Journal of Medical Internet Research

Volume 27 (2025) ISSN 1438-8871

Contents

Editorial

Crisis Text Line and Loris.ai Controversy Highlights the Complexity of Informed Consent on the Internet and Data-Sharing Ethics for Machine Learning and Research (e67878)
Gunther Eysenbach. . . . 37

Reviews

Apps and Digital Resources for Child Neurodevelopment, Mental Health, and Well-Being: Review, Evaluation, and Reflection on Current Resources (e58693)
Kelsie Boulton, Makana Hilton, Emilia Sutton, Adam Guastella. . . . 42

Machine Learning Approaches in High Myopia: Systematic Review and Meta-Analysis (e57644)
Huiyi Zuo, Baoyu Huang, Jian He, Liying Fang, Minli Huang. . . . 56

Social Media Potential and Impact on Changing Behaviors and Actions in Skin Health Promotion: Systematic Review (e54241)
Justyna Brzozowska, Joanna Gotlib. . . . 77

Digital Twins for Clinical and Operational Decision-Making: Scoping Review (e55015)
Vahid Riahi, Ibrahima Diouf, Sankalp Khanna, Justin Boyle, Hamed Hassanzadeh. . . . 94

Effectiveness of Integrated Digital Solutions to Empower Older Adults in Aspects Related to Their Health: Systematic Review and Meta-Analysis (e54466)
Ana Martins, Óscar Ribeiro, Gonçalo Santinha, Telmo Silva, Nelson Rocha, Anabela Silva. . . . 116

Clinical Decision Support Using Speech Signal Analysis: Systematic Scoping Review of Neurological Disorders (e63004)
Upeka De Silva, Samaneh Madanian, Sharon Olsen, John Templeton, Christian Poellabauer, Sandra Schneider, Ajit Narayanan, Rahmina Rubaiat. . . . 148

Enhancing HIV Cognitive Abilities and Self-Management Through Information Technology–Assisted Interventions: Scoping Review (e57363)
Hao Huang, MeiLian Xie, Zhen Yang, AiPing Wang. . . . 176

Participant Contributions to Person-Generated Health Data Research Using Mobile Devices: Scoping Review (e51955)
Shanshan Song, Micaela Ashton, Rebecca Yoo, Zoljargal Lkhagvajav, Robert Wright, Debra Mathews, Casey Taylor. . . . 190

XSL•FO
RenderX

Completeness of Telehealth Interventions Reporting in Randomized Controlled Trials for Caregivers of People With Dementia: Systematic Review (e53737)
Ling Zhu, Yurong Xing, Wenhui Xu, Hongfei Jia, Xiaoxiao Wang, Shiqing Liu, Yaping Ding. . . . 216

Parental Technoference and Child Problematic Media Use: Meta-Analysis (e57636)
Jinghui Zhang, Qing Zhang, Bowen Xiao, Yuxuan Cao, Yu Chen, Yan Li. . . . 229

Digital Health Solutions for Cardiovascular Disease Prevention: Systematic Review (e64981)
Yihan Qi, Emma Mohamad, Arina Azlan, Chenglin Zhang, Yilian Ma, Anqi Wu. . . . 245

Technologies for Interoperable Internet of Medical Things Platforms to Manage Medical Emergencies in Home and Prehospital Care: Scoping Review (e54470)
Mattias Seth, Hoor Jalo, Åsa Högstedt, Otto Medin, Bengt Sjöqvist, Stefan Candefjord. . . . 261

The Use of AI in Mental Health Services to Support Decision-Making: Scoping Review (e63548)
Hassan Auf, Petra Svedberg, Jens Nygren, Monika Nair, Lina Lundgren. . . . 304

Exploring the Use of Smartwatches and Activity Trackers for Health-Related Purposes for Children Aged 5 to 11 years: Systematic Review (e62944)
Lauren Thompson, Sydney Charitos, Jon Bird, Paul Marshall, Amberly Brigden. . . . 325

Rehabilomics Strategies Enabled by Cloud-Based Rehabilitation: Scoping Review (e54790)
Sejun Oh, SangHeon Lee. . . . 345

Implementing Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety in Adults: Systematic Review (e47927)
Daniel Duffy, Derek Richards, Garrett Hisler, Ladislav Timulak. . . . 359

Digital Information Exchange Between the Public and Researchers in Health Studies: Scoping Review (e63373)
Nazli Soltani, Thilo Dietz, Doris Ochterbeck, Jens Dierkes, Katja Restel, Lara Christianson, Karina De Santis, Hajo Zeeb. . . . 378

Effectiveness of Different Intervention Modes in Lifestyle Intervention for the Prevention of Type 2 Diabetes and the Reversion to Normoglycemia in Adults With Prediabetes: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e63975)
Yachen Wang, Xin Chai, Yueqing Wang, Xuejun Yin, Xinying Huang, Qiuhong Gong, Juan Zhang, Ruitai Shao, Guangwei Li. . . . 397

Impact of Online Interactive Decision Tools on Women's Decision-Making Regarding Breast Cancer Screening: Systematic Review and Meta-Analysis (e65974)
Patricia Villain, Laura Downham, Alice Le Bonniec, Charlotte Bauquier, Olena Mandrik, Tom Nadarzynski, Lorie Donelle, Raúl Murillo, Eleni Tolma, Sonali Johnson, Patricia Soler-Michel, Robert Smith. . . . 411

Families' Experiences With Family-Focused Web-Based Interventions for Improving Health: Qualitative Systematic Literature Review (e58774)
Diana Zhu, Aimee Dordevic, Zoe Davidson, Simone Gibson. . . . 436

Metrics for Evaluating Telemedicine in Randomized Controlled Trials: Scoping Review (e67929)
Yuka Sugawara, Yosuke Hirakawa, Masao Iwagami, Ryota Inokuchi, Rie Wakimizu, Masaomi Nangaku. . . . 457

Dynamic Augmented Reality Cues for Telementoring in Minimally Invasive Surgeries: Scoping Review (e63939)
Hawa Hamza, Omar Aboumarzouk, Abdulla Al-Ansari, Nikhil Navkar. . . . 474

XSL•FO
RenderX

Application of Patient-Generated Health Data Among Older Adults With Cancer: Scoping Review (e57379)
Yesol Kim, Geonah Kim, Hyeonmi Cho, Yeonju Kim, Mona Choi. 490

Effectiveness of Virtual Reality Technology Interventions in Improving the Social Skills of Children and Adolescents With Autism: Systematic Review (e60845)
Xipeng Yang, Jinlong Wu, Yudan Ma, Jingxuan Yu, Hong Cao, Aihua Zeng, Rui Fu, Yucheng Tang, Zhanbing Ren. 509

Digital Health Technology Interventions for Improving Medication Safety: Systematic Review of Economic Evaluations (e65546)
Widya Insani, Neily Zakiyah, Irma Puspitasari, Muhammad Permana, Kankan Parmikanti, Endang Rusyaman, Auliya Suwantika. 542

Mapping and Summarizing the Research on AI Systems for Automating Medical History Taking and Triage: Scoping Review (e53741)
Elin Siira, Hanna Johansson, Jens Nygren. 558

Patient-Accessible Electronic Health Records and Information Practices in Mental Health Care Contexts: Scoping Review (e54973)
Timothy Kariotis, Megan Prictor, Kathleen Gray, Shanton Chang. 575

Understanding “Alert Fatigue” in Primary Care: Qualitative Systematic Review of General Practitioners Attitudes and Experiences of Clinical Alerts, Prompts, and Reminders (e62763)
Illin Gani, Ian Litchfield, David Shukla, Gayathri Delanerolle, Neil Cockburn, Anna Pathmanathan. 601

Assessing the Uses, Benefits, and Limitations of Digital Technologies Used by Health Professionals in Supporting Obesity and Mental Health Communication: Scoping Review (e58434)
Amanda Kearns, Anne Moorhead, Maurice Mulvenna, Raymond Bond. 622

Machine Learning in the Management of Patients Undergoing Catheter Ablation for Atrial Fibrillation: Scoping Review (e60888)
Aijing Luo, Wei Chen, Hongtao Zhu, Wenzhao Xie, Xi Chen, Zhenjiang Liu, Zirui Xin. 642

Trends and Gaps in Digital Precision Hypertension Management: Scoping Review (e59841)
Namuun Clifford, Rachel Tunis, Adetimilehin Ariyo, Haoxiang Yu, Hyekyun Rhee, Kavita Radhakrishnan. 663

Interventions for Digital Addiction: Umbrella Review of Meta-Analyses (e59656)
Peng Lu, Jiamin Qiu, Shiqi Huang, Xinman Wang, Shasha Han, Sui Zhu, Youjing Ning, Fang-fang Zeng, Yuan Yuan. 686

Uses of Virtual Care in Primary Care: Scoping Review (e55007)
Payal Agarwal, Glenn Fletcher, Karishini Ramamoorthi, Xiaomei Yao, Onil Bhattacharyya. 701

Assessing the Effectiveness of Digital Health Behavior Strategies on Type 2 Diabetes Management: Systematic Review and Network Meta-Analysis (e63209)
Min Li, Shiyu Liu, Binyang Yu, Ning Li, Aili Lyu, Haiyan Yang, Haiyan He, Na Zhang, Jingru Ma, Meichen Sun, Hong Du, Rui Gao. 724

Simulation-Based Learning Supported by Technology to Enhance Critical Thinking in Nursing Students: Scoping Review (e58744)
Hege Stenseth, Simen Steindal, Marianne Solberg, Mia Ølnes, Anne Sørensen, Camilla Strandell-Laine, Camilla Olaussen, Caroline Farsjø Aure, Ingunn Pedersen, Jaroslav Zlamal, Jussara Gue Martini, Paula Bresolin, Silje Linnerud, Andréa Nes. 751

Virtual Reality Enhanced Exercise Training in Upper Limb Function of Patients With Stroke: Meta-Analytic Study (e66802)
Shiqi Xu, Yanwen Xu, Ruyi Wen, Jun Wang, Yuyu Qiu, Chetwyn Chan. 775

Advancing Health Care With Digital Twins: Meta-Review of Applications and Implementation Challenges (e69544)
Mickaël Ringeval, Faustin Etindele Sosso, Martin Cousineau, Guy Paré. 800

XSL•FO
RenderX

Impact of Patient Online Record Access on Documentation: Scoping Review (e64762)
Eva Meier-Diedrich, Camilla Lyckblad, Gail Davidge, Maria Hägglund, Anna Kharko, Brian McMillan, Charlotte Blease, Julian Schwarz. 812

Digital Interventions for Older People Experiencing Homelessness: Systematic Scoping Review (e63898)
Emily Adams, Eddie Donaghy, Caroline Sanders, Maria Wolters, Lauren Ng, Christa St-Jean, Ryan Galan, Stewart Mercer. 829

Digital Health Interventions to Improve Mental Health in Patients With Cancer: Umbrella Review (e69621)
Chuhan Zhong, Xian Luo, Miaoqin Tan, Jing Chi, Bingqian Guo, Jianyao Tang, Zihan Guo, Shisi Deng, Yujie Zhang, Yanni Wu. 846

Reporting Quality of AI Intervention in Randomized Controlled Trials in Primary Care: Systematic Review and Meta-Epidemiological Study (e56774)
Jinjia Zhong, Ting Zhu, Yafang Huang. 865

Effects of the Modern Digital Information Environment on Maternal Health Care Professionals, the Role of Midwives, and the People in Their Care: Scoping Review (e70108)
Elisabeth Wilhelm, Victoria Vivilaki, Jean Calleja-Agius, Elena Petelos, Maria Tzeli, Paraskevi Giaxi, Elena Triantiafyllou, Eleni Asimaki, Faye Alevizou, Tina Purnat. 880

Features of Structured, One-to-One Videoconference Interventions That Actively Engage People in the Management of Their Chronic Conditions: Scoping Review (e58543)
Yu-Ting Chen, Michelle Lehman, Toni Van Denend, Jacqueline Kish, Yue Wu, Katharine Preissner, Matthew Plow, Tanya Packer. 896

Using Virtual Reality to Improve Outcomes Related to Quality of Life Among Older Adults With Serious Illnesses: Systematic Review of Randomized Controlled Trials (e54452)
Bhagvat Maheta, Alexandra Kraft, Nickolas Interrante, Soraya Fereydooni, Jeremy Bailenson, Brian Beams, Christina Keny, Thomas Osborne, Karleen Giannitrapani, Karl Lorenz. 920

Effectiveness of Digital Diabetes Management Technology on Blood Glucose in Patients With Type 2 Diabetes at Home: Systematic Review and Meta-Analysis (e66441)
Yuping Xiao, Zhenzhen Wang, Lintao Zhang, Nina Xie, Fangyao Chen, Zihao Song, Sha Zhao. 954

Poststroke eHealth Technologies–Based Rehabilitation for Upper Limb Recovery: Systematic Review (e57957)
Margherita Rampioni, Sara Leonzi, Luca Antognoli, Anna Mura, Vera Stara. 967

Enabling Telemedicine From the System-Level Perspective: Scoping Review (e65932)
Xuezhu Li, Lifeng Huang, Hui Zhang, Zhanming Liang. 980

Natural Language Processing Technologies for Public Health in Africa: Scoping Review (e68720)
Songbo Hu, Abigail Oppong, Ebele Mogo, Charlotte Collins, Giulia Occhini, Anna Barford, Anna Korhonen. 997

Facilitators and Barriers to the Implementation of Digital Health Technologies in Hospital Settings in Lower- and Middle-Income Countries Since the Onset of the COVID-19 Pandemic: Scoping Review (e63482)
Sheng Yew, Daksha Trivedi, Nurul Adanan, Boon Chew. 1022

Diagnostic Performance of Artificial Intelligence–Based Methods for Tuberculosis Detection: Systematic Review (e69068)
Seng Hansun, Ahmadreza Argha, Ivan Bakhshayeshi, Arya Wicaksana, Hamid Alinejad-Rokny, Greg Fox, Siaw-Teng Liaw, Branko Celler, Guy Marks. 1044

Application of Machine Learning for Patients With Cardiac Arrest: Systematic Review and Meta-Analysis (e67871)
Shengfeng Wei, Xiangjian Guo, Shilin He, Chunhua Zhang, Zhizhuan Chen, Jianmei Chen, Yanmei Huang, Fan Zhang, Qiangqiang Liu. . . .
1 0 6 7

XSL•FO
RenderX

Factors Influencing eHealth Literacy Worldwide: Systematic Review and Meta-Analysis (e50313)
Zhong Hua, Song Yuqing, Liu Qianwen, Chen Hong. . . . 1096

The State of the Art of eHealth Self-Management Interventions for People With Chronic Obstructive Pulmonary Disease: Scoping Review (e57649)
Eline te Braake, Roswita Vaseur, Christiane Grünloh, Monique Tabak. . . . 1117

Intelligent Robot Interventions for People With Dementia: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e59892)
Wenqi Fan, Rui Zhao, Xiaoxia Liu, Lina Ge. . . . 1140

Examining Challenges to Co-Design Digital Health Interventions With End Users: Systematic Review (e50178)
Anthony Duffy, Nazanin Boroumandzad, Alfredo Sherman, Gregory Christie, Indira Riadi, Sylvain Moreno. . . . 1154

Association of Digital Health Interventions With Maternal and Neonatal Outcomes: Systematic Review and Meta-Analysis (e66580)
Jianing Wang, Nu Tang, Congcong Jin, Jianxue Yang, Xiangpeng Zheng, Qiujing Jiang, Shengping Li, Nian Xiao, Xiaojun Zhou. . . . 1186

Digital Wellness Programs in the Workplace: Meta-Review (e70982)
Saeed Amirabdolahian, Guy Pare, Stefan Tams. . . . 1207

Digital Physical Activity and Sedentary Behavior Interventions for Community-Living Adults: Umbrella Review (e66294)
Eilidh Russell, Alison Kirk, Mark Dunlop, William Hodgson, Mhairi Patience, Kieren Egan. . . . 1236

Enhancing Patient Outcome Prediction Through Deep Learning With Sequential Diagnosis Codes From Structured Electronic Health Record Data: Systematic Review (e57358)
Tuankasfee Hama, Mohanad Alsaleh, Freya Allery, Jung Choi, Christopher Tomlinson, Honghan Wu, Alvina Lai, Nikolas Pontikos, Johan Thygesen. . . . 1253

Patients' Experiences of Digital Health Interventions for the Self-Management of Chronic Pain: Systematic Review and Thematic Synthesis (e69100)
Ashleigh Main, Haruno McCartney, Maryam Ibrar, Fiona Muirhead, Alexandra Mavroeidi, Harleen Rai, Roma Maguire. . . . 1274

Digital Health Education for Chronic Lung Disease: Scoping Review (e53142)
Chao Sun, Huohuo Dai, Rianne M.J.J. van der Kleij, Rong Li, Hengchang Wu, Cynthia Hallensleben, Sofie Willems, Niels Chavannes. . . . 1294

Behavioral Therapy–Based Digital Interventions for Treating Osteoarthritis: Systematic Review and Meta-Analysis (e56227)
Beiyao Zhu, Dian Zhu, Xiao'ao Xue, Hongyi Yang, Shurong Zhang. . . . 1308

Effectiveness of Digital Lifestyle Interventions on Depression, Anxiety, Stress, and Well-Being: Systematic Review and Meta-Analysis (e56975)
Jacinta Brinsley, Edward O'Connor, Ben Singh, Grace McKeon, Rachel Curtis, Ty Ferguson, Georgia Gosse, Iris Willems, Pieter-Jan Marent, Kimberley Szeto, Joseph Firth, Carol Maher. . . . 1330

AI-Derived Blood Biomarkers for Ovarian Cancer Diagnosis: Systematic Review and Meta-Analysis (e67922)
He-Li Xu, Xiao-Ying Li, Ming-Qian Jia, Qi-Peng Ma, Ying-Hua Zhang, Fang-Hua Liu, Ying Qin, Yu-Han Chen, Yu Li, Xi-Yang Chen, Yi-Lin Xu, Dong-Run Li, Dong-Dong Wang, Dong-Hui Huang, Qian Xiao, Yu-Hong Zhao, Song Gao, Xue Qin, Tao Tao, Ting-Ting Gong, Qi-Jun Wu. . . .
1 3 6 2

Facilitators and Barriers to Digital Mental Health Interventions for Depression, Anxiety, and Stress in Adolescents and Young Adults: Scoping Review (e62870)
Shimin Zhu, Yongyi Wang, Yuxi Hu. . . . 1387

XSL•FO
RenderX

Effects of Digital Intelligent Interventions on Self-Management of Patients With Diabetic Foot: Systematic Review (e64400)
Jinyan Zhou, Shanni Ding, Yihong Xu, Hongying Pan. . . . 1408

Physiological Sensors Equipped in Wearable Devices for Management of Long COVID Persisting Symptoms: Scoping Review (e69506)
Shikha Kukreti, Meng-Ting Lu, Chun-Yin Yeh, Nai-Ying Ko. . . . 1426

Efficacy of eHealth Interventions for Hemodialysis Patients: Systematic Review and Meta-Analysis (e67246)
Xu-Hua Zhou, Hui Chen, Weiwei Yang, Li Wang, Lin Chen, Ying Zhu, Yingjun Zhang, Mei Shi, Qin Zhang. . . . 1435

Social Media and eHealth Literacy Among Older Adults: Systematic Literature Review (e66058)
Chenglin Zhang, Emma Mohamad, Arina Azlan, Anqi Wu, Yilian Ma, Yihan Qi. . . . 1450

Use of Mukbang in Health Promotion: Scoping Review (e56147)
Xiao Wang, Yuxue Xiao, Sujin Nam, Ting Zhong, Dongyan Tang, William Li, Peige Song, Wei Xia. . . . 1483

Use of mHealth Technology for Improving Exercise Adherence in Patients With Heart Failure: Systematic Review (e54524)
Pallav Deka, Erin Salahshurian, Teresa Ng, Susan Buchholz, Leonie Klompstra, Windy Alonso. . . . 1812

Evidence on Digital HIV Self-Testing From Accuracy to Impact: Updated Systematic Review (e63110)
Ashlyn Beecroft, Olivia Vaikla, Nora Engel, Thomas Duchaine, Chen Liang, Nitika Pant Pai. . . . 2204

Embedding a Choice Experiment in an Online Decision Aid or Tool: Scoping Review (e59209)
Nyantara Wickramasekera, Phil Shackley, Donna Rowen. . . . 2294

Contribution of Open Access Databases to Intensive Care Medicine Research: Scoping Review (e57263)
Julien Kallout, Antoine Lamer, Julien Grosjean, Gaétan Kerdelhué, Guillaume Bouzillé, Thomas Clavier, Benjamin Popoff. . . . 2716

Early Intervention for Children With Developmental Disabilities and Their Families via Telehealth: Systematic Review (e66442)
Yoomi Shin, Eun Park, Anna Lee. . . . 3356

Use of Remote Assessment Tools to Substitute Routine Outpatient Care: Scoping Review (e65938)
Jayne Barclay, Clair Sullivan, Michael Beckmann, Graeme Mattison, Rebecca Runciman, Elizabeth Martin. . . . 3498

Shared Decision-Making Tools Implemented in the Electronic Health Record: Scoping Review (e59956)
Joni Pierce, Charlene Weir, Teresa Taft, William Richards II, Mary McFarland, Kensaku Kawamoto, Guilherme Del Fiol, Jorie Butler. . . . 3939

Identifying Research Priorities in Digital Education for Health Care: Umbrella Review and Modified Delphi Method Study (e66157)
Alison Potter, Chris Munsch, Elaine Watson, Emily Hopkins, Sofia Kitromili, Iain O'Neill, Judy Larbie, Essi Niittymaki, Catriona Ramsay, Joshua Burke, Neil Ralph. . . . 4132

Machine Learning for the Early Prediction of Delayed Cerebral Ischemia in Patients With Subarachnoid Hemorrhage: Systematic Review and Meta-Analysis (e54121)
Haofuzi Zhang, Peng Zou, Peng Luo, Xiaofan Jiang. . . . 4366

Online-Based Recruitment Methods for Community-Dwelling Older Adults: Scoping Review and Lessons Learned From the PLAN Trial (e55082)
Deborah Min, Ji-Young Yun, Chad Parslow, Anushka Jajodia, Hae-Ra Han. . . . 4692

XSL•FO
RenderX

Effectiveness of Virtual Reality Training in Improving Outcomes for Dialysis Patients: Systematic Review and Meta-Analysis (e58384)
Xin Kang, Yiping Zhang, Chaonan Sun, Jiaxin Zhang, Zhe Che, Jinhui Zang, Rongzhi Zhang. 4882

Extended Reality Interventions for Health and Procedural Anxiety: Panoramic Meta-Analysis Based on Overviews of Reviews (e58086)
Tom Arthur, GJ Melendez-Torres, David Harris, Sophie Robinson, Mark Wilson, Sam Vine. 4897

Use of Augmented Reality for Training Assistance in Laparoscopic Surgery: Scoping Literature Review (e58108)
Francisco Celdrán, Javier Jiménez-Ruescas, Carlos Lobato, Lucía Salazar, Juan Sánchez-Margallo, Francisco Sánchez-Margallo, Pascual González. 4943

Virtual Reality Interventions and Chronic Pain: Scoping Review (e59922)
Michael Ding, Hajar Traiba, Hector Perez. 4963

Current Clinical and Educational Uses of Immersive Reality in Anesthesia: Narrative Review (e62785)
Andrew Fleet, Lilia Kaustov, Elio Belfiore, Bill Kapralos, Clyde Matava, Julian Wiegelmann, Peter Giacobbe, Fahad Alam. 4972

Bias Mitigation in Primary Health Care Artificial Intelligence Models: Scoping Review (e60269)
Maxime Sasseville, Steven Ouellet, Caroline Rhéaume, Malek Sahlia, Vincent Couture, Philippe Després, Jean-Sébastien Paquette, David Darmon, Frédéric Bergeron, Marie-Pierre Gagnon. 5384

Transformers for Neuroimage Segmentation: Scoping Review (e57723)
Maya Iratni, Amira Abdullah, Mariam Aldhaheri, Omar Elharrouss, Alaa Abd-alrazaq, Zahiriddin Rustamov, Nazar Zaki, Rafat Damseh. 5463

AI for IMPACTS Framework for Evaluating the Long-Term Real-World Impacts of AI-Powered Clinician Tools: Systematic Review and Narrative Synthesis (e67485)
Christine Jacob, Noé Brasier, Emanuele Laurenzi, Sabina Heuss, Stavroula-Georgia Mougiakakou, Arzu Cöltekin, Marc Peter. 5512

Artificial Intelligence in Lymphoma Histopathology: Systematic Review (e62851)
Yao Fu, Zongyao Huang, Xudong Deng, Linna Xu, Yang Liu, Mingxing Zhang, Jinyi Liu, Bin Huang. 5566

Effectiveness of AI for Enhancing Computed Tomography Image Quality and Radiation Protection in Radiology: Systematic Review and Meta-Analysis (e66622)
Subo Zhang, Zhitao Zhu, Zhenfei Yu, Haifeng Sun, Yi Sun, Hai Huang, Lei Xu, Jinxin Wan. 5606

Challenges and Opportunities for Data Sharing Related to Artificial Intelligence Tools in Health Care in Low- and Middle-Income Countries: Systematic Review and Case Study From Thailand (e58338)
Aprajita Kaushik, Capucine Barcellona, Nikita Mandyam, Si Tan, Jasper Tromp. 5935

Effect of eHealth Interventions on Body Image of Patients With Cancer: Systematic Review (e55564)
Guohong Huang, Rongrong Wu, Xiuzhi Xu, Yongxia Song, Rong Zheng, Xi Chen, Jingfang Hong. 6130

Applications of Large Language Models in the Field of Suicide Prevention: Scoping Review (e63126)
Glenn Holmes, Biya Tang, Sunil Gupta, Svetha Venkatesh, Helen Christensen, Alexis Whitton. 6554

Twenty-Five Years of Evolution and Hurdles in Electronic Health Records and Interoperability in Medical Research: Comprehensive Review (e59024)
Yun Shen, Jiamin Yu, Jian Zhou, Gang Hu. 6683

Mapping the Landscape of Digital Health Intervention Strategies: 25-Year Synthesis (e59027)
Shiyu Liu, Jingru Ma, Meichen Sun, Chao Zhang, Yujing Gao, Jinghong Xu. 6720

XSL•FO
RenderX

Digital Age Transformation in Patient-Physician Communication: 25-Year Narrative Review (1999-2023) (e60512)
Mingming Song, Joel Elson, Dhundy Bastola. . . . 6757

25 Years of Electronic Health Record Implementation Processes: Scoping Review (e60077)
Harriet Finnegan, Nicola Mountford. . . . 6777

Generative AI Models in Time-Varying Biomedical Data: Scoping Review (e59792)
Rosemary He, Varuni Sarwal, Xinru Qiu, Yongwen Zhuang, Le Zhang, Yue Liu, Jeffrey Chiang. . . . 6792

Original Papers

Virtual Reality Interventions for Older Adults With Mild Cognitive Impairment: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e59195)
Qin Yang, Liuxin Zhang, Fangyuan Chang, Hongyi Yang, Bin Chen, Zhao Liu. . . . 128

Effect of Multimodal App-Based Interventions on Glycemic Control in Patients With Type 2 Diabetes: Systematic Review and Meta-Analysis (e54324)
Emma Bodner, Lena Roth, Kathleen Wiencke, Christian Bischoff, Peter Schwarz. . . . 287

Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis (e62647)
Vivens Mubonanyikuzo, Hongjie Yan, Temitope Komolafe, Liang Zhou, Tao Wu, Nizhuan Wang. . . . 526

Virtual Reality Enhanced Exercise Training in Upper Limb Function of Patients With Stroke: Meta-Analytic Study (e66802)
Shiqi Xu, Yanwen Xu, Ruyi Wen, Jun Wang, Yuyu Qiu, Chetwyn Chan. . . . 775

Unpacking Performance Factors of Innovation Systems and Studying Germany's Attempt to Foster the Role of the Patient Through a Market Access Pathway for Digital Health Applications (DiGAs): Exploratory Mixed Methods Study (e66356)
Sara Gehder, Moritz Goeldner. . . . 1505

Association Between Digital Front Doors and Social Care Use for Community-Dwelling Adults in England: Cross-Sectional Study (e53205)
Jinbao Zhang, Jonathan Prunty, Alison Charles, Julien Forder. . . . 1802

Assessing the Effectiveness of Interactive Robot-Assisted Virtual Health Coaching for Health Literacy and Disease Knowledge of Patients with Chronic Kidney Disease: Quasiexperimental Study (e68072)
Nai-Jung Chen, Ching-Hao Chang, Chiu-Mieh Huang, Fen-He Lin, Li-Ting Lu, Kuan-Yi Liu, Chih-Lin Lai, Chin-Yao Lin, Yi-Chou Hou, Jong-Long Guo. . . . 1827

Studying the Digital Intervention Engagement–Mediated Relationship Between Intrapersonal Measures and Pre-Exposure Prophylaxis Adherence in Sexual and Gender Minority Youth: Secondary Analysis of a Randomized Controlled Trial (e57619)
Michael Williams, Justin Manjourides, Louisa Smith, Crissi Rainer, Lisa Hightow-Weidman, Danielle Haley. . . . 1839

Effectiveness and Safety of the TRIO Optimal Health Management Program in Patients With Type 2 Diabetes Mellitus Initiating Basal Insulin Therapy: Prospective Observational Real-World Study (e67554)
Chenxi Li, Lixin Guo, Lixin Shi, Li Chen, Liming Chen, Yaoming Xue, Hong Li, Yuzhen Liang, Jing Yang, Weimin Wang, Dalong Zhu. . . . 1858

XSL•FO
RenderX

Eliciting Preferences for the Uptake of Smoking Cessation Apps: Discrete Choice Experiment (e37083)
Dorothy Szinay, Rory Cameron, Andy Jones, Jennifer Whitty, Tim Chadborn, Jamie Brown, Felix Naughton. 1880

Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study (e65122)
Gilly Hendrie, Danielle Baird, Genevieve James-Martin, Emily Brindal, Paige Brooker. 1896

Development and Evaluation of a Mindfulness-Based Mobile Intervention for Perinatal Mental Health: Randomized Controlled Trial (e56601)
Sehwan Park, Hee Cho, Jin Park, Kyungmi Chung, Kyungun Jhung. 1913

Improving Outcomes and Quality of Life for Patients With Hand and Foot Eczema: Randomized Study of a Patient-Centered Monitoring App (e52159)
Aimee Bruch, Wanja Weigandt, Yannic Schardt, Raphael Herr, Johannes Benecke, Astrid Schmieder. 1931

Effectiveness of a Web-Based Medication Education Course on Pregnant Women's Medication Information Literacy and Decision Self-Efficacy: Randomized Controlled Trial (e54148)
Suya Li, Hui-Jun Chen, Jie Zhou, Yi-Bei Zhouchen, Rong Wang, Jinyi Guo, Sharon Redding, Yan-Qiong Ouyang. 1949

A Novel Artificial Intelligence–Enhanced Digital Network for Prehospital Emergency Support: Community Intervention Study (e58177)
Ji Kim, Min Kim, Hyeon Kim, Ha Kim, Ji Sung, Hyuk-Jae Chang. 1960

Patient Experiences and Perspectives When MyChart is Introduced in a Large Community Hospital: Mixed Methods Study (e66353)
Shelley Vanderhout, Shipra Taneja, Kamini Kalia, Walter Wodchis, Terence Tang. 1975

A 4-Week Mobile App–Based Telerehabilitation Program vs Conventional In-Person Rehabilitation in Older Adults With Sarcopenia: Randomized Controlled Trial (e67846)
Lu Zhang, Ying Ge, Wowa Zhao, Xuan Shu, Lin Kang, Qiumei Wang, Ying Liu. 1986

Investigating the Efficacy of the Web-Based Common Elements Toolbox (COMET) Single-Session Interventions in Improving UK University Student Well-Being: Randomized Controlled Trial (e58164)
Jeffrey Lambert, Maria Loades, Noah Marshall, Nina Higson-Sweeney, Stella Chan, Arif Mahmud, Victoria Pile, Ananya Maity, Helena Adam, Beatrice Sung, Melanie Luximon, Keren MacLennan, Clio Berry, Paul Chadwick. 2001

Examining Health Care Provider Experiences With Patient Portal Implementation: Mixed Methods Study (e65967)
Shipra Taneja, Kamini Kalia, Terence Tang, Walter Wodchis, Shelley Vanderhout. 2019

Smartphone-Based Intervention Targeting Norms and Risk Perception Among University Students with Unhealthy Alcohol Use: Secondary Mediation Analysis of a Randomized Controlled Trial (e55541)
Joseph Studer, John Cunningham, Elodie Schmutz, Jacques Gaume, Angéline Adam, Jean-Bernard Daeppen, Nicolas Bertholet. 2030

Efficacy of a Blended Low-Intensity Internet-Delivered Psychological Program in Patients With Multimorbidity in Primary Care: Randomized Controlled Trial (e56203)
Alicia Monreal-Bartolomé, Adoración Castro, M Pérez-Ara, Margalida Gili, Fermín Mayoral, María Hurtado, Esperanza Varela Moreno, Cristina Botella, Azucena García-Palacios, Rosa Baños, Yolanda López-Del-Hoyo, Javier García-Campayo, Jesus Montero-Marin. 2046

Effects of Individualized Follow-Up With an App Postcardiac Rehabilitation: Five-Year Follow-Up of a Randomized Controlled Trial (e60256)
Pernille Lunde, Asta Bye, Jostein Grimsmo, Are Pripp, Vibeke Ritschel, Even Jarstad, Birgitta Nilsson. 2070

XSL•FO
RenderX

Evaluating the Clinical Efficacy of an Exergame-Based Training Program for Enhancing Physical and Cognitive Functions in Older Adults With Mild Cognitive Impairment and Dementia Residing in Rural Long-Term Care Facilities: Randomized Controlled Trial (e69109)
Aoyu Li, Wei Qiang, Jingwen Li, Yan Geng, Yan Qiang, Juanjuan Zhao. . . . 2084

Cost Utility Analysis of Internet-Based Cognitive Behavioral Therapy for Major Depressive Disorder: Randomized Controlled Trial (e67567)
Wenjing Zhou, Yan Chen, Herui Wu, Hao Zhao, Yanzhi Li, Guangduoji Shi, Wanxin Wang, Yifeng Liu, Yuhua Liao, Huimin Zhang, Caihong Gao, Jiejing Hao, Gia Le, Roger McIntyre, Xue Han, Ciyong Lu. . . . 2099

Evaluation of the Quality of Delirium Website Content for Patient and Family Education: Cross-Sectional Study (e53087)
Karla Krewulak, Kathryn Strayer, Natalia Jaworska, Krista Spence, Nadine Foster, Scotty Kupsch, Khara Sauro, Kirsten Fiest. . . . 2120

Behavior Change Support Systems for Self-Treating Procrastination: Systematic Search in App Stores and Analysis of Motivational Design Archetypes (e65214)
Jeanine Kirchner-Krath, Manuel Schmidt-Kraepelin, Katharina Schmähl, Christoph Schütz, Benedikt Morschheuser, Ali Sunyaev. . . . 2132

Improving Maternal Mental Health and Weight Control With a Mindfulness Blended Care Approach: Insights From a Randomized Controlled Trial (e56230)
Kathrin Hassdenteufel, Mitho Müller, Harald Abele, Sara Brucker, Johanna Graf, Stephan Zipfel, Armin Bauer, Peter Jakubowski, Jan Pauluschke-Fröhlich, Markus Wallwiener, Stephanie Wallwiener. . . . 2153

Impact of In-Person and Mobile Exercise Coaching on Psychosocial Factors Affecting Exercise Adherence in Inactive Women With Obesity: 20-Week Randomized Controlled Trial (e68462)
Christina Gjestvang, John Kalhovde, Elene Mauseth Tangen, Hege Clemm, Lene Haakstad. . . . 2174

Web-Based Application for Reducing Methamphetamine Use Among Aboriginal and Torres Strait Islander People: Randomized Waitlist Controlled Trial (e58341)
Rachel Reilly, Rebecca McKetin, Federica Barzi, Tayla Degan, Nadine Ezard, Katherine Conigrave, Julia Butt, Yvette Roe, Handan Wand, Brendan Quinn, Wade Longbottom, Carla Treloar, Adrian Dunlop, James Ward. . . . 2191

Utilization and Experiences of Using Quit Now, a Nicotine and Tobacco Smoking Cessation Website: Thematic Analysis (e55592)
Tala Salaheddin, Ramona Sharma, Marcela Fajardo, Cameron Panter, Lauren De Souza, Sheila Matano, Laura Struik. . . . 2220

Internet-Based Cognitive Behavioral Therapy for Preventing Postpartum Depressive Symptoms Among Pregnant Individuals With Depression: Multicenter Randomized Controlled Trial in China (e67386)
Chen-Chi Duan, Chen Zhang, Hua-Lin Xu, Jing Tao, Jia-Le Yu, Dan Zhang, Shan Wu, Xiu Zeng, Wan-Ting Zeng, Zhi-Yin Zhang, Cindy-Lee Dennis, Han Liu, Jia-Ying Wu, Ben Mol, He-Feng Huang, Yan-Ting Wu. . . . 2231

Effects of a Mobile Health Intervention Based on Behavioral Integrated Model on Cognitive and Behavioral Changes in Gestational Weight Management: Randomized Controlled Trial (e55844)
Meng Zhou, Li Wang, Ying Deng, Jinjin Ge, Shiqi Zhao, Hua You. . . . 2247

Coach-Assisted eHealth With Group or Individual Support for Employees With Obesity: Randomized Controlled Trial on Weight, Body Composition, and Health Metrics (e60436)
Siniriikka Männistö, Kirsi Pietiläinen, Joona Muotka, Laura-Unnukka Suojanen, Raimo Lappalainen, Riitta Korpela. . . . 2263

Evaluating Perceptions of the CANreduce 2.0 eHealth Intervention for Cannabis Use: Focus Group Study (e65025)
Daniel Folch-Sanchez, Maria Pellicer-Roca, María Sestelo, Paola Zuluaga, Francisco Arias, Pablo Guzmán Cortez, Salma Amechat, Gustavo Gil-Berrozpe, Estefania Lopez Montes, Clara Mercadé, Francina Fonseca, Laia Miquel, Joan Mestre-Pintó. . . . 2279

XSL•FO
RenderX

Acceptability and Effectiveness of a Fully Web-Based Nutrition and Exercise Program for Individuals With Chronic Disease During COVID-19: Randomized Controlled Trial (e57537)
Puneeta Tandon, Kathleen Ismond, Graeme Purdy, Christofer Cruz, Evelyn Etruw, Kirsten Suderman, Ashley Hyde, Michael Stickland, John Spence, Dale Lien, Rahima Bhanji, Carla Prado, Antonio Miguel-Cruz, Anil Joy, Maryna Yaskina, Margaret McNeely. 2316

Exploring the Discontinuous Usage Behavior of Digital Cognitive Training Among Older Adults With Mild Cognitive Impairment and Their Family Members: Qualitative Study Using the Extended Model of IT Continuance (e66393)
Shangyang Zhang, Min Wu, Ruini Sun, Changjie Cui, Ziqing Zhang, Jing Liao, Ni Gong. 2331

Evaluating a Digital Health Tool Designed to Improve Low Sexual Desire in Women: Mixed-Methods Implementation Science Study (e69828)
Lori Brotto, Kyle Stephenson, Nisha Marshall, Mariia Balvan, Yaroslava Okara, Elizabeth Mahar. 2347

Effect of Clinic-Based and Asynchronous Video-Based Exercise on Clinic and Psychosocial Outcomes in Patients With Knee Osteoarthritis: Quasi-Experimental Study (e58393)
Chidozie Mbada, Henry Awosika, Oluwatobi Sonuga, Micheal Akande, Tadesse Gebrye, Richard Woolf, Francis Fatoye. 2362

Engaging End Users to Inform the Design and Social Marketing Strategy for a Web-Based Sexually Transmitted Infection/Blood-Borne Virus (STI/BBV) Testing Service for Young People in Victoria, Australia: Qualitative Study (e63822)
Ethan Cardwell, Teralynn Ludwick, Shanton Chang, Olivia Walsh, Megan Lim, Rachel Podbury, David Evans, Christopher Fairley, Fabian Kong, Jane Hocking. 2379

Effectiveness of an mHealth- and School-Based Health Education Program for Salt Reduction (EduSaltS) in China: Cluster Randomized Controlled Trial Within Scale-Up (e60092)
Naibo Wang, Chen Wang, Puhong Zhang, Yinghua Li, Feng He, Li Li, Yuan Li, Rong Luo, Dezhi Wan, Lewei Xu, Lifang Deng, Lei Wu. 2390

Efficacy and Acceptability of a Mobile App for Monitoring the Clinical Status of Patients With Chronic Obstructive Pulmonary Disease Receiving Home Oxygen Therapy: Randomized Controlled Trial (e65888)
Anisbed Naranjo-Rojas, Luis Perula-de Torres, Freiser Cruz-Mosquera, Guillermo Molina-Recio. 2406

Digital Frequency Customized Relieving Sound for Chronic Subjective Tinnitus Management: Prospective Controlled Study (e60150)
Dongmei Tang, Yuzhu Peng, Dantong Gu, Yongzhen Wu, Huawei Li. 2422

Efficacy of a Mobile Health–Based Behavioral Treatment for Lifestyle Modification in Type 2 Diabetes Self-Management: Greenhabit Randomized Controlled Trial (e58319)
Ana Ruiz-Leon, Rosa Casas, Sara Castro-Barquero, Sofia Alfaro-González, Petia Radeva, Emilio Sacanella, Francesc Casanovas-Garriga, Ainhoa Pérez-Gesalí, Ramon Estruch. 2435

Long-Term Impact of Using Mobile Phones and Playing Computer Games on the Brain Structure and the Risk of Neurodegenerative Diseases: Large Population-Based Study (e59663)
Yi Xiao, Sirui Zhang, Yuanzheng Ma, Shichan Wang, Chunyu Li, Yan Liang, Huifang Shang. 2452

Cross-Platform Ecological Momentary Assessment App (JTrack-EMA+): Development and Usability Study (e51689)
Mehran Sahandi Far, Jona Fischer, Svea Senge, Robin Rathmakers, Thomas Meissner, Dominik Schneble, Mamaka Narava, Simon Eickhoff, Juergen Dukart. 2467

Relationship Among Macronutrients, Dietary Components, and Objective Sleep Variables Measured by Smartphone Apps: Real-World Cross-Sectional Study (e64749)
Jaehoon Seol, Masao Iwagami, Megane Kayamare, Masashi Yanagisawa. 2476

Oura Ring as a Tool for Ovulation Detection: Validation Analysis (e60667)
Nina Thigpen, Shyamal Patel, Xi Zhang. 2489

XSL•FO
RenderX

Remote Monitoring of Chemotherapy-Induced Peripheral Neuropathy by the NeuroDetect iOS App: Observational Cohort Study of Patients With Cancer (e65615)
Ciao-Sin Chen, Michael Dorsch, Sarah Alsomairy, Jennifer Griggs, Reshma Jagsi, Michael Sabel, Amro Stino, Brian Callaghan, Daniel Hertz. 2 4 9 9

App-Based Ecological Momentary Assessment of Problematic Smartphone Use During Examination Weeks in University Students: 6-Week Observational Study (e69320)
Ji Ahn, InJi Jeong, Sehwan Park, Jooho Lee, Minjeong Jeon, Sangil Lee, Gangho Do, Dooyoung Jung, Jin Park. 2513

Technological-Based Interventions in Cancer and Factors Associated With the Use of Mobile Digital Wellness and Health Apps Among Cancer Information Seekers: Cross-Sectional Study (e63403)
Ogochukwu Ezeigwe, Kenechukwu Nwosu, Oladipo Afolayan, Akpevwe Ojaruega, Jovita Echere, Manali Desai, Modupe Onigbogi, Olajumoke Oladoyin, Nnenna Okoye, Pierre Fwelo. 2532

Good-Quality mHealth Apps for Endometriosis Care: Systematic Search (e49654)
Diksha Sirohi, Cecilia Ng, Niranjan Bidargaddi, Helen Slater, Melissa Parker, M Hull, Rebecca O'Hara. 2544

Smartphone-Based Telemonitoring for Better Oral Health With Toothbrushes: 6-Month Randomized Controlled Trial (e65128)
Jaeyeon Kim, Yiseul Choi, Yoolbin Song, Wonse Park. 2556

Smart Pharmaceutical Monitoring System With Personalized Medication Schedules and Self-Management Programs for Patients With Diabetes: Development and Evaluation Study (e56737)
Jian Xiao, Mengyao Li, Ruwen Cai, Hangxing Huang, Huimin Yu, Ling Huang, Jingyang Li, Ting Yu, Jiani Zhang, Shuqiao Cheng. 2570

Smartphone Application–Based Voice and Speech Training Program for Parkinson Disease: Feasibility and Satisfaction Study With a Preliminary Rater-Blinded Single-Arm Pretest and Posttest Design (e63166)
Sol-Hee Lee, Jiae Kim, Han-Joon Kim. 2582

Continuous Improvement of Chronic Tinnitus Through a 9-Month Smartphone-Based Cognitive Behavioral Therapy: Randomized Controlled Trial (e59575)
Uso Walter, Stefan Pennig, Lothar Bleckmann, Kristina Röschmann-Doose, Thomas Wittig, Jörn Thomsen, Winfried Schlee. 2594

Depression Self-Care Apps' Characteristics and Applicability to Older Adults: Systematic Assessment (e56418)
Ruoyu Yin, Dakshayani Rajappan, Laura Martinengo, Frederick Chan, Helen Smith, Konstadina Griva, Mythily Subramaniam, Lorainne Tudor Car. 2606

Analysis of Factors Influencing the Willingness of Chinese Older Adults to Use mHealth Devices: Nationwide Cross-Sectional Survey Study (e66804)
Mengyao Yan, Wendi Sun, Cheng Tan, Yibo Wu, Yuanli Liu. 2624

Exploring Heart Disease–Related mHealth Apps in India: Systematic Search in App Stores and Metadata Analysis (e53823)
Keerthi Dubbala, Roshan Prizak, Ingrid Metzler, Giovanni Rubeis. 2641

Targeted Behavior Change Communication Using a Mobile Health Platform to Increase Uptake of Long-Lasting Insecticidal Nets Among Pregnant Women in Tanzania: Hati Salama "Secure Voucher" Study Cluster Randomized Controlled Trial (e51524)
Trinity Vey, Eleonora Kinnicutt, Andrew Day, Nicola West, Jessica Sleeth, Kenneth Nchimbi, Karen Yeates. 2660

Spatiotemporal and Behavioral Patterns of Men Who Have Sex With Men Using Geosocial Networking Apps in Shenzhen From Mobile Big Data Perspective: Longitudinal Observational Study (e69569)
Lan Wei, Yongsheng Wu, Lin Chen, Jinquan Cheng, Jin Zhao. 2675

XSL•FO
RenderX

Toward Unsupervised Capacity Assessments for Gait in Neurorehabilitation: Validation Study (e66123)
Aileen Naef, Guichande Duarte, Saskia Neumann, Migjen Shala, Meret Branscheidt, Chris Easthope Awai. 2691

Bridging the Gap in Carbohydrate Counting With a Mobile App: Needs Assessment Survey (e63278)
Asmaa Housni, Alexandra Katz, Lucien Bergeron, Alain Simard, Ashley Finkel, Amélie Roy-Fleming, Meranda Nakhla, Anne-Sophie Brazeau.
2 7 0 4

Analysis of Reddit Discussions on Motivational Factors for Physical Activity: Cross-Sectional Study (e54489)
Michal Shmueli-Scheuer, Yedidya Silverman, Israel Halperin, Yftach Gepner. 2732

Analysis of “Dr Ding Xiang” on WeChat in China to Determine Factors Influencing Readership on Medical Social Media: Observational Study (e65372)
Jiaman Liao, Xueliang Huang, Hao Huang, Cuina Shen, Lixia Li, Yushao Li, Yiqiang Zhan. 2744

Organ Donation Conversations on X and Development of the OrgReach Social Media Marketing Strategy: Social Network Analysis (e59872)
Wasim Ahmed, Mariann Hardey, Josep Vidal-Alaball. 2756

Understanding Human Papillomavirus Vaccination Hesitancy in Japan Using Social Media: Content Analysis (e68881)
Junyu Liu, Qian Niu, Momoko Nagai-Tanima, Tomoki Aoyama. 2766

Traversing Shifting Sands—the Challenges of Caring for Someone With Alzheimer's Disease and the Impact on Care Partners: Social Media Content Analysis (e55468)
Kristian Frederiksen, Julie Hahn-Pedersen, Rebecca Crawford, Ross Morrison, Rose Jeppesen, Lynda Doward, Wendy Weidner. 2783

Understanding Challenges and Emotions of Informal Caregivers of General Older Adults and People With Alzheimer Disease and Related Dementia: Comparative Study (e54847)
Nova Huang, Liang Wong, Shirley Ho, Bryan Timothy. 2793

Living With and Managing Uncomplicated Urinary Tract Infection: Mixed Methods Analysis of Patient Insights From Social Media (e58882)
Melissa Kramer, Jose Polo, Nishant Kumar, Aruni Mulgirigama, Amina Benkiran. 2812

Promoting Public Engagement in Palliative and End-of-Life Care Discussions on Chinese Social Media: Model Development and Analysis (e59944)
Yijun Wang, Han Zheng, Yuxin Zhou, Emeka Chukwusa, Jonathan Koffman, Vasa Curcin. 2825

The Association Between Internet Addiction and the Risk of Suicide Attempts in Chinese Adolescents Aged 11-17 Years: Prospective Cohort Study (e52083)
Sihong Li, Xingyue Jin, Lintong Song, Tianqing Fan, Yanmei Shen, Jiansong Zhou. 2843

Trends in Mental Health Outcomes of College Students Amid the Pandemic (Roadmap mHealth App): Longitudinal Observational Study (e67627)
Gautham Jayaraj, Xiao Cao, Adam Horwitz, Michelle Rozwadowski, Skyla Shea, Shira Hanauer, David Hanauer, Muneesh Tewari, Kerby Shedden, Sung Choi. 2853

Therapeutic Potential of Social Chatbots in Alleviating Loneliness and Social Anxiety: Quasi-Experimental Mixed Methods Study (e65589)
Myungsung Kim, Seonmi Lee, Sieun Kim, Jeong-in Heo, Sangil Lee, Yu-Bin Shin, Chul-Hyun Cho, Dooyoung Jung. 2872

The Impact of Linguistic Signals on Cognitive Change in Support Seekers in Online Mental Health Communities: Text Analysis and Empirical Study (e60292)
Min Li, Dongxiao Gu, Rui Li, Yadi Gu, Hu Liu, Kaixiang Su, Xiaoyu Wang, Gongrang Zhang. 2886

XSL•FO
RenderX

Explainable Predictive Model for Suicidal Ideation During COVID-19: Social Media Discourse Study (e65434)
Salah Bouktif, Akib Khanday, Ali Ouni. 2902

Investigating Smartphone-Based Sensing Features for Depression Severity Prediction: Observation Study (e55308)
Yannik Terhorst, Eva-Maria Messner, Kennedy Opoku Asare, Christian Montag, Christopher Kannen, Harald Baumeister. 2916

Effectiveness of a Web-Based Self-Guided Intervention (MINDxYOU) for Reducing Stress and Promoting Mental Health Among Health Professionals: Results From a Stepped-Wedge Cluster Randomized Trial (e59653)
Yolanda López-del-Hoyo, Selene Fernández-Martínez, Adrian Perez-Aranda, Alicia Monreal-Bartolomé, Alberto Barceló-Soler, Loreto Camarero-Grados, Carilene Armas-Landaeta, José Guzmán-Parra, Vera Carbonell, Daniel Campos, Xinyuan Chen, Javier García-Campayo.
2 9 3 3

Urban-Rural Differences in the Association Between Internet Use Trajectories and Depressive Symptoms in Chinese Adolescents: Longitudinal Observational Study (e63799)
Yujie Liu, Xin Ge, Ying Wang, Xue Yang, Shangbin Liu, Chen Xu, Mi Xiang, Fan Hu, Yong Cai. 2951

Effectiveness of an Internet-Based, Self-Guided, Short-Term Mindfulness Training (ISSMT) Program for Relieving Depressive Symptoms in the Adult Population in China: Single-Blind, Randomized Controlled Trial (e55583)
Tingfei Zhu, Liuyi Zhang, Wenqi Weng, Ruochen Gan, Limin Sun, Yanping Wei, Yueping Zhu, Hongyan Yu, Jiang Xue, Shulin Chen. 2964

Digital Isolation and Dementia Risk in Older Adults: Longitudinal Cohort Study (e65379)
Cheng Deng, Na Shen, Guangzhou Li, Ke Zhang, Shijun Yang. 2981

The Mediating Role of Meaning-Making in the Relationship Between Mental Time Travel and Positive Emotions in Stress-Related Blogs: Big Data Text Analysis Research (e63407)
Yidi Chen, Lei Zheng, Jinjin Ma, Huanya Zhu, Yiqun Gan. 2995

Culturally Adapted Guided Internet-Based Cognitive Behavioral Therapy for Hong Kong People With Depressive Symptoms: Randomized Controlled Trial (e64303)
Jia-Yan Pan, Jonas Rafi. 3009

More Cyberbullying, Less Happiness, and More Injustice—Psychological Changes During the Pericyberbullying Period: Quantitative Study Based on Social Media Data (e64451)
Xingyun Liu, Miao Liu, Xin Kang, Nuo Han, Yuehan Liao, Zhihong Ren. 3027

The Moderated Mediating Effects of Social Media Identity and Loneliness on the Relationship Between Problematic Internet Use and Mental Health in China: Nationwide Cross-Sectional Questionnaire Study (e57907)
Chenxi Liu, Yushu Liu, Chaojie Liu, Rujiao Lin, Xi Wang, Xinyi Zhang, Yibo Wu, Dan Wang. 3044

Competency of Large Language Models in Evaluating Appropriate Responses to Suicidal Ideation: Comparative Study (e67891)
Ryan McBain, Jonathan Cantor, Li Zhang, Olesya Baker, Fang Zhang, Alyssa Halbisen, Aaron Kofner, Joshua Breslau, Bradley Stein, Ateev Mehrotra, Hao Yu. 3060

Health Information Scanning and Seeking in Diverse Language, Cultural and Technological Media Among Latinx Adolescents: Cross-Sectional Study (e64672)
Melissa DuPont-Reyes, Alice Villatoro, Lu Tang. 3068

XSL•FO
RenderX

Online Safety When Considering Self-Harm and Suicide-Related Content: Qualitative Focus Group Study With Young People, Policy Makers, and Social Media Industry Professionals (e66321)
Louise La Sala, Amanda Sabo, Maria Michail, Pinar Thorn, Michelle Lamblin, Vivienne Browne, Jo Robinson. 3082

BePresent Universal Internet-Based Parenting Intervention: Single-Arm Pre-Post Intervention Study (e65391)
Kaisa Mishina, Amit Baumel, Malin Kinnunen, Terja Ristkari, Emmi Heinonen, Susanna Hinkka-Yli-Salomäki, Andre Sourander. 3095

Profiling Generalized Anxiety Disorder on Social Networks: Content and Behavior Analysis (e53399)
Linah Alhazzaa, Vasa Curcin. 3109

Adolescent Cyberbullying and Cyber Victimization: Longitudinal Study Before and During COVID-19 (e70508)
Peter Schulz, Marc-Olivier Boldi, Ann van Ackere. 3132

Transdiagnostic Compulsivity Traits in Problematic Use of the Internet Among UK Residents: Cross-Sectional Network Analysis Study (e66191)
Chang Liu, Samuel Chamberlain, Konstantinos Ioannidis, Jeggan Tiego, Jon Grant, Murat Yücel, Peter Hellyer, Christine Lochner, Adam Hampshire, Lucy Albertella. 3146

Use of e-Mental Health Tools for Suicide Prevention in Clinical Practice by Mental Health Professionals in NSW, Australia: Cross-Sectional Survey (e64746)
Carol Hood, Sally Hunt, Alexandra Metse, Rebecca Hodder, Kim Colyvas, Rachel Sheather-Reid, David Duerden, Jenny Bowman. 3158

US State Public Health Agencies' Use of Twitter From 2012 to 2022: Observational Study (e59786)
Samuel Mendez, Sebastian Munoz-Najar, Karen Emmons, Kasisomayajula Viswanath. 3172

Applying Natural Language Processing Techniques to Map Trends in Insomnia Treatment Terms on the r/Insomnia Subreddit: Infodemiology Study (e58902)
Jack Cummins, Daniel Gottlieb, Tamar Sofer, Danielle Wallace. 3183

Predicting User Engagement in Health Misinformation Correction on Social Media Platforms in Taiwan: Content Analysis and Text Mining Study (e65631)
Hsin-Yu Kuo, Su-Yen Chen. 3194

Forecasting the Incidence of Mumps Based on the Baidu Index and Environmental Data in Yunnan, China: Deep Learning Model Study (e66072)
Xin Xiong, Linghui Xiang, Litao Chang, Irene Wu, Shuzhen Deng. 3212

Spatiotemporal Characteristics and Influential Factors of Electronic Cigarette Web-Based Attention in Mainland China: Time Series Observational Study (e66446)
Zhongmin Zhang, Hengyi Xu, Jing Pan, Fujian Song, Ting Chen. 3225

Analyzing Themes, Sentiments, and Coping Strategies Regarding Online News Coverage of Depression in Hong Kong: Mixed Methods Study (e66696)
Sihui Chen, Cindy Ngai, Cecilia Cheng, Yangna Hu. 3237

Teenager Substance Use on Reddit: Mixed Methods Computational Analysis of Frames and Emotions (e59338)
Xinyu Zhang, Jianfeng Zhu, Deric Kenne, Ruoming Jin. 3253

Engagement With and Use of Health Information on Social Media Among US Latino Individuals: National Cross-Sectional Survey Study (e59387)
Yonaira Rivera, Kathryna Corpuz, Tahilin Karver. 3269

XSL•FO
RenderX

Impact of Demographic and Clinical Subgroups in Google Trends Data: Infodemiology Case Study on Asthma Hospitalizations (e51804)
Diana Portela, Alberto Freitas, Elísio Costa, Mattia Giovannini, Jean Bousquet, João Almeida Fonseca, Bernardo Sousa-Pinto. 3281

Contextualizing Changes in e-Cigarette Use During the Early COVID-19 Pandemic and Accompanying Infodemic ("So Much Contradictory Evidence"): Qualitative Document Analysis of Reddit Forums (e66010)
Shannon Watkins, Katherine Snodgrass, Lexi Fahrion, Emily Shaw. 3295

Optimizing Virtual Follow-Up Care: Realist Evaluation of Experiences and Perspectives of Patients With Breast and Prostate Cancer (e65148)
Sarah Scruton, Geoff Wong, Stephanie Babinski, Lauren Squires, Alejandro Berlin, Julie Easley, Sharon McGee, Ken Noel, Danielle Rodin, Jonathan Sussman, Robin Urquhart, Jacqueline Bender. 3309

Stress Reduction in Perioperative Care: Feasibility Randomized Controlled Trial (e54049)
Haridimos Kondylakis, Irene Giglioli, Dimitrios Katehakis, Hatice Aldemir, Paul Zikas, George Papagiannakis, Santiago Hors-Fraile, Pedro González-Sanz, Konstantinos Apostolakis, Constantine Stephanidis, Francisco Núñez-Benjumea, Rosa Baños-Rivera, Luis Fernandez-Luque, Angelina Kouroubali. 3324

The Long-Term Uptake of Home Spirometry in Regular Cystic Fibrosis Care: Retrospective Multicenter Observational Study (e60689)
Pia Bertram, Martinus Oppelaar, Michiel Bannier, Monique Reijers, Hester van der Vaart, Renske van der Meer, Josje Altenburg, Lennart Conemans, Bart Rottier, Marianne Nuijsink, Lara van den Wijngaart, Peter Merkus, Jolt Roukema. 3346

A New Smartphone-Based Cognitive Screening Battery for Multiple Sclerosis (icognition): Validation Study (e53503)
Stijn Denissen, Delphine Van Laethem, Johan Baijot, Lars Costers, Annabel Descamps, Ann Van Remoortel, Annick Van Merhaegen-Wieleman, Marie D'hooghe, Miguel D'Haeseleer, Dirk Smeets, Diana Sima, Jeroen Van Schependom, Guy Nagels. 3371

Effectiveness of Video Teletherapy in Treating Obsessive-Compulsive Disorder in Children and Adolescents With Exposure and Response Prevention: Retrospective Longitudinal Observational Study (e66715)
Jamie Feusner, Nicholas Farrell, Mia Nunez, Nicholas Lume, Catherine MacDonald, Patrick McGrath, Larry Trusky, Stephen Smith, Andreas Rhode. 3387

Reconstructing Risk Dimensions in Telemedicine: Investigating Technology Adoption and Barriers During the COVID-19 Pandemic in Taiwan (e53306)
Tzu-Chi Wu, Chien-Ta Ho. 3403

Long-Term Monitoring of Individuals With Chronic Obstructive Pulmonary Disease Using Digital Health Technology: Qualitative Study (e63660)
Shih-Ying Chien, Han-Chung Hu, Hsiu-Ying Cho. 3430

Mental Health Changes in US Transgender Adults Beginning Hormone Therapy Via Telehealth: Longitudinal Cohort Study (e64017)
Jae Corman, Jaclyn Hughto, Theresa Shireman, Kellan Baker, Kate Steinle, Michelle Forcier. 3441

Usage Trends and Data Sharing Practices of Healthcare Wearable Devices Among US Adults: Cross-Sectional Study (e63879)
Ranganathan Chandrasekaran, Muhammed Sadiq T, Evangelos Moustakas. 3454

Correlation Between the Online Visiting Time and Frequency Increase in Telemedicine Services Offered by Health Care Providers Before, During, and After the COVID-19 Pandemic in China: Cross-Sectional Study (e65092)
Weiyi Wang, Yuntian Chu, Fangfang Cui, Xiaobing Shi, Xu Zhang, Dongxu Sun, Jinming Shi, Jie Zhao. 3472

XSL•FO
RenderX

Telemedicine Use During the COVID-19 Pandemic in 8 Countries From the International Sexual Health and Reproductive Health Consortium: Web-Based Cross-Sectional Survey Study (e60369)
Rayner Tan, Devon Hensel, Olena Ivanova, Raquel Bravo, Adesola Olumide, Emmanuel Adebayo, Amanda Cleeve, Amanda Gesselman, Sonam Shah, Helen Adesoba, Gifty Marley, Weiming Tang. . . . 3485

Effectiveness of Telehealth Versus In-Person Informed Consent: Randomized Study of Comprehension and Decision-Making (e63473)
Saif Khairat, Paige Ottmar, Prabal Chourasia, Jihad Obeid. . . . 3509

Impact on Patient Outcomes of Continuous Vital Sign Monitoring on Medical Wards: Propensity-Matched Analysis (e66347)
Bradley Rowland, Amit Saha, Vida Motamedi, Richa Bundy, Scott Winsor, Daniel McNavish, William Lippert, Ashish Khanna. . . . 3519

Prevalence of Antibiotic Prescribing for Acute Respiratory Tract Infection in Telehealth Versus Face-to-Face Consultations: Cross-Sectional Analysis of General Practice Registrars' Clinical Practice (e60831)
Yu Gao, Parker Magin, Amanda Tapley, Elizabeth Holliday, Jason Dizon, Katie Fisher, Mieke van Driel, Joshua Davis, Andrew Davey, Anna Ralston, Alison Fielding, Dominica Moad, Katie Mulquiney, Lisa Clarke, Alexandria Turner. . . . 3530

Emotion Forecasting: A Transformer-Based Approach (e63962)
Leire Paz-Arbaizar, Jorge Lopez-Castroman, Antonio Artés-Rodríguez, Pablo Olmos, David Ramírez. . . . 3543

Application of Internet Hospitals in the Disease Management of Patients With Ulcerative Colitis: Retrospective Study (e60019)
Tianzhi Yu, Wanyu Li, Yingchun Liu, Chunjie Jin, Zimin Wang, Hailong Cao. . . . 3562

Factors Associated With the Availability of Virtual Consultations in Primary Care Across 20 Countries: Cross-Sectional Study (e65147)
Gabriele Kerr, Geva Greenfield, Edmond Li, Thomas Beaney, Benedict Hayhoe, Josip Car, Ana Clavería, Claire Collins, Gustavo Gusso, Robert Hoffman, Geronimo Jimenez, Tuomas Koskela, Liliana Laranjo, Heidrun Lingner, Ensieh Memarian, Katarzyna Nessler, Davorina Petek, Rosy Tsopra, Azeem Majeed, Ana Neves. . . . 3571

Effect of a Telemedicine Model on Patients With Heart Failure With Reduced Ejection Fraction in a Resource-Limited Setting in Vietnam: Cohort Study (e67228)
Hoai Nguyen, Hieu Tran, Phuong Tran, Hung Pham, Co Dao, Thanh Le, Loi Do, Ha Nguyen, Thom Vu, James Kirkpatrick, Christopher Reid, Dung Nguyen. . . . 3584

Differences in Telemedicine Use for Patients With Diabetes in an Academic Versus Safety Net Health System: Retrospective Cohort Study (e64635)
Jonathan Shih, Magdalene Kuznia, Sarah Nouri, Elizabeth Sherwin, Kathryn Kemper, Anna Rubinsky, Courtney Lyles, Elaine Khoong. . . . 3598

Topic Modeling of Nursing Issues in the Media During 4 Emerging Infectious Disease Epidemics in South Korea: Descriptive Analysis (e60446)
Jungok Kim, Eun Yun. . . . 3617

The Willingness of Doctors to Adopt Artificial Intelligence–Driven Clinical Decision Support Systems at Different Hospitals in China: Fuzzy Set Qualitative Comparative Analysis of Survey Data (e62768)
Zhongguang Yu, Ning Hu, Qiuyi Zhao, Xiang Hu, Cunbo Jia, Chunyu Zhang, Bing Liu, Yanping Li. . . . 3631

Health and Experiences During the COVID-19 Pandemic Among Children and Young People: Analysis of Free-Text Responses From the Children and Young People With Long COVID Study (e63634)
Natalia Rojas, Sam Martin, Mario Cortina-Borja, Roz Shafran, Lana Fox-Smith, Terence Stephenson, Brian Ching, Anaïs d'Oelsnitz, Tom Norris, Yue Xu, Kelsey McOwat, Emma Dalrymple, Isobel Heyman, Tamsin Ford, Trudie Chalder, Ruth Simmons, CLoCk Consortium, Snehal Pinto Pereira. . . . 3647

XSL•FO
RenderX

Impact of Mobile Phone Usage on Sleep Quality Among Medical Students Across Latin America: Multicenter Cross-Sectional Study (e60630)
Juan Izquierdo-Condoy, Clara Paz, H Nati-Castillo, Ricardo Gollini-Mihalopoulos, Telmo Aveiro-Róbalo, Jhino Valeriano Paucar, Sandra Laura Mamami, Juan Caicedo, Valentina Loaiza-Guevara, Diana Mejía, Camila Salazar-Santoliva, Melissa Villavicencio-Gomezjurado, Cougar Hall, Esteban Ortiz-Prado. 3665

A New Computer-Based Cognitive Measure for Early Detection of Dementia Risk (Japan Cognitive Function Test): Validation Study (e59015)
Hiroyuki Shimada, Takehiko Doi, Kota Tsutsumimoto, Keitaro Makino, Kenji Harada, Kouki Tomida, Masanori Morikawa, Hyuma Makizako. . . 3 6 7 8

Association of Screen Content With Early Development Among Preschoolers in Shanghai: 7-Day Monitoring Study With Auto Intelligent Technology (e65343)
Hao Chen, Yi Sun, Sha Luo, Yingyan Ma, Chenshu Li, Yingcheng Xiao, Yimeng Zhang, Senlin Lin, Yingnan Jia. 3691

Monitoring Sleep Quality Through Low α-Band Activity in the Prefrontal Cortex Using a Portable Electroencephalogram Device: Longitudinal Study (e67188)
Chuanliang Han, Zhizhen Zhang, Yuchen Lin, Shaojia Huang, Jidong Mao, Weiwen Xiang, Fang Wang, Yuping Liang, Wufang Chen, Xixi Zhao. 3702

Investigating the Magnitude and Persistence of COVID-19–Related Impacts on Affect and GPS-Derived Daily Mobility Patterns in Adolescence and Emerging Adulthood: Insights From a Smartphone-Based Intensive Longitudinal Study of Colorado-Based Youths From June 2016 to April 2022 (e64965)
Jordan Alexander, Kelly Duffy, Samantha Freis, Sy-Miin Chow, Naomi Friedman, Scott Vrieze. 3718

The Impact of the COVID-19 Pandemic on Oldest-Old Social Capital and Health and the Role of Digital Inequalities: Longitudinal Cohort Study (e62824)
Luca Valla, Michele Rossi, Alessandra Gaia, Antonio Guaita, Elena Rolandi. 3735

The Association Between the Digital Divide and Health Inequalities Among Older Adults in China: Nationally Representative Cross-Sectional Survey (e62645)
Mengqiu Wu, Yongxi Xue, Chengyu Ma. 3750

Latinx and White Adolescents' Preferences for Latinx-Targeted Celebrity and Noncelebrity Food Advertisements: Experimental Survey Study (e53188)
Marie Bragg, Samina Lutfeali, Daniela Gabler, Diego Quintana Licona, Jennifer Harris. 3772

Colorectal Cancer Racial Equity Post Volume, Content, and Exposure: Observational Study Using Twitter Data (e63864)
Chau Tong, Drew Margolin, Jeff Niederdeppe, Rumi Chunara, Jiawei Liu, Lea Jih-Vieira, Andy King. 3790

Perspectives of Hispanic and Latinx Community Members on AI-Enabled mHealth Tools: Qualitative Focus Group Study (e59817)
Stephanie Kraft, Shaan Chopra, Miriana Duran, Janet Rojina, Abril Beretta, Katherine López, Russell Javan, Benjamin Wilfond, Margaret Rosenfeld, James Fogarty, Linda Ko. 3801

Understanding the Association Between Home Broadband Connection and Well-Being Among Middle-Aged and Older Adults in China: Nationally Representative Panel Data Study (e59023)
Lu Yang, Chris Lynch, John Lee, Brian Oldenburg, Tilahun Haregu. 3816

Digital Competence of Arabic-Speaking Immigrant and Refugee Older Adults Enacting Agency and Navigating Barriers: Qualitative Descriptive Study (e60547)
Jordana Salma, Alesia Au, Ghada Sayadi, Manal Kleib. 3835

XSL•FO
RenderX

Patterns of Internet Use in People Diagnosed With Severe Mental Illness: Qualitative Interview Study (e55072)
Ruth Wadman, Lauren Walker, Olivia Taylor, Paul Heron, Elizabeth Newbronner, Panagiotis Spanakis, Suzanne Crosland, Emily Peckham. . . 3 8 5 2

Noninvasive Oral Hyperspectral Imaging–Driven Digital Diagnosis of Heart Failure With Preserved Ejection Fraction: Model Development and Validation Study (e67256)
Xiaomeng Yang, Zeyan Li, Lei Lei, Xiaoyu Shi, Dingming Zhang, Fei Zhou, Wenjing Li, Tianyou Xu, Xinyu Liu, Songyun Wang, Quan Yuan, Jian Yang, Xinyu Wang, Yanfei Zhong, Lilei Yu. . . . 3867

Enhancing Diagnostic Accuracy of Lung Nodules in Chest Computed Tomography Using Artificial Intelligence: Retrospective Analysis (e64649)
Weiqi Liu, You Wu, Zhuozhao Zheng, Mark Bittle, Wei Yu, Hadi Kharrazi. . . . 3883

MetaAnalysisOnline.com: Web-Based Tool for the Rapid Meta-Analysis of Clinical and Epidemiological Studies (e64016)
János Fekete, Balázs Gy rffy. . . . 3893

Introducing Novel Methods to Identify Fraudulent Responses (Sampling With Sisyphus): Web-Based LGBTQ2S+ Mixed-Methods Study (e63252)
Kinnon MacKinnon, Naail Khan, Katherine Newman, Wren Gould, Gin Marshall, Travis Salway, Annie Pullen Sansfaçon, Hannah Kia, June Lam. . . . 3904

What's Going On With Me and How Can I Better Manage My Health? The Potential of GPT-4 to Transform Discharge Letters Into Patient-Centered Letters to Enhance Patient Safety: Prospective, Exploratory Study (e67143)
Felix Eisinger, Friederike Holderried, Moritz Mahling, Christian Stegemann–Philipps, Anne Herrmann–Werner, Eric Nazarenus, Alessandra Sonanini, Martina Guthoff, Carsten Eickhoff, Martin Holderried. . . . 3923

Enhancing Doctor-Patient Shared Decision-Making: Design of a Novel Collaborative Decision Description Language (e55341)
XiaoRui Guo, Liang Xiao, Xinyu Liu, Jianxia Chen, Zefang Tong, Ziji Liu. . . . 3960

Consumer Engagement With Risk Information on Prescription Drug Social Media Pages: Findings From In-Depth Interviews (e67361)
Jacqueline Amoozegar, Peyton Williams, Kristen Giombi, Courtney Richardson, Ella Shenkar, Rebecca Watkins, Amie O'Donoghue, Helen Sullivan. . . . 3983

Perceptions, Experiences, and Beliefs About Patient Portals Among Women With Limited English Proficiency: Multicultural Qualitative Interview Study (e60699)
Hae-Ra Han, Joyline Chepkorir, Tina Kim, Martha Zamora, Emma Huang. . . . 3995

The Virtual Client Experience Survey for Mental Health and Addictions: Revalidation of a Survey to Measure Client and Family Experiences of Virtual Care (e49844)
Allison Crawford, Anne Kirvan, Marcos Sanches, Amanda Gambin, Denise Canso, Eva Serhal. . . . 4007

Development and Validation of the Digital Sensitivity Scale for Adults: Cross-Sectional Observational Study (e55828)
Hae Park, Minjeong Jeon, Ji Ahn, Kyungmi Chung, Jin Park. . . . 4018

Psychometric Analysis of the eHealth Literacy Scale in Portuguese Older Adults (eHEALS-PT24): Instrument Development and Validation (e57730)
Sara Luz, Paulo Nogueira, Andreia Costa, Adriana Henriques. . . . 4034

Assessing Short-Video Dependence for e-Mental Health: Development and Validation Study of the Short-Video Dependence Scale (e66341)
AnHang Jiang, Shuang Li, HuaBin Wang, HaoSen Ni, HongAn Chen, JunHong Dai, XueFeng Xu, Mei Li, Guang-Heng Dong. . . . 4052

XSL•FO
RenderX

Trust and Acceptance Challenges in the Adoption of AI Applications in Health Care: Quantitative Survey Analysis (e65567)
Janne Kauttonen, Rebekah Rousi, Ari Alamäki. 4067

Development and Validation of the Digital Health Literacy Questionnaire for Stroke Survivors: Exploratory Sequential Mixed Methods Study (e64591)
Qin Ye, Wei Wang, Xuan Zeng, Yuxian Kuang, Bingbing Geng, Song Zhou, Ning Liu. 4089

Global Use, Adaptation, and Sharing of Massive Open Online Courses for Emergency Health on the OpenWHO Platform: Survey Study (e52591)
Jamie Johnston, Nadine Skinner, Anna Tokar, Elham Arabi, Ngouille Ndiaye, Matthew Strehlow, Heini Utunen. 4102

Comparing Virtual Reality–Based and Traditional Physical Objective Structured Clinical Examination (OSCE) Stations for Clinical Competency Assessments: Randomized Controlled Trial (e55066)
Tobias Mühling, Verena Schreiner, Marc Appel, Tobias Leutritz, Sarah König. 4119

Machine Learning–Based Prediction of Substance Use in Adolescents in Three Independent Worldwide Cohorts: Algorithm Development and Validation Study (e62805)
Soeun Kim, Hyejun Kim, Seokjun Kim, Hojae Lee, Ahmed Hammoodi, Yujin Choi, Hyeon Kim, Lee Smith, Min Kim, Guillaume Fond, Laurent Boyer, Sung Baik, Hayeon Lee, Jaeyu Park, Rosie Kwon, Selin Woo, Dong Yon. 4158

Virtual Patient Simulations Using Social Robotics Combined With Large Language Models for Clinical Reasoning Training in Medical Education: Mixed Methods Study (e63312)
Alexander Borg, Carina Georg, Benjamin Jobs, Viking Huss, Kristin Waldenlind, Mini Ruiz, Samuel Edelbring, Gabriel Skantze, Ioannis Parodis.
4 1 7 2

Asynchronous Distance Learning Performance and Knowledge Retention of the National Institutes of Health Stroke Scale Among Health Care Professionals Using Video or e-Learning: Web-based Randomized Controlled Trial (e63136)
Avinash Koka, Loric Stuby, Emmanuel Carrera, Ahmed Gabr, Margaret O'Connor, Nathalie Missilier Peruzzo, Olivier Waeterloot, Friedrich Medlin, Fabien Rigolet, Thomas Schmutz, Patrik Michel, Thibaut Desmettre, Mélanie Suppan, Laurent Suppan. 4185

Digital Gaming and Subsequent Health and Well-Being Among Older Adults: Longitudinal Outcome-Wide Analysis (e69080)
Atsushi Nakagomi, Kazushige Ide, Katsunori Kondo, Koichiro Shiba. 4198

Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives (e48406)
Bryan Cleal, Yanbing Chen, Mandy Wäldchen, Hanne Ballhausen, Drew Cooper, Shane O'Donnell, Christine Knoll, Niklas Krug, Klemens Raile, Tebbe Ubben, Adrian Tappe, Dana Lewis, Ingrid Willaing, Timothy Skinner, Katarina Braune. 4208

Unraveling Online Mental Health Through the Lens of Early Maladaptive Schemas: AI-Enabled Content Analysis of Online Mental Health Communities (e59524)
Beng Ang, Sujatha Gollapalli, Mingzhe Du, See-Kiong Ng. 4227

Understanding the Engagement and Interaction of Superusers and Regular Users in UK Respiratory Online Health Communities: Deep Learning–Based Sentiment Analysis (e56038)
Xiancheng Li, Emanuela Vaghi, Gabriella Pasi, Neil Coulson, Anna De Simoni, Marco Viviani, AD HOC Group. 4244

Harnessing Social Media Data to Understand Information Needs About Kidney Diseases and Emotional Experiences With Disease Management: Topic and Sentiment Analysis (e64838)
Hee Hwang, Nara Kim, Jeong You, Hye Ryu, Seo-Young Kim, Jung Yoon Park, Ki Lee. 4259

XSL•FO
RenderX

Women's Educating and Coping Strategies for Cultivating Supportive Web-Based Spaces for Discussing Sexual and Reproductive Health: Co-Design Study (e62716)
Hyeyoung Ryu, Wanda Pratt. . . . 4279

Examining BRCA Previvors' Social Media Content Creation as a Form of Self and Community Care: Qualitative Interview Study (e67794)
Mariah Wellman, Camilla Owens, Avery Holton, Kimberly Kaphingst. . . . 4303

The Lived Experience of Participating in Online Peer-To-Peer Groups After Acquired Brain Injury: Phenomenological Study (e67658)
Malin Tistad, Lill Hultman, Annica Wohlin Wottrich, Lena von Koch. . . . 4311

Major Adverse Kidney Events in Hospitalized Older Patients With Acute Kidney Injury: Machine Learning–Based Model Development and Validation Study (e52786)
Xiao-Qin Luo, Ning-Ya Zhang, Ying-Hao Deng, Hong-Shen Wang, Yi-Xin Kang, Shao-Bin Duan. . . . 4322

Effectiveness of Remote Patient Monitoring Equipped With an Early Warning System in Tertiary Care Hospital Wards: Retrospective Cohort Study (e56463)
Pavithra Lakshman, Priyanka Gopal, Sheen Khurdi. . . . 4336

A Digital Tool for Clinical Evidence–Driven Guideline Development by Studying Properties of Trial Eligible and Ineligible Populations: Development and Usability Study (e52385)
Shahzad Mumtaz, Megan McMinn, Christian Cole, Chuang Gao, Christopher Hall, Magalie Guignard-Duff, Huayi Huang, David McAllister, Daniel Morales, Emily Jefferson, Bruce Guthrie. . . . 4350

Combining a Risk Factor Score Designed From Electronic Health Records With a Digital Cytology Image Scoring System to Improve Bladder Cancer Detection: Proof-of-Concept Study (e56946)
Sandie Cabon, Sarra Brihi, Riadh Fezzani, Morgane Pierre-Jean, Marc Cuggia, Guillaume Bouzillé. . . . 4384

Risk Factors for Gastrointestinal Bleeding in Patients With Acute Myocardial Infarction: Multicenter Retrospective Cohort Study (e67346)
Yanqi Kou, Shicai Ye, Yuan Tian, Ke Yang, Ling Qin, Zhe Huang, Botao Luo, Yanping Ha, Liping Zhan, Ruyin Ye, Yujie Huang, Qing Zhang, Kun He, Mouji Liang, Jieming Zheng, Haoyuan Huang, Chunyi Wu, Lei Ge, Yuping Yang. . . . 4401

An Easy and Quick Risk-Stratified Early Forewarning Model for Septic Shock in the Intensive Care Unit: Development, Validation, and Interpretation Study (e58779)
Guanghao Liu, Shixiang Zheng, Jun He, Zi-Mei Zhang, Ruoqiong Wu, Yingying Yu, Hao Fu, Li Han, Haibo Zhu, Yichang Xu, Huaguo Shao, Haidan Yan, Ting Chen, Xiaopei Shen. . . . 4414

Perceptions in 3.6 Million Web-Based Posts of Online Communities on the Use of Cancer Immunotherapy: Data Mining Using BERTopic (e60948)
Xingyue Wu, Chun Lam, Ka Hui, Herbert Loong, Keary Zhou, Chun-Kit Ngan, Yin Cheung. . . . 4430

Automated Process for Monitoring of Amiodarone Treatment: Development and Evaluation (e65473)
Birgitta Johansson, Jonas Landahl, Karin Tammelin, Erik Aerts, Christina Lundberg, Martin Adiels, Martin Lindgren, Annika Rosengren, Nikolaos Papachrysos, Helena Filipsson Nyström, Helen Sjöland. . . . 4442

Evaluating the Knowledge and Information-Seeking Behaviors of People Living With Multiple Sclerosis: Cross-Sectional Questionnaire Study (e63763)
Véronique Duguay, Dominique Comeau, Tiffany Turgeon, Nadia Bouhamdani, Mathieu Belanger, Lyle Weston, Tammy Johnson, Nicole Manzer, Melissa Giberson, Ludivine Chamard-Witkowski. . . . 4455

Complete Blood Count and Monocyte Distribution Width–Based Machine Learning Algorithms for Sepsis Detection: Multicentric Development and External Validation Study (e55492)
Andrea Campagner, Luisa Agnello, Anna Carobene, Andrea Padoan, Fabio Del Ben, Massimo Locatelli, Mario Plebani, Agostino Ognibene, Maria Lorubbio, Elena De Vecchi, Andrea Cortegiani, Elisa Piva, Donatella Poz, Francesco Curcio, Federico Cabitza, Marcello Ciaccio. . . . 4466

XSL•FO
RenderX

Stroke Diagnosis and Prediction Tool Using ChatGLM: Development and Validation Study (e67010)
Xiaowei Song, Jiayi Wang, Feifei He, Wei Yin, Weizhi Ma, Jian Wu. 4490

Machine Learning–Based Prediction of Early Complications Following Surgery for Intestinal Obstruction: Multicenter Retrospective Study (e68354)
Pinjie Huang, Jirong Yang, Dizhou Zhao, Taojia Ran, Yuheng Luo, Dong Yang, Xueqin Zheng, Shaoli Zhou, Chaojin Chen. 4502

Machine Learning Models With Prognostic Implications for Predicting Gastrointestinal Bleeding After Coronary Artery Bypass Grafting and Guiding Personalized Medicine: Multicenter Cohort Study (e68509)
Jiale Dong, Zhechuan Jin, Chengxiang Li, Jian Yang, Yi Jiang, Zeqian Li, Cheng Chen, Bo Zhang, Zhaofei Ye, Yang Hu, Jianguo Ma, Ping Li, Yulin Li, Dongjin Wang, Zhili Ji. 4516

Oncology Provider and Patient Perspectives on a Cardiovascular Health Assessment Tool Used During Posttreatment Survivorship Care in Community Oncology (Results from WF-1804CD): Mixed Methods Observational Study (e65152)
Chandylen Nightingale, Emily Dressler, Maura Kepper, Heidi Klepin, Simon Lee, Sydney Smith, Aylin Aguilar, Kimberly Wiseman, Stephanie Sohl, Brian Wells, Joseph DeMari, Alyssa Throckmorton, Lindsey Kulbacki, Jenny Hanna, Randi Foraker, Kathryn Weaver. 4535

Risk Stratification in Immunoglobulin A Nephropathy Using Network Biomarkers: Development and Validation Study (e65563)
Jiaxing Tan, Rongxin Yang, Liyin Xiao, Lingqiu Dong, Zhengxia Zhong, Ling Zhou, Wei Qin. 4549

Two-Year Hypertension Incidence Risk Prediction in Populations in the Desert Regions of Northwest China: Prospective Cohort Study (e68442)
Yinlin Cheng, Kuiying Gu, Weidong Ji, Zhensheng Hu, Yining Yang, Yi Zhou. 4566

Machine Learning to Assist in Managing Acute Kidney Injury in General Wards: Multicenter Retrospective Study (e66568)
Nam-Jun Cho, Inyong Jeong, Se-Jin Ahn, Hyo-Wook Gil, Yeongmin Kim, Jin-Hyun Park, Sanghee Kang, Hwamin Lee. 4583

Factors Influencing Primary Care Physicians' Intent to Refer Patients With Hypertension to a Digital Remote Blood Pressure Monitoring Program: Mixed Methods Study (e64933)
Jennifer Wu, Ross Graham, Julie Çelebi, Kevin Fraser, Geneen Gin, Laurel Dang, Esmatullah Hatamy, Amanda Walker, Courtney Barbato, Ottar Lunde, Lisa Coles, Parag Agnihotri, Cassandra Morn, Ming Tai-Seale. 4598

Detection of Clinically Significant Drug-Drug Interactions in Fatal Torsades de Pointes: Disproportionality Analysis of the Food and Drug Administration Adverse Event Reporting System (e65872)
Huanhuan Ji, Meiling Gong, Li Gong, Ni Zhang, Ruiou Zhou, Dongmei Deng, Ya Yang, Lin Song, Yuntao Jia. 4610

Recruitment of Young Gay, Bisexual, and Other Men Who Have Sex With Men for a Web-Based Human Papillomavirus Vaccination Intervention: Differences in Participant Characteristics and Study Engagement by Recruitment Source in a Randomized Controlled Trial (e64668)
Daniel Marshall, Amy Gower, Mira Katz, José Bauermeister, Abigail Shoben, Paul Reiter. 4625

Urban-Suburban Differences in Public Perspectives on Digitalizing Pediatric Research: Cross-Sectional Survey Study (e60324)
Heping Fang, Ruoling Xian, Juan Li, Yingcun Li, Enmei Liu, Yan Zhao, Yan Hu. 4637

Recruiting Young People for Digital Mental Health Research: Lessons From an AI-Driven Adaptive Trial (e60413)
Wu Zheng, Artur Shvetcov, Aimy Slade, Zoe Jenkins, Leonard Hoon, Alexis Whitton, Rena Logothetis, Smrithi Ravindra, Stefanus Kurniawan, Sunil Gupta, Kit Huckvale, Eileen Stech, Akash Agarwal, Joost Funke Kupper, Stuart Cameron, Jodie Rosenberg, Nicholas Manoglou, Manisha Senadeera, Svetha Venkatesh, Kon Mouzakis, Rajesh Vasa, Helen Christensen, Jill Newby. 4650

XSL•FO
RenderX

Parental Mental Health and Child Maltreatment in the COVID-19 Pandemic: Importance of Sampling in a Quantitative Statistical Study (e52043)
Lara Engelke, Claudia Calvano, Steffi Pohl, Sibylle Winter, Babette Renneberg. 4661

Detecting Deception and Ensuring Data Integrity in a Nationwide mHealth Randomized Controlled Trial: Factorial Design Survey Study (e66384)
Krista Kezbers, Michael Robertson, Emily Hébert, Audrey Montgomery, Michael Businelle. 4682

The Reliability and Quality of Videos as Guidance for Gastrointestinal Endoscopy: Cross-Sectional Study (e58855)
Jinpei Liu, Yifan Qiu, Yilong Liu, Wenping Xu, Weichen Ning, Peimei Shi, Zongli Yuan, Fang Wang, Yihai Shi. 4705

Mechanism Assessment of Physician Discourse Strategies and Patient Consultation Behaviors on Online Health Platforms: Mixed Methods Study (e54516)
Menglei Kong, Yu Wang, Meixuan Li, Zhong Yao. 4717

Factors Influencing Informed Consent Preferences in Digital Health Research: Survey Study of Prospective Participants (e63349)
Brian McInnis, Ramona Pindus, Daniah Kareem, Daniela Vital, Eric Hekler, Camille Nebeker. 4739

Finding Consensus on Trust in AI in Health Care: Recommendations From a Panel of International Experts (e56306)
Georg Starke, Felix Gille, Alberto Termine, Yves Aquino, Ricardo Chavarriaga, Andrea Ferrario, Janna Hastings, Karin Jongsma, Philipp Kellmeyer, Bogdan Kulynych, Emily Postan, Elise Racine, Derya Sahin, Paulina Tomaszewska, Karina Vold, Jamie Webb, Alessandro Facchini, Marcello Ienca. 4751

Publishing Identifiable Patient Photographs in the Digital Age: Focus Group Study of Patients, Doctors, and Medical Students (e59970)
Marija Rogulji , Dina Šimunovi , Ivan Buljan, Marija Žuljevi , Antonela Turi , Ana Maruši 4769

The Association Between COVID-19 Vaccination Uptake and Information-Seeking Behaviors Using the Internet: Nationwide Cross-Sectional Study (e59352)
Kazuya Taira, Misa Shiomi, Takayo Nakabe, Yuichi Imanaka. 4780

Cross-Cultural Sense-Making of Global Health Crises: A Text Mining Study of Public Opinions on Social Media Related to the COVID-19 Pandemic in Developed and Developing Economies (e58656)
Adham Kahlawi, Firas Masri, Wasim Ahmed, Josep Vidal-Alaball. 4793

Description of Weight-Related Content and Recommended Dietary Behaviors for Weight Loss Frequently Reposted on X (Twitter) in English and Japanese: Content Analysis (e64739)
Fumi Oono, Mai Matsumoto, Risa Ogata, Mizuki Suga, Kentaro Murakami. 4808

Data-Sharing Statements Requested from Clinical Trials by Public, Environmental, and Occupational Health Journals: Cross-Sectional Study (e64069)
Yingxin Liu, Jingyi Zhang, Lehana Thabane, Xuerui Bai, Lili Kang, Gregory Lip, Harriette Van Spall, Min Xia, Guowei Li. 4821

Predicting the Risk of HIV Infection and Sexually Transmitted Diseases Among Men Who Have Sex With Men: Cross-Sectional Study Using Multiple Machine Learning Approaches (e59101)
Bing Lin, Jiaxiu Liu, Kangjie Li, Xiaoni Zhong. 4834

Characterizing Public Sentiments and Drug Interactions in the COVID-19 Pandemic Using Social Media: Natural Language Processing and Network Analysis (e63755)
Wanxin Li, Yining Hua, Peilin Zhou, Li Zhou, Xin Xu, Jie Yang. 4851

XSL•FO
RenderX

Health IT Implementation and the Impact of the COVID-19 Pandemic on Clinician-IT Dynamics: Qualitative Study (e57847)
Adeola Bamgboje-Ayodele, Adrian Boscolo, Mitchell Burger, Owen Hutchings, Miranda Shaw, Tim Shaw, Amina Tariq, Sundresan Naicker, Steven McPhail, Melissa Baysari. . . . 4869

Evaluating Whether Nonimmersion Virtual Reality Simulation Training Improves Nursing Competency in Isolation Wards: Randomized Controlled Trial (e63131)
Dandan Zhang, MuLi Fu, Jianzhong Zhang, Yuxuan Li, Li Chen, Yong-Jun Chen, Zhefeng Zhong, Yin-Ping Zhang. . . . 4910

Mass Casualty Incident Training in Immersive Virtual Reality: Quasi-Experimental Evaluation of Multimethod Performance Indicators (e63241)
Anke Baetzner, Yannick Hill, Benjamin Roszipal, Solène Gerwann, Matthias Beutel, Tanja Birrenbach, Markus Karlseder, Stefan Mohr, Gabriel Salg, Helmut Schrom-Feiertag, Marie Frenkel, Cornelia Wrzus. . . . 4925

Neurobiological Mechanisms of Enhanced Pain-Relieving Transcutaneous Electrical Nerve Stimulation via Visuo-Tactile Stimulation in Immersive Virtual Reality: Randomized Controlled Trial (e63137)
Chenxi Wang, Lanqi Gao, Chuan Zhang, Jun Li, Jixin Liu. . . . 4996

Prompt Framework for Extracting Scale-Related Knowledge Entities from Chinese Medical Literature: Development and Evaluation Study (e67033)
Jie Hao, Zhenli Chen, Qinglong Peng, Liang Zhao, Wanqing Zhao, Shan Cong, Junlian Li, Jiao Li, Qing Qian, Haixia Sun. . . . 5007

Using Synthetic Health Care Data to Leverage Large Language Models for Named Entity Recognition: Development and Validation Study (e66279)
Hendrik Šuvalov, Mihkel Lepson, Veronika Kukk, Maria Malk, Neeme Ilves, Hele-Andra Kuulmets, Raivo Kolde. . . . 5023

Association Between Intergenerational Support, Technology Perception and Trust, and Intention to Seek Medical Care on the Internet Among Chinese Older Adults: Cross-Sectional Questionnaire Study (e65065)
Hengjiang Jin, Ying Qu. . . . 5035

Jargon and Readability in Plain Language Summaries of Health Research: Cross-Sectional Observational Study (e50862)
Iain Lang, Angela King, Kate Boddy, Ken Stein, Lauren Asare, Jo Day, Kristin Liabo. . . . 5053

Impact of Internet Usage on Depression Among Older Adults: Comprehensive Study (e65399)
Lin Guo, Yunwei Li, Kai Cheng, Ying Zhao, Wenqiang Yin, Ying Liu. . . . 5070

Measuring Digital Health Literacy in Older Adults: Development and Validation Study (e65492)
SungMin Kim, Choonghee Park, Sunghyeon Park, Dai-Jin Kim, Ye-Seul Bae, Jae-Heon Kang, Ji-Won Chun. . . . 5084

Identifying Profiles of Digital Literacy Among Community-Dwelling Korean Older Adults: Latent Profile Analysis (e57122)
Jiyoung Shin, Hun Kang, Seongmi Choi, Sang Chu, JiYeon Choi. . . . 5099

Digital Health Literacy Questionnaire for Older Adults: Instrument Development and Validation Study (e64193)
Xinxin Wang, Chengrui Zhang, Yue Qi, Ying Xing, Yawen Liu, Jiayi Sun, Wei Luan. . . . 5115

Decomposition and Comparative Analysis of Urban-Rural Disparities in eHealth Literacy Among Chinese University Students: Cross-Sectional Study (e63671)
Yao Yu, Zhenning Liang, Qingping Zhou, Yusupujiang Tuersun, Siyuan Liu, Chenxi Wang, Yuying Xie, Xinyu Wang, Zhuotong Wu, Yi Qian. . . 5 1 3 1

XSL•FO
RenderX

Identification of Intracranial Germ Cell Tumors Based on Facial Photos: Exploratory Study on the Use of Deep Learning for Software Development (e58760)
Yanong Li, Yixuan He, Yawei Liu, Bingchen Wang, Bo Li, Xiaoguang Qiu. 5147

Electronic Health Record Data Collection Practices to Advance Standardization and Interoperability of Patient Preferences for Interpretation Services: Qualitative Study (e62670)
Krysta Heaney-Huls, Rida Shams, Ruth Nwefo, Rachel Kane, Janna Gordon, Alison Laffan, Scott Stare, Prashila Dullabh. 5163

The Prognostic Significance of Sleep and Circadian Rhythm for Myocardial Infarction Outcomes: Case-Control Study (e63897)
Wei-Chih Chin, Pao-Hsien Chu, Lung-Sheng Wu, Kuang-Tso Lee, Chen Lin, Chien-Te Ho, Wei-Sheng Yang, I-Hang Chung, Yu-Shu Huang. . . 5 1 7 5

Mortality Risk Prediction in Patients With Antimelanoma Differentiation–Associated, Gene 5 Antibody–Positive, Dermatomyositis–Associated Interstitial Lung Disease: Algorithm Development and Validation (e62836)
Hui Li, Ruyi Zou, Hongxia Xin, Ping He, Bin Xi, Yaqiong Tian, Qi Zhao, Xin Yan, Xiaohua Qiu, Yujuan Gao, Yin Liu, Min Cao, Bi Chen, Qian Han, Juan Chen, Guochun Wang, Hourong Cai. 5193

Implementation of Medication-Related Technology and Its Impact on Pharmacy Workflow: Real-World Evidence Usability Study (e59220)
Wei-Ning Yu, Yih-Dih Cheng, Yu-Chi Hou, Yow-Wen Hsieh. 5210

An Intelligent System for Classifying Patient Complaints Using Machine Learning and Natural Language Processing: Development and Validation Study (e55721)
Xiadong Li, Qiang Shu, Canhong Kong, Jinhu Wang, Gang Li, Xin Fang, Xiaomin Lou, Gang Yu. 5220

Revealing Patient Dissatisfaction With Health Care Resource Allocation in Multiple Dimensions Using Large Language Models and the International Classification of Diseases 11th Revision: Aspect-Based Sentiment Analysis (e66344)
Jiaxuan Li, Yunchu Yang, Chao Mao, Patrick Pang, Quanjing Zhu, Dejian Xu, Yapeng Wang. 5241

Estimating Patient and Family Costs and CO2 Emissions for Telehealth and In-Person Health Care Appointments in British Columbia, Canada: Geospatial Mixed Methods Study (e56766)
Graham Mainer-Pearson, Kurtis Stewart, Kim Williams, John Pawlovich, Scott Graham, Linda Riches, Sonya Cressman, Kendall Ho. 5257

Ubiquitous News Coverage and Its Varied Effects in Communicating Protective Behaviors to American Adults in Infectious Disease Outbreaks: Time-Series and Longitudinal Panel Study (e64307)
Anqi Shao, Kaiping Chen, Branden Johnson, Shaila Miranda, Qidi Xing. 5270

Stakeholder Consensus on an Interdisciplinary Terminology to Enable the Development and Uptake of Medication Adherence Technologies Across Health Systems: Web-Based Real-Time Delphi Study (e59738)
Alexandra Dima, Urska Nabergoj Makovec, Janette Ribaut, Frederik Haupenthal, Pilar Barnestein-Fonseca, Catherine Goetzinger, Sean Grant, Cristina Jácome, Dins Smits, Ivana Tadic, Job van Boven, Ioanna Tsiligianni, Maria Herdeiro, Fátima Roque, European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE). 5284

New Perspective on Digital Well-Being by Distinguishing Digital Competency From Dependency: Network Approach (e70483)
Si Chen, Omid Ebrahimi, Cecilia Cheng. 5304

Patients' Use of e-Consultations as an Alternative to Other General Practitioner Services: Cross-Sectional Survey Study (e55158)
Eli Kristiansen, Helen Atherton, Bjarne Austad, Trine Bergmo, Børge Norberg, Chris Salisbury, Paolo Zanaboni. 5326

The Role of the Installed Base in Information Exchange Among General Practitioners in Germany: Mixed Methods Study (e65241)
Tim Holetzek, Andreas Häusler, Kathrin Gödde, Michael Rapp, Jacob Spallek, Christine Holmberg. 5343

XSL•FO
RenderX

The Dynamics of Doctor-Patient Communication During Remote Consultations: Qualitative Study Among Norwegian Contract General Practitioners (e57679)
Børge Norberg, Bjarne Austad, Eli Kristiansen, Paolo Zanaboni, Linn Getz. . . . 5359

AI Interventions to Alleviate Healthcare Shortages and Enhance Work Conditions in Critical Care: Qualitative Analysis (e50852)
Nadine Bienefeld, Emanuela Keller, Gudela Grote. . . . 5395

Usefulness of Automatic Speech Recognition Assessment of Children With Speech Sound Disorders: Validation Study (e60520)
Do Kim, Joo Jeong, Dayoung Kang, Taekyung Ahn, Yeonjung Hong, Younggon Im, Jaewon Kim, Min Kim, Dae-Hyun Jang. . . . 5409

A Supervised Explainable Machine Learning Model for Perioperative Neurocognitive Disorder in Liver-Transplantation Patients and External Validation on the Medical Information Mart for Intensive Care IV Database: Retrospective Study (e55046)
Zhendong Ding, Linan Zhang, Yihan Zhang, Jing Yang, Yuheng Luo, Mian Ge, Weifeng Yao, Ziqing Hei, Chaojin Chen. . . . 5420

Ten Machine Learning Models for Predicting Preoperative and Postoperative Coagulopathy in Patients With Trauma: Multicenter Cohort Study (e66612)
Xiaojuan Xiong, Hong Fu, Bo Xu, Wang Wei, Mi Zhou, Peng Hu, Yunqin Ren, Qingxiang Mao. . . . 5433

AI-Driven Innovations for Early Sepsis Detection by Combining Predictive Accuracy With Blood Count Analysis in an Emergency Setting: Retrospective Study (e56155)
Tai-Han Lin, Hsing-Yi Chung, Ming-Jr Jian, Chih-Kai Chang, Hung-Hsin Lin, Chiung-Tzu Yen, Sheng-Hui Tang, Pin-Ching Pan, Cherng-Lih Perng, Feng-Yee Chang, Chien-Wen Chen, Hung-Sheng Shang. . . . 5450

Using Large Language Models to Detect and Understand Drug Discontinuation Events in Web-Based Forums: Development and Validation Study (e54601)
William Trevena, Xiang Zhong, Michelle Alvarado, Alexander Semenov, Alp Oktay, Devin Devlin, Aarya Gohil, Sai Chittimouju. . . . 5481

Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study (e63377)
Godwin Giebel, Pascal Raszke, Hartmuth Nowak, Lars Palmowski, Michael Adamzik, Philipp Heinz, Marianne Tokic, Nina Timmesfeld, Frank Brunkhorst, Jürgen Wasem, Nikola Blase. . . . 5498

Understanding Citizens' Response to Social Activities on Twitter in US Metropolises During the COVID-19 Recovery Phase Using a Fine-Tuned Large Language Model: Application of AI (e63824)
Ryuichi Saito, Sho Tsugawa. . . . 5541

Large Language Models–Supported Thrombectomy Decision-Making in Acute Ischemic Stroke Based on Radiology Reports: Feasibility Qualitative Study (e48328)
Jonathan Kottlors, Robert Hahnfeldt, Lukas Görtz, Andra-Iza Iuga, Philipp Fervers, Johannes Bremm, David Zopfs, Kai Laukamp, Oezguer Onur, Simon Lennartz, Michael Schönfeld, David Maintz, Christoph Kabbasch, Thorsten Persigehl, Marc Schlamann. . . . 5558

Artificial Intelligence in Lymphoma Histopathology: Systematic Review (e62851)
Yao Fu, Zongyao Huang, Xudong Deng, Linna Xu, Yang Liu, Mingxing Zhang, Jinyi Liu, Bin Huang. . . . 5566

Future Use of AI in Diagnostic Medicine: 2-Wave Cross-Sectional Survey Study (e53892)
Bernardo Cabral, Luiza Braga, Carlos Conte Filho, Bruno Penteado, Sandro Freire de Castro Silva, Leonardo Castro, Marcelo Fornazin, Fabio Mota. . . . 5588

Deep Learning–Based Electrocardiogram Model (EIANet) to Predict Emergency Department Cardiac Arrest: Development and External Validation Study (e67576)
Shao-Chi Lu, Guang-Yuan Chen, An-Sheng Liu, Jen-Tang Sun, Jun-Wan Gao, Chien-Hua Huang, Chu-Lin Tsai, Li-Chen Fu. . . . 5616

XSL•FO
RenderX

Large Language Models' Accuracy in Emulating Human Experts' Evaluation of Public Sentiments about Heated Tobacco Products on Social Media: Evaluation Study (e63631)
Kwanho Kim, Soojong Kim. . . . 5631

Investigating Whether AI Will Replace Human Physicians and Understanding the Interplay of the Source of Consultation, Health-Related Stigma, and Explanations of Diagnoses on Patients' Evaluations of Medical Consultations: Randomized Factorial Experiment (e66760)
Weiqi Guo, Yang Chen. . . . 5642

Augmenting Insufficiently Accruing Oncology Clinical Trials Using Generative Models: Validation Study (e66821)
Samer El Kababji, Nicholas Mitsakakis, Elizabeth Jonker, Ana-Alicia Beltran-Bless, Gregory Pond, Lisa Vandermeer, Dhenuka Radhakrishnan, Lucy Mosquera, Alexander Paterson, Lois Shepherd, Bingshu Chen, William Barlow, Julie Gralow, Marie-France Savard, Christian Fesl, Dominik Hlauschek, Marija Balic, Gabriel Rinnerthaler, Richard Greil, Michael Gnant, Mark Clemons, Khaled El Emam. . . . 5655

Perspectives on Using Artificial Intelligence to Derive Social Determinants of Health Data From Medical Records in Canada: Large Multijurisdictional Qualitative Study (e52244)
Victoria Davis, Jinfan Qiang, Itunuoluwa Adekoya MacCarthy, Dana Howse, Abigail Seshie, Leanne Kosowan, Alannah Delahunty-Pike, Eunice Abaga, Jane Cooney, Marjeiry Robinson, Dorothy Senior, Alexander Zsager, Kris Aubrey-Bassler, Mandi Irwin, Lois Jackson, Alan Katz, Emily Marshall, Nazeem Muhajarine, Cory Neudorf, Stephanie Garies, Andrew Pinto. . . . 5683

Generative AI–Enabled Therapy Support Tool for Improved Clinical Outcomes and Patient Engagement in Group Therapy: Real-World Observational Study (e60435)
Johanna Habicht, Larisa-Maria Dina, Jessica McFadyen, Mona Stylianou, Ross Harper, Tobias Hauser, Max Rollwage. . . . 5698

Caregiving Artificial Intelligence Chatbot for Older Adults and Their Preferences, Well-Being, and Social Connectivity: Mixed-Method Study (e65776)
Brooke Wolfe, Yoo Oh, Hyesun Choung, Xiaoran Cui, Joshua Weinzapfel, R Cooper, Hae-Na Lee, Rebecca Lehto. . . . 5713

Machine Learning–Based Explainable Automated Nonlinear Computation Scoring System for Health Score and an Application for Prediction of Perioperative Stroke: Retrospective Study (e58021)
Mi-Young Oh, Hee-Soo Kim, Young Jung, Hyung-Chul Lee, Seung-Bo Lee, Seung Lee. . . . 5726

Public Disclosure of Results From Artificial Intelligence/Machine Learning Research in Health Care: Comprehensive Analysis of ClinicalTrials.gov, PubMed, and Scopus Data (2010-2023) (e60148)
Shoko Maru, Ryohei Kuwatsuru, Michael Matthias, Ross Simpson Jr. . . . 5740

Explainable AI for Intraoperative Motor-Evoked Potential Muscle Classification in Neurosurgery: Bicentric Retrospective Study (e63937)
Qendresa Parduzi, Jonathan Wermelinger, Simon Koller, Murat Sariyar, Ulf Schneider, Andreas Raabe, Kathleen Seidel. . . . 5753

Assessing the Adherence of ChatGPT Chatbots to Public Health Guidelines for Smoking Cessation: Content Analysis (e66896)
Lorien Abroms, Artin Yousefi, Christina Wysota, Tien-Chin Wu, David Broniatowski. . . . 5767

ChatGPT for Univariate Statistics: Validation of AI-Assisted Data Analysis in Healthcare Research (e63550)
Michael Ruta, Tony Gaidici, Chase Irwin, Jonathan Lifshitz. . . . 5778

Laypeople's Use of and Attitudes Toward Large Language Models and Search Engines for Health Queries: Survey Study (e64290)
Tamir Mendel, Nina Singh, Devin Mann, Batia Wiesenfeld, Oded Nov. . . . 5787

Assessment of the Efficiency of a ChatGPT-Based Tool, MyGenAssist, in an Industry Pharmacovigilance Department for Case Documentation: Cross-Over Study (e65651)
Alexandre Benaïche, Ingrid Billaut-Laden, Herivelo Randriamihaja, Jean-Philippe Bertocchio. . . . 5797

XSL•FO
RenderX

Patient Perspectives on Conversational Artificial Intelligence for Atrial Fibrillation Self-Management: Qualitative Analysis (e64325)
Ritu Trivedi, Tim Shaw, Brodie Sheahen, Clara Chow, Liliana Laranjo. . . . 5810

Factors Influencing the Intentions of Patients With Inflammatory Rheumatic Diseases to Use a Digital Human for Medication Information: Qualitative Study (e57697)
Lex Haegens, Victor Huiskes, Bart van den Bemt, Charlotte Bekker. . . . 5822

Clinicians' Perceptions and Potential Applications of Robotics for Task Automation in Critical Care: Qualitative Study (e62957)
Jiafeng Song, Rishika Sridhar, Darlene Rogers, Cheryl Hiddleson, Carolyn Davis, Tina Holden, Shanna Ramsey-Haynes, Lisa Reif, Julie Swann, Craig Jabaley, Mary Gullatte, Rishikesan Kamaleswaran. . . . 5832

Impact of the Smarter Safer Homes Solution on Quality of Life and Health Outcomes in Older People Living in Their Own Homes: Randomized Controlled Trial (e59921)
Wei Lu, David Silvera-Tawil, Hwan-Jin Yoon, Liesel Higgins, Qing Zhang, Mohanraj Karunanithi, Julia Bomke, Joshua Byrnes, Jennifer Hewitt, Vanessa Smallbon, Jill Freyne, Deepa Prabhu, Marlien Varnfield. . . . 5846

Use of Artificial Intelligence, Internet of Things, and Edge Intelligence in Long-Term Care for Older People: Comprehensive Analysis Through Bibliometric, Google Trends, and Content Analysis (e56692)
Shuo-Chen Chien, Chia-Ming Yen, Yu-Hung Chang, Ying-Erh Chen, Chia-Chun Liu, Yu-Ping Hsiao, Ping-Yen Yang, Hong-Ming Lin, Tsung-En Yang, Xing-Hua Lu, I-Chien Wu, Chih-Cheng Hsu, Hung-Yi Chiou, Ren-Hua Chung. . . . 5859

Metaverse Clinic for Pregnant Women With Subclinical Hypothyroidism: Prospective Randomized Study (e64634)
Yuanyuan Zheng, Yizhen Chen, Yan Chen, Liang Lin, Ting Xue, Chuhui Chen, Junping Wen, Wei Lin, Gang Chen. . . . 5878

Pornography-Watching Disorder and Its Risk Factors Among Young Adults: Cross-Sectional Survey (e49860)
Csaba Erd s, Oguz Kelemen, Dávid Pócs, Edit Paulik, András Papp, Edina Horváth, Arbel Golan, Krisztián Széll. . . . 5891

A Comparison of Mobile Social Media Promotion and Volunteer-Driven Strategies for Community Organizations Recruiting Men Who Have Sex with Men for HIV Testing in Zhejiang Province, China: Cross-Sectional Study Based on a Large-Scale Survey (e66702)
Lin He, Shaoqiang Jiang, Tingting Jiang, Wanjun Chen, Jinlei Zheng, Hui Wang, Chengliang Chai. . . . 5905

The Role of the Organization in Promoting Information Security–Related Behavior Among Resident Physicians in Hospitals in Germany: Cross-Sectional Questionnaire Study (e46257)
Judith Kraushaar, Sabine Bohnet-Joschko. . . . 5918

Health Care Professionals and Data Scientists' Perspectives on a Machine Learning System to Anticipate and Manage the Risk of Decompensation From Patients With Heart Failure: Qualitative Interview Study (e54990)
Joana Seringa, Anna Hirata, Ana Pedro, Rui Santana, Teresa Magalhães. . . . 5952

Digital Transformation in Patient Organizations: Interview and Focus Group Study (e62750)
Simon Wallraf, Sara Köthemann, Claudia Wiesemann, Sabine Wöhlke, Marie-Luise Dierks, Marion Schmidt, Henk van Gils-Schmidt, Jonas Lander. . . . 5964

User Personas for eHealth Regarding the Self-Management of Depressive Symptoms in People Living With HIV: Mixed Methods Study (e56289)
Ting Zhao, Chulei Tang, Jun Ma, Huang Yan, Xinyi Su, Xueyuan Zhong, Honghong Wang. . . . 5978

XSL•FO
RenderX

Improving Digital Cancer Care for Older Black Adults: Qualitative Study (e63324)
Paul Wankah, Shivani Chandra, Aisha Lofters, Nebila Mohamednur, Beverley Osei, Tutsirai Makuwaza, Ambreen Sayani. 5994

The Effectiveness of Digital Animation–Based Multistage Education for Patients With Atrial Fibrillation Catheter Ablation: Randomized Clinical Trial (e65685)
Xiaoyu Shi, Yijun Wang, Yuhong Wang, Jun Wang, Chen Peng, Siyi Cheng, Lingpeng Song, Rui Li, Fuding Guo, Zeyan Li, Shoupeng Duan, Xiaomeng Yang, Liping Zhou, Hong Jiang, Lilei Yu. 6006

Associations Among Online Health Information Seeking Behavior, Online Health Information Perception, and Health Service Utilization: Cross-Sectional Study (e66683)
Hongmin Li, Dongxu Li, Min Zhai, Li Lin, ZhiHeng Cao. 6019

Identifying Contextual Factors That Shape Cybersecurity Risk Perception for Assisted Living and Health Care Technologies and Wearables: Mixed Methods Study (e64388)
Alvhild Skjelvik, Nicholas West, Matthias Görges. 6036

Business Venturing in Regulated Markets—Taxonomy and Archetypes of Digital Health Business Models in the European Union: Mixed Methods Descriptive and Exploratory Study (e65725)
Sascha Weimar, Rahel Martjan, Orestis Terzidis. 6054

Evaluation and Comparison of the Academic Quality of Open-Access Mega Journals and Authoritative Journals: Disruptive Innovation Evaluation (e59598)
Yuyan Jiang, Xue-li Liu, Liyun Wang. 6071

Quantifying Public Engagement With Science and Malinformation on COVID-19 Vaccines: Cross-Sectional Study (e64679)
David Grimes, David Gorski. 6085

Subtyping Social Determinants of Health in the "All of Us" Program: Network Analysis and Visualization Study (e48775)
Suresh Bhavnani, Weibin Zhang, Daniel Bao, Mukaila Raji, Veronica Ajewole, Rodney Hunter, Yong-Fang Kuo, Susanne Schmidt, Monique Pappadis, Elise Smith, Alex Bokov, Timothy Reistetter, Shyam Visweswaran, Brian Downer. 6095

Using Structured Codes and Free-Text Notes to Measure Information Complementarity in Electronic Health Records: Feasibility and Validation Study (e66910)
Tom Seinen, Jan Kors, Erik van Mulligen, Peter Rijnbeek. 6118

Changing User Experience of Wearable Activity Monitors Over 7 Years: Repeat Cross-Sectional Survey Study (e56251)
Darcy Beckett, Rachel Curtis, Kimberley Szeto, Carol Maher. 6141

Experiences and Expectations of Immigrant and Nonimmigrant Older Adults Regarding eHealth Services: Qualitative Interview Study (e64249)
Maedeh Ghorbanian Zolbin, Sari Kujala, Isto Huvila. 6154

Satisfactory Evaluation of Call Service Using AI After Ureteral Stent Insertion: Randomized Controlled Trial (e56039)
Ukrae Cho, Yong Gwon, Seung Chong, Ji Han, Do Kim, Seung Doo, Won Yang, Kyeongmin Kim, Sung Shim, Jaehun Jung, Jae Kim. 6171

Effect of Artificial Intelligence Helpfulness and Uncertainty on Cognitive Interactions with Pharmacists: Randomized Controlled Trial (e59946)
Chuan-Ching Tsai, Jin Kim, Qiyuan Chen, Brigid Rowell, X Yang, Raed Kontar, Megan Whitaker, Corey Lester. 6180

XSL•FO
RenderX

Optimizing Engagement With a Smartphone App to Prevent Violence Against Adolescents: Results From a Cluster Randomized Factorial Trial in Tanzania (e60102)
Roselinde Janowski, Lucie Cluver, Yulia Shenderovich, Joyce Wamoyi, Mwita Wambura, David Stern, Lily Clements, G Melendez-Torres, Lauren Baerecke, Abigail Ornellas, Angelique Chetty, Jonathan Klapwijk, Laetitia Christine, Ateamate Mukabana, Esmee Te Winkel, Anna Booij, Gervas Mbosoli, Jamie Lachman. . . . 6193

How Authoritative Media and Personal Social Media Influence Policy Compliance Through Trust in Government and Risk Perception: Quantitative Cross-Sectional Survey Study (e64940)
Hua Zhang, Cheng Yang, Xiuxian Deng, Chunyan Luo. . . . 6276

Public Health Messaging on Twitter During the COVID-19 Pandemic: Observational Study (e63910)
Ashwin Rao, Nazanin Sabri, Siyi Guo, Louiqa Raschid, Kristina Lerman. . . . 6292

Effectiveness of Continuous Glucose Monitoring on Short-Term, In-Hospital Mortality Among Frail and Critically Ill Patients With COVID-19: Randomized Controlled Trial (e67012)
Jiawei Shang, Ziming Yuan, Zuoyan Zhang, Quanhong Zhou, Yan Zou, Wei Wang. . . . 6312

Leveraging Large Language Models for Infectious Disease Surveillance—Using a Web Service for Monitoring COVID-19 Patterns From Self-Reporting Tweets: Content Analysis (e63190)
Jiacheng Xie, Ziyang Zhang, Shuai Zeng, Joel Hilliard, Guanghui An, Xiaoting Tang, Lei Jiang, Yang Yu, Xiufeng Wan, Dong Xu. . . . 6322

The Impact of Stay-At-Home Mandates on Uncertainty and Sentiments: Quasi-Experimental Study (e64667)
Carolina Biliotti, Nicolò Fraccaroli, Michelangelo Puliga, Falco Bargagli-Stoffi, Massimo Riccaboni. . . . 6341

Exploring Psychological Trends in Populations With Chronic Obstructive Pulmonary Disease During COVID-19 and Beyond: Large-Scale Longitudinal Twitter Mining Study (e54543)
Chunyan Zhang, Ting Wang, Caixia Dong, Duwei Dai, Linyun Zhou, Zongfang Li, Songhua Xu. . . . 6357

Application of a 3D Fusion Model to Evaluate the Efficacy of Clear Aligner Therapy in Malocclusion Patients: Prospective Observational Study (e67378)
Chaofeng Liu, Yan Liu, Chunyan Yi, Tao Xie, Jingjun Tian, Peishen Deng, Changyu Liu, Yan Shan, Hangyu Dong, Yanhua Xu. . . . 6373

Examining Intersectionality and Barriers to the Uptake of Video Consultations Among Older Adults From Disadvantaged Backgrounds With Limited English Proficiency: Qualitative Narrative Interview Study (e65690)
Laiba Husain, Trisha Greenhalgh. . . . 6430

Digital Transformation of Rheumatology Care in Germany: Cross-Sectional National Survey (e52601)
Susann May, Robert Darkow, Johannes Knitza, Katharina Boy, Philipp Klemm, Martin Heinze, Nicolas Vuillerme, Pascal Petit, Patricia Steffens-Korbanka, Heike Kladny, Johannes Hornig, Peer Aries, Martin Welcker, Felix Muehlensiepen. . . . 6446

Exploration of Reproductive Health Apps' Data Privacy Policies and the Risks Posed to Users: Qualitative Content Analysis (e51517)
Nina Zadushlivy, Rizwana Biviji, Karmen Williams. . . . 6456

The Combined Effect of Multisensory Stimulation and Therapist Support on Physical and Mental Health of Older Adults Living in Nursing Homes: Pilot Randomized Controlled Trial (e55042)
Sewar Khatib, Yuval Palgi, Yoni Ashar, Natalya Polyvyannaya, Pavel Goldstein. . . . 6473

Patient-Centered mHealth Intervention to Improve Self-Care in Patients With Chronic Heart Failure: Phase 1 Randomized Controlled Trial (e55586)
Spyros Kitsiou, Ben Gerber, Susan Buchholz, Mayank Kansal, Jiehuan Sun, Susan Pressler. . . . 6487

XSL•FO
RenderX

Relative Preference for In-Person, Telehealth, Digital, and Pharmacologic Mental Health Care After the COVID-19 Pandemic: Cross-Sectional Questionnaire Study (e54608)
E Parsons, Zoë Figueroa, Michele Hiserodt, Talea Cornelius, Michael Otto. 6517

Pathways to Sustainable Health Care Development: Study on the Carbon Reduction Potential of Telemedicine in China (e63927)
Zhi Li, Lianrui Xiang, Jing Ning, Wenbo Li, Yong Huang, Xue Xiao. 6527

Application of Large Language Models in Medical Training Evaluation—Using ChatGPT as a Standardized Patient: Multimetric Assessment (e59435)
Chenxu Wang, Shuhan Li, Nuoxi Lin, Xinyu Zhang, Ying Han, Xiandi Wang, Di Liu, Xiaomei Tan, Dan Pu, Kang Li, Guangwu Qian, Rong Yin.
6 5 3 9

Performance Evaluation of Large Language Models in Cervical Cancer Management Based on a Standardized Questionnaire: Comparative Study (e63626)
Warisijiang Kuerbanjiang, Shengzhe Peng, Yiershatijiang Jiamaliding, Yuexiong Yi. 6573

Unveiling GPT-4V's hidden challenges behind high accuracy on USMLE questions: Observational Study (e65146)
Zhichao Yang, Zonghai Yao, Mahbuba Tasmin, Parth Vashisht, Won Jang, Feiyun Ouyang, Beining Wang, David McManus, Dan Berlowitz, Hong Yu. 6592

Retrieval Augmented Therapy Suggestion for Molecular Tumor Boards: Algorithmic Development and Validation Study (e64364)
Eliza Berman, Holly Sundberg Malek, Michael Bitzer, Nisar Malek, Carsten Eickhoff. 6607

Accuracy of Large Language Models for Literature Screening in Thoracic Surgery: Diagnostic Study (e67488)
Zhang-Yi Dai, Fu-Qiang Wang, Cheng Shen, Yan-Li Ji, Zhi-Yang Li, Yun Wang, Qiang Pu. 6618

Assessing Racial and Ethnic Bias in Text Generation by Large Language Models for Health Care–Related Tasks: Cross-Sectional Study (e57257)
John Hanna, Abdi Wakene, Andrew Johnson, Christoph Lehmann, Richard Medford. 6628

Improving Dietary Supplement Information Retrieval: Development of a Retrieval-Augmented Generation System With Large Language Models (e67677)
Yu Hou, Jeffrey Bishop, Hongfang Liu, Rui Zhang. 6637

Large Language Model–Based Assessment of Clinical Reasoning Documentation in the Electronic Health Record Across Two Institutions: Development and Validation Study (e67967)
Verity Schaye, David DiTullio, Benedict Guzman, Scott Vennemeyer, Hanniel Shih, Ilan Reinstein, Danielle Weber, Abbie Goodman, Danny Wu, Daniel Sartori, Sally Santen, Larry Gruppen, Yindalon Aphinyanaphongs, Jesse Burk-Rafel. 6650

Large Language Model–Driven Knowledge Graph Construction in Sepsis Care Using Multicenter Clinical Databases: Development and Usability Study (e65537)
Hao Yang, Jiaxi Li, Chi Zhang, Alejandro Sierra, Bairong Shen. 6666

Building a Digital Health Research Platform to Enable Recruitment, Enrollment, Data Collection, and Follow-Up for a Highly Diverse Longitudinal US Cohort of 1 Million People in the All of Us Research Program: Design and Implementation Study (e60189)
Dave Klein, Aisha Montgomery, Mark Begale, Scott Sutherland, Sherilyn Sawyer, Jacob McCauley, Letheshia Husbands, Deepti Joshi, Alan Ashbeck, Marcy Palmer, Praduman Jain. 6736

XSL•FO
RenderX

Short Papers

How TikTok Influencers Disclose Food and Beverage Brand Partnerships: Descriptive Study (e60891)
Roxanne Dupuis, Aviva Musicus, Brittany Edghill, Emma Keteku, Marie Bragg. 1499

The Effect of Nurse Navigators in Digital Remote Monitoring in Cancer Care: Case Study Using Structural Equation Modeling (e66275)
Etienne Minvielle, Joel Perez-Torrents, Israa Salma, Philippe Aegerter, Marie Ferrua, Charles Ferté, Henri Leleu, Delphine Mathivon, Claude Sicotte, Mario Di Palma, Florian Scotté. 3610

Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical Question Answering Using Reddit Data: Proof-of-Concept Study (e66220)
Sudeshna Das, Yao Ge, Yuting Guo, Swati Rajwal, JaMor Hairston, Jeanne Powell, Drew Walker, Snigdha Peddireddy, Sahithi Lakamana, Selen Bozkurt, Matthew Reyna, Reza Sameni, Yunyu Xiao, Sangmi Kim, Rasheeta Chandler, Natalie Hernandez, Danielle Mowery, Rachel Wightman, Jennifer Love, Anthony Spadaro, Jeanmarie Perrone, Abeed Sarker. 5375

Viewpoints

Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine (e59069)
Kuo Zhang, Xiangbin Meng, Xiangyu Yan, Jiaming Ji, Jingqian Liu, Hua Xu, Heng Zhang, Da Liu, Jingjia Wang, Xuliang Wang, Jun Gao, Yuan-geng-shuo Wang, Chunli Shao, Wenyao Wang, Jiarong Li, Ming-Qi Zheng, Yaodong Yang, Yi-Da Tang. 1524

The CeHRes Roadmap 2.0: Update of a Holistic Framework for Development, Implementation, and Evaluation of eHealth Technologies (e59601)
Hanneke Kip, Nienke Beerlage-de Jong, Lisette van Gemert-Pijnen, Saskia Kelders. 1539

From Theory to Practice: Viewpoint on Economic Indicators for Trust in Digital Health (e59111)
Felix Gille, Laura Maaß, Benjamin Ho, Divya Srivastava. 1556

Era of Generalist Conversational Artificial Intelligence to Support Public Health Communications (e69007)
Emre Sezgin, Ahmet Kocaballi. 1567

Identifying Fraudulent Responses in a Study Exploring Delivery Options for Pregnancies Impacted by Gestational Diabetes: Lessons Learned From a Web-Based Survey (e58450)
Emma Ruby, Serine Ramlawi, Alexa Bowie, Stephanie Boyd, Alysha Dingwall-Harvey, Ruth Rennicks White, Darine El-Chaâr, Mark Walker. .
1 5 7 6

AI Can Be a Powerful Social Innovation for Public Health if Community Engagement Is at the Core (e68198)
Alessandra Bazzano, Andrea Mantsios, Nicholas Mattei, Michael Kosorok, Aron Culotta. 1584

Regulation and Compliance in Telemedicine: Viewpoint (e53558)
Julia Ivanova, Mollie Cummins, Triton Ong, Hiral Soni, Janelle Barrera, Hattie Wilczewski, Brandon Welch, Brian Bunnell. 1589

From Stories to Solutions: A Research Cycle Framework for Enhancing Trustworthiness in Studies of Online Patient Narratives (e58310)
Klay Lamprell, Diana Pulido, Gaston Arnolda, Bróna Easpaig, Yvonne Tran, Jeffrey Braithwaite. 1596

Designing Health Recommender Systems to Promote Health Equity: A Socioecological Perspective (e60138)
Caroline Figueroa, Helma Torkamaan, Ananya Bhattacharjee, Hanna Hauptmann, Kathleen Guan, Gayane Sedrakyan. 1602

Why AI Monitoring Faces Resistance and What Healthcare Organizations Can Do About It: An Emotion-Based Perspective (e51785)
Karl Werder, Lan Cao, Eun Park, Balasubramaniam Ramesh. 1611

XSL•FO
RenderX

Publication Counts in Context: Normalization Using Query and Reference Terms in PubMed (e60616)
Julian Varghese, Lucas Bickmann, Timo Strünker, Nina Neuhaus, Frank Tüttelmann, Sarah Sandmann. 1619

Improving the Implementation of Patient-Reported Outcome Measure in Clinical Practice: Tackling Current Challenges With Innovative Digital Communication Technologies (e60777)
Kelly de Ligt, Saar Hommes, Ruben Vromans, Eva Boomstra, Lonneke van de Poll, Emiel Krahmer. 1624

A Hierarchical Framework for Selecting Reference Measures for the Analytical Validation of Sensor-Based Digital Health Technologies (e58956)
Jessie Bakker, Samantha McClenahan, Piper Fromy, Simon Turner, Barry Peterson, Benjamin Vandendriessche, Jennifer Goldsack. 1634

The Application of Preventive Medicine in the Future Digital Health Era (e59165)
Katherine De la Torre, Sukhong Min, Hyobin Lee, Daehee Kang. 1646

Empowering Health Care Actors to Contribute to the Implementation of Health Data Integration Platforms: Retrospective of the medEmotion Project (e68083)
Marcel Parciak, Noëlla Pierlet, Liesbet Peeters. 1661

Opportunities and Challenges in Using Electronic Health Record Systems to Study Postacute Sequelae of SARS-CoV-2 Infection: Insights From the NIH RECOVER Initiative (e59217)
Hannah Mandel, Shruti Shah, L Bailey, Thomas Carton, Yu Chen, Shari Esquenazi-Karonika, Melissa Haendel, Mady Hornig, Rainu Kaushal, Carlos Oliveira, Alice Perlowski, Emily Pfaff, Suchitra Rao, Hanieh Razzaghi, Elle Seibert, Gelise Thomas, Mark Weiner, Lorna Thorpe, Jasmin Divers, RECOVER EHR Cohort. 1672

Emerging Domains for Measuring Health Care Delivery With Electronic Health Record Metadata (e64721)
Daniel Tawfik, Adam Rule, Aram Alexanian, Dori Cross, A Holmgren, Sunny Lou, Eugenia McPeek Hinz, Christian Rose, Ratnalekha Viswanadham, Rebecca Mishuris, Jorge Rodríguez-Fernández, Eric Ford, Sarah Florig, Christine Sinsky, Nate Apathy. 1684

Health Communication on the Internet: Promoting Public Health and Exploring Disparities in the Generative AI Era (e66032)
Jamal Uddin, Cheng Feng, Junfang Xu. 1693

Assessing Digital Maturity of Hospitals: Viewpoint Comparing National Approaches in Five Countries (e57858)
Kathrin Cresswell, Franziska Jahn, Line Silsand, Leanna Woods, Tim Postema, Marion Logan, Sevala Malkic, Elske Ammenwerth. 1702

Integration of Psychiatric Advance Directives Into the Patient-Accessible Electronic Health Record: Exploring the Promise and Limitations (e68549)
Julian Schwarz, Eva Meier-Diedrich, Matthé Scholten, Lucy Stephenson, John Torous, Florian Wurster, Charlotte Blease. 1711

Lessons Learned From European Health Data Projects With Cancer Use Cases: Implementation of Health Standards and Internet of Things Semantic Interoperability (e66273)
Amelie Gyrard, Somayeh Abedian, Philip Gribbon, George Manias, Rick van Nuland, Kurt Zatloukal, Irina Nicolae, Gabriel Danciu, Septimiu Nechifor, Luis Marti-Bonmati, Pedro Mallol, Stefano Dalmiani, Serge Autexier, Mario Jendrossek, Ioannis Avramidis, Eva Garcia Alvarez, Petr Holub, Ignacio Blanquer, Anna Boden, Rada Hussein. 1722

Enhancing the Innovation Ecosystem: Overcoming Challenges to Introducing Information-Driven Technologies in Health Care (e56836)
Julie Reed, Petra Svedberg, Jens Nygren. 1738

Balancing Risks and Opportunities: Data-Empowered-Health Ecosystems (e57237)
Lan Li, Emma Back, Suna Lee, Rebecca Shipley, Néo Mapitse, Stefan Elbe, Melanie Smallman, James Wilson, Ifat Yasin, Geraint Rees, Ben Gordon, Virginia Murray, Stephen Roberts, Anna Cupani, Patty Kostkova. 1750

Dying in Darkness: Deviations From Data Sharing Ethics in the US Public Health System and the Data Genocide of American Indian and Alaska Native Communities (e70983)
Cason Schmit, Meghan O'Connell, Sarah Shewbrooks, Charles Abourezk, Fallon Cochlin, Megan Doerr, Hye-Chung Kum. 1760

XSL•FO
RenderX

Mindfulness-Based Cognitive Therapy–Game: An Ironic Way to Treat Internet Gaming Disorder (e65786)
Jaehyun Kim, Hayoung Oh, Anderson Yoon. . . . 1771

The Paradigm Shift From Patient to Health Consumer: 20 Years of Value Assessment in Health (e60443)
Eline van den Broek-Altenburg, Adam Atherly. . . . 6708

Tutorial

How to Design, Create, and Evaluate an Instruction-Tuning Dataset for Large Language Model Training in Health Care: Tutorial From a Clinical Perspective (e70481)
Wojciech Nazar, Grzegorz Nazar, Aleksandra Kami ska, Ludmila Danilowicz-Szymanowicz. . . . 1787

Letters to the Editor

Advantages and Inconveniences of a Multi-Agent Large Language Model System to Mitigate Cognitive Biases in Diagnostic Challenges (e69742)
Cedric Bousquet, Divà Beltramin. . . . 6215

Japan's Telepsychiatry Dissemination: Current Status and Challenges (e22849)
Shotaro Kinoshita, Taishiro Kishimoto. . . . 6217

Authors' Reply: Japan's Telepsychiatry Dissemination: Current Status and Challenges (e69798)
Yuka Sugawara, Yosuke Hirakawa, Masao Iwagami, Haruo Kuroki, Shuhei Mitani, Ataru Inagaki, Hiroki Ohashi, Mitsuru Kubota, Soichi Koike, Rie Wakimizu, Masaomi Nangaku. . . . 6219

Authors' Reply: Commentary on "Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line" (e59734)
Anthony Pisani, Carlos Gallo, Madelyn Gould, Nitya Kanuri, John Marcotte, Brian Pascal, David Rousseau, Megan Ranney, Bob Filbin, Shairi Turner. . . . 6221

Response From Crisis Text Line to "Commentary on 'Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line'" (e64015)
Dena Trujillo. . . . 6225

Advancing Insights Into Postoperative Sleep Quality and Influencing Factors (e69193)
Yining Zhao, Xin Hu. . . . 6229

Authors' Reply: Advancing Insights Into Postoperative Sleep Quality and Influencing Factors (e70168)
Chen Shang, Ya Yang, Chengcheng He, Junqi Feng, Yan Li, Meimei Tian, Zhanqi Zhao, Yuan Gao, Zhe Li. . . . 6231

Enhancing the Clinical Relevance of Al Research for Medication Decision-Making (e70657)
Qi Wang, Mingxian Chen. . . . 6234

Authors' Reply: Enhancing the Clinical Relevance of Al Research for Medication Decision-Making (e72007)
Sarah Vordenberg, Julianna Nichols, Vincent Marshall, Kristie Weir, Michael Dorsch. . . . 6236

Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders (e70020)
Alina Yang. . . . 6238

XSL•FO
RenderX

Authors' Reply: "Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders" (e71897)
Callum Lewis, Melody Taba, Tiffany Allen, Patrina Caldwell, S Skinner, Melissa Kang, Hamish Henderson, Liam Bray, Madeleine Borthwick, Philippa Collin, Kirsten McCaffery, Karen Scott. . . . 6240

Advancing Digital Health Integration in Oncology (e70316)
Rai Khan, Hassan Tariq. . . . 6243

Authors' Reply: Advancing Digital Health Integration in Oncology (e72477)
Yura Lee, Ye-Eun Park. . . . 6245

Enhancing Digital Health Interventions for Medication Adherence: Considerations for Broader Applicability and Long-Term Impact (e69204)
ShanShan Du, Yining Zhao. . . . 6247

Impact of Social Media Influencers on Amplifying Positive Public Health Messages (e73062)
Gerard Flaherty, Ryan Mangan. . . . 6249

The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease (e69927)
Siqi Zhang, Tielong Chen. . . . 6251

Authors' Reply: The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease (e70247)
Xinyue Zhang. . . . 6253

Addressing the Digital Divide Among the Older Population Presents a Substantial Challenge (e69482)
Su-hang Xie. . . . 6255

Authors' Reply: Addressing the Digital Divide Among the Older Population Presents a Substantial Challenge (e72565)
Lakshmi Kannan, Tanvi Bhatt. . . . 6257

Corrigenda and Addendas

Correction: Factors Associated With Digital Intervention Engagement and Adherence in Patients With Cancer: Systematic Review (e71370)
Lucile Montalescot, Louise Baussard, Elodie Charbonnier. . . . 6259

Correction: Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study (e71665)
Gilly Hendrie, Danielle Baird, Genevieve James-Martin, Emily Brindal, Paige Brooker. . . . 6261

Correction: Internet-Delivered Psychoeducation (SCOPE) for Transition-Aged Autistic Youth: Pragmatic Randomized Controlled Trial (e72209)
Anna Backman, Lise Roll-Pettersson, Are Mellblom, Elisabet Norman-Claesson, Emma Sundqvist, Eric Zander, Sarah Vigerland, Tatja Hirvikoski. . . . 6263

Correction: Teledentistry Applied to Health and Education Outcomes: Evidence Gap Map (e71577)
Júlia Nascimento da Silva Mulder, Marcelo Ramos Pinto, Isabelle Aníbal, Ana Dornellas, Deise Garrido, Camila Huanca, Ana Haddad, Carmen Mendes Abdala. . . . 6265

Correction: Digital Interventions for Symptoms of Borderline Personality Disorder: Systematic Review and Meta-Analysis (e70850)
Julia Lindsay, Niall McGowan, Thomas Henning, Eli Harriss, Kate Saunders. . . . 6267

XSL•FO
RenderX

Correction: Examining the Effectiveness of Electronic Patient-Reported Outcomes in People With Cancer: Systematic Review and Meta-Analysis (e72687)
Melissa Perry, Sally Taylor, Binish Khatoon, Amy Vercell, Corinne Faivre-Finn, Galina Velikova, Antonia Marsden, Calvin Heal, Janelle Yorke. . 6 2 6 9

Correction: Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives (e72519)
Bryan Cleal, Yanbing Chen, Mandy Wäldchen, Hanne Ballhausen, Drew Cooper, Shane O'Donnell, Christine Knoll, Niklas Krug, Klemens Raile, Tebbe Ubben, Adrian Tappe, Dana Lewis, Ingrid Willaing, Timothy Skinner, Katarina Braune. 6274

Research Letters

Geographical Disparities in Research Misconduct: Analyzing Retraction Patterns by Country (e65775)
Paul Sebo, Melissa Sebo. 6389

Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study (e64993)
Joshua Biro, Jessica Handley, Nathan Cobb, Varsha Kottamasu, Jeffrey Collins, Seth Krevat, Raj Ratwani. 6395

Real-Time fMRI Neurofeedback Modulation of Dopaminergic Midbrain Activity in Young Adults With Elevated Internet Gaming Disorder Risk: Randomized Controlled Trial (e64687)
Anqi Gu, Cheng Chan, Xiaolei Xu, Joseph Dexter, Benjamin Becker, Zhiying Zhao. 6398

Self-Reported Side Effects Associated With Selective Androgen Receptor Modulators: Social Media Data Analysis (e65031)
Aditya Joshi, Diego Kaune, Phillip Leff, Emily Fraser, Sarah Lee, Morgan Harrison, Moustafa Hazin. 6404

Perspectives of Black, Latinx, Indigenous, and Asian Communities on Health Data Use and AI: Cross-Sectional Survey Study (e50708)
Fatuma-Ayaan Rinderknecht, Vivian Yang, Mekaleya Tilahun, Jenna Lester. 6409

Gender Differences in X (Formerly Twitter) Use Among Oncology Physicians at National Cancer Institute–Designated Cancer Centers: Cross-Sectional Study (e66054)
Vivian Tieu, Sungjin Kim, Minji Seok, Leslie Ballas, Mitchell Kamrava, Katelyn Atkins. 6414

eHealth Literacy 3.0: Updating the Norman and Skinner 2006 Model (e70112)
Ariesta Milanti, Cameron Norman, Dorothy Chan, Winnie So, Harvey Skinner. 6420

Primary Care Practice Factors Associated With Telehealth Adoption in the United States: Cross-Sectional Survey Analysis (e70404)
Matthew Mackwood, Elliott Fisher, Rachel Schmidt, A O'Malley, Hector Rodriguez, Stephen Shortell, Ellesse-Roselee Akré, Alena Berube, Karen Schifferdecker. 6424

News and Perspective

The Controversial Push for New Brain and Neurorights (e72270)
Simon Spichak. 6820

XSL•FO
RenderX

Editorial

Crisis Text Line and Loris.ai Controversy Highlights the Complexity of Informed Consent on the Internet and Data-Sharing Ethics for Machine Learning and Research

Gunther Eysenbach[1], MPH, MD

JMIR Publications, Toronto, ON, Canada

Corresponding Author:
Gunther Eysenbach, MPH, MD
JMIR Publications
130 Queens Quay East
Suite 1100-1102
Toronto, ON, M5A 0P6
Canada
Phone: 1 416 786 6970
Email: geysenba@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2019/1/e11507/

Comment on: https://www.jmir.org/2024/1/e42144/

Comment on: https://www.jmir.org/2025/1/e59734/

Comment on: https://www.jmir.org/2025/1/e64015/

Comment on: https://www.jmir.org/2024/1/e67880/

(J Med Internet Res 2025;27:e67878) doi:10.2196/67878

KEYWORDS

data ethics; data sharing; informed consent; disclosure; conflict of interest; transparency; trust

In January 2019, the *Journal of Medical Internet Research* published a viewpoint paper by Anthony Pisani and colleagues [1] related to a noncommercial data-sharing program for academic researchers offered by Crisis Text Line, a not-for-profit technology organization that provides a free 24-7 text line for people in crisis in the United States. Crisis Text Line has been described as a globally prominent online mental health support resource, where people in crisis (eg, with suicidal ideation) can exchange text messages with trained volunteer counselors. As is the case for many internet companies, these digital exchanges are a treasure trove for machine learning, quality improvement, education, and research.

Nancy Lublin, cofounder and former chief executive officer (CEO) of Crisis Text Line (as well as Loris.ai), gave an example in her 2015 TED talk on the internal use of these data [2]:

> *We know that if you text the words "numbs" and "sleeve," there's a 99 percent match for cutting. We know that if you text in the words "mg" and "rubber band," there's a 99 percent match for substance abuse. And we know that if you text in "sex," "oral" and "Mormon," you're questioning if you're gay. Now that's interesting information that a counselor could figure out but that algorithm in our hands means that an automatic pop-up says, "99 percent match for cutting -- try asking one of these questions" to prompt the counselor. Or "99 percent match for substance abuse, here are three drug clinics near the texter." It makes us more accurate.*

Former Crisis Text Line board member, Danah Boyd, recounts her motivation behind opening up data to researchers, which is the focus of the 2019 paper in the *Journal of Medical Internet Research* [3]:

> *From early on, researchers came to Crisis Text Line asking for access to data. This prompted even more reflection. We had significant data and we were seeing trends that had significant implications for far more than our service. (...) This then led to the more*

XSL•FO
RenderX

> *complicated issue of whether or not to allow external researchers to study our data with an eye towards scholarship. (...) Our texters come to us in their darkest hours. Our data was opening up internal questions right and left about how to best support them. We don't have the internal resources to analyze the data to answer all of our questions, to improve our knowledge base in ways that can help texters. I knew that having additional help from researchers could help us learn in ways that would improve training of counselors and help people down the line. I also knew that what we were learning internally might be useful to other service providers in the mental health space and I felt queasy that we were not sharing what we had learned to help others.*

The 2019 paper deals with the question of how these textual data could be ethically shared with academic researchers to answer a variety of research questions. As Pisani and colleagues observed, "few companies are willing to take on the potential work and risks involved in noncommercial data sharing, and the scientific and societal potential of their data goes unrealized" [1], so to Crisis Text Line's credit the nonprofit organization applied for a grant from the Robert Wood Johnson Foundation to create a pilot program for data sharing with academic researchers. The paper described "the process of defining core challenges underlying data sharing in technology-academia partnerships; discusses Crisis Text Line's trial solutions to these challenges; and offers lessons learned that might inform other technology companies' data-sharing partnerships" [1]. This is a viewpoint paper, not an original paper, and the paper in the *Journal of Medical Internet Research* does not use any of the data collected by Crisis Text Line nor does it share the results of the subprojects that Crisis Text Line enabled with their data-sharing program, rather it deals with the meta-question of how such data can be ethically shared with researchers. To that end, Crisis Text Line assembled a data ethics committee consisting of external academic researchers and technology experts (many of whom are coauthors of the paper), convened by then Crisis Text Line's chief data scientist and Crisis Text Line cofounder, Bob Filbin, who is also a coauthor on the paper. The ethics committee advised Crisis Text Line on processes and assessed the research proposals of other researchers who applied for access to Crisis Text Line's data. The paper proposes some general guidelines for other organizations that want to share data with third parties in a *noncommercial* context. The paper is seen by many as an important contribution to the complex question of how organizations can ethically share data with academics while preserving the privacy and integrity of the data, as evidenced by the selection of the article as "best paper of 2019" for the 2020 International Medical Informatics Association (IMIA) Yearbook, Special Section on Ethics in Health Informatics [4].

Data Ethics Called in Question

In November 2021, Tim Reierson, a former volunteer for the Crisis Text Line, and an advocate seeking reform of data ethics at Crisis Text Line and 988 Lifeline, published a 13-page open letter in response to the article published in the *Journal of Medical Internet Research*, where he raised several concerns related to informed consent and an alleged conflict of interest [5]; the full document was sent to us in February 2022.

Reierson's core concern related to *informed consent* was that the paper described the process where Crisis Text Line provided "texters with a link to an easy-to-understand Terms of Service" (Table 1 in [1]), thereby—according to Reierson—"establishing a Terms of Service consent standard." While we as *Journal of Medical Internet Research* editors agree that offering a terms of service (ToS) link does not necessarily equate to informed consent, we do not agree with the interpretation that this viewpoint paper establishes this as a generalizable "informed consent standard." Rather, the guidelines proposed in the viewpoint (Table 1 in [1]) include that researchers "inform users in an unobtrusive way that anonymized data are shared with select research partners", which was implemented by Crisis Text Line by referring users to their ToS. But the guidelines proposed in the viewpoint paper also had other critical requirements. Notably, they also require that academics who use (anonymized) Crisis Text Line data for research obtain institutional review board (IRB) approval for their data analysis projects (Table 1 in [1]: "Establish a review process that includes outside academics and ethics experts").

Informed consent is ideal but not always possible, especially on the internet [6]. There are some circumstances, for example, in urgent or emergency care settings, or in public health practice, where informed consent is considered legally effective when considering contextual variables [7]. Additionally, secondary use of data that has been de-identified or anonymized (as was the case with Crisis Text Line data) is usually allowable in the absence of explicit consent from the data subject [8]. Thus, the critique that the Crisis Text Line clientele can not be expected to read the terms of use is valid, but it is ultimately up to the IRBs to assess the risk—which is a key component mentioned in the Pisani paper [1].

Regarding commercial use, Reierson in his letter to JMIR Publications pointed to the fact that Crisis Text Line had also launched a for-profit subsidiary, Loris.ai, to which Crisis Text Line data was licensed in order to train artificial intelligence (AI) to handle text exchanges for customer service. These events took place after the study period.

Less than 2 months after Reierson made the letter to JMIR Publications public (but before we had seen it), on January 28, 2022, Politico ran a story bringing public scrutiny to the fact that Crisis Text Line had created Loris.ai as a profit-generating entity [9], which was met with widespread public "anger and disgust" [3], and prompted a letter from US Federal Communications Commission's commissioner, Brendan Carr, to Crisis Text Line and Loris.ai demanding they stop this practice immediately [10]. Three days later, Crisis Text Line backed down and "ended the data-sharing relationship with Loris" and requested "that Loris delete the data it has received from Crisis Text Line" [11,12].

XSL•FO
RenderX

Publication Ethics and Corrigendum of the Conflicts of Interest Section

Reierson's concern regarding the *conflict of interest* was that "Crisis Text Line itself had a vested monetary interest in commercial use of the data through its' subsidiary Loris.ai, and therefore vested interest in the 2019 paper's finding [*sic*] that the crisis conversation data is ethically sourced for research purposes." Here, it is important to note that from our point of view as *Journal of Medical Internet Research* editors, the term "finding" is slightly misleading because this is not an empirical study, rather it is a viewpoint paper, largely written by members of the independent Data Ethics Committee consisting of independent and respected ethicists and academics, describing how Crisis Text Line handled data sharing for research and outlining ethical challenges and how they were addressed by the Crisis Text Line's Data Ethics Committee.

From a publication ethics perspective, we are less concerned about the ethics of the arrangement between Crisis Text Line and Loris.ai, but primarily about whether any of the individual contributing authors benefitted (or had the potential to benefit) financially from Loris.ai or had any other ties to Loris.ai, as this would arguably be something that should have been disclosed to the editor and reviewers on submission of the paper [13]. Even though the 2019 paper published in the *Journal of Medical Internet Research* was about *noncommercial* data sharing for academic research, any hypothetical ties between authors and a company that benefits from Crisis Text Line data-sharing guidelines could have influenced the viewpoint.

We shared Tim Reiersons' letter with the corresponding author and received a reply on March 30, 2022, reassuring us that the academic members of the Data Ethics Committee were uncompensated and that there were no further conflicts of interest to disclose. We had no evidence of any wrongdoing that would require a retraction or even issuing an editorial expression of concern. Still, we wanted readers to be aware of the debate (in particular around informed consent) and invited Tim Reierson to submit a condensed commentary for publication alongside the article [14], and we also asked the original authors to respond [15]. We also decided to publish a corrigendum [16] to clarify author relationships with the noncommercial and commercial entities in greater detail [16].

Informed Consent in the Age of AI and Machine Learning

The invited commentary by Reierson highlights complex concerns regarding consent for research, data sharing, and machine learning [14], which are of course amplified given the initial intent of Crisis Text Line to share data with a *commercial* entity, Loris.ai.

Tim Reierson asks legitimate questions on whether it is sufficient to have users in crisis accept a ToS document (which in all likelihood nobody reads). Even former board chair of Crisis Text Line, Danah Boyd, readily admitted that a ToS is not consent [3]:

> *I knew how little data exists in the mental health space, how much we had tried to learn from others, how beneficial knowledge could be to others working in the mental health ecosystem. I also knew that people who came to us in crisis were not consenting to be studied. Yes, there was a terms of service that could contractually permit such use, but I knew darn straight that no one would read it, and advised everyone involved to proceed as such.*

Informed consent is a basic tenet for research; on the other hand, it can be waived when data are deidentified. IRBs routinely grant exemptions to the informed consent ideal when data are deidentified. While such assessments can be made in an academic context, many companies do not have IRBs that could assess the risk or ethical implications if data are analyzed, manually or with machine learning, and the reidentification risks are often not known.

The question of what the role of informed consent plays when academics or companies use data generated on the internet is one of the most vexing questions of our time and a topic of ongoing debate. It is also precisely why we published the 2019 viewpoint paper by Pisani et al [1], which illustrates in an exemplary way how organizations can and should deal with this risk. Ideally, organizations (profit or nonprofit) assemble independent ethics boards, like Crisis Text Line did. Independence is key—the ethics board should consist of independent academics who do not have a stake in the success of the business or organization.

As Danah Boyd further writes, "There have been heated debates in my field about whether or not it is ethical to use corporate trace data without the consent of users to advance scientific knowledge" [3]. Proponents of waiving informed consent requirements for analyzing anonymized big data argue that users implicitly consent to data mining and research by using "free" services—one could argue that people accept this bargain as a trade-off to receiving free services and free information. In the age of generative AI—where large parts of the internet, including discussion boards and other venues that once may have been deemed "private," are scoured by AI bots—users already may have adjusted their expectations regarding privacy and anonymity.

However, when it comes to digital health data, there is an important distinction to be made between noncommercial and commercial use of the data, and perhaps this was the cardinal mistake made by Crisis Text Line—to first communicate that data will never be shared with commercial entities, and then change course, assuming that people who may be fine with giving data (even anonymized) for research purposes to a noncommercial entity (including the volunteers at Crisis Text Line) are also OK with having these anonymized data shared with a commercial entity for machine learning purposes. Research in this journal [17] has shown that people generally distrust data use in a commercial context (technology companies, pharmaceutical companies) more than data sharing for pure research purposes.

XSL•FO
RenderX

The debate over informed consent in the use of "deidentified" internet-generated data for research and commercial use is complex, multifaceted, and constantly evolving.

In a recent 2024 paper [18], the authors argue that especially for digital mental health and other vulnerable populations, the standard to deidentify data to waive informed consent may not be sufficient, and other criteria should be used to address social justice issues. To mitigate the "social risk" of using deidentified data for research without explicit permission from participants, the authors urge researchers to consider the following additional guidelines [18]:

- create socially valuable knowledge,
- fairly share the benefits and burdens of research,
- be transparent about data use,
- create mechanisms for withdrawal of data,
- ensure that stakeholders can provide input into the design and implementation of the research, and
- responsibly report results.

Balancing these ethical considerations with the practical needs of research and innovation is a challenge that requires ongoing dialogue and thoughtful regulation. Ultimately, finding a middle ground that protects individuals' rights while fostering scientific and technological advancement is crucial for navigating this complex issue.

Conclusion

In summary, the commentary of Reierson [14] and the response of Pisani et al [15] highlight the difficult ethical questions nonprofit organizations and companies face when obtaining, analyzing, and sharing data with academics, and even more so for commercial use. The paper by Pisani et al [1] is commendable in that it lays open how data sharing for research has been handled by a nonprofit organization, and the approach of assembling an external data ethics committee to help vet research proposals from other researchers is exemplary. Unfortunately, this has been somewhat tainted by events that occurred after we published the paper, when it emerged that Crisis Text Line made the controversial decision to also share the data with its commercial subsidiary for machine learning purposes. The Data Ethics Committee members who authored the 2019 *Journal of Medical Internet Research* paper were apparently not involved or consulted on that. We see the controversy around Loris.ai as a separate debate that does not invalidate the guidelines and approach for noncommercial data sharing published in the 2019 paper. Many organizations (including nonprofits, hospitals, and research centers) have business development officers and commercialization units that think about how to harness the data they have to advance knowledge and how to commercialize their data in an ethical manner. As such, in the era of generative AI and machine learning, nonprofits creating spin-offs or holding equity in companies that commercialize deidentified data may become increasingly common, and the Crisis Text Line and Loris.ai case may serve as a cautionary lesson on the negative public perception of such arrangements.

Conflicts of Interest

GE is the founder, chief executive officer, and executive editor of JMIR Publications, editor-in-chief of the *Journal of Medical Internet Research*, receives a salary, and owns equity.

References

1. Pisani AR, Kanuri N, Filbin B, Gallo C, Gould M, Lehmann LS, et al. Protecting user privacy and rights in academic data-sharing partnerships: principles from a pilot program at Crisis Text Line. J Med Internet Res 2019 Jan 17;21(1):e11507 [FREE Full text] [doi: 10.2196/11507] [Medline: 30664452]
2. Lublin N. How data from a crisis text line is saving lives. TED. 2015 May. URL: https://www.ted.com/talks/nancy_lublin_how_data_from_a_crisis_text_line_is_saving_lives/transcript?subtitle=en [accessed 2024-12-24]
3. Boyd D. Crisis Text Line, from my perspective. Zephoria. 2022 Jan. URL: https://www.zephoria.org/thoughts/archives/2022/01 [accessed 2024-12-24]
4. Petersen C, Subbian V, Section Editors Special Section on Ethics in Health Informatics of the International Medical Informatics Association Yearbook. Special section on ethics in health informatics. Yearb Med Inform 2020 Aug;29(1):77-80 [FREE Full text] [doi: 10.1055/s-0040-1702014] [Medline: 32823299]
5. Reierson TD. Reader concerns about published work Nov 2021. Reform Crisis Text Line. 2021 Nov 24. URL: https://reformcrisistextline.com/wp-content/uploads/2022/01/ltr_2019_Paper_to_JMIR_11-24-2021_share.pdf [accessed 2024-12-24]
6. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
7. Informed consent FAQs. US Department of Health and Human Services. URL: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/informed-consent/index.html [accessed 2024-12-23]
8. El Emam K, Hintze M. Does anonymization or de-identification require consent under the GDPR? IAPP. 2019 Jan 29. URL: https://iapp.org/news/a/does-anonymization-or-de-identification-require-consent-under-the-gdpr [accessed 2024-12-24]
9. Levine AS. Suicide hotline shares data with for-profit spinoff, raising ethical questions. Politico. 2022 Jan 28. URL: https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617 [accessed 2024-12-24]

XSL•FO
RenderX

10. Bordelon B. FCC Commissioner Carr calls for FTC probe of Crisis Text Line. Benton Institute for Broadband & Society. 2022 Mar 29. URL: https://www.benton.org/headlines/fcc-commissioner-carr-calls-ftc-probe-crisis-text-line [accessed 2024-12-24]
11. An update on data privacy, our community and our service. Crisis Text Line. 2022 Jan 31. URL: https://www.crisistextline.org/blog/2022/01/31/an-update-on-data-privacy-our-community-and-our-service/ [accessed 2024-12-24]
12. Hendel J. Crisis Text Line ends data-sharing relationship with for-profit spinoff. Politico. 2022 Jan 31. URL: https://www.politico.com/news/2022/01/31/crisis-text-line-ends-data-sharing-00004001 [accessed 2024-12-24]
13. Conflicts of interest. COPE. URL: https://publicationethics.org/competinginterests [accessed 2024-12-24]
14. Reierson TD. Commentary on "Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles from a Pilot Program at Crisis Text Line". J Med Internet Res 2024 Dec 30:e42144. [doi: 10.2196/42144] [Medline: 39753221]
15. Pisani AR, Gallo C, Gould MS, Kanuri N, Marcotte JE, Pascal B, et al. Authors' reply: commentary on "Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line". J Med Internet Res 2025 Jan 22:e59734. [doi: 10.2196/59734]
16. JMIR Editorial Office. Correction: protecting user privacy and rights in academic data-sharing partnerships: principles from a pilot program at Crisis Text Line. J Med Internet Res 2024 Dec 20;26:e67880 [FREE Full text] [doi: 10.2196/67880] [Medline: 39705678]
17. Biasiotto R, Viberg Johansson J, Alemu MB, Romano V, Bentzen HB, Kaye J, et al. Public preferences for digital health data sharing: discrete choice experiment study in 12 European countries. J Med Internet Res 2023 Nov 23;25:e47066 [FREE Full text] [doi: 10.2196/47066] [Medline: 37995125]
18. Herington J, Li K, Pisani AR. Expanding the role of justice in secondary research using digital psychological data. Am Psychol 2024 Jan;79(1):123-136. [doi: 10.1037/amp0001190] [Medline: 38236220]

Abbreviations

AI: artificial intelligence
CEO: chief executive officer
IMIA: International Medical Informatics Association
IRB: institutional review board
ToS: terms of service

Edited by T Leung; submitted 23.10.24; this is a non–peer-reviewed article;accepted 24.10.24; published 22.01.25.

Please cite as:
Eysenbach G
Crisis Text Line and Loris.ai Controversy Highlights the Complexity of Informed Consent on the Internet and Data-Sharing Ethics for Machine Learning and Research
J Med Internet Res 2025;27:e67878
URL: https://www.jmir.org/2025/1/e67878
doi:10.2196/67878
PMID:

©Gunther Eysenbach. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Apps and Digital Resources for Child Neurodevelopment, Mental Health, and Well-Being: Review, Evaluation, and Reflection on Current Resources

Kelsie Ann Boulton[1,2,3], PhD; Makana Hilton[1,2,3], MD; Emilia Sutton[1,2,3], MBrainMindSci; Adam John Guastella[1,2,3], PhD

[1]Children's Hospital Westmead Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Child Neurodevelopment and Mental Health Team, Brain and Mind Centre, University of Sydney, Sydney, Australia

[3]Clinic for Autism and Neurodevelopmental (CAN) Research, Brain and Mind Centre, University of Sydney, Sydney, Australia

Corresponding Author:
Kelsie Ann Boulton, PhD
Children's Hospital Westmead Clinical School
Faculty of Medicine and Health
University of Sydney
94 Mallett Street
Sydney, 2050
Australia
Phone: 61 0293510984
Email: kelsie.boulton@sydney.edu.au

Abstract

Background: An increase in the prevalence of neurodevelopmental conditions worldwide, alongside resource constraints within clinical services, has led to increased interest in health information technologies, such as apps and digital resources. Digital tools are often viewed as a solution to bridge this divide and to increase supports for families. There is, however, a paucity of research that has evaluated digital health tools, their potential benefits for child neurodevelopment and associated concerns (eg, mental health, well-being), and their benefit for families.

Objective: This study conducted the first review of existing mobile apps and digital resources targeted at supporting the needs of children with developmental concerns or neurodevelopmental conditions.

Methods: We identified 3435 separate resources, of which 112 (43 apps and 69 digital resources) met the criteria. These resources were categorized according to their purpose or target and were then reviewed based on their engagement, information quality, and evidence base using the Adapted Mobile App Rating Scale.

Results: The most common condition of concern targeted by apps and digital resources was autism (19/112, 17% resources), with retrieved resources focusing on supporting challenging behaviors, promoting speech, language, and social development, and providing options for alternative and assistive communication. Other common areas of concern targeted by apps and digital resources included language and communication (16/112, 14.3%) and attention-deficit/hyperactivity disorder (11/112, 9.8%). Results showed that reviewed resources were engaging, with high levels of accessibility and functionality. Resources had various functions, including developmental or behavioral tasks targeted at children, assistive communication support, scheduling support, journaling, and advice, activities, and strategies for parents. The information quality of resources, such as credibility of source and evidence base was, however, mostly low. Apps and digital resources with good credibility and an existing evidence base were largely developed in partnership with research, health, or government institutions, and were rated significantly higher on overall quality compared with apps and digital resources not developed in partnership with such institutions (apps; t_{41}=–4.35, P<.001; digital resources; t_{67}=–4.95, P<.001).

Conclusions: The lack of evidence base across resources means that it is extremely difficult to provide recommendations to families with respect to apps or digital resources that may support their needs. Frameworks for the development of new tools are discussed, highlighting the novel approaches required to demonstrate the efficacy of tools for improving outcomes for children and families. Such a framework requires collaboration with multiple stakeholders (software developers, researchers, regulatory bodies, clinicians, children, and families) and engagement across multiple levels of expertise (app development, implementation,

XSL•FO
RenderX

and dissemination within services, policy, and clinical regulations), to harness the potential of digital health for improving outcomes and promoting support in child neurodevelopment, which at this juncture remains largely underdeveloped.

(J Med Internet Res 2025;27:e58693) doi:10.2196/58693

KEYWORDS

digital tools; neurodevelopmental conditions; mental health; digital health; implementation; digital interventions; child neurodevelopment; digital technology; mobile phone

Introduction

In recent years, there has been a reported increase in the prevalence of neurodevelopmental conditions (NDCs) [1,2] and a rising awareness of the importance of early diagnosis and intervention in children with signs of developmental delay [3,4]. This has been matched with a growing number of recommendations and guidelines for multidisciplinary supports provided as early as possible [5-7]. Early identification and evidence-based supports are argued to be crucial for enhancing outcomes in the 1 in 10 children diagnosed with a neurodevelopmental condition [4].

Despite this, there is consistent evidence across nations of the considerable delays and barriers children and families face when seeking assessment [8,9] and support [10,11]. In addition, the wide range of multidisciplinary support needs often experienced by children and caregivers are frequently reported to result in fragmented care, where co-occurring conditions (eg, mental health) are often left unaddressed [12-14]. Caregivers can often report feeling overwhelmed and confused, and comment on the lack of coordination of care and difficulty in obtaining trustworthy, accessible information about supports [15-17].

It is unsurprising, given these resource constraints, that many families turn to health information technologies, such as apps and digital tools, to obtain information to supplement their health care [18,19]. The potential utility of such technologies has been reinforced internationally, with global commissioned reports recommending the integration of digital technologies to facilitate and optimize health care for children and adolescents [20,21]. Digital tools may bridge the divide between existing limited resources and growing demand [22], providing more instant access to supports that may be delivered in a flexible manner. Such tools may also bridge geographical barriers, facilitate access to care for diverse communities, and provide additional opportunities for assessment and feedback to support child and family needs across time [23-25].

Despite enthusiasm for the use of innovative digital solutions [26,27], there is an urgent need to understand the digital tools that currently exist for children with NDCs and their families, the platforms that these tools can be accessed on, and whether these tools are reliable, credible and have an existing evidence base to support their implementation. There is also a need for such evaluations to map existing resources so that supports can be strategically developed where gaps exist. For example, it has previously been shown that, despite the promise of digital tools for mental health, only a minority of apps purported to monitor and manage mental health symptoms have clinically validated evidence of effectiveness [28,29]. More recently, a meta-analysis of the efficacy of smartphone apps for symptoms of depression and anxiety reported overall small effects for symptom amelioration [30]. In addition, while there can be many valuable resources available, there are also many poor sources of information [31-33]. As a result, individuals may instead be influenced by factors such as the star rating of an app, which have been shown to have little relationship with clinical utility [34].

In addition to evaluating digital tools in terms of their evidence base and credibility, it is critical to also evaluate these tools in terms of their engagement, aesthetics, and user quality. Worldwide, there is consistent evidence that poor quality of an app or digital resource is associated with abandonment [35]. In contrast, the design and visual features of apps including an engaging interface, and being intuitive and easy to navigate, all increase engagement and use [36]. To illustrate, digital mental health tools have been shown to have continued uptake as low as 0.1% when rolled out at scale and this low rate is believed to be associated with low appeal and poor app maintenance to promote engagement [37]. This means that the ongoing use of a digital tool, no matter the evidence behind it, is likely to fail without high appearance and usability ratings [23,37]. This may be particularly important for families with NDCs as research shows that personal characteristics moderate the type of apps and digital tools that are used [38]. People with NDCs are more likely to learn from well-structured, easy-to-understand, and clear information, further highlighting the need to evaluate the quality and aesthetics of apps and digital resources [39].

The objective of this study was to review available mobile apps and digital resources that are targeted at supporting the needs of children with developmental concerns or NDCs. The study first sought to conduct a systematic review of all existing mobile apps and digital resources and to categorize them according to their purpose or target. Second, the review then sought to evaluate the information quality, utility, and evidence base for each resource by using the Adapted Mobile App Rating Scale (A-MARS) [40] to demonstrate the quality and evidence base of existing tools.

Methods

Design

This review used a stepwise approach in a similar manner to a systematic review. Our approach included a search strategy, assessment against prespecified eligibility criteria, app, and digital resource selection through an initial screening of all identified apps and digital resources, a full review of included apps and digital resources, data extraction and analysis, and

XSL•FO
RenderX

quality assessment of included apps and digital resources using a reliable quality assessment tool, the A-MARS [40,41].

Search Strategy

A search of the main online app stores, iTunes (Apple Inc, Australia) and Google Play (Google Inc, Australia), and a search of digital resources using the Google search engine was conducted from October 2023 to November 2023. To avoid a selection bias from personal Google algorithms, search histories were cleared before each search term as per previous recommendations [42]. Search terms were identified following consultations with researchers with expertise in NDCs, clinicians practicing in neurodevelopmental assessment services, and families with lived experience of NDCs. Search terms were selected based on keywords caregivers of children with developmental concerns or NDCs may use when attempting to access apps and digital resources. The final search terms also demonstrated the best performance in identifying apps and digital resources of interest for this review during preliminary searches. In total, 28 search terms were identified by an expert clinical team based on *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) terms for neurodevelopment and transdiagnostic terms that families might typically use. Search terms included neurodevelopment kids, intellectual disability kids, global developmental delay, communication disorder, language disorder kids, stuttering kids, autism spectrum disorder, autism kids, attention-deficit/hyperactivity disorder (ADHD) kids, ADHD kids, learning disorder kids, motor disorder kids, Tourette's disorder kids, cerebral palsy kids, dyslexia kids, epilepsy kids, anxious kids, depressed kids, sleep problem kids, attention kids, social problems kids, restricted behaviors kids, repetitive behaviors kids, social skills kids, well-being kids, education kids, emotion kids, cognition kids.

Eligibility Criteria

To be considered for inclusion in this review, apps, and digital resources were required to meet the following inclusion criteria: (1) contain keywords of the search term in the title or description, (2) have a focus on the needs of children with developmental concerns or NDCs (eg, social skills, well-being, academic skills, or mental health), (3) be free of charge, and (4) be in the English language. To ensure the resources contained relevant and applicable information that was catered to families of children with developmental concerns or NDCs, a variety of exclusion criteria was applied. Apps and resources were excluded if they (1) had no actionable information for families to use (eg, purely game, social media site, news article, research paper, clinic website, or home assessment), (2) if they were selling products (eg, books, materials, or therapy sessions), (3) if the information was not relevant to our target population (eg, app or resource targeted only at adults or included information was overly general), or (4) if they were in unsuitable formats for review with the A-MARS (ie, PDF document, podcast, or video).

Selection Process

Three authors (MH, ES, and KB) carried out the app store and search engine searches. To balance feasibility and search comprehensiveness, the first 50 resources for each search term were selected from both the app stores and the search engine [43]. A predesigned Microsoft Excel spreadsheet developed for this review was used to enter information about the apps and digital resources. Information entered included the name of the app or digital resource, the name of the app developer or the URL for digital resources, the app store or stores in which the app was available, and the search terms or terms that identified each app or digital resource.

Following removal of duplicates between search terms, 1 rater (MH) conducted an initial screen for eligibility, by reviewing the title, description, and home page of the app or digital resource. Additional screening was conducted on the remaining apps and digital resources to confirm inclusion in the review. This screening was conducted by 2 raters (MH and ES) and involved the installation of the app and a more detailed review of the digital resource, including a review of written information, pictures, and videos. Any disagreements were discussed with a 3rd rater (KB) and resolved by consensus. Reasons for exclusion were recorded.

Data Extraction

Apps and digital resources identified as eligible in the screening phase were then assessed using the A-MARS [40]. The A-MARS consists of 28 items, developed to assess the quality of mobile apps and e-tools (eg, websites and digital resources). The A-MARS provides an evaluation of app and e-tool quality by grading each app or digital resource on several domains, as described below (Textbox 1). Each item was scored using a 5-point Likert scale (1: inadequate, 2: poor, 3: acceptable, 4: good, and 5: excellent). Supplementary questions within the A-MARS capture information pertaining to the health-related quality of the app or digital resource. Eight scores are calculated for the A-MARS, including the mean score for each domain (ie, engagement, functionality, aesthetics, information, subjective quality, and health-related quality), a mean quality score based on the engagement, functionality, aesthetics and information domains, and a mean total score.

XSL•FO
RenderX

Textbox 1. List of Adapted Mobile App Rating Scale subscales and subscale descriptions.

Subscale and description

Engagement (5 items)

- Is the app or digital resource engaging and interesting for the user? Does it have prompts (eg, send alerts, messages, or reminders)?

Functionality (4 items)

- How does the app or digital resource function, does it have a logical flow and design? Is it easy to navigate?

Aesthetics (3 items)

- How appealing is the app or digital resource in terms of its overall visual appeal, graphic design, and stylistic consistency?

Information (6 items)

- Does the app or digital resource contain high-quality information (eg, text, feedback, measures, or references) from a credible source?

Subjective quality (4 items)

- Subjective quality rating of the app or digital resource.

Health-related quality (6 items)

- Does the app or digital resource provide access to other resources, strategies linked to the target issue, or the option for real-time tracking?

All apps and digital resources were rated using the A-MARS instrument. Three expert raters conducted this review: (1) a senior research fellow with a PhD in Psychology and 10 years' experience working with pediatric NDCs and digital health; (2) a senior research assistant with a Master's degree in Brain and Mind Sciences and 3 years' experience in working with NDCs and digital tools to support children with NDCs and their families; and (3) a research affiliate with Bachelors' degrees in Psychology and Medical and Health Sciences and a Postgraduate Diploma in Psychology. Two of these raters were also people with a diagnosed NDC. All raters reviewed the A-MARS in depth before conducting initial pilot ratings. Two apps and 2 digital resources were initially reviewed independently for training purposes. After independently rating the apps and digital resources, the raters met to compare and review results and to resolve discrepancies in ratings. To reach a consensus, the raters reviewed the A-MARS in depth to improve the alignment of ratings. The remaining apps and digital resources were then rated independently by 2 raters. Based on the methodology described by others [40,41], each rater trialed the apps and digital resources for a minimum of 10 minutes and then independently rated their quality using the A-MARS.

Statistical Analyses

Statistical analyses were conducted using SPSS (version 28; IBM Corp). Descriptive statistics were used to summarize the A-MARS domain scores, the mean quality score, and the mean total score across apps and digital resources. Cronbach α and intraclass correlation coefficients were used to calculate the internal consistency and interrater reliability of the A-MARS domain scores, the quality score, and the total score. As this is the first study to our knowledge that has applied the A-MARS in apps and digital resources targeted at child development and NDCs, 2-tailed independent samples *t* tests were also used to compare mean domain scores, quality scores, and total scores between apps and digital resources. Bonferroni corrections were used for these comparisons and a *P* value of <.006 was considered statistically significant. To understand how the involvement of a reputable and credible organization impacted ratings for apps and digital resources, 2-tailed independent samples *t* tests were used to compare domain scores, quality scores, and total scores for apps and digital resources that had been developed in partnership with a university, health or government institution, relative to those that had not been developed in partnership with such institutions.

Results

Overview

The process of identification and inclusion of apps and digital resources is outlined in Figure 1. A total of 3435 apps and digital resources were initially identified. After the removal of duplicates, 2211 apps and digital resources were excluded following initial screening by title and description, and an additional 229 apps and digital resources were excluded following installation and more detailed screening. The most common reason for exclusion was an absence of actionable information that children and families could use (575/2440, 23.6%), followed by not being free of charge (334/2440, 13.6%), and not being relevant to the target population (320/2440, 13.1%). A total of 12 Apple apps, 31 Android apps, and 69 digital resources were included in the final review of this paper.

XSL•FO
RenderX

Figure 1. Flowchart of selection of included resources.

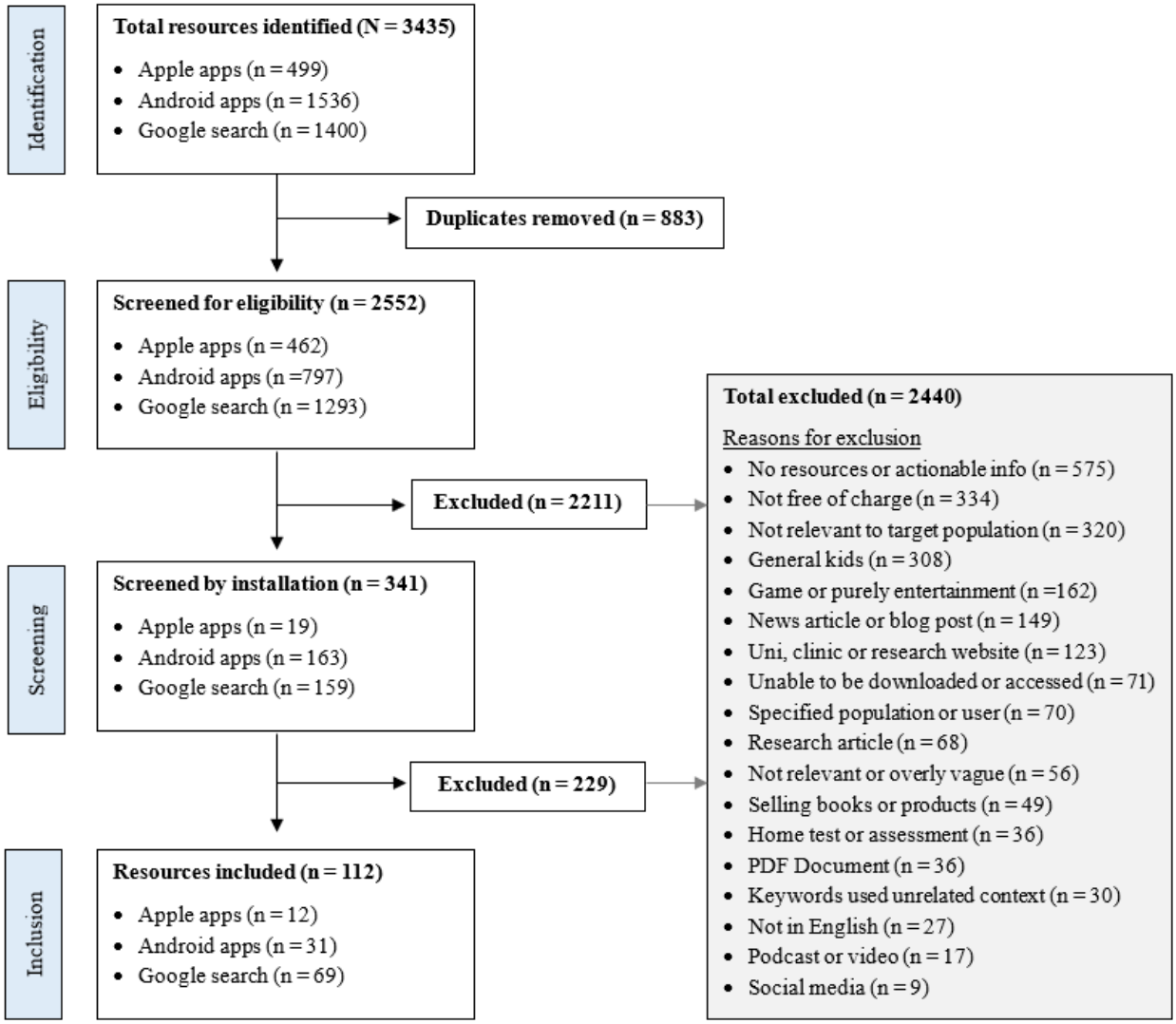

Features of Included Apps and Digital Resources

Table 1 displays the primary area of focus for retrieved apps and digital resources. The most common conditions that the resources targeted were autism, language, or communication difficulties, and ADHD. Both apps and digital resources were retrieved for most searched terms; however, resources retrieved for anxiety, attention, and developmental delay were mainly digital resources. Comparatively, the resources retrieved for epilepsy, cerebral palsy, and child well-being were mainly apps. Retrieved apps and digital resources had a range of functions, including developmental or behavioral tasks targeted at children, assistive communication support, scheduling support, journaling, and advice, activities, and strategies for parents.

XSL•FO
RenderX

Table 1. Primary focus areas included apps and digital resources.

Topic area	Apps, n (%)	Digital resources, n (%)	Apps and digital resources combined, n (%)
Autism	5 (11.6)	14 (20.3)	19 (17.0)
Language and communication	7 (16.3)	9 (13.0)	16 (14.3)
ADHD[a]	6 (14.0)	5 (7.2)	11 (9.8)
Anxiety	1 (2.3)	9 (13.0)	10 (8.9)
Dyslexia	3 (7.0)	6 (8.7)	9 (8.0)
Mood and depression	2 (4.7)	5 (7.2)	7 (6.3)
Social skills	2 (4.7)	4 (5.8)	6 (5.4)
Emotions	2 (4.7)	4 (5.8)	6 (5.4)
Intellectual disability	3 (7.0)	2 (2.9)	5 (4.5)
Epilepsy	5 (11.6)	0 (0)	5 (4.5)
Attention	1 (2.3)	4 (5.8)	5 (4.5)
Sleep	2 (4.7)	2 (2.9)	4 (3.6)
Cerebral palsy	2 (4.7)	1 (1.4)	3 (2.7)
Developmental delay	0 (0)	3 (4.3)	3 (2.7)
Child well-being	2 (4.7)	0 (0)	2 (1.8)
Tourette	0 (0)	1 (1.4)	1 (0.9)

[a]ADHD: attention deficit hyperactivity disorder.

Overall Assessment of Apps and Digital Resources

Across the 43 apps and 69 digital resources, excellent internal consistency and inter-rater reliability was found across the total and sub-domain A-MARS scales. Detailed item and domain statistics for the reviewed apps are displayed in Tables S1 and S2 in Multimedia Appendix 1. Table 2 displays the mean and median scores of the reviewed apps (n=43) and digital resources (n=69) across the 5 A-MARS domains, as well as for the quality and total scores. The apps, taken together, scored higher than the minimum acceptable score of 3.0 across most A-MARS domains, excluding the health-related quality domain (Table 2). Apps typically received low scores on this domain due to a lack of additional resources available (mean 1.81, SD=1.22), and limited access to further support or related information (mean 1.85, SD=1.24). Furthermore, the digital resources scored higher than 3.0 across all A-MARS domains except the subjective quality domain (Table 2).

Table 2. Adapted Mobile App Rating Scale ratings for apps and digital resources.

Apps				Digital resources			Values	
Domain	Mean (SD)	Median (range)	N (%) scoring ≥3.0	Mean (SD)	Median (range)	N (%) scoring ≥3.0	T-statistic (*df*)	*P* value
Engagement	3.67 (0.74)	3.60 (2.40-4.90)	36 (83.7%)	3.19 (0.54)	3.13 (2.00-4.50)	45 (65.2%)	3.97 (110)	<.001
Functionality	4.13 (0.68)	4.25 (2.25-5.00)	40 (93.0%)	4.09 (0.55)	4.13 (2.50-5.00)	66 (95.7%)	0.37 (110)	.72
Aesthetics	3.73 (0.72)	4.00 (1.84-5.00)	38 (88.4%)	3.66 (0.68)	3.67 (2.00-5.00)	62 (89.9%)	0.56 (110)	.57
Information	3.06 (0.95)	3.13 (1.00-4.75)	27 (62.8%)	3.57 (0.59)	3.67 (2.33-4.75)	61 (88.4%)	-3.16 (110)	.002
Subjective quality	3.21 (0.88)	3.50 (1.50-4.88)	26 (60.5%)	2.89 (0.68)	2.88 (1.38-4.25)	27 (39.1%)	2.04 (110)	.04
Health-related quality	2.58 (0.91)	2.33 (1.00-4.83)	13 (30.2%)	3.34 (0.78)	3.33 (2.00-5.00)	45 (65.2%)	-4.74 (110)	<.001
Quality score	3.40 (0.67)	3.40 (2.17-4.88)	32 (74.4%)	3.62 (0.53)	3.66 (2.38-4.69)	60 (87.0%)	–2.01 (110)	.04
Total score	3.40 (0.64)	3.29 (2.00-4.72)	31 (72.1%)	3.48 (0.55)	3.50 (2.32-4.67)	54 (78.3%)	–0.73 (110)	.47

Both apps and digital resources scored higher than 3.0 across quality and total scores. Apps and digital resources scored high on the functionality domain, with consistently high average scores across subcriteria. Of note, included apps scored relatively low in the information quality domain primarily due to questionable source credibility (mean 2.52, SD 0.79). Furthermore, a minority of apps (3/43, 7%) and no digital resources met the criteria for a verifiable evidence base as outlined by the A-MARS (has been trialed or tested and published in scientific literature). Independent sample *t* tests revealed statistically significant differences in mean scores between apps and digital resources on the engagement, information, and health-related quality domains ($P<.006$). While apps had higher ratings relative to digital resources for

XSL•FO
RenderX

engagement, digital resources had higher ratings relative to apps on information and health-related quality domains.

Overall, apps and digital resources that had been developed in partnership with a health, university, or government institution, or where such an institution was listed as a partner, were rated more highly across domains of engagement, functionality, information, subjective quality, and health-related quality, as well as on overall quality and total scores. As shown in Tables 3 and 4, these effects were moderate to large for most comparisons. In total, 19% (8/43) of apps were developed in partnership with a health, university, or government institution, and 48% (33/69) of digital resources were developed in partnership with these institutions.

Table 3. Adapted Mobile App Rating Scale ratings for apps split by partnerships with health, university, or government institutions.

Domain	Apps				
	Institution partnership (n=8), mean (SD)	No institution partnership (n=35), mean (SD)	Effect size	T-statistic (*df*)	*P* value
Engagement	4.23 (0.75)	3.54 (0.68)	0.69	–2.53 (41)	.02
Functionality	4.64 (0.30)	4.01 (0.69)	0.64	–2.53 (41)	.02
Aesthetics	4.15 (0.62)	3.64 (0.73)	0.70	–1.85 (41)	.07
Information	4.30 (0.42)	2.78 (0.80)	0.75	–7.52 (41)	<.001
Subjective quality	4.05 (0.58)	3.02 (0.83)	0.80	–3.29 (41)	.002
Health-related quality	3.53 (1.33)	2.36 (0.63)	0.79	–2.42 (41)	.04
Quality score	4.13 (0.58)	3.23 (0.57)	0.57	–4.04 (41)	<.001
Total score	4.14 (0.51)	3.23 (0.54)	0.54	–4.35 (41)	<.001

Table 4. Adapted Mobile App Rating Scale ratings for digital resources split by partnerships with health, university, or government institutions.

Domain	Digital resources				
	Institution partnership (n=33), mean (SD)	No institution partnership (n=36), mean (SD)	Effect size	T-statistic (*df*)	*P* value
Engagement	3.47 (0.51)	2.93 (0.44)	0.47	–4.76 (67)	<.001
Functionality	4.36 (0.46)	3.83 (0.51)	0.48	–4.58 (67)	<.001
Aesthetics	3.98 (0.62)	3.36 (0.59)	0.6	–4.24 (67)	<.001
Information	3.87 (0.53)	3.30 (0.50)	0.51	–4.60 (67)	<.001
Subjective quality	3.13 (0.70)	2.67 (0.58)	0.64	–2.95 (67)	.004
Health-related quality	3.69 (0.76)	3.02 (0.65)	0.7	–3.95 (67)	<.001
Quality score	3.92 (0.44)	3.35 (0.44)	0.45	–5.18 (67)	<.001
Total score	3.77 (0.50)	3.21 (0.45)	0.47	–4.95 (67)	<.001

Individual Assessment of Apps and Digital Resources

Figures 2 and 3 display scores of the top 5 and bottom 5 apps in terms of their A-MARS ratings. These radar charts show how individual apps and digital resources scored on each A-MARS domain, as well as on the overall quality and total scores. Showing the difference between the top 5 and bottom 5 highlights the disparity between apps and digital resources on A-MARS ratings. In general, the top 5 apps and digital resources were characterized by high levels of factually correct information, relevant resources, good visuals, and engaging and user-friendly interfaces. Highly rated apps and digital resources were developed in collaboration with research, health, or government institutions. In contrast, the bottom 5 apps and digital resources were characterized by less involvement with a reputable institution, defective or inactive links, difficult or cumbersome interfaces, and a lack of information relevant to the purported concern or condition.

XSL•FO
RenderX

Figure 2. Radar chart of the top 5 and bottom 5 ranked apps. ADHD: attention-deficit/hyperactivity disorder.

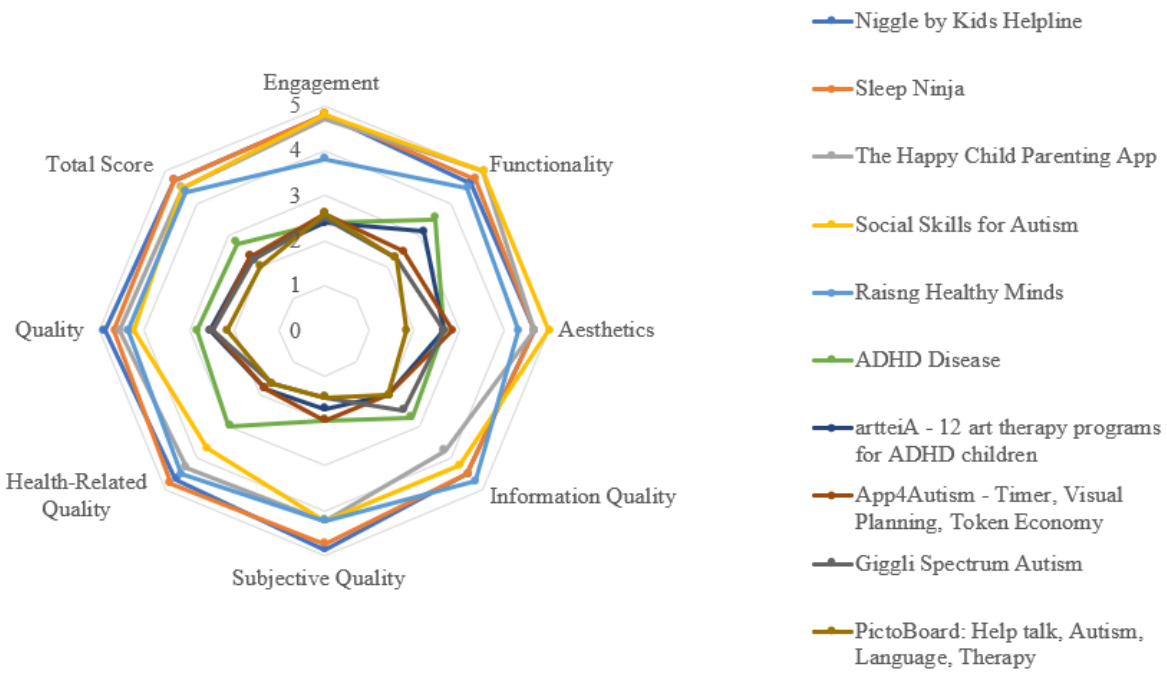

Figure 3. Radar chart of the top 5 and bottom 5 ranked digital resources.

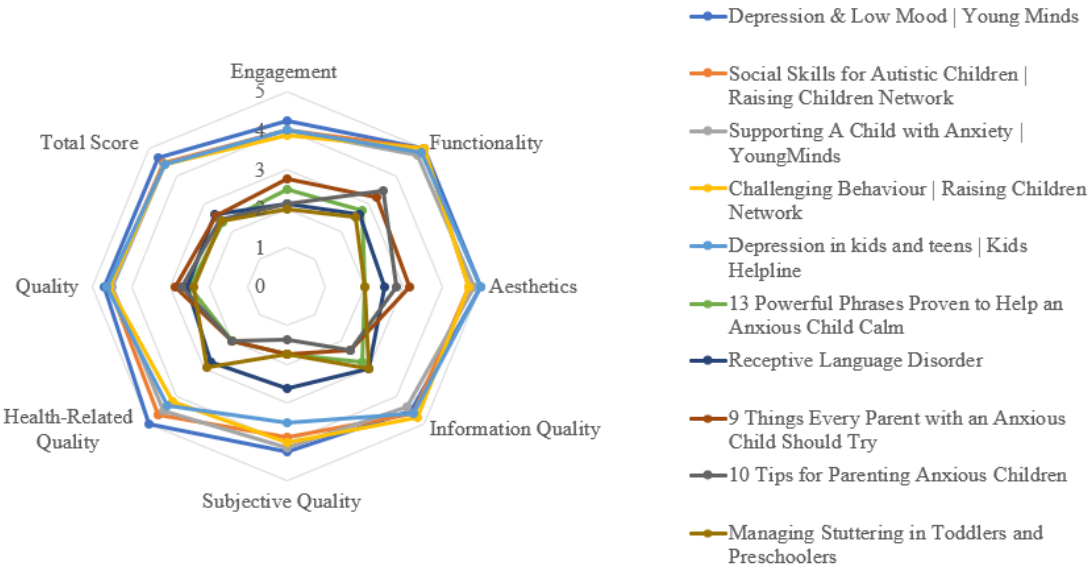

Discussion

This study reviewed existing apps and digital resources that are targeted at the needs of children with developmental concerns or NDCs. To our knowledge, this is the first study that has evaluated apps and digital resources for child neurodevelopment and mental health broadly. From the initial 3435 resources identified, the study reviewed 112 resources in total (43 apps and 69 digital resources). Developed for children with NDCs and their families, the most common resources focus on supports for autism, language or communication difficulties, and ADHD. Most apps and digital resources exceeded the minimum acceptable score of 3.0, suggesting that retrieved resources were engaging, with high levels of accessibility and functionality. However, the results also showed items relating to information quality, such as credibility of source and evidence base, were generally low. Only 7% of apps or digital resources had an established evidence base. Interestingly, mobile apps and digital resources that were developed in partnership with a health, university, or government institution were rated significantly

XSL•FO
RenderX

higher across most domains in comparison to apps and resources that were not developed in partnership with such an organization. Unfortunately, the review showed there was almost no evidence for their efficacy or effectiveness in promoting supportive behaviors or improving outcomes. As a result, there is a significant existing divide between the availability of digital tools that can be integrated into family systems to support needs and evidence that warrants recommendations from clinical teams. While the findings of the review provide an evaluation and collation of existing tools, this conclusion highlights the urgent need to develop new frameworks to test efficacy for improving outcomes and promoting supportive behaviors.

These results show that most apps and digital resources alike met the minimum acceptable criteria in terms of quality (eg, highly engaging and functional). However, we did observe some differences as a function of resource type (ie, apps or digital resources). We found that apps displayed higher ratings on the domain of engagement, suggesting that the reviewed apps were more interesting, customizable, and entertaining, compared to digital resources. In contrast, reviewed digital resources had higher ratings on domains of information and health-related quality compared with apps. This indicates that the quality and credibility of information, as well as any additional resources and strategies provided, was higher in the reviewed digital resources.

Apps and digital resources were rated highly in terms of their engagement and functionality, but somewhat lower on information quality, particularly with respect to credibility of source and existing evidence base. We observed that apps and digital resources that had been developed by or in collaboration with a research, health, or government institutions were rated significantly more highly on information quality and credibility. This aligns with studies that have used evaluation tools like the Mobile App Rating Scale and A-MARS in other fields [40,44]. However, we note that very few apps or digital resources had a scientific evidence base to support their use, which points to the dearth of evidence evaluating the effectiveness of apps and digital resources [40,45]. This lack of evidence base may relate to the iterative, fast-paced nature of technological development and the linear, incremental approach to clinical science and research-based interventions. With established lags of up to 17 years for research to be translated into clinical practice, software developers may opt to move to large-scale dissemination of their product before developing a rigorous evidence base [24]. However, the consequence of this is an abundance of apps and digital resources that have little to no evidence base and may not be beneficial for consumers [27]. To resolve this issue, a novel framework with investment prioritized for development is critical, whereby the effectiveness of apps and digital resources can be rigorously evaluated in a streamlined and timely manner. Such a framework needs to incorporate the swift evaluation of real-world implementation alongside both the evaluation of resource development and the evaluation of efficacy. It is unlikely to be achieved by applying standard academic approaches or randomized controlled trial methods and may need to incorporate adaptive research designs within implementation science methodologies. Integrated research enabling platforms [46-48] that permit rapid implementation of clinical trials for multiple apps and resources within existing ethics and governance frameworks are also required for swift and sustained translation.

Our results show that the most common condition of concern targeted by apps and digital resources was autism (19/112, 17.0% of all apps and digital resources), with retrieved resources focusing on supporting challenging behaviors, promoting speech, language, and social development, and providing options for alternative and assistive communication. Given the growing prevalence of conditions like autism [1], an increase in the number of health information technologies focused on these conditions is relevant and has the potential to provide children and families with a variety of support tools. However, the proportion of resources targeted at comorbid conditions, such as mental health, was lower (10/112, 8.9% targeted at anxiety and 7/112, 6.3% targeted at mood or depression). Mental ill health is one of the fastest-growing health problems in children and adolescents, particularly in the wake of the COVID-19 pandemic [49]. Moreover, children with existing developmental delays and NDCs are more likely to experience mental health conditions relative to their neurotypical peers [12]. Together these findings point to a gap in existing digital resources for children with co-occurring developmental and mental health concerns.

It is unlikely that a single resource will meet the varied needs of children and families. The application and evaluation of digital resources to uplift support in the community may require integrated approaches targeting different needs at different stages of child development and different stages of support seeking. Recent arguments have also been made for the integration of digital navigators that can further facilitate the engagement and implementation of digital health tools within multidisciplinary care teams and standard clinical care [40,50]. With the increasing frequency of digital navigators in other disciplines, such as mental health [51,52], health systems providing child development and assessment services can benefit from adopting a similar approach, to support children and families in accessing reliable, credible, and efficacious digital tools across their health care journey.

While this study provided a first review of apps and digital resources targeted at child neurodevelopment, mental health, and well-being, we note certain limitations. The evaluation of apps and digital resources can be made difficult due to the growing number of app evaluation frameworks [50]. While the evaluation tool we selected, the A-MARS, has sound reliability [53], has been used across disciplines [54,55], and has been adapted for use specifically with digital resources [40], we note that there are challenges common across evaluation tools [50]. While it was deemed the most suitable evaluation tool for this review, we note specific limitations of the A-MARS, such as potential subjectivity in ratings, a relative lack of focus on the clinical effectiveness of health information, and a lack of validation across cultures and languages, which may warrant further evaluation of apps and digital resources with additional evaluation tools. However, there is currently no consensus for the evaluation and regulation of apps and digital resources, making it difficult to conduct standardized evaluations that can provide clear guidance and recommendations to patients and

XSL•FO
RenderX

clinicians. We also note that we restricted our search to apps and digital resources that were available free of charge. While this may have excluded potentially valuable paid resources, we argue that reviewing freely available resources increases the value of this study for use in the broader community and serves to increase equity in access to supports and resources. In a similar vein, we restricted our search to English language apps and digital resources, potentially limiting the applicability of our findings for non–English speaking populations. However, we have detailed a process that future researchers can use to identify and evaluate apps and digital resources in their region of interest, thereby facilitating evaluations in different countries and contexts. Further, we included a large number of search terms, which resulted in many retrieved resources. As a team we discussed restricting the search strategy; however, it was noted that children with developmental concerns or NDCs often experience multiple transdiagnostic concerns, with caregivers likely to search for support across many of these domains. As such, our more extensive search strategy was considered appropriate in order to capture the broad needs and concerns experienced by this patient population.

A crucial next step for this field is to understand how accessible, reliable, evidence-based health information technologies that meet the needs of children, families, and health care providers can be developed and disseminated across clinical settings. This requires collaboration with multiple stakeholders (app developers, researchers, regulatory bodies, clinicians, children, and families) as well as a reconceptualization of how to implement and recommend health information technologies across clinical services for maximum uptake and maintained engagement. At the development level, for example, partnerships between industry, research, and end-users are key to drive the timely co-production of evidence-based, accessible digital tools. Furthermore, given findings from digital mental health [56], a focus on developing tools that include gamification features may be key to optimizing uptake and engagement by children and families. At the implementation and dissemination level, existing models of care may need to be rethought to optimally engage recipients (ie, patients and clinicians) and encourage the integration of digital tools and solutions into clinical practice. Meanwhile, at the policy level, clinical regulation and recommendations require flexibility to enable innovative digital solutions to be adopted and embedded within health services and systems. While complex and multifaceted, such a framework would enable the development of evidence-based digital tools that are primed for accessibility and engagement with children and families and would facilitate the embedding of digital health into clinical service settings in a sustainable manner, uplifting the capacity of services to provide access to reliable, evidence-based digital tools that are likely to be used to support children and families.

In the current technology-driven world, apps and digital resources are being increasingly used and promoted as a source of information or an adjunct to support and therapy. In this review, we found that a large number of apps and digital resources targeted at child neurodevelopment, mental health, and well-being are engaging and functional, but lack an adequate evidence base. Those apps and digital resources that did show good credibility and an evidence base were largely developed in partnership with research, health, or government institutions. As a result, there is a pressing need to recognize, value, and promote this type of collaboration when developing new digital tools, and to also develop a framework where new digital tools can be rigorously evaluated in a timely manner, to promote swift translation from evidence-based research and development into practice. Clinicians and health care professionals can then be supported to recommend reliable, evidence-based apps and digital resources to children and families based on their individual needs, providing an ideal opportunity to evaluate the effectiveness of these tools for enhancing outcomes.

Acknowledgments

This work was supported by a Research Development Grant from the Brain and Mind Centre and an Emerging Researcher Grant from The Research Foundation of Cerebral Palsy Alliance (ERG01921).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KAB and AJG were involved in the study conceptualization and supervision. KAB, MH, and ES were involved in data extraction, data curation, and formal analysis. KAB and AJG wrote the original draft of the manuscript. All authors provided review and feedback on the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interrater reliability and internal consistency of the A-MARS (Adapted Mobile App Rating Scale) items and domain scores.
[DOCX File , 21 KB - jmir_v27i1e58693_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
PRISMA Checklist.
[PDF File (Adobe PDF File), 136 KB - jmir_v27i1e58693_app2.pdf]

References

1. Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2018. MMWR Surveill Summ 2021;70(11):1-16 [FREE Full text] [doi: 10.15585/mmwr.ss7011a1] [Medline: 34855725]
2. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Danielson ML, Bitsko RH, et al. Prevalence and trends of developmental disabilities among children in the United States: 2009-2017. Pediatrics 2019;144(4):e20190811 [FREE Full text] [doi: 10.1542/peds.2019-0811] [Medline: 31558576]
3. Clark MLE, Vinen Z, Barbaro J, Dissanayake C. School age outcomes of children diagnosed early and later with autism spectrum disorder. J Autism Dev Disord 2018;48(1):92-102. [doi: 10.1007/s10803-017-3279-x] [Medline: 28905160]
4. Hadders-Algra M. Early diagnostics and early intervention in neurodevelopmental disorders-age-dependent challenges and opportunities. J Clin Med 2021;10(4):861 [FREE Full text] [doi: 10.3390/jcm10040861] [Medline: 33669727]
5. AADPA. ADHD guideline development group: Australian evidence-based clinical practice guideline for attention deficit hyperactivity. 2022. URL: https://adhdguideline.aadpa.com.au/ [accessed 2024-10-25]
6. NICE. National Institute for Health and Care Excellence: Guidelines. In: Attention Deficit Hyperactivity Disorder: Diagnosis and Management. England, UK: National Institute for Health and Care Excellence (NICE) Copyright © NICE 2018; 2019.
7. Whitehouse AJO, Evans K, Eapen V, Wray J. A national guideline for the assessment and diagnosis of autism spectrum disorders in Australia. In: Summary and Recommendations. Australia: Cooperative Research Centre for Living with Autism; 2018.
8. Straub L, Bateman BT, Hernandez-Diaz S, York C, Lester B, Wisner KL, et al. Neurodevelopmental disorders among publicly or privately insured children in the United States. JAMA Psychiatry 2022;79(3):232-242 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3815] [Medline: 34985527]
9. Boulton KA, Hodge MA, Jewell A, Ong N, Silove N, Guastella AJ. Diagnostic delay in children with neurodevelopmental conditions attending a publicly funded developmental assessment service: findings from the Sydney Child Neurodevelopment Research Registry. BMJ Open 2023;13(2):e069500 [FREE Full text] [doi: 10.1136/bmjopen-2022-069500] [Medline: 36725093]
10. Sapiets SJ, Hastings RP, Totsika V. Predictors of access to early support in families of children with suspected or diagnosed developmental disabilities in the United Kingdom. J Autism Dev Disord 2024;54(4):1628-1641 [FREE Full text] [doi: 10.1007/s10803-023-05996-7] [Medline: 37142908]
11. Boulton KA, Hodge A, Levu K, Ong N, Silove N, Guastella AJ. Access and barriers to supports for children and caregivers attending public child developmental assessment services: findings from the Sydney Child Neurodevelopment Research Registry. Autism Res 2024;17(3):555-567. [doi: 10.1002/aur.3064] [Medline: 38009266]
12. Boulton KA, Guastella AJ, Hodge MA, Demetriou EA, Ong N, Silove N. Mental health concerns in children with neurodevelopmental conditions attending a developmental assessment service. J Affect Disord 2023;335:264-272 [FREE Full text] [doi: 10.1016/j.jad.2023.04.098] [Medline: 37119867]
13. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. Lancet Psychiatry 2017;4(4):339-346. [doi: 10.1016/S2215-0366(16)30376-5] [Medline: 27979720]
14. Kerns CM, Kendall PC, Zickgraf H, Franklin ME, Miller J, Herrington J. Not to be overshadowed or overlooked: functional impairments associated with comorbid anxiety disorders in youth with ASD. Behav Ther 2015;46(1):29-39. [doi: 10.1016/j.beth.2014.03.005] [Medline: 25526833]
15. Carlsson E, Miniscalco C, Kadesjö B, Laakso K. Negotiating knowledge: parents' experience of the neuropsychiatric diagnostic process for children with autism. Int J Lang Commun Disord 2016;51(3):328-338. [doi: 10.1111/1460-6984.12210] [Medline: 26833425]
16. Crane L, Batty R, Adeyinka H, Goddard L, Henry LA, Hill EL. Autism diagnosis in the United Kingdom: perspectives of autistic adults, parents and professionals. J Autism Dev Disord 2018;48(11):3761-3772 [FREE Full text] [doi: 10.1007/s10803-018-3639-1] [Medline: 29948530]
17. Crane L, Chester JW, Goddard L, Henry LA, Hill E. Experiences of autism diagnosis: a survey of over 1000 parents in the United Kingdom. Autism 2016;20(2):153-162. [doi: 10.1177/1362361315573636] [Medline: 25810370]
18. Valentine AZ, Brown BJ, Groom MJ, Young E, Hollis C, Hall CL. A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: a map of the current evidence. Clin Psychol Rev 2020;80:101870 [FREE Full text] [doi: 10.1016/j.cpr.2020.101870] [Medline: 32712216]
19. Gibson AN, Kaplan S, Vardell E. A survey of information source preferences of parents of individuals with autism spectrum disorder. J Autism Dev Disord 2017;47(7):2189-2204. [doi: 10.1007/s10803-017-3127-z] [Medline: 28451948]
20. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Early Childhood Development Interventions Review Group, for the Lancet Early Childhood Development Series Steering Committee. Nurturing care: promoting early childhood development. Lancet 2017;389(10064):91-102. [doi: 10.1016/S0140-6736(16)31390-3] [Medline: 27717615]

XSL•FO
RenderX

21. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, Secretariat of the LancetFinancial Times Commission. The Lancet and Financial Times commission on governing health futures 2030: growing up in a digital world. Lancet 2021;398(10312):1727-1776. [doi: 10.1016/S0140-6736(21)01824-9] [Medline: 34706260]
22. Alon N, Perret S, Torous J. Working towards a ready to implement digital literacy program. Mhealth 2023;9:32 [FREE Full text] [doi: 10.21037/mhealth-23-13] [Medline: 38023777]
23. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
24. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
25. Harvey PD, Depp CA, Rizzo AA, Strauss GP, Spelber D, Carpenter LL, et al. Technology and mental health: state of the art for assessment and treatment. Am J Psychiatry 2022;179(12):897-914. [doi: 10.1176/appi.ajp.21121254] [Medline: 36200275]
26. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
27. Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, et al. eHealth tools that assess and track health and well-being in children and young people: systematic review. J Med Internet Res 2022;24(5):e26015 [FREE Full text] [doi: 10.2196/26015] [Medline: 35550285]
28. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
29. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
30. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry 2024;23(1):139-149 [FREE Full text] [doi: 10.1002/wps.21183] [Medline: 38214614]
31. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
32. Zervogianni V, Fletcher-Watson S, Herrera G, Goodwin M, Pérez-Fuster P, Brosnan M, et al. A framework of evidence-based practice for digital support, co-developed with and for the autism community. Autism 2020;24(6):1411-1422 [FREE Full text] [doi: 10.1177/1362361319898331] [Medline: 32026715]
33. Grant N, Rodger S, Hoffmann T. Evaluation of autism-related health information on the web. J Appl Res Intellect Disabil 2015;28(4):276-282. [doi: 10.1111/jar.12127] [Medline: 25224690]
34. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
35. Lim SL, Bentley PJ, Kanakam N, Ishikawa F, Honiden S. Investigating country differences in mobile app user behavior and challenges for software engineering. IEEE Trans Software Eng 2015;41(1):40-64. [doi: 10.1109/TSE.2014.2360674]
36. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, et al. Augmenting mental health in primary care: a 1-year study of deploying smartphone apps in a multi-site primary care/behavioral health integration program. Front Psychiatry 2019;10:94 [FREE Full text] [doi: 10.3389/fpsyt.2019.00094] [Medline: 30873053]
37. Torous J. A path towards progress: lessons from the hard things about digital mental health. World Psychiatry 2022;21(3):419-420 [FREE Full text] [doi: 10.1002/wps.21003] [Medline: 36073703]
38. Chittaranjan G, Blom J, Gatica-Perez D. Mining large-scale smartphone data for personality studies. Pers Ubiquit Comput 2011;17(3):433-450. [doi: 10.1007/s00779-011-0490-1]
39. Iovannone R, Dunlap G, Huber H, Kincaid D. Effective educational practices for students with autism spectrum disorders. Focus Autism Other Dev Disabl 2003;18(3):150-165. [doi: 10.1177/10883576030180030301]
40. Roberts AE, Davenport TA, Wong T, Moon H, Hickie IB, LaMonica HM. Evaluating the quality and safety of health-related apps and e-tools: adapting the mobile app rating scale and developing a quality assurance protocol. Internet Interv 2021;24:100379 [FREE Full text] [doi: 10.1016/j.invent.2021.100379] [Medline: 33777705]
41. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
42. Piasecki J, Waligora M, Dranseika V. Google search as an additional source in systematic reviews. Sci Eng Ethics 2018;24(2):809-810 [FREE Full text] [doi: 10.1007/s11948-017-0010-4] [Medline: 29249022]

XSL•FO
RenderX

43. D'Urzo KA, Man KE, Bassett-Gunter RL, Latimer-Cheung AE, Tomasone JR. Identifying "real-world" initiatives for knowledge translation tools: a case study of community-based physical activity programs for persons with physical disability in Canada. Transl Behav Med 2019;9(4):797-809. [doi: 10.1093/tbm/iby088] [Medline: 30184185]
44. Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: systematic search in app stores and evaluation. JMIR Mhealth Uhealth 2019;7(9):e12604 [FREE Full text] [doi: 10.2196/12604] [Medline: 31486408]
45. Camacho E, Cohen A, Torous J. Assessment of mental health services available through smartphone apps. JAMA Netw Open 2022;5(12):e2248784 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.48784] [Medline: 36576737]
46. Boulton KA, Coghill D, Silove N, Pellicano E, Whitehouse AJO, Bellgrove MA, et al. A national harmonised data collection network for neurodevelopmental disorders: a transdiagnostic assessment protocol for neurodevelopment, mental health, functioning and well-being. JCPP Adv 2021;1(4):e12048 [FREE Full text] [doi: 10.1002/jcv2.12048] [Medline: 37431407]
47. Frøen JF, Myhre SL, Frost MJ, Chou D, Mehl G, Say L, et al. eRegistries: electronic registries for maternal and child health. BMC Pregnancy Childbirth 2016;16:11 [FREE Full text] [doi: 10.1186/s12884-016-0801-7] [Medline: 26791790]
48. Robotham D, Waterman S, Oduola S, Papoulias C, Craig T, Wykes T. Facilitating mental health research for patients, clinicians and researchers: a mixed-method study. BMJ Open 2016;6(8):e011127 [FREE Full text] [doi: 10.1136/bmjopen-2016-011127] [Medline: 27503859]
49. Samji H, Wu J, Ladak A, Vossen C, Stewart E, Dove N, et al. Review: mental health impacts of the COVID-19 pandemic on children and youth - a systematic review. Child Adolesc Ment Health 2022;27(2):173-189 [FREE Full text] [doi: 10.1111/camh.12501] [Medline: 34455683]
50. Alon N, Torous J. Current challenges for evaluating mobile health applications. J Am Med Inform Assoc 2023;30(3):617-624 [FREE Full text] [doi: 10.1093/jamia/ocac244] [Medline: 36484621]
51. Carpenter-Song E, Acquilano SC, Noel V, Al-Abdulmunem M, Torous J, Drake RE. Individualized intervention to support mental health recovery through implementation of digital tools into clinical care: feasibility study. Community Ment Health J 2022;58(1):99-110 [FREE Full text] [doi: 10.1007/s10597-021-00798-6] [Medline: 33611684]
52. Perret S, Alon N, Carpenter-Song E, Myrick K, Thompson K, Li S, et al. Standardising the role of a digital navigator in behavioural health: a systematic review. Lancet Digit Health 2023;5(12):e925-e932 [FREE Full text] [doi: 10.1016/S2589-7500(23)00152-8] [Medline: 38000876]
53. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
54. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
55. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
56. Fleming T, Poppelaars M, Thabrew H. The role of gamification in digital mental health. World Psychiatry 2023;22(1):46-47 [FREE Full text] [doi: 10.1002/wps.21041] [Medline: 36640378]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
A-MARS: Adapted Mobile App Rating Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
NDC: neurodevelopmental condition

Edited by N Cahill; submitted 22.03.24; peer-reviewed by RM Garad, J Mistry; comments to author 20.04.24; revised version received 29.04.24; accepted 16.09.24; published 01.01.25.

Please cite as:
Boulton KA, Hilton M, Sutton E, Guastella AJ
Apps and Digital Resources for Child Neurodevelopment, Mental Health, and Well-Being: Review, Evaluation, and Reflection on Current Resources
J Med Internet Res 2025;27:e58693
URL: https://www.jmir.org/2025/1/e58693
doi:10.2196/58693
PMID:

XSL•FO
RenderX

©Kelsie Ann Boulton, Makana Hilton, Emilia Sutton, Adam John Guastella. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 01.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Machine Learning Approaches in High Myopia: Systematic Review and Meta-Analysis

Huiyi Zuo[1], MM; Baoyu Huang[1], MM; Jian He[1], MD; Liying Fang[1], MM; Minli Huang[1], MD

Ophthalmology Department, First Affiliated Hospital of GuangXi Medical University, Nanning, China

Corresponding Author:
Minli Huang, MD
Ophthalmology Department
First Affiliated Hospital of GuangXi Medical University
No 6 Shuangyong Road, Nanning, Guangxi
Nanning, 530000
China
Phone: 86 0771 5356507
Email: 420306@sr.gxmu.edu.cn

Abstract

Background: In recent years, with the rapid development of machine learning (ML), it has gained widespread attention from researchers in clinical practice. ML models appear to demonstrate promising accuracy in the diagnosis of complex diseases, as well as in predicting disease progression and prognosis. Some studies have applied it to ophthalmology, primarily for the diagnosis of pathologic myopia and high myopia-associated glaucoma, as well as for predicting the progression of high myopia. ML-based detection still requires evidence-based validation to prove its accuracy and feasibility.

Objective: This study aims to discern the performance of ML methods in detecting high myopia and pathologic myopia in clinical practice, thereby providing evidence-based support for the future development and refinement of intelligent diagnostic or predictive tools.

Methods: PubMed, Cochrane, Embase, and Web of Science were thoroughly retrieved up to September 3, 2023. The prediction model risk of bias assessment tool was leveraged to appraise the risk of bias in the eligible studies. The meta-analysis was implemented using a bivariate mixed-effects model. In the validation set, subgroup analyses were conducted based on the ML target events (diagnosis and prediction of high myopia and diagnosis of pathological myopia and high myopia-associated glaucoma) and modeling methods.

Results: This study ultimately included 45 studies, of which 32 were used for quantitative meta-analysis. The meta-analysis results unveiled that for the diagnosis of pathologic myopia, the summary receiver operating characteristic (SROC), sensitivity, and specificity of ML were 0.97 (95% CI 0.95-0.98), 0.91 (95% CI 0.89-0.92), and 0.95 (95% CI 0.94-0.97), respectively. Specifically, deep learning (DL) showed an SROC of 0.97 (95% CI 0.95-0.98), sensitivity of 0.92 (95% CI 0.90-0.93), and specificity of 0.96 (95% CI 0.95-0.97), while conventional ML (non-DL) showed an SROC of 0.86 (95% CI 0.75-0.92), sensitivity of 0.77 (95% CI 0.69-0.84), and specificity of 0.85 (95% CI 0.75-0.92). For the diagnosis and prediction of high myopia, the SROC, sensitivity, and specificity of ML were 0.98 (95% CI 0.96-0.99), 0.94 (95% CI 0.90-0.96), and 0.94 (95% CI 0.88-0.97), respectively. For the diagnosis of high myopia-associated glaucoma, the SROC, sensitivity, and specificity of ML were 0.96 (95% CI 0.94-0.97), 0.92 (95% CI 0.85-0.96), and 0.88 (95% CI 0.67-0.96), respectively.

Conclusions: ML demonstrated highly promising accuracy in diagnosing high myopia and pathologic myopia. Moreover, based on the limited evidence available, we also found that ML appeared to have favorable accuracy in predicting the risk of developing high myopia in the future. DL can be used as a potential method for intelligent image processing and intelligent recognition, and intelligent examination tools can be developed in subsequent research to provide help for areas where medical resources are scarce.

Trial Registration: PROSPERO CRD42023470820; https://tinyurl.com/2xexp738

(J Med Internet Res 2025;27:e57644) doi:10.2196/57644

KEYWORDS

high myopia; pathological myopia; high myopia-associated glaucoma; machine learning; deep learning

XSL•FO
RenderX

Introduction

Myopia is currently widely regarded as a significant public health issue, leading to substantial vision loss and serving as a risk factor for a range of other serious ocular diseases. It is estimated that by 2050, 4.758 billion people (49.8% of the world population) and 938 million people (9.8% of the world population) will suffer from myopia and high myopia, respectively [1]. A recent meta-analysis study proposed that the global economic burden due to productivity losses from uncorrected myopia and myopic macular degeneration is estimated to reach US $250 billion [2]. Therefore, the prevention of high myopia as well as the diagnosis and treatment of pathological myopia remain a formidable societal challenge.

High myopia is defined as the spherical equivalent ≤–6.0 diopter [3] when the accommodation of the eye is relaxed. However, increased severity of myopia and elongation of the eye's axial length could alter the posterior segment structures, causing posterior scleral staphyloma, myopic macular degeneration, and optic neuropathy related to high myopia, potentially leading to the loss of best-corrected visual acuity [3]. High myopia-related fundus lesions stand as an important contributing factor to blindness across the world as well as in China [4]. The detection of high myopia hinges primarily on artificial auxiliary techniques, like refraction detection, fundus examination, measurement of axial length, and fundus photography. Nevertheless, manual examination and analysis by ophthalmologists are still essential, necessitating a significant investment of time and effort [5]. Additionally, in regions with limited medical resources, the shortage of ophthalmologists and medical equipment impedes the early and accurate identification of high-risk patients with high myopia, resulting in missed opportunities for optimal treatment. Therefore, forecasting the risk of high myopia and precisely diagnosing pathological myopia are currently major research focus.

With the rapid advances in computing technology and the ongoing refinement of statistical theory, machine learning (ML) has gradually been promoted and applied in clinical practice. For instance, ML can not only improve image quality, reduce misregistration, and simulate attenuation correction imaging in core cardiology [6], but also be used for cancer screening (detection of lesions), characterization and grading of tumors, and prognosis prediction, thus facilitating clinical decision-making [7]. Since fundus images are noncontact, noninvasive, low-cost, easily accessible, and easy to process, ML has been extensively used to diagnose common retinal diseases, including diabetic retinopathy [8-10], macular degeneration [10], and glaucoma [11-13]. ML has been applied to various image-processing tasks. Novel techniques for analyzing fundus images of high myopia and pathological myopia are continuously emerging [14,15]. However, the accuracy of these ML detections has not been systematically studied. Consequently, the present study was executed to comprehensively describe the accuracy of ML in detecting different degrees of lesions in high myopia, furnishing an evidence-based reference for subsequent lesion management.

Methods

Study Registration

This study was implemented as per the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines and prospectively registered with PROSPERO (ID: CRD42023470820). The PRISMA checklist is available in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

We established detailed inclusion and exclusion criteria for this systematic review. To enhance visualization, these criteria are presented in tabular form (Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Study type: (1) case-control, cohort, nested case-control, and case-cohort studies and (2) studies reported in English.
- Machine learning (ML): studies that fully constructed ML models for the prediction or diagnosis of high myopia, the diagnosis of pathological myopia, or the diagnosis of high myopia-associated glaucoma.
- Outcome measures: at least one of the following outcome indicators were reported: receiver operating characteristic (ROC), *c*-index, sensitivity, specificity, accuracy, recovery rate, accuracy rate, confusion matrix, F_1-score, and calibration curve.
- Datasets: (1) some studies lacked independent validation sets, and only *k*-fold cross-validation was leveraged to verify the effect of the constructed mode; and (2) in some publicly available datasets, particularly those involving medical imaging, different studies have reported the efficiency of varying ML methods.

Exclusion criteria

- Study type: (1) meta, review, guide, expert opinion; and (2) studies with too few samples (less than 20 cases).
- ML: literature that only executed the risk factor analysis but did not develop a complete ML mode.
- Outcome measures: none of the following outcomes were reported: ROC, *c*-index, sensitivity, specificity, accuracy, recovery rate, accuracy rate, confusion matrix, F_1-score, and calibration curve.

XSL•FO
RenderX

Data Sources and Search Strategy

PubMed, Cochrane, Embase, and Web of Science were thoroughly retrieved up to September 3, 2023, using the form of MeSH (Medical Subjects Headings) + free term, without any restrictions on region or publication period. The specific search strategy is depicted in Multimedia Appendix 2.

Study Selection and Data Extraction

Duplicates were excluded from the retrieved literature, and titles and abstracts were reviewed to delete obviously irrelevant studies. The full texts of the remaining studies were then downloaded and thoroughly read to determine the final included studies in the systematic review. A standard electronic data extraction spreadsheet was prepared prior to extracting data. The extracted data encompassed the title, first author, type of study, year of publication, author's country, patient source, target event, number of cases of the target event, the total number of cases, number of training set cases, the total number of training set cases, method of validation set generation, number of events in the validation set, total number of cases in the validation set, type of models, and modeling variables.

Two researchers (HZ and LF) independently screened the literature and extracted data. Upon completion, their findings were cross-checked. A third reviewer (JH) was consulted for resolution in case of any dissents.

Risk of Bias in Studies

The risk of bias in the eligible studies was appraised by two independent reviewers (HZ and LF) using the prediction model risk of bias assessment tool [16]. This tool is comprised of a large number of questions in four domains (participants, predictors, outcomes, and analysis), which reflect overall bias risk and applicability. The 4 domains involve 2, 3, 6, and 9 specific questions, respectively, and each question may be answered by yes or probably yes, no or probably not, or no information. Following the quality evaluation, a cross-check was carried out. In the event of any disputes, a third researcher (JH) was consulted for resolution.

Synthesis Methods

In some of the original studies included in our research, there was not only 1 validation set. Therefore, the number of models included in the meta-analysis does not equal the number of studies. The meta-analysis of sensitivity and specificity was executed using a bivariate mixed-effects model [17]. Sensitivity and specificity were meta-analyzed as per the diagnostic 2×2 table. However, most included studies did not provide the diagnostic 2×2 table. In such cases, the following two approaches were used to calculate the diagnostic 2×2 table: (1) it was computed based on sensitivity, specificity, and precision, combined with the number of cases; and (2) sensitivity and specificity were extracted based on the optimal Youden index, and then combined with the number of cases for calculation. The meta-analysis was implemented using R (version 4.2.0; R Foundation for Statistical Computing).

Results

Study Selection

A total of 4214 records were retrieved from the databases, of which 582 were duplicates. After reading the titles and abstracts, 3561 studies unrelated to ML in high myopia were excluded, leaving 71 studies. Of these, 13 only conducted image segmentation without constructing ML models, 5 did not provide full extractable outcome indicators, and 8 analyzed risk factors. Ultimately, 45 studies were incorporated into this review. The literature screening process is depicted in Figure 1.

Figure 1. Flowchart of literature screening.

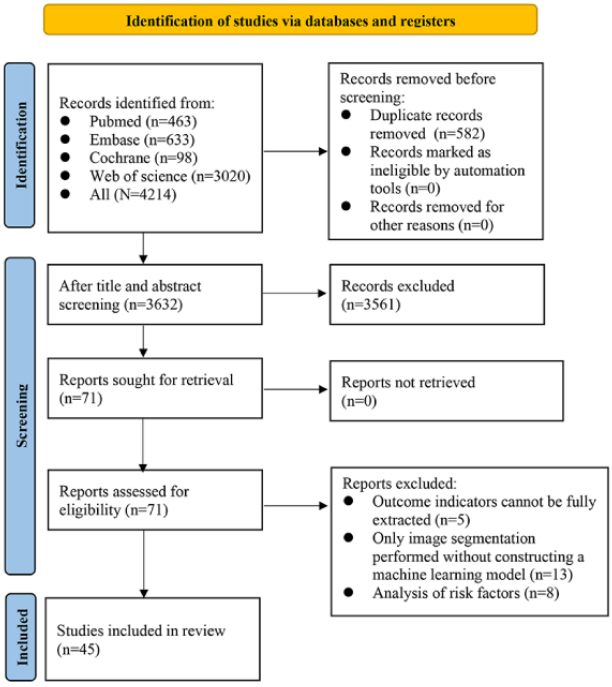

XSL•FO
RenderX

Study Characteristics

The included studies were published from 2010 to 2023. Four of the studies [18-21] were about the prediction of high myopia, and the predicted variables were mainly derived from life characteristics, environmental and genetic factors, and routinely interpretable ocular clinical characteristics. Five of the studies [22-26] were about the diagnosis of high myopia, of which 1 study [22] also involved the diagnosis of pathological lesions of high myopia. Six studies focused on the diagnosis of high myopia-associated glaucoma [27-32]. Out of the included studies, 31 studies focused on the diagnosis of pathological myopia, primarily using optical coherence tomography and fundus imaging to construct artificial intelligence models. Of these, 26 studies [4,15,22,33-55] were based on DL (deep learning), while 5 studies [56-60] required manually coded ML for construction. Additionally, it was noted that in the 45 original studies, all 45 studies included binary classification tasks, with 9 studies [4,33,34,38,39,49,50,52,61] additionally incorporating multiclassification tasks. Regarding validation methods, 31 studies provided an external validation set, and 23 used a combination of internal and external validation sets. In terms of the generation method of validation set, 6 studies [23,24,34,40,47,59] used *k*-fold cross-validation, 29 [15,19-22,25-29,35-38,41,42,45,48-58,61] used random sampling, and 6 [4,18,32,33,44,60] applied a combination of *k*-fold cross-validation and random sampling. The detailed characteristics of the eligible studies are shown in Tables 1 and 2.

XSL•FO
RenderX

Table 1. Fundamental features of included studies.

First author	Year of publication	Country of authors	Study type	Patient source	Target events	Total number of cases
Tang et al [33]	2022	China, United States	Retrospective study	Multicenter	Diagnosis of pathologic myopia	1395 fundus photographs, 895 patients
Li et al [56]	2023	China	Nested case-control study	Single center	Diagnosis and prediction of pathological myopia	20,870 patients
Du et al [57]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia	313 patients with high myopia and 457 eyes
Foo et al [18]	2023	Singapore	Prospective study	Multicenter	Prediction of high myopia	965 children with 1878 eyes and 7456 fundus photographs
Kim et al [58]	2021	Korea	Retrospective study	Multicenter	Diagnosis of pathologic myopia	860 eyes
Zhang et al [59]	2013	Singapore	Retrospective study	Registry database	Diagnosis of pathologic myopia	2258 patients
Zhu et al [34]	2023	China	Retrospective study	Single center	Diagnosis of pathologic myopia	6078 photographs
Wu et al [35]	2022	China	Retrospective study	Single center	Diagnosis of pathologic myopia	1853 photographs
Ye et al [36]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia	1041 patients with pathologic myopia and with 2342 eligible OCT[a] macular images
Wang et al [37]	2023	China	Retrospective study	Single center	Diagnosis of pathologic myopia	7606 patients with 10,347 fundus photographs
Wang et al [19]	2022	China	Prospective, longitudinal, observational study	Wenzhou large-scale survey	Prediction of myopia and high myopia	15,765 patients
Wan et al [4]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia	858 photographs
Wan et al [38]	2023	China	Retrospective study	Single center	Diagnosis of pathologic myopia	1750 photographs
Tan et al [22]	2021	Singapore	Retrospective multicohort study	Multicenter + registry database	Diagnosis of high myopia + pathological myopia	125,421 patients with 251,349 photographs
Sun et al [39]	2023	China	Retrospective multicohort study	Multicenter + registry database	Diagnosis of pathologic myopia	1514 fundus photographs
Sogawa et al [40]	2020	Japan	Retrospective study	Single center	Diagnosis of pathologic myopia	910 patients with 910 images
Du et al [41]	2022	Japan	Retrospective study	Single center	Diagnosis of pathologic myopia	1327 patients with 2400 high myopia eyes and 9176 OCT images
Hou et al [60]	2023	China	Prospective cohort study	Single center	Diagnosis of pathologic myopia	576 patients
Li et al [52]	2022	China	Retrospective cohort study	Multicenter	Pathologic myopia	29,230 patients with 57,148 fundus photographs
Li et al [27]	2021	China	Case-control study	Multicenter	Diagnosis of glaucoma in high myopia	2731 participants with 2731 eyes
Chen et al [20]	2019	China	Prospective study	Single center	Prediction of high myopia	1063 patients
Choi et al [23]	2021	Korea	Retrospective study	Single center	Prediction of high myopia	492 patients with 690 eyes
Cui et al [42]	2021	China, Taiwan	Retrospective study	Registry database	Diagnosis of pathologic myopia	800 images

XSL•FO
RenderX

First author	Year of publication	Country of authors	Study type	Patient source	Target events	Total number of cases
Guan et al [24]	2023	China	Retrospective study	Multicenter	Prediction of high myopia	1,285,609 participants
He et al [61]	2022	China	Retrospective study	Multicenter	Diagnosis of pathologic myopia	2866 patients with 3945 OCT images
Hemelings et al [15]	2021	Belgium	Retrospective study	Registry database	Diagnosis of pathologic myopia	1200 photographs
Rauf et al [44]	2021	Pakistan	Retrospective study	Registry database	Diagnosis of pathologic myopia	840 photographs
Park et al [45]	2022	Korea	Retrospective study	Single center	Diagnosis of pathologic myopia	367 eyes
Lu et al [46]	2021	China	Retrospective study	Single center	Diagnosis of pathologic myopia and diagnosis of pathologic myopia	• 17,330 photographs • 17,330 photographs
Lu et al [47]	2021	China	Retrospective study	Multicenter	Diagnosis of pathologic myopia	32,419 patients with 37,659 images
Liu et al [54]	2010	Singapore	Retrospective study	Single center	Pathologic myopia	80 photographs
Li et al [48]	2022	China, United States	Retrospective study	Single center	Diagnosis of pathologic myopia	1139 patients with 5917 images
Lee et al [28]	2023	Korea	Retrospective study	Single center	Diagnosis of glaucoma in high myopia	260 eyes and 260 images
Kim et al [29]	2023	Korea	Retrospective study	Single center	Diagnosis of glaucoma in high myopia	2607 eyes
Jeong et al [30]	2023	Korea	Retrospective cross-sectional study	Single center	Diagnosis of glaucoma in high myopia	274 patients
Huang et al [21]	2022	China	Case-control study	Single center	Prediction of high myopia	1298 patients
Huang et al [49]	2023	China, United Kingdom	Retrospective study	Single center	Diagnosis of pathologic myopia	1131 patients with 3441 images
Du et al [50]	2021	Japan	Retrospective study	Single center	diagnosis of pathologic myopia	4432 eyes and 7020 images
Crincoli et al [51]	2023	Italy	Case-control study	Multicenter	diagnosis of pathologic myopia	84 patients with 84 eyes and 252 photographs
Asaoka et al [31]	2014	Japan	Case-control study	Multicenter	Diagnosis of glaucoma in high myopia	242 patients and 242 eyes
Bowd et al [32]	2023	United States, Germany	Retrospective study	Single center	Diagnosis of glaucoma in high myopia	593 eyes
Zhao et al [25]	2022	China	Retrospective study	Single center	Prediction of high myopia	546 patients
Liu et al [53]	2010	Singapore	Retrospective study	Single center	Diagnosis of pathologic myopia	80 photographs
Dai et al [26]	2020	China	Retrospective study	Single center	Prediction of high myopia	319 patients with 932 images
Baid et al [55]	2019	India	Retrospective study	Registry database	Diagnosis of pathologic myopia	481 photographs

[a]OCT: optical coherence tomography.

XSL•FO
RenderX

Table 2. Fundamental features of included studies.

Total number of cases in training set	Generation of validation set	Total number of cases in validation set	Total number of cases in test set	Model type	Modeling variables
727 fundus photographs	5-fold cross-validation + random sampling	238 fundus photographs	238 fundus photographs	DL[a]	Fundus photographs
2069 patients	Random sampling	1382 patients	Unclear	ACP[b], ML[c]	Clinical features
319 eyes	Random sampling	138 eyes	Unclear	ML-based radiomics analysis method	Fundus photographs
769 children with 1502 eyes and 5945 photographs	Internal validation (5-fold cross-validation + random sampling) + multicenter external validation	196 children with 376 eyes and 1511 fundus photographs	99 children with 189 eyes and 821 photographs	DL	Fundus photographs + clinical features
602 eyes	Random sampling	258 eyes	unclear	SVM[d], ML	Fundus photographs
2258 patients	Stratified 20-fold cross-validation	unclear	unclear	SVM, ML	SNP[e] + clinical features + fundus photographs
4252 photographs	Stratified 20-fold cross-validation	unclear	1826 photographs	DL	Fundus photographs
1483 photographs	Random sampling	unclear	370 photographs	DL	Fundus photographs
1874 photographs	Internal validation (random sampling) + external validation（multicenter）	468 photographs	450 photographs	DL	Fundus photographs
5003 patients with 7389 photographs	Random sampling	775 patients with 821 photographs	1828 patients with 2137 photographs	DL	Fundus photographs
11,350 patients	Internal validation (random sampling) + external validation(prospective)	4415 patients	6168 patients (prognostic cohort)	LR[f], GBDT[g], NN[h]	Clinical features
758 photographs	5-fold cross-validation + random sampling	100 photographs	Unclear	DL	Fundus photographs
1402 photographs	Random sampling	174 photographs	174 photographs	DL	Fundus photographs
226,686 photographs	Internal validation (random sampling) + external validation（multicenter）	11,303 photographs	213,475 photographs	DL	Fundus photographs
400 fundus photographs	Multicenter	400 fundus photographs	714 fundus photographs	DL	Fundus photographs
Unclear	5-fold cross-validation	Unclear	Unclear	DL	Fundus photographs
7865 photographs	random sampling	1311 photographs	Unclear	DL	Fundus photographs
516 patients	10-fold cross-validation + random sampling	60 patients	Unclear	XGBoost[i], SVM, LR	Clinical features + metabolic characteristics
29,213 photographs	Internal validation (random sampling) + external validation（multicenter）	7302 photographs	16,554 photographs	DCNN[j], DL	Fundus photographs
2223 participants with 2223 eyes	Random sampling	508 participants with 508 eyes	Unclear	FCN[k]	OCT[l] images + clinical features
638 patients	Random sampling	425 patients	Unclear	LR	Genetic factors + clinical features
434 patients with 600 eyes and 1200 images	5-fold cross-validation	Unclear	58 patients with 90 eyes and 180 images	CNN[m], DL	OCT images
400 images	Random sampling	200 images	200 images	DL	Fundus photographs
1600 participants	Internal validation (5-fold cross-validation)	Unclear	400 patients	RF[n], LR, SVM	Clinical features
2380 images	Random sampling	680 photographs	340 photographs	DL	OCT images

XSL•FO
RenderX

Total number of cases in training set	Generation of validation set	Total number of cases in validation set	Total number of cases in test set	Model type	Modeling variables
400 photographs	Random sampling	400 photographs	400 photographs	DL	Fundus photographs
400 photographs	10-fold cross-validation + random sampling	40 photographs	400 photographs	DL	Fundus photographs
293 eyes	random sampling	37 eyes	37 eyes	DL	3D OCT images
11,502 photographs	Unclear	3284 photographs	1642 photographs	DL	Fundus photographs
2457 photographs	Unclear	707 photographs	372 photographs	DL	Fundus photographs
32,010 images	Internal validation (5-fold cross-validation) + external validation (multicenter)	Unclear	732 patients with 1000 images	DL	Fundus photographs
40 photographs	Random sampling	Unclear	40 photographs	SVM, DL	Fundus photographs
838 patients with 4338 images	Internal validation (random sampling) + external validation (prospective)	210 patients with 1167 photographs	91 patients with 174 eyes and 412 photographs	DL	OCT macular images
165 images	Random sampling	46 photographs	49 photographs	DL	OCTA[o] and OCT images
1416 eyes	Internal validation (random sampling) + external validation	471 eyes	720 eyes	DL	OCT images
Unclear	Unclear	Unclear	Unclear	Decision tree	OCT images
325 patients	Random sampling	973 patients	Unclear	DL	Genetic + clinical features
2264 images	Internal validation (random sampling) + external validation (prospective)	501 photographs	604 photographs	DL	OCT images
4140 photographs	Random sampling	1036 photographs	1844 photographs	DL	Fundus photographs
176 photographs	Random sampling	25 photographs	51 photographs	DL	OCT images
Unclear	Unclear	Unclear	Unclear	RF	HRT parameters
347 eyes	5-fold cross-validation + random sampling	87 eyes	159 eyes	CNN	OCT images
928 fundus photographs	Random sampling	232 photographs	Unclear	DL	Fundus photographs
40 photographs	Random sampling	Unclear	40 photographs	SVM or DL	Fundus photographs
792 photographs	Random sampling	Unclear	140 photographs	DL	Fundus photographs
374 photographs	Random sampling	80 photographs	27 photographs	CNN	Fundus photographs

[a]DL: deep learning.

[b]ACP: algorithm of conditional probability.

[c]ML: machine learning.

[d]SVM: support vector machine.

[e]SNP: single nucleic polymorphism.

[f]LR: logistic regression.

[g]GBDT: gradient boosted decision tree.

[h]NN: neural network.

[i]XGBoost: extreme gradient boosting

[j]DCNN: deep convolutional neural networks

[k]FCN: fully connected network

[l]OCT: optical coherence tomography.

[m]CNN: convolutional neural networks

[n]RF: random forest

[o]OCTA: optical coherence tomography angiography.

XSL•FO
RenderX

Risk of Bias in Studies

This review incorporated 67 models. There were 36 retrospective studies [4,15,22-26,28-30,32-42,44-50,52-55,57-59,61] that constructed 39 models, indicating a high bias in the choosing of study participants. Five case-control studies [21,27,31,51,56] constructed 13 models, also showing high bias in the selection of study participants. Since the predictors were evaluated in the context of a known outcome in the case-control studies, there was a high bias in the assessment of predictive factors. Twelve studies [19,20,23,24,27,30,31,56-60] constructed 22 models based on manually coded ML, with a high bias in predictive factors. In terms of statistical analysis, 2 studies [21,45] with 5 models did not meet the requirement of having an event per variable>20%, indicating a high risk of bias. In the statistical analysis, 32 models in 34 studies [4,15,18,21-23,25,26,28,29,32-42,44-55,61] could not estimate event per variable due to the use of the DL method. Additionally, 10 studies [19,20,24,27,30,31,56,58-60] with 29 models in ML did not report on the complexity of the data, rendering it difficult to determine their bias risk. Five studies [20,27,30,31,60] with 11 models were identified as having a high risk of bias in statistical analysis because they did not perform cross-validation to adjust the stability of models with different parameters. In summary, in terms of research participants, 14 models had a low risk of bias; 52 models had a high risk of bias, and 1 model had an unclear risk of bias. In terms of predictors, 37 models had a low risk of bias and 30 models had a high risk of bias. In terms of outcomes, all 67 models had a low risk of bias. In terms of statistical analysis, 3 models had a low risk of bias, 16 models had a high risk of bias, and 48 models had an unclear risk of bias.

Meta-Analysis of ML for Binary Classification Tasks

Pathological Myopia

Twenty studies [26,34-37,39,41,45,47-54,56,58,60,61] reported ML for diagnosing pathological myopia. Modeling algorithms included algorithms of conditional probability, support vector machines (SVMs), logistic regression (LR), extreme gradient boosting, convolutional neural networks (CNNs), and deep convolutional neural networks (DCNNs). The overall sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and summary receiver operating characteristic (SROC) were 0.91 (95% CI 0.89-0.92), 0.95 (95% CI 0.94-0.97), 19.7 (95% CI 13.8-28.2), 0.10 (95% CI 0.08-0.12), 201 (95% CI 122-331), and 0.97 (95% CI 0.95-0.98), respectively. The Deek funnel plot indicated no substantial evidence of publication bias in the included ML models. Assuming that the prior probability of pathological myopia was 20% if the result of ML was pathological myopia, then the probability of true pathological myopia would be 83%. If the result of ML was nonpathological myopia, then the probability of true pathological myopia would be 2% (ie, the probability of true nonpathological myopia was 98%; Figure 2 and Figures S1-S3 in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Forest plot for the meta-analysis of sensitivity and specificity of machine learning in detecting pathological myopia [26,34-37,39,41,45,47-54,56,58,60,61]. Note: the pooled sensitivity and specificity of 44 models from 20 machine learning studies on the diagnosis of pathological myopia were 0.91 (95% CI 0.89-0.92) and 0.95 (95% CI 0.94-0.97), respectively.

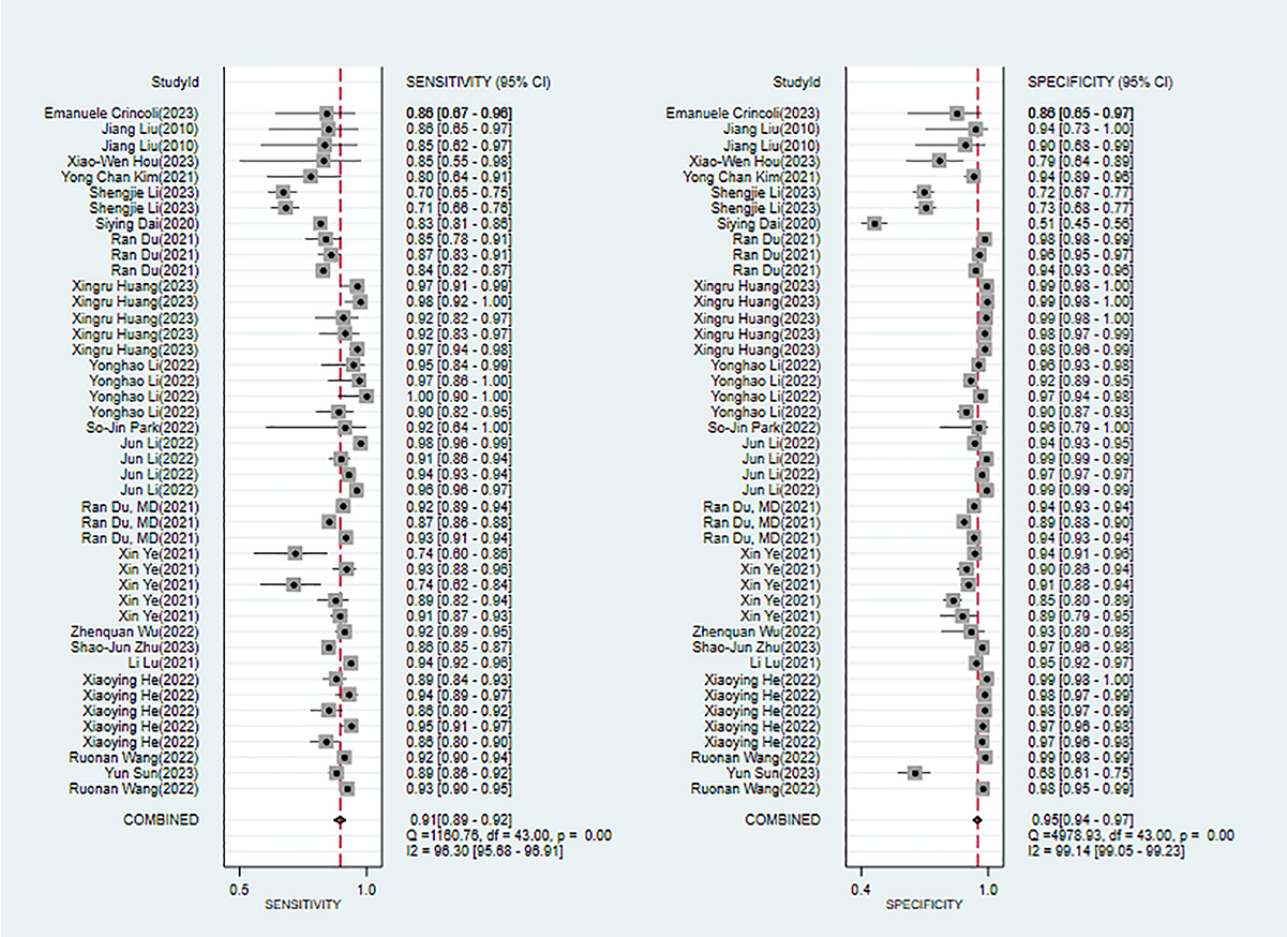

Five studies [53,54,56,58,60] reported conventional ML (non-DL) for diagnosing pathological myopia. Modeling algorithms included algorithms of conditional probability, SVM, extreme gradient boosting, and LR. The overall sensitivity, specificity, PLR, NLR, DOR, and SROC curve were 0.77 (95% CI 0.69-0.84), 0.85 (95% CI 0.75-0.92), 5.2 (95% CI 2.8-9.8), 0.27 (95% CI 0.18-0.39), 20 (95% CI 7-51), and 0.86 (95% CI 0.75-0.92), respectively. The Deek funnel plot indicated the presence of publication bias in the conventional ML (non-DL) models. Assuming that the prior probability of pathological myopia for conventional ML (non-DL) was 20% if the result of conventional ML (non-DL) was pathological myopia, then the probability of true pathological myopia would be 57%. If the result of conventional ML (non-DL) was nonpathological myopia, then the probability of true pathological myopia would be 6% (ie, the probability of true nonpathological myopia was 94%; Figure 3 and Figures S4-S6 in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 3. Forest plot for the meta-analysis of sensitivity and specificity of conventional machine learning (non-deep learning) in detecting pathological myopia [53,54,56,58,60]. Note: the pooled sensitivity and specificity of 6 models from 5 conventional machine learning (non-deep learning) studies on the diagnosis of pathological myopia were 0.77 (95% CI 0.69-0.84) and 0.85 (95% CI 0.75-0.92), respectively.

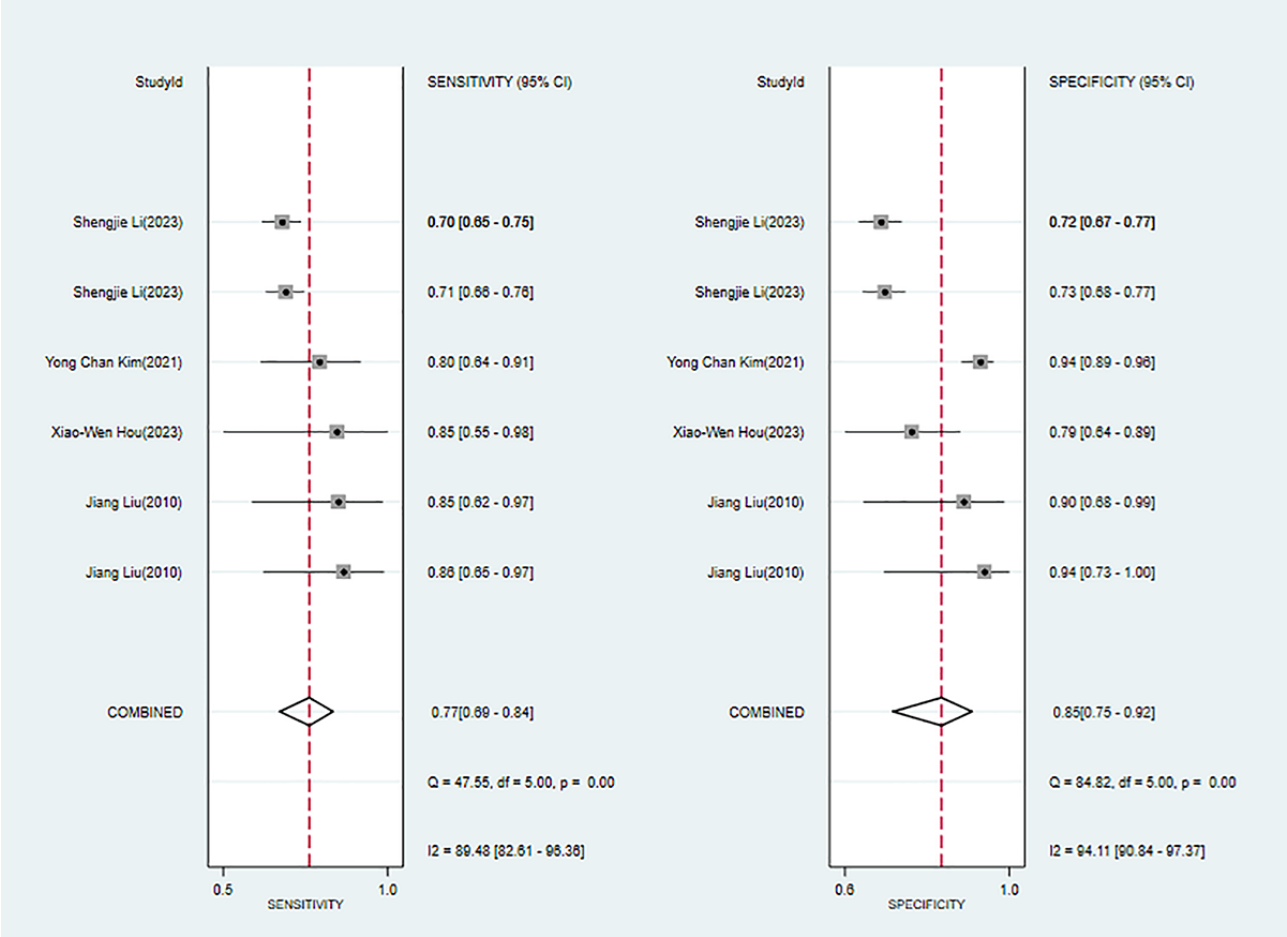

Fifteen studies [26,34-37,39,41,45,47-52,61] mentioned DL for diagnosing pathological myopia. Modeling algorithms included CNN and DCNN. The overall sensitivity, specificity, PLR, NLR, DOR, and SROC were 0.92 (95% CI 0.90-0.93), 0.96 (95% CI 0.95-0.97), 23.7 (95% CI 16.5-34.0), 0.09 (95% CI 0.07-0.11), 271 (95% CI 168-437), and 0.97 (95% CI 0.95-0.98), respectively. The Deek funnel plot revealed no remarkable publication bias in the DL models. Assuming that the prior probability of pathological myopia for DL was 20% if the result of DL was pathological myopia, then the probability of true pathological myopia would be 86%. If the result of DL was nonpathological myopia, then the probability of true pathological myopia would be 2% (ie, the probability of true nonpathological myopia was 98%; Figure 4 and Figures S7-S9 in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 4. Forest plot for the meta-analysis of sensitivity and specificity of deep learning in detecting pathological myopia [26,34-37,39,41,45,47-52,61]. Note: the pooled sensitivity and specificity of 38 models from 15 deep learning studies on the diagnosis of pathological myopia were 0.92 (95% CI 0.90-0.93) and 0.96 (95% CI 0.95-0.97), respectively.

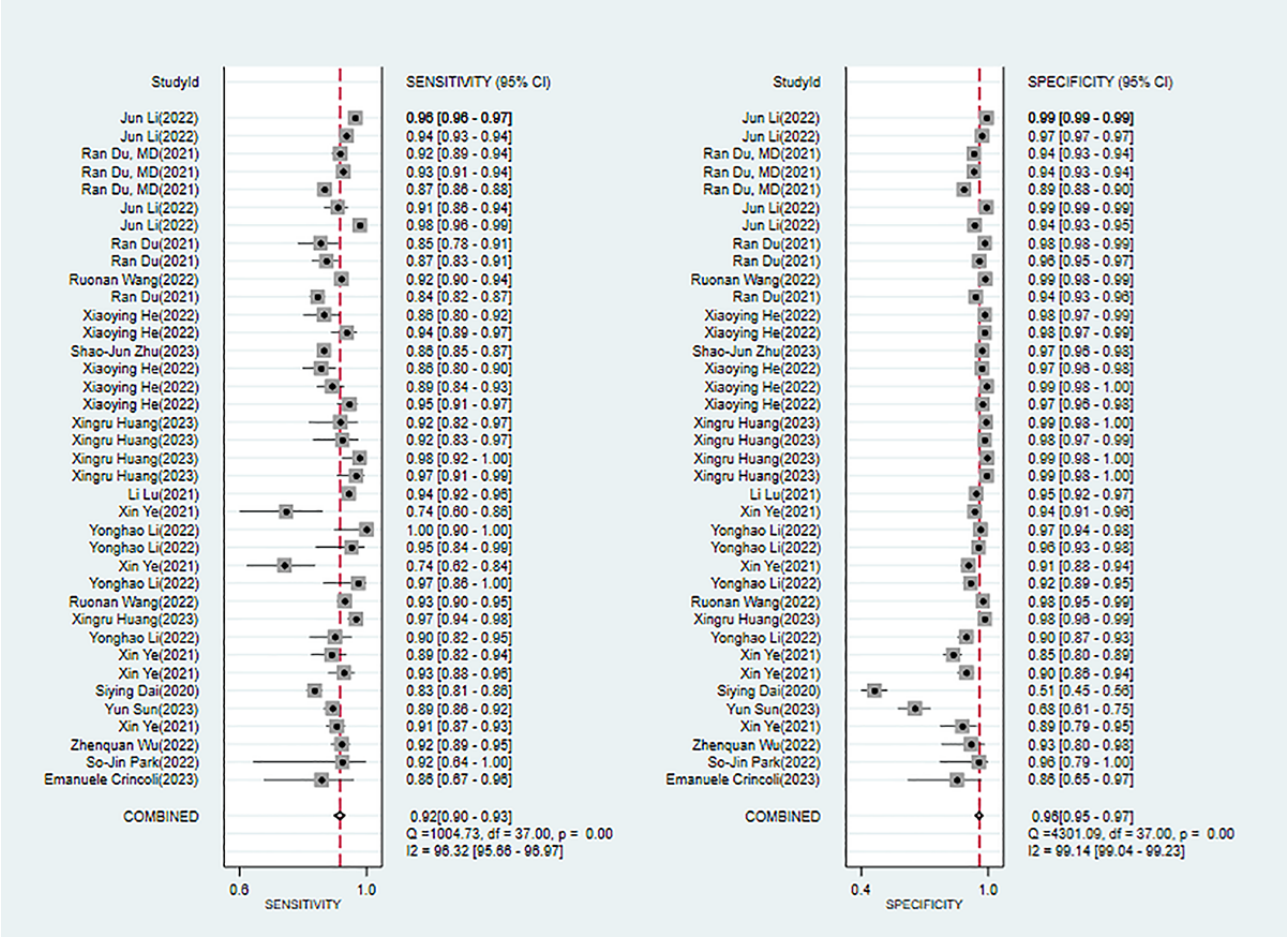

High Myopia

Six studies [4,18,20,23-25] discussed ML for diagnosing and forecasting high myopia. Modeling algorithms included DCNN, CNN, LR, SVM, random forest (RF), and linear mixed models. The sensitivity, specificity, PLR, NLR, DOR, and SROC were 0.94 (95% CI 0.90-0.96), 0.94 (95% CI 0.88-0.97), 16.2 (95% CI 7.7-33.8), 0.06 (95% CI 0.04-0.11), 255 (95% CI 79-822), and 0.98 (95% CI 0.96-0.99), respectively. The Deek funnel plot indicated no substantial evidence of publication bias in the included ML models. Assuming that the prior probability of high myopia for ML was 20% if the result of ML was high myopia, then the probability of true high myopia would be 80%. If the result of ML was non-high myopia, then the probability of true high myopia would be 2% (ie, the probability of true non-high myopia was 98%; Figure 5 and Figures S10-S12 in Multimedia Appendix 3).

Three studies [4,23,25] focused on diagnosing high myopia, while 3 studies [18,20,24] focused on predicting high myopia. Due to the limited number of studies included, we did not perform a meta-analysis for the diagnostic and prediction tasks. In the validation sets of the diagnostic tasks, sensitivity ranged from 0.91 to 1.00 and specificity ranged from 0.85 to 1.00, while in the validation sets of the prediction tasks, these values were 0.85-0.94 and 0.86-0.94, respectively. We found that both diagnostic and prediction tasks demonstrated highly favorable performance.

XSL•FO
RenderX

Figure 5. Forest plot for the meta-analysis of sensitivity and specificity of machine learning in detecting high myopia [4,18,20,23-25]. Note: the pooled sensitivity and specificity of 9 models from 6 machine learning studies on the diagnosis and prediction of high myopia were 0.94 (95% CI 0.90-0.96) and 0.94 (95% CI 0.88-0.97), respectively.

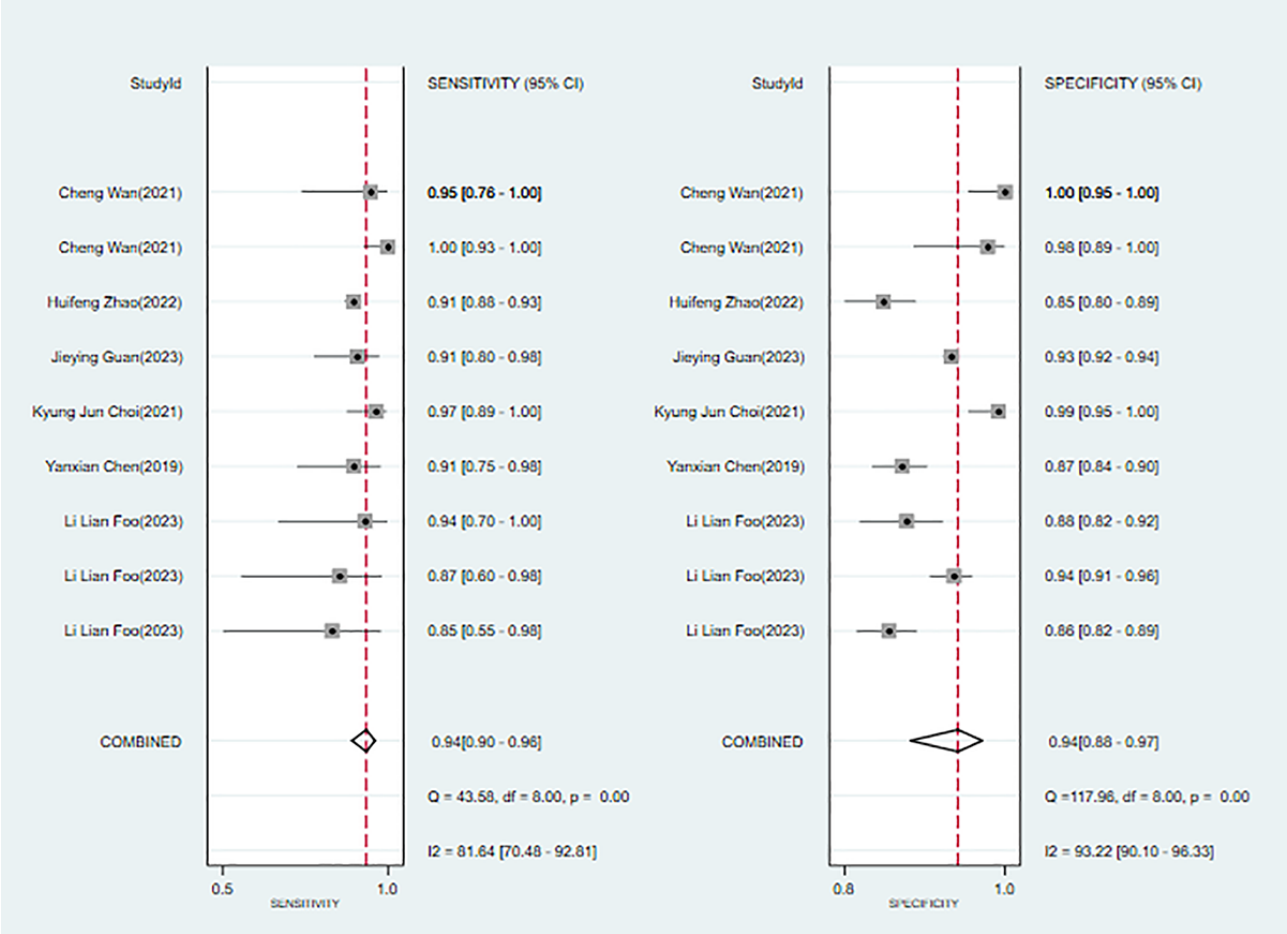

High Myopia–Associated Glaucoma

Six studies [27-32] mentioned ML for diagnosing high myopia-associated glaucoma. Modeling algorithms included Lagrange multiplier, fully connected network, radial basis function network, decision tree, RF, and CNN. The sensitivity, specificity, PLR, NLR, DOR, and SROC curve were 0.92 (95% CI 0.85-0.96), 0.88 (95% CI 0.67-0.96), 7.6 (95% CI 2.4-23.8), 0.09 (95% CI 0.04-0.20), 84 (95% CI 13-555), and 0.96 (95% CI 0.94-0.97), respectively. The Deek funnel plot indicated no substantial evidence of publication bias in the included ML models. Assuming that the prior probability of high myopia–associated glaucoma was 20% if the result of ML was high myopia-associated glaucoma, then the probability of true high myopia–associated glaucoma would be 65%. If the result of ML was non-high myopia–associated glaucoma, then the probability of true high myopia–associated glaucoma would be 2% (ie, the probability of true non-high myopia–associated glaucoma was 98%; Figure 6 and Figures S13-S15 in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 6. Forest plot for the meta-analysis of sensitivity and specificity of machine learning in detecting high myopia-associated glaucoma [27-32]. Note: the pooled sensitivity and specificity of 9 models from 6 machine learning studies on the diagnosis of high myopia-associated glaucoma were 0.92 (95% CI 0.85-0.96) and 0.88 (95% CI 0.67-0.96), respectively.

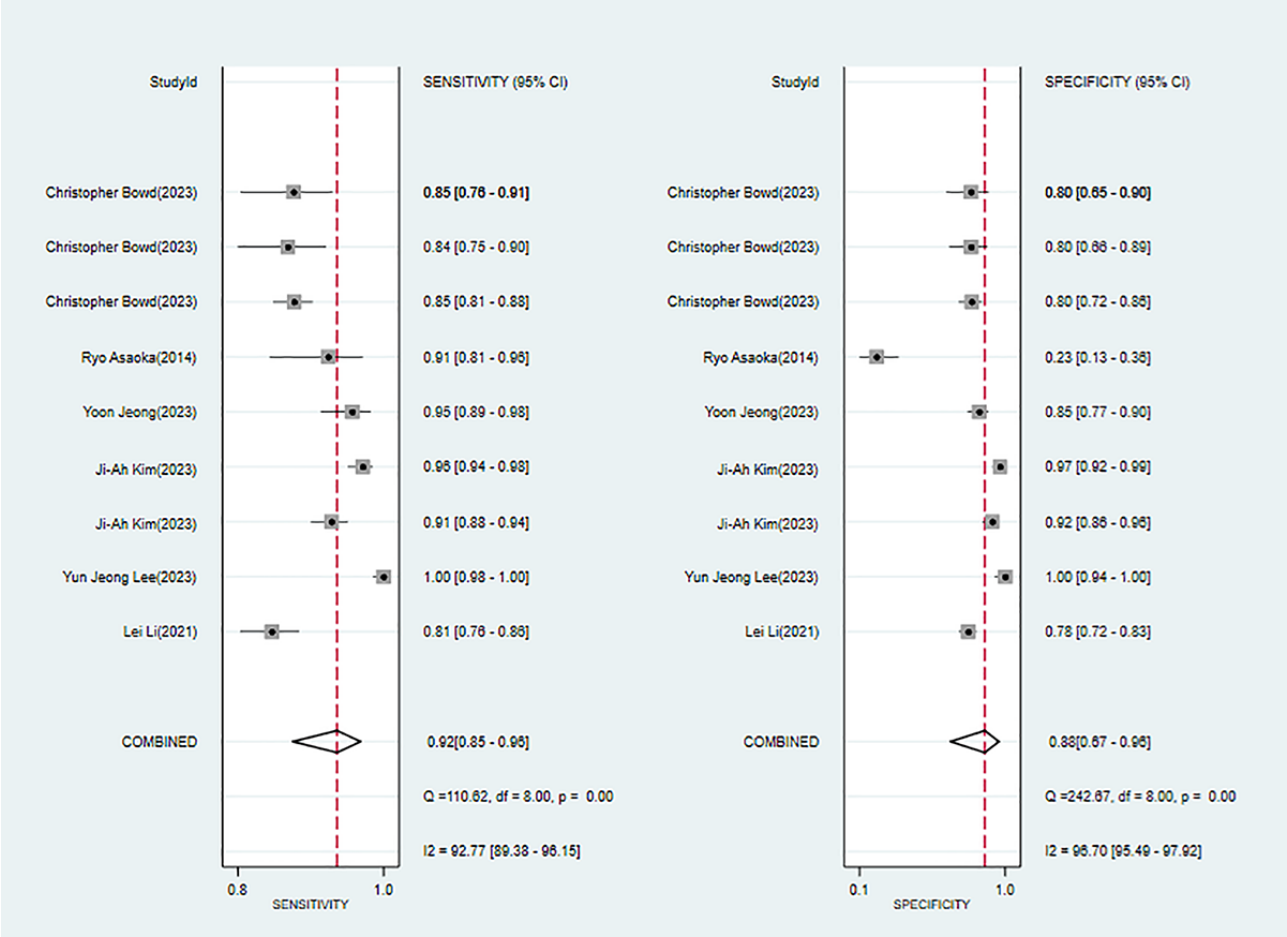

Review of ML for Multiclassification Tasks

Out of the included studies, 9 [4,33,34,38,39,49,50,52,61] used ML for multiclassification tasks. Due to significant variations in the diagnostic differences across these multiclassification tasks, a quantitative analysis was not feasible. Five studies [33,34,38,39,50] focused on fundus images–based DL to detect different types of myopic atrophy maculopathy in high myopia, with an accuracy ranging from 88% to 97%. Two studies [49,61] used optical coherence tomography (OCT) image–based DL to detect different types of myopic traction maculopathy in high myopia, with an accuracy ranging from 91% to 96%. One study [4] used fundus image–based DL to differentiate between normal, low-risk high myopia, and high-risk high myopia, with an accuracy of 99%. One study [52] applied fundus image–based DL to distinguish between normal, fundus tessellation, and pathologic myopia, with an accuracy of 94%, as illustrated in Table 3.

XSL•FO
RenderX

Table 3. Results of machine learning for multiclassification tasks.

First author	Year	Diagnostic purpose	Types of artificial intelligence	Modeling variables	Generation of validation set	Accuracy rate, %
Tang et al [33]	2022	Classification of atrophic macular lesions in myopic	CNNs[a]; DL[b]	Fundus photographs	5-fold cross-validation + random sampling	94
Zhu et al [34]	2023	Classification of atrophic macular lesions in myopic	Neural network; DL	Fundus photographs	Stratified 20-fold cross-validation	90
Wan et al [4]	2021	Normal, low, and high risk of high myopia	DCNNs[c]; DL	Fundus photographs	5-fold cross-validation + random sampling	99
Wan et al [38]	2023	Classification of atrophic macular lesions in myopic	DL	Fundus photographs	Random sampling	95-97
Sun et al [39]	2023	Classification of atrophic macular lesions in myopic	DL	Fundus photographs	External validation (multicenter)	89.2
Li et al [52]	2022	Differential diagnosis of normal, leopard print fundus, and pathological myopia	DCNN; DL	Fundus photographs	Internal validation (random sampling) + external validation (multicenter)	94
He et al [61]	2022	Differential diagnosis of tractive macular degeneration and neovascular macular degeneration in high myopia, and others	DL	OCT[d] images	Random sampling	91-96
Huang et al [49]	2023	Classification of tractive macular degeneration in high myopia	DL	OCT images	Internal validation (random sampling) + external validation (prospective)	96
Du et al [50]	2021	Classification of atrophic macular lesions in myopic	DL	Fundus photographs	Random sampling	88

[a]CNN: convolutional neural network.

[b]DL: deep learning.

[c]DCNN: deep convolutional neural network.

[d]OCT: Optical Coherence Tomography.

Discussion

Summary of the Main Findings

This study comprehensively described the accuracy of ML in detecting high myopia, high myopia-associated glaucoma, and pathologic myopia. ML demonstrated exceptionally favorable performance in detecting high myopia, while DL was highly accurate in diagnosing pathologic myopia.

Comparison With Previous Reviews

Previous studies have also explored the detection accuracy of ML in this field. A systematic review has reported that fundus image– or OCT image–based DL can effectively diagnose and classify myopic maculopathy. Additionally, ML examination of the optic disc area can detect myopic maculopathy that may not be easily identified during clinical examination [14]. A recent meta-analysis based on only 11 studies evaluated the performance of DL in identifying pathological myopia based on fundus images. The SROC, specificity, and sensitivity were found to be 0.9905, 0.959 (95% CI 0.955-0.962), and 0.965 (95% CI 0.963-0.966), respectively [62]. In the previous meta-analysis, the 11 original studies all constructed fundus images-based DL models, and studies on conventional ML (non-DL) were not incorporated. The number of studies included in our review was further expanded, with a total of 20 studies on the performance of ML in diagnosing pathological myopia. Moreover, subgroup analysis was executed between conventional ML (non-DL) and DL. Our finding also indicated that DL demonstrated exceptionally favorable efficiency in detecting pathological myopia.

As the understanding of the etiology of myopia deepens, growing evidence reveals risk factors for the onset or progression of myopia, including age, sex, parental myopia, susceptibility genes, and outdoor activities. For high myopia, early prediction appears to be more beneficial. Among the included studies, one incorporated 135 myopia-related single nucleotide polymorphisms to forecast the progression and onset of high myopia. ML for the prediction of high myopia was

XSL•FO
RenderX

mainly based on genetic factors, environmental factors, and ocular clinical characteristics. ML showed an SROC of 0.96, sensitivity of 0.91, and specificity of 0.87, respectively [20], suggesting that ML methods can effectively identify high-risk individuals with high myopia, thus effectively preventing this condition, especially in minors.

Glaucoma is a significant contributor to irreversible vision impairment and blindness all over the world. A 10-year study in Chinese individuals over the age of 40 years found that every 1 mm increase in axial length increased the risk of open-angle glaucoma by 1.72 times. In comparison to emmetropic and hyperopic eyes, highly myopic eyes had a 7.3 times higher risk of developing open-angle glaucoma [63]. Due to the changes in retinal structure caused by myopia, diagnosing glaucoma in myopic patients, especially those with high myopia, is challenging. Six studies were included to evaluate the diagnosis of high myopia glaucoma. Of them, 3 studies [28,29,32] used fundus OCT image-based DL techniques, while the remaining 3 [27,30,31] used non-DL ML (Lagrange multiplier, fully connected network, radial basis function network, decision tree, RF) approaches using OCT parameters, Heidelberg Retina Tomograph parameters, and ocular biometric parameters of patients. The findings indicated that ML yielded highly promising results in the detection of high myopia glaucoma.

It was also noted that different ML methods, conventional ML and DL, showed significant differences in their ability to identify positive or outcome events. Conventional ML is often used to construct models with interpretable clinical features. Lately, various image-based ML methods have emerged. However, a significant challenge in this context is the requirement for manual annotation to facilitate ML. From this standpoint, manual annotation poses a formidable barrier to effectively mitigating the risk of bias. DL, on the other hand, enables intelligent processing of medical images and has been widely applied in various fields, including detecting diabetic retinopathy [8-10], retinopathy of prematurity [64,65], age-related macular degeneration [10], and glaucoma [11-13]. With the rapid development of ML, imaging data are increasingly becoming a valuable source for medical analysis. Multiple studies have demonstrated that images from various sources, including fundus images [66], external eye appearance [67], and refractive images [68], can effectively estimate a patient's spherical refractive error, indicating the potential of imaging data in predicting the risk of myopia. This study also finds that image-based DL is more accurate than conventional ML, providing a theoretical basis for the creation of future intelligent tools.

Additionally, the dataset used in ML demands considerable attention. Many studies are hampered by a limited number of cases, raising concerns about the robustness of the findings. Additionally, validation methods often depend heavily on internal validation, which may not fully capture the model's generalizability. Hence, incorporating comprehensive patient data is essential for building a robust large-scale database, which will enable the development of ML models that are applicable to a broader population. Among the studies included, 7 [15,22,39,42,44,55,59] established ML models based on publicly available large databases.

Limitations

Although our review includes a larger number of studies than previous meta-analyses and provides an evidence-based basis for subsequent studies, this study has limitations. First, there were few studies on the prediction of high myopia, which limits the interpretation of our results, and clinically interpretable variables for predicting high myopia were not explained. Second, we did not conduct a subgroup analysis on the type of ML (conventional ML vs DL) owing to the insufficient number of included studies based on high myopia glaucoma and high myopia. Third, the majority of the models included in this study were assessed as having a high risk of bias, which may impact the interpretation of our results. Most included studies adopted a retrospective design, which might lead to selection bias.

Conclusions

In conclusion, this study comprehensively reviews and meta-analyzes the performance of ML in the diagnosis and prediction of high myopia, high myopia-associated glaucoma, and pathological myopia, providing valuable guidance and references for future research. Challenges exist within the emerging field of myopia prediction. With the development of new analytical methods and the accumulation of real medical datasets, future research holds the promise of improving the prediction of myopia onset and progression. This advancement brings us closer to the ultimate goal of identifying high-risk individuals promptly and implementing targeted interventions in clinical practice.

Acknowledgments

This work was supported by the Joint Project on Regional High-Incidence Diseases Research of Guangxi Natural Science Foundation under grant (2024GXNSFAA010322); Science and Technology Plan of Qingxiu District, Nanning City (2020016); Medical and Health Appropriate Technology Development and Promotion Application Project of Guangxi Zhuang Autonomous Region (S2018093); and Self-Funded Research Project of Health Commission of Guangxi Zhuang Autonomous Region (Z20210589).

Data Availability

All data generated or analyzed during this study are included in this published article. Further inquiries can be directed to the corresponding author.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v27i1e57644_app1.docx]

Multimedia Appendix 2
Literature search strategy.
[DOCX File , 17 KB - jmir_v27i1e57644_app2.docx]

Multimedia Appendix 3
Additional figures.
[DOCX File , 7517 KB - jmir_v27i1e57644_app3.docx]

References

1. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology 2016;123(5):1036-1042 [FREE Full text] [doi: 10.1016/j.ophtha.2016.01.006] [Medline: 26875007]
2. Naidoo KS, Fricke TR, Frick KD, Jong M, Naduvilath TJ, Resnikoff S, et al. Potential lost productivity resulting from the global burden of myopia: systematic review, meta-analysis, and modeling. Ophthalmology 2019;126(3):338-346 [FREE Full text] [doi: 10.1016/j.ophtha.2018.10.029] [Medline: 30342076]
3. Flitcroft DI, He M, Jonas JB, Jong M, Naidoo K, Ohno-Matsui K, et al. IMI—defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. Invest Ophthalmol Vis Sci 2019;60(3):M20-M30 [FREE Full text] [doi: 10.1167/iovs.18-25957] [Medline: 30817826]
4. Wan C, Li H, Cao GF, Jiang Q, Yang WH. An artificial intelligent risk classification method of high myopia based on fundus images. J Clin Med 2021;10(19):4488 [FREE Full text] [doi: 10.3390/jcm10194488] [Medline: 34640509]
5. Ikuno Y. Overview of the complications of high myopia. Retina 2017;37(12):2347-2351. [doi: 10.1097/IAE.0000000000001489] [Medline: 28590964]
6. Miller RJH. Artificial intelligence in nuclear cardiology. Cardiol Clin 2023;41(2):151-161. [doi: 10.1016/j.ccl.2023.01.004] [Medline: 37003673]
7. Majumder A, Sen D. Artificial intelligence in cancer diagnostics and therapy: current perspectives. Indian J Cancer 2021;58(4):481-492 [FREE Full text] [doi: 10.4103/ijc.IJC_399_20] [Medline: 34975094]
8. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
9. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. PLoS One 2019;14(6):e0217541 [FREE Full text] [doi: 10.1371/journal.pone.0217541] [Medline: 31170223]
10. González-Gonzalo C, Sánchez-Gutiérrez V, Hernández-Martínez P, Contreras I, Lechanteur YT, Domanian A, et al. Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. Acta Ophthalmol 2020;98(4):368-377 [FREE Full text] [doi: 10.1111/aos.14306] [Medline: 31773912]
11. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. JAMA Ophthalmol 2019;137(12):1353-1360 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.3501] [Medline: 31513266]
12. Hood DC, De Moraes CG. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology 2018;125(8):1207-1208. [doi: 10.1016/j.ophtha.2018.04.020] [Medline: 30032794]
13. Mirzania D, Thompson AC, Muir KW. Applications of deep learning in detection of glaucoma: a systematic review. Eur J Ophthalmol 2021;31(4):1618-1642. [doi: 10.1177/1120672120977346] [Medline: 33274641]
14. Li Y, Foo LL, Wong CW, Li J, Hoang QV, Schmetterer L, et al. Pathologic myopia: advances in imaging and the potential role of artificial intelligence. Br J Ophthalmol 2023;107(5):600-606. [doi: 10.1136/bjophthalmol-2021-320926] [Medline: 35288438]
15. Hemelings R, Elen B, Blaschko MB, Jacob J, Stalmans I, De Boever P. Pathological myopia classification with simultaneous lesion segmentation using deep learning. Comput Methods Programs Biomed 2021;199:105920. [doi: 10.1016/j.cmpb.2020.105920] [Medline: 33412285]
16. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]

XSL•FO
RenderX

17. Hsiao V, Massoud E, Jensen C, Zhang Y, Hanlon BM, Hitchcock M, et al. Diagnostic accuracy of fine-needle biopsy in the detection of thyroid malignancy: a systematic review and meta-analysis. JAMA Surg 2022;157(12):1105-1113 [FREE Full text] [doi: 10.1001/jamasurg.2022.4989] [Medline: 36223097]
18. Foo LL, Lim GYS, Lanca C, Wong CW, Hoang QV, Zhang XJ, et al. Deep learning system to predict the 5-year risk of high myopia using fundus imaging in children. NPJ Digit Med 2023;6(1):10 [FREE Full text] [doi: 10.1038/s41746-023-00752-8] [Medline: 36702878]
19. Wang H, Li L, Wang W, Wang H, Zhuang Y, Lu X, et al. Simulations to assess the performance of multifactor risk scores for predicting myopia prevalence in children and adolescents in China. Front Genet 2022;13:861164 [FREE Full text] [doi: 10.3389/fgene.2022.861164] [Medline: 35480319]
20. Chen Y, Han X, Guo X, Li Y, Lee J, He M. Contribution of genome-wide significant single nucleotide polymorphisms in myopia prediction: findings from a 10-year cohort of Chinese twin children. Ophthalmology 2019;126(12):1607-1614 [FREE Full text] [doi: 10.1016/j.ophtha.2019.06.026] [Medline: 31416661]
21. Huang Z, Xu X, Cheng W. Deep learning-based health management model application in extreme myopia eye vision monitoring and risk prediction. J Oncol 2022;2022:4537021. [doi: 10.1155/2022/4537021] [Medline: 36276280]
22. Tan TE, Anees A, Chen C, Li S, Xu X, Li Z, et al. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. Lancet Digit Health 2021;3(5):e317-e329 [FREE Full text] [doi: 10.1016/S2589-7500(21)00055-8] [Medline: 33890579]
23. Choi KJ, Choi JE, Roh HC, Eun JS, Kim JM, Shin YK, et al. Deep learning models for screening of high myopia using optical coherence tomography. Sci Rep 2021;11(1):21663. [doi: 10.1038/s41598-021-00622-x] [Medline: 34737335]
24. Guan J, Zhu Y, Hu Q, Ma S, Mu J, Li Z, et al. Prevalence patterns and onset prediction of high myopia for children and adolescents in southern China via real-world screening data: retrospective school–based study. J Med Internet Res 2023;25:e39507 [FREE Full text] [doi: 10.2196/39507] [Medline: 36857115]
25. Zhao H, Ma Z, Guan Y, Li J, Wei Y. High myopia detection method on fundus images based on curriculum learning. : IEEE; 2022 Presented at: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC); June 27, 2022; Los Alamitos, CA. [doi: 10.1109/compsac54236.2022.00227]
26. Dai S, Chen L, Lei T, Zhou C, Wen Y. Automatic detection of pathological myopia and high myopia on fundus images. : IEEE; 2020 Presented at: 2020 IEEE International Conference on Multimedia and Expo (ICME); July 6-10, 2020; London, United Kingdom p. 6-10 URL: https://ieeexplore.ieee.org/document/9102787 [doi: 10.1109/icme46284.2020.9102787]
27. Li L, Zhu H, Zhang Z, Zhao L, Xu L, Jonas RA, et al. Neural network–based retinal nerve fiber layer profile compensation for glaucoma diagnosis in myopia: model development and validation. JMIR Med Inform 2021;9(5):e22664 [FREE Full text] [doi: 10.2196/22664] [Medline: 34003137]
28. Lee YJ, Sun S, Kim YK, Jeoung JW, Park KH. Diagnostic ability of macular microvasculature with swept-source OCT angiography for highly myopic glaucoma using deep learning. Sci Rep 2023;13(1):5209 [FREE Full text] [doi: 10.1038/s41598-023-32164-9] [Medline: 36997639]
29. Kim JA, Yoon H, Lee D, Kim M, Choi J, Lee EJ, et al. Development of a deep learning system to detect glaucoma using macular vertical optical coherence tomography scans of myopic eyes. Sci Rep 2023;13(1):8040 [FREE Full text] [doi: 10.1038/s41598-023-34794-5] [Medline: 37198215]
30. Jeong Y, Kim YK, Jeoung JW, Park KH. Comparison of optical coherence tomography structural parameters for diagnosis of glaucoma in high myopia. JAMA Ophthalmol 2023;141(7):631-639 [FREE Full text] [doi: 10.1001/jamaophthalmol.2023.1717] [Medline: 37200038]
31. Asaoka R, Iwase A, Tsutsumi T, Saito H, Otani S, Miyata K, et al. Combining multiple HRT parameters using the 'Random Forests' method improves the diagnostic accuracy of glaucoma in emmetropic and highly myopic eyes. Invest Ophthalmol Vis Sci 2014;55(4):2482-2490. [doi: 10.1167/iovs.14-14009] [Medline: 24609628]
32. Bowd C, Belghith A, Rezapour J, Christopher M, Jonas JB, Hyman L, et al. Multimodal deep learning classifier for primary open angle glaucoma diagnosis using wide-field optic nerve head cube scans in eyes with and without high myopia. J Glaucoma 2023;32(10):841-847. [doi: 10.1097/IJG.0000000000002267] [Medline: 37523623]
33. Tang J, Yuan M, Tian K, Wang Y, Wang D, Yang J, et al. An artificial-intelligence-based automated grading and lesions segmentation system for myopic maculopathy based on color fundus photographs. Transl Vis Sci Technol 2022;11(6):16 [FREE Full text] [doi: 10.1167/tvst.11.6.16] [Medline: 35704327]
34. Zhu SJ, Zhan HD, Wu MN, Zheng B, Liu BQ, Zhang SC, et al. Research on classification method of high myopic maculopathy based on retinal fundus images and optimized ALFA-Mix active learning algorithm. Int J Ophthalmol 2023;16(7):995-1004 [FREE Full text] [doi: 10.18240/ijo.2023.07.01] [Medline: 37465510]
35. Wu Z, Cai W, Xie H, Chen S, Wang Y, Lei B, et al. Predicting optical coherence tomography–derived high myopia grades from fundus photographs using deep learning. Front Med 2022;9:842680. [doi: 10.3389/fmed.2022.842680] [Medline: 35308524]
36. Ye X, Wang J, Chen Y, Lv Z, He S, Mao J, et al. Automatic screening and identifying myopic maculopathy on optical coherence tomography images using deep learning. Transl Vis Sci Technol 2021;10(13):10 [FREE Full text] [doi: 10.1167/tvst.10.13.10] [Medline: 34751744]

XSL•FO
RenderX

37. Wang R, He J, Chen Q, Ye L, Sun D, Yin L, et al. Efficacy of a deep learning system for screening myopic maculopathy based on color fundus photographs. Ophthalmol Ther 2023;12(1):469-484 [FREE Full text] [doi: 10.1007/s40123-022-00621-9] [Medline: 36495394]
38. Wan C, Fang J, Hua X, Chen L, Zhang S, Yang W. Automated detection of myopic maculopathy using five-category models based on vision outlooker for visual recognition. Front Comput Neurosci 2023;17:1169464 [FREE Full text] [doi: 10.3389/fncom.2023.1169464] [Medline: 37152298]
39. Sun Y, Li Y, Zhang F, Zhao H, Liu H, Wang N, et al. A deep network using coarse clinical prior for myopic maculopathy grading. Comput Biol Med 2023;154:106556. [doi: 10.1016/j.compbiomed.2023.106556] [Medline: 36682177]
40. Sogawa T, Tabuchi H, Nagasato D, Masumoto H, Ikuno Y, Ohsugi H, et al. Accuracy of a deep convolutional neural network in the detection of myopic macular diseases using swept-source optical coherence tomography. PLoS One 2020;15(4):e0227240 [FREE Full text] [doi: 10.1371/journal.pone.0227240] [Medline: 32298265]
41. Du R, Xie S, Fang Y, Hagino S, Yamamoto S, Moriyama M, et al. Validation of soft labels in developing deep learning algorithms for detecting lesions of myopic maculopathy from optical coherence tomographic images. Asia Pac J Ophthalmol 2022;11(3):227-236 [FREE Full text] [doi: 10.1097/APO.0000000000000466] [Medline: 34937047]
42. Cui J, Zhang X, Xiong F, Chen CL. Pathological myopia image recognition strategy based on data augmentation and model fusion. J Healthc Eng 2021;2021:5549779 [FREE Full text] [doi: 10.1155/2021/5549779] [Medline: 34035883]
43. Burlui D, Badralexe L, Bobocescu E, Brătucu E, Popescu-Băran D, Dragoncea C. [Intestinal carcinoid]. Rev Chir Oncol Radiol O R L Oftalmol Stomatol Chir 1981;30(6):463-467. [Medline: 6211704]
44. Rauf N, Gilani SO, Waris A. Automatic detection of pathological myopia using machine learning. Sci Rep 2021;11(1):16570 [FREE Full text] [doi: 10.1038/s41598-021-95205-1] [Medline: 34400662]
45. Park SJ, Ko T, Park CK, Kim YC, Choi IY. Deep learning model based on 3D optical coherence tomography images for the automated detection of pathologic myopia. Diagnostics 2022;12(3):742 [FREE Full text] [doi: 10.3390/diagnostics12030742] [Medline: 35328292]
46. Lu L, Zhou E, Yu W, Chen B, Ren P, Lu Q, et al. Development of deep learning-based detecting systems for pathologic myopia using retinal fundus images. Commun Biol 2021;4(1):1225 [FREE Full text] [doi: 10.1038/s42003-021-02758-y] [Medline: 34702997]
47. Lu L, Ren P, Tang X, Yang M, Yuan M, Yu W, et al. AI-model for identifying pathologic myopia based on deep learning algorithms of myopic maculopathy classification and "Plus" lesion detection in fundus images. Front Cell Dev Biol 2021;9:719262 [FREE Full text] [doi: 10.3389/fcell.2021.719262] [Medline: 34722502]
48. Li Y, Feng W, Zhao X, Liu B, Zhang Y, Chi W, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. Br J Ophthalmol 2022;106(5):633-639 [FREE Full text] [doi: 10.1136/bjophthalmol-2020-317825] [Medline: 33355150]
49. Huang X, He S, Wang J, Yang S, Wang Y, Ye X. Lesion detection with fine-grained image categorization for myopic traction maculopathy (MTM) using optical coherence tomography. Med Phys 2023;50(9):5398-5409. [doi: 10.1002/mp.16623] [Medline: 37490302]
50. Du R, Xie S, Fang Y, Igarashi-Yokoi T, Moriyama M, Ogata S, et al. Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images. Ophthalmol Retina 2021;5(12):1235-1244. [doi: 10.1016/j.oret.2021.02.006] [Medline: 33610832]
51. Crincoli E, Servillo A, Catania F, Sacconi R, Mularoni C, Battista M, et al. Artificial intelligence's role in differentiating the origin for subretinal bleeding in pathologic myopia. Retina 2023;43(11):1881-1889. [doi: 10.1097/IAE.0000000000003884] [Medline: 37490781]
52. Li J, Wang L, Gao Y, Liang Q, Chen L, Sun X, et al. Automated detection of myopic maculopathy from color fundus photographs using deep convolutional neural networks. Eye Vis 2022;9(1):13 [FREE Full text] [doi: 10.1186/s40662-022-00285-3] [Medline: 35361278]
53. Liu J, Wong DWK, Zhang Z, Lu S, Li H, Saw SM, et al. Automatic classification of pathological myopia in retinal fundus images using PAMELA. 2010 Presented at: SPIE Medical Imaging 2010; February 13-18, 2010; San Diego, CA URL: https://doi.org/10.1117/12.844122 [doi: 10.1117/12.844122]
54. Liu J, Wong DWK, Lim JH, Tan NM, Zhang Z, Li H, et al. Detection of pathological myopia by PAMELA with texture‐based features through an SVM approach. J Healthc Eng 2010;1(1):1-11 [FREE Full text] [doi: 10.1260/2040-2295.1.1.1]
55. Baid U, Baheti B, Dutande P, Talbar S. Detection of pathological myopia and optic disc segmentation with deep convolutional neural networks. : IEEE; 2019 Presented at: TENCON 2019—2019 IEEE Region 10 Conference (TENCON); October 17-20, 2019; Kochi, India URL: https://ieeexplore.ieee.org/document/8929252 [doi: 10.1109/tencon.2019.8929252]
56. Li S, Li M, Wu J, Li Y, Han J, Cao W, et al. Development and validation of a routine blood parameters–based model for screening the occurrence of retinal detachment in high myopia in the context of PPPM. EPMA J 2023;14(2):219-233 [FREE Full text] [doi: 10.1007/s13167-023-00319-3] [Medline: 37275550]
57. Du Y, Chen Q, Fan Y, Zhu J, He J, Zou H, et al. Automatic identification of myopic maculopathy related imaging features in optic disc region via machine learning methods. J Transl Med 2021;19(1):167 [FREE Full text] [doi: 10.1186/s12967-021-02818-1] [Medline: 33902640]

XSL•FO
RenderX

58. Kim YC, Chang DJ, Park SJ, Choi IY, Gong YS, Kim HA, et al. Machine learning prediction of pathologic myopia using tomographic elevation of the posterior sclera. Sci Rep 2021;11(1):6950 [FREE Full text] [doi: 10.1038/s41598-021-85699-0] [Medline: 33772040]
59. Zhang Z, Xu Y, Liu J, Wong DWK, Kwoh CK, Saw SM, et al. Automatic diagnosis of pathological myopia from heterogeneous biomedical data. PLoS One 2013;8(6):e65736 [FREE Full text] [doi: 10.1371/journal.pone.0065736] [Medline: 23799040]
60. Hou XW, Yang JLX, Li DL, Tao YJ, Ke CF, Zhang B, et al. Machine learning–based integration of metabolomics characterisation predicts progression of ,mopic retinopathy in children and adolescents. Metabolites 2023;13(2):301 [FREE Full text] [doi: 10.3390/metabo13020301] [Medline: 36837920]
61. He X, Ren P, Lu L, Tang X, Wang J, Yang Z, et al. Development of a deep learning algorithm for myopic maculopathy classification based on OCT images using transfer learning. Front Public Health 2022;10:1005700 [FREE Full text] [doi: 10.3389/fpubh.2022.1005700] [Medline: 36211704]
62. Prashar J, Tay N. Performance of artificial intelligence for the detection of pathological myopia from colour fundus images: a systematic review and meta-analysis. Eye 2024;38(2):303-314 [FREE Full text] [doi: 10.1038/s41433-023-02680-z] [Medline: 37550366]
63. Wang YX, Yang H, Wei CC, Xu L, Wei WB, Jonas JB. High myopia as risk factor for the 10-year incidence of open-angle glaucoma in the Beijing eye study. Br J Ophthalmol 2023;107(7):935-940 [FREE Full text] [doi: 10.1136/bjophthalmol-2021-320644] [Medline: 35193858]
64. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 2018;136(7):803-810 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.1934] [Medline: 29801159]
65. Mao J, Luo Y, Liu L, Lao J, Shao Y, Zhang M, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. Acta Ophthalmol 2020;98(3):e339-e345 [FREE Full text] [doi: 10.1111/aos.14264] [Medline: 31559701]
66. Han Y, Li W, Liu M, Wu Z, Zhang F, Liu X, et al. Application of an anomaly detection model to screen for ocular diseases using color retinal fundus images: design and evaluation study. J Med Internet Res 2021;23(7):e27822 [FREE Full text] [doi: 10.2196/27822] [Medline: 34255681]
67. Yang Y, Li R, Lin D, Zhang X, Li W, Wang J, et al. Automatic identification of myopia based on ocular appearance images using deep learning. Ann Transl Med 2020;8(11):705 [FREE Full text] [doi: 10.21037/atm.2019.12.39] [Medline: 32617325]
68. Chun J, Kim Y, Shin KY, Han SH, Oh SY, Chung TY, et al. Deep learning-based prediction of refractive error using photorefraction images captured by a smartphone: model development and validation study. JMIR Med Inform 2020;8(5):e16225 [FREE Full text] [doi: 10.2196/16225] [Medline: 32369035]

Abbreviations

CNN: convolutional neural network
DCNN: deep convolutional neural network
DL: deep learning
DOR: diagnostic odds ratio
LR: logistic regression
MeSH: Medical Subjects Headings
ML: machine learning
NLR: negative likelihood ratio
OCT: optical coherence tomography
PLR: positive likelihood ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF: random forest
SROC: summary receiver operating characteristic
SVM: support vector machine

XSL•FO
RenderX

Edited by A Mavragani; submitted 22.02.24; peer-reviewed by TE Komolafe, S Hansun; comments to author 20.05.24; revised version received 02.07.24; accepted 06.11.24; published 03.01.25.

Please cite as:
Zuo H, Huang B, He J, Fang L, Huang M
Machine Learning Approaches in High Myopia: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e57644
URL: https://www.jmir.org/2025/1/e57644
doi:10.2196/57644
PMID:

©Huiyi Zuo, Baoyu Huang, Jian He, Liying Fang, Minli Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Social Media Potential and Impact on Changing Behaviors and Actions in Skin Health Promotion: Systematic Review

Justyna Martyna Brzozowska[1]; Joanna Gotlib[2]

[1]School of Medical & Health Sciences, University of Economics and Human Sciences in Warsaw, Warsaw, Poland

[2]Department of Education and Research in Health Sciences, Faculty of Health Sciences, Medical University of Warsaw, Warsaw, Poland

Corresponding Author:
Justyna Martyna Brzozowska
School of Medical & Health Sciences
University of Economics and Human Sciences in Warsaw
59 Okopowa Street
Warsaw, 01-043
Poland
Phone: 48 604441820
Email: justyna@kardas.pl

Abstract

Background: Social media is used as a tool for information exchange, entertainment, education, and intervention. Intervention efforts attempt to engage users in skin health.

Objective: This review aimed to collect and summarize research assessing the impact of social media on skin health promotion activities undertaken by social media users.

Methods: In accordance with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines, the following scientific databases were searched: Scopus, Web of Science, PubMed, Academic Search Ultimate (via EBSCO), Academic Research Source eJournals (via EBSCO), ERIC (via EBSCO), Health Source: Consumer Edition (via EBSCO), and Health Source: Nursing/Academic Edition (via EBSCO). Using ProQuest Dissertations and Theses, OpenGrey, Grey Literature Report, and MedNar, the search was supplemented with gray literature. Articles on skin care, skin health, skin diseases, skin protection, and educational activities promoting healthy skin on social media were selected for review (search date: February 6, 2023). The following qualification criteria were used: original research; research conducted on social media; and research topics regarding educational activities in skin health promotion, skin care, skin health, skin diseases, and skin protection. To assess the risk of bias, the following tools were used: the Cochrane Collaboration tool for risk-of-bias assessment (randomized controlled trials and quasi-experimental studies) and the Centre for Evidence-Based Medicine checklist (cross-sectional studies).

Results: Altogether, 1558 works were considered, of which 23 (1.48%) qualified, with 3 (13%) studies on acne and 20 (87%) on skin cancer, sunscreen, and tanning. Social media interventions were dealt with in 65% (15/23) of the studies. The review made it possible to investigate cognitive and cognitive-behavioral interventions. In both observational and interventional studies, the most frequently discussed topics were skin exposure and protection against UV radiation and skin cancer. The analyzed research showed that social media is a source of information. Visualization has a strong impact on users. The involvement of social media users is measured through the amount of content shared and contributes to changing attitudes and behaviors regarding skin health.

Conclusions: This review outlined the impact of social media, despite its heterogeneity, on users' skin health behaviors, attitudes, and actions. It identified strategies for digital interventions to promote skin health. In health sciences, a standardized tool is needed to assess the quality of social media digital interventions. This review has several limitations: only articles written in English were considered; ongoing studies were omitted; and there was a small number of interventional studies on acne and a lack of research on daily skin care, education, and antiaging activities on social media. Another limitation, resulting from the topic being too broad, was a failure to perform quantitative data analysis, resulting in the studies that qualified for the review being heterogeneous.

(J Med Internet Res 2025;27:e54241) doi:10.2196/54241

XSL•FO
RenderX

KEYWORDS

skin; social media; prevention; behavioral intervention; skin cancer; sun protection; acne

Introduction

Background

For many people, social media platforms are now an integral part of daily life. Social media browsing is the most popular online activity. According to the statistics, >4.26 billion people used social media in 2021, and this number is likely to grow to almost 6 billion in 2027 [1]. In April 2023, a total of 4.8 billion people, or 59.9% of the world's population, were social media users [2].

Social media is a powerful means of communication, mainly used for an exchange of information and for entertainment. In addition, companies use it for marketing purposes to promote products and services. Currently, there is also a noticeable trend to develop educational platforms, especially on health education. According to Savas et al [3], because of an increase in the number of patients looking for medical information, physicians can use social media such as Facebook as a significant educational tool. Gantenbein et al [4] argue that medical information available on the internet and social media could be useful for most dermatological patients. The authors identified the most important needs of patients: online consultations, medical content on YouTube, and a possibility to chat with a specialist [4]. In turn, according to Banerjee et al [5], social media is a platform for increasing people's commitment and providing information about UV protection, skin cancer prevention, and its early detection.

The aforementioned research confirms the informational and educational role of social media. The aim of information activities is to increase knowledge about skin cancer and increase awareness of the risks of UV radiation. The effect of these activities should be a change in health behaviors (regular skin photoprotection, regular skin self-examination, and regular monitoring of skin lesions by a physician). Engagement in systematic activities by social media users is crucial.

In addition to its educational role, social media is a tool that supports life and work. Rew et al [6] created a prototype of an application based on content concerning social media users' daily activities. The application can assess the condition of the skin and provides advice on improving it, with suggestions for care products.

In addition, social media can be a motivational tool by motivating systematic care and protection activities aimed at healthy and well-groomed skin.

Focusing mainly on the analysis of social media content, previous reviews have discussed their impact on skin care [7] and skin cancer prevention [8]. They have also discussed the effects of social networking sites on health behavior change [9] and on available digital interventions, mainly stressing the importance of sun protection and skin self-examinations [10].

Unlike previous systematic reviews, we sought to understand the nature of motivation to act. We were interested in the following aspects: (1) What dermatological problem that is talked about on social media is the main area of interest for researchers? (2) What drives social media audiences to take goal-directed actions? (3) What is the level of involvement of social media users in achieving this goal? (4) What contributes to perseverance in pursuing a goal? The goal is healthy and well-groomed skin. To achieve this goal, health behavior changes must be made.

We wanted to draw attention to the role of social media as a tool for collecting data and a tool that encourages users to take actions to care for their skin (including preventing the development of skin cancer and alleviating the symptoms of dermatological diseases).

The main parts of skin health preventive measures are promoting skin care, skin protection, and control routines. The involvement and motivation of social media users is important. Therefore, research is needed to conduct a systematic review that evaluates the impact and effectiveness of social media in this field. In addition, such research should not only focus on the information provided by social media but also try to engage users in actions aimed at skin health.

This review helps understand and explores what is already known about social media potential and its impact on users' skin care behaviors and skin health preventive measures. This study suggests the need to optimize social media interventions and develop new studies that will assess long-term change in skin health behaviors. In addition, by increasing the importance of skin protection and its diseases, this review has practical significance. The results of this research can be helpful in practice and in research investigating the role of social media in preventive health care. This review may support the work of dermatologists, cosmetologists, and health educators because it provides information they need, sometimes on complex topics, in a simple and easily available way.

The novelty of this review is that, through the heterogeneity of the research, a broader view of the impact of social media on skin health behaviors is presented. This study provides new observations about the prespecified desired outcome to be achieved. It complements the theory of engagement with digital interventions.

This study differs from previous ones in that it presents social media as an important digital stimulus. It affirms the integration of motivation into specific physical, affective, and cognitive action. The empirical observations concerning, on the one hand, the demonization of the effects of the sun on the skin and, on the other, the image of an attractive tan, which occur in a cause-effect conflict, are surprising.

This review proves that the image, not just the content, is of great importance. There is a noticeable trend in which the image becomes more important than scientific reports. This review also presents the negative side of social media.

XSL•FO
RenderX

Objectives

This paper presents a systematic review of the literature investigating the role of social media platforms as an educational and intervention tool in promoting preventive skin health measures. The aim of this review was to collect and summarize research assessing the impact of social media on skin health promotion activities undertaken by social media users. In the analyzed studies, we focused on the reception of information about the skin (eg, skin cancer, acne, and tanning) by social media users. The literature review comprised (1) studies assessing the impact of the content of social media on its users, (2) social media intervention studies, and (3) studies determining the impact of posts generated by users on their health decisions and their self-monitoring.

The following research questions were formulated:

1. What are the existing data on the use of social media as an educational intervention tool in skin health preventive measures?
2. What activities on social media can be effective in promoting skin health?
3. What intervention strategies are disseminated via social media and what is their role in promoting skin health?

Methods

Design

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines [11,12] and with the Synthesis Without Meta-Analysis reporting guideline [13]. The PRISMA guidelines ensure the highest quality in scientific research. To more fully present the results of the review, a narrative form was additionally adopted.

Information Sources and Search Strategies

This review was conducted using the following databases: Scopus, Web of Science, PubMed, Academic Search Ultimate (via EBSCO), Academic Research Source eJournals (via EBSCO), ERIC (via EBSCO), Health Source: Consumer Edition (via EBSCO), and Health Source: Nursing/Academic Edition (via EBSCO). In addition, the search was supplemented with gray literature using ProQuest Dissertations and Theses, OpenGrey, Grey Literature Report, and MedNar. A reference list search for relevant articles was also conducted.

The search strategy was first constructed by the authors and then consulted with a librarian experienced in scientific information extraction. The developed phrase was as follows:

> *((“social media” OR “Facebook” OR “Instagram” OR “Twitter” OR “Pinterest” OR “YouTube” OR “TikTok” OR “SnapChat”) AND (“skincare” OR “skin care” OR “skin health” OR “skin diagnosis” OR “skin type” OR “beautician” OR “cosmetologist” OR “cosmetician” OR “cosmetic consultation*” OR “cosmetic dermatology” OR “cosmetic product*” OR “dermocosmetic*” OR “cosmetic procedure*” OR “skin cleans*” OR “skin moistur*” OR “skin massage” OR “facial cleans*” OR “facial moistur*” OR “facial massage” OR “photoprotection” OR “sunscreen*” OR “sun protection” OR “skin hydration” OR “skin rejuvenation” OR “skin cancer” OR “skin prevention” OR “acne” OR “rosacea” OR “anti-acne” OR “anti-aging” OR “anti-wrinkle” OR “skin aging” OR “oily skin” OR “dry skin” OR “combination skin” OR “sensitive skin” OR “pigmented skin” OR “pimple*” OR “pustule*” OR “papules” OR “blackheads”))*

The search was conducted on February 6, 2023, and on the same day, the records were imported from the databases into the Mendeley reference manager (Elsevier). The search was not limited to a period to obtain as many results as possible.

Study Selection

After applying the search strategy, the studies were imported and saved to the Mendeley reference manager, and then duplicates were removed. The selection of the studies consisted of 2 stages. The first stage was conducted by one of the authors (JMB), who considered the title and abstract to identify and exclude irrelevant studies. To reach a mutual agreement, the other author (JG) reviewed a randomly selected sample containing 10% of the titles and abstracts. When, based on the title and abstract, the decision whether to include a work was difficult, the full article was then downloaded. The second stage of the selection consisted of an analysis of the full texts of the articles by each author (JMB and JG) separately to determine whether they could be included, and the reasons for exclusion were listed. When the decision to include or exclude a study was unclear, the authors clarified ambiguities and different opinions through discussions. This stage of the article selection process required both authors to be present to resolve voting conflicts. All studies were assessed using the inclusion and exclusion criteria described in the next section.

Inclusion and Exclusion Criteria

All the inclusion and exclusion criteria were defined a priori (Textbox 1).

In particular, the following papers were excluded during the first stage of the selection process: papers written in a language other than English, papers not related to topics of skin care and skin health promotion on social media, conference papers, information from beauty and health blogs, posters of prevention campaigns, and interviews and articles from popular science magazines advertising a specific cosmetic brand. In addition, articles on social media safety rules, product protection, ethics, diet, COVID-19, and tele-dermatology and materials related to patient consent to post photos on social media were eliminated. During the second stage, articles were allocated to the following groups: articles fully meeting the inclusion criteria, comments, case studies, review papers, research protocols, papers in which skin care and skin health preventive measures were not the main topic, papers in which social media was not the main topic (eg, it was used to recruit study participants or to disseminate survey results), papers on plastic surgery and body image perception on social media, papers on social media marketing (eg, dealing with the impact of advertisements encouraging consumers to purchase cosmetic products or with methods of promoting the sale of cosmetic products or assessing the satisfaction with

XSL•FO
RenderX

cosmetic products and with brand image building of skin care products), papers on the psychological effects of social media (eg, self-acceptance of social media users; self-confidence; and the relationship among emotions, emotional states, and the use of social media), papers on social media search for medical information by dermatology patients, and papers focusing exclusively on the analysis of the content most frequently shared on social media (ie, posts and videos).

Articles with educational value and assessing the impact of social media skin-related information on its users were included. This review also included studies that investigated social media interventions (articles on behavioral interventions aimed at skin health promotion, which were included during the primary analysis) and studies that investigated the impact of social media data on users (articles on skin health promotion, which were included during the secondary analysis).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Papers based on original research (ie, research articles)
- Papers on skin care, skin health, skin diseases, and skin protection and papers on educational activities directed at skin health preventive measures
- Papers on social media (eg, Facebook, Instagram, Twitter [subsequently rebranded X], Pinterest, YouTube, TikTok, and Snapchat)

Exclusion criteria

- Papers written in a language other than English
- Review papers
- Comment papers
- Case studies
- Conference papers
- Clinical trial reports
- Information from beauty and health blogs, popular science magazines, interviews, and posters from prevention campaigns
- Papers in which social media coverage of skin care, skin health, and preventive actions were not the main topic

Data Extraction

Research information from the included articles downloaded during the search process was rechecked by both authors (JMB and JG) to ensure that it met the inclusion criteria. The following details were extracted from each paper: authors, year of publication, geographic location of data collection (country where the research took place), type of social media platform, form of information, number and characteristics of study participants, type of study or data collection methods, and main topic and purpose of the research (Multimedia Appendix 1 [14-36]). In addition, one author (JMB) developed a form with questions (Textbox 2) in which both authors independently recorded their responses (summarized in Multimedia Appendix 2 [14-36]). The PRISMA checklist is provided in Multimedia Appendix 3.

Textbox 2. Questions related to qualified publications.

What are the authors and the year of publication?

- Is intervention a topic of the study?
- Does the study evaluate the impact of social media on users?
- Does the study deal with public health campaigns?
- Does the study mention skin self-examination?
- Does the study mention the use of cosmetic materials (eg, sunscreen)?
- Does the study contain a statistical analysis?
- What are the main results of the study?
- What are the conclusions or opinions of the authors on the effectiveness of information delivery strategies on social media?

Quality Assessment

The methodological quality of each study included in this review was assessed separately by both authors (JMB and JG), and their opinions were compared. In case of disagreement, the authors discussed their doubts and came to a consensus. The Cochrane Collaboration tool for risk-of-bias assessment was used to assess the risk of bias in randomized controlled trials and quasi-experimental studies [37]. A total of 6 domains were evaluated, with the plus sign (+) representing a low risk of bias, the en dash (–) representing a high risk of bias, and both (+ and –) representing an unclear risk of bias. Blinding of participants

XSL•FO
RenderX

and staff was not assessed as this was considered impossible in some studies. A study that received a score of ≥4 plus signs was considered to be of high quality.

For cross-sectional studies, the Centre for Evidence-Based Medicine checklist for critical appraisal was used [38]. It consists of 12 questions with possible answers of "yes," "cannot tell," or "no." A cutoff value of 75% was used for the assessment. This means that, with ≥9 affirmative responses, the study was considered to be of high quality. Otherwise, the study was considered to be of low quality.

Data Synthesis Strategy

A narrative synthesis of the included studies was performed in accordance with the recommendations by Popay et al [39]. During data analysis, the authors considered the possibility of classifying the included publications according to the health problem. However, after reanalysis of the selected publications, the authors decided that a better conceptual model was to group them according to the type of intervention, if it was used.

Results

Search Results

A literature search revealed 1558 records (Scopus: n=485, 31.13%; Web of Science: n=305, 19.58%; PubMed: n=270, 17.33%; Academic Search Ultimate via EBSCO: n=234, 15.02%; Academic Research Source eJournals via EBSCO: n=26, 1.67%; ERIC via EBSCO: n=10, 0.64%; Health Source: Consumer Edition via EBSCO: n=35, 2.25%; Health Source: Nursing/Academic Edition via EBSCO: n=62, 3.98%; ProQuest Dissertations and Theses: n=22, 1.41%; OpenGrey: n=4, 0.26%; Grey Literature Report: n=0; MedNar: n=105, 6.74%). During the identification stage, 91.59% (1427/1558) of the records were obtained from scientific databases, and 8.41% (131/1558) of the records were obtained from gray literature. After removing 43.32% (675/1558) of duplicates, 883 records remained. After reviewing the titles and abstracts of these 883 records, 523 (59.2%) were excluded because they were irrelevant to the review topic or were written in a language other than English. A total of 40.8% (360/883) of the publications were selected for full-text download. After analyzing the full texts of the articles, 339 were excluded (n=43, 12.7% comments; n=21, 6.2% reviews; n=1, 0.3% case reports; n=2, 0.6% protocols; n=14, 4.1% articles not focused on skin care and skin health promotion; n=90, 26.5% articles not focused on social media; n=23, 6.8% articles about plastic surgeries; n=32, 9.4% articles about marketing of skin care products on social media; n=8, 2.4% articles about the psychological effect of social media—emotional states; n=4, 1.2% articles about searching for information on social media; and n=101, 29.8% articles about social media post analysis). These studies did not meet the inclusion criteria. A total of 21 articles met the inclusion criteria. In addition, after a detailed analysis of the eligible studies and their references, 2 publications were added. In total, 23 publications were included in this review (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram.

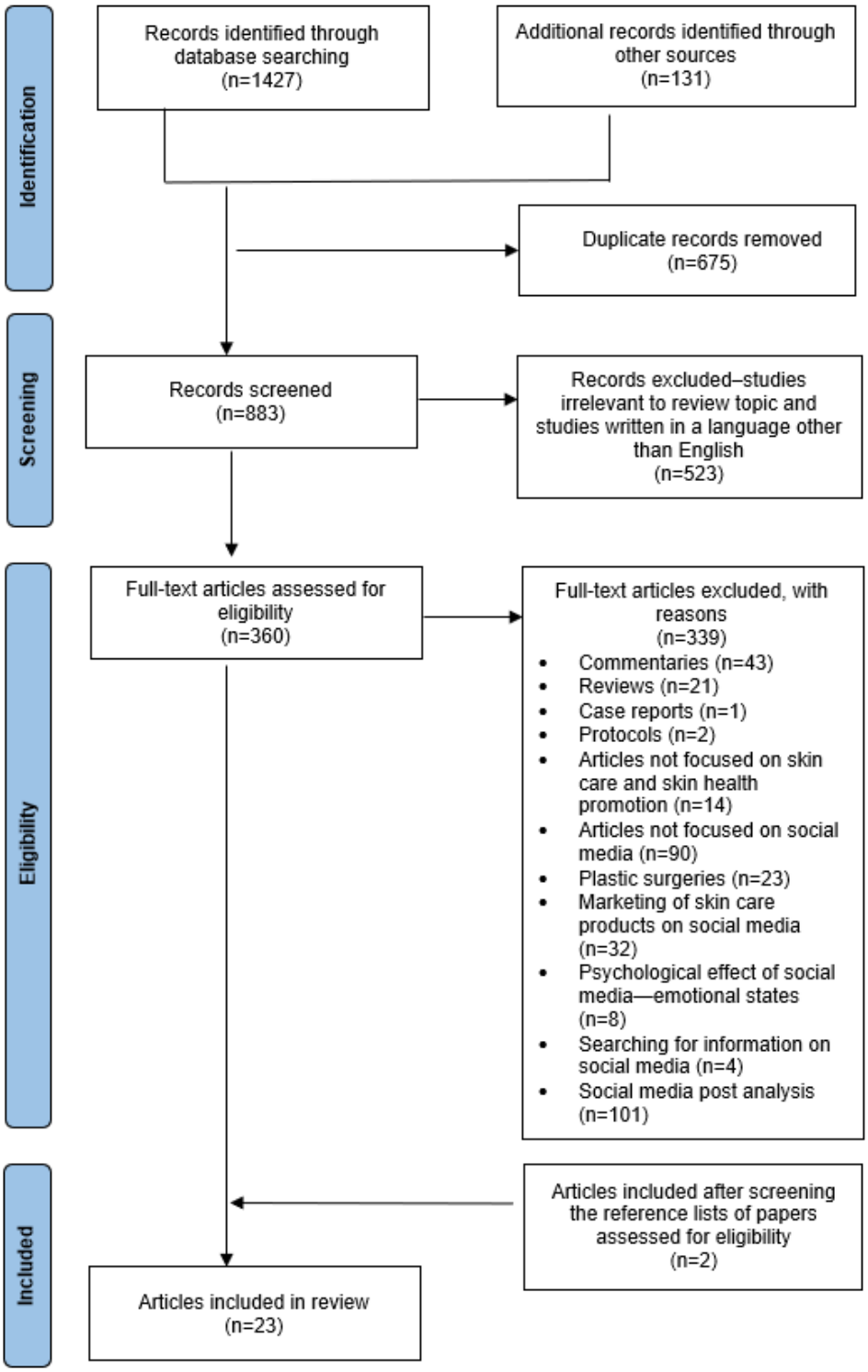

Study Characteristics

The detailed characteristics and main results of the studies included in this systematic review are presented in Multimedia Appendix 1 [14-36] and Multimedia Appendix 2 [14-36]. Most of the research was conducted using Facebook. Of the 23 studies, most (n=14, 61%) were conducted in the United States [14-27], with a smaller number conducted in other countries: the United Kingdom (n=3, 13%) [28-30], Australia (n=3, 13%) [31-33], the Netherlands (n=1, 4%) [34], Denmark (n=1, 4%) [35], and Saudi Arabia (n=1, 4%) [36]. The highest number of studies was published in 2022 (5/23, 22%) [16,17,21,22,36], 17% (4/23) of the studies were published in 2020 [19,27,28,30], 13% (3/23) were published in 2019 [20,26,32], 13% (3/23) were published in 2018 [14,18,24], 13% (3/23) were published in 2017 [29,31,34], 9% (2/23) were published in 2021 [15,25], 9% (2/23) were published in 2011 [33,35], and 4% (1/23) were published in 2016 [23]. The oldest included publications dated back to 2011 (2/23, 9%) [33,35].

The vast majority of the studies were on skin cancer, sunscreen, and tanning (20/23, 87%), with only 13% (3/23) of the studies

XSL•FO
RenderX

on acne [19,27,36]. A total of 17% (4/23) of the studies dealt with skin self-examination [18,19,30,34]. On the other hand, 43% (10/23) of the studies mentioned the use of cosmetic products [14,18,21,22,25,27,28,31,33,36], whereas 35% (8/23) of the studies recommended sunscreen use [14,18,21,22,25,28,31,33]. In addition, 43% (10/23) of the studies discussed the impact of public health campaigns [15-17,20,22,28-30,33,35]. It is worth noting that several publications discussed the results of the same research (eg, 3/23, 13% of the studies by Buller et al [15-17])

Among the 23 studies included in this review, social media interventions were the topic of 15 (65%) [15-18,20-22,24,25,28,29,32-35], among which cognitive effects were assessed by Agha-Mir-Salim et al [28], Damude et al [34], Gough et al [29], and Mingoia et al [32]. On the other hand, both cognitive and behavioral effects were assessed in the studies by Buller et al [15-17], Coups et al [18], Køster et al [35], Morrison et al [20], Myrick et al [21], Pagoto et al [22], Potente et al [33], Stapleton et al [24], and Vraga et al [25]. Comparison groups were used by Agha-Mir-Salim et al [28], Buller et al [15-17], Gough et al [29], Køster et al [35], Mingoia et al [32], Morrison et al [20], Myrick et al [21], Pagoto et al [22], Potente et al [33], and Vraga et al [25].

Quality Appraisal

The risk-of-bias assessment for randomized controlled trials is summarized in Table 1. A total of 35% (8/23) of the studies turned out to be of high quality [15-17,20,22,25,28,32], and 4% (1/23) of the studies were of low quality [29]. A total of 57% (13/23) of cross-sectional studies were of high quality [14,18,19,21,23,24,26,27,30,31,33,35,36], and 4% (1/23) of cross-sectional studies were of low quality [34]. A summary is provided in Multimedia Appendix 2 [14-36].

Table 1. Cochrane risk-of-bias assessment.

Study	Random sequence generation	Allocation concealment	Blinding of participants and personnel or blinding of outcome assessment	Incomplete outcome data	Selecting reporting	Other bias
Agha-Mir-Salim et al [28], 2020	–[a]	+[b]	–	+	+	+
Buller et al [15], 2021	+	+	+	+	+	– (publishing the results of the same study multiple times)
Buller et al [17], 2022	+	+	+	+	+	– (publishing the results of the same study multiple times)
Buller et al [16], 2022	+	+	+	+	+	– (publishing the results of the same study multiple times)
Gough et al [29], 2017	+	–	–	+	+	– (IP addresses were not checked for duplicate users)
Mingoia et al [32], 2019	+	+	–	+	+	+
Morrison et al [20], 2019	+	+	+	+	+	+
Pagoto et al [22], 2022	+	+	+	+	+	+
Vraga et al [25], 2022	+	+	+	+	+	+

[a]High risk of bias.

[b]Low risk of bias.

Data Synthesis

Overview

Among the assumptions we made for this review was that skin health promotion on social media is a broad concept that includes regular activities aimed at protecting the skin against the risk of developing skin cancer, as well as promoting skin care that can ensure good skin condition, prevent skin diseases, slow down the skin aging process, and keep the skin looking nice, which can contribute to one's well-being. However, most of the studies included in this review (20/23, 87%) focused on behaviors associated with skin cancer risk and prevention (exposure to and skin protection from UV radiation). Skin cancer was the main area of interest for researchers, undoubtedly because it is a high-profile problem. There were not enough studies on other skin conditions to draw firm conclusions about the effects of social media. Nevertheless, it is possible to analyze and compare the effect of information and motivational activities on social media on the engagement of users and their behavior changes and actions taken. This will enable social media moderators to develop overall motivational intervention strategies.

This review included observational studies in which social media was used as a tool to disseminate information about the skin, as well as intervention studies in which social media content was manipulated. Despite the fundamental differences in the methodology of these 2 types of research, the main goal was

XSL•FO
RenderX

the same—skin health promotion. Due to differences in research methodology, comparison and summary were difficult.

The results are presented in the following sections in narrative form. A division was made into studies that did not use an intervention and studies that did use an intervention. Then, to synthesize and integrate the results well and draw consistent conclusions, observational studies and then interventional studies were summarized and compared. Attention was drawn to the limitations of observational studies compared to interventional studies.

Assessing the Impact of Social Media on Skin Care and Skin Health Promotion Without an Intervention

The studies included in this review argued that social media is a useful source of knowledge about skin health. According to Bahaj et al [36], most individuals aged between 18 and 25 years choose social media as their first source of acne treatment advice. However, much of social media advice does not follow the current guidelines of the American Academy of Dermatology [40]. The aforementioned authors also found that 74.1% of those social media users had a college degree. There was a positive link between educational attainment and willingness to use social media for advice (P=.002). In addition, a significant link between the severity of acne and search for treatment advice on social media was found (P<.001) [36].

Similarly, in the study by Yousaf et al [27] with 130 participants with acne, 45% (n=58) of them used social media for advice on its treatment. Of these 58 participants, 72% were women (n=42), who were 75% more likely to use social media for advice on acne treatment (prevalence ratio=1.75, 95% CI 1.11-2.76; P=.01). In addition, only 31% of the respondents who consulted their problem on social media managed their condition in line with the American Academy of Dermatology clinical guidelines [27]. However, both Bahaj et al [36] and Yousaf et al [27] observed that the advice available on social media did not comply with the guidelines of the American Academy of Dermatology [40]. In the study by Basch et al [14] on skin cancer and tanning, it was found that people who did not fully adhere to proper sunscreen behavior were more likely than those who followed sun protection rules to believe that social media was an accurate source of health information (10.5% vs 3.3%; P=.046) or a helpful source of health information (62.5% vs 46.7%; P=.02).

According to Mingoia et al [31], the level of social media use was significantly and positively correlated with users' sun exposure (P<.001), sunburn (P<.001), and dissatisfaction with skin tone and negatively correlated with sun protection (P<.001). This meant that viewing photos, posting photos, sharing content, and the number of likes were significantly linked with greater sun exposure, lower sun protection, and greater dissatisfaction with skin tone. It is worth mentioning that the aforementioned study drew attention to the fact that women, to a greater extent than men, are image oriented and more attracted to visualization on social media. For them, the use of images in health messages may be effective in preventive skin health care campaigns [31].

In the study by Stapleton et al [23], 45.8% of participants had a college degree, and 14.8% had a bachelor's degree or higher (a sample of 463 participants). The authors found that higher rates of indoor tanning were associated with Twitter (subsequently rebranded X) and Instagram use among a sample of young adult women. As a result, the authors argued that the aforementioned social media platforms could be a valuable way to provide information about skin cancer prevention among compulsive indoor tanners [23].

In contrast, the sample in the study by Willoughby and Myrick [26] consisted of 502 female college students aged between 18 and 29 years, and it was found that participants who reported more frequent use of social media were more likely to sunbathe outside. Thus, the researchers found that greater use of visual social media platforms such as Instagram, Snapchat, and Pinterest was likely to increase tanning [26].

According to the studies by Bahaj et al [36], Yousaf et al [27], Basch et al [14], Mingoia et al [31], Stapleton et al [23], and Willoughby and Myrick [26], widely shared social media content could influence engagement in risky health behaviors, affecting users' knowledge, attitudes, and actions. Despite the negative impact of social media, the authors noted that it could be a powerful informational, educational, and intervention tool for preventive skin health care campaigns. In addition, the identification of the target group of health advice recipients made it possible to fight disinformation using specific prevention channels (ie, social media platforms). In contrast to the aforementioned studies, the studies by Guckian et al [30] and Martel et al [19] mainly pointed to the positive impact of social media.

Guckian et al [30] found that interventions via social media could have a positive impact on health-related behaviors. According to the authors, social media could motivate patients with distressing cutaneous lesions to visit melanoma screening clinics. However, the authors found that the most common motivating reason for visiting a melanoma screening clinic was skin self-examination. In this study, 65% (162/249) of the patients were regular users of social media, 33% (83/249) saw posts on the internet about cutaneous lesions, 10% (8/83) saw posts from physicians, 36% (30/83) saw posts from health authorities, and 37% (31/83) saw posts from other people (social media users). In addition, 33% (83/249) of patients searched for information about their cutaneous lesions on the internet, whereas one patient did so on social media. In addition, 24% (6/25) of patients said that more posts on cutaneous lesions should be shared [30].

Martel et al [19] found that social media use could be a motivating factor. According to the authors, editing one's own cutaneous lesions on social media, in particular acne and acne scars, contributed to an increase in users' awareness and encouraged them to seek dermatological care. Of 145 people who edited their skin lesions, as many as 128 (88.3%) edited acne and acne scars. This editing increased their awareness of the need for dermatological care (P=.02) [19].

XSL•FO
RenderX

Assessing the Impact of Social Media on Skin Care and Skin Health Promotion After an Intervention

Comparative Studies on the Role of Digital Social Media Interventions and Traditional Media in Health Promotion

The study by Agha-Mir-Salim et al [28] was one of those comparing the effects of digital interventions with those of printed materials. The research involved the SunSafe campaign that aimed to raise awareness of sun exposure risk, melanoma, and the need for sun protection among individuals aged 18 to 29 years. Using flyers and Facebook, the authors found that, after the intervention, the average knowledge score improved in both groups (Facebook=1.82; flyers=3.04; $P<.001$). However, printed flyers turned out to be more effective than Facebook posts (95% CI 0.35-2.09; $P=.006$) [28].

Social Media Interventions as an Educational Strategy

Damude et al [34] assessed the knowledge of patients with melanoma about their illness and their opinion on multiple ways of providing information. YouTube instructional videos presented in the research were designed to stress the need for self-examination of the skin and lymph nodes and demonstrate how to do this properly. It turned out that 63% of patients preferred to receive information in multiple ways, 92% of them preferred verbal instructions from their physician, 62% preferred educational YouTube videos, and 43% preferred to receive instructions via brochures. In addition, the authors concluded that there was an urgent need for educational activities focused on melanoma and for supplementing the methods of education with instructional YouTube videos [34].

In the study on skin cancer prevention by Gough et al [29], a trend toward increasing knowledge about preventive actions in response to social media educational interventions was observed. Those interventions increased awareness that fair-skinned people required the most protection and that skin cancer was the most common form of cancer, with melanoma being its most serious type. In addition, the authors noted a trend of increasing awareness of sun and UV ray exposure, skin cancer risk, and the need for sun protection.

Using an educational strategy and cognitive dissonance, Mingoia et al [32] investigated the effects of a social media intervention on the opinion that *tanned skin is desirable and beneficial for appearance*. The educational strategy was aimed at reducing positive attitudes toward sunbathing and improving the ability to critically analyze social media content [32].

The study by Pagoto et al [22] is another example of a dissonance-based social media intervention to promote sun safety. The intervention involved encouraging survey respondents to create social media posts on healthy skin or a healthy lifestyle. According to the authors, participants in the healthy skin intervention group reported a higher motivation to use protective clothing and sunscreen and a decreased motivation to tan. The authors suggested that this intervention, encouraging social media content creation and response, should enhance cognitive dissonance and affect social norms, which in turn might change attitudes and behaviors concerning tanning [22].

The study by Stapleton et al [24] is another example of a dissonance-based intervention focusing on healthy body image. As part of this intervention, a group of Facebook users was encouraged to publish posts and share their opinions on them. The aim of the dissonance-inducing intervention approach was to engage participants in group discussions related to body image [24].

The topic of skin cancer prevention and the use of tanning beds was also addressed by Morrison et al [20]. To educate the audience, the authors used 3 different short videos aimed at discouraging indoor tanning. The first was a humorous video created in collaboration with a physician, the second was a music video with an Instagram star, and the third one was based on facts about skin cancer. The humorous video received the highest engagement, measured via the number of comments, reactions, and shares, whereas the fact-based video received the lowest engagement [20].

Moreover, in the study by Potente et al [33], an ironic music video was also used, the aim of which was to engage young people in skin cancer prevention. Different forms of entertainment, education, and marketing were used on social media to change attitudes and behaviors regarding sun protection [33].

In contrast to the aforementioned studies, Vraga et al [25] used a video with incorrect information about sunscreen and skin cancer. The results indicated that the disinformation video, compared to an educational video, increased belief in sunscreen myths ($P<.001$) and decreased intent to use sunscreen ($P<.01$) [25].

In the study by Myrick et al [21], an intervention focusing on appearance and self-control–related emotions was used. The results indicated that the intervention focused on appearance contributed to a reduction in the time spent viewing Instagram images of tanned women. In contrast, after the intervention involving self-control emotions, an increase in anticipated pride and in a positive attitude toward solar safety was observed [21].

Digital Interventions on Social Media Based on Social Cognitive Theory, Transportation Theory, and Diffusion of Innovations Theory

Buller et al [15-17] conducted a year-long social media health campaign to reduce mothers' permissiveness of indoor tanning with their teenage daughters. Mothers became less lenient toward indoor tanning with their teenage daughters immediately after the campaign and 6 months after it ended [15,16]. According to the researchers, social media campaigns could motivate mothers to communicate with their daughters and share posts about the danger of indoor tanning [15]. In turn, Buller et al [17] measured engagement levels by measuring reactions (eg, sadness and number of likes) and mothers' comments on posts related to the campaign. They noted that users reacted with comments to 76.4% of posts. Immediately after the campaign ended and 32 months after, mothers who posted reactions or comments were less liberal about indoor tanning with their teenage daughters than mothers who did not engage with the posts [17].

XSL•FO
RenderX

Social Media Behavioral Interventions

In the study by Coups et al [18], a behavioral intervention was conducted via Facebook targeting young patients with melanoma and their families. The content of the intervention consisted of daily posts about skin cancer risk, total cutaneous exam, skin self-examination, and preventive measures (ie, sun protection). Preliminary analyses by the researchers showed an increase in intentions to use the total cutaneous exam, skin self-examination, and sun protection. In addition, Coups et al [18] found that posts that included personal stories, quizzes, skin care exercises, and an opportunity to ask a question to a specialist were perceived by survey respondents as particularly engaging.

An effective intervention was also demonstrated in the study by Køster et al [35], who reported that, in effect, the use of tanning beds decreased in parallel with antisun campaign activities, with an odds ratio of 0.61 (95% CI 0.54-0.69). The campaign was mainly conducted on social media but also in youth magazines and on the radio. It was supplemented with a music video available on the internet and television on skin damage and the negative effects of using cosmetic beds. In addition, respondents to the postcampaign survey were in favor of legislative changes to restrict the access of children and teenagers aged <18 years to tanning salons [35].

Result Summary

To sum up, social media is a place where observational studies and interventional studies can be carried out. Unlike interventional studies, observational studies do not assess cause and effect relationships, require a large sample, and carry a high risk of error. In both observational and interventional studies, the most frequently discussed topics were skin exposure to and protection against UV radiation and skin cancer. Only 13% (3/23) of observational studies focused on acne [19,27,36]. No skin care studies were found that met the inclusion criteria. Observational studies conducted on social media showed that social media is a source of information [14,23,26,27,30,36] and a place where one can edit one's own photos [19,31]. Observational studies on acne showed that dissatisfaction with the appearance of the skin (presence of skin lesions) stimulates individuals to use social media. In the case of observational studies on exposure and protection against UV radiation, it was shown that the use of social media increases dissatisfaction with appearance and, consequently, increases exposure to UV radiation.

Visualization in social media has a strong impact on users. Showing images contributes to taking actions that are both beneficial (in the presence of skin lesions)—seeking dermatological care—and unfavorable (the desire to have a nice tan)—increased exposure to UV radiation (Figure 2).

Figure 2. Summary of observational studies on social media.

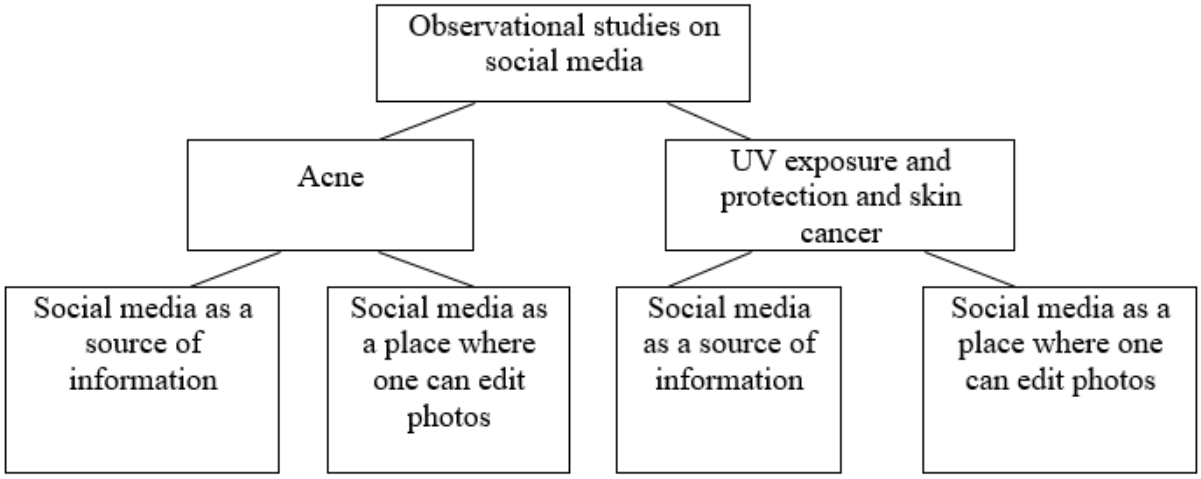

The analyzed social media intervention studies mainly concerned skin exposure to and protection against UV radiation, tanning, skin self-examination, and skin cancer. Most studies (11/23, 48%) used educational interventions on skin cancer prevention [15-18,20,25,28,29,33-35]. One study used an educational intervention on social media literacy and critical content analysis [32]. A total of 13% (3/23) of the studies examined social media users' involvement in generating social media content [15-17,22,24]. These studies pointed out that social media is a motivating factor. Motivation is the readiness to take a specific action, triggered by a need. To take action, one needs the right attitude and the internal conviction that the action is advisable. The aim of educational interventions was to change attitudes toward skin health promotion. However, interventions encouraging participants to create content on social media were intended to engage participants to increase the effectiveness of the intervention and the belief in its purposefulness. A total of 9% (2/23) of the studies focused primarily on body image rather than skin health issues [21,24]. The study by Myrick et al [21] discussed a psychological approach to appearance and emotion self-control (Figure 3). Both in observational and interventional studies, attention was paid to visualization on social media (images, photos, and videos). In addition, attention was paid to the involvement of social media users, which was measured via the amount of content shared, comments, likes, and photos published. Intervention studies showed that educational strategies based on humorous films, interesting personal stories, and opportunities for discussion were more engaging for participants. In observational studies, it was not possible to measure the level of persistence in following recommendations (eg, regular photoprotection). Interventional studies provided this possibility, and their results can be checked after a period.

XSL•FO
RenderX

Figure 3. Summary of interventional studies on social media.

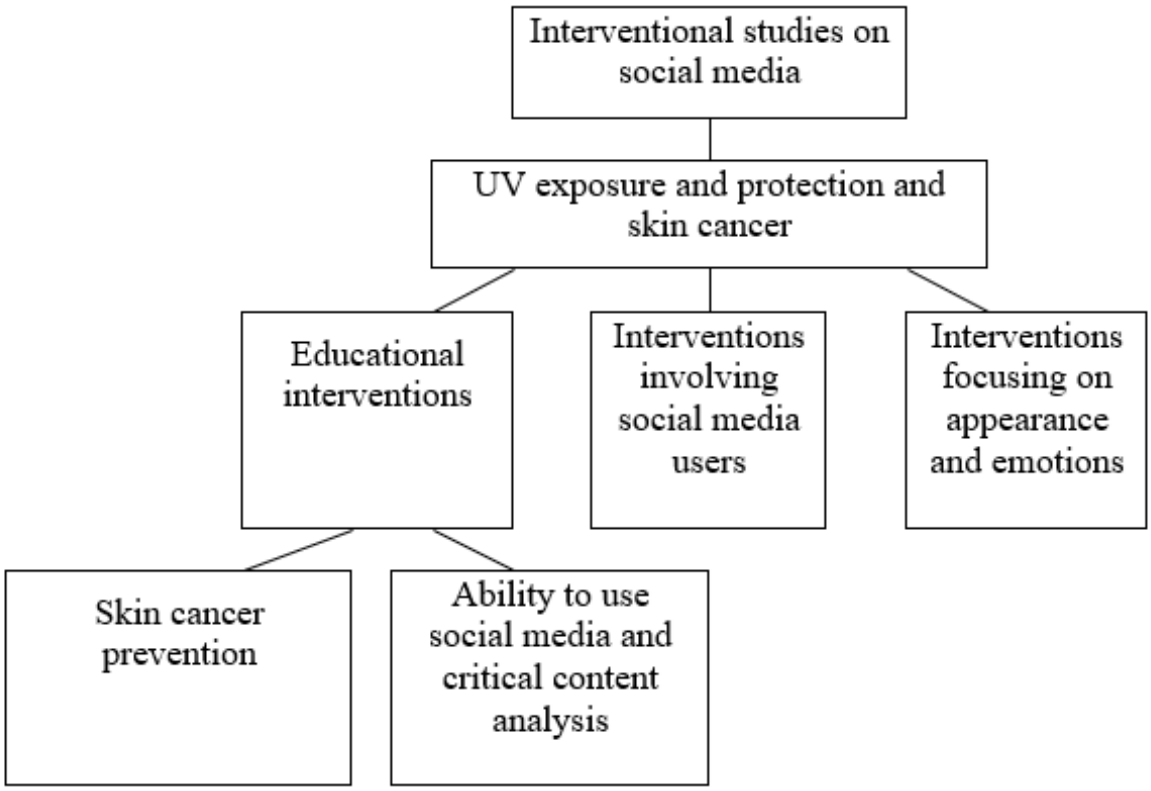

Discussion

Principal Findings

The results of this review indicate that social media is a platform for disseminating health messages on an ongoing basis and for creating interventions targeted at specific groups. However, research on social media impact on education, attitudes, behaviors, and actions concerning skin health promotion was inconclusive and varied. Some studies (2/23, 9%) stressed the negative side of social media, noting that, for example, its acne advice was inconsistent with current guidelines [27,36]. Others (4/23, 17%) drew attention to the fact that social media can promote risky health behaviors concerning sunbathing without sunscreen, which may result in skin cancer [14,23,26,31]. On the other hand, without focusing on digital interventions, some of the studies included in this review (2/23, 9%) dealt with a positive impact of social media on skin care, skin protection, and skin health promotion. They noted that social media increases users' self-awareness and motivates them to take care of their skin [19,30]. According to Benetoli et al [41], social media has positively affected users' interaction with health care professionals. As a result, patients feel more confident and more assertive in their decision-making [41]. Most of the studies discussed in this review (14/23, 61%) argued that social media interventions resulted in positive effects on preventive skin health measures, changing users' behaviors, actions, attitudes, and knowledge levels [15-18,20-22,24,28,29,32-35]. Therefore, it is reasonable to ask how to make social media interventions effective and how to achieve a long-term change in user behaviors.

Agha-Mir-Salim et al [28] investigated the effects of social media digital interventions and found that Facebook was less effective than printed materials in improving awareness of skin cancer and the need for sun protection. However, without information about the level of audience engagement measured through the number of posts and reactions to them, it is difficult to explain the reasons why flyers were more effective than Facebook [28].

User engagement is critical to the success of social media interventions. The assessment of user engagement with social media interventions was presented in 35% (8/23) of the studies, determining the reactions to posts using the number of likes, comments under posts, and shares [17,18,20,22,24,29,32,33]. In addition, the concept of social media engagement plays a key role in health psychology [42]. Systematic reviews evaluating strategies that promote technology-based engagement have shown a potential positive impact of these strategies on health behaviors [43-47]. The success of social media engagement is based on its effect on society. Psychological factors and neural mechanisms are important [48]. Nahum-Shani et al [48] explain how positive engagement develops in response to a digital stimulus facilitating behavior change. By describing the neural basis of decision-making (by explaining its application to digital interventions), the researchers extend the concepts by Samanez-Larkin and Knutson [49] (affect-integration-motivation), creating the concepts of affect-integration-motivation and attention-context-translation of motivation to behavior. These additional elements can be used for targeted engagement of social media users with digital interventions [48].

Behavioral interventions through social media have many advantages, such as a large audience, low costs of disseminating information, and a possibility to create individual content and interact with other users with similar experiences or in a similar

XSL•FO
RenderX

situation. This allows for mutual support, the exchange of information, and an opportunity to interact with health professionals by consulting specialists and obtaining answers [18].

In the observational studies on acne by Bahaj et al [36] and Yousaf et al [27], the authors pointed out that much of the advice available on social media is not consistent with the guidelines of the American Academy of Dermatology. In addition, in the observational study on skin cancer by Basch et al [14], it was shown that people who believed that social media was an accurate source of health information were more likely to fail to follow proper sun protection practices. The aforementioned study showed that social media can increase belief in health myths. The study by Vraga et al [25] showed that videos with erroneous information had a stronger impact on users than educational videos. Social media actions should also be aimed at combating misinformation. The content posted should be in agreement with evidence-based medicine and with the guidelines of the American Academy of Dermatology. In addition, interventions should increase the knowledge of social media users about the dangers of improper skin care and change their positive attitude toward inappropriate activities such as too much tanning. Social media interventions should also contribute to an increase in assertiveness toward inappropriate, harmful content, building self-esteem in users and teaching critical analysis of the content shared. In the case of tanning, interventions should describe alternative options, such as the use of safe moisturizing bronzers.

It seems that effective interventions should not only aim at the health consequences of poor skin care and poor skin protection but also focus on monitoring skin texture. In addition, they should make an effort to shape social norms regarding attractiveness (ie, by questioning the ideal of tanned skin and promoting sun protection). The use of dissonance induction approaches in interventions also seems to be important. The goal of those approaches is to encourage participants to persuade others to engage in appropriate behaviors [22].

Behavioral health interventions concerning appearance and body image often use persuasive techniques based on cognitive dissonance theories. Those techniques encourage participants to engage in discussions and cognitive exercises. Participants endorse views that contradict their previous beliefs and, in the end, accept these views as their own [24]. According to the dissonance theory, when there is a conflict between a belief (eg, a perfect tan is good) and behaviors (eg, persuading others that a perfect tan must be avoided and sunscreen should be used), psychological discomfort (cognitive dissonance) is the result, which motivates an individual to change their original beliefs (ie, the person is motivated to use sunscreen because tanned skin is less important than healthy skin) [22,50]. The effectiveness of cognitive dissonance interventions has been confirmed in studies on eating disorders [51] and perfect tans [52].

Pagoto et al [53] proposed a methodology for adaptation of social media behavioral interventions. It requires that parameters be defined to determine whether the intervention is based on social media completely or partially. The objective of a social network should be defined, too. First, it should be decided whether the main objective of the intervention is to engage users, disseminate information, or both. The type of intervention should be related to the objective. One-way communication is adopted mainly for public health campaigns. In 2-way communication, the content is generated by both the intervention moderators and the audience. Two-way communication allows for discussions and for help with problem-solving and behavioral advice. When designing interventions, it is important to choose the social networking platform and the target population. Another important element when designing an intervention is content conversion, with short, catchy posts, videos, photos, graphics, and links to articles. When the main focus of the intervention is user engagement, a plan should be constructed. It should include the size of the group, the frequency of posting, and the number of engaged participants, and it should specify whether the posts are automated. In the design of interventions, it is also reasonable to create chat groups or use microcounseling, which involves the intervention moderator initiating the discussion. The engagement plan may also include recruiting friends or family members for the intervention. In addition, it is important to train users on how to clarify expectations and encourage them to express their opinions and publish posts [53].

Almost all the intervention studies analyzed in this review (13/15, 87%) were conducted after Pagoto et al [53] developed their methodology for adapting social media interventions. After analyzing the presented interventions, it can be assumed that the authors followed some of the advice by Pagoto et al [53] when designing their interventions. Intervention studies that focus on engaging participants have found greater intervention effectiveness.

In summary, the analysis of the studies included in this review indicates that, when designing social media interventions for skin health, it is necessary to pretest them in a pilot project with comparison groups, including control groups, and measure user engagement on a regular basis. In addition, designing interventions requires focusing on the content that is designed to engage the audience. Absorbing content should not only be educational but also entertaining and discourage abnormal health behaviors. The content should be focused on the benefits according to positive psychology (to increase its attractiveness); it should focus on thinking about the future using goal setting and motivation (to have nice and healthier skin). It is also reasonable to include personal stories, exercises to engage the audience, and various puzzles. When designing interventions, the theories of social psychology they are based on should be considered.

The results of this review may be useful in creating future skin health promotion programs on social media. It is aimed at all those who are ready to broaden their knowledge and effectively engage in activities related to skin care and skin protection.

This study aimed to fill the research gap in assessing the usefulness of social media for skin health promotion. Its main assumption was that it would contribute to the discussion on the importance of the information provided about the skin and

XSL•FO
RenderX

its impact on changing the attitudes and behaviors of recipients of this information.

Strengths and Limitations

As with any other study, this review had its limitations. The first one was that only articles written in English were analyzed, which meant that important publications written in other languages were not included. The second limitation was the omission of studies that were in progress (not completed). One of them was a Facebook intervention targeted at patients with melanoma and their families (randomized controlled trial) as the report on its results was expected to be completed by November 2023 [54]. Therefore, its analysis was not possible, and this review should be updated at a later time. The third limitation is the small number of interventional studies on acne and a lack of research on rosacea and atopic dermatitis, common conditions dealt with in dermatological and cosmetology clinics. In addition, there was a lack of research on daily skin care, education, and antiaging activities on social media. Some papers, such as those dealing with social media posts on esthetic medicine and more or less invasive plastic surgery, were not included in this review. This prevented a broader analysis and assessment of the impact of social media on users' skin care and health activities. The last limitation was the failure to undertake a quantitative data analysis because of too broad a topic, which resulted in the heterogeneity of the studies that qualified for analysis.

Despite its limitations, this paper presents an up-to-date review of social media impact on users' behavior change and the implementation of interventions mainly in the field of skin cancer prevention. In addition, this review identifies research gaps and topics worth focusing on and indicates what should be improved when conducting social media research.

Conclusions and Practical Implications

Social media remote consultations and digital interventions with the use of new technologies in education and research are a relatively new method of skin health promotion. Despite the limited number of studies and their heterogeneity, this review outlined the impact of social media on user behaviors, attitudes, and actions in terms of skin health. It provided insights into social media importance and identified strategies for digital interventions to promote health. In addition, this review underlines the need to conduct research on skin health promotion via social media and on new methods motivating users to take an active role. There is no doubt that more high-quality studies, consistent and with logical conclusions, on social media using standardized interventions are needed.

The observational studies analyzed in this review (8/23, 35%) showed that social media, according to social media users, is a reliable place to obtain information on skin health. Moreover, there was a noticeable tendency of social media influencing the behavior of social media users. On the basis of unpublished own research (by author JMB), of 150 surveyed social media users (people aged ≤25 years), in 2021, a total of 62% (93 people) believed that information from social media regarding skin cancer prevention, skin care and treatment of acne, and antiaging was very important to them. For 44% of respondents (66 people), information on social media influenced their behavior. In a repeated study in 2023, as many as 86% of respondents (129 people) stated that information on skin cancer prevention, skin care and treatment of acne, and antiaging was very important to them. However, 74% of respondents (111 people) stated that information on social media influenced their behavior.

The studies analyzed in this review showed that more frequent use of social media is significantly associated with more frequent exposure to UV radiation. Social media focuses on visualization. Photos and images are a powerful tool to influence users. In intervention studies, there was a noticeable tendency to encourage people to react to posts and create their own. It can be assumed that interventions involving social media users in generating content increase their motivation to take care of their skin. The aim of social media interventions should be education on promoting skin health based on visualization and motivating social media users to take care of their skin (using photoprotection, limiting sunbathing, using protective clothing, conducting skin self-examinations, and having regular skin checkups with a dermatologist).

The availability of social media worldwide provides unlimited possibilities for transmitting and managing information. Analyzing posts and evaluating user engagement by measuring the number of likes, comments, and followers can be a powerful tool for gathering knowledge and studying patients' opinions. After becoming familiar with the expectations of social media users and assessing information resources and their quality and credibility, appropriately targeted actions should be taken to educate, encourage, and motivate users to be active in skin care and skin health promotion in everyday life.

Social media has become a tool supporting the life and work of society. The model of Society 1.0 was based on hunting, Society 2.0 was based on agriculture, Society 3.0 was based on industry, and Society 4.0 is based on information. Social media implements the assumptions of the Society 4.0 model through information activities. The vision of Society 5.0 is a balance between technological progress and human needs. Model 5.0 enables the collection of data (eg, on health status) from patients on an ongoing basis. Social media can be a tool for digital interventions and for Society 5.0 implementation. The implementation of Digital Society 5.0 depends not only on the cooperation between countries but also on combined efforts of various branches of medicine, business, and science, which is the basis for creating innovations at both the national and global level. Innovations are not possible without cooperation; the use of global data (big data); and the coordination of many data systems and knowledge platforms, including social media.

Proper and specialized use of social media by dermatologists, pharmacists, cosmetologists, and health educators can increase awareness concerning skin condition and skin health preventive measures. It can encourage users to conduct a skin self-examination and visit a dermatologist. This review could also enable dermatologists, cosmetologists, pharmacists, and health educators to understand social media potential and its impact on users. This review presents social media use for intervention studies and indicates its educational and

XSL•FO
RenderX

motivational role in actions related to skin health. In addition, those actions have a clinical significance as social media increases user awareness of skin health promotion. As a result, users are not reluctant to seek medical help. Therefore, this review could be a basis for further research on the role of social media interventions in skin health promotion (including sun protection; mole control; and proper skin care for various dermatological diseases, such as acne, rosacea, or atopic dermatitis). It seems that social media, compared to traditional forms of mass communication, has the potential to make a change in skin care and health behaviors and activities among a larger population.

First, future studies should focus on effective social media strategies to promote skin health. Second, research assessing the effects of health and skin care recommendations would be useful. Third, studies assessing the short-term engagement of social media users would be of practical importance, as well as research assessing their commitment to long-term health and skin care habits. The results of this review, in addition to its theoretical value, could have potential applications in online dermatological and cosmetology consultations and in prohealth campaigns concerning increasingly common skin diseases such as acne and melanoma.

This research dealt with studies investigating individually developed social media strategies. It seems that it would still be appropriate to standardize tools assessing the quality of digital health interventions on social media. Such standardization would be helpful both in the design of new digital interventions and in the evaluation of existing ones.

Acknowledgments

The authors would like to thank the Main Library of the University of Siedlce in Poland for access to scientific databases.

Authors' Contributions

JMB contributed to work concept and design, data collection and analysis, visualization, writing the original draft, reviewing and editing the manuscript, critical review, and final approval. JG contributed to data collection and analysis and critical review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the included studies.
[DOCX File , 67 KB - jmir_v27i1e54241_app1.docx]

Multimedia Appendix 2
Findings of the included studies.
[DOCX File , 67 KB - jmir_v27i1e54241_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - jmir_v27i1e54241_app3.docx]

References

1. Number of social media users worldwide from 2017 to 2028. Statista. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2023-05-31]
2. Number of internet and social media users worldwide as of October 2024. Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2023-05-31]
3. Savas JA, Huang KE, Tuchayi SM, Feldman SR. Understanding the influence of social media in medicine: lesson learned from Facebook. Dermatol Online J 2014 Sep 16;20(9):16. [Medline: 25244178]
4. Gantenbein L, Navarini AA, Maul LV, Brandt O, Mueller SM. Internet and social media use in dermatology patients: search behavior and impact on patient-physician relationship. Dermatol Ther 2020 Nov;33(6):e14098. [doi: 10.1111/dth.14098] [Medline: 32725746]
5. Banerjee SC, Rodríguez VM, Greene K, Hay JL. Trending on Pinterest: an examination of pins about skin tanning. Transl Behav Med 2019 Jul 16;9(4):737-748 [FREE Full text] [doi: 10.1093/tbm/iby036] [Medline: 29648613]
6. Rew J, Hwang E, Choi YH, Rho S. Monitoring skin condition using life activities on the SNS user documents. Multimed Tools Appl 2017 Apr 22;77(8):9827-9847. [doi: 10.1007/s11042-017-4623-z]
7. Joshi M, Korrapati NH, Reji F, Hasan A, Kurudamannil RA. The impact of social media on skin care: a narrative review. L viv Clin Bull 2022 Jun 17;1-2(37-38):85-96. [doi: 10.25040/lkv2022.01-02.085]

XSL•FO
RenderX

8. De La Garza H, Maymone MB, Vashi NA. Impact of social media on skin cancer prevention. Int J Environ Res Public Health 2021 May 09;18(9):5002 [FREE Full text] [doi: 10.3390/ijerph18095002] [Medline: 34065061]
9. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
10. Niu Z, Bhurosy T, Heckman CJ. Digital interventions for promoting sun protection and skin self-examination behaviors: a systematic review. Prev Med Rep 2022 Apr;26:101709 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101709] [Medline: 35529530]
11. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
13. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
14. Basch CH, Hillyer GC, Romero RA, MacLean SA, Ethan D. College students' attitudes and behaviors related to sun safety and appearance in relation to health information-seeking behavior and social media use: cross-sectional study. JMIR Dermatol 2018 Dec 19;1(2):e10984. [doi: 10.2196/10984]
15. Buller DB, Pagoto S, Baker K, Walkosz BJ, Hillhouse J, Henry KL, et al. Results of a social media campaign to prevent indoor tanning by teens: a randomized controlled trial. Prev Med Rep 2021 Jun;22:101382 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101382] [Medline: 33996394]
16. Buller DB, Pagoto S, Henry KL, Baker K, Walkosz BJ, Hillhouse J, et al. Persisting effects of a social media campaign to prevent indoor tanning: a randomized trial. Cancer Epidemiol Biomarkers Prev 2022 Apr 01;31(4):885-892 [FREE Full text] [doi: 10.1158/1055-9965.EPI-21-0059] [Medline: 35064063]
17. Buller DB, Pagoto S, Henry KL, Baker K, Walkosz BJ, Hillhouse J, et al. Effects of engagement with a social media campaign for mothers to prevent indoor tanning by teens in a randomized trial. J Health Commun 2022 Jun 03;27(6):394-406 [FREE Full text] [doi: 10.1080/10810730.2022.2113839] [Medline: 35993376]
18. Coups EJ, Manne SL, Pagoto SL, Criswell KR, Goydos JS. Facebook intervention for young-onset melanoma patients and their family members: pilot and feasibility study. JMIR Dermatol 2018 Nov 01;1(2):e3. [doi: 10.2196/derma.9734]
19. Martel J, Powell E, Murina A. The effect of Instagram and photograph editing on seeking dermatologic care. J Cosmet Dermatol 2020 Oct 16;19(10):2732-2735. [doi: 10.1111/jocd.13456] [Medline: 32333461]
20. Morrison L, Chen C, Torres JS, Wehner M, Junn A, Linos E. Facebook advertising for cancer prevention: a pilot study. Br J Dermatol 2019 Oct 04;181(4):858-859 [FREE Full text] [doi: 10.1111/bjd.17993] [Medline: 30972743]
21. Myrick JG, Waldron KA, Cohen O, DiRusso C, Shao R, Cho E, et al. The effects of embedded skin cancer interventions on sun-safety attitudes and attention paid to tan women on Instagram. Front Psychol 2022 Apr 8;13:838297 [FREE Full text] [doi: 10.3389/fpsyg.2022.838297] [Medline: 35465513]
22. Pagoto SL, Waring ME, Groshon LC, Rosen AO, Schroeder MW, Goetz JM. Proof-of-concept feasibility trial of a dissonance-based sun safety intervention for young adult tanners. Ann Behav Med 2022 Aug 02;56(8):830-841 [FREE Full text] [doi: 10.1093/abm/kaab116] [Medline: 35179176]
23. Stapleton JL, Hillhouse J, Coups EJ, Pagoto S. Social media use and indoor tanning among a national sample of young adult nonHispanic White women: a cross-sectional study. J Am Acad Dermatol 2016 Jul;75(1):218-220 [FREE Full text] [doi: 10.1016/j.jaad.2016.01.043] [Medline: 27317521]
24. Stapleton JL, Manne SL, Day AK, Levonyan-Radloff K, Pagoto SL. Healthy body image intervention delivered to young women via Facebook groups: formative study of engagement and acceptability. JMIR Res Protoc 2018 Feb 20;7(2):e54 [FREE Full text] [doi: 10.2196/resprot.9429] [Medline: 29463495]
25. Vraga EK, Bode L, Tully M. The effects of a news literacy video and real-time corrections to video misinformation related to sunscreen and skin cancer. Health Commun 2022 Nov 12;37(13):1622-1630. [doi: 10.1080/10410236.2021.1910165] [Medline: 33840310]
26. Willoughby JF, Myrick JG. Entertainment, social media use and young women's tanning behaviours. Health Educ J 2018 Dec 23;78(3):352-365. [doi: 10.1177/0017896918819643]
27. Yousaf A, Hagen R, Delaney E, Davis S, Zinn Z. The influence of social media on acne treatment: a cross-sectional survey. Pediatr Dermatol 2020 Mar;37(2):301-304 [FREE Full text] [doi: 10.1111/pde.14091] [Medline: 31944359]
28. Agha-Mir-Salim L, Bhattacharyya A, Hart D, Lewandowska M, Spyropoulou E, Stinson L, et al. A randomised controlled trial evaluating the effectiveness of Facebook compared to leaflets in raising awareness of melanoma and harmful sun-related behaviour among young adults. Eur J Cancer Prev 2020 Jan;29(1):89-91. [doi: 10.1097/CEJ.0000000000000519] [Medline: 30998526]

XSL•FO
RenderX

29. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for behavior change: using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
30. Guckian J, Jobling K, Oliphant T, Weatherhead S, Blasdale K. 'I saw it on Facebook!' Assessing the influence of social media on patient presentation to a melanoma screening clinic. Clin Exp Dermatol 2020 Apr 02;45(3):295-301. [doi: 10.1111/ced.14100] [Medline: 31541480]
31. Mingoia J, Hutchinson AD, Gleaves DH, Corsini N, Wilson C. Use of social networking sites and associations with skin tone dissatisfaction, sun exposure, and sun protection in a sample of Australian adolescents. Psychol Health 2017 Dec;32(12):1502-1517. [doi: 10.1080/08870446.2017.1347788] [Medline: 28691513]
32. Mingoia J, Hutchinson AD, Gleaves DH, Wilson C. The impact of a social media literacy intervention on positive attitudes to tanning: a pilot study. Comput Human Behav 2019 Jan;90:188-195. [doi: 10.1016/j.chb.2018.09.004]
33. Potente S, McIver J, Anderson C, Coppa K. “It's a beautiful day for cancer”: an innovative communication strategy to engage youth in skin cancer prevention. Soc Mark Q 2011 Sep 01;17(3):86-105. [doi: 10.1080/15245004.2011.595604]
34. Damude S, Hoekstra-Weebers JE, van Leeuwen BL, Hoekstra HJ. Melanoma patients' disease-specific knowledge, information preference, and appreciation of educational YouTube videos for self-inspection. Eur J Surg Oncol 2017 Aug;43(8):1528-1535. [doi: 10.1016/j.ejso.2017.06.008] [Medline: 28684059]
35. Køster B, Thorgaard C, Philip A, Clemmensen I. Sunbed use and campaign initiatives in the Danish population, 2007-2009: a cross-sectional study. J Eur Acad Dermatol Venereol 2011 Nov 09;25(11):1351-1355. [doi: 10.1111/j.1468-3083.2010.03960.x] [Medline: 21711466]
36. Bahaj RK, Alsaggaf ZH, Abduljabbar MH, Hariri JO. The influence of social media on the treatment of acne in Saudi Arabia. Cureus 2022 Mar;14(3):e23169 [FREE Full text] [doi: 10.7759/cureus.23169] [Medline: 35444887]
37. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
38. Resources. Center for Evidence Based Management. URL: https://cebma.org/resources/ [accessed 2024-09-10]
39. Popay J, Roberts H, AJ S, Petticrew M, Britten N, Arai L, et al. Guidance on the conduct of narrative synthesis in systematic reviews final report. J Epidemiol Community Health 2005;59(Suppl 1):A7 [FREE Full text]
40. Zaenglein AL, Pathy AL, Schlosser BJ, Alikhan A, Baldwin HE, Berson DS, et al. Guidelines of care for the management of acne vulgaris. J Am Acad Dermatol 2016 May;74(5):945-73.e33 [FREE Full text] [doi: 10.1016/j.jaad.2015.12.037] [Medline: 26897386]
41. Benetoli A, Chen TF, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018 Mar;101(3):439-444. [doi: 10.1016/j.pec.2017.08.015] [Medline: 28882545]
42. Nahum-Shani I, Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987 [FREE Full text] [doi: 10.1037/hea0001101] [Medline: 34735165]
43. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
44. De Leon E, Fuentes LW, Cohen JE. Characterizing periodic messaging interventions across health behaviors and media: systematic review. J Med Internet Res 2014 Mar 25;16(3):e93 [FREE Full text] [doi: 10.2196/jmir.2837] [Medline: 24667840]
45. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
46. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
47. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
48. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
49. Samanez-Larkin GR, Knutson B. Decision making in the ageing brain: changes in affective and motivational circuits. Nat Rev Neurosci 2015 May 15;16(5):278-289 [FREE Full text] [doi: 10.1038/nrn3917] [Medline: 25873038]
50. Festinger L. A Theory of Cognitive Dissonance. Redwood City, CA: Stanford University Press; 1957.
51. Stice E, Shaw H, Becker CB, Rohde P. Dissonance-based Interventions for the prevention of eating disorders: using persuasion principles to promote health. Prev Sci 2008 Jun;9(2):114-128 [FREE Full text] [doi: 10.1007/s11121-008-0093-x] [Medline: 18506621]
52. Chait SR, Thompson JK, Jacobsen PB. Preliminary development and evaluation of an appearance-based dissonance induction intervention for reducing UV exposure. Body Image 2015 Jan;12:68-72. [doi: 10.1016/j.bodyim.2014.09.004] [Medline: 25462883]

XSL•FO
RenderX

53. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016 Jan 29;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
54. Manne S, Pagoto S, Peterson S, Heckman C, Kashy D, Berger A, et al. Facebook intervention for young-onset melanoma survivors and families: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Jan 24;12:e39640 [FREE Full text] [doi: 10.2196/39640] [Medline: 36692933]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

Edited by A Mavragani; submitted 02.11.23; peer-reviewed by J Makin, D Buller, A Kinsey; comments to author 20.03.24; revised version received 06.05.24; accepted 09.11.24; published 06.01.25.

Please cite as:
Brzozowska JM, Gotlib J
Social Media Potential and Impact on Changing Behaviors and Actions in Skin Health Promotion: Systematic Review
J Med Internet Res 2025;27:e54241
URL: https://www.jmir.org/2025/1/e54241
doi:10.2196/54241
PMID:

©Justyna Martyna Brzozowska, Joanna Gotlib. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Twins for Clinical and Operational Decision-Making: Scoping Review

Vahid Riahi[1], PhD; Ibrahima Diouf[1], PhD; Sankalp Khanna[2], PhD; Justin Boyle[2], PhD; Hamed Hassanzadeh[2], PhD

[1]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Melbourne, Australia

[2]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia

Corresponding Author:
Vahid Riahi, PhD
Australian e-Health Research Centre
Commonwealth Scientific and Industrial Research Organisation
343 Royal Parade Parkville Vic 3052
Melbourne
Australia
Phone: 61 469842309
Email: vahid.riahi@csiro.au

Abstract

Background: The health care industry must align with new digital technologies to respond to existing and new challenges. Digital twins (DTs) are an emerging technology for digital transformation and applied intelligence that is rapidly attracting attention. DTs are virtual representations of products, systems, or processes that interact bidirectionally in real time with their actual counterparts. Although DTs have diverse applications from personalized care to treatment optimization, misconceptions persist regarding their definition and the extent of their implementation within health systems.

Objective: This study aimed to review DT applications in health care, particularly for clinical decision-making (CDM) and operational decision-making (ODM). It provides a definition and framework for DTs by exploring their unique elements and characteristics. Then, it assesses the current advances and extent of DT applications to support CDM and ODM using the defined DT characteristics.

Methods: We conducted a scoping review following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) protocol. We searched multiple databases, including PubMed, MEDLINE, and Scopus, for original research articles describing DT technologies applied to CDM and ODM in health systems. Papers proposing only ideas or frameworks or describing DT capabilities without experimental data were excluded. We collated several available types of information, for example, DT characteristics, the environment that DTs were tested within, and the main underlying method, and used descriptive statistics to analyze the synthesized data.

Results: Out of 5537 relevant papers, 1.55% (86/5537) met the predefined inclusion criteria, all published after 2017. The majority focused on CDM (75/86, 87%). Mathematical modeling (24/86, 28%) and simulation techniques (17/86, 20%) were the most frequently used methods. Using *International Classification of Diseases, 10th Revision* coding, we identified 3 key areas of DT applications as follows: factors influencing diseases of the circulatory system (14/86, 16%); health status and contact with health services (12/86, 14%); and endocrine, nutritional, and metabolic diseases (10/86, 12%). Only 16 (19%) of 86 studies tested the developed system in a real environment, while the remainder were evaluated in simulated settings. Assessing the studies against defined DT characteristics reveals that the developed systems have yet to materialize the full capabilities of DTs.

Conclusions: This study provides a comprehensive review of DT applications in health care, focusing on CDM and ODM. A key contribution is the development of a framework that defines important elements and characteristics of DTs in the context of related literature. The DT applications studied in this paper reveal encouraging results that allow us to envision that, in the near future, they will play an important role not only in the diagnosis and prevention of diseases but also in other areas, such as efficient clinical trial design, as well as personalized and optimized treatments.

(J Med Internet Res 2025;27:e55015) doi:10.2196/55015

KEYWORDS

digital twin; health care; clinical decision-making; CDM; operational decision-making; ODM; scoping review

XSL•FO
RenderX

Introduction

Background

Digital twins (DTs) are an emerging concept that has recently attracted the attention of researchers and engineers. The development of new technologies, such as virtual reality, blockchain, and the Internet of Things (IoT), is considered a key factor for progress in the DT research field [1]. However, the role of the recent boost in digital transformation and exponential growth in investments by giant tech companies cannot be neglected [2]. DTs have their roots in the field of engineering [3-6] but are quickly expanding to a wide range of applications, such as city planning [7-9], energy [10,11], retail [12,13], and health care [14,15]. Although DTs are an emerging concept, they integrate well-established technologies as underlying components. The origin of the DT goes back to the 1970s when NASA (National Aeronautics and Space Administration) used the "twin" concept by creating simulated environments (ie, mirrored systems) to monitor the spacecraft in the Apollo 13 program [16,17]. Although this instance may not be considered as an advanced DT today due to the lack of data exchange to allow continuous or periodic "twinning" of the digital to the physical [16,18], it represents a very good example of the potential of DT and what DTs can empower in various industries.

A DT of the city of Zurich, Switzerland [7], is a good example of a DT application. In this project, spatial 3D data is transferred to a DT to support urban planning decision-making, for example, simulating climate change and noise issues within urban plans [7]. Another example of a DT was built by Unilever PLC to increase the flexibility and efficiency of their production process [19]. They use DT systems to predict optimal process parameters for new formulations, execute simulations to identify the best operational conditions, and investigate complex what-if scenarios. DT systems have also been applied in the field of water management in Porto, Portugal [20], by creating a virtual representation of the city water network. This DT was developed to predict water quality and flooding issues, simulate and analyze burst pipe scenarios, and ensure the resilience of water infrastructure. The initial results revealed that the developed DT reduced water-supply failures by about 30% and decreased the time needed to repair burst pipes by 8% [20].

There have been several efforts to develop DTs for various purposes in health care. Clinical decision-making (CDM) and operational decision-making (ODM) are 2 of the most important decision-making processes in health care settings. In this context, CDM refers to the decisions made by health professionals in regard to direct health care delivery (diagnosis, test, intervention, etc) based on a patient's health status and related clinical conditions, while ODM refers to the decisions made by managers and administrators after monitoring processes related to operations, such as patient waiting times and hospital revenue [21]. In subsequent sections of this paper, we present a scoping literature review to provide an up-to-date overview of the existing literature relevant to the application of DT systems in health care and to gauge the extent of the designed DT models for CDM and ODM purposes.

There have been several literature reviews on the applications of DTs in health care from different perspectives. Ahmadi-Assalemi et al [22] reviewed recent works in the field of precision health care and discussed the key enabling technologies (eg, IoT and cloud) of DTs. Hassani et al [2] provided a literature review that discussed the value of DTs in health care and proposed some key characteristics, including dynamic, real-time, and bidirectional data connections. Armeni et al [18] identified the opportunities as well as challenges regarding DT implementations in health care (eg, security and privacy, accessibility of the technology, and data collection and management). Elkefi and Asan [23] conducted a systematic review of studies that discussed the contribution of DTs to improving user experience in health care (eg, safety management or information management). They used *digital twin* and *health* as keywords and 17 papers were included in their review. Sun et al [1] provided a systematic review to explore the progress of prominent research on DT technology in medicine. They used the following terms: *digital twin*, *medicine*, *digital health*, and *virtual healthcare*. This resulted in the inclusion of 22 papers in their study after the screening process. They showed that the application of DTs to the cardiovascular system is an attractive area. Sheng et al [24] used structural topic modelling to analyze trends in *DT+healthcare* and their findings show that technology integration and practical applications, for example, IoT and artificial intelligence, are the main focus areas. Katsoulakis et al [25] conducted a scoping review on DTs for health. Using *digital twin* and *health* as search keywords, the authors selected 85 papers in their analysis. They first categorized the selected papers into 8 distinct categories based on their purpose and content, for example, DTs for biomarker and drug discovery as well as DTs in biomanufacturing. Then, they discussed the challenges hindering developments in this field, such as data privacy and security, computing infrastructure, and data quality and accuracy.

This Paper

In this paper, we present a distinctive approach that distinguishes our work from existing literature reviews. The key distinctions are discussed herein. First, the health care industry and academia use various definitions, necessitating research consolidation to establish a unified understanding and ensure future research builds on robust foundations. This study conducts a systematic literature review and thematic analysis of 86 publications on DTs and offers a comprehensive analysis of elements and characterization of the concept, which is absent in existing review papers. This explicit definition of DTs, along with the identification of their elements and characteristics, enables researchers to understand what distinguishes a DT as a unique system, separate from other known ones. Second, DTs represent a burgeoning research field, particularly in health care, prone to misconceptions and misinterpretations. A significant gap in the literature lies in distinguishing established models, such as machine learning and 2D or 3D modeling from true DTs. This paper pioneers an effort to clarify that not all papers claiming to develop DTs genuinely do so. Our analysis reveals that some are simply predictive models or simulations, branded under the DTs trend. This trend risks undermining the integrity of the research. Thus, this paper aims to educate readers on discerning

XSL•FO
RenderX

authentic DTs, emphasizing the essential components required for a system to qualify as such.

What Is a DT?

Overview

There are similarities between DTs and other related technologies, such as simulation, predictive modeling, and 3D modeling, which leads to apparent misconceptions about their definition [2,26]. A summary of DT definitions that are found in the literature is provided in Table 1. It shows the diverse views of DTs across studies and domains. Table 1 also shows the key points of DTs as defined by the respective authors to facilitate the categorization of such diverse definitions and to quickly identify their similarities or differences.

Table 1. Definitions of digital twins (DTs) and the key points in the literature.

Definition of DT	Key points	References
A DT of a human using sensors and other data sources to create an exact virtual copy of a human being	Virtual copy	[18,27-29]
A DT is a digital replica of an object, process, or system that can be used for various purposes	Digital replica	[30-34]
A DT is an instance of the model that mimics a real patient	Mimicking or mirroring	[35-38]
A DT of a patient is a simulation of the patient's trajectory that behaves identically to the patient in terms of outcomes	Simulation	[25,39-41]
A DT is deemed a virtual mapping of a physical entity where simulation clones the dynamic properties of physical entities in the DT	Virtual mapping	[42]
A DT is a digital representation of a physical asset reproducing its data model, behavior, and communication with other physical assets	Digital representation	[43-46]
In newer computer tomographic systems, avatars are fitted as DTs into the surface information of the positioned patients using machine learning	Avatars	[47,48]
DTs are a technology where a personalized computer model is developed that is capable of describing the physiological processes of a human body, tailored to the specific characteristics of a person	Computer model	[49]
The DT is a virtual and dynamic model in the virtual space that is completely uniform and consistent with its corresponding physical entity in the real space	Virtual model	[50,51]
The concept of DTs has emerged to enable modeling and the fusion of individual physical artifacts with digital models reflecting their status in real time	Digital model	[22,52]
DTs are virtual models that represent physical systems and their modifications in real time and there is a bidirectional flow of data, from the physical system to the virtual system and vice versa	Virtual model, bidirectional data flow	[53-56]

From Table 1, we can see that DTs are referred to as a virtual copy, digital replica, mirror, simulation, virtual map, digital representation, image, computer model, and digital and virtual models of a physical entity that are used in different ways, for example, simulating the patient trajectory, personalized treatments, monitoring real-time status, and mirroring the underlying biological systems. As can be observed, the definitions focus on different features of DTs, dependent on the application and purpose of the studies. As this domain could benefit from a universal definition and common understanding of DTs, we first discuss the known elements and underlying characteristics of DTs and then synthesize these into a definition and framework.

DT Architecture

In this section, we will first discuss the elements of DTs and then the characteristics of DTs that distinguish them from other similar technologies.

Real Entity

The aim of a DT is to create a virtual representation of physical entities which is a foundational definition of DTs [57]. In the context of manufacturing, various types of physical entities (eg, vehicles, products, components, etc) and their "real-world" existence have been studied and they are, needless to say, physical. This is a common element in the application of DTs in the field of manufacturing and health care [58]. In the health care context, the patient or hospital environment is typically considered the physical entity [2,15,18]. In ODM and CDM, operational and clinical systems, processes and workflows could also be a target of DTs, which may not be necessarily a physically existing entity. Therefore, real entity (ie, real-world entity) rather than physical entity could be a more generalized terminology.

Virtual Representation

A DT fully describes its real entity from the microatomic level to the macrogeometric level [59,60]. In general, a virtual representation may represent a real entity using 1 or more of the following 4 different dimensions: geometry (ie, the geometric shape, such as size), physics (ie, the physical characteristics and constraints), behavior (the dynamic behavior and responsive mechanism to internal and external changes), and rules (using historical data and domain experts to develop logical abilities, such as reasoning, evaluation, and decision-making) [61]. This representation of a real entity allows broader applications beyond static digital visualizations, for example, computer-aided design and 3D models.

Communication Channels

A bidirectional data connection which is one of the main components of DT systems acts as a communication channel

XSL•FO
RenderX

between the real entity and its virtual counterpart. The nature of this data connection between the real entity and its virtual representation may be interpreted differently. One perspective is that the data flow between the real and virtual product is fully integrated in a way that the real product is controlled by its virtual version and a change in one of them alters the status of the other [62]. An alternative view is that the connection from the real to the virtual product is continuous in a way that the status of the real product is continuously transferred to the virtual product, while the connection of the virtual product to the real one is rather a flow of information and processes that may be applied to the actual object [63]. Regardless of the differences, this live connection is considered one of the main differentiators between DTs and traditional modeling exercises [58].

Computational Models

DTs use computational models that leverage gathered data to analyze, understand, monitor, and predict a system's state and behavior. These models (eg, simulations, machine-learning implementations, and business logic) are the core of DTs [64]. Thus, a crucial step when building DTs is to create high-fidelity models that provide recommendations and capture the corresponding real entity's geometry, physical properties, behaviors, and rules. The multimodel aspects allow the consideration of scenarios that exceed the descriptive capabilities of any single model.

Feedback Mechanism

DTs perform through a closed loop between real and virtual entities. This feedback mechanism is a crucial element of a DT, which could be automated, semiautomated, or human-in-the-loop decision-making depending on several factors, for example, the application, the DT's performance, and the trust level between the DT and stakeholders. Although an automated DT is ideal, in a decision support system, especially in the context of health care where human lives are involved, the decision-making process is risk-oriented and requires a high level of frequent monitoring, consultation, and expertise. Thus, similar to other decision support algorithms, DTs may just provide suggestions and recommendations (eg, a better course of treatment) to decision makers rather than issuing orders. In this situation, the decision maker has the authority to accept or reject any offered suggestions and provide feedback to the virtual entity for further clarifications and improvements that enhance accuracy.

Knowledge Base

The DT is typically considered as a system that is data-driven and thrives on data. A lot of data is captured about the real entity in the development of the virtual representation. This collection of data grows as computational models are run, feedback is provided, resulting decisions are applied to the real entity, and the virtual representation is updated to represent the current state of the real entity. All of this data together constitute a knowledge base that can be used to inform future computational models and related feedback and decision-making processes.

User Interface

One of the advantages of DTs is their ability to handle and manage high information loads in a repetitive manner and provide feedback to decision makers. Considering the scale of their deployment, DTs may have various users (ie, stakeholders), and retrieving and displaying the necessary information for the right user may be challenging [65]. Therefore, a user interface (eg, dashboard, mobile app, etc) allows DTs to represent information in an insightful basis for decision-making and provide information to all users from varied perspectives. The dashboard may be viewed as a control room that provides users (ie, decision makers) a level of interaction to monitor the real-time state of the entities under investigation and view the results of desired experiments. In other words, users can interact with the DT through the user interface.

Considering the already explained elements, DTs have unique characteristics, as presented in Textbox 1.

Textbox 1. Unique characteristics of digital twins (DTs).

- Ongoing updates of virtual representations: a continuous data connection normally is the case when Internet of Things (eg, body sensors) are the source of data collection. However, health data is normally collected from multiple sources, such as electronic health records, laboratory results, and medical images and there is a delay related to capturing those data. As the nature of this connection is to enable a DT to reflect any changes in the state of the real entity in near real time, depending on the application, the frequency of data updates in a DT may vary from seconds to hours or days.
- Close to reality: synchronization is one of the key features of a DT, which is about making sure the virtual representation presented via a DT and the actual counterpart mirror each other as closely as possible. Therefore, building a virtual representation with a high degree of similarity, particularly functional similarity, ensures that when multimodel modeling is carried out on the virtual representation, it responds in the same way we would expect the real entity to respond.
- Ability for feedback to inform decision-making: one of the important aspects of DTs is their ability to analyze and notify findings to a decision maker through an always-active feedback loop. In other words, because DTs have a full knowledge of the real entity's historical performance and an accurate understanding of its future potential, it helps an end user to make an effective decision, for example, by considering multidimensional factors and nonlinear trade-offs that are challenging in reality.
- Multifunctionality: one of the unique aspects of the DTs is their ability to study (eg, simulate and predict) multiple physical properties and the interactions between them. This allows DTs to have many utilizations at the same time, for example, to optimize a process, continuously predict future states (eg, failures), simulate fixes and modifications to identify possible response actions, provide real-time monitoring, support the decision-making process, and provide realistic environments for virtual tests [17].

XSL•FO
RenderX

Our Definition and Framework

A universal and general definition of a DT is a knowledge-driven system that creates a close-to-reality virtual representation (twin) of a real entity (eg, clinical and operational processes). With the help of high-fidelity computational models (eg, simulation, prediction, and 3D modeling) and ongoing updates between virtual and real entities through a feedback mechanism, DTs inform decision makers (eg, administrators and clinicians) via user interfaces, to gain awareness of anticipated problems and gain an improved understanding of potential risks, challenges, and influences. The DT framework is shown in Figure 1.

Figure 1. The proposed universal digital twin (DT) framework.

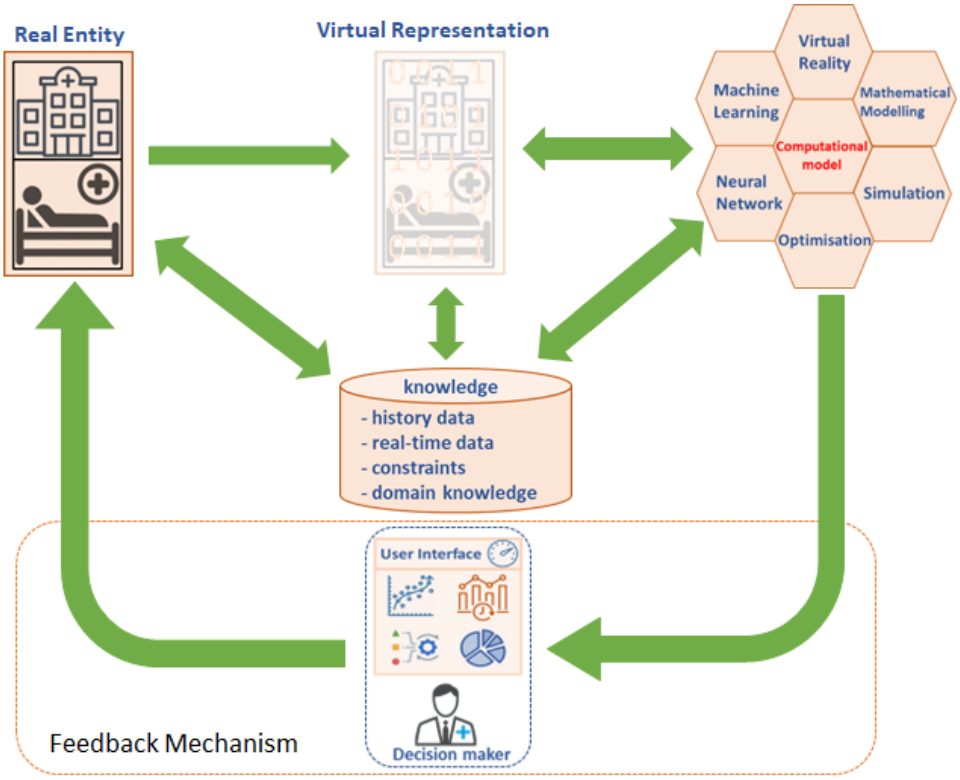

In this framework, 2 environments (ie, real and virtual) are connected through data and a flow of information and processes (eg, dashboard). DTs integrate multiple approaches (eg, simulation, machine learning, and virtual reality) to create a virtual representation of the corresponding real entity. The multifunctionality of the DT provides user-care applications that are shared with the corresponding expert (eg, clinician, nurse, or administrative team) and may be transferred to the actual entity. The remainder of this paper provides a comprehensive review of the current advances and extent of DT applications to support clinical and ODM using this framework and the defined DT characteristics.

Methods

Guideline

We reported the review conducted in this study following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [66] (Multimedia Appendix 1).

Search Strategy

The search was performed through 8 scientific databases, including PubMed, MEDLINE, Scopus, Web of Science, Embase, CINAHL, Cochrane, and gray literature databases in June 2024. As DTs for clinical and ODM within the health care domain is a very new area of research, we included most of the academic databases and gray literature databases (eg, Google Scholar and OpenGrey) with no date restrictions.

To ensure relevant studies were captured in our search process, besides *digital twin**, we provided a list of search terms of which at least 1 needed to be mentioned within the title, abstract, or keywords, such as patient, hospital, Intensive Care Unit or ICU, surgery, clinic, and emergency. The following search terms were used with Boolean operators and wildcards: *digital twin** AND *hospital** OR **patient** OR *health** OR *ICU** OR *ward** OR *emergence** OR *surger** OR *ambulance** OR *clinic** OR *general practi** OR *doctor** OR *nurs**. The search strings used for all databases are shown in Multimedia Appendix 2.

Eligibility Criteria

To identify research papers aligned with our aims, publications had to meet the following criteria: (1) they had to describe the development of a DT model for health care, (2) they had to focus on CDM or ODM in health care (excluding papers focused on medical devices), (3) they had to propose a DT model with experiments, (4) any publication date was considered, (5) only peer-reviewed original research papers were included, and (6) only articles written in English were considered. Papers proposing ideas and frameworks or describing the capabilities of DTs without applications were excluded. Papers published as literature reviews, editorials, and conference posters were

XSL•FO
RenderX

excluded; duplicated papers and conference papers later published as extended journal papers were also excluded.

Data Extraction and Data Analysis

A web-based tool for systematic or scoping reviews, called Rayyan [67], was used to conduct all stages of the screening process. We developed a standardized data extraction form in Microsoft Excel to tabulate specific information. From the included studies (ie, after performing title, abstract, and full-text screening against exclusion and inclusion criteria), the following data were extracted: the country of the first author, publication type, DT characteristics, the environment that DTs were tested within, the sample size, the service, the target area, the main underlying method used in the DTs, the type of data collected to build DTs, the *International Classification of Diseases, 10th Revision* (*ICD-10*) codes, and a summary of included studies. Most included studies are presented in the following section, while the rest are provided in Multimedia Appendix 2.

The *ICD* coding consists of 22 chapters (1 to 22), each covering specific disease areas and courses of treatment. While the papers included in our review did not explicitly provide *ICD-10* codes, we classified them based on the specific disease areas and corresponding courses of treatment mentioned in the studies. For each paper, we first identified the main focus (ie, target area) and then mapped it to the relevant *ICD-10* code, starting from the most specific level and tracing it to the highest chapter level. For example, if a study focused on prostate cancer, it would be assigned the *ICD-10* code C61 (malignant neoplasm of the prostate), which is part of chapter 2, "Neoplasms."

In addition, in our qualitative synthesis, we systematically assessed the characteristics of DT implementations based on the developed framework, focusing on ongoing updates of virtual representations, closeness to reality, and the ability for feedback to inform decision-making. We examined the studies for discussions on the frequency and methods of data updates from real-world entities, considering various sources of data. We noted potential delays in data capture that could affect the real-time functionality of the DT. To evaluate how closely the virtual representation mirrored the actual entity, we examined the validation methods and metrics used to measure functional similarity, ensuring the DT could reliably represent real-world outcomes. We also considered the number and integration of models within each DT, as using multiple model types can enhance accuracy. We examined the mechanisms for ongoing feedback loops within the DTs, focusing on their ability to analyze historical performance and guide future decisions. This included identifying decision-support tools and user interfaces designed to help decision makers act on feedback. We also assessed whether a closed-loop system existed, where the DT actively influences decisions in the real twin. In addition, we explored how feedback is presented to various stakeholders, ensuring that insights from the DT effectively support informed decision-making in clinical and operational contexts. In assessing multifunctionality, we analyzed the range of problems the DTs were designed to address, exploring the distinct functions they served within health care. This involved evaluating the extent to which each DT tackled multiple facets of health care challenges, such as diagnosis, treatment planning, and monitoring. By adopting this structured approach, we ensured a consistent evaluation across studies, capturing the presence of these key characteristics.

Two reviewers collaborated in designing this form to capture the relevant variables effectively. Each reviewer independently extracted data from the included studies, followed by discussions to resolve any discrepancies. In instances where the reviewers could not reach an agreement, a third reviewer was consulted to extract the necessary data for those specific cases.

To synthesize the collated data, we used descriptive statistics to present frequencies and proportions for various characteristics of the included studies. Moreover, we provided a more in-depth qualitative synthesis of findings in the Results section. This qualitative synthesis highlights the nuances of the studies, such as the effectiveness of DT implementations, identified barriers and challenges, and insights into the broader implications for health care practice. The data analysis was performed using both Microsoft Excel and R software.

Results

Selection of Articles

A total of 5537 research papers were retrieved in the first step through all the given search databases. After screening titles and abstracts and removing duplicates, 815 (15%) papers were selected for full-text review. Considering the inclusion criteria, 86 (11%) out of 815 papers were selected for further analysis (ie, CDM: 75/86, 87% and ODM: 11/86, 13%). A total of 729 (89%) papers failed to meet the criteria. Around 34% (246/729) of these papers focused on providing ideas or frameworks and described the potential or capabilities of DTs which was the main reason for their exclusion. Moreover, 49% (355/729) of the papers were not relevant to the scope of this paper, for example, privacy and ethical aspects of DTs in health care, and big data and technology infrastructure required for DT implementations. The review process and results are shown as a PRISMA-ScR diagram in Figure 2.

XSL•FO
RenderX

Figure 2. Flow diagram of the selection process of papers. CDM: clinical decision-making; DT: digital twin; ODM: operational decision-making.

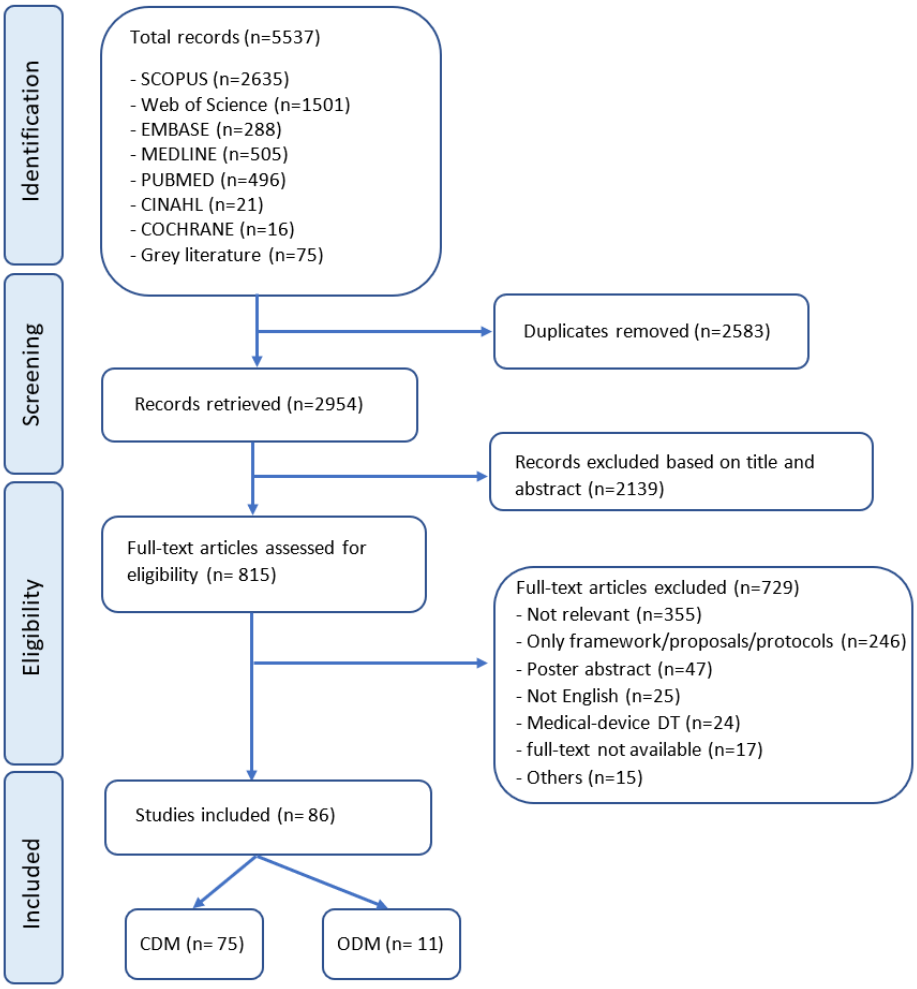

Characteristics of Included Articles

Papers identified through our review reflect the current upward trend of publications in the DT research area as all screened and included studies were published after 2017 as shown in Figure 3. This figure also reveals the difference in research interest among researchers between CDM (more popular) and ODM.

Figure 3. Number of papers on digital twin technology that were screened and included for the review. CDM: clinical decision-making; ODM: operational decision-making.

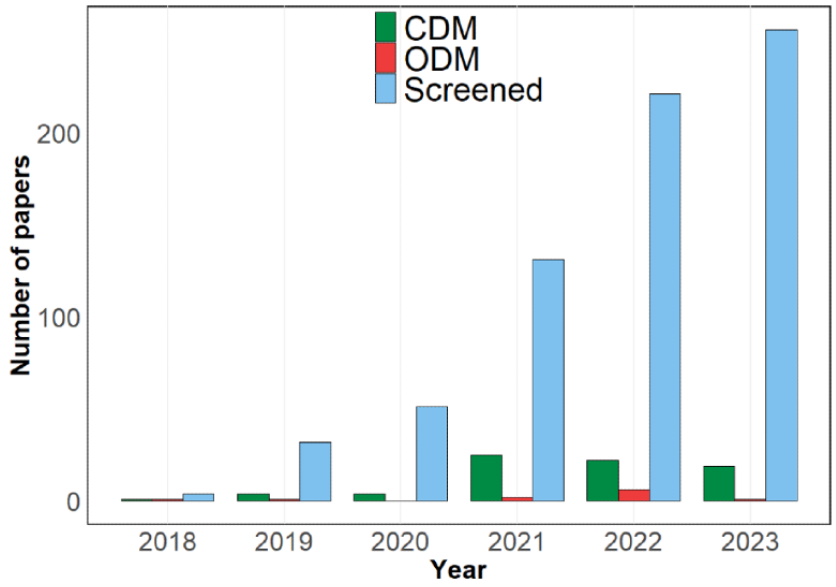

XSL•FO
RenderX

Analysis of the distribution of the papers by the first author's country shows that the United States (17/86, 20%), China (11/86, 13%), and India (10/86, 12%) are the countries with the highest number of published studies. Other countries, including France, Germany, the United Kingdom, South Korea, Austria, Australia, Netherlands, Italy, and Canada contributed to DT research with at least 2 papers each. Out of the 86 papers included, 65 (76%) were published as journal papers and the rest were conference publications.

Data plays an important role within DTs as it allows the physical and virtual entities to communicate together. Besides a small number of studies (5/86, 9%) that used synthetic [40,54,68,69] or randomly generated data [50], most studies used data collected from real entities. Furthermore, analyzing the results shows that the data collected through medical images (eg, magnetic resonance imaging and computed tomography scans) are rich sources to build DT models for various purposes, for example, the His-Purkinje System [31], human vertebra [70], breast cancer [71], tibial plateau fracture [72], the tibiotalar joint [48], knee joint [52], patients with liver tumors [73,74], and the human brain [75].

For further analysis, a summary of the included papers along with their DT and health-related features is provided in Table 2. In this table, we have included characteristics that are featured in the DTs (refer to column "DT characteristics") based on the information reported in the article.

In Table 2, the CDM-based papers were categorized based on the targeted areas of the human body using the *ICD-10* codes. Textbox 2 provides the details of all the chapters in *ICD-10*.

XSL•FO
RenderX

Table 2. The main features of the studied papers (N=86).

Authors	DT[a] characteristics	Environment	Service	*ICD-10*[b] code	Main method
Clinical decision-making					
Gillette et al [31]	CTR[c]	—[d]	Optimizing treatment	Chapter 18	MM[e]
Qi and Cao [76]	CTR	—	Efficient clinical trial	Chapter 2	Simulation
Chakshu and Nithiarasu [45]	CTR	—	Early diagnosis	Chapter 10	NN[f]
Doste et al [40]	CTR	—	Optimizing treatment	Chapter 9	ML[g]
Lai et al [68]	OGU[h], CTR, ATF[i]	Simulated	Personalize treatment	Chapter 21	RL[j]
Tardini et al [77]	OGU, CTR, ATF	Simulated	Optimizing treatment	Chapter 2	RL
Shamanna et al [78]	OGU, CTR, ATF	Deployed	Personalize treatment	Chapter 4	ML
Silfvergren et al [49]	OGU, CTR, MF[k]	Deployed	Early diagnosis	Chapter 21	MM
Shamanna et al [79]	OGU, CTR, ATF	Deployed	Personalize treatment	Chapter 4	ML
Hwang et al [80]	CTR	—	Optimizing treatment	Chapter 11	3D models
Lal et al [81]	CTR	—	Early prevention	Chapter 1	Simulation
Cho et al [82]	CTR	—	Knowledge sharing	Chapter 19	3D models
Shamanna et al [83]	OGU, CTR, ATF	—	Personalize treatment	Chapter 4	ML
Chakshu at al [54]	CTR	—	Early diagnosis	Chapter 9	NN
Ahmadian et al [70]	CTR	—	Early prevention	Chapter 19	NN
Golse et al [84]	OGU, CTR, ATF	Simulated	Early prevention	Chapter 11	MM
Ahmadian et al [85]	CTR, MF	—	Early prevention	Chapter 19	NN
Wu et al [71]	CTR	—	Personalize treatment	Chapter 2	MM
Talukder et al [86]	CTR	—	Accurate prognosis	Chapter 221	ML
Batch et al [87]	CTR	—	Personalize treatment	Chapter 2	NN
Aubert et al [72]	CTR	—	Optimizing treatment	Chapter 19	Simulation
Hernigou et al [48]	CTR	—	Personalize treatment	Chapter 19	NN
Tai et al [88]	OGU, CTR, ATF	Deployed	Optimizing treatment	Chapter 1	NN
Palaniappan and Surendran [55]	CTR	—	Early prevention	Chapter 11	NN
Chen [52]	OGU, CTR, ATF	Deployed	Early prevention	Chapter 19	MM
Baena et al [89]	CTR	—	Early prevention	Chapter 13	Simulation
Barbiero et al [35]	OGU, ATF	Simulated	Knowledge sharing	Chapter 21	NN
Roy et al [90]	OGU	Simulated	Early prevention	Chapter 9	MM
Allen et al [39]	OGU	Simulated	Early prevention	Chapter 9	NN
Hussain et al [91]	—		Early prevention	Chapter 9	ML
Meraghni et al [32]	OGU	Simulated	Early diagnosis	Chapter 2	MM
Scheuermann et al [92]	OGU, CTR, ATF	Deployed	Early prevention	Chapter 21	ML
Semakova et al [93]	—	—	Early prevention	Chapter 9	MM
Ossai and Wickramasinghe [94]	—	—	Early prevention	Chapter 20	BN[l]
Martinez-Velazquez et al [27]	—	—	Early prevention	Chapter 9	NN
Chakshu et al [56]	—	—	Early diagnosis	Chapter 9	MM
Shi et al [73]	OGU, CTR, ATF	Deployed	Optimizing treatment	Chapter 2	MM
Kim et al [95]	—	—	Early diagnosis	Chapter 2	ML
Manocha et al [38]	OGU	Simulated	Real-time monitoring	Chapter 21	NN
Venkatapurapu et al [36]	CTR	—	Optimizing treatment	Chapter 11	Mechanistic model

XSL•FO
RenderX

Authors	DT[a] characteristics	Environment	Service	*ICD-10*[b] code	Main method
Wu et al [69]	OGU, CTR	Simulated	Early diagnosis	Chapter 27	MM
Zhou et al [96]	CTR	—	Personalize treatment	Chapter 21	MM
Lauzeral et al [74]	CTR	—	Personalize treatment	Chapter 1	MM
Wan et al [75]	CTR	—	Accurate prognosis	Chapter 6	ML
Hernigou et al [51]	CTR	—	Knowledge sharing	Chapter 19	3D models
Azzolin et al [97]	OGU, CTR	Simulated	Optimizing treatment	Chapter 9	MM
Ang et al [98]	CTR	—	Optimizing treatment	Chapter 21	MM
van Osta et al [99]	CTR	—	Early diagnosis	Chapter 9	MM
Kardampiki et al [100]	CTR	—	Optimizing treatment	Chapter 9	MM
Jiang et al [101]	OGU, CTR, ATF	Deployed	Real-time monitoring	Chapter 21	Optimization
Kobayashi et al [102]	OGU, CTR, ATF	Deployed	Real-time monitoring	Chapter 5	NLP[m]
Förster et al [103]	CTR	—	Personalize treatment	Chapter 21	Simulation
Fu et al [104]	CTR	—	Personalize treatment	Chapter 28	NN
Gillette et al [105]	CTR	—	Knowledge sharing	Chapter 28	MM
Goodwin et al [106]	OGU, CTR, ATF	Deployed	Optimizing treatment	Chapter 4	Optimization
Yuan et al [107]	—	—	Early diagnosis	Chapter 21	ML
Alcaraz et al [108]	OGU, CTR	Deployed	Optimizing treatment	Chapter 28	MM
Shu et al [109]	OGU, ATF	Deployed	Optimizing treatment	Chapter 19	Simulation
Sarp et al [110]	CTR, MF	—	Real-time monitoring	Chapter 19	NN
Demir et al [111]	CTR	—	Personalize treatment	Chapter 11	MM
Koopsen et al [112]	CTR	—	Optimizing treatment	Chapter 21	MM
Batagov et al [113]	CTR	—	Early diagnosis	Chapter 4	ML
Joshi et al [114]	OGU, CTR, ATF	Deployed	Personalize treatment	Chapter 4	ML
Grieb et al [115]	CTR	—	Early diagnosis	Chapter 2	Graph-based network
Zhang et al [116]	CTR, MF	—	Optimizing treatment	Chapter 4	ML
Rouhollahi [117]	CTR	—	Early prevention	Chapter 17	NN
Serra et al [118]	CTR	—	Early prevention	Chapter 9	Simulation
Chahal [119]	CTR	—	Early diagnosis	Chapter 4	NN
Uyttendaele et al [120]	OGU, CTR, ATF	Deployed	Optimizing treatment	Chapter 28	3D models
Ložek et al [121]	CTR	—	Optimizing treatment	Chapter 9	MM
Thamotharan et al [122]	OGU, CTR, ATF	Deployed	Optimizing treatment	Chapter 4	NN
Cappon et al [123]	CTR	—	Personalize treatment	Chapter 4	Simulation
Salvador et al [124]	CTR, ATF	—	Personalize treatment	Chapter 28	MM
Dubs et al [125]	CTR, ATF	—	Optimizing treatment	Chapter 9	MM
Khan et al [126]	OGU, CTR, ATF	—	Real-time monitoring	Chapter 28	ML
Operational decision-making					
Karakra et al [50]	OGU, ATF	Simulated	Real-time monitoring	—	Simulation
Augusto et al [127]	CTR	—	Facility management	—	Simulation
Pilati et al [53]	OGU, CTR, ATF	Deployed	Efficiency improvement	—	Simulation
Maïzi and Bendavid [128]	OGU	Simulated	Efficiency improvement	—	Simulation
Karakra et al [129]	OGU, ATF	Simulated	Real-time monitoring	—	Simulation
Possik et al [30]	—	—	Efficiency improvement	—	Simulation

XSL•FO
RenderX

Authors	DT[a] characteristics	Environment	Service	*ICD-10*[b] code	Main method
Chen et al [42]	—	—	Knowledge sharing	—	Simulation
Zhong et al [41]	CTR, ATF, MF	—	Medical resource allocation	—	Simulation
Ritzinger et al [37]	OGU, CTR, ATF	Simulated	Medical resource allocation	—	Optimization
Basaglia et al [130]	CTR	—	Facility management	—	Simulation
Zackoff et al [131]	CTR	—	Facility management	—	3D models

[a]DT: digital twin.

[b]ICD: International Classification of Diseases, 10th Revision.

[c]CTR: close to reality.

[d]Not available.

[e]MM: mathematical modeling.

[f]NN: neural network.

[g]ML: machine learning.

[h]OGU: on-going updates.

[i]ATF: ability to feedback.

[j]RL: reinforcement learning.

[k]MF: multifunctional.

[l]BN: Bayesian network.

[m]NLP: natural language programming.

Studies included in this review covered more than half (14) of all 22 chapters. Chapter 19 (“Diseases of the circulatory system”; 14/86, 16%) and chapter 21 (“Factors influencing health status and contact with health services”; 12/86, 14%) include the most studies, followed by chapter 4 (“Endocrine, nutritional, and metabolic diseases”; 10/86, 12%). These findings demonstrate the diversity of health care areas where DT models have been applied.

In terms of health care services in the included studies, for CDM, optimizing treatment (20/86, 23%), early prevention (16/86, 19%), and personalized treatment (15/86, 17%) are the 3 main service areas studied in the included papers, while for ODM, facility management (3/86, 3%) and efficiency improvement (3/86, 3%) are the most popular areas.

Regarding the main computational approach used in the proposed DTs, mathematical modeling (24/86, 28%), for example, moving least square and physiological models is the most common method followed by simulation techniques, for example, discrete event simulation (17/86, 20%) and neural network-based algorithms, for example, deep convolutional neural networks and graph neural networks, (17/86, 20%). Unsurprisingly, simulation-based models, such as discrete event simulation (9/11, 81%) are the most popular method used for ODM (as they can allow decision makers to investigate “what if” scenarios by quantifying the impact of potential changes in the operational settings [132]. The detailed distribution of disease areas, services, and computational methods in the included papers are provided in the Multimedia Appendix 3.

The following was also observed for these included studies: out of 86, only 32 (37%) studies demonstrated ongoing updates of virtual representations in their developed system, although it is the most fundamental characteristic of a DT. Among those 32 studies, 16 (50%) tested the developed system in a simulated environment, and the remainder were tested in real environments. Of the 4 defined characteristics of DTs, multifunctionality is a characteristic that was found to be lacking in most of the studies (81/86, 94%). Among the included papers, all studies missed at least one of the DT’s characteristics, and therefore, did not materialize all 4 characteristics of DT.

Although none of the studies had all the desired DT characteristics, the reported impact of the deployed DTs offers promising insight into their effectiveness. For example, a DT-based program for patients with type 2 diabetes helped all 12 insulin-dependent patients who attended the program to stop insulin injections and helped 38 out of 56 to stop taking metformin [79]. A DT technology used for assisting a surgeon with real-time performance of thermal ablation (a treatment to destroy tumor cells) provided significantly better results compared to other existing methods [73]. A DT for estimating the risk of posthepatectomy liver failure (a leading cause of postoperative death) was developed by modeling blood circulation and estimating 2 important determinants of the disease that could not be predicted before the development of the DT [84].

XSL•FO
RenderX

Textbox 2. Chapters in the International Classification of Diseases, 10th Revision.

- Chapter 1: Certain infectious and parasitic diseases
- Chapter 2: Neoplasms
- Chapter 3: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- Chapter 4: Endocrine, nutritional and metabolic diseases
- Chapter 5: Mental and behavioural disorders
- Chapter 6: Diseases of the nervous system
- Chapter 7: Diseases of the eye and adnexa
- Chapter 8: Diseases of the ear and mastoid process
- Chapter 9: Diseases of the circulatory system
- Chapter 10: Diseases of the respiratory system
- Chapter 1: Diseases of the digestive system
- Chapter 12: Diseases of the skin and subcutaneous tissue
- Chapter 13: Diseases of the musculoskeletal system and connective tissue
- Chapter 14: Diseases of the genitourinary system
- Chapter 15: Pregnancy, childbirth and the puerperium
- Chapter 16: Certain conditions originating in the perinatal period
- Chapter 17: Congenital malformations, deformations and chromosomal abnormalities
- Chapter 18: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
- Chapter 19: Injury, poisoning and certain other consequences of external causes
- Chapter 20: External causes of morbidity and mortality
- Chapter 21: Factors influencing health status and contact with health services
- Chapter 22: Codes for special purposes

Discussion

Principal Findings

The main finding of this study is that while DT systems in health care have shown promising results across multiple areas, they still do not fully incorporate all the desirable characteristics of a comprehensive DT. Nonetheless, DTs are making significant advances in clinical and ODM processes, particularly in areas such as cardiovascular disease, diabetes, cancer, orthopedics, and emergency department operations. The ability to integrate multiple techniques and technologies, such as big data, cloud computing, communication, virtual reality, blockchain, IoT, simulation, prediction, and optimization helps to create a digitally enabled environment for DTs to encapsulate and provide solutions for complex and multidisciplinary problems that were difficult to deal with using traditional methods. Key examples of such solutions include real-time monitoring, treatment optimization, risk factor early intervention, efficient clinical trial design, and improving operational efficiency.

The results of this study showed that the development and deployment of DTs within health care settings still lack maturity. Our review of the literature revealed a lack of a common definition and some inconsistency in the use of the term *digital twin*. Consequently, the focus of this study was shifted toward defining a DT and its underlying elements to support a common understanding and suggest a framework for designing and developing DTs. Building a DT is a complex process in terms of modeling, functionality, domain specificity, and data connectivity process. Therefore, research articles on DTs need to be transparent about their purpose (ie, functionality), the virtual representation that is being created, the data (eg, the frequency of data update, what data elements need to pass from the real entity to the virtual representation and vice versa), the feedback mechanism (eg, what are the outputs, who is receiving that feedback information, etc), and the computational models used (eg, the accuracy of the models and justification for the model's selection). This transparency helps any developed DTs be understandable and their potential for translation be gauged.

Assessing the identified studies against the defined characteristics of DTs revealed that while there are emerging applications of DTs for CDM and ODM, current efforts were not mature enough to include all desired features. As shown in Table 2, most of the reviewed studies did not include ongoing updates of virtual representations in their developed systems. Although they may not materialize a major characteristic of a DT, that is, fed by live data and synchronized with real-life events, they are proof-of-concept studies that provide evidence for the appetite for DTs in health care. One of the main aspects of DTs that is lacking in the included studies is the multifunctionality and scale of the proposed DTs. Studying the input and output data of the developed DTs revealed that they

XSL•FO
RenderX

are mostly designed with a single purpose although DTs are designed to create a complete virtual description of a real entity that is accurate at both the micro and macrolevels. The size of the samples used in the included studies shows that the developed DTs are still in experimental stages, and to fully embrace DT technologies, they need to move from custom expert-driven implementations to accessible robust implementations at scale [133].

As DTs are an emerging research area, especially in health care, it is not surprising to witness misconceptions, misunderstandings, and misrepresentations. It has been discussed that DTs are closely related to other research areas, such as 2D or 3D modeling, system simulation, and digital prototyping and this is a big factor in this confusion [2,26]. Reviewing some of the definitions of DTs over the years provides insight into their growth and development. Considering the fact that researchers are still trying to adopt DT concepts and technologies from manufacturing to health care and figure out how DTs can be implemented in such complex systems, it is understandable that there are still a few steps to go in designing, building, and executing a true DT system within a health care setting. Reviewing the selected papers in this study revealed that DT applications in health care are still in a transformative stage moving from offline systems, for example, digital models and simulations without any real-time connection with the real entity, toward DTs encompassing the aforementioned characteristics, and there is improved awareness of what separates DTs from similar technologies as interest in this field grows.

While none of the reviewed studies possesses all desired DT characteristics, the current state of the deployed models suggests that DTs have the potential for substantial impact once fully developed. They also showcase the versatility of DTs in CDM and ODM applications, from screening, diagnosing, detecting disease, and personal treatment to clinical trial design and optimizing hospital operations. While interest in the applications of DTs in health care is evident, the key focus area might be in understanding how health care operations could benefit from DT use. Developing DTs for designing clinical trials exemplifies how they provide personalized health care by allowing patient-specific responses to therapies to be simulated and treatment plans to be optimized [76]. Nevertheless, the impact of DTs can be extended to population health by providing insights into disease patterns and treatment efficacy across varied populations. Economically, they reduce costs related to traditional clinical trials, such as participant recruitment and logistics, and also position organizations at the forefront of innovation, attracting investment and fostering industry growth. Socially, virtual testing minimizes risks to human participants and enhances trial transparency, as well as addressing disparities by generating inclusive data that accurately represents diverse demographic groups. Thus, DTs promise significant advancements, offering broader societal and economic benefits.

Challenges and Barriers to Implementing DTs in Health Care

Overview

Alongside the promising opportunities, the application of DTs faces several challenges and concerns that could impede their full potential. As with other systems involving the exchange of health information, data security and integrity are essential for preserving patient privacy and ensuring high-fidelity models can be created. To implement DTs widely, significant investments are needed in both digital and physical infrastructure, including scalable cloud computing and high-performance computing for storing and processing the vast amounts of generated data and running complex simulations, and IoT devices with reliable networking for real-time data gathering and supporting rapid data transfer. Such a system upgrade will be particularly challenging due to the need to integrate DTs with existing legacy systems that often lack standardization and interoperability. To facilitate this transformation, it is essential to mitigate these challenges by establishing standardized data formats, communication protocols, and data exchange mechanisms that enhance efficient information flow across systems [134]. Besides, DTs' validation should be first established for an end user to trust the feedback, similar to how predictive performance and clinical validation of artificial intelligence and machine learning models are assessed before use.

There are some other unique aspects related to health care that make the implementation of a DT model challenging but compelling. One important property of DTs is having real time, live, and continuous data flow. However, health data normally is sourced from multiple collections, such as electronic health records, laboratory results, and medical images and there is a delay related to capturing those data. For example, the results of blood tests may take up to 24 hours to be ready or care providers may update coding within a patient record at the end of their stay in the hospital. Capturing accurate data is, in fact, another challenge. Today, the health care industry contributes approximately 30% of the data worldwide and by 2025, it is expected to reach close to 5000 digital device interactions per person per day [135]. This deluge provides an opportunity for developing data-driven models, such as DTs; however, assembling these data from various sources which are normally stored in different formats remains a significant challenge, along with the accuracy of underlying patient-level data stored within datasets. Moreover, the collection and storage of extensive data raises significant data privacy and security concerns, particularly given the sensitive nature of patient information. Solutions involve implementing robust security measures like encryption and access controls to ensure the protection of data during sharing and analysis [18,25,134,136].

Another big difference in DT applications in health care compared to other industries is human involvement, which raises a myriad of ethical issues. It has been discussed that a DT based on a virtual representation of a human may be created in such a way that the virtual version acts on behalf of the actual person and directly affects the person [137]. Therefore, not only does the digital model need informed consent, but also the person

XSL•FO
RenderX

needs to be granted adequate control over their digitally-twinned representation. Trust is another important consideration along the path of the development of DTs in health care. Because of the high level of sensitivity working with health data, DTs as an emerging technology not only need reliable results but also the trust of health care providers.

Future Directions and Policy Considerations for DTs in Health Care

With the rapid growth of DT models, driven by global investments in underlying technologies, such as IoT, and their potential to outperform traditional methods [138], the application of DTs in clinical and ODM is poised to accelerate in the foreseeable future. Effective policy improvement, such as enforcing interoperability standards, providing funding and incentives for technological upgrades, updating data privacy laws, and establishing clear artificial intelligence and machine learning regulatory frameworks will be crucial in facilitating this transmission.

Limitations and Strengths

This study has several limitations. First, the scoping review was limited only to papers published in English. Second, the search process was restricted only to original papers with the term *digital twin* included in either the title, abstract, or keywords which may have resulted in excluding other closely relevant studies that may not have used this term. Third, in the inclusion criteria, we focused on DT applications within ODM and CDM; however, this may have led to excluding a portion of DT applications in other health domains, such as medicine and drug discovery. Fourth, this study reviewed a wide range of DT applications in health care delivery without critically assessing their reporting quality. Therefore, more detailed quantitative analyses can be considered as future research activities that are beyond the scope of this work. Fifth, while our search strategy was comprehensive, some relevant studies indexed under different terminology or metadata, such as MeSH terms, may have been unintentionally missed. Finally, there are articles published between the completion of the search process (June 2024) and the publication of this study that we were not aware of and therefore were not included in this review.

Despite the challenges and limitations highlighted earlier, it is important to acknowledge the strengths of our work. This study offers a comprehensive framework for defining the key characteristics of DTs in health care, providing an essential resource for future research and practical applications. Our review not only outlines the current advancements in DT technology but also pinpoints critical gaps between present capabilities and their full potential, especially in clinical and ODM. The proposed framework serves as a guide for researchers and health care providers in developing DT systems, offering a reference for incorporating all the desirable characteristics of a comprehensive DT to effectively implement and use these technologies. Considering the rapid development of DTs, our review is both timely and highly relevant for researchers and health care practitioners. We adopted a deliberately broad scope, encompassing studies from diverse clinical settings and health care disciplines to ensure a comprehensive understanding of DT applications. By incorporating both academic and gray literature, our findings are further strengthened, offering a solid basis for those working to develop and implement DT systems.

Conclusions

The application of DTs in health care is exponentially growing and it is expected to play a major role in decision-making processes at both operational and clinical levels. This scoping review presents a broad overview of currently developed DT technologies, and the key findings can be summarized as follows: before this study, a common definition of DTs was missing, and researchers often confused other approaches with DTs. In this study, we provided a detailed description of DT elements and characteristics and synthesized these into a universal definition and architecture for DTs for clinical and operational decision support. Second, most prior work has not materialized a major characteristic of DTs: although bidirectional data connection between a real entity and the corresponding virtual one is a fundamental element of a DT, the ongoing update characteristic is lacking in most of the included papers. Third, the implementation of DTs is still in its infancy. Only 19% of the included studies tested the developed DT in a real environment and 95% of them were designed with a single-purpose functionality. Further study is warranted to explore the potential of DTs, for example, integrating ongoing updates into DT models, evaluating practical implementations in health care settings (ie, assessing effectiveness, feasibility, and scalability), and enhancing versatility for broader functionality and seamless integration.

Authors' Contributions

VR, ID, SK, JB, and HH were involved in the concept and design of the review. All authors made major contributions to literature search, data collection, data analysis, and data interpretation. VR created all the figures. VR, ID, SK, JB, and HH wrote and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 498 KB - jmir_v27i1e55015_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Raw data from the review.
[DOCX File , 101 KB - jmir_v27i1e55015_app2.docx]

Multimedia Appendix 3
Distribution of disease areas, services, and computational methods in the included papers.
[DOCX File , 76 KB - jmir_v27i1e55015_app3.docx]

References

1. Sun T, He X, Li Z. Digital twin in healthcare: recent updates and challenges. Digit Health 2023 Jan 03;9:20552076221149651 [FREE Full text] [doi: 10.1177/20552076221149651] [Medline: 36636729]
2. Hassani H, Huang X, MacFeely S. Impactful digital twin in the healthcare revolution. Big Data Cogn Comput 2022 Aug 08;6(3):83. [doi: 10.3390/bdcc6030083]
3. Dang N, Kang H, Lon S, Shim C. 3D digital twin models for bridge maintenance. In: Proceedings of the 10th International Conference on Short and Medium Span Bridges. 2018 Presented at: ICSMSB '18; July 31-August 3, 2018; Montreal, QC p. 73-1-73-9 URL: https://www.researchgate.net/publication/331314334_3D_DIGITAL_TWIN_MODELS_FOR_BRIDGE_MAINTENANCE
4. Ye C, Butler L, Calka B, Iangurazov M, Lu Q, Gregory A. A digital twin of bridges for structural health monitoring. In: Proceedings of the 12th International Workshop on Structural Health Monitoring. 2019 Presented at: SHM '19; September 10-12, 2019; Stanford, CA. [doi: 10.12783/shm2019/32287]
5. Tuegel EJ, Ingraffea AR, Eason TG, Spottswood SM. Reengineering aircraft structural life prediction using a digital twin. Int J Aerosp Eng 2011;2011:1-14. [doi: 10.1155/2011/154798]
6. Rosen R, von Wichert G, Lo G, Bettenhausen KD. About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine 2015;48(3):567-572. [doi: 10.1016/j.ifacol.2015.06.141]
7. Schrotter G, Hürzeler C. The digital twin of the city of Zurich for urban planning. J Photogramm Remote Sens Geoinform Sci 2020 Feb 04;88(1):99-112. [doi: 10.1007/S41064-020-00092-2]
8. Jiang F, Ma L, Broyd T, Chen W, Luo H. Digital twin enabled sustainable urban road planning. Sustain Cities Soc 2022 Mar;78:103645. [doi: 10.1016/j.scs.2021.103645]
9. Deng T, Zhang K, Shen ZJ. A systematic review of a digital twin city: a new pattern of urban governance toward smart cities. J Manag Sci Eng 2021 Jun;6(2):125-134. [doi: 10.1016/j.jmse.2021.03.003]
10. Yu W, Patros P, Young B, Klinac E, Walmsley TG. Energy digital twin technology for industrial energy management: classification, challenges and future. Renew Sustain Energy Rev 2022 Jun;161:112407. [doi: 10.1016/j.rser.2022.112407]
11. Teng SY, Touš M, Leong WD, How BS, Lam HL, Máša V. Recent advances on industrial data-driven energy savings: digital twins and infrastructures. Renew Sustain Energy Rev 2021 Jan;135:110208. [doi: 10.1016/j.rser.2020.110208]
12. Burgos D, Ivanov D. Food retail supply chain resilience and the COVID-19 pandemic: a digital twin-based impact analysis and improvement directions. Transp Res E Logist Transp Rev 2021 Aug;152:102412 [FREE Full text] [doi: 10.1016/j.tre.2021.102412] [Medline: 34934397]
13. Shoji K, Schudel S, Onwude D, Shrivastava C, Defraeye T. Mapping the postharvest life of imported fruits from packhouse to retail stores using physics-based digital twins. Resour Conserv Recycl 2022 Jan;176:105914. [doi: 10.1016/j.resconrec.2021.105914]
14. Liu Y, Zhang L, Yang Y, Zhou L, Ren L, Wang F, et al. A novel cloud-based framework for the elderly healthcare services using digital twin. IEEE Access 2019;7:49088-49101. [doi: 10.1109/access.2019.2909828]
15. Kaul R, Ossai C, Forkan AR, Jayaraman PP, Zelcer J, Vaughan S, et al. The role of AI for developing digital twins in healthcare: the case of cancer care. Wires Data Min Knowl 2022 Nov 21;13(1):e1480. [doi: 10.1002/widm.1480]
16. Barricelli BR, Casiraghi E, Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access 2019;7:167653-167671. [doi: 10.1109/access.2019.2953499]
17. Liu M, Fang S, Dong H, Xu C. Review of digital twin about concepts, technologies, and industrial applications. J Manuf Syst 2021 Jan;58:346-361. [doi: 10.1016/j.jmsy.2020.06.017]
18. Armeni P, Polat I, De Rossi LM, Diaferia L, Meregalli S, Gatti A. Digital twins in healthcare: is it the beginning of a new era of evidence-based medicine? A critical review. J Pers Med 2022 Jul 30;12(8):1255 [FREE Full text] [doi: 10.3390/jpm12081255] [Medline: 36013204]
19. Smith J. Unilever uses virtual factories to tune up its supply chain. The Wall Street Journal. 2019. URL: https://www.wsj.com/articles/unilever-uses-virtual-factories-to-tune-up-its-supply-chain-11563206402 [accessed 2024-04-29]
20. Project focus: how ‘digital twins’ delivered better infrastructure in Chengdu and Porto. Association for Project Management. 2023. URL: https://www.apm.org.uk/resources/find-a-resource/case-studies/case-study-project-focus-how-digital-twins-delivered-better-infrastructure-in-chengdu-and-porto/ [accessed 2024-04-29]
21. More R. Operational decision-making in healthcare using control charts. University of New Haven. URL: https://digitalcommons.newhaven.edu/masterstheses/170/ [accessed 2024-04-29]

XSL•FO
RenderX

22. Ahmadi-Assalemi G, Al-Khateeb H, Maple C, Epiphaniou G, Alhaboby ZA, Alkaabi S. Digital twins for precision healthcare. In: Jahankhani H, Kendzierskyj S, Chelvachandran N, Ibarra J, editors. Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity. Cham, Switzerland: Springer; 2020:133-158.
23. Elkefi S, Asan O. Digital twins for managing health care systems: rapid literature review. J Med Internet Res 2022 Aug 16;24(8):e37641 [FREE Full text] [doi: 10.2196/37641] [Medline: 35972776]
24. Sheng B, Wang Z, Qiao Y, Xie SQ, Tao J, Duan C. Detecting latent topics and trends of digital twins in healthcare: a structural topic model-based systematic review. Digit Health 2023 Oct 12;9:20552076231203672 [FREE Full text] [doi: 10.1177/20552076231203672] [Medline: 37846404]
25. Katsoulakis E, Wang Q, Wu H, Shahriyari L, Fletcher R, Liu J, et al. Digital twins for health: a scoping review. NPJ Digit Med 2024 Mar 22;7(1):77 [FREE Full text] [doi: 10.1038/s41746-024-01073-0] [Medline: 38519626]
26. Boschert S, Rosen R. Digital twin—the simulation aspect. In: Hehenberger P, Bradley D, editors. Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and their Designers. Cham, Switzerland: Springer; 2016:59-74.
27. Martinez-Velazquez R, Gamez R, El Saddik A. Cardio twin: a digital twin of the human heart running on the edge. In: Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications. 2019 Presented at: MeMeA '19; June 26-28, 2019; Istanbul, Turkey p. 1-6 URL: https://ieeexplore.ieee.org/document/8802162 [doi: 10.1109/memea.2019.8802162]
28. Khan A, Milne-Ives M, Meinert E, Iyawa GE, Jones RB, Josephraj AN. A scoping review of digital twins in the context of the COVID-19 pandemic. Biomed Eng Comput Biol 2022 May 24;13:11795972221102115 [FREE Full text] [doi: 10.1177/11795972221102115] [Medline: 35633868]
29. Haleem A, Javaid M, Singh RP, Suman R. Exploring the revolution in healthcare systems through the applications of digital twin technology. Biomed Technol 2023 Dec;4:28-38. [doi: 10.1016/j.bmt.2023.02.001]
30. Possik J, Azar D, Solis AO, Asgary A, Zacharewicz G, Karami A. A distributed digital twin implementation of a hemodialysis unit aimed at helping prevent the spread of the Omicron COVID-19 variant. In: Proceedings of the 2022 IEEE/ACM 26th International Symposium on Distributed Simulation and Real Time Applications. 2022 Presented at: DS-RT '22; September 26-28, 2022; Alès, France p. 168-174 URL: https://ieeexplore.ieee.org/document/9932047 [doi: 10.1109/ds-rt55542.2022.9932047]
31. Gillette K, Gsell MA, Bouyssier J, Prassl AJ, Neic A, Vigmond EJ, et al. Automated framework for the inclusion of a his-Purkinje system in cardiac digital twins of ventricular electrophysiology. Ann Biomed Eng 2021 Dec 24;49(12):3143-3153 [FREE Full text] [doi: 10.1007/s10439-021-02825-9] [Medline: 34431016]
32. Meraghni S, Benaggoune K, Al Masry Z, Terrissa L, Devalland C, Zerhouni N. Towards digital twins driven breast cancer detection. In: Proceedings of the 2021 Computing Conference on Intelligent Computing. 2021 Presented at: IC '21; July 15-16, 2021; Virtual Event p. 87-99 URL: https://link.springer.com/chapter/10.1007/978-3-030-80129-8_7 [doi: 10.1007/978-3-030-80129-8_7]
33. Elayan H, Aloqaily M, Guizani M. Digital twin for intelligent context-aware IoT healthcare systems. IEEE Internet Things J 2021 Dec 1;8(23):16749-16757. [doi: 10.1109/jiot.2021.3051158]
34. Sahal R, Alsamhi SH, Brown KN. Personal digital twin: a close look into the present and a step towards the future of personalised healthcare industry. Sensors (Basel) 2022 Aug 08;22(15):5918 [FREE Full text] [doi: 10.3390/s22155918] [Medline: 35957477]
35. Barbiero P, Viñas Torné R, Lió P. Graph representation forecasting of patient's medical conditions: toward a digital twin. Front Genet 2021 Sep 16;12:652907 [FREE Full text] [doi: 10.3389/fgene.2021.652907] [Medline: 34603366]
36. Venkatapurapu SP, Iwakiri R, Udagawa E, Patidar N, Qi Z, Takayama R, et al. A computational platform integrating a mechanistic model of Crohn's disease for predicting temporal progression of mucosal damage and healing. Adv Ther 2022 Jul 17;39(7):3225-3247 [FREE Full text] [doi: 10.1007/s12325-022-02144-y] [Medline: 35581423]
37. Ritzinger U, Puchinger J, Rudloff C, Hartl RF. Comparison of anticipatory algorithms for a dial-a-ride problem. Eur J Oper Res 2022 Sep;301(2):591-608. [doi: 10.1016/j.ejor.2021.10.060]
38. Manocha A, Afaq Y, Bhatia M. Digital twin-assisted blockchain-inspired irregular event analysis for eldercare. Knowl Based Syst 2023 Jan;260:110138. [doi: 10.1016/j.knosys.2022.110138]
39. Allen A, Siefkas A, Pellegrini E, Burdick H, Barnes G, Calvert J, et al. A digital twins machine learning model for forecasting disease progression in stroke patients. Appl Sci 2021 Jun 16;11(12):5576. [doi: 10.3390/app11125576]
40. Doste R, Lozano M, Jimenez-Perez G, Mont L, Berruezo A, Penela D, et al. Training machine learning models with synthetic data improves the prediction of ventricular origin in outflow tract ventricular arrhythmias. Front Physiol 2022 Aug 12;13:909372 [FREE Full text] [doi: 10.3389/fphys.2022.909372] [Medline: 36035489]
41. Zhong X, Babaie Sarijaloo F, Prakash A, Park J, Huang C, Barwise A, et al. A multidisciplinary approach to the development of digital twin models of critical care delivery in intensive care units. Int J Prod Res 2022 Feb 09;60(13):4197-4213. [doi: 10.1080/00207543.2021.2022235]
42. Chen W, Hong W, Zhang H, Yang P, Tang K. Multi-fidelity simulation modeling for discrete event simulation: an optimization perspective. IEEE Trans Automat Sci Eng 2023 Apr;20(2):1156-1169. [doi: 10.1109/tase.2022.3173296]
43. Croatti A, Gabellini M, Montagna S, Ricci A. On the integration of agents and digital twins in healthcare. J Med Syst 2020 Aug 04;44(9):161 [FREE Full text] [doi: 10.1007/s10916-020-01623-5] [Medline: 32748066]

XSL•FO
RenderX

44. Alazab M, Khan LU, Koppu S, Ramu SP, Iyapparaja M, Boobalan P, et al. Digital twins for healthcare 4.0—recent advances, architecture, and open challenges. IEEE Consumer Electron Mag 2023 Nov 1;12(6):29-37. [doi: 10.1109/mce.2022.3208986]
45. Chakshu NK, Nithiarasu P. An AI based digital-twin for prioritising pneumonia patient treatment. Proc Inst Mech Eng H 2022 Nov 18;236(11):1662-1674 [FREE Full text] [doi: 10.1177/09544119221123431] [Medline: 36121054]
46. Angulo C, Gonzalez-Abril L, Raya C, Ortega JA. A proposal to evolving towards digital twins in healthcare. In: Proceedings of the 8th International Work-Conference on Bioinformatics and Biomedical Engineering. 2020 Presented at: IWBBIO '20; May 6-8, 2020; Granada, Spain p. 418-426 URL: https://link.springer.com/chapter/10.1007/978-3-030-45385-5_37 [doi: 10.1007/978-3-030-45385-5_37]
47. Geissler F, Heiß R, Kopp M, Wiesmüller M, Saake M, Wuest W, et al. Personalized computed tomography - automated estimation of height and weight of a simulated digital twin using a 3D camera and artificial intelligence. Rofo 2021 Apr 03;193(4):437-445 [FREE Full text] [doi: 10.1055/a-1253-8558] [Medline: 33142337]
48. Hernigou P, Olejnik R, Safar A, Martinov S, Hernigou J, Ferre B. Digital twins, artificial intelligence, and machine learning technology to identify a real personalized motion axis of the tibiotalar joint for robotics in total ankle arthroplasty. Int Orthop 2021 Aug 05;45(9):2209-2217. [doi: 10.1007/S00264-021-05175-2]
49. Silfvergren O, Simonsson C, Ekstedt M, Lundberg P, Gennemark P, Cedersund G. Digital twin predicting diet response before and after long-term fasting. PLoS Comput Biol 2022 Sep 12;18(9):e1010469 [FREE Full text] [doi: 10.1371/journal.pcbi.1010469] [Medline: 36094958]
50. Karakra A, Fontanili F, Lamine E, Lamothe J, Taweel A. Pervasive computing integrated discrete event simulation for a hospital digital twin. In: Proceedings of the 15th International Conference on Computer Systems and Applications. 2018 Presented at: AICCSA '18; October 28-November 1, 2018; Aqaba, Jordan p. 1-6 URL: https://ieeexplore.ieee.org/document/8612796 [doi: 10.1109/aiccsa.2018.8612796]
51. Hernigou P, Safar A, Hernigou J, Ferre B. Subtalar axis determined by combining digital twins and artificial intelligence: influence of the orientation of this axis for hindfoot compensation of varus and valgus knees. Int Orthop 2022 May 09;46(5):999-1007. [doi: 10.1007/s00264-022-05311-6] [Medline: 35138455]
52. Chen J. 3D visualization analysis of motion trajectory of knee joint in sports training based on digital twin. Comput Intell Neurosci 2022 Jun 6;2022:3988166-3988169 [FREE Full text] [doi: 10.1155/2022/3988166] [Medline: 35707199]
53. Pilati F, Tronconi R, Nollo G, Heragu SS, Zerzer F. Digital twin of COVID-19 mass vaccination centers. Sustainability 2021 Jul 01;13(13):7396. [doi: 10.3390/su13137396]
54. Chakshu NK, Sazonov I, Nithiarasu P. Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis. Biomech Model Mechanobiol 2021 Apr 16;20(2):449-465 [FREE Full text] [doi: 10.1007/s10237-020-01393-6] [Medline: 33064221]
55. Palaniappan R, Surendran S. A digital twin approach for deepened classification of patients with hepatitis, fibrosis and cirrhosis. J Phys Conf Ser 2022 Sep 01;2335(1):012034. [doi: 10.1088/1742-6596/2335/1/012034]
56. Chakshu NK, Carson J, Sazonov I, Nithiarasu P. A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration-a coupled computational mechanics and computer vision method. Int J Numer Method Biomed Eng 2019 May 20;35(5):e3180 [FREE Full text] [doi: 10.1002/cnm.3180] [Medline: 30648344]
57. Qi Q, Tao F, Hu T, Anwer N, Liu A, Wei Y, et al. Enabling technologies and tools for digital twin. J Manuf Syst 2021 Jan;58:3-21. [doi: 10.1016/j.jmsy.2019.10.001]
58. Jones D, Snider C, Nassehi A, Yon J, Hicks B. Characterising the digital twin: a systematic literature review. CIRP J Manuf Sci Technol 2020 May;29:36-52. [doi: 10.1016/j.cirpj.2020.02.002]
59. Grieves M, Vickers J. Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen FJ, Flumerfelt S, Alves A, editors. Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Cham, Switzerland: Springer; 2017:85-113.
60. Qi J. Mitigating delays and unfairness in appointment systems. Manag Sci 2017 Feb;63(2):566-583. [doi: 10.1287/mnsc.2015.2353]
61. Tao F, Zhang M. Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. IEEE Access 2017;5:20418-20427. [doi: 10.1109/access.2017.2756069]
62. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W. Digital Twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine 2018;51(11):1016-1022. [doi: 10.1016/j.ifacol.2018.08.474]
63. Grieves M. Digital twin: manufacturing excellence through virtual factory replication. Dassault Systèmes. 2014. URL: https://www.3ds.com/fileadmin/PRODUCTS-SERVICES/DELMIA/PDF/Whitepaper/DELMIA-APRISO-Digital-Twin-Whitepaper.pdf [accessed 2024-04-29]
64. Segovia M, Garcia-Alfaro J. Design, modeling and implementation of digital twins. Sensors (Basel) 2022 Jul 20;22(14):5396 [FREE Full text] [doi: 10.3390/s22145396] [Medline: 35891076]
65. Schroeder G, Steinmetz C, Pereira CE, Muller I, Garcia N, Espindola D, et al. Visualising the digital twin using web services and augmented reality. In: Proceedings of the 14th International Conference on Industrial Informatics. 2016 Presented at: INDIN '16; July 19-21, 2016; Poitiers, France p. 522-527 URL: https://ieeexplore.ieee.org/document/7819217/ [doi: 10.1109/indin.2016.7819217]

XSL•FO
RenderX

66. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/m18-0850]
67. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
68. Lai M, Yang H, Gu J, Chen X, Jiang Z. Digital-twin-based online parameter personalization for implantable cardiac defibrillators. In: Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2022 Presented at: EMBC '22; July 11-15, 2022; Scotland, UK p. 20 URL: https://ieeexplore.ieee.org/document/9871142 [doi: 10.1109/embc48229.2022.9871142]
69. Wu Y, Tang R, Kang E, Jiang Z. Model-based clinical assist system for cardiac ablation. In: Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems. 2021 Presented at: ICCPS '21; May 19-21, 2021; Nashville, Tennessee p. 110-120 URL: https://dl.acm.org/doi/10.1145/3450267.3450539 [doi: 10.1145/3450267.3450539]
70. Ahmadian H, Mageswaran P, Walter BA, Blakaj DM, Bourekas EC, Mendel E, et al. Toward an artificial intelligence-assisted framework for reconstructing the digital twin of vertebra and predicting its fracture response. Int J Numer Method Biomed Eng 2022 Jun 26;38(6):e3601 [FREE Full text] [doi: 10.1002/cnm.3601] [Medline: 35403831]
71. Wu C, Jarrett AM, Zhou Z, Elshafeey N, Adrada BE, Candelaria RP. MRI-based digital models forecast patient-specific treatment responses to neoadjuvant chemotherapy in triple-negative breast cancer. Cancer Res 2022;82(18):3394-3404. [doi: 10.1158/0008-5472.can-22-1329]
72. Aubert K, Germaneau A, Rochette M, Ye W, Severyns M, Billot M, et al. Development of digital twins to optimize trauma surgery and postoperative management. A case study focusing on tibial plateau fracture. Front Bioeng Biotechnol 2021 Oct 7;9:722275 [FREE Full text] [doi: 10.3389/fbioe.2021.722275] [Medline: 34692655]
73. Shi Y, Deng X, Tong Y, Li R, Zhang Y, Ren L, et al. Synergistic digital twin and holographic augmented-reality-guided percutaneous puncture of respiratory liver tumor. IEEE Trans Human Mach Syst 2022 Dec;52(6):1364-1374. [doi: 10.1109/thms.2022.3185089]
74. Lauzeral N, Borzacchiello D, Kugler M, George D, Rémond Y, Hostettler A, et al. A model order reduction approach to create patient-specific mechanical models of human liver in computational medicine applications. Comput Methods Programs Biomed 2019 Mar;170:95-106. [doi: 10.1016/j.cmpb.2019.01.003] [Medline: 30712607]
75. Wan Z, Dong Y, Yu Z, Lv H, Lv Z. Semi-supervised support vector machine for digital twins based brain image fusion. Front Neurosci 2021 Jul 9;15:705323 [FREE Full text] [doi: 10.3389/fnins.2021.705323] [Medline: 34305523]
76. Qi T, Cao Y. Virtual clinical trials: a tool for predicting patients who may benefit from treatment beyond progression with pembrolizumab in non-small cell lung cancer. CPT Pharmacometrics Syst Pharmacol 2023 Feb 22;12(2):236-249 [FREE Full text] [doi: 10.1002/psp4.12896] [Medline: 36547213]
77. Tardini E, Zhang X, Canahuate G, Wentzel A, Mohamed AS, Van Dijk L, et al. Optimal treatment selection in sequential systemic and locoregional therapy of oropharyngeal squamous carcinomas: deep Q-learning with a patient-physician digital twin dyad. J Med Internet Res 2022 Apr 20;24(4):e29455 [FREE Full text] [doi: 10.2196/29455] [Medline: 35442211]
78. Shamanna P, Joshi S, Shah L, Dharmalingam M, Saboo B, Mohammed J, et al. Type 2 diabetes reversal with digital twin technology-enabled precision nutrition and staging of reversal: a retrospective cohort study. Clin Diabetes Endocrinol 2021 Nov 15;7(1):21 [FREE Full text] [doi: 10.1186/s40842-021-00134-7] [Medline: 34776010]
79. Shamanna P, Saboo B, Damodharan S, Mohammed J, Mohamed M, Poon T, et al. Reducing HbA1c in type 2 diabetes using digital twin technology-enabled precision nutrition: a retrospective analysis. Diabetes Ther 2020 Nov 25;11(11):2703-2714 [FREE Full text] [doi: 10.1007/s13300-020-00931-w] [Medline: 32975712]
80. Hwang SH, Park JM, Kim JH, Shim JS, Jung WS. A case of full-mouth rehabilitation using a digital twin based on multi-source data in a patient with vertical dimension loss due to multiple tooth wear and fractures. J Implantol Appl Sci 2022 Sep 30;26(3):198-208. [doi: 10.32542/implantology.2022020]
81. Lal A, Li G, Cubro E, Chalmers S, Li H, Herasevich V, et al. Development and verification of a digital twin patient model to predict specific treatment response during the first 24 hours of sepsis. Crit Care Explor 2020 Nov;2(11):e0249 [FREE Full text] [doi: 10.1097/CCE.0000000000000249] [Medline: 33225302]
82. Cho SW, Byun SH, Yi S, Jang WS, Kim JC, Park IY, et al. Sagittal relationship between the maxillary central incisors and the forehead in digital twins of Korean adult females. J Pers Med 2021 Mar 13;11(3):203 [FREE Full text] [doi: 10.3390/jpm11030203] [Medline: 33805617]
83. Shamanna P, Dharmalingam M, Sahay R, Mohammed J, Mohamed M, Poon T, et al. Retrospective study of glycemic variability, BMI, and blood pressure in diabetes patients in the Digital Twin Precision Treatment Program. Sci Rep 2021 Jul 21;11(1):14892 [FREE Full text] [doi: 10.1038/s41598-021-94339-6] [Medline: 34290310]
84. Golse N, Joly F, Combari P, Lewin M, Nicolas Q, Audebert C, et al. Predicting the risk of post-hepatectomy portal hypertension using a digital twin: a clinical proof of concept. J Hepatol 2021 Mar;74(3):661-669. [doi: 10.1016/j.jhep.2020.10.036] [Medline: 33212089]
85. Ahmadian H, Mageswaran P, Walter BA, Blakaj DM, Bourekas EC, Mendel E, et al. A digital twin for simulating the vertebroplasty procedure and its impact on mechanical stability of vertebra in cancer patients. Int J Numer Method Biomed Eng 2022 Jun 07;38(6):e3600 [FREE Full text] [doi: 10.1002/cnm.3600] [Medline: 35347880]

XSL•FO
RenderX

86. Talukder AK, Schriml L, Ghosh A, Biswas R, Chakrabarti P, Haas RE. Diseasomics: actionable machine interpretable disease knowledge at the point-of-care. PLOS Digit Health 2022 Oct 20;1(10):e0000128 [FREE Full text] [doi: 10.1371/journal.pdig.0000128] [Medline: 36812614]
87. Batch KE, Yue J, Darcovich A, Lupton K, Liu CC, Woodlock DP, et al. Developing a cancer digital twin: supervised metastases detection from consecutive structured radiology reports. Front Artif Intell 2022 Mar 2;5:826402 [FREE Full text] [doi: 10.3389/frai.2022.826402] [Medline: 35310959]
88. Tai Y, Zhang L, Li Q, Zhu C, Chang V, Rodrigues JJ, et al. Digital-twin-enabled IoMT system for surgical simulation using rAC-GAN. IEEE Internet Things J 2022 Nov 1;9(21):20918-20931. [doi: 10.1109/jiot.2022.3176300]
89. Baena JP, Estrada OG, Villegas D. Structural analysis of bone by segmentation and finite element analysis in patients with osteoporosis. J Phys Conf Ser 2021 Oct 01;2046(1):012019. [doi: 10.1088/1742-6596/2046/1/012019]
90. Roy D, Mazumder O, Khandelwal S, Sinha A. Wearable sensor driven Cardiac model to derive hemodynamic insights during exercise. In: Proceedings of the 2021 Workshop on Body-Centric Computing Systems. 2021 Presented at: BodySys '21; June 24, 2021; Virtual Event p. 30-35 URL: https://dl.acm.org/doi/10.1145/3469260.3469670#:~:text=The%20model%20incorporates%20the%20real,an%20individual%20while%20performing%20exercises [doi: 10.1145/3469260.3469670]
91. Hussain I, Hossain MA, Park SJ. A healthcare digital twin for diagnosis of stroke. In: Proceedings of the 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health. 2021 Presented at: BECITHCON '21; December 4-5, 2021; Dhaka, Bangladesh p. 18-21 URL: https://ieeexplore.ieee.org/document/9893641 [doi: 10.1109/becithcon54710.2021.9893641]
92. Scheuermann C, Binderberger T, Von Frankenberg N, Werner A. Digital twin: a machine learning approach to predict individual stress levels in extreme environments. In: Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers. 2020 Presented at: UbiComp/ISWC '20; September 12-17, 2020; Virtual Event p. 657-664 URL: https://dl.acm.org/doi/10.1145/3410530.3414316 [doi: 10.1145/3410530.3414316]
93. Semakova A, Zvartau N. Data-driven identification of hypertensive patient profiles for patient population simulation. Procedia Comput Sci 2018;136:433-442. [doi: 10.1016/j.procs.2018.08.269]
94. Ossai C, Wickramasinghe N. A Bayesian network model to establish a digital twin architecture for superior falls risk prediction. In: Proceedings of the 27th Americas Conference on Information Systems. 2021 Presented at: AMCIS '21; August 9-13, 2021; Virtual Event p. 2021 URL: https://aisel.aisnet.org/amcis2021/healthcare_it/sig_health/3/
95. Kim JK, Lee SJ, Hong SH, Choi IY. Machine-learning-based digital twin system for predicting the progression of prostate cancer. Appl Sci 2022 Aug 15;12(16):8156. [doi: 10.3390/app12168156]
96. Zhou C, Chase JG, Knopp J, Sun Q, Tawhai M, Möller K, et al. Virtual patients for mechanical ventilation in the intensive care unit. Comput Methods Programs Biomed 2021 Feb;199:105912. [doi: 10.1016/j.cmpb.2020.105912] [Medline: 33360683]
97. Azzolin L, Eichenlaub M, Nagel C, Nairn D, Sanchez J, Unger L, et al. Personalized ablation vs. conventional ablation strategies to terminate atrial fibrillation and prevent recurrence. Europace 2023 Feb 08;25(1):211-222 [FREE Full text] [doi: 10.1093/europace/euac116] [Medline: 35943361]
98. Ang CY, Lee JW, Chiew YS, Wang X, Tan CP, Cove ME, et al. Virtual patient framework for the testing of mechanical ventilation airway pressure and flow settings protocol. Comput Methods Programs Biomed 2022 Nov;226:107146 [FREE Full text] [doi: 10.1016/j.cmpb.2022.107146] [Medline: 36191352]
99. van Osta N, Kirkels FP, van Loon T, Koopsen T, Lyon A, Meiburg R, et al. Uncertainty quantification of regional cardiac tissue properties in arrhythmogenic cardiomyopathy using adaptive multiple importance sampling. Front Physiol 2021 Sep 30;12:738926 [FREE Full text] [doi: 10.3389/fphys.2021.738926] [Medline: 34658923]
100. Kardampiki E, Vignali E, Haxhiademi D, Federici D, Ferrante E, Porziani S, et al. The hemodynamic effect of modified Blalock–Taussig shunt morphologies: a computational analysis based on reduced order modeling. Electronics 2022 Jun 21;11(13):1930. [doi: 10.3390/electronics11131930]
101. Jiang J, Li Q, Yang F. TCM physical health management training and nursing effect evaluation based on digital twin. Sci Program 2022 Sep 27;2022:1-13. [doi: 10.1155/2022/3907481]
102. Kobayashi T, Fukae K, Imai T, Arai K. Digital twin agent for super-aged society. In: Proceedings of the 2022 IEEE International Conference on Consumer Electronics. 2022 Presented at: ICCE '22; January 7-9, 2022; Las Vegas, NV p. 1-6 URL: https://ieeexplore.ieee.org/document/9730230 [doi: 10.1109/icce53296.2022.9730230]
103. Förster KM, Roth CJ, Hilgendorff A, Ertl-Wagner B, Flemmer AW, Wall WA. In silico numerical simulation of ventilator settings during high-frequency ventilation in preterm infants. Pediatr Pulmonol 2021 Dec 25;56(12):3839-3846. [doi: 10.1002/ppul.25626] [Medline: 34432956]
104. Fu W, Sharma S, Abadi E, Iliopoulos AS, Wang Q, Lo JY, et al. iPhantom: a framework for automated creation of individualized computational phantoms and its application to CT organ dosimetry. IEEE J Biomed Health Inform 2021 Aug;25(8):3061-3072. [doi: 10.1109/jbhi.2021.3063080]

XSL•FO
RenderX

105. Gillette K, Gsell MA, Prassl AJ, Karabelas E, Reiter U, Reiter G, et al. A Framework for the generation of digital twins of cardiac electrophysiology from clinical 12-leads ECGs. Med Image Anal 2021 Jul;71:102080 [FREE Full text] [doi: 10.1016/j.media.2021.102080] [Medline: 33975097]
106. Goodwin GC, Seron MM, Medioli AM, Smith T, King BR, Smart CE. A systematic stochastic design strategy achieving an optimal tradeoff between peak BGL and probability of hypoglycaemic events for individuals having type 1 diabetes mellitus. Biomed Signal Process Control 2020 Mar;57:101813. [doi: 10.1016/j.bspc.2019.101813]
107. Yuan X, Zhang J, Luo J, Chen J, Shi Z, Qin M. An efficient digital twin assisted clustered federated learning algorithm for disease prediction. In: Proceedings of the 95th Vehicular Technology Conference. 2022 Presented at: VTC '22; June 19-22, 2022; Helsinki, Finland p. 1-6 URL: https://ieeexplore.ieee.org/document/9860704 [doi: 10.1109/vtc2022-spring54318.2022.9860704]
108. Alcaraz JC, Moghaddamnia S, Fuhrwerk M, Peissig J. Efficiency of the memory polynomial model in realizing digital twins for gait assessment. In: Proceedings of the 27th European Signal Processing Conference. 2019 Presented at: EUSIPCO '19; September 2-6, 2019; Coruna, Spain p. 1-5 URL: https://ieeexplore.ieee.org/document/8903143 [doi: 10.23919/eusipco.2019.8903143]
109. Shu H, Liang R, Li Z, Goodridge A, Zhang X, Ding H, et al. Twin-S: a digital twin for skull base surgery. Int J Comput Assist Radiol Surg 2023 Jun 09;18(6):1077-1084 [FREE Full text] [doi: 10.1007/s11548-023-02863-9] [Medline: 37160583]
110. Sarp S, Kuzlu M, Zhao Y, Gueler O. Digital twin in healthcare: a study for chronic wound management. IEEE J Biomed Health Inform 2023 Nov;27(11):5634-5643. [doi: 10.1109/jbhi.2023.3299028]
111. Demir O, Uslan I, Buyuk M, Salamci MU. Development and validation of a digital twin of the human lower jaw under impact loading by using non-linear finite element analyses. J Mech Behav Biomed Mater 2023 Dec;148:106207. [doi: 10.1016/j.jmbbm.2023.106207] [Medline: 37922761]
112. Koopsen T, Gerrits W, van Osta N, van Loon T, Wouters P, Prinzen FW, et al. Virtual pacing of a patient's digital twin to predict left ventricular reverse remodelling after cardiac resynchronization therapy. Europace 2023 Dec 28;26(1):e009 [FREE Full text] [doi: 10.1093/europace/euae009] [Medline: 38288616]
113. Batagov A, Dalan R, Wu A, Lai W, Tan CS, Eisenhaber F. Generalized metabolic flux analysis framework provides mechanism-based predictions of ophthalmic complications in type 2 diabetes patients. Health Inf Sci Syst 2023 Dec 29;11(1):18 [FREE Full text] [doi: 10.1007/s13755-023-00218-x] [Medline: 37008895]
114. Joshi S, Shamanna P, Dharmalingam M, Vadavi A, Keshavamurthy A, Shah L, et al. Digital twin-enabled personalized nutrition improves metabolic dysfunction-associated fatty liver disease in type 2 diabetes: results of a 1-year randomized controlled study. Endocr Pract 2023 Dec;29(12):960-970 [FREE Full text] [doi: 10.1016/j.eprac.2023.08.016] [Medline: 37778441]
115. Grieb N, Schmierer L, Kim HU, Strobel S, Schulz C, Meschke T, et al. A digital twin model for evidence-based clinical decision support in multiple myeloma treatment. Front Digit Health 2023 Dec 20;5:1324453 [FREE Full text] [doi: 10.3389/fdgth.2023.1324453] [Medline: 38173909]
116. Zhang Y, Qin G, Aguilar B, Rappaport N, Yurkovich JT, Pflieger L, et al. A framework towards digital twins for type 2 diabetes. Front Digit Health 2024 Jan 26;6:1336050 [FREE Full text] [doi: 10.3389/fdgth.2024.1336050] [Medline: 38343907]
117. Rouhollahi A, Willi JN, Haltmeier S, Mehrtash A, Straughan R, Javadikasgari H, et al. CardioVision: a fully automated deep learning package for medical image segmentation and reconstruction generating digital twins for patients with aortic stenosis. Comput Med Imaging Graph 2023 Oct;109:102289. [doi: 10.1016/j.compmedimag.2023.102289] [Medline: 37633032]
118. Serra D, Franco P, Romero P, Romitti G, Garcia-Fernandez I, Lozano M, et al. Assessment of risk for ventricular tachycardia based on extensive electrophysiology simulations. Annu Int Conf IEEE Eng Med Biol Soc 2023 Jul;2023:1-4. [doi: 10.1109/EMBC40787.2023.10340169] [Medline: 38083190]
119. Chahal Y, Tokas R, Sharma K. Smart solution using digital twin and IoT for diabetic retinopathy. In: Proceedings of the 14th International Conference on Computing Communication and Networking Technologies. 2023 Presented at: ICCCNT '23; July 6-8, 2023; Delhi, India p. 1-6 URL: https://ieeexplore.ieee.org/document/10307620 [doi: 10.1109/icccnt56998.2023.10307620]
120. Uyttendaele V, Knopp JL, Desaive T, Chase JG. Clinical trial validation of the STAR-3D glycemic control framework. IFAC-PapersOnLine 2023;56(2):4758-4763. [doi: 10.1016/j.ifacol.2023.10.1239]
121. Ložek M, Kovanda J, Kubuš P, Vrbík M, Lhotská L, Lumens J, et al. How to assess and treat right ventricular electromechanical dyssynchrony in post-repair tetralogy of Fallot: insights from imaging, invasive studies, and computational modelling. Europace 2024 Feb 01;26(2):e024 [FREE Full text] [doi: 10.1093/europace/euae024] [Medline: 38266248]
122. Thamotharan P, Srinivasan S, Kesavadev J, Krishnan G, Mohan V, Seshadhri S, et al. Human digital twin for personalized elderly type 2 diabetes management. J Clin Med 2023 Mar 07;12(6):2094 [FREE Full text] [doi: 10.3390/jcm12062094] [Medline: 36983097]
123. Cappon G, Vettoretti M, Sparacino G, Favero SD, Facchinetti A, Facchinetti A. ReplayBG: a digital twin-based methodology to identify a personalized model from type 1 diabetes data and simulate glucose concentrations to assess alternative therapies. IEEE Trans Biomed Eng 2023 Nov;70(11):3227-3238. [doi: 10.1109/tbme.2023.3286856]

XSL•FO
RenderX

124. Salvador M, Kong F, Peirlinck M, Parker DW, Chubb H, Dubin AM, et al. Digital twinning of cardiac electrophysiology for congenital heart disease. J R Soc Interface 2024 Jun 05;21(215):20230729. [doi: 10.1098/rsif.2023.0729] [Medline: 38835246]
125. Dubs L, Charitatos V, Buoso S, Wegener S, Winklhofer S, Alkadhi H, et al. Assessment of extracranial carotid artery disease using digital twins - a pilot study. Neuroimage Clin 2023;38:103435 [FREE Full text] [doi: 10.1016/j.nicl.2023.103435] [Medline: 37245493]
126. Khan S, Alzaabi A, Iqbal Z, Ratnarajah T, Arslan T. A novel digital twin (DT) model based on WiFi CSI, signal processing and machine learning for patient respiration monitoring and decision-support. IEEE Access 2023;11:103554-103568. [doi: 10.1109/access.2023.3316508]
127. Augusto V, Murgier M, Viallon A. A modelling and simulation framework for intelligent control of emergency units in the case of major crisis. In: Proceedings of the 2018 Winter Simulation Conference. 2018 Presented at: WSC '18; December 9-12, 2018; Gothenburg, Sweden p. 2495-2506 URL: https://ieeexplore.ieee.org/document/8632438 [doi: 10.1109/wsc.2018.8632438]
128. Maïzi Y, Bendavid Y. Designing a RFID/IoT prototype for improving COVID19 test centers daily operations. In: Proceedings of the 20th International Conference on Modelling and Applied Simulation. 2021 Presented at: MAS '21; September 15-17, 2021; Virtual Event p. 127-134 URL: https://www.cal-tek.eu/proceedings/i3m/2021/mas/016/pdf.pdf [doi: 10.46354/i3m.2021.mas.016]
129. Karakra A, Fontanili F, Lamine E, Lamothe J. A discrete event simulation-based methodology for building a digital twin of patient pathways in the hospital for near real-time monitoring and predictive simulation. digitaltwin 2022 Jan 7;2:1. [doi: 10.12688/digitaltwin.17454.1]
130. Basaglia A, Spacone E, van de Lindt JW, Kirsch TD. A discrete-event simulation model of hospital patient flow following major earthquakes. Int J Disaster Risk Reduct 2022 Mar;71:102825. [doi: 10.1016/j.ijdrr.2022.102825]
131. Zackoff MW, Rios M, Davis D, Boyd S, Roque I, Anderson I, et al. Immersive virtual reality onboarding using a digital twin for a new clinical space expansion: a novel approach to large-scale training for health care providers. J Pediatr 2023 Jan;252:7-10.e3. [doi: 10.1016/j.jpeds.2022.07.031] [Medline: 35973444]
132. Zhang X. Application of discrete event simulation in health care: a systematic review. BMC Health Serv Res 2018 Sep 4;18(1):58. [doi: 10.1186/S12913-018-3456-4]
133. Niederer SA, Sacks MS, Girolami M, Willcox K. Scaling digital twins from the artisanal to the industrial. Nat Comput Sci 2021 May 24;1(5):313-320. [doi: 10.1038/s43588-021-00072-5] [Medline: 38217216]
134. Turab M, Jamil S. A comprehensive survey of digital twins in healthcare in the era of metaverse. BioMedInformatics 2023 Jul 21;3(3):563-584. [doi: 10.3390/biomedinformatics3030039]
135. The healthcare data explosion. RBC Capital Markets. URL: https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion [accessed 2024-04-29]
136. Coorey G, Figtree GA, Fletcher DF, Snelson VJ, Vernon ST, Winlaw D, et al. The health digital twin to tackle cardiovascular disease-a review of an emerging interdisciplinary field. NPJ Digit Med 2022 Aug 26;5(1):126 [FREE Full text] [doi: 10.1038/s41746-022-00640-7] [Medline: 36028526]
137. Braun M. Represent me: please! towards an ethics of digital twins in medicine. J Med Ethics 2021 Mar 15;47(6):394-400. [doi: 10.1136/medethics-2020-106134] [Medline: 33722986]
138. Jakovljevic M, Potapchik E, Popovich L, Barik D, Getzen TE. Evolving health expenditure landscape of the BRICS nations and projections to 2025. Health Econ 2017 Jul 29;26(7):844-852. [doi: 10.1002/hec.3406] [Medline: 27683202]

Abbreviations

CDM: clinical decision-making
DT: digital twin
ICD-10: International Classification of Diseases, 10th Revision
IoT: Internet of Things
ODM: operational decision-making
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

XSL•FO
RenderX

Edited by N Cahill, T Leung; submitted 30.11.23; peer-reviewed by J Ramirez, W Tam; comments to author 17.04.24; revised version received 17.07.24; accepted 28.10.24; published 08.01.25.

Please cite as:
Riahi V, Diouf I, Khanna S, Boyle J, Hassanzadeh H
Digital Twins for Clinical and Operational Decision-Making: Scoping Review
J Med Internet Res 2025;27:e55015
URL: https://www.jmir.org/2025/1/e55015
doi:10.2196/55015
PMID:

©Vahid Riahi, Ibrahima Diouf, Sankalp Khanna, Justin Boyle, Hamed Hassanzadeh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Integrated Digital Solutions to Empower Older Adults in Aspects Related to Their Health: Systematic Review and Meta-Analysis

Ana Isabel Martins[1], PhD; Óscar Ribeiro[2], PhD; Gonçalo Santinha[3], PhD; Telmo Silva[4], PhD; Nelson P Rocha[5], PhD; Anabela G Silva[6], PhD

[1]Center for Health Technology and Services Research, University of Aveiro, Aveiro, Portugal

[2]Department of Education and Psychology, Center for Health Technology and Services Research, University of Aveiro, Aveiro, Portugal

[3]Department of Social, Political and Territorial Sciences, Research Unit in Governance, Competitiveness and Public Policies, University of Aveiro, Aveiro, Portugal

[4]Department of Communication and Art, DigiMedia, University of Aveiro, Aveiro, Portugal

[5]Department of Medical Sciences, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal

[6]School of Health Sciences, Center for Health Technology and Services Research, University of Aveiro, Aveiro, Portugal

Corresponding Author:
Anabela G Silva, PhD
School of Health Sciences
Center for Health Technology and Services Research
University of Aveiro
Campus Universitário de Santiago
Aveiro, 3810-143
Portugal
Phone: 351 234247119
Email: asilva@ua.pt

Abstract

Background: Digital solutions, such as mobile apps or telemonitoring devices, are frequently considered facilitators in the process of empowering older adults, but they can also act as a source of digital exclusion or disempowerment if they are not adequate for older adults' needs and characteristics.

Objective: This study aimed to synthesize and critically evaluate existing evidence on the effectiveness of integrated digital solutions that enable interaction for empowering older adults in aspects related to their health and to explore potential factors (eg, type of technology, participants' characteristics) impacting effectiveness.

Methods: A systematic search was carried out in PubMed, ScienceDirect, SCOPUS, EBSCO, and SciELO using a combination of terms informed by previous reviews on empowerment. Screening of references was performed against predefined inclusion criteria. Data extraction and the methodological quality of included studies using the PEDro (Physiotherapy Evidence Database) scale were performed by 2 authors. The certainty of evidence was graded for the main comparisons and outcomes of the review using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) framework. When at least 3 studies were available within the same domain of empowerment (knowledge, support by others, capacities, and behaviors) and comparison group, a meta-analysis was performed.

Results: A total of 30 manuscripts were included in the review. Regarding knowledge, there was very low certainty of evidence of a medium effect size (ES) favoring the digital intervention group (k=5, ES=0.40, 95% CI 0.07-0.73, I^2=79%). Regarding capacities, there was low certainty of evidence of no between-group differences (k=5, d=0.13, 95% CI –0.02 to 0.29, I^2=0%) when comparing digital solutions against no intervention, low certainty of evidence of a medium ES favoring the digital intervention group (k=13, d=0.29, 95% CI 0.07-0.52, I^2=79%) when comparing digital solutions against usual care, and very low certainty of evidence of no between-group differences (k=4, d=0.97, 95% CI –0.62 to 2.56, I^2=97%) when comparing digital interventions to face-to-face interventions. Regarding social support and behaviors, no meta-analysis was possible, and existing evidence is conflicting.

XSL•FO
RenderX

Conclusions: There is very-low-to-low certainty of evidence that using integrated digital solutions results in increased knowledge and increased capacities (mainly self-efficacy) when compared to usual care and impacts capacities to an extent similar to face-to-face interventions at postintervention. Interestingly, results also suggest, with low certainty of evidence, that there are no differences between using digital solutions and no intervention for improving capacities. Included studies and technologies were diverse, and meta-analysis showed high heterogeneity, which limits the confidence in the results and suggests that further research might affect the conclusions of this review.

Trial Registration: PROSPERO CRD42022346823; https://tinyurl.com/39k29pzc

(J Med Internet Res 2025;27:e54466) doi:10.2196/54466

KEYWORDS

empowerment; older adults; digital health; digital solutions; effectiveness; empowerment; health related; outcomes; systematic review; meta-analysis; synthesize; evaluate; apps; mHealth; mobile health

Introduction

Empowerment is a complex concept with multiple definitions [1,2] and is affected by multiple factors, including individual and contextual factors [3]. One definition refers to empowerment as a social process of recognizing, promoting, and enhancing people's ability to meet their own needs, solve their problems, and mobilize the necessary resources to feel in control of their own lives [4]. In other words, empowerment is a process by which people, organizations, and communities gain mastery over issues of concern to them [3]. When a person or group is empowered, they can make impactful decisions, allowing them to turn those decisions into the actions and results they desire [1,5].

Empowering older adults is of great importance as a means of promoting autonomy and independence by allowing them to make their own decisions, foster social connections that endorse social participation [1], and engage in activities that bring them joy and fulfillment [6]. In addition, empowering older adults contributes to ensuring that they are treated with dignity and respect; to acknowledging their life experiences, wisdom, and contributions to society [7]; and to fostering a sense of self-worth, value, and independence [8]. Empowered older adults feel encouraged to take control of their physical, mental, and emotional health and make informed decisions about their health care and well-being [8,9]. Overall, empowerment means having an opportunity to learn, discuss, decide, and act on decisions across all areas and dimensions of one's own life [10].

Digital solutions, such as mobile apps or telemonitoring systems, are frequently considered facilitators in the process of empowering older adults [11] as they often provide access to a vast amount of information, resources, and services [12] and improve social connection and communication, particularly with family members and friends who do not live nearby [13]. This enables older adults to stay informed, engage in lifelong learning, have access to services and products, make informed decisions about their health and well-being [14], and maintain social connections regardless of physical distance. However, digital solutions can also act as a source of segregation, digital exclusion, or disempowerment for older people as they are not adequate for older adults' needs and characteristics [15-17]. Interaction (ie, the ability to engage with a digital solution) is likely to be a key aspect of the potential of digital solutions for empowering as interactive solutions allow for a more personalized and immersive experience [18]. Nevertheless, several barriers hinder the effective use of digital solutions by older adults, and potentially its ability to empower them, such as the digital divide, the technological complexity, or the lack of flexibility and adequacy to their needs and characteristics, such as age-related or disease-related physical and cognitive constraints [19-21]. Therefore, the potential of digital solutions as a means of empowering older adults might not translate into an effective positive impact, or this impact might vary across different types of digital solutions or depend on the characteristics of older adults. However, and to the best of our knowledge, there is no available synthesis of evidence on the effectiveness of digital solutions for empowering older adults that clarifies the impact of digital solutions or that attempts to identify potential factors affecting their effectiveness. Thus, this study aimed to synthesize and critically evaluate existing evidence on the effectiveness of integrated digital solutions that enable interaction for empowering older adults in aspects related to their health and to explore potential factors (eg, type of technology, participants' characteristics) impacting effectiveness.

Methods

Study Design

This study was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42022346823). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) were followed in this review [22].

Eligibility Criteria

Studies were included in this review if they met the following inclusion criteria:

- Participants had a mean age of 60 years or more.
- Studies reported randomized or quasi-randomized trials aiming primarily at assessing the effectiveness of technology for empowering older adults.
- One of the arms used an integrated digital solution that enables interaction with the users. Examples include complex technologies, such as remote health and social monitoring solutions, and ambient assisted living solutions or simpler solutions, such as mobile phone apps and

XSL•FO
RenderX

websites. A description and examples are presented in Multimedia Appendix 2.
- The comparison was no intervention, usual care, or any intervention (eg, education, written material) that did not use a digital solution.
- The outcome of interest was 1 of the following 4 dimensions of empowerment [23]: (1) patients' capacities (tolerance of uncertainty, skill and technique acquisition, constructive attitudes and approaches, self-monitoring and insight, emotional well-being, positive attitude, self-efficacy, ownership, and self-information search), (2) patients' knowledge (knowledge and confidence in decision-making, informed confidence), (3) patients' behaviors (positive and active engagement in life, health-directed behavior, health service navigation, enabling others, social advocacy, commitment, informed choice, navigation, partnership, and health maintenance), and (4) support by others (social integration and support, integration, client-provider relationship, client-client support, control, doctors' health locus of control, and other people's health locus of control).
- Studies were published in English, Spanish, or Portuguese. Other languages were excluded because of limitations in the language abilities of the research team.

Studies with measurement instruments including individual items on empowerment but not providing individual item scores were excluded.

Search Strategy

A systematic search was carried out in the following electronic databases: PubMed, ScienceDirect, SCOPUS, EBSCO, and SciELO. The search strategy consisted of a combination of terms based on previous reviews on empowerment [24] and technology [25] (for details, see Multimedia Appendix 3). In PubMed, the following search string was used: ((“Empowerment”[Mesh] OR “Patient Participation”[Mesh]) OR “self-efficacy” OR “locus of control” OR perceived autonomy OR perception of autonomy OR overconfidence) AND ((“Technology”[Mesh]) OR “Digital Technology”[Mesh] OR “Mobile apps” OR m-health OR ehealth OR e-health OR “digital solution” OR “Ambient Assisted Living” OR “digital services” OR website OR “digital platform” OR esocial). This string was adapted for the remaining databases. No filters were used. The search was conducted on May 4, 2022, and updated on September 6, 2022, and included papers from each database's inception until August 31, 2023.

Selection Process

After importing the papers into Mendeley (Elsevier), duplicates were removed. Next, the papers were imported to CADIMA (Julius Kühn-Institut), the software used to support screening, which was performed first by title and abstract and then by full text against the eligibility criteria. The entire selection process was performed independently by 2 authors (AIM and AGS). Disagreements or discrepancies during the selection process were resolved by consensus or consultation with a third author (NPR or OR).

Data Extraction and Management

A customized Microsoft Excel form was used for data extraction, which was piloted using 3 studies. The data extraction was manually completed by 1 author (AIM) and checked by a second author (AGS). The following data were extracted from each study: author and date, measurement instruments used to assess the outcomes of interest, intervention (type and duration), characteristics of the sample per group (health condition, sample size, age, sex), comparison, outcomes, and results (at postintervention and follow-up assessments). Data on the period of follow-up and the results of follow-up assessments were collected to allow for analysis of the effectiveness of digital solutions at different time points. When multiple outcomes within the same domain of empowerment were reported in the same paper, the choice of which to include in the meta-analysis was based on similarity with outcomes from other studies in the same meta-analysis. When data were not fully accessible (eg, presented in graphs), they were not extracted and authors were contacted.

Extracted data were summarized under 1 of the 4 domains of empowerment previously defined (patients' capacities, patients' behaviors, patients' knowledge, and support by others). When necessary, the transformation of data, such as the conversion of SEs and CIs into SDs, was performed according to Cochrane recommendations [26].

Methodological Quality Assessment

The quality of studies was assessed using the PEDro (Physiotherapy Evidence Database) scale, which is a valid instrument [27]. Each study was assessed independently by pairs of reviewers (AIM and AGS, NPR and OR), and disagreements were resolved by a third author. The PEDro scale is composed of 11 items, and each item is scored as either present (1) or absent (0). The first item is not considered in the total score, leading to a maximum score of 10. The total score can be interpreted as follows: 0-3, poor; 4-5, fair; 6-8, good; and 9-10, excellent methodological quality [28].

Assessment of the Certainty of Evidence

The certainty of evidence was graded for the main comparisons and outcomes of this review. The overall quality and strength of the evidence per outcome was assessed according to the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) framework. Evidence was assessed against (1) risk of bias, (2) inconsistency of results, (3) indirectness, (4) imprecision, and (5) publication bias, and the overall quality of evidence was classified as high, moderate, low, or very low [29,30]. We started with high evidence as studies included in the meta-analysis were randomized trials, and then, the evidence was downgraded, considering the researchers' assessment of each criterion.

Data Synthesis

When at least 3 studies were available within the same empowerment domain and comparison group (no intervention, usual care, or face-to-face intervention), a meta-analysis was performed. In addition, separate analyses were performed considering whether the digital intervention was the only intervention or was administered together with a face-to-face

XSL•FO
RenderX

intervention. A sensitivity analysis was performed excluding lower-quality studies. Effect sizes (ESs) and their 95% CIs were determined by the standardized mean difference (SMD) and classified according to Cohen's guidelines as small (0.20), medium (0.50), and large (0.80) effects [31]. Statistical heterogeneity was investigated using I^2 statistics interpreted as 25%, 50%, and 75%, reflecting low, moderate, and high statistical heterogeneity, respectively. When at least 10 studies were available, funnel plots were generated to inform publication bias. Follow-up periods were chosen to increase the similarity between studies in the same meta-analysis. The meta-analysis was performed using R version 4.2.2 (R Foundation for Statistical Computing). When no meta-analysis was possible, a qualitative synthesis was performed, indicating the comparisons investigated by the studies and the direction of the effect reported.

Figure 1. Flow diagram of the current review.

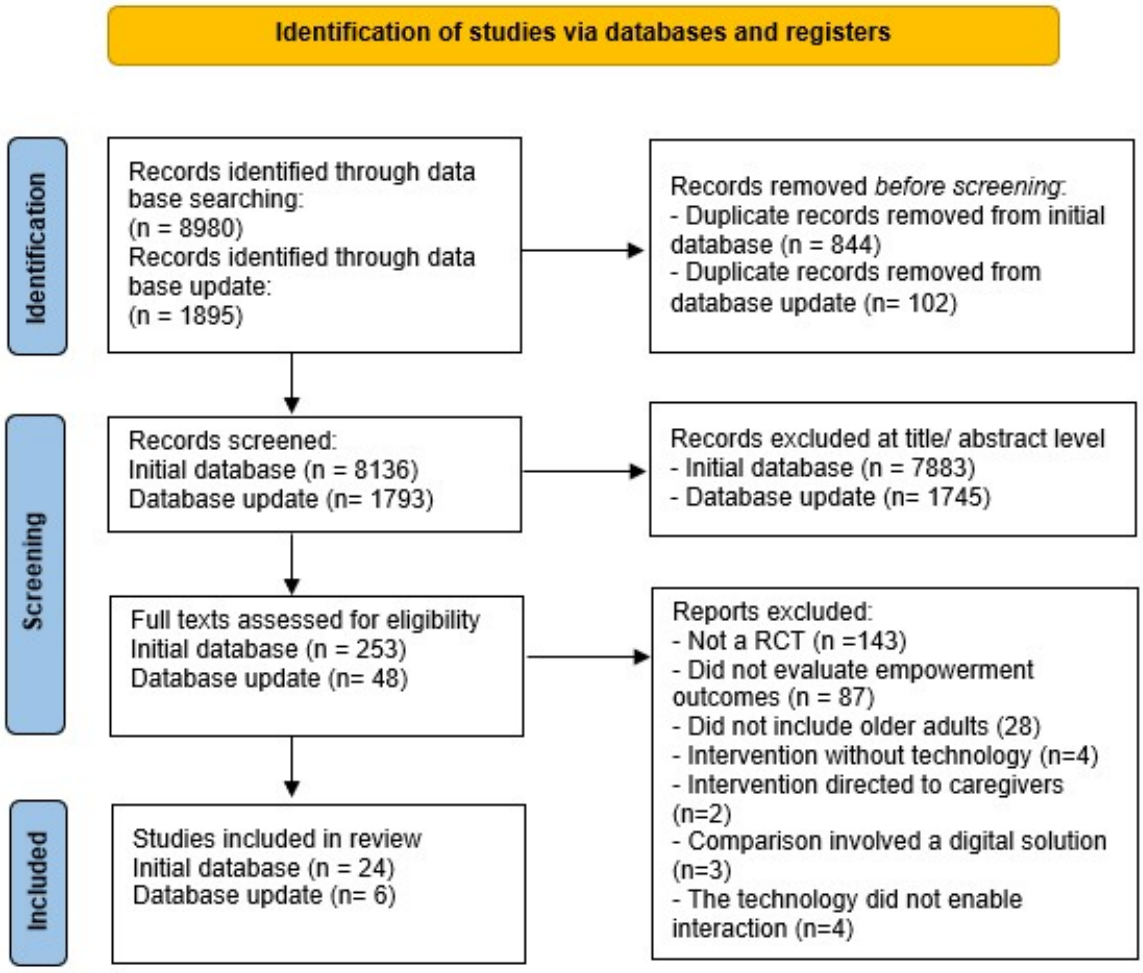

Results

Study Details

A total of 10,875 references were found, of which 946 (8.7%) were duplicates. The remaining 9929 (91.3%) references were screened by title and abstract, and 301 (3%) full-text manuscripts were retrieved and read. Of these, 30 (10%) manuscripts were included in the review (Figure 1). The included manuscripts assessed the impact of using a variety of digital solutions, including the following: websites (n=6, 20%), general mobile apps (n=4, 13.3%), condition-specific mobile apps (n=6, 20%), desktop software (n=1, 3.3%), remote monitoring (n=6, 20%), remote monitoring plus a website (n=1, 3.3%), remote monitoring plus mobile apps (n=5, 16.7%), and remote monitoring, a specific mobile app, plus a website (n=1, 3.3%).

Quality and Certainty of Evidence

The studies scored between 3 and 8 on the PEDro scale. Most studies scored 6-8 (n=20, 66.7%), suggesting good quality; 9 (30%) studies scored between 4 and 5, suggesting fair quality; and 1 (3.3%) study scored 3, suggesting poor quality (Multimedia Appendix 3). The certainty of evidence varied between very low to low and is presented next, along with the results of each meta-analysis (Multimedia Appendix 4).

Impact of Digital Interventions on Empowerment at Postintervention

Data extracted from individual studies are presented in Multimedia Appendix 5.

XSL•FO
RenderX

Patients' Behaviors

Four studies [32-35] assessed the effect of digital solutions on behavior. One study [34] with fair quality compared a digital solution to usual care for patients with chronic obstructive pulmonary disease (COPD) and reported no between-group differences in self-management behavior. Two studies (one with good and the other with fair methodological quality) found the group receiving the digital intervention to report increased self-care behaviors when compared to usual care [32] and to a large written poster [35]. The other study [33], with fair quality, reported no differences between the group receiving the digital intervention and the no-intervention group for perceived participation in enjoyable activities. No meta-analysis was possible due to the unavailability of needed data.

Support by Others

Six studies [35-40] investigated the impact of digital interventions on support by others. Of these, 2 (33.3%) studies (with fair and good methodological quality) compared digital interventions against no intervention: one reported no between-group differences [39], while the other reported that the digital intervention significantly increased social support compared to no intervention [38]. Another study [37], with fair quality, reported no differences in social support between the group receiving the digital intervention and the group receiving usual care. Two studies (fair and good quality) [35,36] used a digital intervention plus a face-to-face component: one reported no significant differences compared to no intervention [36], while the other reported a significant improvement in social support compared to a written large poster [35]. Another study [40] of good quality reported a significant difference at 12 months between the digital intervention and a paper-based health education program, favoring the latter (medium ES).

Patients' Knowledge

Seven studies [32,37,41-45] investigated the impact of digital technologies on patients' knowledge. Of these, 3 (42.9%) studies [32,44,45], 2 (66.7%) of good quality and 1 (33.3%) of fair quality, compared digital solutions against usual care; another 2 (28.6%) studies [41,42], both of fair quality, compared digital solutions against education. There was very low certainty of evidence of a medium ES favoring the experimental group (ES=0.40, 95% CI 0.07-0.73, I^2=79%; Figure 2). The remaining 2 (28.6%) studies [37,43], not included in the meta-analysis, were of fair quality and reported digital solutions to have a greater impact on knowledge compared to usual care.

Figure 2. Plot of the meta-analysis of the standardized mean differences for patients' knowledge (comparison: digital solutions versus usual care or education).

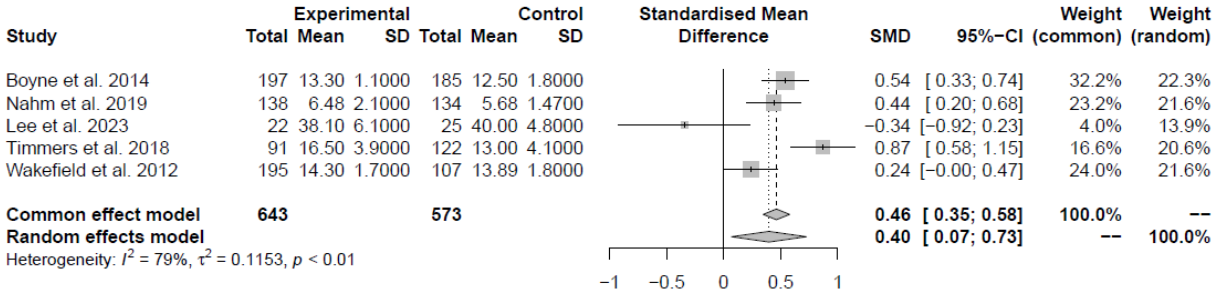

Patients' Capacities

Twenty-five studies investigated the impact of technology on older adults' capacities. A meta-analysis of 5 (20%) studies (n=1, 20%, study with poor, n=1, 20%, study with fair, and 3, 60%, studies with good quality) [39,46-49] that compared digital solutions against no intervention found, with low certainty of evidence, no between-group differences (d=0.13, 95% CI –0.02 to 0.29, I^2=0%; Figure 3). A meta-analysis of 13 (52%) studies (n=10, 76.9%, studies with good and n=3, 23.1%, studies with fair quality) [32,34,40,44,45,50-57] comparing digital solutions against usual care indicated, with low certainty of evidence, a medium ES favoring the digital intervention group (d=0.29, 95% CI 0.07-0.52, I^2=79%; Figure 4). Sensitivity analysis including the 10 (76.9%) studies with good quality did not affect the results (d=0.27, 95% CI 0.02-0.53, I^2=81%).

A meta-analysis of 4 (16%) studies [51,58-60] with good quality that compared digital interventions to face-to-face interventions suggested, with very low certainty of evidence, no between-group differences (d=0.97, 95% CI –0.62 to 2.56, I^2=97%; Figure 5).

Of the 4 (16%) studies [35-37,61] not included in any of the meta-analyses, 3 (75%) of fair-to-good quality reported no between-group differences when the digital solution was compared against usual care/phone calls [37,61] or combined with a face-to-face component and compared against no intervention [36]. The fourth study [35] presented fair quality and reported a significant difference favoring the group that received a digital solution when compared to face-to-face education.

XSL•FO
RenderX

Figure 3. Plot of the meta-analysis of the standardized mean differences for patients' capacities (comparison: digital solutions versus no intervention).

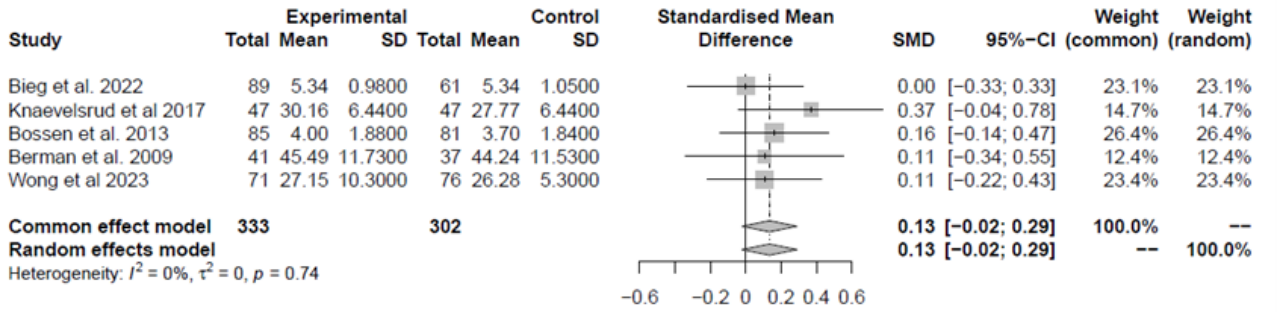

Figure 4. Plot of the meta-analysis of the standardized mean differences for patients' capacities (comparison: digital solutions versus usual care).

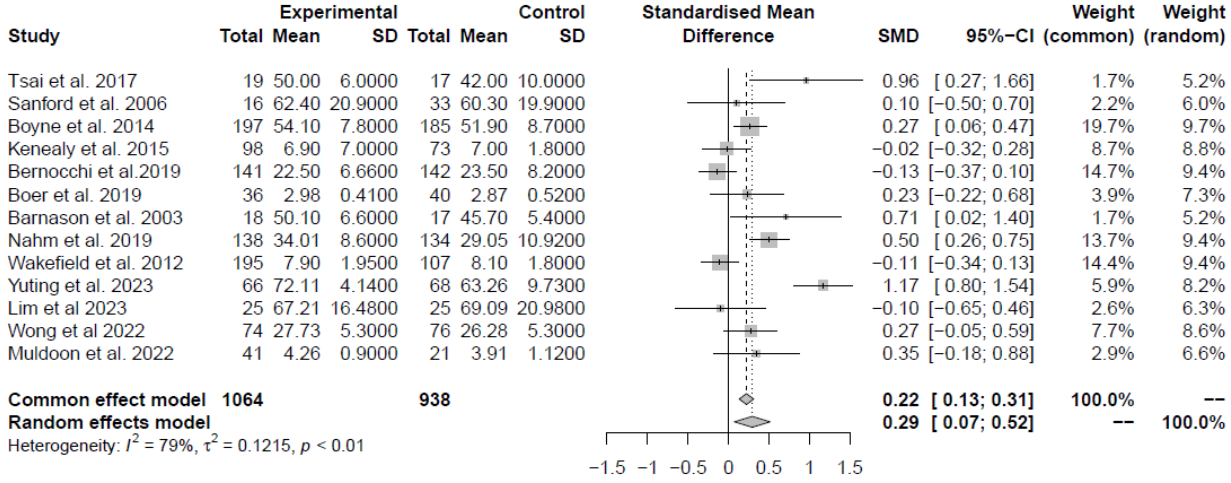

Figure 5. Plot of the meta-analysis of the standardized mean differences for patients' capacities (comparison: digital interventions to face-to-face interventions).

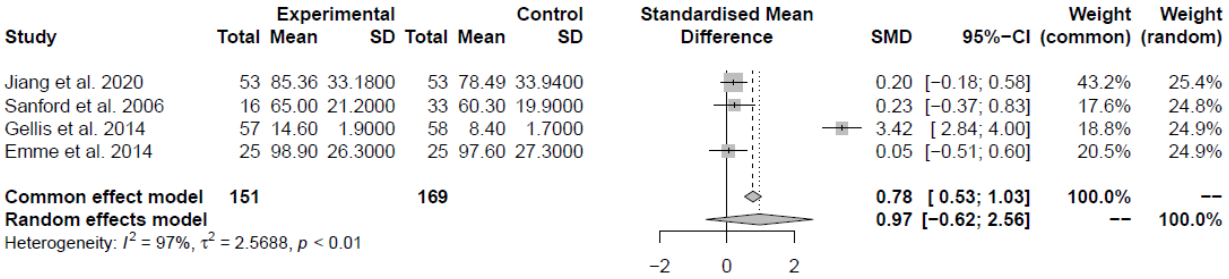

Impact of Digital Interventions on Empowerment at Follow-Up

Patients' Behaviors

Only 1 (3.3%) study [32] assessed the effect of technology on behavior at follow-up (6 and 12 months) and reported that the group receiving the digital intervention reported increased self-care behaviors when compared to usual care.

Support by Others

Two studies [38,39] investigated the impact of digital interventions against no intervention on support by others at up to 12 months' follow-up and reported no between-group differences.

Patients' Knowledge

Three studies investigated the impact of digital interventions against usual care on patients' knowledge at 3 [43], 4 [45], and up to 12 [32] months' follow-up. Two studies [32,43] reported a significant difference in favor of the experimental group, while the third study [45] reported no between-group differences.

XSL•FO
RenderX

Patients' Capacities

Two studies investigated the effect of digital solutions at 12 months' follow-up compared to no intervention: one reported no between-group differences [39], while the other reported significant improvements in self-efficacy in the intervention group [48]. Two studies investigated the effect of digital solutions against face-to-face intervention and reported no between-group differences at 3 [59] and 6 [58] months' follow-up. Three studies investigated the impact of digital solutions against usual care: 2 (66.7%) reported no between-group differences at 4 [45] and 12 [32] months' follow-up, but 1 (33.3%) [53] reported increased self-efficacy in the digital solution group at 3 months' follow-up.

Discussion

Principal Findings

This review synthesized existing evidence on the effectiveness of using integrated digital solutions for empowering older adults. Results suggest with very-low-to-low certainty of evidence that using digital solutions results in increased knowledge and increased capacities when compared to usual care and impacts capacities to an extent similar to face-to-face interventions. Interestingly, results also suggest, with low certainty of evidence, that there are no differences between using digital solutions and no intervention for improving capacities.

Comparison With Prior Work

Our results show an apparent contradiction as they suggest that digital solutions are better than usual care and equal to face-to-face interventions at improving patients' capacities but equal to no intervention. This might be explained by the characteristics of usual care, which, in the studies included, seems to equate to minimal care delivered over 1 or few sessions. As previously reported [62,63], usual care is generally not described in sufficient detail, hindering the interpretation of results. In addition, differences in the degree of interaction and support provided by the digital solutions, in the training given to individuals before the trial, in the amount of interaction with health professionals using the digital solutions, and in the personal and clinical characteristics of the participants might contribute to the differences found. Furthermore, the very-low-to-low certainty of the evidence, the wide CIs for the estimate of the effect, and the high heterogeneity in the usual-care meta-analyses suggest that further research is likely to impact the conclusions of the comparison of digital interventions against no intervention and usual care. Worth highlighting are the 2 studies that compared digital solutions against face-to-face interventions at follow-up and that confirmed the postintervention results by showing no between-group differences. Most studies within the capacity domain assessed self-efficacy (ie, people's beliefs in their own capabilities), which is reduced in older adults receiving care [64] and is an important determinant of health-related behavior change [65].

Digital interventions seem to improve older adults' knowledge to a higher extent than usual care, which, as previously reported, was not detailed in the included studies. The instruments used to assess acquired knowledge are diverse, but most relate to knowledge regarding a specific clinical condition. Other systematic reviews have reported that digital solutions can improve one's knowledge, for example, related to diabetes [66], and are better than usual care for eHealth literacy [67]. A previous study also reported that using digital solutions for seeking health-related information (eg, on specific diseases, medication, symptoms, and health promotion) is common among older adults [21], potentially facilitating adherence to digital interventions that promote knowledge acquisition.

This systematic review with meta-analysis results carefully suggests that the choice of using a digital solution or not might be left to individual preferences and that technology can be used to reach more individuals who would not have access to face-to-face empowerment strategies. Other studies have identified patients' needs and preferences as important aspects to consider when deciding between in-person health care delivery or remote health care delivery [68,69]. In addition, the decision on whether to use a digital solution might depend on the domain of empowerment being specifically targeted as relevant for a particular individual, potentially choosing digital solutions to target knowledge and capacities and face-to-face interventions for other domains. Choosing a mixed approach where some domains of empowerment are targeted with face-to-face interventions, while others are targeted with digital interventions might also be an option. In addition, issues of security and privacy should be considered when choosing digital solutions for older adults, as they often have lower levels of digital literacy and consequently are more exposed to privacy violations, frauds, scams, and phishing attacks [70,71]. Developing digital solutions that have user-friendly interfaces and implementing privacy by design (ie, incorporating privacy features into the design of digital solutions, such as data minimization principles) are essential. Additionally, education and awareness are a means to minimize the risk and to empower older adults with the knowledge and skills to identify online risks [70].

The similarity of results found in this review for the domains of patients' knowledge and capacities might suggest that increased knowledge might promote increased capacities by providing older adults with more knowledge about their health, resources, and management strategies. The study of causality was beyond the scope of this review. This result, rather, identifies a potential topic of study for future studies.

Regarding patients' behaviors and support by others, it was not possible to aggregate the evidence in a meta-analysis due to insufficient studies with a similar comparison. The qualitative analysis of the studies included in our systematic review identified conflicting results regarding the impact of digital interventions. The factors previously identified as potential sources of heterogeneity might also contribute to the conflicting results.

Future studies should detail the intervention administered to the control group, particularly if it is usual care, as well as detail the functionalities of the digital solution provided, including training, interaction with professionals, and adherence to the intervention, as these factors might have an impact on the

XSL•FO
RenderX

effectiveness of the digital intervention to facilitate empowerment. Future systematic reviews and meta-analyses should attempt to identify sources of heterogeneity that might impact the results, including the type of digital solution, the amount of support given to participants, participants' characteristics (age, education, digital literacy), the length of the intervention, and subdomains within each domain of empowerment.

Limitations

The studies included are highly diverse in terms of the type of digital solutions, health conditions, participants' characteristics, and instruments used to assess empowerment. All these aspects are reflected in the high heterogeneity found and undermine the confidence in the results of the meta-analysis. We aimed to run a subgroup meta-analysis considering the type of technology and users' characteristics, but the small number of studies precluded this. For some comparisons, a meta-analysis was not possible due to the reduced number of studies with a similar comparison. This might suggest that empowerment is more researched when related to health domains. Studies that were not based on empowerment theory were not included.

Conclusion

There is very-low-to-low certainty of evidence that using integrated digital solutions results in increased knowledge and increased capacities (mainly self-efficacy) when compared to usual care and that it impacts capacities to an extent similar to face-to-face interventions at postintervention. Interestingly, results also suggest, with low certainty of evidence, that there are no differences between using digital solutions and no intervention for improving capacities. Included studies and technologies are diverse, and the meta-analysis showed high heterogeneity, which limits the confidence in the results and suggests that further research might affect the conclusions of this review. The limited number of studies precluded a subgroup analysis considering the type of technology and user characteristics.

Acknowledgments

This work was supported by the SHAPES (Smart and Health Ageing through People Engaging in Supportive Systems) project funded by the Horizon 2020 Framework Program of the European Union for Research Innovation (grant agreement 857159 -SHAPES – H2020 – SC1-FA-DTS – 2018-2020).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 37 KB - jmir_v27i1e54466_app1.docx]

Multimedia Appendix 2
Description and examples of integrated digital solutions.
[DOCX File , 42 KB - jmir_v27i1e54466_app2.docx]

Multimedia Appendix 3
Search strategy with keywords.
[DOCX File , 15 KB - jmir_v27i1e54466_app3.docx]

Multimedia Appendix 4
Grading of Recommendations Assessment, Development, and Evaluation (GRADE) summary.
[DOCX File , 19 KB - jmir_v27i1e54466_app4.docx]

Multimedia Appendix 5
Methodological quality assessment and extracted data from individual studies.
[DOCX File , 55 KB - jmir_v27i1e54466_app5.docx]

References

1. Tsubouchi Y, Yorozuya K, Tainosyo A, Naito Y. A conceptual analysis of older adults' empowerment in contemporary Japanese culture. BMC Geriatr 2021 Dec 01;21(1):672 [FREE Full text] [doi: 10.1186/s12877-021-02631-x] [Medline: 34852766]
2. Acuña Mora M, Sparud-Lundin C, Moons P, Bratt E. Definitions, instruments and correlates of patient empowerment: a descriptive review. Patient Educ Couns 2022 Feb;105(2):346-355 [FREE Full text] [doi: 10.1016/j.pec.2021.06.014] [Medline: 34140196]

XSL•FO
RenderX

3. Zimmerman MA. Psychological empowerment: issues and illustrations. Am J of Comm Psychol 1995 Oct;23(5):581-599. [doi: 10.1007/bf02506983]
4. Gibson CH. A concept analysis of empowerment. J Adv Nurs 1991 Mar;16(3):354-361. [doi: 10.1111/j.1365-2648.1991.tb01660.x] [Medline: 2037742]
5. Beales S. Expert group meeting on “promoting people’s empowerment in achieving poverty eradication, social integration and productive and decent work for all.” Empowerment and older people: enhancing capabilities in an ageing world. United Nations. 2012. URL: https://www.un.org/esa/socdev/egms/docs/2012/SylviaBeales.pdf [accessed 2024-04-18]
6. Demiris G. Independence and shared decision making: the role of smart home technology in empowering older adults. 2009 Presented at: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009); September 3-6, 2009; Minneapolis, MN p. 6432-6436. [doi: 10.1109/iembs.2009.5333789]
7. Drazic I, Schermuly CC, Büsch V. Empowered to stay active: psychological empowerment, retirement timing, and later life work. J Adult Dev 2023 May 25:1-18 [FREE Full text] [doi: 10.1007/s10804-023-09453-8] [Medline: 37361380]
8. Shearer NBC, Fleury J, Ward KA, O'Brien AM. Empowerment interventions for older adults. West J Nurs Res 2012 Feb 30;34(1):24-51. [doi: 10.1177/0193945910377887] [Medline: 20675621]
9. Tsubouchi Y, Tainosho A, Shimomura K, Kou M, Yorozuya K, Nakashima D, et al. Relationship between empowerment and functioning and disability in older Japanese patients: a covariance structure analysis. Healthcare (Basel) 2022 Dec 23;11(1):44 [FREE Full text] [doi: 10.3390/healthcare11010044] [Medline: 36611504]
10. Haber D. Empowering older adults. In: Health Promotion and Aging. New York, NY: Springer Publishing Company; 2019.
11. Pan American Health Organization and International Telecommunication Union. The Role of Digital Technologies in Aging and Health. Washington, DC: Pan American Health Organization and International Telecommunication Union; 2023.
12. Czaja SJ, Boot WR, Charness N, Rogers WA. Designing for Older Adults. Boca Raton, FL: CRC Press; 2019.
13. Hou B, Li Y, Wang H. Internet use and health status among older adults: the mediating role of social participation. Front Public Health 2022 Nov 25;10:1072398 [FREE Full text] [doi: 10.3389/fpubh.2022.1072398] [Medline: 36504989]
14. Anderson M, Perrin A. Adoption climbs among older Americans. Pew Research Center. 2017. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2024-10-18]
15. Liao S, Chou T, Huang C. Revisiting the development trajectory of the digital divide: a main path analysis approach. Technol Forecast Soc Change 2022 Jun;179:121607. [doi: 10.1016/j.techfore.2022.121607]
16. Wilson-Menzfeld G, Brittain K. Digital exclusion in later life: a narrative review. In: Vulnerable People and Digital Inclusion. Cham: Springer International Publishing; 2022:169-188.
17. Lu X, Yao Y, Jin Y. Digital exclusion and functional dependence in older people: findings from five longitudinal cohort studies. EClinicalMedicine 2022 Dec;54:101708 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101708] [Medline: 36353265]
18. Bruggers CS, Altizer RA, Kessler RR, Caldwell CB, Coppersmith K, Warner L, et al. Patient-empowerment interactive technologies. Sci Transl Med 2012 Sep 19;4(152):152ps16. [doi: 10.1126/scitranslmed.3004009] [Medline: 22993292]
19. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
20. Kebede AS, Ozolins L, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 07;24(12):e40192 [FREE Full text] [doi: 10.2196/40192] [Medline: 36477006]
21. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Pekonen A, Eloranta S, Stolt M, Virolainen P, Leino-Kilpi H. Measuring patient empowerment - a systematic review. Patient Educ Couns 2020 Apr;103(4):777-787. [doi: 10.1016/j.pec.2019.10.019] [Medline: 31767243]
24. Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. PLoS One 2017 Oct;12(10):e0186458 [FREE Full text] [doi: 10.1371/journal.pone.0186458] [Medline: 29040335]
25. Nwagwu W, Onyancha O. Global trends in eHealth research: analysis and visualization of author and indexer-supplied keywords. bioRxiv Preprint posted online 2012. [doi: 10.1101/2020.11.26.399881]
26. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 6.2. London, UK: Cochrane Training; 2021.
27. de Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. Aust J Physiother 2009;55(2):129-133. [doi: 10.1016/s0004-9514(09)70043-1] [Medline: 19463084]
28. Gonzalez GZ, Moseley AM, Maher CG, Nascimento DP, Costa LDCM, Costa LO. Methodologic quality and statistical reporting of physical therapy randomized controlled trials relevant to musculoskeletal conditions. Arch Phys Med Rehabil 2018 Jan;99(1):129-136. [doi: 10.1016/j.apmr.2017.08.485] [Medline: 28962828]
29. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]

XSL•FO
RenderX

30. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE Handbook. 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2024-12-21]
31. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
32. Boyne JJJ, Vrijhoef HJM, Spreeuwenberg M, De Weerd G, Kragten J, Gorgels APM, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
33. Simbrig IM, van der Weegen S, Piazolo F, Kofler M, Sturm N, Hvalič-Touzery S. Impact of using aging-in-place technologies on quality of life: results from a randomized controlled trial in four European countries. Gerontechnology 2022 Jan 01;21(1):1-21. [doi: 10.4017/gt.2023.21.1.792.02]
34. Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, et al. A smart mobile health tool versus a paper action plan to support self-management of chronic obstructive pulmonary disease exacerbations: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 09;7(10):e14408 [FREE Full text] [doi: 10.2196/14408] [Medline: 31599729]
35. Jung H, Lee J. The impact of community-based eHealth self-management intervention among elderly living alone with hypertension. J Telemed Telecare 2016 Jul 09;23(1):167-173. [doi: 10.1177/1357633x15621467]
36. Fanning J, Brooks AK, Hsieh KL, Kershner K, Furlipa J, Nicklas BJ, et al. The effects of a pain management-focused mobile health behavior intervention on older adults' self-efficacy, satisfaction with functioning, and quality of life: a randomized pilot trial. Int J Behav Med 2022 Apr 20;29(2):240-246 [FREE Full text] [doi: 10.1007/s12529-021-10003-3] [Medline: 34018138]
37. Li X, Li T, Chen J, Xie Y, An X, Lv Y, et al. A WeChat-based self-management intervention for community middle-aged and elderly adults with hypertension in Guangzhou, China: a cluster-randomized controlled trial. Int J Environ Res Public Health 2019 Oct 23;16(21):4058 [FREE Full text] [doi: 10.3390/ijerph16214058] [Medline: 31652688]
38. Gustafson DH, Kornfield R, Mares M, Johnston DC, Cody OJ, Yang EF, et al. Effect of an eHealth intervention on older adults' quality of life and health-related outcomes: a randomized clinical trial. J Gen Intern Med 2022 Feb 07;37(3):521-530 [FREE Full text] [doi: 10.1007/s11606-021-06888-1] [Medline: 34100234]
39. Bieg T, Gerdenitsch C, Schwaninger I, Kern BMJ, Frauenberger C. Evaluating active and assisted living technologies: critical methodological reflections based on a longitudinal randomized controlled trial. Comput Hum Behav 2022 Aug;133:107249. [doi: 10.1016/j.chb.2022.107249]
40. Lim ML, Tran M, van Schooten KS, Radford KA, O'Dea B, Baldwin P, et al. A self-guided online cognitive behavioural therapy to reduce fear of falling in older people: a randomised controlled trial. Int J Behav Med 2023 Jun 02;30(3):455-462. [doi: 10.1007/s12529-022-10105-6] [Medline: 35655058]
41. Lee KH, Choi YY, Jung ES. Effectiveness of an oral health education programme using a mobile application for older adults: a randomised clinical trial. Gerodontology 2023 Mar 23;40(1):47-55. [doi: 10.1111/ger.12616] [Medline: 34951060]
42. Timmers T, Janssen L, Pronk Y, van der Zwaard BC, Koëter S, van Oostveen D, et al. Assessing the the efficacy of an educational smartphone or tablet app with subdivided and interactive content to increase patients' medical knowledge: randomized controlled trial. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e10742 [FREE Full text] [doi: 10.2196/10742] [Medline: 30578185]
43. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GY. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: mAF app trial. Am J Med 2017 Dec;130(12):1388-1396.e6 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
44. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hills SL, et al. Outcomes of a home telehealth intervention for patients with diabetes and hypertension. Telemed J E Health 2012 Oct;18(8):575-579. [doi: 10.1089/tmj.2011.0237] [Medline: 22873700]
45. Nahm E, Zhu S, Bellantoni M, Keldsen L, Russomanno V, Rietschel M, et al. The effects of a theory-based patient portal e-learning program for older adults with chronic illnesses. Telemed J E Health 2019 Oct;25(10):940-951 [FREE Full text] [doi: 10.1089/tmj.2018.0184] [Medline: 30431393]
46. Wong AKC, Bayuo J, Wong FKY, Chow KKS, Wong SM, Lau ACK. The synergistic effect of nurse proactive phone calls with an mHealth app program on sustaining app usage: 3-arm randomized controlled trial. J Med Internet Res 2023 May 01;25:e43678 [FREE Full text] [doi: 10.2196/43678] [Medline: 37126378]
47. Berman RL, Iris MA, Bode R, Drengenberg C. The effectiveness of an online mind-body intervention for older adults with chronic pain. J Pain 2009 Jan;10(1):68-79 [FREE Full text] [doi: 10.1016/j.jpain.2008.07.006] [Medline: 18774342]
48. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
49. Knaevelsrud C, Böttche M, Pietrzak RH, Freyberger HJ, Kuwert P. Efficacy and feasibility of a therapist-guided internet-based intervention for older persons with childhood traumatization: a randomized controlled trial. Am J Geriatr Psychiatry 2017 Aug;25(8):878-888. [doi: 10.1016/j.jagp.2017.02.024] [Medline: 28365000]
50. Tsai LLY, McNamara RJ, Moddel C, Alison JA, McKenzie DK, McKeough ZJ. Home-based telerehabilitation via real-time videoconferencing improves endurance exercise capacity in patients with COPD: the randomized controlled TeleR Study. Respirology 2017 May 19;22(4):699-707 [FREE Full text] [doi: 10.1111/resp.12966] [Medline: 27992099]

XSL•FO
RenderX

51. Sanford JA, Griffiths PC, Richardson P, Hargraves K, Butterfield T, Hoenig H. The effects of in-home rehabilitation on task self-efficacy in mobility-impaired adults: a randomized clinical trial. J Am Geriatr Soc 2006 Nov;54(11):1641-1648. [doi: 10.1111/j.1532-5415.2006.00913.x] [Medline: 17087689]
52. Kenealy TW, Parsons MJG, Rouse APB, Doughty RN, Sheridan NF, Hindmarsh JKH, et al. Telecare for diabetes, CHF or COPD: effect on quality of life, hospital use and costs. A randomised controlled trial and qualitative evaluation. PLoS One 2015 Mar 13;10(3):e0116188 [FREE Full text] [doi: 10.1371/journal.pone.0116188] [Medline: 25768023]
53. Barnason S, Zimmerman L, Nieveen J, Schmaderer M, Carranza B, Reilly S. Impact of a home communication intervention for coronary artery bypass graft patients with ischemic heart failure on self-efficacy, coronary disease risk factor modification, and functioning. Heart Lung 2003 May;32(3):147-158. [doi: 10.1016/s0147-9563(03)00036-0] [Medline: 12827099]
54. Bernocchi P, Giordano A, Pintavalle G, Galli T, Ballini Spoglia E, Baratti D, et al. Feasibility and clinical efficacy of a multidisciplinary home-telehealth program to prevent falls in older adults: a randomized controlled trial. J Am Med Dir Assoc 2019 Mar;20(3):340-346. [doi: 10.1016/j.jamda.2018.09.003] [Medline: 30366759]
55. Wong AKC, Wong FKY, Chow KKS, Wong SM, Bayuo J, Ho AKY. Effect of a mobile health application with nurse support on quality of life among community-dwelling older adults in Hong Kong: a randomized clinical trial. JAMA Netw Open 2022 Nov 01;5(11):e2241137 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.41137] [Medline: 36350651]
56. Yuting Z, Xiaodong T, Qun W. Effectiveness of a mHealth intervention on hypertension control in a low-resource rural setting: a randomized clinical trial. Front Public Health 2023 Mar 1;11:1049396 [FREE Full text] [doi: 10.3389/fpubh.2023.1049396] [Medline: 36935728]
57. Muldoon MF, Einhorn J, Yabes JG, Burton D, Irizarry T, Basse J, et al. Randomized feasibility trial of a digital intervention for hypertension self-management. J Hum Hypertens 2022 Aug 08;36(8):718-725. [doi: 10.1038/s41371-021-00574-9] [Medline: 34239050]
58. Jiang Y, Liu F, Guo J, Sun P, Chen Z, Li J, et al. Evaluating an intervention program using WeChat for patients with chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2020 Apr 21;22(4):e17089 [FREE Full text] [doi: 10.2196/17089] [Medline: 32314971]
59. Emme C, Mortensen EL, Rydahl-Hansen S, Østergaard B, Svarre Jakobsen A, Schou L, et al. The impact of virtual admission on self-efficacy in patients with chronic obstructive pulmonary disease - a randomised clinical trial. J Clin Nurs 2014 Nov 30;23(21-22):3124-3137. [doi: 10.1111/jocn.12553] [Medline: 24476457]
60. Gellis ZD, Kenaley BL, Ten Have T. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the Integrated Telehealth Education and Activation of Mood (I-TEAM) study. J Am Geriatr Soc 2014 May;62(5):889-895. [doi: 10.1111/jgs.12776] [Medline: 24655228]
61. Wong AKC, Wong FKY, Chow KKS, Wong SM, Lee PH. Effect of a telecare case management program for older adults who are homebound during the COVID-19 pandemic. JAMA Netw Open 2021 Sep 01;4(9):e2123453 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.23453] [Medline: 34499135]
62. Smelt AF, van der Weele GM, Blom JW, Gussekloo J, Assendelft WJ. How usual is usual care in pragmatic intervention studies in primary care? An overview of recent trials. Br J Gen Pract 2010 Jul 01;60(576):e305-e318. [doi: 10.3399/bjgp10x514819]
63. Paci M, Risaliti F, Pellicciari L. Reporting of "usual care" as the control group in randomized clinical trials of physiotherapy interventions for multiple sclerosis is poor: a systematic review. Neurol Sci 2022 Sep 31;43(9):5207-5216. [doi: 10.1007/s10072-022-06167-9] [Medline: 35639218]
64. Whitehall L, Rush R, Górska S, Forsyth K. The general self-efficacy of older adults receiving care: a systematic review and meta-analysis. Gerontologist 2021 Aug 13;61(6):e302-e317 [FREE Full text] [doi: 10.1093/geront/gnaa036] [Medline: 32373938]
65. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WMP, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
66. Nkhoma DE, Soko CJ, Bowrin P, Manga YB, Greenfield D, Househ M, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021 Oct;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
67. Verweel L, Newman A, Michaelchuk W, Packham T, Goldstein R, Brooks D. The effect of digital interventions on related health literacy and skills for individuals living with chronic diseases: a systematic review and meta-analysis. Int J Med Inform 2023 Sep;177:105114. [doi: 10.1016/j.ijmedinf.2023.105114] [Medline: 37329765]
68. Ladin K, Porteny T, Perugini JM, Gonzales KM, Aufort KE, Levine SK, et al. Perceptions of telehealth vs in-person visits among older adults with advanced kidney disease, care partners, and clinicians. JAMA Netw Open 2021 Dec 01;4(12):e2137193 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.37193] [Medline: 34870680]
69. Tomczyk ?, Mascia ML, Gierszewski D, Walker C. Barriers to digital inclusion among older people: a intergenerational reflection on the need to develop digital competences for the group with the highest level of digital exclusion. IJTEI 2023 Jun 01;9(1):5-26. [doi: 10.24310/innoeduca.2023.v9i1.16433]
70. Quan‐Haase A, Ho D. Online privacy concerns and privacy protection strategies among older adults in East York, Canada. Assoc Info Sci Technol 2020 May 08;71(9):1089-1102. [doi: 10.1002/asi.24364]

XSL•FO
RenderX

71. Ellefsen J, Chen W. Privacy and data security in everyday online services for older adults. 2022 Presented at: DSAI '22: 10th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion; August 31-September 2, 2022; Lisbon, Portugal p. 203-207. [doi: 10.1145/3563137.3563149]

Abbreviations

ES: effect size
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 10.11.23; peer-reviewed by S Zawada, X Yan; comments to author 29.02.24; revised version received 17.04.24; accepted 01.10.24; published 09.01.25.

Please cite as:
Martins AI, Ribeiro Ó, Santinha G, Silva T, Rocha NP, Silva AG
Effectiveness of Integrated Digital Solutions to Empower Older Adults in Aspects Related to Their Health: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e54466
URL: https://www.jmir.org/2025/1/e54466
doi:10.2196/54466
PMID:

©Ana Isabel Martins, Óscar Ribeiro, Gonçalo Santinha, Telmo Silva, Nelson P Rocha, Anabela G Silva. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Reality Interventions for Older Adults With Mild Cognitive Impairment: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Qin Yang[1], BEng; Liuxin Zhang[1], BFA; Fangyuan Chang[1], BEng, MEng, PhD; Hongyi Yang[1], BEng, MEng; Bin Chen[1], BEng, MEng, PhD; Zhao Liu[1], BEng, MEng, PhD

School of Design, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Zhao Liu, BEng, MEng, PhD
School of Design
Shanghai Jiao Tong University
800 Dongchuan Road
Minhang District
Shanghai, 200240
China
Phone: 86 18801971294
Email: hotlz@sjtu.edu.cn

Abstract

Background: Alzheimer disease is incurable, but it is possible to intervene and slow down the progression of dementia during periods of mild cognitive impairment (MCI) through virtual reality (VR) technology.

Objective: This study aimed to analyze the effects of VR interventions on older adults with MCI. The examined outcomes include cognitive abilities, mood, quality of life, and physical fitness, including general cognitive function, memory performance, attention and information processing speed, executive function, language proficiency, visuospatial abilities, depression, daily mobility of individuals, muscle performance, and gait and balance.

Methods: A total of 4 web-based databases (Web of Science, PubMed, Embase, and Ovid) were searched up to December 30, 2023, for randomized controlled trials assessing the self-reported outcomes of VR-based technology on cognition, mood, quality of life, and physical fitness in older adults (aged ≥55 years) with MCI. Two reviewers independently screened the search results and reference lists of the identified papers and related reviews. Data on the intervention components and delivery and behavioral change techniques used were extracted. A meta-analysis, risk-of-bias sensitivity analysis, and subgroup analysis were performed where appropriate to explore potential moderators. The Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach was used to assess the quality of evidence.

Results: This review analyzed 18 studies involving 722 older adults with MCI. VR was delivered through different immersion levels with VR cognitive training, VR physical training, or VR cognitive-motor dual-task training. VR interventions showed significant improvements in memory (standardized mean difference [SMD] 0.2, 95% CI 0.02-0.38), attention and information processing speed (SMD 0.25, 95% CI 0.06-0.45), and executive function (SMD 0.22, 95% CI 0.02-0.42). VR without therapist involvement improved memory as well as attention and information processing speed. VR cognitive training also resulted in significant improvements in attention and information processing speed in older adults with MCI (SMD 0.31, 95% CI 0.05-0.58). In addition, immersive VR had a significant impact on improving attention and information processing speed (SMD 0.25; 95% CI 0.01-0.50) and executive function (SMD 0.25; 95% CI 0.00-0.50). However, the effects of the intervention were very small in terms of general cognitive function, language proficiency, visuospatial abilities, depression, daily living ability, muscle performance, and gait and balance. Quality of evidence varied, with moderate ratings for certain cognitive functions and low ratings for others, based on the GRADE approach.

Conclusions: VR interventions can improve memory, attention and information processing speed, and executive function in older adults with MCI. The quality of evidence is moderate to low, and further research is needed to confirm these findings and explore additional health-related outcomes.

(J Med Internet Res 2025;27:e59195) doi:10.2196/59195

XSL•FO
RenderX

KEYWORDS

Alzheimer disease; virtual reality; VR; mild cognitive impairment; meta-analysis; health care; cognitive function; memory; attention; executive function; older adults

Introduction

Dementia is a medical condition characterized by a notable decrease in cognitive abilities, which disrupts individuals' ability to perform occupational, familial, or social responsibilities [1] and is currently the seventh leading cause of death worldwide [2]. As the worldwide occurrence of Alzheimer disease (AD) rises, the economic burden on society will also increase significantly, with greater disease severity being linked to higher expenses [3,4]. Mild cognitive impairment (MCI) is a transitional stage of cognitive performance that occurs between normal aging and dementia [5]. Treatments can be implemented to slow down the advancement of dementia during the preclinical phase [6]. Conventional nonpharmacologic lifestyle interventions, including moderate-intensity physical activity and autonomic training, encounter challenges such as inadequate adherence, exorbitant treatment expenses, and inequities in health care accessibility [7,8]. Since there are few medications or dietary therapy that can improve cognitive function or slow MCI progression, nonpharmacological treatments have received attention [9,10].

In nonpharmacological interventions, VR has received a lot of attention. It helps the user to create a real sense of presence and immersion in the virtual world by using multiple sensory stimuli (visual, auditory, tactile, and olfactory) [11,12]. This capability has the potential to augment participants' focus and boost the efficacy of training sessions. For patients with MCI, VR has shown potential in enhancing cognitive capacity and motor function [13-15]. VR technology has mostly been used for evaluating and diagnosing the extent of cognitive impairment, as well as for cognitive training and testing [16]. In the assessment of the degree of cognitive impairment, VR-based virtual supermarkets, virtual fire evacuation drills, and voice interaction based on VR and wearable devices have been used to detect and distinguish between older adults with MCI and healthy older adults [15,17,18]. These methods have also been used to differentiate between persons with MCI and those with dementia [19,20], which help clinicians to detect cognitive impairment at an earlier stage compared with current diagnostic models that address different cognitive deficits. Furthermore, in terms of cognitive testing and cognitive training, a growing number of studies, such as the Virtual Environmental Grocery Store [21] and the Virtual Supermarket [22], have discovered that patients' performance in VR tasks is closely related to their performance in traditional neuropsychological tests. This demonstrates the reliability of VR-based assessment tests compared with traditional neuropsychological tests [23-25]. Furthermore, it has the potential to enhance many cognitive capacities such as general cognitive ability, memory, attention, and executive function [26-28]. VR technology can enhance noncognitive aspects of capability or condition in individuals with MCI. For instance, older adults with MCI may experience a reduction in anxiety by engaging with virtual environments such as virtual nature or tourist attractions [29]. Other studies also indicated that VR-based cognitive training (cooking or wayfinding) could effectively enhance instrumental activities of daily living (IADL) in individuals with MCI and AD [30].

Several trials have investigated the impact of VR on older adults with MCI, but the findings have been inconclusive. For example, studies by Baldimtsi et al [31] and Park et al [32] revealed a significant effect of VR on general cognitive abilities. However, the findings from Park et al [33] showed that a 12-week, culture-based VR training program did not improve general cognitive abilities and did not show significant differences in scores on the Mini-Mental Status Examination (MMSE). Also, it is challenging to make definitive conclusions about the effectiveness of interventions because of variations in the content of the interventions and the way they interact with VR. In a study conducted by Yang et al [28], daily life-based VR training games (making juice, shooting crows, finding the number of fireworks, and memorizing objects in the house) were found to positively affect general cognitive performance. However, in a study by Delbroek et al [34], which involved a combination of VR cognitive and motor training for balance, weight-bearing, memory, attention, and dual-tasking, no significant differences in general cognitive performance were observed.

Now, only 1 comprehensive assessment has examined the effects of VR on the cognitive abilities of older adults with MCI [13]. Nevertheless, it solely concentrated on the efficacy of executive function and failed to provide a thorough examination of cognitive capabilities. A comprehensive assessment found that a VR-based neuropsychological intervention was helpful in enhancing cognitive performance in individuals diagnosed with MCI [35]. However, this review did not address the different VR interventions due to the inclusion of only 1 specific intervention based on VR technology and the lack of analysis of multiple VR technology–based interventions. Currently, there is no systematic review of the VR field, and limited knowledge exists regarding the most effective type or combination of VR solutions for improving cognitive capacities, emotional performance, daily functioning, and physical fitness in older individuals with MCI. A broad overview of the evidence through a more nuanced categorization of VR content, professionals, and system-level VR solutions is required. Therefore, this review aims to assess the effects of VR technology–based interventions on older adults, specifically those with MCI. The outcomes include cognition, mood, quality of life, and physical fitness. Also, it is important to identify the most effective combination of the most effective components (eg, VR intervention content, provider, and level of immersion) for use in clinical and managerial responses to the evidence, as well as policy decisions.

XSL•FO
RenderX

Methods

Systematic Literature Review

Guidance published in PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [36] and the Cochrane Handbook of Systematic Reviews [37] was adhered to. The a priori protocol for the review is published in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42024503488).

Search Strategy

Studies were identified by searching web-based databases with support and consultation provided by institutional librarians. Searches included Web of Science, PubMed, Embase, and Ovid using both Medical Subject Headings and free-text keywords relating to MCI, VR interventions, and outcomes. To include studies reflecting the latest advancements in VR technology and methodologies, the focus was on literature published from January 2013 to December 2023. The year 2013 was chosen to ensure consistency in the technological sophistication and usability standards of VR interventions, as VR technology and its applications in cognitive rehabilitation have rapidly evolved over the past decade. One of the two pre-2013 papers were excluded due to unspecified levels of cognitive impairment, which could have affected the identification of cognitive deficits [38]. The other paper was excluded because cognitive deficits were due to conditions such as stroke or cerebral infarction, to maintain sample specificity [39]. Therefore, the outcomes from January 2013 to December 2023 are listed below. Search strategies for the first 3 databases above and dates on which searches were conducted are listed in Multimedia Appendix 1. Searches of the reference list of the previous review papers and included studies were also conducted.

Eligibility Criteria

Type of Studies

Randomized controlled trials (RCTs) of any design, including parallel-group, crossover, and cluster RCTs published in the English language were included in this review.

Population, Intervention, Comparison, Outcome Framework

This study uses the PICO (Population, Intervention, Comparison, Outcome) framework to structure the research question and guide the systematic review and meta-analysis. The PICO approach is a well-established method for formulating research questions and conducting systematic reviews in health care research [40].

Participants

Studies concerning older adults (aged ≥55 years) with a confirmed diagnosis of MCI by neurologic examination or neuropsychological assessment were included. The age cutoff of 55 years and older was chosen because individuals in this age group are eligible for certain age-related benefits and programs in various contexts, such as early retirement plans, senior discounts, and health interventions targeting midlife transitions. This makes it a practical cutoff for researching populations that can access interventions aimed at promoting active aging. No restrictions were applied to the study design. Studies with healthy older adults; persons with schizophrenia, depression, or Parkinson disease; and participants who did not specify the level of cognitive deficits were excluded.

Intervention

The intervention criterion considered the use of virtual environments, virtual interactive means, and VR components, including VR-based physical training, VR-based cognitive training, VR-based cognitive-motor dual-task training, face-to-face cognitive VR rehabilitation system, and VR programs.

Control Condition

Studies with any type of control group were included (inactive controls include educational programs or no intervention; active controls include traditional rehabilitation or any other type of physical activity, physical-cognitive co-training, or video games without VR components).

Outcomes

We included any psychometric, physical functioning measure that was reasonably unidimensional or multidimensional and the relevant subscales associated with the outcomes in Textbox 1.

XSL•FO
RenderX

Textbox 1. Measurement scales related to the outcomes.

Primary Outcomes

- General cognitive function: the Mini-Mental State Examination, the Montreal Cognitive Assessment, Korean version of the Modified Mental State Examination-Dementia Screening Test, the Cognitive Abilities Screening Instrument, Subcategory VST of CNT 4.0, Montreal Cognitive Assessment and GP Cognitive Assessment (dementia screening tool), and Loewenstein Occupational Therapy Cognitive Assessment – Aging
- Memory performance [41]: the Digit Span Test; Rey Auditory Verbal Learning Test; Wechsler Memory Scale, Third Edition; Multifactorial Memory Questionnaire; Chinese Verbal Learning Test; 20-item version of the Everyday Memory Questionnaire (range: 20-180); and Seoul Verbal Learning Test
- Attention and information processing speed: Trajectory Making Test A and B, Symbol Digit Substitution Test, Attention Matrix Test, and Digit Span Test
- Executive function: Korean version of the Executive Function Performance Test, Symbol Digit Substitution Test, Trajectory Making Test B, Stroop Color and Word Test, Executive Interview 25, and Frontal Assessment Battery

Secondary Outcomes

- Language proficiency: Word Fluency Test (category and letter fluency), Verbal Fluency (phonological fluency; semantic fluency), Battery for Analysis of Aphasic Deficits, and Korean version of the Boston Naming Test
- Visuospatial abilities: Rey-Osterrieth Complex Figure Test, Clock Drawing Test, and Weschsler Adult Intelligence Scale Revised Block Design Test
- Depression: 30-item Geriatric Depression Scale, 15-item Geriatric Depression Scale, and Multidimensional Observation Scale for Elderly Subjects
- The daily mobility of individuals: the Lawton Instrumental Activities of Daily Living scale, the Quality of Life in Alzheimer's Disease scale, and the Multidimensional Observation Scale for Elderly Subjects
- Gait and balance: Gait Speed Test, Timed Up and Go Test, Instrumented Timed Up and Go Test, limit of stability (LOS) values (a subcategory of RM Ingénierie's BIORescue system), and Six Minute Walk Test
- Muscle performance: Hand Gauge Measurement, Arm Flexion Test, and Lower Body Strength using the 30-second Sit-Stand Test

Exclusion Criteria

The exclusion criteria were as follows:

1. Not RCTs: Studies that were not RCTs were excluded.
2. Participants' age ≤55 years: Studies where the participants' average age was 65 years or younger were excluded.
3. Not participants with MCI: Studies that did not include participants diagnosed with MCI were excluded.
4. Cognitive impairment caused by other conditions: Studies where cognitive impairment was attributed to other medical conditions, such as stroke, cerebral infarction, traumatic brain injury, or other neurological disorders, were excluded. This ensures that the cognitive impairment under study is specifically related to MCI and not secondary to other health issues.
5. No specific identification of cognitive impairment: Studies where VR was not used in the intervention group, or VR was used in the control group, were excluded.
6. Inappropriate use of VR: Studies where VR was not used in the intervention group, or VR was used in the control group, were excluded.

Minimum Studies for Analysis

Analyses were conducted only if there were at least 3 studies available for each specific outcome measure. This criterion ensures the reliability and validity of the meta-analytic results.

Study Selection and Data Extraction

Two reviewers (QY and ZL) independently screened the abstracts and full text of the search results. Any disagreements were resolved through discussion. Data were extracted by 2 researchers (HY and LZ) and cross-checked by a third researcher (FC). Data items included participant characteristics, intervention device, intervention means, expertise or guidance, duration of intervention, adherence, and fidelity.

Meta-Analysis

We used RevMan (version 5.3; Cochrane Collaboration) for all analyses. Outcome data were expressed as standardized mean difference (SMD) and were pooled in a pairwise fixed-effects model stratified based on the outcome and type of digital medium. We used the mean (SD) of within-group change from baseline to calculate the SMDs. Per Cohen [42], SMDs <0.2 were classified as small, those between 0.2 and 0.8 were classified as medium, and those >0.8 were classified as large. The SMD was converted to the percentage change in the outcome measure by multiplying the SMD with the SD of the control group of the sufficiently powered study (ie, the trial with the largest sample size).

A sensitivity analysis was conducted for the risk-of-bias assessment. Subgroup analyses were conducted to assess the effect of VR on the different immersion effects used, the different intervention components, and the presence or absence of therapist involvement in the intervention. Heterogeneity across pooled studies was examined using the chi-squared test and I^2 statistics. We did not conduct additional exploratory analyses to explore potential moderators because of an insufficient number of studies. The common study characteristics tables were supplemented by additional summary tables including the risk-of-bias assessment; effect estimates; Grading of Recommendations, Assessment, Development, and

XSL•FO
RenderX

Evaluations (GRADE) analysis; and behavior change taxonomy groupings.

Dealing With Missing Data

RevMan calculator was used to estimate missing SDs based on test statistics reported in 2 studies. In addition, sensitivity analyses were conducted to assess the robustness of the results when missing data were imputed. Imputation methods were based on the assumption that data were missing at random.

Risk-of-Bias and GRADE Assessment

Risk of bias was assessed using the risk of bias tool of the Cochrane Handbook for Systematic Reviews [43]. Quality of evidence for outcomes was assessed according to the 5 GRADE domains, including study limitation (risk of bias), inconsistency, indirectness, imprecision, and publication bias [44,45].

Results

Study Selection and Characteristics

A total of 571 titles and abstracts were screened after excluding duplicates, of which 494 records did not meet the inclusion criteria (Figure 1). The full text of 78 potentially eligible records was read, including 18 studies published between 2017-2023. These studies included 722 (range 17-68) older adults with MCI. In total, 18 studies were RCTs [26-28,31-34,46-56]. Specific study characteristics and participant demographics are summarized in Multimedia Appendix 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. MCI: mild cognitive impairment; RCT: randomized controlled trial; VR: virtual reality.

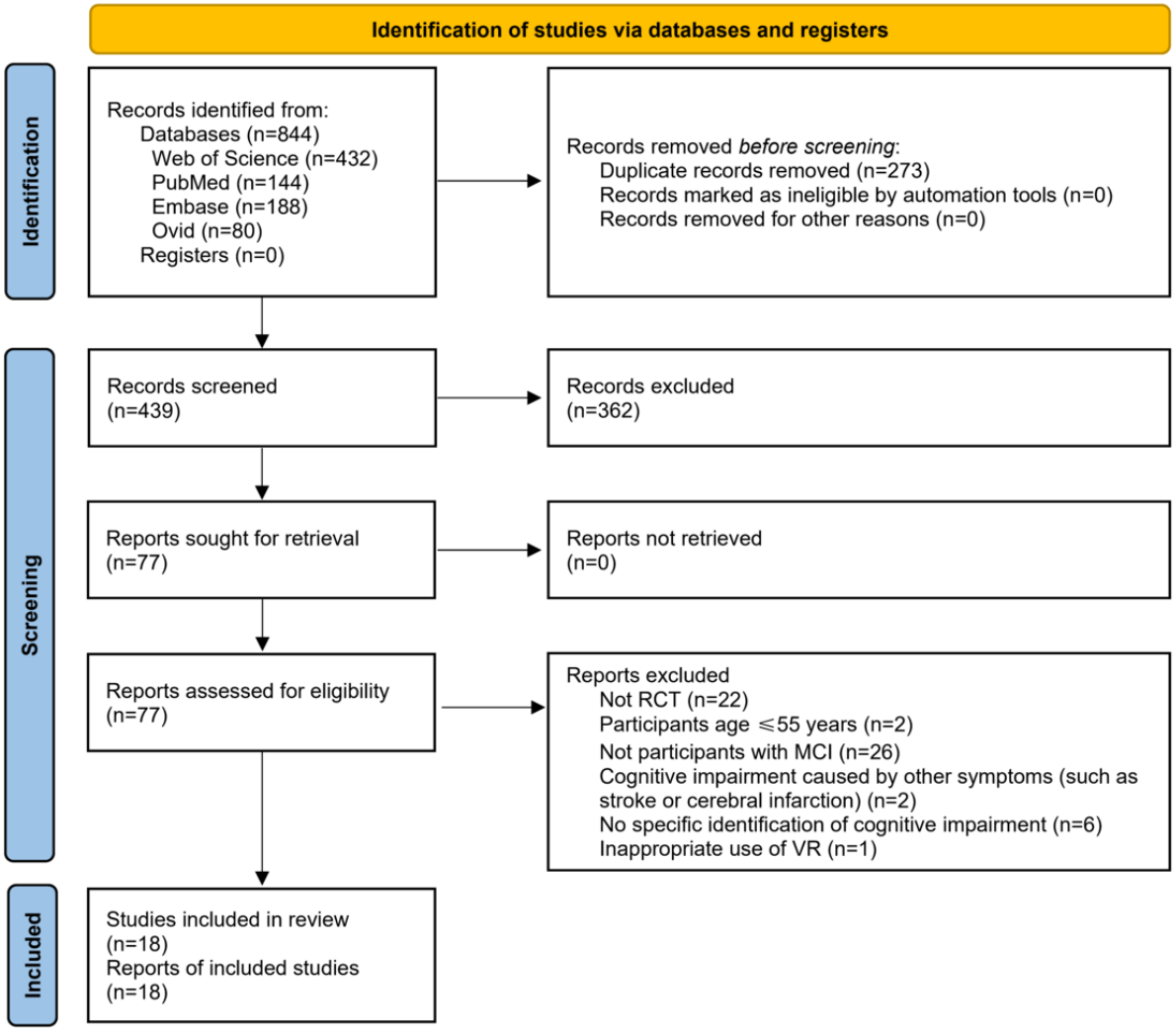

XSL•FO
RenderX

Participants

In total, 18 studies were included, with a total of 8 studies conducted in South Korea; 2 in Hong Kong, China; 2 in Taipei, China; and 6 in mainland China, Greece, Brazil, France, Turkey, and Japan.

The number of participants in the studies ranged from 17 to 68. The median age was 73.15 (IQR 66.07-87.5) years. Most participants in the studies were female (median 63.6%, IQR 16.7%-86.7%).

Nine studies reported participants' years of education, with a median age of 8.56 (IQR 7.20-9.90) years. Ten studies reported participants' baseline MMES scores, with a median score of 26.28 (IQR 25.30-27.4).

Detailed information on the characteristics of the included studies and the demographic data of the study participants are provided in the Multimedia Appendix 2.

Intervention Groups

Nine studies used VR headsets as the medium for VR intervention gear. Among these, 2 studies incorporated additional components, while 7 studies used stand-alone screens or projections. Furthermore, 6 studies were equipped with other components. One study used desktop computers, while 2 studies used tablets. Regarding control and perception, 10 studies used controllers while 4 studies used wearable devices for VR interventions. In terms of cognitive interventions, 15 studies were conducted, with 3 studies focusing on a single cognitive intervention modality and 12 studies using an integrated cognitive intervention system. In 9 of the studies conducted, 2 studies focused solely on motor interventions without involving cognitive interventions, whereas the remaining 7 studies included both motor and cognitive interventions. A total of 14 studies incorporated a gamification element, 9 studies included a feedback component, and 7 studies implemented a progressive difficulty level. Professional input and support were included in all of them. Specifically, 4 studies involved specialists who provided their expertise during the development of the intervention content, while 12 studies had a therapist involved in the intervention throughout its duration. Two studies did not provide relevant descriptions. Regarding the duration of the VR intervention, 3 studies had an intervention period of 4 weeks, 3 studies had an intervention period of 6 weeks, 5 studies had an intervention period of 8 weeks, 6 studies had an intervention period of 12 weeks, and 1 study had an intervention period of 24 weeks. For detailed information on the intervention implementation mechanisms, refer to Multimedia Appendix 2.

Comparison

VR-based interventions were compared with treatment as usual in 2 studies, combined physical and cognitive training group in 3 studies, conventional cognitive rehabilitation group in 3 studies, attention group (health education) in 2 studies, physical exercise group in 2 studies, computer and video games in 1 study, and no intervention in 5 studies.

Risk of Bias

The risk of bias was unclear in 1 study for random sequence generation, 1 study for allocation hiding, and 2 studies for performance bias. A high risk of bias was observed in 8 studies for random sequence generation, 6 studies for allocation concealment, 16 studies for performance bias, and 9 studies for detection bias (Figure 2 [26-28,31-34,46-56]). Due to the nature of the interventions and patient-reported outcome instruments, both detection bias and performance bias were rated high in most studies.

XSL•FO
RenderX

Figure 2. Summary of the risk-of-bias assessment.

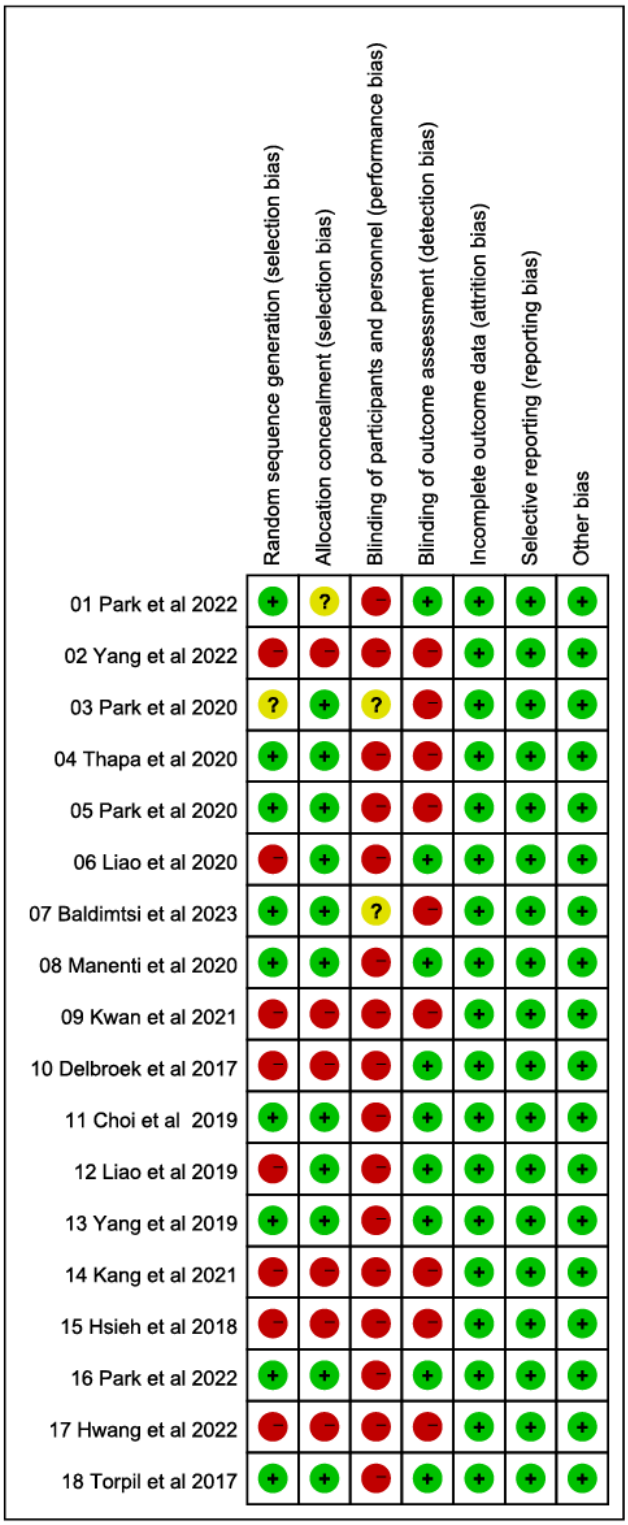

Effects of Interventions

The main comparative results of the study with GRADE ratings are summarized in Multimedia Appendix 3.

Main Outcomes

Primary Outcomes

General Cognitive Ability

A total of 13 studies involving 597 participants reported on the effects of VR technology-based interventions on general cognitive function performance at postintervention time points (range 4-24 weeks) compared with conventional control conditions (Figure 3 [26-28,31-34,47,49,50,52-54]). In

XSL•FO
RenderX

summary, the intervention had a small impact on general cognitive function, but this effect was not statistically significant (SMD 0.09, 95% CI −0.07 to 0.25); there was also no significant heterogeneity (χ^2_{12}=2.51, *P*>.99; I^2=0%). Using the GRADE approach, the quality of evidence was rated moderate because of the high risk of bias in most studies (ie, study limitations).

Figure 3. Forest plot of general cognitive ability: comparison of general cognitive ability improvement at postintervention time points based on virtual reality intervention versus conventional treatment or no intervention control group.

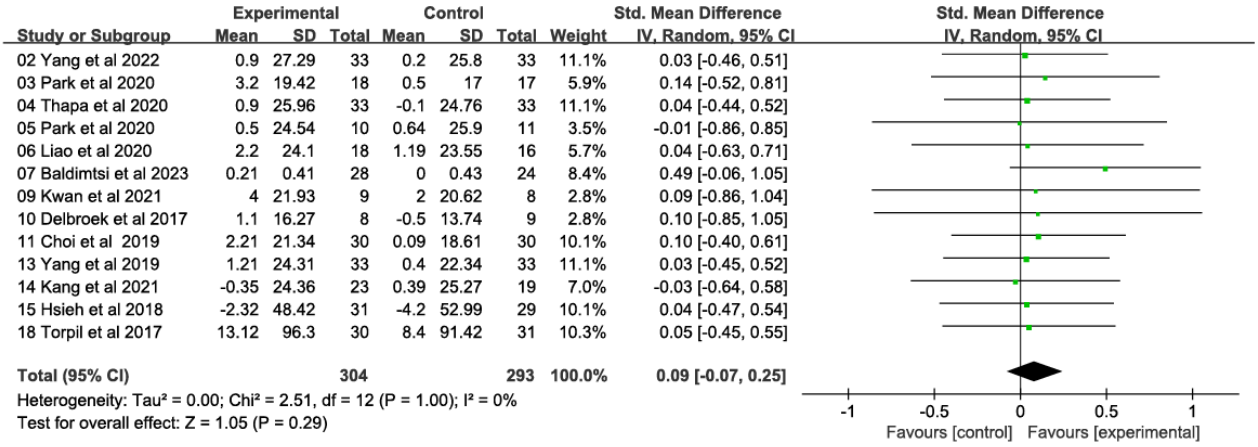

Performance and Memory

Ten studies, with a total of 465 participants, examined the impact of a VR technology–based intervention on performance and memory (Figure 4 [27,31,32,47,48,52-56]). The studies measured these effects at various time points after the intervention, ranging from 4 to 24 weeks. The VR technology–based intervention was compared with a traditional control condition. The overall effect sizes showed statistically significant and moderately sized impacts on performance and memory (SMD 0.20, 95% CI 0.02-0.38). There was some heterogeneity in the results (χ^2_9=2.62, *P*=.08; I^2=0%). The general cognitive function evidence was deemed moderate due to the high risk of bias in most studies (ie, study limitations).

Figure 4. Forest plot of performance and memory: comparison of performance and memory improvement at postintervention time points based on virtual reality intervention versus conventional treatment or no intervention control group.

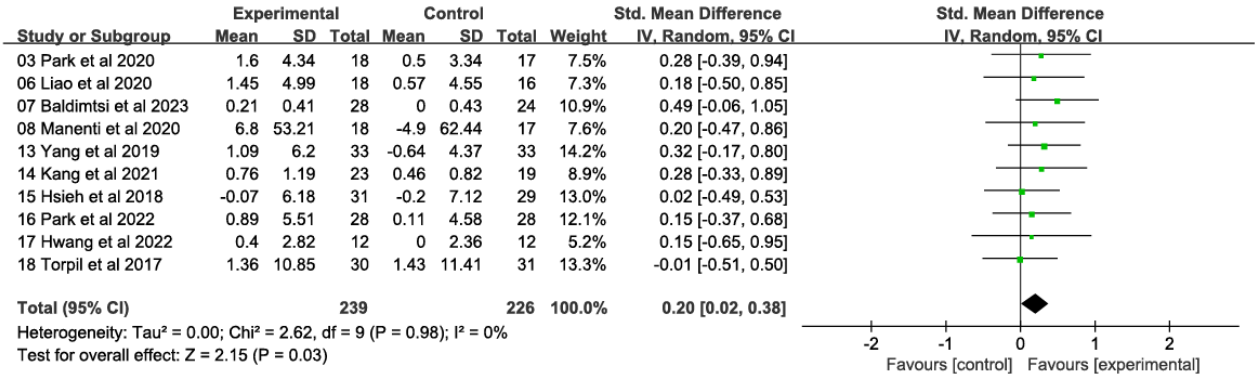

Attention and Information Processing Speed

A total of 9 studies involving 410 participants reported on the effects of a VR technology–based intervention on attention and information processing speed at postintervention time points (range 4-24 weeks) compared with a conventional control condition (Figure 5 [26-28,32,33,48,53,54,56]). Overall effect sizes indicated significant and moderate effects on attention and information processing speed (SMD 0.25, 95% CI 0.06-0.45); heterogeneity (χ^2_8=2.77, *P*=.95; I^2=0%). The quality of evidence for attention and information processing speed was rated as moderate due to the high risk of bias (ie, study limitations).

XSL•FO
RenderX

Figure 5. Forest plot of attention and information processing speed: comparison of attention and information processing speed improvement at postintervention time points based on virtual reality intervention versus conventional treatment or no intervention control group.

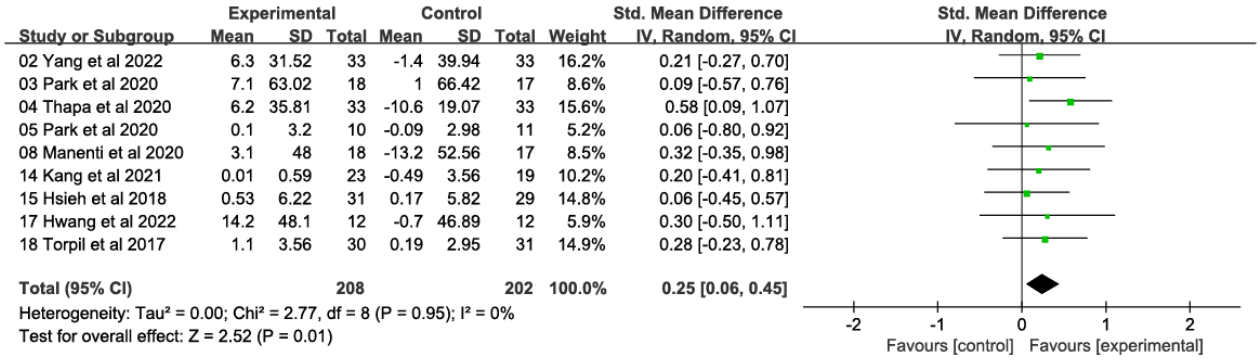

Executive Function

In total, 9 studies, with a total of 382 participants, examined the impact of VR technology–based intervention on executive function (Figure 6 [26,28,31-33,46,47,51,53]). These studies measured the effects at postintervention time points, which ranged from 4 to 12 weeks, and compared them with traditional control conditions. The intervention showed substantial and moderate effects on executive function (SMD 0.22, 95% CI 0.02-0.42). There was no heterogeneity among the studies (χ^2_8=1.28, P>.99; I^2=0%). The evidence on implementation capacity was assessed as intermediate in quality due to the high risk of bias (ie, study limitations).

Figure 6. Forest plot of executive function: comparison of executive function improvement at postintervention time points based on virtual reality intervention versus conventional treatment or no intervention control group.

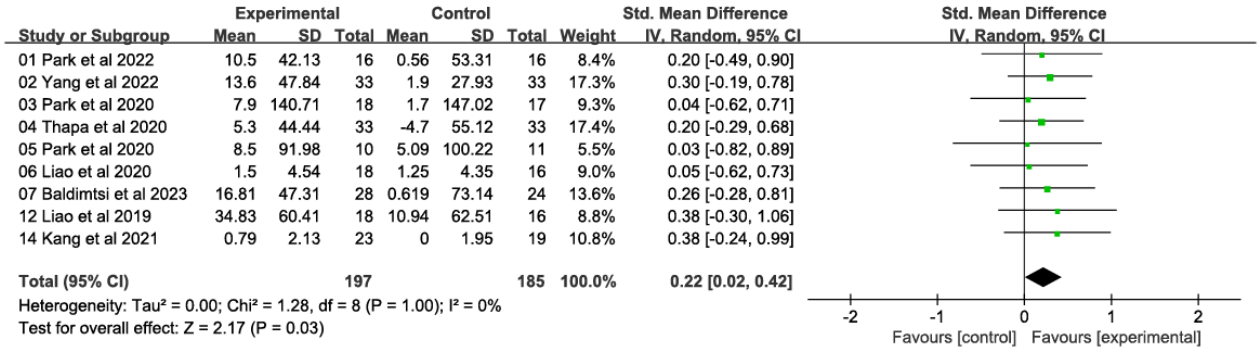

Secondary Outcomes

No significant findings were seen for any of the secondary outcomes (Multimedia Appendix 4). The study found that there were no significant effects on language proficiency (SMD 0.21, 95% CI −0.20 to 0.61) and the quality of evidence was rated as very low. However, there were significant effects on visuospatial abilities, depression, and the daily mobility of Individuals, with effect sizes of 0.24 (95% CI −0.09 to 0.57), 0.06 (95% CI −0.29 to 0.42), and 0.10 (95% CI −0.23 to 0.43), respectively. The quality of evidence for these effects was rated as low. The study found that the impact sizes for gait and balance were statistically negligible, with an SMD of 0.05 and a 95% CI ranging from −0.17 to 0.27. The quality of evidence was considered moderate.

Subgroup Analysis

Therapist Involvement

Performance and Memory

In total, 2 studies involving 118 participants found that interventions to improve performance and memory had significant and moderate effects (SMD 0.39, 95% CI 0.03-0.76). These studies did not involve therapist involvement (Figure 7 [31,32,47,48,52-55]). There was no heterogeneity in the results (χ^2=0.22, P=.64; I^2=0%). On the other hand, 6 studies involving 262 participants did involve therapist involvement but did not find significant effects favoring performance and memory interventions (SMD 0.17, 95% CI −0.07 to 0.41). There was also no heterogeneity in these results (χ^2_5=0.58, P=.99; I^2=0%). Nevertheless, there was no substantial distinction observed between the 2 groups (χ^2_1=1.02, P=.31; I^2=2.1%).

XSL•FO
RenderX

Figure 7. Forest plot of subgroups with and without therapist participation (performance and memory): comparison of performance and memory at postintervention time points based on therapist involvement or no therapist involvement.

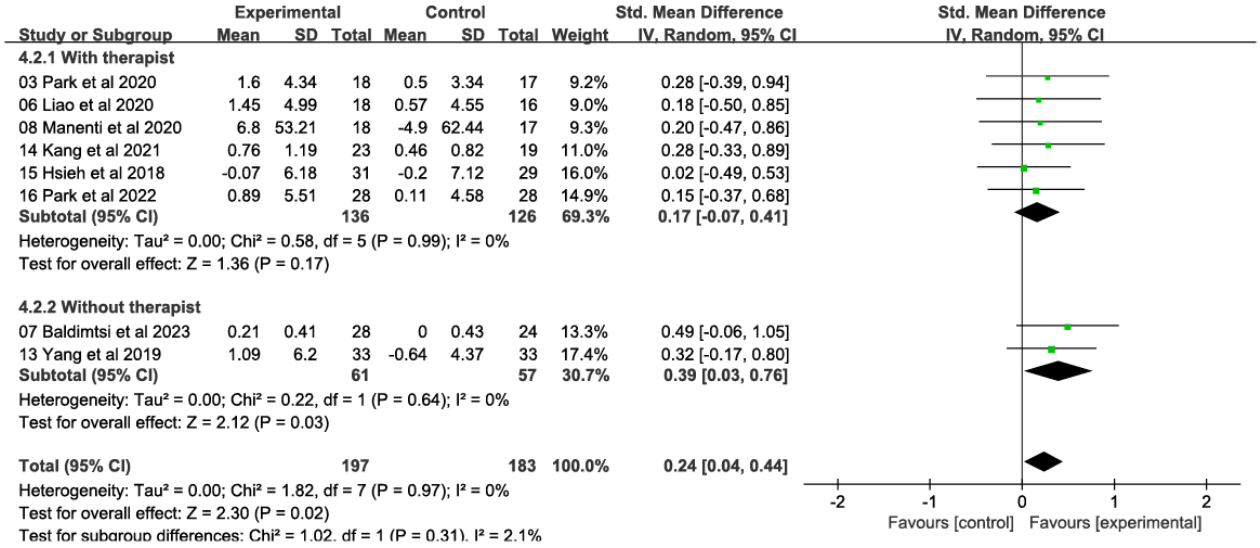

Attention and Information Processing Speed

Two studies involving 132 participants uncovered no therapist involvement results (Figure 8 [26,28,32,33,48,53,54]), with significant and moderate effects observed in favor of interventions to improve attention and information processing speed (SMD 0.39, 95% CI 0.03-0.75); there was no heterogeneity (χ^2_1=1.07, *P*=.30; I^2=7%). In contrast, 5 studies involving 193 participants uncovered results of therapist involvement but did not observe a significant effect in favor of the attention and information processing speed intervention (SMD 0.14, 95% CI –0.14 to 0.43); there was no heterogeneity (χ^2_4=0.46, *P*=.98; I^2=0%). However, the difference between the 2 groups was not significant (χ^2_1=1.16, *P*=.28; I^2=13.5%).

Figure 8. Forest plot of subgroups with and without therapist participation (attention and information processing speed): comparison of attention and information processing speed at postintervention time points based on therapist involvement or no therapist involvement.

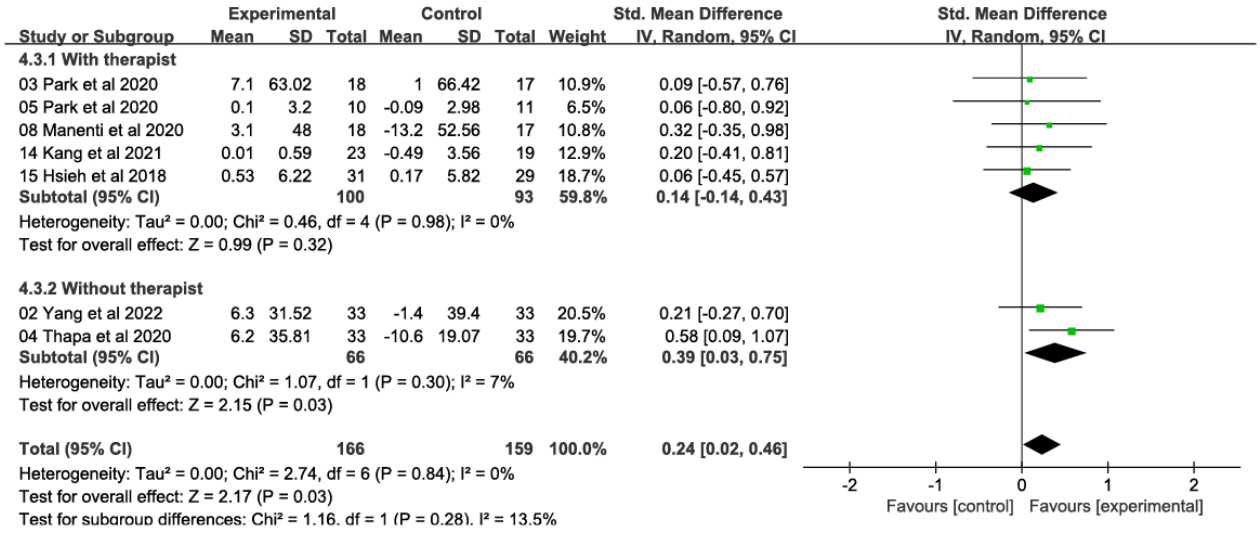

Content of the Intervention

Attention and Information Processing Speed

VR cognitive training (5 studies, 230 participants) produced significant and moderate effects in improving attention and information processing speed (SMD 0.31, 95% CI 0.05-0.58; χ^2_4=1.75, *P*=.78; I^2=0%; Figure 9 [26-28,32,33,48,53,54,56]). The other 3 intervention components—VR physical training (1 study, 60 participants; SMD 0.06, 95% CI −0.45 to 0.57), VR cognitive-motor dual-task training (2 studies, 96 participants; SMD 0.21, 95% CI −0.19 to 0.61; χ^2_1=0.19, *P*=.67; I^2=0%), and VR program (1 study, 24 participants; SMD 0.30, 95% CI −0.50 to 1.11)—did not have a significant effect on attention and information processing speed. The difference between the 4 groups was not significant (χ^2_3=0.84, *P*=.84; I^2=0%).

XSL•FO
RenderX

Figure 9. Forest plot of subgroups by intervention content (attention and information processing speed): comparison of virtual reality cognitive training, virtual reality physical training, virtual reality cognitive-motor dual-task training, and virtual reality program in improving attention and information processing speed at postintervention time points.

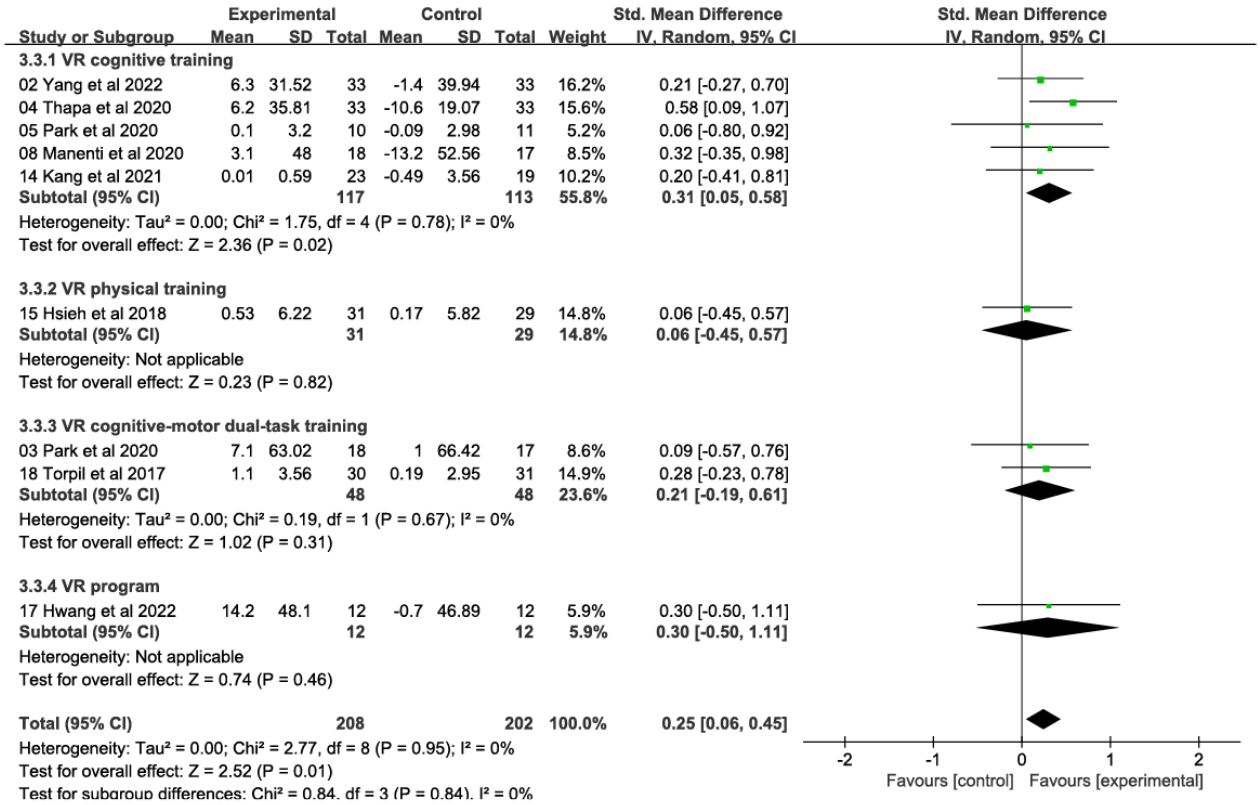

Level of Immersion

Attention and Information Processing Speed

Subgroup analyses based on level of immersion showed that immersive VR interventions (5 studies, 255 participants) produced significant and moderate effects on improving attention and information processing speed (SMD 0.25, 95% CI 0.01-0.50; χ^2_4=2.49, *P*=.65; I^2=0%; Figure 10 [26-28,32,33,48,53,54]). Semi-immersive VR interventions (1 study, 35 participants; SMD 0.09, 95% CI −0.57 to 0.76) and nonimmersive VR interventions (2 studies, 96 participants; SMD 0.29, 95% CI −0.11 to 0.69; χ^2_1=0.01, *P*=.92; I^2=0%) had no significant effects on attention and information processing speed. The difference between the 3 groups was not significant (χ^2_2=0.26, *P*=.88; I^2=0%).

XSL•FO
RenderX

Figure 10. Forest plot of subgroups by immersion level (attention and information processing speed): Immersive, semi-immersive, nonimmersive, and combined immersive and semi-immersive immersion levels improved attention and information processing speed at postintervention.

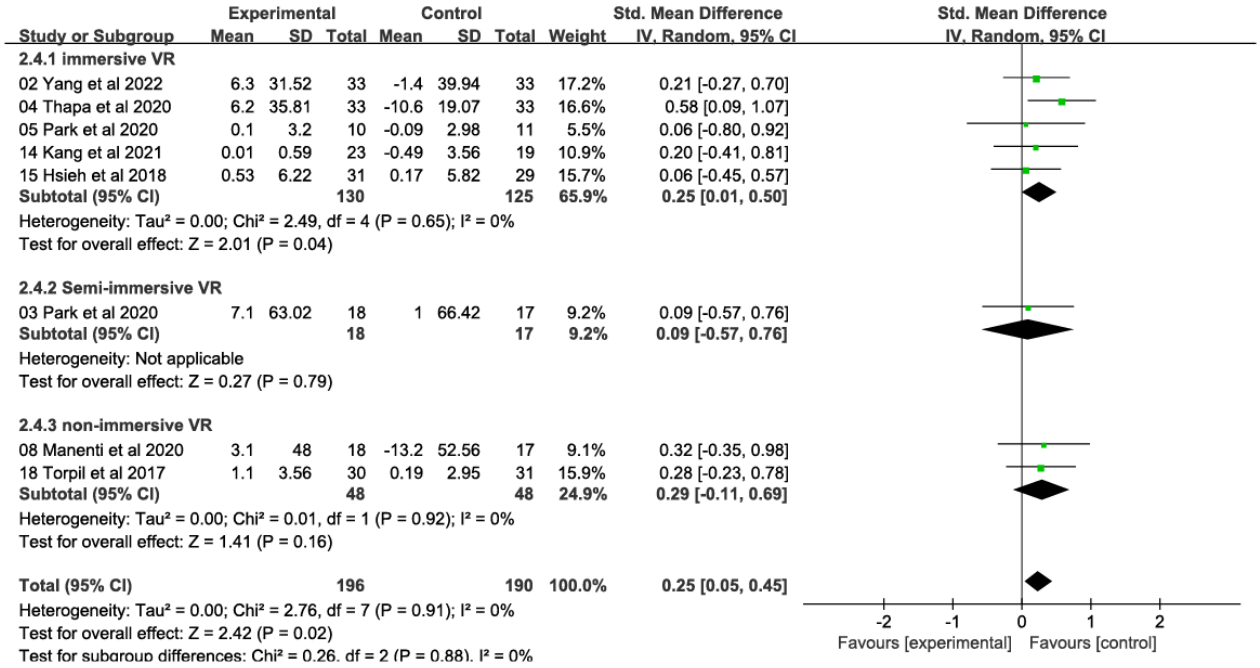

Executive Function

Subgroup analyses based on level of immersion showed that immersive VR interventions (5 studies, 247 participants) produced significant and moderate effects in improving executive function (SMD 0.25, 95% CI 0.00-0.50; χ^2_4=0.67, *P*=.95; I^2=0%; Figure 11 [26,28,31-33,46,47,51,53]). Semi-immersive VR interventions (2 studies, 67 participants; SMD 0.09, 95% CI −0.57 to 0.76; χ^2_1=0.11, *P*=.74; I^2=0%) as well as combined immersive and nonimmersive VR interventions (2 studies, 68 participants; SMD 0.22; 95% CI −0.26 to 0.69; χ^2_1=0.44, *P*=.74; I^2=0%) had no significant effect in executive function. The difference between the 3 groups was not significant (χ^2_2=0.22, *P*=.90; I^2=0%).

Figure 11. Forest plot of subgroups by immersion level (executive ability): comparison of immersion, semi-immersive, nonimmersive, and combined immersive and semi-immersive immersion levels in improving executive ability at postintervention time points.

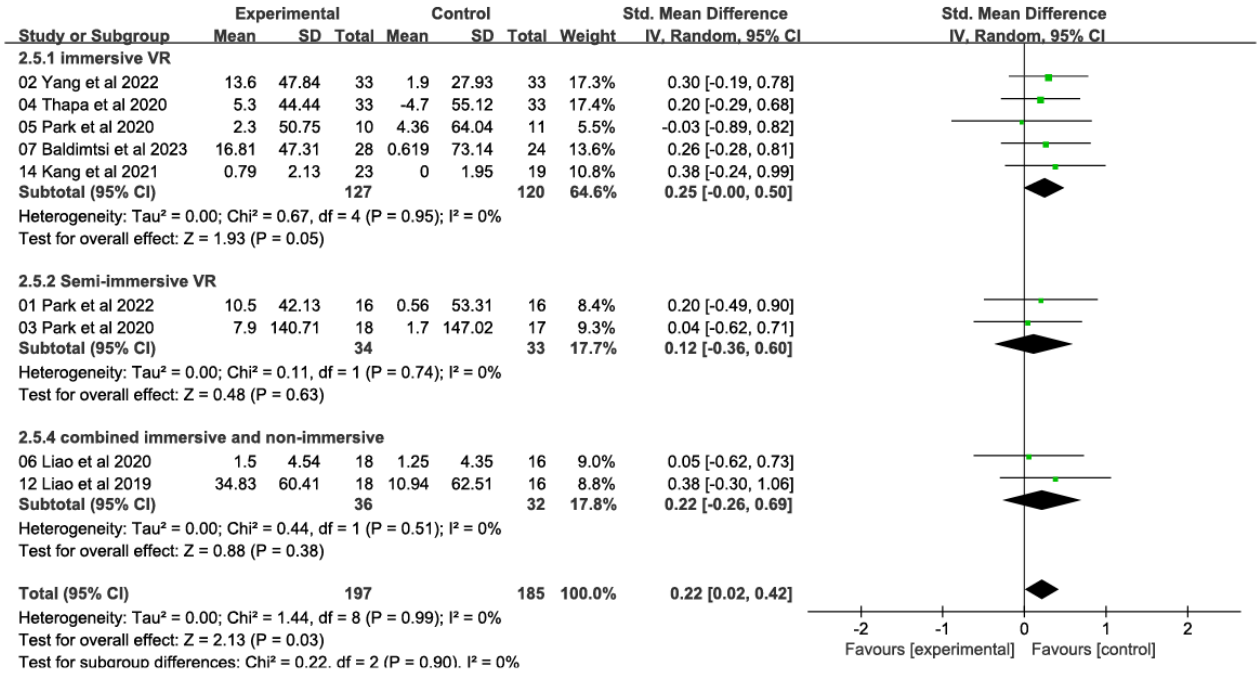

XSL•FO
RenderX

Risk-of-Bias Sensitivity Analysis

There was little evidence of funnel plot asymmetry in the intervention effects for general cognitive function, performance and memory, attention and speed of information processing, and executive function (Multimedia Appendix 5). In addition, excluding studies with no or unclear random allocation, inadequate concealment of treatment allocation, or incomplete outcome data, the effect sizes were as follows: improved general cognitive function (SMD 0.12, 95% CI −0.11 to 0.35); improved performance and memory (SMD 0.30, 95% CI 0.02-0.57); improved attention and information processing speed (SMD 0.41, 95% CI 0.05-0.77); and improved executive function (SMD 0.14, 95% CI −0.16 to 0.44). All the RCTs included in this analysis are parallel RCTs.

Discussion

Principal Findings

The findings of this review indicate that using VR as an intervention resulted in significant improvements in performance and memory (SMD 0.20, 95% CI 0.02-0.38), speed of attention and information processing (SMD 0.25, 95% CI 0.06-0.45), and executive function (SMD 0.22, 95% CI 0.02-0.42). Our study did not discover any notable disparities in general cognitive function and any secondary outcomes. However, the findings should be interpreted with caution as there is a risk of bias in the included studies, and therefore, the overall quality of the evidence is moderate and low.

This review examined how VR impacts performance and memory in older adults with MCI (SMD 0.20, 95% CI 0.02-0.38). Of the 10 included studies that assessed performance and memory, 9 showed positive changes. The results were similar to the study by Zhu et al [14]. This may be due to the fact that memory traces formed under the immersion of a virtual experience have richer content and a more complex associative network, and are part of an extensive autobiographical associative network [57,58]. A study showed that participants in the VR condition mostly relied on the mechanism of autobiographical memory, namely vivid recall, for retrieving memory task items. In contrast, participants in the 2D personal computer-game condition primarily relied on familiarity for retrieval [59]. VR offers advantages such as improved attitudes and heightened desire for training, in comparison with conventional motor or cognitive therapies [60]. Furthermore, VR has the ability to stimulate mirror neurons and improve individual empowerment and self-confidence by integrating perception, cognition, and action [61]. The findings of our study indicate that the use of VR intervention has a notable impact on both performance and memory. However, it is important to note that achieving this good effect may necessitate more comprehensive training during the VR intervention.

The effects of VR on the population with MCI in terms of attention and information processing speed were investigated in 9 included studies, all of which showed positive changes (SMD 0.25, 95% CI 0.06-0.45). The findings align with the research conducted by Yu et al [13], Gomez-Caceres et al [35], and Ren et al [62]. Previous studies demonstrated significant activation of the cuneate lobe, middle occipital gyrus, and regions involved in emotional processing (eg, the insula) during VR experiences [63]. In contrast, the right insula may be activated by a combination of attentional and response control demands that play a role in processing sensory stimuli related to the current target [64]. During VR attention training, engagement in visual attention tasks is associated with a number of significantly enhanced activation clusters, and stereoscopic presentation in VR may facilitate attentional engagement by reducing the metabolic costs of intermediate stages of visual processing [65].

The study examined the impact of VR on the executive function of individuals with MCI (SMD 0.22, 95% CI 0.02-0.42). All 9 of the included studies demonstrated beneficial changes. However, the results conflict with Yan et al [15]. This may be due to the fact that they focused only on VR in the form of combined cognitive and physical interventions, whereas combined interventions may lead to excessive stress and a reduction in the cognitive benefits of cognitive training or physical activity in combined interventions [66]. Most of the VR interventions in our study were goal-oriented tasks with IADL [14,35]. Executive function plays a dominant role in a person's ability to adapt to situations that arise in daily life [67,68]. Therefore, active experience-dependent plasticity triggered by task-oriented training leads to neurological changes in the brain and improves daily life performance [69]. Although this reference concerns poststroke patients, it is hypothesized that similar mechanisms of plasticity might be relevant for individuals with MCI engaged in VR-based cognitive training, based on the shared principles of neuroplasticity. Using IADL as a goal-directed task may improve the effectiveness of executive function interventions. Therefore, future research should consider this variable when developing intervention strategies.

Subgroup Analysis

The results of subgroup analyses show that the VR intervention without therapist involvement had better results in terms of improving performance and memory as well as in terms of attention and speed of information processing compared with the VR intervention with therapist involvement. The results conflict with the study of Manenti et al [48]. This may be due to the bias in this study [48], where the experimental and control groups were not equal in terms of intervention time, intervention measures, and the number of interventions was small. The lack of therapist guidance may have motivated participants to use cognitive strategies more actively, resulting in more effective processing of information and improved performance. Autonomous learning experiences can enhance individuals' cognitive processes, leading to deeper thinking and better integration of knowledge, ultimately resulting in more enduring memory effects. This aligns with the beneficial impact of self-directed learning on the strengthening of memory retention as observed in previous research [70-72]. Furthermore, a study conducted by Plancher et al [73] has noted that the advantages of active conditioning mostly rely on procedural abilities and the consequences of personal engagement in the encoding process. In terms of attention and information processing speed, users in contexts without therapist involvement were more likely to focus on the task itself, reducing the influence of external

XSL•FO
RenderX

incentives and cognitive distractions, resulting in more significant cognitive outcomes. In contexts with therapist involvement, users may be more susceptible to social comparisons and evaluations. The therapist's guidance and feedback may become part of the cognitive load and add additional psychological stress [74,75], which may have a negative impact on attention and information processing speed. In addition, the presence of a therapist introduces extrinsic motivation that may cause users to focus more on the therapist's expectations and feedback than on the task itself, which in turn affects the effectiveness of task performance [76]. It is important to note that this does not imply that the trainer is redundant in all contexts, but rather emphasizes that VR interventions without trainer intervention may be more conducive to achieving significant results in terms of performance and memory, as well as in terms of attention and speed of information processing, in specific contexts. Future research could further explore the interactive effects of different task types and individual differences on the presence or absence of trainer involvement to more fully understand the mechanisms behind this finding. Also, future research should be designed with more attention to user privacy issues to provide users with a stronger sense of security.

The results of the subgroup analysis regarding the content of the VR intervention indicate that VR cognitive training produces significant effects in improving attention and information processing speed compared with VR physical training, VR cognitive-motor dual-task training, and VR programs. This may be due to the fact that VR physical training and VR programs require lower intrinsic cognitive load, and VR cognitive training requires a higher cognitive load. The participants' cognitive systems are more actively engaged, thus stimulating more active cognitive processes [77]. In addition, exercise requires cognitive provision (engagement of attention, memory, and executive function) to fulfill its function [78]. Compared with single-tasking, dual-tasking is performed with higher brain activation in the prefrontal cortex and requires more cognitive resources [79]. Due to the limited capacity of the cognitive system, excessive cognitive load may affect learning and execution of the task [80]. VR cognitive training focuses on a specific cognitive task, whereas VR cognitive motor dual-task training may have introduced more distractions or loads that affected cognitive performance. Therefore, future VR intervention content should carefully consider the relationship between cognitive tasks and motor share, and should be designed to be more suitable for MCI patients.

Our study also find that immersive VR produced significant effects in terms of improving attention and information processing speed and in terms of executive function compared with semi-immersive VR, nonimmersive VR, and combined immersive and semi-immersive VR intervention situations. The results are similar to the study by Zhu et al [14] but conflicted with the study by Yu et al [13], which stated that semi-immersive VR was superior to fully immersive VR and nonimmersive VR in promoting cognitive flexibility. However, in recent years, VR technology has evolved and the use of immersive VR technology has been used to provide a more profound and holistic sensory experience through the creation of immersive virtual environments with more extensive sensory encounter [81]. The increased level of immersion in VR can promote interaction between the patient and the task, thus improving ecological validity [82]. Cognitive training in VR environments can enhance cognitive function in reality by facilitating the transfer of cognitive talents from games to real-life situations, thanks to cognitive function and brain plasticity [69,83]. However, the results should be interpreted with caution, as the included studies are at risk of bias, imprecise, and inconsistent, and therefore, the overall quality of the evidence is low or very low.

Limitations

This review has a number of limitations. First, only studies published in English were considered. Second, it is difficult to blind interventions and outcome assessments; as a result, all studies endure performance and detection bias. In addition, some of the most robust studies were at high risk of bias, reducing the overall quality of the evidence. Third, the sample sizes of most studies were small, limiting the statistical power and generalizability of the results, and secondary outcomes such as language proficiency, depression, and the daily mobility of Individuals were only reported in 4 studies. Fourth, defining thresholds of dysfunction remains a difficult task, and the included studies differed in inconsistency in their inclusion criteria for participants with MCI. Fifth, the age cutoff was set at 55 years and older, with the awareness that the definition of older adults varies across cultures, as some reviews include participants aged ≥40 years [84] and ≥65 years [13]. Given that many researchers have begun advocating against using age as the sole criterion for classifying older adults, future reviews could consider relaxing the inclusion criterion on age and include studies that investigate interventions for active aging. Finally, most studies did not take into account factors that can modulate participants' cognitive functioning, mood, the daily mobility of Individuals, and physical fitness (eg, cognitive reserve, diet, sleep, and physical and cognitive activities outside of the intervention). However, our findings can still provide compelling insights into the future of VR intervention design for older adults with MCI.

Conclusion

This review of 18 RCTs evaluates the effects of VR interventions on cognitive abilities, emotional functioning, daily mobility, and physical fitness in older adults with MCI. We found notable improvements in memory performance, attention, information processing speed, and executive function. VR training, particularly without therapist involvement, led to enhancements in memory, attention, and processing speed. These findings suggest that VR interventions can be an effective tool for cognitive rehabilitation in older adults with MCI, offering a practical approach to support their cognitive health and potentially delay the progression of dementia. Further research is needed to compare VR programs and validate these findings, considering patient diversity, intervention duration, and measurement accuracy.

XSL•FO
RenderX

Acknowledgments

This work was supported by the Natural Science Foundation of Shanghai as part of the Soft Science Project “Digital Rehabilitation Modeling for Patients with Schizophrenia” (Project 23692114900); by the Shanghai Jiao Tong University Medical-Engineering Crossover Fund for the Key Project “Research on Prediction and Intervention of Rehabilitation Status in Patients with Severe Mental Disorders Based on Multimodal Analysis and Artificial Intelligence Agents” (Project YG2024ZD24); and by the Shanghai Jiao Tong University Liberal Arts Youth Talent Cultivation Program (2023QN028). The funding agencies had no direct role in the publication of this article.

Authors' Contributions

QY, ZL, and LZ conceptualized and designed the study. QY, LZ, FC, and HY acquired, analyzed, and interpreted the data. QY drafted the manuscript. QY, ZL, BC, LZ, FC, and HY critically revised the manuscript for important intellectual content. LZ and QY conducted statistical analyses. BC and LZ supervised the study. All authors read and approved the final manuscript. QY and ZL had full access to all data in the study and take responsibility for the integrity of the data and accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 84 KB - jmir_v27i1e59195_app1.pdf]

Multimedia Appendix 2
Characteristics and demographics of included studies.
[PDF File (Adobe PDF File), 819 KB - jmir_v27i1e59195_app2.pdf]

Multimedia Appendix 3
Summary of findings and Grading of Recommendations, Assessment, Development, and Evaluations ratings for the main comparisons.
[PDF File (Adobe PDF File), 84 KB - jmir_v27i1e59195_app3.pdf]

Multimedia Appendix 4
Forest plots of secondary outcomes.
[PDF File (Adobe PDF File), 309 KB - jmir_v27i1e59195_app4.pdf]

Multimedia Appendix 5
Funnel plots of review comparisons.
[PDF File (Adobe PDF File), 137 KB - jmir_v27i1e59195_app5.pdf]

Multimedia Appendix 6
PRISMA checklist.
[PDF File (Adobe PDF File), 188 KB - jmir_v27i1e59195_app7.pdf]

References

1. Hendrie HC. Epidemiology of dementia and Alzheimer's disease. Am J Geriatr Psychiatry 1998;6(2 Suppl 1):S3-18. [doi: 10.1097/00019442-199821001-00002] [Medline: 9581216]
2. Dementia. Worldwide Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2024-11-28]
3. Winblad B, Amouyel P, Andrieu S, Ballard C, Brayne C, Brodaty H, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. Lancet Neurol 2016;15(5):455-532 [FREE Full text] [doi: 10.1016/S1474-4422(16)00062-4] [Medline: 26987701]
4. Cantarero-Prieto D, Leon PL, Blazquez-Fernandez C, Juan PS, Cobo CS. The economic cost of dementia: a systematic review. Dementia (London) 2020;19(8):2637-2657. [doi: 10.1177/1471301219837776] [Medline: 30909718]
5. Petersen RC. Mild cognitive impairment. Continuum (Minneap Minn) 2016;22(2 Dementia):404-418 [FREE Full text] [doi: 10.1212/CON.0000000000000313] [Medline: 27042901]

XSL•FO
RenderX

6. Anderson ND. State of the science on mild cognitive impairment (MCI). CNS Spectr 2019;24(1):78-87. [doi: 10.1017/S1092852918001347] [Medline: 30651152]
7. Eshkoor SA, Hamid TA, Mun CY, Ng CK. Mild cognitive impairment and its management in older people. Clin Interv Aging 2015;10:687-693 [FREE Full text] [doi: 10.2147/CIA.S73922] [Medline: 25914527]
8. Lasaponara S, Marson F, Doricchi F, Cavallo M. A scoping review of cognitive training in neurodegenerative diseases via computerized and virtual reality tools: what we know so far. Brain Sci 2021;11(5):528 [FREE Full text] [doi: 10.3390/brainsci11050528] [Medline: 33919244]
9. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. Neurology 2018;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
10. Pinto C, Subramanyam AA. Mild cognitive impairment: the dilemma. Indian J Psychiatry 2009;51 Suppl 1(Suppl1):S44-S51 [FREE Full text] [Medline: 21416016]
11. Riva G. Psicologia dei nuovi media. Azione, presenza, identità e relazioni. Bologna, Italy: Il Mulino; 2012.
12. Li L, Yu F, Shi D, Shi J, Tian Z, Yang J, et al. Application of virtual reality technology in clinical medicine. Am J Transl Res 2017;9(9):3867-3880 [FREE Full text] [Medline: 28979666]
13. Yu D, Li X, Lai FHY. The effect of virtual reality on executive function in older adults with mild cognitive impairment: a systematic review and meta-analysis. Aging Ment Health 2023;27(4):663-673. [doi: 10.1080/13607863.2022.2076202] [Medline: 35635486]
14. Zhu S, Sui Y, Shen Y, Zhu Y, Ali N, Guo C, et al. Effects of virtual reality intervention on cognition and motor function in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Front Aging Neurosci 2021;13:586999 [FREE Full text] [doi: 10.3389/fnagi.2021.586999] [Medline: 34025384]
15. Yan M, Zhao Y, Meng Q, Wang S, Ding Y, Liu Q, et al. Effects of virtual reality combined cognitive and physical interventions on cognitive function in older adults with mild cognitive impairment: a systematic review and meta-analysis. Ageing Res Rev 2022;81:101708. [doi: 10.1016/j.arr.2022.101708] [Medline: 35953010]
16. Skurla MD, Rahman AT, Salcone S, Mathias L, Shah B, Forester BP, et al. Virtual reality and mental health in older adults: a systematic review. Int Psychogeriatr 2022;34(2):143-155. [doi: 10.1017/S104161022100017X] [Medline: 33757619]
17. Yan M, Yin H, Meng Q, Wang S, Ding Y, Li G, et al. A virtual supermarket program for the screening of mild cognitive impairment in older adults: diagnostic accuracy study. JMIR Serious Games 2021;9(4):e30919 [FREE Full text] [doi: 10.2196/30919] [Medline: 34870610]
18. Wu R, Li A, Xue C, Chai J, Qiang Y, Zhao J, et al. Screening for mild cognitive impairment with speech interaction based on virtual reality and wearable devices. Brain Sci 2023;13(8):1222 [FREE Full text] [doi: 10.3390/brainsci13081222] [Medline: 37626578]
19. Tarnanas I, Schlee W, Tsolaki M, Müri R, Mosimann U, Nef T. Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study. JMIR Serious Games 2013;1(1):e1 [FREE Full text] [doi: 10.2196/games.2778] [Medline: 25658491]
20. Bayahya AY, Alhalabi W, AlAmri SH. Smart health system to detect dementia disorders using virtual reality. Healthcare (Basel) 2021;9(7):810 [FREE Full text] [doi: 10.3390/healthcare9070810] [Medline: 34203116]
21. Parsons TD, Barnett M. Validity of a newly developed measure of memory: feasibility study of the virtual environment grocery store. J Alzheimers Dis 2017;59(4):1227-1235. [doi: 10.3233/JAD-170295] [Medline: 28759971]
22. Zygouris S, Segkouli S, Triantafyllidis A, Giakoumis D, Tsolaki M, Votis K, et al. Usability of the virtual supermarket test for older adults with and without cognitive impairment. J Alzheimers Dis Rep 2022;6(1):229-234 [FREE Full text] [doi: 10.3233/ADR-210064] [Medline: 35719712]
23. Chessa M, Bassano C, Gusai E, Martis AE, Solari F. Human-computer interaction approaches for the assessment and the practice of the cognitive capabilities of elderly people. 2018 Presented at: Proceedings of the European Conference on Computer Vision (ECCV) Workshops; September 8-14, 2018; Munich, Germany. [doi: 10.1007/978-3-030-11024-6_5]
24. da Costa RQM, Pompeu JE, Moretto E, Silva JM, Dos Santos MD, Nitrini R, et al. Two immersive virtual reality tasks for the assessment of spatial orientation in older adults with and without cognitive impairment: concurrent validity, group comparison, and accuracy results. J Int Neuropsychol Soc 2022;28(5):460-472. [doi: 10.1017/S1355617721000655] [Medline: 34080532]
25. Jekel K, Damian M, Wattmo C, Hausner L, Bullock R, Connelly PJ, et al. Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. Alzheimers Res Ther 2015;7(1):17 [FREE Full text] [doi: 10.1186/s13195-015-0099-0] [Medline: 25815063]
26. Thapa N, Park HJ, Yang JG, Son H, Jang M, Lee J, et al. The effect of a virtual reality-based intervention program on cognition in older adults with mild cognitive impairment: a randomized control trial. J Clin Med 2020;9(5):1283 [FREE Full text] [doi: 10.3390/jcm9051283] [Medline: 32365533]
27. Torpil B, Şahin S, Pekçetin S, Uyanık M. The effectiveness of a virtual reality-based intervention on cognitive functions in older adults with mild cognitive impairment: a single-blind, randomized controlled trial. Games Health J 2021;10(2):109-114. [doi: 10.1089/g4h.2020.0086] [Medline: 33058735]

XSL•FO
RenderX

28. Yang JG, Thapa N, Park HJ, Bae S, Park KW, Park JH, et al. Virtual reality and exercise training enhance brain, cognitive, and physical health in older adults with mild cognitive impairment. Int J Environ Res Public Health 2022;19(20):13300 [FREE Full text] [doi: 10.3390/ijerph192013300] [Medline: 36293881]
29. Niki K, Yahara M, Inagaki M, Takahashi N, Watanabe A, Okuda T, et al. Immersive virtual reality reminiscence reduces anxiety in the oldest-old without causing serious side effects: a single-center, pilot, and randomized crossover study. Front Hum Neurosci 2021;14:598161 [FREE Full text] [doi: 10.3389/fnhum.2020.598161] [Medline: 33536887]
30. Son C, Park JH. Ecological effects of VR-based cognitive training on ADL and IADL in MCI and AD patients: a systematic review and meta-analysis. Int J Environ Res Public Health 2022;19(23):15875 [FREE Full text] [doi: 10.3390/ijerph192315875] [Medline: 36497946]
31. Baldimtsi E, Mouzakidis C, Karathanasi EM, Verykouki E, Hassandra M, Galanis E, et al. Effects of virtual reality physical and cognitive training intervention on cognitive abilities of elders with mild cognitive impairment. J Alzheimers Dis Rep 2023;7(1):1475-1490 [FREE Full text] [doi: 10.3233/ADR-230099] [Medline: 38225966]
32. Park JS, Jung YJ, Lee G. Virtual reality-based cognitive-motor rehabilitation in older adults with mild cognitive impairment: a randomized controlled study on motivation and cognitive function. Healthcare (Basel) 2020;8(3):335 [FREE Full text] [doi: 10.3390/healthcare8030335] [Medline: 32932997]
33. Park JH, Liao Y, Kim DR, Song S, Lim JH, Park H, et al. Feasibility and tolerability of a culture-based virtual reality (VR) training program in patients with mild cognitive impairment: a randomized controlled pilot study. Int J Environ Res Public Health 2020;17(9):3030 [FREE Full text] [doi: 10.3390/ijerph17093030] [Medline: 32349413]
34. Delbroek T, Vermeylen W, Spildooren J. The effect of cognitive-motor dual task training with the biorescue force platform on cognition, balance and dual task performance in institutionalized older adults: a randomized controlled trial. J Phys Ther Sci 2017;29(7):1137-1143 [FREE Full text] [doi: 10.1589/jpts.29.1137] [Medline: 28744033]
35. Gómez-Cáceres B, Cano-López I, Aliño M, Puig-Perez S. Effectiveness of virtual reality-based neuropsychological interventions in improving cognitive functioning in patients with mild cognitive impairment: a systematic review and meta-analysis. Clin Neuropsychol 2023;37(7):1337-1370. [doi: 10.1080/13854046.2022.2148283] [Medline: 36416175]
36. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264-269 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
37. Rippey RM, Bill D, Abeles M, Day J, Downing DS, Pfeiffer CA, et al. Computer-based patient education for older persons with osteoarthritis. Arthritis Rheum 1987;30(8):932-935. [doi: 10.1002/art.1780300814] [Medline: 3307793]
38. Optale G, Urgesi C, Busato V, Marin S, Piron L, Priftis K, et al. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabil Neural Repair 2010;24(4):348-357. [doi: 10.1177/1545968309353328] [Medline: 19934445]
39. Kim BR, Chun MH, Kim LS, Park JY. Effect of virtual reality on cognition in stroke patients. Ann Rehabil Med 2011;35(4):450-459 [FREE Full text] [doi: 10.5535/arm.2011.35.4.450] [Medline: 22506159]
40. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123(3):A12-A13. [Medline: 7582737]
41. Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. A cognitive training program based on principles of brain plasticity: results from the improvement in memory with plasticity-based adaptive cognitive training (IMPACT) study. J Am Geriatr Soc 2009;57(4):594-603 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.02167.x] [Medline: 19220558]
42. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge; 1989.
43. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. London, UK: The Cochrane Collaboration; 2011.
44. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
45. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches the GRADE working group. BMC Health Serv Res 2004;4(1):38 [FREE Full text] [doi: 10.1186/1472-6963-4-38] [Medline: 15615589]
46. Park JH. Does the virtual shopping training improve executive function and instrumental activities of daily living of patients with mild cognitive impairment? Asian J Psychiatr 2022;69:102977. [doi: 10.1016/j.ajp.2021.102977] [Medline: 34998232]
47. Liao YY, Tseng HY, Lin YJ, Wang CJ, Hsu WC. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020;56(1):47-57 [FREE Full text] [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
48. Manenti R, Gobbi E, Baglio F, Macis A, Ferrari C, Pagnoni I, et al. Effectiveness of an innovative cognitive treatment and telerehabilitation on subjects with mild cognitive impairment: a multicenter, randomized, active-controlled study. Front Aging Neurosci 2020;12:585988 [FREE Full text] [doi: 10.3389/fnagi.2020.585988] [Medline: 33304267]
49. Kwan RYC, Liu JYW, Fong KNK, Qin J, Leung PKY, Sin OSK, et al. Feasibility and effects of virtual reality motor-cognitive training in community-dwelling older people with cognitive frailty: pilot randomized controlled trial. JMIR Serious Games 2021;9(3):e28400 [FREE Full text] [doi: 10.2196/28400] [Medline: 34383662]

XSL•FO
RenderX

50. Choi W, Lee S. The effects of virtual kayak paddling exercise on postural balance, muscle performance, and cognitive function in older adults with mild cognitive impairment: a randomized controlled trial. J Aging Phys Act 2019;27(4):861-870 [FREE Full text] [doi: 10.1123/japa.2018-0020] [Medline: 31185775]
51. Liao YY, Chen IH, Lin YJ, Chen Y, Hsu WC. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019;11:162 [FREE Full text] [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]
52. Yang HL, Chu H, Kao CC, Chiu HL, Tseng IJ, Tseng P, et al. Development and effectiveness of virtual interactive working memory training for older people with mild cognitive impairment: a single-blind randomised controlled trial. Age Ageing 2019;48(4):519-525. [doi: 10.1093/ageing/afz029] [Medline: 30989165]
53. Kang JM, Kim N, Lee SY, Woo SK, Park G, Yeon BK, et al. Effect of cognitive training in fully immersive virtual reality on visuospatial function and frontal-occipital functional connectivity in predementia: randomized controlled trial. J Med Internet Res 2021;23(5):e24526 [FREE Full text] [doi: 10.2196/24526] [Medline: 33955835]
54. Hsieh CC, Lin PS, Hsu WC, Wang JS, Huang YC, Lim AY, et al. The effectiveness of a virtual reality-based tai Chi exercise on cognitive and physical function in older adults with cognitive impairment. Dement Geriatr Cogn Disord 2018;46(5-6):358-370. [doi: 10.1159/000494659] [Medline: 30537752]
55. Park JH. Effects of virtual reality-based spatial cognitive training on hippocampal function of older adults with mild cognitive impairment. Int Psychogeriatr 2022;34(2):157-163. [doi: 10.1017/S1041610220001131] [Medline: 32616109]
56. Hwang J, Lee S. The effect of virtual reality program on the cognitive function and balance of the people with mild cognitive impairment. J Phys Ther Sci 2017;29(8):1283-1286 [FREE Full text] [doi: 10.1589/jpts.29.1283] [Medline: 28878448]
57. Schöne B, Wessels M, Gruber T. Experiences in virtual reality: a window to autobiographical memory. Curr Psychol 2017;38(3):715-719. [doi: 10.1007/s12144-017-9648-y]
58. Zhu D, Zhu B, Zhao J, Zhang C, He J, Song D, et al. Investigation on the effectiveness of augmented reality memory training game for Chinese older adults: a randomized controlled trial. Games Health J 2024;13(1):5-12. [doi: 10.1089/g4h.2022.0183] [Medline: 38193809]
59. Kisker J, Gruber T, Schöne B. Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: a study on human memory. Curr Psychol 2019;40(7):3190-3197. [doi: 10.1007/s12144-019-00257-2]
60. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
61. Rizzolatti G, Fogassi L. The mirror mechanism: recent findings and perspectives. Philos Trans R Soc Lond B Biol Sci 2014;369(1644):20130420 [FREE Full text] [doi: 10.1098/rstb.2013.0420] [Medline: 24778385]
62. Ren Y, Wang Q, Liu H, Wang G, Lu A. Effects of immersive and non-immersive virtual reality-based rehabilitation training on cognition, motor function, and daily functioning in patients with mild cognitive impairment or dementia: a systematic review and meta-analysis. Clin Rehabil 2024;38(3):305-321. [doi: 10.1177/02692155231213476] [Medline: 38289618]
63. Clemente M, Rey B, Rodriguez-Pujadas A, Barros-Loscertales A, Banos RM, Botella C, et al. An fMRI study to analyze neural correlates of presence during virtual reality experiences. Interacting with Computers 2013;26(3):269-284. [doi: 10.1093/iwc/iwt037]
64. Nagel IE, Chicherio C, Li SC, von Oertzen T, Sander T, Villringer A, et al. Human aging magnifies genetic effects on executive functioning and working memory. Front Hum Neurosci 2008;2:1 [FREE Full text] [doi: 10.3389/neuro.09.001.2008] [Medline: 18958202]
65. Lorentz L, Schüppen A, Suchan B, Binkofski F. Neural correlates of virtual reality-based attention training: an fMRI study. Neuroimage 2023;284:120454 [FREE Full text] [doi: 10.1016/j.neuroimage.2023.120454] [Medline: 37979896]
66. Karssemeijer EGA, Aaronson JA, Bossers WJ, Smits T, Olde Rikkert MGM, Kessels RPC. Positive effects of combined cognitive and physical exercise training on cognitive function in older adults with mild cognitive impairment or dementia: a meta-analysis. Ageing Res Rev 2017;40:75-83. [doi: 10.1016/j.arr.2017.09.003] [Medline: 28912076]
67. Poncet F, Swaine B, Dutil E, Chevignard M, Pradat-Diehl P. How do assessments of activities of daily living address executive functions: a scoping review. Neuropsychol Rehabil 2017;27(5):618-666. [doi: 10.1080/09602011.2016.1268171] [Medline: 28075219]
68. Romero-Ayuso D, Castillero-Perea Á, González P, Navarro E, Molina-Massó JP, Funes MJ, et al. Assessment of cognitive instrumental activities of daily living: a systematic review. Disabil Rehabil 2021;43(10):1342-1358 [FREE Full text] [doi: 10.1080/09638288.2019.1665720] [Medline: 31549907]
69. Cho DR, Lee SH. Effects of virtual reality immersive training with computerized cognitive training on cognitive function and activities of daily living performance in patients with acute stage stroke: a preliminary randomized controlled trial. Medicine (Baltimore) 2019;98(11):e14752 [FREE Full text] [doi: 10.1097/MD.0000000000014752] [Medline: 30882644]
70. Clark I. Formative assessment: assessment Is for self-regulated learning. Educ Psychol Rev 2012;24(2):205-249. [doi: 10.1007/s10648-011-9191-6]
71. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]

XSL•FO
RenderX

72. Luria E, Shalom M, Levy DA. Cognitive neuroscience perspectives on motivation and learning: Revisiting self‐determination theory. Mind Brain Educ 2020;15(1):5-17. [doi: 10.1111/mbe.12275]
73. Plancher G, Tirard A, Gyselinck V, Nicolas S, Piolino P. Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and Alzheimer's disease: influence of active and passive encoding. Neuropsychologia 2012;50(5):592-602. [doi: 10.1016/j.neuropsychologia.2011.12.013] [Medline: 22261400]
74. Tschanz JT, Pfister R, Wanzek J, Corcoran C, Smith K, Tschanz BT, et al. Stressful life events and cognitive decline in late life: moderation by education and age. The cache county study. Int J Geriatr Psychiatry 2013;28(8):821-830 [FREE Full text] [doi: 10.1002/gps.3888] [Medline: 23037866]
75. Peavy GM, Jacobson MW, Salmon DP, Gamst AC, Patterson TL, Goldman S, et al. The influence of chronic stress on dementia-related diagnostic change in older adults. Alzheimer Dis Assoc Disord 2012;26(3):260-266 [FREE Full text] [doi: 10.1097/WAD.0b013e3182389a9c] [Medline: 22037597]
76. Benabou R, Tirole J. Intrinsic and extrinsic motivation. Rev Econ Studies 2003;70(3):489-520. [doi: 10.1111/1467-937X.00253]
77. Sweller J. Cognitive load theory. In: Psychology of Learning and Motivation. volume 55. Washington, DC: Academic Press; 2011:37-76.
78. Koziol LF, Budding DE, Chidekel D. From movement to thought: executive function, embodied cognition, and the cerebellum. Cerebellum 2012;11(2):505-525. [doi: 10.1007/s12311-011-0321-y] [Medline: 22068584]
79. Mancioppi G, Fiorini L, Rovini E, Cavallo F. The use of motor and cognitive dual-task quantitative assessment on subjects with mild cognitive impairment: a systematic review. Mech Ageing Dev 2021;193:111393. [doi: 10.1016/j.mad.2020.111393] [Medline: 33188785]
80. Schnotz W, Kürschner C. A reconsideration of cognitive load theory. Educ Psychol Rev 2007;19:469-508. [doi: 10.1007/s10648-007-9053-4]
81. Rubio-Tamayo JL, Gertrudix M, García F. Immersive environments and virtual reality: systematic review and advances in communication, interaction and simulation. Multimodal Technol. Interact 2017;1(4):21. [doi: 10.3390/mti1040021]
82. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
83. Georgiev DD, Georgieva I, Gong Z, Nanjappan V, Georgiev GV. Virtual reality for neurorehabilitation and cognitive enhancement. Brain Sci 2021;11(2):221 [FREE Full text] [doi: 10.3390/brainsci11020221] [Medline: 33670277]
84. Zhidong C, Wang X, Yin J, Song D, Chen Z. Effects of physical exercise on working memory in older adults: a systematic and meta-analytic review. Eur Rev Aging Phys Act 2021;18(1):18 [FREE Full text] [doi: 10.1186/s11556-021-00272-y] [Medline: 34535084]

Abbreviations

AD: Alzheimer disease
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
IADL: instrumental activities of daily living
LOS: limit of stability
MCI: mild cognitive impairment
PICO: Population, Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SMD: standardized mean difference
VR: virtual reality

Edited by A Coristine; submitted 09.04.24; peer-reviewed by G Saco-Ledo, A Guaita; comments to author 14.05.24; revised version received 03.07.24; accepted 04.11.24; published 10.01.25.

Please cite as:
Yang Q, Zhang L, Chang F, Yang H, Chen B, Liu Z
Virtual Reality Interventions for Older Adults With Mild Cognitive Impairment: Systematic Review and Meta-Analysis of Randomized Controlled Trials
J Med Internet Res 2025;27:e59195
URL: https://www.jmir.org/2025/1/e59195
doi:10.2196/59195
PMID:

XSL•FO
RenderX

©Qin Yang, Liuxin Zhang, Fangyuan Chang, Hongyi Yang, Bin Chen, Zhao Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Clinical Decision Support Using Speech Signal Analysis: Systematic Scoping Review of Neurological Disorders

Upeka De Silva[1], MEng; Samaneh Madanian[1], PhD; Sharon Olsen[2], PhD; John Michael Templeton[3], PhD; Christian Poellabauer[4], PhD; Sandra L Schneider[5], PhD; Ajit Narayanan[1], PhD; Rahmina Rubaiat[6], MSc

[1]Department of Computer Science and Software Engineering, Auckland University of Technology, Auckland, New Zealand

[2]Rehabilitation Innovation Centre, Auckland University of Technology, Auckland, New Zealand

[3]School of Computer Science and Engineering, University of South Florida, Tampa, FL, United States

[4]School of Computing and Information Sciences, Florida International University, Miami, FL, United States

[5]Department of Communicative Sciences & Disorders, St Mary's College, Notre Dame, IN, United States

[6]Knight Foundation of Computing & Information Sciences, Florida International University, Miami, FL, United States

Corresponding Author:
Samaneh Madanian, PhD
Department of Computer Science and Software Engineering
Auckland University of Technology
55 Wellesley Street East, Auckland CBD, Auckland 1010
Auckland, 1010
New Zealand
Phone: 64 09 9219999 ext 6539
Email: sam.madanian@aut.ac.nz

Abstract

Background: Digital biomarkers are increasingly used in clinical decision support for various health conditions. Speech features as digital biomarkers can offer insights into underlying physiological processes due to the complexity of speech production. This process involves respiration, phonation, articulation, and resonance, all of which rely on specific motor systems for the preparation and execution of speech. Deficits in any of these systems can cause changes in speech signal patterns. Increasing efforts are being made to develop speech-based clinical decision support systems.

Objective: This systematic scoping review investigated the technological revolution and recent digital clinical speech signal analysis trends to understand the key concepts and research processes from clinical and technical perspectives.

Methods: A systematic scoping review was undertaken in 6 databases guided by a set of research questions. Articles that focused on speech signal analysis for clinical decision-making were identified, and the included studies were analyzed quantitatively. A narrower scope of studies investigating neurological diseases were analyzed using qualitative content analysis.

Results: A total of 389 articles met the initial eligibility criteria, of which 72 (18.5%) that focused on neurological diseases were included in the qualitative analysis. In the included studies, Parkinson disease, Alzheimer disease, and cognitive disorders were the most frequently investigated conditions. The literature explored the potential of speech feature analysis in diagnosis, differentiating between, assessing the severity and monitoring the treatment of neurological conditions. The common speech tasks used were sustained phonations, diadochokinetic tasks, reading tasks, activity-based tasks, picture descriptions, and prompted speech tasks. From these tasks, conventional speech features (such as fundamental frequency, jitter, and shimmer), advanced digital signal processing–based speech features (such as wavelet transformation–based features), and spectrograms in the form of audio images were analyzed. Traditional machine learning and deep learning approaches were used to build predictive models, whereas statistical analysis assessed variable relationships and reliability of speech features. Model evaluations primarily focused on analytical validations. A significant research gap was identified: the need for a structured research process to guide studies toward potential technological intervention in clinical settings. To address this, a research framework was proposed that adapts a design science research methodology to guide research studies systematically.

Conclusions: The findings highlight how data science techniques can enhance speech signal analysis to support clinical decision-making. By combining knowledge from clinical practice, speech science, and data science within a structured research framework, future research may achieve greater clinical relevance.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63004) doi:10.2196/63004

KEYWORDS

digital health; health informatics; digital biomarker; speech analytics; artificial intelligence; machine learning

Introduction

Background

Clinical decision-making is key to effective patient care. It fundamentally relies on evidence derived from validated assessment tools. These assessments typically combine subjective observations (ie, observable signs and patient-reported symptoms) and objective measurements (ie, physiological measurements [1-4], such as blood pressure and heart rate) [5]. However, these traditional assessments face inherent challenges due to clinical feature complexity, potential clinician bias, varying levels of expertise, and substantial instrumentation costs.

Digital biomarkers emerge as a transformative paradigm in clinical decision support, providing precise, objective measurements that extend beyond traditional assessments. Through smartphone and wearable sensors, these biomarkers capture granular patient data previously inaccessible to clinicians. For example, motion sensors quantify fine motor control through typing patterns and touch screen interactions, whereas accelerometers and gyroscopes measure gait parameters, including stride length variability, postural sway, and turning speed. Cognitive functions can be continuously assessed through patterns of smart device use, including response times, error rates, and daily activity rhythms [6]. Speech signals captured through recording instruments can detect subtle speech variations such as changes in fundamental frequency, rhythmic disturbances, voice quality, articulatory precision, and prosodic features that may indicate psychiatric conditions [7,8]. Among these emerging digital biomarkers, speech feature measurements provide clinical insights through a noninvasive, nonintrusive approach using low-cost smart and wearable digital devices [9] at scale in real time and offline modes [10,11].

Speech production is a complex task that involves the orchestration and coordination of different body systems [12]. Deficiencies in any component of the speech production system could manifest in speech pattern changes [13]. Therefore, these alterations provide objective, quantifiable markers for differential diagnosis and disease progression monitoring. They can also provide insights into normal and pathological biological processes [6,14]. Traditional clinical speech assessment mostly relies on standardized tests administered by speech-language pathologists to assess motor speech production [13,15,16] for conditions such as traumatic brain injury, stroke [17], Parkinson disease (PD), and multiple sclerosis (MS) [18]. In current practices, acoustic measures and auditory perceptual judgments [19] are typically used based on guidelines such as the Darley, Aronson, and Brown system [20,21] in the characterization of motor speech control deficits. Despite these approaches and guidelines, limitations exist in these conventional procedures. Some assessments are time-consuming, require specialized clinical experts [22,23], and heavily rely on clinicians' subjective perceptual judgments. This subjectivity introduces interpretation variability [24] and challenges in maintaining consistent interrater reliability [25,26] although there is evidence showing consistent auditory-perceptual assessments [27]. The environmental, physical, and emotional states of patients during the assessments can also lead to further inconsistencies [22].

Therefore, digital speech signal analysis (hereafter referred to as "speech analysis") offers a promising solution through enhanced objectivity and retest capability [22] with reduced clinical burden and improved accuracy. Opportunities exist in identifying specific speech features or speech patterns related to different health conditions, including neurological diseases [18,28]. Recent advances in artificial intelligence and machine learning (ML) techniques further support detecting subtle changes and fine-grained speech features that can be related to associated health conditions [29].

There is a growing interest in investigating speech as a digital biomarker for clinical decision support. Several studies have explored speech analysis for clinical assessments of specific neurological health conditions. The prosodic aspect of speech production in PD was reviewed in the study by Moro-Velasquez and Dehak [30], whereas Moro-Velasquez et al [31] reviewed the articulatory and phonatory aspects. Early detection of PD using speech features and ML was discussed in the study by Gullapalli and Mittal [32]. Automatic speech assessment in Alzheimer disease (AD) was reviewed in the study by Pulido et al [33], whereas Martínez-Nicolás et al [34] included mild cognitive impairment (MCI) as well. Both the studies by de la Fuente Garcia et al [35] and Petti et al [36] also focused on AD but considered language assessments in addition to speech. Automated speech and language features were reviewed as an indication of deficits in content organization and thought processes considering related neurological impairments such as AD and MCI [15]. A meta-analysis of acoustic features on autism spectrum disorder (ASD) [37] and articulatory impairments in neurodegenerative motor diseases [38] also reviewed and compiled knowledge on speech signal analysis.

Objectives

Despite the growing body of research on speech analysis for different diseases, the field lacks a comprehensive synthesis of available studies, their approaches, and clinical applications for neurological diseases. Therefore, in this research, we aimed to investigate the technology revolution and trends in speech analysis to understand the key concepts and research processes across different neurological conditions. Our review focused on studies that investigated the physical features of speech, focusing on the underlying digital acoustic features instead of the content. We aimed to review the clinical and technical perspectives of the research process in relation to clinical application. Given the interdisciplinary nature of the research field and the premature stage of clinical integration, we emphasize the importance of establishing a suitable research

XSL•FO
RenderX

framework to guide future research. We also proposed a research framework for speech analysis in clinical decision support adapting a design science research process.

Methods

Overview

The review process adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [39] (Multimedia Appendix 1). We conducted a systematic scoping review following a predefined protocol to comprehensively identify, evaluate, and synthesize relevant literature. Our review was guided by the following research questions:

- In which health care disciplines or for which health conditions has speech analysis been investigated? (research question 1)
- What types of clinical outcomes or clinical services might be possible using speech analysis? (research question 2)
  - What data science methods have been used for these speech analyses? (research question 2.1)

No protocol was registered for this review.

Sources and Eligibility Criteria

As a part of a larger research program, we conducted a comprehensive data collection. The review was limited to the Scopus, IEEE Xplore, Google Scholar, MEDLINE via PubMed, SpringerLink, and ScienceDirect databases. Textbox 1 lists the inclusion and exclusion criteria followed during article selection.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Language: publications in English only
- Publication type: peer-reviewed journal articles and conference proceedings
- Publication period: published between January 2010 and December 2022
- Participants: human participants of any age
- Type of research: primary research
- Study characteristics:
  - Carried out analysis of features of speech signals
  - Used features for analysis that were language independent
  - Used data science approaches to derive clinical insights on health conditions that impact speech (statistical analyses and traditional machine learning and deep learning approaches were considered interesting data science approaches)

Exclusion criteria

- Language: non–English-language publications
- Publication type: book chapters, theses, abstracts, editorials, and gray literature
- Publication period: published before January 2010
- Participants: nonhuman participants
- Study characteristics:
  - Analyzed only transcriptions or linguistics features, such as grammar and semantics
  - Investigated speech signal analysis on functional and structural voice disorders
  - Investigated only nonverbal audio sounds, such as breathing, coughing, crying, and snoring
  - Investigated only linguistic features of speech
  - Did not focus on a health care aspect (eg, the studies analyzed speech signals for emotion recognition without concerning a health condition)

Search Strategy and Study Selection

The following keywords and their synonyms were used in building searching queries and extracting articles: “Speech,” “Analysis,” “Health care,” “Data Science,” and “Artificial Intelligence.” The search strategy for the databases is presented in Multimedia Appendix 2.

The identified articles were exported to a reference management software where duplicates were removed. We then screened articles based on the titles and abstracts, followed by a full-text search. After a detailed examination of the retrieved full texts, those that met the eligibility criteria (N=389) were selected for the quantitative analysis of the research landscape. The full texts of articles that met the defined eligibility criteria and focused on a neurological disease with primary data collection (72/389, 18.5%) were included in this analysis and added to NVivo (Lumivero) [40], a qualitative data analysis software. We used qualitative content analysis followed by a mix of deductive and

XSL•FO
RenderX

inductive reasoning [41] to identify key concepts in speech analysis for clinical decision support.

The details of the included research articles can be found in Multimedia Appendix 3 [42-113].

Study Characteristics

The clinical disciplines and the yearly distribution of the overall eligible articles (N=389) are plotted in Figure 1 to understand the potential health care disciplines for digital clinical speech analysis to answer our research question 1. The evaluation of the research field was clearly evident, with an increasing number of articles over the years. Neurological diseases attracted the highest research interest (221/389, 56.8%), including AD, PD, and amyotrophic lateral sclerosis (ALS). Psychiatric disorders such as depression, bipolar disorder, and schizophrenia were among the next most investigated disease categories (98/389, 25.2%), followed by respiratory diseases (40/389, 10.3%). The remaining articles (30/389, 7.7%) investigated other health conditions, such as heart disease and cancer.

Figure 1. Yearly distribution of the eligible articles (N=389) and their clinical disciplines.

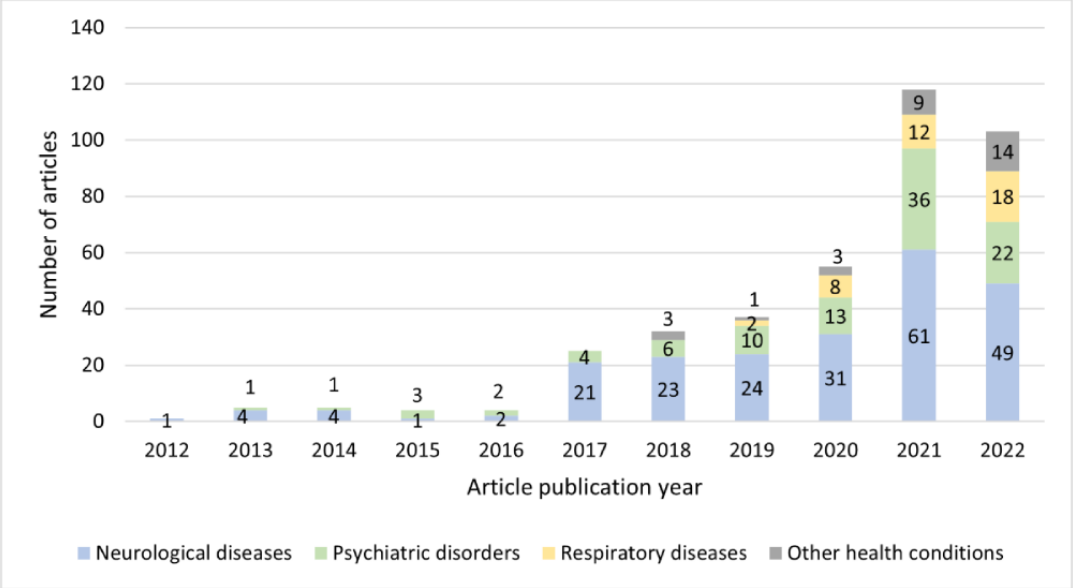

Data sourcing in the studies varied across public datasets, institutionally shared datasets, and primary data collections (Figure 2). In this review, we focused on studies with primary data collection as they described the full research process from problem formulation, data collection, data analysis, and model building to model evaluation. Therefore, the analysis of this review considered 18.5% (72/389) of the studies, which investigated neurological diseases using primary data collection.

Figure 2. Distribution of the eligible articles (N=389) based on their data sourcing approach.

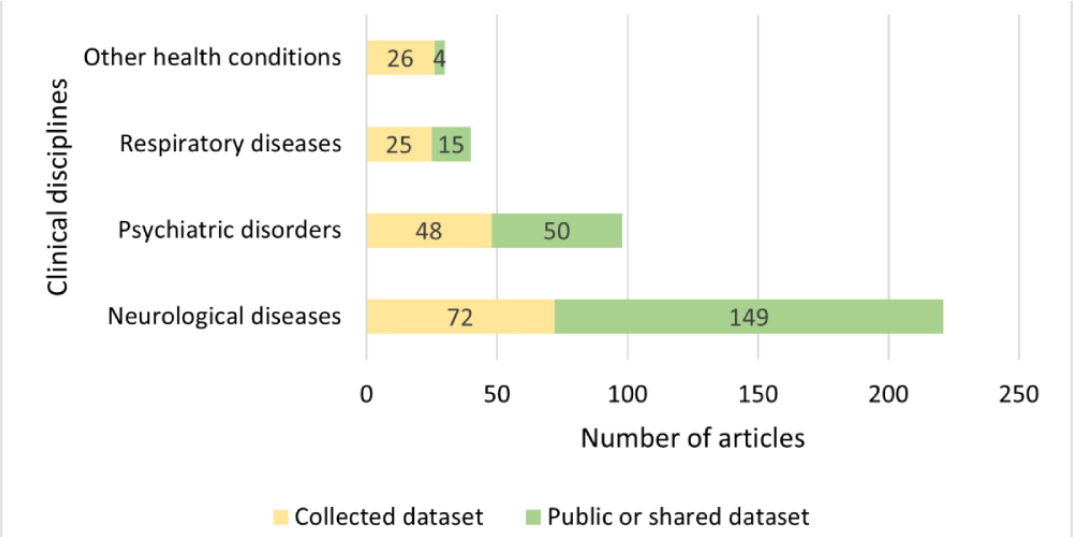

XSL•FO
RenderX

Results

Overview

Figure 3 presents the detailed overview of the literature search and study selection process for this study. The qualitative content analysis of the 18.5% (72/389) of the studies on neurological diseases revealed the following key themes regarding digital speech signal analysis: (1) health condition and clinical purpose, (2) speech data (speech tasks and speech features), (3) data science approaches and evaluations, and (4) clinical applications. The following subsections provide details on these themes. The details of the included research articles can be found in Multimedia Appendix 3 [42-113].

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for the study screening and selection process.

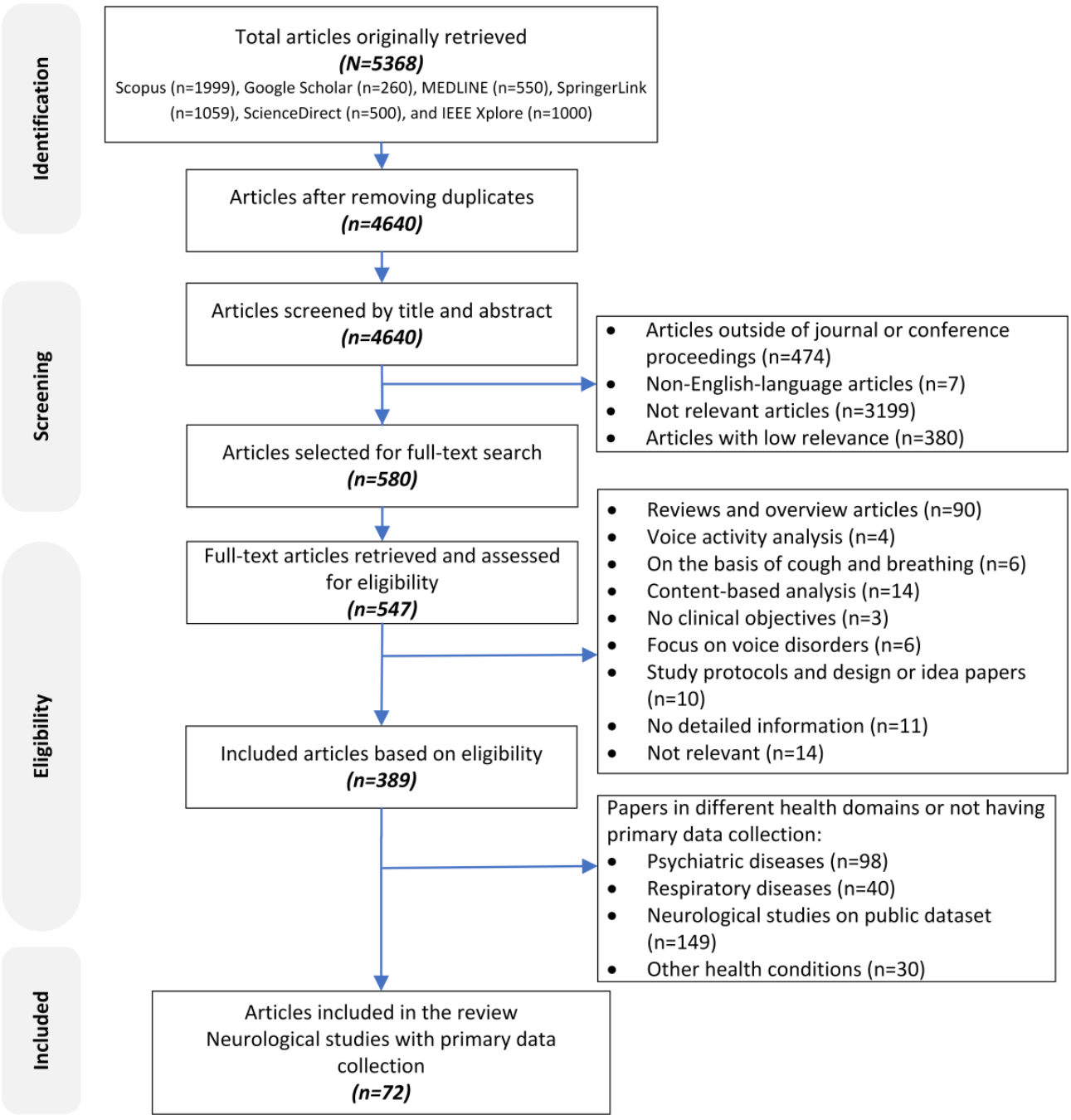

Neurological Conditions and Clinical Purposes

Overview

Many neurological conditions affect the sensorimotor control of speech movements (eg, PD) or cognitive processes, specifically memory and language and perceptual processing (eg, AD and MCI). While thought formulation and motor planning are distinct in speech production [114], alterations in motor planning can occur alongside cognitive impairments, as discussed in various stages of AD [115]. This can cause acoustic changes in the physical speech signal.

Among the neurological diseases, PD (40/72, 56%) was the most investigated disorder, followed by AD and cognitive impairment (12/72, 17%). Other investigated disorders were

XSL•FO
RenderX

MS (5/72, 7%), ALS (4/72, 6%), mild traumatic brain injury (mTBI; 3/72, 4%), Huntington disease (HD; 2/72, 3%), and ASD (2/72, 3%). A total of 4 articles focused on clinical purposes related to apathy (n=1, 25%); intellectual disability (ID; n=1, 25%); essential tremor (ET; n=1, 25%); and a group of central nervous system disorders (CNSDs), including HD and PD (n=1, 25%).

Disease diagnosis, differential diagnosis, severity assessment, and treatment monitoring were the most mentioned clinical purposes in the studies. Figure 4 depicts the distribution of articles with their clinical purposes according to disease categories. Some studies (18/72, 25%) addressed multiple clinical purposes, such as diagnosis and differential diagnosis or diagnosis and severity assessment. Disease diagnosis was widely researched. These studies focused on prevailing clinical challenges such as the lack of definitive objective biomarkers, the need for noninvasive biomarkers, challenges in discriminating diseases with similar symptoms (differential diagnosis), and challenges faced by vulnerable populations such as older adults or rural populations.

Figure 4. Distribution of the main clinical purposes based on disease categories. AD: Alzheimer disease; ALS: amyotrophic lateral sclerosis; ASD: autism spectrum disorder; HD: Huntington disease; MS: multiple sclerosis; mTBI: mild traumatic brain injury; ND: neurological disorder; PD: Parkinson disease.

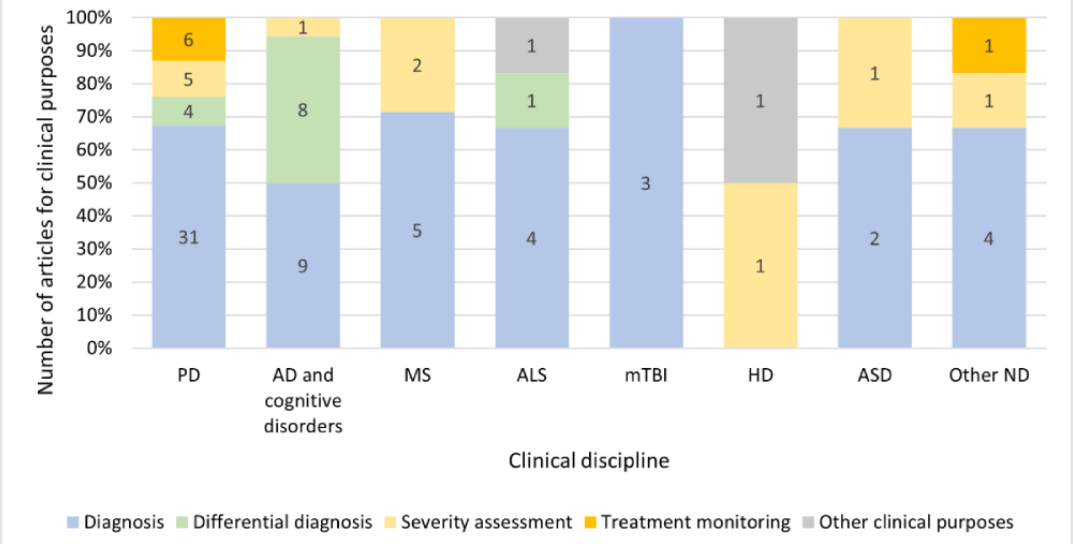

PD Results

PD was the most investigated neurological condition in the studies (40/72, 56%). PD diagnosis, differential diagnosis, treatment monitoring, and severity assessment were explored. Many studies (28/40, 70%) investigated the diagnosis of PD by discriminating parkinsonian speech compared to healthy speech [42-69]. Several studies (3/40, 8%) carried out statistical comparisons of speech features in healthy and PD groups [70-72]. PD severity scores were predicted in the study by Viswanathan and Arjunan [73] for patients with PD and healthy controls. Early detection of PD was investigated in some studies (4/40, 10%) [42,47,57,65], in which participants with early PD were recruited to compare them with healthy participants. Early and mid-aged patients with PD were recruited in the studies by Montaña et al [65] and Wang et al [42], whereas the study by Lim et al [47] focused on patients with early- and advanced-stage PD according to the Hoehn and Yahr staging scale. Patients with PD were assessed in the study by Jeancolas et al [57] if they had been diagnosed with PD within 4 years before the study.

Some studies (4/40, 10%) investigated discrimination of similar-symptom diseases for differential diagnosis. For example, the study by Song et al [44] investigated distinguishing ataxic and hypokinetic dysarthria, which are commonly prevalent in neurodegenerative diseases. Patients with PD and cerebellar ataxia were considered representative cases of diagnosis. A statistical comparison of PD speech and MS speech was performed in the study by Vizza et al [72]. The challenge of differential diagnosis was addressed in the studies by Das et al [74] and Li et al [75] for atypical parkinsonian syndromes (APS) of progressive supranuclear palsy and multiple system atrophy.

In addition to diagnosis, the studies investigated PD severity predictions using the Unified Parkinson's Disease Rating Scale (UPDRS) and Hoehn and Yahr scale. The studies by Viswanathan and Arjunan [73] and Hemmerling and Wojcik-Pedziwiatr [76] predicted UPDRS scores from the speech of patients with PD considering their medication status. Furthermore, Zhang et al [61] and Sztaho et al [67] predicted PD severity level from speech features in relation to the UPDRS and Hoehn and Yahr scale, respectively. The severity of PD symptoms was assessed in the study by Tunc et al [77] through collected speech samples at different points after taking levodopa.

The studies analyzed the speech of patients with PD to assess the impact of antiparkinsonian treatments on speech and voice function. For example, Suppa et al [43] studied speech changes during highly effective and less effective periods from levodopa therapy for PD. PD severity using UPDRS scores was predicted in the study by Hemmerling and Wojcik-Pedziwiatr [76] at different time points after taking levodopa. The effects of dopaminergic medication on speech function were investigated in the studies by Vandana et al [78] and Jain et al [79]. The

XSL•FO
RenderX

impact of assistive speech devices in treating speech impairment was studied by Gaballah et al [80,81] considering these devices' treatment capability outside the clinical facility. In these studies, Gaballah et al [81] investigated discrimination of PD speech and healthy speech under different environments and amplification conditions, whereas Gaballah et al [80] predicted the perceived voice quality of patients with PD with and without assistive speech amplifier devices.

AD and Cognitive Impairment Results

Cognitive decline was assessed through speech analysis, including clinical conditions such as AD, dementia, and MCI. Dementia presents different symptoms at different severity levels of cognitive decline. MCI represents an early stage of cognitive decline without interference with everyday life but can act as a transition stage between healthy aging and dementia when it acts as preclinical AD [116]. AD can evolve over a continuum from normal cognition to MCI due to AD, followed by more severe AD dementia [117]. AD is identified as the most common form of dementia [117].

The studies in this review analyzed speech at different cognitive decline stages for diagnosis, differential diagnosis, and severity assessment (12/72, 17%). The focus was on discriminating impaired cognition from normal cognition. The discrimination of patients with AD [82] and MCI [83-85] from healthy controls was explored. Nagumo et al [85] studied differential diagnosis with global cognitive impairment as they considered patients with MCI, global cognitive impairment, and both forms of cognitive decline. On the other hand, some studies (3/12, 25%) [86-88] investigated both diagnosis and differential diagnosis of AD compared to patients with MCI and healthy participants. Automatic differentiation of patients with MCI and early dementia from healthy controls was studied by Bertini et al [89].

Differential diagnosis of AD and dementia with Lewy bodies was investigated in the study by Yamada et al [90], whereas Sumali et al [91] worked in discriminating between patients with depression and dementia as certain mental disorders (eg, depression) can cause pseudodementia, a temporary decline in mental cognition. While Al-Hameed et al [92] discriminated speech from neurodegenerative diseases, including patients with AD, MCI, and dementia, from speech from functional memory disorder, König et al [93] predicted neuropsychiatric inventory scores in a sample of patients with MCI through speech.

MS Results

MS is a chronic inflammatory disease of the central nervous system that affects cognitive and motor functions causing motor and sensory impairments, visual disabilities, cognitive disorders, and speech and language deficits [118]. Different studies (5/72, 7%) investigated MS diagnosis and severity assessment through speech analysis. Discriminating patients with MS from healthy controls was investigated in several studies (4/5, 80%) [94-97], whereas Fazeli et al [94] explored the relationship between selected speech quality indexes and MS severity. Speech analysis among patients with MS at 2 different disease stages, in addition to comparisons of patients with MS versus healthy patients, was conducted in the study by Vizza et al [98].

ALS Results

ALS is a progressive motor neuron disease that affects upper and lower motor neurons in the motor cortex, the brain stem, and the spinal cord. ALS leads to muscular weakness and spasticity that can result in difficulties with mobility, breathing, and motor speech production [119]. Research on speech in ALS concentrated on identifying differential diagnoses and detecting bulbar involvement (4/72, 6%). Differentiation of speech from 3 participant groups, including patients with ALS with and without bulbar involvement and healthy controls, was conducted in the study by Tena et al [99]. The studies by Illa et al [100] and Likhachov et al [101] investigated discrimination of speech from patients with ALS and healthy participants. The ALS populations in both studies had shown signs of bulbar involvement. Furthermore, Mallela et al [102] investigated the diagnosis of ALS and differential diagnosis of ALS from PD by including ALS patients with a range of speech dysfunction severities.

mTBI Results

Compared to other progressive neurological diseases that occur due to a physical or mental phenomenon originating within the body, mTBIs or concussions are initiated when a person experiences an external force to the head, causing some alteration in brain function [120]. Therefore, the studies in this review on concussion (3/72, 4%) were able to access baseline speech recordings from highly vulnerable populations such as athletes along with the postinjury speech and then aimed to discriminate concussed speech from that of healthy controls [103] and from the individuals' healthy baseline speech [104]. Concussion detection through baseline, postconcussion, and posthealthy speech comparison was also studied by Wall et al [105].

HD Results

HD is a rare severe neurodegenerative disease with a known natural history due to inheritance [106]. One study assessed the severity of HD, whereas the other study investigated the expression of emotions through vocal characteristics in participants with HD and pre-HD using emotion elicitation via speech. Riad et al [106] investigated the prediction of HD severity scores from speech features of a sample of participants with pre-HD and HD. Furthermore, they studied the relationship between speech variables and striatal volumes as well. In addition, the emotion expression capability of patients with HD was studied through speech-based emotion recognition [107].

ASD Results

ASD is a neurodevelopmental disorder that causes difficulties in social communication, particularly in verbal and nonverbal communication [121]. Most people living with ASD exhibit speech and expressive language abnormalities at different levels [121]. The 3% (2/72) of ASD articles in this review focused on discriminating ASD speech from healthy speech [108,109] and estimating autism severity by predicting Autism Diagnostic Observation Schedule scores from speech features [109]. It was noted that comorbid conditions were present in these populations, including attention-deficit/hyperactivity disorder [108] and children with suspicion of ASD, such as children with

XSL•FO
RenderX

other language or developmental delays [109], in addition to children with typical development.

Other Neurological Disease Results

The neurological conditions that were the focus of other studies included ET (1/72, 1%), apathy (1/72, 1%), ID (1/72, 1%), and a broader category of CNSDs (1/72, 1%).

ET is among the most common tremor syndromes, which can encompass voice tremor as well [122]. To complement conventional neurological examination–based assessments, voice tremor in patients with ET was studied by Suppa et al [110] by discriminating the speech of patients with ET who did and did not manifest clinically overt voice tremor. Moreover, they discriminated between patients at baseline and after having medical treatments [110]. Apathy is identified as a motivation disorder that can present in several psychiatric and neurological conditions. Discrimination between patients with and without apathy was done in the study by König et al [111] through speech from patients with mild to moderate neurocognitive disorders. ID is a neurodevelopmental disease similar to ASD that causes cognitive delays in early childhood, resulting in delays in adaptive function, language, and speech [123]. Speech samples from children with typical development and those with ID were compared in the study by Aggarwal et al [112]. Lauraitis et al [113] investigated discrimination of speech impairment in patients with early-stage CNSDs from healthy speech. They included patients with HD, PD, cerebral palsy, stroke, and early dementia in their CNSD population.

Speech Data

Overview

Across the 18.5% (72/389) of neurological studies, speech data were captured through a range of speech tasks, from simple production of sounds or words to spontaneous speech in conversations. The characteristics of the speech tasks impact the reliability of speech features [124]. For example, speech quality features from continuous speech may be less reliable compared to more structured sustained phonation tasks [124]. Therefore, speech tasks and speech features together define speech data in clinical speech analysis [14].

Speech Tasks

Speech tasks across the 18.5% (72/389) of neurological studies can be categorized along a continuum from highly constrained to naturalistic speech elicitation. At the most structured end are sustained phonations and diadochokinetic tasks. In the middle range are reading and picture description tasks, whereas prompted speech tasks elicit the most naturalistic speech production.

Highly constrained tasks frequently appeared in studies of neurological diseases. These included sustained phonation of vowels (/a/, /e/, /i/, /o/, and /u/) and sounds such as “ah,” “eh,” “iuh,” and “iamh.” Similarly structured were oral diadochokinetic tasks, including alternating motion rate and sequential motion rate tasks. Alternating motion rate tasks require rapid repetition of a single syllable (eg, /pa/, /ta/, or /ka/), whereas sequential motion rate tasks involve sequences of different syllables (eg, /pa-ta-ka/).

Semistructured tasks provided a balance between control and naturalistic speech production. Reading tasks typically included reading a set of words, sentences, and passages. Picture description tasks elicited spontaneous speech within a potentially limited vocabulary set associated with the provided visual aid. Activity-related speech included speech fluency tasks, recall and summary tasks, and number-related tasks such as counting and subtraction. On the other hand, prompted speech tasks encourage more free and spontaneous speech within a guided procedure. Monologues on own experiences, conversations including neurological examinations, and spontaneous speech induced by interactive questions were some examples. It was also common for the studies to use multiple speech tasks, combining structured and semistructured tasks. The studies used speech tasks from culturally adapted test batteries such as Thammasat-National Electronics and Computer Technology Center (NECTEC)-Chula’s Thai Language and Cognition Assessment [86].

Overall, structured speech tasks were more widely used for neurodegenerative diseases such as PD and ALS, whereas semistructured speech tasks were common for cognitive impairment–related disorders such as AD, MCI, and dementia. Textbox 2 shows the application of different types of speech tasks in the studies with examples.

XSL•FO
RenderX

Textbox 2. Types of speech tasks used in the included studies.

Structured tasks

- Sustained phonation tasks: the participant is asked to produce a vowel, usually with a steady pitch, for several seconds.
  - Sustained vowel: /a/ for multiple sclerosis (MS) [94,95], /a/ for amyotrophic lateral sclerosis (ALS) [101], /a/ for Parkinson disease (PD) [42,53,54,59,60,62,64,66,71,74,75,77,78], /e/ for PD [42,43,49], and /e/ for essential tremor (ET) [110]
  - Group of vowels (/a/, /e/, /i/, /o/, and /u/): MS [98], ALS [99], and PD [46,55,61,72,76,79]
  - Mix of phonemes and sounds: /s/, "sh," and /f/ (ALS) [102]; continuous phonation task (intellectual disability [ID]) [112]; /a/, /o/, and /m/ (PD) [45]; /a/, /u/, and /m/ (PD) [70,73]; different sounds ("ah," "eh," "iuh," and "iamh") (PD) [51]; and /a/ and /u/ (PD) [56]
- Diadochokinetic tasks: the participant is asked to rapidly repeat alternating syllables (eg, /Pa/, /Ta/, and /Ka/) for several seconds.
  - Monosyllables: /Pa/ and /Ka/ for mild traumatic brain injury (mTBI) [103]; /Pa/ and /Ta/ (PD) [55]; and /Pa/, /Ta/, and /Ka/ (PD) [52,63]
  - Multisyllable sequence: /Pa-Ta-Ka/ for mTBI [103]; /Pa-Pa-Pa/, Ta-Ta-Ta/, Ka-Ka-Ka/, and Pa-Ta-Ka/ and /Ba-Da-Ga/ (ALS) [102]; /Pa-Ta-Ka/, /Pa-Ka-Ta/, /Pe-Ta-Ka/, and /Pe-Ka-Ta/ (PD) [79]; /Pa-Ta-Ka/ (PD) [65,74,75]; and /Pa-Ta-Ka/, /Pe-Ta-Ka/, and /Pa-Ka-Ta/ (PD) [52,63]

Semistructured tasks

- Reading tasks: the participant is provided with a predefined text to read out.
  - Reading a set of words or sentences: set of words (MS) [94,97], set of words and sentences for mTBI [103], set of words and sentences for ALS [100], and reading sentences (PD) [42,43,46,49,50,52,54,55,57,60,63,66]
  - Reading paragraphs: reading a Czech text (MS) [96], Sport Concussion Assessment Tool, Fifth Edition reading paragraph (mTBI) [105], read a short text from predefined poems (central nervous system disorders) [113], read sentences from a French book (Alzheimer disease [AD]) [87], read a short passage (AD) [84], reading texts (PD) [44,52,55,57,58,63,74,79], and reading an article (PD) [47]
- Activity-based speech: The participant is asked to produce speech in response to a semistructured activity.
  - Number-related tasks: digit words (mTBI) [104], counting forward and backward (Huntington disease [HD]) [106], counting [88,90] and subtraction for AD [90], and counting for PD [44,79]
  - Recall and tell: summary of a story (MS) [97], immediate recall and delayed recall of a short film (AD) [83], and imitation of the instructor's voice (ID) [112]
  - Fluency tasks: verbal fluency task (AD) [88,90] and month-remembering task (PD) [79]
- Picture descriptions: the participant is asked to describe a visual aid presented to them.
  - Picture description: cookie theft picture description [90] and picture description [86] for AD, and cookie theft picture description for PD [52,79]
- Prompted speech tasks: the participant is asked to produce a speech with the support of speech prompts.
  - Monologue: talk about the previous day (MS) [97], a monologue (ALS) [100], monologue on a given topic (PD) [63,67,74,75], talk about positive and negative events in life (apathy) [111], talk about a positive event and negative event in life (AD) [93], and talk about the immediate day (AD) [83]
  - Clinical interviews and conversations: interviews by neuropsychologists to describe the last 24 hours and tasks to elicit emotions (HD) [107], conversations during Autism Diagnostic Observation Schedule sessions (autism spectrum disorder) [108,109], conversations (PD) [80,81], and neuropsychological examinations and conversations (AD) [91,92]
  - Spontaneous speech: speech induced through questions (eg, on a picture, about a working day, or about a dream; AD) [89] and 1-minute free talk with an artificial intelligence program (AD) [82]

Speech Features

The studies extracted speech features based on the characteristics of the speech tasks. From a signal processing perspective, the speech features included in the studies can be broadly categorized into 3 main types: fundamental and advanced signal processing–based speech features and audio images (Textbox 3). The primary difference between fundamental and advanced signal processing–based speech features lies in their capacity to correspond to phonetic aspects of speech production. This distinction is important for understanding how these speech features are explained from both a speech science and clinical perspective. Being single-dimensional, both these types of speech features are suitable candidates for statistical analyses, traditional ML approaches, and deep neural networks (DNNs). On the other hand, audio images, a form of multidimensional speech representation, act as candidates for image-specialized neural networks such as convolutional neural networks (CNNs).

XSL•FO
RenderX

Textbox 3. Speech feature types based on signal processing perspective.

Fundamental digital signal processing–based speech features

- Mostly linear signal processing techniques
- Can be described based on signal representation dimension (eg, time domain or frequency domain)
- Can be mapped to a physiological phonetic viewpoint (eg, phonation or articulation) more straightforwardly
- Can be used to create secondary voice indexes, such as voice quality measures (eg, Dysphonia Severity Index)

Advanced digital signal processing–based speech features

- Mostly nonlinear signal processing techniques
- May not be directly mapped to a physiological phonetic viewpoint

Audio images

- 2D representation of speech signals

Speech features can be analyzed through 2 complementary lenses: their representation dimension during speech feature extraction and their physiological phonetic characteristics, which are derived from the field of speech science. Speech signals can be represented and features can be extracted in the time, frequency, time-frequency, and cepstral domains. From the physiological phonetic viewpoint, speech features capture biological and anatomical aspects of speech production, including articulation, phonation, prosody, and speech quality. Within the studies, speech articulation was assessed through time-domain features (eg, AMR [75], diadochokinetic rate, and diadochokinetic period [103]), frequency-domain features (eg, formants [42,76], Bark band energy-based features [42,54], spectral moment, and power spectral moment [76]) as well as cepstral domain features (eg, MFCC [42,45,54]). Similarly, speech phonation was assessed in features from time domain (eg, jitter, shimmer [42,45,62,73,76], amplitude and speed-based glottic cycle features [62,73], Teager-Kaiser energy operator based features [45,77], energy [76]), frequency domain (eg, fundamental frequency [42,48,62,73,76], bark band energy-based features [48], harmonic-to-noise ratio [45,73], noise-to-harmonic ratio [45]) and cepstral domain (eg, MFCC [48,53], linear-frequency cepstral coefficients [53], gamma-tone cepstral coefficients [53], cepstral peak performance [73]). Pause-based features [109] and fundamental frequency [109] are some examples for time domain and frequency domain features for prosody and rhythm assessment.

Different structures have been proposed for categorizing these features. Speech features were categorized as voicing, articulation, and prosodic features in the study by Li et al [75], whereas Wang et al [42] grouped speech features as phonatory, articulatory, prosodic, and cognitive-linguistic features. Phonatory features modeled abnormal patterns in the vocal fold vibrations, whereas articulation features captured deficits in articulatory movements of the lips, tongue, and jaw. Prosodic features, such as speech rate and timing, investigated paralinguistic aspects such as emotions, whereas cognitive-linguistic features considered vocabulary, phrase construction, and word repetitions [42]. The domain of speech feature representation pertains to the complexity and resource demands for feature extraction, whereas phonetic elements relate to speech task characteristics and the biological process involved in speech production.

Speech features are also categorized as linear or base speech features (eg, fundamental frequency, jitter, and shimmer) and nonlinear features, which are mostly derived from advanced signal processing techniques. The studies by Zhang et al [61] and Tunc et al [77] investigated nonlinear speech features such as correlation dimension, recurrence period density entropy, detrended fluctuation analysis, and pitch period entropy. Tunable Q-factor wavelet transform and empirical mode decomposition–based speech features are some more advanced speech features [77]. Different spectrograms, including mel spectrograms [44,54], linear spectrograms [54], and constant-Q transform spectrograms [54] are examples of speech image representations assessed in the included studies.

Secondary voice indexes were another type of speech feature available in the studies. They were indexes defined from other primary speech features. Example measures included the Dysphonia Severity Index, formant centralization ratio, and vowel metrics. The Dysphonia Severity Index is calculated from phonation time, jitter, fundamental frequency, and intensity, whereas the latter 2 are calculated from the measures of formants of vowels [72,94]. Some studies used standard speech feature sets, for example, the extended Geneva Minimalistic Acoustic Parameter Set [107], INTERSPEECH2016 Computational Paralinguistics Challenge speech feature set [49,110], and the emobase feature set [86]. Speech feature embeddings derived from a deep learning (DL) approach were also a type of speech feature representation assessed in the included studies [50,57].

Data Science Approaches

Overview

Descriptive and predictive studies were included among the 18.5% (72/389) of neurological studies. Descriptive studies assessed the relationship between clinical variables and speech features (eg, whether speech features are associated with a disease or its severity), whereas predictive studies investigated the estimation of clinical variables from speech features (eg, predicting disease severity from speech features). Statistical analysis was largely applied in descriptive studies, whereas

XSL•FO
RenderX

predictions were made through traditional ML and DL approaches.

Statistical Analysis: Relationship Between Speech Features and Clinical Measurements

Univariate, bivariate, and multivariate analyses of speech features were observed across the 18.5% (72/389) of neurological studies. Univariate analyses investigated descriptive measures and the statistical significance of speech features in differentiating pathological and healthy groups. Descriptive statistics of speech features within first hour and after 12 hours of medication of PD [78] and at 2 stages of MS [98] are some examples.

Further statistical significance tests showed the differences between speech features, mainly between healthy and impaired speech due to neurological health conditions. Several statistical significance tests were applied depending on the relationship between discriminating groups and the normality of the variables. For example, Suppa et al [110] statistically compared speech with and without voice tremors using the Student *t* test, whereas the paired Student *t* test was used to compare speech features before and after medication. To differentiate AD, dementia with Lewy bodies, and healthy speech, Yamada et al [90] compared speech features between groups after controlling for medication using 1-way analyses of covariance. Both studies explored the relationship between speech features and disease characteristics before building predictive models. The studies used bivariate analyses to explore the association between speech features and clinical variables, such as the correlation between speech measures and UPDRS scores for PD severity [73,76,77]. The study by König et al [111] also assessed the correlation between speech properties and the Apathy Inventory subscales to assess the properties of apathetic speech in people with cognitive disorders. Multivariate analysis was applied to simultaneously explore the relationship among clinical variables, speech features, and other confounding factors such as age and sex. For example, Svoboda et al [96] applied multiple linear regression of speech features, age, and sex to differentiate MS from healthy speech.

Table 1 summarizes the statistical analysis approaches applied in the selected studies.

XSL•FO
RenderX

Table 1. Summary of statistical analysis approaches in the studies with examples.

Technique	Examples
Univariate analysis	
Statistical comparison of speech features between 2 independent groups	• Mann-Whitney U test to compare speech features between patients with MS[a] and healthy groups [94] • Kruskal-Wallis test to compare apathetic and nonapathetic speech [111] • Independent Student t test to compare speech features between patients with MS and healthy groups [95,98] and between patients with and without voice tremor [110] • Kolmogorov-Smirnov 2-sample test to compare speech features in MS and healthy groups [96]
Statistical comparison of speech features between 2 related groups	• Paired Student t test to compare the speech variables of the same patients with and without medication [110]
Statistical comparison of speech features among ≥3 independent groups	• 1-way ANOVA to explore how diseased status (AD[b], DLB[c], and healthy) impacted speech features to support differential diagnosis [90] • 2-way ANOVA to examine the effects of the dementia type (AD and DLB) and disease stage (MCI[d] and dementia) on speech features [90]
Bivariate analysis	
Statistical association between clinical variables and speech features	• Spearman correlation coefficient to assess the relationship between acoustic features and neurological status for MS [94] and assess the relationship between acoustic features and Apathy Inventory subscales [111] • Statistical association between ADOS[e] score and speech features via Pearson correlation coefficient [109]
Multivariate analysis	
Statistical comparison among multiple groups considering confounders on speech features	• Generalized linear regression to assess the statistical significance of speech features, age, and sex on differentiating MS against healthy populations [96]
Dimensionality reduction of speech features	• Application of principal-component analysis and observation of biplots to understand speech feature clusters in 3 discriminative groups related to ALS[f] [99]
Speech feature cluster analysis	• k-means clustering to unveil patterns within speech features obtained from patients with dementia and depression [91]

[a]MS: multiple sclerosis.

[b]AD: Alzheimer disease.

[c]DLB: dementia with Lewy bodies.

[d]MCI: mild cognitive impairment.

[e]ADOS: Autism Diagnostic Observation Schedule.

[f]ALS: amyotrophic lateral sclerosis.

ML Approaches

The studies used traditional ML, DL, and hybrid approaches to predict clinical variables from speech features. Disease diagnosis was formulated as a binary classification, whereas differential diagnosis was extended into multi-class classifications. The severity assessment was primarily treated as a regression problem, whereas treatment monitoring involved either regression or classification of patients based on medication status. Classification problems were the most commonly addressed.

Among traditional ML classifications, the following were the most common algorithms: support vector machine, k-nearest neighbor, random forest, decision tree, Extreme Gradient Boosting, multilayer perceptron, and logistic regression. Linear regression, support vector regression, random forest regression, and artificial neural network regression were among the traditional ML regression techniques, for example, in Autism Diagnostic Observation Schedule score prediction for autism severity assessment [109] and prediction of neuropsychiatric inventory scores [93].

DL approaches were applied as deep feature extractors and end-to-end predictors. In deep feature extractors, DL capability was explored in self-extracting efficient feature representations from speech features in a supervised or unsupervised manner. The studies retrained special DL models and built their own models as feature extractors. Example deep speech feature extractors included the implementation of a standard DNN in the study by Gosztolya et al [97] and an autoencoder network in the study by Bertini et al [89]. Regarding the development of end-to-end DL models, recurrent neural network models and CNNs were among the most promising ones. Different architectures such as CNN with a modified Hybrid Mask U-Net architecture with an adaptive custom loss function for PD assessment [46] and bidirectional long short-term memory neural networks [105] were implemented. Hybrid architectures such

XSL•FO
RenderX

as CNN–long short-term memory networks [102] and personalized convolutional recurrent neural networks [79] were also assessed in the studies. Transfer learning was applied to train computer vision–based approaches for clinical speech analysis. Examples included CNN14 [44] and AlexNet-based CNN for PD assessment [53].

There were fewer research attempts to explore DL approaches for clinical score prediction. Autism severity score prediction using DNN and CNN [109] and PD severity prediction using DNN [67] were among the few examples. Table 2 summarizes the ML and DL approaches applied in the selected studies.

XSL•FO
RenderX

Table 2. Summary of machine learning (ML) and deep learning (DL) approaches used in the studies with examples.

Technique	Examples
Traditional ML: classifiers	
Boosting technologies	• XGBoost[a] for MS[b] diagnosis [96], AD[c] assessment [82], and PD[d] assessment [50,77] • AdaBoost[e] for AD and cognitive impairment assessment [86,92] and AdaBoost for PD assessment [47] • LightGBM[f] for PD assessment [47,50]
KNN[g]	• MS diagnosis [96], ALS[h] assessment [101], AD assessment [87], and PD assessment [42,47,51,54,58,60,67,69]
Simple feed-forward neural networks	• MS diagnosis [96], PD assessment [54], and ID[i] assessment [112] • AD assessment [86], cognitive disorder assessment [92], and PD assessment [49,51,60,61,67]
RF[j]	• MS diagnosis [96], ALS assessment [99], and HD[k] assessment [107] • ID assessment [112], AD and cognitive disorder assessment [82,83,86,92], and PD assessment [47,50,66]
SVM[l]	• MS diagnosis [96], concussion diagnosis [104], ALS assessment [99,100,102], ASD[m] assessment [108], ID assessment [112], ET[n] detection [110], AD and cognitive impairment assessment [83,86-88,90-92], and PD assessment [43,45,47,49-51,53,54,58-61,63,67-69,74,75]
LR[o]	• Concussion diagnosis [103], diagnosis of apathy [111], ALS assessment [99], AD assessment [82,86], and PD assessment [42,47,59,61,75]
NB[p]	• ALS assessment [99], AD and cognitive disorder assessment [83], and PD assessment [42,47,49,54,60,61]
DT[q]	• AD assessment [87] and PD assessment [47,54]
Traditional ML: regressors	
Linear regression	• ASD severity prediction using multiple linear regression [109] • NPI[r] score prediction using linear regression with L1 regularization [93] • UPDRS[s] score prediction for PD assessment [61,76] • HY[t] score prediction for PD assessment [67]
SVR[u]	• ASD severity prediction [109] • NPI score prediction [93] • UPDRS score prediction for PD assessment [61,73,76] • HY score prediction for PD assessment [67]
RF regression	• UPDRS score prediction for PD assessment [73,76]
ANNs[v]	• Prediction of speech impairment in ET [110] • Prediction of the likelihood of speech impairment in PD [43] • HY score prediction for PD assessment [67]
DL and hybrid models: classifiers	
DNNs[w]	• DNN+SVM model for MS diagnosis [97] • ALS assessment [100,102] • AD and cognitive impairment assessment [84] • X-vector for PD assessment [57] • DNN for PD assessment [50,67]
RNNs[x]	• BiLSTM-A[y] for concussion diagnosis [105] • BiLSTM[z] for CNSD[aa] diagnosis [113] and PD assessment [54]
CNNs[ab]	• AlexNet-based CNN for PD assessment [53] • CNN14 for PD assessment [44] • CNN (modified Hybrid Mask U-Net architecture with an adaptive custom loss function) for PD assessment [46] • CNN (ResNet18 architecture) for PD assessment [52] • CNN for PD assessment [54,58,63]

XSL•FO
RenderX

Technique	Examples
CNN+LSTM[ac]	• CNN+LSTM model for ALS assessment [102] • CRNN[ad] for PD assessment [79]
Autoencoder	• auDeep+MLP[ae] network for early dementia assessment [89]
DL and hybrid models: regressors	
DNNs	• ASD severity prediction [109] • DNN regression [80] • HY score prediction for PD assessment [67]
CNNs	• ASD severity prediction [109]

[a]XGBoost: Extreme Gradient Boosting.
[b]MS: multiple sclerosis.
[c]AD: Alzheimer disease.
[d]PD: Parkinson disease.
[e]AdaBoost: Adaptive Boosting.
[f]LightGBM: Light Gradient-Boosting Machine.
[g]KNN: k-nearest neighbor.
[h]ALS: amyotrophic lateral sclerosis.
[i]ID: intellectual disability.
[j]RF: random forest.
[k]HD: Huntington disease.
[l]SVM: support vector machine.
[m]ASD: autism spectrum disorder.
[n]ET: essential tremor.
[o]LR: logistic regression.
[p]NB: naive Bayes.
[q]DT: decision tree.
[r]NPI: neuropsychiatric inventory.
[s]UPDRS: Unified Parkinson's Disease Rating Scale.
[t]HY: Hoehn and Yahr scale.
[u]SVR: support vector regression.
[v]ANN: artificial neural network.
[w]DNN: deep neural network.
[x]RNN: recurrent neural network.
[y]BiLSTM-A: bidirectional long short-term memory–attention.
[z]BiLSTM: bidirectional long short-term memory.
[aa]CNSD: central nervous system disorder.
[ab]CNN: convolutional neural network.
[ac]LSTM: long short-term memory.
[ad]CRNN: convolutional recurrent neural network.
[ae]MLP: multilayer perceptron.

Some studies (24/72, 33%) integrated multiple data science approaches. They first applied statistical techniques to screen and filter speech features before feeding them into ML-based prediction models. Application of unsupervised learning techniques was the least common, with only one study exploring unsupervised clustering. The study applied statistical analysis, unsupervised learning, and supervised classification to address the classification of patients with depression and dementia based on speech, demonstrating comprehensive use of data science approaches in the speech analysis pipeline [91].

Model Evaluations

The analytical performance of algorithms was evaluated based on the model's prediction performance, generalizability, robustness, biases, and fairness. The classification model's prediction performance was assessed through a set of performance metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve. In regression analysis, the root mean square error score was the most widely reported performance metric.

XSL•FO
RenderX

To ensure the model's generalization and robustness, cross-validation with speaker independence was applied to prevent overfitting. Variations of k-fold validations were applied, with *k*=5 and *k*=10 being the most common. In training and testing, speaker independence maintains all speech recordings of the same speaker only in either the training or testing set without splitting between the 2 to avoid overoptimistic performance from information sharing. The studies adapted speaker-independent cross-validation strategies with simple k-fold validation when utterance-level features were considered. Some studies (14/72, 19%) applied leave-one-out cross-validation to increase the training dataset, preserving speaker independence. In this case, data from one participant were kept for testing, whereas all remaining data were used for model training [45,67,83,93]. When low-level speech features such as frame-level speech features were analyzed, speech from the same utterance was treated as a separate input, making speaker independence challenging. To address that challenge, Al-Hameed et al [92] applied a leave-one-group-out cross-validation with segment-level speech features, whereas simple k-fold validation was used with utterance-level features.

The biases and fairness of model predictions were mainly evaluated through cross-corpus testing and confounder assessments. There were several cross-corpus test scenarios in the studies. Training with aged-matched groups and testing on different age groups [91] and recruiting different training and testing cohorts [47] were some examples of maintaining different corpora within data collection. Moreover, the studies combined public and private datasets from different ethnicities and different speech tasks to improve the heterogeneity of the sample populations [46,58,61,63,77]. Speech corpora with different speech recording qualities were also considered [48,52,55,57,59].

The impact of confounding factors such as age and gender were addressed differently in the studies. Mainly, the studies used age-matched disease and control groups in the experiments. Meanwhile, age mismatch between diseased and healthy groups was addressed through age correction of speech features [99]. Application of age and gender as features in their models [84,96] as well as evaluation of model performance for each gender was also conducted [111].

Figure 5 shows the model evaluation criteria commonly considered in the studies.

Figure 5. Model evaluation criteria used in the studies on neurological populations. AUROC: area under the receiver operating characteristic curve; LOGOUT: leave-one-group-out cross-validation; LOOCV: leave-one-out cross-validation; RMSE: root mean square error.

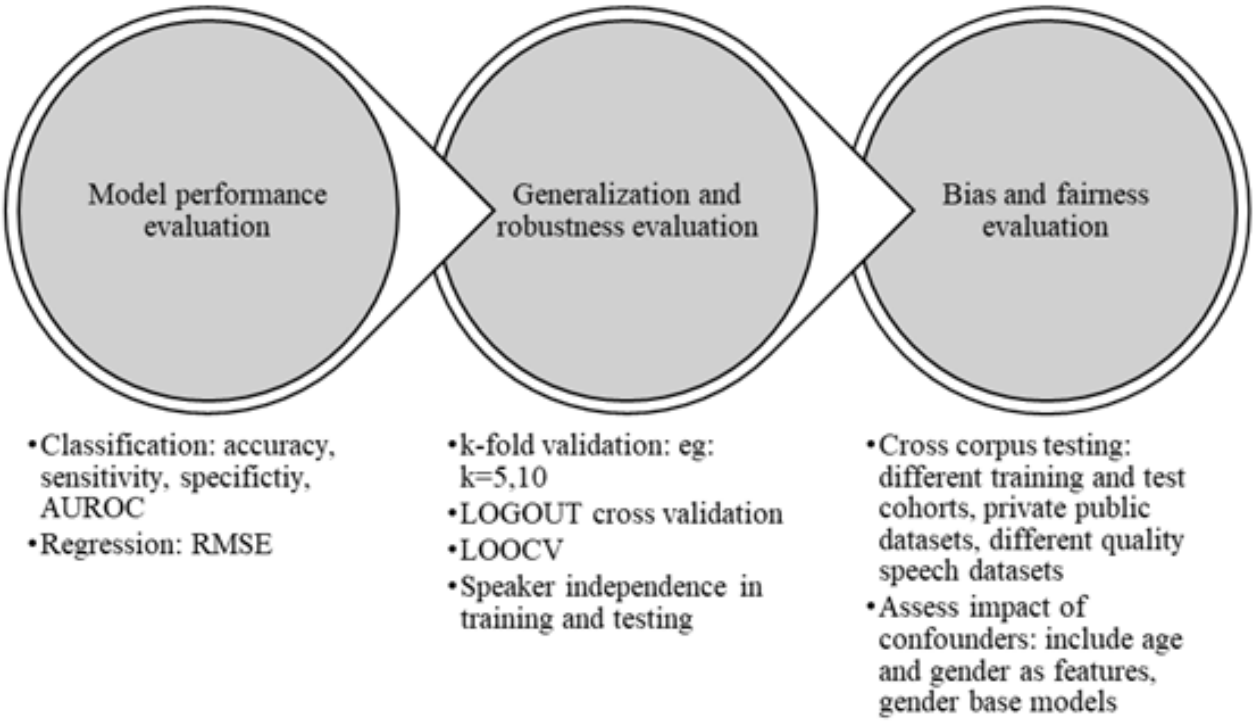

Only one study evaluated model predictions within a clinical setting. To discriminate between ataxia and hypokinetic dysarthria, Song et al [44] compared predictions of their artificial intelligence models and clinical decisions from a group of neurological resident physicians.

Clinical Applications

Several studies (6/72, 8%) extended their research into applications in the PD, ALS, and mTBI domains. For example, Rahman et al [50] proposed a web-based framework to record and analyze speech for PD screening. Meanwhile, Zhang et al [61] deployed their real-time speech analysis tool for PD diagnosis and severity assessment within a mobile app called No Pa for both Android and iOS. In addition, Likhachov et al [101] created a prototype of a mobile app named ALS Expert to assess the voice function of ALS. The studies used mobile apps for their speech data collection to demonstrate the feasibility of remote assessments and remote monitoring. For

XSL•FO
RenderX

example, Vasquez-Correa et al [52] collected smartphone-based speech data via an app called Apkinson, and Laganas et al [48] used a mobile app named iPrognosis to collect speech data from several countries and adopted an on-device feature extraction process. Both studies focused on PD assessment. A lightweight mobile app for speech data collection for mTBI assessment was also developed in the study by Daudet et al [103] to demonstrate the feasibility of on-device speech feature extraction, highlighting the importance of speech analysis at resource constraint devices to support on-field assessments.

Discussion

Principal Findings

On the basis of our review, speech analysis has emerged as a valuable tool across neurological, psychiatric, and respiratory diseases, with a particular focus on PD, AD, and cognitive impairment. Research in speech analysis has spanned the continuum of patient care, addressing diagnosis, differential diagnosis, severity assessment, and treatment monitoring. However, much research was conducted on diagnosing diseases by classifying healthy and diseased populations. Prediction of continuous clinical variables such as clinical scores was explored through regression. Statistical analysis was applied to assess the reliability of speech features and their relationship with clinical variables.

However, there are several limitations in the current research approaches to make their transition into health care settings. Conducting in controlled settings with homogeneous populations, typically single-ethnicity, single-center cohorts was observed as a common limitation across the studies. Early-stage patients—often the most challenging to diagnose—were underrepresented, and longitudinal studies across disease progression were scarce. However, the presence of studies from countries such as Italy, Thailand, China, Spain, and India with speech analysis in local languages is encouraging. This is appealing as speech features may or may not have similar meanings across different languages due to phonetic differences. Therefore, speech assessments within different populations and languages are recommended to assess the generalizability of speech analysis to a wider population [46]. Embedding heterogeneity into study populations and speech recordings through cross-corpus testing, using multiple speech tasks, and using different speech recording conditions were presented in the studies.

The diversity in the speech tasks used and extraction of speech features reflects both the field's complexity and its opportunities. The studies used multiple speech tasks as separate modalities, ranging from highly structured exercises to naturalistic speech. The speech tasks within the studies could largely be categorized into reading tasks, sustained phonations, diadochokinetic tasks, activity-related speech tasks, picture descriptions, and prompted speech tasks. The selection of speech tasks was guided by disease characteristics as structured speech tasks were common for neurodegenerative diseases such as PD and ALS, whereas semistructured speech tasks were common for cognitive impairment–related disorders such as AD, MCI, and dementia. Sustained vowel phonation and diadochokinetic tasks were widely identified in the literature for their ability to represent early signs of neurological diseases [125] and generally demand lower cognitive demands than semistructured speech tasks such as reading tasks. Sustained vowels can be applied across different languages as well due to less linguistic loading related to dialect, region, and language [126]. However, prompted speech tasks such as monologues might be impacted by personality traits, emotional status, sociocultural norms, and the ability to tell stories [124]. In speech feature extraction, the studies reported various associated parameters such as speech frame sizes, window overlapping, audio preprocessing stages, and signal processing algorithms, but no common format exists to present speech feature characteristics to act as biomarkers. When positioning speech features as a biomarker, it is recommended to report parameters associated with speech-processing techniques as they affect the accuracy and robustness of acoustic measures [124]. Adhering to the established guidelines for speech recording and analysis [127,128] along with transparent reporting of the relevant instrumental and computational specifications can lead to reliable data collection and analysis. This facilitates comparisons across studies, accelerates knowledge sharing across interdisciplinary fields, and supports research advancement and reproducibility.

Among data science approaches, statistical analysis approaches were applied to assess the clinical utility of speech features, but the exploration of unsupervised approaches was less present. Traditional ML approaches dominated predictive studies, but there are growing efforts in applying DL approaches with complex network architectures, transfer learning approaches, and audio image analysis. Although transfer learning from computer vision to audio images was explored, there is a void in transferring knowledge between speech analysis in different languages. Despite their limited presence, DL implementations encompassed end-to-end implementations, deep acoustic feature extraction, and transfer learning, which represent the main stems of DL approaches [129]. Among the end-to-end implementations, CNN, long short-term memory, and hybrid implementations were used, but it was noted that no transformer-based neural network models were included in the selected studies. However, recent literature provides evidence of the application of vision transformer–based models to analyze speech signals for neurological diseases. Differentiating PD severity levels using sustained vowels and Swin Transformer [130]; differentiating speech from neurological diseases, including PD and MS, from healthy speech by retraining Google's Vision Transformer Base model [131]; and differentiating PD speech from spinocerebellar degeneration speech by retraining the Patchout faSt Spectrogram Transformer [132] are some examples of vision-based transformers in speech analysis. It is noteworthy to mention that exploration of advanced DL approaches may improve the prediction performances. Additionally, incorporating a range of data science methods such as statistical analysis and unsupervised learning for hypothesis development and testing, conducting confounding factor analysis, and developing interpretable ML models can help researchers present their findings more interpretably and comprehensively to a wider audience.

XSL•FO
RenderX

Model evaluations in the included studies mainly considered analytical validations on prediction performance, generalization and robustness, and biases and fairness. Lack of clinical validations and narrow model assessment scopes can hinder translating research outcomes into clinical practices in the foreseeable future. Only a few studies (6/72, 8%) connected their investigations to clinical supportive applications. However, the development of applications can support further studies and usability assessments of prospective services within clinical and natural environments. In clinical settings, speech assessments can be conducted following standard speech recording protocols in optimal acoustic conditions to convey clinical insights for decision support. Routine health examinations might also be integrated with speech-based assessments to provide cost-effective longitudinal evidence in patient monitoring. Implementation of smartphone apps for speech collection, speech analysis, or transfer of speech data to remote clouds enables speech analysis in natural environments to empower telemedicine platforms. Opportunities exist in using telemedicine in neurological health care to ease the financial and accessibility burdens, such as in the rehabilitation of brain injuries and on-site concussion assessments [133]. Extending research into testable applications helps researchers explore challenges when converting research outcomes into clinical practice. Therefore, we believe that implementing a systematic research process leading to a clinical utility analysis could significantly accelerate advancements in this field. To support this, we organized the key concepts of speech analysis identified in this review into a research framework, aiming to provide more comprehensive guidance for future research.

Proposed Research Framework

Overview

We adapted the research process proposed in the study by Offermann et al [134] to operationalize research in design science. The proposed research process is structured in 3 main phases—problem identification, solution design, and evaluation—supporting both quantitative and qualitative research methods [134]. To build our research framework, we identified key activities and desired outcomes at each phase, focusing on research using primary speech data collection. Table 3 shows the research process, including proposed activities under each subprocess of the main stages of problem identification, solution design, and evaluation. Expected outcomes are proposed at each stage to guide the research progress.

XSL•FO
RenderX

Table 3. Proposed research framework for speech analysis for clinical assessments.

Subprocesses and activities	Outcome
Problem identification	

• Literature research and expert interviews
  • Identify clinical challenges of interesting health conditions
  • Characterize the speech impairment associated with the health condition through evidence from clinical science, speech science, and empirical research
• Identify the problem
  • Identify a research gap
• Pre-evaluate relevance
  • Define the research objective
  • Define research questions
  • Develop a research hypothesis
	• Selected health conditions and clinical purposes
	• Potential clinical applications
	• Characteristics of speech impairment
	• Research hypothesis

Solution design	

• Literature research
  • Identify standard clinical assessments
  • Identify speech assessments associated with specific health condition or similar health conditions
  • Identify study design approaches and research techniques
  • Identify suitable public or accessible datasets
• Design artifacts
  • Define ML[a] problem for the focused clinical purpose
  • Develop the study design of the research
  • Carry out data collection and analysis
	• Study population
	  • Participant recruitment and selection
	• Data collection protocol
	  • Ground truth
	  • Demographic data
	  • Speech tasks and speech features
	• Data analysis
	  • Primary analysis to build speech feature dataset
	  • Secondary analysis to assess the relationship between speech features and clinical variables
	  • Tertiary analysis to build predictive models

Evaluations	

• Refine hypothesis
  • Refine hypothesis based on data analysis
• Laboratory experiments
  • Internal validation (evaluate predictive performance; extend evaluations beyond predictive performance [eg, evaluate biases, fairness, and explainability])
  • External validation (with public dataset if available)
• Case study and expert survey
  • Case study in a clinical environment
  • Qualitative feedback from clinical and speech science experts
	• Internal validation of results
	• External validation of results
	• Case study and clinical feedback
	• Limitations and further improvements

[a]ML: machine learning.

Problem Identification

Problem identification defines the diseases of interest, clinical purpose, and potential clinical applications. Furthermore, it helps characterize the speech impairment associated with the disease to formulate research hypotheses.

As per this review's findings, digital clinical speech analysis addresses different clinical problems. For example, speech features are being researched as objective biomarkers for disorders such as AD, mTBI, and PD in cases in which a definite biomarker is not available. Moreover, speech analysis aims to empower the remote monitoring of patients with PD who are mostly older adults. Speech analysis tries to differentiate overlapping symptoms in different diseases, such as cognitive impairments and symptoms of natural aging. Therefore, researchers can focus on a particular health condition considering existing clinical challenges and evidence of associated speech impairments [114]. To explore clinical challenges and potential speech changes from diseases, existing literature and experiences from health care experts such as clinicians, speech-language pathologists, and frontline health professionals can be referred. Experiences from patients might also be beneficial when the target end applications are considered. Findings of this phase identify a research gap and define research objectives and research questions. Furthermore,

XSL•FO
RenderX

the researcher can qualitatively characterize speech impairments and define research hypotheses for empirical research.

Solution Design

In the solution design phase, specific literature research on current clinical practices and associated clinical speech assessments is beneficial to identify the ground truth [114] and potential speech characteristics to focus on. Study design approaches and state-of-the-art data analysis techniques can be used to ensure the scientific validity and novelty of the research. Clinical problems can be mapped into a predictive modeling problem, and an appropriate speech data collection plan can be developed. Data collection should include patients' demographics, assessment of health conditions through standard clinical measures, and speech recordings. Furthermore, it should define speech recording instances, conditions, recording equipment, and speech tasks. Recording instances define when to collect speech from the participants, such as before or after medication. Recording can be done in either controlled or uncontrolled environments based on the research objectives. For example, if the research objective is the remote monitoring of patients, recordings in an uncontrolled environment would be more appropriate. Speech tasks should be carefully selected to capture appropriate speech characteristics associated with the health condition.

We recommend a comprehensive analysis of speech features to conduct association analysis to clinical parameter predictions. The data analysis workflow can be divided into three main stages—primary analysis, secondary analysis, and tertiary analysis—to (1) explore participant characteristics and extract suitable speech features from speech signals, (2) examine the relationship between speech features and clinical variables, and (3) develop and optimize ML models to predict clinical variables from speech features.

The primary analysis quantifies speech features from the speech signals following appropriate audio preprocessing. Noise reduction, down sampling, and silent period removals are some examples of audio preprocessing steps. Different speech features represent different aspects of speech production. For example, energy-based speech features represent respiratory function, whereas pitch-based features represent phonatory function. On the basis of the anticipated speech impairment and speech task characteristics, a set of representative speech features could be extracted. The secondary analysis then encompasses the relationship between speech features and clinical variables through descriptive statistics, statistical comparisons, association mining, and unsupervised data explorations.

In tertiary analysis, research can exploit speech variability within diseases to build predictive models, typically through classification or regression. Insights derived from secondary analysis can be integrated into predictive model building through feature selection and patient clustering. In addition, the impact of confounding factors such as age and gender should be considered when developing predictive models. Algorithm-dependent advanced ML strategies such as feature selection, data augmentation, and transfer learning can be explored to improve model predictions.

Evaluation

Once the solution has reached a satisfactory state, the evaluation of the proposed solutions or approaches is recommended [134]. During evaluations, the hypothesis can be refined to a more precise level based on the data analysis. For example, data analysis might highlight speech impairments in a particular dimension of speech production, such as phonation or articulation, or a particular stage of disease. Therefore, specific hypotheses can lead to detailed insights. Moving forward, internal validations should extend to different dimensions of model performance, and evaluations can include surveys and case studies for clinical utility assessments.

Through a set of comprehensive experiments and evaluation metrics, internal validations should extend beyond prediction performance to include assessments of biases and fairness, reliability, and explainability of predictions to address challenges that persist in speech analysis [28]. With the aid of publicly available datasets or, if feasible, with external cohorts, external validation of predictions can be conducted. Few case studies can be carried out in a clinical environment to present model assessments to clinical experts and obtain their feedback on model usability.

Finally, study results can be presented including experimental results for internal validations, external validations, and feedback from expert observations. We believe that the research outputs will be more competent and thorough and contribute to long-term research directions, extending the short-term results of the specific research scope.

Limitations

We acknowledge that our research study has certain limitations. Among the eligible studies, we reviewed only 72 articles in the neurological domain. Although we considered only content-independent speech analysis, content analysis can also be relevant for certain clinical conditions. Furthermore, our main findings were synthesized from studies that conducted primary speech data collection. Nonetheless, studies on other clinical disciplines and studies that used publicly available datasets also contribute to the advances in the field. We confined the main concepts to diseases, clinical outcomes, speech tasks, speech features, data science approaches, evaluations, and clinical applications. The highly technical and algorithm-dependent speech feature extraction and feature selection methods were not covered in this review. However, such factors also remain crucial in speech analysis.

Conclusions

This review discussed the main concepts within the growing research field of speech analysis for clinical decision support. The principal findings were presented from both clinical and technical perspectives. The clinical context was addressed through diseases, clinical purposes, and clinical applications, whereas technical aspects were addressed through speech tasks, speech features, data science approaches, and model evaluations.

The main contribution of this research can be summarized as (1) carrying out a comprehensive and extensive systematic scoping literature review followed by qualitative content analysis

XSL•FO
RenderX

on digital clinical speech analysis and (2) presenting a research framework on speech analysis for clinical decision support.

The findings of this research reflect the potential of speech analysis for clinical decision-making and the contribution of data science approaches. Among clinical disciplines, neurological diseases have gained major interest, with PD being the most popular. Interestingly, research efforts are expanding beyond English-speaking populations, but more studies including less represented ethnicities and languages are much warranted. The lack of longitudinal studies also remains as a research gap. Designing experiments to address challenging clinical decision scenarios such as prognosis or early detection might be more appealing for clinical environments. Moreover, given the technical differences in speech features, an interpretable presentation of speech features as a digital biomarker would accelerate research progression and reproducibility. Integration of different data science techniques, including statistical analysis and unsupervised and supervised learning, can make data analysis more comprehensive and interpretable. Model evaluations should expand beyond analytical validations and include more comprehensive evaluations, including clinical utility assessments.

On the basis of the findings of this study, we proposed a research framework for primary research on speech analysis for clinical decision support. We encourage studies to adhere to design science research methodology by integrating both quantitative and qualitative research methods.

Acknowledgments

This research is part of an ongoing international collaboration supported by the Auckland University of Technology Faculty of Design and Creative Technologies and the Auckland University of Technology School of Engineering, Computer and Mathematical Sciences. This research is partially supported by Auckland University of Technology Faculty of Design and Creative Technologies Contestable Grant 2023 and 2024 (principal investigator: SM).

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 3.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 84 KB - jmir_v27i1e63004_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 17 KB - jmir_v27i1e63004_app2.docx]

Multimedia Appendix 3
Overview of the studies included in this review.
[DOCX File , 102 KB - jmir_v27i1e63004_app3.docx]

References

1. King LS. Signs and symptoms. JAMA 1968 Oct 28;206(5):1063. [doi: 10.1001/jama.1968.03150050051011]
2. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nat Rev Neurol 2020 Jul 15;16(7):381-400 [FREE Full text] [doi: 10.1038/s41582-020-0362-2] [Medline: 32541893]
3. Rissardo JP, Caprara AL. Parkinson's disease rating scales: a literature review. Ann Mov Disord 2020;3(1):3. [doi: 10.4103/aomd.aomd_33_19]
4. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord 2017 Jan 15;208:191-197. [doi: 10.1016/j.jad.2016.10.019] [Medline: 27792962]
5. Strimbu K, Tavel A, Tavel MD. What are biomarkers? Curr Opin HIV AIDS 2010 Nov;5(6):463-466 [FREE Full text] [doi: 10.1097/COH.0b013e32833ed177] [Medline: 20978388]
6. Powell D. Walk, talk, think, see and feel: harnessing the power of digital biomarkers in healthcare. NPJ Digit Med 2024 Feb 24;7(1):45 [FREE Full text] [doi: 10.1038/s41746-024-01023-w] [Medline: 38396034]
7. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig Otolaryngol 2020 Feb 31;5(1):96-116 [FREE Full text] [doi: 10.1002/lio2.354] [Medline: 32128436]

XSL•FO
RenderX

8. Flanagan O, Chan A, Roop P, Sundram F. Using acoustic speech patterns from smartphones to investigate mood disorders: scoping review. JMIR Mhealth Uhealth 2021 Sep 17;9(9):e24352 [FREE Full text] [doi: 10.2196/24352] [Medline: 34533465]
9. Madanian S, Parry D, Adeleye O, Poellabauer C, Mirza F, Mathew S. Automatic speech emotion recognition using machine learning: digital transformation of mental health. In: Proceedings of the 2022 Pacific Asia Conference on Information Systems. 2022 Presented at: PACIS '22; July 5-9, 2022; Sydeny, Australia p. 18 URL: https://aisel.aisnet.org/pacis2022/45
10. Deepa P, Khilar R. Speech technology in healthcare. Meas Sens 2022 Dec;24:100565. [doi: 10.1016/j.measen.2022.100565]
11. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for health: the use of vocal biomarkers from research to clinical practice. Digit Biomark 2021 Apr 16;5(1):78-88 [FREE Full text] [doi: 10.1159/000515346] [Medline: 34056518]
12. Docio-Fernandez L, García MC. Speech production. In: Li SZ, Jain A, editors. Encyclopedia of Biometrics. Cham, Switzerland: Springer; 2015:1493-1498.
13. Solomon NP. Evaluation of speech. In: Weissbrod PA, Francis DO, editors. Neurologic and Neurodegenerative Diseases of the Larynx. Cham, Switzerland: Springer; 2020:67-77.
14. Robin J, Harrison JE, Kaufman LD, Rudzicz F, Simpson W, Yancheva M. Evaluation of speech-based digital biomarkers: review and recommendations. Digit Biomark 2020;4(3):99-108 [FREE Full text] [doi: 10.1159/000510820] [Medline: 33251474]
15. Voleti R, Liss JM, Berisha V. A review of automated speech and language features for assessment of cognitive and thought disorders. IEEE J Sel Top Signal Process 2020 Feb;14(2):282-298 [FREE Full text] [doi: 10.1109/jstsp.2019.2952087] [Medline: 33907590]
16. Assadi G. The mental state examination. Br J Nurs 2020 Dec 10;29(22):1328-1332. [doi: 10.12968/bjon.2020.29.22.1328] [Medline: 33325295]
17. Woodford HJ, George J. Cognitive assessment in the elderly: a review of clinical methods. QJM 2007 Aug 02;100(8):469-484. [doi: 10.1093/qjmed/hcm051] [Medline: 17566006]
18. Duffy JR. Motor speech disorders and the diagnosis of neurologic disease: still a well-kept secret? Leader 2008 Nov;13(16):10-13. [doi: 10.1044/leader.ftr1.13162008.10]
19. Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, et al. Evidence-based clinical voice assessment: a systematic review. Am J Speech Lang Pathol 2013 May;22(2):212-226. [doi: 10.1044/1058-0360(2012/12-0014)] [Medline: 23184134]
20. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. J Speech Hear Res 1969 Jun;12(2):246-269. [doi: 10.1044/jshr.1202.246] [Medline: 5808852]
21. Darley FL, Aronson AE, Brown JR. Clusters of deviant speech dimensions in the dysarthrias. J Speech Hear Res 1969 Sep;12(3):462-496 [FREE Full text] [doi: 10.1044/jshr.1203.462] [Medline: 5811846]
22. Baghai-Ravary L, Beet SW. Automatic Speech Signal Analysis for Clinical Diagnosis and Assessment of Speech Disorders. New York, NY: Springer; 2013.
23. Rameau A, Cox SR, Sussman SH, Odigie E. Addressing disparities in speech-language pathology and laryngology services with telehealth. J Commun Disord 2023 Sep;105:106349 [FREE Full text] [doi: 10.1016/j.jcomdis.2023.106349] [Medline: 37321106]
24. Stipancic KL, Golzy M, Zhao Y, Pinkerton L, Rohl A, Kuruvilla-Dugdale M. Improving perceptual speech ratings: the effects of auditory training on judgments of dysarthric speech. J Speech Lang Hear Res 2023 Nov 09;66(11):4236-4258 [FREE Full text] [doi: 10.1044/2023_JSLHR-23-00322] [Medline: 37774738]
25. Allison KM, Russell M, Hustad KC. Reliability of perceptual judgments of phonetic accuracy and hypernasality among speech-language pathologists for children with dysarthria. Am J Speech Lang Pathol 2021 Jun 18;30(3S):1558-1571 [FREE Full text] [doi: 10.1044/2020_AJSLP-20-00144] [Medline: 33647216]
26. Jing L, Grigos MI. Speech-language pathologists' ratings of speech accuracy in children with speech sound disorders. Am J Speech Lang Pathol 2022 Jan 18;31(1):419-430 [FREE Full text] [doi: 10.1044/2021_AJSLP-20-00381] [Medline: 34788553]
27. Bunton K, Kent RD, Duffy JR, Rosenbek JC, Kent JF. Listener agreement for auditory-perceptual ratings of dysarthria. J Speech Lang Hear Res 2007 Dec;50(6):1481-1495. [doi: 10.1044/1092-4388(2007/102)] [Medline: 18055769]
28. Ramanarayanan V, Lammert AC, Rowe HP, Quatieri TF, Green JR. Speech as a biomarker: opportunities, interpretability, and challenges. Perspect ASHA SIGs 2022 Feb 11;7(1):276-283 [FREE Full text] [doi: 10.1044/2021_PERSP-21-00174] [Medline: 33249997]
29. Latif S, Qadir J, Qayyum A, Usama M, Younis S. Speech technology for healthcare: opportunities, challenges, and state of the art. IEEE Rev Biomed Eng 2021;14:342-356. [doi: 10.1109/RBME.2020.3006860] [Medline: 32746367]
30. Moro-Velazquez L, Dehak N. A review of the use of prosodic aspects of speech for the automatic detection and assessment of Parkinson's disease. In: Proceedings of the 1st Workshop on Automatic Assessment of Parkinsonian Speech. 2019 Presented at: AAPS '19; September 20-21, 2019; Cambridge, MA p. 42-59 URL: https://link.springer.com/chapter/10.1007/978-3-030-65654-6_3 [doi: 10.1007/978-3-030-65654-6_3]

XSL•FO
RenderX

31. Moro-Velazquez L, Gomez-Garcia JA, Arias-Londoño JD, Dehak N, Godino-Llorente JI. Advances in Parkinson's disease detection and assessment using voice and speech: a review of the articulatory and phonatory aspects. Biomed Signal Process Control 2021 Apr;66:102418. [doi: 10.1016/j.bspc.2021.102418]
32. Gullapalli AS, Mittal VK. Early detection of Parkinson's disease through speech features and machine learning: a review. In: Senjyu T, Mahalle PN, Perumal T, Joshi A, editors. ICT with Intelligent Applications. Cham, Switzerland: Springer; 2021:203-212.
33. Pulido ML, Hernández JB, Ballester M, González CM, Mekyska J, Smékal Z. Alzheimer's disease and automatic speech analysis: a review. Expert Syst Appl 2020 Jul;150:113213. [doi: 10.1016/j.eswa.2020.113213]
34. Martínez-Nicolás I, Llorente TE, Martínez-Sánchez F, Meilán JJ. Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: a systematic review article. Front Psychol 2021;12:620251 [FREE Full text] [doi: 10.3389/fpsyg.2021.620251] [Medline: 33833713]
35. de la Fuente Garcia S, Ritchie CW, Luz S. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review. J Alzheimers Dis 2020;78(4):1547-1574 [FREE Full text] [doi: 10.3233/JAD-200888] [Medline: 33185605]
36. Petti U, Baker S, Korhonen A. A systematic literature review of automatic Alzheimer's disease detection from speech and language. J Am Med Inform Assoc 2020 Nov 01;27(11):1784-1797 [FREE Full text] [doi: 10.1093/jamia/ocaa174] [Medline: 32929494]
37. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis". Autism Res 2017 Mar 08;10(3):384-407. [doi: 10.1002/aur.1678] [Medline: 27501063]
38. Rowe HP, Shellikeri S, Yunusova Y, Chenausky KV, Green JR. Quantifying articulatory impairments in neurodegenerative motor diseases: a scoping review and meta-analysis of interpretable acoustic features. Int J Speech Lang Pathol 2023 Aug;25(4):486-499. [doi: 10.1080/17549507.2022.2089234] [Medline: 36001500]
39. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
40. NVivo version 1.6.1. QSR International. URL: https://lumivero.com/product/nvivo/ [accessed 2024-04-29]
41. Flick U. The Sage Handbook of Qualitative Data Analysis. Thousand Oaks, CA: Sage Publications; 2014.
42. Wang Q, Fu Y, Shao B, Chang L, Ren K, Chen Z, et al. Early detection of Parkinson's disease from multiple signal speech: based on Mandarin language dataset. Front Aging Neurosci 2022 Nov 10;14:1036588 [FREE Full text] [doi: 10.3389/fnagi.2022.1036588] [Medline: 36438003]
43. Suppa A, Costantini G, Asci F, Di Leo P, Al-Wardat MS, Di Lazzaro G, et al. Voice in Parkinson's disease: a machine learning study. Front Neurol 2022 Feb 15;13:831428 [FREE Full text] [doi: 10.3389/fneur.2022.831428] [Medline: 35242101]
44. Song J, Lee JH, Choi J, Suh MK, Chung MJ, Kim YH, et al. Detection and differentiation of ataxic and hypokinetic dysarthria in cerebellar ataxia and Parkinsonian disorders via wave splitting and integrating neural networks. PLoS One 2022 Jun 3;17(6):e0268337 [FREE Full text] [doi: 10.1371/journal.pone.0268337] [Medline: 35658000]
45. Motin MA, Pah ND, Raghav S, Kumar DK. Parkinson's disease detection using smartphone recorded phonemes in real world conditions. IEEE Access 2022;10:97600-97609. [doi: 10.1109/access.2022.3203973]
46. Maskeliūnas R, Damaševičius R, Kulikajevas A, Padervinskis E, Pribuišis K, Uloza V. A hybrid u-lossian deep learning network for screening and evaluating Parkinson's disease. Appl Sci 2022 Nov 15;12(22):11601. [doi: 10.3390/app122211601]
47. Lim WS, Chiu SI, Wu MC, Tsai SF, Wang PH, Lin KP, et al. An integrated biometric voice and facial features for early detection of Parkinson's disease. NPJ Parkinsons Dis 2022 Oct 29;8(1):145 [FREE Full text] [doi: 10.1038/s41531-022-00414-8] [Medline: 36309501]
48. Laganas C, Iakovakis D, Hadjidimitriou S, Charisis V, Dias SB, Bostantzopoulou S, et al. Parkinson's disease detection based on running speech data from phone calls. IEEE Trans Biomed Eng 2022 May;69(5):1573-1584. [doi: 10.1109/TBME.2021.3116935] [Medline: 34596531]
49. Fayad R, Hajj-Hassan M, Costantini G, Zarazadeh Z, Errico V, Pisani A. Vocal test analysis for assessing Parkinson's disease at early stage. In: Proceedings of the 6th International Conference on Advances in Biomedical Engineering. 2021 Presented at: ICABME '14; October 7-9, 2021; Werdanyeh, Lebanon p. 171-174 URL: https://ieeexplore.ieee.org/document/9604891 [doi: 10.1109/icabme53305.2021.9604891]
50. Rahman W, Lee S, Islam MS, Antony VN, Ratnu H, Ali MR, et al. Detecting Parkinson disease using a web-based speech task: observational study. J Med Internet Res 2021 Oct 19;23(10):e26305 [FREE Full text] [doi: 10.2196/26305] [Medline: 34665148]
51. Cordella F, Paffi A, Pallotti A. Classification-based screening of Parkinson's disease patients through voice signal. In: Proceedings of the 2021 IEEE International Symposium on Medical Measurements and Applications. 2021 Presented at: MeMeA '21; June 23-25, 2021; Lausanne, Switzerland p. 1-6 URL: https://ieeexplore.ieee.org/document/9478683 [doi: 10.1109/memea52024.2021.9478683]
52. Vasquez-Correa JC, Arias-Vergara T, Klumpp P, Perez-Toro PA, Orozco-Arroyave JR, Nöth E. End-2-end modeling of speech and gait from patients with Parkinson's disease: comparison between high quality vs. smartphone data. In: Proceedings

XSL•FO
RenderX

of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. 2021 Presented at: ICASSP '21; June 6-11, 2021; Toronto, ON p. 7298-7302 URL: https://ieeexplore.ieee.org/document/9414729 [doi: 10.1109/icassp39728.2021.9414729]
53. Majda-Zdancewicz E, Potulska-Chromik A, Jakubowski J, Nojszewska M, Kostera-Pruszczyk A. Deep learning vs feature engineering in the assessment of voice signals for diagnosis in Parkinson's disease. Bull Pol Acad Sci Tech Sci 2021;69(3):e137347. [doi: 10.24425/bpasts.2021.137347]
54. Quan C, Ren K, Luo Z. A deep learning based method for Parkinson's disease detection using dynamic features of speech. IEEE Access 2021;9:10239-10252. [doi: 10.1109/access.2021.3051432]
55. Amato F, Borzi L, Olmo G, Artusi CA, Imbalzano G, Lopiano L. Speech impairment in Parkinson's disease: acoustic analysis of unvoiced consonants in Italian native speakers. IEEE Access 2021;9:166370-166381. [doi: 10.1109/access.2021.3135626]
56. Tandjung MD, Wu JC, Wang JC, Li YH. An implementation of FastAI tabular learner model for Parkinson's disease identification. In: Proceedings of the 9th International Conference on Orange Technology. 2021 Presented at: ICOT '21; December 16-17, 2021; Tainan, Taiwan p. 16-17 URL: https://ieeexplore.ieee.org/document/9680650 [doi: 10.1109/icot54518.2021.9680650]
57. Jeancolas L, Petrovska-Delacrétaz D, Mangone G, Benkelfat BE, Corvol JC, Vidailhet M, et al. X-vectors: new quantitative biomarkers for early Parkinson's disease detection from speech. Front Neuroinform 2021 Feb 19;15:578369 [FREE Full text] [doi: 10.3389/fninf.2021.578369] [Medline: 33679361]
58. Goyal J, Khandnor P, Aseri TC. A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals. Expert Syst Appl 2021 Nov;182:115283. [doi: 10.1016/j.eswa.2021.115283]
59. Carrón J, Campos-Roca Y, Madruga M, Pérez CJ. A mobile-assisted voice condition analysis system for Parkinson's disease: assessment of usability conditions. Biomed Eng Online 2021 Nov 21;20(1):114 [FREE Full text] [doi: 10.1186/s12938-021-00951-y] [Medline: 34802448]
60. Ali L, He Z, Cao W, Rauf HT, Imrana Y, Bin Heyat MB. MMDD-ensemble: a multimodal data-driven ensemble approach for Parkinson's disease detection. Front Neurosci 2021 Nov 1;15:754058 [FREE Full text] [doi: 10.3389/fnins.2021.754058] [Medline: 34790091]
61. Zhang L, Qu Y, Jin B, Jing L, Gao Z, Liang Z. An intelligent mobile-enabled system for diagnosing Parkinson disease: development and validation of a speech impairment detection system. JMIR Med Inform 2020 Sep 16;8(9):e18689 [FREE Full text] [doi: 10.2196/18689] [Medline: 32936086]
62. Altay EV, Alatas B. Association analysis of Parkinson disease with vocal change characteristics using multi-objective metaheuristic optimization. Med Hypotheses 2020 Aug;141:109722. [doi: 10.1016/j.mehy.2020.109722] [Medline: 32305812]
63. Vasquez-Correa JC, Arias-Vergara T, Orozco-Arroyave JR, Eskofier B, Klucken J, Noth E. Multimodal assessment of Parkinson's disease: a deep learning approach. IEEE J Biomed Health Inform 2019 Jul;23(4):1618-1630. [doi: 10.1109/JBHI.2018.2866873] [Medline: 30137018]
64. Camnos-Roca Y, Calle-Alonso F, Perez CJ, Naranjo L. Computational diagnosis of Parkinson's disease from speech based on regularization methods. In: Proceedings of the 26th European Signal Processing Conference. 2018 Presented at: EUSIPCO '18; September 3-7, 2018; Rome, Italy p. 1127-1131 URL: https://ieeexplore.ieee.org/document/8553505 [doi: 10.23919/eusipco.2018.8553505]
65. Montaña D, Campos-Roca Y, Pérez CJ. A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease. Comput Methods Programs Biomed 2018 Feb;154:89-97. [doi: 10.1016/j.cmpb.2017.11.010] [Medline: 29249350]
66. Vaiciukynas E, Verikas A, Gelzinis A, Bacauskiene M. Detecting Parkinson's disease from sustained phonation and speech signals. PLoS One 2017 Oct 5;12(10):e0185613 [FREE Full text] [doi: 10.1371/journal.pone.0185613] [Medline: 28982171]
67. Sztahó D, Tulics MG, Vicsi K, Valálik I. Automatic estimation of severity of Parkinson's disease based on speech rhythm related features. In: Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications. 2017 Presented at: CogInfoCom '17; September 11-14, 2017; Debrecen, Hungary p. 11-16 URL: https://ieeexplore.ieee.org/document/8268208 [doi: 10.1109/coginfocom.2017.8268208]
68. Orozco-Arroyave R, Arias-Londoño JD, Vargas-Bonilla JJ, Nöth E. Perceptual analysis of speech signals from people with Parkinson's disease. In: Proceedings of the 5th International Work-Conference on the Interplay Between Natural and Artificial Computation & Natural and Artificial Models in Computation and Biology. 2013 Presented at: IWINAC '13; June 10-14, 2013; Mallorca, Spain p. 201-211 URL: https://link.springer.com/chapter/10.1007/978-3-642-38637-4_21 [doi: 10.1007/978-3-642-38637-4_21]
69. Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform 2013 Jul;17(4):828-834. [doi: 10.1109/JBHI.2013.2245674] [Medline: 25055311]
70. Viswanathan R, Bingham A, Raghav S, Arjunan SP, Jelfs B, Kempster P, et al. Normalized Mutual Information of phonetic sound to distinguish the speech of Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3523-3526. [doi: 10.1109/EMBC.2019.8857112] [Medline: 31946638]

XSL•FO
RenderX

71. Zhang H, Yan N, Wang L, Ng ML. Energy distribution analysis and nonlinear dynamical analysis of phonation in patients with Parkinson's disease. In: Proceedings of the 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. 2017 Presented at: APSIPA-ASC '17; December 12-15, 2017; Kuala Lumpur, Malaysia p. 630-635 URL: https://ieeexplore.ieee.org/document/8282102 [doi: 10.1109/apsipa.2017.8282102]
72. Vizza P, Tradigo G, Mirarchi D, Bossio RB, Lombardo N, Arabia G, et al. Methodologies of speech analysis for neurodegenerative diseases evaluation. Int J Med Inform 2019 Feb;122:45-54. [doi: 10.1016/j.ijmedinf.2018.11.008] [Medline: 30623783]
73. Viswanathan R, Arjunan SP. Estimation of severity in Parkinson's disease using acoustic features of phonatory tasks. IETE J Res 2021 Nov 22;69(9):6292-6303. [doi: 10.1080/03772063.2021.1997361]
74. Das B, Daoudi K, Klempir J, Rusz J. Towards disease-specific speech markers for differential diagnosis in Parkinsonism. In: Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. 2019 Presented at: ICASSP '19; May 12-17, 2019; Brighton, UK p. 378 URL: https://ieeexplore.ieee.org/document/8683887 [doi: 10.1109/icassp.2019.8683887]
75. Li G, Daoudi K, Klempir J, Rusz J. Linear classification in speech-based objective differential diagnosis of parkinsonism. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. 2018 Presented at: ICASSP '18; April 15-20, 2018; Calgary, AB p. 15-20 URL: https://ieeexplore.ieee.org/document/8462681 [doi: 10.1109/icassp.2018.8462681]
76. Hemmerling D, Wojcik-Pedziwiatr M. Prediction and estimation of Parkinson's disease severity based on voice signal. J Voice 2022 May;36(3):439.e9-439.20. [doi: 10.1016/j.jvoice.2020.06.004] [Medline: 32807590]
77. Tunc HC, Sakar CO, Apaydin H, Serbes G, Gunduz A, Tutuncu M, et al. Estimation of Parkinson's disease severity using speech features and extreme gradient boosting. Med Biol Eng Comput 2020 Nov 10;58(11):2757-2773. [doi: 10.1007/s11517-020-02250-5] [Medline: 32910301]
78. Vandana VP, Darshini JK, Vikram VH, Nitish K, Kumar PP, Ravi Y. Speech characteristics of patients with Parkinson's disease-does dopaminergic medications have a role? J Neurosci Rural Pract 2021 Oct;12(4):673-679 [FREE Full text] [doi: 10.1055/s-0041-1735249] [Medline: 34737501]
79. Jain A, Abedinpour K, Polat O, Çalışkan MM, Asaei A, Pfister FM, et al. Voice analysis to differentiate the dopaminergic response in people with Parkinson's disease. Front Hum Neurosci 2021 May 31;15:667997 [FREE Full text] [doi: 10.3389/fnhum.2021.667997] [Medline: 34135742]
80. Gaballah A, Parsa V, Andreetta M, Adams S. Assessment of amplified parkinsonian speech quality using deep learning. In: Proceedings of the 2018 IEEE Canadian Conference on Electrical & Computer Engineering. 2018 Presented at: CCECE '18; May 13-16, 2018; Quebec, QC p. 1-4 URL: https://ieeexplore.ieee.org/document/8447721 [doi: 10.1109/ccece.2018.8447721]
81. Gaballah A, Parsa V, Andreetta M, Adams S. Objective and subjective speech quality assessment of amplification devices for patients with Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 2019 Jun;27(6):1226-1235. [doi: 10.1109/tnsre.2019.2915172]
82. Shimoda A, Li Y, Hayashi H, Kondo N. Dementia risks identified by vocal features via telephone conversations: a novel machine learning prediction model. PLoS One 2021 Jul 14;16(7):e0253988 [FREE Full text] [doi: 10.1371/journal.pone.0253988] [Medline: 34260593]
83. Toth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z, et al. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. Curr Alzheimer Res 2018;15(2):130-138 [FREE Full text] [doi: 10.2174/1567205014666171121114930] [Medline: 29165085]
84. Themistocleous C, Eckerström M, Kokkinakis D. Identification of mild cognitive impairment from speech in Swedish using deep sequential neural networks. Front Neurol 2018 Nov 15;9:975 [FREE Full text] [doi: 10.3389/fneur.2018.00975] [Medline: 30498472]
85. Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, et al. Automatic detection of cognitive impairments through acoustic analysis of speech. Curr Alzheimer Res 2020 Mar 20;17(1):60-68 [FREE Full text] [doi: 10.2174/1567205017666200213094513] [Medline: 32053074]
86. Munthuli A, Vongsurakrai S, Anansiripinyo T, Ellermann V, Sroykhumpa K, Onsuwan C, et al. Thammasat-NECTEC-Chula's Thai language and cognition assessment (TLCA): the Thai Alzheimer's and mild cognitive impairment screening test. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:690-694. [doi: 10.1109/EMBC46164.2021.9630779] [Medline: 34891386]
87. Mirzaei S, El Yacoubi M, Garcia-Salicetti S, Boudy J, Kahindo C, Cristancho-Lacroix V, et al. Two-stage feature selection of voice parameters for early Alzheimer's disease prediction. IRBM 2018 Dec;39(6):430-435. [doi: 10.1016/j.irbm.2018.10.016]
88. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 2015 Mar;1(1):112-124 [FREE Full text] [doi: 10.1016/j.dadm.2014.11.012] [Medline: 27239498]
89. Bertini F, Allevi D, Lutero G, Montesi D, Calzà L. Automatic speech classifier for mild cognitive impairment and early dementia. ACM Trans Comput Healthcare 2021 Oct 15;3(1):1-11. [doi: 10.1145/3469089]

XSL•FO
RenderX

90. Yamada Y, Shinkawa K, Nemoto M, Ota M, Nemoto K, Arai T. Speech and language characteristics differentiate Alzheimer's disease and dementia with Lewy bodies. Alzheimers Dement (Amst) 2022;14(1):e12364 [FREE Full text] [doi: 10.1002/dad2.12364] [Medline: 36320609]
91. Sumali B, Mitsukura Y, Liang KC, Yoshimura M, Kitazawa M, Takamiya A, et al. Speech quality feature analysis for classification of depression and dementia patients. Sensors (Basel) 2020 Jun 26;20(12):3599 [FREE Full text] [doi: 10.3390/s20123599] [Medline: 32604728]
92. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. PLoS One 2019;14(5):e0217388 [FREE Full text] [doi: 10.1371/journal.pone.0217388] [Medline: 31125389]
93. König A, Mallick E, Tröger J, Linz N, Zeghari R, Manera V, et al. Measuring neuropsychiatric symptoms in patients with early cognitive decline using speech analysis. Eur Psychiatry 2021 Oct 13;64(1):e64 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.2236] [Medline: 34641989]
94. Fazeli M, Moradi N, Soltani M, Naderifar E, Majdinasab N, Latifi SM, et al. Dysphonia characteristics and vowel impairment in relation to neurological status in patients with multiple sclerosis. J Voice 2020 May;34(3):364-370. [doi: 10.1016/j.jvoice.2018.09.018] [Medline: 30342799]
95. Fazeli M, Moradi N, Soltani M, Naderifar E, Majdinasab N, Latifi SM. Comparison of dysphonia severity index and its parameters among individuals with multiple sclerosis and healthy subjects. Shiraz E Med J 2018 Jun 12;19(7):e64857 [FREE Full text] [doi: 10.5812/semj.64857]
96. Svoboda E, Bořil T, Rusz J, Tykalová T, Horáková D, Guttmann C, et al. Assessing clinical utility of machine learning and artificial intelligence approaches to analyze speech recordings in multiple sclerosis: a pilot study. Comput Biol Med 2022 Sep;148:105853. [doi: 10.1016/j.compbiomed.2022.105853] [Medline: 35870318]
97. Gosztolya G, Tóth L, Svindt V, Bóna J, Hoffmann I. Using acoustic deep neural network embeddings to detect multiple sclerosis from speech. In: Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. 2022 Presented at: ICASSP '22; May 23-27, 2022; Singapore, Singapore p. 6927-6931 URL: https://ieeexplore.ieee.org/document/9746856 [doi: 10.1109/icassp43922.2022.9746856]
98. Vizza P, Mirarchi D, Tradigo G, Redavide M, Bossio RB, Veltri P. Vocal signal analysis in patients affected by multiple sclerosis. Procedia Comput Sci 2017;108:1205-1214. [doi: 10.1016/j.procs.2017.05.092]
99. Tena A, Claria F, Solsona F, Meister E, Povedano M. Detection of bulbar involvement in patients with amyotrophic lateral sclerosis by machine learning voice analysis: diagnostic decision support development study. JMIR Med Inform 2021 Mar 10;9(3):e21331 [FREE Full text] [doi: 10.2196/21331] [Medline: 33688838]
100. Illa A, Patel D, Yamini B, ss M, Shivashankar N, Veeramani P. Comparison of speech tasks for automatic classification of patients with amyotrophic lateral sclerosis and healthy subjects. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. 2018 Presented at: ICASSP '18; April 15-20, 2018; Calgary, AB p. 6014-6018 URL: https://ieeexplore.ieee.org/document/8461836 [doi: 10.1109/icassp.2018.8461836]
101. Likhachov D, Vashkevich M, Azarov E, Malhina K, Rushkevich Y. A mobile application for detection of amyotrophic lateral sclerosis via voice analysis. In: Proceedings of the 23rd International Conference on Speech and Computer. 2021 Presented at: SPECOM '21; September 27-30, 2021; St. Petersburg, Russia p. 372-383 URL: https://link.springer.com/chapter/10.1007/978-3-030-87802-3_34 [doi: 10.1007/978-3-030-87802-3_34]
102. Mallela J, Illa AS, N SB, Udupa S, Belur Y, Atchayaram N, et al. Voice based classification of patients with amyotrophic lateral sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning. In: Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. 2020 Presented at: ICASSP '20; May 4-8, 2020; Barcelona, Spain p. 6784-6788 URL: https://ieeexplore.ieee.org/document/9053682 [doi: 10.1109/icassp40776.2020.9053682]
103. Daudet L, Yadav N, Perez M, Poellabauer C, Schneider S, Huebner A. Portable mTBI assessment using temporal and frequency analysis of speech. IEEE J Biomed Health Inform 2017 Mar;21(2):496-506. [doi: 10.1109/JBHI.2016.2633509] [Medline: 27913365]
104. Falcone M, Yadav N, Poellabauer C, Flynn P. Using isolated vowel sounds for classification of mild traumatic brain injury. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013 Presented at: ICASSP '13; May 16-31, 2013; Vancouver, BC p. 26-31 URL: https://ieeexplore.ieee.org/document/6639136 [doi: 10.1109/icassp.2013.6639136]
105. Wall C, Powell D, Young F, Zynda AJ, Stuart S, Covassin T, et al. A deep learning-based approach to diagnose mild traumatic brain injury using audio classification. PLoS One 2022;17(9):e0274395 [FREE Full text] [doi: 10.1371/journal.pone.0274395] [Medline: 36170287]
106. Riad R, Lunven M, Titeux H, Cao XN, Hamet Bagnou J, Lemoine L, et al. Predicting clinical scores in Huntington's disease: a lightweight speech test. J Neurol 2022 Sep 14;269(9):5008-5021 [FREE Full text] [doi: 10.1007/s00415-022-11148-1] [Medline: 35567614]
107. Gallezot C, Riad R, Titeux H, Lemoine L, Montillot J, Sliwinski A, et al. Emotion expression through spoken language in Huntington disease. Cortex 2022 Oct;155:150-161 [FREE Full text] [doi: 10.1016/j.cortex.2022.05.024] [Medline: 35986957]

XSL•FO
RenderX

108. MacFarlane H, Salem AC, Chen L, Asgari M, Fombonne E. Combining voice and language features improves automated autism detection. Autism Res 2022 Jul 23;15(7):1288-1300 [FREE Full text] [doi: 10.1002/aur.2733] [Medline: 35460329]
109. Eni M, Dinstein I, Ilan M, Menashe I, Meiri G, Zigel Y. Estimating autism severity in young children from speech signals using a deep neural network. IEEE Access 2020;8:139489-139500. [doi: 10.1109/access.2020.3012532]
110. Suppa A, Asci F, Saggio G, Di Leo P, Zarezadeh Z, Ferrazzano G, et al. Voice analysis with machine learning: one step closer to an objective diagnosis of essential tremor. Mov Disord 2021 Jun 02;36(6):1401-1410. [doi: 10.1002/mds.28508] [Medline: 33528037]
111. König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, et al. Detecting apathy in older adults with cognitive disorders using automatic speech analysis. J Alzheimers Dis 2019;69(4):1183-1193. [doi: 10.3233/JAD-181033] [Medline: 31127764]
112. Aggarwal G, Sharma NV, Kavita, Sinha A. Fisher discriminant ratio based classification of intellectual disability using acoustic features. In: Proceedings of the 2nd International Conference on International Conference. 2020 Presented at: CNC '20; December 29-31, 2020; Gwalior, India p. 301-311 URL: https://link.springer.com/chapter/10.1007/978-981-16-8896-6_24 [doi: 10.1007/978-981-16-8896-6_24]
113. Lauraitis A, Maskeliunas R, Damasevicius R, Krilavicius T. Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features. IEEE Access 2020;8:96162-96172. [doi: 10.1109/access.2020.2995737]
114. Berisha V, Liss JM. Responsible development of clinical speech AI: bridging the gap between clinical research and technology. NPJ Digit Med 2024 Aug 09;7(1):208 [FREE Full text] [doi: 10.1038/s41746-024-01199-1] [Medline: 39122889]
115. Cera ML, Ortiz KZ, Bertolucci PH, Tsujimoto T, Minett T. Speech and phonological impairment across Alzheimer's disease severity. J Commun Disord 2023;105:106364. [doi: 10.1016/j.jcomdis.2023.106364] [Medline: 37453160]
116. Anderson ND. State of the science on mild cognitive impairment (MCI). CNS Spectr 2019 Feb;24(1):78-87. [doi: 10.1017/S1092852918001347] [Medline: 30651152]
117. Davis M, O Connell T, Johnson S, Cline S, Merikle E, Martenyi F, et al. Estimating Alzheimer's disease progression rates from normal cognition through mild cognitive impairment and stages of dementia. Curr Alzheimer Res 2018;15(8):777-788 [FREE Full text] [doi: 10.2174/1567205015666180119092427] [Medline: 29357799]
118. Plotas P, Nanousi V, Kantanis A, Tsiamaki E, Papadopoulos A, Tsapara A, et al. Speech deficits in multiple sclerosis: a narrative review of the existing literature. Eur J Med Res 2023 Jul 24;28(1):252 [FREE Full text] [doi: 10.1186/s40001-023-01230-3] [Medline: 37488623]
119. Masrori P, Van Damme PV. Amyotrophic lateral sclerosis: a clinical review. Eur J Neurol 2020 Oct 07;27(10):1918-1929 [FREE Full text] [doi: 10.1111/ene.14393] [Medline: 32526057]
120. Mayer AR, Quinn DK, Master CL. The spectrum of mild traumatic brain injury: a review. Neurology 2017 Aug 08;89(6):623-632 [FREE Full text] [doi: 10.1212/WNL.0000000000004214] [Medline: 28701496]
121. Vogindroukas I, Stankova M, Chelas EN, Proedrou A. Language and speech characteristics in autism. Neuropsychiatr Dis Treat 2022;18:2367-2377 [FREE Full text] [doi: 10.2147/NDT.S331987] [Medline: 36268264]
122. Hopfner F, Deuschl G. Managing essential tremor. Neurotherapeutics 2020 Oct;17(4):1603-1621 [FREE Full text] [doi: 10.1007/s13311-020-00899-2] [Medline: 32915385]
123. Marrus N, Hall L. Intellectual disability and language disorder. Child Adolesc Psychiatr Clin N Am 2017 Jul;26(3):539-554. [doi: 10.1016/j.chc.2017.03.001] [Medline: 28577608]
124. Schultz BG, Vogel AP. A tutorial review on clinical acoustic markers in speech science. J Speech Lang Hear Res 2022 Sep 12;65(9):3239-3263. [doi: 10.1044/2022_jslhr-21-00647]
125. Karlsson F, Hartelius L. On the primary influences of age on articulation and phonation in maximum performance tasks. Languages 2013;21:174 [FREE Full text] [doi: 10.3390/languages6040174] [Medline: 23844154]
126. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. J Voice 2010 Sep;24(5):540-555. [doi: 10.1016/j.jvoice.2008.12.014] [Medline: 19883993]
127. Rusz J, Tykalova T, Ramig LO, Tripoliti E. Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. Mov Disord 2021 Apr;36(4):803-814. [doi: 10.1002/mds.28465] [Medline: 33373483]
128. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, et al. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. Am J Speech Lang Pathol 2018 Aug 06;27(3):887-905. [doi: 10.1044/2018_AJSLP-17-0009] [Medline: 29955816]
129. van Gelderen L, Tejedor-García C. Innovative speech-based deep learning approaches for Parkinson's disease classification: a systematic review. Applied Sciences 2024 Sep 04;14(17):7873. [doi: 10.3390/app14177873]
130. Malekroodi HS, Madusanka N, Lee BI, Yi M. Leveraging deep learning for fine-grained categorization of Parkinson's disease progression levels through analysis of vocal acoustic patterns. Bioengineering (Basel) 2024 Mar 21;11(3):295 [FREE Full text] [doi: 10.3390/bioengineering11030295] [Medline: 38534569]
131. Soylu E, Gül S, Aslan K, Türko'lu M, Terzi M. Vision transformer based classification of neurological disorders from human speech. J Exper Comp Eng 2023 Jul;3(2):160-174. [doi: 10.62520/fujece.1454309]

XSL•FO
RenderX

132. Eguchi K, Yaguchi H, Kudo I, Kimura I, Nabekura T, Kumagai R, et al. Differentiation of speech in Parkinson's disease and spinocerebellar degeneration using deep neural networks. J Neurol 2024 Feb 21;271(2):1004-1012. [doi: 10.1007/s00415-023-12091-5] [Medline: 37989963]
133. Cardinale AM. The opportunity for telehealth to support neurological healthcare. Telemed J E Health 2018 Mar 20;24(12):969-978. [doi: 10.1089/tmj.2017.0290] [Medline: 29652625]
134. Offermann P, Levina O, Schönherr M, Bub U. Outline of a design science research process. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. 2009 Presented at: DESRIST '09; May 7-8, 2009; Philadelphia, PA p. 1-11 URL: https://dl.acm.org/doi/10.1145/1555619.1555629 [doi: 10.1145/1555619.1555629]

Abbreviations

AD: Alzheimer disease
ALS: amyotrophic lateral sclerosis
ASD: autism spectrum disorder
CNN: convolutional neural network
CNSD: central nervous system disorder
DL: deep learning
DNN: deep neural network
ET: essential tremor
HD: Huntington disease
ID: intellectual disability
MCI: mild cognitive impairment
ML: machine learning
MS: multiple sclerosis
mTBI: mild traumatic brain injury
NECTEC: National Electronics and Computer Technology Center
PD: Parkinson disease
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UPDRS: Unified Parkinson’s Disease Rating Scale

Edited by T Loetscher; submitted 09.06.24; peer-reviewed by V Martel-Sauvageau, S Tayebi Arasteh, F Asci; comments to author 20.08.24; revised version received 30.10.24; accepted 16.11.24; published 13.01.25.

Please cite as:
De Silva U, Madanian S, Olsen S, Templeton JM, Poellabauer C, Schneider SL, Narayanan A, Rubaiat R
Clinical Decision Support Using Speech Signal Analysis: Systematic Scoping Review of Neurological Disorders
J Med Internet Res 2025;27:e63004
URL: https://www.jmir.org/2025/1/e63004
doi:10.2196/63004
PMID:39804693

©Upeka De Silva, Samaneh Madanian, Sharon Olsen, John Michael Templeton, Christian Poellabauer, Sandra L Schneider, Ajit Narayanan, Rahmina Rubaiat. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Enhancing HIV Cognitive Abilities and Self-Management Through Information Technology–Assisted Interventions: Scoping Review

Hao Huang[1,2*], MM; MeiLian Xie[3*], PhD; Zhen Yang[2], MSN; AiPing Wang[2], MSN

[1]School of Nursing, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]First Hospital of China Medical University, Shenyang, China

[3]Capital Medical University, Beijing Ditan Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
AiPing Wang, MSN
First Hospital of China Medical University
The First Affiliated Hospital of China Medical University
No.155, Nanjing North Street, Heping District
Shenyang, 110001
China
Phone: 86 1399889 0800
Email: apwang@cmu.edu.cn

Abstract

Background: HIV/AIDS remains a significant global challenge, and with the rapid advancement of technology, there has been an increasing number of interventions aimed at improving HIV/AIDS cognition and self-management behaviors among patients. However, there is still a lack of detailed literature integrating relevant evidence.

Objective: This study aims to comprehensively review existing research on interventions using modern information methods to improve HIV/AIDS cognition and enhance self-management behaviors among patients. It systematically reports the theoretical frameworks and specific intervention strategies used in current research, providing a comprehensive overview of the development status of relevant studies. We aim to compile existing evidence through this scoping review to identify potential avenues for future research.

Methods: We followed the scoping review framework proposed by the Joanna Briggs Institute for the synthesis and reporting of evidence. Relevant literature was searched using electronic databases, including PubMed, Web of Science, Embase, CINAHL, and Cochrane Library. The time frame for inclusion was from 2018 to December 1, 2023. Inclusion criteria were (1) interventions using modern information technology or new digital media, (2) studies focusing on improving HIV awareness or self-management behaviors among people living with HIV, (3) intervention studies or evaluations of intervention effects, and (4) studies published within the last five years. Two reviewers (HH and MX) independently assessed each study at both the title and abstract screening stage and the full-text review stage, resolving any disagreements through discussion.

Results: A total of 55 studies that met the inclusion criteria were included. The Information-Motivation-Behavioral Skills model, Social Cognitive Theory, Health Belief Model, Theory of Planned Behavior, and Information Systems Research Framework are among the most commonly used theoretical frameworks. Modern information technology interventions are mainly constructed using smartphone apps, SMS text messaging, internet-based platforms, audiovisual materials, and digital health education platforms, with smartphone apps and SMS text messaging being the most widely used intervention media.

Conclusions: Modern information technology is becoming an important tool for health interventions among people living with HIV/AIDS. However, future research should focus on integrating theoretical framework guidance with intervention design, further exploring the diversity of intervention implementations, the applicability of different technological methods, their long-term effects, and how to more effectively combine traditional intervention strategies to maximize intervention outcomes.

(J Med Internet Res 2025;27:e57363) doi:10.2196/57363

KEYWORDS

digital media; self-management; HIV; AIDS; scoping review; technology-assisted interventions; information technology; behaviors; patient; electronic database; information systems; smartphone; app; SMS text messaging; effectiveness; mobile phone

XSL•FO
RenderX

Introduction

Currently, HIV/AIDS remains a severe and intricate global challenge. According to the 2023 report from the United Nations Program on HIV/AIDS [1], there are currently 39 million people living with HIV worldwide, with 29.8 million undergoing antiretroviral therapy (ART). In 2022, there were 1.3 million new HIV infections, and 630,000 individuals succumbed to AIDS-related illnesses. Thanks to collaborative efforts globally, the spread of HIV has slowed, and the annual death toll continues to decline [1,2]. Beyond the typical clinical symptoms of AIDS, patients often grapple with multidimensional challenges in their daily lives, including poor treatment adherence, exacerbated family conflicts, economic burdens due to high treatment costs, and mental health issues [3,4]. Research indicates that reinforcing self-management behaviors and enhancing disease awareness can lead to better treatment adherence, alleviate the burden of disease symptoms, and enable individuals to coexist more effectively with HIV [5,6]. Therefore, strengthening the self-management behaviors and disease awareness of people living with HIV becomes a crucial aspect of improving the quality of life for patients and reducing negative experiences associated with HIV.

Modern information technology rapidly and cost-effectively reduces the barriers to accessing disease-related information. It has been proven that leveraging modern technological means aids in expanding the dissemination of HIV prevention and treatment interventions [7]. Modern information technology, or new digital media, refers to user-controlled, shareable, and digital media [8]. Widely used new digital media, such as SMS text messages, websites, apps, and social media platforms, have significantly improved information sharing and communication. These technologies enable rapid, inexpensive, highly replicable, and widely covered interventions. Digital media achieve greater anonymity, repeatability, time flexibility, and customer sensitivity, meeting diverse learning needs [9]. They break down restrictions related to time and geography, maximizing the impact of information dissemination.

This study will comprehensively review evidence from the last 5 years regarding the use of modern information technology in enhancing HIV disease awareness and self-management. It underscores the enormous potential of contemporary technology in elevating self-management and disease awareness within the HIV community, guiding future research and practice.

Methods

Study Design

This scoping review uses the latest framework proposed by the Joanna Briggs Institute, encompassing the following stages: (1) formulation of research questions; (2) identification of relevant studies; (3) selection of studies; (4) charting the data; and (5) collating, summarizing, and reporting the results [10]. Reporting for this paper follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist, the preferred reporting item for scoping reviews [11]. All members of our research team have undergone systematic training in evidence-based research methodologies and are well-versed in conducting scoping reviews following established guidelines. Additionally, our team includes an expert in HIV research (MX), who has been available for consultation and reference throughout the study to ensure the quality and accuracy of our work. This scoping review has been registered on the Open Science Framework [12].

Formulating Research Questions

This study aims to summarize existing evidence regarding enhancing HIV disease awareness and self-management behaviors among people living with HIV through modern information technology.

Identifying Relevant Studies

We included studies that met the following criteria in this scoping review: (1) intervention involving modern information technology or new digital media, (2) research focusing on improving HIV disease awareness or self-management behaviors of people living with HIV, (3) intervention studies or studies evaluating intervention effects, and (4) studies published within the last five years. The exclusion criteria were (1) studies not reporting research outcomes, (2) gray literature, and (3) studies inaccessible for full-text retrieval.

We limited our literature search to studies published in 5 years to capture the most recent advancements in modern information technology interventions for improving disease awareness and self-management among people living with HIV/AIDS. This 5-year timeframe was chosen because the field is rapidly evolving, and recent studies are more likely to reflect current technologies and practices relevant to our research objectives [13].

Study Selection

English-language literature was retrieved from electronic databases, including PubMed, Web of Science, Cochrane Library, Embase, and CINAHL. The literature search in this study focused on HIV, telemedicine, cognitive training, and self-management. For specific search strings, please refer to Multimedia Appendix 1. EndNote (Clarivate Analytics) software was used to manage the studies and eliminate duplicate publications. After removing duplicates, the literature underwent a title and abstract screening, followed by a full-text review to determine inclusion or exclusion. Both stages were performed independently by two reviewers (HH and MX), and no disagreements were reported.

To ensure consistency among reviewers, we randomly selected 25 titles and abstracts for an initial screening exercise. All reviewers independently screened these based on predefined inclusion criteria. Formal literature screening commenced once a 75% interrater agreement was achieved [14]. Subsequently, HH and MX independently screened the remaining titles and abstracts using the criteria and conducted full-text reviews to determine inclusion or exclusion. In cases of disagreement between the two reviewers, a third reviewer was consulted to discuss and reach a consensus [14]. Excluded studies are documented and reported, and the results and search process

XSL•FO
RenderX

are presented in a flowchart following the PRISMA-ScR guidelines (Multimedia Appendix 2) [11].

Data Extraction and Presentation

We conducted a comprehensive review of the full texts of studies meeting the inclusion criteria, focusing on how they enhanced disease awareness and self-management behaviors of patients with HIV through the application of theoretical frameworks and modern information technology interventions. Specifically, we extracted and reported on the theoretical frameworks and modern information technology used in each study.

Two researchers (HH and MX) independently extracted data and performed content analysis of the included studies. Any disagreements were resolved through discussion, with a third researcher (ZY) serving as an arbiter when necessary. Data were extracted and integrated using mind maps and Excel (Microsoft Corp) spreadsheets. The findings were then presented using a combination of text and tables.

We categorized and extracted the following information from each study: (1) basic study information, which includes authors, publication year, study design, sample size, and study location; (2) intervention methods, which includes details of the modern information technology interventions used in the studies; and (3) conceptual frameworks or models, which includes theoretical frameworks or models applied in the studies.

Collating, Summarizing, and Reporting Results

We reported the results of the included studies by focusing on the following points: (1) the theoretical frameworks or models used in the included literature, and (2) intervention methods based on modern information technology adapted in the included literature.

Results

Study Selection

According to our inclusion criteria, the results of English database searches are as follows: 61 studies were retrieved from PubMed, 44 studies from Web of Science, 361 studies from Cochrane Library, 254 studies from Embase, and 6 studies from CINAHL. After removing duplicates, a total of 648 studies remained. After evaluating titles and abstracts, 101 English-language studies were retained, and after a thorough examination of the complete texts, 55 studies were finally included. The search results are presented as a PRISMA-ScR flowchart (Figure 1).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram.

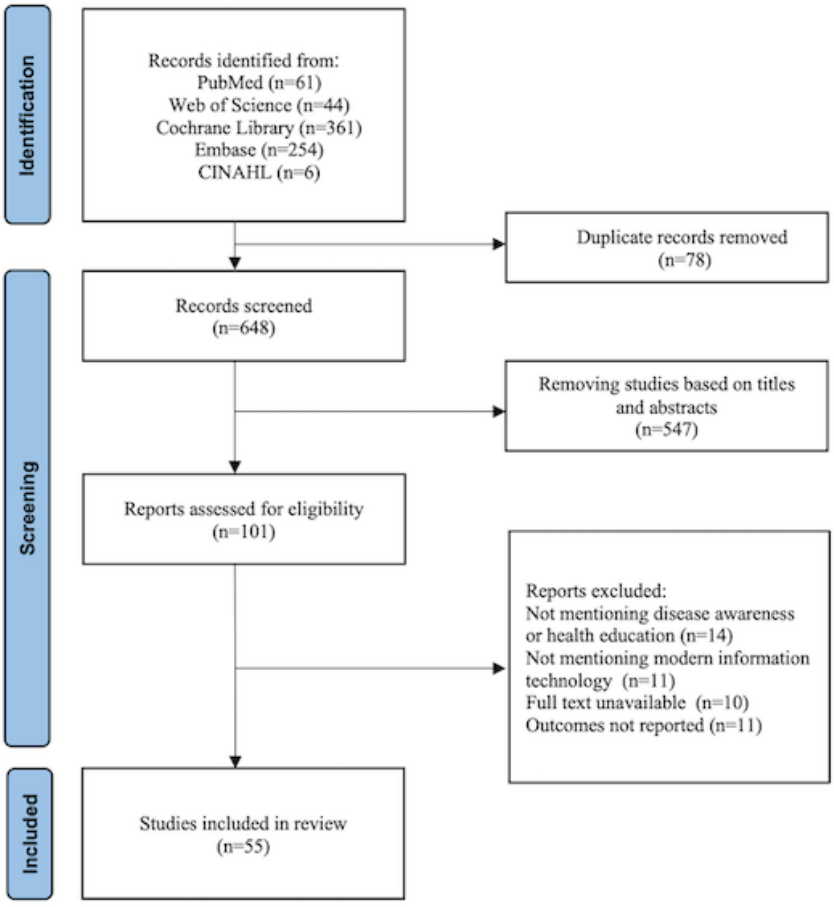

Characteristics of the Studies

Among the 55 studies finally included, the studies ranged from a minimum sample size of 21 participants to a maximum of 2125 participants. Of these, 33 studies were randomized controlled trials, 10 studies were feasibility studies, 5 studies were observational studies, 4 studies were evaluative qualitative studies, and 3 studies were mixed methods studies. Most of the

XSL•FO
RenderX

studies were conducted in the United States (n=28), followed by Kenya (n=4). Other countries include China (n=3), South Africa (n=3), Nigeria (n=3), Uganda (n=2), Thailand (n=2), and others. The general characteristics of the study are presented in Multimedia Appendix 3 [15-69].

Related Theoretical Frameworks

Information-Motivation-Behavioral Skills Theory

Information-Motivation-Behavioral Skills (IMB) theory is the most commonly used theoretical framework for designing interventions, with 7 studies basing their interventions on this theory. The IMB theory posits that information, motivation, and behavioral skills interact, with individuals acquiring information to understand the necessity of healthy behavior [70]. This, in turn, motivates them to desire and guide the adoption of such behavior. Ultimately, behavioral skills are used to transform motivation into concrete actions.

Five studies [15-19] used mobile phones as a platform, operationalizing the IMB model to enhance HIV self-management knowledge and beliefs. They developed apps or tailored SMS text messages aimed at improving HIV self-management behaviors and increasing HIV disease awareness among patients with HIV. Additionally, two studies [20,21] facilitated changes in HIV awareness and self-management behaviors by stimulating motivation, providing relevant information, and enhancing behavioral skills among intervention participants. These changes were implemented through websites and videoconferences.

Social Cognitive Theory

Five studies used the Social Cognitive Theory to develop their interventions. Social Cognitive Theory [71] emphasizes the importance of social learning, suggesting that individuals can acquire new knowledge and skills by observing others and adjusting and internalizing this learning through cognitive regulation and self-efficacy. In line with this theory, existing studies propose methods to enhance self-efficacy, including peer modeling and social persuasion, and build interventions based on these principles. Weitzman et al [22] and Côté et al [23] developed web-based interventions to improve self-management skills. Kalichman et al [24] and Tran et al [25] focused on individual differences among intervention recipients, providing personalized interventions based on the Social Cognitive Theory to enhance medication adherence and self-management skills. Kinuthia et al [26], using the Social Cognitive Theory, crafted intervention messages, integrated treatment adherence with other priorities, and delivered them as SMS text messages to intervention recipients.

Health Belief Model

Three studies developed their interventions based on the Health Belief Model. Despite sharing the same name, the Health Belief Model discussed in this review encompasses studies that different scholars have proposed. First, Tran et al [25] applied the Health Belief Model proposed by Becker [72] in his study. Throughout the research period and subsequent meetings, Tran et al [25] measured perceived severity and self-efficacy for comparison and sent personalized messages as cues for action. The theory-driven design helped in understanding the connections between patient behaviors and identifying clustered behaviors such as alcohol consumption and unemployment. Second, Kinuthia et al [26] applied the Health Belief Model modified by Janz and Becker [73]. Kinuthia used this model to compose messages providing care support to pregnant and postpartum women infected with HIV. Additionally, the Health Belief Model applied by Ditre et al [27] was proposed by Champion and Skinner [74]. Based on this model, the scholars provided computer-generated personalized feedback to modify smoking and opioid misuse behaviors in people living with HIV.

Theory of Planned Behavior

Two studies used the Theory of Planned Behavior to construct their interventions. The Theory of Planned Behavior [75] posits that one can effectively predict whether an individual will adopt a specific behavior by measuring and analyzing behavioral intentions, attitudes, subjective norms, perceived behavioral control, and actual behavior. Ditre et al [27] integrated the Theory of Planned Behavior along with other theories to construct a comprehensive intervention, predicting and improving smoking and opioid misuse behaviors in patients with HIV. Jiao et al [28] used this theory to design an intervention aimed at improving ART adherence among men who have sex with men (MSM).

Information Systems Research Framework

Two studies used the Information Systems Research Framework (ISR) as their theoretical foundation. The ISR [76] advocates for iterative cycles in the design and evaluation process to provide a systematic scientific design approach. Beauchemin et al [29] developed and evaluated the effectiveness of app intervention for improving self-management behaviors in people living with HIV based on the ISR. Similarly, Schnall et al [30] used ISR to create a mobile app to enhance ART adherence among people living with HIV.

Other Theories

In addition to the 5 theories mentioned above, a few have been used by researchers in intervention studies. Ezegbe et al [31], based on Rational Emotive Behavior Therapy Theory [77], used audiovisual interventions to enhance HIV disease awareness among high school students. Nestadt et al [32], based on Social Action Theory [78], developed a cartoon-based program for comprehensive psychosocial responses and treatment adherence intervention for adolescents who are HIV positive. Schnall et al [33], using Social Learning Theory [79], developed a mobile app to examine its effectiveness in reducing sexual risk behaviors among young men attracted to the same sex. Cordova et al [34] conducted a narrative intervention based on the Ecodevelopmental and Empowerment Framework [80,81], aiming to change risky drug use and behaviors among adolescents. Middleton et al [35] and Carroll et al [36], guided by the Capability, Opportunity, Motivation, and Behavior Model [82], developed SMS text messages to increase intervention recipients' knowledge of HIV prevention. Côté et al [23] implemented a web-based intervention guided by the McGill Nursing Model [83]. Sun et al [37], grounded in Cognitive

XSL•FO
RenderX

Behavioral Theory [84], developed an app and tested its effectiveness in promoting HIV prevention among transgender women. Weitzman et al [22] designed web-based interventions based on the Modified AIDS Risk Reduction Model [85], examining their effectiveness for the target population. Chenneville et al [38], using the Cognitive Theory of Multimedia Learning [86] and Dual Coding Theory [87], designed modules to enhance health literacy in young MSM regarding HIV retention care. Cordoba et al [39], guided by the Information System Success [88] framework, evaluated participants' experiences and satisfaction with a self-made app. In the psychological education segment of a comprehensive intervention, Ditre et al [27] primarily used the Transtheoretical Model [89] as guidance. Aladin et al [40], based on the Integrated, Design, Assess, and Share [90] theoretical framework, combined behavioral theories with design thinking and developed a mobile app as an intervention. Tran et al [25], based on the Integrated Theory of Behavior Change [91], developed a smartphone app to promote treatment adherence and self-efficacy among people living with HIV. To better illustrate the findings of this study, we have summarized the frequency of theoretical frameworks applied in the included studies. The results are visually represented in Table 1, providing a clear overview of the theoretical models and their respective utilization.

Table 1. Frequency of theoretical frameworks used in the included studies.

Theoretical framework	Frequency, n
Information-Motivation-Behavioral Skills Theory	7
Social Cognitive Theory	5
Health Belief Model	3
Theory of Planned Behavior	2
Information Systems Research Framework	2
Capability, Opportunity, Motivation, and Behavior Model	2
Rational Emotive Behavior Therapy Theory	1
Social Action Theory	1
Social Learning Theory	1
Ecodevelopmental and Empowerment Framework	1
McGill Nursing Model	1
Cognitive Behavioral Theory	1
The Modified AIDS Risk Reduction Model	1
Cognitive Theory of Multimedia Learning	1
Dual Coding Theory	1
Information System Success	1
Transtheoretical Model	1
Integrated, Design, Assess, and Share	1
Integrated Theory of Behavior Change	1

Intervention Methods Based on Modern Information Technology

Smartphone Apps

Smartphones have become widely popular globally with their high portability and rich features. In this review, it was found that interventions based on smartphone apps are the most favored among researchers. Among the included studies, 25 used smartphone apps as an intervention method.

This category's most popular intervention method involves providing self-management plans or tools to the intervention participants through smartphone apps, facilitating improvements in self-management behaviors. This is commonly observed in efforts to enhance compliance with medical advice, improve awareness of HIV, and ameliorate HIV-related symptoms [15,29,36,41-44,67,68]. Two studies integrated phone and SMS text message modules to build apps to enhance treatment adherence [45,46]. Communication-oriented mobile apps, such as WhatsApp (Meta), WeChat (Tencent), and Line (LY Corporation), were used to create group discussions for implementing interventions [17,47,48]. Finally, one study used the entertainment features of smartphones to develop an iPhone platform game to enhance the participant's compliance [49].

Additionally, some studies have developed apps for targeted populations to implement interventions. For HIV testing behavior, sexual education, and treatment adherence among gay men, four studies have collectively developed three apps for interventions [33,39,50,51]. Two studies have developed apps targeting vulnerable adolescent groups, aiming to improve self-management behaviors and prevent disengagement from care [34,40]. In addressing health, one study developed an app

XSL•FO
RenderX

to promote HIV prevention to meet the needs of transgender women [37].

SMS Text Message

Similar to interventions based on smartphone apps, using the SMS text message functionality of mobile phones for direct interventions has gained favor among many researchers. Sending SMS text messages to improve treatment adherence is the simplest and most direct method. Three studies have systematically sent timed or instant messages to intervention participants, covering medication reminders, peer education, and web-based group discussions, to enhance ART adherence among people living with HIV [16,28,52].

The included literature also takes into account specific populations. For instance, communication strategies in SMS text message content are tailored based on local cultural backgrounds for implementing prevention programs targeting 18- to 22-year-olds [53]. A combination of regular educational sessions and weekly health education SMS text messages are organized to enhance HIV prevention behaviors and reduce risky behaviors in African American youth [54]. Regularly sending SMS text messages for infant health education and promoting treatment adherence is implemented to provide care support for pregnant women and postpartum women living with HIV [26]. Planned SMS text messaging campaigns are carried out to raise HIV awareness among construction workers and encourage HIV testing [35].

Web-Based and Digital Health Education Platforms

In internet-based interventions, most researchers choose to implement intervention programs by constructing websites. Building websites allows for the rapid dissemination of health education. Three studies used websites to create knowledge repositories for providing safety education. These websites comprehensively covered various topics, offering comprehensive and personalized health education to intervention participants through web-based interactions, sexual health information delivery, and modularized intervention methods. This approach aimed to promote HIV prevention and increase awareness of sexual health [55-57]. Using website platforms, two studies relied on teams with medical backgrounds to provide diverse learning methods. They combined social support and emphasized personalized web-based services to facilitate the effective transmission of health information [22,58]. Interventions providing counseling services through remote videoconferences typically use internet platforms to offer web-based video meetings to intervention participants. This involves hosting regular videoconferences and encouraging participant involvement through email, phone calls, and other means to enhance medical adherence and disease awareness among people living with HIV [21,23,59].

Furthermore, there are different forms of internet-based intervention programs. For example, one study created private discussion groups on social media platforms to enhance HIV-related knowledge and treatment literacy, thereby promoting the continuity of HIV care [60]. Other studies constructed internet-based electronic health education programs to strengthen health literacy among people living with HIV in HIV care through digital content [38], developed intervention programs using publicly available internet resources [20], and constructed HIV prevention interventions through video blogs that feature storytelling and digital gaming [61].

Finally, a minority of studies have implemented interventions by constructing digital health education platforms. These studies use computer-based personalized approaches, providing tailored health information and training videos through digital platforms to enhance treatment adherence and HIV-related knowledge and improve health care behaviors among intervention participants [27,62,63].

Performing and Multimedia Interventions

The commonality of audio and video interventions is the use of virtualized digital content to create various intervention plans tailored to different populations and needs. These interventions are delivered in audio and video, emphasizing the dissemination of disease-related knowledge and fostering connections between individuals and communities. Ultimately, they aim to enhance the understanding of intervention participants, boost self-efficacy among people living with HIV, and improve preventive behaviors.

For the adolescent population, considering their capacity for acceptance and understanding, researchers have developed a series of cartoon tutorials to strengthen parent-child communication, problem-solving, and negotiation skills. Participants pair with caregivers to attend the courses, enhancing internal family connections through learning and discussions [32]. For children, researchers have created captivating characters and carefully curated story content. With the assistance of qualified therapists, they deliver digital interventions to children using audio and video formats, achieving intervention goals while providing entertainment [64].

For specific populations, such as the gay and bisexual community, researchers have developed popular web series or short videos with a theme focused on gay, bisexual, and other MSM individuals. These videos capture the curiosity of the target audience while promoting HIV and other sexually transmitted infection testing behaviors [65,66,69]. Finally, some studies combine various communication channels, such as digital games, street interviews, and therapist-guided group psychological discussions, to enhance the effectiveness of performing multimedia interventions, providing participants with knowledge about preventing HIV/AIDS and sexually transmitted infections [18,19,31]. We have visualized the interventions included in the studies through Figure 2 to present the findings of this section more intuitively.

XSL•FO
RenderX

Figure 2. Visualization of interventions included in the studies by type. PFI: personalized feedback intervention.

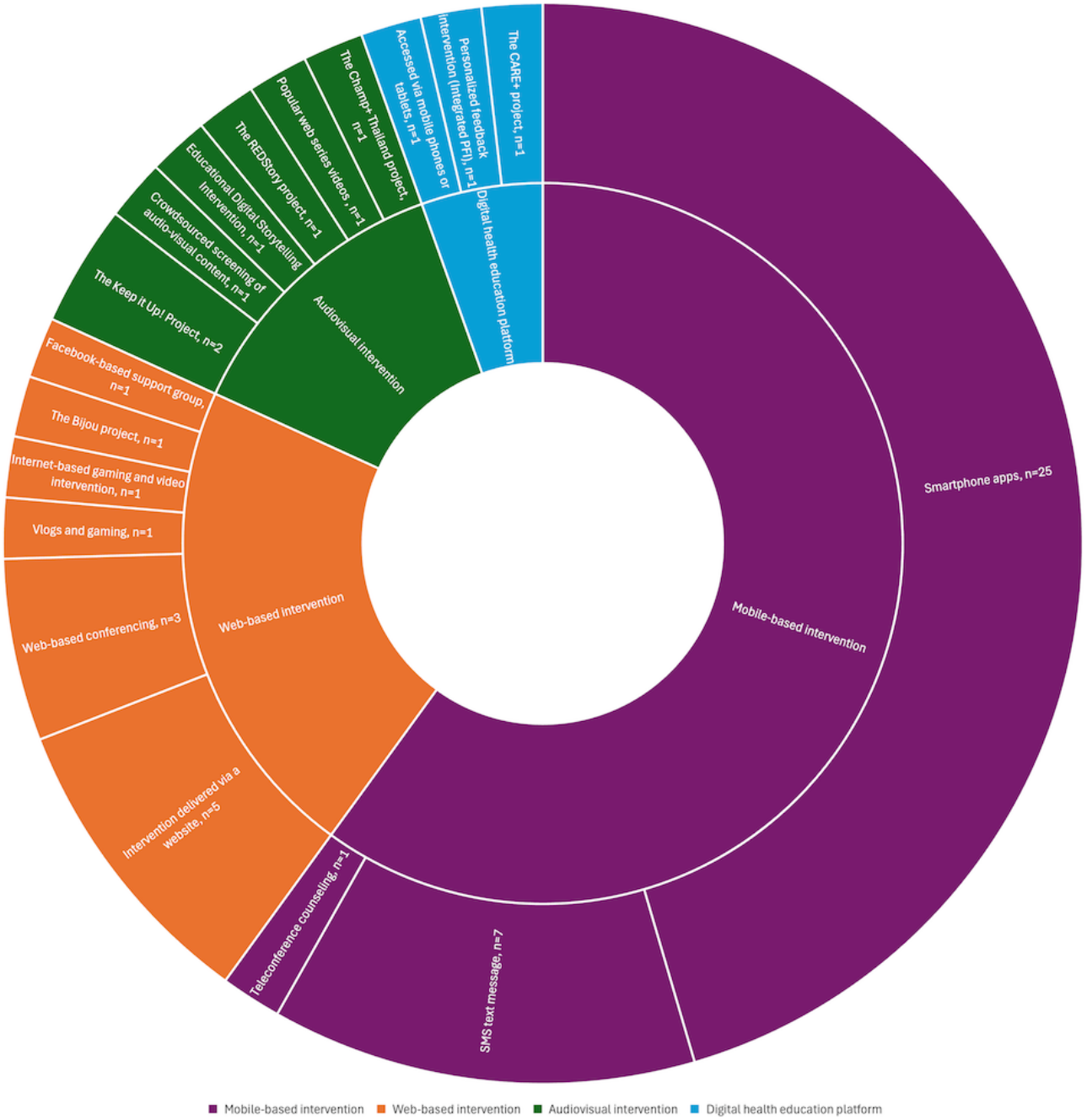

Discussion

Principal Findings

This scoping review reveals that most of the included studies have used one or more theoretical frameworks in the development of their interventions. Currently, the application of modern information technology primarily involves intervention construction, development of smartphone apps, feasibility studies, and analysis to intervene with the relevant population, yielding generally positive feedback. Thanks to the rapid advancement of modern technology, there has been an increasing emergence of modern information technologies, which better improve or innovate medical intervention measures and hold significant potential for the future development of medical technology. Modern information technology (or new media technology) plays a crucial role in HIV/AIDS research. This study explores the potential of information technology in enhancing HIV/AIDS awareness from multiple dimensions and systematically summarizes the theoretical frameworks and specific implementation methods of relevant studies. Through this research, we not only summarize the application cognition of modern information technology in health intervention but also provide new ideas and methods to improve HIV/AIDS awareness among older people, which is meaningful for enhancing HIV/AIDS awareness and inspiring future related research.

As early as 2000, scholars reported that theory-driven intervention measures have been proven to be more effective than non–theory-driven ones [92]. However, among the 55 studies included in this study, only 45% (n=25) of the intervention schemes were based on theory-driven approaches,

XSL•FO
RenderX

with the remaining studies either not reporting the research framework used or not using any theoretical framework. Among these 25 studies, the most commonly used theory was the IMB theory [70]. In health education interventions based on the IMB theory, accurate information about HIV/AIDS is first provided, emphasizing personalized interventions to increase individual awareness levels. Subsequently, by highlighting the serious threat of HIV/AIDS to individual health, individuals' positive motivations for prevention and treatment are stimulated, encouraging proactive behavior change. Finally, by providing practical behavioral skills training, individuals are helped to overcome barriers to implementing healthy behaviors, such as correctly using condoms and undergoing regular HIV testing, to achieve the intervention's objectives. The widespread application of the IMB theory in HIV/AIDS health education reflects its effectiveness in integrating information, motivation, and behavioral skills, providing theoretical support for designing targeted and effective health interventions. Additionally, some studies have integrated multiple theoretical frameworks as comprehensively as possible to construct intervention schemes [22,23,26,38,43]. Therefore, we encourage future research to use theory-driven approaches as much as possible to enhance the quality of interventions and improve intervention outcomes.

We found that the majority of studies used mobile phones as a platform for cognitive interventions, demonstrating diversity and innovation in HIV/AIDS cognitive interventions. From the results of this study, it is evident that the intelligent apps developed by various studies transcend the traditional role of communication tools in cognitive interventions. Leveraging the high portability and ubiquity of smartphones, these interventions offer convenient and personalized ways to reach the target population. By harnessing the mature functionalities of smartphones, such as integrating phone calls, SMS text messaging modules, and chat apps, these interventions provide communication pathways and methods rooted in traditional communication for intervention purposes [17,45-48]. The customization of apps tailored to specific populations, such as MSM [33,39,50,51], vulnerable adolescents [40], and transgender women [37], highlights the flexibility and adaptability of smartphone apps in meeting the diverse health needs of different groups.

Future Directions

Despite the promising future and wide-ranging applicability of smartphone app interventions, as evident from the studies included in this scoping review, only one study [45] mentions the relevance to the population of older people. Older people may face unique challenges and needs related to HIV/AIDS, such as unfamiliarity or discomfort with digital technology and preferences for accessing health information through specific channels. Therefore, developing specialized smartphone app interventions tailored to the population of older people is essential. These interventions should address the unique needs and preferences of older adults by incorporating user-friendly features that align with their specific contexts and lifestyles. For example, simplified navigation systems; larger font sizes; and real-time, one-on-one, web-based support are highly recommended. Additionally, for older people who may face challenges with typing or using smartphones, voice-guided navigation could be an especially effective solution. Such tailored interventions and innovations are crucial for enhancing accessibility and usability, ultimately improving the older people population's engagement with health information and interventions.

Through this review, it is evident that a significant portion of studies prefer using SMS text messaging for interventions, with only one study [16] emphasizing the importance of protecting user privacy. SMS text messaging interventions involve the collection, storage, and use of users' personal information. Therefore, safeguarding user privacy is crucial, especially when sensitive health information is involved. The lack of adequate protection for user privacy may lead to distrust and resistance among users, impacting their acceptance and engagement with the intervention.

Strengths and Limitations

To the best of our knowledge, this scoping review is the first to comprehensively summarize the theoretical frameworks used in interventions using modern information technologies to enhance HIV/AIDS-related cognition and self-management behaviors. Adhering to the Joanna Briggs Institute framework provides a structured and standardized approach to the scoping review process, enhancing the reliability and reproducibility of the findings. Additionally, it provides an updated integration of recent intervention studies leveraging contemporary digital tools, reflecting the rapid advancements in information technology. By systematically consolidating existing intervention strategies, this review offers diverse possibilities for future research and helps prevent the redundancy of developing similar intervention schemes.

This scoping review has several limitations. First, we excluded gray literature, such as conference abstracts and dissertations, from our inclusion and exclusion criteria, which may have restricted the literature retrieved during the search. Second, all the studies included in the review were in English, potentially overlooking high-quality studies written in other languages. Furthermore, some studies using modern information technologies may have used different keywords, which could have resulted in the omission of relevant research from our review. Finally, as a scoping review, this study does not assess the effectiveness of the interventions, which should be addressed in future systematic reviews or meta-analyses.

Conclusions

This study found that modern information technology has been widely used in designing intervention programs to improve cognitive function and self-management behaviors among people living with HIV/AIDS. These interventions primarily use mobile phones as carriers and the internet as the medium for dissemination. Such technological approaches enable personalized, convenient, and real-time interventions, expanding the methods and avenues of traditional interventions. However, we also observed that most of the included studies have not fully used existing theories or conceptual frameworks in constructing their interventions. The lack of support from theoretical frameworks may lead to research failing to achieve the established goals as originally designed, thereby slowing

XSL•FO
RenderX

the problem-solving process and potentially affecting the final effectiveness.

Acknowledgments

This study was funded by the Capital Health Development Research Program (First Issue 2024-2G-2174). All ideas, writing, and original content in this manuscript were conceived and developed exclusively by the authors. No generative artificial intelligence tools or models were used in the creation of any part of this work.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

HH and MX designed and produced this manuscript. ZY, MX, and HH contributed to the screening of literature and the creation of summary tables. AW contributed to the conception and design and critically revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Specific search strings.
[DOCX File , 21 KB - jmir_v27i1e57363_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 117 KB - jmir_v27i1e57363_app2.pdf]

Multimedia Appendix 3
The general information of included literature.
[XLSX File (Microsoft Excel File), 23 KB - jmir_v27i1e57363_app3.xlsx]

References

1. The path that ends AIDS: UNAIDS global AIDS update 2023. UNAIDS. 2023. URL: https://www.unaids.org/en/resources/documents/2023/global-aids-update-2023 [accessed 2024-12-27]
2. UNAIDS global AIDS update 2022 reveals millions of lives are “in danger”. U.S. Embassy in Mozambique. URL: https://mz.usembassy.gov/unaids-in-danger-report-2022-reveals-millions-of-lives-are-at-risk/ [accessed 2023-08-05]
3. Wanjala SW, Nyongesa MK, Mapenzi R, Luchters S, Abubakar A. A qualitative inquiry of experiences of HIV-related stigma and its effects among people living with HIV on treatment in rural Kilifi, Kenya. Front Public Health 2023;11:1188446 [FREE Full text] [doi: 10.3389/fpubh.2023.1188446] [Medline: 37427260]
4. Nyongesa MK, Nasambu C, Mapenzi R, Koot HM, Cuijpers P, Newton CRJC, et al. Psychosocial and mental health challenges faced by emerging adults living with HIV and support systems aiding their positive coping: a qualitative study from the Kenyan coast. BMC Public Health 2022;22(1):76 [FREE Full text] [doi: 10.1186/s12889-021-12440-x] [Medline: 35022012]
5. King E, Kinvig K, Steif J, Qiu AQ, Maan EJ, Albert AY, et al. Mobile text messaging to improve medication adherence and viral load in a vulnerable Canadian population living with human immunodeficiency virus: a repeated measures study. J Med Internet Res 2017;19(6):e190 [FREE Full text] [doi: 10.2196/jmir.6631] [Medline: 28572079]
6. Moucheraud C, Stern AF, Ismail A, Nsubuga-Nyombi T, Ngonyani MM, Mvungi J, et al. Can self-management improve HIV treatment engagement, adherence, and retention? A mixed methods evaluation in Tanzania and Uganda. AIDS Behav 2020;24(5):1486-1494 [FREE Full text] [doi: 10.1007/s10461-019-02672-9] [Medline: 31529290]
7. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
8. Guse K, Levine D, Martins S, Lira A, Gaarde J, Westmorland W, et al. Interventions using new digital media to improve adolescent sexual health: a systematic review. J Adolesc Health 2012;51(6):535-543. [doi: 10.1016/j.jadohealth.2012.03.014] [Medline: 23174462]

XSL•FO
RenderX

9. Marsch LA, Grabinski MJ, Bickel WK, Desrosiers A, Guarino H, Muehlbach B, et al. Computer-assisted HIV prevention for youth with substance use disorders. Subst Use Misuse 2011;46(1):46-56 [FREE Full text] [doi: 10.3109/10826084.2011.521088] [Medline: 21190405]
10. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
11. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
12. Huang H, Yang Z, Xie M, Wang A. Information technology-assisted intervention for HIV cognitive enhancement and self-management: a scoping review. Open Science Framework. 2023 Dec 22. URL: https://osf.io/djg7t [accessed 2025-01-08]
13. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
14. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: scoping reviews (2020 version). In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: JBI; 2020.
15. Shim MS, Kim S, Choi M, Choi JY, Park CG, Kim GS. Developing an app-based self-management program for people living with HIV: a randomized controlled pilot study during the COVID-19 pandemic. Sci Rep 2022;12(1):19401 [FREE Full text] [doi: 10.1038/s41598-022-19238-w] [Medline: 36371442]
16. Aunon FM, Wanje G, Richardson BA, Masese L, Odeny TA, Kinuthia J, et al. Randomized controlled trial of a theory-informed mHealth intervention to support ART adherence and viral suppression among women with HIV in Mombasa, Kenya: preliminary efficacy and participant-level feasibility and acceptability. BMC Public Health 2023;23(1):837 [FREE Full text] [doi: 10.1186/s12889-023-15638-3] [Medline: 37158872]
17. Fan X, Ning K, Liu C, Zhong H, Lau JTF, Hao C, et al. Uptake of an app-based case management service for HIV-positive men who have sex with men in China: process evaluation study. J Med Internet Res 2023;25:e40176 [FREE Full text] [doi: 10.2196/40176] [Medline: 37099367]
18. Madkins K, Moskowitz DA, Moran K, Dellucci TV, Mustanski B. Measuring acceptability and engagement of the keep it up! internet-based HIV prevention randomized controlled trial for young men who have sex with men. AIDS Educ Prev 2019;31(4):287-305 [FREE Full text] [doi: 10.1521/aeap.2019.31.4.287] [Medline: 31361516]
19. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and behavioral outcomes of keep it up!: an eHealth HIV prevention program RCT. Am J Prev Med 2018;55(2):151-158 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]
20. Whiteley LB, Brown LK, Curtis V, Ryoo HJ, Beausoleil N. Publicly available internet content as a HIV/STI prevention intervention for urban youth. J Prim Prev 2018;39(4):361-370. [doi: 10.1007/s10935-018-0514-y] [Medline: 30003459]
21. Navarra AMD, Rosenberg MG, Gormley M, Bakken S, Fletcher J, Whittemore R, et al. Feasibility and acceptability of the adherence connection counseling, education, and support (ACCESS) proof of concept: a peer-led, mobile health (mHealth) Cognitive Behavioral Antiretroviral Therapy (ART) Adherence Intervention for HIV-Infected (HIV+) Adolescents and Young Adults (AYA). AIDS Behav 2023;27(6):1807-1823 [FREE Full text] [doi: 10.1007/s10461-022-03913-0] [Medline: 36574184]
22. Weitzman PF, Zhou Y, Kogelman L, Mack S, Sharir JY, Vicente SR, et al. A web-based HIV/STD prevention intervention for divorced or separated older women. Gerontologist 2020;60(6):1159-1168 [FREE Full text] [doi: 10.1093/geront/gnz098] [Medline: 31403668]
23. Côté J, Rouleau G, Ramirez-Garcia MP, Auger P, Thomas R, Leblanc J. Effectiveness of a web-based intervention to support medication adherence among people living with HIV: web-based randomized controlled trial. JMIR Public Health Surveill 2020;6(2):e17733 [FREE Full text] [doi: 10.2196/17733] [Medline: 32310145]
24. Kalichman SC, Cherry C, Kalichman MO, Eaton LA, Kohler JJ, Montero C, et al. Mobile health intervention to reduce HIV transmission: a randomized trial of behaviorally enhanced HIV treatment as prevention (B-TasP). J Acquir Immune Defic Syndr 2018;78(1):34-42 [FREE Full text] [doi: 10.1097/QAI.0000000000001637] [Medline: 29406429]
25. Tran BX, Bui TM, Do AL, Boyer L, Auquier P, Nguyen LH, et al. Efficacy of a mobile phone-based intervention on health behaviors and HIV/AIDS treatment management: randomized controlled trial. J Med Internet Res 2023;25:e43432 [FREE Full text] [doi: 10.2196/43432] [Medline: 37104001]
26. Kinuthia J, Ronen K, Unger JA, Jiang W, Matemo D, Perrier T, et al. SMS messaging to improve retention and viral suppression in prevention of mother-to-child HIV transmission (PMTCT) programs in Kenya: a 3-arm randomized clinical trial. PLoS Med 2021;18(5):e1003650 [FREE Full text] [doi: 10.1371/journal.pmed.1003650] [Medline: 34029338]
27. Ditre JW, LaRowe LR, Vanable PA, de Vita MJ, Zvolensky MJ. Computer-based personalized feedback intervention for cigarette smoking and prescription analgesic misuse among persons living with HIV (PLWH). Behav Res Ther 2019;115:83-89 [FREE Full text] [doi: 10.1016/j.brat.2018.10.013] [Medline: 30389092]
28. Jiao K, Wang C, Liao M, Ma J, Kang D, Tang W, et al. A differentiated digital intervention to improve antiretroviral therapy adherence among men who have sex with men living with HIV in China: a randomized controlled trial. BMC Med 2022;20(1):341 [FREE Full text] [doi: 10.1186/s12916-022-02538-3] [Medline: 36210434]

XSL•FO
RenderX

29. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. Int J Med Inform 2019;122:37-44 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
30. Schnall R, Sanabria G, Jia H, Cho H, Bushover B, Reynolds NR, et al. Efficacy of an mHealth self-management intervention for persons living with HIV: the wiseApp randomized clinical trial. J Am Med Inform Assoc 2023;30(3):418-426 [FREE Full text] [doi: 10.1093/jamia/ocac233] [Medline: 36469808]
31. Ezegbe B, Eseadi C, Ede MO, Igbo JN, Aneke A, Mezieobi D, et al. Efficacy of rational emotive digital storytelling intervention on knowledge and risk perception of HIV/AIDS among schoolchildren in Nigeria. Medicine (Baltimore) 2018;97(47):e12910 [FREE Full text] [doi: 10.1097/MD.0000000000012910] [Medline: 30461604]
32. Nestadt DF, Saisaengjan C, McKay MM, Bunupuradah T, Pardo G, Lakhonpon S, et al. CHAMP+ Thailand: pilot randomized control trial of a family-based psychosocial intervention for perinatally HIV-infected early adolescents. AIDS Patient Care STDS 2019;33(5):227-236 [FREE Full text] [doi: 10.1089/apc.2019.0021] [Medline: 31067121]
33. Schnall R, Kuhns LM, Pearson C, Batey DS, Bruce J, Hidalgo MA, et al. Efficacy of MyPEEPS mobile, an HIV prevention intervention using mobile technology, on reducing sexual risk among same-sex attracted adolescent males: a randomized clinical trial. JAMA Netw Open 2022;5(9):e2231853. [doi: 10.1001/jamanetworkopen.2022.31853] [Medline: 36129712]
34. Cordova D, Munoz-Velazquez J, Lua FM, Fessler K, Warner S, Delva J, et al. Pilot study of a multilevel mobile health app for substance use, sexual risk behaviors, and testing for sexually transmitted infections and HIV among youth: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(3):e16251 [FREE Full text] [doi: 10.2196/16251] [Medline: 32181747]
35. Middleton M, Somerset S, Evans C, Blake H. Test@Work texts: mobile phone messaging to increase awareness of HIV and HIV testing in UK construction employees during the COVID-19 pandemic. Int J Environ Res Public Health 2020;17(21):7819 [FREE Full text] [doi: 10.3390/ijerph17217819] [Medline: 33114546]
36. Carroll JK, Tobin JN, Luque A, Farah S, Sanders M, Cassells A, et al. "Get Ready and Empowered About Treatment" (GREAT) study: a pragmatic randomized controlled trial of activation in persons living with HIV. J Gen Intern Med 2019;34(9):1782-1789 [FREE Full text] [doi: 10.1007/s11606-019-05102-7] [Medline: 31240605]
37. Sun CJ, Anderson KM, Kuhn T, Mayer L, Klein CH. A sexual health promotion app for transgender women (Trans Women Connected): development and usability study. JMIR Mhealth Uhealth 2020;8(5):e15888 [FREE Full text] [doi: 10.2196/15888] [Medline: 32396131]
38. Chenneville T, Drake H, Gabbidon K, Rodriguez C, Hightow-Weidman L. Bijou: engaging young MSM in HIV care using a mobile health strategy. J Int Assoc Provid AIDS Care 2021;20:23259582211030805 [FREE Full text] [doi: 10.1177/23259582211030805] [Medline: 34242100]
39. Cordoba E, Idnay B, Garofalo R, Kuhns LM, Pearson C, Bruce J, et al. Examining the information systems success (ISS) of a mobile sexual health app (MyPEEPS Mobile) from the perspective of very young men who have sex with men (YMSM). Int J Med Inform 2021;153:104529 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104529] [Medline: 34385097]
40. Aladin B, Thompson M, Addison D, Havens J, McGowan J, Nash D, et al. The YGetIt? Program: a mobile application, PEEP, and digital comic intervention to improve HIV care outcomes for young adults. Health Promot Pract 2023;24(4):658-668. [doi: 10.1177/15248399221150789] [Medline: 36924286]
41. Safdari R, SeyedAlinaghi S, Mohammadzadeh N, Noori T, Rahmati P, Qaderi K, et al. Developing Aysoo: a mobile-based self-management application for people living with HIV. HIV AIDS Rev 2021;21(1):24-30. [doi: 10.5114/hivar.2022.113389]
42. Clouse K, Noholoza S, Madwayi S, Mrubata M, Camlin CS, Myer L, et al. The implementation of a GPS-based location-tracking smartphone app in South Africa to improve engagement in HIV care: randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e44945 [FREE Full text] [doi: 10.2196/44945] [Medline: 37204838]
43. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: a qualitative study. Int J Med Inform 2019;131:103949 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.009] [Medline: 31561192]
44. Barroso J, Madisetti M, Mueller M. A feasibility study to develop and test a cognitive behavioral stress management mobile health application for HIV-related fatigue. J Pain Symptom Manage 2020;59(2):242-253 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.09.009] [Medline: 31539601]
45. Morano JP, Clauson K, Zhou Z, Escobar-Viera CG, Lieb S, Chen IK, et al. Attitudes, beliefs, and willingness toward the use of mHealth tools for medication adherence in the florida mHealth adherence project for people living with HIV (FL-mAPP): pilot questionnaire study. JMIR Mhealth Uhealth 2019;7(7):e12900 [FREE Full text] [doi: 10.2196/12900] [Medline: 31271150]
46. Twimukye A, Naggirinya AB, Parkes-Ratanshi R, Kasirye R, Kiragga A, Castelnuovo B, et al. Acceptability of a mobile phone support tool (Call for Life Uganda) for promoting adherence to antiretroviral therapy among young adults in a randomized controlled trial: exploratory qualitative study. JMIR Mhealth Uhealth 2021;9(6):e17418 [FREE Full text] [doi: 10.2196/17418] [Medline: 34121665]
47. Bergam S, Sibaya T, Ndlela N, Kuzwayo M, Fomo M, Goldstein MH, et al. "I am not shy anymore": a qualitative study of the role of an interactive mHealth intervention on sexual health knowledge, attitudes, and behaviors of South African adolescents with perinatal HIV. Reprod Health 2022;19(1):217 [FREE Full text] [doi: 10.1186/s12978-022-01519-2] [Medline: 36457044]

XSL•FO
RenderX

48. Anand T, Nitpolprasert C, Jantarapakde J, Meksena R, Phomthong S, Phoseeta P, et al. Implementation and impact of a technology-based HIV risk-reduction intervention among Thai men who have sex with men using "Vialogues": a randomized controlled trial. AIDS Care 2020;32(3):394-405 [FREE Full text] [doi: 10.1080/09540121.2019.1622638] [Medline: 31154811]
49. Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. AIDS Care 2018;30(sup4):21-33 [FREE Full text] [doi: 10.1080/09540121.2018.1503224] [Medline: 30626196]
50. Hightow-Weidman L, Muessig KE, Egger JR, Vecchio A, Platt A. Epic allies: a gamified mobile app to improve engagement in HIV care and antiretroviral adherence among young men who have sex with men. AIDS Behav 2021;25(8):2599-2617 [FREE Full text] [doi: 10.1007/s10461-021-03222-y] [Medline: 33740213]
51. Horvath KJ, Lammert S, Danh T, Mitchell JW. The feasibility, acceptability and preliminary impact of mobile application to increase repeat HIV testing among sexual minority men. AIDS Behav 2020;24(6):1835-1850. [doi: 10.1007/s10461-019-02755-7] [Medline: 31823111]
52. Mao L, Buchanan A, Wong HTH, Persson A. Beyond mere pill taking: SMS reminders for HIV treatment adherence delivered to mobile phones of clients in a community support network in Australia. Health Soc Care Community 2018;26(4):486-494. [doi: 10.1111/hsc.12544] [Medline: 29336111]
53. Ybarra ML, Agaba E, Nyemara N. A pilot RCT evaluating InThistoGether, an mHealth HIV prevention program for Ugandan youth. AIDS Behav 2021;25(10):3437-3448 [FREE Full text] [doi: 10.1007/s10461-021-03237-5] [Medline: 33963477]
54. Mayo-Wilson LJ, Coleman J, Timbo F, Latkin C, Brown ERT, Butler AI, et al. Acceptability of a feasibility randomized clinical trial of a microenterprise intervention to reduce sexual risk behaviors and increase employment and HIV preventive practices (EMERGE) in young adults: a mixed methods assessment. BMC Public Health 2020;20(1):1846 [FREE Full text] [doi: 10.1186/s12889-020-09904-x] [Medline: 33267860]
55. Ivanova O, Wambua S, Mwaisaka J, Bossier T, Thiongo M, Michielsen K, et al. Evaluation of the ELIMIKA pilot project: improving ART adherence among HIV positive youth using an eHealth intervention in Mombasa, Kenya. Afr J Reprod Health 2019;23(1):100-110. [doi: 10.29063/ajrh2019/v23i1.10] [Medline: 31034176]
56. Hightow-Weidman LB, LeGrand S, Muessig KE, Simmons RA, Soni K, Choi SK, et al. A randomized trial of an online risk reduction intervention for young black MSM. AIDS Behav 2019;23(5):1166-1177 [FREE Full text] [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
57. Nelson KM, Perry NS, Stout CD, Dunsiger SI, Carey MP. The young men and media study: a pilot randomized controlled trial of a community-informed, online HIV prevention intervention for 14-17-year-old sexual minority males. AIDS Behav 2022;26(2):569-583 [FREE Full text] [doi: 10.1007/s10461-021-03412-8] [Medline: 34342742]
58. Visser M, Kotze M, van Rensburg MJ. An mHealth HIV prevention programme for youth: lessons learned from the iloveLife.mobi programme in South Africa. AIDS Care 2020;32(sup2):148-154. [doi: 10.1080/09540121.2020.1742866] [Medline: 32189510]
59. Periasamy M, Mohankumar V, Shanmugam V, Selvakumar M, Pandian SM, Sridharan L. Redefining venereology practice in Tamil Nadu, South India—Nakshatra health—a networking model. Indian J Sex Transm Dis AIDS 2023;44(1):56-63 [FREE Full text] [doi: 10.4103/ijstd.ijstd_13_23] [Medline: 37457523]
60. Dulli L, Ridgeway K, Packer C, Murray KR, Mumuni T, Plourde KF, et al. A social media-based support group for youth living with HIV in Nigeria (SMART Connections): randomized controlled trial. J Med Internet Res 2020;22(6):e18343 [FREE Full text] [doi: 10.2196/18343] [Medline: 32484444]
61. Hill MJ, Coker S. Novel use of video logs to deliver educational interventions to Black women for disease prevention. West J Emerg Med 2022;23(2):211-221 [FREE Full text] [doi: 10.5811/westjem.2021.12.54012] [Medline: 35302455]
62. Kurth AE, Sidle JE, Chhun N, Lizcano JA, Macharia SM, Garcia MM, et al. Computer-based counseling program (CARE+ Kenya) to promote prevention and HIV health for people living with HIV/AIDS: a randomized controlled trial. AIDS Educ Prev 2019;31(5):395-406 [FREE Full text] [doi: 10.1521/aeap.2019.31.5.395] [Medline: 31550197]
63. Holst C, Stelzle D, Diep LM, Sukums F, Ngowi B, Noll J, et al. Improving health knowledge through provision of free digital health education to rural communities in Iringa, Tanzania: nonrandomized intervention study. J Med Internet Res 2022;24(7):e37666 [FREE Full text] [doi: 10.2196/37666] [Medline: 35900820]
64. Dike IC, Ebizie EN, Njoku OC, Oraelosi CA, Egbe CI, Nnamani AP, et al. Improving knowledge and perception of HIV/AIDS among English language speaking children in rural areas through educational digital storytelling. Medicine (Baltimore) 2021;100(50):e28058 [FREE Full text] [doi: 10.1097/MD.0000000000028058] [Medline: 34918660]
65. Tan RKJ, Koh WL, Le D, Banerjee S, Chio MT, Chan RKW, et al. Effect of a popular web drama video series on HIV and other sexually transmitted infection testing among gay, bisexual, and other men who have sex with men in Singapore: community-based, pragmatic, randomized controlled trial. J Med Internet Res 2022;24(5):e31401 [FREE Full text] [doi: 10.2196/31401] [Medline: 35522470]
66. Tang W, Mao J, Liu C, Mollan K, Zhang Y, Tang S, et al. Reimagining health communication: a noninferiority randomized controlled trial of crowdsourced intervention in China. Sex Transm Dis 2019;46(3):172-178 [FREE Full text] [doi: 10.1097/OLQ.0000000000000930] [Medline: 30741854]

XSL•FO
RenderX

67. Cho H, Porras T, Baik D, Beauchemin M, Schnall R. Understanding the predisposing, enabling, and reinforcing factors influencing the use of a mobile-based HIV management app: a real-world usability evaluation. Int J Med Inform 2018 Sep;117:88-95 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.007] [Medline: 30032969]
68. Aomori M, Matsumoto C, Takebayashi S, Matsuyama N, Uto Y, Tanaka M, et al. Effects of a smartphone app-based diet and physical activity program for men living with HIV who have dyslipidemia: a pilot randomized controlled trial. Jpn J Nurs Sci 2023 Jul;20(3):e12535. [doi: 10.1111/jjns.12535] [Medline: 37060244]
69. Downing MJ, Wiatrek SE, Zahn RJ, Mansergh G, Olansky E, Gelaude D, et al. Video selection and assessment for an app-based HIV prevention messaging intervention: formative research. Mhealth 2023;9:2 [FREE Full text] [doi: 10.21037/mhealth-21-53] [Medline: 36760783]
70. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
71. Bandura A, Freeman WH, Lightsey R. Self-Efficacy: The Exercise of Control. New York City, NY: Springer; 1999.
72. Becker MH. The health belief model and sick role behavior. Health Educ Monogr 1974;2(4):409-419. [doi: 10.1177/109019817400200407]
73. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
74. Champion VL, Skinner CS. The health belief model. Health Behav Health Educ Theory Res Pract 2008;4:45-65 [FREE Full text]
75. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
76. Hevner AR. A three cycle view of design science research. Scand J Inf Syst 2007;19(2):4 [FREE Full text]
77. Ogbuanya TC, Eseadi C, Orji CT, Anyanwu JI, Ede MO, Bakare J. Effect of rational emotive behavior therapy on negative career thoughts of students in technical colleges in Nigeria. Psychol Rep 2018;121(2):356-374. [doi: 10.1177/0033294117724449] [Medline: 28776484]
78. Ewart CK. Social action theory for a public health psychology. Am Psychol 1991;46(9):931-946. [doi: 10.1037//0003-066x.46.9.931] [Medline: 1958012]
79. Bandura A, Walters RH. Social Learning Theory. New York, NY: General Learning Press; 1977.
80. Zimmerman MA, Israel BA, Schulz A, Checkoway B. Further explorations in empowerment theory: an empirical analysis of psychological empowerment. Am J Commun Psychol 1992;20(6):707-727. [doi: 10.1007/bf00942234]
81. Szapocznik J, Coatsworth JD. An Ecodevelopmental Framework for Organizing the Influences on Drug Abuse: A Developmental Model of Risk and Protection. Washington, DC: American Psychological Association; 1999.
82. Ogden J. Celebrating variability and a call to limit systematisation: the example of the behaviour change technique taxonomy and the behaviour change wheel. Health Psychol Rev 2016;10(3):245-250. [doi: 10.1080/17437199.2016.1190291] [Medline: 27189585]
83. Gottlieb L, Rowat K. The McGill model of nursing: a practice-derived model. ANS Adv Nurs Sci 1987;9(4):51-61. [doi: 10.1097/00012272-198707000-00008] [Medline: 3111351]
84. Beck AT. Cognitive Therapy and the Emotional Disorders. Westminster, London: Penguin; 1979.
85. Miller S, Exner TM, Williams SP, Ehrhardt AA. A gender-specific intervention for at-risk women in the USA. AIDS Care 2000;12(5):603-612. [doi: 10.1080/095401200750003789] [Medline: 11218546]
86. Mayer RE. The Cambridge Handbook of Multimedia Learning. Cambridge, United Kingdom: Cambridge University Press; 2005.
87. Paivio A. Mental Representations: A Dual Coding Approach. Oxford, United Kingdom: Oxford University Press; 1990.
88. Petter S, DeLone W, McLean E. Measuring information systems success: models, dimensions, measures, and interrelationships. Eur J Inf Syst 2017;17(3):236-263. [doi: 10.1057/ejis.2008.15]
89. Jo P. The transtheoretical model and stages of change. In: Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass/Wiley; 2008.
90. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
91. Ryan P. Integrated theory of health behavior change: background and intervention development. Clin Nurse Spec 2009;23(3):161-70; quiz 171 [FREE Full text] [doi: 10.1097/NUR.0b013e3181a42373] [Medline: 19395894]
92. Fisher JD, Fisher WA. Theoretical approaches to individual-level change in HIV risk behavior. In: Handbook of HIV Prevention. Berlin, Germany: Springer; 2000:3-55.

Abbreviations

ART: antiretroviral therapy
IMB: Information-Motivation-Behavioral Skills
ISR: Information Systems Research Framework

XSL•FO
RenderX

MSM: men who have sex with men
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by T de Azevedo Cardoso; submitted 15.02.24; peer-reviewed by H Zhang, Z Zhang, D Wu, MS Shim; comments to author 19.11.24; revised version received 01.12.24; accepted 11.12.24; published 13.01.25.

Please cite as:
Huang H, Xie M, Yang Z, Wang A
Enhancing HIV Cognitive Abilities and Self-Management Through Information Technology–Assisted Interventions: Scoping Review
J Med Internet Res 2025;27:e57363
URL: https://www.jmir.org/2025/1/e57363
doi:10.2196/57363
PMID:

©Hao Huang, MeiLian Xie, Zhen Yang, AiPing Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Participant Contributions to Person-Generated Health Data Research Using Mobile Devices: Scoping Review

Shanshan Song[1,2], MA, MSc; Micaela Ashton[3], MS, RN; Rebecca Hahn Yoo[2,4], MSE; Zoljargal Lkhagvajav[2], MD, MPH; Robert Wright[5], MLS; Debra J H Mathews[6,7], MA, PhD; Casey Overby Taylor[1,2,4,8], PhD

[1]Biomedical Informatics & Data Science Section, The Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]Institute for Computational Medicine, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, United States

[3]Johns Hopkins Hospital, Baltimore, MD, United States

[4]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States

[5]Welch Medical Library, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[6]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States

[7]Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[8]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Corresponding Author:
Shanshan Song, MA, MSc
Biomedical Informatics & Data Science Section
The Johns Hopkins University School of Medicine
2024 East Monument St. S 1-200
Baltimore, MD, 21205
United States
Phone: 1 6174170316
Email: ssong41@jhmi.edu

Abstract

Background: Mobile devices offer an emerging opportunity for research participants to contribute person-generated health data (PGHD). There is little guidance, however, on how to best report findings from studies leveraging those data. Thus, there is a need to characterize current reporting practices so as to better understand the potential implications for producing reproducible results.

Objective: The primary objective of this scoping review was to characterize publications' reporting practices for research that collects PGHD using mobile devices.

Methods: We comprehensively searched PubMed and screened the results. Qualifying publications were classified according to 6 dimensions—1 covering key bibliographic details (for all articles) and 5 covering reporting criteria considered necessary for reproducible and responsible research (ie, "participant," "data," "device," "study," and "ethics," for original research). For each of the 5 reporting dimensions, we also assessed reporting completeness.

Results: Out of 3602 publications screened, 100 were included in this review. We observed a rapid increase in all publications from 2016 to 2021, with the largest contribution from US authors, with 1 exception, review articles. Few original research publications used crowdsourcing platforms (7%, 3/45). Among the original research publications that reported device ownership, most (75%, 21/28) reported using participant-owned devices for data collection (ie, a Bring-Your-Own-Device [BYOD] strategy). A significant deficiency in reporting completeness was observed for the "data" and "ethics" dimensions (5 reporting factors were missing in over half of the research publications). Reporting completeness for data ownership and participants' access to data after contribution worsened over time.

Conclusions: Our work depicts the reporting practices in publications about research involving PGHD from mobile devices. We found that very few papers reported crowdsourcing platforms for data collection. BYOD strategies are increasingly popular; this creates an opportunity for improved mechanisms to transfer data from device owners to researchers on crowdsourcing platforms. Given substantial reporting deficiencies, we recommend reaching a consensus on best practices for research collecting PGHD from mobile devices. Drawing from the 5 reporting dimensions in this scoping review, we share our recommendations and justifications for 9 items. These items require improved reporting to enhance data representativeness and quality and empower participants.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e51955) doi:10.2196/51955

KEYWORDS

scoping review; person-generated health data; PGHD; mHealth; mobile device; smartphone; mobile phone; wearable; fitness tracker; smartwatch; BYOD; crowdsourcing; reporting deficiency

Introduction

Collecting Person-Generated Health Data for Research Using Mobile Devices and Their Facilitators

The proliferation of mobile devices boosts the generation of person-generated health data (PGHD). In 2018, more than 80% of Americans owned a smartphone [1,2], with modest growth to 85% by 2021 [3]. Wearable health care devices were used by nearly 28% of the US population by 2020 [4,5]. Among smartphone users, over 50% are collecting "health-associated information" [6]. We use the phrase "person-generated health data (PGHD)" for health-related data created, recorded, or gathered by or from individuals, family members, or other caregivers to help address a health concern inside and outside clinical settings.

PGHD from mobile devices have significant research implications. Mobile devices could facilitate access to large pools of study participant data in a granular, longitudinal, and personal way, potentially accessing high-frequency data at low costs [7]. In this study, "mobile devices" refers to smartphones and wearables, which include fitness trackers and smartwatches; "research" covers biomedical and behavior studies [8].

Because of the above advantages, researchers have made numerous efforts to collect PGHD for mobile device research. Those efforts are enabled by participants willing to share their PGHD for research [9] and by using the informatics infrastructure needed for participants to share their data. Examples of informatics infrastructure to enable PGHD sharing for research dates from 2015 and 2016, with the releases of ResearchKit (Apple Inc) [10,11] and ResearchStack (Cornell Tech and Open mHealth) [12,13] and the launch of mPower [14], Sage Bionetwork's first major smartphone-based health research study. Research practices dealing with PGHD from mobile devices are, in part, guided by the European Union's General Data Protection Regulation (published in 2016; in effect since 2018) and the Food and Drug Administration's guidelines on clinical research, medical devices, and mobile health apps (released in 2013, 2015, and 2019). Between 2000 and 2020, more than 12,000 mobile device–related health research papers were published [15].

Potential Issues With Data Representativeness and Quality When Collecting Person-Generated Health Data From Mobile Devices

Despite the potential advantages of using PGHD for mobile device research, the wide variety of practices to collect PGHD may lead to issues with data quality. These issues may be due to emerging research strategies such as Bring-Your-Own-Device (BYOD; which uses participant-owned devices instead of provisioned devices) that can miss some patient demographics [16], thus leading to low generalizability. Similar issues are noticed in studies using crowdsourcing platforms to collect data [17]. Selection bias with wearables is a significant issue due to differences in access [18,19], which in turn can lead to differences in sharing data for research [20]. Both accessibility and self-selection could lead to lower study sample diversity for race and socioeconomic status. For example, National Institutes of Health's All of Us program allows participants to choose to contribute Fitbit (Google LLC) data through its BYOD subprogram. The All of Us BYOD subprogram has noticed a higher proportion of participants who are White, earn >US $25,000/year, and have college or advanced degrees in the BYOD subgroup compared with all participants [21]. When such data are used in machine learning, the low representation of some patient groups may lead to poor prediction accuracy when applied to those groups [22,23].

Motivation for This Work

To better understand methodological issues such as those related to data representativeness and quality, we aimed to review publications on research collecting PGHD from mobile devices and to characterize the reporting elements of that research thoroughly, particularly those elements relevant to BYOD and crowdsourcing. In doing so, we sought to capture aspects of this field not addressed by previously published reviews. Several reviews explore mobile health research. A review by Cao et al [15] covering the years 2000 to 2020 and a review by El-Sherif et al [24] covering 2020 to February 2021 identified 4 leading contributors (the United States, the United Kingdom, Canada, and Australia). The 2 reviews, however, focused only on bibliometrics. A scoping review by Fischer and Kleen [25] published in 2021 investigated data collection by smartphone apps in longitudinal epidemiological studies, and they found a limited number of studies that integrated apps in data collection. Studies with cross-sectional designs were excluded. A scoping review by Huhn et al [26] published in 2022 summarized wearables in health research and found that most studies were observational and that 93% of the participants were in global health studies. Smartphones were not included in their scope. Other reviews of interest include one that briefly mentioned data collection using mobile devices for health research [27] and another that discussed the application of sensors (eg, inertial measurement units) in health care [28]. Our review differs from these reviews by covering wearables and smartphones and their use to collect PGHD in research. We also attempt to be comprehensive in our characterization of reporting practices, our inclusion of a range of study types (eg, both longitudinal and cross-sectional studies), and our examination of bibliographic information. To the best of our knowledge, ours is the first review of publications on research collecting PGHD from mobile devices and the first to examine BYOD practices and the use of crowdsourcing platforms in this field.

This review characterizes the reporting elements of included publications and discusses the relevance of reporting practices

XSL•FO
RenderX

for supporting reproducible and responsible research. There are no current, ready-to-use guidelines for reporting research involving PGHD from mobile devices. For example, the Future of Privacy Forum's "Best Practices for Consumer Wearables & Wellness Apps & Devices" dates back to 2016 [29]. More recently, the Office of the National Coordinator for Health Information Technology released a white paper in 2018 to conceptualize a data infrastructure for the capture, use, and sharing of PGHD in care delivery and research through 2024 [30]. The final document, however, is still under development. Also relevant are the Structured Template and Reporting Tool for Real-World Evidence (STaRT-RWE) (2021) [31] and the Mobile Health Evidence Reporting and Assessment (mERA) checklist (2016) [32], although both are limited. STaRT-RWE does not specify requirements for mobile device studies, and mERA is limited to health interventions. There are also some focused guidelines for app development (2016) [33] and data integration (2021) [34], but none fully cover what is needed for high-quality reporting of mobile device research. In fact, reporting deficiencies are widely noted in studies involving mobile devices. A systematic review by Olaye et al [35] in 2022 found that 30% of publications failed to report quantitative adherence; such missingness can impede a complete understanding of study data. Our review characterizes the magnitude and types of reporting problems in publications on research collecting PGHD from mobile devices.

Methods

We performed a scoping review following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [36]. The scoping review first involved implementing a comprehensive search and screening strategy. Next, the selected articles were characterized according to their bibliographic details and 5 reporting dimensions.

Search Strategy

A PubMed search was used to find instances of explicitly described participant data contributions (eg, through the use of language such as "data donation," "data sharing," etc) involving wearables, smartphones, or mobile apps (Textbox 1). A MeSH (Medical Subject Headings) term analysis of a preliminary sample of articles was conducted to support PubMed query development (Yale MeSH Analyzer [37]). We selected search terms to balance recall and precision in the search results. We ran the PubMed search on August 7, 2021.

Textbox 1. PubMed search strategy.

- Search concepts
- Research questions:
  - Q1a. Data donation
    - (data[tw] OR record*[tw] OR information[tw]) AND (donat*[tw] OR donor*[tw])
  - Q1b. Data sharing
    - (data[tw] OR record*[tw] OR information[tw]) AND ("Information Dissemination"[Mesh] OR sharing*[tw] OR share*[tw])
  - Q2. Wearables, smartphones, or mobile applications
    - "Wearable Electronic Devices"[Mesh:NoExp] OR "Fitness Trackers"[Mesh] OR wearable sensor*[tw] OR wearable device*[tw] OR wearable technolog*[tw] OR self-tracking[tw] OR self-tracker*[tw] OR fitness tracker*[tw] OR smart watch*[tw] OR smartwatch*[tw] OR "Mobile Applications"[Mesh] OR "Cell Phone"[Mesh:NoExp] OR "Smartphone"[Mesh] OR "Computers, Handheld"[Mesh] OR Mobile health[tw] OR mHealth[tw] OR eHealth[tw] OR e-health[tw] OR e-healthcare[tw] OR mobile application*[tw] OR mobile technolog*[tw] OR app[tw] OR apps[tw] OR cell phone*[tw] OR cellphone*[tw] OR cellular phone*[tw] OR cellular telephone*[tw] OR mobile phone*[tw] OR mobile telephone*[tw] OR smart phone*[tw] OR smartphone*[tw] OR mobile device*[tw] OR personal digital assistant*[tw] OR "Digital Technology"[Mesh] OR digital technolog*[tw]
- Full search strategy
- (Q1a OR Q1b) AND Q2

Screening and Data Charting

Covidence [38] was used to screen publications for eligibility in 2 steps, first by eliminating publications based on a title and abstract review and then by reviewing the full text of the remaining journals. A pre-established set of eligibility criteria was used for screening (Table 1). As this an emerging field and our motivation to see the trends in full landscape, we did not set criteria for time range. Screening decisions were made only after 2 independent reviewers on the screening team (SS, MA, and RHY) agreed. All conflicts were resolved by group discussion. Library journal subscriptions and interlibrary loan were used to obtain full text, with the full text of 1 publication [39] not available from either source. All reviewers conducted pilots for each round of screening.

We charted publications for bibliographic details and along 5 reporting dimensions ("participant," "device," "data," "study," and "ethics"; double quoted here and in the text below). Because our focus is on health data contributions from participants through BYOD and crowdsourcing platforms, we modeled the reporting dimensions on existing work in the areas of citizen science data contribution [40], bioethical approaches to using personal health data [41], and mobile health apps [42]. Upon

XSL•FO
RenderX

reviewing the keywords of included publications (eg, “ethical code,” “consent,” and “informed consent”), we summarized frequent themes of keywords (eg, consent). Furthermore, we refined our charting template by selecting factors related to those frequent themes to improve data adequacy. The final factors included under the 5 reporting dimensions are listed in Table 2. A data charting pilot with a random sample of one-fifth of the final pool of publications was conducted before finalizing the data charting template [43]. A team of 4 reviewers (SS, MA, RHY, and ZL) charted the data using the template built in Covidence [38]. In addition, 2 reviewers independently charted each publication. Consensus was reached by group discussion.

Table 1. Inclusion and exclusion criteria overview.

Theme	Criteria
Phenomenon of interest	Data contribution (ie, data donation or data sharing).
Technology	Wearables, smartphones, or mobile applications: limit the scope of wearables to fitness trackers and smartwatches and include surveys through smartphone applications. Exclude surveys through websites.
Perspective	Individual: individual refers to either patient or nonpatient and exclude data transfer between researcher and researcher, researcher and organization, and organization and organization.
Goal	Collected data is used for research.
Data Type	Data concerning human health.
Study Type	—[a]
Language	English.
Time	—

[a]Not applicable.

XSL•FO
RenderX

Table 2. Definitions of 24 factors under bibliographic details and 5 reporting dimensions.

Factors	Definitions
Bibliographic details (3 factors)	
Publication year	• The year of publication, not the year of submission or acceptance.
Publication location	• The country or area of the affiliation of the first author or first authors.
Publication type	• Research: original research publication. • Protocol: protocol or proposal for original research. • Tech: publication describing a dataset, database, mobile app design, platform, or algorithm; or use case report. • Review: systematic review, scoping review, or any other type of review. • Other: including viewpoint, perspective, editorial, commentary, correspondence, letter, or needs assessment; mostly about ethics and regulations.
Reporting dimensions (21 factors)	
"**Participant"**	
Identity of data subject and data contributor	• This factor examines the question, "Are the data subject and data contributor the same individual?" • Participant in a study could be either data subject or data contributor. • Data subject refers to the person whose data is collected, held, or processed. • Data contributor refers to the individual who provides information for a research study. • Data contributor may contribute data on behalf of data subject. Data contributors may include, but are not limited to, guardians of children, health care providers of patients, and sheriffs or jail administrators of prisoners. • Identity refers to whether the data subject and data contributor are the same individual.
Living status of data subject	• Whether data subject is living or not at the moment of data contribution to research.
Purpose of data generation	• Whether the data contributed to research were generated particularly for this research or generated for other purposes and then used for research. If data were generated for participants themselves during routine use (eg, self-management), this is an example of "other purposes," and it is marked as "for self."
"**Device"**	
Device owner	• Owner of the device used to generate the data, not the owner of device used to transfer the data, although the 2 devices could be the same.
Device type	• Wearable devices include fitness trackers and smartwatches.
Data capture mode	• Passive or active.
"**Data"**	
Data type	• Location data is any type of data revealing location, which may include, but is not limited to, GPS data and self-reported location data. Genetics-biospecimen data may include, but is not limited to, genetics data extracted from biospecimens. • Genetics-non-biospecimen data may include, but is not limited to, genetics data not extracted from biospecimens; for example, data from commercial genetics tests. • Claims or administrative data may include, but is not limited to, insurance claims data and administrative data such as length of stay, date of service, etc. • Clinical data may include, but is not limited to, electronic health records (eg, laboratory results, vital signs, and treatments), health surveys, clinical trial data, and patient or disease registries. • Mental health or lifestyle data are those not included in the categories above. They may include data relevant to, but not limited to, mental health, diet, physical activity, or sleep.
Data owner	• Data ownership is charted based on explicit statements in the text of publications.
Participant access to data after contribution	• Whether participants can access contributed data after data contribution.
"**Study"**	
Research scenario	• Nonprofit, commercial, or public health.
Purpose of data collection	• For primary, secondary, or both primary and secondary data analysis.

XSL•FO
RenderX

Factors	Definitions
Used crowdsourcing platform to collect data	• In our review, we focused on crowdsourcing platforms for data collection. These platforms have 3 essential properties, that are researchers can access potential participants through the platform, researchers can distribute tasks to participants through the platform, and the number of potential participants and enrolled participants is large. Platform examples include Amazon Mechanical Turk and PatientsLikeMe (PatientsLikeMe website).
Research design	• Observational or experimental.
Design duration	• Cross-sectional or longitudinal, not restricted to observational studies.
Attrition reported	• Whether the publication reported attrition or not. In our review, we define attrition as "the loss of eligible participants from research at any time following consent to participate" (adapted from Siddiqi et al [44]).
"Ethics"	
Consent reported	• Whether the publication reported consent or not.
Informed consent reported	• Whether the publication reported informed consent or not.
Consent type	• Consent type is charted based on explicit statements in the publication's text or on the reviewer's best assumptions according to details mentioned in the publication. • Blanket consent: a process by which participants contribute their data without any restrictions [45]. • Broad consent: a process by which participants contribute their data, subject to specified restrictions, for multiple future studies in a broad range, the nature and specificities of which are unknown at the time of consenting [45,46]. • Tiered consent: a process by which participants have the option of giving broad consent for only certain types of research or research uses, including for only specific diseases or indications (eg, neurological diseases), for only publicly funded research, or for only specified institutions or researchers. With tiered consent, participants can also choose whether their data are identifiable or anonymized [47]. • Meta consent: a form of tiered informed consent in which participants can choose a different consent option within different research tiers. For example, a participant could give broad consent to all research conducted in certain institutions but only study-specific consent for genomic research or privately funded research [47]. Suppose a study reports both meta-consent and dynamic consent[a]. In that case, this study will be marked as "meta consent" because dynamic consent is considered to be a characteristic of meta consent by some, but not all, researchers. However, if a study only reports dynamic consent, this study will not be marked as "meta consent" or any other consent because dynamic consent itself is not a consent model [46]. • Explicit consent: a process in which the data subject must give an express statement of consent. An explicit consent statement should specifically refer to (1) the particular dataset that is to be processed, (2) the precise purpose of processing (including any automated decision-making), (3) any risks or implications that might arise for the data subject as a result of the data processing, and (4) any other relevant and specific information that might influence the decision of a data subject to give or not give their consent [48]. Here, explicit consent is used interchangeably with study-specific consent. With study-specific informed consent, research participants consent to their contributed data being used in a single specific research study. If researchers choose this form of consent, they will need to contact data contributors whenever they want to use a particular data point in a new study [47].
Consent subject	• Whether or not the publication reported consent subjects. Consent subjects are individuals who are consented to participate in research studies.
Right to opt-out reported	• Whether or not the publication mentioned the participants' right to opt out of research at any time.
Monetary benefits	• "Unconditional" refers to scenarios where participants can receive monetary benefits as long as they participate, no matter their responses or engagement. • "Conditional" refers to scenarios where participants can receive monetary benefits if they satisfy specific criteria after participation.

[a]Dynamic consent is a process facilitated by collaborative and online digital platforms that allows participants to regularly check research activities and modify their consent for any upcoming research projects [46,47]. Dynamic consent can also be described as an approach to consent that enables ongoing engagement and communication between individuals and the users and custodians of their data [49].

Data Analysis and Visualization

To determine the leading countries and their contributions to various publication types, we compared the counts and percentages of publications by type and location. In addition, we visualized the number of publications by year to evaluate publication trends. We also visualized publication counts by year and location for original research publications. Reporting dimensions, including factors concerning "participant," "device," "data," "study," and "ethics" (Table 2), were analyzed

XSL•FO
RenderX

for original research publications only. We charted the frequency and percentage of original research studies that included these 21 reporting dimension factors.

In addition, 4 factors with multiple-choice options were visualized using Venn (if options ≤3) or alluvial (if options >3) diagrams after excluding publications that did not report these factors. These 4 factors included 1 "device" dimension factor (device owner), 2 "data" dimension factors (data type and data owner), and 1 "study" dimension factor (research scenario). We also analyzed publications according to the frequency of data types collected. We assessed missing factors among original research publications to evaluate changes in reporting deficiencies for each reporting dimension. We compared the percentages of publications missing factors between 2 temporal groups (2010-2016 and 2017-2021). All analyses and visualizations were performed using R (version 4.2.1; R Core Team).

Results

Search and Data Charting

Our search strategy yielded 3206 results. A total of 100 publications remained following screening [50-149]. All 100 articles are included in our data analyses (Figure 1; bibliographic information in Multimedia Appendix 2).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping reviews) flow diagram.

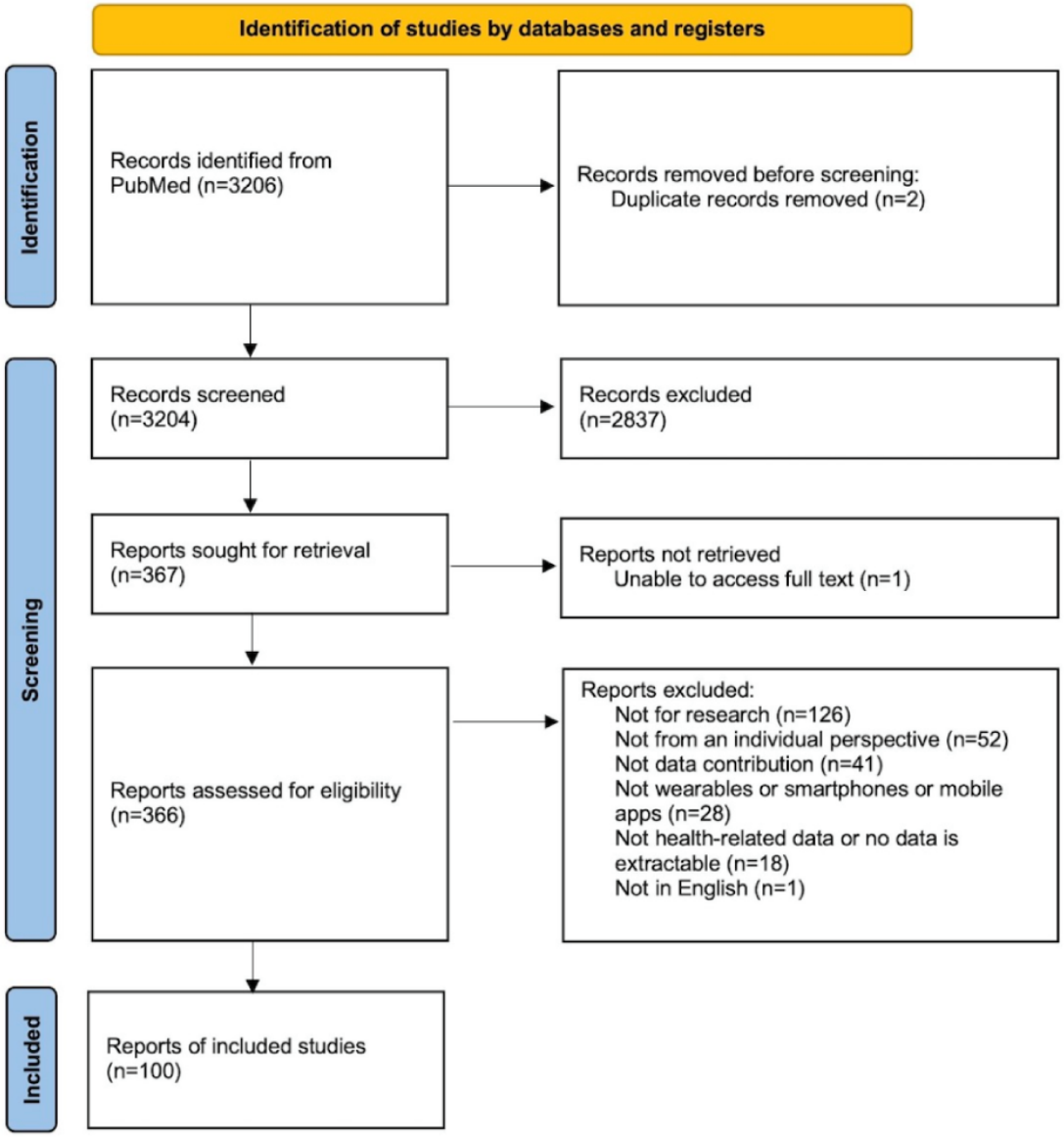

Trends by Publication Year, Location, and Publication Type

Publication year and location were extracted from the 100 included articles and organized by publication type (Figure 2, Multimedia Appendix 3). The majority of publications had a first author from the United States (46%, 46/100), European Union (20%, 20/100), United Kingdom (11%, 11/100), or Canada (6%, 6/100). Publications designated as "Research" (original research) made up the largest proportion of publications (45%, 45/100). These were followed by publications designated as "Tech" (20%, 20/100), "Other" (17%, 17/100), "Review" (13%, 13/100), and "Protocol" (5%, 5/100). Refer to Table 2 for definitions of these publication types.

XSL•FO
RenderX

There was a spike in total publications in 2016, with 11 publications in 2016 compared with 3 or fewer annually beforehand. Since 2016, a minimum of 9 publications have been published annually. There was another spike in output in 2019, with 19 publications. Other trends observed were US-dominant "Tech" publications and diversely geolocated "Review" publications.

Figure 2. Bibliographic details of all publications by year and original research publications by year and location (data from January 2010 to August 2021).

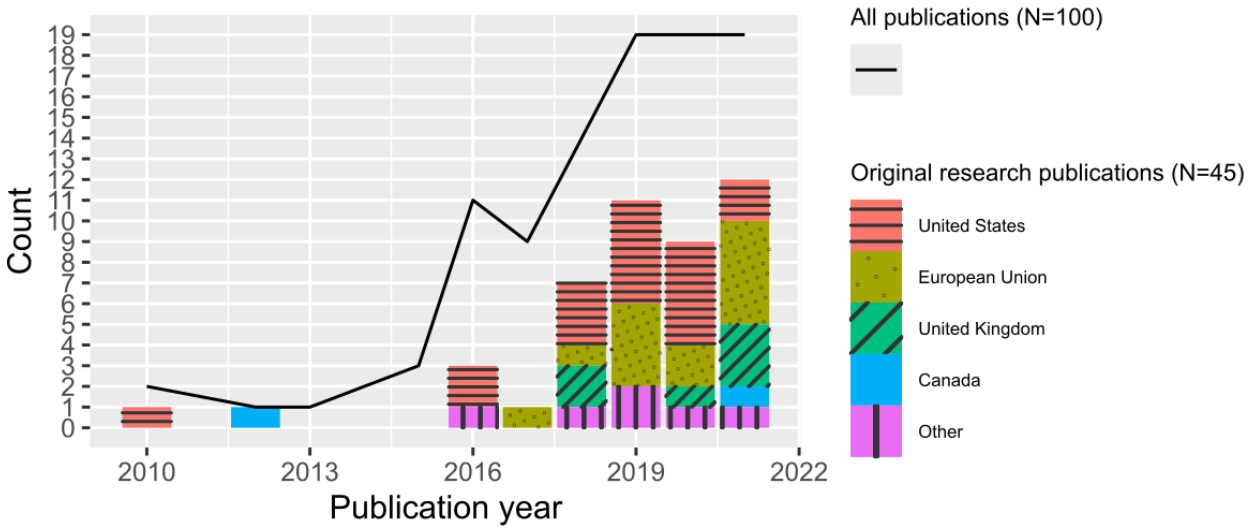

Reporting in Original Research Publications

Reporting of Participant Factors

Most publications (78%, 35/45) declared or implied that the data subject and contributor were the same individual. Almost all data subjects (96%, 43/45) were living at the time of data contribution. In most publications, analyses were conducted using data generated by data subjects for themselves during routine use (56%, 25/45) rather than explicitly for research (11%, 5/45; Table 3).

XSL•FO
RenderX

Table 3. Counts and percentages of original research publications along with reporting factors.

Reporting factors	Original research publications (N=45), n (%)
"Participant" dimension	
Identity of data subject and data contributor (Are the data subject and the data contributor the same individual?)	
Yes	35 (78)
No	0 (0)
Mixed	6 (13)
Not available	4 (8)
Living status of the data subject	
Living	43 (96)
Both living and deceased	1 (2)
Not available	1 (2)
Purpose of data generation	
For research	5 (11)
For self	25 (56)
Not available	15 (33)
"Device" Dimension	
Device owner (multiple choices possible)[a]	
Data subject	19 (42)
Data contributor	18 (40)
Researcher	9 (20)
Not available	17 (38)
Device type	
Wearable	7 (16)
Smartphone	17 (38)
Both	20 (44)
Not available	1 (2)
Data capture mode	
Active	8 (18)
Passive	8 (18)
Both	22 (49)
Not available	7 (16)
"Data" Dimension	
Data type (multiple choices possible)[a]	
Location	18 (40)
Genetics, biospecimen	2 (4)
Genetics, non-biospecimen	3 (7)
Claims or administrative	1 (2)
Clinical	25 (56)
Mental health or lifestyle	38 (84)
Data owner (multiple choices possible)[a]	
Data subject	3 (7)

XSL•FO
RenderX

Reporting factors	Original research publications (N=45), n (%)
Data contributor	4 (9)
Researcher	0 (0)
Company	3 (7)
The public	0 (0)
Not available	41 (91)
"Study" Dimension	
Participant access to data after contribution	
Yes	7 (16)
No	0 (0)
Not available	38 (84)
Research scenario (multiple choices possible)[a]	
Public health	10 (22)
Commercial	16 (36)
Nonprofit	37 (82)
Not available	2 (5)
Purpose of data collection	
Primary	18 (40)
Secondary	0 (0)
Both	12 (27)
Not available	15 (33)
Used crowdsourcing platform to collect data	
Yes	3 (7)
Research design	
Observational	41 (91)
Experimental	4 (9)
Design duration	
Cross-sectional	25 (56)
Longitudinal	18 (40)
Not available	2 (4)
Attrition reported	
Yes	14 (31)
"Ethics" Dimension	
Consent reported	
Yes	41 (91)
Informed consent reported	
Yes	28 (62)
Consent type	
Blanket	0 (0)
Broad	2 (4)
Tiered	1 (2)
Meta	0 (0)
Explicit	5 (11)

XSL•FO
RenderX

Reporting factors	Original research publications (N=45), n (%)
More than one type of consent	2 (4)
Not available	35 (78)
Consent subject	
Data subject	4 (9)
Data contributor	4 (9)
Both	29 (64)
Not available	8 (18)
Right to opt-out reported	
Yes	9 (20)
Monetary benefits	
Unconditional	2 (4)
Conditional	9 (20)
No benefit	6 (13)
Not available	28 (62)

[a]Levels below reporting factors were cleaned to be binary if multiple choices were possible.

Reporting of Device Factors

Among publications reporting the device owner, more reported that devices were owned by participants (data subject or data contributor) than were owned by researchers (75%, 21/28 vs 32%, 9/28; Figure 3A). Furthermore, 2 publications reported a mix of device ownership (owned by data subject, data contributor, and researcher), indicating a transfer of ownership (ie, participants retained the provisioned device after the study). Wearables were more often used with smartphones than used alone (wearables and smartphones 44%, 20/45 vs wearables alone 16%, 7/45; Table 3). Most publications reported a combination of active and passive data capturing, and active-alone and passive-alone data capturing were equally popular (active and passive 49%, 22/45 vs active-alone 18%, 8/45 vs passive-alone 18%, 8/45; Table 3).

Figure 3. Publication counts for 4 reporting factors with multiple choices possible, excluding publications that did not report these factors, (A) device owner, (B) data type, (C) data owner, and (D) research scenario. The data type figure includes a table presenting publication counts by the number of reported data types.

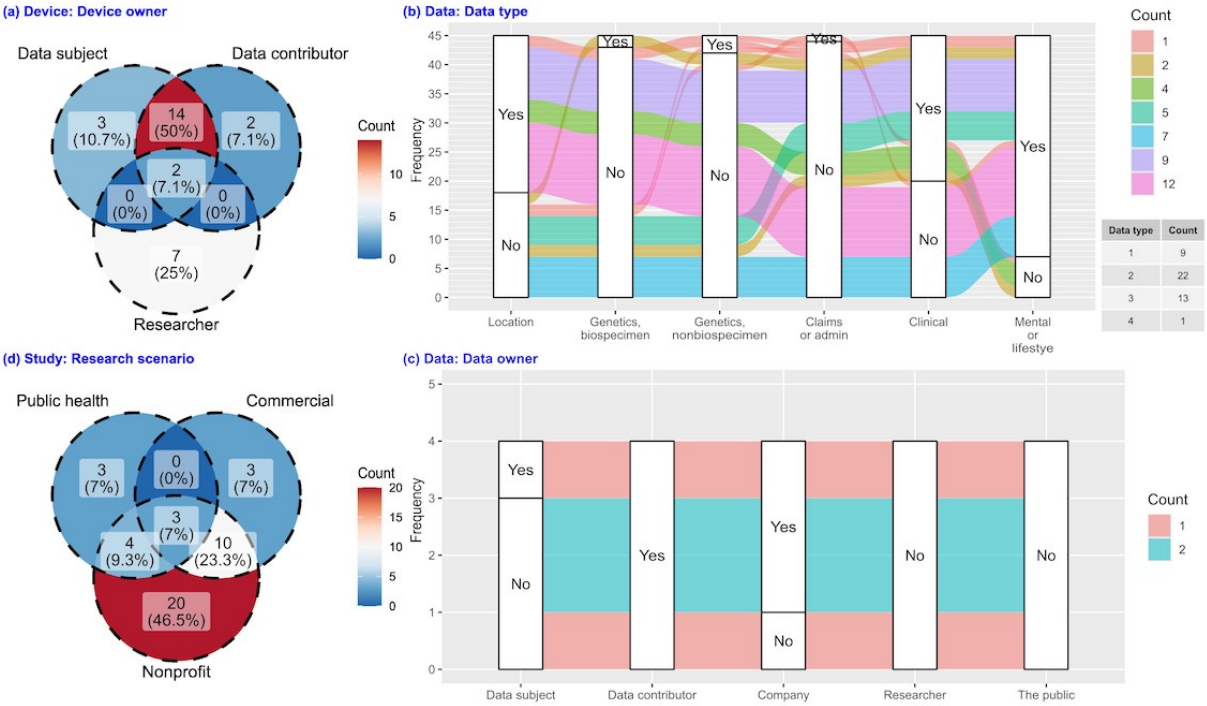

XSL•FO
RenderX

Reporting of Data Factors

With multiple choices possible, the mental health or lifestyle data type (84%, 38/45; Table 3) was the most collected, followed by clinical (56%, 25/45) and location (40%, 18/45) data types. In comparison, genetics data types (biospecimen 4%, 2/45 and nonbiospecimen 7%, 3/45 [Table 3]; in total 11%, 5/45 [Figure 3B]) and insurance claims or administrative data type (2%, 1/45; Table 3) were seldom collected. Most publications reported more than one data type (80%, 36/45; refer to table in Figure 3B). Overall, 40% (18/45) of publications' collected data included clinical data and mental health or lifestyle data (Figure 3B). No dominant data ownership arrangement was observed (data subject 7%, 3/45 vs data contributor 9%, 4/45 vs company 7%, 3/45; Table 3).

Reporting of Study Factors

Most studies that reported research scenarios were conducted within a strictly nonprofit milieu (47%, 20/43; Figure 3D). No study collected data solely for secondary analysis purposes (Table 3). A limited number of publications reported using crowdsourcing platforms to collect data (7%, 3/45; Table 3). Most studies were observational (91%, 41/45; Table 3). A prevalence of cross-sectional designs over longitudinal designs was observed (56%, 25/45 vs 40%, 18/45; Table 3).

Reporting of Ethics Factors

No publication reported blanket or meta-consent (Table 3). Among those publications reporting monetary benefits for participation, more than one-third reported giving no monetary benefits to participants (35%, 6/17; Table 3).

Reporting Deficiencies in Original Research Publications

The number of publications that missed reporting at least 1 factor under the 5 reporting dimensions broke down as follows: 36% (16/45) for the "participant" dimension, 42% (19/45) for the "device" dimension, 96% (43/45) for the "data" dimension, 78% (35/45) for the "study" dimension, and 96% (43/45) for the "ethics" dimension (Figure 4).

Figure 4. Missingness in 21 reporting factors in original research publications (N=45). The red color represents instances of missingness. Publication counts on the right vertical axis are highlighted in red if missingness existed in more than half of the publications for a particular reporting factor (excluding the factor "CrowdsourcingPlatform" for those that used crowdsourcing platforms to collect data). For a definition of each reporting factor, refer to Table 3.

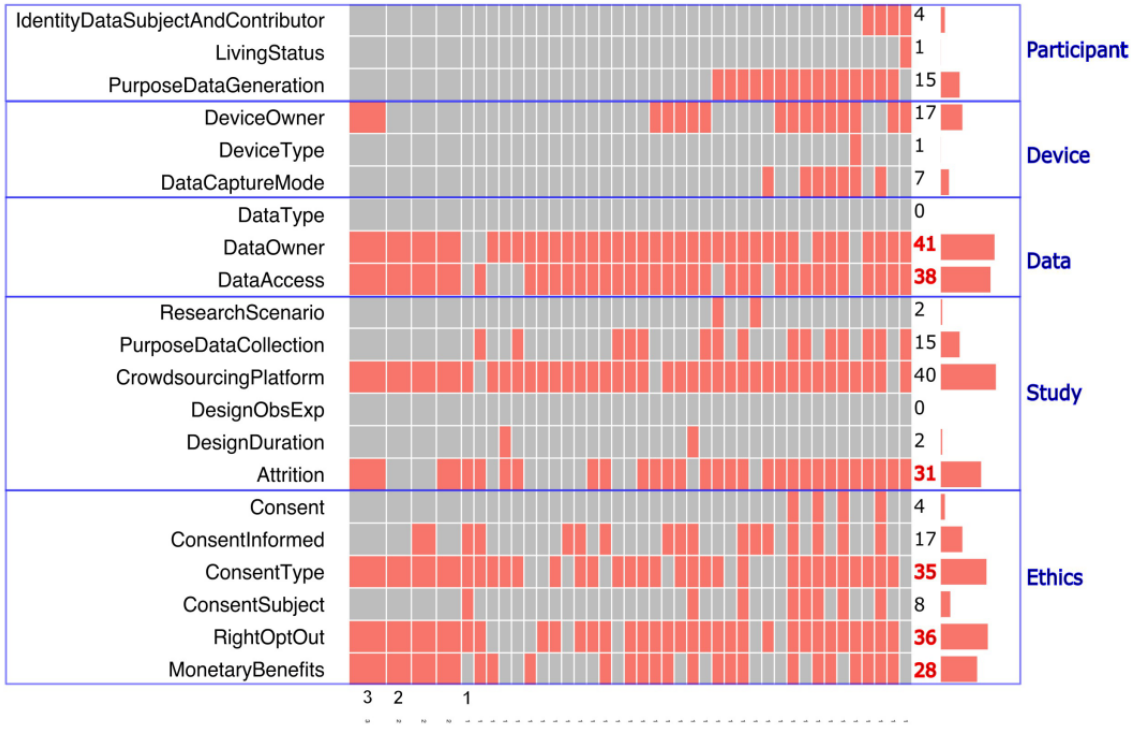

Except for the data type factor ("DataType") and the observational or experimental study design factor ("DesignObsExp"), missingness existed in most factors under the 5 reporting dimensions. In total, 6 factors had missingness in more than half of the original research publications: data owner (91%, 41/45), participant access to data after data contribution (84%, 38/45), right to opt out (80%, 36/45), consent type (78%, 35/45), monetary benefits (62%, 28/45), and attrition (69%, 31/45). These 6 factors were often missing together (Figure 4).

The factor for marking studies that used crowdsourcing platforms to collect data ("CrowdsourcingPlatform") was excluded from the reporting deficiency analysis because using these platforms is optional for research.

Trends in reporting deficiencies were observed in original research publications over time. Reporting deficiencies were measured by the absence, or missingness, of specific reporting factors, with increased missingness reflecting declining reporting performance and decreased missingness reflecting improving reporting performance. The reporting performance of some

XSL•FO
RenderX

reporting factors showed declines and others improvements when compared across the years 2010-2016 and 2017-2021. Furthermore, 2 factors in the "data" dimension exhibited substantial declines in reporting performance: data owner (58% change in missingness; from 2/5, 40% in 2010-2016 to 39/40, 98% in 2017-2021) and participant access to data after contribution (50% change in missingness; from 2/5, 40% in 2010-2016 to 36/40, 90% in 2017-2021). In contrast, 4 factors in the "device" and "ethics" dimensions exhibited substantial improvements in reporting performance: data capture mode (–28% change in missingness; from 2/5, 40% in 2010-2016 to 5/40, 13% in 2017-2021), device type (–20% change in missingness; from 1/5, 20% in 2010-2016 to 0/40, 0% in 2017-2021), consent type (–25% change in missingness; from 5/5, 100% in 2010-2016 to 30/40, 75% in 2017-2021), and consent subject (–25% change in missingness; from 2/5, 40% in 2010-2016 to 6/40, 15% in 2017-2021). The magnitude of change in the percentage of publications with missingness was substantial if the absolute value of the difference was equal to or greater than 20% (Figure 5 and Multimedia Appendix 4).

Figure 5. Shift in the reporting deficiencies of original research publications, measured by the % of missingness across 21 factors between 2010-2016 and 2017-2021 (sorted by % change). For a definition of each reporting factor, refer to Table 3.

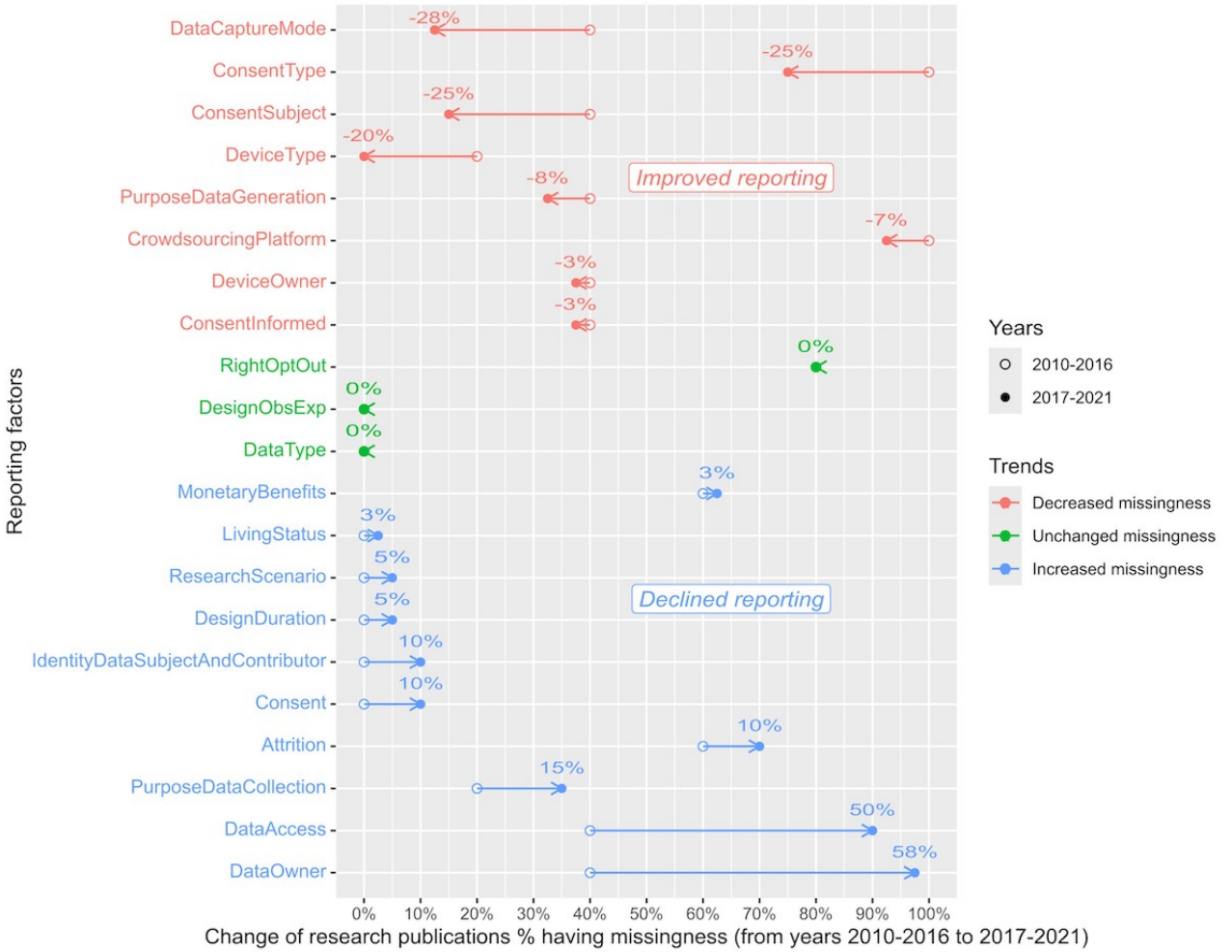

Discussion

Principal Results

We have characterized the reporting practices of published research studies that collected PGHD using mobile devices. Our findings show that the United States has gained momentum in publishing in this area over the past decade (Figure 2). In addition, a high proportion of the original research articles we identified used participant-owned devices or a BYOD strategy if they reported device ownership (75%, 21/28; Figure 3A), with few studies collecting data using crowdsourcing platforms (7%, 3/45; "Used crowdsourcing platform to collect data" in Table 3). We also found reporting deficiencies in the "data" and "ethics" dimensions among original research articles, where most original research articles were deficient in at least 1 factor concerning "data" (96%, 43/45) or "ethics" (96%, 43/45; Figure 4).

Trends by Publication Year, Location, and Publication Type

Our analysis of publications by year and location echoed trends seen in the field, with publication spikes in 2016 and 2019 and stable development afterward (Figure 2). Like others [15,24], we found that the United States is a leading contributor to these publications. Furthermore, the United States dominates contributions in several publication types, including original research (Multimedia Appendix 3).

XSL•FO
RenderX

Reporting in Original Research Publications

In the "participant" reporting dimension, we found that the mobile device data used for research were more often generated by participants for themselves through routine use rather than generated explicitly for research (56%, 25/45 vs 11%, 5/45; "Purpose of data generation" in Table 3). However, the quality of data acquired for routine use may not meet data quality requirements for research. For instance, individuals engaging in routine device use may not be trained in the best way to wear and use the device to ensure high-quality data collection, thereby leading to potential issues with low or missing wear time and user errors that impact the completeness and correctness of the collected data [150]. Systematic data quality assessment tools have been developed specifically for the secondary use of PGHD collected from mobile devices. These tools address the quality issues of routine use data by recommending research reporting clarity on the definition of nonwear time, the threshold for valid records, the definition of data completeness, and the identification of outliers [151].

In the "device" reporting dimension, among publications that reported device ownership, we found that three-quarters of original research publications (21/28; Figure 3A) reported using mobile devices owned by participants (data contributors or data subjects), which implies the popularity of the BYOD strategy. As mentioned in the introduction, BYOD may bring bias into demographics through disparate accessibility to device and self-selection, thus a comprehensive guideline on using BYOD to collect PGHD, which is not available so far, is a need. As an alternative to BYOD, some researchers provide devices to participants (9/28; Figure 3A). In our review, for studies that provided devices to participants, participants in 2 of them (Figure 3A) retained the device after the study (shown in Figure 3A as a mix of device ownership). This strategy may serve as an incentive and help improve device use simultaneously. In our own research [152], however, we found that providing a device when recruiting study participants for cardiovascular wearable studies did not improve sample representativeness in age, race, and education. Thus, more work is needed to understand effective incentives to participate in research that collects PGHD from mobile devices such that recruitment would lead to a more representative study sample.

We found that wearables were predominantly used as secondary devices connected to smartphones rather than as standalone devices (44%, 20/45 vs 16%, 7/45; "Device type" in Table 3), indicating that barriers may exist to the direct collection of data from wearables for research purposes. Evidence to support this supposition is provided by a study published in 2023 [153] that reported that among those who expressed a willingness to share data from a wearable, only 25% deposited their data in the study's research database. The primary reason for the refusal to deposit data was the inconvenience of the data transfer process. Previous research findings [154] have also raised concerns that wearables may be less sustainable than smartphones for remote monitoring over long periods [155]. However, the reliance on a smartphone for data transfer is not ideal if other data are not collected by the smartphone, such as information about heart rate or sleep. A review published in 2022 [156] found that a limited number of mobile health apps use Bluetooth, and an even smaller number use standard Bluetooth Low Energy. This indicates a potentially low capability of smartphone apps to interact with external devices, including different types of wearables.

Researchers adopt mobile devices for research due to their ability to capture data passively, thus reducing recall bias introduced by active approaches (eg, surveys). Our findings, however, suggest that active and passive data capturing were equally popular (18%, 8/45 for each; "Data capture mode" in Table 3) and frequently used in tandem (49%, 22/45; "Data capture mode" in Table 3). Researchers may have hesitated to depend solely on passive data collection because this approach can result in low participant engagement due to privacy concerns [157] and potential data quality issues [150]. An approach that uses both passive and active data collection can offer more comprehensive data and enable cross-validation of responses through data linkages.

In the "data" reporting dimension, a majority of the publications reported collecting multiple data types (80%, 36/45; refer to table in Figure 3B). Furthermore, 40% (18/45) of publications' collected data include clinical data and mental health or lifestyle data (Figure 3B). A subset of these publications indicated that clinical data were collected from participants' electronic health records, reflecting progress made to integrate mobile device data into electronic health records [158]. Studies seeking to connect multiple data types from study participants should exercise caution due to the increased risk of participant reidentification [6].

We found that certain data types were more infrequently collected than others. For example, insurance claim and administrative data were rarely collected (2%, 1/45; "Data type" in Table 3). This gap may be because of difficulties linking these data to other data [159]. Also, despite advancements in informatics infrastructure (eg, the HL7 FHIR [Fast Health Interoperability Resources] API) enabling patients to download these data (eg, Medicare claims and encounter data through Blue Button [160]), only a limited number of beneficiaries have done so. For example, while Blue Button, available since 2010, has been used by over one million Medicare beneficiaries [161], this is a small fraction of the 65 million people covered by Medicare as of 2022 [162]. Genetics data collection was also rare (11%, 5/45; Figure 3B), which is consistent with the prevailing observation that a limited number of genomics mobile apps serve to collect data, compared with other tasks (eg, education and communication) [163,164].

In the "study" reporting dimension, only 7% of original research publications used crowdsourcing platforms to collect data (3/45; "Used crowdsourcing platform to collect data" in Table 3), which may indicate a scarcity of streamlined mechanisms for using these platforms. Given the growing popularity of BYOD research, there is an opportunity for improved mechanisms to leverage such platforms to transfer data from device owners to researchers. We also found that the majority of the studies were observational (91%, 41/45; "Research design" in Table 3), aligning with previous research [26]. Although mobile devices enable facile longitudinal data collection, we discovered a slightly greater prevalence of cross-sectional studies over

XSL•FO
RenderX

longitudinal studies (56%, 25/45 vs 40%, 18/45; "Design duration" in Table 3). This could be due to researchers' apprehensions about device fatigue [165] and stakeholders' concerns regarding privacy [166].

In the "ethics" reporting dimension, we explored factors such as consent models and monetary benefits because they are relevant to the likelihood of participants to share data. For example, consent models, as noticed in a survey conducted by Ko¨ngeter et al [167], have different acceptance rates among patients with cancer. In our scoping review, we found no publication reported blanket or meta consent ("Consent type" in Table 3). Since no studies collected data purely for secondary analysis ("Purpose of data collection" in Table 3), the preference for nonblanket consent was justified. Despite debates about meta consent [168] and dynamic consent [49] as solutions for modern research, our finding that no studies used meta (dynamic) consent aligns with critics who posit that these innovative approaches require highly participatory technological platforms and may introduce bias due to participants' technical competence [169]. Over one-third of the original research publications that reported on monetary benefits provided no monetary benefits to participants (35%, 6/17; "Monetary benefits" in Table 3). Research generally indicates that even small monetary incentives increase consent and response rates [170], irrespective of risk level [171]. However, the impact of monetary benefits on wearable study participation remains undetermined and requires examination on a case-by-case basis (eg, for studies conducted on Amazon Mechanical Turk [9]) given the diversity of mobile device studies.

Prevalence of Reporting Deficiencies in Original Research Publications

Our results show a high level of reporting deficiencies in the "data" and "ethics" dimensions (Figure 4), with 5 reporting factors missing in over half of the original research publications—data ownership (91%, 41/45), participant access to data after data contribution (84%, 38/45), right to opt out (80%, 36/45), consent type (78%, 35/45), and monetary benefits (62%, 28/45). Among these 5 factors, we also noticed a trend over time of significantly declined performance in the reporting of data ownership (40% missingness, 2010-2016 vs 98% missingness, 2017-2021; Figure 5) and participant access to data after contribution (40% missingness, 2010-2016 vs 90% missingness, 2017-2021). However, over time, there was a significantly improved performance in reporting consent type (100% missingness, 2010-2016 vs 75% missingness, 2017-2021). In the "study" dimension, nearly 70% (31/45; Figure 4) of original research publications failed to report study participant attrition.

Recommendations for Reporting Based on Findings From This Review

Our findings highlight several areas to consider when reporting research with PGHD collected using mobile devices. Recommendations for reporting based on findings from this review are summarized in Table 4. Further work is needed, however, to gain consensus on a more comprehensive set of reporting items. Such a consensus on best practices would greatly facilitate responsible and reproducible research and improve research quality and impact [172]. For example, a 2012 systematic review indicated a positive association between journal endorsements of a reporting guideline and the quality of reporting in randomized clinical trials published in those journals [173]. Another study looking at the publications from one journal found that those publications whose papers received editorial interventions designed to ensure adherence to reporting guidelines were more highly cited, with citation counts 43% higher than those publications that had not received the editorial interventions [174].

XSL•FO
RenderX

Table 4. Recommendations for reporting research with PGHD[a] collected using mobile devices.

Reporting dimension	Reporting phenomena identified in this scoping review	Justification to include reporting dimension as a best practice
"Participant" (1 item)	• "Purpose of data generation" item: Many publications reported the use of real-world PGHD from study participants (ie, not collected in controlled or experimental settings).	• More detailed practices for correcting data quality issues that can arise with real-world PGHD should be documented so that others pursuing similar research can learn what works and does not work.
"Device" (2 items)	• "Device owner" item: Many papers reported using a device owned by participants. • "Device type" item: Many publications reported using wearables as secondary devices (ie, used with smartphones instead of as standalone devices).	• The practice of using participant-owned devices should be monitored closely as there is potential for disparities in study participation for some groups due to a lack of device access. • More details on the data transfer process should be given in study designs. It influences the ability to participate in research and the types of data that can be collected.
"Data" (2 items)	• "Data owner" and "participant access to data after contribution" items: Publications largely lack descriptions of data ownership and often fail to describe how study participants can access their shared data.	• If strategies for empowering study participants to maintain data ownership are not documented, studies cannot be performed to measure the effectiveness of these strategies or point to new approaches.
"Study" (1 item)	• "Attrition (reported)" item: Attrition is often not reported.	• Attrition can introduce selection bias and reduce statistical power due to data missingness [35]. Attrition should be documented so that others can evaluate the quality and fairness of data analyses.
"Ethics" (3 items)	• "Consent type," "right to opt out (reported)," and "monetary benefits" items: Descriptions of consent types, the rights of participants to opt out, and monetary benefits (if provided) are largely missing from publications.	• Options provided to participants at the time of recruitment may influence their willingness to participate in research [167,170] and thereby influence the composition of the study population. Such options also inform the ethical assessment of a study.

[a]PGHD: person-generated health data.

Limitations

Our review has limitations related to how we gathered evidence from the literature, our choice of reporting factors, and one of our data analysis methods.

Our PubMed search strategy (Textbox 1) has concepts with explicit data donation and data sharing terms to retrieve publications that describe participant data contributions. Because of this design, the search strategy only retrieved publications whose titles and abstracts had data donation and data sharing terms. Possibly, our search missed relevant publications that described participant data contributions only in the full text. The articles we analyzed were only collected from PubMed. Articles were not retrieved from other databases (eg, Embase or Web of Science), from the reference lists of included articles, or from the gray literature. Thus, it is possible that articles only available through these other sources were overlooked. Our preliminary searches, however, suggested that PubMed provided literature coverage sufficiently comprehensive for our research objectives.

While we selected reporting factors relevant to our goal of reviewing research that collects PGHD using mobile devices with the perspective of an individual's data contribution, we may have excluded relevant factors. For example, we did not include the demographic factors of study populations, which may be more broadly relevant to wearable PGHD research. Detecting the low representation of some groups when using wearables for PGHD collection can indicate where biases exist that could lead to downstream inequities. This need is apparent based on findings that aging populations were underrepresented in studies with wearables [175] and that studies of cardiovascular disease that involve wearables [175]. Such a need requires domain-specific examinations, and we regard it out of the scope of this review, which aims to explore the landscape as the first step. As an example of a second-step effort, we have investigated the studies of cardiovascular disease that involve wearables, and we noticed that they did not reflect the demographics of patient populations in terms of age, race, education, cigarette smoking status, and hypertension status [152].

We lowered the granularity of some factors to avoid issues of imbalance in categorical analysis. For example, when analyzing research scenarios, we consolidated academic institutions, for-profit commercial companies, corporations or foundations, independent research organizations, patient-led groups, citizen scientists, and so on, into 3 categories. Thus, our findings do not necessarily represent the full range of possible ways of describing this research.

Despite these limitations, we are confident that having used strict procedures to reduce bias (eg, 2-level screening, independent charting, and a predefined charting template), this review offers a solid foundation for future research. We also feel that the full PubMed query we provide (Textbox 1) is a valuable contribution to other researchers since it was developed and tested for wearables and smartphones and since publications

XSL•FO
RenderX

in the emerging field of mobile devices and PGHD are not well indexed.

Conclusions

We identified trends and patterns in the reporting of research that collects PGHD using mobile devices. Given the growing interest in using a BYOD model among researchers, there is an opportunity for a broader use of crowdsourcing platforms for data transfer from participant-owned devices. There is also the opportunity to develop best practices that address observed reporting deficiencies in this research and make it more reproducible and responsible. Based on our findings, we recommend 9 items for enhanced reporting.

Acknowledgments

SS, ZL, and COT were supported by National Institutes of Health grant NHGRI R35 HG010714.

Data Availability

The data underlying this article are readily available upon reasonable request to the corresponding author.

Authors' Contributions

All authors had full access to the data in the study, take responsibility for data integrity and analysis accuracy, and have approved the final manuscript. Study design was contributed by SS and COT. Review strategy was contributed by SS, RW, and COT. Data collection was managed by SS, MA, RHY, and ZL. Data analysis was done SS. Drafting was contributed by SS. Intellectual inputs were contributed by RW, DJHM, and COT. All authors contributed to manuscript revision. Supervision was handled by COT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOC File , 124 KB - jmir_v27i1e51955_app1.doc]

Multimedia Appendix 2
Bibliographic information of the articles included in the scoping review.
[DOC File , 327 KB - jmir_v27i1e51955_app2.doc]

Multimedia Appendix 3
Publication types across publication years and publication locations.
[DOC File , 37 KB - jmir_v27i1e51955_app3.doc]

Multimedia Appendix 4
Reporting deficiencies in original research publications over time, as measured by the missingness of reporting factors.
[DOC File , 47 KB - jmir_v27i1e51955_app4.doc]

References

1. Martin M. Computer and Internet Use in the United States: 2018. Suitland, MD: United States Census Bureau; 2021.
2. Mobile technology and home broadband 2021. Pew Research Center. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2023-01-31]
3. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-04-07]
4. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
5. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2023-01-31]
6. Rothstein MA, Wilbanks JT, Beskow LM, Brelsford KM, Brothers KB, Doerr M, et al. Unregulated health research using mobile devices: ethical considerations and policy recommendations. J Law Med Ethics 2020;48(1_suppl):196-226 [FREE Full text] [doi: 10.1177/1073110520917047] [Medline: 32342752]
7. Sim I. Mobile devices and health. N Engl J Med 2019;381(10):956-968. [doi: 10.1056/nejmra1806949]

XSL•FO
RenderX

8. Part 46, Subpart A version of the Final Rule "Federal Policy for the Protection of Human Subjects". 2017. URL: https://tinyurl.com/ypsdr69m [accessed 2024-10-18]
9. Taylor CO, Flaks-Manov N, Ramesh S, Choe EK. Willingness to share wearable device data for research among mechanical turk workers: web-based survey study. J Med Internet Res 2021;23(10):e19789 [FREE Full text] [doi: 10.2196/19789] [Medline: 34673528]
10. IDO T. Apple's ResearchKit frees medical research. Nat Biotechnol 2015;33(4):322. [doi: 10.1038/nbt0415-322]
11. Jardine J, Fisher J, Carrick B. Apple's ResearchKit: Smart Data Collection for the Smartphone era?. London, England: SAGE Publications; 2015:294-296.
12. Misra S. ResearchStack being developed as android's ResearchKit cousin, beta in January 2016. iMedicalApps. 2015. URL: https://www.imedicalapps.com/2015/11/researchkit-cob-open-mhealth/ [accessed 2023-05-25]
13. Wicklund E. ResearchStack goes live, opening mHealth studies to the android ecosystem. Xtelligent Healthcare Media, mHealth Intelligence. 2016. URL: https://mhealthintelligence.com/news/researchstack-goes-live-opening-mhealth-studies-to-the-android-ecosystem [accessed 2023-05-25]
14. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, parkinson disease mobile data collected using ResearchKit. Sci Data 2016;3(1):160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
15. Cao J, Lim Y, Sengoku S, Guo X, Kodama K. Exploring the shift in international trends in mobile health research from 2000 to 2020: bibliometric analysis. JMIR Mhealth Uhealth 2021;9(9):e31097 [FREE Full text] [doi: 10.2196/31097] [Medline: 34494968]
16. Cho PJ, Yi J, Ho E, Shandhi MMH, Dinh Y, Patil A, et al. Demographic imbalances resulting from the bring-your-own-device study design. JMIR Mhealth Uhealth 2022;10(4):e29510 [FREE Full text] [doi: 10.2196/29510] [Medline: 34913871]
17. Mortensen K, Hughes TL. Comparing amazon's mechanical turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med 2018;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
18. Dhingra LS, Aminorroaya A, Oikonomou EK, Nargesi AA, Wilson FP, Krumholz HM, et al. Use of wearable devices in individuals with or at risk for cardiovascular disease in the US, 2019 to 2020. JAMA Netw Open 2023;6(6):e2316634 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.16634] [Medline: 37285157]
19. Vrijheid M, Richardson L, Armstrong BK, Auvinen A, Berg G, Carroll M, et al. Quantifying the impact of selection bias caused by nonparticipation in a case-control study of mobile phone use. Ann Epidemiol 2009;19(1):33-41. [doi: 10.1016/j.annepidem.2008.10.006] [Medline: 19064187]
20. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 2012;14(2):e46 [FREE Full text] [doi: 10.2196/jmir.1988] [Medline: 22397809]
21. Holko M, Ratsimbazafy F, Marginean K, Natarajan K, Cho S, Schilling J. Fitbit' Bring Your Own Device' data in the All of Us Research Program. United States: American Medical Informatics Association; 2020.
22. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019;9(2):010318 [FREE Full text] [doi: 10.7189/jogh.09.020318] [Medline: 31788229]
23. Paulus JK, Kent DM. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med 2020;3:99 [FREE Full text] [doi: 10.1038/s41746-020-0304-9] [Medline: 32821854]
24. El-Sherif DM, Abouzid M. Analysis of mHealth research: mapping the relationship between mobile apps technology and healthcare during COVID-19 outbreak. Global Health 2022;18(1):67 [FREE Full text] [doi: 10.1186/s12992-022-00856-y] [Medline: 35765078]
25. Fischer F, Kleen S. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: scoping review. J Med Internet Res 2021;23(1):e17691 [FREE Full text] [doi: 10.2196/17691] [Medline: 33480850]
26. Huhn S, Axt M, Gunga H, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
27. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022;43(1):117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
28. Vijayan V, Connolly JP, Condell J, McKelvey N, Gardiner P. Review of wearable devices and data collection considerations for connected health. Sensors (Basel) 2021;21(16):5589 [FREE Full text] [doi: 10.3390/s21165589] [Medline: 34451032]
29. Future of Privacy Forum. 2016. URL: https://fpf.org/wp-content/uploads/2016/08/FPF-Best-Practices-for-Wearables-and-Wellness-Apps-and-Devices-Final.pdf [accessed 2022-04-08]
30. Cortez AHP, Mitchell E, Riehl V, Smith P. Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024. Washington, DC: Office of the National Coordinator for Health Information Technology; 2018.
31. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. BMJ 2021;372:m4856 [FREE Full text] [doi: 10.1136/bmj.m4856] [Medline: 33436424]

XSL•FO
RenderX

32. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
33. Chatzipavlou IA, Christoforidou SA, Vlachopoulou M. A recommended guideline for the development of mHealth apps. Mhealth 2016;2:21 [FREE Full text] [doi: 10.21037/mhealth.2016.05.01] [Medline: 28293597]
34. Guide to integrate patient-generated digital health data into electronic health records in ambulatory care settings. AHRQ. URL: https://digital.ahrq.gov/health-it-tools-and-resources/patient-generated-health-data-i-patient-reported-outcomes/practical-guide [accessed 2022-04-08]
35. Olaye IM, Belovsky MP, Bataille L, Cheng R, Ciger A, Fortuna KL, et al. Recommendations for defining and reporting adherence measured by biometric monitoring technologies: systematic review. J Med Internet Res 2022;24(4):e33537 [FREE Full text] [doi: 10.2196/33537] [Medline: 35436221]
36. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
37. Grossetta NHK, Wang L. The Yale MeSH Analyzer. New Haven, CT: Cushing/Whitney Medical Library URL: http://mesh.med.yale.edu/ [accessed 2021-08-17]
38. Covidence Systematic Review Software. Melbourne, Australia: Veritas Health Innovation; 2024.
39. Torous J, Staples P, Slaters L, Adams J, Sandoval L, Onnela J, et al. Characterizing smartphone engagement for schizophrenia: results of a naturalist mobile health study. Clin Schizophr Relat Psychoses 2017 Aug 04 [FREE Full text] [doi: 10.3371/CSRP.JTPS.071317] [Medline: 28777029]
40. Bietz M, Patrick K, Bloss C. Data donation as a model for citizen science health research. Citizen Science: Theory and Practice 2019;4(1). [doi: 10.5334/cstp.178]
41. Krutzinna J, Floridi L. The Ethics of Medical Data Donation. Cham, Switzerland: Springer; 2019.
42. Schmietow B, Marckmann G. Mobile health ethics and the expanding role of autonomy. Med Health Care Philos 2019;22(4):623-630. [doi: 10.1007/s11019-019-09900-y] [Medline: 31011945]
43. Song S, Ashton M, Yoo R, Wright R, Mathews D, Taylor CO. "Data Donation" or "Data Sharing"? A Scoping Review Characterizing Language Use in mHealth Research Involving Person-generated Health Data. Washington, DC: AMIA Annual Symposium 2022; 2022:5-9.
44. Siddiqi AEA, Sikorskii A, Given CW, Given B. Early participant attrition from clinical trials: role of trial design and logistics. Clin Trials 2008;5(4):328-335 [FREE Full text] [doi: 10.1177/1740774508094406] [Medline: 18697847]
45. Wendler D. Broad versus blanket consent for research with human biological samples. Hastings Cent Rep 2013;43(5):3-4 [FREE Full text] [doi: 10.1002/hast.200] [Medline: 24092578]
46. Wiertz S, Boldt J. Evaluating models of consent in changing health research environments. Med Health Care Philos 2022;25(2):269-280 [FREE Full text] [doi: 10.1007/s11019-022-10074-3] [Medline: 35286521]
47. Informed consent: considerations for biobanks. CloudLIMS. URL: https://cloudlims.com/informed-consent-dynamic-broad-tiered-and-meta-consent-for-biobanking/ [accessed 2022-02-23]
48. Explicit Consent. Health Research Board. URL: https://www.hrb.ie/funding/gdpr-guidance-for-researchers/gdpr-and-health-research/consent/explicit-consent/ [accessed 2022-02-23]
49. Budin-Ljøsne I, Teare HJA, Kaye J, Beck S, Bentzen HB, Caenazzo L, et al. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. BMC Med Ethics 2017;18(4):1-10 [FREE Full text] [doi: 10.1186/s12910-016-0162-9] [Medline: 28122615]
50. Ajana B. Digital health and the biopolitics of the quantified self. Digit Health 2017;3:2055207616689509 [FREE Full text] [doi: 10.1177/2055207616689509] [Medline: 29942580]
51. Arab L, Winter A. Automated camera-phone experience with the frequency of imaging necessary to capture diet. J Am Diet Assoc 2010;110(8):1238-1241. [doi: 10.1016/j.jada.2010.05.010] [Medline: 20656101]
52. Atreja A, Khan S, Rogers JD, Otobo E, Patel NP, Ullman T, HealthPROMISE Consortium Group. Impact of the mobile health PROMISE platform on the quality of care and quality of life in patients with inflammatory bowel disease: study protocol of a pragmatic randomized controlled trial. JMIR Res Protoc 2015;4(1):e23 [FREE Full text] [doi: 10.2196/resprot.4042] [Medline: 25693610]
53. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Erratum to: making sense of big data in health research: towards an EU action plan. Genome Med 2016;8(1):118 [FREE Full text] [doi: 10.1186/s13073-016-0376-y] [Medline: 27821178]
54. Augusto DG, Murdolo LD, Chatzileontiadou DSM, Sabatino JJ, Yusufali T, Peyser ND, et al. A common allele of HLA is associated with asymptomatic SARS-CoV-2 infection. Nature 2023 Aug;620(7972):128-136 [FREE Full text] [doi: 10.1038/s41586-023-06331-x] [Medline: 37468623]
55. Bietz MJ, Bloss CS, Calvert S, Godino JG, Gregory J, Claffey MP, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016;23(e1):e42-e48 [FREE Full text] [doi: 10.1093/jamia/ocv118] [Medline: 26335984]

XSL•FO
RenderX

56. Bloem BR, Marks WJ, Silva de Lima AL, Kuijf ML, van Laar T, Jacobs BPF, et al. The personalized Parkinson project: examining disease progression through broad biomarkers in early Parkinson's disease. BMC Neurol 2019;19(1):160 [FREE Full text] [doi: 10.1186/s12883-019-1394-3] [Medline: 31315608]
57. Bocher E, Petit G, Picaut J, Fortin N, Guillaume G. Collaborative noise data collected from smartphones. Data Brief 2017;14:498-503 [FREE Full text] [doi: 10.1016/j.dib.2017.07.039] [Medline: 29071287]
58. Boker SM, Brick TR, Pritikin JN, Wang Y, von Oertzen T, Brown D, et al. Maintained individual data distributed likelihood estimation (MIDDLE). Multivariate Behav Res 2015;50(6):706-720 [FREE Full text] [doi: 10.1080/00273171.2015.1094387] [Medline: 26717128]
59. Bouras A, Simoes EJ, Boren S, Hicks L, Zachary I, Buck C, et al. Non-hispanic white mothers' willingness to share personal health data with researchers: survey results from an opt-in panel. J Particip Med 2020;12(2):e14062 [FREE Full text] [doi: 10.2196/14062] [Medline: 33064096]
60. Bruno E, Böttcher S, Viana PF, Amengual-Gual M, Joseph B, Epitashvili N, et al. Wearable devices for seizure detection: practical experiences and recommendations from the wearables for epilepsy and research (WEAR) international study group. Epilepsia 2021;62(10):2307-2321. [doi: 10.1111/epi.17044] [Medline: 34420211]
61. Buoite Stella A, AJČEVIĆ M, Furlanis G, Cillotto T, Menichelli A, Accardo A, et al. Smart technology for physical activity and health assessment during COVID-19 lockdown. J Sports Med Phys Fitness 2021;61(3):452-460. [doi: 10.23736/s0022-4707.20.11373-2]
62. Burchert S, Kerber A, Zimmermann J, Knaevelsrud C. Screening accuracy of a 14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: comparison with the PHQ-9 depression screening. PLoS One 2021;16(1):e0244955 [FREE Full text] [doi: 10.1371/journal.pone.0244955] [Medline: 33406120]
63. Burkhardt HA, Brandt PS, Lee JR, Karras SW, Bugni PF, Cvitkovic I, et al. StayHome: A FHIR-native mobile COVID-19 symptom tracker and public health reporting tool. Online J Public Health Inform 2021;13(1):e2 [FREE Full text] [doi: 10.5210/ojphi.v13i1.11462] [Medline: 33936522]
64. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
65. Çelik Ertuğrul D, Çelik Ulusoy D. A knowledge-based self-pre-diagnosis system to predict Covid-19 in smartphone users using personal data and observed symptoms. Expert Syst 2022;39(3):e12716 [FREE Full text] [doi: 10.1111/exsy.12716] [Medline: 34177034]
66. Chan YY, Bot BM, Zweig M, Tignor N, Ma W, Suver C, et al. The asthma mobile health study, smartphone data collected using ResearchKit. Sci Data 2018;5:180096 [FREE Full text] [doi: 10.1038/sdata.2018.96] [Medline: 29786695]
67. Chang AR, Bailey-Davis L, Hetherington V, Ziegler A, Yule C, Kwiecen S, et al. Remote dietary counseling using smartphone applications in patients with stages 1-3a chronic kidney disease: a mixed methods feasibility study. J Ren Nutr 2020;30(1):53-60 [FREE Full text] [doi: 10.1053/j.jrn.2019.03.080] [Medline: 31078403]
68. Chen J, Bauman A, Allman-Farinelli M. A study to determine the most popular lifestyle smartphone applications and willingness of the public to share their personal data for health research. Telemed J E Health 2016;22(8):655-665. [doi: 10.1089/tmj.2015.0159] [Medline: 26958742]
69. Cheung C, Bietz MJ, Patrick K, Bloss CS. Privacy attitudes among early adopters of emerging health technologies. PLoS One 2016;11(11):e0166389 [FREE Full text] [doi: 10.1371/journal.pone.0166389] [Medline: 27832194]
70. Chung AE, Sandler RS, Long MD, Ahrens S, Burris JL, Martin CF, et al. Harnessing person-generated health data to accelerate patient-centered outcomes research: the crohn's and colitis foundation of America PCORnet patient powered research network (CCFA partners). J Am Med Inform Assoc 2016;23(3):485-490 [FREE Full text] [doi: 10.1093/jamia/ocv191] [Medline: 26911821]
71. Clarke H, Clark S, Birkin M, Iles-Smith H, Glaser A, Morris MA. Understanding barriers to novel data linkages: topic modeling of the results of the LifeInfo survey. J Med Internet Res 2021;23(5):e24236 [FREE Full text] [doi: 10.2196/24236] [Medline: 33998998]
72. Cooke Bailey JN, Crawford DC, Goldenberg A, Slaven A, Pencak J, Schachere M, et al. Willingness to participate in a national precision medicine cohort: attitudes of chronic kidney disease patients at a Cleveland public hospital. J Pers Med 2018;8(3):21 [FREE Full text] [doi: 10.3390/jpm8030021] [Medline: 29949895]
73. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
74. Deering S, Grade MM, Uppal JK, Foschini L, Juusola JL, Amdur AM, et al. Accelerating research with technology: rapid recruitment for a large-scale web-based sleep study. JMIR Res Protoc 2019;8(1):e10974 [FREE Full text] [doi: 10.2196/10974] [Medline: 30664491]
75. Deering S, Pratap A, Suver C, Borelli AJ, Amdur A, Headapohl W, et al. Real-world longitudinal data collected from the sleep health mobile app study. Sci Data 2020;7(1):418 [FREE Full text] [doi: 10.1038/s41597-020-00753-2] [Medline: 33247114]

XSL•FO
RenderX

76. Eicher-Miller HA, Prapkree L, Palacios C. Expanding the capabilities of nutrition research and health promotion through mobile-based applications. Adv Nutr 2021;12(3):1032-1041 [FREE Full text] [doi: 10.1093/advances/nmab022] [Medline: 33734305]
77. Evans BJ. The perils of parity: should citizen science and traditional research follow the same ethical and privacy principles? J Law Med Ethics 2020;48(1_suppl):74-81 [FREE Full text] [doi: 10.1177/1073110520917031] [Medline: 32342738]
78. Fadda M, Jobin A, Blasimme A, Greshake Tzovaras B, Price Ball M, Vayena E. User perspectives of a web-based data-sharing platform (open humans) on ethical oversight in participant-led research: protocol for a quantitative study. JMIR Res Protoc 2018;7(11):e10939 [FREE Full text] [doi: 10.2196/10939] [Medline: 30487120]
79. Fadrique LX, Rahman D, Vaillancourt H, Boissonneault P, Donovska T, Morita PP. Overview of policies, guidelines, and standards for active assisted living data exchange: thematic analysis. JMIR Mhealth Uhealth 2020;8(6):e15923 [FREE Full text] [doi: 10.2196/15923] [Medline: 32568090]
80. Faurholt-Jepsen M, Þórarinsdóttir H, Vinberg M, Ullum H, Frost M, Bardram J, et al. Automatically generated smartphone data and subjective stress in healthy individuals - a pilot study. Nord J Psychiatry 2020;74(4):293-300. [doi: 10.1080/08039488.2019.1705904] [Medline: 31880486]
81. Fish LA, Jones EJH. A survey on the attitudes of parents with young children on in-home monitoring technologies and study designs for infant research. PLoS One 2021;16(2):e0245793 [FREE Full text] [doi: 10.1371/journal.pone.0245793] [Medline: 33544777]
82. Franklin EF, Nichols HM, House L, Buzaglo J, Thiboldeaux K. Cancer patient perspectives on sharing of medical records and mobile device data for research purposes. J Patient Exp 2020;7(6):1115-1121 [FREE Full text] [doi: 10.1177/2374373520923837] [Medline: 33457553]
83. Freifeld CC, Chunara R, Mekaru SR, Chan EH, Kass-Hout T, Ayala Iacucci A, et al. Participatory epidemiology: use of mobile phones for community-based health reporting. PLoS Med 2010;7(12):e1000376 [FREE Full text] [doi: 10.1371/journal.pmed.1000376] [Medline: 21151888]
84. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. BMC Health Serv Res 2018;18(1):445 [FREE Full text] [doi: 10.1186/s12913-018-3254-z] [Medline: 29898730]
85. Genes N, Violante S, Cetrangol C, Rogers L, Schadt EE, Chan YY. From smartphone to EHR: a case report on integrating patient-generated health data. NPJ Digit Med 2018;1(1):23 [FREE Full text] [doi: 10.1038/s41746-018-0030-8] [Medline: 31304305]
86. Hafen E, Kossmann D, Brand A. Health data cooperatives – citizen empowerment. Methods Inf Med 2018;53(02):82-86. [doi: 10.3414/me13-02-0051]
87. Hamed A, Curran C, Gwaltney C, DasMahapatra P. Mobility assessment using wearable technology in patients with late-onset pompe disease. NPJ Digit Med 2019;2(1):70 [FREE Full text] [doi: 10.1038/s41746-019-0143-8] [Medline: 31341956]
88. Hartmann R, Sander C, Lorenz N, Böttger D, Hegerl U. Utilization of patient-generated data collected through mobile devices: insights from a survey on attitudes toward mobile self-monitoring and self-management apps for depression. JMIR Ment Health 2019;6(4):e11671 [FREE Full text] [doi: 10.2196/11671] [Medline: 30942693]
89. Heidel A, Hagist C. Potential benefits and risks resulting from the introduction of health apps and wearables into the German statutory health care system: scoping review. JMIR Mhealth Uhealth 2020;8(9):e16444 [FREE Full text] [doi: 10.2196/16444] [Medline: 32965231]
90. Heidel A, Hagist C, Schlereth C. Pricing through health apps generated data-digital dividend as a game changer: discrete choice experiment. PLoS One 2021;16(7):e0254786 [FREE Full text] [doi: 10.1371/journal.pone.0254786] [Medline: 34310618]
91. Henderson ML, Thomas AG, Eno AK, Waldram MM, Bannon J, Massie AB, et al. The impact of the mKidney mHealth system on live donor follow-up compliance: protocol for a randomized controlled trial. JMIR Res Protoc 2019;8(1):e11000 [FREE Full text] [doi: 10.2196/11000] [Medline: 30664485]
92. Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA. Consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies during the COVID-19 pandemic: development and usability study. JMIR Public Health Surveill 2021;7(4):e23806 [FREE Full text] [doi: 10.2196/23806] [Medline: 33843598]
93. Herbec A, Brown J, Shahab L, West R. Lessons learned from unsuccessful use of personal carbon monoxide monitors to remotely assess abstinence in a pragmatic trial of a smartphone stop smoking app - a secondary analysis. Addict Behav Rep 2019;9:100122 [FREE Full text] [doi: 10.1016/j.abrep.2018.07.003] [Medline: 31193683]
94. Hershman SG, Bot BM, Shcherbina A, Doerr M, Moayedi Y, Pavlovic A, et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart counts study. Sci Data 2019;6(1):24 [FREE Full text] [doi: 10.1038/s41597-019-0016-7] [Medline: 30975992]
95. Hesse BW, Greenberg AJ, Rutten LJF. The role of Internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol 2016;13(12):767-776. [doi: 10.1038/nrclinonc.2016.78] [Medline: 27273045]
96. Hong SJ, Cho H. Privacy management and health information sharing via contact tracing during the COVID-19 pandemic: a hypothetical study on AI-based technologies. Health Commun 2023;38(5):913-924. [doi: 10.1080/10410236.2021.1981565] [Medline: 34555998]

XSL•FO
RenderX

97. Househ M, Grainger R, Petersen C, Bamidis P, Merolli M. Balancing between privacy and patient needs for health information in the age of participatory health and social media: a scoping review. Yearb Med Inform 2018;27(1):29-36 [FREE Full text] [doi: 10.1055/s-0038-1641197] [Medline: 29681040]
98. Hughes A, Landers D, Arkenau H, Shah S, Stephens R, Mahal A, et al. Development and evaluation of a new technological way of engaging patients and enhancing understanding of drug tolerability in early clinical development: PROACT. Adv Ther 2016;33(6):1012-1024 [FREE Full text] [doi: 10.1007/s12325-016-0335-4] [Medline: 27167621]
99. Katapally TR. A global digital citizen science policy to tackle pandemics like COVID-19. J Med Internet Res 2020;22(5):e19357 [FREE Full text] [doi: 10.2196/19357] [Medline: 32408267]
100. Kim G, Bae JC, Yi BK, Hur KY, Chang DK, Lee M, et al. An information and communication technology-based centralized clinical trial to determine the efficacy and safety of insulin dose adjustment education based on a smartphone personal health record application: a randomized controlled trial. BMC Med Inform Decis Mak 2017;17(1):109 [FREE Full text] [doi: 10.1186/s12911-017-0507-4] [Medline: 28720103]
101. Kim JW, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
102. Kim TK, Choi M. Older adults' willingness to share their personal and health information when adopting healthcare technology and services. Int J Med Inform 2019;126:86-94. [doi: 10.1016/j.ijmedinf.2019.03.010] [Medline: 31029268]
103. Kolovson S, Pratap A, Duffy J, Allred R, Munson S, Areán PA. Understanding participant needs for engagement and attitudes towards passive sensing in remote digital health studies. Int Conf Pervasive Comput Technol Healthc 2020;2020:347-362 [FREE Full text] [doi: 10.1145/3421937.3422025] [Medline: 33717638]
104. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? Open data for healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
105. Labs J, Terry S. Privacy in the coronavirus era. Genet Test Mol Biomarkers 2020;24(9):535-536. [doi: 10.1089/gtmb.2020.29055.sjt] [Medline: 32677850]
106. Laurie GT. Cross-sectoral big data: the application of an ethics framework for big data in health and research. Asian Bioeth Rev 2019;11(3):327-339 [FREE Full text] [doi: 10.1007/s41649-019-00093-3] [Medline: 31632475]
107. Leal Neto O, Dimech GS, Libel M, de Souza WV, Cesse E, Smolinski M, et al. Saúde na copa: the world's first application of participatory surveillance for a mass gathering at FIFA world cup 2014, Brazil. JMIR Public Health Surveill 2017;3(2):e26 [FREE Full text] [doi: 10.2196/publichealth.7313] [Medline: 28473308]
108. Levin HI, Egger D, Andres L, Johnson M, Bearman SK, de Barbaro K. Sensing everyday activity: parent perceptions and feasibility. Infant Behav Dev 2021;62:101511. [doi: 10.1016/j.infbeh.2020.101511] [Medline: 33465730]
109. McKendry RA, Rees G, Cox IJ, Johnson A, Edelstein M, Eland A, et al. Share mobile and social-media data to curb COVID-19. Nature 2020;580(7801):29. [doi: 10.1038/d41586-020-00908-6] [Medline: 32235939]
110. Moore S, Tassé AM, Thorogood A, Winship I, Zawati M, Doerr M. Consent processes for mobile app mediated research: systematic review. JMIR Mhealth Uhealth 2017;5(8):e126 [FREE Full text] [doi: 10.2196/mhealth.7014] [Medline: 28855147]
111. Nakamoto I, Jiang M, Zhang J, Zhuang W, Guo Y, Jin M, et al. Evaluation of the design and implementation of a peer-to-peer COVID-19 contact tracing mobile app (COCOA) in Japan. JMIR Mhealth Uhealth 2020;8(12):e22098 [FREE Full text] [doi: 10.2196/22098] [Medline: 33170801]
112. O'Doherty KC, Christofides E, Yen J, Bentzen HB, Burke W, Hallowell N, et al. If you build it, they will come: unintended future uses of organised health data collections. BMC Med Ethics 2016;17(1):54 [FREE Full text] [doi: 10.1186/s12910-016-0137-x] [Medline: 27600117]
113. Poletti P, Visintainer R, Lepri B, Merler S. The interplay between individual social behavior and clinical symptoms in small clustered groups. BMC Infect Dis 2017;17(1):521 [FREE Full text] [doi: 10.1186/s12879-017-2623-2] [Medline: 28747154]
114. Poudyal A, van Heerden A, Hagaman A, Maharjan SM, Byanjankar P, Subba P, et al. Wearable digital sensors to identify risks of postpartum depression and personalize psychological treatment for adolescent mothers: protocol for a mixed methods exploratory study in rural Nepal. JMIR Res Protoc 2019;8(8):e14734 [FREE Full text] [doi: 10.2196/14734] [Medline: 31512581]
115. Quer G, Gadaleta M, Radin JM, Andersen KG, Baca-Motes K, Ramos E, et al. The physiologic response to COVID-19 vaccination. medRxiv 2021 [FREE Full text] [doi: 10.1101/2021.05.03.21256482] [Medline: 33972954]
116. Radin JM, Peters S, Ariniello L, Wongvibulsin S, Galarnyk M, Waalen J, et al. Pregnancy health in POWERMOM participants living in rural versus urban zip codes. J Clin Transl Sci 2020;4(5):457-462 [FREE Full text] [doi: 10.1017/cts.2020.33] [Medline: 33244436]
117. Rake EA, van Gelder MHJ, Grim DC, Heeren B, Engelen L, van de Belt TH. Personalized consent flow in contemporary data sharing for medical research: a viewpoint. Biomed Res Int 2017;2017:7147212 [FREE Full text] [doi: 10.1155/2017/7147212] [Medline: 28638834]
118. Rendina HJ, Mustanski B. Privacy, trust, and data sharing in web-based and mobile research: participant perspectives in a large nationwide sample of men who have sex with men in the United States. J Med Internet Res 2018;20(7):e233 [FREE Full text] [doi: 10.2196/jmir.9019] [Medline: 29973332]

XSL•FO
RenderX

119. Rieger A, Gaines A, Barnett I, Baldassano CF, Connolly Gibbons MB, Crits-Christoph P. Psychiatry outpatients' willingness to share social media posts and smartphone data for research and clinical purposes: survey study. JMIR Form Res 2019;3(3):e14329 [FREE Full text] [doi: 10.2196/14329] [Medline: 31493326]
120. Rohlman D, Dixon HM, Kincl L, Larkin A, Evoy R, Barton M, et al. Development of an environmental health tool linking chemical exposures, physical location and lung function. BMC Public Health 2019;19(1):854 [FREE Full text] [doi: 10.1186/s12889-019-7217-z] [Medline: 31262274]
121. Rothstein MA, Wilbanks JT, Brothers KB. Citizen science on your smartphone: an ELSI research agenda. J Law Med Ethics 2015;43(4):897-903. [doi: 10.1111/jlme.12327] [Medline: 26711425]
122. Ságvári B, Gulyás A, Koltai J. Attitudes towards participation in a passive data collection experiment. Sensors 2021;21(18):6085. [doi: 10.3390/s21186085]
123. Salamone F, Masullo M, Sibilio S. Wearable devices for environmental monitoring in the built environment: a systematic review. Sensors (Basel) 2021;21(14):4727 [FREE Full text] [doi: 10.3390/s21144727] [Medline: 34300467]
124. Saleheen N, Chakraborty S, Ali N, Mahbubur Rahman MD, Hossain S, Bari R, et al. mSieve: differential behavioral privacy in time series of mobile sensor data. Proc ACM Int Conf Ubiquitous Comput 2016;2016:706-717 [FREE Full text] [doi: 10.1145/2971648.2971753] [Medline: 28058408]
125. Santos-Lozano A, Baladrón C, Martín-Hernández J, Morales JS, Ruilope L, Lucia A. mHealth and the legacy of John Snow. The Lancet 2018;391(10129):1479-1480. [doi: 10.1016/s0140-6736(18)30783-9]
126. Schmitz H, Howe CL, Armstrong DG, Subbian V. Leveraging mobile health applications for biomedical research and citizen science: a scoping review. J Am Med Inform Assoc 2018;25(12):1685-1695 [FREE Full text] [doi: 10.1093/jamia/ocy130] [Medline: 30445467]
127. Seltzer E, Goldshear J, Guntuku SC, Grande D, Asch DA, Klinger EV, et al. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. BMC Med Inform Decis Mak 2019;19(1):157 [FREE Full text] [doi: 10.1186/s12911-019-0886-9] [Medline: 31395102]
128. Simblett SK, Biondi A, Bruno E, Ballard D, Stoneman A, Lees S, RADAR-CNS consortium. Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: a qualitative analysis. Epilepsy Behav 2020;102:106717. [doi: 10.1016/j.yebeh.2019.106717] [Medline: 31785481]
129. Sleigh J. Experiences of donating personal data to mental health research: an explorative anthropological study. Biomed Inform Insights 2018;10:1178222618785131 [FREE Full text] [doi: 10.1177/1178222618785131] [Medline: 30013355]
130. Slotwiner DJ, Tarakji KG, Al-Khatib SM, Passman RS, Saxon LA, Peters NS, et al. Transparent sharing of digital health data: a call to action. Heart Rhythm 2019;16(9):e95-e106 [FREE Full text] [doi: 10.1016/j.hrthm.2019.04.042] [Medline: 31077802]
131. Smith RJ, Grande D, Merchant RM. Transforming scientific inquiry. Academic Medicine 2016;91(4):469-472. [doi: 10.1097/acm.0000000000001022]
132. Staccini P, Fernandez-Luque L. Secondary use of recorded or self-expressed personal data: consumer health informatics and education in the era of social media and health apps. Yearb Med Inform 2017;26(01):172-177. [doi: 10.15265/iy-2017-037]
133. Struminskaya B, Toepoel V, Lugtig P, Haan M, Luiten A, Schouten B. Understanding willingness to share smartphone-sensor data. Public Opin Q 2020;84(3):725-759 [FREE Full text] [doi: 10.1093/poq/nfaa044] [Medline: 34025296]
134. Tajiri E, Yoshimura E, Hatamoto Y, Tanaka H, Shimoda S. Effect of sleep curtailment on dietary behavior and physical activity: a randomized crossover trial. Physiol Behav 2018;184:60-67. [doi: 10.1016/j.physbeh.2017.11.008] [Medline: 29129611]
135. Tannis C, Senerat A, Garg M, Peters D, Rajupet S, Garland E. Improving physical activity among residents of affordable housing: is active design enough? Int J Environ Res Public Health 2019;16(1):151 [FREE Full text] [doi: 10.3390/ijerph16010151] [Medline: 30626013]
136. Taylor SA, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11(2):200-213 [FREE Full text] [doi: 10.1109/TAFFC.2017.2784832] [Medline: 32489521]
137. van Heerden A, Wassenaar D, Essack Z, Vilakazi K, Kohrt BA. In-home passive sensor data collection and its implications for social media research: perspectives of community women in rural South Africa. J Empir Res Hum Res Ethics 2020;15(1-2):97-107. [doi: 10.1177/1556264619881334] [Medline: 31631742]
138. Van Oeveren BT, De Ruiter CJ, Hoozemans MJM, Beek PJ, Van Dieën JH. Inter-individual differences in stride frequencies during running obtained from wearable data. J Sports Sci 2019;37(17):1996-2006. [doi: 10.1080/02640414.2019.1614137] [Medline: 31079578]
139. Vitak J, Zimmer M. More than just privacy: using contextual integrity to evaluate the long-term risks from COVID-19 surveillance technologies. Soc Media Soc 2020;6(3):2056305120948250 [FREE Full text] [doi: 10.1177/2056305120948250] [Medline: 34192036]
140. Vo JDV, Gorbach AM. A platform to record patient events during physiological monitoring with wearable sensors: proof-of-concept study. Interact J Med Res 2019;8(1):e10336 [FREE Full text] [doi: 10.2196/10336] [Medline: 30609977]

XSL•FO
RenderX

141. von Gablenz P, Kowalk U, Bitzer J, Meis M, Holube I. Individual hearing aid benefit in real life evaluated using ecological momentary assessment. Trends Hear 2021;25:2331216521990288 [FREE Full text] [doi: 10.1177/2331216521990288] [Medline: 33787404]
142. Wanyua S, Ndemwa M, Goto K, Tanaka J, K'opiyo J, Okumu S, et al. Profile: the Mbita health and demographic surveillance system. Int J Epidemiol 2013;42(6):1678-1685. [doi: 10.1093/ije/dyt180] [Medline: 24415606]
143. Wasfi R, Poirier Stephens Z, Sones M, Laberee K, Pugh C, Fuller D, et al. Recruiting participants for population health intervention research: effectiveness and costs of recruitment methods for a cohort study. J Med Internet Res 2021;23(11):e21142 [FREE Full text] [doi: 10.2196/21142] [Medline: 34587586]
144. Webster DE, Suver C, Doerr M, Mounts E, Domenico L, Petrie T, et al. The mole mapper study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. Sci Data 2017;4(1):170005 [FREE Full text] [doi: 10.1038/sdata.2017.5] [Medline: 28195576]
145. Welsh JB, Derdzinski M, Parker AS, Puhr S, Jimenez A, Walker T. Real-time sharing and following of continuous glucose monitoring data in youth. Diabetes Ther 2019;10(2):751-755 [FREE Full text] [doi: 10.1007/s13300-019-0571-0] [Medline: 30701468]
146. Wicks P. Patient, study thyself. BMC Med 2018;16(1):217 [FREE Full text] [doi: 10.1186/s12916-018-1216-2] [Medline: 30470218]
147. Witteveen D, de Pedraza P. The roles of general health and COVID-19 proximity in contact tracing app usage: cross-sectional survey study. JMIR Public Health Surveill 2021;7(8):e27892 [FREE Full text] [doi: 10.2196/27892] [Medline: 34081602]
148. Woldaregay AZ, Henriksen A, Issom DZ, Pfuhl G, Sato K, Richard A, et al. User expectations and willingness to share self-collected health data. Stud Health Technol Inform 2020;270:894-898. [doi: 10.3233/SHTI200290] [Medline: 32570511]
149. Yan K, Tracie B, Marie-Ève M, Mélanie H, Jean-Luc B, Benoit T, et al. Innovation through wearable sensors to collect real-life data among pediatric patients with cardiometabolic risk factors. Int J Pediatr 2014;2014:328076 [FREE Full text] [doi: 10.1155/2014/328076] [Medline: 24678323]
150. Cho S, Ensari I, Weng C, Kahn MG, Natarajan K. Factors affecting the quality of person-generated wearable device data and associated challenges: rapid systematic review. JMIR Mhealth Uhealth 2021;9(3):e20738 [FREE Full text] [doi: 10.2196/20738] [Medline: 33739294]
151. Cho S. Data quality assessment for the secondary use of person-generated wearable device data. In: Assessing Self-Tracking Data for Research Purposes. New York: Columbia University; 2021.
152. Song S, Taylor CO. Data representativeness in cardiovascular disease studies that use consumer wearables. In: The 1st International Workshop on Ethics and Bias of Artificial Intelligence in Clinical Applications (EBAIC 2023). 2023 Presented at: 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI); 2023 June 26; Houston, Texas. [doi: 10.1109/ichi57859.2023.00124]
153. Kim J, Im E, Kim H. From intention to action: the factors affecting health data sharing intention and action. Int J Med Inform 2023;175:105071. [doi: 10.1016/j.ijmedinf.2023.105071] [Medline: 37099875]
154. Patel MS, Polsky D, Kennedy EH, Small DS, Evans CN, Rareshide CAL, et al. Smartphones vs wearable devices for remotely monitoring physical activity after hospital discharge: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2020;3(2):e1920677 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20677] [Medline: 32031643]
155. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2(1):45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
156. Philip BJ, Abdelrazek M, Bonti A, Barnett S, Grundy J. Data collection mechanisms in health and wellness apps: review and analysis. JMIR Mhealth Uhealth 2022;10(3):e30468 [FREE Full text] [doi: 10.2196/30468] [Medline: 35262499]
157. Keusch F, Struminskaya B, Antoun C, Couper M, Kreuter F. Willingness to participate in passive mobile data collection. Public Opin Q 2019;83(Suppl 1):210-235 [FREE Full text] [doi: 10.1093/poq/nfz007] [Medline: 31337924]
158. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
159. Dhruva SS, Ross JS, Akar JG, Caldwell B, Childers K, Chow W, et al. Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform. NPJ Digit Med 2020;3(1):60 [FREE Full text] [doi: 10.1038/s41746-020-0265-z] [Medline: 32352038]
160. CARIN consumer directed payer data exchange (CARIN IG for blue button®). HL7 International. 2021. URL: https://build.fhir.org/ig/HL7/carin-bb/ [accessed 2024-10-18]
161. Blue Button® 2.0: improving medicare beneficiary access to their health information. Centers for Medicare & Medicaid Services. 2021. URL: https://web.archive.org/web/20231014191517/https://www.cms.gov/data-research/cms-information-technology/blue-button [accessed 2023-10-14]
162. CMS Fast Facts. Baltimore, MD: Centers for Medicare & Medicaid Services; 2023.
163. Gasteiger N, Vercell A, Davies A, Dowding D, Khan N, Davies A. Patient-facing genetic and genomic mobile apps in the UK: a systematic review of content, functionality, and quality. J Community Genet 2022;13(2):171-182 [FREE Full text] [doi: 10.1007/s12687-022-00579-y] [Medline: 35182377]

XSL•FO
RenderX

164. Talwar D, Yeh YL, Chen WJ, Chen LS. Characteristics and quality of genetics and genomics mobile apps: a systematic review. Eur J Hum Genet 2019;27(6):833-840 [FREE Full text] [doi: 10.1038/s41431-019-0360-2] [Medline: 30809045]
165. Shaw RJ, Steinberg DM, Bonnet J, Modarai F, George A, Cunningham T, et al. Mobile health devices: will patients actually use them? J Am Med Inform Assoc 2016;23(3):462-466 [FREE Full text] [doi: 10.1093/jamia/ocv186] [Medline: 26911820]
166. L'Hommedieu M, L'Hommedieu J, Begay C, Schenone A, Dimitropoulou L, Margolin G, et al. Lessons learned: recommendations for implementing a longitudinal study using wearable and environmental sensors in a health care organization. JMIR Mhealth Uhealth 2019;7(12):e13305 [FREE Full text] [doi: 10.2196/13305] [Medline: 31821155]
167. Köngeter A, Schickhardt C, Jungkunz M, Bergbold S, Mehlis K, Winkler EC. Patients' willingness to provide their clinical data for research purposes and acceptance of different consent models: findings from a representative survey of patients with cancer. J Med Internet Res 2022;24(8):e37665 [FREE Full text] [doi: 10.2196/37665] [Medline: 36006690]
168. Ploug T, Holm S. Meta consent - a flexible solution to the problem of secondary use of health data. Bioethics 2016;30(9):721-732 [FREE Full text] [doi: 10.1111/bioe.12286] [Medline: 27628305]
169. Sheehan M, Thompson R, Fistein J, Davies J, Dunn M, Parker M, et al. Authority and the future of consent in population-level biomedical research. Public Health Ethics 2019;12(3):225-236 [FREE Full text] [doi: 10.1093/phe/phz015] [Medline: 32082417]
170. Abdelazeem B, Abbas KS, Amin MA, El-Shahat NA, Malik B, Kalantary A, et al. The effectiveness of incentives for research participation: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2022;17(4):e0267534 [FREE Full text] [doi: 10.1371/journal.pone.0267534] [Medline: 35452488]
171. Bentley JP, Thacker PG. The influence of risk and monetary payment on the research participation decision making process. J Med Ethics 2004;30(3):293-298 [FREE Full text] [doi: 10.1136/jme.2002.001594] [Medline: 15173366]
172. Altman DG, Simera I. Responsible reporting of health research studies: transparent, complete, accurate and timely. J Antimicrob Chemother 2010;65(1):1-3 [FREE Full text] [doi: 10.1093/jac/dkp410] [Medline: 19900949]
173. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 2012;1(1):60 [FREE Full text] [doi: 10.1186/2046-4053-1-60] [Medline: 23194585]
174. Vilaró M, Cortés J, Selva-O'Callaghan A, Urrutia A, Ribera J, Cardellach F, et al. Adherence to reporting guidelines increases the number of citations: the argument for including a methodologist in the editorial process and peer-review. BMC Med Res Methodol 2019;19(1):112 [FREE Full text] [doi: 10.1186/s12874-019-0746-4] [Medline: 31151417]
175. Guu TW, Muurling M, Khan Z, Kalafatis C, Aarsland D, ffytche D, et al. Wearable devices: underrepresentation in the ageing society. The Lancet Digital Health 2023;5(6):e336-e337. [doi: 10.1016/s2589-7500(23)00069-9]

Abbreviations

BYOD: Bring-Your-Own-Device
FHIR: Fast Health Interoperability Resources
mERA: Mobile Health Evidence Reporting and Assessment
MeSH: medical subject headings
PGHD: person-generated health data
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping reviews
STaRT-RWE: Structured Template and Reporting Tool for Real-World Evidence

Edited by A Mavragani; submitted 17.08.23; peer-reviewed by T Katapally, E Baker; comments to author 02.02.24; revised version received 12.04.24; accepted 27.09.24; published 20.01.25.

Please cite as:
Song S, Ashton M, Yoo RH, Lkhagvajav Z, Wright R, Mathews DJH, Taylor CO
Participant Contributions to Person-Generated Health Data Research Using Mobile Devices: Scoping Review
J Med Internet Res 2025;27:e51955
URL: https://www.jmir.org/2025/1/e51955
doi:10.2196/51955
PMID:

©Shanshan Song, Micaela Ashton, Rebecca Hahn Yoo, Zoljargal Lkhagvajav, Robert Wright, Debra J H Mathews, Casey Overby Taylor. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the

XSL•FO
RenderX

Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Completeness of Telehealth Interventions Reporting in Randomized Controlled Trials for Caregivers of People With Dementia: Systematic Review

Ling Zhu[1], MSN; Yurong Xing[1], MSN; Wenhui Xu[1], MSN; Hongfei Jia[1], MSN; Xiaoxiao Wang[1], MSN; Shiqing Liu[2,3], BSN; Yaping Ding[1], MSN

[1]Department of Basic and Community Nursing, School of Nursing, Nanjing Medical University, NanJing, China

[2]School of Nursing, Nanjing Medical University, NanJing, China

[3]Nursing Department, JiangSu Province Official Hospital, NanJing, China

Corresponding Author:
Yaping Ding, MSN
Department of Basic and Community Nursing
School of Nursing
Nanjing Medical University
101 Longmian Avenue
Jiangning District
NanJing, 211166
China
Phone: 86 13851646554
Email: dingyp@njmu.edu.cn

Abstract

Background: Telehealth interventions can effectively support caregivers of people with dementia by providing care and improving their health outcomes. However, to successfully translate research into clinical practice, the content and details of the interventions must be sufficiently reported in published papers.

Objective: This study aims to evaluate the completeness of a telehealth intervention reporting in randomized controlled trials (RCTs) conducted for caregivers of people with dementia.

Methods: A systematic search of relevant papers was conducted on July 26, 2023, in 9 electronic databases. RCTs of telehealth interventions for caregivers of people with dementia were included. Two independent researchers extracted the descriptive information and assessed the methodological quality (Cochrane risk of bias tool) and the completeness of reporting of the intervention by using the Template for Intervention Description and Replication (TIDieR)-Telehealth checklist, which consists of 12 items.

Results: Thirty-eight eligible RCTs were included finally, and the overall quality of the studies was assessed as moderate. None of the studies completely reported all the TIDieR-Telehealth items. The most frequently reported items were the brief trial name (35/38, 92%), rationale (38/38, 100%), materials and procedures (35/38, 92%), and the modes of delivery (34/38, 90%). The least reported items were the type of location (0/38, 0%), modifications (4/38, 11%), and assessment and improvement of fidelity (9/38, 24%).

Conclusions: Many details of the telehealth interventions in RCTs are reported incompletely. Greater adherence to the TIDieR-Telehealth checklist is essential for improving the reporting quality and for facilitating replicability, which has substantial implications for translation into clinical practice.

(J Med Internet Res 2025;27:e53737) doi:10.2196/53737

KEYWORDS

telehealth; intervention reporting; dementia; caregivers; Template for Intervention Description and Replication; TIDieR-Telehealth checklist

XSL•FO
RenderX

Introduction

According to the World Alzheimer Report 2023, over 55 million people live with dementia worldwide, and this number is expected to reach 139 million by 2050 [1]. Alzheimer disease is ranked as the seventh leading cause of mortality and poses the highest burden of disease globally [2], and dementia care has become the focus of global health services [3]. To perform activities of daily living, most people with dementia require care from informal caregivers such as family members and friends or from formal caregivers [4]. Studies have shown that caregivers for people with dementia may experience emotional stress that probably results in adverse effects on the patients [5,6]. Reports indicate that these caregivers frequently experience symptoms such as depression, anxiety, and stress [7]. Currently, there are many interventions for supporting caregivers for dementia care. With the development of the society and the progress of technology, telehealth as an alternative medium for delivering health care has boomed.

Telehealth is defined as the use of medical information exchanged via electronic communication from one location to another [8]. Various types of telehealth interventions have been widely used in medical care, which have been shown to improve the quality of care and provide better outcomes for patients [9,10]. Caregivers of people with dementia require favorable interventions because of the complexity of dementia care, and telehealth can increase the opportunities for caregivers to receive interventions. Studies have shown that telehealth interventions have benefits for caregivers, such as improving their perceived competency, reducing their burden, and relieving their depression and stress [11,12].

Randomized controlled trials (RCTs) are currently recognized as the gold standard for evaluating clinical treatments and for generating high-quality evidence for the effectiveness and efficacy of interventions [13]. One challenge identified in the reporting of clinical trials is the quality of the descriptions of interventions [14]. The reproducibility of RCT results is fundamental to improving evidence-based care and patient outcomes, but insufficiently published descriptions of interventions will affect the reproducibility of trials, raising concerns about the validity and reliability of the findings [15].

Although the CONSORT (Consolidated Standards of Reporting Trials) statement offers general recommendations for intervention reporting, it lacks detailed guidance on critical intervention components such as the particular techniques and modes of delivery [12]. To address this gap, the Template for Intervention Description and Replication (TIDieR) checklist was published in 2014 to emphasize the importance of adequate reporting of interventions in clinical trials, enabling detailed descriptions of the intervention content for the reproducibility of trials [15,16]. In 2022, the TIDieR-Telehealth checklist, as an extension of the original TIDieR checklist, provided additional guidance specific to reporting of telehealth interventions with the goal of improving reporting quality and maximizing reproducibility and implementation in clinical trials [17].

Previous studies using the TIDieR checklist for assessment have shown incomplete intervention reporting [15,16,18], raising concerns about the comprehensiveness of telehealth intervention reports for dementia caregivers. To our knowledge, no study has yet examined the quality of reporting of interventions within dementia care interventions or identified which components are the most underreported. Therefore, the purpose of this study was to assess the completeness of telehealth intervention reporting in RCTs for caregivers of people with dementia.

Methods

Study Design

This review assesses the completeness of telehealth intervention reporting in RCTs for caregivers of people with dementia by using the TIDieR-Telehealth checklist.

Search Strategy

Related papers were retrieved from PubMed, Embase, Cochrane Library, Web of Science, CINAHL, China National Knowledge Infrastructure, Chinese Biomedical Literature database, Wanfang database, and VIP Chinese Science and Technology Periodicals Full-Text database, without any restriction on the date of publication (from inception to July 26, 2023). Only RCTs published in English or Chinese were included. The search strategy can be found in Multimedia Appendix 1.

Inclusion Criteria

The criteria for the inclusion of the published papers were as follows: (1) participants were caregivers providing care for people with dementia at any stage (formal or informal caregivers); (2) telehealth interventions involved any treatments that were delivered remotely via text messaging, videoconferencing, audio-only communication, mobile apps, and other telecommunication tools [17]; (3) the control group received usual care, standard care, waiting list, face-to-face, or any other nontelehealth interventions; (4) all eligible studies were analyzed for completeness of reporting of the intervention, regardless of the outcomes; and (5) only RCTs.

Study Selection

Two independent researchers, systematically trained in evidence-based research, screened the titles and abstracts of the selected papers by using the inclusion criteria to identify potentially eligible studies. Full texts were obtained and again screened for final inclusion. Disagreements were resolved by a third researcher.

Data Extraction

Two independent researchers extracted information regarding the characteristics, methodological quality, and completeness of reporting of the intervention. A third researcher resolved disagreements arising at any stage. Regarding the study characteristics, we extracted the following data: first author, year of publication, sample size, age and gender of included participants, severity of dementia, intervention characteristics, outcomes and measurement tools, assessment times, risk of bias, and completeness of reporting.

XSL•FO
RenderX

Risk of Bias Assessment

The Cochrane Risk of Bias tool [19] was used to assess the methodological quality of the eligible trials on random sequence generation, allocation concealment, blinding of participants and personnel, blinding outcome assessment, incomplete outcome data, selective reporting, and other biases. Each item was rated as low risk, high risk, or unclear risk. In addition, the overall risk of bias of the study was classified as A (low risk in all domains), B (low risk in some domains), or C (no low-risk domains) [20]. Due to the high risk of bias and low quality of grade C studies, those studies were not included [20-22].

Completeness of Reporting Assessment

The completeness of reporting was extracted using the TIDieR-Telehealth checklist, which has 12 items such as brief name, why, what (materials and procedures), who provided, how, where, when, and how much, tailoring, modifications, and how well (planned, actual). Considering that there are multiple aspects included in the items, each item was scored on a 3-point scale: 1 if all the criteria were met, 0.5 if some criteria were included, and 0 if no criteria were met or mentioned. The full score of each study was 12. It should be noted that item 10 was considered adequately reported if there was explicit mention of modifications during the study and was considered not adequately reported if no explicit mention of modifications was made during the study.

Data Analysis

We performed a descriptive analysis to describe the included trials, the methodological quality, and the reporting of the interventions. Overall scores were calculated on the completeness of reporting for each study, and numbers and percentages were used to summarize the adherence to each item of the checklist for all the included trials. Additionally, we used a histogram to represent the number of trials with different scores on each item. The methodological quality of the eligible trials was shown by a risk-of-bias summary graph created in Review Manager version 5.4.

Results

Study Selection

A total of 5133 results were obtained after the preliminary screening. After excluding 1893 duplicate studies and 3155 irrelevant studies that did not meet the criteria (as shown by their titles and abstracts), 38 studies were included after full-text review (Multimedia Appendix 2). The literature screening process and results are shown in Figure 1.

Figure 1. Flow diagram for the screening of studies. CBM: Chinese Biomedical literature; CNKI: China National Knowledge Infrastructure.

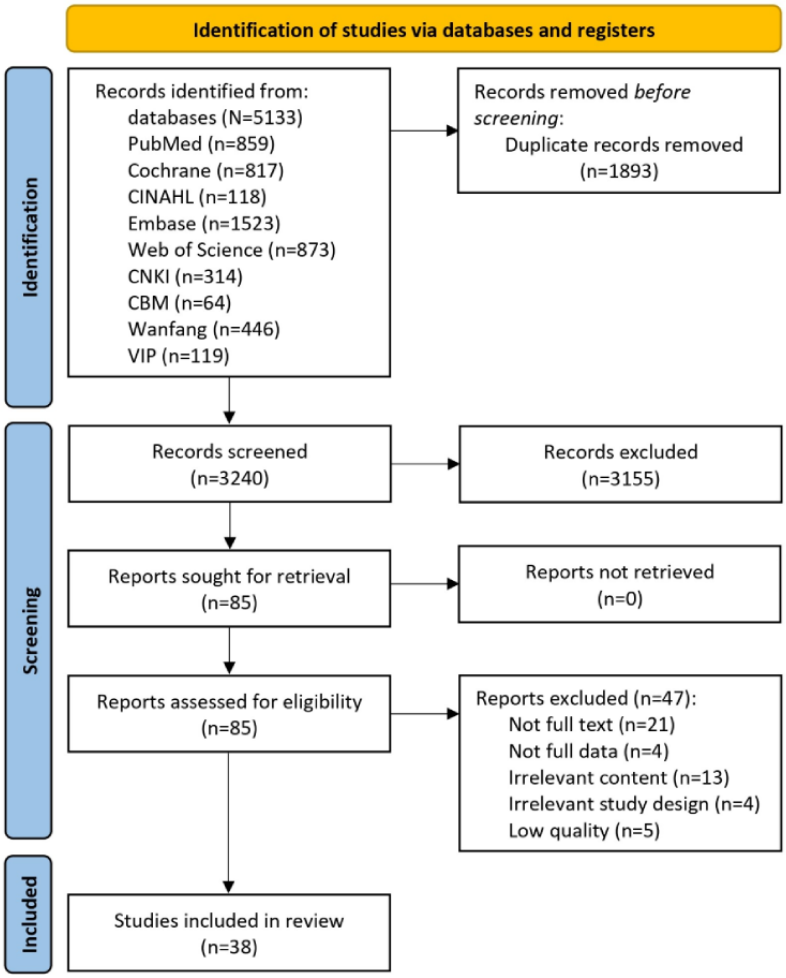

XSL•FO
RenderX

Description of the Included Studies

The characteristics of the eligible RCTs are described in Multimedia Appendix 3 [23-60]. Considering the recent publication of the telehealth checklist, our study mainly conducted a retrospective evaluation of the completeness of reporting in the relevant papers. All 38 papers included in this review were published from 2003 to 2023 in 12 countries, namely, the United States (n=16) [32-40,43-45,48,55,56,58], Germany (n=6) [25,30,46,47,54,57], China (n=4) [23,42,52,59], Canada (n=2) [31,60], Spain (n=2) [49,50], Netherlands (n=2) [26,27], Belgium (n=1) [51], France (n=1) [28], Greece (n=1) [41], Portugal (n=1) [53], India (n=1) [24], and Italy (n=1) [29]. The total sample size was 4001 (range 14-273), with an average age range of 37-72 years. Seven studies [25,29,32,33,35,41,55] did not specify the proportion of females, while the other studies [23,24,26-28,30,31,34,36-40,42-54,56-60] included 2676 (79.4%) females. All the trial interventions were delivered by the internet or telephone, with 15 studies [23,25,29,30,35,42,44,45,47,51,54,55,57-59] using the telephone to provide education or support to caregivers. Only 17 of the 23 studies that conducted interventions through the internet mentioned specific modalities of the intervention, of which 13 [24,27,28,31,36,37,40,41,49,50,52,53,60] provided caregivers with knowledge, skills, and support through a web page or app, and 3 [32,33,43] studies provided support via a computer-telephone integration system. One study [34] provided a monitoring system for caregivers.

Methodological Quality

The risks of bias for the included studies is shown in Figure 2 [23-60]. According to the risk of bias assessed by the Cochrane handbook, 38 papers were qualified, with 3 studies graded as A [30,31,53] and 35 as B. In general, the methodological quality of these RCTs was moderate. Specifically, 23 studies [24-28,30,31,34,37,40,42,43,45-49,51,53,55,57,59,60] described the generation of random sequences. Allocation concealment was not reported in 15 papers [23,27,29,32,33,36,38,39,41,44,46,49,50,54,58], and 22 papers [23,25,27,29,32-36,38-41,44-46,49,50,54,55,58,59] provided no information on the blinding of participants and personnel. As data were often collected online, the detection bias in 16 papers [26,30,31,33-35,37,38,40,42,43,49,51,53,59,60] was determined to be low risk. Researchers in 21 studies [23,26-28,30-32,34,44,47-51,53-57,59,60] reported completeness of data. In addition, reporting bias was uncertain in most papers (n=26) [23,24,26,27,29,32,33,35-39,41-46,48,50,51,54,55,58-60]. The overall quality of the included studies was moderate, and further high-quality studies are needed.

XSL•FO
RenderX

Figure 2. Risk of bias graph.

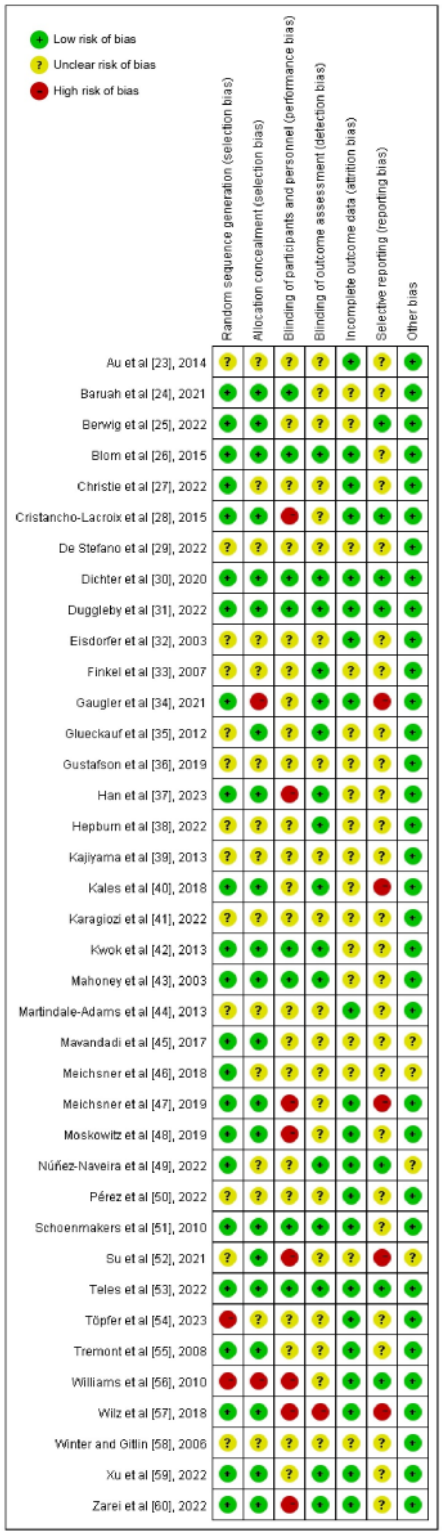

Completeness of Reporting

The item scores for each trial are shown in Table 1 [23-60]. None of the papers described all the items on the TIDieR-Telehealth checklist. The completeness of the reporting score varied between 5 and 11.5, with a mean of 7.7. Figure 3 shows the distribution of studies with different scores in each item of the TIDieR-Telehealth checklist. Nearly all trials (n=35) reported the name, the use of physical or informational materials, and the procedures of the interventions (item 1, item 3, item 4). All trials described a rationale (item 2). More than half (n=22) of the included studies [23,25,26,29,30,35,37,38,41,42,45-47,50-52,54,55,57-60] mentioned intervention providers (item 5), and 68% (15/22) of the studies [26,29,30,35,37,45-47,50,51,54,55,57-59] provided detailed information of the providers, including their expertise, background, and any specific training given. All trials described the mode of delivery (item 6), and 4 papers [26,39,53,54] did

XSL•FO
RenderX

not mention whether the interventions were provided individually or in a group. Only 7 studies [29,32,34,35,45,56,57] mentioned the types of locations where the intervention occurred (item 7). The majority of the studies (37/38, 98%) mentioned the number of times the intervention was delivered, while fewer than half (17/37) [23,25,28-30,32,35,37,38,41,42,44,50,55-57,60] reported the schedule and duration (item 8). Nearly half (n=18) of the studies [23,32,33,35,38,40-42,44-46,49-53,55,57] reported the tailoring of the interventions (item 9). The least reported item was item 10, with only 4 papers [25,28,35,48] mentioning modification of the intervention. Item 11, that is, strategies to improve fidelity/adherence, was reported in 25 (66%) of the 38 studies [23-27,31,32,34-38,40,45-48,51-55,57,59,60]. Only 13 (34%) studies [24,28,31,33,35-37,44-46,54,57,59] provided a complete description of the actual fidelity of the intervention delivery (item 12).

Figure 3. Distribution of studies with different scores in each item of the Template for Intervention Description and Replication-Telehealth checklist. Numbers in bars represent the number of interventions rated as 1 (adequate reporting), 0.5 (partial reporting), and 0 (no reporting). TIDieR-Telehealth: Template for Intervention Description and Replication-Telehealth.

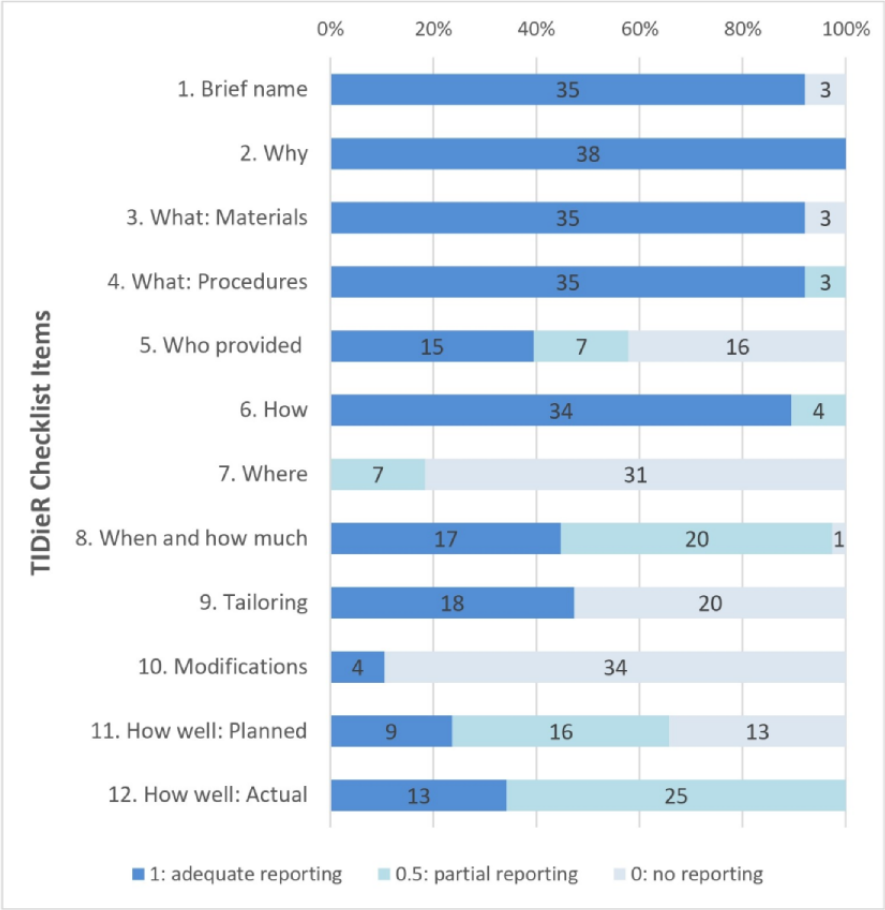

XSL•FO
RenderX

Table 1. Template for Intervention Description and Replication-Telehealth scores [23-60].

Author	Template for Intervention Description and Replication-Telehealth items												Final score
	1 (n=35)	2 (n=38)	3 (n=35)	4 (n=36.5)	5 (n=18.5)	6 (n=36)	7 (n=3.5)	8 (n=27)	9 (n=18)	10 (n=4)	11 (n=17)	12 (n=25.5)	
Au et al [23]	1	1	1	1	0.5	1	0	1	1	0	0.5	0.5	8.5
Baruah et al [24]	1	1	1	1	0	1	0	0.5	0	0	0.5	1	7
Berwig et al [25]	1	1	1	1	0.5	1	0	1	0	1	0.5	0.5	8.5
Blom et al [26]	1	1	1	1	1	0.5	0	0.5	0	0	1	0.5	7.5
Christie et al [27]	1	1	1	1	0	1	0	0.5	0	0	1	0.5	7
Cristancho-Lacroix et al [28]	1	1	1	1	0	1	0	1	0	1	0	1	8
De Stefano et al [29]	1	1	0	1	1	1	0.5	1	0	0	0	0.5	7
Dichter et al [30]	1	1	1	1	1	1	0	1	0	0	0	0.5	7.5
Duggleby et al [31]	1	1	1	1	0	1	0	0.5	0	0	1	1	7.5
Eisdorfer et al [32]	1	1	1	1	0	1	0.5	1	1	0	0.5	0.5	8.5
Finkel et al [33]	1	1	1	1	0	1	0	0.5	1	0	0	1	7.5
Gaugler et al [34]	1	1	1	1	0	1	0.5	0.5	0	0	0.5	0.5	7
Glueckauf et al [35]	1	1	1	1	1	1	0.5	1	1	1	1	1	11.5
Gustafson et al [36]	1	1	1	1	0	1	0	0.5	0	0	0.5	1	7
Han et al [37]	1	1	1	1	1	1	0	1	0	0	0.5	1	8.5
Hepburn et al [38]	1	1	1	1	0.5	1	0	1	1	0	0.5	0.5	8.5
Kajiyama et al [39]	1	1	1	1	0	0.5	0	0.5	0	0	0	0.5	5.5
Kales et al [40]	0	1	1	1	0	1	0	0.5	1	0	0.5	0.5	6.5
Karagiozi et al [41]	1	1	1	1	0.5	1	0	1	1	0	0	0.5	8
Kwok et al [42]	1	1	1	1	0.5	1	0	1	1	0	0	0.5	8
Mahoney et al [43]	1	1	1	1	0	1	0	0.5	0	0	0	0.5	6
Martin-dale-Adams et al [44]	1	1	1	1	0	1	0	1	1	0	0	1	8
Mavandadi et al [45]	1	1	1	1	1	1	0.5	0.5	1	0	0.5	1	9.5

XSL•FO
RenderX

Author	Template for Intervention Description and Replication-Telehealth items												Final score
	1 (n=35)	2 (n=38)	3 (n=35)	4 (n=36.5)	5 (n=18.5)	6 (n=36)	7 (n=3.5)	8 (n=27)	9 (n=18)	10 (n=4)	11 (n=17)	12 (n=25.5)	
Meichsner et al [46]	1	1	1	1	1	1	0	0.5	1	0	1	1	9.5
Meichsner et al [47]	1	1	1	1	1	1	0	0.5	0	0	0.5	0.5	7.5
Moskowitz et al [48]	1	1	1	1	0	1	0	0.5	0	1	0.5	0.5	7.5
Núñez-Naveira et al [49]	1	1	1	1	0	1	0	0.5	1	0	0	0.5	7
Pérez et al [50]	0	1	1	1	1	1	0	1	1	0	0	0.5	7.5
Schoenmakers et al [51]	0	1	0	0.5	1	1	0	0.5	1	0	0.5	0.5	6
Su et al [52]	1	1	1	1	0.5	1	0	0.5	1	0	0.5	0.5	8
Teles et al [53]	1	1	1	1	0	0.5	0	0.5	1	0	1	1	8
Töpfer et al [54]	1	1	1	0.5	1	0.5	0	0.5	0	0	0.5	0.5	6.5
Tremont et al [55]	1	1	1	1	1	1	0	1	1	0	1	0.5	9.5
Williams et al [56]	1	1	1	1	0	1	0.5	1	0	0	0	0.5	7
Wilz et al [57]	1	1	1	1	1	1	0.5	1	1	0	1	1	10.5
Winter and Gitlin [58]	1	1	0	0.5	1	1	0	0	0	0	0	0.5	5
Xu et al [59]	1	1	1	1	1	1	0	0.5	0	0	1	1	8.5
Zarei et al [60]	1	1	1	1	0.5	1	0	1	0	0	0.5	0.5	7.5

Discussion

Principal Findings

This review is a retrospective study evaluating the completeness of reporting on telehealth interventions for caregivers of people with dementia. The principal finding was that none of the telehealth interventions were reported in enough detail to satisfy every TIDieR-Telehealth checklist item, despite the checklist being published in 2022. This finding confirms the need for the introduction of the TIDieR-Telehealth checklist and suggests that there is a large shortfall in the reporting of information needed to accurately replicate telehealth interventions. In particular, the least frequently reported items were the ones referring to the type of location of the interventions (fully reported by no study), as well as the modifications (fully reported by 4 studies [25,28,35,48]), and fidelity assessment during the study (fully reported by 9 studies [26,27,31,35,46,53,55,57,59]). Importantly, these elements may be fundamental for enabling interventions to be adequately replicated, compared, and transferable in clinical practice [16].

Similar to our findings, a previous systematic review showed that the reporting of telehealth-delivered dietary interventions in chronic disease is inadequate, although this was assessed using the TIDieR checklist [61]. Given that the TIDieR-Telehealth checklist was published in 2022, there are currently no studies describing the completeness of telehealth interventions. However, some studies evaluated using the TIDieR checklist also indicated inadequate completeness of reporting across trials of weight management, cardiac rehabilitation, and manipulation and mobilization techniques [15,16,18]. Compared to those studies, the proportion of papers with complete descriptions of the intervention in our study was different. It might be that telehealth intervention had its unique evaluation standard: the TIDieR-Telehealth checklist. Another explanation might be that our interpretation of the

XSL•FO
RenderX

TIDieR-Telehealth checklist and guide produced a more stringent evaluation than previous studies of interventions reporting quality, and we adopted a more detailed way of keeping the score.

Unlike other studies [18,62], the names of the included trials were not fully reported as required, because the TIDieR-Telehealth checklist requires that the name of the trial should include the word "telehealth," which differs from the requirements in the TIDieR checklist. Most studies (>88%) reported intervention materials, procedures, and modes of delivery, which were similar to the findings of previous studies [15,18]. Information on intervention providers, such as category, experience, background, and any specific training given, was fully described in only 15 studies [26,29,30,35,37,45-47,50,51,54,55,57-59]. These descriptions are important for telehealth interventions because the delivery of telehealth interventions requires training to ensure the normativity of the study, and the expertise and disciplinary background of the provider might affect the outcome of a trial [18]. None of the trials fully described the location of the intervention, which may be related to the characteristic that telehealth is not limited by the location. However, the authors of the TIDieR checklist and guide emphasized that the location(s) of a trial might impact intervention feasibility and adherence. Nearly half of the studies fully reported the duration and number of interventions. Some interventions using remote monitoring systems or telehealth support programs were designed for participants to use when needed; therefore, the length and the number of interventions are not described. However, the arrangement of these interventions has a great influence on the effect of the intervention; therefore, the evaluation of the fidelity of intervention is particularly important. Tailoring, as an important part of the intervention, was mentioned in 18 trials [23,32,33,35,38,40-42,44-46,49-53,55,57]. The application parameters of technology are no longer predetermined but are constantly adjusted due to the feedback of the patients and the tailoring of the therapist to the responses of the individual patient [63]. Frequent shortcomings were items related to the reporting of modifications and the fidelity of interventions. In this study, only 4 trials (11%) [25,28,35,48] mentioned modifications in the interventions, which are essential for refining the intervention content and enhancing the replicability and the validity of the trial. Moreover, only 9 studies [26,27,31,35,46,53,55,57,59] fully described the methods for assessing and improving the fidelity of intervention. The assessment of fidelity remains a challenge in intervention trials, particularly in telehealth, where maintaining consistency can be difficult. Given that fidelity substantially affects treatment outcomes, this is an important factor to consider when interpreting study results [14,17].

Limitation and Strengths

Different from other studies [15,61], our research employs a 3-point method, accounting for both fully and partially reported items. From these statistics, we can identify the missing parts in a report and urge that study to improve the completeness of that report. Moreover, one limitation of our study is that we did not compare the completeness of a report before and after the publication of the TIDieR-Telehealth checklist, as this checklist has been published only recently and related studies were limited. More trials are needed to confirm the usefulness of this checklist.

Conclusion

We found the reporting of interventions according to the TIDieR-Telehealth checklist to be inadequate in our sample of related trials. Fundamental details were often not described, affecting research validity and reproducibility. Future RCTs on telehealth interventions for dementia should adhere to standardized reporting. It is necessary to consider other mechanisms to improve TIDieR adherence or to find other solutions to improve intervention reporting. Meanwhile, telehealth interventions, as an intervention method in the new era, have brought many benefits to patients and caregivers but also posed challenges to intervention providers. Therefore, the next step in the implementation of telehealth interventions is how to properly utilize the benefits of telehealth to maximize its value.

Acknowledgments

We thank all the members of our study team for their cooperation and the original authors of the included studies for their work. This work was supported by the Project of Nursing Science funded by the 4th Priority Discipline Development Program of Jiangsu Higher Education Institutions (Jiangsu Education Department 2023, No 11) and the Connotation Construction Project of Nanjing Medical University for Priority Academic of Nursing Science (2022-12).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 24 KB - jmir_v27i1e53737_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

XSL•FO
RenderX

[DOCX File , 31 KB - jmir_v27i1e53737_app2.docx]

Multimedia Appendix 3
Characteristics of the eligible randomized controlled trials [26-63].
[DOCX File , 29 KB - jmir_v27i1e53737_app3.docx]

References

1. Long S, Benoist C, Weidner W. World Alzheimer Report 2023. London, England: Alzheimer Disease International; 2023 Sep 21. URL: https://www.alzint.org/resource/world-alzheimer-report-2023/ [accessed 2025-01-03]
2. Alzheimer's Association. 2024 Alzheimer's disease facts and figures. Alzheimers Dement 2024 May 30;20(5):3708-3821 [FREE Full text] [doi: 10.1002/alz.13809] [Medline: 38689398]
3. Zhu XN, Chen S, He M, Dong YY, Fang SY, Atigu YM, et al. Life experience and identity of spousal caregivers of people with dementia: A qualitative systematic review. Int J Nurs Stud 2024 Jun;154:104757. [doi: 10.1016/j.ijnurstu.2024.104757] [Medline: 38552470]
4. Isik AT, Soysal P, Solmi M, Veronese N. Bidirectional relationship between caregiver burden and neuropsychiatric symptoms in patients with Alzheimer's disease: A narrative review. Int J Geriatr Psychiatry 2019 Sep;34(9):1326-1334. [doi: 10.1002/gps.4965] [Medline: 30198597]
5. Meng XF, Su JP, Li HH, Ma DF, Zhao YN, Li YJ, et al. Effectiveness of caregiver non-pharmacological interventions for behavioural and psychological symptoms of dementia: An updated meta-analysis. Ageing Res Rev 2021 Nov;71:101448. [doi: 10.1016/j.arr.2021.101448] [Medline: 34416379]
6. França AB, Gordon AL, Samra R, Rodolpho Duarte ESR, Jacinto AF. Symptoms of mood disorders in family carers of older people with dementia who experience caregiver burden: a network approach. Age Ageing 2020 Jul 01;49(4):628-633. [doi: 10.1093/ageing/afaa008] [Medline: 32091573]
7. Reuben DB, Romero T, Evertson LC, Jennings LA. Overwhelmed: a dementia caregiver vital sign. J Gen Intern Med 2022 Aug;37(10):2469-2474 [FREE Full text] [doi: 10.1007/s11606-021-07054-3] [Medline: 34389938]
8. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/nejmsr1503323]
9. Lu JF, Chen CM, Hsu CY. Effect of home telehealth care on blood pressure control: A public healthcare centre model. J Telemed Telecare 2019 Jan;25(1):35-45. [doi: 10.1177/1357633x17734258]
10. Li JJ, Liu Y, Jiang JJ, Peng XC, Hu XL. Effect of telehealth interventions on quality of life in cancer survivors: A systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021 Oct;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]
11. Zhu AY, Cao WT, Zhou YH, Xie AA, Cheng Y, Chu SF. Telehealth intervention for carers of dementia patients-a systematic review and meta-analysis of randomized controlled trials. Front Aging Neurosci 2021 Feb 10;13:612404 [FREE Full text] [doi: 10.3389/fnagi.2021.612404] [Medline: 33643022]
12. Saragih ID, Tonapa SI, Porta CM, Lee BO. Effects of telehealth intervention for people with dementia and their carers: A systematic review and meta-analysis of randomized controlled studies. J Nurs Scholarsh 2022 Nov;54(6):704-719. [doi: 10.1111/jnu.12797] [Medline: 35769007]
13. Cuschieri S. The CONSORT statement. Saudi J Anaesth 2019 Apr;13(Suppl 1):S27-S30 [FREE Full text] [doi: 10.4103/sja.SJA_559_18] [Medline: 30930716]
14. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
15. Pool J, Maissan F, de Waele N, Wittink H, Ostelo R. Completeness of the description of manipulation and mobilisation techniques in randomized controlled trials in neck pain; A review using the TiDieR checklist. Musculoskelet Sci Pract 2020 Mar;45:102098. [doi: 10.1016/j.msksp.2019.102098] [Medline: 32056823]
16. Ben-Porat T, Yousefi R, Stojanovic J, Boucher VG, Fortin A, Lavoie K, et al. Completeness of intervention reporting of nutrition-focused weight management interventions adjunct to metabolic and bariatric surgery: effect of the TIDieR checklist. Int J Obes (Lond) 2024 Jul;48(7):901-912. [doi: 10.1038/s41366-024-01506-6] [Medline: 38459257]
17. Rhon DI, Fritz JM, Kerns RD, McGeary DD, Coleman BC, Farrokhi S, et al. TIDieR-telehealth: precision in reporting of telehealth interventions used in clinical trials - unique considerations for the Template for the Intervention Description and Replication (TIDieR) checklist. BMC Med Res Methodol 2022 Jun 02;22(1):161 [FREE Full text] [doi: 10.1186/s12874-022-01640-7] [Medline: 35655144]
18. McGrath A, McHale S, Hanson CL, McLelland C, Hamilton DF. Completeness of intervention reporting in randomised trials of technology-enabled remote or hybrid exercise-based cardiac rehabilitation: a systematic review using the TIDieR framework. Disabil Rehabil 2024 Sep;46(19):4350-4358. [doi: 10.1080/09638288.2023.2274887] [Medline: 37899659]

XSL•FO
RenderX

19. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
20. Guo G, Lan MJ, Zeng F, Liang JSY, Guo LY, Cai LY, et al. Effects of remote interventions with digital health technologies in lung transplant patients: a meta-analysis. Chin J Nurs. 2024 Jun 10. URL: https://zh.zhhlzzs.com/EN/10.3761/j.issn.0254-1769.2024.11.016 [accessed 2025-01-03]
21. Linares-Espinós E, Hernández V, Domínguez-Escrig J, Fernández-Pello S, Hevia V, Mayor J, et al. Methodology of a systematic review. Actas Urológicas Españolas (English Edition) 2018 Oct;42(8):499-506. [doi: 10.1016/j.acuroe.2018.07.002]
22. Bindels BJJ, Mercier C, Gal R, Verlaan J, Verhoeff JJC, Dirix P, et al. Stereotactic body and conventional radiotherapy for painful bone metastases: a systematic review and meta-analysis. JAMA Netw Open 2024 Mar 05;7(2):e2355409 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.55409] [Medline: 38345820]
23. Au A, Wong MK, Leung LM, Leung P, Wong A. Telephone-assisted pleasant-event scheduling to enhance well-being of caregivers of people with dementia: a randomised controlled trial. Hong Kong Med J 2014 Jun;20(3 Suppl 3):30-33 [FREE Full text] [Medline: 25001033]
24. Baruah U, Varghese M, Loganathan S, Mehta KM, Gallagher-Thompson D, Zandi D, et al. Feasibility and preliminary effectiveness of an online training and support program for caregivers of people with dementia in India: a randomized controlled trial. Int J Geriatr Psychiatry 2021 Apr;36(4):606-617. [doi: 10.1002/gps.5502] [Medline: 33491811]
25. Berwig M, Lessing S, Deck R. Telephone-based aftercare groups for family carers of people with dementia - results of the effect evaluation of a randomised controlled trial. BMC Health Serv Res 2022 Feb 11;22(1):177 [FREE Full text] [doi: 10.1186/s12913-022-07490-9] [Medline: 35144607]
26. Blom MM, Zarit SH, Groot Zwaaftink RBM, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
27. Christie HL, Dam AEH, van Boxtel M, Köhler S, Verhey F, de Vugt ME. Lessons learned from an effectiveness evaluation of inlife, a web-based social support intervention for caregivers of people with dementia: randomized controlled trial. JMIR Aging 2022 Dec 07;5(4):e38656 [FREE Full text] [doi: 10.2196/38656] [Medline: 36476485]
28. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud AS. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015 May 12;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
29. De Stefano M, Esposito S, Iavarone A, Carpinelli Mazzi M, Siciliano M, Buonanno D, et al. Effects of phone-based psychological intervention on caregivers of patients with early-onset Alzheimer's disease: a six-months study during the COVID-19 emergency in Italy. Brain Sci 2022 Feb 24;12(3):310 [FREE Full text] [doi: 10.3390/brainsci12030310] [Medline: 35326267]
30. Dichter MN, Albers B, Trutschel D, Ströbel AM, Seismann-Petersen S, Wermke K, et al. TALKING TIME: A pilot randomized controlled trial investigating social support for informal caregivers via the telephone. BMC Health Serv Res 2020 Aug 25;20(1):788 [FREE Full text] [doi: 10.1186/s12913-020-05523-9] [Medline: 32838773]
31. Duggleby W, Ploeg J, McAiney C, Peacock S, Fisher K, Ghosh S, et al. Web-based intervention for family carers of persons with dementia and multiple chronic conditions (My Tools 4 Care): pragmatic randomized controlled trial. J Med Internet Res 2018 Jun 29;20(6):e10484 [FREE Full text] [doi: 10.2196/10484] [Medline: 29959111]
32. Eisdorfer C, Czaja SJ, Loewenstein DA, Rubert MP, Argüelles S, Mitrani VB, et al. The effect of a family therapy and technology-based intervention on caregiver depression. Gerontologist 2003 Aug;43(4):521-531 [FREE Full text] [doi: 10.1093/geront/43.4.521] [Medline: 12937331]
33. Finkel S, Czaja SJ, Schulz R, Martinovich Z, Harris C, Pezzuto D. E-care: a telecommunications technology intervention for family caregivers of dementia patients. Am J Geriatr Psychiatry 2007 May;15(5):443-448. [doi: 10.1097/JGP.0b013e3180437d87] [Medline: 17463195]
34. Gaugler JE, Zmora R, Mitchell LL, Finlay J, Rosebush CE, Nkimbeng M, et al. Remote activity monitoring for family caregivers of persons living with dementia: a mixed methods, randomized controlled evaluation. BMC Geriatr 2021 Dec 18;21(1):715 [FREE Full text] [doi: 10.1186/s12877-021-02634-8] [Medline: 34922475]
35. Glueckauf RL, Davis WS, Willis F, Sharma D, Gustafson DJ, Hayes J, et al. Telephone-based, cognitive-behavioral therapy for African American dementia caregivers with depression: initial findings. Rehabil Psychol 2012 May;57(2):124-139. [doi: 10.1037/a0028688] [Medline: 22686551]
36. Gustafson DH, Gustafson DH, Cody OJ, Chih M, Johnston DC, Asthana S. Pilot test of a computer-based system to help family caregivers of dementia patients. J Alzheimers Dis 2019;70(2):541-552. [doi: 10.3233/jad-190052]
37. Han A, Yuen HK, Jenkins J. The feasibility and preliminary effects of a pilot randomized controlled trial: Videoconferencing acceptance and commitment therapy in distressed family caregivers of people with dementia. J Health Psychol 2023 May;28(6):554-567 [FREE Full text] [doi: 10.1177/13591053221141131] [Medline: 36591636]
38. Hepburn K, Nocera J, Higgins M, Epps F, Brewster GS, Lindauer A, et al. Results of a randomized trial testing the efficacy of tele-savvy, an online synchronous/asynchronous psychoeducation program for family caregivers of persons living with dementia. Gerontologist 2022 Apr 20;62(4):616-628 [FREE Full text] [doi: 10.1093/geront/gnab029] [Medline: 33640979]

XSL•FO
RenderX

39. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
40. Kales HC, Gitlin LN, Stanislawski B, Myra Kim H, Marx K, Turnwald M, et al. Effect of the WeCareAdvisor™ on family caregiver outcomes in dementia: a pilot randomized controlled trial. BMC Geriatr 2018 May 10;18(1):113 [FREE Full text] [doi: 10.1186/s12877-018-0801-8] [Medline: 29747583]
41. Karagiozi K, Margaritidou P, Tsatali M, Marina M, Dimitriou T, Apostolidis H, et al. Comparison of on site versus online psycho education groups and reducing caregiver burden. Clin Gerontol 2022;45(5):1330-1340. [doi: 10.1080/07317115.2021.1940409] [Medline: 34219617]
42. Kwok T, Wong B, Ip I, Chui K, Young D, Ho F. Telephone-delivered psychoeducational intervention for Hong Kong Chinese dementia caregivers: a single-blinded randomized controlled trial. Clin Interv Aging 2013 Aug;8:1191-1197. [doi: 10.2147/cia.s48264]
43. Mahoney DF, Tarlow BJ, Jones RN. Effects of an automated telephone support system on caregiver burden and anxiety: findings from the REACH for TLC intervention study. Gerontologist 2003 Aug;43(4):556-567. [doi: 10.1093/geront/43.4.556] [Medline: 12937334]
44. Martindale-Adams J, Nichols LO, Burns R, Graney MJ, Zuber J. A trial of dementia caregiver telephone support. Can J Nurs Res 2013 Dec;45(4):30-48. [doi: 10.1177/084456211304500404] [Medline: 24617278]
45. Mavandadi S, Wright EM, Graydon MM, Oslin DW, Wray LO. A randomized pilot trial of a telephone-based collaborative care management program for caregivers of individuals with dementia. Psychol Serv 2017 Feb;14(1):102-111. [doi: 10.1037/ser0000118] [Medline: 28134558]
46. Meichsner F, Theurer C, Wilz G. Acceptance and treatment effects of an internet-delivered cognitive-behavioral intervention for family caregivers of people with dementia: A randomized-controlled trial. J Clin Psychol 2019 Apr;75(4):594-613. [doi: 10.1002/jclp.22739] [Medline: 30597537]
47. Meichsner F, Töpfer NF, Reder M, Soellner R, Wilz G. Telephone-based cognitive behavioral intervention improves dementia caregivers' quality of life. Am J Alzheimers Dis Other Demen 2019 Jun;34(4):236-246 [FREE Full text] [doi: 10.1177/1533317518822100] [Medline: 30636429]
48. Moskowitz JT, Cheung EO, Snowberg KE, Verstaen A, Merrilees J, Salsman JM, et al. Randomized controlled trial of a facilitated online positive emotion regulation intervention for dementia caregivers. Health Psychol 2019 May;38(5):391-402 [FREE Full text] [doi: 10.1037/hea0000680] [Medline: 31045422]
49. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT platform to help informal caregivers of people with dementia: a pilot randomized controlled study. Biomed Res Int 2016;2016:1-13 [FREE Full text] [doi: 10.1155/2016/5726465] [Medline: 28116300]
50. Pérez V, Menéndez-Crispín EJ, Sarabia-Cobo C, de Lorena P, Fernández-Rodríguez A, González-Vaca J. Mindfulness-based intervention for the reduction of compassion fatigue and burnout in nurse caregivers of institutionalized older persons with dementia: a randomized controlled trial. Int J Environ Res Public Health 2022 Sep 11;19(18):11441 [FREE Full text] [doi: 10.3390/ijerph191811441] [Medline: 36141714]
51. Schoenmakers B, Buntinx F, Delepeleire J. Supporting family carers of community-dwelling elder with cognitive decline: a randomized controlled trial. Int J Family Med 2010;2010:184152-184152 [FREE Full text] [doi: 10.1155/2010/184152] [Medline: 22332005]
52. Su HF, Koo M, Lee WL, Sung HC, Lee RP, Liu WI. A dementia care training using mobile e-learning with mentoring support for home care workers: a controlled study. BMC Geriatr 2021 Feb 16;21(1):126 [FREE Full text] [doi: 10.1186/s12877-021-02075-3] [Medline: 33593287]
53. Teles S, Ferreira A, Paúl C. Feasibility of an online training and support program for dementia carers: results from a mixed-methods pilot randomized controlled trial. BMC Geriatr 2022 Mar 01;22(1):173 [FREE Full text] [doi: 10.1186/s12877-022-02831-z] [Medline: 35232389]
54. Töpfer NF, Wrede N, Theurer C, Wilz G. Face-to-face versus telephone-based cognitive-behavioral therapy for family caregivers of people with dementia. J Clin Psychol 2023 Oct;79(10):2270-2287. [doi: 10.1002/jclp.23538] [Medline: 37222452]
55. Tremont G, Davis JD, Bishop DS, Fortinsky RH. Telephone-delivered psychosocial intervention reduces burden in dementia caregivers. Dementia (London) 2008;7(4):503-520 [FREE Full text] [doi: 10.1177/1471301208096632] [Medline: 20228893]
56. Williams VP, Bishop-Fitzpatrick L, Lane JD, Gwyther LP, Ballard EL, Vendittelli AP, et al. Video-based coping skills to reduce health risk and improve psychological and physical well-being in Alzheimer's disease family caregivers. Psychosom Med 2010 Nov;72(9):897-904. [doi: 10.1097/psy.0b013e3181fc2d09]
57. Wilz G, Reder M, Meichsner F, Soellner R. The Tele.TAnDem intervention: Telephone-based CBT for family caregivers of people with dementia. Gerontologist 2018 Mar 19;58(2):e118-e129. [doi: 10.1093/geront/gnx183] [Medline: 29190357]
58. Winter L, Gitlin LN. Evaluation of a telephone-based support group intervention for female caregivers of community-dwelling individuals with dementia. Am J Alzheimers Dis Other Demen 2006;21(6):391-397 [FREE Full text] [doi: 10.1177/1533317506291371] [Medline: 17267370]

XSL•FO
RenderX

59. Xu XY, Kwan RYC, Miao J, Chai A, Leung AYM. Telephone-based behavioral activation for improving sleep quality in family caregivers of people with dementia: a pilot randomized controlled trial. Behav Ther 2022 Sep;53(5):887-899. [doi: 10.1016/j.beth.2022.02.007] [Medline: 35987546]
60. Zarei S, Lakhanpal G, Sadavoy J. Tele-mindfulness for dementia's family caregivers: a randomized trial with a usual care control group. Curr Alzheimer Res 2022;19(5):364-372. [doi: 10.2174/1567205019666220514131015] [Medline: 35570513]
61. Warner MM, Kelly JT, Reidlinger DP, Hoffmann TC, Campbell KL. Reporting of telehealth-delivered dietary intervention trials in chronic disease: systematic review. J Med Internet Res 2017 Dec 11;19(12):e410 [FREE Full text] [doi: 10.2196/jmir.8193] [Medline: 29229588]
62. Ishaq I, Skinner IW, Mehta P, Verhagen AP. Description of massage interventions in randomised clinical trials for neck pain; a review using the TIDieR checklist. Clin Rehabil 2024 Mar;38(3):375-392. [doi: 10.1177/02692155231210377] [Medline: 37908084]
63. Tuttle N, Hazle C. An empirical, pragmatic approach applying reflection in interaction approach to manual therapy treatments. Physiother Theory Pract 2021 Jul;37(7):775-786. [doi: 10.1080/09593985.2019.1650405] [Medline: 31385748]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
RCT: randomized controlled trial
TIDieR: Template for Intervention Description and Replication
TIDieR-Telehealth: Template for Intervention Description and Replication-Telehealth

Edited by T de Azevedo Cardoso; submitted 18.10.23; peer-reviewed by Y Zhao, PW Stubbs, D Neal; comments to author 16.05.24; revised version received 10.07.24; accepted 25.11.24; published 20.01.25.

Please cite as:
Zhu L, Xing Y, Xu W, Jia H, Wang X, Liu S, Ding Y
Completeness of Telehealth Interventions Reporting in Randomized Controlled Trials for Caregivers of People With Dementia: Systematic Review
J Med Internet Res 2025;27:e53737
URL: https://www.jmir.org/2025/1/e53737
doi:10.2196/53737
PMID:

©Ling Zhu, Yurong Xing, Wenhui Xu, Hongfei Jia, Xiaoxiao Wang, Shiqing Liu, Yaping Ding. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Parental Technoference and Child Problematic Media Use: Meta-Analysis

Jinghui Zhang[1], MA; Qing Zhang[1], BEd; Bowen Xiao[2], PhD; Yuxuan Cao[1], BEd; Yu Chen[1], BEd; Yan Li[1], Prof Dr

[1]Shanghai Institute of Early Childhood Education, Shanghai Normal University, Shanghai, China

[2]Psychology Department, Carleton University, Ottawa, ON, Canada

Corresponding Author:
Yan Li, Prof Dr
Shanghai Institute of Early Childhood Education
Shanghai Normal University
100 Guilin Road, Xuhui District
Shanghai, 200234
China
Phone: 86 18939892301
Email: liyan@shnu.edu.cn

Abstract

Background: Parental technoference, the interruption of parent-child interactions by technology, has been associated with negative outcomes in children's media use. However, the magnitude of this relationship and its moderating factors remain unclear.

Objective: This study aims to systematically examine the relationship between parental technoference and child problematic media use, as well as to identify moderating factors such as age, parental technoference group, study design, and type of problematic media use.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, a comprehensive literature search was conducted up to August 2024 across multiple databases, including Web of Science, EBSCO, ProQuest, PubMed, PsycINFO, and China National Knowledge Infrastructure, using predefined search strings. A total of 53 studies with a total of 60,555 participants (mean age of 13.84, SD 1.18 years) were included. Inclusion criteria comprised studies involving children under the age of 22 years, assessing the association between parental technoference and child problematic media use with valid measures, and reporting necessary statistical data. Exclusion criteria included studies focusing on other child outcomes, having sample sizes <30, or being case reports or review papers. The risk of bias was assessed using the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. A random-effects meta-analysis was performed using R (version 4.2.1; R Foundation for Statistical Computing) with the *meta* and *metafor* packages to evaluate the association and conduct moderator analyses.

Results: The meta-analysis identified a significant positive association between parental technoference and child problematic media use (r=0.296, 95% CI 0.259-0.331). Moderator analyses revealed that both parental technoference group (P<.001) and study design (P=.008) significantly influenced this relationship. Specifically, the association was stronger when both parents engaged in technoference compared to when only 1 parent did, and in cross-sectional studies compared to longitudinal studies. Age, gender, publication status, and type of problematic media use did not significantly moderate the relationship (all P>.05).

Conclusions: This meta-analysis provides robust evidence of the association between parental technoference and child problematic media use. The findings highlight the need for family-based interventions and underscore the importance of longitudinal research to understand the temporal dynamics of this relationship better.

Trial Registration: PROSPERO CRD42023471997; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=471997

(J Med Internet Res 2025;27:e57636) doi:10.2196/57636

KEYWORDS

parental technoference; child; problematic media use; meta-analysis; children; youth; adolescents; phubbing; screen distraction; systematic reviews; interventions; mental health; digital health

XSL•FO
RenderX

Introduction

Background

Problematic media use in children refers to excessive screen engagement that disrupts functioning and development, transcending mere time metrics [1,2]. Globally, problematic media use is on the rise [3,4], with significant implications for the developmental health of children and adolescents [5-10]. A key factor implicated in the rise of problematic media use among youth is parental technoference. Technoference, a term coined by combining "technology" and "interference," refers to the disruption or disturbance of daily interpersonal communication and shared time caused by digital and mobile technology devices [11,12]. Parental technoference, which refers to the interruptions in parent-child interactions due to digital and electronic media devices, is a concern that has been increasingly recognized within the domain of family dynamics [13-15]. Numerous empirical studies have explored the relationship between parental technoference and child problematic media use, yet the findings have been inconsistent, possibly due to variations in research methodologies and sample characteristics [2,16-18]. Despite the growing body of literature, there is a conspicuous absence of systematic reviews or meta-analysis studies aimed at synthesizing these findings to address the discrepancies and draw reliable conclusions. To fill this gap, this study used a meta-analysis method to quantitatively integrate the extant research on this topic. Our objectives are twofold: to establish a more robust and reliable estimate of the effect size of parental technoference on child problematic media use, and to identify potential moderators that may influence the strength of this relationship. Through this work, we aim to contribute valuable insights into the dynamics of family media use and to promote healthier media habits that support the well-being of children and adolescents.

Association Between Parental Technoference and Child Problematic Media Use

Parental technoference may affect child problematic media use, and this association can be explained through various theories and psychological processes. First, the interactional theory of childhood problematic media use [19] posits that problematic media use in children is influenced by both individual and contextual factors. It suggests that children's media use patterns are shaped by their interactions with their environment, including parental behaviors. Specifically, parental technoference may lead to unclear media use rules in the family environment and reduce interaction and communication between parents and children [11,20-22]. In such cases, children may be more inclined to engage in excessive media use to compensate for the reduced parent-child interaction caused by parental technoference. Indeed, research has found that poor parent-child relationships are a significant risk factor for children's problematic media use [23,24]. Second, based on the accept-rejection theory [25], children's behavior is influenced by their perception of acceptance or rejection from significant others, especially their parents. When parents are preoccupied with technology and fail to provide emotional availability and responsiveness, children may perceive this as rejection or neglect. In response, children may seek alternative sources of comfort and stimulation, such as excessive media use, to compensate for the lack of emotional connection with their parents. Thus, parental technoference can contribute to problematic media use as a coping mechanism for perceived parental rejection. Finally, social learning theory [26] emphasizes the role of observational learning and modeling in shaping behavior. According to this theory, children learn by observing and imitating the behaviors of others, particularly their parents. When parents engage in excessive media use or prioritize technology over other activities, children are more likely to adopt similar behaviors. Parental technoference, by modeling problematic media use, may contribute to the development of problematic media use in children through observational learning processes. In conclusion, parental technoference can influence child problematic media use through multiple theoretical perspectives. It disrupts parent-child interactions, may contribute to children's perception of rejection, and models problematic media use.

Extensive empirical research has investigated the relationship between parental technoference and child problematic media use, including smartphone addiction [18,27-32], internet gaming disorder [33-37], social networking site addiction [2,38], and short-form video addiction [39]. The findings consistently indicate a significant positive correlation between parental technoference and child problematic media use, raising concerns for children across different age groups, from children with an average age of 10.33 (SD 0.98) years [37] to adolescents with an average age below 20 years [39-41]. However, there are fair-sized discrepancies in the magnitude of association between the same parental technoference and child problematic media use in the literature. For example, the correlation between parental technoference and child problematic media use ranged from 0.11 [17] to 0.54 [16]. These conflicts may be due to differences in study design, study characteristics, and result characteristics.

Impact of Moderator Variables

Gender differences in children's susceptibility to parental technoference and its association with problematic media use have been extensively documented in the literature. For instance, Xie et al [31] found that gender moderates the relationship between parental technoference and smartphone addiction in children, with a greater impact on smartphone addiction observed in boys compared to girls. However, Wang et al [42] did not find gender differences in the relationship between parental technoference and child problematic smartphone use. This discrepancy may be due to the uncertain nature of gender differences in problematic media use. Some studies have reported a significant correlation between gender and children's problematic media use. Among these studies, some have indicated a more severe tendency toward problematic media use in boys compared to girls [36], while others have found the opposite [35]. Additionally, some studies have not established a significant relationship between parental technoference and child problematic media use [2,29,43]. Overall, the impact of parental technoference on child problematic media use may differ between boys and girls, although a definitive conclusion

XSL•FO
RenderX

is currently lacking. Therefore, in this study, we examined the moderating effect of gender.

In addition, the age of children may also play a moderating role in the relationship between parental technoference and child problematic media use. On one hand, parental technoference may be related to the age of the child. Previous studies have found a significant positive correlation [2], a significant negative correlation [37], and a nonsignificant correlation [29,35,36] between parental technoference and child age. This variation may stem from the fact that each empirical study has only investigated a limited age range, failing to examine the relationship between parental technoference and child age from a broader perspective. On the other hand, the age of children may also influence their problematic media use. For example, researchers have found a significant positive correlation between child age and problematic media use [28,31,44], indicating that as children get older, their problematic media use becomes more severe. In summary, both parental technoference and child problematic media use vary with the child's age. Therefore, in this study, we also examined whether child age moderates the association between parental technoference and child problematic media use.

Moreover, this study aims to explore the separate effects of parental technoference on problematic media use in children. It is well documented that fathers and mothers often have different parenting roles and practices, which can influence children's development in unique ways [45,46]. For example, McDaniel and Radesky [15] observed a positive association between mother technoference in parenting and children's externalizing and internalizing behaviors, while father technoference in parenting did not show a significant correlation with the child's problem behavior. Thus, it is imperative to separately examine the influences of father and mother technoferences on children's development. Notably, existing research has predominantly focused on the relationship between technoference from both parents and child problematic media use [2,18,28,29,31,42,44], without distinguishing the specific impacts of fathers and mothers. Therefore, this study aims to fill this gap by exploring the moderating effects of parental technoference groups on child problematic media use.

Finally, we examine the impact of different types of child problematic media use, publication status, and study design on the relationship between parental technoference and child problematic media use. Specifically, we investigated whether various types of problematic media use influence this relationship. Furthermore, we compared the findings of published studies with those of unpublished master's thesis to determine the consistency of conclusions. Additionally, we examined potential differences in conclusions between cross-sectional and longitudinal research in the context of parental technoference studies.

Objectives

As discussed earlier, the empirical research investigating the relationship between parental technoference and child problematic media use has yielded inconsistent results. To our knowledge, there are no meta-analysis studies in this area to explain these differences. Therefore, this study performed a meta-analysis to synthesize the findings of previous studies examining the association between parental technoference and child problematic media use to better understand the associations between these constructs. The meta-analysis aims to address two core research questions.

- What is the overall relationship between parental technoference and child problematic media use?
- Are there any moderating variables that influence the relationship between parental technoference and child problematic media use?

We also explored potential moderating variables, such as the child's age, gender, type of problematic media use, parental technoference group, publication status, and research design.

Methods

Study Design

The meta-analysis followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines [47]. The PRISMA–Individual Participant Data checklist for this review is presented in Multimedia Appendix 1. To ensure transparency and avoid unintentional duplication of effort, the protocol for this meta-analysis was registered in the International Prospective Register for Systematic Reviews (PROSPERO CRD42023471997).

Search Strategy

A systematic literature search was conducted based on the PRISMA statement [47]. Multiple electronic databases, including Web of Science, EBSCO, ProQuest, PubMed, PsycINFO, and China National Knowledge Infrastructure, were searched using a predefined search string. The search terms consisted of three elements: (1) parent (eg, parent* OR parental* OR caregiver* OR guardian* OR dad* OR father* OR mom* OR mother* OR family); (2) child (eg, child* OR infant* OR baby OR babies OR toddler* OR preschool* OR kid* OR youth* OR teen* OR adolescent* OR young*); and (3) technoference (eg, technoference* OR phubbing* OR "technology interference*" OR "distraction with phone*" OR "digital distraction*" OR "smartphone distraction*" OR "device distraction*" OR "technology interruption*" OR "digital interruption*" OR "smartphone interruption*" OR "device interruption*" OR "parental media use*" OR "parental smartphone use*" OR "parental device use*"). The search strategy for all databases is presented in Multimedia Appendix 2. To avoid potential bias, outcome terms related to child problematic media use were not included in the search strategy, as suggested by Frandsen et al [48]. The last search was conducted in August 2024. Additionally, to ensure comprehensive coverage, the reference lists of eligible study reports and relevant reviews in the field of parental technoference were manually searched. This iterative process continued until no further studies could be identified.

Inclusion and Exclusion Criteria

The inclusion criteria for the literature in this meta-analysis were as follows: (1) studies involving children under the age of 22 years; (2) investigations examining the association

XSL•FO
RenderX

between parental technoference and child problematic media use; (3) use of a valid and reliable measure to assess parental technoference and child problematic media use; (4) clear reporting of correlation coefficients (*r*) between parental technoference and child problematic media use, or provision of 2-tailed *t* test values, *F* test values, or chi-square test values that could be converted into *r*; (5) reporting of the sample size; (6) inclusion of cross-sectional or longitudinal study designs; and (7) availability of studies written in either Chinese or English. Conversely, studies were excluded if they (1) assessed child outcome indicators other than problematic media use; (2) had a sample size of <30; or (3) were case reports or review papers.

Selection Procedure

The study selection process was conducted in multiple stages following standard systematic review procedures. First, duplicate studies were removed. Subsequently, four authors (JZ, QZ, Y Cao, and Y Chen) working in pairs independently screened the remaining articles. Each record was assessed independently by 2 authors in a 2-stage process: initial screening of titles and abstracts, followed by full-text evaluation of potentially eligible studies. The screening process was conducted under blinded conditions to minimize selection bias. Any discrepancies between the raters were resolved through discussion with the corresponding author until a consensus was reached.

Data Extraction

The following data were extracted: (1) first author names and publication year, (2) correlation coefficient, (3) the number of study samples, (4) gender distribution of children (measured by "boy ratio"), (5) average age of children, (6) measurement instrument, (7) parental technoference group (father technoference vs mother technoference vs parental technoference), (8) participant's country, and (9) publication type (journal paper vs thesis).

In the process of data extraction, we followed specific guidelines. First, effect sizes were generated based on independent samples, with each independent sample contributing 1 effect size for a specific parental technoference and child problematic media use. Second, when 2 effect sizes from the same sample could be classified into 2 different subgroups for moderator analysis, they were considered separate. For example, if a sample reported correlation coefficients for both father and mother technoferences with children's problematic media use separately, these were treated as independent when assessing the moderating effect of the parental technoference category. In other scenarios, to maintain the independence of effect sizes, when multiple effect sizes for the same variable within a group were reported, we used the average correlation to address dependency issues. Third, for longitudinal studies, only 1 longitudinal correlation between the initial measurement of parental technoference and the final measurement of child problematic media use was retained. Lastly, if participant characteristics, such as gender, were reported separately, they were coded as such.

To ensure coding reliability, 2 authors independently conducted the coding, and agreement levels were assessed. Interrater reliability was calculated using the intraclass correlation coefficient (ICC) for continuous variables and Cohen κ for categorical variables. For continuous variables, intercoder reliability was calculated for correlation coefficients (ICC=0.982), sample (ICC=0.997), gender (ICC=0.996), and age (ICC=1.000). For categorical variables, intercoder reliability was calculated for publication year (κ=1.000), publication status (κ=1.000), measurement of parental technoference (κ=1.000), parental technoference group (κ=1.000), and research design (κ=1.000). These results indicated excellent interrater reliability, reflecting a high level of agreement between the coders on the characteristics of the studies. Any discrepancies were resolved through discussion.

Risk of Bias Assessment

The risk of bias for all included studies was assessed using the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies [49]. This tool comprises 14 items, each offering 5 response choices: yes, no, cannot be determined, not reported, and not applicable. A score of 1 was assigned for "yes," while the other response choices did not receive any points. The total score was used to categorize the quality of the literature as good (total score>7), fair (total score between 5 and 7), or poor (total score<5). A high-quality study is generally characterized by minimal risk of bias. Two authors independently conducted the coding process, and their agreement on the total score demonstrated a high level of consistency (ICC=0.987). Any discrepancies in the coding process were resolved through consensus discussions. The specific details of the quality assessment for each study can be found in Multimedia Appendix 3 [2,16-18,27-44,50-80].

Statistical Analyses

The meta-analysis was conducted using R (version 4.2.1-win; R Foundation for Statistical Computing) with the *meta* and *metafor* packages. Effect sizes were measured using correlation coefficients (*r*). Before the meta-analysis, all correlation coefficients were transformed into Fisher *z* scores. After the analysis, Fisher *z* values were converted back to Pearson correlation coefficients for easier interpretation. Given the anticipated heterogeneity in measurement methods for parental technoference and child problematic media use, as well as variations in participant characteristics (eg, gender, age, cultural background) across the included studies, we used a random effects model for our meta-analysis. Heterogeneity was assessed using the Q and I^2 test statistics. A significant Q test or an I^2 value above 75% indicated substantial heterogeneity, supporting the use of a random effects model [81].

For the moderator analysis, continuous moderators were analyzed using meta-regression, while categorical moderators were analyzed using subgroup analysis. The significance of moderators was assessed using the Q statistic. As recommended by Huang [82], each subgroup should include a minimum of 3 studies for the analysis of categorized moderating variables.

Publication bias refers to the tendency for significant results to be more likely to be published, while nonsignificant results may remain unpublished. To address publication bias, this study included both published journal papers and unpublished theses when selecting literature, which helped to control the influence

XSL•FO
RenderX

of publication bias on the research findings. Additionally, to ensure the reliability of the meta-analysis results, funnel plot and Egger regression intercept were used to assess the presence of publication bias. If the funnel plot exhibited a symmetrical inverted funnel shape and the Egger regression intercept was nonsignificant, publication bias was considered negligible.

Results

Study Characteristics

Out of 1232 initially identified studies (Figure 1), 420 duplicates were excluded. After screening the titles and abstracts, 506 studies were excluded, leaving 306 studies for full-text screening. Following predefined inclusion and exclusion criteria, 253 studies were excluded. Consequently, 53 studies met all eligibility criteria and were included in the meta-analysis.

Figure 1. Flow diagram of study search.

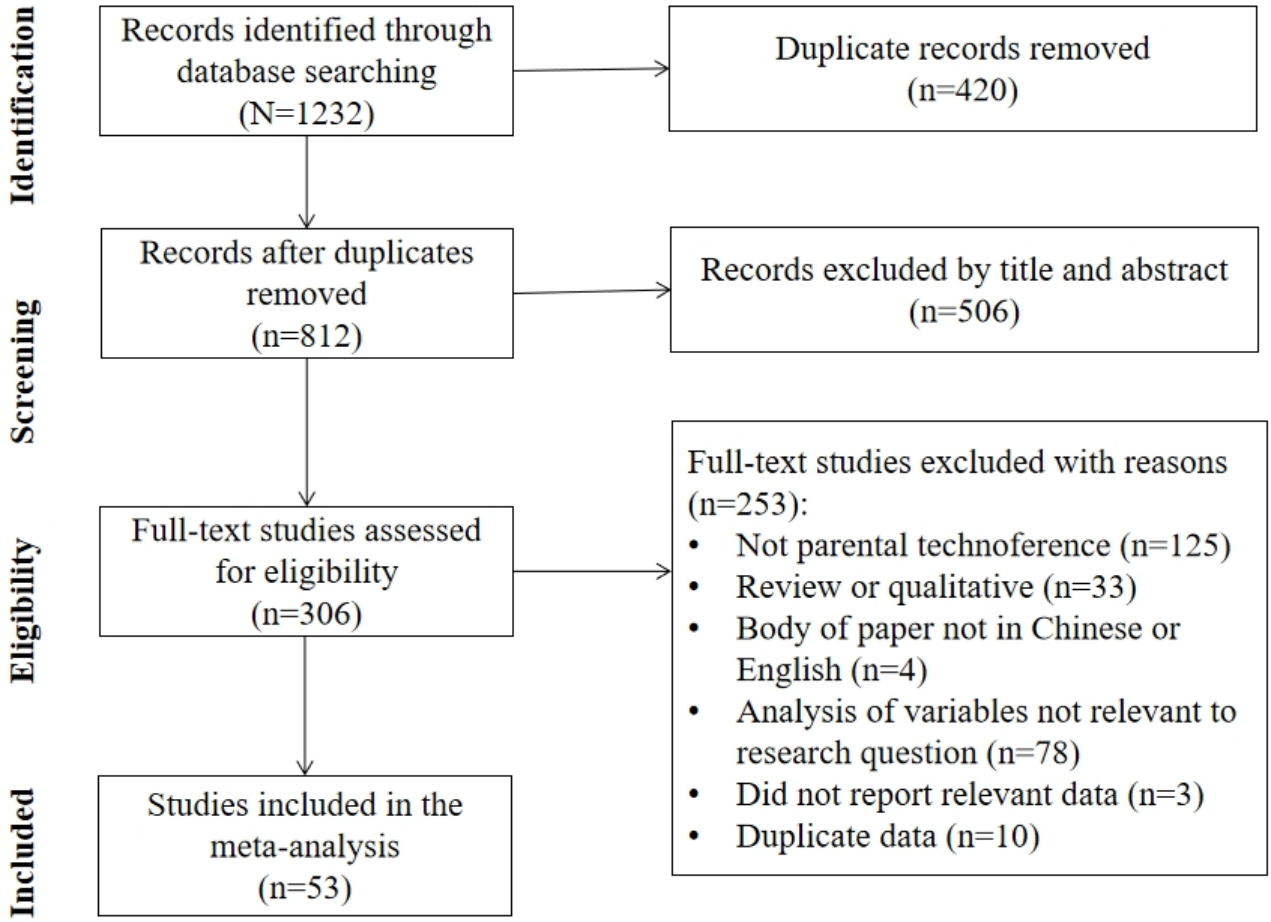

The specific information of the literature included in the meta-analysis is shown in Table 1. The meta-analysis included a total of 60 effect sizes from 53 studies, encompassing 60,555 participants. The sample sizes of the studies ranged from 227 to 4172, with an average participant age of 13.84 (SD 1.18) years. The publication years of the included studies varied from 2018 to 2024. Of the total studies, 46 were cross-sectional, while 7 were longitudinal. The literature consisted of 44 published journal papers and 9 unpublished master's theses. The primary measurement method used in parental technoference research was questionnaire surveys. Commonly used questionnaires included the Partner Phubbing Scale, Phubbing Scale, and the Generic Scale of Being Phubbed. The analysis directly assessed parental technoference in 44 studies, while 7 studies evaluated father technoference and mother technoference separately, 1 study exclusively evaluated mother technoference, and 1 study solely evaluated father technoference. Regarding child problematic media use, the analysis covered problematic smartphone problem use, problematic internet use, problematic social media use, internet gaming disorder, and short-form video addiction, with 34 studies, 8 studies, 5 studies, 5 studies, and 1 study, respectively. Furthermore, 50 of the studies were conducted in China, while the remaining studies were from the United States, Italy, and the Netherlands. In terms of literature quality, 6 studies were rated as "fair," while the remaining 47 were rated as "good."

XSL•FO
RenderX

Table 1. Overview of studies included in the meta-analysis.

Authors (publication year)	Participants, n	Age (y), mean (SD)	Gender (boys), %	Technoference type	Questionnaire survey	Indicators	Publication type	Research design	Country	Literature quality
Chen et al (2023) [50]	728	—[a]	0.50	PT[b]	PPS[c]	PMU[d]	Journal	Cross-sectional research	China	Good
Dai et al (2024) [51]	1912	11.03 (1.13)	0.53	FT[e] and MT[f]	PPS	PIU[g]	Journal	Cross-sectional research	China	Good
Deng and Hong (2023) [33]	855	12.25 (0.58)	0.53	FT and MT	PPS	PIU	Journal	Cross-sectional research	China	Fair
Ding et al (2018) [52]	312	12.89 (0.71)	0.55	PT	PPS	PMU	Journal	Longitudinal research	China	Good
Ding et al (2019) [53]	555	12.86 (0.71)	0.52	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Ding et al (2020) [54]	574	12.90 (0.71)	0.53	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Ding et al (2022) [55]	812	—	0.43	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Dong et al (2022) [2]	2286	13.46 (0.93)	0.49	PT	PPS	PSMU[h]	Journal	Cross-sectional research	China	Good
Fu et al (2020) [56]	2238	13.89 (2.44)	0.50	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Geng et al (2021) [27]	1447	16.15 (0.65)	0.40	FT and MT	GSBP[i]	PMU	Journal	Longitudinal research	China	Good
Geurts et al (2022) [57]	403	13.51 (2.15)	0.47	PT	SCS[j]	PSMU	Journal	Cross-sectional research	Netherlands	Good
Han (2021) [58]	582	13.53 (0.83)	0.48	PT	PPS	PIU	Thesis	Cross-sectional research	China	Good
Hernandez et al (2024) [59]	227	—	0.45	PT	SCS	PIU	Journal	Cross-sectional research	America	Good
Hong et al (2019) [44]	1721	13.69 (1.64)	—	PT	PPS	PMU	Journal	Longitudinal research	China	Good
Ji (2022) [60]	2090	16.00 (1.32)	0.44	PT	PPS	PMU	Thesis	Cross-sectional research	China	Good
Liu et al (2019) [43]	602	15.09 (2.89)	0.52	PT	PPS	PSMU	Journal	Cross-sectional research	China	Fair
Liu et al (2020) [28]	3051	13.08 (0.89)	0.49	PT	TILES[k]	PMU	Journal	Cross-sectional research	China	Good
Liu et al (2024) [61]	495	13.39 (0.77)	0.45	PT	PPS	PIU	Journal	Cross-sectional research	China	Good
Liu and Wu (2024) [62]	2465	10.43 (0.99)	0.53	PT	TILES	PMU	Journal	Cross-sectional research	China	Fair
Liu et al (2022) [63]	1202	14.54 (1.79)	-	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Liu (2021) [64]	822	13.66 (0.96)	0.50	FT and MT	PPS	PMU	Thesis	Cross-sectional research	China	Good
Ma (2023) [65]	1173	17.00 (0.91)	0.62	PT	PPS	PMU	Thesis	Longitudinal research	China	Good
Ma et al (2024) [66]	786	13.71 (1.35)	0.53	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Mi et al (2023) [16]	780	14.04 (0.93)	0.50	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Mu et al (2022) [17]	242	—	0.34	PT	PPS	SFVA[l]	Journal	Cross-sectional research	China	Fair

XSL•FO
RenderX

Authors (publication year)	Participants, n	Age (y), mean (SD)	Gender (boys), %	Technoference type	Questionnaire survey	Indicators	Publication type	Research design	Country	Literature quality
Niu et al (2020) [29]	726	14.55 (1.37)	0.51	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Ou and Zhu (2021) [40]	636	19.60 (1.40)	0.45	PT	PPS	PMU	Journal	Cross-sectional research	China	Fair
Pivetta et al (2024) [67]	557	15.62 (1.54)	0.69	FT and MT	PPS	IGD[m]	Journal	Longitudinal research	Italy	Good
Qiao and Liu (2020) [30]	1354	16.10 (0.96)	0.46	PT	TILES	PMU	Journal	Cross-sectional research	China	Good
Qu and Zhang (2020) [34]	318	11.56 (0.92)	0.49	FT and MT	TDIS[n]	IGD	Journal	Cross-sectional research	China	Good
Shao et al (2024) [68]	3023	10.59 (0.32)	0.57	PT	TTS[o]	PMU	Journal	Cross-sectional research	China	Good
Shen et al (2022) [35]	809	13.15 (0.62)	0.49	PT	PPS	IGD	Journal	Cross-sectional research	China	Good
Tang et al (2024) [69]	742	12.97 (0.64)	0.55	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Wang and Lei (2022) [39]	549	17.01 (2.09)	0.55	PT	PPS	PIU	Journal	Cross-sectional research	China	Good
Wang et al (2022) [38]	4172	16.41 (0.77)	0.52	FT	GSBP	PSMU	Journal	Cross-sectional research	China	Good
Wang et al (2024) [70]	234	5.02 (1.10)	0.46	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Wang et al (2023) [42]	3519	16.42 (0.77)	0.47	MT	GSBP	PSMU	Journal	Cross-sectional research	China	Good
Wang et al (2023) [71]	2260	12.76 (0.58)	0.50	PT	PPS	PMU	Journal	Longitudinal research	China	Good
Wu (2022) [72]	780	—	0.50	PT	PPS	PMU	Thesis	Cross-sectional research	China	Good
Xiao (2020) [73]	452	16.90 (0.93)	0.43	PT	PS[p]	PMU	Thesis	Cross-sectional research	China	Good
Xie et al (2019) [31]	1007	13.85 (1.53)	0.49	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Xie et al (2021) [36]	779	13.15 (0.61)	0.49	PT	PPS	IGD	Journal	Cross-sectional research	China	Good
Yang et al (2022) [74]	2220	13.97 (1.92)	0.46	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Yang et al (2022) [75]	341	13.41 (0.68)	0.52	PT	PPS	PMU	Journal	Longitudinal research	China	Good
Yang (2022) [76]	301	—	0.32	PT	PPS	PIU	Thesis	Cross-sectional research	China	Fair
Yang (2022) [77]	944	13.89 (0.87)	0.45	PT	PPS	PMU	Thesis	Cross-sectional research	China	Good
Zhang and Zhang (2020) [41]	411	19.82 (1.34)	0.38	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Zhang et al (2021) [18]	471	13.46 (1.11)	0.40	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Zhao et al (2022) [32]	931	13.54 (1.08)	—	PT	PPS	PMU	Journal	Cross-sectional research	China	Good
Zhong et al (2023) [78]	2174	13.25 (0.93)	0.51	PT	TILES	PMU	Journal	Cross-sectional research	China	Good
Zhou et al (2022) [37]	1021	10.33 (0.98)	0.56	PT	PPS	IGD	Journal	Cross-sectional research	China	Good

XSL•FO
RenderX

Authors (publication year)	Participants, n	Age (y), mean (SD)	Gender (boys), %	Technoference type	Questionnaire survey	Indicators	Publication type	Research design	Country	Literature quality
Zhu and Jiang (2022) [79]	1034	13.45 (0.68)	0.53	FT and MT	PPS	PIU	Journal	Cross-sectional research	China	Good
Zhu (2023) [80]	430	—	0.48	PT	PPS	PMU	Thesis	Cross-sectional research	China	Good

[a]Data missing.
[b]PT: parental technoference.
[c]PPS: Partner Phubbing Scale.
[d]PMU: problematic mobile use.
[e]FT: father technoference.
[f]MT: mother technoference.
[g]PIU: problematic internet use.
[h]PSMU: problematic social media use.
[i]GSBP: Generic Scale of Being Phubbed.
[j]SCS: Self-constructed Scale.
[k]TILES: Technology Interference in Life Examples Scale.
[l]SFVA: short-form video addiction.
[m]IGD: internet gaming disorder.
[n]TDIS: Teenagers-perceived Digital Interference Scale.
[o]TTS: The Technoference Scale.
[p]PS: Phubbing Scale.

Analysis of Publication Bias

Funnel plots show that effect sizes were roughly evenly distributed on either side of the total effect (Figure 2). The Egger test showed no intercept values in the outcome categories were statistically significant. Specifically, the intercept was –0.78 (P=.71). Collectively, these analyses suggest that the findings of this meta-analysis are robust and unlikely to be substantially influenced by publication bias.

Figure 2. Funnel plot of the association between parental technoference and child problematic media use.

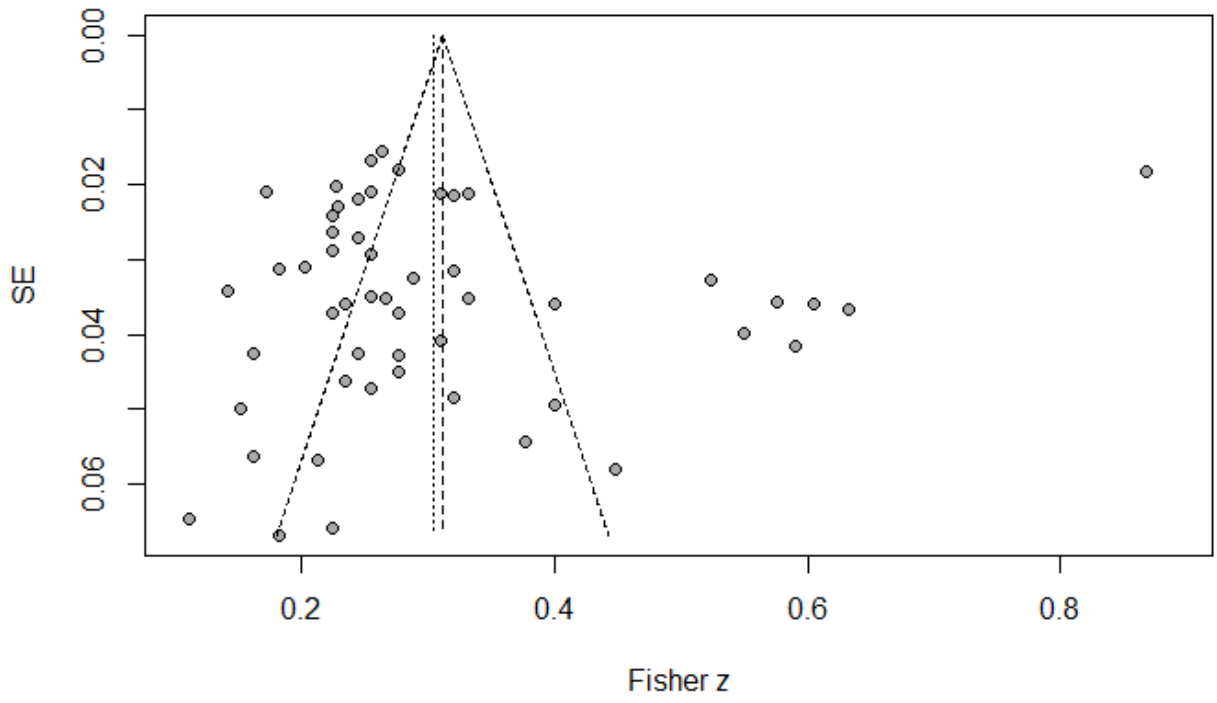

XSL•FO
RenderX

Overall Relation Between Parental Technoference and Child Problematic Media Use

The homogeneity test revealed significant heterogeneity among the effect sizes (Q_{52}=1558.22; P<.001; I^2=96.7%). These findings indicate that there was substantial variability in the effect sizes, suggesting the presence of real differences. Consequently, a random effects model was used to examine both the overall effect and the moderating effect.

The analysis showed a significant positive correlation between parental technoference and child problematic media use (r=0.296; 95% CI 0.259-0.331), as shown in Figure 3 [2,16-18,27-44,50-80]. According to Lipsey and Wilson [83], correlation coefficients can be categorized as low (|r|≤0.1), medium (0.1<|r|<0.4), or high (|r|≥0.4). In this study, the mean effect sizes of the correlation between parental technoference and child problematic media use were considered medium.

Figure 3. Forest plot for correlation between parental technoference and child problematic media use.

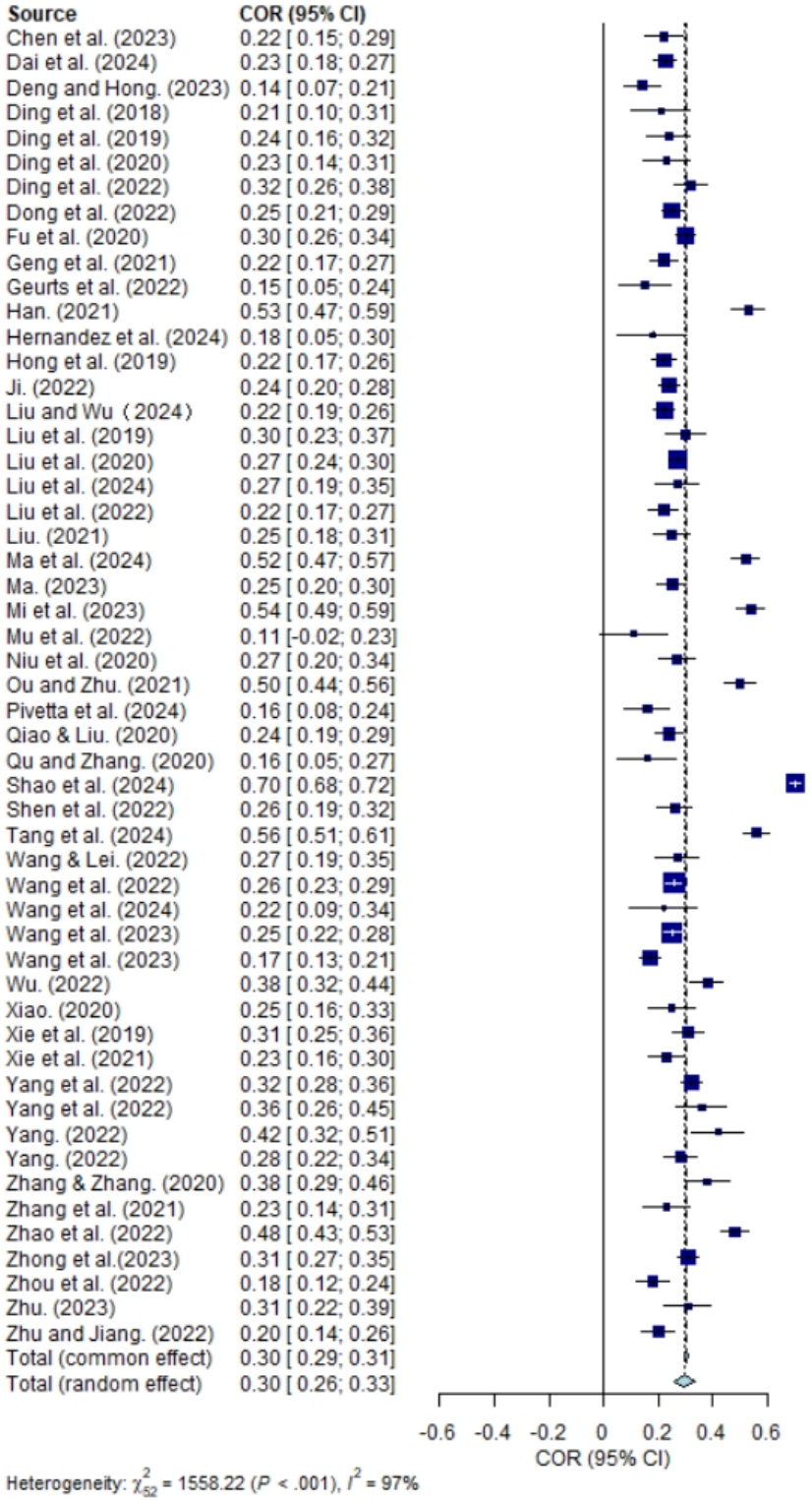

XSL•FO
RenderX

Analysis of Moderator Variables

The results of the meta-regression analysis indicated that the moderating effects of gender (B=0.123; SE 0.341; 95% CI 0.546 to 0.792; z=0.360; P=.72) and age (B=0.004; SE 0.010; 95% CI 0.015 to 0.022; z=0.368; P=.71) was found not significant.

Table 2 presents the detailed analysis results for the categorized moderating variables. The analysis revealed no significant moderating effects for publication status (P=.39) or the type of problematic media use (P=.24). However, the parental technoference group emerged as a significant moderator (P<.001), indicating that the effect size of technoference on children's problematic media use was larger when both parents engaged in technoference compared to when only one parent exhibited such behavior. Furthermore, the study design demonstrated a significant moderating effect (P=.008). Specifically, cross-sectional studies yielded larger effect sizes for the relationship between parental technoference and child problematic media use compared to longitudinal studies.

Table 2. Categorical moderator analysis of parental phubbing and internalizing problems.

Moderator variables	Studies, n	*r*	95% CI	*Q* test (*df*)	*P* value
Parental technoference group				18.08 (2)	<.001
Both parents	44	0.313	0.272-0.354		
Father	8	0.211	0.182-0.239		
Mother	8	0.208	0.171-0.246		
Publication status				0.73 (1)	.39
Journal paper	44	0.290	0.248-0.330		
Thesis	9	0.325	0.254-0.392		
Study design				6.84 (1)	.008
Cross-sectional	46	0.306	0.265-0.345		
Longitudinal	7	0.223	0.174-0.270		
Type of PMU[a]				1.38 (1)	.24
General problematic media use	42	0.305	0.261-0.348		
Specific problematic media use	12	0.211	0.202-0.316		

[a]PMU: problematic mobile use.

Sensitivity Analysis

Furthermore, sensitivity analysis was performed to assess the impact of each study using the leave-one-out method. The sensitivity analysis results indicated that the effect size r remained significant, ranging from 0.275 to 0.307, even after excluding any single sample. This suggests that the findings of this study remained relatively stable and reliable.

Discussion

Principal Findings

Despite the growing body of empirical research exploring the relationship between parental technoference and child problematic media use, inconsistencies in research conclusions persist. No meta-analysis had been conducted to systematically synthesize these findings—until now. This study used a meta-analytical approach to quantitatively aggregate existing research, examining the relationship between parental technoference and child problematic media use, as well as identifying moderating factors within this relationship. The results indicated a positive association between parental technoference and child problematic media use. Furthermore, moderator analyses showed that the parental technoference group and study design significantly influenced this relationship.

Parental Technoference and Child Problematic Media Use

The present meta-analysis provides compelling evidence for a significant positive correlation between parental technoference and child problematic media use. This finding aligns with the interactional theory of childhood problematic media use [19], which posits that children's media use patterns are influenced by their interactions with their environment, including parental behaviors. Parental technoference, characterized by reduced parent-child interaction and unclear media usage rules [11,20-22], may lead children to engage in excessive media use as a means of compensating for the lack of parental attention and communication.

Moreover, the accept-rejection theory [25] offers another perspective on the relationship between parental technoference and child problematic media use. When parents are preoccupied with technology, children may perceive this as a form of rejection or neglect. Consequently, children may turn to excessive media use as an alternative source of comfort and stimulation to cope with the perceived lack of emotional connection with their parents.

Furthermore, social learning theory [26] suggests that children learn by observing and imitating the behaviors of their parents. When parents engage in excessive media use or prioritize technology over other activities, they inadvertently model

XSL•FO
RenderX

problematic media use behaviors. As a result, children may adopt similar problematic media use patterns through observational learning processes.

This meta-analysis, in conjunction with the theoretical perspectives discussed, provides a more comprehensive understanding of the potential mechanisms underlying the relationship between parental technoference and child problematic media use. However, it is important to acknowledge that the strength of this association may vary across studies due to methodological differences, such as sample characteristics, measurement tools, and cultural contexts [16,17]. Future research should aim to further elucidate these factors and their role in modulating the relationship between parental technoference and child problematic media use.

In conclusion, the meta-analysis substantiates the theoretical assertions that parental technoference is a significant contributor to child problematic media use. This relationship can be understood through the lenses of the interactional theory of childhood problematic media use, accept-rejection theory, and social learning theory, which collectively suggest that parental technoference may disrupt parent-child interactions, contribute to children's perception of rejection, and model problematic media use behaviors.

Moderator Analysis

The significant moderating effect of the parental technoference group on the relationship between parental technoference and child problematic media use suggests a compounded risk when both parents are involved in technoference. This finding aligns with Zhang et al [84], who reported a stronger association between parental technoference and children’s internalizing problems when both parents, rather than one, engaged in technoference. Our finding extends this understanding to problematic media use, indicating that the combined technoference of both parents may create a more pervasive environment of digital distraction, which could amplify negative outcomes in children’s media behaviors. This could be due to a cumulative lack of parental attention and modeling of poor media habits, indicating the necessity for interventions that address family media use as a whole. Future research should explore the mechanisms through which the combined technoference of both parents exerts a greater influence on child outcomes, potentially informing targeted family-based strategies to mitigate problematic media use in children.

The significant moderating effect of study design, with cross-sectional studies yielding larger effect sizes compared to longitudinal studies, suggests that the impact of parental technoference on child problematic media use may diminish over time. This finding highlights the potential for adaptation or the development of coping mechanisms in children exposed to parental technoference. It also emphasizes the importance of longitudinal research to capture the dynamic nature of this relationship and to inform long-term intervention strategies.

The study did not find significant moderating effects for child gender, age, publication status, or type of problematic media use. This indicates that parental technoference exerts a consistent impact on child problematic media use across different genders, ages, and types of media use issues. Such consistency underscores the pervasive influence of parental technoference, suggesting its stable impact on children’s media behaviors irrespective of individual or contextual differences.

Limitations and Future Directions

This meta-analysis provides valuable insights into the relationship between parental technoference and child problematic media use. However, several limitations warrant consideration. The predominance of Chinese participants in the sample limits the generalizability of findings to other cultural contexts, particularly those with differing parenting dynamics and societal values. Future research should encompass a broader cultural spectrum to assess the global applicability of these results. Moreover, the reliance on cross-sectional studies in the analyzed literature presents challenges in establishing causality and may introduce recall biases. To address these limitations, future studies should use experimental methods to isolate the direct impact of parental technoference on child outcomes. Additionally, longitudinal designs spanning extended periods would enable the examination of developmental trajectories and the dynamic interplay between parental technoference and child problematic media use over time. Implementing these methodological approaches in future research will enhance our understanding of the causal mechanisms and long-term implications of parental technoference, ultimately informing more effective interventions to promote healthy family media use practices.

Conclusions

This meta-analysis provides robust evidence for a significant positive association between parental technoference and child problematic media use. The findings reveal that this relationship is particularly pronounced when both parents engage in technoference and are stronger in cross-sectional studies compared to longitudinal ones. These results underscore the pervasive influence of parental digital distraction on children’s media habits across various demographics and types of media use. The study highlights the critical need for family-based interventions targeting parental technology use and emphasizes the importance of longitudinal research to understand the temporal dynamics of this relationship better. By addressing parental technoference, we may significantly impact children’s problematic media use, potentially improving developmental outcomes and overall well-being in the digital age. Future research should focus on developing targeted interventions and exploring the underlying mechanisms of this relationship across diverse cultural contexts.

XSL•FO
RenderX

Acknowledgments

This study was supported by STI 2030—major projects (2022ZD0209000, 2022ZD0209001, 2022ZD0209002, 2022ZD0209003, 2022ZD0209004, 2022ZD0209005, and 2021ZD0200516). The funding agents had no role in the study design, the collection, analysis, and interpretation of the data, or in the writing of the manuscript.

Data Availability

The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

The execution of this paper was a collaborative effort on behalf of all authors. JZ and YL conceptualized and designed the study. JZ, QZ, and YL conducted the data collection and analysis. JZ, QZ, and Y Cao drafted the initial manuscript before BX and Y Chen contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-IPD (Preferred Reporting Items for Systematic Review and Meta-Analyses of Individual Participant Data) checklist.
[DOCX File , 27 KB - jmir_v27i1e57636_app1.docx]

Multimedia Appendix 2
Search strategy for all databases.
[DOCX File , 20 KB - jmir_v27i1e57636_app2.docx]

Multimedia Appendix 3
Quality assessment rating for each study included in the meta-analysis.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v27i1e57636_app3.xlsx]

References

1. Domoff SE, Harrison K, Gearhardt AN, Gentile DA, Lumeng JC, Miller AL. Development and validation of the Problematic Media Use Measure: a parent report measure of screen media "addiction" in children. Psychol Pop Media Cult 2019;8(1):2-11 [FREE Full text] [doi: 10.1037/ppm0000163] [Medline: 30873299]
2. Dong W, Li S, Wang X. Parental phubbing and Chinese adolescents' SNSs addiction: loneliness as a mediator and self-esteem as a moderator. Curr Psychol 2022;42:16560-16572. [doi: 10.1007/s12144-022-04066-y]
3. Lozano-Blasco R, Robres AQ, Sánchez AS. Internet addiction in young adults: a meta-analysis and systematic review. Comput Hum Behav 2022;130:107201 [FREE Full text] [doi: 10.1016/j.chb.2022.107201]
4. Olson JA, Sandra DA, Colucci ES, Bikaii AA, Chmoulevitch D, Nahas J, et al. Smartphone addiction is increasing across the world: a meta-analysis of 24 countries. Comput Hum Behav 2022;129:107138 [FREE Full text] [doi: 10.1016/j.chb.2021.107138]
5. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
6. Huang C. A meta-analysis of the problematic social media use and mental health. Int J Soc Psychiatry 2022;68(1):12-33. [doi: 10.1177/0020764020978434] [Medline: 33295241]
7. Marino C, Gini G, Vieno A, Spada MM. The associations between problematic facebook use, psychological distress and well-being among adolescents and young adults: a systematic review and meta-analysis. J Affect Disord 2018;226:274-281. [doi: 10.1016/j.jad.2017.10.007] [Medline: 29024900]
8. Sohn SY, Rees P, Wildridge B, Kalk NJ, Carter B. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. BMC Psychiatry 2019;19(1):356 [FREE Full text] [doi: 10.1186/s12888-019-2350-x] [Medline: 31779637]
9. Yigiter MS, Demir S, Dogan N. The relationship between problematic social media use and depression: a meta-analysis study. Curr Psychol 2023;43(9):7936-7951. [doi: 10.1007/s12144-023-04972-9]
10. Shannon H, Bush K, Villeneuve PJ, Hellemans KG, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. JMIR Ment Health 2022;9(4):e33450 [FREE Full text] [doi: 10.2196/33450] [Medline: 35436240]

XSL•FO
RenderX

11. Komanchuk J, Toews AJ, Marshall S, Mackay LJ, Hayden KA, Cameron JL, et al. Impacts of parental technoference on parent-child relationships and child health and developmental outcomes: a scoping review. Cyberpsychol Behav Soc Netw 2023;26(8):579-603. [doi: 10.1089/cyber.2022.0278] [Medline: 37347957]
12. McDaniel BT, Coyne SM. "Technoference": the interference of technology in couple relationships and implications for women's personal and relational well-being. Psychol Pop Media Cult 2016;5(1):85-98. [doi: 10.1037/ppm0000065]
13. McDaniel BT. Parent distraction with phones, reasons for use, and impacts on parenting and child outcomes: a review of the emerging research. Human Behav and Emerg Tech 2019;1(2):72-80. [doi: 10.1002/hbe2.139]
14. McDaniel BT, Coyne SM. Technology interference in the parenting of young children: implications for mothers' perceptions of coparenting. Soc Sci J 2019;53(4):435-443. [doi: 10.1016/j.soscij.2016.04.010]
15. McDaniel BT, Radesky JS. Technoference: parent distraction with technology and associations with child behavior problems. Child Dev 2018;89(1):100-109 [FREE Full text] [doi: 10.1111/cdev.12822] [Medline: 28493400]
16. Mi Z, Cao W, Diao W, Wu M, Fang X. The relationship between parental phubbing and mobile phone addiction in junior high school students: a moderated mediation model. Front Psychol 2023;14:1117221 [FREE Full text] [doi: 10.3389/fpsyg.2023.1117221] [Medline: 37123292]
17. Mu H, Jiang Q, Xu J, Chen S. Drivers and consequences of short-form video (SFV) addiction amongst adolescents in China: stress-coping theory perspective. Int J Environ Res Public Health 2022;19(21):14173 [FREE Full text] [doi: 10.3390/ijerph192114173] [Medline: 36361050]
18. Zhang Y, Ding Q, Wang Z. Why parental phubbing is at risk for adolescent mobile phone addiction: a serial mediating model. Child Youth Serv Rev 2021;121:105873. [doi: 10.1016/j.childyouth.2020.105873]
19. Domoff SE, Borgen AL, Radesky JS. Interactional theory of childhood problematic media use. Hum Behav Emerg Technol 2020;2(4):343-353 [FREE Full text] [doi: 10.1002/hbe2.217] [Medline: 36381426]
20. Beamish N, Fisher J, Rowe H. Parents' use of mobile computing devices, caregiving and the social and emotional development of children: a systematic review of the evidence. Australas Psychiatry 2019;27(2):132-143. [doi: 10.1177/1039856218789764] [Medline: 30234373]
21. Braune-Krickau K, Schneebeli L, Pehlke-Milde J, Gemperle M, Koch R, von Wyl A. Smartphones in the nursery: parental smartphone use and parental sensitivity and responsiveness within parent-child interaction in early childhood (0-5 years): a scoping review. Infant Ment Health J 2021;42(2):161-175 [FREE Full text] [doi: 10.1002/imhj.21908] [Medline: 33452702]
22. Kildare CA, Middlemiss W. Impact of parents mobile device use on parent-child interaction: a literature review. Comput Hum Behav 2017;75:579-593 [FREE Full text] [doi: 10.1016/j.chb.2017.06.003]
23. Schneider LA, King DL, Delfabbro PH. Family factors in adolescent problematic internet gaming: a systematic review. J Behav Addict 2017;6(3):321-333 [FREE Full text] [doi: 10.1556/2006.6.2017.035] [Medline: 28762279]
24. Zhu Y, Deng L, Wan K. The association between parent-child relationship and problematic internet use among English- and Chinese-language studies: a meta-analysis. Front Psychol 2022;13:885819 [FREE Full text] [doi: 10.3389/fpsyg.2022.885819] [Medline: 36110283]
25. Rohner RP, Lansford JE. Deep structure of the human affectional system: introduction to interpersonal acceptance–rejection theory. J Family Theory Rev 2017;9(4):426-440. [doi: 10.1111/jftr.12219]
26. Bandura A. Model of causality in social learning theory. In: Freeman A, Mahoney MJ, DeVito P, Martin D, editors. Cognition and Psychotherapy. New York, NY: Springer Publishing Co; 2004:81-99.
27. Geng J, Lei L, Ouyang M, Nie J, Wang P. The influence of perceived parental phubbing on adolescents' problematic smartphone use: a two-wave multiple mediation model. Addict Behav 2021;121:106995. [doi: 10.1016/j.addbeh.2021.106995] [Medline: 34090204]
28. Liu Q, Wu J, Zhou Z, Wang W. Parental technoference and smartphone addiction in Chinese adolescents: the mediating role of social sensitivity and loneliness. Child Youth Serv Rev 2020;118:105434. [doi: 10.1016/j.childyouth.2020.105434]
29. Niu G, Yao L, Wu L, Tian Y, Xu L, Sun X. Parental phubbing and adolescent problematic mobile phone use: the role of parent-child relationship and self-control. Child Youth Serv Rev 2020;116:105247. [doi: 10.1016/j.childyouth.2020.105247]
30. Qiao L, Liu Q. The effect of technoference in parent-child relationships on adolescent smartphone addiction: the role of cognitive factors. Child Youth Serv Rev 2020;118:105340. [doi: 10.1016/j.childyouth.2020.105340]
31. Xie X, Chen W, Zhu X, He D. Parents' phubbing increases adolescents' mobile phone addiction: roles of parent-child attachment, deviant peers, and gender. Child Youth Serv Rev 2019;105:104426. [doi: 10.1016/j.childyouth.2019.104426]
32. Zhao J, Ye B, Luo L, Yu L. The effect of parent phubbing on Chinese adolescents' smartphone addiction during COVID-19 pandemic: testing a moderated mediation model. Psychol Res Behav Manag 2022;15:569-579 [FREE Full text] [doi: 10.2147/PRBM.S349105] [Medline: 35281997]
33. Deng N, Hong Y. The influence of parents' "phubbing" on internet addiction behavior of junior middle school students: the mediating role of psychological resilience [Article in Chinese]. Mental Health Educ Prim Secondary Sch 2023(22):14-18 [FREE Full text]
34. Qu M, Zhang Y. The relationship between parental electronic device interference and adolescent online gaming addiction [Article in Chinese]. Chin J Health Stat 2020;37(02):212-214 [FREE Full text]
35. Shen X, Xie X, Wu S. Do adolescents addict to internet games after being phubbed by parents? The roles of maladaptive cognition and self-esteem. Curr Psychol 2022;42(3):2255-2267. [doi: 10.1007/s12144-022-03255-z]

XSL•FO
RenderX

36. Xie X, Guo Q, Wang P. Childhood parental neglect and adolescent internet gaming disorder: from the perspective of a distal-proximal-process-outcome model. Child Youth Serv Rev 2021;120(6):105564. [doi: 10.1016/j.childyouth.2020.105564]
37. Zhou J, Li X, Gong X. Parental phubbing and internet gaming addiction in children: mediating roles of parent-child relationships and depressive symptoms. Cyberpsychol Behav Soc Netw 2022;25(8):512-517. [doi: 10.1089/cyber.2022.0021] [Medline: 35759389]
38. Wang P, Hu H, Mo PKH, Ouyang M, Geng J, Zeng P, et al. How is father phubbing associated with adolescents' social networking sites addiction? Roles of narcissism, need to belong, and loneliness. J Psychol 2022;156(5):331-348. [doi: 10.1080/00223980.2022.2034726] [Medline: 35482939]
39. Wang H, Lei L. The relationship between parental phubbing and short-form videos addiction among Chinese adolescents. J Res Adolesc 2022;32(4):1580-1591. [doi: 10.1111/jora.12744] [Medline: 35253320]
40. Ou Q, Zhu W. The influence of parents' phubbing behavior on college students' smartphone addiction. J Campus Life Ment Health 2021;19(04):326-329. [doi: 10.19521/j.cnki.1673-1662.2021.04.005]
41. Zhang L, Zhang L. Revision of the Chinese version of Generic Scale of Phubbing. Chin J Clin Psychol 2020;28(01):29-32+6. [doi: 10.16128/j.cnki.1005-3611.2020.01.007]
42. Wang P, Ouyang M, Yin Y, Li B. Mother phubbing and adolescents' problematic SNS use: the mediating role of perceived burdensomeness and the moderating role of need to belong. Front Psychol 2023;14:1098707 [FREE Full text] [doi: 10.3389/fpsyg.2023.1098707] [Medline: 37359861]
43. Liu RD, Wang J, Gu D, Ding Y, Oei TP, Hong W, et al. The effect of parental phubbing on teenager's mobile phone dependency behaviors: the mediation role of subjective norm and dependency intention. Psychol Res Behav Manag 2019;12:1059-1069 [FREE Full text] [doi: 10.2147/PRBM.S224133] [Medline: 31819681]
44. Hong W, Liu RD, Ding Y, Oei TP, Zhen R, Jiang S. Parents' phubbing and problematic mobile phone use: the roles of the parent-child relationship and children's self-esteem. Cyberpsychol Behav Soc Netw 2019;22(12):779-786. [doi: 10.1089/cyber.2019.0179] [Medline: 31747305]
45. Collins WA, Russell G. Mother-child and father-child relationships in middle childhood and adolescence: a developmental analysis. Dev Rev 1991;11(2):99-136. [doi: 10.1016/0273-2297(91)90004-8]
46. McBride B, Mills G. A comparison of mother and father involvement with their preschool age children. Early Child Res Q 1993;8(4):457-477 [FREE Full text] [doi: 10.1016/s0885-2006(05)80080-8]
47. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. referred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data: the PRISMA-IPD statement. JAMA 2015;313(16):1657-1665. [doi: 10.1001/jama.2015.3656] [Medline: 25919529]
48. Frandsen TF, Bruun Nielsen MF, Lindhardt CL, Eriksen MB. Using the full PICO model as a search tool for systematic reviews resulted in lower recall for some PICO elements. J Clin Epidemiol 2020;127:69-75. [doi: 10.1016/j.jclinepi.2020.07.005] [Medline: 32679315]
49. Quality assessment tool for observational cohort and crosssectional studies. National Heart, Lung, and Blood Institute. 2014. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2023-09-09]
50. Chen S, Qiu D, Li X, Zhao Q. Discrepancies in adolescent-parent perceptions of parental phubbing and their relevance to adolescent smartphone dependence: the mediating role of parent-child relationship. Behav Sci 2023;13(11):888 [FREE Full text] [doi: 10.3390/bs13110888] [Medline: 37998635]
51. Dai B, Lin Y, Lai X, He J, Du M, Hou X, et al. The effects of self-esteem and parental phubbing on adolescent internet addiction and the moderating role of the classroom environment: a hierarchical linear model analysis. BMC Psychiatry 2024;24(1):55 [FREE Full text] [doi: 10.1186/s12888-023-05489-y] [Medline: 38243210]
52. Din Q, Kong L, ZhangY, Zhou Z, Hu W. Parents phubbing and mobile phone addiction in junior high school students: a cross-lagged analysis. Chin J Clin Psychol 2018;26(05):952-955. [doi: 10.16128/j.cnki.1005-3611.2018.05.025]
53. Ding Q, Zhang Y, Zhou Z. The relationship between parents phubbing and high school students mobile phone addiction: the moderating effect of parental monitoring [Article in Chinese]. Chin J Spec Educ 2019:66-71 [FREE Full text]
54. Ding Q, Wang Z, Zhang Y. Revision of the Chinese version of Parents Phubbing Scale in adolescents. Chin J Clin Psychol 2020;28(05):942-5+896. [doi: 10.16128/j.cnki.1005-3611.2020.05.017]
55. Ding Q, Luo X, Huang L, Zhang Y. Interparental conflict affects adolescent mobile phone addiction: based on the spillover hypothesis and emotional security theory. Chin J Clin Psychol 2022;30(03):653-657 [FREE Full text] [doi: 10.16128/j.cnki.1005-3611.2022.03.031]
56. Fu X, Liu J, Liu R, Ding Y, Hong W, Jiang S. The impact of parental active mediation on adolescent mobile phone dependency: a moderated mediation model. Comput Hum Behav 2020;107(6):106280. [doi: 10.1016/j.chb.2020.106280]
57. Geurts SM, Koning IM, Vossen HGM, van den Eijnden RJJM. Rules, role models or overall climate at home? Relative associations of different family aspects with adolescents' problematic social media use. Compr Psychiatry 2022;116:152318 [FREE Full text] [doi: 10.1016/j.comppsych.2022.152318] [Medline: 35537295]
58. Han Y. The Influence of Parents' Phubbing on Junior High School Students' Problematic Internet Use [master's thesis]. Taiyuan, China: Shanxi University; 2021.
59. Hernandez JM, Ben-Joseph EP, Reich S, Charmaraman L. Parental monitoring of early adolescent social technology use in the US: a mixed-method study. J Child Fam Stud 2023;33(3):759-776. [doi: 10.1007/s10826-023-02734-6]

XSL•FO
RenderX

60. Ji H. The Relationship Between Parents' Phubbing Behavior and Problematic Smartphone Use in Middle School Students [master's thesis]. Tianjin, China: Tianjin Normal University; 2022.
61. Liu S, Wu P, Han X, Wang M, Kan Y, Qin K, et al. Mom, dad, put down your phone and talk to me: how parental phubbing influences problematic internet use among adolescents. BMC Psychol 2024;12(1):125 [FREE Full text] [doi: 10.1186/s40359-024-01620-0] [Medline: 38443976]
62. Liu Q, Wu J. What children learn in a digital home: the complex influence of parental mediation and smartphone interference. Educ Inf Technol 2023;29(5):6273-6291. [doi: 10.1007/s10639-023-12071-2]
63. Liu QQ, Yang XJ, Nie YG. Interactive effects of cumulative social-environmental risk and trait mindfulness on different types of adolescent mobile phone addiction. Curr Psychol 2022;42(20):1-17. [doi: 10.1007/s12144-022-02899-1]
64. Liu Q. The Relationship Between Parental Phubbing and Middle School Students' Smartphone Addiction: The Role of Parent-Child Relationship and Intentional Self-Regulation [master's thesis]. Nanning, China: Guangxi University for Nationalities; 2021.
65. Ma F. The Impact of Parents' Phubbing Behavior on Mobile Phone Dependence Among Secondary Vocational Students: Evidence From Cross-Sectional, Follow-up and Intervention Studies [master's thesis]. Guangzhou, China: Guangzhou University; 2023.
66. Ma S, Bi X, Cui H, Ma Y. Parental phubbing and mobile phone addiction among Chinese adolescents: a moderated mediation model. Front Psychol 2024;15:1379388 [FREE Full text] [doi: 10.3389/fpsyg.2024.1379388] [Medline: 38860052]
67. Pivetta E, Marino C, Bottesi G, Pastore M, Vieno A, Canale N. The influence of maternal and paternal phubbing on adolescents' problematic gaming: a two-wave mediation model. Comput Hum Behav 2024;152:108058. [doi: 10.1016/j.chb.2023.108058]
68. Shao T, Zhu C, Lei H, Jiang Y, Wang H, Zhang C. The relationship of parent-child technoference and child problematic smartphone use: the roles of parent-child relationship, negative parenting styles, and children's gender. Psychol Res Behav Manag 2024;17:2067-2081 [FREE Full text] [doi: 10.2147/PRBM.S456411] [Medline: 38800525]
69. Tang Y, Xu J, Zhao Y, Davis T, Zhang H. Parental phubbing and adolescent smartphone addiction: depression and perceived school climate matter. Cyberpsychol Behav Soc Netw 2024;27(4):287-293. [doi: 10.1089/cyber.2023.0300] [Medline: 38407868]
70. Wang Y, Xia Q, Gu L, Li F. Chinese college teachers' work-family conflict and their preschool children's problem behaviors: the mediating effects of parental phubbing and children's problematic media use. Early Child Dev Care 2024;194(3):478-491. [doi: 10.1080/03004430.2024.2329685]
71. Wang X, Qiao Y, Wang S. Parental phubbing, problematic smartphone use, and adolescents' learning burnout: a cross-lagged panel analysis. J Affect Disord 2023;320:442-449. [doi: 10.1016/j.jad.2022.09.163] [Medline: 36206880]
72. Wu M. The Relationship Between Parental Phubbing and Smartphone Addiction in Junior High School Students [master's thesis]. Qingdao, China: Qingdao University; 2022.
73. Xiao M. The Influence of Parents' Phubbing on Adolescent Mobile Phone Addiction Tendency: The Chain Mediated Effect of Family Cohesion and Negative Emotion [master's thesis]. Harbin, China: Harbin Normal University; 2020.
74. Yang Y, Gao W, Fan C, Tao T, Wang L, Bai L. The current status and influencing factors of smartphone dependency among adolescents [Article in Chinese]. Mental Health Edu Prim Secondary Sch 2022:10-18 [FREE Full text]
75. Yang Y, Gao W, Tao T, Fan C, Wang L. Cross-lagged analysis of parental phubbing behavior, anxiety, depression, and mobile phone addiction of adolescents [Article in Chinese]. Chin J Health Psychol 2022:1-14 [FREE Full text] [doi: 10.13342/j.cnki.cjhp.2022.12.022]
76. Yang J. The Effect of Family Cumulative Risk on Network Dependence of Rural College Students [master's thesis]. Changchun, China: Northeast Normal University; 2022.
77. Yang M. The Relationship Between Stress Life Events and Smartphone Addiction in Junior High School Students: A Moderated Mediation Model [master's thesis]. Guangzhou, China: Guangzhou University; 2022.
78. Zhong D, Zhou Z, Qi D, Liu Q. Parental technoference and adolescent smartphone addiction: the mediating role of self-disgust and social sensitivity. Chin J Clin Psychol 2023;31(06):1433-1437. [doi: 10.16128/j.cnki.1005-3611.2023.06.027]
79. Zhu Y, Jiang Z. [Parents' phubbing and problematic internet use in junior high school students: chain mediation of parent-child cohesion and relatedness need satisfaction]. Chin J Clin Psychol 2022;30(02):434-8+87. [doi: 10.16128/j.cnki.1005-3611.2022.02.037]
80. Zhu M. The Effect of Parental Parents' Phubbing on Middle School Students' Mobile Phone Addiction: A Moderated Mediation Model and Intervention Study [master's thesis]. Shihezi, China: Shihezi University; 2023.
81. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006;11(2):193-206. [doi: 10.1037/1082-989X.11.2.193] [Medline: 16784338]
82. Huang C. Social network site use and Big Five personality traits: a meta-analysis. Comput Hum Behav 2019;97:280-290. [doi: 10.1016/j.chb.2019.03.009]
83. Lipsey MW, Wilson DB. Practical Meta-Analysis. Thousand Oaks, CA: Sage Publications, Inc; 2001.
84. Zhang J, Dong C, Jiang Y, Zhang Q, Li H, Li Y. Parental phubbing and child social-emotional adjustment: a meta-analysis of studies conducted in China. Psychol Res Behav Manag 2023;16:4267-4285 [FREE Full text] [doi: 10.2147/PRBM.S417718] [Medline: 37877136]

XSL•FO
RenderX

Abbreviations

ICC: intraclass correlation coefficient
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T de Azevedo Cardoso; submitted 23.02.24; peer-reviewed by T Soluoku, J Mistry; comments to author 01.04.24; revised version received 10.04.24; accepted 28.10.24; published 22.01.25.

Please cite as:
Zhang J, Zhang Q, Xiao B, Cao Y, Chen Y, Li Y
Parental Technoference and Child Problematic Media Use: Meta-Analysis
J Med Internet Res 2025;27:e57636
URL: https://www.jmir.org/2025/1/e57636
doi:10.2196/57636
PMID:

©Jinghui Zhang, Qing Zhang, Bowen Xiao, Yuxuan Cao, Yu Chen, Yan Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Health Solutions for Cardiovascular Disease Prevention: Systematic Review

Yihan Qi[1], MSocSci; Emma Mohamad[1,2], PhD; Arina Anis Azlan[1,2], PhD; Chenglin Zhang[1], MSocSci; Yilian Ma[1], MSocSci; Anqi Wu[1], MSocSci

[1]Centre for Research in Media and Communication, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[2]Universiti Kebangsaan Malaysia, Komunikasi Kesihatan (Healthcomm) - UKM Research Group, Selangor, Malaysia

Corresponding Author:
Emma Mohamad, PhD
Centre for Research in Media and Communication
Faculty of Social Sciences and Humanities
Universiti Kebangsaan Malaysia
UKM
Selangor, 43600
Malaysia
Phone: 60 389215457
Email: emmamohamad@ukm.edu.my

Abstract

Background: Cardiovascular disease (CVD) is a major global health issue, with approximately 70% of cases linked to modifiable risk factors. Digital health solutions offer potential for CVD prevention; yet, their effectiveness in covering the full range of prevention strategies is uncertain.

Objective: This study aimed to synthesize current literature on digital solutions for CVD prevention, identify the key components of effective digital interventions, and highlight critical research gaps to inform the development of sustainable strategies for CVD prevention.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we conducted a comprehensive search in Web of Science, Scopus, and PubMed to identify original English-language studies published between January 2000 and May 2024 that examined primary or secondary CVD prevention through digital solutions. The exclusion criteria included: telephone-only interventions, abstract-only publications, methodology-focused studies without primary data, studies without participants or specific groups, and studies with no follow-up period. The literature search used the string with terms like “digital health,” “mHealth,” “mobile health,” “text message,” “short message service,” “SMS,” “prevention,” “prevent,” “cardiovascular disease,” “CVD,” etc. Study bias was assessed using the RoB 2 (Cochrane Collaboration) and the ROBINS-I tool (Cochrane Collaboration). Data on prevention components, prevention types, study design, population, intervention, follow-up duration, personnel, and delivery settings were extracted.

Results: A total of 2871 studies were identified through the search. After excluding ineligible studies, 30 studies remained, including 24 randomized controlled trials. The reviewed digital solutions for CVD prevention focused on baseline assessment (29/30, 97%), physical activity counseling (18/30, 60%), tobacco cessation (14/30, 47%), blood pressure management (13/30, 43%), and medication adherence (10/30, 33%). The technologies used were categorized into 3 types, smartphones and wearables (16/30, 53%), email and SMS communications (12/30, 40%), and websites or web portals (3/30, 10%). The majority of the study outcomes addressed blood pressure (14/30, 47%), exercise capacity (12/30, 40%), weight (12/30, 40%), and lipid profile (11/30, 37%), while fewer focused on nicotine dependence (9/30, 30%), medication use (8/30, 27%), quality of life (7/30, 23%), dietary habits (5/30, 17%), intervention adherence (4/30, 13%), waist circumference (4/30, 13%), and blood glucose levels (2/30, 7%).

Conclusions: Digital solutions can address challenges in traditional CVD prevention by improving preventive behaviors and monitoring health indicators. However, most evaluated interventions have focused on medication use, quality of life, dietary habits, adherence, and waist circumference. Further studies are needed to assess the long-term impact of more comprehensive interventions on key cardiovascular outcomes.

(J Med Internet Res 2025;27:e64981) doi:10.2196/64981

XSL•FO
RenderX

KEYWORDS

cardiovascular disease prevention; cardiovascular outcomes; digital technologies; remote care; mobile phone

Introduction

Cardiovascular disease (CVD) remains a significant worldwide health issue, making a contribution to global mortality. Approximately 70% of CVD cases are due to modifiable risk factors [1], these include lifestyle-related elements such as physical inactivity, poor dietary habits, high levels of blood pressure, and tobacco use. CVD prevention encompasses three stages: primary, secondary, and tertiary prevention. Primary prevention targets high-risk individuals without CVD [2], secondary prevention focuses on those with established CVD, and tertiary prevention is for individuals seriously affected by CVD and aims to enhance their life expectancy [3]. Despite the evidence supporting the effectiveness of CVD prevention initiatives, there is a significant difference between the potential benefits and the actual participation of individuals. Participation rates in preventive strategies still remain low, and adherence to lifestyle changes and medical recommendations is still unstable [4,5]. This arises from challenges such as socioeconomic inequalities, differences in the timing of participation, limited awareness or understanding of preventive strategies, restricted access to health care resources, and barriers related to transportation and distance [6]. Furthermore, barriers in health care systems, such as insufficient funding, fragmented care delivery, and inadequate integration, further hinder the adoption of CVD prevention measures [3].

Recent technological advances in CVD prevention offer solutions to the limitations of traditional facility-based measures. Mobile apps, wearable devices, telemedicine, and remote monitoring systems can improve individual engagement and adherence. These approaches can provide personalized instructions, real-time monitoring, and remote consultations, making it easier to manage their cardiovascular health. Moreover, they can reach broader populations, including those with limited access to health care facilities. This review employs digital health solutions to incorporate a range of technologies to facilitate care delivery. This way acknowledges the variety of digital tools available, including eHealth, mobile health (mHealth), SMS, wearable devices, mobile apps, and telemedicine.

Current studies have shown that digital solutions, such as mobile apps, hold promise for CVD prevention, but their effectiveness has mainly been demonstrated in limited settings, and broad implementation in clinical practice remains rare [7,8]. In addition, there is limited understanding of how comprehensively these digital solutions address the key components of CVD prevention and achieve targeted health outcomes. While these innovations hold the potential to enhance CVD prevention by improving accessibility, efficiency, and individual engagement, significant barriers remain. Challenges such as regulatory constraints, interoperability issues, and systemic limitations must be addressed to facilitate their effective integration into routine healthcare practices.

This study aims to evaluate the current landscape of digital technologies for CVD prevention, focusing on the comprehensiveness and effectiveness of digital solutions. Specifically, we serve multiple purposes, first, it integrates existing literature on digital solutions for CVD prevention; second, it identifies key components of CVD prevention that are effectively addressed through digital solutions; and third, it charts the gaps that need attention to facilitate the sustainable integration of digital solutions for CVD prevention into clinical practice.

Methods

Study Design

A systematic review was conducted following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [9] (Multimedia Appendix 1) to ensure transparency and consistency in reporting. This review aimed to answer two key research questions, that are (1) “What types of digital technologies are utilized in CVD prevention studies, including their sample sizes, intervention durations, follow-up periods, and primary findings?” and (2) “How comprehensive and effective are these CVD prevention solutions?”

Eligibility Criteria

All studies published in English from January 2000 to May 2024 that examined digital solutions aimed at CVD prevention were considered for inclusion. Since Frank introduced the term “digital health” in the 2000s [10], it has transformed health care practices, bringing innovative approaches to CVD prevention worldwide. This review focused on the period from January 2000 to May 2024 (present) to capture the comprehensive applications of digital health technologies in CVD prevention, with an emphasis on primary and secondary prevention. In contrast, tertiary interventions, including coronary angioplasty, stenting, and bypass surgery, were excluded because their primary focus was on halting disease progression rather than prevention, which was beyond the objectives and scope of this study. In this review, the term “digital” refers to advanced technologies that facilitate remote, interactive, and personalized interventions. These technologies went beyond basic telephonic communication and included internet-based platforms, wearable devices that monitored and provided real-time physiological feedback, as well as mobile apps designed to track health behaviors, deliver educational content, and offer virtual guidance. As noted in the American College of Cardiology Scientific Statement on CVD Prevention [11], studies that only relied on telephone interventions were excluded. Specifically, studies were included if they (1) were original research using telemedicine or digital methods specifically targeting primary or secondary CVD prevention, (2) reported findings on feasibility and usability, and (3) were published in English. In contrast, studies were excluded if they (1) relied exclusively on telephone interventions, (2) were only available as abstracts without full-text access, (3) only focused on methodology

XSL•FO
RenderX

without primary data, (4) did not involve participants or specific groups, or (5) lacked a follow-up period, such as cross-sectional studies without outcome tracking over time.

Information Source

A comprehensive search was conducted for studies published between January 2000 and May 2024, using the Web of Science, Scopus, and PubMed databases for the selection process.

Search Strategy

The search strategy used the PICO (Population, Intervention, Comparison, Outcome)-based search string, combining keywords and abbreviations related to digital health care technologies, prevention strategies, and CVD. Terms included "digital health," "mHealth," "mobile health," "text message," "short message service," "SMS," "prevention," "prevent," "cardiovascular disease," "CVD," etc. A comprehensive list of search terms was provided in Multimedia Appendix 2. 2 researchers (YQ and CZ) independently screened and identified relevant studies. Any discrepancies were resolved through discussion, with additional input from other researchers (EM and AAA) as needed to reach a consensus.

Selection Process

To ensure reliability and remove duplicates, 2 researchers (YQ and CZ) independently imported the articles into EndNote 21 (Clarivate) and conducted the initial screening. Discrepancies in screening decisions were resolved by a third researcher (EM). Following title and abstract screening, 2 researchers (YQ and CZ) independently conducted a full-text review, with any discrepancies resolved through consultation with the third researcher (EM).

Data Collection, Data Items, and Data Synthesis

Data collection was conducted by 2 researchers (YQ and CZ), and any disagreements were resolved through consultation with a third researcher (EM). Our approach follows the American Heart Association consensus statement [12] and Guide to Primary Prevention of Cardiovascular Diseases [13] about the components of primary and secondary CVD prevention. In addition, we divided the digital solutions outlined in each study into two groups: (1) those that operated independently and (2) those that enhanced traditional approaches to CVD prevention. According to Wongvibulsin et al [14], studies were classified as standalone solutions if the intervention primarily involved remote teaching, with the initial face-to-face meeting used for onboarding, baseline assessment, or outcome evaluation, provided that the primary intervention was delivered remotely. Moreover, we observed other characteristics including CVD prevention types, study countries, study designs applied, participation and population, duration of intervention or follow-up, personnel used, and delivery settings. Due to the significant heterogeneity among the publications, a quantitative synthesis and meta-analysis were not feasible. As a result, this study emphasized a qualitative synthesis, and data collection was conducted using Microsoft Excel 2023.

Risk of Bias Assessment and Reporting Bias Assessment

The risk of bias for all included studies was independently assessed by 2 researchers (YQ and CZ), who also verified each other's findings. Any disagreements were resolved through consultation with a third researcher (EM). Methodological rigor was ensured for each study type by the Cochrane Collaboration's Risk of Bias (RoB) 2 tool [15] for randomized studies and the ROBINS-I tool [16] for nonrandomized studies. For randomized studies, we evaluated risks associated with the randomization process, deviations from intended interventions, missing outcome data, outcome measurement, and result selection. Each risk was classified as "low risk," "some concerns," or "high risk," based on standardized criteria and expert judgment. For nonrandomized studies, we examined risks related to confounding bias, selection bias, intervention classification bias, deviations from intended interventions, missing data bias, outcome measurement bias, and reporting bias. Each risk was rated as "low," "moderate," "serious," or "critical," based on standardized criteria and expert judgment.

Results

Study Screening

Following PRISMA guidelines, we conducted a comprehensive search on the Web of Science, Scopus, and PubMed, identifying a total of 2871 articles published in English. During the initial title and abstract screening, 1542 duplicate records were identified and removed, resulting in 1329 articles proceeding to the next stage. After further title and abstract review, 1068 articles were excluded due to their lack of relevance to telemedicine or digital solutions for primary or secondary CVD prevention, leaving 261 articles for full-text screening. Among these, 231 articles were subsequently excluded for reasons such as telephone-only interventions, abstract-only publications, no primary data, no participants, and no follow-up period. When articles met multiple exclusion criteria, they were classified under the first applicable category based on priority. Finally, 30 studies were included. A PRISMA flowchart of the screening process is provided in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. CVD: cardiovascular disease.

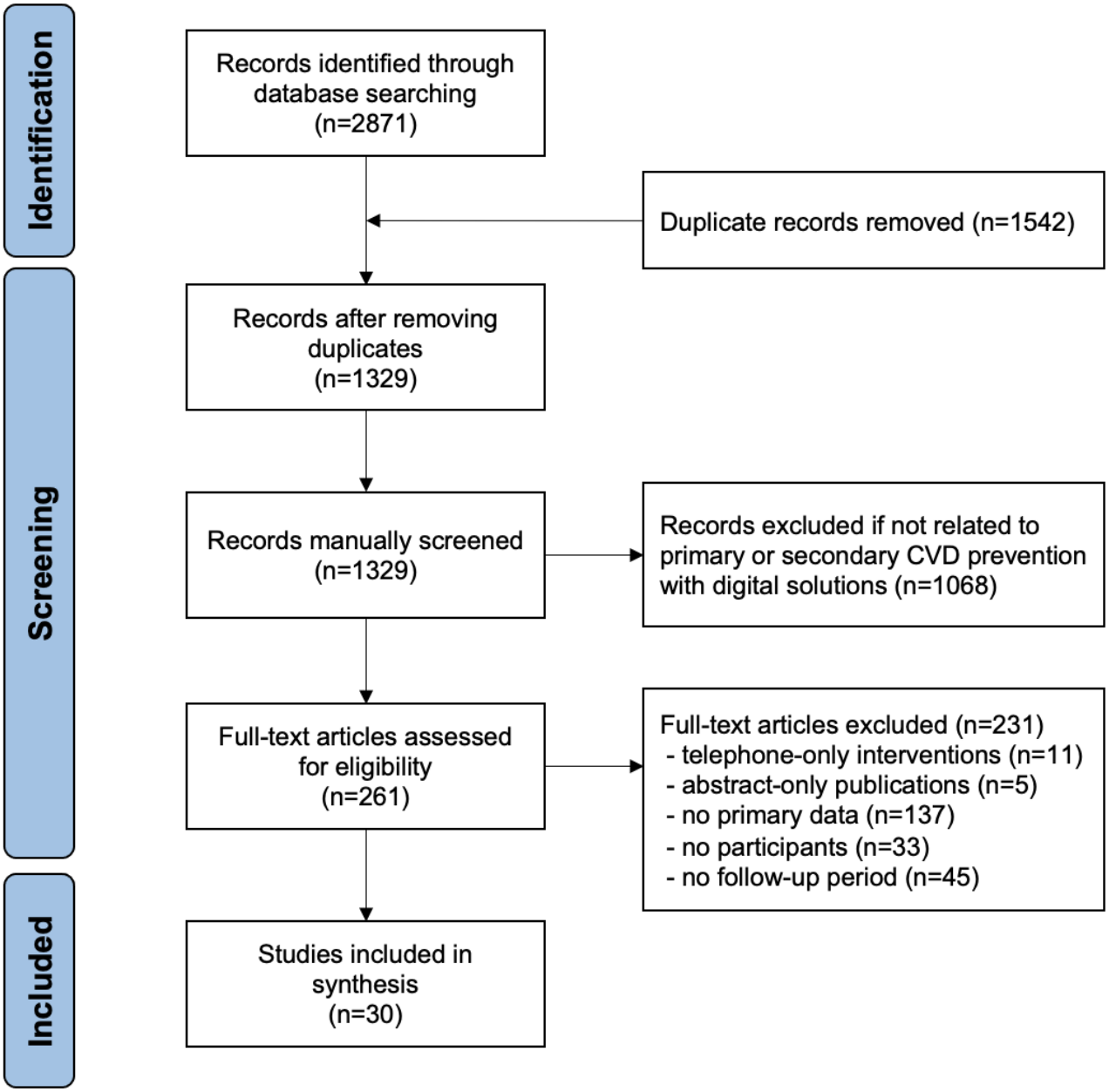

Characteristics of the Eligible Studies

This review included 30 studies meeting the inclusion criteria detailed in Multimedia Appendix 3, with their characteristics detailed in Table 1. Of the 30 studies, 24 (80%) were randomized controlled trials, 6 (20%) were nonrandomized, and 23 (77%) focused on secondary prevention. The majority were conducted in the Americas (9/30, 30%). Of the 30 studies, 16 (53%) studies used smartphones and wearables, 12 (40%) studies used email or SMS communications, and 3 (10%, 3/30) studies used websites or web portals. Interventions are most often directed by research team staff (26/30, 87%), followed by health coaches (3/30, 10%), physicians (3/30, 10%), general health professionals (3/30, 10%), nurses (2/30, 7%), dietitians (1/30, 3%), pharmacists (1/30, 3%), and community health workers (1/30, 3%). The median sample size was 465.5 (IQR 162.75-751.5), with a median follow-up period of 6 (IQR 3.5-12) months and a median intervention duration of 6 (IQR 3-12) months. In total, 28 out of 30 (93%) studies were standalone digital CVD prevention interventions. A bar chart was used to compare the number of studies that included each component of CVD prevention (Figure 2). The most common components of CVD prevention were baseline assessment (29/30, 97%), physical activity counseling (18/30, 60%), tobacco cessation (14/30, 47%), blood pressure management (13/30, 43%), and medication adherence (10/30, 33%). Approximately one-third of the studies covered additional components of CVD prevention, such as disease knowledge (9/30, 30%) and exercise training (9/30, 30%). Less than one-third of the studies focused on nutrition counseling (8/30, 27%), lipid (8/30, 27%), psychological management (7/30, 23%), diabetes (6/30, 20%), weight (6/30, 20%), heart rate (5/30, 17%), blood glucose (1/30, 3%), and alcohol use (1/30, 3%).

XSL•FO
RenderX

Table 1. Summary of the eligible studies about digital solutions for CVD prevention in the systematic review.

Category and subcategory	Studies, n
Research type	
Randomized controlled trial	24
Nonrandomized study	6
Prevention type	
Secondary	23
Primary	7
Location of study, by continent	
North and South America	9
Asia	8
Oceania	8
Europe	5
Publication year	
2015-2019	17
2020-2024	13
Technology use	
Smartphones and wearables	16
Email-SMS communications	12
Websites or web portals	3
Personnel	
Research team staffs	26
Health coaches	3
Physicians	3
General health professionals	3
Nurses	2
Dietitians	1
Pharmacists	1
Community health workers	1

XSL•FO
RenderX

Figure 2. Number of eligible studies with components of CVD prevention in the systematic review. CVD: cardiovascular disease.

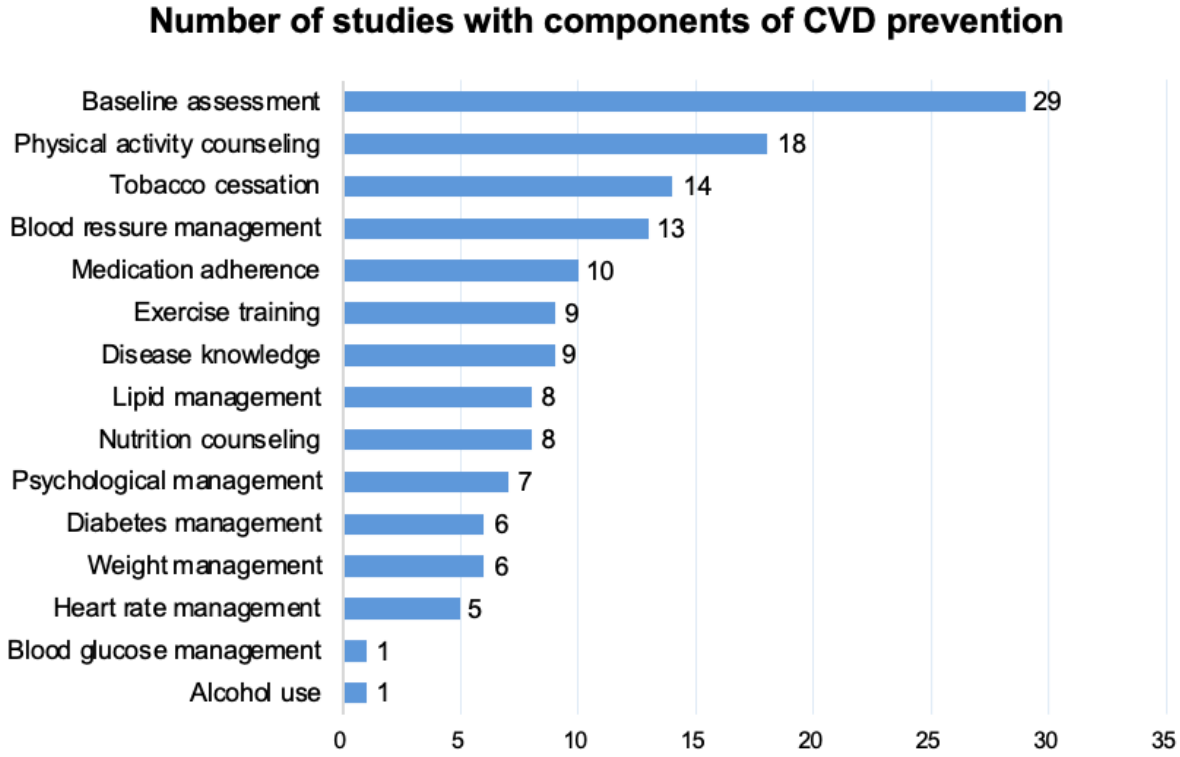

Bias Reporting of the Eligible Studies

In the assessment of bias among the 24 randomized controlled trials using the RoB 2 tool, 2 [17,18] studies were found to have a high risk of bias, 14 [19-32] had some concerns for risk of bias, and 8 [33-40] were rated as having a low risk of bias. Although all studies applied randomization, 9 [19,20,22,23,25,27,28,30,32] did not clearly specify the implementation details of the randomization process, and 2 [17,18] reported that participants were not blinded to their allocation due to unavoidable research limitations. For the 6 nonrandomized studies, evaluated with the ROBINS-I tool, 1 [41] study was judged to have a serious risk of bias, 4 [42-45] were assessed as having moderate risk, and 1 [46] was rated as low risk. 4 [41-44] of these studies presented concerns due to insufficient control of confounding factors in before-and-after designs, and 1 [41] study exhibited issues with missing data. Figures 3 [17-40] and 4 [41-46] summarized the results of the risk of bias assessments for the eligible studies using the RoB 2 and ROBINS-I tools.

XSL•FO
RenderX

Figure 3. Risk of bias summary using RoB tool. RoB: risk of bias [17-23,25-27,29,30,32-36,38,39,40,42,43,45,46].

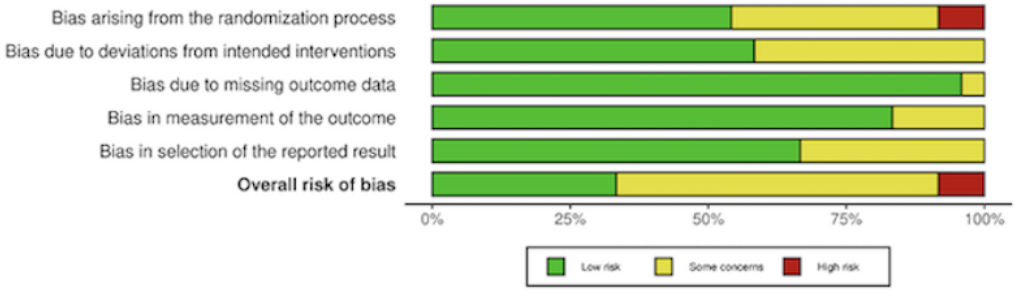

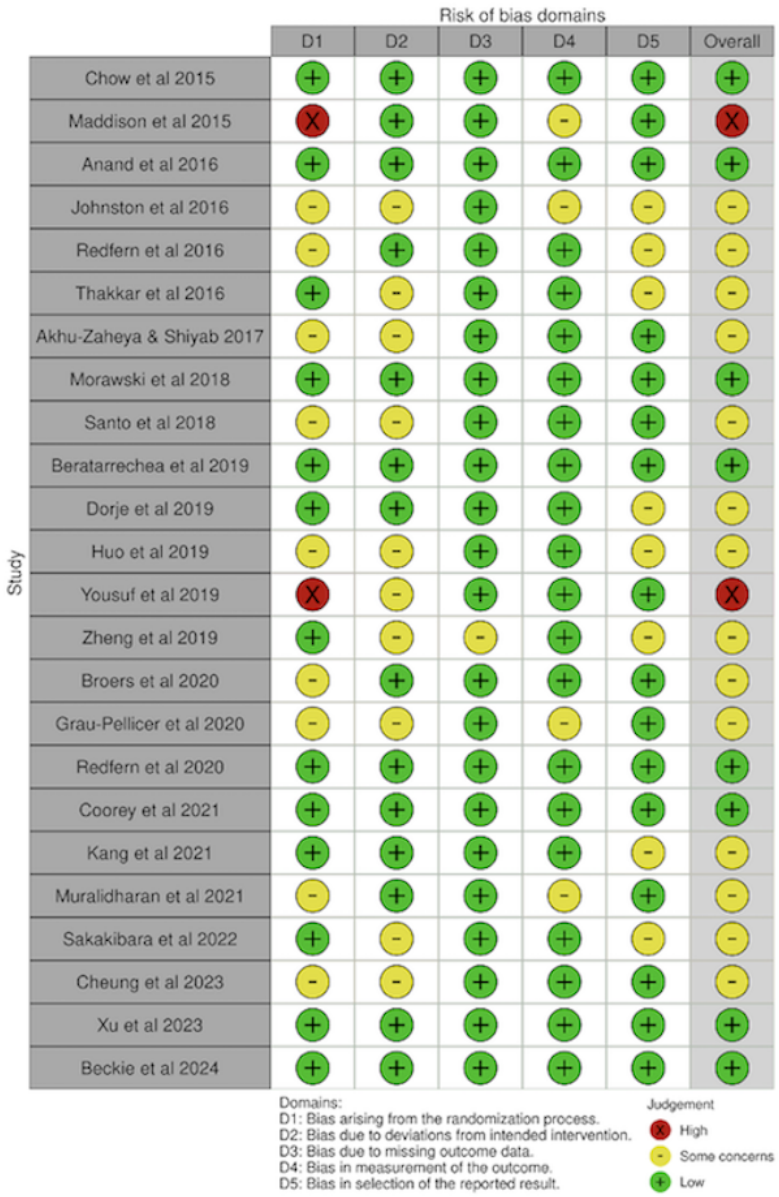

XSL•FO
RenderX

Figure 4. Risk of bias summary using the ROBINS-I tool [24,28,31,37,41,44].

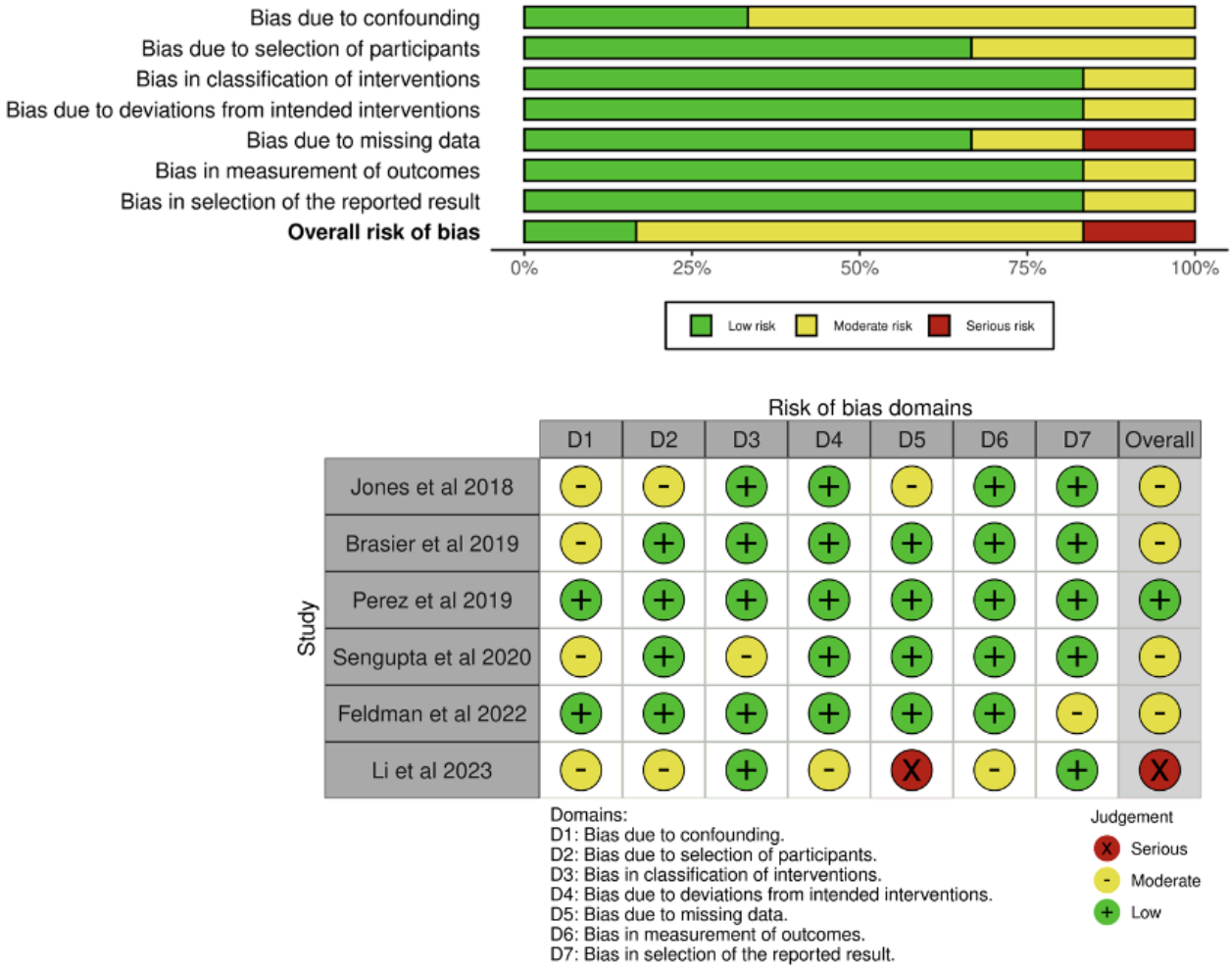

Technology Use in Digital Solutions

The analysis of 30 selected studies categorized the use of technology in digital health solutions for CVD prevention into 3 main categories: smartphones and wearables (53%, 16/30), email-SMS communications (40%, 12/30), and websites or web portals 10% (3/30). In addition, 3% (1/30) study incorporated both email-SMS communications and websites or web portals.

More than half of the selected studies integrated smartphones or mobile devices with wearable technology for monitoring and managing patients' health conditions. Notable researchers including Johnston et al [19], Morawski et al [35], Beratarrechea et al [36], Brasier et al [43], Dorje et al [24], Perez et al [46], Grau-Pellicer et al [28], Sengupta et al, [44], Redfern et al [37], Broers et al [27], Kang et al [29], Muralidharan et al [30], and Li et al [41] focused on the efficacy of smartphone applications in preventing CVD. For instance, Johnston et al [19] evaluated an interactive patient support tool through a smartphone application designed to enhance treatment adherence and cardiovascular health among patients with myocardial infarction. According to Grau-Pellicer et al [28], they evaluated the effectiveness of a mHealth application on physical activity adherence and suggested that this technology provides a way to promote adherence to home exercise programs poststroke. Sengupta et al [44], Feldman et al [45], and Beckie et al [40] examined interventions that use smartphones and smartwatches to enhance CVD prevention. Xu et al [39] explored the smartphone-based gamification intervention for CVD prevention. The use of email-SMS communications was also considerable, indicating the popularity of interventions such as the TEXT ME program [20,21,23,26,33]. In addition, the use of email or SMS programs [18,25,34], and SMS intervention [22,32,42], illustrated the various approaches employed in the field. There was relatively limited research on the use of websites or portals. Coorey et al [38] introduced a purpose-built, versatile web-based application intervention. Sakakibara et al [31] developed an attention control memory training program delivered through the web. Maddison et al [17] took a multifaceted approach, integrating email-SMS communications and websites or portals into their interventions.

Key Outcomes

Across the 30 studies, the most common key findings are summarized in Table 2. The most frequently evaluated outcome was blood pressure (14/30, 47%). Other commonly assessed outcomes included exercise capacity (12/30, 40%), weight (12/30, 40%), lipid profile (11/30, 37%), nicotine dependence (9/30, 30%), medication use (8/30, 27%), quality of life (7/30, 23%), dietary habits (5/30, 17%), intervention adherence (4/30, 13%), waist (4/30, 13%), and blood glucose (2/30, 7%). We have distinguished between findings related to enhancing

XSL•FO
RenderX

preventive behaviors and those related to monitoring health indicators for a more nuanced understanding of the results. For instance, blood pressure was categorized as a health indicator, whereas exercise capacity was categorized as a preventive behavior for CVD prevention. After the analysis of the key findings, it was found that individuals using digital solutions generally performed better when receiving interventions compared with individuals using traditional CVD prevention measures. For instance, due to reminders and personalized feedback, digital interventions promoted more consistent medication use [19,22,32,35,36,41]. Furthermore, as a result of the enhanced engagement and support provided by these technologies, the quality of life indicators of participants using digital solutions improved significantly [17,18,27,28,31]. Dietary habits had also improved substantially, with digital platforms often providing tailored nutrition advice and monitoring that was more effective than traditional measures [18,22-24,40]. Intervention adherence was notably higher, which could be attributed to the real-time feedback provided by digital tools [19,26,41,44]. Waist circumference also showed a slight reduction among those using digital interventions [30,40,42].

Table 2. Summary of key findings by thematic outcomes.

Key findings	Studies, n	Summary
Blood pressure (health indicator)	14	8 of the 14 studies reported that the digital solutions improved blood pressure management and were no worse than the control group [18,20,24,33,35,37,40,42]. In contrast, 6 studies concluded that it did not have a statistically significant influence on blood pressure management [25,26,29,30,32,34].
Exercise capacity (preventive behavior)	12	7 studies examined exercise capacity as outcomes showing that the performance of the intervention group was comparable to that of the control group [20,21,24,28,33,37,39]. Notably, in 5 of these studies, no significant difference in exercise capacity was observed [17-19,25,26].
Weight (preventive behavior)	12	7 of the 12 studies, digital solutions were effective in addressing weight management [18,20,23,33,39,42,44]. However, 5 studies found no significant change or a slight decrease in body weight or BMI after the digital intervention [19,24-26,37].
Lipid profile (health indicator)	11	4 of the 11 studies reported elevated high-density lipoprotein levels in the intervention group [18,23,24,37]. 3 other studies showed reductions in low-density lipoprotein and total cholesterol levels [20,33,42]. No significant changes were observed in the remaining 4 studies [25,26,29,32].
Nicotine dependence (health indicator)	9	Of the 9 studies included, 4 showed a positive trend towards improvement in smoking habits, as evidenced by a reduction in nicotine dependence scores [20,29,33,36], while the remaining 5 did not change significantly [19,22,24,26,37].
Medication use (preventive behavior)	8	6 studies found that medication use was beneficial and had a positive effect [19,22,32,35,36,41]. However, 2 studies reported slight or insignificant changes due to medication [37,45].
Quality of life (preventive behavior)	7	5 studies documented an improvement in quality of life for participants in the intervention group [17,18,27,28,31]. Conversely, 2 studies failed to observe significant differences in quality of life [19,24].
Dietary habits (preventive behavior)	5	Of the 5 studies reviewed, all reported improvements in dietary habits [18,22-24,40].
Intervention adherence (preventive behavior)	4	4 studies provided evidence to support the idea that digital solutions can improve individuals' adherence to interventions [19,26,41,44].
Waist (health indicator)	4	3 studies showed a slight reduction in waist circumference after the digital intervention [30,40,42]. However, there was also 1 study in which no significant effects were observed [34].
Blood glucose (health indicator)	2	1 study showed the effectiveness of interventions in managing blood glucose levels [25]. In contrast, 1 study found no significant benefit in this regard [18].

Notably, the digital solution showed efficacy comparable to the traditional control group across outcomes, including nicotine dependence, blood pressure, blood glucose levels, exercise capacity, and weight management. For example, digital interventions had the same impact on nicotine dependence as traditional interventions, with nearly half of the studies showing a positive trend toward improving smoking habits [20,29,33,36], and the remaining half showing no significant change [19,22,24,26,37]. Similarly, blood pressure, blood glucose levels, exercise capacity, and weight management results were comparable to those of the traditional control group. However, the effectiveness of digital solutions varies when it comes to lipid profiles. Of the 11 studies reviewed, 4 reported elevated high-density lipoprotein levels in the intervention group [18,23,24,37]. 3 other studies showed reductions in low-density lipoprotein and cholesterol levels [20,33,42]. No significant changes were observed in the remaining 4 studies [25,26,29,32].

Study outcomes varied across follow-up times. Dietary habits and intervention adherence yielded positive results at various intervals, including 3, 6, and 12 months [18,22-24,40], as well

XSL•FO
RenderX

as 3, 5, 6, and 8 months [19,26,41,44]. Waist circumference decreased at 3 months in studies by Beckie et al [40], Jones et al [42], and Muralidharan et al [30], but Anand et al [34] found no significant effects at 12 months. Quality of life was assessed at 6-month intervals [17,18,27,31] and at 2 months [28], with Johnston et al [19] and Dorje et al [24] finding no significant differences in the digital intervention versus usual care groups at 6 and 12 months. Medication uses improved health outcomes over 3, 5, 6, and 12 months [19,22,32,35,36,41], although Feldman et al [45] and Redfern et al [37] observed minimal changes over longer periods (12 and 43 months). Studies on nicotine dependence showed varied results, with some studies indicating positive trends over longer follow-ups [20,29,33,36], while others reported different outcomes [19,22,24,26,37]. Positive blood pressure outcomes were reported within 3 months by Beckie et al [40], Jones et al [42], and Morawski et al [35], with improvements observed at 6 months by Chow et al [33] and Yousuf et al [18]. Dorje et al [24] and Redfern et al [20,37] noted significant benefits at 12 months, whereas Anand et al [34], Kang et al [29], and Cheung et al [32] respectively found no significant impact at 12 and 6 months. Muralidharan et al [30], Huo et al [25], and Zheng et al [26] did not observe notable improvements over 3, 6, and 8 months. Few studies exist on blood glucose levels, with Huo et al [25] showing effectiveness at 6 months, but Yousuf et al [18] found no significant benefits during the same period. For exercise capacity, positive findings were reported at 6 and 12 months by Chow et al [33], Thakkar et al [21], Dorje et al [24], and Redfern et al [20,37], with shorter follow-ups also supporting these findings. However, no significant differences were found between digital intervention and usual care groups at 6 or 8 months [17-19,25,26]. For weight management, significant reductions were noted at 6 and 12 months by Chow et al [33] and Redfern et al [20], while Jones et al [42] and Sengupta et al [44] reported favorable outcomes at 3 months. However, minimal to no improvements were found by Dorje et al [24] and Huo et al [25] at 6 months. Johnston et al [19] and Redfern et al [37] also showed no significant changes at 6 and 12 months, respectively. Studies with 3 to 6-month follow-ups predominantly showed no significant changes in lipid profiles, with minimal impacts observed within 6 months [23,32,42]. Conversely, longer follow-ups tended to show more pronounced effects, such as significant increases in high-density lipoprotein reported by Dorje et al [24] and Redfern et al [37] and reductions in low-density lipoprotein and total cholesterol observed by Redfern et al [20] and Chow et al [33]. However, Huo et al [25] and Kang et al [29] found no significant lipid changes in long-term studies.

The review examined the comprehensive standalone interventions and found that the interventions in 7 studies identified as standalone included 5 components or more for CVD prevention, excluding baseline assessments [18,23,24,26,29,30,37]. Outcomes associated with most of the key outcomes associated with the prevention of CVD components were heterogeneous throughout the intervention. Among these 7 studies, Redfern et al [37] documented outcomes regarding medication use, showing no significant effects. Regarding blood glucose, Yousuf et al [18] did not find a significant benefit. With regard to waist circumference, Muralidharan et al [30] demonstrated a reduction in waist circumference after the digital intervention. Santo et al [23], Dorje et al [24], and Yousuf et al [18] observed improvements in mood within the intervention group, while Redfern et al [37] found similar results. Nevertheless, Kang et al [29] and Zheng et al [26] discovered that the digital solutions did not significantly impact lipid profiles. Regarding weight management, Santo et al [23], Yousuf et al [18], and Redfern et al [37] observed that digital solutions were effective. In contrast, Dorje et al [24] and Zheng et al [26] reported that there was no significant change, or even a slight decrease, in body weight or BMI after the digital intervention. Santo et al [23], Dorje et al [24], and Yousuf et al [18] all recorded improvements in dietary habits. Dorje et al [24], Yousuf et al [18], Redfern et al [37], and Muralidharan et al [30] reported that the digital solution significantly enhanced blood pressure management and was no worse than the control group. However, Kang et al [29] found that digital solutions did not significantly affect blood pressure management. In terms of exercise capacity, Dorje et al [24] and Redfern et al [37] showed that the intervention group was similar to the control group. Yousuf et al [18] and Zheng et al [26] discovered that there was no difference in exercise capacity between the digital intervention group and the control group. For nicotine dependence, Kang et al [29] showed a positive trend toward improving smoking habits, while Dorje et al [24], Zheng et al [26], and Redfern et al [37] showed no significant change. In terms of quality of life, Yousuf et al [18] reported enhanced quality of life, while Dorje et al [24] reported no significant difference in the quality of life of intervention groups. Overall, interventions taken and reported outcomes varied.

Discussion

The Role of Digital Solutions for CVD Prevention

The study highlights the potential of digital technologies to enhance health care delivery and expand CVD prevention by enhancing preventive behaviors and monitoring health indicators. These digital solutions can significantly improve individual outcomes and facilitate the wider implementation of preventive approaches. Unlike previous systematic reviews, which often focused narrowly on specific aspects of digital health, our study provides a comprehensive assessment of digital solutions, including primary and secondary CVD prevention. Our review demonstrates the varied effectiveness of these digital solutions. However, variations in potential risks of bias, as assessed, should be considered when interpreting the findings. Based on existing studies, our review supports remote CVD prevention through internet-based platforms and digital devices, which offer innovative ways to monitor, educate, and engage individuals. Our findings show that using digital solutions to prevent CVD is both feasible and effective, and the results are comparable to traditional approaches. Whether as a complement or substitute for traditional approaches, digital solutions show great potential in improving individual health outcomes [19,21-28,30,33,35,36,39-43]. Notably, the integration of digital technologies into CVD prevention can make more efficient use of health care resources, reduce the burden on health care systems, and provide cost-effective solutions for individuals [17,18].

XSL•FO
RenderX

The results of our review are consistent with previous studies, including Gray et al [47], which demonstrated the effectiveness of remote consultation, smartphone applications, wearables, remote monitoring, and predictive analytics in influencing individual behavior. These digital solutions can greatly contribute to the primary and secondary prevention of CVD and play a role in preventing and managing disease. In addition, Moshawrab et al [48] showed that wearable devices were highly accurate in detecting, predicting, and even treating CVD. Their study highlighted the potential of digital health solutions to improve individual outcomes and optimize the use of health care resources, further supporting their integration into standard CVD prevention strategies. Our review and the included studies highlight the role of effective communication strategies in the success of digital health solutions for CVD prevention. Effective communication influences user engagement, adherence, and the overall impact of these interventions. Specific strategies, such as using interactive features to boost engagement, providing reminders to support adherence, offering support through virtual coaching, tailoring experiences to individual needs, and implementing feedback mechanisms, can significantly motivate individuals to participate and commit to preventive behaviors. By fostering a supportive environment, effective communication strengthens relationships between health care providers and patients, empowering individuals to take an active role in managing their health and promoting health outcomes. However, the effectiveness of communication strategies in digital health solutions is hindered by barriers related to digital literacy and access. Many individuals, particularly those at higher risk of CVD, may lack the skills required to use digital tools effectively. Without sufficient digital literacy, the potential benefits of remote consultations, smartphone applications, wearables, and remote monitoring systems may remain inaccessible for some individuals. The digital divide, marked by disparities in digital skills and access, poses a significant obstacle to the equitable and effective implementation of digital health solutions [49]. To address these barriers, targeted strategies to improve digital literacy are essential. Initiatives may include specialized training programs designed to help users develop the skills necessary to engage effectively with digital health tools, with a focus on older adults and underserved populations who may face greater challenges. Furthermore, designing user-friendly interfaces and simplifying onboarding processes can facilitate initial engagement and reduce usability obstacles. Health care providers also play a role in offering technical support, instructions, and tutorials that guide users in understanding and integrating digital tools into their daily routines. Enhancing digital literacy and access is important to enabling effective communication, thereby amplifying the impact of digital health solutions for CVD prevention. By improving digital skills and accessibility, individuals can more effectively set baselines, customize treatment plans, monitor progress, and receive tailored support, which strengthens their capacity to engage with preventive behaviors and optimize health outcomes.

Limitations and Future Research Recommendations

Some limitations appeared in our review. First, although this review aimed to address the effects and factors surrounding specific phenomena, we did not use meta-analysis methods to summarize empirical evidence due to the heterogeneity of the studies. The reviewed articles exhibited significant variation in study designs, main outcomes, technologies, and control groups. For instance, some studies focused on a single digital health intervention, such as a mobile health application, while others looked at a combination or multiple uses of various digital technologies. In addition, when considering control groups, some studies used standard care in people with traditional CVD prevention, while others did not include control groups. Second, this review used RoB 2 and ROBINS-I tools to assess the risk of bias in the included studies. While both tools were widely recognized for evaluating bias, they might be influenced by subjective judgment, particularly in complex cases where interpretations of criteria can vary among evaluators. Third, our review was limited in scope and time frame. While we included studies published between 2000 and 2024, the earliest studies meeting our inclusion criteria were from 2015, likely because digital solutions for disease prevention represent a relatively new field that has emerged with recent technological advances. Furthermore, our search was limited to 3 databases and did not include reference list searches and gray literature. As this is a rapidly evolving field, studies published after the search date may have been missed. Fourth, our review only highlighted primary and secondary prevention of CVD. However, tertiary prevention, which involved advanced medical procedures and interventions aimed at managing and mitigating the long-term effects of diagnosed CVD, also played a role in the integrated management of CVD. Therefore, we suggested several strategies for future research. First, it is crucial that future studies standardize the objectives of interventions to ensure consistency in study design, outcome measures, and control groups. Specifically, researchers should clearly define whether their digital solutions are intended as comprehensive standalone programs or as complementary tools to traditional prevention approaches. Second, future research should also aim to standardize and provide detailed descriptions of the specific CVD prevention components targeted by the intervention. This includes specifying whether the intervention's goal is to enhance preventive behaviors, monitor key health indicators (eg, blood pressure and cholesterol), or address both aspects simultaneously. Third, it is essential to incorporate robust measures of user engagement and behavior modification in evaluating digital solutions. Researchers should use validated metrics such as compliance rates, user satisfaction surveys, and long-term health outcomes to assess the impact of digital interventions on user behavior and sustained engagement. Finally, exploring the cost-effectiveness of digital solutions for CVD prevention, including tertiary prevention, compared to traditional approaches is crucial. Future studies should assess not only direct costs but also factors such as scalability, maintenance, and initial setup costs.

We have identified several key aspects for future research in the field of digital interventions for CVD prevention. First, while digital solutions effectively address factors such as eating habits, interventions targeting other risk factors, including blood glucose control, are still lacking. Moreover, further evaluation is needed to assess the ongoing efficacy of these digital solutions, so studies using a larger population and implementing suitable control groups are needed [8,50]. Second, future

XSL•FO
RenderX

research should focus on integrating multifaceted interventions to address a wide range of CVD risk factors, including stress management. It is crucial to study the synergies of combining digital tools such as mobile apps, wearables, and telemedicine. Understanding the psychological and behavioral mechanisms behind user engagement, such as motivation, digital literacy, and personalization, is also critical to improving adherence and sustained behavior change [51]. Third, it is crucial to evaluate the incidence of adverse events among individuals undergoing these interventions compared to traditional methods [14]. Rigorous monitoring and reporting of adverse events will provide insight into the potential risks associated with these technologies, helping to refine protocols and ensure individual safety. Finally, exploring the potential of artificial intelligence (AI) to improve communication effectiveness by personalizing and tailoring interventions in real time could significantly improve user experience [52,53]. As an emerging field, AI has shown great potential in fields such as health care and education. In the field of CVD prevention, its ability to analyze large datasets, predict individual outcomes, and tailor interventions to individual health conditions holds great potential for strengthening prevention measures and optimizing patient care pathways [52-54]. As research in this field continues to advance, the integration of AI into clinical practice could lead to more effective, personalized, and accessible solutions to combat CVD worldwide.

Conclusions

Our study highlights the potential of digital technologies to alleviate challenges associated with traditional CVD prevention approaches by enhancing preventive behaviors and monitoring health indicators. The widespread implementation of digital solutions in CVD prevention is expected to have a significant impact, increasing accessibility, affordability, and cost-effectiveness, and improving individual outcomes beyond what can be achieved with traditional approaches. However, interventions evaluated focused primarily on medication use, quality of life, dietary habits, intervention adherence, and waist circumference, with limited attention to other components of CVD prevention. In addition, our study primarily assessed the technical aspects and comprehensiveness of digital procedures in the prevention of CVD. However, the complexities associated with the certification of CVD prevention programs are beyond the scope of this review. Future studies should aim to explore more comprehensive CVD prevention interventions to assess their long-term and sustained impact.

Acknowledgments

This study would not have been possible without the support of the Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia.

This study did not receive external funding.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

YQ handled conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, drafting of the original manuscript, and editing the manuscript. YM and AW contributed to conceptualization. CZ supported methodology and data curation. EE and AAA were responsible for conceptualization, methodology, supervision, and reviewing and editing the manuscript. No generative AI was used in any part of the manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 76 KB - jmir_v27i1e64981_app1.pdf]

Multimedia Appendix 2
Search Terms.
[DOCX File , 15 KB - jmir_v27i1e64981_app2.docx]

Multimedia Appendix 3
Characteristics of the eligible studies about digital solutions for CVD prevention in the systematic review.
[DOCX File , 34 KB - jmir_v27i1e64981_app3.docx]

References

XSL•FO
RenderX

1. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. The Lancet 2020;395(10226):795-808. [doi: 10.1016/s0140-6736(19)32008-2]
2. Slater K, Colyvas K, Taylor R, Collins CE, Hutchesson M. Primary and secondary cardiovascular disease prevention interventions targeting lifestyle risk factors in women: A systematic review and meta-analysis. Front Cardiovasc Med 2022;9:1010528 [FREE Full text] [doi: 10.3389/fcvm.2022.1010528] [Medline: 36439996]
3. Collier J, Kienzler H. Barriers to cardiovascular disease secondary prevention care in the West Bank, Palestine - a health professional perspective. Confl Health 2018;12:27 [FREE Full text] [doi: 10.1186/s13031-018-0165-x] [Medline: 30026794]
4. Lönn A, Kallings LV, Andersson G, Paulsson S, Wallin P, Eriksson JS, et al. Lifestyle-related habits and factors before and after cardiovascular diagnosis: a case control study among 2,548 Swedish individuals. Int J Behav Nutr Phys Act 2023;20(1):41 [FREE Full text] [doi: 10.1186/s12966-023-01446-w] [Medline: 37020317]
5. Rippe JM. Lifestyle strategies for risk factor reduction, prevention, and treatment of cardiovascular disease. Am J Lifestyle Med 2019;13(2):204-212 [FREE Full text] [doi: 10.1177/1559827618812395] [Medline: 30800027]
6. Aerts N, Anthierens S, Van Bogaert P, Peremans L, Bastiaens H. Prevention of cardiovascular diseases in community settings and primary health care: a pre-implementation contextual analysis using the consolidated framework for implementation research. Int J Environ Res Public Health 2022;19(14):8467 [FREE Full text] [doi: 10.3390/ijerph19148467] [Medline: 35886317]
7. Islam SMS, Maddison R. Digital health approaches for cardiovascular diseases prevention and management: lessons from preliminary studies. Mhealth 2021;7:41 [FREE Full text] [doi: 10.21037/mHealth-2020-6] [Medline: 34345618]
8. Santo K, Redfern J. Digital health innovations to improve cardiovascular disease care. Curr Atheroscler Rep 2020;22(12):71 [FREE Full text] [doi: 10.1007/s11883-020-00889-x] [Medline: 33009975]
9. PRISMA Statement. PRISMA statement. URL: https://www.prisma-statement.org [accessed 2024-06-01]
10. Frank SR. Digital health care--the convergence of health care and the Internet. J Ambul Care Manage 2000;23(2):8-17. [doi: 10.1097/00004479-200004000-00003] [Medline: 10848396]
11. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American association of cardiovascular and pulmonary rehabilitation, the American heart association, and the American college of cardiology. Circulation 2019;140(1):133-153. [doi: 10.1161/cir.0000000000000663]
12. Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, et al. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007;115(20):2675-2682. [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
13. Grundy SM, Balady GJ, Criqui MH, Fletcher G, Greenland P, Hiratzka LF, et al. Guide to primary prevention of cardiovascular diseases. Circulation 1997;95(9):2329-2331. [doi: 10.1161/01.CIR.95.9.2329]
14. Wongvibulsin S, Habeos EE, Huynh PP, Xun H, Shan R, Porosnicu Rodriguez KA, et al. Digital health interventions for cardiac rehabilitation: systematic literature review. J Med Internet Res 2021;23(2):e18773 [FREE Full text] [doi: 10.2196/18773] [Medline: 33555259]
15. Sterne JAC, Savovi J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. A Revised Tool for Assessing Risk of Bias in Randomised Trials. BMJ York. 2019. URL: https://doi.org/10.1136/bmj.l4898 [accessed 2024-11-13]
16. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
17. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
18. Yousuf H, Reintjens R, Slipszenko E, Blok S, Somsen GA, Tulevski II, et al. Effectiveness of web-based personalised e-Coaching lifestyle interventions. Neth Heart J 2019;27(1):24-29 [FREE Full text] [doi: 10.1007/s12471-018-1200-7] [Medline: 30488381]
19. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. Am Heart J 2016;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
20. Redfern J, Santo K, Coorey G, Thakkar J, Hackett M, Thiagalingam A, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLoS One 2016;11(10):e0163929 [FREE Full text] [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
21. Thakkar J, Redfern J, Thiagalingam A, Chow CK. Patterns, predictors and effects of texting intervention on physical activity in CHD - insights from the TEXT ME randomized clinical trial. Eur J Prev Cardiol 2016;23(17):1894-1902. [doi: 10.1177/2047487316664190] [Medline: 27512051]

XSL•FO
RenderX

22. Akhu-Zaheya LM, Shiyab WY. The effect of short message system (SMS) reminder on adherence to a healthy diet, medication, and cessation of smoking among adult patients with cardiovascular diseases. Int J Med Inform 2017;98:65-75. [doi: 10.1016/j.ijmedinf.2016.12.003] [Medline: 28034414]
23. Santo K, Hyun K, de Keizer L, Thiagalingam A, Hillis GS, Chalmers J, et al. The effects of a lifestyle-focused text-messaging intervention on adherence to dietary guideline recommendations in patients with coronary heart disease: an analysis of the TEXT ME study. Int J Behav Nutr Phys Act 2018;15(1):45 [FREE Full text] [doi: 10.1186/s12966-018-0677-1] [Medline: 29792202]
24. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
25. Huo X, Krumholz HM, Bai X, Spatz ES, Ding Q, Horak P, et al. Effects of mobile text messaging on glycemic control in patients with coronary heart disease and diabetes mellitus: a randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019;12(9):e005805. [doi: 10.1161/CIRCOUTCOMES.119.005805] [Medline: 31474119]
26. Zheng X, Spatz ES, Bai X, Huo X, Ding Q, Horak P, et al. Effect of text messaging on risk factor management in patients with coronary heart disease: the CHAT randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019;12(4):e005616. [doi: 10.1161/CIRCOUTCOMES.119.005616] [Medline: 30998400]
27. Broers ER, Widdershoven J, Denollet J, Lodder P, Kop WJ, Wetzels M, et al. Personalized eHealth program for life-style change: results from the "Do Cardiac Health Advanced New Generated Ecosystem (Do CHANGE 2)" randomized controlled trial. Psychosom Med 2020;82(4):409-419. [doi: 10.1097/psy.0000000000000802]
28. Grau-Pellicer M, Lalanza J, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Top Stroke Rehabil 2020;27(5):354-368. [doi: 10.1080/10749357.2019.1691816] [Medline: 31790639]
29. Kang SH, Baek H, Cho J, Kim S, Hwang H, Lee W, et al. Management of cardiovascular disease using an mHealth tool: a randomized clinical trial. NPJ Digit Med 2021;4(1):165 [FREE Full text] [doi: 10.1038/s41746-021-00535-z] [Medline: 34862449]
30. Muralidharan S, Ranjani H, Anjana RM, Gupta Y, Ambekar S, Koppikar V, et al. Change in cardiometabolic risk factors among asian Indian adults recruited in a mHealth-based diabetes prevention trial. Digit Health 2021;7:20552076211039032 [FREE Full text] [doi: 10.1177/20552076211039032] [Medline: 34567611]
31. Sakakibara BM, Lear SA, Barr SI, Goldsmith CH, Schneeberg A, Silverberg ND, et al. Telehealth coaching to improve self-management for secondary prevention after stroke: a randomized controlled trial of stroke coach. Int J Stroke 2022;17(4):455-464 [FREE Full text] [doi: 10.1177/17474930211017699] [Medline: 33949270]
32. Cheung NW, Redfern J, Thiagalingam A, Hng TM, Marschner S, Haider R, SupportMe Investigators. Effect of mobile phone text messaging self-management support for patients with diabetes or coronary heart disease in a chronic disease management program (SupportMe) on blood pressure: pragmatic randomized controlled trial. J Med Internet Res 2023;25:e38275 [FREE Full text] [doi: 10.2196/38275] [Medline: 37327024]
33. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
34. Anand SS, Samaan Z, Middleton C, Irvine J, Desai D, Schulze KM, South Asian Heart Risk Assessment Investigators. A digital health intervention to lower cardiovascular risk: a randomized clinical trial. JAMA Cardiol 2016;1(5):601-606. [doi: 10.1001/jamacardio.2016.1035] [Medline: 27438754]
35. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
36. Beratarrechea A, Abrahams‐Gessel S, Irazola V, Gutierrez L, Moyano D, Gaziano TA. Using mHealth tools to improve access and coverage of people with public health insurance and high cardiovascular disease risk in argentina: a pragmatic cluster randomized trial. JAHA 2019;8(8):e011799. [doi: 10.1161/jaha.118.011799]
37. Redfern J, Coorey G, Mulley J, Scaria A, Neubeck L, Hafiz N, et al. A digital health intervention for cardiovascular disease management in primary care (CONNECT) randomized controlled trial. NPJ Digit Med 2020;3(1):117 [FREE Full text] [doi: 10.1038/s41746-020-00325-z] [Medline: 32964140]
38. Coorey G, Peiris D, Scaria A, Mulley J, Neubeck L, Hafiz N, et al. An internet-based intervention for cardiovascular disease management integrated with primary care electronic health records: mixed methods evaluation of implementation fidelity and user engagement. J Med Internet Res 2021;23(4):e25333 [FREE Full text] [doi: 10.2196/25333] [Medline: 33900204]
39. Xu L, Tong Q, Zhang X, Yu T, Lian X, Yu T, et al. Smartphone-based gamification intervention to increase physical activity participation among patients with coronary heart disease: a randomized controlled trial. J Telemed Telecare 2024;30(9):1425-1436. [doi: 10.1177/1357633X221150943] [Medline: 36794484]
40. Beckie TM, Sengupta A, Dey AK, Dutta K, Ji M, Chellappan S. A mobile health behavior change intervention for women with coronary heart disease: a randomized controlied pilot study. J Cardiopulm Rehabil Prev 2024;44(1):40-48. [doi: 10.1097/hcr.0000000000000804]

XSL•FO
RenderX

41. Li DM, Lu XY, Yang PF, Zheng J, Hu HH, Zhou Y, et al. Coordinated patient care via mobile phone-based telemedicine in secondary stroke prevention: a propensity score-matched cohort study. J Nurs Care Qual 2023;38(3):E42-E49. [doi: 10.1097/NCQ.0000000000000693] [Medline: 36827597]
42. Jones AR, Moser DK, Hatcher J. Using text messages to promote health in african-Americans: #HeartHealthyandCancerFree. Ethn Health 2018;23(3):307-320. [doi: 10.1080/13557858.2016.1263289] [Medline: 27897049]
43. Brasier N, Raichle CJ, Dörr M, Becke A, Nohturfft V, Weber S, et al. Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). Europace 2019;21(1):41-47 [FREE Full text] [doi: 10.1093/europace/euy176] [Medline: 30085018]
44. Sengupta A, Beckie T, Dutta K, Dey A, Chellappan S. A mobile health intervention system for women with coronary heart disease: usability study. JMIR Form Res 2020;4(6):e16420 [FREE Full text] [doi: 10.2196/16420] [Medline: 32348270]
45. Feldman K, Duncan RG, Nguyen A, Cook-Wiens G, Elad Y, Nuckols T, et al. Will Apple devices' passive atrial fibrillation detection prevent strokes? Estimating the proportion of high-risk actionable patients with real-world user data. J Am Med Inform Assoc 2022;29(6):1040-1049 [FREE Full text] [doi: 10.1093/jamia/ocac009] [Medline: 35190832]
46. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381(20):1909-1917 [FREE Full text] [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
47. Gray R, Indraratna P, Lovell N, Ooi SY. Digital health technology in the prevention of heart failure and coronary artery disease. Cardiovasc Digit Health J 2022;3(6 Suppl):S9-S16 [FREE Full text] [doi: 10.1016/j.cvdhj.2022.09.002] [Medline: 36589760]
48. Moshawrab M, Adda M, Bouzouane A, Ibrahim H, Raad A. Smart wearables for the detection of cardiovascular diseases: a systematic literature review. Sensors (Basel) 2023;23(2):828 [FREE Full text] [doi: 10.3390/s23020828] [Medline: 36679626]
49. Campanozzi LL, Gibelli F, Bailo P, Nittari G, Sirignano A, Ricci G. The role of digital literacy in achieving health equity in the third millennium society: a literature review. Front Public Health 2023;11:1109323 [FREE Full text] [doi: 10.3389/fpubh.2023.1109323] [Medline: 36891330]
50. Zwack CC, Haghani M, Hollings M, Zhang L, Gauci S, Gallagher R, et al. The evolution of digital health technologies in cardiovascular disease research. NPJ Digit Med 2023;6(1):1 [FREE Full text] [doi: 10.1038/s41746-022-00734-2] [Medline: 36596833]
51. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
52. Haq IU, Haq I, Xu B. Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging. Cardiovasc Diagn Ther 2021;11(3):911-923 [FREE Full text] [doi: 10.21037/cdt.2020.03.09] [Medline: 34295713]
53. Mohsen F, Al-Saadi B, Abdi N, Khan S, Shah Z. Artificial intelligence-based methods for precision cardiovascular medicine. J Pers Med 2023;13(8):1268 [FREE Full text] [doi: 10.3390/jpm13081268] [Medline: 37623518]
54. Parsa S, Somani S, Dudum R, Jain SS, Rodriguez F. Artificial intelligence in cardiovascular disease prevention: is it ready for prime time? Curr Atheroscler Rep 2024;26(7):263-272. [doi: 10.1007/s11883-024-01210-w] [Medline: 38780665]

Abbreviations

AI: artificial intelligence
CVD: cardiovascular disease
mHealth: mobile health
PICO: Population, Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RoB: Risk of Bias

Edited by N Cahill; submitted 02.08.24; peer-reviewed by M Nomali, Z Ehtesham; comments to author 15.10.24; revised version received 18.11.24; accepted 30.11.24; published 23.01.25.

Please cite as:
Qi Y, Mohamad E, Azlan AA, Zhang C, Ma Y, Wu A
Digital Health Solutions for Cardiovascular Disease Prevention: Systematic Review
J Med Internet Res 2025;27:e64981
URL: https://www.jmir.org/2025/1/e64981
doi:10.2196/64981
PMID:39847411

XSL•FO
RenderX

©Yihan Qi, Emma Mohamad, Arina Anis Azlan, Chenglin Zhang, Yilian Ma, Anqi Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Technologies for Interoperable Internet of Medical Things Platforms to Manage Medical Emergencies in Home and Prehospital Care: Scoping Review

Mattias Seth[1], MSc; Hoor Jalo[1], MSc; Åsa Högstedt[2], MA; Otto Medin[3]; Bengt Arne Sjöqvist[1], PhD; Stefan Candefjord[1], PhD

[1]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

[2]Prehospen – Centre for Prehospital Research, Faculty of Caring Science, Work Life and Social Welfare, University of Borås, Borås, Sweden

[3]InterSystems Corp, Stockholm, Sweden

Corresponding Author:
Mattias Seth, MSc
Department of Electrical Engineering
Chalmers University of Technology
Hörsalsvägen 11
Gothenburg, 412 58
Sweden
Phone: 46 739728529
Email: mattias.seth@chalmers.se

Abstract

Background: The aging global population and the rising prevalence of chronic disease and multimorbidity have strained health care systems, driving the need for expanded health care resources. Transitioning to home-based care (HBC) may offer a sustainable solution, supported by technological innovations such as Internet of Medical Things (IoMT) platforms. However, the full potential of IoMT platforms to streamline health care delivery is often limited by interoperability challenges that hinder communication and pose risks to patient safety. Gaining more knowledge about addressing higher levels of interoperability issues is essential to unlock the full potential of IoMT platforms.

Objective: This scoping review aims to summarize best practices and technologies to overcome interoperability issues in IoMT platform development for prehospital care and HBC.

Methods: This review adheres to a protocol published in 2022. Our literature search followed a dual search strategy and was conducted up to August 2023 across 6 electronic databases: IEEE Xplore, PubMed, Scopus, ACM Digital Library, Sage Journals, and ScienceDirect. After the title, abstract, and full-text screening performed by 2 reviewers, 158 articles were selected for inclusion. To answer our 2 research questions, we used 2 models defined in the protocol: a 6-level interoperability model and a 5-level IoMT reference model. Data extraction and synthesis were conducted through thematic analysis using Dedoose. The findings, including commonly used technologies and standards, are presented through narrative descriptions and graphical representations.

Results: The primary technologies and standards reported for interoperable IoMT platforms in prehospital care and HBC included cloud computing (19/30, 63%), representational state transfer application programming interfaces (REST APIs; 17/30, 57%), Wi-Fi (17/30, 57%), gateways (15/30, 50%), and JSON (14/30, 47%). Message queuing telemetry transport (MQTT; 7/30, 23%) and WebSocket (7/30, 23%) were commonly used for real-time emergency alerts, while fog and edge computing were often combined with cloud computing for enhanced processing power and reduced latencies. By contrast, technologies associated with higher interoperability levels, such as blockchain (2/30, 7%), Kubernetes (3/30, 10%), and openEHR (2/30, 7%), were less frequently reported, indicating a focus on lower level of interoperability in most of the included studies (17/30, 57%).

Conclusions: IoMT platforms that support higher levels of interoperability have the potential to deliver personalized patient care, enhance overall patient experience, enable early disease detection, and minimize time delays. However, our findings highlight a prevailing emphasis on lower levels of interoperability within the IoMT research community. While blockchain, microservices, Docker, and openEHR are described as suitable solutions in the literature, these technologies seem to be seldom used in IoMT platforms for prehospital care and HBC. Recognizing the evident benefit of cross-domain interoperability, we advocate a stronger focus on collaborative initiatives and technologies to achieve higher levels of interoperability.

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.2196/40243

(J Med Internet Res 2025;27:e54470) doi:10.2196/54470

KEYWORDS

Internet of Medical Things; enabling technologies; standards; cross-domain interoperability; scoping review; technology; medical emergency; internet; prehospital care; gerontology; global population; chronic disease; multimorbidity; health care system; home-based care; innovation; digital health; health informatics; telehealth; artificial intelligence

Introduction

Background

The aging world population and an increased prevalence of chronic conditions and multimorbidity have placed greater pressure on health care systems and professionals [1]. In Europe, over 50 million people have at least 1 chronic disease [2]; and in the United States, it is estimated that the number of people aged 50 years and older with at least 1 chronic disease will increase by 100%, from 71 million in 2020 to 142 million by 2050 [3]. The incidence of life-threatening falls among older adults (aged ≥70 y) continues to rise globally [4]; together with other age-related medical emergencies, such as stroke [5], these events claim millions of lives each year [5,6]. The notable changes in global demographics, combined with the prevalence of frailty among older adults, underscore the importance of expanding the human resources available in the public health sector. Nevertheless, due to economic and occupational constraints, this may not be a feasible solution [7]. To alleviate the pressure on health care systems, more sustainable initiatives need to be implemented. One potential solution often discussed in the literature is the transition to home-based care (HBC) [8,9]. HBC covers a wide continuum of care and involves delivering increasingly complex health care services to individuals in their own residences, allowing them to maintain their independence as an alternative to relying on residential, long-term, or institutional nursing care [10,11]. This transition is often supported by older adults because the majority choose to remain in their own homes for as long as possible [12,13].

To support the transitioning to HBC, there is a growing demand for new technological innovations and collaborative initiatives [14]. This often involves modifications of peoples' homes and the use of medical equipment that not only facilitates long-term health monitoring but also enables the management of medical emergencies that demand rapid medical attention [15-17]. This is particularly important in the case of older adults (aged ≥65 y) because their inability to manually initiate an emergency alarm after a sudden deterioration in physical condition (eg, acute stroke or hip dislocation after a fall) may lead to long-term consequences and increased mortality [15]. Studies have shown that Internet of Medical Things (IoMT; Figure 1) platforms have the potential to support the transitioning to HBC by streamlining workflows, reducing costs and time delays, and improving patient well-being [18-20]. While previous research has demonstrated the usefulness of stand-alone platforms for the monitoring of chronic conditions and detection of abnormalities, there has been limited focus on achieving higher levels of interoperability within IoMT [21-24]. Although stand-alone platforms may be feasible in certain situations, the lack of interoperability with external systems such as electronic health records (EHRs), emergency medical services, and public safety answering points (PSAPs) may lead to data becoming trapped within silos, potentially resulting in delayed information transfer, suboptimal decision-making, and patient harm [7,21,25,26]; for example, a study conducted by Magrabi et al [27] showed that approximately 20% of reported patient safety hazards were linked to deficient information transfer. Furthermore, Hyvämäki et al [26] showed that inadequate documentation and use of information in HBC plays a significant role in interorganizational health information exchange–related incidents, resulting in, for example, delayed care and patient harm.

XSL•FO
RenderX

Figure 1. In this study, the Internet of Medical Things (IoMT) was defined as a network of interconnected medical devices and health care systems that use the internet to collect, transmit, and exchange health care data. The figure illustrates device-to-device (thin lines), device-to-system (thin lines), and system-to-system (thick lines) communication, with 2 IoMT platforms (A and B) interacting with each other.

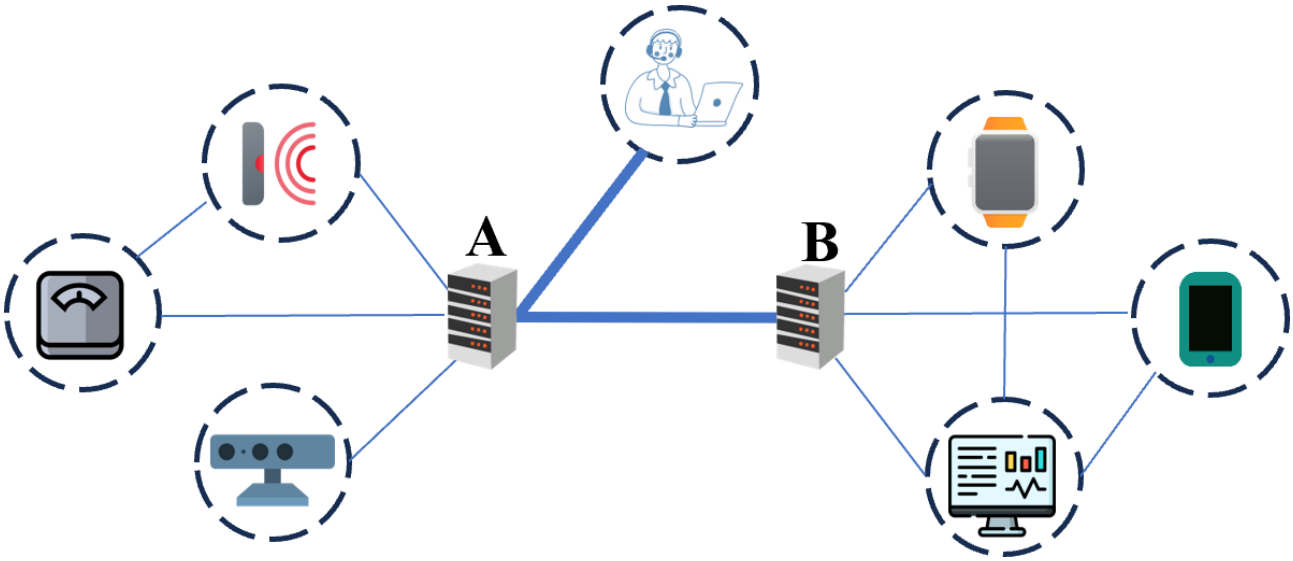

At Chalmers University of Technology in Gothenburg, Sweden, the Care@Distance research group is active in the field of remote and prehospital digital health. The group is dedicated to improving health care delivery through interoperable cutting-edge solutions, encompassing clinical decision support systems, artificial intelligence (AI), modern IT, and innovative user interactions. While interoperability offers numerous advantages, we acknowledge that current medical data still contain nonstandard elements [28,29], and vendors continue to use their own proprietary solutions [29]; for example, most inpatient EHRs include over 5000 variables [28,29], making it difficult to ensure a shared understanding of medical concepts across domains and organizations [29]. One system may code "A fall on and from stairs and steps" as W10 in the *International Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*), while another system uses Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) code 900000000000448009 for "Fall on stairs." Establishing a shared understanding of these 2 terms requires a translation [21]. As the number of unique systems and standards increases, the need for separate translations also increases, especially considering that SNOMED CT includes over 340,000 medical concepts (clinical findings, procedures, substances, etc) [29]. Due to these complex translation processes within health care, involving various file formats (eg, text, video, images, and audio), communication protocols, and semantics, over 80% of all medical data tend to be overlooked or discarded [19]. This not only hampers communication between systems but also limits the use of AI, international cooperation, and research [29]. As the technological landscape expands, navigating among the available technologies and standards is becoming increasingly challenging. Hence, in this study, we summarize existing knowledge and best practices to overcome interoperability issues in IoMT platform development to manage medical emergencies within HBC and prehospital care settings.

Objectives

This scoping review aims to summarize and map the enabling technologies that can be used to develop interoperable platforms to manage medical emergencies in HBC and prehospital care. We have proceeded from a 6-level interoperability model comprising device, network, syntactical, semantic, cross-platform, and cross-domain interoperability and a 5-level IoMT reference model comprising perception, transport, processing, application, and business layers [11]. These models provide adequate granularity and context to be applied within the context of IoMT [10]. Relevant actors and domains within the IoMT domain include homes, sensor providers, emergency medical services, PSAPs, social security services, and hospitals. The aim is to describe technologies and their use in an accessible way, enabling cross-disciplinary discussion between clinicians and engineers. This scoping review can potentially serve as a guide for software developers, clinicians, and other practitioners aiming to develop interoperable IoMT platforms.

Methods

Overview

This scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines for scoping reviews (refer to Multimedia Appendix 1 for the PRISMA-ScR checklist) [30] and adheres to the methodology outlined in the research protocol [11]. Given the broad spectrum of research questions addressed in this scoping review, the methodology includes 4 distinct search strategies, referred to as strategies A to D in the research protocol [11]. The relationships between these search strategies are illustrated in Figure 2.

This approach allows each research question to be addressed more systematically by separating the search terms, search periods, and goals within each search strategy. Search strategies A and B were completed with the research protocol [11] (Figure 2) and addressed the following research questions:

1. What are the current challenges of developing a real-time IoMT platform for managing medical emergencies such as falls?

XSL•FO
RenderX

2. What is interoperability? How can it be defined in the context of IoMT?
3. What types of models are used to visualize the different layers of interoperability? When talking about medical devices in an IoMT setting, which model is preferable and why?
4. Which reference model with corresponding protocols can best describe and define the structure of key aspects of the information being managed in a real-time IoMT system? How is the model being used today?

These 4 research questions were addressed in the research protocol [11], resulting in the definition of a 6-level interoperability model and an IoMT reference model to be used as reference materials in this scoping review. On the basis of these definitions, this scoping review will proceed with search strategies C and D (Figure 2), focusing on answering the following research questions:

1. Have any studies examined which current technologies are associated with the layers in the IoMT reference model, comprising device, network, syntactical, semantic, cross-platform, and cross-domain interoperability, and how these are being used to fulfill the set of rules defined by each layer? If so, what are the results?
2. How can interoperability solutions be mapped to the layers in the interoperability model?
3. What recommendations regarding enabling technologies can be given to clinicians and practitioners who want to develop IoMT platforms that can aggregate, store, and process data from relevant actors in prehospital care and HBC settings?

Research questions 1 and 2 are addressed in search strategy C, and research question 3 is addressed in search strategy D in this review.

Figure 2. This scoping review used a 4-strategy approach to streamline the search process. Strategies A and B were conducted as part of the research protocol [11], while strategies C and D are addressed in this scoping review. HBC: home-based care; IoMT: Internet of Medical Things.

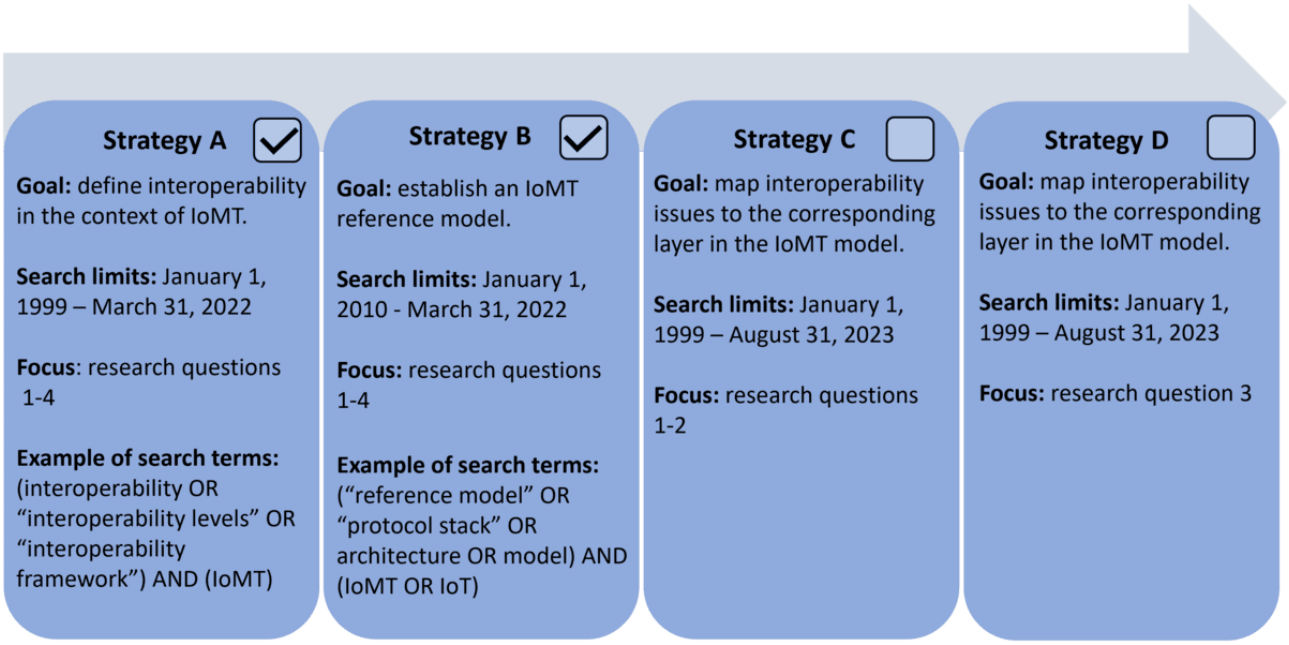

Search and Screening Process

Six electronic databases were used in this study: IEEE Xplore, PubMed, Scopus, ACM Digital Library, Sage Journals, and ScienceDirect [11]. In the published protocol, we had proposed using the Google Scholar database, but we replaced it with the ACM Digital Library in this review to ensure higher precision in search results [31,32] (refer to Multimedia Appendix 2 for the search terms used for search strategies C and D). The retrieved articles were assessed for inclusion based on predefined eligibility criteria and underwent a 2-step screening process using the web application Rayyan (Rayyan Systems Inc) [33]. The screening was performed by HJ and MS, and any disagreements were resolved through discussion between them [11].

Eligibility Criteria

The eligibility criteria were originally defined in the research protocol [11]. However, a minor adjustment was necessary to streamline the screening process and enhance clarity (Textbox 1). The last inclusion criterion was updated to emphasize a stronger focus on technologies that address interoperability issues.

XSL•FO
RenderX

Textbox 1. Eligibility criteria.

Inclusion criteria

- Published peer-reviewed journals and conference papers
- Written in English
- Published between January 1,1999, and August 31, 2023
- Studies describing or reporting the development or design of Internet of Medical Things (IoMT) systems with a focus on technology
- Studies reporting challenges and barriers to integrating IoMT platforms into prehospital care or home-based care settings with a focus on technology
- Studies describing the enabling technologies that can be used to solve interoperability issues in IoMT platform development

Exclusion criteria

- Full-text articles that were unavailable or not written in English
- Conference abstracts, book reviews, commentaries, and editorial articles
- Studies focusing on hardware, project management processes, or regulatory compliance
- Studies reporting on the design or development of IoT applications with no focus on health data (eg, Industry 4.0, including the automotive, food, and manufacturing industries)
- Studies describing the design or development of machine learning methods to achieve interoperability

Data Extraction

In search strategy C, data were extracted to map interoperability solutions to their respective interoperability models, categorizing them based on the specific level of interoperability addressed. The mapping was based on the 6-level interoperability model and the 5-level IoMT reference model established in the research protocol [11]. Similar mappings that could be found in the literature acted as a reference and, combined with the expertise of our research group, were used to validate the mappings.

In search strategy D, previous efforts in IoMT platform development for prehospital care and HBC were examined. A thematic analysis approach was used to identify and summarize the technologies and standards used in these development processes (Textbox 2). This analysis was performed by 1 reviewer (MS) using Microsoft Excel (version 2403) and the web application Dedoose (version 9.0.107; SocioCultural Research Consultants LLC). The thematic analysis was based on the framework outlined by Braun and Clarke [34]. The framework was adapted to this study and included five steps:

1. Reading through the literature and recording valuable aspects of the data, including the technologies and standards used
2. Organizing data into meaningful groups and creating codes relevant to the research questions (each technology was assigned a code, represented by a short descriptive summary of its intended use; Textbox 2)
3. Collating codes and assigning different keywords to them (each code could be assigned multiple keywords; the keywords were used to facilitate thematic analysis, aiming to conceptualize the functionality of each level in the interoperability model)
4. Assigning a nonoverlapping theme to each technology based on the codes and keywords (technologies with similar codes and keywords were assigned the same theme; a theme was defined as one of the following interoperability levels: device, network, syntactical, semantic, cross-platform, or cross-domain interoperability)
5. Compiling the thematic insights into a coherent review

Textbox 2. An example of the thematic analysis process. The applicability of the mapping was confirmed by our research group through tests with multiple articles [35].

Blockchain technology mapped to cross-domain interoperability

- Technology: blockchain
- Description in literature: “integration of blockchain into healthcare applications, including all aspects of privacy, validity of safety, and access to patient and electronic health records”
- Code assigned: blockchain can be used to preserve privacy and integrity of health data
- Keywords: security, data privacy, and integrity
- Assigned theme: cross-domain interoperability

From the thematic analysis, a list of commonly used technologies was compiled. Each technology was accompanied by a descriptive overview of its application area and function, along with a mapping to one of the levels in the interoperability model defined in the research protocol [11]. For each study, the main areas of interest were identified, such as citation details (eg, author, year of publication, and country of origin) and key study characteristics (eg, type of IoMT platform, platform

XSL•FO
RenderX

requirements, interoperability challenges, and the technologies and standards used to achieve different levels of interoperability). The results were summarized using narrative descriptions as well as figures and tables. The data extraction process ensured that sufficient data were gathered to answer each research question.

Results

Study Selection

The study selection flowchart is presented in Figure 3. A total of 2949 articles were identified (n=2306, 78.2% retrieved through search strategy C and n=643, 21.8% retrieved through search strategy D). After screening and snowballing, 158 articles were included in this study (n=84, 53.2% from strategy C, n=29, 18.4% from strategy D, and n=45, 28.5% from snowballing). Of the 33 articles excluded after title and abstract screening, 7 (21%) were left out because they were in non-English languages.

Figure 3. Study selection flowchart.

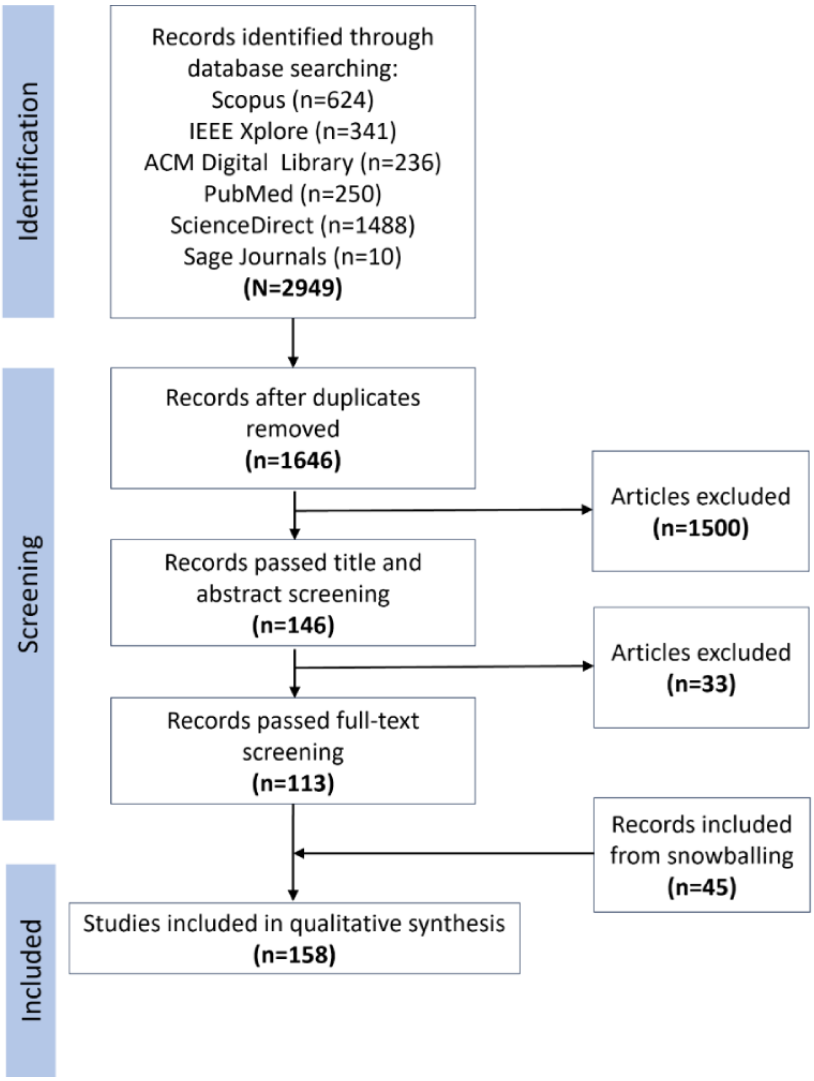

Subsections

This section is divided into 5 subsections (Figure 4). In the first, we provide an overview of the technologies and standards applicable to IoMT, emphasizing their relevance in addressing various levels of interoperability, as introduced in the research protocol [11]. In the second subsection, we present common interoperability requirements identified in the included studies, focusing on their application in prehospital care and HBC scenarios. Next, in the third subsection, we explore the interoperability challenges associated with IoMT platform development. In the fourth subsection, we summarize previous efforts in IoMT platform development and the technologies and standards used in these initiatives. Finally, in the fifth subsection, we draw upon our research findings to present actionable insights and recommendations for addressing interoperability challenges.

XSL•FO
RenderX

Figure 4. The main Results section is divided into 5 parts: overview (a summary of the enabling technologies that address various levels of interoperability), requirements (mapping Internet of Medical Things [IoMT] platform needs to the enabling technologies), challenges (presenting common interoperability challenges), strategies (outlining strategies to overcome these challenges), and recommendations (offering our research group's suggestions for addressing interoperability issues in IoMT platform development).

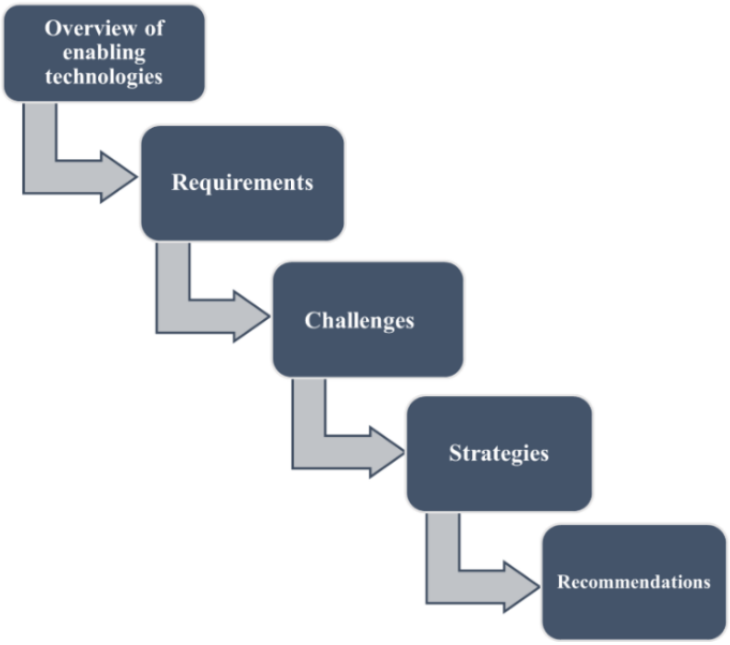

An Overview of the Enabling Technologies That Address Interoperability Issues Within IoMT

In the following subsections, we have categorized and summarized the enabling technologies based on the specific interoperability level they address (refer to Multimedia Appendix 2 for a description of the enabling technologies).

Device Interoperability

Device interoperability is the foundational level of interoperability [36] and involves ensuring that heterogeneous devices can physically connect and communicate in a restricted network, typically a personal area network (PAN) [36]. Solving issues at this level is often the first step in IoMT platform development processes because interoperable sensors are a prerequisite for subsequent analysis and decision-making processes. However, the widespread use of vendors' proprietary solutions often hinders the achievement of this level of interoperability [37]. Consequently, patients receiving hypertension management services from a particular company, for example, are often required to use a blood pressure device provided by the same company [38].

A variety of standards and protocols, including the Institute of Electrical and Electronics Engineers (IEEE) 11073 personal health device (PHD) [38-40], Zigbee [41], and Bluetooth Low Energy (BLE) [42], play a crucial role in addressing interoperability among medical devices, such as weighing scales, blood pressure monitors, and blood glucose monitors [39,43]. These technologies and devices can often be mapped to the perception layer in the IoMT reference model [44], a layer that is responsible for gathering raw data [45].

Network Interoperability

Network interoperability focuses on information exchange over the internet, including the ability of different networks or devices on separate networks to communicate [46]. Typically, network interoperability involves extended communication compared to device interoperability, which includes communication within local area networks (LANs) or wide area networks (WANs).

Technologies that focus on solving issues at the network interoperability level can primarily be mapped to the transport or processing layer in the IoMT reference model and include IP [47], user datagram protocol [48], transmission control protocol, IPv6 over low-power wireless PAN (6LoWPAN) [48,49], software-defined networking [36,50], gateways [51-53], message queuing telemetry transport (MQTT) [54], and WebSocket [55].

Syntactic interoperability

Syntactic interoperability, the third level of interoperability, involves data formats and data structures [36]. In the health care sector, both unstructured (eg, images, audio, and video streams) and structured data are used, which means that technologies used to address syntactic interoperability must be able to process diverse data types. Without syntactic interoperability, data might be sent to a system that is unable to process and use the information [56].

Technologies that focus on syntactic interoperability issues can be mapped to the processing layer in the IoMT reference model and include JSON [57], Health Level 7 version 2 (HL7v2) [58], and XML [59].

Semantic Interoperability

Semantic interoperability refers to the ability of different computer systems to have a common understanding of message contents, enabling them to share data with unambiguous, shared meaning [25]. This level of interoperability is essential for enabling automatic data processing and decision-making in IoMT settings [60]. Shared semantics within health care can

XSL•FO
RenderX

help to avoid knowledge mismanagement, clinical misinterpretation, misdiagnosis of a patient's illness, and even patient deaths [61]; for example, a system that receives "123" as input from another system (Figure 5) cannot interpret the data without additional information. For the receiving system to process and use the received data, the value "123" must be complemented by relevant metadata tags, such as "systolic blood pressure" or "patient ID." To achieve this contextual enrichment, well-established standards such as RxNorm [62], openEHR [63], *ICD* [37], Logical Observation Identifiers Names and Codes (LOINC) [62,64], and SNOMED CT [62] can be used. These standards offer a structured framework for associating informative labels and classifications with data [62].

Figure 5. An example of the lack of semantic interoperability. If a system receives insufficient data, it cannot interpret the information correctly. In the absence of semantic interoperability between the sending and receiving systems and without additional context or information, the receiving system cannot interpret the value "123".

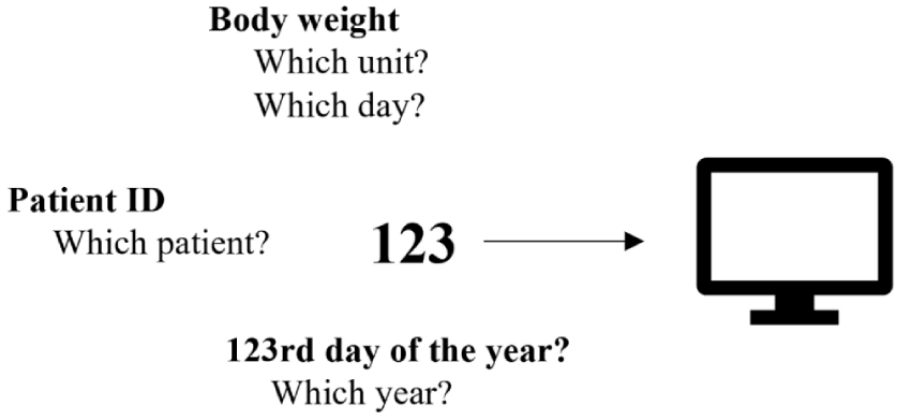

Cross-Platform Interoperability

Cross-platform interoperability denotes the ability of different platforms within a single domain to work together seamlessly. It could be various systems (eg, PSAPs, EHRs, mobile apps, and laboratory information systems) running on different platforms (web, mobile, desktop, etc) located in different hospital wards or hospitals. Cross-platform interoperability challenges emerge at this level primarily due to the presence of a wide range of operating systems (Windows, Android, Linux, iOS, etc), programming languages, data structures, architectures, and access methods for both entities and data (eg, application programming interfaces [APIs]) [36]. Technologies that focus on addressing these levels of interoperability issues can be mapped to the application layer in the IoMT reference model and include Fast Healthcare Interoperability Resources (FHIR) [65-68], representational state transfer APIs (REST APIs) [66], microservices [7,69], Docker [70], Kubernetes [70], and cloud services [67,69-72].

Cross-Domain Interoperability

Patients are likely to receive medical attention from several institutions and across various domains over their lifetime. Hence, communicating vital information across organizational and national boundaries is essential to ensure proper patient care and treatment [29]. This level of communication is enabled by cross-domain interoperability, which represents the highest level of interoperability [36].

In this context, a domain refers to a sociotechnical system defined by shared objectives and interests. These domains, or systems, are often separable from other systems by social, technical, and legal boundaries (Figures 6 and 7). Examples of domains include a medical device manufacturing company or a hospital. Each domain might have distinct goals, processes, security policies, and terminologies. Technologies and standards such as BioPortal [73], blockchain [74], ontology mediation [13,75], Web Ontology Language (OWL) [13], ontologies [25,74,76], General Data Protection Regulation (GDPR) [77,78], and Health Insurance Portability and Accountability Act (HIPAA) [78,79] can help achieve cross-domain interoperability.

XSL•FO
RenderX

Figure 6. Ontologies define domain knowledge, and semantics provide data meaning. Semantics can tag data (eg, "systolic"), while ontologies provide context, such as relating "123" to "blood pressure" and defining "elevated systolic level" within the domain.

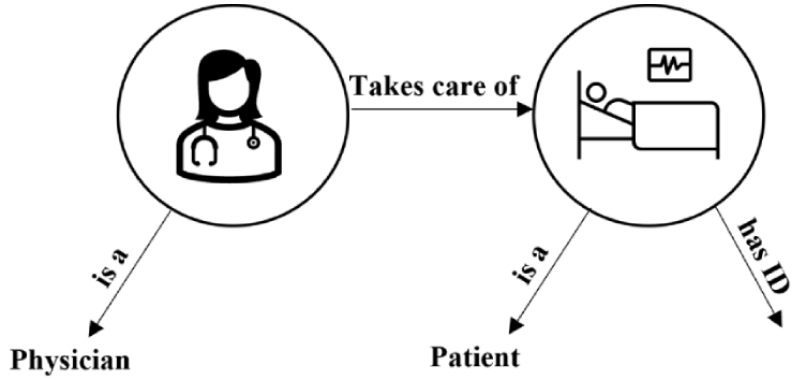

Figure 7. Technologies and standards such as blockchain, the General Data Protection Regulation (GDPR), and ontologies can enable data sharing between 2 separate domains.

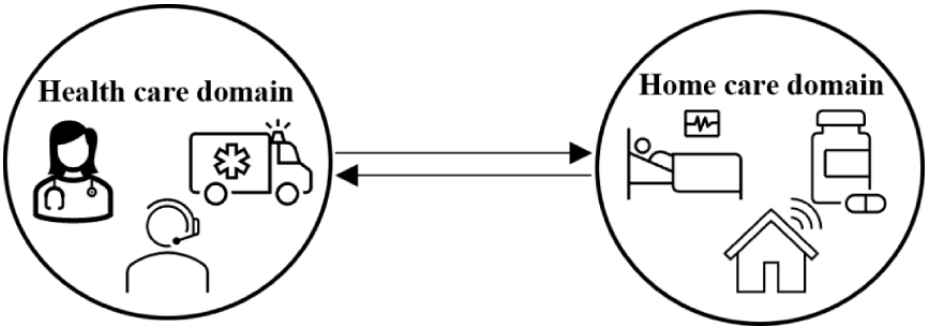

Common IoMT Interoperability Requirements for Prehospital Care and HBC

IoMT platforms for prehospital care and HBC have specific requirements. These platforms are required to operate in time-critical environments, where large amounts of sensitive data need to be exchanged in real time across various domains. In this subsection, we have summarized common interoperability requirements identified in the included studies and grouped them based on their potential application in IoMT settings (Table 1).

XSL•FO
RenderX

Table 1. The enabling technologies for interoperability and their application in Internet of Medical Things (IoMT) for prehospital care and home-based care. Suitable technologies are mapped to various IoMT platform requirements and potential use, with supporting references provided in the last column.

IoMT platform requirement	Potential use	Suitable technologies	References
Real-time text communication	Alerts, notifications, chats, and data visualization	MQTT[a], WebSocket, and webhooks	[19,54,80-88]
Real-time videoconferences or audioconferences	Video and voice calls between patients and caregiver	WebRTC[b] and VOIP[c]	[83,89]
Common data formats	Data interoperability and sharing	JSON, XML, CSV, and SenML[d]	[7,23,24,60,77,80,83,84,87,90,91]
Common semantics	Automatic data exchange, processing, and interpretation; persistent data storage and enabling use of AI[e]	FHIR[f], LOINC[g], SNOMED CT[h], RxNorm, openEHR, ontologies, BioPortal, DMTO[i], and HL7 CDA[j]	[19,23,24,60,77,82,88,92-94]
Data privacy and security	Encrypted data communication, access controls, and consent management services	Blockchain, HTTPS, TLS[k], SSL[l], GDPR[m], and HIPAA[n]	[7,24,54,77,79,82-85,87,89,92,95,96]
Extended sensor communication	Allowing devices in LAN[o] or PAN[p] to send their data over internet (WAN[q]; eg, allowing monitoring applications outdoors)	Gateways, LPWAN[r] (LoRa[s] or LoRaWAN), and 3G, 4G, or 5G	[19,24,60,80,83,91,95,97-99]
System-to-system communication	Allowing systems located on different networks to connect over WANs	HTTP, REST API[t], API, AMQP[u], and SOAP[v]	[7,19,23,24,60,77,79,80,82-89,92,94,96,100-102]
Allowing sensors from different vendors to communicate	Allowing sensors from different vendors to adhere to common data format and semantics	OneM2M and IEEE[w] 11073	[19,24,94]
Translation or transcoding between ontologies in different domains or creation of common ontologies	Interpreting and processing information coming from a separate domain (eg, using AI)	OWL[x], RDF[y], OM[z], Semantic Web, and HeTOP[aa]	[23,24,60,82,83,99,100,102]
Timely data processing and immediate responses in emergency situations	Minimizing delays and information loss in latency-sensitive and real-time applications (eg, heart monitoring systems)	Fog and edge computing	[7,19,79,80,83,85,91,97,98]
Device-to-device communication in PANs or LANs	Creating a smaller network of integrated medical devices that can collect patients' vital signs	Zigbee, BLE[ab], Bluetooth, Z-Wave, 6LoWPAN[ac], ANT+[ad], NFC[ae], CoAP[af], Wi-Fi, RFID[ag], and IEEE 11073 PHD[ah]	[7,19,23,24,80,83,84,89-91,94,95,97,99,100,103]
Modular system	Enhancing scalability and flexibility in IoMT; ensures consistent deployment and efficient application management	Microservices, Docker, and Kubernetes	[7,77,79,80,83,85]
Data availability and processing power	Improving administration and availability of data; enabling large-scale storage, computationally intensive data processing and advanced analysis tasks	Cloud services (computing and storage, etc)	[7,19,23,24,77,79,80,83,87,88,91,93,95,97,100]
Common infrastructures or middleware	Platforms, frameworks, and infrastructure for creating interoperable, standardized, and scalable solutions	FIWARE, IHE[ai], Continua, and OpenICE[aj]	[54,80,94]

[a]MQTT: message queuing telemetry transport.

[b]WebRTC: Web Real-Time Communication.

[c]VOIP: voice over IP.

[d]SenML: sensor markup language.

[e]AI: artificial intelligence.

[f]FHIR: Fast Healthcare Interoperability Resources.

[g]LOINC: Logical Observation Identifiers Names and Codes.

[h]SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms.

[i]DMTO: Diabetes Mellitus Treatment Ontology.

[j]HL7 CDA: Health Level 7 clinical document architecture.

[k]TLS: transport layer security.

XSL•FO
RenderX

[l]SSL: secure sockets layer.
[m]GDPR: General Data Protection Regulation.
[n]HIPAA: Health Insurance Portability and Accountability Act.
[o]LAN: local area network.
[p]PAN: personal area network.
[q]WAN: wide area network.
[r]LPWAN: low-power wide area network.
[s]LoRa: long range.
[t]REST API: representational state transfer application programming interface.
[u]AMQP: advanced message queuing protocol.
[v]SOAP: simple object access protocol.
[w]IEEE: Institute of Electrical and Electronics Engineers.
[x]OWL: Web Ontology Language.
[y]RDF: resource description framework.
[z]OM: ontology for units of measure.
[aa]HeTOP: Health Terminology/Ontology Portal.
[ab]BLE: Bluetooth Low Energy.
[ac]6LoWPAN: IPv6 over low-power wireless personal area network.
[ad]ANT+: Advanced and Adaptive Network Technology+.
[ae]NFC: near field communication.
[af]CoAP: constrained application protocol.
[ag]RFID: radio frequency identification.
[ah]PHD: personal health device.
[ai]IHE: Integrating the Healthcare Enterprise.
[aj]OpenICE: Open Integrated Clinical Environment.

Common Interoperability Challenges in IoMT

Achieving interoperability is not a trivial task. The higher the interoperability level, the more complex the endeavor due to the involvement of additional technologies and standards. Previous research has reported recurring interoperability challenges frequently encountered in IoMT platform development for prehospital care and HBC. Some of these challenges are reported in Table 2.

Table 2. Common interoperability challenges in Internet of Medical Things (IoMT) platform development for prehospital care and home-based care. Each challenge is mapped to studies (references) that discuss it in the context of previous IoMT platform developments efforts.

Challenges	References
Latency	[11,13,22,75,79,81-85,91,101]
Privacy or security	[11,13,24,60,79,82,83,85,91,101-103]
Volume and complexity of data	[11,22,75,82-84,91,94,97,98,100,101,104]
Multiple or proprietary protocols	[11,56,60,79,83,85,91,98]
Different terminologies or semantics	[19,23,24,60,82,88,100,102]
Poor connectivity	[84,95,97,103]

Among the examined studies (n=25), we could identify 7 frequently reported challenges associated with interoperability. Among the 7 reported challenges, 12 (21%) were related to latency problems, 8 (14%) to proprietary protocols, 13 (23%) to the volume and complexity of data, 12 (21%) to privacy and security concerns, 8 (14%) to semantic coding issues and 4 (7%) to poor connectivity.

Common Strategies to Overcome Interoperability Issues in IoMT Platform Development

In this section, we provide an overview of the technologies and standards used frequently to address interoperability challenges in IoMT platform development for prehospital care and HBC (Table 3).

XSL•FO
RenderX

Table 3. Summary of Internet of Medical Things (IoMT) platform development for home-based care and prehospital care as presented in 30 (19%) of the 158 studies reviewed, highlighting each IoMT platform's primary purpose and objective, publication year, and the technologies and standards used to address interoperability issues.

IoMT platform	Objective	Publication year	Technologies and standards
System for sleep monitoring [80]	Real-time, remote monitoring of patients with sleep apnea to support diagnosis and treatment	2018	Fog and cloud computing, MQTT[a], Zigbee, BLE[b], CoAP[c], 6LoWPAN[d], FIWARE, JSON, CSV, HDFS[e], REST API[f], gateways, microservices, and Docker
IoT[g]-based message broker system [18]	Individual patient vital signs monitoring in potential emergency or prehospital applications	2022	RFID[h], MQTT, REST API, Wi-Fi, CSV, XML, PDF, and JPEG
Platform for management of chronic diseases [23]	Care plan management tool integrated with clinical decision support services, EHRs[i], and sensors	2019	FHIR STU3[j], HL7 CDA[k], CSV, XML, SOAP[l], JSON, SNOMED CT[m], LOINC[n] and WHO ATC[o], local versions of *ICD-10*[p], REST API, cloud computing, HeTOP[q], and SMART[r] on FHIR
RePulmo[s] [54]	Open-source platform for secure and accurate remote pulmonary data monitoring	2019	MQTT, OpenICE[t], TLS[u], and API
SPIDEP[v] [7]	Support the early diagnosis of infectious diseases in older people	2020	Microservices, Docker, Kubernetes, cloud and edge computing, REST API, JSON, gateways, HTTPS, TLS, Wi-Fi, and SMART on FHIR
Semantic IoMT platform for eHealth [60]	An interoperable IoMT platform	2020	openEHR, OWL[w], ontologies, protocol converter, Bluetooth, Zigbee, Wi-Fi, SenML[x], gateways, JSON, REST API, cloud computing, and Semantic Web
IoMT platform for aggregation, processing, and sharing [19]	A platform to cover the domain of health care, following widely adopted standards, enabling semantic interoperability, and considering a big data approach	2019	REST API; MQTT; CoAP; openEHR; FHIR; SenML; Wi-Fi; cloud, fog, and edge computing; OneM2M; gateways, and Wi-Fi
Platform for health care promotion and cardiovascular disease prevention [96]	Smart health care platform oriented to multiple point-of-care scenarios for health care promotion and cardiovascular disease prevention	2021	REST API, TCP[y], JSON, XML, SOAP, Bluetooth, Wi-Fi, ANT+[z], HTTPS, TLS, cloud computing, and Wi-Fi
SemPryv [84]	SemPryv supports REST APIs to consume and produce interoperable streams of health care data, following the HL7 FHIR standard and using the Pryv.io platform	2023	FHIR, ontologies, SNOMED CT, LOINC, BioPortal, REST API, WebSocket, webhooks, RxNorm, UCUM[aa], Docker, GDPR[ab], Kubernetes, and RDF[ac]
LoRaWAN[ad]-based NXTGeUH[ae] [103]	System for vital signs monitoring and fall detection	2018	Zigbee, LoRaWAN, TCP/IP, Bluetooth, and gateways
Monitoring platform for patients with diabetes [90]	Support treatment, monitoring, and data collection	2021	API, Bluetooth, JSON, and NFC[af]
VITASENIOR-MT [83]	Remote monitoring of health parameters of older people	2019	Cloud and fog computing, microservices, CoAP, JSON, BLE, WebSocket, REST API, MQTT, gateways, 6LoWPAN, AMQP[ag], Bluetooth, WebRTC[ah], GDPR, and HTTPS
Apnea MedAssist 2 [104]	Monitoring for Alzheimer disease diagnosis and rehabilitation	2013	Bluetooth, Wi-Fi, cloud computing, and 4G
IC-SMART[ai] [100]	Monitoring for Alzheimer disease diagnosis and rehabilitation	2019	OWL, ontologies, gateways, REST API, RFID, Wi-Fi, Bluetooth, cloud computing, Docker, and HTTP
HealthFog [101]	Automatic diagnosis of heart diseases	2020	REST API; cloud, edge, and fog computing; gateways; CSV; blockchain; and HTTP
HABITAT[aj] [102]	Platform for independent older people	2019	RFID, HTTPS, WebSocket, RDF, JSON, Semantic Web, ontologies, and SPARQL[ak]

XSL•FO
RenderX

IoMT platform	Objective	Publication year	Technologies and standards
SensorHub [84]	A central platform that acts as a bridge between sensors and a user's smartphone	2022	BLE, Bluetooth, REST API, HTTPS, WebSocket, CSV, JSON, and Docker
Energy-aware IoT-based architecture [85]	On-wrist fall detection system	2023	MQTT, HTTPS, Wi-Fi, Docker, and API
Real-time remote monitoring system [91]	Low-cost health system for continuous monitoring of ECG[al]	2017	Bluetooth, Wi-Fi, gateways, cloud and fog computing, JSON, XML, API, and TCP
Fall detection system [97]	A system that can be used to monitor, for example, cardiovascular diseases or diabetes	2019	LoRa; edge, fog, and cloud computing; BLE; and gateways
ChainSDI [79]	Securing health care applications at home	2020	Edge and cloud computing, blockchain, Wi-Fi, Kubernetes, Docker, HIPAA[am], HTTP, and API
SmartHabits [77]	A home care assistant system to support informal caregivers caring for persons living alone	2019	REST API, JSON, AMQP, gateways, cloud computing, ontologies, Z-Wave, HTTP, microservices, GDPR, and Wi-Fi
DALÍ [87]	A personal assistant platform for older people and those who are visually impaired	2020	JSON, MQTT, HTTPS, REST API, cloud computing, WebSocket, and SSL[an]
Integrated health monitoring system [94]	Integrated monitoring system	2015	FHIR, REST API, IEEE[ao] 11073, Zigbee, Bluetooth, IHE[ap], Continua, SOAP, JSON, and XML
A mobile health monitoring-and-treatment system [24]	Platform for diabetes monitoring and to provide customized, long-term, and real-time treatment plans	2019	FHIR, ontologies, OWL, JSON, HTTPS, REST API, Bluetooth, cloud computing, SNOMED CT, LOINC, RxNorm, BioPortal, DMTO[aq], IEEE 11073, Wi-Fi, 3G, 4G, or 5G, and BFO[ar]
Patient monitoring system in smart homes [98]	Remote monitoring platform for monitoring patients in smart homes	2018	Cloud and fog computing, RFID, gateways, and Wi-Fi
VitalCore [88]	Allowing patients access to their health data in real time	2021	REST API, WebSocket, and HL7 data
Health care system for older people [99]	An ontology-based IoMT platform to alleviate problems related to chronic diseases	2021	Ontologies, gateways, Wi-Fi, Zigbee, LoRaWAN, RDF, and OWL
We-Care [95]	Platform to assist older people in their homes and trigger alarms in case of emergency situations; works offline	2017	Bluetooth, 6LoWPAN, gateways, UDP, Wi-Fi, cloud computing, and API

XSL•FO
RenderX

IoMT platform	Objective	Publication year	Technologies and standards
Remote pain monitoring system [81]	A scalable IoT system for real-time automatic pain assessment using facial expressions	2018	Gateways, UDP, TCP, Wi-Fi, WebSocket, TLS, and cloud computing

[a]MQTT: message queuing telemetry transport.
[b]BLE: Bluetooth Low Energy.
[c]CoAP: constrained application protocol.
[d]6LoWPAN: IPv6 over low-power wireless personal area network.
[e]HDFS: Hadoop Distributed File System.
[f]REST API: representational state transfer application programming interface.
[g]IoT: Internet of Things.
[h]RFID: radio frequency identification.
[i]EHR: electronic health record.
[j]FHIR STU3: Fast Healthcare Interoperability Resources Standard for Trial Use, version 3.
[k]HL7 CDA: Health Level 7 clinical document architecture.
[l]SOAP: simple object access protocol.
[m]SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms.
[n]LOINC: Logical Observation Identifiers Names and Codes.
[o]WHO ATC: World Health Organization Anatomical Therapeutic Chemical.
[p]ICD-10: International Classification of Diseases, Tenth Revision.
[q]HeTOP: Health Terminology/Ontology Portal.
[r]SMART: Substitutable Medical Applications, Reusable Technologies.
[s]RePulmo: remote pulmonary monitoring system.
[t]OpenICE: Open Integrated Clinical Environment.
[u]TLS: transport layer security.
[v]SPIDEP: System for Prediagnosis and Telecare of Infectious Diseases in Elderly People.
[w]OWL: Web Ontology Language.
[x]SenML: sensor markup language.
[y]TCP: transmission control protocol.
[z]ANT+: Advanced and Adaptive Network Technology+.
[aa]UCUM: Unified Code for Units of Measure.
[ab]GDPR: General Data Protection Regulation.
[ac]RDF: resource description framework.
[ad]LoRaWAN: long-range wide area network.
[ae]NXTGeUH: Next Generation Ubiquitous Healthcare.
[af]NFC: near field communication.
[ag]AMQP: advanced message queuing protocol.
[ah]WebRTC: Web Real-Time Communication.
[ai]IC-SMART: Internet of Things Cloud-Enabled Seamless Monitoring for Alzheimer Diagnosis and Rehabilitation.
[aj]HABITAT: Home Assistance Based on the Internet of Things for the Autonomy of Everybody.
[ak]SPARQL: SPARQL Protocol and RDF Query Language.
[al]ECG: electrocardiography.
[am]HIPAA: Health Insurance Portability and Accountability Act.
[an]SSL: secure sockets layer.
[ao]IEEE: Institute of Electrical and Electronics Engineers.
[ap]IHE: Integrating the Healthcare Enterprise.
[aq]DMTO: Diabetes Mellitus Treatment Ontology.
[ar]BFO: Basic Formal Ontology.
[as]UDP: user datagram protocol.

As can be seen in Table 3, IoMT platform development often requires a combination of technologies and standards. Simply specifying “Bluetooth,” “Wi-Fi,” or “SNOMED CT” is not sufficient because no single standard, protocol, or technology can address higher levels of interoperability [100]; for example, SNOMED CT can be complemented by more domain-specific standards such as LOINC [29], while Bluetooth can be used in conjunction with a gateway, MQTT, and FHIR to enhance interoperability within IoMT [19]. Table 2 reveals variations in the frequency of the technology and standards used. Consequently, we have summarized the most used technologies and standards in Figure 8. Technologies and standards that

XSL•FO
RenderX

appeared in 2 or fewer articles were excluded from Figure 8. These include Web Real-Time Communication (WebRTC), openEHR, voice over IP, webhooks, and Substitutable Medical Applications, Reusable Technologies (SMART) on FHIR, as well as the interoperability guidelines and frameworks such as FIWARE, Integrating the Healthcare Enterprise (IHE), Continua, and Open Integrated Clinical Environment (OpenICE). The top 5 reported technologies were cloud computing (19/37, 51%), REST APIs (17/37, 46%), Wi-Fi (17/37, 46%), gateways (15/37, 41%), and JSON (14/37, 38%).

Figure 8. Occurrences of common technologies addressing interoperability in Internet of Medical Things (IoMT) platforms for prehospital care and home-based care, grouped by application area (each color). Only technologies and standards reported in 2 or more separate articles are included. 6LoWPAN: IPv6 over low-power wireless personal area network; API: application programming interface; BLE: Bluetooth Low Energy; CoAP: constrained application protocol; FHIR: Fast Healthcare Interoperability Resources; GDPR: General Data Protection Regulation; LOINC: Logical Observation Identifiers Names and Codes; MQTT: message queuing telemetry transport; OWL: Web Ontology Language; RDF: resource description framework; REST API: representational state transfer application programming interface; RFID: radio frequency identification; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms; TLS: transport layer security.

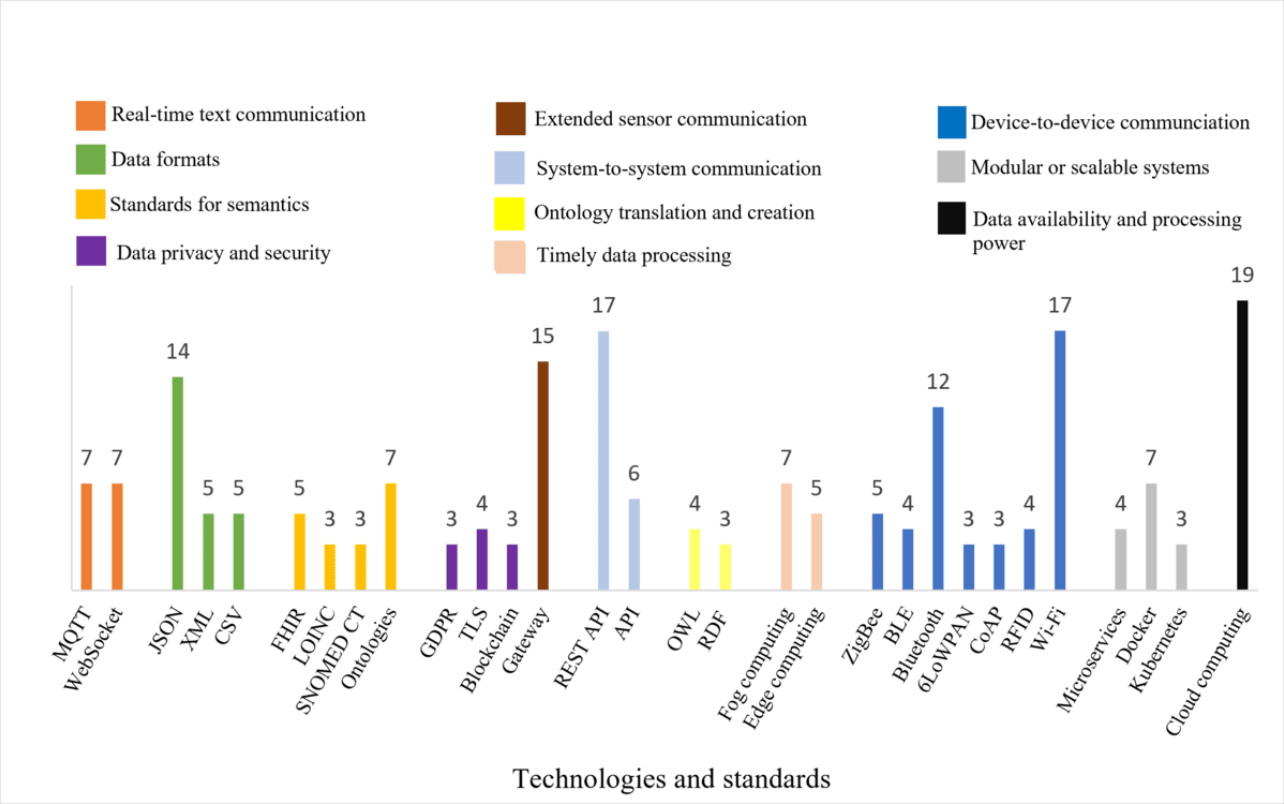

Recommendations

Our recommendations (Textbox 3) are based on the findings presented in Table 3 together with discussions within our research group. These recommendations focus on technologies and standards that support cross-domain interoperability across all layers of the IoMT reference model, spanning from data collection by sensors to cloud-based data processing and secure cross-domain data exchange. In broad terms, an IoMT platform for HBC and prehospital care should include, in our opinion, the capabilities presented in Textbox 3.

XSL•FO
RenderX

Textbox 3. Capabilities required by an Internet of Medical Things (IoMT) platform for home-based care and prehospital care.

Compatibility for medical device integration and secure and reliable real-time data transfer to the cloud

- The IoMT platform should ensure seamless integration of medical devices and enable secure, real-time transmission of data to the cloud.
- Recommended technologies and standards: Bluetooth Low Energy (BLE), Zigbee, Institute of Electrical and Electronics Engineers (IEEE) 11073 personal health device (PHD), Open Integrated Clinical Environment (OpenICE), and FIWARE

Real-time and low latency data processing and communication

- The IoMT platform should be capable of processing and visualizing data in real time with minimal latency to support timely decision-making and data processing.
- Recommended technologies and standards: message queuing telemetry transport (MQTT), WebSocket, and Web Real-Time Communication (WebRTC)

Data persistence for reliable data storage

- The IoMT platform should ensure that data are stored in a consistent and secure way.
- Recommended technologies and standards: openEHR and blockchain

Open data standards and data exchange mechanisms for secure and efficient system-to-system communication.

- The IoMT platform should adhere to open data standards and secure exchange mechanisms, facilitating efficient communication between different systems.
- Recommended technologies and standards: JSON and representational state transfer application programming interface (REST API)
- Common semantics and domain knowledge descriptions for automatic knowledge extraction

Common semantics and domain descriptions enable automatic extraction of knowledge from data, enhancing system understanding.

- Recommended technologies and standards: Fast Healthcare Interoperability Resources (FHIR), Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT), Web Ontology Language (OWL), resource description framework (RDF), Logical Observation Identifiers Names and Codes (LOINC), and BioPortal

Security measures to protect patient data and ensure data confidentiality

- The IoMT platform should ensure the privacy and integrity of patient data throughout its life cycle.
- Recommended technologies and standards: transport layer security (TLS), General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), blockchain, and Open Authorization 2.0 (OAuth2)

Scalability to accommodate a growing number of connected devices and generated data volumes

- The platform should scale effectively to handle an increasing number of devices and growing data volumes.
- Recommended technologies and standards: Docker, Kubernetes, cloud services, and fog and edge computing

Discussion

Principal Findings and Study Contribution

This scoping review provides insights into the enabling technologies that can address interoperability issues in IoMT settings. Although the primary focus has been on solutions for prehospital care and HBC, the results are also applicable to other settings with similar characteristics and requirements. The results show that higher levels of interoperability in IoMT can be achieved by combining various technologies and standards from multiple interoperability levels.

On the basis of the studies (n=30) presented in Table 3, it seems that contemporary IoMT studies on prehospital care and HBC tend to focus on lower levels of interoperability. This conclusion is in line with previous research conducted by Rubí and Gondim [19]. One possible explanation for the predominant focus on device integration and lower levels of interoperability among the reviewed studies is our use of “IoMT” as a search term. Although IoMT is broad and includes integration with medical devices as well as medical applications, the term is somewhat ambiguous and might introduce some biases toward device-centric approaches. We believe that future research could benefit from using a more refined definition to emphasize a shifted focus toward higher levels of interoperability. Therefore, we propose the introduction of the term Internet of Medical Systems (IoMS) to address this gap. This new terminology could help distinguish studies focusing on device-to-device integration from those focusing on integration with EHRs and other health applications.

We acknowledge that some may argue that technologies and standards such as blockchain, fog computing, GDPR, and the like may primarily focus on aspects other than interoperability (eg, security, communication, networking, data transfer, or architectural design). However, proceeding from the 6-level interoperability model [11], our interpretation of interoperability extends beyond mere data exchange and interpretation. Even if two systems possess the capabilities to exchange and interpret

XSL•FO
RenderX

information, we believe that it is not enough to achieve higher levels of interoperability within health care without considering privacy and security aspects. However, our findings indicate varying levels of focus on technologies such as blockchain and SMART on FHIR within the research community, with some of the studies (4/30, 13%) analyzing these technologies in depth, while others (26/30, 87%) tend to overlook them. As an example, of the 30 included studies, 8 (27%) reported the use of authentication and authorization mechanisms [79,80,90,91,95,96,101,105], but only 7% (2/30) reported the use of Open Authorization (OAuth) or SMART on FHIR as mechanisms for authentication and authorization [7,23].

On the basis of our findings, it seems that more comprehensive platforms tend to incorporate a wider range of technologies compared to simpler platforms (Table 3); for example, platforms that reported interoperability capabilities with external systems or the use of AI seem to more frequently use technologies and standards to semantically code the data, including ontologies, FHIR, LOINC, *ICD*, openEHR, RxNorm, and SNOMED CT. However, this relationship was not confirmed statistically, and we acknowledge that other factors may influence platform complexity, such as regulatory requirements, available technologies, and expertise among developers.

Table 3 reveals that 30% (9/30) of the IoMT platform development studies for prehospital care and HBC reported the use of semantic frameworks or ontologies. As the creation of ontologies is a necessary but time-consuming and labor-intensive task, research has examined the potential to achieve ontology alignment by using AI; for example, Dam et al [106] used augmented intelligence to map data across various domains to facilitate the integration of diverse data sources. According to Tangi et al [107], AI has the potential to not only establish a common language and foster a shared understanding of data but also clean and structure the data. While we acknowledge that using AI for interoperability is an emerging trend, such studies have not been incorporated in this review.

In the reviewed literature, semantic standards and ontologies were described as essential for establishing connections with external systems and enabling the use of AI. Interestingly, standards for image and video formats, such as Digital Imaging and Communications in Medicine (DICOM), have not been mentioned in the examined studies. Interoperability enhancing organizations and initiatives such as FIWARE, IHE, Continua, and OpenICE were reported to guide researchers and developers [88], but these organizations and initiatives were only mentioned in 3 (10%) of the 30 studies.

Fog and edge computing were reported to be used to overcome problems related to limited bandwidth, often associated with high data rate applications, such as fall detection systems or electrocardiography monitoring [97]. Our findings suggest that fog and edge computing are often combined with cloud computing, especially in AI-intensive applications that require more processing power. To improve the performance of AI algorithms and enable machines to correlate data, studies reported the use of ontologies, resource description framework (RDF), and OWL [24,99]. Furthermore, platforms requiring real-time functionalities (alerts or notifications) reported the use of MQTT (7/30, 23%) and WebSocket (7/30, 23%) [82,85].

While we have documented the use of commonly used technologies and standards in this study, it is important to acknowledge that the accuracy of these findings depends on the transparency and thoroughness of the explanations within the included studies; for instance, some of the studies (30/63, 48%) offered detailed insights into their IoMT platform development processes and the technologies and standards used, while others (33/63, 52%) focused on broad aspects of the topic, as in the studies by Rakhman et al [108] and Shim et al [109]. As a result, some of the studies (33/63, 52%) were excluded from Table 2. Furthermore, because technologies and standards can have versatile applications, specifying their exact application area can often be challenging. Therefore, the findings presented in Table 3 should be considered as a guidance rather than absolute truths.

Mapping the Enabling Technologies to the Interoperability Model

In this study, the mapping was performed based on a 6-level interoperability model defined in the study protocol [11]. To facilitate the mapping process, any inconsistencies between our model and the levels described in the included studies were addressed using a flowchart model [11]. This was done to manage different naming conventions across different models; for instance, one study may label a specific interoperability level as “technical interoperability,” [105] while others may refer to it as “foundational interoperability” [110]. Similarly, the specific functions and components of each level may not be consistent across different models or frameworks, which can create ambiguity and uncertainty about how to define and interpret interoperability. Hence, proceeding from a different interoperability model [71] would most likely result in a somewhat different mapping outcome.

It is important to note that although we mapped the enabling technologies to specific interoperability levels, it is uncommon for a technology to address a single interoperability level; for example, the FHIR standard not only specifies data exchange formats and semantics but also supports REST APIs, enabling it to address syntactic, semantic, and cross-platform interoperability concurrently. Hence, in this study, we mapped FHIR to cross-platform interoperability, which represents the highest level of interoperability addressed by FHIR. However, saying that the FHIR standard *solves* cross-platform interoperability is not completely true because it needs to be combined with technologies on lower interoperability levels. We argue that the correct phrasing should be that FHIR *can* solve cross-platform interoperability issues if it is appropriately combined with other relevant technologies and standards.

Another aspect worth mentioning regarding the mapping is the importance of examining the primary focus of the technologies and standards; for example, the IEEE 11073 PHD standard involves data models that specify how a measurement and observations of vital signs should be represented, including descriptions of the different units, data types, and semantic meanings [38,40]. On the basis of this description, it is likely that the IEEE 11073 PHD standard should be categorized at the

XSL•FO
RenderX

semantic interoperability level. However, due to its primary focus on device-to-device communication within LANs or PANs, the IEEE 11073 PHD standard does not fully meet the requirements for network interoperability, let alone semantic interoperability. Hence, the IEEE 11073 PHD standard was mapped to the device interoperability level in this study.

Summarizing the Enabling Technologies That Address Interoperability Issues

In our study, we found that technologies such as BLE and Zigbee, which primarily target device interoperability, are not sufficient for IoMT applications in prehospital or HBC settings due to their restricted physical communication range (ie, 20-100 meters) [41,111]. For physicians to remotely monitor patients and for devices to automatically send alarms during medical emergencies, the first step is to extend the communication range by enabling network interoperability [49]. Otherwise, the physician's system would need to be in the same room as the patient, which obviates the concept of remote monitoring; for example, by combining BLE with long-range wide area network (LoRaWAN) or 4G or 5G—or 6LoWPAN with CoAP—long-range data transmission can be achieved [112]. This allows physicians to remotely monitor patients residing several kilometers away from the hospital [25]. However, choosing a particular combination of technologies from different interoperability levels does not ensure the achievement of the desired level of interoperability. This is because interoperability initiatives require collaboration among all involved stakeholders; for instance, a standard such as IEEE 11073 can be adopted in several ways, which means that simply specifying the use of IEEE 11073 may not be sufficient to achieve device interoperability. Instead, a standard should be tailored to meet the specific requirements of a particular use case or application. This can be achieved using profiles to ensure that the standard is implemented in a way that suits the needs.

Although close collaboration among stakeholders is essential to achieve higher levels of interoperability, studies have shown this endeavor to be challenging; for example, a study conducted in 2021 by Everson et al [113] revealed that more than half (55%) of health information exchange organizations in the United States reported varying degrees of intentional information blocking. The authors showed that the most prevalent form of information blocking involved the refusal to share information, which 14% of the health information exchange organizations routinely observed among EHR vendors. The reason for blocking information according to Everson et al [113] had to do with regional competition among vendors. By limiting the sharing of information with other actors, vendors can ensure that health care providers stick to their platform [113,114]. While this approach benefits vendors, it creates disadvantages for patients and the health care industry. The findings of Everson et al [113] suggest that while technologies and standards can streamline data exchange in health care, the reluctance of organizations to engage in such efforts impedes progress toward higher levels of interoperability.

Comments on Our Recommendations

The recommendations provided in this study draw on insights from previous IoMT platform development projects and our research group's expertise. They focus on technologies and standards that together can support cross-platform or cross-domain interoperability. We acknowledge that multiple technologies and standards can be used to develop IoMT platforms. We are also aware that there is no "one right way" of adopting a specific technology or standard. Rather, we argue that interoperability initiatives and development projects should be facilitated by close collaboration among the stakeholders involved.

With this study and our recommendations, we hope to bring together the research, developer, and health care communities by highlighting relevant technologies and standards to foster collaboration and improve interoperability initiatives within the health care domain. While we recognize the importance of independence in driving innovation, we hope to see more IoMT platforms using common standards and best practices already available in the market. We further hope to see an increased engagement in interoperability initiatives such as FIWARE, IHE, and OpenICE.

We acknowledge the advantages of IoMT platforms capable of delivering personalized patient care and enhancing patient safety. However, we do not believe that technologies and standards are the only solutions. To achieve the highest level of interoperability, we argue that it is necessary to have a profound understanding of organizations' working processes, routines, and policies. Hence, we look forward to delving into these challenges in close collaboration with relevant stakeholders and developing an IoMS platform that supports cross-domain interoperability.

Limitations and Biases

This scoping review adheres to the published protocol [11] and focuses on technologies addressing interoperability issues within the context of IoMT platform development in HBC and prehospital care settings. Hence, studies focusing on hospital systems [115], platforms for intensive care units [81], integration platforms [93], or platforms for research or clinical trials [92,116] were omitted. Furthermore, most of the studies (33/63, 52%) displayed a deficiency in information or a lack of transparency regarding the development process and the technologies used; therefore, they were excluded from Table 1 and Table 3. As the focus of this study was to identify and summarize best practices and frequently used technologies, we did not assess the effectiveness of technology implementations in detail.

To manage the scope of this review process effectively, the reviewers made a deliberate decision to limit the included studies to those specifically addressing interoperability in an IoMT context. This was necessary to avoid overwhelming screening efforts with an unmanageable number of studies. We are aware that we might have missed relevant articles with this strategy. Furthermore, we are aware that technologies used in other sectors, for example, Industry 4.0, manufacturing, and transportation, could also be applicable to IoMT, especially considering that these domains share some common characteristics (eg, requirements for real-time, cross-domain information exchange). Another limitation of this study might be the use of the term "IoMT" in our literature search. Although

XSL•FO
RenderX

IoMT is often defined as the interconnection of medical devices and applications, there seems to be a predominant focus on devices and lower levels of interoperability among these studies. This limitation could potentially be eliminated by the introduction and use of the new term IoMS.

Although recent advancements in AI have demonstrated its potential to enhance interoperability, studies reporting the use of AI for this purpose were excluded from this review. Another limitation of this study is that, to expedite the process, the coding scheme was tested by only 1 reviewer (MS).

Conclusions

In this study, we have demonstrated that the highest level of interoperability can be theoretically achieved through a strategic combination of various technologies and standards. Furthermore, we have provided a summary of the relevant technologies and standards that can be used in IoMT platforms to overcome interoperability issues in HBC and prehospital care settings.

Despite the availability of innovative and suitable technologies, the IoMT research community has reported limited interest in adopting technologies and standards such as Docker, blockchain, SMART on FHIR, HIPAA, and GDPR to achieve cross-domain interoperability. Most of the studies (17/30, 57%) have primarily focused on lower levels of interoperability (up to the semantic interoperability level). This observation highlights a significant research gap, particularly in achieving cross-domain interoperability within IoMT for HBC and prehospital care. To emphasize the need for higher levels of interoperability and to support future research, we advocate for the introduction of the term IoMS. In addition, the reluctance of organizations and vendors to share information and participate in interoperability initiatives highlights the importance of considering these aspects when addressing interoperability challenges, especially in procurement processes [117].

Acknowledgments

This study was supported by the following funding bodies: the Kamprad Family Foundation for Entrepreneurship, Research & Charity; the European Interreg project “Kontiki – AI som Beslutsstöd for patienter och helsetjenesten” (“Kontiki – AI as decision support for patients and health care”); and the strategic innovation program IoT Sweden, a joint venture by Vinnova, Formas, and the Swedish Energy Agency, through the IoT project “ASAP PoC – bättre militära och civila prehospitala Point-of-Care beslut med hjälp av datafusion och AI” (“ASAP-PoC – better civil and military prehospital point-of-care decisions using data fusion and AI”; 2022-03748). The funding bodies did not have a role in the design of the study or in the writing of the manuscript.

Data Availability

All datasets generated or analyzed during this study are included in this published paper and its supplementary information files.

Conflicts of Interest

OM is a sales engineer at InterSystems (Stockholm, Sweden). All other authors declare no other conflicts of interest.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 498 KB - jmir_v27i1e54470_app1.pdf]

Multimedia Appendix 2
Search strategies for electronic databases and mapping result.
[XLSX File (Microsoft Excel File), 31 KB - jmir_v27i1e54470_app2.xlsx]

References

1. Holman HR. The relation of the chronic disease epidemic to the health care crisis. ACR Open Rheumatol 2020 Mar;2(3):167-173 [FREE Full text] [doi: 10.1002/acr2.11114] [Medline: 32073759]
2. Brennan P, Perola M, van Ommen GJ, Riboli E, European Cohort Consortium. Chronic disease research in Europe and the need for integrated population cohorts. Eur J Epidemiol 2017 Sep;32(9):741-749 [FREE Full text] [doi: 10.1007/s10654-017-0315-2] [Medline: 28986739]
3. Ansah JP, Chiu C. Projecting the chronic disease burden among the adult population in the United States using a multi-state population model. Front Public Health 2022;10:1082183 [FREE Full text] [doi: 10.3389/fpubh.2022.1082183] [Medline: 36711415]
4. Ren L, Peng Y. Research of fall detection and fall prevention technologies: a systematic review. IEEE Access 2019;7:77702-77722. [doi: 10.1109/access.2019.2922708]
5. Feigin VL, GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]

XSL•FO
RenderX

6. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021 [FREE Full text] [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
7. Calderon-Gomez H, Mendoza-Pitti L, Vargas-Lombardo M, Gomez-Pulido JM, Castillo-Sequera JL, Sanz-Moreno J, et al. Telemonitoring system for infectious disease prediction in elderly people based on a novel microservice architecture. IEEE Access 2020;8:118340-118354. [doi: 10.1109/access.2020.3005638]
8. Landers S, Madigan E, Leff B, Rosati RJ, McCann BA, Hornbake R, et al. The future of home health care: a strategic framework for optimizing value. Home Health Care Manag Pract 2016 Nov;28(4):262-278 [FREE Full text] [doi: 10.1177/1084822316666368] [Medline: 27746670]
9. Storman D, Jemioło P, Swierz MJ, Sawiec Z, Antonowicz E, Prokop-Dorner A, et al. Meeting the unmet needs of individuals with mental disorders: scoping review on peer-to-peer web-based interactions. JMIR Ment Health 2022 Dec 05;9(12):e36056 [FREE Full text] [doi: 10.2196/36056] [Medline: 36469366]
10. Young HM, Nesbitt TS. Increasing the capacity of primary care through enabling technology. J Gen Intern Med 2017 Apr;32(4):398-403 [FREE Full text] [doi: 10.1007/s11606-016-3952-3] [Medline: 28243871]
11. Seth M, Jalo H, Högstedt Å, Medin O, Björner U, Sjöqvist BA, et al. Technologies for interoperable internet of medical things platforms to manage medical emergencies in home and prehospital care: protocol for a scoping review. JMIR Res Protoc 2022 Sep 20;11(9):e40243 [FREE Full text] [doi: 10.2196/40243] [Medline: 36125863]
12. Luker JA, Worley A, Stanley M, Uy J, Watt AM, Hillier SL. The evidence for services to avoid or delay residential aged care admission: a systematic review. BMC Geriatr 2019 Aug 08;19(1):217 [FREE Full text] [doi: 10.1186/s12877-019-1210-3] [Medline: 31395018]
13. Villanueva-Miranda I, Nazeran H, Martinek R. A semantic interoperability approach to heterogeneous internet of medical things (IoMT) platforms. In: Proceedings of the 20th International Conference on e-Health Networking, Applications and Services. 2018 Presented at: Healthcom 18; September 17-20, 2018; Ostrava, Czech Republic p. 1-5 URL: https://ieeexplore.ieee.org/document/8531103 [doi: 10.1109/healthcom.2018.8531103]
14. Albahri OS, Albahri AS, Mohammed KI, Zaidan AA, Zaidan BB, Hashim M, et al. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. J Med Syst 2018 Mar 22;42(5):80. [doi: 10.1007/s10916-018-0943-4] [Medline: 29564649]
15. Fleming J, Brayne C, Cambridge City over-75s Cohort (CC75C) study collaboration. Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. BMJ 2008 Nov 17;337:a2227 [FREE Full text] [doi: 10.1136/bmj.a2227] [Medline: 19015185]
16. Lei N, Kareem M, Moon SK, Ciaccio EJ, Acharya UR, Faust O. Hybrid decision support to monitor atrial fibrillation for stroke prevention. Int J Environ Res Public Health 2021 Jan 19;18(2):813 [FREE Full text] [doi: 10.3390/ijerph18020813] [Medline: 33477887]
17. Simonsen SA, Andresen M, Michelsen L, Viereck S, Lippert FK, Iversen HK. Evaluation of pre-hospital transport time of stroke patients to thrombolytic treatment. Scand J Trauma Resusc Emerg Med 2014 Nov 13;22:65 [FREE Full text] [doi: 10.1186/s13049-014-0065-z] [Medline: 25391354]
18. Tsao YC, Cheng FJ, Li YH, Liao LD. An IoT-based smart system with an MQTT broker for individual patient vital sign monitoring in potential emergency or prehospital applications. Emerg Med Int 2022;2022:7245650 [FREE Full text] [doi: 10.1155/2022/7245650] [Medline: 35132364]
19. S Rubí JN, L Gondim PR. IoMT platform for pervasive healthcare data aggregation, processing, and sharing based on OneM2M and OpenEHR. Sensors (Basel) 2019 Oct 03;19(19):4283 [FREE Full text] [doi: 10.3390/s19194283] [Medline: 31623304]
20. Winderbank-Scott P, Barnaghi P. A non-invasive wireless monitoring device for children and infants in pre-hospital and acute hospital environments. In: Proceedings of the 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). 2017 Presented at: iThings-GreenCom-CPSCom-SmartData '07; June 21-23, 2017; Exeter, UK p. 591-597 URL: https://ieeexplore.ieee.org/document/8276812 [doi: 10.1109/ithings-greencom-cpscom-smartdata.2017.93]
21. Jaleel A, Mahmood T, Hassan MA, Bano G, Khurshid SK. Towards medical data interoperability through collaboration of healthcare devices. IEEE Access 2020;8:132302-132319. [doi: 10.1109/access.2020.3009783]
22. Yang Y, Li X, Qamar N, Liu P, Ke W, Shen B, et al. Medshare: a novel hybrid cloud for medical resource sharing among autonomous healthcare providers. IEEE Access 2018;6:46949-46961. [doi: 10.1109/access.2018.2865535]
23. Laleci Erturkmen GB, Yuksel M, Sarigul B, Arvanitis TN, Lindman P, Chen R, et al. A collaborative platform for management of chronic diseases via guideline-driven individualized care plans. Comput Struct Biotechnol J 2019;17:869-885 [FREE Full text] [doi: 10.1016/j.csbj.2019.06.003] [Medline: 31333814]
24. El-Sappagh S, Ali F, Hendawi A, Jang J, Kwak K. A mobile health monitoring-and-treatment system based on integration of the SSN sensor ontology and the HL7 FHIR standard. BMC Med Inform Decis Mak 2019 May 10;19(1):97 [FREE Full text] [doi: 10.1186/s12911-019-0806-z] [Medline: 31077222]

XSL•FO
RenderX

25. Adel E, El-Sappagh S, Barakat S, Kwak KS, Elmogy M. Semantic architecture for interoperability in distributed healthcare systems. IEEE Access 2022;10:126161-126179. [doi: 10.1109/access.2022.3223676]
26. Hyvämäki P, Sneck S, Meriläinen M, Pikkarainen M, Kääriäinen M, Jansson M. Interorganizational health information exchange-related patient safety incidents: a descriptive register-based qualitative study. Int J Med Inform 2023 Jun;174:105045 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105045] [Medline: 36958225]
27. Magrabi F, Ong MS, Runciman W, Coiera E. An analysis of computer-related patient safety incidents to inform the development of a classification. J Am Med Inform Assoc 2010;17(6):663-670 [FREE Full text] [doi: 10.1136/jamia.2009.002444] [Medline: 20962128]
28. Bates DW, Samal L. Interoperability: what is it, how can we make it work for clinicians, and how should we measure it in the future? Health Serv Res 2018 Oct;53(5):3270-3277 [FREE Full text] [doi: 10.1111/1475-6773.12852] [Medline: 29527678]
29. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79 [FREE Full text] [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
30. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
31. Halevi G, Moed H, Bar-Ilan J. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—review of the literature. J Informetr 2017 Aug;11(3):823-834. [doi: 10.1016/j.joi.2017.06.005]
32. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020 Mar 28;11(2):181-217 [FREE Full text] [doi: 10.1002/jrsm.1378] [Medline: 31614060]
33. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Khatri S, Alzahrani FA, Ansari MT, Agrawal A, Kumar R, Khan RA. A systematic analysis on blockchain integration with healthcare domain: scope and challenges. IEEE Access 2021;9:84666-84687. [doi: 10.1109/access.2021.3087608]
36. Noura M, Atiquzzaman M, Gaedke M. Interoperability in internet of things: taxonomies and open challenges. Mobile Netw Appl 2018 Jul 21;24(3):796-809. [doi: 10.1007/S11036-018-1089-9]
37. Roehrs A, da Costa CA, Righi RD, Rigo SJ, Wichman MH. Toward a model for personal health record interoperability. IEEE J Biomed Health Inform 2019 Mar;23(2):867-873. [doi: 10.1109/JBHI.2018.2836138] [Medline: 29993759]
38. Lim JH, Park C, Park S, Lee K. ISO/IEEE 11073 PHD message generation toolkit to standardize healthcare device. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:1161-1164. [doi: 10.1109/IEMBS.2011.6090272] [Medline: 22254521]
39. Sloane EB, Thalassinidis A, Silva RJ. ISO/IEEE 11073, IHE, and HL7: fostering standards-based safe, reliable, secure and interoperable biomedical technologies. In: Proceedings of the annual IEEE Conference on Technical, Professional, and Student. 1073 Presented at: SoutheastCon '18; April 19-22, 2018; Saint Petersburg, FL p. 1 URL: https://ieeexplore.ieee.org/document/8479122 [doi: 10.1109/secon.2018.8479051]
40. Clarke M, de Folter J, Verma V, Gokalp H. Interoperable end-to-end remote patient monitoring platform based on IEEE 11073 PHD and ZigBee health care profile. IEEE Trans Biomed Eng 2018 May;65(5):1014-1025. [doi: 10.1109/TBME.2017.2732501] [Medline: 28796600]
41. Wang ZQ, Huang ZH. Wearable health status monitoring device for electricity workers using ZigBee-based wireless sensor network. In: Proceedings of the 7th International Conference on Biomedical Engineering and Informatics. 2014 Presented at: BMEI '14; October 14-16, 2014; Dalian, China p. 602-606 URL: https://ieeexplore.ieee.org/document/7002845 [doi: 10.1109/bmei.2014.7002845]
42. Caranguian LP, Pancho-Festin S, Sison LG. Device interoperability and authentication for telemedical appliance based on the ISO/IEEE 11073 personal health device (PHD) standards. In: Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012 Presented at: PHD '12; August 28-September 1, 2012; San Diego, CA p. 1270-1273 URL: https://ieeexplore.ieee.org/document/6346169 [doi: 10.1109/embc.2012.6346169]
43. Javaid S, Zeadally S, Fahim H, He B. Medical sensors and their integration in wireless body area networks for pervasive healthcare delivery: a review. IEEE Sensors J 2022 Mar 1;22(5):3860-3877. [doi: 10.1109/jsen.2022.3141064]
44. Askar NA, Habbal A, Mohammed AH, Sajat MS, Ziyodulla Yusupov ZY, Kodirov D. Architecture, protocols, and applications of the internet of medical things (IoMT). J Commun 2022;17(11):900-918. [doi: 10.12720/jcm.17.11.900-918]
45. Sethi P, Sarangi SR. Internet of things: architectures, protocols, and applications. J Electr Comput Eng 2017;2017:1-25. [doi: 10.1155/2017/9324035]
46. Noura M, Atiquzzaman M, Gaedke M. Interoperability in internet of things infrastructure: classification, challenges, and future work. In: Proceedings of the 3rd International Conference on IoT as a Service. 2018 Presented at: IoTaaS '17; September 20-22, 2017; Taichung, Taiwan p. 11-18 URL: https://link.springer.com/chapter/10.1007/978-3-030-00410-1_2 [doi: 10.1007/978-3-030-00410-1_2]

XSL•FO
RenderX

47. Ahsan Chishti M, Majid Ahanger A, Qureshi S, Mir AH. Performance analysis of source specific multicast over internet protocol version 6 with internet protocol version 4 in a test bed. In: Proceedings of the 10th Consumer Communications and Networking Conference. 2013 Presented at: CCNC '13; January 11-14, 2013; Las Vegas, NV p. 956-961 URL: https://ieeexplore.ieee.org/document/6488590 [doi: 10.1109/ccnc.2013.6488590]
48. Touati F, Tabish R, Ben Mnaouer A. Towards u-health: an indoor 6LoWPAN based platform for real-time healthcare monitoring. In: Proceedings of the 6th Joint IFIP Wireless and Mobile Networking Conference. 2013 Presented at: WMNC '13; April 23-25, 2013; Dubai, United Arab Emirates p. 1-4 URL: https://ieeexplore.ieee.org/abstract/document/6548958 [doi: 10.1109/wmnc.2013.6548958]
49. Li M, Moll E, Chituc CM. IoT for Healthcare: an architecture and prototype implementation for the remote e-health device management using Continua and LwM2M protocols. In: Proceedings of the 44th Annual Conference of the IEEE Industrial Electronics Society. 2018 Presented at: IECON '18; October 21-23, 2018; Washington, DC p. 2018-2044 URL: https://ieeexplore.ieee.org/document/8591635 [doi: 10.1109/iecon.2018.8591635]
50. Tamri R, Rakrak S. The SDN-MQTT for an interoperable smart home. In: Proceedings of the 3rd International Conference of Computer Science and Renewable Energies. 2021 Jan 25 Presented at: ICCSRE '20; December 22-24, 2020; Agadir, Morocco p. 01031 URL: https://www.e3s-conferences.org/articles/e3sconf/abs/2021/05/e3sconf_iccsre2021_01031/e3sconf_iccsre2021_01031.html [doi: 10.1051/e3sconf/202122901031]
51. Rahman T, Chakraborty SK. Provisioning technical interoperability within ZigBee and BLE in IoT environment. In: Proceedings of the 2nd International Conference on Electronics, Materials Engineering & Nano-Technology. 2018 Presented at: IEMENTech '18; May 4-5, 2018; Kolkata, India p. 1-4 URL: https://ieeexplore.ieee.org/document/8465272 [doi: 10.1109/iementech.2018.8465272]
52. Fernandez-Lopez H, Afonso JA, Correia JH, Simoes R. Remote patient monitoring based on ZigBee: lessons from a real-world deployment. Telemed J E Health 2014 Jan;20(1):47-54 [FREE Full text] [doi: 10.1089/tmj.2013.0059] [Medline: 24229401]
53. Huang YS, Shih M, Shau YW, Lin WT. A distributed continua AHD system with ZigBee/PAN-IF gateway and continua QoS control mechanism. J Sens Actuator Netw 2012 Jul 25;1(2):97-110. [doi: 10.3390/jsan1020097]
54. Nguyen H, Ivanov R, DeMauro S, Weimer J. RePulmo: a remote pulmonary monitoring system. SIGBED Rev 2019 Aug 16;16(2):46-50. [doi: 10.1145/3357495.3357501]
55. Lomotey R, Kazi R, Deters R. Near real-time medical data dissemination in m-Health. In: Proceedings of the International Conference on Management of Emergent Digital EcoSystems. 2012 Presented at: MEDES '12; October 28-31, 2012; Addis Ababa, Ethiopia p. 67-74 URL: https://dl.acm.org/doi/10.1145/2457276.2457290 [doi: 10.1145/2457276.2457290]
56. Umberfield EE, Staes CJ, Morgan TP, Grout RW, Mamlin BW, Dixon BE. Syntactic interoperability and the role of syntactic standards in health information exchange. In: Dixon BE, editor. Health Information: Exchange Navigating and Managing a Network of Health Information Systems. New York, NY: Academic Press; 2023:236.
57. Lv T, Yan P, He W. Survey on JSON data modelling. J Phys Conf Ser 2018 Aug 30;1069:012101. [doi: 10.1088/1742-6596/1069/1/012101]
58. Lu X, Gu Y, Zhao J, Yu N, Jia W. Research and implementation of medical information format conversion based on HL7 Version 2.x. In: Proceedings of the 2011 International Conference on Computer Science and Service System. 2011 Presented at: CSSS '11; June 27-29, 2011; Nanjing, China p. 2440-2443 URL: https://ieeexplore.ieee.org/document/5974909 [doi: 10.1109/csss.2011.5974909]
59. Lubamba C, Bagula A. Cyber-healthcare cloud computing interoperability using the HL7-CDA standard. In: Proceedings of the 2017 IEEE Symposium on Computers and Communications. 2017 Presented at: ISCC '17; July 3-6, 2017; Heraklion, Greece p. 105-110 URL: https://ieeexplore.ieee.org/document/8024513 [doi: 10.1109/iscc.2017.8024513]
60. Rubí JN, Gondim PR. Interoperable internet of medical things platform for e-health applications. Int J Distrib Sens Netw 2020 Jan 07;16(1):155014771988959. [doi: 10.1177/1550147719889591]
61. Iroju O, Soriyan A, Gambo I. Ontology matching: an ultimate solution for semantic interoperability in healthcare. Int J Comput Appl 2012 Aug 30;51(21):7-14. [doi: 10.5120/8325-1707]
62. Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. Yearb Med Inform 2018 Aug;27(1):129-139 [FREE Full text] [doi: 10.1055/s-0038-1667077] [Medline: 30157516]
63. Min L, Tian Q, Lu X, An J, Duan H. An openEHR based approach to improve the semantic interoperability of clinical data registry. BMC Med Inform Decis Mak 2018 Mar 22;18(Suppl 1):15 [FREE Full text] [doi: 10.1186/s12911-018-0596-8] [Medline: 29589572]
64. Vorisek CN, Lehne M, Klopfenstein SA, Mayer PJ, Bartschke A, Haese T, et al. Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review. JMIR Med Inform 2022 Jul 19;10(7):e35724 [FREE Full text] [doi: 10.2196/35724] [Medline: 35852842]
65. Zampognaro P, Paragliola G, Falanga V. Definition of an FHIR-based multiprotocol IoT home gateway to support the dynamic plug of new devices within instrumented environments. J Reliable Intell Environ 2021 Dec 07;8(4):319-331. [doi: 10.1007/S40860-021-00161-2]

XSL•FO
RenderX

66. Bender D, Sartipi K. HL7 FHIR: an Agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013 Presented at: CBMS '13; June 20-22, 2013; Porto, Portugal p. 326-331 URL: https://ieeexplore.ieee.org/document/6627810 [doi: 10.1109/cbms.2013.6627810]
67. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019 Jun;94:103188 [FREE Full text] [doi: 10.1016/j.jbi.2019.103188] [Medline: 31063828]
68. Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The use of FHIR in digital health - a review of the scientific literature. Stud Health Technol Inform 2019 Sep 03;267:52-58. [doi: 10.3233/SHTI190805] [Medline: 31483254]
69. Bayramcavus A, Kaya MC, Dogru AH. Interoperability of microservice-based systems. In: Proceedings of the 13th International Conference on Electrical and Electronics Engineering. 2021 Presented at: ELECO '21; November 25-27, 2021; Bursa, Turkey p. 594-598 URL: https://ieeexplore.ieee.org/document/9677712 [doi: 10.23919/eleco54474.2021.9677712]
70. Goethals T, Kerkhove D, van Hoye L, Sebrechts M, de Turck F, Volckaert B. FUSE: a microservice approach to cross-domain federation using docker containers. In: Proceedings of the 9th International Conference on Cloud Computing and Services Science. 2019 Presented at: CLOSER '19; May 2-4, 2019; Heraklion, Crete Greece p. 90-99 URL: https://dl.acm.org/doi/10.5220/0007706000900099 [doi: 10.5220/0007706000900099]
71. Torab-Miandoab A, Samad-Soltani T, Jodati A, Rezaei-Hachesu P. Interoperability of heterogeneous health information systems: a systematic literature review. BMC Med Inform Decis Mak 2023 Jan 24;23(1):18 [FREE Full text] [doi: 10.1186/s12911-023-02115-5] [Medline: 36694161]
72. Sowmya S, Deepika P, Naren J. Layers of cloud – IaaS, PaaS and SaaS: a survey. ResearchGate. 2014. URL: https://www.researchgate.net/publication/264458816 [accessed 2024-04-29]
73. Salvadores M, Alexander PR, Musen MA, Noy NF. BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. Semant Web 2013;4(3):277-284. [Medline: 25214827]
74. Abou-Nassar EM, Iliyasu AM, El-Kafrawy PM, Song OY, Bashir AK, El-Latif AA. DITrust chain: towards blockchain-based trust models for sustainable healthcare IoT systems. IEEE Access 2020;8:111223-111238. [doi: 10.1109/access.2020.2999468]
75. de Bruijn J, Ehrig M, Feier C, Martíns‐Recuerda F. Ontology mediation, merging, and aligning. In: Davies J, Studer R, Warren P, editors. Semantic Web Technologies: Trends and Research in Ontology‐based Systems. Hoboken, NJ: John Wiley & Sons; 2006:95-113.
76. Novo O, Francesco MD. Semantic interoperability in the IoT. ACM Trans Internet Things 2020 Mar 02;1(1):1-25. [doi: 10.1145/3375838]
77. Grgurić A, Mošmondor M, Huljenić D. The SmartHabits: an intelligent privacy-aware home care assistance system. Sensors (Basel) 2019 Feb 21;19(4):907 [FREE Full text] [doi: 10.3390/s19040907] [Medline: 30795587]
78. Hameed SS, Hassan WH, Abdul Latiff L, Ghabban F. A systematic review of security and privacy issues in the internet of medical things; the role of machine learning approaches. PeerJ Comput Sci 2021;7:e414 [FREE Full text] [doi: 10.7717/peerj-cs.414] [Medline: 33834100]
79. Li P, Xu C, Jin H, Hu C, Luo Y, Cao Y, et al. ChainSDI: a software-defined infrastructure for regulation-compliant home-based healthcare services secured by blockchains. IEEE Syst J 2020 Jun;14(2):2042-2053. [doi: 10.1109/jsyst.2019.2937930]
80. Yacchirema DC, Sarabia-Jacome D, Palau CE, Esteve M. A smart system for sleep monitoring by integrating IoT with big data analytics. IEEE Access 2018;6:35988-36001. [doi: 10.1109/access.2018.2849822]
81. Yang G, Jiang M, Ouyang W, Ji G, Xie H, Rahmani AM, et al. IoT-based remote pain monitoring system: from device to cloud platform. IEEE J Biomed Health Inform 2018 Nov;22(6):1711-1719. [doi: 10.1109/JBHI.2017.2776351] [Medline: 29990259]
82. Calbimonte JP, Aidonopoulos O, Dubosson F, Pocklington B, Kebets I, Legris P, et al. Decentralized semantic provision of personal health streams. J Web Semantics 2023 Apr;76:100774. [doi: 10.1016/j.websem.2023.100774]
83. Mendes D, Jorge D, Pires G, Panda R, António R, Dias P, et al. VITASENIOR-MT: a distributed and scalable cloud-based telehealth solution. In: Proceedings of the 5th World Forum on Internet of Things. 2019 Presented at: WF-IoT '19; April 15-18, 2019; Limerick, Ireland p. 767-772 URL: https://ieeexplore.ieee.org/document/8767184/authors#authors [doi: 10.1109/wf-iot.2019.8767184]
84. Chromik J, Kirsten K, Herdick A, Kappattanavar AM, Arnrich B. SensorHub: multimodal sensing in real-life enables home-based studies. Sensors (Basel) 2022 Jan 05;22(1):65 [FREE Full text] [doi: 10.3390/s22010408] [Medline: 35009950]
85. Othmen F, Baklouti M, Lazzaretti AE, Hamdi M. Energy-aware IoT-based method for a hybrid on-wrist fall detection system using a supervised dictionary learning technique. Sensors (Basel) 2023 Mar 29;23(7):6 [FREE Full text] [doi: 10.3390/s23073567] [Medline: 37050627]
86. Pathak N, Misra S, Mukherjee A, Kumar N. HeDI: healthcare device interoperability for IoT-Based e-Health platforms. IEEE Internet Things J 2021 Dec 1;8(23):16845-16852. [doi: 10.1109/jiot.2021.3052066]
87. Meliones A, Maidonis S. DALÍ: a digital assistant for the elderly and visually impaired using alexa speech interaction and TV display. In: Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2020 Presented at: PETRA '20; June 30-July 3, 2020; Corfu, Greece p. 1-9 URL: https://dl.acm.org/doi/10.1145/3389189.3397972 [doi: 10.1145/3389189.3397972]

XSL•FO
RenderX

88. Choi H, Lor A, Megonegal M, Ji X, Watson A, Weimer J, et al. VitalCore: analytics and support dashboard for medical device integration. IEEE Int Conf Connect Health Appl Syst Eng Technol 2021 Dec;2021:82-86 [FREE Full text] [doi: 10.1109/chase52844.2021.00016] [Medline: 35582521]
89. Macis S, Loi D, Raffo L. The HEREiAM tele-social-care platform for collaborative management of independent living. In: Proceedings of the 2016 International Conference on Collaboration Technologies and Systems. 2016 Presented at: CTS '16; October 31-November 4, 2016; Orlando, FL p. 506-510 URL: https://ieeexplore.ieee.org/document/7871032 [doi: 10.1109/cts.2016.0094]
90. Sousa AL, Lopes J, Guimarães T, Santos MF. mHealth: monitoring platform for diabetes patients. Procedia Comput Sci 2021;184:911-916. [doi: 10.1016/j.procs.2021.03.113]
91. Gia TN, Jiang M, Sarker VK, Rahmani AM, Westerlund T, Liljeberg P, et al. Low-cost fog-assisted health-care IoT system with energy-efficient sensor nodes. In: Proceedings of the 13th International Wireless Communications and Mobile Computing Conference. 2017 Presented at: IWCMC '17; June 26-30, 2017; Valencia, Spain p. 1765-1770 URL: https://ieeexplore.ieee.org/document/7986551/ [doi: 10.1109/IWCMC.2017.7986551]
92. Bonetto M, Nicolò M, Gazzarata R, Fraccaro P, Rosa R, Musetti D, et al. I-Maculaweb: a tool to support data reuse in ophthalmology. IEEE J Transl Eng Health Med 2016;4:3800110 [FREE Full text] [doi: 10.1109/JTEHM.2015.2513043] [Medline: 27170913]
93. Sahay RL, Akhtar W, Fox R. PPEPR: plug and play electronic patient records. In: Proceedings of the 2008 ACM symposium on Applied computing. 2008 Presented at: SAC '08; March 16-20, 2008; Ceara, Brazil p. 2298-2304 URL: https://dl.acm.org/doi/10.1145/1363686.1364232 [doi: 10.1145/1363686.1364232]
94. Franz B, Schuler A, Krauss O. Applying FHIR in an integrated health monitoring system. Eur J Biomed Inform 2015;11(02):en51-en55 [FREE Full text] [doi: 10.24105/ejbi.2015.11.2.8]
95. Pinto S, Cabral J, Gomes T. We-care: an IoT-based health care system for elderly people. In: Proceedings of the 2017 IEEE International Conference on Industrial Technology. 2017 Presented at: ICIT '17; March 22-25, 2017; Toronto, ON p. 1378-1383 URL: https://ieeexplore.ieee.org/document/7915565 [doi: 10.1109/icit.2017.7915565]
96. Gomez-Garcia CA, Askar-Rodriguez M, Velasco-Medina J. Platform for healthcare promotion and cardiovascular disease prevention. IEEE J Biomed Health Inform 2021 Jul;25(7):2758-2767. [doi: 10.1109/JBHI.2021.3051967] [Medline: 33449888]
97. Pena Queralta J, Gia TN, Tenhunen H, Westerlund T. Edge-AI in LoRa-based health monitoring: fall detection system with fog computing and LSTM recurrent neural networks. In: Proceedings of the 42nd International Conference on Telecommunications and Signal Processing. 2019 Presented at: TSP '19; July 1-3, 2019; Budapest, Hungary p. 601-604 URL: https://ieeexplore.ieee.org/document/8768883 [doi: 10.1109/tsp.2019.8768883]
98. Verma P, Sood SK. Fog assisted-IoT enabled patient health monitoring in smart homes. IEEE Internet Things J 2018 Jun;5(3):1789-1796. [doi: 10.1109/jiot.2018.2803201]
99. Gupta S, Singh U. Ontology-based IoT healthcare systems (IHS) for senior citizens. Int J Big Data Anal Healthc 2018;6(2):1-17 [FREE Full text] [doi: 10.4018/ijbdah.287604]
100. Kaur PD, Sharma P. IC-SMART: IoTCloud enabled seamless monitoring for Alzheimer diagnosis and rehabilitation SysTem. J Ambient Intell Human Comput 2019 Oct 11;11(8):3387-3403. [doi: 10.1007/S12652-019-01534-5]
101. Tuli S, Basumatary N, Gill SS, Kahani M, Arya RC, Wander GS, et al. HealthFog: an ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments. Future Gener Comput Syst 2020 Mar;104:187-200. [doi: 10.1016/j.future.2019.10.043]
102. Borelli E, Paolini G, Antoniazzi F, Barbiroli M, Benassi F, Chesani F, et al. HABITAT: an IoT solution for independent elderly. Sensors (Basel) 2019 Mar 12;19(5):23 [FREE Full text] [doi: 10.3390/s19051258] [Medline: 30871107]
103. Patel WD, Pandya S, Koyuncu B, Ramani B, Bhaskar S, Ghayvat H. NXTGeUH: LoRaWAN based NEXT generation ubiquitous healthcare system for vital signs monitoring and falls detection. In: Proceedings of the 2018 IEEE Pune Section Conference. 2018 Presented at: PUNECON '18; November 30-December 2, 2018; Pune, India p. 1-8 URL: https://ieeexplore.ieee.org/document/8745431 [doi: 10.1109/punecon.2018.8745431]
104. Shi C, Nourani M, Gupta G, Tamil T. Apnea MedAssist II: a smart phone based system for sleep apnea assessment. In: Proceedings of the 2013 IEEE International Conference on Bioinformatics and Biomedicine. 2013 Presented at: BIBM '13; December 18-21, 2013; Shanghai, China p. 572-577 URL: https://ieeexplore.ieee.org/abstract/document/6732560 [doi: 10.1109/bibm.2013.6732560]
105. Strassner J, Diab WW. A semantic interoperability architecture for internet of things data sharing and computing. In: Proceedings of the 2016 IEEE 3rd Conference on World Forum on Internet of Things. 2016 Presented at: WF-IoT '16; December 12-14, 2016; Reston, VA p. 609-614 URL: https://ieeexplore.ieee.org/document/7845422 [doi: 10.1109/wf-iot.2016.7845422]
106. Dam TA, Fleuren LM, Roggeveen LF, Otten M, Biesheuvel L, Jagesar AR, Dutch ICU Data Sharing Against COVID-19 Collaborators. Augmented intelligence facilitates concept mapping across different electronic health records. Int J Med Inform 2023 Nov;179:105233 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105233] [Medline: 37748329]

XSL•FO
RenderX

107. Tangi L, Combetto M, Martin Bosch J, Rodriguez Müller P. Artificial Intelligence for interoperability in the European public sector. Office of the European Union. 2022. URL: https://op.europa.eu/en/publication-detail/-/publication/3f94b728-640c-11ee-9220-01aa75ed71a1/language-en [accessed 2024-04-29]
108. Rakhman AZ, Nugroho LE, Widyawan, Kurnianingsih. Fall detection system using accelerometer and gyroscope based on smartphone. In: Proceedings of the 1st International Conference on Information Technology, Computer, and Electrical Engineering. 2014 Presented at: ICITACEE '14; November 8, 2014; Semarang, Indonesia p. 99-104 URL: https://ieeexplore.ieee.org/document/7065722 [doi: 10.1109/icitacee.2014.7065722]
109. Shim J, Shim MH, Baek YS, Han TD. The development of a detection system for seniors' accidental fall from bed using cameras. In: Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication. 2011 Presented at: ICUIMC '11; February 21-23, 2011; Seoul, Korea p. 1-4 URL: https://dl.acm.org/doi/10.1145/1968613.1968734 [doi: 10.1145/1968613.1968734]
110. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. P T 2017 Sep;42(9):572-575 [FREE Full text] [Medline: 28890644]
111. Ashfaq Z, Rafay A, Mumtaz R, Hassan Zaidi SM, Saleem H, Raza Zaidi SA, et al. A review of enabling technologies for internet of medical things (IoMT) ecosystem. Ain Shams Eng J 2022 Jun;13(4):101660. [doi: 10.1016/j.asej.2021.101660]
112. Al-Kashoash HA, Kemp AH. Comparison of 6LoWPAN and LPWAN for the internet of things. Aust J Electr Electron Eng 2017 Dec 27;13(4):268-274. [doi: 10.1080/1448837x.2017.1409920]
113. Everson J, Patel V, Adler-Milstein J. Information blocking remains prevalent at the start of 21st Century Cures Act: results from a survey of health information exchange organizations. J Am Med Inform Assoc 2021 Mar 18;28(4):727-732 [FREE Full text] [doi: 10.1093/jamia/ocaa323] [Medline: 33410891]
114. Everson J, Adler-Milstein J. Engagement in hospital health information exchange is associated with vendor marketplace dominance. Health Aff (Millwood) 2016 Jul 01;35(7):1286-1293. [doi: 10.1377/hlthaff.2015.1215] [Medline: 27385246]
115. Afonso A, Alvaréz C, Ferreira D, Oliveira D, Peixoto H, Abelha A, et al. OpenEHR based bariatric surgery follow-up. Procedia Comput Sci 2022;210:271-276. [doi: 10.1016/j.procs.2022.10.149]
116. Vandenberk T, Storms V, Lanssens D, De Cannière H, Smeets CJ, Thijs IM, et al. A vendor-independent mobile health monitoring platform for digital health studies: development and usability study. JMIR Mhealth Uhealth 2019 Oct 29;7(10):e12586 [FREE Full text] [doi: 10.2196/12586] [Medline: 31663862]
117. Seth M, Jalo H, Lee E, Bakidou A, Medin O, Björner U, et al. Reviewing challenges in specifying interoperability requirement in procurement of health information systems. Stud Health Technol Inform 2024 Jan 25;310:8-12. [doi: 10.3233/SHTI230917] [Medline: 38269755]

Abbreviations

6LoWPAN: IPv6 over low-power wireless personal area network
API: application programming interface
BLE: Bluetooth Low Energy
DICOM: Digital Imaging and Communications in Medicine
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
GDPR: General Data Protection Regulation
HBC: home-based care
HIPAA: Health Insurance Portability and Accountability Act
HL7v2: Health Level 7 version 2
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
IEEE: Institute of Electrical and Electronics Engineers
IHE: Integrating the Healthcare Enterprise
IoMS: Internet of Medical Systems
IoMT: Internet of Medical Things
ISO: International Organization for Standardization
LAN: local area network
LOINC: Logical Observation Identifiers Names and Codes
LoRaWAN: long-range wide area network
MQTT: message queuing telemetry transport
OAuth: Open Authorization
OpenICE: Open Integrated Clinical Environment
OWL: Web Ontology Language
PAN: personal area network
PHD: personal health device

XSL•FO
RenderX

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PSAP: public safety answering point
RDF: resource description framework
REST API: representational state transfer application programming interface
SMART: Substitutable Medical Applications, Reusable Technologies
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
WAN: wide area network

Edited by A Mavragani; submitted 10.11.23; peer-reviewed by E Vashishtha, TAR Sure, E Burner, S Rizvi; comments to author 14.03.24; revised version received 09.05.24; accepted 20.11.24; published 23.01.25.

Please cite as:
Seth M, Jalo H, Högstedt Å, Medin O, Sjöqvist BA, Candefjord S
Technologies for Interoperable Internet of Medical Things Platforms to Manage Medical Emergencies in Home and Prehospital Care: Scoping Review
J Med Internet Res 2025;27:e54470
URL: https://www.jmir.org/2025/1/e54470
doi:10.2196/54470
PMID:

©Mattias Seth, Hoor Jalo, Åsa Högstedt, Otto Medin, Bengt Arne Sjöqvist, Stefan Candefjord. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of Multimodal App-Based Interventions on Glycemic Control in Patients With Type 2 Diabetes: Systematic Review and Meta-Analysis

Emma Bodner[1*], MSc; Lena Roth[2*], MSc; Kathleen Wiencke[1,3], MSc; Christian Bischoff[1], MSc; Peter EH Schwarz[2], Dr med habil

[1]Sidekick Health Germany GmbH, Hamburg, Germany

[2]Department for Prevention and Care of Diabetes, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[3]Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[*]these authors contributed equally

Corresponding Author:
Lena Roth, MSc
Department for Prevention and Care of Diabetes
Faculty of Medicine Carl Gustav Carus
Technische Universität Dresden
Fetscherstraße 74
Dresden, 01307
Germany
Phone: 49 49 351 458 2715
Email: lena_maria_klara.roth@tu-dresden.de

Abstract

Background: Digital technologies for type 2 diabetes mellitus (T2DM) care hold great potential to improve patients' health in the long term. Only a subset of telemedicine offerings are digital interventions that meet the criteria for prescribable digitale Gesundheitsanwendung (digital health apps; DiGAs) in Germany. Digital treatments further provide vast amounts of patient data that are important to generate evidence.

Objective: This systematic review aims to analyze the efficacy of multimodal digital therapies that mainly meet the DiGA criteria for patients with T2DM and to elicit the potential of such therapies. This includes evidence from randomized controlled trials (RCTs) as well as from real-world data. The outcome of interest was a reduction in glycated hemoglobin (hemoglobin A_{1c} [HbA_{1c}]; long-term blood glucose measurements).

Methods: A systematic literature search was conducted in the literature bases PubMed, LIVIVO, and Cochrane, based on the predefined PICO (Population; Intervention; Control; Outcome) scheme. Identified studies were assessed for risk of bias, pragmatism, and overall quality of evidence. Meta-analyses were conducted for between group differences using RCTs only, and for within-group differences using RCTs and non-RCTs, to examine the effect of the interventions on HbA_{1c}.

Results: In total, 795 records were identified, of which 24 were eligible for this systematic review and 23 studies were eligible for the meta-analysis. The results of the meta-analyses showed significant and clinically relevant reductions in HbA_{1c} in patients with T2DM. Regarding the between-group difference for HbA_{1c} reduction, the pooled effect of the RCTs showed a reduction of –0.36% (95% CI –0.59% to –0.14%; P<.001), favoring app-based interventions. The average mean within-group reduction in HbA_{1c} was –0.79 (95% CI –1.02 to –0.55), with no significant difference between RCTs (–0.69, 95% CI –1.13 to –0.24) and non-RCTs (–0.87, 95% CI –1.16 to –0.57; P<.01, differences between RCTs and RCTs P=.44). A pragmatism rating showed that both study types were on average (very) pragmatic, that is, close to usual care. However, the overall quality of evidence was low to very low.

Conclusions: This systematic review shows that digital therapies that mainly meet the DiGA criteria can effectively improve HbA_{1c} in patients with T2DM. The integration of digital health care into usual care holds great potential and should be considered as a complementary option to usual care in the future.

Trial Registration: PROSPERO CRD42023440203; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=440203

XSL•FO
RenderX

(J Med Internet Res 2025;27:e54324) doi:10.2196/54324

KEYWORDS

app-based interventions; diabetes mellitus type 2; digital health; real world evidence; pragmatic trials; ehealth; mhealth; HbA1c; systematic review; meta-analysis; interventions; app; glycemic; patients; type 2 diabetes

Introduction

Background

Type 2 diabetes mellitus (T2DM) is one of the leading causes of death as well as disability-adjusted life years [1]. The rapidly rising global prevalence of T2DM in adults, especially in rural and high-income areas, was estimated to be 10.5% in 2021 [2]. Those levels were previously estimated to occur in 2030 [3]. The metabolic disease is characterized by elevated blood glucose levels, which are associated with an increased risk for vascular and cardiovascular complications, such as heart disease, chronic kidney disease, diabetic retinopathy, and diabetic foot ulcers [4,5]. In addition, T2DM is related to the occurrence of mental health disorders, such as depression [6]. This causes T2DM to be associated with an increased mortality rate in patients enduring the disease [4,7]. Consequently, the chronic condition poses a major health burden on the physical and mental health of those affected as well as on their families [3].

There are several known risk factors for developing T2DM. In particular, conditions such as obesity, high blood pressure, and hyperlipidemia, are associated with relevant lifestyle factors such as high-caloric diets, sedentary behavior, smoking, and alcohol consumption [8,9]. To counteract the multiple risk factors of the disease, guidelines describe lifestyle modification measures as the foundation of all therapeutic interventions in patients with T2DM [10]. Behavioral interventions toward a healthy lifestyle play an important role in preventing the onset of T2DM in individuals at high risk and decreasing the development and progression of diabetic complications for patients with manifested T2DM [11].

Multimodal Therapy

A multimodal therapeutic approach for people with obesity and T2DM combines nutritional medical intervention, exercise, and behavioral therapy support and addresses treatment adherence to enhance the long-term success of such interventions [12]. These components of interdisciplinary treatment can be communicated to patients digitally. For example, eating behavior can be tracked and then a reduction in calorie intake and optimization of food composition can be implemented using app features. In addition, training sessions or tasks to increase physical activity can also be integrated. To increase therapy engagement, interventions should not only be multidisciplinary but also patient-centered and address as many risk factors as possible [5,11].

In line with that, behavioral interventions can also reduce the pharmacotherapy escalation. In a trial applying a multimodal approach, including nutritional and behavioral counseling, educational content, digital coaching, and medication management, most participants, 42.7%, decreased their medication, while another 8% eliminated their medications. Hemoglobin A_{1c} (HbA_{1c}; long-term blood glucose) levels were also significantly reduced with 56.1% of participants achieving a HbA_{1c} value of <6.5% [13].

Blood glucose control is a key target for such interventions, to lower cardiovascular risk as well as mortality from T2DM [5]. In adults with T2DM, the HbA_{1c} target value is set individually, depending on various factors, such as diabetes duration, comorbidities, and patient preference, but it lies within the corridor of 6.5%-8.5% [10]. Accordingly, guidelines for the treatment of T2DM formulate lifestyle interventions to manage indicators such as weight, blood pressure, and blood lipids as the foundation of the prevention and treatment of T2DM and also promote self-management and patient autonomy as a therapeutic goal [10,14]. Due to the chronic nature of T2DM, patients must manage their disease in their daily lives independently of medical care [15].

Digital Therapy

In this context, digital technologies, for example, tracking and providing a feedback loop with health care professionals (HCPs) are already included in guidelines as having great potential to improve diabetes care [5,16].

With the Digital Healthcare Act (January 2020) Germany was the first country to use a legal health claim for insured individuals to receive evidence-based treatment in the form of digitale Gesundheitsanwendung (digital health apps [DiGA]), that is, apps prescribed by health care providers [17]. Generally, DiGAs are not primary prevention tools but support the monitoring, discovery, and treatment of diseases, injuries, and disabilities. Other countries, in particular France and Belgium, are following the example and pursuing easier market access for digital solutions in the health care sector [18].

The potential of digital interventions to reduce HbA_{1c} levels has been shown in previous systematic reviews [19-21]. For instance, one umbrella review reports mean reductions in patients with T2DM using telehealth interventions of between −0.01% and −1.13% [21]. The medical purpose of a DiGA is achieved through the interaction of the patient with digital technology [22]. This means that DiGAs are not purely digital communication channels between patients and HCPs, but treatment is largely carried out independently by the patient through an app [20,22]. This means that therapy with DiGAs is independent of the prescription site in which it is used, as the driver of the therapeutic effect is the patient's interaction with the application. By focusing on apps that meet the definition of DiGA, we aim to analyze the evidence for these technologies in diabetes care. This also allows us to expand the current evidence base for the digitalization of usual care, for example, delivering care through phone calls or chats.

Evidence Based on Different Study Designs

For a medical device to be approved as a DiGA in Germany, its effectiveness has to be shown with clinical evidence

XSL•FO
RenderX

generated by randomized controlled trials (RCTs) [23]. Especially before approval or after a device is already approved and on the market, digital applications, including DiGAs, offer new opportunities to generate evidence based on vast amounts of actual user and patient data. In this context, real-world data (RWD) and real-world evidence are increasingly recognized as complementary to RCTs. For example, user data allows us to continuously investigate the efficacy after approval as part of postmarket research, offering additional insights into how medical devices perform [24]. The European Medicines Agency, defines RWD as "routinely collected data relating to a patient's health status or the delivery of health care from a variety of sources other than traditional clinical trials" [25]. One type of such data is health records directly tracked in-app. Evidence generated with RWD is often considered pragmatic and externally valid, making it cost- and time-efficient, particularly for regulatory purposes, such as postmarket surveillance under the European Medical Device Regulation [23,24,26,27].

RCTs provide high internal validity by controlling for confounding variables and minimizing bias through more controlled circumstances and more rigid protocols. Non-RCTs contribute by offering evidence generated outside a study setting and, while less controlled, still provide valuable information on how interventions perform in the real world. Both approaches offer opportunities for implementation in pragmatic settings [28]. To better distinguish between the internal and external validity of trials and the potential balance between them, we use both the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) and Pragmatic Explanatory Continuum Indicator Summary (PRECIS-2) tools, respectively.

Research Question

Current evidence on digital therapies exists for a broad range of intervention types but not specifically for apps that meet the definition of a DiGA addressing T2DM. The objective of this systematic review was to analyze the efficacy of such and to elicit their potential, including explanatory as well as pragmatic studies.

Methods

Search Strategy and Eligibility Criteria

A systematic literature search was conducted in January 2023 to identify evidence on the efficacy of multimodal, app-based lifestyle interventions that meet the definition of DiGAs in reducing HbA_{1c} in patients with T2DM. The search was performed in the electronic databases PubMed, LIVIVO, and Cochrane and the search strategy was individually tailored for each database. A combination of search terms, MeSH (Medical Subject Headings) terms, and appropriate Boolean operators was developed (Multimedia Appendix 1) based on the PICO (Population; Intervention; Control; Outcome) scheme (Table 1) to identify relevant studies. To generate a broader evidence base, not only RCTs but also other study designs were considered eligible (eg, observational studies or analyzing user data). Study selection was carried out per the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [29] and the title-abstract screening and full-text screening of the search results were performed using CADIMA [30]. The screening was independently conducted by EB and LR, and disagreements were resolved through discussion and consensus, involving a third party if necessary.

Table 1. Inclusion and exclusion criteria according to the PICO[a]-scheme.

	Inclusion	Exclusion
Population	Adult patients with diagnosed diabetes mellitus type 2 and HbA_{1c}[b] ≥6.5%	Age <18 years, type 1 diabetes, gestational diabetes, or focus on special subgroups of populations (eg, low income)
Intervention	App-based lifestyle interventions meeting the digitale Gesundheitsanwendung (digital health app) definition, that is, primarily stand-alone (additional human support was possible, if not as the main driver) or multimodal, that is, covering at least two areas of diabetes care (diet, exercise, self-management, etc)	Interventions not primarily app-based (ie, the need for human resources to deliver intervention) or not multimodal (ie, covering only one area of diabetes care, eg, diet intervention), pharmacological interventions, or interventions for prevention
Control	All kinds of control groups or no control group	N/A[c]
Outcome	Glycemic control (HbA_{1c})	HbA_{1c} not reported
Study design	Systematic reviews[d], meta-analyses[d], randomized controlled trials, observational studies, or pilot studies	Study protocols, case reports, surveys, qualitative studies, narrative literature reviews, cross-sectional studies, scoping reviews, economic analyses, or books

[a]PICO: Population, Intervention, Control, Outcome.

[b]HbA_{1c}: hemoglobin A_{1c} (long-term blood glucose).

[c]Not applicable as control was not defined as an exclusion criteria.

[d]Identified systematic reviews were screened for studies that met the inclusion criteria of the present review.

XSL•FO
RenderX

Grading of Pragmatism (PRECIS-2)

Both RCTs and non-RCTs were graded by EB, KW, and LR using the PRECIS-2 tool [31] to represent how explanatory or pragmatic the trials were. The PRECIS-2 tool covers nine domains: "eligibility," "recruitment," "organization," "setting," "flexibility (adherence)," "flexibility (delivery)," "follow-up," "primary endpoint," and "primary analysis." All domains are rated on a 5-point Likert scale from 0 (very explanatory) to 5 (very pragmatic). Generally, a trial is more pragmatic the less strict its protocol is or the fewer additional resources are used, and is therefore closer to usual care. On the other hand, a very explanatory trial would differ vastly from the usual care setting, using strict protocols, a lot of extra resources, specific patients as well as study settings or collecting a vast amount of extra data.

To grade the pragmatism of the studies included in this systematic review, a mean over all domains for each study and across all studies within the respective study type (RCT vs non-RCT) was calculated and was used to compare pragmatism between domains, studies, and study types.

Grading of Evidence (GRADE) and Risk of Bias Assessment

Risk of bias (RoB) assessments for the included studies were independently performed by 2 reviewers, and disagreements were resolved through discussion and consensus. The RCTs were assessed by CB and EB using the RoB 2 tool [32] and nonrandomized and observational studies were assessed by EB and LR, using the Risk Of Bias in Non-Randomized Studies of Interventions (ROBINS-I) tool [33]. The quality of evidence was evaluated by EB, KW, and LR using the GRADE tool [34]. This established tool was used to assess the overall certainty of evidence according to RoB, imprecision, inconsistency, indirectness, and publication bias (see Multimedia Appendix 2) [34]. This was carried out using the GRADEpro Guideline Development Tool (GRADEpro GDT) for the three outcomes: changes in HbA_{1c} between groups in RCTs, within groups in RCTs, and within groups in non-RCTs. In addition, publication bias was evaluated with the help of funnel plots using the R (R Foundation) function *meta::forest.meta* [35].

Statistical Analysis

The data analysis, that is, pooling the results, was performed with R (version 4.3.0) in the RStudio environment.

Two approaches were used to pool the results of the studies. First, the unstandardized between group differences were used to assess the effect of the intervention group (IG) or unit compared with the control group (CG) or unit. Second, with the within-group mean differences (MDs) the absolute change in HbA_{1c} postintervention was quantified. Both effect sizes were not standardized because the included studies measured the relevant outcome on the same scale, that is, HbA_{1c} (in percentage) [36]. The results (mean [SD]) with the highest statistical quality were extracted: (1) for RCTs intention-to-treat (ITTs) analyses that analyze all participants as randomized independent of protocol adherence (with imputation or complete cases analysis) were preferred over per-protocol analyses including only patients adhering to the protocol and hence hold the potential of biased "best case" results and (2) generally adjusted values were preferred over the potentially confounded.

While *meta::metacont* can calculate the mean between group difference and the respective 95% CIs based on each within-group MD and corresponding SD, the within-group MD and corresponding SE needed to be precalculated before effects could be pooled with *meta::metagen* [37,38]. The function further allowed us to calculate prediction intervals for the expected intervention effect in a single new study [39]. Random-effects models were used to pool results because study heterogeneity cannot be excluded (ie, due to different observation periods and different digital apps). Missing within-group MDs and SDs for the IG (and CG) were calculated from the available information reported in the studies per the following suggested procedures. MDs between IG and CG were calculated based on the formula for within-group differences, that is, the mean of the follow-up value subtracted by the mean of the baseline value [36]. In case 95% CIs, *t* statistic values, or *P* values were available, the SD was calculated based on the formulas provided by Higgins et al [40]. If only baseline or follow-up SDs were available, one was substituted with the other; under the assumption that the intervention did not alter the variance [40]. Assuming that the correlation coefficient between the pre- and postvalue of the IGs was similar, missing SDs for the within-group differences were calculated by the mean of the baseline and follow-up SDs [40].

For heterogeneity assessment, the I^2 statistic was used. It quantifies the amount of variation in the results that is not random. Generally, an I^2 below 40% can be considered low, an I^2 between 30%-60% moderate, and an I^2 between 50%-70% or even 70%-90% high [41]. In case of high heterogeneity outlier analysis based on the study by Harrer et al [42] was conducted and reported.

Results

Study Selection

Overall, 21 trials and 3 systematic reviews met the inclusion criteria and were found to be eligible. Systematic reviews were not included in this review and meta-analysis but were used to identify further eligible trials. Their content was screened and thereby, two further trials were identified and included as cross-references. In total, 23 studies were included in this systematic review and were eligible for meta-analysis (see PRISMA flow diagram in Figure 1 and checklist in Multimedia Appendix 3).

XSL•FO
RenderX

Figure 1. PRISMA flow chart of included and excluded studies within this systematic review and meta-analysis. *The content of identified systematic reviews (n=3) was screened to identify further eligible trials that were included as crossreferences. Systematic Reviews were not included in this study. PICO: Population; Intervention; Control; Outcome; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

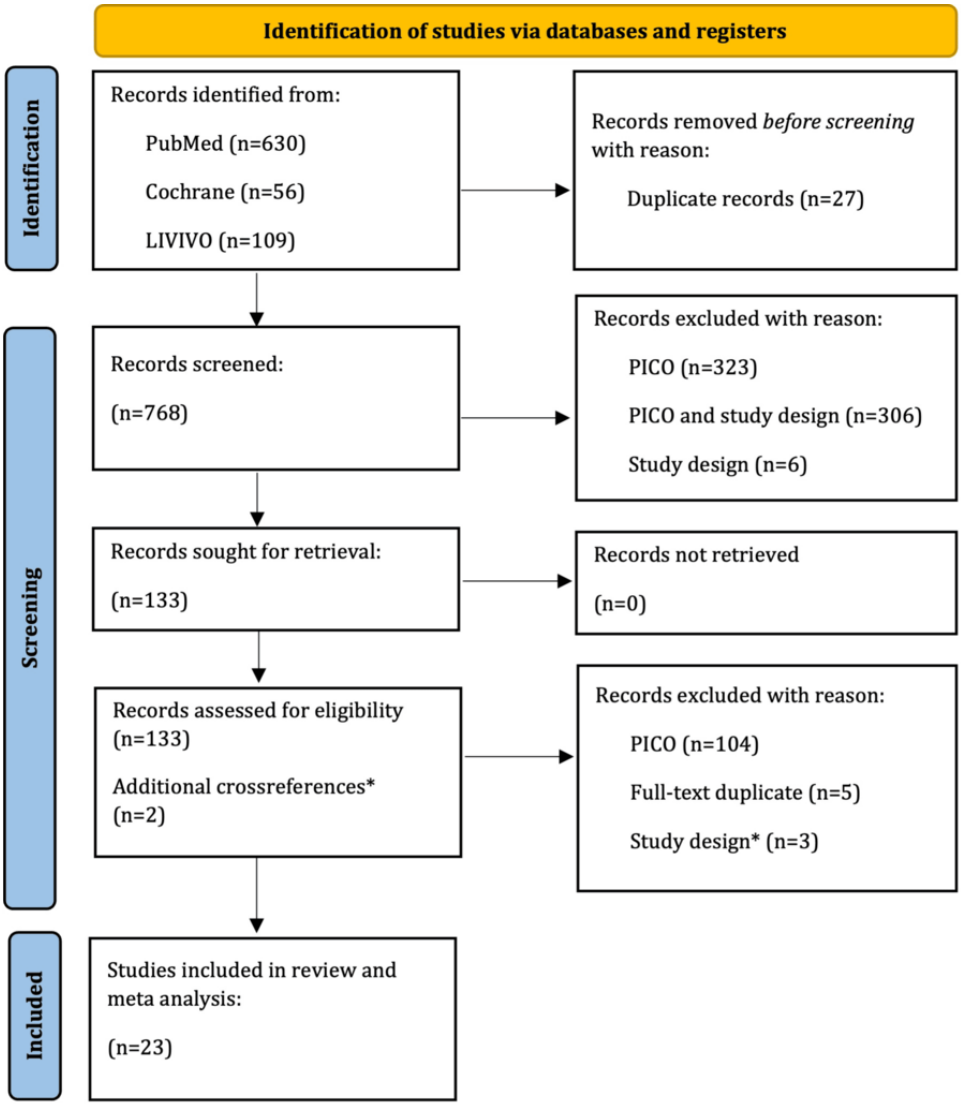

Characteristics of Included Studies

A total of 23 studies were eligible for the review, 12 RCTs and 11 non-RCTs. Of those final studies, 3 were performed as secondary analyses looking at specific subgroups from another study (Lee et al 2021 [43] and Lim et al 2022 [44]) or presented preliminary results of a trial (Torbjørnsen et al 2014 [45]). They were hence not included in the meta-analysis because the main results are covered by other included studies. Another study (Dixon et al 2020 [46]) only looked at HbA_{1c} changes based on different HbA_{1c} baseline categories and was also not included in the meta-analysis because a result for the whole sample could not be obtained and including several subgroups would overrepresent the single study.

Overall, nine of the studies included in the meta-analysis were RCTs and 10 were non-RCTs. Most of the non-RCTs were single-arm observational (pilot) studies and two had an intraindividual CG (Table 2). The average calculated pragmatism of each study is reported in Table 2, while detailed information on each domain is provided in Multimedia Appendix 4. Detailed information on the content and the features of the apps is provided in Multimedia Appendix 5. Most RCTs were performed in Asia, most non-RCTs were performed in the United States. The patient characteristics of study participants and app users are provided in Table 3.

XSL•FO
RenderX

Table 2. Overview of studies included in the systematic review and meta-analysis.

Reference	PRECIS-2[a,b]	Intervention	Intervention duration	Country
RCTs[c]				
[47]	4.3	Sidekick Health (Sidekick Health)	6 months	Iceland
[45,48]	4.4	Few Touch Application (owned by the Norwegian Centre for Integrated Care and Telemedicin)	12 months	Norway
[43,49]	3.7	Switch (Huraypositive Inc)	6 months	South Korea
[50]	4.4	LIBIT (Huraypositiv) + Medilarm (GST Korea)	3 months	South Korea
[51]	4	iCareD (Medical Excellence Inc)	26 weeks	South Korea
[44,52]	3.8	nBuddy Diabetes (HeartVoice)	6 months	Singapore
[53]	4.3	N/A[d] (developed for research)	12 weeks	Indonesia
[54]	4.2	DialBetics (Department of Ubiquitous Health Informatics, NTT DOCOMO)	12 weeks	Japan (Tokyo)
[55]	4.7	N/A (developed for research)	6 months	China
Non-RCTs				
[56][e]	4.3	Wellthy CARE (Wellthy Therapeutics Pvt Ltd)	4 months	India
[57][f]	4.3	Time2Focus (Focus-Complementary Medicine)	12 weeks	United States
[58][e]	4.3	FareWell (Better Therapeutics LLC)	12 weeks	United States
[59][g]	4.3	Vitadio (Vitadio Health)	3 months	Germany
[46][h]	4.9	Onduo Virtual Diabetes Clinic (Verily Life Sciences LLC)	N/A	United States
[60][h]	4.9	BlueStar (WellDoc Inc)	N/A	Mainly the United States
[61][f]	4.3	N/A	12 week	South Korea
[62][e]	4.1	GlycoLeap (Holmusk, Inc)	6 months	Singapore
[63][e]	4.4	Onduo Virtual Diabetes Clinic	4 months	United States
[64][e]	4.4	Vida Health (Vida Health)	N/A	United States
[65][g]	4.7	Vida Health (Vida Health)	N/A	Mainly the United States

[a]The Pragmatic Explanatory Continuum Indicator Summary tool was used to rate pragmatism. Numbers are the mean over all domains per study. The highest grade (5) indicates a very pragmatic trial and the lowest grade (1) indicates a very explanatory trial.

[b]PRECIS-2: Pragmatic Explanatory Continuum Indicator Summary.

[c]RCT: randomized controlled trial.

[d]N/A: not (publicly) available.

[e]Observational study.

[f]Observational pilot study.

[g]Intraindividual observational study.

[h]Analysis of user data.

XSL•FO
RenderX

Table 3. Patient characteristics[a].

Reference	n/N in analysis[b]	Dropout, n (%)	Baseline HbA_{1c} (%), mean (SD)	Age (year), mean (SD)	Gender female (%)	Disease duration (year), mean (SD)	Medication[c]
RCTs[d]							
[47]	30/37	IG[e]: 3 (17); CG[f]: 4 (21)	IG: 7.7 (2); CG: 7.8 (1.9)	51.2 (10.6)	62	IG: 4.9 (5.1); CG: 7.4 (4.4)	O/I[g]
[45,48]	IG: 80/101; CG: 79/101	IG: 11 (24); IG+HCP[h]: 10 (20); CG: 8 (18)	IG: 8.1 (1.1); IG+HCP: 8.2 (1.1); CG: 8.3 (1.2)	57 (12)	41	IG: 11.2 (7.3); IG + HCP: 9.6 (8.4); CG: 9.4 (5.5)	O/I
[43,49]	IG: 136/148; CG: N/A[i]/72	IG: 10 (13.5); CG: 2 (2.7)	IG: 8.2 (1.5); CG: 8 (1.2)	IG: 51.4 (7.9); CG: 52.6 (7.9)	IG: 42; CG: 31	IG: 7.5 (4.8); CG: 9.3 (6)	O/I
[50]	48/50	IG: 2 (8); CG: 0 (0)	IG: 7.4 (0.6); CG: 7.6 (0.8)	IG: 56 (8.1); CG: 63 (8.5)	IG: 56; CG: 56	IG: 7.9 (6.3); CG: 10.8 (8)	O/I
[51]	269/269	IG: 12 (13.2); IG+HCP: 7 (7.7); CG: 16 (18.4)	IG: 8.7 (1.3); IG + HCP: 8.8 (1.4); CG: 8.6 (1.1)	IG: 51.3 (13.1); IG + HCP: 53.6 (11.7); CG: 52.6 (12.1)	IG: 56; IG + HCP: 58; CG: 57	IG: 10.9 (8.3); IG + HCP: 11.9 (7.8); CG: 11.5 (8.2)	O/I
[44,52]	IG: 204/204; CG: N/A/171	IG: 5 (5.1); CG: 4 (3.8)	IG: 7.4 (1.2); CG: 7.5 (1.3)	IG: 51.6 (9.4); CG: 50.8 (10)	IG: 33.3; CG: 37.1	IG: 5.2 (4.5); CG: 4.2 (3.6)	O[j]
[53]	60/60	0	IG: 8 (2); CG: 8.6 (3)	IG: 56.2 (7.63); CG: 54.5 (9.2)	IG: 80; CG: 63	N/A	N/A
[54]	54/54	IG: 3 (11.1); CG: 2 (7.4)	IG: 7.1 (1); CG: 7 (0.9)	IG: 57.1 (10.2); CG: 57.4 (9.4)	IG: 26; CG: 22	IG: 9.6 (7); CG: 8.5 (8)	O/I
[55]	120/120	0	IG: 8.6 (2.3); CG: 8.7 (2.3)	IG: 45.1 (8.7); CG: 45.8 (8.4)	IG: 45; CG: 48	N/A	O/I
Non-RCTs							
[56]	102/102	N/A	8.5	50.8	31	N/A	N/A
[57]	201/201	N/A	IG: 9 (1.2); CG: 9 (1.2)	IG: 56.1 (11); CG: 58.5 (10.8)	IG. 49; CG: 53	N/A	O/I
[58]	97/118	9 (7.6)	8.1 (1.6)	50.7 (9.4)	81	2.6 (1.6)	1.4 (0.9)[k]
[59]	42/60	18 (30)	IG: 7.9 (1.0); CG: 8.2 (1.3)	57 (7.4)	45	7.6 (6.4)	7.7 (1.7) [k]
[46]	N/A/740	N/A	7.7 (1.7)	53.8 (8.8)	63	N/A	7.7 (1.7) [k]
[60]	372/3142	N/A	N/A	N/A	50	N/A	O/I
[61]	29/29	N/A	7.7 (0.7)	53.9 (9.1)	31	N/A	O/I
[62]	83/100	17 (17)	8.8 (1.6)	53.5 (9.6)	50	9.3 (7.3)	O
[63]	55/60	5 (8.3)	8.9 (1)	57.3 (11.6)	40	N/A	O/I
[64]	1128/1934	N/A	9.8 (1.7)	54.1 (10)	65	N/A	N/A
[65]	211/950	78[l]	8.8 (1.6)	59.3 (11.3)	59	N/A	N/A

[a]Patient characteristics are described on the initially included patients at baseline (when available). However, information was not always available for the whole sample (eg, medication or duration of disease). When possible, sample characteristics were provided for the intervention groups and control groups separately.

[b]n refers to the patients included in the analysis whose result is used for the meta-analysis and does not necessarily correspond to the sample after dropout.

[c]Medication is provided as reported in the studies, either as the number of medications (mean [SD]) or the type of medication.

XSL•FO
RenderX

[d]RCT: randomized controlled trial.

[e]IG: intervention group.

[f]CG: control group.

[g]O/I: oral and insulin or a mix of both.

[h]HCP: health care professionals.

[i]N/A: not applicable.

[j]O: oral medication only.

[k]Mean (SD).

[l]Drop out for analysis of real-world users was defined as the users initially included in this study but excluded for the analysis due to missing (follow-up) HbA_{1c} values.

Grading of Pragmatism

To rate the level of pragmatism of the included studies, the PRECIS-2 tool was used (Multimedia Appendix 4). The mean score of pragmatism for RCTs was 4.2 and for non-RCTs 4.5, that is, both study types were on average (very) pragmatic, that is, close to usual care. As a routinely collected parameter in usual care, the HbA_{1c}, is a very pragmatic parameter, resulting in high ratings in the domain of "primary outcome" in all studies. The "eligibility," that is, inclusion and exclusion criteria were also rated as rather pragmatic, as they were close to real-world conditions in almost all studies for the use of a health app; including mostly patients who would be eligible under real-world conditions as well. The organization was mostly rated (very) pragmatic as well, carrying out the study in the usual care environment.

Due to more extensive data collection, RCTs had a slightly lower average score in the domain of "follow-up," mostly due to patient-reported outcomes. Non-RCTs, on the other hand, were rated as less pragmatic than RCTs in the domains of "primary analysis" and "recruitment" due to insufficient inclusion of all available data and sometimes extensive recruitment strategies. The highest rating of pragmatism was achieved when studies analyzed actual user data.

RoB Assessment

Only 3 RCTs were rated as having a low RoB (Figure 2 [47-55]). All of the remaining RCTs were rated as having some concerns, with domain 5 raising concerns in all of them, due to missing study protocols. Missing blinding of participants was present in all studies and some studies also lacked blinding of the outcome assessor (domain 4). We nevertheless decided to rate the RoB as low in this respect, as the outcome of interest, the HbA_{1c}, is a physiological parameter that is less likely to be affected by knowledge of the received intervention, when compared to patient-reported outcomes [26].

Figure 2. Cochrane risk of bias assessment of randomized controlled trials included in the meta-analysis [47-55]. D: domain.

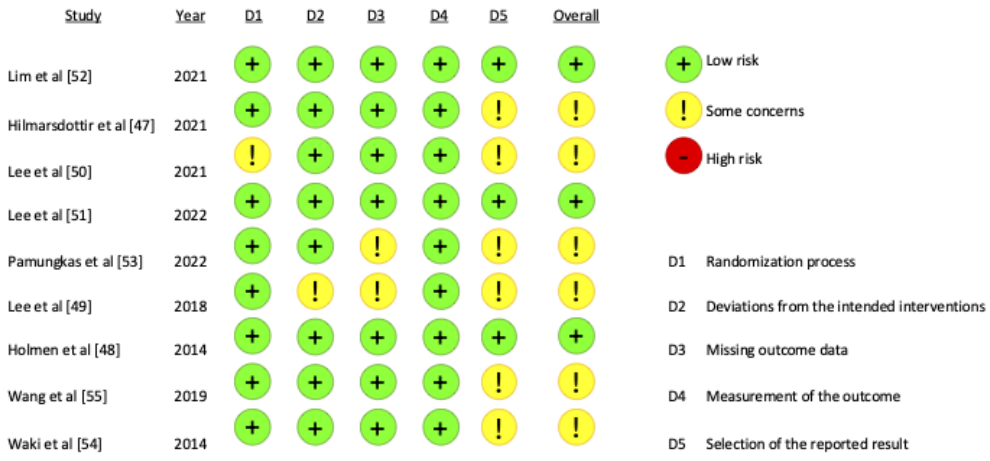

All of the non-RCTs had at least a serious (n=4), if not a critical (n=5) RoB according to ROBINS-I; the detailed ratings can be found in Multimedia Appendix 6. Due to missing information on clear time frames for assessment as well as patient flow, one study could not be ranked [57]. All of the ranked studies had a serious RoB due to confounding because no adjustment for possible confounders was performed or, as in the case of Berman et al [58], adjustment for postintervention variables was performed. The categories leading to critical risks were due to bias in the selection of participants. This particularly relates to observational studies that analyzed app users but excluded all participants who did not have a follow-up HbA_{1c}, leading to exclusion rates of up to almost 80% [60] (Table 3). As it has to be assumed that reporting and continuation of the intervention is influenced by the intervention itself, the RoB of selection was judged to be critical. Another problem of RWD that led to a critical selection bias, was that start and follow-up times did not coincide for all users. Again, the lack of statistical adjustments to correct for missing data resulted in a selection bias and also did not offer evidence that the results were robust to missing data. Two of the studies also had a serious RoB in

XSL•FO
RenderX

the category of measurement outcome due to self-reported HbA_{1c} values.

Grading of Evidence

The grading of evidence was based on the meta-analyses results for the between group differences using 9 RCTs and the within-group differences using 12 RCTs as well as 11 observational studies. Using the GRADE Framework, the levels of certainty for the HbA_{1c} outcomes were rated as low quality of evidence for the RCTs and very low quality of evidence for the observational studies (Table 4).

Table 4. Summary of findings table.

Certainty assessment							Patients, n		Effect		Certainty
Studies, n	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other factors	App-based intervention	Standard care	Relative (95% CI)	Absolute (95% CI)	
9	RCTs[a] (between groups)	Serious[b]	Not serious	Serious[c]	Not serious	None	523	522	—[d]	MD[e] –0.36% (–0.59 to –0.14)	2/4 low
10	Non-RCTs	Very serious[f]	Very serious[g]	Very serious[h]	Not serious	None	2358	271	—	MD –0.87% (–1.16 to –0.57)	1/4 very low
9	RCTs (within-group)	Serious[b]	Not serious	Serious[c]	Not serious	None	523	522	—	MD –0.69% (–1.13 to –0.24)	2/4 low

[a]RCT: randomized controlled trial.

[b]Mainly moderate risk of bias (RoB2-Tool).

[c]Mainly Asia, widely varying intervention duration.

[d]Not applicable.

[e]MD: mean difference.

[f]Serious to critical Risk Of Bias in Non-Randomized Studies of Interventions (ROBINS-I-Tool).

[g]High heterogeneity I^2 82%-94%.

[h]Mainly in the United States, widely varying intervention duration and intervention duration are often unclear.

For the between group differences in RCTs, this rating is due to a serious RoB, illustrated by the low to moderate RoB 2 rating. It also reflects a serious indirectness of the results, which arises from the RCTs being conducted primarily in Asia with widely varying intervention durations of 3 to 12 months. However, imprecision was not detected and neither was inconsistency, based on a low heterogeneity score of I^2=19%. The same applies to publication bias, which was also not detected. For the RCTs used for the within-group differences, the same grading applies.

For the non-RCTs, the very low quality of evidence arose from a very serious RoB rating, pictured by the results of the ROBINS-I-Tool, which showed a serious to critical RoB in the non-RCTs (Multimedia Appendix 6). A high heterogeneity score of I^2=82%-94% showed a high, very serious inconsistency. Due to studies being primarily conducted in the United States with intervention durations widely varying or remaining unclear, as is the nature of RWD, the quality was further downgraded for very serious indirectness. It was further downgraded for publication bias because funnel plots showed some asymmetry in favor of positive results. Imprecision however was not detected.

Results of Meta-Analysis

Between Group Differences

To assess the effect of app-based interventions compared to usual care, the results of the RCTs were pooled. The meta-analysis was performed independent of the intervention period because the number of studies that fell in the same category was too small. Most studies reported results after 6 months, two after 3 months, and one after 12 months. The average MD between IG and CG in the reduction of HbA_{1c} was –0.36% (95% CI –0.59% to –0.14%), favoring app-based interventions (Figure 3). The random effects model yielded significance implying that the group differences were significant (P<.001). The between-study heterogeneity can be assumed to be low with an I^2 of 19% and the test for heterogeneity does not yield significance (P=.27). The prediction interval ranged from –0.57% to –0.16% and hence is very similar to the effect size, indicating robust results.

XSL•FO
RenderX

Figure 3. Meta-analysis of change in HbA1c after app-based lifestyle intervention meeting DiGA definition. DiGA: digitale Gesundheitsanwendung (digital health app); HbA1c: hemoglobin A1c (long-term blood glucose); MD: mean difference.

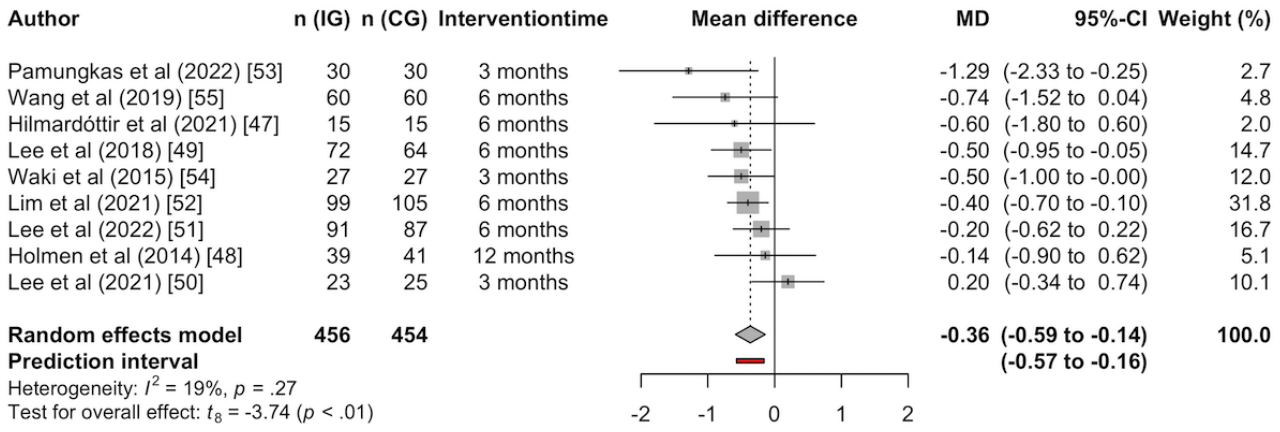

Within-Group Differences

To compare the effects of RCTs and non-RCTs, pre-post effects within the IG of the apps were considered due to a lack of a CG in most non-RCTs. Again, the analysis was performed independent of the duration of the intervention. Non-RCTs had varying observation periods that, on average, were shorter compared to RCTs. The average mean pre-post reduction in HbA_{1c} levels is –0.79 (95% CI –1.02 to –0.55), not differing significantly (P=.44) between RCTs (–0.69, 95% CI –1.13 to –0.24) and non RCTs (–0.87, 95% CI –1.16 to –0.57; Figure 4). The high heterogeneity (I^2=82%-94%) and the prediction interval that also included positive values indicate a potential lack of robustness of the results. The heterogeneity remains moderate to high (RCTs: 55%; non-RCTs: 59%) in the outlier analysis (Multimedia Appendix 7).

Figure 4. Meta-analysis of the pre-post effects within the IG of RCTs and non-RCTs. IG: intervention group; MD: mean difference; RCT: randomized controlled trial.

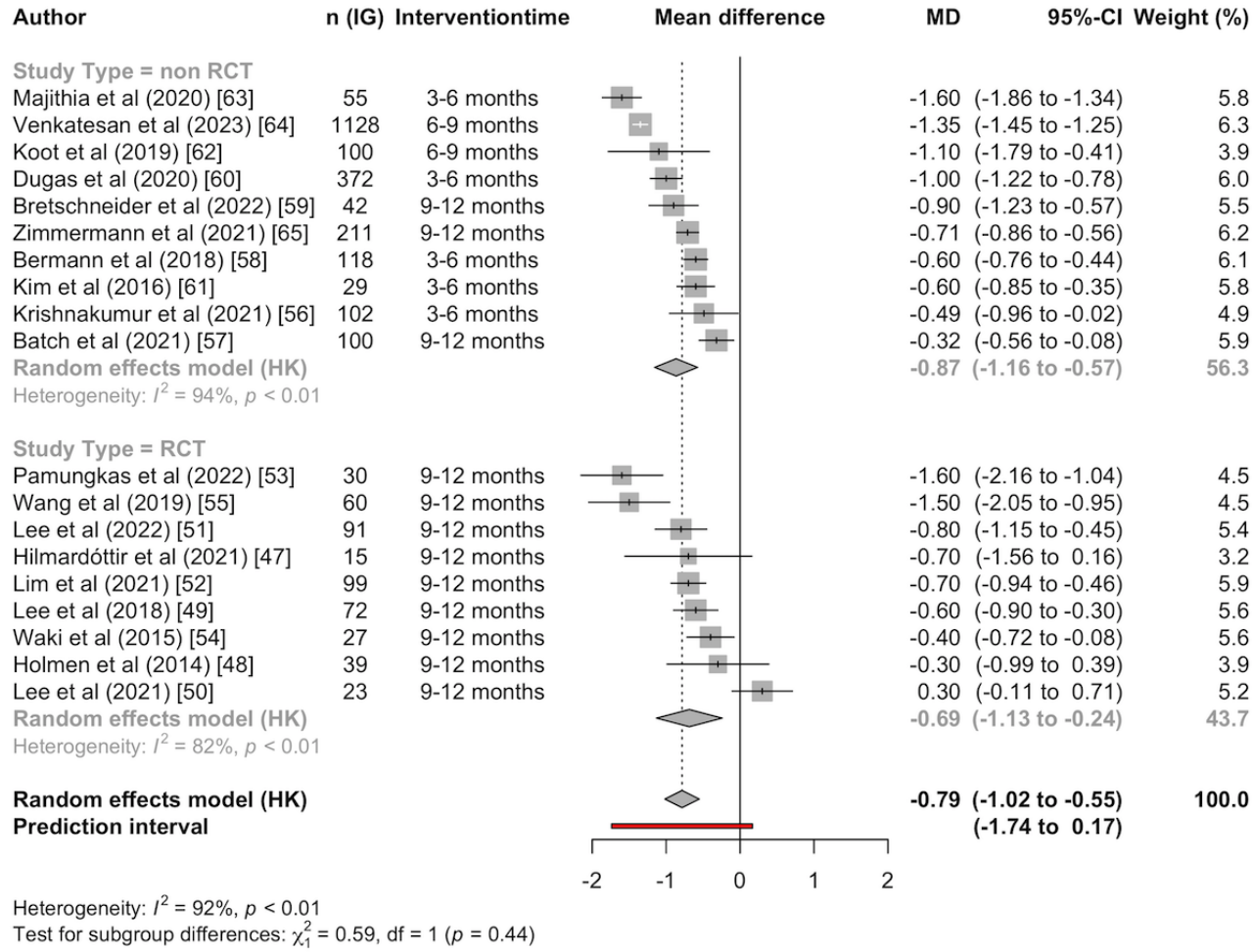

XSL•FO
RenderX

Discussion

Principal Findings

Overall, this systematic review and meta-analysis aimed at giving an overview of the current evidence regarding the efficacy and effectiveness of app-based interventions for the treatment of T2DM. For this, RCTs as well as non-RCTs were included. To our knowledge, this review is the first to include RCTs and non-RCTs for a pooled pre-post effect of digital lifestyle therapy on the HbA_{1c} in patients living with diabetes. Evidence from 9 RCTs shows that app-based interventions for the treatment of T2DM support patients in significantly reducing their HbA_{1c} levels by, on average, –0.36% and had favorable results compared to CGs receiving usual diabetes care. Looking at the within-group differences, patients were able to reduce their HbA_{1c} values on average by –0.79% after 3-12 months. While non-RCTs showed slightly higher reductions compared to the RCTs, the differences between study types were not significant. However, the heterogeneity (I^2 statistic) for the within-group differences across study types, as well as within-study types was high, likely due to differences in the study duration and setting, methodology, and patients' characteristics [66]. In contrast, the RCTs' between-subject analysis resulted in a low heterogeneity. This shows the importance of controlling for baseline characteristics in statistical models to analyze outcomes. The pre-post effects only adjusted for confounding factors within their study population. In contrast, between group differences cancel out several effects unrelated to the intervention. First, the effect of the patient population characteristics itself. Second, the effect of the "usual care" that potentially differs between study sites (nationality, resources, study information provided to the participants, etc). Third, both groups are followed up over the same period, which allows controls for unexpected or spontaneous outcome-related events influencing the disease process of the whole population. As a result, only the between group differences show the isolated efficacy of the intervention, while within-group differences show the effectiveness of the intervention including external factors that might influence study outcomes, that is, the clinical effect under real-world conditions. This review shows that non-RCTs are likely to produce similar outcomes to RCTs per the effectiveness of interventions and as such can complement the evidence from RCTs. To address the limiting factor of a missing comparator in non-RCTs, appropriate artificial CGs could be used, such as matching methods or intraindividual cohorts. In fact, some of the included non-RCTs used the latter method.

Comparison to Prior Work

The average additional mean HbA_{1c} reduction of –0.36% achieved with app-based interventions in comparison to usual care only, is comparable to previous systematic reviews on tele-medical treatment of diabetes that found mean group-differences of –0.44% [19] and –0.52% [21]. To reduce long-term diabetes-related complications, an HbA_{1c} reduction of 0.3% is considered clinically meaningful [67-69]. Stronger reductions in previous reviews could be due to broader inclusion criteria because the current review only included studies on apps that meet the DiGA definition and thus largely function through their technology independently of HCPs. Systematic reviews that analyze the efficacy of app-based interventions for type 1 diabetes mellitus and T2DM show that additional remote access to HCPs [19] as well as a higher frequency and intensity of feedback and interaction are associated with greater HbA_{1c} reductions [20]. Whether or not the feedback is automated or manual, on the other hand, might not be relevant [70].

Pragmatism (PRECIS-2) Versus Quality of the Study (GRADE)

RCTs are designed such that the result is highly controlled, providing strong internal validity. However, this does not mean an RCT cannot be implemented in a real-world setting. The reasons for choosing a non-RCT design can be very diverse, with pragmatic implementation being a major aspect. In our study, the concept of pragmatism was used to judge the external validity of studies with the help of the PRECIS-2 tool. The most pragmatic studies were those that analyzed actual users of the app-based interventions [46,60]. Both RCTs and non-RCTs had, on average, high grades for pragmatism indicating (very) pragmatic trials and good external validity. Due to a lack of transparency and suitable statistical methods to account for missing data, non-RCTs had slightly lower grades than RCTs in the domains "primary analysis" and "recruitment."

The GRADE ratings in both RCTs and non-RCTs further showed potential for improvement in future trials aiming for analyzing in-app user data or studies without a CG. First of all, improved statistical methods should be used to control for baseline covariates. Especially in studies analyzing user data or data collected from health records, clear information on data selection and suitable statistical methods are lacking. Particularly, in the absence of a CG, the handling of missing data needs to be clearly justified (eg, patients who do not use the app or do not provide any data). One way to improve the quality might be the use of preregistered analysis that predefines statistical methods and inclusion or exclusion criteria of patients [23]. Additional sensitivity analyses under different assumptions or conditions may increase the interpretability of results, as might blinding that so far is a major limitation of most digital interventions.

While non-RCTs in the context of explanatory trials show several limitations, they do have strengths in depicting results in a real-world context as pragmatic trials. However, the results of the PRECIS-2 rating showed that non-RCTs and RCTs included in our study can be classified as similarly pragmatic, and as such hold the potential to analyze interventions under real-world conditions to complement RCTs conducted in a controlled setting. Especially, in the context of fast-changing digital interventions, non-RCTs offer a cost- and time-efficient addition to RCTs. It is important to note that, for both RCTs and non-RCTs, the results of the study depend on the specific recruitment strategy and patient characteristics. The findings offer insights into the shortcomings of both RCTs and non-RCTs concerning pragmatism, however, they should be read with care, and translation of the evidence to other care settings should be made with great caution.

XSL•FO
RenderX

Limitations

This study has several limitations. First, it cannot be ruled out that relevant publications were overseen in the course of the conducted search. However, this was counteracted by building extensive search terms, based on the PICO scheme, and by using text words as well as MeSH terms. In addition, the search was also conducted in three different databases, which might have further contributed to limiting the risk of selection bias. Second, the present results are limited to studies without long-term, follow-up periods, generating short-term rather than long-term evidence of the interventions. Other reviews suggest that the initialization of a treatment and as such the immediate short-term outcomes of treatment duration might be associated with better treatment success [21] making long-term and follow-up effects of particular interest. Third, even though the comparator of the included RCTs was theoretically the same, that is, usual care, in practice diabetes care differs between countries and could not be considered thoroughly which might have influenced the results. Due to the limited number of studies that fit the inclusion criteria, the heterogeneity between studies could not be accounted for in further subgroup analyses, for example, based on the time of the interventions. In fact, the results of the GRADE rating further resulted in (very) low evidence for the results of meta-analysis. Therefore, the results should be evaluated with caution. Finally, it is important to note that non-RCTs including real-world evidence exist on a broad continuum of possible settings. Unlike RCTs, which are more clearly defined per their study conditions, they vary greatly in their design and context. The PRECIS-2 tool was used to obtain an estimate of how pragmatic the delivery of digital care was in the different studies, including RCTs and non-RCTs. The PRECIS-2 tool was originally developed for designing pragmatic RCTs. As the authors leave freedom and flexibility in the usage of the tool, we applied it for the retrospective grading for RCTs as well as for non-RCTs. Downgrading of specific dimensions within the PRECIS tool depends on the deviation from the usual care condition, which can also vary across different settings. This makes it difficult to apply coherent criteria to a set of different studies.

Conclusions

Overall, it can be concluded that app-based lifestyle interventions that meet the definition of DiGA can effectively reduce HbA_{1c} in patients with T2DM. This has been shown not only in RCTs but also in non-RCTs. While the latter still have several limitations per their design and statistical analyses, non-RCTs that implement suitable designs and methodologies have the potential to become an important source of complementary evidence, for example, in the context of postmarket analyses or piloting studies.

A DiGA approval requires high-quality evidence with minimal sources for biases, such as RCTs, before manufacturers can claim their medical product as a DiGA. Non-RCTs and in particular the analysis of in-app data can complement evidence from RCTs as a cost and time-efficient source of evidence to continuously monitor clinical outcomes of the medical device after being placed on the market.

Beyond the monitoring of the clinically relevant endpoints of DiGAs, in-app data can be a relevant addition to understanding patients' needs and support postmarket analysis. As such the combination of evidence generated by both RCTs and non-RCTs is gaining relevance to develop the potential of DiGAs, for example, implemented in usual care as hybrid models. Moreover, introducing digital care solutions in the health care system may pave the way for artificial intelligence to further enhance the treatment of T2DM worldwide. Yet, future studies should aim for more methodological transparency and appropriate statistical evaluation procedures and methodologies to account for current limitations of non-RCTs, such as the missing comparator.

Acknowledgments

The authors acknowledge the support of Sidekick Health Germany GmbH (formerly aidhere GmbH) in funding this research.

Data Availability

Data sharing does not apply to this article as no data sets were generated or analyzed during this study.

Authors' Contributions

EB and LR designed this study and conducted the systematic literature search; including the critical evaluation, data extraction, and presentation of the results. Different gradings were performed by EB, LR, CB, and KW as mentioned in the text. LR was responsible for the meta-analysis and its visual representation. KW advised on improvements. EB and LR drafted this paper. All authors critically evaluated and reviewed the article.

Conflicts of Interest

EB, CB, and KW are employees of Sidekick Health Germany GmbH (formerly aidhere GmbH).

Multimedia Appendix 1
Search strings for systematic review.
[DOCX File , 21 KB - jmir_v27i1e54324_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Publication bias.
[PDF File (Adobe PDF File), 365 KB - jmir_v27i1e54324_app2.pdf]

Multimedia Appendix 3
PRISMA checklist for systematic reviews. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[PDF File (Adobe PDF File), 101 KB - jmir_v27i1e54324_app3.pdf]

Multimedia Appendix 4
PRECIS-2 rating. PRECIS: Pragmatic Explanatory Continuum Indicator Summary.
[DOCX File , 32 KB - jmir_v27i1e54324_app4.docx]

Multimedia Appendix 5
Characteristics of included apps.
[DOCX File , 30 KB - jmir_v27i1e54324_app5.docx]

Multimedia Appendix 6
Risk of bias assessment for non-RCTs. RCT: randomized controlled trial.
[DOCX File , 31 KB - jmir_v27i1e54324_app6.docx]

Multimedia Appendix 7
Outlier analyses.
[DOCX File , 20 KB - jmir_v27i1e54324_app7.docx]

References

1. Global Health Estimates 2019: deaths by cause, age, sex, by country and by region, 2000-2019. Geneva: World Health Organization; 2020. URL: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates [accessed 2023-10-02]
2. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119 [FREE Full text] [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
3. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9 edition. Diabetes Res Clin Pract 2019;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
4. Emerging Risk Factors Collaboration, Sarwar N, Gao P, Seshasai SRK, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010;375(9733):2215-2222 [FREE Full text] [doi: 10.1016/S0140-6736(10)60484-9] [Medline: 20609967]
5. Schwarz PEH, Timpel P, Harst L, Greaves CJ, Ali MK, Lambert J, et al. Blood sugar regulation for cardiovascular health promotion and disease prevention: JACC health promotion series. J Am Coll Cardiol 2018;72(15):1829-1844 [FREE Full text] [doi: 10.1016/j.jacc.2018.07.081] [Medline: 30286928]
6. Teli M, Thato R, Rias YA. Predicting factors of health-related quality of life among adults with type 2 diabetes: a systematic review. SAGE Open Nurs 2023;9:23779608231185921 [FREE Full text] [doi: 10.1177/23779608231185921] [Medline: 37448972]
7. Cosentino, Grant PJ, Aboyans V, Bailey C, Ceriello A, Delgado V. 2019 ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: the task force for diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC) and the European Association for the Study of Diabetes (EASD). Rev Esp Cardiol Engl Ed 2020;73(5):404.
8. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, on behalf of the American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of care in diabetes-2023. Diabetes Care 2023;46(Suppl 1):S19-S40 [FREE Full text] [doi: 10.2337/dc23-S002] [Medline: 36507649]
9. Global report on diabetes. Geneva: World Health Organization; 2016. URL: https://apps.who.int/iris/handle/10665/204871 [accessed 2023-02-14]
10. Arzneimittelkommission Der Deutschen Apotheker (AMK), Arzneimittelkommission Der Deutschen Ärzteschaft (AkdÄ), Deutsche Dermatologische Gesellschaft E. V. (DDG), Deutsche Diabetes Gesellschaft E. V. (DDG), Deutsche Gesellschaft Der Plastischen RUÄCEV (DGPRAEC), Deutsche Gesellschaft Für Allgemeinmedizin Und Familienmedizin E. V. (DEGAM), et al. NVL Typ-2-Diabetes ? Teilpublikation der Langfassung, 2. Auflage [Internet]. Bundesärztekammer

XSL•FO
RenderX

(BÄK); Kassenärztliche Bundesvereinigung (KBV); Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). URL: https://www.leitlinien.de/mdb/downloads/nvl/diabetes-mellitus/diabetes-2aufl-vers1.pdfDOI: [accessed 2023-01-18]
11. Landgraf R, Heinemann L, Schleicher E, Gerdes C, Petersmann A, Müller-Wieland D, et al. Definition, klassifikation, diagnostik und differenzialdiagnostik des diabetes mellitus: update 2022. Diabetol Stoffwechs 2022;17(S 02):S98-S110. [doi: 10.1055/a-1789-5615]
12. Aberle J, Lautenbach A, Meyhöfer S, Schmid SM, Selig L, Blüher M. Adipositas und diabetes. Diabetol Stoffwechs 2020;15(S 01):S139-S147. [doi: 10.1055/a-1179-2778]
13. McKenzie AL, Hallberg SJ, Creighton BC, Volk BM, Link TM, Abner MK, et al. A novel intervention including individualized nutritional recommendations reduces hemoglobin A1c level, medication use, and weight in type 2 diabetes. JMIR Diabetes 2017;2(1):e5 [FREE Full text] [doi: 10.2196/diabetes.6981] [Medline: 30291062]
14. National Institute for Health and Care Excellence (NICE). Type 2 diabetes in adults: management. United Kingdom: NICE; 2015. URL: https://www.nice.org.uk/guidance/ng28/resources/type-2-diabetes-in-adults-management-pdf-1837338615493 [accessed 2022-06-29]
15. Schram MT, Baan CA, Pouwer F. Depression and quality of life in patients with diabetes: a systematic review from the European Depression In Diabetes (EDID) research consortium. Curr Diabetes Rev 2009;5(2):112-119 [FREE Full text] [doi: 10.2174/157339909788166828] [Medline: 19442096]
16. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, on behalf of the American Diabetes Association. 7. Diabetes technology: standards of care in diabetes-2023. Diabetes Care 2022;46(Suppl 1):S111-S127 [FREE Full text] [doi: 10.2337/dc23-S007] [Medline: 36507635]
17. Digitale-Versorgung-Gesetz. Bundesregierung. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/D/Digitale-Versorgung-Gesetz_DVG_Kabinett.pdf [accessed 2020-01-01]
18. Kramer U. DiGA-fast track – blueprint für digitalturbo in Europa? HealthOn. 2023. URL: https://healthon.de/blogs/2023/03/14/diga-fast-track-blueprint-fuer-digitalturbo-europa [accessed 2024-04-16]
19. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR mHealth uHealth 2017;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
20. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. npj Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
21. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]
22. BfArM. Das fast-track-verfahren für digitale gesundheitsanwendungen (DiGA) nach § 139e SGB V. Bundesinstitut für Arzneimittel und Medizinprodukte. 2023. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.html?nn=597198 [accessed 2023-09-04]
23. Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: the changing landscape of the regulatory approval process. Pharmacoepidemiol Drug Saf 2020;29(10):1201-1212 [FREE Full text] [doi: 10.1002/pds.4932] [Medline: 31823482]
24. Wicherski J, Schneider K, Zinserling J, Heß S, Haenisch B, Broich K. Real-world-Daten in der Arzneimittelregulation – aktuelle Entwicklungen und Ausblick. Präv Gesundheitsf 2023:1-8. [doi: 10.1007/s11553-022-01010-7]
25. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. Clin Pharmacol Ther 2019;106(1):36-39 [FREE Full text] [doi: 10.1002/cpt.1426] [Medline: 30970161]
26. Hemkens LG. [Benefit assessment of digital health applications-challenges and opportunities]. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 2021;64(10):1269-1277 [FREE Full text] [doi: 10.1007/s00103-021-03413-x] [Medline: 34524477]
27. Khosla S, Tepie MF, Nagy MJ, Kafatos G, Seewald M, Marchese S, et al. The alignment of real-world evidence and digital health: realising the opportunity. Ther Innov Regul Sci 2021;55(4):889-898 [FREE Full text] [doi: 10.1007/s43441-021-00288-7] [Medline: 33914297]
28. Lange S, Lauterberg J. Pragmatischere randomisierte Studien mit Fokus auf Registerbasierung. Präv Gesundheitsf 2022:1-9. [doi: 10.1007/s11553-022-00974-w]
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
30. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid 2018;7(1):12. [doi: 10.1186/s13750-018-0124-4]
31. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147 [FREE Full text] [doi: 10.1136/bmj.h2147] [Medline: 25956159]
32. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]

XSL•FO
RenderX

33. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
34. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
35. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Chapter 9 publication bias. Doing meta-analysis in R. 2022. URL: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pub-bias.html [accessed 2023-09-25]
36. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Chapter 3 effect sizes. Doing meta-analysis with R: a hands-on guide. 2022. URL: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/effects.html#effect-sizes-in-control-group-designs [accessed 2023-06-27]
37. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019;22(4):153-160 [FREE Full text] [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
38. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Chapter 4 pooling effect sizes. Doing meta-analysis with R: a hands-on guide. 2022. URL: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pooling-es.html [accessed 2023-06-27]
39. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009;172(1):137-159 [FREE Full text] [doi: 10.1111/j.1467-985X.2008.00552.x] [Medline: 19381330]
40. Higgins JPT, Tianjing L, Deeks JJ. Chapter 6: choosing effect measures and computing estimates of effect. Cochrane Handbook for Systematic Reviews of Interventions.: Cochrane; 2022. URL: https://training.cochrane.org/handbook/current/chapter-06 [accessed 2023-06-27]
41. Schünemann H, Brożek J, Guyatt G, Oxman A. Handbook for grading quality of evidence and strength of recommendations. 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2023-06-27]
42. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Chapter 5 between-study heterogeneity. Doing meta-analysis with R: a hands-on guide. 2022. URL: https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/heterogeneity.html [accessed 2023-06-27]
43. Lee MK, Lee DY, Ahn HY, Park CY. A novel user utility score for diabetes management using tailored mobile coaching: secondary analysis of a randomized controlled trial. JMIR mHealth uHealth 2021;9(2):e17573. [doi: 10.2196/17573] [Medline: 33625363]
44. Lim SL, Tay MHJ, Ong KW, Johal J, Yap QV, Chan YH, et al. Association between mobile health app engagement and weight loss and glycemic control in adults with type 2 diabetes and prediabetes (D'LITE Study): prospective cohort study. JMIR Diabetes 2022;7(3):e35039 [FREE Full text] [doi: 10.2196/35039] [Medline: 36178718]
45. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A low-intensity mobile health intervention with and without health counseling for persons with type 2 diabetes, part 1: baseline and short-term results from a randomized controlled trial in the Norwegian part of RENEWING HEALTH. JMIR mHealth uHealth 2014;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
46. Dixon RF, Zisser H, Layne JE, Barleen NA, Miller DP, Moloney DP, et al. A virtual type 2 diabetes clinic using continuous glucose monitoring and endocrinology visits. J Diabetes Sci Technol 2020;14(5):908-911 [FREE Full text] [doi: 10.1177/1932296819888662] [Medline: 31762302]
47. Hilmarsdóttir E, Sigurðardóttir Á, Arnardóttir RH. A digital lifestyle program in outpatient treatment of type 2 diabetes: a randomized controlled study. J Diabetes Sci Technol 2021;15(5):1134-1141 [FREE Full text] [doi: 10.1177/1932296820942286] [Medline: 32680441]
48. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR mHealth uHealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
49. Lee DY, Park J, Choi D, Ahn HY, Park SW, Park CY. The effectiveness, reproducibility, and durability of tailored mobile coaching on diabetes management in policyholders: a randomized, controlled, open-label study. Sci Rep 2018;8(1):3642 [FREE Full text] [doi: 10.1038/s41598-018-22034-0] [Medline: 29483559]
50. Lee SE, Park SK, Park YS, Kim KA, Choi HS, Oh SW. Effects of short-term mobile application use on weight reduction for patients with type 2 diabetes. J Obes Metab Syndr 2021;30(4):345-353 [FREE Full text] [doi: 10.7570/jomes21047] [Medline: 34875628]
51. Lee EY, Cha SA, Yun JS, Lim SY, Lee JH, Ahn YB, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022;24(7):e37430 [FREE Full text] [doi: 10.2196/37430] [Medline: 35900817]
52. Lim SL, Ong KW, Johal J, Han CY, Yap QV, Chan YH, et al. Effect of a smartphone app on weight change and metabolic outcomes in Asian adults with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2021;4(6):e2112417 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12417] [Medline: 34081137]

XSL•FO
RenderX

53. Pamungkas RA, Usman AM, Chamroonsawasdi K, Abdurrasyid. A smartphone application of diabetes coaching intervention to prevent the onset of complications and to improve diabetes self-management: a randomized control trial. Diabetes Metab Syndr 2022;16(7):102537. [doi: 10.1016/j.dsx.2022.102537] [Medline: 35724489]
54. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
55. Wang Y, Li M, Zhao X, Pan X, Lu M, Lu J, et al. Effects of continuous care for patients with type 2 diabetes using mobile health application: a randomised controlled trial. Int J Health Plann Manage 2019;34(3):1025-1035. [doi: 10.1002/hpm.2872] [Medline: 31368137]
56. Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, et al. Evaluating glycemic control in patients of South Asian origin with type 2 diabetes using a digital therapeutic platform: analysis of real-world data. J Med Internet Res 2021;23(3):e17908 [FREE Full text] [doi: 10.2196/17908] [Medline: 33764306]
57. Batch BC, Spratt SE, Blalock DV, Benditz C, Weiss A, Dolor RJ, et al. General behavioral engagement and changes in clinical and cognitive outcomes of patients with type 2 diabetes using the Time2Focus mobile app for diabetes education: pilot evaluation. J Med Internet Res 2021;23(1):e17537 [FREE Full text] [doi: 10.2196/17537] [Medline: 33470947]
58. Berman MA, Guthrie NL, Edwards KL, Appelbaum KJ, Njike VY, Eisenberg DM, et al. Change in glycemic control with use of a digital therapeutic in adults with type 2 diabetes: cohort study. JMIR Diabetes 2018;3(1):e4 [FREE Full text] [doi: 10.2196/diabetes.9591] [Medline: 30291074]
59. Bretschneider MP, Klásek J, Karbanová M, Timpel P, Herrmann S, Schwarz PEH. Impact of a digital lifestyle intervention on diabetes self-management: a pilot study. Nutrients 2022;14(9):1810 [FREE Full text] [doi: 10.3390/nu14091810] [Medline: 35565782]
60. Dugas M, Wang W, Crowley K, Iyer AK, Peeples M, Shomali M, et al. Engagement and outcomes associated with contextual annotation features of a digital health solution. J Diabetes Sci Technol 2022;16(4):804-811 [FREE Full text] [doi: 10.1177/1932296820976409] [Medline: 33355003]
61. Kim EK, Kwak SH, Baek S, Lee SL, Jang HC, Park KS, et al. Feasibility of a patient-centered, smartphone-based, diabetes care system: a pilot study. Diabetes Metab J 2016;40(3):192-201 [FREE Full text] [doi: 10.4093/dmj.2016.40.3.192] [Medline: 27098508]
62. Koot D, Goh PSC, Lim RSM, Tian Y, Yau TY, Tan NC, et al. A mobile lifestyle management program (GlycoLeap) for people with type 2 diabetes: single-arm feasibility study. JMIR mHealth uHealth 2019;7(5):e12965 [FREE Full text] [doi: 10.2196/12965] [Medline: 31127720]
63. Majithia AR, Kusiak CM, Armento Lee A, Colangelo FR, Romanelli RJ, Robertson S, et al. Glycemic outcomes in adults with type 2 diabetes participating in a continuous glucose monitor-driven virtual diabetes clinic: prospective trial. J Med Internet Res 2020;22(8):e21778 [FREE Full text] [doi: 10.2196/21778] [Medline: 32856597]
64. Venkatesan A, Zimmermann G, Rawlings K, Ryan C, Voelker L, Edwards C. Improvements in glycemic control and depressive symptoms among adults with type 2 diabetes: retrospective study. JMIR Form Res 2023;7:e41880 [FREE Full text] [doi: 10.2196/41880] [Medline: 36701178]
65. Zimmermann G, Venkatesan A, Rawlings K, Scahill MD. Improved glycemic control with a digital health intervention in adults with type 2 diabetes: retrospective study. JMIR Diabetes 2021;6(2):e28033 [FREE Full text] [doi: 10.2196/28033] [Medline: 34075880]
66. Cuijpers P, Weitz E, Cristea IA, Twisk J. Pre-post effect sizes should be avoided in meta-analyses. Epidemiol Psychiatr Sci 2017;26(4):364-368 [FREE Full text] [doi: 10.1017/S2045796016000809] [Medline: 27790968]
67. Lind M, Odén A, Fahlén M, Eliasson B. The shape of the metabolic memory of HbA1c: re-analysing the DCCT with respect to time-dependent effects. Diabetologia 2010;53(6):1093-1098. [doi: 10.1007/s00125-010-1706-z] [Medline: 20237754]
68. Lind M, Odén A, Fahlén M, Eliasson B. A systematic review of HbA1c variables used in the study of diabetic complications. Diabetes Metab Syndr Clin Res Rev 2008;2(4):282-293. [doi: 10.1016/j.dsx.2008.04.006]
69. Brinkworth GD, Wycherley TP, Taylor PJ, Thompson CH. A health care professional delivered low carbohydrate diet program reduces body weight, haemoglobin A1c, diabetes medication use and cardiovascular risk markers-a single-arm intervention analysis. Nutrients 2022;14(20):4406 [FREE Full text] [doi: 10.3390/nu14204406] [Medline: 36297093]
70. Shen Y, Wang F, Zhang X, Zhu X, Sun Q, Fisher E, et al. Effectiveness of internet-based interventions on glycemic control in patients with type 2 diabetes: meta-analysis of randomized controlled trials. J Med Internet Res 2018;20(5):e172 [FREE Full text] [doi: 10.2196/jmir.9133] [Medline: 29735475]

Abbreviations

CG: control group
DiGA: digitale Gesundheitsanwendung (digital health app)
GRADE: Grading of Recommendations, Assessment, Development and Evaluation
HbA_{1c}: hemoglobin A1c (long-term blood glucose)
HCP: health care professional

XSL•FO
RenderX

IG: intervention group
ITT: intention-to-treat
MD: mean difference
MeSH: Medical Subject Headings
PICO: Population; Intervention; Control; Outcome
PRECIS-2: Pragmatic Explanatory Continuum Indicator Summary
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB: Risk of Bias
ROBINS-I: Risk Of Bias in Non-Randomized Studies of Interventions
RWD: real-world data
T2DM: type 2 diabetes mellitus

Edited by A Coristine; submitted 06.11.23; peer-reviewed by V Mohan, H Li, H Napierala; comments to author 10.02.24; revised version received 17.05.24; accepted 30.10.24; published 24.01.25.

Please cite as:
Bodner E, Roth L, Wiencke K, Bischoff C, Schwarz PEH
Effect of Multimodal App-Based Interventions on Glycemic Control in Patients With Type 2 Diabetes: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e54324
URL: https://www.jmir.org/2025/1/e54324
doi:10.2196/54324
PMID:

©Emma Bodner, Lena Roth, Kathleen Wiencke, Christian Bischoff, Peter EH Schwarz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Use of AI in Mental Health Services to Support Decision-Making: Scoping Review

Hassan Auf[1], MPH; Petra Svedberg[1], PhD; Jens Nygren[1], PhD; Monika Nair[1], PhD; Lina E Lundgren[2], PhD

[1]Halmstad University, School of Health and Welfare, Halmstad, Sweden

[2]School of Business, Innovation and Sustainability, Halmstad University, Halmstad, Sweden

Corresponding Author:
Hassan Auf, MPH
Halmstad University
School of Health and Welfare
Box 823, Kristian IV:s väg 3
Halmstad, 30118
Sweden
Phone: 46 35167100
Email: hassan.auf@hh.se

Abstract

Background: Recent advancements in artificial intelligence (AI) have changed the care processes in mental health, particularly in decision-making support for health care professionals and individuals with mental health problems. AI systems provide support in several domains of mental health, including early detection, diagnostics, treatment, and self-care. The use of AI systems in care flows faces several challenges in relation to decision-making support, stemming from technology, end-user, and organizational perspectives with the AI disruption of care processes.

Objective: This study aims to explore the use of AI systems in mental health to support decision-making, focusing on 3 key areas: the characteristics of research on AI systems in mental health; the current applications, decisions, end users, and user flow of AI systems to support decision-making; and the evaluation of AI systems for the implementation of decision-making support, including elements influencing the long-term use.

Methods: A scoping review of empirical evidence was conducted across 5 databases: PubMed, Scopus, PsycINFO, Web of Science, and CINAHL. The searches were restricted to peer-reviewed articles published in English after 2011. The initial screening at the title and abstract level was conducted by 2 reviewers, followed by full-text screening based on the inclusion criteria. Data were then charted and prepared for data analysis.

Results: Of a total of 1217 articles, 12 (0.99%) met the inclusion criteria. These studies predominantly originated from high-income countries. The AI systems were used in health care, self-care, and hybrid care contexts, addressing a variety of mental health problems. Three types of AI systems were identified in terms of decision-making support: diagnostic and predictive AI, treatment selection AI, and self-help AI. The dynamics of the type of end-user interaction and system design were diverse in complexity for the integration and use of the AI systems to support decision-making in care processes. The evaluation of the use of AI systems highlighted several challenges impacting the implementation and functionality of the AI systems in care processes, including factors affecting accuracy, increase of demand, trustworthiness, patient-physician communication, and engagement with the AI systems.

Conclusions: The design, development, and implementation of AI systems to support decision-making present substantial challenges for the sustainable use of this technology in care processes. The empirical evidence shows that the evaluation of the use of AI systems in mental health is still in its early stages, with need for more empirically focused research on real-world use. The key aspects requiring further investigation include the evaluation of the use of AI-supported decision-making from human-AI interaction and human-computer interaction perspectives, longitudinal implementation studies of AI systems in mental health to assess the use, and the integration of shared decision-making in AI systems.

(J Med Internet Res 2025;27:e63548) doi:10.2196/63548

KEYWORDS

artificial intelligence; AI; mental health; decision-making; shared decision-making; implementation; human-computer interaction

XSL•FO
RenderX

Introduction

Background

There has been an increasing interest in the use of artificial intelligence (AI) systems in health care in recent years as part of supporting health care professionals and individuals with health problems to facilitate care processes. This technology is taking advantage of the abundance of health care data to assess and support decision-making across multiple dimensions. Numerous studies have investigated the prospective utility of AI systems in augmenting decision-making across various health care domains. These encompass early detection, diagnostics, treatment modalities, and self-care initiatives [1-3]. While extensive research has examined the ethics, performance, and potential applications of AI models [4-6], there remains a notable gap in the literature regarding the integrative and empirical use of implemented AI systems in health processes.

AI systems for decision-making are equipped to enhance health care and self-care processes. These systems not only have the potential to improve the efficiency of health care processes but also act as a bridge, integrating health care with self-care by enabling seamless information exchange between the 2 domains [7]. From a health care viewpoint, AI systems hold transformative potential for the future, particularly in the identification and diagnosis of health problems, as it can introduce more objectivity and precision [8]. Self-care, in contrast, refers to the personal management of health and daily life activities outside the formal health care system, while self-help encompasses the use of external resources for the management of health and well-being, both within and beyond the formal health care system [9].

From the standpoint of clinical workflows and end users, AI systems present several advantages, including personalized decision-making, enhanced patient-physician communication, and support for shared decision-making (SDM) [10-12]. However, concerns have been raised by both clinicians and patients regarding the widespread adoption of AI systems, underscoring the need for cultural adjustments within health care to facilitate acceptance [10]. Clinicians have expressed reservations about integrating AI systems into decision-making processes, given the challenges around accepting a second opinion without validation from another clinician [12]. Beyond issues of acceptability, the implementation of AI in SDM poses various challenges, including potential delays in decision-making, communication barriers between health care professionals and patients, and uncertainty surrounding the role of AI in decision-making [13]. Research into the use of AI systems in the context of SDM is still in its nascent stages, emphasizing the need for further exploration into how AI systems can effectively support SDM processes in the future [14].

The potential use of AI for SDM holds particular significance in the context of mental health care given the challenges in the identification and management of mental health problems faced by the contemporary mental health care institutions. However, the complex and heterogeneous nature of mental health problems presents considerable obstacles for AI systems in delivering accurate diagnostics, effective treatment guidance, and tailored interventions. Instead, AI systems currently demonstrate greater efficacy in aiding the identification of symptoms of mental health problems rather than diagnosing specific disorders [10,15]. Given this context, it becomes imperative to understand the interplay between the complexities of mental health problems and AI decision-making support mechanisms, emphasizing the incorporation of the perspectives of end users. Such an inclusive approach is essential for guiding the effective implementation of AI systems in mental health settings [16]. Numerous studies underscore the necessity for a nuanced understanding of AI system use within real-world mental health care to inform future implementation strategies [1,17-19].

Objectives

Despite the potential promises and inherent challenges associated with AI systems in supporting decision-making processes within mental health care, there remains a notable research gap concerning their practical utility. Fundamental questions persist regarding the types of decisions AI systems can support in real-world environments, the extent to which these systems engage patients in the decision-making process, and the methodologies used for their evaluation. Given this backdrop, this study aims to explore the empirical evidence for the use of AI systems to support different types of decision-making in mental health. This includes how it has been researched, applied, and evaluated for use and implementation, contributing to the understanding and improvement of future design, development, and implementation of AI systems in mental health.

Methods

Study Design

A scoping review approach was considered the preferable method to address the explorative nature of the study and to systematically map the available empirical studies within the study domain. The design of this scoping review was guided by the 5-phase framework put forward by Arksey and O'Malley [20] to ensure quality and a systematic approach to the study. The framework phases included defining the research questions (RQs); identifying relevant studies; selecting the study; charting the data; and collating, summarizing, and reporting results. The reporting process was conducted following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [21], as shown in Multimedia Appendix 1.

Identifying the RQs

A multidisciplinary research team comprising experts from health care, nursing, AI systems, and implementation research engaged in discussions in the initial phase of the study to identify the scope of the research and RQs. Four main concepts were defined to guide the study: mental health, AI, decision-making, and implementation (Table 1).

XSL•FO
RenderX

Table 1. Operational definitions of the main concepts used in the article.

Concept	Operational definition	Specification
Mental health	• It is conceptualized as including mental health problems, mental illnesses, and disorders for patients and individuals seeking care [22].	• This enabled the exploration of broader types of users with mental health problems, not only patients in clinical settings but also individuals seeking care in a nonclinical environment. The term patient is used to refer to a person receiving or registered with a health care provider. The term individuals refers to people seeking care regardless of being registered in health care.
AI[a] systems	• The AI concept refers to systems described to use AI types relevant to health care, including ML[b], neural networks, deep learning, computational intelligence, supervised ML, or robotics in connection to supporting decision-making utility [23].	• Exploring the AI models as part of a system facilitates understanding the processes within which the AI is integrated into the care settings.
Decision-making	• Decision-making is defined as a cyclical process from situation awareness to acting on a decision [24,25]. • Decision-making support provided by the AI systems are systems that "provide clinicians, staff, and patients with knowledge, patient-specific information, and recommendations" and are "designed to assist decision-makers and interactively support all phases of a human decision-making process." [26,27]. • Decision-making in mental health may include a collaboration between the patients and health care professionals to reach a decision. Shared decision-making in this study is defined as "an approach where clinicians and patients make decisions together using the best available evidence" [28].	• This concept allowed the investigation of AI systems from a decision-making support utility perspective in health care and self-care settings.
Implementation	• The criteria for the level of implementation of AI systems in practice were based on a TRL[c] of ≥6 [28]. A TRL of 6 indicates that the technology has advanced beyond the basic development phase and is now being demonstrated in a relevant real-world or simulated environment, meaning it is now ready for operational testing within its intended context.	This study focuses on researching AI systems specifically in terms of their use and implementation within care settings, with an emphasis on how these technologies are used and adopted in practice. Thus, AI systems with a TRL of ≤5 were excluded, as they are not sufficiently developed to allow end-user interaction or for meaningful evaluation in real-world care environments.

[a]AI: artificial intelligence.

[b]ML: machine learning.

[c]TRL: technology readiness level.

The following three RQs were formulated to address the aim of the study:

1. What are the characteristics of research on AI systems used in relation to support decision-making in mental health?
2. Which types of technologies, decisions, actors, and user flows of AI systems to support decision-making are described?
3. How were the AI-based decision support systems evaluated in research in the mental health context, and what discernible consideration might enable or hinder the adoption or implementation of these systems?

Identifying Relevant Studies

A search strategy was developed within the research team with the assistance of a librarian. Literature searches were carried out using the 5 main health research databases: PubMed, Scopus, PsycINFO, Web of Science, and CINAHL. Keywords used in the searches were guided by the following concepts and Medical Subject Headings terms: *mental health*, *artificial intelligence*, *decision-making*, and *implementation*. Synonyms were joined by the Boolean operator OR; next, we combined the search strings for each keyword with the Boolean operator AND (Multimedia Appendix 2). Eligibility criteria were followed for the search and study selection, as shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Eligibility criteria.

Inclusion criteria

- Written in English
- Peer-reviewed
- Publication date between January 2011 and September 2022
- All artificial intelligence (AI)–based decision support systems intended for use in relation to mental health interventions
- AI tools with user interaction with patients or health professionals aiming to support decision-making
- All mental health care settings and providers
- Empirical study design

Exclusion criteria

- Interventions that use AI for other purposes than to support decision-making
- Interventions that deliver support for other than mental health professionals and care receivers
- Proof-of-concept, viewpoints, or validation studies not related to the implementation and use of AI

Study Selection (Screening)

All articles identified from the searches were uploaded to EndNote (version 20.1; Clarivate) to remove duplicates. The remaining articles were then uploaded to Rayyan (Rayyan Systems Inc), a web-based tool used to screen articles on the abstract level and identify eligible studies for full-text screening. Rayyan was selected as it provides a systematic approach and facilitates organized teamwork [29]. In the first phase of screening, researchers (HA and MN) reviewed the abstracts and held regular follow-up meetings to ensure criteria consistency. In the second phase, full-text articles were reviewed by 2 reviewers in the research team and screened to meet the eligibility criteria and address the RQs. Throughout both phases, articles where there were uncertainties about meeting the inclusion criteria were discussed among the full research team until a consensus was reached. The screening process followed a decision tree process for excluding the studies that did not meet the inclusion criteria, then illustrated in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.

Charting the Data (Data Extraction)

A template sheet was developed to extract data from each of the included articles regarding the characteristics of research to respond to the first RQs, including the author's name, publication date, title, aim, country, study design, care settings, mental health problems, and types of AI. A descriptive summary was performed to answer RQ1 concerning the characteristics.

Collating, Summarizing, and Reporting Results

The reporting of the results for RQ2 was performed in 3 parts. First, we categorized the AI systems described in the included articles according to their utility in supporting decision-making in mental health. Second, we mapped the AI system's utility of supporting decision-making in relation to the end user's interaction and use. This mapping was carried out in terms of types of interaction and decision-making stages as per a conceptual model [24,25], which describe decision-making as a cyclical process spanning from situational awareness to the execution of decisions. Third, we conducted an explorative analysis, mapping data flow and user interaction in relation to the AI system's basic structure as elucidated by Sremac et al [30], comprising 3 primary phases: training data, AI model, and AI output (decision). In the identified articles, the study by Sadeh-Sharvit et al [31] did not provide sufficient data in the reporting for RQs 2 and 3.

An abductive analysis was conducted for RQ3 [32] combining inductive and deductive reasoning when looking at the data by firstly conducting inductive thematic analysis and connecting the emerged themes deductively to the human, organization, and technology-fit (HOT-fit) framework [33]. This framework provides 3 essential components when an AI system is being used; the human factor focuses on the individual user experience of the AI system; the organization factor focuses on the structure and environment evaluation of use; and the technology factor focuses on the quality of use of the system and the information provided by the AI system. These factors consist of subcategories or dimensions, and each dimension includes several elements. First, the content in each article was collated by the first author (HA) and inductively abstracted into themes. The framework was then used to deductively map the themes into the factors and their included elements. Two new elements, *trustworthiness* and *explainability*, were found in the abductive analysis and were included because they were relevant to the study objectives. Explainability refers to the ability of the AI system to provide clear and understandable reasoning of its output or decision support provided to the intended end users [34]. Trustworthiness refers to the extent to which the AI system is trusted by the end users to use, including being reliable, ethical, and dependable for use and in supporting decision-making [35]. The analysis was iteratively and repeatedly discussed between all the authors until consensus was achieved.

XSL•FO
RenderX

Results

Overview and Study Characteristics

The literature search resulted in 1773 articles between the years 2011 and 2022. Following the removal of duplicates, 1217 articles remained for abstract-level screening. Of these, 1159 (95.23%) articles were excluded based on the inclusion and exclusion criteria, leaving 58 (4.77%) articles for full-text screening. A total of 12 (0.99%) articles were found to meet our study eligibility criteria after the completion of the screening process as seen in the PRISMA flowchart (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart—study search and selection process. TRL: technology readiness level.

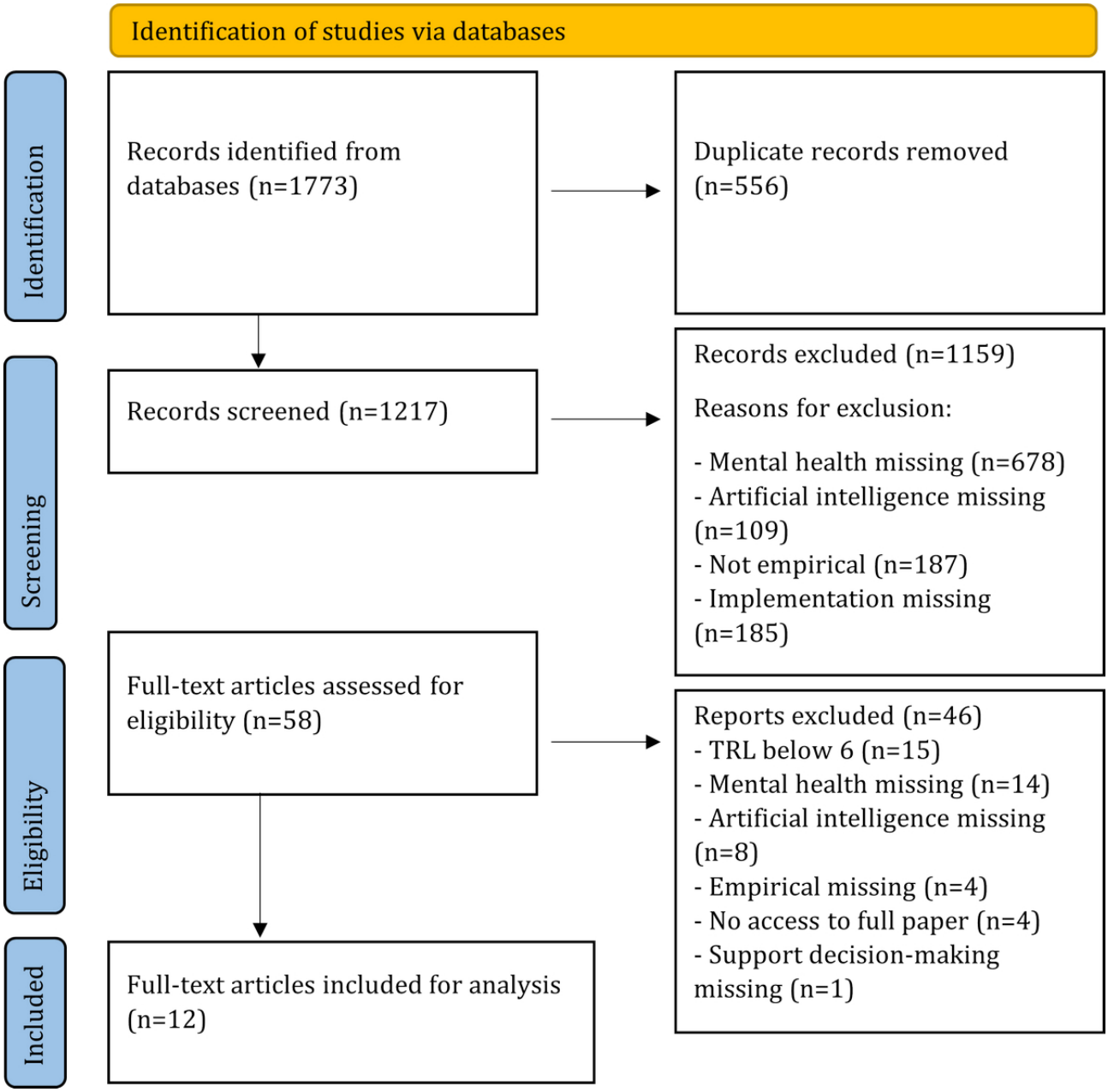

The articles included were published between 2020 and 2022, and a variety of designs were used (Table 2). Most of the articles were from the United States (7/12, 58%); 25% (3/12) were from Canada, 8% (1/12) were from the United Kingdom, and 8% (1/12) were from China. The included studies encompass 3 settings of investigation: health care, self-care (remotely), and simulation environment. The health care studies took place in primary care, emergency department, and specialized mental health care. The self-care studies focused on AI systems using mobile apps or wearables with sensors. Simulation studies were conducted in the psychiatry department within a university-based setting. The spectrum of severity and types of mental health problems found to be supported by the AI systems varied. Depression was the most common mental health problem targeted by the AI system support (5/12, 42%), followed by substance misuse (2/12, 17%), autism spectrum disorder (2/12, 17%), suicide and mental health crisis (2/12, 17%), and psychotherapy for different mental health problems (1/12, 8%).

XSL•FO
RenderX

Table 2. General characteristics of the included articles (N=12).

Study, year, and country	Study aim	Study design	Care setting	Mental health problems	Type of AI[a] (ML[b], DL[c], etc)
Dosovitsky et al [36], 2020; United States	To understand how users engage and are redirected through a chatbot for depression (Tess) to provide design recommendations.	Descriptive study	Self-care and online access	Depression	Not specified (using AI or AI model)
Benrimoh et al [37], 2021; Canada	To explore the use of a simulation center environment in evaluating the usability of (Aifred), particularly its impact on the physician–patient interaction.	Simulation-based study design	Simulation in a specialized care context (research and development)	Depression (major depressive disorder)	DL
Dimeff et al [38], 2021; United States	To evaluate the feasibility, acceptability, and effectiveness of (Jaspr Health) among suicide adults in emergency departments.	Pilot RCT[d]	Emergency department	Suicide	Not specified (using AI or AI model)
Jacobs et al[e] [39], 2021; United States	To investigate the influence of ML and AI on clinician decisions in major depressive disorder treatment.	Experimental study design	Primary and specialized health care	Depression (major depressive disorder)	ML
Popescu et al[e] [40], 2021; Canada	To examine the feasibility of an AI-powered CDSS[f], which combines the operationalized 2016 Canadian Network for Mood and Anxiety Treatments guidelines with a neural network–based individualized treatment remission prediction.	Longitudinal feasibility study	Primary and specialized health care	Depression (major depressive disorder)	DL
Prochaska et al [41], 2021; United States	To evaluate the efficacy and use of Woebot-SUDs[g] in managing SUD relative to a waitlist group in an RCT.	RCT	Self-care and online access	SUD	Not specified (using AI or AI model)
Prochaska et al [42], 2021; United States	To examine the feasibility, acceptability, and preliminary efficacy of Woebot-SUDs in managing SUD.	Development and usability study	Self-care and online access	SUD	Not specified (using AI or AI model)
Deng et al [43], 2022; China	To evaluate the effectiveness of sensory management recommendation system device for children with autism.	Experimental study	Self-care (wearables and app)	Autism	ML
Garriga et al [44], 2022; United Kingdom	To explore the added value of an implemented ML model in a clinical context.	Prospective cohort study	Specialized and primary care	A range of mental health problems	ML
Megerian et al [45], 2022; United States	To test the accuracy of an AI-based software as a medical device designed to aid primary care health care providers in diagnosing autism spectrum disorder.	A double-blinded, multisite, prospective, active comparator cohort study	Primary care	Autism	ML
Sadeh-Sharvit et al [31], 2022; United States	To explore therapists' use of a standard component that is pertinent across most behavioral treatments—prompting clients to summarize their treatment session as a means for consolidating and augmenting their understanding of the session and the treatment plan.	Retrospective study	Specialized care—behavioral health treatments	A range of mental health problems	ML and NLP
Tanguay-Sela et al [46], 2022; Canada	To evaluate the utility of a CDSS as perceived by physicians participating in simulated clinical interactions.	Simulation-based study design	Simulation in a primary and specialized care context (research and development)	Depression (major depressive disorder)	DL

[a]AI: artificial intelligence.

[b]ML: machine learning.

[c]DL: deep learning.

[d]RCT: randomized controlled trial.

[e]Jacobs et al [39] and Popescu et al [40] did not explicitly state the study aim in the articles.

[f]CDSS: clinical decision support system.

[g]SUD: substance use disorder.

XSL•FO
RenderX

Decision Support Utility, Data Flow, and User's Interaction With the AI System

Decision Support Utility

In our findings, 3 types of AI systems emerged based on their decision support utility and the form of output they provide to support decision-making in mental health, as shown in Multimedia Appendix 3. These three types are as follows: (1) *diagnostic and predictive AI*, including AI systems providing support in the identification of mental health states with diagnostic or prognostic output in binary or categorical forms; (2) *treatment selection AI*, including AI systems providing options for treatment with information explaining each option, designed to assess health care practitioners and patients in the therapy selection process; and (3) *self-help AI*, including AI systems designed to support patient and individuals in managing their own mental health, either in health care setting or through self-care practices. The self-help AI in the reviewed studies were all based on conversational agents, which address various mental health problems through psychoeducation and behavioral therapy methods.

These 3 types of AI systems differ in relation to the intended end user's engagement with the AI output to make a decision and with the extent of actionability of decisions the AI output is assisting (Figure 2 [31,36-46]). In Figure 2, *awareness* represents the extent to which the AI supports the end user's understanding of the situation related to decision-making, while *acting on a decision* refers to more direct and explicit AI support on which decision to be made. It could be seen that the 3 systems have different interaction dynamics. The figure shows that different AI systems require varied engagement of end users, including patients and health care professionals, to finalize a decision. Moreover, the urgency and requirement to act on a health-related decision in a care process provided by the AI systems differ as well (Figure 2).

Figure 2. Mapping the decision-support utility of the artificial intelligence (AI) systems on two dimensions: (1) the extent of actionable output the AI system provides and (2) the types and variety of potential end users using the AI system output. * Indicates the studies included: (1) Dosovitsky et al [36]; (2) Benrimoh et al [37]; (3) Dimeff et al [38]; (4) Jacobs et al [39]; (5) Popescu et al [40]; (6) Prochaska et al [41]; (7) Prochaska et al [42]; (8) Deng et al [43]; (9) Garriga et al [44]; (10) Megerian et al [45]; and (12) Tanguay-Sela et al [46].

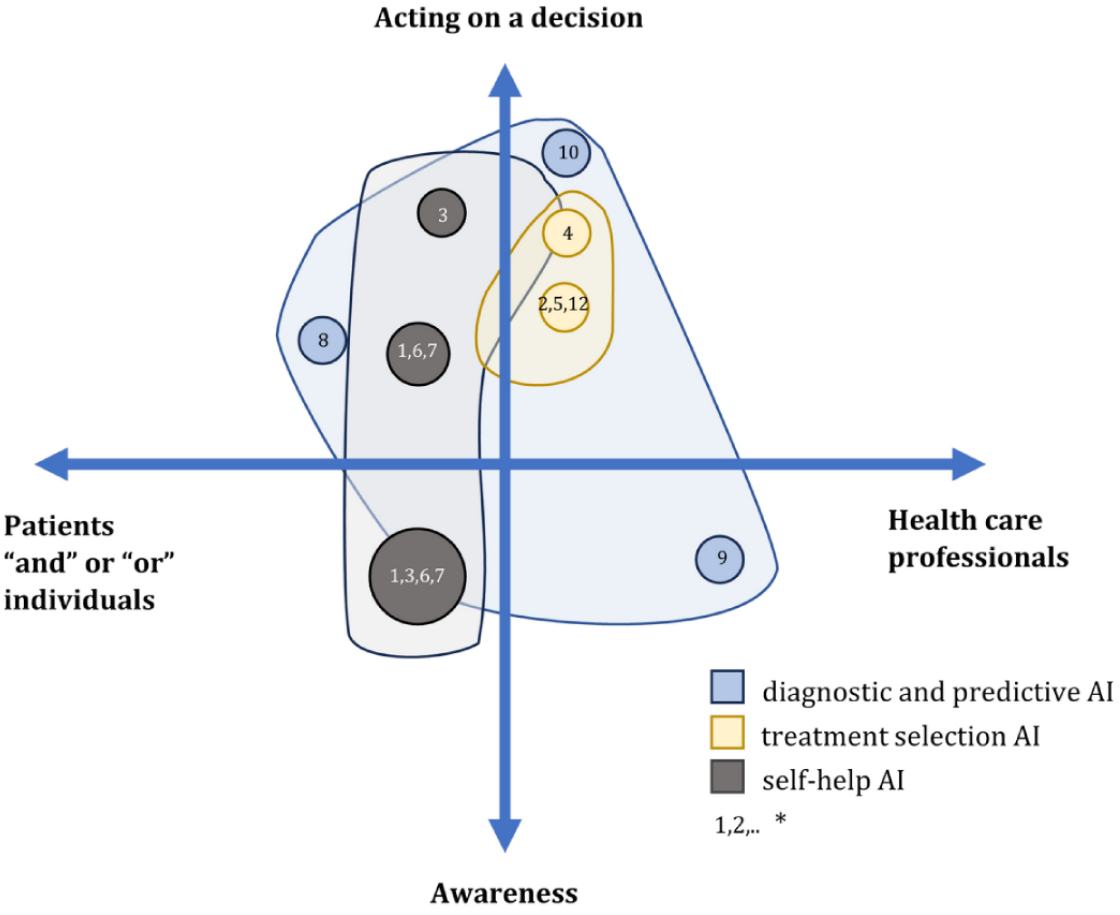

The dot's locations represent the positions of the *intended primary end users* who are expected to interact directly with the AI system interface, and the intended primary decision-making utility (as represented in the 2 dimensions). The dot size reflects the number of AI systems that share the same location. The transparent background highlights the potential variety of involvement of secondary end users who may use the AI output alongside the primary intended user to make a decision.

In reference to Figure 2, the findings show diagnostic and predictive AI systems had the broadest potential number of primary end users from the health care and patients that needed to follow-up with the same AI output until reaching a decision without the need to interact with the AI interface. Treatment

XSL•FO
RenderX

selection AI systems had a broader need for interaction through the same interface by both the health care professionals and the patients. The systems in self-help AI varied between supporting psychoeducation (closer to awareness) and psychotherapy (closer to acting on a therapeutic decision). Compared to the other types of AI systems, this was the most multifunctional type in relation to supporting various decisions.

Data Flow and User's Interaction With the AI System

It can be seen in a closer analysis of the use of AI systems for mental health issues that the 3 systems are different in the flow of data, both as input and output to the AI model; user interaction with the AI system; and the dynamics of the decision makers using the AI output. Four data flow and interaction modalities were found as seen in Figure 3.

Figure 3. Illustration of the 4 data flow and interaction modalities of the 3 types of artificial intelligence (AI) systems. EHR: electronic health record.

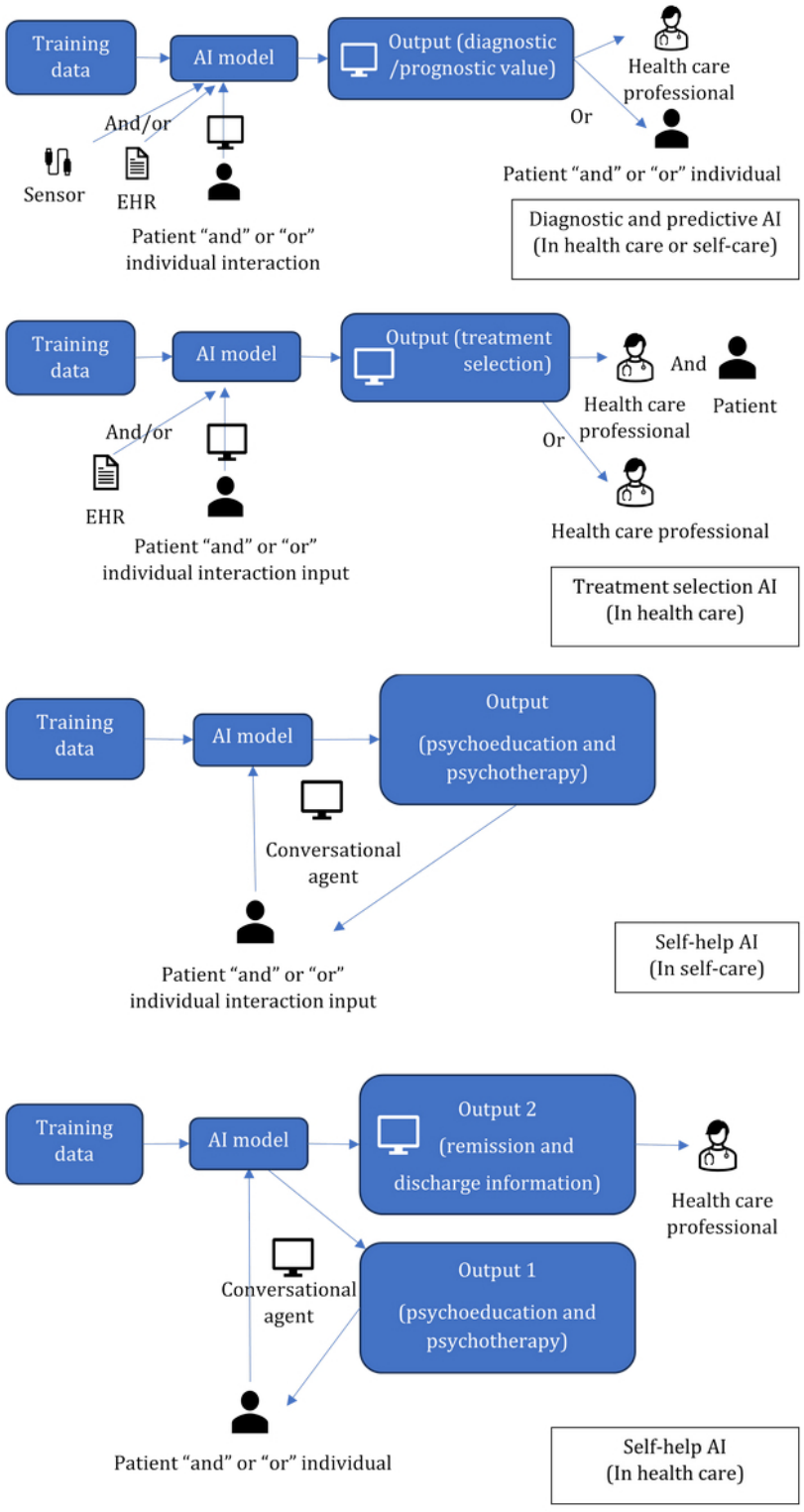

XSL•FO
RenderX

Mapping the structure of the AI system's use journeys shows 3 key differences: the variations in AI input data sources, the use of AI output interface, and the different user interaction interfaces along the AI use journey. First, the sources of AI input data varied in the 3 systems, and a broader range of sources of data were used in the diagnostic and predictive AI. This indicates a more diverse array of new input being incorporated into the AI data analysis. In contrast, self-help AI systems, which were based on conversational interactions, incorporated the end user's textual interaction with a loop of self-help outputs. Second, the use of the AI output interface differed in the 3 AI systems, with more complicated dynamics of use in the treatment selection AI, where both patients and health care professionals potentially needed to interact with the same AI output interface. This is in contrast to the other 2 AI system types, where the interface was only used by 1 type of user at a time, either the health care professional or patient. Third, the interfaces used in the user journey from data input to output revealed variations in the dynamics of interaction. The interfaces for data input and output in the diagnostic and predictive AI and the treatment selection AI were separated, indicating a clear distinction between how information is entered into the system and how it is presented. In the self-help AI systems in health care settings, the input and output were integrated into a single interface for the patient's or individual's use, which in addition showed the potential for providing an extended interface for output for the health care professional to follow-up with the patient interaction.

Evaluation of the Implementation and Use of the AI Systems to Support Decision-Making

Overview

The findings revealed a varied distribution in how the studies evaluated AI-based decision support systems. The human factor in the self-help AI had a predominant focus compared with the other 2 AI types. AI factors were more balanced in representation in diagnostic and predictive AI and treatment selection. The organization factor emerged as the least evaluated aspect in terms of frequency in all AI system types (Table 3). The assessment of the AI systems using the HOT-fit elements, as detailed in Table 3 and Multimedia Appendix 4 [33], reveals the different focus of evaluation across the 3 types of AI systems, illustrating potential challenges and enablers for use and implementation.

XSL•FO
RenderX

Table 3. Summary of the human, organization, and technology-fit (HOT-fit) elements found in the 3 types of artificial intelligence (AI) systems and key findings of potential challenges and enablers of use and implementation.

HOT-fit elements	AI system type	Key findings, challenges (–), and enablers (+)
Human		
Amount and duration	Diagnostic and predictive, treatment selection, and self-help	• (–) Prominent inconsistent duration of use found in self-help AI in self-care settings
Motivation to use	Self-help	• (–) Potential lack of motivation for long-term of use without external support
Acceptance	Self-help	• Different acceptance according to user's socioeconomic status
Recurring use	Self-help	• (–) Unstable recurring use can disturb the evaluation of self-care conversational agent's effectiveness
Expectations and belief	Self-help	• Most users expressed that the conversational agent met some of their goals or needs, only few users expressed that it met most or all their needs
Resistance and reluctance	Self-help	• (–) Potential onboarding resistance and discontinuation after the first or second step in self-care settings
Percentage used	Treatment selection and self-help	• (–) Low completion rates in self-help AI in self-care settings
Voluntaries of use	Self-help	• Users are more likely to engage when initiated by the conversational agent
Knowledge and expertise	Treatment selection	• (+) Knowledge about the AI model can potentially improve the ability to explain the AI output • (–) Previous knowledge about the AI in general may reduce the trust
Overall satisfaction	Diagnostic and predictive and self-help	• (+) Overall satisfaction was relatively high in relation to the amount of help the end users got from the system, the consistency of the system, and if they would use it again or recommend it for others
Perceived usefulness	Diagnostic and predictive, treatment selection, and self-help	• (+) When evaluated most of participants perceived that the AI systems can successfully assist them
Satisfaction with specific functions	Treatment selection and self-help	• Self-help AI satisfaction varied according to the self-help topic • Treatment selection AI satisfaction was related to information the AI provided and can help with the communication with the patient
Decision making satisfaction	Treatment selection	• (–) When evaluated only 60% expressed that the model is somehow satisfactory to support treatment decisions
Organization		
Communication	Treatment selection	• (+) Potential participation in improving the communication of options or decisions
Clinical process	Diagnostic and predictive and treatment selection	• (+) Positive responses on the ability to transform current clinical practice • (–) Can lead to increase of workload by introducing additional steps into their practice
Localization	Diagnostic and predictive	• Potential change of AI performance when used in different environments in relation to stability
Technology		
Ease of use	Diagnostic and predictive and self-help	• (–) Potential difficulty requiring the need for technical support during the use of AI system
Technical support	Diagnostic and predictive	• (–) Expressed need for assistance to get introduced to the system to be able to start using it

XSL•FO
RenderX

HOT-fit elements	AI system type	Key findings, challenges (–), and enablers (+)
Usefulness of system features and functions	Diagnostic and predictive, treatment selection, and self-help	• (+) Providing options, identification of cases and prioritization support features were found useful by participants of the different systems • (–) The need for time to onboard to use the different features were expressed
Data accuracy	Diagnostic and predictive	• (–) Outdated data and misrepresentation of data can lead to lower AI model accuracy
Trustworthiness	Treatment selection	• (–) Affected by the end user's knowledge, the AI processing methods, and onboarding use
Explainability	Diagnostic and predictive and treatment selection	• (–) Lack of explainability can affect trust negatively
Usefulness	Diagnostic and predictive, treatment selection, and self-help	• (–) Potential misalignment between what the end users found useful compared with AI intended purpose of use
Relevance	Diagnostic and predictive	• (+) Most end users found the self-help service relevant to their wants.
Reliability	Treatment selection	• (–) Potential misalignment of AI processing method compared with the health care method
Accuracy	Treatment selection	• Potentially affected by the environment used in, and health care professional's confidence in their decision.
Empathy	Treatment selection and self-help	• (–) Potential skepticism by health care professionals "AI interprets data, but people are not data"

In relation to the 3 types of AI systems, the study identified the following findings.

Diagnostic and Predictive AI

In relation to the evaluation of the use and implementation of diagnostic and predictive AI systems, the articles had a greater focus on evaluating technology elements compared with a less detailed evaluation of human and organization factors.

The human factor was considered only in terms of the amount and duration of use of the AI system, overall satisfaction, and perceived usefulness in the studies by Garriga et al [44] and Megerian et al [45] indicating positive outcomes in this regard.

In relation to the investigation of the technology factor, the main elements were accuracy and ease of use. The accuracy of the AI system was associated with several issues and considerations in relation to the relevance of the data used, illustrating the importance of including different stakeholders in the development of AI models to ensure that relevant data are included. However, in the study by Garriga et al [44] with a predictive AI system based on electronic health record data, the health care professionals highlighted that there was outdated information fed to the AI model and important qualitative data from the patients were not included. While in the study by Deng et al [43], the data used in the diagnostic AI model for children with autism were only based on the child's input, excluding the input from their caregivers. The study illustrated that caregivers were concerned about not having their input for better outcomes.

The environment in which the AI system is used is found to contribute to affecting the accuracy. Deng et al [43] found that the prediction of the autism symptoms was less accurate when using the AI system in a less comfortable environment for the patients, and it gave more accurate output when implementing strategies that adjust the environment for the use of the AI system. Another consideration found affecting accuracy was health care professionals' confidence in their own decisions. Megerian et al [45] showed that greater confidence in one's decision in ruling out the disease over AI output was associated with less accuracy of negative diagnostics (compared with false positive accuracy, which was notably higher).

The ease-of-use element was evaluated in 2 studies reporting that the AI system was relatively easy to use, with the possible need for technical support due to use difficulties [43,44]. The design and ease of use of the AI system was found to influence the health care professional's adoption of decisions provided by the AI system. The low level of understanding of the AI output by a few health care professionals in the study by Garriga et al [44] was a reason for them to not follow-up with the decisions of the AI system.

The evaluation of the organization factor was mainly focused on the clinical process element. Garriga et al [44] reported that the predictive AI tool was used as an additional resource to identify cases, and it led to additional workload due to the need for contacting more patients, arranging visits, and reviewing patient cases in a multidisciplinary team for medical SDM. While in terms of the clinical process change in this paper, the health care professionals reported that the tool did not change the method they followed to assess risk; it rather notified them of the unseen cases that might been missed in the standard practice processes. The identified facilitators for the adoption of the AI system in the study by Garriga et al [44] for the clinical processes were good training and technical support. In contrast, the barriers reported were related to concerns about the

XSL•FO
RenderX

responsibility of acting on the AI output and the potential increase in workload and time when using the AI system.

Treatment Selection AI

Several elements affecting the use and implementation of AI systems were evaluated in this AI type within the 3 factors of the HOT-fit model.

Duration of use, users' knowledge, and perceived usefulness were the main elements in the human factor found affecting the use of AI. The willingness of health care professionals to use the *treatment selection AI* systems was ≤5 minutes per session in the study by Benrimoh et al [37]. Moreover, 40% of the physicians in the study by Benrimoh et al [37] stated that they would only use the AI system in complicated or treatment-resistant patients, while 50% stated they would use it for all cases. The frequency and amount of use of the AI systems in the study by Jacobs et al [39] regarding *treatment selection AI* were associated with the amount of clinician's knowledge in machine learning (ML). The increased familiarity of the health care professionals about ML was associated with the reduction of use of the ML recommendation, despite the higher level of perceived utility of ML. The health care professionals, in the 4 studies looking at the contexts in which treatment selection AI was used, expressed two main ways in which they found the AI system useful: (1) improving patient-clinician communication, which included improving patient understanding for treatment options, making communication less personal, and improving trust in treatment and (2) improving clinical practice such as potentially saving time, providing more objectivity to the practice, confirming or suggesting options, using it as a guideline, or as a source of extra information.

The main technology-related elements in this type of AI system were related to accuracy and trustworthiness. Consistent with the findings in the diagnostic and preventive AI system, the increased familiarity of ML was associated with a decrease in accuracy, despite the increase in clinician confidence in their decisions [39]. Three studies [37,40,46] investigated trustworthiness of the AI system and most of the participants had a high level of trust in using the AI system. Trust in the AI model was reported in the study by Tanguay-Sela et al [46] to be improved with sessions only for participants who reported their trust as unsure in the first session of use, while ratings of either low or high levels of trust did not change for participants who reported low or high levels of trust in the first session (ie, a low level of trust did not improve with increased sessions of use). A lack of trust in particular AI features were found to be an aspect affecting the overall trust of the AI system. The participants in the study by Tanguay-Sela et al [46] reported that their trust in the AI system was reduced due to them finding one of the features of the AI analysis irrelevant. A lack of explainability and evidence of the used models were mentioned in the study by Tanguay-Sela et al [46] as potentially affecting trust in the AI system.

AI system use in terms of the organization factor has been found to affect clinical processes, particularly in patient-physician communication. Some health care professionals in the study by Tanguay-Sela et al [46] had negative opinions about the use of the AI tool stating that it was negatively interfering with the interaction flow with the patient, while others saw it as a valuable tool in the communication process. Benrimoh et al [37] showed that physicians turned their screen to include the patients in the medication selection process even though it was not required in the study design. The AI system was reported in the study by Tanguay-Sela et al [46] to be potentially used in clinical processes other than direct treatment selection for the health care professionals. It was mentioned that the AI system can support clinical processes such as in assisting communication and SDM by visualizing the different options and being able to motivate options for the patients. For this to be valid, the study illustrated that the AI tool needs to be well understood by the physicians, and with transparent explanation of the options, for them to be able to communicate it to the patients.

Self-Help AI

The human factor was the center of focus of evaluation in this type of AI system. Duration of use, percentage used, frequency of use, acceptability, and user satisfaction were the main elements evaluated.

More consistent duration of use and better completion rates of self-help content were found in the conversational agents with health care–supervised type [38], compared with the unguided self-care AI agents [36,41,42]. The duration of use in the study by Dosovitsky et al [36] was inconsistent ranging from 31 minutes to 48 hours to finish 1 module of content. The frequency of use in the studies by Dosovitsky et al [36] and Prochaska et al [42] was inconsistent as well, with a connection found in relation to the intrinsic motivation and passive engagement of the users with the conversational agents [36], where it was found that the users would be more likely to engage in a session when it is initiated by the conversational agent. Satisfaction scored relatively high in all studies looking at satisfaction and perceived usefulness, but in relation to acceptability of use for self-care conversational agents it was found to vary according to the severity of the mental health condition, the socioeconomic status, such as cultural or racial background, marital status, and education [41].

Technology elements such as ease of use, usefulness, and empathy were investigated only in score-based evaluation showing positive outcomes, while organization elements in this AI system type were not found to be investigated.

Discussion

Principal Findings

This study explored the empirical evidence surrounding the use and implementation of AI systems for decision-making support in mental health. The identified studies primarily focused on AI systems developed for health care and self-care settings, with articles predominantly sourced from high-income countries. The state of research on AI systems in mental health is in the preimplementation stage in clinical processes. Although studies have evaluated the use of AI systems and their potential impact on clinical processes, none of the studies have explored the full adoption or long-term implementation of AI systems in care

XSL•FO
RenderX

settings. Of the 12 studies, 8 (67%) evaluated the use of AI systems in real-time care use, while 4 (33%) studies focused on simulations or clinical case studies.

A total of 3 main types of AI systems were categorized according to their utility in mental health: diagnostics and predictive AI, treatment selection AI, and self-help AI. These systems exhibited differences in their use process and interaction with end users. The findings about the types of AI systems and the variety of end users engaged in using these systems were found to be consistent with previous research [47-49], which underscored the potential use of AI in mental health, and emphasized the importance of understanding the sustainable design, which enables the collaboration in decision-making and the integration of these technologies into clinical processes. The AI systems provided support to a diverse range of actors both in health care, including primary care physicians, community psychiatric nurses, psychiatrists, and occupational therapists, as well as outside the health care services, including patients, individuals with mental health problems, and caregivers.

Evaluating the AI systems through the lens of HOT-fit highlighted several challenges in the integration and use of AI systems in care flow. These challenges included potential issues affecting accuracy and use biases, increased demand, physician-patient communication, and engagement with the AI system in self-care. However, despite the existing insights, there remains a pronounced lack of evidence that can help assess the integrative use and implementation of AI systems in different care contexts.

AI System Design for Decision Support and Interaction

The studies demonstrated the diverse interaction dynamics between end users of AI decision-making support systems. Interactions were influenced by the types of users and the data flow within the AI systems. There were variations in how data flows through the 3 types of AI systems. From the methods used to input data, to the types of interfaces presented to users, to the inner functionality of the AI models, and finally, to the nature of the output data—each system painted a unique picture. These variations align with the 4 central human-computer interaction (HCI) aspects identified by Rundo et al [50] for investigating decision-support systems, emphasizing the necessity of integrating HCI principles into AI technology development and implementation to ensure they effectively serve user needs, enhance decision-making outcomes, and ensure the sustainable use of AI in decision-making in mental health care.

Diagnostic and predictive AI systems in the articles showed the greatest diversity in data sources, which can complicate user understanding and create a *black-box* effect due to lack of explainability, which can reduce user acceptance and decision-making quality [51,52]. Furthermore, how the AI model functions to support decision-making based on data input is crucial for acceptance and sustainable use [12,53]. Nevertheless, none of the studies showed how the AI model is being sustainably trained or how the end user's use of the AI systems contributed to the continual learning process of the AI model. This process is a fundamental part of AI or ML apps that provide continual adaptation of the AI model for improvement of its performance with further use [54]. End user–inclusive strategies, such as the human-in-the-loop approach, where health care professionals and patients actively participate in providing feedback to enhance the AI model's functionality, can be useful in addressing this knowledge gap. This ensures that both humans and technology are aligned toward common long-term goals, promoting their joint improvements and effectiveness [54,55].

Three main types of interfaces for end users were discernible when examining the use and data flow within the AI systems: data collection interfaces, AI output visualization interfaces, and combined input and output interfaces in which interaction loops take place as in self-help AI type. These interfaces were designed to be used by a single user (patient or health care professional) or multiple end users. The differences in interface types and end-user interaction can pose challenges in examining these systems and ensuring fluency in care flows. For example, while self-help AI interaction loop interfaces can simplify the flow of use and enhance automation, compared with separated input and output interfaces, it can also complicate measuring interaction effectiveness in the health care [56-58]. Moreover, when multiple end users with different backgrounds use a single AI interface for decision-making, there is a growing need for interfaces to provide on-demand customized explanations to fit the understanding and reasoning of multiple stakeholders [59]. This also points to the need for co-design approaches to involve stakeholders from the early stages of development [60-62].

Using AI system interfaces to support decision-making presents varied dynamics of interaction, not only between the end users and the AI system but also among the end users themselves. Interestingly, none of the reviewed studies included an AI system explicitly designed to facilitate SDM between health care professionals and patients nor did they assess the role of SDM when using AI systems. This absence raises questions about the extent to which collaboration between health care professionals and patients, often highlighted as crucial for improving decision quality in mental health care, is being integrated into AI-assisted processes [51,52,63]. Drake et al [64] emphasized the need to incorporate SDM into the clinical workflows for effective implementation of patients' involvement [64]. As AI systems become more integrated into clinical decision-making workflows, they could alter traditional clinician-patient interactions, introducing new modalities such as clinician-AI, clinician-patient, and patient-AI interactions [65]. Each of these modalities may bring normative challenges that affect the roles of agency, transparency, trustworthiness, and responsibility among end users. Future research should explore how to thoughtfully integrate SDM in the design, implementation, and practical use of AI systems in care processes, while also addressing the broader normative shifts that may arise in clinical workflows because of AI integration.

Types of AI Systems for Supporting Decision-Making

The evaluation of the 3 types of AI systems described in the articles using the HOT-fit framework highlighted several challenges impacting the use and potential implementation of these systems in mental health care processes. General challenges were observed affecting the accuracy of the AI systems due to the use of nonrepresentative or skewed data

XSL•FO
RenderX

input, in addition to the selective use of the AI system participating in potential future biases. Other challenges found more in one AI system than another included the increase in usual care process steps when using the predictive AI systems, the need for adjustment for patient-physician communication when using the treatment selection AI type, and user engagement issues when using the conversational agents in self-help AI.

It was noted in some studies that health care professionals had a selective use of the AI systems for cases with specific severity, or the systems were preferred to be used by specific health care professionals more than others. This selective use with the continuous learning of the AI model may lead to feedback loop biases. This type of bias refers to the distortion of the AI output affected by the skewed use for specific cases over an extended span of time. Consequently, the AI output becomes more accurate to the selected cases and less toward the general use, which can affect the overall accuracy [66]. Furthermore, users in some of the studies described that the AI systems are useful for purposes that deviate from the AI model's primary objective. If the potential deviated use was not considered during the implementation efforts, this issue can lead to *concept drift* bias. This refers to the bias that may happen when the AI processing method differs from the actual use of the AI system leading to degradation of the AI output performance over time [67]. This underscores the need for including health care professionals and patients in the development and implementation phases to ensure that the AI system is designed and implemented in alignment with the sustainable use dynamics [68,69]. Co-design methods for analyzing sociotechnical scenarios can facilitate this by engaging stakeholders in early phases using prototype drafts to validate key AI-related concepts such as its goal and potential risks [62]. Early stakeholder involvement is also essential for building trust and ensuring implementation and adoption. Research shows that collaboration among a diverse, interdisciplinary team, including clinicians, potential AI users, hospital leaders, quality improvement teams, human resources, and IT departments is key to achieving effective implementation. This collective expertise ensures that the AI system meets clinical needs, aligns with organizational goals, integrates smoothly into workflows, and addresses technical, operational, and user concerns [70].

In diagnostic and predictive AI, the utility of the AI systems in detecting mental health state and identifying potentially hidden cases can disrupt common clinical processes by increasing demand and requiring additional steps or time to validate the decision support presented by the AI system. Nevertheless, none of the studies evaluated how this disruption of the care process can be adjusted to and integrated into the usual care routines. If not addressed in the future, this can lead to a dilemma of consequences between undertrusting or overtrusting the AI output in decision-making. Undertrusting of the AI output may require additional steps to validate the decisions provided by the AI system leading to additional workload and time consumption or abandoning the AI system use [12,71]. In contrast, the overreliance on AI systems, particularly when they have a high success rate, can lead to *automation bias*, where decision makers favor AI recommendations over other sources information including their own [72]. This may result in increased dependency on AI, discarding both health care professionals' and patients' judgments and preferences. Such a shift could move clinical care away from person-centered care toward the paternalized style of decision-making raising ethical and safety challenges [73,74]. A prominent example is a commercial AI model implemented in the health care system in the United States, which led to a racial bias discriminating access to health care for millions of patients [75]. These concerns highlight the need for caution in scaling and adopting AI for critical decision-making tasks without robust evidence. Initiatives such as the AI Act aim to address these ethical issues by setting requirements for AI development and implementation. Current recommendations emphasize that final decisions in diagnosis and care should remain with human judgment, with AI serving as an assistive tool, a view supported by numerous studies in mental health care [8,76-78].

Treatment selection AI systems were the only type evaluating the patient-physician relationship when using the AI system. Health care professionals used the AI system in their communication with the patients in some of the studies to discuss options and guidelines. However, this integration of AI decision support can disrupt the communication process, introducing sociorelational challenges into the clinical processes that need to be addressed in the design and implementation of the AI to avoid such challenges [13]. The average time willingness to use the AI system was limited to 5 minutes in one of the studies. This time constraint may present a challenge of integrating the AI support into the patient-physician communication. It highlights the need for efficient and explainable AI that can facilitate communication within a short span of time and tailored to the individualized care processes considering both technical and end-user preferences [79].

A main challenge identified in the studies with self-help AI systems based on conversational agents is inconsistency in user engagement, including frequency of use, duration, and completion rates. The self-help AI systems, primarily used for psychotherapy and psychoeducation, face distinct challenges in both self-care and health care settings. Low engagement can reduce the effectiveness of self-help AI systems, particularly because behavioral change and cognitive behavioral therapy (CBT) techniques require long-term engagement to attain effectiveness [80,81]. While conversational agents show great potential in facilitating mental health services [82,83], several studies indicate a lack of clarity about effectiveness and evidence-based design for therapy and mental health assessment [84-88]. This challenge highlights the need for additional research on long-term effectiveness and reliability of conversational agents across various mental health applications. Jabir et al [89] examined how conversational agents are typically evaluated and found that effectiveness assessment usually includes at least 1 clinical outcome and 1 user experience outcome. Given the need for sustained engagement with CBT-based conversational agents and the findings of our study, we recommend that future studies incorporate a third type of outcome measurement to better validate effectiveness. This additional measurement should focus on behavioral change, which is a fundamental aspect of CBT and chronic health conditions [90,91]. Cole-Lewis et al [92] demonstrated that

XSL•FO
RenderX

engagement with technology can lead to real-world behavior changes and consequently supporting health outcomes. Including behavioral change as an outcome is important to ensure that the mediated behaviors are successfully targeted not just the clinical outcomes or the user engagement. This is crucial because long-term life changes may mislead the association between self-help AI effectiveness and other positive life-related factors. Using behavior change techniques can serve as validated instruments for assessing health or decision-making behavior change, thus guiding the evaluation of conversational agents [93-97].

Overall, from the findings it is evident that successful integration and implementation of AI systems in existing mental health care workflows requires substantial efforts to create evidence regarding its sustainable use. There is a lack of empirical studies evaluating the long-term implementation of AI in mental health care. Nair et al [70] identified strategies and barriers for successful implementation that lead to sustainable use in 3 essential areas (planning, implementing, and sustaining the use). When integrating AI systems for decision-making support, the evaluation of changes in clinical workflows needs to be considered proactively during the planning phase to avoid unseen organizational or increased demand issues when put in place [68]. This step requires the inclusion of multidisciplinary individuals who will engage with the workflow using simulations or prototypes to enable the detection of any barriers or potential risks before putting the integrated AI systems in place [98,99]. Regarding implementation strategies and ensuring sustainable use, it is important to investigate approaches relevant to the context of mental health care processes addressing resistance of change, staff training, and monitoring of AI performance. This can help create evidence participating in planning for sustainable AI systems. Future studies, including case studies and longitudinal implementation research, are necessary to generate evidence for how to successfully plan, implement, and integrate AI for sustainable use in mental health. These studies can support proposing strategies for integrating AI decision-making utility in mental health practices, with the consideration to, including all relevant stakeholders in the research process.

Strengths and Limitations

Some methodological considerations are worth noting. The literature search was conducted using 5 relevant databases, which were selected based on the study's focus and the research field. However, it is important to acknowledge that using different databases may yield varying results, potentially affecting the comprehensiveness of the review. The keywords used in the search strategy were thoroughly discussed within the research group and with librarian assistance to align with the study's aims. Nonetheless, the evolving nature of interdisciplinary field, such as health care, implementation science, and AI technology means that the literature may use a broad range of terminology. This variability could impact the identification and inclusion of relevant studies, as different keywords might yield different sets of literature.

Furthermore, the decision to include only AI systems with technology readiness level ≥6 intended to focus the exploration on AI systems that are close to the implementation phase and investigating the AI use in care settings. While this criterion ensured relevance to real-world applications, it also narrowed the scope of the review, excluding studies that focus on the earlier developmental stages of AI models. This limitation means that the review does not address the potential capabilities of emerging AI systems that are still in the research or proof of concept phase but might have profound implications for mental health applications in the future. In addition, the limited number of studies available in implementation phase restricts the ability to draw generalized conclusions about the long-term effectiveness or scalability of these AI systems in mental health.

Moreover, 2 researchers independently conducted the selection and review, enhancing the reliability and reducing potential bias in study inclusion. However, assessing SDM in relation to these AI systems posed a challenge, as most studies did not report on the interaction dynamics between health care professionals, patients, and the AI system in a mutual decision-making context. This lack of detailed reporting on SDM limits the understanding of how AI influences or facilitates collaborative care in mental health.

Implications and Future Research

This study sheds light on the knowledge gaps regarding the empirical evidence of the use and integration of AI systems in mental health decision-making processes, offering insights that can guide the development and design of human-centered AI systems. It underscores the need to consider the utility of AI systems and end-user perspectives in the planning for the development and implementation of AI systems to support decision-making in mental health, in addition to the need for more empirical evidence to validate the factors affecting behavioral sustainable use. Furthermore, this study presents the current state of complexity in which the HCI perspective is fundamental in combination with co-design methodologies to ensure person-centered care when implementing AI systems in care processes.

Future research is advised to explore the integration of support for SDM in the AI system's design and development, addressing the gap identified in the current literature. In addition, there is a need for conceptual refinement of the current SDM process due to the disruptive effect of AI use. As the care process is evolving from the traditional physician-patient SDM, potentially adding the AI component as a third decision maker in the SDM dynamics. This adjustment is leading to changes in the essential elements of the process, and the 3-talk model, in particular when it comes to the technology-human SDM dynamics [28,100]. Conducting implementation studies are needed for creating robust empirical evidence, and to address potential long-term challenges by identifying barriers and facilitators of AI implementation and sustainable use in mental health care.

Conclusions

The scoping review demonstrated the diverse range of uses of AI in the mental health field, including health care settings, self-care contexts, and hybrid approaches. However, there was a lack of empirical evidence from an implementation perspective on how the AI systems should be integrated into clinical

XSL•FO
RenderX

processes. None of the studies discussed how to adjust clinical processes to accommodate aspects such as patient-physician communication and SDM.

The results showed that the utility of AI systems in supporting decision-making in mental health varies and illustrates the complexity of evaluating the utility of decision-making, influenced by the type of actors involved in decisions, the level of decision actionability, and the dynamics of data flow through the AI system. The evaluation of the use of AI systems in care settings through human, organization, and technology lenses revealed several challenges for the implementation of AI systems for sustainable use in care settings. As a consequence, these use- and implementation-related challenges may impact the accuracy of the decisions supported by the AI system, disrupt physician-patient communication, pose trust issues, and present engagement problems impacting the effectiveness and adoption of AI decision support. Future studies are needed to address stakeholder's needs in AI systems design, evaluate the implementation of AI systems in clinical processes for sustainable use, and assess the integration of SDM in AI systems.

Acknowledgments

This research study was part of Center for Applied Intelligent Systems Research (CAISR) Health profile and was funded by the Knowledge Foundation (grant 20200208 01H) and Halmstad University. The Knowledge Foundation was not involved in any aspect of study design, data collection, analysis, interpretation of data, or in the writing and publication of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 498 KB - jmir_v27i1e63548_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOC File , 43 KB - jmir_v27i1e63548_app2.doc]

Multimedia Appendix 3
Artificial intelligence systems characteristics.
[DOCX File , 41 KB - jmir_v27i1e63548_app3.docx]

Multimedia Appendix 4
All elements used in the human, organization, and technology-fit framework with reference to the articles.
[DOCX File , 22 KB - jmir_v27i1e63548_app4.docx]

References

1. Bertl M, Ross P, Draheim D. A survey on AI and decision support systems in psychiatry – uncovering a dilemma. Expert Syst Appl 2022 Sep;202:117464. [doi: 10.1016/j.eswa.2022.117464]
2. Maslej MM, Kloiber S, Ghassemi M, Yu J, Hill SL. Out with AI, in with the psychiatrist: a preference for human-derived clinical decision support in depression care. Transl Psychiatry 2023 Jun 16;13(1):210 [FREE Full text] [doi: 10.1038/s41398-023-02509-z] [Medline: 37328465]
3. Russ TC, Woelbert E, Davis KA, Hafferty JD, Ibrahim Z, Inkster B, et al. How data science can advance mental health research. Nat Hum Behav 2019 Jan 10;3(1):24-32. [doi: 10.1038/s41562-018-0470-9] [Medline: 30932051]
4. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
5. Grzenda A. Artificial intelligence in mental health. In: Eyre HA, Berk M, Lavretsky H, Reynolds C, editors. Convergence Mental Health: A Transdisciplinary Approach to Innovation. Oxford, UK: Oxford Academic Press; 2021:137.
6. Warrier U, Warrier A, Khandelwal K. Ethical considerations in the use of artificial intelligence in mental health. Egypt J Neurol Psychiatr Neurosurg 2023 Oct 27;59(1):36. [doi: 10.1186/S41983-023-00735-2]
7. Ji M, Xie W, Huang R, Qian X. Forecasting the suitability of online mental health information for effective self-care developing machine learning classifiers using natural language features. Int J Environ Res Public Health 2021 Sep 24;18(19):10048 [FREE Full text] [doi: 10.3390/ijerph181910048] [Medline: 34639348]
8. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 2019 Nov 07;21(11):116 [FREE Full text] [doi: 10.1007/s11920-019-1094-0] [Medline: 31701320]

XSL•FO
RenderX

9. Richmond RS, Connolly M. A delineation of self-management and associated concepts. Int J Healthc Manag 2020 Aug 25;14(4):1576-1588. [doi: 10.1080/20479700.2020.1810963]
10. Lee EE, Torous J, de Choudhury M, Depp CA, Graham SA, Kim HC, et al. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging 2021 Sep;6(9):856-864 [FREE Full text] [doi: 10.1016/j.bpsc.2021.02.001] [Medline: 33571718]
11. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak 2023 Apr 20;23(1):73 [FREE Full text] [doi: 10.1186/s12911-023-02162-y] [Medline: 37081503]
12. Viswanathan S, Rollwage M, Harper R. Promises and challenges of AI-enabled mental healthcare: a foundational study. In: Proceedings of the 36th Conference on Neural Information Processing Systems. 2022 Presented at: NeurIPS '22; November 28-December 9, 2022; Virtual Event p. 1-6 URL: https://openreview.net/pdf?id=TfmEdrBTVs5
13. Triberti S, Durosini I, Pravettoni G. A "third wheel" effect in health decision making involving artificial entities: a psychological perspective. Front Public Health 2020 Apr 28;8:117 [FREE Full text] [doi: 10.3389/fpubh.2020.00117] [Medline: 32411641]
14. Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, Ghadiri P, Grad R, Poenaru D, et al. Application of artificial intelligence in shared decision making: scoping review. JMIR Med Inform 2022 Aug 09;10(8):e36199 [FREE Full text] [doi: 10.2196/36199] [Medline: 35943793]
15. Yan WJ, Ruan QN, Jiang K. Challenges for artificial intelligence in recognizing mental disorders. Diagnostics (Basel) 2022 Dec 20;13(1):2 [FREE Full text] [doi: 10.3390/diagnostics13010002] [Medline: 36611294]
16. Golden G, Popescu C, Israel S, Perlman K, Armstrong C, Fratila R, et al. Applying artificial intelligence to clinical decision support in mental health: what have we learned? Health Policy Technol 2024 Jun;13(2):100844. [doi: 10.1016/j.hlpt.2024.100844]
17. Espejo G, Reiner W, Wenzinger M. Exploring the role of artificial intelligence in mental healthcare: progress, pitfalls, and promises. Cureus 2023 Sep;15(9):e44748 [FREE Full text] [doi: 10.7759/cureus.44748] [Medline: 37809254]
18. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020 Oct;3(3):326-331 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa033] [Medline: 33215066]
19. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021 Apr 22;23(4):e25759 [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Granlund M, Imms C, King G, Andersson AK, Augustine L, Brooks R, et al. Definitions and operationalization of mental health problems, wellbeing and participation constructs in children with NDD: distinctions and clarifications. Int J Environ Res Public Health 2021 Feb 09;18(4):1656 [FREE Full text] [doi: 10.3390/ijerph18041656] [Medline: 33572339]
23. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
24. Johansen ML, O'Brien JL. Decision making in nursing practice: a concept analysis. Nurs Forum 2016;51(1):40-48. [doi: 10.1111/nuf.12119] [Medline: 25639525]
25. Irani E, Hirschman KB, Cacchione PZ, Bowles KH. Home health nurse decision-making regarding visit intensity planning for newly admitted patients: a qualitative descriptive study. Home Health Care Serv Q 2018;37(3):211-231 [FREE Full text] [doi: 10.1080/01621424.2018.1456997] [Medline: 29578844]
26. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
27. Gupta JN, Forgionne GA, Mora MT. Intelligent Decision-Making: Support Systems Foundations, Applications and Challenges. Cham, Switzerland: Springer; 2007.
28. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
30. Sremac S, Kazimieras Zavadskas E, Matić B, Kopić M, Stević Ž. Neuro-fuzzy inference systems approach to decision support system for economic order quantity. Econ Res Ekon Istraz 2019 Jun 08;32(1):1114-1137. [doi: 10.1080/1331677X.2019.1613249]
31. Sadeh-Sharvit S, Rego SA, Jefroykin S, Peretz G, Kupershmidt T. A comparison between clinical guidelines and real-world treatment data in examining the use of session summaries: retrospective study. JMIR Form Res 2022 Aug 16;6(8):e39846 [FREE Full text] [doi: 10.2196/39846] [Medline: 35972782]

XSL•FO
RenderX

32. Thompson J. A guide to abductive thematic analysis. Qual Rep 2022 May 20;27(5):1410-1421 [FREE Full text] [doi: 10.46743/2160-3715/2022.5340]
33. Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for health information systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008 Jun;77(6):386-398. [doi: 10.1016/j.ijmedinf.2007.08.011] [Medline: 17964851]
34. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ. XAI-explainable artificial intelligence. Sci Robot 2019 Dec 18;4(37):eaay7120. [doi: 10.1126/scirobotics.aay7120] [Medline: 33137719]
35. Thiebes S, Lins S, Sunyaev A. Trustworthy artificial intelligence. Electron Mark 2020 Oct 01;31(2):447-464. [doi: 10.1007/S12525-020-00441-4]
36. Dosovitsky G, Pineda BS, Jacobson NC, Chang C, Escoredo M, Bunge EL. Artificial intelligence chatbot for depression: descriptive study of usage. JMIR Form Res 2020 Nov 13;4(11):e17065. [doi: 10.2196/17065] [Medline: 33185563]
37. Benrimoh D, Tanguay-Sela M, Perlman K, Israel S, Mehltretter J, Armstrong C, et al. Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician-patient interaction. BJPsych Open 2021 Jan 06;7(1):e22 [FREE Full text] [doi: 10.1192/bjo.2020.127] [Medline: 33403948]
38. Dimeff LA, Jobes DA, Koerner K, Kako N, Jerome T, Kelley-Brimer A, et al. Using a tablet-based app to deliver evidence-based practices for suicidal patients in the emergency department: pilot randomized controlled trial. JMIR Ment Health 2021 Mar 01;8(3):e23022 [FREE Full text] [doi: 10.2196/23022] [Medline: 33646129]
39. Jacobs M, Pradier MF, McCoy Jr TH, Perlis RH, Doshi-Velez F, Gajos KZ. How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection. Transl Psychiatry 2021 Feb 04;11(1):108 [FREE Full text] [doi: 10.1038/s41398-021-01224-x] [Medline: 33542191]
40. Popescu C, Golden G, Benrimoh D, Tanguay-Sela M, Slowey D, Lundrigan E, et al. Evaluating the clinical feasibility of an artificial intelligence-powered, web-based clinical decision support system for the treatment of depression in adults: longitudinal feasibility study. JMIR Form Res 2021 Oct 25;5(10):e31862 [FREE Full text] [doi: 10.2196/31862] [Medline: 34694234]
41. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. Drug Alcohol Depend 2021 Oct 01;227:108986 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108986] [Medline: 34507061]
42. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
43. Deng L, Rattadilok P. A sensor and machine learning-based sensory management recommendation system for children with autism spectrum disorders. Sensors (Basel) 2022 Aug 03;22(15):5803 [FREE Full text] [doi: 10.3390/s22155803] [Medline: 35957356]
44. Garriga R, Mas J, Abraha S, Nolan J, Harrison O, Tadros G, et al. Machine learning model to predict mental health crises from electronic health records. Nat Med 2022 Jun;28(6):1240-1248 [FREE Full text] [doi: 10.1038/s41591-022-01811-5] [Medline: 35577964]
45. Megerian JT, Dey S, Melmed RD, Coury DL, Lerner M, Nicholls CJ, et al. Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. NPJ Digit Med 2022 May 05;5(1):57 [FREE Full text] [doi: 10.1038/s41746-022-00598-6] [Medline: 35513550]
46. Tanguay-Sela M, Benrimoh D, Popescu C, Perez T, Rollins C, Snook E, et al. Evaluating the perceived utility of an artificial intelligence-powered clinical decision support system for depression treatment using a simulation center. Psychiatry Res 2022 Feb;308:114336. [doi: 10.1016/j.psychres.2021.114336] [Medline: 34953204]
47. Ettman CK, Galea S. The potential influence of AI on population mental health. JMIR Ment Health 2023 Nov 16;10:e49936 [FREE Full text] [doi: 10.2196/49936] [Medline: 37971803]
48. Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key considerations for incorporating conversational AI in psychotherapy. Front Psychiatry 2019;10:746 [FREE Full text] [doi: 10.3389/fpsyt.2019.00746] [Medline: 31681047]
49. Shimada K. The role of artificial intelligence in mental health: a review. Sci Insights 2023 Nov 29;43(5):1119-1127. [doi: 10.15354/si.23.re820]
50. Rundo L, Pirrone R, Vitabile S, Sala E, Gambino O. Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine. J Biomed Inform 2020 Aug;108:103479 [FREE Full text] [doi: 10.1016/j.jbi.2020.103479] [Medline: 32561444]
51. Khan B, Fatima H, Qureshi A, Kumar S, Hanan A, Hussain J, et al. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. Biomed Mater Devices 2023 Feb 08:1-8 [FREE Full text] [doi: 10.1007/s44174-023-00063-2] [Medline: 36785697]
52. Joyce DW, Kormilitzin A, Smith KA, Cipriani A. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. NPJ Digit Med 2023 Jan 18;6(1):6 [FREE Full text] [doi: 10.1038/s41746-023-00751-9] [Medline: 36653524]

XSL•FO
RenderX

53. Higgins O, Short BL, Chalup SK, Wilson RL. Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: an integrative review. Int J Ment Health Nurs 2023 Aug;32(4):966-978. [doi: 10.1111/inm.13114] [Medline: 36744684]
54. Lee CS, Lee AY. Clinical applications of continual learning machine learning. Lancet Digit Health 2020 Jun;2(6):e279-e281 [FREE Full text] [doi: 10.1016/S2589-7500(20)30102-3] [Medline: 33328120]
55. Chandler C, Foltz PW, Elvevåg B. Improving the applicability of AI for psychiatric applications through human-in-the-loop methodologies. Schizophr Bull 2022 Sep 01;48(5):949-957 [FREE Full text] [doi: 10.1093/schbul/sbac038] [Medline: 35639561]
56. Heer J. Agency plus automation: designing artificial intelligence into interactive systems. Proc Natl Acad Sci U S A 2019 Feb 05;116(6):1844-1850 [FREE Full text] [doi: 10.1073/pnas.1807184115] [Medline: 30718389]
57. Ooge J, Stiglic G, Verbert K. Explaining artificial intelligence with visual analytics in healthcare. Wiley Interdiscip Rev Data Min Knowl Discov 2021 Nov 28;12(1):e1427. [doi: 10.1002/widm.1427]
58. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
59. Jacobs M, He JF, Pradier MF, Lam B, Ahn AC, McCoy TH, et al. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-14 URL: https://dl.acm.org/doi/10.1145/3411764.3445385 [doi: 10.1145/3411764.3445385]
60. Patrickson B, Musker M, Thorpe D, van Kasteren Y, Bidargaddi N. In-depth co-design of mental health monitoring technologies by people with lived experience. Future Internet 2023 May 25;15(6):191. [doi: 10.3390/fi15060191]
61. Hogg HD, Al-Zubaidy M, Technology Enhanced Macular Services Study Reference Group, Talks J, Denniston AK, Kelly CJ, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. J Med Internet Res 2023 Jan 10;25:e39742 [FREE Full text] [doi: 10.2196/39742] [Medline: 36626192]
62. Zicari RV, Ahmed S, Amann J, Braun SA, Brodersen J, Bruneault F, et al. Co-design of a trustworthy AI system in healthcare: deep learning based skin lesion classifier. Front Hum Dyn 2021 Jul 13;3:96. [doi: 10.3389/fhumd.2021.688152]
63. Gotz D, Borland D. Data-driven healthcare: challenges and opportunities for interactive visualization. IEEE Comput Grap Appl 2016 May;36(3):90-96. [doi: 10.1109/mcg.2016.59]
64. Drake RE, Cimpean D, Torrey WC. Shared decision making in mental health: prospects for personalized medicine. Dialogues Clin Neurosci 2009;11(4):455-463 [FREE Full text] [doi: 10.31887/DCNS.2009.11.4/redrake] [Medline: 20135903]
65. Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics 2020 Apr 03;47(12):e3 [FREE Full text] [doi: 10.1136/medethics-2019-105860] [Medline: 32245804]
66. Ferrara E. The butterfly effect in artificial intelligence systems: implications for AI bias and fairness. Mach Learn Appl 2024 Mar;15:100525. [doi: 10.1016/j.mlwa.2024.100525]
67. Schwartz R, Schwartz R, Vassilev A, Greene K, Perine L, Burt A, et al. Towards a standard for identifying and managing bias in artificial intelligence. National Institute of Standards and Technology. URL: https://www.nist.gov/publications/towards-standard-identifying-and-managing-bias-artificial-intelligence [accessed 2024-04-29]
68. Magrabi F, Ammenwerth E, McNair JB, de Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform 2019 Aug;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
69. Byrne MF, Parsa N, Greenhill AT, Chahal D, Ahmad O, Bagci U. Making AI work: designing and evaluating AI systems in healthcare. In: Byrne MF, Parsa N, Greenhill AT, Chahal D, Ahmad O, Bagci U, editors. AI in Clinical Medicine: A Practical Guide for Healthcare Professionals. Hoboken, NJ: John Wiley & Sons; 2023.
70. Nair M, Svedberg P, Larsson I, Nygren JM. A comprehensive overview of barriers and strategies for AI implementation in healthcare: mixed-method design. PLoS One 2024;19(8):e0305949 [FREE Full text] [doi: 10.1371/journal.pone.0305949] [Medline: 39121051]
71. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. Med Phys 2020 Jun 17;47(5):e228-e235 [FREE Full text] [doi: 10.1002/mp.13562] [Medline: 32418341]
72. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127 [FREE Full text] [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
73. Ferrario A, Gloeckler S, Biller-Andorno N. AI knows best? Avoiding the traps of paternalism and other pitfalls of AI-based patient preference prediction. J Med Ethics 2023 Mar;49(3):185-186. [doi: 10.1136/jme-2023-108945] [Medline: 36788037]
74. Lorenzini G, Arbelaez Ossa L, Shaw DM, Elger BS. Artificial intelligence and the doctor-patient relationship expanding the paradigm of shared decision making. Bioethics 2023 Jun;37(5):424-429. [doi: 10.1111/bioe.13158] [Medline: 36964989]
75. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
76. Göndöcs D, Dörfler V. AI in medical diagnosis: AI prediction and human judgment. Artif Intell Med 2024 Mar;149:102769 [FREE Full text] [doi: 10.1016/j.artmed.2024.102769] [Medline: 38462271]

XSL•FO
RenderX

77. Rebelo AD, Verboom DE, dos Santos NR, de Graaf JW. The impact of artificial intelligence on the tasks of mental healthcare workers: a scoping review. Comput Human Behav 2023 Aug;1(2):100008. [doi: 10.1016/j.chbah.2023.100008]
78. Alhuwaydi A. Exploring the role of artificial intelligence in mental healthcare: current trends and future directions - a narrative review for a comprehensive insight. Risk Manag Healthc Policy 2024;17:1339-1348 [FREE Full text] [doi: 10.2147/RMHP.S461562] [Medline: 38799612]
79. Amann J, Vetter D, Blomberg SN, Christensen HC, Coffee M, Gerke S, et al. To explain or not to explain?-artificial intelligence explainability in clinical decision support systems. PLOS Digit Health 2022 Feb 17;1(2):e0000016 [FREE Full text] [doi: 10.1371/journal.pdig.0000016] [Medline: 36812545]
80. Reddy YJ, Sudhir PM, Manjula M, Arumugham SS, Narayanaswamy JC. Clinical practice guidelines for cognitive-behavioral therapies in anxiety disorders and obsessive-compulsive and related disorders. Indian J Psychiatry 2020;62(8):230. [doi: 10.4103/psychiatry.indianjpsychiatry_773_19]
81. Lincoln T, Jung E, Wiesjahn M, Schlier B. What is the minimal dose of cognitive behavior therapy for psychosis? An approximation using repeated assessments over 45 sessions. Eur Psychiatry 2016 Oct 23;38:31-39. [doi: 10.1016/j.eurpsy.2016.05.004] [Medline: 27642702]
82. Rollwage M, Habicht J, Juechems K, Carrington B, Viswanathan S, Stylianou M, et al. Using conversational AI to facilitate mental health assessments and improve clinical efficiency within psychotherapy services: real-world observational study. JMIR AI 2023 Dec 13;2:e44358 [FREE Full text] [doi: 10.2196/44358] [Medline: 38875569]
83. Rollwage M, Habicht J, Juchems K, Carrington B, Hauser TU, Harper R. Conversational AI facilitates mental health assessments and is associated with improved recovery rates. BMJ Innov 2024 Mar 01;10(1-2):4-12. [doi: 10.1136/bmjinnov-2023-001110]
84. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 2023 Dec 19;6(1):236 [FREE Full text] [doi: 10.1038/s41746-023-00979-5] [Medline: 38114588]
85. Alfano L, Malcotti I, Ciliberti R. Psychotherapy, artificial intelligence and adolescents: ethical aspects. J Prev Med Hyg 2023 Dec;64(4):E438-E442 [FREE Full text] [doi: 10.15167/2421-4248/jpmh2023.64.4.3135] [Medline: 38379752]
86. Saqib H, Sumyia AS, Saqib A. AI chatbots and psychotherapy: a match made in heaven? J Pak Med Assoc 2023 Oct 30;73(11):2321. [doi: 10.47391/jpma.9608]
87. Balcombe L. AI chatbots in digital mental health. Informatics 2023 Oct 27;10(4):82. [doi: 10.3390/informatics10040082]
88. Denecke K, Abd-Alrazaq A, Househ M. Artificial intelligence for chatbots in mental health: opportunities and challenges. In: Househ M, Borycki E, Kushniruk A, editors. Multiple Perspectives on Artificial Intelligence in Healthcare: Opportunities and Challenges. Cham, Switzerland: Springer; 2021:115-128.
89. Jabir AI, Martinengo L, Lin X, Torous J, Subramaniam M, Tudor Car L. Evaluating Conversational Agents for Mental Health: Scoping Review of Outcomes and Outcome Measurement Instruments. J Med Internet Res 2023 Apr 19;25:e44548 [FREE Full text] [doi: 10.2196/44548] [Medline: 37074762]
90. Dean E, Söderlund A. What is the role of lifestyle behaviour change associated with non-communicable disease risk in managing musculoskeletal health conditions with special reference to chronic pain? BMC Musculoskelet Disord 2015 Apr 13;16:87 [FREE Full text] [doi: 10.1186/s12891-015-0545-y] [Medline: 25888381]
91. Benzo RP. Self-management programs and the pursuit of behavior change. Respir Care 2024 May 28;69(6):678-685. [doi: 10.4187/respcare.11987] [Medline: 38806226]
92. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
93. Kok G, Gottlieb NH, Peters GJ, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an intervention mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
94. Agbadjé TT, Elidor H, Perin MS, Adekpedjou R, Légaré F. Towards a taxonomy of behavior change techniques for promoting shared decision making. Implement Sci 2020 Aug 20;15(1):67 [FREE Full text] [doi: 10.1186/s13012-020-01015-w] [Medline: 32819410]
95. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
96. Marlatt GA. Taxonomy of high-risk situations for alcohol relapse: evolution and development of a cognitive-behavioral model. Addiction 1996 Dec;91(s12):S37-S49. [doi: 10.1111/j.1360-0443.1996.tb02326.x]
97. Guven E, Gokce G. How effective are efficiency studies?: an evaluation in the context of cognitive behavioral therapies. Turk Psikiyatri Derg 2018;29(1):21. [doi: 10.5080/U22706]
98. Scheinker D, Brandeau ML. Implementing analytics projects in a hospital: successes, failures, and opportunities. INFORMS J Appl Anal 2020 May;50(3):176-189. [doi: 10.1287/inte.2020.1036]

XSL•FO
RenderX

99. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020 Jul 15;8(7):e15182 [FREE Full text] [doi: 10.2196/15182] [Medline: 32673244]
100. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns 2006 Mar;60(3):301-312. [doi: 10.1016/j.pec.2005.06.010] [Medline: 16051459]

Abbreviations

AI: artificial intelligence
HCI: human-computer interaction
HOT-fit: human, organization, and technology-fit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RQ: research question
SDM: shared decision-making

Edited by T de Azevedo Cardoso; submitted 26.06.24; peer-reviewed by S Mahed Mousavi, RA El Arab, D Patel; comments to author 19.09.24; revised version received 28.10.24; accepted 25.11.24; published 24.01.25.

Please cite as:
Auf H, Svedberg P, Nygren J, Nair M, Lundgren LE
The Use of AI in Mental Health Services to Support Decision-Making: Scoping Review
J Med Internet Res 2025;27:e63548
URL: https://www.jmir.org/2025/1/e63548
doi:10.2196/63548
PMID:

©Hassan Auf, Petra Svedberg, Jens Nygren, Monika Nair, Lina E Lundgren. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Exploring the Use of Smartwatches and Activity Trackers for Health-Related Purposes for Children Aged 5 to 11 years: Systematic Review

Lauren Thompson[1], BSc, MSc; Sydney Charitos[1], MSc; Jon Bird[1], BSc, DPhil; Paul Marshall[1], BSc, DPhil; Amberly Brigden[1], BSc, MSc, PhD

University of Bristol, Bristol, United Kingdom

Corresponding Author:
Lauren Thompson, BSc, MSc
University of Bristol
1 Cathedral Square, Trinity Street
Bristol, BS1 5DD
United Kingdom
Phone: 44 01173746653
Email: lauren.thompson@bristol.ac.uk

Abstract

Background: Digital health interventions targeting behavior change are promising in adults and adolescents; however, less attention has been given to younger children. The proliferation of wearables, such as smartwatches and activity trackers, that support the collection of and reflection on personal health data highlights an opportunity to consider novel approaches to supporting health in young children (aged 5-11 y).

Objective: This review aims to investigate how smartwatches and activity trackers have been used across child health interventions (for children aged 5-11 y) for different health areas, specifically to identify the population characteristics of those being targeted, describe the characteristics of the devices being used, and report the feasibility and acceptability of these devices for health-related applications with children.

Methods: We searched 10 databases (CINAHL, Embase, ACM Digital Library, IEEE Xplore, Cochrane Library, PsycINFO, Web of Science, PubMed, Scopus, and MEDLINE) to identify relevant literature in March 2023. The inclusion criteria for studies were as follows: (1) peer-reviewed, empirical studies; (2) published in English; (3) involved a child aged 5 to 11 years using a smartwatch for health-related purposes. Two researchers independently screened articles to assess eligibility. One researcher extracted data relating to the 3 aims and synthesized the results using narrative and thematic synthesis.

Results: The database searches identified 3312 articles, of which 15 (0.45%) were included in this review. Three (20%) articles referred to the same intervention. In 77% (10/13) of the studies, the devices were used to target improvements in physical activity. Other applications included using smartwatches to deliver interventions for emotional regulation and asthma management. In total, 9 commercial devices were identified, many of which delivered minimal data feedback on the smartwatch or activity tracker, instead relying on a partner app running on a linked parental smartphone with greater functionality. Of the 13 studies, 8 (62%) used devices designed for adults rather than children. User feedback was positive overall, demonstrating the acceptability and feasibility of using these devices with children. However, the studies often lacked a child-focused approach, with 3 (23%) studies gathering user feedback only from parents.

Conclusions: Interventions involving smartwatches and activity trackers for children aged 5 to 11 years remain limited, primarily focusing on enhancing physical activity, with few studies investigating other health applications. These devices often provide limited data feedback and functionality to support children's independent engagement with the data, relying on paired smartphone apps managed by caregivers, who control access and facilitate children's interaction with the data. Future research should adopt child-centered methods in the design and evaluation of these technologies, integrating children's perspectives alongside their caregivers, to ensure that they are not only feasible and acceptable but also meaningful and effective for young children.

Trial Registration: PROSPERO CRD42022373813, https://tinyurl.com/4kxu8zss

(J Med Internet Res 2025;27:e62944) doi:10.2196/62944

XSL•FO
RenderX

KEYWORDS

children; systematic review; wearable activity trackers; smartwatches; feasibility; mobile phone

Introduction

Background

Health during childhood is a key predictor of health outcomes in adolescence and adulthood [1-5]. Poor physical health during childhood that continues into adulthood can result in higher health care costs and poorer quality of life [6]. As the prevalence of childhood ill health increases [7], it is important to consider novel approaches to help improve health outcomes in children. Behavioral interventions can be used to promote positive lifestyle behaviors (eg, engaging in physical activity [PA]) as well as support the management of chronic conditions (eg, promoting treatment adherence) [8].

For adults and adolescents, there is evidence that digital health interventions are effective in promoting health behavior change while also being cost-effective, appealing, and accessible [9-11]. However, in a systematic review of digital health interventions for children, of the 17 digital interventions identified, only 5 were found to be promising in producing a positive impact on clinical outcomes [12]. The lack of effective interventions indicates a need for progress in the field of digital health interventions for children, investigating technological innovations that meet the specific needs and contexts of children.

The development of wearables and sensors presents innovative solutions for delivering behavioral interventions. There are now a large number of commercial products that enable the real-time tracking and monitoring of personal health information and behaviors such as step count, heart rate, and sleep [13]. Devices such as smartwatches are well-suited for use as health intervention tools, as they are widely available and allow users to reflect on the health information collected via sensors in real time to gain a better understanding of their health and engage in self-management behaviors [14]. In this review, we define a smartwatch as a wrist-worn device that uses sensors to collect at least 2 types of personal health data (eg, steps and heart rate) and has an interactive display. With the popularity of adult trackers, there is also an established market for wearables and smartwatches for children [15]. These devices may be more appropriate for younger users, as they can be used independently from mobile devices, as device ownership is uncommon in young children [16].

There is a growing body of evidence for activity tracker and smartwatch health interventions for adults and adolescents, but despite the promise, there has been inadequate attention to their use with children [17,18]. Designing and deploying wearable health interventions for children needs specific attention due to the distinct developmental characteristics of this age group (social, emotional, and cognitive) and their greater reliance on parents and carers for support [8,19,20]. This includes considering how wearable hardware and software should be designed to meet the distinct needs of children. It also includes considering wider systems, such as family systems, for example, considering how multiple devices should be integrated into one intervention to meet the needs of different users (eg, adult and child) and how issues such as wearable data sharing and data privacy should be navigated. To understand these technical and sociotechnical issues, understanding the end users' experience, such as the technology's feasibility and acceptability, is essential [21].

Few reviews have focused on the design and user experience of wearables for children. Systematic reviews of the acceptability and feasibility of wearable devices to promote PA in children have identified that wearables can increase children's motivation for PA using behavior change techniques but noted novelty effects and uncertainties over long-term use and impact [17,18]. However, these reviews had a narrow focus on wearables for PA and a broad age range; they were not designed to explore developmentally sensitive design implications for children across health contexts. Previous work has identified common design adaptations for children's devices such as being more colorful, having fewer health metrics (eg, calories burned not included), and including features, such as games, to boost children's engagement [15,22]. However, these did not use a systematic review methodology and are limited to presenting a description of design adaptations without analysis of the end users' experiences or engagement with these design adaptations.

Aims

This systematic review aimed to investigate how smartwatches and activity trackers have been used across the full range of child health interventions (for children aged 5-11 y), spanning public health and prevention interventions to interventions for clinical populations, focusing on their acceptability and feasibility for children. This review aimed to build on the established literature on digital interventions for children by investigating design considerations to support young children's engagement with the smartwatch and the data collected for their health. Therefore, we focused on devices with an interactive display that supports data feedback. In this review, we investigated the following research questions (RQs): (1) What are the population characteristics of those being targeted in the health-related smartwatch and activity tracker interventions for children aged 5 to 11 years? (2) What are the characteristics of the smartwatches and activity trackers being used for health-related applications with children aged 5 to 11 years (eg, device type and features included)? (3) What is the feasibility and acceptability of using smartwatches and activity trackers for health-related applications with children aged 5 to 11 years?

Methods

This review was registered in the PROSPERO database and follows the guidance of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23]. See Multimedia Appendix 1 for the full checklist.

Search Strategy

The literature search was conducted in March 2023 in the following databases: CINAHL, ACM Digital Library, PsycINFO, Embase, MEDLINE, Cochrane Library, IEEE

XSL•FO
RenderX

Xplore, PubMed, Scopus, and Web of Science. The search included key terms for (1) children aged 5 to 11 years, such as "school child" or "minor" and (2) smartwatches, such as "smartwatch" or "wearable electronic device" (Multimedia Appendix 2 provides an example of the full list of search terms used in Embase).

Inclusion and Exclusion Criteria

Studies were included if they fulfilled the following criteria: (1) peer-reviewed or empirical studies with any study design; (2) published in English; and (3) the research involved a child (aged 5-11 y) using a smartwatch for health-related purpose, including behavior change intervention (physical, mental, or social). We use the term intervention to apply to any activity undertaken to improve health [24]. For the purpose of this review, a smartwatch is defined as a wrist-worn device that uses sensors to collect at least 2 types of personal health data (eg, steps and heart rate) and has an interactive display. We also extended this to include activity trackers with a visual display (excluding those that do not reflect on the data via the device), as they are largely similar to smartwatches but are typically cheaper with fewer sensors and can be used without a smartphone [25]. We selected devices that collect at least 2 types of personal health information and incorporate a visual, interactive display, thereby excluding simpler devices such as pedometers, which lack these functionalities and therefore do not meet our criteria of being a smartwatch or activity tracker. The devices could be used in conjunction with a partnering app if the child engages with the wearable device independently as part of its use.

Studies were excluded if they fulfilled the following criteria: (1) the study did not use a smartwatch or activity tracker with an interactive display; (2) the children in the study only wore the smartwatch or activity tracker as a passive data collection tool, and they did not otherwise interact with the device, by which we mean actively collect or view the data it collected as part of an intervention or feasibility assessment (eg, accuracy or validation studies); (3) the study included children aged <5 years or >11 years or do not report on the eligible age range specifically; or (4) the article was considered a review on previously reported studies, protocol, nonempirical or nonscientific paper (eg, book chapters), or when the full text was not available.

Screening Procedures

The search results were imported into Zotero (Corporation for Digital Scholarship) and duplicates were removed. Each title and abstract were independently screened by 2 authors (LT and SC) for relevance. If an inclusion decision could not be made based on the title and abstract, the articles were included for full-text review. Articles deemed relevant at the title and abstract screening stage were independently double-screened against the inclusion and exclusion criteria using the data management platform Rayaan (Rayyan Systems Inc), with reasons for exclusion recorded. Disagreements at both stages were discussed and resolved in meetings by the reviewers (LT and SC), with persistent disagreements being resolved through discussion with the wider research team (AB and JB). If the full text did not contain the information needed, 2 attempts were made to contact authors by email; if the information was not provided, the study was excluded.

Data Extraction and Synthesis

Data related to our RQs were extracted by the primary author (LT), as described in Textbox 1. Due to the heterogeneity in the study designs and the data reported, we synthesized data using narrative and thematic synthesis [26,27]. The extracted data are summarized in Textbox 1. For qualitative data, open coding was conducted by the primary author on the available qualitative data referring to participants' experiences of using the device, which included quotations from participants or authors' interpretations of participants' experiences. Themes were developed top-down using the technology acceptance model (TAM) [28], in which codes generated were grouped under each relevant theme of the TAM (perceived ease of use and perceived usefulness). The remaining analysis was data driven (bottom-up) to develop themes for any codes that fell outside the scope of the TAM. Wherever possible, quotes from the participants of the included studies (primary data) have been presented. In addition, to understand the characteristics of the device relevant to behavior change, we used the taxonomy of app features [29] and coded whether the features were present on the smartwatch device or partner app in each research study as shown in Multimedia Appendix 3.

XSL•FO
RenderX

Textbox 1. Data extraction categories and details extracted.

General study information

- Publication details: authors, year of publication, country of study, and publication location
- Study aim
- Sample size

Research question 1: What are the population characteristics of those being targeted in the health-related smartwatch or activity tracker interventions?

- Health behavior or condition targeted
- Stakeholders involved (eg, target child, parent, or other)
- Child sample characteristics: age and gender
- Child's device experience

Research question 2: What are the characteristics of the smartwatch or activity trackers?

- Intervention summary: a description of the wearable element of the intervention
- Device details: name, type (smartwatch or activity tracker), technology readiness level [30] (consumer device or experimental prototype), and if there is a partner app or device
- Device rationale and justification: information that the authors provide describing why the device was selected
- Smartwatch software details: description of features, functionality, and interaction modality and whether the software used was modified or bespoke, including any justification or rationale for design choices
- Software on partner app: using the taxonomy of app features [29] to identify app features used to promote behavior change
- Data sharing features: who the data is shared with and if data sharing is part of the interventions

Research question 3: What is the feasibility and acceptability of using a smartwatch and activity tracker for health-related applications with children aged 5 to 11 years?

- Qualitative data relating to acceptability and feasibility and who reported these data (eg, child or parent)
- Quantitative data relating to acceptability and feasibility (eg, measures of use and who reported these data)
- Any information captured about acceptability and feasibility but not captured through formal qualitative or quantitative data collection, that is, authors' interpretations in the results and discussion

Quality Assessment

Two authors conducted quality assessments independently, discussing disagreements to reach a consensus. As the studies varied in study design, we used different quality assessment tools. We used the Critical Appraisal Skills Programme systematic review checklist [31] for qualitative studies. This is a 10-item checklist with 3 responses: "yes," "no," and "cannot tell." The Quality of Survey Studies in Psychology checklist [32] was used to assess articles that collected user feedback via surveys and questionnaires. This checklist comprises 20 items under 4 categories: introduction, participants, data, and ethics. Finally, the National Heart, Lung, and Blood Institute pre-post and controlled studies assessment tools were used. These are 12- and 14-item checklists with 5 responses: "yes," "no," "cannot determine," "not applicable," and "not reported" [33,34]. Full details of the quality assessment can be found in Multimedia Appendix 4 [35-49].

Results

In this section, we summarize the data extracted to address our RQs as outlined in Textbox 1.

Search Results

After deduplication, 3312 (69%) of the 4772 articles were identified from the database searches. After abstract and title screening, 150 (4.5%) articles were considered for a full-text review, and of these, 15 (10%) were identified as eligible for inclusion. In total, 3 articles reported on the same intervention; therefore, we identified 13 studies detailing children's (aged 5-11 y) use of smartwatches for health-related purposes. A summary of the study details can be found in Table 1. Table 2 provides a summary of the population characteristics. We will refer to these 13 distinct studies unless otherwise indicated. Most (11/13, 85%) studies focused on using a single device, but 2 (15%) studies [35,36] used multiple devices during the study period. Additional devices not meeting the study criteria (n=5) are not reported here. The most common reasons for exclusion were participants being outside the target age range, wrong publication type, and wrong device type. The PRISMA flow diagram can be found in Figure 1.

XSL•FO
RenderX

Table 1. Summary of included studies.

Study	Location	Sample size	Summary
O'Brien et al [37], 2020	NR[a]	1 child	Investigated the acceptability of a smartwatch intervention that sends visual directives to children with autism spectrum disorder
Buchele Harris and Chen [38], 2018	NR	116 children	Investigated the impact of a technology-enhanced physical activity classroom intervention
Saksono et al [39], 2018	United States	9 children and 11 parents	Explored the experiences of low SES[b] families using physical activity tracking technologies
Schoeppe et al [40], 2020	Australia	40 families; 58 children	Examined the short-term effects of an activity tracker and app intervention to increase physical activity in families (Step it Up Family program)
Schoeppe et al [41], 2023	Australia	19 families	Explored families' experience and satisfaction with a physical activity intervention using wearable activity trackers and apps (Step it Up Family program)
Schoeppe et al [42], 2022	Australia	40 families; 58 children	Explored the feasibility of a family-based activity tracker intervention to increase physical activity (Step it Up Family program)
Torrado et al [43], 2017	Spain	2 children	Assessed the usability of a personalized emotional self-regulation tool
Creaser et al [44], 2022	United Kingdom	36 adults and 29 target children	A mixed methods observational study that explored the acceptability of using wearables in a family setting
Masteller et al [36], 2017	NR	16 children	Explored children's perceptions of activity trackers and their associated websites
Wing et al [45], 2022	United States	137 children	Explored the usability and acceptability of Fitbit Charge with families via pre- and poststudy questionnaires
Hosseini et al [46], 2017	United States	1 adult and 1 child	Pilot study to assess the feasibility of the mobile health asthma system in a real-world setting
Jackson et al [47], 2022[c]	United States	64 children; 35 children; 39 children; 56 children; 38 children; 25 children	Evaluated the impact of Fitbit Charge HR on children's activity levels during school
Schaefer et al [35], 2014	United States	25 children	A qualitative study to assess the acceptability and compliance of wearable activity trackers
Duck et al [48], 2021	United States	35 children	A quasi-experiment to assess the effectiveness of an activity tracker with altruistic motivation at increasing physical activity behavior
Saksono et al [49], 2019	United States	16 parents and 15 children	A qualitative study investigating how parents and children interact with their activity trackers in a naturalistic low SES context

[a]NR: not reported.

[b]SES: socioeconomic status.

[c]This study reported several conditions with unique participant demographics and study procedures.

XSL•FO
RenderX

Table 2. Summary of population characteristics (research question 1).

Study	Health focus	Age range (y)	Motivation for demographics?	Stakeholders involved in the intervention	Setting	Device experience
O'Brien et al [37], 2020	Autism spectrum disorder	9	No	Teacher	School	__[a]
Buchele Harris and Chen [38], 2018	Physical activity	10-11	No	Teacher	School	—
Saksono et al [39], 2018	Physical activity	5-11	No	Parents	Family intervention	—
Schoeppe et al [40], 2020; Schoeppe et al [41], 2023; Schoeppe et al [42], 2022	Physical activity	6-10	Yes; children aged 6 to 10 years were targeted as this age range is important for forming physical activity behaviors	Parents and siblings	Family intervention	Required to have none for eligibility
Torrado et al [43], 2017	Autism spectrum disorder	10	No	Parents and teachers	School	Yes
Creaser et al [44], 2022	Physical activity	5-9	No	Parents and siblings	Family intervention	Yes; 4 children owned a device; duration of use ranged from <1 month and >2 years
Masteller et al [36], 2017	Physical activity	6-11	Yes; elementary school children were the target demographic for the wearable	Parent	—	—
Wing et al [45], 2022	Physical activity	9-10	No	Parents	—	—
Hosseini et al [46], 2017	Asthma	7	No	—	—	—
Jackson et al [47], 2022	Physical activity	9-10	No	—	School	—
Schaefer et al [35], 2014	Physical activity	7-10	No	Parents	—	—
Duck et al [48], 2021	Physical activity	9-10	No	—	School	—
Saksono et al [49], 2018	Physical activity	6-11	No	Parents	—	—

[a]Not applicable.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

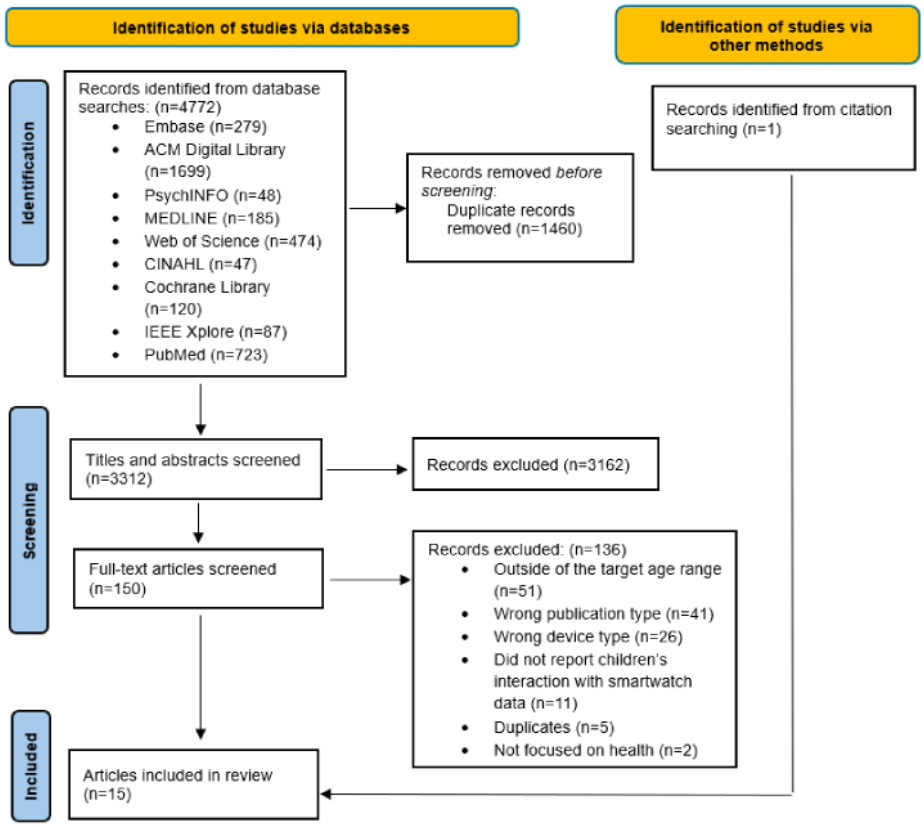

RQ1: What Are the Population Characteristics of Those Being Targeted in Health-Related Smartwatch and Activity Tracker Interventions?

Health Behaviors Targeted

Most studies (10/13, 77%) targeted PA behavior. The remaining 23% (3/13) of the studies targeted specific health conditions or recruited participants from specific demographic groups, including autism spectrum disorder (ASD; n=2, 15%) and asthma (n=1, 8%).

Demographics of Child Participants

Of the 13 studies with >1 participant, where information was provided, the mean child age ranged from 6 to 10 years, and all but one reported that at least 50% of the sample was female. Only 2 (15%) of the studies reported their motivation for the selected age range of participants. Reasons included (1) targeting elementary school-aged children as they were the target demographic of the wearable being used [36] and (2) focusing on children aged 6 to 10 years due to the developmental importance of forming PA behaviors, parental social support, and role modeling. The authors also reported that children aged <5 years were excluded due to concerns about their ability to comprehend and engage with the smartwatch and partner app [40,42].

Stakeholders Involved in the Intervention

In total, 3 of the interventions were family-based PA interventions [39-42,44], one of which extended beyond immediate family (parents and siblings) to include stepparents, cousins, and grandparents [44].

Of the 13 studies, 5 (38%) took place within a school setting in which teachers played a role in data collection or had a facilitating role in the intervention [37,38,43,47,48].

Device Experience

In total, 2 (15%) of the 13 studies reported on children's previous experience using smartwatches or activity trackers. In one study, 86.2% of the target group had never owned one. Only 4 children currently owned one, and the duration of use ranged from <1 month to >2 years [44]. In one study [43], both children had previous experience wearing a smartwatch so neither they nor their classroom peers found them distracting. In contrast, child participants in the Step It Up Family intervention [40-42] had to have no previous experience using an activity tracker to be eligible to take part in the study.

RQ2: What Are the Characteristics of Smartwatch and Activity Trackers Used for Health-Related Applications With Children Aged 5 to 11 Years?

Device Details

A total of 9 different devices were used. All of the studies used commercial devices, with Fitbit, Garmin, and UNICEF Kid Power Band being the most common. Three of the selected devices (UNICEF Kid Power Band, Garmin Vivofit Jr, and Movband) are specifically designed for children. Seven devices were described as having a partner app. A summary of the devices and the features reported can be found in Tables 3 and 4.

XSL•FO
RenderX

Table 3. Summary of device characteristics (research question 2)[a].

Study	Device	Bespoke software design?	Partner app; partner app owner	Motivation for device choice or design considerations reported?	Duration of use	Wearable part of a wider intervention?
O'Brien et al [37], 2020	Apple Watch	No	No	Yes; SMS text messages were used to deliver intervention via smartwatch rather than images due to the speed of transfer compared to images	16 days	No
Buchele Harris and Chen [38], 2018	Fitbit Charge HR	No	No	No	5 school days, for 4 weeks	Yes
Saksono et al [39], 2018	UNICEF Kid Power Band	No	Yes; not reported	Yes; accuracy, ease of use, comfort, on-band display and battery life, tailored for children	2 months	No
Schoeppe et al [40], 2020; Schoeppe et al [41], 2023; Schoeppe et al [42], 2022	Garmin Vivofit Jr	No	Yes; parent's phone	Yes; decision based on feasibility demonstrated previously	6 weeks	Yes
Torrado et al [43], 2017	LG Watch Urbane	Yes	Yes; parent's phone	Yes; they used recognized pictograms that children had familiarity with	4 hours a day, for 9 days (total of 36 h wear)	No
Creaser et al [44], 2022	Fitbit Alta HR	No	Yes; not reported	No	4 weeks	No
Masteller et al [36], 2017	Movband	No	Yes; website	No	4 days	No
Wing et al [45], 2022	Fitbit Charge HR	No	Yes; parent's phone or child's if they had their own	No	22 days	No
Hosseini et al [46], 2017	Samsung Gear Live	Yes	Yes; parent's phone	Yes; an animated dragon was considered to be more engaging for children	Several weeks	Yes
Jackson et al [47], 2022	Fitbit Charge HR	No	No	Yes; previous research had determined it appropriate to use with children	8 days to 8 weeks	No
Schaefer et al [35], 2014	Polar Active	No	No	No	5 days	No
Duck et al [48], 2021	UNICEF Kid Power Band	No	Yes; child's tablet	Yes; the device reward system used the theory of altruistic motivation	10 weeks	Yes
Saksono et al [49], 2019	UNICEF Kid Power Band	No	Yes; not reported	Yes; accuracy, ease of use, comfort, on-band display and battery life, tailored for children	2 months	No

[a]In this paper "bespoke software" design refers to any feature that is not present in existing commercial devices.

XSL•FO
RenderX

Table 4. Summary of the wearable devices used (research question 2).

Study	Device	Device type	Target user	Interaction modality	Measures on display	Features in the partner app
Saksono et al [39], 2018; Duck et al [48], 2021; Saksono et al [49], 2019	UNICEF Kid Power Band	Activity tracker	Children	Not reported	Not reported	Reward points
Creaser et al [44], 2022	Fitbit Alta HR	Activity tracker	Adults	Touch (tap display and vibration)	Daily steps, distance traveled, calories burned, active minutes, and heart rate	Goal setting, reminders, challenges, rewards, information sharing, and tracking of further health behaviors (eg, sleep)
Torrado et al [43], 2017	LG Watch Urbane	Smartwatch	Adults	Touch (tap display and vibration), auditory	Personalized self-regulation strategy (eg, pictogram of listening to music)	—[a]
O'Brien et al [37], 2020	Apple Watch	Smartwatch	Adults	Touch (vibration), auditory	Not reported	Not reported
Schoeppe et al [40], 2020; Schoeppe et al [41], 2023; Schoeppe et al [42], 2022	Garmin Vivofit Jr	Activity tracker	Children	Touch (vibration)	Daily steps, goal progress (60 min of activity)	Family leaderboard, challenges, goal setting, personal name, animal images on display, rewards (virtual coins), and virtual interactive game
Hosseini et al [46], 2017	Samsung Gear Live	Smartwatch	Adults	Touch (tap display)	Asthma attack risk (happy, neutral, or sad dragon)	Trend lines for each sensor and risk level, enabling users to focus on points of interest and review data over the last several hours. The app also had a study coordinator and physician interface
Masteller et al [36], 2017	Movband	Activity tracker	Children	Not reported	Daily steps (moves)	Displays collected points, steps, miles, and activity averages
Schaefer et al [35], 2014	Polar Active	Smartwatch	Adults	Not reported	Daily activity bar (amount of time in moderate-to-vigorous intensity zones) and animated figure that indicated activity intensity	—
Buchele Harris and Chen [38], 2018; Wing et al [45],2022; Jackson et al [47], 2022	Fitbit Charge HR	Activity tracker	Adults	Not reported	Daily steps, distance traveled, calories burned, and heart rate	Not reported

[a]Not applicable.

Device Rationale and Justification

Of the 13 studies, 5 (38%) reported their motivation for the wearable chosen for the study. The Garmin Vivofit Jr was selected due to the author's review of previous research, which demonstrated high feasibility for monitoring PA in children [40]. Similarly, the Fitbit Charge HR was selected as it had previously been considered appropriate for children [47]. The UNICEF Kid Power wristbands were selected in 2 (15%) studies due to their accuracy, battery life, comfort, on-band display, and age-appropriate design [39,49]. This device was also used in an additional study that investigated the impact of altruistic motivation due to the charity component of the device [48].

Intervention Summary

Of the 15 articles, 11 (73%) reported on behavior change, with 3 (23%) relating to the same intervention [37-44,47-49]. Of these, most (n=9, 69%) used the wearables to motivate an increase in PA. The devices were primarily used to provide immediate activity feedback to users, with step count being the

XSL•FO
RenderX

most common data point displayed on all devices. Moreover, 2 of the 13 (15%) studies used smartwatches to deliver personalized self-regulation strategies to children with ASD [37,43]. One (8%) study used the wearable to measure and visualize asthma attack risk via a dragon animation [46]. In 2 (15%) of the studies, the wearables were used primarily for research purposes. Wing et al [45] aimed to identify the parameters for determining valid wear time (a measure of how much the device is worn to determine typical patterns of activity) of the smartwatch, and Schaefer et al [35] investigated the feasibility of different wearables to inform future research on measuring PA in children.

The study periods ranged from 4 days to 2 months. Most (9/13, 69%) studies were stand-alone wearable interventions while 4 (31%) studies used the wearables as part of a wider intervention [38,40,46,47]. These included additional behavior change techniques such as motivational and educational messages (n=3, 23%) or required additional equipment as part of the wearable intervention such as a spirometer (n=1, 8%).

Smartwatch Software

Using the taxonomy of app features [29], the most common behavior change strategy incorporated into the device was activity feedback (eg, displaying step count), which was reported in 77% (10/13) of the studies. Other features included having preset activity goals and reminders. One study used the built-in SMS text messaging functionality of the Apple Watch to deliver the intervention.

In total, 2 (15%) of the 13 studies used a bespoke or modified software design using an existing commercial smartwatch [43,46]. In one study, a smartwatch system was used to assist with behavior challenges due to emotional dysregulation. The system displays self-regulation activities selected by the children's caregivers (eg, parents and teachers) via a partner device on the watch face [43]. Hosseini et al [46] described the development of a smartwatch system, which calculates and communicates asthma attack risk. Animated dragon graphics were designed based on previous research, which has shown that animations are a more engaging, age-appropriate design for children [46].

Software on Partner App

Overall, 73% (11/15) of the studies (3 refer to the same intervention) described the use of a partner app or web-based dashboard linked to the wearable. Of these, 55% (6/11) reported that the wearable was paired with a parent's smartphone [40-43,45,46] and 27% (3/11) did not report this information [39,44,49]. For the remaining 18% (2/11) of the studies, children had primary control via a website [36] or were given a compatible tablet [48]. In one study, children were not allowed to use the Fitbit mobile app or web-based dashboard alone (including on their own devices) in line with Fitbit's privacy policy [45].

The partner apps commonly had more features than the wearable device, such as access to greater data granularity, games, and access rewards. Some of the apps also allowed users to compare their performance against others and set activity goals.

Data Sharing Features

Of the 15 studies, 7 (47%; 3 reporting on the same intervention) explicitly reported that the primary data collected via the wearable were shared with other stakeholders [38,40-42,45-47]. These included parents or carers, clinicians, teachers, and peers. Data sharing was a component of the intervention to target behavior change. This included step count being shared in the classroom with peers and teachers [38,47] or among the wider family unit via step-count leaderboards within the partner app to try and encourage increased PA [40]. As part of the wearable system for asthma [46], the authors described an additional interface for physicians to access deidentified data collected via the wearable system, though this was not evaluated. In 3 interventions, parents had control over the wearable's partner app and thus had access to the data collected via the wearable [40-42,46]. Only 2 (13%) of the 15 articles in this review reported any user feedback relating to data sharing [41,44].

RQ3: What Is the Feasibility and Acceptability of Using Smartwatch and Activity Trackers for Health-Related Applications with Children Aged 5 to 11 Years?

Overview

In total, 3 (20%) of the 15 included studies did not report on the wearable's acceptability and feasibility or collect any user feedback on this [38,47,48]. 6 of the 15 articles (40%) reported gathering user feedback directly from children [35,36,44-46,49]. Feedback was also provided by parents (n=7, 47%), and in one study, treatment acceptability was completed by staff at the children's school [37].

It was not possible to statistically synthesize quantitative measures of acceptability and feasibility (response rate and wear time) as there was heterogeneity across the studies in how these statistics were reported. Instead, we provide a descriptive overview.

Quantitative Data on Usability, Acceptability, and Feasibility

Total Device Wear Time

Of the 13 studies, 5 (38%) [35,36,38,39,45] explicitly reported instructing participants on wear time. This included encouraging participants to wear the wearables as much as possible and specifying a minimum time for wearing or engaging with the partnering app or website. Only 1 (8%) study reported contacting participants if there were no new data for 3 days [45].

In total, 2 (13%) of the 15 studies reported quantitative data on device wear time. In one article, the authors reported the average number of valid recording days (≥1000 steps/d) per week as an indicator of wearable use across the 4-week intervention. This showed that wearable use was generally high, ranging from 71% (wk 3) to 91% (wk 2) [44]. Similarly, in the other article, the authors reported that during the 42-day intervention, the mean number of recording days from the children's wearables was 36.5 (SD 8.3) days [40].

In total, 2 (13%) of the 15 studies used self-reported data to indicate device use. First, parental logging was used to capture

XSL•FO
RenderX

when the device was removed, which authors reported as the children's level of compliance with the study protocol. The device was reported as being used 98% of the time during the 5 days of data collection [35]. Second, the children's self-reports indicated that 74% of them checked the device several times per day for updates on their activity. In addition, 62% of children reported removing the device daily and 29% reported forgetting to put it back on [45].

Software Use

In total, 2 (13%) of the 15 articles quantitatively reported on partner app use. One study detailed children's use as indicated by parents in a web-based survey. While the article primarily focused on parents' use, the authors reported that 42% of children used the Garmin app 2 to 6 times per week during the study [42]. When using the Fitbit Charge HR [45], 80% of children reported using the associated app to see activity information during the study period.

Experience Using the Devices

Of the 15 studies, 5 (33%) provided quantitative user feedback on the acceptability and feasibility of using the devices and associated software [35,37,42,44,45], most of which were collected via user surveys or questionnaires. Moreover, 1 (7%) article reported the frequency of common interview responses [35].

After using the Fitbit Charge HR, 98% of children reported feeling comfortable wearing the device among their peers, 87% enjoyed wearing the device over the deployment period, and the majority (98%) would be interested in wearing the device for longer in the future [45]. Most children (23/24, 96%) rated the Movband as the preferred device compared to the wrist-worn tracker without a display, and the hip-worn tracker (not discussed in this review) due to its comfort (10/23, 43%) and its feedback features (8/23, 35%) [35]. Parents reported the perceived usefulness of the Garmin Vivofit JR device and partner app, with 85.9% and 76.6% of parents rating it as very or quite useful, respectively [42]. In the visual directive intervention for ASD by O'Brien et al [37], the overall intervention was evaluated by school staff via a treatment acceptability survey with an average score of 42.75 out of 50. All participants agreed that it was an acceptable method for school staff to deliver directives; however, 2 participants did not think that the treatment was likely to result in permanent improvement.

In total, 2 (13%) of the 15 studies described experiencing data loss due to difficulties with the wearables, such as syncing errors leading to data corruption, user error, or insufficient wear [9,15]. In a study, the authors reported that 1 family withdrew their participation due to issues with setting up the devices [44].

Qualitative Data on Usability, Acceptability, and Feasibility

Overview

Of the 15 studies, 7 (47%) included qualitative data on users' experience of using the devices and therefore were included in the thematic synthesis [35,36,39,41-44,49]. Of the 7 studies which collected qualitative data, 5 (71%) provided quotations from participants, which can be found in Table 5. We identified 3 themes: the perceived ease of use (TAM), perceived usefulness (TAM), and the perceptions of engagement.

XSL•FO
RenderX

Table 5. Participant quotes extracted from studies (primary data).

Theme, study, and supporting quotes	Participant
Perceived ease of use	
Schoeppe et al [41], 2023	
"Sometimes it wouldn't sync properly or I thought it was syncing and I am not sure if it actually did"	Parent
"My son was getting a bit of a rash from wearing it all the time.. it would come undone really easily"	Parent
"I didn't find it particularly easy but then I am not really technology proficient"	Parent
Creaser et al [44], 2022	
"It hurt and it—it made a mark on my arm"	Child aged 5 years
Perceived usefulness	
Creaser et al [44], 2022	
"I don't really understand the calories that I've burnt. But when I am older I'll probably understand more"	Child aged 10 years
"It really did depend if I was in class or not [if they responded to 'reminders to move'] because I couldn't just run out of class to get the steps...sometimes if fit into my break so that would be good"	Child aged 9 years
"He's only 5 if he was a little bit older, I would consider getting him one"	Parent
"Maybe when I am older, but I don't think so right now"	Child aged 10 years
"It had like 2 more steps to do so I did those two more steps and then noticed that it wasn't picking it up"	Child aged 10 years
"Well I didn't really want to compare them because I thought they might be a bit more than me...I was going to do a thing where at the end of the day there's a winner for how many—for the biggest amount of steps, but I quitted that because I saw mum did a lot of steps at work...it made me jealous"	Child aged 7 years
Masteller et al [36], 2017	
"The MovBand I didn't really like it because you couldn't realty do anything active, you just look at it which wasn't really fun to me. It did have one advantage which was seeing if you're average or not average"	Child aged 9 years
"This one I thought was the least interesting because you could barely do anything"	Child aged 10 years
Schoeppe et al [41], 2023	
"My son often got told by the teacher to put it away as it was a bit distracting for him in the classroom"	Parent
"It actually made us a little more aware of our child's sleep patterns. Something we've investigated a little further"	Parent
"Kids were disappointed when swimming and cycling did not record as steps but they could have still entered them"	Parent
"Difficulty came when the kids wanted to win...[child] is cheating because he is moving his arm around...but he is going to be top of the leaderboard"	Parent
Saksono et al [39], 2018	
"I don't think hers is accurate. Because sometimes if tells you, 'she's at 10,000 steps,' I'm like 'you didn't do 10,000 steps. Like I moved more than you did'"	Parent
Saksono et al [49] , 2019	
"We just look at them [...] I mean, it's pretty much like just look at the, see what the numbers say. Um and it's I leave it"	Parent
"I've told them, but they be so into their the world. [...] They just be like brushing me off. [...] They're used to walking around, and moving and activities. [...] They don't know that they're completing the things on the challenges [in the Kid Power app]"	Parent
Experience using the devices	
Creaser et al [44], 2022	
"I already saw by the end of the week the novelty was wearing off"	Parent
"I liked uh I liked going on walks because you could check how much steps you've done"	Child aged 10 years
"I like the firework when you get to 10,000"	Child aged 6 years
"I didn't really like the sleep bit in it because of that...because they [parents] could check when I was sleeping"	Child aged 12 years

XSL•FO
RenderX

Theme, study, and supporting quotes	Participant
Saksono et al [49], 2019	
"I give [Kid Power app a] 10 [out of 10] because like I love walking. And like I love, love to walk and it gives you missions. And points, which I really love points"	Child aged 9 years
Schoeppe et al [41], 2023	
"The kids absolutely loved the kids app with the little activities and getting coins for things"	Parent

Theme 1: Perceived Ease of Use

In total 2 of the 15 (13%) studies reported that children reported finding wearing the devices uncomfortable and, in some cases, wearing them caused skin irritation [41,44]. In addition, parents described finding the devices and partner apps difficult to use and revealed practical and technical challenges with the devices, such as syncing them and the partner app [41,44]

Theme 2: Perceived Usefulness

Most parents and children expressed a positive experience of using wearables overall, with many considering using a device again in the future. However, in a study, there was some discussion regarding the appropriateness of using these devices with younger children. For example, it was highlighted that children cannot respond to the hourly "reminder to move" prompts when they conflict with their school schedule. Children in the study also described challenges with interpreting the data output and visual display of the Fitbit device [44]. Some parents also highlighted some negative consequences of using the devices, such as the devices becoming distracting during school and causing negative emotions or sibling conflict due to competition elements [41,44].

Articles reported that users liked receiving feedback on their activity, with many families describing how this helped increase motivation for engaging in PA [39,41,42,44]. However, children and parents highlighted their concerns regarding the accuracy of the devices to be able to detect and record PA [39,44]. In addition, children were disappointed that not all activities (eg, swimming) could be tracked using some of the devices and thus did not contribute to their overall activity summary [41,44].

In a study, parents described facing challenges when trying to get their child to engage with how the wearable works and to explore the additional features of the companion app. In addition, families reported having limited discussions about the PA data collected via the wearables, including the meaning of the data within the context of their overall health and well-being [49].

Theme 3: Perceptions of Engagement

Overall, children enjoyed using the wearables and their partner apps. Step count was frequently reported as a popular feature along with gamified features, such as rewards [35,36,41,44,49]. Devices with these additional features were preferred over those that supported little or no interaction [36]. However, some parents reported that children's engagement with the devices declined over time due to a novelty effect resulting in a loss of interest [41,44].

Only 1 (7%) of the 15 studies described participant comments in relation to sharing data among participants. Parents liked that the trackers captured additional data (eg, sleep patterns) and described using the partner app to monitor their child often without their involvement [41]. However, there was no exploration of children's comfort with sharing their data with other stakeholders.

Discussion

Principal Findings

This systematic review aimed to answer three RQs by identifying (1) the population characteristics of children targeted in health-related wearable interventions, (2) the characteristics of the devices used, and (3) the feasibility and acceptability of using these devices in wearable interventions for children aged 5 to 11 years. We found that most (10/13, 77%) of the studies focused on PA, with only 23% (3/13) focusing on clinical conditions. The devices were typically used to provide activity feedback to support behavior change, with additional features often available on an accompanying app. Many (10/13, 77%) studies included wider stakeholders (parents, teachers, and peers) and data sharing, either as part of the intervention or implicitly due to the device being linked to a partner app. The quantitative and qualitative user feedback shows that using smartwatches for health interventions is acceptable and feasible with children aged 5 to 11 years, though there were common usability issues (eg, discomfort, technical issues, and short battery life). Overall, several studies included in this review lacked sufficient detail for data extraction, were considered not to have a rigorous data analysis procedure, and thus are considered to be of poor quality (Multimedia Appendix 3 provides full quality assessment). Sample sizes varied from n=1 to n=137 and small sizes limit our ability to generalize their findings, which should be rectified in future research. There was a limited focus on child-centered design, with many studies using commercial adult devices that require parent involvement rather than developing bespoke child-friendly designs. Furthermore, when gathering user feedback, children's views were often not the focus, with 3 of the 12 (25%) studies not gathering feedback from children but instead focusing on their parents.

Comparison With Prior Work

All of the studies in this review used existing commercial smartwatches and activity-tracking devices. These provide limited data feedback, often only reporting step count on the device. More detailed information was generally only available through a partner app, which was usually under the control of a parent. This is consistent with previous research, which found that children were frustrated by the limited data feedback available in real time due to it being on the app rather than the device [50]. The devices and their partner apps included a

XSL•FO
RenderX

number of behavior change techniques, such as feedback or monitoring and goal setting, both of which are consistent with previous child and adult research [51,52]. The studies also identified usability issues, including discomfort, technical challenges such as syncing issues, and logistical challenges such as poor battery life, which is consistent with previous research [50,53,54].

Despite this, the studies in this review reported high levels of engagement and enjoyment using these devices and their partner apps. In total, 2 (13%) of the 15 studies also reported a novelty effect, indicated by declining use across the study period. This effect has also been observed with adults' use of wearable devices, where it is common to see initial high engagement levels, followed by abandonment after a few months [55]. One potential way to combat abandonment and support motivation is through the use of gamification, as research has shown that gamified apps have longer use than nongamified apps for PA [56]. Furthermore, while the children were mostly content with wearing the devices, in one of the studies participants described that children were not actively engaging with the data collected via the device. This highlights an important distinction between engagement with the data collection process and the use of or reflection on the data collected. According to the staged-based model of personal informatics by Li et al [57], reflection is an important distinct stage that takes place before action, in which individuals can enact behavior change with their new information. Previous research has shown that the way people reflect on their data can be impacted by both the phase of self-tracking and the feedback they receive such as data visualizations, which help individuals gain insights from their data [58]. This is particularly relevant as children can struggle to understand the data collected and visualized on smartwatches [59], which, in turn, may impact their engagement. For example, research with children aged 10 to 15 years has shown that children can misinterpret biofeedback metrics, including heart rate and calories [60,61]. Therefore, further work is needed to understand how to support children's sensemaking of personal health data collected via wearables.

To ensure that children understand and can reflect on the data being collected via smartwatches, it is important to consider their needs and preferences by including them and other stakeholders in a user-centered design process [62]. Considering the end users' experience and including these individuals in the design process has been shown to produce technology, which is more engaging and acceptable to users [63,64]. This is particularly relevant when using children's wearables in health interventions, as research has shown that children conceptualize health differently from adults. It was found that children extended their understanding of health beyond traditional health measures to include other metrics, such as time spent playing outside [59], highlighting the importance of exploring children's health-tracking interests. However, many studies in this review did not seem to take a user-centered approach to the design and choice of devices in their studies. For example, a few (3/13, 23%) studies in this review outlined the rationale for their choice of wearable and failed to detail the specific design features of the device or software that supported children to use it in relation to their health. In addition, most of the devices used across the studies were adult devices and not designed specifically for use by children. Moreover, the device partner apps included additional features and behavior change techniques, such as leaderboards, games, and rewards; however, these were often only accessible via a parent's smartphone as young children did not have their own smartphone. This can lead to the issue identified in previous research of parents acting as "gatekeepers" to their child's interaction with an intervention. This dynamic can result in children no longer being the focus of the design and generates additional "invisible" work for parents [50,51].

However, including parents in pediatric behavioral interventions has previously been shown to be a key component for achieving sustained behavioral change [8,12]. In this review, 77% (10/13) of the studies included additional stakeholders such as parents and teachers, but it was rare that they considered the child's perspective on their inclusion. Ethical and privacy issues around collecting and sharing self-tracked data via wearables are rarely discussed [65]. Potapov and Marshall [66] found that teenagers had concerns about the consequences of sharing their data with others; however, this has not been thoroughly investigated in younger children. In this review, privacy concerns were not directly explored or raised by younger children. However, previous research that has investigated the impact of sharing children's PA and sleep data with their parents or carer found that children reported being uncomfortable with the surveillance. Furthermore, shared access to the data appeared to influence trust between children and their parents or carers negatively, as parents used the data to assert further parental control and impose additional restrictions [67]. Similarly, when using co-design methods with families with children aged 7 to 15 years, it was found that parents and children raised concerns about the impact of data sharing on children's privacy and autonomy, to which the authors suggested the need for child-controlled data sharing options to help set boundaries and minimize conflict [68]. As these privacy concerns were raised in studies with children older than the focus of this review, this may be indicative of a change that occurs during the transition to adolescence, whereby children are working toward increased independence and autonomy [69]. A robust investigation of end users' views (children, parents, carers, and other stakeholders) on wearable data sharing and data privacy is an important research gap, with significant implications for design consideration for wearable interventions for children. This should be considered in future research, where it would be valuable to investigate both children's current preferences for data sharing and privacy as well as their shifting views as they mature and the sociotechnical design considerations that could meet these changing needs.

In this review, most (9/13, 69%) studies focused on using smartwatches and activity trackers to promote PA. Future work should consider the potential of using these devices for more novel health applications. One growing use case for smartwatches within health care and health care research is just-in-time adaptive interventions (JITAIs) [70,71]. JITAIs use mobile and sensing technology to offer personalized and timely support [72] and have been used in a variety of adult and adolescent health domains including PA [70] and substance use behavior [71,73]. In this review, 23% (3/13) of the studies used

XSL•FO
RenderX

the devices to deliver interventions in real time. This included using the watch display and SMS text messages to deliver emotional regulation strategies to children with ASD. However, most of the devices in this review only provided basic data summaries on the watch face. Only one study reported the use of a novel visualization using data analytics, in which an animated dragon was used to convey a risk score as part of a JITAI. This highlights the immense potential for novel uses of smartwatches in pediatric health and further work is needed to explore how smartwatch interventions can be adapted to be developmentally appropriate to convey health information to children in a way that is meaningful to support behavior change.

Strengths and Limitations

Digital health is an interdisciplinary field, and a key strength of this review is the interdisciplinary approach taken by searching both health science databases as well as human-computer interaction venues. The review aimed to investigate how wearables are designed and used for children (5-11 y) to understand developmentally sensitive design implications. To investigate this, we looked across the broad spectrum of child health behaviors and conditions, which means our review is not limited to PA interventions, making the findings relevant to a broader scope of researchers working within pediatric wearable interventions. However, we excluded studies that included our target age group (5-11 y) but also included older and younger children (eg, 5-18 y) that did not report the results separately, as it was not feasible to stratify the results by age. We also conducted a narrative synthesis, which has previously been criticized for lacking transparency [74]. To mitigate this, we have included primary data where possible in our qualitative summary. The limited qualitative data available in the articles also resulted in our synthesized themes having limited richness.

Conclusions

This systematic review addresses the gap in understanding children's wearables by expanding the scope of the review to a broad range of health applications, with a focus on identifying developmentally sensitive design features and understanding the experience of younger children. We identified a total of 15 articles (13 distinct studies) that involved children aged 5 to 11 years using a smartwatch for health-related purposes. Most of these were using commercial smartwatches and activity trackers as tools to help increase PA, but other novel applications included uses to support children with ASD and asthma. Of those that gathered user feedback, experiences using the devices were positive, although there was an indication that the children may not have been engaging with the visualization of their data. Overall, the studies lacked a child-centered focus with (1) many (9/15, 60%) studies not including children's opinions reflecting on the designs, (2) most of the functionality and data visualization was placed on partner apps usually under the control of parents, and (3) most smartwatches and activity trackers being used were adult devices. Further work is needed to better understand children's perspectives on the design and evaluation of these devices while also investigating other potential areas of research beyond PA tracking for children, such as managing chronic health conditions. In addition, future research should provide richer details about the devices used, including the rationale for the choice of device, and who has access to the data to consider how this aligns with their research aim and user experience.

Acknowledgments

LT and SC are funded by the Engineering and Physical Sciences Research Council as part of PhD studentships.

Authors' Contributions

LT conceptualized this work in collaboration with the coauthors. LT carried out the search strategy and conducted the article screening with SC. LT and AB performed the quality assessment. LT wrote the initial manuscript. SC, AB, JB, and PM reviewed, refined, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 71 KB - jmir_v27i1e62944_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 14 KB - jmir_v27i1e62944_app2.docx]

Multimedia Appendix 3
Taxonomy of app features (as reported).
[DOCX File , 22 KB - jmir_v27i1e62944_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Quality assessment.
[XLSX File (Microsoft Excel File), 20 KB - jmir_v27i1e62944_app4.xlsx]

References

1. Kessler RC, McLaughlin KA, Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, et al. Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. Br J Psychiatry 2010 Nov 02;197(5):378-385 [FREE Full text] [doi: 10.1192/bjp.bp.110.080499] [Medline: 21037215]
2. Wadsworth ME, Kuh DJ. Childhood influences on adult health: a review of recent work from the British 1946 national birth cohort study, the MRC National Survey of Health and Development. Paediatr Perinat Epidemiol 1997 Jan 25;11(1):2-20. [doi: 10.1046/j.1365-3016.1997.d01-7.x] [Medline: 9018723]
3. Galobardes B, Smith GD, Lynch JW. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. Annals of Epidemiology 2006 Feb;16(2):91-104. [doi: 10.1016/j.annepidem.2005.06.053] [Medline: 16257232]
4. Nader PR, O'Brien M, Houts R, Bradley R, Belsky J, Crosnoe R, National Institute of Child Health and Human Development Early Child Care Research Network. Identifying risk for obesity in early childhood. Pediatrics 2006 Sep;118(3):e594-e601. [doi: 10.1542/peds.2005-2801] [Medline: 16950951]
5. Dalton ED, Hammen CL, Brennan PA, Najman JM. Pathways maintaining physical health problems from childhood to young adulthood: the role of stress and mood. Psychol Health 2016 Nov 08;31(11):1255-1271 [FREE Full text] [doi: 10.1080/08870446.2016.1204448] [Medline: 27329508]
6. Delaney L, Smith JP. Childhood health: trends and consequences over the life course. Future Child 2012 Mar;22(1):43-63 [FREE Full text] [doi: 10.1353/foc.2012.0003] [Medline: 22550685]
7. Van Cleave J, Gortmaker SL, Perrin JM. Dynamics of obesity and chronic health conditions among children and youth. JAMA 2010 Feb 17;303(7):623-630. [doi: 10.1001/jama.2010.104] [Medline: 20159870]
8. Brigden A, Parslow RM, Linney C, Higson-Sweeney N, Read R, Loades M, et al. How are behavioural interventions delivered to children (5-11 years old): a systematic mapping review. BMJ Paediatr Open 2019 Dec 10;3(1):e000543 [FREE Full text] [doi: 10.1136/bmjpo-2019-000543] [Medline: 31909219]
9. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
10. Wang J, Wang Y, Wei C, Yao N, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
11. McDermott MS, While AE. Maximizing the healthcare environment: a systematic review exploring the potential of computer technology to promote self-management of chronic illness in healthcare settings. Patient Educ Couns 2013 Jul;92(1):13-22. [doi: 10.1016/j.pec.2013.02.014] [Medline: 23566427]
12. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020 Jul 31;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]
13. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb 2;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
14. Butler S, Sculley D, Santos D, Girones X, Singh-Grewal D, Coda A. Using digital health technologies to monitor pain, medication adherence and physical activity in young people with juvenile idiopathic arthritis: a feasibility study. Healthcare (Basel) 2024 Feb 02;12(3):392 [FREE Full text] [doi: 10.3390/healthcare12030392] [Medline: 38338277]
15. Zehrung R, Huang L, Lee B, Choe EK. Investigating opportunities to support kids' agency and well-being: a review of kids' wearables. arXiv Preprint posted online April 13, 2021 [FREE Full text]
16. Radesky JS, Weeks HM, Ball R, Schaller A, Yeo S, Durnez J, et al. Young children's use of smartphones and tablets. Pediatrics 2020;146(1):12 [FREE Full text]
17. Reeder B, David A. Health at hand: a systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
18. Jat AS, Grønli TM. Smart watch for smart health monitoring: a literature review. In: Proceedings of the 9th International Work-Conference on Bioinformatics and Biomedical Engineering. 2022 Presented at: IWBBIO '22; June 27-30, 2022; Gran Canaria, Spain p. 256-268 URL: https://link.springer.com/chapter/10.1007/978-3-031-07704-3_21 [doi: 10.1007/978-3-031-07704-3_21]
19. Grave J, Blissett J. Is cognitive behavior therapy developmentally appropriate for young children? A critical review of the evidence. Clin Psychol Rev 2004 Aug;24(4):399-420. [doi: 10.1016/j.cpr.2004.03.002] [Medline: 15245828]
20. Erickson SJ, Gerstle M, Feldstein SW. Brief interventions and motivational interviewing with children, adolescents, and their parents in pediatric health care settings: a review. Arch Pediatr Adolesc Med 2005 Dec 01;159(12):1173-1180. [doi: 10.1001/archpedi.159.12.1173] [Medline: 16330743]

XSL•FO
RenderX

21. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
22. Gram-Hansen SB. Family wearables – what makes them persuasive? Behav Inf Technol 2019 Nov 26;40(4):385-397. [doi: 10.1080/0144929X.2019.1694993]
23. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009 Aug 18;151(4):W65-W94 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00136] [Medline: 19622512]
24. Smith PG, Morrow RH, Ross DA. Field Trials of Health Interventions: A Toolbox. Oxford, UK: Oxford Academic Press; 2015.
25. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
26. Popay J, Roberts H, Sowden A, Petticrew M. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. Lancaster University. 2006. URL: https://tinyurl.com/a3fsm46a [accessed 2024-04-29]
27. Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris J, et al. Qualitative evidence. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: Cochrane; 2019:525-545.
28. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man Mach Stud 1993 Mar;38(3):475-487. [doi: 10.1006/imms.1993.1022]
29. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: a taxonomy of app features based on self-determination theory. Int J Hum Comput Stud 2020 Aug;140:102449. [doi: 10.1016/j.ijhcs.2020.102449]
30. Jansen-Kosterink S, Broekhuis M, van Velsen L. Time to act mature-Gearing eHealth evaluations towards technology readiness levels. Digit Health 2022 Jul 11;8:20552076221113396 [FREE Full text] [doi: 10.1177/20552076221113396] [Medline: 35847525]
31. C.A.S.P. qualitative studies checklist. Critical Appraisal Skills Programme (CASP). URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2024-04-29]
32. Protogerou C, Hagger MS. A checklist to assess the quality of survey studies in psychology. Methods Psychol 2020 Dec;3:100031. [doi: 10.1016/j.metip.2020.100031]
33. Quality assessment of controlled intervention studies. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2024-04-29]
34. Quality assessment tool for before-after (pre-post) studies with no control group. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2024-04-29]
35. Schaefer SE, Van Loan M, German JB. A feasibility study of wearable activity monitors for pre-adolescent school-age children. Prev Chronic Dis 2014 May 22;11:E85 [FREE Full text] [doi: 10.5888/pcd11.130262] [Medline: 24854236]
36. Masteller B, Sirard J, Freedson P. The physical activity tracker testing in youth (P.A.T.T.Y.) study: content analysis and children's perceptions. JMIR Mhealth Uhealth 2017 Apr 28;5(4):e55 [FREE Full text] [doi: 10.2196/mhealth.6347] [Medline: 28455278]
37. O'Brien AM, Schlosser RW, Shane H, Wendt O, Yu C, Allen AA, et al. Providing visual directives via a smart watch to a student with autism spectrum disorder: an intervention note. Augment Altern Commun 2020 Dec 11;36(4):249-257. [doi: 10.1080/07434618.2020.1862299] [Medline: 33423542]
38. Buchele Harris H, Chen W. Technology-enhanced classroom activity breaks impacting children's physical activity and fitness. J Clin Med 2018 Jun 29;7(7):165 [FREE Full text] [doi: 10.3390/jcm7070165] [Medline: 29966308]
39. Saksono H, Castaneda-Sceppa C, Hoffman J, El-Nasr MS, Morris V, Parker AG. Family health promotion in low-SES neighborhoods: a two-month study of wearable activity tracking. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-13 URL: https://dl.acm.org/doi/10.1145/3173574.3173883 [doi: 10.1145/3173574.3173883]
40. Schoeppe S, Salmon J, Williams SL, Power D, Alley S, Rebar AL, et al. Effects of an activity tracker and app intervention to increase physical activity in whole families-the step it up family feasibility study. Int J Environ Res Public Health 2020 Oct 20;17(20):7655 [FREE Full text] [doi: 10.3390/ijerph17207655] [Medline: 33092219]
41. Schoeppe S, Waters K, Salmon J, Williams SL, Power D, Alley S, et al. Experience and satisfaction with a family-based physical activity intervention using activity trackers and apps: a qualitative study. Int J Environ Res Public Health 2023 Feb 14;20(4):3327 [FREE Full text] [doi: 10.3390/ijerph20043327] [Medline: 36834022]
42. Schoeppe S, Salmon J, Williams S, Power D, Waters K, Alley S, et al. Feasibility of using activity trackers and apps to increase physical activity in whole families: the step it up family intervention. Digit Health 2022 Oct 06;8:20552076221129083 [FREE Full text] [doi: 10.1177/20552076221129083] [Medline: 36225986]

XSL•FO
RenderX

43. Torrado JC, Gomez J, Montoro G. Emotional self-regulation of individuals with autism spectrum disorders: smartwatches for monitoring and interaction. Sensors (Basel) 2017 Jun 11;17(6):1359 [FREE Full text] [doi: 10.3390/s17061359] [Medline: 28604607]
44. Creaser AV, Hall J, Costa S, Bingham DD, Clemes SA. Exploring families' acceptance of wearable activity trackers: a mixed-methods study. Int J Environ Res Public Health 2022 Mar 15;19(6):3472 [FREE Full text] [doi: 10.3390/ijerph19063472] [Medline: 35329166]
45. Wing D, Godino JG, Baker FC, Yang R, Chevance G, Thompson WK, et al. Recommendations for identifying valid wear for consumer-level wrist-worn activity trackers and acceptability of extended device deployment in children. Sensors (Basel) 2022 Nov 26;22(23):9189 [FREE Full text] [doi: 10.3390/s22239189] [Medline: 36501894]
46. Hosseini A, Buonocore C, Hashemzadeh S, Hojaiji H, Kalantarian H, Sideris C, et al. Feasibility of a secure wireless sensing smartwatch application for the self-management of pediatric asthma. Sensors (Basel) 2017 Aug 03;17(8):1780 [FREE Full text] [doi: 10.3390/s17081780] [Medline: 28771168]
47. Jackson SL, Kraft GL, Khoo M, Rockie P, Karin C. Enhancing the fitness and academics of children using technology in the schools (FACTS). J Phys Educ Sport 2022;22(11):2810-2817 [FREE Full text]
48. Duck AA, Hall KC, Klamm M, Temple M, Robinson JC. Physical activity and fitness: the feasibility and preliminary effectiveness of wearable activity tracker technology incorporating altruistic motivation in youth. J Spec Pediatr Nurs 2021 Jan 24;26(1):e12313. [doi: 10.1111/jspn.12313] [Medline: 32970924]
49. Saksono H, Castaneda-sceppa C, Hoffman J, El-Nasr MS, Morris V, Parker A. Social reflections on fitness tracking data: a study with families in low-SES neighborhoods. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: the chi conference on human factors in computing systems; May 4-9, 2019; Glasgow, UK p. 14 URL: https://dl.acm.org/doi/fullHtml/10.1145/3290605.3300543 [doi: 10.1145/3290605.3300543]
50. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental perspectives of a wearable activity tracker for children younger than 13 years: acceptability and usability study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
51. Oygür I, Epstein DA, Chen Y. Raising the responsible child: collaborative work in the use of activity trackers for children. Proc ACM Hum Comput Interact 2020 Oct 15;4(CSCW2):1-23. [doi: 10.1145/3415228]
52. Düking P, Tafler M, Wallmann-Sperlich B, Sperlich B, Kleih S. Behavior change techniques in wrist-worn wearables to promote physical activity: content analysis. JMIR Mhealth Uhealth 2020 Nov 19;8(11):e20820 [FREE Full text] [doi: 10.2196/20820] [Medline: 33211023]
53. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
54. Fraser H, Thompson L, Crawley E, Ridd MJ, Brigden A. "not one size fits all": the challenges of measuring paediatric health-related quality of life and the potential role of digital ecological momentary assessment: a qualitative study. Qual Life Res 2024 Feb 27;33(2):443-452 [FREE Full text] [doi: 10.1007/s11136-023-03535-6] [Medline: 37889385]
55. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Human Behav 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
56. Maher CA, Olds T, Vandelanotte C, Plotnikoff R, Edney SM, Ryan JC, et al. Gamification in a physical activity app: what gamification features are being used, by whom, and does it make a difference? Games Health J 2022 Jun 01;11(3):193-199. [doi: 10.1089/g4h.2021.0207] [Medline: 35501981]
57. Li I, Day A, Forlizzi J. A stage-based model of personal informatics systems. In: Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; April 10-15, 2010; Atlanta, GA p. 557-566 URL: https://dl.acm.org/doi/10.1145/1753326.1753409 [doi: 10.1145/1753326.1753409]
58. Choe EK, Lee B, Zhu H, Riche NH, Baur D. Understanding self-reflection: how people reflect on personal data through visual data exploration. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthca. 2017 Presented at: PervasiveHealth '17; May 23-26, 2017; Barcelona, Spain p. 173-182 URL: https://dl.acm.org/doi/10.1145/3154862.3154881 [doi: 10.1145/3154862.3154881]
59. Ananthanarayan S, Siek K, Eisenberg M. A craft approach to health awareness in children. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems. 2016 Presented at: DIS '16; June 4-8, 2016; Brisbane, Australia p. 724-735 URL: https://dl.acm.org/doi/10.1145/2901790.2901888 [doi: 10.1145/2901790.2901888]
60. Ankrah EA, Cibrian FL, Silva LM, Tavakoulnia A, Beltran JA, Schuck SE, et al. Me, my health, and my watch: how children with ADHD understand smartwatch health data. ACM Trans Comput Hum Interact 2023 Sep 12;30(4):1-25. [doi: 10.1145/3577008]
61. Cibrian FL, Lakes KD, Tavakoulnia A, Guzman K, Schuck S, Hayes GR. Supporting self-regulation of children with ADHD using wearables: tensions and design challenges. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-13 URL: https://dl.acm.org/doi/10.1145/3313831.3376837 [doi: 10.1145/3313831.3376837]

XSL•FO
RenderX

62. Wiljén A, Chaplin JE, Crine V, Jobe W, Johnson E, Karlsson K, et al. The development of an mHealth tool for children with long-term illness to enable person-centered communication: user-centered design approach. JMIR Pediatr Parent 2022 Mar 08;5(1):e30364 [FREE Full text] [doi: 10.2196/30364] [Medline: 35258466]
63. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012 Sep;Suppl(s2):49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
64. Birnie KA, Campbell F, Nguyen C, Lalloo C, Tsimicalis A, Matava C, et al. iCanCope PostOp: user-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents. JMIR Form Res 2019 Apr 22;3(2):e12028 [FREE Full text] [doi: 10.2196/12028] [Medline: 31008704]
65. Epstein DA, Caldeira C, Figueiredo MC, Lu X, Silva LM, Williams L, et al. Mapping and taking stock of the personal informatics literature. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Dec 18;4(4):1-38. [doi: 10.1145/3432231]
66. Potapov K, Marshall P. LifeMosaic: co-design of a personal informatics tool for youth. In: Proceedings of the 2020 Interaction Design and Children Conference. 2020 Presented at: IDC '20; June 21-24, 2020; London, UK p. 519-531 URL: https://dl.acm.org/doi/10.1145/3392063.3394429 [doi: 10.1145/3392063]
67. Jørgensen MS, Nissen FK, Paay J, Kjeldskov J, Skov MB. Monitoring children's physical activity and sleep: a study of surveillance and information disclosure. In: Proceedings of the 28th Australian Conference on Computer-Human Interaction. 2016 Presented at: OzCHI '16; November 29- December 2, 2016; Launceston, Australia p. 50-58 URL: https://dl.acm.org/doi/10.1145/3010915.3010936 [doi: 10.1145/3010915.3010936]
68. Oygur Ilhan I, Chen Y, Epstein DA. Co-designing for the co-use of child-owned wearables. In: Proceedings of the 22nd Annual ACM Interaction Design and Children Conference. 2023 Presented at: IDC '23; June 19-23, 2023; Chicago, IL p. 603-607 URL: https://tinyurl.com/bdeumsmr [doi: 10.1145/3585088.3593868]
69. Daddis C. Desire for increased autonomy and adolescents' perceptions of peer autonomy: "everyone else can; why can't I?". Child Dev 2011;82(4):1310-1326. [doi: 10.1111/j.1467-8624.2011.01587.x] [Medline: 21466540]
70. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
71. Yang MJ, Sutton SK, Hernandez LM, Jones SR, Wetter DW, Kumar S, et al. A just-in-time adaptive intervention (JITAI) for smoking cessation: feasibility and acceptability findings. Addict Behav 2023 Jan;136:107467 [FREE Full text] [doi: 10.1016/j.addbeh.2022.107467] [Medline: 36037610]
72. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
73. Coughlin LN, Nahum-Shani I, Philyaw-Kotov ML, Bonar EE, Rabbi M, Klasnja P, et al. Developing an adaptive mobile intervention to address risky substance use among adolescents and emerging adults: usability study. JMIR Mhealth Uhealth 2021 Jan 15;9(1):e24424 [FREE Full text] [doi: 10.2196/24424] [Medline: 33448931]
74. Campbell M, Katikireddi SV, Sowden A, McKenzie JE, Thomson H. Improving conduct and reporting of narrative synthesis of quantitative data (ICONS-Quant): protocol for a mixed methods study to develop a reporting guideline. BMJ Open 2018 Feb 26;8(2):e020064. [doi: 10.1136/bmjopen-2017-020064]

Abbreviations

ASD: autism spectrum disorder
JITAI: just in-time adaptive intervention
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
TAM: technology acceptance model

Edited by T de Azevedo Cardoso; submitted 05.06.24; peer-reviewed by S Honorio, TA Swain, M Khudair; comments to author 27.09.24; revised version received 18.10.24; accepted 06.12.24; published 27.01.25.

Please cite as:
Thompson L, Charitos S, Bird J, Marshall P, Brigden A
Exploring the Use of Smartwatches and Activity Trackers for Health-Related Purposes for Children Aged 5 to 11 years: Systematic Review
J Med Internet Res 2025;27:e62944
URL: https://www.jmir.org/2025/1/e62944
doi:10.2196/62944
PMID:

XSL•FO
RenderX

©Lauren Thompson, Sydney Charitos, Jon Bird, Paul Marshall, Amberly Brigden. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Rehabilomics Strategies Enabled by Cloud-Based Rehabilitation: Scoping Review

Sejun Oh[1,2], PT, PhD; SangHeon Lee[3], MD, PhD

[1]Korea University Anam Hospital, Seoul, Republic of Korea

[2]Human Behavior & Genetic Institute, Associate Research Center, Korea University, Seoul, Republic of Korea

[3]Department of Physical Medicine and Rehabilitation, Korea University Anam Hospital, Seoul, Republic of Korea

Corresponding Author:
SangHeon Lee, MD, PhD
Department of Physical Medicine and Rehabilitation, Korea University Anam Hospital
145, Anam-Ro, Seongbuk-Gu
Seoul, 02841
Republic of Korea
Phone: 82 2286 1419
Fax: 82 920 5269
Email: rmlsh@korea.ac.kr

Abstract

Background: Rehabilomics, or the integration of rehabilitation with genomics, proteomics, metabolomics, and other "-omics" fields, aims to promote personalized approaches to rehabilitation care. Cloud-based rehabilitation offers streamlined patient data management and sharing and could potentially play a significant role in advancing rehabilomics research. This study explored the current status and potential benefits of implementing rehabilomics strategies through cloud-based rehabilitation.

Objective: This scoping review aimed to investigate the implementation of rehabilomics strategies through cloud-based rehabilitation and summarize the current state of knowledge within the research domain. This analysis aims to understand the impact of cloud platforms on the field of rehabilomics and provide insights into future research directions.

Methods: In this scoping review, we systematically searched major academic databases, including CINAHL, Embase, Google Scholar, PubMed, MEDLINE, ScienceDirect, Scopus, and Web of Science to identify relevant studies and apply predefined inclusion criteria to select appropriate studies. Subsequently, we analyzed 28 selected papers to identify trends and insights regarding cloud-based rehabilitation and rehabilomics within this study's landscape.

Results: This study reports the various applications and outcomes of implementing rehabilomics strategies through cloud-based rehabilitation. In particular, a comprehensive analysis was conducted on 28 studies, including 16 (57%) focused on personalized rehabilitation and 12 (43%) on data security and privacy. The distribution of articles among the 28 studies based on specific keywords included 3 (11%) on the cloud, 4 (14%) on platforms, 4 (14%) on hospitals and rehabilitation centers, 5 (18%) on telehealth, 5 (18%) on home and community, and 7 (25%) on disease and disability. Cloud platforms offer new possibilities for data sharing and collaboration in rehabilomics research, underpinning a patient-centered approach and enhancing the development of personalized therapeutic strategies.

Conclusions: This scoping review highlights the potential significance of cloud-based rehabilomics strategies in the field of rehabilitation. The use of cloud platforms is expected to strengthen patient-centered data management and collaboration, contributing to the advancement of innovative strategies and therapeutic developments in rehabilomics.

(J Med Internet Res 2025;27:e54790) doi:10.2196/54790

KEYWORDS

cloud-based; health; rehabilitation; rehabilomics; strategies

XSL•FO
RenderX

Introduction

Background

Rehabilitation is a crucial component of health care aimed at improving the functional capacity and quality of life for individuals facing various health challenges [1,2]. Traditionally, standardized approaches have been used in rehabilitation, sometimes failing to fully consider the unique needs and characteristics of each patient. However, the need for more personalized rehabilitation has become important after COVID-19 [3]. Advancements in genomics [4], epigenomics [5], proteomics [6], metabolomics [7], and related fields have made personalized rehabilitation approaches more feasible [8,9]. The convergence of rehabilitation and *omics* sciences has given rise to the concept of *rehabilomics* [10]. This concept uses an individual's biological characteristics or biomarkers to tailor rehabilitation approaches, providing a pathway for personalized rehabilitation programs, and enabling the integration of "omics" for studying rehabilitation processes and outcomes [10,11].

Rehabilomics seeks to leverage cutting-edge molecular and genomic information to tailor rehabilitation interventions for optimal patient outcomes [10,12]. By understanding the genetic, proteomic, and metabolic profiles of patients, health care providers can design rehabilitation programs that are not only more effective, but also better aligned with an individual's genetic and physiological makeup [11-13]. This paradigm shift holds immense promise for improving rehabilitation outcomes and enhancing patient experiences [14].

Rehabilomics can build systematic patient information based on large data using cloud-based solutions, use data from medical experts through patient consent, and establish customized treatment and rehabilitation strategies for patients [10,15]. In addition, evaluation- and monitoring-based rehabilitation interventions can be performed at home by collecting quantitative and objective information [16,17]. Owing to the nature of the cloud, there are advantages, such as simultaneous data sharing through the network; however, if the security measures are insufficient, there may be problems, such as data loss or patient information leakage; therefore, it is necessary to prepare for establishing data management system functions and security [18].

Simultaneously, the proliferation of cloud computing technologies has revolutionized data storage, sharing, and analysis in the health care domain based on the Fourth Industrial Revolution [19,20]. Cloud-based solutions offer scalable, secure, and cost-effective platforms for managing vast amounts of patient data, including genomic and clinical information [21]. Particularly, blockchain-based cloud solutions ensure a user-centric electronic health record system, offering enhanced security controls and a platform that securely safeguards health care data, emphasizing privacy and stability [22,23]. This technological convergence of rehabilomics and cloud computing opens up new horizons for data-driven rehabilitation practices, facilitating data sharing among health care professionals (rehabilitation medicine specialists, physical therapists, occupational therapists, speech-language pathologists, and other advocates of rehabilitation-related professionals), researchers, and patients [24].

Despite these exciting possibilities, there remains a need to comprehensively explore the current landscape of cloud-based rehabilitation in the context of rehabilomics [25].

This Review

This scoping review aims to fill this gap by providing an overview of the existing literature and shedding light on key trends, challenges, and opportunities in this emerging field. By examining the integration of cloud-based technologies with rehabilomics strategies, we aim to contribute to understanding how these approaches can be harnessed to optimize rehabilitation interventions and ultimately improve patient outcomes. Therefore, this study examined the latest trends through a scoping review of cloud-based rehabilitation. Accordingly, we sought to derive strategies for rehabilomics. The rehabilomics approach, which uses an individual's genomic, proteomic, and metabolomic information to deliver personalized rehabilitation, has tremendous potential to enhance the effectiveness and patient experience of rehabilitation services. Moreover, the integration of cloud computing technology can facilitate data sharing and collaboration among health care providers, researchers, and patients, thereby opening opportunities to develop innovative, data-driven rehabilitation solutions.

Methods

Literature Search Strategy

The primary objective of this scoping review was to investigate and synthesize current literature trends in cloud-based rehabilitation for rehabilomics strategies. To initiate the research, we conducted searches across major academic databases, including CINAHL, Embase, Google Scholar, PubMed, MEDLINE, ScienceDirect, Scopus, and Web of Science. The search terms and syntax were designed to encompass terms such as "cloud-based rehabilitation," "rehabilomics," "cloud computing in rehabilitation," "personalized rehabilitation," and "omics technologies in rehabilitation." The search included articles that were published between January 2000 and July 2024. The literature review was conducted using a 2-step process. First, we examined the abstracts as a preliminary step toward gaining an initial understanding. This was followed by a more detailed review of the full-text of the selected articles, based on insights gathered from screening the abstract. The study protocol was registered in the Open Science Framework [26,27]. The research questions were as follows:

- Have cloud-based rehabilitation systems been developed and evaluated recently?
- What rehabilomics strategies are based on primary outcomes and user perceptions of cloud-based rehabilitation systems?

Inclusion and Exclusion Criteria

The study adhered to established reporting guidelines (Multimedia Appendix 1), including the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Extension for Scoping Reviews (PRISMA-ScR). Scoping

XSL•FO
RenderX

reviews aim to summarize the breadth of evidence on a topic and identify gaps in future studies. Consistent with best practices for scoping reviews, the authors designed and registered a priori protocol in the Open Science Framework, ensuring rigor and transparency throughout the review process.

Inclusion and exclusion criteria were established to select articles from the search results. The inclusion criteria included papers discussing the integration of cloud computing and rehabilomics strategies, applications in the fields of rehabilitation and health care, data sharing and collaboration, and personalized therapy. Papers unrelated to relevant topics or those lacking sufficient information were excluded. Two independent reviewers conducted a thorough examination of the papers, and the final selections were made through discussion.

Studies were only included when (1) full texts were available and (2) cloud-based rehabilitation was used. Owing to the broad nature of this scoping review, we were not limited by participants, study design, intervention, or outcomes. Exclusion criteria were as follows: (1) full text was not available, (2) papers which were not in English, (3) cloud-based rehabilitation was not used in the study, (4) unclear rehabilomics, and (5) studies focusing only on rehabilitation.

Data Extraction and Analysis

Relevant information was extracted and analyzed from the selected papers. This includes details such as paper titles, authors, publication years, research topics, methodologies, key findings, and conclusions. Extracted data were synthesized in relation to key themes and subjected to statistical and content analyses to ascertain current research trends in cloud-based rehabilitation and rehabilomics strategies. The data were analyzed by 2 researchers, of whom one had >15 years of experience as a physical therapist and 20 years of experience as a rehabilitation medicine specialist. The data were analyzed through a basic literature search and review, a detailed review and screening, and a final review and decision.

Reporting of Findings

The results from this scoping review are presented as a comprehensive overview of key themes, research trends, and related topics. This review summarizes the current research trends in cloud-based rehabilitation and rehabilomics strategies and offers suggestions for future research. Overall, 214,040 databases were searched, 2915 records were screened, 58 were retrieved, and 31 were assessed for eligibility.

Results

Characteristics of Included Studies

These studies cover a wide range of topics related to cloud-based rehabilitation and rehabilomics strategies (Multimedia Appendix 2). Figure 1 provides an overview of the selection process based on the PRISMA diagram to identify studies using databases and registers.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram to identify studies using databases and registers.

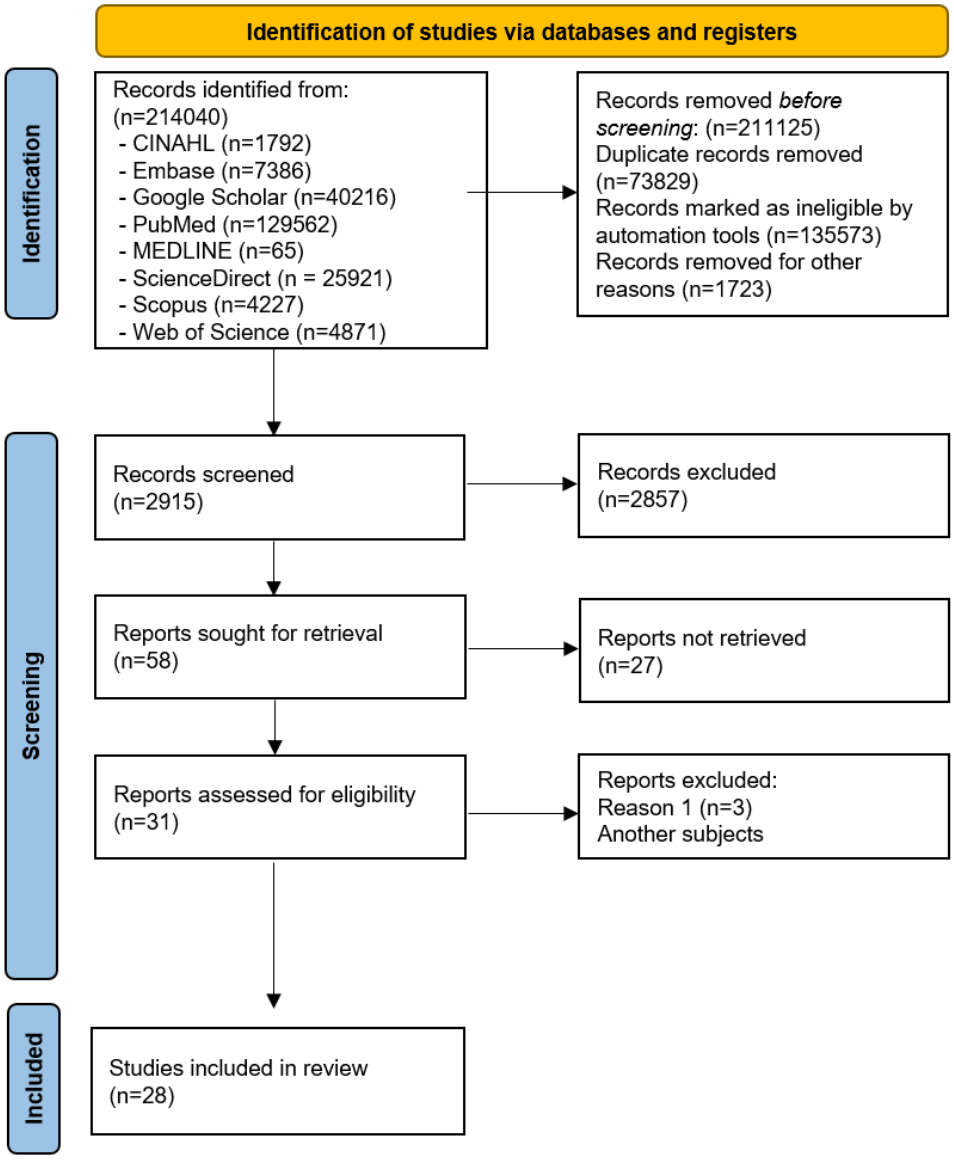

Key Themes

The Analysis of the Included Studies Revealed Several Key Themes

Integration of Cloud Computing

A total of 28 of the included studies [16,17,28-53] explored the integration of cloud computing technologies in the context of rehabilitation and rehabilomics. Cloud platforms are used for data storage, analysis, and sharing to facilitate collaborative research efforts.

Personalized Rehabilitation

Out of 28 studies, 16 (57%) studies [16,17,28,29,33-35,38,39,42,46,48,49,52,53] focused on the concept of personalized rehabilitation by leveraging omics data to tailor rehabilitation programs for individual patients. This approach aims to improve treatment efficacy and patient outcomes.

Data Security and Privacy

Out of 28 studies, 12 (43%) studies [16,17,28,34-36,41,44,47,49,52,53] addressed the critical issues of data security and patient privacy in cloud-based rehabilitation. Ensuring the confidentiality and integrity of patient data has become a key concern.

Interdisciplinary Collaboration

Collaboration between researchers from diverse backgrounds, including rehabilitation specialists, geneticists, and data scientists, was a recurring theme [16,17,28-53]. Such interdisciplinary collaboration is essential for advancing this field.

Cloud-Based Rehabilitation

Overview

We identified and included a total of 28 studies focused on cloud-based rehabilitation. These studies can be categorized into platform, software, and disease and health topics (Figure 2).

XSL•FO
RenderX

Because of the synthesis of the contents of the categories of these studies, the platform included cloud-based frameworks [28], cloud-based systems [2,3,14,23,29], Internet of Things (IoT)–based systems [12,30], user performance evaluation, virtual reality (VR) platforms [19,31], and edge-cloud–based computing. Software included rehabilitation gaming systems [20,22], noninvasive telerehabilitation exercise monitoring [31], Parkinson disease (PD) assessment systems [6], wearable sensors [10], mobile computing [5], deep learning-assisted systems [16], facial paralysis rehabilitation [25], and automation-empowered virtual rehabilitation [33]. Diseases and health conditions included stroke, poststroke disability rehabilitation [8,21,24], neurorehabilitation [13], PD diagnosis and monitoring [17], psychometric test validation [47], developmental language disorders in children [51], and diagnosis of sports injuries [52]. It also included health monitoring and fall detection in older people [35].

Figure 2. Trends of cloud-based rehabilitation.

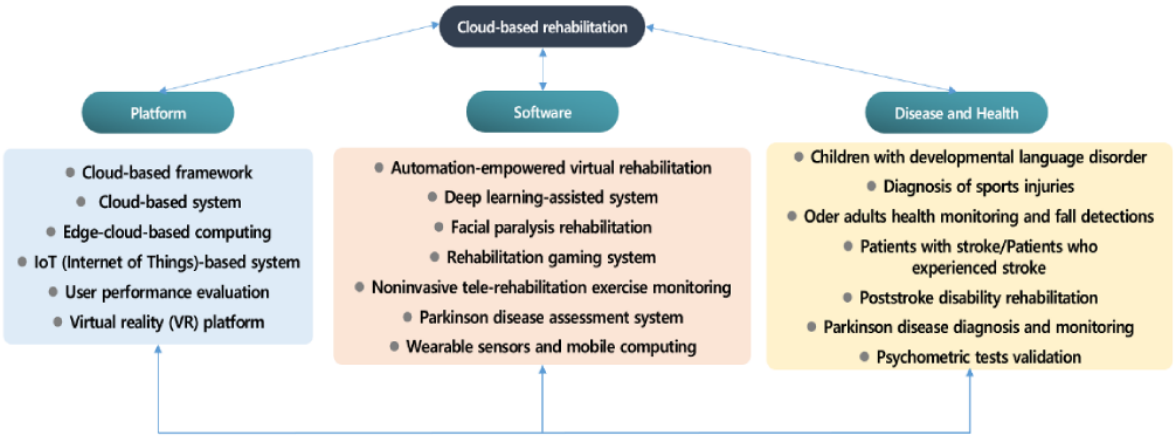

Cloud

CRehab

This refers to a cloud-based framework for rehabilitation management [28].

The authors introduced "CRehab," a cloud-based framework featuring web-based and augmented reality support aimed at managing the rehabilitation processes. This framework targets the administrative, communication, and patient-motivation aspects by leveraging cloud capabilities. The primary objective was to devise an architecture adaptable to diverse rehabilitation types, irrespective of the diagnosis, environment, or patient demographics.

Rehab-aaService

This refers to a cloud-based motor rehabilitation digital assistant [29]. The integration of body sensor networks with cloud computing for motor rehabilitation enables remote monitoring, reduces costs, and provides comfort and safety to patients. Rehab-aaService, the newly introduced motor rehabilitation digital assistant, is based on wearable motion sensor nodes, a personal mobile device, and a cloud-based backend, offering enhanced support and a more comprehensive rehabilitation experience.

Open-Source Cloud-Based Platform for Health Monitoring and Fall Detection in Older People

The authors designed an open-source eHealth monitoring and fall detection system that integrated medical sensors, a microcontroller, and cloud communication [35]. This system stores data in the cloud to aid diagnosis, real-time monitoring, and rehabilitation facilitation. In addition, it features a fall detection system that promptly alerts medical personnel in the event of unexpected falls.

Platforms

Cloud-Based Rehabilitation Exergames System

A low-cost rehabilitation gaming system using Microsoft Kinect was developed for individuals with chronic arm pain [30]. This system encompasses goal-oriented exercises, enabling accurate performance measures for quantitative evaluation of the treatment plan. It notably boosts user motivation, hastens muscle recovery, and provides therapists with a means to assess patient progress.

Development of a VR Platform for Cloud-Based Neurorehabilitation

The authors designed a cloud-based VR platform for neurorehabilitation to analyze its impact on the sense of agency and ownership in healthy individuals [40]. The experimental platform featured a virtual avatar that replicated the individual's movements using motion capture technology and a head-mounted display. The findings revealed that perceived arm length changed according to the displayed arm length of the avatar, demonstrating the considerable potential of the VR system.

SIGVERSE: a Cloud-Based VR Platform for Human-Robot Interaction Research

The authors introduced a cloud-based VR platform to explore human-robot interaction and obtain social intelligence for robots [50]. This successfully showcased the system's practicality in real human-robot interaction scenarios, indicating its potential for fostering social intelligence development.

Edge-Cloud–Based Wearable Computing for Automation-Empowered Virtual Rehabilitation

This study referred to technology-enhanced rehabilitative care for patients with stroke through the use of edge cloud and IoT

XSL•FO
RenderX

technology [53]. This involved the development of an edge-cloud model to track the vital signs and rehabilitation progress of patients in real time. The system exhibited remarkable operational efficiency and significantly improved rehabilitation treatment outcomes, paving the way for more advanced and intelligent medical service systems.

Hospitals and Rehabilitation Centers

Cloud-Based Rehabilitation System for Patients With Stroke

A cloud-based rehabilitation system focused on improving motor function in patients with stroke was developed. This study used data from healthy participants to accurately assess hand function [33]. Using the autoregressive integrated moving average model with the dynamic time-warping algorithm, this study made predictions regarding the rehabilitation and recovery of patients with stroke. Over a 10-week period involving 3 patients, the system showed significant improvements in upper limb control.

Monitoring of Upper Limb Rehabilitation and Recovery after Stroke: An Architecture for a Cloud-Based Therapy Platform

The Limbs Alive project was introduced at the Newcastle University and is dedicated to using video games for therapeutic movements to assist patients with stroke [34]. The authors developed a cloud-based therapy platform that gathers data from therapeutic games played by patients with stroke, estimates performance metrics, and delivers results to patients and clinicians via web applications.

System-Level Design of a Cloud-Based Training Device for Upper Limb Spasticity Rehabilitation

A robotic part-task trainer was developed to evaluate upper limb spasticity using cloud-stored clinical data [39]. It enables the remote coaching of trainees and continuous updates by rehabilitation physicians, thereby fostering a dynamic learning environment.

Cloud-Based Training and Evaluation System for Facial Paralysis Rehabilitation

This study referred to focusing on aiding facial paralysis rehabilitation via cloud-based information and communication technologies [46]. The proposed system aims to support patients and physicians by delivering rehabilitation training, automating progress reviews, and evaluating the results. It involves training a client for patient rehabilitation and data collection, leveraging machine learning on a cloud platform to analyze outcomes.

Telehealth

Cloud-Based Noninvasive Telerehabilitation Exercise Monitoring

The authors developed a conceptualized cloud-based telerehabilitation framework using cost-effective, noninvasive Microsoft Kinect for rehabilitation exercises [31]. This framework enables patients to engage in rehabilitation exercises at home while maintaining high-quality care standards. The initial experiments validated the potential of the system for training healthy participants to mimic physical rehabilitation exercise motions.

Mobile Cloud–Based PD Assessment System

The authors designed "PD Dr," a mobile health app aimed at gathering PD-related motion data and evaluating symptoms [16]. The system was examined using information from 40 patients with PD to validate its efficiency in capturing significant symptoms and estimate their severity. It demonstrated robust connections between the collected motion features and the severity of PD, hand-resting tremors, and gait difficulty.

Cloud-Based PD Diagnosis and Monitoring Framework

The authors suggested a cloud-based framework designed to detect and monitor patients with PD, particularly in low-resource settings [17]. The goal was to provide remote diagnosis for patients in areas with limited health care access, achieving high accuracy in the process.

Remote Cloud-Based Automated Stroke Rehabilitation Assessment Using Wearables

These studies showcase a broad spectrum of cloud-based rehabilitation applications, ranging from pediatric developmental language disorder rehabilitation to stroke rehabilitation assessments using wearable technology [45]. These efforts aim to enhance rehabilitation outcomes through innovative technological approaches and data-driven analyses by incorporating cloud computation and sensor-based data for accurate assessment.

IoT Cloud-Based Telerehabilitation Service for Smart Cities

This study was centered on telerehabilitation as a service by leveraging cloud computing, IoT, and big data analytics [49]. It emphasizes the management of significant health care data generated from remote rehabilitation devices within the hospital cloud using NoSQL databases.

Home and Community

Integration of Wireless Wearable Sensors and Mobile Computing with Cloud-Based Service for Patient Rehabilitation Monitoring

The authors proposed a system that used wearable wireless sensors and handheld devices to track body motion, focusing on patients with stroke [36]. This setup allowed patients to conduct rehabilitation exercises at home while health care providers remotely monitored their progress. This system integrates cloud services for remote monitoring, eliminating the need for expensive hardware and software licenses for data storage. Initial user tests demonstrate the feasibility of the system.

IoT-based System for Physical Rehabilitation Monitoring

The authors described an IoT-based wearable system designed for monitoring physical rehabilitation using accelerometer and gyroscope sensors [37]. This system can record movement data to characterize patient movement and estimate their condition, thereby enabling remote monitoring through a cloud-based service with accurate capabilities for monitoring elbow rehabilitation characteristics.

XSL•FO
RenderX

Cloud-Based Platform for Autonomous Home Rehabilitation With Exergames

The authors developed a cloud-based platform that facilitates independent at-home rehabilitation through customizable exercise games [43]. This system uses a flexible database and structured rehabilitation data to allow clinicians to review and fine-tune exercise parameters efficiently. The platform incorporates parameterized exercise games designed to offer personalized goal-oriented rehabilitation.

Cloud-Based Smart Home Environment for Home Health Care

The authors developed "CoSHE," a cloud-based smart home environment for monitoring the health of older populations within their homes [44]. The system incorporates wearable sensors and environmental data to provide real-time information to remote caregivers. The integration of contextual information successfully enhances understanding of an individual's health status.

User Performance Evaluation and Real-Time Guidance in Cloud-Based Physical Therapy Monitoring and Guidance System

A cloud-based physical therapy monitoring and guidance system was developed to address the limitations of traditional therapy owing to the restricted interaction time [48]. The system leverages the actions of a physical therapist recorded as an avatar to guide patients in real time for self-training. By aligning the physical therapy and user motion sequences with a gesture-based dynamic time-warping algorithm, it evaluates user performance based on the physical therapy criteria and provides real-time guidance, optimizing patient learning.

Disease and Disability

Speech-Controlled Cloud-Based Wheelchair Platform for Disabled Persons

A prototype speech-controlled cloud-based wheelchair platform using WebKit Speech Application Programming Interface in the cloud was developed, which introduced a graphical user interface compatible with the web and mobile devices that enabled live video streaming [32]. It achieved speech recognition accuracy between 60% and 97% and recorded latencies spanning from 0.4 to 1.3 seconds. Through clinical testing, it exhibited promising outcomes, signaling potential for future implementation.

Cloud-Based Architecture Proposal for Aphasia Rehabilitation

The authors designed a cloud-based architecture specifically tailored for aphasia rehabilitation in Romanian-speaking patients [38]. The system comprises an application database, logic, and interfaces for patients and therapists. Using a virtual assistant, the system guided patients through exercises crafted by therapists. Treatment adjustments were made based on the obtained scores and the data were stored in a statistical module for further assessment.

Comparison of Python and Java Capabilities for Somatosensory Games in Cloud-Based Rehabilitation

The authors investigated the performance and stability of cloud-based somatosensory games in rehabilitation systems, while handling various user loads [41]. In this study, Python and Java programming languages were compared, and the findings indicated that Python consumed fewer resources in both the simulation and monitoring systems.

Cloud-Supported Framework for Poststroke Disability Rehabilitation

The authors suggested a cloud-based rehabilitation framework that combined augmented reality and sensor gloves for patients after an episode of stroke [42]. Notably, considerable enhancements in finger strength were observed after 6 weeks of therapeutic exercise. Sensor gloves were used to recognize the real-time conditions of the patients engaging in rehabilitative exercises.

Cloud-Based Flexible Solution for Psychometric Test Validation, Administration, and Evaluation

A unified cloud-based resource was developed to manage psychometric tests used for attitude evaluation, personal selection, educational purposes, rehabilitation, and diagnosis of cognitive disorders [47]. This resource aims to simplify the process of validation, standardization, and reorganization of tests traditionally conducted using complex statistical methods.

Effectiveness of Cloud-Based Rehabilitation in Children With Developmental Language Disorders During the COVID-19 Pandemic

This study referred to the application of cloud-based rehabilitation for children with developmental language disorders during the COVID-19 pandemic [51]. The authors investigated the effectiveness of the JingYun Rehab Cloud Platform on language and cognitive outcomes in 162 children with developmental language disorders through a prospective cohort study. Patients who underwent remote cloud-based rehabilitation were compared with a control group that received home-based rehabilitation. The results indicated significantly better language abilities among the children in the cloud-based rehabilitation group.

Cloud-Based Deep Learning–Assisted System for Diagnosis of Sports Injuries

The authors proposed a deep learning-assisted system incorporating IoT sensors in a body area network to diagnose sports injuries using cloud computing resources. IoT sensors gather data for diagnosis of injury and cloud computing to provide flexible computing resources [52]. They developed a brain injury monitoring framework and optimized neural networks to forecast sports injuries and evaluated the model performance using accuracy, precision, recall, and F_1-score metrics.

Our study established a strategy for personalized rehabilomics tailored toward individual needs by integrating clinical information from hospitals and extending it further to genomics, epigenomics, proteomics, and metabolomics for intervention, measurement, assessment, and prognosis (Figure 3). We

XSL•FO
RenderX

emphasized the active use of clouds and platforms. In addition, we derived a strategic direction emphasizing the integration of hospitals and homes, promoting active community-based rehabilitation and linking it with individualized care providers. The first aspect involves a cloud that encompasses rehabilomics big data along with genomics, epigenomics, proteomics, metabolomics, and personalized medicine. The second aspect includes game-based platforms; robot–human-based platforms; and the use of VR, augmented reality, mixed reality, and extended reality. The third aspect involves hospitals and rehabilitation centers conducting diagnosis, measurement, evaluation, and prognosis, and using and linking such rehabilitation data. The fourth aspect is telehealth, which actively uses IoT-based systems, wearable sensors, and monitoring and communication. The fifth aspect involves home and community, using individual rehabilitation, and actively using information and rehabilitation between caregivers and care receivers. The sixth aspect focuses on disease and disability, actively using rehabilitation and data from children to the older adults, and targeting neurological and musculoskeletal disorders. All these processes, interconnected and used, ultimately lead to the development of a strategy for individualized rehabilitation.

Figure 3. Rehabilomics strategies enabled by cloud-based rehabilitation. AR: augmented reality; MR: mixed reality; VR: virtual reality; XR: extended reality.

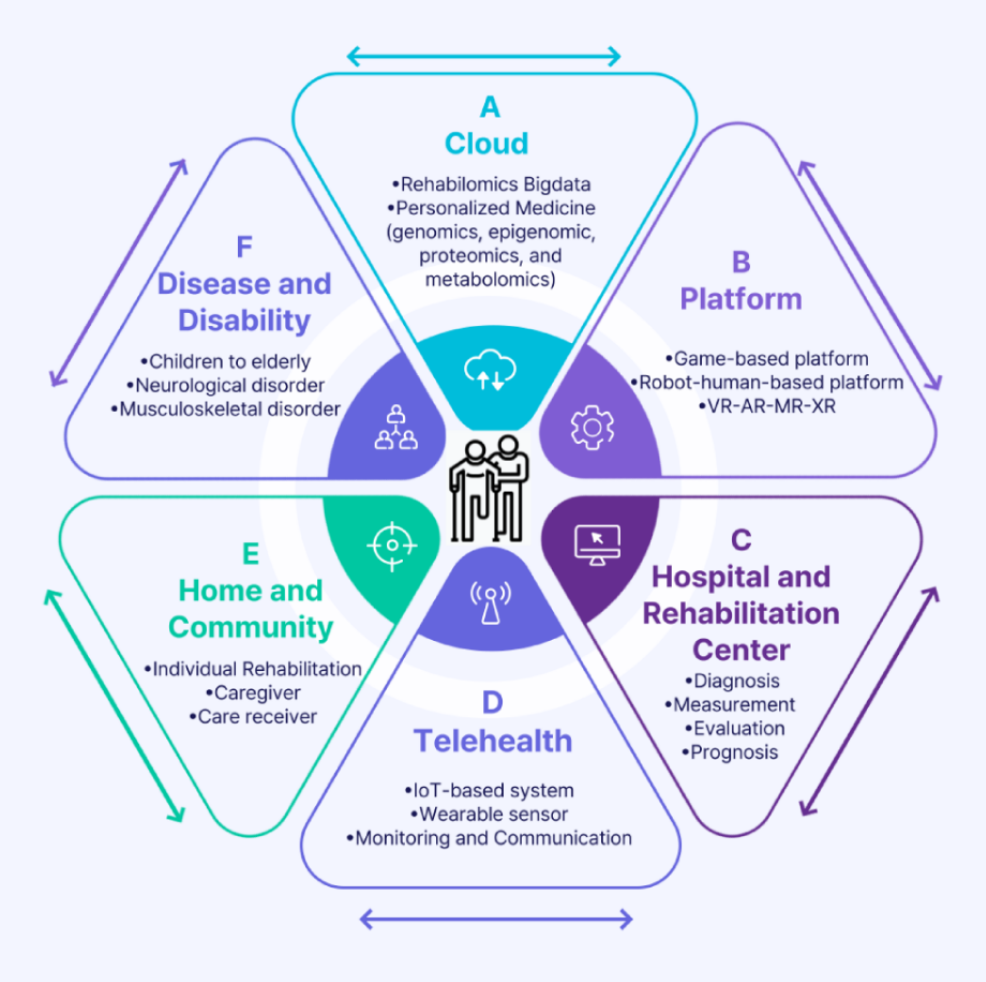

Discussion

Principal Findings

Future Prospects of Cloud-Based Rehabilitation

The findings of this scoping review provide valuable insights into the prospects of cloud-based rehabilitation. Cloud computing technology offers innovative opportunities in the health care domain [44,54] and is increasingly being leveraged as part of rehabilomics strategies. This integrative approach has the potential to open new horizons in rehabilitation therapy and enhance patient-centered personalized treatment. Particularly, telehealth [55] and telemedicine rehabilitation [56,57] offer the advantage of providing medical support and monitoring through connectivity, especially for patients with chronic conditions, such as PD [16,17], stroke [42], and traumatic brain injury [11,12,58]. Daily routines and rehabilitation data are aggregated in the cloud, enabling individuals to review their personal data [49]. With their consent, patients can share these data with local community hospitals when seeking medical consultation or treatment, allowing health care professionals to assess their rehabilitation patterns at home [48].

Cloud-based rehabilitation related to the large data of rehabilomics and personalized medicine is needed [10,59]. This can be achieved through construction by a software company with cloud-related technology or using an existing cloud. Identifying aspects of cloud-based rehabilitation platforms to

XSL•FO
RenderX

integratively manage patients' exercise data, symptom reporting, and progress tracking will be helpful [60].

For example, cloud-based rehabilitation based on rehabilomics provides acute and subacute hospital treatment for patients with neurological damage, such as stroke and rehabilitation at home in the chronic phase [33,34,42,45,61,62].

Cloud-based rehabilitation based on rehabilomics continuously records cloud-based medical devices, rehabilitation exercises, and medications and visits the hospital during midterm evaluations to check records with rehabilitation experts [15,28,59]. This may help establish future intervention directions. In addition, combining web-based systems, such as telehealth, [63,64] with cloud-based rehabilitation, hospitals, and clinics allows intervention or evaluation to be performed in real time by linking with the medical system and is beneficial to both patients and caregivers [65,66]. Carrying out steady and sustainable rehabilitation is motivating, and it is cloud-based. A significant advantage exists in sharing and viewing various data in real time.

Hospitals can use these individualized data to collaborate, diagnose, and treat patients. Through the analysis of aggregated data among patient groups, they can offer research insights and develop tailored rehabilitation prescriptions. This facilitates the provision of personalized rehabilitation programs [48]. Particularly, it seems necessary to upgrade to an extensive and secure personal data system [18].

Security and Privacy Concerns

One of the prominent challenges associated with cloud-based rehabilitation is data security [67] and privacy. Ensuring confidentiality and integrity of patient data is of paramount importance and requires further research and policy development. Establishing stringent data security and privacy protocols between cloud service providers and health care institutions is essential, and patients must be confident that their data will be handled securely. Specifically, the protection of big data based on health care delivery and chronic disease management is necessary, particularly concerning systems integrating personal medical and rehabilitation information, such as application [68].

A cloud-based rehabilitation system needs to be integrated with electronic medical records and personal health records, based on individual consent for personal medical information and information protection policies founded on international standards, such as International Organization for Standardization and International Electrotechnical Commission 27018 for secure cloud services [69-71].

Enhancing Interdisciplinary Collaboration

The successful implementation of rehabilomics strategies underscores the importance of enhanced interdisciplinary collaboration [10,15]. Close cooperation between rehabilitation specialists, geneticists, data scientists, and health care information technology experts is expected to contribute significantly to the advancement of cloud-based rehabilitation and personalized therapy. Opportunities to foster dialogue and collaboration among researchers, clinical practitioners, and health care administrators should also be explored.

Future Research Directions

The consideration of future research directions in the field of cloud-based rehabilitation and rehabilomics strategies is crucial. Further research should prioritize the safety and efficiency of cloud services [67,72] based on blockchain technology [23,73-75], with a focus on data sharing and the development of standardized protocols [76,77]. More extensive clinical studies and evaluations are needed to validate the effectiveness of individualized rehabilitation programs and identify best practices for improving patient outcomes.

Strengths of the Study

This study holds significant value for charting the development trajectory of cloud-based rehabilitation in the context of rehabilomics strategies. This study provides important insights by systematically analyzing the key factors for realizing personalized rehabilitation and emphasizing the importance of data security and interdisciplinary collaboration. Future studies should focus on developing standardized data-sharing protocols and rigorously evaluate the clinical efficacy of these approaches. Enhanced patient experience and treatment outcomes will serve as a driving force for advancing cloud-based rehabilitation. The findings of this study offer valuable information and directions for relevant researchers and clinical practitioners.

Limitations

First, although the detailed variables and algorithms of each rehabilomics model were not known in the review process, various interventions and rehabilitation strategies according to each disease, body type, constitution, and gene are needed based on the models and strategies presented in this study. Second, this study did not involve a scoping review conducted by a librarian. However, the author received training on review search strategies from the affiliated library and proceeded with the research.

Conclusions

This scoping review sheds light on the evolving landscape of cloud-based rehabilitation in the context of rehabilomics strategies. The integration of cloud computing technologies holds immense promise in revolutionizing rehabilitation practices and advancing personalized therapy. However, critical challenges must be addressed, particularly in the domains of data security and interdisciplinary collaboration.

Secure handling of patient data in cloud-based rehabilitation remains a paramount concern. As the field continues to expand, the development and implementation of robust data security and privacy measures are becoming increasingly nonnegotiable. This involves establishing stringent protocols, fostering transparency, and building trust among patients regarding data protection.

The success of rehabilomics strategies relies heavily on effective interdisciplinary collaboration. Collaboration between rehabilitation specialists, geneticists, data scientists, and health care IT experts should be encouraged. Bridging the knowledge

XSL•FO
RenderX

gaps between these diverse fields is crucial for achieving comprehensive patient-centered care.

Future studies should prioritize the refinement of cloud services, emphasize data integrity, and develop standardized protocols for data sharing. Moreover, it is essential to conduct robust clinical studies to evaluate the efficacy of personalized rehabilitation interventions. These efforts will not only provide evidence of improved patient outcomes but also guide the implementation of best practices in clinical settings.

When harnessed effectively and responsibly, cloud-based rehabilitation has the potential to transform the field and usher in a new era of patient-centric data-driven therapy. As researchers and health care practitioners, it is our collective responsibility to navigate the challenges, collaborate across disciplines, and advance this promising frontier for better patient care.

Acknowledgments

The authors would like to thank Editage for the technical editing of the manuscript. This research was supported by a grant (grant RS-2022-KH129263) from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Authors' Contributions

SO participated in conceptualization, data curation, formal analysis, investigation, methodology, resources, visualization, writing the original draft, reviewing, and editing. SL participated in conceptualization, formal analysis, methodology, project administration, supervision, validation, writing, reviewing, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 107 KB - jmir_v27i1e54790_app1.docx]

Multimedia Appendix 2
Studies of cloud-based rehabilitation (N=28).
[DOCX File , 39 KB - jmir_v27i1e54790_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File (Adobe PDF File), 4085 KB - jmir_v27i1e54790_app3.pdf]

References

1. Shrestha B, Basnet BB, Adhikari G. A questionnaire study on the impact on oral health-related quality of life by conventional rehabilitation of edentulous patient. BDJ Open 2020 Jan 30;6(1):3 [FREE Full text] [doi: 10.1038/s41405-020-0029-5] [Medline: 32577306]
2. Seves BL, Hoekstra F, Hettinga FJ, Dekker R, van der Woude LH, Hoekstra T. Trajectories of health-related quality of life among people with a physical disability and/or chronic disease during and after rehabilitation: a longitudinal cohort study. Qual Life Res 2021 Jan 28;30(1):67-80 [FREE Full text] [doi: 10.1007/s11136-020-02647-7] [Medline: 32986126]
3. Sun T, Guo L, Tian F, Dai T, Xing X, Zhao J, et al. Rehabilitation of patients with COVID-19. Expert Rev Respir Med 2020 Dec 12;14(12):1249-1256. [doi: 10.1080/17476348.2020.1811687] [Medline: 32799694]
4. Ambrosio F, Kleim JA. Regenerative rehabilitation and genomics: frontiers in clinical practice. Phys Ther 2016 Apr;96(4):430-432. [doi: 10.2522/ptj.2016.96.4.430] [Medline: 27037291]
5. Petrie MA, Taylor EB, Suneja M, Shields RK. Genomic and epigenomic evaluation of electrically induced exercise in people with spinal cord injury: application to precision rehabilitation. Phys Ther 2022 Jan 01;102(1):pzab243 [FREE Full text] [doi: 10.1093/ptj/pzab243] [Medline: 34718779]

XSL•FO
RenderX

6. Catitti G, Cufaro MC, De Bellis D, Cicalini I, Vespa S, Tonelli F, et al. Extracellular vesicles in regenerative processes associated with muscle injury recovery of professional athletes undergoing sub maximal strength rehabilitation. Int J Mol Sci 2022 Nov 29;23(23):14913 [FREE Full text] [doi: 10.3390/ijms232314913] [Medline: 36499243]
7. Melvin SD, March DT, Marshall K, Carroll AR, van de Merwe JP. Improving rehabilitation outcomes using metabolomics: health, recovery and biomarkers of mortality in sick and injured green turtles (Chelonia mydas). Biol Conserv 2021 Feb;254:108943. [doi: 10.1016/j.biocon.2020.108943]
8. Zhan Q. The implication of precision medicine to nursing science. Interdiscip Nurs Res 2023;2(1):1-3 [FREE Full text] [doi: 10.1097/NR9.0000000000000017]
9. Koen N, Du Preez I, Loots DT. Metabolomics and personalized medicine. Adv Protein Chem Struct Biol 2016;102:53-78. [doi: 10.1016/bs.apcsb.2015.09.003] [Medline: 26827602]
10. Cao W, Zhang X, Qiu H. Rehabilomics: a state-of-the-art review of framework, application, and future considerations. Front Neurol 2023 Mar 8;14:1103349 [FREE Full text] [doi: 10.3389/fneur.2023.1103349] [Medline: 36970504]
11. Wagner AK, Kumar RG. TBI rehabilomics research: conceptualizing a humoral triad for designing effective rehabilitation interventions. Neuropharmacology 2019 Feb;145(Pt B):133-144 [FREE Full text] [doi: 10.1016/j.neuropharm.2018.09.011] [Medline: 30222984]
12. Wagner AK. TBI rehabilomics research: an exemplar of a biomarker-based approach to precision care for populations with disability. Curr Neurol Neurosci Rep 2017 Sep 19;17(11):84. [doi: 10.1007/s11910-017-0791-5] [Medline: 28929311]
13. Wagner AK, Sowa G. Rehabilomics research: a model for translational rehabilitation and comparative effectiveness rehabilitation research. Am J Phys Med Rehabil 2014 Oct;93(10):913-916. [doi: 10.1097/PHM.0000000000000114] [Medline: 24901762]
14. French MA, Roemmich RT, Daley K, Beier M, Penttinen S, Raghavan P, et al. Precision rehabilitation: optimizing function, adding value to health care. Arch Phys Med Rehabil 2022 Jun;103(6):1233-1239. [doi: 10.1016/j.apmr.2022.01.154] [Medline: 35181267]
15. Wagner AK. A rehabilomics framework for personalized and translational rehabilitation research and care for individuals with disabilities: perspectives and considerations for spinal cord injury. J Spinal Cord Med 2014 Jul 16;37(5):493-502. [doi: 10.1179/2045772314y.0000000248]
16. Pan D, Dhall R, Lieberman A, Petitti DB. A mobile cloud-based Parkinson's disease assessment system for home-based monitoring. JMIR Mhealth Uhealth 2015 Mar 26;3(1):e29 [FREE Full text] [doi: 10.2196/mhealth.3956] [Medline: 25830687]
17. Mamun KA, Alhussein M, Sailunaz K, Islam MS. Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications. Future Gener Comput Syst 2017 Jan;66:36-47. [doi: 10.1016/j.future.2015.11.010]
18. Anciaux N, Bonnet P, Bouganim L, Nguyen B, Pucheral P, Sandu Popa I, et al. Personal data management systems: the security and functionality standpoint. Inf Syst 2019 Feb;80:13-35. [doi: 10.1016/j.is.2018.09.002]
19. Youssef A. A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments. Int J Ambient Syst Appl 2014 Jun 30;2(2):1-11. [doi: 10.5121/ijasa.2014.2201]
20. Aceto G, Persico V, Pescapé A. Industry 4.0 and health: internet of things, big data, and cloud computing for healthcare 4.0. J Ind Inf Integr 2020 Jun;18:100129. [doi: 10.1016/j.jii.2020.100129]
21. Goli-Malekabadi Z, Sargolzaei-Javan M, Akbari MK. An effective model for store and retrieve big health data in cloud computing. Comput Methods Programs Biomed 2016 Aug;132:75-82. [doi: 10.1016/j.cmpb.2016.04.016] [Medline: 27282229]
22. Omar AA, Bhuiyan MZ, Basu A, Kiyomoto S, Rahman MS. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Gener Comput Syst 2019 Jun;95:511-521. [doi: 10.1016/j.future.2018.12.044]
23. Nguyen DC, Pathirana PN, Ding M, Seneviratne A. Blockchain for secure EHRs sharing of mobile cloud based e-health systems. IEEE Access 2019;7:66792-66806. [doi: 10.1109/access.2019.2917555]
24. Rajabion L, Shaltooki AA, Taghikhah M, Ghasemi A, Badfar A. Healthcare big data processing mechanisms: the role of cloud computing. Int J Inf Manag 2019 Dec;49:271-289. [doi: 10.1016/j.ijinfomgt.2019.05.017]
25. Bonnechère B. Integrating rehabilomics into the multi-omics approach in the management of multiple sclerosis: the way for precision medicine? Genes (Basel) 2022 Dec 24;14(1):63 [FREE Full text] [doi: 10.3390/genes14010063] [Medline: 36672802]
26. Elliott F, Errington TM, Bagby C, Frazier M, Geiger BJ, Liu Y, et al. An introduction to the Open Science Framework. Open Science Framework Home. 2012 May 31. URL: https://osf.io/4znzp/ [accessed 2024-12-26]
27. Oh S, Lee S. Rehabilomics strategies enabled by cloud-based rehabilitation: a scoping review. JMIR Preprints Preprint posted online on November 21, 2023 [FREE Full text] [doi: 10.2196/54790]
28. Fardoun HM, Altalhi AH, Cipres AP, Castillo JR, Albiol-Pérez S. CRehab: a cloud-based framework for the management of rehabilitation processes. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. 2013 Presented at: PervasiveHealth 2013; May 5-8, 2013; Venice, Italy. [doi: 10.4108/icst.pervasivehealth.2013.252252]
29. Fortino G, Gravina R. Rehab-aaService: a cloud-based motor rehabilitation digital assistant. In: Proceedings of the REHAB 2014. 2014 Presented at: REHAB 2014; May 20, 2014; Oldenburg, Germany URL: https://doi.org/10.4108/icst.pervasivehealth [doi: 10.4108/icst.pervasivehealth.2014.255273]

XSL•FO
RenderX

30. Hoda M, Dong H, Ahmed D, El Saddik A. Cloud-based rehabilitation exergames system. In: Proceedings of the IEEE International Conference on Multimedia and Expo Workshops. 2014 Presented at: ICMEW 2014; July 14-18, 2014; Chengdu, China. [doi: 10.1109/icmew.2014.6890660]
31. Li S, Pathirana PN. Cloud-based non-invasive tele-rehabilitation exercise monitoring. In: Proceedings of the IEEE Conference on Biomedical Engineering and Sciences. 2014 Presented at: IECBES 2014; December 8-10, 2014; Kuala Lumpur, Malaysia. [doi: 10.1109/iecbes.2014.7047527]
32. Škraba A, Stojanović R, Zupan A, Kološvari A, Kofjač D. Speech-controlled cloud-based wheelchair platform for disabled persons. Microprocess Microsyst 2015 Nov;39(8):819-828. [doi: 10.1016/j.micpro.2015.10.004]
33. Hoda M, Hoda Y, Hage A, Alelaiwi A, El Saddik A. Cloud-based rehabilitation and recovery prediction system for stroke patients. Cluster Comput 2015 Apr 28;18(2):803-815. [doi: 10.1007/s10586-015-0448-6]
34. Woodman S, Hiden H, Turner M, Dowsland S, Watson P. Monitoring of upper limb rehabilitation and recovery after stroke: an architecture for a cloud-based therapy platform. In: Proceedings of the IEEE 11th International Conference on e-Science. 2015 Presented at: eScience 2015; August 31-September 4, 2015; Munich, Germany. [doi: 10.1109/escience.2015.29]
35. Khalifeh AF, Saleh A, AL-Nuimat M, Tair DA. An open source cloud based platform for elderly health monitoring and fall detection. In: Proceedings of the 4th Workshop on ICTs for Improving Patients Rehabilitation Research Techniques. 2016 Presented at: REHAB '16; October 13-4, 2016; Lisbon, Portugal URL: https://doi.org/10.1145/3051488.3051513 [doi: 10.1145/3051488.3051513]
36. Causo A, Liu L, Krishnasamy G, Yeo SH, Chen IM. Integration of wireless wearable sensors and mobile computing with cloud-based service for patient rehabilitation monitoring. In: Proceedings of the International Conference on Intelligent Unmanned Systems 2012. 2012 Presented at: ICIUS 2012; October 22-24, 2012; Singapore, Singapore.
37. Bilic D, Uzunovic T, Golubovic E, Ustundag BC. Internet of things-based system for physical rehabilitation monitoring. In: Proceedings of the XXVI International Conference on Information, Communication and Automation Technologies. 2017 Presented at: ICAT 2017; October 26-28, 2017; Sarajevo, Bosnia and Herzegovina. [doi: 10.1109/icat.2017.8171641]
38. Cârstoiu D, Oltean VE, Nica SM, Spiridon G. A cloud-based architecture proposal for rehabilitation of aphasia patients. Rev Roum Sci Technol 2017;62(3):332-337 [FREE Full text]
39. Yee J, Low CY, Hanapiah FA, Zakaria NA, Mohammad U, bin Abd Rahman R. System-level design of a cloud-based training device for upper limb spasticity rehabilitation. In: Proceedings of the 2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation. 2017 Presented at: ROMA 2017; September 19-21, 2017; Kuala Lumpur, Malaysia. [doi: 10.1109/roma.2017.8231732]
40. Inamura T, Unenaka S, Shibuya S, Ohki Y, Oouchida Y, Izumi SI. Development of VR platform for cloud-based neurorehabilitation and its application to research on sense of agency and ownership. Adv Robot 2016 Dec 22;31(1-2):97-106 [FREE Full text] [doi: 10.1080/01691864.2016.1264885]
41. Cheng IJ, Cheng PH, Tsai SY, Chen PH, Lai YJ, Yang FC, et al. A comparison of Python and Java capabilities on somatosensory game for cloud-based rehabilitation system. In: Proceedings of the IEEE 4th Global Conference on Consumer Electronics. 2015 Presented at: GCCE 2015; October 27-30, 2015; Osaka, Japan. [doi: 10.1109/gcce.2015.7398726]
42. Hossain MS, Hoda M, Muhammad G, Almogren A, Alamri A. Cloud-supported framework for patients in post-stroke disability rehabilitation. Telemat Inform 2018 Jul;35(4):826-836. [doi: 10.1016/j.tele.2017.12.001]
43. Borghese NA, Pezzera M, Mainetti R, Essenziale J, Cazzaniga R, Reggiori B, et al. A cloud-based platform for effective supervision of autonomous home rehabilitation through exer-games. In: Proceedings of the IEEE 6th International Conference on Serious Games and Applications for Health. 2018 Presented at: SeGAH 2018; May 16-18, 2018; Vienna, Austria. [doi: 10.1109/segah.2018.8401383]
44. Pham M, Mengistu Y, Do H, Sheng W. Delivering home healthcare through a Cloud-based Smart Home Environment (CoSHE). Future Gener Comput Syst 2018 Apr;81:129-140. [doi: 10.1016/j.future.2017.10.040]
45. Halloran S, Shi JQ, Guan Y, Chen X, Dunne-Willows M, Eyre J. Remote cloud-based automated stroke rehabilitation assessment using wearables. In: Proceedings of the IEEE 14th International Conference on e-Science. 2018 Presented at: e-Science 2018; October 29-November 1, 2018; Amsterdam, The Netherlands. [doi: 10.1109/escience.2018.00063]
46. Jiang Z, Dai W, Wang W, Wang W. A cloud-based training and evaluation system for facial paralysis rehabilitation. In: Proceedings of the IEEE 16th International Conference on Industrial Informatics. 2018 Presented at: INDIN 2018; July 18-20, 2018; Porto, Portugal. [doi: 10.1109/indin.2018.8471934]
47. Sciuto GL, Samuele R, Napoli C. A cloud-based flexible solution for psychometric tests validation, administration and evaluation. In: Proceedings of the Symposium for Young Scientists in Technology, Engineering and Mathematics. 2019 Presented at: SYSTEM 2019; May 30, 2019; Gliwice, Poland.
48. Wei W, Lu Y, Rhoden E, Dey S. User performance evaluation and real-time guidance in cloud-based physical therapy monitoring and guidance system. Multimed Tools Appl 2017 Nov 6;78(7):9051-9081. [doi: 10.1007/s11042-017-5278-5]
49. Celesti A, Lay-Ekuakille A, Wan J, Fazio M, Celesti F, Romano A, et al. Information management in IoT cloud-based tele-rehabilitation as a service for smart cities: comparison of NoSQL approaches. Measurement 2020 Feb;151:107218. [doi: 10.1016/j.measurement.2019.107218]
50. Inamura T, Mizuchi Y. SIGVerse: a cloud-based VR platform for research on multimodal human-robot interaction. Front Robot AI 2021 May 31;8:549360 [FREE Full text] [doi: 10.3389/frobt.2021.549360] [Medline: 34136534]

XSL•FO
RenderX

51. Yi A, Chen Z, Ling W, Yin X, Li Y, Yan J, et al. Effectiveness of cloud-based rehabilitation in children with developmental language disorder during the COVID-19 pandemic: a prospective cohort study. Medicine (Baltimore) 2022 Aug 19;101(33):e30056 [FREE Full text] [doi: 10.1097/MD.0000000000030056] [Medline: 35984120]
52. Wu X, Zhou J, Zheng M, Chen S, Wang D, Anajemba J, et al. Cloud-based deep learning-assisted system for diagnosis of sports injuries. J Cloud Comp 2022 Nov 23;11(1):82 [FREE Full text] [doi: 10.1186/s13677-022-00355-w]
53. Lv Z, Singh AK. Edge-cloud-based wearable computing for automation empowered virtual rehabilitation. IEEE Trans Automat Sci Eng 2024 Jul;21(3):3896-3909. [doi: 10.1109/tase.2023.3289908]
54. Gu D, Yang X, Deng S, Liang C, Wang X, Wu J, et al. Tracking knowledge evolution in cloud health care research: knowledge map and common word analysis. J Med Internet Res 2020 Feb 25;22(2):e15142 [FREE Full text] [doi: 10.2196/15142] [Medline: 32130115]
55. Douglass H, Lowman J, Causey-Upton R. Clinician perspectives of telehealth pre–COVID-19: a systematic review and qualitative metasynthesis. Perspect ASHA SIGs 2023 Apr 05;8(2):396-411. [doi: 10.1044/2022_PERSP-22-00074]
56. Lal H, Mohanta S, Kumar J, Patralekh MK, Lall L, Katariya H, et al. Telemedicine-rehabilitation and virtual reality in orthopaedics and sports medicine. Indian J Orthop 2023 Jan 22;57(1):7-19 [FREE Full text] [doi: 10.1007/s43465-022-00766-6] [Medline: 36660485]
57. Zhang L, Maitinuer A, Lian Z, Li Y, Ding W, Wang W, et al. Home based pulmonary tele-rehabilitation under telemedicine system for COPD: a cohort study. BMC Pulm Med 2022 Jul 24;22(1):284 [FREE Full text] [doi: 10.1186/s12890-022-02077-w] [Medline: 35871652]
58. Kobeissy FH, Guingab-Cagmat JD, Razafsha M, O'Steen L, Zhang Z, Hayes RL, et al. Leveraging biomarker platforms and systems biology for rehabilomics and biologics effectiveness research. PM R 2011 Jun;3(6 Suppl 1):S139-S147. [doi: 10.1016/j.pmrj.2011.02.012] [Medline: 21703572]
59. Kasoju N, Remya NS, Sasi R, Sujesh S, Soman B, Kesavadas C, et al. Digital health: trends, opportunities and challenges in medical devices, pharma and bio-technology. CSIT 2023 Apr 11;11(1):11-30 [FREE Full text] [doi: 10.1007/s40012-023-00380-3]
60. Brocal F, Sebastián MA, González C. Chapter 2 - Advanced manufacturing processes and technologies. In: Roig B, Roig K, Thireau V, editors. Management of Emerging Public Health Issues and Risks: Multidisciplinary Approaches to the Changing Environment. New York, NY: Academic Press; 2019:31-64.
61. Arntz A, Weber F, Handgraaf M, Lällä K, Korniloff K, Murtonen KP, et al. Technologies in home-based digital rehabilitation: scoping review. JMIR Rehabil Assist Technol 2023 Jul 27;10:e43615 [FREE Full text] [doi: 10.2196/43615] [Medline: 37253381]
62. Epalte K, Grjadovojs A, Bērziņa G. Use of the digital assistant Vigo in the home environment for stroke recovery: focus group discussion with specialists working in neurorehabilitation. JMIR Rehabil Assist Technol 2023 Apr 14;10:e44285 [FREE Full text] [doi: 10.2196/44285] [Medline: 37058334]
63. Erdt M, Yusof SB, Chai L, Md Salleh SU, Liu Z, Sarim HB, et al. Characterization of telecare conversations on lifestyle management and their relation to health care utilization for patients with heart failure: mixed methods study. J Med Internet Res 2024 Oct 30;26:e46983 [FREE Full text] [doi: 10.2196/46983] [Medline: 39476370]
64. Waranski M, Garbsch R, Kotewitsch M, Teschler M, Schmitz B, Mooren FC. A behavioral change-based mobile intervention for promoting regular physical activity in medical rehabilitation maintenance of patients with coronary artery disease: controlled trial. J Med Internet Res 2024 Oct 08;26:e56480 [FREE Full text] [doi: 10.2196/56480] [Medline: 39378432]
65. Pak SS, Janela D, Freitas N, Costa F, Moulder R, Molinos M, et al. Comparing digital to conventional physical therapy for chronic shoulder pain: randomized controlled trial. J Med Internet Res 2023 Aug 18;25:e49236 [FREE Full text] [doi: 10.2196/49236] [Medline: 37490337]
66. Wang H, Zhao Y, Yu L, Liu J, Zwetsloot IM, Cabrera J, et al. A personalized health monitoring system for community-dwelling elderly people in Hong Kong: design, implementation, and evaluation study. J Med Internet Res 2020 Sep 30;22(9):e19223 [FREE Full text] [doi: 10.2196/19223] [Medline: 32996887]
67. Hariharasitaraman S, Balakannan SP. A dynamic data security mechanism based on position aware Merkle tree for health rehabilitation services over cloud. J Ambient Intell Human Comput 2019 Jul 27. [doi: 10.1007/s12652-019-01412-0]
68. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
69. de Hert P, Papakonstantinou V, Kamara I. The cloud computing standard ISO/IEC 27018 through the lens of the EU legislation on data protection. Comput Law Secur Rev 2016 Feb;32(1):16-30. [doi: 10.1016/j.clsr.2015.12.005]
70. Lalanne V, Munier M, Gabillon A. Information security risk management in a world of services. In: Proceedings of the International Conference on Social Computing. 2013 Presented at: SocialCom 2013; September 8-14, 2013; Alexandria, VA. [doi: 10.1109/socialcom.2013.88]
71. Arafat M. Information security management system challenges within a cloud computing environment. In: Proceedings of the 2nd International Conference on Future Networks and Distributed Systems. 2018 Presented at: ICFNDS '18; June 26-27, 2018; Amman, Jordan URL: https://doi.org/10.1145/3231053.3231127 [doi: 10.1145/3231053.3231127]
72. Ermakova T, Fabian B, Kornacka M, Thiebes S, Sunyaev A. Security and privacy requirements for cloud computing in healthcare. ACM Trans Manage Inf Syst 2020 May 03;11(2):1-29 [FREE Full text] [doi: 10.1145/3386160]

XSL•FO
RenderX

73. Lakhan A, Mohammed MA, Nedoma J, Martinek R, Tiwari P, Kumar N. Blockchain-enabled cybersecurity efficient IIOHT cyber-physical system for medical applications. IEEE Trans Netw Sci Eng 2023 Sep 1;10(5):2466-2479. [doi: 10.1109/tnse.2022.3213651]
74. Lakhan A, Mohammed MA, Nedoma J, Martinek R, Tiwari P, Kumar N. DRLBTS: deep reinforcement learning-aware blockchain-based healthcare system. Sci Rep 2023 Mar 13;13(1):4124 [FREE Full text] [doi: 10.1038/s41598-023-29170-2] [Medline: 36914679]
75. Si H, Sun C, Li Y, Qiao H, Shi L. IoT information sharing security mechanism based on blockchain technology. Future Gener Comput Syst 2019 Dec;101:1028-1040. [doi: 10.1016/j.future.2019.07.036]
76. Chen Y. IoT, cloud, big data and AI in interdisciplinary domains. Simul Model Pract Theory 2020 Jul;102:102070 [FREE Full text] [doi: 10.1016/j.simpat.2020.102070]
77. Qiu H, Qiu M, Liu M, Memmi G. Secure health data sharing for medical cyber-physical systems for the healthcare 4.0. IEEE J Biomed Health Inform 2020 Sep;24(9):2499-2505. [doi: 10.1109/jbhi.2020.2973467]

Abbreviations

CoSHE: cloud-based smart home environment
IoT: Internet of Things
PD: Parkinson disease
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension For Scoping Reviews
VR: virtual reality

Edited by A Coristine; submitted 21.11.23; peer-reviewed by HY Shi, Y Chu, P Talero, KL Lin; comments to author 15.05.24; revised version received 06.09.24; accepted 20.11.24; published 28.01.25.

Please cite as:
Oh S, Lee S
Rehabilomics Strategies Enabled by Cloud-Based Rehabilitation: Scoping Review
J Med Internet Res 2025;27:e54790
URL: https://www.jmir.org/2025/1/e54790
doi:10.2196/54790
PMID:

©Sejun Oh, SangHeon Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Implementing Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety in Adults: Systematic Review

Daniel Duffy[1], PhD; Derek Richards[1], PhD; Garrett Hisler[1], PhD; Ladislav Timulak[2], PhD

[1]Amwell Science, Amwell, Boston, MA, United States

[2]Trinity College Dublin, Dublin, Ireland

Corresponding Author:
Daniel Duffy, PhD
Amwell Science, Amwell
75th State St., 26th Floor
Boston, MA, 02109
United States
Phone: 1 617 204 3500
Email: Daniel.Duffy@amwell.com

Abstract

Background: Scientific implementation findings relevant to the implementation of internet-delivered cognitive behavioral therapy (iCBT) for depression and anxiety in adults remain sparse and scattered across different sources of published information. Identifying evidence-based factors that influence the implementation of iCBT is key to successfully using iCBT in real-world clinical settings.

Objective: This systematic review evaluated the following: (1) aspects that research articles postulate as important for the implementation of iCBT and (2) aspects relevant to the day-to-day running of iCBT services. A mixed methods systematic review using a convergent synthesis design was conducted to bring together evidence across this sparse literature consisting of divergent scientific article types to investigate the implementation of iCBT for depression and anxiety in adults.

Methods: We searched the PsycINFO, PsycArticles, MEDLINE, CINAHL Complete, and Embase databases for any published peer-reviewed scientific articles that report on the implementation of iCBT for depression or anxiety disorders in adults. A total of 40 articles spanning the case study, commentary, meta-analysis, mixed methods study, pilot randomized controlled trial, randomized controlled trial, qualitative study, quantitative study, review, and systematic review article types were identified as eligible for this mixed methods review. Data were analyzed qualitatively using the descriptive-interpretive approach.

Results: The first domain highlighted the impact of therapist and patient attitudes when implementing iCBT, the superiority of guided iCBT over unguided iCBT, its noninferiority to equivalent face-to-face treatments, and its utility outside of the original target of mild-to-moderate depression and anxiety. In total, 3 subdomains were identified under the second domain: (1) the management of iCBT in the workplace, detailing the importance of managing the iCBT service, related staff, and their motivations for using it; (2) the practice of iCBT in the workplace, describing the therapeutic aspects of iCBT provision, such as the provision of support, the background of supporters, and screening procedures; and (3) contextual considerations, detailing the impact of governmental legislation on therapy conducted over the internet, the lack of an iCBT workforce as a limiting factor, and the cost estimates associated with iCBT provision.

Conclusions: Broadly, the findings describe several aspects that should be taken into account when researchers or practitioners implement iCBT as part of their work. However, the findings should be interpreted with caution, as the articles reviewed spanned many article types, and few of the included studies were directly focused on evaluating the implementation of iCBT. While findings provide insight into important factors to consider during iCBT implementation, these findings and their limitations highlight the need for more implementation-specific research in this area.

(J Med Internet Res 2025;27:e47927) doi:10.2196/47927

KEYWORDS

mixed methods systematic review; internet-delivered cognitive behavioral therapy; iCBT; implementation science; implementation research; depression; anxiety

XSL•FO
RenderX

Introduction

Background

Internet-delivered cognitive behavioral therapy (iCBT) for depression and anxiety has been developed to help increase access to evidence-based therapies. There is empirical support for its use in treating depression and anxiety [1-4]. End users experience it positively [5] and find it to be satisfactory and acceptable [1,6,7]. However, disseminating iCBT at scale remains a challenge [8,9], and COVID-19 has brought its relevance to light now more so than ever [10-12]. A 2019 commentary [13] discussed the evidence-to-practice gap in digital mental health treatments. The authors postulated that the reason for this gap is a lack of knowledge in the field of iCBT around implementing interventions within routine care. They suggested the adoption of implementation science methodologies to bridge this evidence-to-practice gap.

Implementation science has been defined as the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice to improve the quality and effectiveness of health services and care [14]. Central to this definition is the problem statement behind it: it takes almost 17 years for health care research to achieve its intended benefit, which is termed as the "evidence-to-practice gap" [15,16]. As a newly emerging academic field, implementation science is largely integrative; it borrows and adapts theories from multiple fields and uses these to understand the determinant mechanisms as to why (or why not) a specific implementation succeeds [17]. Implementation science theories provide a framework that allows for implementation plans to be developed and relevant outcomes to be measured [18], and it has been posited that using these methodologies in future studies of iCBT could generate learnings relevant to its real-world application [13,19,20].

In a recent review focusing on determinants of implementation for eHealth interventions [8], 37 determinants associated with successful implementation were identified. However, it should be noted that "eHealth interventions" in this case contained a wide variety of digitally enabled interventions, including iCBT and psychotherapy delivered via videoconferencing. When comparing iCBT and other eHealth interventions, "complexity" is a factor for consideration, that is, the degree to which an intervention contains multiple components that require interaction from many individuals, from various levels within an organization, to enact the intervention effectively [21]. iCBT's level of complexity is highlighted in service illustration articles by Titov et al [22,23]; for example, therapists' skill set to operate iCBT efficiently (technical knowledge and constructing written messages), revised services delivery pathways, adherence to regulatory frameworks, and newly aligned clinical governance procedures are some elements of how delivering iCBT may differ from more traditional or less complex services. Conversely, although administering psychological therapy through videoconferencing software may require some altering of specific therapeutic skills and technical upskilling [24], relative complexity across other areas may be lower (eg, referral pathways and wider system integration). Similarly, some authors have illustrated the need for both iCBT [25,26] and telehealth-specific competency frameworks [27], further illustrating the need for specialized skills to extend the traditional therapist skill set.

Attempts to mobilize implementation science information on eHealth interventions, generally to the point of having pragmatic, clinical relevance for iCBT, have been sparse [28]. As a consequence, the availability of implementation findings relevant to iCBT remains low. This study is a mixed methods systematic review that aimed to account for literature that specifically references or can inform factors relevant to the implementation of iCBT, specifically for depression and anxiety treatment in adults. The benefit of a mixed methods systematic review over traditional systematic literature reviewing is that it seeks to extract relevant information across qualitative, quantitative, review, and illustration-based articles. Mixed methods synthesis affords a way to effectively capture this information and synthesize it qualitatively to produce relevant insights into the implementation of iCBT. Therefore, a convergent integrated approach to the mixed methods review was chosen due to its appropriateness over other review methods for the subject; a traditional systematic review on the implementation of iCBT for depression and anxiety would not be appropriate due to insufficient qualitative or quantitative findings to generate insights [29]. Furthermore, there are no restrictions imposed on the type of evidence included within the synthesis, which aligns with the anticipated variety of articles that would be identified [30]. The disorder domains of depression and anxiety were chosen due to them being the most substantive areas of research for iCBT.

A mixed methods systematic review departs from and complements previous work in the ways mentioned subsequently [8]. First, it will specifically focus on iCBT-based interventions, which can be considered relatively "complex" in their administration [21]. Second, it will provide a rich description of the current "practice behind the science" by focusing on reportage within the methods, results, and discussion sections of the articles. Third, it will contribute to the existing literature regarding specific implementation strategies that are associated with the use of iCBT [31]. Finally, it will allow the interpretation of research findings in a way that is hoped to be productive for future implementations specific to iCBT for the treatment of depression and anxiety.

Review Objectives

The overarching objective of the review was centered on the pragmatic question of "What can we learn from published peer-reviewed literature about the implementation of iCBT for depression and anxiety?" This objective was further broken down into 2 domains of focus, on which data extraction and subsequent data analysis were based (refer to the Methods section).

The first domain was centered on *implementation insights derived from iCBT research*. This objective and domain center on understanding the novel information that is often presented in published research and how this information can have relevance and be mobilized for the benefit of iCBT implementation.

XSL•FO
RenderX

The second domain was *implementation process considerations for the successful implementation of iCBT*, which consisted of establishing the strategies that are used within the literature to facilitate the implementation of iCBT. According to implementation science literature, implementation strategies are methods used to facilitate the implementation of an intervention, where strategies can consist of training packages, management approaches, developing protocols for intervention use, etc [31,32].

Methods

Design

Because implementation science information on eHealth interventions that has pragmatic, clinical relevance for iCBT has been sparse and scattered across different article types, a mixed methods systematic review was conducted. This mixed methods systematic review used a convergent integrated approach and was conducted to identify literature that was central to the review objective mentioned earlier [30,33,34]. The convergent integrated approach to conducting a mixed methods systematic review consists of "qualitizing" numerical or statistical findings; that is, quantitative findings are extracted and allocated textual descriptions to allow for integration and simultaneous synthesis with other qualitative data. The resulting qualitative data were then analyzed using the descriptive-interpretive approach [35]. This review was not registered, and a review protocol was not prepared.

Search Strategy

The general search strategy used was as follows: ("ICBT" OR "CCBT" OR "internet-delivered CBT" OR "internet-delivered cognitive behavioural therapy" OR "internet-delivered cognitive behaviour therapy" OR "internet-based cognitive behaviour therapy" OR "internet-based cognitive behavioural therapy" OR "internet-administered cognitive behaviour therapy" OR "internet-administered cognitive behavioural therapy" OR "internet-supported cognitive behaviour therapy" OR "internet-supported cognitive behavioural therapy") AND ("Anx*" OR "depress*" OR "low mood" OR "GAD" OR "phobia" OR "SAD") AND "Implement*." Databases searched included PsycINFO, PsycArticles, MEDLINE, CINAHL Complete, and Embase. A full description of terms and derivatives for the different databases is included in Multimedia Appendix 1. Search engine limitations required that the search date had to start from 2007. The search was initially conducted in June 2020 (January 1, 2007-June 1, 2020) and further repeated in September 2021 (June 1, 2020-August 31, 2021) to identify any new or relevant publications. The inclusion and exclusion criteria are given in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Study provides reports on outcomes related broadly to the 2 domains of interest:
  - Implementation insights derived from internet-delivered cognitive behavioral therapy (iCBT) research
  - Implementation process considerations for the successful implementation of iCBT
- The study types included in the review:
  - Empirical research, encompassing pre-post experimental (eg, feasibility or randomized controlled trial), case study, observational, or qualitative designs in naturalistic, nonefficacy settings
  - Review-type studies, including systematic, meta, umbrella, narrative, and scoping reviews
  - Service-illustration articles that report on the effectiveness of iCBT clinics over time or describe their operating model
- Studies targeting adult patient populations, mental health care workers (eg, clinicians, therapists, and service managers), or prospective users of iCBT
- The study must be conducted in reference to iCBT (eg, patients undertaking iCBT, clinicians or therapists, or patients reporting on their views of iCBT).
- The study must be primarily conducted in reference to depression and anxiety disorders (eg, patients undertaking iCBT for depression and anxiety, clinicians and therapists, or patients reporting on their views of iCBT for depression and anxiety)

Exclusion criteria

- Nonpeer reviewed research
- Research not in the English language
- Protocols
- Dissertations (due to the difficulty in identifying and accessing these at a wide scale)
- Book chapters
- Conference presentations and abstracts
- Research with participants aged <18 years
- Studies reporting only on clinical effectiveness data with no information on the implementation of iCBT reported on in the study.

XSL•FO
RenderX

Screening

The screening consisted of two steps: (1) review the title and abstract and (2) review the full-paper. We chose to first review the titles and abstracts of all identified records due to the nature of this review and the wide range of study types that were anticipated to result from the search. For example, it was noted throughout the reviewing process that articles frequently cited the terms "implementation" or "feasibility" in the title but failed to provide any relevant information under these constructs when abstracts were reviewed. Where articles provided inadequate information in their abstracts (eg, "the results inform the feasibility of implementing iCBT within XYZ context") to apply the inclusion criteria, DR acted as a second reviewer for these abstracts and consulted with DD (primary reviewer) to make a decision on inclusion or exclusion. Once step 1 was completed, full texts of all articles were reviewed by DD at full text to discern their relevance to the domains of interest. During this review, articles were rejected during full text screening for the following two reasons: (1) incorrect record specification from the databases (eg, conference presentations being mislabeled) and (2) provided little (eg, minor comments relating to the implementation of iCBT within "future research" sections) or no information on implementation. Once all articles were screened and the final dataset established, data extraction commenced.

Data Analysis

Meaning Unit Extraction

Meaning unit extraction began by identifying qualitative meaning units within the methodology, results, and discussion sections of articles relevant to the study objective, that is, learning about the implementation of iCBT for depression and anxiety. Meaning units are discrete data chunks (either paragraphs or sentences) that contribute stand-alone meaning toward a particular research question or objective [36]. Throughout the mixed methods systematic review process, relevant quantitative findings were translated (or "qualitized") to qualitative meaning units. The resulting qualitative meaning units were identified and subsequently extracted to an Excel (Microsoft Corporation) file for analysis and assigned relevant identifiers. In addition to assigning identifiers, each meaning unit was also assigned brief, textual summary labels that provided a way to quickly identify the information being conveyed by longer meaning units.

To address the review objective of learning about the implementation of iCBT for depression and anxiety, 2 main focus areas were identified that allowed for the structuring of the relevant data to guide meaning unit extraction.

The first focus area was the implementation process considerations for the successful implementation of iCBT. It consisted of the following questions: (1) What strategies do articles report on that are related to the process of implementing iCBT (eg, training clinicians or therapists, screening procedures, referral pathways, and service operations)? (2) Do articles report on the impact of these strategies on specific stakeholder groups (eg, patients and clinicians or therapists)? and (3) Do articles acknowledge or cite factors within the context of the implementation (eg, governmental policy, service infrastructure, and funding)?

The second focus area was the implementation insights derived from iCBT research. It involved the following questions: (1) What implications do authors of the included studies cite as important for the future of the implementation of iCBT? and (2) How do authors interpret their findings in discussion sections of articles, and can these interpretations have implications for how iCBT is implemented?

Category Generation

Meaning units within these 2 focus areas were then further analyzed for similarities and clustered together into categories and subcategories grouping meaning units of similar meaning. The categories and subcategories were named to capture the meaning of the meaning units they contained. The process of organizing and naming categories and subcategories was an ongoing activity led by DD and involving feedback from DR and another colleague. The consensus was agreed upon through group discussion, where the core component of each subcategory was established (eg, this subcategory describes x), and the revised category names were generated based on this shared understanding. As per the descriptive-interpretive approach, it was important that progress was continuously audited through group meetings with DD, DR, JP, and LT.

Results

Overview

A total of 40 eligible articles published between 2010 and 2021 were included in the mixed methods synthesis. Multimedia Appendix 2 provides the types of articles included in this mixed methods systematic review and brief summaries of these included articles. Article types included in this review were case study (1/40, 2%), commentary (4/40, 10%), meta-analysis (5/40, 12%), mixed methods study (3/40, 8%), pilot randomized controlled trial (1/40, 2%), randomized controlled trial (2/40, 5%), qualitative study (5/40, 12%), quantitative study (13/40, 32%), review (3/40, 8%), and systematic review (3/40, 8%). Multimedia Appendix 3 [4,9,22,23,28,37-71] includes a numbered list of all articles analyzed as part of this mixed methods systematic review. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram constructed to illustrate the search findings is illustrated in Figure 1 and the PRISMA checklist is provided in Multimedia Appendix 4.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram for mixed methods systematic review search. DOI: digital object identifier; iCBT: internet-delivered cognitive behavioral therapy.

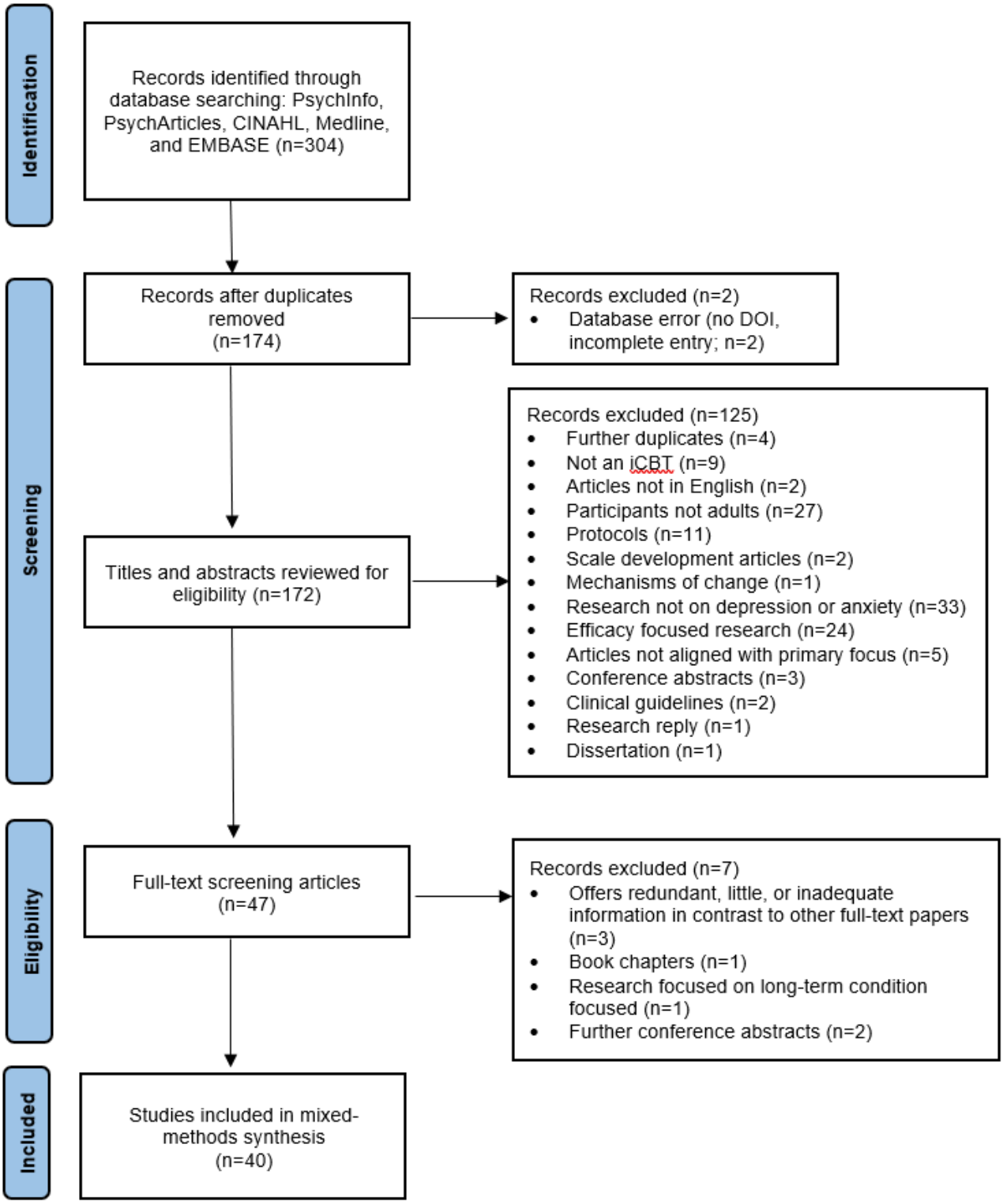

Domain and Category Structure

Two domains (and lower subdomains, categories, and subcategories) were identified: (1) implementation insights derived from iCBT research and (2) implementation processes related to the successful implementation of iCBT in care settings (Tables 1 and 2). Subcategories are discussed as talking points within each category as opposed to adding further organizational levels within this study.

XSL•FO
RenderX

Table 1. Categories and subcategories identified under the domain implementation insights derived from internet-delivered cognitive behavioral therapy (iCBT) research and illustration of the number of contributing articles (N=40).

Category and subcategory	Articles, n (%)	Study reference number
Clinician attitudes toward iCBT		
Negative attitudes toward iCBT can impact referral rates and patient outcome	10 (25)	[4,5,13,23,25-27,29,34,36]
Positive attitudes toward iCBT can increase acceptability and help to grow iCBT in service	3 (8)	[29,31,34]
Patient attitudes toward iCBT		
Positive attitudes toward iCBT content, support, privacy, and convenience of iCBT can foster engagement	10 (25)	[3,8,27,29,32,33,36,72-74]
Attitudes as moderators of clinical outcome, perceived helpfulness, and adherence	3 (8)	[18,36,73]
Negative attitudes relate to preference for face-to-face therapy and issues with the utility of iCBT to patient needs	5 (12)	[25,26,36,72,73]
The delivery of internet-delivered therapies can be helped by technological and clinical augmentation	7 (18)	[11,16,23-25,27,34]
Specific patient characteristics need to be considered when implementing iCBT		
Age is negatively associated with adherence and clinical outcomes in guided iCBT and not associated with symptom deterioration in unguided iCBT	4 (10)	[1,20,36,74]
The relationship between gender or sex and adherence is unclear in iCBT overall, but gender or sex is not associated with symptom deterioration in unguided iCBT	2 (5)	[1,36]
Patient technological literacy is tentatively positively associated with adherence and clinical outcomes in iCBT	2 (5)	[36,75]
Medication and alcohol use are not associated with iCBT adherence	1 (2)	[36]
Racial or ethnic minority group membership is negatively associated with adherence to iCBT	1 (2)	[35]
The relationship between adherence and marital status, employment status, and education level is mixed overall, but they are not associated with symptom deterioration in unguided iCBT	3 (8)	[1,36,74]
Having a lower income is positively associated with dropout	1 (2)	[74]
Comorbidity of disorders can moderate treatment outcome	1 (2)	[74]
Making sudden clinical gains is associated with greater improvements after the treatment	1 (2)	[24]
The severity of depression can positively impact clinical outcomes and adherence	2 (5)	[10,36]
Symptoms of depression can negatively impact iCBT adherence	2 (5)	[12,18]
Chronic mental health problems are negatively associated with iCBT adherence	1 (2)	[74]
Guided iCBT is superior to unguided iCBT in regard to symptom outcomes and adherence	11 (28)	[1,7,17,23-25,27,34,36,73,74]
iCBT is as effective as face-to-face delivery of the same protocol, yet preference is often for face-to-face treatment	13 (32)	[6,11,13,14,16,23-25,27,29,36,72]
iCBT appears to be effective beyond the original target of mild-to-moderate depression and anxiety	10 (25)	[8,10,13,14,23,26,27,29,32,33]
Conducting future research that has relevance for iCBT implementation is important		
More implementation research is needed to understand the uptake of iCBT within routine care	9 (22)	[7,18,23-25,27,28,33,34]
More research is needed on adverse events to understand the negative effects of iCBT	2 (5)	[19,23]
More research is needed to understand the relationship between adherence and iCBT	4 (10)	[8,23,28,36]

XSL•FO
RenderX

Table 2. Subdomains, categories, and the contributing articles identified under the domain implementation processes related to the successful implementation of internet-delivered cognitive behavioral therapy (N=40).

Subdomain and category and subcategory	Articles, n (%)	Studies
Management of iCBT in the workplace		
Successful training of supporters is important for the provision of iCBT	8 (20)	[4,14,16,22,26,29,31,34]
Training stakeholders within the health system is important in creating awareness of iCBT	2 (5)	[14,29]
Effective management of risk and adverse event management in iCBT is important for its delivery	10 (25)	[5,8,12,14,22,23,27,30,33,75]
iCBT should be delivered through secure and interoperable systems that facilitate clinician and client access	11 (28)	[4,11,12,15,24,27,29,30,34-36]
Operational considerations for managing iCBT and related staff are important		
Effective management and leadership support facilitate the implementation	7 (18)	[12,14,22,24,29,31,34]
Management of workplace resources is necessary to facilitate staff time to use iCBT	4 (10)	[4,26,29,31]
Staff motivation to use iCBT needs to be fostered	4 (10)	[26,27,29,34]
Use of routine monitoring of iCBT to convey intervention effectiveness and enhance its delivery	5 (12)	[27,29,33,34,74]
Effective marketing and service promotion enhance the uptake of iCBT	6 (15)	[4,18,21,29,33,35]
Staff recruitment and retention in iCBT is a challenge that needs to be mitigated against	2 (5)	[4,14]
Scaling of iCBT within services is challenging and requires multiple considerations (eg, infrastructure, funding, proper testing, and governance)	7 (18)	[4,12,14,15,22,33,74]
The practice of iCBT in the workplace		
Appropriate referral pathways and management of waiting times are important for the delivery of iCBT	19 (48)	[2,4,9,12,14,18,20-23,27,30-33,35,73-75]
Screening and inclusion criteria for accessing iCBT need to be thoroughly defined	21 (52)	[8,9,11-14,16,18,20,22,24,26,29,30,32,33,35,72-75]
Considerations of the level of support for patients are crucial in the provision of iCBT		
Positive impact of support on patients	11 (28)	[2,3,5,10,13,23-25,33,36,73]
The quality of support impacts the success of iCBT provision	13 (32)	[8,12,16,22-24,27,29,30,34,36,73,74]
Appropriate considerations should be given to the mediums and modalities of support to fit service and user needs	20 (50)	[2,8,11-13,16,21-27,30,32-35,73,74]
The time demand associated with the provision of support needs to fit service and user needs	18 (45)	[2,7,11,12,18,21-24,26,27,30,32-35,73,74]
The optimal personal and professional background of the supporter needs to be considered in the provision of iCBT	17 (43)	[4,11,12,14-16,21,23-27,30-33,35]
Contextual considerations		
Governmental and health care regulations affect the implementation of iCBT	10 (25)	[4,12-15,24,29,31,33,34]
Lack of workforce availability for iCBT as a limiting factor in the provision of iCBT	4 (10)	[5,27,31,34]
Considering the cost estimates associated with iCBT for patients and health care providers before implementing	11 (28)	[4,8,12-14,22,26,27,29,31,75]

Domain 1: Implementation Insights Derived From iCBT Research

This domain includes categories that contribute to the success of iCBT in either research or routine practice settings or further learning to inform it.

Category 1: Clinician Attitudes Toward iCBT Affect Patient Outcomes and Implementation of iCBT

Clinician attitudes toward iCBT are mixed [48]. Negative attitudes hinder the successful dissemination of iCBT to clients [58]. These attitudes consist of skepticism about the effectiveness and quality of iCBT [9,38], technological limitations of iCBT [56], the inability to generate a therapeutic alliance through this medium [9], preference for face-to-face

XSL•FO
RenderX

therapy [42,65,67], the perceived lower priority of the intervention in the workplace [28], and its highly standardized nature being incompatible with other psychological interventions [28,58]. Such negative attitudes can arise from a lack of iCBT exposure or training [59] and can be transferred to patients and undermine patient outcomes [28]. Accordingly, there is a need to engage with these negative attitudes to successfully implement iCBT [28,65].

In contrast, positive attitudes toward iCBT can increase acceptability and help grow iCBT in service. Positive attitudes include acknowledging the benefits of iCBT in terms of time efficiency, cost-effectiveness, evidence base, program design quality, accessibility, and ability to bridge treatment gaps for those on waiting lists [28]. Professionals with more experience implementing iCBT regard iCBT more positively in terms of its applicability to service provision [62]. Even health care professionals with little exposure to iCBT are generally positive and accepting toward iCBT but have biases around suitability and large knowledge gaps [65].

Category 2: Patient Attitudes Toward iCBT Affect Engagement, Adherence, and Outcomes

The first subcategory under this larger category relates to patients reporting positive attitudes toward the iCBT treatment, its content, and the associated therapist support they receive [28,41,63,67,69,71]. They also report strong motivations to seek out iCBT [45] and acknowledge advantages in terms of convenience, cost, privacy, and self-directed nature [38,67-69]. A second subcategory relates to attitudes, where greater positive initial attitudes predict greater symptom reduction and adherence, and improvement or decline in attitudes during treatment predicts better or worse adherence and outcomes [52,67,69]. The third subcategory relates to negative attitudes, which can be a barrier to treatment success [59]. Such negative patient attitudes include skepticism toward the effectiveness and credibility of iCBT [68,69] as well as toward motivation and accountability to progress through iCBT [68]. There is also a reported preference for face-to-face therapies over iCBT [58,67-69]. Interestingly, offering iCBT as a waiting list treatment can create “unfavorable comparisons” between iCBT and face-to-face therapy, resulting in negative perceptions of iCBT [69].

Category 3: The Delivery of iCBT can be Augmented by Technological and Clinical Design Factors

iCBT has been augmented with novel design elements or treatment strategies to understand their utility and benefit. Such elements and strategies include integrating sensors [57], gamification [57], transdiagnostic elements [38,57,58], iCBT as an add-on or adjunct to existing care pathways [47,56,65], incorporating “persuasive technology” components [57], iCBT use in a blended care model [50,56,57], and its use as a first-line intervention to promote interest in further mental health care [65].

Category 4: Patient Characteristics Related to iCBT Outcomes

A variety of demographic, medical, and technological factors have been implicated to affect iCBT and are reported in the subsequent subcategories. There is mixed evidence for age influencing iCBT outcomes wherein null or negative associations have been observed between age and adherence [67,71] and symptom reduction [39,54]. Mixed results are also reported for gender and sex. Females can have greater or similar iCBT adherence to males [67], and sex was found to be associated with symptom deterioration in unguided iCBT [39]. Perceived technological literacy has been posited to promote adherence [67,70] and clinical outcomes [70], though the evidence is unclear regarding the impact of tech literacy on iCBT outcomes [67]. A review reported that patient-reported medication and alcohol use were not associated with iCBT adherence [67]. Racial and ethnic minority group membership has been negatively associated with adherence to iCBT [66]. Both positive and negative associations have been observed for marital status, employment status [39,67,71], and education level [67] with adherence. Individuals with both lower income levels and marital status of single were more likely to drop out from iCBT [71]. Comorbidity of mental health disorders can reduce the effect of iCBT treatment [45]. Chronic mental health problems are negatively associated with iCBT adherence [71].

Making sudden, large clinical gains on symptom measures at the start of treatment is associated with greater improvements after the treatment [57]. The severity of depression can positively impact clinical outcomes of depression; for example, higher pretreatment severity results in greater effect sizes [4,67]. Symptoms of depression have also been found to negatively impact iCBT adherence, where symptoms like low motivation have been found to be negatively associated with iCBT adherence [23,52]. Chronic mental health problems are further associated with iCBT adherence, where years of living with chronic mental health conditions was found to be negatively associated with adherence to iCBT [71].

Category 5: Guided iCBT as Superior to Unguided iCBT in Regards to Symptom Outcomes and Adherence

Guided iCBT shows superiority over unguided iCBT in terms of adherence and clinical outcomes [44,51,56-58,65,67,71]. However, an individual participant data meta-analysis [39] postulates that the small effects achieved by unguided iCBT are superior to control groups (or no intervention) and can be best used when implemented at scale, such as at the public health level. The therapist element of guided iCBT, in particular, is posited to improve adherence to iCBT [65,67]. Guided iCBT support also fulfills an expressed need to navigate through and explain therapeutic content when patients encounter difficulties [69].

Category 6: iCBT Is as Effective as Face-to-Face Delivery of the Same Protocol, Yet Preference is Often for Face-to-Face Treatment

iCBT produces similar adherence [38,57] and clinical outcomes [48,50,56,57] to equivalent face-to-face therapy. iCBT has advantages over face-to-face therapy in terms of time efficiency, access rates [28,57], and its ability to deliver standardized treatment [22,47]. However, patients demonstrate a preference for face-to-face treatment over iCBT, which can be a reason for the dropout of iCBT [28,43,58,67,68]. In such instances, these

XSL•FO
RenderX

preferences for iCBT over face-to-face treatment can be reduced by introducing a time delay when accessing treatment [43].

Category 7: iCBT Appears to be Effective Beyond the Original Target of Mild-to-Moderate Depression and Anxiety

iCBT is not typically offered for severe presentations of depression and anxiety [59], but real-world data illustrate that a large proportion of patients seen by iCBT clinics have symptoms in this range [22]. Patients with severe symptoms at baseline can make large clinical gains [4,38,45,56], show comparable adherence rates to patients with less severe symptoms [38], and, in some cases, produce larger gains than patients with more severe symptoms [38]. Studies requiring greater treatment-seeking behaviors tend to recruit individuals with more severe symptoms of depression, illustrating the willingness and motivation of this cohort to initiate treatment [38,56]. After treatment initiation, the effect of higher pretreatment severity on adherence and completion is unclear, with a study positing that higher pretreatment severity may be associated with lower iCBT completion rates [64]. Similarly, patients whose symptoms are in the subclinical range also benefit from iCBT [56,63]. Those with suicidal ideation are also found to benefit from iCBT [38,48,56]. iCBT may also be applicable to conditions where depression is secondary to the presenting problem (eg, addiction, trauma, schizophrenia, and bipolar disorder [28,38]).

Category 8: Future Research on Implementation Critical to Advancing iCBT

Several key areas critical to advancing the implementation of iCBT were apparent. First, a plethora of articles stated a lack of research that details the process of implementing iCBT in naturalistic settings; therefore, more research is needed to understand and improve iCBT uptake [38,44,52,56-58,60,64,65]. Second, more research is needed regarding adverse events within iCBT and that current reporting of the adverse events in studies is poor [53,56]. Third, more research is needed to understand the relationship between adherence and iCBT outcomes [45,56,67]. It was also suggested that the definition of "dropout" should not be conceptualized in a binary way because varying dosages of iCBT have been found to produce positive clinical change when less than the intended program is completed. A further study stated that high rates of dropout observed in iCBT research should similarly be expected for practical implementations [60].

Domain 2: Implementation Processes Related to the Successful Implementation of iCBT

This domain includes categories of factors that are important for the successful implementation of iCBT.

Subdomain 1: Management of iCBT Day-to-Day Workplace Operations

This subdomain consists of factors important for managing the day-to-day operations of iCBT, with categories pertaining to the training of staff, risk management, marketing and service promotion, IT infrastructure, working with other services, and managing the staff who work in the provision of iCBT.

Category 1: Successful Training of Supporters Is Important for the Provision of iCBT

Effective training of supporters in iCBT requires technical training in the use of the program [22,50], developing competencies around computer skills and web-based written communication [9,22,62], and practice providing support to fictional patients [50]. Training on writing skills (as many iCBTs rely on supporters messaging the patients on the iCBT platform) should occur before patient interactions rather than through trial and error during their interactions with patients [9]. Training supports, including a manual [59,65], giving clinicians access to training resources [28,65], and providing them with feedback on their written reviews [50], were considered helpful. A study stated that there are limited opportunities for support training in iCBT [37].

Category 2: Educating Stakeholders Within the Health System Is Important to Create Awareness of iCBT

Training and educating other relevant stakeholders (eg, nonclinical staff, referral providers, and patients) about the benefits of iCBT is important to create awareness of the intervention and its clinical effectiveness [28] increasing the perceived viability of the intervention as a treatment option [22].

Category 3: Effective Management of Risk and Adverse Events in iCBT Is Important for Its Delivery

Successfully implemented iCBT has to be supported by a risk monitoring system (eg, suicidal risk measures) that alerts clinicians to risk (eg, triggered automated messages within the iCBT software) and allows clinicians to act on indicated risk (eg, clinician contacts identified risk cases; [22,23,37,42,45,56,64,70]).

Category 4: iCBT Requires Secure, Interoperable Systems That Facilitate Clinician and Client Access

Reviewed articles suggested that iCBT interventions should be hosted on secure servers [28,38,66], should be optimized to run on a variety of mediums (tablets, desktops, and phones) [38], should be integrated with larger patient databases [9,47,61], and should have security standards that adhere to relevant governing bodies [23,57]. Internet connection difficulty [49,65], enabling service computers to access iCBT and its related websites, a lack of integration of iCBT apps with health care records [65], and providing patients access to technology to use iCBT have been cited as limiting factors and may contribute to patient dropout [49,67].

Category 5: Operational Considerations for Managing iCBT and Related Staff Are Important for the Successful Implementation

This category details operational factors that impact the successful implementation of iCBT within clinics or workplaces. First, reviewed articles suggest that management and leadership are important to implement iCBT [28] and include activities such as developing guidelines and service procedures [23,28,37,57], change management [22], and planning implementation and engaging stakeholders [28,65]. In addition, management of workplace resources is necessary to facilitate staff time to use iCBT. For instance, clinicians with already

XSL•FO
RenderX

high workloads may experience time shortages for administering or reviewing the program [28,59]. iCBT clinic managers worry about clinicians balancing iCBT and face-to-face work workloads [9], and it may be necessary to have a dedicated workforce to support iCBT delivery [62].

Second, staff motivation to use iCBT needs to be fostered, as motivation to use iCBT is "essential" [59] because iCBT proponents facilitate the implementation of iCBT [28]. However, fostering this motivation and changing the way clinicians practice is difficult [28,38,65]. One way to foster this motivation may be through routine monitoring of iCBT outcomes, which can provide persuasive evidence of intervention effectiveness to stakeholders as well as enhance its delivery. Services in Australia and Canada reported that they regularly conducted audits of service effectiveness [38,64,71], and such routine monitoring data were used so that staff could evaluate and understand the effectiveness of iCBT in their service [28,65] and to ensure compliance with treatment manual [64].

Third, effective marketing and service promotion are essential to spreading iCBT initiatives [9,28]. Advertisement campaigns (eg, web-based and print media) are frequently successful in recruiting participants for trials and routine care [52,55,64,66], though such marketing campaigns can require a large quantity of resources [28]. Recruitment and retention of therapists in iCBT-related positions is also critical. Clinician recruitment can be an issue as some therapists believe that iCBT limits professional freedom due to its highly structured working requirements and that the working conditions are not attractive [9,22].

Fourth, scaling of iCBT within services is challenging and requires consideration of multiple factors. The physical infrastructure (eg, internet) must be in place [49], funding needs to be procured [9], service decision makers must be convinced of the intervention's feasibility [9], there needs to be evaluation frameworks for existing and new iCBT programs [22,37], and governance frameworks (eg, clinical, IT, and organizational) must be implemented that adhere to the wider legislative context [23]. Exploring new service pathways that are developed when considering iCBT services may allow for existing iCBT services to scale their offering [64]. A study stated that iCBT services should start with a small offering (eg, minimally monitored iCBT) and then acquire human and financial resources over time to build out the service [71].

Subdomain 2: The Practice of iCBT in the Workplace

Category 1: Appropriate Referral Pathways Are Necessary for the Delivery of iCBT

Reviewed articles suggested that successful implementation of iCBT requires the development of appropriate referral pathways, though there are many different referral strategies. Such pathways include self-referral [9,22,23,37,54,64], health care provider referral [22,23,38,40,46,52,54,56,61,62,64,66,70], access to pathways through marketing materials [52,55], contacting patients by email [63], contacting patients on waiting lists for face-to-face services [69], or clients applying through a secure website [71].

Category 2: Screening and Inclusion and Exclusion Criteria for iCBT Are Wide Ranging

Successfully implemented programs required patients to complete a web-based [22,23,28,37,54,63,64,66,68], in-person [23,48,61], or phone screening assessment [57,64,66]. As part of the screening, patients were asked about demographic information, mental health symptoms, commitment to iCBT, treatment history, level of risk, internet access, and language proficiency [28,37,45,48,61,63,64,66,68-71].

A variety of inclusion and exclusion criteria for accessing iCBT are reported as follows: no severe depression [37,59,63,71], no severe anxiety [63], no chronic or recurrent depression [47], no dementia [47], no past history of psychotic symptoms [28,64], aged >18 years [23,45,63,64,66,70,71], aged between 18 and 65 years [52], diagnosis of disorder via psychiatric interview or exceeding cutoff on established measure [23,37,45-47,50,52,64], no comorbid substance abuse [23,45,46,52,64,66] or use of benzodiazepines [46], no suicide risk [23,37,45,46,64,66,69-71], no bipolar disorder, psychosis, or obsessive compulsive disorder [23,45,46,52,66,69,71], adequate understanding of program language [45,46,50,52,66,71], no developmental disorders or other cognitive disabilities [69], no comorbidities or nonpsychiatric diseases that could cause depressive symptoms [70], no concurrent treatment [47,64,66,70,71], no change in medication before 1 month of commencing treatment [71], no email address or technological means to access treatment [46,82,70,71], patients with low motivation [37,64], and patients to far outside of the geographic location of the clinic [37,64,66,70].

Category 3: Use of, and Processes for, Providing Support That Are Crucial to the Provision of iCBT

Supported iCBT achieves positive clinical outcomes for patients [4,40,56-58], provides superior clinical and adherence outcomes over unguided interventions [4,40,56,64,67], and is regarded positively by the patients [41,69]. Therapeutic alliance is implicated as a mechanism behind the positive effects of supported interventions [48]; however, its effects are still unclear in iCBT, as it has been associated with positive outcomes [42] or no effect [48].

The purpose of support in iCBT is to "recognize and reinforce the participants' work with the self-help material" [56] and promote engagement with the intervention [65,71]. The supporter in iCBT is posited to assume the role of a motivator, where the iCBT platform delivers the core treatment elements [38,56], and involves therapists monitoring patient progress [23,28,37,69], responding to their iCBT-related needs [28,67,69], and guiding the user through the initial setup [45,61]. Through written support, clinicians can encourage and affirm patients by expressing interest in the thoughts, feelings, and behaviors that patients share [50,57]. The quality of support affects client outcomes. For instance, leniency toward patient accountability (eg, homework completion) is associated with poorer patient outcomes [57]. Moreover, within written messages, the studies found that misspellings were frequent, emojis and emoticons were seldom used, and less detailed, shorter messages were associated with fewer web-based sessions completed [50].

XSL•FO
RenderX

iCBT support has been delivered in many ways [56]: in person [55,59], over email [38,64,71], by telephone [23,40,55,64,69,71], through the iCBT platform [23,45,50,55,63,66], through videoconferencing software [63], or by SMS text message [23]. Support can occur in real-time [58], on an "on-demand" basis [37], or asynchronously [48,58,64,66]. Support can be delivered weekly [23,47,55-57,61,64-66] or constantly through ongoing therapist monitoring [63]. Some programs incorporate homework assignments to inform clinicians when conducting support sessions [23,37]. In total, 3 programs implemented "step-wise" access to modules, where new content could not be accessed without completing a supported session [37], was unlocked 7 days after the completion of the previous module [69], or released gradually over an 8-week period [64].

Such decisions around how support is delivered and its frequency likely influence clinician time demand. Time spent delivering support varies widely [56], ranging from 10 to 100 minutes per session [23,38,40,44,55-57,59,61,64,69,71] and up to 8 hours per individual per course of treatment [38]. Support can be delivered weekly [23,47,55-57,64-66] or constantly through ongoing therapist monitoring [63]. Some programs involve 6 to 12 supporter telephone calls [40], emails from the supporter [52], or that the supporter contacts the patient ≥1 time a week for 8 weeks [47]. The end of treatment for some programs was based on a specific time or the number of support sessions received; for example, iCBT was cited across articles to be delivered over a varying course of 7 to 20 weeks [23,37,44,49,55,61,63,64].

Category 4: iCBT Can Be Successfully Implemented Across Supporters With a Variety of Personal and Professional Backgrounds

Successful implementation of iCBT included supporters who were volunteer peer-supporters with lived experience of the mental health condition [61], trained volunteers [49], psychologically trained experts (unspecified qualifications; [38]), clinical psychologists [9,23,47,50,61-64,66], psychiatrists [9], registered or provisionally registered mental health professionals [22,61], graduate students of psychology [22,66], trained health care professionals [49], psychologists in training [50,62], psychotherapists [62], social workers [62,64], mental health nurses [50], nurses [64], therapists with training on addictions [64], trained technicians [55,56,58], and general practitioners [59]. There is evidence that untrained technicians [56,58] or novice clinicians [57] achieve equal outcomes compared to trained clinicians, and that support from a technician is more effective than a waitlist control group [55].

Subdomain 3: Contextual Considerations for iCBT Implementation

Category 1: Governmental and Health Care Regulations Affect Implementation of iCBT

Governmental and health care regulations influence how iCBT can be implemented. An example of this is in Canada, where iCBT has been recognized by the Canadian government, which provides specific funding streams for iCBT services and research [28,64,65]. Other countries have implemented policies that incentivize the use of iCBT to improve access to psychological therapy [9,49]. Other regulations impacting iCBT include limitations placed on therapeutic contact taking place over the internet [48,49], requiring iCBT clinics to adhere to existing frameworks for the delivery of therapy [22,23,57] and policies around the delivery of therapy over the internet [57,62].

Category 2: Lack of Workforce Availability for iCBT as a Limiting Factor in the Provision of iCBT

The lack of workforce availability can limit the implementation of iCBT, particularly because as access to mental health care increases, so does the demand for services [38,42]. One study in Sweden observed only 1 to 2 therapists participating in iCBT initiatives among implementing services and further commented that due to face-to-face resources being expensive and scarce, a dedicated iCBT workforce could resolve this issue in terms of resource and cost [62]. Another study stated that the presence of trained iCBT professionals in certain health sectors (eg, veteran care in the United States) is rare [65].

Category 3: Cost Models and Cost-Effectiveness of iCBT Maybe an Asset for Successful Implementation

iCBT was provided to patients through the following 5 cost models: free of charge [59], through publicly funded health care systems [9,22,23,28,37,45,48], subsidized by health care providers [23,38], at a cost to patients when they are not within certain catchment areas or countries [38], or as part of insurance plans [48]. The establishment of reimbursement systems for iCBT was cited as an important factor for cost estimates in the future [9,48]. Moreover, it was hypothesized that as iCBT cost-effectiveness becomes more salient, health care providers (public or private) will advocate for it as a first-line intervention to efficiently gatekeep therapeutic resources [38]. iCBT did not incur extra costs to public health care systems [28] and was cost-effective (depending on the "willingness to pay" standards of the health care body) [70]. One study posited that a dedicated iCBT workforce should be developed to create a less expensive alternative [62].

Discussion

Principal Findings

Our mixed methods systematic review highlights the knowledge we have gained from the available literature on the implementation of iCBT. Some of our key findings regarding the process for implementing iCBT include the practice of iCBT, with special reference to determining client eligibility and effectively supporting patients in iCBT. The management of iCBT in the workplace, especially staff and operational considerations, also surfaced as an important process to consider when implementing. Other related findings include the importance of staff training, the management of treatment pathways, security, and factors for consideration within the wider context that impact the implementation of iCBT. In terms of implementation insights, this review has highlighted that clinician and patient attitudes toward iCBT can influence its ability to achieve intended outcomes and the need to continually tailor iCBT for patient benefit and that further research can help to develop our understanding of implementing iCBT successfully.

XSL•FO
RenderX

As would be expected, the practice of iCBT was highlighted as important to the implementation process for iCBT within a mental health service. For instance, what constitutes eligibility for an iCBT intervention manifested in two categories: (1) screening and inclusion criteria for accessing iCBT need to be thoroughly defined and (2) consideration of the usefulness of iCBT beyond the original target of mild-to-moderate depression and anxiety. Historically, eligibility for iCBT has been characterized by low symptom presence (mild to moderate) and no significant risk issues. This approach was sensible while establishing iCBT's safety and effectiveness as an intervention, subsequently resulting in a well-validated evidence base supporting iCBT for treating depression and anxiety. Consequently, the preponderance of historical eligibility seems to be an artifact in need of revision. This is particularly important in light of the growing body of literature to support iCBT's applicability to more severe presentations of mental health difficulties [72-75]. In addition, real-world data from iCBT clinics highlight that a substantial proportion of patients accessing these services have presentations within the moderate to severe range [22].

There is a clear need for services to consider the populations they serve (eg, general severity levels and client demographics) and tailor their model of iCBT provision. Still, despite the available evidence, clinical guidelines lag in their support for iCBT in extended service delivery pathways [76]. This situation poses some difficulty for certain services or health systems to innovate their use of digital interventions (eg, the improving access to psychological therapies program in England, which offers treatment based on National Institute for Health and Care Excellence guidelines). Specifically in the English context, the original guideline for the use of iCBT was rolled out in 2004 [77,78] and was updated in 2022. Since then, technologies and research have developed and would suggest the utility of iCBT for the broader population.

Our results highlight the importance of the operational aspects of iCBT services. First, the importance of effective management and leadership to support the implementation of iCBT was identified through the review. Transformational leadership approaches, that is, leadership styles associated with motivating and compelling employees to participate in a shared vision [79], have been found to be associated with increased levels of innovation climate, further defined as an organizational climate that is conducive to the adoption of novel, evidence-based practices [80]. Implementing iCBT requires leaders to navigate interactions across multiple levels of a service and motivate staff to ensure the vision of iCBT is fulfilled. However, the current studies identified do not illustrate in depth the effects of leadership nor was it their primary or secondary focus. Despite this, it is still important that this finding was communicated through this small pool of studies, and more research is needed to inform this gap in knowledge.

Training staff in iCBT and increasing their motivation to use it were both cited as important. As an in-service activity, training clinicians and therapists in the use of evidence-based practice has a substantial literature base [81,82]. However, our findings highlight variance in the training delivered to therapists charged with delivering iCBT, ranging from hours to up to a year of continued education, and the components of the training were also not described at length across articles. The wider literature on training stands in contrast to what we identified; training programs for evidence-based practice tend to produce better outcomes (eg, competency, evidence-based practice use, and positive attitudes) when multicomponent approaches are used (eg, workshop, follow-up, and audit of skills acquired) [81]. To date, no systematic evaluation of iCBT training programs has been conducted, and it has also been cited by one of the included studies that training programs for these interventions are rare [37]. Similarly, we identified that staff motivation to use iCBT needs to be fostered. This motivation can also be developed through training initiatives, where implementers can illustrate the benefits that iCBT brings to routine clinical practice (eg, improves patient symptoms and access to care, is usable, and is not time consuming), and this activity may potentially influence motivation around intervention use [83,84].

Furthermore, routine monitoring of the intervention and its outcomes was also cited as important for the continued development of iCBT within the service. This activity can allow supporters in iCBT to reflect on their own practice to improve service provision, with an article stating that clinicians who administer iCBT desire comprehensive updates regarding iCBT to understand its impact on wider service outcomes [28]. This activity is reminiscent of the construct "reflexive monitoring" from normalization process theory [84], where individual and group reflections on processes around a specific evidence-based practice can lead to revisions in practice that are adapted to best suit the needs and structures of the service context. The results regarding the operational aspects of iCBT, despite not being widely reported across the literature, indicate that factors associated with evidence-based practice success in the implementation literature are being considered when iCBT is implemented, which is a promising finding. More widespread reporting of this information could be beneficial to practicing professionals when making choices about using iCBT with their patients.

Patients tend to be positive about iCBT and the support they receive, but clinician attitudes generally lean toward the negative. From service illustrations, we can infer that clinicians receive significant exposure to iCBT when it is implemented [22,23,38], and its effectiveness is grounded in the literature. Negative attitudes can result in the abandonment of the implementation effort due to a lack of acceptance or misunderstandings around the perceived value of the treatment [85,86]. A better understanding of negative clinician attitudes can be attained if iCBT were to be interpreted as a novel, evidence-based practice. A literature search around clinicians' attitudes toward evidence-based practice provides some insights, including that clinicians rate "other" sources of information (eg, colleague opinion and previous experience) as more impactful than published evidence on their decisions for treatment [87-89]. Fostering attitudes conducive to the uptake of evidence-based practice has been associated with transformational leadership styles [80] and systematic training initiatives that highlight how the evidence-based practice is integrated with the wider service system [82], both of which were evident within the current review. However, where there is a disconnect between clinicians

XSL•FO
RenderX

and service management or staff does not understand the relative advantage (from the diffusion of innovations) [90], iCBT over existing practice can subsequently create barriers to evidence-based practice uptake [90,91]. This disconnect is well documented in implementation science theories, such as readiness for change [92] and implementation climate [93,94], both of which also emphasize the role of attitudes in evidence-based practice use and implementation.

iCBTs vary widely in their delivery (guided and self-guided), support time frames, and those who provide the support, but the take-home insight is that patients receive the interventions well in terms of satisfaction and clinical outcomes achieved. This malleability of iCBT, where it can assume many forms yet achieve the intended results, underlines the scalability of the intervention. A narrative review of factors associated with scaling public health interventions described that, once an intervention has proven its effectiveness in both small- and large-scale trials, management and practice factors, such as having systems for monitoring intervention performance, funding, and interacting with stakeholders within the wider health care system, become important for the scaling process [95]. Despite this, an "implementation gap" remains, where effectiveness reductions and high levels of attrition occur when we transition from efficacy settings to real-world service provision [96]. This further creates a treatment gap within users assigned iCBT as a treatment option, where implementation factors (eg, abandonment due to encountering bugs and not having a care provider to explore these with) may cause attrition, lowering the promise of scalability purported by these treatments [97].

Limitations

In total, 4 main limitations were identified as part of this mixed methods systematic review. First, we used a targeted search strategy to produce a dataset that the authors acknowledge is incomplete due to a lack of proper use of terminology within the field to reference implementation. Therefore, we acknowledge that this review is not definitive on the implementation of iCBT and only reports on relevant factors within the articles identified. There is already a movement to standardize the reporting of digitally delivered psychological treatments in research studies (eg, use of CONSORT [Consolidated Standards of Reporting Trials] for eHealth) [98], and perhaps this should be succeeded by an attempt to standardize how we report implementation learnings too.

The second limitation consisted of the "blind spots" associated with the development of the analytic framework that may have resulted from the background of the researchers. The authors mainly come from a background in psychology, and none would consider themselves to be implementation specialists. Other review types (eg, realist, scoping, or narrative reviews) conducted by different research groups may uncover nuances that were otherwise unidentified by this review.

The third limitation of the study relates to the limitations present in the original papers included in this review. The heterogeneous nature of the articles included prevented a formal quality appraisal from being conducted. For instance, we had considered using a tool, such as the Critical Appraisal Skills Programme checklist, but this tool did not provide for a quality review of the narrative-type articles (eg, Schröder et al [56], Andersson et al [57], Andersson et al [58], and others) that were included. Furthermore, there are issues around assessing the quality of qualitative evidence within review-type studies and whether studies should be excluded based on perceived quality, which further compounds the issue [99-101]. Thus, we did not assess the quality of the included articles, and no articles were excluded based on methodological flaws. Relatedly, a few articles had a primary objective of exploring a facet of the implementation of iCBT for depression and anxiety, which is important to note when interpreting the results.

The fourth and final limitation of the study is whether an expansion of search terms could have been used. For instance, the term "computerized CBT" was not used in the search terms. However, while the search string did not include this term, we did include an abbreviated term for this, "ccbt," and it is notable that by using this abbreviated term we identified specific studies that referenced computerized cognitive behavioral therapy (eg, Wells et al [44], Wright et al [4], Kenicer et al [49], Wright et al [51], and Grist and Cavanagh [54]). Regardless, it is possible that the search string may not have completely captured all possible studies.

Conclusions

This mixed methods systematic review has identified several strategies for consideration when attempting to implement iCBT. Broadly, these strategies emphasize the importance of effective leadership, managing staff and operations associated with the practice of iCBT, implementing and developing professionals to provide the supported component of iCBT, accounting for context, and deriving implementation insights from novel research contributions in iCBT for patient benefit. Future research into iCBT in real-world settings should endeavor to supply appropriate supplemental information that details the efforts associated with implementing the intervention within care pathways. In tandem, efforts could be made to standardize practices that can support the transferability of learning and scalability through the use of a standardized lexicon of terms that are appropriately used.

Acknowledgments

The authors would like to acknowledge Angel Enrique, senior manager and digital health scientist at SilverCloud Health, for his feedback on early iterations of this review. This research was supported with joint funding from the Irish Research Council and SilverCloud Health under the employment-based postgraduate program (grant EBPPG/2018/198).

XSL•FO
RenderX

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Authors' Contributions

DD was responsible for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, writing the original draft, and reviewing and editing the manuscript. DR was responsible for supervision, formal analysis, funding acquisition, and reviewing and editing the manuscript. GH was responsible for conceptualization, visualization, writing the original draft, and reviewing and editing the manuscript. LT was responsible for funding, supervision, and reviewing and editing the manuscript.

Conflicts of Interest

DD, DR, and GH are employees of SilverCloud Health (now a subsidiary of American Well Corporation). DR and GH are shareholders of American Well Corporation, a company specializing in the development of telehealth solutions for physical and mental health conditions. This study formed part of the doctoral research conducted by DD, which was funded by the Irish Research Council and SilverCloud Health as part of the employment-based postgraduate research program and was hosted by Trinity College Dublin, Ireland. LT is a professor in counseling psychology at Trinity College Dublin and an external research consultant to SilverCloud Health.

Multimedia Appendix 1
Search terms used within mixed methods systematic review.
[DOCX File , 34 KB - jmir_v27i1e47927_app1.docx]

Multimedia Appendix 2
Description of included articles within mixed methods systematic review.
[DOCX File , 36 KB - jmir_v27i1e47927_app2.docx]

Multimedia Appendix 3
References analyzed as part of this mixed methods systematic review.
[DOCX File , 257 KB - jmir_v27i1e47927_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 54 KB - jmir_v27i1e47927_app4.pdf]

References

1. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
2. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3(3):CD011565 [FREE Full text] [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
3. Romijn G, Batelaan N, Kok R, Koning J, van Balkom A, Titov N, et al. Internet-delivered cognitive behavioral therapy for anxiety disorders in open community versus clinical service recruitment: meta-analysis. J Med Internet Res 2019 Apr 17;21(4):e11706 [FREE Full text] [doi: 10.2196/11706] [Medline: 30994462]
4. Wright JH, Owen JJ, Richards D, Eells TD, Richardson T, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. J Clin Psychiatry 2019 Mar 19;80(2):11 [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
5. Jardine J, Earley C, Richards D, Timulak L, Palacios JE, Duffy D, et al. The experience of guided online therapy: a longitudinal, qualitative analysis of client feedback in a naturalistic RCT. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-15 URL: https://dl.acm.org/doi/10.1145/3313831.3376254 [doi: 10.1145/3313831.3376254]
6. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The acceptability of computer-aided cognitive behavioural therapy: a pragmatic study. Cogn Behav Ther 2009;38(4):235-246. [doi: 10.1080/16506070802561256] [Medline: 19306147]
7. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]

XSL•FO
RenderX

8. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
9. Folker AP, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: a comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018 Mar;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
10. Druss BG, Cohen AN, Brister T, Cotes RO, Hendry P, Rolin D, et al. Supporting the mental health workforce during and after COVID-19. Psychiatr Serv 2021 Oct 01;72(10):1222-1224. [doi: 10.1176/appi.ps.202000509] [Medline: 33882690]
11. Andersson G, Berg M, Riper H, Huppert JD, Titov N. The possible role of internet-delivered psychological interventions in relation to the COVID-19 pandemic. Clin Psychol Eur 2020 Sep 30;2(3):e3941 [FREE Full text] [doi: 10.32872/cpe.v2i3.3941] [Medline: 36398150]
12. Rodriguez-Villa E, Naslund J, Keshavan M, Patel V, Torous J. Making mental health more accessible in light of COVID-19: scalable digital health with digital navigators in low and middle-income countries. Asian J Psychiatr 2020 Dec;54:102433. [doi: 10.1016/j.ajp.2020.102433] [Medline: 33271713]
13. Lipschitz J, Hogan TP, Bauer MS, Mohr DC. Closing the research-to-practice gap in digital psychiatry: the need to integrate implementation science. J Clin Psychiatry 2019 May 14;80(3):2969 [FREE Full text] [doi: 10.4088/JCP.18com12659] [Medline: 31091029]
14. Eccles MP, Mittman BS. Welcome to implementation science. Implementation Sci 2006 Feb 22;1(1):6. [doi: 10.1186/1748-5908-1-1]
15. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. Yearb Med Inform 2018 Mar 07;09(01):65-70. [doi: 10.1055/S-0038-1637943]
16. Grant J, Cottrell R, Cluzeau F, Fawcett G. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. BMJ 2000 Apr 22;320(7242):1107-1111 [FREE Full text] [doi: 10.1136/bmj.320.7242.1107] [Medline: 10775218]
17. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
18. Smith JD, Polaha J. Using implementation science to guide the integration of evidence-based family interventions into primary care. Fam Syst Health 2017 Jun;35(2):125-135. [doi: 10.1037/fsh0000252] [Medline: 28617015]
19. Kadesjö Banck J, Bernhardsson S. Experiences from implementation of internet-delivered cognitive behaviour therapy for insomnia in psychiatric health care: a qualitative study applying the NASSS framework. BMC Health Serv Res 2020 Aug 08;20(1):729 [FREE Full text] [doi: 10.1186/s12913-020-05596-6] [Medline: 32771024]
20. van der Vaart R, Worm-Smeitink M, Bos Y, Wensing M, Evers A, Knoop H. Implementing guided ICBT for chronic pain and fatigue: a qualitative evaluation among therapists and managers. Internet Interv 2019 Dec;18:100290 [FREE Full text] [doi: 10.1016/j.invent.2019.100290] [Medline: 31737491]
21. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. Health Technol Assess 2021 Sep;25(57):1-132 [FREE Full text] [doi: 10.3310/hta25570] [Medline: 34590577]
22. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1239 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
23. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
24. Richards D, Viganó N. Online counseling: a narrative and critical review of the literature. J Clin Psychol 2013 Sep;69(9):994-1011 [FREE Full text] [doi: 10.1002/jclp.21974] [Medline: 23630010]
25. Friesen LN, Hadjistavropoulos HD, Pugh NE. A qualitative examination of psychology graduate students' experiences with guided Internet-delivered cognitive behaviour therapy. Internet Interv 2014 Apr;1(2):41-48. [doi: 10.1016/j.invent.2014.04.001]
26. Terpstra JA, van der Vaart R, Spillekom-van Koulil S, van Dam A, Rosmalen JG, Knoop H, et al. Becoming an eCoach: training therapists in online cognitive-behavioral therapy for chronic pain. Patient Educ Couns 2018 Sep;101(9):1702-1707. [doi: 10.1016/j.pec.2018.03.029] [Medline: 29661704]
27. Hilty DM, Chan S, Torous J, Luo J, Boland RJ. A telehealth framework for mobile health, smartphones, and apps: competencies, training, and faculty development. J Technol Behav Sci 2019 Apr 15;4(2):106-123. [doi: 10.1007/S41347-019-00091-0]
28. Hadjistavropoulos HD, Nugent MM, Dirkse D, Pugh N. Implementation of internet-delivered cognitive behavior therapy within community mental health clinics: a process evaluation using the consolidated framework for implementation research. BMC Psychiatry 2017 Sep 12;17(1):331 [FREE Full text] [doi: 10.1186/s12888-017-1496-7] [Medline: 28899365]

XSL•FO
RenderX

29. Goldsmith MR, Bankhead CR, Austoker J. Synthesising quantitative and qualitative research in evidence-based patient information. J Epidemiol Community Health 2007 Mar;61(3):262-270 [FREE Full text] [doi: 10.1136/jech.2006.046110] [Medline: 17325406]
30. Stern C, Lizarondo L, Carrier J, Godfrey C, Rieger K, Salmond S, et al. Methodological guidance for the conduct of mixed methods systematic reviews. JBI Evid Synth 2020 Oct;18(10):2108-2118. [doi: 10.11124/JBISRIR-D-19-00169] [Medline: 32813460]
31. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
32. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci 2013 Dec 01;8:139 [FREE Full text] [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]
33. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017 Mar 23;6(1):61 [FREE Full text] [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]
34. Pluye P, Hong QN. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. Annu Rev Public Health 2014;35:29-45. [doi: 10.1146/annurev-publhealth-032013-182440] [Medline: 24188053]
35. Elliott R, Timulak L. Descriptive and interpretive approaches to qualitative research. In: Miles J, Gilbert P, editors. A Handbook of Research Methods for Clinical and Health Psychology. Oxford, UK: Oxford Academic Press; 2005:147-160.
36. Elliott Jr RK, Timulak L. Essentials of Descriptive-Interpretive Qualitative Research: A Generic Approach. New York, NY: American Psychological Association; 2021.
37. El Alaoui S, Hedman E, Kaldo V, Hesser H, Kraepelien M, Andersson E, et al. Effectiveness of Internet-based cognitive-behavior therapy for social anxiety disorder in clinical psychiatry. J Consult Clin Psychol 2015 Oct;83(5):902-914 [FREE Full text] [doi: 10.1037/a0039198] [Medline: 26009780]
38. Andrews G, Williams AD. Up-scaling clinician assisted internet cognitive behavioural therapy (iCBT) for depression: a model for dissemination into primary care. Clin Psychol Rev 2015 Nov;41:40-48. [doi: 10.1016/j.cpr.2014.05.006] [Medline: 25043445]
39. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. Psychol Med 2018 Nov 15;48(15):2456-2466 [FREE Full text] [doi: 10.1017/S0033291718000648] [Medline: 29540243]
40. Gellatly J, Chisnall L, Seccombe N, Ragan K, Lidbetter N, Cavanagh K. @Home eTherapy service for people with common mental health problems: an evaluation. Behav Cogn Psychother 2018 Jan 16;46(1):115-120. [doi: 10.1017/S1352465817000297] [Medline: 28506333]
41. Richards D, Murphy T, Viganó N, Timulak L, Doherty G, Sharry J, et al. Acceptability, satisfaction and perceived efficacy of "space from depression" an internet-delivered treatment for depression. Internet Interv 2016 Sep;5:12-22 [FREE Full text] [doi: 10.1016/j.invent.2016.06.007] [Medline: 30135802]
42. Woods AP, Stults CB, Terry RL, Rego SA. Strengths and limitations of internet-based cognitive-behavioral treatments for anxiety disorders. Pragmat Case Stud Psychother 2017 Nov 12;13(3):271-283. [doi: 10.14713/pcsp.v13i3.2015]
43. Peynenburg VA, Mehta S, Hadjistavropoulos HD. Postsecondary student perceptions and preferences for the treatment of depression and anxiety: comparison of internet-delivered cognitive behaviour therapy to face-to-face cognitive behaviour therapy and medication. Can J Behav Sci 2020 Jul;52(3):220-230. [doi: 10.1037/cbs0000165]
44. Wells MJ, Owen JJ, McCray LW, Bishop LB, Eells TD, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression in primary care. Prim Care Companion CNS Disord 2018 Mar 01;20(2):11. [doi: 10.4088/pcc.17r02196]
45. Mathiasen K, Riper H, Andersen TE, Roessler KK. Guided internet-based cognitive behavioral therapy for adult depression and anxiety in routine secondary care: observational study. J Med Internet Res 2018 Nov 28;20(11):e10927 [FREE Full text] [doi: 10.2196/10927] [Medline: 30487118]
46. Nordgreen T, Gjestad R, Andersson G, Carlbring P, Havik OE. The effectiveness of guided internet-based cognitive behavioral therapy for social anxiety disorder in a routine care setting. Internet Interv 2018 Sep;13:24-29 [FREE Full text] [doi: 10.1016/j.invent.2018.05.003] [Medline: 30206515]
47. Whiteside U, Richards J, Steinfeld B, Simon G, Caka S, Tachibana C, et al. Online cognitive behavioral therapy for depressed primary care patients: a pilot feasibility project. Perm J 2014;18(2):21-27 [FREE Full text] [doi: 10.7812/TPP/13-155] [Medline: 24867546]
48. Andersson G, Hedman E. Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings. Verhaltenstherapie 2013 Aug 22;23(3):140-148. [doi: 10.1159/000354779]
49. Kenicer D, McClay CA, Williams C. A national survey of health service infrastructure and policy impacts on access to computerised CBT in Scotland. BMC Med Inform Decis Mak 2012 Sep 08;12(1):102 [FREE Full text] [doi: 10.1186/1472-6947-12-102] [Medline: 22958309]
50. Mol M, Dozeman E, Provoost S, van Schaik A, Riper H, Smit JH. Behind the scenes of online therapeutic feedback in blended therapy for depression: mixed-methods observational study. J Med Internet Res 2018 May 03;20(5):e174 [FREE Full text] [doi: 10.2196/jmir.9890] [Medline: 29724708]

XSL•FO
RenderX

51. Wright JH, McCray LW, Eells TD, Gopalraj R, Bishop LB. Computer-assisted cognitive-behavior therapy in medical care settings. Curr Psychiatry Rep 2018 Sep 07;20(10):92. [doi: 10.1007/s11920-018-0947-2] [Medline: 30194498]
52. Schröder J, Berger T, Meyer B, Lutz W, Späth C, Michel P, et al. Impact and change of attitudes toward internet interventions within a randomized controlled trial on individuals with depression symptoms. Depress Anxiety 2018 May 28;35(5):421-430. [doi: 10.1002/da.22727] [Medline: 29489038]
53. Gullickson KM, Hadjistavropoulos HD, Dear BF, Titov N. Negative effects associated with internet-delivered cognitive behaviour therapy: an analysis of client emails. Internet Interv 2019 Dec;18:100278 [FREE Full text] [doi: 10.1016/j.invent.2019.100278] [Medline: 31890627]
54. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251. [doi: 10.1007/S10879-013-9243-y]
55. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014 May 20;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
56. Schröder J, Berger T, Westermann S, Klein JP, Moritz S. Internet interventions for depression: new developments. Dialogues Clin Neurosci 2022 Apr 01;18(2):203-212. [doi: 10.31887/dcns.2016.18.2/jschroeder]
57. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb 02;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
58. Andersson G. The promise and pitfalls of the internet for cognitive behavioral therapy. BMC Med 2010 Dec 07;8(1):82 [FREE Full text] [doi: 10.1186/1741-7015-8-82] [Medline: 21138574]
59. Wilhelmsen M, Høifødt RS, Kolstrup N, Waterloo K, Eisemann M, Chenhall R, et al. Norwegian general practitioners' perspectives on implementation of a guided web-based cognitive behavioral therapy for depression: a qualitative study. J Med Internet Res 2014 Sep 10;16(9):e208 [FREE Full text] [doi: 10.2196/jmir.3556] [Medline: 25208886]
60. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013 Apr 15;13(1):113 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
61. Cavanagh K, Seccombe N, Lidbetter N. The implementation of computerized cognitive behavioural therapies in a service user-led, third sector self help clinic. Behav Cogn Psychother 2011 Jul 22;39(4):427-442. [doi: 10.1017/S1352465810000858] [Medline: 21338531]
62. Brantnell A, Woodford J, Baraldi E, van Achterberg T, von Essen L. Views of implementers and nonimplementers of internet-administered cognitive behavioral therapy for depression and anxiety: survey of primary care decision makers in Sweden. J Med Internet Res 2020 Aug 12;22(8):e18033 [FREE Full text] [doi: 10.2196/18033] [Medline: 32784186]
63. Calero Elvira A, Santacreu Ivars M, Marchena Giráldez C, Shih P. Internet-based cognitive behavioural therapy programme with and without videoconference guidance sessions: a randomized controlled trial to treat work-related symptoms of anxiety and depression. Clin Psychol Psychother 2021 Sep 23;28(5):1230-1242. [doi: 10.1002/cpp.2571] [Medline: 33586819]
64. Hadjistavropoulos HD, Peynenburg V, Thiessen DL, Nugent M, Karin E, Staples L, et al. Utilization, patient characteristics, and longitudinal improvements among patients from a provincially funded transdiagnostic internet-delivered cognitive behavioural therapy program: observational study of trends over 6 years. Can J Psychiatry 2022 Mar 12;67(3):192-206 [FREE Full text] [doi: 10.1177/07067437211006873] [Medline: 33840264]
65. Leung LB, Dyer KE, Yano EM, Young AS, Rubenstein LV, Hamilton AB. Collaborative care clinician perceptions of computerized cognitive behavioral therapy for depression in primary care. Transl Behav Med 2020 Aug 07;10(3):565-572 [FREE Full text] [doi: 10.1093/tbm/ibz122] [Medline: 32766864]
66. Lindegaard T, Seaton F, Halaj A, Berg M, Kashoush F, Barchini R, et al. Internet-based cognitive behavioural therapy for depression and anxiety among Arabic-speaking individuals in Sweden: a pilot randomized controlled trial. Cogn Behav Ther 2021 Jan 30;50(1):47-66 [FREE Full text] [doi: 10.1080/16506073.2020.1771414] [Medline: 32603632]
67. Treanor CJ, Kouvonen A, Lallukka T, Donnelly M. Acceptability of computerized cognitive behavioral therapy for adults: umbrella review. JMIR Ment Health 2021 Jul 06;8(7):e23091 [FREE Full text] [doi: 10.2196/23091] [Medline: 34255714]
68. McCall HC, Sison AP, Burnett JL, Beahm JD, Hadjistavropoulos HD. Exploring perceptions of internet-delivered cognitive behaviour therapy among public safety personnel: informing dissemination efforts. Int J Environ Res Public Health 2020 Aug 19;17(17):6026 [FREE Full text] [doi: 10.3390/ijerph17176026] [Medline: 32824994]
69. Pedersen MK, Mohammadi R, Mathiasen K, Elmose M. Internet-based cognitive behavioral therapy for anxiety in an outpatient specialized care setting: a qualitative study of the patients' experience of the therapy. Scand J Psychol 2020 Dec 23;61(6):846-854. [doi: 10.1111/sjop.12665] [Medline: 32705703]
70. Piera-Jiménez J, Etzelmueller A, Kolovos S, Folkvord F, Lupiáñez-Villanueva F. Guided internet-based cognitive behavioral therapy for depression: implementation cost-effectiveness study. J Med Internet Res 2021 May 11;23(5):e27410 [FREE Full text] [doi: 10.2196/27410] [Medline: 33973857]
71. Robichaud M, Talbot F, Titov N, Dear BF, Hadjistavropoulos HD, Hadjistavropoulos T, et al. Facilitating access to iCBT: a randomized controlled trial assessing a translated version of an empirically validated program using a minimally monitored

XSL•FO
RenderX

delivery model. Behav Cogn Psychother 2020 Mar 16;48(2):185-202. [doi: 10.1017/S135246581900047X] [Medline: 31416489]
72. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-delivered cognitive behavior therapy as a prequel to face-to-face therapy for depression and anxiety: a naturalistic observation. Front Psychiatry 2019 Jan 9;10:902 [FREE Full text] [doi: 10.3389/fpsyt.2019.00902] [Medline: 31998149]
73. Richards D, Duffy D, Burke J, Anderson M, Connell S, Timulak L. Supported internet-delivered cognitive behavior treatment for adults with severe depressive symptoms: a secondary analysis. JMIR Ment Health 2018 Oct 02;5(4):e10204 [FREE Full text] [doi: 10.2196/10204] [Medline: 30279154]
74. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013 Feb 26;346:f540 [FREE Full text] [doi: 10.1136/bmj.f540] [Medline: 23444423]
75. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interv 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
76. The NICE Guideline on the treatment and management of depression (CG90). The National Institute for Health and Care Excellence. 2009. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC548164/#:~:text=The%20guidelines%20do%20not%20recommend,also%20present%20for%20psychotherapeutic%20interventions [accessed 2024-04-29]
77. Anxiety: Management of anxiety (panic disorder, with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care (CG22). The National Institute for Health and Care Excellence. 2004. URL: https://www.nice.org.uk/guidance/cg22 [accessed 2024-04-29]
78. Depression: management of depression in primary and secondary care (CG23). The National Institute for Health and Care Excellence. 2004. URL: https://www.nice.org.uk/guidance/cg23 [accessed 2024-04-29]
79. Judge TA, Bono JE. Five-factor model of personality and transformational leadership. J Appl Psychol 2000;85(5):751-765. [doi: 10.1037//0021-9010.85.5.751]
80. Aarons GA, Sommerfeld DH. Leadership, innovation climate, and attitudes toward evidence-based practice during a statewide implementation. J Am Acad Child Adolesc Psychiatry 2012 Apr;51(4):423-431 [FREE Full text] [doi: 10.1016/j.jaac.2012.01.018] [Medline: 22449648]
81. Frank HE, Becker-Haimes EM, Kendall PC. Therapist training in evidence-based interventions for mental health: a systematic review of training approaches and outcomes. Clin Psychol (New York) 2020 Sep;27(3):e12330 [FREE Full text] [doi: 10.1111/cpsp.12330] [Medline: 34092941]
82. Beidas RS, Kendall PC. Training therapists in evidence-based practice: a critical review of studies from a systems-contextual perspective. Clin Psychol (New York) 2010 Mar;17(1):1-30 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01187.x] [Medline: 20877441]
83. Sivakumar A, Pan RY, Choi D, Wang AH, Yu CH. "If we got a win-win, you can sell it to everybody": a qualitative study employing normalization process theory to identify critical factors for eHealth implementation and scale-up in primary care. Can J Diabetes 2022 Mar;46(2):181-188 [FREE Full text] [doi: 10.1016/j.jcjd.2021.08.005] [Medline: 35283061]
84. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
85. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
86. Rooshenas L, Elliott D, Wade J, Jepson M, Paramasivan S, Strong S, ACST-2 study group, By-Band-Sleeve study group, Chemorad study group, CSAW study group, Optima prelim study group, POUT study group. Conveying equipoise during recruitment for clinical trials: qualitative synthesis of clinicians' practices across six randomised controlled trials. PLoS Med 2016 Oct;13(10):e1002147 [FREE Full text] [doi: 10.1371/journal.pmed.1002147] [Medline: 27755555]
87. von Ranson KM, Robinson KE. Who is providing what type of psychotherapy to eating disorder clients? A survey. Int J Eat Disord 2006 Jan;39(1):27-34. [doi: 10.1002/eat.20201] [Medline: 16231336]
88. Stewart RE, Chambless DL. Does psychotherapy research inform treatment decisions in private practice? J Clin Psychol 2007 Mar;63(3):267-281. [doi: 10.1002/jclp.20347] [Medline: 17211876]
89. Stewart RE, Chambless DL, Stirman SW. Decision making and the use of evidence based practice: is the three-legged stool balanced? Pract Innov (Wash D C) 2018 Mar;3(1):56-67 [FREE Full text] [doi: 10.1037/pri0000063] [Medline: 32219174]
90. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
91. Gadolin C, Andersson T. Healthcare quality improvement work: a professional employee perspective. Int J Health Care Qual Assur 2017 Jun 12;30(5):410-423. [doi: 10.1108/IJHCQA-02-2016-0013] [Medline: 28574326]
92. Weiner BJ. A theory of organizational readiness for change. Implement Sci 2009 Oct 19;4:67 [FREE Full text] [doi: 10.1186/1748-5908-4-67] [Medline: 19840381]

XSL•FO
RenderX

93. Weiner BJ, Belden CM, Bergmire DM, Johnston M. The meaning and measurement of implementation climate. Implement Sci 2011 Jul 22;6:78 [FREE Full text] [doi: 10.1186/1748-5908-6-78] [Medline: 21781328]
94. Klein KJ, Sorra JS. The challenge of innovation implementation. Acad Manag Rev 1996 Oct;21(4):1055-1080. [doi: 10.5465/amr.1996.9704071863]
95. Milat AJ, Bauman A, Redman S. Narrative review of models and success factors for scaling up public health interventions. Implement Sci 2015 Aug 12;10:113 [FREE Full text] [doi: 10.1186/s13012-015-0301-6] [Medline: 26264351]
96. Chambers D, Cantrell A, Booth A. Factors that facilitate the implementation of interventions to reduce preventable hospital admissions with a focus on cardiovascular or respiratory conditions: an evidence map and realist synthesis. National Institute for Health and Care Research. URL: https://njl-admin.nihr.ac.uk/document/download/2033157 [accessed 2019-12-05]
97. Pihlaja S, Lahti J, Lipsanen JO, Ritola V, Gummerus EM, Stenberg JH, et al. Scheduled telephone support for internet cognitive behavioral therapy for depression in patients at risk for dropout: pragmatic randomized controlled trial. J Med Internet Res 2020 Jul 23;22(7):e15732 [FREE Full text] [doi: 10.2196/15732] [Medline: 32706658]
98. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
99. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, et al. Integrating qualitative research with trials in systematic reviews. BMJ 2004 Apr 24;328(7446):1010-1012 [FREE Full text] [doi: 10.1136/bmj.328.7446.1010] [Medline: 15105329]
100. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration. 2011 Sep 06. URL: https://handbook-5-1.cochrane.org/ [accessed 2024-04-29]
101. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
iCBT: internet-delivered cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 05.04.23; peer-reviewed by L Yang, J Abbas, A AL-Asadi, M Attridge; comments to author 24.02.24; revised version received 28.06.24; accepted 06.08.24; published 28.01.25.

Please cite as:
Duffy D, Richards D, Hisler G, Timulak L
Implementing Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety in Adults: Systematic Review
J Med Internet Res 2025;27:e47927
URL: https://www.jmir.org/2025/1/e47927
doi:10.2196/47927
PMID:

©Daniel Duffy, Derek Richards, Garrett Hisler, Ladislav Timulak. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Information Exchange Between the Public and Researchers in Health Studies: Scoping Review

Nazli Soltani[1], MSc; Thilo Dietz[2,3], PhD; Doris Ochterbeck[1], PhD; Jens Dierkes[3], PhD; Katja Restel[3], MA; Lara Christianson[1], MLS; Karina Karolina De Santis[1], PhD; Hajo Zeeb[1,4], MD, PhD, Prof Dr

[1]Department of Prevention and Evaluation, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

[2]Institute of Medical Sociology, Health Services Research, and Rehabilitation Science (IMVR), University of Cologne, Cologne, Germany

[3]Department of Research and Publication Support, University and City Library, University of Cologne, Cologne, Germany

[4]Faculty of Human and Health Sciences, University of Bremen, Bremen, Germany

Corresponding Author:
Karina Karolina De Santis, PhD
Department of Prevention and Evaluation
Leibniz Institute for Prevention Research and Epidemiology – BIPS
Achterstrasse 30
Bremen, 28359
Germany
Phone: 49 42121856908
Email: desantis@leibniz-bips.de

Abstract

Background: Information exchange regarding the scope and content of health studies is becoming increasingly important. Digital methods, including study websites, can facilitate such an exchange.

Objective: This scoping review aimed to describe how digital information exchange occurs between the public and researchers in health studies.

Methods: This scoping review was prospectively registered and adheres to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Eligibility was defined using the population (public and researchers), concept (digital information exchange), and context (health studies) framework. Bibliographic databases (MEDLINE, PsycINFO, CINAHL, and Web of Science), bibliographies of the included studies, and Google Scholar were searched up to February 2024. Studies published in peer-reviewed journals were screened for inclusion based on the title, abstract, and full text. Data items charted from studies included bibliographic and PCC (Population, Concept, and Context) characteristics. Data were processed into categories that inductively emerged from the data and were synthesized into main themes using descriptive statistics.

Results: Overall, 4072 records were screened, and 18 studies published between 2010 and 2021 were included. All studies evaluated or assessed the preferences for digital information exchange. The target populations included the public (mainly adults with any or specific diseases), researchers, or both. The digital information exchange methods included websites, emails, forums, platforms, social media, and portals. Interactivity (ie, if digital information exchange is or should be active or passive) was addressed in half of the studies. Exchange content included health information or data with the aim to inform, recruit, link, or gather innovative research ideas from participants in health studies. We identified 7 facilitators and 9 barriers to digital information exchange. The main facilitators were the consideration of any stakeholder perspectives and needs to clarify expectations and responsibilities, the use of modern or low-cost communication technologies and public-oriented language, and continuous communication of the health study process. The main barriers were that information exchange was not planned or not feasible due to inadequate resources, highly complex technical language was used, and ethical concerns (eg, breach of anonymity if study participants are brought together) were raised. Evidence gaps indicate that new studies should assess the methods and the receiver (ie, public) preferences and needs that are required to deliver and facilitate interactive digital information exchange.

Conclusions: Few studies addressing digital information exchange in health studies could be identified in this review. There was little focus on interactivity in such an exchange. Digital information exchange was associated with more barriers than facilitators, suggesting that more effort is required to improve such an exchange between the public and researchers. Future studies should investigate interactive digital methods and the receiver preferences and needs required for such an exchange.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63373) doi:10.2196/63373

KEYWORDS

health information; information exchange; communication; knowledge translation; dissemination; digital technology; research participant; scoping review

Introduction

Public Engagement With Health Research

Public engagement with research can improve the visibility, quality, and impact of such research. The term "public engagement" is used to describe the various ways in which the public (ie, the general population) can be involved in the design, conduct, and dissemination of research. It may be employed to inform and inspire, to advise and listen, or to collaborate with the public. Engagement is by definition a 2-way process involving interaction and listening and intending to generate mutual benefit [1]. A conceptual review of 142 peer-reviewed articles published between 1996 and 2019 showed that the term "public engagement" converges with the term "public involvement." This is supported by the increasing use of the combination term "patient and public involvement and engagement" [2].

In the context of health research with human participants, the public (ie, the general population) should have a say on how to safely participate in studies and how the data collected in such studies is used [3]. Integrating public engagement in the whole research process includes sharing research with nonacademic audiences by contributing to dissemination plans and materials for the public and supporting the use of research in practice [4]. It has been suggested that enhanced involvement of the public (eg, study participants or other interested laypeople) can improve executive procedures such as recruitment and data collection [5,6]. For clinical trials, information exchange, including bidirectional sharing of information, such as data and feedback, between researchers and the public creates a more satisfying trial experience for all involved (ie, study participants and researchers) [7]. The dissemination of study results is defined as one of the key aspects of patient and public involvement and engagement in health research, where patients should provide input on the communication strategy of the study and be co-authors of study publications [8]. A systematic review on youth involvement in health research identified several benefits of such involvement, including that it provides unique perspectives that help to improve research questions and priorities and that it leads to improved translation, dissemination, and uptake of research findings [9].

Furthermore, incorporating the perspectives of the public in research right from the beginning can help researchers to identify and focus on issues relevant to the target group. This might not only improve the proximity of research to practice and thus enhance its translation, but also increase the acceptance and willingness to participate in health studies [1,8]. A mixed-methods study using quantitative surveys and qualitative interviews with researchers and patients assessed patient experiences and perceptions of engagement in health studies [10]. Survey results identified key factors influencing engagement, such as communication quality and participant expectations, highlighting areas for improved collaboration. Interviews provided deeper insights into patient experiences, emphasizing the importance of mutual respect, transparency, and addressing barriers, including unclear roles or logistical challenges of communication. These findings suggest that well-structured frameworks are needed to plan meaningful patient involvement that is required to enhance the effectiveness of patient-researcher partnerships [10].

Digital Information Exchange

Owing to the great advancements in technology during the past decades and the increased accessibility of technological tools and devices, health information exchange is now possible with less effort and expense than before. It can thus help transform the role of study participants from passive observers and subjects in research studies to active members of study teams, who can have an impact on the conduct and outcome of a research study. For example, citizen scientists (participants in a study on arthritis) recorded their daily pain intensity in a smartphone app [11]. The GPS location of their phones was then linked to local weather data to look for possible associations between weather conditions and pain intensity. It was thus possible to collect large datasets and develop a tool to assist patients in better managing their respective health condition [11]. However, there are various barriers associated with the use of digital methods for health information exchange, including poor digital health literacy, ethical and legal issues, low trust in online information sources, and poor technological affinity. For example, in the context of COVID-19, it was difficult for the public to obtain appropriate and accurate online information due to uncertainty about the disease and a plethora of inconsistent content [12]. When it comes to the dissemination of research results, the complexity of the research process and the use of appropriate plain rather than technical language are problematic. Well-designed websites built in association with medical professionals, which have a clear interface and contain quality health information, can draw the attention of patients and lead them to access trustworthy information [13]. Having broad access to additional health information over the internet, which patients can use for discussions with health care professionals, can lead to more patient involvement and engagement in medical decision-making [14].

To develop and establish participation and information exchange opportunities using digital methods, it is vital to explore what is available and, in particular, if these methods are accepted by the target group, that is, the public, including study participants and any laypeople. The goal of such participatory research is to assess the wants and needs of all stakeholders involved in health research.

Digital Portals for Health Information Exchange

The lexical definition of a portal describes a door, passageway, or entrance area that allows entry into a large or grand building,

XSL•FO
RenderX

thus providing access to a new separate area [15]. This meaning can also be applied to digital infrastructure. In this context, an internet-based digital portal aggregates information from numerous sources and makes it accessible to a diverse group of users [16]. Thus, digital portals offer an opportunity for digital information exchange in health research.

The early literature distinguishes between different types of health portals [17] that provide access to a range of functions [18]. The first developments of patient portals for access to personal health information can be traced back to the 1990s [19-21]; however, their more widespread use started later in about 2006 [22]. The definitions of the term "patient portal" and its functions vary. A patient portal refers to an application provided by a health care institution (eg, a hospital) that allows patients to access general health information [23]. However, a patient portal could also provide secure access to the patient's electronic health record (EHR) [20,24-26]. Studies indicate the benefits of patient portals, as they can enhance doctor-patient communication [27] and link patient information with other information on the internet in the case of digital portals [23]. Further studies indicate other benefits in terms of active patient participation and increased satisfaction and treatment adherence [28-31]. The use of a patient portal depends on the expected benefit to the user and the cognitive resources that need to be invested to use it. On average, users are younger, chronically ill, and able to understand digital health information compared to patients who do not use such portals but are aware of their existence [32].

The onset of the COVID-19 pandemic was associated with the increased availability of "open health data." This type of data can originate from public authorities (eg, health authorities), clinical records (eg, EHR), or public health–related surveys [33]. The increasing availability of these data has been accompanied by the development and organization of digital networks and portals for data exchange. These digital portals allow, in part, unrestricted, free, and unlimited access to open health data in machine-readable format for all interested parties [33]. While patients may be interested in the development and progression of a disease, authorities, governments, and medical institutions may need information on the spread of diseases [34]. To meet these demands, there is an increasing need for sophisticated and user-oriented digital portals with health data [33].

According to nonsystematic searches of the internet via Google, we identified 5 examples of digital portals with health data (Table 1). These digital portals are available in German, English, or multiple languages. The target population of the portals included either the public or researchers. Overall, 4 portals focused on data exchange, while 1 portal allowed patients to search for health information using a filter option.

Table 1. Examples of digital health portals.

Portal name	Targeted population	Portal aim	Portal focus	Search terms used to locate the portal in Google
Gesund.bund.de: Verlässliche Informationen für Ihre Gesundheit [35]	Patients, family members, care takers, and the public (interested parties)	Detailed search for information on diseases and the provision of assistance in the event of a disease	Information exchange	"Patientenportal" OR "patient information portal"
European Health Information Portal [36]	Researchers	Search for existing study records; data exchange	Data exchange	"Patientenportal" OR "patient information portal"
Health Data Lab [37]	Researchers and legislators	Access and exchange of health and care data	Data exchange	"health study data" OR "health data portal"
HealthData.gov [38]	Researchers, entrepreneurs, and legislators	Identification of trial and health data	Data exchange	"health study data" OR "health data portal"
Add Health – The National Longitudinal Study of Adolescent to Adult Health/Add Health Navigator [38,39]	Public (participants and interested parties) and researchers	Access study data by searching for records using the "Add Health Navigator"; study-specific portal for data sharing and exchange	Data exchange	"health study data" OR "health data portal"

NFDI4Health (The National Research Data Infrastructure for Personal Health Data) Initiative in Germany

The National Research Data Infrastructure for Personal Health Data (NFDI4Health) is one example of a nationwide digital health portal established in Germany that has the overarching goal of providing new opportunities for the scientific use of personal health data while respecting privacy requirements. The central element of NFDI4Health is the operationalization of the FAIR (findable, accessible, interoperable, reusable) principles [40] around health data, especially those from epidemiological and clinical trials, to promote their reuse and strengthen the reputation of collaborating researchers. One important objective of NFDI4Health is the transfer of research results and study knowledge to the public, which should strengthen the interaction among the public, researchers, and health research institutions [41]. As part of this objective, a digital portal called Research Dialogue [42] has been developed to provide the public with online access to study information, results, and aggregated data from research projects listed in NFDI4Health. In the long term, this digital portal should enable an interactive (ie, bidirectional) exchange between the public and researchers to improve the understanding of health data and health research, ensure a low-threshold exchange of information, and promote participation in clinical studies in Germany. This study was designed to identify digital evidence-based methods from other studies that could be used in the development of the Research Dialogue portal for digital information exchange between the public and researchers.

XSL•FO
RenderX

Study Aims and Objectives

This study aimed to describe how digital information exchange occurs between the public and researchers in health studies, using a scoping review methodology. A scoping review is particularly applicable when a body of scientific evidence needs to be described and mapped into categories without a focus on health outcomes. This review has the following 6 objectives:

1. Studies: What are the characteristics of assessments on digital information exchange between the public and researchers in health studies?
2. Population: What populations were addressed in these studies?
3. Concept: What methods of digital information exchange were described in these studies?
4. Context: What content and aim of digital information exchange were addressed in these studies?
5. Factors associated with digital information exchange: What facilitators and barriers associated with digital information exchange were identified in these studies?
6. Evidence gaps: What evidence gaps and ideas for future research were identified in these studies?

Methods

Study Design

This scoping review adheres to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [43]. The PRISMA-ScR checklist is presented in Multimedia Appendix 1.

Protocol and Registration

A protocol for this review was prospectively registered [44]. There were no changes between the protocol and the content of this review. At the time of protocol registration, preliminary searches were conducted to design the search strategy.

Eligibility Criteria

The eligibility for this scoping review was based on the PCC (Population, Concept, and Context) framework (Textbox 1). The 3 components of this framework were defined as follows. “Population” includes the public that can include patients, study participants, or any laypeople (ie, the general population). This group is considered as the primary receivers of health information. “Population” can also include researchers (ie, any experts, such as academics, health care professionals, or other relevant stakeholders) who are primary generators and providers of health information (ie, they exchange or plan to exchange health information with the public). “Concept” includes the actual or planned health information exchange using any digital methods (eg, a website with health study content that can be accessed by the population). “Context” includes the health information from specific or any health studies.

XSL•FO
RenderX

Textbox 1. Eligibility criteria.

Inclusion criteria

- Population: Public (ie, laypeople, general population, study participants, and patients) of any age and with any health status (ie, healthy, at risk for any disease, or with any disease) or researchers (eg, any experts, such as academics, health care professionals, or any relevant stakeholders) who exchange or plan to exchange information with the public
- Concept: Digital health information exchange or need for health information using any digital methods (eg, websites accessed using any devices, such as computers or smartphones)
- Context: Health study; (1) Focus on information exchange in a specific health study; (2) Focus on public or researcher perspectives on information exchange or need for health information in any health study
- Study design: Primary study with any design (eg, observational study or survey) or data type (quantitative or qualitative); reports on digital portals with health content
- Study type: Published in a peer-reviewed journal
- Language: English or German
- Access: Full text available

Exclusion criteria

- Population: (1) No focus on the public (eg, focus on information exchange among researchers, health care professionals, politicians, stakeholders, students, or other population groups); (2) Nonhuman populations (eg, veterinary health; meta-research studies on published articles or research methods)
- Concept: (1) Digital health information exchange in health care settings (eg, for clinical data collection or recording); (2) Nondigital health information exchange (eg, information exchange on paper or in face-to-face settings)
- Context: (1) No focus on health studies; (2) Focus on specific topics in the health context, including clinical treatment, online support groups, individual clinical data, electronic health records, and genome data; development of health systems, health frameworks, digital health technologies, clinical guidelines, procedures, or tools; marketing and sales of health products; ethical and legal issues in consent and data sharing; teaching, training, and health education for health care professionals; participant recruitment; health literacy; and quality of health information
- Study design: Nonprimary study (eg, literature review, comment, editorial, correction, or study protocol)
- Study type: Other study types (eg, conference papers, dissertations, or books)
- Language: Language other than English or German
- Access: Full text not available

Information Sources

The information sources for this scoping review include 4 international bibliographic databases (MEDLINE, PsycINFO, CINAHL, and Web of Science), bibliographies of any included studies, and the internet (eg, Google Scholar).

Search Strategy

The search syntax was developed by the team (NS and KKDS) with support from an experienced information specialist (LC). The search terms addressed the PCC criteria (Textbox 1) and included relevant synonyms and subject terms (Textbox 2). The search, using Boolean connectors, adjacency operators, and truncation, was designed in MEDLINE and adapted to other databases (Multimedia Appendix 2). The keyword search was limited to the title and abstract fields, and there were no language or other limits applied.

XSL•FO
RenderX

Textbox 2. Example search terms used in database searches. Population, concept, and context terms were combined with the Boolean operator "AND" in the search.

Population: public

- Search terms: patient, subject, public, citizen, participant, lay, user, general population
- Subject terms: patients, "research subjects"

Concept: digital information exchange

- Search terms: digital or online health education, digital or online health information, digital or online health knowledge, health education or health information or health knowledge portal, health education or health information or health knowledge platform, health education or health information or health knowledge website, digital information exchange or communication or dissemination, online information exchange or communication or dissemination, information portal or information website or information platform access
- Subject terms: "health information exchange," "consumer health information," "information seeking behavior," "information dissemination"

Context: health study

- Search terms: health research, health study, health trial, medical research, medical study, medical trial, epidemiological research, epidemiological study, epidemiological trial

The search was performed (by LC) from database inception to February 1, 2024. The search documentation is shown in Multimedia Appendix 2. All search results were exported to the Systematic Review Accelerator Deduplicator (Bond University Institute for Evidence-Based Healthcare) for deduplication and EndNote 20 (Clarivate) for management. Manual searches of bibliographies of the included studies and the internet via Google Scholar were performed (by NS, JD, and KR) up to February 2024. All search results were exported and managed in EndNote 20.

Study Selection

Screening was performed in EndNote by 3 researchers (NS, KKDS, and a team assistant). Overall, 50% of all studies were screened by at least 2 researchers. As recommended by the PRISMA-ScR guidelines [43], the screening procedure was discussed in the team to increase the consistency among the screeners. The eligibility criteria (Textbox 1) were pilot tested on randomly selected studies (at least one study per criterion) until consensus was reached. Title and abstract screening was supported by the *smartgroups* function in EndNote that was used to identify studies for exclusion (eg, reviews, books, or conference papers). Two researchers (KKDS and a team assistant) checked and confirmed the exclusion of such studies identified by EndNote. Screening of all titles and abstracts was performed by 1 researcher (KKDS) and checked by another researcher (NS or a team assistant). Screening of full-text articles was performed by 2 researchers (NS and KKDS) independently, and final consensus was reached by discussion within the team (NS, KKDS, and a team assistant). Studies were included if they addressed any of the objectives of the review and if they provided details of any digital exchange methods or the intention to exchange the research findings with the public.

Data Charting

A data charting form was developed and calibrated within the team. Data were charted by 2 researchers (NS and TD) who extracted author statements from the included studies and agreed on the final statement selection by discussion. The extracted data were processed into meaningful categories by 2 researchers (NS and TD), and a final consensus was reached by discussion with a third researcher (KKDS). Data processing was subsequently discussed within the team (NS, TD, and KKDS), and a consensus was reached by discussion.

Extracted data included quantitative information (eg, publication year) and qualitative information (eg, description of digital information exchange methods) per study. In scoping reviews, qualitative information is organized by categorizing data into main themes to provide structured insights [43]. In this scoping review, 2 researchers (NS and TD) read all extracted data and then classified them into meaningful categories that were inductively identified in the data. For example, we classified the methods of digital information exchange as "active" or "passive" based on the description of digital exchange extracted from each individual study. The categories were discussed within the team (NS, TD, and KKDS) and calibrated by discussion until a consensus was reached. The final categories were entered into the data charting form for each study for further analysis. The main themes were identified in the process of data synthesis.

Data Items

A list of data items (Table 2) was developed by 1 researcher (NS) and discussed within the team (NS, TD, and KKDS).

XSL•FO
RenderX

Table 2. Data items based on the objectives of this scoping review.

Objective and data item	Data item content or definition
Objective 1 (Studies: design and aim)	
1. Bibliographic characteristics	First author, title, publication year, corresponding author continent, funding source
2. Study characteristics	Design (ie, primary study or report), aim (eg, evaluation or preferences for digital information)
Objective 2 (population)	
3. Population characteristics	Type: Public (ie, laypeople, general population, study participants, and patients) who are primary receivers of digital health information or researchers (eg, any experts, such as academics, health care professionals, or any relevant stakeholders) who are primary providers of digital health information, age, location (continent), health focus (eg, specific disease or any disease)
Objective 3 (concept: digital information exchange)	
4. Modality	Modality according to study authors (eg, website or other)
5. Interactivity	Interactivity implicitly or explicitly described by study authors (ie, passive exchange or interactive exchange)
Objective 4 (context: health study)	
6. Digital exchange content	Content (eg, data exchange, information exchange, or both); if data exchange, type of exchanged data (eg, aggregated study data or individual patient data, or both)
7. Digital exchange aim	Aim (eg, to inform the public about the health study), format (eg, video or other)
Objective 5 (factors associated with digital information exchange)	
8. Facilitators	Factors that could facilitate digital information exchange
9. Barriers	Factors that could hinder digital information exchange
Objective 6 (evidence gaps)	
10. Evidence gaps	Evidence gaps and ideas for future research with a focus on information exchange addressed in studies (eg, conclusion)

Critical Appraisal

Critical appraisal was not performed because this scoping review aimed to describe central topics and questions related to digital information exchange rather than to evaluate the effectiveness of such exchange.

Data Synthesis

The data processed into categories in each study were synthesized for all 18 studies using absolute and relative frequencies or means and SDs in Excel (version 10; Microsoft Corp) to address the 6 objectives of this scoping review. For example, 2 researchers (NS and TD) listed all categories relevant for objective 3 (methods of digital information exchange), counted all studies out of 18 that were assigned into each category, and noted their citations. The main themes in the data were identified by 3 researchers (NS, TD, and KKDS) by discussion. We read all individual categories assigned to each objective and clustered them based on their content. For example, we clustered the categories “active” or “passive” digital information exchange into a main theme “interactivity” as part of the methods of digital information exchange. For objectives 5 (factors associated with digital information exchange) and 6 (evidence gaps), we clustered the categories from individual studies into 3 main themes (ie, facilitators, barriers, and evidence gaps) and reported example author statements to further explain the meaning of each theme. These 3 themes were either explicitly mentioned by the study authors or were derived based on our interpretation of categories assigned to individual studies.

Results

Study Selection

The electronic and manual searches identified 5218 records (Figure 1). Following the removal of duplicates, 4072 records were screened based on the title and abstract and 48 records were screened based on the full text. Overall, 492 of the 4072 records were selected for exclusion using EndNote (ie, nonprimary studies, including reviews and study protocols; other study types, including conference papers, dissertations, and books; and retracted articles or duplicates that were not detected by Deduplicator). Further, 3532 of the 4072 records were excluded as they did not fulfill at least 1 eligibility criterion (Textbox 1). The specific reasons for exclusion of these records were not documented, as suggested by the PRISMA-ScR guidelines [43]. Full-text screening identified 18 of 48 studies [7,45-61] that met the inclusion criteria and were included in this scoping review. The list of the 30 excluded studies with individual reasons for exclusion is presented in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram.

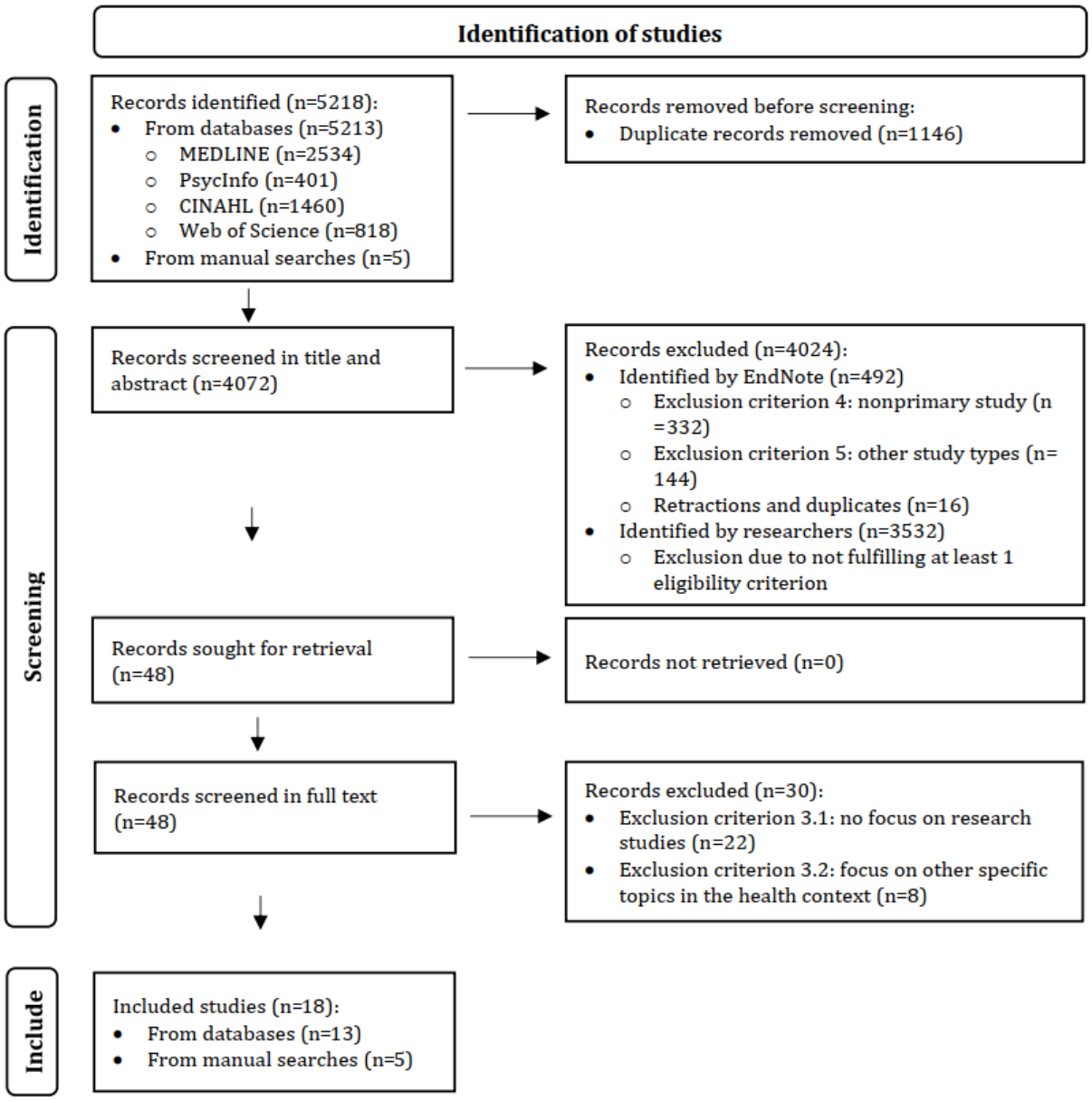

Bibliographic Characteristics of the Included Studies

All charted and processed data from the 18 studies are reported in Multimedia Appendix 4. The 18 studies were published between 2010 and 2021 (Figure 2). The studies originated from Europe (9/18, 50%), North America (7/18, 39%), and Australia (2/18, 11%). Funding source or a lack of funding was reported in 17 (94%) studies.

XSL•FO
RenderX

Figure 2. Characteristics of the 18 included studies.

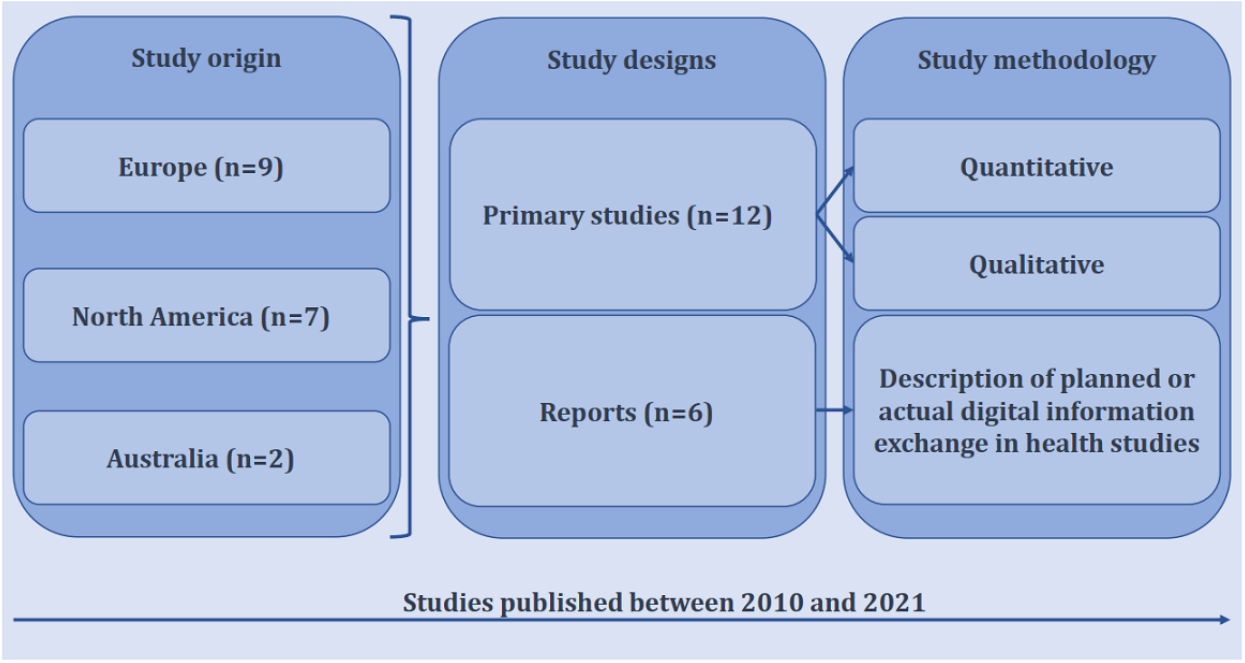

Objective 1: Study Designs and Aims

The 18 studies included 12 (67%) primary studies [7,45,46,48,50,51,53,54,57,59-61] with quantitative or qualitative data and 6 (33%) reports [47,49,52,55,56,58], which described the actual or planned digital information exchange in health studies (Figure 2). Overall, 12 (67%) studies [45-47,49,54-61] aimed to evaluate the existing strategies for digital information exchange. The remaining 6 (33%) studies [7,48,50-53] assessed preferences for such information exchange.

Objective 2: Population Characteristics

The PCC characteristics of the included studies are summarized in Figure 3.

Figure 3. Characteristics of the 18 included studies based on objectives 2-6 in the scoping review.

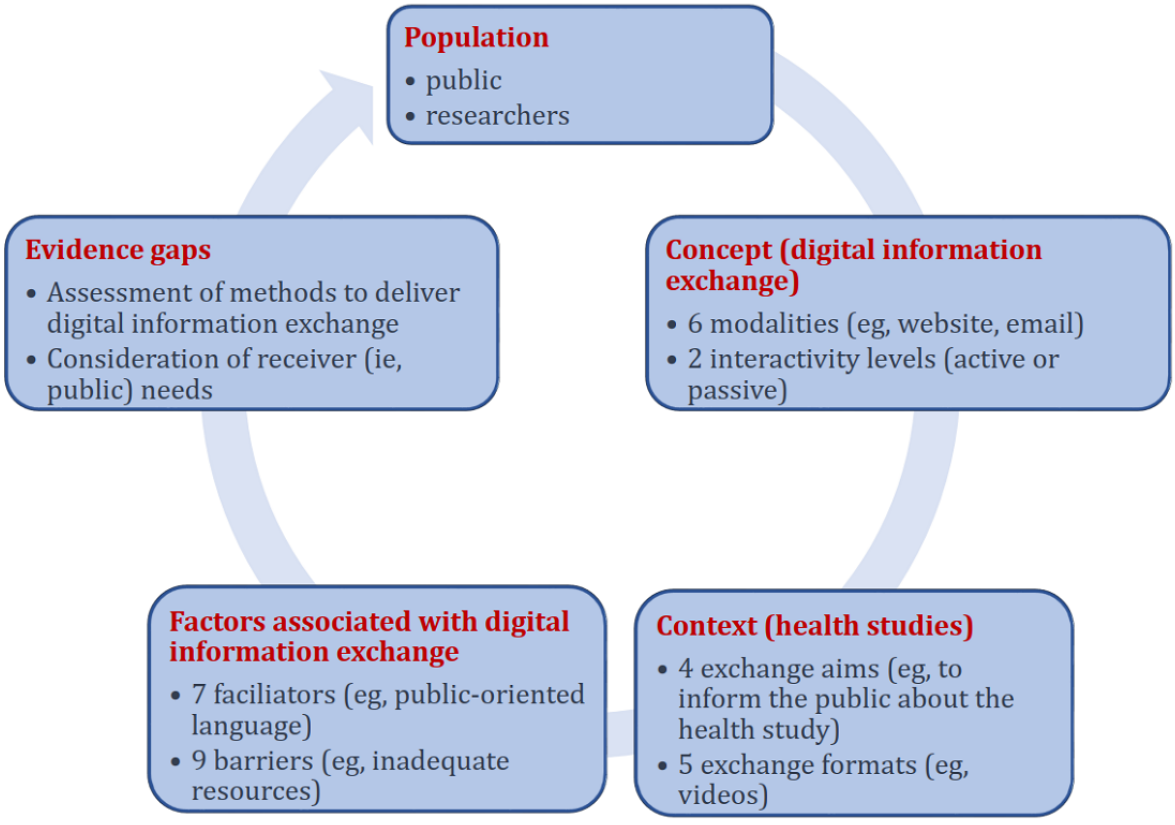

XSL•FO
RenderX

All 18 studies addressed digital information exchange for 1 to 2 (mean 1.3, SD 0.5) population types. The addressed populations included the public in 10 (56%) studies [45-47,50,51,54-57,59], the researchers in 2 (11%) [52,53], and both the public and researchers in 6 (33%) [7,48,49,58,60,61].

Among the 18 studies, 9 (50%) [7,46,47,49,54,56,58,59,61] focused on populations in Europe, 4 (22%) [7,51,53,55] focused on populations in North America, and 1 (6%) each focused on populations in Asia [50], Australia [48], or Africa [7]. Overall, 4 (22%) studies [45,52,57,60] did not state the region of the assessed population.

Moreover, among the 18 studies, 7 (39%) [48,50,51,59,60] addressed only adults or adolescents (ie, people aged 18 years or 16 years or older) and 3 (17%) [47,54,55] addressed people of any age (ie, children and adults). In 8 (44%) studies [45,46,49,52,56-58,61], the age of the addressed population was not mentioned.

Most of the 18 studies focused on clinical populations. Overall, 8 (44%) studies [7,47,49,51,53-56] addressed any diseases, 5 (28%) [46,48,57,58,60] addressed specific diseases (eg, rare diseases or chronic kidney disease), and 3 (17%) [45,59,61] addressed any cancer diagnoses. In 2 (11%) studies [50,52], the health focus was not mentioned.

Objective 3: Concept (Digital Information Exchange Methods)

Among the 18 studies, 13 (72%) [45-47,49,51,54-61] described 1 to 6 (mean 1.6, SD 0.7) exchange modalities. All 13 studies described methods of digital information exchange using diverse terminology. According to this terminology, digital information exchange occurred through websites, emails, forums, platforms, social media, or portals (Table 3). Some of those studies (6/13, 46%) [47,54-56,58,61] described multiple modalities of digital information exchange.

Table 3. Digital information exchange methods.

Variable	Studies (N=18), n (%)	Study citation
Modality		
Website	10 (56)	[45,47,49,54-59,61]
Email	3 (17)	[47,51,54]
Forum	3 (17)	[47,55,58]
Platform	2 (11)	[46,60]
Social media	2 (11)	[54,56]
Portal	1 (6)	[61]
Not mentioned	5 (28)	[7,48,50,52,53]
Interactivity		
Active (eg, bidirectional communication via a website with a discussion forum)	4 (22)	[46,56,58,61]
Passive (eg, 1-directional dissemination of information though a website)	3 (17)	[47,55,60]
Any (active or passive)	2 (11)	[45,54]
Not mentioned	9 (50)	[7,48-53,57,59]

Interactivity in the method of digital information exchange was implicitly or explicitly addressed in 9 (50%) studies and not addressed in the remaining 9 (50%) studies. We classified the interactivity of digital information exchange as active, passive, or any (Table 3).

Objective 4: Context (Content and Aim of Digital Information Exchange)

Among the 18 studies, 11 (61%) [45-47,49,55-57,59-61] mentioned exchange content (ie, data or information exchange regarding the health studies) and 7 (39%) [7,48,50-54] did not mention exchange content. Studies that addressed data exchange (3/11, 27%) [47,55,59] mentioned a digital exchange of individual and aggregated study data (1/3, 33%) [47] or only aggregated study data (2/3, 67%) [55,59].

Among the 11 studies [45-47,49,55-61] that addressed exchange content, there were 1 to 3 (mean 1.7, SD 0.7) exchange aims. The exchange aims focused on four themes: (1) to inform the public about the health study, (2) to recruit participants for the health study, (3) to link study participants with each other, and (4) to gather innovative research ideas from study participants (Table 4).

XSL•FO
RenderX

Table 4. Aim and format of digital information exchange.

Variable	Studies (N=18), n (%)	Study citation
Exchange aim		
To inform the public about the health study	10 (56)	[45,47,49,55-61]
To recruit participants for the health study	4 (22)	[49,58,60,61]
To link study participants with each other	3 (17)	[46,47,61]
To gather innovative research ideas from study participants	2 (11)	[46,58]
Not mentioned	7 (39)	[7,48,50-54]
Format		
Video	6 (33)	[46,47,55,56,58,61]
Audio	3 (17)	[46,58,61]
Picture	3 (17)	[46,47,58]
Infographic	1 (6)	[55]
Plain language summary	1 (6)	[58]
Not mentioned	12 (67)	[7,45,48-50,52-54,57,59,60]

Within these 11 studies that addressed exchange content, 6 (33%) [46,47,55,56,58,61] mentioned the format of exchange. Those 6 studies mentioned 1 to 4 (mean 2.3, SD 0.9) exchange formats that included videos, audio data, pictures, infographics, or plain language summaries (PLSs) (Table 4).

Objective 5: Factors Associated With Digital Information Exchange

Among the 18 studies, we identified 7 facilitators of digital information exchange in 17 (94%) [7,46-61] and 9 barriers of digital information exchange in 16 (89%) [7,45,47,48,50-61]. There were 1 to 3 (mean 1.9, SD 0.8) facilitators and 1 to 5 (mean 2.3, SD 1.2) barriers of digital information exchange per study.

Among all 7 facilitators (Table 5; Multimedia Appendix 5), the most commonly identified facilitators in 6 of the 17 studies (35%) were: (1) consideration of any stakeholder perspectives and needs (by clarifying expectations and responsibilities), (2) use of modern or low-cost communication technologies, (3) use of public-oriented language, and (4) continuous communication of the health study process. Further facilitators identified in less than 6 of the 17 studies were: (1) appropriate dissemination of study results to participants, (2) appropriate amount of information content, and (3) interactive co-design of communication.

XSL•FO
RenderX

Table 5. Factors associated with digital information exchange.

Factors	Studies (N=18), n (%)	Study citation
Facilitators of digital information exchange		
Consideration of any stakeholder perspectives and needs to clarify the expectations and responsibilities	6 (33)	[46,48,55,56,58,59]
Use of modern or low-cost communication technologies	6 (33)	[7,47,51,52,55,60]
Use of public-oriented language	6 (33)	[47,50,54,55,57,61]
Continuous communication of the health study process	6 (33)	[7,48,49,53,57,58]
Appropriate dissemination of study results to participants	4 (22)	[7,50,51,53]
Appropriate amount of information content	3 (17)	[47,54,61]
Interactive co-design of communication	2 (11)	[54,58]
Barriers of digital information exchange		
Information exchange not planned in study design or inadequate resources for planned information exchange	9 (50)	[7,48,50-55,61]
Too complex technical language for the public	6 (33)	[45,47,54,56,60,61]
Ethical concerns (eg, breach of anonymity if study participants are brought together)	5 (28)	[51-54,59]
Information exchange not wished from the public (eg, due to anticipated burden of potentially worrying research results, such as the detection of serious clinical symptoms)	4 (22)	[51,52,54,59]
Lack of interactive communication features	4 (22)	[45,54,58,61]
Low trust in digital health information	3 (17)	[58-60]
Poor health literacy among the public	2 (11)	[53,56]
Difficulties with finding or accessing digital health information	2 (11)	[57,60]
No possibility to contact the researchers	1 (6)	[45]

Among all 9 barriers (Table 5; Multimedia Appendix 5), the most commonly identified barriers in 5 to 9 of the 16 studies were (1) information exchange not planned in study design or inadequate resources for planned information exchange, (2) too complex technical language for the public, and (3) ethical concerns. Other barriers identified in less than 5 of the 16 studies were (1) information exchange not wished from the public, (2) lack of interactive communication features, (3) low trust in digital health information, (4) poor health literacy among the public, (5) difficulties with finding or accessing digital health information, and (6) no possibility to contact the researchers.

Objective 6: Evidence Gaps

Among the 18 studies, 9 (50%) [7,45-47,51-53,57,60] identified evidence gaps and provided suggestions for future research. Within these 9 studies, there were 1 to 2 (mean 1.2, SD 0.4) evidence gaps. These evidence gaps indicate that new studies are needed to address two themes: (1) assessment of methods to deliver and facilitate interactive digital information exchange and (2) consideration of receiver (ie, public) needs in the context of digital information exchange (Table 6).

Table 6. Evidence gaps and ideas for future research.

Evidence gaps	Studies (N=18), n (%)	Example statements from the included studies	Study citation
Assessment of methods to deliver and facilitate interactive digital information exchange	7 (39)	"[…], the quality of that information is highly variable, […], few sites offer interactive features. More research is needed to determine the best way of harnessing the power of the Internet—and its progeny, social media—to communicate high-yield cancer-related clinical trial information." ([1], page 1631)	[45-47,51,53,57,60]
Consideration of receiver (ie, public) needs in the context of digital information exchange	3 (17)	"Studies […] identify participants' specific goals with respect to using the information that they receive. Those studies could help researchers and funders gauge the extent to which they should focus on […] referring participants to specific resources, etc." ([53], page 3)	[7,46,52]
No gaps identified	9 (50)	—[a]	[48-50,54-56,58,59,61]

[a]Not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

The aim of this scoping review was to identify strategies for digital information exchange between the public and researchers in health studies. Our search identified only 18 peer-reviewed studies or reports that addressed various aspects of this topic. In general, very few details regarding such digital information exchange were reported. All studies described the actual or planned digital information exchange and focused mainly on receivers of health information, who were predominantly adult and clinical populations. The modality of digital information exchange was referred to using heterogeneous terminology, including websites, platforms, or portals. Exchange content included mostly health information and seldom study or individual patient data. Despite suggesting that digital information exchange should aim to inform, recruit, link, or gather innovative research ideas from participants in health studies, only half of the included studies addressed the issue of interactivity (ie, if digital information exchange is or should be active or passive). In addition, there were more barriers than facilitators associated with digital information exchange identified in the included studies. Thus, a key finding of this scoping review is the notable gap in evidence around standardized methods for digital information exchange in health studies. It is unclear how such digital information exchange should be designed and implemented across various platforms and audiences. Since digital platforms are increasingly used to share health research, evidence-based guidance on the best practices is needed to ensure that digital information exchange is effective, engaging, and accessible to a wide range of audiences, and contributes to long-term health benefits. This review identified several facilitators and barriers impacting digital information exchange that could be considered by researchers who plan such an exchange in their health studies. Overall, new studies are needed to assess the methods and public needs required to design, implement, and facilitate interactive digital information exchange.

General Benefits of Digital Information Exchange in Health Studies

The past literature suggests that public participation and interest in clinical or medical trials has a positive impact on the visibility and influence of the trials [1,62] and that the aspect of participation has gained importance in recent decades, both scientifically and politically [8,63,64]. For example, digital information exchange between the public and researchers can enhance the inclusivity, accessibility, and acceptance of research among the public and ensure that research delivery aligns with their needs [62]. In addition, the COVID-19 pandemic has led to an increase in the development and use of digital portals with health data [33]. Digital information exchange methods like apps can facilitate communication between the public and health care providers by providing users access to health information, with consideration of challenges, such as privacy and security concerns, reliability, and accuracy of health information, through implementation to maximize their effectiveness [65].

Content and Format of Digital Information Exchange in Health Studies

The studies included in this review suggest that digital information exchange should aim to inform, recruit, link, or gather innovative research ideas from participants in health studies. While most studies described the digital exchange of health information (eg, information on study content), only 3 studies addressed digital exchange of data [47,55,59]. Such data could include individual patient data [47] or aggregated study data [55,59]. In general, sharing of study data may be problematic due to ethical and legal reasons [66]. The investigation of study participant preferences for receiving research results [51] showed a spectrum of positive and negative responses, including anxiety, anger, guilt, relief, and pleasure, associated with receiving research results. Despite some negative responses, a median of 90% of participants reported the need to receive study results [51]. However, generalizing the results of studies on participant reactions and preferences regarding receiving research results is challenging due to relatively small sample sizes in such studies and a focus on specific and sensitive issues relevant for the targeted populations. Thus, different target populations may have potentially different needs for obtaining any study results or individual data.

Several formats of digital information exchange are available due to technological advancements, including not only text-based information but also other formats identified in this review, such as videos, blogs, and graphics. Given the effectiveness of PLSs based on comprehension, understanding, and enjoyment over graphical abstracts or published abstracts [67] and the availability of various PLS templates for several decades [68], it is surprising that this form of digital information exchange was mentioned in only 1 study in this review [58]. We have shown that studies with health content, such as Cochrane reviews that have mandatory PLSs according to Cochrane guidelines, are frequently mentioned online via channels accessible by the public (eg, social media, blogs, news, Wikipedia, and others) [69]. Although it is unclear how the public uses such information, exposure to information on health studies written in a nontechnical language could positively affect digital information exchange in terms of informing the public about health research. Health studies planning the digital information exchange of health data should consider the needs of different groups of relevant stakeholders. For example, target group–specific language should be used, and a lack of such language was frequently identified as a barrier to digital information exchange in this scoping review. This addresses both the lack of consideration mentioned above and the added value of this form of communication, which has been confirmed in previous studies [57,67].

According to the findings of this review, videos were the most used digital information exchange method. As suggested by others [67], if the research findings are of public relevance, researchers might consider investing time and money into creating a video of their results, as videos make viewers feel confident and positive about the presented research. However, other formats and methods, such as communication via interactive apps with gamification [70], could also be used in digital information exchange, enabling direct feedback options.

XSL•FO
RenderX

Facilitators and Barriers of Digital Information Exchange in Health Studies

The results of this scoping review showed that websites, forums, and portals are the main methods of digital information exchange for connecting with the desired audience online. Despite the benefits of interaction as a form of participation (by or for the public) [71], only half of the identified studies provided information on the possibility of active or passive interaction (Table 3). Thus, despite the social relevance and the potential benefits, digital interaction between information providers (ie, researchers) and receivers (ie, the public) was neglected in the identified studies. Another review [72], which included 28 studies on patient participation, indicated that an average of 90% of participants were interested in their own or general study results. Another study also showed that individuals who requested study results did not receive feedback on these requests [51]. There could be several explanations for this finding as evident in the list of facilitators and barriers associated with digital information exchange identified in this review (Table 5). The main barrier to dissemination was that communication was not planned in the initial study design. Research suggests that even when community-based participatory research is conducted with public involvement as a central element, no more information is disseminated than is actually published in the form of a scientific publication [73]. One reason for this could be that study participants explicitly express no interest in receiving their own study results. This could be due to anticipated burden of potentially worrying research results, such as the detection of serious clinical symptoms [74].

Other key aspects that should be considered to improve the communication of study results and the communication of health information are adequate interaction with the information provided and an appropriate volume of such information. The latter is also a barrier to the dissemination of information and can be related to the poor health literacy among the public. Information can be burdensome when it seems too complex and not tangible [75]. The filtering function could be used to counteract potential information overload [76]. This approach suggests that interactivity rather than a passive provision of health information is necessary. Incorporating interactivity when sharing research studies with the public can increase engagement, improve understanding, and make complex information more accessible [77]. Examples of interactive communication methods are interactive graphs and maps that show data over time (eg, COVID-19 spread) to help users explore trends [78], interactive quizzes [79], and serious games related to study content that can improve health literacy or online forums where people can discuss the research findings, give feedback, and see responses [80].

This scoping review also identified several facilitators that can strengthen communication between researchers and the public as well as increase trust in the information presented online (on websites or portals). In particular, the trustworthiness of information is a relevant aspect to be addressed in future studies [81]. To improve trust in health information online, such information could be discussed with health professionals who continue to be important reference persons [82].

Evidence Gaps

Evidence gaps in the included studies suggest that appropriate methods required to deliver and facilitate interactive digital information exchange need to be identified in future research. Furthermore, the preferences and needs of the receiver (ie, public) to facilitate such digital information exchange should be investigated. Interestingly, the consideration of the preferences and needs of the target populations (ie, the public being the health information receivers) was mentioned as a requirement for future studies already several years ago [7,46,52] and was the most commonly reported facilitator of digital information exchange in this review (Table 4).

Strengths and Limitations

The main strength of this review is that it addresses a highly relevant topic on how information generated in health studies can be exchanged between the public and researchers using digital methods. In particular, we identified a detailed list of facilitators and barriers associated with digital information exchange that could be considered in health studies. The literature search was supervised and performed by an experienced information specialist, and a standardized methodological approach was used to select studies and extract, process, and synthesize the data.

The main limitation of this review is the difficulty in finding relevant studies in bibliographic databases. Identifying the most suitable terminology for our objectives was challenging because some of the less specific search terms, such as public and health studies, frequently appear in diverse contexts within the health literature. Despite numerous iterations, our search syntax identified only 18 studies published up to 2021 that met the inclusion criteria. It cannot be ruled out that we missed other published studies that did not include our search terms in their titles, abstracts, or other searched fields. However, the peer-reviewed literature on this topic might be truly limited, as it is possible that researchers do not report how they share research results with the public in their studies even if they do so. Thus, we would like to encourage researchers working on health studies to comment on and share their experiences regarding digital information exchange from such studies in academic articles. Other research designs, such as comprehensive assessments of existing online data sharing databases and portals and qualitative interviews with health information providers and receivers, could be used to further investigate digital information exchange in health studies as the field evolves.

Furthermore, there was relatively little information available on the details of digital information exchange in the included studies. Despite this, we identified various factors that may help other researchers to facilitate digital information exchange and reduce the barriers in such an exchange. Although our list of such facilitators and barriers is extensive, further factors may exist that need to be identified in future research. Finally, probably due to the fact that the included studies were somewhat dated (ie, published up to 2021), we could not identify any studies discussing more recent methods of digital information exchange, such as study-related communication via apps or the use of artificial intelligence in supporting such communication.

XSL•FO
RenderX

One example of a relevant app is PIA (Prospective Monitoring and Management App) developed for monitoring, managing observational epidemiological studies, and systematically collecting user feedback [70].

Conclusions

This scoping review describes how digital information exchange occurs between the public and researchers in health studies. Our search strategy identified only 18 studies on this topic that were published mostly in Western countries (Europe and North America) between 2010 and 2021. In these studies, there was little focus on interactivity in digital information exchange, and more barriers than facilitators of such an exchange were identified. The facilitators and barriers associated with digital information exchange offer crucial insights for gaining a deeper understanding of how effective digital information exchange can be designed and implemented. Given the evidence gaps identified in the included studies, further research is needed to explore how to improve digital information exchange between the public and researchers. Future studies should investigate interactive digital methods and receiver preferences and needs required for such an exchange.

Acknowledgments

We thank Lisa Stiens (Leibniz Institute for Prevention Research and Epidemiology) for her assistance with screening. The study was performed as part of the NFDI4Health (National Research Data Infrastructure for Personal Health Data) Consortium [41]. The consortium is publicly funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) under project number 442326535.

Data Availability

The datasets generated and analyzed during this study are reported in the article and multimedia appendices.

Authors' Contributions

NS conceptualized the study; developed the methodology; selected the studies; extracted, processed, and analyzed the data; visualized the results; wrote the first draft of the manuscript; and reviewed and edited the manuscript. TD extracted, processed, and analyzed the data; visualized the results; wrote the first draft of the manuscript; and reviewed and edited the manuscript. DO performed pilot data extraction and reviewed and edited the manuscript. JD and KR conceptualized the study, developed the methodology, selected the studies, and reviewed and edited the manuscript. LC developed the methodology, performed bibliographic searches, and reviewed and edited the manuscript. KKDS developed the methodology, selected the studies, wrote the first draft of the manuscript, and reviewed and edited the manuscript. HZ conceptualized the study, developed the methodology, and reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 35 KB - jmir_v27i1e63373_app1.docx]

Multimedia Appendix 2
Search strategy.
[XLSX File (Microsoft Excel File), 30 KB - jmir_v27i1e63373_app2.xlsx]

Multimedia Appendix 3
List of included and excluded studies.
[DOCX File , 138 KB - jmir_v27i1e63373_app3.docx]

Multimedia Appendix 4
Data file.
[XLSX File (Microsoft Excel File), 74 KB - jmir_v27i1e63373_app4.xlsx]

Multimedia Appendix 5
Facilitators and barriers of digital information exchange with examples.
[DOCX File , 63 KB - jmir_v27i1e63373_app5.docx]

References

XSL•FO
RenderX

1. Blackburn S, Clinch M, de Wit M, Moser A, Primdahl J, van Vliet E, et al. Series: public engagement with research. Part 1: the fundamentals of public engagement with research. Eur J Gen Pract 2023 Dec;29(1):2232111 [FREE Full text] [doi: 10.1080/13814788.2023.2232111] [Medline: 37578421]
2. Wilkinson C, Gibson A, Biddle M, Hobbs L. Public involvement and public engagement: an example of convergent evolution? Findings from a conceptual qualitative review of patient and public involvement, and public engagement, in health and scientific research. PEC Innov 2024 Dec;4:100281 [FREE Full text] [doi: 10.1016/j.pecinn.2024.100281] [Medline: 38638421]
3. Gibson A, Britten N, Lynch J. Theoretical directions for an emancipatory concept of patient and public involvement. Health (London) 2012 Sep 25;16(5):531-547. [doi: 10.1177/1363459312438563] [Medline: 22535648]
4. Swaithes L, Campbell L, Anthierens S, Skrybant M, Schiphof D, French H, et al. Series: Public engagement with research. Part 4: Maximising the benefits of involving the public in research implementation. Eur J Gen Pract 2023 Dec;29(1):2243037 [FREE Full text] [doi: 10.1080/13814788.2023.2243037] [Medline: 37609798]
5. Hersh W, Totten A, Eden K, Devine B, Gorman P, Kassakian S, et al. Health information exchange. Evid Rep Technol Assess (Full Rep) 2015 Dec(220):1-465. [doi: 10.23970/AHRQEPCERTA220] [Medline: 30307736]
6. Sadoughi F, Nasiri S, Ahmadi H. The impact of health information exchange on healthcare quality and cost-effectiveness: A systematic literature review. Comput Methods Programs Biomed 2018 Jul;161:209-232. [doi: 10.1016/j.cmpb.2018.04.023] [Medline: 29852963]
7. Dietrich J, Alivojvodic J, Seliverstov I, Metcalf M, Jakee K. Improving information exchange with clinical trial participants: a proposal for industry. Ther Innov Regul Sci 2017 Sep 30;51(5):542-550. [doi: 10.1177/2168479017725109] [Medline: 30231690]
8. Aiyegbusi OL, McMullan C, Hughes SE, Turner GM, Subramanian A, Hotham R, et al. Considerations for patient and public involvement and engagement in health research. Nat Med 2023 Aug 20;29(8):1922-1929. [doi: 10.1038/s41591-023-02445-x] [Medline: 37474660]
9. Sanchez S, Thorburn R, Rea M, Kaufman P, Schwartz R, Selby P, et al. A systematic review of theories, models and frameworks used for youth engagement in health research. Health Expect 2024 Feb;27(1):e13975 [FREE Full text] [doi: 10.1111/hex.13975] [Medline: 39102670]
10. Babatunde S, Ahmed S, Santana MJ, Nielssen I, Zelinsky S, Ambasta A. Working together in health research: a mixed-methods patient engagement evaluation. Res Involv Engagem 2023 Aug 01;9(1):62 [FREE Full text] [doi: 10.1186/s40900-023-00475-w] [Medline: 37528438]
11. Dixon WG, Beukenhorst AL, Yimer BB, Cook L, Gasparrini A, El-Hay T, et al. How the weather affects the pain of citizen scientists using a smartphone app. NPJ Digit Med 2019 Oct 24;2(1):105 [FREE Full text] [doi: 10.1038/s41746-019-0180-3] [Medline: 31667359]
12. Jo W, Lee J, Park J, Kim Y. Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: structural topic model and network analysis. J Med Internet Res 2020 Jun 02;22(6):e19455 [FREE Full text] [doi: 10.2196/19455] [Medline: 32463367]
13. Chiu Y, Tsai C, Liang J. Laypeople's online health information search strategies and use for health-related problems: cross-sectional survey. J Med Internet Res 2022 Sep 02;24(9):e29609 [FREE Full text] [doi: 10.2196/29609] [Medline: 36053561]
14. Bussey LG, Sillence E. The role of internet resources in health decision-making: a qualitative study. Digit Health 2019 Nov 07;5:2055207619888073 [FREE Full text] [doi: 10.1177/2055207619888073] [Medline: 31741741]
15. Oxford English Dictionary. URL: https://doi.org/10.1093/OED/2685171445 [accessed 2025-01-06]
16. Tatnall A. Web Portals: The New Gateways to Internet Information and Services. Hershey, PA: Idea Group Inc; 2005.
17. Dias C. Corporate portals: a literature review of a new concept in information management. Int J Inf Manage 2001 Aug;21(4):269-287. [doi: 10.1016/s0268-4012(01)00021-4]
18. van Brakel P. Information portals: a strategy for importing external content. Electron Libr 2003;21(6):591-600 [FREE Full text] [doi: 10.1108/02640470310509153]
19. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
20. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
21. Mandl KD, Simons WW, Crawford WCR, Abbett JM. Indivo: a personally controlled health record for health information exchange and communication. BMC Med Inform Decis Mak 2007 Sep 12;7:25 [FREE Full text] [doi: 10.1186/1472-6947-7-25] [Medline: 17850667]
22. Weitzman ER, Kaci L, Mandl KD. Acceptability of a personally controlled health record in a community-based setting: implications for policy and design. J Med Internet Res 2009 Apr 29;11(2):e14 [FREE Full text] [doi: 10.2196/jmir.1187] [Medline: 19403467]
23. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]

XSL•FO
RenderX

24. Bourgeois FC, Mandl KD, Shaw D, Flemming D, Nigrin DJ. Mychildren's: integration of a personally controlled health record with a tethered patient portal for a pediatric and adolescent population. AMIA Annu Symp Proc 2009 Nov 14;2009:65-69 [FREE Full text] [Medline: 20351824]
25. Lockwood MB, Dunn-Lopez K, Pauls H, Burke L, Shah SD, Saunders MA. If you build it, they may not come: modifiable barriers to patient portal use among pre- and post-kidney transplant patients. JAMIA Open 2018 Oct;1(2):255-264 [FREE Full text] [doi: 10.1093/jamiaopen/ooy024] [Medline: 31984337]
26. Turner K, Hong Y, Yadav S, Huo J, Mainous AG. Patient portal utilization: before and after stage 2 electronic health record meaningful use. J Am Med Inform Assoc 2019 Oct 01;26(10):960-967 [FREE Full text] [doi: 10.1093/jamia/ocz030] [Medline: 30947331]
27. Ross SE, Lin C. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003;10(2):129-138 [FREE Full text] [doi: 10.1197/jamia.m1147] [Medline: 12595402]
28. Beal LL, Kolman JM, Jones SL, Khleif A, Menser T. Quantifying patient portal use: systematic review of utilization metrics. J Med Internet Res 2021 Feb 25;23(2):e23493 [FREE Full text] [doi: 10.2196/23493] [Medline: 33629962]
29. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
30. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
31. Tulu B, Trudel J, Strong DM, Johnson SA, Sundaresan D, Garber L. Patient portals: an underused resource for improving patient engagement. Chest 2016 Jan;149(1):272-277. [doi: 10.1378/chest.14-2559] [Medline: 26066707]
32. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
33. Wu D, Ma L, Zhang H. Evaluation of open health data portals for COVID-19 from the perspective of the user experience. Electron Libr 2021 Jun 10;39(2):296-317. [doi: 10.1108/el-01-2021-0011]
34. Ji X, Chun S, Geller J. Monitoring Public Health Concerns Using Twitter Sentiment Classifications. 2013 Presented at: IEEE International Conference on Healthcare Informatics; September 09-11, 2013; Philadelphia, PA. [doi: 10.1109/ichi.2013.47]
35. Verlässliche Informationen für Ihre Gesundheit. Bundesministerium für Gesundheit. URL: https://gesund.bund.de/ [accessed 2024-04-29]
36. European Health Information Portal. URL: https://www.healthinformationportal.eu/ [accessed 2024-04-29]
37. Health Data Lab. URL: https://www.healthdatalab.de/ [accessed 2024-04-29]
38. HealthData. URL: https://healthdata.gov/ [accessed 2024-04-29]
39. Add Health Navigator. URL: https://addhealth-navigator.cpc.unc.edu/ [accessed 2024-04-29]
40. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
41. NFDI4Health. URL: https://www.nfdi4health.de/en/ [accessed 2024-05-28]
42. ResearchDialogue - Shaping Health Together. NFDI4Health. URL: https://www.nfdi4health.de/en/community-eng/researchdialogue.html [accessed 2024-05-28]
43. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
44. Soltani N, Ochterbeck D, Christianson L, Restel K, Dierkes J, Zeeb H. Digital exchange of information with patient/public in health research studies. A scoping review (study protocol). Open Science Framework (OSF). URL: https://osf.io/T9BKU/ [accessed 2025-01-06]
45. Abel G, Cronin A, Earles K, Gray S. Accessibility and quality of online cancer-related clinical trial information for naïve searchers. Cancer Epidemiol Biomarkers Prev 2015 Oct;24(10):1629-1631. [doi: 10.1158/1055-9965.EPI-15-0274] [Medline: 26265204]
46. Bullinger AC, Rass M, Adamczyk S, Moeslein KM, Sohn S. Open innovation in health care: analysis of an open health platform. Health Policy 2012 May;105(2-3):165-175. [doi: 10.1016/j.healthpol.2012.02.009] [Medline: 22440194]
47. Exley K, Cano N, Aerts D, Biot P, Casteleyn L, Kolossa-Gehring M, et al. Communication in a Human biomonitoring study: Focus group work, public engagement and lessons learnt in 17 European countries. Environ Res 2015 Aug;141:31-41 [FREE Full text] [doi: 10.1016/j.envres.2014.12.003] [Medline: 25499539]
48. Gutman T, Tong A, Howell M, Dansie K, Hawley C, Craig J, et al. Principles and strategies for involving patients in research in chronic kidney disease: report from national workshops. Nephrol Dial Transplant 2020 Sep 01;35(9):1585-1594. [doi: 10.1093/ndt/gfz076] [Medline: 31093667]
49. Hägglund M, Scott Duncan T, Kai-Larsen K, Hedlin G, Krakau I. IntegrIT - Towards utilizing the Swedish national health information exchange platform for clinical research. Stud Health Technol Inform 2017;235:146-150. [Medline: 28423772]

XSL•FO
RenderX

50. Keys J, Monk J, Woolley K. Sharing results with clinical trial participants: insights from an online survey of Chinese consumers. Chin Med J (Engl) 2016 Apr 20;129(8):1007-1008 [FREE Full text] [doi: 10.4103/0366-6999.179787] [Medline: 27064050]
51. Long CR, Stewart MK, Cunningham TV, Warmack TS, McElfish PA. Health research participants' preferences for receiving research results. Clin Trials 2016 Dec 23;13(6):582-591 [FREE Full text] [doi: 10.1177/1740774516665598] [Medline: 27562368]
52. Long CR, Stewart MK, McElfish PA. Health research participants are not receiving research results: a collaborative solution is needed. Trials 2017 Oct 02;18(1):449 [FREE Full text] [doi: 10.1186/s13063-017-2200-4] [Medline: 28969665]
53. Long CR, Purvis RS, Flood-Grady E, Kimminau KS, Rhyne RL, Burge MR, et al. Health researchers' experiences, perceptions and barriers related to sharing study results with participants. Health Res Policy Syst 2019 Mar 04;17(1):25 [FREE Full text] [doi: 10.1186/s12961-019-0422-5] [Medline: 30832733]
54. Lucas PJ, Allnock D, Jessiman T. How are European birth-cohort studies engaging and consulting with young cohort members? BMC Med Res Methodol 2013 Apr 11;13:56 [FREE Full text] [doi: 10.1186/1471-2288-13-56] [Medline: 23578172]
55. Monnard K, Benjamins MR, Hirschtick JL, Castro M, Roesch PT. Co-creation of knowledge: a community-based approach to multilevel dissemination of health information. Health Promot Pract 2021 Mar 30;22(2):215-223. [doi: 10.1177/1524839919865228] [Medline: 31470741]
56. Mosconi P, Antes G, Barbareschi G, Burls A, Demotes-Mainard J, Chalmers I, et al. A European multi-language initiative to make the general population aware of independent clinical research: the European Communication on Research Awareness Need project. Trials 2016 Jan 12;17(1):19 [FREE Full text] [doi: 10.1186/s13063-015-1146-7] [Medline: 26758509]
57. Penlington M, Silverman H, Vasudevan A, Pavithran P. Plain language summaries of clinical trial results: a preliminary study to assess availability of easy-to-understand summaries and approaches to improving public engagement. Pharmaceut Med 2020 Dec 28;34(6):401-406 [FREE Full text] [doi: 10.1007/s40290-020-00359-4] [Medline: 33113147]
58. Rubinelli S, Collm A, Glässel A, Diesner F, Kinast J, Stucki G, et al. Designing interactivity on consumer health websites: PARAFORUM for spinal cord injury. Patient Educ Couns 2013 Dec;93(3):459-463 [FREE Full text] [doi: 10.1016/j.pec.2013.09.015] [Medline: 24215943]
59. Sarradon-Eck A, Sakoyan J, Desclaux A, Mancini J, Genre D, Julian-Reynier C. "They should take time": disclosure of clinical trial results as part of a social relationship. Soc Sci Med 2012 Sep;75(5):873-882. [doi: 10.1016/j.socscimed.2012.04.022] [Medline: 22658623]
60. Suart CE, Graham KJ, Suart TN, Truant R. Development of a knowledge translation platform for ataxia: Impact on readers and volunteer contributors. PLoS One 2020 Sep 1;15(9):e0238512 [FREE Full text] [doi: 10.1371/journal.pone.0238512] [Medline: 32870931]
61. Wallwiener M, Wallwiener CW, Brucker SY, Hartkopf AD, Fehm TN, Kansy JK. The Brustkrebs-Studien.de website for breast cancer patients: User acceptance of a German internet portal offering information on the disease and treatment options, and a clinical trials matching service. BMC Cancer 2010 Dec 02;10(1):663 [FREE Full text] [doi: 10.1186/1471-2407-10-663] [Medline: 21126358]
62. Mader LB, Harris T, Kläger S, Wilkinson IB, Hiemstra TF. Inverting the patient involvement paradigm: defining patient led research. Res Involv Engagem 2018;4:21 [FREE Full text] [doi: 10.1186/s40900-018-0104-4] [Medline: 30002875]
63. Geißler J, Huber S. Vom Patienten zum Akteur in der Nationalen Dekade gegen Krebs. Forum 2023 Aug 11;38(4):293-298. [doi: 10.1007/s12312-023-01227-3]
64. Prebeg M, Patton M, Desai R, Smith M, Krause K, Butcher N, et al. From participants to partners: reconceptualising authentic patient engagement roles in youth mental health research. Lancet Psychiatry 2023 Feb;10(2):139-145. [doi: 10.1016/S2215-0366(22)00377-7] [Medline: 36502816]
65. Okolo CA, Babawarun O, Arowoogun JO, Adeniyi AO, Chidi R. The role of mobile health applications in improving patient engagement and health outcomes: A critical review. Int J Sci Res Arch 2024 Feb 28;11(1):2566-2574. [doi: 10.30574/ijsra.2024.11.1.0334]
66. Thorogood A, Joly Y, Knoppers BM, Nilsson T, Metrakos P, Lazaris A, et al. An implementation framework for the feedback of individual research results and incidental findings in research. BMC Med Ethics 2014 Dec 23;15(1):88 [FREE Full text] [doi: 10.1186/1472-6939-15-88] [Medline: 25539799]
67. Bredbenner K, Simon SM. Video abstracts and plain language summaries are more effective than graphical abstracts and published abstracts. PLoS One 2019;14(11):e0224697 [FREE Full text] [doi: 10.1371/journal.pone.0224697] [Medline: 31743342]
68. Root J, Stableford S. Easy-to-read consumer communications: a missing link in Medicaid managed care. J Health Polit Policy Law 1999 Feb;24(1):1-26. [doi: 10.1215/03616878-24-1-1] [Medline: 10342253]
69. De Santis KK, Kirstein M, Kien C, Griebler U, McCrabb S, Jahnel T. Online dissemination of Cochrane reviews on digital health technologies: a cross-sectional study. Syst Rev 2024 May 15;13(1):133 [FREE Full text] [doi: 10.1186/s13643-024-02557-6] [Medline: 38750593]

XSL•FO
RenderX

70. Heise J, Dey R, Emmerich M, Kemmling Y, Sistig S, Krause G, et al. Putting digital epidemiology into practice: PIA-Prospective Monitoring and Management Application. Inform Med Unlocked 2022;30:100931. [doi: 10.1016/j.imu.2022.100931]
71. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. JAMA 2004 Oct 06;292(13):1593-1601. [doi: 10.1001/jama.292.13.1593] [Medline: 15467062]
72. Shalowitz D, Miller F. Communicating the results of clinical research to participants: attitudes, practices, and future directions. PLoS Med 2008 May 13;5(5):e91 [FREE Full text] [doi: 10.1371/journal.pmed.0050091] [Medline: 18479180]
73. Chen PG, Diaz N, Lucas G, Rosenthal MS. Dissemination of results in community-based participatory research. Am J Prev Med 2010 Oct;39(4):372-378. [doi: 10.1016/j.amepre.2010.05.021] [Medline: 20837290]
74. Miller FA, Christensen R, Giacomini M, Robert JS. Duty to disclose what? Querying the putative obligation to return research results to participants. J Med Ethics 2008 Mar;34(3):210-213. [doi: 10.1136/jme.2006.020289] [Medline: 18316466]
75. Khaleel I, Wimmer BC, Peterson GM, Zaidi STR, Roehrer E, Cummings E, et al. Health information overload among health consumers: A scoping review. Patient Educ Couns 2020 Jan;103(1):15-32. [doi: 10.1016/j.pec.2019.08.008] [Medline: 31451363]
76. Hall A, Walton G. Information overload within the health care system: a literature review. Health Info Libr J 2004 Jun 10;21(2):102-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2004.00506.x] [Medline: 15191601]
77. Mancone S, Corrado S, Tosti B, Spica G, Diotaiuti P. Integrating digital and interactive approaches in adolescent health literacy: a comprehensive review. Front Public Health 2024;12:1387874 [FREE Full text] [doi: 10.3389/fpubh.2024.1387874] [Medline: 39444982]
78. WHO COVID-19 dashboard. World Health Organization. URL: https://data.who.int/dashboards/covid19/cases [accessed 2024-11-11]
79. Engage Users. Health Literacy Online. URL: https://odphp.health.gov/healthliteracyonline/engage/full/ [accessed 2024-11-18]
80. Public Engagement Methods and Tools. Engage2020. URL: https://engage2020.eu/media/D3-2-Public-Engagement-Methods-and-Tools-3.pdf [accessed 2025-01-06]
81. Boyer C, Gaudinat A, Hanbury A, Appel RD, Ball MJ, Carpentier M, et al. Accessing reliable health information on the web: a review of the HON approach. Stud Health Technol Inform 2017;245:1004-1008. [Medline: 29295252]
82. Kwon JH, Kye S, Park EY, Oh KH, Park K. What predicts the trust of online health information? Epidemiol Health 2015;37:e2015030 [FREE Full text] [doi: 10.4178/epih/e2015030] [Medline: 26212505]

Abbreviations

EHR: electronic health record
NFDI4Health: National Research Data Infrastructure for Personal Health Data
PCC: Population, Concept, and Context
PLS: plain language summary
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by A Mavragani; submitted 18.06.24; peer-reviewed by A Patel, K Sterba; comments to author 04.11.24; revised version received 25.11.24; accepted 12.12.24; published 28.01.25.

Please cite as:
Soltani N, Dietz T, Ochterbeck D, Dierkes J, Restel K, Christianson L, De Santis KK, Zeeb H
Digital Information Exchange Between the Public and Researchers in Health Studies: Scoping Review
J Med Internet Res 2025;27:e63373
URL: https://www.jmir.org/2025/1/e63373
doi:10.2196/63373
PMID:

©Nazli Soltani, Thilo Dietz, Doris Ochterbeck, Jens Dierkes, Katja Restel, Lara Christianson, Karina Karolina De Santis, Hajo Zeeb. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Different Intervention Modes in Lifestyle Intervention for the Prevention of Type 2 Diabetes and the Reversion to Normoglycemia in Adults With Prediabetes: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Yachen Wang[1*], BMed; Xin Chai[1*], BMed; Yueqing Wang[1*], MPH; Xuejun Yin[1,2], PhD; Xinying Huang[1,3], MPH; Qiuhong Gong[4], MD; Juan Zhang[1], PhD; Ruitai Shao[1], PhD; Guangwei Li[1], MD

[1]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]The George Institute for Global Health, University of New South Wales, Newtown, Australia

[3]Department of Neurology, Beijing Tiantan Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[4]Fuwai Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Ruitai Shao, PhD
School of Population Medicine and Public Health
Chinese Academy of Medical Sciences & Peking Union Medical College
31 Yard, Beijige Santiao, Dongcheng District, Beijing, China
Beijing, 100730
China
Phone: 86 13911644595
Email: shaoruitai@sph.pumc.edu.cn

Abstract

Background: Lifestyle interventions have been acknowledged as effective strategies for preventing type 2 diabetes mellitus (T2DM). However, the accessibility of conventional face-to-face interventions is often limited. Digital health intervention has been suggested as a potential solution to overcome the limitation. Despite this, there remains a significant gap in understanding the effectiveness of digital health for individuals with prediabetes, particularly in reducing T2DM incidence and reverting to normoglycemia.

Objective: This study aimed to assess the effectiveness of different intervention modes of digital health, face-to-face, and blended interventions, particularly the benefits of digital health intervention, in reducing T2DM incidence and facilitating the reversion to normoglycemia in adults with prediabetes compared to the usual care.

Methods: We conducted a comprehensive search in 9 electronic databases, namely MEDLINE, Embase, ACP Journal Club, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Cochrane Clinical Answers, Cochrane Methodology Register, Health Technology Assessment, and NHS Economic Evaluation Database through Ovid, from the inception to October 2024. This review included randomized controlled trials (RCTs) that studied the effectiveness of lifestyle interventions in adults with prediabetes. The overall intervention effect was synthesized using a random-effects model. The I^2 statistic was used to assess heterogeneity across the RCTs. We performed a subgroup analysis to explore the effectiveness of digital health, face-to-face, and blended interventions compared with the control group, which received usual care.

Results: From an initial 7868 records retrieved from 9 databases, we identified 54 articles from 31 RCTs. Our analysis showed that face-to-face interventions demonstrated a significant 46% risk reduction in T2DM incidence (risk ratio [RR] 0.54, 95% CI 0.47-0.63; I^2=43%; P<.001), and a 46% increase in the reversion to normoglycemia (RR 1.46, 95% CI 1.11-1.91; I^2=82%; P=.006), when compared with the control group. On the other hand, digital health interventions, compared with the control group, were associated with a 12% risk reduction in T2DM incidence (RR 0.88, 95% CI 0.77-1.01; I^2=0.6%; P=.06). Moreover, the blended interventions combining digital and face-to-face interventions suggested a 37% risk reduction in T2DM incidence (RR 0.63, 95% CI 0.49-0.81;I^2<0.01%; P<.001) and an 87% increase in the reversion to normoglycemia (RR 1.87, 95% CI 1.30-2.69; I^2=23%;

XSL•FO
RenderX

P=.001). However, no significant effect on the reversal of prediabetes to normoglycemia was observed from the digital health interventions.

Conclusions: Face-to-face interventions have consistently demonstrated promising effectiveness in both reductions in T2DM incidence and reversion to normoglycemia in adults with prediabetes. However, the effectiveness of digital health interventions in these areas has not been sufficiently proven. Given these results, further research is required to provide more definitive evidence of digital health and blended interventions in T2DM prevention in the future.

Trial Registration: PROSPERO CRD42023414313; https://tinyurl.com/55ac4j4n

(J Med Internet Res 2025;27:e63975) doi:10.2196/63975

KEYWORDS

mobile phone; prediabetic state; digital health intervention; intervention mode; lifestyle intervention; type 2 diabetes mellitus; meta-analysis; systematic review; review

Introduction

Type 2 diabetes mellitus (T2DM) is a significant global public health issue. Up to 2021, there were 529 million individuals with T2DM [1], and 6.7 million deaths attributable to T2DM [2] with a cost of over a trillion dollars in health care worldwide [3].

Prediabetes is a high-risk condition for developing T2DM [4]. Without appropriate interventions, 5% to 18% of prediabetes will progress to T2DM annually [5]. In 2018, the estimated prevalence of prediabetes among Chinese adults was 38%, highlighting the urgent need for early interventions to prevent the progression to T2DM.

According to the World Health Organization (WHO) definition provided in 1998, a lifestyle intervention aims to modify an individual's way of living and improve their physical and psychological health by changing patterns of behavior that are harmful to health [6]. There is substantial evidence indicating that lifestyle interventions can prevent T2DM progression [7-9], and even reverse prediabetes to normoglycemia [10,11]. However, most of this evidence stems from face-to-face interventions, which are time-consuming and reach a limited audience. Digital health intervention, defined as a discrete functionality of digital technology, is applied to achieve health objectives and is implemented within digital health applications, including SMS, smartphones, and wearable devices, to achieve health objectives. Meanwhile, it offers immense opportunities for self-management and the prevention of chronic diseases [12-14].

While previous meta-analyses have demonstrated the effectiveness of digital health interventions in promoting physical activity and dietary changes, leading to weight loss and improved glucose control in individuals with prediabetes [15-17], evidence regarding their effectiveness in reducing T2DM incidence and reverting prediabetes to normoglycemia remains limited. Therefore, T2DM incidence and reversion to normoglycemia are essential endpoints for providing evidence on the application of digital health interventions in T2DM prevention.

This systematic review and meta-analysis aims to evaluate the effectiveness of face-to-face, digital health, and blended interventions in reducing T2DM incidence and reversing prediabetes to normoglycemia based on randomized controlled trials (RCTs).

Methods

Overview

The systematic review followed the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) 2020 checklist [18]. The review protocol was registered at PROSPERO (International Prospective Register of Systematic Reviews; CRD42023414313).

Search Strategy and Eligible Criteria

Electronic searches of 9 databases (MEDLINE, Embase, ACP Journal Club, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Cochrane Clinical Answers, Cochrane Methodology Register, Health Technology Assessment, and NHS Economic Evaluation Database) were done through Ovid. We performed the initial search on May 31, 2022, and updated the search on October 2, 2024, with no language restrictions. The search terms were developed in consultation with librarians incorporating MeSH (Medical Subject Headings) terms and text words and piloted before being used in the current review. The search strategy was: "prediabetes or high-risk of diabetes" and "lifestyle intervention or digital health intervention" and "RCTs." The detailed search strategy can be found in Multimedia Appendix 1 [15,19]. The reference lists of the included studies and previous reviews were thoroughly hand-searched to identify any additional relevant research. The eligibility criteria were guided by the Population, Intervention, Comparison, Outcome, and Study Design (PICOS) framework, as illustrated in Textbox 1.

XSL•FO
RenderX

Textbox 1. Population, Intervention, Comparison, Outcome, and Study Design (PICOS) to determine the eligibility

Inclusion criteria

- Population
- Prediabetes, including impaired fasting glucose, impaired glucose tolerance, impaired fasting glucose and impaired glucose tolerance, elevated hemoglobin A_{1c}, and intermediate hyperglycemia by different guidelines (Multimedia Appendix 2 [20-25])
- Intervention
- Lifestyle intervention (see "Definition and Examples of Lifestyle Intervention") lasted for at least 1 year.
- Comparison
- Usual care.
- Outcome
- Type 2 diabetes mellitus incidence or reversion to normoglycemia.
- Study design
- Randomized controlled trial.

Exclusion criteria

- Population
- High risk of diabetes identified by risk scores.
- Intervention
- Pharmacological intervention or a combination of pharmacological and lifestyle intervention.
- Comparison
- Pharmacological intervention.
- Outcome
- Intermediate outcomes such as weight loss.
- Study design
- Nonrandomized controlled trial.

A total of 2 authors (YCW and XC) independently screened the titles and abstracts of studies and assessed the full text for eligibility. Disagreements, if any, were resolved through discussion with a third author (XJY). For efficient management of the selection process, all retrieved studies were imported into Covidence.

Definition and Examples of Lifestyle Intervention

According to the 1998 World Health Organization definition, a lifestyle intervention aims to modify an individual's way of living and improve their physical and psychological health by changing patterns of behavior that are harmful to health [6].

For example, in our meta-analysis, the behavior pattern changes mainly include nutrition, physical activity, and weight management.

Data Extraction and Quality Assessment

Data extraction was carried out independently using standard data extraction forms. Extracted data included (1) study information (author, year of publication, and study location); (2) study participants (prediabetes phenotype, diagnostic criteria, ethics, age, the proportion of male participants, baseline, BMI, fasting plasma glucose [FPG], and sample size); (3) intervention and control details (the intervention treatment, the control group, intervention settings, delivery format, intervention mode, intervention duration, and the follow-up duration was defined from the intervention year to the first year in which reported T2DM incidence occurred, as well as the time to reversion to normoglycemia); (4) for the studies applying digital health interventions, the theoretical framework, intervention adherence, digital health intervention categories, and its intervention components categorized by the WHO [26]; (5) outcome measurements; and (6) key study results including T2DM events, proportion or numbers of individuals reverting to normoglycemia and effect size between intervention group and control group.

The risk of bias was assessed using the revised Cochrane Risk-of-Bias tool (RoB 2.0) for randomized trials [27]. The assessment considered the following elements: (1) randomization process, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of the outcome, (5) selection of the reported result, and (6) overall bias.

Data extraction and quality assessment were independently carried out by 2 authors (YCW and XC). Any disagreements between the 2 authors were resolved through discussion or with a third author (XJY).

XSL•FO
RenderX

Synthesis of Results

Results were sought for both baseline and end-of-intervention or postintervention outcome data, specifically focusing on the data that first observed the primary outcome. If the original article provided 95% CI or SE without SD, we converted them into SD, following the guidelines in the Cochrane Handbook for Systematic Reviews of Interventions [27]. To facilitate the interpretation of the findings, we used standardization techniques to unify measurement units across studies. Both blood glucose and blood lipids indicators were measured in mmol/L, except hemoglobin A_{1c} (HbA_{1c}) in percentage.

The primary outcomes of this systematic review and meta-analysis were the incidence of T2DM and the reversion from prediabetes to normoglycemia, both compared with the usual care, expressed in terms of risk ratios (RRs), hazard ratios, or odds ratios. Secondary outcomes encompassed the changes in blood glucose, specifically FPG, 2-hour postprandial glucose, and HbA_{1c}. This study also examined variations in key indicators including body weight, BMI, waist circumference, blood pressure, and blood lipid profiles. The lipid profiles encompass triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and total cholesterol.

Statistical Analysis

Intervention effects were defined as the differences in outcomes between the intervention and control groups. For dichotomized outcomes, synthesized measures were presented as RRs; and for continuous outcomes, we presented mean differences. The pooled estimates were synthesized by random-effect models and presented in the forest plots. The I^2 statistic was used to assess statistical heterogeneity between trials. For studies with 2 or more intervention groups, each group was treated independently and annotated in the forest plots [28].

A prespecified subgroup analysis was conducted to assess the effectiveness of digital health interventions, face-to-face lifestyle interventions, and the combined application compared with the usual care. The subgroups were categorized into 3 intervention modes [29], face-to-face, digital health, and blended. Based on the provision of behavioral support, digital health interventions can be divided into stand-alone interventions and digital support interventions (Textbox 2). Sensitivity analysis was conducted to verify the robustness of the pooled RRs. Funnel plot and Egger test were used to investigate the publication bias [30]. All analyses were carried out using Stata (version 16.0; StataCorp). We considered *P* less than .05 as significant.

Textbox 2. Categories of examples of intervention modes based on the provision of behavioral supports.

(i) Face-to-face interventions

- In-person support from a counselor face-to-face: for example, group sessions on lifestyle modification.

(ii) Digital health interventions

- Stand-alone interventions: participants were not offered ongoing support from a counselor. For example, self-management or self-monitor smartphone apps.
- Digital support intervention: remote support with behavioral support through mobile health technologies. For example, emails, text messaging prompts, or phone calls with a counselor.

(iv) Blended interventions

- Combination of face-to-face intervention and digital health intervention.

Ethical Considerations

This review does not involve human subject information, primary data collection, or any form of experimentation on individuals.

Results

Study Selection

The database search yielded 7868 articles from the 9 databases and identified 37 articles as eligible. We manually identified an additional 209 articles from the reference list of relevant reviews, of which 17 articles were deemed eligible. Finally, a total of 54 articles from 31 studies met the eligibility criteria and were included in the review (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) flow diagram describing the literature review process. T2DM: type 2 diabetes mellitus.

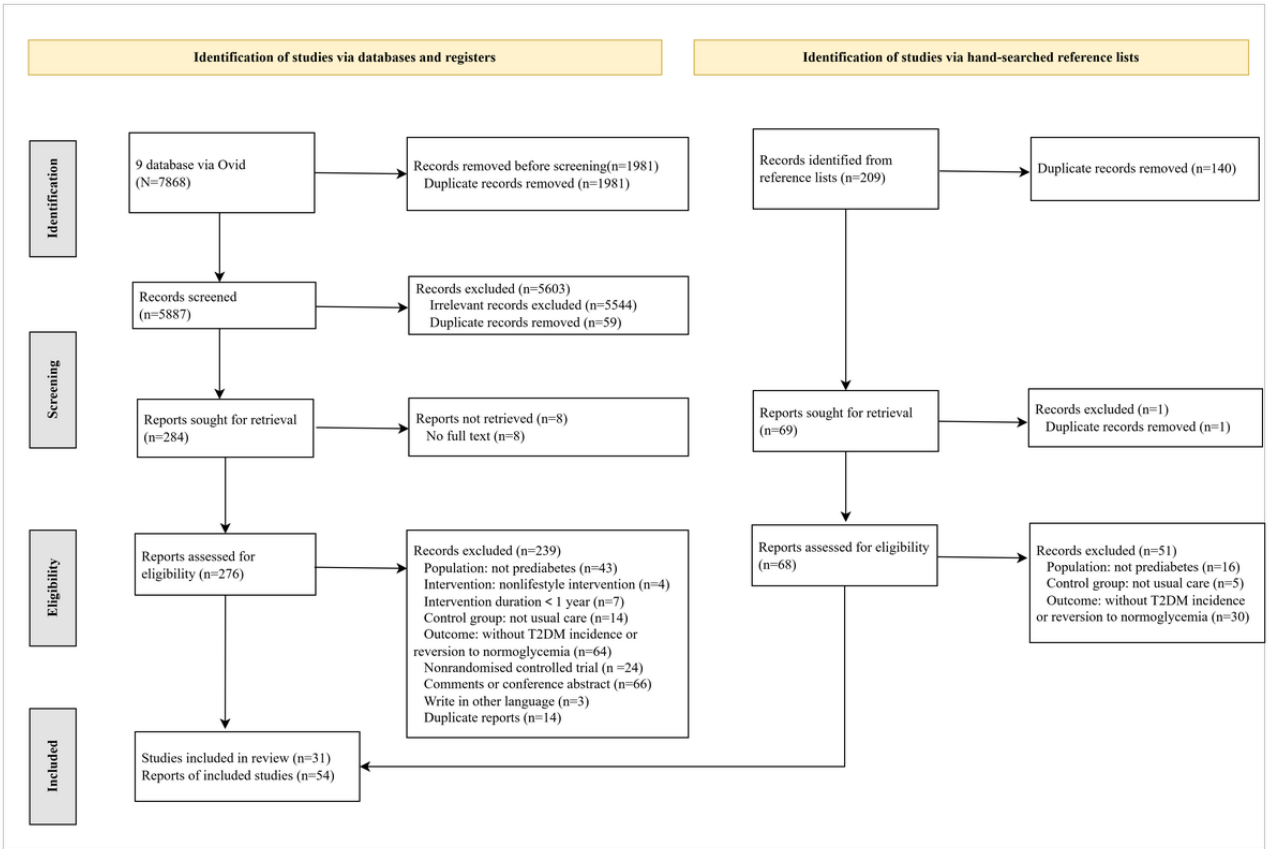

Study Characteristics

Baseline characteristics of the included studies are shown in Multimedia Appendix 3 [7-11,31-56]. The systematic review and meta-analysis included 23,684 individuals with prediabetes in total, with studies ranging from 48 to 2865 participants. Prediabetes identifications were primarily based on the WHO criteria since 1999 (15/31, 48%), which defined prediabetes by FPG of 6.1 to 6.9 mmol/L or 2-hour postprandial glucose of 7.8 to 11.0 mmol/L (Multimedia Appendix 2 [20-25]). Among the included studies, the majority (18/31, 58%) reported an average participant age exceeding 50 years. Among 31 studies, 30 studies reported the T2DM incidence as the outcome [7-11,31-54,57], and 10 studies reported the reversion to normoglycemia as the outcome [8-11,36,38,39,44,53,55].

Description of Interventions

Intervention characteristics are outlined in Multimedia Appendix 3 [7-11,31-56], including intervention content, comparative interventions, settings, delivery method, intervention duration, and follow-up duration. Intervention durations and follow-up durations across the studies varied, ranging from 1 to 6 years. A majority of the studies (22/31, 70% [7-10,35-38,40-44,46-52,55,57]) used face-to-face intervention, 6 studies used digital health intervention [10,11,32,34,45,54], and 6 studies used blended intervention [31,33,39,50,53,57]. No study with digital health intervention used stand-alone intervention. In the face-to-face interventions, group sessions (22/28, 79%) were the most common delivery format applied (Multimedia Appendix 3 [7-11,31-56]).

Of all the studies that included digital health components, SMS text messaging (5/12, 42%) and telephone contacts (6/12, 50%) were the most used. One of these studies used a web-based approach [11], and 3 studies used wearable technology [11,31,58]. According to the WHO classification of digital health interventions, targeted client communication was the most used digital health category (11/12, 92%), followed by personal health tracking (4/12, 33%). Transmitting targeted health information to clients was the intervention mostly used in targeted client communication including behavior change communication, health promotion communication, and client-centered SMS text messaging. Among the 12 studies that included digital health interventions, only 5 were published within the last 5 years [10,11,45,50,53]. The transtheoretical model of behavioral change and social cognitive theory was the most common theoretical method used in digital health interventions.

Main Outcomes

Type 2 Diabetes Mellitus Incidence

The pooled estimate of 30 studies revealed the significant effectiveness of lifestyle interventions in reducing the incidence of T2DM (RR 0.60, 95% CI 0.53-0.69; I^2=55%; P<.001; Figure 2).

XSL•FO
RenderX

Figure 2. The forest plot on the reduction of type 2 diabetes mellitus incidence by intervention mode subgroups (face-to-face, blended, and digital health interventions) compared with the usual care group [7-11,26-30,51-70]. Multiple intervention arms within a single study were treated as distinct intervention groups. DL: DerSimonian-Laird, a method to estimate the heterogeneity variance.

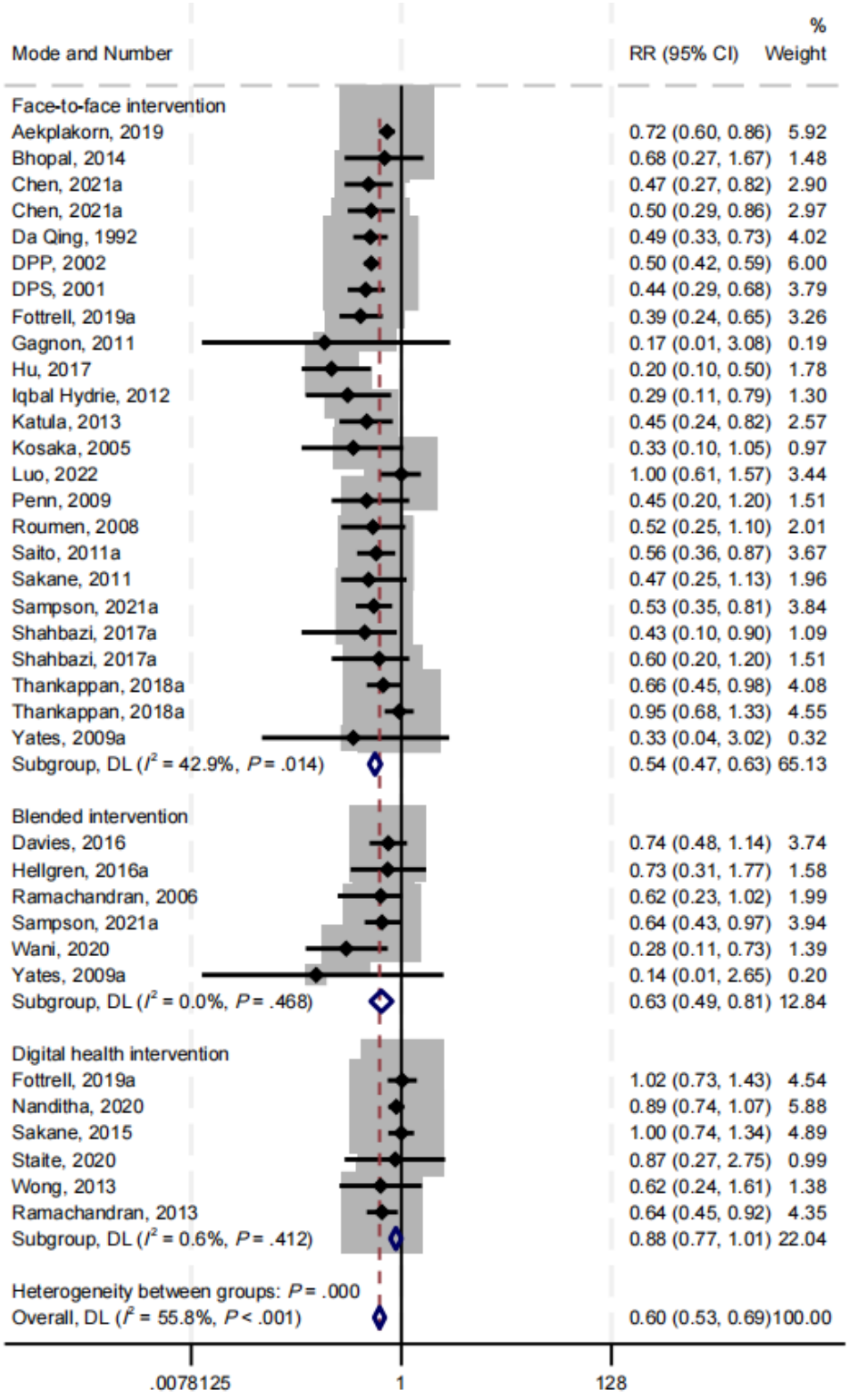

In subgroup analysis, face-to-face interventions were found to reduce the incidence of T2DM among adults with prediabetes by 46% (RR 0.54, 95% CI 0.47-0.63; I^2=43%; P<.001) compared with the control group (Figure 2). The digital health interventions (RR 0.88, 95% CI 0.77-1.01; I^2=0.6%; P=.06) observed a 12% risk reduction in T2DM incidence compared with the control group, although no statistical difference was observed. Notably, one study reported a 36% risk reduction in T2DM incidence through SMS test messaging [32]. However, no statistically significant difference was observed in the other two similar studies, despite observing a reduction in T2DM incidence [33,57]. Studies using telephone contacts and voice messages did not report any significant reduction in T2DM incidence [10,34]. The blended interventions (RR 0.63, 95%

XSL•FO
RenderX

CI 0.49-0.81; I^2<0.01%; P<.001) observed a 37% risk reduction in T2DM incidence compared with the control group.

Reversion From Prediabetes to Normoglycemia

A total of 10 studies were included in the meta-analysis of reversion from prediabetes to normoglycemia [8-11,36,38,39,44,53,55]. Compared with the control group, the lifestyle intervention group demonstrated a 44% increase in the rate of reversion from prediabetes to normoglycemia (RR 1.44, 95% CI 1.15-1.81; I^2=84%; P<.001; Figure 3).

Figure 3. The forest plot on reversion to normoglycemia by intervention mode subgroups (face-to-face, blended, and digital health interventions) compared with the usual care group [8-11,52,54,55,60,69,71]. Multiple intervention arms within a single study were treated as distinct intervention groups. DL: DerSimonian-Laird, a method to estimate the heterogeneity variance.

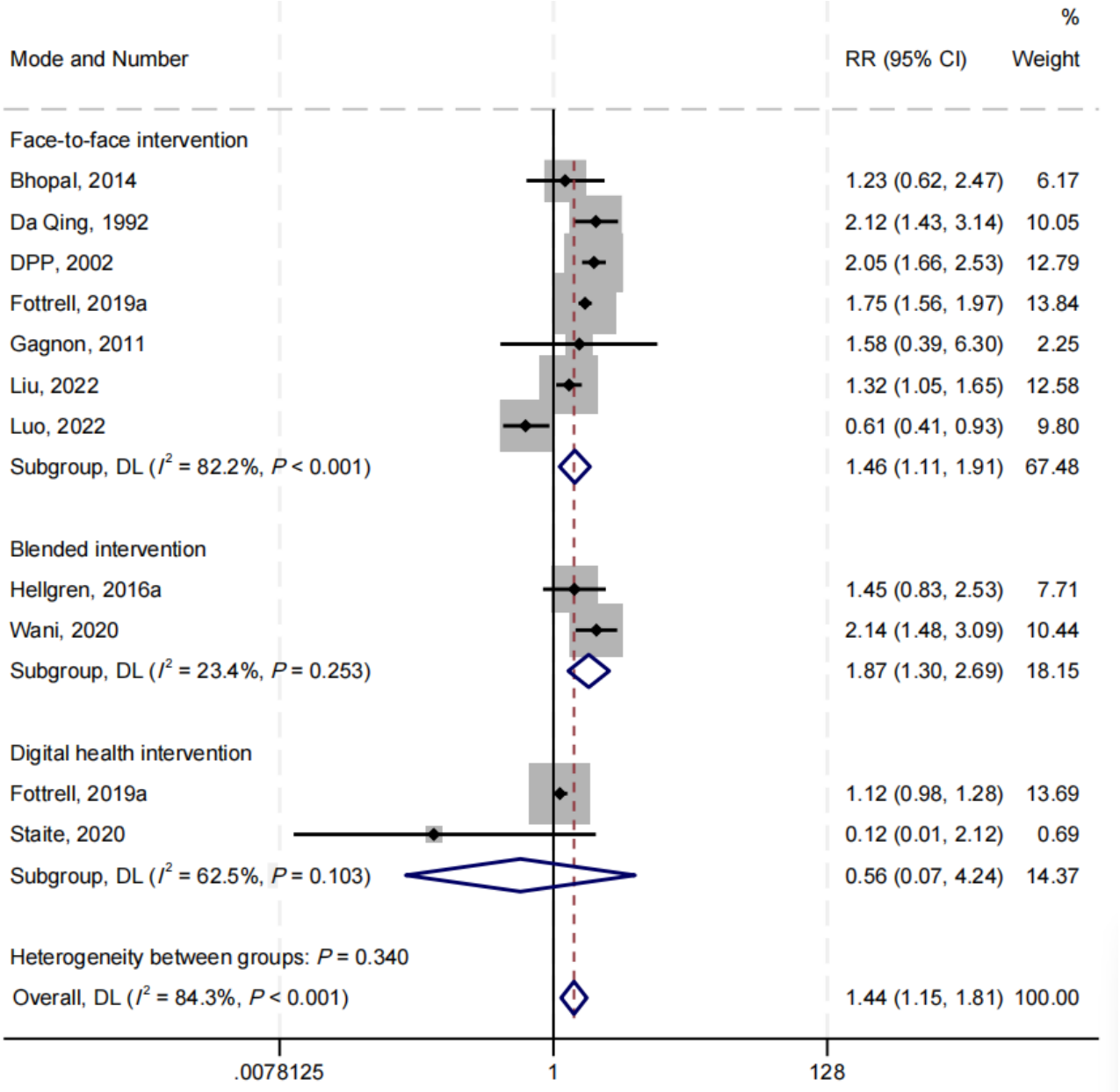

In subgroup analysis (Figure 3), the face-to-face intervention and blended intervention group exhibited significant effectiveness by 46% (RR 1.46, 95% CI 1.11-1.91; I^2=82%; P=.006) and 87% (RR 1.87, 95% CI 1.30-2.69; I^2=23%; P=.001) increase in reversion to normoglycemia compared with the control group, respectively. A total of 2 studies used digital health interventions: one used voice messages [10], while the other incorporated wearable technology, web-based education, and motivational SMS text messages [11]. In the digital health intervention group, neither of the 2 studies observed significant effectiveness of digital health interventions on the reversion of prediabetes to normoglycemia. One study [10] observed the potential effectiveness on reversion but produced no statistical difference (RR 1.12, 95% CI 0.98-1.28) [10]. In another study [11], the CI was too wide to determine the effect (RR 0.12, 95% CI 0.01-2.12).

Secondary Outcomes

Face-to-face interventions' standard mean difference (SMD=–0.25, 95% CI –0.35 to –0.15; I^2=48%; P<.001)

XSL•FO
RenderX

demonstrated a significant decrease in weight loss than digital health interventions (SMD=–0.07, 95% CI –0.17 to 0.03; I^2=48%; P=.15). Similarly, reductions in waist circumference, BMI, FPG, 2-hour postprandial glucose, HbA_{1c}, systolic blood pressure, diastolic blood pressure, and TG were more effectively pronounced in the face-to-face intervention. Contrarily, low-density lipoprotein cholesterol showed a significant decrease through digital health interventions (SMD=–0.10, 95% CI –0.19 to –0.02; I^2=0%; P=0.01) while no effectiveness was observed in face-to-face interventions (SMD=–0.13, 95% CI –0.41 to 0.14; I^2=81%; P=.34; Multimedia Appendix 4).

Risk-of-Bias Assessment and Publication Bias

Among 30 studies that reported T2DM incidence as the outcome, 10 studies were classified as “low risk” [9,10,32,36,42,44,45,48,52,56], 7 studies as having “some concerns” [7,8,31,33,39,40,54], and the remaining 13 studies as “high risk” [11,34,35,37,38,41,43,46,47,49-51,53], assessed by RoB 2.0 judgment. Among the 10 studies that reported reversion to normoglycemia as the outcome, 4 studies were classified as “low risk” [9,10,36,44], 2 studies as having “some concerns” [8,39], and the remaining 4 studies as “high risk” [11,38,53,55] (Multimedia Appendix 4). The high risk of bias primarily stemmed from the biased selection of reported results, followed by missing outcome data. In other domains, most studies were considered low risk. Publication bias was detected for T2DM incidence (Egger test P=.02), but not for prediabetes reversion (Egger test P=.66). Funnel plots suggested that studies with small sample sizes were less likely to be published (Multimedia Appendix 4).

Sensitivity Analysis

The sensitivity analysis found no significant change after excluding studies one by one in both T2DM prevention and reversion to normoglycemia. After excluding studies with certain concerns and those deemed to be at high risk of bias, sensitivity analysis indicates that our meta-analysis results for T2DM incidence are stable. However, the number of studies on reversion to normoglycemia was too small to conduct a meta-analysis (Multimedia Appendix 4).

Discussion

Principal Findings

Our systematic review and meta-analysis demonstrated that face-to-face and blended interventions can effectively prevent T2DM incidence and promote reversion to normoglycemia. Digital health interventions would be a possible solution in T2DM prevention based on our review; however, more evidence is required to substantiate the effectiveness. There is limited evidence to support digital health intervention in reversion to normoglycemia among adults with prediabetes.

The findings in face-to-face interventions showed consistent effectiveness in T2DM prevention and reversion to normoglycemia, which is similar to the previous meta-analysis reported [59,60]. In addition, the study found that the face-to-face intervention group showed greater improvements in blood glucose, blood lipid levels, and other indicators compared with the digital health group, which is similar to previous studies [15].

The results of the blended interventions, which combine face-to-face and digital health interventions, showed significant effectiveness in reducing T2DM incidence and promoting the reversion to normoglycemia. This is in accordance with our earlier observations, which showed that blended intervention can effectively help people lose weight [61].

Although no statistical significance is observed in digital health intervention, the risk reduction in T2DM incidence indicates its potential effectiveness. These results are similar to those reported by previous systematic reviews among older adults with prediabetes including 3 RCTs [62], which contrast with earlier studies that reported the effectiveness of glucose indicators or lifestyle modification [63] in digital health intervention.

The inconsistency between studies using T2DM incidence and those focusing on weight loss and lifestyle changes as outcomes in digital health interventions may stem from several factors. First, it necessitates a multiyear observation period to ascertain the effectiveness of these interventions on the T2DM incidence and the reversal of hyperglycemia [7-9]. However, the studies of digital health intervention included in this meta-analysis had relatively short intervention periods ranging from 1 to 3.8 years, which may not be sufficient for observing the effectiveness. Second, it is recognized that adherence is an important indicator of expected good results [15,32]; however, 10 of 12 studies with digital health intervention included in this meta-analysis have not provided adherence information, which may explain why the digital health intervention has not reported the significant changes [10,11,31,32,39,45,50,53,54,56]. Finally, a possible explanation for the results may be the lack of adequate studies, resulting in CIs that are too wide to detect significance. Therefore, we should interpret the results with caution in digital health interventions.

Implications

Face-to-face lifestyle intervention, even when effective, adherence is not ideal in the real world [64]. Recently, studies have used digital health intervention in the Diabetes Prevention Program (DPP) in the real world, called the digital DPP [65,66]. The results showed that, in comparison to traditional face-to-face small-group education, the dDPP exhibits good reach and adherence [67]. Also, promising effectiveness in weight loss was found in dDPP; however, it is unclear in the T2DM incidence in the real world [65,68]. It is recognized that behavioral changes necessitated by such interventions require a more extended time period to assess their impact on T2DM incidence [69,70], although the American Diabetes Association recommends using digital health interventions in diabetes prevention programs for its effectiveness in glycemic control [71]. Therefore, more studies are needed to explore the effectiveness of digital health and blended interventions, observing outcomes related to T2DM incidence and reversion to normoglycemia, in order to provide more evidence for T2DM prevention.

XSL•FO
RenderX

Moreover, a prevailing challenge associated with digital health interventions is the observed decline in adherence over extended periods of use, which may account for the generally short intervention durations characteristic of digital health interventions [65]. Thus, digital health interventions should use certain behavioral change techniques or adhere to a theoretical model, as demonstrated in previous studies [72-74], as most studies used the transtheoretical model of behavioral change and social cognitive theory in lifestyle modification. Future research should not only focus on lifestyle modification but also collect information on adherence using digital health interventions.

In addition, for high-risk patients, the dDPP group had, on average, lower costs [75]. For social care use, cost-effectiveness analyses also support the preferential selection of DPPs over face-to-face small-group education per quality-adjusted life years [68]. Further research is required to develop accessible, effective, and cost-efficient diabetes prevention interventions incorporating digital health.

Furthermore, although the studies included in our meta-analysis did not use machine learning or generative artificial intelligence, these technologies have now been extensively applied in chronic disease prevention and control. In the future, machine learning with data from wearable devices will be capable of making informed judgments and providing individualized recommendations for the precision prevention and treatment of diabetes through the application of generative artificial intelligence [76]. However, it is crucial to pay attention to its fidelity to maximize its value and promote health care equity, given the challenges of compatibility between different artificial intelligence devices, standardization of data, and risks to data security [77,78].

Finally, even though there are several digital health interventions, digital support interventions have been used in diabetes prevention research limited to T2DM incidence and normoglycemia for the primary aim. For instance, a recent meta-analysis of diabetes management found that smartphone app and SMS test message interventions, but not website-based interventions, were associated with better glycemic control [15], which was defined as the primary aim of the research. Once the results with T2DM incidence and reversion of normoglycemia are available sufficiently, analyses can be performed to further explore the effective components within digital health interventions.

Strengths and Limitations

To the best of our knowledge, this is the first meta-analysis to assess the effectiveness of digital health and blended interventions on T2DM prevention using T2DM incidence and reversion to normoglycemia instead of weight loss and changes in blood glucose levels. We conducted a quantitative synthesis to explore the effectiveness of digital health interventions in T2DM prevention, which is different from other systematic reviews constrained by a limited number of studies to conduct a meta-analysis.

The study also has some limitations. First, digital health has been widely used in health research; however, 5 of 12 studies encompassed digital health interventions were published within the last 5 years, which does not meet the expectation in novel studies. To ensure robust findings, we conducted sensitivity analyses to exclude studies with potential biases from small samples or weak designs. Second, publication bias was found in the study for the T2DM incidence, and these results need to be interpreted with caution. Third, the limited studies on digital health interventions restrict our ability to conduct subgroup analyses on the effectiveness of their components. We have summarized the components and their theoretical bases in our results to aid further research. Finally, our research does not directly compare the effectiveness of different intervention modes; instead, it focuses on comparisons within control groups due to the lack of direct comparative studies in the literature.

Conclusion

While digital health interventions for lifestyle changes suggest potential effectiveness in preventing T2DM, further evidence is needed to confirm their effectiveness. There is limited evidence supporting the effectiveness of digital health interventions in reversing prediabetes. Face-to-face lifestyle interventions are effective but limited by accessibility and low adherence. With the rapid annual increase of diabetes globally, there is an urgent need to prevent T2DM, particularly among those with prediabetes. Future research is needed to consolidate the effectiveness of interventions incorporating digital health in T2DM prevention and to increase the accessibility of lifestyle interventions.

Acknowledgments

This study is supported by Construction of Medical Base for Chronic Diseases at Peking Union Medical College (WH10022022010), the nonprofit Central Research Institute Fund of Chinese Academy of Medical Sciences (Grant No. 2021-RC330-004), CAMS Innovation Fund for Medical Sciences (CIFMS) (Grant No. 2023-I2M-2-001), and the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (Grant No. 2022-ZHCH330-01). The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

YCW, XC, RS, XY, JZ, QG, YQW, and XH developed the search strategy, selected which trials to include, and extracted data from the randomized controlled trials. YCW and XC conducted the quality assessment of trials. YCW carried out the analysis, interpreted the analysis, drafted the final review, and kept the review up to date. YCW and YQW wrote the main manuscript text. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 17 KB - jmir_v27i1e63975_app1.docx]

Multimedia Appendix 2
Diagnosis and classification of prediabetes in different guidelines.
[DOCX File , 21 KB - jmir_v27i1e63975_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies.
[DOCX File , 165 KB - jmir_v27i1e63975_app3.docx]

Multimedia Appendix 4
Assessment of secondary outcomes, evaluation of publication bias, and sensitivity analysis.
[DOCX File , 710 KB - jmir_v27i1e63975_app4.docx]

Multimedia Appendix 5
PRISMA checklist.
[PDF File (Adobe PDF File), 171 KB - jmir_v27i1e63975_app5.pdf]

References

1. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the global burden of disease study 2021. Lancet 2023;402(10397):203-234 [FREE Full text] [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
2. Liu M, Liu SW, Wang LJ, Bai YM, Zeng XY, Guo HB, et al. Burden of diabetes, hyperglycaemia in China from to 2016: findings from the 1990 to 2016, global burden of disease study. Diabetes Metab 2019;45(3):286-293. [doi: 10.1016/j.diabet.2018.08.008] [Medline: 30196138]
3. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. J Epidemiol Glob Health 2020;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
4. Wang T, Zhao Z, Wang G, Li Q, Xu Y, Li M, et al. Age-related disparities in diabetes risk attributable to modifiable risk factor profiles in Chinese adults: a nationwide, population-based, cohort study. Lancet Healthy Longev 2021;2(10):e618-e628 [FREE Full text] [doi: 10.1016/S2666-7568(21)00177-X] [Medline: 36098018]
5. Echouffo-Tcheugui JB, Perreault L, Ji L, Dagogo-Jack S. Diagnosis and management of prediabetes: a review. JAMA 2023;329(14):1206-1216. [doi: 10.1001/jama.2023.4063] [Medline: 37039787]
6. Deshpande SN, Simkin DR. Complementary and integrative approaches to prevention and treatment of child and adolescent obesity. Child Adolesc Psychiatr Clin N Am 2023;32(2):395-419. [doi: 10.1016/j.chc.2022.08.013] [Medline: 37147044]
7. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001;344(18):1343-1350. [doi: 10.1056/NEJM200105033441801] [Medline: 11333990]
8. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The da qing IGT and diabetes study. Diabetes Care 1997;20(4):537-544. [doi: 10.2337/diacare.20.4.537] [Medline: 9096977]
9. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
10. Fottrell E, Ahmed N, Morrison J, Kuddus A, Shaha SK, King C, et al. Community groups or mobile phone messaging to prevent and control type 2 diabetes and intermediate hyperglycaemia in Bangladesh (DMagic): a cluster-randomised

XSL•FO
RenderX

controlled trial. Lancet Diabetes Endocrinol 2019;7(3):200-212 [FREE Full text] [doi: 10.1016/S2213-8587(19)30001-4] [Medline: 30733182]
11. Staite E, Bayley A, Al-Ozairi E, Stewart K, Hopkins D, Rundle J, et al. A wearable technology delivering a web-based diabetes prevention program to people at high risk of type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(7):e15448 [FREE Full text] [doi: 10.2196/15448] [Medline: 32459651]
12. American Diabetes Association. 3. Prevention or delay of type 2 diabetes: standards of medical care in diabetes-2020. Diabetes Care 2020;43(Suppl 1):S32-S36. [doi: 10.2337/dc20-S003] [Medline: 31862746]
13. Chinese Society of Endocrinology, Chinese Diabetes Society, Chinese Endocrinologist Association. Intervention for adults with pre-diabetes: a Chinese expert consensus (2023 edition). Chinese Journal of Diabetes Mellitus. 2023 (06) 2023 Jun 20:484-494. [doi: 10.3760/cma.j.cn115791-20230509-00188]
14. WHO Guidelines Approved by the Guidelines Review Committee. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019.
15. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, DigiCare4You Consortium. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
16. Edney S, Chua XH, Müller AM, Kui KY, Müller-Riemenschneider F. mHealth interventions targeting movement behaviors in Asia: A scoping review. Obes Rev 2022;23(4):e13396. [doi: 10.1111/obr.13396] [Medline: 34927346]
17. Chen C, Ding S, Wang J. Digital health for aging populations. Nat Med 2023;29(7):1623-1630. [doi: 10.1038/s41591-023-02391-8] [Medline: 37464029]
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Yang M, Duan Y, Liang W, Peiris D, Baker JS. Effects of face-to-face and eHealth blended interventions on physical activity, diet, and weight-related outcomes among adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2023;20(2):1560 [FREE Full text] [doi: 10.3390/ijerph20021560] [Medline: 36674317]
20. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2004 Jan;27 Suppl 1:S5-S10. [doi: 10.2337/diacare.27.2007.s5] [Medline: 14693921]
21. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2010 Jan;33 Suppl 1(Suppl 1):S62-S69 [FREE Full text] [doi: 10.2337/dc10-S062] [Medline: 20042775]
22. World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. URL: https://iris.who.int/handle/10665/66040 [accessed 2024-12-19]
23. World Health Organization. WHO Expert Committee on Diabetes Mellitus: second report. World Health Organ Tech Rep Ser 1980;646:1-80. [Medline: 6771926]
24. International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care 2009 Jul;32(7):1327-1334 [FREE Full text] [doi: 10.2337/dc09-9033] [Medline: 19502545]
25. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 1997 Jul;20(7):1183-1197. [doi: 10.2337/diacare.20.7.1183] [Medline: 9203460]
26. world health organization. Classification of digital health interventions v1.0. 2018. URL: https://www.who.int/publications/i/item/WHO-RHR-18.06 [accessed 2018-01-01]
27. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
28. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the cochrane handbook for systematic reviews of interventions. Cochrane Database Syst Rev 2019;10(10):ED000142 [FREE Full text] [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
29. MacPherson MM, Merry KJ, Locke SR, Jung ME. mHealth prompts within diabetes prevention programs: a scoping review. Mhealth 2022;8:20 [FREE Full text] [doi: 10.21037/mhealth-21-22] [Medline: 35449504]
30. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]
31. Davies MJ, Gray LJ, Troughton J, Gray A, Tuomilehto J, Farooqi A, Let's Prevent Diabetes Team. A community based primary prevention programme for type 2 diabetes integrating identification and lifestyle intervention for prevention: the let's prevent diabetes cluster randomised controlled trial. Prev Med 2016;84:48-56. [doi: 10.1016/j.ypmed.2015.12.012] [Medline: 26740346]
32. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]

XSL•FO
RenderX

33. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V, Indian Diabetes Prevention Programme (IDPP). The Indian diabetes prevention programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). Diabetologia 2006;49(2):289-297. [doi: 10.1007/s00125-005-0097-z] [Medline: 16391903]
34. Sakane N, Kotani K, Takahashi K, Sano Y, Tsuzaki K, Okazaki K, et al. Effects of telephone-delivered lifestyle support on the development of diabetes in participants at high risk of type 2 diabetes: J-DOIT1, a pragmatic cluster randomised trial. BMJ Open 2015;5(8):e007316 [FREE Full text] [doi: 10.1136/bmjopen-2014-007316] [Medline: 26289448]
35. Aekplakorn W, Tantayotai V, Numsangkul S, Tatsato N, Luckanajantachote P, Himathongkam T. Evaluation of a community-based diabetes prevention program in Thailand: a cluster randomized controlled trial. J Prim Care Community Health 2019;10:2150132719847374 [FREE Full text] [doi: 10.1177/2150132719847374] [Medline: 31079517]
36. Bhopal RS, Douglas A, Wallia S, Forbes JF, Lean MEJ, Gill JMR, et al. Effect of a lifestyle intervention on weight change in south Asian individuals in the UK at high risk of type 2 diabetes: a family-cluster randomised controlled trial. Lancet Diabetes Endocrinol 2014;2(3):218-227 [FREE Full text] [doi: 10.1016/S2213-8587(13)70204-3] [Medline: 24622752]
37. Chen X, Zhao S, Hsue C, Dai X, Liu L, Miller JD, et al. Effects of aerobic training and resistance training in reducing cardiovascular disease risk for patients with prediabetes: a multi-center randomized controlled trial. Prim Care Diabetes 2021;15(6):1063-1070. [doi: 10.1016/j.pcd.2021.08.013] [Medline: 34649825]
38. Gagnon C, Brown C, Couture C, Kamga-Ngande CN, Hivert MF, Baillargeon JP, et al. A cost-effective moderate-intensity interdisciplinary weight-management programme for individuals with prediabetes. Diabetes Metab 2011;37(5):410-418. [doi: 10.1016/j.diabet.2011.01.003] [Medline: 21489843]
39. Hellgren MI, Jansson P, Wedel H, Lindblad U. A lifestyle intervention in primary care prevents deterioration of insulin resistance in patients with impaired glucose tolerance: a randomised controlled trial. Scand J Public Health 2016;44(7):718-725. [doi: 10.1177/1403494816663539] [Medline: 27550085]
40. Hu Z, Qin L, Xu H. One-year results of a synthetic intervention model for the primary prevention of T2D among elderly individuals with prediabetes in Rural China. Int J Environ Res Public Health 2017;14(4):417 [FREE Full text] [doi: 10.3390/ijerph14040417] [Medline: 28420105]
41. Iqbal Hydrie MZ, Basit A, Shera AS, Hussain A. Effect of intervention in subjects with high risk of diabetes mellitus in Pakistan. J Nutr Metab 2012;2012:867604 [FREE Full text] [doi: 10.1155/2012/867604] [Medline: 22888411]
42. Katula JA, Vitolins MZ, Morgan TM, Lawlor MS, Blackwell CS, Isom SP, et al. The healthy living partnerships to prevent diabetes study: 2-year outcomes of a randomized controlled trial. Am J Prev Med 2013;44(4 Suppl 4):S324-S332 [FREE Full text] [doi: 10.1016/j.amepre.2012.12.015] [Medline: 23498294]
43. Kosaka K, Noda M, Kuzuya T. Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. Diabetes Res Clin Pract 2005;67(2):152-162. [doi: 10.1016/j.diabres.2004.06.010] [Medline: 15649575]
44. Luo Y, Wang H, Zhou X, Chang C, Chen W, Guo X, et al. A randomized controlled clinical trial of lifestyle intervention and pioglitazone for normalization of glucose status in Chinese with prediabetes. J Diabetes Res 2022;2022:2971382 [FREE Full text] [doi: 10.1155/2022/2971382] [Medline: 35036447]
45. Nanditha A, Thomson H, Susairaj P, Srivanichakorn W, Oliver N, Godsland IF, et al. A pragmatic and scalable strategy using mobile technology to promote sustained lifestyle changes to prevent type 2 diabetes in India and the UK: a randomised controlled trial. Diabetologia 2020;63(3):486-496 [FREE Full text] [doi: 10.1007/s00125-019-05061-y] [Medline: 31919539]
46. Penn L, White M, Oldroyd J, Walker M, Alberti KGMM, Mathers JC. Prevention of type 2 diabetes in adults with impaired glucose tolerance: the European diabetes prevention RCT in newcastle upon Tyne, UK. BMC Public Health 2009;9:342 [FREE Full text] [doi: 10.1186/1471-2458-9-342] [Medline: 19758428]
47. Roumen C, Corpeleijn E, Feskens EJM, Mensink M, Saris WHM, Blaak EE. Impact of 3-year lifestyle intervention on postprandial glucose metabolism: the SLIM study. Diabet Med 2008;25(5):597-605. [doi: 10.1111/j.1464-5491.2008.02417.x] [Medline: 18445174]
48. Saito T, Watanabe M, Nishida J, Izumi T, Omura M, Takagi T, Zensharen Study for Prevention of Lifestyle Diseases Group. Lifestyle modification and prevention of type 2 diabetes in overweight Japanese with impaired fasting glucose levels: a randomized controlled trial. Arch Intern Med 2011;171(15):1352-1360. [doi: 10.1001/archinternmed.2011.275] [Medline: 21824948]
49. Sakane N, Kotani K, Okazaki K, Sato J, Suzuki S, Morita S. Recruitment and baseline characteristics of J-DOIT1 (Japan diabetes outcome trial 1). Diabetes Technol Ther 2011;13(2):267-268.
50. Sampson M, Clark A, Bachmann M, Garner N, Irvine L, Howe A, Norfolk Diabetes Prevention Study (NDPS) Group. Lifestyle intervention with or without lay volunteers to prevent type 2 diabetes in people with impaired fasting glucose and/or nondiabetic hyperglycemia: a randomized clinical trial. JAMA Intern Med 2021;181(2):168-178 [FREE Full text] [doi: 10.1001/jamainternmed.2020.5938] [Medline: 33136119]
51. Shahbazi S, Vahdat Shariatpanahi Z. Prevention of type 2 diabetes mellitus by changes in diet among subjects with abnormal glucose metabolism: a randomized clinical trial. Int J Diabetes Dev Ctries 2018;38(1):69-74. [doi: 10.1007/s13410-017-0548-3]

XSL•FO
RenderX

52. Thankappan KR, Sathish T, Tapp RJ, Shaw JE, Lotfaliany M, Wolfe R, et al. A peer-support lifestyle intervention for preventing type 2 diabetes in India: a cluster-randomized controlled trial of the Kerala diabetes prevention program. PLoS Med 2018;15(6):e1002575 [FREE Full text] [doi: 10.1371/journal.pmed.1002575] [Medline: 29874236]
53. Wani K, Alfawaz H, Alnaami AM, Sabico S, Khattak MNK, Al-Attas O, et al. Effects of a 12-month intensive lifestyle monitoring program in predominantly overweight/obese Arab adults with prediabetes. Nutrients 2020;12(2):464 [FREE Full text] [doi: 10.3390/nu12020464] [Medline: 32059477]
54. Wong CKH, Fung CSC, Siu SC, Lo YYC, Wong KW, Fong DYT, et al. A short message service (SMS) intervention to prevent diabetes in Chinese professional drivers with pre-diabetes: a pilot single-blinded randomized controlled trial. Diabetes Res Clin Pract 2013;102(3):158-166. [doi: 10.1016/j.diabres.2013.10.002] [Medline: 24466598]
55. Liu Y, Guo H, Wang Q, Chen J, Xuan Y, Xu J, et al. Short-term effects of lifestyle intervention in the reversion to normoglycemia in people with prediabetes. Prim Care Diabetes 2022;16(1):168-172. [doi: 10.1016/j.pcd.2021.12.009] [Medline: 34930688]
56. Yates T, Davies M, Khunti K. Preventing type 2 diabetes: can we make the evidence work? Postgrad Med J 2009;85(1007):475-480. [doi: 10.1136/pgmj.2008.076166] [Medline: 19734515]
57. Yates T, Davies M, Gorely T, Bull F, Khunti K. Effectiveness of a pragmatic education program designed to promote walking activity in individuals with impaired glucose tolerance: a randomized controlled trial. Diabetes Care 2009;32(8):1404-1410 [FREE Full text] [doi: 10.2337/dc09-0130] [Medline: 19602539]
58. Yates T, Griffin S, Bodicoat DH, Brierly G, Dallosso H, Davies MJ, et al. Promotion of physical activity through structured Education with differing Levels of ongoing Support for people at high risk of type 2 diabetes (PROPELS): study protocol for a randomized controlled trial. Trials 2015;16:289 [FREE Full text] [doi: 10.1186/s13063-015-0813-z] [Medline: 26130075]
59. Hemmingsen B, Gimenez-Perez G, Mauricio D, Roqué I Figuls M, Metzendorf M, Richter B. Diet, physical activity or both for prevention or delay of type 2 diabetes mellitus and its associated complications in people at increased risk of developing type 2 diabetes mellitus. Cochrane Database Syst Rev 2017;12(12):CD003054 [FREE Full text] [doi: 10.1002/14651858.CD003054.pub4] [Medline: 29205264]
60. Glechner A, Keuchel L, Affengruber L, Titscher V, Sommer I, Matyas N, et al. Effects of lifestyle changes on adults with prediabetes: a systematic review and meta-analysis. Prim Care Diabetes 2018;12(5):393-408 [FREE Full text] [doi: 10.1016/j.pcd.2018.07.003] [Medline: 30076075]
61. Cheah KJ, Manaf ZA, Mat Ludin AF, Razalli NH. Potential role of hybrid weight management intervention: a scoping review. Digit Health 2024;10:20552076241258366 [FREE Full text] [doi: 10.1177/20552076241258366] [Medline: 39184018]
62. Jeem YA, Andriani RN, Nabila R, Emelia DD, Lazuardi L, Koesnanto H. The use of mobile health interventions for outcomes among middle-aged and elderly patients with prediabetes: a systematic review. Int J Environ Res Public Health 2022;19(20):13638 [FREE Full text] [doi: 10.3390/ijerph192013638] [Medline: 36294218]
63. Singh B, Ahmed M, Staiano AE, Gough C, Petersen J, Vandelanotte C, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024;7(1):179 [FREE Full text] [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
64. Cannon MJ, Masalovich S, Ng BP, Soler RE, Jabrah R, Ely EK, et al. Retention among participants in the national diabetes prevention program lifestyle change program, 2012-2017. Diabetes Care 2020;43(9):2042-2049 [FREE Full text] [doi: 10.2337/dc19-2366] [Medline: 32616617]
65. Hawkes RE, Miles LM, Ainsworth B, Ross J, Meacock R, French DP. Engagement with a nationally-implemented digital behaviour change intervention: usage patterns over the 9-month duration of the national health service digital diabetes prevention programme. Internet Interv 2023;33:100647 [FREE Full text] [doi: 10.1016/j.invent.2023.100647] [Medline: 37502122]
66. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
67. Waring ME, Pagoto SL, Moore Simas TA, Blackman Carr LT, Eamiello M, Libby B, et al. Delivering a postpartum weight loss intervention via facebook or in-person groups: results from a randomized pilot feasibility trial. JMIR Mhealth Uhealth 2023;11:e41545 [FREE Full text] [doi: 10.2196/41545] [Medline: 37103991]
68. Park S, Ward T, Sudimack A, Cox S, Ballreich J. Cost-effectiveness analysis of a digital diabetes prevention program (dDPP) in prediabetic patients. J Telemed Telecare 2023:1357633X231174262. [doi: 10.1177/1357633X231174262] [Medline: 37287252]
69. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
70. Mastellos N, Gunn LH, Felix LM, Car J, Majeed A. Transtheoretical model stages of change for dietary and physical exercise modification in weight loss management for overweight and obese adults. Cochrane Database Syst Rev 2014;2014(2):CD008066 [FREE Full text] [doi: 10.1002/14651858.CD008066.pub3] [Medline: 24500864]

XSL•FO
RenderX

71. American Diabetes Association Professional Practice Committee. 5. Facilitating positive health behaviors and well-being to improve health outcomes: standards of care in diabetes-2024. Diabetes Care 2024;47(Suppl 1):S77-S110. [doi: 10.2337/dc24-S005] [Medline: 38078584]
72. Elsman EBM, Leerlooijer JN, Ter Beek J, Duijzer G, Jansen SC, Hiddink GJ, et al. Using the intervention mapping protocol to develop a maintenance programme for the SLIMMER diabetes prevention intervention. BMC Public Health 2014;14:1108 [FREE Full text] [doi: 10.1186/1471-2458-14-1108] [Medline: 25346512]
73. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
74. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
75. Smith DH, O'Keeffe-Rosetti M, Fitzpatrick SL, Mayhew M, Firemark AJ, Gruß I, et al. Costs and cost-effectiveness of implementing a digital diabetes prevention program in a large, integrated health system. Perm J 2022;26(3):74-82 [FREE Full text] [doi: 10.7812/TPP/22.029] [Medline: 36038966]
76. Makroum MA, Adda M, Bouzouane A, Ibrahim H. Machine learning and smart devices for diabetes management: systematic review. Sensors (Basel) 2022;22(5):1843 [FREE Full text] [doi: 10.3390/s22051843] [Medline: 35270989]
77. Walter JR, Xu S, Rogers JA. From lab to life: how wearable devices can improve health equity. Nat Commun 2024;15(1):123 [FREE Full text] [doi: 10.1038/s41467-023-44634-9] [Medline: 38167483]
78. Gong E, Wang H, Zhu W, Galea G, Xu J, Yan LL, et al. Bridging the digital divide to promote prevention and control of non-communicable diseases for all in China and beyond. BMJ 2024;387:e076768. [doi: 10.1136/bmj-2023-076768] [Medline: 39424328]

Abbreviations

DPP: Diabetes Prevention Program
FPG: fasting plasma glucose
HbA1c: hemoglobin A1c
MeSH: Medical Subject Headings
PICOS: Population, Intervention, Comparison, Outcome, and Study design
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RoB 2.0: Cochrane Risk-of-Bias tool
RR: risk ratio
SMD: standard mean difference
T2DM: type 2 diabetes mellitus
WHO: World Health Organization

Edited by A Coristine; submitted 04.07.24; peer-reviewed by M Nagpal, K Rawlings, SA Sebastian; comments to author 16.09.24; revised version received 10.11.24; accepted 19.11.24; published 29.01.25.

Please cite as:
Wang Y, Chai X, Wang Y, Yin X, Huang X, Gong Q, Zhang J, Shao R, Li G
Effectiveness of Different Intervention Modes in Lifestyle Intervention for the Prevention of Type 2 Diabetes and the Reversion to Normoglycemia in Adults With Prediabetes: Systematic Review and Meta-Analysis of Randomized Controlled Trials
J Med Internet Res 2025;27:e63975
URL: https://www.jmir.org/2025/1/e63975
doi:10.2196/63975
PMID:

©Yachen Wang, Xin Chai, Yueqing Wang, Xuejun Yin, Xinying Huang, Qiuhong Gong, Juan Zhang, Ruitai Shao, Guangwei Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Impact of Online Interactive Decision Tools on Women's Decision-Making Regarding Breast Cancer Screening: Systematic Review and Meta-Analysis

Patricia Villain[1], PhD; Laura Downham[1], PhD; Alice Le Bonniec[1], PhD; Charlotte Bauquier[2], PhD; Olena Mandrik[3], PhD; Tom Nadarzynski[4], PhD; Lorie Donelle[5], PhD; Raúl Murillo[6], PhD; Eleni L Tolma[7], PhD; Sonali Johnson[8], PhD; Patricia Soler-Michel[9], PhD; Robert Smith[10], PhD

[1]International Agency for Research on Cancer, World Health Organization, Lyon, France
[2]Pôle de Psychologie Sociale, Inserm U1296, Université Lumière Lyon 2, Lyon, France
[3]Sheffield Centre for Health and Related Research, The University of Sheffield, Sheffield, United Kingdom
[4]School of Social Sciences, University of Westminster, London, United Kingdom
[5]College of Nursing, University of South Carolina, Columbia, SC, United States
[6]Centro Javeriano de Oncología, Hospital Universitario San Ignacio, Bogota, Colombia
[7]Department of Primary Care and Population Health, University of Nicosia Medical School, Nicosia, Cyprus
[8]Union for International Cancer Control, Geneva, Switzerland
[9]Centre Régional de Coordination des Dépistages des Cancers Auvergne-Rhône-Alpes, Site Rhône & Métropole de Lyon, Lyon, France
[10]Cancer Screening, American Cancer Society, Atlanta, GA, United States

Corresponding Author:
Patricia Villain, PhD
International Agency for Research on Cancer
World Health Organization
25 Avenue Tony Garnier CS 90627
Lyon, 69366
France
Phone: 33 4 72 73 84 40
Fax: 33 4 72 73 84 40
Email: patriciavillain1@gmail.com

Abstract

Background: The online nature of decision aids (DAs) and related e-tools supporting women's decision-making regarding breast cancer screening (BCS) through mammography may facilitate broader access, making them a valuable addition to BCS programs.

Objective: This systematic review and meta-analysis aims to evaluate the scientific evidence on the impacts of these e-tools and to provide a comprehensive assessment of the factors associated with their increased utility and efficacy.

Methods: We followed the 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and conducted a search of MEDLINE, PsycINFO, Embase, CINAHL, and Web of Science databases from August 2010 to April 2023. We included studies reporting on populations at average risk of breast cancer, which utilized DAs or related e-tools, and assessed women's participation in BCS by mammography or other key cognitive determinants of decision-making as primary or secondary outcomes. We conducted meta-analyses on the identified randomized controlled trials, which were assessed using the revised Cochrane Risk of Bias 2 (RoB 2) tool. We further explored intermediate and high heterogeneity between studies to enhance the validity of our results.

Results: In total, 22 different e-tools were identified across 31 papers. The degree of tailoring in the e-tools, specifically whether the tool was fully tailored or featured with tailoring, was the most influential factor in women's decision-making regarding BCS. Compared with control groups, tailored e-tools significantly increased women's long-term participation in BCS (risk ratio 1.14, 95% CI 1.07-1.23, P<.001, I^2=0%). Tailored-to-breast-cancer-risk e-tools increased women's level of worry (mean difference 0.31, 95% CI 0.13-0.48, P<.001, I^2=0%). E-tools also improved women's adequate knowledge of BCS, with features-with-tailoring e-tools designed and tested with the general population being more effective than tailored e-tools designed for or tested with

XSL•FO
RenderX

non-BCS participants (χ^2_1=5.1, *P*=.02). Features-with-tailoring e-tools increased both the rate of women who intended not to undergo BCS (risk ratio 1.88, 95% CI 1.43-2.48, *P*<.001, I^2=0%) and the rate of women who had made an informed choice regarding their intention to undergo BCS (risk ratio 1.60, 95% CI 1.09-2.33, *P*=.02, I^2=91%). Additionally, these tools decreased the proportion of women with decision conflict (risk ratio 0.77, 95% CI 0.65-0.91, *P*=.002, I^2=0%). Shared decision-making was not formally evaluated. This review is limited by small sample sizes, including only a few studies in the meta-analysis, some with a high risk of bias, and high heterogeneity between the studies and e-tools.

Conclusions: Features-with-tailoring e-tools could potentially negatively impact BCS programs by fostering negative intentions and attitudes toward BCS participation. Conversely, tailored e-tools may increase women's participation in BCS but, when tailored to risk, they may elevate their levels of worry. To maximize the effectiveness of e-tools while minimizing potential negative impacts, we advocate for an "on-demand" layered approach to their design.

(J Med Internet Res 2025;27:e65974) doi:10.2196/65974

KEYWORDS

breast cancer screening; decision-making; online interactive; decision aid; average risk; shared decision-making; screening participation; cognitive determinants; women

Introduction

Breast cancer constitutes 11.6% (2.309 million new cases) of global cancer incidence and 6.9% (665,684 deaths) of global cancer mortality [1].

Mammography is the most widely used screening method in breast cancer screening (BCS) programs for detecting breast cancer. These programs primarily target women at average risk of the disease. A woman is considered at average risk if she lacks risk factors associated with a significantly increased likelihood of breast cancer, such as a personal history of the disease, a strong family history, a genetic mutation known to elevate risk (eg, a *BRCA* mutation), or prior high-dose radiation therapy to the chest at a young age [2-4]. Conversely, a woman is considered at high or very high risk of breast cancer if she has 1 or more of these risk factors. Such high-risk women do not participate in standard BCS programs; instead, they undergo more personalized screening options [2,5].

Despite the proven efficacy of mammography-based BCS and national health policies promoting regular screening for women at average risk, participation in BCS remains suboptimal in many countries, falling well below the recommended rates set by the European Union and other organizations [6-8].

Numerous barriers to women undergoing BCS have been identified in the literature, such as fears of a breast cancer diagnosis or the mammography procedure itself [9-11]. For some women, participation in a BCS program may also be perceived as a dilemma due to ongoing debates surrounding BCS. Indeed, scientific and medical discussions continue about specific aspects of BCS, including the optimal screening interval [12,13]. Furthermore, public discourse often highlights controversies and potential risks associated with BCS, such as overdiagnosis and overtreatment [14,15].

The use of decision aids (DAs), either independently or as part of a shared decision-making (SDM) process, appears particularly suitable and important for both women and BCS programs. These tools help women make an "informed choice" about their participation in BCS, enabling them to make a reasoned and justifiable decision after receiving reliable and sufficient information about the procedure or examination and its associated risks [16]. DAs are evidence-based cognitive tools designed to elucidate the decision-making process and help patients clarify their values and preferences [17-19]. Similarly, SDM is described as an interactive, balanced, step-by-step discussion between health professionals (HPs) and patients, often incorporating DAs to support the dialogue [20,21].

Patients faced with decisions about cancer screening (including BCS) who are exposed to DAs tend to be better informed and more aligned with their values [19,22-26]. Although promising, evidence regarding the effectiveness of SDM remains limited, particularly in real-world settings where its implementation is often hindered by the busy and demanding schedules of HPs [27-30]. Additionally, there is insufficient evidence on when it is more beneficial for patients to engage in an SDM discussion with an HP rather than relying solely on a DA [20,31-34].

Global interest in digital health and eHealth tools, including telehealth, has surged in recent years, largely driven by the COVID-19 pandemic [35-37]. Online delivery offers the potential to enhance access to DAs for patients and HPs, particularly for those in underserved communities, while also enabling more integrated implementation of SDM. As a result, incorporating online DAs and related tools into BCS programs could support women in making informed decisions. A review of web-based DAs found these tools to be effective in increasing women's knowledge and ability to make informed choices while reducing decisional conflict or dilemmas about BCS [38].

However, to better inform future BCS programs about the utility and implementation of online tools for women's decision-making, it is necessary to go beyond Yu et al's review [38] and provide additional evidence. First, it is important to evaluate not just web-based DAs but also all types of online interactive tools (e-tools) to fully explore their potential for supporting women's decision-making about BCS [38-40]. An overall assessment of the effect of these e-tools, including web-based DAs, on women's participation in BCS and important cognitive determinants in decision-making (eg, women's attitudes or perceptions of risk) was lacking [19,22,25,26,38]. Additionally, there was a need for a comprehensive evaluation

XSL•FO
RenderX

of factors associated with the greater utility and effectiveness of these e-tools to support more informed BCS programs. This includes exploring aspects related to information delivery (eg, mobile vs computer-based e-tools, degree of message tailoring), user characteristics (eg, nonadherent to BCS vs the general population), and the methods used to evaluate these e-tools (ie, the outcomes assessed to measure their impact on decision-making). Finally, evidence of the effect of these e-tools on the SDM process was also missing.

Through a combined systematic review and meta-analysis, our research aimed to address these evidence gaps and synthesize the scientific literature on e-tools designed to support decision-making in the context of BCS by mammography. The primary objective of this review was to inform future BCS programs about the utility and implementation of such e-tools.

Methods

Review Design

This review was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42020164479) [41]. The main deviation from the original protocol was the inclusion of studies without comparator arms. This change was made for 2 reasons: (1) a preliminary review of the databases indicated that the total number of studies might be limited if only those with comparators were included, and (2) to provide the most comprehensive assessment of the effects of e-tools, including web-based DAs, on women's participation in BCS and important cognitive determinants in decision-making. The type of design for each study included in this review was considered when assessing the overall certainty of the evidence using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) methodology (outlined below). A workshop held in December 2019, organized by PV, on developing and implementing e-tools as DAs to support women's decision-making about BCS, played a significant role in shaping the development of this review. We conducted our systematic review in accordance with the 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [42] (see Table S1 in Multimedia Appendix 1).

Study Search, Selection, and Data Extraction

We systematically searched the MEDLINE (via PubMed), PsycINFO, Embase, CINAHL, and Web of Science databases from August 2010 to August 2020, with an update in April 2023 (see Appendix S1 in Multimedia Appendix 1). We limited the search to articles published from 2010 onward to minimize the risk of including e-tools that may have become technologically obsolete—those using outdated technology and no longer considered useful, efficient, or functioning well compared with newer alternatives [43]. Additional websites (eg, the International Patient Decision Aid Standards [IPDAS] website) were also searched nonsystematically. The following inclusion criteria were applied: (1) women considered to be at average risk of breast cancer, that is, those without risk factors associated with a significantly increased risk of the disease [2-4]; (2) the tool, whether identified as a DA or not, was used for women's decision-making about BCS, aimed to inform users at least about BCS by mammography, and was an e-tool defined as one that runs on the internet, computer, phone/tablet, or other electronic device and is interactive (ie, allowing users to "participate in modifying the form and content of a mediated environment in real time") [39,40]; and (3) the study, of any design, reported as primary or secondary outcomes any of the following: women's participation in BCS by mammography (behavior), intention to participate, knowledge, attitudes, self-efficacy, worry, perceptions of risk, regrets, decisional conflict, and informed choice—measures shown to reflect decision-making and behavior and to be important in decision-making [19,22,25,26]. Exclusion criteria were as follows: (1) participants at high risk of breast cancer, that is, those with 1 or more risk factors associated with a significantly increased risk of the disease [2-4]; (2) screening limited to breast examination by an HP or self-examination; (3) non–peer-reviewed publications; and (4) studies focused on BCS cessation.

The search yielded 16,061 records, which were managed using Covidence software [44]. Three independent reviewers (PV, ALB, and CB) assessed titles and abstracts, followed by full-text evaluation against the study inclusion criteria. A total of 227 papers were selected for full-text eligibility assessment, and 196 were subsequently excluded. A data extraction form was used (see Appendix S2 in Multimedia Appendix 1); it was completed by a main reviewer (PV or ALB) and independently assessed by 2 additional reviewers (PV, LD, ALB, or CB).

Qualitative and Quantitative Syntheses

Extracted data were synthesized both qualitatively and quantitatively, as detailed later.

The qualitative synthesis aimed to provide an overview of the characteristics of the e-tools and a description of the reported outcomes and the instruments used to measure them. These are reported in the Results section. Additionally, the characteristics of the studies included in this review and the study populations are reported in Tables S2 and S3, respectively, in Multimedia Appendix 1.

Quantitative synthesis was conducted to report the effect of the e-tools on women's decision-making about BCS. Main results are reported in the Results section, with additional results reported in Appendix S4 and Figures S1-S8 in Multimedia Appendix 1.

Description of the e-Tools

The main descriptive variables were identified through discussions and consensus among the authors. These variables were considered important as they could potentially influence the effects of e-tools on women's decision-making about BCS. The e-tools identified in this review were classified as either "tailored"—when the e-tools provided individualized or personalized notifications, messages, or information based on an individual's assessment [40,45,46]—or "features with tailoring"—when the e-tools presented some degree of tailored information but not based on an individual assessment [40].

XSL•FO
RenderX

Effect of the e-Tools on Women's Decision-Making About BCS

Overview

The effect of the e-tools was assessed through the following outcomes: women's participation in BCS by mammography (behavior), intention to participate, knowledge, attitudes, self-efficacy, worry, perceptions of risk, regrets, decisional conflict, and informed choice. These outcomes have been shown to reflect decision-making and behavior and are particularly important in the context of BCS [19,22-26,38]. The effects of the e-tools on SDM or communication with HPs were also evaluated.

For each outcome, quantitative synthesis was conducted and reported through a 2-step process. First, where possible, we performed meta-analyses with the randomized controlled trials (RCTs) identified in this review, following Cochrane guidance [47]. Additionally, the synthesis without meta-analysis (SWiM) guideline was used to synthesize evidence from RCTs not included in the meta-analyses or other study designs (see Table S2 in Multimedia Appendix 1) [48]. Second and finally, for each outcome, the overall certainty of the evidence collected through the meta-analyses and SWiM exercise was independently evaluated by PV and LD using the GRADE methodology [49-51]. Results from the risk-of-bias assessment (discussed later) were used to downgrade the overall certainty of the evidence, if necessary. Because of the low total number of studies included in this review, we did not downgrade the overall certainty of the evidence, assessed with GRADE, based on the imprecision (number of studies, SD) criteria, except when only 2 studies (meta-analyses) or 1 study were available. This applied to the "intention not to participate," "worry," "perception of risk," and "decisional conflict" outcomes assessed through meta-analyses, as well as the "self-efficacy," "regret," and "discussion/SDM" outcomes. Any discrepancies between the 2 independent reviewers were resolved through the intervention of a third reviewer (RS).

Meta-Analyses and Risk of Bias

Meta-analyses were conducted with RCTs following Cochrane guidance [47] (see also Part A in Appendix S3 in Multimedia Appendix 1).

Data were extracted from the original publications, and additional information was obtained by directly contacting the authors when necessary. Only data related to BCS by mammography, as reported by the authors for "women," were extracted.

PV and LD, along with ALB or CB, independently performed the risk-of-bias assessment for all RCTs included in this review using the revised Cochrane Risk-of-Bias 2 (RoB 2) tool for RCTs [52]. Any discrepancies between the 2 independent reviewers were resolved through the intervention of a third reviewer. As a result of the low number of studies, none were excluded based on RoB 2 results. The level of risk for each RoB 2 domain (A: randomization process; B: deviations from the intended protocol; C: missing data; D: outcome measurement; E: reporting results) and overall risk (F) were evaluated for each RCT and reported on the forest plots (discussed below) as low (green), moderate (yellow), or high (red).

To minimize sources of heterogeneity between studies, we chose to include only RCTs that reported the most similar outcomes based on their definitions and the instruments used to measure them in the meta-analyses described in this review. Therefore, contrary to Yu et al's review [38], we reported the decisional conflict outcome only for RCTs using the "SURE" scale and did not report the pooled effect for the "regret" outcome.

We used Review Manager software (RevMan, version 5.4.1; Cochrane) [53] to conduct meta-analyses, perform subgroup analyses, and display forest plots, including risk-of-bias assessments.

We reported pooled estimates as mean difference (MD) for continuous variables (adequate knowledge, worry, and accurate perception of risk) and relative risk/risk ratio (RR) for dichotomous variables (other outcomes). Heterogeneity or inconsistency between studies was assessed using the I^2 statistic. A fixed-effect (FE) model was applied when low heterogeneity was found ($I^2 \leq 30$), while a random-effects (RE) model was used in other cases. Results from meta-analyses with intermediate or high heterogeneity between studies ($I^2 \geq 35$) were further explored to strengthen the validity of our findings.

Exploration of Heterogeneity and Subgrouping Analysis

Where meta-analyses showed intermediate or high heterogeneity between studies ($I^2 \geq 35$), further analyses were conducted. The main results are reported in the text (see also Figures S1-S8 and Appendix S4 in Multimedia Appendix 1). Two approaches were used to explore the causes of intermediate/high heterogeneity [47]: (1) exclusion of 1 study when it differed from the others in the meta-analysis based on either the type of e-tool used (eg, web application or mobile app) or the control used (eg, a website or a video). We also explored the exclusion of studies with a high risk of bias(es). Exclusion of Lee et al's [54] study, which used a mobile app downloaded on a mobile phone, or Roberto et al's study [55], which used a website as a control, in the meta-analysis assessing short-term participation or adequate knowledge, respectively, was effective in reducing heterogeneity (reported); (2) additionally, subgroup analyses were conducted based on various variables that could have explained the source of heterogeneity, for example, characteristics of the e-tools such as the type or degree of tailoring, the type of outcomes or instruments, or characteristics of the study populations such as age, ethnicity/rurality, and nonparticipation in previous BCS by mammography (see also Table S3 in Multimedia Appendix 1).

Besides exploring heterogeneity, subgroup analysis was also conducted to examine the effect of different variables (described above) on the reported outcomes.

Results

Overview of Included Studies and e-Tools

We identified 31 published papers/studies (Figure 1; also see Table S2 in Multimedia Appendix 1) [54-84], representing 22

XSL•FO
RenderX

different e-tools [54-75]. The majority of these were designed and evaluated in the United States (n=16) and tested through RCTs (n=14). Study populations are detailed in Table S3 in Multimedia Appendix 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram. The diagram was generated using the Covidence software [44].

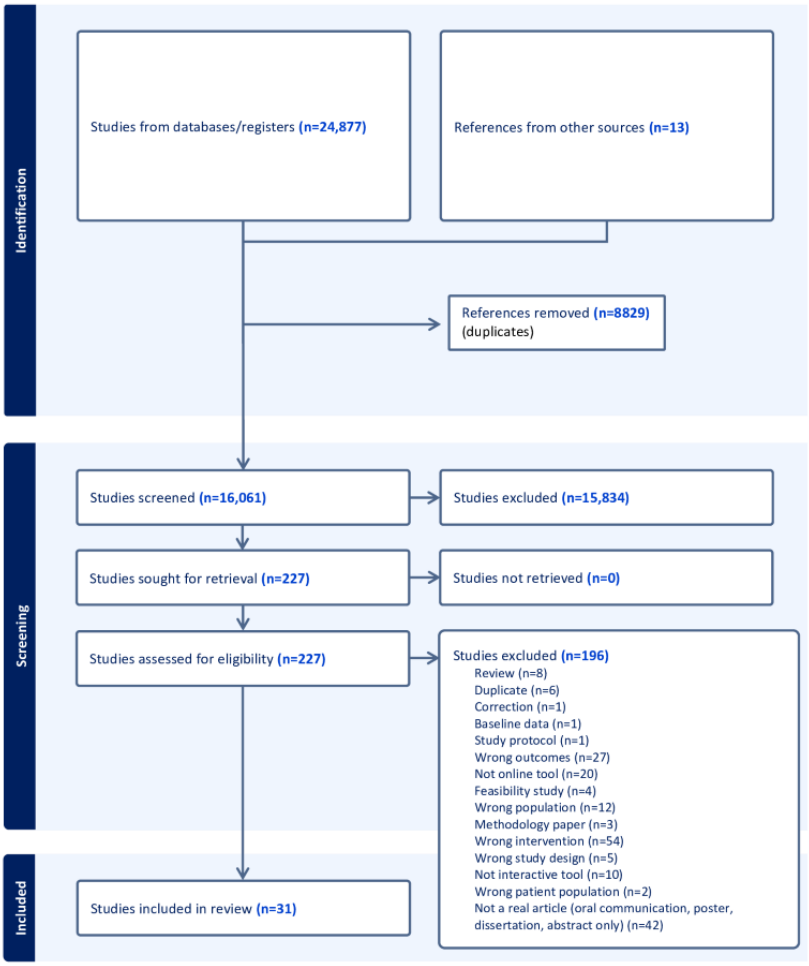

Description of the e-Tools

Among the 22 e-tools, the majority (n=16) were web applications, either standalone or integrated into patient portals [55,56,58,59,61,63,65,67-75] (Table 1; see also Table S4 in Multimedia Appendix 1).

Two e-tools used artificial intelligence (AI)–based virtual doctors [58,64].

Most of the e-tools (16/22; Table 1) were tailored to the individual's (1) breast cancer risk (n=7) [63,65,71-75]; (2) barriers, beliefs, or screening status identified at baseline (n=8) [56,57,59,61,64,66-68]; or (3) preferences regarding the content, number, and timing of daily messages [54]. Of these 16 tailored e-tools, 8 targeted nonparticipants in BCS (see Table S3 in Multimedia Appendix 1) [54,56,57,59,63,66-68]; 11 out of the 16 tailored e-tools included specific features designed to facilitate or enhance opportunities and the quality of discussions about BCS between women and HPs (Table 1) [54,59,61,63-65,67,72-75].

The remaining 6 e-tools (out of 22) were not tailored to individual users and were classified as "features-with-tailoring" e-tools (Table 1) [55,58,60,62,69,70]. None of these tools were evaluated specifically with BCS nonattenders (see Table S3 in Multimedia Appendix 1). The tailoring features in these tools were designed to either (1) address BCS screening barriers for specific subpopulations (n=2) [60,62] or adapt the communication style of an AI-based virtual HP to meet the needs of specific user groups [58], or (2) prompt users to identify their personal beliefs and values regarding BCS (Table 1) [55,69,70].

XSL•FO
RenderX

Table 1. Description of the e-tools[a] used to help women's decision-making about breast cancer screening by mammography.

Descriptive variables	Characteristics of e-tool
Theoretical model[b]	• Not reported [55,58,59,61,62,67,69,70,74,75] • The Transtheoretical Model of Behavior Change [56,57,64,66,68], the Health Belief Model [54,56,57,66], the Decisional Conflict Theory [72,73], the Exemplification Theory [63,71], the Elaboration Likelihood Model of Persuasion [71], the Self-Regulation Model [65], the Theory of Planned Behavior [56], the Conceptual Framework to Measure Engagement [60], and the Fogg Behavioral Model [54]

Decision aids[c]	• Identified as decision aids [55,61,63,67,69-75]
	  • All except 2 e-tools [61,71] were reported to be developed based on international standards: either the International Patient Decision Aid Standards Framework [55,63,69,70,72-74] or the Ottawa Decision Support Framework [67,75]
	  • Quality assessment against International Patient Decision Aid Standards criteria was performed by the authors themselves [63,75] or by others and through a systematic review [85] in which 4 e-tools were evaluated and assessed to be of either high overall quality [69,70,75] or below the overall mean quality [72]
	• Not identified as decision aids [54,56-60,62,64-66,68]

Type[d]	• Standalone web-based applications [55,56,58,63,65,68-71,74,75], which were optimized for smartphones [72]; Fissler et al's e-tool [58] used an artificial intelligence–based virtual doctor • Patient portal: either a new one [61] or "embedded" or "linked" to an existing one [59,67,73] • A mobile app either on a smartphone with a global positioning system [54] or on an iPad [64]; Walsh et al's e-tool [64] used an artificial intelligence–based virtual doctor • Interactive DVD [57,66] • Social media–based: either Facebook [60] or WhatsApp [62]
Information provided	• Breast cancer screening and breast cancer [54,55,57,58,60,62,63,68-75] • Breast cancer screening and breast cancer with additional information[e], such as other cancer screening [56,59,61,64,66,67], or other prevention messages (eg, diet) [59,61,64,65]

Degree of tailoring[f]	• Tailored
	  • To address woman' breast cancer risk[g] [63,65,71-75]; some e-tools also allowing to clarify value[h] [63,65,72-75]
	  • To address individuals' barriers, beliefs, or screening status at baseline [56,57,59,61,64,66-68], including the stage of change based on the Transtheoretical Model [68]
	  • To adapt the content, number, and timing of daily messages to each individual [54]
	• Features-with-tailoring
	  • To address the screening barriers specific to a group of the targeted population [60,62]
	  • To target the communication style of the factious health professional to the needs of a specific group of patients [58]
	  • To allow individuals to clarify their values[h] [55,69,70]

XSL•FO
RenderX

Descriptive variables	Characteristics of e-tool

Access to health professionals[i]	• The e-tools were part of clinical pathways or referred to actual health care services [55,59,61,63,65,67,70,72,73,75]
	  • Being a patient portal or embedded in an existing one [59,61,67,73]
	  • patients with medical preappointments were automatically invited to use the tool [67]
	  • patient being directly informed of a care gap (eg, delay in getting breast cancer screening), and could directly book an appointment through the portal [59]
	  • both patients and health professionals were informed automatically when the former were at a higher risk of breast cancer [73]
	  • individual recommendations targeting 18 preventive health services were provided [61]
	  • Being available before/after a booked medical appointment: before [72,73], or most probably only or at the first instance at clinics [63,75], or in advance of a prescreening appointment made by breast cancer screening programs [55,70]
	  • Offering counseling sessions online or discussions with health professionals [65]
	• Used artificial intelligence–based virtual doctor to simulate an appointment between a health professional and their patient [58,64], designed to encourage the user to continue [64] or to improve [58] discussion with a health professional
	• Providing a personalized report
	  • All being downloadable [55,63,64,67,70,72-75] or probably being downloadable [61]
	  • Either a summary with a list of questions [61,63,64,67,72-75] or the results of the value clarification exerciseh [55,70]
	  • When the tools were part of clinical pathways, reports were put directly in the patient folder [63,64], a specific appointment to discuss was offered [72], or the report was sent to the health professional through the patient portal [61,67,73]
	  • Women were encouraged to either share or discuss this report with their health professional [61,63,67,72-75], or with people they trust [55], or to talk with a health professional about any concerns [64,70] or in case of indecision about screening [70]
	• Being tailored e-tools based on risk: women at higher risk receive automatic messages to follow-up including talking with health professionals [65,72-75], and health professionals are informed automatically [73]
	• Includes an option to contact a health navigator for assistance in navigating through cancer screening information, addressing technical problems, and providing transportation and interpretation services [54]
	• With a global positioning system facilitating cancer screening clinics localization and other information [54]

[a]The e-tools were categorized into groups and subgroups based on the descriptive variables as indicated (see also Table S4 in Multimedia Appendix 1).

[b]Any psycho-theoretical model reported to be used to develop the e-tool.

[c]We classified the e-tools as "identified as decision aids" when the authors explicitly used "DA" or "decision-making" to describe the e-tool, or when the e-tool contained a decision aid/decision-making module or a link to decision aids. The remaining tools were classified as "not identified as decision aids." A rapid assessment of these tools against the 5 previously described decision aids criteria [86], that is, (1) information about options, decisions, and outcomes, including benefits and harms; (2) evidence-based information on options; (3) probabilities; (4) value clarification; and (5) decision guidance, did not provide sufficient information to determine whether or not these tools could be classified as decision aids.

[d]The tools were classified as previously described [40,87].

[e]In this group of e-tools, the study population was either only women [56,65,66] or both men and women [59,61,64,67].

[f]Tools were classified as either "tailored" or "features-with-tailoring" e-tools, as previously defined [40,45,46]. Tailored tools provided individualized or personalized notifications based on an individual's assessment. This assessment could be conducted through participants' responses to questionnaires, which were either external to the e-tool and completed before its use or integrated into the e-tool itself, addressing factors such as the individual's breast cancer risk or other characteristics. By contrast, "features-with-tailoring" e-tools offered some degree of personalization or tailoring, but it is not based on individual assessments.

[g]Risk was estimated using 2 validated instruments: the Breast Cancer Risk Assessment Tool for personal risk estimates and the Breast Cancer Genetics Referral Screening Tool for assessing familial breast cancer risk. In 1 study, a different instrument was used, and it was unclear whether the risk estimation was communicated to the women involved [74]. Among the 7 tailored-to-risk tools, 4 utilized risk assessment to exclude women with above-average breast cancer risk from accessing the decision aid, instead directing them to receive messages for appropriate follow-up based on best practices [72-75]. For the remaining 3 tools, risk assessment was conducted online to personalize the features of the tools according to each woman's individual risk [63,65,71]. One tool provided 6 levels of tailored-to-risk messages, and its effects were evaluated [71].

[h]The tools aimed to encourage users to identify their personal beliefs and values regarding breast cancer screening. The methods used were all explicit, meaning they required active user participation—such as moving sliders, assigning weights to scales, or inputting numbers—to express their values, preferences, and concerns [88].

[i]Specific features aimed to facilitate access to or communication with health professionals.

XSL•FO
RenderX

Effect of the e-Tools on Women's Decision-Making About BCS

Overview of the Outcomes

The effect of the e-tools on women's decision-making about BCS was assessed through various outcomes in the studies included in this review. A comprehensive overview of these outcomes and the instruments used to measure them are provided in Table 2.

The synthesized results regarding the effects of the e-tools are presented in the following paragraphs, with the outcomes listed in the same order as in Table 2. We encourage readers to refer to this table for the definitions of each outcome.

XSL•FO
RenderX

Table 2. Outcomes and instruments to assess the effect of e-tools on women's decision-making about breast cancer screening by mammography.

Outcomes[a]	Definitions[b] and instruments[c]

Participation (behavior) [54-57,59,61,63-67,70]	• Authors reported the rates of women who have participated in breast cancer screening by mammography.
	• Participation in breast cancer screening was reported by the authors as "participation rate" [55], "adherence" [57], "uptake" [54,63,70], "receipt" [54,56], "screening behavior" [65], "getting screened" [67], "being up to date" [64,66,67], or "screening gap closure" [59].
	• Instruments used to measure women's participation were self-report questionnaires/surveys [54,55,65,70], health records [59,63,67], or both, with self-reports used to supplement missing medical records [56,57,61,64,66].
	• In 1 study [70], the authors differentiated participation in the breast cancer screening program, which was reported, and opportunistic screening.
	• Depending on the study, participation was assessed as follows:
	  • At short-term (up to 6 months): measured at 0.5-2 months [55], at 3 months [59,67,70], at 4 months [61], or at 6 months [54,56,57]. In Krist et al [67], a non–randomized controlled trial, measures were done after each invitation phase.
	  • At long-term (12-16 months): measured at 12 months [63,65,66], at 14 months [64], or at 16 months [61]. It must be highlighted that in Champion et al's [66] study, participation after the 12-month window was classified as "no screening."

Intention to undergo breast cancer screening [54,55,60,68-74,76]	• Intention, as defined in the Theory of Planned Behavior [89], is assumed to capture the motivational factors that influence a behavior; it indicates how hard people are willing to try or how much effort they are planning to exert to perform the behavior. Intention assessment was reported as the rates of women with intention/willingness to undergo breast cancer screening. Reder and Kolip [70] differentiated the breast cancer screening program and opportunistic screening and reported intention to participate in the breast cancer screening program. • In the randomized controlled trials, intention was assessed just after using the e-tool [54,68-74] or 7-10 days after [55]. • Generally, intention was assessed through 1 or 2 items in a questionnaire/survey with anchors defined by the authors themselves to be either cross-checked [54,55,68,70-74,76] or evaluated with a Likert scale [60,69]. • Intention was formulated either as a general statement [55,73] or as a more specific one and anchored to a specific time or age. Specific time(s) was/were linked to the near future either now [69], in the next 3 months [70], in the coming/next year [60,68], or at different possible times (within 1 month, 3 months, or up to 1 year) [54]. In 1 study it was evaluated for a more distant future (in the next 1-2 years) [74]. When anchored to age, women's intention was evaluated to start or continue in their forties [71,72], or to start at 50 years [71]. We used the cumulative value of intention to assess Lee et al's [54] and Seitz et al's [71] tools in the meta-analysis (see Part A in Appendix S3 in Multimedia Appendix 1): intention "to get screened within 3 months to 1 year" and intention "to start or continue in the forties, or to start at 50 years old" were reported.
Intention *not* to undergo breast cancer screening (negative intention) [69,70]	• It was assessed in 2 randomized controlled trials and reported as the rates of women with intention/willingness not to undergo breast cancer screening [69,70].
Knowledge [54,55,62-65,67,69,70,72,77]	• Knowledge was assessed about breast cancer screening (most studies) and, in some studies, about breast cancer. • With Walsh et al's [64] and Krist et al's [67] e-tools, participants were asked to evaluate whether the tool improved their own knowledge; the authors reported rates of participants presenting different degrees of agreement [64,67,77]. With Eden et al's [72] e-tool, women rated how much new information they learned from the e-tool on a scale from 1 (lowest) to 10 (highest), with the authors reporting median and IQR. • In the remaining 7 e-tools (ie, 6 randomized controlled trials [54,55,63,65,69,70] and 1 mixed method study [62]), the instrument used to measure knowledge was more complex and heterogeneous; the instrument was a series of 5-18 items created or adapted by the authors to assess participants' knowledge accompanied by anchors. The format of the anchors was either multichoice responses [62,63], both multichoice and numerical responses [55,69,70], true/false responses [54], or response scales of 1-4 [65]. Lee et al [54] used a validated knowledge instrument, which was revised to reflect the current American Cancer Society's breast cancer screening guidelines. Adequate knowledge was reported as a quantitative value to reflect the number of items the participants answered correctly [54,55,62,63,65,69,70]. In some studies, the percentage of women with adequate knowledge was also reported; those were defined as women providing correct answers to at least half plus 1 of the total number of questions [55,69,70].

XSL•FO
RenderX

Outcomes[a]	Definitions[b] and instruments[c]
Attitudes about breast cancer screening [55,58,69,70]	• Refers to the degree to which a woman has a favorable or unfavorable evaluation or appraisal of an undergoing breast cancer screening program [16,89]. • In 4 randomized controlled trials [55,58,69,70], the instrument to measure attitudes toward undergoing breast cancer screening was adapted from Marteau et al's [16] definition and measure. It was either 4 items [55,58,70] or 6 items from Dormandy et al's scale [69,90]; all these items were linked to a Likert scale. In 1 of those randomized controlled trials [58], the authors reported attitudes score, with higher scores reflecting higher positive attitudes toward undergoing breast cancer screening. In the other 3 randomized controlled trials [55,69,70], Marteau et al's [16] method was used. A predefined threshold was set to differentiate women with positive attitudes [55,70] or values [69], that is, women scoring higher than the threshold, from women with negative attitudes [55,70] or values [69], and the percentage of women with positive attitude was reported.
Self-efficacy [54,72]	• Confidence in one's ability to take action as described in the Health Belief Model [91]. Also includes the concept of perceived behavioral control in the Theory of Planned Behavior models, referring to the perceived ease or difficulty of performing the behavior and assuming to reflect past experience as well as anticipated impediments and obstacles [89]. • Instruments used to measure self-efficacy were validated instruments. They were either the 8 related items from the Champion's Health Belief Model Scale [54] or the Decision Self-Efficacy Scale [72]. The level of self-efficacy was reported. One component of the Decision Self-Efficacy scale measures self-confidence or belief in one's ability to make decisions and participate in shared decision-making.
Worry [63,67,69,71,75,78]	• Assessment of emotional reactions related to breast cancer was exclusively based on assessing changes in anxiety, worry, or fear either in general, "feeling fear and worry" [69,71,78], or specifically breast cancer [63,67,75]. • In Krist et al's [67] study, the level of fear and worry was assessed through a series of questions related to breast cancer prompted by the e-tool. The authors then evaluated whether discussion with health professionals was helpful to reduce women's fears and worries if fears/worries were identified. • For other e-tools, instruments used to measure worry were heterogeneous. It was a validated instrument: either the Lerman Breast Cancer Worry Scale [63] or adapted from the McCaul Breast Cancer Worry Scale [75]; a Likert scale was used to rate the answers, and the means of scoring was reported. In other cases, it was created by the authors themselves and was generally 1-2 questions linked to a Likert scale [69,71,78]. Seitz et al [71,78] adapted a previous instrument.
Risk perceptions [54,63,71,75]	• Personal risk estimation to develop a breast cancer. It could also include "perceived susceptibility" from the Health Belief Model (ie, beliefs about the chances of experiencing a risk or getting a condition or disease) [91]. • In the randomized controlled trial by Lee et al [54], risk perception was reported through "perceived susceptibility" from the HBM model. In 3 additional e-tools, all tailored-to-risk (Table 1), the authors evaluated women's risk perceptions differently [63,71,75]. In Elkin et al's [75], the participants were all at low-to-average risk based on their personalized risk estimates; the authors assessed the rates of women with this adequate "accurate" perception of their own risk. To evaluate the 2 other e-tools, instruments were adapted from previous studies/scales. Participants were asked to evaluate their chance of getting breast cancer as a percentage (0% being no chance and 100% being sure of getting cancer) [71], or as frequency (ie, a number out of 1000) [63,71]. The difference between the risk perception evaluated by the women itself and objective risk estimates was measured and reported by the authors, with a decrease of difference reflecting an increase in women's accurate perception of risk.
Quality of decision	
Regret [63,70]	• The effect of the e-tool was evaluated on decision regret [70] or on anticipated regret [63]. • Decision regret describes regret associated with self-recrimination around having made a bad decision or regret associated with the knowledge that another choice would have resulted in a better outcome. • Anticipated regret measure is future oriented; anticipated inaction regret was associated with engaging in protective health behaviors, while anticipated action regret was associated with nonengaging. Findings suggest that anticipated inaction regret is more strongly felt and has more reliable associations with behavioral intentions and health behaviors.

XSL•FO
RenderX

Outcomes[a]	Definitions[b] and instruments[c]
Decisional conflict [55,63,70,72,74,75]	• Personal uncertainty about which option to choose. • Decisional conflict is a multidimensional outcome measuring either 4 dimensions (ie, being "SURE[d]") or 5 (previous dimensions plus "effective decision-making"), and measured with the SURE scale and the Decisional Conflict Scale, respectively. These 2 instruments are validated and negatively correlated scales [92]. • With the SURE scale, a score of ≤3 indicates decisional conflict and a score equal to 4 an absence of conflict; the authors reported the number of women with a decisional conflict. This instrument was used in 2 randomized controlled trials [55,70]; whereas in the study by Reder and Kolip [70], the instrument was specifically about the intention to undergo breast cancer screening, this was not the case in the study by Roberto et al's [55] in which the instrument was more general and about "decision." • The Decisional Conflict Scale used was either the low literacy version (ie, including 10 questions with 3 response categories) [72,74] or the traditional one (ie, including 16 statements with 5 response categories) [63,75]; Decisional Conflict Scale scoring ranged from 0 (no decisional conflict) to 100 (ie, extremely high decisional conflict or poor decision process); means of scoring was reported to reflect the decisional conflict.
Informed choice [55,69,70]	• Based on Marteau et al's [16] definition: "an informed choice to undergo a screening test occurs when an individual has a positive attitude towards undergoing a test, has relevant knowledge about the test and undergoes it. An informed choice to decline a test occurs when an individual holds a negative attitude towards undergoing a test, has relevant knowledge about the test and does not undergo it. The choices that occur when individuals do not have relevant knowledge or when their attitudes are not reflected in their behavior, are uninformed." • Following Marteau et al's [16] definition, only 1 study [70] reported informed choice regarding participation in breast cancer screening. Informed choice was usually reported toward intention to undergo breast cancer screening [55,69,70]. "Informed choice about intention" was then reported in our meta-analysis as the measure of informed choice. Except for Roberto et al [55], results regarding informed choice were reported in the original scientific publications only for participants who had made a decision about their intention to participate, excluding those who remained undecided [69,70]. This was made to strictly follow Marteau et al's [16] definition (ie, measure of informed choice only among those who made either a positive or a negative decision). In our meta-analysis, we reported informed choice over the total number of women for whom data related to informed choice measurement (ie, knowledge, attitude, and intention) were available.
Shared decisions and communications with health professionals [64,67,72,75]	• Eden et al [72] assessed the readiness of women to make a shared decision using the Graham Preparation for Decision Making scale. A higher level of scoring indicated that the women felt the decision aid prepared them to make a shared decision with clinical providers. • Among the 22 e-tools, including the 6 identified as preparing for or facilitating shared decision-making [63,64,72-75], none was evaluated against whether a shared decision-making discussion occurred. Instead, the tools were assessed by measuring women's readiness to discuss mammography [64], whether this discussion occurred [64,75], or the characteristics of the discussion including its quality [67]. The perspective of the health professionals was also evaluated regarding the 2 latest variables [67,75].
Other outcomes	• Few studies assessed outcomes related to the stage of adoption or stage of decision other than behavior or intention, such as "having made a decision" [70] or "readiness to be tested" [64]; data were not reported here. Perceived quality of life was reported in 1 study [65]. Satisfaction with the decision was not reported. Among implementation outcomes, "Reach" (ie, how the tool reached target populations) was the most reported after efficacy/effectiveness (detailed above) and through a variety of instruments.

[a]The outcomes are reported in the following order: first, we present participation in breast cancer screening (behavior), followed by related determinants/factors based on the main theories, models, and concepts used to explain or understand cancer screening behavior [89,91]. Among those, intention, identified as the most proximal determinant of behavior in the Theory of Planned Behavior [89], is reported first. Next, we include outcomes related to the "quality of decision," such as "regret" and the composite outcomes "decisional conflict" and "informed choice." The outcomes reported last are those that were less frequently assessed in the included studies.

[b]Although reported by the authors, we did not include some studies; the reasons are detailed in Part B in Appendix S3 in Multimedia Appendix 1.

[c]"Instruments" were any item (eg, scale, questionnaire, or medical records) used to measure the outcomes.

[d]SURE: Sure of myself (Uncertainty), Understand information, Risk-benefit ratio (value clarity), and Encouragement.

Participation in BCS (Behavior)

Our meta-analysis found that e-tools, including a mix of tailored-not-to-risk e-tools (ie, tailored e-tools but not to the individual's breast cancer risk; n=4) and features-with-tailoring tools (n=2), did not significantly increase women's participation in BCS up to 6 months after using the e-tools (RR 1.09, 95% CI 0.97-1.23, P=.16, I^2=78%, RE model) compared with the control group, which included either usual care [54,56,57,61,70]

XSL•FO
RenderX

or a control website [55] (see Figure S1A in Multimedia Appendix 1). Excluding Lee et al's study [54] significantly reduced heterogeneity (see the "Methods" section) without altering the overall pooled result (RR 1.03, 95% CI 0.98-1.08, P=.28, I^2=14%, FE model; see Figure S1B and S1C in Multimedia Appendix 1). No subgroup differences were observed between tailored and features-with-tailoring e-tools (χ^2_1=1.2, P=.28; see Figure S1D in Multimedia Appendix 1). Two interventional studies without concurrent controls that assessed tailored-not-to-risk e-tools reported an increase in BCS participation among users compared with nonusers [59,67].

Our meta-analysis further demonstrated that e-tools (n=5), all tailored to either risk (n=2) or other variables (n=3), increased women's participation in BCS assessed at long-term follow-up (12-16 months after tool use; RR 1.14, 95% CI 1.07-1.23, P<.001, I^2=0%, FE model) compared with the control group, which included usual care [61,63,65,66] or a control video [64] (Figure 2A). Subgroup analysis indicated a significant difference in effect between tailored-to-risk e-tools and other tailored e-tools, while highlighting a high heterogeneity within one subgroup (χ^2_1=23.54, P<.001, risk-based group: I^2=0%, non-risk–based group: I^2=98%; see Figure S2 in Multimedia Appendix 1).

Figure 2. Forest plots of studies comparing the effect of e-tools with that of control on women's (A) participation in breast cancer screening (BCS) by mammography at long term [61,63-66], (B) intention to undergo BCS [54,55,68-71], (C) intention *not* to undergo BCS [69,70], and (D) adequate knowledge about BCS [54,63,69,70]. We reported in (B) the data corresponding to Lin et al.'s "tailored message intervention" (TMI) [68] and Seitz et al.'s "extended information with untailored exemplars" e–tools [71]. Data for the other Lin et al.'s [68] and Seitz et al's [71] e-tools are presented in in Appendix S4 in Multimedia Appendix 1. In (D) subgrouping was based on tailoring complexity of the e–tools (ie, tailored e-tools compared with features–with–tailoring e–tools), which end up being identical to subgrouping based on previous breast cancer screening status (ie, nonparticipants in BCS [tailored e–tools] compared with general population [features–with–tailoring e–tools]). Control was usual care, except in the studies by Lin and Wang [68], Roberto et al [55], and Seitz et al [71] (website), and Walsh et al [64] (video). The risk of bias was assessed using the Risk-of-Bias 2 (RoB 2) tool [52]. The risk level for each domain of RoB 2 (A: randomization process, B: deviations from the intended protocol, C: missing data, D: outcome measurement, and E: reporting results) and the (F) overall risk were evaluated as low (green), moderate (yellow), or high (red). M-H: Mantel-Haenszel.

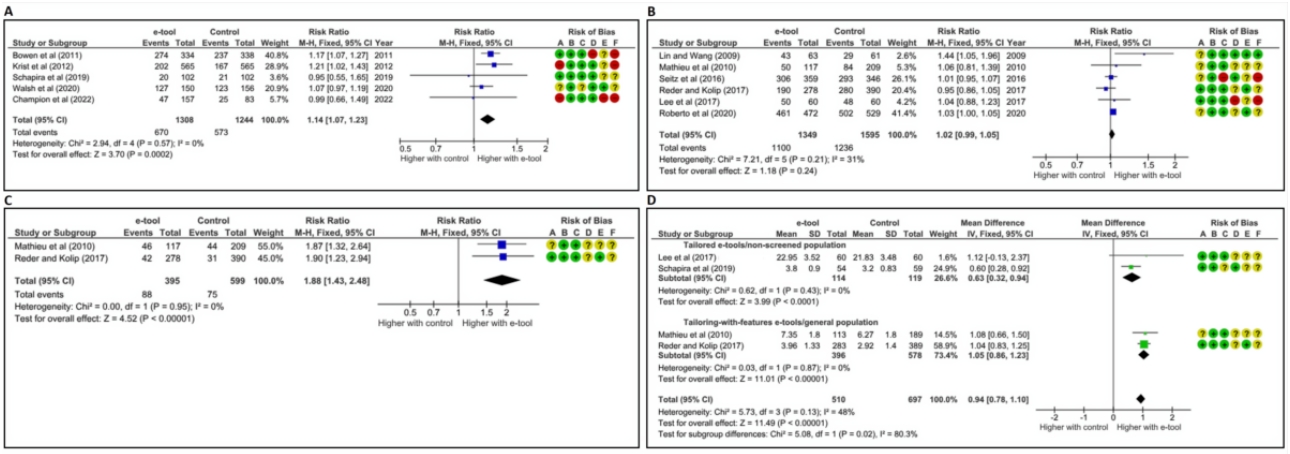

Intention

Our meta-analysis, which incorporated 6 RCTs, showed that a mix of tailored (n=3, with 2 not based on risk) and features-with-tailoring (n=3) e-tools did not affect the rate of women intending to undergo BCS (RR 1.02, 95% CI 0.99-1.05, P=.24, I^2=31%, FE model) compared with the control group, which included either usual care [54,69,70] or a control website [55,68,71] (Figure 2B; also see Appendix S4 in Multimedia Appendix 1). No subgroup differences were observed between tailored and features-with-tailoring e-tools (χ^2_1=0.6, P=.43; see Figure S3 in Multimedia Appendix 1). Similarly, no significant difference in intention was reported in 2 pre-post studies evaluating tailored-to-risk e-tools [38,72,74]. In 2 other studies with no control and small sample sizes (N=8 or 49), participants either expressed an intent to undergo BCS [60] or showed no change in intention [73].

Moreover, in 2 RCTs [69,70], the effect of 2 features-with-tailoring e-tools on the rate of women with negative intention (willingness not to undergo BCS) was assessed. Our meta-analysis showed that these tools increased the rate compared with usual care (RR 1.88, 95% CI 1.43-2.48, P<.001, I^2=0%, FE model; Figure 2C).

Knowledge

One RCT evaluating a tailored e-tool reported a positive correlation between increased knowledge and participation in BCS [65]. Our meta-analysis, conducted with 5 RCTs assessing a mix of tailored (n=3) and features-with-tailoring (n=2) e-tools, demonstrated that the e-tools increased women's adequate knowledge about BCS (MD 0.75, 95% CI 0.31-1.19, P<.001, I^2=89%, RE model; see Figure S4A in Multimedia Appendix 1) compared with the control group, which consisted of either usual care [54,63,69,70] or a control website [55]. We conducted subanalyses to explore the sources of heterogeneity, with all analyses not altering the overall result (see Figure S4B and S4C in Multimedia Appendix 1). Excluding Roberto et al's [55] study (see the "Methods" section) and subgroup analysis based on the degree of tailoring of the e-tools successfully reduced heterogeneity. Our findings indicated that adequate knowledge was higher with the features-with-tailoring e-tools, which were designed and tested with the general population, compared with the tailored e-tools, which were designed for and tested with

XSL•FO
RenderX

women who were nonparticipants in BCS (χ^2_1=5.1, *P*=.02; Figure 2D) [54,63,69,70]. In a pre-post study [62], a significant increase in adequate knowledge was reported with 1 features-with-tailoring e-tool (*P*=.001). For 2 tailored e-tools, with no reported control group, 48%-64% of users indicated increased knowledge or reported receiving new information, while approximately 15% either disagreed or had some level of disagreement with that statement [64,67,77].

Attitudes

Our meta-analysis showed that features-with-tailoring e-tools, compared with the control group (either usual care [69,70] or a control website [55]), had no significant effect on the rate of women with positive attitudes toward undergoing BCS (ie, being favorable to undergo BCS; RR 0.98, 95% CI 0.95-1.01, *P*=.19, I^2=0%, FE model; see Figure S5 in Multimedia Appendix 1). In another RCT with a 2×2 design that tested a features-with-tailoring e-tool featuring a virtual AI physician, the communication style of the physician and whether the women's needs were made salient were found to influence women's attitudes toward undergoing BCS [58].

Self-Efficacy

The impact of e-tools on women's level of self-efficacy regarding BCS was reported for 2 tailored e-tools. One study, an RCT, found no significant difference in self-efficacy compared with a brochure [54]. By contrast, a pre-post study [72] demonstrated a significant increase in self-efficacy.

Worry

Emotional changes, including anxiety, worry, or fear, were assessed either in general [69,71,78] or specifically related to breast cancer [63,67,75] (Table 2). Our meta-analysis, conducted with 2 RCTs evaluating tailored-to-risk tools, showed that compared with the control group (usual care [63] or a control website [71]), the e-tools increased women's level of worry (MD 0.31, 95% CI 0.13-0.48, *P*<.001, I^2=0%, FE model; Figure 3A), irrespective of the complexity of the tailored messages (see Figure S6 in Multimedia Appendix 1). In a prospective single-arm study, a tailored-to-risk tool that predicted breast cancer risk for women also increased their frequency of worry, which correlated with their level of risk (*P*<.01 by analysis of variance) [75]. In 2 other studies not comparing with controls [67,69], women reported not being particularly worried about breast cancer after using a features-with-tailoring tool [69]. Additionally, 80.9% of users of a tailored-not-to-risk tool indicated that the clinician helped reduce their fears and worries [67].

Figure 3. Forest plots of studies comparing the effect of e-tools with that of control on women's (A) level of worry [63,71], (B) accuracy of perception of individual breast cancer risk [63,71], (C) decisional conflict [55,70], and (D) informed choice [55,69,70]. We reported in (A) and (B) the data corresponding to Seitz et al's [71] "extended information with untailored exemplars" e-tool, which was the most comparable to Schapira et al's [63] e-tool. Data for the other Seitz et al's [71] e-tools are presented in Figures S6 and S7 in Multimedia Appendix 1. The control group was usual care, except for Seitz et al's [71] and Roberto et al's [55] e-tools, which used a website as the control. The risk of bias was assessed using the Risk-of-Bias 2 (RoB 2) tool [52]. The risk level for each domain of RoB 2 (A: randomization process, B: deviations from the intended protocol, C: missing data, D: outcome measurement, and E: reporting results) and the (F) overall risk were evaluated as low (green), moderate (yellow), or high (red). M-H: Mantel-Haenszel.

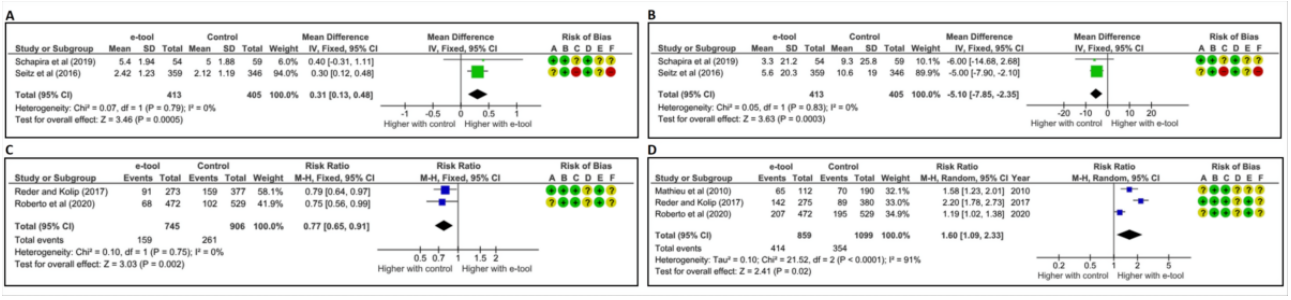

Perceptions of Risk of Breast Cancer

Two tailored-to-risk e-tools were evaluated for their effect on women's accuracy in perceiving their breast cancer risk. Accuracy was determined by measuring the difference between the risk perception assessed by the woman herself and her objective risk estimates provided by the e-tool. Our meta-analysis showed a significant decrease in this difference, indicating that women's accurate estimation of breast cancer risk improved with the use of the e-tools compared with the control group, which consisted of usual care [63] or a control website [71]. A significant improvement in the accuracy of breast cancer risk perception was observed when extended information with untailored examples was used in the e-tool (MD –5.10, 95% CI –7.85 to 2.35, *P*<.001, I^2=0%, FE model; Figure 3B), but not with other types of tailored messages (see Figure S7 in Multimedia Appendix 1). In another study using a tailored-to-risk tool without a control group, 70% of users reported having an "accurate perception" of their breast cancer risk, correctly identifying themselves as at low or average risk [75]. Additionally, the perceived susceptibility of breast cancer was assessed using a tailored-not-to-risk e-tool, which showed no significant difference compared with a control brochure ($F_{1,118}$=0.73, *P*=.01, effect size = 0.01) [54].

Regret

One RCT [70] assessed women's regret about their decision to undergo BCS. The study found no significant difference in regret between the group using the features-with-tailoring e-tool and the usual care group. Additionally, anticipated regret for decisions to delay or initiate mammography was evaluated using

XSL•FO
RenderX

a tailored-to-risk e-tool in another RCT [63], which also showed no significant difference compared with usual care.

Decisional Conflict

Our meta-analysis, which included 2 RCTs comparing 2 features-with-tailoring e-tools with either usual care [70] or a control website [55], found that the e-tools significantly reduced the rate of women experiencing decisional conflict (ie, uncertainty regarding the choice to be made or the decision to undergo BCS). The results showed an RR of 0.77 (95% CI 0.65-0.91, P=.002, I^2=0%, FE model; Figure 3C).

Other results obtained with tailored-to-risk tools were more mixed [63,72,74,75]. One RCT reported no change in the average decisional conflict (RR –0.34, 95% CI –0.71 to 0.03, P=.07) [63]. For 2 other e-tools tested through pre-post study designs [72,74], Yu et al [38] reported a significant decrease in decisional conflict. In a prospective single-arm study, women were found not to have difficulties in implementing decisions, as indicated by their average decisional conflict score [75].

Informed Choice

Our meta-analysis showed that, compared with the control group (either usual care [69,70] or a control website [55]), 3 features-with-tailoring tools increased the rate of women who made an informed choice regarding their intention to undergo BCS by mammography (RR 1.60, 95% CI 1.09-2.33, P=.02, I^2=91%; RE model; Figure 3D). Subanalyses to reduce heterogeneity were conducted but did not yield a significant decrease (data not shown).

Shared Decision-Making and Communications With Health Professionals

The effect of the tools on discussions about BCS during appointments between HPs and women was evaluated with 4 tailored tools [64,67,72,75]. One study reported that an additional 17.2% of women discussed mammography with their HPs compared with a control video group (P<.01) [64]. Another study found that women felt well-prepared to make shared decisions with clinical providers [72]. In 2 other studies comparing e-tool users with nonusers, HPs reported no significant difference in the rates of mammography discussions (90% vs 92%) [75], and the rates of medical “wellness” appointments did not increase (16.3% vs 21.5%) [67].

Grading of the Available Evidence

We summarized and graded the evidence detailed above in Table 3.

With a moderate to high level of certainty, the evidence supports the findings that e-tools (1) increased women’s long-term participation in BCS, intention not to participate, adequate knowledge, worry, and informed choice, while decreasing decisional conflict, and (2) had no effect on short-term participation in BCS, intention, and positive attitudes toward undergoing BCS (Table 3). Additionally, evidence indicates that the tailoring nature of the e-tools is the most influential factor in driving their effectiveness (Table 3).

XSL•FO
RenderX

Table 3. Grade summary of certainty of evidence related to the outcomes used to assess the effects of e-tools.

Outcome and e-tool effect (compared with control)	Main source of evidence (number of e-tools and type of control)	Certainty of evidence (GRADE[a])[b] (reasons for downgrading)	E-tool or population main characteristics
Participation at short term in BCS[c]			
No effect	• Meta-analysis with RCTs[d] (5 tools, control was either usual care [4 tools] or website [1 tool])	⊗⊗⊗◯ Moderate[e] (downgrading due to the presence of 2 high-risk of bias studies in the meta-analysis)	Mix of tailored-not-to-risk and features-with-tailoring e-tools, no differences between the 2 types of e-tools
Participation at long term in BCS[c]			
Increased	• Meta-analysis with RCTs (5 tools, control was either usual care [4 tools] or video [1 tool])	⊗⊗⊗◯ Moderate[e] (downgrading due to 3 high-risk of bias RCTs)	All tailored tools, but not only based on risk. Results suggest a difference between tailored-to-risk and other tailored e-tools
Intention to undergo BCS[c]			
No effect	• Meta-analysis with RCTs (6 tools, control was either usual care [3 tools] or website [3 tools]) • Yu et al's [38] meta-analysis with pre-post study design (2 tools)	⊗⊗⊗⊗ High[f]	Mix of tailored and features-with-tailoring e-tools, no differences between the 2 types of tools, further suggested by Yu et al's results with 2 tailored-to-risk e-tools
Intention not to participate (negative intention) in BCS[c]			
Increased	• Meta-analysis with RCTs (2 tools, control was usual care [2 tools])	⊗⊗⊗◯ Moderate[e] (downgrading due to very low number of RCTs [2 studies])	Features-with-tailoring e-tools
Adequate knowledge			
Increased	• Meta-analysis with RCTs (4 tools, control was usual care [4 tools])	⊗⊗⊗⊗ High[f]	The increase is independent of whether the e-tool is tailored or featured-with-tailoring; however, adequate knowledge is higher with features-with-tailoring e-tools assessed with the general population compared with tailored tools assessed with the nonscreened population
Positive attitude			
No effect	• Meta-analysis with RCTs (3 tools, control was usual care [2 tools] or website [1 tool])	⊗⊗⊗⊗ High[f]	Features-with-tailoring e-tools
Self-efficacy			
No effect	• 1 RCT and 1 pre-post study	⊗◯◯◯ Very low[g] (downgrading due to very low number of studies, contradictory results obtained with 1 RCT and 1 pre-post study)	Tailored e-tools

XSL•FO
RenderX

Outcome and e-tool effect (compared with control)	Main source of evidence (number of e-tools and type of control)	Certainty of evidence (GRADE[a])[b] (reasons for downgrading)	E-tool or population main characteristics
Worry			
Increased	• Meta-analysis with RCTs (2 tools, control was either usual care [1 tool] or website [1 tool]) • Prospective single-arm study showing correlation with risk and worry [75]	⊗⊗⊗◯ Moderate[e] (downgrading due to low number of RCTs [2 studies] including 1 with high bias; upgrading due to correlation result)	Tailored-to-risk e-tools
Accurate risk perception			
Increased	• Meta-analysis with RCTs (2 tools, control was either usual care [1 tool] or website [1 tool])	⊗⊗◯◯ Low[h] (downgrading due to low number of RCTs [2 studies], including 1 with a high risk of bias)	Tailored-to-risk e-tools but with specific types of tool messages. The increase seems to be specific to tailored-to-risk e-tools as no effect on perceived susceptibility was obtained with tailored-not-to-risk e-tool
Regret (decision regret or anticipated regret)			
No effect	• 1 RCT for each type of regret (1 tool, control was usual care)	⊗◯◯◯ Very low[g] (downgrading due to very low number of RCTs [n=1] for each type of regret)	1 features-with-tailoring and 1 tailored e-tool
Decisional conflict			
Decreased	• Meta-analysis with RCTs (2 tools, control was either usual care [1 tool] or website [1 tool])	⊗⊗⊗◯ Moderate[e] (downgrading due to 2 RCTs only)	Features-with-tailoring e-tools
No effect	• 1 RCT and Yu et al's [38] meta-analysis with a pre-post study design (2 tools)	⊗◯◯◯ Very low[g] (downgrading due to the very low number of RCTs [1 study] and contradictory results with Yu et al's [38] meta-analysis with pre-post studies)	Tailored-to-risk e-tools
Informed choice about the intention to undergo BCS[c]			
Increased	• Meta-analysis with RCTs (3 tools, control was either usual care [2 tools] or website [1 tool])	⊗⊗⊗◯ Moderate[e] (downgrading due to inconsistency, ie, high I^2)	Features-with-tailoring e-tools
Discussion/SDM[i]			

XSL•FO
RenderX

Outcome and e-tool effect (compared with control)	Main source of evidence (number of e-tools and type of control)	Certainty of evidence (GRADE[a])[b] (reasons for downgrading)	E-tool or population main characteristics
Increase women's discussion about mammography	• 1 RCT 1 tool	⊗◯◯◯ Very low[g] (Downgrading due to the very low number of RCT[d] [1 study])	1 tailored e-tool
No evidence about SDM	• Not applicable	◯◯◯◯ No formal evidence	Not applicable

[a]GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

[b]We adopted the GRADE Working Group methodology of grading evidence [49-51]. Our baseline statement was high for meta-analyses with RCTs (our results), moderate for RCTs without meta-analysis, and low for non-RCTs. Where indicated, we decreased the certainty in the evidence to at least one level based on the assessment of the following domains: risk of bias, imprecision (number of studies, SD), inconsistency (inconsistent effect between several studies, I^2), indirectness, and publication bias; additionally, downgrading and reasons why it was performed are indicated. As a result of the low total number of studies selected in this review, we did not downgrade based on imprecision (number of studies, SD) criteria except where only 2 studies were available and as detailed in the table. We graded evidence as in footnotes e, f, g, and h.

[c]BCS: breast cancer screening

[d]RCT: randomized controlled trial.

[e]Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

[f]High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

[g]Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

[h]Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

[i]SDM: shared decision-making.

Discussion

Principal Findings

This is the first systematic review to provide a comprehensive overview of the effects of e-tools on women's decision-making about BCS. The findings indicate that e-tools increase women's long-term participation in BCS, intention not to participate, adequate knowledge, worry, and informed choice. They also reduce women's decisional conflict. However, e-tools have no effect on women's short-term participation in BCS, intention, or positive attitudes toward undergoing BCS. Additionally, the review identifies variables that could influence these effects. Among all the variables explored, the degree of tailoring in e-tools (ie, whether the tools were fully "tailored" or "featured with tailoring") appeared to be the most influential (Table 3).

While all e-tools were shown to increase knowledge about BCS, the extent of the increase depended on the degree of tailoring or the study population. The increase was greater with "featured-with-tailoring" e-tools assessed in the general population than with tailored e-tools assessed in non-BCS participants. However, the observation that 1 tailored e-tool was more effective in less-educated women [57,79] supports the notion that the degree of tailoring plays a key role in driving a greater or lesser effect on knowledge. Our results suggest that complex e-tools (ie, tailored tools) are not necessarily more effective in ensuring that women are adequately informed about BCS than less complex e-tools (ie, features-with-tailoring). This finding aligns with recent studies, particularly those involving women from low socioeconomic backgrounds, which indicate that complex information about cancer screening may not be essential for women to be adequately informed [93,94].

Our results indicate that features-with-tailoring e-tools influence women's decisions about undergoing BCS. These tools reduce decisional conflict regarding BCS as well as increase both the intention not to undergo BCS and informed choice about this intention (Table 3). However, they have no effect on positive attitudes toward BCS. The reduction in decisional conflict concerning BCS aligns with findings from a previous meta-analysis of web-based tools [38]. By contrast, a meta-analysis that included both e-tools and printed DAs did not report this effect [23], which may be attributable to the quality of one of the assessed printed DAs [95,96]. An increase in women's informed choice regarding BCS has been consistently reported in all meta-analyses conducted to date on DAs used for BCS, regardless of whether the tools were web-based or printed [23,24,38]. Similar to our findings, these reported effects were associated with features-with-tailoring tools, as defined in our study [23,24,38]. However, a more detailed analysis of our results suggests that this increase in informed choice may stem from a rise in both negative intentions and negative attitudes toward BCS. As defined by Marteau et al [16], individuals are considered to have made an informed choice about BCS if they possess adequate knowledge and hold either positive attitudes and intentions or negative attitudes and intentions. Our results showed that while the e-tools by Reder and Kolip [70], Mathieu et al [69], and Roberto et al [55] increased informed choice (Figure 3D) and knowledge (see Figure S8A in Multimedia Appendix 1), they did not increase either positive attitudes toward BCS (see Figure S5 in Multimedia Appendix 1) or the intention to undergo BCS (see

XSL•FO
RenderX

Figure S8B in Multimedia Appendix 1). It is likely that the observed increase in informed choice among women using these e-tools was driven by an increase in negative intentions, as evidenced by 2 of the 3 tools (Figure 2C), and possibly by negative attitudes toward participation in BCS, as reported in other studies [24]. Interestingly, these 3 e-tools are among the highest-rated BCS DAs based on the IPDAS criteria. This was formally evaluated for the e-tools by Reder and Kolip [70] and Mathieu et al [69] in a recent systematic review [85]. Although Roberto et al's [55] e-tool was not formally assessed in the same review, it is expected to be of similarly high quality due to its strong resemblance to Reder and Kolip's [70] tool [85]. Our results suggest that even high-quality e-tools, as assessed by IPDAS criteria, and their ability to achieve higher levels of adequate knowledge, increase informed choice, and reduce decisional conflict, do not guarantee the promotion of positive intentions and attitudes toward undergoing BCS. On the contrary, these tools may increase negative intentions and attitudes. BCS programs should take this finding into consideration.

Our results on participation in BCS (Table 3) strongly suggest that only tailored e-tools have a significant effect on participation. Two factors may explain this finding regarding tailored-to-risk e-tools. First, these e-tools, when appropriately designed, increase women's accurate perception of their own breast cancer risk (as shown by our results). Second, they may also heighten women's levels of worry (Table 3), which is likely a key driver of increased participation in BCS. Indeed, while fear of breast cancer has been reported as a barrier, it has also been identified as a facilitator of BCS participation [9,10,97]. BCS programs should address women's feelings of worry if they aim to implement tailored-to-risk e-tools. However, further studies are needed to assess women's acceptability of feeling worried in comparison to their perceived benefits of using these e-tools. In addition, it is important to note that the tailored e-tools evaluated in this review incorporate multiple features that allow for their effective integration into clinical pathways or health care services (Table 1). These features can be seen as opportunities to reduce barriers and increase access to BCS participation, such as by facilitating helpful discussions with health care providers (HPs) or simplifying the appointment process [9,10,97]. Additionally, these features could help mitigate the impact of the tools on increasing worry [67].

Among the 22 e-tools investigated in this research, including the 6 identified as preparing for or facilitating SDM [63,64,72-75], none were evaluated using any validated SDM instruments [98-106]. There was no evidence to indicate whether the e-tools had an effect on SDM. Although this review suggests that e-tools, particularly tailored e-tools, may improve SDM by enhancing the quality of appointments with health care providers (HPs), this is likely due to their multiple features, which enable effective integration into clinical pathways or health care services (Table 1).

Recommendations for Future Developments

While all types of e-tools appeared effective in increasing users' knowledge about BCS, our results highlight important dilemmas for BCS programs that are using or planning to use e-tools to support women at average risk of breast cancer [2-4] in making decisions about BCS. Although features-with-tailoring e-tools can potentially increase informed choice and reduce decisional conflict, they may also be perceived as negatively impacting BCS programs by fostering negative intentions and attitudes toward undergoing BCS. Conversely, tailored e-tools, despite lacking evidence to support their effects on informed choice and decisional conflict, would increase women's participation in BCS but, when tailored to risk, these tools may also heighten their levels of worry. One potential approach to minimize risks to both women's well-being (worry) and BCS programs would be an "on-demand" model. In this model, women would "tailor" their own information needs, deciding on the nature and amount of information they feel is necessary. In this approach, we recommend that (1) all women be provided with a minimum of important, carefully assessed information about BCS in both text and audio formats, with multiple language options; and (2) women who do not wish to participate in BCS be given the option, through the e-tool, to obtain estimates of their breast cancer risk or, at the very least, to have a discussion with health care providers (HPs) about breast cancer risk in general. Some women may be unaware of the average breast cancer risk for women or of their own increased risk [2-4]. We also suggest that e-tools be well integrated into clinical pathways or health care services in various ways to reduce access-related barriers, especially those specific to low socioeconomic groups, such as health literacy. Notably, embedding e-tools in a health portal (Table 1) would facilitate appointments and communication with health care providers (HPs). AI tools, which have not been fully explored in the context of BCS according to this review, could support the "on-demand" approach. However, both the efficacy of such AI-based tools and potential ethical issues or biases need to be carefully assessed [107-110].

Limitations

This study has several limitations that should be considered in future research. Caution is needed when generalizing our results due to (1) the limited number of studies included, particularly in the meta-analyses, with some outcomes assessed by only 2 RCTs (eg, "intention not to undergo BCS," "level of worry," "accuracy of perception of individual breast cancer risk," and "decisional conflict"), some of which were at high risk of bias; (2) the high heterogeneity between the studies and e-tools; (3) the fact that almost all studies were conducted in high-income countries, mostly in the United States, which may limit the applicability of the results to other settings or low- and middle-income countries; and (4) the lack of assessment of implementation factors (eg, comparing users vs nonusers or completers vs noncompleters) and their impact on decision-making outcomes. Additionally, (1) we did not thoroughly review the content of the BCS information provided in the e-tools evaluated in this review, so we cannot draw conclusions about the tangible value of the increase in knowledge [95]; and (2) we did not have enough information to determine whether some of the e-tools could be classified as DAs. However, we applied a robust methodology to minimize some of these limitations, notably by comprehensively exploring the variables that could explain the heterogeneity of the results,

XSL•FO
RenderX

using RoB 2 and GRADE assessments. A review is currently underway to explore implementation outcomes.

Conclusions

Although no evidence was available to assess the efficacy of the tools in supporting SDM, this review demonstrates that the e-tools designed to assist women's decision-making regarding BCS do impact this process. The degree of tailoring of the e-tools (ie, whether they are tailored, particularly to risk, or include features with tailoring) appears to be the most significant factor influencing decision-making. This review provides valuable insights for BCS programs when implementing such e-tools and offers directions for future development.

Acknowledgments

The authors would thank Dr. Marie-Anne Durand (Unisanté, Lausanne) for her valuable comments and participation. We are also grateful to all the study authors with whom we communicated to gather additional data for our meta-analyses. We thank our IARC colleague Mrs. Karima Bendeddouche for her support with the submission process. Our thanks also go to Teresa Lee (IARC Knowledge Manager) and Latifa Bouanzi (IARC Information Assistant) for their help in developing the research question and building it across all databases. This review is dedicated to the memory of our kind and highly professional colleague from WHO/Geneva, Tomas John Allen (Librarian, WHO), who helped identify the search strategy for Embase.

Data Availability

Multimedia Appendix 1, which cites additional references [111-115], contains extra information about the methods used and reports additional meta-analyses and subanalyses. The complete qualitative data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

PV conceived the review, conducted the analyses, including meta-analyses, and wrote the publication. RS supervised the study and the writing. PV and ALB performed the initial data search and PV updated the search. PV and LD, ALB, or CB extracted the data. Risk-of-bias of the RCTs was assessed by PV and LD, ALB, or CB. Grade assessment was performed by PV and LD. All authors contributed to establishing the data extraction form. All authors critically reviewed the preliminary results and the final report. The main contributors at the 2019 workshop on e-DAs and in further discussions are all listed as co-authors or in the Acknowledgments (Marie-Anne Durand). PV/IARC obtained funding from the Institut National du Cancer, France (INCa, grant DEP18-066). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and additional methods and analysis.
[PDF File (Adobe PDF File), 3280 KB - jmir_v27i1e65974_app1.pdf]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology: breast cancer (v. 6.2024). NCCN. 2024 Oct 23. URL: https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1419 [accessed 2024-12-18]
3. Oeffinger KC, Fontham ETH, Etzioni R, Herzig A, Michaelson JS, Shih YT, American Cancer Society. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 2015 Oct 20;314(15):1599-1614 [FREE Full text] [doi: 10.1001/jama.2015.12783] [Medline: 26501536]
4. US Preventive Services Task Force, Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, et al. Screening for breast cancer: US Preventive Services Task Force recommendation statement. JAMA 2024 Jun 11;331(22):1918-1930. [doi: 10.1001/jama.2024.5534] [Medline: 38687503]
5. Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, American Cancer Society Breast Cancer Advisory Group. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin 2007;57(2):75-89 [FREE Full text] [doi: 10.3322/canjclin.57.2.75] [Medline: 17392385]
6. The Council of the European Union. Council recommendation of 9 December 2022 on strengthening prevention through early detection: a new EU approach on cancer screening replacing Council recommendation 2003/878/EC 2022/C 473/01.

XSL•FO
RenderX

Publication Office of the European Union. 2022 Dec 09. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022H1213(01) [accessed 2024-12-18]
7. Zielonke N, Kregting LM, Heijnsdijk EAM, Veerus P, Heinävaara S, McKee M, EU-TOPIA collaborators. The potential of breast cancer screening in Europe. Int J Cancer 2021 Jan 15;148(2):406-418 [FREE Full text] [doi: 10.1002/ijc.33204] [Medline: 32683673]
8. IARC Working Group on the Evaluation of Cancer-Preventive Interventions. Breast Cancer Screening: IARC Handbooks of Cancer Prevention (Vol. 15). Lyon, France: International Agency for Research on Cancer; 2016.
9. Mandrik O, Tolma E, Zielonke N, Meheus F, Ordóñez-Reyes C, Severens JL, et al. Systematic reviews as a "lens of evidence": determinants of participation in breast cancer screening. J Med Screen 2021 Jun;28(2):70-79 [FREE Full text] [doi: 10.1177/0969141320930743] [Medline: 32517538]
10. Mascara M, Constantinou C. Global perceptions of women on breast cancer and barriers to screening. Curr Oncol Rep 2021 May 03;23(7):74. [doi: 10.1007/s11912-021-01069-z] [Medline: 33937940]
11. Özkan İ, Taylan S. Barriers to women's breast cancer screening behaviors in several countries: a meta-synthesis study. Health Care Women Int 2021 Sep;42(7-9):1013-1043. [doi: 10.1080/07399332.2020.1814777] [Medline: 32877315]
12. Autier P. Screening for breast cancer: worries about its effectiveness. Rev Prat 2013 Dec;63(10):1369-1377. [Medline: 24579331]
13. Pivot X, Viguier J, Touboul C, Morère JF, Blay J, Coscas Y, et al. Breast cancer screening controversy: too much or not enough? Eur J Cancer Prev 2015 Jun;24 Suppl:S73-S76. [doi: 10.1097/CEJ.0000000000000145] [Medline: 26016791]
14. Bour C. Is the socio-cultural environment favorable to women making an informed decision regarding breast cancer screening in France? BMJ Evidence-Based Medicine 2022 Jun;27(Suppl 1):A2 [FREE Full text] [doi: 10.1136/bmjebm-2022-PODabstracts.3]
15. Mandrik O, Zielonke N, Meheus F, Severens JLH, Guha N, Herrero Acosta R, et al. Systematic reviews as a 'lens of evidence': determinants of benefits and harms of breast cancer screening. Int J Cancer 2019 Aug 15;145(4):994-1006 [FREE Full text] [doi: 10.1002/ijc.32211] [Medline: 30762235]
16. Marteau TM, Dormandy E, Michie S. A measure of informed choice. Health Expect 2001 Jun;4(2):99-108 [FREE Full text] [doi: 10.1046/j.1369-6513.2001.00140.x] [Medline: 11359540]
17. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International Patient Decision Aids Standards (IPDAS) Collaboration. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
18. Joseph-Williams N, Newcombe R, Politi M, Durand M, Sivell S, Stacey D, et al. Toward minimum standards for certifying patient decision aids: a modified Delphi consensus process. Med Decis Making 2014 Aug;34(6):699-710. [doi: 10.1177/0272989X13501721] [Medline: 23963501]
19. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4(4):CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
20. Burnside ES, Schrager S, DuBenske L, Keevil J, Little T, Trentham-Dietz A, et al. Team science principles enhance cancer care delivery quality improvement: interdisciplinary implementation of breast cancer screening shared decision making. JCO Oncol Pract 2023 Jan;19(1):e1-e7 [FREE Full text] [doi: 10.1200/OP.22.00355] [Medline: 36126243]
21. Légaré F, Stacey D, Pouliot S, Gauvin F, Desroches S, Kryworuchko J, et al. Interprofessionalism and shared decision-making in primary care: a stepwise approach towards a new model. J Interprof Care 2011 Jan;25(1):18-25 [FREE Full text] [doi: 10.3109/13561820.2010.490502] [Medline: 20795835]
22. McAlpine K, Lewis KB, Trevena LJ, Stacey D. What Is the Effectiveness of Patient Decision Aids for Cancer-Related Decisions? A Systematic Review Subanalysis. JCO Clin Cancer Inform 2018 Dec;2:1-13 [FREE Full text] [doi: 10.1200/CCI.17.00148] [Medline: 30652610]
23. Martínez-Alonso M, Carles-Lavila M, Pérez-Lacasta MJ, Pons-Rodríguez A, Garcia M, Rué M, InforMa Group. Assessment of the effects of decision aids about breast cancer screening: a systematic review and meta-analysis. BMJ Open 2017 Oct 06;7(10):e016894 [FREE Full text] [doi: 10.1136/bmjopen-2017-016894] [Medline: 28988175]
24. Ivlev I, Hickman EN, McDonagh MS, Eden KB. Use of patient decision aids increased younger women's reluctance to begin screening mammography: a systematic review and meta-analysis. J Gen Intern Med 2017 Jul;32(7):803-812 [FREE Full text] [doi: 10.1007/s11606-017-4027-9] [Medline: 28289963]
25. Légaré F, Adekpedjou R, Stacey D, Turcotte S, Kryworuchko J, Graham I, et al. Interventions for increasing the use of shared decision making by healthcare professionals. Cochrane Database Syst Rev 2018 Jul 19;7(7):CD006732. [doi: 10.1002/14651858.CD006732.pub4] [Medline: 30025154]
26. Stacey D, Volk RJ. The International Patient Decision Aid Standards (IPDAS) Collaboration: evidence update 2.0. Med Decis Making 2021 Aug 20;41(7):729-733. [doi: 10.1177/0272989x211035681]
27. Hashem F, Calnan MW, Brown PR. Decision making in NICE single technological appraisals: how does NICE incorporate patient perspectives? Health Expect 2018 Feb;21(1):128-137 [FREE Full text] [doi: 10.1111/hex.12594] [Medline: 28686809]

XSL•FO
RenderX

28. Holmes-Rovner M, Srikanth A, Henry SG, Langford A, Rovner DR, Fagerlin A. Decision aid use during post-biopsy consultations for localized prostate cancer. Health Expect 2018 Feb;21(1):279-287 [FREE Full text] [doi: 10.1111/hex.12613] [Medline: 28881105]
29. Leppin AL, Kunneman M, Hathaway J, Fernandez C, Montori VM, Tilburt JC. Getting on the same page: communication, patient involvement and shared understanding of "decisions" in oncology. Health Expect 2018 Feb;21(1):110-117 [FREE Full text] [doi: 10.1111/hex.12592] [Medline: 28636280]
30. Menear M, Garvelink MM, Adekpedjou R, Perez MMB, Robitaille H, Turcotte S, et al. Factors associated with shared decision making among primary care physicians: findings from a multicentre cross-sectional study. Health Expect 2018 Feb;21(1):212-221 [FREE Full text] [doi: 10.1111/hex.12603] [Medline: 28768060]
31. DuBenske LL, Schrager SB, Hitchcock ME, Kane AK, Little TA, McDowell HE, et al. Key elements of mammography shared decision-making: a scoping review of the literature. J Gen Intern Med 2018 Oct;33(10):1805-1814 [FREE Full text] [doi: 10.1007/s11606-018-4576-6] [Medline: 30030738]
32. Schrager SB, Phillips G, Burnside E. A simple approach to shared decision making in cancer screening. Fam Pract Manag 2017;24(3):5-10 [FREE Full text] [Medline: 28671358]
33. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns 2006 Mar;60(3):301-312. [doi: 10.1016/j.pec.2005.06.010] [Medline: 16051459]
34. Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, Pieterse AH. Key components of shared decision making models: a systematic review. BMJ Open 2019 Dec 17;9(12):e031763 [FREE Full text] [doi: 10.1136/bmjopen-2019-031763] [Medline: 31852700]
35. Ruco A, Dossa F, Tinmouth J, Llovet D, Jacobson J, Kishibe T, et al. Social media and mHealth technology for cancer screening: systematic review and meta-analysis. J Med Internet Res 2021 Jul 30;23(7):e26759 [FREE Full text] [doi: 10.2196/26759] [Medline: 34328423]
36. World Health Organization (WHO). Global strategy on digital health 2020-2025. WHO. Geneva, Switzerland: World Health Organization (WHO); 2021. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2024-12-18]
37. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022 Sep 09;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
38. Yu L, Li P, Yang S, Guo P, Zhang X, Liu N, et al. Web-based decision aids to support breast cancer screening decisions: systematic review and meta-analysis. J Comp Eff Res 2020 Oct;9(14):985-1002. [doi: 10.2217/cer-2020-0052] [Medline: 33025800]
39. Stout P, Villegas J, Kim H. Enhancing learning through use of interactive tools on health-related websites. Health Educ Res 2001 Dec;16(6):721-733. [doi: 10.1093/her/16.6.721] [Medline: 11780710]
40. World Health Organization (WHO). Classification of digital health interventions v1.0. WHO. 2018. URL: https://www.who.int/publications/i/item/WHO-RHR-18.06 [accessed 2024-12-18]
41. Le Bonniec A, Bauquier C, Mandrik O, Villain P. Effectiveness of online or interactive tools used as decision aids or shared decision making for women to decide about breast cancer screening: a systematic review. The International Prospective Register of Systematic Reviews (PROSPERO). 2020. URL: https://www.crd.york.ac.uk/prospero/ [accessed 2024-12-18]
42. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
43. Straw I, Brass I, Mkwashi A, Charles I, Soares A, Steer C. Insights from a clinically orientated workshop on health care cybersecurity and medical technology: observational study and thematic analysis. J Med Internet Res 2024 Jul 11;26:e50505 [FREE Full text] [doi: 10.2196/50505] [Medline: 38990611]
44. Veritas Health Innovation. Covidence systematic review software. Covidence. Melbourne, Victoria, Australia: Veritas Health Innovation URL: https://www.covidence.org/ [accessed 2024-12-18]
45. Briñol P, Petty R. Fundamental processes leading to attitude change: implications for cancer prevention communications. Journal of Communication 2006;56(suppl_1):S81-S104. [doi: 10.1111/j.1460-2466.2006.00284.x]
46. Bull FC, Kreuter MW, Scharff DP. Effects of tailored, personalized and general health messages on physical activity. Patient Educ Couns 1999 Feb;36(2):181-192. [doi: 10.1016/s0738-3991(98)00134-7] [Medline: 10223022]
47. Deeks JJ, Higgins JPT, Altman DG, McKenzie JE, Veroniki AA. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.5. London, UK: Cochrane; 2024.
48. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
49. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013 Jul;66(7):726-735. [doi: 10.1016/j.jclinepi.2013.02.003] [Medline: 23570745]
50. Siemieniuk R, Guyatt G. What is GRADE? BMJ. 2019. URL: https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/ [accessed 2024-12-18]

XSL•FO
RenderX

51. Zeng L, Brignardello-Petersen R, Hultcrantz M, Mustafa RA, Murad MH, Iorio A, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. J Clin Epidemiol 2022 Oct;150:216-224 [FREE Full text] [doi: 10.1016/j.jclinepi.2022.07.014] [Medline: 35934265]
52. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
53. The Cochrane Collaboration. Review Manager (RevMan), version 5.4. Cochrane. URL: https://revman.cochrane.org [accessed 2024-12-18]
54. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: pilot randomized controlled trial. JMIR Mhealth Uhealth 2017 Nov 07;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]
55. Roberto A, Colombo C, Candiani G, Satolli R, Giordano L, Jaramillo L, et al. A dynamic web-based decision aid to improve informed choice in organised breast cancer screening. A pragmatic randomised trial in Italy. Br J Cancer 2020 Sep 17;123(5):714-721 [FREE Full text] [doi: 10.1038/s41416-020-0935-2] [Medline: 32546834]
56. Champion VL, Christy SM, Rakowski W, Lairson DR, Monahan PO, Gathirua-Mwangi WG, et al. An RCT to increase breast and colorectal cancer screening. Am J Prev Med 2020 Aug;59(2):e69-e78 [FREE Full text] [doi: 10.1016/j.amepre.2020.03.008] [Medline: 32690203]
57. Champion VL, Rawl SM, Bourff SA, Champion KM, Smith LG, Buchanan AH, et al. Randomized trial of DVD, telephone, and usual care for increasing mammography adherence. J Health Psychol 2016 Jun;21(6):916-926 [FREE Full text] [doi: 10.1177/1359105314542817] [Medline: 25070967]
58. Fissler T, Bientzle M, Cress U, Kimmerle J. The impact of advice seekers' need salience and doctors' communication style on attitude and decision making: a web-based mammography consultation role play. JMIR Cancer 2015 Sep 08;1(2):e10 [FREE Full text] [doi: 10.2196/cancer.4279] [Medline: 28410160]
59. Henry SL, Shen E, Ahuja A, Gould MK, Kanter MH. The online personal action plan: a tool to transform patient-enabled preventive and chronic care. Am J Prev Med 2016 Jul;51(1):71-77. [doi: 10.1016/j.amepre.2015.11.014] [Medline: 26826751]
60. Klippert H, Schaper A. Using Facebook to communicate mammography messages to rural audiences. Public Health Nurs 2019 Mar 29;36(2):164-171. [doi: 10.1111/phn.12556] [Medline: 30370687]
61. Krist AH, Woolf SH, Rothemich SF, Johnson RE, Peele JE, Cunningham TD, et al. Interactive preventive health record to enhance delivery of recommended care: a randomized trial. Ann Fam Med 2012 Jul 09;10(4):312-319 [FREE Full text] [doi: 10.1370/afm.1383] [Medline: 22778119]
62. Pereira AAC, Destro JR, Picinin Bernuci M, Garcia LF, Rodrigues Lucena TF. Effects of a WhatsApp-delivered education intervention to enhance breast cancer knowledge in women: mixed-methods study. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17430 [FREE Full text] [doi: 10.2196/17430] [Medline: 32706726]
63. Schapira MM, Hubbard RA, Seitz HH, Conant EF, Schnall M, Cappella JN, et al. The impact of a risk-based breast cancer screening decision aid on initiation of mammography among younger women: report of a randomized trial. MDM Policy Pract 2019;4(1):2381468318812889 [FREE Full text] [doi: 10.1177/2381468318812889] [Medline: 30729166]
64. Walsh J, Potter M, Salazar R, Ozer E, Gildengorin G, Dass N, et al. PreView: a randomized trial of a multi-site intervention in diverse primary care to increase rates of age-appropriate cancer screening. J Gen Intern Med 2020 Feb;35(2):449-456. [doi: 10.1007/s11606-019-05438-0] [Medline: 31728894]
65. Bowen DJ, Robbins R, Bush N, Meischke H, Ludwig A, Wooldridge J. Effects of a Web-based intervention on women's breast health behaviors. Transl Behav Med 2011 Mar;1(1):155-164 [FREE Full text] [doi: 10.1007/s13142-011-0028-0] [Medline: 21643515]
66. Champion VL, Monahan PO, Stump TE, Biederman EB, Vachon E, Katz ML, et al. The effect of two interventions to increase breast cancer screening in rural women. Cancers (Basel) 2022 Sep 07;14(18):4354 [FREE Full text] [doi: 10.3390/cancers14184354] [Medline: 36139515]
67. Krist AH, Woolf SH, Hochheimer C, Sabo RT, Kashiri P, Jones RM, et al. Harnessing information technology to inform patients facing routine decisions: cancer screening as a test case. Ann Fam Med 2017 May 08;15(3):217-224 [FREE Full text] [doi: 10.1370/afm.2063] [Medline: 28483886]
68. Lin Z, Wang S. A tailored web-based intervention to promote women's perceptions of and intentions for mammography. J Nurs Res 2009 Dec;17(4):249-260. [doi: 10.1097/JNR.0b013e3181c15a38] [Medline: 19955881]
69. Mathieu E, Barratt AL, McGeechan K, Davey HM, Howard K, Houssami N. Helping women make choices about mammography screening: an online randomized trial of a decision aid for 40-year-old women. Patient Educ Couns 2010 Oct;81(1):63-72. [doi: 10.1016/j.pec.2010.01.001] [Medline: 20149953]
70. Reder M, Kolip P. Does a decision aid improve informed choice in mammography screening? Results from a randomised controlled trial. PLoS One 2017;12(12):e0189148 [FREE Full text] [doi: 10.1371/journal.pone.0189148] [Medline: 29236722]
71. Seitz HH, Gibson L, Skubisz C, Forquer H, Mello S, Schapira MM, et al. Effects of a risk-based online mammography intervention on accuracy of perceived risk and mammography intentions. Patient Educ Couns 2016 Oct;99(10):1647-1656 [FREE Full text] [doi: 10.1016/j.pec.2016.05.005] [Medline: 27178707]

XSL•FO
RenderX

72. Eden KB, Scariati P, Klein K, Watson L, Remiker M, Hribar M, et al. Mammography decision aid reduces decisional conflict for women in their forties considering screening. J Womens Health (Larchmt) 2015 Dec;24(12):1013-1020 [FREE Full text] [doi: 10.1089/jwh.2015.5256] [Medline: 26360918]
73. Eden KB, Ivlev I, Bensching KL, Franta G, Hersh AR, Case J, et al. Use of an online breast cancer risk assessment and patient decision aid in primary care practices. J Womens Health (Larchmt) 2020 Jun 01;29(6):763-769 [FREE Full text] [doi: 10.1089/jwh.2019.8143] [Medline: 32159424]
74. Scariati P, Nelson L, Watson L, Bedrick S, Eden KB. Impact of a decision aid on reducing uncertainty: pilot study of women in their 40s and screening mammography. BMC Med Inform Decis Mak 2015 Nov 10;15:89 [FREE Full text] [doi: 10.1186/s12911-015-0210-2] [Medline: 26554555]
75. Elkin EB, Pocus VH, Mushlin AI, Cigler T, Atoria CL, Polaneczky MM. Facilitating informed decisions about breast cancer screening: development and evaluation of a web-based decision aid for women in their 40s. BMC Med Inform Decis Mak 2017 Mar 21;17(1):29 [FREE Full text] [doi: 10.1186/s12911-017-0423-7] [Medline: 28327125]
76. Lin Z, Effken JA. Effects of a tailored web-based educational intervention on women's perceptions of and intentions to obtain mammography. J Clin Nurs 2010 May 08;19(9-10):1261-1269. [doi: 10.1111/j.1365-2702.2009.03180.x] [Medline: 20345827]
77. Arora M, Gerbert B, Potter M, Gildengorin G, Walsh J. PRE-VIEW: development and pilot testing of an interactive video doctor plus provider alert to increase cancer screening. ISRN Prev Med 2013;2013:935487 [FREE Full text] [doi: 10.5402/2013/935487] [Medline: 24409373]
78. Seitz HH, Schapira MM, Gibson LA, Skubisz C, Mello S, Armstrong K, et al. Explaining the effects of a decision intervention on mammography intentions: the roles of worry, fear and perceived susceptibility to breast cancer. Psychol Health 2018 May;33(5):682-700 [FREE Full text] [doi: 10.1080/08870446.2017.1387261] [Medline: 29073799]
79. Gathirua-Mwangi WG, Monahan PO, Stump T, Rawl SM, Skinner CS, Champion VL. Mammography adherence in African-American women: results of a randomized controlled trial. Ann Behav Med 2016 Feb;50(1):70-78 [FREE Full text] [doi: 10.1007/s12160-015-9733-0] [Medline: 26416127]
80. Lin ZC, Effken JA, Li YJ, Kuo CH. Designing a tailored web-based educational mammography program. Comput Inform Nurs 2011;29(1):16-23. [doi: 10.1097/NCN.0b013e3181f9db1c] [Medline: 21099540]
81. Lee HY, Lee MH, Gao Z, Sadak K. Development and evaluation of culturally and linguistically tailored mobile app to promote breast cancer screening. J Clin Med 2018 Jul 24;7(8):181 [FREE Full text] [doi: 10.3390/jcm7080181] [Medline: 30042291]
82. Klein KA, Watson L, Ash JS, Eden KB. Evaluation of risk communication in a mammography patient decision aid. Patient Educ Couns 2016 Jul;99(7):1240-1248 [FREE Full text] [doi: 10.1016/j.pec.2016.02.013] [Medline: 26965020]
83. Skinner CS, Buchanan A, Champion V, Monahan P, Rawl S, Springston J, et al. Process outcomes from a randomized controlled trial comparing tailored mammography interventions delivered via telephone vs. DVD. Patient Educ Couns 2011 Nov;85(2):308-312 [FREE Full text] [doi: 10.1016/j.pec.2010.10.024] [Medline: 21112173]
84. Reder M, Soellner R, Kolip P. Do women with high eHealth literacy profit more from a decision aid on mammography screening? Testing the moderation effect of the eHEALS in a randomized controlled trial. Front Public Health 2019;7:46 [FREE Full text] [doi: 10.3389/fpubh.2019.00046] [Medline: 30931291]
85. Hild S, Johanet M, Valenza A, Thabaud M, Laforest F, Ferrat E, DEDICACES Group, the French National College of General Practitioners. Quality of decision aids developed for women at average risk of breast cancer eligible for mammographic screening: Systematic review and assessment according to the International Patient Decision Aid Standards instrument. Cancer 2020 Jun 15;126(12):2765-2774 [FREE Full text] [doi: 10.1002/cncr.32858] [Medline: 32267546]
86. Elwyn G, O'Connor AM, Bennett C, Newcombe RG, Politi M, Durand M, et al. Assessing the quality of decision support technologies using the International Patient Decision Aid Standards instrument (IPDASi). PLoS One 2009;4(3):e4705 [FREE Full text] [doi: 10.1371/journal.pone.0004705] [Medline: 19259269]
87. Keary T. Web-based application. Technopedia. URL: https://www.techopedia.com/definition/26002/web-based-application [accessed 2024-12-18]
88. Witteman HO, Ndjaboue R, Vaisson G, Dansokho SC, Arnold B, Bridges JFP, et al. Clarifying values: an updated and expanded systematic review and meta-analysis. Med Decis Making 2021 Oct;41(7):801-820 [FREE Full text] [doi: 10.1177/0272989X211037946] [Medline: 34565196]
89. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
90. Dormandy E, Michie S, Hooper R, Marteau TM. Informed choice in antenatal Down syndrome screening: a cluster-randomised trial of combined versus separate visit testing. Patient Educ Couns 2006 Apr;61(1):56-64. [doi: 10.1016/j.pec.2005.02.006] [Medline: 16533677]
91. Rosenstock IM. The health belief model: explaining health behavior through expectancies. In: Glanz K, Lewis FM, Rimer BK, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass/Wiley; 1990:39-62.
92. The Ottawa Hospital Patient Decison Aids Reseach Group. Decisional Conflict Scale. The Ottawa Hospital Patient Decison Aids Reseach Group. URL: https://decisionaid.ohri.ca/eval_dcs.html [accessed 2024-12-18]

XSL•FO
RenderX

93. Tyldesley-Marshall N, Grove A, Ghosh I, Kudrna L, Ayorinde A, Singh M, et al. Investigating informed choice in screening programmes: a mixed methods analysis. BMC Public Health 2022 Dec 12;22(1):2319 [FREE Full text] [doi: 10.1186/s12889-022-14685-6] [Medline: 36510247]
94. Guigon L, Sánchez LXG, Petit A, Bonniec AL, Basu P, Rodrigue CM, et al. Would shared decision-making be useful in breast cancer screening programmes? A qualitative study using focus group discussions to gather evidence from French women with different socioeconomic backgrounds. BMC Public Health 2024 Feb 07;24(1):404 [FREE Full text] [doi: 10.1186/s12889-024-17876-5] [Medline: 38326802]
95. Sasieni PD, Smith RA, Duffy SW. Informed decision-making and breast cancer screening. J Med Screen 2015 Dec;22(4):165-167 [FREE Full text] [doi: 10.1177/0969141315587344] [Medline: 26018778]
96. Hersch J, Barratt A, Jansen J, Irwig L, McGeechan K, Jacklyn G, et al. Use of a decision aid including information on overdetection to support informed choice about breast cancer screening: a randomised controlled trial. Lancet 2015 Apr 25;385(9978):1642-1652. [doi: 10.1016/S0140-6736(15)60123-4] [Medline: 25701273]
97. Lemmo D, Martino ML, Vallone F, Donizzetti AR, Freda MF, Palumbo F, et al. Clinical and psychosocial constructs for breast, cervical, and colorectal cancer screening participation: a systematic review. Int J Clin Health Psychol 2023;23(2):100354 [FREE Full text] [doi: 10.1016/j.ijchp.2022.100354] [Medline: 36415605]
98. Nicolai J, Moshagen M, Eich W, Bieber C. The OPTION scale for the assessment of shared decision making (SDM): methodological issues. Z Evid Fortbild Qual Gesundhwes 2012;106(4):264-271. [doi: 10.1016/j.zefq.2012.03.002] [Medline: 22749073]
99. Ubbink DT, van Asbeck EV, Aarts JW, Stubenrouch FE, Geerts PA, Atsma F, et al. Comparison of the CollaboRATE and SDM-Q-9 questionnaires to appreciate the patient-reported level of shared decision-making. Patient Educ Couns 2022 Jul;105(7):2475-2479 [FREE Full text] [doi: 10.1016/j.pec.2022.03.007] [Medline: 35331573]
100. Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. Patient Educ Couns 2010 Jul;80(1):94-99. [doi: 10.1016/j.pec.2009.09.034] [Medline: 19879711]
101. Melbourne E, Sinclair K, Durand MA, Légaré F, Elwyn G. Developing a dyadic OPTION scale to measure perceptions of shared decision making. Patient Educ Couns 2010 Feb;78(2):177-183. [doi: 10.1016/j.pec.2009.07.009] [Medline: 19647970]
102. Martin LR, DiMatteo MR, Lepper HS. Facilitation of patient involvement in care: development and validation of a scale. Behav Med 2001;27(3):111-120. [doi: 10.1080/08964280109595777] [Medline: 11985184]
103. Lerman CE, Brody DS, Caputo GC, Smith DG, Lazaro CG, Wolfson HG. Patients' perceived involvement in care scale: relationship to attitudes about illness and medical care. J Gen Intern Med 1990 Jan;5(1):29-33. [doi: 10.1007/bf02602306]
104. Degner L, Kristjanson L, Bowman D, Sloan J, Carriere K, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. JAMA 1997 May 14;277(18):1485-1492. [Medline: 9145723]
105. Barr PJ, Thompson R, Walsh T, Grande SW, Ozanne EM, Elwyn G. The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. J Med Internet Res 2014 Jan 03;16(1):e2 [FREE Full text] [doi: 10.2196/jmir.3085] [Medline: 24389354]
106. The National Health Service (NHS). Measuring shared decision making, a review of research evidence. NHS. 2012. URL: https://www.england.nhs.uk/wp-content/uploads/2013/08/7sdm-report.pdf [accessed 2024-12-18]
107. Parviainen J, Rantala J. Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care Philos 2022 Mar;25(1):61-71 [FREE Full text] [doi: 10.1007/s11019-021-10049-w] [Medline: 34480711]
108. Pallardy C. The chatbot will see you now: 4 ethical concerns of AI in health care. Information Week. 2023. URL: https://www.informationweek.com/machine-learning-ai/the-chatbot-will-see-you-now-4-ethical-concerns-of-ai-in-health-care [accessed 2024-12-18]
109. Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: ethical issues with using chatbots in mental health. Digit Health 2023;9:20552076231183542 [FREE Full text] [doi: 10.1177/20552076231183542] [Medline: 37377565]
110. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850 [FREE Full text] [doi: 10.2196/27850] [Medline: 34847056]
111. Reeves BC, Deeks JJ, Higgins JPT, Shea B, Tugwell P, Wells GA. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.2. London, UK: Cochrane; 2024.
112. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Metzendorf MI, et al. Chapter 4: Searching for and selecting studies [last updated September 2024]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.5. London, UK: Cochrane; 2024.
113. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]

XSL•FO
RenderX

114. Champion VL, Christy SM, Rakowski W, Gathirua-Mwangi WG, Tarver WL, Carter-Harris L, et al. A randomized trial to compare a tailored web-based intervention and tailored phone counseling to usual care for increasing colorectal cancer screening. Cancer Epidemiol Biomarkers Prev 2018 Dec;27(12):1433-1441 [FREE Full text] [doi: 10.1158/1055-9965.EPI-18-0180] [Medline: 30181203]
115. Biederman E, Baltic R, Katz ML, Rawl S, Vachon E, Monahan PO, et al. Increasing breast, cervical, and colorectal cancer screening among rural women: baseline characteristics of a randomized control trial. Contemp Clin Trials 2022 Dec;123:106986. [doi: 10.1016/j.cct.2022.106986] [Medline: 36328234]

Abbreviations

AI: artificial intelligence
BCS: breast cancer screening
DA: decision aid
FE: fixed effect
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
HP: health professional
IPDAS: International Patient Decision Aid Standards
MD: mean difference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RE: random effect
RoB 2: Risk-of-Bias 2
RR: relative risk/risk ratio
SDM: shared decision-making
SWiM: synthesis without meta-analysis

Edited by G Eysenbach, T de Azevedo Cardoso; submitted 30.08.24; peer-reviewed by R Trivedi; comments to author 18.10.24; revised version received 27.10.24; accepted 29.10.24; published 29.01.25.

Please cite as:
Villain P, Downham L, Le Bonniec A, Bauquier C, Mandrik O, Nadarzynski T, Donelle L, Murillo R, Tolma EL, Johnson S, Soler-Michel P, Smith R
Impact of Online Interactive Decision Tools on Women's Decision-Making Regarding Breast Cancer Screening: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e65974
URL: https://www.jmir.org/2025/1/e65974
doi:10.2196/65974
PMID:

©Patricia Villain, Laura Downham, Alice Le Bonniec, Charlotte Bauquier, Olena Mandrik, Tom Nadarzynski, Lorie Donelle, Raúl Murillo, Eleni L Tolma, Sonali Johnson, Patricia Soler-Michel, Robert Smith. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Families' Experiences With Family-Focused Web-Based Interventions for Improving Health: Qualitative Systematic Literature Review

Diana Zhu[1], BA, MS, MDiet; Aimee L Dordevic[1], PhD; Zoe E Davidson[1], BHlthSc, PhD; Simone Gibson[2,3], BSc, PhD

[1]Department of Nutrition Dietetics and Food, Monash University, Melbourne, Australia

[2]School of Clinical Sciences, Monash University, Melbourne, Australia

[3]Monash Centre for Scholarship in Health Education, Monash University, Melbourne, Australia

Corresponding Author:
Diana Zhu, BA, MS, MDiet
Department of Nutrition Dietetics and Food
Monash University
264 Ferntree Gully Rd
Level 1
Melbourne, 3168
Australia
Phone: 61 3 9902 4270
Email: diana.zhu@monash.edu

Abstract

Background: eHealth interventions can favorably impact health outcomes and encourage health-promoting behaviors in children. More insight is needed from the perspective of children and their families regarding eHealth interventions, including features influencing program effectiveness.

Objective: This review aimed to explore families' experiences with family-focused web-based interventions for improving health.

Methods: Five databases were searched on October 26, 2022—updated on October 24, 2023—for studies reporting qualitative data on participating children or their caregivers' experiences with web-based programs. Study identification was performed in duplicate and studies were independently appraised for quality. Thematic synthesis was undertaken on qualitative data extracted from the results section of each included article.

Results: Of 5524 articles identified, 28 articles were included. The studies examined the experiences of school-aged children (aged 5-18 years) and their caregivers (mostly mothers) with 26 web-based interventions that were developed to manage 17 different health conditions or influence health-supporting behaviors. Six themes were identified on families' experiences: connecting with others, agency of learning, program reputability or credibility, program flexibility, meeting participants' needs regarding program content or delivery, and impact on lifestyle.

Conclusions: Families positively perceived family-focused web-based interventions, finding value in quality connections and experiencing social support; intervention features aligned with behavioral and self-management principles. Key considerations were highlighted for program developers and health care professionals on ways to adapt eHealth elements to meet families' health-related needs. Continued research examining families' experiences with eHealth interventions is needed, including the experiences of families from diverse populations and distinguishing the perspectives of children, their caregivers, and other family members, to inform the expansion of family-focused eHealth interventions in health care systems.

Trial Registration: PROSPERO CRD42022363874; https://tinyurl.com/3xxa8enz

(J Med Internet Res 2025;27:e58774) doi:10.2196/58774

KEYWORDS

eHealth; family based; qualitative; pediatric health; internet; mobile phone

XSL•FO
RenderX

Introduction

Pediatric Health

A growing number of children are engaging in health risk behaviors or living with a health condition [1-6]. Children's management of health conditions greatly influences their physical, emotional, and psychosocial growth; development; and well-being as well as their health into adulthood [1,3,7-9].

The family unit, namely children and their immediate family members (ie, caregivers and siblings), plays an important role in shaping children's health-related behaviors [6-8,10,11]. Children also assume varied degrees of responsibility for their health with age and into adulthood [3,6]. Long-term multicomponent, multidisciplinary interventions incorporating behavioral change and self-management techniques (eg, disease education, goal setting, and self-monitoring) and involving the family unit are recommended for treating or managing childhood health conditions or encouraging health-promoting behaviors in children [3,12]. Such services are traditionally administered in person; offer limited enrollment with strict eligibility criteria; require extensive time from trained health care professionals (eg, physicians, allied health practitioners, and nurses); and have limited accessibility, particularly in areas of lower socioeconomic advantage or regional or remote areas [3,5,13].

Pediatric eHealth Interventions

Using digital technologies to adapt conventional services to a web-based setting—offering eHealth interventions—has the potential to overcome the limited accessibility and reach of traditional services for treating childhood health conditions or to influence health-supporting behaviors. Studies have suggested that the use of digital technologies (eg, mobile or video communication platforms or websites) allow the continuation of health care when conventional or specialized care, beyond primary care, is unavailable [5,14].

Existing literature has considered the breadth of eHealth interventions delivered through various eHealth modalities. A 2021 systematic review reported that in the last decade and up until the beginning of the COVID-19 pandemic, there has been a surge in eHealth interventions, largely developed to treat mental illnesses and noncommunicable diseases and delivered through telehealth platforms or mobile phones [15]. Another systematic review in 2021 [4] found that eHealth technologies were mostly used for monitoring, tracking, and reporting purposes. Existing research on eHealth interventions for treating or managing specific health conditions or encouraging health-promoting behaviors in children has focused on evaluating program effectiveness [3,9,12,16-18]. Studies have found that such interventions have improved condition-specific outcomes (eg, disease markers, symptom management, and adherence to disease management plans) and health behavior changes in children [3,12,17-20]. Web-based programs were proposed as a potentially favorable type of eHealth intervention for children and their caregivers [4,21].

With an increased number of families with children living with health conditions needing treatment, urgent action is needed to optimize the development and delivery of eHealth interventions for children and families. Although the evidence base on the effectiveness of such interventions is expanding, there is limited research exploring the experiences of eHealth interventions for children and their families. The needs, values, and perceptions of program end users (ie, children and their caregivers) are essential for program development; participants are key informants of their health, including engagement in health-supporting behaviors or medical treatment [22,23]. Understanding their experiences with eHealth interventions can provide valuable insight on the program's potential to impact the participating child's health, including on health outcomes overtime, sustainability of changes, design features to maximize program effectiveness, uptake and engagement, and mechanisms underlying children's and their families' health. There is a developing body of literature exploring participants' lived experiences with family-focused eHealth interventions to prevent or treat health conditions in children.

This qualitative systematic review aims to synthesize the viewpoints of children and their families on their experiences with family-focused web-based programs for improving health. Findings from this review will inform the development of eHealth interventions and enhance our understanding of the ability of eHealth interventions to meet the health-related needs of children and their families.

Methods

The qualitative systematic literature review was conducted and reported in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and the Enhancing Transparency in Reporting the Synthesis of Qualitative Research statements [24,25]. The review protocol was prospectively registered via PROSPERO (CRD42022363874). The PRISMA checklist for the review is included in Multimedia Appendix 1.

Eligibility Criteria

Eligibility criteria were determined using the Population, Intervention, Comparator, Outcome and Study design framework.

Population

Studies involving the family unit (at least a child aged ≤18 years and a caregiver) were included in this review. There were no limits on study participants regarding gender, ethnic or medical background, and locality.

Intervention

Included study interventions were web-based programs targeted at the family unit. A web-based program was defined as an eHealth intervention where the primary component of the program was completed on the web, including web-based modules and activities. The web-based intervention included the active participation of both the participating child and at least one caregiver (ie, the program included activities for both the participating child and caregiver to complete). The eHealth intervention could have been accessed via multiple technological modalities, such as computers, phones, and tablets. Interventions may have included other adjunct eHealth components (eg, a mobile app for monitoring, tracking, or reporting purposes;

XSL•FO
RenderX

email; and phone messaging or calls) or health care services (eg, feedback or support from medical and allied health practitioners). eHealth interventions delivered solely through smartphone apps were not included.

Studies on eHealth interventions where the web-based program was used to support conventional face-to-face health care interventions (including telehealth) or where the web-based component was not the primary part of the intervention were excluded. Studies on web-based interventions used as decision-making, screening, or assessment tools, where the program content was delivered primarily through live sessions (eg, videoconferencing sessions), or on programs targeting the caregiver or child exclusively were also excluded.

Comparator

A comparator was not specified for this review.

Outcomes

Included studies described the child or caregiver's perceptions of participating in a web-based program (eg, perceptions of the intervention as a whole or specific intervention features; reflections on occurrences or attitudes before, during, or after the intervention) using qualitative methods, such as interviews, focus groups, and open-ended responses retrieved through surveys.

Study Design

All study designs were considered for inclusion if they included a qualitative component exploring participants' experiences and were published in English. Review articles, doctoral theses, and conference abstracts were excluded.

Search Strategy

Ovid MEDLINE, Ovid Embase, Cochrane Library, Scopus, and CINAHL were searched for articles that met the eligibility criteria in October 2022 and updated in October 2023. Search results were limited to human studies published in English and within the last decade to capture evidence on the most up-to-date eHealth developments and updated versions of web-based interventions.

The search used both keyword and subheading search terms related to families, web-based programs, experiences or perspectives, and qualitative research methods. The search strategy for each database is included in Multimedia Appendix 2.

To test the validity of the search strategy, 3 key articles that met the inclusion criteria were identified [26-28]. The search strategy was developed with senior researchers with experience in pediatric research or interventions and qualitative research and confirmed with a university librarian. The reference lists of included articles were hand searched for additional relevant articles.

Study Selection and Data Extraction

Retrieved citations were exported into EndNote 20 software (Clarivate). Duplicates were removed and the remaining results were imported into and managed using Covidence (Veritas Health Innovation). Titles and abstracts were screened independently by at least 2 authors (DZ, SG, and ZED), after which full-text articles were screened independently against the eligibility criteria by 2 authors (DZ, SG, and ZED). Disagreements on the inclusion of articles were resolved by consensus.

Data were extracted using a bespoke data extraction template that was piloted by 2 authors (DZ and SG) before data extraction. Information extracted from articles included publication details (authors and year and country of publication); study aim; participant characteristics (participating children's age health status and the participating caregivers); study design (qualitative methodology and sample size); intervention characteristics (purpose and key features of the intervention and level of guidance or support provided throughout the intervention); and outcomes (themes and representative quotes relating to participants' experiences). Data were extracted independently by one author (DZ) and confirmed with a second author (SG); 3 papers were selected randomly where SG independently extracted data to compare and confirm the data extraction process.

Quality Appraisal

Studies were appraised using the Critical Appraisal Skills Programme qualitative checklist [29]. Included studies were independently assessed by one author (DZ) and confirmed with a second author (SG). Three papers were selected at random and independently appraised by both authors, after which assessment results were discussed, discrepancies were addressed, and the first author appraised the remaining papers.

Data Synthesis

Thematic synthesis [30] was undertaken on the extracted data about participating children and caregivers' experiences with the intervention using inductive line-by-line coding of the results and representative quotes extracted from the included studies. Four papers were selected at random and independently and manually coded by 2 authors (DZ and SG). Codes were discussed by both authors (DZ and SG) and refined, whereby the coding process was confirmed. The remaining papers were then coded by DZ, where new codes identified throughout the process were discussed and further verified by SG. One author (DZ) then independently developed descriptive categories from these codes that were discussed and critically reviewed regularly between 2 authors (DZ and SG) until a consensus was reached. These descriptive categories were then synthesized into overarching themes.

Individual and collaborative reflexivity were undertaken during the planning and implementation of the review methodology [31]. Authors were dietitians or nutritionists and researchers, with clinical or research experience in pediatric nutrition, weight management or lifestyle programs, or education. Most authors have been developing their knowledge of or capacity in qualitative methodology, and one author has extensive experience with qualitative research. All authors reflected on and acknowledged their personal and professional experiences by engaging in reflexive writing (eg, researcher notes and journaling) or team discussions throughout the review process. Authors collaborated as a team to resolve disagreements during

XSL•FO
RenderX

the screening process and when analyzing and reporting the data. During team discussions, authors also communicated their expertise in the review methodology, assumptions made during decision-making, and expectations of results.

Results

Search Results and Study Characteristics

The search identified 5524 articles after the removal of duplicates, of which 28 articles were included in the qualitative synthesis (refer to Figure 1 for the PRISMA diagram). Two articles included in this review reported findings from the same study [32,33]. The studies included in this review examined 26 distinct interventions, where 1 study explored 3 versions of an intervention (ie, adapted to 3 cultures) [34], 1 study examined 2 versions of an intervention (ie, partly guided versus entirely self-guided versions) [35], 5 studies evaluated 2 separate interventions (ie, including the conductance of process evaluations on the web-based program) [26,27,36-38], and the remaining studies investigated one web-based program each [28,39-56].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram outlining the study selection process.

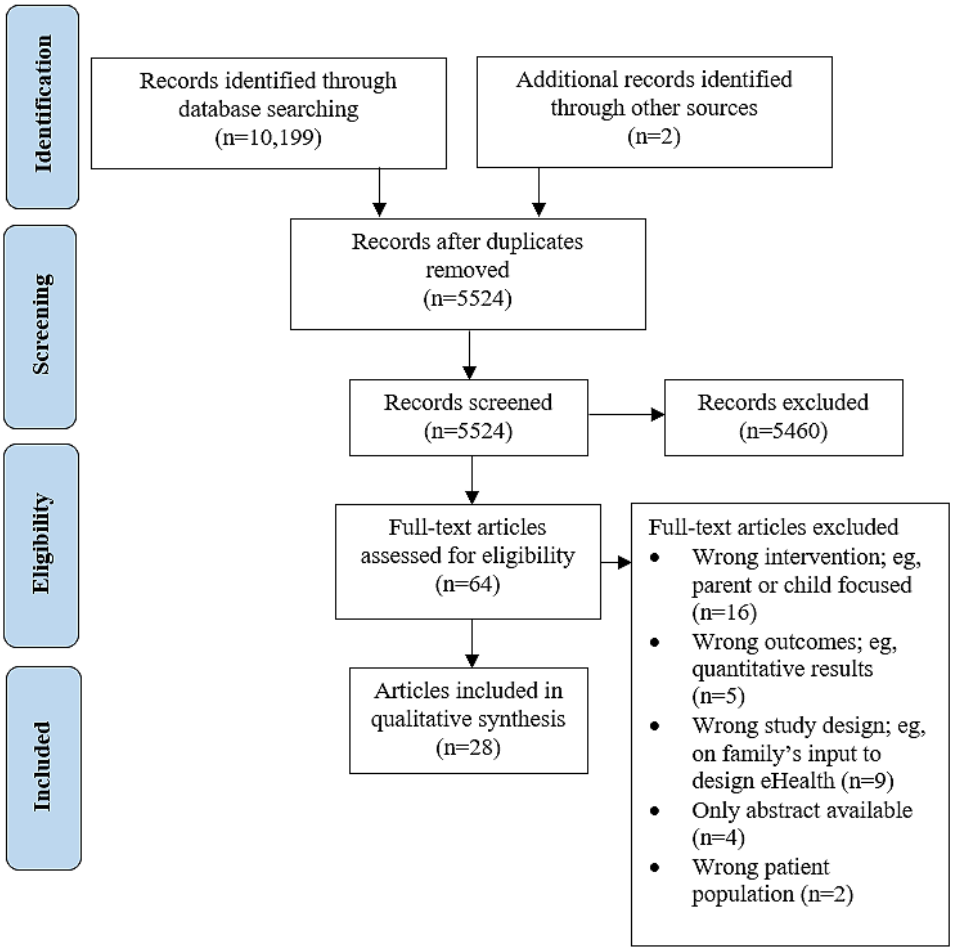

The characteristics of the included articles are summarized in Table 1. Studies examined web-based programs delivered in the United States [39-46], Canada [28,47-49], United Kingdom [34,50,51], England [32,33], Sweden [35,52-54], Australia [37,38,55], New Zealand [34,56], Italy [34], and Spain [26,27,36]. Participants included children in their school-aged years (aged 5-18 years): primary school–aged children [26,27,32,33,36,42,50,56], secondary school–aged children [28,34,35,37-41,43-46,51,52,54], or both primary and secondary school–aged children [47-49,53,55]. Children in 24 studies were formally diagnosed or screened to be at higher than normal risk for a health-related condition or risk factor by a health care professional (eg, physicians, allied health practitioners, and nurses) [26-28,32-45,47-49,51-54,56]. One study included families with children who may have self-identified with having a health disorder [55]. Of the included studies, 15 reported the gender of the participating caregivers; mostly mothers participated in the intervention program [32,35,37,39-44,48,50,51,53-55]. A total of 8 studies included information on the parent or caregiver's academic background, where all caregivers noted completing higher education [32,40,42-44,46,48,49]. Only 2 studies reported the socioeconomic statuses of the families [44,46], where families from both studies mostly reported having an average annual income.

XSL•FO
RenderX

Table 1. Study design of included studies and program characteristics.

Study	Study aim	Program characteristics (partly guided or self-guided; frequency or duration of intervention)	Family characteristics	Qualitative data collection method
Guagliano et al [50], 2019; United Kingdom	To assess the feasibility and acceptability of a web-based physical activity intervention (FRESH[a])	Self-guided; weekly challenges; 6 wk	• n=12 families • Children[b]: 8-10 y (average 8.3 y) • Caregivers[b]: 39.8 y (average) • Mostly mothers	Feedback questionnaire and semistructured focus with families at 6 wk
Hatfield et al [37], 2018; Australia	To determine the effectiveness, usability, and barriers and facilitators related to an online transition program for adolescents with autism spectrum disorder program (BOOST-A[c])	Partly guided; 2 mo	• n=39 families • Children[b]: secondary school–aged (8-11 y; average 14.8 y); mostly boys • Caregivers[b]: mostly mothers; mostly with moderate-high SES[d]	Feedback questionnaire and semistructured interviews with families within 2 mo post program completion
Jogova et al [47], 2013; Canada	To conduct a process evaluation of an online healthy lifestyle program for children with obesity (LiGHT[e])	Partly guided; 4 wk	• n=20 families participated in the intervention • Children[f]: 10-17 y (average 13-15.5 y); mostly girls	2-h focus groups with families and feedback questionnaires collected from families post program
Bevan Jones et al [51], 2020; United Kingdom	To evaluate the feasibility, acceptability, and potential impact of an online psychosocial program for children with a history of depression (MoodHwb[g])	Self-guided; 2 mo	• n=44 children; n=31 caregivers • Children[b]: ≥13 y (average 16.3 y); mostly girls • Caregivers[b]: mostly mothers	Feedback questionnaire and semistructured interviews with families post program completion (2 mo)
Khan et al [32][h], 2021; England	To explore the fidelity and experiences of families with a behavioral intervention for children with tics disorder (ORBIT[i])	Partly guided; 10-12 wk	• n=112 families received intervention • Children[f]: 9-16 y (average 12y); mostly boys with moderately severe symptoms • Caregivers[f]: mostly mothers; completed higher education[j]	Feedback questionnaire and semistructured interviews with families at post program completion (3 mo)
Lalloo et al [39], 2022; United States	To characterize families' engagement with a pain management intervention for children with SCD[k] (iCanCope with SCD)	Self-guided	• n=56 families; n=1 child received intervention • Children[b]: 12-18 y (average 14.8 y); mostly girls • Caregivers[b]: mostly mothers	Semistructured interviews with families within 6 mo post program completion
Lenhard et al [52], 2016; Sweden	To describe participants' experiences with a cognitive behavioral therapy program for children with obsessive compulsive disorder (BiP OCD[l])	Partly guided	• n=21 families received intervention • Children[f]: 12-17 y	Interviews with families at 3-6 mo post program completion
Muller et al [55], 2024; Australia	To gain parents' insights on their engagement with a program for the management of childhood anxiety (BRAVE[m] Self-Help)	Self-guided	• n=14 families received family version • Children[b]: 3-17 y • Caregivers[f]: average 44y; mostly mothers	Semistructured interviews with caregivers at 6 mo
Murray et al [40], 2022; United States	To evaluate the feasibility and acceptability of a pain management program for children undergoing spinal fusion surgery	Partly guided; weekly or fortnightly modules; presurgery period: 4-6 wk, postsurgery period: 6-8 wk	• n=13 families received the intervention • Children[b]: 12-17 y (average 14.3 y); mostly girls • Caregivers[b]: mostly mothers; mostly completed higher education and with high SES	Semistructured interviews with families 3 mo postsurgery

XSL•FO
RenderX

Study	Study aim	Program characteristics (partly guided or self-guided; frequency or duration of intervention)	Family characteristics	Qualitative data collection method
Nieto et al [36], 2015; Spain	To assess the feasibility of a pain management program for children with functional abdominal pain (DARWeb[n])	Self-guided; weekly modules	• n=15 families received intervention • Children[b]: 9-14 y; mostly girls	Feedback questionnaire and semistructured interviews with families at 2 wk post program completion
Nieto et al [26], 2019; Spain	To explore the impact and families' perspectives of a pain management program for children with functional abdominal pain (DARWeb)	Self-guided; weekly modules	• n=22 families • Children:f 9-14 y (average 11.23 y); mostly girls and with low pain severity	Semistructured interviews with families at post program completion
Nieto et al [27], 2019; Spain	To evaluate the efficacy of a pain management program for children with functional abdominal pain (DARWeb)	Self-guided; weekly modules	• n=25 families received intervention • Children[b]: 9-15 y (average 11.28 y); mostly girls	Feedback questionnaire and semistructured interviews with families post program completion (11 wk)
O'Sullivan et al [48], 2018; Canada	To evaluate the acceptability of a self-management program for Irish children with JIA[o] and their families (TTC[p])	Self-guided; weekly modules;12 wk	• n=20 families received intervention • Children[f]: 12-18 y (average 14.19 y); mostly girls • Caregivers[f]: mostly 40-49 y; mostly mothers who completed higher education	Focus groups or interviews with families at 2 wk program commencement
Palermo et al [41], 2018; United States	To evaluate the acceptability and feasibility of a cognitive behavioral intervention for children with sickle cell disease and their families (Web-MAP[q])	Partly guided	• n=15 families received CBT[r] version • Children[b]: 11-18 y (average 14.8 y); mostly girls • Caregivers[b]: mostly mothers with low to moderate SES	Semistructured with families at post program completion
Sonney et al [42], 2020; United States	To evaluate the feasibility, acceptability and efficacy of a sleep intervention for children with asthma and their families (SKIP[s])	Self-guided; weekly modules	• n=29 families received the intervention • Children[b]: 6-11 y (average 8.3 y) • Caregivers[b]: mostly mothers who completed higher education	Feedback questionnaire and semistructured interviews with families at 12 wk
Stasiak et al [56], 2018; New Zealand	To evaluate the feasibility and acceptability of a cognitive behavioral intervention for children with mild to moderate anxiety related to the experience of a natural disaster and their families in primary care (BRAVE-ONLINE[t])	Partly guided; weekly modules	• n=42 families received the intervention • Children[b]: 7-15 y (average 11.1 y); mostly with generalized anxiety disorder • Caregivers[b]: mostly with moderate-high SES	Feedback questionnaires with families at post program completion (12 wk)
Stinson et al [28], 2015; Canada	To explore the usability of a self-management program for children with cancer and their families (Teens Taking Charge: Managing Cancer Online)	Self-guided	• n=22 children, n=15 caregivers received intervention • Children[b]: 12-18 y (average 15.2 y); mostly boys	Semistructured interviews with families following receipt of intervention
Thompson et al [43], 2019; United States	To evaluate the feasibility and acceptability of a self-management program for children with type 1 diabetes and their families (FTO[u])	Self-guided; fortnightly modules; 3 mo	• n=27 families received intervention • Children[f]: 10-15 y; mostly girls • Caregivers[f]: mostly mothers who completed higher education and with moderate-high SES	Phone interviews with families post program completion
Thorén et al [53], 2021; Sweden	To explore parents' experiences with a lifestyle program for children with obesity (Web-COP[v])	Partly guided; weekly sessions; in-person group sessions–4 wk, web-based modules–12 wk	• n=51 families received intervention • Children[b]: 5-13 y; mostly girls • Caregivers[f]: mostly mothers with history of obesity	Semistructured interviews with caregivers at 2-4 mo post program completion

XSL•FO
RenderX

Study	Study aim	Program characteristics (partly guided or self-guided; frequency or duration of intervention)	Family characteristics	Qualitative data collection method
Wade et al [44], 2017; United States	To assess the feasibility and acceptability of a problem-solving and communication skills training program for children with traumatic brain injury and their families (TOPS[w])	Partly guided	• n=49 families received the family version • Children[b]: 11-18 y (average 14.7 y); mostly boys with moderate-severe brain injury • Caregivers[b]: mostly mothers who completed higher education and with moderate-high SES	Satisfaction surveys and interviews with families at 6 mo
Wade et al [34], 2021; New Zealand, United Kingdom, and Italy	To adapt a problem-solving and communication skills training program for children with traumatic brain injury and their families (TOPS) in New Zealand, United Kingdom, and Italy	Partly guided; New Zealand: 5 wk	• Childrenb: New Zealand: 12-17 y • Italy and United Kingdom: adolescent years	Focus groups with caregivers or families following receipt of the intervention in Italy and New Zealand; feedback questionnaire with families in the United Kingdom at post program completion (1 mo)
Yuen et al [45], 2016; United States	To develop and evaluate the usability of a psychoeducation intervention for children affected by a natural disaster and their families (BBN[x])	Self-guided	• n=24 children accessed intervention module • Children[f]: 12-17 y (average 14.12 y); mostly girls with some degree of PTSD[y]	Children were observed and provided feedback (verbal and written) while completing selected program module; feedback questionnaire and interviews with children after completing selected program module
Simonsson et al [54], 2021; Sweden	To explore the experiences of families with a treatment program for children with nonsuicidal self-injury disorder and their families (online ERITA[z])	Self-guided; 12 wk	• n=25 families received the intervention • Children[f]: 14-17 y; mostly girls • Caregivers[f]: 43-55 y; mostly mothers	Semistructured interviews with families post program completion
Lee et al [46], 2023; United States	To evaluate the effectiveness and usability of an intervention on human papillomavirus vaccination for Hmong-American families (Hmong Promoting Vaccines—HmongHPV website)	Partly guided; daily modules; 1 wk	• n=30 families received intervention • Children[f]: 12-16 y (mostly <15 y) • Caregivers[f]: mostly completed higher education and with average SES	Surveys with families post intervention completion (1 and 5 wk); interviews with families at 6 wk post intervention completion
Andersson et al [35], 2024; Sweden	To explore families' experiences completing an intervention for children with depression (online BA[aa])	Self-guided; 10 wk	• n=11 families received intervention programs • Children[f]: 13-17 y (average 15.2 y); mostly boys • Caregivers[f]: all mothers	Semistructured interviews with families at post program completion
Connan et al [49], 2019; Canada	To assess the usability of an intervention on the gluten free diet for children with celiac disease and type 1 diabetes and their families	Self-guided	• n=20 families recruited • Children[f]: mostly secondary school–aged years (average 13.4-13.5 y); mostly girls • Caregivers[f]: mostly completed higher education	Families were observed and engaged in usability interviews during and post module completion
Khan et al [33][h], 2022; England	To explore the factors influencing the efficacy and engagement of families with a behavioral intervention for children with tics disorder (ORBIT)	Partly guided; 10-12 wk	• n=112 families received intervention • Childrenb: 9-17 y (average 12.2 y); mostly boys with moderately severe symptoms	Semistructured interviews with families post program completion

XSL•FO
RenderX

Study	Study aim	Program characteristics (partly guided or self-guided; frequency or duration of intervention)	Family characteristics	Qualitative data collection method
Hatfield et al [38], 2017; Australia	To assess the feasibility of an online transition program for adolescents with autism spectrum disorder program (BOOST-A)	Partly guided	• n=6 families received intervention • Children[f]: secondary school–aged years (10 and 11 y); mostly boys	Surveys with families immediately following completion of each module

[a]FRESH: Families Reporting Every Step to Health.

[b]Baseline characteristics regardless of allocation.

[c]BOOST-A: Better Outcomes and Successful Transitions for Autism.

[d]SES: socioeconomic status; based on the Socio-Economic Indexes for Areas decile or higher SES with an annual income ≥US $70,000.

[e]LiGHT: Living Green, Healthy and Thrifty program.

[f]Characteristics of participants who engaged in the research's qualitative data collection component.

[g]MoodHwb: *Hwb* translates to hub, lift, or boost in Welsh.

[h]Same study protocol.

[i]ORBIT: Online Remote Behavioural Intervention for Tics.

[j]Higher education refers to completed tertiary education.

[k]SCD: sickle cell disease.

[l]BiP OCD: BarnInternetProjektet obsessive-compulsive disorder.

[m]BRAVE: Body signs, Relax, Activate helpful thoughts, Victory over fears, Enjoy yourself.

[n]DARWeb: Dolor Abdominal Recurrente web-based intervention.

[o]JIA: juvenile idiopathic arthritis.

[p]TTC: Teens Taking Charge.

[q]Web-MAP: Web-based Management of Adolescent Pain.

[r]CBT: cognitive behavioral therapy.

[s]SKIP: Sleep Intervention for Kids and Parents.

[t]BRAVE-ONLINE: Body signs, Relax, Activate helpful thoughts, Victory over fears, Enjoy yourself - online program.

[u]FTO: Family Teamwork Online.

[v]Web-COP: Web-based childhood obesity prevention.

[w]TOPS: Teen Online Problem Solving.

[x]BBN: Bounce Back Now.

[y]PTSD: posttraumatic stress disorder.

[z]ERITA: Emotion Regulation Individual Therapy for Adolescents.

[aa]BA: Behavioural Activation.

Most of the web-based programs were developed for the management of a medical condition (24 distinct interventions for the treatment or management of 17 unique health-related conditions: autism [37,38], overweight or obesity [47,53], depression [35,51], tic disorder [32,33], sickle cell disease [39,41], obsessive compulsive disorder [52], anxiety [55,56], spinal fusion [40], functional abdominal pain [26,27,36], juvenile idiopathic arthritis [48], sleep disturbance related to having asthma [42], cancer [28], type 1 diabetes [43,49], celiac disease [49], traumatic brain injury [34,44], nonsuicidal self-injury [54], and posttraumatic stress disorder [45]). Two interventions were developed to influence health-related behaviors in school-aged children (unrestricted to a medical diagnosis or condition) [46,50]. All but 2 interventions were entirely technology based (no in-person elements) [37,53]. A total of 19 studies reported on the intervention program's frequency [26,27,34,36,40,42,43,46,48,50,53,56] or length [32,33,35,37,39,40,43,46-48,50,51,53,54], where interventions mostly included weekly modules and lasted 4 to 12 weeks.

A total of 13 interventions were self-guided [26-28,35,36,39,42,43,45,48-51,54,55], of which 7 interventions included other eHealth components [26,27,36,39,42,43,50,54,55]. The eHealth components had personalized functions in 3 programs (ie, pedometer [50] and a mobile app used for monitoring or tracking purposes [39,54]) and automated functions in 4 programs (ie, email reminders [26,27,36,42,43,55]). Studies also examined partly guided programs [32-35,37,40,41,44,46,47,52,53,56]. A total of 12 interventions included adjunctive support from professionals (eg, doctoral or postdoctoral research fellows, psychologists, and exercise specialists) who used eHealth technologies such as a videoconferencing platform [34,44], a phone [32,33,35,40,47,56], an email [47], and a built-in messaging platform [32,33,35,52,56]. Adjunct features of included programs are summarized in Table 2.

XSL•FO
RenderX

Table 2. Key features of the intervention program (family-based version) of included studies.

Study	Purpose of intervention program	Key features of the intervention program							
		Email[a]	Phone[b]	Video-conf[c]	Built-in messaging platform	Smartphone app	Other eHealth monitoring technology[d]	In-person sessions	Support from a health care professional or expert
Guagliano et al [50], 2019	To increase physical activity among families						✓ (pedometer)		
Hatfield et al [37], 2018; [38], 2017	To support children with autism to prepare for life outside of school							__[e]	✓(champions [var[f]])
Jogova et al [47], 2013	To support children with obesity to build a healthy lifestyle behaviors or habits	✓ (resp[g])	✓ (resp)						✓ (exercise specialist)
Bevan Jones et al [51], 2020	Psychosocial program to support children with a history of depression								
Khan et al [32], 2021; [33], 2022	To support symptom management for children with tics disorder					✓ (stopwatch)			✓ (therapist)
Lalloo et al [39], 2022	To support symptom management for children with sickle cell disease					✓			
Lenhard et al [52], 2016	Treatment program for children with obsessive compulsive disorder				✓				✓ (psychologist)
Muller et al [55], 2024	Treatment program for children with anxiety	✓ (rem[h])							
Murray et al [40], 2022	To support symptom management for children undergoing spinal fusion surgery		✓ (call)						✓ (postdoctoral research fellows in psych[i])
Nieto et al [36], 2015; [26,27], 2019	To support pain management for children with functional abdominal pain	✓ (rem)							
O'Sullivan et al [48], 2018	To support symptom management for Irish children with juvenile idiopathic arthritis								
Palermo et al [41], 2018	To support symptom management for children with sickle cell disease				✓				✓ (therapist; MS[j] level or postdoctoral research fellows in psychology)
Sonney et al [42], 2020	To improve sleep in children with asthma	✓ (rem)							
Stasiak et al [56], 2018	Treatment program for children with mild to moderate anxiety related to the experience of a natural disaster		✓ (call)		✓				✓ (therapist)
Stinson et al [28], 2015	To support the management of symptoms for children with cancer								

XSL•FO
RenderX

Study	Purpose of intervention program	Key features of the intervention program							
		Email[a]	Phone[b]	Video-conf[c]	Built-in messaging platform	Smartphone app	Other eHealth monitoring technology[d]	In-person sessions	Support from a health care professional or expert
Thompson et al [43], 2019	To support the management of type 1 diabetes in children	✓ (rem)							
Thorén et al [53], 2021	To support lifestyle changes in children with obesity							✓	
Wade et al [44], 2017	To enhance problem-solving and communication skills in children with traumatic brain injury			✓					✓ (therapist; psychologist or graduate student)
Wade et al [34], 2021	To enhance problem-solving and communication skills in children with traumatic brain injury in New Zealand, United Kingdom, or Italy			✓ (Italy and United Kingdom)					✓ (therapist)
Yuen et al [45], 2016	Treatment program for children affected by a natural disaster								
Simonsson et al [54], 2021	Treatment program for children with nonsuicidal self-injury disorder					✓			
Lee et al [46], 2023	To improve the human papillomavirus vaccine rates among Hmong-American families		✓ (rem)				✓ (GPS locator)		✓ (Hmong-American health navigator)
Andersson et al [35], 2024	Treatment program for children with depression		✓ (therapist-supported version–phone calls)		✓ (therapist-supported version)				✓ (therapist-supported version—clinical psychologist)
Connan et al [49], 2019	Education program on the gluten free diet for children with coeliac disease and type 1 diabetes and their families								

[a]Reminder (automatic) or responsive communication.

[b]Reminder (responsive communication) or session call.

[c]Videoconf: videoconferencing.

[d]For example, pedometer, stopwatch, or GPS locator.

[e]Cannot tell.

[f]var: variable.

[g]resp: responsive.

[h]rem: reminder.

[i]psych: psychology.

[j]MS: master's.

Qualitative data were mostly collected using semistructured interviews conducted at post program completion (3 months post program completion or shorter) [26-28,32,33,35-37,39-46,48,50-55]. A total of 23 studies examined the experiences of the family unit [26-28,32-37,39-44,46-52,54-56], with only 1 study focusing solely on the point of views of the participating caregivers [53], and 1 study reporting on feedbacks of the participating children [45]. All except 1 study involved the delivery of the web-based program in its entirety to families; 1 study explored families' experiences with the intervention

XSL•FO
RenderX

following their completion of one representative program module [45].

Quality Appraisal

The Critical Appraisal Skills Programme checklist [29] completed for each study is summarized in Table 3. All studies were clear about their research aims, included qualitative methodology appropriately, and discussed the value of the research. All except 1 study [34] clearly described the recruitment strategy or the collection of all data. Similarly, all except 2 studies [38,44] clearly described the research design. All studies considered ethical issues; however, 1 study was unclear whether ethics approval was obtained [45]. Most studies adequately described the analysis process [26,27,35-37,40,42,45-48,51-55]. Most studies also provided a clear statement of findings [26-28,33,35-37,40-43,45,47-49,51-55]. It was unclear whether most studies adequately considered the relationship between researcher and participants; most studies lacked reports of authors' reflexivity or information on whether the researcher critically examined their own role, potential bias and influence during study design, data collection, analysis, and presentation [26-28,32-34,36,38-51,53,54,56].

XSL•FO
RenderX

Table 3. The Critical Appraisal Skills Programme qualitative studies checklist.

Study	Aim	Method	Design	Recruitment	Data collection	Relationship	Ethical	Data analysis	Findings	Value
Guagliano et al [50], 2019	Yes	Yes	Yes	Yes	Yes	No	Yes	Uncertain	Uncertain	Yes
Hatfield et al [37], 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Jogova et al [47], 2013	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Yes	Yes	Yes
Bevan Jones et al [51], 2020	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Yes	Yes	Yes
Khan et al [32], 2021	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Uncertain	Uncertain	Yes
Lalloo et al [39], 2022	Yes	Yes	Yes	Yes	Yes	No	Yes	Uncertain	Uncertain	Yes
Lenhard et al [52], 2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Muller et al [55], 2024	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Murray et al [40], 2022	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Nieto et al [36], 2015	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Nieto et al [26], 2019	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Nieto et al [27], 2019	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
O'Sullivan et al [48], 2018	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Palermo et al [41], 2018	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
Sonney et al [42], 2020	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Stasiak et al [56], 2018	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Uncertain	Yes
Stinson et al [28], 2015	Yes	Yes	Yes	Yes	Yes	No	Yes	Uncertain	Yes	Yes
Thompson et al [43], 2019	Yes	Yes	Yes	Yes	Yes	No	Yes	Uncertain	Yes	Yes
Thorén et al [53], 2021	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Yes	Yes	Yes
Wade et al [34], 2021	Yes	Yes	Yes	Uncertain	Yes	No	Yes	Uncertain	Uncertain	Yes
Wade et al [44], 2017	Yes	Yes	Uncertain	Yes	Uncertain	No	Yes	No	Uncertain	Yes
Yuen et al [45], 2016	Yes	Yes	Yes	Yes	Yes	No	Uncertain	Yes	Yes	Yes
Simonsson et al [54], 2021	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Yes	Yes	Yes
Lee et al [46], 2023	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Yes	Uncertain	Yes
Andersson et al [35], 2024	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

XSL•FO
RenderX

Study	Aim	Method	Design	Recruitment	Data collection	Relationship	Ethical	Data analysis	Findings	Value
Connan et al [49], 2019	Yes	Yes	Yes	Yes	Yes	Uncertain	Yes	Uncertain	Yes	Yes
Khan et al [33], 2022	Yes	Yes	Yes	Yes	Yes	No	Yes	Uncertain	Yes	Yes
Hatfield et al [38], 2017	Yes	Yes	Uncer-tain	Yes	Yes	No	Yes	Uncertain	Uncertain	Yes

Results of Data Syntheses

The key themes identified on families' experiences with family-focused web-based health programs were (1) connecting with others, (2) agency of learning, (3) program reputability or credibility, (4) program flexibility, (5) meeting participants' needs regarding program content or delivery, and (6) impact on lifestyle. Themes reported by authors of included studies are summarized in Multimedia Appendix 3 [26-28,32-56].

Theme 1: Connecting With Others

Connecting with others related to forming quality relationships and established by the constructs of the web-based program and the influence of interpersonal relationships on families' experiences with the intervention. Building new or strengthening existing relationships encouraged the uptake of and engagement with treatment [32,36,39,42,43,50,52-55]. Subthemes included connecting with own family members (internal), other families (external), experts or health care professionals, and artificial intelligence (AI; eg, relational or conversational agent and chatbot).

Connecting With Own Family Members (Internal)

Families reflected on connecting with own family members during the intervention. The interventions created an opportunity for family members to involve themselves or realize their roles in making changes to support health collectively [28,35,37,39,42,43,47,50,52-54] or complete the intervention [27,32,33,36,39,43,47,50,52,53]. Families described the active participation of the entire (ie, immediate) family or at least one other family member (eg, sibling, other parent, or grandparents) in the program, including having "within-family competitions" [50] and nonparticipating siblings using program resources [53]. Parents and children expressed the importance of involving the family unit or parents to achieve goals or engage in the web-based program [27,32,35,36,39,42,43,50,52-54]. For instance, parents were key supporters of participating children and motivated program engagement [27,32,35,36,39,42,50,52,53]. Some parents reported that their motivation to engage with the program was dependent on their child's engagement [55]. Children also positively perceived their parent's involvement throughout the program; children felt less alone and found it beneficial to have their parent partake with them [35,52].

Families described improved relationships with each other post program completion [35,44,50,54]. Parents suggested that the program helped them better understand their child and their experiences [35,54], "opened lines of communication" [47] or allowed them to "have common language" [54] that was not possible before the program. Similarly, parents wanted the intervention to provide more structured opportunities for them to converse or complete an activity with their child [46,54].

Connecting With Other Families (External)

Families expressed the benefits of having the opportunity to connect with other families during the intervention [36,39,47,48,50,53,55]. Parents recognized that establishing a sense of community or network was an important feature of conventional treatment programs and recommended these elements to be preserved in eHealth interventions [47,55]. Families suggested that building relationships and interacting with other families, whether through group sessions or eHealth technologies or program features, supports their engagement [53,55] with and enjoyment [39] of the program. Parents indicated that having the opportunity to connect with other families throughout the program was reassuring for them and helped them know that they are "not alone with [their] struggles" [55]. Families suggested that eHealth intervention features, including shared forums, web-based chats, and support groups, enable the exchange of ideas [36,39,47,53] and shared experiences [47,48,53,55].

Connecting With Experts, Health Care Professionals, or Responsive AI

Families conveyed the importance of receiving support from professionals [32,35,37,44,47,52-56] or AI [43]. Families appreciated the responsive support received from health care professionals or AI to navigate the treatment [37,55], address queries [33,35,43,47,52-54], or maintain their confidence or motivation throughout the program [32,35,52,55]. Families described connecting with or feeling supported by a health care professional—actual human [32,52]—or AI [43], like they "had never been before." For instance, families who participated in partly guided web-based programs, reflected on the increased accessibility and strengthened connections established with experts, including increased "self-disclosure" [52] and the likelihood of families to "ask direct questions at any time" [53] or any questions freely [54]. Some children noted that program materials may be used to enhance sessions with experts in conventional services (ie, during hybrid care) or serve as an alternate intervention in between sessions [51].

Families also expressed that having interactions and support from human experts preserved the merit of conventional services [54], including the notion that professionals "wanted to help them" [54] and were on their side [52]. Participants noted favorable features of these relationships (both human-to-human or human-to-AI): the responders being prompt [35,43,47,52,54], optimistic [47], and personable [35,43,52,54]. Conversely, some families suggested the need for in-person contact with experts to establish or re-establish engagement with treatment [52,55].

XSL•FO
RenderX

Having contact with an actual person was preferred by some families over interacting with AI [43,52,54]. Families from the study on a self-guided web-based program with inbuilt support from a relational agent noted the limited ability of AI to meet unique needs [43]. Similarly, families who completed partly guided intervention programs noted their ambivalence with having all things digital or receiving responses from "robots" (automated response) [52,54].

Theme 2: Agency of Learning

Agency of learning related to an individual's sense of self-responsibility for their learning and includes their intention, self-efficacy, autonomy, and motivation. The increase in learning agency noticed by participating children or caregivers in themselves or others influenced program uptake, engagement, or sustainability [32,35,37,39,42,47,50-54]. Families often reported that the participating children had low agency, namely low self-confidence or motivation for or ownership of plans to improve their health in the beginning of the intervention [37,39,47,50,53,55]. Parents also conveyed similar reluctance and low motivation in their family to engage with program activities, related to having negative preconceived thoughts (eg, having apprehension about completing an activity well) [37,50]. Conversely, some families perceived that the structure or contents of the intervention aligned with their intent and respected their autonomy, which supported their engagement and positive experiences [52,54,55]. Some children enjoyed "feeling that [they] had the treatment to [themselves] and [helped themselves] rather than just [receiving] help" [54]. Families described that the program provided resources, such as new knowledge and ideas, that empowered them to take active roles to improve their health; they gained an understanding [26,27,46,53,54] of and ways to address [37,39,42,46,51-54] their symptoms, condition, or situation. Families also reflected that goal setting was facilitated by the intervention and strengthened learner agency [39,40,42,47,51]; goal setting helped families recognize that change was "manageable and possible" [47]. They suggested achieving goals or making progress further empowered them [42,47,51].

However, some children recounted that increased self-responsibility influenced negative experiences throughout the intervention [54,55]. Whereas children appreciated the independence to manage and have ownership of their treatment or changes, their perceived failure to achieve goals and complete activities increased feelings of "shame and guilt and decreased self-confidence" [54]. Families also expressed that the self-guided nature of the intervention (ie, intervention encouraging independence) was counterproductive when they had "doubts about their own capabilities" [54] or "whether they were completing the program correctly [or] on the right track" [55]; some families experienced moments of decreased self-confidence during the program [54,55]. Some children expressed that these negative experiences reminded them of their disorder and decreased agency [54]. When participants had low learning agency, they were poorly engaged with the intervention [26,50,53,55].

Theme 3: Program Reputability or Credibility

Knowledge of the program developers or affiliates impacted families' engagement with the intervention. For instance, families perceived programs developed by well-known institutions or health care professionals as trustworthy and were therefore motivated to uptake or engage with the web-based program [47,55]. Similarly, families noted being more likely to participate in and trust programs recommended to them by "friends, allied health professionals, hospitals and educational facilities" [55]. Families expressed the desire to learn from these sources; participating in the program was their opportunity to access the evidence-based knowledge [55]. Families also regarded the intervention as credible when the information presented matched their prior knowledge or lived experiences [28].

Theme 4: Program Flexibility

Program flexibility relates to the adaptability of program components into families' ways of living and family's reflections on the ease of program use. Families expressed that they were able to adapt the intervention to their lifestyles [27,35,36,39,43-45,47,50,52-54]. For instance, families noted that they were able to choose when to engage with the program or at their own pace, including using program resources and scheduling services when deemed most necessary, while adhering to the overall timeline of the web-based program [27,35,36,47,52-54,56]. Engaging in the program also required an acceptable amount of their time [39,43,47] or was a valuable use of their time [42,47]. Families reported that they preferred the web-based delivery of the intervention [27,35,36,52,56]; the program supported their ability to "prioritize other activities in life, as well as [fit] the treatment into [their] weekly schedule" [52]. Families suggested that program features, such as the compatibility of the intervention across multiple devices [40,55] and being able to freely navigate content (eg, return to completed sections or skipping content) [46,49,55] were favorable and supported program engagement.

On the other hand, some families reflected on difficulty participating in the intervention due to a lack of time [26,35,39,50,53,55], competing priorities [26,32,34,35,39,50,55], or a lack of energy from a demanding schedule [39]. Families also described the incompatibility of the intervention program to their preferred platform as an inconvenience, which further "wasted" time (eg, families mentioned website features that did not work on their preferred digital device) [32,36-38,51,56].

Theme 5: Meeting Participants' Needs Regarding Program Content or Delivery

Families reflected on whether the intervention met their needs regarding the intervention program's content or delivery, including the appropriateness of program content for children or a subset of children with respect to their age or disease and symptom severity. Families expressed that positive experiences with the program were related to learning new, tangible information (eg, ideas and knowledge) [26,27,33,35-37,39-44,46-49,52-56]. Before participating in the intervention program, families shared that most of their condition-specific knowledge

XSL•FO
RenderX

were retrieved from searching the internet [49]. Families conveyed that including condition-specific elements [28,36,37,39-41,43,44,48,51,52] and age-appropriate language or content [39,47,51] supported program engagement. Families suggested that program design features often perceived to support children's participation or understanding of the content included, having greater use of friendly and complementary visuals [27,28,41,42,45,47,49,51,56], videos or animations [33,37-39,42,47], interactive content particularly for younger primary school–aged children [28,33,39,40,42,46,47,49,53], simple wording [28,35,39,47,51], affirmative or respectful language [51,53], and customizable features [28,34,39,45-48, 51], having rewards particularly physical rewards [33,50, 51], and being easy to navigate [28,39,41,42,45,49,51,53,55,56]. Families described that relatable and personable features of the program, such as the use of storylines featuring children as the main characters or case studies portraying relatable scenarios further motivated or encouraged their participation [27,28,34,36, 40,43-45,47,51,52,54,55].

Parents suggested that program elements that were unappealing to children included having an excessive amount of text [27,36,47] or using advanced language or jargon [47]. Families described the intervention as unsupportive of their needs when they found the program's content redundant [27,32,39,42,47], too simple [47], lengthy [38,39,49], or irrelevant [35,39,55]. Families also wanted the intervention to include further or more specific information on their condition, health-related behaviors, or skills [35,37,43,49,51]. Families found the program irrelevant when they had prior education or lived experiences with the disorder or situation for a longer duration [35,55]. Such families often perceived the intervention as an "immediate lifeline" [55] or early intervention for participants who were first diagnosed with or having an early onset of a condition or experiencing mild symptoms [28,39,48,51,55]. Conversely, some families perceived the intervention being more suitable for children experiencing more severe symptoms [26,35]. Families recommended the development of different versions of the program to meet the needs of children in different age groups; children in older age groups found the program too simple and unengaging [32,33,47].

Theme 6: Impact on Lifestyle

Families reflected on the sustained health-related behavioral changes they noticed or actively made during or following program completion, which impacted their lifestyle. Families described that the intervention was a first step that helped them improve engagement in health-supporting behaviors or establish health-supporting routines, including healthy eating, physical activity, and sleep [40,42,50,53]. For instance, parents noted that because their participation in the intervention program, they set new boundaries at home to support healthy behaviors or habits [53]. Some families reported that they continued practicing techniques or skills learned from the program, which in turn enhanced their quality of life [27,35,36,40,54] or allowed them to "return to regular activities" [35,40], even "normal life at 100%" [27].

Discussion

Principal Findings

This review identified 28 articles on 26 studies, resulting in 6 themes describing children or caregivers' perception of family- and web-based eHealth interventions. Descriptions of each theme encompassed families' reflections on program features regarded as valuable to them and that influenced their uptake of or engagement with the intervention.

Comparison With Existing Literature

On the basis of the studies included in this review, families mostly had positive experiences using web-based programs supporting health. This review found that families positively perceived elements of the web-based program that imitated features of their conventional counterparts delivered in person. Existing literature on eHealth interventions for children and their families suggests that as with essential features of conventional interventions, eHealth interventions should incorporate behavioral or self-management principles involving the family unit, such as symptom management, lifestyle behaviors, relationship management, and psychosocial management [3,12]. Similarly, this review found that families favored self-guided features aligned with behavioral and self-management principles of the web-based programs, including goal setting; self-monitoring and self-reflection tasks; having interactive, tangible, relatable, and credible content; and personable activities that are easily adaptable to daily living (ie, themes: agency of learning, program reputability or credibility, program flexibility, and meeting participants' needs regarding program content or delivery). The web-based nature of web-based eHealth programs further empowers families during treatment by enabling flexibility to access resources when needed and unconstrained by locality and in whether or how resources (eg, information) are used [57]. Families also shared that such features enabling this flexibility encouraged their engagement with the intervention or health-supporting behaviors.

The included studies highlighted the importance of quality relationships to bolster the health and well-being of children and their families. Evidence suggests that children with chronic health conditions feel lonelier compared to children lacking health conditions [58]. Loneliness detrimentally impacts health, including the trajectory of health conditions, and hinders children from seeking help or communicating with others [58]. The proliferated use of technology affects social interactions, and passive use (ie, using technology in a way that decreases opportunities for social connectedness) also risks exacerbated feelings of loneliness [59]. This review found that web-based eHealth interventions can promote connections for families—including families with children with health conditions—within and outside the family unit (ie, theme: connecting with others). As suggested by other literature examining children's perspectives of digital technologies, opportunities to engage with web-based communications preserved or enhanced children's relationships with others (eg, family members and friends) [57]. Findings from this review suggest that web-based eHealth programs can cultivate quality

XSL•FO
RenderX

connections with others, regardless of the absence of in-person interactions, and may therefore influence health-related outcomes in children. Our findings extend existing evidence on the health-supporting effects of social connections by providing insight into the benefits of designing web-based programs involving the family unit (ie, subtheme: connecting with own family members). A web-based program that involves children and their families, or at least a caregiver, can satisfy the sense of relatedness required for children to engage in the desired health-promoting activities or behaviors. Existing literature recognizes the importance of the social environment and personable support to facilitate learning and influence children's self-management of or responsibility for their health [3,60-62]. Web-based program features that enable social interactions encourage active learning; individuals are more likely to apply learned behaviors when informational and emotional support are received from others [60-62]. Children, in particular, learn through modeling or observing others [61,62]. School-aged children, particularly primary school–aged children (ie, children <12 years of age), significantly depend on their caregivers to manage their health; caregiver involvement is commonly characteristic of conventional behavioral interventions for children [8]. Depending on the level of guidance provided by the web-based program, direct digital translations of conventional health care services may lack humanized elements, including psychosocial support or human connections, which may deter achieving intended health outcomes [63]. We also found that families valued human connections and were more likely to engage with the intervention when the program supported or enhanced their relationships with others, including children with their caregivers and vice versa, families with other families with children undergoing similar health conditions or situations, and health care professionals (ie, theme: connecting with others).

Novel insights were gained on families' relationships with health care professionals or the role of health care professionals in family-focused web-based programs (ie, subtheme: connecting with experts, health care professionals, or responsive AI). It was found that families improved their communication with a health care professional throughout the web-based program. Participating children may have found comfort with the web-based nature of the intervention—related to them being able to participate in the intervention at a physical environment of their choice—which supported them to feel safe and actively communicate or build a rapport with others [57,61]. Web-based programs can cultivate collaborative relationships or social engagements between participating families and involved health care professionals, supported by features including web-based discussion forums, inbuilt messaging platforms, and phone or videoconferencing platforms. This collaborative relationship further increases the likelihood that children engage with the intervention (ie, the learning process) and assume increased responsibility for their health [60,61].

This review found that families may prefer support received from human-to-human over human-to-AI interactions throughout the intervention (ie, subtheme: connecting with experts, health care professionals, or responsive AI). This is consistent with previous literature on the limitations of AI in eHealth interventions to support the specialized or therapeutic needs of patients [63,64]. However, this review also noted families' appreciation for responsive AI when included in web-based programs due to its rapid responses and accessibility. Generative and responsive AI is rapidly developing, with increased efforts to humanize AI features and evidence that AI technology has the potential to provide compassionate health care [65,66]. It is important to note that the limitations of AI found in this review may not be applicable in the future [65]. Involving families in the development of AI features in eHealth interventions is essential for ensuring acceptable family-centered education [67].

This review also identified unmet needs experienced by families with the web-based program (ie, theme: meeting participants' needs regarding program content or delivery). Families noted unsatisfied informational needs, where the content was described as being too simple or irrelevant to them, given their age, education background, or experience with the pediatric health condition. Children have individual and dynamic needs with age and stage or degree of health conditions [8,60,64]. This finding suggests that web-based programs may satisfy the foundational needs of children and their families when managing health conditions. However, more complex and specialized needs, including diminishing motivation or troubleshooting decreased self-confidence or self-management experienced during eHealth interventions, may require support from a health care professional or more sophisticated integration of contemporary AI [12,64]. On the other hand, this finding may also suggest that web-based eHealth interventions can create opportunities for families to realize their unmet health-related needs or intentions to seek further support from health care professionals [57].

Strengths and Limitations

This is the first qualitative systematic review, to the researchers' knowledge, that synthesizes research reported in the last decade on the experiences of both the participating children and caregivers with family-focused web-based intervention programs. Whereas this review applied a search strategy that involved Boolean searching versus, for example, subject searching and was translated across databases to identify articles that met the eligibility criteria, the evidence base on family-focused web-based programs to improve children's health is still developing. The search method of this review was therefore chosen to maximize flexibility. The search strategy was also confirmed with a university librarian to ensure focus. A limitation of the identified studies is the lack of clarity on whether researchers engaged in reflexivity, which is a core component to ensure transparency of the research process and quality of qualitative research. Limitations of this review also include that findings are representative of research from economically advanced countries, where resources, including access to the internet, may be more available. It is unclear whether this review equitably captures the perspectives of families from diverse populations, as not all studies have reported on characteristics that were found to influence participants' experiences with eHealth interventions, including caregivers' education background or socioeconomic information [64,65]. Similarly, families in the identified studies may have

XSL•FO
RenderX

also disproportionally conveyed positive feedback with the web-based programs. This review also explored the experiences of the family unit. Most studies in this review reported on the combined perspectives of participating children and their caregivers. Children, caregivers, and other family members may have different understandings of or priorities and motivators with children's health that may influence their experiences with treatment or health management, including eHealth interventions. More research is needed to distinguish these perspectives of children, caregivers, caregiver-child dyads, and other family members on family-focused eHealth interventions and across disease trajectories and children's developmental stages. This review focused on web-based programs directed at the family unit, which may exclude families' experiences with other types of family-focused eHealth interventions such as smartphone app–based programs. This review also presents a snapshot of findings on families' experiences with web-based eHealth programs in the last decade. With progressive advancements in digital technologies and AI and families' adaptations to the digital culture, continued exploration of this phenomenon is needed to inform the development of family-focused eHealth interventions.

Implications for Practice

On the basis of the findings of this review, key considerations when developing eHealth interventions for improving children's health and incorporating eHealth in health care systems include those outlined in Textbox 1.

Textbox 1. Key considerations for developing eHealth interventions to improve children's health.

- Children and their families highly regard eHealth interventions directed at the family unit and that create opportunities for cultivating quality social connections within or external to participating families. Web-based education may be supported by interventions that leverage features recognized by families to influence children's agency or self-management skills relative to improving their health: goal setting, self-monitoring and self-reflection tasks, interactive, tangible, relatable and evidence-based content, and personable activities that are easily adaptable to daily living.
- Family members, namely caregivers, were identified as social factors to support children's engagement in health-supporting activities and behaviors. Self-guided eHealth interventions require features that are positively perceived by families and related to foundational or informational needs of families lacking prior experiences with the health condition and treatment.
- Families highlighted that web-based programs need to be adaptable to meet children's changing needs throughout the program, including features that respond to fluctuations in motivation and their often-dynamic experience of their health conditions.
- eHealth (ie, web-based) interventions may present opportunities for families to recognize unmet health-related needs.
- Responsive features have the potential to humanize self-guided eHealth interventions. Advancements in responsive AI may serve similar functions as health care professionals in web-based programs, including meeting families' health-related needs.

Conclusions

This review synthesized the evidence on families' experiences with family-focused eHealth (ie, web-based) interventions. Insights were gained on the potential of eHealth interventions to satisfy health-related needs of children and their families. Key considerations presented in this review highlight the need for program developers and health care systems to adapt the elements of eHealth interventions to child developmental stages and the complex dynamism of childhood health conditions. In doing so, the benefits of eHealth interventions may be maximized to meet the growing number of families needing treatment for childhood health conditions. More research is needed to equitably tailor family-focused eHealth (ie, web-based) interventions to diverse populations and explore this phenomenon in the context of novel advancements in digital technologies and AI. Future research could also distinguish the perspectives of children, their caregivers, and other family members on eHealth interventions. These considerations can be used to enhance the design of family-focused eHealth interventions and better inform their inclusion in treatment plans to improve children's health.

Acknowledgments

The authors would like to thank the university librarian Mario Sos for supporting the refinement of the search strategy of this review.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ZED, SG, and DZ conceptualized and designed the review, and screened for articles that met the predefined eligibility criteria. DZ extracted, analyzed, and interpreted the data, and wrote the original draft. SG verified the analysis process and interpretation of the data. All authors contributed to the drafting of the manuscript and have approved the final version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

The PhD stipend of DZ is provided by the Better Health Company for the evaluation of a family-focused web-based healthy lifestyle program. ZED is the senior investigator on the research funded by the Better Health Company. Better Health Company had no role in any aspect of this review. The authors conceptualized, completed, and submitted the review independently. All other authors declare no other conflict of interest.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v27i1e58774_app1.docx]

Multimedia Appendix 2
Search strategy per database.
[DOCX File , 18 KB - jmir_v27i1e58774_app2.docx]

Multimedia Appendix 3
Themes or subthemes, categories, or codes reported in the results section of included studies.
[DOCX File , 26 KB - jmir_v27i1e58774_app3.docx]

References

1. Compas BE, Jaser SS, Dunn MJ, Rodriguez EM. Coping with chronic illness in childhood and adolescence. Annu Rev Clin Psychol 2012 Apr 27;8(1):455-480 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032511-143108] [Medline: 22224836]
2. Judson L. Global childhood chronic illness. Nurs Adm Q 2004;28(1):60-66. [doi: 10.1097/00006216-200401000-00013] [Medline: 14986511]
3. Stinson J, Wilson R, Gill N, Yamada J, Holt J. A systematic review of internet-based self-management interventions for youth with health conditions. J Pediatr Psychol 2009 Jun 23;34(5):495-510. [doi: 10.1093/jpepsy/jsn115] [Medline: 19029142]
4. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021 Nov 17;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
5. Perrin JM, Anderson LE, Van Cleave J. The rise in chronic conditions among infants, children, and youth can be met with continued health system innovations. Health Aff (Millwood) 2014 Dec;33(12):2099-2105. [doi: 10.1377/hlthaff.2014.0832] [Medline: 25489027]
6. Nightingale R, McHugh G, Kirk S, Swallow V. Supporting children and young people to assume responsibility from their parents for the self-management of their long-term condition: an integrative review. Child Care Health Dev 2019 Mar 20;45(2):175-188 [FREE Full text] [doi: 10.1111/cch.12645] [Medline: 30690751]
7. Dall'Oglio I, Gasperini G, Carlin C, Biagioli V, Gawronski O, Spitaletta G, et al. Self-care in pediatric patients with chronic conditions: a systematic review of theoretical models. Int J Environ Res Public Health 2021 Mar 28;18(7):3513 [FREE Full text] [doi: 10.3390/ijerph18073513] [Medline: 33800684]
8. Brigden A, Parslow RM, Linney C, Higson-Sweeney N, Read R, Loades M, et al. How are behavioural interventions delivered to children (5-11 years old): a systematic mapping review. BMJ Paediatr Open 2019 Dec 10;3(1):e000543 [FREE Full text] [doi: 10.1136/bmjpo-2019-000543] [Medline: 31909219]
9. de Sousa D, Fogel A, Azevedo J, Padrão P. The effectiveness of web-based interventions to promote health behaviour change in adolescents: a systematic review. Nutrients 2022 Mar 16;14(6):1258 [FREE Full text] [doi: 10.3390/nu14061258] [Medline: 35334915]
10. Davison KK, Lawson HA, Coatsworth JD. The family-centered action model of intervention layout and implementation (FAMILI): the example of childhood obesity. Health Promot Pract 2012 Jul 31;13(4):454-461. [doi: 10.1177/1524839910377966] [Medline: 21632465]
11. Catarino M, Charepe Z, Festas C. Promotion of self-management of chronic disease in children and teenagers: scoping review. Healthcare (Basel) 2021 Nov 27;9(12):1642 [FREE Full text] [doi: 10.3390/healthcare9121642] [Medline: 34946368]
12. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949. [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
13. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008 May 15;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
14. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]

XSL•FO
RenderX

15. da Fonseca MH, Kovaleski F, Picinin CT, Pedroso B, Rubbo P. E-health practices and technologies: a systematic review from 2014 to 2019. Healthcare (Basel) 2021 Sep 10;9(9):1192 [FREE Full text] [doi: 10.3390/healthcare9091192] [Medline: 34574966]
16. Brown HE, Atkin AJ, Panter J, Wong G, Chinapaw MJ, van Sluijs EM. Family-based interventions to increase physical activity in children: a systematic review, meta-analysis and realist synthesis. Obes Rev 2016 Apr;17(4):345-360 [FREE Full text] [doi: 10.1111/obr.12362] [Medline: 26756281]
17. Azevedo LB, Stephenson J, Ells L, Adu-Ntiamoah S, DeSmet A, Giles EL, et al. The effectiveness of e-health interventions for the treatment of overweight or obesity in children and adolescents: a systematic review and meta-analysis. Obes Rev 2022 Feb 07;23(2):e13373. [doi: 10.1111/obr.13373] [Medline: 34747118]
18. Kracht CL, Hutchesson M, Ahmed M, Müller AM, Ashton LM, Brown HM, et al. E- and mHealth interventions targeting nutrition, physical activity, sedentary behavior, and/or obesity among children: a scoping review of systematic reviews and meta-analyses. Obes Rev 2021 Dec 02;22(12):e13331 [FREE Full text] [doi: 10.1111/obr.13331] [Medline: 34476890]
19. Fowler LA, Grammer AC, Staiano AE, Fitzsimmons-Craft EE, Chen L, Yaeger LH, et al. Harnessing technological solutions for childhood obesity prevention and treatment: a systematic review and meta-analysis of current applications. Int J Obes (Lond) 2021 May 24;45(5):957-981 [FREE Full text] [doi: 10.1038/s41366-021-00765-x] [Medline: 33627775]
20. Chai LK, Farletti R, Fathi L, Littlewood R. A rapid review of the impact of family-based digital interventions for obesity prevention and treatment on obesity-related outcomes in primary school-aged children. Nutrients 2022 Nov 15;14(22):4837 [FREE Full text] [doi: 10.3390/nu14224837] [Medline: 36432522]
21. Burrows T, Hutchesson M, Chai LK, Rollo M, Skinner G, Collins C. Nutrition interventions for prevention and management of childhood obesity: what do parents want from an eHealth program? Nutrients 2015 Dec 15;7(12):10469-10479 [FREE Full text] [doi: 10.3390/nu7125546] [Medline: 26694456]
22. Darbyshire P, MacDougall C, Schiller W. Multiple methods in qualitative research with children: more insight or just more? Qual Res 2005 Nov 01;5(4):417-436. [doi: 10.1177/1468794105056921]
23. Vaknin O, Zisk-Rony RY. Including children in medical decisions and treatments: perceptions and practices of healthcare providers. Child Care Health Dev 2011 Jul;37(4):533-539. [doi: 10.1111/j.1365-2214.2010.01153.x] [Medline: 20854447]
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
25. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12(1):181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
26. Nieto R, Boixadós M, Hernández E, Beneitez I, Huguet A, McGrath P. Quantitative and qualitative testing of DARWeb: an online self-guided intervention for children with functional abdominal pain and their parents. Health Informatics J 2019 Dec 04;25(4):1511-1527 [FREE Full text] [doi: 10.1177/1460458218779113] [Medline: 29865899]
27. Nieto R, Boixadós M, Ruiz G, Hernández E, Huguet A. Effects and experiences of families following a web-based psychosocial intervention for children with functional abdominal pain and their parents: a mixed-methods pilot randomized controlled trial. J Pain Res 2019;12:3395-3412 [FREE Full text] [doi: 10.2147/JPR.S221227] [Medline: 32099447]
28. Stinson J, Gupta A, Dupuis F, Dick B, Laverdière C, LeMay S, et al. Usability testing of an online self-management program for adolescents with cancer. J Pediatr Oncol Nurs 2015;32(2):70-82. [doi: 10.1177/1043454214543021] [Medline: 25037173]
29. CASP qualitative checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist/ [accessed 2023-12-19]
30. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
31. Olmos-Vega FM, Stalmeijer RE, Varpio L, Kahlke R. A practical guide to reflexivity in qualitative research: AMEE guide no. 149. Med Teach 2022 Apr 07:1-11. [doi: 10.1080/0142159X.2022.2057287] [Medline: 35389310]
32. Khan K, Hollis C, Hall CL, Murray E, Davies EB, Andrén P, et al. Fidelity of delivery and contextual factors influencing children's level of engagement: process evaluation of the online remote behavioral intervention for tics trial. J Med Internet Res 2021 Jun 21;23(6):e25470 [FREE Full text] [doi: 10.2196/25470] [Medline: 34152270]
33. Khan K, Hollis C, Hall CL, Davies EB, Murray E, Andrén P, et al. Factors influencing the efficacy of an online behavioural intervention for children and young people with tics: process evaluation of a randomised controlled trial. J Behav Cogn Ther 2022 Sep;32(3):197-206. [doi: 10.1016/j.jbct.2022.02.005]
34. Wade SL, Jones KM, Corti C, Adlam AR, Limond J, Bardoni A, et al. Adapting intervention approaches to new contexts: three case studies of international adaptation of the Teen Online Problem Solving (TOPS) program. Rehabilitation Psychology 2021 Nov;66(4):356-365. [doi: 10.1037/rep0000414]
35. Andersson R, Vigerland S, Ahlen J, Widström H, Unger I, Serlachius E, et al. "Therapy without a therapist?" the experiences of adolescents and their parents of online behavioural activation for depression with and without therapist support. Eur Child Adolesc Psychiatry 2024 Jan 17;33(1):105-114 [FREE Full text] [doi: 10.1007/s00787-023-02142-7] [Medline: 36650254]

XSL•FO
RenderX

36. Nieto R, Hernández E, Boixadós M, Huguet A, Beneitez I, McGrath P. Testing the feasibility of DARWeb: an online intervention for children with functional abdominal pain and their parents. Clin J Pain 2015 Jun;31(6):493-503. [doi: 10.1097/AJP.0000000000000199] [Medline: 25551478]
37. Hatfield M, Falkmer M, Falkmer T, Ciccarelli M. Process evaluation of the BOOST-A™ transition planning program for adolescents on the autism spectrum: a strengths-based approach. J Autism Dev Disord 2018 Feb 10;48(2):377-388. [doi: 10.1007/s10803-017-3317-8] [Medline: 29019012]
38. Hatfield M, Murray N, Ciccarelli M, Falkmer T, Falkmer M. Pilot of the BOOST-A™: an online transition planning program for adolescents with autism. Aust Occup Ther J 2017 Dec 19;64(6):448-456. [doi: 10.1111/1440-1630.12410] [Medline: 28722152]
39. Lalloo C, Nishat F, Zempsky W, Bakshi N, Badawy S, Ko YJ, et al. Characterizing user engagement with a digital intervention for pain self-management among youth with sickle cell disease and their caregivers: subanalysis of a randomized controlled trial. J Med Internet Res 2022 Aug 30;24(8):e40096 [FREE Full text] [doi: 10.2196/40096] [Medline: 36040789]
40. Murray CB, Bartlett A, Meyyappan A, Palermo TM, Aaron R, Rabbitts J. A pilot feasibility and acceptability study of an internet-delivered psychosocial intervention to reduce postoperative pain in adolescents undergoing spinal fusion. Can J Pain 2022 Apr 13;6(1):12-23 [FREE Full text] [doi: 10.1080/24740527.2021.2009334] [Medline: 35434455]
41. Palermo TM, Dudeney J, Santanelli JP, Carletti A, Zempsky WT. Feasibility and acceptability of internet-delivered cognitive behavioral therapy for chronic pain in adolescents with sickle cell disease and their parents. J Pediatr Hematol Oncol 2018 Mar;40(2):122-127 [FREE Full text] [doi: 10.1097/MPH.0000000000001018] [Medline: 29176462]
42. Sonney JT, Thompson HJ, Landis CA, Pike KC, Chen ML, Garrison MM, et al. Sleep intervention for children with asthma and their parents (SKIP Study): a novel web-based shared management pilot study. J Clin Sleep Med 2020 Jun 15;16(6):925-936 [FREE Full text] [doi: 10.5664/jcsm.8374] [Medline: 32056537]
43. Thompson D, Callender C, Gonynor C, Cullen KW, Redondo MJ, Butler A, et al. Using relational agents to promote family communication around type 1 diabetes self-management in the diabetes family teamwork online intervention: longitudinal pilot study. J Med Internet Res 2019 Sep 13;21(9):e15318 [FREE Full text] [doi: 10.2196/15318] [Medline: 31538940]
44. Wade SL, Narad ME, Kingery KM, Taylor HG, Stancin T, Kirkwood MW, et al. Teen online problem solving for teens with traumatic brain injury: rationale, methods, and preliminary feasibility of a teen only intervention. Rehabil Psychol 2017 Aug;62(3):290-299 [FREE Full text] [doi: 10.1037/rep0000160] [Medline: 28836809]
45. Yuen EK, Gros K, Welsh KE, McCauley J, Resnick HS, Danielson CK, et al. Development and preliminary testing of a web-based, self-help application for disaster-affected families. Health Informatics J 2016 Sep 26;22(3):659-675 [FREE Full text] [doi: 10.1177/1460458215579292] [Medline: 25933798]
46. Lee HY, Xiong S, Sur A, Khang T, Vue B, Culhane-Pera KA, et al. Evaluating human papillomavirus eHealth in Hmong adolescents to promote vaccinations: pilot feasibility study. JMIR Form Res 2023 Jun 20;7:e38388 [FREE Full text] [doi: 10.2196/38388] [Medline: 37338961]
47. Jogova M, Song JE, Campbell AC, Warbuton D, Warshawski T, Chanoine JP. Process evaluation of the Living Green, Healthy and Thrifty (LiGHT) web-based child obesity management program: combining health promotion with ecology and economy. Can J Diabetes 2013 Apr;37(2):72-81. [doi: 10.1016/j.jcjd.2013.03.359] [Medline: 24070796]
48. O'Sullivan G, O'Higgins S, Caes L, Saetes S, McGuire BE, Stinson J. Self-management needs of Irish adolescents with Juvenile Idiopathic Arthritis (JIA): how can a Canadian web-based programme meet these needs? Pediatr Rheumatol Online J 2018 Nov 08;16(1):68 [FREE Full text] [doi: 10.1186/s12969-018-0287-0] [Medline: 30409209]
49. Connan V, Marcon MA, Mahmud FH, Assor E, Martincevic I, Bandsma RH, et al. Online education for gluten-free diet teaching: development and usability testing of an e-learning module for children with concurrent celiac disease and type 1 diabetes. Pediatr Diabetes 2019 May 10;20(3):293-303. [doi: 10.1111/pedi.12815] [Medline: 30652421]
50. Guagliano JM, Brown HE, Coombes E, Hughes C, Jones AP, Morton KL, et al. The development and feasibility of a randomised family-based physical activity promotion intervention: the Families Reporting Every Step to Health (FRESH) study. Pilot Feasibility Stud 2019 Feb 9;5(1):21 [FREE Full text] [doi: 10.1186/s40814-019-0408-7] [Medline: 30788135]
51. Bevan Jones R, Thapar A, Rice F, Mars B, Agha SS, Smith D, et al. A digital intervention for adolescent depression (MoodHwb): mixed methods feasibility evaluation. JMIR Ment Health 2020 Jul 17;7(7):e14536 [FREE Full text] [doi: 10.2196/14536] [Medline: 32384053]
52. Lenhard F, Vigerland S, Engberg H, Hallberg A, Thermaenius H, Serlachius E. "On my own, but not alone" - adolescents' experiences of internet-delivered cognitive behavior therapy for obsessive-compulsive disorder. PLoS One 2016 Oct 6;11(10):e0164311 [FREE Full text] [doi: 10.1371/journal.pone.0164311] [Medline: 27711249]
53. Thorén A, Janson A, Persson M. 'Now she prefers jeans, like everyone else…' - parents' experiences of group- and web-based treatment of children's obesity. Acta Paediatr 2021 Jun 15;110(6):1869-1879. [doi: 10.1111/apa.15798] [Medline: 33554379]
54. Simonsson O, Engberg H, Bjureberg J, Ljótsson B, Stensils J, Sahlin H, et al. Experiences of an online treatment for adolescents with nonsuicidal self-injury and their caregivers: qualitative study. JMIR Form Res 2021 Jul 23;5(7):e17910 [FREE Full text] [doi: 10.2196/17910] [Medline: 34297001]
55. Muller JL, Tomlin L, March S, Jackson B, Budden T, Law KH, et al. Understanding parent perspectives on engagement with online youth-focused mental health programs. Psychol Health 2024 May 27;39(5):613-630. [doi: 10.1080/08870446.2022.2090561] [Medline: 35758102]

XSL•FO
RenderX

56. Stasiak K, Merry SN, Frampton C, Moor S. Delivering solid treatments on shaky ground: feasibility study of an online therapy for child anxiety in the aftermath of a natural disaster. Psychother Res 2018 Jul;28(4):643-653. [doi: 10.1080/10503307.2016.1244617] [Medline: 27781568]
57. Bitto Urbanova L, Madarasova Geckova A, Dankulincova Veselska Z, Capikova S, Holubcikova J, van Dijk JP, et al. Technology supports me: perceptions of the benefits of digital technology in adolescents. Front Psychol 2022 Jan 30;13:970395 [FREE Full text] [doi: 10.3389/fpsyg.2022.970395] [Medline: 36798642]
58. Maes M, Van den Noortgate W, Fustolo-Gunnink SF, Rassart J, Luyckx K, Goossens L. Loneliness in children and adolescents with chronic physical conditions: a meta-analysis. J Pediatr Psychol 2017 Jul 01;42(6):622-635. [doi: 10.1093/jpepsy/jsx046] [Medline: 28340072]
59. Nowland R, Necka EA, Cacioppo JT. Loneliness and social internet use: pathways to reconnection in a digital world? Perspect Psychol Sci 2018 Jan;13(1):70-87. [doi: 10.1177/1745691617713052] [Medline: 28937910]
60. Ryan P, Sawin KJ. The individual and family self-management theory: background and perspectives on context, process, and outcomes. Nurs Outlook 2009 Jul;57(4):217-25.e6 [FREE Full text] [doi: 10.1016/j.outlook.2008.10.004] [Medline: 19631064]
61. Li S, Hong YC, Craig SD. A systematic literature review of social learning theory in online learning environments. Educ Psychol Rev 2023 Nov 11;35(4):108. [doi: 10.1007/S10648-023-09827-0]
62. Hajli M, Bugshan H, Lin X, Featherman M. From e-learning to social learning – a health care study. Eur J Train Dev 2013;37(9):2012-2863. [doi: 10.1108/ejtd-10-2012-0062]
63. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Nov 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]
64. Griffin AC, Xing Z, Khairat S, Wang Y, Bailey S, Arguello J, et al. Conversational agents for chronic disease self-management: a systematic review. AMIA Annu Symp Proc 2020;2020:504-513 [FREE Full text] [Medline: 33936424]
65. Lind SJ. Can AI-powered avatars replace human trainers? An empirical test of synthetic humanlike spokesperson applications. J Workplace Learn 2025 Jan 2;37(1):19-40. [doi: 10.1108/jwl-04-2024-0075]
66. Morrow E, Zidaru T, Ross F, Mason C, Patel KD, Ream M, et al. Artificial intelligence technologies and compassion in healthcare: a systematic scoping review. Front Psychol 2023 Jan 17;13:971044 [FREE Full text] [doi: 10.3389/fpsyg.2022.971044] [Medline: 36733854]
67. Adus S, Macklin J, Pinto A. Exploring patient perspectives on how they can and should be engaged in the development of artificial intelligence (AI) applications in health care. BMC Health Serv Res 2023 Oct 26;23(1):1163 [FREE Full text] [doi: 10.1186/s12913-023-10098-2] [Medline: 37884940]

Abbreviations

AI: artificial intelligence
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung; submitted 25.03.24; peer-reviewed by A Jolliff, L Ontai; comments to author 27.09.24; revised version received 22.11.24; accepted 30.11.24; published 30.01.25.

Please cite as:
Zhu D, Dordevic AL, Davidson ZE, Gibson S
Families' Experiences With Family-Focused Web-Based Interventions for Improving Health: Qualitative Systematic Literature Review
J Med Internet Res 2025;27:e58774
URL: https://www.jmir.org/2025/1/e58774
doi:10.2196/58774
PMID:

©Diana Zhu, Aimee L Dordevic, Zoe E Davidson, Simone Gibson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Metrics for Evaluating Telemedicine in Randomized Controlled Trials: Scoping Review

Yuka Sugawara[1], MD, PhD; Yosuke Hirakawa[1], MD, PhD; Masao Iwagami[2], MSc, MD, PhD; Ryota Inokuchi[3], MD, PhD; Rie Wakimizu[4], PhD; Masaomi Nangaku[1], MD, PhD

[1]Division of Nephrology and Endocrinology, The University of Tokyo, Tokyo, Japan

[2]Department of Health Services Research, Institute of Medicine, University of Tsukuba, Ibaraki, Japan

[3]Department of Clinical Engineering, The University of Tokyo Hospital, Tokyo, Japan

[4]Department of Child Health and Development Nursing, Institute of Medicine, University of Tsukuba, Ibaraki, Japan

Corresponding Author:
Masaomi Nangaku, MD, PhD
Division of Nephrology and Endocrinology
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 3815 5411
Email: mnangaku@m.u-tokyo.ac.jp

Abstract

Background: Telemedicine involves medical, diagnostic, and treatment-related services using telecommunication technology. Not only does telemedicine contribute to improved patient quality of life and satisfaction by reducing travel time and allowing patients to be seen in their usual environment, but it also has the potential to improve disease management by making it easier for patients to see a doctor. Recently, owing to IT developments, research on telemedicine has been increasing; however, its usefulness and limitations in randomized controlled trials remain unclear because of the multifaceted effects of telemedicine. Furthermore, the specific metrics that can be used as cross-disciplinary indicators when comparing telemedicine and face-to-face care also remain undefined.

Objective: This review aimed to provide an overview of the general and cross-disciplinarity metrics used to compare telemedicine with in-person care in randomized controlled trials. In addition, we identified previously unevaluated indicators and suggested those that should be prioritized in future clinical trials.

Methods: MEDLINE and Embase databases were searched for publications that met the inclusion criteria according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis Extension for Scoping Reviews). Original, English-language articles on randomized controlled trials comparing some forms of telemedicine with face-to-face care from January 2019 to March 2024 were included, and the basic information and general metrics used in these studies were summarized.

Results: Of the 2275 articles initially identified, 79 were included in the final analysis. The commonly used metrics that can be used across medical specialties were divided into the following 3 categories: (1) patient-centeredness (67/79, 85%), including patient satisfaction, workload, and quality of life; (2) patient outcomes (57/79, 72%), including general clinical parameters such as death, admission, and adverse events; and (3) cost-effectiveness (40/79, 51%), including cost assessment and quality-adjusted life year. Notably, only 25 (32%) of 79 studies evaluated all the 3 categories. Other metrics, such as staff convenience, system usability, and environmental impact, were extracted as indicators in different directions from the three categories above, although few previous reports have evaluated them (staff convenience: 8/79, 10%; system usability: 3/79, 4%; and environmental impact: 2/79, 3%).

Conclusions: A significant variation was observed in the metrics used across previous studies. Notably, general indicators should be used to enhance the understandability of the results for people in other areas, even if disease-specific indicators are used. In addition, indicators should be established to include all three commonly used categories of measures to ensure a comprehensive evaluation: patient-centeredness, patient outcomes, and cost-effectiveness. Staff convenience, system usability, and environmental impact are important indicators that should be used in future trials. Moreover, standardization of the evaluation metrics is desired for future clinical trials and studies.

XSL•FO
RenderX

Trial Registration: Open Science Forum Registries YH5S7; https://doi.org/10.17605/OSF.IO/YH5S7

(J Med Internet Res 2025;27:e67929) doi:10.2196/67929

KEYWORDS

patient experience; patient-reported outcome; quality of life; quality-adjusted life year; telehealth; eHealth; mobile phone; metrics; telemedicine; systematic review; scoping review; review; telecommunications; database; health care; patient-centeredness; patient satisfaction; patient outcome; clinical parameter; cost-effectiveness; evaluation metrics; mHealth; mobile health

Introduction

Background

Telemedicine involves any type of medical, diagnostic, or treatment-related service using telecommunication technology [1]. In Japan, the term "online medical care," defined more specifically by the government, is used instead of telemedicine, which refers to the act of examining and diagnosing a patient, communicating the results of the diagnosis, and prescribing medical treatment in real time involving a doctor and a patient through telecommunication technology [2]. Recent developments in telecommunications technology have been remarkable. The telemedicine environment is changing with the spread of smartphones, improvements in data transmission speeds, and the development of wearable devices. Furthermore, although the use of telemedicine had been attempted earlier, it was adopted widely during the COVID-19 pandemic because of social distancing [3]. These findings indicate that the telemedicine content and environment have changed over the past few years.

The reports on the implementation of telemedicine in various medical fields are limited. Compared with standard face-to-face medical care, telemedicine reduces the risk of infection by minimizing physical contact and offers advantages to patients, such as shorter visit times, including waiting and travel times, which can improve patient satisfaction rates [4]. Ease of access to medical care provided by telemedicine may also improve disease management and clinical outcomes [5]. Telemedicine can be effective in many aspects of health care [6,7]. Therefore, the metrics that should be used to evaluate telemedicine should be carefully considered.

For evaluating disease management, using disease-specific indicators, such as hemoglobin A_{1c} (HbA_{1c}) for diabetes, is necessary to determine the usefulness of telemedicine. Additionally, general indicators, including non–disease-specific indicators, are required to compare results among studies on different diseases. Telemedicine can also influence patient satisfaction and quality of life (QoL). Furthermore, the use of patient-reported outcomes (PROs) was recently proposed for the evaluation of "patient-centeredness," the concept in which medical care is provided "for the benefit of the patient" [8].

Telemedicine affects various aspects of medical care; thus, it requires multiple measures rather than a single measure for efficacy evaluations. It is possible that indicators that have not been widely used in the past should be used in the future. A variety of metrics have been used across previous studies, and reports summarizing the general metrics used for evaluating telemedicine remain lacking.

Telemedicine does not oppose face-to-face care but rather complements it, similar to the relationship between outpatient, inpatient, and home care. However, when examining the usefulness of telemedicine, it is commonly compared with face-to-face care, which is the well-established, long-standing method of medical care. Further research into telemedicine, particularly randomized controlled trials (RCTs), is essential to generate solid evidence. To identify the metrics that should be used in future RCTs, it is essential to first understand the actual use of general metrics in past RCTs.

Objectives

In this study, we conducted a scoping review of articles comparing telemedicine with in-person care in RCTs and summarized the metrics used to understand their usefulness. This study also aimed to identify indicators that were not thoroughly evaluated in previous clinical trials but will be useful in the future and suggest which indicators should be prioritized in future clinical trials.

Methods

Literature Search

This systematic literature review was conducted and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis Extension for Scoping Review) guidelines [9]. The completed PRISMA checklist is presented in Multimedia Appendix 1. MEDLINE and Embase databases were systematically searched on March 15, 2024, for peer-reviewed full papers published between January 1, 2019, and March 14, 2024. We searched only the MEDLINE and Embase databases because we assumed that most of the articles of interest (metrics used in RCTs) would be included in them. Owing to the significant advancements in telemedicine in recent years, older publications were deemed outdated, and the search was limited to the most recent publications from the past 5 years since 2019. The search string used for the search in MEDLINE contained the following terms related to telemedicine: (telemedicine OR "online medical care" OR teleconsultation OR "online consultation" OR "telemedical consultation") AND ("randomized controlled trial") AND (outcome OR effectiveness) AND (control OR conventional OR face-to-face). Multimedia Appendix 2 lists the search terms used in the MEDLINE and Embase databases.

The review was registered with Open Science Forum Registries (YH5S7).

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) studies published between January 1, 2019, and March 14, 2024; (2) RCTs

XSL•FO
RenderX

comparing some form of telemedicine with a standard of care (face-to-face care); (3) studies involving telemedicine, including online monitoring (ie, information measured using the device is automatically sent to the medical facility) or synchronous communication through telephone or video; and (4) studies involving both disease-specific metrics (eg, HbA_{1c} for diabetes) and general (ie, non–disease-specific) cross-disciplinary metrics.

The following studies were excluded: (1) protocols and proposals; (2) systematic reviews, commentaries, and scoping reviews; and (3) gray literature.

Study Selection

Two investigators (YS and YH) independently screened the titles and abstracts of the identified studies using the Covidence systematic review software (Veritas Health Innovation). Subsequently, the investigators (YS and YH) independently reviewed the full text of the selected studies based on the inclusion and exclusion criteria. Disagreements regarding the classification of a reference were resolved by a third reviewer (MI) by conducting an additional review and confirming the final classification.

Data Collection and Summary

Data regarding study details were extracted using the Covidence software (Veritas Health Innovation) and Microsoft Excel. One reviewer (YS) extracted the additional data from eligible studies, including the content of telemedicine and the disease-specific and general (ie, non–disease-specific) indicators for evaluating telemedicine. All extracted data were cross-checked for accuracy by a second reviewer (YH), and any discrepancies were resolved through discussion. The extracted data elements included the study title, year of publication, country, target disease or population, total sample size, type of telemedicine used, and disease-specific and general metrics used for evaluating telemedicine. A meta-analysis was deemed inappropriate because of heterogeneity in the study design, populations, and outcome measures for quantitative studies, and a narrative synthesis of quantitative study results was conducted. The evaluation metrics were categorized through the following process. Initially, all unique metrics identified by a single author (YS) were reviewed and confirmed by two additional authors (YH and MI). Following this, the 3 authors engaged in collaborative discussions to group the metrics into primary and subcategories based on shared directional perspectives.

This study examined the actual use of metrics for evaluating telemedicine and did not focus on the results or conclusions (superiority or noninferiority of telemedicine compared with face-to-face care) of the included studies.

Results

Literature Search

Initially, 2275 studies were considered relevant according to the inclusion criteria. However, 606 duplicate studies or those marked as ineligible by automation tools were removed before screening. Furthermore, 1507 studies were excluded after screening the titles and abstracts, and 83 studies were excluded after reading the articles in greater depth during the assessment of extracted data. Finally, 79 studies were included in the analysis. The PRISMA flow diagram summarizing the article selection process is presented in Figure 1. A list of all 79 studies and their information is provided in Multimedia Appendix 3 [7,10-87].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) 2020 flow diagram for scoping review.

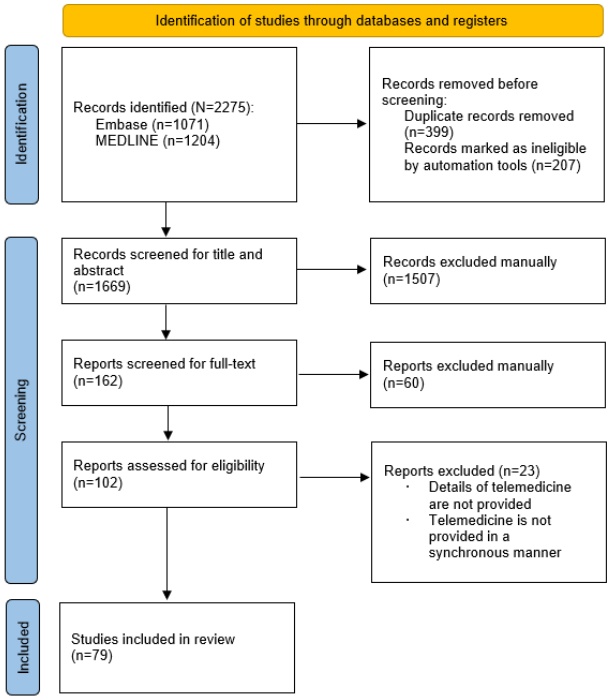

XSL•FO
RenderX

Study Description

Table 1 presents the characteristics of the 79 included studies. After 2020, a total of 10 - 20 papers comparing telemedicine and face-to-face care in RCTs were consistently published annually. These studies were conducted and published commonly in the United States (11/79, 14%), Australia (10/79, 13%), and European countries (France: 6/79, 8%; Norway: 6/79, 8%; and: Spain 6/79, 8%), with China (5/79, 6%) and Japan (5/79, 6%) being the most common in Asia.

The most common departments in which the studies were conducted were internal medicine (39/79, 49%), psychiatry (11/79, 14%), and rehabilitation (9/79, 11%). Most dealt with outpatient care (76/79, 96%); however, a few reports dealt with telemedicine from home as an alternative to inpatient care (3/79, 4%). The specific contents were as follows: early discharge from the hospital with telemedicine versus discharge on a regular schedule (study number 57 [65]), home care under telemonitoring versus long-term hospitalization (study number 71 [79]), and initiation of treatment at home versus in the hospital (study number 74 [82]).

Regarding the targeted diseases and populations, the most common was postoperative follow-up (11/79, 14%), followed by chronic diseases, such as coronary artery disease (7/79, 9%), diabetes (6/79, 8%), and chronic obstructive pulmonary disease (4/79, 5%). The median total sample size was 106 (IQR 59 - 240), with 10 and 2185 being the minimum and maximum numbers, respectively.

When examining the details of telemedicine, online medical care was the most common (69/79, 87%; Table 1), including telepsychiatry involving cognitive behavior therapy. Furthermore, online monitoring and telerehabilitation were also common practices. In several studies, telemedicine included apps for patient education and the ability to make emergency calls. Online monitoring (information measured by the device automatically sent to the medical facility) or synchronous communication through telephone or video was considered telemedicine in this review.

XSL•FO
RenderX

Table 1. Characteristics of the included studies (n=79).

Characteristics	Value, n (%)
Publication year	
2019	9 (11)
2020	15 (19)
2021	18 (23)
2022	18 (23)
2023	12 (15)
2024 (From January to March)	7 (9)
Publication country[a]	
United States	11 (14)
Australia	10 (13)
France	6 (8)
Norway	6 (8)
Spain	6 (8)
China	5 (6)
Germany	5 (6)
Japan	5 (6)
The Netherlands	5 (6)
United Kingdom	3 (4)
Department[a]	
Internal medicine	39 (49)
Psychiatry	11 (14)
Rehabilitation	9 (11)
Surgery	5 (6)
Orthopedics	4 (5)
Pediatrics	3 (4)
Anesthesiology	2 (3)
Gynecology	2 (3)
Urology	2 (3)
Setting	
Outpatient care	76 (96)
Inpatient care	3 (4)
Target disease or population[a]	
Postoperative	11 (14)
Coronary disease	7 (9)
Diabetes mellitus	6 (8)
Sleep apnea	5 (6)
Chronic obstructive pulmonary disease	4 (5)
Mental disorders	4 (5)
Rheumatoid arthritis	3 (4)
Brain tumor	2 (3)
Chronic conditions	2 (3)

XSL•FO
RenderX

Characteristics	Value, n (%)
Chronic pain	2 (3)
Heart failure	2 (3)
Inflammatory bowel disease	2 (3)
Insomnia	2 (3)
Orthopedic consultation	2 (3)
Patients with pacemakers	2 (3)
Parkinson disease	2 (3)
Primary care for children	2 (3)
Detailed contents of telemedicine	
Online medical care[b]	69 (87)
Telepsychiatry including cognitive-behavior therapy	10 (13)
Online monitoring	17 (22)
Telerehabilitation	11 (14)

[a]Two or more reports.

[b]According to the following definition by the Japanese government: the act of examining and diagnosing a patient, communicating the results of the diagnosis, and prescribing medical treatment in real time involving a doctor and a patient through telecommunication technology

General Metrics for Evaluating Telemedicine

We identified 25 unique metrics that can be generally used (patient satisfaction, patient workload, patient time spent in medical visits, absence from work, travel distance, waiting time, QoL, patient choice of subsequent type of care, death, admission, hospital days, emergency department visits, frequency of visits or contacts, duration of visits or contacts, treatment adherence, change in diagnosis or treatment, adverse effects or safety, cost, quality-adjusted life years [QALY], staff satisfaction, staff work, staff time spent per patient, system usability, environmental impact, and feasibility for future clinical trials). A median of 4 (IQR 3-5, with a minimum of 1 and a maximum of 10) metrics were used per study. We classified commonly used metrics into three categories: (1) patient-centeredness (67/79, 85%), (2) patient outcomes (57/79, 72%), and (3) cost-effectiveness (40/79, 51%) (Table 2). Notably, 32% (25/79) evaluated all these categories.

We further subdivided the metrics into nine subcategories and examined the numbers and combinations of uses per article. Patient centeredness was divided into three categories: (1) patient satisfaction, (2) patient work, and (3) QoL. Patient outcomes were classified as (4) general clinical parameters. Cost-effectiveness was divided into two categories: (5) costs, and (6) QALY. Other metrics were categorized into three categories: (7) staff satisfaction, (8) staff work, and (9) feasibility for future clinical trial. Most studies (69/79, 87%) used two or more general metrics (ie, non–disease-specific metrics), and the median number of the nine subcategories used in each study was 3 (IQR 2-4, with a minimum of 1 and a maximum of 6) metrics. The most frequently used combinations in the nine subcategories were as follows: general clinical parameters and patient satisfaction (7/79, 9%); general clinical parameters (5/79, 6%); QoL and patient satisfaction (5/79, 6%); QoL and general clinical parameters (5/79, 6%); costs, QoL, and general clinical parameters (5/79, 6%); and costs, QALY, and QoL (5/79, 6%)

Table 2. General metrics for evaluating telemedicine (n=79).

Metric	Value, n (%)
Patient-centeredness	67 (85)
Patient outcomes	57 (72)
Cost-effectiveness	40 (51)
Other	40 (51)

The evaluation metrics by department are summarized in Table 3. With regard to internal medicine, psychiatry, and rehabilitation, for which a relatively large number of articles were included in this review, the utilization rate of patient outcomes was high in internal medicine (33/39, 85%) but was rather low in psychiatry (5/11, 45%) and rehabilitation (5/9, 56%). In contrast, the utilization rate of patient-centeredness was very high in psychiatry (10/11, 91%) and rehabilitation (8/9, 89%) compared with that in internal medicine (31/39, 79%).

XSL•FO
RenderX

Table 3. Utilization rate of general metrics by department.

Department.	Patient-centeredness, n (%)	Patient outcomes, n (%)	Cost-effectiveness, n (%)
Internal medicine (n=39)	31 (79)	33 (85)	21 (54)
Psychiatry (n=11)	10 (91)	5 (45)	7 (64)
Rehabilitation (n=9)	8 (89)	5 (56)	3 (33)
Surgery (n=5)	5 (100)	5 (100)	2 (40)
Orthopedics (n=4)	4 (100)	2 (50)	3 (75)
Pediatrics (n=3)	2 (67)	1 (33)	2 (67)
Anesthesiology (n=2)	2 (100)	1 (50)	1 (50)
Gynecology (n=2)	1 (50)	1 (50)	0 (0)
Urology (n=2)	2 (100)	2 (100)	1 (50)

Patient-Centeredness

Patient satisfaction, work associated with medical visits, and QoL were considered metrics relevant to patient-centeredness, and these metrics were evaluated in most studies (67/79, 85%). The metrics used for evaluating patient-centeredness are summarized in Table 4.

Table 4. General metrics for evaluating patient-centeredness (n=67).

Metric	Value, n (%)
Patient satisfaction	41 (61)
Patient workload	17 (25)
Time spent in visit or contact	9 (13)
Absence from work	7 (10)
Travel distance	7 (10)
Waiting time	3 (4)
Quality of life	45 (67)

More than half (41/67, 61%) of the studies evaluated patient satisfaction. Table 5 presents the questionnaires used to evaluate patient satisfaction. Approximately half of the studies that addressed patient satisfaction used their own set of questions, whereas the other half used previously developed questionnaires, such as the Client Satisfaction Questionnaire [88], Telemedicine Satisfaction Questionnaire [89], and Outpatient Experiences Questionnaire [90]. In addition, 2 other studies evaluated patient experience (study numbers 48 [56] and 68 [76] in Multimedia Appendix 3), which refers to the patient's specific "experience" of health care services. This concept is an evolution of patient satisfaction and is internationally recognized as an important quality indicator in health care [91]. Only one study evaluated "PROs."

XSL•FO
RenderX

Table 5. Questionnaires used to assess patient satisfaction and quality of life (QoL). Regarding questionnaires for QoL, two reports used both the EQ-5D and SF-36[a] QoL measures, and other two used both the EQ-5D and SF-12[b]. Others included general and disease-specific QoL measures other than EQ-5D, SF-36, and SF-12.

Questionnaires	Value, n (%)
Patient satisfaction (n=41)	
Questions developed for the study	19 (46)
Client Satisfaction Questionnaire [88]	3 (7)
Telemedicine Satisfaction Questionnaire [89]	3 (7)
Outpatient Experiences Questionnaire [90]	2 (5)
Satisfad10 Questionnaire [92]	2 (5)
Other questionnaires from previous studies	12 (29)
Quality-of-life (n=45)	
EQ-5D-3L/5L [93,94]	26 (58)
SF-36[a] [95]	9 (20)
SF-12[b] [96]	4 (9)
Others	10 (22)

[a]SF-36: Medical Outcomes Study (MOS) Short Form 36-Item Health Survey.

[b]SF-12: MOS Short Form 12-Item Health Survey.

Some studies (17/67, 25%) focused on work or burden associated with medical visits, including factors such as the time spent or travel distance for medical visits or contacts. Other related indicators, such as absence from work and waiting time, were also investigated in a few studies.

More than half of the studies evaluated QoL (45/67, 67%). Table 5 summarizes the metrics used to measure QoL. The most frequently used measure was the EQ-5D-3L/5L [93,94], which was used by more than half of the studies that evaluated QoL. The Medical Outcomes Study Short Form 36-Item Health Survey (SF-36) [95] and Medical Outcomes Study Short Form 12-Item Health Survey (SF-12) [96] were also frequently used. In addition to these general QoL measures, some studies also used disease-specific QoL measures.

Patient Outcomes

When comparing telemedicine and face-to-face care in RCTs, various clinical metrics for patient outcomes, including disease-specific ones, are used to assess their superiority or noninferiority. Table 6 summarizes general metrics that appear to be applicable for evaluating various disease populations. The most commonly used general clinical parameters were admission (26/57, 46%), adherence (23/57, 40%), adverse events or safety (21/57, 37%), and frequency of visits or contacts (17/57, 30%).

Table 6. General metrics for evaluating patient outcomes (n=57).

Metrics	Value, n (%)
Death	10 (18)
Admission	26 (46)
Hospital days	8 (14)
Emergency department visits	10 (18)
Frequency of visits or contacts	17 (30)
Duration of visit or contact	8 (14)
Treatment adherence	23 (40)
Change in diagnosis or treatment	4 (7)
Adverse events or safety	21 (37)

Cost-Effectiveness

The following ranges of costs were considered when evaluating cost-effectiveness: medical costs (fees for consultation, examination, medication, injection, surgery, and hospitalization), nonmedical costs (costs indirectly incurred by medical interventions, such as transportation, welfare equipment, or home improvements), and indirect costs (loss due to injury, illness, or death). The method of calculating medical costs and the range of costs included in the calculations differed across studies, partly because of differences in the reimbursement systems of the countries in which the studies were conducted

XSL•FO
RenderX

and differences in perspectives (patient, insurer, or society). QALY was evaluated in 28% (11/40) of the studies that evaluated cost-effectiveness.

Staff Convenience

Staff convenience including staff satisfaction, staff workload, and time spent per patient was evaluated in a few studies (8/79, 10%). Staff satisfaction was not frequently evaluated (4/79, 5%) compared with patient satisfaction (41/79, 52%). Similarly, fewer studies examined staff workload (5/79, 6%) than patient workload (17/79, 22%). Notably, the time spent per patient was evaluated to assess the staff work in the studies.

System Usability

A small number of studies (study numbers 8 [18], 14 [24], and 49 [57] in Multimedia Appendix 3) have evaluated the user-friendliness of telemedicine systems using items such as usability questionnaires or the number of technical challenges encountered.

Others

Other metrics that were not included in either of the aforementioned categories were greenhouse gas impacts (study numbers 6 [16] and 18 [27] in Multimedia Appendix 3), mental health of caregivers (study number 12 [22]), health literacy (study number 79 [87]), and self-efficacy (study number 17 [26]). Details of each are provided in Multimedia Appendix 3.

Discussion

Principal Findings

In this study, we conducted a scoping review of RCTs comparing telemedicine with face-to-face care published in the last 5 years and summarized the cross-disciplinary measures used in these studies. Notably, several studies used multiple measures, and metrics related to patient-centeredness, patient outcomes, and cost-effectiveness were commonly used. However, the measures used and their combinations varied across studies, and only a few studies evaluated staff satisfaction, system usability, and environmental impacts. These results highlight the requirement to standardize evaluation indicators for comparing telemedicine with face-to-face care. Although establishing evaluation measures for telemedicine is challenging owing to its multifaceted impact, the evaluation indicators can be classified into three main categories: (1) patient-centeredness, (2) patient outcomes, and (3) cost-effectiveness.

We believe that in addition to using disease-specific indicators, these three categories of general metrics should be used in telemedicine evaluations. From the papers analyzed in this study (Multimedia Appendix 3), some pertinent examples from the identified studies are as follows: Gayot et al [10] evaluated QoL for patient-centeredness, unplanned hospitalization and death for patient outcomes, and direct costs and incremental cost-effectiveness ratios for cost-effectiveness in addition to disease-specific metrics, and Mínguez Clemente et al [11] evaluated patient satisfaction for patient-centeredness; admission, hospital days, frequency of visits, and adherence for patient outcomes; and cost per patient for cost-effectiveness while using other parameters (study numbers 16 [10] and 73 [11] in Multimedia Appendix 3, respectively). Our results suggest that within a single study, at least three indicators, each belonging to a different category, should be simultaneously evaluated to ensure a comprehensive assessment of the impact of telemedicine. However, given that different diseases have different management approaches, our results indicate that the choice of metrics differs by clinical specialty (Table 3). Therefore, they need to be considered more deeply and individually in special circumstances. In particular, studies of replacing inpatient care with telemedicine were not well represented in this review and may require separate consideration.

In this study, we summarized the evaluation metrics commonly used in previous reports, identified 3 categories that should be covered, and described other 3 indicators that should be evaluated in the future. However, it did not provide a more rigorous standardization of which indicators should be used for each of the three categories. We believe that this should be discussed among the expert panel and finally decided using approaches such as the Delphi method [97].

Patient-Centeredness

Patient-centeredness is a concept closely related to PROs [8]. Although measures of patient prognosis are relatively objective, PROs are subjective to the patient, with no intervention from a physician or measurement using instruments. However, considering such subjective evaluations as evidence is debatable [98]. Patient-centeredness is a crucial concept based on the premise that medical care is performed “for the patient,” while also emphasizing the need for standardization and clear validation. In this review, we adopted a broad perspective of patient-centeredness, considering not only studies using the term “PROs” but also the following terms as relevant: patient satisfaction, patient work associated with medical visits, and QoL.

Patient satisfaction, experience, and PROs were the main indicators used to assess patient-centeredness. Patient satisfaction is an abstract concept of “satisfaction” that also depends on individual expectations, making the standardization of measurement scales challenging and complicating the identification of specific issues from the results [99,100]. Although several questionnaires have been developed to assess patient satisfaction [88-90], their use varies across studies, and in approximately half of the studies, questions used to evaluate satisfaction were originally designed in those studies. In contrast, patient experience deals with specific “experiences” and is easier to standardize than patient satisfaction, with more convenience in identifying specific issues from the results [99,101]. PROs are even broader measures that reflect various aspects of patient health, as they are reported directly by the patient without additional interpretation by a health care professional or anyone else [8]. They may be divided into disease-specific and comprehensive PROs. Patient satisfaction, patient experience, and PROs are interrelated and cannot be studied separately. Assessing patient experience or PROs may be preferable over patient satisfaction to help identify issues.

Among the various QoL scales, EQ-5D-3L/L [93,94], SF-36 [95], and SF-12 [96] are used most commonly. Notably, even

XSL•FO
RenderX

while using the disease-specific QoL assessment scales, the aforementioned general scales may be used in combination.

Patient Outcomes

The indicators of patient outcomes can be divided into disease-specific and cross-disciplinary outcomes. In this study, disease-specific indicators are listed only in Multimedia Appendix 3 and are not tabulated as they vary across fields. Moreover, even if disease-specific indicators (eg, HbA_{1c} for diabetes) are evaluated, they may be difficult to understand for readers in other fields. Therefore, cross-disciplinary general indicators should be used in addition to disease-specific indicators. The specific parameters that should be used as general indicators depend on the severity and progression pattern of the disease but can include death, hospitalization, emergency visits, and frequency of general outpatient visits.

Cost-Effectiveness

The most common cost-effectiveness evaluation was the calculation and comparison of the costs of telemedicine and usual face-to-face care. From the health care payer's perspective, only medical costs are important, whereas from the patient's perspective, only the copayment of medical costs is important. Furthermore, from a societal perspective, all medical, nonmedical, and indirect costs should be included in the analysis [102-104]. When comparing telemedicine to traditional face-to-face care, a broader spectrum of parameters must be evaluated. These encompass the costs associated with software development and hardware implementation, as well as the potential economic advantages such as reduced transportation expenses and fewer missed workdays, which are not typically considered in comparisons of in-person treatment modalities. Therefore, evaluating not only medical costs but also nonmedical and indirect costs is crucial. However, standardizing the calculation method is difficult because costs may differ depending on the reimbursement system, prices, and social environment in each country.

The QALY is a cost-effectiveness index based on QoL and survival time and is considered a common effectiveness index, even if the target diseases are different [105]. However, although various factors are associated with evaluating the cost-effectiveness of telemedicine, QALY evaluates only QoL and survival time. Therefore, this should be considered when using QALY.

Other Metrics

In addition to the metrics mentioned earlier, future studies should also consider other metrics for evaluation. The first candidate is the metric for staff convenience. We all know that health care is "for the patient"; as noted earlier, patient-centeredness has been well investigated in many studies. On the other side of the coin, its convenience for health care professionals has not been considered important or thoroughly evaluated. However, it is an undeniable fact that the health care providers (medical professionals) are the counterparts of the health care recipients (patients), and it is easy to imagine that a system that health care providers find inconvenient is unlikely to be widely adopted. Particularly in telemedicine, health care professionals are accustomed to a well-established and easy-to-use form of care, which is face-to-face care. Therefore, evaluating whether telemedicine is easy to use for health care professionals is vital. Such information would aid in assessing how favorable telemedicine is likely to be received by health care providers when implemented in general clinical practice. Careful examination of the reasons why staff members felt telemedicine was inconvenient would also provide insight into how to improve it.

Another important factor is whether the system is easy to use, as the most important component of telemedicine is the digital equipment and internet connection. The degree of digital literacy varies among individuals, and if the system is difficult to use, telemedicine may even worsen access to health care for some people, as symbolized by the term "digital divide." Although it has not been well evaluated so far, the evaluation metrics of system usability are also considered important for ensuring equitable access to health care.

In addition, a small number of studies evaluated the impact of greenhouse gases and environmental costs. Given the seriousness of global warming, considering a health care delivery system that is sustainable for the entire planet is also important. The Greenhouse Gas Protocol divides carbon emissions into 3 categories: scope 1, direct emissions secondary to energy use; scope 2, indirect emissions secondary to purchased electricity; and scope 3, indirect emissions [106]. Most emissions related to the health care system reportedly correspond to scope 3, which includes indirect emissions occurring as a consequence of health care activities such as disposables, equipment (medical and nonmedical), and pharmaceuticals, as it includes a wider range of emission sources than other scopes [107]. Therefore, it may be necessary to consider a wide range of targets to assess the environmental impacts of telemedicine. There is a scoping review on the inclusion of environmental impacts in evaluations of health care economics [108]. Greenhouse gas emissions, energy use, water use, and physical waste impacts were frequently used in the past, and 10 methods for assessing environmental impacts have been described. However, it remains unknown which metrics should be evaluated and in which units, creating an issue for discussion in the future.

Limitations

Although our scoping review was thorough, it has some limitations. First, we searched only the MEDLINE and Embase databases for peer-reviewed, full-text articles, although we assumed that most of the articles in our scope would be included within them because we focused on the metrics used in RCTs. Second, only papers published in 2019-2024 were included because the content of telemedicine has changed rapidly with technological advancements and the emergence of COVID-19, and it is unclear whether it is appropriate to apply the content of the distant past to current analyses. This limited the scope of the search and emphasized the most recent data. Third, a narrative synthesis is a secondary analysis of data that focuses on the interpretations presented by the authors of original papers and is not based on primary data. Furthermore, our findings represent an interpretation of the data and should be viewed as a heuristic theory.

XSL•FO
RenderX

Conclusions

In studies comparing telemedicine with face-to-face care from 2019 to 2024, the most commonly used metrics were patient-centeredness, patient outcomes, and cost-effectiveness. These are important indicators of telemedicine and should be evaluated simultaneously in future clinical studies. In addition, our study also indicates that staff satisfaction could be an important evaluation metric for future clinical trials.

Acknowledgments

This study was supported by MHLW 23IA2001. We would also like to thank Editage for the English language editing.

Data Availability

All data analyzed during this study are included in the article and its supplementary files.

Authors' Contributions

YS, YH, and MN contributed to study design and conception. YS, YH, RI, MI, and RW conducted the investigation. YS, YH, and MI were involved in writing (original draft). RI, RW, and MN contributed to writing (review and editing). MN handled the funding acquisition and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 85 KB - jmir_v27i1e67929_app1.docx]

Multimedia Appendix 2
Search terms used for the MEDLINE and Embase databases.
[PDF File (Adobe PDF File), 98 KB - jmir_v27i1e67929_app2.pdf]

Multimedia Appendix 3
List of all the 79 papers and their information.
[XLSX File (Microsoft Excel File), 128 KB - jmir_v27i1e67929_app3.xlsx]

References

1. Telehealth, telemedicine, and telecare: what's what? Federal Communications Commission. URL: https://www.fcc.gov/general/telehealth-telemedicine-and-telecare-whats-what [accessed 2024-10-01]
2. Guidelines for the appropriate implementation of online medical care. Ministry of Health, Labour and Welfare. 2022. URL: https://www.mhlw.go.jp/content/12601000/000901835.pdf [accessed 2024-10-01]
3. OECD. The COVID-19 Pandemic and the future of telemedicine, OECD health policy studies. Paris: OECD Publishing; 2023. URL: https://www.oecd.org/en/publications/the-covid-19-pandemic-and-the-future-of-telemedicine_ac8b0a27-en.html [accessed 2024-10-01]
4. Anawade PA, Sharma D, Gahane S. A comprehensive review on exploring the impact of telemedicine on healthcare accessibility. Cureus 2024;16(3):e55996 [FREE Full text] [doi: 10.7759/cureus.55996] [Medline: 38618307]
5. Satoh M, Tatsumi Y, Nakayama S, Shinohara Y, Kawazoe M, Nozato Y, et al. Self-measurement of blood pressure at home using a cuff device for change in blood pressure levels: systematic review and meta-analysis. Hypertens Res 2024. [doi: 10.1038/s41440-024-01981-4] [Medline: 39572787]
6. Cardoso LB, Couto P, Correia P, Lopes PC, Fernandes JCH, Fernandes GVO, et al. Impact of digital innovations on health literacy applied to patients with special needs: a systematic review. Information 2024;15(11):663 [FREE Full text] [doi: 10.3390/info15110663]
7. Kishimoto T, Kinoshita S, Kitazawa M, Hishimoto A, Asami T, Suda A, et al. Live two-way video versus face-to-face treatment for depression, anxiety, and obsessive-compulsive disorder: a 24-week randomized controlled trial. Psychiatry Clin Neurosci 2024;78(4):220-228. [doi: 10.1111/pcn.13618] [Medline: 38102849]
8. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013;346:f167 [FREE Full text] [doi: 10.1136/bmj.f167] [Medline: 23358487]
9. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

XSL•FO
RenderX

10. Gayot C, Laubarie-Mouret C, Zarca K, Mimouni M, Cardinaud N, Luce S, et al. Effectiveness and cost-effectiveness of a telemedicine programme for preventing unplanned hospitalisations of older adults living in nursing homes: the GERONTACCESS cluster randomized clinical trial. BMC Geriatr 2022;22(1):991 [FREE Full text] [doi: 10.1186/s12877-022-03575-6] [Medline: 36550496]
11. Mínguez Clemente P, Pascual-Carrasco M, Mata Hernández C, Malo de Molina R, Arvelo LA, Cadavid B, et al. Follow-up with telemedicine in early discharge for COPD exacerbations: randomized clinical trial (TELEMEDCOPD-Trial). COPD 2021;18(1):62-69 [FREE Full text] [doi: 10.1080/15412555.2020.1857717] [Medline: 33307857]
12. Wang QP, Chang WY, Han MM, Hu YX, Lin SS, Gu YC. Application of telemedicine system for older adults postoperative patients in community: a feasibility study. Front Public Health 2024;12:1291916 [FREE Full text] [doi: 10.3389/fpubh.2024.1291916] [Medline: 38435285]
13. Sim HW, Koh KWL, Poh SC, Chan SP, Marchesseau S, Singh D, et al. Remote intensive management to improve antiplatelet adherence in acute myocardial infarction: a secondary analysis of the randomized controlled IMMACULATE trial. J Thromb Thrombolysis 2024;57(3):408-417. [doi: 10.1007/s11239-023-02931-6] [Medline: 38300500]
14. Sten-Gahmberg S, Pedersen K, Harsheim IG, Løyland HI, Snilsberg Ø, Iversen T, et al. Pragmatic randomized controlled trial comparing a complex telemedicine-based intervention with usual care in patients with chronic conditions. Eur J Health Econ 2024;25(7):1275-1289. [doi: 10.1007/s10198-023-01664-w] [Medline: 38291176]
15. Gordon LG, Jones S, Parker G, Chambers S, Aitken JF, Foote M, et al. Cost-utility analysis of a telehealth psychological support intervention for people with primary brain tumor: telehealth making sense of brain tumor. Psychooncology 2024;33(1):e6243. [doi: 10.1002/pon.6243] [Medline: 37946565]
16. Morau E, Chevallier T, Serrand C, Perin M, Gricourt Y, Cuvillon P. Teleconsultation compared with face-to-face consultation in the context of pre-anesthesia evaluation: TELANESTH, a randomized controlled single-blind non-inferiority study. J Clin Anesth 2024;92:111318. [doi: 10.1016/j.jclinane.2023.111318] [Medline: 37944402]
17. Wright AA, Poort H, Tavormina A, Schmiege SJ, Matulonis UA, Campos SM, et al. Pilot randomized trial of an acceptance-based telehealth intervention for women with ovarian cancer and PARP inhibitor-related fatigue. Gynecol Oncol 2023;177:165-172. [doi: 10.1016/j.ygyno.2023.08.020] [Medline: 37708581]
18. Tümtürk İ, Bakırhan S, Özden F, Gültaç E, Kılınç CY. Effect of telerehabilitation-based exercise and education on pain, function, strength, proprioception, and psychosocial parameters in patients with knee osteoarthritis: a randomized controlled clinical trial. Am J Phys Med Rehabil 2024;103(3):222-232. [doi: 10.1097/PHM.0000000000002335] [Medline: 37678215]
19. Pei G, Ou Q, Lao M, Wang L, Xu Y, Tan J, et al. APAP treatment acceptance rate and cost-effectiveness of telemedicine in patients with obstructive sleep apnea: a randomized controlled trial. Nat Sci Sleep 2023;15:607-622 [FREE Full text] [doi: 10.2147/NSS.S416221] [Medline: 37560381]
20. Fridriksson B, Berndtson M, Hamnered H, Faeder E, Zou D, Hedner J, et al. Beneficial effects of early intervention telemedicine-based follow-up in sleep apnea: a randomized controlled multicenter trial. Ann Am Thorac Soc 2023;20(10):1499-1507. [doi: 10.1513/AnnalsATS.202208-723OC] [Medline: 37463309]
21. Ballesta S, Chillarón JJ, Inglada Y, Climent E, Llauradó G, Pedro-Botet J, et al. Telehealth model versus in-person standard care for persons with type 1 diabetes treated with multiple daily injections: an open-label randomized controlled trial. Front Endocrinol (Lausanne) 2023;14:1176765 [FREE Full text] [doi: 10.3389/fendo.2023.1176765] [Medline: 37441496]
22. Ownsworth T, Chambers S, Jones S, Parker G, Aitken JF, Foote M, et al. Evaluation of the telehealth making sense of brain tumor psychological support intervention for people with primary brain tumor and their caregivers: a randomized controlled trial. Psychooncology 2023;32(9):1385-1394 [FREE Full text] [doi: 10.1002/pon.6189] [Medline: 37409906]
23. So H, Chow E, Cheng IT, Lau SL, Li TK, Szeto CC, et al. Telemedicine for follow-up of systemic lupus erythematosus during the 2019 coronavirus pandemic: a pragmatic randomized controlled trial. J Telemed Telecare 2023:1357633X231181714 [FREE Full text] [doi: 10.1177/1357633X231181714] [Medline: 37357745]
24. Lundgren KM, Langlo KAR, Salvesen Ø, Zanaboni P, Cittanti E, Mo R, et al. Feasibility of telerehabilitation for heart failure patients inaccessible for outpatient rehabilitation. ESC Heart Fail 2023;10(4):2406-2417 [FREE Full text] [doi: 10.1002/ehf2.14405] [Medline: 37221704]
25. Babar M, Zhu D, Loloi J, Laudano M, Ohmann E, Abraham N, et al. Comparison of patient satisfaction and safety outcomes for postoperative telemedicine vs face-to-face visits in Urology: results of the Randomized Evaluation and Metrics Observing Telemedicine Efficacy (REMOTE) trial. Urol Pract 2022;9(5):371-378. [doi: 10.1097/UPJ.0000000000000323] [Medline: 37145727]
26. Zanaboni P, Dinesen B, Hoaas H, Wootton R, Burge AT, Philp R, et al. Long-term telerehabilitation or unsupervised training at home for patients with chronic obstructive pulmonary disease: a randomized controlled trial. Am J Respir Crit Care Med 2023;207(7):865-875 [FREE Full text] [doi: 10.1164/rccm.202204-0643OC] [Medline: 36480957]
27. Muschol J, Heinrich M, Heiss C, Hernandez AM, Knapp G, Repp H, et al. Economic and environmental impact of digital health app video consultations in follow-up care for patients in orthopedic and trauma surgery in Germany: randomized controlled trial. J Med Internet Res 2022;24(11):e42839 [FREE Full text] [doi: 10.2196/42839] [Medline: 36333935]
28. Fink T, Chen Q, Chong L, Hii MW, Knowles B. Telemedicine versus face-to-face follow up in general surgery: a randomized controlled trial. ANZ J Surg 2022;92(10):2544-2550 [FREE Full text] [doi: 10.1111/ans.18028] [Medline: 36069322]

XSL•FO
RenderX

29. DE Lima AP, Pereira DG, Nascimento IO, Martins TH, Oliveira AC, Nogueira TS, et al. Cardiac telerehabilitation in a middle-income country: analysis of adherence, effectiveness and cost through a randomized clinical trial. Eur J Phys Rehabil Med 2022;58(4):598-605 [FREE Full text] [doi: 10.23736/S1973-9087.22.07340-3] [Medline: 35634888]
30. Treskes RW, van den Akker-van Marle ME, van Winden L, van Keulen N, van der Velde ET, Beeres S, et al. The box-eHealth in the outpatient clinic follow-up of patients with acute myocardial infarction: cost-utility analysis. J Med Internet Res 2022;24(4):e30236 [FREE Full text] [doi: 10.2196/30236] [Medline: 35468091]
31. Irgens I, Midelfart-Hoff J, Jelnes R, Alexander M, Stanghelle JK, Thoresen M, et al. Videoconferencing in pressure injury: randomized controlled telemedicine trial in patients with spinal cord injury. JMIR Form Res 2022;6(4):e27692 [FREE Full text] [doi: 10.2196/27692] [Medline: 35438645]
32. Bernard L, Valsecchi V, Mura T, Aouinti S, Padern G, Ferreira R, et al. Management of patients with rheumatoid arthritis by telemedicine: connected monitoring. a randomized controlled trial. Joint Bone Spine 2022;89(5):105368. [doi: 10.1016/j.jbspin.2022.105368] [Medline: 35248737]
33. Laurberg T, Schougaard LMV, Hjollund NHI, Lomborg KE, Hansen TK, Jensen AL. Randomized controlled study to evaluate the impact of flexible patient-controlled visits in people with type 1 diabetes: the DiabetesFlex trial. Diabet Med 2022;39(5):e14791. [doi: 10.1111/dme.14791] [Medline: 35028992]
34. Frielitz FS, Eisemann N, Werner K, Hiort O, Katalinic A, Lange K, et al. Direct costs of healthcare for children with type 1 diabetes using a CGM system: a health economic analysis of the VIDIKI telemedicine study in a German setting. Exp Clin Endocrinol Diabetes 2022;130(9):614-620. [doi: 10.1055/a-1708-3134] [Medline: 34979571]
35. Krzyzanowska MK, Julian JA, Gu CS, Powis M, Li Q, Enright K, et al. Remote, proactive, telephone based management of toxicity in outpatients during adjuvant or neoadjuvant chemotherapy for early stage breast cancer: pragmatic, cluster randomised trial. BMJ 2021;375:e066588 [FREE Full text] [doi: 10.1136/bmj-2021-066588] [Medline: 34880055]
36. Brouwers RWM, van der Poort EKJ, Kemps HMC, van den Akker-van Marle ME, Kraal JJ. Cost-effectiveness of cardiac telerehabilitation with relapse prevention for the treatment of patients with coronary artery disease in the netherlands. JAMA Netw Open 2021;4(12):e2136652 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.36652] [Medline: 34854907]
37. Taguchi K, Numata N, Takanashi R, Takemura R, Yoshida T, Kutsuzawa K, et al. Clinical effectiveness and cost-effectiveness of videoconference-based integrated cognitive behavioral therapy for chronic pain: randomized controlled trial. J Med Internet Res 2021;23(11):e30690 [FREE Full text] [doi: 10.2196/30690] [Medline: 34813489]
38. Yang L, Xu J, Kang C, Bai Q, Wang X, Du S, et al. Effects of mobile phone-based telemedicine management in patients with type 2 diabetes mellitus: a randomized clinical trial. Am J Med Sci 2022;363(3):224-231. [doi: 10.1016/j.amjms.2021.09.001] [Medline: 34534510]
39. Damery S, Jones J, O'Connell Francischetto E, Jolly K, Lilford R, Ferguson J. Remote consultations versus standard face-to-face appointments for liver transplant patients in routine hospital care: feasibility randomized controlled trial of myVideoClinic. J Med Internet Res 2021;23(9):e19232 [FREE Full text] [doi: 10.2196/19232] [Medline: 34533461]
40. Contal O, Poncin W, Vaudan S, de Lys A, Takahashi H, Bochet S, et al. One-year adherence to continuous positive airway pressure with telemonitoring in sleep apnea hypopnea syndrome: a randomized controlled trial. Front Med (Lausanne) 2021;8:626361 [FREE Full text] [doi: 10.3389/fmed.2021.626361] [Medline: 33959620]
41. Shdaifat MBM, Khasawneh RA, Alefan Q. Clinical and economic impact of telemedicine in the management of pediatric asthma in Jordan: a pharmacist-led intervention. J Asthma 2022;59(7):1452-1462. [doi: 10.1080/02770903.2021.1924774] [Medline: 33941032]
42. Gomis-Pastor M, Mirabet Perez S, Roig Minguell E, Brossa Loidi V, Lopez Lopez L, Ros Abarca S, et al. Mobile health to improve adherence and patient experience in heart transplantation recipients: the mHeart trial. Healthcare (Basel) 2021;9(4):463 [FREE Full text] [doi: 10.3390/healthcare9040463] [Medline: 33919899]
43. Giusto LL, Derisavifard S, Zahner PM, Rueb JJ, Deyi L, Jiayi L, et al. Telemedicine follow-up is safe and efficacious for synthetic midurethral slings: a randomized, multi-institutional control trial. Int Urogynecol J 2022;33(4):1007-1015 [FREE Full text] [doi: 10.1007/s00192-021-04767-1] [Medline: 33877376]
44. Bonnaud G, Haennig A, Altwegg R, Caron B, Boivineau L, Zallot C, et al. Real-life pilot study on the impact of the telemedicine platform EasyMICI-MaMICI on quality of life and quality of care in patients with inflammatory bowel disease. Scand J Gastroenterol 2021;56(5):530-536. [doi: 10.1080/00365521.2021.1894602] [Medline: 33691075]
45. Batalik L, Dosbaba F, Hartman M, Konecny V, Batalikova K, Spinar J. Long-term exercise effects after cardiac telerehabilitation in patients with coronary artery disease: 1-year follow-up results of the randomized study. Eur J Phys Rehabil Med 2021;57(5):807-814 [FREE Full text] [doi: 10.23736/S1973-9087.21.06653-3] [Medline: 33619944]
46. Harkey K, Kaiser N, Zhao J, Hetherington T, Gutnik B, Matthews BD, et al. Postdischarge virtual visits for low-risk surgeries: a randomized noninferiority clinical trial. JAMA Surg 2021;156(3):221-228 [FREE Full text] [doi: 10.1001/jamasurg.2020.6265] [Medline: 33439221]
47. Flynn A, Preston E, Dennis S, Canning CG, Allen NE. Home-based exercise monitored with telehealth is feasible and acceptable compared to centre-based exercise in Parkinson's disease: a randomised pilot study. Clin Rehabil 2021;35(5):728-739. [doi: 10.1177/0269215520976265] [Medline: 33272025]

XSL•FO
RenderX

48. Chapoutot M, Peter-Derex L, Schoendorff B, Faivre T, Bastuji H, Putois B. Telehealth-delivered CBT-I programme enhanced by acceptance and commitment therapy for insomnia and hypnotic dependence: a pilot randomized controlled trial. J Sleep Res 2021;30(1):e13199. [doi: 10.1111/jsr.13199] [Medline: 33020985]
49. Pers YM, Valsecchi V, Mura T, Aouinti S, Filippi N, Marouen S, et al. A randomized prospective open-label controlled trial comparing the performance of a connected monitoring interface versus physical routine monitoring in patients with rheumatoid arthritis. Rheumatology (Oxford) 2021;60(4):1659-1668. [doi: 10.1093/rheumatology/keaa462] [Medline: 33020846]
50. Arnedt JT, Conroy DA, Mooney A, Furgal A, Sen A, Eisenberg D. Telemedicine versus face-to-face delivery of cognitive behavioral therapy for insomnia: a randomized controlled noninferiority trial. Sleep 2021;44(1):zsaa136. [doi: 10.1093/sleep/zsaa136] [Medline: 32658298]
51. Watanabe E, Yamazaki F, Goto T, Asai T, Yamamoto T, Hirooka K, et al. Remote management of pacemaker patients with biennial in-clinic evaluation: continuous home monitoring in the Japanese at-home study: a randomized clinical trial. Circ Arrhythm Electrophysiol 2020;13(5):e007734 [FREE Full text] [doi: 10.1161/CIRCEP.119.007734] [Medline: 32342703]
52. Treskes RW, van Winden LAM, van Keulen N, van der Velde ET, Beeres SLMA, Atsma DE, et al. Effect of smartphone-enabled health monitoring devices vs regular follow-up on blood pressure control among patients after myocardial infarction: a randomized clinical trial. JAMA Netw Open 2020;3(4):e202165 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.2165] [Medline: 32297946]
53. Laver K, Liu E, Clemson L, Davies O, Gray L, Gitlin LN, et al. Does telehealth delivery of a dyadic dementia care program provide a noninferior alternative to face-to-face delivery of the same program? A randomized, controlled trial. Am J Geriatr Psychiatry 2020;28(6):673-682 [FREE Full text] [doi: 10.1016/j.jagp.2020.02.009] [Medline: 32234275]
54. Kane LT, Thakar O, Jamgochian G, Lazarus MD, Abboud JA, Namdari S, et al. The role of telehealth as a platform for postoperative visits following rotator cuff repair: a prospective, randomized controlled trial. J Shoulder Elbow Surg 2020;29(4):775-783. [doi: 10.1016/j.jse.2019.12.004] [Medline: 32197766]
55. Augestad KM, Sneve AM, Lindsetmo RO. Telemedicine in postoperative follow-up of STOMa PAtients: a randomized clinical trial (the STOMPA trial). Br J Surg 2020;107(5):509-518. [doi: 10.1002/bjs.11491] [Medline: 32100297]
56. Noel K, Messina C, Hou W, Schoenfeld E, Kelly G. Tele-transitions of care (TTOC): a 12-month, randomized controlled trial evaluating the use of telehealth to achieve triple aim objectives. BMC Fam Pract 2020;21(1):27 [FREE Full text] [doi: 10.1186/s12875-020-1094-5] [Medline: 32033535]
57. Nelson M, Bourke M, Crossley K, Russell T. Telerehabilitation is non-inferior to usual care following total hip replacement - a randomized controlled non-inferiority trial. Physiotherapy 2020;107:19-27. [doi: 10.1016/j.physio.2019.06.006] [Medline: 32026820]
58. Lopez-Villegas A, Catalan-Matamoros D, Peiro S, Lappegard KT, Lopez-Liria R. Cost-utility analysis of telemonitoring versus conventional hospital-based follow-up of patients with pacemakers. The NORDLAND randomized clinical trial. PLoS One 2020;15(1):e0226188 [FREE Full text] [doi: 10.1371/journal.pone.0226188] [Medline: 31995558]
59. Ruiz de Adana MS, Alhambra-Expósito MR, Muñoz-Garach A, Gonzalez-Molero I, Colomo N, Torres-Barea I, Diabetes Group of SAEDYN (Andalusian Society of Endocrinology, Diabetes,Nutrition). Randomized study to evaluate the impact of telemedicine care in patients with type 1 diabetes with multiple doses of insulin and suboptimal hba in Andalusia (Spain): PLATEDIAN study. Diabetes Care 2020;43(2):337-342 [FREE Full text] [doi: 10.2337/dc19-0739] [Medline: 31831473]
60. Guo P, Qiao W, Sun Y, Liu F, Wang C. Telemedicine technologies and tuberculosis management: a randomized controlled trial. Telemed J E Health 2020;26(9):1150-1156. [doi: 10.1089/tmj.2019.0190] [Medline: 31794684]
61. Avila A, Claes J, Buys R, Azzawi M, Vanhees L, Cornelissen V. Home-based exercise with telemonitoring guidance in patients with coronary artery disease: does it improve long-term physical fitness? Eur J Prev Cardiol 2020;27(4):367-377. [doi: 10.1177/2047487319892201] [Medline: 31787026]
62. Murase K, Tanizawa K, Minami T, Matsumoto T, Tachikawa R, Takahashi N, et al. A randomized controlled trial of telemedicine for long-term sleep apnea continuous positive airway pressure management. Ann Am Thorac Soc 2020;17(3):329-337. [doi: 10.1513/AnnalsATS.201907-494OC] [Medline: 31689141]
63. Nelson M, Russell T, Crossley K, Bourke M, McPhail S. Cost-effectiveness of telerehabilitation versus traditional care after total hip replacement: a trial-based economic evaluation. J Telemed Telecare 2021;27(6):359-366. [doi: 10.1177/1357633X19869796] [Medline: 31530065]
64. Haghnia Y, Samad-Soltani T, Yousefi M, Sadr H, Rezaei-Hachesu P. Telepsychiatry- based care for the treatment follow-up of Iranian war veterans with post-traumatic stress disorder: a randomized controlled trial. Iran J Med Sci 2019;44(4):291-298 [FREE Full text] [doi: 10.30476/IJMS.2019.44944] [Medline: 31439972]
65. Bednarski BK, Nickerson TP, You YN, Messick CA, Speer B, Gottumukkala V, et al. Randomized clinical trial of accelerated enhanced recovery after minimally invasive colorectal cancer surgery (RecoverMI trial). Br J Surg 2019;106(10):1311-1318 [FREE Full text] [doi: 10.1002/bjs.11223] [Medline: 31216065]
66. Yaron M, Sher B, Sorek D, Shomer M, Levek N, Schiller T, et al. A randomized controlled trial comparing a telemedicine therapeutic intervention with routine care in adults with type 1 diabetes mellitus treated by insulin pumps. Acta Diabetol 2019;56(6):667-673. [doi: 10.1007/s00592-019-01300-1] [Medline: 30783823]

XSL•FO
RenderX

67. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
68. Coker TR, Porras-Javier L, Zhang L, Soares N, Park C, Patel A, et al. A telehealth-enhanced referral process in pediatric primary care: a cluster randomized trial. Pediatrics 2019;143(3):e20182738. [doi: 10.1542/peds.2018-2738] [Medline: 30770523]
69. Xu L, Yi H, Pi M, Zhang C, Keenan BT, Glick HA, et al. Telemedicine management of obstructive sleep apnea disorder in China: a randomized, controlled, non-inferiority trial. Sleep Breath 2024;28(3):1173-1185. [doi: 10.1007/s11325-024-02994-6] [Medline: 38225441]
70. Guaracha-Basáñez GA, Contreras-Yáñez I, Estrada González VA, Pacheco-Santiago LD, Valverde-Hernández SS, Pascual-Ramos V. Impact of a hybrid medical care model in the rheumatoid arthritis patient-reported outcomes: a non-inferiority crossover randomized study. J Telemed Telecare 2024;30(6):931-940. [doi: 10.1177/1357633X221122098] [Medline: 36046945]
71. Ahmed MAEK, Zakaria MF, Elaziz AAEA, Fouad MM, Elbokl AM, Swelam MS. Assessment of the role of telemedicine in the outcome of multiple sclerosis patients. Egypt J Neurol Psychiatry Neurosurg 2023;59(1):99. [doi: 10.1186/s41983-023-00690-y]
72. Strassberger-Nerschbach N, Magyaros F, Maria W, Ehrentraut H, Ghamari S, Schenk A, et al. Quality comparison of remote anesthetic consultation versus on-site consultation in children with sedation for a magnetic resonance imaging examination-a randomized controlled trial. Paediatr Anaesth 2023;33(8):647-656. [doi: 10.1111/pan.14679] [Medline: 37069740]
73. Mayet S, Gledhill A, McCaw I, Hashmani Z, Drozdova Z, Arshad S, et al. Telemedicine in addictions: feasibility randomised controlled trial. Heroin addiction and related clinical problems 2023;25(3):27-36.
74. Mooney SS, Gill GK, Readman E. Virtual clinics in gynaecology - can we shorten the wait? A randomised controlled trial implementing a novel care pathway for postmenopausal bleeding. Aust N Z J Obstet Gynaecol 2022;62(5):732-739. [doi: 10.1111/ajo.13573] [Medline: 35754324]
75. Whittington JR, Hughes DS, Rabie NZ, Ounpraseuth ST, Nembhard WN, Chauhan SP, et al. Detection of fetal anomalies by remotely directed and interpreted ultrasound (teleultrasound): a randomized noninferiority trial. Am J Perinatol 2022;39(2):113-119. [doi: 10.1055/s-0041-1739352] [Medline: 34808687]
76. Davidson L, Haynes SC, Favila-Meza A, Hoch JS, Tancredi DJ, Bares AD, et al. Parent experience and cost savings associated with a novel tele-physiatry program for children living in rural and underserved communities. Arch Phys Med Rehabil 2022;103(1):8-13. [doi: 10.1016/j.apmr.2021.07.807] [Medline: 34425088]
77. Sydow H, Prescher S, Koehler F, Koehler K, Dorenkamp M, Spethmann S, et al. Non-invasive telemedical interventional management and its cost-effectiveness in patients with heart failure: economic results of the TIM-HF2 trial. Eur J Heart Fail 2022;24(Supplement 2):262.
78. Macfarlane GJ, Beasley M, Scott N, Chong H, McNamee P, McBeth J, et al. Maintaining musculoskeletal health using a behavioural therapy approach: a population-based randomised controlled trial (the MAmMOTH Study). Ann Rheum Dis 2021;80(7):903-911. [doi: 10.1136/annrheumdis-2020-219091] [Medline: 33526434]
79. Raso MG, Arcuri F, Liperoti S, Mercurio L, Mauro A, Cusato F, et al. Telemonitoring of patients with chronic traumatic brain injury: a pilot study. Front Neurol 2021;12:598777. [doi: 10.3389/fneur.2021.598777] [Medline: 33868141]
80. Tönnies J, Hartmann M, Wensing M, Szecsenyi J, Peters-Klimm F, Brinster R, et al. Mental health specialist video consultations versus treatment-as-usual for patients with depression or anxiety disorders in primary care: randomized controlled feasibility trial. JMIR Ment Health 2021;8(3):e22569 [FREE Full text] [doi: 10.2196/22569] [Medline: 33709931]
81. Mínguez Clemente P, Pascual-Carrasco M, Mata Hernández C, Malo de Molina R, Arvelo LA, Cadavid B, et al. Follow-up with telemedicine in early discharge for COPD exacerbations: randomized clinical trial (TELEMEDCOPD-Trial). COPD 2021;18(1):62-69 [FREE Full text] [doi: 10.1080/15412555.2020.1857717] [Medline: 33307857]
82. Duiverman ML, Vonk JM, Bladder G, van Melle JP, Nieuwenhuis J, Hazenberg A, et al. Home initiation of chronic non-invasive ventilation in COPD patients with chronic hypercapnic respiratory failure: a randomised controlled trial. Thorax 2020;75(3):244-252. [doi: 10.1136/thoraxjnl-2019-213303] [Medline: 31484786]
83. de Jong MJ, Boonen A, van der Meulen-de Jong AE, Romberg-Camps MJ, van Bodegraven AA, Mahmmod N, et al. Cost-effectiveness of telemedicine-directed specialized vs standard care for patients with inflammatory bowel diseases in a randomized trial. Clin Gastroenterol Hepatol 2020;18(8):1744-1752. [doi: 10.1016/j.cgh.2020.04.038] [Medline: 32335133]
84. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient reported outcomes with remote orthopaedic consultations by telemedicine: a randomised controlled trial. J Telemed Telecare 2019;25(8):451-459. [doi: 10.1177/1357633X18783921] [Medline: 29973130]
85. Barker A, Cameron P, Flicker L, Arendts G, Brand C, Etherton-Beer C, et al. Evaluation of RESPOND, a patient-centred program to prevent falls in older people presenting to the emergency department with a fall: a randomised controlled trial. PLoS Med 2019;16(5):e1002807. [doi: 10.1371/journal.pmed.1002807] [Medline: 31125354]
86. Sekimoto S, Oyama G, Hatano T, Sasaki F, Nakamura R, Jo T, et al. A randomized crossover pilot study of telemedicine delivered via ipads in Parkinson's disease. Parkinsons Dis 2019;2019:9403295. [doi: 10.1155/2019/9403295] [Medline: 30723541]

XSL•FO
RenderX

87. Bohingamu Mudiyanselage S, Stevens J, Watts JJ, Toscano J, Kotowicz MA, Steinfort CL, et al. Personalised telehealth intervention for chronic disease management: a pilot randomised controlled trial. J Telemed Telecare 2019;25(6):343-352. [doi: 10.1177/1357633X18775850] [Medline: 29793387]
88. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
89. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the telemedicine satisfaction questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
90. Garratt AM, Bjaertnes Ø, Krogstad U, Gulbrandsen P. The outpatient experiences questionnaire (OPEQ): data quality, reliability, and validity in patients attending 52 Norwegian hospitals. Qual Saf Health Care 2005;14(6):433-437 [FREE Full text] [doi: 10.1136/qshc.2005.014423] [Medline: 16326790]
91. Aoki T. Significance and prospects of patient experience (PX) assessment [Article in Japanese]. Journal of the Society for Healthcare Quality and Safety 2022;17(4):393-398 [FREE Full text]
92. Morales Asencio JM, Bonill de Las Nieves C, Celdrán Mañas M, Morilla Herrera JC, Martín Santos FJ, Contreras Fernández E, et al. Design and validation of a home care satisfaction questionnaire: SATISFAD. Gac Sanit 2007;21(2):106-113 [FREE Full text] [doi: 10.1157/13101036] [Medline: 17419926]
93. Stavem K, Augestad LA, Kristiansen IS, Rand K. General population norms for the EQ-5D-3 L in Norway: comparison of postal and web surveys. Health Qual Life Outcomes 2018;16(1):204 [FREE Full text] [doi: 10.1186/s12955-018-1029-1] [Medline: 30340499]
94. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095-1108. [doi: 10.1097/00005650-199711000-00002] [Medline: 9366889]
95. Loge JH, Kaasa S, Hjermstad MJ, Kvien TK. Translation and performance of the Norwegian SF-36 health survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. J Clin Epidemiol 1998;51(11):1069-1076. [doi: 10.1016/s0895-4356(98)00098-5] [Medline: 9817124]
96. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International quality of life assessment. J Clin Epidemiol 1998;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7] [Medline: 9817135]
97. Green KC, Armstrong JS, Graefe A. Methods to elicit forecasts from groups: delphi and prediction markets compared. SSRN Journal 2008;8:17-20. [doi: 10.2139/ssrn.1153124]
98. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. Health Expect 2021;24(4):1015-1024 [FREE Full text] [doi: 10.1111/hex.13254] [Medline: 33949755]
99. Friedel AL, Siegel S, Kirstein CF, Gerigk M, Bingel U, Diehl A, et al. Measuring patient experience and patient satisfaction-how are we doing it and why does it matter? A comparison of European and U.S. American approaches. Healthcare (Basel) 2023;11(6):797 [FREE Full text] [doi: 10.3390/healthcare11060797] [Medline: 36981454]
100. Bleich SN, Ozaltin E, Murray CKL. How does satisfaction with the health-care system relate to patient experience? Bull World Health Organ 2009;87(4):271-278 [FREE Full text] [doi: 10.2471/blt.07.050401] [Medline: 19551235]
101. Oben P. Understanding the patient experience: a conceptual framework. J Patient Exp 2020;7(6):906-910 [FREE Full text] [doi: 10.1177/2374373520951672] [Medline: 33457518]
102. Meltzer MI, Shapiro CN, Mast EE, Arcari C. The economics of vaccinating restaurant workers against hepatitis A. Vaccine 2001;19(15-16):2138-2145. [doi: 10.1016/s0264-410x(00)00396-0] [Medline: 11228386]
103. Sanders GD, Maciejewski ML, Basu A. Overview of cost-effectiveness analysis. JAMA 2019;321(14):1400-1401. [doi: 10.1001/jama.2019.1265] [Medline: 30855638]
104. Cost-effectiveness analysis. U.S. Department of Veterans Affairs. URL: https://www.herc.research.va.gov/include/page.asp?id=cost-effectiveness-analysis [accessed 2024-10-01]
105. Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technol Assess Health Care 1989;5(4):559-575. [doi: 10.1017/s0266462300008461] [Medline: 2634630]
106. Greenhouse gas protocol. World Resources Institute. URL: https://www.wri.org/initiatives/greenhouse-gas-protocol [accessed 2024-11-27]
107. Rodríguez-Jiménez L, Romero-Martín M, Spruell T, Steley Z, Gómez-Salgado J. The carbon footprint of healthcare settings: a systematic review. J Adv Nurs 2023;79(8):2830-2844. [doi: 10.1111/jan.15671] [Medline: 37198974]
108. Williams JTW, Bell KJL, Morton RL, Dieng M. Methods to include environmental impacts in health economic evaluations and health technology assessments: a scoping review. Value Health 2024;27(6):794-804 [FREE Full text] [doi: 10.1016/j.jval.2024.02.019] [Medline: 38462223]

Abbreviations

HbA1c: hemoglobin A1c
IQR: interquartile range

XSL•FO
RenderX

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analysis Extension for Scoping Reviews
PRO: patient-reported outcome
QALY: quality-adjusted life year
QoL: quality of life
RCT: randomized control trial
SF-12: Medical Outcomes Study Short Form 12-Item Health Survey
SF-36: Medical Outcomes Study Short Form 36-Item Health Survey

Edited by N Cahill; submitted 25.10.24; peer-reviewed by M Boarini, JCH Fernandes, A Adebisi; comments to author 18.11.24; revised version received 12.12.24; accepted 23.12.24; published 31.01.25.

Please cite as:
Sugawara Y, Hirakawa Y, Iwagami M, Inokuchi R, Wakimizu R, Nangaku M
Metrics for Evaluating Telemedicine in Randomized Controlled Trials: Scoping Review
J Med Internet Res 2025;27:e67929
URL: https://www.jmir.org/2025/1/e67929
doi:10.2196/67929
PMID:

©Yuka Sugawara, Yosuke Hirakawa, Masao Iwagami, Ryota Inokuchi, Rie Wakimizu, Masaomi Nangaku. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Dynamic Augmented Reality Cues for Telementoring in Minimally Invasive Surgeries: Scoping Review

Hawa Hamza[1], BSc; Omar M Aboumarzouk[1], PhD; Abdulla Al-Ansari[1], MD; Nikhil V Navkar[1], PhD

Department of Surgery, Hamad Medical Corporation, Doha, Qatar

Corresponding Author:
Nikhil V Navkar, PhD
Department of Surgery
Hamad Medical Corporation
PO Box 3050
Doha, 3050
Qatar
Phone: 974 77606674
Email: nnavkar@hamad.qa

Abstract

Background: Remote surgeons use telementoring technologies to provide real-time guidance during minimally invasive surgeries (MIS). Such technologies are continuously improving with the integration of dynamic augmented reality (AR) cues. This includes virtual overlays of hand gestures, pointers, and surgical tools onto the operating surgeon's view. The operating surgeon comprehends this augmented information to operate on the patient. Thus, understanding these dynamic AR cues (either during surgical training or live surgery) is crucial.

Objective: In this paper, we aimed to review the existing telementoring technologies that use dynamic AR cues during MIS. This review describes the MIS (including surgery type, specialty, procedure, and clinical trial), the telementoring system, the dynamic AR cues generated by these systems, and evaluation of the technology in terms of technical aspects, user perceptions, skills gained, and patient outcomes.

Methods: A scoping review was conducted using PubMed, Web of Science, Scopus, IEEE Xplore, and ACM Digital Library databases. The search terms included "telementoring," "minimally invasive surgery," and "augmented reality" without restrictions imposed on the publication year. Articles covering telementoring using dynamic AR cues during MIS, including laparoscopic and robot-assisted, were identified.

Results: A total of 21 articles were included and categorized based on type of surgery, the telementoring technology used, and evaluation of the technology. Most of the articles reported on laparoscopic suturing performed using synthetic phantoms. Hand gestures and surgical tools were the most frequently used dynamic AR cues (10 articles on each cue), while the mentors and mentees primarily consisted of experienced surgeons and medical students. The studies assessing the telementoring technologies were either descriptive (7 articles) or analytical (14 articles) where it was compared against no cue, prerecorded visual cue, in-person guidance, audio cue, or static AR cue. Outcomes were measured mostly using skills gained (13 articles) and user perception about the telementoring system.

Conclusions: In general, telementoring using dynamic AR cues resulted in positive outcomes during MIS. In some cases, they were considered on par with conventional methods such as audio cues and in-person guidance. Further randomized controlled trials are required to objectively assess its clinical benefits.

(J Med Internet Res 2025;27:e63939) doi:10.2196/63939

KEYWORDS

minimally invasive surgery; surgeons; laparoscopic; telementoring; teleproctoring; telemedicine; augmented reality; dynamic visual cue; technologies; robotics; patient outcomes; communication; scoping review; PRISMA

XSL•FO
RenderX

Introduction

One of the telemedicine technologies that is gaining considerable interest is telementoring, during which an expert medical professional (mentor) can provide real-time instructions to a novice (mentee) even if they are remotely located [1]. A growing application of telementoring technologies is seen in the field of surgery as a result of rising demands for specialized surgical expertise coupled with limited training opportunities [2]. Minimally invasive surgery (MIS) is one such specialized surgical field, which has emerged as the standard practice [3]. During MIS, which includes laparoscopic and robot-assisted procedures, surgeons use long, thin surgical instruments inserted through small incisions on the body. The operative view is captured through a camera scope and displayed on a screen. MIS is increasingly adopted due to the benefits it provides to patients, in terms of significant reduction in pain, surgical site infection, scarring, and recovery time as compared with the highly invasive open approach [4]. Despite these pronounced advantages, its widespread implementation is hindered due to the challenges in adequate skills acquisition. Surgeons performing MIS have an indirect view of the operative view through the screen projecting the live camera scope feed and experience a longer learning curve as compared with open surgery [5]. Hence, specialized training is needed to develop sufficient expertise. While basic MIS skills are learned outside the operating room, intraoperative training is vital for attaining proficiency [5]. Unlike open surgeries where the expert surgeon can easily point at critical structures with their hands over the operative view, mentors must rely on verbal instructions during MIS, which are prone to communication errors [6]. In some cases, the mentor must take complete control of the instruments from the mentee to show the required surgical technique. While the dual surgeon console in robot-assisted surgeries facilitates such takeovers by the mentor, the cost can be prohibitively high. Furthermore, many countries experience challenges in implementing MIS due to inadequate training programs and high costs associated with bringing in expert surgeons [7]. Telementoring technologies that allow an expert surgeon to efficiently guide the operating surgeon play a major role in addressing this gap [8].

Implementing telementoring technologies during MIS is relatively easier, as compared with open surgeries, since a live view of the operating field is already captured through the camera scopes [9,10]. During conventional telementoring, video captured from the operating room is transmitted in real time to an expert mentor at the remote location. Verbal instructions (audio cues) from the mentor are transferred to the mentee [11]. The mentor can also play surgical videos (prerecorded visual cues) to guide the mentee on a particular step [12]. However, such cues may not be sufficient for intraoperative instructions [6]. To overcome the communication challenges posed by verbal instructions, expert surgeons need a way to point at anatomical structures on the screen displaying the operative view to the mentee [13]. For this purpose, the mentor may use augmented reality (AR), which involves the use of computer-generated virtual objects that are superimposed on the mentee's view of the operative field [14]. The AR cues may be rendered to the mentee on a visualization screen, or the surgeon's console with stereoscopic vision during robot-assisted surgery. In some cases, external AR devices such as Microsoft HoloLens head-mounted display (HMD) or iPad tablets may also be used [15]. The cues could be in the form of 2-dimensional (2D) annotations (static AR cues) demonstrating an incision line or highlighting a specific anatomy [16,17]. Such forms of telementoring have been applied extensively for MIS demonstrating improved communication and performance [18]. Nonetheless, static AR cues are limited while compared with in-person guidance as complex tool-tissue interactions may not be accurately depicted [19,20].

Telementoring can be enhanced with the use of dynamic AR cues to effectively communicate surgical steps. As opposed to static AR cues, which remain stationary on the mentee's field of view, dynamic AR cues move in real time under the mentor's control [21]. These virtual cues can be in the form of hand gestures, pointers, or surgical tools. They are often 3-dimensional (3D) in nature, providing depth perception as the mentor guides the mentee through surgical sub steps. In cases where the mentor and mentee have similar macro skills, such as proficiency in critical anatomy and movement of surgical tools, these dynamic cues can aid the mentee's development of micro skills related to a particular unfamiliar surgical technique [22]. Dynamic AR cues are not only useful for telementoring when the mentor is situated in a remote location, but can also be used to enhance in-person guidance during MIS where the mentor and mentee are in the same operating room [23].

Several reviews related to telementoring in health care have been previously published; however, they did not focus on applications for surgery in specific [15,24]. Other reviews covering surgical applications of telemedicine and telementoring have been presented [1,11,25-30]. These articles largely covered implementations of telementoring systems where video feed from the mentee is sent for audio feedback from the remote mentor. Some of them did not focus on MIS and explored applications in open surgery as well [11,29,30]. Additional review articles focused on specific use of telementoring for certain specialties such as head and neck [31], robotic [32], and urology [33]. With respect to MIS, Nickel et al [23] provided a concise clinical summary, whereas others reported on the use of audio and static AR cues [16,17,19,34,35]. While these reviews offer a comprehensive understanding of telementoring during MIS, they do not highlight the developments in the use of dynamic AR cues. To the best of our knowledge, there has not yet been a systematic examination specifically on the use of dynamic AR cues used by remote mentors during telementoring in MIS. Such a synthesis providing an overview of the different types of dynamic AR cues as well as the surgical procedures, clinical settings, and factors used for testing is needed. This will help guide future research and developments on telementoring in MIS and ensure that they are relevant and practical.

Through this scoping review we aim to provide a comprehensive summary of dynamic AR cues used for telementoring during MIS. A survey of the various types of surgeries, participants, dynamic AR cues, and evaluation of the telementoring technology is presented to guide future developments on

XSL•FO
RenderX

telementoring during MIS. We aim to address the following research questions (RQs), which are yet to be answered by review articles published in the field:

- RQ1: What are the MIS specialties and procedures where dynamic AR cues have been used for telementoring? What are the clinical trials or settings they were tested on?
- RQ2: What are the different types of dynamic AR cues used for MIS telementoring? What information does the mentor convey to the mentee using these cues?
- RQ3: What factors are used to evaluate the use of dynamic AR cues during MIS telementoring? What are the common outcomes reported?

Methods

Overview

The review was conducted based on the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines [36].

Eligibility Criteria

To be considered eligible for the review, articles must satisfy the following criteria (Textbox 1).

Textbox 1. Eligibility criteria for scoping review.

Inclusion criteria

- Original article.
- Reporting telementoring technologies using dynamic augmented reality cue.
- Reporting telementoring technologies during minimally invasive surgeries (including, endoscopic, laparoscopic, and robot-assisted).
- Published in English.

Exclusion criteria

- Articles on nonsurgical applications.
- Open surgery (ie, not minimally invasive surgeries).
- Microsurgery.
- Studies without telementoring technologies.
- Reports of telementoring without the use of augmented reality.
- Telementoring using static augmented reality cues (annotations) only.
- Articles without enough information (such as editorials and comments).
- Review articles.

Search

A thorough search of scientific databases including PubMed, Web of Science, and Scopus was conducted. Other databases such as “IEEE *Xplore*” and “ACM Digital Library” covering technical fields were also included. The search terms comprised of, but were not restricted to, “telementoring,” “teleproctoring,” “augmented reality,” “overlay,” and “minimally invasive surgery.” There were no restrictions imposed on the year of publication. The latest search was conducted on March 11, 2024. The strategy used for searching in PubMed is provided in Multimedia Appendix 1. Additional records were found through citation searches and scanning review articles.

Study Selection

Through the searches, a total of 460 scientific records were identified. Duplicate removal, title and abstract and full-text screening were done using the Rayyan web app [37]. Two reviewers [HH] and [NN] were independently responsible for title and abstract and full-text screening process. Resolution of disagreements on eligibility of an article was done through consensus and discussion with a third reviewer [OA] when necessary.

Data Extraction

Data items extracted from the articles included year of publication, type of surgery, surgical specialty and procedure, phase of clinical trial, number of participants and their experience level, telementoring technology, type of dynamic AR cue, comparator (if applicable), factor measured, and outcome of the study. The data extraction form with the variables to be extracted from each eligible article were jointly developed by 2 reviewers [HH] and [NN] for this scoping review. The reviewers were independently responsible for extracting the data from each study. Resolution of disagreements on results of the data extraction was achieved through discussion with a third reviewer [OA] when necessary.

Quality Assessment

Methodological quality assessment of the included studies was conducted based on the Medical Education Research Study Quality Instrument (MERSQI) [38]. The tool is designed to appraise the quality of medical education research, which was found to be appropriate for the topic addressed by this scoping review. The quality assessment tool was not used as an exclusion criterion, but instead, it was used after the screening process to gauge the quality and reliability of the articles reporting the use of dynamic AR cues during minimally invasive surgical

XSL•FO
RenderX

telementoring [39]. Each article was assessed according to study design, type of data obtained, evaluation instrument used, analysis of the data, and outcomes measured to give a score out of 18. A higher score given to an article indicates a superior study design. Similar to the study selection and data extraction process, 2 reviewers [HH] and [NN] were responsible for quality assessment of the included studies, while disagreements were resolved through discussions with a third reviewer [OA] when necessary.

Data Synthesis

The data extracted from the articles were presented using a descriptive table with information on surgery, participants (performing the procedure), dynamic AR cues, comparator, outcomes measured, and primary findings of the study. Study characteristics such as surgical procedure, clinical trial, type of dynamic AR cue, and publication year were visually summarized as counts. A meta-analysis was not conducted since a wide variety of outcomes measured, clinical setting, and type of participants was observed.

Results

Overview

A total of 21 articles reporting on the use of dynamic AR cues during telementoring in MIS were included in the review (Figure 1). An average MERSQI score of 12.9 was observed in the included articles. Table 1 presents a summary of the included articles. The table elements are described in detail in the following sections. They include details about the MIS procedures, the various dynamic AR cues used for telementoring, and evaluation of the technologies and their outcomes.

Figure 1. Flowchart depicting identification, screening, and inclusion of articles on the use of dynamic augmented reality cues during minimally invasive surgical telementoring.

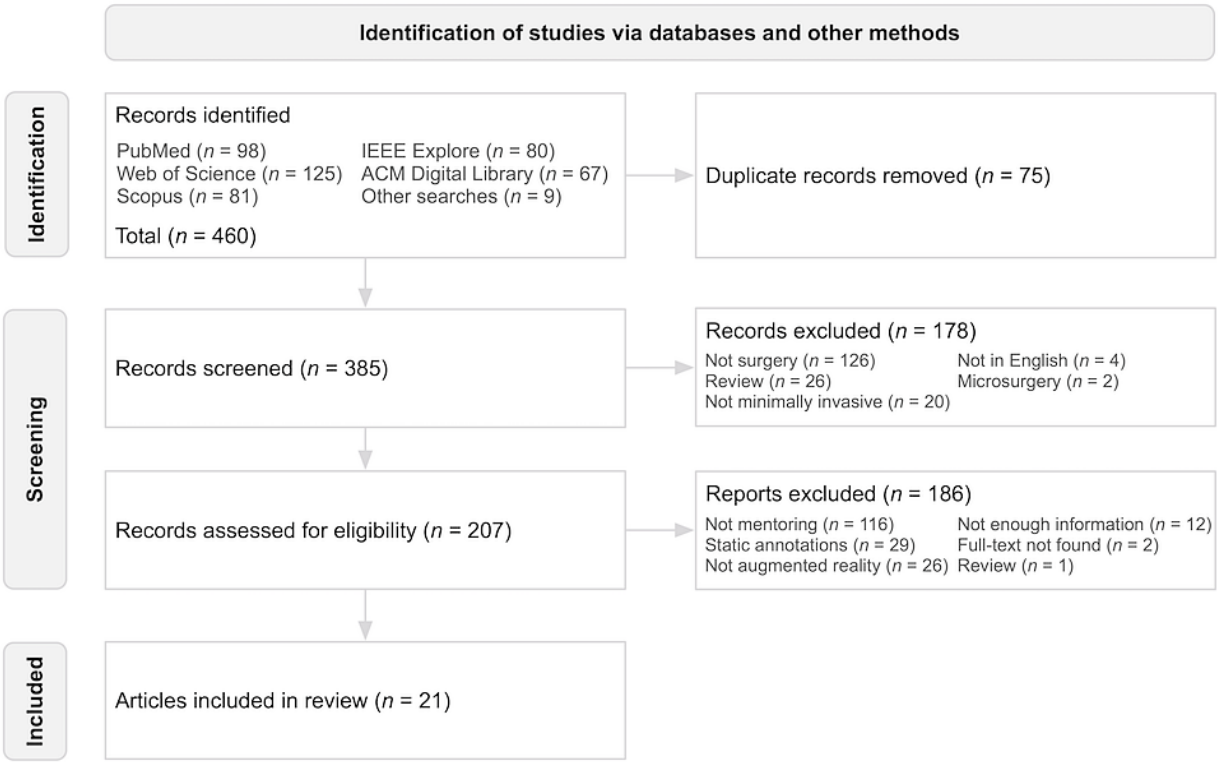

XSL•FO
RenderX

Table 1. Studies reporting dynamic augmented reality cues during minimally invasive surgery telementoring.

Study	MERSQI[a] score	Minimally invasive surgery			Mentor and mentee	Intervention		Evaluation	
		Type[b]	Procedure	Clinical trial		Technology	Dynamic AR[c] cue	Comparator	Outcome
Cizmic et al [6]	15.5	L	Cholecystectomy	Animal (cadaver)	Doctor trained by surgeon (n=1); medical students (n=40)	iSurgeon	Hand gesture	Audio cue	Skills: improved GOALS[d], OSATS[e] scores and fewer complications, no difference in operating time
Felinska et al [18]	14.5	L	Cholecystectomy; basic tasks	Animal (cadaver), virtual simulator, synthetic phantom	Mentor (n=1); medical students (n=40)	iSurgeon	Hand gesture	Audio cue	Skills: improved participant gaze behavior, lower number of errors, higher OSATS score in iSurgeon group
Huettl et al [40]	16	L	Cholecystectomy	Human (live)	Mentors (n=13); surgical trainees (n=15)	HoloPointer	Pointer	Audio cue	Perceptions, skills: reduced corrections, improved subjective performance, favorable user ratings; no difference in GOALS score, CVS[f], and operating time
Long et al [41]	10	R	Prostatectomy; suturing, basic task	Human (live), synthetic phantom	Expert surgeon (n=1); trainee (n=1)	In-house developed system	Hand gesture, surgical tool	—[g]	Technical: latency within standard acceptable latency (330 ms)
Shabir et al [42]	12.5	L	Suturing	Synthetic phantom	Mentors (n=2); mentees (n=12)	In-house developed system	Surgical tool	In-person guidance	Skills: longer operating time observed, but comparable performance
Shabir et al [20]	11.5	L	Salpingectomy	Virtual simulator	Mentor (n=1); mentees (n=12)	In-house developed system	Surgical tool	In-person guidance	Skills: lower complication percentage during learning stage; lowered operating time during testing stage
Shabir et al [43]	12.5	L	Suturing	Synthetic phantom	Mentor (n=1); mentees (n=18)	In-house developed system	Surgical tool	Prerecorded visual cue, in-person guidance	Skills: lower error count, no difference in operating time
Lowry et al [44]	15.5	L	Suturing, basic task	Synthetic phantom	Mentor (n=1); premedical, medical students (n=30)	MVR[h] HelpLightning platform	Surgical tool	No cue, in-person guidance	Skills: improvement compared to group who received no cue, similar FLS[i] skills among telementored and in-person guided groups
Shabir et al [22]	10	L	Basic task	Synthetic phantom	Mentor (n=1); mentee (n=6)	In-house developed system	Surgical tool	—	Technical: latency of 260 ms observed to send information from mentee to mentor and 132 ms vice versa

XSL•FO
RenderX

Study	MERSQI[a] score	Minimally invasive surgery			Mentor and mentee	Intervention		Evaluation	
		Type[b]	Procedure	Clinical trial		Technology	Dynamic AR[c] cue	Comparator	Outcome
Wild et al [13]	15.5	L	Cholecystectomy	Animal (cadaver)	Expert surgeon (n=1); medical students (n=60)	iSurgeon	Hand gesture	Audio cue	Perceptions, skills: faster operating time, lower error rate, improved GOALS and OSATS scores, lower complications, and reduced workload (NASA-TLX[j])
Youssef et al [45]	9	R	Prostatectomy	Human (live)	Expert urology mentors (n=2); senior urology fellows (n=2)	Proximie	Hand gesture	—	Perceptions: favorable ratings provided by mentors and mentees
El-Asmar et al [46]	16	R	Aquablation for benign prostatic obstruction	Human (live)	Mentors (n=3); surgeons (n=not specified)	Proximie	Hand gesture	In-person guidance	Patient outcome: no difference in operating time, drop in hemoglobin, urinary retention, and complications; increased anesthesia use and cauterization under telementor guidance
Heinrich et al [47]	14.5	L	Cholecystectomy	Virtual simulator	Expert surgeon (n=1); surgical trainees (n=10)	HoloPointer	Pointer	Audio cue	Skills: improvement in economy of movement, error rate, and performance
Shabir et al [48]	11	L, R	Suturing	Synthetic phantom	Mentor (n=1); mentee (n=1)	In-house developed system	Surgical tool	—	Technical: latency of 1.56 s to transfer information from operating room to remote area and 0.089 s vice versa
Feng et al [49]	13.5	L	Cholecystectomy	Synthetic phantom	Fellow trained by surgeon (n=1); residents (n=6)	Virtual pointer	Pointer	Audio cue	Skills: improvement in turn frequency (number of turns per second)
Feng et al [50]	14.5	L	Cholecystectomy	Synthetic phantom	Fellow trained by surgeon (n=1); surgical trainees (n=7)	Virtual pointer	Pointer	Audio cue	Skills: improvement in economy of movement and subjective performance, no difference in errors and operating time
Jarc et al [21]	11	R	Suturing, basic task	Animal (live)	Expert surgeons (n=6); medical students, residents (n=7)	In-house developed system	Hand gesture, pointer, surgical tool	—	Perceptions: tool rated favorably by mentors and mentees
Kowalewski et al [51]	12.5	L	Suturing, basic task	Synthetic phantom	Surgeons, mentors (n=4); novice, intermediate, expert mentees (n=30)	iSurgeon	Hand gesture	—	Skills: positive correlation observed between iSurgeon parameters and OSATS score
Davis et al [52]	10	L	Ventriculostomy	Human (live)	Expert surgeon (n=1); expert surgeon (n=1)	VIPAR[k]	Hand gesture, pointer	—	Technical: latency of 237 ms in visual cue; successful implementation of VIPAR

XSL•FO
RenderX

Study	MERSQI[a] score	Minimally invasive surgery			Mentor and mentee	Intervention		Evaluation	
		Type[b]	Procedure	Clinical trial		Technology	Dynamic AR[c] cue	Comparator	Outcome
Jarc et al [53]	11	R	Basic task	Animal (live)	Surgeons (n=26); surgical, nonsurgical trainees (n=26)	In-house developed system	Hand gesture, pointer, surgical tool	Static AR cue (2D telestration)	Perceptions: participants preferred 3D tools over 2D telestration
Vera et al [54]	14.5	L	Suturing	Synthetic phantom	Mentor (n=1); premedical, medical students (n=19)	ART[l]	Surgical tool	In-person guidance	Skills: steeper learning curve, faster operating time, fewer failed attempts in ART group

[a]MERSQI: Medical Education Research Study Quality Instrument.

[b]types: L: laparoscopic; R: robot-assisted.

[c]AR: Augmented Reality.

[d]GOALS: Global Operative Assessment of Laparoscopic Skills.

[e]OSATS: Objective Structured Assessments of Technical Skills.

[f]CVS: Critical view of safety.

[g]Not applicable.

[h]MVR: merged virtual reality.

[i]FLS: fundamentals of laparoscopic surgery.

[j]NASA-TLX: NASA (National Aeronautics and Space Administration) Task Load Index.

[k]VIPAR: virtual interactive presence and augmented reality.

[l]ART: augmented reality telementoring.

MIS (Research Question 1)

Article on MIS included 2 categories, namely, manual or conventional laparoscopic (n=16 articles) and robot-assisted (n=6). These were subdivided based on surgical specialties, which consisted of general, gynecology, neurosurgery, urology, and nonspecific specialties. Figure 2 depicts the surgical procedures and the corresponding clinical settings under which the telementoring technologies were tested. Each vertical block represents an article using dynamic augmented reality cues. Articles using multiple dynamic cues are represented as segmented vertical blocks. The surgical procedures identified included cholecystectomy [6,13,18,40,47,49,50], salpingectomy [20], ventriculostomy [52], aquablation [46], and prostatectomy [41,45]. Nonspecific procedures consisted of basic tasks (such as cutting, knot tying, ligation, peg transfer, and tool motion) [18,21,22,41,44,51,53] as well as suturing [21,41-44,48,51,54]. The procedures were conducted under four different clinical settings: virtual simulator, synthetic phantom, animal (cadaver or live), or human (live). Virtual simulators present highly realistic computer-generated surgical procedures. However, they often have limited haptic feedback providing tactile sensation [55]. Synthetic phantoms consist of low or high-fidelity tissue models. Depending on the material it is made of, it may even provide realistic tissue interaction [56]. Animal models include euthanized cadavers or anesthetized live animals, while surgeries on real patients were categorized as human (live). All the 4 clinical settings were used for telementoring during cholecystectomy procedures. Maximum number of articles covered laparoscopic suturing on synthetic phantoms (n=6). A range of dynamic AR cues were used in articles covering robot-assisted basic tasks performed on animal models.

XSL•FO
RenderX

Figure 2. Depiction of surgical procedures and the corresponding clinical settings under which the telementoring technologies were tested.

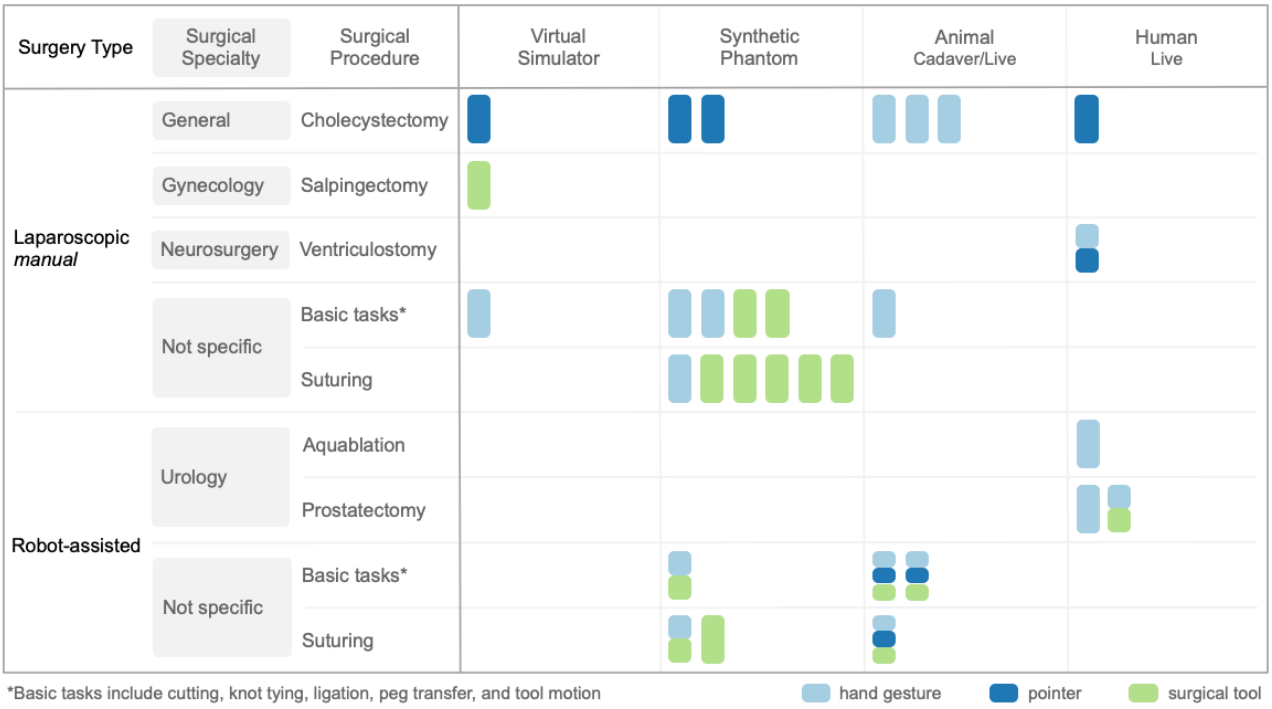

Telementoring Technology (Research Question 2)

The mentors used various forms of dynamic AR cues that were overlaid on the mentee's view of the operating field (Figure 3). These visual cues were divided into 3 types, hand gesture (displaying mentor's hand position and orientation), pointer (in the form of dynamic annotations), and surgical tool (showing the type of tool, position, or required motion). As depicted in Figure 4 [6,13,18,20-22,41-46,48,51-54], most of the articles covered hand gestures and surgical tools (n=10 articles on each cue). The telementoring system developed by Jarc et al [21,53] used all the 3 types of dynamic AR cues. Some of the pros and cons of each cue are summarized in Table 2. The following paragraphs provide a detailed description of the three dynamic AR cues.

Figure 3. Various dynamic augmented reality cues overlayed on the mentee's operative field.

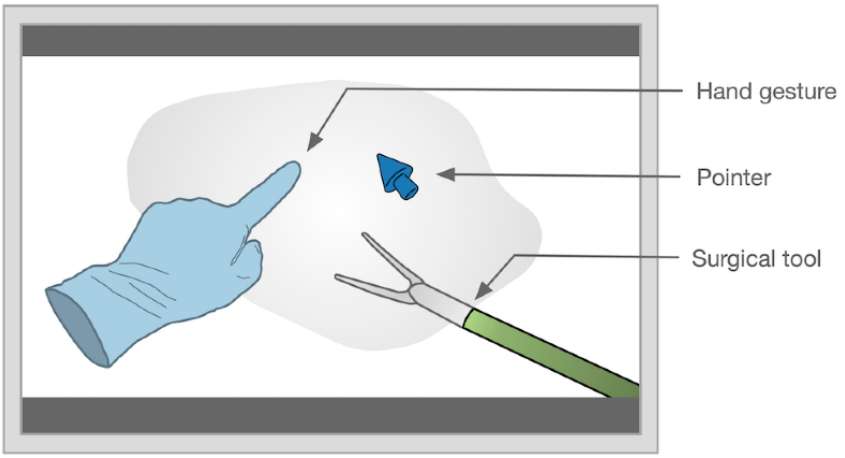

XSL•FO
RenderX

Figure 4. Number of articles reporting telementoring systems using dynamic augmented reality cues, specifically, hand gestures, pointers, and surgical tools [6,13,18,20-22,40-48,51-54]. ART: augmented reality telementoring; MVR: merged virtual reality; VIPAR: virtual interactive presence and augmented reality.

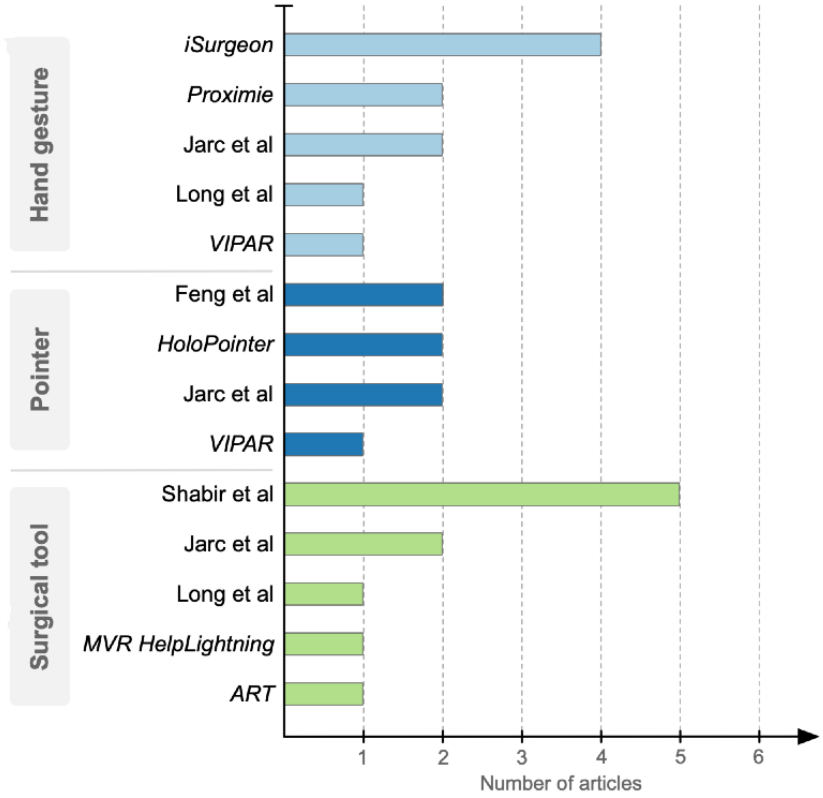

Table 2. Summary of common pros and cons of each dynamic augmented reality cue.

Dynamic augmented reality cue	Pros	Cons
Hand gesture	• Natural and intuitive mode of communication [18]. • Orientation and position can be conveyed [21]. • Grasping tool actions can be mimicked using open and close gestures with fingers [53].	• Mentee's view of the operative field may get obstructed [43]. • Complex tool tip movements may not be conveyed [48].
Pointer	• Critical anatomical structures and dissection plane can be indicated [52]. • Mentors can use hands-free interaction when pointing is enabled through head tracking [47].	• Tool tip orientation may not be conveyed [48]. • Complex instructions such as grasping and other tool movements may not be replicated [48]. • Verbal clarification of the instruction may be required [53].
Surgical tool	• Exact tool to be used, tool tip orientation, and position can be conveyed [44,54]. • Required motion and open and close gestures of the tool can be conveyed [53]. • Decreases the need for takeover of the surgery by the mentor [54].	• Mentee's view of the operative field may get obstructed [43]. • Implementation of the technology may be challenging [22].

Hand Gesture

Hand gesture was one of the most used dynamic AR cues among the included articles. It enhances the mentoring experience with the ability to point at critical anatomical structures, demonstration of grasping gestures by opening and closing fingers, and depiction of orientation through hand rotation movements [13,53]. In the iSurgeon system, the mentor's hand gestures were captured using an overhead depth video camera, which was then projected on to the mentee's view of the operative field [6,13,18,51]. In this way, the mentor was able to support the mentee by communicating the dissection plain, target anatomy, and indicating the next steps of the procedures [6]. Similarly, the hand gesture overlays were also used in telementoring systems such as Proximie and Virtual Interactive Presence and Augmented Reality (VIPAR) [46,52]. In the VIPAR system, the video camera on iPad tablet devices were used to capture the mentee's operative view as well as the mentor's hand gestures. Both the video feeds were then combined allowing a transparent overlay of the mentor's hand

XSL•FO
RenderX

to appear in the mentee's view [52]. The telementoring system developed by Jarc et al [21,53] integrated custom wireless input devices and commercial game controllers allowing the mentor to manipulate the orientation and position of a virtual hand appearing in the mentee's operative view. The mentor was also able to indicate grasping gestures (open and close) using the input device. In this way, the mentee was able to understand and translate the hand gestures to orient, position, and manipulate the surgical tool accordingly. Consequently, the use of virtual hands for telementoring was rated as highly effective by both the mentors and mentees [53]. As reflected by its frequent usage observed in the included articles, hand gestures provide a natural and intuitive mode of communication for the mentor making it easier for the mentee to interpret and follow the instructions.

Pointer

The virtual pointer was used to indicate critical anatomy or a specific region in the mentee's operative view. Frequency of its use as dynamic AR cue during telementoring was relatively lower compared with hand gestures and surgical tools among the included articles. Nevertheless, the use of pointers showed significant improvement in quality of instructions compared with verbal feedback alone [50]. The HoloPointer system used an AR HMD device (Microsoft HoloLens) to track head movements [40]. In doing so, the mentor was able to point at various regions of the operating field view in a hands-free manner [47]. In other studies, the mentors used hand movements, which were captured by a motion sensing device (Microsoft Kinect) to control the virtual pointer as described by Feng et al [49,50]. The VIPAR system, similar to its implementation for hand gestures, used an iPad tablet device to capture the pointing device used by the mentor and combine it with the mentee's operative view [52]. The 3D pointers appearing on the mentee's view were manipulated by the mentor using game controllers in the system presented by Jarc et al [21,53]. However, the use of 3D pointers received low ratings from the mentors. Overall, virtual pointers allowed the mentors to clarify verbal instructions by directing the mentee's attention to a particular region of interest. However, it was not used as frequently, since the mentor is not able to convey precise instructions, such as tool tip orientation and motion.

Surgical Tool

Virtual surgical tools provided a close-to-reality visual cue during telementoring. It enhances the realism of the mentee's training environment as the mentor can accurately convey the precise tool to be used, orientation and position of the tool tip, as well as the necessary movements (such as open and close gestures). Hence, the surgical tool as a dynamic AR cue was frequently used among the included studies. This was primarily done in 2 ways, by superimposing live video feed with real surgical tool from the mentor's side on to the mentee's view or by displaying 3D models of the surgical tool, which the mentor manipulates using various input devices. One of the earliest implementations, the augmented reality telementoring (ART) system, used a laparoscopic box simulator with an embedded camera at the remote mentor's side. As the mentor maneuvered the real surgical tool, the live video feed was superimposed on the mentee's view of the operative field. This resulted in improved understanding of instructions and quicker skills acquisition. The mentees also reported that the cue did not obstruct the operative view [54]. With respect to the use of 3D surgical tool models, Shabir et al [43,48] reported a telementoring system where the mentor did not use real tools and instead controlled the virtual tools appearing on the mentee's view using a haptic device (Touch, 3D Systems). Similarly, Long et al [41] used an AR HMD device that tracked the mentor's hand movements. These movements were then translated to manipulate the virtual tool instructing the mentee. Following the approaches for hand gestures and virtual pointers, Jarc et al [21,53] enabled the mentors to manipulate the ghost tools using custom input devices and game controllers. The virtual surgical tool models enabled the mentor to convey intricate tool wrist orientation information. Overall, surgical tools enhance the instructional experience by replicating complex motions during different surgery phases and overcoming communication barriers between the remote mentor and mentee. Despite the considerable benefits, the use of surgical tools as dynamic AR cues for telementoring in MIS is still in the early stages of clinical trial.

Evaluation of Technology (Research Question 3)

The telementoring technologies using dynamic AR cues were evaluated with participants having varying expertise in MIS. Majority of the studies had experienced surgeons as mentors [13,21,41,45,47,51-53], whereas medical students [6,13,18,21,44,54] and surgical trainees [40,47,50,53] frequently formed the mentee group. The evaluation studies were identified as having either descriptive or analytical study designs. The descriptive studies (n=7 articles) presented the usability or feasibility of the proposed dynamic AR cue, whereas analytical studies (n=15 articles) used comparators to measure effectiveness. Such analytical studies compared outcomes of telementoring using dynamic AR cue against performance with no cue (independent performance without mentoring), prerecorded visual cue (video guidance), in-person guidance, audio cue, or static AR cue (2D telestration). Audio cue [6,13,18,40,47,49,50] and in-person guidance [20,42-44,46,54] were the most frequently used comparators.

The telementoring studies assessed the effectiveness of dynamic AR cues using a range of outcome measures. These were categorized broadly into factors based on technical evaluation and some of the items specified under the MERSQI outcomes domain [38]. Accordingly, the factors measured included technical, perceptions, skills, and patient outcomes. Technical factors (n=4 articles) included assessment of information latency between mentor and mentee locations [22,41,48,52]. Perceptions (n=5 articles) consisted of preferences, subjective performance assessment, user ratings, and workload measured using NASA (National Aeronautics and Space Administration) Task Load Index (NASA-TLX) [13,21,40,45,53]. Skills (n=13 articles) encompassed Critical View of Safety (CVS), economy of movement, fundamentals of laparoscopic surgery (FLS), gaze behavior, Global Operative Assessment of Laparoscopic Skills (GOALS), number of errors, operating time, Objective Structured Assessment of Technical Skills (OSATS), and turn frequency measuring communication skills

XSL•FO
RenderX

[6,13,18,20,40,42-44,47,49-51,54]. Patient outcomes (n=1 article) included hemoglobin level, urinary retention, complications, and anesthesia use in a real surgery setting [46].

Discussion

One of the aims of the scoping review was to present a comprehensive summary of the use of dynamic AR cues for telementoring during MIS. Through systematic searches of the scientific literature, it was observed that dynamic AR cues for telementoring were more frequently implemented for laparoscopic surgeries (as compared with robot-assisted surgeries). It was mainly used for teaching suturing skills on synthetic phantoms and remote mentoring during simulated cholecystectomy procedures. The mentors primarily consisted of expert surgeons, while medical students and surgical trainees dominated the mentee group. The use of hand gestures and surgical tools as dynamic AR cues were more frequent as compared with virtual pointers. Hand gestures and surgical tools allowed the mentor to convey information such as position, orientation, and tool tip motions to the mentee. To display the dynamic AR cues to the mentee, a small proportion of the studies used iPad tablets (n=1, 5%) Microsoft HoloLens (n=3, 14%), and the surgeon's console with stereoscopic vision for robot-assisted surgery (n=4, 19%). However, most of the articles used a standard visualization screen (n=13, 62%) through which the mentee was able to see the AR cues. Among the included articles, at least 14 (67%) of them had a study design where the use of dynamic AR cue was compared with different modes of instruction with the most common comparators being audio cues and in-person guidance. The key factor used for evaluation was skills. The most frequently reported outcome measures were time taken for task completion and number of errors.

In general, the results of the included papers suggest that participants subjected to telementoring with dynamic AR cues experienced considerable beneficial effects when compared with groups receiving no cue, prerecorded visual cue, in-person guidance, audio cue, or static AR cue. More specifically, dynamic AR cues resulted in significant improvement of FLS score when compared with no cue [44] and lowered error count (P<.05) when compared with prerecorded visual cues [43]. While comparing against in-person guidance, some studies reported longer operating time [42], increased anesthesia use, and cauterization [46] with the use of dynamic AR cues. At the same time, few others noted a reduced complication percentage [20], lower operating time (P=.01), steeper learning curve, and fewer failed attempts [54]. Furthermore, several other studies found no significant difference in performance [42,44], number of errors [54], operating time [43,46], and clinical outcomes [46]. This indicates that telementoring using dynamic AR cues was similar to in-person guidance. While comparing with audio cues, some studies reported no significant difference in CVS, GOALS score [40], number of errors [50], and operating time [6,40,50]. However, the majority of studies demonstrated significant advantages of dynamic AR cues over audio cues. These were in terms of subjective performance [40,47,50], user ratings [40], NASA-TLX score (P<.022) [13], economy of movement [47,50], error rates [13,18,47], gaze behavior (P<.01) [18], GOALS score [6,13], operating time (P<.001) [13], OSATS score (P<.01) [6,13,18], turn frequency (P<.0001) [49], and complications [6,13]. These results show that dynamic AR cues were comparable, and significantly superior to telementoring with audio cues alone. In addition, a study by Jarc et al [53] noted that users preferred dynamic AR cues over static 2D telestrations. Overall, dynamic AR cues resulted in superior outcomes, and in some cases, they were considered on par with conventional methods such as in-person guidance and audio cues.

The extent to which the above-mentioned findings may be used to determine the benefits of the technology is limited. This is because majority of the analytical studies with a comparator were conducted on synthetic phantoms, mostly assessing suturing [42-44,54]. Only 2 of the analytical studies were conducted in real patient settings comparing dynamic AR cue with audio during cholecystectomy [40] and in-person guidance during aquablation [46]. In addition, the outcome measures used to assess the different dynamic AR cues were not consistently used across the studies. The effectiveness of hand gestures was primarily assessed through OSATS, pointers through subjective performance, and surgical tools through latency and operating time. As a result, it is difficult to objectively compare between the different dynamic AR cues and determine the optimal one. Nevertheless, the articles published in the past few years reveal an upward trend in the use of hand gestures and surgical tools (Figure 5). This strongly indicates a rising interest in the use of dynamic AR cues for telementoring in MIS as well as the relevance of using hand gestures and surgical tools.

The implementation of telementoring using dynamic AR cues requires consideration of several technical factors. The Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) issued a set of guidelines that telementoring systems are recommended to comply with. SAGES recommends that a minimum bandwidth of 40 megabits per second must be maintained. In addition, the latency (delay in sending mentee's operative view to the remote mentor) must fall below 450 milliseconds [57]. One of the methods used to reduce latency is by encoding the video frames at the mentee's location and decoding it at the mentor's side. However, during this process, there is a risk of lowered video quality experienced by the mentor. To measure the extent of visual degradation experienced by the mentor, metrics such as structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), and mean square error (MSE) are often used [58,59]. SSIM is considered more effective since it is more closely related to the visual quality perceived by the human eye, as compared with PSNR and MSE [22]. The delay experienced by the mentee in receiving mentor's instructions can be reduced when point cloud data (depicting hand gestures or surgical tool) is used instead of video data. Harnessing 5G and future 6G capabilities that provide extremely low latencies can ensure little to no disruption [23]. While several studies included in the review reported information latency between mentor and mentee, they were mostly conducted on synthetic phantoms [22,41,48]. Further research presenting technical results during actual live MIS is imperative.

XSL•FO
RenderX

Figure 5. Number of articles using dynamic augmented reality cues in minimally invasive surgical telementoring.

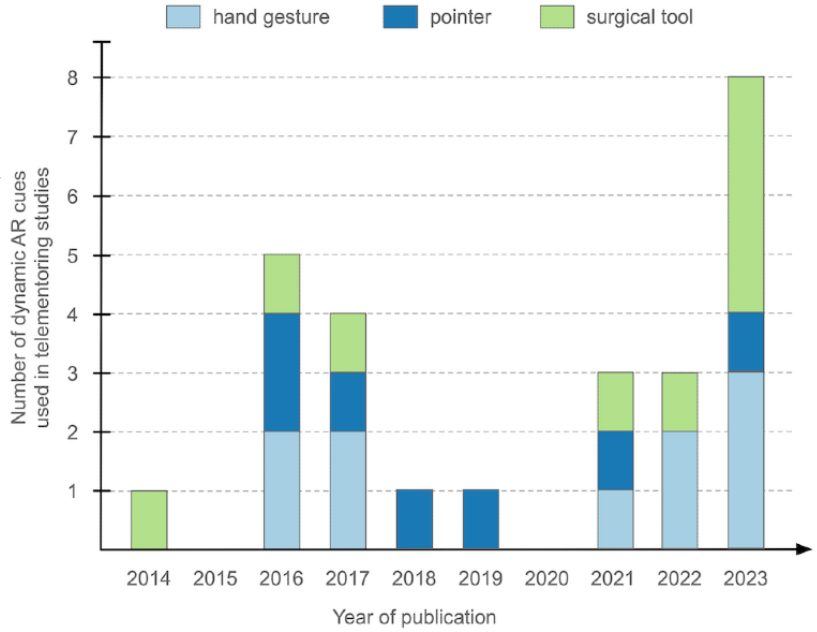

Ethical, safety, and legal concerns must be addressed while implementing telementoring systems. Informed patient consent must be ensured [23]. Every effort should be made to minimize cybersecurity risks and ensure compliance with guidelines on telemedicine issued by HIPAA (Health Insurance Portability and Accountability Act) [31]. The computer systems at both the mentor and mentee locations must be secure with monitoring schemes in place to detect and prevent potential breaches of sensitive patient data. To protect the data while being transferred between the mentor and mentee, appropriate data encryption and virtual private networks must be used [57]. In addition, a clearly outlined process for licensure and credentialing of surgeons is also required for remotely located mentors [8]. Generally, the mentee (who will be a competent operating surgeon) holds primary medical responsibility and is legally liable. Nonetheless, the SAGES states that during telementoring, the remote mentor retains shared responsibility for the patient [60]. Consequently, the mentor must hold the necessary medical license and surgical privileges in the region where the surgery is taking place. This becomes a barrier to adopting telementoring nationally in places like the United States and Canada as licensing is usually at the state or provincial level. However, this would not be a problem in the European Union (EU) where surgeons licensed under one EU member state can practice in other EU states as well [22]. Furthermore, licensing requirements would not be an issue when dynamic AR cues are used to enhance in-person guidance.

Telementoring using dynamic AR cues addresses the need for efficient ways to remotely guide an operating surgeon. This can be potentially replaced with telesurgery where the remote surgeon performs the surgery through robot-assisted surgical systems. While successful telesurgery cases were reported since the early 2000s, significant progress was made only in the recent years owing to the advances in surgical robot systems as well as telecommunication technologies [61]. However, several factors prevent widespread adoption of telesurgery. Although 5G networks are available, 3G or 4G is the standard in most countries leading to major concerns about latency. In addition, more research is needed to assess the cost-effectiveness of telesurgical services. A long-term sustainable approach must be explored for implementation and maintenance. Nonetheless, several cases of telesurgery for surgical specialties such as general, urology, and neurosurgery have been reported [62]. This provides hope for using remote technologies to reduce the global burden of death and disease that can be prevented or treated through surgical care.

This review is limited by the inclusion of articles published in English only, making it less comprehensive. A risk of bias assessment of the included studies was not carried out. A comparative analysis between the different dynamic AR cues was not conducted due to the variety in outcome measures reported. Instead, overall summary potentially guiding future research on dynamic AR cues used for MIS telementoring is presented.

Although further research is needed to establish its benefits over conventional mentoring, the use of dynamic AR cues in MIS telementoring proves to be promising. Future developments may focus on evaluating the technology through multi-institutional randomized controlled trials. In addition to measures of perceptions and skills, objective clinical outcomes that reliably assess the effectiveness of telementoring should also be evaluated.

Acknowledgments

We would like to thank the Clinical Academic Department for their support and guidance. Research reported in this publication was supported by the Qatar Research, Development and Innovation (QRDI) Council Academic Research Grant (ARG) award

XSL•FO
RenderX

(ARG01-0430-230047). All opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of our sponsors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used for PubMed and other databases.
[DOCX File , 17 KB - jmir_v27i1e63939_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
[DOCX File , 85 KB - jmir_v27i1e63939_app2.docx]

References

1. Fernandes RD, Ghasroddashti A, Sorefan-Mangou F, Williams E, Choi K, Fasola L, et al. Educational effectiveness of telementoring as a continuing professional development intervention for surgeons in practice: a systematic review. Ann Surg Open 2023;4(4):e341 [FREE Full text] [doi: 10.1097/AS9.0000000000000341] [Medline: 38144497]
2. Netzer I, Kirkpatrick AW, Nissan M, McKee JL, McBeth P, Dobron A, et al. Rubrum coelis: the contribution of real-time telementoring in acute trauma scenarios-a randomized controlled trial. Telemed J E Health 2019;25(11):1108-1114. [doi: 10.1089/tmj.2018.0173] [Medline: 30707651]
3. Sparn MB, Teixeira H, Chatziisaak D, Schmied B, Hahnloser D, Bischofberger S. Virtual reality simulation training in laparoscopic surgery - does it really matter, what simulator to use? Results of a cross-sectional study. BMC Med Educ 2024;24(1):589 [FREE Full text] [doi: 10.1186/s12909-024-05574-0] [Medline: 38807093]
4. Wagenaar S, Nederhoed JH, Hoksbergen AW, Bonjer HJ, Wisselink W, van Ramshorst GH. Minimally invasive, laparoscopic, and robotic-assisted techniques versus open techniques for kidney transplant recipients: a systematic review. Eur Urol 2017;72(2):205-217. [doi: 10.1016/j.eururo.2017.02.020] [Medline: 28262412]
5. Chang SY, Chan KS, Oo AM. Can computerized simulation be used to assess surgical proficiency in laparoscopic colorectal surgeries? A systematic review. Surg Innov 2024;31(2):195-211. [doi: 10.1177/15533506241232791] [Medline: 38373603]
6. Cizmic A, Müller F, Wise PA, Häberle F, Gabel F, Kowalewski KF, et al. Telestration with augmented reality improves the performance of the first ten ex vivo porcine laparoscopic cholecystectomies: a randomized controlled study. Surg Endosc 2023;37(10):7839-7848 [FREE Full text] [doi: 10.1007/s00464-023-10360-y] [Medline: 37612445]
7. Rollan L, Gigena C, Diaz Saubidet I, Valverde S, Bellia-Munzon G, Millán C. Innovation and challenges: minimally invasive surgery training in Latin America. Cir Pediatr 2024;37(4):147-151 [FREE Full text] [doi: 10.54847/cp.2024.04.10] [Medline: 39506407]
8. SAGES Digital Surgery Working Group, Ali JT, Yang G, Green CA, Reed BL, Madani A, et al. Defining digital surgery: a SAGES white paper. Surg Endosc 2024;38(2):475-487. [doi: 10.1007/s00464-023-10551-7] [Medline: 38180541]
9. Rafiq A, Moore JA, Zhao X, Doarn CR, Merrell RC. Digital video capture and synchronous consultation in open surgery. Ann Surg 2004;239(4):567-573. [doi: 10.1097/01.sla.0000118749.24645.45] [Medline: 15024319]
10. Eadie LH, Seifalian AM, Davidson BR. Telemedicine in surgery. Br J Surg 2003;90(6):647-658. [doi: 10.1002/bjs.4168] [Medline: 12808611]
11. Erridge S, Yeung DKT, Patel HRH, Purkayastha S. Telementoring of surgeons: a systematic review. Surg Innov 2019;26(1):95-111 [FREE Full text] [doi: 10.1177/1553350618813250] [Medline: 30465477]
12. Gasques D, Johnson JG, Sharkey T, Feng Y, Wang R, Xu ZR, et al. ARTEMIS: a collaborative mixed-reality system for immersive surgical telementoring. : Association for Computing Machinery; 2021 Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 8 - 13 May 2021; Yokohama, Japan p. 1-14. [doi: 10.1145/3411764.3445576]
13. Wild C, Lang F, Gerhäuser AS, Schmidt MW, Kowalewski KF, Petersen J, et al. Telestration with augmented reality for visual presentation of intraoperative target structures in minimally invasive surgery: a randomized controlled study. Surg Endosc 2022;36(10):7453-7461 [FREE Full text] [doi: 10.1007/s00464-022-09158-1] [Medline: 35266048]
14. Lin Z, Xu Z, Liu H, Wang X, Ai X, Zhou C, et al. Robotic manipulator‐assisted omnidirectional augmented reality for endoluminal intervention telepresence. Advanced Intelligent Systems 2024;6(1). [doi: 10.1002/aisy.202470002]
15. Dinh A, Yin AL, Estrin D, Greenwald P, Fortenko A. Augmented reality in real-time telemedicine and telementoring: scoping review. JMIR Mhealth Uhealth 2023;11:e45464 [FREE Full text] [doi: 10.2196/45464] [Medline: 37071458]
16. Collins JW, Chen J, Hung A. Telementoring for minimally invasive surgery. Robotic Surgery 2021:361-378. [doi: 10.1007/978-3-030-53594-0_22]
17. Budrionis A, Augestad KM, Patel HR, Bellika JG. An evaluation framework for defining the contributions of telestration in surgical telementoring. Interact J Med Res 2013;2(2):e14 [FREE Full text] [doi: 10.2196/ijmr.2611] [Medline: 23887078]

XSL•FO
RenderX

18. Felinska EA, Fuchs TE, Kogkas A, Chen ZW, Otto B, Kowalewski KF, et al. Telestration with augmented reality improves surgical performance through gaze guidance. Surg Endosc 2023;37(5):3557-3566 [FREE Full text] [doi: 10.1007/s00464-022-09859-7] [Medline: 36609924]
19. Bilgic E, Turkdogan S, Watanabe Y, Madani A, Landry T, Lavigne D, et al. Effectiveness of telementoring in surgery compared with on-site mentoring: a systematic review. Surg Innov 2017;24(4):379-385. [doi: 10.1177/1553350617708725] [Medline: 28494684]
20. Shabir D, Padhan J, Mohammed A, Al-Ansari A, Tsiamyrtzis P, Navkar NV, et al. Telementoring system assessment integrated with laparoscopic surgical simulators. : IEEE; 2023 Presented at: 2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS); 2023 June 02-04; Tainan, Taiwan.
21. Jarc AM, Stanley AA, Clifford T, Gill IS, Hung AJ. Proctors exploit three-dimensional ghost tools during clinical-like training scenarios: a preliminary study. World J Urol 2017;35(6):957-965. [doi: 10.1007/s00345-016-1944-x] [Medline: 27671899]
22. Shabir D, Abdurahiman N, Padhan J, Anbatawi M, Trinh M, Balakrishnan S, et al. Preliminary design and evaluation of a remote tele-mentoring system for minimally invasive surgery. Surg Endosc 2022;36(5):3663-3674 [FREE Full text] [doi: 10.1007/s00464-022-09164-3] [Medline: 35246742]
23. Nickel F, Cizmic A, Chand M. Telestration and augmented reality in minimally invasive surgery: an invaluable tool in the age of COVID-19 for remote proctoring and telementoring. JAMA Surg 2022;157(2):169-170. [doi: 10.1001/jamasurg.2021.3604] [Medline: 34705030]
24. Bui DT, Barnett T, Hoang H, Chinthammit W. Tele-mentoring using augmented reality technology in healthcare: a systematic review. AJET 2021;37(4):68-88. [doi: 10.14742/ajet.6243]
25. Huang EY, Knight S, Guetter CR, Davis CH, Moller M, Slama E, et al. Telemedicine and telementoring in the surgical specialties: a narrative review. Am J Surg 2019;218(4):760-766. [doi: 10.1016/j.amjsurg.2019.07.018] [Medline: 31350010]
26. Jin ML, Brown MM, Patwa D, Nirmalan A, Edwards PA. Telemedicine, telementoring, and telesurgery for surgical practices. Curr Probl Surg 2021;58(12):100986. [doi: 10.1016/j.cpsurg.2021.100986] [Medline: 34895561]
27. Gerardo R, Lele P, Sundaram K, Ponsky T. Surgical telementoring: feasibility, applicability, and how to. J Surg Oncol 2021;124(2):241-245. [doi: 10.1002/jso.26511] [Medline: 34245573]
28. A Butt K, Augestad KM. Educational value of surgical telementoring. J Surg Oncol 2021;124(2):231-240 [FREE Full text] [doi: 10.1002/jso.26524] [Medline: 34245572]
29. Pfennig M, Lee A, Mi M. How does telementoring impact medical education within the surgical field? A scoping review. Am J Surg 2022;224(3):869-880 [FREE Full text] [doi: 10.1016/j.amjsurg.2022.04.038] [Medline: 35545476]
30. Hamza H, Al-Ansari A, Navkar NV. Technologies used for telementoring in open surgery: a scoping review. Telemed J E Health 2024;30(7):1810-1824. [doi: 10.1089/tmj.2023.0669] [Medline: 38546446]
31. Yang A, Kim D, Hwang PH, Lechner M. Telemedicine and telementoring in rhinology, otology, and laryngology: a scoping review. OTO Open 2022;6(1):2473974X211072791 [FREE Full text] [doi: 10.1177/2473974X211072791] [Medline: 35274073]
32. Seeliger B, Collins JW, Porpiglia F, Marescaux J. The role of virtual reality, telesurgery, and teleproctoring in robotic surgery. In: Robotic Urologic Surgery. Heidelberger Platz: Springer International Publishing; 2022:61-77.
33. Ayoub CH, El-Asmar JM, Abdulfattah S, El-Hajj A. Telemedicine and telementoring in urology: a glimpse of the past and a leap into the future. Front Surg 2022;9:811749 [FREE Full text] [doi: 10.3389/fsurg.2022.811749] [Medline: 35273996]
34. Augestad KM, Han H, Paige J, Ponsky T, Schlachta CM, Dunkin B, et al. Educational implications for surgical telementoring: a current review with recommendations for future practice, policy, and research. Surg Endosc 2017;31(10):3836-3846. [doi: 10.1007/s00464-017-5690-y] [Medline: 28656341]
35. Hung AJ, Chen J, Shah A, Gill IS. Telementoring and telesurgery for minimally invasive procedures. J Urol 2018;199(2):355-369. [doi: 10.1016/j.juro.2017.06.082] [Medline: 28655529]
36. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
37. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
38. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA 2007;298(9):1002-1009. [doi: 10.1001/jama.298.9.1002] [Medline: 17785645]
39. Glossop SC, Bhachoo H, Murray TM, Cherif RA, Helo JY, Morgan E, et al. Undergraduate teaching of surgical skills in the UK: systematic review. BJS Open 2023;7(5):zrad083 [FREE Full text] [doi: 10.1093/bjsopen/zrad083] [Medline: 37819804]
40. Huettl F, Heinrich F, Boedecker C, Vradelis L, Ludt A, Kneist W, et al. Real-time augmented reality annotation for surgical education during laparoscopic surgery: results from a single-center randomized controlled trial and future aspects. J Am Coll Surg 2023;237(2):292-300. [doi: 10.1097/XCS.0000000000000712] [Medline: 37042553]

XSL•FO
RenderX

41. Long Y, Li C, Dou Q. Robotic surgery remote mentoring via AR with 3D scene streaming and hand interaction. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 2023;11(4):1027-1032. [doi: 10.1080/21681163.2022.2145498]
42. Shabir D, Balakrishnan S, Padhan J, Yaacoub E, Mohammed A, Deng Z, et al. Evaluating a remote tele-mentoring gameplay setup for teaching laparoscopic suturing skills. : IEEE; 2023 Presented at: 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH); 2023 August 28-30; Athens, Greece.
43. Shabir D, Balakrishnan S, Padhan J, Abinahed J, Yaacoub E, Mohammed A, et al. Tele-mentoring using augmented reality: a feasibility study to assess teaching of laparoscopic suturing skills. : IEEE; 2023 Presented at: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS); 2023 June 22-24; L'Aquila, Italy. [doi: 10.1109/cbms58004.2023.00193]
44. Lowry B, Johnson GGRJ, Vergis A. Merged virtual reality teaching of the fundamentals of laparoscopic surgery: a randomized controlled trial. Surg Endosc 2022;36(9):6368-6376 [FREE Full text] [doi: 10.1007/s00464-021-08939-4] [Medline: 34981231]
45. Youssef SC, Sabbubeh B, Haram K, Noël J, Aydın A, Challacombe B, et al. Augmented reality robot-assisted radical prostatectomy with PROXIMIE: preliminary clinical experience. Urology Video Journal 2022;16:100187. [doi: 10.1016/j.urolvj.2022.100187]
46. El-Asmar JM, Labban M, El-Hajj A. Integration of aquablation through telemetry: an alternative to onsite proctoring? World J Urol 2021;39(9):3473-3479 [FREE Full text] [doi: 10.1007/s00345-021-03603-x] [Medline: 33547926]
47. Heinrich F, Huettl F, Schmidt G, Paschold M, Kneist W, Huber T, et al. HoloPointer: a virtual augmented reality pointer for laparoscopic surgery training. Int J Comput Assist Radiol Surg 2021;16(1):161-168 [FREE Full text] [doi: 10.1007/s11548-020-02272-2] [Medline: 33095424]
48. Shabir D, Abdurahiman N, Padhan J, Trinh M, Balakrishnan S, Kurer M, et al. Towards development of a tele-mentoring framework for minimally invasive surgeries. Int J Med Robot 2021;17(5):e2305. [doi: 10.1002/rcs.2305] [Medline: 34256415]
49. Feng Y, Li K, Semsar A, Mcgowan H, Mun J, Zahiri HR, et al. Communication cost of single-user gesturing tool in laparoscopic surgical training. 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; 4 - 9 May 2019; Glasgow Scotland Uk. [doi: 10.1145/3290605.3300841]
50. Feng Y, McGowan H, Semsar A, Zahiri HR, George IM, Turner T, et al. A virtual pointer to support the adoption of professional vision in laparoscopic training. Int J Comput Assist Radiol Surg 2018;13(9):1463-1472. [doi: 10.1007/s11548-018-1792-9] [Medline: 29796835]
51. Kowalewski KF, Hendrie JD, Schmidt MW, Garrow CR, Bruckner T, Proctor T, et al. Development and validation of a sensor- and expert model-based training system for laparoscopic surgery: the iSurgeon. Surg Endosc 2017;31(5):2155-2165. [doi: 10.1007/s00464-016-5213-2] [Medline: 27604368]
52. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
53. Jarc AM, Shah SH, Adebar T, Hwang E, Aron M, Gill IS, et al. Beyond 2D telestration: an evaluation of novel proctoring tools for robot-assisted minimally invasive surgery. J Robot Surg 2016;10(2):103-109 [FREE Full text] [doi: 10.1007/s11701-016-0564-1] [Medline: 26914650]
54. Vera AM, Russo M, Mohsin A, Tsuda S. Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. Surg Endosc 2014;28(12):3467-3472. [doi: 10.1007/s00464-014-3625-4] [Medline: 24962856]
55. Syamlan A, Fathurachman, Denis K, Vander Poorten E, Pramujati B, Tjahjowidodo T. Haptic/virtual reality orthopedic surgical simulators: a literature review. Virtual Reality 2022;26(4):1795-1825. [doi: 10.1007/s10055-022-00666-y]
56. Higgins M, Leung S, Radacsi N. 3D printing surgical phantoms and their role in the visualization of medical procedures. Annals of 3D Printed Medicine 2022;6:100057. [doi: 10.1016/j.stlm.2022.100057]
57. Bogen EM, Schlachta CM, Ponsky T. White paper: technology for surgical telementoring-SAGES project 6 technology working group. Surg Endosc 2019;33(3):684-690. [doi: 10.1007/s00464-018-06631-8] [Medline: 30617422]
58. Shabir D, Anbatawi M, Abdurahiman N, Trinh M, Padhan J, Al-Ansari A, et al. Benchmarking network performance of augmented reality based surgical telementoring systems. : IEEE; 2022 Presented at: 2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE); 2022 November 07-09; Taichung, Taiwan. [doi: 10.1109/BIBE55377.2022.00061]
59. Hassan A, Ghafoor M, Tariq SA, Zia T, Ahmad W. High efficiency video coding (HEVC)-based surgical telementoring system using shallow convolutional neural network. J Digit Imaging 2019;32(6):1027-1043 [FREE Full text] [doi: 10.1007/s10278-019-00206-2] [Medline: 30980262]
60. Schlachta CM, Nguyen NT, Ponsky T, Dunkin B. Project 6 summit: SAGES telementoring initiative. Surg Endosc 2016;30(9):3665-3672. [doi: 10.1007/s00464-016-4988-5] [Medline: 27270593]
61. Dohler M, Saikali S, Gamal A, Moschovas MC, Patel V. The crucial role of 5G, 6G, and fiber in robotic telesurgery. J Robot Surg 2024;19(1):4. [doi: 10.1007/s11701-024-02164-6] [Medline: 39549181]

XSL•FO
RenderX

62. Reddy SK, Saikali S, Gamal A, Moschovas MC, Rogers T, Dohler M, et al. Telesurgery a systematic literature review and future directions. Ann Surg 2024. [doi: 10.1097/SLA.0000000000006570] [Medline: 39435556]

Abbreviations

AR: augmented reality
ART: augmented reality telementoring
CVS: Critical View of Safety
EU: European union
FLS: fundamentals of laparoscopic surgery
GOALS: Global Operative Assessment of Laparoscopic Skills
HIPAA: Health Insurance Portability and Accountability Act
HMD: head-mounted display
MERSQI: Medical Education Research Study Quality Instrument
MIS: minimally invasive surgery
MSE: mean square error
NASA-TLX: NASA (National Aeronautics and Space Administration)Task Load Index
OSATS: Objective Structured Assessment of Technical Skills
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PSNR: peak signal-to-noise ratio
RQ: research question
SAGES: Society of American Gastrointestinal and Endoscopic Surgeons
SSIM: structural similarity index measure
VIPAR: Virtual Interactive Presence and Augmented Reality

Edited by T de Azevedo Cardoso; submitted 03.07.24; peer-reviewed by R Li, X Cheng; comments to author 04.11.24; revised version received 21.11.24; accepted 23.12.24; published 03.02.25.

Please cite as:
Hamza H, Aboumarzouk OM, Al-Ansari A, Navkar NV
Dynamic Augmented Reality Cues for Telementoring in Minimally Invasive Surgeries: Scoping Review
J Med Internet Res 2025;27:e63939
URL: https://www.jmir.org/2025/1/e63939
doi:10.2196/63939
PMID:39899360

©Hawa Hamza, Omar M Aboumarzouk, Abdulla Al-Ansari, Nikhil V Navkar. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Application of Patient-Generated Health Data Among Older Adults With Cancer: Scoping Review

Yesol Kim[1,2], PhD; Geonah Kim[3], MSN; Hyeonmi Cho[4,5], PhD; Yeonju Kim[1], MPH; Mona Choi[4,6], PhD

[1]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

[2]College of Nursing, Gyeongsang National University, Jinju, Republic of Korea

[3]Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea

[4]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[5]Research Institute of AI and Nursing Science, College of Nursing, Gachon University, Incheon, Republic of Korea

[6]Yonsei Evidence Based Nursing Centre, A JBI Affiliated Group, Seoul, Republic of Korea

Corresponding Author:
Mona Choi, PhD
Mo-Im Kim Nursing Research Institute
College of Nursing
Yonsei University
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3341
Fax: 82 2 2227 8303
Email: monachoi@yuhs.ac

Abstract

Background: The advancement of information and communication technologies has spurred a growing interest in and increased applications of patient-generated health data (PGHD). In particular, PGHD may be promising for older adults with cancer who have increased survival rates and experience a variety of symptoms.

Objective: This scoping review aimed to identify the characteristics of research on PGHD as applied to older adults with cancer and to assess the current use of PGHD.

Methods: Guided by Arksey and O'Malley as well as the JBI (Joanna Briggs Institute) methodology for scoping reviews, 6 electronic databases were searched: PubMed, Embase, CINAHL, Cochrane Library, Scopus, and Web of Science. In addition, the reference lists of the selected studies were screened to identify gray literature. The researchers independently screened the literature according to the predefined eligibility criteria. Data from the selected studies were extracted, capturing study, participant, and PGHD characteristics.

Results: Of the 1090 identified studies, 88 were selected. The publication trend gradually increased, with a majority of studies published since 2017 (69/88, 78%). Almost half of the studies were conducted in North America (38/88, 43%), followed by Europe (30/88, 34%). The most common setting in which the studies were conducted was the participant's home (69/88, 78%). The treatment status varied; the median sample size was 50 (IQR 33.8-84.0). The devices that were used to measure the PGHD were classified as research-grade wearable devices (57/113, 50.4%), consumer-grade wearable devices (28/113, 24.8%), or smartphones or tablet PCs for mobile apps (23/113, 20.4%). More than half of the studies measured physical activity (69/123, 56.1%), followed by patient-reported outcomes (23/123, 18.7%), vital signs (13/123, 10.6%), and sleep (12/123, 9.8%). The PGHD were mainly collected passively (63/88, 72%), and active collection methods were used from 2015 onward (20/88, 23%). In this review, the stages of PGHD use were classified as follows: (1) identification, monitoring, review, and analysis (88/88, 100%); (2) feedback and reporting (32/88, 39%); (3) motivation (30/88, 34%); and (4) education and coaching (19/88, 22%).

Conclusions: This scoping review provides a comprehensive summary of the overall characteristics and use stages of PGHD in older adults with various types and stages of cancer. Future research should emphasize the use of PGHD, which interacts with patients to provide patient-centered care through patient engagement. By enhancing symptom monitoring, enabling timely interventions, and promoting patient involvement, PGHD have the potential to improve the well-being of older adults with cancer,

XSL•FO
RenderX

contributing to better health management and quality of life. Therefore, our findings may provide valuable insights into PGHD that health care providers and researchers can use for geriatric cancer care.

Trial Registration: Open Science Framework Registry OSF.IO/FZRD5; https://doi.org/10.17605/OSF.IO/FZRD5

(J Med Internet Res 2025;27:e57379) doi:10.2196/57379

KEYWORDS

patient-generated health data; wearable devices; patient-reported outcomes; patient-centered care; older adults; cancer; scoping review

Introduction

Background

Advancements in early detection and treatment modalities have led to a consistent increase in the survival rates of patients diagnosed with cancer. In the United States, the number of survivors of cancer reached 18.1 million in 2022, and this figure is projected to increase to 26 million by 2040 [1]. With rising cancer survival rates, there is a notable trend of aging among survivors of cancer. In 2022, over two-thirds (67%) of US survivors of cancer were aged 65 years or older [1].

Older adults often experience multiple chronic health conditions, and survivors of cancer show an even higher prevalence of these issues compared to those without cancer [2]. Common co-occurring conditions in older patients with cancer include musculoskeletal conditions, such as osteoarthritis and sciatica; mental health disorders, such as depression and anxiety; and cardiovascular diseases, such as hypertension and coronary artery disease [2,3]. These conditions may exacerbate susceptibility to functional decline, with older survivors of cancer facing an accelerated deterioration in grip strength, gait speed, and overall physical performance compared to their peers without cancer [4]. Moreover, the presence of multiple chronic conditions and functional limitations in older survivors of cancer are associated with a reduced quality of life, increased risk of early mortality, and significant economic burden [5,6].

Given the challenges associated with both cancer and aging, innovative approaches to monitoring and managing the health of older survivors of cancer are essential for improving their health and well-being outcomes. The collection and sharing of pertinent health data can offer survivors of cancer and their health care providers an opportunity to enhance their health [7]. The rise of consumer technologies, such as smartphone apps and wearable devices, has paved the way for the creation and use of tools that enable individuals to gather their own health-related data [8]. These data are related to their physical behavior (eg, step count or sleep patterns), physiological parameters (eg, heart rate or blood pressure), biochemical markers (eg, blood glucose levels), and environmental factors (eg, air quality or temperature) in which they are present.

With the advancement of information and communication technologies, the interest in and the applications of patient-generated health data (PGHD) are growing. PGHD refer to “health-related data—including health history, symptoms, biometric data, treatment history, lifestyle choices, and other information—created, recorded, gathered, or inferred by or from patients or their designees (ie, care partners or those who assist them)” [9]. PGHD are distinct from data sourced from clinical environments or interactions with health care providers. The primary distinction is that PGHD are recorded by patients and not providers, and their sharing is directed by the patients [9]. Patients can collect these data either passively or actively through mobile health apps, wearable devices (eg, fitness bands), medical devices (eg, blood glucose monitors), or survey instruments on mobile devices [7]. Passive data collection occurs when data are gathered automatically with minimal patient input, typically using automated systems or sensors. By contrast, active data collection requires patients to be actively involved, typically by responding to surveys.

PGHD enable the consistent, long-term monitoring of diverse health-related information that can be easily exchanged between patients and health care providers [10]. These data can complement the information that health care systems typically gather through providers, offering individualized baseline indicators. With PGHD, both patients and health care providers can quickly discern changes in the health status, allowing for more appropriate and prompt interventions [11].

PGHD have found applications in monitoring and managing various long-term health conditions. For instance, PGHD can be used to identify undiagnosed obesity and diabetes, assess the risk of these conditions, and predict glycemic events in patients with diabetes [12]. In the context of HIV care, PGHD help alleviate financial and temporal pressures by reducing the need for frequent in-person consultations [11]. Continuous monitoring through PGHD enhances medication adherence and increases patients’ understanding of their health status [11]. For survivors of cancer, who face higher posttreatment health risks, PGHD enable real-time tracking of health parameters such as activity levels, symptom progression, and medication adherence, facilitating early detection of issues and more personalized interventions [7]. Beyond individual care, PGHD can also benefit larger survivors of cancer communities by contributing to population-level insights, informing public health strategies, enhancing survivorship guidelines, and supporting research on cancer survivorship.

While there is the potential for using PGHD in older adults with cancer, few review studies have focused on PGHD in these patients. A comprehensive understanding of the characteristics and use of PGHD in older adults with cancer will help improve patient engagement in health care and enable researchers to apply PGHD in geriatric cancer care. This will ultimately lead to the effective application of PGHD in enhancing the health and well-being of older adults with cancer.

XSL•FO
RenderX

Objectives

This scoping review aimed to identify the characteristics of research on PGHD as applied to older adults with cancer and to assess the current use of PGHD. The main research question for the scoping review was "What is the present landscape of research concerning PGHD in older adults with cancer?"

The specific research questions were as follows: (1) What are the characteristics of the studies conducted on PGHD in older adults with cancer? (2) What are the characteristics of older adults with cancer who have been the subject of PGHD-related research? and (3) What are the characteristics of PGHD that have been gathered from older adults with cancer?

Methods

Design

This scoping review was conducted following the guidelines of Arksey and O'Malley [13] and the JBI (Joanna Briggs Institute) methodology for scoping reviews [14]. This scoping review was also performed according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Review) guidelines Multimedia Appendix 1 [15]. The population, concept, and context for this scoping review were as follows: (1) population—older adults with cancer, (2) concept—PGHD, which were collected using devices; and (3) context—no limitation for setting. The protocol of the scoping review was registered in the Open Science Framework Registry.

Search Strategy

In this study, 6 electronic databases—PubMed, Embase, CINAHL, Cochrane Library, Scopus, and Web of Science—were searched for relevant literature. Considering the population, concept, and context of this scoping review, the search strategy was developed by combining keywords and index terms from each database (Multimedia Appendix 2). PGHD as a concept was introduced as a Medical Subject Headings (MeSH) term by PubMed in 2018. To retrieve relevant literature before the introduction of PGHD as a MeSH term, we further constructed PGHD-specific search terms by combining the keywords "patient-generated" and "devices." In addition, the search was limited to the index terms included in the titles and abstracts of the studies to ensure proper identification of the literature. An experienced librarian at the researcher's university confirmed this search strategy. A literature search was conducted using a confirmed search strategy on October 21, 2022. After the literature search, the reference lists of the selected studies were screened to identify the gray literature.

Eligibility Criteria and Study Selection

To assess the eligibility of studies on PGHD in older adults with cancer, the inclusion criteria were as follows: (1) studies that included older adults with a mean or median age of 65 years or older; (2) studies that included older adults with cancer, regardless of the type of cancer or treatment status; and (3) studies that applied PGHD, including wearable devices that measure patients' physical activity or physiological features or mobile phones that measure patient-reported outcomes (PROs). The exclusion criteria were as follows: (1) studies in which PGHD were selectively used in some of the participants; (2) studies comparing results using PGHD for validation of other tools; (3) publications other than original studies, such as editorials, letters, or protocols; (4) publications that were not peer reviewed, such as abstracts, conference proceedings, dissertations, or theses; and (5) studies written in languages other than English or Korean.

The studies identified in the literature search were exported to the reference management software program (EndNote X9; Clarivate Analytics). After deleting duplicate results using EndNote, the results were extracted with a Microsoft Excel program (Microsoft Corp) for literature screening. Two researchers (Yesol K and GK) independently selected studies based on the inclusion and exclusion criteria. The titles and abstracts were initially reviewed; subsequently, the full texts were reviewed to select studies for final inclusion. If the two researchers did not agree on a particular study, a consensus was reached through discussion and the participation of a third researcher (MC).

Data Extraction

Data from the selected studies were extracted using Microsoft Excel in a structured data charting format, which was constructed by the researchers. It comprised the study, participant, and PGHD characteristics. First, the study characteristics included published year, country, continent, study design, and dataset. Second, the participant characteristics included the study setting, type of cancer, treatment status, and sample size. Lastly, the PGHD characteristics included the type of device that was used to measure PGHD, type of assessment by PGHD, on-body locations of the wearable devices that measure PGHD, duration of wear of the devices that measure PGHD, type of PGHD collection, and type of PGHD use. The types of PGHD collection were categorized as passive (using sensor-based devices that automatically acquire and transfer PGHD from patients), active (requiring active patient engagement to collect PGHD that are not easily collected automatically), and combined methods [16]. The types of PGHD use were categorized into the following four categories: (1) identification, monitoring, review, and analysis; (2) feedback and reporting; (3) motivation; and (4) education and coaching [16]. Two researchers (Yesol K and GK) independently extracted the data, and disagreements, if any, were resolved by consensus through discussion and the participation of a third researcher (MC).

Data Analysis

The data extracted from this scoping review are summarized and presented using descriptive statistics. The results were mapped and are presented in the form of tables, charts, or figures according to the purpose and research questions of this scoping review.

Results

Study Selection

The literature selection process is illustrated in Figure 1 as a PRISMA (Preferred Reporting Items for Systematic Reviews

XSL•FO
RenderX

and Meta-Analyses) flow diagram. A total of 1090 articles were retrieved from 6 database searches. After the elimination of duplicates, the titles and abstracts of 659 (60.5%) articles were reviewed. Subsequently, the full texts of 367 (33.7%) articles were reviewed to determine eligibility for this review. The reasons for the exclusion of articles from the full-text review are listed in Figure 1. After the screening, 71 (6.5%) articles were selected. A review of the references of the 71 selected articles led to the additional selection of 17 articles, resulting in a selection of a total of 88 articles for the review [17-104].

Figure 1. PRISMA flow diagram of this study. PGHD: patient-generated health data; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

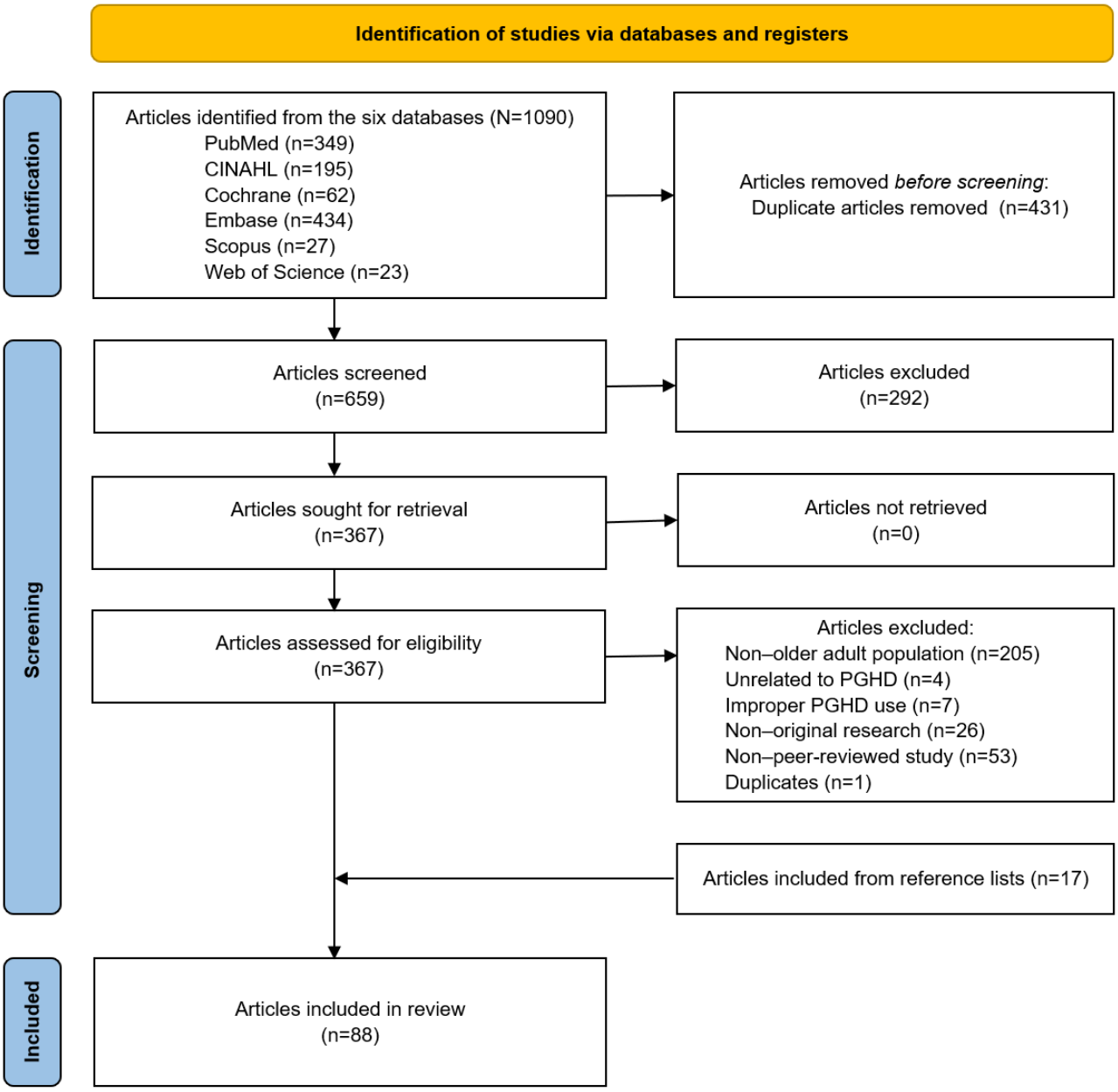

Study Characteristics

The studies included in this scoping review were published between 2006 and 2022 (Multimedia Appendix 3). There has been a discernible upward trend in publication volume, with more than 70% of the articles published since 2017 (69/88, 78%) [36-104]. Geographically, North America emerged as the leading contributor to the body of research, accounting for 43% (38/88) of the studies. The study designs were categorized into nonexperimental and experimental designs. Among studies with nonexperimental designs, a survey-based study design was most frequently adopted (39/88, 44%). Among studies with experimental designs, a pre-experimental design was most frequently adopted (28/88, 32%). Data derived from original research constituted the basis for more than 60% of the studies (59/88, 67%; Table 1).

XSL•FO
RenderX

Table 1. Characteristics of included studies (n=88).

Study characteristics	Studies, n (%)	Reference numbers
Continents		
Asia	8 (9)	[37,48,53,76,82,97,99,102]
Europe	30 (34)	[17,18,20,28,32,33,35,38,40,42,43,45,50,58,61,66,69,71,73,74,78,79,83-85,89,90,96,101,104]
North America	38 (43)	[21-23,29,30,36,41,44,46,47,49,51,54-57,60,62-65,67,68,70,75,77,80,81,86,87,91-95,98,100,103]
South America	1 (1)	[52]
Africa	0 (0)	N/A[a]
Oceania	11 (13)	[19,24-27,31,34,39,59,72,88]
Study design		
Nonexperimental design		
Survey	39 (44)	[17-19,21-23,26-28,30,34,35,37-39,41,44,46-49,51,55,57,59,62,67,68,73,74,76,78,81,83,89,95,97,99,102]
Qualitative study	3 (3)	[33,42,96]
Methodological research	2 (2)	[91,94]
Experimental design		
True experimental design	13 (15)	[25,31,50,56,60,70,82,84,86-88,93,98]
Quasi-experimental design	3 (3)	[43,58,66]
Pre-experimental design	28 (32)	[20,24,29,32,36,40,45,52-54,61,63-65,69,71,72,75,77,79,80,85,90,92,100,101,103,104]
Dataset		
Original research	59 (67)	[17,18,20,24,26-32,37,39,41,43,44,46-50,52-54,57,60,61,63,64,66-70,73-75,77,79-87,90-93,96-103]
Other research	24 (27)	[21,25,33-36,38,40,42,45,51,56,58,59,65,71,72,76,78,88,89,94,95,104]
National data or cohort data	5 (6)	[19,22,23,55,62]

[a]N/A: not applicable.

Participant Characteristics

The participant characteristics are presented in Table 2. Most studies were conducted at the participants' homes (69/88, 78%). Among the included studies, none focused solely on older adults living alone. Only 15 (17%) of the 88 studies distinguished participant characteristics based on whether they lived alone or with others, and most participants in these studies lived with others. Of the 88 studies, 64 (73%) specified the selection criteria for study participants according to cancer type. These studies included 76 cases of various types of cancer, the most common of which was prostate cancer (22/76, 29%). Regarding gender proportions, 66 (75%) of the 88 studies had a higher proportion of men, while 20 (23%) studies had a higher proportion of women. Of these, 18 (20%) studies focused exclusively on men, all involving patients with prostate cancer. In contrast, 3 (3%) studies focused exclusively on women, all involving patients with breast cancer. Half of the studies were conducted on patients who received active treatment (29/88, 33%) or surgery (15/88, 17%); a quarter were conducted on survivors of cancer (22/88, 25%). In this review, survivors of cancer are defined as individuals who have completed their primary cancer treatments—such as surgery, chemotherapy, or radiation therapy—and are in the posttreatment phase. The median sample size was 50 (IQR 33.8-84.0), and more than 40% of the studies included fewer than 50 (39/88, 44%) participants. In the studies included in this review, the mean age of participants ranged from 65 (SD 7) to 76.9 (SD 10.5) years.

XSL•FO
RenderX

Table 2. Characteristics of participants (n=88).

Participant characteristics	Studies, n (%)	Reference numbers
Setting		
Home	69 (78)	[18-27,30-35,37-52,54-65,67,69-72,74,76,78-80,82-84,86-90,92,93,95-97,100,104]
Hospital (outpatient)	3 (3)	[29,36,53]
Hospital (inpatient)	5 (6)	[17,62,73,99,102]
Home and hospital (combined)	11 (13)	[28,68,75,77,81,85,91,94,98,101,103]
Type of cancer[a]		
Prostate	22 (29)	[21-23,25,31,32,34,38,40,42,43,50,51,54,56,60,61,70,82,85,88,92]
Gastrointestinal	21 (28)	[20,23,28,33,35,38,41,45,50,58,64-66,68,71,72,77,78,91,94,101]
Lung	18 (24)	[20,26,27,30,36,37,46,48,49,53,57,75-77,86,94,97,101]
Bladder, urethra, or urothelial	5 (7)	[63,66,83,92,98]
Others	10 (13)	[19,23,59,66,72,79,80,84,100,103]
Gender proportions		
Higher proportion of men	66 (75)	[18,20-22,25-32,34-44,46-48,50,51,53-56,58-61,63-66,68-71,73,74,76-79,82-85,88-92,95-99,101,102]
Higher proportion of women	20 (23)	[17,19,23,24,33,49,52,57,62,67,72,75,80,81,86,87,93,94,100,104]
Equal proportion of men and women	1 (1)	[103]
Not reported	1 (1)	[45]
Treatment status		
Diagnosed with cancer with no limitation of treatment	11 (13)	[26,27,29,44,48,53,59,62,76,79,88]
Received active treatment	29 (33)	[21,28,30,32,33,36,37,39,40,42,43,45,51,52,54,56,64,65,67,78,80,82-85,92,97,103,104]
Underwent surgery	15 (17)	[58,63,66,71,74,75,77,80,81,89-91,94,98,101]
Survivor of cancer	22 (25)	[19,20,22-25,31,34,35,38,41,49,50,55,60,61,72,87,93,95,96,100]
Received end-of-life and palliative care	4 (5)	[47,73,99,102]
Not reported	7 (8)	[17,18,46,57,69,70,86]
Sample size (participants)		
≤49	39 (44)	[21,24,28-30,32,33,39,41,42,45,46,52-54,63,68,71-73,76,77,79,82,83,85,86,88,90,93,95-100,102-104]
50-99	28 (32)	[17,18,20,26,27,37,40,47,48,56-60,64,65,69,74,75,78,80,81,87,89,91,92,94,101]
100-149	11 (13)	[22,23,25,34,43,44,55,61,66,67,84]
≥150	10 (11)	[19,31,35,36,38,49-51,62,70]

[a]There were a total number of 76 cases of various types of cancer among 64 studies.

PGHD Characteristics

The characteristics of the PGHD identified in each study are summarized in Table 3. In this review, research-grade wearable devices were used in more than half of the cases (57/113, 50.4%): these included the ActivPal Monitor (PAL Technologies Ltd), Stepwatch Activity Monitor (Modus Health), and ActiGraph GT3x Activity Monitor. Consumer-grade wearable devices were used in 28 (24.8%) cases: these included Fitbit, Apple Watch, Garmin band, Microsoft Band, Samsung Gear, and Xiaomi MiBand. Smartphones, tablet PCs, and computers were used to collect data using mobile apps or web-based surveys in 25 (22.1%) cases. Other devices, including smart mattresses and an instant camera, were used in 3 (2.7%) cases.

XSL•FO
RenderX

Table 3. Characteristics of patient-generated health data (n=88).

Patient-generated health data characteristics	Studies, n (%)	Reference numbers
Device type[a]		
Research-grade wearable devices	57 (50.4)	[17-28,30,31,34,35,37-39,41,47-51,54-56,59-67,72,73,76,78,82,87,88,93,95,101]
Consumer-grade wearable devices	28 (24.8)	[45,46,57,60,68,70,72,74,77,80,86,87,89-91,93-95,97,98,100,103]
Smartphones, tablet PCs, or computers for mobile apps or surveys	25 (22.1)	[29,32,33,36,40,42-44,52,53,58,69-71,75,79,81,83-85,89,91,92,96,104]
Other devices	3 (2.7)	[99,100,102]
Assessment type[b]		
Physical activity	69 (56.1)	[17-28,31,34,35,38,39,41,45,46,48-50,52,54-57,59-68,70,72,74,76,77,80-82,86-91,93-95,98,100-103]
Sleep	12 (9.8)	[17,18,30,37,47,48,51,57,73,80,91,103]
Vital signs	13 (10.6)	[28,45,63,78,89-91,97,99]
Patient-reported outcomes	23 (18.7)	[29,32,33,36,40,42-44,53,58,69-71,75,79,83-85,89,91,92,96,104]
Others	6 (4.9)	[45,63,78,90,91,100]
On-body location of wearable device[c]		
Wrist	35 (29.9)	[17,18,21,28,30,37,41,45,47,48,51,54,57,61-63,72-74,76-78,80,86,87,89-91,93,94,97,100,103]
Waist	11 (9.4)	[26,27,49,60,67,68,70,82,93,95]
Hip	13 (11.1)	[19,22,23,31,34,38,39,50,52,59,65,72,88]
Lower extremity (leg, ankle, or foot)	8 (6.8)	[20,35,66,67,87,95,101]
Others (neck, chest, arm, finger, skin, or forehead)	13 (11.1)	[18,28,45,55,63,66,78,90,95]
Nonbody (cloth or mattress)	7 (6.0)	[45,46,60,63,98,99,102]
Not admitted (mobile apps or camera)	30 (25.6)	[24,25,29,32,33,36,40-44,53,56,58,64,69-71,75,79,81,83-85,89,91,92,96,100,104]
Wearing period of device[a]		
≤1 week	47 (41.6)	[17-23,26,29,34-39,41,44,46-49,52,53,55,57,59,62,63,66,67,73,76,78,82,87,95,97,100]
>1 week or ≤4 weeks (1 month)	23 (20.4)	[27,28,30,32,50,51,54,61,64,65,68,72,75,88-90,93,94,96]
5-12 weeks (3 months)	25 (22.1)	[24,31,40,42,43,45,54,60,74,77,79,85,86,89-93,98,101,103]
>12 weeks (3 months)	16 (14.2)	[25,33,56,58,69-72,80,81,83,84,87,99,102,104]
Not reported	2 (1.8)	[18,70]
Collection type		
Passive collection	63 (72)	[17-28,30,31,34,35,37-39,41,45-52,54-57,59-68,72-74,76-78,80-82,86-88,93-95,97-99,101-103]
Active collection	20 (23)	[29,32,33,36,40,42-44,53,58,69,71,75,79,83-85,92,96,104]
Combined collection	5 (6)	[38,70,89,91,100]
Collection support by health care providers	20 (23)	[32,36,40,42-44,47,58,74,83,86,87,89,90,92,97,98,101,103,104]

[a]The total number is 113 due to multiple categories.

[b]The total number is 123 due to multiple categories.

[c]The total number is 117 due to multiple categories.

Regarding the type of assessment by PGHD, more than half of the assessments (69/123, 56.1%) focused on physical activity measured through step count, moderate-to-vigorous physical activity, and sedentary behavior. Sleep, including the sleep-wake cycle, circadian rhythm, and movement in bed, were measured in 12 (9.8%) cases. There were 23 (18.7%) cases of PROs, including symptoms and quality of life, and 6 (4.9%) cases of other assessments, such as body weight and exposure to light intensity.

Regarding the on-body locations of wearable devices that measure PGHD, approximately one-third were worn on the wrist (35/116, 30.2%). Additionally, mobile apps or cameras were used without attaching them to the body in 25.9% (30/116) of cases. Regarding the wearing period of the device, the device was worn for ≤1 week in 47 (41.6%) cases and for >12 weeks (3 months) in 16 (14.2%) cases.

PGHD were collected in various ways, including passive, active, and combined collection (Table 3 and Figure 2). PGHD were primarily passively collected from participants (63/88, 72%),

XSL•FO
RenderX

constituting most of the included studies published between 2006 and 2022. Furthermore, in 20 (22%) studies, PGHD were actively collected from participants from 2015 onward. However, few studies combined passive and active collection methods for PGHD (5/88, 6%), which have been used in research since 2020.

Figure 2. Type of patient-generated health data collection by year.

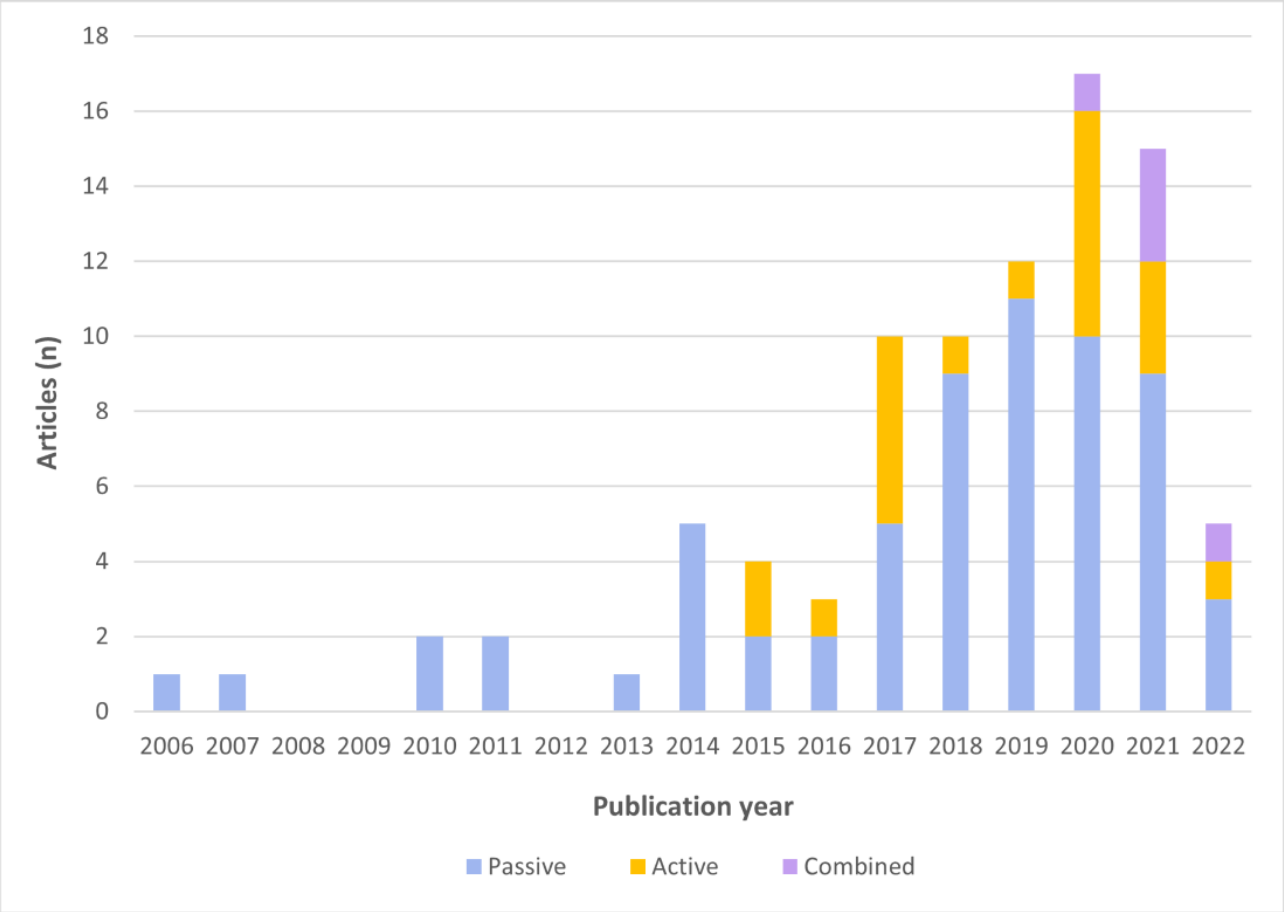

In detail, 25 studies actively collected PGHD, including combined collection methods. Among these, smartphones, tablet PCs, or computers were the predominant devices for active data collection through mobile apps or surveys (23/25, 92%). Except for 1 study that used computers, of the 22 studies using smartphones and tablet PCs, 6 (27%) used the Android (Google LLC) operating system, 5 (23%) used iOS (Apple Inc), 3 (14%) used both, and 8 (36%) did not specify the operating system. Regarding device ownership, 10 (45%) studies provided devices to participants, 3 (14%) studies relied on participants' own devices, 4 (18%) studies used both approaches, and 5 (23%) studies did not report on this aspect.

Among the studies included in this review, approximately a quarter (21/88, 24%) had a sample size of over 100 participants. The PGHD monitoring period varied from 1 day to 24 weeks, with over 60% (13/21) of studies reporting a duration of 1 week or less. Additionally, more than three-quarters of these studies (16/21) collected PGHD passively, primarily using wearables, including both research-grade and consumer-grade devices. Of the 88 studies, 67 (76%) reported on participants' adherence to PGHD collection, with adherence rates ranging from 36% to 100%. More than half of these studies (36/67, 54%) reported an adherence rate above 80%. In addition, the dropout rates reported in 57 longitudinal studies ranged from 0% to 72.4%. Approximately one-fifth of these studies (13/57, 23%) completed PGHD collection with full participant retention.

In this review, health care providers supported PGHD collection from participants in only 20 (23%) studies. Health care providers assisted with the synchronization of the devices, troubleshooting of technical issues, and so forth (10/20, 50%) [42,44,74,86,87,89,98,101,103,104]. They also provided reminders to participants, encouraging PGHD collection (9/20, 45%) [32,36,40,43,58,83,92,97,101]. In addition, health care providers directly assisted with data collection (1/20, 5%) [47] and requested additional information if participants did not respond appropriately (1/20, 5%) [90]. In the case of support methods, phone calls (11/20, 55%) [40,42,43,47,83,86,87,89,90,92,97], text messages (1/20, 5%) [86], app push notifications (4/20, 20%) [32,40,43,58], emails (1/20, 5%) [83], and in-person contact (2/20, 10%) [44,47] were identified.

In this study, the stages of PGHD use were classified as follows: (1) identification, monitoring, review, and analysis; (2) feedback and reporting; (3) motivation; and (4) education and coaching (Multimedia Appendix 4). Identification, monitoring, review, and analysis of PGHD were implemented across all studies (88/88, 100%). Other stages were implemented by interacting with the participants and health care providers. Feedback and

XSL•FO
RenderX

reporting were applied in 34 (39%) studies by using PGHD to aid participants in checking their data and receiving technical advice [24,26,32,33,38,40,42,43,52-54,56,58,60,63,69-72,74,75,79,82,83,86-90,92,93,98,101,104]. PGHD was used as a motivator for participants in approximately one-third of the studies (30/88, 34%) [24,32,33,40,42,43,53,54,56,58,60,63,65,70,71,74,75,77,79,82,86-88,92,93,96,98,100,101,103]. It included directly encouraging participants to generate data, reminding them, setting goals, and supporting them in the process of generating PGHD. Approximately one-fifth of the studies (19/88, 22%) provided education and coaching using PGHD, enhancing the attainment of additional information, knowledge, and skills [25,32,40,53,56,58,70-72,75,77,79,82,87,88,92,93,100,104].

Discussion

Principal Findings

This scoping review presents a broad overview and summary of the studies on PGHD use among older adults with cancer. Following are the key findings of this scoping review, which included a total of 88 studies published between 2006 and 2022. A variety of devices were used to collect PGHD, including research-grade wearable devices, consumer-grade wearable devices, and smartphones or tablet PCs. These devices assessed different PGHD in older adults with cancer, including physical activity, sleep, vital signs, and PROs. In most studies, PGHD were collected passively, requiring no additional effort from the patient, other than wearing the device. Active and mixed PGHD collection methods have recently been applied to research. In most studies, there was no support from health care providers when PGHD were collected from older adults with cancer. PGHD were used for the identification, monitoring, review, and analysis in all studies included in the review. However, fewer than half of the studies identified advanced use of PGHD, such as feedback and reporting, motivation, and education and coaching.

Study Characteristics

Since 2017, an increasing number of studies have been published on the applications of PGHD in older adults with cancer. This may be attributed to the fact that the MeSH term for PGHD was introduced in PubMed in 2018. Given the current publication trend, this can be considered an early stage of use of PGHD-related research in older adults with cancer, indicating that there is a need for further research. Approximately one-third of the included studies used secondary sources for PGHD; most were original, while 5 used cohort data. These 5 studies used the following cohort data: the Baltimore Longitudinal Study of Aging [55] and the National Health and Nutrition Examination Survey [19,22,23,62] from the United States, which measured the participants' physical activity. The advantage of using data from a cohort is that PGHD can be collected from a relatively large number of people and can be used by many researchers simultaneously, which can have a positive impact on patient health care. Therefore, it is recommended to consider including PGHD when constructing a large cohort study.

Participants Characteristics

Approximately 80% (69/88) of the studies reviewed in this analysis focused on the application of PGHD in home settings, underscoring its well-established use outside of hospital environments. Our findings indicate that it is valuable to provide health care providers with continuously collected data on health status at home through PGHD [105]. PGHD can offer patients valuable insights into their health by allowing them to monitor their health at home between hospital visits or during self-care [106]. Moreover, the application of PGHD eliminates time and location constraints, thereby enhancing patient accessibility and participation in their health care [107].

In this review, the cancer treatment statuses of the participants differed. Notably, the cancer survivorship phase, which constituted a quarter of this review, often lacked ongoing health care beyond periodic hospital follow-up visits. Multiple studies have highlighted the unmet need of survivors of cancer for continued care postactive treatment [108,109]. However, PGHD offers several benefits for survivors of cancer. Petersen [7] indicated that it is beneficial for survivors of cancer to apply for PGHD to improve and promote their health. For instance, the collection of PGHD at home, encompassing various health aspects, such as symptoms, can assist health care providers in identifying and managing health issues effectively [7]. In other words, PGHD have the potential to improve patient engagement among survivors who do not frequently receive hospital care and allow health care providers to access patients in a clinical setting. This makes it a promising tool for enhancing the quality of life and health outcomes of survivors of cancer, thereby encouraging future studies to actively use PGHD in their management.

PGHD Characteristics

We found that wearable devices were used to collect approximately three-quarters of the PGHD. These wearable devices have expanded the scope of research beyond the confines of electronic health records by enabling integrated data acquisition from patients [110]. Previous reviews reporting on the use of wearable devices across various health domains indicated that approximately 86% data were collected using consumer-grade wearable devices, including Fitbits, whereas approximately 14% were from research-grade devices [111]. In contrast, in this study, research-grade wearable devices were used in approximately half of the cases and consumer-grade in approximately a quarter. This difference can be attributed to the perceived suitability and higher measurement accuracy of research-grade wearable devices in research settings, particularly among older adults with cancer, which is the target population of this review [112]. Despite this, recent clinical studies have shown an increase in the use of consumer-grade wearable devices, and their reliability and validity have been established [113]. Given their relative affordability and lower burden of wear, consumer-grade wearable devices, which enable continuous usage, are considered to have a high potential for widespread use [114]. Therefore, rather than distinguishing between the types of wearable devices, the focus should be on the selection of the most appropriate device for collecting

XSL•FO
RenderX

PGHD, considering the specific circumstances and environments of older adults with cancer.

In this review, more than half of the collected PGHD items focused on physical activity. Previous studies have shown that older adults with cancer tend to engage in less physical activity compared to middle-aged patients [115]. Moreover, physical activity has been demonstrated to enhance physical function and overall quality of life in older adults with cancer [116,117]. This underscores the importance of prioritizing increased physical activity as a critical factor in improving the health of this population. PGHD can increase health awareness and patient engagement, leading to positive effects on physical activity [107]. Essentially, it can help older adults with cancer in making health behavior changes that enhance physical activity while supporting their autonomy [7]. Moreover, self-monitoring of physical activity using PGHD can motivate individuals to engage in these activities [118]. Therefore, future research using PGHD to monitor and encourage physical activity among older adults with cancer may improve their health outcomes.

This review identified PGHD that measure PROs, including symptoms, in older adults with cancer, which is crucial for enhancing treatment and health outcomes [119]. While visiting a hospital in person and objectively measuring their condition is challenging for patients with cancer, self-reporting conditions, including symptoms, can be more beneficial [120]. Patients were able to report the symptoms they experienced in real time from their homes through PGHD, and contrary to expectations, they did not find it burdensome to report daily [40]. PGHD offer a means for patients with cancer to communicate their condition directly to health care providers, thereby eliminating time and distance constraints and benefiting both the patients and providers [7]. Especially, older patients with cancer are at a higher risk of reporting more severe symptoms [121]. Therefore, enabling older patients with cancer to assess their condition, including symptoms, through PGHD without time and distance limitations can facilitate the early detection of deterioration and personalized care [119]. However, none of the studies included in this review actively involved older adults with cancer in selecting the PGHD to be collected. Therefore, future studies should consider selecting and collecting PGHD items based on participants' needs to enhance engagement.

In addition, based on the data collected through the PROs, reports or alerts about symptoms were provided to patients, and health care providers offered advice or interventions for symptom management [32,42,79]. This indicates that patients, even when not in the hospital, can experience gratitude and psychological comfort, knowing that they are connected to and receiving care from health care providers and are managing their health issues [40]. PGHD enabled patients to gain confidence in self-management of their symptoms, suggesting a positive impact on their health management.

This review categorizes PGHD collection into passive, active, and combined methods, with the passive collection being predominantly used (63/88, 72%). This differs from a prior review of PGHD for disease prevention and health promotion, in which active or partially active methods were more common in healthy populations [16]. Passive methods that use wearable devices for objective data collection without an added burden have been in use since 2006. Given the focus on older adults with cancer in this review, simpler passive methods have been widely adopted. In addition, health care providers tend to view passively collected PGHD as more reliable and more challenging to manipulate, as indicated in previous studies [122].

In contrast, approximately one-fifth of the studies used the active collection method, which showed a gradual increase over time. This method, which involves direct and active participation of the patients in data collection, allows patients to report their subjective states, contributing to enhanced patient engagement [123]. However, a previous review identified conflicting characteristics between passive and active collection [16]. Passive approaches were found to be less burdensome for patients, but may reduce their participation, whereas active approaches require voluntary effort, but motivate patient involvement [16]. Considering these characteristics, a combined collection approach may be ideal for PGHD. In this review, only approximately 6% (5/88) of the studies used a combined collection method, which started to appear in 2020. For example, in the study by Low et al [91], passive data collection using wearable and smartphone sensors for PGHD, combined with active data collection of patient-reported symptoms, facilitated personalized and adaptive symptom assessments to enable early detection of adverse events. This approach may ultimately enhance patient well-being by enabling timely interventions throughout long-term cancer treatment and recovery. Thus, future research should consider using a combined data collection approach to comprehensively assess the health status of older adults with cancer.

An adequate level of digital literacy is essential for patients to generate PGHD, comprehend their significance, and receive motivation for behavioral changes to enhance adherence. However, concerns have arisen about the potential disparities among those unable to collect or use PGHD for various reasons. Age is a contributing factor, as older adults may struggle to adopt new technologies owing to their unfamiliarity [124]. Moreover, patients with cancer often experience higher physical and psychological symptom burdens than do those without cancer, potentially reducing their inclination to use digital health technologies [125]. Therefore, future studies may require additional strategies to identify and enhance digital literacy among older adults with cancer. A tailored approach that integrates demographic factors (eg, age or gender), geographic factors (eg, place of residence), socioeconomic factors (eg, education level), behavioral factors (eg, information-seeking behavior), and social factors (eg, social support) could promote more effective use of wearable devices and mobile apps for PGHD collection and use [126].

Regarding the PGHD use stages, all studies used PGHD at the first stage, primarily collecting and analyzing PGHD, to identify and monitor the health status of older adults with cancer for research purposes. This aligns with a previous review of disease prevention and health promotion, in which many studies reviewed and analyzed PGHD [16]. However, the first stage of PGHD use can be considered as merely using the gathered information for analysis. To facilitate bidirectional rather than unidirectional communication through PGHD, patients must be

XSL•FO
RenderX

able to check or access their generated data. The studies included in this review initially identified and analyzed PGHD, and then PGHD were used for advanced use depending on the purpose of the study. In only approximately one-third of the studies, the patients were provided with information and feedback on PGHD to help them understand their condition; PGHD motivated patients to achieve health-related goals and was used for education or coaching interventions. Therefore, offering actionable recommendations based on PGHD is expected to increase patient engagement in health behaviors, especially among older adults. This implies that the use of PGHD should not be limited to the generation of data through wearable devices but should extend to accessing this data and receiving feedback that motivates behavior change. Similarly, when collecting PROs including symptoms via mobile apps, a bidirectional structure is necessary in which patients receive advice or guidance in addition to reporting their information.

To promote the widespread use of PGHD in research involving older adults with cancer, it is crucial to increase adherence to digital devices for PGHD measurements. This necessitates identifying the needs, barriers, and facilitators for older adults with cancer regarding PGHD, and providing education where there are gaps to encourage adherence. In particular, patients with lower digital literacy encounter greater barriers to using PGHD [16,127]; this can negatively impact the collection and use of PGHD, emphasizing the need for initiatives to improve digital literacy.

Cultural considerations are also essential for adopting PGHD among older adults with cancer. A survey conducted across four European countries revealed that over half of the participants recognized the usefulness of PGHD and expressed willingness to share it with health care providers to improve their own health [128]. In contrast, for community-dwelling older adults in urban areas of South Korea, family or peer influence was reported as both a facilitator and a barrier to adopting digital technologies [129]. Considering the unique cultural attitudes and behavioral intentions that influence technology adoption is crucial for effectively collecting and using PGHD among older adults with cancer [130].

Moreover, the quality of PGHD is a significant concern, influenced by user-related factors, device and technology, and data governance [110]. Researchers should not only focus on collecting PGHD but also on enhancing data quality for effective use. Additionally, the establishment of standards or frameworks for assessing the quality, reliability, and accuracy of PGHD from various devices is essential. Meeting these requirements will enable the sharing and integration of PGHD with health care providers for use in patient care and health management in clinical settings. Ultimately, future research should aim to integrate PGHD from nonhospital settings into health care systems, such as electronic health records or research registries, to ensure continuity of care.

Limitations

This scoping review endeavored to apply a rigorous methodology; however, it has a few limitations. First, the PubMed MeSH term for PGHD has only been included since 2018, indicating that some studies, corresponding to PGHD, may not have been searched. To compensate for this, search terms corresponding to PGHD such as wearable devices and patient report were included. Furthermore, even if PGHD were not explicitly stated, studies deemed to be PGHD through discussions among researchers were included in the review. Additionally, the references of the selected studies were reviewed to ensure that there were no omissions. Second, this review did not include gray literature in the literature selection process. By excluding theses and conference papers, and focusing only on peer-reviewed studies, it is possible that recent literature containing the latest information on PGHD was excluded. Third, this study included older patients with various types of cancer at different treatment stages, suggesting a need for caution regarding the heterogeneity related to cancer in the application of PGHD. Therefore, careful interpretation of the results of this review is warranted, and future research should focus on understanding the use of PGHD according to the specific types of cancer and treatment stages. Finally, although the inclusion criteria of this review focused on studies with a mean or median age of 65 years or older, some studies may have included participants younger than 65 years. This could affect the generalizability of the findings to the geriatric population, as the presence of younger participants may not fully reflect the health characteristics and challenges faced by older adults.

Conclusions

This scoping review investigated and summarized the overall characteristics and use stages of PGHD among older adults with various types and stages of cancer. Research involving PGHD in older adults with cancer is gradually increasing and is primarily conducted at home through wearable devices or mobile apps. PGHD encompass a collection of physiological variables, such as physical activity, sleep, and vital signs, along with PROs, using both passive and active methods, and have been used in various aspects of research. However, while PGHD collected from older adults with cancer were universally used for analysis within the studies, their use for providing feedback, motivation, and interventions that could impact patients and researchers was limited. These findings suggest that future research should emphasize the use of PGHD, which interact with patients to provide patient-centered care through patient engagement. Furthermore, health care, especially in areas such as physical activity and symptom management, is critical for older adults with cancer, and PGHD can positively influence their active participation in health care. Therefore, our findings may provide insights into PGHD that can be used by health care providers and researchers in future research in the context of geriatric cancer care.

XSL•FO
RenderX

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No.2020R1A6A1A03041989) and the Brain Korea 21 FOUR Project funded by the National Research Foundation (NRF) of Korea, Yonsei University College of Nursing.

Data Availability

The protocol of the scoping review was registered in the Open Science Framework Registry. All data generated in the review is available in the main text and in Multimedia Appendix 3 amd Multimedia Appendix 4.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 520 KB - jmir_v27i1e57379_app1.pdf]

Multimedia Appendix 2
Search strategies for PubMed, EMBASE, and CINAHL.
[DOCX File , 24 KB - jmir_v27i1e57379_app2.docx]

Multimedia Appendix 3
Proportion of studies that focused on various stages of patient-generated health data use (n=88).
[PNG File , 28 KB - jmir_v27i1e57379_app3.png]

Multimedia Appendix 4
Proportion of studies that focused on various stages of patient-generated health data use (n=88).
[PNG File , 68 KB - jmir_v27i1e57379_app4.png]

References

1. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal AB, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 2022;72(5):409-436 [FREE Full text] [doi: 10.3322/caac.21731] [Medline: 35736631]
2. Kenzik KM, Kent EE, Martin MY, Bhatia S, Pisu M. Chronic condition clusters and functional impairment in older cancer survivors: a population-based study. J Cancer Surviv 2016;10(6):1096-1103 [FREE Full text] [doi: 10.1007/s11764-016-0553-4] [Medline: 27229869]
3. Soto-Perez-de-Celis E, Li D, Yuan Y, Lau YM, Hurria A. Functional versus chronological age: geriatric assessments to guide decision making in older patients with cancer. Lancet Oncol 2018;19(6):e305-e316. [doi: 10.1016/S1470-2045(18)30348-6] [Medline: 29893262]
4. Siddique A, Simonsick EM, Gallicchio L. Functional decline among older cancer survivors in the Baltimore longitudinal study of aging. J Am Geriatr Soc 2021;69(11):3124-3133 [FREE Full text] [doi: 10.1111/jgs.17369] [Medline: 34346072]
5. Corbett T, Bridges J. Multimorbidity in older adults living with and beyond cancer. Curr Opin Support Palliat Care 2019;13(3):220-224. [doi: 10.1097/SPC.0000000000000439] [Medline: 31157655]
6. Sehl M, Lu X, Silliman R, Ganz PA. Decline in physical functioning in first 2 years after breast cancer diagnosis predicts 10-year survival in older women. J Cancer Surviv 2013;7(1):20-31 [FREE Full text] [doi: 10.1007/s11764-012-0239-5] [Medline: 23232922]
7. Petersen C. Patient-generated health data: a pathway to enhanced long-term cancer survivorship. J Am Med Inform Assoc 2016;23(3):456-461 [FREE Full text] [doi: 10.1093/jamia/ocv184] [Medline: 26714765]
8. Demiris G, Iribarren SJ, Sward K, Lee S, Yang R. Patient generated health data use in clinical practice: a systematic review. Nurs Outlook 2019;67(4):311-330 [FREE Full text] [doi: 10.1016/j.outlook.2019.04.005] [Medline: 31277895]
9. Shapiro M, Johnston D, Wald J, Mon D. Patient-generated health data. RTI International. 2012. URL: https://www.rti.org/sites/default/files/resources/patientgeneratedhealthdata.pdf [accessed 2023-09-21]
10. Chung AE, Basch EM. Potential and challenges of patient-generated health data for high-quality cancer care. J Oncol Pract 2015;11(3):195-197 [FREE Full text] [doi: 10.1200/JOP.2015.003715] [Medline: 25852139]
11. Hewitt C, Lloyd KC, Tariq S, Durrant A, Claisse C, Kasadha B, et al. Patient-generated data in the management of HIV: a scoping review. BMJ Open 2021;11(5):e046393 [FREE Full text] [doi: 10.1136/bmjopen-2020-046393] [Medline: 34011598]

XSL•FO
RenderX

12. Nagpal MS, Barbaric A, Sherifali D, Morita PP, Cafazzo JA. Patient-generated data analytics of health behaviors of people living with type 2 diabetes: scoping review. JMIR Diabetes 2021;6(4):e29027 [FREE Full text] [doi: 10.2196/29027] [Medline: 34783668]
13. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
14. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020;18(10):2119-2126 [FREE Full text] [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic patient-generated health data to facilitate disease prevention and health promotion: scoping review. J Med Internet Res 2019;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]
17. Fernandes R, Stone P, Andrews P, Morgan R, Sharma S. Comparison between fatigue, sleep disturbance, and circadian rhythm in cancer inpatients and healthy volunteers: evaluation of diagnostic criteria for cancer-related fatigue. J Pain Symptom Manage 2006;32(3):245-254 [FREE Full text] [doi: 10.1016/j.jpainsymman.2006.03.014] [Medline: 16939849]
18. Fouladiun M, Körner U, Gunnebo L, Sixt-Ammilon P, Bosaeus I, Lundholm K. Daily physical-rest activities in relation to nutritional state, metabolism, and quality of life in cancer patients with progressive cachexia. Clin Cancer Res 2007;13(21):6379-6385 [FREE Full text] [doi: 10.1158/1078-0432.CCR-07-1147] [Medline: 17975150]
19. Lynch BM, Dunstan DW, Healy GN, Winkler E, Eakin E, Owen N. Objectively measured physical activity and sedentary time of breast cancer survivors, and associations with adiposity: findings from NHANES (2003-2006). Cancer Causes Control 2010;21:283-288 [FREE Full text] [doi: 10.1007/s10552-009-9460-6] [Medline: 19882359]
20. Maddocks M, Byrne A, Johnson CD, Wilson RH, Fearon KCH, Wilcock A. Physical activity level as an outcome measure for use in cancer cachexia trials: a feasibility study. Support Care Cancer 2010;18:1539-1544 [FREE Full text] [doi: 10.1007/s00520-009-0776-2] [Medline: 19956982]
21. Hanisch LJ, Gehrman PR. Circadian rhythm of hot flashes and activity levels among prostate cancer patients on androgen deprivation therapy. Aging Male 2011;14(4):243-248 [FREE Full text] [doi: 10.3109/13685538.2011.582528] [Medline: 21627561]
22. Lynch BM, Dunstan DW, Winkler E, Healy GN, Eakin E, Owen N. Objectively assessed physical activity, sedentary time and waist circumference among prostate cancer survivors: findings from the national health and nutrition examination survey (2003-2006). Eur J Cancer Care (Engl) 2011;20:514-519 [FREE Full text] [doi: 10.1111/j.1365-2354.2010.01205.x] [Medline: 20597954]
23. Loprinzi PD, Lee H, Cardinal BJ. Objectively measured physical activity among US cancer survivors: considerations by weight status. J Cancer Surviv 2013;7:493-499 [FREE Full text] [doi: 10.1007/s11764-013-0293-7] [Medline: 23686709]
24. Frensham LJ, Zarnowiecki DM, Parfitt G, King S, Dollman J. The experiences of participants in an innovative online resource designed to increase regular walking among rural cancer survivors: a qualitative pilot feasibility study. Support Care Cancer 2014;22:1923-1929 [FREE Full text] [doi: 10.1007/s00520-014-2177-4] [Medline: 24573604]
25. Galvão DA, Spry N, Denham J, Taaffe DR, Cormie P, Joseph D, et al. A multicentre year-long randomised controlled trial of exercise training targeting physical functioning in men with prostate cancer previously treated with androgen suppression and radiation from TROG 03.04 RADAR. Eur Urol 2014;65(5):856-864 [FREE Full text] [doi: 10.1016/j.eururo.2013.09.041] [Medline: 24113319]
26. Granger CL, Denehy L, McDonald CF, Irving L, Clark RA. Physical activity measured using global positioning system tracking in non-small cell lung cancer: an observational study. Integr Cancer Ther 2014;13(6):482-492 [FREE Full text] [doi: 10.1177/1534735414542484] [Medline: 25006040]
27. Granger CL, McDonald CF, Irving L, Clark RA, Gough K, Murnane A, et al. Low physical activity levels and functional decline in individuals with lung cancer. Lung Cancer 2014;83(2):292-299 [FREE Full text] [doi: 10.1016/j.lungcan.2013.11.014] [Medline: 24360323]
28. Roche VP, Mohamad-Djafari A, Innominato PF, Karaboué A, Gorbach A, Lévi FA. Thoracic surface temperature rhythms as circadian biomarkers for cancer chronotherapy. Chronobiol Int 2014;31(3):409-420 [FREE Full text] [doi: 10.3109/07420528.2013.864301] [Medline: 24397341]
29. Aktas A, Hullihen B, Shrotriya S, Thomas S, Walsh D, Estfan B. Connected health: cancer symptom and quality-of-life assessment using a tablet computer: a pilot study. Am J Hosp Palliat Care 2015;32(2):189-197 [FREE Full text] [doi: 10.1177/1049909113510963] [Medline: 24212101]
30. Dean GE, Abu Sabbah EA, Yingrengreung S, Ziegler P, Chen H, Steinbrenner LM, et al. Sleeping with the enemy: sleep and quality of life in patients with lung cancer. Cancer Nurs 2015;38(1):60-70. [doi: 10.1097/NCC.0000000000000128] [Medline: 25486204]

XSL•FO
RenderX

31. Livingston PM, Craike MJ, Salmon J, Courneya KS, Gaskin CJ, Fraser SF, et al. Effects of a clinician referral and exercise program for men who have completed active treatment for prostate cancer: a multicenter cluster randomized controlled trial (ENGAGE). Cancer 2015;121(15):2646-2654 [FREE Full text] [doi: 10.1002/cncr.29385] [Medline: 25877784]
32. Sundberg K, Eklöf AL, Blomberg K, Isaksson A, Wengström Y. Feasibility of an interactive ICT-platform for early assessment and management of patient-reported symptoms during radiotherapy for prostate cancer. Eur J Oncol Nurs 2015;19(5):523-528. [doi: 10.1016/j.ejon.2015.02.013] [Medline: 25813529]
33. Drott J, Vilhelmsson M, Kjellgren K, Berterö C. Experiences with a self-reported mobile phone-based system among patients with colorectal cancer: a qualitative study. JMIR mHealth uHealth 2016;4(2):e66 [FREE Full text] [doi: 10.2196/mhealth.5426] [Medline: 27282257]
34. Gaskin CJ, Craike M, Mohebbi M, Salmon J, Courneya KS, Broadbent S, et al. Associations of objectively measured moderate-to-vigorous physical activity and sedentary behavior with quality of life and psychological well-being in prostate cancer survivors. Cancer Causes Control 2016;27(9):1093-1103 [FREE Full text] [doi: 10.1007/s10552-016-0787-5] [Medline: 27469939]
35. van Roekel EH, Winkler EA, Bours MJ, Lynch BM, Willems PJ, Meijer K, et al. Associations of sedentary time and patterns of sedentary time accumulation with health-related quality of life in colorectal cancer survivors. Prev Med Rep 2016;4:262-269 [FREE Full text] [doi: 10.1016/j.pmedr.2016.06.022] [Medline: 27419042]
36. Basch E, Pugh SL, Dueck AC, Mitchell SA, Berk L, Fogh S, et al. Feasibility of patient reporting of symptomatic adverse events via the patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE) in a chemoradiotherapy cooperative group multicenter clinical trial. Int J Radiat Oncol Biol Phys 2017;98(2):409-418 [FREE Full text] [doi: 10.1016/j.ijrobp.2017.02.002] [Medline: 28463161]
37. Chang WP, Lin CC. Changes in the sleep-wake rhythm, sleep quality, mood, and quality of life of patients receiving treatment for lung cancer: a longitudinal study. Chronobiol Int 2017;34(4):451-461. [doi: 10.1080/07420528.2017.1293678] [Medline: 28282238]
38. Golsteijn RHJ, Bolman C, Peels DA, Volders E, de Vries H, Lechner L. A web-based and print-based computer-tailored physical activity intervention for prostate and colorectal cancer survivors: a comparison of user characteristics and intervention use. J Med Internet Res 2017;19(8):e298 [FREE Full text] [doi: 10.2196/jmir.7838] [Medline: 28835353]
39. Jeffery E, Lee YG, McVeigh J, Straker L, Wooding T, Newton RU, et al. Feasibility of objectively measured physical activity and sedentary behavior in patients with malignant pleural effusion. Support Care Cancer 2017;25:3133-3141. [doi: 10.1007/s00520-017-3721-9] [Medline: 28455548]
40. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
41. Lawrence LM, Stone MR, Rainham DG, Keats MR. Environments associated with moderate-to-vigorous physical activity and sedentary behavior of colorectal cancer survivors. Int J Behav Med 2017;24:120-126. [doi: 10.1007/s12529-016-9575-6] [Medline: 27364123]
42. Nyman MH, Frank C, Langius-Eklöf A, Blomberg K, Sundberg K, Wengström Y. Patients' perspective on participation in care with or without the support of a smartphone app during radiotherapy for prostate cancer: qualitative study. JMIR mHealth uHealth 2017;5(7):e107 [FREE Full text] [doi: 10.2196/mhealth.6829] [Medline: 28747294]
43. Sundberg K, Wengström Y, Blomberg K, Hälleberg-Nyman M, Frank C, Langius-Eklöf A. Early detection and management of symptoms using an interactive smartphone application (Interaktor) during radiotherapy for prostate cancer. Support Care Cancer 2017;25(7):2195-2204 [FREE Full text] [doi: 10.1007/s00520-017-3625-8] [Medline: 28236145]
44. Thomas S, Walsh D, Shrotriya S, Aktas A, Hullihen B, Estfan B, et al. Symptoms, quality of life, and daily activities in people with newly diagnosed solid tumors presenting to a medical oncologist. Am J Hosp Palliat Med 2017;34(7):611-621. [doi: 10.1177/1049909116649948] [Medline: 27217423]
45. Wynter-Blyth V, MacKinnon T. Streamlining perioperative care for oesophago-gastric cancer surgery patients using home remote monitoring. Primary Health Care 2017;27(5):27-31. [doi: 10.7748/phc.2017.e1225]
46. Bade BC, Brooks MC, Nietert SB, Ulmer A, Thomas DD, Thomas DD, et al. Assessing the correlation between physical activity and quality of life in advanced lung cancer. Integr Cancer Ther 2018;17(1):73-79 [FREE Full text] [doi: 10.1177/1534735416684016] [Medline: 28024420]
47. Bernatchez MS, Savard J, Savard MH, Aubin M, Ivers H. Sleep-wake difficulties in community-dwelling cancer patients receiving palliative care: subjective and objective assessment. Palliat Support Care 2018;16(6):756-766. [doi: 10.1017/S1478951517000815] [Medline: 28933313]
48. Chang WP, Smith R, Lin CC. Age and rest-activity rhythm as predictors of survival in patients with newly diagnosed lung cancer. Chronobiol Int 2018;35(2):188-197. [doi: 10.1080/07420528.2017.1391278] [Medline: 29144183]
49. D'Silva A, Gardiner PA, Boyle T, Bebb DG, Johnson ST, Vallance JK. Associations of objectively assessed physical activity and sedentary time with health-related quality of life among lung cancer survivors: a quantile regression approach. Lung Cancer 2018;119:78-84. [doi: 10.1016/j.lungcan.2018.03.010] [Medline: 29656757]

XSL•FO
RenderX

50. Golsteijn RHJ, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018;15:106 [FREE Full text] [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
51. Gonzalez BD, Small BJ, Cases MG, Williams NL, Fishman MN, Jacobsen PB, et al. Sleep disturbance in men receiving androgen deprivation therapy for prostate cancer: the role of hot flashes and nocturia. Cancer 2018;124(3):499-506 [FREE Full text] [doi: 10.1002/cncr.31024] [Medline: 29072790]
52. Soto-Perez-De-Celis E, Kim H, Rojo-Castillo MP, Sun CL, Chavarri-Guerra Y, Navarrete-Reyes AP, et al. A pilot study of an accelerometer-equipped smartphone to monitor older adults with cancer receiving chemotherapy in Mexico. J Geriatr Oncol 2018;9(2):145-151. [doi: 10.1016/j.jgo.2017.09.008] [Medline: 29017891]
53. Tang FW, Chan CW, Choy YP, Loong HH, Chow KM, So WK. A feasibility study on using tablet personal computers for self-reported symptom assessment in newly diagnosed lung cancer patients. Int J Nurs Pract 2018;24(4):e12658. [doi: 10.1111/ijn.12658] [Medline: 29642280]
54. Trinh L, Arbour-Nicitopoulos KP, Sabiston CM, Berry SR, Loblaw A, Alibhai SMH, et al. RiseTx: testing the feasibility of a web application for reducing sedentary behavior among prostate cancer survivors receiving androgen deprivation therapy. Int J Behav Nutr Phys Act 2018;15:49 [FREE Full text] [doi: 10.1186/s12966-018-0686-0] [Medline: 29880049]
55. Wanigatunga AA, Gresham GK, Kuo PL, Martinez-Amezcua P, Zipunnikov V, Dy SM, et al. Contrasting characteristics of daily physical activity in older adults by cancer history. Cancer 2018;124(24):4692-4699 [FREE Full text] [doi: 10.1002/cncr.31745] [Medline: 30276791]
56. Alibhai SMH, Santa Mina D, Ritvo P, Tomlinson G, Sabiston C, Krahn M, et al. A phase II randomized controlled trial of three exercise delivery methods in men with prostate cancer on androgen deprivation therapy. BMC Cancer 2019;19:2 [FREE Full text] [doi: 10.1186/s12885-018-5189-5] [Medline: 30606137]
57. Fujisawa D, Temel JS, Greer JA, El-Jawahri A, Traeger L, Jacobs JM, et al. Actigraphy as an assessment of performance status in patients with advanced lung cancer. Palliat Support Care 2019;17(5):574-578. [doi: 10.1017/S1478951518001074] [Medline: 30739635]
58. Gustavell T, Sundberg K, Segersvärd R, Wengström Y, Langius-Eklöf A. Decreased symptom burden following surgery due to support from an interactive app for symptom management for patients with pancreatic and periampullary cancer. Acta Oncol 2019;58(9):1307-1314 [FREE Full text] [doi: 10.1080/0284186X.2019.1633473] [Medline: 31284797]
59. Jeffery E, Lee YCG, Newton RU, Lyons-Wall P, McVeigh J, Nowak AK, et al. Body composition and nutritional status in malignant pleural mesothelioma: implications for activity levels and quality of life. Eur J Clin Nutr 2019;73(10):1412-1421. [doi: 10.1038/s41430-019-0418-9] [Medline: 30886321]
60. Kenfield SA, van Blarigan EL, Ameli N, Lavaki E, Cedars B, Paciorek AT, et al. Feasibility, acceptability, and behavioral outcomes from a technology-enhanced behavioral change intervention (prostate 8): a pilot randomized controlled trial in men with prostate cancer. Eur Urol 2019;75(6):950-958 [FREE Full text] [doi: 10.1016/j.eururo.2018.12.040] [Medline: 30638635]
61. Lemanska A, Poole K, Griffin BA, Manders R, Saxton JM, Turner L, et al. Community pharmacy lifestyle intervention to increase physical activity and improve cardiovascular health of men with prostate cancer: a phase II feasibility study. BMJ Open 2019;9(6):e025114 [FREE Full text] [doi: 10.1136/bmjopen-2018-025114] [Medline: 31186245]
62. Loprinzi PD, Nooe A. Objectively measured physical activity and all-cause mortality among cancer survivors: national prospective cohort study. South Med J 2019;112(4):234-237. [doi: 10.14423/SMJ.0000000000000956] [Medline: 30943543]
63. Metcalf M, Glazyrine V, Glavin K, Dahlgren A, Michael C, Bechtel M, et al. The feasibility of a health care application in the treatment of patients undergoing radical cystectomy. J Urol 2019;201(5):902-908. [doi: 10.1097/JU.0000000000000050] [Medline: 30694938]
64. Ngo-Huang A, Parker NH, Bruera E, Lee RE, Simpson R, O'Connor DP, et al. Home-based exercise prehabilitation during preoperative treatment for pancreatic cancer is associated with improvement in physical function and quality of life. Integr Cancer Ther 2019;18:1534735419894061 [FREE Full text] [doi: 10.1177/1534735419894061] [Medline: 31858837]
65. Parker NH, Ngo-Huang A, Lee RE, O'Connor DP, Basen-Engquist KM, Petzel MQ, et al. Physical activity and exercise during preoperative pancreatic cancer treatment. Support Care Cancer 2019;27:2275-2284 [FREE Full text] [doi: 10.1007/s00520-018-4493-6] [Medline: 30334105]
66. Porserud A, Aly M, Nygren-Bonnier M, Hagströmer M. Objectively measured mobilisation is enhanced by a new behaviour support tool in patients undergoing abdominal cancer surgery. Eur J Surg Oncol 2019;45(10):1847-1853 [FREE Full text] [doi: 10.1016/j.ejso.2019.04.013] [Medline: 31030805]
67. Zahiri M, Chen KM, Zhou H, Nguyen H, Workeneh BT, Yellapragada SV, et al. Using wearables to screen motor performance deterioration because of cancer and chemotherapy-induced peripheral neuropathy (CIPN) in adults - toward an early diagnosis of CIPN. J Geriatr Oncol 2019;10(6):960-967 [FREE Full text] [doi: 10.1016/j.jgo.2019.01.010] [Medline: 30665876]
68. Barkley R, Khalil M, Shen P, Levine EA, Votanopoulos K, Clark CJ. Feasibility of low-cost accelerometers in measuring functional recovery after major oncologic surgery. J Surg Oncol 2020;121:279-285 [FREE Full text] [doi: 10.1002/jso.25789] [Medline: 31782174]

XSL•FO
RenderX

69. Buergy D, Siefert V, Neumaier C, Ganslandt T, Sperk E, Blessing M, et al. Prospective trial on telemonitoring of geriatric cancer patients using handheld devices. Strahlenther Onkol 2020;196(3):205-212. [doi: 10.1007/s00066-019-01548-0] [Medline: 31740981]
70. Chan JM, van Blarigan EL, Langlais CS, Zhao S, Ramsdill JW, Daniel K, et al. Feasibility and acceptability of a remotely delivered, web-based behavioral intervention for men with prostate cancer: four-arm randomized controlled pilot trial. J Med Internet Res 2020;22(12):e19238 [FREE Full text] [doi: 10.2196/19238] [Medline: 33382378]
71. Gustavell T, Sundberg K, Langius-Eklöf A. Using an interactive app for symptom reporting and management following pancreatic cancer surgery to facilitate person-centered care: descriptive study. JMIR mHealth uHealth 2020;8(6):e17855 [FREE Full text] [doi: 10.2196/17855] [Medline: 32554375]
72. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: exploratory analysis from a randomised controlled trial. PLoS One 2020;15(10):e0240967 [FREE Full text] [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
73. Jakobsen G, Engstrøm M, Thronæs M, Løhre ET, Kaasa S, Fayers P, et al. Sleep quality in hospitalized patients with advanced cancer: an observational study using self-reports of sleep and actigraphy. Support Care Cancer 2020;28:2015-2023. [doi: 10.1007/s00520-019-04998-5] [Medline: 31392550]
74. Jonker LT, Hendriks S, Lahr MM, van Munster BC, de Bock GH, van Leeuwen BL. Postoperative recovery of accelerometer-based physical activity in older cancer patients. Eur J Surg Oncol 2020;46(11):2083-2090 [FREE Full text] [doi: 10.1016/j.ejso.2020.06.012] [Medline: 32682650]
75. Kneuertz PJ, Jagadesh N, Perkins A, Fitzgerald M, Moffatt-Bruce SD, Merritt RE, et al. Improving patient engagement, adherence, and satisfaction in lung cancer surgery with implementation of a mobile device platform for patient reported outcomes. J Thorac Dis 2020;12(11):6883-6891 [FREE Full text] [doi: 10.21037/jtd.2020.01.23] [Medline: 33282391]
76. Kuo LC, Chang WP, Huang HC, Lin CC. Association of time-varying rest-activity rhythm with survival in older adults with lung cancer. Cancer Nurs 2020;43(1):45-51. [doi: 10.1097/NCC.0000000000000647] [Medline: 30299421]
77. Lafaro KJ, Raz DJ, Kim JY, Hite S, Ruel N, Varatkar G, et al. Pilot study of a telehealth perioperative physical activity intervention for older adults with cancer and their caregivers. Support Care Cancer 2020;28(8):3867-3876 [FREE Full text] [doi: 10.1007/s00520-019-05230-0] [Medline: 31845007]
78. Lévi F, Komarzynski S, Huang Q, Young T, Ang Y, Fuller C, et al. Tele-monitoring of cancer patients' rhythms during daily life identifies actionable determinants of circadian and sleep disruption. Cancers (Basel) 2020;12(7):1938 [FREE Full text] [doi: 10.3390/cancers12071938] [Medline: 32708950]
79. Maguire R, Connaghan J, Arber A, Klepacz N, Blyth KG, McPhelim J, et al. Advanced symptom management system for patients with malignant pleural mesothelioma (ASyMSmeso): mixed methods study. J Med Internet Res 2020;22(11):e19180 [FREE Full text] [doi: 10.2196/19180] [Medline: 33180025]
80. Manda S, Yimer HA, Noga SJ, Girnius S, Yasenchak CA, Charu V, et al. Feasibility of long-term proteasome inhibition in multiple myeloma by in-class transition from bortezomib to ixazomib. Clin Lymphoma Myeloma Leuk 2020;20(11):e910-e925 [FREE Full text] [doi: 10.1016/j.clml.2020.06.024] [Medline: 32912820]
81. Panda N, Solsky I, Huang EJ, Lipsitz S, Pradarelli JC, Delisle M, et al. Using smartphones to capture novel recovery metrics after cancer surgery. JAMA Surg 2020;155(2):123-129 [FREE Full text] [doi: 10.1001/jamasurg.2019.4702] [Medline: 31657854]
82. Park S, Kim K, Ahn HK, Kim JW, Min G, Chung BH, et al. Impact of lifestyle intervention for patients with prostate cancer. Am J Health Behav 2020;44(1):90-99. [doi: 10.5993/AJHB.44.1.10] [Medline: 31783936]
83. Taarnhøj GA, Lindberg H, Dohn LH, Omland LH, Hjøllund NH, Johansen C, et al. Electronic reporting of patient-reported outcomes in a fragile and comorbid population during cancer therapy - a feasibility study. Health Qual Life Outcomes 2020;18:225 [FREE Full text] [doi: 10.1186/s12955-020-01480-3] [Medline: 32653005]
84. Tolstrup LK, Bastholt L, Dieperink KB, Möller S, Zwisler A, Pappot H. The use of patient-reported outcomes to detect adverse events in metastatic melanoma patients receiving immunotherapy: a randomized controlled pilot trial. J Patient Rep Outcomes 2020;4:88 [FREE Full text] [doi: 10.1186/s41687-020-00255-0] [Medline: 33125537]
85. Appleyard S, Larkin M, Stewart E, Minton O, Gilbert D. Digital medicine in men with advanced prostate cancer - a feasibility study of electronic patient-reported outcomes in patients on systemic treatment. Clin Oncol (R Coll Radiol) 2021;33(12):751-760 [FREE Full text] [doi: 10.1016/j.clon.2021.04.008] [Medline: 33966948]
86. Bade BC, Gan G, Li F, Lu L, Tanoue L, Silvestri GA, et al. "Randomized trial of physical activity on quality of life and lung cancer biomarkers in patients with advanced stage lung cancer: a pilot study". BMC Cancer 2021;21:352 [FREE Full text] [doi: 10.1186/s12885-021-08084-0] [Medline: 33794808]
87. Blair CK, Harding E, Wiggins C, Kang H, Schwartz M, Tarnower A, et al. A home-based mobile health intervention to replace sedentary time with light physical activity in older cancer survivors: randomized controlled pilot trial. JMIR Cancer 2021;7(2):e18819 [FREE Full text] [doi: 10.2196/18819] [Medline: 33847588]
88. Evans HEL, Galvão DA, Forbes CC, Girard D, Vandelanotte C, Newton RU, et al. Acceptability and preliminary efficacy of a web- and telephone-based personalised exercise intervention for individuals with metastatic prostate cancer: the ExerciseGuide pilot randomised controlled trial. Cancers (Basel) 2021;13(23):5925 [FREE Full text] [doi: 10.3390/cancers13235925] [Medline: 34885036]

XSL•FO
RenderX

89. Jonker LT, Lahr MMH, Oonk MHM, de Bock GH, van Leeuwen BL. Post-discharge telemonitoring of physical activity, vital signs, and patient-reported symptoms in older patients undergoing cancer surgery. Ann Surg Oncol 2021;28(11):6512-6522 [FREE Full text] [doi: 10.1245/s10434-021-09707-3] [Medline: 33641013]
90. Jonker LT, Plas M, de Bock GH, Buskens E, van Leeuwen BL, Lahr MMH. Remote home monitoring of older surgical cancer patients: perspective on study implementation and feasibility. Ann Surg Oncol 2021;28(1):67-78 [FREE Full text] [doi: 10.1245/s10434-020-08705-1] [Medline: 32602060]
91. Low CA, Li M, Vega J, Durica KC, Ferreira D, Tam V, et al. Digital biomarkers of symptom burden self-reported by perioperative patients undergoing pancreatic surgery: prospective longitudinal study. JMIR Cancer 2021;7(2):e27975 [FREE Full text] [doi: 10.2196/27975] [Medline: 33904822]
92. Msaouel P, Oromendia C, Siefker-Radtke AO, Tannir NM, Subudhi SK, Gao J, et al. Evaluation of technology-enabled monitoring of patient-reported outcomes to detect and treat toxic effects linked to immune checkpoint inhibitors. JAMA Netw Open 2021;4(8):e2122998 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.22998] [Medline: 34459906]
93. Pinto BM, Kindred M, Franco R, Simmons V, Hardin J. A 'novel' multi-component approach to promote physical activity among older cancer survivors: a pilot randomized controlled trial. Acta Oncol 2021;60(8):968-975 [FREE Full text] [doi: 10.1080/0284186X.2021.1896032] [Medline: 33689553]
94. Rossi LA, Melstrom LG, Fong Y, Sun V. Predicting post-discharge cancer surgery complications via telemonitoring of patient-reported outcomes and patient-generated health data. J Surg Oncol 2021;123(5):1345-1352 [FREE Full text] [doi: 10.1002/jso.26413] [Medline: 33621378]
95. Sada YH, Poursina O, Zhou H, Workeneh BT, Maddali SV, Najafi B. Harnessing digital health to objectively assess cancer-related fatigue: the impact of fatigue on mobility performance. PLoS One 2021;16(2):e0246101 [FREE Full text] [doi: 10.1371/journal.pone.0246101] [Medline: 33636720]
96. Schuit AS, van Zwieten V, Holtmaat K, Cuijpers P, Eerenstein SEJ, Leemans CR, et al. Symptom monitoring in cancer and fully automated advice on supportive care: Patients' perspectives on self-management strategies and the eHealth self-management application Oncokompas. Eur J Cancer Care (Engl) 2021;30(6):e13497 [FREE Full text] [doi: 10.1111/ecc.13497] [Medline: 34339081]
97. Shih CH, Chou PC, Chou TL, Huang TW. Measurement of cancer-related fatigue based on heart rate variability: observational study. J Med Internet Res 2021;23(7):e25791 [FREE Full text] [doi: 10.2196/25791] [Medline: 36260384]
98. Strother M, Koepsell K, Song L, Faerber J, Bernard J, Malkowicz S, et al. Financial incentives and wearable activity monitors to increase ambulation after cystectomy: a randomized controlled trial. Urol Oncol 2021;39(7):434.e31-434.e38 [FREE Full text] [doi: 10.1016/j.urolonc.2020.11.035] [Medline: 33308975]
99. Tanaka H, Fukui S, Maeda I, Hatano Y, Higuchi A, Higami Y, et al. The change over time of vital signs with consideration for opioid use in the last 2 weeks of life among cancer patients in a palliative care unit: continuous measurement of vital signs using a non-wearable monitor. Cancer Med 2021;10(24):8799-8807 [FREE Full text] [doi: 10.1002/cam4.4382] [Medline: 34845870]
100. Bentley JR, Yu X, Karmarkar AM, Downer B, Prochaska J, Lyons EJ. Feasibility and thematic analysis of narrative visualization materials with physical activity monitoring among breast cancer survivors. BMC Cancer 2022;22(1):553 [FREE Full text] [doi: 10.1186/s12885-022-09629-7] [Medline: 35578196]
101. de Leeuwerk ME, Botjes M, van Vliet V, Geleijn E, de Groot V, van Wegen E, et al. Self-monitoring of physical activity after hospital discharge in patients who have undergone gastrointestinal or lung cancer surgery: mixed methods feasibility study. JMIR Cancer 2022;8(2):e35694 [FREE Full text] [doi: 10.2196/35694] [Medline: 35749165]
102. Higami Y, Higuchi A, Tanaka H, Moriki Y, Utsumi M, Yamakawa M, et al. Nonwearable actigraphy to assess changes in motor activity before and after rescue analgesia in terminally ill patients with cancer: a pilot study. Int J Nurs Pract 2022;28(4):e13019 [FREE Full text] [doi: 10.1111/ijn.13019] [Medline: 34651388]
103. Oswald LB, Li X, Carvajal R, Hoogland AI, Gudenkauf LM, Hansen DK, et al. Longitudinal collection of patient-reported outcomes and activity data during CAR-T therapy: feasibility, acceptability, and data visualization. Cancers (Basel) 2022;14(11):2742 [FREE Full text] [doi: 10.3390/cancers14112742] [Medline: 35681722]
104. Passardi A, Foca F, Caffo O, Tondini CA, Zambelli A, Vespignani R, et al. A remote monitoring system to optimize the home management of oral anticancer therapies (ONCO-TreC): prospective training-validation trial. J Med Internet Res 2022;24(1):e27349 [FREE Full text] [doi: 10.2196/27349] [Medline: 35080505]
105. Purswani JM, Dicker AP, Champ CE, Cantor M, Ohri N. Big data from small devices: the future of smartphones in oncology. Semin Radiat Oncol 2019;29(4):338-347 [FREE Full text] [doi: 10.1016/j.semradonc.2019.05.008] [Medline: 31472736]
106. Cohen DJ, Keller SR, Hayes GR, Dorr DA, Ash JS, Sittig DF. Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project HealthDesign. JMIR Hum Factors 2016;3(2):e26 [FREE Full text] [doi: 10.2196/humanfactors.5919] [Medline: 27760726]
107. Omoloja A, Vundavalli S. Patient generated health data: benefits and challenges. Curr Probl Pediatr Adolesc Health Care 2021;51(11):101103 [FREE Full text] [doi: 10.1016/j.cppeds.2021.101103] [Medline: 34799255]
108. Burg MA, Adorno G, Lopez EDS, Loerzel V, Stein K, Wallace C, et al. Current unmet needs of cancer survivors: analysis of open-ended responses to the American Cancer Society Study of Cancer Survivors II. Cancer 2015;121(4):623-630 [FREE Full text] [doi: 10.1002/cncr.28951] [Medline: 25581252]

XSL•FO
RenderX

109. Mayer DK, Nasso SF, Earp JA. Defining cancer survivors, their needs, and perspectives on survivorship health care in the USA. Lancet Oncol 2017;18(1):e11-e18. [doi: 10.1016/S1470-2045(16)30573-3] [Medline: 28049573]
110. Cho S, Ensari I, Weng C, Kahn MG, Natarajan K. Factors affecting the quality of person-generated wearable device data and associated challenges: rapid systematic review. JMIR mHealth uHealth 2021;9(3):e20738 [FREE Full text] [doi: 10.2196/20738] [Medline: 33739294]
111. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR mHealth uHealth 2022;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
112. Yavelberg L, Zaharieva D, Cinar A, Riddell MC, Jamnik V. A pilot study validating select research-grade and consumer-based wearables throughout a range of dynamic exercise intensities in persons with and without type 1 diabetes: a novel approach. J Diabetes Sci Technol 2018;12(3):569-576 [FREE Full text] [doi: 10.1177/1932296817750401] [Medline: 29320885]
113. Bai Y, Welk GY, Nam YH, Lee JA, Lee JM, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016;48(1):151-158. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
114. Strain T, Wijndaele K, Pearce M, Brage S. Considerations for the use of consumer-grade wearables and smartphones in population surveillance of physical activity. J Meas Phys Behav 2022;5(1):8-14 [FREE Full text] [doi: 10.1123/jmpb.2021-0046]
115. Tarasenko Y, Chen C, Schoenberg N. Self-reported physical activity levels of older cancer survivors: results from the 2014 National Health Interview Survey. J Am Geriatr Soc 2017;65(2):e39-e44 [FREE Full text] [doi: 10.1111/jgs.14589] [Medline: 27943255]
116. Blair CK, Morey MC, Desmond RA, Cohen HJ, Sloane R, Snyder DC, et al. Light-intensity activity attenuates functional decline in older cancer survivors. Med Sci Sports Exerc 2014;46(7):1375-1383 [FREE Full text] [doi: 10.1249/MSS.0000000000000241] [Medline: 24389524]
117. Courneya KS, Karvinen KH. Exercise, aging, and cancer. Appl Physiol Nutr Metab 2007;32(6):1001-1007 [FREE Full text] [doi: 10.1139/H07-074] [Medline: 18059572]
118. Jim HSL, Hoogland AI, Brownstein NC, Barata A, Dicker AP, Knoop H, et al. Innovations in research and clinical care using patient-generated health data. CA Cancer J Clin 2020;70(3):182-199 [FREE Full text] [doi: 10.3322/caac.21608] [Medline: 32311776]
119. Pritchett JC, Patt D, Thanarajasingam G, Schuster A, Snyder C. Patient-reported outcomes, digital health, and the quest to improve health equity. Am Soc Clin Oncol Educ Book 2023;43:e390678 [FREE Full text] [doi: 10.1200/EDBK_390678] [Medline: 37290027]
120. Morgan R, Sail K, Witte L. Patient-reported outcome data. In: Richesson RL, Andrews JE, Fultz Hollis K, editors. Clinical Research Informatics. Cham, Switzerland: Springer International Publishing; 2023:291-308.
121. Bubis LD, Davis LE, Canaj H, Gupta V, Jeong Y, Barbera L, et al. Patient-reported symptom severity among 22,650 cancer outpatients in the last six months of life. J Pain Symptom Manage 2020;59(1):58-66.e4 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.08.016] [Medline: 31430522]
122. Kim B, Ghasemi P, Stolee P, Lee J. Clinicians and older adults' perceptions of the utility of patient-generated health data in caring for older adults: exploratory mixed methods study. JMIR Aging 2021;4(4):e29788 [FREE Full text] [doi: 10.2196/29788] [Medline: 34738913]
123. Nghiem J, Adler DA, Estrin D, Livesey C, Choudhury T. Understanding mental health clinicians' perceptions and concerns regarding using passive patient-generated health data for clinical decision-making: qualitative semistructured interview study. JMIR Form Res 2023;7:e47380 [FREE Full text] [doi: 10.2196/47380] [Medline: 37561561]
124. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR mHealth uHealth 2022;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
125. Hasnan S, Aggarwal S, Mohammadi L, Koczwara B. Barriers and enablers of uptake and adherence to digital health interventions in older patients with cancer: a systematic review. J Geriatr Oncol 2022;13(8):1084-1091 [FREE Full text] [doi: 10.1016/j.jgo.2022.06.004] [Medline: 35752605]
126. Zhang Y, Xu P, Sun Q, Baral S, Xi L, Wang D. Factors influencing the e-health literacy in cancer patients: a systematic review. J Cancer Surviv 2023;17(2):425-440 [FREE Full text] [doi: 10.1007/s11764-022-01260-6] [Medline: 36190672]
127. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021;27(4):267-278 [FREE Full text] [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
128. Karampela M, Ouhbi S, Isomursu M. Connected health user willingness to share personal health data: questionnaire study. J Med Internet Res 2019;21(11):e14537 [FREE Full text] [doi: 10.2196/14537] [Medline: 31774410]
129. Baek J, Kim H, Choi S, Hong S, Kim Y, Kim E, et al. Digital literacy and associated factors in older adults living in urban South Korea: a qualitative study. Comput Inform Nurs 2024;42(3):226-239. [doi: 10.1097/CIN.0000000000001109] [Medline: 38300124]
130. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]

XSL•FO
RenderX

Abbreviations

JBI: Joanna Briggs Institute
MeSH: Medical Subject Headings
PGHD: patient-generated health data
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Review
PRO: patient-reported outcome

Edited by A Mavragani; submitted 14.02.24; peer-reviewed by S Mildrum Chana, A Billis; comments to author 12.06.24; revised version received 05.11.24; accepted 16.12.24; published 04.02.25.

Please cite as:
Kim Y, Kim G, Cho H, Kim Y, Choi M
Application of Patient-Generated Health Data Among Older Adults With Cancer: Scoping Review
J Med Internet Res 2025;27:e57379
URL: https://www.jmir.org/2025/1/e57379
doi:10.2196/57379
PMID:

©Yesol Kim, Geonah Kim, Hyeonmi Cho, Yeonju Kim, Mona Choi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Virtual Reality Technology Interventions in Improving the Social Skills of Children and Adolescents With Autism: Systematic Review

Xipeng Yang[1*], MA; Jinlong Wu[2*], PhD; Yudan Ma[3*], PhD; Jingxuan Yu[1], MA; Hong Cao[1], MA; Aihua Zeng[3], MA; Rui Fu[4], MA; Yucheng Tang[1], PhD; Zhanbing Ren[1], PhD

[1]College of Physical Education, Shenzhen University, Shenzhen, China

[2]College of Physical Education, Southwest University, Chongqing, China

[3]School of Public Teaching, Shanwei Institute of Technology, Shanwei, China

[4]Department of Education, Shenzhen University, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Zhanbing Ren, PhD
College of Physical Education
Shenzhen University
No. 3688,Nanhai Avenue Nanshan District
Shenzhen, 518000
China
Phone: 86 13318881919
Email: rzb@szu.edu.cn

Abstract

Background: Virtual reality (VR) technology has shown significant potential in improving the social skills of children and adolescents with autism spectrum disorder (ASD).

Objective: This study aimed to systematically review the evidence supporting the effectiveness of VR technology in improving the social skills of children and adolescents with ASD.

Methods: The search for eligible studies encompassed 4 databases: PubMed, Web of Science, IEEE, and Scopus. Two (XY and JW) researchers independently assessed the extracted studies according to predefined criteria for inclusion and exclusion. These researchers also independently extracted information regarding gathered data on the sources, samples, measurement methods, primary results, and data related to the main results of the studies that met the inclusion criteria. The quality of the studies was further evaluated using the Physiotherapy Evidence Database scale.

Results: This review analyzed 14 studies on using VR technology interventions to improve social skills in children and adolescents with ASD. Our findings indicate that VR interventions have a positive effect on improving social skills in children and adolescents with ASD. Compared with individuals with low-functioning autism (LFA), those with high-functioning autism (HFA) benefited more from the intervention. The duration and frequency of the intervention may also influence its effectiveness. In addition, immersive VR is more suitable for training complex skills in individuals with HFA. At the same time, nonimmersive VR stands out in terms of lower cost and flexibility, making it more appropriate for basic skill interventions for people with LFA. Finally, while VR technology positively enhances social skills, some studies have reported potential adverse side effects. According to the quality assessment using the Physiotherapy Evidence Database scale, of the 14 studies, 6 (43%) were classified as high quality, 4 (29%) as moderate quality, and 4 (29%) as low quality.

Conclusions: This systematic review found that VR technology interventions positively impact social skills in children and adolescents with ASD, with particularly significant effects on the enhancement of complex social skills in individuals with HFA. For children and adolescents with LFA, progress was mainly observed in basic skills. Immersive VR interventions are more suitable for the development of complex skills. At the same time, nonimmersive VR, due to its lower cost and greater flexibility, also holds potential for application in specific contexts. However, the use of VR technology may lead to side effects such as dizziness, eye fatigue, and sensory overload, particularly in immersive settings. These potential issues should be carefully addressed

XSL•FO
RenderX

in intervention designs to ensure user comfort and safety. Future research should focus on optimizing individualized interventions and further exploring the long-term effects of VR interventions.

Trial Registration: International Platform of Registered Systematic Review and Meta-analysis Protocols INPLASY202420079U1; https://inplasy.com/inplasy-2024-2-0079/

(J Med Internet Res 2025;27:e60845) doi:10.2196/60845

KEYWORDS

VR technology; autism spectrum disorder; children; adolescents; social skills; virtual reality; VR

Introduction

Background

Autism spectrum disorder (ASD) is a neurodevelopmental condition marked by challenges in social interaction and communication, as well as by the presence of repetitive and stereotyped interests and behaviors [1]. The precise etiology of ASD remains elusive; however, studies indicate that a combination of genetic, environmental, and neurodevelopmental factors may contribute to its development [2]. Statistics from the US Centers for Disease Control and Prevention have shown a marked rise in the prevalence of ASD diagnoses in recent decades, with 1 in 77 children currently being diagnosed [3]. Social impairment is often characterized by difficulty interpreting social cues, including changes in facial expressions, body language, and tone of voice. This can lead to children and adolescents with ASD displaying a reduced interest in, or maladaptive responses to, social situations [4]. Currently, early intervention is an effective way to improve the social skills of children and adolescents with ASD [5,6].

Individuals with ASD experience significant deficits in social skills. These deficits manifest as difficulties in understanding others' emotions and intentions, establishing and maintaining social relationships, and lacking flexibility in social interactions [7]. Such challenges often lead to increased social isolation and a decreased quality of life for those with ASD [8]. The severity of ASD symptoms varies; individuals with high-functioning autism (HFA) usually experience milder social impairments, while those with low-functioning autism (LFA) face more severe communication challenges [9]. Various interventions have been developed to address social skill deficits in individuals with ASD, achieving varying levels of success. For example, applied behavior analysis (ABA) is a therapeutic approach designed to improve social skill acquisition and enhance social motivation among individuals with ASD. Research by Roane et al [10] suggests that individualized, ABA-based interventions can significantly improve developmental outcomes for individuals with ASD, leading to progressive enhancements in social skills [11]. In addition, speech therapy offers personalized language and communication support, helping individuals gradually improve social communication skills. In the Social Communication Intervention Project (SCIP) led by Adams et al [12], it was found that speech therapy could enhance expressive language and comprehension, enabling individuals to gradually improve social skills by practicing conversations, learning communication techniques, and using assistive tools [13]. Cognitive behavioral therapy (CBT) is also widely used to address critical symptoms and socio-emotional challenges in children and adolescents with ASD. A meta-analysis by Wang et al [14] revealed significant improvements in both autistic symptoms and socio-emotional issues following CBT. Despite the considerable benefits of these traditional interventions, they have certain limitations. The high costs and the need for specific therapeutic environments and specialized therapists make these services inaccessible for many families, limiting their reach and frequency [15,16]. In addition, the lack of personalization in traditional methods may impact their effectiveness, especially for individuals with varying needs and responses [17]. These limitations underscore the importance of exploring alternative, more adaptable, individualized, and cost-effective intervention strategies.

Virtual reality (VR) technology, as a promising tool for promoting health, has gradually been applied in various fields, including gaming, education, health care, and psychotherapy. In gaming and entertainment, VR technology provides users with a highly immersive experience, allowing them to be fully immersed in a virtual world, which enhances interactivity and engagement [18,19]. VR simulates laboratories and historical scenes in education, enabling students to engage in hands-on learning within a safe virtual environment, thereby promoting deeper understanding and knowledge retention [20]. VR has been widely used in health care for rehabilitation, pain management, and psychotherapy [21,22]. Research by Rizzo and Koenig [23] suggests that VR technology provides a highly personalized interactive environment for psychotherapy and cognitive rehabilitation, helping patients better engage in treatment. Moreover, VR aids motor rehabilitation, particularly for patients with Parkinson disease and children with balance disorders, significantly enhancing recovery outcomes [24]. VR interventions can be categorized into immersive and nonimmersive types, depending on the realism of the simulated scenes [25]. Immersive VR typically uses head-mounted displays to fully engage the user's senses, offering a highly realistic experience, thus making it particularly suitable for interventions requiring deep involvement [26]. In contrast, nonimmersive VR is experienced through traditional computer displays, offering greater flexibility and accessibility, thus making it more appropriate for everyday educational settings [27]. For ASD interventions, Parsons and Cobb [28] conducted an in-depth study on the application of VR technology, noting that VR can provide a safe virtual environment for children with ASD to practice social skills without real-world pressures. Therefore, VR technology holds significant potential as an innovative and adaptable intervention for addressing social skill deficits in children and adolescents with ASD.

XSL•FO
RenderX

Objective

Although VR technology has shown great potential in the fields of education and therapy, especially in improving social skills in children and adolescents with ASD, empirical research remains limited. Therefore, this systematic review aims to evaluate the effects of VR interventions in ASD treatment and address the following key research questions: (1) What are the differences in the positive effects of VR interventions on the social skills of children and adolescents with HFA versus LFA? (2) How do different types of VR interventions impact the social skills of children and adolescents with ASD? and (3) Are there potential adverse effects of VR interventions on children and adolescents with ASD?

To answer these questions, this study aims to achieve the following objectives: (1) to compare the positive effects of VR interventions on the social skills of children and adolescents with HFA and LFA, (2) to analyze the impact of different types of VR interventions on the social skills of children and adolescents with ASD, and (3) to evaluate the potential adverse effects of VR interventions on children and adolescents with ASD.

Methods

Recruitment

Our research protocol was registered in the International Platform of Registered Systematic Review and Meta-analysis Protocols (ID: INPLASY202420079U1). This study followed the PRISMA (Preferred Reporting Items for Systematic Evaluation) guidelines to ensure transparency in the research process [29,30].

Search Strategy

A comprehensive search was conducted across 4 electronic databases: PubMed, Web of Science, IEEE, and Scopus, as of April 11, 2024, to systematically identify pertinent studies concerning the impact of VR technology on improving the social abilities of children and adolescents with ASD. The search focused explicitly on peer-reviewed articles published in English. The search strategy incorporated keywords such as VR, ASD, social skills, communication skills, children, and adolescents. Multimedia Appendix 1 provides detailed information about the search terms used. Previous reviews have described the selection of search terms for VR interventions, social skills, and children and adolescents with ASD [31]. In addition, researchers manually searched the included studies' reference lists to identify relevant articles.

Eligibility Criteria

The criteria for inclusion in the study were developed using the PICOS (Population, Intervention, Comparison, Outcomes, and Study design) principles [32], as follows:

- Population: The participants were children and adolescents aged 3-18 years, diagnosed by a physician or by a reliable ASD assessment tool. Mixed studies on children and adolescents with attention deficit/hyperactivity disorder, Down syndrome, or other neurodevelopmental disorders were excluded.
- Intervention: This study investigated the implementation of VR technology as an intervention in clinical settings or clinical trials.
- Comparison: No restrictions were placed on the control group.
- Outcomes: All studies used a validated tool or methodology to assess the social competence of children and adolescents with ASD.
- Study design: There were no restrictions on trial design, including randomized and nonrandomized controlled trials.

Study Selection and Data Extraction

Two researchers independently conducted a multistep search and assessed the studies based on the titles, abstracts, and full texts. To evaluate the risk of bias in the included studies, we used the Cochrane Risk of Bias tool [33]. This tool evaluates bias across several domains, including random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias. Two (XY and JW) researchers independently assessed each study, resolving disagreements through discussion or consulting a third researcher [34].

The team created a data extraction form to collect information about each study's characteristics, including details of the literature (authors and year), participant characteristics (sample size and age range), intervention components (location, intervention design, frequency, and duration), measurement instruments, and data related to primary outcomes. To increase transparency, we have included the PRISMA checklist in the Multimedia Appendix 2. This checklist provides a comprehensive overview of the methodology followed throughout the study, ensuring all relevant criteria are addressed and documented. This addition will allow readers to understand our research process better and replicate it.

Quality Assessment

The methodological quality of each included study was independently assessed by 2 researchers according to the Physiotherapy Evidence Database (PEDro) scale [35]. The decision to use the PEDro scale was based on its widespread use and recognition in assessing the quality of clinical trials, particularly in physiotherapy and rehabilitation [35]. Despite the challenges posed by VR studies, the PEDro scale remains a valuable tool due to its comprehensive assessment of key methodological aspects, such as randomization and allocation concealment, which are critical for reducing bias in clinical research [36].

The PEDro scale consists of 11 scoring criteria: eligibility criteria (not scored), random allocation, concealed allocation, baseline comparability, blinding of subjects, blinding of therapists, blinding of assessors, sufficient follow-up, intention-to-treat analysis, between-group comparisons, and point estimates with variability [36]. PEDro scores ranged from 0 to 10, with a median score of 5. However, scoring individuals and therapists participating in VR technology trials can be challenging, and blinding participants or therapists in VR studies is often not feasible due to the nature of the technology, as both

XSL•FO
RenderX

parties are usually aware of the ongoing intervention [37]. Therefore, considering the limitations of using VR technology, the scoring system can be categorized into 3 categories: high quality ≥6, appropriate quality=4-5, and low quality ≤3 [38]. In total, 2 (XY and JW) researchers independently assessed the methodological quality of the included studies according to the 11 PEDro criteria and calculated a total score to determine the quality of the study [39].

Disagreements regarding quality ratings should be deliberated until a consensus is achieved; if consensus proves elusive, a third researcher makes the final decision following a mutual discussion.

Results

Search Results

Figure 1 illustrates the specific search process, during which we initially retrieved 724 relevant studies: 30 from PubMed, 236 from Web of Science, 262 from IEEE, 196 from Scopus, and 5 additional studies (manual screening of reference lists and consultations with experts in the field). After excluding duplicates, 532 studies were included in the final analysis. First, 396 irrelevant studies were excluded based on title and abstract screening. The researchers removed 91 incomplete studies, leaving 45 studies for analysis. The final exclusion of reports and abstracts, studies lacking trial data, books, patents, and non-English language studies resulted in the inclusion of 14 studies.

Figure 1. The flow chart of literature screening.

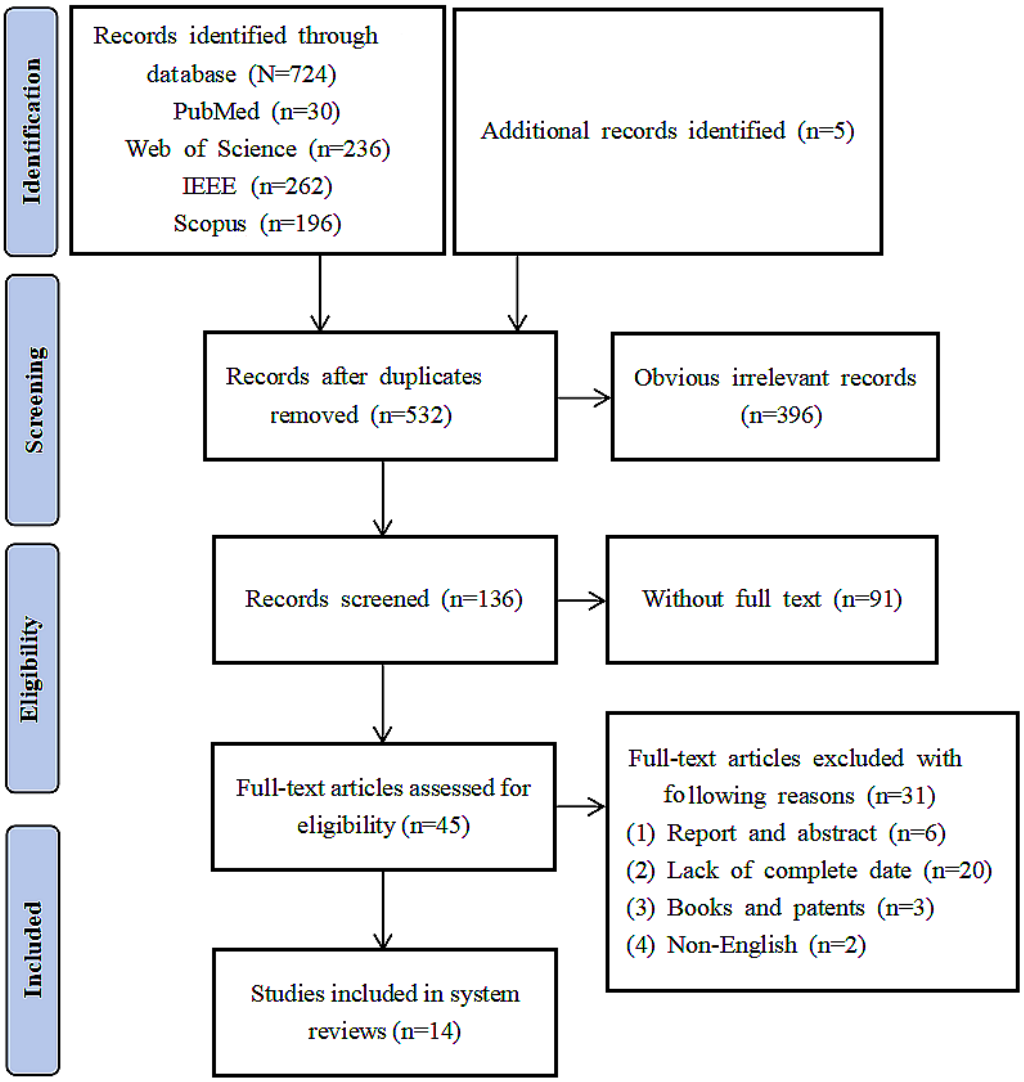

Research Characteristics

A total of 14 relevant studies were included in this review [40-53], with detailed characteristics of each study presented in Tables 1 and 2.

XSL•FO
RenderX

Table 1. Study characteristics.

Author	Country or region	Age range (year)	Diagnostic methods	Sample size, n (IG[a] and CG[b])	HFA[c] or LFA[d]	Setting
Matsentidou and Poullis [40]	Cyprus	9-10	NR[e]	12	HFA	Clinical
Cheng et al [41]	Taiwan	10-13	NR	3	HFA	School
Ip et al [42]	Hong Kong	6-12	NR	72 (36 and 36)	HFA	School
Rosenfield et al [43]	United States	NR	NR	2	HFA	School
Junaidi et al [44]	Indonesia	13-18	NR	7	LFA	Clinical
Yeh and Meng [45]	Taiwan	7-14	NR	10	HFA	School
Tsai et al [46]	Taiwan	7-9	DSM-5[f]	3	HFA	Clinical
Soltani Kouhbanani et al [47]	Iran	6-12	DSM-5	30 (15 and 15)	HFA and LFA	Clinical
Moon and Ke [48]	United States	10-14	NR	15	HFA	School
Elkin et al [49]	United States	6-17	DSM-5	10	HFA and LFA	Clinical
Frolli et al [50]	Italy	9-10	DSM-5	60 (30 and 30)	HFA	Clinical
Ke et al [51]	United States	10-14	NR	7	HFA	Clinical
Zhao et al [52]	China	3-5	DSM-5	44 (22 and 22)	HFA and LFA	Clinical
Moon and Ke [53]	United States	12-13	DSM-5	4	HFA	Clinical

[a]IG: intervention group.

[b]CG: control group.

[c]HFA: high-functioning autism.

[d]LFA: low-functioning autism.

[e]NR: not reported.

fDSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

XSL•FO
RenderX

Table 2. Intervention information and main findings

References	Immersive or nonimmersive	Intervention	Frequency (length)	Type	Measures	Main findings
Matsentidou and Poullis [40]	Immersive	CAVE[a]	NR[b]	Chronic	Behavior observation	Results are "encouraging"
Cheng et al [41]	Immersive	3D-SU[c]	6 wk	Chronic	Social events card; social behaviors scale	Improved nonverbal communication, social initiations, and social cognition behaviors
Ip et al [42]	Immersive	CAVE	2/wk; 14 wk	Chronic	Psychoeducational Profile (Third Edition)	Social interactions ($P<.05$); emotional expression ($P<.01$)
Rosenfield et al [43]	Immersive	VR[d]	NR	Chronic	Behavior observation	To help effectively
Junaidi et al [44]	Nonimmersive	VR	NR	Acute	Behavior observation	Supports shopping skill development effectively
Yeh and Meng [45]	Nonimmersive	VR-SST[e]	1/wk; 8 wk	Chronic	Social skills evaluation sheet	Answer speed ($P<.05$); conversation and expression ($P<.05$); the completeness of sentence-structure ($P<.05$); conversation etiquette ($P<.05$)
Tsai et al [46]	Immersive	CAVE	5 wk	Chronic	Social Skills Training Scales	TPP-RPG[f] system improved social skills ($P<.01$), maintained after intervention
Soltani Kouhbanani et al [47]	Nonimmersive	VR	7/wk; 12 wk	Chronic	Vineland Adaptive Behavior Scale	Social skills ($P<.01$); behavioral symptoms ($P<.01$)
Moon and Ke [48]	Nonimmersive	VR-SST	8 wk	Chronic	Jaccard Index	Treatment integrity scores improved ($P<.05$)
Elkin et al [49]	Immersive	VR-SC[g]	1 time	Chronic	Behavior observation	Enhances propensity for social interaction through gaze behavior analysis
Frolli et al [50]	Immersive	3D projection of primary emotions	3/wk; 12 wk	Chronic	Image testing	Primary emotions ($P=.25$); emotions and situations for primary emotions ($P<.05$)
Ke et al [51]	Nonimmersive	VR-SST	2/wk; 15 wk	Chronic	Simulator Sickness Questionnaire; Social Communication Questionnaire	Simulator Sickness Questionnaire Self ($P>.05$); Simulator Sickness Questionnaire Parent ($P>.05$); Social Communication Questionnaire ($P>.05$)
Zhao et al [52]	Nonimmersive	VR	3/wk; 12 wk	Chronic	Psychoeducational Profile (Third Edition)	Social interaction ($P<.01$); emotional expression ($P<.01$)
Moon and Ke [53]	Nonimmersive	VR-SST	3/wk; 8 wk	Chronic	Simulator Sickness Questionnaire	Social-interaction initiation ($P<.05$); responding to social interactions ($P>.05$); interpersonal negotiation ($P<.05$); self-identity ($P>.05$); Simulator Sickness Questionnaire Self ($P>.05$)

[a]CAVE: cave automatic virtual environment.

[b]NR: not reported.

[c]3D-SU: 3D Immersive Virtual Environment System to Enhance Social Understanding.

[d]VR: virtual reality.

[e]VR-SST: virtual reality social skills training.

[f]TPP-RPG: third-person perspective role-playing game.

[g]VR-SC: virtual reality safe course.

The 14 studies included 279 children and adolescents with ASD aged 3-18 years [40-53]. Six studies used the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) to diagnose children and adolescents with ASD [46,47,49,50,52,53], while 8 studies had clinical diagnoses made by physicians (Table 1) [40,45,48,51].

Of these 14 studies, 5 were from the United States [43,48,49,51,53]; 1 from Hong Kong [42]; 1 from Iran [47]; 1 from Cyprus [40]; 1 from Indonesia [44]; 3 from Taiwan, China [41,45,46]; 1 from Italy [50]; and 1 from China [52]. In terms of the trial design, 1 study used a single case trial [44,50]; 4 used a randomized controlled trial [42,46,50,52]; and 9 used a single-group, before-and-after controlled trial [40,41,43,45,46,48,49,51,53] (Table 1).

Regarding trial sites, 9 studies were conducted in institutions or clinics [40,44,46,47,49-53], and 5 were conducted in homes or schools (Table 1) [41-43,45,48].

XSL•FO
RenderX

In terms of intervention modalities, 3 studies used a cave automatic virtual environment [40-42,46], 1 study used a 3D Immersive Virtual Environment System to Enhance Social Understanding [41], 4 studies used VR Social Skills Training [45,48,51,53], 1 study used VR SAFE Course [49], 1 study used a 3D projection of primary emotions [50], and 5 studies did not report the specific VR technology and content used (Table 2) [43,44,47,50,52].

The 14 studies had a minimum intervention period of 1 day [49] and a maximum of 15 weeks [51]. Notably, 3 studies did not report detailed information on the duration of the intervention (Table 2) [40,43,44].

To assess social competence, 2 studies used the Psychoeducational Profile (Third Edition) [42,52], 2 used the Simulator Sickness Questionnaire [51,53], 1 used the Social Events Card and Social Behaviors Scale [41], 1 used the Vineland Adaptive Behaviour Scale [47], and 1 used the Jaccard Index scores [48]. One study used image testing [50], 1 study used the Social Communication Questionnaire [51], 1 study used Social Skills Training scales and the Likert Scale [46], 1 study used speech data mining [53], 1 study used Social Skills Performance Evaluation Sheet [45], and 4 studies used behavioral observation and analysis (Table 2) [40,43,44,49].

Quality Evaluation

Details of the quality assessments of the included studies are presented in Table 3. The 14 studies had an average PEDro quality (overall study quality) score of 5.21, of which 6 had an overall study quality score of ≥7. All 14 studies met the eligibility criteria and point measures criteria, reflecting a general consistency in study design and a general focus on outcome assessment. Random allocation, concealed allocation, and between-group comparison were implemented in 6 studies each. In contrast, a similar-at-baseline comparison was implemented in 10 studies, demonstrating that a proportion of the studies attempted to adopt the methodological criteria of randomization and control in their design to enhance the validity and credibility of the findings.

No studies implemented subject blinding, therapist blinding, or assessor blinding simultaneously, suggesting that the use of blinding was very limited in these studies, which may be due to the specific design and implementation settings of the studies. Most studies (n=11) reported a participant attrition dropout rate. Still, they did not perform intention-to-treat analyses, which may have affected the interpretation and generalization of the findings.

XSL•FO
RenderX

Table 3. Methodological quality assessment of included studies.

References	Eligibility criteria	Random allocation	Concealed allocation	Similar at baseline	Subject blinded	Therapist blinded	Assessor blinded	Dropout rate	Intention-to-Treat analysis	Between-group comparison	Point measures	Test of statistical significance	Overall study quality
Matsentidou and Poullis [40]	1[a]	0[b]	0	0	0	0	0	0	0	0	1	0	2
Cheng et al [41]	1	0	0	0	0	0	0	1	0	0	1	1	4
Ip et al [42]	1	1	1	1	0	0	1	1	0	1	1	1	9
Rosenfield et al [43]	1	0	0	0	0	0	0	0	0	0	1	0	2
Junaidi et al [44]	1	0	0	1	0	0	0	1	0	0	1	0	4
Yeh and Meng [45]	1	1	1	1	0	0	0	1	0	1	1	0	7
Tsai et al [46]	1	0	0	1	0	0	0	0	0	0	1	0	3
Soltani Kouhbanani et al [47]	1	0	0	0	0	0	0	0	0	0	1	0	2
Moon and Ke [48]	1	1	1	1	1	0	0	1	0	1	1	1	9
Elkin et al [49]	1	1	1	1	0	1	0	1	0	1	1	0	8
Frolli et al [50]	1	1	1	1	0	0	0	1	0	1	1	1	8
Ke et al [51]	1	1	1	1	0	0	0	1	0	1	1	0	7
Zhao et al [52]	1	0	0	1	0	0	0	1	0	0	1	0	4
Moon and Ke [53]	1	0	0	1	0	0	0	1	0	0	1	0	4

[a]1: Yes.

[b]0: No.

Main Findings

Impact of VR Interventions on the Social Skills of Children and Adolescents With ASD

Among the 14 studies included [40-53], 71% (n=10) of the participants were individuals with HFA [40-43,45,46,48,50,51,53], 21% (n=3) included both participants with HFA and LFA [47,49,52], and 7% (1) involved only participants with LFA (Table 1) [44]. We found that VR interventions significantly improved the social skills of children and adolescents with HFA, while the improvements for participants with LFA were more limited (Table 4).

XSL•FO
RenderX

Table 4. The main results and conclusions of the included studies.

References	Main results	Adverse effects	Conclusions
Matsentidou and Poullis [40]	75% success in street crossing	Yes	VR[a] improves social skills
Cheng et al [41]	Increases in initiation, cognition, non-verbal	NR[b]	3D-SU[c] enhances social initiation skills
Ip et al [42]	Significant improvements in emotion and social interaction	Yes	VR enhances emotional regulation and social skills
Rosenfield et al [43]	Reduced response time in conversation	Yes	VR facilitates conversation practice
Junaidi et al [44]	Moderate effectiveness; high readability	Yes	VR moderately effective for clarity
Yeh and Meng [45]	Improved answer speed and conversation skills	NR	VR enhances conversation etiquette
Tsai et al [46]	Increased correct social responses	Yes	TPP-RPG[d] improves social reciprocity
Soltani Kouhbanani et al [47]	Maintained social improvements at follow-up	NR	VR + risperidone enhances social behavior
Moon and Ke [48]	Enhanced initiation and negotiation skills	NR	Enhanced initiation and negotiation skills
Elkin et al [49]	Gaze behavior varies with ASD[e] severity	NR	VR differentiates gaze behavior by severity
Frolli et al [50]	Faster emotion recognition	NR	VR accelerates social skills acquisition
Ke et al [51]	Improved social initiation and flexibility	Yes	VR enhances social interaction
Zhao et al [52]	Increased cognitive and social scores	NR	VR + rehab boosts cognitive or social skills
Moon and Ke [53]	Improved social interaction scores	NR	VR prompts improve social skills

[a]VR: virtual reality.

[b]NR: not reported.

[c]3D-SU: 3D Immersive Virtual Environment System to Enhance Social Understanding.

[d]TPP-RPG: Third-Person Perspective Role-Playing Game.

[e]ASD: autism spectrum disorder.

Studies focusing on HFA consistently demonstrated that VR social skills training systems significantly improved the social skills of children and adolescents with HFA [40-43,45-53]. Some studies also found that VR training effectively enhanced eye contact and the quality of social interactions in individuals with HFA [49]. Other research indicated that VR training helped HFA children and adolescents improve cognitive flexibility, social initiative, and conflict resolution skills [45,49,53]. In comparison, the effects of VR interventions on children and adolescents with LFA were less pronounced [40-53]. While VR-based shopping skills training was found to help individuals with LFA improve basic life skills to some extent, the overall progress was slower, and the degree of improvement was not as significant as in individuals with HFA [44]. Although some studies found that LFA participants did not perform as well as participants with HFA in VR settings, others still reported that VR interventions helped them progress in self-care skills [47,52]. Furthermore, the research emphasized that continuous improvements in adaptive prompts and personalized training could offer more support for individuals with LFA [52].

Impact of Different VR Intervention Approaches on the Social Skills of Children and Adolescents With ASD

Of the included studies, 64% (9/14) were conducted in institutional or clinical settings [40,44,46,47,49-53], while 36% (5/14) were conducted in-home or school environments (Table 1) [41-43,45,48]. We found that interventions conducted in institutional or clinical settings generally had greater control and professional oversight (Table 4) [40,44,46,47,49-53]. These settings often had specialized equipment and supervision by therapists, resulting in more significant intervention effects, which were better suited for highly customized interventions for children and adolescents with ASD [40,44,46]. However, the primary advantages of home- or school-based interventions were their convenience and real-world applicability [42,45]. Although these interventions had relatively less control, they still demonstrated benefits in terms of flexibility and cost-effectiveness [41-43,45,48].

Regarding intervention duration and frequency, 79% (11/14) of the studies reported this information [41,42,45-53], while 21% (3/14) did not [40,43,44]. The intervention duration ranged from as short as 1 day to as long as 15 weeks (Table 2) [49,51]. Compared to short-term interventions, we found that long-term interventions led to more significant improvements in social skills and cognitive flexibility (Table 4) [42,45-53]. In addition, higher-frequency interventions generally produced more noticeable improvements, while lower-frequency interventions, although beneficial, did not yield results as strong as high-frequency ones [52,53]. Lower-frequency interventions may not sufficiently reinforce the skills being taught, leading to weaker transfer effects [45]. Overall, the most effective intervention duration was found to be 6 to 15 weeks, with a frequency of at least 2-3 sessions per week [42,45-53]. This high-frequency, long-term intervention model provided participants ample opportunities to practice, ensuring effective social skills development and long-term retention [42,45-53].

Of the studies, 50% (7/14) used immersive VR technology [40-43,46,49,50], while 50% (7/14) used nonimmersive VR

XSL•FO
RenderX

systems (Table 2) [44,45,47,48,51-53]. We found that immersive VR was more effective for complex social skills training, especially in participants with HFA (Table 4) [40-43,46,49,50]. Immersive VR technology, by providing rich multisensory stimulation, allowed participants to better grasp emotional expression and social interaction skills within simulated environments [40]. In more complex school scenarios, immersive VR better supports participants in practicing and transferring learned skills to real-life situations [42]. Finally, immersive VR training experiences received positive feedback, and no instances of motion sickness were reported [43]. While nonimmersive VR systems, such as desktop-based VR, did not offer fully immersive experiences, they still demonstrated advantages in terms of flexibility, cost, and adaptability [44,45,47,48,51-53].

Adverse Effects of VR Interventions on Children and Adolescents With ASD

Although VR technology significantly improved social skills, 6 studies reported adverse side effects associated with its use (Table 4) [40,42-44,46,51]. One study found that some children and adolescents with HFA experienced dizziness and nausea while using the cave automatic virtual environment system [40]. Another study reported that 71% (10/14) of children and adolescents with LFA experienced anxiety and fear when using head-mounted displays for the first time, and most participants felt fatigued after using VR devices [44]. In addition, 2 other studies also reported anxiety-related issues [46,51]. While overall, the use of VR technology for social skills training was positive, some children reported discomfort, although no cases of motion sickness were observed [43]. Another study indicated that children and adolescents with HFA experienced headaches and eye fatigue, particularly after prolonged use of VR devices [42].

Discussion

Principal Findings

This systematic review analyzed the effects of VR interventions on the social skills of children and adolescents with ASD. The results showed that VR interventions significantly impacted individuals with HFA, while the effects on those with LFA were more limited. The study also indicated that the duration and frequency of the interventions had a notable impact on their effectiveness. Moreover, immersive VR was found to be more suitable for training complex skills in individuals with HFA. At the same time, nonimmersive VR excelled in cost and flexibility, making it more appropriate for basic skill interventions in individuals with LFA. Finally, although VR interventions showed positive outcomes in improving social skills, some studies reported potential side effects, highlighting the need to account for individual differences when designing intervention programs and implementing measures to reduce adverse effects.

Impact of VR Interventions on the Social Skills of Children and Adolescents With ASD

The differences in the effectiveness of VR interventions for children and adolescents with HFA and LFA can be attributed to variations in cognitive, social, and sensory processing skills [40-43,45,46,48,50,51,53]. Individuals with HFA generally have higher cognitive abilities and language skills, enabling them to understand better and adapt to complex social situations within VR environments [54]. This finding aligns with earlier studies that highlighted the need for more personalized intervention strategies based on ASD symptom severity [9-11]. These previous studies suggested that traditional interventions like ABA and CBT were more effective for high-functioning individuals due to their adaptability and cognitive skills, which is similarly observed in VR settings [9-11]. The advantages demonstrated by children and adolescents with HFA in VR interventions are partly due to their stronger abstract thinking skills and higher sensory tolerance [55]. Multisensory stimuli, such as the combination of visual, auditory, and tactile inputs, activate key brain regions, including the mirror neuron system and the prefrontal cortex, further enhancing social interaction and emotional understanding [56]. In addition, individuals with HFA are better at handling contextual changes within VR environments, adapting to feedback mechanisms in virtual scenarios, allowing them to quickly learn and transfer skills to real-life situations [42]. This adaptability explains why children with HFA generally exhibit faster progress in VR interventions. In contrast, individuals with LFA's slower cognitive and language development hinders their progress in VR interventions [44,47,49,52]. This limitation reflects findings from traditional interventions, where children with LFA often progressed more in structured, basic tasks than in complex social or emotional skills [12,13,49]. Children with LFA often struggle with processing multisensory stimuli in virtual environments, making them prone to discomfort, such as sensory overload or anxiety reactions [44,57]. This sensory sensitivity confuses complex social scenarios, making it difficult for them to keep pace with the intervention [58]. Moreover, for individuals with LFA, cognitive and social difficulties limit their ability to master complex emotional expression or social interaction skills [49]. They tend to show progress in more structured, more straightforward tasks, such as life skills training or behavior management [44,52]. These limitations suggest that individuals with LFA require specialized support and simplified task designs to ensure gradual improvement in controlled settings.

Future interventions should be more personalized to bridge the gap in VR intervention outcomes between individuals with HFA and LFA [59]. For individuals with HFA, continued focus on complex social scenario training is essential. For individuals with LFA, interventions should emphasize managing sensory stimuli, reducing the risk of sensory overload, simplifying social interaction tasks, and offering more structured and incremental training content. The effectiveness of VR interventions is also significantly influenced by the setting in which they are conducted. Institutional and clinical environments generally offer specialized equipment and supervision by therapists, providing participants with highly customized support and real-time feedback [40,44,46,47,49-53]. These settings help ensure the quality and consistency of interventions, especially in emotional regulation and complex social skills training [60]. However, the limitations of institutional and clinical environments include high costs and limited scalability [40,47]. In contrast, home and school environments, although lacking

XSL•FO
RenderX

immediate professional supervision, offer convenience and real-world applicability advantages, making them more suitable for skill transfer to everyday activities [41-43,45,48]. In addition, home and school settings reduce participants' anxiety, as children are more relaxed and receptive to interventions in familiar environments [61]. The flexibility and accessibility of these environments make them suitable for long-term interventions, though they may compromise the depth and complexity of the intervention [61]. Future research should explore how to incorporate more professional support into home and school environments to ensure the effectiveness of interventions is not compromised.

Impact of Different VR Intervention Approaches on the Social Skills of Children and Adolescents With ASD

The duration and frequency of interventions significantly impact the effectiveness of VR interventions. Longer intervention durations and higher frequencies generally have more significant and lasting effects, particularly in social skills and emotional regulation [41,46-48]. This observation is consistent with the findings in traditional therapies, where prolonged and frequent sessions in interventions like ABA and CBT have been linked to more sustainable outcomes in social skills improvement [11,14]. Shorter intervention durations or lower frequencies may not be sufficient to consolidate the skills learned, limiting the maintenance of intervention effects [42,45,49-53]. Long-term, high-frequency interventions provide more practice opportunities, helping participants gradually master and transfer these skills [62]. However, high-frequency and long-term interventions may also lead to fatigue, particularly in individuals with LFA who are more sensitive to sensory overload [44]. Therefore, future intervention designs should balance the duration and frequency of interventions, ensuring sufficient practice opportunities while avoiding sensory fatigue and overstimulation in participants. Further research should explore the combination of optimal intervention duration and frequency to ensure both effectiveness and sustainability.

The differences in the effectiveness of immersive and nonimmersive VR interventions stem from their varying levels of sensory stimulation and interaction methods [63]. Immersive VR, with its multisensory stimulation and highly realistic virtual environments, is particularly effective for children with HFA, enhancing their adaptability in complex social interactions [40-43,46,49,50]. However, its intense sensory input may pose challenges for individuals with LFA, potentially causing anxiety or sensory overload [44,45,47,48,51-53]. As a lower-stimulus alternative, nonimmersive VR supports adaptive training for children with LFA, focusing on basic skills and behavior management [64]. Its ease of use and scalability make it suitable for broader and long-term applications, particularly in educational settings where lower-stimulus environments are often preferable [15,17]. Future research should prioritize integrating the strengths of immersive and nonimmersive VR to balance engagement and usability, ensuring interventions can address the diverse needs of individuals with ASD effectively.

Practical Implications and Adverse Effects

VR technology interventions hold significant potential for clinical and educational applications in children and adolescents with ASD [65]. In clinical practice, VR can serve as a complementary tool to traditional interventions, providing therapists with a flexible and engaging means to meet the diverse needs of individuals with ASD [66]. For instance, the benefits of VR as an adjunct to speech therapy align with findings from the SCIP intervention in traditional therapy, where tailored approaches proved beneficial for improving social communication skills in children with ASD [12]. For instance, traditional speech therapy interventions, such as SCIP, have been shown to effectively improve social communication skills in children with ASD [12]. Integrating VR into these interventions could enhance the effects of traditional speech therapy, offering more engaging and personalized treatment sessions [67]. In addition, a meta-analysis of CBT's effectiveness in addressing social-emotional challenges in children with ASD has shown significant improvements in core symptoms following CBT [14]. By incorporating VR into CBT, children can practice coping strategies in simulated real-life emotional and social scenarios, potentially enhancing outcomes [68]. Finally, VR could be combined with ABA to create personalized social scenarios, allowing therapists to offer more dynamic and adaptable intervention plans while children practice social behaviors in realistic but controlled environments [11,69].

Despite the many advantages of VR interventions, there are challenges to its widespread application, particularly regarding accessibility, cost, and the training required for educators and therapists to implement these technologies [70,71] effectively. Previous studies on traditional therapy have also highlighted these issues, where high costs and professional training needs limit scalability [15,16]. Adequate training for therapists and educators is crucial to maximizing the potential of VR interventions [72]. Research has highlighted that while immersive VR offers powerful therapeutic effects, there are challenges related to user comfort and sensory overload, especially for children with ASD. It is essential to carefully manage symptoms such as dizziness, nausea, and sensory overload, emphasizing the importance of gradual exposure in VR-based interventions [73]. In addition, studies have reported adverse effects such as eye fatigue, headaches, anxiety, and disorientation during VR use [40,42-44,46,51]. These reactions are particularly concerning for children with ASD, who may exhibit heightened sensitivity to sensory stimuli [74]. For instance, immersive VR systems with complex visual and auditory elements have been found to increase the risk of dizziness and nausea in approximately 15% of users [75]. Such adverse effects may reduce engagement and limit VR interventions' scalability [76]. To address these challenges, intervention designs must incorporate strategies to minimize adverse effects. These include shortening session durations, integrating regular breaks, and adapting VR content to reduce sensory overload [77]. For example, designing environments with simpler visual elements or offering less immersive options for highly sensitive children could significantly mitigate these effects [78]. Furthermore, providing psychological support during interventions, such as teaching coping mechanisms for

XSL•FO
RenderX

anxiety, may enhance user comfort and overall feasibility [79]. Future research should also systematically evaluate the prevalence and impact of these adverse effects, offering data-driven recommendations for optimizing VR-based interventions.

Moving forward, research should focus on improving the affordability and user-friendliness of VR interventions, ensuring that they are not limited to professional settings but can be widely applied in both clinical and educational environments [80].

Limitations

Despite the many strengths of this study, several limitations require further discussion. The sample size of this systematic review is relatively small, including only 14 studies, which may limit the generalizability of the findings. The included studies used various scales and tools to assess the social skills of children and adolescents with ASD, and this diversity in assessment tools may have led to inconsistencies in the pooled results. Due to differences in evaluation tools and reported data types, this study could not provide a quantitative summary. There was significant heterogeneity in intervention protocols, including variations in VR content, session length, and frequency. This lack of standardization limits comparisons between studies and highlights the need for uniform intervention protocols in future research. Many of the included studies had small sample sizes, which reduced statistical power and increased the likelihood of underestimating the effects of VR interventions. Future research should recruit larger and more diverse samples to enhance the reliability and generalizability of findings. Finally, some studies lacked detailed reporting of key methodological elements, such as randomization and blinding, which increases the potential for bias. Adopting rigorous designs and transparent reporting practices, guided by frameworks such as PRISMA or CONSORT (Consolidated Standards of Reporting Trials), will help strengthen the evidence base for future studies. By addressing these limitations, this study underscores the importance of more robust and standardized research to advance the field of VR interventions for ASD.

Conclusion

This systematic review highlights the significant potential of VR technology interventions in enhancing social skills for children and adolescents with ASD (Textbox 1).

In conclusion, long-term, high-frequency VR interventions (especially those lasting 6 to 15 weeks with 2-3 sessions per week) demonstrate the most effectiveness. Future research should focus on optimizing VR interventions to meet the personalized needs of different subgroups of patients with ASD and conduct large-scale longitudinal studies to validate their long-term impact.

Textbox 1. Key insights on virtual reality (VR) interventions for children and adolescents with autism spectrum disorder (ASD).

- Differences in VR intervention effects for individuals with high-functioning autism (HFA) and low-functioning autism (LFA): VR interventions show considerable improvements in complex emotional regulation and social interaction skills for children and adolescents with HFA. In contrast, improvements for individuals with LFA have been relatively limited, primarily supporting basic life skills. This suggests the need for more tailored VR approaches for different ASD subgroups.
- Impact of different VR intervention types: Immersive VR proves more effective for enhancing complex social skills, particularly in children with HFA, as it provides a deeply engaging experience. Meanwhile, due to its flexibility and lower cost, nonimmersive VR offers valuable benefits in specific contexts, particularly in home or school settings. Institutional and clinical settings have shown more specialized and effective results, whereas interventions in home or school environments provide greater adaptability.
- Potential adverse effects of VR interventions: Although VR technology holds promise, adverse effects such as dizziness and fatigue may arise, especially during immersive VR sessions. These potential issues should be carefully addressed in the design of VR-based interventions.
- Practical recommendations for stakeholders: Therapists, educators, and developers should collaborate to optimize VR interventions for children with ASD. Therapists can integrate VR as a complementary tool to traditional therapies, tailoring interventions and gradually increasing exposure to reduce sensory overload. Educators can use VR for social skills training in controlled environments with proper training. Developers should focus on creating cost-effective, user-friendly VR platforms with customizable features to ensure accessibility in clinical and educational settings. These efforts will enhance the scalability and impact of VR interventions.

Acknowledgments

This research was supported by the Special Funding for Key Areas in Biomedicine and Health of Ordinary Colleges and Universities in Guangdong Province (2022ZDZX2086; 2024ZDZX2108) and the Humanities and Social Sciences Project of the Chinese Ministry of Education (24YJAZH106). Meanwhile, we express our sincere gratitude to Dr Zhongtang Wu, who serves in the College of Physical Education at Shenzhen University, for his invaluable technical expertise and assistance, which was paramount to the success of our research.

Data Availability

All data generated or analyzed during this study are included in this published paper (and Multimedia Appendices 1 and 2).

XSL•FO
RenderX

Authors' Contributions

ZR was responsible for the research protocol and critically redrafted the paper. XY and JW performed literature searches. XY and JW performed study identification and screening. XY, JW, YM, and YT extracted and analyzed the data from the included journals. XY and JW wrote the first draft of the paper. JY, HC, AZ, and RF, contributed to refining the ideas, carrying out additional analyses, and finalizing this paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search research strategy.
[DOCX File , 12 KB - jmir_v27i1e60845_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 22 KB - jmir_v27i1e60845_app2.docx]

References

1. First MB. Diagnostic and statistical manual of mental disorders, 5th edition, and clinical utility. J Nerv Ment Dis 2013;201(9):727-729. [doi: 10.1097/NMD.0b013e3182a2168a] [Medline: 23995026]
2. Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 2017;56(6):466-474. [doi: 10.1016/j.jaac.2017.03.013] [Medline: 28545751]
3. Pain O, Pocklington AJ, Holmans PA, Bray NJ, O'Brien HE, Hall LS, et al. Novel insight into the etiology of autism spectrum disorder gained by integrating expression data with genome-wide association statistics. Biol Psychiatry 2019;86(4):265-273 [FREE Full text] [doi: 10.1016/j.biopsych.2019.04.034] [Medline: 31230729]
4. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The heritability of autism spectrum disorder. JAMA 2017;318(12):1182-1184 [FREE Full text] [doi: 10.1001/jama.2017.12141] [Medline: 28973605]
5. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. Pediatrics 2011;128(3):e488-e495 [FREE Full text] [doi: 10.1542/peds.2010-2825] [Medline: 21844053]
6. Rogers S, Vismara LA. Evidence-based comprehensive treatments for early autism. J Clin Child Adolesc Psychol 2008;37(1):8-38. [doi: 10.1080/15374410701817808] [Medline: 18444052]
7. Mehling M, Tassé MJ. Severity of autism spectrum disorders: current conceptualization, and transition to DSM-5. J Autism Dev Disord 2016;46(6):2000-2016. [doi: 10.1007/s10803-016-2731-7] [Medline: 26873143]
8. Hirota T, King BH. Autism spectrum disorder: a review. JAMA 2023;329(2):157-168. [doi: 10.1001/jama.2022.23661] [Medline: 36625807]
9. Aldred C, Green J, Adams C. A new social communication intervention for children with autism: pilot randomised controlled treatment study suggesting effectiveness. J Child Psychol Psychiatry 2004;45(8):1420-1430. [doi: 10.1111/j.1469-7610.2004.00848.x] [Medline: 15482502]
10. Roane H, Fisher WW, Carr JE. Applied behavior analysis as treatment for autism spectrum disorder. J Pediatr 2016;175:27-32. [doi: 10.1016/j.jpeds.2016.04.023] [Medline: 27179552]
11. Weiss MJ, Harris SL. Teaching social skills to people with autism. Behav Modif 2001;25(5):785-802. [doi: 10.1177/0145445501255007] [Medline: 11573340]
12. Adams C, Lockton E, Freed J, Gaile J, Earl G, McBean K, et al. The social communication intervention project: a randomized controlled trial of the effectiveness of speech and language therapy for school-age children who have pragmatic and social communication problems with or without autism spectrum disorder. Int J Lang Commun Disord 2012;47(3):233-244. [doi: 10.1111/j.1460-6984.2011.00146.x] [Medline: 22512510]
13. Pennington L, Goldbart J, Marshall J. Speech and language therapy to improve the communication skills of children with cerebral palsy. Cochrane Database Syst Rev 2004;2004(2):CD003466 [FREE Full text] [doi: 10.1002/14651858.CD003466.pub2] [Medline: 15106204]
14. Wang X, Zhao J, Huang S, Chen S, Zhou T, Li Q, et al. Cognitive behavioral therapy for autism spectrum disorders: a systematic review. Pediatrics 2021;147(5):e2020049880. [doi: 10.1542/peds.2020-049880] [Medline: 33888566]
15. Sapiets SJ, Totsika V, Hastings RP. Factors influencing access to early intervention for families of children with developmental disabilities: a narrative review. J Appl Res Intellect Disabil 2021;34(3):695-711 [FREE Full text] [doi: 10.1111/jar.12852] [Medline: 33354863]

XSL•FO
RenderX

16. Maglione MA, Gans D, Das L, Timbie J, Kasari C, Technical Expert Panel, et al. Nonmedical interventions for children with ASD: recommended guidelines and further research needs. Pediatrics 2012;130 Suppl 2:S169-S178. [doi: 10.1542/peds.2012-0900O] [Medline: 23118248]
17. Lord C, Charman T, Havdahl A, Carbone P, Anagnostou E, Boyd B, et al. The Lancet Commission on the future of care and clinical research in autism. Lancet 2022;399(10321):271-334. [doi: 10.1016/S0140-6736(21)01541-5] [Medline: 34883054]
18. Wu J, Loprinzi PD, Ren Z. The rehabilitative effects of virtual reality games on balance performance among children with cerebral palsy: a meta-analysis of randomized controlled trials. Int J Environ Res Public Health 2019;16(21):4161 [FREE Full text] [doi: 10.3390/ijerph16214161] [Medline: 31661938]
19. Fodor L, Cote CD, Cuijpers P, Szamoskozi , David D, Cristea IA. The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: a meta-analysis. Sci Rep 2018;8(1):10323 [FREE Full text] [doi: 10.1038/s41598-018-28113-6] [Medline: 29985400]
20. Wu J, Ma Y, Ren Z. Rehabilitative effects of virtual reality technology for mild cognitive impairment: a systematic review with meta-analysis. Front Psychol 2020;11:1811 [FREE Full text] [doi: 10.3389/fpsyg.2020.01811] [Medline: 33101098]
21. Wu J, Sun Y, Zhang G, Zhou Z, Ren Z. Virtual reality-assisted cognitive behavioral therapy for anxiety disorders: a systematic review and meta-analysis. Front Psychiatry 2021;12:575094 [FREE Full text] [doi: 10.3389/fpsyt.2021.575094] [Medline: 34366904]
22. Beheshti J, Bilal D, Druin A, Large A. Testing children's information retrieval systems: Challenges in a new era. Proc. Am. Soc. Info. Sci. Tech 2011;47(1):1-4 [FREE Full text] [doi: 10.1002/meet.14504701222]
23. Rizzo AS, Koenig ST. Is clinical virtual reality ready for primetime? Neuropsychology 2017;31(8):877-899. [doi: 10.1037/neu0000405] [Medline: 29376669]
24. Wu J, Zhang H, Chen Z, Fu R, Yang H, Zeng H, et al. Benefits of virtual reality balance training for patients with parkinson disease: systematic review, meta-analysis, and meta-regression of a randomized controlled trial. JMIR Serious Games 2022;10(1):e30882 [FREE Full text] [doi: 10.2196/30882] [Medline: 35230242]
25. Bevilacqua R, Maranesi E, Riccardi GR, Donna VD, Pelliccioni P, Luzi R, et al. Non-immersive virtual reality for rehabilitation of the older people: a systematic review into efficacy and effectiveness. J Clin Med 2019;8(11):1882 [FREE Full text] [doi: 10.3390/jcm8111882] [Medline: 31694337]
26. Horvat N, Martinec T, Perišić M, Škec S. Comparing design review outcomes in immersive and non-immersive collaborative virtual environments. Procedia CIRP 2022;109:173-178 [FREE Full text] [doi: 10.1016/j.procir.2022.05.232]
27. Smith C, Read JE, Bennie C, Hale L, Milosavljevic S. Can non-immersive virtual reality improve physical outcomes of rehabilitation? Physical Therapy Reviews 2013;17(1):1-15 [FREE Full text] [doi: 10.1179/1743288x11y.0000000047]
28. Parsons S, Cobb S. State-of-the-art of virtual reality technologies for children on the autism spectrum. Eur. J. Spec. Needs Educ 2011;26(3):355-366 [FREE Full text] [doi: 10.1080/08856257.2011.593831]
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
30. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
31. Mesa-Gresa P, Gil-Gómez H, Lozano-Quilis J, Gil-Gómez J. Effectiveness of virtual reality for children and adolescents with autism spectrum disorder: an evidence-based systematic review. Sensors (Basel) 2018;18(8):2486 [FREE Full text] [doi: 10.3390/s18082486] [Medline: 30071588]
32. Morgan R, Whaley P, Thayer KA, Schünemann HJ. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environ Int 2018;121(Pt 1):1027-1031 [FREE Full text] [doi: 10.1016/j.envint.2018.07.015] [Medline: 30166065]
33. Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. The revised cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. J Clin Epidemiol 2020;126:37-44. [doi: 10.1016/j.jclinepi.2020.06.015] [Medline: 32562833]
34. Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, Fuentes J, et al. Poor reliability between cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. PLoS One 2014;9(5):e96920 [FREE Full text] [doi: 10.1371/journal.pone.0096920] [Medline: 24824199]
35. Herbert R, Moseley A, Sherrington C. PEDro: a database of randomised controlled trials in physiotherapy. Health Inf Manag 1998;28(4):186-188. [doi: 10.1177/183335839902800410] [Medline: 10387366]
36. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83(8):713-721. [Medline: 12882612]
37. Sherrington C, Moseley AM, Herbert RD, Elkins MR, Maher CG. Ten years of evidence to guide physiotherapy interventions: Physiotherapy Evidence Database (PEDro). Br J Sports Med 2010;44(12):836-837. [doi: 10.1136/bjsm.2009.066357] [Medline: 19846426]

XSL•FO
RenderX

38. Armijo-Olivo S, da Costa BR, Cummings GG, Ha C, Fuentes J, Saltaji H, et al. PEDro or cochrane to assess the quality of clinical trials? A meta-epidemiological study. PLoS One 2015;10(7):e0132634 [FREE Full text] [doi: 10.1371/journal.pone.0132634] [Medline: 26161653]
39. Cheung M, Vijayakumar R. A guide to conducting a meta-analysis. Neuropsychol Rev 2016;26(2):121-128. [doi: 10.1007/s11065-016-9319-z] [Medline: 27209412]
40. Matsentidou S, Poullis C. Immersive visualizations in a VR Cave environment for the training and enhancement of social skills for children with autism. 2014 Presented at: International Conference on Computer Vision Theory and Applications (VISAPP); January 5-8, 2014; Lisbon, Portugal p. 230-236 URL: https://doi.org/10.5220/0004680002300236
41. Cheng Y, Huang CL, Yang CS. Using a 3D immersive virtual environment system to enhance social understanding and social skills for children with autism spectrum disorders. Focus Autism Other Dev Disabl 2015;30(4):222-236 [FREE Full text] [doi: 10.1177/1088357615583473]
42. Ip HHS, Wong SWL, Chan DFY, Byrne J, Li C, Yuan VSN, et al. Enhance emotional and social adaptation skills for children with autism spectrum disorder: a virtual reality enabled approach. Comput. Educ 2018;117:1-15 [FREE Full text] [doi: 10.1016/j.compedu.2017.09.010]
43. Rosenfield N, Lamkin K, Re J, Day K, Boyd L, Linstead E. A virtual reality system for practicing conversation skills for children with autism. Multimodal Technol. Interact 2019;3(2):28 [FREE Full text] [doi: 10.3390/mti3020028]
44. Junaidi AR, Alamsyah Y, Hidayah O, Mulyawati NW. Development of virtual reality content to improve social skills in children with low function autism. 2020 Presented at: 6th International Conference on Education and Technology (ICET); October 17, 2020; Malang, Indonesia p. 115-119 URL: https://doi.org/10.1109/ICET51153.2020.9276607
45. Yeh CC, Meng YR. Preliminary study on the application of virtual reality social skills course to improve the abilities of social skills for elementary and junior high school students with high functional autism. Commun. Comput. Inf. Sci 2020:183-193 [FREE Full text]
46. Tsai WT, Lee IJ, Chen C. Inclusion of third-person perspective in CAVE-like immersive 3D virtual reality role-playing games for social reciprocity training of children with an autism spectrum disorder. Univ Access Inf Soc 2020;20(2):375-389 [FREE Full text] [doi: 10.1007/s10209-020-00724-9]
47. Soltani Kouhbanani S, Khosrorad R, Zarenezhad S, Arabi SM. Comparing the effect of risperidone, virtual reality and risperidone on social skills, and behavioral problems in children with autism: a follow-up randomized clinical trial. Arch Iran Med 2021;24(7):534-541 [FREE Full text] [doi: 10.34172/aim.2021.76] [Medline: 34488318]
48. Moon J, Ke F. Exploring the treatment integrity of virtual reality-based social skills training for children with high-functioning autism. Interact. Learn. Environ 2019;29(6):939-953 [FREE Full text] [doi: 10.1080/10494820.2019.1613665]
49. Elkin TD, Zhang Y, Reneker JC. Gaze fixation and visual searching behaviors during an immersive virtual reality social skills training experience for children and youth with autism spectrum disorder: a pilot study. Brain Sci 2022;12(11):1568 [FREE Full text] [doi: 10.3390/brainsci12111568] [Medline: 36421892]
50. Frolli A, Savarese G, Di Carmine F, Bosco A, Saviano E, Rega A, et al. Children on the autism spectrum and the use of virtual reality for supporting social skills. Children (Basel) 2022;9(2):181 [FREE Full text] [doi: 10.3390/children9020181] [Medline: 35204903]
51. Ke F, Moon J, Sokolikj Z. Virtual reality–based social skills training for children with autism spectrum disorder. J Spec Educ Technol 2020;37(1):49-62 [FREE Full text] [doi: 10.1177/0162643420945603]
52. Zhao J, Zhang X, Lu Y, Wu X, Zhou F, Yang S, et al. Virtual reality technology enhances the cognitive and social communication of children with autism spectrum disorder. Front Public Health 2022;10:1029392 [FREE Full text] [doi: 10.3389/fpubh.2022.1029392] [Medline: 36276341]
53. Moon J, Ke F. Effects of adaptive prompts in virtual reality-based social skills training for children with autism. J Autism Dev Disord 2024;54(8):2826-2846. [doi: 10.1007/s10803-023-06021-7] [Medline: 37246166]
54. Levin L. Social and academic abilities in children with high-functioning autism spectrum disorders. Sch. Soc. Work J 2014;39(1):96 [FREE Full text]
55. Lai CLE, Lau Z, Lui SSY, Lok E, Tam V, Chan Q, et al. Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. Autism Res 2017;10(5):911-939. [doi: 10.1002/aur.1723] [Medline: 27874266]
56. Fernández MS. Mirror neurons and its impact on neurological diseases. Universidad de Cantabria. 2016. URL: https://repositorio.unican.es/xmlui/handle/10902/8743 [accessed 2025-08-01]
57. Reichow B, Servili C, Yasamy MT, Barbui C, Saxena S. Non-specialist psychosocial interventions for children and adolescents with intellectual disability or lower-functioning autism spectrum disorders: a systematic review. PLoS Med 2013;10(12):e1001572; discussion e1001572 [FREE Full text] [doi: 10.1371/journal.pmed.1001572] [Medline: 24358029]
58. Robertson CE, Baron-Cohen S. Sensory perception in autism. Nat Rev Neurosci 2017;18(11):671-684. [doi: 10.1038/nrn.2017.112] [Medline: 28951611]
59. Serret S, Hun S, Iakimova G, Lozada J, Anastassova M, Santos A, et al. Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new serious game: a pilot study. Mol Autism 2014;5:37 [FREE Full text] [doi: 10.1186/2040-2392-5-37] [Medline: 25018866]

XSL•FO
RenderX

60. Mazon C, Fage C, Sauzéon H. Effectiveness and usability of technology-based interventions for children and adolescents with ASD: a systematic review of reliability, consistency, generalization and durability related to the effects of intervention. Comput. Hum. Behav 2019;93:235-251 [FREE Full text] [doi: 10.1016/j.chb.2018.12.001]
61. McLoone JK, Rapee RM. Comparison of an anxiety management program for children implemented at home and school: lessons learned. Sch. Ment. Health. J. Multidiscip. Res 2012;4(4):231-242 [FREE Full text] [doi: 10.1007/s12310-012-9088-7]
62. Howard SJ, Vasseleu E, Batterham M, Neilsen-Hewett C. Everyday practices and activities to improve pre-school self-regulation: cluster RCT evaluation of the PRSIST program. Front Psychol 2020;11:137 [FREE Full text] [doi: 10.3389/fpsyg.2020.00137] [Medline: 32116939]
63. Wu B, Yu X, Gu X. Effectiveness of immersive virtual reality using head‐mounted displays on learning performance: a meta‐analysis. Brit J Educational Tech 2020;51(6):1991-2005 [FREE Full text] [doi: 10.1111/bjet.13023]
64. Panerai S, Gelardi D, Catania V, Rundo F, Tasca D, Musso S, et al. Functional living skills: a non-immersive virtual reality training for individuals with major neurocognitive disorders. Sensors (Basel) 2021;21(17):5751 [FREE Full text] [doi: 10.3390/s21175751] [Medline: 34502642]
65. Dailey-Hebert A, Estes JS, Choi DH. This history and evolution of virtual reality. In: Choi DH, Dailey-Hebert A, Estes JS, editors. Current and Prospective Applications of Virtual Reality in Higher Education. Beijing, China: Information Science Reference/IGI Global; 2021:1-20.
66. Tan B, Shi J, Yang S, Loh H, Ng D, Choo C, et al. The use of virtual reality and augmented reality in psychosocial rehabilitation for adults with neurodevelopmental disorders: a systematic review. Front Psychiatry 2022;13:1055204 [FREE Full text] [doi: 10.3389/fpsyt.2022.1055204] [Medline: 36590624]
67. Picard RW, Lane JK, El Kaliouby R, Goodwin M, Hoque ME. Exploring speech therapy games with children on the autism spectrum. Antwerp, Belgium: International Speech Communication Association; 2009 Presented at: 10th Annual Conference of the International Speech Communication Association; September 6-10, 2009; Brighton, United Kingdom URL: https://hdl.handle.net/1721.1/56580
68. Lindner P. Better, virtually: the past, present, and future of virtual reality cognitive behavior therapy. J Cogn Ther 2020;14(1):23-46 [FREE Full text] [doi: 10.1007/s41811-020-00090-7]
69. Leaf JB, Cihon JH, Leaf R, McEachin J, Liu N, Russell N, et al. Concerns about ABA-based intervention: an evaluation and recommendations. J Autism Dev Disord 2022;52(6):2838-2853 [FREE Full text] [doi: 10.1007/s10803-021-05137-y] [Medline: 34132968]
70. Zhao W, Xu S, Zhang Y, Li D, Zhu C, Wang K. The application of extended reality in treating children with autism spectrum disorder. Neurosci Bull 2024;40(8):1189-1204. [doi: 10.1007/s12264-024-01190-6] [Medline: 38498091]
71. Pirker J, Dengel A. The potential of 360° virtual reality videos and real VR for education-a literature review. IEEE Comput Graph Appl 2021;41(4):76-89. [doi: 10.1109/MCG.2021.3067999] [Medline: 33755559]
72. Wiebe A, Kannen K, Selaskowski B, Mehren A, Thöne AK, Pramme L, et al. Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. Clin Psychol Rev 2022;98:102213 [FREE Full text] [doi: 10.1016/j.cpr.2022.102213] [Medline: 36356351]
73. Didehbani N, Allen T, Kandalaft M, Krawczyk D, Chapman S. Virtual reality social cognition training for children with high functioning autism. Comput. Hum. Behav 2016;62:703-711 [FREE Full text] [doi: 10.1016/j.chb.2016.04.033]
74. Posar A, Visconti P. Sensory abnormalities in children with autism spectrum disorder. Jornal de Pediatria 2018;94(4):342-350. [doi: 10.1016/j.jped.2017.08.008]
75. Conner NO, Freeman HR, Jones JA, Luczak T, Carruth D, Knight AC, et al. Virtual reality induced symptoms and effects: concerns, causes, assessment and mitigation. Virtual Worlds 2022;1(2):130-146 [FREE Full text] [doi: 10.3390/virtualworlds1020008]
76. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
77. Indovina P, Barone D, Gallo L, Chirico A, de Pietro G, Giordano A. Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: a comprehensive literature review. Clin J Pain 2018;34(9):858-877. [doi: 10.1097/AJP.0000000000000599] [Medline: 29485536]
78. Leekam S, Prior MR, Uljarevic M. Restricted and repetitive behaviors in autism spectrum disorders: a review of research in the last decade. Psychol Bull 2011;137(4):562-593. [doi: 10.1037/a0023341] [Medline: 21574682]
79. Bernaerts S, Bonroy B, Daems J, Sels R, Struyf D, Gies I, et al. Virtual reality for distraction and relaxation in a pediatric hospital setting: an interventional study with a mixed-methods design. Front Digit Health 2022;4:866119 [FREE Full text] [doi: 10.3389/fdgth.2022.866119] [Medline: 35712230]
80. Zhang M, Ding H, Naumceska M, Zhang Y. Virtual reality technology as an educational and intervention tool for children with autism spectrum disorder: current perspectives and future directions. Behav Sci (Basel) 2022;12(5):138 [FREE Full text] [doi: 10.3390/bs12050138] [Medline: 35621435]

XSL•FO
RenderX

Abbreviations

ABA: applied behavior analysis
ASD: autism spectrum disorder
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
HFA: high-functioning autism
LFA: low-functioning autism
PEDro: Physiotherapy Evidence Database
PICOS: Population, Intervention, Comparison, Outcomes, and Study design
PRISMA: Preferred Reporting Items for Systematic Evaluation
SCIP: Social Communication Intervention Project
VR: virtual reality

Edited by F Xie; submitted 23.05.24; peer-reviewed by F Jorge, N Luke , X Ding, P Li; comments to author 25.08.24; revised version received 10.09.24; accepted 20.12.24; published 05.02.25.

Please cite as:
Yang X, Wu J, Ma Y, Yu J, Cao H, Zeng A, Fu R, Tang Y, Ren Z
Effectiveness of Virtual Reality Technology Interventions in Improving the Social Skills of Children and Adolescents With Autism: Systematic Review
J Med Internet Res 2025;27:e60845
URL: https://www.jmir.org/2025/1/e60845
doi:10.2196/60845
PMID:

©Xipeng Yang, Jinlong Wu, Yudan Ma, Jingxuan Yu, Hong Cao, Aihua Zeng, Rui Fu, Yucheng Tang, Zhanbing Ren. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis

Vivens Mubonanyikuzo[1*], MSC; Hongjie Yan[2*], PhD; Temitope Emmanuel Komolafe[3*], PhD; Liang Zhou[4], PhD; Tao Wu[1], PhD; Nizhuan Wang[5], PhD

[1]College of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China

[2]Department of Neurology, Affiliated Lianyungang Hospital of Xuzhou Medical University, Lianyungang, China

[3]Collaborative Research Center, Shanghai University of Medicine & Health Sciences, Shanghai, China

[4]Department of Radiology, Jiading District Central Hospital Affiliated to Shanghai University of Medicine & Health Sciences, Shanghai, China

[5]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, China (Hong Kong)

[*]these authors contributed equally

Corresponding Author:
Nizhuan Wang, PhD
Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University
11 Yuk Choi Rd, Hung Hom
Hong Kong
Kowloon, 999077
China (Hong Kong)
Phone: 852 27667681
Email: wangnizhuan1120@gmail.com

Abstract

Background: Alzheimer disease (AD) is a progressive condition characterized by cognitive decline and memory loss. Vision transformers (ViTs) are emerging as promising deep learning models in medical imaging, with potential applications in the detection and diagnosis of AD.

Objective: This review systematically examines recent studies on the application of ViTs in detecting AD, evaluating the diagnostic accuracy and impact of network architecture on model performance.

Methods: We conducted a systematic search across major medical databases, including China National Knowledge Infrastructure, CENTRAL (Cochrane Central Register of Controlled Trials), ScienceDirect, PubMed, Web of Science, and Scopus, covering publications from January 1, 2020, to March 1, 2024. A manual search was also performed to include relevant gray literature. The included papers used ViT models for AD detection versus healthy controls based on neuroimaging data, and the included studies used magnetic resonance imaging and positron emission tomography. Pooled diagnostic accuracy estimates, including sensitivity, specificity, likelihood ratios, and diagnostic odds ratios, were derived using random-effects models. Subgroup analyses comparing the diagnostic performance of different ViT network architectures were performed.

Results: The meta-analysis, encompassing 11 studies with 95% CIs and *P* values, demonstrated pooled diagnostic accuracy: sensitivity 0.925 (95% CI 0.892-0.959; *P*<.01), specificity 0.957 (95% CI 0.932-0.981; *P*<.01), positive likelihood ratio 21.84 (95% CI 12.26-38.91; *P*<.01), and negative likelihood ratio 0.08 (95% CI 0.05-0.14; *P*<.01). The area under the curve was notably high at 0.924. The findings highlight the potential of ViTs as effective tools for early and accurate AD diagnosis, offering insights for future neuroimaging-based diagnostic approaches.

Conclusions: This systematic review provides valuable evidence for the utility of ViT models in distinguishing patients with AD from healthy controls, thereby contributing to advancements in neuroimaging-based diagnostic methodologies.

Trial Registration: PROSPERO CRD42024584347; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=584347

(J Med Internet Res 2025;27:e62647) doi:10.2196/62647

XSL•FO
RenderX

KEYWORDS

diagnostic accuracy; vision transformer; Alzheimer disease; detection; neuroimaging, meta-analysis; neuroimaging; deep learning; medical database; diagnostic; clinical implementation; machine learning; magnetic resonance imaging; neural networks

Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disorder characterized by deterioration in cognitive function and memory impairment. As the global population ages, AD is emerging as a significant public health concern, affecting millions of individuals worldwide [1]. AD poses a mounting socioeconomic burden due to the escalating costs associated with medical care and long-term support for patients with these diseases and their caregivers. Consequently, there is an urgent need to develop accurate and early diagnostic methods to facilitate timely intervention and management of this debilitating condition [2].

The early and accurate detection of AD is imperative for initiating timely therapeutic interventions and proper care planning before irreversible neurological damage occurs. However, a definitive diagnosis of AD can only be established postmortem through the identification of amyloid plaques and neurofibrillary tangles in the brain tissue. Consequently, researchers have turned their attention to leveraging novel imaging techniques, such as magnetic resonance imaging (MRI) and 18-fluorodeoxyglucose and positron emission tomography, by the use of advanced computational techniques like machine learning (ML) and deep learning (DL) approaches to facilitate the earlier identification of AD. These advanced neuroimaging modalities and computational methods hold the potential to aid in the presymptomatic detection and monitoring of AD progression, which enables more effective disease management strategies [3]. DL and ML approaches have been widely applied to various diagnostic tasks, although the diagnostic accuracy of DL remains uncertain in some areas. To address this gap, meta-analysis methods have been proposed [4] to provide a more comprehensive understanding. For example, studies have estimated the diagnostic accuracy of DL models for COVID-19 detection [5] and evaluated ML models for osteoporosis diagnosis in the hip bone [6]. Similarly, both ML and DL techniques have been extensively used for the detection and diagnostic evaluation of AD. Odusami et al [7] undertook a systematic review and meta-analysis. They used ML models alongside multimodal neuroimaging data to classify various stages of AD progression. The findings from their investigation were highly encouraging, demonstrating pooled estimates for sensitivity of 94.6% and specificity of 93.5% in classifying patients with AD from healthy controls. This study demonstrates the considerable promise of ML algorithms when combined with multimodal neuroimaging biomarkers for differentiating patients with AD from cognitively normal (CN) individuals. This capability holds potential advantages for enabling early diagnosis and disease monitoring, essential for effective management and treatment. The application of ML in AD diagnosis opens a new chapter for research with the capacity to significantly impact patient care and management. The extensive meta-analysis conducted by Wang et al [8] explored the concept of deep neural networks and heat map visualizations on multimodal neuroimaging data, encompassing structural MRI and functional imaging, to unravel patterns linked with AD. Through the application of deep neural networks and heat map visualizations, the study successfully pinpointed and emphasized the particular brain regions mostly impacted by the disease. This outcome provides valuable insights into the fundamental mechanisms underlying AD. Qu et al [9] conducted a comprehensive review and meta-analysis comparing the diagnostic efficacy of generative adversarial network (GAN)–based and non-GAN DL approaches in AD diagnosis. Their findings revealed that GAN-based DL models are more superior and effective than the non-GAN models in distinguishing individuals with AD from those with just CN individuals. This superiority was demonstrated through significant enhancements in accuracy, sensitivity, specificity, and the area under the curve of the summary receiver operating characteristic. However, when differentiating between progressive mild cognitive impairment (MCI) and stable MCI, the GAN method did not demonstrate a notable improvement in either the accuracy or the sensitivity. Despite this, it displayed marginally better specificity and area under the curve of the summary receiver operating characteristic curve compared with the non-GAN method. These results highlight the potential of GAN-based DL techniques to enhance AD diagnosis. Nonetheless, their applicability may be constrained in certain diagnostic contexts. Vision transformers (ViTs) are a novel type of DL model that has been effectively used in various types of computer vision tasks due to their ability to learn long-range connections within the images [10]. The ViTs process image data using the self-attention mechanism, which allows the model to connect sensitive regions of the image, which ultimately improves computing efficiency and contextual comprehension. Dosovitskiy et al [11] proposed a novel approach in computer vision, demonstrating the efficacy of ViT for image recognition tasks. Deviating from conventional methods that rely on convolutional neural networks (CNN), ViTs directly apply transformer architecture to sequences of image patches. By leveraging pretraining on extensive data sets, ViTs achieve competitive performance on various image classification benchmarks with lesser computational complexity. This pioneering study emphasizes the potential of transformer-based models in revolutionizing image recognition tasks, with scalable and promising results. Thereafter, various novel ViT models were developed, with enhancements aimed at reducing their computational cost and extending them to additional vision tasks, such as object detection [12] and image segmentation [13]. ViTs are well-designed to serve as the fundamental models for computer vision tasks in the medical field. Therefore, many studies have been published recently focusing on this area of research [14-17].

CNN models process images as pixel arrays, whereas ViT splits the input images into visual tokens. The visual transformer breaks an image into fixed-size patches, embeds each one appropriately, and incorporates positional embeddings as input

XSL•FO
RenderX

for the transformer encoder. In addition, ViT models offer nearly four times the computational efficiency and accuracy compared with CNNs [18]. The self-attention layer in ViT enables the model to embed information globally across the entire image. It also learns from the training data to encode the relative positions of image patches, allowing it to reconstruct the image's structure effectively.

Despite the promising outcomes documented in the scientific literature regarding the utilization of ViTs for AD image classification and analysis, a comprehensive meta-analysis examining the advantages and performance of these ViTs is notably absent. This research gap underscores the necessity for a systematic review and meta-analysis to consolidate existing evidence systematically; this systematic review with meta-analysis of ViT network architecture aims to address this evidence gap. The goal of this review and meta-analysis is to evaluate and summarize all available evidence to quantitatively assess the diagnostic test accuracy (DTA) of ViTs for detecting AD using MRI and positron emission tomography (PET) brain images, also providing critical insights that can inform future research directions and clinical practice in the field of AD diagnosis and management.

Methods

Data Sources and Literature Query

In this systematic review and meta-analysis, relevant studies on AD and ViT-based models were systematically gathered from multiple databases known for extensive research on these topics. These databases included CENTRAL (Cochrane Central Register of Controlled Trials), China National Knowledge Infrastructure, PubMed, Scopus, ScienceDirect, and Web of Science. In addition, we searched other sources like Google Scholar for gray literature and conducted citation searches. In total, 24,693 records were identified. The selected literature encompassed publications from January 1, 2020, to March 1, 2024, specifically focusing on studies using ViTs for AD prediction with neuroimaging data. The selected studies used MRI and PET as imaging data. The search strategy used MeSH terms and other selected keywords across different databases. Search strings can be found in Multimedia Appendix 1. A comprehensive outline of the search methodology, adhering to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guideline of reporting systematic reviews [19], can be found in Multimedia Appendix 2.

Eligibility Criteria

In this systematic review and meta-analysis, we considered studies using ViTs for AD image classification, specifically those published since their inception in 2020 [11]. These studies were required to report detailed performance metrics, including sensitivity, specificity, and a 2×2 confusion matrix, for ViT models used either independently or in hybrid models with other DL architectures. The MRI and PET image data sets used in the studies could be either unimodal or multimodal. Studies were excluded based on the following criteria: (1) conference abstracts, (2) duplicate studies, (3) incomplete papers lacking full text, (4) studies lacking classification results for AD and normal control (NC), (5) studies not using ViT models for AD detection, (6) literature reviews, and (7) studies not adopting neuroimaging for AD and NC prediction. Studies included in this review showcased the application of ViT models for AD image classification and provided comprehensive performance metrics.

Selection of Involved Studies

Two teams of reviewers (VM and TEK and HY and NW) independently conducted the screening process. Using Rayyan (Rayyan Systems Inc) [20], a freely available software, duplicates were removed based on titles or abstracts, streamlining the screening process. The initial review involved titles and abstracts, followed by a thorough examination of full texts for further screening. Eligible studies were included in the final analysis. In instances where the 2 review teams were unable to reach an agreement, the disagreement was resolved by consulting a more experienced third reviewer (LZ), who facilitated consensus.

Data Collection Process

Two researchers (VM and TEK) independently adopted a self-developed data extraction form to meticulously extract specific parameters regarding network architecture. These include the country of study, network model type, data type, methods, number of training epochs, batch sizes, and other relevant information. In addition, we diligently extracted the (2×2) confusion matrix, consisting of true positive, false negative, true negative, and false positive. These data formed the foundation for computing various DTA metrics, including sensitivity, specificity, diagnostic odds ratios (DORs), recall, precision, F_1-score, and positive and negative likelihood ratios (LR+ and LR–). This rigorous approach ensures a comprehensive analysis of ViT performance in AD detection using neuroimages. These metrics were used to assess the performance of models used in studies. The formulas for estimating some of these parameters are presented in Equations 1-4.



Risk of Bias and Quality Assessment

The quality assessment of the included studies was conducted using the Quality Assessment of Diagnostic Accuracy Studies-2 tool [21], a validated instrument designed to rigorously evaluate both the methodological integrity and reporting quality of diagnostic accuracy studies. This tool systematically appraises key domains, including patient selection, index test, reference standard, and flow and timing, thereby offering a structured framework for a comprehensive assessment of study quality. About 31% of the outcomes in the included studies lacked details on patient selection, leading to a high risk of bias, whereas the rest were considered low risk. In addition, 37% of

XSL•FO
RenderX

the outcomes had unclear information on how the index test was performed, resulting in an unclear risk of bias for the test. Since the studies focused on AD, there was a mismatch between the review question and the index test. Furthermore, about 31% of the outcomes from all included studies did not clarify if the review question matched the targeted condition, raising concerns about the applicability of the index test. About 31% of the outcomes lacked clear information on the interval between the index test and reference standard, as well as on their execution, leading to unclear bias regarding flow and timing. The remaining studies were considered low risk for these factors. Figure 2 shows the overall study of the risk of bias evaluation.

Figure 2. Summary of QUADAS-2 plots across the 16 included studies in the domain. (A) Risk of bias plot. (B) Risk of bias for applicability concerns. QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies-2.

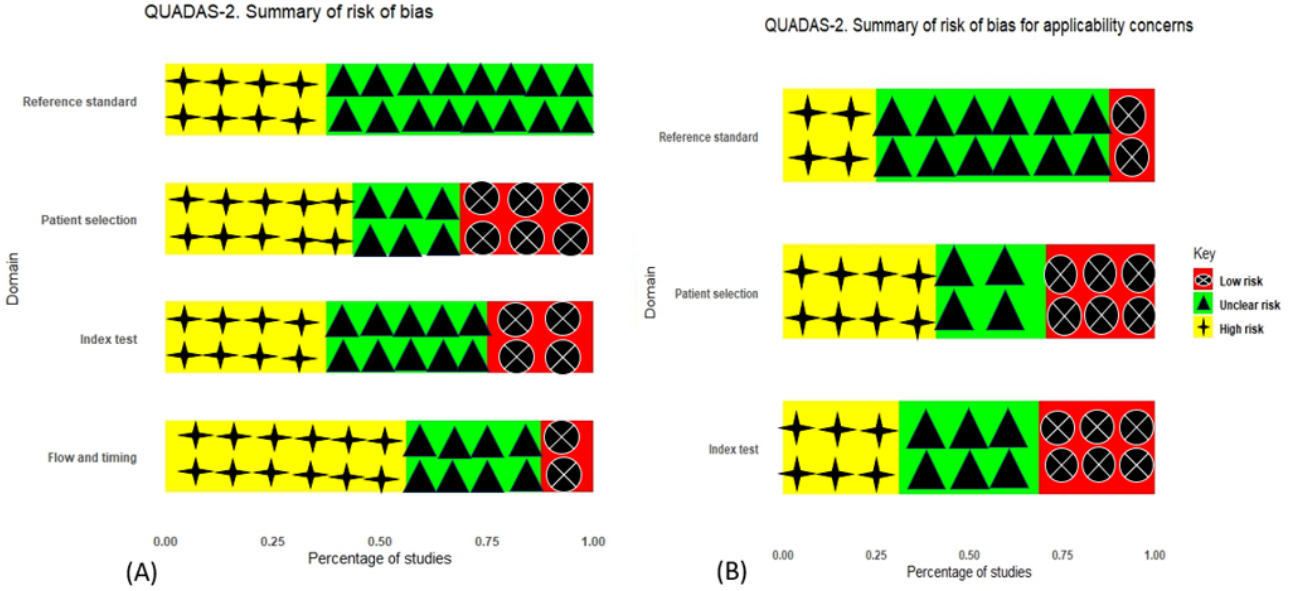

Statistical Analysis

The analysis combining data from multiple studies was conducted independently for sensitivity and specificity to consolidate the diagnostic accuracy findings of each network model. In this study, we used a random-effects model [22] to analyze the diagnostic performance of various ML models for medical diagnosis.

The primary outcomes are sensitivity, specificity, and the summary receiver operating characteristic curve, to provide a comprehensive assessment of diagnostic performance across studies. We computed point estimates and their corresponding 95% CIs for sensitivity and specificity to ensure consistency and comparability. Secondary outcomes, such as positive and negative likelihood ratios, were also explored. To assess statistical heterogeneity, the Cochran Q test and the I^2 statistic were used. For the Q statistic, values ranging from 0% to 40% imply insignificant heterogeneity, 30% to 60% connote moderate heterogeneity, and 75% to 100% mean considerable heterogeneity. Publication bias was assessed using funnel plots, with significance set at $P<.05$ for 2-sided tests. Furthermore, we conducted subgroup analyses by stratifying studies based on the training approach, distinguishing between those using the ViT model alone and hybrid models. This meticulous approach allowed for a robust evaluation of the effectiveness and potential biases within the literature.

Results

General Overview of All Included Studies

The initial search across databases yielded a total of 24,693 publications. After removing duplicates and screening the titles and abstracts, 11 papers [23-33] met the eligibility criteria, and 16 reports were included for qualitative synthesis and meta-analysis [23-33]. In addition, 4 of these papers [24,26,30,32] reported more than one study, resulting in a total of 16 outcomes that were used in the analysis as shown in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA flowchart for search strategy. Records were identified from CENTRAL, China National Knowledge Infrastructure, PubMed, Scopus, ScienceDirect, and Web of Science. AD: Alzheimer disease; CN: cognitively normal; CNN: convolutional neural network; HC: healthy controls; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

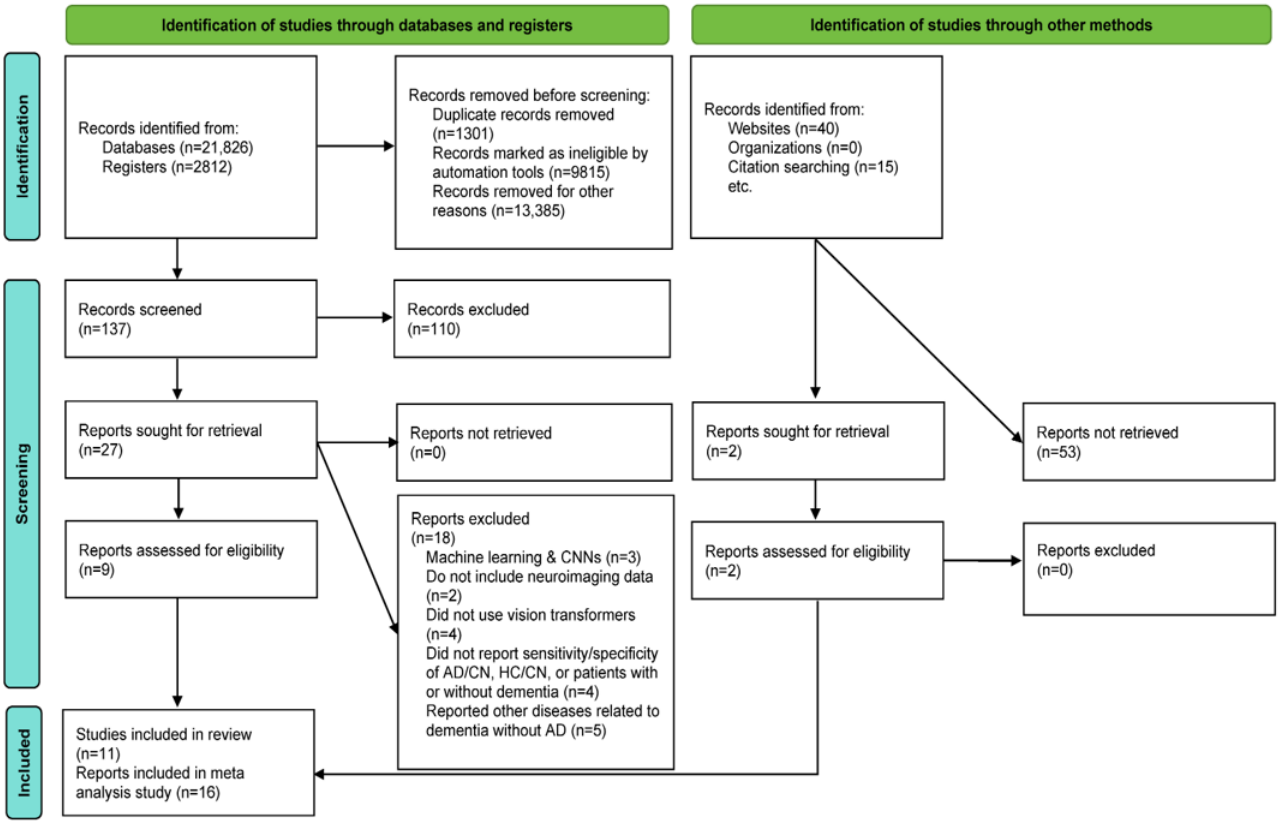

The ViT-based network architectures were classified into 2 categories. Among the reported studies, 7 of 16 outcomes (64%) used ViT models alone during the training and testing phases, whereas the remaining studies used hybrid models that integrated ViT with other DL architectures for training and testing. Various public data sets were used for training and evaluating the models, including the Alzheimer's Disease Neuroimaging Initiative; Alzheimer's Disease Neuroimaging Initiative Grand Opportunity Study; UK Biobank; Australian Imaging, Biomarkers, and Lifestyle Study; and others. The neuroimaging modalities used in these studies encompassed functional MRI, structural MRI, and PET, specifically fluorodeoxyglucose-PET.

The characteristics of the included studies are summarized in Table 1. The studies originated from China, representing 36% (4/11) of the total. The United States is the second most frequent at 27% (3/11). Lithuania and Korea are tied for third at 9% (1/11) each, whereas Colombia and Thailand each account for 9% (1/11). The most common objective across these studies was the diagnosis of AD versus NCs. Some studies used different datasets, and in such cases, we included each study separately. In addition, certain studies reported on multiple network architectures.

XSL•FO
RenderX

Table 1. Comprehensive summary of the characteristics of all the studies included in this systematic review.

Study	Country of study	Network architecture model (ViT[a]/hybrid)	Data (imaging method), MRI[b]/PET[c]	Training epochs/batch size	Type of task	Significance and the limitations of the study	Brief study summary
Alejandro et al 2023 [23]	Colombia	ViT	MRI	250/NA[d]	Classification	Significance: new image preprocessing techniques and spatial data augmentation. Limitations: small data set	This study uses deep learning to analyze MRI scans and detect AD[e] with up to 89% accuracy in early stages.
Odusam et al 2023_1 [24][f]; Odusam et al 2023_2 [24][f]	Lithuania	ViT	MRI+PET	NA/1024	Detection	Significance: using transfer learning, which improves the performance of the proposed model; an image fusion technique has been proposed to fuse multimodal images for AD diagnosis. Limitations: limited data set (ADNI[g] only used), the fusion parameters in the study were not optimized to their full potential.	Multimodal MRI+PET fusion for improved AD.
Zhu et al 2022 [25]	China	ViT	MRI	16/NA	Detection	Significance: integrates representation learning, feature distilling, and classifier modeling into a unified framework. Limitations: the fixed 5×3×5 size of the feature extraction patch is less suitable because brain disease-related structural changes vary in size, so a dynamic patch size would be more appropriate.	BraInf, a novel model, uses long-term connections in brain scans to detect AD with near-perfect accuracy.
Tang et al 2023_1[f] [26]; Tang et al 2023_2 [26][f]; Tang et al 2023_3[f] [26]	China	Hybrid	MRI+PET	128/300	Detection or classification	Significance: in the study CsAGP[h] a dual-branch vision transformer using cross-attention and graph pooling for multilevel feature interaction and representation was introduced to learn a shared feature representation by interacting with features at multiple levels. Limitations: the proposed CsAGP is slice-based and considers only axial view slices. Since 2D images cannot include all the information from a full brain scan. Also, this study has not conducted a time processing comparison.	A dual-branch vision transformer model using cross-attention and graph pooling for AD detection from multimodal images, achieving high classification accuracy rates through extensive experiments on the ADNI database.

XSL•FO
RenderX

Study	Country of study	Network architecture model (ViT[a]/hybrid)	Data (imaging method), MRI[b]/PET[c]	Training epochs/batch size	Type of task	Significance and the limitations of the study	Brief study summary
Liu et al 2023 [27]	China	Hybrid	MRI	50/NA	Classification	Significance: a multimodal method used to handle missing data in various clinical settings. Limitations: reliance on a GPU[i] may pose challenges in clinical environments where access to GPUs is limited.	This work introduces a cascaded multimodal mixing transformer (3MT) model for AD classification that effectively handles incomplete data, leveraging cross-attention and a novel modality dropout mechanism for robust and state-of-the-art performance on multimodal data sets.
Dhinagar et al 2023 [28]	United States	ViT	MRI	16/NA	Classification	Significance: variants of the ViT architecture used, 3D MRI images. Limitations: small data set used.	Application of ViTs on structural MRI scans for AD detection, achieving high classification accuracy.
Pan et al 2022 [29]	China	Hybrid	MRI	NA/NA	Classification	Significance: GAN[j] and transformer architecture used in AD detection. Limitations: one dataset used.	A Cross-Modal Transformer GAN (CT-GAN) framework that effectively fuses structural and functional brain imaging data to enhance AD prediction was introduced.
Khatri et al 2024_1 [30][f]; Khatri et al 2024_2 [30][f]	Korea	ViT; hybrid	PET	32/NA	Classification	Significance: ViT architecture by pretraining the feature extractor using the self-distillation with no labels (DINO) and extreme learning machine (ELM) as classifier models was introduced. Limitation: training process is complex.	RMTnet, a deep learning model combining recurrent neural networks (RNNs) and transformers for AD diagnosis using FDG-PET, was introduced.
Khan et al 2024 [31]	Thailand	Hybrid	MRI+PET	100/32	Detection	Significance: a novel model of mixed-transformer with furthered U-Net architecture. Limitations: lack of model interpretability.	The model combines MRI and PET scans with mixed-transformer and semantic segmentation techniques, achieving superior performance metrics compared with existing models.
Aghdam et al 2024_1 [32][f]; Aghdam et al 2024_2 [32][f]	United States	ViT; hybrid	MRI	20/NA	Detection	Significance: diagnosis of AD on white matter of T1-weighted structural data. Limitations: small data sets.	PVTAD[k], a new method for AD diagnosis that applies the Pyramid Vision Transformer to the white matter of T1-weighted structural MRI data, was introduced.

XSL•FO
RenderX

Study	Country of study	Network architecture model (ViT[a]/hybrid)	Data (imaging method), MRI[b]/PET[c]	Training epochs/batch size	Type of task	Significance and the limitations of the study	Brief study summary
Huang et al 2023 [33]	United States	Hybrid	MRI	50/16	Classification	Significance: combining CNN[l] and ViT. Limitations: low accuracy.	A self-supervised learning approach for dementia diagnosis from MRI scans was introduced, achieving state-of-the-art 92% accuracy.

[a]ViT: vision transformers.

[b]MRI: magnetic resonance imaging.

[c]PET: positron emission tomography.

[d]NA: not available.

[e]AD: Alzheimer disease.

[f]Studies that reported more than 1 outcome, for 3 outcomes (1,2,3) and 2 outcomes (1,2).

[g]ADNI: Alzheimer's Disease Neuroimaging Initiative.

[h]CsAGP: cross-attention and graph pooling algorithm.

[i]GPU: graphic processing unit.

[j]GAN: generative adversarial network.

[k]PVTAD: pyramid vision transformer for Alzheimer disease.

[l]CNN: convolutional neural networks.

The Overall Diagnostic Test Accuracy

The overall DTA for all included studies is presented in Figure 3; the pooled sensitivity of univariate analysis of 11 studies [23-33] with 16 outcomes was 0.925 (95% CI 0.892-0.959, I^2=84.2% for 16 outcomes). The pooled specificity of univariate analysis was 0.957 (95% CI 0.932-0.981, I^2=76% for 16 outcomes), as shown in Figure 4.

Figure 3. Forest plot of pooled sensitivity from 11 studies [23-33] with 16 outcomes using a meta-analysis of the proportion random-effect model.

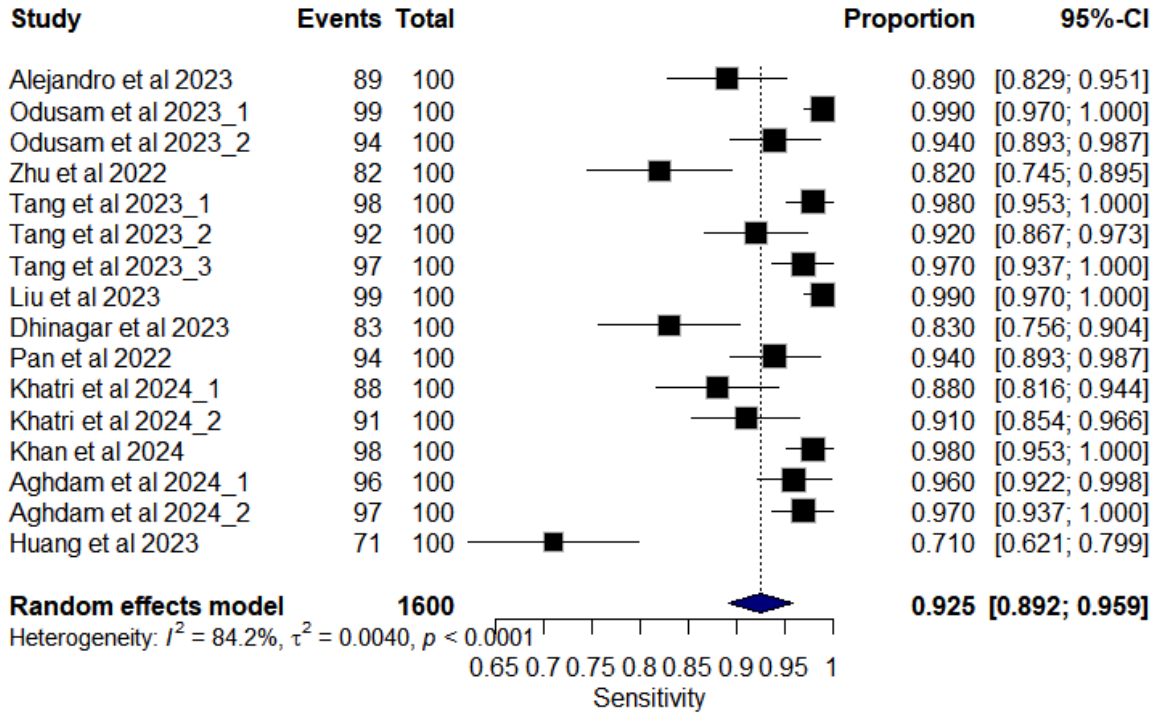

XSL•FO
RenderX

Figure 4. Forest plot of pooled specificity from 11 studies [23-33] with 16 outcomes using a meta-analysis of the proportion random-effect model.

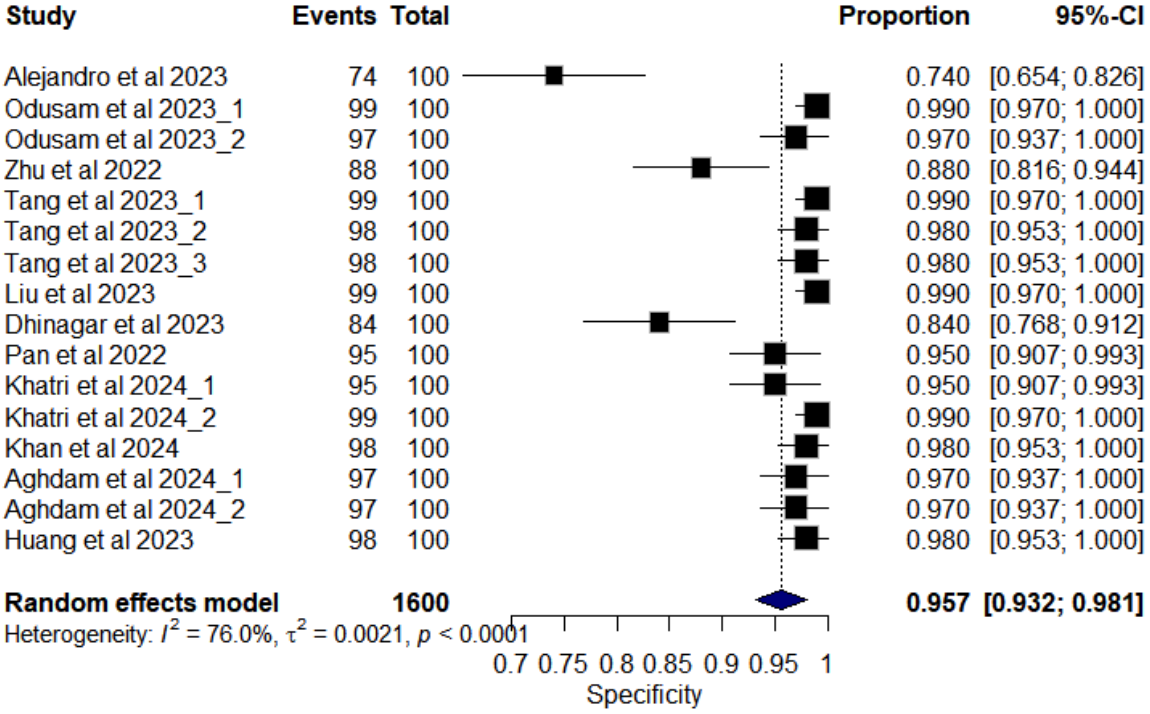

As depicted in Figure 5, the positive likelihood ratio (LR+) showed a pooled value of 21.84 (95% CI 12.26-38.91, I^2=64.9% for 16 outcomes). Similarly, the negative likelihood ratio (LR–) showed a pooled value of 0.084 (95% CI 0.054-0.133, I^2=40.2% for 16 outcomes), as shown in Figure 6. The summary receiver operating characteristic curve of the bivariate model has an AUC of 0.924 (Figure 7), indicating excellent overall diagnostic performance.

Figure 5. Forest plot of pooled positive likelihood ratio (LR) from 11 studies [23-33] with 16 outcomes. The risk ratio (RR) with 95% CI for each outcome is presented alongside the corresponding study weight.

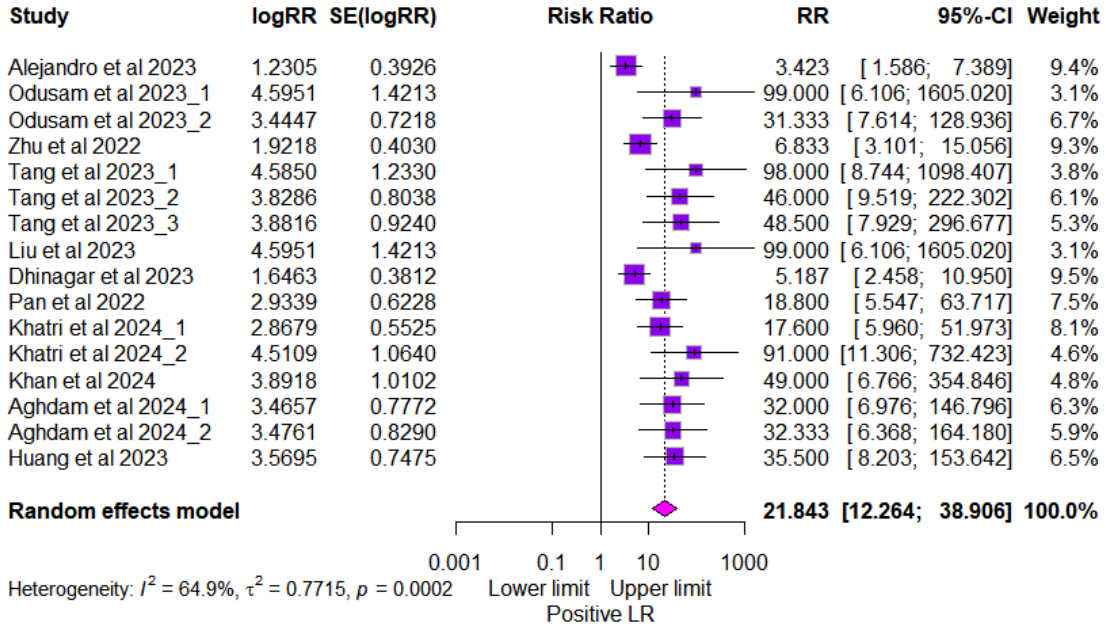

XSL•FO
RenderX

Figure 6. Forest plot of pooled negative likelihood ratio (LR) from 11 studies [23-33] with 16 outcomes. The risk ratio (RR) with 95% CI for each outcome is presented alongside the corresponding study weight.

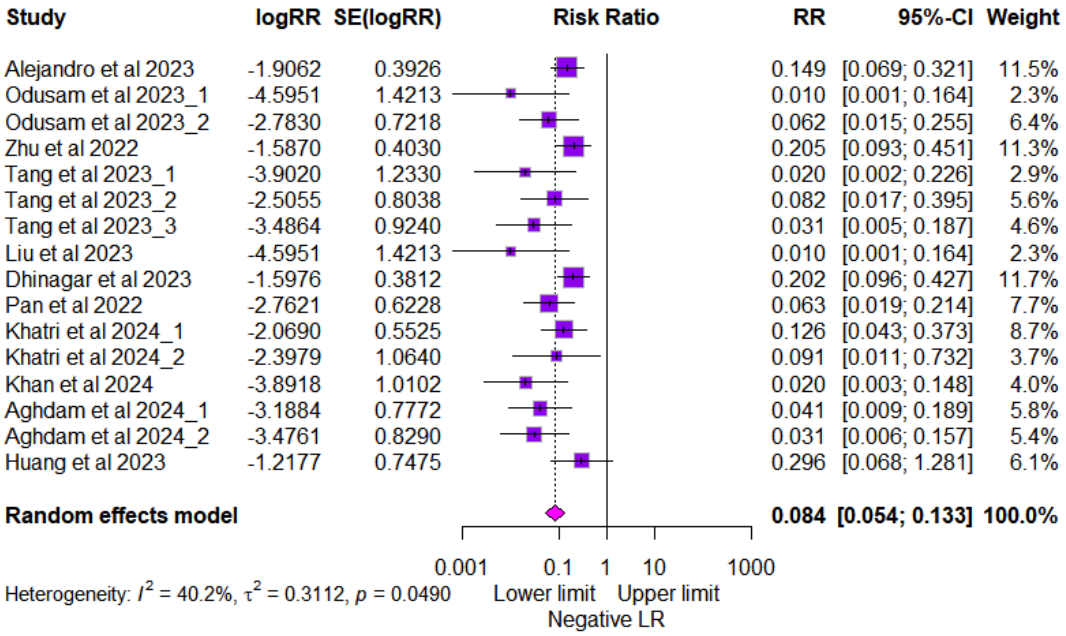

Figure 7. A plot of bivariate SROC for the diagnosis of AD using the ViT model. Green points represent individual studies included in the analysis, while the blue curve represents the fitted SROC curve based on the bivariate model. AD: Alzheimer disease; AUC: area under the curve; ViT: vision transformer; SROC: summary receiver operating characteristic curve; SE(AUC): standard error of the area under the curve.

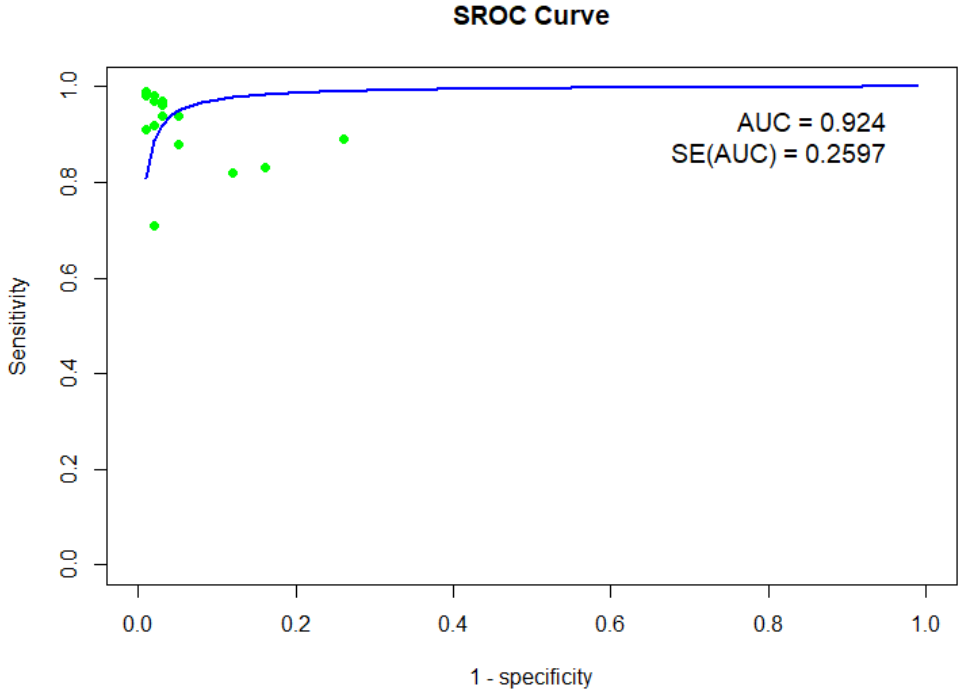

DTA for the Subgroup Analysis Based on Network Architecture

Based on the network architecture, the model was divided into 2 models: ViT models and hybrid networks that incorporate the ViT model with other DL models. The ViT model alone has a sensitivity of 0.908 (95% CI 0.860-0.957, I^2=86.9% for 7 of 16 outcomes), whereas the hybrid models have a sensitivity of 0.937 (95% CI 0.890-0.985, I^2=83.1% for 9 of 16 outcomes). The pooled sensitivity values indicate that there is no significant difference between the 2 subgroups of the network model, with a P value of .40, as shown in Figure 8. The pooled specificity of the ViT model alone exhibits a pooled specificity of 0.912 (95% CI 0.849-0.976, I^2=88.4% across 7 of 16 outcomes). In contrast, the hybrid model demonstrates superior pooled specificity at 0.984 (95% CI 0.975-0.992, I^2=0% across 9 of 16 outcomes). These results reveal that there is a statistically

XSL•FO
RenderX

significant difference in the specificity of the 2 categories of network models with a *P* value of .03 as shown in Figure S1 in Multimedia Appendix 3.

In the subgroup analysis comparing the ViT model and the hybrid model in our meta-analysis, the pooled DOR for the ViT model was 152.681 (95% CI 34.969-666.632, I^2=87.4% for 7 of 16 outcomes), whereas for the hybrid model, it was notably higher at 880.203 (95% CI 337.53-2052.12, I^2=50% for 9 of 16 outcomes). In addition, the *P* value for the subgroup difference was found to be .04, indicating a statistically significant difference between the 2 models. Given the substantially higher pooled DOR for the hybrid model and the significant subgroup difference, it shows that the hybrid model appears to be more effective in diagnosing AD based on our meta-analysis findings (Figure S2 in Multimedia Appendix 3).

Figure 8. Subgroup analysis plot of pooled sensitivity based on the network architecture category. NB: Model 0 represents the ViT model alone, whereas Model 1 adapts the ViT model with other deep learning (DL) models.

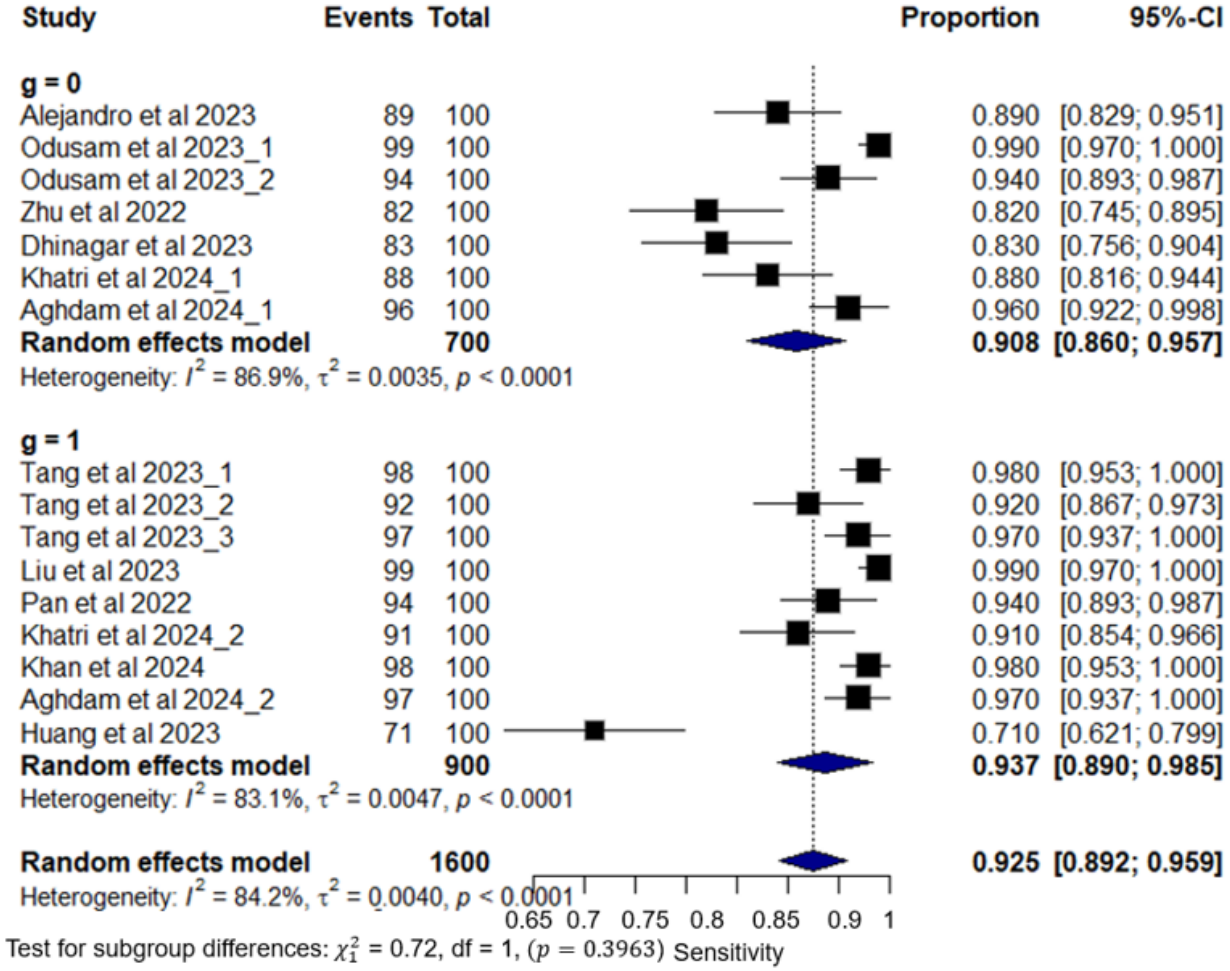

Assessment of Study Quality

Assessment of the review domains on risk of bias and applicability concerns was performed after reviewing studies. The review domains included patient selection, index test, reference standard, and flow and timing.

Patient Selection

There were 4 of 16 outcomes from 3 studies [28,30,33] (25%) that raised low concerns for patient selection bias and 5 of 16 outcomes from 2 studies [26,32] (31.25%) with high concerns, whereas 7 of 16 outcomes from 6 studies [23-25,27,29,31] (43.75%) were unclear, due to the studies not providing the detailed information about dataset used during training, testing, and validation.

Index Test

There were 6 of 16 outcomes from 4 studies [26,28,29,33] (38%) that raised low concerns, 4 of 16 outcomes from 3 studies [25,27,32] (25%) with high concerns, and 6 of 16 outcomes from 4 studies [23,24,30,31] (38%) that were unclear, primarily due to inadequate reporting of methodological details.

Reference Standard

There were 10 of 16 outcomes (62%) from 7 studies [25,26,28,29,31-33] with low concerns and 6 of 16 outcomes (38%) from 4 studies [23,24,27,30] that were unclear.

Flow and Timing

We conducted applicability concerns based on the interval between the index test and reference standard, whether all patients were included in the analysis, and whether the same data set used in studies received the same reference standard.

XSL•FO
RenderX

The results showed 7 of 16 outcomes from 4 studies [28,30,31,33] (43.75%) raise low concerns for patient selection bias, 3 of 16 outcomes from 2 studies [26,32] (18.75%) raise high concerns, and 6 of 16 outcomes from 5 studies [23-25,27,29] (38%) were unclear; for the index test, 4 of 16 outcomes from 4 studies [26,28,29,33] (25%) raise low concerns, 5 of 16 outcomes from 4 studies [25,27,31,32] (31.25%) raise high concerns, and 5 of 16 outcomes from 3 studies [23,24,30] (31.25%) were unclear; for reference standard, 10 of 16 outcomes from 7 studies [25,26,28,29,31-33] (62%) raise low concerns, 1 of 16 outcomes from 1 study [30] (6%) raises high concerns, and 4 of 16 outcomes from 3 studies [23,24,27] (25%) were unclear, which can be found in Multimedia Appendix 4.

Discussion

Principal Findings

In the current systematic review and meta-analysis, the DTA of the vision transformer (ViT) in detecting AD was evaluated. The pooled DTA metrics across 11 studies revealed robust performance in detecting AD. The pooled sensitivity was 0.925 (95% CI 0.892-0.959), indicating the model's ability to correctly identify individuals with AD, whereas the specificity was 0.957 (95% CI 0.932-0.981), demonstrating its capacity to accurately classify individuals without AD. In addition, the positive likelihood ratio was 21.84 (95% CI 12.26-38.91), showing the strength of a positive test result, whereas the negative likelihood ratio was 0.08 (95% CI 0.05-0.13), indicating the robustness of a negative test result. The area under the curve was 0.924, further confirming the model's overall diagnostic accuracy. Comparatively, the study conducted by Odusami et al [7] also assessed diagnostic performance in AD detection using ML and DL models. Their meta-analysis reported pooled sensitivity and specificity estimates of 94.60% (95% CI 90.76-96.89) and 93.49% (95% CI 91.60-94.90), respectively, for differentiating AD from CN individuals. Notably, while our specificity estimate was slightly higher than that reported by Odusami et al [7], both studies demonstrate high accuracy in AD detection.

The comparison suggests that both approaches achieve strong diagnostic performance, with minimal variations in specificity estimates. These differences may originate from variations in sample characteristics, imaging protocols, or model architectures used across studies. Overall, the findings corroborate the effectiveness of ML and DL approaches in AD detection, providing valuable insights for clinical practice and future research directions.

Qu et al (2022) [9] conducted a meta-analysis on AD versus CN classification across 11 studies, revealing GAN-based DL with a pooled sensitivity of 0.88, specificity of 0.93, and an AUC of 0.96. In our study, using the ViT-based models for AD detection, we observed promising results with a pooled sensitivity of 0.940, specificity of 0.962, and an impressive AUC of 0.9874. Our findings suggest that ViT models offer high diagnostic accuracy in distinguishing patients with AD from healthy controls, potentially enabling early and accurate diagnosis. Notably, our results further indicate slightly superior sensitivity and specificity compared with GAN-based approaches, indicating the potential superiority of ViT models in AD detection.

Hu et al [34] conducted a meta-analysis assessing the diagnostic performance of MRI-based ML in AD detection. Their findings revealed a DOR of 43.34 (95% CI 26.89-69.84), suggesting strong discriminatory power, as the odds of a positive result were 43.34 times higher in patients with AD compared with those without AD. Comparatively, our analysis indicates that ViT-based models outperformed other types of DL models. Hu et al [30] reported an LR+ of 7.15 (95% CI 5.40-9.47) and an LR– of 0.17 (95% CI 0.12-0.22). These likelihood ratios imply that a positive result is associated with a higher likelihood of AD, whereas a negative result suggests a lower likelihood. In contrast, our estimated results for ViT models yielded a pooled LR+ of 26.02 and LR– of 0.07. This indicates that a positive ViT model result is 26.02 times more likely in individuals with AD compared with those without, whereas a negative result is more likely in patients without AD. Overall, the findings highlight the superior diagnostic accuracy of ViT models in AD detection compared with other types of DL models.

The meta-analysis revealed substantial heterogeneity across the studies due to differences in countries, ViT-based architectures, and data types. This heterogeneity could be a potential source of bias and highlights the need for more studies to obtain more consistent results in AD detection. The high sensitivity and specificity of the ViT-based models imply that ViT models can be a valuable tool in clinical scenarios to aid radiologists and neurologists in interpreting neuroimages for AD, potentially reducing the number of false negatives and false positives. The use of ViT, which captures direct relationships between image areas, may offer advantages in analyzing the complex network of the brain. Based on a search of academic literature databases, this is the first systematic review and meta-analysis that focuses on the diagnosis of AD using ViT-based models. Therefore, this review serves to highlight the potential of ViT and neuroimaging in enhancing the precision and productivity of AD diagnosis, providing a plethora of innovative solutions to address the challenges inherent in early detection and classification. However, this study faced some limitations. First, the pioneering nature of this systematic review presents a challenge for comparative analysis with other DL models. This is because, to the best of our knowledge, there have been no previous systematic and meta-analysis studies that focus specifically on ViT in the context of AD. The pioneering nature of this study means that there is a lack of existing literature to draw comparisons and establish a benchmark for the diagnostic accuracy of ViT-based models. The study acknowledges this issue and recognizes the need for further research to fill this gap. In the future, more studies are needed; specifically, there is a need to establish a more comprehensive understanding of ViTs diagnostic accuracy and their performance relative to other DL models in diagnosing AD.

Second, the reliance on public datasets for the included studies means that the diagnostic performance of ViT methods is based on data that may not fully represent the complexity and variability of real-world AD cases. The use of such datasets, although useful for initial research, does not account for the potential subtleties and specificities of patient data obtained

XSL•FO
RenderX

from hospitals. This limitation is further compounded by the challenges associated with patient privacy and the complexities of preprocessing hospital data, which can hinder the validation of ViT methods in a real-world context. The study emphasizes the need for future research to address these challenges and validate the findings on more representative and diverse datasets. Researchers should also conduct further evaluations of ViTs using hospital data to better align with real-world applications [35].

Third, the limited number of studies available has restricted the research to binary classification tasks, which differentiate patients with AD from CN individuals. This focus does not capture the full spectrum of AD's progression, including the various stages and the transition from MCI to late MCI and stable MCI [15]. The study acknowledges the need for a broader application of ViT-based DL methods to detect and track the various stages of AD, which is crucial for a more comprehensive understanding of the disease and its progression. In addition, there is a notable gap in studies that use multimodal data to predict the progression of AD.

Furthermore, the study highlights a significant gap in the current models' ability to provide expert-level interpretations [36]. The absence of a reference standard in the form of expert diagnosis makes it difficult to conclusively determine the superiority of ViT-based models in real-world health care settings. This limitation underscores the importance of future research efforts that should focus on enhancing the transparency and explainability of ViT-based models. In addition, there is a need to explore the integration of these models with expert opinions to improve the accuracy of AD detection and diagnosis. By doing so, the research can move toward a more collaborative approach that combines the strengths of both artificial intelligence and human expertise.

Third, most ViTs require the use of GPUs for training, which may not be readily available in many hospitals. Therefore, there is a need to develop smaller models that can be trained in more affordable and accessible environments.

Finally, the findings of this study have important implications for clinical practice. ViT models show high diagnostic accuracy in AD detection, offering the potential for early diagnosis and personalized treatment. However, several barriers may hinder their implementation in routine clinical settings. These include the need for large, well-labeled datasets, the risk of overfitting, and the challenges of integrating ViT models into existing clinical workflows. In addition, the computational resources required for training, such as GPUs, may not be readily available in many hospitals. Addressing these challenges is essential to realize the full potential of ViTs in clinical applications.

Conclusion

The systematic review and meta-analysis highlight the strong diagnostic capability of ViT-based models in detecting AD from MRI and PET brain images. Although there are limitations, such as the lack of comparative analysis with other DL models, ViT models demonstrate high sensitivity and specificity. ViTs have the potential to significantly enhance image analysis in AD diagnosis by improving feature extraction, enabling early detection of subtle brain changes, supporting personalized diagnoses, and facilitating the integration of additional data for more accurate diagnoses. However, the effective application of ViT models comes with challenges, including the need for large datasets, the risk of overfitting, and the integration of these models into clinical workflows. To fully harness their potential in clinical settings, continued research is essential, particularly to improve model transparency and incorporate expert opinions. This study highlights the promising role of ViTs in advancing early AD detection and classification, paving the way for innovative solutions in neuroimaging analysis.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82001160), the “HaiYan Plan” Scientific Research Funding Project of Lianyungang City (2017-QD-009), the First People's Hospital of Lianyungang–Advanced Technology Support Project (XJ1811), the Project of Huaguoshan Mountain Talent Plan-Doctors for Innovation and Entrepreneurship, and the Hong Kong Polytechnic University Start-up Fund (P0053210). The code is available on GitHub.

Authors' Contributions

VM, HY, NW, and TEK contributed to data acquisition, analysis, and the drafting and review of the manuscript. HY and NW secured funding for the study. TEK and NW coordinated project administration and supervision and served as co-corresponding authors. LZ and TW provided expertise in data interpretation as well as manuscript drafting and reviewing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms used in each database.
[DOCX File , 19 KB - jmir_v27i1e62647_app1.docx]

Multimedia Appendix 2
PRISMA checklist.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 290 KB - jmir_v27i1e62647_app2.pdf]

Multimedia Appendix 3
Supplementary figures for subgroup analysis of diagnostic test accuracy (DTA) vision transformers based on network architecture.
[DOCX File , 337 KB - jmir_v27i1e62647_app3.docx]

Multimedia Appendix 4
Supplementary table for Risk of Bias assessment.
[DOCX File , 20 KB - jmir_v27i1e62647_app4.docx]

References

1. Chavan AB, Patil SR, Patel AM, Chaugule SV, Gharal SK. A comprehensive review on Alzheimer's disease its pathogenesis, epidermiology, diagnostics and treatment. Journal for Research in Applied Sciences and Biotechnology 2023;2(4):66-72 [FREE Full text] [doi: 10.55544/jrasb.2.4.8]
2. Bateman DR, Brady E, Wilkerson D, Yi EH, Karanam Y, Callahan CM. Comparing crowdsourcing and friendsourcing: a social media-based feasibility study to support Alzheimer disease caregivers. JMIR Res Protoc 2017;6(4):e56 [FREE Full text] [doi: 10.2196/resprot.6904] [Medline: 28396304]
3. Qi W, Zhu X, He D, Wang B, Cao S, Dong C, et al. Mapping knowledge landscapes and emerging trends in AI for dementia biomarkers: bibliometric and visualization analysis. J Med Internet Res 2024;26:e57830 [FREE Full text] [doi: 10.2196/57830] [Medline: 39116438]
4. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021;4(1):65 [FREE Full text] [doi: 10.1038/s41746-021-00438-z] [Medline: 33828217]
5. Komolafe TE, Cao Y, Nguchu BA, Monkam P, Olaniyi EO, Sun H, et al. Diagnostic test accuracy of deep learning detection of COVID-19: a systematic review and meta-analysis. Acad Radiol 2021;28(11):1507-1523 [FREE Full text] [doi: 10.1016/j.acra.2021.08.008] [Medline: 34649779]
6. Rahim F, Zaki Zadeh A, Javanmardi P, Emmanuel Komolafe T, Khalafi M, Arjomandi A, et al. Machine learning algorithms for diagnosis of hip bone osteoporosis: a systematic review and meta-analysis study. Biomed Eng Online 2023;22(1):68 [FREE Full text] [doi: 10.1186/s12938-023-01132-9] [Medline: 37430259]
7. Odusami M, Maskeliūnas R, Damaševičius R, Misra S. Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: a systematic review and meta-analysis. Cogn Neurodyn 2024;18(3):775-794 [FREE Full text] [doi: 10.1007/s11571-023-09993-5] [Medline: 38826669]
8. Wang D, Honnorat N, Fox PT, Ritter K, Eickhoff SB, Seshadri S, Alzheimer's Disease Neuroimaging Initiative, et al. Deep neural network heatmaps capture Alzheimer's disease patterns reported in a large meta-analysis of neuroimaging studies. Neuroimage 2023;269:119929 [FREE Full text] [doi: 10.1016/j.neuroimage.2023.119929] [Medline: 36740029]
9. Qu C, Zou Y, Ma Y, Chen Q, Luo J, Fan H, et al. Diagnostic performance of generative adversarial network-based deep learning methods for Alzheimer's disease: a systematic review and meta-analysis. Front Aging Neurosci 2022;14:841696 [FREE Full text] [doi: 10.3389/fnagi.2022.841696] [Medline: 35527734]
10. Duong LT, Le NH, Tran TB, Ngo VM, Nguyen PT. Detection of tuberculosis from chest x-ray images: boosting the performance with vision transformer and transfer learning. Expert Syst Appl 2021;184:115519 [FREE Full text] [doi: 10.1016/j.eswa.2021.115519]
11. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929 2020 [FREE Full text]
12. Beal J, Kim E, Tzeng E, Park DH, Zhai A, Kislyuk D. Toward transformer-based object detection. arXiv Preprint posted online on December 17, 2020 [FREE Full text]
13. Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, et al. Alzheimer disease. Nat Rev Dis Primers 2021;7(1):33 [FREE Full text] [doi: 10.1038/s41572-021-00269-y] [Medline: 33986301]
14. Shin H, Jeon S, Seol Y, Kim S, Kang D. Vision transformer approach for classification of Alzheimer's disease using 18F-Florbetaben brain images. Appl Sci 2023;13(6):3453 [FREE Full text] [doi: 10.3390/app13063453]
15. Hoang GM, Kim UH, Kim JG. Vision transformers for the prediction of mild cognitive impairment to Alzheimer's disease progression using mid-sagittal sMRI. Front Aging Neurosci 2023;15:1102869 [FREE Full text] [doi: 10.3389/fnagi.2023.1102869] [Medline: 37122374]
16. Dhinagar NJ, Thomopoulos SI, Laltoo E, Thompson PM. Efficiently training vision transformers on structural MRI scans for Alzheimer's disease detection. Annu Int Conf IEEE Eng Med Biol Soc 2023;2023:1-6. [doi: 10.1109/EMBC40787.2023.10341190] [Medline: 38083552]
17. Hu Z, Wang Z, Jin Y, Hou W. VGG-TSwinformer: transformer-based deep learning model for early Alzheimer's disease prediction. Comput Methods Programs Biomed 2023;229:107291. [doi: 10.1016/j.cmpb.2022.107291] [Medline: 36516516]

XSL•FO
RenderX

18. Krishnan PT, Krishnadoss P, Khandelwal M, Gupta D, Nihaal A, Kumar TS. Enhancing brain tumor detection in MRI with a rotation invariant vision transformer. Front Neuroinform 2024;18:1414925 [FREE Full text] [doi: 10.3389/fninf.2024.1414925] [Medline: 38957549]
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
22. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-188 [FREE Full text] [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
23. Mora-Rubio A, Bravo-Ortíz MA, Quiñones Arredondo S, Saborit Torres JM, Ruz GA, Tabares-Soto R. Classification of Alzheimer's disease stages from magnetic resonance images using deep learning. PeerJ Comput Sci 2023;9:e1490 [FREE Full text] [doi: 10.7717/peerj-cs.1490] [Medline: 37705614]
24. Odusami M, Maskeliūnas R, Damaševičius R. Optimized convolutional fusion for multimodal neuroimaging in Alzheimer's disease diagnosis: enhancing data integration and feature extraction. J Pers Med 2023;13(10):1496 [FREE Full text] [doi: 10.3390/jpm13101496] [Medline: 37888107]
25. Zhu J, Tan Y, Lin R, Miao J, Fan X, Zhu Y, et al. Efficient self-attention mechanism and structural distilling model for Alzheimer's disease diagnosis. Comput Biol Med 2022;147:105737 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.105737] [Medline: 35785662]
26. Tang C, Wei M, Sun J, Wang S, Zhang Y, Alzheimer's Disease Neuroimaging Initiative. CsAGP: detecting Alzheimer's disease from multimodal images via dual-transformer with cross-attention and graph pooling. J King Saud Univ Comput Inf Sci 2023;35(7):101618 [FREE Full text] [doi: 10.1016/j.jksuci.2023.101618] [Medline: 38559705]
27. Liu L, Liu S, Zhang L, To XV, Nasrallah F, Chandra SS. Cascaded multi-modal mixing transformers for Alzheimer's disease classification with incomplete data. Neuroimage 2023;277:120267 [FREE Full text] [doi: 10.1016/j.neuroimage.2023.120267] [Medline: 37422279]
28. Dhinagar NJ, Thomopoulos SI, Laltoo E, Thompson PM. Efficiently training vision transformers on structural MRI scans for Alzheimer's disease detection. Annu Int Conf IEEE Eng Med Biol Soc 2023;2023:1-6. [doi: 10.1109/EMBC40787.2023.10341190] [Medline: 38083552]
29. Pan J, Jing C, Zuo Q, Nieuwoudt M, Wang S. Cross-modal transformer GAN: a brain structure-function deep fusing framework for Alzheimer's disease. In: Advances in Brain Inspired Cognitive Systems. Singapore: Springer; 2022:82-91.
30. Khatri U, Kwon GR. RMTnet: recurrence meet transformer for Alzheimer's disease diagnosis using FDG-PET. TechRxiv. 2024. URL: https://www.techrxiv.org/users/714999/articles/699440-rmtnet-recurrence-meet-transformer-for-alzheimer-s-disease-diagnosis-using-fdg-pet?commit=a15de5eee3b5b66dcbd6cf0b02e25a60501e351c [accessed 2024-01-22]
31. Khan AA, Mahendran RK, Perumal K, Faheem M. Dual-3DMAD: mixed transformer based semantic segmentation and triplet pre-processing for early multi-class Alzheimer's diagnosis. IEEE Trans Neural Syst Rehabil Eng 2024;32:696-707. [doi: 10.1109/TNSRE.2024.3357723] [Medline: 38261494]
32. Akhavan Aghdam M, Bozdag S, Saeed F, Alzheimer's Disease Neuroimaging Initiative. bioRxiv. Pvtad: Alzheimer's disease diagnosis using pyramid vision transformer applied to white matter of t1-weighted structural MRI data Preprint posted online on December 4, 2023 [FREE Full text] [doi: 10.1101/2023.11.17.567617]
33. Huang Z. Self-supervised dementia prediction from MRI scans with metadata integration. 2023. URL: https://projects.horizoninspires.com/index.php/harp/preprint/view/120 [accessed 2023-10-25]
34. Hu J, Wang Y, Guo D, Qu Z, Sui C, He G, et al. Diagnostic performance of magnetic resonance imaging-based machine learning in Alzheimer's disease detection: a meta-analysis. Neuroradiology 2023;65(3):513-527. [doi: 10.1007/s00234-022-03098-2] [Medline: 36477499]
35. Surianarayanan C, Lawrence JJ, Chelliah PR, Prakash E, Hewage C. Convergence of artificial intelligence and neuroscience towards the diagnosis of neurological disorders-a scoping review. Sensors (Basel) 2023;23(6):3062 [FREE Full text] [doi: 10.3390/s23063062] [Medline: 36991773]
36. Arras L, Osman A, Samek W. CLEVR-XAI: a benchmark dataset for the ground truth evaluation of neural network explanations. Inf Fusion 2022;81:14-40. [doi: 10.1016/j.inffus.2021.11.008]

Abbreviations

AD: Alzheimer disease
CENTRAL: Cochrane Central Register of Controlled Trials
CNN: convolutional neural networks
DL: deep learning

XSL•FO
RenderX

DOR: diagnostic odds ratio
DTA: diagnostic test accuracy
LR+: positive likelihood ratio
LR–: negative likelihood ratio
ML: machine learning
MRI: magnetic resonance imaging
NC: normal control
PET: positron emission tomography
ViTs: vision transformers

Edited by B Puladi; submitted 28.05.24; peer-reviewed by T Chen, MH Alshayeji; comments to author 25.08.24; revised version received 02.10.24; accepted 29.11.24; published 05.02.25.

Please cite as:
Mubonanyikuzo V, Yan H, Komolafe TE, Zhou L, Wu T, Wang N
Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e62647
URL: https://www.jmir.org/2025/1/e62647
doi:10.2196/62647
PMID:

©Vivens Mubonanyikuzo, Hongjie Yan, Temitope Emmanuel Komolafe, Liang Zhou, Tao Wu, Nizhuan Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Health Technology Interventions for Improving Medication Safety: Systematic Review of Economic Evaluations

Widya Norma Insani[1,2], PhD; Neily Zakiyah[1,2], PhD; Irma Melyani Puspitasari[1,2], Prof Dr; Muhammad Yorga Permana[3], PhD; Kankan Parmikanti[4], MStat; Endang Rusyaman[4], Prof Dr; Auliya Abdurrohim Suwantika[1,2], Prof Dr

[1]Department of Pharmacology and Clinical Pharmacy, Universitas Padjadjaran, Sumedang, Indonesia

[2]Centre of Excellence for Pharmaceutical Care Innovation, Universitas Padjadjaran, Sumedang, Indonesia

[3]School of Business & Management, Bandung Institute of Technology, Bandung, Indonesia

[4]Department of Mathematics, Universitas Padjadjaran, Sumedang, Indonesia

Corresponding Author:
Widya Norma Insani, PhD
Department of Pharmacology and Clinical Pharmacy
Universitas Padjadjaran
Jl Raya Bandung Sumedang KM 21, Jatinangor
Sumedang, 45363
Indonesia
Phone: 62 7796200
Email: widya.insani@unpad.ac.id

Abstract

Background: Medication-related harm, including adverse drug events (ADEs) and medication errors, represents a significant iatrogenic burden in clinical care. Digital health technology (DHT) interventions can significantly enhance medication safety outcomes. Although the clinical effectiveness of DHT for medication safety has been relatively well studied, much less is known about the cost-effectiveness of these interventions.

Objective: This study aimed to systematically review the economic impact of DHT interventions on medication safety and examine methodological challenges to inform future research directions.

Methods: A systematic search was conducted across 3 major electronic databases (ie, PubMed, Scopus, and EBSCOhost). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed for this systematic review. Two independent investigators conducted a full-text review after screening preliminary titles and abstracts. We adopted recommendations from the Panel on Cost-Effectiveness in Health and Medicine for data extraction. A narrative analysis was conducted to synthesize clinical and economic outcomes. The quality of reporting for the included studies was assessed using the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) guidelines.

Results: We included 13 studies that assessed the cost-effectiveness (n=9, 69.2%), cost-benefit (n=3, 23.1%), and cost-utility (n=1, 7.7%) of DHT for medication safety. Of the included studies, more than half (n=7, 53.9%) evaluated a clinical decision support system (CDSS)/computerized provider order entry (CPOE), 4 (30.8%) examined automated medication-dispensing systems, and 2 (15.4%) focused on pharmacist-led outreach programs targeting health care professionals. In 12 (92.3%) studies, DHT was either cost-effective or cost beneficial compared to standard care. On average, DHT interventions reduced ADEs by 37.12% (range 8.2%-66.5%) and medication errors by 54.38% (range 24%-83%). The key drivers of cost-effectiveness included reductions in outcomes, the proportion of errors resulting in ADEs, and implementation costs. Despite a significant upfront cost, DHT showed a return on investment within 3-4.25 years due to lower cost related with ADE treatment and improved workflow efficiency. In terms of reporting quality, the studies were classified as good (n=10, 76.9%) and moderate (n=3, 23.1%). Key methodological challenges included short follow-up periods, the absence of alert compliance tracking, the lack of ADE and error severity categorization, and omission of indirect costs.

Conclusions: DHT interventions are economically viable to improve medication safety, with a substantial reduction in ADEs and medication errors. Future studies should prioritize incorporating alert compliance tracking, ADE and error severity classification, and evaluation of indirect costs, thereby increasing clinical benefits and economic viability.

(J Med Internet Res 2025;27:e65546) doi:10.2196/65546

XSL•FO
RenderX

KEYWORDS

digital health technology; drug safety; adverse drug events; medication errors; patient safety

Introduction

Medication-related harm, including adverse drug events (ADEs) and medication errors, represents a significant burden in clinical care [1]. A previous systematic review showed that 3% of patients in various health care settings experience preventable medication-related harm worldwide, with a quarter classified as severe or potentially life threatening [2]. Unsafe medication practices can occur at various stages of the medication process, including prescribing errors, where nearly 50% of medication errors occur (eg, inappropriate drugs for age and condition, incorrect dosage, contraindication, and drug-drug interactions (DDIs) overlooked) [3,4]; dispensing errors; administration errors; and inadequate monitoring (eg, hypokalemia due to inadequate renal and electrolyte monitoring among diuretic users) [5,6]. Additional costs associated with medication errors have been estimated at US $42 billion annually, indicating a significant burden of unsafe medication practices for patients and the health system [7].

Digital health technology (DHT) applies information and communication technology to enhance health care outcomes [8]. DHT interventions range from a clinical decision support system (CDSS)/computerized provider order entry (CPOE) or electronic prescribing, automated medication-dispensing systems, telemedicine, and mobile health (mHealth) apps to telephone or text message reminders [9-12]. DHT may assist in supporting clinical decisions and enhancing the monitoring of medication use [13,14]. Several studies have shown that DHT may confer benefits by reducing ADEs and medication errors in both hospital and community settings [15-17]. Previous systematic reviews have examined the clinical impact of DHT interventions (ie, CPOE) to improve medication safety [18,19], but none have examined the economic impact of such interventions.

Given the increasing application of DHT, there is an urgent need to determine the economic benefits of DHT to allow clinicians to decide whether to implement DHT strategies [20]. Although DHT interventions can potentially improve clinical outcomes, their additional expense for infrastructure, implementation, and maintenance may hinder their adoption compared to standard care [21,22]. This study aimed to systematically review the economic impact of DHT interventions on medication safety and examine the methodological challenges in these interventions. Such understanding can better inform resource allocation and policy decisions for strengthening health system capacity [23].

Methods

Search Strategy

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guideline. A systematic search was performed across 3 major electronic databases (PubMed, Scopus, and EBSCOhost) to identify studies investigating economic evaluations of DHT to improve medication safety outcomes (ie, reduction in ADEs and medication errors). The search strategies included 3 categories of terms related to:

- DHT interventions, defined as strategies that use information and communication technology to enhance health outcomes (ie, reduction in ADEs and medication errors). To ensure comprehensive coverage of all available evidence, we included a broad range of DHT interventions, including a CDSS/CPOE, electronic prescribing, telemedicine, telepharmacy, automated medication-dispensing systems, mHealth apps, and telephone or text message reminders. Comparators included paper-based prescribing, traditional floor stock storage, and standard pharmaceutical care.
- Medication safety outcomes (ie, changes in the ADE and medication error rate). An ADE is defined as “an injury resulting from the use of a drug,” which includes harm resulting from either errors or medication-inherent effects [24]. The term “medication error” refers to “any error in the process of prescribing, dispensing, or administering a drug, which may or may not result in harm” [25]. A previous study showed that although medication errors are common, only around 1% result in actual harm or ADEs, which might be because these errors have little potential for injury or they are intercepted before an adverse outcome occurs [24,26].
- Economic evaluations (ie, cost-benefit analysis [CBA], cost-effectiveness analysis [CEA], cost-utility analysis [CUA], and cost minimization analysis [CMA]).

Full details of the search strategy are provided in Table 1. An additional search of reference lists of relevant reviews and included studies was performed.

XSL•FO
RenderX

Table 1. Search strategy of DHT[a] interventions for medication safety.

Type of key terms	Search strategy
Terms related to DHT interventions	Telemedicine[MeSH[b]] OR telepharmacy[tiab] OR computerized provider order entry[tiab] OR computerized physician order entry[tiab] OR computer provider entry[tiab] OR clinical decision support[tiab] OR automated medication system[tiab] OR automated pharmacy system[tiab] OR bar coding[tiab] OR electronic medication order entry[tiab] OR electronic medication management system[tiab] OR electronic prescribing[tiab] OR ePrescribing OR electronic prescription[tiab] OR electronic medication administration records[tiab] OR electronic system*[tiab] OR automated dispensing[tiab] OR computerized reminder system[tiab] OR information technology[tiab] OR medication ordering entry[tiab] OR electronic medication ordering and administration system[tiab] OR remote consultation[MeSH] OR electronic consult*[tiab] OR digital technolog*[tiab] OR teleconsult*[tiab] OR mhealth[tiab] OR m-health[tiab] OR multimedia[tiab] OR virtual[tiab] OR mobile health[tiab] OR telemedicine[tiab] OR electronic health record[tiab] OR telehealth[tiab] OR telecare[tiab] OR telehealth care[tiab] OR mobile health intervention*[tiab] OR mobile applications[tiab] OR mobile telemedicine[tiab] OR mcare[tiab] OR m-care[tiab] OR mobile communication[tiab] OR mobile technolog*[tiab] OR multimedia technolog*[tiab] OR mobile devic*[tiab] OR app[tiab] OR apps[tiab] OR mobile app*[tiab] OR website*[tiab] OR internet consultation*[tiab] OR internet monitoring[tiab] OR video consultation*[tiab] OR video monitoring[tiab] OR telephone*[tiab] OR mobile phone*[tiab] OR smart phone*[tiab] OR smart-phone*[tiab] OR text messag*[tiab] OR text messaging[tiab] OR SMS[tiab] OR short messag*[tiab] OR multimedia messag*[tiab] OR multi-media messag*[tiab] OR website platform[tiab] OR web-based medication platform[tiab] OR web-based application[tiab] OR web-based tool[tiab] OR electronic health[tiab] OR ehealth[tiab] OR e-health[tiab]
Terms related to medication safety outcomes	Drug-related side effects and adverse reactions [MeSH] OR adverse drug reaction*[tiab] OR adverse drug event*[tiab] OR drug related problem*[tiab] OR medication related problem[tiab] OR drug therapy problem*[tiab] OR drug safety[tiab] OR medication safety[tiab] OR medication error*[tiab] OR prescribing error*[tiab] OR prescription error*[tiab] OR dispensing error*[tiab] OR administration error*[tiab] OR inappropriate prescribing[tiab] OR inappropriate medication*[tiab] OR drug complication*[tiab]
Terms related to economic evaluations	Cost and cost analysis[MeSH] OR cost-benefit analysis[MeSH] OR cost-effectiveness[tiab] OR cost utility[tiab] OR cost minimi*[tiab] OR economic evaluation[tiab] OR economic analysis[tiab]

[a]DHT: digital health technology.

[b]MeSH: Medical Subject Headings.

Inclusion and Exclusion Criteria

We included studies that reported health economic evaluations of DHT to improve medication safety outcomes (ie, ADEs and medication errors) compared to standard care. Detailed information related to inclusion criteria is provided in Table 2. We excluded studies where the intervention did not include DHT, studies that did not report medication safety outcomes, studies without full economic evaluation, non-English studies, conference abstracts, and editorials.

XSL•FO
RenderX

Table 2. PICO[a] framework and inclusion criteria of the study.

Framework item	Inclusion criteria
Population/problem	Patients receiving medication at any point of care, including in hospital and community settings
Intervention	Any DHT[b] intervention, including: • CDSS[c]/CPOE[d] • Automated medication-dispensing system • Telepharmacy, telemedicine • mHealth[e] app • Other
Comparator	Standard care (eg, paper-based prescribing, traditional floor stock storage)
Outcome	Reduction in ADEs[f] and medication errors
Study type	A full economic evaluation of DHT to improve medication safety categorized as CBA[g], CEA[h], CUA[i], and CMA[j]

[a]PICO: population/problem, intervention, comparator, outcome.
[b]DHT: digital health technology.
[c]CDSS: clinical decision support system.
[d]CPOE: computerized provider order entry.
[e]mHealth: mobile health.
[f]ADE: adverse drug event.
[g]CBA: cost-benefit analysis.
[h]CEA: cost-effectiveness analysis.
[i]CUA: cost-utility analysis.
[j]CMA: cost minimization analysis.

Screening and Data Extraction

The database-screening results were exported to the Mendeley reference manager library and examined for duplicates. Two investigators (authors WNI and NZ) independently conducted a full-text review after screening the preliminary titles and abstracts. Any discrepancies between the 2 reviewers were resolved through discussion. We adopted recommendations from the Panel on Cost-Effectiveness in Health and Medicine for data extraction. Data extracted included (1) study characteristics; (2) clinical and economic outcomes, including the incremental cost-effectiveness ratio (ICER), the cost-benefit ratio (CBR), and the return on investment (ROI); (3) cost components; (4) methodological challenges; and (5) quality of reporting. All monetary values were converted to 2024 US dollar values using the Campbell and Cochrane Economics Methods Group–Evidence for Policy & Practice Information Centre Cost Converter [27].

Quality of Reporting

Reporting quality was assessed using the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) 2022 checklist. The checklist included 28 items, classified into 6 categories: (1) title and abstract, (2) introduction, (3) methods (including choice of model, health outcomes, and measurement of effectiveness), (4) results, (5) discussion, and (6) others. Based on the reporting quality, the included studies were categorized as excellent (score 100%), good (score 75%-99%), moderate (score 50%-74%), and low (score ≤49%) [28].

Results

Study Selection

A total of 408 citations were retrieved from the electronic databases. After removing duplicates, 355 (87%) papers remained for evaluation based on titles and abstracts, yielding 49 (13.8%) records eligible for full-text assessment. A total of 13 (26.5%) economic evaluation studies [10,12,15,17,29-37] were finally included in this systematic review. Most of the studies included CEA (n=9, 69.2%) [12,15,17,29-34], followed by CBA (n=3, 23.1%) [10,35,36], and CUA (n=1, 7.7%) [37]. The included studies were predominantly conducted in hospital inpatient settings (n=10, 76.9%) [10,12,17,30-36]. The study selection flowchart based on PRISMA guidelines is provided in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA flowchart of study selection. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

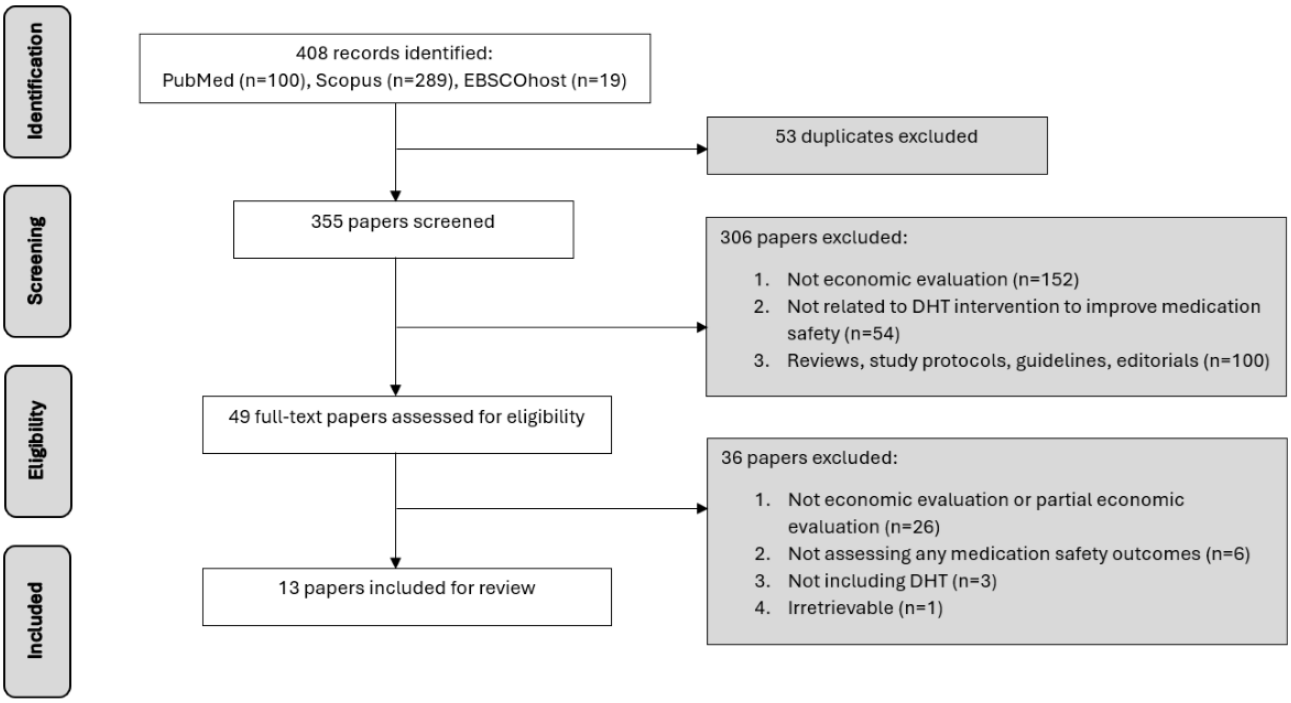

Overview of Key Characteristics

The most frequently implemented DHT intervention was the CDSS/CPOE or electronic medication record system. This intervention typically replaced traditional paper-based prescribing, with a safety alerts feature to assist clinicians in making informed decisions during the prescribing process [12,17,29,30,33-35]. A wide range of features were used in the included studies, ranging from basic DDI to the addition of medication administration tracking and various alerts, including pregnancy contraindications, allergy checks, dosage checks, therapeutic duplications, and potentially inappropriate medication (PIM) usage in vulnerable populations at increased risk of ADEs [12,17,29,30,33-35]. One study [30] compared a structured pharmacist medication review supported by a CDSS with a review conducted without CDSS support for older hospitalized patients. The intervention included medication reconciliation and personalized pharmaceutical care (Table 3).

Of the 13 studies, 4 (30.8%) [10,31,32,36] showed that the implementation of automated medication-dispensing systems has been increasingly used to reduce dispensing and administration errors, enhance workflow efficiency, and improve stock inventory tracking. These systems typically incorporated automated individual unit-dose dispensing, which ensured precise medication packaging tailored to patient-specific prescriptions, unlike manual systems where nurses prepare doses manually [31,32,36]. The systems were often integrated with barcode scanning technology, either at the hospital pharmacy or at the patient bedside level, enabling verification of both the medication and the patient's identity to prevent administration errors [31,32]. Additional functionalities included structured drug storage and controlled access (eg, automated dispensing cabinets [ADCs] and medicine carousel systems with rotating shelves or bins to minimize the time spent searching for medications and medication selection errors) [10,32,36]. One intervention also improved inventory management by providing automated real-time stock monitoring [10]. Furthermore, some systems were linked with electronic medication administration records and other CDSS tools, allowing better coordination between dispensing and prescribing processes [31,32,36].

XSL•FO
RenderX

Table 3. Overview of key characteristics of the 13 included studies.

Author	Country	DHT[a] type	Target population	Sample size	Type of study	Time horizon
Vermeulen et al [12]	The Netherlands	CDSS[b]/CPOE[c]	Patients admitted to the internal medicine, gastroenterology, or geriatric ward	1195	CEA[d]	Admission to discharge
Westbrook et al [33]	Australia	CDSS/CPOE	Patients admitted to the cardiology ward	1202	CEA	15 years
Wu et al [34]	Canada	CDSS/CPOE	Patients admitted to all wards	74,351	CEA	10 years
Avery et al [15]	United Kingdom	IT-based pharmacist outreach	Targeted patients based on conditions and medications in general practices	480,942	CEA	6 months
Berdot et al [10]	France	Automated medication dispending	Patients admitted to all wards	70,421	CBA[e]	1 year
Forrester et al [29]	United States	CDSS/CPOE	Patients in multidisciplinary outpatient clinics	10,080	CEA	5 years
Elliot et al [37]	United Kingdom	IT-based pharmacist outreach	Targeted patients based on conditions and medications in general practices	480,942	CUA[f]	6 months
Gallagher et al [30]	Ireland	CDSS and pharmacist review	Older hospitalized patients	737	CEA	Admission to discharge or 10-day follow-up
Li et al [35]	China	CDSS/CPOE	Inpatients and outpatients	620,000	CBA	6 years
Maviglia et al [36]	United States	Barcode dispensing	Patients admitted to all wards	175,000	CBA	5 years
Nuckols et al [17]	United States	CPOE/CDSS	Patients admitted to all wards	4891[g]	CEA	10 years
Risør et al [31]	Denmark	Automated medication dispensing	Patients admitted to the hematological ward	1336	CEA	6 months
Risør et al [32]	Denmark	Automated medication dispensing	Patients admitted to acute wards	90,000[h]	CEA	6 months

[a]DHT: digital health technology.

[b]CDSS: clinical decision support system.

[c]CPOE: computerized provider order entry.

[d]CEA: cost-effectiveness analysis.

[e]CBA: cost-benefit analysis.

[f]CUA: cost-utility analysis.

[g]Number of acute care hospitals in the United States.

[h]Total number of doses.

A technology-based pharmacist outreach was implemented in 2 (15.4%) studies [15,37], involving pharmacists engaging directly with other health care professionals to target specific high-risk prescribing errors, such as prescribing nonsteroidal anti-inflammatory drugs (NSAIDs) without proton pump inhibitors (PPIs) for patients with a history of peptic ulcers, beta-blockers for patients with asthma, and angiotensin-converting enzyme (ACE) inhibitors or diuretics without proper monitoring of renal function and electrolytes. The interventions involved a CDSS-supported feedback system and pharmacist educational outreach for general practice staff. This intervention model extended beyond simple error reporting by providing pharmacist-intensive support and guidance to other health care professionals.

Clinical Effectiveness Estimates

The DHT interventions were effective in reducing ADEs, with an average reported effectiveness of 37.12% (range 8.2%-66.5%) across the included studies (Multimedia Appendix 2 and Table 4) [12,17,29,30,34,35]. Medication errors reduced by an average of 54.38% (range 24%-83%) [10,12,15,17,29,31-33,36]. The variability in the effectiveness of DHT in reducing ADEs and medication errors can be influenced by differences in DHT features and the target population. A comprehensive CDSS/CPOE with both prescription entry and administration tracking resulted in greater reduction in ADEs compared to a CDSS/CPOE with only basic alerts [12,34]. Nevertheless, none of the studies provided data on alert compliance tracking by health care professionals. Higher effectiveness was also observed in DHT interventions targeting high-risk populations, such as older and pediatric hospitalized patients. These populations are more prone to ADEs and medication errors resulting from comorbidities, concomitant medications, and differences in physiological characteristics. Consequently, interventions targeting these populations showed greater absolute reductions in ADEs and medication errors compared to interventions targeting lower-risk populations, as the elevated baseline risk provides more room for improvement [29,30].

XSL•FO
RenderX

Table 4. Examples of ADEs[a] and medication errors assessed in the included studies.

Safety outcomes	Examples	Reduction (%), mean (range)
ADEs	• Dyspepsia associated with NSAIDs[b], antiplatelets, and corticosteroids use • Severe hypotension resulting from ACE[c] inhibitors in patients with volume depletion • Acute kidney injury caused by NSAIDs prescribed without proper renal monitoring • Hyperkalemia due to the concurrent use of ACE inhibitors and potassium-sparing diuretics • Bleeding complications caused by an inappropriate dose of anticoagulants • Hypoglycemia resulting from insulin overdose • Bradycardia from incorrect beta-blocker dispensing • Falls and fractures caused by benzodiazepine use in older patients • Renal toxicity from aminoglycosides prescribed without dose adjustment for kidney function • Hypotension and syncope caused by inappropriate antihypertensive drug combinations • Cardiac arrhythmia caused by QT-prolonging drug combinations • Severe allergic reactions from failure to recognize and document allergies	37.12 (8.20-66.50)
Medication errors	• Incorrect drug dosing due to incorrect weight-based calculation and transcription error • Failure to adjust medication doses for patients with renal impairment • NSAIDs without PPIs[d] in patients with ulcer history • Beta-blockers prescribed to patients with asthma • Dispensing the incorrect medication, strength, or dosage form • Errors related to supply failure or expired drugs • Monitoring errors, such as patients on ACE inhibitors or diuretics without renal and electrolyte monitoring in the past 15 months, patients on warfarin without a recorded International Normalized Ratio (INR) check in the past 12 weeks, patients on methotrexate without a full blood count or liver function test in the past 3 months	54.38 (24.00-83.00)

[a]ADE: adverse drug event.

[b]NSAID: nonsteroidal anti-inflammatory drug.

[c]ACE: angiotensin-converting enzyme.

[d]PPI: proton pump inhibitor.

Cost-Effectiveness Estimates

CDSS/CPOE Intervention

Implementation of a CDSS/CPOE to replace paper-based prescribing was associated with an average cost ranging from US $25.64 to $81.36 per patient, depending on the complexity and features of the system. The key cost components for a CDSS/CPOE included initial investments in hardware and software, licensing fees, staff training, ongoing system maintenance, and periodic updates to ensure compatibility with clinical workflows (Table 5) [12,17,29,30,33-35]. The implementation cost of a CDSS/CPOE depends on the hospital size, influenced by the need for more extensive hardware infrastructure and software to support a higher number of workstations, higher provider and patient volume, and more comprehensive clinical decision support rules to support various specialties [17]. Most studies evaluating a CDSS/CPOE reported the interventions as cost-effective from perspectives of societal, hospital, and health care systems, with 3 (23.1%) studies [12,17,33] identifying cost-saving results. Only 1 (7.7%) study showed an exceptionally high ICER per ADE prevented without stating a willingness-to-pay (WTP) threshold, highlighting that more data on the effectiveness of CPOE in reducing ADEs are needed [34]. This study was conducted in 2006; therefore, the data on effectiveness may not reflect the most current findings, as presented in more recent studies [12,29,30,33,37]. Although the development of an electronic medical record system with a CDSS requires an initial high investment, 1 (7.7%) study [35] showed that it yields a positive cost-benefit ratio of 1.45, indicating that every US $1 spent generates US $1.45 in benefits, with the ROI achieved within 3 years, driven by a 40% reduction in ADEs as additional treatments and hospitalizations due to ADEs reduced and efficiency improved from transitioning to electronic systems [35].

XSL•FO
RenderX

Table 5. Cost components associated with the DHT[a] interventions included in this review.

Author	DHT type	Currency, year	Cost components	Discount rate (%)
Vermeulen et al [12]	CDSS[b]/CPOE[c]	Euro, 2009	• Software, hardware, and maintenance specific to the CDSS/CPOE • Costs for personnel involved in setting up and maintaining the system • Maintenance and operational costs: ongoing expenses for system updates and operational requirements	Not reported (follow-up period did not exceed 1 year)
Westbrook et al [33]	CDSS/CPOE	AU $, 2012-2013	• Software license fees, infrastructure upgrades, and equipment specific to the electronic medication management system • Personnel training and configuration • Ongoing operating costs: annual software licensing fees, routine system maintenance, and regular training sessions for personnel	5
Wu et al [34]	CDSS/CPOE	US $, 2007	• Software and hardware setup • System configuration and testing • Training sessions for doctors, nurses, and support staff on using the new electronic system • Ongoing operational and maintenance costs: software updates and maintenance, licensing fees, and periodic staff training • Workload costs (considered in sensitivity analysis)	5
Avery et al [15]	IT-based pharmacist outreach	UK pound sterling, 2012	• Report generation • Pharmacist training • Error management activities, including review of patient medical records, consultations with general practitioners, patient follow-up (eg, patient counselling and medication adjustment, if needed), and training for general practice staff by pharmacists	Not reported (follow-up period did not exceed 1 year)
Berdot et al [10]	Automated medication-dispensing system	Euro, 2015	• Purchase of dispensing units, including software and licenses • Immobilized drug stock • Annual maintenance • Labor costs: pharmacy technician wages for ADC[d]-related tasks	Not reported
Forrester et al [29]	CDSS/CPOE	US $, 2010	• Hardware, software, and system setup • Administrative costs: prescription processing, including chart pulls and queuing • Incentives: financial incentives related to meaningful use and pay-for-performance criteria • Maintenance costs • Personnel costs for implementation, training, and support	3
Elliot et al [37]	IT-based pharmacist outreach	UK pound sterling, 2012	• Report generation • Pharmacist training • Error management activities, including review of patient medical records, consultations with general practitioners, patient follow-up (eg, patient counselling and medication adjustment, if needed0, and training for general practice staff by pharmacists	3.5
Gallagher et al [30]	CDSS and pharmacist review	Euro, 2012	• Pharmacist time and training: cost of pharmacists applying structured medication reviews and the CDSS • Health care staff review time: physician and nurse time for reviewing care plans • Hospital inpatient day: cost of inpatient care per day • Software and training costs	Not reported (follow-up period did not exceed 1 year)
Li et al [35]	CDSS/CPOE	US $, 2009	• Hardware and software costs • Implementation costs: workflow setup, training, and transition • Maintenance costs: ongoing system support and utilities	10
Maviglia et al [36]	Barcode-dispensing system	US $, 2009	• Planning cost: workflow redesign and stakeholder engagement • Software development: linking the barcode system with the existing CPOE system, medication inventory, dose verification tracking, interface design, etc • Equipment purchase and infrastructure changes • Training cost	3

XSL•FO
RenderX

Author	DHT type	Currency, year	Cost components	Discount rate (%)
Nuckols et al [17]	CDSS/CPOE	US $, 2012	• Implementation cost: hardware, software, training, and technical support • Provider workflow cost	3
Risor et al [31]	Automated medication-dispensing system	Euro, 2017	• System integration: development of interfaces between electronic medication administration records, scanners, and the barcode-scanning system • Equipment purchase: automated medication-dispensing machine and barcode-scanning devices • Operational costs: cost of dose packaging, pharmaceutical services for prescription checks, and additional labor costs • Training cost: training for nurses and pharmacy staff on AMS[e] processes	Not reported (follow-up period did not exceed one year)
Risor et al [32]	Automated medication-dispensing system	Euro, 2018	• System implementation: establishment of electronic medication administration records tailored to AMS types • Equipment purchase: automated medication-dispensing machines, scanners, and ADCs (complex automated medication system [CAMS] only) • Operational costs: maintenance costs, dose bag handling, and additional pharmacy labor for prescription checks (only for patient-specific automated medication system [PSAMS] and CAMS) • Training cost: education programs tailored to the AMS complexity used in each ward	Not reported (follow-up period did not exceed one year)

[a]DHT: digital health technology.

[b]CDSS: clinical decision support system.

[c]CPOE: computerized provider order entry.

[d]ADC: automated dispensing cabinet.

[e]AMS: automated medication system.

Automated Medication-Dispensing System

Implementation of automated medication-dispensing systems to replace traditional floor stock systems was cost-effective and beneficial, with the ICER ranging from US $0.33 to $62.00 per medication error avoided [10,31,32,36]. Key cost components for an automated medication-dispensing system included purchase of dispensing cabinets or medicine carousel systems with rotating shelves or bins, equipment for the repackaging center, software development, a barcode scanner, labor for restocking and packaging, training expenses, and system integration with the existing CPOE and pharmacy system. A reported net benefit of US $5,379,938.17 was achieved over 5 years, with the ROI attained within 4.25 years after initiation [10,36]. The key drivers of the cost-effectiveness included handling costs and differences in the rates and types of medication errors [31,32,36]. Different types of dispensing errors have varying cost consequences; for example, the administration of an incorrect drug, dosage, or strength can have more severe consequences than procedural or administrative errors, which may potentially but not necessarily result in actual harm [32].

IT-Based Pharmacist Outreach

The implementation of a pharmacist-led DHT intervention was cost-effective in reducing clinically significant prescribing and monitoring errors in primary care, with an ICER of US $129.8 per medication error prevented and US $7788 per quality-adjusted life year (QALY) per practice [15,37]. In addition, this intervention not only improved health outcomes (ie, additional QALYs) but also reduced overall costs compared to usual care, indicating its potential for cost savings [37]. Key cost components included pharmacist training, the time spent reviewing patient records and consulting with general practitioners (GPs), report generation from the CDSS, facilitation of practice meetings, patient follow-up, and system integration with the existing IT infrastructure. The number of patients per general practice was a key driver of cost-effectiveness in this intervention. In larger practices, intervention costs (eg, pharmacist training and CDSS report generation) are distributed across a broader patient base. In addition, the targeted nature of this intervention, which addressed specific prescribing errors (ie, NSAIDs without PPIs, beta-blockers for asthma, and monitoring errors) increased the efficiency and impact of pharmacists' and GPs' time in practices with higher patient volumes due to economies of scale [15,37,38].

Methodological Characteristics and Challenges in the Included Studies

Of the 13 studies 6 (46.2%) [10,12,17,31,32,35] used a quasi-experimental design (ie, before-after and interrupted-time-series approaches, with multiple data points tracking ADEs and medication errors before and after DHT initiation). A decision analytic model was used in 3 (23.1%) studies [29,33,35], focusing on short-term, event-specific analysis, while a Markov model was used in 1 (7.7%) study [37]. Randomized controlled trial (RCT)–based evaluation was used in 2 (15.4%) studies [15,30]. Challenges related to methodological aspects of evaluating DHT interventions for medication safety are presented in Table 6. Methodological challenges included short follow-up times, the absence of CDSS alert compliance tracking, a lack of ADE severity classification, and the omission of indirect costs (eg, productivity loss and caregiver time).

XSL•FO
RenderX

Table 6. Methodological challenges in DHT[a] intervention evaluation for medication safety.

DHT type and methodological issues	Description
CDSS[b] /CPOE[c]	
Study design	• The short-term duration of the study (hospital stay until discharge or 10-day follow-up) did not capture the medium- or long-term impact of the intervention [12,30]. • Studies were not able to track whether or how many CDSS alerts were acknowledged and acted upon by health care professionals, limiting accurate assessment of DHT's effectiveness [12,17,29,30,33-35]. • There was no standard care comparator. Before-after designs may not fully account for external factors influencing the observed trend [12,33,35]. The use of a parallel control group for direct comparison could help mitigate confounding effects.
Cost data	• Studies included a societal perspective but lacked data on indirect patient costs (eg, productivity loss, caregiver time) [17].
Clinical data	• There was a lack of detailed ADE[d] severity classification, which may introduce variability in assessing ADE consequences and the benefits of the intervention [33]. • Few studies assessed direct patient-centered outcomes (eg, hospital readmissions, QALY[e]) [17].
Automated medication-dispensing system	
Study design	• Studies did not analyze whether identified dispensing errors were tracked or immediately corrected (real-time monitoring), relying on retrospective analysis [10,31,32,36]. • Studies were conducted only in hospital settings, limiting generalizability to different settings.
Cost data	• Studies did not explore how saved handling time could be reallocated to other productive tasks (ie, opportunity cost).
Clinical data	• Studies focused on refill errors and urgent deliveries but did not systematically track errors or discrepancies between prescription and dispensing records [10].
Pharmacist-led IT intervention	
Study design	• Studies focused on specific high-risk prescribing and monitoring errors but did not directly assess the proportion of errors leading to ADEs [15,37].

[a]DHT: digital health technology.

[b]CDSS: clinical decision support system.

[c]CPOE: computerized provider order entry.

[d]ADE: adverse drug event.

[e]QALY: quality-adjusted life year.

Quality of Reporting

Based on the quality assessment using the CHEERS checklist, most of the included studies (n=10, 76.92%) [12,15,17,29-33,36,37] were rated as good, and the remaining 3 (23.1%) studies [10,34,35] were rated as moderate. Adherence to the checklist items varied across the section. Most studies did not report any approach to engagement with patients in the study design and results, except for the RCT-based study [30]. The rationale for selecting the model was inadequately reported in over a third of the studies, indicating issues in validating the selected methodology [10,30,31,34,36]. All studies reported how to measure ADEs and medication errors; however, the valuation of these safety outcomes was not reported in some studies [10,12,33]. The components of the outcomes included costs related to additional treatments and hospital stays due to the occurrence of ADEs and medication errors. Although most studies conducted sensitivity analysis to assess the robustness of the findings, only 1 (7.7%) study [32] assessed the heterogeneity of the outcomes based on different types of automated medication-dispensing systems. Most studies only focused on aggregated data without exploring subgroup differences, such as variations within patient populations, hospital wards, or intervention types.

Discussion

Principal Findings

This is the first systematic review that assessed the economic evaluations of DHT interventions to improve medication safety outcomes. More than half of the studies (n=7, 53.9%) focused on a CDSS/CPOE, less than a third (n=4, 30.8%) on automated medication-dispensing systems, and the remaining (n=2, 15.4%) on pharmacist-led outreach programs targeting health care professionals. On average, DHT interventions reduced ADEs by 37.12% and medication errors by 54.38%. In 92.3% (12/13) of the included studies, the DHT was either cost-effective or cost beneficial compared to standard care. Despite a significant upfront cost, DHT showed an ROI within 3-4.25 years. Key methodological challenges included short follow-up periods, a

XSL•FO
RenderX

lack of ADE severity categorization, the absence of alert compliance tracking, and the omission of indirect costs.

A CDSS/CPOE has been increasingly used to support clinicians in making informed decisions by providing recommendations based on patient data and clinical guidelines [39]. Nevertheless, the effectiveness of a CDSS is often compromised by alert fatigue, which occurs when clinicians override a large number of potentially irrelevant drug safety alerts (eg, drug interactions that are not clinically significant, flagging dosages outside the standard guideline when the prescribed dosage is appropriate for the patient condition) [40-42]. A previous study showed that alert fatigue might be reduced by using a CDSS/CPOE with interactive features, such as incorporating tiered safety alerts with varying priority levels, offering action plans for clinicians (eg, dose reduction), and requiring clinicians to justify overriding an alert [41]. In our review, none of the studies tracked CDSS/CPOE alert compliance, which may influence their effectiveness. A previous study showed drug allergy alerts had the highest compliance [43]. Further DHT evaluation studies should investigate how alert compliance affects the outcomes to ensure optimal utility of DHT.

In addition, tailoring clinical roles within a CDSS can also enhance alert acceptance [44-46]. A pharmacist-mediated CDSS has improved prescriber acceptance by filtering irrelevant advice and providing actionable drug safety recommendations [44]. This approach leverages pharmacists' expertise related to medication to support the decision-making process and reduce alert fatigue among prescribers [47-49]. Several included studies combined a CDSS with pharmacist-led interventions, such as structured medication review and targeted training for health care professionals for error correction, highlighting the central role of pharmacists in medication safety [15,30,37].

Advancements in CDSSs/CPOE have significantly reduced prescribing errors, but administration errors, such as administering the wrong dose, using the incorrect route of administration, and failing to administer a scheduled dose, still present substantial room for improvement [50]. This review found that a CDSS/CPOE with integrated prescription entry and administration tracking that address prescribing errors, while also monitoring and mitigating administration errors, achieves a greater reduction in ADEs compared to systems with basic features [12,34]. The development and implementation of more comprehensive systems that target all stages of medication management, including administration, are essential to reduce medication errors and ADEs.

Different types of automated medication-dispensing systems were used in the included studies. Several factors need to be considered when selecting an automated medication-dispensing system, including the volume of prescriptions and types of medication the device can handle (eg, oral tablets, liquid medications, injectables), integration with the existing information technology system, and the presence of an error-checking mechanism [51,52]. Almaki et al [53] and Tsao et al [54] showed that integration with existing digital infrastructure is key to supporting a seamless workflow process to ensure all components of medication management (ie, prescribing, dispensing, and inventory control) are interconnected, enhancing efficiency and medication error prevention.

All the included studies focused on hospital-based automated medication-dispensing systems, with limited investigation in outpatient settings [10,31,32,36]. Williams et al [52] demonstrated that automated medication-dispensing systems for chronic medication regimens, such as antiretroviral therapy, enable efficient one-time password (OTP)–based medication collection and reduced waiting times, benefiting high-volume, resource-limited outpatient clinics. Further studies may adopt this system for less complex medication regimens to improve patient satisfaction and reduce the staff burden, enabling health care providers to focus on other critical tasks, such as optimization of drug therapy [54,55].

Targeting medication errors with a substantial clinical impact was a key strategy in the technology-based pharmacist-led outreach [15,37]. As around 1% of medication errors lead to actual harm, prioritizing interventions is essential [24]. The National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) developed an index to standardize the classification of medication errors based on severity (ie, no error, error-no harm, error-harm, and error-death), which has been adopted in the hospital setting [56,57]. An initiative to standardize the interception of medication errors using this severity classification index has proven effective to prevent clinically significant patient harm and provide substantial cost savings to the health system [58].

Various economic models were used in the included studies (eg, quasi-experimental design, RCT, decision analytic model), indicating the complexity and multifaceted nature of DHT for medication safety. Sculpher et al [59] emphasized the importance of placing an RCT within a broader framework of evidence synthesis and decision analysis in economic evaluation studies to balance internal validity and broader applicability. Quasi-experimental design can be practical for assessing real-world effectiveness and scalability. However, the assessment of confounders that might affect the observed trend and the lack of guidance in conducting economic evaluations alongside quasi-experimental trials necessitate careful interpretation of the findings [60].

Strengths and Limitations

This study is the first systematic review investigating economic evaluations of DHT interventions to improve medication safety outcomes. We conducted a rigorous literature search with a comprehensive strategy encompassing a wide range of DHT intervention types to provide a thorough review of different strategies to improve medication safety. We also included evaluations on methodological challenges in DHT intervention assessment to inform future research direction.

Nevertheless, our review has several potential limitations. First, the generalizability of findings may be restricted to high-income settings, as most included studies were conducted in such contexts. Second, heterogeneity in health care settings, study designs, and economic evaluation methods may hinder direct comparisons across studies. Despite this, a narrative synthesis was developed to integrate and interpret the findings. Third, the

XSL•FO
RenderX

exclusion of non-English studies may have limited the comprehensiveness of the evidence base. Finally, there was a lack of medication error and ADE severity classification among the included studies. Since different medication errors and ADEs have varying cost implications, this variability makes the interpretation of overall cost-effectiveness less straightforward.

Conclusion

DHT interventions are economically viable to improve medication safety, with substantial reduction in ADEs and medication errors. On average, DHT interventions reduced ADEs by 37.12% and medication errors by 54.38%. In 92.3% of the included studies, DHT was either cost-effective or cost beneficial compared to standard care. Despite a significant upfront cost, DHT showed an ROI within 3-4.25 years. The key drivers of cost-effectiveness include reductions in outcomes, the proportion of errors resulting in ADEs, and implementation costs. Key methodological challenges included short follow-up periods, the absence of compliance tracking, the lack of ADE severity categorization, and the omission of indirect costs. Future studies should prioritize incorporating alert compliance tracking, ADE and medication error severity classification, and the evaluation of indirect costs, thereby increasing clinical benefits and economic viability.

Acknowledgments

This study was supported by an Academic Leadership Grant (ALG) from Padjadjaran University in 2024 through AAS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[PDF File (Adobe PDF File), 66 KB - jmir_v27i1e65546_app1.pdf]

Multimedia Appendix 2
Clinical and economic outcomes of the included studies.
[DOCX File , 27 KB - jmir_v27i1e65546_app2.docx]

References

1. World Health Organization. Global burden of preventable medication-related harm in health care: a systematic review. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/9789240088887 [accessed 2024-04-01]
2. Hodkinson A, Tyler N, Ashcroft DM, Keers RN, Khan K, Phipps D, et al. Preventable medication harm across health care settings: a systematic review and meta-analysis. BMC Med 2020 Nov 06;18(1):313 [FREE Full text] [doi: 10.1186/s12916-020-01774-9] [Medline: 33153451]
3. Velo GP, Minuz P. Medication errors: prescribing faults and prescription errors. Br J Clin Pharmacol 2009 Jun 25;67(6):624-628 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03425.x] [Medline: 19594530]
4. Wheeler AJ, Scahill S, Hopcroft D, Stapleton H. Reducing medication errors at transitions of care is everyone's business. Aust Prescr 2018 Jun 01;41(3):73-77 [FREE Full text] [doi: 10.18773/austprescr.2018.021] [Medline: 29922001]
5. Piazza G, Nguyen TN, Cios D, Labreche M, Hohlfelder B, Fanikos J, et al. Anticoagulation-associated adverse drug events. Am J Med 2011 Dec;124(12):1136-1142 [FREE Full text] [doi: 10.1016/j.amjmed.2011.06.009] [Medline: 22114827]
6. Schepkens H, Vanholder R, Billiouw J, Lameire N. Life-threatening hyperkalemia during combined therapy with angiotensin-converting enzyme inhibitors and spironolactone: an analysis of 25 cases. Am J Med 2001 Apr 15;110(6):438-441. [doi: 10.1016/s0002-9343(01)00642-8] [Medline: 11331054]
7. Aitken M. Advancing the Responsible Use of Medicines: Applying Levers for Change. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2012.
8. Rowlands D. What is digital health? Health Informatics Society of Australia. 2019. URL: https://www.hisa.org.au/wpcontent/uploads/2019/12/What_is_Digital_Health.pdf [accessed 2025-01-23]
9. Afreen N, Padilla-Tolentino E, McGinnis B. Identifying potential high-risk medication errors using telepharmacy and a web-based survey tool. Innov Pharm 2021 Feb 12;12(1):9 [FREE Full text] [doi: 10.24926/iip.v12i1.3377] [Medline: 34007681]
10. Berdot S, Blanc C, Chevalier D, Bezie Y, Lê LMM, Sabatier B. Impact of drug storage systems: a quasi-experimental study with and without an automated-drug dispensing cabinet. Int J Qual Health Care 2019 Apr 01;31(3):225-230 [FREE Full text] [doi: 10.1093/intqhc/mzy155] [Medline: 30020459]
11. Ong SW, Jassal SV, Porter EC, Min KK, Uddin A, Cafazzo JA, et al. Digital applications targeting medication safety in ambulatory high-risk ckd patients: randomized controlled clinical trial. CJASN 2021 Mar 18;16(4):532-542. [doi: 10.2215/cjn.15020920]

XSL•FO
RenderX

12. Vermeulen K, van Doormaal J, Zaal R, Mol P, Lenderink A, Haaijer-Ruskamp F, et al. Cost-effectiveness of an electronic medication ordering system (CPOE/CDSS) in hospitalized patients. Int J Med Inform 2014 Aug;83(8):572-580. [doi: 10.1016/j.ijmedinf.2014.05.003] [Medline: 24929633]
13. Food and Drug Administration. What Is Digital Health?. Silver Spring, MD: Food and Drug Administration; 2020.
14. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407 [FREE Full text] [doi: 10.3390/ijerph20043407] [Medline: 36834105]
15. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M, et al. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. Lancet 2012 Apr;379(9823):1310-1319. [doi: 10.1016/s0140-6736(11)61817-5]
16. Leon C, Hogan H, Jani YH. Identifying and mapping measures of medication safety during transfer of care in a digital era: a scoping literature review. BMJ Qual Saf 2024 Feb 19;33(3):173-186 [FREE Full text] [doi: 10.1136/bmjqs-2022-015859] [Medline: 37923372]
17. Nuckols TK, Asch SM, Patel V, Keeler E, Anderson L, Buntin MB, et al. Implementing computerized provider order entry in acute care hospitals in the United States could generate substantial savings to society. Jt Comm J Qual Patient Saf 2015 Aug;41(8):341-350 [FREE Full text] [doi: 10.1016/s1553-7250(15)41045-1] [Medline: 26215523]
18. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. J Am Med Inform Assoc 2008 Sep 01;15(5):585-600. [doi: 10.1197/jamia.m2667]
19. Ranji SR, Rennke S, Wachter RM. Computerised provider order entry combined with clinical decision support systems to improve medication safety: a narrative review. BMJ Qual Saf 2014 Sep 12;23(9):773-780. [doi: 10.1136/bmjqs-2013-002165] [Medline: 24728888]
20. Park T, Kim H, Song S, Griggs SK. Economic evaluation of pharmacist-led digital health interventions: a systematic review. Int J Environ Res Public Health 2022 Sep 22;19(19):11996 [FREE Full text] [doi: 10.3390/ijerph191911996] [Medline: 36231307]
21. Clarke M, Fursse J, Brown-Connolly NE, Sharma U, Jones R. Evaluation of the National Health Service (NHS) direct pilot telehealth program: cost-effectiveness analysis. Telemed J E Health 2018 Jan;24(1):67-76. [doi: 10.1089/tmj.2016.0280] [Medline: 28723244]
22. Lopez-Villegas A, Catalan-Matamoros D, Peiro S, Lappegard KT, Lopez-Liria R. Cost-utility analysis of telemonitoring versus conventional hospital-based follow-up of patients with pacemakers. The NORDLAND randomized clinical trial. PLoS One 2020 Jan 29;15(1):e0226188 [FREE Full text] [doi: 10.1371/journal.pone.0226188] [Medline: 31995558]
23. Slawomirski L, Auraaen A, Klazinga NS. The economics of patient safety. Organisation for Economic Co-operation and Development. 2017. URL: https://www.oecd.org/en/publications/the-economics-of-patient-safety_5a9858cd-en.html [accessed 2025-01-23]
24. Bates DW, Boyle DL, Vliet MBV, Schneider J, Leape L. Relationship between medication errors and adverse drug events. J Gen Intern Med 1995 Apr;10(4):199-205. [doi: 10.1007/bf02600255]
25. Leape L. Preventing adverse drug events. Am J Health Syst Pharm 1995 Feb 15;52(4):379-382. [doi: 10.1093/ajhp/52.4.379] [Medline: 7757863]
26. Montesi G, Lechi A. Prevention of medication errors: detection and audit. Br J Clin Pharmacol 2009 Jun 25;67(6):651-655 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03422.x] [Medline: 19594533]
27. CCEMG: EPPI-Centre cost converter. EPPI Centre. 2022. URL: https://eppi.ioe.ac.uk/costconversion [accessed 2025-01-23]
28. Zakiyah N, van Asselt ADI, Roijmans F, Postma MJ. Economic evaluation of family planning interventions in low and middle income countries: a systematic review. PLoS One 2016 Dec 19;11(12):e0168447 [FREE Full text] [doi: 10.1371/journal.pone.0168447] [Medline: 27992552]
29. Forrester SH, Hepp Z, Roth JA, Wirtz HS, Devine EB. Cost-effectiveness of a computerized provider order entry system in improving medication safety ambulatory care. Value Health 2014 Jun;17(4):340-349 [FREE Full text] [doi: 10.1016/j.jval.2014.01.009] [Medline: 24968993]
30. Gallagher J, O'Sullivan D, McCarthy S, Gillespie P, Woods N, O'Mahony D, et al. Structured pharmacist review of medication in older hospitalised patients: a cost-effectiveness analysis. Drugs Aging 2016 Apr 9;33(4):285-294. [doi: 10.1007/s40266-016-0348-3] [Medline: 26861468]
31. Risør BW, Lisby M, Sørensen J. Cost-effectiveness analysis of an automated medication system implemented in a Danish hospital setting. Value Health 2017 Jul;20(7):886-893 [FREE Full text] [doi: 10.1016/j.jval.2017.03.001] [Medline: 28712617]
32. Risør BW, Lisby M, Sørensen J. Comparative cost-effectiveness analysis of three different automated medication systems implemented in a Danish hospital setting. Appl Health Econ Health Policy 2018 Feb 8;16(1):91-106. [doi: 10.1007/s40258-017-0360-8] [Medline: 29119475]
33. Westbrook J, Gospodarevskaya E, Li L, Richardson K, Roffe D, Heywood M, et al. Cost-effectiveness analysis of a hospital electronic medication management system. J Am Med Inform Assoc 2015 Jul;22(4):784-793 [FREE Full text] [doi: 10.1093/jamia/ocu014] [Medline: 25670756]

XSL•FO
RenderX

34. Wu RC, Laporte A, Ungar WJ. Cost-effectiveness of an electronic medication ordering and administration system in reducing adverse drug events. J Eval Clin Pract 2007 Jun 18;13(3):440-448. [doi: 10.1111/j.1365-2753.2006.00738.x] [Medline: 17518812]
35. Li K, Naganawa S, Wang K, Li P, Kato K, Li X, et al. Study of the cost-benefit analysis of electronic medical record systems in general hospital in China. J Med Syst 2012 Oct 3;36(5):3283-3291. [doi: 10.1007/s10916-011-9819-6] [Medline: 22212632]
36. Maviglia SM. Cost-benefit analysis of a hospital pharmacy bar code solution. Arch Intern Med 2007 Apr 23;167(8):788. [doi: 10.1001/archinte.167.8.788]
37. Elliott RA, Putman KD, Franklin M, Annemans L, Verhaeghe N, Eden M, et al. Cost effectiveness of a pharmacist-led information technology intervention for reducing rates of clinically important errors in medicines management in general practices (PINCER). Pharmacoeconomics 2014 Jun 18;32(6):573-590. [doi: 10.1007/s40273-014-0148-8] [Medline: 24639038]
38. Bernet PM, Singh S. Economies of scale in the production of public health services: an analysis of local health districts in Florida. Am J Public Health 2015 Apr;105(S2):S260-S267. [doi: 10.2105/ajph.2014.302350]
39. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 06;3(1):17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
40. Acheampong F, Anto B, Koffuor G. Medication safety strategies in hospitals--a systematic review. Int J Risk Saf Med 2014;26(3):117-131. [doi: 10.3233/jrs-140623]
41. Hussain M, Reynolds T, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. J Am Med Inform Assoc 2019 Oct 01;26(10):1141-1149 [FREE Full text] [doi: 10.1093/jamia/ocz095] [Medline: 31206159]
42. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006 Mar 01;13(2):138-147. [doi: 10.1197/jamia.m1809]
43. Naeem A, Alwadie A, Alshehri A, Alharbi L, Nawaz M, AlHadidi R, et al. The overriding of computerized physician order entry (CPOE) drug safety alerts fired by the clinical decision support (CDS) tool: evaluation of appropriate responses and alert fatigue solutions. Cureus 2022 Nov;14(11):e31542 [FREE Full text] [doi: 10.7759/cureus.31542] [Medline: 36540455]
44. Joosten H, Drion I, Boogerd KJ, van der Pijl EV, Slingerland RJ, Slaets JPJ, et al. Optimising drug prescribing and dispensing in subjects at risk for drug errors due to renal impairment: improving drug safety in primary healthcare by low eGFR alerts. BMJ Open 2013 Jan 24;3(1):e002068 [FREE Full text] [doi: 10.1136/bmjopen-2012-002068] [Medline: 23355668]
45. Niedrig D, Krattinger R, Jödicke A, Gött C, Bucklar G, Russmann S. Development, implementation and outcome analysis of semi-automated alerts for metformin dose adjustment in hospitalized patients with renal impairment. Pharmacoepidemiol Drug Saf 2016 Oct 14;25(10):1204-1209 [FREE Full text] [doi: 10.1002/pds.4062] [Medline: 27418265]
46. Smith T, Philmon CL, Johnson GD, Ward WS, Rivers LL, Williamson SA, et al. Antimicrobial stewardship in a community hospital: attacking the more difficult problems. Hosp Pharm 2014 Oct 01;49(9):839-846 [FREE Full text] [doi: 10.1310/hpj4909-839] [Medline: 25477615]
47. Armando LG, Miglio G, de Cosmo P, Cena C. Clinical decision support systems to improve drug prescription and therapy optimisation in clinical practice: a scoping review. BMJ Health Care Inform 2023 May 02;30(1):e100683 [FREE Full text] [doi: 10.1136/bmjhci-2022-100683] [Medline: 37130626]
48. Brixner D, Biltaji E, Bress A, Unni S, Ye X, Mamiya T, et al. The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy. J Med Econ 2016 Nov 11;19(3):213-228 [FREE Full text] [doi: 10.3111/13696998.2015.1110160] [Medline: 26478982]
49. Jokanovic N, Tan EC, Sudhakaran S, Kirkpatrick CM, Dooley MJ, Ryan-Atwood TE, et al. Pharmacist-led medication review in community settings: an overview of systematic reviews. Res Social Adm Pharm 2017 Jul;13(4):661-685. [doi: 10.1016/j.sapharm.2016.08.005] [Medline: 27665364]
50. Devine E, Hansen R, Wilson-Norton J, Lawless N, Fisk A, Blough D, et al. The impact of computerized provider order entry on medication errors in a multispecialty group practice. J Am Med Inform Assoc 2010;17(1):78-84. [doi: 10.1197/jamia.m3285]
51. Zheng WY, Lichtner V, Van Dort BA, Baysari MT. The impact of introducing automated dispensing cabinets, barcode medication administration, and closed-loop electronic medication management systems on work processes and safety of controlled medications in hospitals: a systematic review. Res Social Adm Pharm 2021 May;17(5):832-841 [FREE Full text] [doi: 10.1016/j.sapharm.2020.08.001] [Medline: 32891535]
52. Williams V, Haumba S, Ngwenya-Ngcamphalala F, Mafukidze A, Musarapasi N, Byarugaba H, et al. Implementation of the automated medication dispensing system-early lessons from Eswatini. Int J Public Health 2023 Oct 12;68:1606185 [FREE Full text] [doi: 10.3389/ijph.2023.1606185] [Medline: 37901592]
53. Almalki A, Jambi A, Elbehiry B, Albuti H. Improving inpatient medication dispensing with an automated system. Glob J Qual Saf Healthc 2023;6(4):117-125. [doi: 10.36401/jqsh-23-15]

XSL•FO
RenderX

54. Tsao NW, Lo C, Babich M, Shah K, Bansback NJ. Decentralized automated dispensing devices: systematic review of clinical and economic impacts in hospitals. Can J Hosp Pharm 2014 Mar 30;67(2):138-148 [FREE Full text] [doi: 10.4212/cjhp.v67i2.1343] [Medline: 24799724]
55. Gray J, Ludwig B, Temple J, Melby M, Rough S. Comparison of a hybrid medication distribution system to simulated decentralized distribution models. Am J Health Syst Pharm 2013 Aug 01;70(15):1322-1335. [doi: 10.2146/ajhp120512] [Medline: 23867489]
56. About medication errors. National Coordinating Council for Medication Error Reporting and Prevention. 2001 Feb. URL: https://www.nccmerp.org/about-medication-errors [accessed 2025-01-23]
57. Snyder RA, Abarca J, Meza JL, Rothschild JM, Rizos A, Bates DW. Reliability evaluation of the adapted National Coordinating Council Medication Error Reporting and Prevention (NCC MERP) index. Pharmacoepidemiol Drug Saf 2007 Sep 24;16(9):1006-1013. [doi: 10.1002/pds.1423] [Medline: 17523185]
58. Cabri A, Barsegyan N, Postelnick M, Schulz L, Nguyen V, Szwak J, et al. Pharmacist intervention on prescribing errors: use of a standardized approach in the inpatient setting. Am J Health Syst Pharm 2021 Nov 23;78(23):2151-2158 [FREE Full text] [doi: 10.1093/ajhp/zxab278] [Medline: 34283219]
59. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ 2006 Jul 20;15(7):677-687. [doi: 10.1002/hec.1093] [Medline: 16491461]
60. Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. Soc Sci Med 2019 Jan;220:353-361 [FREE Full text] [doi: 10.1016/j.socscimed.2018.11.032] [Medline: 30513485]

Abbreviations

ACE: angiotensin-converting enzyme
ADC: automated dispensing cabinet
ADE: adverse drug event
AMS: automated medication system
CBA: cost-benefit analysis
CBR: cost-benefit ratio
CDSS: clinical decision support system
CEA: cost-effectiveness analysis
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CMA: cost minimization analysis
CPOE: computerized provider order entry
CUA: cost-utility analysis
DDI: drug-drug interaction
DHT: digital health technology
GP: general practitioner
ICER: incremental cost-effectiveness ratio
MeSH: Medical Subject Headings
mHealth: mobile health
NSAID: nonsteroidal anti-inflammatory drug
PPI: proton pump inhibitor
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY: quality-adjusted life year
RCT: randomized controlled trial
ROI: return on investment

Edited by A Schwartz; submitted 20.08.24; peer-reviewed by J Aarts, J Waterson; comments to author 25.09.24; revised version received 03.12.24; accepted 08.01.25; published 05.02.25.

Please cite as:
Insani WN, Zakiyah N, Puspitasari IM, Permana MY, Parmikanti K, Rusyaman E, Suwantika AA
Digital Health Technology Interventions for Improving Medication Safety: Systematic Review of Economic Evaluations
J Med Internet Res 2025;27:e65546
URL: https://www.jmir.org/2025/1/e65546
doi:10.2196/65546
PMID:

XSL•FO
RenderX

©Widya Norma Insani, Neily Zakiyah, Irma Melyani Puspitasari, Muhammad Yorga Permana, Kankan Parmikanti, Endang Rusyaman, Auliya Abdurrohim Suwantika. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Mapping and Summarizing the Research on AI Systems for Automating Medical History Taking and Triage: Scoping Review

Elin Siira[1], PhD; Hanna Johansson[1], MSc; Jens Nygren[1], PhD

School of Health and Welfare, Halmstad University, Halmstad, Sweden

Corresponding Author:
Elin Siira, PhD
School of Health and Welfare
Halmstad University
Box 823
Halmstad, 301 18
Sweden
Phone: 46 70 692 46 13
Email: elin.siira@hh.se

Abstract

Background: The integration of artificial intelligence (AI) systems for automating medical history taking and triage can significantly enhance patient flow in health care systems. Despite the promising performance of numerous AI studies, only a limited number of these systems have been successfully integrated into routine health care practice. To elucidate how AI systems can create value in this context, it is crucial to identify the current state of knowledge, including the readiness of these systems, the facilitators of and barriers to their implementation, and the perspectives of various stakeholders involved in their development and deployment.

Objective: This study aims to map and summarize empirical research on AI systems designed for automating medical history taking and triage in health care settings.

Methods: The study was conducted following the framework proposed by Arksey and O'Malley and adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines. A comprehensive search of 5 databases—PubMed, CINAHL, PsycINFO, Scopus, and Web of Science—was performed. A detailed protocol was established before the review to ensure methodological rigor.

Results: A total of 1248 research publications were identified and screened. Of these, 86 (6.89%) met the eligibility criteria. Notably, most (n=63, 73%) studies were published between 2020 and 2022, with a significant concentration on emergency care (n=32, 37%). Other clinical contexts included radiology (n=12, 14%) and primary care (n=6, 7%). Many (n=15, 17%) studies did not specify a clinical context. Most (n=31, 36%) studies used retrospective designs, while others (n=34, 40%) did not specify their methodologies. The predominant type of AI system identified was the hybrid model (n=68, 79%), with forecasting (n=40, 47%) and recognition (n=36, 42%) being the most common tasks performed. While most (n=70, 81%) studies included patient populations, only 1 (1%) study investigated patients' views on AI-based medical history taking and triage, and 2 (2%) studies considered health care professionals' perspectives. Furthermore, only 6 (7%) studies validated or demonstrated AI systems in relevant clinical settings through real-time model testing, workflow implementation, clinical outcome evaluation, or integration into practice. Most (n=76, 88%) studies were concerned with the prototyping, development, or validation of AI systems. In total, 4 (5%) studies were reviews of several empirical studies conducted in different clinical settings. The facilitators and barriers to AI system implementation were categorized into 4 themes: technical aspects, contextual and cultural considerations, end-user engagement, and evaluation processes.

Conclusions: This review highlights current trends, stakeholder perspectives, stages of innovation development, and key influencing factors related to implementing AI systems in health care. The identified literature gaps regarding stakeholder perspectives and the limited research on AI systems for automating medical history taking and triage indicate significant opportunities for further investigation and development in this evolving field.

(J Med Internet Res 2025;27:e53741) doi:10.2196/53741

KEYWORDS

scoping review; artificial intelligence; AI; medical history taking; triage; health care; automation

XSL•FO
RenderX

Introduction

Background

In health care, efficient management of the flow of patients through the health care system is important to ensure that care is accessible and appropriate, given the needs of patients [1]. With the introduction of artificial intelligence (AI) systems, there has been growing interest in using these technologies to improve patient flow [2]. Research in this area remains in its early stages, constituting a disjointed field [3]. In addition, AI systems hold great promise to make patient flows more efficient at the admission of patients by improving triage [4,5].

Traditionally, triage involves a health care professional taking the patient's medical history to systematically decide the optimal prioritization and assess the appropriate treatment for the patient. There are different types of triage, each with its distinctive elements, such as inpatient triage, emergency department triage, incident triage, military triage, and mass casualty triage, all of which possess the aforementioned characteristics [6]. AI systems could potentially increase the efficiency of triage processes by replacing health care professionals in taking the patient's medical history and directing patients to the most appropriate treatment. This could effectively reduce health care professionals' workload and allow for a more optimal allocation of their time [4,5]. Despite the high hopes and AI systems showing good performance in studies, only a small number have been implemented in operational health care systems [7,8].

To fully understand how AI systems for automating medical history taking and triage could be used to create value in the health care context, it is important to understand the entire development and implementation process, including data collection, algorithm development, model validation, and real-world implementation [9]. Each of these steps presents unique challenges and considerations that must be carefully addressed to ensure successful integration into health care. It is essential to determine at which stage of the innovation development process the AI systems are halted or hindered before they are integrated into health care practices [7,8]. Many technical and nontechnical factors affect the implementation process. Technical challenges, such as data quality and interoperability, algorithm accuracy, and system reliability, must be carefully evaluated and addressed [9-11]. In addition, ethical considerations, regulatory frameworks, organizational barriers, and patient acceptance play critical roles in the implementation and use of these technologies [10]. By studying these factors, we can identify the barriers that hinder the effective implementation of AI systems and facilitate strategies to overcome them.

However, it is equally important to recognize that the successful implementation of AI systems in health care extends beyond the technical and regulatory aspects. The perspectives of a wide range of stakeholders must be considered as they can offer valuable insights and contribute to the overall effectiveness and acceptance of these systems [12,13]. Stakeholders include health care providers, patients, decision makers, regulatory bodies, administrators, and other context-based relevant actors. Each stakeholder group brings unique perspectives, priorities, and concerns, which must be integrated into the development and deployment of AI systems for medical history taking and triage [12,13]. By incorporating such diverse perspectives, we can ensure that these technologies align with the needs, values, and expectations of the stakeholders they seek to serve.

In light of these considerations, it is crucial to conduct a scoping review to map and summarize empirical research on AI systems for automating medical history taking and triage. By systematically reviewing the literature, we can identify the current state of knowledge across various stages of the innovation development process, gain insights into the hindering and facilitating factors affecting the implementation and use of AI systems, and consider the perspectives of different stakeholders. Such a comprehensive review will not only help identify gaps in knowledge but also guide future research efforts and inform evidence-based practice for integrating AI systems for automating medical history taking and triage in health care settings.

Aim and Research Questions

The aim of the scoping review was to map and summarize empirical research on AI systems for automating medical history taking and triage. The following research questions (RQs) guided this review:

1. What are the characteristics of research publications on AI systems for automated medical history taking and triage in health care?
2. Whose perspective (researcher, health care professional, or patient) of the AI systems is described?
3. At which stage of the innovation development process are the AI systems studied?
4. What facilitating factors and barriers are considered in relation to the introduction of the AI systems?

Methods

Study Design

The scoping review was designed according to the framework of Arksey and O'Malley [14], which has been widely used to explore, summarize, and draw conclusions about the overall state of research on new technology and health care (refer to the studies by Sharma et al [15], Gama et al [16], and Wikström et al [17]). The characteristics of included research publications (RQ1) and whose perspective is described in the publications (RQ2) are questions commonly addressed in scoping reviews. To investigate at which stages of the innovation development process the AI systems in the research publications were studied (RQ3), we adopted the Technology Readiness Level (TRL) scale for clinical readiness by Fleuren et al [7]. The scale is based on the original TRL scale developed by the National Aeronautics and Space Administration to evaluate technological maturity. It measures the clinical applicability of AI (ie, machine learning) systems on a scale of 1 to 9 (1 being the least mature and 9 being the most mature) [7,18]. To identify and report patterns in facilitating factors and barriers to the introduction of AI systems (RQ4), we used techniques from thematic analysis (ie, we coded the data and generated themes) [19]. The scoping review was reported according to the PRISMA-ScR (Preferred

XSL•FO
RenderX

Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines [20] (Multimedia Appendix 1), and a scoping review protocol was established before conducting the review.

Search Strategy

To cover any literature relevant to the scope of the review, we used 4 key terms: *health care*, *AI systems*, *medical history-taking*, and *triage* (Textbox 1). To systematically capture literature related to the key terms and synonyms, they were searched using both index terms (ie, subject headings) and free text, as outlined in Multimedia Appendix 2. To generate an extensive coverage of relevant studies, as recommended in scoping reviews [14], 5 major databases for health care–related research were searched: PubMed, CINAHL, PsycINFO, Scopus, and Web of Science Core Collection. The search was conducted in October 2022. All the databases were searched for studies published between January 2000 and September 2022. The search strings were first piloted in one of the databases (PubMed) and then further developed with the support of a research librarian before systematically searching all 5 databases. Textbox 2 provides an illustration of one of the search strings.

Textbox 1. Key terms used in search and their definitions.

- Healthcare: "...all the organizations, institutions and resources that are devoted to producing health actions" [21]
- Artificial intelligence (AI) systems: "...a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." [22]
- Medical history-taking: The process of acquiring information from a patient on medical conditions and past treatment [23,24]
- Triage: The process upon which the optimal prioritization and assessment of the appropriate next step for the patient can be decided [6]

Textbox 2. Search string used for the PubMed database (publications were filtered for publication date from January 1, 2000, to September 30, 2022).

((“neural networks, computer”[MeSH Terms:noexp] OR “artificial intelligence”[MeSH Terms:noexp] OR “deep learning”[MeSH Terms] OR “supervised machine learning”[MeSH Terms] OR “artificial intelligence”[Title/Abstract] OR “deep learning”[Title/Abstract] OR “supervised machine learning”[Title/Abstract]) AND (“delivery of health care”[MeSH Terms] OR “health care”[Title/Abstract] OR “healthcare”[Title/Abstract] OR “medical”[Title/Abstract] OR “clinical”[Title/Abstract])) AND (“Medical History Taking”[MeSH Terms] OR “anamnes*”[All Fields] OR “Triage”[MeSH Terms])

Eligibility Criteria

The eligibility criteria were set to identify research publications that presented empirical studies or systematically conducted reviews of empirical studies that focused on AI systems for automating medical history taking and triage in health care. A retrieved study was eligible for inclusion if it (1) contained a title and abstract; (2) was written in English; (3) was published between January 1, 2000, and September 30, 2022; (4) was peer reviewed; (5) analyzed empirical data or systematically reviewed empirical data; and focused on (6) health care, (7) AI, and (8) medical history taking or triage (Textbox 3). Before using the eligibility criteria for study selection in all databases, we piloted the criteria for the studies retrieved from one of the databases (PubMed). The criteria were not altered after piloting.

XSL•FO
RenderX

Textbox 3. Eligibility criteria.

Inclusion criteria
Contained title and abstract: Yes
English language: Yes
Publication date: January 1, 2000, to September 30, 2022
Peer reviewed: Yes
Empirical study: Analyzed empirical data or systematically reviewed articles with empirical data
Focused on health care: Yes
Focused on artificial intelligence: Yes
Focused on medical history taking or triage: Yes
Exclusion criteria
Contained title and abstract: No
English language: No
Publication date: Before December 31, 1999
Peer reviewed: No
Empirical study: Analyzed based on theoretical or hypothetical reasoning
Focused on health care: No
Focused on artificial intelligence: No
Focused on medical history taking or triage: No

Study Selection

First, studies identified through the search strategy were imported into Endnote software (version 20; Clarivate) and exported to the Rayyan web application (Rayyan Systems, Inc) [25] for controlled removal of duplicates. After the removal of duplicates, 2 of the authors (ES and HJ) independently screened the retrieved titles and abstracts for checking their eligibility for inclusion. When screening studies for inclusion, the authors checked the retrieved studies for inclusion according to the inclusion criteria in chronological order, that is, if criterion 1 was met, the authors checked if the retrieved study met inclusion criterion 2, and so on. Conference proceedings were excluded because they are often preliminary and based on limited analyses [26]. Editorials, commentaries, viewpoints, discussion and opinion papers, and book chapters were excluded because they often do not undergo a rigorous and impartial peer-review process. Narrative and integrative reviews were excluded because they did not systematically analyze or review empirical data. Any disagreements during the screening of titles and abstracts were discussed with a third author (JN) until a consensus was reached. Full texts of the studies that met the inclusion criteria after screening titles and abstracts were assessed for eligibility. ES and HJ independently screened the full texts of the studies. Disagreements concerning the full-text screening were discussed among all the authors until a consensus was reached.

Data Extraction

A data extraction template was developed by one of the authors (ES) based on the RQs. The template was piloted with 5 studies before extracting the data from all included research publications. The template was not altered after piloting. To extract data relevant to the RQs, the template contained columns for information on the characteristics of the research publications (RQ1; authors, publication year, country, clinical setting, sample population, study aim, study design, type of AI system, and main task performed by AI system), the perspective described in the research publications (RQ2; ie, the researchers', health care professionals', or patients' perspectives), the TRLs (1-9) of the AI systems in the research (RQ3), and facilitators and barriers to the introduction of the AI systems (RQ4). To identify the type of AI system and the main task performed by the AI system (RQ1), we adopted the operational terms presented in the Organisation for Economic Co-operation and Development Framework for the Classification of AI systems [27].

We extracted data in relation to RQs 1 to 3 from all included studies, whereas data in relation to RQ4 were extracted only from the studies deemed to have TRLs equal to 6 or higher (ie, they reported a demonstration of an AI system in a real health care environment) and for literature reviews. In total, 2 (3%) of the 77 studies with TRLs ≥6 [28,29] and 2 (33%) of the 6 literature reviews did not report the introduction of AI systems in real health care environments [30,31]. Consequently, data from these studies did not address RQ4.

Data on the characteristics of the research publications (RQ1) were extracted from the entire article; data on the perspective described in the research publications (RQ2) and the TRLs (RQ3) were extracted from the methods, results, and discussion sections of the articles. Finally, data on facilitators and barriers to the introduction of the AI systems (RQ4) were extracted from the results, discussion, and conclusions sections of the research publications.

XSL•FO
RenderX

Data Analysis

Following the framework proposed by Arksey and O'Malley [14], we began the analysis with a basic numerical analysis of the extracted data and produced tables to report the distribution of the characteristics of the studies (RQ1), the perspective described in the studies (RQ2), and their TRLs (RQ3). Thereafter, to provide a narrative account of the facilitating factors and barriers to the introduction of the AI systems (RQ4), we coded and organized the extracted data thematically [19] This involved an inductive iterative process in which we coded the extracted text on the facilitating factors and barriers and collated the codes into potential themes, which we then revised and refined. ES and HJ independently coded the data, compiled all the codes, and developed the final themes. The thematic analysis was finalized through a group discussion among all 3 authors.

Results

Overview

A total of 2008 records were retrieved from the electronic databases. After removing duplicates (n=760, 37.85%), 1248 records remained and were included in the screening. Of these, 133 (10.66%) records were assessed for eligibility in full text, and 86 (64.7%) were deemed eligible for inclusion. The PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram gives the details on the identification and screening of studies (Figure 1) [32].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. AI: artificial intelligence.

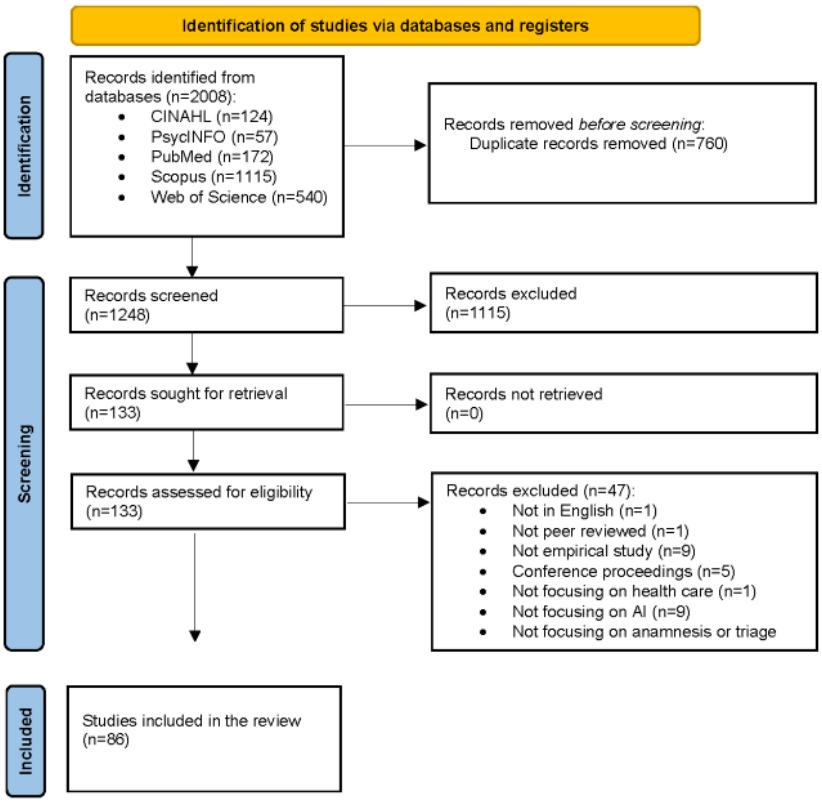

RQ1: Characteristics of the Research Publications

The included studies were published between 2000 and 2022, but most (63/86, 73%) of the studies were published between 2020 and 2022. The remaining studies (16/86, 19%) were published predominantly between 2018 and 2019. Only a small number of studies (7/86, 8%) were published between 2000 and 2017. Most of the studies were from Asia (26/86, 30%), particularly China (8/86, 9%); North America (25/86, 29%), predominantly the United States (23/86, 27%); and Europe (22/86, 26%), primarily the United Kingdom (9/86, 10%). A small number of studies (13/86, 15%) were from Africa (1/86, 1%), Oceania (5/86, 6%), and South America (4/86, 5%) or did not specify the country of origin (3/86, 3%). The most common clinical contexts in focus in the included studies were emergency care (32/86, 37%), followed by radiology (12/86, 14%) and primary care (6/86, 7%). In addition, a large number of studies (15/86, 17%) did not specify the clinical context in focus. The sample population in most studies consisted of patients (70/86, 81%), which included datasets of different types of patient records. In the remaining studies, the sample population was health care professionals (9/86, 10%), citizens (2/86, 2%), or unspecified (5/86, 6%). In these studies, the sample size was small, and there was a higher degree of involvement of the participants in the study. Regarding study design, a large number of studies were retrospective studies (31/86, 36%) or did not specify any specific design (29/86, 34%). Other study designs were, for example, prospective studies (7/86, 8%), different types of pilot studies (4/86, 5%), or reviews (4/86, 5%). The most common type of AI system was hybrid models (68/86, 79%), that is, they combined statistical and symbolic approaches. Most (40/86, 47%) of the AI systems studied were used for forecasting, that is, they predicted future outcomes based on past and present behavior or recognition (36/86, 42%), which

XSL•FO
RenderX

involved identifying and categorizing data into specific classifications or performing image segmentation and object recognition. Definitions of the types of AI systems and their main tasks used in this review can be found in the Organisation for Economic Co-operation and Development Framework for the Classification of AI systems [27]. Multimedia Appendix 3 provides an overview of the characteristics of all the 86 included studies [28-31,33-114].

RQ2: Perspectives Described in the Publications (Among Researchers, Health Care Professionals, or Patients)

Although the sample population in a large number of the included studies consisted of patients, these studies did not portray their perspectives. In most (83/86, 97%) of the included studies, researchers' perspectives were portrayed (ie, researchers' description and interpretation of the AI system and the results of the particular study. Only 1 (1%) study reflected the patients' perspectives [87]. It described patient ratings of the usability and interface of an AI system for medical history taking in an emergency department. The same study also portrayed physicians' and 10 nurses' ratings for the same AI system. Overall, patients, physicians, and nurses were strongly positive about the AI system as a support for patient-clinician communication [87]. Another study described health care professionals' perspectives [63]. More specifically, it explored how emergency triage nurses understood, contextualized, and incorporated an AI-based decision support system into their work on triaging patients. Initially, nurses expressed apprehension that the AI system lacked the cultural and contextual understanding necessary for patient triage. However, they later discovered that the system helped them provide safe care [63]. In total, 2% (2/86) of the studies presented health care professionals' perspectives [63,87]. Additional studies were conducted using health care professionals as the sample population. However, these studies focused solely on health are professionals' decision-making performance in relation to triage with AI systems, without providing insights into their perspectives [40,44,49,59,77,80,95]. In addition, 1% (1/86) of the studies portrayed citizens' (ie, men and women aged 18-80 y) perspectives on an AI system with a conversational user interface for self-guided medical history taking [46]. The citizens completed a usability test and questionnaire concerning the usability, design, and functionality of the AI system. Overall, the citizens stated that they felt comfortable with the system; however, the authors noted that the level of technical competencies and educational levels were high among the participants, limiting the generalization to patient groups [46].

RQ3: Stages of the Innovation Development Process (TRLs)

Most (76/86, 88%) of the included studies did not demonstrate or validate AI systems for medical history taking and triage in clinical environments (ie, they were concerned with clinical problem identification; proposal of a model or solution; or model prototyping, development, or validation. These studies scored ≤5 on the TRL scale for clinical readiness [7]. A few (6/86, 7%) studies scored ≥6 on the TRL scale (ie, they described real-time testing of AI models, workflow implementation, evaluation of clinical outcomes, or model integration). In total, 5% (4/86) of the studies, which were reviews of several different empirical studies, were not considered applicable to the TRL scale [30,31,52,57]. An overview of the included studies and their TRL scores is presented in Table 1.

Table 1. Technology Readiness Level (TRL) by Fleuren et al [7] of the included studies (n=82).

TRLs	Clinical definitions	Studies, n (%)
1	Clinical problem identification	0 (0)
2	Proposal of model or solution	0 (0)
3	Model prototyping	8 (9)
4	Model development	50 (58)
5	Model validation	18 (21)
6	Real-time model testing	1 (1)
7	Workflow implementation	2 (2)
8	Clinical outcome evaluation	1 (1)
9	Model integration	2 (2)

No studies were classified as TRL 1 or 2. A small number (8/86, 9%) of studies achieved TRL 3, indicating prototyping or model development of an AI system [46,82,86,93,98,103,105,108]. Examples of studies in this category include an AI-based field triage tool [82], triage of medical referrals [105], and a self-diagnosis system for emergency dental care [93]. A substantial number (50/86, 58%) of studies achieved TRL 4, that is, they demonstrated the potential of AI systems or optimized and validated these using clinical data. Examples of these studies include triage of adult chest radiographs [36], triage using digitalized patient histories in primary care [49], and triage of patients with COVID-19 under limited health care resources [69]. Several (18/86, 21%) studies were categorized as TRL 5. This means that they validated AI models using realistic datasets other than the population used to train or test the AI model. This particular type of validation data was either retrospective [48] or prospective [40]. One study attained a TRL of 6; it investigated the testing of an AI model in real time, specifically, an algorithm for automated diabetic retinopathy screening [28]. In addition, 2% (2/86) of the studies examined AI models integrated into clinical workflows and assessed their outcomes, resulting in an assessment of TRL 7 [29,87]. The

XSL•FO
RenderX

studies had a common theme of emergency care and focused on a symptom-taking tool [87] and a rapid and laboratory-free COVID-19 triage [29], respectively. One study achieved a score of 8 on the TRL scale, that is, it evaluated the clinical outcomes of an implemented AI model. This study examined a medical history–taking system that generated differential diagnoses in an outpatient department [65]. In total, 2% (2/86) of the studies focused on the postimplementation phase of AI systems and thus achieved the TRL 9 [63,81]. One study focused on the impact of cultural embeddedness when implementing an AI system for emergency care triage [63]. Another study investigated the user demographics and levels of triage acuity of a symptom checker chatbot used within a large integrated health system [81].

RQ4: Facilitating Factors and Barriers to the Introduction of the AI Systems

From the 6 retrieved studies that reported the introduction of AI systems for automating medical history taking and triage in health care, we identified 4 themes relating to facilitating factors and barriers to their introduction [52,57,63,65,81,87]. The themes were related to technical aspects, contextual and cultural factors, end users, and evaluation (Table 2). The theme denoted as *technical aspects* pertained to the AI systems' design, integration, requirements, and performance. The theme of *contextual and cultural* factors was associated with the organization and patient populations in which the AI systems were introduced. The theme *end users* encompassed the viewpoints and experiences of patients and health care professionals with regard to AI systems. The theme *evaluation and regulation* covered various aspects of evaluating, regulating, and setting guidelines for AI systems.

Table 2. Facilitating factors and barriers to the introduction of artificial intelligence (AI) systems.

	Facilitating factors	Barriers
Technical aspects	• Flexible design of AI system [52] • Integration with existing technical systems [87]	• Requirements of AI system [52] • Insufficient performance of AI systems [63,65,87] • Lack of integration with existing technical systems [57]
Context and culture	• Understanding of a setting's organizational context, clinical workload, and cultural competence [63]	• Delays in the start-up of AI systems [57] • Highly mixed patient populations, which was a challenge for the AI system [87]
End users	• Consideration of patient's perspectives [57] • Adequate training and the possibility to understand the AI system [52,63,81] • Innovative methods for promoting use among health care professionals [57] • Communication of benefits of using the AI system [52,63]	• Lack the acceptance of diagnosis provided by the AI system [65] • Negative receptivity among health care professionals [63] • Health care professionals' lack of time [65] • Low comprehensiveness among specific patient groups [87] • Lack of eHealth literacy among patients [57,87] • Insufficient language skills among patients [87] • Low number of users [57] • Lack of opportunities for patients to learn how to use the system [87]
Evaluation	• Formulation of evaluation guidelines and regulations [57] • Initial and ingoing and evaluation [57] • User-focused evaluation [87]	• Lack of regulation and guidelines for evaluation [57] • Lack of external validation [52]

Of the 6 studies, 5 (80%) reported factors related to technical aspects of the introduction of AI systems for automating medical history taking and triage that either facilitated or hindered its introduction into clinical settings [52,57,63,65,87]. Technical facilitators included a flexible design that allows the AI system to be implemented in different geographical areas [52]. Furthermore, doctors and nurses perceived that integrating AI systems with electronic health records would enhance their usefulness [87]. Technical barriers involved the amount of information required by AI systems, which might not be possible to collect in the short time required for the triage process in emergency departments [52]. An additional issue to consider was the inadequate performance of AI systems. This was due to their lack of cultural understanding of the context and the patient population they served [63], limited accuracy, particularly when dealing with patients with complex presentations, such as older adults [65], and the absence of optimization for certain patient groups, such as those with neurological symptoms [87]. The lack of integration with established health care systems constituted an additional barrier associated with the technical attributes of the AI systems [57].

A total of 3 (50%) of the 6 studies reported facilitating barriers related to context and culture [57,63,87]. Facilitating factors included greater integration of the AI system into the clinical workflow, as nurses and physicians believed this would enhance its usefulness [87]. In addition, the knowledge of the organizational context, clinical workload and culture, skills, and recognized behavior change techniques should be considered

XSL•FO
RenderX

when implementing AI systems. Contextual and cultural barriers included delays in the start-up of AI systems due to strategic decision-making processes [57] and the use of these systems in settings with highly mixed patient populations, such as emergency departments. This posed a challenge for AI systems, as reported by physicians, ultimately impacting their usefulness [87].

Of the 6 studies, 5 (80%) reported facilitating factors and barriers in relation to end users [52,57,63,81,87]. The consideration of patients' perspectives in adapting the AI system for accessibility and usefulness as well as innovative approaches to promote use among health care professionals, such as having superusers [57], were identified as facilitating factors. Other factors that facilitated the introduction of AI systems for automating medical history taking and triage included providing adequate training for health care professionals, as this is crucial given the systems' dependence on their users, and enabling them to effectively use the system [52,63,81] while also supporting patients through familiar health delivery mechanisms [81]. In addition, effective communication regarding the potential benefits of AI systems in enhancing patient care, resource use [52], and patient outcomes [63] was found to facilitate their adoption. Notably, the significance of the latter aspect was emphasized for its widespread uptake [63]. End-user barriers to the uptake of the AI system included physicians' nonacceptance of AI-generated diagnosis [65], negative receptivity toward rapid implementation [63], and health care professionals constrained by limited time to use the AI system [87]. In addition, a lack of comprehension among certain patient groups (such as those with neurological symptoms), inadequate eHealth literacy [57,87], linguistic deficiencies in patients (specifically, in expressing complaints in a way that the AI system can understand) [87], and limited opportunities for patients to acquire the necessary skills to use AI tools, due to limited occasions for using it [57,87], were additional barriers linked to end users.

Facilitating factors and barriers related to evaluation were reported in 50% (3/6) of the studies [52,57,87]. Factors suggested to facilitate the introduction of AI systems were the formulation and implementation of evaluation guidelines and regulations to ease the development and use of AI systems [57]. Initial and ongoing assessments of AI systems and evaluations focused on users were facilitating factors for assessing the safety and effectiveness of introducing AI systems in complex environments [87]. Barriers related to evaluation included the current lack of evaluation regulations and guidelines [57] as well as the validation of AI systems [52].

Discussion

Principal Findings

The goal of this study was to analyze and compile existing empirical studies on AI systems used for automating medical history taking and triage in health care. This review shows that research on AI systems to automate medical history taking and triage in clinical settings is still at an early stage, articulating a large gap to be bridged between model development and the bedside, similar to what has been highlighted in previous research on AI in health care [7,15,16,115,116]. There is a paucity of studies demonstrating AI systems to automate medical history taking and triage in clinical settings, representing the perspectives of patients or health care professionals, and providing information on facilitators and barriers to the introduction of these AI systems. The findings on the state of research in this area should be understood in light of some of the characteristics of the included research studies, indicating that research on AI systems to automate medical history taking and triage in health care remains relatively new and underdeveloped, and is mainly conducted in a nonclinical research environment of prototyping, development, and validation. While it is crucial to establish a solid research foundation that tests and validates AI systems before their implementation in clinical settings, which could potentially impact care quality, safety, clinical outcomes, and working conditions for health care professionals, it is also essential to conduct research that puts these systems into practical use to showcase their potential [117]. However, the current level of research on AI systems for automating medical history taking and triage remains inadequate.

Characteristics of the Research Publications

The study revealed a notable surge in research on AI systems for automating medical history taking and triage over the past 3 years, indicating a growing interest in leveraging AI technologies in relation to managing patient flows in health care settings. Geographically, research was predominantly concentrated in Asia, North America, and Europe, aligning with regions known for both advanced health care infrastructure and research capabilities. The prevalence of studies focusing on emergency and primary care underscores the potential of AI systems to automate medical history taking and triage in first-line care, which globally faces challenges in meeting the needs of the people it is intended to serve [118,119]. Many (12/86, 14%) studies also focused on radiology, illustrating the potential of AI systems to assist radiologists in image triage, a finding that is not surprising, given the progress of AI in image recognition tasks [120]. However, a significant (15/86, 17%) proportion of studies did not specify the clinical context, suggesting the need for more context-specific investigations to better understand the applicability of AI systems for medical history taking and triage. A large number (31/86, 36%) of studies also had a retrospective design, which, together with the large group of non-context–specific studies, may indicate a trend toward horizontal research rather than toward the implementation of AI systems for automating medical history taking and triage in clinical settings. As previous studies have suggested, this type of research and development may be partly due to the challenges of conducting studies that actually implement AI in real clinical settings [116]. The most common type of AI system was hybrid models that performed tasks related to forecasting or recognition. These findings resonate with previous research on the ways in which machine learning has been used to predict the flow of patients through health care systems [3]. Most (70/86, 81%) studies used patient datasets as their sample population, comprising patient records that were used to test, validate, and prototype AI models. This also meant that they focused on the technical capabilities of AI systems to

XSL•FO
RenderX

automate medical history taking and triage based on a specific set of patient records, rather than their usability for health care professionals or patients.

Perspectives Described in the Publications (Among Researchers, Health Care Professionals, or Patients)

While the promise of AI systems for automating medical history taking and triage to transform health care is well recognized, this study highlights an imbalance in stakeholder perspectives in the current research. Researchers' viewpoints overwhelmingly dominated the literature. Despite providing insights into the description and interpretation of AI systems designed for medical history taking and triage, there has been limited exploration of the critical viewpoints stemming from patients and health care professionals regarding their practical application in clinical settings. Previous research has argued that the imbalance between different stakeholder groups' perspectives in research on AI in health care should be addressed to meet the distinct needs of different groups [121]. Furthermore, the importance of the acceptability of interventions in routine health care practice is widely recognized in the fields of intervention, innovation, implementation, and improvement science [117]. Therefore, it is essential to consider the views of patients and health care professionals regarding the use of AI systems to automate medical history taking and triage. This will help to ensure the successful implementation of these systems. The positive feedback from patients on usability and support in patient-clinician communication from AI systems for automating medical history taking and triage emphasizes the potential benefits of involving patients in the design and evaluation of such systems. Previous research investigating a rule-based, automated medical history–taking app [122] even deemed patient involvement in the development process of these types of tools necessary. One additional study delved into health care professionals' perspectives, focusing on concerns and adjustments when integrating AI systems for automating medical history taking and triage into the triaging processes. The study found that nurses were uneasy with the AI system's acontextual nature [63]. To increase trust in AI systems, researchers suggest that focusing solely on technical aspects, such as objectivity, efficiency, and accuracy, is insufficient. AI systems must also be aligned with the values and principles of the clinical environment in which they are used. This is crucial because people are more inclined to trust and embrace AI systems when they perceive them as aligning with the inherent value, significance, and utility of the context in which they are deployed [123].

Stages of the Innovation Development Process: TRLs

The application of the TRL framework [7] provided a structured understanding of AI systems for automating medical history taking and triage development stages. Most (76/86, 88%) studies focused on lower TRLs, which aligns with the prevalence of studies that emphasize developmental work and potential demonstrations in the broader field of applying AI in health care [15]. This disparity underscores a critical gap in our understanding. This highlights the pressing need for more comprehensive investigations into the practical implementation and validation of AI systems in real health care settings. These higher TRL evaluations are pivotal in ensuring that AI technologies not only show promise in controlled environments but also prove their effectiveness and reliability when deployed in complex and dynamic safety-critical settings in health care practice [124]. Such research can contribute significantly to building trust and confidence in AI systems within the medical domain and promoting their implementation and use in practice.

Facilitating Factors and Barriers to the Introduction of AI Systems

The introduction of AI systems for automating medical history taking and triage is a complex process influenced by several factors, both hindering and facilitating its implementation. One set of barriers described lies in technical aspects, encompassing system performance and integration. The effectiveness and seamless integration of these AI systems into existing health care workflows will significantly impact their adoption and use. In addition to the technical issues, contextual and cultural factors are also important. The organizational context within health care settings as well as the diversity of patient populations shaped adoption. Different health care institutions, stakeholder roles, and patient perspectives may have unique requirements and expectations regarding AI integration in health care [9,125]. Recognizing the importance of stakeholder and patient perspectives highlights the need for tailoring AI systems to cater to context- and situation-specific needs and concerns [9,126,127]. Regulatory and evaluative considerations add another layer of complexity to the integration process. The development of guidelines and user-focused evaluations is essential to ensure the safety, efficacy, and ethical use of AI systems in health care [128-130]. These considerations underscore the need for a well-structured and carefully monitored approach to the implementation of AI in medical history taking and triage.

Limitations

The study's strengths include a thorough review of records with rigorous screening. The search strategy covered 5 electronic databases, ensuring comprehensiveness. However, it missed the opportunity to explore gray literature, which could have revealed additional cases that may present ongoing or completed work not yet in the research literature. We decided not to include conference proceedings in this review due to their often preliminary nature and limited analysis [26]. Upon examining the conference papers identified in the search, it became evident that they represented a diverse array of formats and scopes and varied significantly in how they were quality assessed and made accessible. This heterogeneity posed a challenge for their integration into a pooled analysis alongside more uniform and rigorously peer-reviewed papers. Nevertheless, in instances where the evidence is limited or contradictory, which is often the case in research on AI, the incorporation of conference proceedings can prove advantageous for systematic reviews [26]. Therefore, it is recommended that future reviews in this domain of research consider the incorporation of high-quality conference proceedings. We did not manually search the reference lists of the included studies to add more content to the search because an as stringent as possible review of the current literature was sought. Identifying studies focusing on

XSL•FO
RenderX

AI systems can be a complex task due to the lack of consensus on their definition. To ensure clarity for the reader, we have aimed to be transparent about how we define AI systems in this review. To maintain clarity, we based our inclusion on how the original authors described their systems as AI, relying on their accuracy. Because AI definitions vary, this introduces subjectivity, potentially affecting the consistency and comparability of the included articles. We suggest that future research consider this challenge in greater depth. We used a search strategy similar to that used in previous scoping reviews to identify AI systems [15,16]. The studies reviewed in this study concentrate on medical history taking and triage, both crucial processes for managing patient flow by assessing a patient's condition and referring them to the appropriate care. However, it is important to note that this approach may have some limitations. Therefore, we recommend that future research conduct separate literature reviews for each of these patient flow management processes. The eligibility criteria for study publication dates were set to identify any early studies on the topic of this review, despite AI applications in health care being a more recent phenomenon [131]. To ensure consistency, all full-text screening decisions were confirmed in pairs and discussed at biweekly meetings. This approach encouraged open discussions to address doubts, exclusion criteria, and differing interpretations. Conflicting interpretations and uncertainties were resolved through discussion.

Conclusions

This scoping review contributes to the understanding of AI system implementation in health care by providing insights into current trends, stakeholder perspectives, innovation development stages, and influencing factors. It highlights gaps in patient perspectives and the underrepresentation of higher TRLs, indicating opportunities for further research. To realize the full potential of AI systems for automating medical history taking and triage in health care settings, it is crucial to integrate patients' voices into AI development, enhance real-world validation, and address the multifaceted challenges highlighted by end users and regulations. As health care systems evolve, embracing AI technologies while ensuring alignment with clinical and patient needs remains a key challenge for future research and implementation efforts. These multifaceted factors collectively shape the landscape of AI system adoption in the context of automating medical history taking and triage. Therefore, a holistic and inclusive approach is necessary to ensure the effectiveness and ethical use of these types of AI systems.

Acknowledgments

This project has received funding from the Swedish Innovation Agency and the Knowledge Foundation.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Study conception and design were formulated by ES and JN. ES and HJ were responsible for data collection. ES, HJ, and JN contributed to the analysis and interpretation of results. All authors collaborated in the preparation of the draft and approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 530 KB - jmir_v27i1e53741_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 20 KB - jmir_v27i1e53741_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies.
[DOCX File , 194 KB - jmir_v27i1e53741_app3.docx]

References

1. Kollberg B, Dahlgaard JJ, Brehmer PO. Measuring lean initiatives in health care services: issues and findings. Int J Product Perform Manag 2006 Dec 12;56(1):7-24. [doi: 10.1108/17410400710717064]

XSL•FO
RenderX

2. Ellahham S, Ellahham N. Use of artificial intelligence for improving patient flow and healthcare delivery. J Comput Sci Syst Biol 2019;12(3):1-6 [FREE Full text]
3. El-Bouri R, Taylor T, Youssef A, Zhu T, Clifton DA. Machine learning in patient flow: a review. Prog Biomed Eng (Bristol) 2021 Apr 22;3(2):022002 [FREE Full text] [doi: 10.1088/2516-1091/abddc5] [Medline: 34738074]
4. Cecula P, Yu J, Dawoodbhoy FM, Delaney J, Tan J, Peacock I, et al. Applications of artificial intelligence to improve patient flow on mental health inpatient units - narrative literature review. Heliyon 2021 Apr;7(4):e06626 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06626] [Medline: 33898804]
5. Dawoodbhoy FM, Delaney J, Cecula P, Yu J, Peacock I, Tan J, et al. AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units. Heliyon 2021 May;7(5):e06993 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06993] [Medline: 34036191]
6. Iserson KV, Moskop JC. Triage in medicine, part I: concept, history, and types. Ann Emerg Med 2007 Mar;49(3):275-281. [doi: 10.1016/j.annemergmed.2006.05.019] [Medline: 17141139]
7. Fleuren LM, Thoral P, Shillan D, Ercole A, Elbers PW, Right Data Right Now Collaborators. Machine learning in intensive care medicine: ready for take-off? Intensive Care Med 2020 Jul;46(7):1486-1488. [doi: 10.1007/s00134-020-06045-y] [Medline: 32399747]
8. Naseri Jahfari A, Tax D, Reinders M, van der Bilt I. Machine learning for cardiovascular outcomes from wearable data: systematic review from a technology readiness level point of view. JMIR Med Inform 2022 Jan 19;10(1):e29434 [FREE Full text] [doi: 10.2196/29434] [Medline: 35044316]
9. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022 Jul 01;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
10. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
11. Ilicki J. Challenges in evaluating the accuracy of AI-containing digital triage systems: a systematic review. PLoS One 2022 Dec 27;17(12):e0279636 [FREE Full text] [doi: 10.1371/journal.pone.0279636] [Medline: 36574438]
12. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
13. Kueper JK, Terry A, Bahniwal R, Meredith L, Beleno R, Brown JB, et al. Connecting artificial intelligence and primary care challenges: findings from a multi stakeholder collaborative consultation. BMJ Health Care Inform 2022 Jan;29(1):e100493 [FREE Full text] [doi: 10.1136/bmjhci-2021-100493] [Medline: 35091423]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022 Oct 05;24(10):e40238 [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
16. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022 Jan 27;24(1):e32215 [FREE Full text] [doi: 10.2196/32215] [Medline: 35084349]
17. Wikström L, Schildmeijer K, Nylander EM, Eriksson K. Patients' and providers' perspectives on e-health applications designed for self-care in association with surgery - a scoping review. BMC Health Serv Res 2022 Mar 23;22(1):386 [FREE Full text] [doi: 10.1186/s12913-022-07718-8] [Medline: 35321707]
18. Héder M. From NASA to EU: the evolution of the TRL scale in public sector innovation. Innov J 2017 Aug;22(2):1 [FREE Full text]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. The world health report 2000: health systems: improving performance. World Health Organization. 2000 Jun 14. URL: https://www.who.int/publications/i/item/924156198X [accessed 2022-10-01]
22. Recommendation of the council on artificial intelligence. Organisation for Economic Co-Operation and Development. 2019. URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText [accessed 2023-01-01]
23. Nichol JR, Sundjaja JH, Nelson G. Medical history. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2024.
24. Svensk MeSH. Karolinska Institutet. URL: https://mesh.kib.ki.se/term/D008487/medical-history-taking [accessed 2022-10-01]
25. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
26. Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. Syst Rev 2019 Nov 07;8(1):264 [FREE Full text] [doi: 10.1186/s13643-019-1188-0] [Medline: 31699124]

XSL•FO
RenderX

27. OECD framework for the classification of AI systems. Organisation for Economic Co-Operation and Development. URL: https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html [accessed 2023-03-03]
28. Heydon P, Egan C, Bolter L, Chambers R, Anderson J, Aldington S, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. Br J Ophthalmol 2021 May;105(5):723-728 [FREE Full text] [doi: 10.1136/bjophthalmol-2020-316594] [Medline: 32606081]
29. Soltan AA, Yang J, Pattanshetty R, Novak A, Yang Y, Rohanian O, et al. Real-world evaluation of rapid and laboratory-free COVID-19 triage for emergency care: external validation and pilot deployment of artificial intelligence driven screening. Lancet Digit Health 2022 Apr;4(4):e266-e278 [FREE Full text] [doi: 10.1016/S2589-7500(21)00272-7] [Medline: 35279399]
30. Joudar SS, Albahri AS, Hamid RA. Triage and priority-based healthcare diagnosis using artificial intelligence for autism spectrum disorder and gene contribution: a systematic review. Comput Biol Med 2022 Jul;146:105553. [doi: 10.1016/j.compbiomed.2022.105553] [Medline: 35561591]
31. Sánchez-Salmerón R, Gómez-Urquiza JL, Albendín-García L, Correa-Rodríguez M, Martos-Cabrera MB, Velando-Soriano A, et al. Machine learning methods applied to triage in emergency services: a systematic review. Int Emerg Nurs 2022 Jan;60:101109. [doi: 10.1016/j.ienj.2021.101109] [Medline: 34952482]
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
33. Abe D, Inaji M, Hase T, Takahashi S, Sakai R, Ayabe F, et al. A prehospital triage system to detect traumatic intracranial hemorrhage using machine learning algorithms. JAMA Netw Open 2022 Jun 01;5(6):e2216393 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.16393] [Medline: 35687335]
34. Ahmad R, Xie L, Pyle M, Suarez MF, Broger T, Steinberg D, et al. A rapid triage test for active pulmonary tuberculosis in adult patients with persistent cough. Sci Transl Med 2019 Oct 23;11(515):eaaw8287. [doi: 10.1126/scitranslmed.aaw8287] [Medline: 31645455]
35. Ahmed MM, Sayed AM, Khafagy GM, El Sayed IT, Elkholy YS, Fares AH, et al. Accuracy of the traditional COVID-19 phone triaging system and phone triage-driven deep learning model. J Prim Care Community Health 2022;13:21501319221113544 [FREE Full text] [doi: 10.1177/21501319221113544] [Medline: 35869692]
36. Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated triaging of adult chest radiographs with deep artificial neural networks. Radiology 2019 Apr;291(1):196-202 [FREE Full text] [doi: 10.1148/radiol.2018180921] [Medline: 30667333]
37. Ayling RM, Wong A, Cotter F. Use of ColonFlag score for prioritisation of endoscopy in colorectal cancer. BMJ Open Gastroenterol 2021 Jun 03;8(1):e000639 [FREE Full text] [doi: 10.1136/bmjgast-2021-000639] [Medline: 34083226]
38. Azeez D, Ali MA, Gan KB, Saiboon I. Comparison of adaptive neuro-fuzzy inference system and artificial neutral networks model to categorize patients in the emergency department. Springerplus 2013;2:416 [FREE Full text] [doi: 10.1186/2193-1801-2-416] [Medline: 24052927]
39. Azeez D, Gan KB, Mohd Ali MA, Ismail MS. Secondary triage classification using an ensemble random forest technique. Technol Health Care 2015;23(4):419-428. [doi: 10.3233/THC-150907] [Medline: 25791174]
40. Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, Sangar D, et al. A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. Front Artif Intell 2020;3:543405 [FREE Full text] [doi: 10.3389/frai.2020.543405] [Medline: 33733203]
41. Chang YH, Shih HM, Wu JE, Huang FW, Chen WK, Chen DM, et al. Machine learning-based triage to identify low-severity patients with a short discharge length of stay in emergency department. BMC Emerg Med 2022 May 20;22(1):88 [FREE Full text] [doi: 10.1186/s12873-022-00632-6] [Medline: 35596154]
42. Choi SW, Ko T, Hong KJ, Kim KH. Machine learning-based prediction of Korean triage and acuity scale level in emergency department patients. Healthc Inform Res 2019 Oct;25(4):305-312 [FREE Full text] [doi: 10.4258/hir.2019.25.4.305] [Medline: 31777674]
43. Dehghani Soufi M, Samad-Soltani T, Shams Vahdati S, Rezaei-Hachesu P. Decision support system for triage management: a hybrid approach using rule-based reasoning and fuzzy logic. Int J Med Inform 2018 Jun;114:35-44. [doi: 10.1016/j.ijmedinf.2018.03.008] [Medline: 29673601]
44. Delshad S, Dontaraju VS, Chengat V. Artificial intelligence-based application provides accurate medical triage advice when compared to consensus decisions of healthcare providers. Cureus 2021 Aug;13(8):e16956 [FREE Full text] [doi: 10.7759/cureus.16956] [Medline: 34405077]
45. Dembrower K, Wåhlin E, Liu Y, Salim M, Smith K, Lindholm P, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit Health 2020 Sep;2(9):e468-e474 [FREE Full text] [doi: 10.1016/S2589-7500(20)30185-0] [Medline: 33328114]
46. Denecke K, Hochreutener SL, Pöpel A, May R. Self-anamnesis with a conversational user interface: concept and usability study. Methods Inf Med 2018 Nov;57(5-06):243-252. [doi: 10.1055/s-0038-1675822] [Medline: 30875703]
47. Duceau B, Alsac JM, Bellenfant F, Mailloux A, Champigneulle B, Favé G, et al. Prehospital triage of acute aortic syndrome using a machine learning algorithm. Br J Surg 2020 Jul;107(8):995-1003. [doi: 10.1002/bjs.11442] [Medline: 32043569]

XSL•FO
RenderX

48. Dyer T, Chawda S, Alkilani R, Morgan TN, Hughes M, Rasalingham S. Validation of an artificial intelligence solution for acute triage and rule-out normal of non-contrast CT head scans. Neuroradiology 2022 Apr;64(4):735-743. [doi: 10.1007/s00234-021-02826-4] [Medline: 34623478]
49. Entezarjou A, Bonamy AK, Benjaminsson S, Herman P, Midlöv P. Human- versus machine learning-based triage using digitalized patient histories in primary care: comparative study. JMIR Med Inform 2020 Sep 03;8(9):e18930 [FREE Full text] [doi: 10.2196/18930] [Medline: 32880578]
50. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. PLoS One 2020 Apr 2;15(4):e0230876 [FREE Full text] [doi: 10.1371/journal.pone.0230876] [Medline: 32240233]
51. Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A, et al. Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing. PLoS One 2020 Mar 3;15(3):e0229331 [FREE Full text] [doi: 10.1371/journal.pone.0229331] [Medline: 32126097]
52. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
53. Gao Z, Qi X, Zhang X, Gao X, He X, Guo S, et al. Developing and validating an emergency triage model using machine learning algorithms with medical big data. Risk Manag Healthc Policy 2022;15:1545-1551 [FREE Full text] [doi: 10.2147/RMHP.S355176] [Medline: 36017058]
54. Giavina-Bianchi M, Cordioli E, Dos Santos AP. Accuracy of deep neural network in triaging common skin diseases of primary care attention. Front Med (Lausanne) 2021;8:670300 [FREE Full text] [doi: 10.3389/fmed.2021.670300] [Medline: 34513863]
55. Goncharov M, Pisov M, Shevtsov A, Shirokikh B, Kurmukov A, Blokhin I, et al. CT-based COVID-19 triage: deep multitask learning improves joint identification and severity quantification. Med Image Anal 2021 Jul;71:102054 [FREE Full text] [doi: 10.1016/j.media.2021.102054] [Medline: 33932751]
56. Goto T, Camargo CAJ, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. JAMA Netw Open 2019 Jan 04;2(1):e186937 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.6937] [Medline: 30646206]
57. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
58. Gross DP, Steenstra IA, Shaw W, Yousefi P, Bellinger C, Zaïane O. Validity of the work assessment triage tool for selecting rehabilitation interventions for workers' compensation claimants with musculoskeletal conditions. J Occup Rehabil 2020 Sep;30(3):318-330. [doi: 10.1007/s10926-019-09843-4] [Medline: 31267266]
59. Harada Y, Katsukura S, Kawamura R, Shimizu T. Efficacy of artificial-intelligence-driven differential-diagnosis list on the diagnostic accuracy of physicians: an open-label randomized controlled study. Int J Environ Res Public Health 2021 Feb 21;18(4):2086 [FREE Full text] [doi: 10.3390/ijerph18042086] [Medline: 33669930]
60. Hwang S, Lee B. Machine learning-based prediction of critical illness in children visiting the emergency department. PLoS One 2022 Feb 17;17(2):e0264184 [FREE Full text] [doi: 10.1371/journal.pone.0264184] [Medline: 35176113]
61. Ivanov O, Wolf L, Brecher D, Lewis E, Masek K, Montgomery K, et al. Improving ED emergency severity index acuity assignment using machine learning and clinical natural language processing. J Emerg Nurs 2021 Mar;47(2):265-78.e7 [FREE Full text] [doi: 10.1016/j.jen.2020.11.001] [Medline: 33358394]
62. Jiang H, Mao H, Lu H, Lin P, Garry W, Lu H, et al. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. Int J Med Inform 2021 Jan;145:104326. [doi: 10.1016/j.ijmedinf.2020.104326] [Medline: 33197878]
63. Jordan M, Hauser J, Cota S, Li H, Wolf L. The impact of cultural embeddedness on the implementation of an artificial intelligence program at triage: a qualitative study. J Transcult Nurs 2023 Jan;34(1):32-39. [doi: 10.1177/10436596221129226] [Medline: 36214065]
64. Joseph JW, Leventhal EL, Grossestreuer AV, Wong ML, Joseph LJ, Nathanson LA, et al. Deep-learning approaches to identify critically ill patients at emergency department triage using limited information. J Am Coll Emerg Physicians Open 2020 Oct;1(5):773-781 [FREE Full text] [doi: 10.1002/emp2.12218] [Medline: 33145518]
65. Kawamura R, Harada Y, Sugimoto S, Nagase Y, Katsukura S, Shimizu T. Incidence of diagnostic errors among unexpectedly hospitalized patients using an automated medical history-taking system with a differential diagnosis generator: retrospective observational study. JMIR Med Inform 2022 Jan 27;10(1):e35225 [FREE Full text] [doi: 10.2196/35225] [Medline: 35084347]
66. Kerr E, McGinnity TM, Coleman S, Shepherd A. Human vital sign determination using tactile sensing and fuzzy triage system. Expert Syst Appl 2021 Aug;175:114781. [doi: 10.1016/j.eswa.2021.114781]
67. Kim CK, Choi JW, Jiao Z, Wang D, Wu J, Yi TY, et al. An automated COVID-19 triage pipeline using artificial intelligence based on chest radiographs and clinical data. NPJ Digit Med 2022 Jan 14;5(1):5 [FREE Full text] [doi: 10.1038/s41746-021-00546-w] [Medline: 35031687]

XSL•FO
RenderX

68. Kim D, You S, So S, Lee J, Yook S, Jang DP, et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. PLoS One 2018;13(10):e0206006 [FREE Full text] [doi: 10.1371/journal.pone.0206006] [Medline: 30352077]
69. Kim J, Lim H, Ahn JH, Lee KH, Lee KS, Koo KC. Optimal triage for COVID-19 patients under limited health care resources with a parsimonious machine learning prediction model and threshold optimization using discrete-event simulation: development study. JMIR Med Inform 2021 Nov 02;9(11):e32726 [FREE Full text] [doi: 10.2196/32726] [Medline: 34609319]
70. Knitza J, Janousek L, Kluge F, von der Decken CB, Kleinert S, Vorbrüggen W, et al. Machine learning-based improvement of an online rheumatology referral and triage system. Front Med (Lausanne) 2022;9:954056 [FREE Full text] [doi: 10.3389/fmed.2022.954056] [Medline: 35935756]
71. Kwon JM, Lee Y, Lee Y, Lee S, Park H, Park J. Validation of deep-learning-based triage and acuity score using a large national dataset. PLoS One 2018;13(10):e0205836 [FREE Full text] [doi: 10.1371/journal.pone.0205836] [Medline: 30321231]
72. Kyono T, Gilbert FJ, Schaar MV. Triage of 2D mammographic images using multi-view multi-task convolutional neural networks. ACM Trans Comput Healthcare 2021 Jul 15;2(3):1-24 [FREE Full text] [doi: 10.1145/3453166]
73. Larsson A, Berg J, Gellerfors M, Gerdin Wärnberg M. The advanced machine learner XGBoost did not reduce prehospital trauma mistriage compared with logistic regression: a simulation study. BMC Med Inform Decis Mak 2021 Jun 21;21(1):192 [FREE Full text] [doi: 10.1186/s12911-021-01558-y] [Medline: 34148560]
74. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Ann Emerg Med 2018 May;71(5):565-74.e2. [doi: 10.1016/j.annemergmed.2017.08.005] [Medline: 28888332]
75. Libório MP, Kritski A, Almeida IN, Miranda PF, Mesquita JR, Mota RM, et al. Impact of a computer system as a triage tool in the management of pulmonary tuberculosis in a HIV reference center in Brazil. Rev Soc Bras Med Trop 2022;55:e0451 [FREE Full text] [doi: 10.1590/0037-8682-0451-20] [Medline: 35946632]
76. Liu Y, Gao J, Liu J, Walline JH, Liu X, Zhang T, et al. Development and validation of a practical machine-learning triage algorithm for the detection of patients in need of critical care in the emergency department. Sci Rep 2021 Dec 15;11(1):24044 [FREE Full text] [doi: 10.1038/s41598-021-03104-2] [Medline: 34911945]
77. Love SM, Berg WA, Podilchuk C, López Aldrete AL, Gaxiola Mascareño AP, Pathicherikollamparambil K, et al. Palpable breast lump triage by minimally trained operators in Mexico using computer-assisted diagnosis and low-cost ultrasound. J Glob Oncol 2018 Dec;4:1-9. [doi: 10.1200/jgo.17.00222]
78. Lowres N, Duckworth A, Redfern J, Thiagalingam A, Chow CK. Use of a machine learning program to correctly triage incoming text messaging replies from a cardiovascular text-based secondary prevention program: feasibility study. JMIR Mhealth Uhealth 2020 Jun 16;8(6):e19200 [FREE Full text] [doi: 10.2196/19200] [Medline: 32543439]
79. Majidian M, Tejani I, Jarmain T, Kellett L, Moy R. Artificial intelligence in the evaluation of telemedicine dermatology patients. J Drugs Dermatol 2022 Feb 01;21(2):191-194. [doi: 10.36849/jdd.6277] [Medline: 35133107]
80. Morrill J, Qirko K, Kelly J, Ambrosy A, Toro B, Smith T, et al. A machine learning methodology for identification and triage of heart failure exacerbations. J Cardiovasc Transl Res 2022 Feb;15(1):103-115 [FREE Full text] [doi: 10.1007/s12265-021-10151-7] [Medline: 34453676]
81. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
82. Nederpelt CJ, Mokhtari AK, Alser O, Tsiligkaridis T, Roberts J, Cha M, et al. Development of a field artificial intelligence triage tool: confidence in the prediction of shock, transfusion, and definitive surgical therapy in patients with truncal gunshot wounds. J Trauma Acute Care Surg 2021 Jun 01;90(6):1054-1060. [doi: 10.1097/TA.0000000000003155] [Medline: 34016929]
83. Nsengiyumva NP, Hussain H, Oxlade O, Majidulla A, Nazish A, Khan AJ, et al. Triage of persons with tuberculosis symptoms using artificial intelligence-based chest radiograph interpretation: a cost-effectiveness analysis. Open Forum Infect Dis 2021 Dec;8(12):ofab567 [FREE Full text] [doi: 10.1093/ofid/ofab567] [Medline: 34917694]
84. Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel AS, Naheyan T, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. Lancet Digit Health 2021 Sep;3(9):e543-e554 [FREE Full text] [doi: 10.1016/S2589-7500(21)00116-3] [Medline: 34446265]
85. Raita Y, Goto T, Faridi MK, Brown DF, Camargo CAJ, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019 Feb 22;23(1):64 [FREE Full text] [doi: 10.1186/s13054-019-2351-7] [Medline: 30795786]
86. Sanders NW, Mann NH3. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. Comput Biol Med 2000 Sep;30(5):287-298. [doi: 10.1016/s0010-4825(00)00013-5] [Medline: 10913774]

XSL•FO
RenderX

87. Scheder-Bieschin J, Blümke B, de Buijzer E, Cotte F, Echterdiek F, Nacsa J, et al. Improving emergency department patient-physician conversation through an artificial intelligence symptom-taking tool: mixed methods pilot observational study. JMIR Form Res 2022 Feb 07;6(2):e28199 [FREE Full text] [doi: 10.2196/28199] [Medline: 35129452]
88. Scheetz LJ, Zhang J, Kolassa JE. Using crash scene variables to predict the need for trauma center care in older persons. Res Nurs Health 2007 Aug;30(4):399-412. [doi: 10.1002/nur.20203] [Medline: 17654535]
89. Senda A, Endo A, Kinoshita T, Otomo Y. Development of practical triage methods for critical trauma patients: machine-learning algorithm for evaluating hybrid operation theatre entry of trauma patients (THETA). Eur J Trauma Emerg Surg 2022 Dec;48(6):4755-4760. [doi: 10.1007/s00068-022-02002-0] [Medline: 35616704]
90. Shazzadur Rahman AA, Langley I, Galliez R, Kritski A, Tomeny E, Squire SB. Modelling the impact of chest X-ray and alternative triage approaches prior to seeking a tuberculosis diagnosis. BMC Infect Dis 2019 Jan 28;19(1):93 [FREE Full text] [doi: 10.1186/s12879-019-3684-1] [Medline: 30691448]
91. Shiraz A, Egawa N, Pelt DM, Crawford R, Nicholas AK, Romashova V, et al. Cervical cell lift: a novel triage method for the spatial mapping and grading of precancerous cervical lesions. EBioMedicine 2022 Aug;82:104157 [FREE Full text] [doi: 10.1016/j.ebiom.2022.104157] [Medline: 35863292]
92. Singh VK, Shrivastava U, Bouayad L, Padmanabhan B, Ialynytchev A, Schultz SK. Machine learning for psychiatric patient triaging: an investigation of cascading classifiers. J Am Med Inform Assoc 2018 Nov 01;25(11):1481-1487 [FREE Full text] [doi: 10.1093/jamia/ocy109] [Medline: 30380082]
93. Smith K, Clark A, Dyson K, Kruger E, Lejmanoski L, Russell A, et al. Guided self diagnosis: an innovative approach to triage for emergency dental care. Aust Dent J 2006 Mar;51(1):11-15 [FREE Full text] [doi: 10.1111/j.1834-7819.2006.tb00394.x] [Medline: 16669471]
94. Spasic I, Button K. Patient triage by topic modeling of referral letters: feasibility study. JMIR Med Inform 2020 Nov 06;8(11):e21252 [FREE Full text] [doi: 10.2196/21252] [Medline: 33155985]
95. Swaminathan S, Qirko K, Smith T, Corcoran E, Wysham NG, Bazaz G, et al. A machine learning approach to triaging patients with chronic obstructive pulmonary disease. PLoS One 2017;12(11):e0188532 [FREE Full text] [doi: 10.1371/journal.pone.0188532] [Medline: 29166411]
96. Tadesse GA, Javed H, Thanh NL, Thi HD, Tan LV, Thwaites L, et al. Multi-modal diagnosis of infectious diseases in the developing world. IEEE J Biomed Health Inform 2020 Jul;24(7):2131-2141. [doi: 10.1109/JBHI.2019.2959839] [Medline: 31944967]
97. Tan Y, Bacchi S, Casson RJ, Selva D, Chan W. Triaging ophthalmology outpatient referrals with machine learning: a pilot study. Clin Exp Ophthalmol 2020 Mar;48(2):169-173. [doi: 10.1111/ceo.13666] [Medline: 31648398]
98. Tao X, Chu X, Guo B, Pan Q, Ji S, Lou W, et al. Scrutinizing high-risk patients from ASC-US cytology via a deep learning model. Cancer Cytopathol 2022 Jun;130(6):407-414 [FREE Full text] [doi: 10.1002/cncy.22560] [Medline: 35290728]
99. Tsai DJ, Tsai SH, Chiang HH, Lee CC, Chen SJ. Development and validation of an artificial intelligence electrocardiogram recommendation system in the emergency department. J Pers Med 2022 Apr 27;12(5):700 [FREE Full text] [doi: 10.3390/jpm12050700] [Medline: 35629122]
100. Vaghefi E, Yang S, Xie L, Hill S, Schmiedel O, Murphy R, et al. THEIA™ development, and testing of artificial intelligence-based primary triage of diabetic retinopathy screening images in New Zealand. Diabet Med 2021 Apr;38(4):e14386 [FREE Full text] [doi: 10.1111/dme.14386] [Medline: 32794618]
101. van de Leur RR, Blom LJ, Gavves E, Hof IE, van der Heijden JF, Clappers NC, et al. Automatic triage of 12-lead ECGs using deep convolutional neural networks. J Am Heart Assoc 2020 May 18;9(10):e015138 [FREE Full text] [doi: 10.1161/JAHA.119.015138] [Medline: 32406296]
102. Verburg E, van Gils CH, van der Velden BH, Bakker MF, Pijnappel RM, Veldhuis WB, et al. Deep learning for automated triaging of 4581 breast MRI examinations from the DENSE trial. Radiology 2022 Jan;302(1):29-36. [doi: 10.1148/radiol.2021203960] [Medline: 34609196]
103. Wang C, Feng F. ERNIE based intelligent triage system. J Intell Fuzzy Syst 2022 Aug 10;43(4):5013-5022. [doi: 10.3233/jifs-212140]
104. Wang M, Xia C, Huang L, Xu S, Qin C, Liu J, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. Lancet Digit Health 2020 Oct;2(10):e506-e515. [doi: 10.1016/s2589-7500(20)30199-0]
105. Wee CK, Zhou X, Sun R, Gururajan R, Tao X, Li Y, et al. Triaging medical referrals based on clinical prioritisation criteria using machine learning techniques. Int J Environ Res Public Health 2022 Jun 16;19(12):7384 [FREE Full text] [doi: 10.3390/ijerph19127384] [Medline: 35742633]
106. Wolff P, Ríos SA, Graña M. Setting up standards: a methodological proposal for pediatric Triage machine learning model construction based on clinical outcomes. Expert Syst Appl 2019 Dec;138:112788. [doi: 10.1016/j.eswa.2019.07.005]
107. Xie F, Ong ME, Liew JN, Tan KB, Ho AF, Nadarajan GD, et al. Development and assessment of an interpretable machine learning triage tool for estimating mortality after emergency admissions. JAMA Netw Open 2021 Aug 02;4(8):e2118467 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18467] [Medline: 34448870]

XSL•FO
RenderX

108. Xiong C, Yang M, Kozar R, Zhang L. Integrating transportation data with emergency medical service records to improve triage decision of high-risk trauma patients. J Transp Health 2021 Sep;22:101106 [FREE Full text] [doi: 10.1016/j.jth.2021.101106]
109. Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A deep learning model to triage screening mammograms: a simulation study. Radiology 2019 Oct;293(1):38-46. [doi: 10.1148/radiol.2019182908] [Medline: 31385754]
110. Yang H, Gonçalves T, Quaresma P, Vieira R, Veladas R, Pinto CS, et al. Clinical trial classification of SNS24 calls with neural networks. Future Internet 2022 Apr 26;14(5):130. [doi: 10.3390/fi14050130]
111. Yao LH, Leung KC, Tsai CL, Huang CH, Fu LC. A novel deep learning-based system for triage in the emergency department using electronic medical records: retrospective cohort study. J Med Internet Res 2021 Dec 27;23(12):e27008 [FREE Full text] [doi: 10.2196/27008] [Medline: 34958305]
112. Yu JY, Jeong GY, Jeong OS, Chang DK, Cha WC. Machine learning and initial nursing assessment-based triage system for emergency department. Healthc Inform Res 2020 Jan;26(1):13-19 [FREE Full text] [doi: 10.4258/hir.2020.26.1.13] [Medline: 32082696]
113. Zhong Q, Li Z, Wang W, Zhang L, He K. Integrated medical resource consumption stratification in hospitalized patients: an Auto Triage Management model based on accurate risk, cost and length of stay prediction. Sci China Life Sci 2022 May;65(5):988-999 [FREE Full text] [doi: 10.1007/s11427-021-1987-5] [Medline: 34632536]
114. Zmiri D, Shahar Y, Taieb-Maimon M. Classification of patients by severity grades during triage in the emergency department using data mining methods. J Eval Clin Pract 2012 Apr;18(2):378-388. [doi: 10.1111/j.1365-2753.2010.01592.x] [Medline: 21166962]
115. Svedberg P, Reed J, Nilsen P, Barlow J, Macrae C, Nygren J. Toward successful implementation of artificial intelligence in health care practice: protocol for a research program. JMIR Res Protoc 2022 Mar 09;11(3):e34920 [FREE Full text] [doi: 10.2196/34920] [Medline: 35262500]
116. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med 2021 Jul;47(7):750-760 [FREE Full text] [doi: 10.1007/s00134-021-06446-7] [Medline: 34089064]
117. Nilsen P, Reed J, Nair M, Savage C, Macrae C, Barlow J, et al. Realizing the potential of artificial intelligence in healthcare: learning from intervention, innovation, implementation and improvement sciences. Front Health Serv 2022 Sep 15;2:961475 [FREE Full text] [doi: 10.3389/frhs.2022.961475] [Medline: 36925879]
118. Hanson K, Brikci N, Erlangga D, Alebachew A, De Allegri M, Balabanova D, et al. The Lancet Global Health Commission on financing primary health care: putting people at the centre. Lancet Glob Health 2022 May;10(5):e715-e772 [FREE Full text] [doi: 10.1016/S2214-109X(22)00005-5] [Medline: 35390342]
119. Lindner G, Woitok BK. Emergency department overcrowding : analysis and strategies to manage an international phenomenon. Wien Klin Wochenschr 2021 Mar;133(5-6):229-233. [doi: 10.1007/s00508-019-01596-7] [Medline: 31932966]
120. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
121. Camaradou JC, Hogg HD. Commentary: patient perspectives on artificial intelligence; what have we learned and how should we move forward? Adv Ther 2023 Jun;40(6):2563-2572 [FREE Full text] [doi: 10.1007/s12325-023-02511-3] [Medline: 37043172]
122. Nilsson E, Sverker A, Bendtsen P, Eldh AC. A human, organization, and technology perspective on patients' experiences of a chat-based and automated medical history-taking service in primary health care: interview study among primary care patients. J Med Internet Res 2021 Oct 18;23(10):e29868 [FREE Full text] [doi: 10.2196/29868] [Medline: 34661544]
123. Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J. Implementing AI in healthcare-the relevance of trust: a scoping review. Front Health Serv 2023;3:1211150 [FREE Full text] [doi: 10.3389/frhs.2023.1211150] [Medline: 37693234]
124. Cummings ML. Rethinking the maturity of artificial intelligence in safety - critical settings. AI Mag 2021 Mar;42(1):6-15 [FREE Full text] [doi: 10.1002/j.2371-9621.2021.tb00005.x]
125. Landers C, Vayena E, Amann J, Blasimme A. Stuck in translation: stakeholder perspectives on impediments to responsible digital health. Front Digit Health 2023 Feb 6;5:1069410 [FREE Full text] [doi: 10.3389/fdgth.2023.1069410] [Medline: 36815171]
126. Bergquist M, Rolandsson B, Gryska E, Laesser M, Hoefling N, Heckemann R, et al. Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. Eur Radiol 2024 Jan;34(1):338-347 [FREE Full text] [doi: 10.1007/s00330-023-09967-5] [Medline: 37505245]
127. Neher M, Petersson L, Nygren JM, Svedberg P, Larsson I, Nilsen P. Innovation in healthcare: leadership perceptions about the innovation characteristics of artificial intelligence-a qualitative interview study with healthcare leaders in Sweden. Implement Sci Commun 2023 Jul 18;4(1):81 [FREE Full text] [doi: 10.1186/s43058-023-00458-8] [Medline: 37464420]
128. Katirai A. The ethics of advancing artificial intelligence in healthcare: analyzing ethical considerations for Japan's innovative AI hospital system. Front Public Health 2023;11:1142062 [FREE Full text] [doi: 10.3389/fpubh.2023.1142062] [Medline: 37529426]

XSL•FO
RenderX

129. Petersson L, Vincent K, Svedberg P, Nygren JM, Larsson I. Ethical perspectives on implementing AI to predict mortality risk in emergency department patients: a qualitative study. Stud Health Technol Inform 2023 May 18;302:676-677. [doi: 10.3233/SHTI230234] [Medline: 37203776]
130. Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JF. Guidelines for artificial intelligence in medicine: literature review and content analysis of frameworks. J Med Internet Res 2022 Aug 25;24(8):e36823 [FREE Full text] [doi: 10.2196/36823] [Medline: 36006692]
131. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

Abbreviations

AI: artificial intelligence
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
RQ: research question
TRL: Technology Readiness Level

Edited by A Mavragani; submitted 17.10.23; peer-reviewed by LH Yao, Y Harada, E Nilsson; comments to author 10.02.24; revised version received 15.04.24; accepted 27.12.24; published 06.02.25.

Please cite as:
Siira E, Johansson H, Nygren J
Mapping and Summarizing the Research on AI Systems for Automating Medical History Taking and Triage: Scoping Review
J Med Internet Res 2025;27:e53741
URL: https://www.jmir.org/2025/1/e53741
doi:10.2196/53741
PMID:

©Elin Siira, Hanna Johansson, Jens Nygren. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Patient-Accessible Electronic Health Records and Information Practices in Mental Health Care Contexts: Scoping Review

Timothy Kariotis[1,2], BA, BSocSc, BNutrSc, MPH, MBA, MHCM; Megan Prictor[2], LLB, PhD; Kathleen Gray[3], MLibSc, MEnvSc, PhD; Shanton Chang[1], BCom, PhD

[1]Faculty of Engineering and Information Technology, University of Melbourne, Carlton, Australia

[2]Melbourne Law School, University of Melbourne, Melbourne, Australia

[3]Centre for Digital Transformation of Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Timothy Kariotis, BA, BSocSc, BNutrSc, MPH, MBA, MHCM
Faculty of Engineering and Information Technology
University of Melbourne
700 Swanston Street
Carlton, 3053
Australia
Phone: 61 488300223
Email: Timothy.kariotis@unimelb.edu.au

Abstract

Background: Patients are increasingly being provided with access to their electronic health records. However, in mental health care contexts, concerns have been raised due to a perception that such access would pose risks to patients, third parties, and the therapeutic relationship. These perceived risks may affect the information practices of health care professionals (HCPs) and patients, such as how they document, share, and use information in mental health care services with a patient-accessible electronic health record (PAEHR). Although there is growing research interest in PAEHRs, no study has specifically examined how they impact information practices. Understanding the impacts on information practices may help explain other outcomes of implementing PAEHRs and inform their design.

Objective: This scoping review aimed to explore the research on PAEHRs in mental health care contexts and how PAEHRs affect information practices of HCPs and patients in this context.

Methods: A scoping review was considered the most appropriate method due to the relatively recent adoption of PAEHRs in mental health care contexts and the heterogeneous nature of the evidence base. A comprehensive search of electronic databases was conducted for original empirical studies that described the use of PAEHRs or associated systems in mental health care contexts. One author reviewed all full texts, with 3 other authors reviewing a subset of studies. The study characteristics and findings were documented, and a thematic synthesis and textual narrative analysis were used to develop descriptive and analytical themes that addressed the research questions.

Results: A total of 66 studies were considered eligible and included in the analysis. The impact of PAEHRs on information practices in mental health care contexts included the following: (1) they may change how HCPs document patient information, including a reduction in detail and a focus on information perceived by HCPs as objective rather than subjective; (2) they may negatively impact workflows due to changes in documentation practices and limited guidance for HCPs on how to use PAEHRs; and (3) they may contribute to improved communication between HCPs and patients, including constructive disagreements regarding what is documented in the health record. The changes to HCP information practices were influenced by a concern for the therapeutic relationship and patient safety. Furthermore, PAEHRs supported new information practices for patients, such as using their PAEHR to prepare for clinical encounters.

Conclusions: We have identified several ways in which PAEHRs shape the information practices of HCPs and patients in the mental health context. These findings can inform the design of PAEHRs to promote information practices that contribute to improving the quality of mental health care. Further research is necessary to understand how changes in information practices due to the implementation of PAEHRs impact clinical outcomes and patient experiences of care.

(J Med Internet Res 2025;27:e54973) doi:10.2196/54973

XSL•FO
RenderX

KEYWORDS

patient-accessible electronic health records; patient portals; psychiatry; mental health; health informatics; mental illness; scoping review

Introduction

Background

Patients in many high-income countries, such as Australia, the United States, the United Kingdom, France, and the Netherlands, have a legal right to request access to their health records [1,2]. Traditionally, patients have had to apply for a copy of their paper-based health records, which can hinder easy access and efficient use of their health information [3]. There is a growing movement to make it easier for patients to access their health information by providing electronic access to their health records. This shift toward online access has been enabled by the widespread adoption of electronic health records (EHRs) [2,4].

Providing patients with access to their health records can improve their disease management, self-care, and understanding of their health conditions [5], while facilitating more productive discussions between health care professional (HCPs) and patients [5-7]. Mental health care contexts have been slower to adopt patient-accessible EHRs (PAEHRs), and in some cases limitations have been placed on patients with mental health conditions accessing their PAEHR [8-12]. There are various concerns regarding the use of PAEHRs in mental health care contexts, including the impact on third parties, such as relatives whose information may be documented in the record and the risk that patients will experience distress from reading their PAEHR [11,13].

The digitalization of patient health records has impacted the information practices of HCPs and patients, changing the way they seek, document, and share information [14]. We use the concept of information practices to capture how information activities, such as documentation, are situated in and shaped by specific contexts [15]. Østensen et al [16] defined information practice as "a socially constructed practice that determines how information is produced, organized, disseminated, distributed, reproduced and circulated in the community, and which specific types of information are legitimized." The concept of information practice is used in this study because it encourages consideration of factors unique to mental health care contexts, such as stigma and risk management, which may shape HCPs' and patients' information activities [17,18].

Several technologies are available to provide patients with access to parts or all of their health record, including electronic personal health records (ePHRs) and patient portals. A range of initiatives, such as OpenNotes and Blue Button, implemented predominantly in the United States and across several Nordic countries, have promoted immediate patient access to their EHRs via patient portals [19-21]. Both personal EHRs (pEHRs) and patient portals can be defined in comparison to EHRs, which are the digital clinical records held by a health service and accessible only by HCPs [22,23]. Patient portals provide patients with access to some parts of the EHR and a range of other functions, such as booking appointments [24]. In comparison, pEHRs are health records controlled by the patient, which may be linked to an EHR or stand-alone entities [25]. pEHRs may include a range of functions, such as capturing patient-generated data, such as biometric data from wearable devices, and providing access to educational material [26,27]. Beyond these simple definitions, there is substantial heterogeneity in the types, functions, architecture, and content of pEHRs and patient portals [26,28]. In this paper, because the focus is on patients' online access to their health records, we use the term PAEHRs. This term captures the range of technologies that give people complete or partial access to the information contained in their EHR. Other terms may be mentioned in this paper when referring to results from specific studies.

Several recent reviews have explored the use of PAEHRs [29-32], with 2 specifically in mental health contexts [9,24]. However, these studies have not considered the impact of PAEHRs on information practices and have usually focused solely on patient portals. Zhang et al [24] found that the factors for successful implementation and adoption of patient portals in mental health care contexts include education for patients, the perceived value of the portal by patients, how easy it was for patients to use the portal, attitudes of HCPs toward using the portal, and adequate staffing and staff training to support the use of the portal. Schwarz et al [9] found that patients reported various positive experiences of using patient portals, such as increased empowerment and trust in clinicians, while negative experiences were related to inaccuracies in their records or disrespectful language. HCPs reported more concerns than benefits, including increased documentation burden and possible harm from patients reading their records. These studies reflect the growing prevalence and interest in PAEHRs in mental health care contexts. However, research is needed to consider the impact of PAEHRs on information practices because changes to information practices, such as HCPs documenting less information, may impact the quality and safety of care provided in mental health care settings.

This paper builds on our previous scoping review where we found that EHRs promote standardized and formalized documentation practices that often conflict with the prevalent use of narrative free-text information in mental health care settings [15]. The review also found that HCPs had concerns about documenting sensitive information in EHRs due to the increased ease of access by other HCPs [15]. The previous scoping review suggests that changing what is documented, how it is documented, and who can access information in the health record may alter clinicians' information practices. Other studies have found that adopting EHRs affects patients' intention to disclose certain information [33,34]. Thus, it could be assumed that the introduction of PAEHRs may impact HCPs' and patients' information practices.

This review has the following objectives and research questions:

1. How do PAEHRs change the information practices of patients and HCPs in mental health care contexts?

XSL•FO
RenderX

2. What effects do changes in information practices have on other aspects of care, including the therapeutic relationship, risk management, information sharing, and patient experience?

A Note on Language

In this study, we used the term *patient* when referring to people accessing and using mental health services. We chose this over other terms, such as client or service user, mainly because many of the PAEHR technologies reference "patients" in their name (eg, patient portal), and we wanted to avoid causing confusion by using another term. We acknowledge that the term *patient* can be considered disempowering and that the terminology in this space is not settled.

The title of this paper refers to *mental health care contexts*, which is used to capture the broad range of clinical and nonclinical services people may access when experiencing mental health issues [35].

Methods

Overview

Scoping reviews facilitate consideration of a broad range of evidence, methods, and study types in relatively new or heterogeneous research fields [36]. Because PAEHRs in mental health care contexts are relatively new and information practices have received limited consideration in the literature, we considered a scoping review an appropriate method to address the research objectives [37,38]. We were informed by the framework for scoping reviews by Arksey and O'Malley [36] and the relevant criteria outlined in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [39]. No protocol was published for this review.

Study Selection

The inclusion and exclusion criteria are outlined in Textbox 1. We chose to include soon-to-be-implemented PAEHRs, which we defined as systems that had been announced or were planned for implementation, for 3 reasons. First, if we were to publish an updated review, it would provide us with the opportunity to compare pre- and postimplementation studies. Second, perceptions of a PAEHR may influence practices as much as using a PAEHR [40,41]. Third, in our initial reading of the literature, we noticed that often PAEHRs had low use by HCPs or had only been recently implemented. Thus, there may be little difference between perceptions in the lead-up to implementation and perceptions after implementation if the PAEHR has not been widely used.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Articles describing original empirical research
- Research conducted in a mental health care context (eg, psychiatric hospital) or with participants who have a lived experience of mental illness (eg, patients in a primary health care context with a diagnosis of depression)
- Studies published up to September 2023
- Studies focused on the implementation of a system for allowing patients with mental health conditions to access their health record or about the perception of an imminent implementation of such a system or studies that sought input from people with involvement in the design, adoption, use, or evaluations of patient-accessible electronic health records

Exclusion criteria

- Studies with no methodology
- Studies with no full text in English
- Protocol papers, commentaries, or posters
- Studies exploring health care professionals' or patients' general perceptions about access to online mental health records
- Studies where the participants are children or adolescents, or the study is conducted in the youth mental health system
- Studies conducted in forensic mental health care settings
- Scoping reviews or systematic reviews

Search Strategy

We searched Scopus, Embase, Web of Science, MEDLINE via PubMed, PsycINFO, and CINAHL using a combination of key terms summarized in Textbox 2. A search was also conducted on Google and ResearchGate. The search strategy was developed iteratively alongside further identification of key terms in the literature and hand searching reference lists of relevant studies. The initial search was undertaken in late 2018 and updated in December 2022, with new papers continually identified until September 2023, when the final draft was completed.

XSL•FO
RenderX

Textbox 2. Example search strategy used in Scopus.

(TITLE-ABS-KEY (“patient portal” OR “open*notes” OR “personal electronic record” OR “personal electronic health record” OR “personal electronic medical record” OR “personal medical record” OR “personal health record” OR “pEHR” OR “pEMR” OR “patient*controlled personal health records” OR “shared medical record” OR “patient controlled record” OR “patient*held medical record” OR “patient*controlled journal” OR “patient*controlled electronic health record” OR “blue button” OR “health record portal” OR “consumer portal” OR “patient access to health record” OR “PHR” OR “ePHR” OR “patient accessible record” OR “patient*shared record” OR “patient*carried record” OR “patient*held record” OR “patient internet portal” OR “patient accessible electronic health record” OR “patient access to record”) AND TITLE-ABS-KEY (“psychiatr*” OR “mental illness” OR “mental health” OR “mental health services” OR “behavioral health*” OR “mental healthcare” OR “mental health care” OR “mental health nurs*” OR “schizophrenia” OR “bipolar” OR “depression” OR “anxiety” OR “personality disorder” OR “psychosocial”))

Screening Process

After removing duplicates, TK screened the titles and abstracts against the inclusion criteria to identify relevant studies for full-text screening. TK reviewed all full-text articles using the inclusion and exclusion criteria. TK identified 23% (15/66) of the studies for which it was unclear whether they met the inclusion criteria. The other 3 authors (MP, KG, and SC) each reviewed one-third (5/15, 33%) of these studies to determine whether they should be included. The authors worked together to resolve disagreements and determine the final articles for inclusion in the review. There were a few disagreements regarding studies in which the use of the PAEHR was minimal, as it was part of a larger study or intervention. Furthermore, 23% (15/66) of the studies that did not clearly meet the inclusion criteria were included after a discussion between the authors.

Analysis of Included Studies

Due to the breadth of the study designs and objectives, covering a range of qualitative and quantitative methods, we adopted a textual narrative and thematic synthesis approach, as has been done in similar studies [15,42,43]. The textual analysis involved tabulating study findings alongside study characteristics and conclusions in a spreadsheet. The study characteristics we planned to document included the country where the study took place, year of publication, study design and method, participant characteristics, research focus, type of PAEHR, implementation status, and functions of the PAEHR. TK extracted these data from each paper into a spreadsheet, which was discussed with the research team to ensure its completeness.

Using the thematic synthesis approach described by Thomas and Harden [44], we developed descriptive themes by coding both direct participant quotes and researcher interpretations from the qualitative studies into a mind map. Our research questions framed this coding process, in that we coded findings related to how HCPs and patients used the PAEHR to manage information, how the PAEHR changed how HCPs and patients managed information, and findings related to issues raised in the Introduction section of this paper. Related codes were clustered together to develop descriptive themes. We intended these descriptive themes to stay as close to the original findings as possible. We integrated the quantitative data extracted during the textual synthesis into the descriptive themes. Finally, using our review questions, we developed analytical themes that sought to consider the impact of the PAEHRs on the information practices of HCPs and patients. This process is necessarily subjective and relied on the researchers’ judgment and insights [44,45]. To ensure the validity of the themes, TK undertook the initial descriptive analysis before seeking feedback from the research team to ensure the themes aligned with their understanding of the included studies and research questions. Similarly, TK developed the analytical themes before seeking feedback from the research team.

Results

Overview

The search strategy identified 1713 studies. Another 24 studies were identified through hand searching and other sources, including Google and ResearchGate. After removing duplicates, of the 1737 studies, 54.92% (n=954) were identified for screening. After screening the titles and abstracts, of the 954 studies, 11.2% (n=107) met the inclusion criteria. Of the 107 articles reviewed in the full text, 38.3% (n=41) were excluded, and 61.7% (n=66) were included (Figure 1). Details of the included studies are outlined in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. PAEHR: patient-accessible electronic health record.

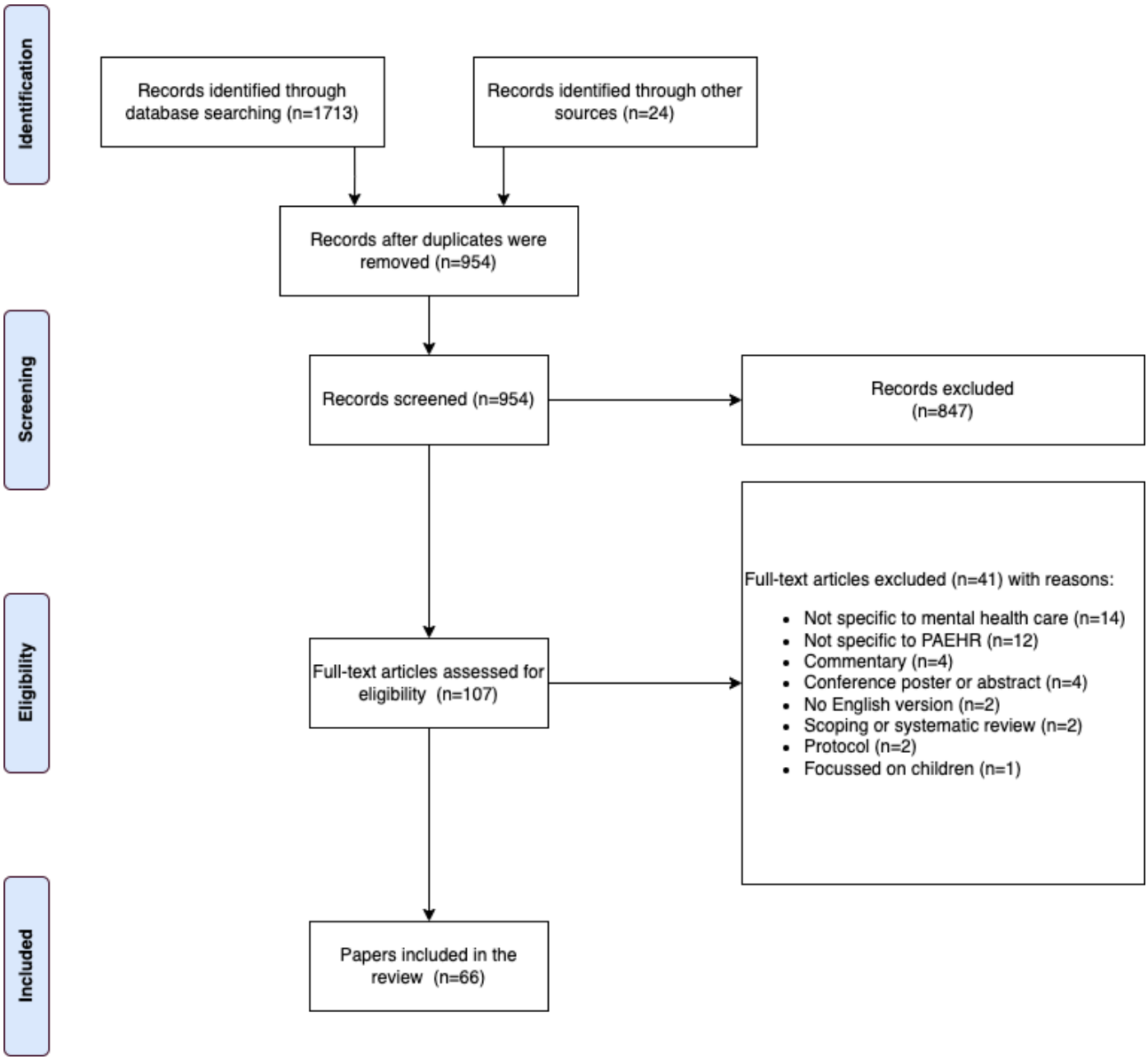

Study Characteristics

The following sections report on the characteristics of the included studies that were identified in the textual synthesis.

Country

The top 3 countries where the studies were conducted were the United States (29/66, 44%), Canada (13/66, 20%), and Sweden (9/66, 14%). Table 1 provides a more detailed breakdown of the countries where the included studies were conducted. The main difference across countries was whether there was a national patient portal system, such as in Sweden, compared to a patient portal implemented across a specific service or set of services, such as the US Veterans Affairs My HealtheVet.

XSL•FO
RenderX

Table 1. Countries where the studies were conducted (N=66).

Countries	Studies, n (%)
United States	29 (44)
Canada	13 (20)
Sweden	9 (14)
Norway	4 (6)
The Netherlands	2 (3)
United Kingdom	3 (5)
Germany	2 (3)
New Zealand	1 (2)
Multiple countries	3 (5)

Year of Publication

The included studies were published in the period from 2009 to 2023 (the cutoff year for inclusion). The highest number of studies was published in 2018 (11/66, 17%), followed by 2019 (10/66, 15%) and 2022 (10/66, 15%). There is an overall upward trend of studies published on PAEHRs in mental health care contexts since 2009, with some fluctuations in recent years, as illustrated in Figure 2. Due to the varied level of detail reported in the included studies on the functions and technical features of the PAEHRs, we did not evaluate how PAEHRs have changed over time.

Figure 2. Increasing trend in publications on patient-accessible electronic health records in mental health care contexts.

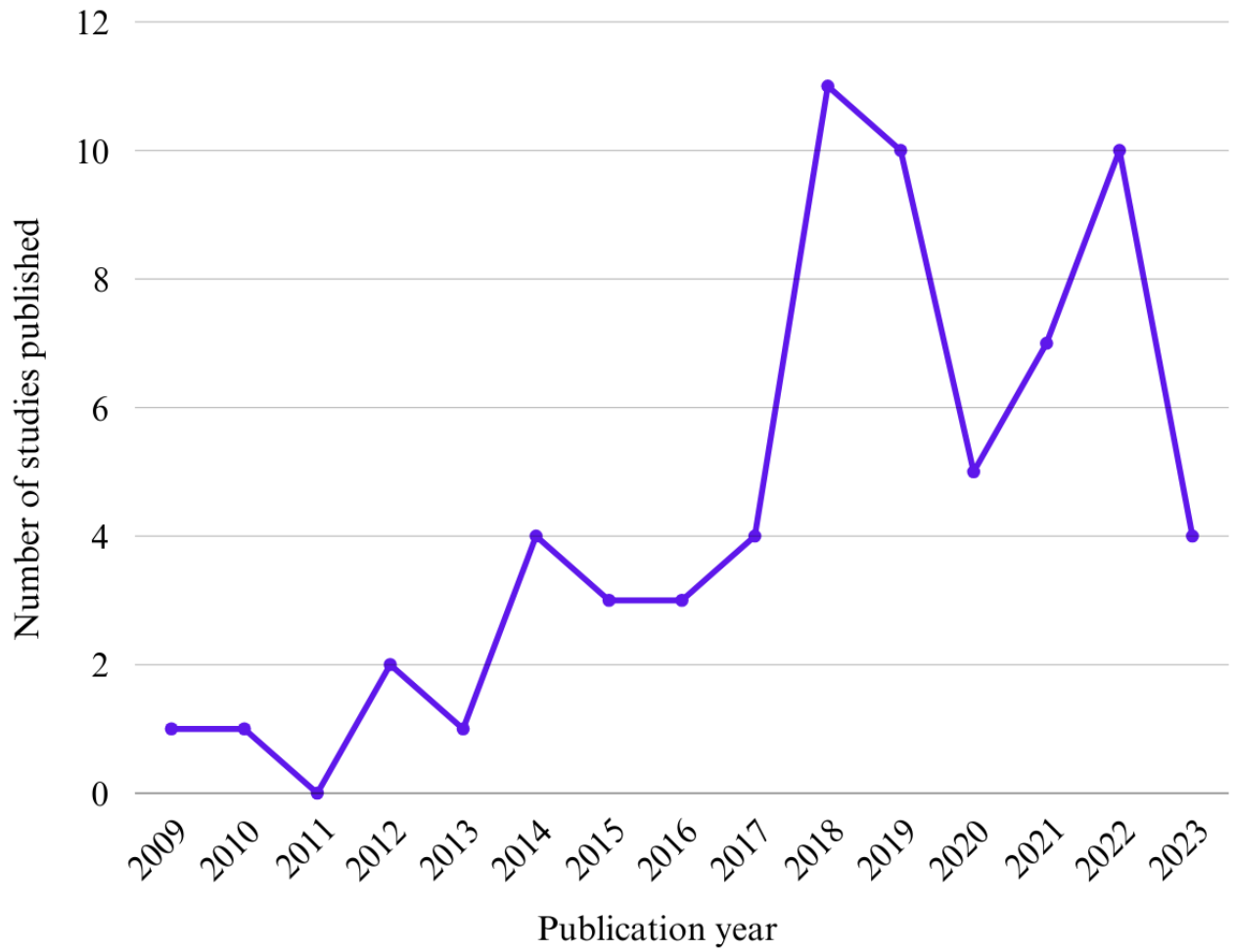

Study Design and Method

There was a mix of quantitative (29/66, 44%), qualitative (22/66, 33%), and mixed methods (15/66, 23%) approaches used in the included studies (Table 2). The main methods used were surveys (27/66, 41%), interviews (18/66, 27%), and interventions or trials involving a PAEHR (13/66, 20%).

XSL•FO
RenderX

Table 2. Study design and method (N=66).

Designs and methods	Studies, n (%)
Study design	
Quantitative	29 (44)
Qualitative	22 (33)
Mixed methods	15 (23)
Research methods[a]	
Survey	27 (41)
Interview	18 (27)
Focus group	7 (11)
Interventions or trials involving a PAEHR[b] (eg, randomized controlled trials)[c]	13 (20)
Secondary data analysis	7 (11)
Design	2 (3)
Observation	1 (2)
Document analysis	1 (2)

[a]Some studies included multiple methods and thus were counted twice.

[b]PAEHR: patient-accessible electronic health record.

[c]We did not report the specific methods of intervention studies; instead, we just categorized them as interventions.

Participants

Most studies (37/66, 56%) included only patients as participants, while 27% (18/66) of the studies included only HCPs, and 15% (10/66) of the studies included both patients and HCPs. Table 3 provides a more detailed breakdown of participants. Nurses were the most represented HCPs, followed by social workers, psychologists, psychiatrists, and physicians. Family members, carers, and peer workers were rarely involved as participants in the included studies.

Table 3. Participants in the included studies (N=66).

Participants	Studies, n (%)
Study sample (high level)	
HCPs[a]	18 (27)
Patients	37 (56)
Both HCPs and patients	10 (15)
Other or not specified	1 (2)
Study sample (detailed)	
Patients	48 (73)
Family and carers	3 (5)
Peer workers	3 (5)
Social workers	17 (26)
Psychiatrists	14 (21)
Psychologists	16 (24)
Mental health nurses, nurse practitioners, nurses, and nurse assistants	19 (29)
Physicians, doctors, general practitioners, and family physicians	14 (21)
Allied health—not including psychologists (occupational therapists, counselors, and physiotherapists)	5 (8)
Administrative staff	7 (11)

[a]HCP: health care professional.

XSL•FO
RenderX

Research Focus

We compared the aims of the included studies and grouped them according to similar topic areas, as outlined in Table 4. Most studies focused on the general experiences and perceptions of HCPs (18/66, 27%) and patients (13/66, 20%) toward a PAEHR. Several studies also explored specific clinical interventions or models of care using a PAEHR (12/66, 18%) and adoption and use rates across different demographic groups (11/66, 17%).

Table 4. Research focus (N=66).

Research focus	Studies, n (%)[a]
HCPs'[b] experience and perspectives of PAEHR[c]	18 (27)
Patients' experience and perspectives of PAEHR	13 (20)
Interventions using a PAEHR	12 (18)
Adoption and use of PAEHR by different patient groups	11 (17)
Design and usability of PAEHR for patients	7 (11)
Barriers and facilitators to patients using PAEHR	4 (6)
HCP or patient perspectives of the impact of PAEHR on therapeutic relationships	5 (8)
Exploring how education or training can support the use of PAEHR by an HCP or patient	3 (5)
Describing policies related to PAEHR	2 (3)
Exploring the impact of PAEHRs on transparency in the clinical encounter	2 (3)
Perceived benefits, harms, or risks of PAEHRs	2 (3)
Other (exploring errors in PAEHR, changing HCP practices, family or carer perceptions, reasons patients do not adopt PAEHR, perspectives on the needs of patients and clinicians when using OpenNotes, and impact of PAEHR on medication adherence)	6 (9)

[a]Some studies included multiple research aims and thus were counted twice.

[b]HCP: health care professional.

[c]PAEHR: patient-accessible electronic health record.

PAEHR System Used

There were several challenges in defining the PAEHR system used in each study. Most studies (26/66, 40%) discussed patient portals or OpenNotes generally without specifying a specific product or whether the system was a patient portal of ePHR (Table 5). Some studies discussed systems that appeared to be implemented at the national or health system level, although it was difficult to discern across studies if the same system was being discussed.

Table 5. Patient-accessible electronic health record (PAEHR) systems reported in the included studies (N=66).

Type of PAEHR	Patient portal or ePHR[a] (categorized by author)	Studies, n (%)
Lawson Smart Record	ePHR	4 (6)
Veterans Affairs My HealtheVet	Patient portal	13 (20)
Patient Portal—general (including OpenNotes)	Patient portal	26 (40)
Swedish Journalen	Patient portal	8 (12)
Helsenorge.no [46-49]	Patient portal	4 (6)
ePHR—general	ePHR	6 (9)
My Health Locker	ePHR	2 (3)
EPIC MyChartPortal	Patient portal	1 (2)
My Medical Record	Patient portal	1 (2)
MyCare (Cerner patient portal)	Patient portal	1 (2)

[a]ePHR: electronic personal health record.

Implementation Status

Most studies (51/66, 77%) reported on a PAEHR that had been implemented (Table 6).

XSL•FO
RenderX

Table 6. Implementation status of patient-accessible electronic health record (PAEHR; N=66).

Implementation status of PAEHR	Studies, n (%)
Implemented PAEHR	51 (77)
Soon-to-be-implemented PAEHR	5 (8)
PAEHR being designed as part of the study	3 (5)
Not applicable (expert perspective)	2 (3)
Partial implementation, including if some participants were using or had access to the PAEHR	3 (5)
Pre-post implementation	1 (2)
Unclear	1 (2)

Functionality of PAEHR

The included studies did not provide enough information for a thorough analysis of the functions available across the different PAEHRs. Therefore, we only address specific elements of functionality related to the themes discussed in the results of the thematic synthesis discussed in the Thematic Results section.

Study Quality

Due to the heterogeneity of the included studies and the lack of evaluation instruments that could be applied systematically across these studies, a quality evaluation was not deemed appropriate for this scoping review. It is optional for scoping reviews to evaluate the quality of the included studies [50]. One quality issue we identified during the review of the included studies was a lack of detail regarding the functionality and technical specifications of the PAEHRs, which Talmon et al [51] recommend should be reported in health informatics studies. We identified a lack of reporting on functionality and technical specifications in our previous scoping review of EHRs [15].

During the peer-review process, a reviewer identified that 4 articles were published in journals that they considered predatory. We concluded that publishing in a predatory journal is not necessarily an indication of the quality of the study, and thus, we decided to include these studies in the review.

Thematic Results

Overview

In the following sections, we report on the findings of the thematic synthesis of the 66 included studies. The analysis led to the development of 8 themes and 12 subthemes. The themes are outlined in Table 7. Illustrative quotes for each theme are provided in Multimedia Appendix 3.

XSL•FO
RenderX

Table 7. Summary of themes.

Themes	Subthemes
Theme 1: patient-accessible electronic health records (PAEHRs) change the purpose of the health record	—[a]
Theme 2: health care professionals (HCPs) change their information practices to protect the therapeutic relationship and patient safety	• Subtheme 2.1: PAEHRs introduce risks to the therapeutic relationship • Subtheme 2.2: PAEHRs introduce risks to patient safety • Subtheme 2.3: trust, transparency, and accuracy of information can mitigate the risks posed by the PAEHRs to the therapeutic relationship and patient safety • Subtheme 2.4: PAEHRs offer opportunities to enhance accountability and communication between HCPs and patients regarding the documentation of information
Theme 3: PAEHRs disrupt HCPs' documentation practices	• Subtheme 3.1: HCPs document less detailed information in PAEHRs, particularly relating to sensitive information • Subtheme 3.2: PAEHRs require HCPs to document information that patients will understand • Subtheme 3.3: PAEHRs require HCPs to take a person-centered approach when documenting potentially subjective information • Subtheme 3.4: HCPs limit the documentation of uncertain information in PAEHRs • Subtheme 3.5: HCPs may omit content or restrict access to PAEHRs
Theme 4: PAEHRs introduce changes to HCPs' information workflows	—
Theme 5: HCPs require tailored training, education, and guidelines on documenting information in a PAEHR	—
Theme 6: patients are empowered with new information practices if they are supported to use their PAEHR	• Subtheme 6.1: patients need support to use their PAEHR • Subtheme 6.2: PAEHRs must be easy for patients to navigate, access, and understand
Theme 7: PAEHRs raise new concerns about information privacy and security for HCPs, patients, and third parties	• Subtheme 7.1: PAEHRs pose challenges to managing third-party data • Subtheme 7.2: PAEHRs pose challenges to managing third-party access to patient information

[a]No subthemes.

Theme 1: PAEHRs Change the Purpose of the Health Record

HCPs perceived that the introduction of a PAEHR changed the purpose of the health record, with HCPs concerned by the increasing array of audiences they had to consider when documenting information [49,52-54]. This change was framed as a culture shift [55], with one HCP in the study by Denneson et al [53] describing it as "antithetical to the way that many of us have been, literally, trained and learned to think about our field." HCPs thought the way information was documented previously, which assumed that only other HCPs would access it, was inappropriate for patients and would pose risks such as worry, offense, or distress for patients [47,49,52,56]. Conversely, there was a perception that writing only for the patients may reduce the clinical value of the EHR [57]. HCPs in several studies also perceived a shift in control over the health records due to the ability of patients to access their health records in real time via their PAEHR [49,56,57]. Although patients may have previously been able to request access to their records, clinicians described having had more control over what was released and when it was released [49,53]. HCPs also felt they had less control over what patients might do with the information in their PAEHR, such as sharing it with third parties [58].

Theme 2: HCPs Change Their Information Practices to Protect the Therapeutic Relationship and Patient Safety

Overview

HCPs were concerned that PAEHRs would damage the therapeutic relationship because patients may not understand what is documented in their PAEHR or experience distress from reading their PAEHR. Some HCPs and patients recognized that PAEHRs could enhance the therapeutic relationship through improved transparency and communication about what was documented. However, the perceived risks appeared to be significant enough to change HCPs' information practices. These perceived risks are part of a broader perception, outlined in theme 1, that the purpose of the record changes when patients have access to it via a PAEHR and that this puts into question previous practices around what is documented, for whom, and for what purpose.

XSL•FO
RenderX

Subtheme 2.1: PAEHRs Introduce Risks to the Therapeutic Relationship

HCPs across several studies reported concerns that PAEHRs would negatively impact the therapeutic relationship due to patients not understanding what was written in their record or perceiving something written in their record to be a negative judgment about them [53,58-61]. This risk shaped many of the documentation changes described in theme 3. HCPs described a risk that patients might see a disconnect between what was written in their record and their in-person clinical experience [53]. For example, what HCPs document is not always the same information they would share with patients, such as clinical formulations [53]. HCPs in the study by van Rijt et al [58] perceived a risk that patients might feel "insulted, misinterpret the information given, or feel unheard," which in turn may negatively impact a patient's trust in the HCP. In comparison, patients in the included studies focused more on the benefits of the PAEHR for the therapeutic relationship, as considered in subthemes 2.3 and 2.4.

Subtheme 2.2: PAEHRs Introduce Risks to Patient Safety

HCPs across several studies reported a concern for patient safety, specifically that patients might find the PAEHR overwhelming and distressing or be triggered by reading their records [49,53,55,57-59,62]. For example, Turvey et al [62] reported that 55% (16/29) of HCPs in their study reported having patients who experienced significant distress after reading their records, while 29% (8/29) and 21% (6/29) of HCPs reported experiences of patients terminating treatment after accessing their records or reporting having engaged in "negative and/or self-destructive behavior toward themselves or others," respectively. These results do not specify how many patients had negative experiences of reading their PAEHR. Rather these results suggest that HCPs are coming across examples of patients' having negative experiences, which is shaping their perception of PAEHRs. These concerns could be considered through the lens of HCPs' ethical need to keep patients safe, with HCPs in the study by Jonnergård et al [63] identifying "patient safety" as the most important issue to be addressed in implementing OpenNotes. In addition, there were some concerns that PAEHRs might lead to increased violent behavior by patients. However, studies that explored these concerns did not find evidence of increased violent behavior from the adoption of PAEHRs in mental health contexts [64,65].

Patients were concerned about the risk of being upset by access to their PAEHR or experiencing anxiety when thinking about what might be written in their PAEHR [57,66]. For example, 11% (4/37) of patients reported "sometimes experiencing significant distress" after reading their record in the study by Turvey et al [62]. Denneson et al [67] found in patient survey that 26.4% (47/178) of patients sometimes, and 8.4% (15/178) often or always, "experienced stress or worry after reading their record," while 18% (32/178) of patients sometimes, and 8.4% (15/178) often or always, reported "feeling upset." A common reason reported for patients feeling upset after accessing their PAEHR was because they made the patients' problems seem smaller than they were [67]. Some patients in the study by Schwarz et al [68] also reported a fear of reading their records because of how confronting it may be to revisit the experiences they had shared with their HCPs during clinical sessions. Alternatively, accessing information in the PAEHR between appointments could reduce the anxiety of waiting to see the HCP [37]. Interestingly, the 1 study that compared patients with mental health diagnoses to patients without mental health diagnoses found no difference in perceptions, including the likelihood to report worry, toward viewing their records [69].

Subtheme 2.3: Trust, Transparency, and Accuracy of Information Can Mitigate the Risks Posed by the PAEHRs to the Therapeutic Relationship and Patient Safety

Trust and transparency mediated the potential negative impact of PAEHRs on the therapeutic relationship. Several studies found that PAEHRs could support the therapeutic relationship through improved transparency and trust [37,38,52,54,57,68], particularly when information in the PAEHR aligned with what was discussed in the clinical encounter [38,66,67]. For example, the study by Cromer et al [38] reported a negative impact on patients' trust in HCPs when they perceived low transparency about what was documented and when there was outdated information, mistakes, details that did not align with their recollection, or missing information. Similarly, patients in the study by Pisciotta et al [61] reported that the quality of information documented in their PAEHR was important to patients, including accurate notes that were not copied and pasted from previous sessions. Schwarz et al [68] described this as a "trust, but verify" situation, where patients valued the opportunity to confirm that HCPs understood them and their needs. Accuracy was also important, as patients were worried that inaccurate records could negatively affect their future treatment [38].

Subtheme 2.4: PAEHRs Offer Opportunities to Enhance Accountability and Communication Between HCPs and Patients About What Is Documented

Communication between HCPs and patients about the PAEHR and what was being documented was identified as important for building trust and strengthening the therapeutic relationship [38,50,52,53,57,66,67,70]. HCPs in the study by Denneson et al [53] discussed the role of communication, disagreements, and listening as important to consider when documenting in the PAEHR. In the study by Cromer et al [38], patients spoke of trust being built when HCPs listened to them, approached the relationship as an equal, and focused on the patient's strengths. Several studies suggested that communication about what was documented, even disagreements, could positively contribute to the therapeutic relationship [38,53,58,62]. Interestingly, Pisciotta et al [61] found that patients and HCPs avoided discussing health records for different reasons. HCPs were worried about patients challenging their records or requesting changes, and patients were worried about appearing difficult or offending the HCP. However, both HCPs and patients agreed that having a conversation about information before it is documented in the PAEHR is beneficial [61].

Although both HCPs and patients fear disagreements over what is documented in the PAEHR, disagreements could also provide an opportunity to improve accountability and communication [37,46,49,58,68,71]. Disagreements may come about due to legitimate errors in what was documented, a misunderstanding

XSL•FO
RenderX

of what was discussed and documented, or an actual disagreement between the HCP and patient as to what they understood was true [46,49,52,53,62]. However, these disagreements could improve the PAEHR by addressing errors or misunderstandings and providing opportunities to discuss issues that were previously left unspoken [46,54,58,71]. Patients commonly identified errors in their PAEHR; in a national Swedish survey by Bärkås et al [72], half of the respondents (1586/3131, 50.56%) from mental health care contexts reported finding an error, while a third (1089/3131, 34.78%) reported finding an omission. In addition, Bärkås et al [72] found that patients in mental health care settings were more likely to identify errors and rate them as serious than patients in other health care settings. Some HCPs saw the PAEHR as a way to increase accountability by having a feedback loop where patients could review their health record and report inaccuracies, errors, or misinterpretations [53,54]. Knowing that patients may disagree with what is written in the PAEHR may motivate clinicians to proactively discuss what they are documenting with patients and consider patients' needs when documenting information [53,61,62].

Although HCPs were concerned that some patients may try to dictate what should be written in their PAEHR and the care they receive [54], there was limited evidence of this occurring. For example, only 15.4% (~51/332) of HCPs in the study by Johansen et al [47] reported receiving patient feedback. However, this may be due to patients' low awareness of the PAEHR, as outlined in theme 6.

Theme 3: PAEHRs Disrupt HCPs' Documentation Practices

Overview

HCPs reported that giving patients access to their mental health records would lead to or had led to changes in how they documented information [47,49,53,57,58,61,65,73,74]. These changes were usually framed as an attempt to protect the therapeutic relationship from potential misunderstandings or to protect patients from distress caused by reading their health record. As outlined in theme 2, HCPs appear to take a risk-averse approach to PAEHRs due to a perception that patients may react negatively or experience distress from reading their PAEHR. However, HCPs also seem to recognize that certain ways of framing information about patients may trigger a negative reaction, disagreement, or distress. There was a lack of guidance, support, or training for HCPs on managing these risks when documenting information in a PAEHR [53,54,56,75].

Subtheme 3.1: HCPs Document Less Detailed Information in PAEHRs, Particularly Relating to Sensitive Information

Clinicians addressed perceived risks posed by the PAEHR by omitting or "watering down" information that the patient might perceive negatively [53,57-59,65,70,76]. Petersson et al [65], in exploring the adoption of OpenNotes in Swedish Psychiatric Care settings, found that 22% (147/667) and 17.7% (117/662) of HCPs surveyed were less candid in their documentation and took more time to edit notes, respectively. Similarly, 63% (127/~208) of HCPs in the study by Dobscha et al [59] reported that they are or plan to be less detailed in their documentation. One reason reported for including less detail was a perception that more detailed notes increased the potential for misinterpretation and disagreements [61]. An HCP in the study by Zanaboni et al [49] described it as not writing "things that you can keep to yourself," such as longer hypotheses. However, some HCPs considered that the benefit of documenting complete information outweighed the risks and that any harm could be repaired through discussions with the patient [49]. Patients preferred detailed notes that provided a complete picture of their appointment and mental health, as they perceived the health record to be a reflection of the HCPs' understanding of them [61].

HCPs and patients agreed that sensitive information required careful management in how it was documented in PAEHRs [54,58,61,65,71,74-78]. Sensitive information was defined broadly, including information that patients may not want to be shared beyond the clinical encounter, information that may be triggering or distressing for the patient, information that might be stigmatizing, and information related to the risk of self-harm or suicide [54,58,71,75,78,79]. It was common for studies to refer to mental health information generally as sensitive information. Some clinicians reported excluding specific details from the PAEHR that could be triggering, traumatic, or cause harm to patients [58,71,76]. For example, participants in the study by Kassam et al [54] discussed not documenting specific instances of physical abuse reported by patients experiencing domestic violence. Instead, the HCPs may imply domestic violence in the record by focusing on the safety planning they had discussed with the patient. There was agreement across HCPs and patients in the study by Pisciotta et al [61] that traumatic information should not be recorded in detail in the PAEHR.

HCPs identified various risks to documenting less detailed information, including that it might impact the clinical value of the EHR, particularly if the information is not detailed enough for other HCPs [54,58,62]. HCPs in the study by van Rijt et al [58] considered that they are responsible for what is documented and for what is not documented, as well as the impacts these documentation choices have on the patient. Patients may also rely on their health record to access certain services or support, which may require detailed information [61,67].

Subtheme 3.2: PAEHRs Require HCPs to Document Information That Patients Will Understand

One way that HCPs changed their documentation was to reduce or reframe the use of medical terminology or jargon that patients might not understand [54,55,61,71,80]. This change was something that both HCPs reported doing [49,54,61,66,71] and that patients wanted HCPs to do [61,66,68]. In one of the few reported examples of HCPs receiving guidelines for using a PAEHR, the guidelines recommended that HCPs limit the use of abbreviations, medical language, or euphemisms, even if they may be commonly used in the medical profession [81]. However, medical terminology may also be used by HCPs to communicate with other health HCPs or to disguise information from the patient [49,56]. Another approach to managing medical terminology was proactively discussing it with the patients to manage any potential confusion the terms might cause [54].

XSL•FO
RenderX

Subtheme 3.3: PAEHRs Require HCPs to Take a Person-Centered Approach When Documenting Potentially Subjective Information

HCPs reported that they changed their documentation to be more objective and less open for interpretation in the presence of a PAEHR [58,71,76]. For some HCPs, this change was related to a concern that information in the record may offend patients [47,52]. For example, HCPs discussed the challenge of documenting a patient's behavioral or physical presentation, which the patient could perceive as a subjective judgment [54]. Participants in the study by Blease et al [71] recommended that clinicians avoid demeaning, embarrassing, or stigmatizing terms such as "patient complains of." HCPs were conscious of medical terms that had stigmatized social meanings, such as "psychotic," that could lead to misunderstandings with patients [62,76]. This concern meant that HCPs might not document information related to the clinician-patient relationship, the patient's personality, or certain symptoms such as paranoia and may instead provide less detailed objective information [54,70]. However, it was recognized that some mental health assessments require subjective input from the HCP and that not including this information in the EHR might impact future episodes of care and cause confusion among the care team [54].

HCPs across several studies acknowledged that PAEHRs encourage them to write more person-centered and empathetic information [53,54,61]. In the study by Kassam et al [54], HCPs reported that the PAEHR made them conscious of language that "may be perceived as judgmental and demoralizing." Recognizing that a negative tone could come across as judgmental, some clinicians sought to write in a tone that conveyed empathy [61]. This need formed part of a broader consideration that HCPs should not just highlight the deficits and issues in the record but also the strengths and unique attributes of the patient [61]. Some HCPs sought to document more information on patients' strengths and recovery when using a PAEHR [54,61]. Patients have reported wanting HCPs to include such information in their record [61]. There was an acknowledgment by HCPs that while some information may be perceived as offensive, other information was explicitly disrespectful, and PAEHRs might help minimize such language [70].

It was important to patients that that the information documented in their PAEHR was respectful and empathetic. Participants in the studies by Cromer et al [38] and Fagerlund et al [46] wanted their health record to reflect them as a whole person rather than just containing "surface level observations." Patients thought that HCPs could show respect by being mindful of their tone and words that may appear as judging or labeling [61]. In the study by Bärkås et al [72], a third of patients in a national survey reported feeling offended by something written in their PAEHR.

Subtheme 3.4: HCPs Limit the Documentation of Uncertain Information in PAEHRs

HCPs discussed the challenge of documenting uncertain or unfinalized information in the PAEHR, such as diagnostic hypotheses or information that had not been confirmed with the patient [49,56-58,71,78]. For example, HCP might avoid documenting the assessment process leading to a final diagnosis or treatment plan because a patient might read those initial notes without knowing that it is not a final decision [49,53,56]. HCPs may withhold documenting suspected diagnoses due to the risk of upsetting or alarming the patient [54] or wait to discuss the diagnosis face-to-face with the patient before documenting it [49]. Some studies reported cases of patients discovering a diagnosis in their record that had not been discussed with them previously [38,61]. However, documenting uncertain information could play an important role in collaborative care arrangements and handovers, where multiple HCPs contribute to an assessment and need access to this information [56]. This information may also form an important part of the clinician's work to assess a patient and formulate a diagnosis or treatment plan [57].

Subtheme 3.5: HCPs May Omit Content or Restrict Access to PAEHRs

HCPs discussed various scenarios where information might be withheld from patients or omitted from the PAEHR [47,49,52,56,58,59,62,70,78]. These scenarios included where patients were experiencing delusions or suicidal thoughts that could be exacerbated by access to their records [54] or where a patient's knowledge of a treatment plan in advance, particularly involutory treatment, may hinder the provision of that treatment [58]. This practice may occur either through actively omitting information from the PAEHR, maintaining information in a separate informal record not sanctioned by the health service, or recording the information in a section of the PAEHR that patients cannot access. Some PAEHRs offered functionality to restrict or hide information from patients [47,53,60]. Some PAEHRs provided functionality to limit patient access to information, including allowing for a delay to be implemented before patients could view information [49,78,82], allowing only "signed" notes that an HCP had validated to be viewed by patients [47,52,78,82], or providing specific templates where HCPs could document sensitive information that patients could not view [65,74,82]. Some HCPs in the study by Zanaboni et al [49] mentioned keeping a physical folder with information that had not yet been included in the EHR. The risk of such unsanctioned records is that the information may be lost if not entered into the EHR or inaccessible by other HCPs [56]. HCPs discussed the need for discretion in determining when to withhold information from patients, particularly when there is a justified reason for doing so [54]. Other HCPs thought that patients should have access to all their health information and that all patients should have equal access to their PAEHR [53], a view also held by some patients [83]. Limiting patient access to information, particularly mental health information, could perpetuate stigma and distrust of services [54].

HCPs across several studies identified certain patients that they considered at increased risk of negative outcomes from accessing their PAEHR [47,49,59,65,71,76]. The groups reported were those with psychosis, paranoia, high levels of suspicion, or personality disorders [49,65,71]. In a pilot study of OpenNotes by Peck et al [76], where HCPs could choose which patients to enroll, HCPs reported choosing patients based on the severity of their illness, length of time in treatment, and psychiatric diagnosis. Similarly, in the study by Zanaboni et al

XSL•FO
RenderX

[49], HCPs in an outpatient setting discussed how they do not treat "the most challenging patients." HCPs in the study by Kassam et al [54] reported that their documentation practices may change with specific patients in mind, including their diagnosis and clinical program. HCPs may also consider the context, such as whether patient access during a crisis may negatively shape their engagement with mental health care services, particularly planned involuntary treatment [58]. Although HCPs across several studies believed that a certain subgroup of patients, either due to diagnosis or current symptoms, may be at greater risk of harm from accessing their health record, there seemed to be a lack of guidance or support for making these decisions. Thus, it may be that the available evidence does not accurately reflect the experiences of certain groups of patients if HCPs are actively not encouraging certain patients to use their PAEHR, or because studies are not being undertaken in settings with patients who have more complex conditions. This finding that HCPs may avoid involving certain patients in using their PAEHR may also explain why, in some studies, patients with certain diagnoses had lower odds of being registered for a PAEHR [84-87].

Theme 4: PAEHRs Introduce Changes to HCPs' Information Workflows

HCPs across several studies reported that PAEHRs would or had increased their workload due to changes in documenting information, answering patient questions, and reassuring patients [10,47,49,52,55,58,59,65,68,80]. However, in other studies, HCPs did not report any change to their workflow or workload [49,52,59,65,71] or suggested that changes in documentation may only require more work initially to adapt their documentation style [54]. For example, only 14.5% (86/594) of HCPs in the study by Petersson et al [65] reported that appointments took longer, and 18% (106/588) reported spending more time answering patient questions. The variation in the impact of PAEHRs on HCP workflows may be related to how the PAEHR was implemented and the level of support provided to HCPs and patients. One workflow change reported was the order of communication and documentation of information [49]. As previously discussed in subtheme 3.5, HCPs across several studies had previously used the health record to document preliminary information. The adoption of a PAEHR led HCPs to document this preliminary information elsewhere or not at all or to have earlier discussions with the patient about this information [49,56-58,71,78]. Another workflow issue raised by HCPs in the study by van Rijt et al [58] was the time required to explain what is written in the PAEHR and update the PAEHR if a patient disagreed with what was documented. In some cases, workflow issues were related to technical issues with the PAEHR, such as slow sign-in or troubleshooting technical issues for patients [88]. Interestingly, patients in the study by Schwarz et al [68] were concerned that the PAEHR might undermine the trust established with their HCP partly due to an increased HCP workload.

Theme 5: HCPs Require Tailored Training, Education, and Guidelines Documenting Information in a PAEHR

There was a clear lack of support and guidelines for HCPs across the included studies [49,53,75]. As outlined in theme 1, PAEHRs were viewed by HCPs as contrary to how they were trained. Several studies discussed the need for HCPs to receive training and support in using and promoting PAEHRs, such as how to write notes appropriately for patients [53,61,73,75,89,90]. Some of the issues identified as requiring guidance included describing sensitive matters and writing information "that may potentially be perceived in a negative light in a way that is clinically appropriate and respectful" [71]. When guidance was unavailable, HCPs took a risk-averse approach by adopting the information practices outlined in theme 3, such as reducing the level of detail documented about patients.

The few studies that included training for HCPs found mixed results regarding its effectiveness. Dobscha et al [73], reporting on a study of web-based education programs for HCPs using OpenNotes, found that after training, clinicians were better able to communicate, educate, and advise patients on accessing and reading their notes. The training also reduced HCPs' concerns regarding the potential negative outcomes of OpenNotes. However, HCPs in the study by Zanaboni et al [49] who also completed internet-based learning modules discussed how they would like more formal training in using the PAEHR, as their memories of the online training were vague. HCPs in the study by Jonnergård et al [63] identified that HCPs preferred to receive information about OpenNotes via channels that allowed for dialogue and rich information, such as workplace meetings.

Theme 6: Patients Are Empowered With New Information Practices if They Are Supported to Use Their PAEHR

Overview

Many studies reported improvements in patient empowerment or the perception of improved empowerment with the adoption of PAEHRs. Empowerment included improvements in self-monitoring; medication management and self-management of conditions; increased engagement with health care providers, health screening, and care plans; improved understanding of and insight into one's mental health; and feeling more prepared to engage with HCPs [37,50,52,54,57,59,67,68,73,76,77,90-95]. For example, patients in the study by Shin et al [66] reported that seeing their health record made them feel involved in the health care process, which was empowering. HCPs in the study by Kassam et al [54] thought PAEHRs helped dismantle power imbalances between patients and HCPs, fostering patient empowerment. Many of these studies were patients' self-reports or clinicians' perceptions of increased empowerment. However, Kipping et al [92] found that, when comparing patient portal users with nonusers, those who used the portal had increased appointment attendance, self-reported autonomy, and patient activation. Furthermore, in a randomized control trial, Druss et al [96,97] found that personal health record users had increased engagement with preventative health services.

PAEHRs could also support patients' engagement with information by acting as a reference tool, supporting their preparation for and engagement in clinical encounters, reminding them of previous discussions with HCPs, and helping them keep track of their treatment [37,49,52,54,57-59,66,68,70,98]. PAEHRs may also improve

XSL•FO
RenderX

patients' knowledge of their health status, treatment, and medications [47,55,58,67,95]; increase the number of contacts with HCPs and the completion of follow-up assessments [94]; encourage patients to ask more questions about their treatment [62]; and help prepare them to discuss their treatment with an HCP [37,50,58]. The transparency promoted by the PAEHR, as discussed in subtheme 2.3, can also help patients track how their care is progressing [83]. Patients may also use the PAEHR to validate communications from their HCP and clarify information after a clinical encounter [52,66]. However, some patients were concerned that PAEHRs would replace face-to-face conversations with their HCPs [66].

In most studies that explored the rates of PAEHR adoption, the engagement of patients was low [84,85,92,99,100]. Connolly et al [101] found in a study of veterans with depression that only 21.9% (~669/3053) were registered for their My HealtheVet platform. HCPs across several studies also reported low PAEHR adoption and patient engagement with their PAEHR [49,52,57,93]. There were mixed findings on the demographic groups that were more likely to access their PAEHR. Connolly et al [101] found that patients with more severe symptoms were more likely to register for and download information from their patient portal. Some studies suggest that people with mental health conditions may be more likely to use a PAEHR [72,102], but this was not consistent across all studies [103]. PAEHR use may increase over time, with Onyeaka et al [12] finding in a national survey in the United States of people with depression and anxiety that patient portal use had increased from approximately a third to a half of respondents over 3 years. There were variable results regarding whether people with different diagnoses were more likely to use a PAEHR. Etingen et al [84] suggested that differences in use across diagnoses may be due to some diagnoses being more likely to be associated with other comorbidities, which may encourage or facilitate engagement with the PAEHR.

Interestingly, in studies that explored the functions of PAEHRs used by patients, viewing their health records is not always the primary use [12,84,101]. In many cases, other functions, such as medication refills and booking and viewing appointments, are used most [84]. Robotham [90] and Forchuk et al [98] also found that patients value visual representations of, and the ability to track changes in, their health information.

Only a few studies discussed patients documenting information in their PAEHR [37,55,66]. Patient-documented information is not a feature of all PAEHR systems, but some patients saw value in this functionality. For example, patients in the study by Shin et al [66] discussed how being able to report information in the PAEHR might help HCPs better understand their needs. In the study by Durocher et al [37], 1 participant thought it would be valuable if they could flag notes in the PAEHR and provide their interpretations of certain events.

Subtheme 6.1: Patients Need Support to Use Their PAEHR

Patients were more likely to access their PAEHR when HCPs spoke to them about accessing their PAEHR and when they understood its purpose [87]. However, in several studies, HCPs took a passive approach by letting patients figure out how to access and read their health record themselves [47,52,59,65]. In some studies, it appeared that not all HCPs had discussed the PAEHR with their patients; for example, 40% (82/205) of participants in the study by Dobscha et al [59] had not discussed it with their patients, and only a few participants reported having multiple conversations with patients about it. Similarly, in the study by Johansen et al [47], 50.8% (~169/332) of the HCPs surveyed reported that they informed patients that they could read their own EHR. In the study by Petersson et al [65] study, only 26.8% (180/671) of the HCPs encouraged patients to read their PAEHR, while 15.6% (105/671) initiated a discussion with patients about something in their PAEHR, and only 10% (67/670) used the PAEHR actively in treatment. This issue was also reflected in the study by Fagerlund [46], where patients could not recall being informed about the PAEHR and instead had heard about it through advertisements or media coverage. In addition, Leung et al [104] found that 50% (~51/103) of patients surveyed were aware that they could access their health records, while Bärkås et al [72] found that 64.77% (2028/3131) of patients in a national survey in Sweden had not been encouraged by anyone to read their PAEHR.

It is unclear why HCPs are not discussing the PAEHR with patients. Earlier themes might suggest that this finding is related to the perceived risks of PAEHR to patients or the lack of support for HCPs [53,59]. Some HCPs in the study by Zanaboni et al [49] suggested that the PAEHR is the patient's tool, and therefore, the patient had to take the lead in using the PAEHR, rather than it being driven by the HCP. This perception of the PAEHR not being the HCP's responsibility may be supported by a lack of training or information for HCPs [49].

Several studies implemented a PAEHR as part of an intervention or model of care, which appeared to lead to increased use by patients. For example, Weisner et al [105] found that patients in an intervention that involved six 45-minute sessions with a psychologist focused on patient education and activation had greater log-ins to a patient portal and checking of information in the portal than those not involved in the intervention. Kelly et al [93] found in a randomized controlled trial that patients were more likely to access a PAEHR when it was implemented alongside support from a peer navigator. Druss et al [97] also found that peer specialist involvement might contribute to higher PAEHR use.

Providing educational resources and support may help patients engage with their PAEHR. Kelly et al [93], when comparing low engagement in their study with higher engagement in the study by Druss et al [96], noted that the latter provided more intense resources and support to patients. Denneson et al [106] found in implementing a web-based education program for patients who accessed their health record on the internet that the training improved patients' interactions and communication with HCPs and their own "activation." The web-based program included information on what is included in their health record and how to use and communicate with clinicians about them.

Subtheme 6.2: PAEHRs Must Be Easy for Patients to Navigate, Access, and Understand

Several studies explored design requirements for PAEHRs and identified issues and opportunities to make it easier for patients to use their PAEHR [12,61,73,75,89,90,98,107,108]. Several

XSL•FO
RenderX

studies identified login and account setup processes as an issue for patients, with recommendations that the process should be simplified, including addressing issues such as remembering passwords [12,75,83,89,98,109]. Patients also raised basic accessibility issues such as font size and colors [89,98,107] and navigation issues, including the number of links that patients had to navigate [79,107]. For example, in the study by van den Heuvel [91], patients who had dropped out of using an ePHR cited issues with the user friendliness as the main reason, along with the ePHR requiring too much work and effort. HCPs may need to support patients in navigating and understanding their PAEHRs, particularly patients with low health literacy or digital literacy, cognitive issues, and memory issues [68,71,107,108]. For HCPs to support patients, they need to be aware of what patients see in the record and how it functions for patients [58]. A lack of technology skills could be a barrier to patients engaging with their PAEHR [12,68], while some patients may also lack access to technology or the internet needed to use their PAEHR [12]. There is a risk that the benefits of PAEHRs may primarily benefit patients who are literate, educated, and affluent [10].

Theme 7: PAEHRs Raise New Concerns About Information Privacy and Security for HCPs, Patients, and Third Parties

Overview

Both patients and HCPs shared general concerns about privacy and security threats introduced by the PAEHR [49,52,57,62,68,76,104,107]. Some of these concerns were related more broadly to EHRs that multiple HCPs could access rather than the PAEHR itself [52,76]. For example, some patients in the studies by O'Neill et al [13] and Peck et al [76] raised concerns about their health record being available to other HCPs without their explicit consent. General concerns were also expressed about data hacking and breaches, but no participants reported experiences of these issues [10,62,68,76,77,79,104,107]. The most specific privacy threats identified were managing third-party data in the PAEHR and third-party access to the PAEHR, as outlined in the subsequent sections.

Subtheme 7.1: PAEHRs Pose Challenges to Managing Third-Party Data

An issue reported by HCPs was how to protect the privacy of third parties, such as family members, when documenting information they had provided in the health record information [49]. For example, HCPs might gather information about a patient from their family, which, when recorded in the PAEHR, would become visible to the patient. This risk might lead to HCPs reporting less information from family members or carers in the health record, which, as 1 participant in the study by Erlingsdóttir et al [57] said, will be a "big problem because it is important information that otherwise falls out of the system." Another approach reported by participants in the study by Zanaboni et al [49] was to deidentify information provided by third parties.

Subtheme 7.2: PAEHRs Pose Challenges to Managing Third-Party Access to Patient Information

Both HCPs and patients raised concerns that PAEHRs could lead to third parties, such as family members, inappropriately accessing a patient's health record [49,57,68,71]. Participants in the study by van Rijt et al [58] outlined the risk that patients could be influenced to share copies of their records or login details with family members who they may not want to have access to this information. There was also the broader issue of whether HCPs should record information assuming that the patient might give a third-party access to their PAEHR. For example, 1 participant in the study by Denneson et al [53] stated the following:

> *...someone said, "I smoked meth for 40 years and my wife doesn't know." And I was like, gosh, do I put this in the note? Because I don't know if he is going to give his wife access to his notes and then see something that was delivered in confidence...*

Such concerns might prompt HCPs to deny access altogether to certain information in the PAEHR to manage the risk of third-party access [49].

A few studies explored the potential benefits and issues with family members or carers being given access to PAEHRs [54,68,77,91,98,104]. One of the few studies that involved family and carers found that they thought it would be helpful to support the patient if they had access to the PAEHR [104]. Similarly, in the study by Leung et al [77], some patients reported that family or carer access could also be beneficial if it helped their family members or carers to better understand their health issues. However, in the same study, 43.7% (~43/103) of patients were interested in family members having access. The study by Schwarz et al [68] found that family access could enhance the exchange of information about the patient's illness and help family members participate in the patient's care. Van den Heuvel et al [91] reported that 23.1% (~9/39) of patients preferred informal caregivers to have access compared to 25.6% (10/39) finding it adverse. Similarly, Leung et al [104] found, in interviewing patients about a patient portal, that less than half of those interviewed were interested in family members having access to their records. There may also be scenarios where family members support the patient in using the PAEHR [110], which raises questions about how to ensure that the patient has the option to maintain privacy in those situations.

Discussion

Principal Findings

This scoping review explored how PAEHRs implemented in mental health care contexts change the information practices of HCPs and patients. We found that HCPs change their documentation practices in the presence of a PAEHR, including reducing the level of detail they document. HCPs changed their information practices because they were concerned that the PAEHR would negatively impact the therapeutic relationship and patient safety, and there was a lack of guidelines as to how these risks should be managed. There were clear benefits for patients who could use the PAEHR to understand and manage

XSL•FO
RenderX

their care. However, patients may struggle to use the PAEHR without support and consideration of its ease of use.

The findings of this review suggest that HCPs perceive that PAEHRs change the purpose of the health record by requiring them to write with consideration of the patient as an audience rather than solely a clinical audience. It is a long-running issue that in an increasingly connected health care system, while technical interoperability between EHRs may increase, information is not necessarily written for multiple audiences [111]. The findings of this review suggest that digital health technologies that allow records created or updated by HCPs to be indiscriminately accessed by multiple types of users risk HCPs adopting information practices outside formal systems and processes, such as by creating separate records that are not officially recognized by the health service [49,112]. Omitting information from a patient's health record not only fragments the record but also poses a risk to future care if information is lost or unavailable when needed. Potential technical advances, such as natural language processing and generative artificial intelligence, may solve this issue by supporting HCPs in translating their notes for different audiences [113-115].

HCPs in the included studies were concerned that PAEHRs would negatively impact the therapeutic relationship. The concern for the therapeutic relationship is not surprising, given it is a strong predictor of outcomes in mental health care [116]. HCPs and patients also thought that the PAEHR could strengthen the therapeutic relationship by contributing to transparency, communication, and trust. This finding aligns with evidence that mutual trust, demonstration of mutual respect, and shared decision-making were key components of a beneficial therapeutic relationship [117]. This finding may also explain why HCPs and patients saw benefits in the PAEHR facilitating communication about disagreements, as it helps build mutual trust [118]. However, both HCPs and patients expressed concerns that discussing what is documented in the PAEHR may add to workloads. A lack of time available for HCPs and patients to actively review and discuss the PAEHR may limit its use in strengthening the therapeutic relationship.

Risk to patient safety due to patients experiencing distress from reading their PAEHR was a key concern for both HCPs and patients. One of the key strategies used by HCPs to manage risk was to not document or limit the documentation of information that they perceived may be distressing for patients. This focus on risk is expected, given risk management has a prominent role in mental health care, as HCPs must consider and manage risks posed by patients to themselves or others [119]. One way HCPs manage risk is by avoiding situations that may involve heightened risks [119,120]. In this review, HCPs avoided risk by not documenting information they perceived could pose a risk to patients. This approach to avoiding risk may have been reinforced by a lack of guidelines for managing risk. HCPs are usually required to take a systematic approach to risk assessment, informed by organizational policies [121], which appeared to be lacking in relation to PAEHR adoption in the included studies. Evidence also suggests that there is value in involving patients in identifying and managing risks [122]. However, patient involvement in HCPs' decisions to limit the documentation of information in the PAEHR appeared to be limited.

Patients wanted a greater breadth of information documented in their PAEHR, including strength-based and recovery-focused information. In contrast, HCPs managed the perceived risks of PAEHRs by reducing the depth of information they documented. These findings align with the growing literature on person-centered documentation, which seeks to consider the role patients can play in information documentation practices. Lyadhl et al [123] reflects that health record documentation practices are embedded in a biomedical model of care that prioritizes certain types of information to support communication between HCPs, rather than communication with patients. Person-centered documentation requires consideration of the patient's goals, needs, resources, and narrative. "Collaborative documentation" is a model that supports HCPs and patients in working together to determine what should be documented during the clinical encounter [124,125]. It is a form of shared decision-making that seeks to improve the transparency of what is documented in the patient's EHR [124]. Adopting a model such as collaborative documentation may address some of the concerns that PAEHRs will lead to disagreements about what is documented in the PAEHR.

PAEHRs can support new patient information practices, but there are barriers to patients using their PAEHR. PAEHR use by patients was low in the included studies, which is concerning given that recent studies in other health care settings have found generally high use rates [126,127]. The lower rates of PAEHR use in mental health care contexts may be partly explained by the slower adoption of PAEHRs in mental health care contexts, as the use of PAEHRs tends to increases over time [12,128]. It is also worth noting that many of the studies that considered use rates were conducted before the COVID-19 pandemic. The pandemic may have accelerated the adoption and use of PAEHRs due to the limits on in-person contact and the rapid adoption of telehealth [129].

Patient uptake of PAEHRs also relies on endorsement from HCPs [130]. However, our findings suggest that the lack of guidance for HCPs paired with the perceived risks to the therapeutic relationship and patient safety may act as a barrier to HCPs actively engaging patients in using their PAEHR. It was also identified that HCPs may not proactively engage certain groups of patients whom they perceive as at greater risk of negative outcomes from using their PAEHR. This finding may partly explain why, in some studies, patients with certain diagnoses were less likely to use their PAEHR. There is value in considering which patients may experience a negative response to a accessing their PAEHR, noting that some patients in the included studies did have negative experiences. Patient readiness measures, which have been found to be valid in measuring the likelihood of patients using health information technology [131,132], could form part of guidelines for HCPs to identify patients who need further support to use their PAEHR.

The time and resource pressures on HCPs are well documented, as are the impacts of these pressures on HCPs' ability to engage with new digital health technologies [133,134]. In some mental

XSL•FO
RenderX

health care services, other professionals could support patients to use their PAEHR, if HCPs are not provided with the time and resources to do this role. For example, peer workers are increasingly involved in mental health care services [135] and could play a role in supporting patients to use their PAEHR, as well as role-modeling how can be integrated into [136]. Similarly, care workers or "health care navigators" who in some health care systems support patients to move between health care services [137] may be well positioned to support patients in using their PAEHR. However, in all these cases, consideration will need to be given to how confidentiality of patient information is managed.

Managing third-party access to the PAEHR, particularly coercive access, was a key issue for HCPs but was not raised as a key issue by patients. It may be that patients at risk of having third parties inappropriately access their records were not participants in the included studies. The few studies that asked patients about family access found that they were uncomfortable with family members having access to their PAEHR. The issue of third-party access could be addressed through initiatives such as dynamic consent, where patients are given real-time control over managing their data through a personalized digital interface [138]. Australia's national PAEHR, My Health Record, presents another approach for patients to manage third-party access by allowing them to nominate a trusted person to access their record and determine the level of access available to them [139]. Health care services could also provide kiosks for patients to access their PAEHR in the health care clinic if at-home access poses a risk of unauthorized third-party access. However, these approaches do not necessarily address the issue of third parties who may coerce patients into providing access to their PAEHR.

The findings of this review also raise questions as to how inequalities in patients' resources and digital literacy may shape their interactions with a PAEHR. In the mental health care context, patients can experience socioeconomic disadvantages, which may limit their access to technology, meaning they may rely on smartphones or public computers [96,107,140]. PAEHRs should be designed to maintain privacy and accessibility, regardless of the location or type of device patients use to access them. There is a risk that the benefits of PAEHRs may disproportionately favor patients with more resources, rather than those facing the greatest health challenges [141,142]. These issues pose a more serious question about what the health system expects of patients when a PAEHR is available. Wynia and Dunn [143] outline 2 assumptions underlying PAEHRs: that more information for patients equals better decision-making and that patients should take greater responsibility for their health care. Lucivero [144] posits that PAEHRs may enroll patients into providing free labor for the health system where they act as intermediaries between HCPs by using their PAEHR to fill gaps in information. However, this role may be undermined if HCPs do not support patients to use their PAEHR, such as if HCPs are unwilling to act on information patients share with them, or if patients do not have the resources to fulfill this role.

Comparison With Prior Work

The findings of this review align and extend the findings of recent reviews by Zhang et al [24] and Schwarz et al [9] in mental health care contexts. Some of the factors for successful implementation of patient portals identified by Zhang et al [24], such as training and resourcing, are also issues that we identified as contributing to some of the changes to HCPs' information practices. Similar to the findings of Schwarz et al [9], we found that HCPs were more likely to express concerns about PAEHRs compared to patients. Our study extends the research into PAEHRs in mental health care contexts by framing the perceived risks of PAEHRs to the therapeutic relationship and patient safety as key factors in HCPs changing their information practice, further fueled by a lack of guidance and training. Several recent reviews of PAEHRs in other health care contexts also support the findings of this study. In a systematic review, Tapuria et al [31] found that patient portals had a range of benefits, such as reassuring patients and reducing anxiety, improving the physician-patient relationship, and improving patient outcomes. However, they also identified concerns regarding security, privacy and confidentiality, and patients experiencing anxiety with access to their records. In a recent systematic review of facilitators and barriers to electronic portal use, Powell [145] found that HCP encouragement is a key factor in both the initial and continued use of the PAEHR by patients. The findings of our review suggest that PAEHRs in mental health care contexts raise similar issues to their use in other health care contexts. However, we also found that that in mental health care contexts, HCPs still harbor significant concerns about how to protect the therapeutic relationship and patient safety—2 issues that have distinctive implications in the mental health care context.

What Do the Findings Mean for Future Research?

Future research should consider how changes to HCP and patient information practices after the adoption of a PAEHR impacts clinical and patient experience outcomes. This study found that the use of PAEHRs led to changes in what is documented in the EHR and a potential "watering down" of information. Future research should explore the impact of these changes on collaborative care and clinical outcomes in the presence of a PAEHR. If these changes to information practices significantly impact the ability of other HCPs to provide appropriate care, it will strengthen the case for increased training, support, and guidance for how HCPs should document information in the presence of a PAEHR. It would also be valuable for more studies to explore the implementation and use of PAEHRs in practice and how HCP and patient use change over time, particularly given that use of PAEHRs tends to increase over time. There are very few studies that involved HCPs and patients in the process of designing a PAEHR. Given the findings of this review, there is a clear need for more studies that actively involve HCPs and patients in decisions about how PAEHRs are designed and implemented.

Limitations

This review is limited in part due to the heterogeneity of the field, particularly the lack of common definitions of PAEHRs, which means that there may be studies that were missed in the

XSL•FO
RenderX

search strategy. Furthermore, our search strategy did not include gray literature, meaning evaluations of some PAEHR systems that have not been published in academic journals may have been missed. This review was also limited by having only 1 author screen all the studies and extract the data for analysis. A pragmatic approach to screening was used, where only studies with unclear eligibility for inclusion were screened by a second author.

Conclusions

This scoping review explored the impact of PAEHRs on information practices in mental health contexts. HCPs reported making various changes to their documentation practices to minimize the risk of PAEHRs harming the therapeutic relationship and patient safety. However, PAEHRs could also promote communication between HCPs and patients about what was being documented. PAEHRs also introduced new information practices for patients that allowed them to manage their health better and engage with health care services. PAEHRs are still considerably new in mental health care contexts, and patient adoption and use of established PAEHRs remain relatively low. Research is required on the impact of documentation changes on clinical outcomes and patient experience as well as approaches to improve adoption and use rates of PAEHRs. The findings of this study point to a lack of guidelines and support for HCPs and patients in the implementation of PAEHRs.

Acknowledgments

TK was supported by an Australian Government Research Training Program stipend and Research Training Program Fee-Offset Scholarship through the University of Melbourne.

Generative artificial intelligence was not used in the research, preparation, writing, or editing of this manuscript.

Authors' Contributions

TCK was involved in conceptualization, writing the original draft, and investigation. MP, KG and SC were involved in reviewing and editing the draft and study supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 109 KB - jmir_v27i1e54973_app1.docx]

Multimedia Appendix 2
Details of the included studies.
[DOCX File , 37 KB - jmir_v27i1e54973_app2.docx]

Multimedia Appendix 3
Illustrative quotes.
[DOCX File , 24 KB - jmir_v27i1e54973_app3.docx]

References

1. Tavakoli N, Isfahani SS, Piri Z, Amini A. Patient access to electronic health record: a comparative study on laws, policies and procedures in selected countries. Med Arch 2013;67(1):63-67. [doi: 10.5455/medarh.2013.67.63-67] [Medline: 23678844]
2. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy Technol 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]
3. Ross SE, Lin CT. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003;10(2):129-138 [FREE Full text] [doi: 10.1197/jamia.m1147] [Medline: 12595402]
4. Caligtan CA, Dykes PC. Electronic health records and personal health records. Semin Oncol Nurs 2011 Aug;27(3):218-228. [doi: 10.1016/j.soncn.2011.04.007] [Medline: 21783013]
5. Esch T, Mejilla R, Anselmo M, Podtschaske B, Delbanco T, Walker J. Engaging patients through open notes: an evaluation using mixed methods. BMJ Open 2016 Jan 29;6(1):e010034 [FREE Full text] [doi: 10.1136/bmjopen-2015-010034] [Medline: 26826154]
6. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017 Apr 01;24(e1):e166-e172 [FREE Full text] [doi: 10.1093/jamia/ocw108] [Medline: 27497795]

XSL•FO
RenderX

7. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
8. Kahn MW, Bell SK, Walker J, Delbanco T. A piece of my mind. Let's show patients their mental health records. JAMA 2014 Apr 02;311(13):1291-1292. [doi: 10.1001/jama.2014.1824] [Medline: 24691603]
9. Schwarz J, Bärkås A, Blease C, Collins L, Hägglund M, Markham S, et al. Sharing clinical notes and electronic health records with people affected by mental health conditions: scoping review. JMIR Ment Health 2021 Dec 14;8(12):e34170 [FREE Full text] [doi: 10.2196/34170] [Medline: 34904956]
10. Strudwick G, Clark C, Sanches M, Strauss J. Predictors of mental health professionals' perceptions of patient portals. AMIA Annu Symp Proc 2018;2018:989-997 [FREE Full text] [Medline: 30815142]
11. Laugharne R, Stafford A. Access to records and client held records for people with mental illness. Psychiatr Bull 2018 Jan 02;20(6):338-341 [FREE Full text] [doi: 10.1192/pb.20.6.338]
12. Onyeaka H, Ajayi KV, Muoghalu C, Eseaton PO, Azuike CO, Anugwom G, et al. Access to online patient portals among individuals with depression and anxiety. Psychiatry Res Commun 2022 Dec;2(4):100073 [FREE Full text] [doi: 10.1016/j.psycom.2022.100073]
13. O'Neill S, Blease C, Delbanco T. Open notes become law: a challenge for mental health practice. Psychiatr Serv 2021 Jul 01;72(7):750-751. [doi: 10.1176/appi.ps.202000782] [Medline: 33971748]
14. Clavier V, Paganelli C. Knowledge organization and studies of information practices: contribution to the characterization of health information. In: Paganelli C, Clavier V, editors. Information Practices and Knowledge in Health. Hoboken, NJ: John Wiley & Sons; 2022.
15. Kariotis TC, Prictor M, Chang S, Gray K. Impact of electronic health records on information practices in mental health contexts: scoping review. J Med Internet Res 2022 May 04;24(5):e30405 [FREE Full text] [doi: 10.2196/30405] [Medline: 35507393]
16. Østensen E, Bragstad LK, Hardiker NR, Hellesø R. Nurses' information practice in municipal health care-a web-like landscape. J Clin Nurs 2019 Jul 23;28(13-14):2706-2716. [doi: 10.1111/jocn.14873] [Medline: 30938870]
17. Langan J. Mental health, risk communication and data quality in the electronic age. Br J Soc Work 2009;39(3):467-487 [FREE Full text] [doi: 10.1093/bjsw/bcm129]
18. Grando MA, Murcko A, Mahankali S, Saks M, Zent M, Chern D, et al. A study to elicit behavioral health patients' and providers' opinions on health records consent. J Law Med Ethics 2017 Jun 14;45(2):238-259 [FREE Full text] [doi: 10.1177/1073110517720653] [Medline: 30976154]
19. Mohsen MO, Aziz HA. The blue button project: engaging patients in healthcare by a click of a button. Perspect Health Inf Manag 2015;12(Spring):1d [FREE Full text] [Medline: 26755898]
20. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
21. Blease C, Salmi L, Rexhepi H, Hägglund M, DesRoches CM. Patients, clinicians and open notes: information blocking as a case of epistemic injustice. J Med Ethics 2021 May 14;48(10):785-793 [FREE Full text] [doi: 10.1136/medethics-2021-107275] [Medline: 33990427]
22. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]
23. Garets D, David M. Electronic medical records vs. electronic health records: yes, there is a difference. HIMSS Analytics. 2006 Jan 26. URL: https://www.aao.org/asset.axd?id=8e9b1f20-0ed6-4d2b-92f8-e28fbaf378ec&t=634962799822530000 [accessed 2023-10-29]
24. Zhang T, Shen N, Booth R, LaChance J, Jackson B, Strudwick G. Supporting the use of patient portals in mental health settings: a scoping review. Inform Health Soc Care 2022 Jan 02;47(1):62-79. [doi: 10.1080/17538157.2021.1929998] [Medline: 34032528]
25. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
26. Harahap NC, Handayani PW, Hidayanto AN. Functionalities and issues in the implementation of personal health records: systematic review. J Med Internet Res 2021 Jul 21;23(7):e26236 [FREE Full text] [doi: 10.2196/26236] [Medline: 34287210]
27. Jung SY, Kim JW, Hwang H, Lee K, Baek RM, Lee HY, et al. Development of comprehensive personal health records integrating patient-generated health data directly from Samsung s-health and Apple health apps: retrospective cross-sectional observational study. JMIR Mhealth Uhealth 2019 May 28;7(5):e12691 [FREE Full text] [doi: 10.2196/12691] [Medline: 31140446]
28. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]

XSL•FO
RenderX

29. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021 Sep 08;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
30. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020 Nov 11;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]
31. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 02;46(2):192-204. [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
32. Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E. How the use of a patient-accessible health record contributes to patient-centered care: scoping review. J Med Internet Res 2021 Jan 11;23(1):e17655 [FREE Full text] [doi: 10.2196/17655] [Medline: 33427683]
33. DePuccio MJ, Di Tosto G, Walker DM, McAlearney AS. Patients' perceptions about medical record privacy and security: implications for withholding of information during the COVID-19 pandemic. J Gen Intern Med 2020 Oct;35(10):3122-3125 [FREE Full text] [doi: 10.1007/s11606-020-05998-6] [Medline: 32737791]
34. Cherif E, Mzoughi M. Electronic health record adopters: a typology based on patients' privacy concerns and perceived benefits. Public Health 2022 Jun;207:46-53. [doi: 10.1016/j.puhe.2022.03.010] [Medline: 35486983]
35. Lee SJ, Crowther E, Keating C, Kulkarni J. What is needed to deliver collaborative care to address comorbidity more effectively for adults with a severe mental illness? Aust N Z J Psychiatry 2013 Apr;47(4):333-346. [doi: 10.1177/0004867412463975] [Medline: 23076548]
36. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
37. Durocher K, Shin HD, Lo B, Chen S, Ma C, Strudwick G. Understanding the role of patient portals in fostering interprofessional collaboration within mental health care settings: mixed methods study. JMIR Hum Factors 2023 Jul 19;10:e44747 [FREE Full text] [doi: 10.2196/44747] [Medline: 37467024]
38. Cromer R, Denneson LM, Pisciotta M, Williams H, Woods S, Dobscha SK. Trust in mental health clinicians among patients who access clinical notes online. Psychiatr Serv 2017 May 01;68(5):520-523 [FREE Full text] [doi: 10.1176/appi.ps.201600168] [Medline: 28142383]
39. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
40. Tubaishat A. Perceived usefulness and perceived ease of use of electronic health records among nurses: application of Technology Acceptance Model. Inform Health Soc Care 2018 Dec 18;43(4):379-389. [doi: 10.1080/17538157.2017.1363761] [Medline: 28920708]
41. Morris MG, Dillon A. How user perceptions influence software use. IEEE Softw 1997;14(4):58-65. [doi: 10.1109/52.595956]
42. Lakshman R, Ogilvie D, Ong KK. Mothers' experiences of bottle-feeding: a systematic review of qualitative and quantitative studies. Arch Dis Child 2009 Aug;94(8):596-601 [FREE Full text] [doi: 10.1136/adc.2008.151910] [Medline: 19602520]
43. Sartain SA, Stressing S, Prieto J. Patients' views on the effectiveness of patient-held records: a systematic review and thematic synthesis of qualitative studies. Health Expect 2015 Dec;18(6):2666-2677 [FREE Full text] [doi: 10.1111/hex.12240] [Medline: 25059439]
44. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
45. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodol 2009 Aug 11;9:59 [FREE Full text] [doi: 10.1186/1471-2288-9-59] [Medline: 19671152]
46. Fagerlund AJ, Kristiansen E, Simonsen RA. Experiences from using patient accessible electronic health records - a qualitative study within Sámi mental health patients in Norway. Int J Circumpolar Health 2022 Dec 17;81(1):2025682 [FREE Full text] [doi: 10.1080/22423982.2022.2025682] [Medline: 35037572]
47. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health professionals' experience with patients accessing their electronic health records: results from an online survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]
48. Kristiansen E, Johansen M, Zanaboni P. Healthcare personnels' experience with patients' online access to electronic health records: differences between professions, regions, and somatic and psychiatric healthcare. In: Proceedings of the 17th Scandinavian Conference on Health Informatics. 2019 Presented at: SHI 2019; November 12-13, 2019; Oslo, Norway.
49. Zanaboni P, Kristiansen E, Lintvedt O, Wynn R, Johansen MA, Sørensen T, et al. Impact on patient-provider relationship and documentation practices when mental health patients access their electronic health records online: a qualitative study among health professionals in an outpatient setting. BMC Psychiatry 2022 Jul 28;22(1):508 [FREE Full text] [doi: 10.1186/s12888-022-04123-7] [Medline: 35902841]
50. Fisher B, Bhavnani V, Winfield M. How patients use access to their full health records: a qualitative study of patients in general practice. J R Soc Med 2009 Dec;102(12):539-544 [FREE Full text] [doi: 10.1258/jrsm.2009.090328] [Medline: 19966130]

XSL•FO
RenderX

51. Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI--statement on reporting of evaluation studies in health informatics. Int J Med Inform 2009 Jan;78(1):1-9. [doi: 10.1016/j.ijmedinf.2008.09.002] [Medline: 18930696]
52. Chimowitz H, O'Neill S, Leveille S, Welch K, Walker J. Sharing psychotherapy notes with patients: therapists' attitudes and experiences. Soc Work 2020 Apr 01;65(2):159-168. [doi: 10.1093/sw/swaa010] [Medline: 32236447]
53. Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK. A qualitative analysis of how online access to mental health notes is changing clinician perceptions of power and the therapeutic relationship. J Med Internet Res 2017 Jun 14;19(6):e208 [FREE Full text] [doi: 10.2196/jmir.6915] [Medline: 28615152]
54. Kassam I, Shin HD, Durocher K, Lo B, Shen N, Mehta R, et al. "I think it's something that we should lean in to": the use of OpenNotes in Canadian psychiatric care contexts by clinicians. Digit Health 2022;8:20552076221144106 [FREE Full text] [doi: 10.1177/20552076221144106] [Medline: 36532111]
55. Mayhew C, Strudwick G, Waddell J. Clinical nurse specialists' perceptions of a mental health patient portal. Clin Nurse Spec 2018;32(6):313-322. [doi: 10.1097/nur.0000000000000406]
56. Petersson J, Backman C. Off the record: the invisibility work of doctors in a patient-accessible electronic health record information service. Sociol Health Illn 2021 Jun;43(5):1270-1285. [doi: 10.1111/1467-9566.13294] [Medline: 34013985]
57. Erlingsdóttir G, Petersson L, Jonnergård K. A theoretical twist on the transparency of open notes: qualitative analysis of health care professionals' free-text answers. J Med Internet Res 2019 Sep 25;21(9):e14347 [FREE Full text] [doi: 10.2196/14347] [Medline: 31573905]
58. van Rijt AM, Hulter P, Weggelaar-Jansen AM, Ahaus K, Pluut B. Mental health care professionals' appraisal of patients' use of web-based access to their electronic health record: qualitative study. J Med Internet Res 2021 Aug 27;23(8):e28045 [FREE Full text] [doi: 10.2196/28045] [Medline: 34448705]
59. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
60. O'Neill S, Chimowitz H, Leveille S, Walker J. Embracing the new age of transparency: mental health patients reading their psychotherapy notes online. J Ment Health 2019 Oct;28(5):527-535. [doi: 10.1080/09638237.2019.1644490] [Medline: 31364902]
61. Pisciotta M, Denneson LM, Williams HB, Woods S, Tuepker A, Dobscha SK. Providing mental health care in the context of online mental health notes: advice from patients and mental health clinicians. J Ment Health 2019 Feb;28(1):64-70. [doi: 10.1080/09638237.2018.1521924] [Medline: 30468100]
62. Turvey CL, Fuhrmeister LA, Klein DM, Moeckli J, Howren MB, Chasco EE. Patient and provider experience of electronic patient portals and secure messaging in mental health treatment. Telemed J E Health 2022 Feb;28(2):189-198 [FREE Full text] [doi: 10.1089/tmj.2020.0395] [Medline: 33887164]
63. Jonnergård K, Petersson L, Erlingsdóttir G. Communicating the implementation of open notes to health care professionals: mixed methods study. JMIR Med Inform 2021 Aug 16;9(8):e22391 [FREE Full text] [doi: 10.2196/22391] [Medline: 34398794]
64. Åkerstedt US, Cajander Å, Moll J, Ålander T. On threats and violence for staff and patient accessible electronic health records. Cogent Psychol 2018 Sep 27;5(1):1518967. [doi: 10.1080/23311908.2018.1518967]
65. Petersson L, Erlingsdóttir G. Open notes in Swedish psychiatric care (part 2): survey among psychiatric care professionals. JMIR Ment Health 2018 Jun 21;5(2):e10521 [FREE Full text] [doi: 10.2196/10521] [Medline: 29929946]
66. Shin HD, Durocher K, Lo B, Chen S, Ma C, Wiljer D, et al. Impact of a mental health patient portal on patients' views of compassion: a mixed-methods study. BMC Digit Health 2023 Jan 24;1(1) [FREE Full text] [doi: 10.1186/s44247-022-00002-z]
67. Denneson LM, Chen JI, Pisciotta M, Tuepker A, Dobscha SK. Patients' positive and negative responses to reading mental health clinical notes online. Psychiatr Serv 2018 May 01;69(5):593-596 [FREE Full text] [doi: 10.1176/appi.ps.201700353] [Medline: 29493408]
68. Schwarz J, Meier-Diedrich E, Neumann K, Heinze M, Eisenmann Y, Thoma S. Reasons for acceptance or rejection of online record access among patients affected by a severe mental illness: mixed methods study. JMIR Ment Health 2024 Feb 05;11:e51126 [FREE Full text] [doi: 10.2196/51126] [Medline: 38315523]
69. Klein JW, Peacock S, Tsui JI, O'Neill SF, DesRoches CM, Elmore JG. Perceptions of primary care notes by patients with mental health diagnoses. Ann Fam Med 2018 Jul;16(4):343-345 [FREE Full text] [doi: 10.1370/afm.2287] [Medline: 29987083]
70. Blease C, Kharko A, Hägglund M, O'Neill S, Wachenheim D, Salmi L, et al. The benefits and harms of open notes in mental health: a Delphi survey of international experts. PLoS One 2021 Oct 13;16(10):e0258056 [FREE Full text] [doi: 10.1371/journal.pone.0258056] [Medline: 34644320]
71. Blease C, Torous J, Kharko A, DesRoches CM, Harcourt K, O'Neill S, et al. Preparing patients and clinicians for open notes in mental health: qualitative inquiry of international experts. JMIR Ment Health 2021 Apr 16;8(4):e27397 [FREE Full text] [doi: 10.2196/27397] [Medline: 33861202]

XSL•FO
RenderX

72. Bärkås A, Kharko A, Blease C, Cajander Å, Johansen Fagerlund A, Huvila I, et al. Errors, omissions, and offenses in the health record of mental health care patients: results from a nationwide survey in Sweden. J Med Internet Res 2023 Nov 03;25:e47841. [doi: 10.2196/47841] [Medline: 37921861]
73. Dobscha SK, Kenyon EA, Pisciotta MK, Niederhausen M, Woods S, Denneson LM. Impacts of a web-based course on mental health clinicians' attitudes and communication behaviors related to use of OpenNotes. Psychiatr Serv 2019 Jun 01;70(6):474-479. [doi: 10.1176/appi.ps.201800416] [Medline: 30890047]
74. Petersson L, Erlingsdóttir G. Open notes in Swedish psychiatric care (part 1): survey among psychiatric care professionals. JMIR Ment Health 2018 Feb 02;5(1):e11 [FREE Full text] [doi: 10.2196/mental.9140] [Medline: 29396386]
75. Gasteiger N, Fleming T, Day K. Converging perspectives of providers and student users on extending a patient portal into a university-based mental health service: a qualitative study. Internet Interv 2020 Mar;19:100304 [FREE Full text] [doi: 10.1016/j.invent.2020.100304] [Medline: 31993336]
76. Peck P, Torous J, Shanahan M, Fossa A, Greenberg W. Patient access to electronic psychiatric records: a pilot study. Health Policy Technol 2017 Sep;6(3):309-315. [doi: 10.1016/j.hlpt.2017.06.003]
77. Strudwick G, Booth RG, McLean D, Leung K, Rossetti S, McCann M, et al. Identifying indicators of meaningful patient portal use by psychiatric populations. Inform Health Soc Care 2020 Oct 01;45(4):396-409. [doi: 10.1080/17538157.2020.1776291] [Medline: 32603617]
78. Bärkås A, Scandurra I, Rexhepi H, Blease C, Cajander Å, Hägglund M. Patients' access to their psychiatric notes: current policies and practices in Sweden. Int J Environ Res Public Health 2021 Aug 30;18(17):9140 [FREE Full text] [doi: 10.3390/ijerph18179140] [Medline: 34501730]
79. Matthews EB, Savoy M, Paranjape A, Washington D, Hackney T, Galis D, et al. Acceptability of health information exchange and patient portal use in depression care among underrepresented patients. J Gen Intern Med 2022 Nov;37(15):3947-3955 [FREE Full text] [doi: 10.1007/s11606-022-07427-2] [Medline: 35132548]
80. Hochwarter S, Fehler G, Muente C, Eisenmann Y, Heinze M, Hibsch C, et al. Design of a patient-accessible electronic health record system in mental health. Stud Health Technol Inform 2022 May 25;294:583-584. [doi: 10.3233/SHTI220534] [Medline: 35612155]
81. Bärkås A, Scandurra I, Hägglund M. Analysis of voluntary user feedback of the Swedish National PAEHR Service. Stud Health Technol Inform 2019 Aug 21;264:1126-1130. [doi: 10.3233/SHTI190401] [Medline: 31438100]
82. Bärkås A, Hägglund M, Moll J, Cajander Å, Rexhepi H, Hörhammer I, et al. Patients' access to their psychiatric records - a comparison of four countries. Stud Health Technol Inform 2022 May 25;294:510-514. [doi: 10.3233/SHTI220511] [Medline: 35612132]
83. Hägglund M, Scandurra I. Usability of the Swedish accessible electronic health record: qualitative survey study. JMIR Hum Factors 2022 Jun 23;9(2):e37192 [FREE Full text] [doi: 10.2196/37192] [Medline: 35737444]
84. Etingen B, Hogan TP, Martinez RN, Shimada S, Stroupe K, Nazi K, et al. How do patients with mental health diagnoses use online patient portals? An observational analysis from the Veterans Health Administration. Adm Policy Ment Health 2019 Sep;46(5):596-608. [doi: 10.1007/s10488-019-00938-x] [Medline: 31065908]
85. Abel EA, Shimada SL, Wang K, Ramsey C, Skanderson M, Erdos J, et al. Dual use of a patient portal and clinical video telehealth by veterans with mental health diagnoses: retrospective, cross-sectional analysis. J Med Internet Res 2018 Nov 07;20(11):e11350 [FREE Full text] [doi: 10.2196/11350] [Medline: 30404771]
86. Shimada SL, Brandt CA, Feng H, McInnes DK, Rao SR, Rothendler JA, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. J Med Internet Res 2014 Dec 12;16(12):e272 [FREE Full text] [doi: 10.2196/jmir.3751] [Medline: 25498515]
87. Dobscha SK, Denneson LM, Pisciotta MK, Bourne DS, Chen JI, Philipps-Moses D, et al. Predictors of viewing progress notes among users of VA's electronic health portal who receive mental health care. JAMIA Open 2018 Jul;1(1):122-127 [FREE Full text] [doi: 10.1093/jamiaopen/ooy007] [Medline: 31984324]
88. Forchuk C, Rudnick A, Hoch J, Donelle L, Campbell R, Osaka W, et al. Mental health engagement network: innovating community-based mental healthcare. J Gen Pract 2014;02(01) [FREE Full text] [doi: 10.4172/2329-9126.1000143]
89. Eyasu T, Leung K, Strudwick G. Guiding improvements in user experience: results of a mental health patient portal user interface assessment. Stud Health Technol Inform 2019;257:110-114. [Medline: 30741182]
90. Robotham D, Mayhew M, Rose D, Wykes T. Electronic personal health records for people with severe mental illness; a feasibility study. BMC Psychiatry 2015 Aug 06;15:192 [FREE Full text] [doi: 10.1186/s12888-015-0558-y] [Medline: 26245805]
91. van den Heuvel SC, Meije D, Regeer EJ, Sinnema H, Riemersma RF, Kupka RW. The user experiences and clinical outcomes of an online personal health record to support self-management of bipolar disorder: a pretest-posttest pilot study. J Affect Disord 2018 Oct 01;238:261-268. [doi: 10.1016/j.jad.2018.05.069] [Medline: 29894931]
92. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
93. Kelly EL, Braslow JT, Brekke JS. Using electronic health records to enhance a peer health navigator intervention: a randomized pilot test for individuals with serious mental illness and housing instability. Community Ment Health J 2018 Nov 3;54(8):1172-1179 [FREE Full text] [doi: 10.1007/s10597-018-0282-4] [Medline: 29725878]

XSL•FO
RenderX

94. Pecina J, North F, Williams MD, Angstman KB. Use of an on-line patient portal in a depression collaborative care management program. J Affect Disord 2017 Jan 15;208:1-5. [doi: 10.1016/j.jad.2016.08.034] [Medline: 27736688]
95. Blease C, Dong Z, Torous J, Walker J, Hägglund M, DesRoches CM. Association of patients reading clinical notes with perception of medication adherence among persons with serious mental illness. JAMA Netw Open 2021 Mar 01;4(3):e212823 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2823] [Medline: 33760088]
96. Druss BG, Ji X, Glick G, von Esenwein SA. Randomized trial of an electronic personal health record for patients with serious mental illnesses. Am J Psychiatry 2014 Mar;171(3):360-368. [doi: 10.1176/appi.ajp.2013.13070913] [Medline: 24435025]
97. Druss BG, Li J, Tapscott S, Lally CA. Randomized trial of a mobile personal health record for behavioral health homes. Psychiatr Serv 2020 Aug 01;71(8):803-809 [FREE Full text] [doi: 10.1176/appi.ps.201900381] [Medline: 32362226]
98. Forchuk C, Reiss JP, O'Regan T, Ethridge P, Donelle L, Rudnick A. Client perceptions of the mental health engagement network: a qualitative analysis of an electronic personal health record. BMC Psychiatry 2015 Oct 14;15:250 [FREE Full text] [doi: 10.1186/s12888-015-0614-7] [Medline: 26467210]
99. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
100. Tsai J, Rosenheck RA. Use of the internet and an online personal health record system by US veterans: comparison of Veterans Affairs mental health service users and other veterans nationally. J Am Med Inform Assoc 2012;19(6):1089-1094 [FREE Full text] [doi: 10.1136/amiajnl-2012-000971] [Medline: 22847305]
101. Connolly SL, Etingen B, Shimada SL, Hogan TP, Nazi K, Stroupe K, et al. Patient portal use among veterans with depression: associations with symptom severity and demographic characteristics. J Affect Disord 2020 Oct 01;275:255-259 [FREE Full text] [doi: 10.1016/j.jad.2020.06.073] [Medline: 32734917]
102. Kim EH, Kim Y. Digital divide: use of electronic personal health record by different population groups. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:1759-1762. [doi: 10.1109/IEMBS.2010.5626732] [Medline: 21096415]
103. Raps SJ, Chen D, Bakken S, Caban J, Engler MB. Baseline eHealth behaviors of service members: a retrospective, cross-sectional analysis of patient portal use before the pandemic. Mil Med 2023 Jul 22;188(7-8):e2598-e2605 [FREE Full text] [doi: 10.1093/milmed/usac168] [Medline: 35766515]
104. Leung K, Clark C, Sakal M, Friesen M, Strudwick G. Patient and family member readiness, needs, and perceptions of a mental health patient portal: a mixed methods study. Stud Health Technol Inform 2019;257:266-270. [Medline: 30741207]
105. Weisner CM, Chi FW, Lu Y, Ross TB, Wood SB, Hinman A, et al. Examination of the effects of an intervention aiming to link patients receiving addiction treatment with health care: the LINKAGE clinical trial. JAMA Psychiatry 2016 Aug 01;73(8):804-814 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0970] [Medline: 27332703]
106. Denneson LM, Pisciotta M, Hooker ER, Trevino A, Dobscha SK. Impacts of a web-based educational program for veterans who read their mental health notes online. J Am Med Inform Assoc 2019 Jan 01;26(1):3-8 [FREE Full text] [doi: 10.1093/jamia/ocy134] [Medline: 30445648]
107. Ennis L, Robotham D, Denis M, Pandit N, Newton D, Rose D, et al. Collaborative development of an electronic Personal Health Record for people with severe and enduring mental health problems. BMC Psychiatry 2014 Nov 18;14:305 [FREE Full text] [doi: 10.1186/s12888-014-0305-9] [Medline: 25403285]
108. Hilton JF, Barkoff L, Chang O, Halperin L, Ratanawongsa N, Sarkar U, et al. A cross-sectional study of barriers to personal health record use among patients attending a safety-net clinic. PLoS One 2012;7(2):e31888 [FREE Full text] [doi: 10.1371/journal.pone.0031888] [Medline: 22363761]
109. Forchuk C, Rudnick A, Hoch J, Donelle L, Godin M, Osoka W, et al. Mental Health Engagement Network (MHEN). Int J Adv Life Sci 2013;5(1&2):1-10 [FREE Full text]
110. Forchuk C, Donelle L, Ethridge P, Warner L. Client perceptions of the mental health engagement network: a secondary analysis of an intervention using smartphones and desktop devices for individuals experiencing mood or psychotic disorders in Canada. JMIR Ment Health 2015;2(1):e1 [FREE Full text] [doi: 10.2196/mental.3926] [Medline: 26543906]
111. Gray LC, Berg K, Fries BE, Henrard JC, Hirdes JP, Steel K, et al. Sharing clinical information across care settings: the birth of an integrated assessment system. BMC Health Serv Res 2009 Apr 29;9(1):71 [FREE Full text] [doi: 10.1186/1472-6963-9-71] [Medline: 19402891]
112. Salomon RM, Blackford JU, Rosenbloom ST, Seidel S, Clayton EW, Dilts DM, et al. Openness of patients' reporting with use of electronic records: psychiatric clinicians' views. J Am Med Inform Assoc 2010;17(1):54-60 [FREE Full text] [doi: 10.1197/jamia.M3341] [Medline: 20064802]
113. Horak J, Novak A, Voumik LC. Healthcare generative artificial intelligence tools in medical diagnosis, treatment, and prognosis. Contemp Read Law Soc Justice 2023;15(1):81-98. [doi: 10.22381/CRLSJ15120235]
114. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. J Biomed Inform 2017 Sep;73:14-29 [FREE Full text] [doi: 10.1016/j.jbi.2017.07.012] [Medline: 28729030]

XSL•FO
RenderX

115. Emerencia A, van der Krieke L, Petkov N, Aiello M. Assessing schizophrenia with an interoperable architecture. In: Proceedings of the First International Workshop on Managing Interoperability and Complexity in Health Systems. 2011 Presented at: MIXHS '11; October 28, 2011; Glasgow, UK. [doi: 10.1145/2064747.2064766]
116. Priebe S, Mccabe R. Therapeutic relationships in psychiatry: the basis of therapy or therapy in itself? Int Rev Psychiatry 2008 Dec;20(6):521-526. [doi: 10.1080/09540260802565257] [Medline: 19085408]
117. Farrelly S, Lester H. Therapeutic relationships between mental health service users with psychotic disorders and their clinicians: a critical interpretive synthesis. Health Soc Care Community 2014 Sep;22(5):449-460. [doi: 10.1111/hsc.12090] [Medline: 24428515]
118. Woltmann EM, Whitley R. Shared decision making in public mental health care: perspectives from consumers living with severe mental illness. Psychiatr Rehabil J 2010;34(1):29-36. [doi: 10.2975/34.1.2010.29.36] [Medline: 20615842]
119. Felton A, Repper J, Avis M. Therapeutic relationships, risk, and mental health practice. Int J Ment Health Nurs 2018 Jun;27(3):1137-1148. [doi: 10.1111/inm.12430] [Medline: 29280259]
120. Ahmed N, Barlow S, Reynolds L, Drey N, Begum F, Tuudah E, et al. Mental health professionals' perceived barriers and enablers to shared decision-making in risk assessment and risk management: a qualitative systematic review. BMC Psychiatry 2021 Nov 25;21(1):594 [FREE Full text] [doi: 10.1186/s12888-021-03304-0] [Medline: 34823487]
121. Muir-Cochrane E, Gerace A, Mosel K, O'Kane D, Barkway P, Curren D, et al. Managing risk: clinical decision-making in mental health services. Issues Ment Health Nurs 2011;32(12):726-734. [doi: 10.3109/01612840.2011.603880] [Medline: 22077745]
122. Deering K, Pawson C, Summers N, Williams J. Patient perspectives of helpful risk management practices within mental health services. A mixed studies systematic review of primary research. J Psychiatr Ment Health Nurs 2019 Jun;26(5-6):185-197. [doi: 10.1111/jpm.12521] [Medline: 31046161]
123. Lydahl D, Britten N, Wolf A, Naldemirci Ö, Lloyd H, Heckemann B. Exploring documentation in person-centred care: a content analysis of care plans. Int J Older People Nurs 2022 Sep;17(5):e12461 [FREE Full text] [doi: 10.1111/opn.12461] [Medline: 35393772]
124. Stanhope V, Ingoglia C, Schmelter B, Marcus SC. Impact of person-centered planning and collaborative documentation on treatment adherence. Psychiatr Serv 2013 Jan;64(1):76-79. [doi: 10.1176/appi.ps.201100489] [Medline: 23280459]
125. Maniss S, Pruit A. Collaborative documentation for behavioral healthcare providers: an emerging practice. J Human Serv Train Res Pract 2018;3(1) [FREE Full text]
126. Liu SK, Osborn AE, Bell S, Mecchella JN, Hort S, Batsis JA. Patient characteristics and utilization of an online patient portal in a rural academic general internal medicine practice. BMC Med Inform Decis Mak 2022 Feb 16;22(1):42 [FREE Full text] [doi: 10.1186/s12911-022-01778-w] [Medline: 35172805]
127. El Yaman N, Zeitoun J, Diab R, Mdaihly M, Diab R, Kobeissi L, et al. Utilization of patient portals: a cross-sectional study investigating associations with mobile app quality. BMC Med Inform Decis Mak 2023 Sep 05;23(1):177 [FREE Full text] [doi: 10.1186/s12911-023-02252-x] [Medline: 37670277]
128. Hong YA, Jiang S, Liu PL. Use of patient portals of electronic health records remains low from 2014 to 2018: results from a national survey and policy implications. Am J Health Promot 2020 Jul;34(6):677-680. [doi: 10.1177/0890117119900591] [Medline: 32030989]
129. Singh S, Polavarapu M, Arsene C. Changes in patient portal adoption due to the emergence of COVID-19 pandemic. Inform Health Soc Care 2023 Apr 03;48(2):125-138. [doi: 10.1080/17538157.2022.2070069] [Medline: 35473512]
130. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
131. Koopman RJ, Mehr DR, Petroski G, Campbell JD, Erdelez S, Headrick L. Patient readiness to use internet health resources. The Agency for Healthcare Research and Quality. 2014. URL: https://digital.ahrq.gov/sites/default/files/docs/publication/k08hs017948-koopman-final-report-2014.pdf [accessed 2023-10-29]
132. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014 Jan 28;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
133. Tsai CH, Eghdam A, Davoody N, Wright G, Flowerday S, Koch S. Effects of electronic health record implementation and barriers to adoption and use: a scoping review and qualitative analysis of the content. Life (Basel) 2020 Dec 04;10(12):327 [FREE Full text] [doi: 10.3390/life10120327] [Medline: 33291615]
134. Ajami S, Bagheri-Tadi T. Barriers for adopting electronic health records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134 [FREE Full text] [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]
135. Gillard S, Holley J. Peer workers in mental health services: literature overview. Adv Psychiatr Treat 2018 Jan 02;20(4):286-292. [doi: 10.1192/apt.bp.113.011940]
136. Miller KE, Kuhn E, Yu J, Owen JE, Jaworski BK, Taylor K, et al. Use and perceptions of mobile apps for patients among VA primary care mental and behavioral health providers. Prof Psychol Res Pract 2019 Jun;50(3):204-209. [doi: 10.1037/pro0000229]
137. Waid J, Halpin K, Donaldson R. Mental health service navigation: a scoping review of programmatic features and research evidence. Soc Work Mental Health 2021 Jan 11;19(1):60-79. [doi: 10.1080/15332985.2020.1870646]

XSL•FO
RenderX

138. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141-146 [FREE Full text] [doi: 10.1038/ejhg.2014.71] [Medline: 24801761]
139. Nominated representatives. Australian Digital Health Agency. URL: https://www.digitalhealth.gov.au/initiatives-and-programs/my-health-record/manage-your-record/privacy-and-access/nominated-representatives [accessed 2023-10-29]
140. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 04;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]
141. Richwine C, Johnson C, Patel V. Disparities in patient portal access and the role of providers in encouraging access and use. J Am Med Inform Assoc 2023 Jan 18;30(2):308-317 [FREE Full text] [doi: 10.1093/jamia/ocac227] [Medline: 36451262]
142. Antonio MG, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? A scoping review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):871-883 [FREE Full text] [doi: 10.1093/jamia/ocz054] [Medline: 31066893]
143. Wynia M, Dunn K. Dreams and nightmares: practical and ethical issues for patients and physicians using personal health records. J Law Med Ethics 2010;38(1):64-73. [doi: 10.1111/j.1748-720X.2010.00467.x] [Medline: 20446985]
144. Lucivero F. Lessons about so-called "difficult" patients from the UK controversy over patient access to electronic health records. AMA J Ethics 2017 Apr 01;19(4):374-380 [FREE Full text] [doi: 10.1001/journalofethics.2017.19.4.stas1-1704] [Medline: 28430572]
145. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]

Abbreviations

EHR: electronic health record
ePHR: electronic personal health record
HCP: health care professional
PAEHR: patient-accessible electronic health record
pEHR: personal electronic health record
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by T de Azevedo Cardoso; submitted 29.11.23; peer-reviewed by GD Karampatakis, MJ Tan; comments to author 08.02.24; revised version received 09.04.24; accepted 25.11.24; published 07.02.25.

Please cite as:
Kariotis T, Prictor M, Gray K, Chang S
Patient-Accessible Electronic Health Records and Information Practices in Mental Health Care Contexts: Scoping Review
J Med Internet Res 2025;27:e54973
URL: https://www.jmir.org/2025/1/e54973
doi:10.2196/54973
PMID:39918859

©Timothy Kariotis, Megan Prictor, Kathleen Gray, Shanton Chang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Understanding “Alert Fatigue” in Primary Care: Qualitative Systematic Review of General Practitioners Attitudes and Experiences of Clinical Alerts, Prompts, and Reminders

Illin Gani[1*], MPH; Ian Litchfield[1*], PhD; David Shukla[1], MBChB, MRCGP; Gayathri Delanerolle[1], PhD; Neil Cockburn[1], MBChB; Anna Pathmanathan[2], MPH

[1]Department of Health Sciences, University of Birmingham, Birmingham, United Kingdom

[2]Population Health Sciences, Centre for Academic Primary Care, University of Bristol, Bristol, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Ian Litchfield, PhD
Department of Health Sciences
University of Birmingham
Room 134
Murray Learning Centre
Birmingham, B15 2TT
United Kingdom
Phone: 44 01214146006
Email: i.litchfield@bham.ac.uk

Abstract

Background: The consistency and quality of care in modern primary care are supported by various clinical reminders (CRs), which include “alerts” describing the consequences of certain decisions and “prompts” that remind users to perform tasks promoting desirable clinical behaviors. However, not all CRs are acted upon, and many are disregarded by general practitioners (GPs), a chronic issue commonly referred to as “alert fatigue.” This phenomenon has significant implications for the safety and quality of care, GP burnout, and broader medicolegal consequences. Research on mitigating alert fatigue and optimizing the use of CRs remains limited. This review offers much-needed insight into GP attitudes toward the deployment, design, and overall effectiveness of CRs.

Objective: This systematic review aims to synthesize current qualitative research on GPs’ attitudes toward CRs, enabling an exploration of the interacting influences on the occurrence of alert fatigue in GPs, including the deployment, design, and perceived efficacy of CRs.

Methods: A systematic literature search was conducted across the Health Technology Assessment database, MEDLINE, MEDLINE In-Process, Embase, CINAHL, Conference Proceedings Citation Index, PsycINFO, and OpenGrey. The search focused on primary qualitative and mixed methods research conducted in general or family practice, specifically exploring GPs’ experiences with CRs. All databases were searched from inception to December 31, 2023. To ensure structured and practicable findings, we used a directed content analysis of the data, guided by the 7 domains of the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework, including domains related to Technology, Adopter attitudes, and Organization.

Results: A total of 9 studies were included, and the findings were organized within the 7 domains. Regarding Condition and Value Proposition, GPs viewed CRs as an effective way to maintain or improve the safety and quality of care they provide. When considering the attributes of the Technology, the efficacy of CRs was linked to their frequency, presentation, and the accuracy of their content. Within Adopters, concerns were raised about the accuracy of CRs and the risk that their use could diminish the value of GP experience and contextual understanding. From an Organization perspective, the need for training on the use and benefits of CRs was highlighted. Finally, in the context of the Wider system and their Embedding Over Time, suggestions included sharing best practices for CR use and involving GPs in their design.

Conclusions: While GPs acknowledged that CRs, when used optimally, can enhance patient safety and quality of care, several concerns emerged regarding their design, content accuracy, and lack of contextual nuance. Suggestions to improve CR adherence included providing coherent training, enhancing their design, and incorporating more personalized content.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42016029418; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=29418
International Registered Report Identifier (IRRID): RR2-10.1186/s13643-017-0627-z

(J Med Internet Res 2025;27:e62763) doi:10.2196/62763

KEYWORDS

primary care; general practitioners; alert fatigue; computer decision support systems; fatigue; qualitative; systematic review; quality of care; clinical behaviors; behaviors; database; family practice; family; algorithm; patient safety; patient

Introduction

Primary care is facing unprecedented pressure across the globe, caught between growing patient demand, issues of workforce recruitment and retention, and increasingly complex options for treatment [1,2]. To improve the consistency and quality of care in this dynamic and highly pressured environment, health systems worldwide have introduced guidelines for care informed by the latest clinical evidence [3,4]. These guidelines are promoted and supported in high-income countries through a range of clinical reminders (CRs) that are presented to general practitioners (GPs) via their clinical management systems and consist of "alerts" that remind them of the consequences of making a certain decision, and "prompts" that remind the user to perform a task that promotes desirable clinical behaviors [5-10]. These alerts can relate to potential diagnoses, drug-drug interactions, and allergies, and prompts linked to patient-specific needs, further information, and recommendations for high-quality care (see Figure 1; see also [11]).

Figure 1. Types of clinical reminders and alerts used in clinical management systems in primary care (Cecil et al [11]).

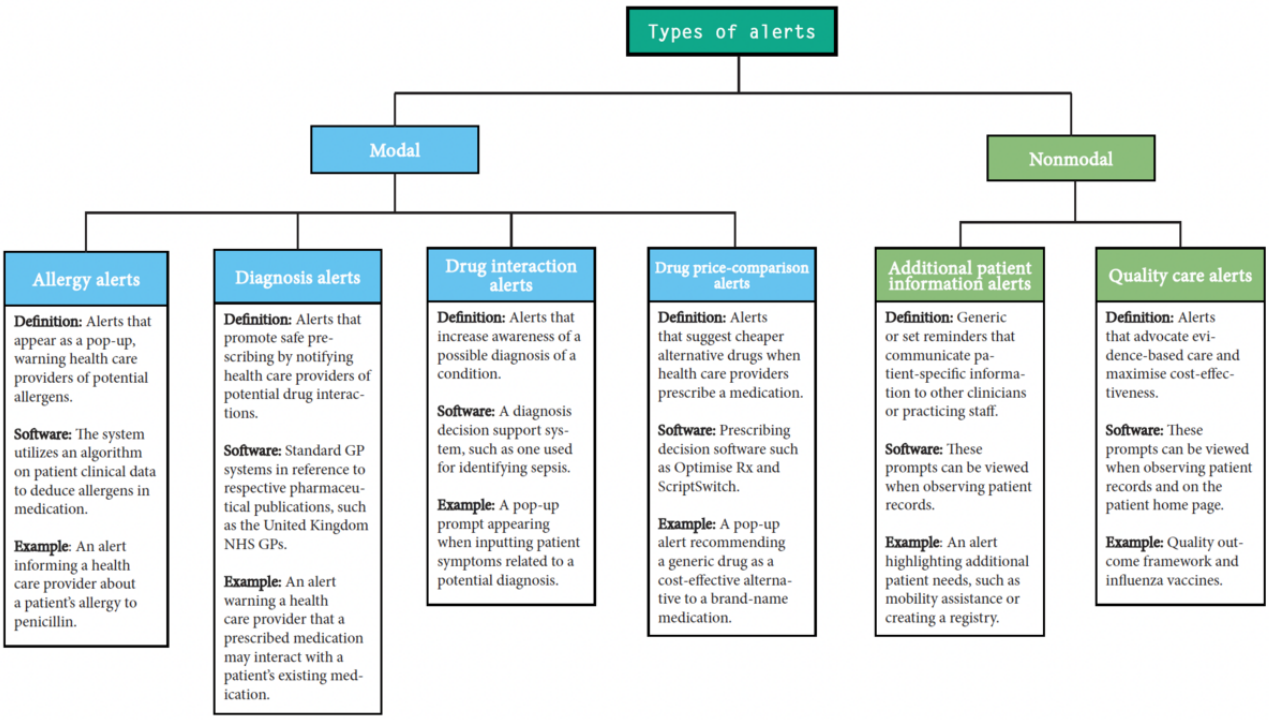

The growing sophistication of medical knowledge and the electronic systems that manage patient data and inform clinical decision-making have led to a steady increase in the activation of CRs at the point of care and the cognitive load placed on GPs [12-15]. Not all CRs are acted upon, and some may justifiably be considered irrelevant or inapplicable [16,17], but increasingly GPs are disregarding pertinent or urgent CRs [18-20], a chronic negligence commonly referred to as "alert fatigue" [21].

Despite the implications of alert fatigue for the safety and quality of care, GP burnout, and the medico-legal consequences for individuals and their organizations, the proliferation of CRs in primary care is reliant on evidence limited to their effects on a specific condition [22,23]. What remains underexplored is how their use might be optimized in the context of the cumulative effect of multiple CRs on a single health care provider [19,20,22,24]. There is then a need to better understand GP attitudes toward the deployment, design, and overall effectiveness of CRs, as well as broader implications for their impact on patient-clinician communication [22,25]. The systematic review presented here aims to address this by synthesizing current qualitative research on GPs' attitudes toward CRs describing the findings within the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework [26]. This framework was designed to study unfolding digital health technology programs and so provides the opportunity for offering structured insight into the future design, use, and positioning of CRs in primary care.

XSL•FO
RenderX

Methods

Applying the NASSS Framework to Digital Health Implementation

The review used best practices in systematic review methodology and directed content analysis [27,28]. We describe the results through the lens of the NASSS framework developed to support, guide, and monitor the implementation of digital technologies in health and social care [26,29,30]. The framework uses 7 domains to describe various aspects of the technology, end users, organizational structure, and broader policy-driven context [26]. The domains are defined alongside their potential impact on implementation in Table 1.

XSL•FO
RenderX

Table 1. The NASSS[a] framework: domains, definitions, and influences on implementation (adapted from [26,29-31]).

Domain	Definition	Influences on implementation
Condition	The condition(s) for which the innovation or technology has been designed. These can be physical, mental, or psychosocial in nature.	• The complexity of the condition. • Its metabolic volatility, association with comorbidities, and impact on cognitive function [32].
Technology	The technology/ies or other innovation that is/are being introduced. Include(s) both hardware and software and can include a novel protocol or pathway—or some combination of these.	• Its material properties, functionality, dependability, and speed [33]. • The knowledge needed to use them [34]. • The knowledge generated by the technology [35]. • The supply model and the relationship with the care provider [36]. • The ownership of intellectual property [37].
Value Proposition	The value (financial or otherwise) that the new technology and care model generates. This includes, for commercial stakeholders, the return on investment; for patients, improvements in comfort or quality of life; for health care organizations, improvements in quality, safety, inclusivity, and efficiency of the care delivered.	• Provision of value to a range of stakeholders, suppliers, patients, the health care system, and taxpayers or insurers. • Formulating a credible business plan where efficacy or cost-effectiveness studies are unavailable or contested [38,39].
Adopters	The intended users of the technology or other innovation. This includes patients/laypersons, professionals, administrative, and support staff.	• Acceptability to service users and their families/carers [40]. • Attitudes toward new and emerging technologies [41]. • Influence of sociocultural factors such as poverty or social exclusion [42]. • Acceptability to staff. • Impact on roles, professional traditions, and codes of conduct [26].
Organizations	The cultural and organizational characteristics of the organizations involved. This includes structure, capacity, and capability to adopt new ways of working, as well as resources of staff and infrastructure.	• The organization's general capacity to innovate [43]. • Readiness for this particular technology [26]. • The decisions around funding the intervention (including the presence of interorganizational agreements or speculative cross-system savings in the funding decision [44]). • The extent of the change needed, including the potential disruption to existing routines [45]. • The work required in implementation, including staff engagement, fidelity of implementation, and evaluation [46].
Wider System	The national and local context for the introduction of the technology or program	• The impact of national and local policies and objectives [47]. • The support of regulatory or professional bodies [48]. • Sociocultural factors including public perceptions of the technology [49]. • The presence of interorganizational networking and collaborative initiatives in supporting implementation [50].
Embedding Over Time	The key changes and uncertainties expected to affect the integration of the technology over the next 3-5 years.	• The ability of the technology to adapt to changing contexts [51]. • The resilience and cultural stability of the organizations involved [52].

[a]NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability.

Search Strategy

The databases searched were the Health Technology Assessment Database, MEDLINE, MEDLINE in Process, Embase, CINAHL, Conference Proceedings Citation Index, PsycINFO, and OpenGrey, from January 1, 1960, to December 31, 2023, a date range chosen to reflect the introduction of electronic or digitally supported health care. The search was conducted by the author IG in March 2024 and no language or location limits were applied. Medical Subject Headings (MeSH) terms and free-text words were used against the Sample, Phenomenon of Interest, Design, Evaluation, Research type framework [53] to

XSL•FO
RenderX

collect relevant studies from the perspective of GPs in a primary care setting (see Multimedia Appendix 1). The search strategy is described in the PROSPERO registration and the previously published protocol [54]. The EndNote (Clarivate Plc) bibliographic software was used to organize all retrieved studies and remove any duplicates.

Eligibility Criteria

Studies included were primary qualitative research conducted in primary care settings and published in English. The research had to have a focus on general or family practice and be qualitative in nature because we were specifically interested in GPs' experience with and perspectives of CRs (see Multimedia Appendix 2 for detailed eligibility criteria).

Study Selection and Data Extraction

We used the recommended full dual review approach, that is, 2 independent reviewers (IG and AP) screened the title, abstract, and then the full texts, as this has been shown to increase the number of eligible papers identified and more broadly minimize error in the selection process. The data were managed using an Excel spreadsheet (Microsoft Corporation), with any disagreements on the inclusion of a study consensually resolved through discussion with a third reviewer (IL). A data extraction form was designed to summarize study characteristics. The data extracted included information on study outcomes, study design, participants, and key findings.

Quality Appraisal

The quality and bias of included papers were determined by the Critical Appraisal Skills Programme checklist for qualitative studies and the Mixed Methods Appraisal Tool for mixed method studies [55]. Although the Mixed Methods Appraisal Tool cautions against the approach of labeling studies at low, medium, and high risk of bias, we used a scoring system to determine this labeling, for ease of interpretation between the 2 different tools. The quality appraisal was undertaken independently by 2 authors (IL and IG) who between them agreed on the final value.

Data Synthesis

The data were coded and managed via NVivo12 (Lumivero, LLC). We used a directed content analysis to populate the NASSS framework, based on the unconstrained matrix approach. This allowed the development and inclusion of constructs directly relevant to emergent themes within the established framework [56,57]. The analysis was conducted by IG and IL who both independently analyzed a sample of the same 3 papers. They then met to discuss their analysis and any discrepancies in the allocation of data before IL analyzed the remainder of the studies, with the final analysis and data allocation consensually agreed upon by all authors. In completing the analysis, we merged the domains of "Wider system" and "Embedding over time" due to the concurrence that arose in the data.

Results

Study Characteristics

The screening and study selection are outlined in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 2; also see Multimedia Appendix 3). The study selection and data extraction templates can be found in Multimedia Appendix 4. Full-text studies were excluded because of duplication, or lack of detail on the care provider, setting, or digital tool (see Multimedia Appendix 5 for a detailed description). A total of 9 studies were included, 5 from the United Kingdom [58-62] and 1 each from Australia [63], Belgium [64], the Netherlands [65], and Norway [66]. Seven of these were considered high quality [58-62,65,66], 1 medium [64], and 1 low [63]. The papers were published between 2003 and 2021. The key characteristics of included studies and findings are summarized in Table 2. This includes the quality assessment (see Multimedia Appendix 6 for a detailed description). The directed content analysis enabled the mapping of the data within the NASSS framework. Within each domain, we determined a series of emergent constructs relevant to the implementation of CRs in primary care. These findings are summarized in Table 3 and presented below within each domain of the framework alongside exemplar quotes taken from the studies identified.

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the systematic review.

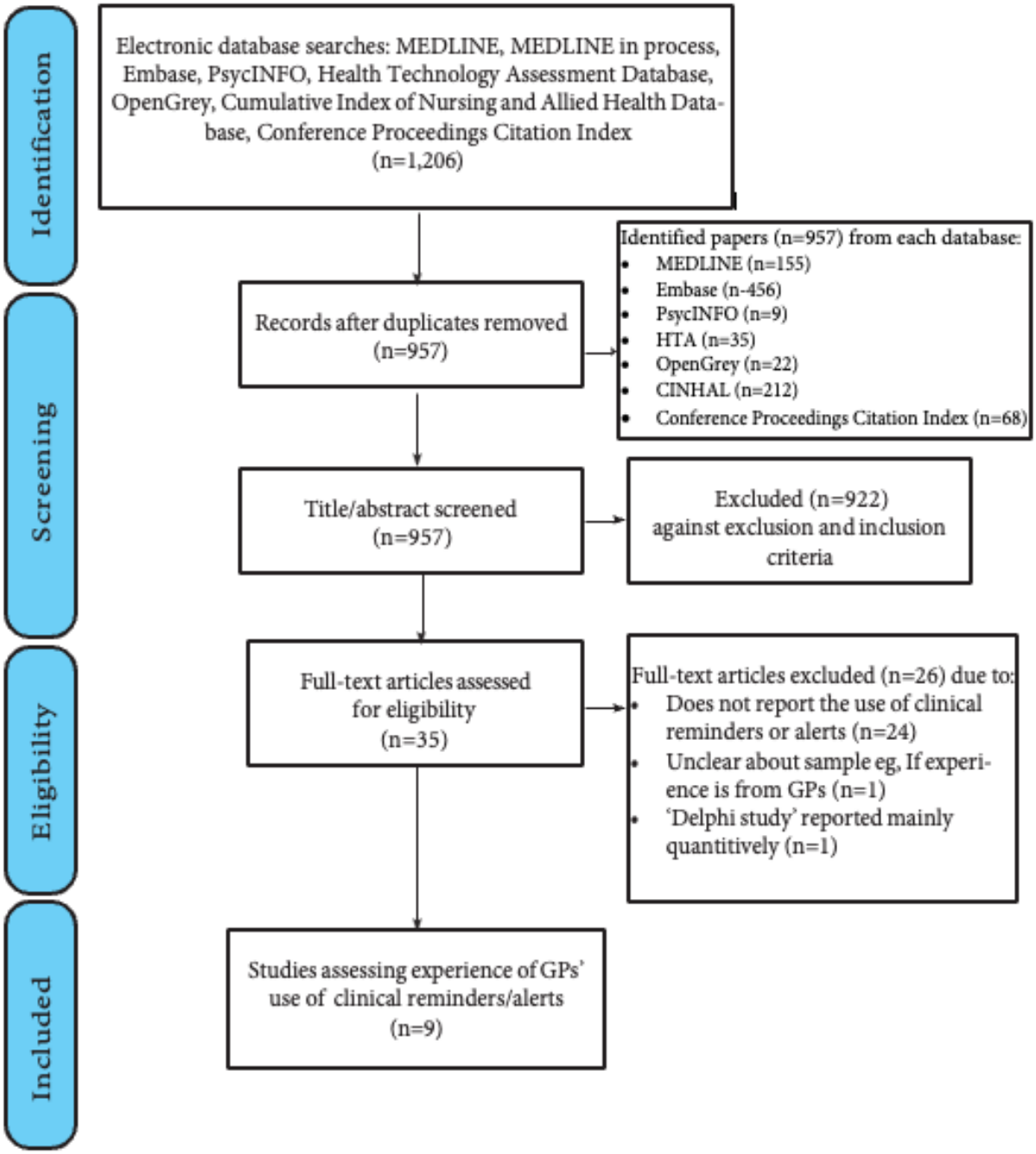

XSL•FO
RenderX

Table 2. Key characteristics and findings of identified studies.

Reference	Country of study	Study outcomes	Study design (quality of study)	Purpose of CR[a,b]	Participants/setting	Key findings
Ahearn and Kerr [63]	Australia	To determine GPs'[c] perceptions of strengths and weaknesses of prescription systems	Focus groups (low)	Medication	14 GPs; 1 focus group in a rural area and 2 focus groups in an urban area	• CRs affected prescribing behavior. • GPs recognized the need for training. • Considered helpful features of clinical decision support systems.
Bindels et al [65]	The Netherlands	To investigate the experiences of GPs with the "GRIF system" (automated test ordering combined with a feedback system)	Group discussion and interview (high)	Diagnosis	24 GPs; mean age 49 (SD 4.9) years, mean experience 19 (SD 7) years; males comprised 75%	• The system increased attention to and promotion of guidelines.
Christensen and Grimsmo [66]	Norway	To study primary care physicians' experiences of electronic prescription reminders	Focus groups; observations of doctor-patient encounters; and National Questionnaire Survey (high)	Medication	24 GPs (15 females + 9 males)	• The ability to present relevant patient information and medical knowledge was appreciated.
Dikomitis et al [61]	United Kingdom	To obtain views from GPs who piloted the electronic risk assessment tools	Telephone interviews (high)	Diagnosis	23 GPs (13 males + 10 females)	• Key issues over the integration of CRs into existing digital platforms.
Ford et al [59]	United Kingdom	To understand the features of individual clinical decision support systems that were barriers and facilitators to their use	Semistructured interview (high)	Medication	11 GPs (7 males + 4 females); age group 30-39 years, n=8; age group 40-49 years, n=2; and age group 50+ years, n=1	• Issues emerged relating to the accuracy of CRs and their usability.
Jeffries et al [58]	United Kingdom	To understand the factors that influenced clinical decision support system implementation	Semistructured interviews (high)	Medication	14 GPs across 4 regions in the United Kingdom	• There was a shared recognition of the need to use CRs and their potential to support patient care but felt CRs needed to be more relevant.
Heselmans et al [64]	Belgium	To assess users' perceptions toward the EBMeDS system (a computer decision support system provided by EBPNet, a national computerized point-of-care information service in Belgium)	Survey (medium)	Management, medication, and diagnosis	Survey issued to 334 (Dutch-speaking) GPs	• Reminders considered too general and inflexible, with questions over their reliability.
Holt et al [60]	United Kingdom	To investigate the implementation of the CRs in UK general practice	Semistructured interview (high)	Diagnosis	7 GPs (no characteristics described from the sample)	• The advantages of CRs were recognized but with caveats over their specificity.

XSL•FO
RenderX

Reference	Country of study	Study outcomes	Study design (quality of study)	Purpose of CR[a,b]	Participants/setting	Key findings
Slight et al [62]	United Kingdom	To examine the causes of prescribing and monitoring errors in general practice	Semistructured interviews and focus groups (high)	Medication	33 GPs participated in semistructured interviews and 46 contributed in focus groups	• The need for prior training and improved information technology infrastructure and hardware.

[a]CR: clinical reminder.

[b]Medication, diagnosis, or ongoing management.

[c]GP: general practitioner.

XSL•FO
RenderX

Table 3. Summary of findings by NASSS[a] domain.

NASSS domain and emergent constructs		Definition	Summary of findings (and review papers)
Condition			
	Medication	• To support safe prescribing and medication adherence.	• Issues accommodated by CRs[b] included drug-drug interactions, drug allergies, and dose-related recommendations [58,59,62-64,66].
	Diagnosis	• To support prompt and accurate diagnosis of a range of chronic and acute conditions.	• Diagnostic use included both acute [60,61] and chronic conditions [64,65].
	Ongoing management	• Providing reminders and alerts for the safe and successful long-term management of (chronically ill) patients.	• CRs included references to reviews of patient medication, symptoms, and testing regimes [64].
Technology			
	Frequency and content of CRs	• The number of CRs and the relevance and accuracy of their content.	• The overabundance of CRs was described [59,61,63,64]. • GPs[c] described a preference for accompanying information that aided comprehension [59].
	Design and integration of CRs	• The presentation and readability of CRs and their integration with existing clinical dashboards.	• GPs described how CRs requiring a response disrupted their workflow [59,63]. • A preference was expressed for CRs with clear color-coded designations of risk or otherwise readily understandable formatting [59]. Dropdown boxes were seen as overly complicated [62].
	Digital infrastructure	• The capacity, capability, and compatibility of hardware and software systems and the interoperability of data.	• The quality of connectivity and the capacity of practice software systems impacted the timely appearance of alerts, frustrating clinicians and increasing the risk of errors [59,60,62,66].
Value proposition			
	Overall quality and safety	• CRs capable of presenting timely prompts with the necessary actions.	• CRs were valued for their ability to provoke a decision [58,61,63], particularly in busy environments [59,60]. Their ability to act as educational tools was mooted [58]. • The constant interruption of tasks by CRs was considered a risk to patient safety [59].
	Self-efficacy	• CRs can provide reassurance for clinicians that they have not missed anything significant.	• The increasing complexity of primary care meant that clinicians welcomed the additional support [59,61,63].
Adopter			

XSL•FO
RenderX

NASSS domain and emergent constructs		Definition	Summary of findings (and review papers)
	Engagement, and compliance in CRs	• The degree to which clinicians follow the recommendations of CRs and the factors that influence this decision.	• There were concerns over the accuracy and relevance of CRs, with a need for more transparent evidence of their efficacy, as well as self-doubt if the CR was ignored [59,63].
	Digital literacy and acceptability	• Ability to understand, use, and incorporate digital technologies in health care.	• Recent cohorts of GPs were more familiar and trusting of digitalization [59]. • There were concerns that training, experience, and context would be subjugated to computer-generated recommendations [59,61].
	Patient-clinician relationship	• The impact on patient interactions, communication, and consultation behaviors.	• CRs could disrupt consultations and inhibit patient communication [61,63].
Organization			
	Training in the purpose and use of CRs	• Importance of providing adequate training and ensuring clinicians understand the purpose of CRs.	• The importance of training was recognized when introducing CRs [58-61,63].
	Practice capacity and workflow	• The capacity, workflow, and protocols of practice organizations.	• Shortening consultation times precluded GPs from interacting with many of the CRs generated [58,60].
Wider system adapting and embedding over time			
	Adaptation over time	• The ability of CRs to incorporate technological advances and reflect the evolving needs and preferences of patients and clinicians.	• The need for greater and earlier engagement of software developers with GPs was described [58,63,65]. • A number of suggestions for future CR design were shared across studies including better integration with existing software systems, patient-specific content, clear graphical formats, and the ability to customize frequency and content [59-61,63,67].
	Systemwide collaboration	• Unified approaches to the long-term introduction of technology including procurement, infrastructure, and shared learning.	• There was a recognized need for improved interoperability and shared data between settings and organizations [63,66] and for continued learning from peer experience was voiced [59].

[a]NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability.

[b]CR: clinical reminder.

[c]GP: general practitioner.

Condition

The studies identified described how CRs were used to optimize care and manage patient safety in 3 key areas: medication [58,59,62-64,66]; diagnosis, both broadly [64,65] and for specific conditions, namely, atrial fibrillation and stroke [60], and cancer [61]. CRs were also used in the ongoing management of chronically ill patients, providing reminders of reviews or tests [64].

Technology

Challenges in Technological Aspects of CRs: Alerts, Design, and Infrastructure

Seven studies described issues relating to various technological aspects of CRs, specifically the frequency and content of alerts, their design and integration with existing interfaces, and the broader digital infrastructure and capability including the interoperability of data.

Frequency and Content of CRs

Four studies described how the inability to process a large number of alerts was a concern among GPs [59,61,63,64]. As one GP described:

XSL•FO
RenderX

> *There are so many things (...) so many things popping up, so many things prompting you. You don't probably respond to all the prompts, because there's a box here, a box there, a box everywhere, and you don't see everything (...) It's such a busy screen you don't respond to everything...* [61]

The value of presenting accompanying patient-centered information alongside the clinical data of the CR was described in one study. As a GP elaborated:

> *It [the CR] actually comes up with a nice diagram that it can use with your patients, that's got smiley faces of different colours to illustrate risk (...) it's just a different way to help augment that communication with them. For things like that it's quite helpful because patients are wary of tablets, quite rightly.* [59]

Design and Integration of CRs

Three studies described how pop-up CRs that required the checking of drop-down lists or otherwise had to be actioned before any next step were overly disruptive and actually inhibited their engagement with the CRs content [59,62,63]:

> *[when they appear]...you can't go looking in the notes, you can't input anything else, this is now taking priority, so you either suspend it and then re-open it and suspend it and re-open it.* [59]

Digital Infrastructure

The capacity of practice information technology (IT) systems and hardware was also a consideration, for example, some CRs slowed the operating speed of practice computers [59,60,62,66].

> *Our problem here is our computers are rubbish. They work really, really slowly (...) it takes a long time to come through so we scroll down an option, a pick list - it can pick the wrong thing! Which is very frustrating...* [62]

Value Proposition

GPs' Perspectives on CR Quality, Safety, and Educational Impact

GPs described their perspectives on the quality and safety of CRs, their potential to support continuing education and training, and their impact on self-efficacy.

Overall Safety and Quality

In 5 studies GPs reported that CRs were capable of improving patient safety by promoting reflection on their decision [58-61,63]:

> *...even if you immediately dismiss it, at least that millisecond you've thought about it, and I think that is going to be useful at some point, but for how many people I don't know* [61]

GPs also appreciated that CRs could improve the quality of care by providing useful reminders of tasks or actions [59,60]:

> *...in a patient with diabetes where there are 17 things to do and I've just missed out two or three of them (...) it's really nice to have "somebody" say "Oh, don't forget, there's the urine test, and you haven't yet sent the patient for the eye test that you thought somebody else was doing, and it looks like nobody's done after all so maybe get that done?"* [59]

GPs in one study reported how CRs could improve how they practice in the long term by their ability to promote and reinforce appropriate prescribing behaviors.

> *...and it's actually educational, because a lot of them you know the warning before it comes up because you've seen it so many times so you're already aware of it. So I mean some of the things I know now are because of it, so there is an educational element to it [...]* [58]

Self-Efficacy

There were conflicting accounts regarding the extent to which CRs enhance GPs' confidence in practicing safely. In one study, GPs described how the presence of CRs was a source of reassurance, supporting their ability to provide safe care in the increasingly complex environment of modern primary care [61,63].

> *...work in general practice is getting bigger and bigger, and more work's going to come to general practice. So, although it [the CR] can cause irritation, the flip side of that is you can relax a little bit more and that you don't have to remember absolutely everything (...)* [63]

In another study, a GP reflected upon how having to choose a predetermined response outside of their original decision led them to question their ability [59]:

> *...in order to dismiss it you have to select what action you've taken such as, well have you called an ambulance, have you sent them into hospital...you're then panicking thinking gosh I don't want, I don't want to lie, no I haven't sent them into hospital but is that a judgement, is that them saying because I haven't done that I'm a rubbish doctor?* [59]

Adopter

Factors Influencing GPs' Engagement With CRs: Digital Literacy and Consultation Impact

From the perspective of those adopting CRs, GPs described the factors affecting their engagement with CRs, the influence of digital literacy, and the impact of CRs on their consultation.

Engagement and Compliance in CRs

In one study, GPs expressed hesitation in relying on a computer-generated algorithm [59]. Concerns surfaced in 3 studies over the accuracy and relevance of the recommendations associated with CRs, with examples including unrecognized contraindications [58,63,65]:

> *[drug-drug interactions] were mentioned in printed textbooks but were not picked up by the interaction checking facility of their prescribing software* [63]

XSL•FO
RenderX

One study described a preference for greater transparency between the clinical evidence and the presentation of the CR. As one GP described:

> *[We're] inherently distrustful of anything that hasn't got somebody saying "Oh! I've done this trial, and this trial, and we've used it!" and I think that's what we've been taught at medical school - to be sceptical, until we've got the evidence.* [59]

Digital Literacy and Acceptability

The belief emerged that the new generation of GPs more familiar with digital technology in everyday life would be more accepting of computer-generated reminders [59].

> *I think to a certain extent there is a bit of a cultural change in the sense that the newer generation of doctors coming through are much more used to computers telling them things in their own lives and so they're much more accepting of the idea that the computer might give them helpful information.* [59]

GPs in 3 studies described how the reliance on CRs risked undermining the accrual of contextual knowledge of a particular patient [59,61,67].

> *A lot of the patients who I kind of see regularly with chronic problems, I kind of know if they're well or not because I know the patient and I just can tell if they're themselves or if they're not because you have a constant, you know, you lose that with a computer.* [59]

Related to this was the feeling that CRs were overly prescriptive and one GP felt that the value of their input was diminished relative to the diktats of the CR:

> *I think it feels frustrating...when the system is so process-driven that there's no sort of autonomy of the GP to, you know, change it or, kind of, put in what's relevant.* [59]

Patient-Clinician Relationship

Two studies reported that GPs understood the importance of focusing on patients during consultations, acknowledging that referral to, and interaction with, CRs can be an impediment to patient-centered interaction [61,63]:

> *I don't use any separate clinical decision support tool in the consultation because of time problems, losing eye contact with patients* [63]

Organization

Organizational Influences on CR Adoption: Training and System Capacity

In considering organizational influences, evidence emerged of the importance of training and the impact of existing practice capacity and systems.

Training and Purpose

GPs in 5 studies reflected on the importance of training not only to understand the purpose of CRs but also their functionality [58-61,63]. One GP bemoaned the typical lack of training before the introduction of digital tools:

> *...so, yeah things just happen [. . .] They often will appear before you have any training or knowledge of it whatsoever. Then if you're lucky maybe there's a launch event six months after you've already started using it.* [58]

Practice Capacity and Workflow

GPs noted how the pressure on their time restricted their ability to interact with CRs that take time to resolve [58,60]. As one GP explained:

> *We might have six or seven of them [reminders], and once you've got (...) worked through about three or four, and then there's this anticoagulation...you think, well that's another 5 or 10 minutes of consultation (...) so you leave it.* [60]

In another example, one GP likened the appearance of a CR to a colleague bursting through the door during a consultation:

> *It really is like an interruption, you know, no GP wants somebody to just burst in with the door opening or the phone ringing. In the same vein no GP really wants a big thing to just pop up on the screen that they didn't call up.* [59]

Wider System and Embedding Over Time

Enhancing CR Implementation: Systemwide Collaboration and Adaptation Strategies

The potential impact of improved systemwide collaboration was described, and a number of suggestions were made as to how CRs may be adapted to ensure more sustained implementation.

Systemwide Collaboration

It was suggested in 2 studies that more closely integrated data sets across settings would support the quality of the CRs [63,66].

> *I want a dynamic electronic health communication with the possibility of a written dialog and forwarding missing information.* [66]

The importance of sharing best practices and recommendations for certain CR systems between peers was also described:

> *I think there is a degree of sharing these kind of protocols and algorithms amongst GP practises and we certainly, you know, GPs talk to each other and say "oh yeah we've got this thing that alerts for that particular problem", "oh, have you? Right great can we, can we bring it in?"* [59]

Adaptation Over Time

Three studies described how GPs would have liked greater engagement with the software developers in the design of the CRs [58,63,65]. Suggestions for ways in which CRs might be improved included the acquisition of patient-specific alerts, better integration into practice workflows, and customizability informed by GP feedback [59-61,63]. In terms of their design and content, there were also suggestions for the inclusion of

XSL•FO
RenderX

additional details on dosage or color-grading CRs according to severity [63].

Discussion

General Findings

A total of 9 studies published since 2003 were identified, reporting research conducted in Europe, predominantly the United Kingdom, and Australia, 7 of which were judged to be of high quality. The findings from these studies were effectively categorized within the domains of the NASSS framework. Regarding Condition and Value Proposition, the use of CRs in diagnosis, prescribing, and the ongoing management of chronic conditions was seen as an effective means of maintaining or improving safe and high-quality care, albeit with contrasting effects on clinician self-efficacy. In terms of Technology attributes, the efficacy of CRs was found to be directly linked to their frequency, the presentation and accuracy of their content, and the speed and reliability of IT infrastructure and hardware. With respect to Adopter attitudes, GPs raised concerns about the accuracy of CRs and the potential loss of benefits derived from their clinical experience and consideration of individual patient contexts, in favor of algorithm-based decisions. From an Organization perspective, the importance of training and the ability of CRs to integrate seamlessly with existing workflows were emphasized. Finally, in the context of the Wider System, several suggestions were made to improve the usability of CRs, including the sharing of best practices and involving GPs earlier in the design process to ensure successful embedding over time.

Specific Findings and Comparison With Existing Literature

Condition/Value Proposition

Managing the medications of a growing number of patients with multiple long-term conditions is a major contributor to GP workload in the United Kingdom and other countries [68,69]. The studies we identified suggest that CRs can play a positive role in supporting this management [58,59,62-64,66], and it is widely recognized that CRs present a valuable opportunity to reduce adverse drug reactions [70].

The potential of CRs to support diagnosis was highlighted in 4 of the identified studies [60,61,64,65]. This ability has been previously acknowledged as a significant advantage in primary care, where diagnostic errors are a critical concern, particularly for diseases with low incidence but high health risk [71-73].

Primary care is expected to take on an increasing role in managing chronic conditions across various countries and health systems, including the United Kingdom [74,75]. However, this capacity is often hindered by limited consultation times, high patient volumes, and inadequate communication with specialists [75,76]. In such contexts, CRs that draw on patient-specific data to remind GPs to test, review, or conduct periodic interventions appear highly valuable, though only 1 study explicitly examined this functionality [64].

The ability of CRs to prompt reflection on clinical decision-making was viewed positively in 5 of the studies [58-61,63]. The benefits of reflective practice on clinical decisions are well established, particularly when the feedback on decisions, as with CRs, is immediate [77].

Technology

The studies we identified confirmed that GPs felt overwhelmed by the number, complexity, and often poor functionality of CRs [59,61,63,64]. This cognitive overload caused by multiple appearances of CRs reflects previous research suggesting it can hinder clinician interaction with clinical decision support systems [18-20,78]. The cognitive burden associated with CRs also stems from the complexity of their content [59,62,63]. Therefore, CRs should be designed to accurately convey risk and relevant information in a clear, easily interpretable format that adheres to best practices in digital interface design [79-82]. Additionally, they should be better integrated into existing software and workflows [83,84].

Artificial intelligence has begun to be utilized in clinical decision support systems for individual conditions, and it is reasonable to assume it will play a role in refining the specificity of CRs in primary care [85]. However, efforts to minimize the negative impact of CRs, such as improving specificity [18] or allowing clinicians the autonomy to disable certain CRs [83], must remain mindful of patient safety implications [86].

The ability of GPs to respond promptly to CRs was also influenced by the speed and reliability of practice IT infrastructure and hardware [59,60,62,66]. Ensuring that existing IT systems can handle the growing prevalence and complexity of digital health care is widely acknowledged as a critical need in both high- and low- and middle-income countries. However, this challenge has yet to be universally addressed [87,88].

Adopter

GP concerns about CRs that enforce process-driven decisions at the expense of context and experience were described in 4 of the studies we identified [58,59,63,65]. Notably, lasting cynicism expressed by GPs who had experienced prior failures with CRs (described in 2 of the identified studies [59,63]) has been previously recognized in primary care, where digital interventions have failed to meet expectations [89].

A framework for technology developers has emerged that aims to enhance clinician confidence in clinical decision support systems by adhering to a predetermined set of criteria, focusing on understanding and sustaining trust among end users [90,91]. Evidence from one study we identified suggested that resistance to digital interventions may be lower among the generation of clinicians who grew up alongside the proliferation of digital technologies [59]. It is hoped that trust in CRs can be fostered through greater and earlier engagement with clinician end users by software developers and senior decision makers [92-94]. This aligns with long-standing recommendations for increased user involvement in the development of digital health tools [67,89].

The negative disruption of patient consultations by CRs was also highlighted [61,63]. It is already known that technology can interrupt the flow of conversation, reduce nonverbal communication cues, and compromise the overall quality of patient interactions [95,96]. Concerns expressed by GP

XSL•FO
RenderX

participants in the identified studies reflect existing evidence that clinicians highly value uninterrupted in-person consultations, as they build patient trust and foster a constructive provider-patient relationship [97,98].

Organization

The lack of training emerged as a barrier to engagement and compliance with CRs in the studies we identified [58-61,63] and has previously been highlighted as central to their effective use, including sharing with GPs the processes behind their generation [99,100]. However, targeted training to enhance engagement with CRs has yet to be systematically implemented in the high-income countries where they are most prevalent [101].

The pressure on GPs to address increasingly complex care needs and rising patient expectations within shortening consultation times is widely acknowledged [102]. GPs in the identified studies reported that managing multiple CRs was unfeasible within the constraints of limited consultation times [58,60]. Some studies suggest that specific CRs, such as those supporting preventive care, could be delivered to patients before the consultation to alleviate the time pressure on GPs [103].

Wider System and Embedding Over Time

The need for systemwide coordination and collaboration in introducing CRs into primary care, as described in the review's findings [59,63,66], has also been recognized in other health care settings [104-106]. Similarly, the calls for earlier clinician engagement in the design and development of CRs identified in this review [58,63,65] have been previously acknowledged [107-109], alongside the input of clinical informaticists [110]. While there are recognized challenges in engaging time-pressured GPs during the early stages of software development, sustainable adoption of health technology requires understanding its social and organizational contexts. This can only be achieved by involving end users throughout its design and development [111,112].

Strengths and Limitations

The recognized systematic review methodology was applied to identify relevant papers [27,113], including the dual-review process [114]. The comprehensive NASSS framework successfully accommodated all of the data identified [26] and its use demonstrated both the transparency and rigor of our approach but also allowed us to structure and contextualize our data with existing research to avoid distorting our conclusions [28]. We acknowledge that there were a number of potential limitations. First, all the studies identified originated in high-income countries and were presented in English; however, we would suggest that the lessons learned have the potential for broader applicability for digital interventions in health systems of low- and middle-income countries that in reality share many of the same, if accentuated, barriers to technology-enabled care [111,112]. Second, it is possible that the use of an a priori framework might limit the analysis, with some data unable to be allocated within its confines; however, taking the unconstrained matrix approach suggested by Elo and Kyngäs [56] allowed the development and inclusion of emergent themes within the established framework. This meant we could maintain alignment with our established objectives and provide a systematic description of novel findings.

Conclusions

If utilized correctly CRs can enhance patient safety, workflow management, preventative care, and reduce cognitive overload. Based on the findings of the review several recommendations can be made for the future design and implementation of CRs (see Table 4). These include improved graphical designs better suited to the purpose of the CR, more nuanced and patient-specific content, greater clinician autonomy regards if and when CRs are seen, and earlier and more consistent engagement with the workforce in their design and development.

Table 4. Recommendations for future CR[a] design and implementation.

Domain	Recommendation
Technology	• Visual-driven user interface to present information more effectively and reduce information overload. • More directly applicable data (eg, dosage or risk). • The capacity of the infrastructure.
Adopter	• GP[b]-endorsed or GP-recommended CRs would promote and increase their uptake. • Ensuring that CRs transparently reflect the latest clinical evidence.
Organization	• Greater engagement with the workforce in the introduction of CRs. • Additional training would help the uptake of CRs and using them effectively, accommodating the age and experience of the clinician. • Enable organizations to better incorporate CRs in existing/novel workflows.
Wider systems	• GP-stakeholders collaboration would enable feedback and co-design of alert utilization.
Adaptation	• Alert customization would allow GPs to deactivate unnecessary options to change output and will provide flexibility.

[a]CR: clinical reminder.

[b]GP: general practitioner.

XSL•FO
RenderX

Acknowledgments

The main funding for this project was provided by the Engineering and Physical Sciences Research Council (grant EP/M014401).

Data Availability

This is a systematic review and the data it uses are in the public domain.

Authors' Contributions

IL was responsible for conceptualization, funding acquisition, methodology, project administration, supervision, formal analysis, writing—original draft, writing/review, and editing. IG was responsible for the methodology, investigation, data curation, formal analysis, writing—original draft, and writing—review & editing. DS was responsible for validation and writing—review and editing. GD was responsible for writing—review and editing. NC was responsible for writing—review and editing. AP was responsible for data curation, formal analysis, validation, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 33 KB - jmir_v27i1e62763_app1.docx]

Multimedia Appendix 2
Eligibility criteria.
[DOCX File , 20 KB - jmir_v27i1e62763_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist (v17.12.24).
[PDF File (Adobe PDF File), 71 KB - jmir_v27i1e62763_app3.pdf]

Multimedia Appendix 4
Study selection and data extraction templates.
[DOCX File , 26 KB - jmir_v27i1e62763_app4.docx]

Multimedia Appendix 5
Excluded studies.
[DOCX File , 29 KB - jmir_v27i1e62763_app5.docx]

Multimedia Appendix 6
Quality appraisal.
[DOCX File , 30 KB - jmir_v27i1e62763_app6.docx]

References

1. Macdonald N, Clements C, Sobti A, Rossiter D, Unnithan A, Bosanquet N. Tackling the elective case backlog generated by Covid-19: the scale of the problem and solutions. J Public Health (Oxf) 2020 Nov 23;42(4):712-716 [FREE Full text] [doi: 10.1093/pubmed/fdaa155] [Medline: 32930797]
2. Henley J, Connolly K, Jones S, Giuffrida A. A ticking time bomb?: Health care under threat across western Europe. The Guardian. 2022. URL: https://www.theguardian.com/society/2022/dec/14/a-ticking-time-bomb-healthcare-under-threat-across-western-europe [accessed 2024-07-01]
3. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay C, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Int J Technol Assess Health Care 2005 Mar 02;21(1):149-149. [doi: 10.1017/s0266462305290190]
4. Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. J Contin Educ Health Prof 2004;24 Suppl 1(S1):S31-S37. [doi: 10.1002/chp.1340240506] [Medline: 15712775]

XSL•FO
RenderX

5. Brokel J, Schwichtenberg T, Wakefield D, Ward M, Shaw M, Kramer J. Evaluating clinical decision support rules as an intervention in clinician workflows with technology. Computers, Informatics, Nursing 2011;29(1):36-42. [doi: 10.1097/ncn.0b013e3181f9dbb1]
6. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, with the HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36-32 [FREE Full text] [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
7. Saleem JJ, Patterson ES, Militello L, Anders S, Falciglia M, Wissman JA, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. Journal of the American Medical Informatics Association 2007 Sep 01;14(5):632-640. [doi: 10.1197/jamia.m2163]
8. Ivers N, Jamtvedt G, Flottorp S, Young J, Odgaard-Jensen J, French S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jun 13;2012(6):CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
9. Baker R, Camosso-Stefinovic J, Gillies C, Shaw E, Cheater F, Flottorp S, et al. Tailored interventions to address determinants of practice. Cochrane Database Syst Rev 2015 Apr 29;2015(4):CD005470 [FREE Full text] [doi: 10.1002/14651858.CD005470.pub3] [Medline: 25923419]
10. Dexheimer JW, Talbot TR, Sanders DL, Rosenbloom ST, Aronsky D. Prompting clinicians about preventive care measures: a systematic review of randomized controlled trials. Journal of the American Medical Informatics Association 2008 May 01;15(3):311-320. [doi: 10.1197/jamia.m2555]
11. Cecil E, Dewa LH, Ma R, Majeed A, Aylin P. General practitioner and nurse practitioner attitudes towards electronic reminders in primary care: a qualitative analysis. BMJ Open 2021 Jul 12;11(7):e045050 [FREE Full text] [doi: 10.1136/bmjopen-2020-045050] [Medline: 34253661]
12. Moxey A, Robertson J, Newby D, Hains I, Williamson M, Pearson SA. Computerized clinical decision support for prescribing: provision does not guarantee uptake. J Am Med Inform Assoc 2010;17(1):25-33 [FREE Full text] [doi: 10.1197/jamia.M3170] [Medline: 20064798]
13. Litvin CB, Davis KS, Moran WP, Iverson PJ, Zhao Y, Zapka J. The use of clinical decision‐support tools to facilitate geriatric education. J American Geriatrics Society 2012 May 29;60(6):1145-1149. [doi: 10.1111/j.1532-5415.2012.03960.x]
14. Trevena L, McCaffery K, Salkeld G, Glasziou P, Del MC, Doust J. Clinical decision-making tools: how effective are they in improving the quality of health care? Analysis & Policy Observatory. 2014. URL: https://apo.org.au/node/53712 [accessed 2025-01-30]
15. Elwyn G, Scholl I, Tietbohl C, Mann M, Edwards AG, Clay C, et al. "Many miles to go …": a systematic review of the implementation of patient decision support interventions into routine clinical practice. BMC Med Inform Decis Mak 2013 Nov 29;13(S2):e. [doi: 10.1186/1472-6947-13-s2-s14]
16. Magnus D, Rodgers S, Avery AJ. GPs' views on computerized drug interaction alerts: questionnaire survey. J Clin Pharm Ther 2002 Oct;27(5):377-382. [doi: 10.1046/j.1365-2710.2002.00434.x] [Medline: 12383140]
17. Militello L, Patterson ES, Tripp-Reimer T, Asch SM, Fung CH, Glassman P, et al. Clinical reminders: why don't they use them? In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2004 Sep 01 Presented at: The Human Factors and Ergonomics Society Annual Meeting; September 1, 2004; New Orleans, LO p. 1651-1655. [doi: 10.1177/154193120404801505]
18. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. Journal of the American Medical Informatics Association 2006 Mar 01;13(2):138-147. [doi: 10.1197/jamia.m1809]
19. Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS. Physicians' decisions to override computerized drug alerts in primary care. Arch Intern Med 2003 Nov 24;163(21):2625-2631. [doi: 10.1001/archinte.163.21.2625] [Medline: 14638563]
20. Hsieh TC, Kuperman GJ, Jaggi T, Hojnowski-Diaz P, Fiskio J, Williams DH, et al. Characteristics and consequences of drug allergy alert overrides in a computerized physician order entry system. J Am Med Inform Assoc 2004 Nov 01;11(6):482-491. [doi: 10.1197/jamia.m1556]
21. Feldstein A, Simon SR, Schneider J, Krall M, Laferriere D, Smith DH, et al. How to Design Computerized Alerts to Ensure Safe Prescribing Practices. The Joint Commission Journal on Quality and Safety 2004 Nov;30(11):602-613. [doi: 10.1016/s1549-3741(04)30071-7] [Medline: 15565759]
22. Bell H, Garfield S, Khosla S, Patel C, Franklin BD. Mixed methods study of medication-related decision support alerts experienced during electronic prescribing for inpatients at an English hospital. Eur J Hosp Pharm 2019 Nov 19;26(6):318-322 [FREE Full text] [doi: 10.1136/ejhpharm-2017-001483] [Medline: 31798854]
23. Thomas Craig KJ, Willis V, Gruen D, Rhee K, Jackson G. The burden of the digital environment: a systematic review on organization-directed workplace interventions to mitigate physician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):985-997 [FREE Full text] [doi: 10.1093/jamia/ocaa301] [Medline: 33463680]
24. Teich JM, Osheroff JA, Pifer EA, Sittig DF. Clinical Decision Support in Electronic Prescribing: Recommendations and an Action Plan. J Am Med Inform Assoc 2005 Jul 01;12(4):365-376. [doi: 10.1197/jamia.m1822]
25. Saleem JJ, Patterson ES, Militello L, Render ML, Orshansky G, Asch SM. Exploring barriers and facilitators to the use of computerized clinical reminders. J Am Med Inform Assoc 2005 Jul 01;12(4):438-447. [doi: 10.1197/jamia.m1777]

XSL•FO
RenderX

26. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
27. Assarroudi A, Heshmati Nabavi F, Armat MR, Ebadi A, Vaismoradi M. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. J Res Nurs 2018 Mar 10;23(1):42-55 [FREE Full text] [doi: 10.1177/1744987117741667] [Medline: 34394406]
28. McKenzie J, Brennan S, Higgins J, Thomas J, Chandler J, Cumpston M. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.5. London, UK: Cochrane; 2022.
29. Abimbola S, Patel B, Peiris D, Patel A, Harris M, Usherwood T, et al. The NASSS framework for ex post theorisation of technology-supported change in healthcare: worked example of the TORPEDO programme. BMC Med 2019 Dec 30;17(1):233 [FREE Full text] [doi: 10.1186/s12916-019-1463-x] [Medline: 31888718]
30. Litchfield I, Glasby J, Parkinson S, Hocking L, Tanner D, Roe B, et al. "Trying to find people to fit the tech…": a qualitative exploration of the lessons learnt introducing artificial intelligence-based technology into English social care. Health & Social Care in the Community 2023 May 16;2023:1-11. [doi: 10.1155/2023/9174873]
31. Greenhalgh T, Maylor H, Shaw S, Wherton J, Papoutsi C, Betton V, et al. The NASSS-CAT tools for understanding, guiding, monitoring, and researching technology implementation projects in health and social care: protocol for an evaluation study in real-world settings. JMIR Res Protoc 2020 May 13;9(5):e16861 [FREE Full text] [doi: 10.2196/16861] [Medline: 32401224]
32. Schaink AK, Kuluski K, Lyons RF, Fortin M, Jadad AR, Upshur R, et al. A scoping review and thematic classification of patient complexity: offering a unifying framework. J Comorb 2012 Oct 10;2(1):1-9 [FREE Full text] [doi: 10.15256/joc.2012.2.15] [Medline: 29090137]
33. Proctor E, Ramsey AT, Saldana L, Maddox TM, Chambers DA, Brownson RC. FAST: a framework to assess speed of translation of health innovations to practice and policy. Glob Implement Res Appl 2022 Jun 02;2(2):107-119 [FREE Full text] [doi: 10.1007/s43477-022-00045-4] [Medline: 35669171]
34. Schoville RR, Titler MG. Guiding healthcare technology implementation: a new integrated technology implementation model. Comput Inform Nurs 2015 Mar;33(3):99-107; quiz E1. [doi: 10.1097/CIN.0000000000000130] [Medline: 25799235]
35. Belay M, Desta A, Smithson S, Meshesha M. Investigate knowledge management technology implementation for supporting decision making in Ethiopian health sectors. BMC Med Inform Decis Mak 2021 May 05;21(1):146 [FREE Full text] [doi: 10.1186/s12911-021-01507-9] [Medline: 33952240]
36. Priestman W, Collins R, Vigne H, Sridharan S, Seamer L, Bowen D, et al. Lessons learned from a comprehensive electronic patient record procurement process-implications for healthcare organisations. BMJ Health Care Inform 2019 May 09;26(1):e000020 [FREE Full text] [doi: 10.1136/bmjhci-2019-000020] [Medline: 31072821]
37. Terry N. Appification, AI, and healthcare's new iron triangle. J Health Care L & Pol'y 2017;20:117 [FREE Full text]
38. Lehoux P, Miller FA, Daudelin G, Denis J. Providing value to new health technology: the early contribution of entrepreneurs, investors, and regulatory agencies. Int J Health Policy Manag 2017 Sep 01;6(9):509-518 [FREE Full text] [doi: 10.15171/ijhpm.2017.11] [Medline: 28949463]
39. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
40. Perski O, Short C. Acceptability of digital health interventions: embracing the complexity. Transl Behav Med 2021 Jul 29;11(7):1473-1480 [FREE Full text] [doi: 10.1093/tbm/ibab048] [Medline: 33963864]
41. Rahman MS, Ko M, Warren J, Carpenter D. Healthcare Technology Self-Efficacy (HTSE) and its influence on individual attitude: an empirical study. Computers in Human Behavior 2016 May;58:12-24. [doi: 10.1016/j.chb.2015.12.016]
42. Lythreatis S, Singh SK, El-Kassar A. The digital divide: a review and future research agenda. Technological Forecasting and Social Change 2022 Feb;175:121359. [doi: 10.1016/j.techfore.2021.121359]
43. Caccia-Bava MDC, Guimaraes T, Harrington S. Hospital organization culture, capacity to innovate and success in technology adoption. J Health Organ Manag 2006;20(2-3):194-217. [doi: 10.1108/14777260610662735] [Medline: 16869354]
44. Cresswell K, Sheikh A, Franklin BD, Krasuska M, Nguyen HT, Hinder S, et al. Theoretical and methodological considerations in evaluating large-scale health information technology change programmes. BMC Health Serv Res 2020 May 27;20(1):477 [FREE Full text] [doi: 10.1186/s12913-020-05355-7] [Medline: 32460830]
45. Fitzgerald L, McDermott A. Challenging Perspectives on Organizational Change in Health Care. Abingdon, Oxon: Routledge; 2017.
46. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
47. Frønsdal KB, Facey K, Klemp M, Norderhaug IN, Mørland B, Røttingen J. Health technology assessment to optimize health technology utilization: using implementation initiatives and monitoring processes. Int J Technol Assess Health Care 2010 Jun 29;26(3):309-316. [doi: 10.1017/s0266462310000309]

XSL•FO
RenderX

48. Ofori-Asenso R, Hallgreen CE, De Bruin ML. Improving interactions between health technology assessment bodies and regulatory agencies: a systematic review and cross-sectional survey on processes, progress, outcomes, and challenges. Front Med (Lausanne) 2020 Oct 16;7:582634 [FREE Full text] [doi: 10.3389/fmed.2020.582634] [Medline: 33178721]
49. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010 Nov 02;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
50. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Quarterly 2004 Dec 09;82(4):581-629. [doi: 10.1111/j.0887-378x.2004.00325.x]
51. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
52. Hollnagel E, Braithwaite J, Wears R. Delivering Resilient Health Care. Abingdon, Oxon: Routledge; 2018.
53. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct 24;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
54. Backman R, Bayliss S, Moore D, Litchfield I. Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence. Syst Rev 2017 Dec 13;6(1):255 [FREE Full text] [doi: 10.1186/s13643-017-0627-z] [Medline: 29237488]
55. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
56. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr 18;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
57. Bingham AJ, Witkowsky P. Deductive and inductive approaches. In: Analyzing and Interpreting Qualitative Data: After the Interview. Thousand Oaks, CA: Sage Publications; 2021.
58. Jeffries M, Salema N, Laing L, Shamsuddin A, Sheikh A, Avery A, et al. The implementation, use and sustainability of a clinical decision support system for medication optimisation in primary care: a qualitative evaluation. PLoS One 2021 May 3;16(5):e0250946 [FREE Full text] [doi: 10.1371/journal.pone.0250946] [Medline: 33939750]
59. Ford E, Edelman N, Somers L, Shrewsbury D, Lopez Levy M, van Marwijk H, et al. Barriers and facilitators to the adoption of electronic clinical decision support systems: a qualitative interview study with UK general practitioners. BMC Med Inform Decis Mak 2021 Jun 21;21(1):193 [FREE Full text] [doi: 10.1186/s12911-021-01557-z] [Medline: 34154580]
60. Holt TA, Dalton AR, Kirkpatrick S, Hislop J, Marshall T, Fay M, et al. Barriers to a software reminder system for risk assessment of stroke in atrial fibrillation: a process evaluation of a cluster randomised trial in general practice. Br J Gen Pract 2018 Nov 05;68(677):e844-e851. [doi: 10.3399/bjgp18x699809]
61. Dikomitis L, Green T, Macleod U. Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences. Prim Health Care Res Dev 2015 Mar 3;16(06):548-555. [doi: 10.1017/s1463423615000109]
62. Slight SP, Howard R, Ghaleb M, Barber N, Franklin BD, Avery AJ. The causes of prescribing errors in English general practices: a qualitative study. Br J Gen Pract 2013 Oct 01;63(615):e713-e720. [doi: 10.3399/bjgp13x673739]
63. Ahearn MD, Kerr SJ. General practitioners' perceptions of the pharmaceutical decision-support tools in their prescribing software. Med J Aust 2003 Jul 07;179(1):34-37. [doi: 10.5694/j.1326-5377.2003.tb05415.x] [Medline: 12831382]
64. Heselmans A, Aertgeerts B, Donceel P, Geens S, Van de Velde S, Ramaekers D. Family physicians' perceptions and use of electronic clinical decision support during the first year of implementation. J Med Syst 2012 Dec 9;36(6):3677-3684. [doi: 10.1007/s10916-012-9841-3] [Medline: 22402980]
65. Bindels R, Hasman A, Derickx M, Van Wersch JWJ, Winkens RAG. User satisfaction with a real-time automated feedback system for general practitioners: a quantitative and qualitative study. Int J Qual Health Care 2003 Dec 01;15(6):501-508. [doi: 10.1093/intqhc/mzg076] [Medline: 14660533]
66. Christensen T, Grimsmo A. Expectations for the next generation of electronic patient records in primary care: a triangulated study. Inform Prim Care 2008;16(1):21-28 [FREE Full text] [doi: 10.14236/jhi.v16i1.671] [Medline: 18534074]
67. Mahmud H, Islam AN, Ahmed SI, Smolander K. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Technological Forecasting and Social Change 2022 Feb;175:121390. [doi: 10.1016/j.techfore.2021.121390]
68. Carrier H, Zaytseva A, Bocquier A, Villani P, Verdoux H, Fortin M, et al. GPs' management of polypharmacy and therapeutic dilemma in patients with multimorbidity: a cross-sectional survey of GPs in France. Br J Gen Pract 2019 Feb 25;69(681):e270-e278. [doi: 10.3399/bjgp19x701801]
69. Reeve J. Avoiding harm: Tackling problematic polypharmacy through strengthening expert generalist practice. Brit J Clinical Pharma 2020 Sep 09;87(1):76-83. [doi: 10.1111/bcp.14531]
70. Armando LG, Miglio G, de Cosmo P, Cena C. Clinical decision support systems to improve drug prescription and therapy optimisation in clinical practice: a scoping review. BMJ Health Care Inform 2023 May 02;30(1):e100683 [FREE Full text] [doi: 10.1136/bmjhci-2022-100683] [Medline: 37130626]
71. Lehtovuori T, Heikkinen AM, Raina M, Kauppila T. The effect of electronic reminders on the recording of diagnoses in primary care: a quasi-experimental before and after study. SAGE Open Med 2020 Apr 28;8:2050312120918267 [FREE Full text] [doi: 10.1177/2050312120918267] [Medline: 32435481]

XSL•FO
RenderX

72. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. BMJ Qual Saf 2017 Jun 16;26(6):484-494 [FREE Full text] [doi: 10.1136/bmjqs-2016-005401] [Medline: 27530239]
73. Round T, Steed L, Shankleman J, Bourke L, Risi L. Primary care delays in diagnosing cancer: what is causing them and what can we do about them? J R Soc Med 2013 Nov 09;106(11):437-440 [FREE Full text] [doi: 10.1177/0141076813504744] [Medline: 24108536]
74. Nolte E, Knai C, Saltman R. Assessing chronic disease management in European health systems. In: European Observatory on Health Systems and Policies. Copenhagen, Denmark: European Observatory on Health Systems and Policies; 2014.
75. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 09;19(1):11 [FREE Full text] [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]
76. Foo KM, Sundram M, Legido-Quigley H. Facilitators and barriers of managing patients with multiple chronic conditions in the community: a qualitative study. BMC Public Health 2020 Feb 27;20(1):273 [FREE Full text] [doi: 10.1186/s12889-020-8375-8] [Medline: 32106838]
77. Prakash S, Sladek RM, Schuwirth L. Interventions to improve diagnostic decision making: a systematic review and meta-analysis on reflective strategies. Medical Teacher 2018 Sep 23;41(5):517-524. [doi: 10.1080/0142159x.2018.1497786]
78. Westerbeek L, Ploegmakers KJ, de Bruijn G, Linn AJ, van Weert JC, Daams JG, et al. Barriers and facilitators influencing medication-related CDSS acceptance according to clinicians: a systematic review. Int J Med Inform 2021 Aug;152:104506 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104506] [Medline: 34091146]
79. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. J Biomed Inform 2012 Dec;45(6):1202-1216 [FREE Full text] [doi: 10.1016/j.jbi.2012.09.002] [Medline: 22995208]
80. Wiklund M, Kendler J, Hochberg L, Weinger M. Technical basis for user interface design of health IT. National Institute of Standards and Technology (NIST). 2015. URL: https://nvlpubs.nist.gov/nistpubs/gcr/2015/NIST.GCR.15-996.pdf [accessed 2025-01-30]
81. Silva S, Sousa Santos B, Madeira J. Using color in visualization: a survey. Computers & Graphics 2011 Apr;35(2):320-333. [doi: 10.1016/j.cag.2010.11.015]
82. Franconeri SL, Padilla LM, Shah P, Zacks JM, Hullman J. The science of visual data communication: what works. Psychol Sci Public Interest 2021 Dec 15;22(3):110-161. [doi: 10.1177/15291006211051956] [Medline: 34907835]
83. Olakotan OO, Mohd Yusof M. The appropriateness of clinical decision support systems alerts in supporting clinical workflows: a systematic review. Health Informatics J 2021 Apr 15;27(2):14604582211007536 [FREE Full text] [doi: 10.1177/14604582211007536] [Medline: 33853395]
84. Sharma D, Bhaskar S. Addressing the Covid-19 burden on medical education and training: the role of telemedicine and tele-education during and beyond the pandemic. Front Public Health 2020 Nov 27;8:589669 [FREE Full text] [doi: 10.3389/fpubh.2020.589669] [Medline: 33330333]
85. Wang L, Chen X, Zhang L, Li L, Huang Y, Sun Y, et al. Artificial intelligence in clinical decision support systems for oncology. Int J Med Sci 2023;20(1):79-86 [FREE Full text] [doi: 10.7150/ijms.77205] [Medline: 36619220]
86. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation. Health Aff (Millwood) 2011 Dec;30(12):2310-2317. [doi: 10.1377/hlthaff.2010.1111] [Medline: 22147858]
87. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022:10.31478/202206e [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
88. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
89. Ziebland S, Hyde E, Powell J. Power, paradox and pessimism: On the unintended consequences of digital health technologies in primary care. Soc Sci Med 2021 Nov;289:114419. [doi: 10.1016/j.socscimed.2021.114419] [Medline: 34619631]
90. Jones C, Thornton J, Wyatt JC. Enhancing trust in clinical decision support systems: a framework for developers. BMJ Health Care Inform 2021 Jun 04;28(1):e100247 [FREE Full text] [doi: 10.1136/bmjhci-2020-100247] [Medline: 34088721]
91. O'Neill O. Linking Trust to Trustworthiness. International Journal of Philosophical Studies 2018 Apr 25;26(2):293-300. [doi: 10.1080/09672559.2018.1454637]
92. McDermott I, Checkland K, Coleman A, Osipovič D, Petsoulas C, Perkins N. Engaging GPs in commissioning: realist evaluation of the early experiences of Clinical Commissioning Groups in the English NHS. J Health Serv Res Policy 2017 Jan 08;22(1):4-11 [FREE Full text] [doi: 10.1177/1355819616648352] [Medline: 27151153]
93. Duncan P, Ridd MJ, McCahon D, Guthrie B, Cabral C. Barriers and enablers to collaborative working between GPs and pharmacists: a qualitative interview study. Br J Gen Pract 2020 Feb 10;70(692):e155-e163. [doi: 10.3399/bjgp20x708197]
94. Sherrill AM, Wiese CW, Abdullah S, Arriaga RI. Overcoming clinician technophobia: what we learned from our mass exposure to telehealth during the COVID-19 pandemic. J Technol Behav Sci 2022 Aug 19;7(4):547-553 [FREE Full text] [doi: 10.1007/s41347-022-00273-3] [Medline: 36034538]

XSL•FO
RenderX

95. Kumarapeli P, de Lusignan S. Using the computer in the clinical consultation; setting the stage, reviewing, recording, and taking actions: multi-channel video study. J Am Med Inform Assoc 2013 Jun 01;20(e1):e67-e75 [FREE Full text] [doi: 10.1136/amiajnl-2012-001081] [Medline: 23242763]
96. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst 2020;1(3):e1.
97. Caldwell G. The process of clinical consultation is crucial to patient outcomes and safety: 10 quality indicators. Clin Med (Lond) 2019 Nov;19(6):503-506 [FREE Full text] [doi: 10.7861/clinmed.2019-0263] [Medline: 31732592]
98. Crook RL, Iftikhar H, Moore S, Lowdon P, Modarres P, Message S. A comparison of in-person versus telephone consultations for outpatient hospital care. Future Healthc J 2022 Jul;9(2):154-160 [FREE Full text] [doi: 10.7861/fhj.2022-0006] [Medline: 35928204]
99. Kersting C, Weltermann B. Electronic reminders to facilitate longitudinal care: a mixed-methods study in general practices. BMC Med Inform Decis Mak 2016 Nov 24;16(1):148 [FREE Full text] [doi: 10.1186/s12911-016-0387-z] [Medline: 27881130]
100. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
101. Johnson MP, Zheng K, Padman R. Modeling the longitudinality of user acceptance of technology with an evidence-adaptive clinical decision support system. Decision Support Systems 2014 Jan;57:444-453. [doi: 10.1016/j.dss.2012.10.049]
102. Salisbury H. Helen Salisbury: the 10 minute appointment. BMJ 2019 Jun 04;365:l2389. [doi: 10.1136/bmj.l2389] [Medline: 31164335]
103. Frank OR. Clinical software packages can be improved to help general practitioners increase the provision of preventive care. Aust. J. Prim. Health 2022 Apr 28;28(3):204-206. [doi: 10.1071/py21204]
104. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 06;3(1):17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
105. Chen W, O'Bryan CM, Gorham G, Howard K, Balasubramanya B, Coffey P, et al. Barriers and enablers to implementing and using clinical decision support systems for chronic diseases: a qualitative systematic review and meta-aggregation. Implement Sci Commun 2022 Jul 28;3(1):81 [FREE Full text] [doi: 10.1186/s43058-022-00326-x] [Medline: 35902894]
106. Meunier P, Raynaud C, Guimaraes E, Gueyffier F, Letrilliart L. Barriers and facilitators to the use of clinical decision support systems in primary care: a mixed-methods systematic review. Ann Fam Med 2023 Jan 23;21(1):57-69 [FREE Full text] [doi: 10.1370/afm.2908] [Medline: 36690490]
107. Schwartz J, Moy A, Rossetti S, Elhadad N, Cato K. Clinician involvement in research on machine learning-based predictive clinical decision support for the hospital setting: A scoping review. J Am Med Inform Assoc 2021 Mar 01;28(3):653-663 [FREE Full text] [doi: 10.1093/jamia/ocaa296] [Medline: 33325504]
108. Trinkley KE, Blakeslee WW, Matlock DD, Kao DP, Van Matre AG, Harrison R, et al. Clinician preferences for computerised clinical decision support for medications in primary care: a focus group study. BMJ Health Care Inform 2019 Apr 17;26(1) [FREE Full text] [doi: 10.1136/bmjhci-2019-000015] [Medline: 31039120]
109. Sanz M, Acha B, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021 Apr 30;21(2):16 [FREE Full text] [doi: 10.5334/ijic.5573] [Medline: 33981193]
110. Brown J, Hewner S. The role of telehealth and clinical informatics in data driven primary care redesign. J Inform Nurs 2022;6(4):jin_21N4_A3. [Medline: 35733915]
111. Aarts J, Callen J, Coiera E, Westbrook J. Information technology in health care: socio-technical approaches. In: Int J Med Inform. Information technology in health care: socio-technical approaches. Elsevier; Jun 2010:389-390.
112. Bate P, Robert G. Bringing User Experience to Health Care Improvement: The Concepts, Methods and Practices of Experience-Based Design. Abingdon, Oxon, UK: Radcliffe Publishing; 2007.
113. Cumpston M, Chandler J. Chapter II: Planning a Cochrane review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.5. London, UK: Cochrane; 2022.
114. Stoll CR, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. Res Synth Methods 2019 Dec 18;10(4):539-545 [FREE Full text] [doi: 10.1002/jrsm.1369] [Medline: 31272125]

Abbreviations

CR: clinical reminder
GP: general practitioner
IT: information technology
NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

Edited by N Cahill; submitted 03.06.24; peer-reviewed by R Marshall, V P?rez Jover; comments to author 27.09.24; revised version received 08.10.24; accepted 23.12.24; published 07.02.25.

Please cite as:
Gani I, Litchfield I, Shukla D, Delanerolle G, Cockburn N, Pathmanathan A
Understanding "Alert Fatigue" in Primary Care: Qualitative Systematic Review of General Practitioners Attitudes and Experiences of Clinical Alerts, Prompts, and Reminders
J Med Internet Res 2025;27:e62763
URL: https://www.jmir.org/2025/1/e62763
doi:10.2196/62763
PMID:39918864

©Illin Gani, Ian Litchfield, David Shukla, Gayathri Delanerolle, Neil Cockburn, Anna Pathmanathan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Assessing the Uses, Benefits, and Limitations of Digital Technologies Used by Health Professionals in Supporting Obesity and Mental Health Communication: Scoping Review

Amanda Kearns[1], MPH, MBS; Anne Moorhead[2], PhD; Maurice Mulvenna[3], PhD; Raymond Bond[3], PhD

[1]Life and Health Sciences, Institute for Nursing and Health Research, Ulster University, Belfast, United Kingdom

[2]School of Communication and Media, Institute for Nursing and Health Research, Ulster University, Belfast, United Kingdom

[3]School of Computing, Ulster University, Belfast, United Kingdom

Corresponding Author:
Amanda Kearns, MPH, MBS
Life and Health Sciences
Institute for Nursing and Health Research
Ulster University
2-24 York Street
Belfast, BT15 1AP
United Kingdom
Phone: 44 7706848477
Email: kearns-a7@ulster.ac.uk

Abstract

Background: Obesity and mental health issues present interconnected public health challenges that impair physical, social, and mental well-being. Digital technologies offer potential for enhancing health care communication between health professionals (HPs) and individuals living with obesity and mental health issues, but their effectiveness is not fully understood.

Objective: This scoping review aims to identify and understand the different types of technologies used by HPs in supporting obesity and mental health communication.

Methods: A comprehensive scoping review, which followed a validated methodology, analyzed studies published between 2013 and 2023 across 8 databases. The data extraction focused on HPs' use of communication technologies, intervention types, biopsychosocial considerations, and perceptions of technology use. The review was guided by the following research question: "What are the uses, benefits, and limitations of digital technologies in supporting communication between HPs and persons living with obesity and mental health issues?"

Results: In total, 8 studies—featuring web-based platforms, social media, synchronous video calls, telephone calls, automated SMS text messaging, and email—met the inclusion criteria. Technologies such as virtual learning collaborative dashboards and videoconferencing, supported by automated SMS text messaging and social media (Facebook and WhatsApp groups), were commonly used. Psychologists, dietitians, social workers, and health coaches used digital tools to facilitate virtual appointments, diet and mental health monitoring, and motivational and educational support through group therapy, 1-on-1 sessions, and hybrid models. Benefits included enhanced access to care and engagement, personalized digital cognitive behavioral therapy, perceived stigma reduction, privacy, and improved physical health outcomes in weight reduction. However, improvements in mental health outcomes were not statistically significant in studies reporting *P* values ($P \geq .05$). The limitations included engagement difficulties due to conflicting personal family and work commitments; variable communication mode preferences, with some preferring in-person sessions; and misinterpretations of SMS text messaging prompts. Conflicts arose from cultural and individual differences, weight stigma, and confusion over HP roles in obesity and mental health care.

Conclusions: Digital technologies have diversified the approaches HPs can take in delivering education, counseling, and motivation to individuals with obesity and mental health issues, facilitating private, stigma-reduced environments for personalized care. While the interventions were effective in obesity management, the review revealed a shortfall in addressing mental health needs. This highlights an urgent need for digital tools to serve as media for a deeper engagement with individuals' complex biopsychosocial needs. The integration of data science and technological advancements offers promising avenues for tailored digital solutions. The findings advocate the importance of continued innovation and adaptation in digital health care communication strategies, with clearer HP roles and an interdisciplinary, empathetic approach focused on individual needs.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e58434) doi:10.2196/58434

KEYWORDS

digital communication; digital technology; digital transformation; health professional; mental health; obesity; complex needs; artificial intelligence; AI; PRISMA

Introduction

Background

Obesity and mental health issues, two of the greatest global public health challenges [1], are intricately linked in a bidirectional and complex relationship that impacts both physical and mental well-being [2]. People experiencing depression are found to have a 58% increased risk of becoming obese [3]. Conversely, obesity is associated not only with physiological challenges but also social and mental health burdens [4], with persons who are obese facing an increased risk of developing depression over time. This relationship is particularly pronounced among people with serious mental illnesses (SMIs) such as schizophrenia or bipolar disorders, where obesity prevalence ranges between 45% and 55% [5]. Mental disorders account for the leading cause of global disability and determine how individuals make healthy choices [6], impacting personal well-being and quality of life [7]. The intertwined nature of obesity and mental health problems has a substantial impact on socioeconomic and health care systems [7]. The increasing prevalence of these conditions in society [8] calls for an urgent need to address this dual challenge to improve global health outcomes.

Health professionals (HPs) play an integral role in communicating mental health and obesity risks to the public [9]. Effective communication has been shown to positively impact the relationship between the HP and service users, engagement, and patient outcomes [10]. Communication regarding weight and mental health is especially challenging because communication training for HPs focuses on the medical aspects rather than the complexities of the conditions [11]. HPs require appropriate guidelines to identify users at risk of mental health issues [12]. Technology advances have the potential to transform health care, as evidenced by their critical role in communicating health messages during the COVID-19 pandemic [13]. According to Thomas [14], health communication is defined as "the systematic study and practical application of communication strategies in health, to inform and influence the attitudes, knowledge, and practices of individuals and communities." Within this domain, the co-occurrence of obesity and mental health issues presents a challenge. While there is an increasing acknowledgment of the importance of these interrelated health issues, a distinct gap exists in the literature. Currently, there is no review that addresses the intersection of mental health issues and obesity.

The rates of coexisting obesity and mental health issues are rising [15,16] owing to a multitude of physical, psychological, and environmental factors, such as medications, unhealthy diet, alcohol misuse, sedentary lifestyles, and low mood affecting adherence to weight management (WM) interventions [3,17-19]. Despite this growing concern, research studies often overlook the interplay between obesity and mental health issues, particularly in how these conditions are addressed and communicated by HPs. Current multidisciplinary guidelines for HPs dealing with obesity and mental illness recommend that they monitor and address patients' lifestyles; yet, research has found that most HPs have limited lifestyle-related knowledge and skills and that lifestyle treatment protocols are lacking [16]. There is a lack of understanding among HPs of how to integrate digital health solutions into everyday practices and procedures [20]. Existing research shows a need for obesity and mental health communication support for HPs. Obesity and mental health issues are acknowledged as a global health priority, but medical and sociocultural characteristics of the population are not taken into consideration [21]. Thus, research is required to explore the use of digital communication aids among HPs and the growing challenges presented by aging populations with multimorbidities.

Objectives

This scoping review aims to identify and understand the different types of technologies that are used by HPs in supporting obesity and mental health communication. This scoping review was guided by the following broad central research question: "What are the uses, benefits, and limitations of digital technologies in supporting communication between HPs and persons living with obesity and mental health issues?"

Methods

Scoping Review Framework

A scoping review applying a systematic approach was conducted as per the guidance of Munn et al [22] on choosing between different approaches when conducting evidence synthesis. This novel method was selected to address the broad research question guiding this review: to scope the literature for knowledge gaps [23] and gain conceptual clarity on the adoption of digital technologies in health care communication by thematically analyzing the available literature [24]. The framework developed by Arksey and O'Malley [25] guided this review, as recommended by Levac et al [23] and Daudt et al [26]. A scoping review applying a systematic approach was conducted, as per the guidance of Munn et al [22] on choosing between different approaches when conducting evidence synthesis, and the Cochrane Handbook for Systematic Reviews of Interventions [27] for systematic review methodology. This involved the following steps: (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, and (5) summarizing and reporting the results [25]. The review process applied the rigorous standards of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [28] (Multimedia Appendix 1) as the latest and most advanced approach for transparent and comprehensive reporting of scoping reviews.

XSL•FO
RenderX

Research Question

The purpose of this scoping review is to present the results of existing literature addressing the role of digital technologies used by HPs in facilitating communication with individuals living with obesity and mental health issues. The guiding research question was as follows: "What are the uses, benefits, and limitations of digital technologies in supporting communication between HPs and persons living with obesity and mental health issues?"

Search Strategy

The inclusion criteria followed the JBI scoping review guide [29], using *population*, *concept*, and *context*. The review included studies of all designs to identify the best evidence available to address the research objective. To support a thorough search, no time frame limit was applied, and all languages were included if translation was possible using Google Translate [30].

Eight electronic databases—Ovid MEDLINE, CINAHL, Scopus, ScienceDirect, PsycINFO, IEEE Xplore, ACM Digital Library, and ClinicalTrials.gov—were systematically searched in April 2023. These databases were chosen to capture articles in health care, psychology, and computer science fields. In addition, manual citation searching was performed to identify any relevant sources that might have been overlooked by the database searches.

Due to the novelty of the topic concerned, a broad search strategy was developed with high sensitivity to the following key terms: (1) "health professional," (2) "digital technology," (3) "obesity," and (4) "mental health." On the basis of the literature, benchmark definitions for these concepts were developed. A total of 144 synonyms were formulated for the search strings (Multimedia Appendix 2) with the guidance of a trained librarian. From the database searches and manual citation searching, 1656 hits were identified (Multimedia Appendix 3).

Study Selection

Following the framework outlined by Arksey and O'Malley [25] and aligning with the exploratory nature of scoping reviews, this study included primary research articles on digital technology interventions supporting HP communication with individuals facing obesity and mental health challenges. Exclusions were made for sources not directly related to HP communication support, such as those focused on technology use during HP training, which were deemed to be outside the scope of this review. In addition, conference papers, design protocols, and abstracts without full studies were excluded. The full eligibility criteria are presented in Textbox 1.

Textbox 1. Eligibility criteria for the scoping review.

Inclusion criteria

- Population: health professionals (HPs) and clients living with obesity and mental health issues
- Concept: studies that implement digital technology interventions aimed at enhancing communication processes
- Context: comorbidities of obesity and mental health issues
- Study characteristics: primary research studies; no time limits applied; all languages included if translation possible using Google Translate

Exclusion criteria

- Population: intervention without a focus on HP communication support
- Concept: digital technology intervention for clients with obesity and mental health issues, with no clear supporting role of HPs
- Context: broad or general comorbidities, not including obesity and mental health issues specifically
- Study characteristics: conference papers and design protocols

Articles identified from the systematic search were managed using the Covidence software platform (Veritas Health Innovation Ltd) [31]. Titles and abstracts were screened by 2 reviewers (AK and AM) independently (n=1078). The full text of relevant articles identified during the title and abstract screening was examined for eligibility against the inclusion and exclusion criteria by 2 independent reviewers (AK and AM). Ultimately, of the 1078 articles screened, 8 (0.74%) studies exemplifying the use of technology in HP communication for obesity and mental health issues were selected for analysis.

Charting and Extracting Data

Adhering to the JBI guidelines, a tailored data extraction template focused on study details, participant characteristics, interventions, and outcomes. The Covidence template was checked for suitability by 2 reviewers (AM and MM) before data extraction commenced. One author (AK) extracted data, which were verified by a second author (AM); disagreements were resolved by discussion. A thematic analysis identified key concepts related to HP communication technologies, including types of technology use, HP roles, biopsychosocial considerations, and perceptions regarding benefits and limitations in obesity and mental health contexts.

Summarizing and Reporting Results

The data charting process identified key themes, which were mapped around the biopsychosocial theoretical model developed by Engel [32] to facilitate an understanding of the results. Data synthesis involved comparing methodologies and results across the 8 studies to identify trends and gaps.

XSL•FO
RenderX

Quality Appraisal

The primary aim of the scoping review was to map out the diversity of technology use in HP communication; however, to enhance the reliability of the findings on technology use in HP communication for obesity and mental health issues, a detailed quality appraisal of the included studies was conducted. This involved using the revised Cochrane risk-of-bias tool for randomized trials (RoB 2) for randomized controlled trials (RCTs) [33] and the Risk of Bias in Nonrandomized Studies of Interventions (The Cochrane Collaboration) tool for nonrandomized studies [34]. This dual approach helped critically evaluate the evidence base, identifying studies with lower risks of bias that could provide more dependable insights into the effectiveness of digital technology interventions for managing obesity and mental health issues.

Eight original research studies integrating technology for HP communication support were critiqued for study design, implementation, and reporting [35-42]. The studies were rigorously assessed by 2 independent reviewers (AK and AM), using a grading key for risk of bias (ROB): (1) low ROB, (2) some concerns, and (3) high ROB.

Results

Selected Studies

This scoping review included 8 studies: 4 (50%) RCTs, 1 (12%) pilot RCT, and 3 (38%) qualitative feasibility studies. They were conducted in the United States (5/8, 62%), the United Kingdom (1/8, 12%), South Korea (1/8, 12%), and Iran (1/8, 12%; the percentages do not add up to 100% due to rounding). The study selection process is detailed in the PRISMA-ScR flow diagram (Figure 1), outlining the inclusion and exclusion steps taken during the literature review.

Key data from each study, including publication year, country, design, sample size, aim, context, demographic details, key findings, and limitations, were systematically charted (Table 1).

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram outlining the literature search and study selection procedures. HP: health professional.

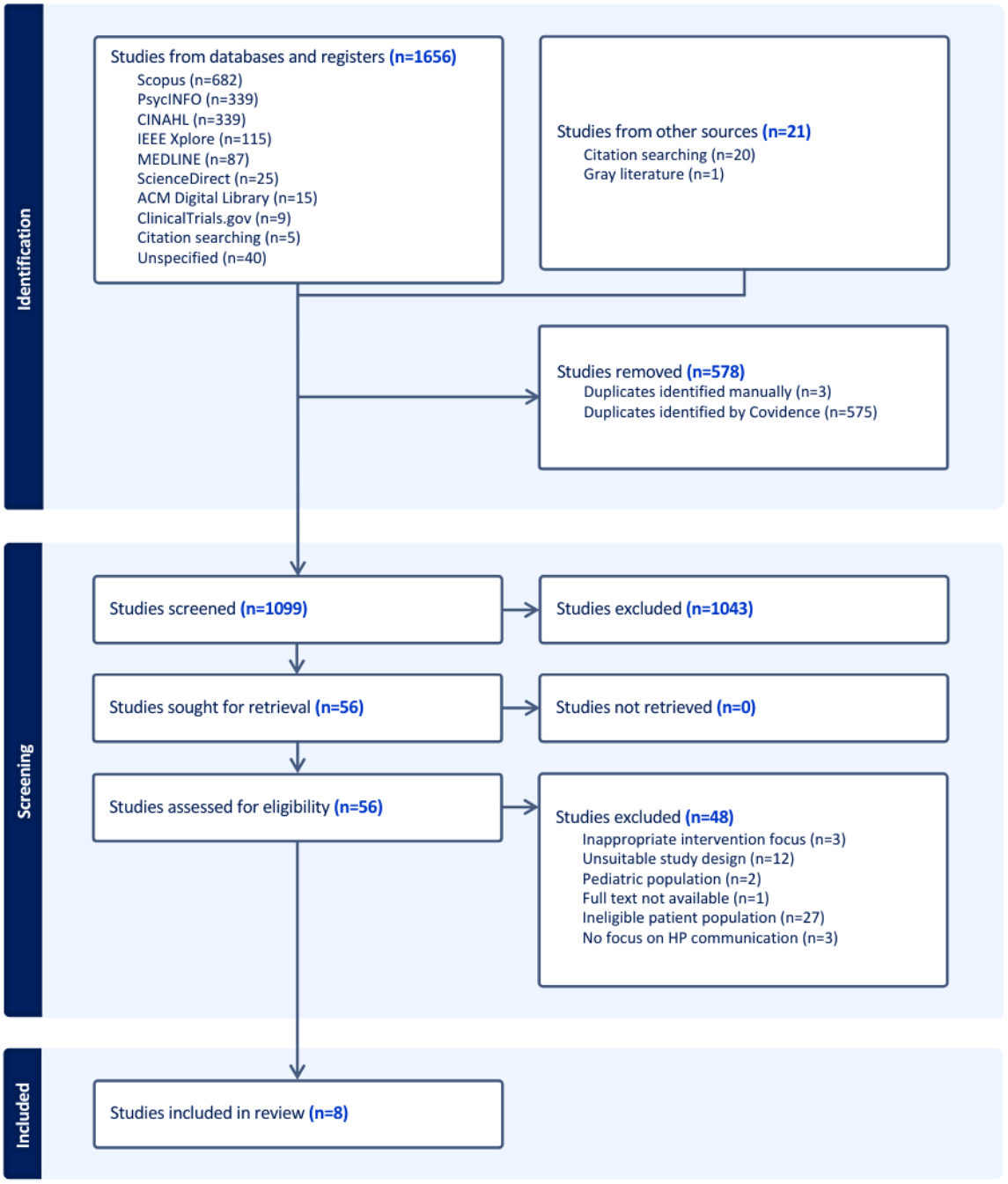

XSL•FO
RenderX

Table 1. Data extraction charting for the included scoping review studies (n=8)[a].

Study; year; country; study design	Aim	Participants, n; context; sex: %; age (y)	Key findings and conclusions	Comments
Lee et al [35]; 2023; United Kingdom; qualitative feasibility	Enhance web-based WM[b] program uptake for SMI[c] using VLC[d]	17; web-based, participants with SMI; female: 76; ≥18 (mean 48, SD not reported; range 29-70)	Promising engagement through digitally enhanced augmented WM program, complemented by brief, targeted education and low-intensity check-ins	Small size; potential bias
Bartels et al [38]; 2022; United States; hybrid RCT[e]	Compare web-based VLC vs TA[f] in the InSHAPE health promotion program for individuals with SMI	HPs[g]: 55; CMHC[h], with patients with SMI: 1219; female: % not reported; ≥18 (mean and SD not reported)	VLC as effective as TA in SMI obesity management	Sample representativeness; no cost data
Aschbrenner et al [36]; 2022; United States; 2-arm RCT	Evaluate PeerFIT mHealth[i] online Facebook group vs 1-on-1 mHealth (BEAT[j]) coaching for young adults with SMI who were overweight or obese	150; CMHC, young adults with SMI; female: 55; 18-35 (mean 28.46, SD 4.5)	PeerFIT and mHealth coaching equally effective, indicating that low-intensity mHealth coaching is a scalable, beneficial intervention option	Blinding issues; COVID-19 impact on retention; low engagement; lack of text or activity tracker data access; unexplored intervention format preferences
Nicol et al [41]; 2022; United States; qualitative feasibility	Test web-based iOTA[k] method (SMS text messaging and in-person coaching) for adults with SMI undergoing obesity treatment	26; CMHC, adults with SMI; female: 60; 16-75 (mean 48.5, SD 15.67)	iOTA method feasible; engagement varied; outcomes influenced by symptom severity, indicating a need for tailored adaptations	Small heterogeneous sample; implementation challenges; more comprehensive symptom assessments needed
Haddad et al [39]; 2022; United States; qualitative feasibility	Adapt mHealth obesity treatment for adults with SMI using web-based SMS text messaging I-Corps[l] methods	29; CMHC, various stakeholders, participants with SMI unspecified; sex not reported; age not reported	Feasibility of adapted mHealth obesity treatment; stakeholder insights: need for microadjustments in clinical settings and clarity in health coaching responsibilities	Limited consumer input; reporting bias; no specific differentiation in SMI diagnosis
Abedishargh et al [37]; 2021; Iran; 3-arm RCT	Assess ICBT[m] on WhatsApp for mental health issues of women with overweight	90; online support group, participants with depression or anxiety; female: 100; 18-50, 38.2% 18-30 (mean and SD not reported)	ICBT reduced BMI but not mental health issues	No follow-up; pandemic influence on short duration of therapy; online implementation challenges; sample selection process not detailed
Yu et al [42]; 2021; United States; pilot RCT	Evaluate videoconferencing-based treatment for BED[n]	18; web-based, participants with obesity with BED; female: 100; 20-73 (mean 32, SD 15.27)	Videoconferencing-based treatment improved shape concern but not weight	Small size; high dropout rate; technical issues; potential self-reported bias
Kim et al [40]; 2020; South Korea; 2-arm RCT	Test digital CBT[o] for WM in adults with depression or anxiety using intensive, multifactorial, daily psychologist coaching	70; online, participants with depression or anxiety; female: 100; 18-39 (mean 21.8, SD 3.3)	Effective CBT integration in digital WM combined with multidisciplinary human support	Generalizability; self-report inaccuracies; feasibility: intensive and costly, requiring daily intervention by trained therapists

[a]Summary: study design: 4 randomized controlled trials (RCTs), 1 pilot RCT, and 3 qualitative feasibility studies; conducted in the United States (5/8, 62%); aim: assessed digital interventions including web-based virtual learning collaborative dashboards, SMS text messaging, and online support groups for improving engagement and effectiveness in serious mental illness and obesity; participants: n=17–150; contexts included community mental health centers (50%) and online communities (50%); primarily female, 16–75 years (mean 35.8, SD 9.68; % female unknown in some studies); key findings: interventions improved engagement and effectiveness but faced limitations like small samples, biases, and generalizability issues; comments: limitations included small sample sizes, biases, and feasibility concerns.

[b]WM: weight management.

[c]SMI: serious mental illness.

[d]VLC: virtual learning collaborative.

[e]RCT: randomized controlled trial.

[f]TA: technical assistance.

[g]HP: health professional.

[h]CMHC: community mental health center.

[i]mHealth: mobile health.

[j]BEAT: Basic Education Supported by Activity Tracking.

[k]iOTA: interactive obesity treatment approach.

XSL•FO
RenderX

[l]I-Corps: Innovation Corps.

[m]ICBT: internet-based cognitive behavioral therapy.

[n]BED: binge eating disorder.

[o]CBT: cognitive behavioral therapy.

Participant Characteristics

The participant pool varied, ranging from 17 to 150 participants per study. Half of the studies (4/8, 50%) were conducted at community mental health centers [36,38,39,41], and the other half (4/8, 50%) involved web-based recruitment of the general population [35,37,40,42]. Most participants were female, with sex data reported in 75% (6/8) of studies (range 55%-100%). Recruitment ages ranged from 16 to 75 years, with a mean age of 35.8 years. Among the 4 studies that reported SDs (50%), the average SD was 9.68 years (Table 1). The studies used various tools to assess mental health conditions such as SMI, depression, anxiety, and binge eating disorder, providing a diverse insight into mental health evaluations in the context of digital technology use in health care (Table 2).

Table 2. Participant mental health (MH) characteristics and assessment tools in the included studies (n=8).

Participant criteria for MH issues	Studies, n (%)	Specific MH assessment tools	Tools used (n=14), n (%)
SMI[a] (as defined by, eg, schizophrenia, schizophreniform disorder, schizoaffective disorder, bipolar disorder, or depression with psychosis) [41]	1 (12)	• Clinical Global Impression severity score of >5	1 (7)
Depression, anxiety (with stress or self-esteem issues) [37,40]	2 (25)	• Situational Motivation Scale • Korean version of the Beck Depression Inventory-II • Taylor Anxiety Inventory • Rosenberg Self-Esteem Scale • Automatic Thoughts Questionnaire-30 • Depression Anxiety Stress Scales	6 (43)
BED[b] [42]	1 (12)	• Eating Disorder Examination Questionnaire 6.0 • Eating Attitudes Test-26 • Three-Factor Eating Questionnaire–Revised 18 • Yale Food Addiction Scale • Symptom Checklist-90–Revised • Beck Depression Inventory-II • Impact of Weight on Quality of Life–Lite	7 (50)
SMI but no specified assessment tools, that is, no specificity of MH issues [35,39] or measured fidelity and cardiovascular risk reduction only [36,38]	4 (50)	• None	0 (0)

[a]SMI: serious mental illness.

[b]BED: binge eating disorder.

Uses of Digital Technologies in Obesity and Mental Health Communication

The investigation into the use of digital technologies across the 8 included studies highlights their role in addressing obesity and mental health issues. These technologies facilitated virtual appointments, diet and mental health monitoring, and the provision of motivational and educational content. In terms of professional use, psychologists or counselors were consistent users across all studies [35-42], with additional involvement from dietitians [38,42], social workers [36,39], and trained health coaches [35,38] (Figure 2 [35-42]).

XSL•FO
RenderX

Figure 2. Mapping of the uses of digital technologies and key themes from the 8 included studies [35-42] regarding the role of digital technologies in supporting communication between health professionals (HPs) and participants with obesity and mental health issues. WW: WeightWatchers.

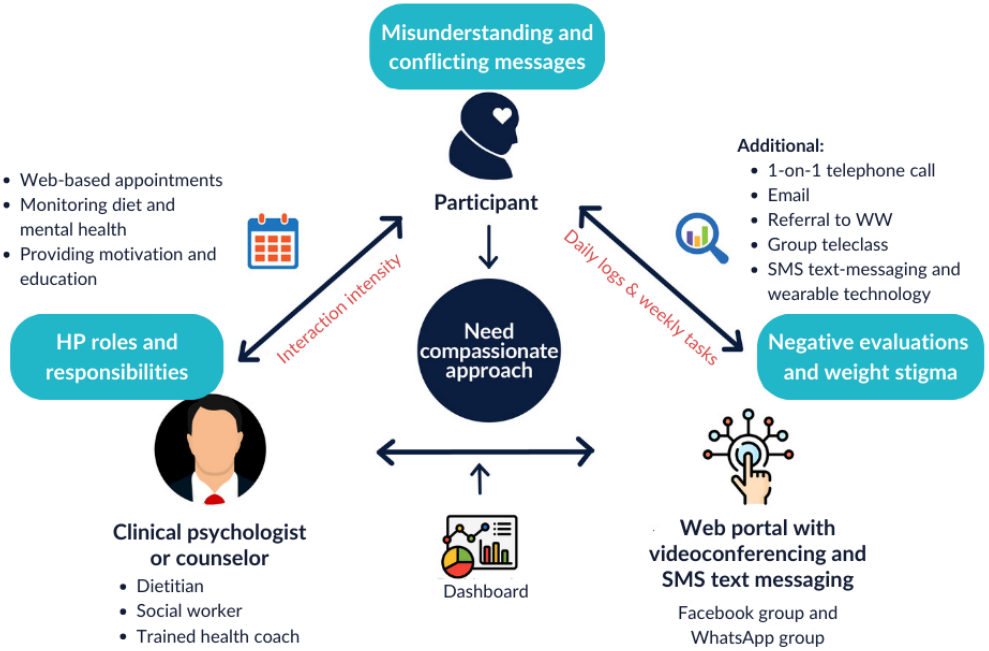

The primary digital interventions varied. They included web-based platforms with synchronous video calls (4/8, 50%), automated SMS text messaging (2/8, 25%), and social media (Facebook and WhatsApp groups; 2/8, 25%). Various supplementary support methods were used, including 1-on-1 telephone calls (3/8, 38%), email (3/8, 38%), group teleclasses (2/8, 25%) [37,42], referrals to WeightWatchers (2/8, 25%), and the integration of SMS text messaging with wearable technology (1/8, 12%). The most frequently reported tool was the virtual learning collaborative (VLC) dashboard with its videoconferencing feature [35,38,40,42]. The most common supplementary support methods used were 1-on-1 telephone calls [35,38,40] and email [35,38,42].

The 8 studies illustrated diversity in their approach to digital interventions, with 4 (50%) focusing on group therapy, 2 (25%) on 1-on-1 interventions, and 2 (25%) using a hybrid model that combined both methods. HP interaction frequencies varied, with the most common being daily interactions (3/8, 38%). Typical session durations were either 1 hour (2/8, 25%) or 1.5 hours (2/8, 25%), and the interventions spanned an average duration of 3 months (3/8, 38%, SD 1.63 months), with the total number of sessions ranging from 4 to 52 sessions with an HP (Table 3).

XSL•FO
RenderX

Table 3. Data extraction results for the included studies on digital interventions and health professional (HP) communication (n=8)[a].

Study; year	Intervention name	Secondary support	HPs involved	Contact time	Interaction	Duration (mo)	Sessions, n	Adherence	Theory
Main digital intervention: web portal VLC[b] dashboard with videoconferencing									
Lee et al [35]; 2023	WHEEL[c]; Meet Your Mentor (single 1.5-h, web-based, 1-on-1 consultation)	Telephone call or email mentor check-ins plus referral to WW[d] group	Psychologist	90-min plus 15-min interim booster support	Weekly (fixed day and time) for 12 wk	3	12	Good	PBA[e] (based on the MRC[f] model)
Bartels et al [38]; 2022	InSHAPE VLC	Telephone call or email mentor check-ins plus referral to WW group	Trained fitness instructor receives instruction from health psychologist, dietitian, and senior-level InSHAPE program manager and trainer	4-d InSHAPE training plus 90-min follow-up sessions	Monthly	12-18	50	Good	VLC
Yu et al [42]; 2021	Fruit Street videoconferencing program	Group teleclass and email	Psychologist and dietitian	1 h	Weekly	3	12	Good	CBT[g]
Kim et al [40]; 2020	Noom coach and InBody	1-on-1 telephone calls	Psychologist	Weekly group call plus 15-min interim booster support	Daily	2	8	Good	CBT+MI[h]
Main digital intervention: SMS text messaging									
Haddad et al [39]; 2022	Innovation Corps iOTA[i]	—[j]	Psychologist, prescribing physician, nurse, and social worker	—	—	—	—	Not reported	VLC
Nicol et al [41]; 2022	Working for You iOTA	—	Trained health coach at mental health clubhouse; no specific roles	—	Monthly	3	4	Adequate	Chronic care model
Main digital intervention: Facebook or WhatsApp groups									
Aschbrenner et al [36]; 2022	PeerFIT	SMS text messaging and wearable technology	Psychologist, counselor, social worker, or fitness professionals	1 h	Daily	12	52	Poor	SCT[k]+DPP[l]
Abedishargh et al [37]; 2021	Weight loss	Group teleclass	Psychologist plus peer-led interventions	2 h	Daily	1.5	42	Good	CBT

[a]Table 3 summarizes various digital interventions (web portals, virtual learning collaborative [VLC], SMS text messaging, and social media), with mixed delivery methods (1-on-1 telephone calls, mentor check-ins, and group classes) and diverse health professionals' involvement (psychologists, dietitians, social workers, and fitness instructors). Contact time ranged from 15-minute booster sessions to 90-minute consultations, with interaction frequencies varying from daily to monthly. Intervention durations ranged from 1.5 to 18 months, with 4 to 52 sessions. Adherence varied, and theories included cognitive behavioral therapy, VLC, motivational interviewing, social cognitive theory, Diabetes Prevention Program, and person-based approach.

[b]VLC: virtual learning collaborative.

[c]WHEEL: Weight Change for People With Serious Mental Illness.

[d]WW: WeightWatchers.

XSL•FO
RenderX

[e]PBA: person-based approach.

[f]MRC: Medical Research Council.

[g]CBT: cognitive behavioral therapy.

[h]MI: motivational interviewing.

[i]iOTA: interactive obesity treatment approach.

[j]Not applicable.

[k]SCT: social cognitive theory.

[l]DPP: Diabetes Prevention Program.

Content-wise, the digital interventions drew upon a range of theoretical frameworks and strategies, including cognitive behavioral therapy (CBT; 2/8, 25%), CBT plus motivational interviewing (1/8, 12%), VLC (2/8, 25%), the person-based approach following the Medical Research Council model (1/8, 12%), the chronic care model (1/8, 12%), and the social cognitive theory supplemented with Diabetes Prevention Program content (1/8, 12%; Table 3).

Quality Appraisal

The studies varied in quality. Of the 5 RCT studies, 2 (40%) were rated as high quality with low ROB [38,40], and 3 (60%) were deemed to have an acceptable ROB [36,37,42] based on the ROB 2 (Table 4), while the qualitative feasibility studies were consistently rated as high quality and low ROB [35,39,41] based on the Risk of Bias in Nonrandomized Studies of Interventions tool (Table 5).

Table 4. Quality assessment of 5 included randomized controlled trial studies using the Cochrane risk-of-bias tool for randomized trials (ROB 2)[a].

Study; year	Risk of bias: methods			Risk of bias: results		Overall risk of bias
	Study design	Participants	Assessments and data collection	Reporting quality	Clinical efficacy	
Bartels et al [38]; 2022	++[b]	++	++	++	++	++
Aschbrenner et al [36]; 2022	++	+[c]	+	+	+	+
Abedishargh et al [37]; 2021	++	++	+	+	+	+
Yu et al [42]; 2021	++	+	+	+	+	+
Kim et al [40]; 2020	++	++	++	++	++	++

[a]The studies demonstrated a low risk of bias in methods and results overall, though some concerns were noted in assessments, reporting quality, and clinical efficacy. Variability in risk levels was observed across study design and participant selection

[b]++: high quality and low risk of bias.

[c]+: acceptable risk of bias.

Table 5. Quality assessment of 3 included qualitative feasibility studies using the Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I) tool[a].

Study; year	Risk of bias						
	Confounding	Participant selection	Intervention classification	Deviations from intended interventions	Missing data	Measurement of outcomes	Selection of reported result
Lee et al [35]; 2023	Low	Low	Low	Low	Low	Moderate	Low
Nicol et al [41]; 2022	Moderate	Low	Low	Moderate	Low	Moderate	Low
Haddad et al [39]; 2022	Moderate	Low	Low	Moderate	Low	Low	Low

[a]Across the studies, deviations from intended interventions were assessed as moderate risk of bias, while low risk of bias was consistently noted for participant selection, intervention classification, missing data, and selection of reported results. Risk of bias for outcome measurement was moderate in 67% (2/3) of studies.

Benefits of Digital Technologies in Obesity and Mental Health Communication

The digital interventions provide benefits such as improved access to personalized care via digital CBT, tailored feedback, and enhanced physical health outcomes, including weight loss and BMI improvement. The study by Haddad et al [39] unveiled that some service users had previous negative in-person interactions with HPs, perceiving limited benefit relative to the "additional burden of time and mental energy" invested, and some participants preferred online intervention delivery because it helped reduce stigma (Multimedia Appendix 4 [35-42]). The digital format offered a supportive and private environment conducive to client adherence, motivation, engagement, and

XSL•FO
RenderX

comfort, thereby reducing the stigma associated with seeking help [36,37]. Participants with mood disorders reported experiences of shame and negative self-evaluation [33], showing the importance of addressing self-esteem issues within digital health platforms. Personalization and engagement emerged as key factors in the success of digital interventions [39]. Programs with digital CBT, adapted to individual characteristics and supported by intensive daily monitoring, were shown to improve participant engagement and adherence with minimal dropout rates: as noted by Bartels et al [38], "only one drop out for personal reasons, and 80% remained active until the end of the treatment session. Authors owe this to motivation construct."

Efforts to enhance the implementation and effectiveness of digital interventions include comprehensive training for health mentors, involving psychologists, dietitians, and program managers [37]. This indicates the significance of a multidisciplinary approach and the need for sustainable funding and training solutions: and training solutions. For instance, Abedishargh et al [37] highlighted the challenge of "finding a funding source to pay for a dietitian," while Bartels et al [38] emphasized the importance of "training changes and sustainability of a wellness program." The collaboration within VLC programs, intensive skill-building training, and the involvement of mentors and leadership emphasize the value of mentorship and a compassionate approach in program delivery [36]. Significant improvements in WM were observed in 7 (88%) of the 8 studies [35-41], showing the efficacy of digital interventions in supporting better physical health outcomes.

The fact that there was no significant improvement in mental health outcomes suggests a potential inadequacy in the ability of the digital interventions to simultaneously address mental health needs and obesity. Yu et al [42] observed no significant changes in weight and binge eating episodes among participants. Conversely, Nicol et al [41] demonstrated that individuals with less severe symptoms and higher engagement achieved statistically significant weight loss (P<.001) compared to those with more severe symptoms. Nonetheless, clarity regarding the inclusion criteria for individuals with SMI was lacking in the studies, and there was inconsistency in the measurement of outcomes related to mental health status (Table 2). These findings indicate a need for further research that adopts a comprehensive and multidisciplinary approach to digital health interventions.

Limitations of Digital Technologies in Obesity and Mental Health Communication

Challenges in implementing digital technologies include misinterpretations of SMS text messaging prompts [41], variability in user engagement, and adherence [35,36,38,42]. These challenges highlight the importance of flexible and adaptable interventions, personalized to individual lifestyles and preferences, and a clear understanding of HP roles in digital interventions. The recurrent call for dietetic expertise [38-40,42] identifies a gap in incorporating comprehensive nutritional guidance within these interventions, advocating for a multidisciplinary strategy that includes dietitians to offer balanced and scientifically sound dietary advice tailored to the needs of individuals managing obesity and mental health issues (Table 6).

Table 6. Key further research questions identified from the scoping review.

Research question	Relevance	Directions for future research
How can digital interventions be tailored to better address the mental health needs of individuals with comorbid obesity and mental health issues?	Psychosocial	Investigate adaptive features and content within digital interventions to provide comprehensive mental health support alongside obesity management.
How can personalized care be enhanced in digital interventions for obesity and mental health care?	Psychosocial	Examine methods for deep personalization in digital interventions, considering the complexity of comorbid conditions and individual user needs, focusing on unique needs, engagement, and outcomes.
What are the optimal strategies for enhancing user engagement and adherence to digital interventions?	Psychosocial	Develop and test engagement strategies that account for individual preferences for, and barriers to, sustained use of digital health tools, while incorporating robust adherence measures.
How can digital interventions be designed to support a more compassionate and empathetic approach to obesity and mental health care?	Psychosocial	Explore the integration of features that enhance empathy, compassion, and effective communication between clients and HPs[a] in digital health platforms, promoting patient satisfaction, trust, and improved outcomes.
What are the optimal roles and responsibilities of HPs in digital interventions to ensure high fidelity and participant retention?	Social	Determine the best practices for multidisciplinary collaboration in digital health interventions, optimizing HP roles, involvement, and training needs.

[a]HP: health professional.

Variability in intervention intensity and the roles of HPs present challenges in achieving effective program outcomes. The comparison between a nutritionist-led "treatment as usual" group and a psychologist-led internet-based CBT intervention highlights this issue [37]. The daily delivery of internet-based CBT demonstrated a more substantial reduction in BMI compared to the weekly sessions conducted by nutritionists, suggesting that the interaction frequency rather than the professional role might influence outcomes. Nonetheless, neither approach significantly impacted participants' stress, anxiety, or depression levels [35]. Furthermore, low adherence to tools such as food diaries, perceived as the "least helpful component"

XSL•FO
RenderX

[42], indicates that such tools may trigger feelings of shame or self-criticism among users. Future research should explore the integration of 1-on-1 technical assistance and peer-group problem-solving to address these challenges [38], emphasizing the need for personalization, improved communication, empathetic care, and the integration of multidisciplinary care teams to address both physical and mental health issues in digital health care settings.

The studies collectively revealed certain key themes: misunderstanding and conflicting messages, negative evaluations and weight stigma, the need for a compassionate and empathetic approach, and confusion about HP roles and responsibilities in the context of obesity and mental health care (Table 3). These themes, along with detailed narratives, benefits, and limitations associated with obesity and mental health communication derived from the 8 included studies, are documented in Multimedia Appendix 4. This compilation aligns with the biopsychosocial theoretical model developed by Engel [32], illustrating the integrated approach needed to address these issues effectively.

Discussion

Overview

This scoping review reveals a significant research gap in the application of digital technologies for health care communication between HPs and individuals with comorbid obesity and mental health issues [43] (Figure 3). Despite the extensive search strategy, only 8 studies were retrieved for this scoping review, showing the scarcity of published research. Significantly, these papers were published after 2020, highlighting the novelty of the interdisciplinary research focus and emerging interest after the COVID-19 pandemic, which has accelerated the integration of digital technologies into health care settings. This trend is observable across various countries, emphasizing the universal shift toward digital modalities in addressing complex health care needs associated with obesity and mental health, particularly in the postpandemic era.

Although prior studies have explored the use of technology in the realms of obesity [17,44,45], no research has yet investigated its role in managing comorbid obesity and mental health communication. This review suggests the necessity for more primary research to evaluate whether digital technologies can enhance health communication practices, improve care, and engage users; however, it also highlights a lack of existing studies addressing the combined challenges of obesity and mental health issues. Further research should aim to understand how technological advances can support communication between HPs and individuals living with obesity and mental health issues to deliver person-centered care that meets the biopsychosocial needs of this population [20]. Figure 4 [35-42] portrays the complexities of these biopsychosocial factors and the mapping of the key research themes from the 8 included studies to the biopsychosocial theoretical model.

Figure 3. The role of digital technologies in supporting bidirectional communication between health professionals and persons living with obesity and mental health issues.

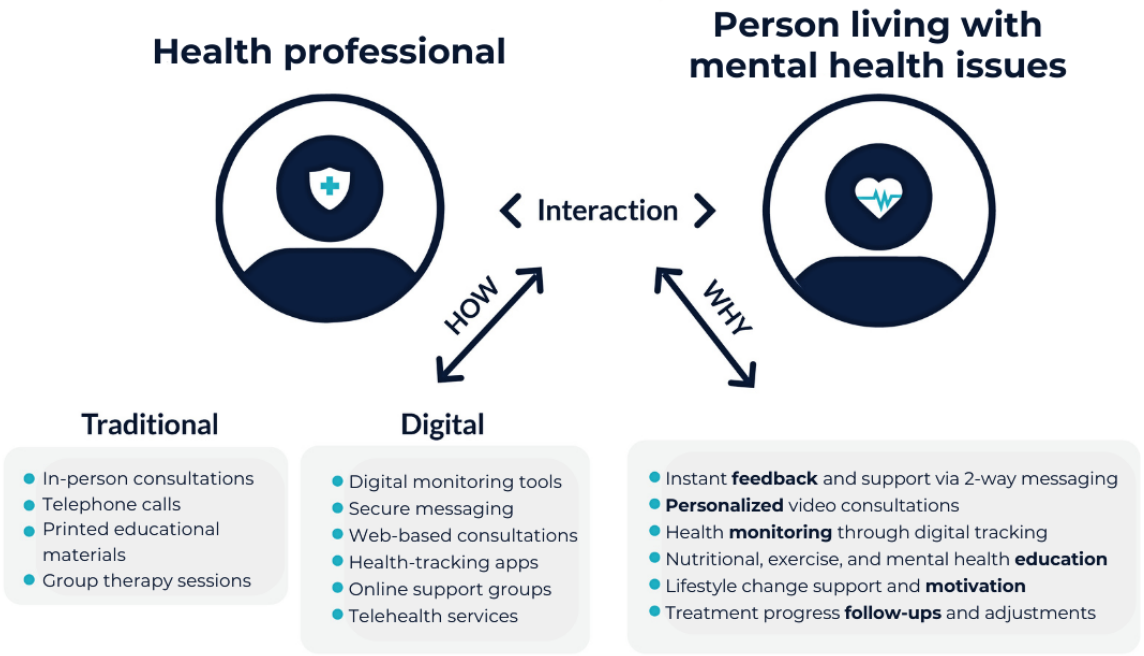

XSL•FO
RenderX

Figure 4. Mapping of the key themes related to obesity and mental health complexities from the 8 included studies [35-42] to the biopsychosocial theoretical model. EQ: emotional quotient; HP: health professional.

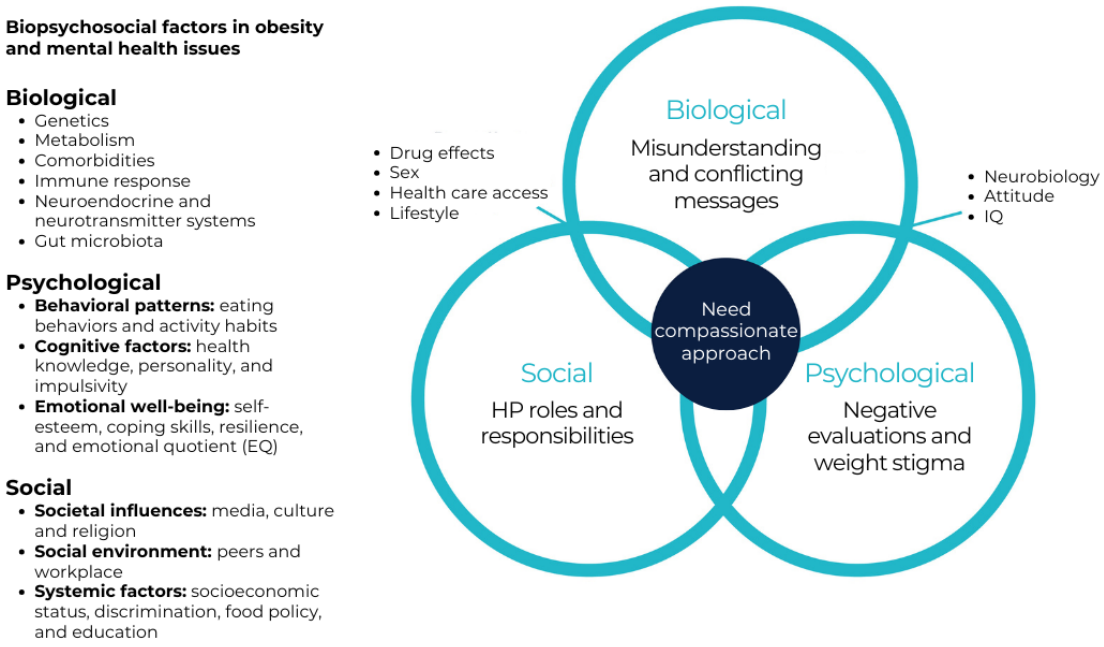

Uses of Digital Technologies in Health Care Communication Among Those With Obesity and Mental Health Issues

Although research in this area remains limited, the available evidence suggests that digital platforms have a growing role in bridging the communication gap between HPs and clients, especially in a post–COVID-19 era marked by disruptions in physical access to HPs [46]. Digital platforms, including those providing web-based interfaces, social media networks, and synchronous video call features (Figure 3), facilitate services such as virtual appointments and health monitoring across different professional domains. The findings complement the World Health Organization report [47], highlighting that tele–mental health has emerged as an effective medium for delivering education and psychological treatments, offering an alternative or supplementary approach to traditional therapies [46].

However, this review reveals variability in participant preferences for intervention delivery, highlighting the potential benefits of combining regular HP visits with extended mobile health coaching [35,40,41]. In terms of the frequency and scheduling of HP sessions, some participants found that weekly scheduled telephone check-ins were helpful [35,42], whereas others wanted more flexibility around the competing demands of family or work commitments [35,36,42]. Such clarity is vital for ensuring a comprehensive and unified approach to treating obesity and mental health issues, supporting a more cohesive and integrated care experience for participants.

Benefits of Digital Technologies in Obesity and Mental Health Communication

Digital technologies enhance obesity and mental health care by offering web-based counseling, personalized care, and tailored interventions [35,41], facilitating a more effective response to individual needs. This personalized approach is significant in ensuring that each individual's unique health journey is adequately considered and managed (Multimedia Appendix 4). A key element in the success of digital health care interventions is the formation of a therapeutic alliance within web-based settings. The "Meet Your Mentor" program exemplifies the advantages of digital health care, providing a platform for clients to share their stories in a private, in-depth, initial, 1-on-1, 1.5-hour consultation [35]. This indicates the value of individualized attention in digital interventions to facilitate a deeper connection and understanding between the client and the provider.

The shift toward digital health care platforms has significantly reduced the necessity for in-person interactions, mitigating the stigma and negative perceptions often linked to certain health conditions [48]. This change has enhanced client engagement by alleviating the discomfort of activities such as physical weigh-ins, which are more daunting in a group setting [35]. Digital modalities provide anonymity, supporting an environment that encourages open and honest communication, essential for overcoming the challenges of the stigma prevalent in traditional settings [48]. Prior studies support the effectiveness of digital health care in promoting better client-provider interactions through the privacy and detachment it offers compared to face-to-face consultations [35,49].

Furthermore, digital platforms support social learning and compassionate care through peer-to-peer interactions [36,37],

XSL•FO
RenderX

promoting mutual support and understanding among participants [50]. Digital technologies have been effectively used to maintain connectivity and engagement between participants, with higher adherence observed in interventions that were tailored to unique individual needs [35,40]. A recent scoping review identified the need to explore effective communication techniques that could be adopted to maintain interpersonal client-provider connections in digitally enabled health care environments [49]; however, the evidence is still unclear, and no studies focus on populations with obesity and mental health issues. This review also identifies a need for more comprehensive interventions to simultaneously address the complex challenges posed by comorbid mental health issues and obesity.

Limitations of Digital Technologies in Obesity and Mental Health Communication

Despite the potential of digital technology use for health care communication in the context of obesity and mental health issues, there are certain limitations, as this review reveals. Variability in adherence to digital interventions suggests that not all components are equally effective across different settings. The COVID-19 pandemic restrictions further highlighted engagement and retention issues [36,37], showing the need for adaptable research designs to accommodate unpredictable circumstances [36,41]. Issues such as the inconsistent use of self-reported food diaries [40,42] point to the need for interventions that consider individual preferences and behaviors (Table 6). Nicol et al [41] reported the need for more comprehensive symptom assessments, while Aschbrenner et al [36] noted poor adherence, suggesting that young adults prefer individually tailored mobile health coaching over group-based sessions. The variability in engagement and adherence, especially in group-delivered sessions, points to a need for more customized approaches that consider individual preferences and lifestyles.

The current body of evidence identifies a gap in addressing the social complexities associated with comorbid obesity and mental health issues. The studies included in this review lack sufficient detail on communication preferences within interventions [38], highlighting the need for research into effective, compassionate communication strategies to achieve person-centered care [49]. The complexity of various HP roles involved in implementing digital technologies, compounded by conflicting messages and weight stigma [44,51,52], highlights the necessity for empathy and an understanding of the unique challenges faced by individuals with comorbid conditions. Recent research has explored compassionate digital health care [49] but not specifically in the context of obesity and mental health issues.

Future research should focus on tailoring digital interventions to meet the mental health needs of those with comorbid conditions, enhancing personalization, developing engagement strategies, incorporating compassionate care approaches, and defining optimal roles for HPs in digital settings (Table 6). These areas represent critical steps toward improving the effectiveness of digital health care interventions for individuals facing the dual challenges of obesity and mental health issues.

Interdisciplinary and Person-Centered Care

The integration of HP disciplines and the use of diverse digital technologies in interventions (Table 3) align with the biopsychosocial model of health [32], offering holistic care for those with comorbid obesity and mental health issues. Government initiatives, such as the UK Department of Health's Mental Health Strategy 2021-2031, support digital, multidisciplinary, person-centered care [53-55]; yet, organizational barriers often prevent effective care delivery [39]. This review identified a potential disconnect between policy recommendations and their practical implementation in health care settings [39], while the recollection of weight stigma experiences, known to harm mental and physical health, underscores the need for action at the policy level [51].

This review points to a lack of technology acceptance theoretical models in current research [4], suggesting a need for frameworks, such as the Unified Theory of Acceptance and Use of Technology 2 formulated by Venkatesh et al [56], to understand the adoption of digital technology in obesity and mental health care. The studies included in this review did not use advanced technologies such as artificial intelligence and virtual reality (VR) [53,57,58], which could offer personalized immersive interventions and create new avenues for engagement. Innovative treatment modalities such as exposure therapy in VR [59] and detailed behavioral data collection through digital phenotyping [57] could help better understand and treat mental health conditions. This disparity suggests a gap in using emerging technologies to enhance the quality and reach of digitally supported health care services [53].

The interdisciplinary nature of interventions aimed at improving physical and mental health among individuals with SMI is evident in the collaborative efforts across various disciplines, including mental health services, fitness professionals, and digital technology experts [35,36,38-40]. Programs such as InSHAPE and PeerFIT demonstrate the benefits of community-based partnerships and professional diversity [36,38] by providing health-mentoring sessions at local community fitness facilities. Integrating individuals with SMI into the broader community promotes social inclusion and reduces stigma. However, this review also points to a deficiency in compassionate communication through digital health care, emphasizing the necessity for strategies that support empathy and understanding in treating obesity and mental health issues.

The PeerFIT and Weight Change for People With Serious Mental Illness interventions illustrate the value of interdisciplinary collaboration, engaging professionals from various backgrounds, such as mental health, fitness, psychology, and social work, trained in health coaching [35,36]. These initiatives, alongside the stakeholder engagement strategy proposed by Haddad et al [39] show the effectiveness of integrating diverse expertise and community involvement in health interventions. Such approaches ensure that the interventions are responsive to the real-world needs of individuals with SMI, leveraging digital technology for more accessible, inclusive health care solutions. This emphasizes the need for interdisciplinary and person-centered principles in developing next-generation communication tools, aiming to

XSL•FO
RenderX

comprehensively address the multifaceted needs of those with comorbid conditions (Table 7).

Recent research has acknowledged that technology use influences how HPs perform their work and interact with clients [49]; however, little is known about how compassion is conveyed through digital health care modalities, especially in the context of obesity and mental health care. Despite the widespread adoption of digital health technologies and their potential to impact client outcomes positively [49,55,57,60], strategies for compassionate digital communication for complex health issues are underexplored, signaling a need for research on compassionate, digital approaches for obesity and mental health support (Table 7).

This review emphasizes the need for sociocultural sensitivity in tailoring messages and goals in digital interventions. While most of the included studies reported improvements in physical health outcomes such as weight and BMI, the lack of significant progress in mental health outcomes was a significant finding. This discrepancy raises questions about the comprehensiveness of these interventions in addressing the dual challenges of obesity and mental health issues [42]. Individuals with mental health issues may have trouble engaging with routine WM interventions, and bespoke options are rarely offered in routine practice [35].

Table 7. Proposed requirements for next-generation communication tools based on the findings from the reviewed studies (n=8).

Requirement	Description	Justification
Comprehensive care	• Content addressing both physical and mental health needs, including modules for digital CBT[a], stress management, dietary tracking, and physical activity monitoring with real-time feedback and support	• Ensures holistic care by addressing the intertwined nature of physical and mental health, improving overall well-being
Personalization and tailoring	• Customized content and dynamic intervention adjustments using AI[b] and machine learning based on user data (eg, behavior, cognition, and emotion)	• Personalization and feedback based on individual user data (eg, behavior, cognition, and emotion) have been shown to improve engagement and adherence [39,40]
Enhanced engagement and adherence	• Engaging interfaces, gamification, reminders, incentives, and progress-tracking features	• Increases user motivation and sustained tool use, crucial for long-term health management
Compassionate design	• Features promoting empathy, including personalized messaging, web-based support groups, videoconferencing, secure messaging, and forums for peer support • Training for HPs[c] on compassionate care delivery	• A compassionate approach addresses psychological and social needs in health care, supporting a sense of community and belonging [35,36,38,42]
Multidisciplinary support	• Tools for HPs to monitor progress, provide feedback, and collaborate across disciplines	• Enhances care fidelity and patient retention through comprehensive, coordinated care approaches [38,39,42]
Stigma reduction	• Anonymous participation options and educational content to counteract stereotypes and misinformation	• Reduces barriers to seeking help by mitigating stigma and promoting a more inclusive care environment
Anonymity and privacy	• Ensuring user anonymity and privacy to reduce fears of judgment and stigma	• Encourages wider participation by providing a secure and confidential user experience. Anonymity encourages participation by mitigating concerns about judgment [35,37]
Integration with existing health care systems	• Compatibility with health care platforms for streamlined care coordination. Adaptability across health care settings and comprehensive training for users and providers	• Seamless integration with health care systems can enhance the efficacy and sustainability of digital interventions [37,38] • Compliance with health care regulations and data protection standards
Technical assistance	• Continuous support and education for users and providers to maximize digital tool benefits and adapt to updates	• Ensures that users and providers can effectively use the full range of tool capabilities, enhancing overall intervention success

[a]CBT: cognitive behavioral therapy.

[b]AI: artificial intelligence.

[c]HP: health professional.

Digital technologies play a vital role in extending health care reach, especially to communities considered marginalized, by providing anonymous, unbiased access to services, thus reducing pressure on health care systems [60]. They serve as an accessible

XSL•FO
RenderX

bridge for support while individuals await further assistance [57,60]. However, achieving their full potential requires ongoing innovation, tailored designs, and an inclusive and interdisciplinary strategy that equally addresses physical and mental health [12,18,58,61].

To address these gaps, future research should explore more extensive interventions with larger participant groups and extended follow-up assessments to gauge the sustained impacts of digital interventions. Customizing digital tools to fit individual needs, aligned with policy support, is essential [47,54]. These tools should facilitate deeper engagement with the complex needs of those facing both obesity and mental health challenges, moving beyond mere information dissemination. High-quality data use could improve direct client interactions [57], although the included studies lack detailed insights into the specific characteristics and preferences of this demographic, highlighting a gap in effective digital communication strategies for HPs catering to individuals with comorbid obesity and mental health issues.

Limitations

Study limitations included the small sample of 8 selected studies, with limited sample sizes (ranging from 17 to 150 participants), which may limit the generalizability of the findings. A common reliance on self-reported data, such as food diaries, introduces susceptibility to inaccuracies and social desirability bias. Of the 8 included studies, 7 reported intervention durations. Most (6/7, 86%) were of short duration (≤12 weeks) and lacked long-term follow-ups, limiting insights into the sustainability of outcomes. One study did not specify its duration, highlighting the need for standardized reporting. The interventions noted feasibility challenges related to the labor intensity required from HPs and implementation costs [38]. The predominance of female participants and varied study designs, including RCTs and feasibility studies, offer varied perspectives but also raise questions about the broader applicability of the results.

Digital interventions, while beneficial, risk exacerbating social isolation or excluding certain populations due to reduced physical contact and digital accessibility issues, such as a lack of digital skills, limited resources, or unwillingness to engage with digital platforms. There was a lack of exploration of participant preferences for intervention format, which is crucial for tailoring interventions to individual needs and enhancing engagement and effectiveness. This highlights a need to consider digital inclusion in future digital health initiatives.

Methodologically, inconsistencies in the inclusion criteria for SMI and the measurement of mental health outcomes were noted (Table 2). This ambiguity hinders the ability to definitively assess the impact of digital interventions on mental health issues, pointing to the need for more standardized criteria in future research. In addition, the focus on diagnostic criteria may limit the understanding of underlying processes common across weight and mental disorders, suggesting a shift toward a more process-oriented approach in future studies [19].

Conclusions

This scoping review sheds light on the evolving role of digital technologies in health care communication for individuals living with obesity and mental health issues. It reveals the use of diverse modalities to bridge communication gaps; deliver education, motivation, and counseling; and enhance care accessibility and multidisciplinary support. However, it also highlights limitations related to inconsistent adherence, challenges in implementation, and the absence of technology acceptance theoretical models in current research. A key finding is the tangible improvement in physical health outcomes in obesity management through digital means. However, the effectiveness of these technologies in concurrently addressing mental health issues is less conclusive, marking a critical area for further investigation. The limited but insightful studies show the interdisciplinary and evolving nature of this field, accelerated by recent advancements in digital health care delivery.

The findings advocate for future research to be more extensive and inclusive, with an emphasis on the long-term impacts of digital interventions. This includes conducting follow-up assessments to understand the sustained effects and integrating innovative technologies such as artificial intelligence, VR, and machine learning to enrich mental health interventions [53,59]. The application of frameworks such as the Unified Theory of Acceptance and Use of Technology 2 model [56] could provide valuable insights into the adoption and use of digital technology by HPs in addressing complex health issues. While digital tools offer promise for enhancing health care communication in obesity management, their full potential in addressing concurrent mental health issues is yet to be realized. The review calls for custom digital solutions that address the biopsychosocial needs of those facing both obesity and mental health challenges, advocating for an interdisciplinary approach that encompasses various HP roles and comprehensive nutrition support [40]. This research is pivotal for advancing clinical practices and informing health policy, emphasizing the ongoing need for innovation and adaptation in digital health care communication strategies.

In conclusion, the journey toward using the full potential of digital technologies in health care communication is ongoing, with enhanced accessibility, reduced stigma, and personalized care as key benefits. Nevertheless, there remains a need to develop more comprehensive and integrated digital solutions to effectively tackle the complex interplay of obesity and mental health issues [36]. The vital steps for moving forward include addressing participant preferences; encouraging collaborative design [57]; and ensuring that interventions provide consistency, cohesion, and choice for participants [54].

Acknowledgments

This study is funded by the Department of the Economy (Northern Ireland) PhD Scholarship scheme.

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published paper (and its supplementary information files).

Authors' Contributions

AK led the systematic search and study selection, adhering to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, and was responsible for data extraction and result synthesis, formal analysis, and manuscript writing. AM assisted in designing the study and the methodology, guided the thematic analysis, oversaw the quality appraisal of the included studies, assessed the risk of bias, and contributed to manuscript review and editing. AM, MM, and RB supervised the research process. All authors engaged in discussions to synthesize the results and to draw meaningful conclusions. Each author reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 20 KB - jmir_v27i1e58434_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 23 KB - jmir_v27i1e58434_app2.docx]

Multimedia Appendix 3
Complete search strategy and database results.
[DOCX File , 22 KB - jmir_v27i1e58434_app3.docx]

Multimedia Appendix 4
Table of results mapping biopsychosocial constructs to the research themes.
[DOCX File , 18 KB - jmir_v27i1e58434_app4.docx]

References

1. Mental health. World Health Organization. 2022 Jun 17. URL: https://tinyurl.com/d2nmnbu7 [accessed 2023-04-12]
2. Casanova F, O'Loughlin J, Martin S, Beaumont RN, Wood AR, Watkins ER, et al. Higher adiposity and mental health: causal inference using Mendelian randomization. Hum Mol Genet 2021 Nov 30;30(24):2371-2382 [FREE Full text] [doi: 10.1093/hmg/ddab204] [Medline: 34270736]
3. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010 Mar;67(3):220-229. [doi: 10.1001/archgenpsychiatry.2010.2] [Medline: 20194822]
4. Rentrop V, Damerau M, Schweda A, Steinbach J, Schüren LC, Niedergethmann M, et al. Predicting acceptance of e-mental health interventions in patients with obesity by using an extended unified theory of acceptance model: cross-sectional study. JMIR Form Res 2022 Mar 17;6(3):e31229 [FREE Full text] [doi: 10.2196/31229] [Medline: 35297769]
5. Looijmans A, Jörg F, Bruggeman R, Schoevers RA, Corpeleijn E. Multimodal lifestyle intervention using a web-based tool to improve cardiometabolic health in patients with serious mental illness: results of a cluster randomized controlled trial (LION). BMC Psychiatry 2019 Nov 05;19(1):339 [FREE Full text] [doi: 10.1186/s12888-019-2310-5] [Medline: 31690281]
6. Mental disorders. World Health Organization. 2022 Jun 08. URL: https://tinyurl.com/52s37f4x [accessed 2025-01-29]
7. Milano W, Ambrosio P, Carizzone F, De Biasio V, Di Munzio W, Foia MG, et al. Depression and obesity: analysis of common biomarkers. Diseases 2020 Jun 14;8(2):23 [FREE Full text] [doi: 10.3390/diseases8020023] [Medline: 32545890]
8. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Best practices for designing chatbots in mental healthcare – a case study on iHelpr. In: Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI). 2018 Presented at: HCI '18; July 4-6, 2018; Belfast, UK. [doi: 10.14236/ewic/hci2018.129]
9. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
10. Blackburn M, Stathi A, Keogh E, Eccleston C. Raising the topic of weight in general practice: perspectives of GPs and primary care nurses. BMJ Open 2015 Aug 07;5(8):e008546 [FREE Full text] [doi: 10.1136/bmjopen-2015-008546] [Medline: 26254471]

XSL•FO
RenderX

11. Quigley F, Moorhead A, Bond R, McAloon T, Zheng H. Identifying nutritional myths when healthcare professionals communicate about weight and obesity in healthcare settings. Proc Nutr Soc 2020 Oct 19;79(OCE3):E775. [doi: 10.1017/S0029665120007612]
12. Marx W, Manger SH, Blencowe M, Murray G, Ho FY, Lawn S, et al. Clinical guidelines for the use of lifestyle-based mental health care in major depressive disorder: World Federation of Societies for Biological Psychiatry (WFSBP) and Australasian Society of Lifestyle Medicine (ASLM) taskforce. World J Biol Psychiatry 2023 Jun;24(5):333-386 [FREE Full text] [doi: 10.1080/15622975.2022.2112074] [Medline: 36202135]
13. Iyengar K, Upadhyaya GK, Vaishya R, Jain V. COVID-19 and applications of smartphone technology in the current pandemic. Diabetes Metab Syndr 2020;14(5):733-737 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.033] [Medline: 32497963]
14. Thomas RK. Health Communication. Heidelberg, Germany: Springer; 2006.
15. Drew RJ, Morgan PJ, Collins CE, Callister R, Kay-Lambkin F, Kelly BJ, et al. Behavioral and cognitive outcomes of an online weight loss program for men with low mood: a randomized controlled trial. Ann Behav Med 2022 Oct 03;56(10):1026-1041. [doi: 10.1093/abm/kaab109] [Medline: 34964449]
16. Looijmans A, Jörg F, Bruggeman R, Schoevers R, Corpeleijn E. Design of the Lifestyle Interventions for severe mentally ill Outpatients in the Netherlands (LION) trial; a cluster randomised controlled study of a multidimensional web tool intervention to improve cardiometabolic health in patients with severe mental illness. BMC Psychiatry 2017 Mar 21;17(1):107 [FREE Full text] [doi: 10.1186/s12888-017-1265-7] [Medline: 28327086]
17. Shannon HH, Joseph R, Puro N, Darrell E. Use of technology in the management of obesity: a literature review. Perspect Health Inf Manag 2019 Oct 01;16(Fall):1c [FREE Full text] [Medline: 31908626]
18. Patsalos O, Keeler J, Schmidt U, Penninx BW, Young AH, Himmerich H. Diet, obesity, and depression: a systematic review. J Pers Med 2021 Mar 03;11(3):176 [FREE Full text] [doi: 10.3390/jpm11030176] [Medline: 33802480]
19. Treasure J, Bektas S, Mutwalli H, Dhopatkar N, Himmerich H. Novel approaches to tackling emotional loss of control of eating across the weight spectrum. Proc Nutr Soc 2022 Sep;81(3):255-263. [doi: 10.1017/S0029665122000994] [Medline: 35260206]
20. Ahmadvand A, Gatchel R, Brownstein J, Nissen L. The biopsychosocial-digital approach to health and disease: call for a paradigm expansion. J Med Internet Res 2018 May 18;20(5):e189 [FREE Full text] [doi: 10.2196/jmir.9732] [Medline: 29776900]
21. Ma J, Rosas LG, Lv N, Xiao L, Snowden MB, Venditti EM, et al. Effect of integrated behavioral weight loss treatment and problem-solving therapy on body mass index and depressive symptoms among patients with obesity and depression: the RAINBOW randomized clinical trial. JAMA 2019 Mar 05;321(9):869-879 [FREE Full text] [doi: 10.1001/jama.2019.0557] [Medline: 30835308]
22. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
23. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
24. Pollock D, Davies EL, Peters MD, Tricco AC, Alexander L, McInerney P, et al. Undertaking a scoping review: a practical guide for nursing and midwifery students, clinicians, researchers, and academics. J Adv Nurs 2021 Apr 04;77(4):2102-2113 [FREE Full text] [doi: 10.1111/jan.14743] [Medline: 33543511]
25. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
26. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013 Mar 23;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
27. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Hoboken, NJ: John Wiley & Sons; 2019.
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
29. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
30. Google Translate. Google. URL: https://translate.google.com/?hl=mr&sl=auto&tl=mr&op=translate [accessed 2023-04-11]
31. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2025-01-29]
32. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 08;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]
33. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]

XSL•FO
RenderX

34. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
35. Lee C, Waite F, Piernas C, Aveyard P. Development and initial evaluation of a behavioural intervention to support weight management for people with serious mental illness: an uncontrolled feasibility and acceptability study. BMC Psychiatry 2023 Mar 01;23(1):130 [FREE Full text] [doi: 10.1186/s12888-023-04517-1] [Medline: 36859248]
36. Aschbrenner KA, Naslund JA, Gorin AA, Mueser KT, Browne J, Wolfe RS, et al. Group lifestyle intervention with mobile health for young adults with serious mental illness: a randomized controlled trial. Psychiatr Serv 2022 Feb 01;73(2):141-148. [doi: 10.1176/appi.ps.202100047] [Medline: 34189933]
37. Abedishargh N, Ramezani Farani A, Gharraee B, Farahani H. Effectiveness of internet-based cognitive behavioral therapy in weight loss, stress, anxiety, and depression via virtual group therapy. Iran J Psychiatry Behav Sci 2021 Sep 01;15(3):e113096. [doi: 10.5812/ijpbs.113096]
38. Bartels SJ, Aschbrenner KA, Pratt SI, Zubkoff L, Jue K, Williams G, et al. Virtual learning collaborative compared to technical assistance as a strategy for implementing health promotion in routine mental health settings: a hybrid type 3 cluster randomized trial. Adm Policy Ment Health 2022 Nov 20;49(6):1031-1046. [doi: 10.1007/s10488-022-01215-0] [Medline: 35987829]
39. Haddad R, Badke D'Andrea C, Ricchio A, Evanoff B, Morrato EH, Parks J, et al. Using innovation-corps (I-Corps™) methods to adapt a mobile health (mHealth) obesity treatment for community mental health settings. Front Digit Health 2022 May 27;4:835002 [FREE Full text] [doi: 10.3389/fdgth.2022.835002] [Medline: 35721796]
40. Kim M, Kim Y, Go Y, Lee S, Na M, Lee Y, et al. Multidimensional cognitive behavioral therapy for obesity applied by psychologists using a digital platform: open-label randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 30;8(4):e14817 [FREE Full text] [doi: 10.2196/14817] [Medline: 32352391]
41. Nicol G, Jansen M, Haddad R, Ricchio A, Yingling MD, Schweiger JA, et al. Use of an interactive obesity treatment approach in individuals with severe mental illness: feasibility, acceptability, and proposed engagement criteria. JMIR Form Res 2022 Dec 13;6(12):e38496 [FREE Full text] [doi: 10.2196/38496] [Medline: 36512399]
42. Yu Z, Roberts B, Snyder J, Stuart K, Wilburn J, Pudwill H, et al. A pilot study of a videoconferencing-based binge eating disorder program in overweight or obese females. Telemed J E Health 2021 Mar 01;27(3):330-340. [doi: 10.1089/tmj.2020.0070] [Medline: 32503392]
43. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
44. Hinchliffe N, Capehorn MS, Bewick M, Feenie J. The potential role of digital health in obesity care. Adv Ther 2022 Oct 04;39(10):4397-4412 [FREE Full text] [doi: 10.1007/s12325-022-02265-4] [Medline: 35925469]
45. Lam C, Milne-Ives M, Harrington R, Jani A, Helena van Velthoven M, Harding T, et al. Internet of things-enabled technologies as an intervention for childhood obesity: a systematic review. PLOS Digit Health 2022 Apr;1(4):e0000024 [FREE Full text] [doi: 10.1371/journal.pdig.0000024] [Medline: 36812526]
46. Nutley SK, Falise AM, Henderson R, Apostolou V, Mathews CA, Striley CW. Impact of the COVID-19 pandemic on disordered eating behavior: qualitative analysis of social media posts. JMIR Ment Health 2021 Jan 27;8(1):e26011 [FREE Full text] [doi: 10.2196/26011] [Medline: 33465035]
47. World mental health report: transforming mental health for all. World Health Organization. 2022 Jun 16. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2025-01-29]
48. FitzGerald C, Mumenthaler C, Berner D, Schindler M, Brosch T, Hurst S. How is physicians' implicit prejudice against the obese and mentally ill moderated by specialty and experience? BMC Med Ethics 2022 Aug 24;23(1):86 [FREE Full text] [doi: 10.1186/s12910-022-00815-7] [Medline: 36002822]
49. Ali S, Kleib M, Paul P, Petrovskaya O, Kennedy M. Compassionate nursing care and the use of digital health technologies: a scoping review. Int J Nurs Stud 2022 Mar;127:104161. [doi: 10.1016/j.ijnurstu.2021.104161] [Medline: 35032743]
50. Aschbrenner KA, Naslund JA, Reed JD, Fetter JC. Renewed call for lifestyle interventions to address obesity among individuals with serious mental illness in the COVID-19 era and beyond. Transl Behav Med 2021 Jul 29;11(7):1359-1364. [doi: 10.1093/tbm/ibab076] [Medline: 34160055]
51. Brown A, Flint SW, Batterham RL. Pervasiveness, impact and implications of weight stigma. EClinicalMedicine 2022 May;47:101408 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101408] [Medline: 35497065]
52. Fruh SM, Graves RJ, Hauff C, Williams SG, Hall HR. Weight bias and stigma: impact on health. Nurs Clin North Am 2021 Dec;56(4):479-493 [FREE Full text] [doi: 10.1016/j.cnur.2021.07.001] [Medline: 34749889]
53. Preparing the healthcare workforce to deliver the digital future. National Health Service. 2019. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2024-12-23]
54. Mental health strategy 2021-2031. Department of Health, Northern Ireland. 2021 Jun 29. URL: https://www.health-ni.gov.uk/publications/mental-health-strategy-2021-2031 [accessed 2024-01-17]
55. Mental health digital playbook. National Health Service England. URL: https://tinyurl.com/mr49ws8r [accessed 2024-01-17]

XSL•FO
RenderX

56. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
57. Bond RR, Mulvenna MD, Potts C, O'Neill S, Ennis E, Torous J. Digital transformation of mental health services. Npj Ment Health Res 2023 Aug 22;2(1):13 [FREE Full text] [doi: 10.1038/s44184-023-00033-y] [Medline: 38609479]
58. Gudzune KA, Jerome GJ, Dalcin AT, Gao R, Mace E, Fink T, et al. Scaling a behavioral weight-loss intervention for individuals with serious mental illness using the enhanced replicating effective programs framework: a preconditions phase proof-of-concept study. Obes Sci Pract 2023 Dec;9(6):618-630 [FREE Full text] [doi: 10.1002/osp4.690] [Medline: 38090682]
59. Brown T, Nauman Vogel E, Adler S, Bohon C, Bullock K, Nameth K, et al. Bringing virtual reality from clinical trials to clinical practice for the treatment of eating disorders: an example using virtual reality cue exposure therapy. J Med Internet Res 2020 Apr 23;22(4):e16386 [FREE Full text] [doi: 10.2196/16386] [Medline: 32324145]
60. Maximising the potential of digital in mental health. NHS Confederation. URL: https://tinyurl.com/4b4nt68m [accessed 2024-01-17]
61. Plackett B. The vicious cycle of depression and obesity. Nature 2022 Aug 24;608(7924):S42-S43. [doi: 10.1038/d41586-022-02207-8] [Medline: 36002500]

Abbreviations

CBT: cognitive behavioral therapy
HP: health professional
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
ROB: risk of bias
SMI: serious mental illness
VLC: virtual learning collaborative
VR: virtual reality
WM: weight management

Edited by A Mavragani; submitted 15.03.24; peer-reviewed by A Jain, J Mistry; comments to author 02.04.24; revised version received 11.04.24; accepted 28.10.24; published 10.02.25.

Please cite as:
Kearns A, Moorhead A, Mulvenna M, Bond R
Assessing the Uses, Benefits, and Limitations of Digital Technologies Used by Health Professionals in Supporting Obesity and Mental Health Communication: Scoping Review
J Med Internet Res 2025;27:e58434
URL: https://www.jmir.org/2025/1/e58434
doi:10.2196/58434
PMID:

©Amanda Kearns, Anne Moorhead, Maurice Mulvenna, Raymond Bond. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Machine Learning in the Management of Patients Undergoing Catheter Ablation for Atrial Fibrillation: Scoping Review

Aijing Luo[1,2,3,4], MD; Wei Chen[1,2,3,4], BEng; Hongtao Zhu[1,3,4,5], ME; Wenzhao Xie[2,3,4,6], MD; Xi Chen[1,4,5], ME; Zhenjiang Liu[1,4,7], MD; Zirui Xin[1,3,4,5], MD

[1]The Second Xiangya Hospital, Central South University, Changsha, China

[2]School of Life Sciences, Central South University, Changsha, China

[3]Key Laboratory of Medical Information Research (Central South University), College of Hunan Province, Changsha, China

[4]Clinical Research Center For Cardiovascular Intelligent Healthcare In Hunan Province, Changsha, China

[5]Information and Network Center, The Second Xiangya Hospital, Central South University, Changsha, China

[6]The Third Xiangya Hospital, Central South University, Changsha, China

[7]Department of Cardiology, The Second Xiangya Hospital, Central South University, Changsha, China

Corresponding Author:
Zirui Xin, MD
The Second Xiangya Hospital
Central South University
No. 139 Renmin Middle Road
Changsha, 410011
China
Phone: 86 15211017166
Email: xinzirui@csu.edu.cn

Abstract

Background: Although catheter ablation (CA) is currently the most effective clinical treatment for atrial fibrillation, its variable therapeutic effects among different patients present numerous problems. Machine learning (ML) shows promising potential in optimizing the management and clinical outcomes of patients undergoing atrial fibrillation CA (AFCA).

Objective: This scoping review aimed to evaluate the current scientific evidence on the application of ML for managing patients undergoing AFCA, compare the performance of various models across specific clinical tasks within AFCA, and summarize the strengths and limitations of ML in this field.

Methods: Adhering to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, relevant studies published up to October 7, 2023, were searched from PubMed, Web of Science, Embase, the Cochrane Library, and ScienceDirect. The final included studies were confirmed based on inclusion and exclusion criteria and manual review. The PROBAST (Prediction model Risk Of Bias Assessment Tool) and QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) methodological quality assessment tools were used to review the included studies, and narrative data synthesis was performed on the modeled results provided by these studies.

Results: The analysis of 23 included studies showcased the contributions of ML in identifying potential ablation targets, improving ablation strategies, and predicting patient prognosis. The patient data used in these studies comprised demographics, clinical characteristics, various types of imaging (9/23, 39%), and electrophysiological signals (7/23, 30%). In terms of model type, deep learning, represented by convolutional neural networks, was most frequently applied (14/23, 61%). Compared with traditional clinical scoring models or human clinicians, the model performance reported in the included studies was generally satisfactory, but most models (14/23, 61%) showed a high risk of bias due to lack of external validation.

Conclusions: Our evidence-based findings suggest that ML is a promising tool for improving the effectiveness and efficiency of managing patients undergoing AFCA. While guiding data preparation and model selection for future studies, this review highlights the need to address prevalent limitations, including lack of external validation, and to further explore model generalization and interpretability.

(J Med Internet Res 2025;27:e60888) doi:10.2196/60888

XSL•FO
RenderX

KEYWORDS

atrial fibrillation; catheter ablation; deep learning; patient management; prognosis; quality assessment tools; cardiac arrhythmia; public health; quality of life; severe medical condition; electrocardiogram; electronic health record; morbidity; mortality; thromboembolism; clinical intervention

Introduction

Background

Atrial fibrillation (AF) is the most common type of cardiac arrhythmia [1], and its incidence has risen steadily over the past 3 decades, making it an urgent public health problem worldwide [2,3]. While AF substantially impacts patients' quality of life, it also increases the risk of serious medical conditions [4-7], such as stroke and thromboembolism. This contributes to long-term rises in both morbidity and mortality rates [8] as well as a growing strain on health care costs [9].

Catheter ablation (CA) represents the most effective clinical intervention currently available for AF [10]. When patients receive their first CA treatment, the effectiveness rate can reach 60%-80% within 1 year [11,12]. However, the long-term efficacy is not optimal , and the success rate decreases to 50%-60% at 3 to 5 years [13-15]. The type and degree of AF, cardiac stroma, obesity, alcohol consumption, and obstructive sleep apnea syndrome are all known to be potential contributors to high rates of recurrence in the long term [16,17]. In addition, the surgical modality and ablation strategy used in treatment have an impact, particularly for those patients who require additional non–pulmonary vein (PV) ablations [18]. Consequently, developing tailored management strategies that consider different patient conditions, optimizing the surgical approach and enhancing its long-term efficacy, is a challenge that needs to be resolved in this field.

In clinical practice, to better manage patients, clinical risk scores (such as APPLE [19]) specifically for predicting postoperative recurrence of AF CA (AFCA) have emerged in the past decade, or attempts have been made to use other scores (such as $CHADS_2$ [20] and CHA_2DS_2-VASc [21]) to evaluate patients' postoperative rhythm outcomes after AFCA [22]. These risk scores typically use a limited number of risk factors as predictors. Although this simplicity facilitates their adoption and use in clinical settings, it concurrently caps their predictive effectiveness. In addition, these scoring models predominantly rely on the evaluation of specific biomarkers in previous literature when selecting predictors. This approach may easily lead to the neglect of some potential predictors [23], thus failing to truly capture the complexity of the AFCA recurrence mechanism and making it difficult to truly meet the pursuit of efficient and personalized patient management.

Rapidly developing machine learning (ML) techniques hold significant promise as a formidable support in grappling with the complexities of this challenge. In recent years, the ability of ML to analyze large biomedical datasets, such as electronic health records (EHRs), medical imaging, and multiomics data, has made it increasingly popular in medical research [24,25]. ML models have become promising tools for enhancing the precision and efficiency of AF diagnosis and treatment. By analyzing complex, multidimensional datasets, they help identify individuals at high risk for AF, uncover risk factors and biomarkers associated with the disease's progression, and guide personalized interventions for prognosis. One branch of ML, deep learning (DL), stands out due to its powerful capabilities in automatic feature recognition and prediction. In processing complex data types such as images and electrophysiological signals, DL has demonstrated performance comparable to that of human experts [26,27]. Several studies have attempted to apply ML technology to the management of patients undergoing AFCA. As the application prospects of ML technology in this field become increasingly clear, a comprehensive overview and evaluation of the existing relevant literature is needed.

Objectives

The purpose of this scoping review is to systematically review the current progress in using ML technology in the management of patients undergoing AFCA. Specifically, the studies discussed in this review explore a wide range of ML models and multiple data modalities, including EHR, imaging omics, and cardiac electrophysiological signals. We will summarize how these studies prepare data, select models, and evaluate results across different clinical tasks. In addition, we will emphasize the advantages and limitations of ML technology in this field based on the integration and comparison of these study findings.

Methods

This scoping review was conducted following the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines [28].

Information Sources and Search Strategies

A comprehensive search was conducted to identify relevant studies across 5 databases as of October 7, 2023: PubMed, Web of Science, Embase, the Cochrane Library, and ScienceDirect. The specific search strategies used for each database are detailed in Table S1 in Multimedia Appendix 1. In addition to obtaining literature through search formula, additional relevant articles (3/246, 1.2%) were manually identified by using a forward-backward snowballing approach.

Study Inclusion and Exclusion Criteria

In line with the thematic focus and objectives of our review, the studies included in this scoping review were required to meet the eligibility criteria outlined in Textbox 1. Studies were excluded from this scoping review if they met any of the exclusion criteria.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria for published studies applying machine learning (ML) technology for the management of patients undergoing catheter ablation for atrial fibrillation.

Inclusion criteria

- Article type: articles (peer reviewed and formally published)
- Study topic: specializes in the management of patients undergoing catheter ablation for atrial fibrillation
- Study participants: real patients with atrial fibrillation
- Study design: ML techniques were clearly applied and described; results of the study and the ML model performance outcome metrics were explicitly presented
- Language of publication: English

Exclusion criteria

- Article type: reviews; meta-analyses; editorials; case reports; editorial materials; other types of publications
- Study topic: atrial fibrillation disease detection; risk prediction of new onset of atrial fibrillation
- Study participants: only used fictional patient data
- Study design: did not apply ML techniques; did not explicitly report on the model outcome metrics
- Language of publication: non-English

Study Selection

After removing duplicates, all retrieved studies were manually screened by title and abstract to exclude those clearly not aligned with the objectives of this review. Studies that passed this initial screening were then downloaded in full and assessed based on the inclusion and exclusion criteria during a full-text review. This process was independently conducted by 2 reviewers, with any disagreements resolved through consultation with a third reviewer to reach a consensus.

Data Extraction

Two reviewers independently extracted data from the selected studies, and a third reviewer was responsible for assessing and addressing any discrepancies and biases in data extraction. The following data items were extracted: author names and publication years, data sources, dataset sizes, ML models, model variables, model tasks (such as prediction or detection), and model results. If multiple ML models were used in a study, we will report only the one with the best performance.

Quality Assessment

We selected the widely recognized Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [29] and Prediction model Risk Of Bias Assessment Tool (PROBAST) [30] as methodological quality assessment tools. We divided the included studies into 2 subgroups for independent evaluation based on different research objectives, namely, subgroup A: studies that aim at image classification or segmentation, assessed using an adapted version of QUADAS-2; and subgroup B: with the goal of predicting patient prognosis, assessed using PROBAST. This assessment was independently conducted by 2 reviewers, and any disagreements or discrepancies that arise during the assessment process were resolved with the involvement of a third reviewer.

During our review, we adjusted some signal questions in QUADAS-2 as needed. Tables S2 and S3 in Multimedia Appendix 1 report the specific signal questions and assessment instructions for the tools used in the quality assessment process for this review.

Results

Study Selection and Study Characterization

A total of 246 studies were identified from different databases using the method described in the Information Sources and Search Strategies section (Figure 1). After excluding duplicate publications (65/246, 26.4%), nonarticle publications (32/246, 13%), and studies that did not match the research objectives of this scoping review (115/246, 46.7%), 34 (13.8%) studies remained for full-text assessment. Ultimately, 23 studies were deemed eligible and were included for data extraction and synthesis. Table S4 in Multimedia Appendix 1 provides a detailed list of the studies excluded after full-text assessment, along with the reasons for their exclusion.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of study selection.

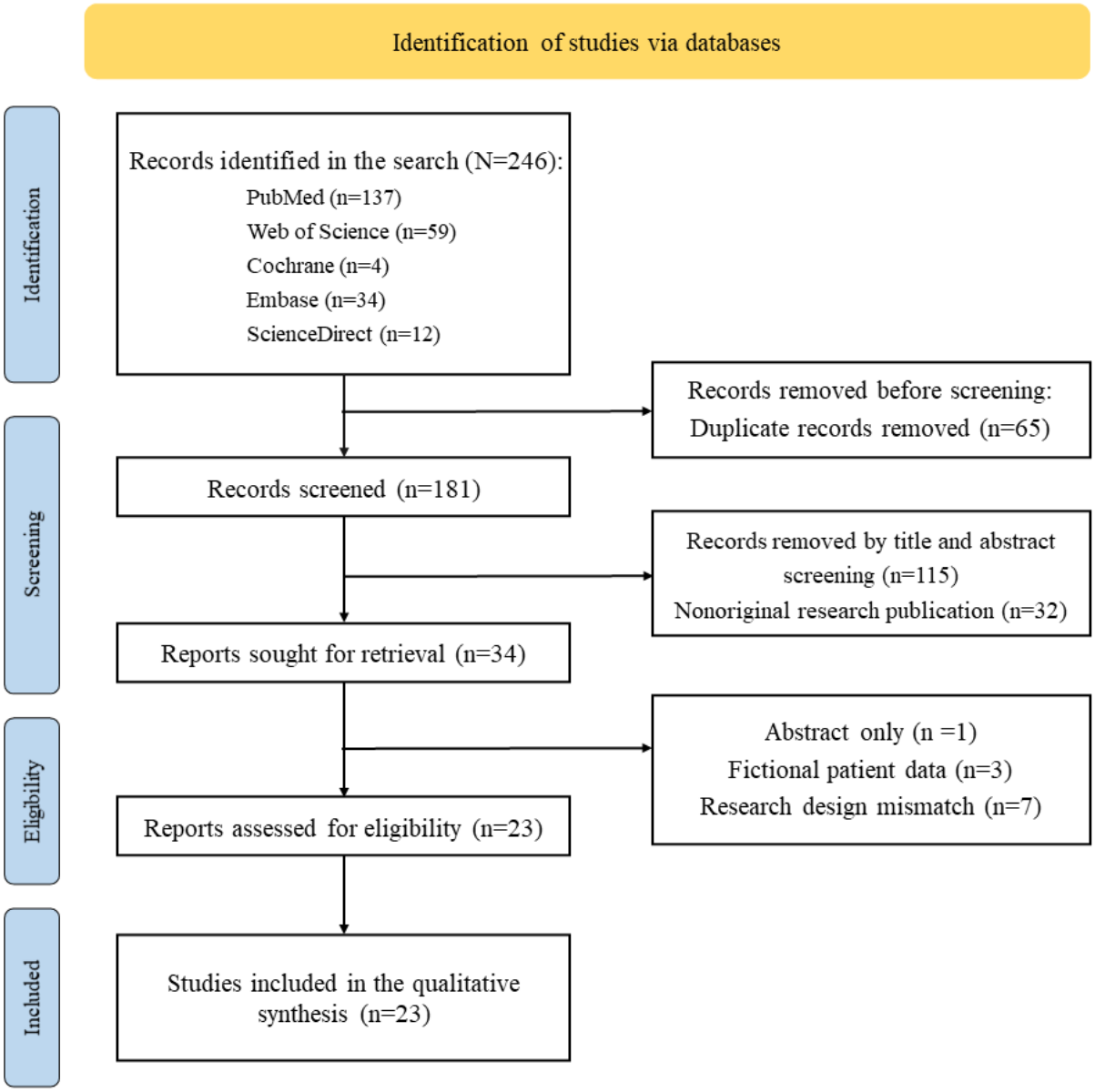

Table 1 provides the basic characteristics of the studies included in this review. The publication of these studies was concentrated within the past 5 years, which can be attributed to the maturation of ML technologies and the growing interest in precision medicine among researchers in the field of AF. The included studies were conducted in 9 different countries, with the United States (7/23, 30%) having the largest number of studies, followed by China (5/23, 22%) and the United Kingdom (4/23, 17%). This diversity in publication countries reflects the global interest in using ML models for the management of patients undergoing AFCA. Regarding the types of models used, most of the included studies (14/23, 61%) used DL techniques.

XSL•FO
RenderX

Table 1. Basic characteristics of the included studies (N=23).

Characteristic	Studies, n (%)
Country of study	
United States	7 (30)
China	5 (22)
United Kingdom	4 (17)
Korea	2 (9)
Spain	1 (4)
The Netherlands	1 (4)
Canada	1 (4)
Italy	1 (4)
Japan	1 (4)
Year of publication	
2023	3 (13)
2022	7 (30)
2021	6 (26)
2020	6 (26)
2019	1 (4)
Type of models	
Machine learning	23 (100)
Deep learning	14 (61)
Key medical tasks	
Patient prognosis prediction	16 (70)
Ablation targets identification	3 (13)
Ablation strategy improvement	4 (17)

Regarding the application direction of ML in AFCA patient management, this review divides the included studies into 3 subgroups corresponding to 3 key medical tasks based on the purpose of the model. The first and most common focus is the development of prognostic models for patients undergoing AFCA (16/23, 70%). These studies aim to provide accurate, timely, and personalized postoperative risk predictions for patients undergoing AFCA [31-46]. Although these studies share similar objectives, they use diverse types of data, with EHRs being the primary data source. The second key task focuses on the identification of ablation targets (3/23, 13%), aiming to use the feature extraction capabilities of DL models to identify potential ablation targets, such as rotors [47,48] and focal sources [49], from electrophysiological and medical imaging data. Third, some studies focus on using preoperative imaging data of patients to improve ablation strategies (4/23,17%), assisting physicians in recognizing interpatient differences to guide ablation treatments [50-53].

Characteristics of the Data in the Included Studies

When developing ML models for patient management, key factors to consider include dataset size, data modality, and resource availability. Table 2 outlines the wide range of data modalities covered in this review, and we note that there is a wide variation in the choice of variables and the number of features across studies, which highlights the diversity and complexity of factors associated with influencing patient management in AFCA.

XSL•FO
RenderX

Table 2. Summary of data use of the included studies.

Study	Patient cohort		Input data for the model	Data sources
	Patients, n	Time frame		
Liu et al [50], 2020	358	October 2004 to December 2017	PVCT[a]	Taipei Veterans General Hospital
Firouznia et al [31], 2021	203	2013 to 2016	CT[b], age, LAV[c], AF[d] type, and catheter ablation technique	Cleveland Clinic
Atta-Fosu et al [32], 2021	68	July 2015 to November 2016	CT, age, sex, LAV, LVEF[e], BMI, sinus rhythm at the time of ablation, AF type, and catheter ablation technique	Cleveland Clinic
Muizniece et al [51], 2021	122	2011 to 2016	LGE-MRI[f]	2018 STACOM Split Challenge and St. Thomas Hospital
Muffoletto et al [52], 2021	122	2011 to 2016	LGE-MRI	2018 STACOM Split Challenge and St. Thomas Hospital
Shade et al [33], 2020	32	December 2011 to December 2015	Personalized atrial computational modeling and LGE-MRI	Hopkins University Hospital
Roney et al [34], 2022	100	NR[g]	Different types of fibrosis in the atrial matrix, fibrograms, AF induction protocols, ERP[h] values, PVI[i] size, visual fibrosis score, total surface area of left atrium, total surface area of the PVs, fibrotic area of PVs, LVEF, BMI, age, female, congestive HF[j], hypertension, diabetes mellitus, stroke, coronary artery disease, and type of AF	St. Thomas Hospital
Yang et al [53], 2020	190	2011 to 2018	LGE-MRI	Local regional ethics committee
Hwang et al [35], 2020	606	July 2008 to July 2019	Transthoracic echocardiography	Chang Gung Memorial Hospital
Liao et al [49], 2021	78	NR	uEGM[k]	Focal Source and Trigger randomized controlled trial (Chauhan et al [54])
Ríos-Muñoz et al [47], 2022	48	NR	uEGM, bEGM[l], and uLAT[m]	Gregorio Marañón University General Hospital
Alhusseini et al [48], 2020	35	Up to December 2017	EGM[n]	Comparison of Algorithms for Rotational Evaluation in Atrial Fibrillation Registry
An et al [36], 2022	12	January 2015 to January 2019	3D atrial endocardial unipolar and bipolar voltage map	the Dallas VA Medical Center
Tang et al [37], 2022	156	2015 to 2017	EGM, ECG[o], age, gender, height, weight, BMI, ethnicity and race, hypertension, hyperlipidemia, TIA[p], stroke, CAD[q], diabetes mellitus, chronic kidney disease, congestive HF, OSA[r], type of AF, previous history of AF ablation, LVEF, LAD[s], LAV, LASA[t], and LASI[u]	Stanford University Tertiary Referral Center
Li et al [38], 2019	14	NR	BSPM[v]	West China Hospital of Sichuan University
Saiz-Vivo et al [39], 2021	74	October 2005 to July 2014	Clinical data and heart rate variability characteristics extracted from ICM[w]	Reveal LINQ Usability Study and clinical data from a single center

XSL•FO
RenderX

Study	Patient cohort		Input data for the model	Data sources
	Patients, n	Time frame		
Jiang et al [40], 2023	1618	January 2012 to May 2019	12-lead ECG	Guangdong Provincial People's Hospital
Lee et al [41], 2022	177	February 2017 to October 2020	Age, sex, height, weight, hypertension, AF type, AF duration, LAD, LAMI[x], LVEF, and eGFR[y]	Chungbuk National University Hospital
Park et al [42], 2022	1872	2009 to 2018	Age, female, AF type, BMI, heart failure, hypertension, diabetes mellitus, stroke or TIA, vascular disease, left atrium dimension, LVEF, EEM[z], creatinine, hemoglobin, and preablation PR interval	Yonsei AF Ablation Cohort Database
Hung et al [43], 2020	11,334	January 2013 to December 2013	Age, total hospital discharges, number of diagnoses, number of chronic diseases, length of hospitalization, number of surgeries, gender, diabetes, hypertension, hypothyroidism, COPD[aa], renal failure, depression, peripheral vascular disease, obesity, size of hospital beds, type of hospital, and discharge status	National Readmission Database 2013
Saglietto et al [44], 2023	3128	April 2012 to April 2015	LVEDV[ab], estimated glomerular filtration rate, BMI, age, pre- and post–left atrium diameters, LVEF, CHAD2 DS2-VASc score, dyslipidemia, AFL[ac], type of ablation procedure, type of AF, structural heart disease, hypertension, baseline sinus rhythm, gender, abnormal ECG, heart failure, CAD, and smoking history	European Society of Cardiology-European Heart Rhythm Association Long-Term Registry for Atrial Fibrillation Ablation
Ma et al [45], 2023	471	January 2018 to December 2020	Age, male, BMI, diastolic blood pressure, systolic blood pressure, smoking history, CHA2DS2-VASc, HABLED score, AF duration, hypertension, CAD, type 2 diabetes mellitus, chronic HF, atrial septal defect, LAD, LVEF, and white blood cell counts	The First Affiliated Hospital of Air Force Medical University

XSL•FO
RenderX

Study	Patient cohort		Input data for the model	Data sources
	Patients, n	Time frame		
Zhou et al [46], 2022	310	Jun 2016 - Oct 2019	NT-proBNP[ad], paroxysmal AF, LAAV[ae], and LAV	Toyo University Ohashi Medical Center

[a]PVCT: pulmonary vein computed tomography.
[b]CT: computed tomography.
[c]LAV: left atrial volume.
[d]AF: atrial fibrillation.
[e]LVEF: left ventricular ejection fraction.
[f]LGE-MRI: late gadolinium-enhanced cardiac magnetic resonance.
[g]NR: not reported.
[h]ERP: effective refractory period.
[i]PVI: pulmonary vein isolation.
[j]HF: heart failure.
[k]uEGM: unipolar endocardial electrogram.
[l]bEGM: bipolar endocardial electrogram.
[m]uLAT: unipolar localized activation time.
[n]EGM: electrogram.
[o]ECG: electrocardiogram.
[p]TIA: transient ischemic attack.
[q]CAD: coronary artery disease.
[r]OSA: obstructive sleep apnea.
[s]LAD: left atrial diameter.
[t]LASA: left atrial surface area.
[u]LASI: left atrial spherical index.
[v]BSPM: body surface potential map.
[w]ICM: implantable cardiac monitor.
[x]LAMI: left atrial mass index.
[y]eGFR. estimated glomerular filtration rate.
[z]EEM: peak transmitral flow velocity (E), and tissue Doppler echocardiography of the peak septal mitral annular velocity (em).
[aa]COPD: chronic obstructive pulmonary disease.
[ab]LVEDV: left ventricular end-diastolic volume.
[ac]AFL: atrial flutter.
[ad]NT-proBNP: N-terminal pro-brain natriuretic peptide.
[ae]LAAV: left atrial appendage volume.

The characteristics of the data used in the included studies are provided in Table 3, which summarizes dataset size, data modality, data source, and whether the data were multicenter. There was substantial variation among the included studies in the size of the dataset used to develop the ML model, with patient numbers ranging from 8 to 5872, but few studies adequately described their methodology or reasons for determining the sample size, in contrast to studies of randomized controlled trials based on statistical methods. In terms of data resource availability, only 17% (4/23) of the studies used publicly available datasets of patients with AF (2013 Nationwide Readmissions Database [43], 2018 Atria Segmentation Data [51,52], and European Society of Cardiology-European Heart Rhythm Association Long-Term Registry for AF Ablation [44]) for model development or validation, while most of the remaining studies (19/23, 83%) used private single-center retrospective datasets. In addition, 17% (4/23) of the studies used multicenter datasets in model development or external datasets for validation after model development [39,44,51,52].

XSL•FO
RenderX

Table 3. Characteristics of the data used in the included studies (N=23).

Characteristics of used data	Studies, n (%)	References
Size of dataset		
<50	5 (22)	[33,36,38,47,48]
50-100	4 (17)	[32,34,39,49]
101-200	5 (22)	[37,41,51-53]
201-500	4 (17)	[31,45,46,50]
501-1000	1 (4)	[35]
>1000	4 (17)	[40,42-44]
Data modality		
Electronic health records	10 (43)	[31,32,34,37,41-46]
Medical imaging	9 (39)	[31-35,50-53]
Electrophysiological signal	7 (30)	[36-38,40,47-49]
Biomedical simulation	2 (9)	[33,34]
Other	1 (4)	[39]
Multimodal input used	4 (17)	[31,32,34,37]
Availability of data source		
Private	19 (83)	[31-42,45-50,53]
Public	4 (17)	[43,44,51,52]
Multicenter data used		
Yes	4 (17)	[39,44,51,52]
No	19 (83)	[31-38,40-43,45-50,53]

To provide a more intuitive integration and comparison of the data modalities of the included studies, an evidence heat map was used to visualize the results (Figure 2), which highlights and reveals the medical imaging data and electrophysiological signal data involved in the included studies. In studies aimed at predicting patient prognosis, the most commonly used data type was the EHR (10/16, 63%), and it is worth noting that a small number of studies (4/16, 25%) in this segment also used multimodal data, combining either imaging data or electrocardiographic signals (or both) with clinical features to build model, and reported superior model performance relative to their constructed unimodal models. Intracardiac electrograms (EGMs) are a key data modality in studies focused on identifying ablation targets, whereas imaging data are commonly used in studies aimed at improving ablation strategies.

XSL•FO
RenderX

Figure 2. Evidence heat map of data modality by machine learning model (grouped by medical task).

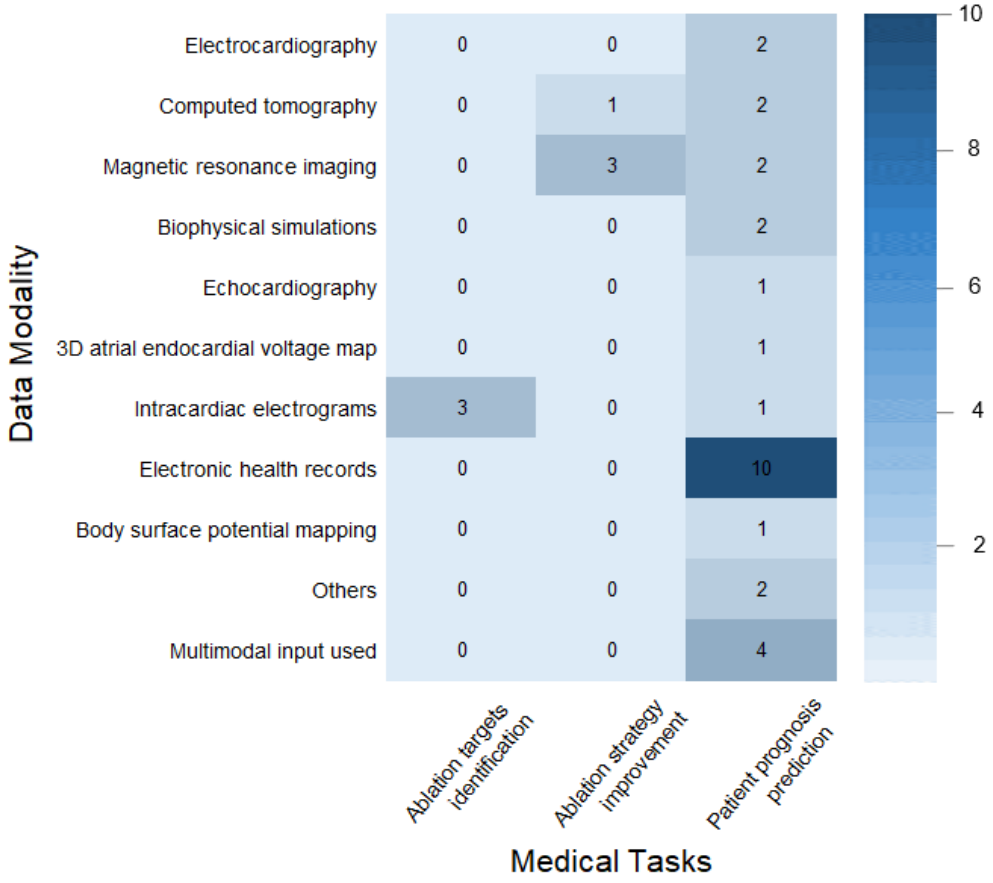

Modeling Approaches and Performance

Distinct medical tasks and research objectives necessitate the application of various ML techniques. Table 4 provides the ML models constructed in each study and the specific values of their performance indicators. It should be noted that if >1 ML model was constructed in 1 study at the same time, only the optimal model performance among them will be reported in this review, and the complete information of the ML models is provided in Table S2 in Multimedia Appendix 1. In this review, if a study reports multiple ML models simultaneously, we will report only the optimal model performance among them. Table 4 provides details on the used ML techniques, medical tasks, and model performances in the included studies.

The classical ML models covered in this review are diverse, with random forest (6/23, 26%) and support vector machines (6/23, 26%) appearing most frequently, and in some studies, although they are not the best-performing models, they have been applied for feature selection or as comparative models. Of the 23 studies, 5 (22%) used the k-nearest neighbors algorithm, 3 (13%) used logistic regression, and 2 (9%) used the decision tree. In addition, 17% (4/23) of the studies used integrated learning models such as XGBoost and AdaBoost. It is important to note that DL models represented by convolutional neural network (CNN) and RestNet were the most frequently used as performance-optimized models in all the included studies, particularly in those studies that used image data and electrical signal data as model inputs, and 61% (14/23) of the studies reported performance-optimized models belonging to the DL type. Specifically, CNN and its variants were applied in 30% (7/23) of the studies, and ResNet was applied in 9% (2/23) of the studies.

In terms of model performance evaluation indicators, the main ones used were area under the receiver operating characteristic curve (AUC; 17/23, 74%), sensitivity and specificity (11/23, 48%), and accuracy (11/23, 48%), while some studies chose to assess the model performance with metrics such as F_1-score. From the information in Tables 4 and 5, it can be seen that the performance of ML models in the included studies was good overall. Among them, 74% (17/23) of the studies reported the AUC values of the ML models, and the values of this performance indicator ranged from 0.5 to 0.98. The average AUC for ML models predicting patient prognosis was 0.8 (SD=0.11), with a median of 0.83 (IQR 0.74-0.86), and 13% (3/23) of the studies had a model AUC >0.90. Furthermore, 48% (11/23) of the studies reported model sensitivity and specificity, with sensitivity ranging between 0.71 and 0.97 and specificity ranging from 0.79 to 0.99. In terms of accuracy, 48% (11/23) of the studies reported this statistic, with values ranging from 0.72 to 0.99. All models aimed at improving ablation strategies reported accuracy, with an average of 0.87.

XSL•FO
RenderX

Table 4. Performance for each machine learning model according to the medical tasks (N=23).

Medical tasks and proposed model	AUC[a]	Sensitivity	Specificity	Accuracy	Other
Patient prognosis prediction					
RF[b] [31]	0.87	NR[c]	NR	NR	NR
RF [42]	0.965	0.96	0.89	NR	NR
RF [44]	0.721	NR	NR	NR	NR
RF [45]	0.667	NR	NR	NR	NR
XGboost [32]	0.78	NR	NR	NR	NR
QDA[d] [33]	0.82	0.82	0.89	NR	NR
SVM[e] [34]	0.85	NR	NR	NR	F_1-score=0.8
CNN[f] [35]	0.861	0.803	0.789	0.796	NR
CNN [37]	0.859	NR	NR	NR	NR
CNN-SVM [38]	NR	0.88	0.96	0.96	NR
CNN [40]	0.84	0.723	0.95	0.92	F_1-score=0.71
CNN [46]	0.76	NR	NR	NR	CI[g]=0.76
KNN[h] [43]	0.91	0.713	0.991	0.854	PR=0.886
OWV[i] [39]	NR	0.76	0.87	0.82	NR
MLP[j] [36]	0.5	NR	NR	0.875	F_1-score=0.933
MLP [41]	0.766	NR	NR	NR	NR
Ablation targets identification					
ResNet [49]	0.980	0.781	0.822	NR	NR
CRNN[k] [47]	0.81	NR	NR	NR	MCC[l]=0.680
CNN [48]	NR	0.97	0.93	0.95	NR
Ablation strategy improvement					
ResNet [50]	0.88	0.75	0.957	0.886	NR
Reinforcement Q-learning [51]	NR	NR	NR	0.72	NR
MVVT[m] [53]	NR	0.92	0.99	0.99	DI[n]=93.11
CNN [52]	NR	NR	NR	0.865	NR

[a]AUC: area under the curve.
[b]RF: random forest.
[c]NR: not reported.
[d]QDA: quadratic discriminant analysis.
[e]SVM: support vector machine.
[f]CNN: convolutional neural network.
[g]CI: C-index.
[h]KNN: K-nearest neighbor.
[i]OWV: optimal weighted voting.
[j]MLP: multilayer perceptron.
[k]CRNN: convolutional recurrent neural network.
[l]MCC: Matthews correlation coefficient.
[m]MVTT: fully automatic multiview dual-task recursive attention model.
[n]DI: Dice score.

XSL•FO
RenderX

Table 5. Statistical validation of the machine learning models.

Statistics	Studies, n (%)	Medical tasks involved (fraction) and corresponding study reference
AUC (%[a])		
<70	2 (9)	• Patient prognosis prediction (2/2): [36,45]
70-79	4 (17)	• Patient prognosis prediction (4/4): [32,41,44,46]
80-85	4 (17)	• Patient prognosis prediction (3/4): [33,34,40] • Ablation targets identification (1/4): [47]
86-90	4 (17)	• Patient prognosis prediction (3/4): [32,35,37] • Ablation strategy improvement (1/4): [50]
>90	3 (13)	• Patient prognosis prediction (2/3): [42,43] • Ablation targets identification (1/3): [49]
Sensitivity (%[b])		
<85	7 (30)	• Patient prognosis prediction (5/7): [33,35,39,40,43] • Ablation targets identification (1/7): [49] • Ablation strategy improvement (1/7): [50]
85-89	1 (4)	• Patient prognosis prediction (1/1): [38]
90-95	1 (4)	• Ablation strategy improvement(1/1): [53]
>95	2 (9)	• Patient prognosis prediction (1/2): [42] • Ablation targets identification (1/2): [48]
Specificity (%[c])		
<85	2 (9)	• Patient prognosis prediction (1/2): [35] • Ablation targets identification (1/2): [49]
85-89	3 (13)	• Patient prognosis prediction (3/3): [33,39,42]
90-95	2 (9)	• Patient prognosis prediction (1/2): [40] • Ablation targets identification (1/2): [48]
>95	4 (17)	• Patient prognosis prediction (2/4): [38,43] • Ablation strategy improvement (2/4): [50,53]
Accuracy (%[d])		
<80	2 (9)	• Patient prognosis prediction (1/2): [35] • Ablation strategy improvement (1/2): [51]
80-86	2 (9)	• Patient prognosis prediction (2/2): [39,43]
87-92	4 (17)	• Patient prognosis prediction (2/4): [36,40] • Ablation strategy improvement (2/4): [50,52]
>92	3 (13)	• Patient prognosis prediction (1/3): [38] • Ablation targets identification (1/3): [48] • Ablation strategy improvement (1/3): [53]

[a]AUC: area under the curve; AUC values were mentioned in 17 included studies.

[b]Sensitivity was mentioned in 11 included studies.

[c]Specificity was mentioned in 11 included studies.

[d]Accuracy was mentioned in 11 included studies.

XSL•FO
RenderX

Findings of the Included Studies

Identifying AFCA Targets

Three studies discussed the application of ML techniques in identifying potential targets for CA in AF [47-49]. Clinically, electrophysiologists must resort to time-consuming and complex invasive cardiac electrophysiological examinations to confirm ablation targets; deep neural networks can help improve this process. Liao et al [49] applied ResNet to classify raw intracardiac EGMs during AF, accurately identifying patients' putative focal source target site with performance approximating that of human experts, obtaining an AUC of 0.98. Two other studies [47,48] applied CNN to detecting rotors in the intracardiac EGMs, with one of the them taking into account the interpretability of DL model and using gradient-weighted class-activation mapping (Grad-CAM) to explore how the model classifies AF electrical patterns [48], and the other designed the convolutional recurrent neural network from the perspective of improving model, while experimenting with multiple types of input signals, resulting in improved classification performance of the model.

Improving CA Strategies for AF

Due to varying degrees of individualization among patients, the development of a rational and effective strategy for AFCA relies on the judgment and surgical experience of the physician [16]. Notably, 17% (4/23) of the studies demonstrated the potential of DL to improve this process [50-53]. A common factor among these studies is that they were all based on imaging data. Liu et al [50] used the patient's preablation PV computed tomography (CT) images to construct a ResNet-based prediction model, which identified potential non-PV triggering factors before ablation treatment with an AUC of 0.88, thereby prompting physicians to perform necessary additional ablation on the basis of PV isolation. In 3 other studies, cardiac late gadolinium-enhanced cardiac magnetic resonance (LGE-MRI) was used [51-53]. Muffoletto et al [52] and Muizniece et al [51] used CNN and reinforcement learning (Q-Learning) algorithms, respectively, to directly engage in the process of formulating personalized ablation strategies for patients with AF and achieved success rates of 79% and 72%, respectively, in their respective test sets. Yang et al [53] proposed a joint segmentation method based on the multiview two-task recursive attention model to provide clinicians with segmented left ventricular (LV) anatomical structures and LV scars directly and simultaneously from LGE-MRI, which can be used to provide valuable guidance for ablation treatment.

Prognosis for Patients With AFCA

The high long-term recurrence rate of CA has been troubling patients and clinicians. Of the 23 included studies, 16 (70%) focused on the prognosis of patients after AFCA. Furthermore, 35% (8/23) of the studies [38,40-46] attempted to predict the risk of postprocedural recurrence of AF using easily accessible preprocedural patient clinical characteristics, with commonly used variables including age, gender, BMI, comorbidities, and electrocardiogram (ECG), but their modeling performance varied widely, with reported AUC ranging from 0.667 to 0.965. Four studies [31,32,34,37] chose to fuse features from multiple modalities as the input of ML model for recurrence prediction of AF. For example, Tang et al [37] proposed a CNN-based multimodal fusion framework to predict AF recurrence 1 year after CA using patients' clinical features, EGMs, ECG, and their combinations. The comparison of multiple experimental results showed that the recurrence prediction AUC of the multimodal feature fusion model was improved by 13.2% (0.86 vs 0.76) compared with the single-modality model. Independent prediction of AF recurrence using various types of examination images is also a feasible approach, but the relative performance is not particularly impressive. In contrast, several other studies have shown that models combining radiomics data with common clinical features exhibit predictive performance with better results [31,32,34].

Quality Assessment

The risk of bias and applicability of the included studies in subgroup A (5/23, 22%; targeting patients classification and image segmentation) were confirmed by using the adapted QUADAS-2 tool. The results presented in Figure 3 indicate that the "index test" domain is the most susceptible to bias. All studies exhibit a high risk of bias in this domain because they rely solely on single-center data for model development and lack external validation, which is essential to fully assess the model's generalizability. The applicability of 60% (3/5) of the studies showed unclear or high risk of bias in different fields.

For subgroup B (16/23, 70%; targeting patient prognosis prediction), the result indicated that the overall risk of bias and applicability was determined to be low in 25% (4/16) of the studies and high in 69% (11/16) of the studies, and the latter was also due to insufficient consideration of model generalization issues in the assessment of the "analysis" domain, resulting in a high risk of bias. Furthermore, 12% (2/16) of the studies were rated as uncertain risk in terms of applicability due to unclear inclusion criteria for the study participants. The results are shown in Figure 4.

The detailed risk of bias and applicability assessment results for both subgroups are shown in Figure S1 and Table S5 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 3. Summary of the assessment results of the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool on 5 patient classification and image segmentation studies in subgroup A.

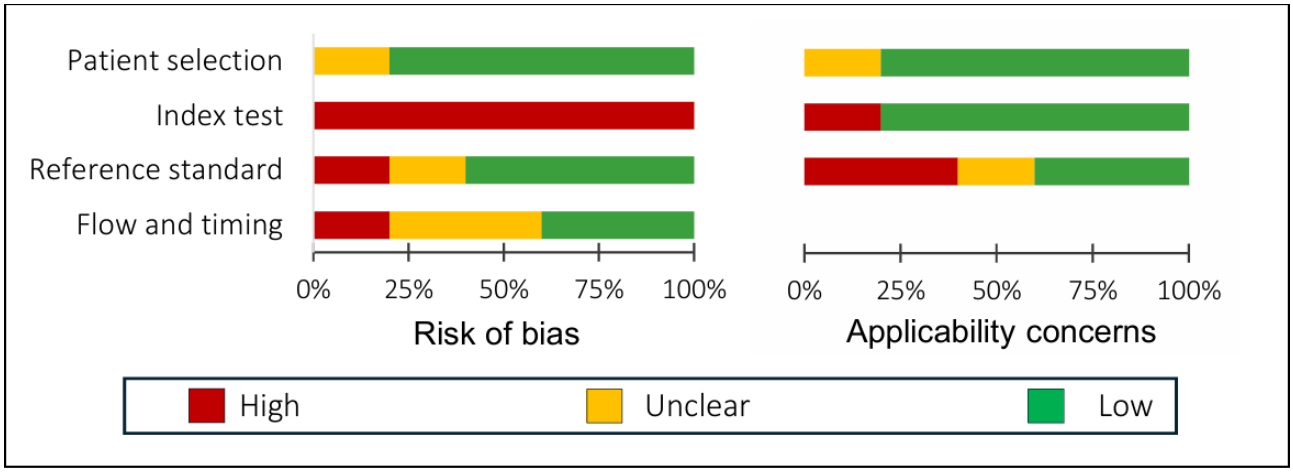

Figure 4. Summary of the assessment results of the Prediction model Risk Of Bias Assessment Tool (PROBAST) on 16 patient prognosis prediction studies in subgroup B.

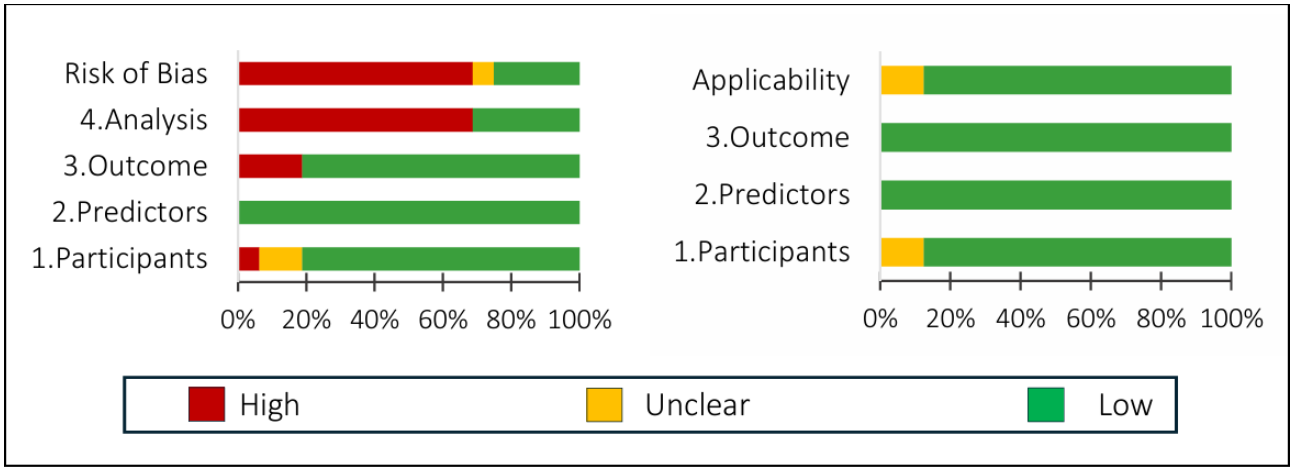

Discussion

Principal Findings

ML has shown great advantages in processing and analyzing medical data and has been increasingly applied to solve various medical problems [24], including those in the field of AF. Compared with several existing ML-related reviews in this field [55-57], we focus more on the research of applying ML in the management of patients undergoing AFCA. We systematically evaluated the 23 highly relevant studies and found that ML performed well overall in AFCA patient management. Models constructed in several studies performed as well as human clinicians in identifying potential ablation targets (rotors and focal sources) and assisting in improving ablation strategies (left atrium scar segmentation) [48,49,53]. It should be noted that comparisons between all included studies are hampered to some extent by the heterogeneity of ML results, and some findings may require discussion at a smaller granularity.

General Overview of the Studies

The results of our review suggest that the application of ML techniques to the prognosis of patients undergoing AFCA seems promising and has received the most attention from researchers within the scope of this review (16/23, 70%). Compared with traditional prediction models, ML models have a better prediction effect on the long-term recurrence of AF after CA. In the past decade, the construction of AF recurrence prediction models was mainly based on 2 methods; the first is to directly use clinical risk scores, such as APPLE [19], CAAP-AF [58], and MB-LATER [59]. The second is to use some specific biomarkers of patients with AF (such as LA volume [60] and LA sphericity index [61]) or to combine the scoring system with clinical variables and biomarkers for statistical inference. Several different evaluation studies have shown that the AUC range for the prediction of AF recurrence of the models constructed using the abovementioned methods is 0.55 to 0.67 [19,20,22], which is lower than the results of ML models reported in most of the included studies. Most of these ML models achieved performance values >0.8 in terms of AUC, accuracy, and F_1-score indicators. The reason for the poor predictive performance of the traditional methods may lie in the limitations of the predictor dimensions, as scoring and statistical models usually rely on previous literature to select a limited number of clinical variables and specific markers due to the need to fully consider clinical ease of use and interpretability, which also leads to some extent to the neglect of the interactions between variables.

XSL•FO
RenderX

In contrast, ML models are capable of handling much larger and diverse datasets, particularly including imaging and electrophysiologic signal data that were not directly amenable to statistical analysis in the past. Through comprehensive learning of complex data patterns and nonlinear relationships, ML models can achieve more accurate predictive results, which will have many positive impacts for effective intervention in the long-term health of patients undergoing AFCA. On the one hand, the predictive results provided by ML models will guide postprocedure medical management and expectations [37] and help clinicians to take interventions in advance, including medication, lifestyle modifications, or regular follow-ups; on the other hand, the application of ML models can also improve patient engagement in health management and treatment adherence, as in the case of the web calculator developed by Saglietto et al [44], which allows the free personalized prediction of the probability of recurrence after 1 year of AFCA treatment, serving as a decision-support tool that facilitates patient-centered care and joint decision-making between physicians and patients.

Similarly, optimizing CA strategies and identifying potential ablation targets could also benefit from the integration of ML. Because of the different mechanisms of AF triggering and maintenance among patients and the difficulty of identifying patients with one or more of several different mechanisms [16], physicians need to develop ablation strategies and identify ablation targets with the help of different means, such as imaging (including MRI and CT) or invasive electrophysiologic examinations. When dealing with some specific patient groups, such as patients with non–PV-triggered AF, this process not only tests the physician's clinical experience but also raises the patient's exposure time to radiation [62]. Several studies in this review demonstrated that the DL models can identify potential ablation targets based on the patient's EGM data, with 82.5% and 80.4% accuracy[47,49], or guide CA according to personalized appropriate strategies based on LGE-MRI images [51,52], which is conducive to guiding clinicians to implement targeted treatment and shorten the duration of radiation exposure. In addition, using ML technology and patient preoperative data to determine whether additional ablation is needed on the basis of PV isolation is also meaningful for improving patient selection and reducing the risk of postoperative recurrence in patients [31,50]. With the help of ML technology, physicians were given the opportunity to intuitively gain insights from these data into the mechanisms that initiate or maintain arrhythmias in patients, thereby efficiently locating ablation targets, reducing uncertainty in CA treatment and late recurrence, optimizing the allocation of medical resources, and ultimately improving patients' clinical outcomes.

In terms of model training, using multimodal data from patients with AF to develop ML models might be a promising approach to improve model performance. Models benefit from multimodal data that more fully reflect the demographic characteristics, clinical features, and actual cardiac condition with electrophysiological characteristics of patients with AF, as reported in some previous studies [31,32,34,37]. In traditional prognostic studies involving patients who have undergone ablation, factors such as AF type, ejection fraction values, LA volume [63], and obstructive sleep apnea [10] have been acknowledged as potential predictors that may be independently or jointly associated with recurrence. However, after comprehensively analyzing the cardiac morphological structure, electrophysiological structure, and clinical characteristics with the help of ML, factors that have been ignored, such as fractal characteristics [31,32], have attracted attention. This suggests that ML will reveal previously unknown associations and patterns that will drive new insights and innovations in the management of patients undergoing CA for AF. It is worth noting that the use of ML techniques to process multimodal data is common across the medical field (eg, oncology [64] and lymphocytosis [65]), but it needs to be further popularized in the field of AF, particularly for the management of patients undergoing AFCA.

In addition, compared with classical ML methods, DL seems to have more obvious advantages in the field of AFCA patient management. As shown in Tables 1 and 4, >60% (14/23, 61%) of the studies included in this scoping review used DL technology, and a notable characteristic of this subset of studies is that many involved data modalities beyond general clinical features, encompassing imaging data (CT, echocardiography, etc) or electrophysiological signal (ECG, EGM, etc). This is attributed to DL's demonstrated prowess in complex feature learning, time series processing, and other aspects, contributing to its gradual emergence as a mainstream technology in the field of computer-aided detection and diagnosis [66]. In a blinded, randomized controlled trial designed by He et al [67] for DL algorithm, the initial assessment of LV ejection fraction by DL was even superior to that of sonographers. Therefore, we believe that in the future, there will be more data-driven instances of applying DL to AFCA patient management to promote substantive progress in this field, such as optimizing treatment strategies, reducing the risk of complications, and conducting early intervention in patient populations prone to recurrence.

Existing Limitations of the ML Models

We also noted some general limitations to the current use of ML techniques in AFCA patient management, which have hindered its more comprehensive implementation. In using the adapted QUADAS-2 and PROBAST to assess the risk of bias for all included studies, we found that most of the studies (19/23, 83%) were conducted based on single-center, retrospective data and lacked external validation, which raises concerns as to whether ML models can be generalized to a high level. Limited by data availability, some ML models with excellent results will not be perfectly reproduced on other datasets. In addition, the absence of an external validation set may result in overestimating the model's performance, preventing an accurate assessment of the ML model's true capability [68]. Therefore, how to break through the limitations of single-center studies is an issue that has to be considered when further promoting the in-depth use of ML.

In addition, there remain some difficulties in effectively comparing and evaluating the results of ML models applied to the management of patients undergoing AFCA. We found that even for the similar medical task (eg, patient prognosis prediction or ablation strategy improvement), existing studies

XSL•FO
RenderX

still showed significant heterogeneity in the reporting and analysis of results [36,46], which hinders objective comparisons between the results of different ML models.

It is also worth noting that when considering the use and introduction of ML techniques from a practical perspective, we should correctly recognize that there are differences in the level of trust and acceptance of ML among different individuals (both physicians and patients). This phenomenon is partly caused by the "black box" problem of some DL models [69], which was rarely discussed in the studies included in this scoping review. Without the ability to understand how models make decisions, physicians will be more likely to trust their own clinical experience and medical knowledge to confirm treatment decisions based on interviews, symptom judgments, and clinical examinations, as they may not be willing to take responsibility for suboptimal results of ML models [70]. More critically, the "black box" prevents clinicians and researchers from fully understanding or explaining the principles behind model decisions [55], which is a limitation that cannot be ignored for research with the original intention of explaining medical problems.

Future Research

In future studies, the composition of the patient cohort and the construction of the dataset are primary issues that need to be considered. The challenges of conducting multicenter studies primarily stem from the legal, ethical, and privacy constraints associated with sharing raw AFCA patient data [71]. Distributed learning, represented by paradigms such as integrated learning and federated learning, would be a strategy for future studies to realize effective collaboration among multicenters [72]. In a network of distributed learning nodes, individual institutions do not need to share patient data externally. Instead, they contribute to constructing the final model across multiple centers by updating the model (eg, adjusting the weighting parameters) [73].

In addition, efforts should be accelerated to develop standardized protocols and consensus guidelines for the evaluation of ML models, thus helping clinicians in selecting more appropriate ML techniques to address specific clinical tasks in the management of patients undergoing AFCA [74,75].

To address the issue of limited clinical interpretability of ML models, commonly used techniques in medical research, such as Shapley Additive Explanation and attention mechanisms [45,76-78], can generate input importance maps to highlight the features that contribute most to model detection. We believe that more research will be devoted to opening this "black box" in the future [79]. Whether it is physicians, patients, or researchers, the real hope is that ML will be able to prompt as much human-understandable auxiliary information as possible to prove its usefulness as an additional decision point in the complex, changing, practice-based patient management environment of AFCA.

Strengths and Limitations

To the best of the authors' knowledge, this is the first scoping review of studies applying ML techniques to the management of patients undergoing AFCA. We summarized the various ML methods, variables, model results, and clinical application scenarios that have emerged; emphasized the huge potential for further in-depth application of ML and the limitations that need to be overcome; and provided guidance for future research in terms of data preparation and model selection. Moreover, some sources of evidence were critically evaluated using modified QUADAS-2 and PROBAST to make the conclusions of the review more reliable and convincing. This study has several limitations. First, the results of the included studies could not be synthesized into quantitative analysis. Although all studies reported model performance, different indicators were not directly comparable. Second, this review only focused on studies published in English and did not cover other languages and gray literature, which may have led to the omission of some important findings in this field.

Conclusions

The rapid development of ML has injected vitality into the transformation of AFCA patient management to become more precise and reliable. It could improve the identification of ablation targets, ablation strategies, and prediction of patient prognosis, which will effectively improve treatment effectiveness and patient management efficiency. Despite its broad prospects, the widespread and in-depth application of ML technology in AFCA patient management still requires consideration of issues from data and the model itself, such as data quality, model generalization, and model interpretability.

Acknowledgments

This study was supported by the Key Laboratory of Medical Information Research (Central South University), college of Hunan Province, as part of the project Clinical Research Center for Cardiovascular Intelligent Healthcare in Hunan Province (agreement 2021SK4005).

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1 and 2.

Authors' Contributions

AL and ZX contributed to the conception and design of the study. WC, HZ, WX, XC, and ZL developed the search strategy. WC, XC, ZL, and ZX contributed to the literature selection and data extraction. AL, WC, HZ, WX, and ZX contributed to risk of bias

XSL•FO
RenderX

and applicability evaluation. AL, WC, HZ, WX, and ZX wrote the first draft of the manuscript and edited the manuscript. All authors contributed to critical revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 104 KB - jmir_v27i1e60888_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 516 KB - jmir_v27i1e60888_app2.pdf]

References

1. Brundel BJ, Ai X, Hills MT, Kuipers MF, Lip GY, de Groot NM. Atrial fibrillation. Nat Rev Dis Primers 2022 Apr 07;8(1):21. [doi: 10.1038/s41572-022-00347-9] [Medline: 35393446]
2. Dong XJ, Wang BB, Hou FF, Jiao Y, Li HW, Lv SP, et al. Global burden of atrial fibrillation/atrial flutter and its attributable risk factors from 1990 to 2019. Europace 2023 Mar 30;25(3):793-803 [FREE Full text] [doi: 10.1093/europace/euac237] [Medline: 36603845]
3. Elliott AD, Middeldorp ME, Van Gelder IC, Albert CM, Sanders P. Epidemiology and modifiable risk factors for atrial fibrillation. Nat Rev Cardiol 2023 Jun;20(6):404-417. [doi: 10.1038/s41569-022-00820-8] [Medline: 36600003]
4. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. J Am Coll Cardiol 2014 Dec 02;64(21):e1-76 [FREE Full text] [doi: 10.1016/j.jacc.2014.03.022] [Medline: 24685669]
5. Seiffge DJ, Werring DJ, Paciaroni M, Dawson J, Warach S, Milling TJ, et al. Timing of anticoagulation after recent ischaemic stroke in patients with atrial fibrillation. Lancet Neurol 2019 Jan;18(1):117-126 [FREE Full text] [doi: 10.1016/S1474-4422(18)30356-9] [Medline: 30415934]
6. Pabel S, Knierim M, Stehle T, Alebrand F, Paulus M, Sieme M, et al. Effects of atrial fibrillation on the human ventricle. Circ Res 2022 Apr;130(7):994-1010. [doi: 10.1161/circresaha.121.319718]
7. Frederiksen TC, Dahm CC, Preis SR, Lin H, Trinquart L, Benjamin EJ, et al. The bidirectional association between atrial fibrillation and myocardial infarction. Nat Rev Cardiol 2023 Sep 17;20(9):631-644. [doi: 10.1038/s41569-023-00857-3] [Medline: 37069297]
8. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol 2014;6:213-220 [FREE Full text] [doi: 10.2147/CLEP.S47385] [Medline: 24966695]
9. Burdett P, Lip GY. Atrial fibrillation in the UK: predicting costs of an emerging epidemic recognizing and forecasting the cost drivers of atrial fibrillation-related costs. Eur Heart J Qual Care Clin Outcomes 2022 Mar 02;8(2):187-194. [doi: 10.1093/ehjqcco/qcaa093] [Medline: 33346822]
10. Dewire J, Calkins H. Update on atrial fibrillation catheter ablation technologies and techniques. Nat Rev Cardiol 2013 Oct 27;10(10):599-612. [doi: 10.1038/nrcardio.2013.121] [Medline: 23979215]
11. Cappato R, Calkins H, Chen SA, Davies W, Iesaka Y, Kalman J, et al. Updated worldwide survey on the methods, efficacy, and safety of catheter ablation for human atrial fibrillation. Circ Arrhythm Electrophysiol 2010 Feb;3(1):32-38. [doi: 10.1161/circep.109.859116]
12. Wazni OM, Marrouche NF, Martin DO, Verma A, Bhargava M, Saliba W, et al. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of symptomatic atrial fibrillation: a randomized trial. JAMA 2005 Jun 01;293(21):2634-2640. [doi: 10.1001/jama.293.21.2634] [Medline: 15928285]
13. Ouyang F, Tilz R, Chun J, Schmidt B, Wissner E, Zerm T, et al. Long-term results of catheter ablation in paroxysmal atrial fibrillation: lessons from a 5-year follow-up. Circulation 2010 Dec 07;122(23):2368-2377. [doi: 10.1161/circulationaha.110.946806]
14. Medi C, Sparks PB, Morton JB, Kistler PM, Halloran K, Rosso R, et al. Pulmonary vein antral isolation for paroxysmal atrial fibrillation: results from long-term follow-up. J Cardiovasc Electrophysiol 2011 Feb;22(2):137-141. [doi: 10.1111/j.1540-8167.2010.01885.x] [Medline: 20812937]
15. Gaita F, Caponi D, Scaglione M, Montefusco A, Corleto A, Di Monte F, et al. Long-term clinical results of 2 different ablation strategies in patients with paroxysmal and persistent atrial fibrillation. Circ Arrhythm Electrophysiol 2008 Oct;1(4):269-275. [doi: 10.1161/circep.108.774885]

XSL•FO
RenderX

16. Latchamsetty R, Morady F. Atrial fibrillation ablation. Annu Rev Med 2018 Jan 29;69(1):53-63. [doi: 10.1146/annurev-med-041316-090015] [Medline: 28806148]
17. Parameswaran R, Al-Kaisey AM, Kalman JM. Catheter ablation for atrial fibrillation: current indications and evolving technologies. Nat Rev Cardiol 2021 Mar 13;18(3):210-225. [doi: 10.1038/s41569-020-00451-x] [Medline: 33051613]
18. Lee KN, Roh SY, Baek YS, Park H, Ahn J, Kim D, et al. Long-term clinical comparison of procedural end points after pulmonary vein isolation in paroxysmal atrial fibrillation. Circ Arrhythm Electrophysiol 2018 Feb;11(2):e005019. [doi: 10.1161/circep.117.005019]
19. Kornej J, Hindricks G, Shoemaker MB, Husser D, Arya A, Sommer P, et al. The APPLE score: a novel and simple score for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation. Clin Res Cardiol 2015 Oct 17;104(10):871-876 [FREE Full text] [doi: 10.1007/s00392-015-0856-x] [Medline: 25876528]
20. Jacobs V, May HT, Bair TL, Crandall BG, Cutler M, Day JD, et al. The impact of risk score (CHADS2 versus CHA2DS2-VASc) on long-term outcomes after atrial fibrillation ablation. Heart Rhythm 2015 Apr;12(4):681-686. [doi: 10.1016/j.hrthm.2014.12.034] [Medline: 25546809]
21. Kornej J, Hindricks G, Kosiuk J, Arya A, Sommer P, Husser D, et al. Comparison of CHADS2, R2CHADS2, and CHA2DS2-VASc scores for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation: the Leipzig Heart Center AF ablation registry. Circ Arrhythm Electrophysiol 2014 Apr;7(2):281-287. [doi: 10.1161/circep.113.001182]
22. Mulder MJ, Kemme MJ, Hopman LH, Kuşgözoğlu E, Gülçiçek H, van de Ven PM, et al. Comparison of the predictive value of ten risk scores for outcomes of atrial fibrillation patients undergoing radiofrequency pulmonary vein isolation. Int J Cardiol 2021 Dec 01;344:103-110 [FREE Full text] [doi: 10.1016/j.ijcard.2021.09.029] [Medline: 34555444]
23. Mohsen F, Al-Absi HR, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. NPJ Digit Med 2023 Oct 25;6(1):197 [FREE Full text] [doi: 10.1038/s41746-023-00933-5] [Medline: 37880301]
24. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
25. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
26. Gao Y, Zeng S, Xu X, Li H, Yao S, Song K, et al. Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study. Lancet Digit Health 2022 Mar;4(3):e179-e187 [FREE Full text] [doi: 10.1016/S2589-7500(21)00278-8] [Medline: 35216752]
27. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. Circulation 2022 Jan 11;145(2):122-133. [doi: 10.1161/circulationaha.121.057480]
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
29. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
30. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019 Jan 01;170(1):W1-33 [FREE Full text] [doi: 10.7326/M18-1377] [Medline: 30596876]
31. Firouznia M, Feeny AK, LaBarbera MA, McHale M, Cantlay C, Kalfas N, et al. Machine learning–derived fractal features of shape and texture of the left atrium and pulmonary veins from cardiac computed tomography scans are associated with risk of recurrence of atrial fibrillation postablation. Circ Arrhythm Electrophysiol 2021 Mar;14(3):e009265. [doi: 10.1161/circep.120.009265]
32. Atta-Fosu T, LaBarbera M, Ghose S, Schoenhagen P, Saliba W, Tchou PJ, et al. A new machine learning approach for predicting likelihood of recurrence following ablation for atrial fibrillation from CT. BMC Med Imaging 2021 Mar 09;21(1):45 [FREE Full text] [doi: 10.1186/s12880-021-00578-4] [Medline: 33750343]
33. Shade JK, Ali RL, Basile D, Popescu D, Akhtar T, Marine JE, et al. Preprocedure application of machine learning and mechanistic simulations predicts likelihood of paroxysmal atrial fibrillation recurrence following pulmonary vein isolation. Circ Arrhythm Electrophysiol 2020 Jul;13(7):e008213. [doi: 10.1161/circep.119.008213]
34. Roney CH, Sim I, Yu J, Beach M, Mehta A, Alonso Solis-Lemus J, et al. Predicting atrial fibrillation recurrence by combining population data and virtual cohorts of patient-specific left atrial models. Circ Arrhythm Electrophysiol 2022 Feb;15(2):e010253. [doi: 10.1161/circep.121.010253]
35. Hwang YT, Lee HL, Lu CH, Chang P, Wo H, Liu H, et al. A novel approach for predicting atrial fibrillation recurrence after ablation using deep convolutional neural networks by assessing left atrial curved m-mode speckle-tracking images. Front Cardiovasc Med 2020 Jan 22;7:605642 [FREE Full text] [doi: 10.3389/fcvm.2020.605642] [Medline: 33553257]
36. An Q, McBeth R, Zhou H, Lawlor B, Nguyen D, Jiang S, et al. Prediction of type and recurrence of atrial fibrillation after catheter ablation via left atrial electroanatomical voltage mapping registration and multilayer perceptron classification: a

XSL•FO
RenderX

retrospective study. Sensors (Basel) 2022 May 27;22(11):4058 [FREE Full text] [doi: 10.3390/s22114058] [Medline: 35684678]
37. Tang S, Razeghi O, Kapoor R, Alhusseini MI, Fazal M, Rogers AJ, et al. Machine learning–enabled multimodal fusion of intra-atrial and body surface signals in prediction of atrial fibrillation ablation outcomes. Circ Arrhythm Electrophysiol 2022 Aug;15(8):e010850. [doi: 10.1161/circep.122.010850]
38. Li Z, Feng X, Wu Z, Yang C, Bai B, Yang Q. Classification of atrial fibrillation recurrence based on a convolution neural network with SVM architecture. IEEE Access 2019;7:77849-77856. [doi: 10.1109/access.2019.2920900]
39. Saiz-Vivo J, Corino VD, Hatala R, de Melis M, Mainardi LT. Heart rate variability and clinical features as predictors of atrial fibrillation recurrence after catheter ablation: a pilot study. Front Physiol 2021 May 25;12:672896 [FREE Full text] [doi: 10.3389/fphys.2021.672896] [Medline: 34113264]
40. Jiang J, Deng H, Liao H, Fang X, Zhan X, Wei W, et al. An artificial intelligence-enabled ECG algorithm for predicting the risk of recurrence in patients with paroxysmal atrial fibrillation after catheter ablation. J Clin Med 2023 Mar 01;12(5):1933. [doi: 10.3390/jcm12051933] [Medline: 36902719]
41. Lee DI, Park MJ, Choi JW, Park S. Deep learning model for predicting rhythm outcomes after radiofrequency catheter ablation in patients with atrial fibrillation. J Healthc Eng 2022 Sep 10;2022:2863495-2863410 [FREE Full text] [doi: 10.1155/2022/2863495] [Medline: 36124238]
42. Park JW, Kwon OS, Shim J, Hwang I, Kim YG, Yu HT, et al. Machine learning-predicted progression to permanent atrial fibrillation after catheter ablation. Front Cardiovasc Med 2022 Feb 16;9:813914 [FREE Full text] [doi: 10.3389/fcvm.2022.813914] [Medline: 35252393]
43. Hung M, Lauren E, Hon E, Xu J, Ruiz-Negrón B, Rosales M, et al. Using machine learning to predict 30-day hospital readmissions in patients with atrial fibrillation undergoing catheter ablation. J Pers Med 2020 Aug 09;10(3):82 [FREE Full text] [doi: 10.3390/jpm10030082] [Medline: 32784873]
44. Saglietto A, Gaita F, Blomstrom-Lundqvist C, Arbelo E, Dagres N, Brugada J, et al. AFA-Recur: an ESC EORP AFA-LT registry machine-learning web calculator predicting atrial fibrillation recurrence after ablation. Europace 2023 Feb 08;25(1):92-100 [FREE Full text] [doi: 10.1093/europace/euac145] [Medline: 36006664]
45. Ma Y, Zhang D, Xu J, Pang H, Hu M, Li J, et al. Explainable machine learning model reveals its decision-making process in identifying patients with paroxysmal atrial fibrillation at high risk for recurrence after catheter ablation. BMC Cardiovasc Disord 2023 Feb 17;23(1):91 [FREE Full text] [doi: 10.1186/s12872-023-03087-0] [Medline: 36803424]
46. Zhou X, Nakamura K, Sahara N, Takagi T, Toyoda Y, Enomoto Y, et al. Deep learning-based recurrence prediction of atrial fibrillation after catheter ablation. Circ J 2022 Jan 25;86(2):299-308 [FREE Full text] [doi: 10.1253/circj.CJ-21-0622] [Medline: 34629373]
47. Ríos-Muñoz GR, Fernández-Avilés F, Arenal Á. Convolutional neural networks for mechanistic driver detection in atrial fibrillation. Int J Mol Sci 2022 Apr 11;23(8):4216 [FREE Full text] [doi: 10.3390/ijms23084216] [Medline: 35457044]
48. Alhusseini MI, Abuzaid F, Rogers AJ, Zaman JA, Baykaner T, Clopton P, et al. Machine learning to classify intracardiac electrical patterns during atrial fibrillation. Circ Arrhythm Electrophysiol 2020 Aug;13(8):e008160. [doi: 10.1161/circep.119.008160]
49. Liao S, Ragot D, Nayyar S, Suszko A, Zhang Z, Wang B, et al. Deep learning classification of unipolar electrograms in human atrial fibrillation: application in focal source mapping. Front Physiol 2021 Jul 30;12:704122 [FREE Full text] [doi: 10.3389/fphys.2021.704122] [Medline: 34393823]
50. Liu CM, Chang SL, Chen HH, Chen WS, Lin YJ, Lo LW, et al. The clinical application of the deep learning technique for predicting trigger origins in patients with paroxysmal atrial fibrillation with catheter ablation. Circ Arrhythm Electrophysiol 2020 Nov;13(11):e008518. [doi: 10.1161/CIRCEP.120.008518] [Medline: 33021404]
51. Muizniece L, Bertagnoli A, Qureshi A, Zeidan A, Roy A, Muffoletto M, et al. Reinforcement learning to improve image-guidance of ablation therapy for atrial fibrillation. Front Physiol 2021 Aug 25;12:733139. [doi: 10.3389/fphys.2021.733139] [Medline: 34512401]
52. Muffoletto M, Qureshi A, Zeidan A, Muizniece L, Fu X, Zhao J, et al. Toward patient-specific prediction of ablation strategies for atrial fibrillation using deep learning. Front Physiol 2021 May 26;12:674106 [FREE Full text] [doi: 10.3389/fphys.2021.674106] [Medline: 34122144]
53. Yang G, Chen J, Gao Z, Li S, Ni H, Angelini E, et al. Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. Future Gener Comput Syst 2020 Jun;107:215-228 [FREE Full text] [doi: 10.1016/j.future.2020.02.005] [Medline: 32494091]
54. Chauhan VS, Verma A, Nayyar S, Timmerman N, Tomlinson G, Porta-Sanchez A, et al. Focal source and trigger mapping in atrial fibrillation: randomized controlled trial evaluating a novel adjunctive ablation strategy. Heart Rhythm 2020 May;17(5):683-691. [doi: 10.1016/j.hrthm.2019.12.011] [Medline: 31991116]
55. Olier I, Ortega-Martorell S, Pieroni M, Lip GY. How machine learning is impacting research in atrial fibrillation: implications for risk prediction and future management. Cardiovasc Res 2021 Jun 16;117(7):1700-1717 [FREE Full text] [doi: 10.1093/cvr/cvab169] [Medline: 33982064]

XSL•FO
RenderX

56. Wegner FK, Plagwitz L, Doldi F, Ellermann C, Willy K, Wolfes J, et al. Machine learning in the detection and management of atrial fibrillation. Clin Res Cardiol 2022 Sep;111(9):1010-1017 [FREE Full text] [doi: 10.1007/s00392-022-02012-3] [Medline: 35353207]
57. Siontis KC, Yao X, Pirruccello JP, Philippakis AA, Noseworthy PA. How will machine learning inform the clinical care of atrial fibrillation? Circ Res 2020 Jun 19;127(1):155-169. [doi: 10.1161/circresaha.120.316401]
58. Winkle RA, Jarman JW, Mead RH, Engel G, Kong MH, Fleming W, et al. Predicting atrial fibrillation ablation outcome: the CAAP-AF score. Heart Rhythm 2016 Nov;13(11):2119-2125 [FREE Full text] [doi: 10.1016/j.hrthm.2016.07.018] [Medline: 27435586]
59. Mujović N, Marinković M, Marković N, Shantsila A, Lip GY, Potpara TS. Prediction of very late arrhythmia recurrence after radiofrequency catheter ablation of atrial fibrillation: the MB-LATER clinical score. Sci Rep 2017 Jan 20;7(1):40828 [FREE Full text] [doi: 10.1038/srep40828] [Medline: 28106147]
60. Chollet L, Iqbal SU, Wittmer S, Thalmann G, Madaffari A, Kozhuharov N, et al. Impact of atrial fibrillation phenotype and left atrial volume on outcome after pulmonary vein isolation. Europace 2024 Mar 30;26(4):258 [FREE Full text] [doi: 10.1093/europace/euae071] [Medline: 38597211]
61. Shi J, Xu S, Chen L, Wu B, Yang K, Chen S, et al. Impact of left atrial sphericity index on the outcome of catheter ablation for atrial fibrillation. J Cardiovasc Transl Res 2021 Oct 06;14(5):912-920. [doi: 10.1007/s12265-020-10093-6] [Medline: 33409962]
62. Della Rocca DG, Di Biase L, Mohanty S, Trivedi C, Gianni C, Romero J, et al. Targeting non-pulmonary vein triggers in persistent atrial fibrillation: results from a prospective, multicentre, observational registry. Europace 2021 Dec 07;23(12):1939-1949. [doi: 10.1093/europace/euab161] [Medline: 34417816]
63. Balk EM, Garlitski AC, Alsheikh-Ali AA, Terasawa T, Chung M, Ip S. Predictors of atrial fibrillation recurrence after radiofrequency catheter ablation: a systematic review. J Cardiovasc Electrophysiol 2010 Nov 29;21(11):1208-1216. [doi: 10.1111/j.1540-8167.2010.01798.x] [Medline: 20487117]
64. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer 2022 Feb 18;22(2):114-126 [FREE Full text] [doi: 10.1038/s41568-021-00408-3] [Medline: 34663944]
65. Sahasrabudhe M, Sujobert P, Zacharaki EI, Maurin E, Grange B, Jallades L, et al. Deep multi-instance learning using multi-modal data for diagnosis of lymphocytosis. IEEE J Biomed Health Inform 2021 Jun;25(6):2125-2136. [doi: 10.1109/JBHI.2020.3038889] [Medline: 33206611]
66. Feeny AK, Chung MK, Madabhushi A, Attia ZI, Cikes M, Firouznia M, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. Circ Arrhythm Electrophysiol 2020 Aug;13(8):e007952 [FREE Full text] [doi: 10.1161/CIRCEP.119.007952] [Medline: 32628863]
67. He B, Kwan AC, Cho JH, Yuan N, Pollick C, Shiota T, et al. Blinded, randomized trial of sonographer versus AI cardiac function assessment. Nature 2023 Apr 05;616(7957):520-524 [FREE Full text] [doi: 10.1038/s41586-023-05947-3] [Medline: 37020027]
68. Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns (N Y) 2020 Nov 13;1(8):100129 [FREE Full text] [doi: 10.1016/j.patter.2020.100129] [Medline: 33294870]
69. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May 13;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
70. Rasheed K, Qayyum A, Ghaly M, Al-Fuqaha A, Razi A, Qadir J. Explainable, trustworthy, and ethical machine learning for healthcare: a survey. Comput Biol Med 2022 Oct;149:106043 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.106043] [Medline: 36115302]
71. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging 2018 Oct;9(5):745-753 [FREE Full text] [doi: 10.1007/s13244-018-0645-y] [Medline: 30112675]
72. Kirienko M, Sollini M, Ninatti G, Loiacono D, Giacomello E, Gozzi N, et al. Distributed learning: a reliable privacy-preserving strategy to change multicenter collaborations using AI. Eur J Nucl Med Mol Imaging 2021 Nov 13;48(12):3791-3804 [FREE Full text] [doi: 10.1007/s00259-021-05339-7] [Medline: 33847779]
73. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med 2020 Sep 14;3(1):119 [FREE Full text] [doi: 10.1038/s41746-020-00323-1] [Medline: 33015372]
74. Chakraborty C, Bhattacharya M, Pal S, Lee S. From machine learning to deep learning: advances of the recent data-driven paradigm shift in medicine and healthcare. Curr Res Biotechnol 2024;7:100164. [doi: 10.1016/j.crbiot.2023.100164]
75. Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare. Injury 2023 May;54 Suppl 3:S69-S73. [doi: 10.1016/j.injury.2022.01.046] [Medline: 35135685]
76. Elul Y, Rosenberg AA, Schuster A, Bronstein AM, Yaniv Y. Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning-based ECG analysis. Proc Natl Acad Sci U S A 2021 Jun 15;118(24):e1541 [FREE Full text] [doi: 10.1073/pnas.2020620118] [Medline: 34099565]

XSL•FO
RenderX

77. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. Crit Care 2020 Jul 31;24(1):478 [FREE Full text] [doi: 10.1186/s13054-020-03179-9] [Medline: 32736589]
78. Zhao Y, Cao LY, Zhao YX, Wang F, Xie L, Xing H, et al. Medical record data-enabled machine learning can enhance prediction of left atrial appendage thrombosis in nonvalvular atrial fibrillation. Thromb Res 2023 Mar;223:174-183. [doi: 10.1016/j.thromres.2023.01.001]
79. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019 Apr 02;9(4):e1312 [FREE Full text] [doi: 10.1002/widm.1312] [Medline: 32089788]

Abbreviations

AF: atrial fibrillation
AFCA: atrial fibrillation catheter ablation
AUC: area under the receiver operating characteristic curve
CA: catheter ablation
CNN: convolutional neural network
CT: computed tomography
DL: deep learning
ECG: electrocardiogram
EGM: electrogram
EHR: electronic health record
LGE-MRI: late gadolinium-enhanced cardiac magnetic resonance
LV: left ventricular
ML: machine learning
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PROBAST: Prediction model Risk Of Bias Assessment Tool
PV: pulmonary vein
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies-2

Edited by A Mavragani; submitted 24.05.24; peer-reviewed by S Mao, WC Lee, L Dall'Olio; comments to author 02.12.24; revised version received 21.12.24; accepted 30.12.24; published 10.02.25.

Please cite as:
Luo A, Chen W, Zhu H, Xie W, Chen X, Liu Z, Xin Z
Machine Learning in the Management of Patients Undergoing Catheter Ablation for Atrial Fibrillation: Scoping Review
J Med Internet Res 2025;27:e60888
URL: https://www.jmir.org/2025/1/e60888
doi:10.2196/60888
PMID:

©Aijing Luo, Wei Chen, Hongtao Zhu, Wenzhao Xie, Xi Chen, Zhenjiang Liu, Zirui Xin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Trends and Gaps in Digital Precision Hypertension Management: Scoping Review

Namuun Clifford[1], MSN; Rachel Tunis[2], BS; Adetimilehin Ariyo[1], BS; Haoxiang Yu[3], MS; Hyekyun Rhee[1], PhD; Kavita Radhakrishnan[1], MSEE, PhD

[1]School of Nursing, The University of Texas at Austin, Austin, TX, United States

[2]School of Information, The University of Texas at Austin, Austin, TX, United States

[3]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Namuun Clifford, MSN
School of Nursing
The University of Texas at Austin
1710 Red River St
Austin, TX, 78701
United States
Phone: 1 (512) 471 7913
Email: namuun.clifford@utexas.edu

Abstract

Background: Hypertension (HTN) is the leading cause of cardiovascular disease morbidity and mortality worldwide. Despite effective treatments, most people with HTN do not have their blood pressure under control. Precision health strategies emphasizing predictive, preventive, and personalized care through digital tools offer notable opportunities to optimize the management of HTN.

Objective: This scoping review aimed to fill a research gap in understanding the current state of precision health research using digital tools for the management of HTN in adults.

Methods: This study used a scoping review framework to systematically search for articles in 5 databases published between 2013 and 2023. The included articles were thematically analyzed based on their precision health focus: personalized interventions, prediction models, and phenotyping. Data were extracted and summarized for study and sample characteristics, precision health focus, digital health technology, disciplines involved, and characteristics of personalized interventions.

Results: After screening 883 articles, 46 were included; most studies had a precision health focus on personalized digital interventions (34/46, 74%), followed by prediction models (8/46, 17%) and phenotyping (4/46, 9%). Most studies (38/46, 82%) were conducted in or used data from North America or Europe, and 63% (29/46) of the studies came exclusively from the medical and health sciences, with 33% (15/46) of studies involving 2 or more disciplines. The most commonly used digital technologies were mobile phones (33/46, 72%), blood pressure monitors (18/46, 39%), and machine learning algorithms (11/46, 24%). In total, 45% (21/46) of the studies either did not report race or ethnicity data (14/46, 30%) or partially reported this information (7/46, 15%). For personalized intervention studies, nearly half (14/30, 47%) used 2 or less types of data for personalization, with only 7% (2/30) of the studies using social determinants of health data and no studies using physical environment or digital literacy data. Personalization characteristics of studies varied, with 43% (13/30) of studies using fully automated personalization approaches, 33% (10/30) using human-driven personalization, and 23% (7/30) using a hybrid approach.

Conclusions: This scoping review provides a comprehensive mapping of the literature on the current trends and gaps in digital precision health research for the management of HTN in adults. Personalized digital interventions were the primary focus of most studies; however, the review highlighted the need for more precise definitions of *personalization* and the integration of more diverse data sources to improve the tailoring of interventions and promotion of health equity. In addition, there were significant gaps in the reporting of race and ethnicity data of participants, underuse of wearable devices for passive data collection, and the need for greater interdisciplinary collaboration to advance precision health research in digital HTN management.

Trial Registration: OSF Registries osf.io/yuzf8; https://osf.io/yuzf8

(J Med Internet Res 2025;27:e59841) doi:10.2196/59841

XSL•FO
RenderX

KEYWORDS

precision health; hypertension; digital health; prediction models; personalization; phenotyping; machine learning; algorithms; mobile apps; mobile health

Introduction

Background

Hypertension (HTN) is the leading preventable cause of cardiovascular disease and premature mortality worldwide, affecting an estimated 1.28 billion adults [1]. It surpasses smoking, diabetes, and obesity as the most significant modifiable risk factor, contributing to 54% of stroke and 47% of ischemic heart disease cases [2,3]. Despite mounting evidence that antihypertensive treatment can reduce morbidity and mortality, HTN remains underdiagnosed and undertreated [4,5]. Globally, nearly half of adults with HTN are unaware that they have the condition, and only 21% have their blood pressure (BP) under recommended levels [1]. Significant disparities exist in HTN prevalence and management. While 82% of individuals with HTN live in low- and middle-income countries, only 7.7% achieve BP control, compared to 28.4% in high-income countries [4,6]. These disparities have been linked to various social and environmental determinants that disproportionately affect individuals from racial and ethnic minority groups [7]. Given the considerable global health burden and inequities stemming from suboptimal HTN management, there is an urgent demand for innovative and scalable solutions for effective HTN prevention and control.

Challenges in HTN Management

Despite the availability of effective treatments, only 21% of adults with HTN worldwide have their BP under control [6]. Longitudinal analyses have shown that up to 50% of patients stop taking their prescribed medications within 1 year; this is attributed to factors such as sociodemographics, medication side effects, lack of knowledge, comorbidities, lack of access to care, and patient-clinician relationship [8,9]. Current HTN treatment approaches are based on evidence from randomized controlled trials (RCTs) that reflect the mean results for the average patient [10]. This “one-size-fits-all” approach fails to consider the wide variation in an individual’s genetic, biological, behavioral, sociodemographic, and environmental factors that profoundly influence HTN treatment adherence and outcomes. This lack of personalization contributes to suboptimal treatment adherence and poor overall HTN control [9]. Effective management strategies must incorporate more nuanced approaches that adapt to the unique characteristics and contexts of individuals within diverse populations.

Digital Precision Health for HTN Management

The emergence of precision health as a paradigm to empower individuals, predict and prevent disease before it starts, and personalize care addresses these challenges and presents a promising road map for transforming HTN management [11]. Expanding on precision medicine’s focus on personalized medical care, precision health goes beyond treatment, emphasizing health promotion and disease prevention for a more proactive approach to addressing population health [11-14]. Interventions in precision health are customized to the individual’s unique variations in genetic, biological, behavioral, sociocultural, and environmental determinants to improve health outcomes [14-16]. While personalized care has always been a goal in clinical practice, it is only with recent technological advancements in artificial intelligence, data analytics, and digital tools that the depth of personalization and predictive capabilities in disease risk or treatment response are becoming truly attainable [16]. Digital health technologies, including mobile apps, wearable devices, and remote monitoring, enable the harnessing of data to tailor interventions, predict disease risk, and engage patients more effectively in their care [11]. These tools present an opportunity to overcome traditional care barriers, facilitating more accessible, accurate, timely, and personalized HTN prevention and management strategies.

Objectives

Although existing research has examined the effectiveness of digital interventions in improving health outcomes for HTN, there is a notable lack of prior reviews that integrate these findings in the context of precision health [12-15]. The primary objective of this scoping review is to synthesize the existing studies pertaining to precision health using various digital health technologies for the management of HTN in adults. Specifically, we aimed to address the following questions:

1. Which aspects of precision health are addressed (ie, prediction models, personalization, and phenotyping), and what are the characteristics of these studies?
2. Which digital health tools are being used in precision HTN management?
3. What are the characteristics of participants in precision health studies for HTN management?
4. For personalized interventions, what types of data are used for personalization, and what are the characteristics of personalization?

By mapping out the current state of digital precision health for HTN management, we seek to provide a foundation for future research, policy, and the development of effective and equitable interventions to improve the prevention, diagnosis, and management of HTN on a global scale.

Methods

Design

This scoping literature review was conducted following the framework for scoping reviews developed by Arksey and O’Malley [17]. This framework consists of five stages: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results. Thematic analysis of findings were conducted based on the 3 categories of studies included in the review: personalized interventions, phenotyping, and prediction models. The review used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews)

XSL•FO
RenderX

reporting guidelines to increase methodological transparency (Multimedia Appendix 1) [17]. The review protocol was registered with the OSF Registries (osf.io/yuzf8).

Search Methods

In July 2023, assisted by a research librarian, we conducted comprehensive searches across 5 databases: PubMed, CINAHL, Web of Science, Embase, and Inspec. We filtered the search for peer-reviewed articles published between 2013 and 2023 in the English language. This 10-year time frame was chosen to capture the most recent and relevant advancements in digital health technologies and their application in HTN management. Search terms included "precision health" (eg, *precision medicine* OR *personalized* OR *tailored* OR *individualized*) AND "digital health" (*telemedicine* OR *telehealth* OR *mobile apps* OR *wearable electronic devices* OR *electronic health*) AND "hypertension" (*high blood pressure* OR *elevated blood pressure*). Medical subject headings terms were adapted across databases. The full search strategy is provided in Multimedia Appendix 2.

Final search results were transferred to an EndNote (version 20) database, and duplicates were removed. The remaining articles were imported into Covidence, a web-based software platform designed for conducting literature reviews. Blinded screening of title, abstract, and full texts was conducted by 3 authors (NC, RT, and AA), with each study independently reviewed by 2 of these authors. Any conflicts in opinion were reconciled through discussion or, if needed, by involving a third reviewer (HR).

Eligibility Criteria

This scoping review sought to comprehensively map key concepts within the identified research area by including all types of research designs of original, peer-reviewed research papers. In addition, studies were included if they (1) sampled adults aged ≥18 years, (2) contained a diagnosis of HTN, (3) had a precision health focus, and (4) used a digital health technology. If studies included participants with >1 diagnosis (ie, HTN and diabetes mellitus), and HTN management was a primary focus of the intervention, they were included. To identify studies with a focus on precision health, if the term precision health or its synonyms (ie, personalized, tailored, individualized, predictive, and phenotype) were in the title, abstract, or main text, they were included. For the detailed eligibility criteria, refer to Textbox 1.

Textbox 1. Eligibility criteria for scoping review.

Inclusion criteria

- Study type: original peer-reviewed research paper
- Period: studies published between January 1, 2013, and July 20, 2023
- Language: English
- Population: adults aged ≥18 years with a diagnosis of hypertension
- Had a precision health focus (ie, predictive, personalized, or phenotyping)
- Used a digital health technology (eg, mobile phones, telemedicine, wearable devices, or electronic health records)

Exclusion criteria

- Study types: case studies, editorials, opinion pieces, gray literature, dissertations, literature reviews, and trial protocols
- Period: studies published before January 1, 2013, or after July 20, 2023
- Language: any language other than English
- Population: pediatric or pregnant patients
- Pulmonary artery hypertension diagnosis
- Interventions targeting health care providers

Studies were excluded if they involved pediatric or pregnant participants, had a diagnosis of pulmonary artery HTN, or reported interventions targeting health care providers. Pregnant individuals were excluded due to the unique physiological and treatment differences in managing HTN during pregnancy, such as gestational HTN or preeclampsia, which is considerably outside the scope of this review. Furthermore, we excluded case studies, editorials, opinion pieces, gray literature, dissertations, literature reviews, and trial protocols, as these are secondary or nonempirical sources lacking original research data essential for our analysis.

Data Extraction

We developed 3 separate data extraction templates, each containing fields for all identified key data elements. The first template was used to extract data on study characteristics, including study location, design, sample characteristics, precision health focus, digital health technology used, and disciplines involved. The other templates were used to extract data on personalization characteristics and types of data used for personalization of tailored intervention studies. For each included study, the first author (NC) and 1 of the 3 coauthors (RT, AA, and HY) independently extracted data using the standardized templates. This dual-reviewer approach was used to minimize errors and biases in the data extraction process.

XSL•FO
RenderX

Any discrepancies between reviewers in the data extraction were resolved through discussion, with a third author available for consultation if consensus could not be reached.

Data Synthesis

Included studies were categorized based on their precision health focus: personalization, phenotyping, and prediction models. We conducted thematic analysis to identify recurring themes and patterns across the studies, with themes derived both deductively from the research objectives and inductively from the study findings [18]. Studies were cross-compared to highlight differences and similarities in methodologies, designs, and outcomes. Synthesizing the extracted data allowed us to identify gaps in the current research landscape, generating recommendations for future research directions which we discuss in this study.

Results

Search Results

The search yielded 883 studies after the removal of duplicates; after title and abstract screening, we identified 104 (12%) studies for full-text review. Of the 104 studies, a final sample of 46 (44%) studies met the inclusion criteria for this review (Figure 1). In alignment with the scoping review methodology, we did not conduct a quality appraisal of the included studies, as our goal was to rapidly identify and synthesize the existing evidence of digital precision health research for HTN management [16].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection.

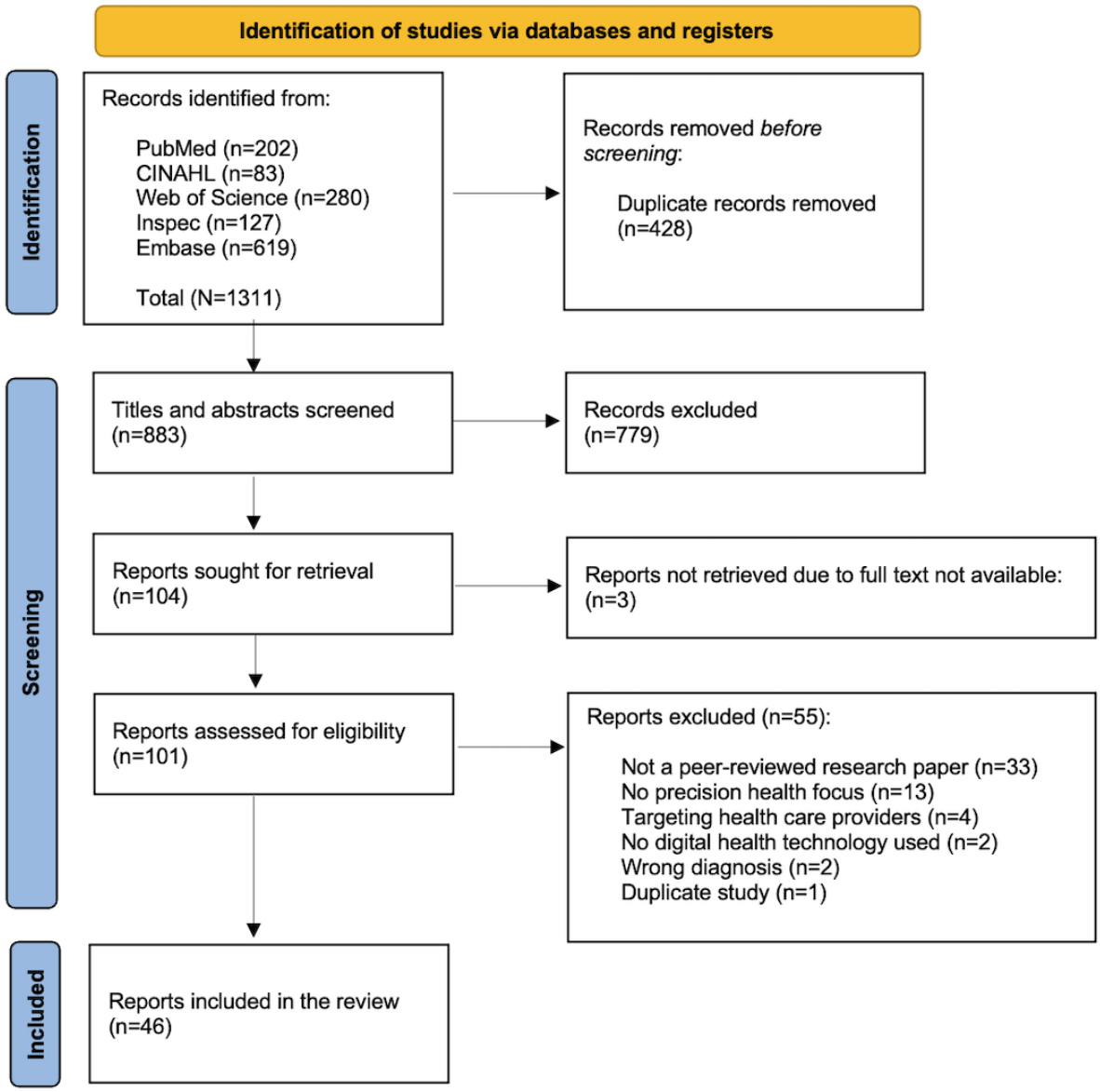

Summary of the Included Studies

Table 1 provides an overview of the included studies. Most studies were conducted in or used data from North America, predominantly from the United States (30/46, 65%), with 1 (2%) study conducted in Canada. Of the 46 studies, 7 (15%) were conducted in or used data from Europe, followed by 6 (13%) from Asia and 1 (2%) from Brazil in South America. Of the 46 studies, 1 (2%) study on genome sequencing used data from both the United States and the United Kingdom [19]. Nearly half of the studies (21/46, 46%) were RCTs, of which 8 (17%) were pilot or feasibility RCTs. There was a diverse

XSL•FO
RenderX

mix of studies comprising qualitative, mixed methods, and analytical observational (eg, cohort and cross-sectional) designs in addition to the RCTs.

In terms of precision health focus, most studies (34/46, 74%) focused on personalization, 8 (17%) focused on prediction models, and 4 (9%) focused on phenotyping (Figure 2). Disciplines involved were predominantly within the health sciences (29/46, 63%), with over a third of studies (15/46, 33%) featuring interdisciplinary teams comprising ≥2 disciplines (eg, medical and health sciences, informatics, and computer and electrical engineering). All included studies (46/46, 100%) sampled populations with HTN, with additional diagnoses of diabetes mellitus (16/46, 35%), hyperlipidemia (7/46, 15%), and classifications of overweight or obesity (3/46, 7%).

Table 1. Summary of the included studies (N=46).

Characteristics and categories	Studies, n (%)
Study location	
North America (the United States and Canada)	31 (67)
Europe (Spain, Ireland, the United Kingdom, and Sweden)	7 (15)
Asia (South Korea, Japan, Lebanon, Hong Kong, and China)	6 (13)
South America (Brazil)	1 (2)
Multiple locations (the United States and the United Kingdom)	1 (2)
Study design	
RCTs[a] (including pilot or feasibility)	21 (46)
Analytical observational	16 (35)
Qualitative	5 (11)
Mixed methods	3 (7)
Proof of concept	1 (2)
Precision health focus	
Personalization	34 (74)
Prediction models	8 (17)
Phenotyping	4 (9)
Disciplines involved	
Health sciences	29 (63)
Computer and electrical engineering	1 (2)
Informatics and communication sciences	1 (2)
Interdisciplinary team (≥2 disciplines)	15 (33)
Clinical conditions of the sample	
Hypertension	46 (100)
Diabetes mellitus	16 (35)
Hyperlipidemia	7 (15)
Overweight or obesity	3 (7)
Others (bipolar disorder, kidney transplant, and stroke)	3 (7)

[a]RCT: randomized controlled trial.

XSL•FO
RenderX

Figure 2. Studies published by year and precision health focus.

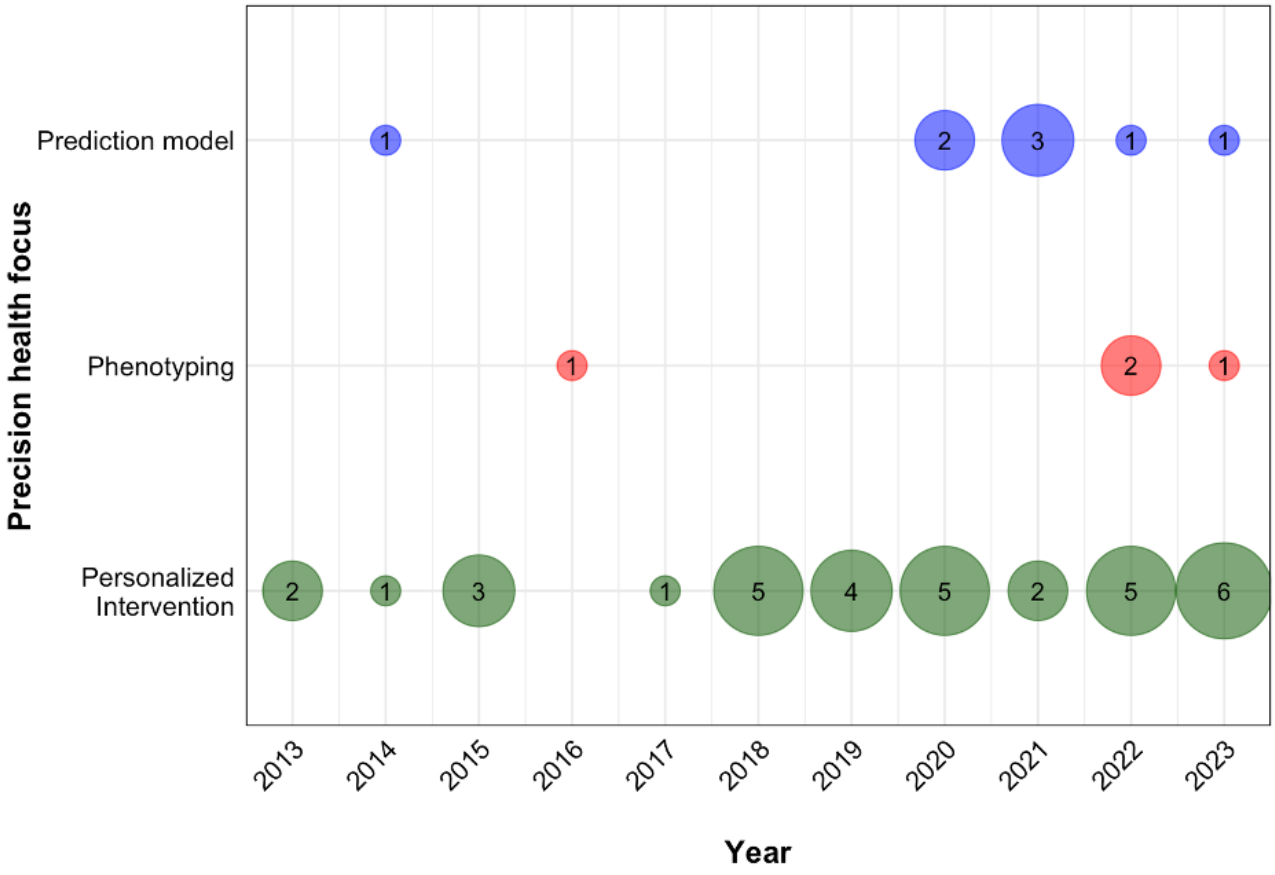

Study Characteristics

Study characteristics are presented in Multimedia Appendix 3, and a summary of the sample characteristics is presented in Table 2. Sample sizes in the studies ranged from 7 to 764,135. Nearly half (21/46, 46%) of the studies had a sample size <100, followed by 11 (24%) studies with a sample size between 100 and 499 and 6 (13%) with a sample size >10,000. Most studies (33/46, 72%) included participants with a mean age between 50 and 69 years. The range of female participants across all studies was 1.4% to 100%, with a median of 51% (IQR 40-62). Of the 46 studies, 14 (30%) did not report the race or ethnicity of participants.

XSL•FO
RenderX

Table 2. Summary of sample characteristics from the included studies (N=46).

Characteristic and category	Studies
Sample size, n (%)	
<99	21 (46)
100-499	11 (23.9)
500-999	3 (6.5)
1000-9999	5 (10.87)
>10,000	6 (13.04)
Mean age[a] (y), n (%)	
18-29	0 (0)
30-39	0 (0)
40-49	5 (11)
50-59	18 (39)
60-69	15 (33)
>70	1 (2)
Not reported	2 (4)
Race and ethnicity[b]	
African American or Black, n (median %[c], IQR)	21 (39.7, 11.8-53.8)
American Indian or Alaska Native, n (median %[c], IQR)	0 (0, 0)
Asian, n (median %[c], IQR)	7 (1.0, 0-2.8)
Hawaiian or Pacific Islander, n (median %[c], IQR)	0 (0, 0)
Hispanic or Latinx, n (median %[c], IQR)	17 (5.2, 2.1-14.3)
Non-Hispanic White, n (median %[c], IQR)	23 (50.9, 23.0-74.6)
Other, n (median %[c], IQR)	12 (2.8, 0-4.7)
Not reported, n	14
Partially reported, n	7
Gender[d]	
Female, n (median %, IQR)	39 (51, 40-62)
Not reported, n	2

[a]Mean age: Studies that cited another study or database for participant age (n=6) were excluded from the breakdown above.

[b]Race and ethnicity: Studies that cited another study or database for race and ethnicity (n=4) were excluded from the breakdown above.

[c]Median %: median percentage across the reported studies.

[d]Gender: Studies that cited another study or database for gender (n=5) were excluded from the breakdown above.

Digital Personalization for HTN Management

Overview

Of the 46 studies, 34 (74%) had a precision health focus of personalization in HTN management (Table 3). These studies tested or assessed digital interventions that were tailored to the individual or group characteristics of its users. Most studies (30/34, 88%) tested a personalized digital intervention and were primarily quantitative in design (eg, cohort, RCTs and mixed methods). Of the 34 personalization studies, 4 (13%) were qualitative and described participants' views on technology-based interventions. Sample sizes ranged from 11 to 10,803, and the most commonly reported primary outcomes included BP (21/34, 62%), medication adherence (4/34, 12%), participants' views or feedback (4/34, 12%), and feasibility and satisfaction of the intervention (3/34, 9%). Digital technologies used most often in these studies included mobile phones (30/34, 88%), followed by BP monitors (17/34, 50%), wearable devices (4/34, 12%), electronic medication trays or pillboxes (3/34, 9%), web platforms (2/34, 6%), and tablets (2/34, 6%).

XSL•FO
RenderX

Table 3. Summary of studies on digital personalization for hypertension (n=34).

First author	Study design	Sample size	Digital technology	Primary outcomes
Beran et al [20]	Mixed methods	450	Mobile phone and BP[a] monitor	BP
Blood et al [21]	Cohort	10,803	Mobile phone and BP monitor	BP
Bosworth et al [22]	RCT[b]	428	Mobile phone and BP monitor	CVD[c] risk score
Brewer et al [23]	Mixed methods	16	Mobile phone	BP
Chandler et al [24]	RCT	56	Mobile phone, BP monitor, and electronic medication trays and pill boxes	BP
Choudhry et al [25]	RCT	4078	Mobile phone	Medication adherence
David et al [26]	Secondary analysis of RCT	231	Mobile phone and BP monitor	BP
Davidson et al [27]	RCT	38	Mobile phone, BP monitor, electronic medication tray	BP
Glynn et al [28]	Qualitative	50	Views on technology	Views on technology
Guthrie et al [29]	Cohort	172	Mobile phone and BP monitor	BP
Hellem et al [30]	Qualitative	86	Mobile phone	Feedback on the design of the mobile app intervention
Jeong et al [31]	Pilot RCT	35	Mobile phone and BP monitor	Health behavior and BP
Kario et al [32]	RCT	390	Mobile phone and BP monitor	BP
Kassavou et al [33]	RCT	135	Mobile phone	Medication adherence
Klein et al [34]	Cohort	38	Mobile phone	Medication adherence
Leitner et al [35]	RCT	38	Wearable device, mobile phone, and BP monitor	BP
Lewinski et al [36]	Pilot RCT	118	Mobile phone	BP
Lv et al [37]	Pre-post	149	Mobile phone, BP monitor web-based system, and wearable device	BP
McBride et al [38]	Qualitative	11	Mobile phone and BP monitor	Feedback using app
McGillicuddy et al [39]	Pilot RCT	20	Mobile phone, BP monitor, electronic medication tray	BP and medication adherence
Naqvi et al [40]	Pilot RCT	50	Electronic tablet and BP monitor	BP
Payne Riches et al [41]	Pilot RCT	47	Mobile phone	Follow-up, fidelity, and app use
Petrella et al [42]	RCT	149	Mobile phone, BP monitor, and wearable device	BP
Rodriguez et al [43]	RCT	544	Mobile phone	BP
Rodriguez et al [44]	RCT	533	Mobile phone	Stage of change and DASH[d] score
Saleh et al [45]	Mixed methods	606	Mobile phone	Patient satisfaction
Schoenthaler et al [46]	Pilot RCT	42	Electronic tablet	Intervention acceptance and BP
Shea et al [47]	Pilot RCT	140	Mobile phone	BP
Steinberg et al [48]	Pilot RCT	140	Mobile phone	Diet tracking and DASH score
Thiboutot et al [49]	RCT	500	Website	BP
Van Emmenis et al [50]	Qualitative	20	Mobile phone	Views on the mobile app intervention
Wang et al [51]	Pilot RCT	49	Mobile phone and BP monitor	Feasibility of app use
Willis et al [52]	Cohort	7752	Mobile phone and BP monitor	BP
Zhang et al [53]	RCT	192	Wearable device and mobile phone	BP

[a]BP: blood pressure.

[b]RCT: randomized controlled trial.

[c]CVD: cardiovascular disease.

[d]DASH: dietary approaches to stop hypertension.

XSL•FO
RenderX

Of the 34 studies, 4 (12%) qualitative studies highlighted participants' perceived importance of personalization in digital interventions. Participants emphasized the need for digital interventions to be customizable, allowing for tailoring based on their preferences, such as personalizing functions for SMS text messages, the inclusion of all medications in the app, and the ability to adjust reminder settings [28,50]. A study on the design of a mobile app intervention found distinct differences in preferences between participants from a federally qualified health center and a university cardiovascular clinic; the former group placed greater value on social support themes and addressing health-related social needs [30]. McBride et al [38] found that participants using a mobile app for home BP monitoring felt the app's visual feedback and medication reminders improved their understanding of their condition as well as their sense of control and responsibility, resulting in improved self-management practices.

Personalization Characteristics of the Digital Interventions

Modalities and Approaches

For the 30 studies that tested a personalized digital intervention (Table 4), the modality of intervention delivery occurred exclusively through mobile apps for half of the studies (n=16, 53%), phone calls only (n=5, 17%), SMS text messages only (n=2, 7%), and web- or tablet-based only (n=2, 7%). Of the 30 studies, 6 (20%) used ≥2 modalities (ie, phone calls, texts, emails, and mobile apps) for intervention delivery. Of the 30 studies, 19 (63%) used personnel (eg, case managers, patient navigators, and clinicians) for some aspect of the intervention design and delivery, while 11 (37%) studies relied solely on digital technology for intervention delivery.

XSL•FO
RenderX

Table 4. Personalization characteristics of the digital interventions (n=30).

First author	Personnel	Human vs automated tailoring	Personalized elements of intervention	Study findings
Beran et al [20]	Pharmacists	Hybrid	• Tailored education • Individualized medication management	CS[a]
Blood et al [21]	Pharmacists, patient navigators, and NPs[b] and MDs[c]	Hybrid	• Tailored education • Individualized medication management	SS[d]
Bosworth et al [22]	Pharmacists	Human	• Tailored education • Individualized medication management	NS[e]
Brewer et al [23]	Community health workers and clinicians	Hybrid	• BP[f] and medication tracking on an app • Culturally tailored education • Assessment of SDoH[g] needs and referrals for services (housing and utilities)	CS
Chandler et al [24]	—[h]	Automated	• Reminders, feedback, and motivational messages based on treatment adherence, and values, beliefs, or goals	SS
Choudhry et al [25]	Pharmacists	Hybrid	• Tailored education • Assessment of SDoH needs and referrals for social work and affordable prescription options	SS
David et al [26]	—	Automated	• Tailored education • Personalized feedback based on BP, goals, and treatment adherence	SS
Davidson et al [27]	—	Automated	• Culturally tailored SMS text messages based on treatment adherence and values, beliefs, or goals	SS
Guthrie et al [29]	Multidisciplinary team	Hybrid	• AI[i]-assisted tailored feedback and education on self-management and treatment adherence	SS
Jeong et al [31]	Nurses	Human	• Tailored education addressing self-management behaviors and treatment adherence	CS
Kario et al [32]	Health professionals and chatbot based virtual nurses	Hybrid	• Personalized lifestyle program based on age, sex, lifestyle, social background, and behavior patterns	SS
Kassavou et al [33]	Clinicians	Automated	• Tailored messages on improving medication adherence based on patient beliefs, attitudes, self-efficacy, and emotional state • Medication refill reminders sent via SMS text messages	SS
Klein et al [34]	—	Automated	• Tailored SMS text messages to improve medication adherence comprising educational and motivational content	CS
Leitner et al [35]	—	Automated	• AI-driven personalized lifestyle recommendations based on clinical, behavior, and psychological data	CS
Lewinski et al [36]	Case manager	Human	• Tailored education on self-management behaviors and treatment adherence	NS
Lv et al [37]	Case manager, dietitian, and pharmacist	Human	• Individualized BP management plan • Tailored feedback on BP, medicines, weight, steps, and diet	SS
McGillicuddy et al [39]	—	Automated	• Tailored reminders, feedback, and summary reports including positive reinforcement and suggestions for improvement • Tailored to medication dosing schedule and BP goals	SS

XSL•FO
RenderX

First author	Personnel	Human vs automated tailoring	Personalized elements of intervention	Study findings
Naqvi et al [40]	NPs or MDs and pharmacists	Human	• Individualized consultations regarding symptoms and medication changes • BP infographics tailored to BP at discharge, available in English and Spanish	SS
Payne Riches et al [41]	Clinicians	Hybrid	• Individualized goal setting and feedback on participants' food choices to help them identify lower-salt options	CS
Petrella et al [42]	Exercise specialists	Human	• Tailored exercise program prescribed by an exercise specialist based on fitness level	NS
Rodriguez et al [43]	Counselors	Human	• Tailored counseling based on the stage of change for adherence to exercise, diet, and medication adherence • Individualized assessment of barriers to behavior changes and solutions	SS
Rodriguez et al [44]	Counselors	Human	• Tailored counseling based on the stage of change for DASH[j] diet, exercise, and medication adherence	SS
Saleh et al [45]	—	Automated	• Personalized reminders for appointments, laboratory tests, and examinations • Content tailored to language and health literacy	SS
Schoenthaler et al [46]	Research assistant	Automated	• Personalized list of adherence intervention strategies based on unique barriers to treatment adherence	SS
Shea et al [47]	—	Automated	• Tailored educational texts based on participants' priority topics and most recent BP reading	CS
Steinberg et al [48]	—	Automated	• Tailored messages based on adherence to the DASH diet • Content was tailored for women and provided behavioral tips to reinforce dietary change and provide social support	NS
Thiboutot et al [49]	—	Automated	• Customized feedback and recommendations on questions to discuss with their health care providers	NS
Wang et al [51]	—	Automated	• Individual monitoring of treatment response and tailored recommendations for follow-up based on BP levels	CS
Willis et al [52]	MDs	Human	• Individualized remote treatment based on BP levels	SS
Zhang et al [53]	Health facilitators	Human	• Tailored feedback on self-management behaviors	CS

[a]CS: clinically significant but not statistically significant.
[b]NP: nurse practitioner.
[c]MD: medical doctor.
[d]SS: statistically significant.
[e]NS: no statistical or clinical significance.
[f]BP: blood pressure.
[g]SDoH: social determinants of health.
[h]Not applicable.
[i]AI: artificial intelligence.
[j]DASH: dietary approaches to stop hypertension.

Personalization strategies varied, with 13 (43%) of the 30 studies using a fully automated personalization approach without the need for human involvement. For example, interventions used automated SMS text messages to improve treatment adherence and provide feedback to participants [24,26,34] or used an artificial intelligence–driven system to generate weekly tailored lifestyle recommendations [35]. Of the 30 studies, 10 (33%) used human-driven personalization, in which study personnel

XSL•FO
RenderX

were the primary avenues for personalization of the intervention. For example, nurses and case managers offered patients personalized education on self-management behaviors and treatment adherence [31,36], and pharmacists and physicians provided individualized medication adjustments based on treatment response [22,52]. Of the 30 studies, 7 (23%) used a hybrid approach (ie, both the technology and human study personnel contributed to the personalization of the intervention).

Most personalized intervention studies (25/30, 83%) had either statistically significant positive outcomes (16/30, 53%) or improved or clinically meaningful outcomes, although not statistically significant (9/30, 30%). The studies with nonsignificant, yet positive findings tended to be pilot studies or had small sample sizes (ie, <40 participants). Among the 25 studies with significant positive outcomes, 16 (64%) involved personnel in intervention delivery, with involvement levels varying from comprehensive multidisciplinary teams (pharmacists, nurse practitioners, physicians, and patient navigators) offering patient education, medication management, and monitoring [21], to minimal, such as research assistants giving participants tablet use instructions [46]. In addition, 48% (12/25) of these studies used an automated tailoring approach, followed by 28% (7/25) of the studies using human-led tailoring and 24% (6/25) using a combination of both human-led and automated tailoring.

Personalized Elements of the Interventions

Personalized interventions included tailored education on self-management (ie, medication adherence, diet, and exercise), along with motivational messages and feedback aligned with each participant's adherence, values, beliefs, and goals. For instance, a culturally adapted mobile health intervention sent the following message to a Hispanic man hoping to find a wife and start a family: "Get back on track taking your meds for building that stronger, healthier body for when you meet that special woman" [24]. Studies also provided personalized reminders for medications, appointments, laboratory tests, and individualized medication management based on treatment adherence and BP levels. Of the 30 studies, only 2 (7%) provided assessment and referrals for social determinants of health needs (ie, housing, utilities, and affordable prescription options) [23,25].

Data Used for Personalization

Tailored intervention studies used a range of data types for personalization of the intervention (Table 5). From 8 categories of data available for personalization (eg, demographic, clinical, and behavioral), just over half the studies (16/30, 53%) used ≥3 types of data, with 14 (47%) studies using ≤2 data types. Of the 30 studies, 6 (20%) used 5 categories of data for personalization, with no studies using >5 data types. The 3 most common types of data used for personalization were behavioral (28/30, 93%), clinical (24/30, 80%), and psychological (17/30, 57%). Behavioral data were typically used to monitor patients' adherence to key behaviors relevant to HTN management, such as taking medication, physical activity, and diet. Collecting real-time information on participants' health behaviors allowed for individualized feedback in many studies. For example, in the study by Steinberg et al [48], participants reported their food intake and could receive texts with feedback and guidance, such as "You did best with reducing saturated fat and boosting your fiber intake and seemed to struggle with getting enough potassium and magnesium...to get more magnesium, try dried fruit as a snack!" Similarly, clinical data such as BP or medication-related information were used to offer feedback tailored to participant's clinical metrics. For instance, Wang et al [51] developed a telehealth system that used decision rules to adjust care based on BP levels; if BP was optimal and participants were adherent to medications without side effects, follow-up appointments could be deferred and medication refills could be automatically prescribed. Psychological data included information on participants' attitudes, beliefs, values, and goals. For example, a few studies assessed participants' "stage of change" [43,44] and readiness to modify behaviors [25], while others took into account their values, attitudes, and beliefs [24,27,34] or their perceived barriers to self-management [25,36,43,46].

XSL•FO
RenderX

Table 5. Data types used for personalization in digital interventions (n=30).

Study	Demographic (n=4)	Socioeconomic (n=6)	Clinical (n=24)	Behavioral (n=28)	Psychologic (n=17)	Health literacy (n=6)	Physical environment	Cultural (n=4)
Beran et al [20]			✓	✓				
Blood et al [21]			✓					
Bosworth et al [22]	✓		✓	✓	✓	✓		
Brewer et al [23]		✓	✓	✓		✓		✓
Chandler et al [24]			✓	✓	✓			✓
Choudhry et al [25]		✓	✓	✓	✓			
David et al [26]			✓	✓				
Davidson et al [27]			✓	✓	✓	✓		✓
Guthrie et al [29]			✓	✓	✓			
Jeong et al [31]		✓	✓	✓	✓			
Kario et al [32]	✓	✓	✓	✓	✓			
Kassavou et al [33]				✓	✓			
Klein et al [34]				✓	✓			
Leitner et al [35]			✓	✓	✓			
Lewinski et al [36]		✓	✓	✓	✓	✓		
Lv et al [37]			✓	✓	✓			
McGillicuddy et al [39]			✓	✓	✓			
Naqvi et al [40]	✓		✓	✓				
Payne Riches et al [41]				✓	✓			
Petrella et al [42]			✓	✓				
Rodriguez et al [43]			✓	✓	✓			
Rodriguez et al [44]				✓	✓			
Saleh et al [45]			✓	✓				
Schoenthaler et al [46]		✓		✓	✓			✓
Shea et al [47]			✓	✓				
Steinberg et al [48]	✓			✓				
Thiboutot et al [49]			✓					
Wang et al [51]			✓	✓				
Willis et al [52]			✓	✓				
Zhang et al [53]			✓	✓				

The other types of data used for personalization were all used by less than one-third of the studies: health literacy (6/30, 20%), social and economic (6/30, 20%), demographics (4/30, 13%), and cultural (4/30, 13%). No studies obtained physical environment data for personalization of the intervention. For health literacy, the prevailing strategy involved designing interventions for a low-literacy population as a whole, rather than tailoring it at the individual participant level, and none of the studies tailored interventions to the level of digital literacy [36,45,54]. Similarly, studies implementing cultural tailoring primarily focused on specific populations as opposed to tailoring at the individual level. For example, Brewer et al [23] and Schoenthaler et al [46] both conducted studies with only Black participants and offered culturally tailored content geared for this population, such as education on the impact of racism and discrimination on BP. However, not all studies applied cultural tailoring broadly. Davidson et al [27] assessed cultural values and beliefs at an individual level, sending customized SMS text messages based on these insights.

Digital Phenotyping

There were 4 studies with a precision health focus on phenotyping [19,55-57] using a range of demographic, behavioral, clinical, and genetic data to apply phenotyping in diverse ways to enhance understanding and management of HTN (Table 6). Sample sizes ranged from 13 for a qualitative study [57] to 764,135 in a whole-genome sequencing analysis [19]. Digital tools used included mobile phones (2/4, 50%), web

XSL•FO
RenderX

platform (1/4, 25%), electronic health record (EHR; 1/4, 25%), machine learning (ML) algorithms (1/4, 25%), BP monitor (1/4, 25%), and genomic databases (1/4, 25%).

Table 6. Summary of the studies on digital phenotyping for HTN[a].

First author	Study design	Sample size	Digital technology	Description of the study
Bakre et al [55]	Secondary data analysis	11,934	Mobile phone and web platform	The study analyzed demographic, dietary, and clinical data of participants with stage 2 HTN who used a digital nutrition platform to identify characteristics associated with greater reductions in BP[b].
Chen et al [56]	Secondary data analysis	2521	EHR[c] and ML[d] algorithm	The study developed an ML framework to identify predictive features and cluster patients with HTN into 4 clinically meaningful groups of disease severity.
Hellem et al [57]	Qualitative	13	Mobile phone and BP monitor	The study examined participants' determinants of engagement from a digital intervention through interviews, phenotyping participants into high engagers, low engagers, and early enders.
Kelly et al [19]	Genome sequencing	764,135	Genomic databases	This is the first study to identify a promising but unconfirmed intergenic rare variant associated with BP. This variant lowered SBP[e] by an average of 33 mm Hg in carriers compared with noncarriers, and all participants were of Asian ancestry.

[a]HTN: hypertension.

[b]BP: blood pressure.

[c]EHR: electronic health record.

[d]ML: machine learning.

[e]SBP: systolic blood pressure.

Of the 4 studies, 2 (50%) used secondary data analysis to apply phenotyping. Bakre et al [55] applied phenotyping by analyzing demographic, dietary, and clinical data of participants with stage 2 HTN who used a digital nutrition platform, identifying characteristics associated with a decrease in BP. They found that participants who achieved greater BP reductions had higher reductions in weight and dietary improvements, suggesting the effectiveness of digital nutrition guidance on management of HTN [55]. Chen et al [56] used EHR data to develop an ML framework to phenotype patients with HTN into 4 clinically meaningful groups, aiming to improve personalization of care and resource allocation. They used demographic and clinical data for cluster analysis, identifying 2 groups with mild disease and 2 with more severe disease profiles.

The qualitative study by Hellem et al [57] explored participants' determinants of engagement from a SMS text messaging intervention to reduce BP among patients with HTN. Through interviews with participants purposely sampled from 3 engagement categories—high engagers, low engagers, and early enders—this study phenotyped individuals based on a mix of social, psychological, and environmental factors, including digital literacy and access. They found that high engagers had a better understanding of the intervention, the least number of social needs, and the greatest social support. In contrast, early enders and low engagers expressed greater amounts of social needs and less social support [57]. Finally, the study by Kelly et al [19] phenotyped participants at a genetic level, offering a direct link between genotype and phenotypic expression of BP, presenting a road map for targeted genetic interventions for personalized management of HTN.

Digital Prediction Tools

There were 8 studies with a precision health focus on prediction models (Table 7). These studies collectively explored advanced ML techniques for personalized risk prediction for HTN management. Study designs were primarily ML model development (7/8, 88%), with 1 (13%) proof-of-concept study [58] and 1 (13%) hybrid design comprising both ML model development and RCT [59]. Sample sizes varied from 7 for the proof-of-concept study [58] to 245,499 for the study using EHR data [60]. Digital technologies used for these studies included the following: ML algorithms (8/8, 100%), EHR (5/8, 63%), wearable devices (4/8, 50%), mobile phone and web platform (1/8, 13%), and BP monitor (1/8, 13%). Various ML models were used for each of these studies, with half of the studies using random forest models (4/8, 50%), with others including online recurrent extreme learning machine and long short-term memory.

XSL•FO
RenderX

Table 7. Summary of the studies on prediction models for HTN[a] (n=8).

First author	Study design	Sample size	Digital technology	Study description	ML[b] models
Abrar et al [61]	ML model development	35	ML algorithms	Development of a personalized BP[c] prediction model tailored to individual physiology and lifestyle factors	OR-ELM[d]
Bernal et al [58]	Proof of concept	7	ML algorithms, wearable device, mobile phone, and web platform	Development of an intelligent web-based ecosystem that integrates clinical, behavioral, and environmental data to predict adverse high BP events	Bayesian ridge, support vector regression, and random forest
Bertsimas et al [62]	ML model development	19,926	ML algorithms and EHR[e]	Development of ensemble ML models for personalized predictions and treatment recommendations, with significant improvements in HTN through optimal treatment prescriptions based on individual patient data	Classification models (eg, multivariate logistic regression and random forests) and regression models (eg, support vector regression and optimal regression trees)
Cano et al [63]	ML model development	86	ML algorithms, wearable device, and EHR	Development of a system for improved discrimination between healthy individuals, individuals with prehypertension, and individuals with HTN using PPG[f] and ECG[g] signals, focusing on improving the accuracy of HTN detection	A total of 37 different classification models were used: logistic regression, support vector machines, and nearest neighbors, with the Coarse Tree model achieving the highest discrimination
Chiang et al [59]	RCT[h] and ML model development	25	ML algorithms, wearable device, and BP monitor	Development of a personalized BP model for each individual using clinical and behavioral data, identifying the most important lifestyle factors impacting BP trend, and providing precise recommendations for HTN management	Random forest with Shapley value–based feature selection model outperformed other models in prediction accuracy
Hu et al [64]	ML model development	42,792	ML algorithms, wearable device, and EHR	Development of an ML model for personalized antihypertensive medication classes for patients with HTN using a robust algorithm that accommodates outliers in EHR data	The distributionally linear regression–informed k-nearest neighbors model resulted in a 14.22 mm Hg reduction in SBP[i] on average
Jimeng et al [65]	ML model development	1294	ML algorithms and EHR	Development of ML model to predict transition points in HTN control status, aiming to inform personalized management strategies	Logistic regression, naïve Bayes, and random forests
Ye et al [60]	ML model development	245,499	ML algorithms and EHR	Development of deep learning models to predict individualized HTN treatment pathways based on EHR data	LSTM[j] and bidirectional LSTM

[a]HTN: hypertension.

[b]ML: machine learning.

[c]BP: blood pressure.

[d]OR-ELM: online recurrent extreme learning machine.

[e]EHR: electronic health record.

[f]PPG: photoplethysmography.

[g]ECG: electrocardiogram.

[h]RCT: randomized controlled trial.

[i]SBP: systolic blood pressure.

[j]LTSM: long short-term memory.

Emphasizing personalized health care, these studies demonstrate improved accuracy in HTN control by tailoring recommendations to individual patient profiles, challenging traditional one-size-fits-all approaches. One key difference lies in the scale and nature of the datasets used, with some researchers analyzing extensive patient data over extended periods [62], whereas others focus on smaller, more controlled patient groups [59]. In addition, the scope of each study varies; some studies aim to improve BP prediction accuracy, while others focus on optimizing treatment pathways [64], integrating lifestyle data [59], or providing real-time [58] and personalized health care recommendations [59].

Discussion

Principal Findings

This scoping review provided a comprehensive mapping of the literature on the trends and gaps in digital precision health research for the management of HTN in adults. Analysis of 46 studies identified 3 main categories of precision health focus:

XSL•FO
RenderX

personalization, phenotyping, and prediction. The predominance of personalization (34/46, 74% of studies), compared to prediction (8/46, 17%) and phenotyping (4/46, 9%), highlights a strong emphasis on tailored interventions but reveals limited attention to predictive modeling and comprehensive phenotyping, which are essential for predicting HTN outcomes and identifying and categorizing patient subgroups.

Gaps in Digital Precision Health Research

The review revealed geographic, disciplinary, and demographic gaps in digital precision health research for HTN. Most studies (30/46, 65%) reviewed were conducted in the United States, followed by Europe (7/46, 15%), Asia (6/46, 13%), and South America (1/46, 2%), with no studies from Africa or Oceania. This contrasts with high HTN prevalence rates in Africa and Oceania, underscoring the need for more region-specific research to better understand local challenges and develop culturally sensitive and effective strategies for HTN prevention and management [5].

Disciplinary gaps were evident, with >60% (29/46) of the studies originating solely from the medical and health sciences. Less than a third of the studies (15/46) involved interdisciplinary teams, highlighting the need for collaborative efforts across disciplines such as engineering, informatics, and social sciences. Transdisciplinary research, in which team members from diverse fields develop unified conceptual frameworks that transcend their respective disciplinary views, is essential for advancing precision health research in HTN [66].

Demographic reporting was inconsistent and insufficient across studies. While 96% (44/46) of the studies reported participants' gender and age, 45% (21/46) did not report race and ethnicity, or only did so partially. Of those that reported race and ethnicity, the median percentages of participant representation were skewed heavily toward non-Hispanic White (51%), followed by African American (40%), with underrepresentation of Hispanic or Latinx (5%) and Asian (1%) groups. This lack of diversity raises concerns about generalizability of findings, concern for bias, and underscores the need to for transparency in reporting demographic data, which is especially salient in the context of precision health [67].

Personalized Digital Interventions: Tools, Modalities, and Data Types

Mobile phones were the predominant tools for digital HTN interventions (30/34, 88%), while only 50% (17/34) incorporated home BP monitors and 12% (4/34) used wearable devices. This highlights an opportunity to integrate wearable sensor devices for richer data collection (eg, sleep, heart rate, and physical activity), fostering a more holistic understanding of HTN and its management. Wearable BP monitoring devices is a growing area of research and offers noninvasive, continuous BP monitoring through methods such as photoplethysmography [68]. These devices offer a more comprehensive view of cardiovascular health over time, enabling continuous monitoring, personalized treatment plans, and enhanced predictive capabilities for early intervention. However, further research is needed to enhance their accuracy, calibration, and validation for clinical use [68].

Traditional communication methods such as phone calls and texting were still used (5/30, 17% and 2/30, 7% of studies, respectively), demonstrating their relevance especially for reaching underserved populations with limited digital literacy or access to advanced technology. However, only 20% (6/30) of the studies used multiple intervention delivery modalities (ie, phone calls, texting, and apps), suggesting the need for more integrative approaches to improve access, engagement, and efficacy of digital interventions by catering to different user preferences and needs.

Most studies used behavioral (28/30, 93%), clinical (24/30, 80%), and psychological (17/30, 57%) data for personalization, but only 7% (2/30) of the studies incorporated social determinants such as housing needs or prescription affordability [23,25]. No studies included physical environment or digital literacy data, highlighting key gaps in personalized digital interventions for HTN. Prior studies have documented poor digital literacy and lack of support and training as significant barriers to the use of digital health tools by populations considered disadvantaged [69]. This underscores the need to incorporate diverse sources of data to provide the additional context that can help address disparities in digital intervention engagement and efficacy. Health equity–focused frameworks, such as the Digital Health Equity framework [70], can be used to guide the development and implementation of digital interventions to ensure holistic data collection and personalization, mitigate intervention-generated inequalities, and better address existing health disparities in HTN.

Defining Personalization Strategies

This review identified a critical need for greater clarity and specificity in the use of terminology such as "personalization" or "tailoring" to improve the understanding and application of tailored interventions. As personalization is so broadly defined, it remains challenging to assess the most effective characteristics of personalization for digital interventions. For example, Steinberg et al [48] wrote that "only 28% of participants said they felt the texts were personalized, despite the use of an algorithm designed to personalize messages about intake of specific DASH nutrients." Such findings indicate the importance of research that specifically delineates which strategies and approaches to personalization are most impactful for users; specifically, there is a need for tools to characterize these approaches to personalization. Fan and Poole [71] provide a useful classification scheme for how personalization can be implemented. Two of the dimensions they describe, in particular, are highly relevant to our corpus: the target of the intervention (either individual or categorical) and the automation approach (either explicit or configured by the user or implicit or automated by the system).

Studies in this review used both individual and categorical targets. For instance, in the study by Kario et al [32], personalization was implemented broadly to group characteristics through tailored BP infographics developed for the target population. This is in contrast to personalization at the individual level, such as in the study by Brewer et al [23], in which participants were assessed for individual social needs and referred for needed services (ie, housing and utilities). While

XSL•FO
RenderX

categorical-level personalization strategies can be effective and impactful, we caution against their widespread use in personalization discussions, as precision health primarily focuses on individual-level differences within a patient population. Further research is needed to better understand the characteristics and levels of personalization that are most effective for diverse populations.

In addition, the distinction of whether personalization is automated or human driven is an important lens in understanding an intervention and its effects, and studies in this review leveraged both these approaches as well as hybrid versions of them. While automated models can streamline interventions, they may lack the adaptability of human-led approaches.

Hybrid methods, combining automated and human input, could address the limitations of each approach alone. For example, Schoenthaler et al [46] mentioned the following in their paper: "despite high acceptability, one-third of intervention participants recommended including a health educator as an adjunct to the mHealth intervention, suggesting that some in-person contact is important and could not be replaced by the design of this intervention." Several papers in our review demonstrate hybrid approaches to personalization, which could guide future research. For example, the intervention developed by Choudhry et al [25] involved a pharmacist-led, semistructured consultation with specific guidelines for mapping barriers to adherence and developing corresponding response strategies. This human-driven, semistructured consultation is paired with automated, system-driven intervention elements, such as SMS text messages. Such hybrid approaches to the design of digital health interventions present a promising avenue for future research.

Digital Phenotyping

The 4 studies focusing on phenotyping within digital precision HTN management illuminate the potential for leveraging a wide spectrum of data—demographic, clinical, behavioral, and genetic—to identify unique phenotypes of participants with HTN to tailor treatment strategies. Multiple study designs, including qualitative interviews and whole-genome analysis, have been used to enhance understanding of HTN. Leveraging secondary data from EHRs and past RCTs for analysis using ML algorithms can help cluster phenotypes of patients with HTN and explain variations in treatment response [72]. Classifying individuals into meaningful groups based on HTN phenotypes is a crucial step toward personalized care and advancing precision health.

By contrast, qualitative approaches, despite having smaller sample sizes, have the ability to provide a more contextual, holistic, and nuanced understanding of the social and psychological determinants of HTN treatment adherence and outcomes from digital interventions, which is crucial for understanding disparities. Future areas of research should integrate the use of multiple sources of data, such as surveys, wearable devices, remote BP monitoring, and ecological momentary assessments, to derive digital phenotypes of participants' self-management behaviors and provide more effective interventions [73].

Prediction Models

The 8 studies with a precision health focus on prediction models for digital HTN management highlight significant advancements and challenges in the field. These studies collectively demonstrate the innovative use of advanced ML techniques for personalized risk prediction and intervention strategies. Using a wide array of digital tools, including EHR data, wearable devices, mobile platforms, and remote monitoring, these studies are shifting the paradigm beyond one-size-fits-all approaches to care to offer more personalized and preventive care. ML models can help identify factors to predict individuals' risk of HTN, potentially preventing the development of high BP through early intervention. There are a growing number of studies using diverse ML methods (eg, support vector machine, deep learning, and XGBoost) and data types (eg, genetic, behavioral, and sociodemographic) to accurately predict the development of HTN [74-76].

Despite the potential of prediction models to transform HTN care, these studies also illuminate the hurdles that must be overcome to realize this potential fully. For instance, the generalizability of findings is a significant concern, with varying sample sizes and potential biases from using specific datasets that may not be representative of the population at large. A recent analysis of 63 HTN research studies using ML methods found that only 46% of studies described the participant demographics, and none of the studies provided a rigorous assessment of algorithmic bias, with only 6 studies acknowledging a risk of bias [77]. Algorithmic bias occurs when the diversity of the input dataset used for model development does not match that of the target population, resulting in inaccurate predictions for underrepresented groups and disparities in outcomes [77]. The effectiveness of ML models to accurately predict risk and personalize care relies heavily on the quality of data on which they are trained. The lack of representativeness of participants in training data for these models poses significant concerns regarding their generalizability and potential for bias [67]. To address these limitations, reporting guidelines such as CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) or TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence) should be used when reporting ML prediction models for HTN; these guidelines provide specific criteria for authors to follow, ensuring transparency, completeness, and standardization in study reporting [78,79]. These guidelines emphasize detailed reporting of model development, training, validation, performance metrics, data sources, and patient characteristics which enable comparisons across studies. Furthermore, the prediction models reviewed focused primarily on development, underscoring a key research gap in model deployment and real-world implementation. Advancing precision health in HTN management requires research on best practices for deploying, evaluating, and monitoring models in clinical care.

Strengths and Limitations

This scoping review on digital precision HTN management followed a systematic approach based on the framework

XSL•FO
RenderX

developed by Arksey and O'Malley [17], ensuring a rigorous and comprehensive exploration of the field. By adopting a precision health lens, the review covers a broad spectrum of studies, including personalized interventions, prediction models, and phenotyping, offering valuable insights into these areas. The detailed analysis of personalized intervention studies, particularly regarding the characteristics of personalization and data used for personalization, adds depth to the understanding of how digital tools can optimize HTN care. In addition, the interdisciplinary composition of the review team, drawing expertise from nursing, informatics, and computer science, further strengthens this review by incorporating diverse perspectives.

However, the broad scope of this review and the heterogeneity of included studies introduce certain limitations. For example, a detailed examination of intervention effectiveness and statistical significance was limited due to diverse study designs, such as underpowered pilot studies or qualitative designs. The absence of a quality assessment for the included studies may also compromise the validity of the conclusions drawn, as it overlooks methodological rigor and potential biases within studies. Finally, our search terms included "precision health" and only the "personalization" derivatives (eg, *precision medicine* OR *personalized* OR *tailored* OR *individualized*), which could have limited the number of studies on prediction models or phenotyping. However, the review team chose to include the studies on prediction models and phenotyping as these were considered to be important aspects of precision health research. Future reviews could emphasize a more comprehensive search of these specific aspects of precision health research for digital HTN management.

Conclusions

This scoping review of 46 studies synthesized the current state of precision health research encompassing personalization, phenotyping, and prediction using digital health tools for the management of HTN in adults. The findings from our review demonstrate that the majority (34/46, 74%) of the included studies had a precision health focus of personalization in digital HTN management, followed by prediction models and phenotyping. Our analysis highlighted significant gaps in reporting of participant race and ethnicity data, geographic distribution of research in digital HTN management, and use of wearable devices for capturing passive data. Furthermore, the analysis of personalization characteristics in interventions underscores the need for the integration of multiple sources of data for personalization, such as social determinants of health, physical environment, and digital literacy, to promote health equity. There is a need for more precision in the use of the term "personalization" as well as further research that explores the impact of specific types of personalization on health outcomes. Finally, greater interdisciplinary collaboration and ultimately a transdisciplinary approach are needed to meaningfully advance the field of precision health for HTN risk prediction, prevention, and management.

Acknowledgments

The authors extend their sincere gratitude to Imelda Vetter, health sciences librarian, for her expert guidance in systematic database searches to identify eligible studies for this scoping review. Research reported in this publication was completed with support from the National Institute of Nursing Research of the National Institutes of Health (Award Number T32NR019035, Precision Health Intervention Methodology Training in Self-Management of Multiple Chronic Conditions). The funding agency played no role in the design, data collection, analysis, or writing of the manuscript.

Conflicts of Interest

Authors NC and AA were supported by the National Institute of Nursing Research of the National Institutes of Health (Award Number T32NR019035, Precision Health Intervention Methodology Training in Self-Management of Multiple Chronic Conditions). The funding agency played no role in the design, data collection, analysis, or writing of the manuscript.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 104 KB - jmir_v27i1e59841_app1.pdf]

Multimedia Appendix 2
Database search terms.
[DOCX File , 17 KB - jmir_v27i1e59841_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies.
[DOCX File , 50 KB - jmir_v27i1e59841_app3.docx]

References

1. Hypertension. World Health Organization. 2023 Mar 16. URL: https://www.who.int/news-room/fact-sheets/detail/hypertension [accessed 2024-04-29]

XSL•FO
RenderX

2. Lawes CM, Vander Hoorn S, Rodgers A, International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. Lancet 2008 May 03;371(9623):1513-1518. [doi: 10.1016/S0140-6736(08)60655-8] [Medline: 18456100]
3. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1923-1994 [FREE Full text] [doi: 10.1016/S0140-6736(18)32225-6] [Medline: 30496105]
4. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980 [FREE Full text] [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
5. Zhou B, Perel P, Mensah GA, Ezzati M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. Nat Rev Cardiol 2021 Nov 28;18(11):785-802 [FREE Full text] [doi: 10.1038/s41569-021-00559-8] [Medline: 34050340]
6. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. Circulation 2016 Aug 09;134(6):441-450 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.018912] [Medline: 27502908]
7. Schutte AE, Jafar TH, Poulter NR, Damasceno A, Khan NA, Nilsson PM, et al. Addressing global disparities in blood pressure control: perspectives of the International Society of Hypertension. Cardiovasc Res 2023 Mar 31;119(2):381-409 [FREE Full text] [doi: 10.1093/cvr/cvac130] [Medline: 36219457]
8. Vrijens B, Vincze G, Kristanto P, Urquhart J, Burnier M. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. BMJ 2008 May 17;336(7653):1114-1117 [FREE Full text] [doi: 10.1136/bmj.39553.670231.25] [Medline: 18480115]
9. Hamrahian SM, Maarouf OH, Fülöp T. A critical review of medication adherence in hypertension: barriers and facilitators clinicians should consider. Patient Prefer Adherence 2022;16:2749-2757 [FREE Full text] [doi: 10.2147/PPA.S368784] [Medline: 36237983]
10. Dzau VJ, Hodgkinson CP. Precision hypertension. Hypertension 2024 Apr;81(4):702-708. [doi: 10.1161/hypertensionaha.123.21710]
11. Minor L, Rees M. Discovering Precision Health: Predict, Prevent, and Cure to Advance Health and Well-Being. Hoboken, NJ: Wiley Blackwell; 2020.
12. Gambhir SS, Ge TJ, Vermesh O, Spitler R. Toward achieving precision health. Sci Transl Med 2018 Feb 28;10(430):52 [FREE Full text] [doi: 10.1126/scitranslmed.aao3612] [Medline: 29491186]
13. Olstad DL, McIntyre L. Reconceptualising precision public health. BMJ Open 2019 Sep 13;9(9):e030279 [FREE Full text] [doi: 10.1136/bmjopen-2019-030279] [Medline: 31519678]
14. Fu MR, Kurnat-Thoma E, Starkweather A, Henderson WA, Cashion AK, Williams JK, et al. Precision health: a nursing perspective. Int J Nurs Sci 2020 Jan 10;7(1):5-12 [FREE Full text] [doi: 10.1016/j.ijnss.2019.12.008] [Medline: 32099853]
15. Hickey KT, Bakken S, Byrne MW, Bailey DC, Demiris G, Docherty SL, et al. Precision health: advancing symptom and self-management science. Nurs Outlook 2019;67(4):462-475 [FREE Full text] [doi: 10.1016/j.outlook.2019.01.003] [Medline: 30795850]
16. Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision health: the role of the social and behavioral sciences in advancing the vision. Ann Behav Med 2020 Nov 01;54(11):805-826 [FREE Full text] [doi: 10.1093/abm/kaaa018] [Medline: 32338719]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Kelly TN, Sun X, He KY, Brown MR, Taliun SA, Hellwege JN, Samoan Obesity, Lifestyle,Genetic Adaptations Study (OLaGA) Group,‡ NHLBI Trans-Omics for Precision Medicine TOPMed) Consortium. Insights from a large-scale whole-genome sequencing study of systolic blood pressure, diastolic blood pressure, and hypertension. Hypertension 2022 Aug;79(8):1656-1667 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.122.19324] [Medline: 35652341]
20. Beran M, Asche SE, Bergdall AR, Crabtree B, Green BB, Groen SE, et al. Key components of success in a randomized trial of blood pressure telemonitoring with medication therapy management pharmacists. J Am Pharm Assoc (2003) 2018;58(6):614-621. [doi: 10.1016/j.japh.2018.07.001] [Medline: 30077564]
21. Blood AJ, Cannon CP, Gordon WJ, Mailly C, MacLean T, Subramaniam S, et al. Results of a remotely delivered hypertension and lipid program in more than 10 000 patients across a diverse health care network. JAMA Cardiol 2023 Jan 01;8(1):12-21 [FREE Full text] [doi: 10.1001/jamacardio.2022.4018] [Medline: 36350612]
22. Bosworth HB, Olsen MK, McCant F, Stechuchak KM, Danus S, Crowley MJ, et al. Telemedicine cardiovascular risk reduction in veterans: the CITIES trial. Am Heart J 2018 May;199:122-129. [doi: 10.1016/j.ahj.2018.02.002] [Medline: 29754649]
23. Brewer LC, Jones C, Slusser JP, Pasha M, Lalika M, Chacon M, et al. mHealth intervention for promoting hypertension self-management among African American patients receiving care at a community health center: formative evaluation of

XSL•FO
RenderX

the FAITH! hypertension app. JMIR Form Res 2023 Jun 16;7:e45061 [FREE Full text] [doi: 10.2196/45061] [Medline: 37115658]
24. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mHealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019 Apr 06;16(7):1226 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
25. Choudhry NK, Isaac T, Lauffenburger JC, Gopalakrishnan C, Lee M, Vachon A, et al. Effect of a remotely delivered tailored multicomponent approach to enhance medication taking for patients with hyperlipidemia, hypertension, and diabetes: the STIC2IT cluster randomized clinical trial. JAMA Intern Med 2018 Sep 01;178(9):1182-1189 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3189] [Medline: 30083727]
26. David CN, Iochpe C, Harzheim E, Sesin GP, Gonçalves MR, Moreira LB, et al. Effect of mobile health interventions on lifestyle and anthropometric characteristics of uncontrolled hypertensive participants: secondary analyses of a randomized controlled trial. Healthcare (Basel) 2023 Apr 08;11(8):1069 [FREE Full text] [doi: 10.3390/healthcare11081069] [Medline: 37107903]
27. Davidson T, Favella A, McGillicuddy J, Mueller M, Brunner-Jackson B, Anderson A, et al. Evaluation of an mHealth medication regimen self-management program for African American and Hispanic uncontrolled hypertensives. J Pers Med 2015 Nov 17;5(4):389-405 [FREE Full text] [doi: 10.3390/jpm5040389] [Medline: 26593951]
28. Glynn L, Casey M, Walsh J, Hayes PS, Harte RP, Heaney D. Patients' views and experiences of technology based self-management tools for the treatment of hypertension in the community: a qualitative study. BMC Fam Pract 2015 Sep 09;16:119 [FREE Full text] [doi: 10.1186/s12875-015-0333-7] [Medline: 26354752]
29. Guthrie NL, Berman MA, Edwards KL, Appelbaum KJ, Dey S, Carpenter J, et al. Achieving rapid blood pressure control with digital therapeutics: retrospective cohort and machine learning study. JMIR Cardio 2019 Mar 12;3(1):e13030 [FREE Full text] [doi: 10.2196/13030] [Medline: 31758792]
30. Hellem AK, Casetti A, Bowie K, Golbus JR, Merid B, Nallamothu BK, et al. A community participatory approach to creating contextually tailored mHealth notifications: myBPmyLife project. Health Promot Pract 2024 May 27;25(3):417-427. [doi: 10.1177/15248399221141687] [Medline: 36704967]
31. Jeong S, Choi H, Gwon SH, Kim J. Telephone support and telemonitoring for low-income older adults. Res Gerontol Nurs 2018 Jul 01;11(4):198-206. [doi: 10.3928/19404921-20180502-01] [Medline: 29767806]
32. Kario K, Nomura A, Harada N, Okura A, Nakagawa K, Tanigawa T, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021 Oct 21;42(40):4111-4122 [FREE Full text] [doi: 10.1093/eurheartj/ehab559] [Medline: 34455443]
33. Kassavou A, Mirzaei V, Brimicombe J, Edwards S, Massou E, Prevost AT, et al. A highly tailored text and voice messaging intervention to improve medication adherence in patients with either or both hypertension and type 2 diabetes in a UK primary care setting: feasibility randomized controlled trial of clinical effectiveness. J Med Internet Res 2020 May 19;22(5):e16629 [FREE Full text] [doi: 10.2196/16629] [Medline: 32427113]
34. Klein P, Aebi ME, Sajatovic M, Depp C, Moore D, Blixen C, et al. Differential medication attitudes to antihypertensive and mood stabilizing agents in response to an automated text-messaging adherence enhancement intervention. J Behav Cogn Ther 2020 Apr;30(1):57-64 [FREE Full text] [doi: 10.1016/j.jbct.2020.03.015] [Medline: 33409504]
35. Leitner J, Chiang PH, Khan B, Dey S. An mHealth lifestyle intervention service for improving blood pressure using machine learning and IoMTs. In: Proceedings of the 2022 IEEE International Conference on Digital Health. 2022 Presented at: ICDH '22; July 10-16, 2022; Barcelona, Spain p. 142-150 URL: https://ieeexplore.ieee.org/document/9861082 [doi: 10.1109/icdh55609.2022.00030]
36. Lewinski AA, Patel UD, Diamantidis CJ, Oakes M, Baloch K, Crowley MJ, et al. Addressing diabetes and poorly controlled hypertension: pragmatic mHealth self-management intervention. J Med Internet Res 2019 Apr 09;21(4):e12541 [FREE Full text] [doi: 10.2196/12541] [Medline: 30964439]
37. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized hypertension management using patient-generated health data integrated with electronic health records (EMPOWER-H): six-month pre-post study. J Med Internet Res 2017 Sep 19;19(9):e311 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
38. McBride CM, Morrissey EC, Molloy GJ. Patients' experiences of using smartphone apps to support self-management and improve medication adherence in hypertension: qualitative study. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e17470 [FREE Full text] [doi: 10.2196/17470] [Medline: 33112251]
39. McGillicuddy JW, Gregoski MJ, Weiland AK, Rock RA, Brunner-Jackson BM, Patel SK, et al. Mobile health medication adherence and blood pressure control in renal transplant recipients: a proof-of-concept randomized controlled trial. JMIR Res Protoc 2013 Sep 04;2(2):e32 [FREE Full text] [doi: 10.2196/resprot.2633] [Medline: 24004517]
40. Naqvi IA, Strobino K, Kuen Cheung Y, Li H, Schmitt K, Ferrara S, et al. Telehealth after stroke care pilot randomized trial of home blood pressure telemonitoring in an underserved setting. Stroke 2022 Dec;53(12):3538-3547. [doi: 10.1161/strokeaha.122.041020]
41. Payne Riches S, Piernas C, Aveyard P, Sheppard JP, Rayner M, Albury C, et al. A mobile health salt reduction intervention for people with hypertension: results of a feasibility randomized controlled trial. JMIR Mhealth Uhealth 2021 Oct 21;9(10):e26233 [FREE Full text] [doi: 10.2196/26233] [Medline: 34673535]

XSL•FO
RenderX

42. Petrella RJ, Stuckey MI, Shapiro S, Gill DP. Mobile health, exercise and metabolic risk: a randomized controlled trial. BMC Public Health 2014 Oct 18;14:1082 [FREE Full text] [doi: 10.1186/1471-2458-14-1082] [Medline: 25326074]
43. Rodriguez MA, Friedberg JP, Wang B, Fang Y, Natarajan S. Sustain ability of a tailored behavioral intervention to improve hypertension control: outcomes of a randomized controlled trial. J Gen Intern Med 2015(7):S261-S262 [FREE Full text]
44. Rodriguez MA, Friedberg JP, DiGiovanni A, Wang B, Wylie-Rosett J, Hyoung S, et al. A tailored behavioral intervention to promote adherence to the DASH diet. Am J Health Behav 2019 Jul 01;43(4):659-670. [doi: 10.5993/AJHB.43.4.1] [Medline: 31239010]
45. Saleh S, Farah A, El Arnaout N, Dimassi H, El Morr C, Muntaner C, et al. mHealth use for non-communicable diseases care in primary health: patients' perspective from rural settings and refugee camps. J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii52-ii63 [FREE Full text] [doi: 10.1093/pubmed/fdy172] [Medline: 30307516]
46. Schoenthaler A, Leon M, Butler M, Steinhaeuser K, Wardzinski W. Development and evaluation of a tailored mobile health intervention to improve medication adherence in black patients with uncontrolled hypertension and type 2 diabetes: pilot randomized feasibility trial. JMIR Mhealth Uhealth 2020 Sep 23;8(9):e17135 [FREE Full text] [doi: 10.2196/17135] [Medline: 32965230]
47. Shea S, Thompson JL, Schwartz JE, Chen Y, de Ferrante M, Vanderbeek AM, et al. The Retail Outlet Health Kiosk Hypertension Trial (ROKHYT): pilot results. Am J Hypertens 2022 Jan 05;35(1):103-110 [FREE Full text] [doi: 10.1093/ajh/hpab129] [Medline: 34382648]
48. Steinberg DM, Kay MC, Svetkey LP, Askew S, Christy J, Burroughs J, et al. Feasibility of a digital health intervention to improve diet quality among women with high blood pressure: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2020 Dec 07;8(12):e17536. [doi: 10.2196/17536] [Medline: 33284116]
49. Thiboutot J, Sciamanna CN, Falkner B, Kephart DK, Stuckey HL, Adelman AM, et al. Effects of a web-based patient activation intervention to overcome clinical inertia on blood pressure control: cluster randomized controlled trial. J Med Internet Res 2013 Sep 04;15(9):e158 [FREE Full text] [doi: 10.2196/jmir.2298] [Medline: 24004475]
50. Van Emmenis M, Jamison J, Kassavou A, Hardeman W, Naughton F, A'Court C, et al. Patient and practitioner views on a combined face-to-face and digital intervention to support medication adherence in hypertension: a qualitative study within primary care. BMJ Open 2022 Feb 28;12(2):e053183 [FREE Full text] [doi: 10.1136/bmjopen-2021-053183] [Medline: 35228280]
51. Wang SQ, Leung M, Leung SY, Han J, Leung W, Hui E, et al. Correction: safety, feasibility, and acceptability of telemedicine for hypertension in primary care: a proof-of-concept and pilot randomized controlled trial (SATE-HT). J Med Syst 2023 Jul 05;47(1):68. [doi: 10.1007/s10916-023-01965-w] [Medline: 37405511]
52. Willis M, Darwiche G, Carlsson M, Nilsson A, Wohlin J, Lindgren P. Real-world long-term effects on blood pressure and other cardiovascular risk factors for patients in digital therapeutics. Blood Press Monit 2023 Apr 01;28(2):86-95 [FREE Full text] [doi: 10.1097/MBP.0000000000000633] [Medline: 36729897]
53. Zhang Y, Tao Y, Zhong Y, Thompson J, Rahmani J, Bhagavathula AS, et al. Feedback based on health advice via tracing bracelet and smartphone in the management of blood pressure among hypertensive patients: a community-based RCT trial in Chongqing, China. Medicine (Baltimore) 2022 Jul 15;101(28):e29346 [FREE Full text] [doi: 10.1097/MD.0000000000029346] [Medline: 35839004]
54. Naqvi IA, Strobino K, Li J, Schmitt K, Garcon M, Li H, et al. Abstract 108: home blood pressure telemonitoring-enhanced versus usual post-acute stroke care in an underserved setting: the telehealth after stroke care pilot randomized clinical trial. Stroke 2022 Feb;53(Suppl_1):33. [doi: 10.1161/str.53.suppl_1.108]
55. Bakre S, Shea B, Langheier J, Hu EA. Blood pressure control in individuals with hypertension who used a digital, personalized nutrition platform: longitudinal study. JMIR Form Res 2022 Mar 17;6(3):e35503 [FREE Full text] [doi: 10.2196/35503] [Medline: 35297775]
56. Chen R, Sun J, Dittus RS, Fabbri D, Kirby J, Laffer CL, et al. Patient stratification using electronic health records from a chronic disease management program. IEEE J Biomed Health Inform (Forthcoming) 2016 Jan 04 [FREE Full text] [doi: 10.1109/JBHI.2016.2514264] [Medline: 26742152]
57. Hellem AK, Whitfield C, Casetti A, Robles MC, Dinh M, Meurer W, et al. Engagement in self-measured blood pressure monitoring among medically underresourced participants (the reach out trial): digital framework qualitative study. JMIR Cardio 2023 Apr 07;7:e38900. [doi: 10.2196/38900] [Medline: 37027200]
58. Bernal SL, Valverde JM, Celdrán AH, Pérez GM. SENIOR: an intelligent web-based ecosystem to predict high blood pressure adverse events using biomarkers and environmental data. Appl Sci 2021 Mar 11;11(6):2506. [doi: 10.3390/app11062506]
59. Chiang PH, Wong M, Dey S. Using wearables and machine learning to enable personalized lifestyle recommendations to improve blood pressure. IEEE J Transl Eng Health Med 2021;9:1-13. [doi: 10.1109/jtehm.2021.3098173]
60. Ye XY, Zeng QT, Facelli JC, Brixner DI, Conway M, Bray BE. Predicting optimal hypertension treatment pathways using recurrent neural networks. Int J Med Inform 2020 Jul;139:104122 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104122] [Medline: 32339929]
61. Abrar S, Loo CK, Kubota N, Tahir GA. A personalised blood pressure prediction system using gaussian mixture regression and online recurrent extreme learning machine. In: Proceedings of the 2020 International Symposium on Community-centric

XSL•FO
RenderX

Systems. 2020 Presented at: CcS '20; September 23-26, 2020; Tokyo, Japan p. 23-26 URL: https://ieeexplore.ieee.org/document/9231328 [doi: 10.1109/ccs49175.2020.9231328]
62. Bertsimas D, Borenstein AR, Dauvin A, Orfanoudaki A. Ensemble machine learning for personalized antihypertensive treatment. Nav Res Logist 2021 Dec 10;69(5):669-688. [doi: 10.1002/nav.22040]
63. Cano J, Hornero F, Quesada A, Martinez-Rodrigo A, Alcaraz R, Rieta JJ. Improved discrimination between healthy and hypertensive individuals combining photoplethysmography and electrocardiography. In: Proceedings of the 2021 Conference on Computing in Cardiology. 2021 Presented at: CinC '21; September 13-15, 2021; Brno, Czech Republic p. 1-4 URL: https://ieeexplore.ieee.org/document/9662894 [doi: 10.23919/cinc53138.2021.9662894]
64. Hu Y, Huerta J, Cordella N, Mishuris RG, Paschalidis IC. Personalized hypertension treatment recommendations by a data-driven model. BMC Med Inform Decis Mak 2023 Mar 01;23(1):44. [doi: 10.1186/s12911-023-02137-z] [Medline: 36859187]
65. Sun J, McNaughton CD, Zhang P, Perer A, Gkoulalas-Divanis A, Denny JC, et al. Predicting changes in hypertension control using electronic health records from a chronic disease management program. J Am Med Inform Assoc 2014 Mar 01;21(2):337-344 [FREE Full text] [doi: 10.1136/amiajnl-2013-002033] [Medline: 24045907]
66. Mabry PL, Olster DH, Morgan GD, Abrams DB. Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. Am J Prev Med 2008 Aug;35(2 Suppl):S211-S224. [doi: 10.1016/j.amepre.2008.05.018] [Medline: 18619402]
67. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA 2019 Dec 24;322(24):2377-2378 [FREE Full text] [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
68. Konstantinidis D, Iliakis P, Tatakis F, Thomopoulos K, Dimitriadis K, Tousoulis D, et al. Wearable blood pressure measurement devices and new approaches in hypertension management: the digital era. J Hum Hypertens 2022 Nov 23;36(11):945-951 [FREE Full text] [doi: 10.1038/s41371-022-00675-z] [Medline: 35322181]
69. Hood CM, Gennuso KP, Swain GR, Catlin BB. County health rankings: relationships between determinant factors and health outcomes. Am J Prev Med 2016 Feb;50(2):129-135. [doi: 10.1016/j.amepre.2015.08.024] [Medline: 26526164]
70. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
71. Fan H, Poole M. What is personalization? Perspectives on the design and implementation of personalization in information systems. J Organ Comput Electron Commer 2006:16-202. [doi: 10.1207/s15327744joce1603&4_2]
72. Chaikijurajai T, Laffin LJ, Tang WH. Artificial intelligence and hypertension: recent advances and future outlook. Am J Hypertens 2020 Nov 03;33(11):967-974 [FREE Full text] [doi: 10.1093/ajh/hpaa102] [Medline: 32615586]
73. Radhakrishnan K, Kim MT, Burgermaster M, Brown RA, Xie B, Bray MS, et al. The potential of digital phenotyping to advance the contributions of mobile health to self-management science. Nurs Outlook 2020;68(5):548-559. [doi: 10.1016/j.outlook.2020.03.007] [Medline: 32402392]
74. Pei Z, Liu J, Liu M, Zhou W, Yan P, Wen S, et al. Risk-predicting model for incident of essential hypertension based on environmental and genetic factors with support vector machine. Interdiscip Sci 2018 Mar;10(1):126-130. [doi: 10.1007/s12539-017-0271-2] [Medline: 29380342]
75. Maxwell A, Li R, Yang B, Weng H, Ou A, Hong H, et al. Deep learning architectures for multi-label classification of intelligent health risk prediction. BMC Bioinformatics 2017 Dec 28;18(Suppl 14):523 [FREE Full text] [doi: 10.1186/s12859-017-1898-z] [Medline: 29297288]
76. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res 2018 Jan 30;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]
77. du Toit C, Tran TQ, Deo N, Aryal S, Lip S, Sykes R, et al. Survey and evaluation of hypertension machine learning research. J Am Heart Assoc 2023 May 02;12(9):e027896 [FREE Full text] [doi: 10.1161/JAHA.122.027896] [Medline: 37119074]
78. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020 Sep 09;26(9):1364-1374 [FREE Full text] [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]
79. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:1-14 [FREE Full text] [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]

Abbreviations

BP: blood pressure
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
EHR: electronic health record
HTN: hypertension
ML: machine learning
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

RCT: randomized controlled trial
TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence

Edited by A Mavragani; submitted 23.04.24; peer-reviewed by N Chapman, WX Lai; comments to author 22.10.24; revised version received 12.11.24; accepted 16.12.24; published 10.02.25.

Please cite as:
Clifford N, Tunis R, Ariyo A, Yu H, Rhee H, Radhakrishnan K
Trends and Gaps in Digital Precision Hypertension Management: Scoping Review
J Med Internet Res 2025;27:e59841
URL: https://www.jmir.org/2025/1/e59841
doi:10.2196/59841
PMID:

©Namuun Clifford, Rachel Tunis, Adetimilehin Ariyo, Haoxiang Yu, Hyekyun Rhee, Kavita Radhakrishnan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Interventions for Digital Addiction: Umbrella Review of Meta-Analyses

Peng Lu[1], JSD; Jiamin Qiu[2], BM; Shiqi Huang[2], BS; Xinman Wang[1], UG; Shasha Han[3], MD; Sui Zhu[2], MD; Youjing Ning[1], UG; Fang-fang Zeng[2*], MD; Yuan Yuan[1*], JSD

[1]Department of Chinese International Education, Chinese Language and Culture College, Huaqiao University, Xiamen, China

[2]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[3]Department of Neonatology and Pediatrics, The First Affiliated Hospital of Jinan University, Jinan University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Yuan Yuan, JSD
Department of Chinese International Education
Chinese Language and Culture College
Huaqiao University
No 8 Jiageng Road
Xiamen, 361021
China
Phone: 86 6068002
Email: yuanyuan361021@163.com

Abstract

Background: Numerous studies have explored interventions to reduce digital addiction outcomes, but inconclusive evidence makes it difficult for decision-makers, managers, and clinicians to become familiar with all available literature and find appropriate interventions.

Objective: This study aims to summarize and assess the certainty of evidence of interventions proposed to decrease digital addiction from published meta-analyses.

Methods: An umbrella review of published meta-analyses was performed. We searched PubMed, Cochrane Library, Web of Science, and Embase for meta-analyses published up to February 2024. Eligible studies evaluated interventions using randomized controlled trials, nonrandomized controlled trials, or quasi-experimental studies and were assessed for methodological quality using Assessment of Multiple Systematic Reviews version 2. A random effects model was used to analyze data, considering heterogeneity and publication bias. Grading of Recommendations, Assessment, Development, and Evaluations was applied to assess evidence with certainty.

Results: A total of 5 studies assessing 21 associations were included in the umbrella review, of which 4 (80%) were high-quality meta-analyses. Weak evidence was observed in 19 associations, whereas null associations appeared in the remaining 2 associations. These associations pertained to 8 interventions (group counseling, integrated internet addiction [IA] prevention programs, psychosocial interventions, reality therapy, self-control training programs, cognitive behavioral therapy, interventions to reduce screen time in children, and exercise) and 9 outcomes (self-control, self-esteem, internet gaming disorder symptoms, time spent gaming, IA scores, screen use time, interpersonal sensitivity longlines, anxiety, and depression). Cognitive behavioral therapy reduces anxiety (standardized mean difference [SMD] 0.939, 95% CI 0.311 to 1.586), internet gaming disorder symptoms (SMD 1.394, 95% CI 0.664 to 2.214), time spent gaming (SMD 1.259, 95% CI, 0.311 to 2.206), and IA scores (SMD –2.097, 95% CI –2.814 to –1.381). Group counseling had a large effect on improving self-control (SMD 1.296, 95% CI 0.269 to 2.322) and reducing IA levels (SMD –1.417, 95% CI –1.836 to –0.997). Exercise intervention reduced IA scores (SMD –2.322, 95% CI –3.212 to –1.431), depression scores (SMD –1.421, 95% CI –2.046 to –797), and interpersonal sensitivity scores (SMD –1.433, 95% CI –2.239 to –0.627).

Conclusions: The evidence indicates that current interventions to reduce digital addiction are weak. Data from more and better-designed studies with larger sample sizes are needed to establish robust evidence.

Trial Registration: PROSPERO CRD42024528173; crd.york.ac.uk/PROSPERO/display_record.php?RecordID=528173

XSL•FO
RenderX

(J Med Internet Res 2025;27:e59656) doi:10.2196/59656

KEYWORDS

digital addiction; interventions; umbrella reviews; systematic review; internet addiction; loneliness; quality of life; well-being; internet; psychological; pharmacological treatment; cognitive behavioral therapy

Introduction

The World Health Organization has officially recognized digital addiction as a global problem in which excessive digital activities and internet use lead to the inability to manage time, decreased energy and concentration during the day, and procrastination of bedtime and shortened total sleep time due to disrupted sleep patterns at night, thereby reducing an individual's subjective happiness [1,2]. A survey in 31 countries revealed that 6% of internet users aged 12-41 years had symptoms of digital addiction [3]. The prevalence of digital addiction varies globally, with an average of 4.6% in Western countries and 8.9% in other countries [4]. Digital addiction has negative effects on personal work and life; damages personal physical and mental health; causes physical health problems such as fatigue and a lack of sleep; causes negative emotions such as anxiety, depression, and anger; reduces work concentration and work efficiency; damages social relationships; leads to loneliness; reduces life satisfaction; and affects the growth of adolescents [5-8].

Many studies in the fields of psychology and medicine have focused on the harms of digital addiction and potential intervention strategies. Moreover, numerous studies have pointed to personal traits and emotional self-regulatory mechanisms (eg, self-esteem and self-control) as key factors associated with the propensity to become digitally addicted [9-11]. In addition, elevated "fear of missing out" awareness compels individuals to consistently use digital networks to maintain social connections, which may lead to inappropriate use of digital media and the emergence of anxiety symptoms that exacerbate digital addiction [12,13]. Resilience is an expression of positive adaptation in the face of great adversity, which can neutralize the harmful effects of stress and thus reduce the tendency to become addicted to the internet; however, when faced with stress that exceeds their resilience, individuals may be inclined to overuse the internet as a coping strategy to alleviate stress [14,15]. Owing to the magnitude of the impact of psychological disorders on digital addiction, for individuals, the prevention and treatment of digital addiction are mainly psychological and pharmacological [1]. Beute et al [16] report that cognitive behavioral therapy (CBT) can play a role in treating adolescent internet addiction (IA)**.** CBT and CBT-based psychotherapies are the psychotherapy most commonly used for digital addiction [17]. Similarly, psychosocial intervention, a unique type of psychotherapy, is believed to alleviate digital addiction by enhancing an individual's self-esteem and self-regulation [18]. In addition, adjuvant therapies such as sports interventions can be effective in preventing and treating IA [19]. Alternative therapy uses alternative methods such as reading to reduce dependence on the internet, thereby achieving the purpose of relieving emotions. Many researchers have also tried to combine two or more interventions to achieve better therapeutic effects.

Currently, there are several meta-analyses on the prevention and intervention of digital addiction [20-22]. Liu et al [20] combine 58 randomized controlled trials (RCTs) and report that group counseling programs, CBT, and sports interventions reduce IA symptoms. Matthew et al [22] report that CBT was a short-term intervention for reducing internet gaming disorder (IGD) and depression symptoms, but its role in reducing actual time spent gaming was unknown. A study by Zhang et al [21] indicates that exercise can effectively reduce the symptoms of anxiety, depression, and loneliness among internet-addicted students and is beneficial to mental health. However, the heterogeneity of most meta-analyses on the treatment and prevention of digital addiction, due to reasons such as lack of a consistent definition, different outcome measurements, and wide variation in study design, limit the clinical impact of such results. Therefore, it is necessary to summarize the current evidence related to digital addiction intervention methods and conduct an overall assessment of their quality based on comprehensive evidence grading standards by using umbrella reviews which could consolidate the highest quality level of evidence by merging and assessing the meta-analyses [23,24].

Accordingly, this study conducted an umbrella analytic review of relevant meta-analyses with the purpose of evaluating the effects of various intervention modalities for digital addiction; systematically examined, organized, and assessed the strength of evidence and potential degree of bias from multiple reviews and meta-analyses; and conducted a literature review of interventions for digital addiction to identify the most effective interventions and provide recommendations to guide future health-related research and develop effective therapeutic interventions.

Methods

Study Design

This umbrella review was conducted in accordance with the registered protocol (Open Science Framework Registry) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [25]. Our study protocol has been registered with PROSPERO.

Search Strategy and Selection Criteria

We searched the following databases for English language studies published up to February 2024: PubMed, the Cochrane Library, Web of Science, and Embase. We used key terms related to digital addiction and interventions. The following search strategy was used: ("Internet" OR "digital" OR "screen" OR "net" OR "online" OR "media" OR "electronic device" OR "electronic gadgets" OR "computer" OR "mobile" OR "phone" OR "smartphone" OR "television" OR "TV" OR "video" OR

XSL•FO
RenderX

"Facebook" OR "game" OR "gaming") AND ("addict" OR "use" OR "dependent" OR "overuse" OR "abuse" OR "disorder" OR "excessive" OR "effects" OR "habits" OR "intervene" OR "treatment" OR "therapy" OR "compulsive" OR "heavy") AND ("meta-analy" OR "meta-analytic" OR "metaanaly" OR "meta-regression" OR "meta synthesis" OR "meta-analysis" OR "meta-analysis") AND ("intervention" OR "interfere" OR "intervene" OR "treatment" OR "therapy" OR "cure" OR "therapeutic" OR "exercise therapy" OR "counseling"). Studies were published in full peer-reviewed literature without language or date restrictions.

The studies included in the meta-analysis had to meet the following inclusion criteria: (1) research topics related to digital addiction intervention; (2) participants diagnosed with digital addiction; (3) a study design limited to RCTs, nonrandomized controlled trials, and quasi-experimental studies; and (4) different forms of digital addiction with similar effects.

The exclusion criteria were as follows: (1) duplicate publications; (2) full-text documents could not be obtained, and other experimental-related basic research was performed; and (3) meta-analysis without any effect size. When 2 or more meta-analyses were performed for the same association, only the most recent systematic review with the largest number of individual studies was included to avoid duplication of samples (Multimedia Appendix 1 [18,20,21,26-39]).

Study Selection

Documentation records were extracted and imported into Covidence (Veritas Health Innovation). After excluding duplicates, titles and abstracts were assessed independently by 2 researchers. The full texts of the relevant studies were independently reviewed by 2 researchers to ensure that the inclusion and exclusion criteria were met. In the event of discrepancies, a final assessment was made by a third independent researcher.

Data Extraction and Methodological Quality Assessment

Data extraction was independently completed by 2 researchers. The information extracted from each meta-analysis was as follows: first author; publication year; design and number of included studies; statistical estimates of the meta-analysis; number of original studies; sample size and region; intervention; comparison; any relevant subgroup analyses; and intervention-related outcome indicators. Any disputes during data entry were resolved through discussion with a third researcher. When a meta-analysis contained multiple intervention outcome indicators, each indicator was extracted and analyzed separately.

Two reviewers independently assessed the methodological quality of each systematic review via the Assessment of Multiple Systematic Reviews version 2 (AMSTAR 2) rating scale [40]. AMSTAR 2 consists of 16 items and is a measurement tool used to evaluate meta-analyses. The scale has good consistency, reliability, construct validity, and feasibility. It covers key questions of the included studies, proposals, literature searches, literature screening, data extraction, basic characteristics of the included original studies, data analysis, and conflicts of interest. The overall quality of a systematic review is informed by assessing how well the study meets each of the criteria. Overall quality is classified as "high," "medium," "low," or "very low" [40].

Data Analysis

The umbrella review compiles the evidence of multiple clinical-related issues in the system reviews and meta-analyses and then comprehensively evaluates the strength of the meta-analyses related to specific issues. First, the combined effect values and their 95% CIs associated with digital addiction interventions in each meta-analysis and their 95% CIs were extracted, and the combined effect sizes, 95% CIs, and *P* values were reestimated after the normalization of the mean difference (MD) and Hedges *g* to the standardized mean difference (SMD) via a random effects model (DerSimonian and Laird) by referring to the analyses of the previous omnibus review. I^2 and Cochran *Q* were used to evaluate the heterogeneity of the evidence in the literature, with $I^2 \geq 50\%$ or Cochran *Q* test *P* value of less than .10 indicating significant heterogeneity in the evidence. The estimated values of the publication bias of various studies through Egger inspection and evaluation, *P* value of less than .10, indicate that there is a significant publication bias. When the results indicate that there is a significant bias, if the total estimated effect value of the meta-analysis is greater than the effect value of the original research with the largest proportion of its effect, it indicates that the meta-analysis has a small-scale research effect. By determining whether the number of statistical results (*P* value of less than .05) in each meta-analysis (*P* value of less than .05) is greater than its expected quantity (E), the meta-analysis is excessive E, and *P* value of less than .10 indicates excessive significance. The rest of the analyses used a difference of *P* value of less than .05 to determine whether the difference was statistically significant.

The results are presented in forest plots and tables, and all analyses and mapping were performed via Stata (version 17.7, StataCorp) and R (version 4.2.1; R Foundation for Statistical Computing).

Determining the Credibility of Evidence

The Grading of Recommendations, Assessments, Development, and Evaluations (GRADE) approach was used to evaluate the strength of the evidence in each included study. According to the statistical analysis results, the credibility of each piece of evidence is further divided into 5 levels: I, II, III, IV, and NS, which represent "convincing evidence," "highly suggestive evidence," "suggestive evidence," "weak evidence," and "insignificant evidence," respectively [41].

Results

Study Characteristics

We retrieved 284 meta-analyses from PubMed, 277 from Embase, 319 from Web of Science, and 75 from Cochrane. Two investigators independently screened the studies, and 5 meta-analyses were ultimately included in this umbrella review through 2 rounds of literature screening (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of the study selection process.

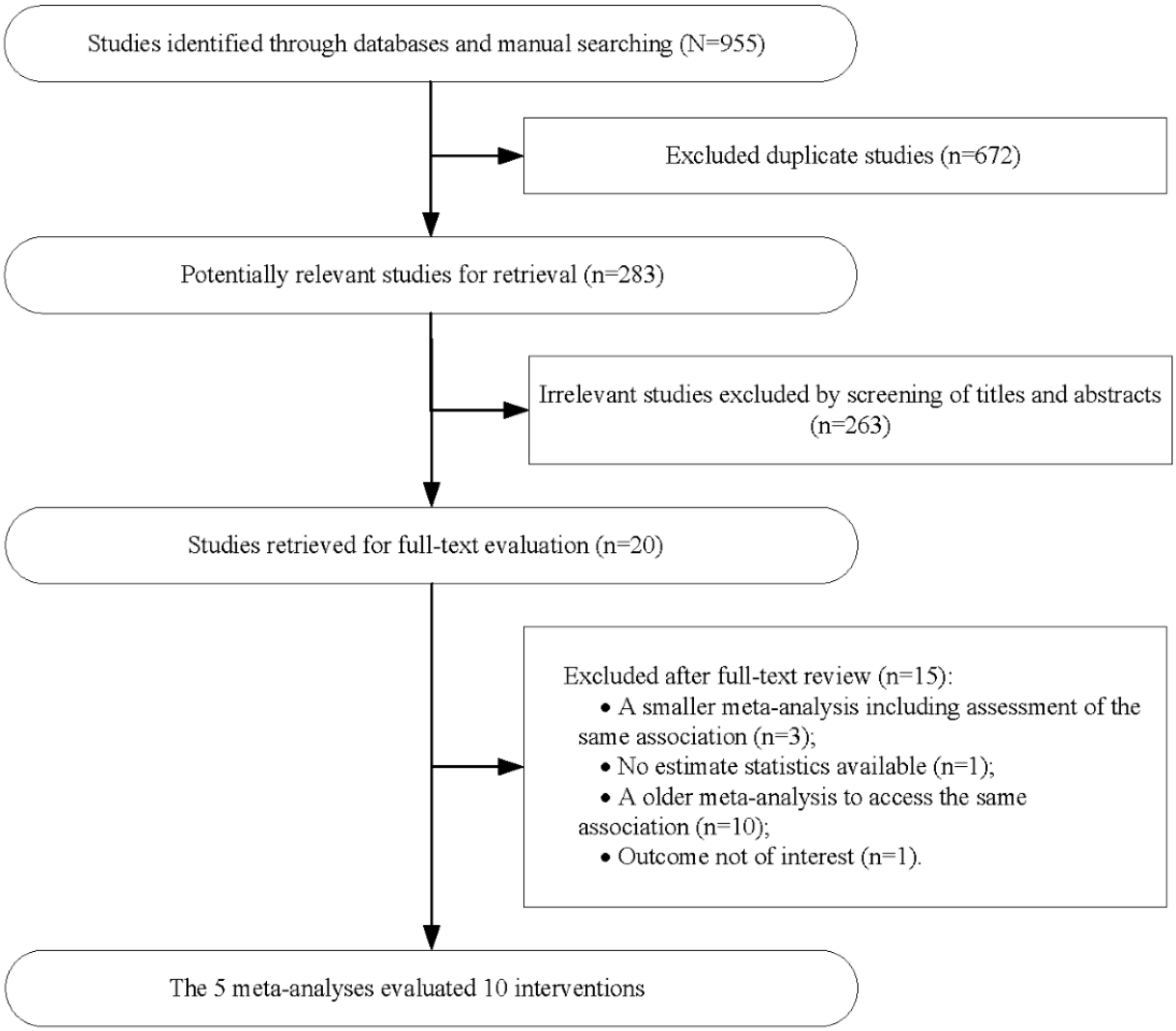

The number of studies included in the 5 meta-analyses ranged from 11 to 59, with a median of 7 (IQR 6-11). The sample size ranged from 580 to 4656, with a median of 407 (IQR 107-1490). The original research covered multiple regions, including Asia, Europe, Oceania, and the Americas. AMSTAR 2 conducted a methodological quality assessment of the 5 meta-analyses included in this umbrella review. Among them, 4 studies were rated as “high,” and 1 study was rated as “low” (Table 1).

XSL•FO
RenderX

Table 1. Characteristics and quality assessment of included meta-analyses to access interventions of digital addiction.

Author (year)	Regions of the original study	Study design	Studies (participants)	Population	Exposure	Follow-up	Comparison	Outcomes (type of ES[a])	SE (95% CI)	AMSTAR 2[b]
Yeun (2016) [18]	East Asia	NRCT[c] and RCT[d]	37 (1490)	IA[e]	Psychosocial intervention	NR[f]	Control	• Improving self-control (SMD[g]) • Self-esteem (MD[h])	• 0.29 (0.11-0.47) • 3.58 (2.03-5.12)	High
Stevens (2018) [22]	China, South Korea, Japan, Germany, United States, and Brazil	NR	12 (580)	IGD[i]	CBT[j]	3-6 months	Control	• Reducing IGD symptoms (Hedges *g*) • Depression (Hedges *g*) • Anxiety (Hedges *g*)	• 0.92 (0.5-1.34) • 0.8 (0.21-1.38) • 0.55 (0.17-0.93)	High
Martin (2021) [42]	United States, Canada, Australia, Switzerland, Norway, New Zealand, and Japan	RCT, quasi-experimental	11 (4656)	Children	Interventions to reduce screen time in children	NR	Control	Screen use time (MD)	-0.56 (-0.92 to -0.2)	Low
Zhang (2022) [26]	China, Korea, and Iran	RCT	59 (3832)	IA	CBT, group counseling, sports intervention, and internet-based intervention	1 week to 5 months	No-treatment control group	Reducing the total scores of IA (SMD)	–1.9 (–2.26 to –1.55)	High
Zhang (2023) [21]	China	RCT	39 (2408)	IA student	Exercises	8-24 months	NR	• Loneliness (SMD) • Anxiety (SMD) • Depression (SMD) • Interpersonal sensitivity (SMD)	• –0.96 (–1.5 to –0.41) • –1.79 (–2.37 to –1.22) • –1.5 (–2.19 to –0.81) • –1.34 (–2.05 to –0.64)	High

[a]ES: effect statistics.

[b]AMSTAR 2: assessment of multiple systematic reviews version 2.

[c]NRCT: nonrandomized controlled trial.

[d]RCT: randomized controlled trial.

[e]IA: internet addiction.

[f]NR: not reported.

[g]SMD: standardized mean difference.

[h]MD: mean difference.

[i]IGD: internet gaming disorder.

[j]CBT: cognitive behavioral therapy.

Quantitative Synthesis and Evidence Grading of the Meta-Analysis

Five meta-analyses evaluated 9 interventions: CBT, group counseling, exercise, integrated IA prevention programs, psychosocial interventions, reality therapy, self-control training programs, and interventions to reduce screen time in child and art therapy. The general characteristics of the studies included in the meta-analysis are presented in Table 1. Four of the selected meta-analyses were graded as “high quality” on the

XSL•FO
RenderX

basis of the AMSTAR 2 score, and only 1 review was graded as "low quality." However, the evidence class of most meta-analyses included evidence rated as "weak evidence" (19/21, 90%) and "not significant" (2/21, 10%). A summary of the quality appraisals of the meta-analyses and the AMSTAR 2 scores are presented in Table 1 and Multimedia Appendix 2 [18,21,22,26,42].

Cognitive Behavioral Therapy

CBT is effective in treating substance abuse, gambling, and affective and eating disorders. CBTs are based on the cognitive behavioral model, which holds that thoughts determine feelings; thus, changing one's thoughts can help with behavioral change [43]. CBT can reduce anxiety, depression, IGD symptoms, time spent gaming, and IA scores, as shown in this review. Two meta-analyses previously examined the effectiveness of CBT for patients with digital addiction [22,44]. For example, Stevens et al [22] report that CBT demonstrated high efficacy in reducing IGD symptoms (SMD 1.394, 95% CI 0.664-2.214) and depression (SMD 0.797, 95% CI 0.254-1.341) and moderate efficacy in reducing anxiety (SMD 0.939, 95% CI 0.311-1.586) and time spent on gaming (SMD 1.259, 95% CI 0.311-2.206). However, owing to the evidence base, there is a need for more rigorous studies to determine the potential long-term benefits of CBT for IGD [22]. The heterogeneity in the effects of CBT on reducing anxiety (I^2=92.5), depression (I^2=87.1), IGD symptoms (I^2=96.6), and time spent gaming (I^2=95.9) was high, although the effects of CBT on depression exhibited publication bias, albeit the evidence was weak. Another review explored the impact of CBT on IA levels [44] and found CBT to be significantly superior to routine intervention and paroxetine combined with buspirone drug therapy (*P* value of less than .05) on the IA test scale and also superior to routine intervention and no intervention (*P* value of less than .05) on the basis of the Revised Chen Internet Addiction scale and Young Diagnostic Questionnaire scale (*P* value of less than .05) [40]. Finally, the heterogeneity for the effects of CBT on reducing IA scores was high (I^2=92.5; Figure 2 and Table 2).

Figure 2. Forest plot of interventions for digital addiction. CBT: cognitive behavioral therapy; IA: internet addiction; IGD: internet gaming disorder; NS: not significant.

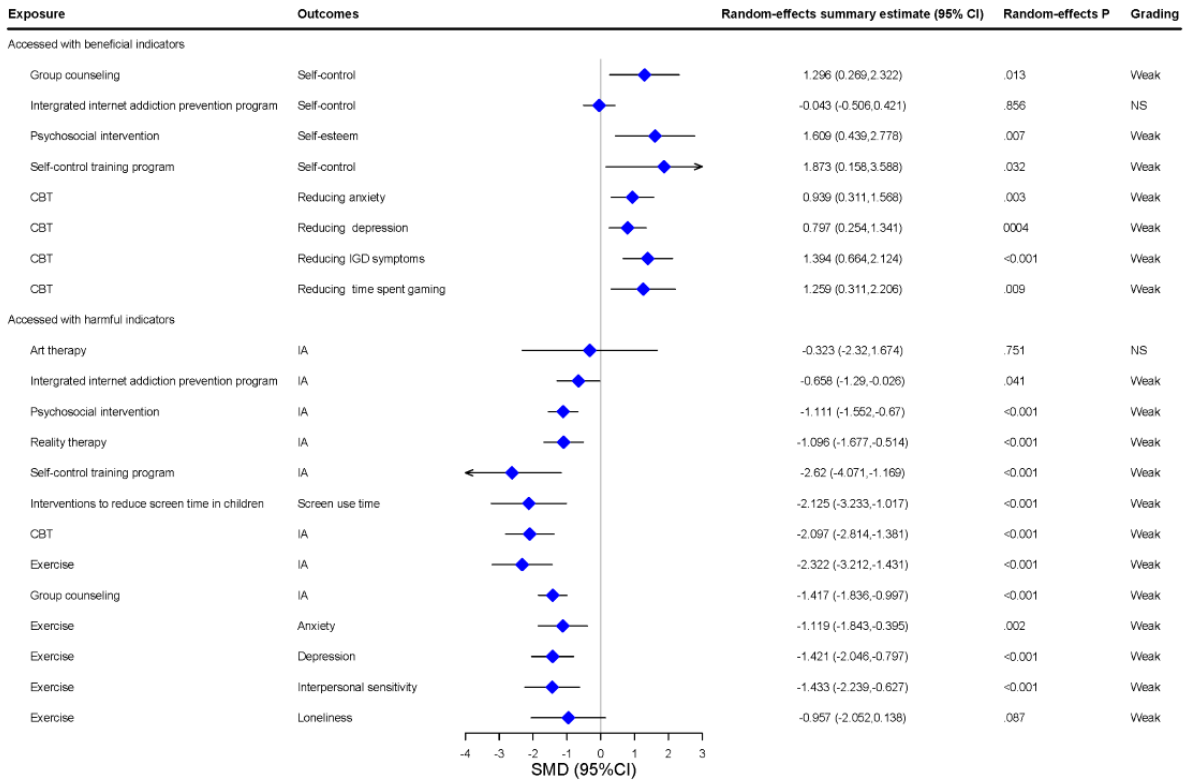

XSL•FO
RenderX

Table 2. Quantitative synthesis and evidence grading of meta-analyses to access interventions of digital addiction.

Study (year)	Exposures	Comparison	Outcomes	Studies (participants)	Original effect metrics	Random-effects summary estimate (95% CI)	Random effects, *P* value	I^2 (%)	95% prediction interval	Egger test>*P* value	Largest study estimate (95% CI)	Significant studies		Grading
												O/E[a]	*P* value	
Yeun and Han 2016) [18]	Art therapy	Control	IA[b]	3 (52)	SMD[c]	–0.323 (–2.32 to 1.674)	.751	90	(<0.001, >1000)	0.172	18.432 (2.909-114.686)	2/2.62	—[d]	NS[e]
Yeun and Han (2016) [18]	Group counseling	Control	Self-control	3 (70)	SMD	1.296 (0.269-2.322)	.013	66.9	(<0.001, >1000)	0.154	2.806 (0.865-9.266)	2/0.78	.17	Weak
Yeun and Han (2016) [18]	Integrated IA prevention program	Control	IA	4 (461)	SMD	–0.658 (–1.29 to –0.026)	.041	88.1	(0.002, 57.896)	0.027	1.095 (0.722-1.690)	3/0.23	<.01	Weak
Yeun and Han (2016) [18]	Integrated IA prevention program	Control	Self-control	4 (291)	SMD	–0.043 (–0.506 to 0.421)	.856	73.4	(0.025, 34.189)	0.319	0.320 (0.164-0.636)	1/2.64	—	NS
Yeun and Han (2016) [18]	Psychosocial intervention	Control	IA	16 (625)	SMD	–1.111 (–1.552 to –0.67)	<.001	82.5	(0.006, 3.123)	0.004	1.135 (0.581-2.178)	13/0.88	<.01	Weak
Yeun and Han (2016) [18]	Psychosocial intervention	Control	Self-esteem	6 (142)	MD[e]	1.609 (0.439-2.778)	.007	87.7	(0.010, >1000)	0.359	11.064 (2.490-49.162)	4/5.08	—	Weak
Yeun and Han (2016) [18]	Reality therapy	Control	IA	2 (54)	SMD	–1.096 (–1.677 to –0.514)	<.001	0	—	—	0.182 (0.050-0.672)	2/1.28	.54	Weak
Yeun and Han (2016) [18]	Self-control training program	Control	IA	5 (107)	SMD	–2.62 (–4.071 to –1.169)	<.001	85.6	(<0.001, 140.7)	0.020	0.503 (0.100-2.517)	4/0.67	<.01	Weak
Yeun and Han (2016) [18]	Self-control training program	Control	Self-control	3 (67)	SMD	1.873 (0.158, 3.588)	.032	86.8	(<0.001, >1000)	0.202	1.919 (0.435, 8.618)	2/0.39	.05	Weak
Stevens et al (2018) [22]	CBT[g]	Control	Reducing anxiety	7 (586)	Hedges *g*	0.939 (0.311-1.568)	.003	92.5	(0.097, 310.259)	0.216	2.34 (1.288-4.255)	4/3.34	.72	Weak
Stevens et al (2018) [22]	CBT	Control	Reducing depression	6 (303)	Hedges *g*	0.797 (0.254-1.341)	.004	87.1	(0.139, 128.859)	0.935	1.135 (0.592-2.178)	4/0.34	<.01	Weak
Stevens et al (2018) [22]	CBT	Control	Reducing IGD[h] symptoms	11 (1164)	Hedges *g*	1.394 (0.664-2.124)	<.001	96.6	(0.072, >1000)	0.833	8.618 (6.222-11.938)	9/10.63	—	Weak

XSL•FO
RenderX

Study (year)	Exposures	Comparison	Outcomes	Studies (participants)	Original effect metrics	Random-effects summary estimate (95% CI)	Random effects, *P* value	I^2 (%)	95% prediction interval	Egger test>*P* value	Largest study estimate (95% CI)	Significant studies		Grading
												O/E[a]	*P* value	
Stevens et al (2018) [22]	CBT	Control	Reducing time spent gaming	4 (397)	Hedges *g*	1.259 (0.311-2.206)	.009	95.9	(0.003, >1000)	0.226	74.277 (44.746-123.298)	4/4	≥.99	Weak
Martin et al (2021) [42]	Interventions to reduce screen time in children	Control	Screen use time	7 (1905)	MD	–2.125 (–3.233 to –1.017)	<.001	98.7	(<0.001, 34)	0.698	0.020 (0.014-0.029)	45450	—	Weak
Zhang et al (2022) [26]	CBT	No intervention	Total scores of IA	7 (491)	SMD	–2.097 (–2.814 to –1.381)	<.001	90.7	(<0.001, 2.094)	0.239	0.048 (0.022-0.102)	7/6.96	≥.99	Weak
Zhang et al (2022) [26]	Exercise	No intervention	Total scores of IA	4 (285)	SMD	–2.322 (–3.212 to –1.431)	<.001	88.1	(<0.001, 27.397)	0.247	0.031 (0.012, 0.079)	4/4	≥.99	Weak
Zhang et al (2022) [26]	Group counseling	No intervention	Total scores of IA	11 (416)	SMD	–1.417 (–1.836 to –0.997)	<.001	70.3	(0.006, 0.978)	0.545	0.215 (0.082-0.550)	10/7.18	.11	Weak
Zhang et al (2023) [21]	Exercise	Control	Anxiety	9 (362)	SMD	–1.119 (–1.843 to –0.395)	.002	87	(0.003, 7.315)	0.492	0.207 (0.071-0.613)	6/6.24	—	Weak
Zhang et al (2023) [21]	Exercise	Control	Depression	4 (102)	SMD	–1.421 (–2.046 to –0.797)	<.001	70.7	(0.001, 9.428)	0.073	0.176 (0.104-0.625)	4/2.45	.16	Weak
Zhang et al (2023) [21]	Exercise	Control	Interpersonal sensitivity	5 (144)	SMD	–1.433 (–2.239 to –0.627)	<.001	79.9	(<0.001, 14.804)	0.106	0.819 (0.215-3.128)	4/0.3	<.01	Weak
Zhang et al (2023) [21]	Exercise	Control	Loneliness	8 (416)	SMD	–0.957 (–2.052 to 0.138)	.087	93.1	(<0.001, 5.319)	0.841	0.108 (0.034-0.350)	6/7.16	—	Weak

[a]O/E: Observed number of studies with positive findings/expected number of studies with positive findings.

[b]IA: internet addiction.

[c]SMD: standardized mean difference.

[d]—: not applicable.

[e]NS: not significant.

[f]MD: mean difference.

[g]CBT: cognitive behavioral therapy.

[h]IGD: internet gaming disorder.

Group Counseling

The main components of group counseling include identifying previous solutions and acknowledging problems while also realizing that exceptions to the problem are key to the solution, with a focus on the present and future [45]. Two meta-analyses previously explored the effectiveness of group counseling on digital addiction. Yeun et al [18] reported that group counseling

XSL•FO
RenderX

had a large effect on improving self-control among school-aged children with IA (SMD 1.296, 95% CI 0.269-2.322). The heterogeneity for the effects of group counseling on improving self-control was high (I^2=66.9), but the evidence was weak. Another review reported that group counseling reduced IA levels (SMD –1.417, 95% CI –1.836 to –0.997) [44]. The heterogeneity for the effects of group counseling on reducing IA scores was high (I^2=95.9; Figure 2 and Table 2).

Exercises

Exercise is an important intervention method that positively impacts individual cognition, emotion, and physical fitness [5]. Two meta-analyses examined this intervention's effectiveness in reducing digital addiction outcomes. One review reported that sports interventions reduced IA scores (SMD –2.322, 95% CI –3.212 to –1.431) [44]. The heterogeneity for the effects of exercise on reducing IA scores was high (I^2=88.1), although the evidence was weak. Another review conducted by Zhang et al [21] included a total of 2408 students from 39 RCTs. Six exercise types (team sport, double sport, single sport, team + double sport, team + single sport, and team + double + single sport) were compared on the basis of their effectiveness in reducing IA and maintaining mental health. The meta-analysis results revealed that, compared with the control group, exercising significantly improved loneliness, anxiety, depression, and interpersonal sensitivity (*P* value of less than .05) [21]. The heterogeneity in the effects of exercise on improving loneliness (I^2=93.1), anxiety (I^2=87), depression (I^2=70.7), and interpersonal sensitivity (I^2=79.9) was high (I^2=95.9; Figure 2 and Table 2).

Psychosocial Interventions

Psychosocial interventions capitalize on psychological or social actions to produce changes in psychological, social, biological, and functional outcomes [46]. One review examined the efficacy of psychosocial interventions on school-aged children's IA, self-control, and self-esteem and revealed that psychosocial interventions had a strong effect on reducing IA (SMD –1.111, 95% CI –1.552 to –0.67) and improving self-esteem (SMD 1.609, 95% CI 0.439-2.778) [18]. The effect on IA had high heterogeneity (I^2=82.5) and published bias (*P*=.004), and the heterogeneity of self-esteem was high (I^2=87.7; Figure 2 and Table 2).

Integrated IA Prevention Program

The integrated IA prevention (IIAP) program combines 2 or more interventions, including CBT, play, art therapy, conflict, stress management, and CBT. [18] A meta-analysis of data on the efficacy of IIAP for self-control and IA revealed that IIAP had no effect on self-control (SMD –0.043, 95% CI –0.506 to 0.421), although the evidence was insignificant [18]. IIAP was found to reduce the IA, and while the evidence was weak, the effect showed high heterogeneity (I^2=88.1), publication bias (*P*=.027), and excessive significance (observed number of studies with positive findings=3; expected number of studies with positive findings=0.23; *P* value of less than .01; Figure 2 and Table 2).

Self-Control Training Program

A meta-analysis of data on the efficacy of self-control training programs for self-control and IA reported in 5 studies revealed that self-control training programs improved self-control (SMD 1.873, 95% CI 0.158 to 3.588) and reduced IA (SMD –2.62, 95% CI –4.071 to –1.169) [18], but the evidence was weak. In addition, the effects were highly heterogeneous and highly significant (*P* value of less than .1). There was a publication bias in terms of the effect on IA (*P*=.02; Figure 2 and Table 2).

Reality Therapy

Weak evidence has indicated that in a review of the effect of reality therapy on IA based on two trials, reality therapy reduced IA (SMD –1.096, 95% CI –1.677 to 0.514) [18] without heterogeneity between the 2 trials (Figure 2 and Table 2).

Interventions to Reduce Screen Time in Children

A meta-analysis suggested that interventions for screen use and sleep can achieve a small reduction in screen time (SMD –2.125, 95% CI –3.233 to –1.017) [47]. Heterogeneity was high (I^2=98.7), and the evidence was weak (Figure 2 and Table 2).

Discussion

Principal Findings

This umbrella review provides an overview of the relationships between digital addiction intervention-related outcome indicators and intervention methods derived from 5 meta-analyses. Four of the selected meta-analyses were of "high" quality. We found that most interventions (eg, CBT and exercise) can improve related outcomes, such as lifestyle, emotional health, mental health, and social relationships, among people with digital addiction. However, this conclusion was derived from weak evidence and has significant heterogeneity and excessive significance. The evidence in this review was rated as "weak evidence" (19/21, 90%) or "not significant" (2/21, 10%). The evidence was downgraded to "weak" considering that most of the included studies were RCTs, which generally involved a limited number of participants, that is, fewer than 1000 in total after pooled analyses, easily causing the small sample sizes limitation, and second, the articles were considered to have exhibited heterogeneity as well as publication bias, suggesting that more high-quality, credible research is needed to further clarify the associations between intervention methods and digital addiction outcomes.

Comparison With Prior Work

On the basis of the results of this umbrella review, CBT not only reduces IA scores but also reduces IGD symptoms, depression, and anxiety. CBT, the psychotherapy most commonly used for digital addiction, can improve symptoms of depression and anxiety by reshaping negative thoughts to help patients understand routine and problematic behaviors and motivate the creation of more adaptive, goal-oriented routines [48]. Compared with other interventions (such as drug replacement therapy), CBT has the advantage of being less invasive and effective in the short term. Furthermore, CBT is the most effective intervention for reducing internet use and improving self-perception [49]. CBT emphasizes the connection

XSL•FO
RenderX

among behaviors, thoughts, and emotions, prompting patients to pay attention to these behaviors [20]. A previous meta-analysis revealed that CBT was associated with positive changes in depression, anxiety, and psychoticism in patients with IA [20]. Ding KY et al [48] also reported that most interventions for digital addiction in children and adolescents were CBT- or CBT-based interventions, which can improve anxiety, depression, and digital addiction-related symptoms. However, another meta-analysis did not conclude that the CBT intervention had a significant effect on the severity of IA [50]. Our analysis of the effect of CBT on digital addiction suggested that genuine heterogeneity likely exists. The high heterogeneity might be related not only to the potential bias in the original studies but also to the differences across the studies included in this meta-analysis. These associations are both supported by a weak level of evidence; thus, more studies are needed to further document the effect of CBT on patients with digital addiction.

There is evidence that group counseling is one of the main treatments for addiction. One of the advantages of group programs is their economic value since they target many students simultaneously [51]. A previous study reported that groups of fewer than 14 people presented effect sizes exceeding 90% [52]. Group counseling programs enable group members to understand and help each other, increase social participation and interpersonal interactions among patients, improve adaptability, develop personalities, and eliminate the symptoms of digital addiction [53]. Group counseling programs can gain patient support and insight through experiencing similar cognitions and emotions, providing a major modality in the treatment of digital addiction [54]. In this review, we conclude that group counseling improved self-control and reduced IA scores among patients with digital addiction. A meta-analysis suggested that group counseling programs improve relationships, health problems, time management, tolerance, and compulsive internet use [20]. A previous meta-analysis by Park et al [55] identified the beneficial effects of group counseling programs for adolescents with IA. However, group counseling may not be truly beneficial until the barriers related to social anxiety, social isolation, and a lack of social skills are overcome. Therefore, developing a protocol for appropriate methods on the basis of the characteristics of the participants is necessary.

Exercise, including running, badminton, basketball, table tennis, and other outdoor sports games, was found to be helpful in reducing IA scores and improving loneliness, anxiety, depression, and interpersonal sensitivity in patients with digital addiction. Exercise is an important intervention method that positively impacts individual cognition, emotion, and physical fitness, in addition to psychological interventions. Digital addiction alters the neural structure, reduces the activity of the dopaminergic system, and limits neurocognitive functions [56]. However, exercise-based interventions can effectively improve the autonomic nervous system in patients with IA and normalize, to a certain extent, the structure of specific parts of the central nervous system [57]. Moreover, it can increase plasma glial cell-derived neurotrophic factor and glucocorticoid levels [58], neurotransmitter release [57], and telomere protection [59]. Exercise-based interventions have been shown to reduce the prevalence and symptoms of IA, and participation in physical activity may prevent IA. Exercise has replaced most internet experiences and improved an individual's physical and mental health [19]. Long-term physical exercise has been found to significantly reduce the degree of IA and depression and improve sleep quality among college students with IA [44]. The meta-analysis included in this review of exercise interventions included only exercise type classification, that is, it did not compare exercise intensity or exercise strength. Thus, more studies on exercise interventions are needed to demonstrate their effects on digital addiction.

Digital addiction may constitute impulsive behavior, although the consequences are negative impacts and involve continuous participation [60]. The impulse theory suggests that the behavior of addiction is the result of the generally excessive active brain reward system. Individuals with excessively active brain reward systems may show strong reactions to the prompts of predicting potential rewards, thereby explaining the pursuit of novelty, impulses, and powerful motivation to obtain internet or other addictive stimuli [61]. Therefore, impulses have been proposed as potential marks and treatment targets of digital addiction [62]. Reality therapy helps individuals control their behavior and consider other options related to internet use [63]. As reality therapy is based on the theory of selecting, the theory holds that people are responsible for their own lives, their feelings, their actions, and their behaviors [64]. In a sense, reality therapy helps individuals reflect on their behavior, evaluate their choices, choose more effective choices, and directly target the choice and self-control of target guidance [64]. In this method, the participants are required to determine the goals of their behavior and realize what they are actually doing. They were then guided to determine whether their behavior promoted or hindered their ability to achieve their original goals. Finally, they are encouraged to seek more appropriate and healthier alternatives to replace their current behavior, achieve their goals, and formulate plans to change negative behaviors [63]. Real-world therapy has been widely used to treat addictive diseases, including diseases involving drugs, sex, and food [63]. For people with addiction and impulses, reality therapy can help them improve their self-control and thereby reduce their problems. Research has shown that reality therapy can increase self-esteem and decrease IA [64]. Similar impacts have also been proven in the study of material addiction in real-world therapy [65]. In this review, we found that real-world therapy reduces IA scores. Despite the study being graded as having high methodological quality, the level of research evidence was weak due to the small number of participants included in the study.

IIAPs are characterized by the simultaneous or sequential delivery of several different types of interventions, including pharmacology, psychotherapy, and group counseling. A composite approach may help solve complex problems, whereas combinations of different therapies with different therapeutic targets and mechanisms may produce synergistic effects. The effects of CBT in combination with other psychosocial or pharmacological therapies for IA have been studied and determined that the combination of CBT with other therapies affects IA severity, somatization, paranoid ideation, psychosis,

XSL•FO
RenderX

obsessive-compulsive traits, interpersonal sensitivity, depression, and anxiety [66]. In this umbrella review, the assessments of IIAPs had high between-study heterogeneity, with public bias and excessive salience, and the evidence of the effect on IA scores was weak. The IIAP showed a null evidence class of effects on self-control, suggesting that there is no clear scientific evidence to support the link between IIAP and self-control in patients with digital addiction.

The evidence reviewed here suggests that CBT, group counseling, exercise, IIAP programs, psychosocial interventions, reality therapy, self-control training programs, and interventions to reduce screen time may be beneficial to patients with digital addiction. Nonetheless, more robust population-based studies are essential to substantiate these potential benefits, as the current evidence is largely inconclusive or weak. It is imperative to conduct further research to evaluate the long-term efficacy and cost-effectiveness of these interventions, as well as to investigate their suitability and generalizability across diverse cultural and socioeconomic settings.

Limitations

This umbrella review has certain limitations. First, this review was based on the results of published and available meta-analyses and meta-analyses, so other results that have not been evaluated by meta-analysis and missing data in the original literature could not be considered. However, our results were not significantly affected because repeated meta-analytic evaluations yielded similar results. Second, throughout the study, original studies covering multiple regions were included in this analysis, and although we were not able to describe subgroups of regions based on the meta-analysis that was conducted, a cross-cultural analysis of multiple geographic sources provides a better description of the effects of various interventions on digital addiction. Third, studies suggest that gender plays a role in digital addiction, as evidenced by the greater susceptibility of males, but the meta-analysis included did not stratify gender to examine the differences in the efficacy of interventions for digital addiction and therefore cannot be discussed further in this study [67]. Fourth, the statistical methods we used to test for the presence of bias indicate only its existence, not its exact source. Finally, the methodological quality of the meta-analyses included in this umbrella review and the quality of evidence for most prognosis-related outcome indicators were not sufficient. Therefore, more high-quality studies are needed in the future to further verify the results of this study.

Acknowledgments

This work was supported by Huaqiao University's Academic Project Supported by the Fundamental Research Funds for the Central Universities, Humanities and Social Sciences Youth Foundation of the Ministry of Education of China (24YJC710055), and Guangdong Basic and Applied Basic Research Foundation (2021A1515012428). The funding sponsors had no role in the design of the study, in the collection, analysis, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Conflicts of Interest

None declared.

Authors' Contributions

Conceptualization: FZ, YY
Data curation: PL, FZ, YY
Formal analysis: PL, JQ, S Huang
Writing – original draft: PL, JQ
Visualization: JQ
Writing – review & editing: PL, JQ, S Huang, XW, S Han, SZ, YN, FZ, YY
Note: FZ and YY contributed equally to this work and are co-corresponding authors. Correspondence should be addressed to FZ at zengffjnu@126.com or YY at yuanyuan361021@163.com.

Multimedia Appendix 1
Excluded studies with reasons for exclusion.
[DOCX File , 54 KB - jmir_v27i1e59656_app1.docx]

Multimedia Appendix 2
Result of the Assessment of Multiple Systematic Reviews version 2 (AMSTAR-2).
[DOCX File , 17 KB - jmir_v27i1e59656_app2.docx]

References

1. Dresp-Langley B, Hutt A. Digital addiction and sleep. Int J Environ Res Public Health 2022;19(11):6910. [doi: 10.3390/ijerph19116910] [Medline: 35682491]

XSL•FO
RenderX

2. Cain N, Gradisar M. Electronic media use and sleep in school-aged children and adolescents: a review. Sleep Med 2010;11(8):735-742. [doi: 10.1016/j.sleep.2010.02.006] [Medline: 20673649]
3. Cheng C, Li AY. Internet addiction prevalence and quality of (real) life: a meta-analysis of 31 nations across seven world regions. Cyberpsychol Behav Soc Netw 2014;17(12):755-760 [FREE Full text] [doi: 10.1089/cyber.2014.0317] [Medline: 25489876]
4. Pan Y, Chiu Y, Lin Y. Systematic review and meta-analysis of epidemiology of internet addiction. Neurosci Biobehav Rev 2020;118:612-622. [doi: 10.1016/j.neubiorev.2020.08.013] [Medline: 32853626]
5. Ding K, Shen Y, Liu Q, Li H. The effects of digital addiction on brain function and structure of children and adolescents: a scoping review. Healthcare 2023;12(1):15 [FREE Full text] [doi: 10.3390/healthcare12010015] [Medline: 38200921]
6. Rumpf H, Achab S, Billieux J, Bowden-Jones H, Carragher N, Demetrovics Z, et al. Including gaming disorder in the ICD-11: the need to do so from a clinical and public health perspective. J Behav Addict 2018;7(3):556-561 [FREE Full text] [doi: 10.1556/2006.7.2018.59] [Medline: 30010410]
7. Kuss DJ, Griffiths MD, Karila L, Billieux J. Internet addiction: a systematic review of epidemiological research for the last decade. Curr Pharm Des 2014;20(25):4026-4052. [doi: 10.2174/13816128113199990617] [Medline: 24001297]
8. Greenfield DN. Treatment considerations in internet and video game addiction: a qualitative discussion. Child Adolesc Psychiatr Clin N Am 2018;27(2):327-344. [doi: 10.1016/j.chc.2017.11.007] [Medline: 29502754]
9. Geng Y, Gu J, Wang J, Zhang R. Smartphone addiction and depression, anxiety: the role of bedtime procrastination and self-control. J Affect Disord 2021;293:415-421. [doi: 10.1016/j.jad.2021.06.062] [Medline: 34246950]
10. Gagné JP, Radomsky AS. Beliefs about losing control, obsessions, and caution: an experimental investigation. Behav Res Ther 2020;126:103574. [doi: 10.1016/j.brat.2020.103574] [Medline: 32045733]
11. Mei S, Yau YHC, Chai J, Guo J, Potenza MN. Problematic internet use, well-being, self-esteem and self-control: data from a high-school survey in China. Addict Behav 2016;61:74-79 [FREE Full text] [doi: 10.1016/j.addbeh.2016.05.009] [Medline: 27249805]
12. Elhai JD, Rozgonjuk D, Liu T, Yang H. Fear of missing out predicts repeated measurements of greater negative affect using experience sampling methodology. J Affect Disord 2020;262:298-303. [doi: 10.1016/j.jad.2019.11.026] [Medline: 31733919]
13. Rozgonjuk D, Sindermann C, Elhai JD, Montag C. Fear of missing out (FoMO) and social media's impact on daily-life and productivity at work: Do Whatsapp, Facebook, Instagram, and Snapchat use disorders mediate that association? Addict Behav 2020;110:106487. [doi: 10.1016/j.addbeh.2020.106487] [Medline: 32674020]
14. Li R, Fu W, Liang Y, Huang S, Xu M, Tu R. Exploring the relationship between resilience and internet addiction in Chinese college students: the mediating roles of life satisfaction and loneliness. Acta Psychol 2024;248:104405 [FREE Full text] [doi: 10.1016/j.actpsy.2024.104405] [Medline: 39067239]
15. Liao Y, Huang T, Lin S, Wu C, Chang K, Hsieh S, et al. Mediating role of resilience in the relationships of physical activity and mindful self-awareness with peace of mind among college students. Sci Rep 2023;13(1):10386 [FREE Full text] [doi: 10.1038/s41598-023-37416-2] [Medline: 37369802]
16. Beutel ME, Brähler E, Glaesmer H, Kuss DJ, Wölfling K, Müller KW. Regular and problematic leisure-time internet use in the community: results from a German population-based survey. Cyberpsychol Behav Soc Netw 2011;14(5):291-296. [doi: 10.1089/cyber.2010.0199] [Medline: 21067277]
17. Lindenberg K, Kindt S, Szász-Janocha C. Effectiveness of cognitive behavioral therapy-based intervention in preventing gaming disorder and unspecified internet use disorder in adolescents: a cluster randomized clinical trial. JAMA Netw Open 2022;5(2):e2148995 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.48995] [Medline: 35179587]
18. Yeun YR, Han SJ. Effects of psychosocial interventions for school-aged children's internet addiction, self-control and self-esteem: meta-analysis. Healthc Inform Res 2016;22(3):217-230 [FREE Full text] [doi: 10.4258/hir.2016.22.3.217] [Medline: 27525163]
19. Li S, Wu Q, Tang C, Chen Z, Liu L. Exercise-based interventions for internet addiction: neurobiological and neuropsychological evidence. Front Psychol 2020;11:1296 [FREE Full text] [doi: 10.3389/fpsyg.2020.01296] [Medline: 32670157]
20. Liu J, Nie J, Wang Y. Effects of group counseling programs, cognitive behavioral therapy, and sports intervention on internet addiction in East Asia: a systematic review and meta-analysis. Int J Environ Res Public Health 2017;14(12):1470 [FREE Full text] [doi: 10.3390/ijerph14121470] [Medline: 29182549]
21. Zhang Y, Li G, Liu C, Chen H, Guo J, Shi Z. Mixed comparison of interventions for different exercise types on students with internet addiction: a network meta-analysis. Front Psychol 2023;14:1111195 [FREE Full text] [doi: 10.3389/fpsyg.2023.1111195] [Medline: 37303910]
22. Stevens MWR, King DL, Dorstyn D, Delfabbro PH. Cognitive-behavioral therapy for internet gaming disorder: a systematic review and meta-analysis. Clin Psychol Psychother 2019;26(2):191-203. [doi: 10.1002/cpp.2341] [Medline: 30341981]
23. Kim MS, Kim WJ, Khera AV, Kim JY, Yon DK, Lee SW, et al. Association between adiposity and cardiovascular outcomes: an umbrella review and meta-analysis of observational and Mendelian randomization studies. Eur Heart J 2021;42(34):3388-3403 [FREE Full text] [doi: 10.1093/eurheartj/ehab454] [Medline: 34333589]
24. Papatheodorou S, Evangelou E. Umbrella reviews: what they are and why we need them. Methods Mol Biol 2022;2345:135-146. [doi: 10.1007/978-1-0716-1566-9_8] [Medline: 34550588]

XSL•FO
RenderX

25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Zhang X, Zhang J, Zhang K, Ren J, Lu X, Wang T, et al. Effects of different interventions on internet addiction: a meta-analysis of random controlled trials. J Affect Disord 2022;313:56-71. [doi: 10.1016/j.jad.2022.06.013] [Medline: 35760188]
27. Kim J, Lee S, Lee D, Shim S, Balva D, Choi KH, et al. Psychological treatments for excessive gaming: a systematic review and meta-analysis. Sci Rep 2022;12(1):20485 [FREE Full text] [doi: 10.1038/s41598-022-24523-9] [Medline: 36443408]
28. Chang CH, Chang YC, Yang L, Tzang RF. The comparative efficacy of treatments for children and young adults with internet addiction/internet gaming disorder: an updated meta-analysis. Int J Environ Res Public Health 2022;19(5):2612 [FREE Full text] [doi: 10.3390/ijerph19052612] [Medline: 35270305]
29. Andrade ALM, Di Girolamo Martins G, Scatena A, Lopes FM, de Oliveira WA, Kim HS, et al. The effect of psychosocial interventions for reducing co-occurring symptoms of depression and anxiety in individuals with problematic internet use: a systematic review and meta-analysis. Int J Ment Health Addict 2022:1-22 [FREE Full text] [doi: 10.1007/s11469-022-00846-6] [Medline: 35677712]
30. Lu GL, Ding YM, Zhang YM, Huang HT, Liang YP, Chen CR. The correlation between mobile phone addiction and coping style among Chinese adolescents: a meta-analysis. Child Adolesc Psychiatry Ment Health 2021;15(1):60 [FREE Full text] [doi: 10.1186/s13034-021-00413-2] [Medline: 34654451]
31. Goslar M, Leibetseder M, Muench HM, Hofmann SG, Laireiter AR. Treatments for internet addiction, sex addiction and compulsive buying: a meta-analysis. J Behav Addict 2020;9(1):14-43 [FREE Full text] [doi: 10.1556/2006.2020.00005] [Medline: 32359229]
32. Liu S, Xiao T, Yang L, Loprinzi PD. Exercise as an alternative approach for treating smartphone addiction: a systematic review and meta-analysis of random controlled trials. Int J Environ Res Public Health 2019;16(20):3912 [FREE Full text] [doi: 10.3390/ijerph16203912] [Medline: 31618879]
33. Chun J, Shim H, Kim S. A meta-analysis of treatment interventions for internet addiction among Korean adolescents. Cyberpsychol Behav Soc Netw 2017;20(4):225-231. [doi: 10.1089/cyber.2016.0188] [Medline: 28263660]
34. Augner C, Vlasak T, Aichhorn W, Barth A. Tackling the 'digital pandemic': the effectiveness of psychological intervention strategies in problematic internet and smartphone use—a meta-analysis. Aust N Z J Psychiatry 2022;56(3):219-229. [doi: 10.1177/00048674211042793] [Medline: 34477009]
35. Augner C, Vlasak T, Aichhorn W, Barth A. Psychological online interventions for problem gambling and gambling disorder—a meta-analytic approach. J Psychiatr Res 2022;151:86-94. [doi: 10.1016/j.jpsychires.2022.04.006] [Medline: 35472684]
36. Winkler A, Dörsing B, Rief W, Shen Y, Glombiewski JA. Treatment of internet addiction: a meta-analysis. Clin Psychol Rev 2013;33(2):317-329. [doi: 10.1016/j.cpr.2012.12.005] [Medline: 23354007]
37. Danielsen PA, Mentzoni RA, Låg T. Treatment effects of therapeutic interventions for gaming disorder: a systematic review and meta-analysis. Addict Behav 2024;149:107887 [FREE Full text] [doi: 10.1016/j.addbeh.2023.107887] [Medline: 37826910]
38. Jiang Y, Liu T, Qin D, Wang Z, He X, Chen Y. Effects of non-pharmacological interventions on youth with internet addiction: a systematic review and meta-analysis of randomized controlled trials. Front Psychiatry 2023;14:1327200 [FREE Full text] [doi: 10.3389/fpsyt.2023.1327200] [Medline: 38274427]
39. Li Z, Xia X, Sun Q, Li Y. Exercise intervention to reduce mobile phone addiction in adolescents: a systematic review and meta-analysis of randomized controlled trials. Front Psychol 2023;14:1294116 [FREE Full text] [doi: 10.3389/fpsyg.2023.1294116] [Medline: 38192396]
40. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
41. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
42. Martin KB, Bednarz JM, Aromataris EC. Interventions to control children's screen use and their effect on sleep: a systematic review and meta-analysis. J Sleep Res 2021;30(3):e13130. [doi: 10.1111/jsr.13130] [Medline: 32567219]
43. Young KS. Cognitive behavior therapy with internet addicts: treatment outcomes and implications. Cyberpsychol Behav 2007;10(5):671-679. [doi: 10.1089/cpb.2007.9971] [Medline: 17927535]
44. Zhang W, Xu R. Effect of exercise intervention on internet addiction and autonomic nervous function in college students. Biomed Res Int 2022;2022:5935353 [FREE Full text] [doi: 10.1155/2022/5935353] [Medline: 36105927]
45. Pu Y, Liu Y, Qi Y, Yan Z, Zhang X, He Q. Five-week of solution-focused group counseling successfully reduces internet addiction among college students: a pilot study. J Behav Addict 2023;12(4):964-971 [FREE Full text] [doi: 10.1556/2006.2023.00064] [Medline: 37966485]

XSL•FO
RenderX

46. Montgomery P, Grant S, Hopewell S, Macdonald G, Moher D, Michie S, et al. Protocol for CONSORT-SPI: an extension for social and psychological interventions. Implement Sci 2013;8:99 [FREE Full text] [doi: 10.1186/1748-5908-8-99] [Medline: 24004579]
47. Martin KB, Bednarz JM, Aromataris EC. Interventions to control children's screen use and their effect on sleep: a systematic review and meta-analysis. J Sleep Res 2021;30(3):e13130 [FREE Full text] [doi: 10.1111/jsr.13130] [Medline: 32567219]
48. Ding K, Li H. Digital addiction intervention for children and adolescents: a scoping review. Int J Environ Res Public Health 2023;20(6):4777 [FREE Full text] [doi: 10.3390/ijerph20064777] [Medline: 36981687]
49. Kim S, Yim H, Jo S, Jung K, Lee K, Park M. The effects of group cognitive behavioral therapy on the improvement of depression and anxiety in adolescents with problematic internet use. Soa Chongsonyon Chongsin Uihak 2018;29(2):73-79 [FREE Full text] [doi: 10.5765/jkacap.2018.29.2.73] [Medline: 32595298]
50. Malinauskas R, Malinauskiene V. A meta-analysis of psychological interventions for internet/smartphone addiction among adolescents. J Behav Addict 2019;8(4):613-624 [FREE Full text] [doi: 10.1556/2006.8.2019.72] [Medline: 31891316]
51. Fisher G, Harrison T. Substance Abuse: Information for School Counselors, Social Workers, Therapists, and Counselors. Boston, MA: Allyn & Bacon; 1996.
52. Ik-Soo O, Chul K. Meta-analysis on the effects of the prevention and intervention programs for internet addiction. J Korean Assoc Inf Educ 2009;13(4):529-537 [FREE Full text]
53. Kim JU. The effect of a R/T group counseling program on the internet addiction level and self-esteem of Internet addiction university students. Forum Soc Econ 2015;6:41-49.
54. Park SY, Kim SM, Roh S, Soh M, Lee SH, Kim H, et al. The effects of a virtual reality treatment program for online gaming addiction. Comput Methods Programs Biomed 2016;129:99-108. [doi: 10.1016/j.cmpb.2016.01.015] [Medline: 26860055]
55. Park SM. A meta-analysis on the effects of adolescent internet addiction group counseling program in Korea. Korean J Couns Psychother 2009;21:607-623 [FREE Full text]
56. Rosenbaum S, Vancampfort D, Steel Z, Newby J, Ward PB, Stubbs B. Physical activity in the treatment of post-traumatic stress disorder: a systematic review and meta-analysis. Psychiatry Res 2015;230(2):130-136. [doi: 10.1016/j.psychres.2015.10.017] [Medline: 26500072]
57. Herrera JJ, Fedynska S, Ghasem PR, Wieman T, Clark PJ, Gray N, et al. Neurochemical and behavioural indices of exercise reward are independent of exercise controllability. Eur J Neurosci 2016;43(9):1190-1202 [FREE Full text] [doi: 10.1111/ejn.13193] [Medline: 26833814]
58. Alves RL, Cardoso BRL, Ramos IPR, Oliveira BDS, Dos Santos ML, de Miranda AS, et al. Physical training improves exercise tolerance, cardiac function and promotes changes in neurotrophins levels in chagasic mice. Life Sci 2019;232:116629. [doi: 10.1016/j.lfs.2019.116629] [Medline: 31276687]
59. Lin X, Zhou J, Dong B. Effect of different levels of exercise on telomere length: a systematic review and meta-analysis. J Rehabil Med 2019;51(7):473-478 [FREE Full text] [doi: 10.2340/16501977-2560] [Medline: 31093683]
60. Dong G, Zhou H, Zhao X. Impulse inhibition in people with internet addiction disorder: electrophysiological evidence from a Go/NoGo study. Neurosci Lett 2010;485(2):138-142. [doi: 10.1016/j.neulet.2010.09.002] [Medline: 20833229]
61. Luijten M, Schellekens AF, Kühn S, Machielse MWJ, Sescousse G. Disruption of reward processing in addiction : an image-based meta-analysis of functional magnetic resonance imaging studies. JAMA Psychiatry 2017;74(4):387-398. [doi: 10.1001/jamapsychiatry.2016.3084] [Medline: 28146248]
62. Irvine MA, Worbe Y, Bolton S, Harrison NA, Bullmore ET, Voon V. Impaired decisional impulsivity in pathological videogamers. PLoS One 2013;8(10):e75914 [FREE Full text] [doi: 10.1371/journal.pone.0075914] [Medline: 24146789]
63. Xu L, Wu L, Geng X, Wang Z, Guo X, Song K, et al. A review of psychological interventions for internet addiction. Psychiatry Res 2021;302:114016. [doi: 10.1016/j.psychres.2021.114016] [Medline: 34087672]
64. Kim JU. A reality therapy group counseling program as an Internet addiction recovery method for college students in Korea. Int J Real Ther 2007;26:3-9.
65. Karimyar Jahromi M, Mosallanejad L. The impact of reality therapy on metacognition, stress and hope in addicts. Glob J Health Sci 2014;6(6):281-287 [FREE Full text] [doi: 10.5539/gjhs.v6n6p281] [Medline: 25363118]
66. Du Y, Jiang W, Vance A. Longer term effect of randomized, controlled group cognitive behavioural therapy for internet addiction in adolescent students in Shanghai. Aust N Z J Psychiatry 2010;44(2):129-134. [doi: 10.3109/00048670903282725] [Medline: 20113301]
67. Zewde EA, Tolossa T, Tiruneh SA, Azanaw MM, Yitbarek GY, Admasu FT, et al. Internet addiction and its associated factors among African high school and university students: systematic review and meta-analysis. Front Psychol 2022;13:847274 [FREE Full text] [doi: 10.3389/fpsyg.2022.847274] [Medline: 35386896]

Abbreviations

AMSTAR 2: Assessment of Multiple Systematic Reviews version 2
CBT: cognitive behavioral therapy
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
IA: internet addiction

XSL•FO
RenderX

IGD: internet gaming disorder
IIAP: integrated internet addiction prevention
MD: mean difference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by A Schwartz; submitted 18.04.24; peer-reviewed by M Marti-Vilar, A Rizvi; comments to author 01.08.24; revised version received 25.08.24; accepted 28.11.24; published 11.02.25.

Please cite as:
Lu P, Qiu J, Huang S, Wang X, Han S, Zhu S, Ning Y, Zeng FF, Yuan Y
Interventions for Digital Addiction: Umbrella Review of Meta-Analyses
J Med Internet Res 2025;27:e59656
URL: https://www.jmir.org/2025/1/e59656
doi:10.2196/59656
PMID:

©Peng Lu, Jiamin Qiu, Shiqi Huang, Xinman Wang, Shasha Han, Sui Zhu, Youjing Ning, Fang-fang Zeng, Yuan Yuan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Uses of Virtual Care in Primary Care: Scoping Review

Payal Agarwal[1,2], MSc, MD; Glenn George Fletcher[3], MSc; Karishini Ramamoorthi[1], MSc; Xiaomei Yao[3,4], MSc, MD; Onil Bhattacharyya[1,2], MD, PhD

[1]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[3]Program in Evidence-Based Care, Department of Oncology, McMaster University, Hamilton, ON, Canada

[4]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Onil Bhattacharyya, MD, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville Street
Toronto, ON, M5S 1B2
Canada
Phone: 1 4163236400
Email: Onil.bhattacharyya@wchospital.ca

Abstract

Background: The COVID-19 pandemic catalyzed an uptake in virtual care. However, the rapid shift left unanswered questions about the impact of virtual care on the quality of primary care and its appropriateness and effectiveness. Moving forward, health care providers require guidance on how best to use virtual care to support high-quality primary care.

Objective: This study aims to identify and summarize clinical studies and systematic reviews comparing virtual care and in-person care in primary care, with a focus on how virtual care can support key clinical functions such as triage, medical assessment and treatment, counseling, and rehabilitation in addition to the management of particular conditions.

Methods: We conducted a scoping review following an established framework. Comprehensive searches were performed across the following databases: Embase, MEDLINE, PsycInfo, Emcare, and Cochrane Database of Systematic Reviews. Other well-known websites were also searched. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines were followed. Articles were selected by considering article type, language, care provided, intervention, mode of care delivery, and sample size.

Results: A total of 13,667 articles were screened, and 219 (1.6%) articles representing 170 studies were included in the review. Of the 170 studies included, 142 (83.5%) were primary studies, and 28 (16.5%) were systematic reviews. The studies were grouped by functions of primary care, including triage (16/170, 9.4%), medical assessment and treatment of particular conditions (63/170, 37.1%), rehabilitation (17/170, 10%), and counseling (74/170, 43.5%). The studies suggested that many primary care functions could appropriately be conducted virtually. Virtual rehabilitation was comparable to in-person care and virtual counseling was found to be equally effective as in-person counseling in several contexts. Some of the studies indicated that many general primary care issues could be resolved virtually without the need for any additional follow-up, but data on diagnostic accuracy were limited. Virtual triage is clinically appropriate and led to fewer in-person visits, but overall impact on efficiency was unclear. Many studies found that virtual care was more convenient for many patients and provided care equivalent to in-person care for a range of conditions. Studies comparing appropriate antibiotic prescription between virtual and in-person care found variable impact by clinical condition. Studies on virtual chronic disease management observed variability in impact on overall disease control and clinical outcomes.

Conclusions: Virtual care can be safe and appropriate for triage and seems equivalent to in-person care for counseling and some rehabilitation services; however, further studies are needed to determine specific contexts or medical conditions where virtual care is appropriate for diagnosis, management outcomes, and other functions of primary care. Virtual care needs to be adapted to fit a new set of patient and provider workflows to demonstrate positive impacts on experience, outcomes, and costs of care.

(J Med Internet Res 2025;27:e55007) doi:10.2196/55007

XSL•FO
RenderX

KEYWORDS

primary care; telemedicine; telehealth; virtual care; virtual health; virtual medicine; remote consultation; telephone consultation; video consultations; medical informatics; digital health; digital technology; digital intervention; COVID-19; SARS-CoV-2; coronavirus infections; PRISMA

Introduction

Background

The COVID-19 pandemic led to a significant uptake of virtual care use, with increasing recognition that virtual care will continue to play a key role in Canada's health care system [1]. Before the onset of the pandemic, 1.2% of all primary care visits in Canada occurred virtually [2]. By contrast, between January 2021 and March 2022, there was a significant uptake of virtual care use in primary care, with 38% of all family physician appointments in Canada being conducted virtually [3]. Similar trends are also observed globally. In the United States, the proportion of primary care physicians using virtual care increased from 5% before the pandemic to 46% during the pandemic [4,5]. In addition, economic evaluations of virtual care suggest that virtual care may be more cost-effective in some circumstances, reducing the cost per episode of care and the cost to attend the appointment (eg, travel and parking costs) compared to in-person care [6]. These cost savings and conveniences for patients contribute to their positive perceptions about using virtual care. By contrast, some studies have found that neither patients nor providers perceive the quality of virtual visits to be better than that of in-person visits [7-10]. In particular, the rapid shift in care delivery without care redesign has raised concerns about appropriateness, effectiveness, and equity [3]. Furthermore, as health systems return to prepandemic visit volumes, there are concerns that these new care modalities could increase workload, leading to increased burnout and eventually reduced staff resource capacity [11].

A broad range of stakeholders have suggested that patients should be the focal point of decision-making and that virtual care should be built into streamlined workflows that cover all aspects of primary care, with meaningful incentives to drive system impact [12]. However, there is little empirical evidence to support putting these principles into practice. Reviews to date looking at outcomes related to virtual primary care have shown mixed or uncertain outcomes, with minimal guidance to drive care decisions [13]. Nevertheless, strong continued interest from patients suggests that virtual care can provide benefits in supporting quality primary care [13]. Studies have suggested that high-quality, sustainable models for primary virtual care require mechanisms to support triage and shared decision-making on the use of different modalities in clinical practice [12,14]. Consensus guidelines have also suggested that the use of virtual care in primary care should be based on multiple factors, including clinical appropriateness, patient preferences, and equity [15,16]. Primary care involves many functions, such as triaging incoming requests, diagnosing acute conditions, managing chronic conditions, providing counseling, and supporting rehabilitation (among others). Despite the ongoing use of virtual care by many providers, questions remain about which primary care functions are best suited to this mode of care.

This paper is based on an evidence summary developed by the Program in Evidence-Based Care (PEBC) at McMaster University at the request of the Population Health and Value-Based Health Systems portfolio of Ontario Health (a government agency in Canada) as one of the inputs into a guidance document, "Clinically Appropriate Use of Virtual Care in Primary Care" [16]. This paper focuses on a subset of the results of this broader review to explore various primary care use cases for virtual care.

Objectives

The main objective of this review was to identify and summarize clinical studies of virtual care use in primary care. Of particular interest were studies indicating the circumstances when synchronous virtual care (primarily videoconference or telephone interactions) was likely to be equivalent, superior, or inferior to in-person care to inform decisions about appropriate modes of interaction with patients. Rather than make definitive statements on the benefits or harms of virtual care or the value of virtual-only primary care, we have organized the findings in ways that highlight potential benefits across a range of primary care functions to indicate which functions of primary care may be most suitable for virtual care. We have also suggested ways of conducting and reviewing future studies.

Methods

Literature Search and Screening

The literature search strategy and screening were performed according to systematic review methodology. The review protocol was developed by the coauthors and approved at the PEBC and by the sponsor (Population Health and Value-Based Health Systems, Ontario Health) before the commencement of the review; a post hoc decision was made to analyze and present the results as a scoping review rather than as a systematic review. A scoping review includes a predefined protocol and systematic approach to literature searches, but instead of focusing on the rigor of included studies, it reviews the literature to identify or map the state of knowledge and gaps and the research conducted, as well as identify further research needs [17-19]. This scoping review follows the guidelines formulated by Arksey and O'Malley [17] and aligns with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) statement [20]. The PRISMA checklist was also used to support reporting (refer to Multimedia Appendix 1) Research team members, using an inductive process, grouped the results into different categories, representing distinct use cases in primary care. While a formal assessment of the strength or quality of evidence was not conducted for each study—an aspect that differentiates a scoping review from a systematic review—characteristics such as study design (eg, randomized controlled trial [RCT] and prospective or retrospective comparative study), sample size, and type of comparison were extracted. In interpreting and discussing the

XSL•FO
RenderX

results, we considered factors such as study size, trial design, consistency of results, and whether the studies addressed very narrow, specific issues or broader topics.

Definitions

The following subsections provide definitions and explanations of the key terms used throughout this paper.

Virtual Care

The term *virtual care* was defined as synchronous remote interaction between patients and clinicians to replace all or a portion of face-to-face (in-person) interactions. Care is considered synchronous when the patient and clinician are both present at the same time, and there is 2-way communication without delays in response (ie, in real time). It can include both video and telephone interactions. By contrast, asynchronous communication does not take place in real time. Asynchronous communication was included only if it was used as part of synchronous communication.

Primary Care

Primary care was defined as "the first point of contact between a patient and the health care system and includes illness prevention, health promotion, diagnosis, treatment, and rehabilitation and counseling" [21]. It included care by family or general practice physicians and nurse practitioners, general practice pediatricians and geriatricians, midwives, psychologists, psychotherapists, social workers, pharmacists, and physiotherapists. Care provided by dentists, psychiatrists, or medical specialists normally seen only by referral or in a hospital setting was excluded. The definition encompassed a variety of health care providers to reflect the broadest range of staff typically found in a multidisciplinary primary care team.

Triage

Triage refers to the initial decision process within primary care practice upon first patient contact, resulting in the allocation of patients to either in-person or virtual family practice appointments or, in some cases, to administrative assistance or other types of care. Articles categorized under triage included those that discuss signposting—a strategy designed to direct patients to the right provider at the right time and the right place at the first point of contact with primary care [22]. As the first point of contact for many patients, triage is often considered a core function of primary care.

Research Question

To explore the effectiveness of virtual care in primary care settings, this study addresses the following research question:

When using virtual modalities to seek or deliver primary care, are there differences in the outcomes of interest (either in general or for specific medical conditions or purposes of appointments) between

- synchronous virtual care by telephone or video compared to exclusive in-person care;
- synchronous virtual care by telephone or video plus asynchronous care (SMS text or secure messaging and email) compared to exclusive in-person care?

Outcomes of Interest

Outcomes of interest from a patient perspective included equity and accessibility (race, ethnicity, socioeconomic status, urban or rural or remote residence, age, gender, computer literacy, and mobility), availability of appointments (time, location, and wait time to get an appointment), and on-time appointments. Outcomes relevant to multiple stakeholders—patients, care providers, and health systems—included disease stability, improvement or deterioration, complications, or death; satisfactory resolution of the reason for visit; subsequent in-person visit to primary care practitioner because virtual care was not appropriate for assessment or treatment; referral to specialists, emergency department use, or hospital admissions; rates of blood tests, clinical laboratory tests, and diagnostic imaging; prescription of antibiotics or other medications; overall health of patients of primary practice team; and continuity of care.

Search Strategy

Embase, MEDLINE, PsycInfo, Emcare, and Cochrane Database of Systematic Reviews databases using the Ovid platform were searched from 2014 to November 24, 2021, using terms related to the concepts of *virtual care* and *primary care* (refer to Multimedia Appendix 2 [6,23-240] for the full search strategy). Due to time constraints for the project, a pragmatic decision was made to begin the database search in 2014. Therefore, publications before 2014 were not screened and only included in the literature review if cited elsewhere. Websites that the PEBC routinely consults to identify guidelines and reviews were also searched from 2015 to March 2022 (Multimedia Appendix 2); we did not search the websites of health care organizations devoted to primary care or specific diseases. Articles cited in other publications were included, regardless of publication year, if they met the other inclusion criteria.

Inclusion and Exclusion Criteria and Screening of Primary Studies

The inclusion and exclusion criteria are outlined in Textbox 1. Articles were screened based on these criteria, which included categories such as article type, language, care provided, intervention, mode of delivery, and sample size.

A post hoc decision was made to include noncomparative studies that did not initially meet the inclusion criteria, but only if they were categorized as relating to triage. This adjustment was made because triage is a core component of first-contact care provided by primary care practitioners.

A review of the titles and abstracts of the primary literature, followed by data extraction for studies meeting the specified criteria, was conducted by GGF, a professional health research methodologist at the PEBC at McMaster University. The coauthors were consulted in cases of uncertainty. An independent audit of the extracted data was conducted by Jilian Sing (refer to the Acknowledgments section). Discrepancies were noted and addressed.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Study type: randomized controlled trial; other comparative studies (virtual vs in-person care)
- Language: abstract or full text in English
- Care provided: primary care study (refer to the definition of *primary care* in the Definitions subsection); care provided in a continuous primary care practice or a clinic that offered both walk-in and virtual care
- Intervention: virtual care provided by the same clinician or primary care team responsible for in-person care
- Mode of care delivery: synchronous virtual care (telephone or video), which replaced in-person care, including replacing of a subset of in-person appointments
- Sample size: sample size ≥30 (this is often considered sufficient for the central limit theorem to hold true, enabling the detection of differences between 2 study groups) [241]

Exclusion criteria

- Article type: trial registries or other study or review protocols without published results; editorials or commentaries; noncomparative surveys or questionnaires about patient or clinician experience; case studies; case series without a comparison group; conference abstracts
- Language: not English
- Care provided: study investigated the role of education, materials, lifestyle adaptation, exercise, diet, and so on, and the main study comparison was not virtual versus in-person delivery of these interventions; replacement of in-person care provided by 1 professional or treatment team or unit with virtual or in-person care provided by a different professional or team
- Intervention: study related to virtual-only practices (eg, “walk-in” telephone clinics or hotlines); app or remote patient monitoring;
- Mode of care delivery: SMS Text, email, or other asynchronous interventions, unless as a supplement to synchronous virtual care components
- Sample size: sample size <30

Systematic Reviews

Publications described as being systematic reviews or meta-analyses by their authors (primarily in the title or abstract) were included if they reported on primary studies meeting the inclusion criteria (Textbox 1). Reviews were also considered to be systematic reviews if the authors reported the databases searched, search strategy, and inclusion and exclusion criteria; the results section included a list of the reviewed studies along with extracted data; and the review was not otherwise classified by the authors. Systematic reviews on virtual care that passed initial screening but predominantly included studies not meeting the inclusion criteria were excluded; these reviews were used only to identify additional primary studies not found in the database search. Only reviews published in English were considered, with no restrictions placed on the publication date of individual studies.

Results

Overview

The database search identified 14,916 citations. In addition, 26 studies were identified through website searches, and 69 studies were found from the reference lists of other publications (Figure 1). The included studies were categorized into primary care functions, including triaging of incoming requests, diagnosis of a range of conditions through general primary care, rehabilitation, and counseling. A breakdown of the 170 included studies is presented in Table 1. Data for each study have been extracted and are reported in the tables in Multimedia Appendix 2. Primary studies on triage (24 publications representing 14 studies) are summarized in Table S1 in Multimedia Appendix 2 [23-46], and systematic reviews on triage are summarized in Table S2 in Multimedia Appendix 2 [47,48]. As the number of studies addressing specific types of virtual care was very limited, we decided not to subdivide results further by this category.

For studies on virtual care other than counseling, there were 64 publications representing 62 primary studies and 19 publications representing 18 systematic reviews (Tables S3-S13 in Multimedia Appendix 2 [6,49-130]). These have been subdivided by the categories listed in Table 1 and consisted of 11 RCTs and 18 possibly prospective nonrandomized trials; the rest were retrospective studies that generally used chart review or registry data. Only a few of the studies used multivariate analysis to control for possibly confounding factors. Table S3 [49-68] and Table S4 [6,69-73] in Multimedia Appendix 2 summarize studies and systematic reviews, respectively, in general primary care. Table S5 in Multimedia Appendix 2 [74-82] summarizes studies conducted during COVID-19–related restrictions; Table S6 in Multimedia Appendix 2 [83-90] and Table S7 in Multimedia Appendix 2 [91,92] summarize studies and systematic reviews of minor infections, respectively; and Table S8 in Multimedia Appendix 2 [93-96] summarizes studies of COVID-19 diagnosis and management. Table S9 in Multimedia Appendix 2 [97-106] and Table S10 in Multimedia Appendix 2 [107,108] summarize studies and reviews of virtual care in chronic disease management, respectively, while Table S11 in Multimedia Appendix 2 [109-113] reports on the use of virtual care in medical abortion. Clinical studies and systematic reviews in rehabilitation are covered in Table S12 in Multimedia Appendix 1 [114-121] and Table S13 in Multimedia Appendix 2 [107,122-130], respectively.

XSL•FO
RenderX

Primary counseling studies (102 publications representing 66 studies) are summarized in Table S14 in Multimedia Appendix 2 [131-232], while systematic reviews on counseling (n=8) are reported in Table S15 in Multimedia Appendix 2 [233-240]. The number of studies on counseling was slightly more than the number of other studies combined. Most of the counseling studies (50/65, 77%) were RCTs, providing stronger and higher-quality evidence on this topic.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram showing the number of studies identified, screened, assessed for eligibility, and included in the final analysis. *Studies were excluded for the following reasons: did not meet our definition of primary care or virtual care (3134/11,026, 28.42%); an additional intervention was being studied (2076/11,026, 18.83%); not full publications of completed studies (protocols, trial registrations, or conference abstracts; 2060/11,026, 18.68%); no comparison group for the outcomes of interest (1318/11,026, 11.95%); focused on the education of clinicians or interprofessional consultation (864/11,026, 7.84%); asynchronous interventions or not involving patient-clinician contact (821/11,026, 7.44%); case studies or <30 patients per group (362/11,026, 3.28%); letters, editorials, or commentary (195/11,026, 1.77%); virtual reality (142/11,026, 1.29%); and other (54/11,026, 0.49%).

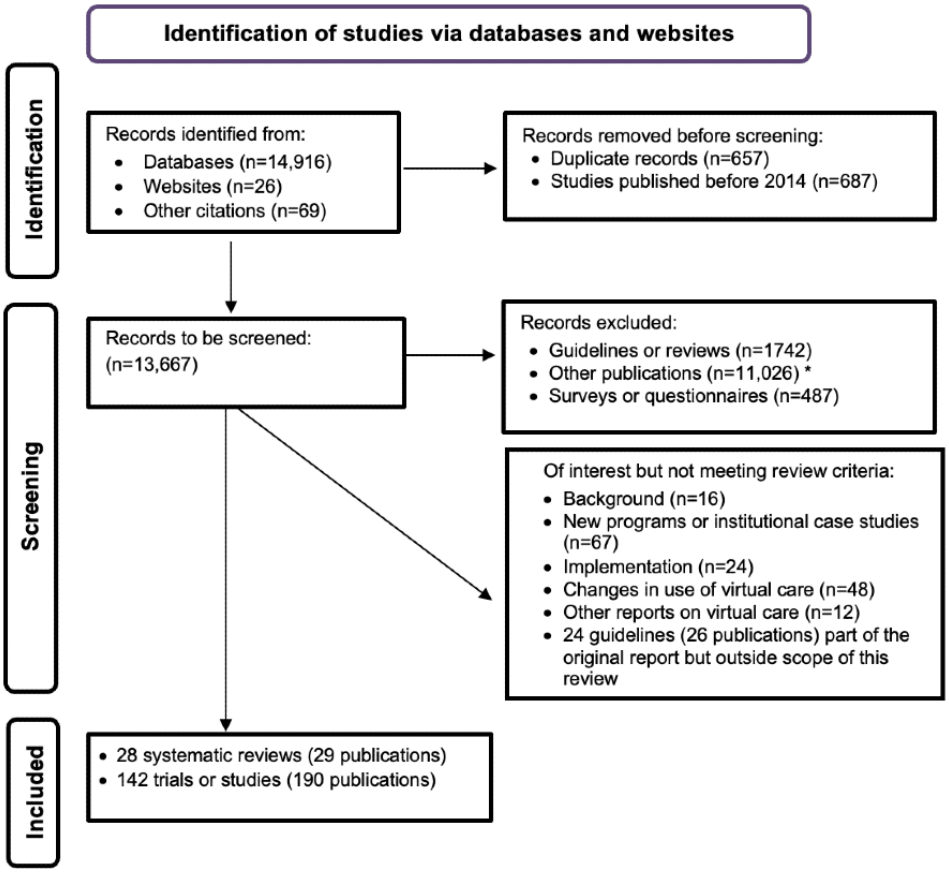

Table 1. Breakdown of included studies by primary care function.

Reason for consultation	Studies, n (%)		
	Primary studies (n=142)	Systematic reviews (n=28)	Total (n=170)
Triage	14 (9.9)	2 (7.1)	16 (9.4)
Medical assessment and treatment	54 (38)	9 (32.1)	63 (37.1)
General primary care	19 (35.2)	5 (55.6)	24 (38.1)
Primary care during COVID-19–related restrictions	9 (16.7)	0 (0)	9 (14.3)
Minor infections	8 (14.8)	2 (22.2)	10 (15.9)
COVID-19 management	4 (7.4)	0 (0)	4 (6.3)
Chronic disease management	9 (16.7)	2 (22.2)	11 (17.5)
Medical abortion	5 (3.2)	0 (0)	5 (7.9)
Rehabilitation	8 (5.6)	9 (32.1)	17 (10)
Counseling	66 (46.5)	8 (28.6)	74 (43.5)

XSL•FO
RenderX

Triage

Overview

The 14 primary studies involving triage (Table S1 in Multimedia Appendix 2 [23-46]) used various designs and often did not allow a direct comparison of virtual versus in-person care. Of these 14 studies, 10 (71%) were conducted in the United Kingdom, while 1 (7%) study each was conducted in Canada, Israel, and the Netherlands, 2 studies were conducted in Denmark. The 2 systematic reviews—1 (50%) on telephone triage [47] and 1 (50%) on remote triage [48]—are summarized in Table S2 in Multimedia Appendix 1.

Telephone Triage

The most common practice involved telephone triage by a nurse or other office staff and the assignment of patients to telephone consultation if appropriate and if the patient did not object. Both patient and disease characteristics determined the modality of care. Of the 16 studies on triage, only 2 (13%) were of this design [42,44]; most reports of this type were excluded because there was generally no comparison group. A variation used by 3 (19%) of the 16 studies was a telephone-first approach, where a general practitioner (GP) spoke by telephone with all patients (except those with administrative queries) and either addressed the issue during the call or scheduled an in-person visit [24-29,39]. Of the 16 studies, 2 (13%) involved triage by GPs in Denmark [43,46]. Of these 2 studies, 1 (50%) found that approximately 40% of the calls resulted in face-to-face consultation [43], while 1 (50%) reported medication prescription rates [46]. In 2 (13%) of the 16 studies, patients accessed the medical office using a web-based form, and patients with nonadministrative concerns were allocated to a GP who decided on the mode of contact [30,31,45].

The only RCT was the ESTEEM trial, and 7 (44%) of the 16 publications on triage covered various aspects of this trial [32-38]. Conducted in England, this trial, which used a combination of the aforementioned approaches, may serve as the most significant source of insights into triage. Patients (n=15,394) requesting same-day appointments were randomized to triage by a nurse, triage by a GP, or usual care. Estimated overall contact durations (triage+subsequent contacts on the same day) were 10.3 minutes for GP triage, 14.8 minutes for nurse triage, and 9.6 minutes for usual care. The number of deaths, emergency hospital admissions, and accident and emergency department attendances were not different [36]. Triage by a nurse practitioner resulted in fewer in-person visits than triage by practice nurses (odds ratio 0.19, 95% CI 0.07-0.49); nurse practitioners were more likely to definitively manage patients within the triage consultation [33].

Signposting

Another study from the United Kingdom [41] used signposting in 2 practices to reduce GP telephone workload; only patients for whom a physician visit was judged essential were given a telephone appointment. The remaining callers were referred to allied health professionals, alternative services, nonmedical staff, or self-help [41]. Approximately 40% to 60% of telephone appointments were eliminated. The proportion of essential consultations taking place at the 2 practices increased from 28.6% and 27.3% at baseline to 82.6% and 71.4%, respectively.

Emergency Assessment

A total of 4 (25%) out of 16 studies on triaging explored the use of virtual care to assist in triaging and managing emergency calls within health systems. Of the 16 studies on triaging, 1 (6%) looked at adding GP support to emergency calls, either in person or by telephone [40]. The study found that 8.1% of all calls to emergency services could be triaged to ambulance personnel plus GP support. Of these patients, 78% were not transported to hospital, although a transfer was more often avoided if the patient was assessed in person by the GP. The remaining study, which was conducted in British Columbia reported experience using HealthLink BC Emergency iDoctor-in-assistance [23]. It looked at patients who called an 811 nurse care navigation service. Callers who were triaged to seek care within 24 hours were referred to a physician by videoconferencing. Of the callers directed to videoconferencing with HealthLink BC Emergency iDoctor-in-assistance physicians, 33.8% were advised to attempt home treatment, 38.3% to contact a primary care physician within 1 week, 15% to attend an emergency department immediately, and 7.1% to contact their primary health care provider right away.

Medical Assessment and Treatment

General Primary Care

A total of 24 (38%) of the 63 studies on medical assessment and treatment were on general primary care. These studies in general primary care are summarized in Table S3 in Multimedia Appendix 2 [49-68] and systematic reviews in Table S4 in Multimedia Appendix 2 [6,69-73]. Of the 19 primary studies classified as general primary care, 15 (79%) involved general practice contacts for all or several reasons [49-58,60-65], while 1 (5%) reviewed those with acute deterioration within 3 days of a primary care visit [59] and found a slightly higher rate of self-referral to emergency services after telephone consultations compared to in-person visits.

Of the 15 studies involving general practice contacts for all or several reasons, 1 (7%) focused on the 11 most common illnesses managed via telehealth: sinusitis, upper respiratory infection, urinary tract infection, conjunctivitis, bronchitis, pharyngitis, influenza, cough, dermatitis, digestive symptoms (nausea, vomiting, or diarrhea), and ear pain [65]. This large retrospective study used insurance company claims and focused on the treatment of acute non-urgent conditions. It found that follow-up rates for both virtual and in-person care were similar, suggesting comparable clinical resolution of these issues through virtual modalities, although antibiotic use was higher and laboratory testing lower for the virtual care group [65].

Of the 2 RCTs that looked at a broad range of virtual primary care issues, 1 (50%) crossover study involved evaluations of patients by 2 physicians [49,50]: patients were randomized to either 1 virtual and 1 in-person assessment the same day or 2 in-person assessments. Follow-up for chronic diseases such as hypertension, elevated cholesterol level, and diabetes and consultations for acute diseases such as upper respiratory illness or sinusitis and musculoskeletal complaints were the most

XSL•FO
RenderX

common. The diagnostic agreements between the physicians were 84% between face-to-face and virtual visits and 80% between the 2 face-to-face visits. This study was among the few that directly compared the reliability of diagnosis for common conditions via in-person or virtual care.

The other RCT randomized patients to either a face-to-face appointment on that day or a callback that morning by a physician who either offered advice or arranged an in-person visit [51]. Telephone consultations were shorter (6.7 vs 8.2 min; *P*=.002) but resulted in slightly more follow-up consultations within 2 weeks (0.6 vs 0.4 consultations; *P*=.01), thus offsetting the time savings. Blood pressure was more frequently measured in person (13.3% vs 6.6%), but there were no significant differences in patient perceptions or other secondary outcomes.

Of the 24 studies on general primary care, 1 (4%) [67] compared video to telephone communication in emergency paramedicine response and found that video enhanced clinical evaluation 85% of the time; the odds ratio for emergency transport was 0.80 (95% CI 0.62-1.03).

Of the 9 systematic reviews on medical assessment and treatment, 5 (56%) were on virtual care in general without specifying any diseases or conditions [6,69-73]. The reviews by Totten et al [71-73] for the US Agency for Healthcare Research and Quality [71-73] concluded that telehealth consultations are effective in providing services or improving outcomes, although evidence is stronger for some applications. The reviews found evidence of effectiveness for counseling and the management of chronic conditions. However, overall impact, including cost-effectiveness, varied significantly based on how virtual care was integrated into the larger health service delivery ecosystem.

Primary Care During COVID-19–Related Restrictions

Of the 63 studies on medical assessment and treatment, 9 (14%) reported the use of virtual care due to COVID-19–related restrictions and are summarized in Table S5 in Multimedia Appendix 1 [74-82]. Of these 9 studies, 1 (11%) found that there was a rapid increase in the use of virtual care—from 31% in April 2019 to 90% in April 2020. Of the visits conducted virtually, 89% were by telephone and 1% by video [80]. Another study found that 71% of the video visits included visual observation–dependent findings that could not be assessed using a telephone; however, common barriers to using video visits were the lack of appropriate equipment and patient preferences [74]. Among the studies looking at primary care during the COVID-19 pandemic, there were variable results on the impact of age on the use of video visits: the study by Eberly et al [75] found that older individuals were less likely to use video visits, whereas Schenker et al [76] found that age was a significant positive predictor of having video visits.

Minor Infections

Of the 63 studies on medical assessment and treatment, 10 (16%) discussed antibiotic prescription (primary studies: n=8, 80%; Table S6 in Multimedia Appendix 1 [83-90]; systematic reviews: n=2, 20%; Table S7 in Multimedia Appendix 1 [91,92]). The 2 reviews did not have sufficient evidence to draw strong conclusions about the impact of virtual care on antibiotic prescribing but noted variable effects by condition, with higher rates for some conditions and lower rates for others. Of the 8 primary studies, 5 (62%) retrospective studies evaluated antibiotic prescribing for acute respiratory infections [83,85-88], while 1 (12%) evaluated antibiotic prescribing for urinary symptoms or infections [84]; in addition, 4 (50%) studies found similar or lower antibiotic prescribing rates with virtual compared to in-person care, while 1 (12%) reported higher antibiotic prescribing rates for telemedicine than in-person visits. This varied by condition within the same health care organization, with a chart review finding lower rates of antibiotic prescribing for sinusitis by SMS text and telephone than face-to-face visits and similar rates for urinary tract infections across all 3 modes [83,84]. By contrast, a review of claims data with matched populations showed higher rates of prescribing for acute respiratory infections for direct-to-consumer telemedicine versus urgent care or care provided by the patient's usual primary care provider [86]. A similar pattern was found in a study comparing retail health clinics with urgent care and primary care providers [65].

Some of the other general studies [55,63,65] also noted antibiotic prescribing as one of the outcomes. 1 of the 10 studies on minor infections (10%) compared the outcomes of patient-selected modality with the same provider, finding higher rates of antibiotic prescribing for in-person visits (10.6% for video vs 9.7% for telephone vs 13.5% for in-person visits) [55]. Of the 10 studies, 2 (20%) evaluated and found higher antibiotic prescribing rates with telephone consultations for patients with conjunctivitis, a condition that usually requires visual assessment [89,90].

COVID-19 Management

Of the 4 studies that reported on the assessment of patients with symptoms consistent with COVID-19 infection (Table S8 in Multimedia Appendix 2 [93-96]), 3 (75%) compared remote assessment (by telephone, video, or unspecified) with in-person care. Some studies conducted during the COVID-19 pandemic used antibiotic prescribing as one of the outcomes. Of these 4 studies, 1 (25%) [94] found that in-person visits involved more testing for influenza and higher antibiotic prescribing rates.

Chronic Disease Management

Of the 63 studies focused on medical assessment and treatment, 11 (18%) focused on chronic disease management (primary studies: n=9, 82%; Table S9 in Multimedia Appendix 1 [97-106]; systematic reviews: n=2, 18%; Table S10 in Multimedia Appendix 1 [107,108]). Of the 9 primary studies, 3 (33%) RCTs were conducted in patients with asthma [97-99], and all found higher rates of follow-up appointments in the telephone groups; 2 (22%) retrospective studies found similar diabetes control using either telephone [100] or videoconferencing [101] compared to face-to-face or usual care; and, by contrast, 1 (11%) retrospective study examining the use of telehealth by community health care workers to support Hispanic patients with low-income status with uncontrolled diabetes (defined by glycated hemoglobin level ≥9%) reported that patients were much less likely to achieve glycemic control with telephone visits alone compared to either exclusively in-person visits or a mix of in-person and virtual visits [102].

XSL•FO
RenderX

However, 1 in-person visit plus ≥1 telephone visit was worse (longer time to diabetes control) than 1 in-person visit alone [102], suggesting that adjustment for the confounders was not adequate.

Medical Abortion

Of the 63 studies on medical assessment and treatment, 5 (8%) explored the use of virtual care in medical abortion assessment or follow-up (Table S11 in Multimedia Appendix 1 [109-113]). These studies found no difference in successful abortion rates or adverse events, reported similar satisfaction, and suggested that providing care by telephone improved access compared to in-person care.

Rehabilitation

Rehabilitation studies (17/63, 27%) included those focusing on physiotherapy and cardiac rehabilitation. Of the 17 studies, 8 (47%) were primary studies (Table S12 in Multimedia Appendix 2 [114-121]), and 9 (53%) were systematic reviews (Table S13 in Multimedia Appendix 2 [107,122-130]). Specifically, of the 8 primary studies, 2 (25%) involved cardiac rehabilitation. The noninferiority RCT compared remotely monitored telerehabilitation to center-based programs and found similar benefits remote noninferior in terms of improvement in maximal oxygen consumption and health state [114]. The quasi-experimental study compared home-based videoconference rehabilitation to in-hospital rehabilitation; there was similar improvement in the 6-minute walk test, and the authors concluded that video rehabilitation was feasible [115].

Of the 8 primary studies, 2 RCTs (25%) [116,117] and 1 (12%) retrospective study [118] found that telerehabilitation after total knee or total hip and knee arthroplasty was noninferior or comparable to traditional therapy. Studies on movement assessment in children [119] and physiotherapy assessment [121] found 92% and 83% agreement between video and in-person assessment, respectively; it is unclear whether this is sufficient. A study on physical activity coaching [120] found no significant differences between control (no intervention), in-person, and telephone intervention.

Counseling Studies

Of the 8 systematic reviews on counseling (Table S15 in Multimedia Appendix 2 [233-240]), 3 (37%) focused on psychotherapy [233,234,236]; 1 (12%) explored interventions targeting smoking, nutrition, alcohol consumption, physical activity, and obesity [235]; 1 (12%) each focused on depression [237], posttraumatic stress disorder [238], and suicide [239]; and 1 (12%) investigated traumatic stress during a pandemic [240]. Overall, these reviews concluded that virtual care is an established medium for delivering counseling, with evidence suggesting equivalence to in-person care in many areas of counseling.

Most of the counseling studies (50/74, 68%; Table S14 in Multimedia Appendix 2 [131-232]) were RCTs. Of the 50 RCTs, 19 (38%) were designed as noninferiority studies. Many studies compared counseling by either telephone or videoconferencing to the same counseling provided in person (face-to-face). Of the 74 counseling studies, 3 (4%) [177,191-193] used real-time SMS text or chat conversations (all other studies involving SMS text-based communication were asynchronous and had been excluded). Cognitive behavioral therapy (CBT) was used in 18 (24%) of the 74 studies for treating depression, anxiety, pain, insomnia, eating disorders such as bulimia nervosa, and obsessive-compulsive disorder. Components of CBT may be present in some other studies that were described as behavioral studies, although they were not explicitly labeled as CBT. The overall evidence indicates that outcomes for telephone and video CBT are similar or noninferior to those for face-to-face CBT, although some differences for specific populations were noted. Other commonly used counseling types were problem-solving therapy, behavioral activation therapy, behavioral treatment interventions, prolonged exposure therapy, and cognitive processing therapy, with the latter 2 names used especially for posttraumatic stress disorder. For lifestyle-type interventions for weight loss, smoking cessation, and alcohol use, the type of counseling was less likely to fit within the traditional psychotherapy categories. While evidence for CBT is the greatest, the overall evidence suggests a role for telephone or video counseling in many other areas.

Discussion

Principal Findings

This scoping review synthesized studies that compared in-person and virtual care in primary care, with a mix of studies conducted before and during the COVID-19 pandemic. The studies were categorized into primary care functions, including triaging of incoming patient requests, medical assessment and treatment (including general primary care, minor infections, COVID-19 management, chronic disease management, and medical abortion), rehabilitation, and counseling. The review showed strong evidence that virtual care is equivalent to in-person care for counseling and comparable for some types of rehabilitation. The results suggest that virtual triaging is clinically appropriate and may lead to fewer in-person visits, especially when conducted by a physician or a nurse practitioner. However, for other primary care functions, outcomes from the included studies varied. Several studies demonstrated that many general primary care issues could be resolved virtually without the need for in-person follow-up, but only a few looked at diagnostic accuracy. A study that focused on diagnostic agreement between modalities for common primary care conditions found strong alignment [49,50]. Studies on chronic disease management observed variability in their benefit in using virtual care. Studies comparing appropriate antibiotic prescription between virtual and in-person care found variable impact depending on the clinical condition, whether patients chose a modality or were randomized to one, and the clinical context. A few of the studies measured the impact on emergency department visits or hospitalizations and found no difference, but they did not adjust for the type or severity of the illness [53,55]. Overall, the variation in impact may be a result of the relative effectiveness of the modality for a given condition or use case, but in some of the studies the differences may be driven by how the modality is selected, the practice context, and health care provider incentives within a health care organization.

XSL•FO
RenderX

The evidence for the overall experience and value of virtual care is evolving, but our review suggests some advantages for patients and unclear impacts on provider workload and system costs. This aligns with previous reviews that have found uncertain outcomes for virtual primary care as a whole [6,13]. In our review, most patients reported a similar overall experience between virtual and in-person interactions; however, certain benefits of virtual care were noted, including in terms of cost, travel time, convenience, waiting time, and access to more specialized care [39,42,69,70]. Therapeutic alliance was similar between in-person and virtual counseling in the studies reviewed. This contrasts with an earlier systematic review of virtual care in cancer, where some clinicians reported that they found it more difficult to comfort patients during vulnerable times, found virtual care more stressful, and had concerns about therapeutic alliance [242].

The included studies did not assess the impact on continuity of care or health equity, both of which are key components of primary care and patient experience. Other studies that aim to explore continuity of care after a primary care visit only do so by examining primary care as a whole; for example, the study by Reed et al [243] found a slightly higher percentage of in-person return visits and emergency department visits after video or telephone visits compared to in-person primary care visits, but the authors did not examine the differences based on the type of primary care function. Given that continuity of care is a defining characteristic of primary care and has been shown to be instrumental in driving primary care's positive impact on equity and quality of care, future studies should explore the impact of virtual care use on different primary care functions in the context of its impact on continuity of care [244,245]. In addition, the impact of virtual care on provider workload and system costs was less clear in our review. Telephone visits tended to be shorter than in-person interactions, but it was unclear how the addition of virtual care affected overall workloads. An initial telephone-based clinical triage by nonphysicians resulted in a higher proportion of visits being deemed essential by primary care physicians, suggesting that triage reduced low-value appointments with a physician [48]. Having said that, although the overall workload was not addressed, other studies in the literature highlight the value of virtual triage to physicians, notably that it can improve a clinician's experience by streamlining patient-clinician communication and could help clinicians better manage their time by reducing their administrative workload [246].

To our knowledge, this is one of the first comprehensive reviews to examine the appropriateness of virtual care in terms of key functions of primary care. Studies examining virtual care use in primary care tend to examine primary care as a whole; for example, the study by Reed et al [243] found that 50% of primary care visits used telemedicine (19% by video and 31% by telephone); however, the impact on these numbers based on primary care function was not studied. Similarly, previous reviews have also focused on examining virtual care use more broadly through categories such as cost and health care use. However, given the shift in practice since the COVID-19 pandemic, it is now more important to understand how and when in the clinical process virtual care should be used [6,13]. In addition, these reviews only included publications from 2020 onward and did not include RCTs, which are a focus of our review. Our review shows that there is significant variability in evidence across different clinical use cases in primary care, making it difficult to provide broad conclusions about the appropriateness of virtual primary care for specific conditions. Instead, our work provides evidence that the impact and outcomes of virtual modalities in primary care should be examined according to the various interaction points along a patient's journey; for example, there is strong evidence for equivalence between in-person and virtual care for counseling, and thus counseling could be more readily integrated into sustainable, high-quality primary care workflows. By contrast, the use of virtual care for situations that are likely to result in antibiotic prescribing requires further study to understand whether it is indeed equivalent to in-person care. Our review suggests that virtual triage can play a role in supporting primary care workflows, but more research is required to determine the best implementation model. A rapid review by Barnabe et al [247] highlighted some implications for health care organizations to consider before implementing virtual triage. These include ensuring that the platforms are locally based; that they are flexible, responsive, and tailored to regional circumstances; and that implementation is carried out, keeping partnerships and collaborations in mind [247]. These results contrast with those of other reviews on the use of virtual modalities in primary care, which analyzed clinical appropriateness and outcomes across all use cases and were generally unable to make strong recommendations to guide clinical care [15]. Several authors have suggested that a more nuanced approach, one that looks at different uses of primary care, might ultimately provide more insight into driving decisions around patient care and system planning [6,248,249].

Virtual care during the COVID-19 pandemic has expanded to cover a wide range of conditions, and our review suggests that there is still limited evidence to broadly define the optimal mode of interaction (in-person, video, telephone, or asynchronous contact); however, the evidence does provide some guidance. Our review suggests that virtual care is likely at least as good as in-person care for a wide range of primary care issues and should be offered as part of routine practice. Overall, for any given visit, the choice of modality remains a question of clinical judgment and shared decision-making between a patient and their care team. Patient factors, including access to technology, comfort with virtual visits, and maintaining a strong therapeutic relationship, must be considered when deciding the best modality for a given interaction [250]. Virtual care should be embedded in larger service models that allow for easy escalation to in-person follow-up when required. Enabling virtual triage can provide clinics with certain advantages, including improving overall access, better prioritization for the most appropriate modality, and determining whether any information should be provided in advance of the visit. Furthermore, providing counseling, self-management support, education, and ongoing monitoring virtually also offers advantages over regular in-person appointments, including improved convenience and similar outcomes for many patient populations. Virtual care could also be offered to support chronic disease management but may not work well for all patients. Options such as providing

XSL•FO
RenderX

smartphones, tablet devices, or computers and limited telephone or internet service were used in some of the studies for individuals with limited technology accessibility [139,152-154,167,195]. However, health organizations and systems would benefit from continued investment in understanding how to better integrate and optimize virtual care to achieve the quintuple aim of improving patient experience, improving population health, improving clinician well-being, reducing health care costs, and advancing health equity [251], using human-centered methods such as co-design, service design, and participatory research or other frameworks [252]. Now that the practice of virtual care is so widespread, ongoing research can help in understanding how it can be leveraged to improve access and quality of care [248].

Limitations

This review has several limitations. First, it was restricted to studies comparing virtual care to in-person care instead of care (regardless of delivery method) to an objective standard. Therefore, it is applicable only to clinical areas in which there is established in-person care for evaluation and treatment. While many of the included studies were designed to measure the noninferiority of virtual care to in-person care, in-person treatment should not necessarily be the gold standard, especially for counseling studies where effectiveness is often based on a validated tool to measure disease, symptom intensity, or patient-reported outcomes. Second, this is a scoping review rather than a systematic review; therefore, we did not evaluate the methodological quality of the studies or determine which studies might provide more reliable results. In addition, we did not include gray literature in our search to ensure the reliability of the evidence underpinning the recommendations provided in the paper. Having said that, although we supplemented our search using trusted websites, it is possible that the exclusion of gray literature resulted in underrepresented perspectives from policy makers and community organizations that are not captured in academic publications. This could mean that insights from the practical applications of real-world data could have been missed. Furthermore, articles were only screened by a single reviewer, which could have introduced bias in the results if personal biases impacted the interpretation of the literature. Although these limitations were mitigated through an audit conducted by another member of the team, the likelihood of errors is increased when only 1 reviewer is involved. Third, the study contexts varied substantially. Most were cross-sectional studies focused on the initial consultation or ongoing chronic disease management and did not assess the impact on continuity of care, a key component of primary care. Finally, given the time frame for the database search, it is possible that new literature has become available that contributes additional evidence to this topic. Nonetheless, the study demonstrates a novel approach to organizing the literature on virtual care, moving from the question of whether virtual care is as effective as in-person care for a given health problem to thinking about how virtual care supports the core functions of primary care. It argues for a different way of conducting studies in the future and provides a framework for organizing this research. Much of the included literature covers the period during the COVID-19 pandemic, when most ambulatory care was virtual in many jurisdictions. This context is not as broadly applicable to the current state of care. However, we believe that our findings still provide valuable insights into how virtual care can be leveraged in primary care and can suggest appropriate ways of structuring future studies to better understand when and how voice, video, and SMS text-based interactions are best used in clinical care.

Conclusions

This scoping review found no consistent differences in the appropriateness of care or patient outcomes between virtual and in-person care across the included studies. Virtual care is perceived as equivalent to in-person care for many common uses, despite some variation by condition. The overall impact of virtual care will likely depend on whether it is used for triage, diagnosis, treatment, or counseling and whether patients or providers determine the choice of modality. Moreover, we found that the overall value of virtual care for the health system, including potential cost savings, remains unclear. Our findings highlight that although the rates of virtual care use may decrease as concerns about COVID-19 infection lessen, for functions such as triage, counseling, and rehabilitation, virtual care may remain the norm. Further studies, incorporating options for patient and provider choice, are needed to determine the optimal use of virtual care from a resource and outcome perspective. This comprehensive review is one of the first that aims to understand how virtual care can be leveraged in the various components of routine primary care functions. Given the recent changes in practice, this evidence base will continue to grow, and timely reviews will be needed to keep up with the literature. These modalities are no longer just a temporary measure to get through the pandemic and now need to be adapted to fit a new set of patient and provider workflows to demonstrate positive impacts on experience, outcomes, and costs of care. Future reviews should focus on when to use virtual modes of communication in a patient's primary care journey.

Acknowledgments

The authors would like to thank Population Health and Value-Based Health Systems, Ontario Health, for funding this study. The funder played no part in the collection, analysis, or interpretation of the data; the writing of the report; or the decision to submit the paper for publication. The authors also thank Jilian Sing for screening and summarizing reviews and for supporting the data audit.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 99 KB - jmir_v27i1e55007_app1.pdf]

Multimedia Appendix 2
Search strategy and data extraction tables.
[DOC File , 528 KB - jmir_v27i1e55007_app2.doc]

References

1. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: a repeated cross-sectional study. CMAJ Open 2021 Feb 17;9(1):E107-E114 [FREE Full text] [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
2. Glazier RH, Green ME, Wu FC, Frymire E, Kopp A, Kiran T. Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada. CMAJ 2021 Feb 08;193(6):E200-E210 [FREE Full text] [doi: 10.1503/cmaj.202303] [Medline: 33558406]
3. Canadian Institute for Health Information. The expansion of virtual care in canada: New data and information. Ottawa, ON: CIHI; 2023.
4. Callaghan T, McCord C, Washburn D, Goidel K, Schmit C, Nuzhath T, et al. The changing nature of telehealth use by primary care physicians in the United States. J Prim Care Community Health 2022 Jul 06;13:21501319221110418 [FREE Full text] [doi: 10.1177/21501319221110418] [Medline: 35795898]
5. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022 Dec;49(4):517-530 [FREE Full text] [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
6. Hui D, Dolcine B, Loshak H. Approaches to evaluations of virtual care in primary care. Can J Health Technol 2022 Jan;2(1). [doi: 10.51731/cjht.2022.238]
7. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Mar 17;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20x101020]
8. Chan S, O'Riordan A, Appireddy R. Exploring the determinants and experiences of senior stroke patients with virtual care. Can J Neurol Sci 2021 Jan 27;48(1):87-93. [doi: 10.1017/cjn.2020.162] [Medline: 32713397]
9. Appireddy R, Khan S, Leaver C, Martin C, Jin A, Durafourt BA, et al. Home virtual visits for outpatient follow-up stroke care: cross-sectional study. J Med Internet Res 2019 Oct 07;21(10):e13734 [FREE Full text] [doi: 10.2196/13734] [Medline: 31593536]
10. Donelan K, Barreto EA, Sossong S, Michael C, Estrada JJ, Cohen AB, et al. Patient and clinician experiences with telehealth for patient follow-up care. Am J Manag Care 2019 Jan;25(1):40-44 [FREE Full text] [Medline: 30667610]
11. Shachak A, Alkureishi MA. Virtual care: a 'Zoombie' apocalypse? J Am Med Inform Assoc 2020 Nov 01;27(11):1813-1815 [FREE Full text] [doi: 10.1093/jamia/ocaa185] [Medline: 32940711]
12. Shaw J, Jamieson T, Agarwal P, Griffin B, Wong I, Bhatia RS. Virtual care policy recommendations for patient-centred primary care: findings of a consensus policy dialogue using a nominal group technique. J Telemed Telecare 2018 Oct;24(9):608-615. [doi: 10.1177/1357633X17730444] [Medline: 28945161]
13. Lachance CC, Severn M, Kim J. Evidence base for virtual primary care. Can J Health Technol 2023 May 30;3(5) [FREE Full text] [doi: 10.51731/cjht.2023.661]
14. Girdhari R, Krueger P, Wang R, Meaney C, Domb S, Larsen D, et al. Electronic communication between family physicians and patients: findings from a multisite survey of academic family physicians in Ontario. Can Fam Physician 2021 Jan 22;67(1):39-46 [FREE Full text] [doi: 10.46747/cfp.670139] [Medline: 33483396]
15. Nayakarathna R, Neilson H, MacDougall D, Cowling T. Virtual care use in primary care or specialty care settings. Can J Health Technol 2022 Aug 18;2(8). [doi: 10.51731/cjht.2022.421]
16. Clinically appropriate use of virtual care in primary care : guidance reference document. Ontario Health. 2022 Nov 1. URL: https://www.ontariohealth.ca/sites/ontariohealth/files/2022-11/ClinicallyAppropriateUseVirtualCarePrimaryCare.pdf [accessed 2023-11-02]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
19. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]

XSL•FO
RenderX

20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Primary care payment models in Ontario. Ontario Ministry of Health. URL: https://www.ontario.ca/page/primary-care-payment-models-ontario [accessed 2024-11-01]
22. Brunton L, Tazzyman A, Ferguson J, Hodgson D, Nelson PA. The challenges of integrating signposting into general practice: qualitative stakeholder perspectives on care navigation and social prescribing in primary care. BMC Prim Care 2022 Apr 01;23(1):66 [FREE Full text] [doi: 10.1186/s12875-022-01669-z] [Medline: 35365072]
23. Ho K, Lauscher HN, Stewart K, Abu-Laban RB, Scheuermeyer F, Grafstein E, et al. Integration of virtual physician visits into a provincial 8-1-1 health information telephone service during the COVID-19 pandemic: a descriptive study of HealthLink BC Emergency iDoctor-in-assistance (HEiDi). CMAJ Open 2021 Jun 15;9(2):E635-E641 [FREE Full text] [doi: 10.9778/cmajo.20200265] [Medline: 34131026]
24. Newbould J, Exley J, Ball S, Corbett J, Pitchforth E, Roland M. GPs' and practice staff's views of a telephone first approach to demand management: a qualitative study in primary care. Br J Gen Pract 2019 Apr 23;69(682):e321-e328. [doi: 10.3399/bjgp19x702401]
25. Newbould J, Ball S, Abel G, Barclay M, Brown T, Corbett J, et al. A 'telephone first' approach to demand management in English general practice: a multimethod evaluation. Health Serv Deliv Res 2019 May;7(17):1-158. [doi: 10.3310/hsdr07170] [Medline: 31063292]
26. Ball SL, Newbould J, Corbett J, Exley J, Pitchforth E, Roland M. Qualitative study of patient views on a 'telephone-first' approach in general practice in England: speaking to the GP by telephone before making face-to-face appointments. BMJ Open 2018 Dec 31;8(12):e026197 [FREE Full text] [doi: 10.1136/bmjopen-2018-026197] [Medline: 30598491]
27. Newbould J, Abel G, Ball S, Corbett J, Elliott M, Exley J, et al. Evaluation of telephone first approach to demand management in English general practice: observational study. BMJ 2017 Sep 27;358:j4197 [FREE Full text] [doi: 10.1136/bmj.j4197] [Medline: 28954741]
28. Miller D, Loftus AM, O'Boyle PJ, McCloskey M, O'Kelly J, Mace D, et al. Impact of a telephone-first consultation system in general practice. Postgrad Med J 2019 Nov;95(1129):590-595. [doi: 10.1136/postgradmedj-2019-136557] [Medline: 31326942]
29. Jiwa M, Mathers N, Campbell M. The effect of GP telephone triage on numbers seeking same-day appointments. Br J Gen Pract 2002 May;52(478):390-391 [FREE Full text] [Medline: 12014537]
30. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
31. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Mar 19;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
32. Holt TA, Fletcher E, Warren F, Richards S, Salisbury C, Calitri R, et al. Telephone triage systems in UK general practice: analysis of consultation duration during the index day in a pragmatic randomised controlled trial. Br J Gen Pract 2016 Feb 25;66(644):e214-e218. [doi: 10.3399/bjgp16x684001]
33. Varley A, Warren FC, Richards SH, Calitri R, Chaplin K, Fletcher E, et al. The effect of nurses' preparedness and nurse practitioner status on triage call management in primary care: a secondary analysis of cross-sectional data from the ESTEEM trial. Int J Nurs Stud 2016 Jun;58:12-20 [FREE Full text] [doi: 10.1016/j.ijnurstu.2016.02.001] [Medline: 27087294]
34. Warren FC, Calitri R, Fletcher E, Varley A, Holt TA, Lattimer V, et al. Exploring demographic and lifestyle associations with patient experience following telephone triage by a primary care doctor or nurse: secondary analyses from a cluster randomised controlled trial. BMJ Qual Saf 2015 Sep 18;24(9):572-582 [FREE Full text] [doi: 10.1136/bmjqs-2015-003937] [Medline: 25986572]
35. Calitri R, Warren FC, Wheeler B, Chaplin K, Fletcher E, Murdoch J, et al. Distance from practice moderates the relationship between patient management involving nurse telephone triage consulting and patient satisfaction with care. Health Place 2015 Jul;34:92-96 [FREE Full text] [doi: 10.1016/j.healthplace.2015.04.002] [Medline: 25982703]
36. Campbell JL, Fletcher E, Britten N, Green C, Holt T, Lattimer V, et al. The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial). Health Technol Assess 2015 Feb;19(13):1-212, vii [FREE Full text] [doi: 10.3310/hta19130] [Medline: 25690266]
37. Campbell JL, Fletcher E, Britten N, Green C, Holt TA, Lattimer V, et al. Telephone triage for management of same-day consultation requests in general practice (the ESTEEM trial): a cluster-randomised controlled trial and cost-consequence analysis. The Lancet 2014 Nov;384(9957):1859-1868. [doi: 10.1016/s0140-6736(14)61058-8]
38. Murdoch J, Varley A, Fletcher E, Britten N, Price L, Calitri R, et al. Implementing telephone triage in general practice: a process evaluation of a cluster randomised controlled trial. BMC Fam Pract 2015 Apr 10;16(1):47 [FREE Full text] [doi: 10.1186/s12875-015-0263-4] [Medline: 25887747]

XSL•FO
RenderX

39. Lawless M, Wright E, Davidson J. A collaborative approach to improving patient access in general practice: impact of three different pilot schemes in 12 general practices in Greenwich. London J Prim Care (Abingdon) 2016 Jun 10;8(4):56-65 [FREE Full text] [doi: 10.1080/17571472.2016.1173946] [Medline: 28250835]
40. Villarreal M, Leach J, Ngianga-Bakwin K, Dale J. Can a partnership between general practitioners and ambulance services reduce conveyance to emergency care? Emerg Med J 2017 Jul 18;34(7):459-465. [doi: 10.1136/emermed-2015-204924] [Medline: 28420688]
41. Siddiqui F, Sidhu B, Tahir MA. Using 'active signposting' to streamline general practitioner workload in two London-based practices. BMJ Open Qual 2017 Oct 21;6(2):e000146 [FREE Full text] [doi: 10.1136/bmjoq-2017-000146] [Medline: 29450292]
42. Elliott M, Jones S, Johnson C, Wallace C. What are the benefits of nurse-led triage in primary care? Prim Health Care 2020. [doi: 10.7748/phc.2020.e1607]
43. Huibers L, Moth G, Carlsen AH, Christensen MB, Vedsted P. Telephone triage by GPs in out-of-hours primary care in Denmark: a prospective observational study of efficiency and relevance. Br J Gen Pract 2016 Jul 18;66(650):e667-e673. [doi: 10.3399/bjgp16x686545]
44. Jansen T, Hek K, Schellevis FG, Kunst AE, Verheij RA. Income-related differences in out-of-hours primary care telephone triage using national registration data. Emerg Med J 2021 Jun 14;38(6):460-466. [doi: 10.1136/emermed-2020-209649] [Medline: 33853937]
45. Eccles A, Hopper M, Turk A, Atherton H. Patient use of an online triage platform: a mixed-methods retrospective exploration in UK primary care. Br J Gen Pract 2019 Mar 25;69(682):e336-e344. [doi: 10.3399/bjgp19x702197]
46. Christensen MB, Nørøxe KB, Moth G, Vedsted P, Huibers L. Drug prescriptions in Danish out-of-hours primary care: a 1-yearpopulation-based study. Scand J Prim Health Care 2016 Dec 02;34(4):453-458 [FREE Full text] [doi: 10.1080/02813432.2016.1248622] [Medline: 27804314]
47. Lake R, Georgiou A, Li J, Li L, Byrne M, Robinson M, et al. The quality, safety and governance of telephone triage and advice services - an overview of evidence from systematic reviews. BMC Health Serv Res 2017 Aug 30;17(1):614 [FREE Full text] [doi: 10.1186/s12913-017-2564-x] [Medline: 28854916]
48. Rushton S, Boggan JC, Lewinski AA, Gordon AM, Shoup JP, Van Voorhees E, et al. Effectiveness of Remote Triage: A Systematic Review. Washington, DC: Department of Veterans Affairs; 2019.
49. Dixon RF, Stahl JE. A randomized trial of virtual visits in a general medicine practice. J Telemed Telecare 2009;15(3):115-117. [doi: 10.1258/jtt.2009.003003] [Medline: 19364890]
50. Stahl JE, Dixon RF. Acceptability and willingness to pay for primary care videoconferencing: a randomized controlled trial. J Telemed Telecare 2010 Apr 12;16(3):147-151. [doi: 10.1258/jtt.2009.090502] [Medline: 20386035]
51. Sutcliffe L. Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices. Prim Health Care 2004 Jun;14(5):8. [doi: 10.7748/phc.14.5.8.s14]
52. McKinstry B, Hammersley V, Burton C, Pinnock H, Elton R, Dowell J, et al. The quality, safety and content of telephone and face-to-face consultations: a comparative study. Qual Saf Health Care 2010 Aug 29;19(4):298-303. [doi: 10.1136/qshc.2008.027763] [Medline: 20430933]
53. Gujral K, Scott JY, Ambady L, Dismuke-Greer CE, Jacobs J, Chow A, et al. A primary care telehealth pilot program to improve access: associations with patients' health care utilization and costs. Telemed J E Health 2022 May 01;28(5):643-653. [doi: 10.1089/tmj.2021.0284] [Medline: 34559017]
54. Llorian ER, Mason G. Healthcare utilization and telemedicine: an evaluation using linked administrative data from Manitoba. J Telemed Telecare 2021 Jan 17;29(4):282-290. [doi: 10.1177/1357633x20981227]
55. Reed M, Huang J, Graetz I, Muelly E, Millman A, Lee C. Treatment and follow-up care associated with patient-scheduled primary care telemedicine and in-person visits in a large integrated health system. JAMA Netw Open 2021 Nov 01;4(11):e2132793 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.32793] [Medline: 34783828]
56. Graetz I, Huang J, Muelly E, Gopalan A, Lee C, Reed ME. Patient choice of telemedicine increases timeliness of primary care visits. AMIA Annu Symp Proc 2020;2020:502-503 [FREE Full text] [Medline: 33936423]
57. Reed ME, Huang J, Graetz I, Lee C, Muelly E, Kennedy C, et al. Patient characteristics associated with choosing a telemedicine visit vs office visit with the same primary care clinicians. JAMA Netw Open 2020 Jun 01;3(6):e205873 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5873] [Medline: 32585018]
58. Ryskina KL, Shultz K, Zhou Y, Lautenbach G, Brown RT. Older adults' access to primary care: gender, racial, and ethnic disparities in telemedicine. J Am Geriatr Soc 2021 Oct;69(10):2732-2740 [FREE Full text] [doi: 10.1111/jgs.17354] [Medline: 34224577]
59. Cecil E, Bottle A, Majeed A, Aylin P. Factors associated with potentially missed acute deterioration in primary care: cohort study of UK general practices. Br J Gen Pract 2021 Feb 01;71(708):e547-e554. [doi: 10.3399/bjgp.2020.0986]
60. Gonzalez F, Cimadevila B, Garcia-Comesaña J, Cerqueiro S, Andion E, Prado J, et al. Telephone consultation in primary care: a retrospective two-year observational analysis of a public healthcare system. J Health Org Manag 2018 Mar 26;32(2):321-337. [doi: 10.1108/jhom-08-2017-0201]
61. Ross AA, Yap TL, Nest JV, Martin K, Edie AH. Increasing primary care access close to home for residents of remote communities in Northern Alberta. Healthc Q 2016 Oct 31;19(3):61-66. [doi: 10.12927/hcq.2016.24863] [Medline: 27808026]

XSL•FO
RenderX

62. Bernstein P, Ko KJ, Israni J, Cronin AO, Kurliand MM, Shi JM, et al. Urgent and non-emergent telehealth care for seniors: findings from a multi-site impact study. J Telemed Telecare 2023 Aug;29(7):566-575 [FREE Full text] [doi: 10.1177/1357633X211004321] [Medline: 33866894]
63. Lovell T, Albritton J, Dalto J, Ledward C, Daines W. Virtual vs traditional care settings for low-acuity urgent conditions: an economic analysis of cost and utilization using claims data. J Telemed Telecare 2019 Jul 25;27(1):59-65. [doi: 10.1177/1357633x19861232]
64. Ohta M, Ohira Y, Uehara T, Keira K, Noda K, Hirukawa M, et al. How accurate are first visit diagnoses using synchronous video visits with physicians? Telemed J E Health 2017 Feb;23(2):119-129. [doi: 10.1089/tmj.2015.0245] [Medline: 27351424]
65. Gordon AS, Adamson WC, DeVries AR. Virtual visits for acute, nonurgent care: a claims analysis of episode-level utilization. J Med Internet Res 2017 Feb 17;19(2):e35 [FREE Full text] [doi: 10.2196/jmir.6783] [Medline: 28213342]
66. Grech CK, Laux MA, Burrows HL, Macy ML, Pomeranz ES. Pediatric emergency department resource utilization among children with primary care clinic contact in the preceding 2 days: a cross-sectional study. J Pediatr 2017 Sep;188:245-51.e2. [doi: 10.1016/j.jpeds.2017.06.007] [Medline: 28690005]
67. Abrashkin KA, Washko JD, Li T, Berkowitz J, Poku A, Zhang J, et al. Video or telephone? A natural experiment on the added value of video communication in community paramedic responses. Ann Emerg Med 2021 Jan;77(1):103-109. [doi: 10.1016/j.annemergmed.2020.04.026] [Medline: 32534834]
68. Al Mazrouei N, Ibrahim RM, Al Meslamani AZ, Abdel-Qader DH, Mohamed Ibrahim O. Virtual pharmacist interventions on abuse of over-the-counter medications during COVID-19 versus traditional pharmacist interventions. J Am Pharm Assoc (2003) 2021 May;61(3):331-339 [FREE Full text] [doi: 10.1016/j.japh.2021.02.003] [Medline: 33676838]
69. Carrillo de Albornoz S, Sia KL, Harris A. The effectiveness of teleconsultations in primary care: systematic review. Fam Pract 2022 Jan 19;39(1):168-182 [FREE Full text] [doi: 10.1093/fampra/cmab077] [Medline: 34278421]
70. Gray C, Mason J, Loshak H. An overview of direct-to-patient virtual visits in Canada. Can J Health Technol 2021 Jun;1(6). [doi: 10.51731/cjht.2021.80]
71. Totten AM, Hansen RN, Wagner J, Stillman L, Ivlev I, Davis-O'Reilly C, et al. Telehealth for Acute and Chronic Care Consultations [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2019.
72. Totten AM, McDonagh MS, Wagner JH. The evidence base for telehealth: reassurance in the face of rapid expansion during the COVID-19 pandemic. Agency for Healthcare Research and Quality. 2020. URL: https://effectivehealthcare.ahrq.gov/products/telehealth-expansion/white-paper [accessed 2022-01-31]
73. Totten AM, Womack DM, Eden KB, McDonagh M, Griffin JC, Grusing S, et al. Telehealth: mapping the evidence for patient outcomes from systematic reviews. Agency for Healthcare Research and Quality. 2016. URL: https://www.drugsandalcohol.ie/32090/ [accessed 2022-01-31]
74. Schifeling CH, Shanbhag P, Johnson A, Atwater RC, Koljack C, Parnes BL, et al. Disparities in video and telephone visits among older adults during the COVID-19 pandemic: cross-sectional analysis. JMIR Aging 2020 Nov 10;3(2):e23176 [FREE Full text] [doi: 10.2196/23176] [Medline: 33048821]
75. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SA, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640. [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
76. Schenker RB, Laguna MC, Odisho AY, Okumura MJ, Burnett H. Are we reaching everyone? A cross-sectional study of telehealth inequity in the COVID-19 pandemic in an urban academic pediatric primary care clinic. Clin Pediatr (Phila) 2022 Jan;61(1):26-33 [FREE Full text] [doi: 10.1177/00099228211045809] [Medline: 34514898]
77. Lopez Segui F, Hernandez Guillamet G, Pifarré Arolas H, Marin-Gomez FX, Ruiz Comellas A, Ramirez Morros AM, et al. Characterization and identification of variations in types of primary care visits before and during the COVID-19 pandemic in Catalonia: big data analysis study. J Med Internet Res 2021 Sep 14;23(9):e29622 [FREE Full text] [doi: 10.2196/29622] [Medline: 34313600]
78. van der Velden AW, Bax EA, Bongard E, Munck Aabenhus R, Anastasaki M, Anthierens S, et al. Primary care for patients with respiratory tract infection before and early on in the COVID-19 pandemic: an observational study in 16 European countries. BMJ Open 2021 Jul 29;11(7):e049257 [FREE Full text] [doi: 10.1136/bmjopen-2021-049257] [Medline: 34326052]
79. Brown CL, Montez K, Amati JB, Simeonsson K, Townsend JD, Orr CJ, et al. Impact of COVID-19 on pediatric primary care visits at four academic institutions in the Carolinas. Int J Environ Res Public Health 2021 May 27;18(11):5734 [FREE Full text] [doi: 10.3390/ijerph18115734] [Medline: 34071783]
80. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021 Jan 17;71(704):e166-e177. [doi: 10.3399/bjgp.2020.0948]
81. Sigurdsson EL, Blondal AB, Jonsson JS, Tomasdottir MO, Hrafnkelsson H, Linnet K, et al. How primary healthcare in Iceland swiftly changed its strategy in response to the COVID-19 pandemic. BMJ Open 2020 Dec 07;10(12):e043151 [FREE Full text] [doi: 10.1136/bmjopen-2020-043151] [Medline: 33293329]

XSL•FO
RenderX

82. Mohamed Ibrahim O, Ibrahim RM, Abdel-Qader DH, Al Meslamani AZ, Al Mazrouei N. Evaluation of telepharmacy services in light of COVID-19. Telemed J E Health 2021 Jun 01;27(6):649-656. [doi: 10.1089/tmj.2020.0283] [Medline: 33030986]
83. Penza KS, Murray MA, Myers JF, Furst JW, Pecina JL. Management of acute sinusitis via e-visit. Telemed J E Health 2021 May 10;27(5):532-536. [doi: 10.1089/tmj.2020.0047] [Medline: 32522103]
84. Murray MA, Penza KS, Myers JF, Furst JW, Pecina JL. Comparison of eVisit management of urinary symptoms and urinary tract infections with standard care. Telemed J E Health 2020 May 01;26(5):639-644. [doi: 10.1089/tmj.2019.0044] [Medline: 31313978]
85. Shi Z, Mehrotra A, Gidengil CA, Poon SJ, Uscher-Pines L, Ray KN. Quality of care for acute respiratory infections during direct-to-consumer telemedicine visits for adults. Health Aff (Millwood) 2018 Dec;37(12):2014-2023 [FREE Full text] [doi: 10.1377/hlthaff.2018.05091] [Medline: 30633682]
86. Ray KN, Shi Z, Gidengil CA, Poon SJ, Uscher-Pines L, Mehrotra A. Antibiotic prescribing during pediatric direct-to-consumer telemedicine visits. Pediatrics 2019 May 08;143(5):e20182491 [FREE Full text] [doi: 10.1542/peds.2018-2491] [Medline: 30962253]
87. Ray KN, Martin JM, Wolfson D, Schweiberger K, Schoemer P, Cepullio C, et al. Antibiotic prescribing for acute respiratory tract infections during telemedicine visits within a pediatric primary care network. Acad Pediatr 2021 Sep;21(7):1239-1243. [doi: 10.1016/j.acap.2021.03.008] [Medline: 33741531]
88. Ewen E, Willey VJ, Kolm P, McGhan WF, Drees M. Antibiotic prescribing by telephone in primary care. Pharmacoepidemiol Drug Saf 2015 Feb 31;24(2):113-120. [doi: 10.1002/pds.3686] [Medline: 25079292]
89. Frost HM, Sebastian T, Durfee J, Jenkins TC. Ophthalmic antibiotic use for acute infectious conjunctivitis in children. J AAPOS 2021 Dec;25(6):350.e1-350.e7 [FREE Full text] [doi: 10.1016/j.jaapos.2021.06.006] [Medline: 34737083]
90. Penza KS, Murray MA, Myers JF, Maxson J, Furst JW, Pecina JL. Treating pediatric conjunctivitis without an exam: an evaluation of outcomes and antibiotic usage. J Telemed Telecare 2018 Aug 28;26(1-2):73-78. [doi: 10.1177/1357633x18793031]
91. Bakhit M, Baillie E, Krzyzaniak N, van Driel M, Clark J, Glasziou P, et al. Antibiotic prescribing for acute infections in synchronous telehealth consultations: a systematic review and meta-analysis. BJGP Open 2021 Sep 08;5(6):BJGPO.2021.0106. [doi: 10.3399/bjgpo.2021.0106]
92. Han SM, Greenfield G, Majeed A, Hayhoe B. Impact of remote consultations on antibiotic prescribing in primary health care: systematic review. J Med Internet Res 2020 Nov 09;22(11):e23482 [FREE Full text] [doi: 10.2196/23482] [Medline: 33031045]
93. Phillips JC, Lord RW, Davis SW, Burton AA, Kirk JK. Comparing telehealth to traditional office visits for patient management in the COVID-19 pandemic: a cross-sectional study in a respiratory assessment clinic. J Telemed Telecare 2021 Feb 01;29(5):374-381. [doi: 10.1177/1357633x21990197]
94. Tarn DM, Hintz C, Mendez-Hernandez E, Sawlani SP, Bholat MA. Using virtual visits to care for primary care patients with COVID-19 symptoms. J Am Board Fam Med 2021 Feb 23;34(Supplement):S147-S151. [doi: 10.3122/jabfm.2021.s1.200241]
95. Riese A, Kelly JM, Chu TC, Heinly A, Kamath S, Golova N, et al. Visits for possible COVID-19 in a pediatric primary care practice early in the pandemic. R I Med J (2013) 2021 Aug 02;104(6):43-48 [FREE Full text] [Medline: 34323879]
96. Irving G, Lawson D, Tinsley A, Parr H, Whittaker C, Jones H, et al. Evaluation of a 'drop box' doorstep assessment service to aid remote assessments for COVID-19 in general practice. BMJ Open Qual 2021 Mar 29;10(1):e001081 [FREE Full text] [doi: 10.1136/bmjoq-2020-001081] [Medline: 33781992]
97. Gruffydd-Jones K, Hollinghurst S, Ward S, Taylor G. Targeted routine asthma care in general practice using telephone triage. Br J Gen Pract 2005 Dec;55(521):918-923 [FREE Full text] [Medline: 16378560]
98. Pinnock H, Adlem L, Gaskin S, Harris J, Snellgrove C, Sheikh A. Accessibility, clinical effectiveness, and practice costs of providing a telephone option for routine asthma reviews: phase IV controlled implementation study. Br J Gen Pract 2007 Sep;57(542):714-722 [FREE Full text] [Medline: 17761059]
99. Pinnock H, Bawden R, Proctor S, Wolfe S, Scullion J, Price D, et al. Accessibility, acceptability, and effectiveness in primary care of routine telephone review of asthma: pragmatic, randomised controlled trial. BMJ 2003 Mar 01;326(7387):477-479 [FREE Full text] [doi: 10.1136/bmj.326.7387.477] [Medline: 12609944]
100. Al Harthi T, Anwar H, Al Lawati A, Al Shuriqi F, Al Rashdi F, Al Mahrouqi A, et al. The impact of Covid-19 on diabetes care in Muscat governorate: a retrospective cohort study in primary care. J Prim Care Community Health 2021 Oct 30;12:21501327211051930 [FREE Full text] [doi: 10.1177/21501327211051930] [Medline: 34719302]
101. Lu AD, Gunzburger E, Glorioso TJ, Smith WB2, Kenney RR, Whooley MA, et al. Impact of longitudinal virtual primary care on diabetes quality of care. J Gen Intern Med 2021 Sep;36(9):2585-2592 [FREE Full text] [doi: 10.1007/s11606-020-06547-x] [Medline: 33483815]
102. Turner BJ, Liang Y, Ramachandran A, Poursani R. Telephone or visit-based community health worker care management for uncontrolled diabetes mellitus: a longitudinal study. J Community Health 2020 Dec 29;45(6):1123-1131 [FREE Full text] [doi: 10.1007/s10900-020-00849-1] [Medline: 32472457]

XSL•FO
RenderX

103. Wickström H, Öien RF, Midlöv P, Anderberg P, Fagerström C. Pain and analgesics in patients with hard-to-heal ulcers: using telemedicine or standard consultations. J Wound Care 2021 Jun 01;30(Sup6):S23-S32. [doi: 10.12968/jowc.2021.30.sup6.s23]
104. Wickström HL, Öien RF, Fagerström C, Anderberg P, Jakobsson U, Midlöv PJ. Comparing video consultation with inperson assessment for Swedish patients with hard-to-heal ulcers: registry-based studies of healing time and of waiting time. BMJ Open 2018 Feb 15;8(2):e017623 [FREE Full text] [doi: 10.1136/bmjopen-2017-017623] [Medline: 29449288]
105. Eibl JK, Gauthier G, Pellegrini D, Daiter J, Varenbut M, Hogenbirk JC, et al. The effectiveness of telemedicine-delivered opioid agonist therapy in a supervised clinical setting. Drug Alcohol Depend 2017 Jul 01;176:133-138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.01.048] [Medline: 28535455]
106. Rivera J, Shcherbakova N, Vala C, Capoccia K. Community pharmacists' interventions and documentation during medication therapy management encounters delivered face-to-face versus via telephone: the devil is in the details. Res Social Adm Pharm 2020 Oct;16(10):1447-1451. [doi: 10.1016/j.sapharm.2019.12.020] [Medline: 31889640]
107. Ho C, Argáez C. Telehealth-Delivered Opioid Agonist Therapy for the Treatment of Adults with Opioid Use Disorder: Review of Clinical Effectiveness, Cost-Effectiveness, and Guidelines [Internet]. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2018.
108. Barth J, Nickel F, Kolominsky-Rabas PL. Diagnosis of cognitive decline and dementia in rural areas - a scoping review. Int J Geriatr Psychiatry 2018 Mar;33(3):459-474. [doi: 10.1002/gps.4841] [Medline: 29314221]
109. Grossman D, Grindlay K. Safety of medical abortion provided through telemedicine compared with in person. Obstet Gynecol 2017 Oct;130(4):778-782. [doi: 10.1097/AOG.0000000000002212] [Medline: 28885427]
110. Grossman DA, Grindlay K, Buchacker T, Potter JE, Schmertmann CP. Changes in service delivery patterns after introduction of telemedicine provision of medical abortion in Iowa. Am J Public Health 2013 Jan;103(1):73-78. [doi: 10.2105/AJPH.2012.301097] [Medline: 23153158]
111. Grossman D, Grindlay K, Buchacker T, Lane K, Blanchard K. Effectiveness and acceptability of medical abortion provided through telemedicine. Obstet Gynecol 2011 Aug;118(2 Pt 1):296-303. [doi: 10.1097/AOG.0b013e318224d110] [Medline: 21775845]
112. Grossman D, Baum SE, Andjelic D, Tatum C, Torres G, Fuentes L, et al. A harm-reduction model of abortion counseling about misoprostol use in Peru with telephone and in-person follow-up: a cohort study. PLoS One 2018 Jan 10;13(1):e0189195 [FREE Full text] [doi: 10.1371/journal.pone.0189195] [Medline: 29320513]
113. Cameron ST, Glasier A, Dewart H, Johnstone A, Burnside A. Telephone follow-up and self-performed urine pregnancy testing after early medical abortion: a service evaluation. Contraception 2012 Jul;86(1):67-73. [doi: 10.1016/j.contraception.2011.11.010] [Medline: 22225841]
114. Maddison R, Rawstorn JC, Stewart RA, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan 27;105(2):122-129 [FREE Full text] [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
115. Scalvini S, Zanelli E, Comini L, Dalla Tomba M, Troise G, Febo O, et al. Home-based versus in-hospital cardiac rehabilitation after cardiac surgery: a nonrandomized controlled study. Phys Ther 2013 Aug;93(8):1073-1083. [doi: 10.2522/ptj.20120212] [Medline: 23599353]
116. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty: a noninferiority randomized controlled trial. J Bone Joint Surg Am 2015 Jul 15;97(14):1129-1141. [doi: 10.2106/JBJS.N.01066] [Medline: 26178888]
117. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396 [FREE Full text] [doi: 10.2340/16501977-1119] [Medline: 23474735]
118. Kuether J, Moore A, Kahan J, Martucci J, Messina T, Perreault R, et al. Telerehabilitation for total hip and knee arthroplasty patients: a pilot series with high patient satisfaction. HSS J 2019 Oct 21;15(3):221-225 [FREE Full text] [doi: 10.1007/s11420-019-09715-w] [Medline: 31624476]
119. Nicola K, Waugh J, Charles E, Russell T. The feasibility and concurrent validity of performing the Movement Assessment Battery for Children - 2nd Edition via telerehabilitation technology. Res Dev Disabil 2018 Jun;77:40-48. [doi: 10.1016/j.ridd.2018.04.001] [Medline: 29656273]
120. Ewald B, Stacey F, Johnson N, Plotnikoff RC, Holliday E, Brown W, et al. Physical activity coaching by Australian Exercise Physiologists is cost effective for patients referred from general practice. Aust N Z J Public Health 2018 Feb;42(1):12-15 [FREE Full text] [doi: 10.1111/1753-6405.12733] [Medline: 29165855]
121. Cottrell MA, O'Leary SP, Swete-Kelly P, Elwell B, Hess S, Litchfield MA, et al. Agreement between telehealth and in-person assessment of patients with chronic musculoskeletal conditions presenting to an advanced-practice physiotherapy screening clinic. Musculoskelet Sci Pract 2018 Dec;38:99-105. [doi: 10.1016/j.msksp.2018.09.014] [Medline: 30366292]
122. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31;2015(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]

XSL•FO
RenderX

123. Anderson L, Sharp GA, Norton RJ, Dalal H, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6(6):CD007130 [FREE Full text] [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
124. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]
125. Cottrell MA, Galea OA, O'Leary SP, Hill AJ, Russell TG. Real-time telerehabilitation for the treatment of musculoskeletal conditions is effective and comparable to standard practice: a systematic review and meta-analysis. Clin Rehabil 2017 May 02;31(5):625-638. [doi: 10.1177/0269215516645148] [Medline: 27141087]
126. Jiang S, Xiang J, Gao X, Guo K, Liu B. The comparison of telerehabilitation and face-to-face rehabilitation after total knee arthroplasty: a systematic review and meta-analysis. J Telemed Telecare 2016 Dec 27;24(4):257-262. [doi: 10.1177/1357633x16686748]
127. Shukla H, Nair SR, Thakker D. Role of telerehabilitation in patients following total knee arthroplasty: evidence from a systematic literature review and meta-analysis. J Telemed Telecare 2016 Jul 09;23(2):339-346. [doi: 10.1177/1357633x16628996]
128. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. Heart 2016 Aug 01;102(15):1183-1192. [doi: 10.1136/heartjnl-2015-308966] [Medline: 26936337]
129. Mani S, Sharma S, Omar B, Paungmali A, Joseph L. Validity and reliability of internet-based physiotherapy assessment for musculoskeletal disorders: a systematic review. J Telemed Telecare 2016 Mar 31;23(3):379-391. [doi: 10.1177/1357633x16642369]
130. Telehealth for Speech and Language Pathology: A Review of Clinical Effectiveness, Cost-Effectiveness, and Guidelines [Internet]. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2015.
131. Alegría M, Ludman E, Kafali EN, Lapatin S, Vila D, Shrout PE, et al. Effectiveness of the Engagement and Counseling for Latinos (ECLA) intervention in low-income Latinos. Med Care 2014 Nov;52(11):989-997 [FREE Full text] [doi: 10.1097/MLR.0000000000000232] [Medline: 25310525]
132. Alcántara C, Li X, Wang Y, Canino G, Alegría M. Treatment moderators and effectiveness of Engagement and Counseling for Latinos intervention on worry reduction in a low-income primary care sample. J Consult Clin Psychol 2016 Nov;84(11):1016-1022 [FREE Full text] [doi: 10.1037/ccp0000146] [Medline: 27631958]
133. Kirkness CJ, Cain KC, Becker KJ, Tirschwell DL, Buzaitis AM, Weisman PL, et al. Randomized trial of telephone versus in-person delivery of a brief psychosocial intervention in post-stroke depression. BMC Res Notes 2017 Oct 10;10(1):500 [FREE Full text] [doi: 10.1186/s13104-017-2819-y] [Medline: 29017589]
134. Byun E, Becker KJ, Kohen R, Kirkness CJ, Mitchell PH. Brief psychosocial intervention to address poststroke depression may also benefit fatigue and sleep-wake disturbance. Rehabil Nurs 2021;46(4):222-231. [doi: 10.1097/RNJ.0000000000000304] [Medline: 33443981]
135. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. JAMA 2012 Jun 06;307(21):2278-2285 [FREE Full text] [doi: 10.1001/jama.2012.5588] [Medline: 22706833]
136. Kalapatapu RK, Ho J, Cai X, Vinogradov S, Batki SL, Mohr DC. Cognitive-behavioral therapy in depressed primary care patients with co-occurring problematic alcohol use: effect of telephone-administered vs. face-to-face treatment-a secondary analysis. J Psychoactive Drugs 2014 May 06;46(2):85-92 [FREE Full text] [doi: 10.1080/02791072.2013.876521] [Medline: 25052784]
137. Stiles-Shields C, Kwasny MJ, Cai X, Mohr DC. Therapeutic alliance in face-to-face and telephone-administered cognitive behavioral therapy. J Consult Clin Psychol 2014 Apr;82(2):349-354 [FREE Full text] [doi: 10.1037/a0035554] [Medline: 24447003]
138. Stiles-Shields C, Corden ME, Kwasny MJ, Schueller SM, Mohr DC. Predictors of outcome for telephone and face-to-face administered cognitive behavioral therapy for depression. Psychol Med 2015 Nov;45(15):3205-3215 [FREE Full text] [doi: 10.1017/S0033291715001208] [Medline: 26077620]
139. Choi NG, Marti CN, Bruce ML, Hegel MT, Wilson NL, Kunik ME. Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older adults. Depress Anxiety 2014 Aug 05;31(8):653-661 [FREE Full text] [doi: 10.1002/da.22242] [Medline: 24501015]
140. Choi NG, Marti CN, Conwell Y. Effect of problem-solving therapy on depressed low-income homebound older adults' death/suicidal ideation and hopelessness. Suicide Life Threat Behav 2016 Jun 12;46(3):323-336 [FREE Full text] [doi: 10.1111/sltb.12195] [Medline: 26456016]
141. Egede LE, Acierno R, Knapp RG, Lejuez C, Hernandez-Tejada M, Payne EH, et al. Psychotherapy for depression in older veterans via telemedicine: a randomised, open-label, non-inferiority trial. Lancet Psychiatry 2015 Aug;2(8):693-701. [doi: 10.1016/s2215-0366(15)00122-4]
142. Egede LE, Acierno R, Knapp RG, Walker RJ, Payne EH, Frueh BC. Psychotherapy for depression in older veterans via telemedicine: effect on quality of life, satisfaction, treatment credibility, and service delivery perception. J Clin Psychiatry 2016 Dec 08;77(12):1704-1711 [FREE Full text] [doi: 10.4088/JCP.16m10951] [Medline: 27835713]

XSL•FO
RenderX

143. Egede LE, Gebregziabher M, Walker RJ, Payne EH, Acierno R, Frueh BC. Trajectory of cost overtime after psychotherapy for depression in older veterans via telemedicine. J Affect Disord 2017 Jan 01;207:157-162. [doi: 10.1016/j.jad.2016.09.044] [Medline: 27721190]
144. Egede LE, Dismuke CE, Walker RJ, Acierno R, Frueh BC. Cost-effectiveness of behavioral activation for depression in older adult veterans. J Clin Psychiatry 2018 Aug 28;79(5):17m11888. [doi: 10.4088/jcp.17m11888]
145. Egede LE, Walker RJ, Payne EH, Knapp RG, Acierno R, Frueh BC. Effect of psychotherapy for depression via home telehealth on glycemic control in adults with type 2 diabetes: subgroup analysis of a randomized clinical trial. J Telemed Telecare 2018 Oct;24(9):596-602. [doi: 10.1177/1357633X17730419] [Medline: 28945160]
146. Luxton DD, Pruitt LD, Wagner A, Smolenski DJ, Jenkins-Guarnieri MA, Gahm G. Home-based telebehavioral health for U.S. military personnel and veterans with depression: a randomized controlled trial. J Consult Clin Psychol 2016 Nov;84(11):923-934. [doi: 10.1037/ccp0000135] [Medline: 27599225]
147. Smolenski DJ, Pruitt LD, Vuletic S, Luxton DD, Gahm G. Unobserved heterogeneity in response to treatment for depression through videoconference. Psychiatr Rehabil J 2017 Sep;40(3):303-308. [doi: 10.1037/prj0000273] [Medline: 28604014]
148. Pruitt LD, Vuletic S, Smolenski DJ, Wagner A, Luxton DD, Gahm GA. Predicting post treatment client satisfaction between behavioural activation for depression delivered either in-person or via home-based telehealth. J Telemed Telecare 2018 Jul 05;25(8):460-467 [FREE Full text] [doi: 10.1177/1357633x18784103]
149. Bouchard S, Allard M, Robillard G, Dumoulin S, Guitard T, Loranger C, et al. Videoconferencing psychotherapy for panic disorder and agoraphobia: outcome and treatment processes from a non-randomized non-inferiority trial. Front Psychol 2020 Aug 21;11:2164 [FREE Full text] [doi: 10.3389/fpsyg.2020.02164] [Medline: 32973638]
150. Milosevic I, Cameron DH, Milanovic M, McCabe RE, Rowa K. Face-to-face versus video teleconference group cognitive behavioural therapy for anxiety and related disorders: a preliminary comparison. Can J Psychiatry 2022 May 23;67(5):391-402 [FREE Full text] [doi: 10.1177/07067437211027319] [Medline: 34159838]
151. Watts S, Marchand A, Bouchard S, Gosselin P, Langlois F, Belleville G, et al. Telepsychotherapy for generalized anxiety disorder: impact on the working alliance. J Psychother Integr 2020 Jun;30(2):208-225. [doi: 10.1037/int0000223]
152. Acierno R, Gros DF, Ruggiero KJ, Hernandez-Tejada MA, Knapp RG, Lejuez CW, et al. Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. Depress Anxiety 2016 May 10;33(5):415-423. [doi: 10.1002/da.22476] [Medline: 26864655]
153. Acierno R, Knapp R, Tuerk P, Gilmore AK, Lejuez C, Ruggiero K, et al. A non-inferiority trial of prolonged exposure for posttraumatic stress disorder: in person versus home-based telehealth. Behav Res Ther 2017 Feb;89:57-65 [FREE Full text] [doi: 10.1016/j.brat.2016.11.009] [Medline: 27894058]
154. Gros DF, Allan NP, Lancaster CL, Szafranski DD, Acierno R. Predictors of treatment discontinuation during prolonged exposure for PTSD. Behav Cogn Psychother 2017 Jul 03;46(1):35-49. [doi: 10.1017/s135246581700039x]
155. Acierno R, Jaffe AE, Gilmore AK, Birks A, Denier C, Muzzy W, et al. A randomized clinical trial of in-person vs. home-based telemedicine delivery of prolonged exposure for PTSD in military sexual trauma survivors. J Anxiety Disord 2021 Oct;83:102461. [doi: 10.1016/j.janxdis.2021.102461] [Medline: 34391978]
156. White CN, Kauffman BY, Acierno R. Factors contributing to veterans' satisfaction with PTSD treatment delivered in person compared to telehealth. J Telemed Telecare 2021 Jan 26;29(6):426-434. [doi: 10.1177/1357633x20987704]
157. Morland LA, Mackintosh MA, Glassman LH, Wells SY, Thorp SR, Rauch SA, et al. Home-based delivery of variable length prolonged exposure therapy: a comparison of clinical efficacy between service modalities. Depress Anxiety 2020 Apr 24;37(4):346-355. [doi: 10.1002/da.22979] [Medline: 31872563]
158. Morland LA, Mackintosh MA, Greene CJ, Rosen CS, Chard KM, Resick P, et al. Cognitive processing therapy for posttraumatic stress disorder delivered to rural veterans via telemental health. J Clin Psychiatry 2014 May 15;75(05):470-476. [doi: 10.4088/jcp.13m08842]
159. Morland LA, Mackintosh MA, Rosen CS, Willis E, Resick P, Chard K, et al. Telemedicine versus in-person delivery of cognitive processing therapy for women with posttraumatic stress disorder: a randomized noninferiority trial. Depress Anxiety 2015 Nov 03;32(11):811-820. [doi: 10.1002/da.22397] [Medline: 26243685]
160. Glassman LH, Mackintosh MA, Talkovsky A, Wells SY, Walter KH, Wickramasinghe I, et al. Quality of life following treatment for PTSD: comparison of videoconferencing and in-person modalities. J Telemed Telecare 2017 Nov 16;25(2):123-127. [doi: 10.1177/1357633x17740610]
161. Liu L, Thorp SR, Moreno L, Wells SY, Glassman LH, Busch AC, et al. Videoconferencing psychotherapy for veterans with PTSD: results from a randomized controlled non-inferiority trial. J Telemed Telecare 2019 Jun 19;26(9):507-519. [doi: 10.1177/1357633x19853947]
162. Maieritsch KP, Smith TL, Hessinger JD, Ahearn EP, Eickhoff JC, Zhao Q. Randomized controlled equivalence trial comparing videoconference and in person delivery of cognitive processing therapy for PTSD. J Telemed Telecare 2016 Jun;22(4):238-243. [doi: 10.1177/1357633X15596109] [Medline: 26231819]
163. Morland LA, Greene CJ, Rosen CS, Foy D, Reilly P, Shore J, et al. Telemedicine for anger management therapy in a rural population of combat veterans with posttraumatic stress disorder: a randomized noninferiority trial. J Clin Psychiatry 2010 Jul;71(7):855-863. [doi: 10.4088/JCP.09m05604blu] [Medline: 20122374]

XSL•FO
RenderX

164. Morland LA, Greene CJ, Grubbs K, Kloezeman K, Mackintosh MA, Rosen C, et al. Therapist adherence to manualized cognitive-behavioral therapy for anger management delivered to veterans with PTSD via videoconferencing. J Clin Psychol 2011 Jun 25;67(6):629-638. [doi: 10.1002/jclp.20779] [Medline: 21360528]
165. Greene CJ, Morland LA, Macdonald A, Frueh BC, Grubbs KM, Rosen CS. How does tele-mental health affect group therapy process? Secondary analysis of a noninferiority trial. J Consult Clin Psychol 2010 Oct;78(5):746-750 [FREE Full text] [doi: 10.1037/a0020158] [Medline: 20873910]
166. Valentine LM, Donofry SD, Broman RB, Smith ER, Rauch SA, Sexton MB. Comparing PTSD treatment retention among survivors of military sexual trauma utilizing clinical video technology and in-person approaches. J Telemed Telecare 2019 Apr 11;26(7-8):443-451. [doi: 10.1177/1357633x19832419]
167. Kelleher SA, Winger JG, Dorfman CS, Ingle KK, Moskovich AA, Abernethy AP, et al. A behavioral cancer pain intervention: a randomized noninferiority trial comparing in-person with videoconference delivery. Psychooncology 2019 Aug 19;28(8):1671-1678 [FREE Full text] [doi: 10.1002/pon.5141] [Medline: 31162756]
168. Check DK, Winger JG, Jones KA, Somers TJ. Predictors of response to an evidence-based behavioral cancer pain management intervention: an exploratory analysis from a clinical trial. J Pain Symptom Manage 2021 Aug;62(2):391-399 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.12.020] [Medline: 33387606]
169. Herbert MS, Afari N, Liu L, Heppner P, Rutledge T, Williams K, et al. Telehealth versus in-person acceptance and commitment therapy for chronic pain: a randomized noninferiority trial. J Pain 2017 Feb;18(2):200-211 [FREE Full text] [doi: 10.1016/j.jpain.2016.10.014] [Medline: 27838498]
170. Levy RL, Langer SL, van Tilburg MA, Romano JM, Murphy TB, Walker LS, et al. Brief telephone-delivered cognitive behavioral therapy targeted to parents of children with functional abdominal pain: a randomized controlled trial. Pain 2017 Apr;158(4):618-628 [FREE Full text] [doi: 10.1097/j.pain.0000000000000800] [Medline: 28301859]
171. Chavooshi B, Mohammadkhani P, Dolatshahee B. Telemedicine vs. in-person delivery of intensive short-term dynamic psychotherapy for patients with medically unexplained pain: a 12-month randomized, controlled trial. J Telemed Telecare 2016 Jul 09;23(1):133-141. [doi: 10.1177/1357633x15627382]
172. Arnedt JT, Conroy DA, Mooney A, Furgal A, Sen A, Eisenberg D. Telemedicine versus face-to-face delivery of cognitive behavioral therapy for insomnia: a randomized controlled noninferiority trial. Sleep 2021 Jan 21;44(1):zsaa136. [doi: 10.1093/sleep/zsaa136] [Medline: 32658298]
173. Crow SJ, Mitchell JE, Crosby RD, Swanson SA, Wonderlich S, Lancanster K. The cost effectiveness of cognitive behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2009 Jun;47(6):451-453 [FREE Full text] [doi: 10.1016/j.brat.2009.02.006] [Medline: 19356743]
174. Mitchell JE, Crosby RD, Wonderlich SA, Crow S, Lancaster K, Simonich H, et al. A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2008 May;46(5):581-592 [FREE Full text] [doi: 10.1016/j.brat.2008.02.004] [Medline: 18374304]
175. Marrone S, Mitchell JE, Crosby R, Wonderlich S, Jollie-Trottier T. Predictors of response to cognitive behavioral treatment for bulimia nervosa delivered via telemedicine versus face-to-face. Int J Eat Disord 2009 Apr 24;42(3):222-227 [FREE Full text] [doi: 10.1002/eat.20603] [Medline: 18951452]
176. Ertelt TW, Crosby RD, Marino JM, Mitchell JE, Lancaster K, Crow SJ. Therapeutic factors affecting the cognitive behavioral treatment of bulimia nervosa via telemedicine versus face-to-face delivery. Int J Eat Disord 2011 Dec 15;44(8):687-691 [FREE Full text] [doi: 10.1002/eat.20874] [Medline: 22072405]
177. Zerwas SC, Watson HJ, Hofmeier SM, Levine MD, Hamer RM, Crosby RD, et al. CBT4BN: a randomized controlled trial of online chat and face-to-face group therapy for bulimia nervosa. Psychother Psychosom 2017;86(1):47-53 [FREE Full text] [doi: 10.1159/000449025] [Medline: 27883997]
178. Watson HJ, Levine MD, Zerwas SC, Hamer RM, Crosby RD, Sprecher CS, et al. Predictors of dropout in face-to-face and internet-based cognitive-behavioral therapy for bulimia nervosa in a randomized controlled trial. Int J Eat Disord 2017 May;50(5):569-577 [FREE Full text] [doi: 10.1002/eat.22644] [Medline: 27862108]
179. Watson HJ, McLagan N, Zerwas SC, Crosby RD, Levine MD, Runfola CD, et al. Cost-effectiveness of internet-based cognitive-behavioral treatment for bulimia nervosa: results of a randomized controlled trial. J Clin Psychiatry 2018;79(1):16m11314 [FREE Full text] [doi: 10.4088/JCP.16m11314] [Medline: 29228517]
180. Lindegaard Pedersen J, Pedersen PU, Damsgaard EM. Nutritional follow-up after discharge prevents readmission to hospital - a randomized clinical trial. J Nutr Health Aging 2017 May 26;21(1):75-82 [FREE Full text] [doi: 10.1007/s12603-016-0745-7] [Medline: 27999853]
181. Lovell K, Cox D, Haddock G, Jones C, Raines D, Garvey R, et al. Telephone administered cognitive behaviour therapy for treatment of obsessive compulsive disorder: randomised controlled non-inferiority trial. BMJ 2006 Oct 28;333(7574):883 [FREE Full text] [doi: 10.1136/bmj.38940.355602.80] [Medline: 16935946]
182. Turner CM, Mataix-Cols D, Lovell K, Krebs G, Lang K, Byford S, et al. Telephone cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled non-inferiority trial. J Am Acad Child Adolesc Psychiatry 2014 Dec;53(12):1298-307.e2 [FREE Full text] [doi: 10.1016/j.jaac.2014.09.012] [Medline: 25457928]
183. Nair A, Turner C, Heyman I, Mataix-Cols D, Lovell K, Krebs G, et al. Moderators and predictors of outcomes in telephone delivered compared to face-to-face cognitive behaviour therapy for paediatric obsessive-compulsive disorder: preliminary

XSL•FO
RenderX

evidence from a non-inferiority RCT. Cogn Behav Ther 2019 Sep 17;48(5):353-368. [doi: 10.1080/16506073.2018.1513555] [Medline: 30221589]
184. Tie H, Krebs G, Lang K, Shearer J, Turner C, Mataix-Cols D, et al. Cost-effectiveness analysis of telephone cognitive-behaviour therapy for adolescents with obsessive-compulsive disorder. BJPsych Open 2019 Jan;5(1):e7 [FREE Full text] [doi: 10.1192/bjo.2018.73] [Medline: 30762502]
185. Watson M, White C, Lynch A, Mohammed K. Telephone-delivered individual cognitive behavioural therapy for cancer patients: an equivalence randomised trial. Psychooncology 2017 Mar;26(3):301-308. [doi: 10.1002/pon.4338] [Medline: 27943570]
186. Rodrigues P, Watson M, White C, Lynch A, Mohammed K, Sagoo GS. Cost-effectiveness analysis of telephone-based cognitive behaviour therapy compared to treatment as usual CBT for cancer patients: evidence from a small, randomised controlled trial. Psychooncology 2021 Oct;30(10):1691-1698 [FREE Full text] [doi: 10.1002/pon.5751] [Medline: 34153136]
187. Lleras de Frutos M, Medina JC, Vives J, Casellas-Grau A, Marzo JL, Borràs JM, et al. Video conference vs face-to-face group psychotherapy for distressed cancer survivors: a randomized controlled trial. Psychooncology 2020 Dec 07;29(12):1995-2003. [doi: 10.1002/pon.5457] [Medline: 32618395]
188. Guzman D, Ann-Yi S, Bruera E, Wu J, Williams JL, Najera J, et al. Enhancing palliative care patient access to psychological counseling through outreach telehealth services. Psychooncology 2020 Jan 10;29(1):132-138. [doi: 10.1002/pon.5270] [Medline: 31707735]
189. Morgan RD, Patrick AR, Magaletta PR. Does the use of telemental health alter the treatment experience? Inmates' perceptions of telemental health versus face-to-face treatment modalities. J Consult Clin Psychol 2008 Feb;76(1):158-162. [doi: 10.1037/0022-006X.76.1.158] [Medline: 18229993]
190. Ospina-Pinillos L, Davenport T, Iorfino F, Tickell A, Cross S, Scott EM, et al. Using new and innovative technologies to assess clinical stage in early intervention youth mental health services: evaluation study. J Med Internet Res 2018 Sep 10;20(9):e259 [FREE Full text] [doi: 10.2196/jmir.9966] [Medline: 30201602]
191. Fukkink RG, Hermanns JM. Children's experiences with chat support and telephone support. J Child Psychol Psychiatry 2009 Jun;50(6):759-766. [doi: 10.1111/j.1469-7610.2008.02024.x] [Medline: 19207634]
192. Fukkink R, Hermanns J. Counseling children at a helpline: chatting or calling? J Community Psychol 2009 Oct 20;37(8):939-948. [doi: 10.1002/jcop.20340]
193. King R, Bambling M, Reid W, Thomas I. Telephone and online counselling for young people: a naturalistic comparison of session outcome, session impact and therapeutic alliance. Couns Psychother Res 2007 Feb 17;6(3):175-181. [doi: 10.1080/14733140600874084]
194. Murphy L, Parnass P, Mitchell DL, Hallett R, Cayley P, Seagram S. Client satisfaction and outcome comparisons of online and face-to-face counselling methods. Br J Soc Work 2009 Apr 15;39(4):627-640. [doi: 10.1093/bjsw/bcp041]
195. Laver K, Liu E, Clemson L, Davies O, Gray L, Gitlin LN, et al. Does telehealth delivery of a dyadic dementia care program provide a noninferior alternative to face-to-face delivery of the same program? A randomized, controlled trial. Am J Geriatr Psychiatry 2020 Jun;28(6):673-682 [FREE Full text] [doi: 10.1016/j.jagp.2020.02.009] [Medline: 32234275]
196. Wade SL, Cassedy AE, McNally KA, Kurowski BG, Kirkwood MW, Stancin T, et al. A randomized comparative effectiveness trial of family-problem-solving treatment for adolescent brain injury: parent outcomes from the coping with head injury through problem solving (CHIPS) study. J Head Trauma Rehabil 2019;34(6):E1-E9. [doi: 10.1097/HTR.0000000000000487] [Medline: 31033747]
197. Corona LL, Stainbrook JA, Simcoe K, Wagner L, Fowler B, Weitlauf AS, et al. Utilization of telemedicine to support caregivers of young children with ASD and their Part C service providers: a comparison of intervention outcomes across three models of service delivery. J Neurodev Disord 2021 Sep 15;13(1):38 [FREE Full text] [doi: 10.1186/s11689-021-09387-w] [Medline: 34525940]
198. Kalichman SC, Katner H, Eaton LA, Banas E, Hill M, Kalichman MO. Comparative effects of telephone versus in-office behavioral counseling to improve HIV treatment outcomes among people living with HIV in a rural setting. Transl Behav Med 2021 Apr 07;11(3):852-862 [FREE Full text] [doi: 10.1093/tbm/ibaa109] [Medline: 33200772]
199. Kalichman SC, Katner H, Eaton LA, Hill M, Ewing W, Kalichman MO. Randomized community trial comparing telephone versus clinic-based behavioral health counseling for people living with HIV in a rural setting. J Rural Health 2022 Sep 08;38(4):728-739. [doi: 10.1111/jrh.12618] [Medline: 34494681]
200. Phanuphak N, Anand T, Jantarapakde J, Nitpolprasert C, Himmad K, Sungsing T, et al. What would you choose: online or offline or mixed services? Feasibility of online HIV counselling and testing among Thai men who have sex with men and transgender women and factors associated with service uptake. J Int AIDS Soc 2018 Jul;21 Suppl 5(Suppl Suppl 5):e25118 [FREE Full text] [doi: 10.1002/jia2.25118] [Medline: 30033644]
201. Delahanty LM, Chang Y, Levy DE, Porneala B, Dushkin A, Bissett L, et al. Design and participant characteristics of a primary care adaptation of the Look AHEAD Lifestyle Intervention for weight loss in type 2 diabetes: the REAL HEALTH-diabetes study. Contemp Clin Trials 2018 Aug;71:9-17 [FREE Full text] [doi: 10.1016/j.cct.2018.05.018] [Medline: 29803816]

XSL•FO
RenderX

202. Delahanty LM, Levy DE, Chang Y, Porneala BC, Goldman V, McCarthy J, et al. Effectiveness of lifestyle intervention for type 2 diabetes in primary care: the REAL HEALTH-diabetes randomized clinical trial. J Gen Intern Med 2020 Sep 21;35(9):2637-2646 [FREE Full text] [doi: 10.1007/s11606-019-05629-9] [Medline: 31965526]
203. Duke DC, Wagner DV, Ulrich J, Freeman KA, Harris MA. Videoconferencing for teens with diabetes: family matters. J Diabetes Sci Technol 2016 Jul;10(4):816-823 [FREE Full text] [doi: 10.1177/1932296816642577] [Medline: 27075708]
204. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
205. Riley AR, Duke DC, Freeman KA, Hood KK, Harris MA. Depressive symptoms in a trial behavioral family systems therapy for diabetes: a post hoc analysis of change. Diabetes Care 2015 Aug;38(8):1435-1440. [doi: 10.2337/dc14-2519] [Medline: 26015558]
206. Nevanperä N, Keränen AM, Ukkola O, Laitinen J. Effects of group counseling transmitted through videoconferencing on changes in eating behaviors. J Nutr Educ Behav 2015 Nov;47(6):555-9.e1. [doi: 10.1016/j.jneb.2015.07.004] [Medline: 26323164]
207. Harder VS, Musau AM, Musyimi CW, Ndetei DM, Mutiso VN. A randomized clinical trial of mobile phone motivational interviewing for alcohol use problems in Kenya. Addiction 2020 Jun 03;115(6):1050-1060 [FREE Full text] [doi: 10.1111/add.14903] [Medline: 31782966]
208. Lohr PA, Aiken AR, Forsyth T, Trussell J. Telephone or integrated contraception counselling before abortion: impact on method choice and receipt. BMJ Sex Reprod Health 2018 Apr 03;44(2):114-121 [FREE Full text] [doi: 10.1136/bmjsrh-2017-101818] [Medline: 29921634]
209. James EL, Ewald BD, Johnson NA, Stacey FG, Brown WJ, Holliday EG, et al. Referral for expert physical activity counseling: a pragmatic RCT. Am J Prev Med 2017 Oct;53(4):490-499. [doi: 10.1016/j.amepre.2017.06.016] [Medline: 28818417]
210. Andrews M, Baker AL, Halpin SA, Lewin TJ, Richmond R, Kay-Lambkin FJ, et al. Early therapeutic alliance, treatment retention, and 12-month outcomes in a healthy lifestyles intervention for people with psychotic disorders. J Nerv Ment Dis 2016 Dec;204(12):894-902. [doi: 10.1097/NMD.0000000000000585] [Medline: 27575791]
211. Baker AL, Richmond R, Kay-Lambkin FJ, Filia SL, Castle D, Williams JM, et al. Randomized controlled trial of a healthy lifestyle intervention among smokers with psychotic disorders. Nicotine Tob Res 2015 Aug 05;17(8):946-954. [doi: 10.1093/ntr/ntv039] [Medline: 25744962]
212. Baker AL, Richmond R, Kay-Lambkin FJ, Filia SL, Castle D, Williams JM, et al. Randomised controlled trial of a healthy lifestyle intervention among smokers with psychotic disorders: outcomes to 36 months. Aust N Z J Psychiatry 2018 Mar 14;52(3):239-252. [doi: 10.1177/0004867417714336] [Medline: 28610482]
213. Berndt N, Bolman C, Lechner L, Max W, Mudde A, de Vries H, et al. Economic evaluation of a telephone- and face-to-face-delivered counseling intervention for smoking cessation in patients with coronary heart disease. Eur J Health Econ 2016 Apr 22;17(3):269-285. [doi: 10.1007/s10198-015-0677-x] [Medline: 25796578]
214. Byaruhanga J, Paul CL, Wiggers J, Byrnes E, Mitchell A, Lecathelinais C, et al. Connectivity of real-time video counselling versus telephone counselling for smoking cessation in rural and remote areas: an exploratory study. Int J Environ Res Public Health 2020 Apr 22;17(8):2891 [FREE Full text] [doi: 10.3390/ijerph17082891] [Medline: 32331356]
215. Byaruhanga J, Wiggers J, Paul CL, Byrnes E, Mitchell A, Lecathelinais C, et al. Acceptability of real-time video counselling compared to other behavioural interventions for smoking cessation in rural and remote areas. Drug Alcohol Depend 2020 Dec 01;217:108296 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108296] [Medline: 32980788]
216. Byaruhanga J, Paul CL, Wiggers J, Byrnes E, Mitchell A, Lecathelinais C, et al. The short-term effectiveness of real-time video counselling on smoking cessation among residents in rural and remote areas: an interim analysis of a randomised trial. J Subst Abuse Treat 2021 Dec;131:108448. [doi: 10.1016/j.jsat.2021.108448] [Medline: 34098302]
217. Carlson LE, Lounsberry JJ, Maciejewski O, Wright K, Collacutt V, Taenzer P. Telehealth-delivered group smoking cessation for rural and urban participants: feasibility and cessation rates. Addict Behav 2012 Jan;37(1):108-114. [doi: 10.1016/j.addbeh.2011.09.011] [Medline: 21968227]
218. Nomura A, Tanigawa T, Muto T, Oga T, Fukushima Y, Kiyosue A, et al. Clinical efficacy of telemedicine compared to face-to-face clinic visits for smoking cessation: multicenter open-label randomized controlled noninferiority trial. J Med Internet Res 2019 Apr 26;21(4):e13520 [FREE Full text] [doi: 10.2196/13520] [Medline: 30982776]
219. Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. J Med Internet Res 2015 May 08;17(5):e113 [FREE Full text] [doi: 10.2196/jmir.3975] [Medline: 25956257]
220. Appel LJ, Clark JM, Yeh HC, Wang NY, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. N Engl J Med 2011 Nov 24;365(21):1959-1968. [doi: 10.1056/nejmoa1108660]
221. Dalcin AT, Jerome GJ, Fitzpatrick SL, Louis TA, Wang NY, Bennett WL, et al. Perceived helpfulness of the individual components of a behavioural weight loss program: results from the Hopkins POWER trial. Obes Sci Pract 2015 Oct 09;1(1):23-32 [FREE Full text] [doi: 10.1002/osp4.6] [Medline: 27668085]

XSL•FO
RenderX

222. Daumit GL, Janssen EM, Jerome GJ, Dalcin AT, Charleston J, Clark JM, et al. Cost of behavioral weight loss programs implemented in clinical practice: the POWER trial at Johns Hopkins. Transl Behav Med 2020 Feb 03;10(1):103-113 [FREE Full text] [doi: 10.1093/tbm/iby120] [Medline: 30855082]
223. Befort CA, VanWormer JJ, Desouza C, Ellerbeck EF, Gajewski B, Kimminau KS, et al. Effect of behavioral therapy with in-clinic or telephone group visits vs in-clinic individual visits on weight loss among patients with obesity in rural clinical practice: a randomized clinical trial. JAMA 2021 Jan 26;325(4):363-372 [FREE Full text] [doi: 10.1001/jama.2020.25855] [Medline: 33496775]
224. Kurz D, McCrea-Robertson S, Nelson-Brantley H, Befort C. Rural engagement in primary care for optimizing weight reduction (REPOWER): a mixed methods study of patient perceptions. Patient Educ Couns 2022 Jul;105(7):2371-2381. [doi: 10.1016/j.pec.2021.11.028] [Medline: 34865892]
225. Fujii H, Yokoyama T, Yoshimi I, Mizushima S. A randomized controlled trial to evaluate the effects of health guidance with video call as compared to face-to-face health guidance. Int Med J 2017;24(2):186.
226. Harrigan M, Cartmel B, Loftfield E, Sanft T, Chagpar AB, Zhou Y, et al. Randomized trial comparing telephone versus in-person weight loss counseling on body composition and circulating biomarkers in women treated for breast cancer: the lifestyle, exercise, and nutrition (LEAN) study. J Clin Oncol 2016 Mar 01;34(7):669-676 [FREE Full text] [doi: 10.1200/JCO.2015.61.6375] [Medline: 26598750]
227. Harvey-Berino J. Changing health behavior via telecommunications technology: using interactive television to treat obesity. Behav Ther 1998;29(3):505-519. [doi: 10.1016/s0005-7894(98)80046-4]
228. Lukenbill T, San Giovanni CB, Simpson A, Chew M, Basco W, Roberts J. Assessing anthropometric and laboratory outcomes of a paediatric telehealth weight management program. J Telemed Telecare 2021 Feb 09;29(5):399-405. [doi: 10.1177/1357633x20986022]
229. Lutes LD, Dinatale E, Goodrich DE, Ronis DL, Gillon L, Kirsh S, et al. A randomized trial of a small changes approach for weight loss in veterans: design, rationale, and baseline characteristics of the ASPIRE-VA trial. Contemp Clin Trials 2013 Jan;34(1):161-172. [doi: 10.1016/j.cct.2012.09.007] [Medline: 23041618]
230. Damschroder LJ, Lutes LD, Kirsh S, Kim HM, Gillon L, Holleman RG, et al. Small-changes obesity treatment among veterans: 12-month outcomes. Am J Prev Med 2014 Nov;47(5):541-553. [doi: 10.1016/j.amepre.2014.06.016] [Medline: 25217098]
231. Vimalananda V, Damschroder L, Janney CA, Goodrich D, Kim HM, Holleman R, et al. Weight loss among women and men in the ASPIRE-VA behavioral weight loss intervention trial. Obesity (Silver Spring) 2016 Sep 04;24(9):1884-1891 [FREE Full text] [doi: 10.1002/oby.21574] [Medline: 27488278]
232. Lutes LD, Damschroder LJ, Masheb R, Kim HM, Gillon L, Holleman RG, et al. Behavioral treatment for veterans with obesity: 24-month weight outcomes from the ASPIRE-VA small changes randomized trial. J Gen Intern Med 2017 Apr 7;32(Suppl 1):40-47 [FREE Full text] [doi: 10.1007/s11606-017-3987-0] [Medline: 28271430]
233. Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E. Live psychotherapy by video versus in-person: a meta-analysis of efficacy and its relationship to types and targets of treatment. Clin Psychol Psychother 2021 Nov 05;28(6):1535-1549. [doi: 10.1002/cpp.2594] [Medline: 33826190]
234. Thomas N, McDonald C, de Boer K, Brand RM, Nedeljkovic M, Seabrook L. Review of the current empirical literature on using videoconferencing to deliver individual psychotherapies to adults with mental health problems. Psychol Psychother 2021 Sep 23;94(3):854-883 [FREE Full text] [doi: 10.1111/papt.12332] [Medline: 33620133]
235. Byaruhanga J, Atorkey P, McLaughlin M, Brown A, Byrnes E, Paul C, et al. Effectiveness of individual real-time video counseling on smoking, nutrition, alcohol, physical activity, and obesity health risks: systematic review. J Med Internet Res 2020 Sep 11;22(9):e18621 [FREE Full text] [doi: 10.2196/18621] [Medline: 32915156]
236. Irvine A, Drew P, Bower P, Brooks H, Gellatly J, Armitage CJ, et al. Are there interactional differences between telephone and face-to-face psychological therapy? A systematic review of comparative studies. J Affect Disord 2020 Mar 15;265:120-131 [FREE Full text] [doi: 10.1016/j.jad.2020.01.057] [Medline: 32090733]
237. Ho C, Severn M. e-Therapy Interventions for the Treatments of Patients with Depression: A Review of Clinical Effectiveness [Internet]. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2018.
238. Turgoose D, Ashwick R, Murphy D. Systematic review of lessons learned from delivering tele-therapy to veterans with post-traumatic stress disorder. J Telemed Telecare 2018 Oct;24(9):575-585. [doi: 10.1177/1357633X17730443] [Medline: 28958211]
239. Jobes DA, Crumlish JA, Evans AD. The COVID-19 pandemic and treating suicidal risk: the telepsychotherapy use of CAMS. J Psychother Integr 2020 Jun;30(2):226-237. [doi: 10.1037/int0000208]
240. Rosen CS, Glassman LH, Morland LA. Telepsychotherapy during a pandemic: a traumatic stress perspective. J Psychother Integr 2020 Jun;30(2):174-187. [doi: 10.1037/int0000221]
241. Mendenhall W, Beaver RJ, Beaver BM. Introduction to Probability and Statistics. Pacific Grove, CA: Brooks/Cole Publishing; 2009.
242. Singh S, Fletcher GG, Yao X, Sussman J. Virtual care in patients with cancer: a systematic review. Curr Oncol 2021 Sep 08;28(5):3488-3506 [FREE Full text] [doi: 10.3390/curroncol28050301] [Medline: 34590602]

XSL•FO
RenderX

243. Reed M, Huang J, Somers M, Hsueh L, Graetz I, Millman A, et al. Telemedicine versus in-person primary care: treatment and follow-up visits. Ann Intern Med 2023 Oct;176(10):1349-1357. [doi: 10.7326/M23-1335] [Medline: 37844311]
244. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005 Oct 03;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
245. Saultz JW, Lochner J. Interpersonal continuity of care and care outcomes: a critical review. Ann Fam Med 2005 Mar 01;3(2):159-166 [FREE Full text] [doi: 10.1370/afm.285] [Medline: 15798043]
246. Gellert GA, Rasławska-Socha J, Marcjasz N, Price T, Heyduk A, Mlodawska A, et al. The role of virtual triage in improving clinician experience and satisfaction: a narrative review. Telemed Rep 2023 Jul 01;4(1):180-191 [FREE Full text] [doi: 10.1089/tmr.2023.0020] [Medline: 37529770]
247. Barnabe J, Petrie S, Peters P. Virtual triage and teletriage in rural Canada and Australia: A rapid review. Spatial Determinants of Health Lab Ottawa. 2020. URL: https://carleton.ca/determinants/wp-content/uploads/VirtualTriage.pdf [accessed 2024-07-29]
248. Jonnagaddala J, Godinho MA, Liaw ST. From telehealth to virtual primary care in Australia? A rapid scoping review. Int J Med Inform 2021 Jul;151:104470 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104470] [Medline: 34000481]
249. Neves AL, Li E, Gupta PP, Fontana G, Darzi A. Virtual primary care in high-income countries during the COVID-19 pandemic: policy responses and lessons for the future. Eur J Gen Pract 2021 Dec 25;27(1):241-247 [FREE Full text] [doi: 10.1080/13814788.2021.1965120] [Medline: 34431426]
250. Agarwal P, Wang R, Meaney C, Walji S, Damji A, Gill N, et al. Sociodemographic differences in patient experience with primary care during COVID-19: results from a cross-sectional survey in Ontario, Canada. BMJ Open 2022 May 09;12(5):e056868 [FREE Full text] [doi: 10.1136/bmjopen-2021-056868] [Medline: 35534055]
251. Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. JAMA 2022 Feb 08;327(6):521-522. [doi: 10.1001/jama.2021.25181] [Medline: 35061006]
252. Segal JB, Davis S, Dukhanin V. Working framework for appropriate use of virtual care in primary care. J Am Board Fam Med 2022 May 31;35(3):629-633 [FREE Full text] [doi: 10.3122/jabfm.2022.03.210469] [Medline: 35641060]

Abbreviations

CBT: cognitive behavioral therapy
GP: general practitioner
PEBC: Program in Evidence-Based Care
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by A Mavragani; submitted 06.12.23; peer-reviewed by A Lu, S Khonsari; comments to author 15.03.24; revised version received 10.08.24; accepted 05.12.24; published 14.02.25.

Please cite as:
Agarwal P, Fletcher GG, Ramamoorthi K, Yao X, Bhattacharyya O
Uses of Virtual Care in Primary Care: Scoping Review
J Med Internet Res 2025;27:e55007
URL: https://www.jmir.org/2025/1/e55007
doi:10.2196/55007
PMID:

©Payal Agarwal, Glenn George Fletcher, Karishini Ramamoorthi, Xiaomei Yao, Onil Bhattacharyya. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Assessing the Effectiveness of Digital Health Behavior Strategies on Type 2 Diabetes Management: Systematic Review and Network Meta-Analysis

Min Li[1], MMS; Shiyu Liu[2], PhD; Binyang Yu[3], PhD; Ning Li[1], PhD; Aili Lyu[1], PhD; Haiyan Yang[1], MMS; Haiyan He[1], MMS; Na Zhang[1], MMS; Jingru Ma[2], MMS; Meichen Sun[2], MMS; Hong Du[2], MMS; Rui Gao[1], PhD

[1]School of Nursing, Health Science Center, Xi 'an Jiaotong University, Xi 'an, China

[2]School of Public Health, Xi'an Jiaotong University, Xi 'an, China

[3]Graduate School, Beijing University of Chinese Medicine, Beijing, China

Corresponding Author:
Rui Gao, PhD
School of Nursing
Health Science Center
Xi 'an Jiaotong University
74 Yanta West Road
Xi 'an, 710061
China
Phone: 86 18966606582
Email: gaorui@xjtu.edu.cn

Abstract

Background: Various mobile technologies and digital health interventions (DHIs) have been developed for type 2 diabetes mellitus (T2DM) management. Strategies are crucial for ensuring the effectiveness of DHIs. However, there is currently a lack of categorization and summarization of the strategies used in DHIs for T2DM.

Objective: This study aims to (1) identify and categorize the strategies used in DHIs for T2DM management; (2) assess the effectiveness of these DHI strategies; and (3) compare and rank the efficacy of different strategy combinations on glycated hemoglobin A_{1c} (HbA_{1c}) levels, fasting blood glucose (FBG) levels, BMI, and weight loss.

Methods: Relevant randomized controlled trials (RCTs) were extracted from PubMed, Web of Science, and Scopus databases. Three rounds of screening and selection were conducted. The strategies were identified and categorized based on the principles of behavior change techniques and behavior strategies. The synthesis framework for the assessment of health IT was used to structure the evaluation of the DHI strategies qualitatively. A network meta-analysis was performed to compare the efficacy of different strategy combinations. The data quality was assessed using the Cochrane Risk of Bias tool.

Results: A total of 52 RCTs were included, identifying 63 strategies categorized into 19 strategy themes. The most commonly used strategies were *guide*, *monitor*, *management*, and *engagement*. Most studies reported positive or mixed outcomes for most indicators based on the synthesis framework for the assessment of health IT. Research involving a medium or high number of strategies was found to be more effective than research involving a low number of strategies. Of 52 RCTs, 27 (52%) were included in the network meta-analysis. The strategy combination of *communication*, *engagement*, *guide*, and *management* was most effective in reducing HbA_{1c} levels (mean difference [MD] –1.04, 95% CI –1.55 to –0.54), while the strategy combination of *guide*, *management*, and *monitor* was effective in reducing FBG levels (MD –0.96, 95% CI –1.86 to –0.06). The strategy combination of *communication*, *engagement*, *goal setting*, *management*, and *support* was most effective for BMI (MD –2.30, 95% CI –3.16 to –1.44) and weight management (MD –6.50, 95% CI –8.82 to –4.18).

Conclusions: Several DHI strategy combinations were effective in reducing HbA_{1c} levels, FBG levels, BMI, and weight in T2DM management. Health care professionals should be encouraged to apply these promising strategy combinations in DHIs during clinical care. Future research should further explore and optimize the design and implementation of strategies.

Trial Registration: PROSPERO CRD42024544629; https://tinyurl.com/3zp2znxt

(J Med Internet Res 2025;27:e63209) doi:10.2196/63209

XSL•FO
RenderX

KEYWORDS

T2DM; type 2 diabetes mellitus; digital health interventions; behavior strategy; strategy combinations; effectiveness; network meta-analysis

Introduction

The Severity of Type 2 Diabetes

Type 2 diabetes mellitus (T2DM) has become a serious public health problem worldwide. It is a progressive disease that can impair health-related quality of life [1,2] while also imposing substantial economic burdens on individuals, health systems, and society [3]. By 2040, there will be 642 million patients with diabetes worldwide, with the incidence of T2DM on the rise across all regions [4]. T2DM accounts for >90% of the diagnosis in all types of diabetes, and it can cause various complications [5]. The goal of diabetes treatment is to prevent or delay complications and optimize quality of life [6]. Poor glucose control is associated with the occurrence of complications [7]. Glycated hemoglobin A_{1c} (HbA_{1c}) can be used to determine glucose control levels [8]. Continuous monitoring of fasting blood glucose (FBG) can provide an intuitive reflection of changes in the patient's glucose levels. Obesity is closely related to T2DM, and weight gain is an independent risk factor for T2DM [9]. Controlling glucose levels and body weight within the normal range can effectively reduce the complications of patients with diabetes [10]. Many studies have shown that T2DM can be slowed down, stopped, or even reversed by changing lifestyle (eg, low-calorie diet and increasing physical activity) [11,12]. This can reduce long-term complications and may extend life expectancy [13].

The Effect of Digital Health Interventions

Maintaining glycemic control is a challenge for both patients and health care providers, making it difficult to encourage or motivate patients to make long-term lifestyle changes, explain their self-monitoring of blood glucose data, provide immediate feedback, and understand their lifestyle [14]. Recognizing that patients require more self-management support, various mobile technologies (ie, mobile health care) and digital health interventions (DHIs) have been developed [15-18], including mobile apps, SMS, wearable and ambient sensors, and social media. These technologies can provide early support for the improvement of health behaviors in patients with diabetes, encouraging patients with T2DM to eat healthily, engage in physical exercise, and collect and analyze personal data to assess clinical conditions. The data reported by patients can be used to customize personalized feedback information, including health promotion, motivation, encouragement, reminders, and emotional support information [19,20]. Relevant systematic reviews and meta-analyses have confirmed the effectiveness of DHIs on behavior change, blood glucose control, and weight loss in patients with diabetes [21-24].

The Critical Role of Strategies in DHIs for T2DM Management

Strategies play a pivotal role in enhancing the efficacy of DHIs for patients with T2DM, with the careful selection and implementation of appropriate strategies being crucial to DHI success [25]. Within the realm of digital interventions, using suitable strategies, such as user-centered participatory design, can heighten user engagement, thereby rendering intervention measures more appealing and efficacious [26]. Furthermore, mobile technology interventions informed by strategies, such as health behavior theory, hold the potential for more comprehensive mechanisms of behavior change, fostering a health care approach that is both impactful and sustainable [27]. Through the identification and categorization of diverse strategies, researchers and practitioners can gain deeper insights into the effective methodologies and techniques used in DHIs. This involves assessing the strengths and limitations of various strategies and their applicability to specific health domains or populations [28].

Current Findings on DHI Strategies for T2DM Management

There have been studies that systematically summarize the strategies used in DHIs currently. For example, various engagement strategies have been reported in DHIs for mental health promotion, including personalization, human and social support, gamification, personalized feedback, and reminders, which work best to promote engagement [29]. A study indicated that the effectiveness of internet-based interventions correlates with the extensive use of theory, particularly the theory of planned behavior; the incorporation of a greater number of behavior change techniques; and the integration of additional methods for interacting with participants [25]. A systematic review and meta-analysis of lifestyle interventions in postpartum women found that the provision of certain strategies including problem-solving, goal setting of the outcome, reviewing the outcome goal, providing feedback, and self-monitoring of behavior was associated with greater decreases in energy intake [30]. In addition, the review analyzed loneliness reduction strategies including improving social skills, enhancing social support, increasing opportunities for social interaction, and addressing deficits in social cognition [31]. It was suggested in a narrative umbrella review that credible sources, social support, prompts and cues, graded tasks, goals and planning, feedback and monitoring, and human coaching and personalization components increased the effectiveness of DHIs targeting the prevention and management of noncommunicable diseases [32]. Moreover, substantial strategies have been found to be effective in improving recruitment, reducing loss to follow-up, and enhancing retention during intervention in some trials [33-36]. However, there is a lack of categorization and summarization of the strategies used in DHIs of T2DM at present. Although previous research has compared the effectiveness of 5 strategies for T2DM management, it has primarily focused on telemedicine within DHIs and compared several independent strategies, without providing a systematic summary of the strategies used in DHIs [37]." Meanwhile, most interventional studies incorporate different numbers and combinations of strategies rather than focusing on individual strategies, and it is not yet known which strategy combination has the best effect.

XSL•FO
RenderX

Therefore, it is essential to summarize and compare strategies and strategy combinations, identify the content of these strategies, and further determine the optimal number and form of strategy combinations to provide more practical guidance for precise and personalized digital health management of T2DM.

Study Objectives

This study aims to (1) identify and categorize the strategies used in DHIs on T2DM management; (2) assess the effectiveness of these DHI strategies; and (3) compare and rank the efficacy of different strategy combinations on HbA_{1c} levels, FBG levels, BMI, and weight loss.

Methods

Protocol and Registration

This systematic review followed the PRISMA-NMA (Preferred Reporting Items for Systematic Review and Meta-Analyses extension for Network Meta-Analysis) 2020 statement for designing and reporting [38]. The PRISMA checklist for this study can be found in Multimedia Appendix 1. The protocol for this study has been registered with PROSPERO (CRD42024544629).

Data Sources

We conducted a comprehensive search for papers published in English from PubMed, Web of Science, and Scopus databases. This search used a combination of DHI-related terms, along with database-specific subject headings and filters, to ensure thoroughness and focus. The detailed search strategy is provided in Multimedia Appendix 2. The time span was from January 1, 1999, to March 10, 2024.

Data Selection and Extraction

All the searched records were imported into EndNote X9 (Clarivate) to eliminate duplicate studies. The first round of screening and selection focused on identifying DHIs. Our criteria were as follows:

1. The means of intervention should be digital, primarily including wearable devices, telemedicine, electronic health records, electronic medical records, mobile phone apps, web pages, blogs, emails, SMS text messages, social media, and similar technologies.
2. The intervention must be health related, encompassing areas such as health behavior improvement, disease treatment, and health education.
3. We selected studies that used randomized controlled trials (RCTs) as the intervention as eligible, including both individual and cluster RCTs, as they provide the highest level of evidence for evaluating interventions.

The second round of screening and selection concentrated on identifying specific strategies within the selected DHIs. A strategy is defined as a specific approach or technique used within the intervention to promote health behavior change or improve health outcomes. We used the framework of potential strategies (Multimedia Appendix 3) based on behavior change techniques [39] and behavior strategies [30].

The third round of screening and selection focused on identifying the population with T2DM based on the second round of work. A data extraction form was developed to facilitate electronic comparison of entries. The extracted data include the author, year of publication, study setting (ie, country), characteristics of the participants, details of the interventions, strategies, and outcomes. The inclusion criteria were as follows:

1. Population—patients diagnosed with T2DM and aged ≥18 years
2. Comparisons—the effectiveness of DHIs was compared to no intervention
3. Outcomes—the outcomes for the overview of the effectiveness of DHI strategies included, but were not limited to, acceptability, usability, satisfaction, appropriateness, and efficiency. The outcome of the meta-analysis included the changes in HbA_{1c} (%) level, FBG (mmol/L) level, BMI (kg/m^2), and weight loss (kg).
4. Study design—we only included RCTs.

During the second and third screening, full manuscripts of the studies identified as potentially relevant were obtained and assessed by 6 independent reviewers according to the inclusion criteria. Any discrepancies were resolved by discussion or through adjudication by a senior researcher.

Quality Appraisal

Two independent reviewers assessed the individual quality of the final selected studies using the Cochrane Collaboration Risk of Bias tool [40], with any discrepancies being resolved by consensus. The quality evaluation items of each trial included selection bias (random sequence generation and allocation concealment), performance bias (blinding of participants and personnel), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), reporting bias (selective reporting), and other bias. Each item was scored as low, high, or unclear risk of bias.

Data Analysis and Synthesis

Methods of Analysis

The variability in some outcomes, such as changes in physical activity, diabetes self-efficacy, medication adherence, quality of life, acceptability, and satisfaction, precluded the possibility of conducting a meta-analysis for these specific measures. Therefore, we reported a structured analysis of the findings to draw conclusions about the effectiveness of different DHI strategies on T2DM management. If a certain outcome measure had a statistically significant ($P<.05$) improvement compared to the control group or over time, it was considered effective. If the outcome measures did not show significant changes over time or there was no statistically significant difference from the control group, it was considered that there was not enough evidence to prove their effectiveness. The synthesis framework for the assessment of health IT (SF/HIT) was used to structure the evaluation of the studies because it included a whole system set of outcome variables [41]. These comprised variables such as adherence or attendance, acceptability, effectiveness, satisfaction, and perceived ease of use or usefulness. Following

XSL•FO
RenderX

the framework, evidence for each of the outcome variables was coded as *positive or mixed* or *neutral or negative.* If the study did not address the outcome in question, it was coded as neutral or negative.

For the meta-analysis, the outcome included HbA_{1c} levels, FBG levels, BMI, and weight loss. These were continuous variables, and thus, network estimates were presented as mean differences (MDs) with 95% CIs. We assumed that a P value<.05 indicated statistical significance. We measured the level of heterogeneity with the I^2 statistics; an I^2<50% was considered to have no significant heterogeneity, in which case we would use a fixed-effects model to calculate the pooled effect sizes. Otherwise, a random-effects model would be used. Stata software (version 15.1, *network package* and *network graphs package*; StataCorp) was used to conduct network meta-analysis [42,43]. The *network package* performed the network meta-analysis based on the frequentist framework using random-effects models. The approach was to test the research hypothesis, as this was simpler than the problem of establishing prior probability [44]. This approach is not complex and has few limitations for ordinary researchers using network meta-analysis [45]. A network diagram with nodes and lines was constructed to represent different interventions, where the size of nodes represents the number of populations, and the thickness of lines between nodes represents the number of studies. The results of the network meta-analysis were summarized based on all possible pairwise comparisons, including mixed comparisons (ie, the combined effect of direct and indirect comparisons) and indirect comparisons. The effect of different interventions was estimated based on the surface under the cumulative ranking curve (SUCRA). The SUCRA value ranges from 0% to 100%, where a SUCRA value of 100% indicates that the treatment was the most effective, and the smaller the value, the poorer the treatment effect.

Assessment of Inconsistency

The node-splitting test was used to assess the local inconsistency between direct and indirect comparisons. Differences between direct and indirect coefficients (assessed via the P value) were used to estimate inconsistency: if P<.05, local inconsistency was considered to exist [46]. If inconsistency was observed, nontransitivity was also suspected to exist, and potential modifiers influencing treatment effects were examined.

Risk of Bias Across Studies

The risk of publication bias in network meta-analysis was analyzed using the Egger test [47]. The symmetry of the generated funnel plots was assessed visually using the Egger test, together with adjusted rank correlation and regression asymmetry tests [48,49].

Sensitivity Analyses

We used the method of removing individual studies separately.

Results

Characteristics of Included Studies

The schematic flow for the selection of the included studies is presented in Figure 1. A total of 12,372 studies were identified, of which only 52 (0.42%) trials [14,50-100] met the eligibility criteria and were included. The characteristics of the included studies are provided in Multimedia Appendix 4 [14,50-100]. Among the 52 included studies from the literature, the publication year was between 2000 and 2023. The number of publications has also increased over the years, with the highest publication volume in 2022 and 2023. Most research was conducted in the United States (20/52, 38%), followed by China (5/52, 10%). The minimum sample size was 30, and the maximum was 1926. The duration of intervention ranged from 3 to 24 months, with most studies lasting 12 months (15/52, 29%). The outcome indicators in intervention mainly included HbA_{1c} levels, FBG levels, BMI, blood pressure, weight loss, waist circumference, low-density lipoprotein cholesterol levels, high-density lipoprotein cholesterol levels, total cholesterol levels, diabetes self-efficiency, diabetes medication adherence, quality of life, self-management, depression, diabetes distress, and other outcomes.

XSL•FO
RenderX

Figure 1. Flow diagram for the search and selection of the included studies.

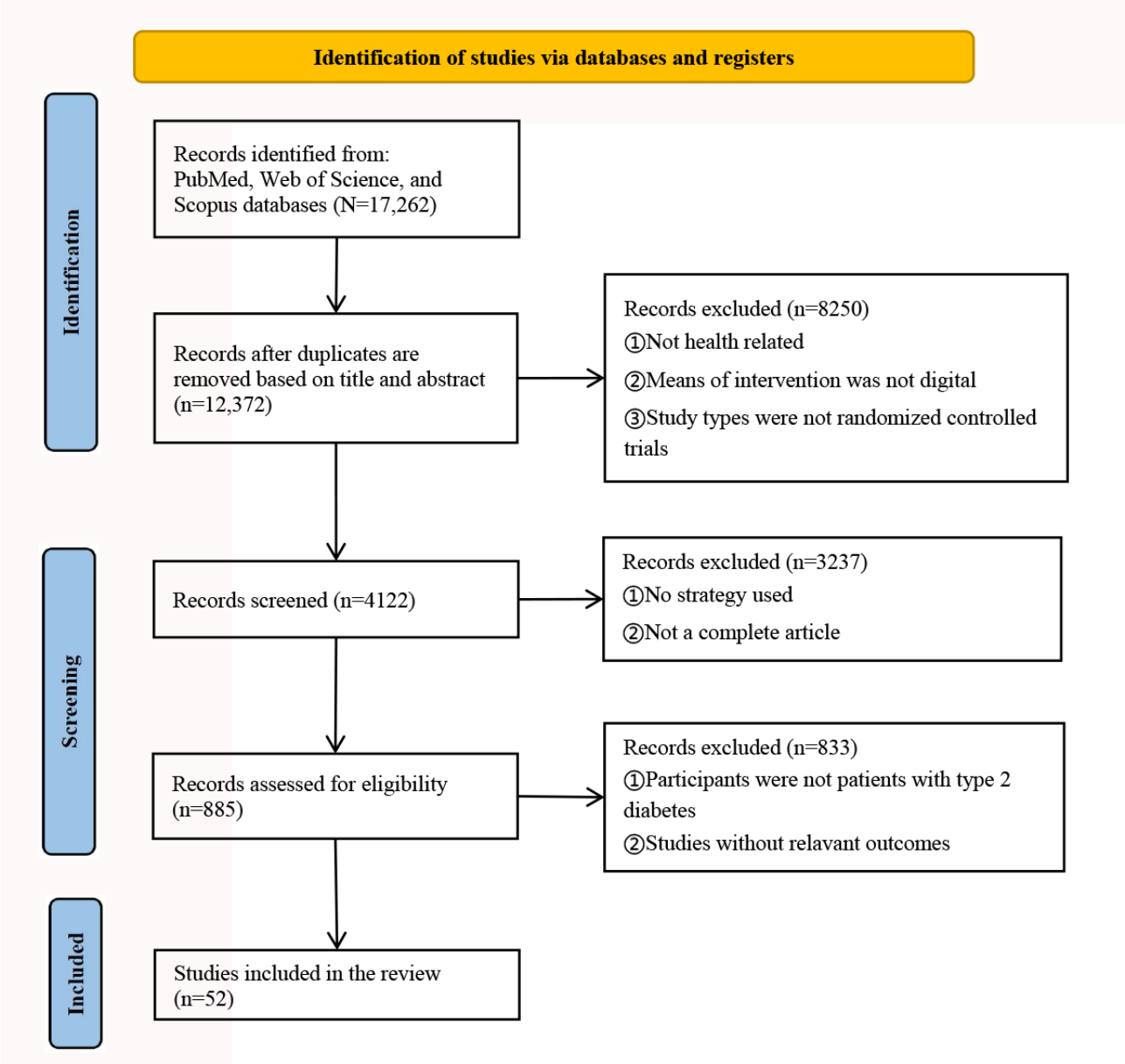

Synthesis of Different DHI Strategies

Among the 52 studies included, we identified 63 different strategies, and they were categorized into 19 themes. Among the identified strategies, *monitor* and *guide* were the most frequently used, appearing in 77% (40/52) and 67% (35/52) of the total papers, respectively. Following closely was *management* (32/52, 61%) and *engagement* (28/52, 54%). More than 40% of the studies used *stimulate*,*communication*, and *goal setting*. In addition, more than one-fifth of the research used *support*,*shape*, *feedback*, *prompt*, *action*, and *tailor*. The usage frequencies were notably lower for cues, *identity*, *reward*, *model or demonstrate*, and *restructure*. The studies ranged from using a minimum of 1 to a maximum of 32 strategies. We categorized studies into 3 groups based on the number of strategies used: low (1 to 3 strategies), medium (4 to 6 strategies), and high (≥7 strategies). Among the 52 studies, 18 (35%) were categorized as high-strategy study, 23 (44%) as medium-strategy study, and 11 (21%) as low-strategy study. In these 52 studies, 37 different combinations of strategies were identified based on the thematic strategies used in each study. Detailed information about these strategies and strategy combinations is provided in Multimedia Appendix 5. The strategy themes and the number of identified studies are listed in Table 1.

XSL•FO
RenderX

Table 1. The strategy themes and the number of identified studies (N=52).

	Theme	Studies, n (%)
A	Action planning	12 (23)
B	Communication	23 (44)
C	Cues	4 (8)
D	Engagement	28 (54)
E	Feedback	13 (25)
F	Goal setting	24 (46)
G	Guide	35 (67)
H	Identity	4 (8)
I	Management	32 (61)
J	Model or demonstrate	2 (4)
K	Monitor	40 (77)
L	Prompt	12 (23)
M	Restructure	2 (4)
N	Reward	4 (8)
O	Shape	15 (29)
P	Stimulate	22 (42)
Q	Support	19 (36)
R	Tailor	11 (21)
S	Others	7 (13)

Evaluation of the Effectiveness of Different DHI Strategies

In this assessment, we encoded positive or mixed as 1 and neutral or negative as 0. Overall, most studies have reported positive or mixed outcomes for most outcome indicators. The number of positive or mixed results that studies achieved ranged from 6 to 11. In total, 41 (79%) of the 52 studies have reported having ≥10 positive or mixed results as stipulated in the SF/HIT. For the preventive care; efficiency; perceived ease of use or usefulness; safety, privacy, or security; acceptability; appropriateness; and satisfaction domains nearly all studies (>50) have reported positive or mixed results. Following closely were the process of service delivery or performance, effectiveness, and adherence or attendance domains, with 49 (94%), 47 (90%), and 45 (86%) of the 52 studies reporting positive or mixed results. However, in terms of cost-effectiveness, most studies (35/52, 67%) either did not report relevant results or reported neutral or negative findings. Out of 18 high-strategy studies, 16 (89%) reported ≥10 positive or mixed results. Out of 23 medium-strategy studies, 20 (87%) reported ≥10 positive or mixed results. However, 7 (64%) out of the 11 low-strategy studies reported having ≥10 positive or mixed results. Meanwhile, it is noteworthy that the 2 studies with the fewest positive or mixed results were both high-strategy studies. The evaluation summary of SF/HIT results is presented in Multimedia Appendix 6.

Risk of Bias and Quality Assessments of Included Studies

We conducted a synthesis of bias risk assessment results using RevMan software (version 5.4; Cochrane), as shown in Figure 2. Overall, all included RCTs exhibited a relatively low risk of bias. All studies reported the generation of random sequences and described specific randomization methods assessed as low risk. There were no incomplete outcome data across all studies assessed as low risk. Selective reporting was not observed in any of the studies assessed as low risk. Regarding allocation concealment, 36% (19/52) of the studies reported using this method, assessed as low risk; 54% (28/52) did not provide such reporting, assessed as unclear risk; and 10% (5/52) explicitly stated not using allocation concealment, assessed as high risk. For blinding of participants and personnel, most studies (41/52, 79%) did not use blinding, assessed as high risk, often due to the nature of the intervention precluding blinding; 15% (8/52) of the studies implemented blinding for participants, assessed as low risk, while 6% (3/52) did not report blinding status, assessed as unclear risk. For the blinding of outcome assessment, 46% (24/52) of the studies did not use blinding, resulting in high risk, whereas 33% (17/52) reported using blinding for outcome assessment, assessed as low risk, with the remainder (11/52, 21%) assessed as unclear risk. The risk of bias in each study is presented in Multimedia Appendix 7.

XSL•FO
RenderX

Figure 2. Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.

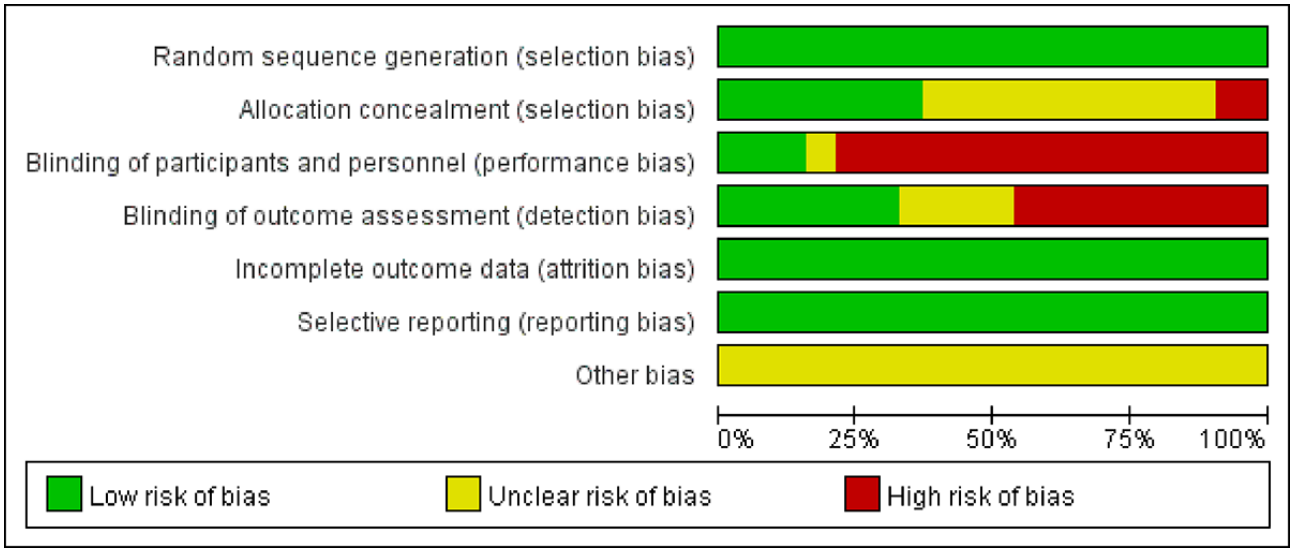

Meta-Analysis

Effects of Strategy Combinations in Reducing HbA1c Levels

Of 52 included RCTs, 27 (52%) RCTs [14,50,53-55,57,59,60,66,68,70,72,74,77,78,80,82,84,85,87,88,93,94,96-98,100] assessed the effects of 12 different DHI strategy combinations on HbA_{1c} levels in patients with T2DM. These 12 strategy combinations encompassed (1) ABCDEFGHIJKLMNOQR (action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor); (2) ABDEFGIKLOPQ (action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support); (3) FGIKQ (goal setting, guide, management, monitor, and support); (4) BGKR (communication, guide, monitor, and tailor); (5) EGIKPR (feedback, guide, management, monitor, stimulate, and tailor); (6) EKP (feedback, monitor, and stimulate); (7) DGIKPQR (engagement, guide, management, monitor, stimulate, support, and tailor); (8) BDFIQ (communication, engagement, goal setting, management, and support); (9) BDGI (communication, engagement, guide, and management); (10) BDGKNOP (communication, engagement, guide, monitor, reward, shape, and stimulate); (11) DFGKP (engagement, goal setting, guide, monitor, and stimulate); and (12) GIK (guide, management, and monitor). The specific strategy represented by the letters in each strategy combination is provided in Table 1.

The network evidence plot for HbA_{1c} levels is shown in Figure 3A. The SUCRA probability ranking for the reducing effect of HbA_{1c} levels in different DHI strategy combinations is shown in Figure 4A. Of the 12 DHI strategy combinations, the possibility of DHI-BDGI (communication, engagement, guide, and management) being the best strategy combination was the highest. The SUCRA value predicted the possibility of different strategy combinations as the best way, and the effects were ranked as follows: DHI-BDGI (91.8%)>DHI-DFGKP (76%)>DHI-DGIKPQR (73.2%)>DHI-FGIKQ (67.4%)>DHI-ABDEFGIKLOPQ (62.1%) > DHI-EGIKPR (52.8%)>DHI-GIK (52.1%)>DHI-EKP (40.5%) > DHI-BDGKNOP (39.4%) > DHI-ABCDEFGHIJKLMNOQR (30.1%)>DHI-BDFIQ (27.7%)>DHI-BGKR (24.1%)>usual (12.8%). Compared with the usual care groups, DHI-BDGI (MD –1.04, 95% CI –1.55 to –0.54), DHI-DFGKP (MD –0.76, 95% CI –1.36 to –0.16), DHI-FGIKQ (MD –0.61, 95% CI –1.01 to –0.22), DHI-ABDEFGIKLOPQ (MD –0.55, 95% CI –0.92 to –0.18), and DHI-GIK (MD –0.44, 95% CI –0.85 to –0.03) strategy combinations were statistically significantly effective in reducing HbA_{1c} levels (Figure 5A). On the basis of the interval estimation of direct and indirect comparison (Table 2), DHI-BDGI was more effective than the DHI-EKP, DHI-BDFIQ, and DHI-BGKR strategy combinations.

XSL•FO
RenderX

Figure 3. Evidence network of outcome indicators: (A) evidence network for glycated hemoglobin A1c, (B) evidence network for fasting blood glucose, (C) evidence network for BMI, and (4) evidence network for weight loss. ABDEFGIKLOPQ: action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; ABCDEFGHIJKLNO: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, reward, and shape; ABCDEFGHIJKLMNOQR: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor; BDFGIKLOPQ: communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; BDFIQ: communication, engagement, goal setting, management, and support; BDGI: communication, engagement, guide, and management; BDGKNOP: communication, engagement, guide, monitor, reward, shape, and stimulate; BGKR: communication, guide, monitor, and tailor; DFGKP: engagement, goal setting, guide, monitor, and stimulate; DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor; DHI: digital health intervention; EGIKPR: feedback, guide, management, monitor, stimulate, and tailor; EKP: feedback, monitor, and stimulate; FGIKQ: goal setting, guide, management, monitor, and support; GIK: guide, management, and monitor.

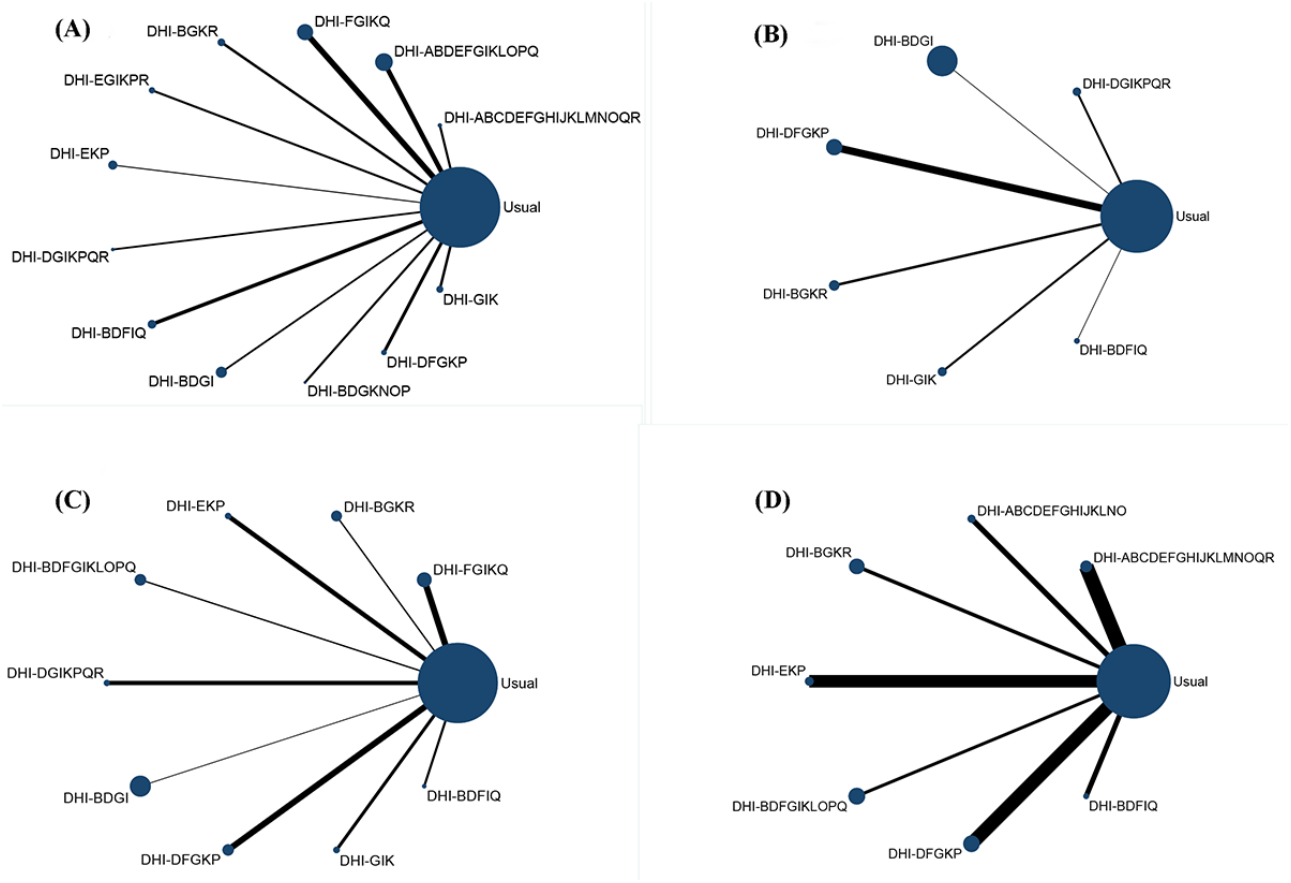

XSL•FO
RenderX

Figure 4. Surface under the cumulative ranking curve rank of outcome indicators: (A) surface under the cumulative ranking curve rank for glycated hemoglobin A1c, (B) surface under the cumulative ranking curve rank for fasting blood glucose, (C) surface under the cumulative ranking curve rank for BMI, and (D) surface under the cumulative ranking curve rank for weight loss. ABDEFGIKLOPQ: action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; ABCDEFGHIJKLNO: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, reward, and shape; ABCDEFGHIJKLMNOQR: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor; BDFGIKLOPQ: communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; BDFIQ: communication, engagement, goal setting, management, and support; BDGI: communication, engagement, guide, and management; BDGKNOP: communication, engagement, guide, monitor, reward, shape, and stimulate; BGKR: communication, guide, monitor, and tailor; DFGKP: engagement, goal setting, guide, monitor, and stimulate; DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor; DHI: digital health intervention; EGIKPR: feedback, guide, management, monitor, stimulate, and tailor; EKP: feedback, monitor, and stimulate; FGIKQ: goal setting, guide, management, monitor, and support; GIK: guide, management, and monitor.

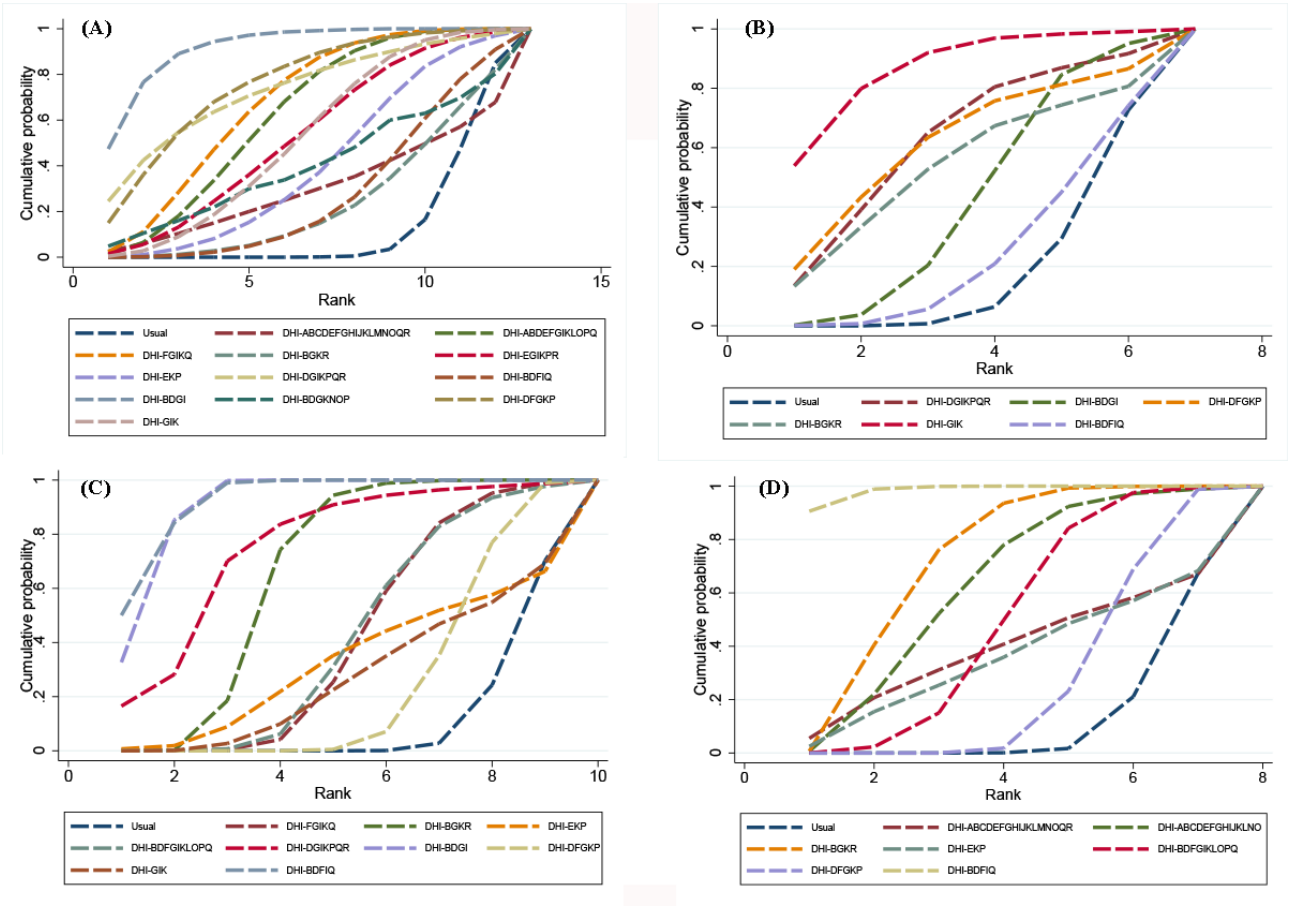

XSL•FO
RenderX

Figure 5. Forest plots of network meta-analysis results: (A) forest plots for glycated hemoglobin A1c, (B) forest plots for fasting blood glucose; (C) forest plots for BMI, and (D) forest plots for weight loss. ABDEFGIKLOPQ: action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; ABCDEFGHIJKLNO: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, reward, and shape; ABCDEFGHIJKLMNOQR: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor; BDFGIKLOPQ: communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; BDFIQ: communication, engagement, goal setting, management, and support; BDGI: communication, engagement, guide, and management; BDGKNOP: communication, engagement, guide, monitor, reward, shape, and stimulate; BGKR: communication, guide, monitor, and tailor; DFGKP: engagement, goal setting, guide, monitor, and stimulate; DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor; DHI: digital health intervention; DL: DerSimonian and Laird method; EGIKPR: feedback, guide, management, monitor, stimulate, and tailor; EKP: feedback, monitor, and stimulate; FGIKQ: goal setting, guide, management, monitor, and support; GIK: guide, management, and monitor; IV: inverse variance method; MD: mean difference.

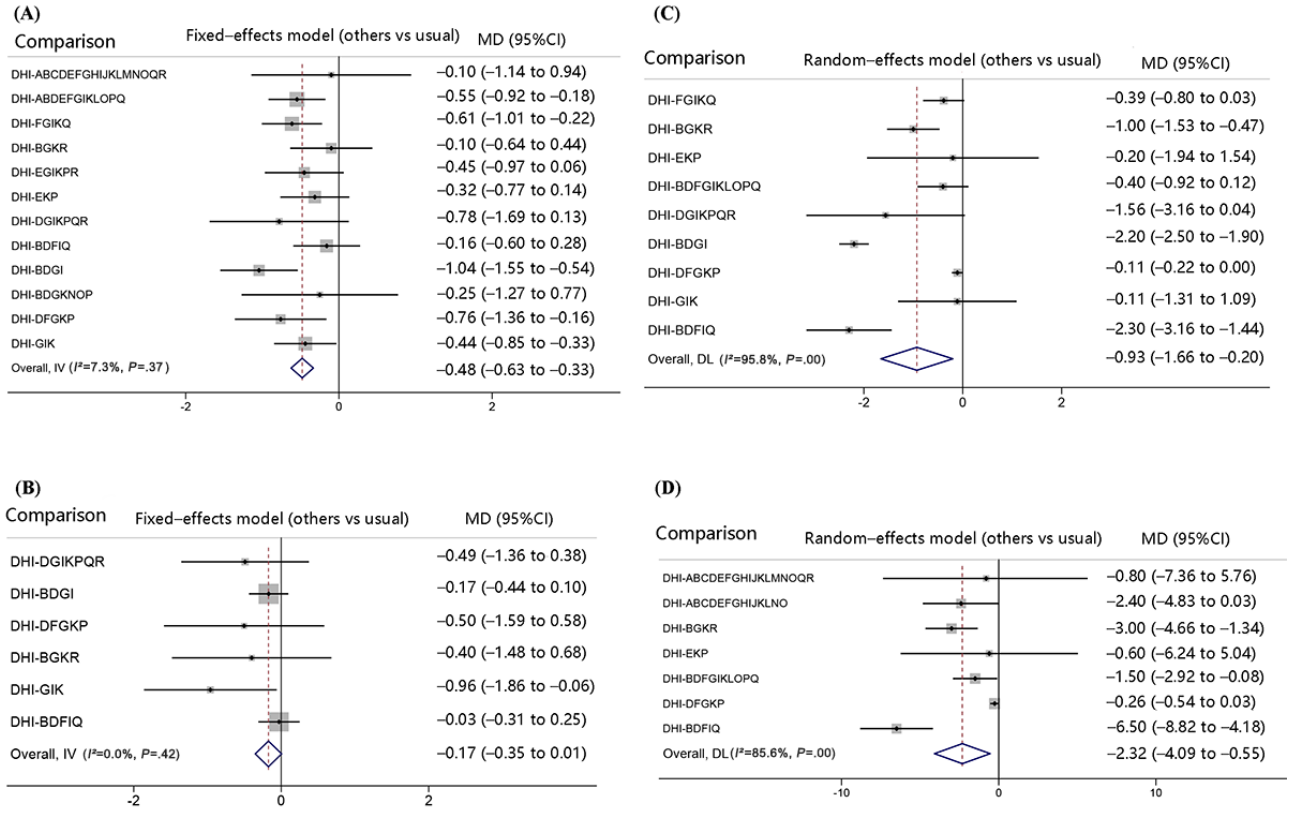

XSL•FO
RenderX

Table 2. Network meta-analysis of glycated hemoglobin A1c levels.

DHI[a] strategy combinations	MDs[b] (95% CIs)												
	DHI-BD-GI[c]	DHI-DFGKP[d]	DHI-DGIKPQR[e]	DHI-FGIKQ[f]	DHI-ABDE-FGIK-LOPQ[g]	DHI-EGIKPR[h]	DHI-GIK[i]	DHI-EKP[j]	DHI-BDG-KNOP[k]	DHI-ABCDE-FGHI-JKLMNO-QR[l]	DHI-BD-FIQ[m]	DHI-BGKR[n]	Usu-al
DHI-DFGKP	–0.28 (–1.06 to 0.50)	0	__[o]	—	—	—	—	—	—	—	—	—	—
DHI-DGIKPQR	–0.26 (–1.30 to 0.78)	0.02 (–1.07 to 1.11)	0	—	—	—	—	—	—	—	—	—	—
DHI-FGIKQ	–0.43 (–1.07 to 0.21)	–0.15 (–0.86 to 0.57)	–0.17 (–1.16 to 0.82)	0	—	—	—	—	—	—	—	—	—
DHI-ABDEFGIK-LOPQ	–0.50 (–1.12 to 0.13)	–0.21 (–0.92 to 0.49)	–0.23 (–1.21 to 0.75)	–0.07 (–0.61 to 0.48)	0	—	—	—	—	—	—	—	—
DHI-EGIKPR	–0.59 (–1.31 to 0.13)	–0.31 (–1.10 to 0.48)	–0.33 (–1.37 to 0.72)	–0.16 (–0.81 to 0.49)	–0.10 (–0.73 to 0.54)	0	—	—	—	—	—	—	—
DHI-GIK	–0.60 (–1.24 to 0.03)	–0.32 (–1.05 to 0.40)	–0.34 (–1.34 to 0.65)	–0.17 (–0.74 to 0.39)	–0.11 (–0.66 to 0.44)	–0.01 (–0.67 to 0.64)	0	—	—	—	—	—	—
DHI-EKP	–0.73 (–1.41 to –0.05)	–0.45 (–1.19 to 0.30)	–0.46 (–1.48 to 0.55)	–0.30 (–0.90 to 0.30)	–0.23 (–0.82 to 0.35)	–0.14 (–0.82 to 0.55)	–0.12 (–0.73 to 0.48)	0	—	—	—	—	—
DHI-BDGKNOP	–0.79 (–1.93 to 0.35)	–0.51 (–1.69 to 0.67)	–0.53 (–1.90 to 0.84)	–0.36 (–1.46 to 0.73)	–0.30 (–1.38 to 0.79)	–0.20 (–1.35 to 0.94)	–0.19 (–1.29 to 0.91)	–0.07 (–1.18 to 1.05)	0	—	—	—	—
DHI-ABCDE-FGHIJKLMNO-QR	–0.94 (–2.10 to 0.22)	–0.66 (–1.86 to 0.54)	–0.68 (–2.06 to 0.70)	–0.51 (–1.63 to 0.60)	–0.45 (–1.56 to 0.66)	–0.35 (–1.52 to 0.81)	–0.34 (–1.46 to 0.78)	–0.22 (–1.35 to 0.92)	–0.15 (–1.61 to 1.31)	0	—	—	—
DHI-BDFIQ	–0.88 (–1.55 to –0.22)	–0.60 (–1.34 to 0.14)	–0.62 (–1.63 to 0.39)	–0.45 (–1.04 to 0.13)	–0.39 (–0.96 to 0.18)	–0.29 (-0.97 to 0.38)	–0.28 (–0.87 to 0.31)	–0.16 (–0.78 to 0.47)	–0.09 (–1.20 to 1.02)	0.06 (–1.07 to 1.19)	0	—	—
DHI-BGKR	–0.94 (–1.68 to –0.21)	–0.66 (–1.46 to 0.14)	–0.68 (–1.73 to 0.37)	–0.51 (–1.18 to 0.15)	–0.45 (–1.10 to 0.20)	−0.35 (–1.10 to 0.39)	–0.34 (–1.01 to 0.33)	–0.22 (–0.92 to 0.49)	–0.15 (–1.30 to 1.00)	0.00 (–1.17 to 1.17)	–0.06 (–0.75 to 0.63)	0	—
Usual	–1.04 (–1.55 to –0.54)	–0.76 (–1.36 to –0.16)	–0.78 (–1.69 to 0.13)	–0.61 (–−1.01 to –0.22)	–0.55 (–0.92 to –0.18)	–0.45 (–0.97 to 0.06)	–0.44 (–0.85 to –0.03)	–0.32 (–0.77 to 0.14)	–0.25 (–1.27 to 0.77)	–0.10 (–1.14 to 0.94)	–0.16 (–0.60 to 0.28)	–0.10 (–0.64 t 0.44)	0

[a]DHI: digital health intervention.

[b]MD: mean difference.

[c]BDGI: communication, engagement, guide, and management.

[d]DFGKP: engagement, goal setting, guide, monitor, and stimulate.

XSL•FO
RenderX

[e]DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor.

[f]FGIKQ: goal setting, guide, management, monitor, and support.

[g]ABDEFGIKLOPQ: action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support.

[h]EGIKPR: feedback, guide, management, monitor, stimulate, and tailor.

[i]GIK: guide, management, and monitor.

[j]EKP: feedback, monitor, and stimulate.

[k]BDGKNOP: communication, engagement, guide, monitor, reward, shape, and stimulate.

[l]ABCDEFGHIJKLMNOQR: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor.

[m]BDFIQ: communication, engagement, goal setting, management, and support.

[n]BGKR: communication, guide, monitor, and tailor.

[o]Not applicable.

Effects of Strategy Combinations on FBG Levels

Of 52 included RCTs, 7 (13%) RCTs [74,78,84,85,87,97,98] assessed the effects of 6 different DHI strategy combinations on FBG levels in patients with T2DM. These 6 strategy combinations encompassed (1) DGIKPQR (engagement, guide, management, monitor, stimulate, support, and tailor); (2) BDGI (communication, engagement, guide, and management); (3) DFGKP (engagement, goal setting, guide, monitor, and stimulate); (4) BGKR (communication, guide, monitor, and tailor); (5) GIK (guide, management, and monitor); and (6) BDFIQ (communication, engagement, goal setting, management, and support).

The network evidence plot for FBG levels is shown in Figure 3B. The SUCRA probability ranking for the reducing effect of FBG levels of different DHI strategy combinations is shown in Figure 4B. Of the 6 strategy combinations, the possibility of DHI-GIK (guide, management, and monitor) being the best strategy combination was the highest. The SUCRA value predicted the possibility of different strategy combinations as the best way, and the effects were ranked as follows: DHI-GIK (86.7%)>DHI-DGIKPQR (62.8%)>DHI-DFGKP (61.6%)>DHI-BGKR (53.6%)>DHI-BDGI (42.7%)>DHI-BDFIQ (24.4%)>usual (18.2%). Compared with the usual care groups, only the DHI-GIK strategy combination (MD –0.96, 95% CI –1.86 to –0.06) was statistically significantly effective in reducing FBG levels (Figure 5B). On the basis of the interval estimation of direct and indirect comparison (Table 3), no strategy combination was superior to the others.

Table 3. Network meta-analysis of fasting blood glucose levels.

DHI[a] strategy combinations	MDs[b] (95% CIs)						
	DHI-GIK[c]	DHI-DGIKPQR[d]	DHI-DFGKP[e]	DHI-BGKR[f]	DHI-BDGI[g]	DHI-BDFIQ[h]	Usual
DHI-DGIKPQR	–0.47 (–1.72 to 0.78)	0	__[i]	—	—	—	—
DHI-DFGKP	–0.46 (-1.87 to 0.95)	0.01 (–1.38 to 1.40)	0	—	—	—	—
DHI-BGKR	–0.56 (–1.97 to 0.85)	–0.09 (–1.47 to 1.29)	–0.10 (–1.64 to 1.43)	0	—	—	—
DHI-BDGI	–0.79 (–1.73 to 0.15)	–0.32 (–1.23 to 0.59)	–0.33 (–1.45 to 0.79)	–0.23 (–1.34 to 0.88)	0	—	—
DHI-BDFIQ	–0.93 (–1.87 to 0.01)	–0.46 (–1.37 to 0.45)	–0.47 (–1.60 to 0.65)	–0.37 (–1.49 to 0.75)	–0.14 (–0.53 to 0.25)	0	—
Usual	–0.96 (–1.86 to –0.06)	–0.49 (–1.36 to 0.38)	–0.50 (–1.59 to 0.58)	–0.40 (–1.48 to 0.68)	–0.17 (–0.44 to 0.10)	–0.03 (–0.31 to 0.25)	0

[a]DHI: digital health intervention.

[b]MD: mean difference.

[c]GIK: guide, management, and monitor.

[d]DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor.

[e]DFGKP: engagement, goal setting, guide, monitor, and stimulate.

[f]BGKR: communication, guide, monitor, and tailor.

[g]BDGI: communication, engagement, guide, and management.

[h]BDFIQ: communication, engagement, goal setting, management, and support.

[i]Not applicable.

XSL•FO
RenderX

Effects of Strategy Combinations on BMI

Of 52 included RCTs, 11 (21%) RCTs [60,66,70,72-74,78,84,85,97,98] assessed the effects of 9 different DHI strategy combinations on BMI levels in patients with T2DM. These 9 strategy combinations encompassed (1) FGIKQ (goal setting, guide, management, monitor, and support); (2) BGKR (communication, guide, monitor, and tailor); (3) EKP (feedback, monitor, and stimulate); (4) BDFGIKLOPQ (communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support); (5) DGIKPQR (engagement, guide, management, monitor, stimulate, support, and tailor); (6) BDGI (communication, engagement, guide, and management); (7) DFGKP (engagement, goal setting, guide, monitor, and stimulate); (8) GIK (guide, management, and monitor); and (9) BDFIQ (communication, engagement, goal setting, management, and support).

The network evidence plot for BMI is shown in Figure 3C. The SUCRA probability ranking for the reducing effect of BMI of different DHI strategy combinations is shown in Figure 4C. Of the 9 strategy combinations, DHI-BDFIQ (communication, engagement, goal setting, management, and support) was most likely to be the optimal strategy combination. On the basis of the SUCRA values, the effects were ranked as follows: DHI-BDFIQ (92.6%)>DHI-BDGI (90.9%)>DHI-DGIKPQR (75.2%)>DHI-BGKR (65.1%)>DHI-BDFGIKLOPQ (41.4%)>DHI-FGIKQ (40.9%)>DHI-EKP (32%)>DHI-GIK (26.8%)>DHI-DFGKP (24.3%)>Usual (10.8%). Compared with the usual care groups, DHI-BDFIQ (MD –2.30, 95% CI –3.16 to –1.44), DHI-BDGI (MD –2.20, 95% CI –2.50 to –1.90) and DHI-BGKR (MD –1.00, 95% CI –1.53 to –−0.47) strategy combinations were statistically significantly effective in reducing BMI (Figure 5C). On the basis of the interval estimation of direct and indirect comparison (Table 4), DHI-BDFIQ was superior to all other strategy combinations except for DHI-BDGI and DHI-DGIKPQR. In addition, except for DHI-BDFIQ and DHI-DGIKPQR, DHI-BDGI was superior to the rest of strategy combinations. DHI-BGKR was also superior to the DHI-DFGKP strategy combination.

XSL•FO
RenderX

Table 4. Network meta-analysis of BMI.

DHI[a] strategy combinations	MDs[b] (95% CIs)									
	DHI-BD-FIQ[c]	DHI-BD-GI[d]	DHI-DGIKPQR[e]	DHI-BGKR[f]	DHI-BD-FGIK-LOPQ[g]	DHI-FGIKQ[h]	DHI-EKP[i]	DHI-GIK[j]	DHI-DFGKP[k]	Usual
DHI-BDGI	–0.10 (–1.01 to 0.81)	0	—[l]	—	—	—	—	—	—	—
DHI-DGIKPQR	–0.74 (–2.56 to 1.08)	–0.64 (–2.27 to 0.99)	0	—	—	—	—	—	—	—
DHI-BGKR	–1.30 (–2.31 to –0.29)	–1.20 (–1.81 to –0.59)	–0.56 (–2.25 to 1.13)	0	—	—	—	—	—	—
DHI-BDFGIK-LOPQ	–1.90 (–2.90 to –0.90)	–1.80 (–2.40 to –1.20)	–1.16 (–2.85 to 0.53)	–0.60 (–1.34 to 0.14)	0	—	—	—	—	—
DHI-FGIKQ	–1.91 (–2.87 to –0.96)	–1.81 (–2.33 to –1.30)	–1.17 (–2.83 to 0.48)	–0.61 (–1.29 to 0.06)	–0.01 (–0.68 to 0.65)	0	—	—	—	—
DHI-EKP	–2.10 (–4.04 to –0.16)	–2.00 (–3.76 to –0.24)	–1.36 (–3.72 to 1.00)	–0.80 (–2.62 to 1.02)	–0.20 (–2.01 to 1.61)	–0.19 (1.97 to 1.60)	0	—	—	—
DHI-GIK	−2.19 (−3.67 to −0.71)	−2.09 (−3.33 to −0.85)	–1.45 (–3.45 to 0.55)	–0.89 (–2.20 to 0.42)	–0.29 (–1.60 to 1.02)	–0.28 (–1.55 to 0.99)	–0.09 (–2.20 to 2.02)	0	—	—
DHI-DFGKP	−2.19 (−3.06 to −1.32)	−2.09 (−2.41 to −1.77)	–1.45 (–3.06 to 0.16)	–0.89 (–1.43 to –0.35)	–0.29 (–0.82 to 0.24)	–0.28 (–0.71 to 0.15)	–0.09 (–1.83 to 1.65)	–0.00 (–1.21 to 1.20)	0	—
Usual	−2.30 (−3.16 to −1.44)	−2.20 (−2.50 to −1.90)	–1.56 (–3.16 to 0.04)	–1.00 (–1.53 to –0.47)	–0.40 (–0.92 to 0.12)	–0.39 (–0.80 to 0.03)	–0.20 (–1.94 to 1.54)	–0.11 (–1.31 to 1.09)	–0.11 (–0.22 to 0.00)	0

[a]DHI: digital health intervention.

[b]MD: mean difference.

[c]BDFIQ: communication, engagement, goal setting, management, and support.

[d]BDGI: communication, engagement, guide, and management.

[e]DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor.

[f]BGKR: communication, guide, monitor, and tailor.

[g]BDFGIKLOPQ: communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support.

[h]FGIKQ: goal setting, guide, management, monitor, and support.

[i]EKP: feedback, monitor, and stimulate.

[j]GIK: guide, management, and monitor.

[k]DFGKP: engagement, goal setting, guide, monitor, and stimulate.

[l]Not applicable.

Effects of Strategy Combinations on Weight Loss

Of 52 included RCTs, 8 (15%) RCTs [50,52,70,72,73,84,85,98] assessed the effects of 7 different DHI strategy combinations on weight loss in patients with T2DM. These 7 strategy combinations encompassed (1) ABCDEFGHIJKLMNOQR (action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor); (2) ABCDEFGHIJKLNO (action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, reward, and shape); (3) BGKR (communication, guide, monitor, and tailor); (4) EKP (feedback, monitor, and stimulate); (5) BDFGIKLOPQ (communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support); (6) DFGKP (engagement, goal setting, guide, monitor, and stimulate); (7) BDFIQ (communication, engagement, goal setting, management, and support).

XSL•FO
RenderX

The network evidence plot for weight loss is shown in Figure 3D. The SUCRA probability ranking for the effect of weight loss of different DHI strategy combinations is shown in Figure 4D. Of the 7 strategy combinations, DHI-BDFIQ (communication, engagement, goal setting, management, support) was most likely to be the optimal strategy combination. On the basis of the SUCRA values, the effects were ranked as follows: DHI-BDFIQ (98.5%)>DHI-BGKR (72.9%)>DHI-ABCDEFGHIJKLNO (63%)>DHI-BDFGIKLOPQ (49.8%)>DHI-ABCDEFGHIJKLMNOQR (39.2%)>DHI-EKP (36.2%)>DHI-DFGKP (27.5%)>Usual (12.9%). Compared with the usual care groups, DHI-BDFIQ (MD –6.50, 95% CI –8.82 to –4.18), DHI-BGKR (MD –3.00, 95% CI –4.66 to –1.34) and DHI-BDFGIKLOPQ (MD –1.50, 95% CI –2.92 to –0.08) strategy combinations were statistically significantly effective in weight loss (Figure 5D). On the basis of the interval estimation of direct and indirect comparison (Table 5), DHI-BDFIQ was superior to DHI-BGKR, DHI-ABCDEFGHIJKLNO, DHI-BDFGIKLOPQ, and DHI-DFGKP strategy combinations. In addition, DHI-BGKR was also superior to the DHI-DFGKP strategy combination.

Table 5. Network meta-analysis of weight loss.

DHI[a] strategy combinations	MDs[b] (95% CIs)							
	DHI-BDFIQ[c]	DHI-BGKR[d]	DHI-ABCDE-FGHIJKLNO[e]	DHI-BDFGIK-LOPQ[f]	DHI-ABCDEFGHI-JKLMNOQR[g]	DHI-EKP[h]	DHI-DFGKP[i]	Usual
DHI-BGKR	–3.50 (–6.35 to –0.65)	0	__[j]	—	—	—	—	—
DHI-ABCDE-FGHIJKLNO	–4.10 (–7.46 to –0.74)	–0.60 (–3.55 to 2.35)	0	—	—	—	—	—
DHI-BDFGIK-LOPQ	–5.00 (–7.71 to –2.29)	–1.50 (–3.69 to 0.69)	–0.90 (–3.72 to 1.92)	0	—	—	—	—
DHI-ABCDE-FGHIJKLMNO-QR	–5.70 (–12.66 to 1.26)	–2.20 (–8.97 to 4.57)	–1.60 (–8.60 to 5.40)	–0.70 (–7.41 to 6.01)	0	—	—	—
DHI-EKP	–5.90 (–12.00 to 0.20)	–2.40 (–8.28 to 3.48)	–1.80 (–7.94 to 4.34)	–0.90 (–6.72 to 4.92)	–0.20 (–8.85 to 8.45)	0	—	—
DHI-DFGKP	–6.24 (–8.58 to –3.91)	–2.74 (–4.43 to –1.05)	–2.14 (–4.59 to 0.31)	–1.24 (–2.69 to 0.20)	–0.54 (–7.11 to 6.03)	–0.34 (–5.99 to 5.30)	0	—
Usual	–6.50 (–8.82 to –4.18)	–3.00 (–4.66 to –1.34)	–2.40 (–4.83 to 0.03)	–1.50 (–2.92 to –0.08)	–0.80 (–7.36 to 5.76)	–0.60 (–6.24 to 5.04)	–0.26 (–0.54 to 0.03)	0

[a]DHI: digital health intervention.

[b]MD: mean difference.

[c]BDFIQ: communication, engagement, goal setting, management, and support.

[d]BGKR: communication, guide, monitor, and tailor.

[e]ABCDEFGHIJKLNO: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, reward, and shape.

[f]BDFGIKLOPQ: communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support.

[g]ABCDEFGHIJKLMNOQR: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor.

[h]EKP: feedback, monitor, and stimulate.

[i]DFGKP: engagement, goal setting, guide, monitor, and stimulate.

[j]Not applicable.

Heterogeneity and Consistency Analysis

The global inconsistency and local inconsistency were assessed using the node-splitting test for inconsistency analysis. The results showed no statistical inconsistency in each outcome comparison (P>.05). High heterogeneity between studies was only observed in BMI and weight loss outcomes. However, we did not find significant publication bias for these 2 outcomes through Egger test (P=.29 for BMI and P=.11 for weight loss). The net split function analyses found no statistically significant inconsistencies when assessing differences between direct and indirect effects. More details are provided in Table 6.

XSL•FO
RenderX

Table 6. The results of heterogeneity and inconsistency.

Outcome	Heterogeneity		Inconsistency (*P* value)
	I^2 (%)	*P* value	
HbA_{1c}[a]	7.3	.37	—[b]
FBG[c]	0.0	.42	—
BMI	95.8	<.001	—
Weight loss	85.6	<.001	—

[a]HbA_{1c}: glycated hemoglobin A_{1c}.

[b]Data not available.

[c]FBG: fasting blood glucose.

Assessment of Publication Bias

We conducted a comparison-adjusted funnel plot of trials included in the network meta-analysis for each result (Figure 6): (A) HbA_{1c}, (B) FBG, (C) BMI, and (D) weight loss. The included studies were generally symmetrically distributed in the upper and middle parts of the funnel and around the left and right sides of the midline, indicating a low possibility of publication bias. Individual studies were distributed at the bottom, which may be related to the small sample size.

XSL•FO
RenderX

Figure 6. Comparison-adjusted funnel plots of trials included in the network meta-analysis: (A) comparison-adjusted funnel plots for glycated hemoglobin A1c, (B) comparison-adjusted funnel plots for fasting blood glucose, (C) comparison-adjusted funnel plots for BMI, and (D) comparison-adjusted funnel plots for weight loss. ABDEFGIKLOPQ: action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; ABCDEFGHIJKLNO: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, reward, and shape; ABCDEFGHIJKLMNOQR: action planning, communication, cues, engagement, feedback, goal setting, guide, identity, management, model or demonstrate, monitor, prompt, restructure, reward, shape, support, and tailor; BDFGIKLOPQ: communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support; BDFIQ: communication, engagement, goal setting, management, and support; BDGI: communication, engagement, guide, and management; BDGKNOP: communication, engagement, guide, monitor, reward, shape, and stimulate; BGKR: communication, guide, monitor, and tailor; DFGKP: engagement, goal setting, guide, monitor, and stimulate; DGIKPQR: engagement, guide, management, monitor, stimulate, support, and tailor; DHI: digital health intervention; EGIKPR: feedback, guide, management, monitor, stimulate, and tailor; EKP: feedback, monitor, and stimulate; FGIKQ: goal setting, guide, management, monitor, and support; GIK: guide, management, and monitor.

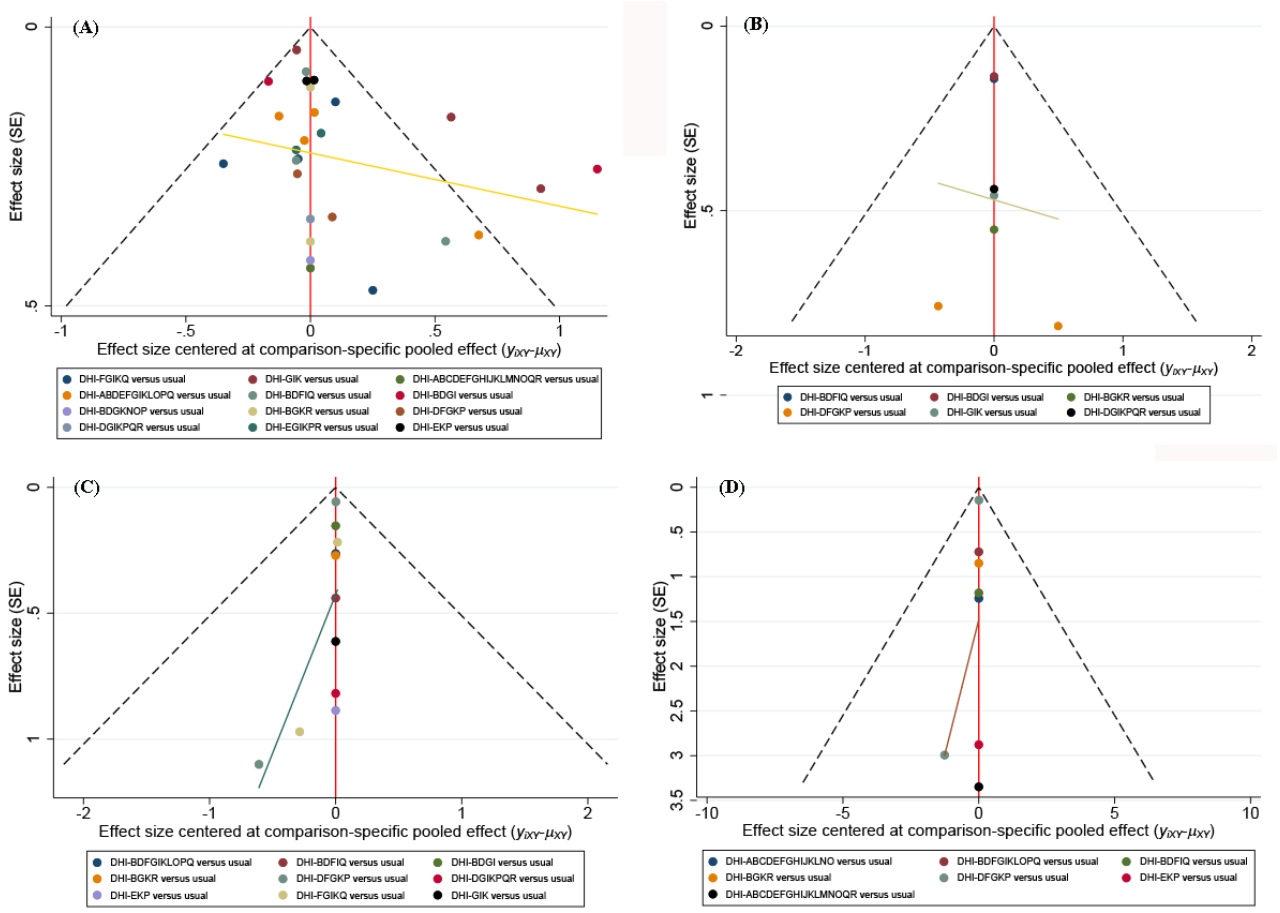

Sensitivity Analysis

We used the method of removing individual studies separately. Our results were generally stable and credible. As shown in Figure 7, the 2 vertical bright lines, except for the vertical axis, represent the CI of the merged population. Each study corresponds to 2 yellow short lines and a yellow circle, representing the CI of the remaining studies combined after excluding this study. If all yellow circles are within the range of 2 vertical bright lines, it indicates that this set of data is stable. Although the results showed that the combined effect value of HbA_{1c} levels, BMI, and weight loss appeared outside the CI of the original combined effect value, all CIs did not include 0; they were significant in treatment effects. The details are shown in Figure 7.

XSL•FO
RenderX

Figure 7. Sensitivity analyses using the method of removing individual studies separately: (A) sensitivity analyses for glycated hemoglobin A1c, (B) sensitivity analyses for fasting blood glucose, (C) sensitivity analyses for BMI, and (D) sensitivity analyses for weight loss.

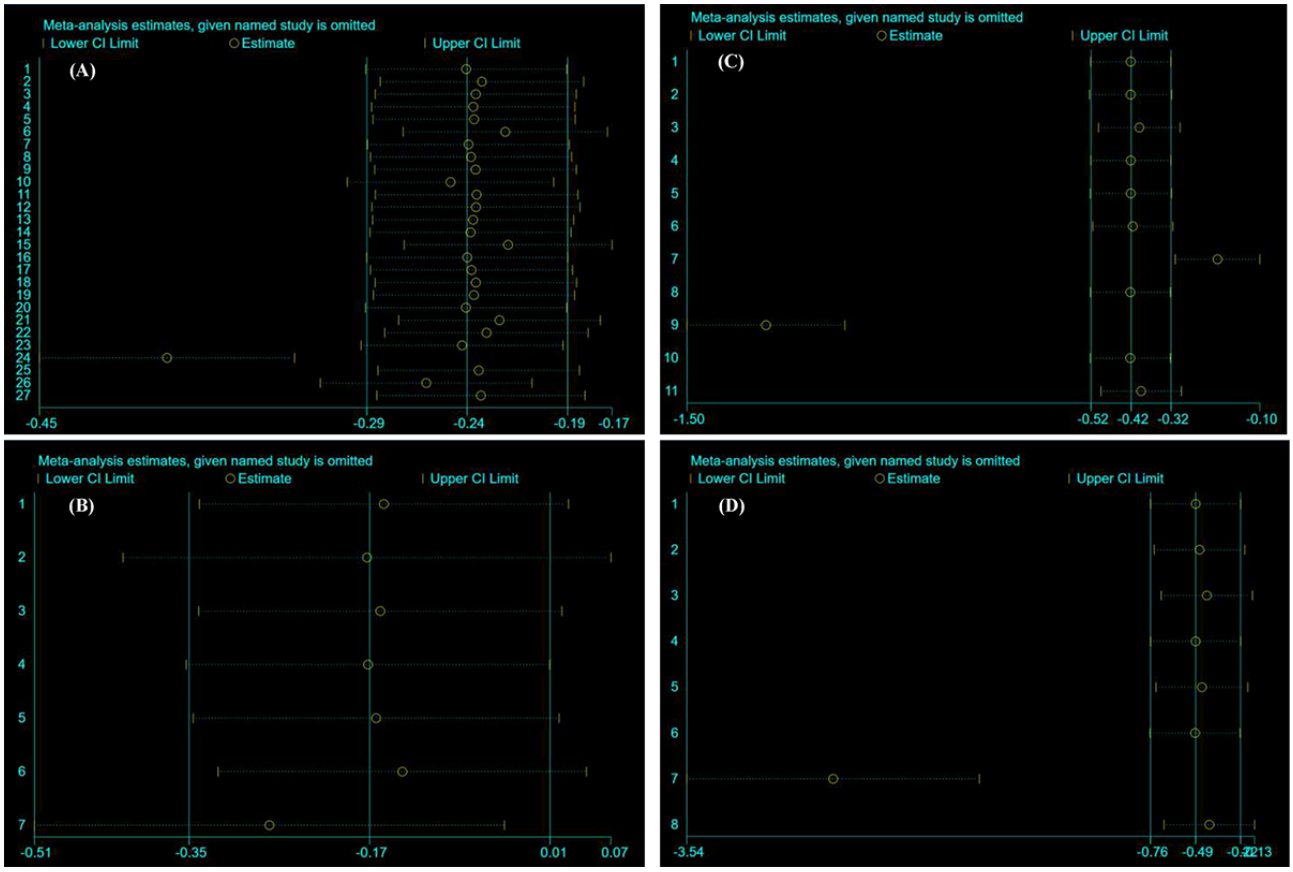

Discussion

Principal Findings

Our study offered a comprehensive insight into the strategies used in DHIs for T2DM management. A total of 52 RCTs were included, identifying 63 strategies categorized into 19 strategy themes. The most commonly used strategies were *guide*, *monitor*, *management*, and *engagement*. Most studies reported positive or mixed outcomes for most indicators based on the SF/HIT. Research involving a medium or high number of strategies was found to be more effective than research involving a low number of strategies. A total of 27 RCTs were included in the network meta-analysis. The strategy combination composed of *communication*, *engagement*, *guide*, and *management* was most effective in reducing HbA_{1c} levels, while the strategy combination that included *guide*, *management*, and *monitor* was effective in reducing FBG levels. A strategy combination composed of *communication*, *engagement*, *goal setting*, *management*, and *support* was most effective for BMI and weight management.

In the identified strategies, *monitor* and *guide* were the most frequently used ones, accounting for 77% (40/52) and 67% (35/52) of the total papers, respectively. Following closely were *management* (32/52, 61%) and *engagement* (28/52, 54%) strategies. The frequency of the use of these strategies indicates the trend and focus in this field. First, the *monitor* strategy reflects the role of digital tools in monitoring patients' physiological indicators and behavior habits in diabetes management. It allows health care professionals and patients to track key parameters such as blood glucose level, weight, and exercise volume in real time, which helps to adjust the treatment plan in time and reduce the risk of complications [101]. Second, the use of a *guide* strategy reflects the value of digital health management in providing personalized and real-time advice and guidance to patients. Through intelligent algorithms and personalized settings, digital platforms can provide patients with accurate nutritional advice, exercise plans, and medication guidance based on their individual characteristics and historical data, helping them better manage their diseases and improve treatment outcomes [102]. These results are consistent with previous research that highlights the crucial role of *monitor* and *guide* strategies in DHIs [103,104]. On this basis, the *management* strategy emphasizes the important role of digital platforms in assisting health care professionals in disease management and treatment decision-making. Through data analysis and predictive models, digital platforms can provide health care professionals with more comprehensive and timely patient information, help them develop more effective treatment plans, and improve the accuracy and efficiency of clinical decision-making [105]. Furthermore, the *engagement* strategy emphasizes the promoting role of digital health management in patient participation and self-management. By providing personalized educational content, social support, and behavioral motivation, digital platforms can stimulate patients' treatment motivation and enhance their awareness and self-management abilities toward the disease, thereby improving treatment compliance and long-term efficacy [106].

XSL•FO
RenderX

The number of strategies used in each study varies greatly, ranging from 1 to 32. Among 52 studies, 18 (35%) were categorized as high-strategy study, 23 (44%) as medium-strategy study, and 11 (21%) as low-strategy study. In digital health management, the use of a different number of strategies in research may reflect the diversity of research design, objectives, and methods. Many studies have chosen a medium number of strategies to comprehensively consider multiple aspects and evaluate the effectiveness of disease management. Research using a low number of strategies may focus more on the intervention effects in specific aspects to further explore the mechanisms of action of specific strategies. On the contrary, studies using a high number of strategies may aim to explore the combined effects of multiple intervention methods to achieve more comprehensive disease management. However, the results of the studies we included did not indicate that more strategies have more effective effects. Therefore, we further evaluated the impact of different numbers of strategies and compared the effect of different strategy combinations to provide more practical guidance for accurate and personalized digital health management.

We assessed the effectiveness of DHI strategies qualitatively based on the SF/HIT. Overall, most (≥50/52, ≥96%) studies have reported positive or mixed outcomes for most outcome indicators, such as preventive care; efficiency; perceived ease of use or usefulness; safety, privacy or security; acceptability; appropriateness; satisfaction process of service delivery or performance; effectiveness; and adherence or attendance. However, in terms of cost-effectiveness, most (50/52, 96%) studies have not reported positive results. This may be due to the introduction and development of digital health management tools, which provide more convenient, safe, and efficient management methods for patients with diabetes [60]. These tools include remote monitoring devices, mobile apps, and web-based platforms that can help patients better monitor blood glucose levels, manage diet, and engage in exercise, thereby improving the effectiveness of preventive care and service delivery processes and enhancing patient acceptance and satisfaction with treatment plans [105]. However, the negative results in terms of cost-effectiveness may be due to the significant investment required for the implementation and operation of digital health management tools, and the long-term cost-effectiveness has not been fully demonstrated [59,62,73]. In high-strategy and medium-strategy studies, most (36/41, 88%) studies reported positive outcomes, while low-strategy (7/11, 64%) studies reported fewer positive outcomes. This reflects the investment of more resources and technology, which can more comprehensively meet the needs of patients and improve treatment effectiveness. In contrast, low-strategy research may not have achieved the same effect due to resource constraints or insufficient technology. This was consistent with research by Brueton et al [35], which indicates that using a combination of strategies rather than a single approach can be more effective in enhancing participant retention. Future research should focus on cost-effectiveness analysis, explore the long-term economic benefits of digital health management tools, and propose more effective low-cost strategies to promote the sustainable development of diabetes management strategies.

In the meta-analysis, we included 27 studies evaluating the efficacy of 12 different DHI strategy combinations on HbA_{1c} levels. We found several high-quality combinations that can significantly reduce the HbA_{1c} levels, which reflects that, in diabetes management, using a specific strategy combination may have more advantages than using a single strategy. Among numerous strategy combinations, the SUCRA probability ranking suggested that DHI-BDGI (communication, engagement, guide, and management) might be the most effective strategy combination, followed by DHI-DFGKP (engagement, goal setting, guide, monitor, and stimulate); DHI-FGIKQ (goal setting, guide, management, monitor, and support); and DHI-ABDEFGIKLOPQ (action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support). The effect of only using the combination of DHI-GIK (guide, management, and monitor) was relatively weak. First, in the current medical practice, the treatment of diabetes is no longer a simple drug treatment but a comprehensive management, including lifestyle intervention, psychological support, drug treatment, and other comprehensive interventions [5]. Among these, communication and patient engagement are considered crucial factors that can enhance patients' understanding and compliance with treatment and improve treatment effectiveness [106]. Meanwhile, guidance and management provide specific action guidelines and treatment plans to help patients better control their blood glucose levels [102,105]. The combination of these strategies complements each other and can comprehensively cover all aspects of the treatment process, maximizing the therapeutic effect. Second, these 2 combinations of DHI-DFGKP (engagement, goal setting, guide, monitor, and stimulate) and DHI-FGIKQ (goal setting, guide, management, monitor, and support) also showed good results. They emphasize aspects such as patient engagement, goal setting, guidance, monitoring, and support. The combination of these strategies makes the treatment process more systematic and orderly, helping patients better understand and execute treatment plans and improving the success rate of treatment [59]. However, the combination of DHI-ABDEFGIKLOPQ (action planning, communication, engagement, feedback, goal setting, guide, management, monitor, prompt, shape, stimulate, and support) encompasses more strategies and may seem more comprehensive, but there may be some challenges in practical applications. This complex combination may require more resources and time to implement while also increasing the cognitive burden on patients, which may affect treatment compliance. Finally, the combination of DHI-GIK (guide, management, and monitor), which only includes a few strategies, was relatively weak. This may be due to a lack of communication and patient engagement, reflecting insufficient communication between health care professionals and patients in real clinical practice and a lack of enthusiasm and participation from patients during treatment [78,106]. In addition, relying solely on guidance and management while ignoring the psychological and lifestyle factors of patients can easily lead to poor treatment outcomes.

At the same time, we included 7 studies evaluating the efficacy of 6 different DHI strategy combinations on FBG levels. Only

XSL•FO
RenderX

DHI-GIK (guide, management, and monitor) was shown to have an effect. The advantage of this effective strategy combination is that it integrates 3 key elements: guidance, management, and monitoring. These strategies can provide systematic treatment support, help patients develop and execute effective treatment plans, and monitor treatment outcomes in a timely manner, thereby better controlling FBG levels [97]. Our result was consistent with research that has shown that multidimensional interventions, including guidance, management, and monitoring, are critical to the success of diabetes management [5]. However, other strategy combinations may lack sustained guidance and monitoring or may not fully motivate patients to actively participate in treatment, resulting in poor effect compared with DHI-GIK (guide, management, and monitor). Meanwhile, this result may also be influenced by the bias caused by the insufficient amount of original studies on FBG levels in our meta-analysis. Therefore, there is a certain difference in the results between studies evaluating FBG and HbA_{1c} levels.

In general, our research results highlight the importance of comprehensive strategy combination in blood glucose management of T2DM. Future research can further explore other combinations of strategies to find more effective management models for diabetes. There is a need to strengthen communication and interaction between health care professionals and patients, as well as to improve patient participation and compliance. Furthermore, it is essential to strengthen the long-term tracking and evaluation of diabetes management strategies.

This review showed that DHI-BDFIQ (communication, engagement, goal setting, management, and support) and DHI-BGKR (communication, guide, monitor, and tailor) combination strategies were significantly effective for both BMI and weight loss. Among these, DHI-BDFIQ (communication, engagement, goal setting, management, and support) was considered the optimal combination. First, it covers the following key aspects: effective communication promotes cooperation and understanding between patients and medical teams; active engagement enhances the patient's participation and execution of treatment, such as physical activity and dietary intake control; clear goal setting helps guide the direction and progress of treatment [98]; effective management can provide systematic treatment plans and continuous management support; and continuous support can help patients overcome difficulties and maintain a positive attitude during the treatment process [59], to help patients better manage their weight. On the other hand, DHI-BGKR (communication, guide, monitor, and tailor) emphasizes personalized guidance and monitoring, which is more targeted, allowing patients to adjust their weight management plan according to their own situation [56,64,70], thereby reducing weight and BMI. In addition, the combination strategy of DHI-BDGI (communication, engagement, guide, and management) was found to be effective for BMI, and DHI-BDFGIKLOPQ (communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support) was effective for weight loss. Our research results have indicated that DHI-BDGI (communication, engagement, guide, and management) was the most effective combination strategy in reducing the HbA_{1c} levels, and therefore, it has a good effect on controlling BMI. Studies have shown that comprehensive treatment plans, including communication, engagement, guidance, and management, can effectively improve blood glucose control and weight management in patients with T2DM [107,108]. The effectiveness of these strategies may stem from their ability to comprehensively influence the patient's lifestyle and behavioral habits, thereby promoting the achievement of treatment goals [73,79,94]. The combination of DHI-BDFGIKLOPQ (communication, engagement, goal setting, guide, management, monitor, prompt, shape, stimulate, and support) includes multiple strategies, from goal setting to behavior shaping and support. While it has a positive effect on weight loss, it may encounter some challenges in practical applications, leading to a relatively weak overall effect. Therefore, the effect was relatively weak.

Our research emphasizes the advantages of the combination of communication, patient engagement, goal setting, personalized management, external support, and continuous monitoring strategies in the DHIs of BMI and weight management of patients with T2DM. Future research can further explore the applicability of these strategy combinations in different outcomes, such as quality of life, complication rate, and other indicators. In addition, precise personalized intervention based on artificial intelligence and big data analysis is necessary. At the same time, more interactive and personalized digital health platforms can be developed to improve patient engagement and treatment compliance.

Limitations

This study has several limitations. First, despite conducting a comprehensive search, it is possible that some relevant papers published in non-English languages or from nonindexed sources might have been overlooked. Second, we conducted 3 rounds of data screening and extraction, which could have introduced inconsistencies in certain cases. In addition, in our quantitative analysis, we only performed meta-analyses on specific outcome measures, thereby limiting the generalizability of our conclusions. Regarding BMI and weight loss outcomes, there was substantial heterogeneity among studies, which we addressed through sensitivity analyses and publication bias assessments. Furthermore, the number of original studies included in the meta-analyses was relatively small, potentially leading to some degree of bias in the conclusions. Finally, due to the nature of the interventions, blinding was not feasible in most studies. This might have exaggerated the estimated intervention effects in the network meta-analysis and contributed to the low to moderate quality of the evidence and the low methodology quality of included studies.

Conclusions

Our research provided a comprehensive analysis and summary of the strategies used in DHIs for T2DM. We identified 63 strategies and categorized them into 19 strategy themes. *Guide*, *monitor*, *management*, and *engagement* were the most commonly used strategies. Most studies reported positive or mixed outcomes for most outcome indicators based on the SF/HIT. Research involving a medium or high number of strategies was found to be more effective than research involving a low number of strategies. The strategy combination composed

XSL•FO
RenderX

of *communication*, *engagement*, *guide*, and *management* was most effective in reducing HbA_{1c} levels, while the strategy combination that included *guide*, *management*, and *monitor* was effective in reducing FBG levels. The strategy combination composed of *communication*, *engagement*, *goal setting*, *management*, and *support* was most effective in BMI and weight management. Future research should further confirm the effectiveness of these strategies in other indicators and populations, explore more strategy combinations, and optimize the design and implementation of strategies for patients with diabetes. Furthermore, it is necessary to develop more interactive and personalized digital health platforms. Finally, the cost-effectiveness analysis of strategy use should be strengthened to provide more effective guidance for disease management and clinical practice of T2DM.

Acknowledgments

This study was supported by the Key Research and Development Plan of Shaanxi Province, China, (2021SF-144) and National Higher Medical Education Association, Nursing Education Branch, Education Science Fund Project (GJHL160008). The authors would like to sincerely appreciate all the authors who participated in this systematic review and meta-analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 28 KB - jmir_v27i1e63209_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 13 KB - jmir_v27i1e63209_app2.docx]

Multimedia Appendix 3
Behavior change techniques.
[DOCX File , 51 KB - jmir_v27i1e63209_app3.docx]

Multimedia Appendix 4
Characteristics of the included studies.
[XLSX File (Microsoft Excel File), 45 KB - jmir_v27i1e63209_app4.xlsx]

Multimedia Appendix 5
Detailed information about strategies and strategy combinations.
[XLSX File (Microsoft Excel File), 28 KB - jmir_v27i1e63209_app5.xlsx]

Multimedia Appendix 6
Evaluation of the effectiveness of different digital health intervention strategies based on the synthesis framework for the assessment of health IT.
[XLSX File (Microsoft Excel File), 23 KB - jmir_v27i1e63209_app6.xlsx]

Multimedia Appendix 7
Risk of bias and quality assessments of included studies.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v27i1e63209_app7.xlsx]

References

1. Väätäinen S, Keinänen-Kiukaanniemi S, Saramies J, Uusitalo H, Tuomilehto J, Martikainen J. Quality of life along the diabetes continuum: a cross-sectional view of health-related quality of life and general health status in middle-aged and older Finns. Qual Life Res 2014 Sep;23(7):1935-1944. [doi: 10.1007/s11136-014-0638-3] [Medline: 24510623]
2. Jalkanen K, Aarnio E, Lavikainen P, Jauhonen HM, Enlund H, Martikainen J. Impact of type 2 diabetes treated with non-insulin medication and number of diabetes-coexisting diseases on EQ-5D-5 L index scores in the Finnish population. Health Qual Life Outcomes 2019 Jul 08;17(1):117 [FREE Full text] [doi: 10.1186/s12955-019-1187-9] [Medline: 31286982]

XSL•FO
RenderX

3. Williams R, Karuranga S, Malanda B, Saeedi P, Basit A, Besançon S, et al. Global and regional estimates and projections of diabetes-related health expenditure: results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract 2020 Apr;162:108072. [doi: 10.1016/j.diabres.2020.108072] [Medline: 32061820]
4. Magliano DJ, Boyko EJ, IDF Diabetes Atlas 10th Edition Scientific Committee. IDF diabetes atlas: 10th edition. International Diabetes Federation. 2021. URL: https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf [accessed 2025-01-24]
5. Chinese Elderly Type 2 Diabetes Prevention and Treatment of Clinical Guidelines Writing Group, Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Society, Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Health Care Society, Geriatric Professional Committee of Beijing Medical Award Foundation, National Clinical Medical Research Center for Geriatric Diseases (PLA General Hospital). [Clinical guidelines for prevention and treatment of type 2 diabetes mellitus in the elderly in China (2022 edition)]. Zhonghua Nei Ke Za Zhi 2022 Jan 01;61(1):12-50. [doi: 10.3760/cma.j.cn112138-20211027-00751] [Medline: 34979769]
6. American Diabetes Association. 4. Comprehensive medical evaluation and assessment of comorbidities: standards of medical care in diabetes-2021. Diabetes Care 2021 Jan;44(Suppl 1):S40-S52. [doi: 10.2337/dc21-S004] [Medline: 33298415]
7. Bain SC, Bekker Hansen B, Hunt B, Chubb B, Valentine WJ. Evaluating the burden of poor glycemic control associated with therapeutic inertia in patients with type 2 diabetes in the UK. J Med Econ 2020 Jan;23(1):98-105 [FREE Full text] [doi: 10.1080/13696998.2019.1645018] [Medline: 31311364]
8. Fellinger P, Rodewald K, Ferch M, Itariu B, Kautzky-Willer A, Winhofer Y. HbA1c and glucose management indicator discordance associated with obesity and type 2 diabetes in intermittent scanning glucose monitoring system. Biosensors (Basel) 2022 Apr 29;12(5):288 [FREE Full text] [doi: 10.3390/bios12050288] [Medline: 35624589]
9. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
10. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008 Oct 09;359(15):1577-1589 [FREE Full text] [doi: 10.1056/NEJMoa0806470] [Medline: 18784090]
11. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018 Feb 10;391(10120):541-551 [FREE Full text] [doi: 10.1016/S0140-6736(17)33102-1] [Medline: 29221645]
12. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an intensive lifestyle intervention on glycemic control in patients with type 2 diabetes: a randomized clinical trial. JAMA 2017 Aug 15;318(7):637-646 [FREE Full text] [doi: 10.1001/jama.2017.10169] [Medline: 28810024]
13. Gong Q, Zhang P, Wang J, Ma J, An Y, Chen Y, et al. Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30-year results of the Da Qing Diabetes Prevention Outcome Study. Lancet Diabetes Endocrinol 2019 Jun;7(6):452-461 [FREE Full text] [doi: 10.1016/S2213-8587(19)30093-2] [Medline: 31036503]
14. Lee EY, Cha SA, Yun JS, Lim SY, Lee JH, Ahn YB, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022 Jul 28;24(7):e37430 [FREE Full text] [doi: 10.2196/37430] [Medline: 35900817]
15. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
16. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
17. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
18. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
19. Rhee SY, Kim C, Shin DW, Steinhubl SR. Present and future of digital health in diabetes and metabolic disease. Diabetes Metab J 2020 Dec;44(6):819-827 [FREE Full text] [doi: 10.4093/dmj.2020.0088] [Medline: 33389956]
20. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
21. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]

XSL•FO
RenderX

22. Sherrington A, Newham JJ, Bell R, Adamson A, McColl E, Araujo-Soares V. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obes Rev 2016 Jun;17(6):541-551 [FREE Full text] [doi: 10.1111/obr.12396] [Medline: 26948257]
23. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
24. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
25. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
26. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
27. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
28. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144 [FREE Full text] [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
29. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
30. Lim S, Hill B, Pirotta S, O'Reilly S, Moran L. What are the most effective behavioural strategies in changing postpartum women's physical activity and healthy eating behaviours? A systematic review and meta-analysis. J Clin Med 2020 Jan 16;9(1):237 [FREE Full text] [doi: 10.3390/jcm9010237] [Medline: 31963150]
31. Masi CM, Chen HY, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug;15(3):219-266 [FREE Full text] [doi: 10.1177/1088868310377394] [Medline: 20716644]
32. Mair JL, Salamanca-Sanabria A, Augsburger M, Frese BF, Abend S, Jakob R, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835 [FREE Full text] [doi: 10.1093/abm/kaad041] [Medline: 37625030]
33. Hwang DA, Lee A, Song JM, Han HR. Recruitment and retention strategies among racial and ethnic minorities in web-based intervention trials: retrospective qualitative analysis. J Med Internet Res 2021 Jul 12;23(7):e23959 [FREE Full text] [doi: 10.2196/23959] [Medline: 34255658]
34. Treweek S, Pitkethly M, Cook J, Fraser C, Mitchell E, Sullivan F, et al. Strategies to improve recruitment to randomised trials. Cochrane Database Syst Rev 2018 Feb 22;2(2):MR000013 [FREE Full text] [doi: 10.1002/14651858.MR000013.pub6] [Medline: 29468635]
35. Brueton VC, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. Cochrane Database Syst Rev 2013 Dec 03(12):MR000032 [FREE Full text] [doi: 10.1002/14651858.MR000032.pub2] [Medline: 24297482]
36. Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol 2007 Aug;60(8):757-765 [FREE Full text] [doi: 10.1016/j.jclinepi.2006.11.023] [Medline: 17606170]
37. Lee SW, Chan CK, Chua SS, Chaiyakunapruk N. Comparative effectiveness of telemedicine strategies on type 2 diabetes management: a systematic review and network meta-analysis. Sci Rep 2017 Oct 04;7(1):12680 [FREE Full text] [doi: 10.1038/s41598-017-12987-z] [Medline: 28978949]
38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
39. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. Br J Health Psychol 2015 Feb;20(1):130-150. [doi: 10.1111/bjhp.12102] [Medline: 24815766]
40. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019 Oct 03;10(10):ED000142 [FREE Full text] [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
41. Christopoulou SC, Kotsilieris T, Anagnostopoulos I. Assessment of health information technology interventions in evidence-based medicine: a systematic review by adopting a methodological evaluation framework. Healthcare (Basel) 2018 Aug 31;6(3):109 [FREE Full text] [doi: 10.3390/healthcare6030109] [Medline: 30200307]

XSL•FO
RenderX

42. Lin L, Zhang J, Hodges JS, Chu H. Performing arm-based network meta-analysis in R with the pcnetmeta package. J Stat Softw 2017 Aug;80:5 [FREE Full text] [doi: 10.18637/jss.v080.i05] [Medline: 28883783]
43. Xu C, Niu Y, Wu J, Gu H, Zhang C. Software and package applicating for network meta-analysis: a usage-based comparative study. J Evid Based Med 2018 Aug;11(3):176-183. [doi: 10.1111/jebm.12264] [Medline: 29266878]
44. Bhatnagar N, Lakshmi PV, Jeyashree K. Multiple treatment and indirect treatment comparisons: an overview of network meta-analysis. Perspect Clin Res 2014 Oct;5(4):154-158 [FREE Full text] [doi: 10.4103/2229-3485.140550] [Medline: 25276624]
45. Shim S, Yoon BH, Shin IS, Bae JM. Network meta-analysis: application and practice using Stata. Epidemiol Health 2017;39:e2017047 [FREE Full text] [doi: 10.4178/epih.e2017047] [Medline: 29092392]
46. Spineli LM. An empirical comparison of Bayesian modelling strategies for missing binary outcome data in network meta-analysis. BMC Med Res Methodol 2019 Apr 24;19(1):86 [FREE Full text] [doi: 10.1186/s12874-019-0731-y] [Medline: 31018836]
47. Sedgwick P, Marston L. How to read a funnel plot in a meta-analysis. BMJ 2015 Sep 16;351:h4718. [doi: 10.1136/bmj.h4718] [Medline: 26377337]
48. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
49. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
50. Glasgow RE, Toobert DJ. Brief, computer-assisted diabetes dietary self-management counseling: effects on behavior, physiologic outcomes, and quality of life. Med Care 2000 Nov;38(11):1062-1073. [doi: 10.1097/00005650-200011000-00002] [Medline: 11078048]
51. Keyserling TC, Samuel-Hodge CD, Ammerman AS, Ainsworth BE, Henríquez-Roldán CF, Elasy TA, et al. A randomized trial of an intervention to improve self-care behaviors of African-American women with type 2 diabetes: impact on physical activity. Diabetes Care 2002 Sep;25(9):1576-1583. [doi: 10.2337/diacare.25.9.1576] [Medline: 12196430]
52. Tate DF, Jackvony EH, Wing RR. Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. JAMA 2003 Apr 09;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
53. Young RJ, Taylor J, Friede T, Hollis S, Mason JM, Lee P, et al. Pro-active call center treatment support (PACCTS) to improve glucose control in type 2 diabetes: a randomized controlled trial. Diabetes Care 2005 Feb;28(2):278-282. [doi: 10.2337/diacare.28.2.278] [Medline: 15677779]
54. Kim HS, Jeong HS. A nurse short message service by cellular phone in type-2 diabetic patients for six months. J Clin Nurs 2007 Jun;16(6):1082-1087. [doi: 10.1111/j.1365-2702.2007.01698.x] [Medline: 17518883]
55. Lorig K, Ritter PL, Villa F, Piette JD. Spanish diabetes self-management with and without automated telephone reinforcement: two randomized trials. Diabetes Care 2008 Mar;31(3):408-414. [doi: 10.2337/dc07-1313] [Medline: 18096810]
56. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
57. Powers BJ, Olsen MK, Oddone EZ, Bosworth HB. The effect of a hypertension self-management intervention on diabetes and cholesterol control. Am J Med 2009 Jul;122(7):639-646 [FREE Full text] [doi: 10.1016/j.amjmed.2008.12.022] [Medline: 19559166]
58. Nelson LA, Greevy RA, Spieker A, Wallston KA, Elasy TA, Kripalani S, et al. Effects of a tailored text messaging intervention among diverse adults with type 2 diabetes: evidence from the 15-month REACH randomized controlled trial. Diabetes Care 2021 Jan;44(1):26-34 [FREE Full text] [doi: 10.2337/dc20-0961] [Medline: 33154039]
59. Feng Y, Zhao Y, Mao L, Gu M, Yuan H, Lu J, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 20;11:e40420 [FREE Full text] [doi: 10.2196/40420] [Medline: 36939825]
60. Ruissen MM, Torres-Peña JD, Uitbeijerse BS, Arenas de Larriva AP, Huisman SD, Namli T, et al. Clinical impact of an integrated e-health system for diabetes self-management support and shared decision making (POWER2DM): a randomised controlled trial. Diabetologia 2023 Dec 29;66(12):2213-2225 [FREE Full text] [doi: 10.1007/s00125-023-06006-2] [Medline: 37775611]
61. Mansberger SL, Gleitsmann K, Gardiner S, Sheppler C, Demirel S, Wooten K, et al. Comparing the effectiveness of telemedicine and traditional surveillance in providing diabetic retinopathy screening examinations: a randomized controlled trial. Telemed J E Health 2013 Dec;19(12):942-948 [FREE Full text] [doi: 10.1089/tmj.2012.0313] [Medline: 24102102]
62. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse coaching and mobile health compared with usual care to improve diabetes self-efficacy for persons with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]
63. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A fully automated web-based program improves lifestyle habits and HbA1c in patients with type 2 diabetes and abdominal obesity: randomized trial of patient e-coaching nutritional support (the ANODE study). J Med Internet Res 2017 Nov 08;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]

XSL•FO
RenderX

64. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015 Mar 03;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
65. Schillinger D, Handley M, Wang F, Hammer H. Effects of self-management support on structure, process, and outcomes among vulnerable patients with diabetes: a three-arm practical clinical trial. Diabetes Care 2009 Apr;32(4):559-566 [FREE Full text] [doi: 10.2337/dc08-0787] [Medline: 19131469]
66. Davis RM, Hitch AD, Salaam MM, Herman WH, Zimmer-Galler IE, Mayer-Davis EJ. TeleHealth improves diabetes self-management in an underserved community: diabetes TeleCare. Diabetes Care 2010 Aug;33(8):1712-1717 [FREE Full text] [doi: 10.2337/dc09-1919] [Medline: 20484125]
67. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
68. Wild SH, Hanley J, Lewis SC, McKnight JA, McCloughan LB, Padfield PL, et al. Supported telemonitoring and glycemic control in people with type 2 diabetes: the Telescot Diabetes Pragmatic Multicenter randomized controlled trial. PLoS Med 2016 Jul;13(7):e1002098 [FREE Full text] [doi: 10.1371/journal.pmed.1002098] [Medline: 27458809]
69. Crowley MJ, Tarkington PE, Bosworth HB, Jeffreys AS, Coffman CJ, Maciejewski ML, et al. Effect of a comprehensive telehealth intervention vs telemonitoring and care coordination in patients with persistently poor type 2 diabetes control: a randomized clinical trial. JAMA Intern Med 2022 Sep 01;182(9):943-952 [FREE Full text] [doi: 10.1001/jamainternmed.2022.2947] [Medline: 35877092]
70. Hesseldal L, Christensen JR, Olesen TB, Olsen MH, Jakobsen PR, Laursen DH, et al. Long-term weight loss in a primary care-anchored eHealth lifestyle coaching program: randomized controlled trial. J Med Internet Res 2022 Sep 23;24(9):e39741 [FREE Full text] [doi: 10.2196/39741] [Medline: 36149735]
71. Lavikainen P, Mattila E, Absetz P, Harjumaa M, Lindström J, Järvelä-Reijonen E, et al. Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J Med Internet Res 2022 Feb 24;24(2):e31530 [FREE Full text] [doi: 10.2196/31530] [Medline: 35200147]
72. Nagata T, Aoyagi SS, Takahashi M, Nagata M, Mori K. Effects of feedback from self-monitoring devices on lifestyle changes in workers with diabetes: 3-month randomized controlled pilot trial. JMIR Form Res 2022 Aug 09;6(8):e23261 [FREE Full text] [doi: 10.2196/23261] [Medline: 35943766]
73. Thorsen IK, Yang Y, Valentiner LS, Glümer C, Karstoft K, Brønd JC, et al. The effects of a lifestyle intervention supported by the InterWalk smartphone app on increasing physical activity among persons with type 2 diabetes: parallel-group, randomized trial. JMIR Mhealth Uhealth 2022 Sep 28;10(9):e30602 [FREE Full text] [doi: 10.2196/30602] [Medline: 36170002]
74. Yin W, Liu Y, Hu H, Sun J, Liu Y, Wang Z. Telemedicine management of type 2 diabetes mellitus in obese and overweight young and middle-aged patients during COVID-19 outbreak: a single-center, prospective, randomized control study. PLoS One 2022;17(9):e0275251 [FREE Full text] [doi: 10.1371/journal.pone.0275251] [Medline: 36174028]
75. Calikoglu F, Bagdemir E, Celik S, Idiz C, Ozsarı H, Issever H, et al. Telemedicine as a motivational tool to optimize metabolic control in patients with diabetes in Turkey: a prospective, randomized, controlled TeleDiab trial. Telemed J E Health 2023 Apr 01;29(4):518-530. [doi: 10.1089/tmj.2022.0028] [Medline: 36067018]
76. Cheung NW, Redfern J, Thiagalingam A, Hng TM, Marschner S, Haider R, et al. Effect of mobile phone text messaging self-management support for patients with diabetes or coronary heart disease in a chronic disease management program (SupportMe) on blood pressure: pragmatic randomized controlled trial. J Med Internet Res 2023 Jun 16;25:e38275 [FREE Full text] [doi: 10.2196/38275] [Medline: 37327024]
77. Gerber BS, Biggers A, Tilton JJ, Smith Marsh DE, Lane R, Mihailescu D, et al. Mobile health intervention in patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023 Sep 05;6(9):e2333629 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.33629] [Medline: 37773498]
78. Han CY, Zhang J, Ye XM, Lu JP, Jin HY, Xu WW, et al. Telemedicine-assisted structured self-monitoring of blood glucose in management of T2DM results of a randomized clinical trial. BMC Med Inform Decis Mak 2023 Sep 14;23(1):182 [FREE Full text] [doi: 10.1186/s12911-023-02283-4] [Medline: 37710226]
79. Nelson LA, Spieker AJ, Greevy RA, Roddy MK, LeStourgeon LM, Bergner EM, et al. Glycemic outcomes of a family-focused intervention for adults with type 2 diabetes: main, mediated, and subgroup effects from the FAMS 2.0 RCT. Diabetes Res Clin Pract 2023 Dec;206:110991. [doi: 10.1016/j.diabres.2023.110991] [Medline: 37925077]
80. Whitehouse CR, Knowles M, Long JA, Mitra N, Volpp KG, Xu C, et al. Digital health and community health worker support for diabetes management: a randomized controlled trial. J Gen Intern Med 2023 Jan 17;38(1):131-137 [FREE Full text] [doi: 10.1007/s11606-022-07639-6] [Medline: 35581452]
81. Zamanillo-Campos R, Fiol-deRoque MA, Serrano-Ripoll MJ, Mira-Martínez S, Ricci-Cabello I. Development and evaluation of DiabeText, a personalized mHealth intervention to support medication adherence and lifestyle change behaviour in patients with type 2 diabetes in Spain: a mixed-methods phase II pragmatic randomized controlled clinical trial. Int J Med Inform 2023 Aug;176:105103. [doi: 10.1016/j.ijmedinf.2023.105103] [Medline: 37267809]

XSL•FO
RenderX

82. Lee DY, Yoo SH, Min KP, Park CY. Effect of voluntary participation on mobile health care in diabetes management: randomized controlled open-label trial. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e19153 [FREE Full text] [doi: 10.2196/19153] [Medline: 32945775]
83. Bender MS, Cooper BA, Park LG, Padash S, Arai S. A feasible and efficacious mobile-phone based lifestyle intervention for Filipino Americans with type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017 Dec 12;2(2):e30 [FREE Full text] [doi: 10.2196/diabetes.8156] [Medline: 30291068]
84. Fortmann AL, Gallo LC, Garcia MI, Taleb M, Euyoque JA, Clark T, et al. Dulce digital: an mHealth SMS-based intervention improves glycemic control in Hispanics with type 2 diabetes. Diabetes Care 2017 Oct;40(10):1349-1355 [FREE Full text] [doi: 10.2337/dc17-0230] [Medline: 28600309]
85. Lee JY, Wong CP, Tan CS, Nasir NH, Lee SW. Telemonitoring in fasting individuals with type 2 diabetes mellitus during Ramadan: a prospective, randomised controlled study. Sci Rep 2017 Aug 31;7(1):10119 [FREE Full text] [doi: 10.1038/s41598-017-10564-y] [Medline: 28860546]
86. Quinn CC, Swasey KK, Crabbe JC, Shardell MD, Terrin ML, Barr EA, et al. The impact of a mobile diabetes health intervention on diabetes distress and depression among adults: secondary analysis of a cluster randomized controlled trial. JMIR Mhealth Uhealth 2017 Dec 07;5(12):e183 [FREE Full text] [doi: 10.2196/mhealth.8910] [Medline: 29217502]
87. Ramadas A, Chan CK, Oldenburg B, Hussein Z, Quek KF. Randomised-controlled trial of a web-based dietary intervention for patients with type 2 diabetes: changes in health cognitions and glycemic control. BMC Public Health 2018 Jun 08;18(1):716 [FREE Full text] [doi: 10.1186/s12889-018-5640-1] [Medline: 29884161]
88. Sarayani A, Mashayekhi M, Nosrati M, Jahangard-Rafsanjani Z, Javadi M, Saadat N, et al. Efficacy of a telephone-based intervention among patients with type-2 diabetes; a randomized controlled trial in pharmacy practice. Int J Clin Pharm 2018 Apr;40(2):345-353. [doi: 10.1007/s11096-018-0593-0] [Medline: 29435911]
89. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth app intervention for persons with type 2 diabetes and its associations with initial self-management: randomized controlled trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
90. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
91. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
92. MacPherson MM, Merry KJ, Locke SR, Jung ME. Effects of mobile health prompts on self-monitoring and exercise behaviors following a diabetes prevention program: secondary analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 05;7(9):e12956 [FREE Full text] [doi: 10.2196/12956] [Medline: 31489842]
93. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
94. Torbjørnsen A, Jenum AK, Småstuen MC, Arsand E, Holmen H, Wahl AK, et al. A low-intensity mobile health intervention with and without health counseling for persons with type 2 diabetes, part 1: baseline and short-term results from a randomized controlled trial in the Norwegian part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]
95. Vervloet M, van Dijk L, de Bakker DH, Souverein PC, Santen-Reestman J, van Vlijmen B, et al. Short- and long-term effects of real-time medication monitoring with short message service (SMS) reminders for missed doses on the refill adherence of people with type 2 diabetes: evidence from a randomized controlled trial. Diabet Med 2014 Jul;31(7):821-828. [doi: 10.1111/dme.12439] [Medline: 24646343]
96. Wakefield BJ, Koopman RJ, Keplinger LE, Bomar M, Bernt B, Johanning JL, et al. Effect of home telemonitoring on glycemic and blood pressure control in primary care clinic patients with diabetes. Telemed J E Health 2014 Mar;20(3):199-205 [FREE Full text] [doi: 10.1089/tmj.2013.0151] [Medline: 24404819]
97. Zhou P, Xu L, Liu X, Huang J, Xu W, Chen W. Web-based telemedicine for management of type 2 diabetes through glucose uploads: a randomized controlled trial. Int J Clin Exp Pathol 2014;7(12):8848-8854 [FREE Full text] [Medline: 25674254]
98. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
99. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming clinical inertia: a randomized clinical trial of a telehealth remote monitoring intervention using paired glucose testing in adults with type 2 diabetes. J Med Internet Res 2015 Jul 21;17(7):e178 [FREE Full text] [doi: 10.2196/jmir.4112] [Medline: 26199142]
100. Nicolucci A, Cercone S, Chiriatti A, Muscas F, Gensini G. A randomized trial on home telemonitoring for the management of metabolic and cardiovascular risk in patients with type 2 diabetes. Diabetes Technol Ther 2015 Aug;17(8):563-570. [doi: 10.1089/dia.2014.0355] [Medline: 26154338]

XSL•FO
RenderX

101. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
102. Dixon RF, Zisser H, Layne JE, Barleen NA, Miller DP, Moloney DP, et al. A virtual type 2 diabetes clinic using continuous glucose monitoring and endocrinology visits. J Diabetes Sci Technol 2020 Sep;14(5):908-911 [FREE Full text] [doi: 10.1177/1932296819888662] [Medline: 31762302]
103. Ridho A, Alfian SD, van Boven JF, Levita J, Yalcin EA, Le L, et al. Digital health technologies to improve medication adherence and treatment outcomes in patients with tuberculosis: systematic review of randomized controlled trials. J Med Internet Res 2022 Feb 23;24(2):e33062 [FREE Full text] [doi: 10.2196/33062] [Medline: 35195534]
104. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
105. Kerr D, Edelman S, Vespasiani G, Khunti K. New digital health technologies for insulin initiation and optimization for people with type 2 diabetes. Endocr Pract 2022 Aug;28(8):811-821 [FREE Full text] [doi: 10.1016/j.eprac.2022.04.006] [Medline: 35452813]
106. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021 Mar 25;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
107. Roy K, Iqbal S, Gadag V, Bavington B. Relationship between psychosocial factors and glucose control in adults with type 2 diabetes. Can J Diabetes 2020 Oct;44(7):636-642. [doi: 10.1016/j.jcjd.2020.01.005] [Medline: 32205074]
108. Alaofè H, Hounkpatin WA, Djrolo F, Ehiri J, Rosales C. Knowledge, attitude, practice and associated factors among patients with type 2 diabetes in Cotonou, Southern Benin. BMC Public Health 2021 Feb 12;21(1):339 [FREE Full text] [doi: 10.1186/s12889-021-10289-8] [Medline: 33579243]

Abbreviations

DHI: digital health intervention
FBG: fasting blood glucose
HbA1c: glycated hemoglobin A1c
MD: mean difference
PRISMA-NMA: Preferred Reporting Items for Systematic Review and Meta-Analyses extension for Network Meta-Analysis
RCT: randomized controlled trial
SF/HIT: synthesis framework for the assessment of health IT
SUCRA: surface under the cumulative ranking curve
T2DM: type 2 diabetes mellitus

Edited by A Mavragani; submitted 12.06.24; peer-reviewed by KR Quimby, A Hornsten; comments to author 22.11.24; revised version received 04.12.24; accepted 16.12.24; published 14.02.25.

Please cite as:
Li M, Liu S, Yu B, Li N, Lyu A, Yang H, He H, Zhang N, Ma J, Sun M, Du H, Gao R
Assessing the Effectiveness of Digital Health Behavior Strategies on Type 2 Diabetes Management: Systematic Review and Network Meta-Analysis
J Med Internet Res 2025;27:e63209
URL: https://www.jmir.org/2025/1/e63209
doi:10.2196/63209
PMID:

©Min Li, Shiyu Liu, Binyang Yu, Ning Li, Aili Lyu, Haiyan Yang, Haiyan He, Na Zhang, Jingru Ma, Meichen Sun, Hong Du, Rui Gao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Simulation-Based Learning Supported by Technology to Enhance Critical Thinking in Nursing Students: Scoping Review

Hege Vistven Stenseth[1], MSc; Simen A Steindal[1,2], MSc, PhD; Marianne Trygg Solberg[1,3], MSc, PhD; Mia Alexandra Ølnes[1], BScN, BA; Anne Lene Sørensen[1], MSc; Camilla Strandell-Laine[1,4], PhD; Camilla Olaussen[1], PhD; Caroline Farsjø Aure[1], PhD; Ingunn Pedersen[5], MSc; Jaroslav Zlamal[1,6], MSc; Jussara Gue Martini[7], PhD; Paula Bresolin[7], PhD; Silje Christin Wang Linnerud[1], MSc; Andréa Aparecida Gonçalves Nes[1,8], PhD

[1]Department of Graduate Studies, Faculty of Nursing, Lovisenberg Diaconal University College, Oslo, Norway

[2]Institute of Nursing, Faculty of Health Sciences, VID Specialized University, Oslo, Norway

[3]Department of Health and Nursing Science, Faculty of Health and Sports Sciences, University of Agder, Grimstad, Norway

[4]Novia University of Applied Sciences, Turku, Finland

[5]Nord University, Bodø, Norway

[6]Clinical Simulation Center, Vestfold Hospital Trust, Tønsberg, Norway

[7]Postgraduate Program in Nursing, Universidade Federal de Santa Catarina, Florianópolis, Brazil

[8]Department of Caring and Ethics, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

Corresponding Author:
Hege Vistven Stenseth, MSc
Department of Graduate Studies
Faculty of Nursing
Lovisenberg Diaconal University College
Lovisenbergata 15 B
Oslo, 0456
Norway
Phone: 47 93840198
Email: hege.vistven@ldh.no

Abstract

Background: Critical thinking is a crucial skill in the nursing profession and must be fostered through nursing education. Simulation-based learning (SBL) with technological modalities is a pedagogical approach to enhance critical thinking skills for nursing students. The use of technology in SBL to achieve critical thinking skills is diverse. No previous scoping review has systematically mapped studies on SBL supported by technology to enhance critical thinking in nursing students.

Objective: This scoping review aimed to systematically map research on the use of SBL supported by technology to enhance critical thinking in nursing students.

Methods: This scoping review was conducted according to the framework by Arksey and O'Malley and was reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. A systematic, comprehensive literature search was performed in the LILACS, ERIC, MEDLINE, Embase, PsycINFO, and Web of Science databases in 2021 and repeated in 2023 and 2024. Pairs of authors independently assessed titles, abstracts, and full-text papers and extracted data from the included studies. The data underwent summative and thematic analysis and were categorized according to the findings.

Results: In total, 4 main categories of technology applied in SBL were identified: computer-based simulations, human-patient simulators, virtual reality or immersive virtual reality, and others. The findings revealed a shift across time in the technology used for SBL to enhance critical thinking, from human patient simulators to computer-based simulations. A dominant part of the included studies published after 2018 (21/44, 48%) incorporated a combination of asynchronous and synchronous learning activities. The theoretical foundation of the studies revealed a range of scientific theories and conceptual frameworks and models. Enablers of or barriers to the enhancement of critical thinking skills in nursing students were identified within the following themes: affinity for and availability of technology, realism, accessibility, engagement and motivation, validation, return on investment, and enhanced critical thinking through SBL using technology.

XSL•FO
RenderX

Conclusions: There has been a noticeable shift in the technology and use of technology in SBL. Descriptions of the applied technology and pedagogical considerations are pivotal for comparing or synthesizing research results. There has been a trend toward a blended educational approach combining synchronous and asynchronous learning activities. User technological proficiency and the perceived quality of the technology are imperative in the development of critical thinking. Realism, engagement, and motivation play pivotal roles in the enhancement of critical thinking in technologically supported SBL. The establishment of robust theoretical foundations of research and standardized research practices will strengthen the evidence obtained from the research conducted.

(J Med Internet Res 2025;27:e58744) doi:10.2196/58744

KEYWORDS

critical thinking; simulation-based learning; technologically supported simulation-based learning; nursing education; nursing students; review

Introduction

Background

Critical thinking skills are vital for professional nurses to perform nursing care in a safe, evidence-based manner. Nurses need to be able to analyze, summarize, and evaluate information and initiate appropriate actions. Critical thinking skills empower nurses to manage the uncertainties inherent in nursing practice, thus contributing to safe, effective care across diverse clinical settings [1-4]. Critical thinking includes purposeful and self-regulatory judgment in a process of interpretation, analysis, evaluation, and inference [5]. Critical thinking also includes evidential, conceptual, methodological, criteriological, or contextual considerations based on the relevant judgment [5]. Critical thinking is driven by internal motivation and characterized by reflection, self-monitoring, and self-correction. This process cultivates a reflective judgment on what to do, believe, or make sense of in any given context [6]. However, critical thinking is a concept that has no established or agreed upon definition [7]. As Cassum et al [8] have pointed out, individuals may often define critical thinking based on their own understanding and choose to emphasize certain aspects of critical thinking. Several studies have pointed out that critical thinking is used synonymously with other concepts, such as clinical decision-making, analytical thinking, clinical judgment, critical judgment, creative thinking, problem-solving, reflective thinking, and diagnostic reasoning or higher-order thinking [9,10]. This is mirrored in a study by Geng [11], who identified 64 definitions of critical thinking. The common characteristics of critical thinking in these definitions were judgment, argument, questioning, information processing, problem-solving, and meta-cognition. These characteristics can be found both in definitions of critical thinking, for example, in the definition by Facione [5], and in definitions of other concepts, such as clinical reasoning, decision-making, or clinical judgment [12,13]. Although achieving consensus on the definition of critical thinking poses challenges for identifying studies related to this concept, there is no doubt about its vital importance in the field of nursing.

To facilitate the development of critical thinking skills in nursing students, educators could encourage active student participation in the learning process. Simulation-based learning (SBL) is one pedagogical approach based on principles of active learning [1-3]. Bland et al [14] have conceptualized and defined SBL in nursing education as “a dynamic process involving the creation of a hypothetical opportunity that incorporates an authentic representation of reality, facilitates active student engagement and integrates the complexities of practical and theoretical learning with opportunity for repetition, feedback, evaluation and reflection.” SBL offers possibilities for experiential learning and deliberate practice through the integration of theoretical and practical knowledge, application of skills, and development of the ability to reflect and give and receive feedback [15]. Furthermore, SBL aims to replicate features of clinical practice that require learners to use critical thinking and clinical reasoning, involving cognitive, psychomotor, and affective skills [16,17].

Technological solutions for SBL in nursing education are rapidly evolving, and different simulation modalities are applied according to pedagogical and practical considerations [15,18]. SBL in nursing education can be divided into 2 main categories: SBL in the physical learning environment and SBL in the virtual learning environment. The physical learning environment is typically represented by simulation laboratories within universities but can include SBL in the clinical environment and context as well. The virtual learning environment refers to a digital or computerized environment that provides opportunities for independence in learning and diverse approaches to interacting with technology and others within the virtual world [19]. The types of simulation technology or equipment used as part of the simulation activity are often referred to using the term *simulation modality* [20]. The modality is the platform and resources used to support the simulated experience and may incorporate a variety of technologies [21,22]. One common modality for SBL is human-patient simulators (HPSs), which are advanced physical simulators with humanlike features and responses driven by computers [22]. Other modalities often applied in virtual learning environments can be categorized as technology-enhanced simulation (TES). TES is a blanket term for SBL with direct or assisted interaction with an electronic medium presented through computers or other technology that provides learners with a virtual environment in which they complete certain tasks, use information, assess, make clinical decisions, and observe results [20]. TES modalities include computer-assisted simulation, serious games, and computer simulation games, collectively referred to as computer-based simulation (CBS). CBSs are interactive simulation games in which learners are assigned roles, typically as the nurse, and

XSL•FO
RenderX

nursing students are involved in a scenario targeted to interpreting information and problem-solving [23]. Virtual reality (VR) and immersive VR (IVR) are the most recently introduced TES modalities, with different levels of immersion. These modalities seek to replicate specific clinical settings in a virtual environment. IVR distinguishes itself from other VR technologies by using special headsets that immerse the participants in a virtual world [24-26]. IVR experiences are often designed to promote specific learning outcomes in scenarios with patient-nurse interactions in a suitable clinical environment [27].

Several previous reviews have examined the prevalence and effectiveness of SBL to enhance critical thinking in nursing education [16,18,28-30]. SBL is the predominant learning activity that uses digital tools in nursing education [18]. A scoping review on teaching strategies for the development of clinical reasoning in advanced nursing clinical practice reported that SBL is the most frequently encountered teaching strategy [28]. Systematic reviews have examined the effectiveness of SBL on clinical reasoning and critical thinking skills in nursing education [16,29,30]. These reviews have determined that there is insufficient evidence to form definitive conclusions on the effect of SBL on clinical reasoning and critical thinking. However, the use of technology to enhance critical thinking in SBL has not been thoroughly addressed in these reviews.

Reviews have also examined the prevalence of SBL supported by technology and compared SBL that uses different technologies or SBL that uses technology versus traditional teaching methods [15,26,31,32]. A systematic review and meta-analysis of 12 randomized controlled trials concluded that VR offers higher potential than conventional teaching methods in nursing education for advancing outcomes of overall satisfaction, theoretical knowledge, and practice proficiencies without providing an advantage in enhancing critical thinking skills [31]. This review limited the examination to randomized controlled trials, excluding studies with other research designs [31]. Another systematic review examined game-based learning in undergraduate nursing education and found serious games to be the most used game type. Game-based learning was effective in achieving learning outcomes, particularly in the cognitive domain [32]. However, critical thinking was not a specific outcome, and only mixed methods studies were included in the review. In addition, a scoping review explored the use of virtual simulation (VS), an interactive recreation of real-life clinical scenarios using computer technology, to enhance diagnostic reasoning in health care professional education. VS was found to be at least as effective if not superior to traditional SBL, but only 3 of the 12 included studies were related to nursing education [26]. Another literature review also supported the effectiveness of VS in improving skills, learning, and critical thinking in nursing education [15]. However, critical thinking was the least explored outcome, the literature search was limited to 2 databases, and solely English-language papers were included.

Moreover, reviews have investigated the effectiveness of SBL in enhancing critical thinking in nursing education incorporating technology [33,34]. A systematic review and meta-analysis focused on VS as a tool to enhance clinical reasoning in nursing education and found that VS can improve clinical reasoning skills [34].

A systematic review comparing VR to traditional SBL practices concluded that the effectiveness of VR on clinical reasoning educational outcomes was similar or superior to that of traditional methods, although the evidence is limited [33]. A meta-analysis on technology in the SBL learning environment for the development of complex skills such as critical thinking across different educational domains stated that recent technologies can greatly enhance the effect of SBL [4]. A qualitative literature review explored nursing students' experiences with virtual screen-based simulations and found that a comfortable atmosphere; engagement and feedback; and improved skills, including critical thinking and decision-making skills, positively impacted learning in VS. Negative experiences with VS stemmed from the need for more support, poor design, and lack of authenticity [35].

Objectives

In nursing education, SBL supported by technology is increasingly applied as a learning strategy given the collective demand for active and flexible educational approaches and the growing interest and need for technological solutions [36]. Understanding the enablers and barriers associated with these solutions may be of great value for selecting suitable technological modalities for SBL to promote the development of critical thinking. Previous reviews have not comprehensively examined the range and use of technology for supporting SBL in enhancing nursing students' critical thinking skills. Therefore, there is a need for a broad, comprehensive literature review, such as a scoping review, including scientific papers in several languages and that use diverse research methods. Such a review is important to summarize the scope and variety of published peer-reviewed studies in this area while also identifying potential gaps in the existing peer-reviewed studies [37]. To the best of our knowledge, no scoping review has investigated the variety and use of technology in SBL for enhancing nursing students' critical thinking skills. Consequently, this scoping review aimed to systematically map studies on the use of SBL supported by technology to enhance critical thinking in nursing students.

Methods

Overview

This scoping review was conducted according to the methodological framework by Arksey and O'Malley [37], which includes the following steps: (1) identifying the research questions (RQs); (2) identifying relevant studies; (3) selecting studies; (4) charting the data; and (5) collating, summarizing, and reporting the results. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines were followed for the reporting of this review [38,39]. The research protocol was published in *JMIR Research Protocols* [40], and the deviations from the protocol are described in Multimedia Appendix 1.

Identifying the RQs

The following RQs were asked:

XSL•FO
RenderX

1. What variety of technology is used in SBL to enhance critical thinking skills in nursing education? (RQ 1)
2. How is technology used in SBL to enhance critical thinking skills in nursing education? (RQ 2)
3. What barriers and enablers do nursing students report in the use of technology in SBL to enhance critical thinking? (RQ 3)

Identifying Relevant Studies

The Sample, Phenomenon of Interest, Design, Evaluation, and Research Type framework determined the inclusion and exclusion criteria [41] (Textbox 1).

Textbox 1. Eligibility criteria for inclusion of the selected studies.

Inclusion criteria

- Sample: papers focused on undergraduate and postgraduate nursing students
- Phenomenon of interest (the main aim of this scoping review was to systematically map studies on the use of simulation-based learning [SBL] supported by technology to enhance critical thinking in nursing students. As such, it is warranted to use critical thinking as a concept interchangeably with other concepts): SBL supported by technology to stimulate critical thinking, clinical decision-making, analytical thinking, creative thinking, problem-solving, reflective thinking, diagnostic reasoning, clinical reasoning, or clinical judgment in educational or institutional contexts; SBL supported by technology, including human patient simulator–based modalities, virtual reality (VR), immersive VR, virtual simulation, augmented reality, or computer-based simulation
- Design: studies with quantitative, qualitative, or mixed methods designs
- Evaluation: undergraduate and postgraduate nursing students' perspectives and experiences regarding the use of technology in SBL to enhance critical thinking or similar concepts
- Research type: studies of any research type published in Portuguese, Spanish, English, Norwegian, Swedish, or Danish in peer-reviewed journals

Exclusion criteria

- Sample: papers focused health care students other than nursing students
- Phenomenon of interest: SBL that does not use technology; SBL that uses technology but is not related to critical thinking or similar concepts; SBL in clinical practice not related to education
- Evaluation: nurse educators' perspectives and experiences regarding the use of technology in SBL to stimulate critical thinking
- Research type: case studies, case reports, clinical guidelines, all types of reviews, master's and PhD theses, conference proceedings and abstracts, letters, comments, discussions, editorials, and book chapters

Selecting Studies

A systematic search was conducted in CINAHL, ERIC, Embase, LILACS, Ovid MEDLINE, PsycINFO, and Web of Science on June 28, 2021. The database search was updated on March 17, 2023, and November 13, 2024. Each database was searched from its inception.

The search strategy in Ovid MEDLINE, using Medical Subject Headings and text words, was designed by an experienced research librarian (MAØ) in collaboration with the rest of the research team and encompassed 3 elements: SBL, technology, and nursing students and nursing education. A second research librarian (Kari Larsen Mariussen) reviewed the search strategy using the Peer Review of Electronic Search Strategies Checklist [42]. The search strategy in Ovid MEDLINE is outlined in Multimedia Appendix 2. In addition, we manually examined the reference lists of the included papers to determine whether any studies mentioned in those references were relevant to our review. Moreover, we conducted forward citation searching using the Google Scholar platform to identify relevant studies that had cited the included papers.

The identified citations were exported to EndNote (Clarivate Analytics) for removal of duplicates using the method described by Bramer et al [43]. The citations were randomly divided into 8 groups and transferred to the web application Rayyan (Qatar Computing Research Institute) [44] for storage, organization, and blinding of the study selection process. HVS and AAGN screened the titles and abstracts of 10% of the papers to pilot-test the eligibility criteria and deemed that the eligibility criteria did not require modification. In total, 8 pairs of authors and nonauthor contributors (HVS and CFA, SCWL and SAS, MTS and JZ, Andrea Mohallem and Fernando Riegel [nonauthors], PB and JGM, ALS and CS-L, CO and HVS, and IP and AAGN) performed the study selection process. Pairs of authors independently assessed whether titles and abstracts and then full-text articles met the eligibility criteria. When there was any doubt regarding inclusion, a third author (HVS or AAGN) independently assessed the full-text paper, and the decision was based on a negotiated consensus. The second updated search in 2024 was conducted by 2 pairs of authors (HVS and SAS and AAGN and MTS) consistent with the previous procedure for study selection.

Charting the Data

A standardized data collection form was developed in Microsoft Excel for data extraction from the included papers, including the following data: authors, year, and country; aim; sample; design; technology; simulation procedures; scenario design; and results related to the RQs. HVS and AAGN piloted the data collection form on 5 of the included papers, and some small revisions were made. Pairs of authors (Andrea Mohallem and Fernando Riegel, SCWL and SAS, PB and JGM, CO and CFA, IP and AAGN, JZ and MTS, and ALS and CS-L) extracted data

XSL•FO
RenderX

from the full-text papers. One author extracted the data, and the other checked their accuracy against the papers. Disagreement or uncertainties among pairs of authors were resolved through an independent assessment by HVS, and the agreement was based on negotiated consensus. The extracted data were also checked by the first (HVS) and last (AAGN) authors to ensure quality and consistency in data extraction.

Collating, Summarizing, and Reporting the Results

Data from the included papers regarding RQ 1 and RQ 2 were analyzed by HVS and AAGN using summative analysis. Regarding RQ 3, data were analyzed using an inductive approach to organize the results thematically, a method previously used in scoping reviews [45-47]. HVS and AAGN read the extracted data several times to identify patterns of similarities and differences related to the RQ, and the patterns were organized into subthemes and themes using a low level of interpretation and abstraction. An example of inductive thematic analysis for RQ 3 includes participants highlighting the ability to attend the SBL regardless of time and place, which was categorized under subtheme Flexibility and Theme Affinity for technology. Uncertainties of owns performance was categorized under subtheme Need for feedback and theme Validation.

SAS provided feedback on the preliminary thematic groupings. The data were then analyzed deductively from the perspective of the 6 dimensions (system quality, information quality, service quality, use and intention to use, user satisfaction, and net benefits) of the DeLone and McLean (D&M) [48,49] model of information system (IS) success. These themes were sorted accordingly and subsequently positioned within the predetermined categories of the framework of the D&M IS Success Model through deductive analysis. This model aims to evaluate the success of an IS and suggests that its dimensions are interrelated and influence each other [48]. The model has been widely used in research and practice to assess the success of various types of ISs, including enterprise systems, e-commerce platforms, and educational technology [19,49,50]. Studies exploring learners' (nursing students') experiences with technology have provided valuable insights into how technology affects their ability to develop critical thinking skills. User satisfaction, a key indicator of IS success, served as the basis for evaluating the technology's effectiveness [50]. The reported enablers and barriers for critical thinking enhancement through SBL using technology were sorted according to the results of the inductive analysis. The outcomes of all the studies were related to the development of critical thinking among nursing students through learning activities with technologically supported SBL. Therefore, the outcomes would be synonymous with learning and the achievement of learning outcomes.

Results

Overview

After the screening of 6297 records and 315 (5%) full-text papers, a total of 96 (1.52%) studies were included. Citations with additional information are listed in Multimedia Appendix 3 [51-146]. The study selection process and reasons for exclusion of full-text papers are presented in Figure 1 [39].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram of the study selection process for this scoping review.

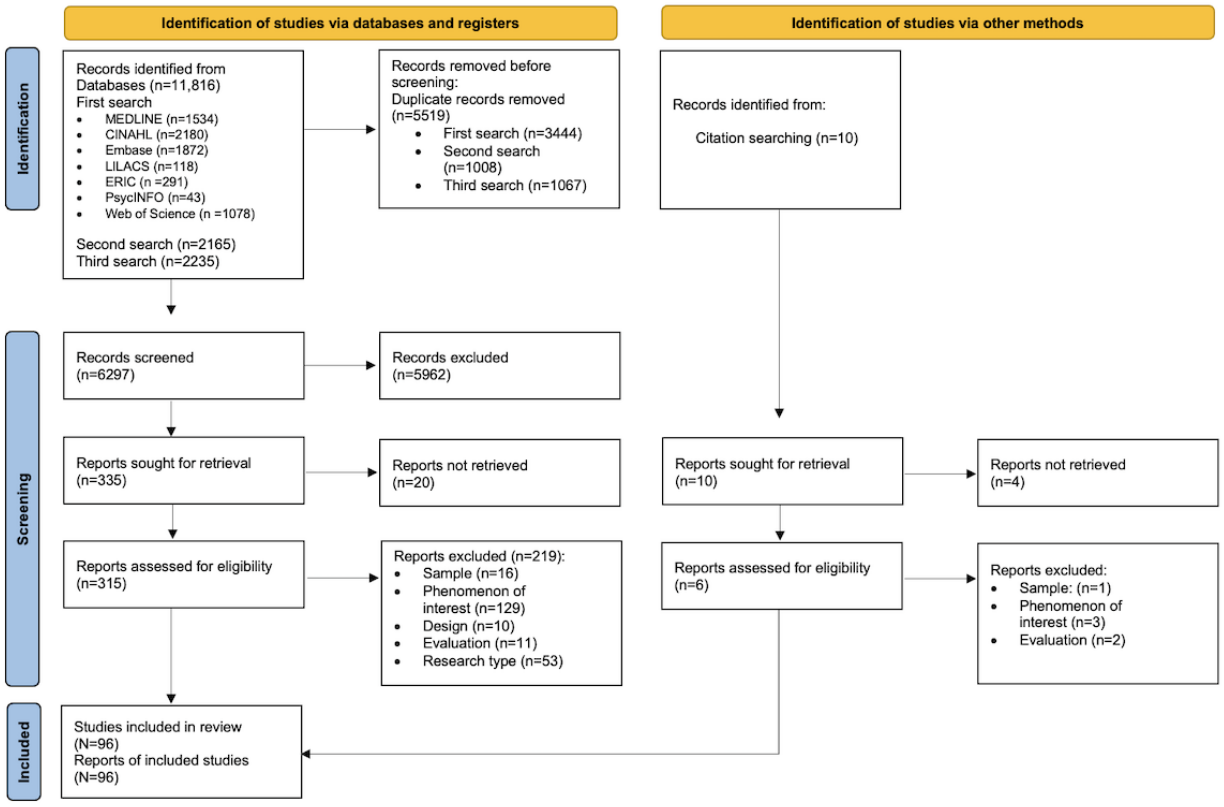

Sociodemographic Findings

The studies originated from 20 countries: the United States (42/96, 44%) [51-92], South Korea (12/96, 12%) [93-104], Australia (6/96, 6%) [105-110], China (6/96, 6%) [111-116], Turkey (6/96, 6%) [117-122], Canada (4/96, 4%) [123-126], Finland (3/96, 3%) [127-129], Saudi Arabia (3/96, 3%) [130-132], Taiwan (3/96, 3%) [133-135], France (1/96, 1%) [136], Hong Kong (1/96, 1%) [137], Indonesia (1/96, 1%) [138], Iran (1/96, 1%) [139], Israel (1/96, 1%) [140], Lebanon and Jordan (1/96, 1%) [141], Norway (1/96, 1%) [142], Portugal (1/96, 1%) [143], Sweden (1/96, 1%) [144], Thailand (1/96, 1%) [145], and the United Kingdom (1/96, 1%) [146].

The years of publication ranged from 1986 to 2024. Most studies (66/96, 69%) used a quantitative design [51, 52, 54-69, 71-74, 78, 86, 89-96, 98-104, 111, 114-119, 121, 127-130, 132-137, 139-141, 143-146]. A mixed methods, multimethod, or blended research approach was applied in 20% (19/96) of the studies [53,77,83-85,87,97,106,107,109,110,112,113,122-126,138], and 11% (11/96) of the studies incorporated a qualitative design [58,75,79,81,82,88,105,108,120,131,142].

Most studies (92/96, 96%) included undergraduate nursing students. The studies also included Master of Science nursing students [83]; nurse practitioner master's degree students [86]; postgraduate nurse anesthesia students [82]; and nursing students preparing at the diploma, associate, and baccalaureate levels [54]. The sample sizes ranged from 8 to 675 participants. The reported ages of the participants (73/96, 76% of the studies) ranged from 15 to 56 years. All studies that reported demographic data on gender (67/96, 70%) had a majority of female participants, ranging from 57% to 100% except for one study that reported 52% male participants [131]. A total of 23% (22/96) of the studies did not report demographic data regarding age or gender [59, 63, 66, 67, 73, 78, 80, 82, 83, 85, 88-91, 97, 98, 101, 105-107, 129, 146]. In total, 10% (10/96) of the studies reported age but not gender [76, 96, 102, 103, 109, 126-128, 133, 134], and 4% (4/96) of the studies reported data on gender but not on age [77,112,121,137].

Theoretical Foundations and Measurement of Critical Thinking

More than half (56/96, 58%) of the studies [51-55, 61, 63, 65, 69, 74-77, 80, 81, 83-92, 97, 100, 102-105, 107, 109, 111, 112, 114, 115, 118-126, 128-131, 135-137, 139, 140, 145] presented a theory or conceptual framework as a foundation for the study conducted. Scientific theories such as learning theory (24/96, 25%), nursing theory (7/96, 7%), and philosophical-methodological theory (2/96, 2%) were applied alone or in combination 33 times across the studies. Specific theoretical models or conceptual frameworks were presented as the theoretical foundation alone or in combination 36 times across the studies.

Most studies (54/96, 56%) used one or more validated instruments to evaluate the development of critical thinking skills or equivalent concepts [51, 55-57, 60, 64, 65, 67, 69, 72, 74, 76, 77, 84-87, 91-93, 95-104, 111-119, 121-124, 126, 129, 132-137, 139, 141, 144]. Across these 54 studies, 42 validated

XSL•FO
RenderX

instruments were used to assess outcomes. The California Critical Thinking Disposition Inventory, including its Chinese version, was used in 13% (7/54) of these studies to measure critical thinking. The Lasater Clinical Judgment Rubric was the second most frequently used instrument, applied in 11% (6/54) of the studies. The Critical Thinking Disposition Tool was used in 9% (5/54) of the studies, whereas the Simulation Effectiveness Tool–Modified was used in 7% (4/54) of the studies, including one instance of using the Turkish version. Notably, 29 instruments were used in only one study each. A detailed overview is provided in Multimedia Appendix 3.

Range and Features of the Technology Applied in SBL for Enhancing Nursing Students' Critical Thinking (RQ 1)

The technology in SBL to enhance critical thinking was divided into 4 main categories. The first category of applied technology was CBS (44/96, 46%) [52, 53, 55, 59, 67, 76, 78, 83, 85, 86, 89, 90, 93, 96, 97, 99, 101, 104, 105, 108, 109, 111, 114, 117-125, 127, 128, 130-133, 136-140, 143]. A wide variety of terms was used in the studies, and details are presented in Multimedia Appendix 3. The second category of applied technology was HPSs (26/96, 27%) [51, 57, 60, 62-66, 68-73, 80-82, 84, 87, 91, 94, 100, 107, 141, 145, 146]. The third category of applied technology in SBL was VR or IVR (7/96, 7%) [75,77,95,98,103,110,126].

The fourth category comprised studies (19/96, 20%) that used several or combinations of technologies or technology not consistent with the first 3 categories. In these studies, technology was applied through a combination of CBS, an HPS and case studies [116], a combination of CBS and IVR simulation [115,129], a combination of CBS and simulated electronic health records [61], multimedia technologies combined with an HPS [58], videotaped vignettes combined with an HPS [54], FaceTime technology [79], an interactive videodisc system [134], telehealth simulation and real-time video and audio technology [142], telehealth simulation and an HPS [135], mixed reality [106], a combination of CBS and HPSs [56,102,110,112,113,144], a combination of an HPS and interactive case studies [74], and a VS designed in Microsoft PowerPoint [88]. An overview is provided in Table 1.

The analysis revealed an evolution of the technology used in the included studies across time, from most studies (21/32, 66%) based on HPSs as the modality before 2018 to TESs using CBS as the major modality category in studies after 2018 (Figure 2).

The technology applied was diversely described in the studies. Some studies using HPS technology (25/96, 26%) gave descriptions of features and functions, but these studies generally provided limited description of the HPSs, which were often described only by the HPS's name and producer [51, 62-64, 68-71, 81, 82, 87, 91, 94, 96, 100, 102, 107, 113, 116, 135, 141, 144-146]. For CBS and VR or IVR, there was generally a richer description of the technology used, with descriptions of features that could contribute to enhancing critical thinking in SBL through learning. Several studies (23/96, 24%) used commercially developed technology featuring preprogrammed scenarios with virtual patients and VS [52, 53, 59, 61, 67, 75, 83, 86, 89, 92, 93, 96, 99, 105, 106, 109-117, 121, 122, 125, 126, 131, 133, 143, 144]. Some studies (16/96, 17%) investigated technological solutions or software for CBS, VR, or IVR that was self-developed or developed in a cooperation between expert nursing educators and IT engineers [78, 85, 88, 90, 98, 101-104, 108, 118, 120, 124, 130, 133, 136]. One study combined already developed commercial platforms and developed its own scenarios using various technologies [97].

Table 1. Categories of technology applied in simulation-based learning to enhance critical thinking in the included studies (N=96).

Category of applied technology	Studies, n (%)
CBS[a] [52,53,55,59,67,76,78,83,85,86,89,90,93,96,97,99,101,104,105,108,109,111,114,117-125,127,128,130-133,136-140,143]	44 (46)
HPS[b] [51,57,60,62-66,68-73,80-82,84,87,91,94,100,107,141,145,146]	26 (27)
VR[c] or IVR[d] [75,77,95,98,103,110,126]	7 (7)
Other—combination of technologies, technologies not consistent with the first 3 categories, or unclear description of technology [54,56,58,61,74,79,88,102,106,110,112,113,115,116,129,134,135,142,144]	19 (20)

[a]CBS: computer-based simulation.

[b]HPS: human-patient simulator.

[c]VR: virtual reality.

[d]IVR: immersive virtual reality.

XSL•FO
RenderX

Figure 2. Evolution of the applied technology in simulation-based learning to enhance critical thinking from 1986 to 2024. CBS: computer-based simulation; HPS: human-patient simulator; IVR: immersive virtual reality; VR: virtual reality.

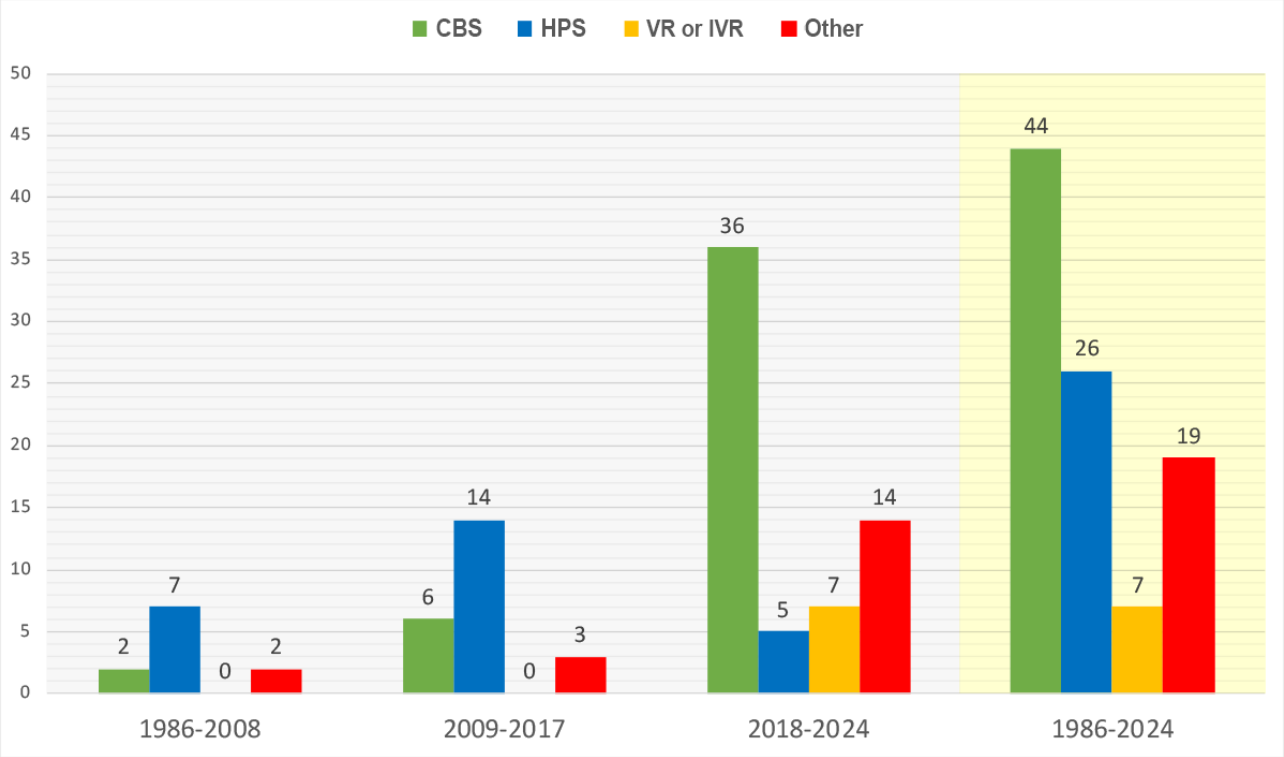

Use of Technology in SBL Activities to Enhance Critical Thinking (RQ 2)

Technology Integration Into SBL

HPS technology was integrated into SBL mainly by nursing students playing out active roles in real-world clinical scenarios, with HPSs representing patients. SBL was typically described as being built around clinical scenarios, with nursing students as nurses caring for the patient by assessing and taking necessary measures in the scenario [51, 62-64, 68-71, 81, 82, 91, 94, 96, 100, 107, 141, 144, 146]. SBL using telehealth technology in the studies was built around clinical telehealth scenarios [79,135,142]. The digital technology used in TES, CBS, and VR or IVR was usually built around clinical cases and scenarios in which nursing students took on the role of a nurse caring for a patient. Nursing students applied their knowledge and skills to assess, solve problems, and make clinical decisions [53, 59, 67, 78, 83, 85-90, 93, 96, 101-104, 106, 109, 112, 114, 115, 118, 120-122, 124, 125, 128, 130, 131, 133, 134, 136, 140, 143]. Several studies (34/61, 56%) incorporated scenario design and game mechanics aimed at developing clinical reasoning or clinical judgment and critical thinking skills [61, 75, 77, 85-90, 92, 96,97,99, 101-104, 109, 111, 112, 114, 116, 117, 119, 122, 124-126, 130-132, 137, 139, 144]. Some studies (12/61, 20%) stated that they used a narrative approach, constructing patient stories that evolved over time following a cause-and-effect structure to demonstrate the consequences of decisions and actions, thereby contributing to the development of critical thinking [77,85,88-90,105,111,117,122,128,131,143].

Synchronous and Asynchronous Learning Activities

The studies used technology in SBL to promote critical thinking through both synchronous and asynchronous learning activities. The learning activities in SBL that used an HPS were predominantly conducted synchronously, featuring scenarios in real time that aimed to enhance critical thinking and the clinical reasoning process through experiential learning [51, 56, 57, 60, 62-66, 68-73, 81, 82, 84, 87,91, 94, 100, 107, 113, 116, 135, 141, 144, 145]. These learning activities were conducted in groups ranging from pairs to larger numbers of nursing students. Asynchronous learning activities were often designed using TESs in SBL intended for self-directed, individual learning, providing more flexibility and independence for nursing students [52, 53, 55, 59, 61, 67, 78, 83, 85, 88, 92, 99, 104, 108, 113, 114, 116-122, 125, 127-129, 131-134, 140, 144]. This approach also gave nursing students the possibility of repetition and repetitive learning [61,99,109,127,134,144]. A third approach to using technology in SBL was to combine asynchronous and synchronous learning activities. The most common technologies applied for the combined approach were CBS and VR or IVR [77, 86, 90, 93, 96-99, 101-103, 105, 110-112, 115, 124, 126, 130, 136-139, 144], with SBL most commonly conducted as a combination of individual performance in CBS, VR, or IVR followed by group reflections and debriefing facilitated by trained faculty. One study reported conducting all parts of SBL using CBS synchronously on the web via a videoconferencing platform [89].

Other combinations included videotaped vignettes of a scenario combined with individual performance testing [54], an HPS in combination with multimedia in prerecorded simulated assessments, offering eighter synchronous or asynchronous debriefing for online students using a learning management

XSL•FO
RenderX

system [58], and guided workshops of CBS content before asynchronous individual learning [109]. The studies that used a combined approach were predominantly published after 2019 as 50% (8/16) of the included studies in the first updated search (2021-2023) and 46% (13/28) in the second updated search (2023-2024) reported using this approach compared to 6% (5/52) of the studies in the initial search (1986-2021). Of the 96 included studies, 6 (6%) did not give sufficient descriptions of the learning activities [75,76,80,95,123,146].

Feedback for Developing Critical Thinking Skills

The feedback described in studies that used HPS technology in SBL was provided in 2 ways: feedback provided by the HPSs giving human patient responses to actions taken during the scenario, including both physical reactions and verbal interactions, and feedback provided by educators or peers during the debriefing phase of SBL (21/96, 22%) [51,56,57,60,62-66,68-74,81,84,87,91,94,100,107,112,135,141]. In the SBL modality that used FaceTime and audiovisual equipment for telehealth scenarios, feedback was also provided during debriefing by trained facilitators [79,135,142]. A total of 3% (3/96) of the studies reported using video recordings during debriefing to evaluate, reflect, and provide feedback on student performance [62,72,100]. One study reported feedback provided through asynchronous discussion within a learning management system [58]. TESs using CBS and VR or IVR technology provided other and additional feedback models compared to SBL using HPS technology, with both audio and visual automated responses. Some TESs provided feedback in real time as the game unfolded or cumulatively at the end of the scenario or gaming session [59, 88, 89, 96, 109, 110, 112, 114, 115, 121, 122, 125, 127-129, 131]. Feedback in both CBS and VR or IVR was also delivered indirectly through patients' responses to actions and communication with a cause-and-effect approach that demonstrated the consequences of decisions and actions taken [75,103,110,117,129,131,143]. Feedback was given in diverse forms, such as written messages or verbal feedback from a patient, via facilitators or peers during the SBL session [90,115,127-129], quizzes within the CBS [102], and pop-up messages or summaries on execution compared to expected nursing performance at the end of the VR, IVR, or CBS session [52,55,95,102,103,108,109,120,136]. CBSs could provide feedback tailored to debriefing purposes [93,117,143]. In some studies (9/61, 15%), TESs provided students with a performance score, typically in the form of points or a percentage reflecting the accuracy of their actions when the game was completed [53,61,78,86,108,127,128,136]. Several studies (10/61, 16%) using TESs combined feedback from CBSs or IVR with face-to-face synchronous debriefing and group discussions [86,89,90,101-104,110,114,124].

Enablers and Barriers of the Applied Technology in SBL to Enhance Critical Thinking (RQ 3)

Overview

From the inductive data analyses, seven distinct themes emerged: (1) affinity for and availability of technology, (2) realism, (3) accessibility, (4) engagement and motivation, (5) validation, (6) return on investment, and (7) enhanced critical thinking through SBL using technology. The studies explored aspects of developing critical thinking skills through learning in SBL using technological modalities such as HPSs, CBSs, VR or IVR, and others. An overview of the dimensions of the D&M IS Success Model [48], the identified themes, their respective subthemes, and examples of both barriers and enablers is provided in Table 2.

XSL•FO
RenderX

Table 2.

Dimension of the D&M IS Success Model, main theme, and subthemes	Examples of measures for IS success enablers	Examples of measures for IS success barriers
System quality—ease of learning, ease of use (usability), intuitiveness, flexibility, reliability, response times, availability, and desirable characteristics		
Affinity for and availability of technology		
Understanding technology	• Well-structured and logical design [87,90,131].	• Challenges in understanding computer or accessory context [85,108,121] • Outdated software [138]
Understanding game features	• Easy to navigate [131]	• Challenges in navigating the game setting [92,108,123,138]
Flexibility	• Possibility of repetition [58,75,97,103,109,110,112,125] • Native language [120] • Independent of time and space [103,109,121,131]	• Foreign language [124,138]
Information quality—relevance, completeness, accuracy, conciseness, personalization, understandability, currency, timeliness, and security		
Realism		
Real-life scenarios	• Provides all aspects of the patient and context, giving a realistic experience [58, 75, 84, 88, 90, 102, 103, 106, 108, 112, 122, 124-126, 131, 138, 142]	• Low flexibility for changing a situation (compromised reality) [58,81,110] • Situations lacking resemblance to real life [82,113,123]
Physiological realism	• Replication of real human physiological features [75,81,107,131,138]	• Limited human features of the patient presented by technology [81,82,87,107]
Theory-to-practice transition	• Being able to visualize patient status and the effect of measures [75,81,107,108,131]	• Technology with low and intermediate fidelity [81,82] • Context perceived as not relevant [123]
Service quality—the quality of the support that system users receive from the service providers; this could entail IT support as well as other types of user support and empathy from staff		
Accessibility		
Technical accessibility	• Technology functioning as expected [142]	• Technical problems that cause interruptions [92,110,120,124] • Technical issues such as internet access and connection [90,120,138]
Use (behavior) and intention to use (attitude)—amount, frequency, use patterns, appropriateness, and purpose of use; attitude toward using and reusing the system		
Engagement and motivation		
Preferred learning styles	• Visual learning [106-108,120,122,125] • Autonomy in learning [75,77,97,106,130,131,138] • Being challenged [77,81,84,92,107,108,122,126,132,136]	• Perceived as not sufficient for “hands-on” learning [110,120,123] • Limited variability [113,121]
Physiological and psychological experiences	• Safe learning environment [87,88,109,110,112,122,124,126] • Immersion [92,97,103,108,110,112,126] • Sparked curiosity [108,113,126] • Activation of thoughts and feelings [92,108,122,131]	• Negative physiological reactions [89,106,110] • Confusion [82] • Feeling of incompetence and frustration and low sense of mastery and self-efficacy [82,108]
Experienced use	• Time to become familiar with features of the technology [79,142] • Proper training and instructions [79,110,142] • Appropriate taxonomical level [97,107]	• Limited time to become familiar with technology [103,107,138] • Viewing the learning tool negatively [108,123] • Content not adapted to the taxonomical level [89,107,123] • Inappropriate time frame allotted to tasks [123]
Validation		

XSL•FO
RenderX

Dimension of the D&M IS Success Model, main theme, and subthemes	Examples of measures for IS success enablers	Examples of measures for IS success barriers
Need for feedback	• Feedback provided on the actions taken [79,82,97,103,107,109,112,138] • Feedback enables engagement [102,107,109,125,138] • Provision of examples and explanation of how to perform patient care and why [53,79,125,138]	• No feedback offered [58] • Insufficient feedback offered [81,85]
Satisfaction—overall opinions about the system		
Return on investment		
Student investment	• No information available	• No value for money [123]
Net benefits—impacts on individual users, groups, organizations, industries, and nations		
Enhanced critical thinking through SBL [a] using technology		
Meta-cognition	• Reflecting on one's own actions and thinking [58,75,81,83,84,92,97,109,123,124,126] • Self-reflection and self-evaluation [58,75,83,97,126,131]	• Insufficient debriefing provided [107,112]

[a]SBL: simulation-based learning.

System Quality: Affinity and Availability

This theme pertains primarily to TESs, in which students interact with technology such as CBS and VR or IVR in self-directed, individual learning activities. Well-structured and logical technological platforms with easy navigation were reported as perceived enablers for achieving outcomes in SBL [87,90,131]. Challenges in understanding technological and game features [85,108,121], outdated software [138], and challenges navigating the technological platform [92,108,123,138] were reported as barriers because they limited options and influenced the interaction and the flow of the simulation [92,108,138]. Disorientation and difficulties obtaining information were viewed as time-consuming and hindering to nursing students' learning process, which caused frustration and negative experiences [92,108,123,138]. Flexibility encompassed the use of TESs individually and regardless of time and space [109,121,131] and the possibility of repetition [58,75,97,103,109,110,112,125], which was perceived positively in terms of motivation to use the applied technology and improve performance [58,75,97]. Interacting with TESs in the students' native language enhanced the availability of learning situations [120]. Options in a foreign language and lack of multiple language options were reported perceived barriers that made learning less available to nursing students [124,138].

Information Quality: Realism

Technology that provided all aspects of patient context with a realistic experience was reported as likely to enable learning and the development of critical thinking [58, 75, 84, 88, 90, 102, 103, 106, 108, 112, 122, 125, 126, 131, 138, 142]. Realistic replication of the human physiological features of the patient [75,81,107,131,138], realistic changes in patients' conditions, and being able to visually inspect the patient's status and see the effect of measures [75,81,107,108,131] were reported as increasing perceived realism; theory-to-practice transition; and, hence, nursing students' development of critical thinking skills. Low flexibility for changes in the scenario [58,81,110], limited resemblance to real-life situations [82,113,123], and failure to replicate realistic human features of the patient [81,82,87,107] were factors of technology that were reported to compromise realism in SBL [58,81,82,107,110,123] and cause confusion for nursing students.

Service Quality: Accessibility

When technology functions as expected, it may enhance critical thinking by making learning opportunities accessible [142]. Technical problems caused interruptions that interfered with learning and the smooth development of the scenario in SBL and were reported as a barrier to the development of critical thinking skills by negatively influencing the accessibility of the learning opportunity [82,120]. Another barrier reported, especially with TESs conducted individually, was technical issues related to internet access and connection [120,138].

Use and Intention to Use: Engagement, Motivation, and Validation

The preferred learning style may influence nursing students' engagement and motivation for interacting with technology in SBL and achieve critical thinking outcomes. Nursing students reported that the visual elements of the technology enabled the development of critical thinking through increased engagement from visual learners [106-108,120,122,125]. The perceived autonomy of self-directed learning in TESs, with the opportunity to try out behaviors and apply knowledge and skills without being supervised [75,77,97,106,130,131,138], as well as being challenged on skills and knowledge [77, 81, 84, 92, 97, 107, 108, 126, 132, 138], increased nursing students engagement and motivation. Some nursing students perceived that SBL using HPSs, CBSs, or VR or IVR did not constitute hands-on learning and, hence, was not appropriate for gaining enhanced critical thinking skills outside the clinical environment [110,120,123]. Across several studies (13/42, 31%), nursing students reported that TESs provided a safe learning environment

XSL•FO
RenderX

[87,88,109,110,112,122,124,126] where a high level of immersion [92,97,103,108,110,112,126], sparked curiosity [108,113,126], and the activation of thoughts and feelings [92,108,122,131] enabled involvement and engagement and facilitated the development of critical thinking for nursing students. Limited variability and opportunity to explore inhibited motivation for some nursing students [113,121]. Negative physiological reactions from immersive technology, such as dizziness and headaches, influenced nursing students' engagement and motivation [89,106,110]. Emotions such as confusion, frustration, feelings of incompetency, and a low sense of mastery related to technology led to decreased engagement and loss of motivation and, hence, critical thinking development [82,108].

Experienced use of technology in SBL influenced engagement and motivation. Time to become familiar with features of the technology [79,142], proper training [79,110,142], and appropriate taxonomical level [97,107] were factors associated with enhanced motivation to engage in learning reported by nursing students. Failure to become familiar with the technology [103,107,138], content not adapted to the appropriate taxonomical level [89,107,123], and an inappropriate time frame for conducting a task [123] resulted in decreased engagement and motivation for nursing students, leading to a negative perception of the technology [108,123].

Validation and feedback with provision of examples and explanations of patient care [53,79,125,138] and feedback on actions taken [79,82,97,103,107,109,112,138] were perceived as valuable for nursing students in SBL, enabling critical thinking development [79,82,97,105,107,138]. Feedback enabled motivation and engagement [102,107,109,125,138]. Insufficient or absence of feedback led to uncertainties regarding nursing students' own performance and the accuracy of their problem-solving, ultimately influencing the learning experience and development of critical thinking skills [58,81,85].

Satisfaction: Return on Investment

Chircop et al [123] found that nursing students were not satisfied with the CBS program they were required to buy and that several expressed that the CBS was not perceived as providing value for the money spent.

Net Benefits: SBL Using Technology

Nursing students reported that they perceived that SBL using technology can enhance critical thinking development through opportunities for meta-cognition and engagement in reflection [58,75,77,97,123,138]. Nursing students expressed a perceived value of opportunities to think about their own actions and thinking [58,75,81,83,84,92,97,109,123,124,126], engage in self-reflection and self-evaluation, and consider the benefits of receiving and giving feedback in debriefing sessions with peers and faculty [58,75,83,97,126,131]. Insufficient or no debriefing could decrease nursing students' opportunities for meta-cognition and development of critical thinking [107,112]

Discussion

Principal Findings

This scoping review aimed to systematically map studies on the use of SBL supported by technology to enhance critical thinking in nursing students. We identified 4 main categories of applied technology in SBL to enhance critical thinking as well as a shift across time in the use of the technology applied: from HPSs to CBSs. The technology applied in SBL to enhance critical thinking skills was closely connected to pedagogical considerations and educational approaches. This technology was applied in asynchronous and synchronous learning approaches. Learning approaches and use of technology in SBL have developed over time and trend toward a more blended or combined educational approach after 2019. Our main findings on RQ 3 were related to the information quality and use and intention to use dimensions of the D&M IS Success Model, with themes associated with realism as well as engagement and motivation as prominent findings. The included studies were diverse in origin, which indicates that the topic of interest is relevant worldwide. However, 44% (42/96) of the included studies originated in the United States. This poses the question of whether the use of technology in SBL to enhance critical thinking is more emphasized in the United States or whether this is an expression of a geographical bias in conducting research on the topic. Another concern is about the global applicability. Although a significant proportion of the studies in this review (42/96, 44%) were from the United States, it is important to consider whether and how the findings can be transferred to other educational contexts. Educational systems vary significantly worldwide, and factors such as cultural differences, resource availability, and teaching methods can impact the effectiveness of SBL supported by technology. For instance, in countries with limited resources, access to advanced technology may be a challenge. This may affect the implementation and outcomes of SBL. In addition, cultural differences in learning styles and teaching methods may require adaptations to ensure that SBL methods are effective [147]. It is also crucial to consider how different health care systems and nursing education requirements can influence the transferability of the findings. For example, in countries with different nursing education standards, adaptations may be necessary to ensure that SBL is relevant and effective [148].

Range of Technology

The included studies spanned a 38-year time frame, necessitating consideration of the significant technological development during this period. The technologies applied in SBL must be evaluated within the context of the studies' time frames as they might be outdated by present standards. From a historical perspective, the findings showed a shift in the technology applied in SBL to enhance critical thinking from HPS to TES modalities, such as CBS and VR or IVR. From 2018 to 2023, the predominant technology was CBS, with even more studies conducted using VR or IVR technology than HPSs. This showcases the evolving landscape of educational technology for SBL and active learning in nursing education. Although there is an increased use of health care and telehealth technology in SBL for future registered nurses [149], our findings suggest

XSL•FO
RenderX

that there is limited research on the use of telehealth technology, which is in line with the results of a study by Giuffrida et al [28]. The lack of studies on the use of telehealth in SBL to enhance critical thinking is striking considering the global emphasis on technology and technological advancements as well as the impact of the COVID-19 pandemic and its consequences. The limited number of studies regarding telehealth technology in SBL could be due to publication delay [150], and more research could emerge in the near future.

The descriptions of the applied technology varied across the studies. To consider the validity and generalizability or transferability of the findings in relation to technology as well as compare findings across studies, detailed descriptions of the technology are necessary [151,152]. Our review suggests that studies incorporating a digital technology into TESs generally provided a richer description of the technology involved than those that used HPSs. This may be due to the technology being more novel and complex and researchers acknowledging the importance of providing this information.

Use of Technology in SBL Activities to Enhance Critical Thinking

The use of technology in terms of learning design and teaching strategies in SBL to enhance critical thinking varied across the included studies. Synchronous learning was predominant in SBL that used HPSs, featuring real-time scenarios with physical presence. Conversely, digital technology such as TES leaned toward asynchronous activities, fostering self-directed learning with flexibility for learners. Wong et al [153] argue that developing skills to optimize self-directed learning is essential, with senior nursing students potentially being more adapted to mastering this skill compared to their junior counterparts, who may require more support, guidance, and validation. The need for validation, especially in self-directed learning within SBL, indicates the importance of aligning technological tools with educational objectives and providing comprehensive support for nursing students. As highlighted by the International Nursing Association of Clinical Simulation and Learning [154], there is a need to align tools or modalities in SBL with specific learning outcomes in the design phase of SBL. According to Nadelson et al [155], several nursing programs are looking to provide nursing students with more opportunities for online SBL, and the authors indicate that one important pedagogical choice is whether this should be provided as a synchronous or asynchronous learning activity. Research suggests that both formats yield similar outcomes, with TES offering efficiency through standardized content and the possibility of individualized training [156]. The potential for asynchronous, individualized training and multiple iterations as offered by TES could make these modalities resource efficient due to lower staff costs [25,26,157], with time freed for supportive teaching activities [158]. Hence, pedagogical considerations and cost efficiency should be determined before the choice of technological modality.

The evolution of teaching strategies was evident in the included studies published after 2019, with an increased number of studies that reported combining asynchronous and synchronous aspects in learning activities. According to Kalanlar [149], findings across 30 international nursing education programs reveal that 81% transitioned from traditional teaching to online classes during the pandemic, and of these, 61% conducted SBL using TES, 12% conducted SBL using HPSs, and 10% conducted SBL using telehealth. However, nursing students have expressed low satisfaction with excessive asynchronous activities and pointed out quality differences compared to tactile learning, thus highlighting the challenges of online SBL [149,159]. In a national postpandemic survey in the United States, nursing students reported that the virtual environment and online SBL were inadequate to prepare them for clinical reasoning in direct patient care [160]. The shift to online nursing education accelerated by the COVID-19 pandemic posed challenges to both educators and students, and this may have prompted the shift to a more blended teaching approach [161]. This evolution tends toward combined approaches in teaching critical thinking skills using technology in SBL, typically through individual asynchronous scenarios conducted virtually with TESs followed by synchronous debriefing. This may signal a shift toward a more balanced and comprehensive teaching strategy that considers aspects of social constructivism in learning, in which meaningful interactions influence learning positively [162]. The trend of applying a blended learning approach may represent a recognition that technology is not yet capable of replacing an experienced facilitator in guiding proper reflection processes. Debriefing, integral to SBL, facilitates reflection and learning, with feedback playing a crucial role in developing critical thinking skills [163,164]. Our review identified several methods of receiving feedback for the development of critical thinking through SBL using different technological modalities and tools. Consistent with a systematic review of VR in nursing education [15], we found that TESs offered more diverse feedback models and debriefing practices compared with HPSs, such as scoring systems and real-time feedback on actions taken and communication. This prompts exploration of the factors influencing the choice of educational strategies applied in nursing education to achieve the desired outcome and the optimal technology and design for SBL for effective critical thinking development.

Enablers and Barriers of the Applied Technology in SBL to Enhance Critical Thinking

An evident finding is the crucial role of realism, including real-life scenarios and the replication of human features in simulated patients, for nursing students' learning enhancement. Our findings showed that nursing students' perceptions of realistic patients, with both consistent human and physiological features and contexts, enable learning and enhance critical thinking skills. This could link to nursing students' experience of learning opportunities' relevance in SBL, connecting them to real nursing experiences. Nursing students value realistic features in the SBL environment to perform effectively. The term *relational realism* refers to the degree of realism experienced when interacting with HPSs in a simulated clinical setting [165]. Learning activities using an HPS in SBL may fail if there is no perceived relational realism, leading to learners being reluctant to interact and engage. In designing SBL, educators need to consider several aspects of reality, such as physiological, psychological, and conceptual realism, as well

XSL•FO
RenderX

as technological simulation modalities and tools [21]. Furthermore, experiences of immersion were reported by nursing students to enhance realism and, hence, learning and critical thinking skill development through engagement, motivation, and emotional activation. The literature supports that immersion and interactivity are key to virtual technology learning experiences [166]. The effectiveness of VR or IVR technologies relies on the users' sense of presence [167]. Motivation theories highlight the impact of emotions on students' engagement in learning activities, suggesting the importance of realism factors, as noted by Dubovi [168]. These insights emphasize the importance of creating SBL environments that mirror real-life patients and contexts closely for nurturing critical thinking skills in nursing education.

Our findings highlight the pivotal role of engagement, motivation, and personalized learning styles, emphasizing the need for adaptive and learner-centered approaches in designing and implementing technology in SBL through different modalities and tools. A systematic review on the use of VR among nursing students and registered nurses identified 2 disadvantages with the virtual world: a lack of realism and technical issues [169]. Our scoping review suggests that technical issues and problems could compromise realism and be barriers to learning and the development of critical thinking skills as these issues limit the availability of learning opportunities in the learning environment. The absence of realism may diminish immersion, consequently reducing engagement and the achievement of learning outcomes. Our results emphasize the importance of nursing students' technological proficiency in effectively interacting with technology. Harerimana and Mtshali [170] found that students' computer skills influenced their perception of the effectiveness of teaching and learning using technology. Consequently, using teaching and learning strategies to facilitate students' development of computer proficiency and technological literacy may enhance learning using technology in other areas, such as the development of critical thinking skills [46]. A lack of technological proficiency can be a major source of frustration for nursing students, potentially leading to a negative perception of technology [169] and decreasing their motivation and engagement [170]. Studies have identified poorly structured and designed digital learning interventions and technical problems as major causes of negative outcomes when introducing digital learning interventions in higher education [170,171]. They emphasize that technical difficulties diminish the value of the learning experience, decrease motivation, and contribute to feelings of inequality among students. This is particularly crucial in synchronous learning situations, in which interaction is essential [171]. Coupled with providing comprehensive technical support, ensuring technological functionality and quality is crucial for motivating and engaging nursing students in technology-based learning activities to enhance their critical thinking skills.

Theoretical Foundation for Pedagogy and Outcome Evaluation

The variety in theoretical foundations and the diverse outcome measures identified in our review pose a challenge regarding the comparison of research outcomes. More than half (56/96, 58%) of the studies in our scoping review incorporated an underlying theory or conceptual framework to inform their interventions. Establishing a scientific theoretical foundation for future studies may secure more robust research outcomes in the field of SBL with the support of technology to enhance critical thinking. Bauce et al [172] highlight that theory is needed in research to support pedagogy and emphasize that theory has served to underpin pedagogy and provide a foundation for outcome evaluation. The use of scientific theories in SBL research is important in formulating precise RQs, guiding study design and analysis, and contributing to a cumulative research program. It is crucial to differentiate between scientific theories, which are explanatory and descriptive, and a conceptual framework, which represents some aspects of reality [173]. Conceptual frameworks have a different scientific meaning from that of a scientific theory, and thus, the studies conducted based on them are different in nature. Varpio et al [174] underscore the need for a standardized understanding of terms such as *theory*, *theoretical framework*, and *conceptual framework* within health profession education and research. This is essential for enhancing the clarity and rigor of research reporting, subsequently fostering a more robust theoretical foundation for future studies in the field of SBL supported by technology to augment critical thinking skills. Furthermore, there was diversity in the validated instruments and self-developed questionnaires used across the included studies to measure outcomes. Adib-Hajbaghery and Sharifi [29] question the use of general measurement instruments to assess the outcome of critical thinking when evaluating effects in relation to specialized nursing interventions. This underscores the necessity for standardization and consensus on global research practices, facilitating more meaningful cross-study analyses and enabling the development of comprehensive reviews and meta-analyses.

Future Research Directions

The extensive body of research, comprising 96 studies included in this scoping review, highlights the potential for future systematic reviews and meta-analyses of both quantitative and qualitative scientific studies on SBL using technology to enhance critical thinking. This would provide deeper insights into the effectiveness of SBL strategies and identify best practices for integrating technology to foster critical thinking development. Furthermore, there is a need for consistency and standardization in the use of scientific theory to enhance the clarity and rigor of research reporting and establish a robust theoretical foundation for future studies in the field of SBL supported by technology to augment nursing students' critical thinking skills.

Future research should investigate SBL in relation to different learning styles, variations in SBL outcomes among nursing students at different stages of their education, generational differences in technology use, content and design of SBL, and the choice of technology applied to achieve the outcome of enhanced critical thinking. Studies should also explore factors influencing educators' choice of technology and design for SBL in nursing education to achieve outcomes and the enhancement of critical thinking skills in nursing students. Despite the increasing emphasis on health care technology and telehealth,

XSL•FO
RenderX

there is limited research on such technology use in SBL in nursing education to enhance nursing students' critical thinking skills. This calls for further research to inform nursing education on the pedagogical aspects of SBL using health care technology generally and telehealth technology specifically to enhance nursing students' critical thinking skills. Furthermore, studies comparing and evaluating the impact of different technologies on higher-order thinking outcomes may provide valuable insights that could inform nursing education.

Strengths and Limitations

The strengths of this scoping review were that our scoping review protocol was peer reviewed and published, and we used an acknowledged methodological framework and reported the review according to PRISMA-ScR checklist [38,39] to enhance transparency (Multimedia Appendix 4). A comprehensive search strategy was developed in collaboration with an experienced research librarian. The inclusion of various concepts often considered synonymous with critical thinking that share several similar factors was important due to the vague and diverse definitions of critical thinking. However, the fact that the studies may have used these concepts in ways that are not conceptually equivalent must be considered. Our inclusion criteria were limited to specified languages. Consequently, we may have introduced selection bias. Other limitations may be the dominant population of undergraduate students in the included studies, which might limit transferability to other contexts and populations. Furthermore, the dominance of included studies originating in the United States should be considered when interpreting the findings. However, we believe that, by including studies from 20 different countries, this reflects SBL with technology to enhance critical thinking in nursing education as a universal theme and the review exhibits global trends in nursing education and the use of technology in SBL to enhance critical thinking. Therefore, an understanding of each country's educational system is imperative in interpreting the findings. In addition, the variability in theoretical foundations and outcome measures across the studies may have hindered cross-study comparisons in this review. Finally, the findings need to be interpreted with caution due to the diversity in research contexts, technology used, and SBL interventions in the included studies and because the methodological quality of the included studies was not appraised and the data were organized and not synthesized.

Conclusions

This scoping review highlights a global relevance through a range of diverse studies. Over time, there was a noticeable shift from HPSs to TESs such as CBS and VR or IVR, emphasizing the importance of detailed descriptions of the applied technology to ensure validity and comparability across studies. After 2018, a trend toward blended educational approaches combining asynchronous and synchronous learning activities emerged. The feedback provided to nursing students in SBL to enhance critical thinking varied in delivery, with technology-driven feedback tailored to learning objectives playing a pivotal role in enhancing critical thinking. This highlights the need for educators to carefully select and tailor technology-based tools and programs to deliver feedback aligned with learning objectives and outcomes of critical thinking development. The establishment of robust theoretical foundations of research and standardized research practices will strengthen the evidence obtained from the research conducted. Realism, user proficiency, and the perceived quality of the technology significantly influence engagement, motivation, and the development of critical thinking in nursing students. This review provides valuable insights for educators, designers, and technology developers to enhance the effectiveness of technologically supported SBL in nursing education. Our findings can inform evidence-based practices, guide the design of effective educational interventions, and contribute to the ongoing discourse on the integration of technology into SBL in nursing education.

Acknowledgments

The authors acknowledge and extend their gratitude to Kari Larsen Mariussen for peer reviewing the search strategy and conducting the second updated search in 2024. They extend their acknowledgment for the vast amount of input to Andrea Mohallem, who chose not to continue in the process of completing the scoping review, and to Fernando Riegel for a substantial and invaluable contribution to the process until he sadly passed away before completion of this work. The authors extend their deepest gratitude and appreciation for his memory and express their condolences to his family and loved ones. This research was funded by the Educational Research Focusing on Competence Development and Quality in Nursing Education research group at Lovisenberg Diaconal University College, Oslo, Norway. The authors received no grants from any funding agency in the commercial, public, or nonprofit sector.

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Authors' Contributions

HVS contributed to project administration, data curation, investigation, formal analysis, conceptualization, visualization, and writing—review and editing. SAS contributed to data curation, conceptualization, visualization, writing—original draft, and writing—review and editing. MTS contributed to data curation, visualization, writing—original draft, and writing—review and editing. MAØ contributed to resource search. ALS contributed to data curation and writing—review and editing. CS-L contributed to data curation and writing—review and editing. CO contributed to data curation and writing—review and editing. CFA contributed

XSL•FO
RenderX

to data curation and writing—review and editing. IP contributed to data curation and writing—review and editing. JZ contributed to data curation and writing—review and editing. JGM contributed to data curation and writing—review and editing. PB contributed to data curation and writing—review and editing. SCWL contributed to data curation and writing—review and editing. AAGN contributed to data curation, conceptualization, investigation, formal analysis, supervision, visualization, writing—original draft, and writing—review and editing. All the authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Deviations from the published protocol.
[DOCX File , 24 KB - jmir_v27i1e58744_app1.docx]

Multimedia Appendix 2
Search strategy in Ovid MEDLINE.
[DOCX File , 60 KB - jmir_v27i1e58744_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies.
[DOCX File , 103 KB - jmir_v27i1e58744_app3.docx]

Multimedia Appendix 4
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 108 KB - jmir_v27i1e58744_app4.docx]

References

1. Von Colln-Appling C, Giuliano D. A concept analysis of critical thinking: a guide for nurse educators. Nurse Educ Today 2017 Feb;49:106-109. [doi: 10.1016/j.nedt.2016.11.007] [Medline: 27902948]
2. Riegel F, Crossetti MD. Theoretical frameworks and instruments for evaluation of critical thinking in nursing and education. Rev Gaucha Enferm 2018 Jul 23;39:e2017-e0097 [FREE Full text] [doi: 10.1590/1983-1447.2018.2017-0097] [Medline: 30043943]
3. Gross Forneris S, Fey MK. Critical conversations: the NLN guide for teaching thinking. Nurs Educ Perspect 2016;37(5):248-249. [doi: 10.1097/01.NEP.0000000000000069] [Medline: 27740554]
4. Chernikova O, Heitzmann N, Stadler M, Holzberger D, Seidel T, Fischer F. Simulation-based learning in higher education: a meta-analysis. Rev Educ Res 2020 Jun 15;90(4):499-541. [doi: 10.3102/0034654320933544]
5. Facione P. Critical thinking: a statement of expert consensus for purposes of educational assessment and instruction. Research Findings and Recommendations. The Delphi Report. Millbrae, CA: California Academic Press; 1990. URL: https://philarchive.org/archive/FACCTA [accessed 2025-02-10]
6. Facione PA, Crossetti MD, Riegel F. Holistic critical thinking in the nursing diagnostic process. Rev Gaucha Enferm 2017;38(3):e75576 [FREE Full text] [doi: 10.1590/1983-1447.2017.03.75576] [Medline: 29641771]
7. Andreucci-Annunziata P, Riedemann A, Cortés S, Mellado A, del Río MT, Vega-Muñoz A. Conceptualizations and instructional strategies on critical thinking in higher education: a systematic review of systematic reviews. Front Educ 2023 Mar 9;8:1141686. [doi: 10.3389/feduc.2023.1141686]
8. Cassum SH, McGrath JP, Gul RB, Dilshad A, Syeda K. Multidimensionality of critical thinking: a holistic perspective from multidisciplinary educators in Karachi, Pakistan. J Nurs Educ Pract 2013 Jan 09;3(7):9-23. [doi: 10.5430/jnep.v3n7p9]
9. Wang S. An exploration into research on critical thinking and its cultivation: an overview. Theory Pract Lang Stud 2017 Dec 03;7(12):1266-1280. [doi: 10.17507/tpls.0712.14]
10. Peixoto T, Peixoto N. Critical thinking of nursing students in clinical teaching: an integrative review. Rev Enf Ref 2017 Jun 14;IV Série(13):125-138. [doi: 10.12707/RIV16029]
11. Geng F. An content analysis of the definition of critical thinking. Asian Soc Sci 2014 Sep 29;10(19):124-128. [doi: 10.5539/ass.v10n19p124]
12. Guerrero JG. Practice rationale care model: the art and science of clinical reasoning, decision making and judgment in the nursing process. Open J Nurs 2019;09(02):79-88. [doi: 10.4236/ojn.2019.92008]
13. Reji RK, Saini SK. Critical thinking and decision making: essential skills in nursing. Int J Res Pharm Sci 2022 Mar 21;13(1):61-67. [doi: 10.26452/ijrps.v13i1.21]
14. Bland AJ, Topping A, Wood B. A concept analysis of simulation as a learning strategy in the education of undergraduate nursing students. Nurse Educ Today 2011 Oct;31(7):664-670. [doi: 10.1016/j.nedt.2010.10.013] [Medline: 21056920]

XSL•FO
RenderX

15. Foronda CL, Fernandez-Burgos M, Nadeau C, Kelley CN, Henry MN. Virtual simulation in nursing education: a systematic review spanning 1996 to 2018. Simul Healthc 2020 Feb;15(1):46-54. [doi: 10.1097/SIH.0000000000000411] [Medline: 32028447]
16. Theobald KA, Tutticci N, Ramsbotham J, Johnston S. Effectiveness of using simulation in the development of clinical reasoning in undergraduate nursing students: a systematic review. Nurse Educ Pract 2021 Nov;57:103220. [doi: 10.1016/j.nepr.2021.103220] [Medline: 34781195]
17. Wang Y, Li X, Liu Y, Shi B. Mapping the research hotspots and theme trends of simulation in nursing education: a bibliometric analysis from 2005 to 2019. Nurse Educ Today 2022 Sep;116:105426. [doi: 10.1016/j.nedt.2022.105426] [Medline: 35679627]
18. Chatzea VE, Logothetis I, Kalogiannakis M, Rovithis M, Vidakis N. Digital educational tools for undergraduate nursing education: a review of serious games, gamified applications and non-gamified virtual reality simulations/tools for nursing students. Information 2024 Jul 15;15(7):410. [doi: 10.3390/info15070410]
19. Halonen R, Thomander H, Laukkanen E. DeLone and McLean IS success model in evaluating knowledge transfer in a virtual learning environment. Int J Inf Sys Soc Change 2010;1(2):36-48. [doi: 10.4018/jissc.2010040103]
20. Molloy MA, Holt J, Charnetski M, Rossler K. Healthcare simulation standards of best PracticeTM simulation glossary. Clin Simul Nurs 2021 Sep;58:57-65. [doi: 10.1016/j.ecns.2021.08.017]
21. Watts PI, McDermott DS, Alinier G, Charnetski M, Ludlow J, Horsley E, et al. Healthcare simulation standards of best PracticeTM simulation design. Clin Simul Nurs 2021 Sep;58:14-21. [doi: 10.1016/j.ecns.2021.08.009]
22. Cole R, Flenady T, Heaton L. High fidelity simulation modalities in preregistration nurse education programs: a scoping review. Clin Simul Nurs 2023 Jul;80:64-86. [doi: 10.1016/j.ecns.2023.04.007]
23. Jraidi I, Chaouachi M, Ben Khedher A, Lajoie SP, Frasson C. Understanding clinical reasoning through visual Scanpath and brain activity analysis. Computation 2022 Jul 28;10(8):130. [doi: 10.3390/computation10080130]
24. Cant RP, Cooper SJ. Use of simulation-based learning in undergraduate nurse education: an umbrella systematic review. Nurse Educ Today 2017 Feb;49:63-71. [doi: 10.1016/j.nedt.2016.11.015] [Medline: 27902949]
25. Cant RP, Cooper SJ. Simulation in the Internet age: the place of web-based simulation in nursing education. An integrative review. Nurse Educ Today 2014 Dec;34(12):1435-1442. [doi: 10.1016/j.nedt.2014.08.001] [Medline: 25156144]
26. Duff E, Miller L, Bruce J. Online virtual simulation and diagnostic reasoning: a scoping review. Clin Simul Nurs 2016 Sep;12(9):377-384. [doi: 10.1016/j.ecns.2016.04.001]
27. Koivisto J, Haavisto E, Niemi H, Haho P, Nylund S, Multisilta J. Design principles for simulation games for learning clinical reasoning: a design-based research approach. Nurse Educ Today 2018 Jan;60:114-120. [doi: 10.1016/j.nedt.2017.10.002] [Medline: 29096383]
28. Giuffrida S, Silano V, Ramacciati N, Prandi C, Baldon A, Bianchi M. Teaching strategies of clinical reasoning in advanced nursing clinical practice: a scoping review. Nurse Educ Pract 2023 Feb;67:103548. [doi: 10.1016/j.nepr.2023.103548] [Medline: 36708638]
29. Adib-Hajbaghery M, Sharifi N. Effect of simulation training on the development of nurses and nursing students' critical thinking: a systematic literature review. Nurse Educ Today 2017 Mar;50:17-24. [doi: 10.1016/j.nedt.2016.12.011] [Medline: 28011333]
30. Lapkin S, Levett-Jones T, Bellchambers H, Fernandez R. Effectiveness of patient simulation manikins in teaching clinical reasoning skills to undergraduate nursing students: a systematic review. Clin Simul Nurs 2010 Nov;6(6):e207-e222. [doi: 10.1016/j.ecns.2010.05.005]
31. Liu K, Zhang W, Li W, Wang T, Zheng Y. Effectiveness of virtual reality in nursing education: a systematic review and meta-analysis. BMC Med Educ 2023 Sep 28;23(1):710 [FREE Full text] [doi: 10.1186/s12909-023-04662-x] [Medline: 37770884]
32. Ozdemir EK, Dinc L. Game-based learning in undergraduate nursing education: a systematic review of mixed-method studies. Nurse Educ Pract 2022 Jul;62:103375. [doi: 10.1016/j.nepr.2022.103375] [Medline: 35749962]
33. Rourke S. How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review. Int J Nurs Stud 2020 Feb;102:103466. [doi: 10.1016/j.ijnurstu.2019.103466] [Medline: 31783192]
34. Sim JJ, Rusli KD, Seah B, Levett-Jones T, Lau Y, Liaw SY. Virtual simulation to enhance clinical reasoning in nursing: a systematic review and meta-analysis. Clin Simul Nurs 2022 Aug;69:26-39 [FREE Full text] [doi: 10.1016/j.ecns.2022.05.006] [Medline: 35754937]
35. Killam LA, Silva A, Gordon R, Tyerman J, Luctkar-Flude M. Virtual screen-based clinical simulation: an integrative review of student experiences. Teach Learn Nurs 2024 Apr;19(2):e359-e368. [doi: 10.1016/j.teln.2023.12.017]
36. Wyres M, Taylor N. COVID-19: using simulation and technology-enhanced learning to negotiate and adapt to the ongoing challenges in UK healthcare education. BMJ Simul Technol Enhanc Learn 2020 May 04;6(6):317-319 [FREE Full text] [doi: 10.1136/bmjstel-2020-000642] [Medline: 35515494]
37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

XSL•FO
RenderX

38. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
40. Stenseth HV, Steindal SA, Solberg MT, Ølnes MA, Mohallem A, Sørensen AL, et al. Simulation-based learning supported by technology to enhance critical thinking in nursing students: protocol for a scoping review. JMIR Res Protoc 2022 Apr 04;11(4):e36725 [FREE Full text] [doi: 10.2196/36725] [Medline: 35373777]
41. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
42. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
43. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243 [FREE Full text] [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]
44. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
45. Steindal SA, Nes AA, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' experiences of telehealth in palliative home care: scoping review. J Med Internet Res 2020 May 05;22(5):e16218 [FREE Full text] [doi: 10.2196/16218] [Medline: 32369037]
46. Nes AA, Steindal SA, Larsen MH, Heer HC, Lærum-Onsager E, Gjevjon ER. Technological literacy in nursing education: a scoping review. J Prof Nurs 2021;37(2):320-334 [FREE Full text] [doi: 10.1016/j.profnurs.2021.01.008] [Medline: 33867086]
47. Skedsmo K, Nes AA, Stenseth HV, Hofsø K, Larsen MH, Hilderson D, et al. Simulation-based learning in palliative care in postgraduate nursing education: a scoping review. BMC Palliat Care 2023 Mar 29;22(1):30 [FREE Full text] [doi: 10.1186/s12904-023-01149-w] [Medline: 36991463]
48. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
49. Sabeh HN, Husin MH, Kee DM, Baharudin AS, Abdullah R. A systematic review of the DeLone and McLean model of information systems success in an e-learning context (2010–2020). IEEE Access 2021;9:81210-81235. [doi: 10.1109/access.2021.3084815]
50. Zaineldeen S, Hongbo L, Koffi A. Review of the DeLone and McLean model of information systems success' background and it's an application in the education setting, and association linking with technology acceptance model. Int J Res Soc Sci 2022;10(09):99-109 [FREE Full text]
51. Blakeslee JR. Effects of high-fidelity simulation on the critical thinking skills of baccalaureate nursing students: a causal-comparative research study. Nurse Educ Today 2020 Sep;92:104494. [doi: 10.1016/j.nedt.2020.104494] [Medline: 32544764]
52. Weatherspoon DL, Wyatt TH. Testing computer-based simulation to enhance clinical judgment skills in senior nursing students. Nurs Clin North Am 2012 Dec;47(4):481-491. [doi: 10.1016/j.cnur.2012.07.002] [Medline: 23137600]
53. Donovan LM, Argenbright CA, Mullen LK, Humbert JL. Computer-based simulation: effective tool or hindrance for undergraduate nursing students? Nurse Educ Today 2018 Oct;69:122-127. [doi: 10.1016/j.nedt.2018.07.007] [Medline: 30048812]
54. Fero LJ, O'Donnell JM, Zullo TG, Dabbs AD, Kitutu J, Samosky JT, et al. Critical thinking skills in nursing students: comparison of simulation-based performance with metrics. J Adv Nurs 2010 Oct;66(10):2182-2193 [FREE Full text] [doi: 10.1111/j.1365-2648.2010.05385.x] [Medline: 20636471]
55. Weatherspoon DL, Phillips K, Wyatt TH. Effect of electronic interactive simulation on senior Bachelor of Science in nursing students' critical thinking and clinical judgment skills. Clin Simul Nurs 2015 Feb;11(2):126-133. [doi: 10.1016/j.ecns.2014.11.006]
56. Brown D, Chronister C. The effect of simulation learning on critical thinking and self-confidence when incorporated into an electrocardiogram nursing course. Clin Simul Nurs 2009 Jan;5(1):e45-e52. [doi: 10.1016/j.ecns.2008.11.001]
57. Wood RY, Toronto CE. Measuring critical thinking dispositions of novice nursing students using human patient simulators. J Nurs Educ 2012 Jun;51(6):349-352. [doi: 10.3928/01484834-20120427-05] [Medline: 22533501]
58. Rush KL, Dyches CE, Waldrop S, Davis A. Critical thinking among RN-to-BSN distance students participating in human patient simulation. J Nurs Educ 2008 Nov;47(11):501-507. [doi: 10.3928/01484834-20081101-07] [Medline: 19010048]
59. Lowdermilk DL, Fishel AH. Computer simulations as a measure of nursing students' decision-making skills. J Nurs Educ 1991 Jan;30(1):34-39. [doi: 10.3928/0148-4834-19910101-09] [Medline: 1847408]

XSL•FO
RenderX

60. Ravert P. Patient simulator sessions and critical thinking. J Nurs Educ 2008 Dec;47(12):557-562. [doi: 10.3928/01484834-20081201-06] [Medline: 19112746]
61. Everett-Thomas R, Joseph L, Trujillo G. Using virtual simulation and electronic health records to assess student nurses' documentation and critical thinking skills. Nurse Educ Today 2021 Apr;99:104770. [doi: 10.1016/j.nedt.2021.104770] [Medline: 33516978]
62. Schoening AM, Sittner BJ, Todd MJ. Simulated clinical experience: nursing students' perceptions and the educators' role. Nurse Educ 2006;31(6):253-258. [doi: 10.1097/00006223-200611000-00008] [Medline: 17108788]
63. Mahoney AE, Hancock LE, Iorianni-Cimbak A, Curley MA. Using high-fidelity simulation to bridge clinical and classroom learning in undergraduate pediatric nursing. Nurse Educ Today 2013 Jun;33(6):648-654. [doi: 10.1016/j.nedt.2012.01.005] [Medline: 22341995]
64. Goodstone L, Goodstone MS, Cino K, Glaser CA, Kupferman K, Dember-Neal T. Effect of simulation on the development of critical thinking in associate degree nursing students. Nurs Educ Perspect 2013;34(3):159-162. [doi: 10.5480/1536-5026-34.3.159] [Medline: 23914457]
65. Sullivan-Mann J, Perron CA, Fellner AN. The effects of simulation on nursing students' critical thinking scores: a quantitative study. Newborn Infant Nurs Rev 2009 Jun;9(2):111-116. [doi: 10.1053/j.nainr.2009.03.006]
66. Parr MB, Sweeney NM. Use of human patient simulation in an undergraduate critical care course. Crit Care Nurs Q 2006;29(3):188-198. [doi: 10.1097/00002727-200607000-00003] [Medline: 16862020]
67. Fogg N, Kubin L, Wilson CE, Trinka M. Using virtual simulation to develop clinical judgment in undergraduate nursing students. Clin Simul Nurs 2020 Nov;48:55-58. [doi: 10.1016/j.ecns.2020.08.010]
68. Powell-Laney S, Keen C, Hall K. The use of human patient simulators to enhance clinical decision-making of nursing students. Educ Health (Abingdon) 2012 Jul;25(1):11-15. [doi: 10.4103/1357-6283.99201] [Medline: 23787379]
69. Smith A, Lollar J, Mendenhall J, Brown H, Johnson P, Roberts S. Use of multiple pedagogies to promote confidence in triage decision making: a pilot study. J Emerg Nurs 2013 Nov;39(6):660-666. [doi: 10.1016/j.jen.2011.12.007] [Medline: 22421315]
70. Burns HK, O'Donnell J, Artman J. High-fidelity simulation in teaching problem solving to 1st-year nursing students. Clin Simul Nurs 2010 May;6(3):e87-e95. [doi: 10.1016/j.ecns.2009.07.005]
71. Lewis DY, Ciak AD. The impact of a simulation lab experiencefor nursing students. Nurs Educ Perspect 2011;32(4):256-258. [Medline: 21923007]
72. Kiernan LC, Olsen DM. Improving clinical competency using simulation technology. Nursing 2020 Jul;50(7):14-19 [FREE Full text] [doi: 10.1097/01.NURSE.0000668448.43535.4f] [Medline: 32472824]
73. Rhodes ML, Curran C. Use of the human patient simulator to teach clinical judgment skills in a baccalaureate nursing program. Comput Inform Nurs 2005;23(5):256-263. [doi: 10.1097/00024665-200509000-00009] [Medline: 16166827]
74. Hudson S, Penkalski MR. High-fidelity simulation versus case study: which is best for practical nursing students? Nurs Educ Perspect 2022;43(1):49-50. [doi: 10.1097/01.NEP.0000000000000767] [Medline: 34939771]
75. Pardue KT, Holt K, Dunbar DM, Baugh N. Exploring the development of nursing clinical judgment among students using virtual reality simulation. Nurse Educ 2023;48(2):71-75. [doi: 10.1097/NNE.0000000000001318] [Medline: 36332204]
76. Badowski D, Rossler KL, Reiland N. Exploring student perceptions of virtual simulation versus traditional clinical and manikin-based simulation. J Prof Nurs 2021;37(4):683-689 [FREE Full text] [doi: 10.1016/j.profnurs.2021.05.005] [Medline: 34187664]
77. Cieslowski B, Haas T, Oh KM, Chang K, Oetjen CA. The development and pilot testing of immersive virtual reality simulation training for prelicensure nursing students: a quasi-experimental study. Clin Simul Nurs 2023 Apr;77:6-12. [doi: 10.1016/j.ecns.2023.02.001]
78. Thiele JE, Baldwin JH, Hyde RS, Sloan B, Strandquist GA. An investigation of decision theory: what are the effects of teaching cue recognition? J Nurs Educ 1986 Oct;25(8):319-324. [doi: 10.3928/0148-4834-19861001-05] [Medline: 3023568]
79. Powers K, Neustrup W, Thomas C, Saine A, Sossoman LB, Ferrante-Fusilli FA, et al. Baccalaureate nursing students' experiences with multi-patient, standardized patient simulations using telehealth to collaborate. J Prof Nurs 2020;36(5):292-300. [doi: 10.1016/j.profnurs.2020.03.013] [Medline: 33039061]
80. Montenery SM, Walker M, Sorensen E, Thompson R, Kirklin D, White R, et al. Millennial generation student nurses' perceptions of the impact of multiple technologies on learning. Nurs Educ Perspect 2013;34(6):405-409. [doi: 10.5480/10-451] [Medline: 24475603]
81. Lasater K. High-fidelity simulation and the development of clinical judgment: students' experiences. J Nurs Educ 2007 Jun;46(6):269-276. [doi: 10.3928/01484834-20070601-06] [Medline: 17580739]
82. Henrichs B, Rule A, Grady M, Ellis W. Nurse anesthesia students' perceptions of the anesthesia patient simulator: a qualitative study. AANA J 2002 Jun;70(3):219-225. [Medline: 12078470]
83. Kleinheksel A. Transformative learning through virtual patient simulations: predicting critical student reflections. Clin Simul Nurs 2014 Jun;10(6):e301-e308. [doi: 10.1016/j.ecns.2014.02.001]
84. Kuiper R, Heinrich C, Matthias A, Graham MJ, Bell-Kotwall L. Debriefing with the OPT model of clinical reasoning during high fidelity patient simulation. Int J Nurs Educ Scholarsh 2008;5:Article17. [doi: 10.2202/1548-923X.1466] [Medline: 18454731]

XSL•FO
RenderX

85. Aller L, Shelestak D, Phillips L, Reed J, Allen B. Measuring nursing student development through computer-based simulation activities. Nurse Educ 2023 Apr 10;48(6):298-303. [doi: 10.1097/NNE.0000000000001423] [Medline: 37043289]
86. Beebe SL, McNelis AM, El-Banna M, Dreifuerst KT, Zhou QP. Nailing the diagnosis: using screen-based simulation to improve factors of diagnostic reasoning in family nurse practitioner education. Clin Simul Nurs 2024 Jun;91:101528. [doi: 10.1016/j.ecns.2024.101528]
87. Cole B. Enhancing clinical reasoning and student confidence through pediatric simulation. J Pediatr Nurs 2024 Sep;78:e432-e437. [doi: 10.1016/j.pedn.2024.08.005] [Medline: 39122581]
88. Elliott EJ, Firkins JL. Use of scenario-based activities for learning and confidence in accelerated bachelor nursing students in acute settings. Int J Nurs Educ 2024 Oct 22;16(4):13-19. [doi: 10.37506/ps1p9p93]
89. Fogg N, Yousef MG, Thompson A, Bauman EB, Kardong-Edgren S. Fostering clinical decision-making using screen-based simulations aligned with the NCSBN clinical judgment measurement model. Clin Simul Nurs 2023 Nov;84:101452. [doi: 10.1016/j.ecns.2023.101452]
90. Legge TR, Holthaus A, Hallmark B, Alexander K. Preparing for the next generation NCLEX using a virtual simulation. J Nurs Educ 2024 Jul;63(7):485-489. [doi: 10.3928/01484834-20230713-01] [Medline: 37499252]
91. Perez A, Andrews A, Luebbert R. Differences in community health nursing learning outcomes based on simulation modality. Clin Simul Nurs 2024 Nov;96:101606. [doi: 10.1016/j.ecns.2024.101606]
92. Williams R, Helmer B, Elliott A, Robinson D, Jimenez FA, Faragher ME. Navigating the virtual frontier: a virtual patient simulation pilot study in prelicensure baccalaureate nursing education. Clin Simul Nurs 2024 Sep;94:101589. [doi: 10.1016/j.ecns.2024.101589]
93. Kang SJ, Hong CM, Lee H. The impact of virtual simulation on critical thinking and self-directed learning ability of nursing students. Clin Simul Nurs 2020 Dec;49:66-72. [doi: 10.1016/j.ecns.2020.05.008]
94. Kim JY, Kim EJ. Effects of simulation on nursing students' knowledge, clinical reasoning, and self-confidence: a quasi-experimental study. Korean J Adult Nurs 2015;27(5):604. [doi: 10.7475/kjan.2015.27.5.604]
95. Lee H, Han JW. Development and evaluation of a virtual reality mechanical ventilation education program for nursing students. BMC Med Educ 2022 Nov 10;22(1):775 [FREE Full text] [doi: 10.1186/s12909-022-03834-5] [Medline: 36357886]
96. Park S, Hur HK, Chung C. Learning effects of virtual versus high-fidelity simulations in nursing students: a crossover comparison. BMC Nurs 2022 Apr 27;21(1):100 [FREE Full text] [doi: 10.1186/s12912-022-00878-2] [Medline: 35473614]
97. Rim D, Shin H. Development and assessment of a multi-user virtual environment nursing simulation program: a mixed methods research study. Clin Simul Nurs 2022 Jan;62:31-41. [doi: 10.1016/j.ecns.2021.10.004]
98. Jeong Y, Lee H, Han JW. Development and evaluation of virtual reality simulation education based on coronavirus disease 2019 scenario for nursing students: a pilot study. Nurs Open 2022 Mar;9(2):1066-1076 [FREE Full text] [doi: 10.1002/nop2.1145] [Medline: 34851042]
99. Jung MJ, Roh YS. Mediating effects of cognitive load on the relationship between learning flow and clinical reasoning skills in virtual simulation learning. Clin Simul Nurs 2022 Mar;64:16-23. [doi: 10.1016/j.ecns.2021.12.004]
100. Shin H, Ma H, Park J, Ji ES, Kim DH. The effect of simulation courseware on critical thinking in undergraduate nursing students: multi-site pre-post study. Nurse Educ Today 2015 Apr;35(4):537-542. [doi: 10.1016/j.nedt.2014.12.004] [Medline: 25549985]
101. Kim GM, Lim JY, Kim EJ, Yeom M. Impact of virtual reality mental health nursing simulation on nursing students' competence. J Multidiscip Healthc 2024;17:191-202 [FREE Full text] [doi: 10.2147/JMDH.S435986] [Medline: 38226028]
102. Lee E, Baek G. Development and effects of a virtual reality simulation nursing education program combined with clinical practice based on an information processing model. Comput Inform Nurs 2023 Dec 01;41(12):1016-1025. [doi: 10.1097/CIN.0000000000001051] [Medline: 37647622]
103. Lee E, Baek G, Hwang Y. Effectiveness of the patient's severity classification competency promotion virtual reality program of nursing students during the COVID-19 pandemic period. Healthcare (Basel) 2023 Apr 13;11(8):1122 [FREE Full text] [doi: 10.3390/healthcare11081122] [Medline: 37107957]
104. Lee E, Baek G. Development and effects of adult nursing education programs using virtual reality simulations. Healthcare (Basel) 2024 Jun 30;12(13):1313 [FREE Full text] [doi: 10.3390/healthcare12131313] [Medline: 38998848]
105. Peddle M, Mckenna L, Bearman M, Nestel D. Development of non-technical skills through virtual patients for undergraduate nursing students: an exploratory study. Nurse Educ Today 2019 Feb;73:94-101. [doi: 10.1016/j.nedt.2018.11.008] [Medline: 30610960]
106. Frost J, Delaney L, Fitzgerald R. Exploring the application of mixed reality in Nurse education. BMJ Simul Technol Enhanc Learn 2020;6(4):214-219 [FREE Full text] [doi: 10.1136/bmjstel-2019-000464] [Medline: 35520006]
107. Wotton K, Davis J, Button D, Kelton M. Third-year undergraduate nursing students' perceptions of high-fidelity simulation. J Nurs Educ 2010 Nov;49(11):632-639. [doi: 10.3928/01484834-20100831-01] [Medline: 20795614]
108. Volejnikova-Wenger S, Andersen P, Clarke KA. Student nurses' experience using a serious game to learn environmental hazard and safety assessment. Nurse Educ Today 2021 Mar;98:104739. [doi: 10.1016/j.nedt.2020.104739] [Medline: 33418087]
109. Finn J, Bradley L. vSim® gerontology and inquiry-based learning enhancing clinical reasoning and preparation for practice. Teach Learn Nurs 2023 Oct;18(4):e146-e150. [doi: 10.1016/j.teln.2023.05.002]

XSL•FO
RenderX

110. Kiegaldie D, Shaw L. Virtual reality simulation for nursing education: effectiveness and feasibility. BMC Nurs 2023 Dec 19;22(1):488 [FREE Full text] [doi: 10.1186/s12912-023-01639-5] [Medline: 38114940]
111. Luo Y, Geng C, Pei X, Chen X, Zou Z. The evaluation of the distance learning combining webinars and virtual simulations for senior nursing students during the COVID-19 period. Clin Simul Nurs 2021 Aug;57:31-40 [FREE Full text] [doi: 10.1016/j.ecns.2021.04.022] [Medline: 35915811]
112. Han F, Jin S. Evaluation of the combination of virtual simulation and in-person simulation among undergraduate nursing students: a mixed methods study. Nurse Educ Pract 2024 Feb;75:103899. [doi: 10.1016/j.nepr.2024.103899] [Medline: 38277799]
113. Yang J, Zhou WJ, Zhou SC, Luo D, Liu Q, Wang AL, et al. Integrated virtual simulation and face-to-face simulation for clinical judgment training among undergraduate nursing students: a mixed-methods study. BMC Med Educ 2024 Jan 05;24(1):32 [FREE Full text] [doi: 10.1186/s12909-023-04988-6] [Medline: 38183036]
114. Li G, Gao L, Yin H, Jia Y, Zhang X, Tian H, et al. Implementation and evaluation of a virtual reality-based cognitive assessment and rehabilitation simulation course in undergraduate nursing students: a pre-post study. Clin Simul Nurs 2023 Aug;81:101430. [doi: 10.1016/j.ecns.2023.101430]
115. Li Y, Lv Y, Dorol RD, Wu J, Ma A, Liu Q, et al. Integrative virtual nursing simulation in teaching cardiopulmonary resuscitation: a blended learning approach. Australas Emerg Care 2024 Mar;27(1):37-41 [FREE Full text] [doi: 10.1016/j.auec.2023.07.006] [Medline: 37558556]
116. Tong LK, Li YY, Au ML, Ng WI, Wang SC, Liu Y, et al. The effects of simulation-based education on undergraduate nursing students' competences: a multicenter randomized controlled trial. BMC Nurs 2024 Jun 17;23(1):400 [FREE Full text] [doi: 10.1186/s12912-024-02069-7] [Medline: 38886708]
117. Sahin G, Basak T. The effect of virtual patient simulation on nursing students' clinical decision making and problem-solving skills. J Educ Res Nurs 2021 Jun 17;18(2):178-182. [doi: 10.5152/jern.2021.26504]
118. Durmaz A, Dicle A, Cakan E, Cakir S. Effect of screen-based computer simulation on knowledge and skill in nursing students' learning of preoperative and postoperative care management: a randomized controlled study. Comput Inform Nurs 2012 Apr;30(4):196-203. [doi: 10.1097/NCN.0b013e3182419134] [Medline: 22228217]
119. Karaduman GS, Basak T. Is virtual patient simulation superior to human patient simulation: a randomized controlled study. Comput Inform Nurs 2023 Jun 01;41(6):467-476. [doi: 10.1097/CIN.0000000000000957] [Medline: 36633879]
120. Edeer AD, Sarıkaya A. Views, perceptions and recommendations of nursing students on screen-based computer simulation: qualitative. Int J Psychol Educ Stud 2018 May 01;5(2):21-29. [doi: 10.17220/IJPES.2018.02.3]
121. Calik A, Kapucu S. Comparative effectiveness of developed serious game versus standardized patients' simulation in nursing education. Games Health J 2024 Dec 02;13(6):466-476. [doi: 10.1089/g4h.2024.0037] [Medline: 39093826]
122. Onay T, Gulpinar MA, Saracoglu M, Akdeniz E. Determining the effectiveness of a virtual service /patient-based education program on patient care and clinical decision-making in nursing: a quasi-experimental study. North Clin Istanb 2024;11(5):422-433. [doi: 10.14744/nci.2024.78095] [Medline: 39431025]
123. Chircop A, Cobbett S, Schofield RE, Boudreau C, Egert A, Filice S, et al. Multi-jurisdictional evaluation of sentinel city virtual simulation for community health nursing clinical education. Qual Adv Nurs Educ 2022 Oct 14;8(4):2. [doi: 10.17483/2368-6669.1352]
124. Hamidi Y, Tyerman JJ, Domingue JL, Luctkar-Flude M. The perceived effectiveness of a suicide assessment virtual simulation module for undergraduate nursing students. Clin Simul Nurs 2024 Apr;89:101509. [doi: 10.1016/j.ecns.2024.101509]
125. Maheu-Cadotte MA, Dubé V, Lavoie P. Development and contribution of a serious game to improve nursing students' clinical reasoning in acute heart failure: a multimethod study. Comput Inform Nurs 2023 Jun 01;41(6):410-420. [doi: 10.1097/CIN.0000000000000966] [Medline: 36729896]
126. Vihos J, Chute A, Carlson S, Shah M, Buro K, Velupillai N. Virtual reality simulation in a health assessment laboratory course: a mixed-methods explanatory study examining student satisfaction and self-confidence. Nurse Educ 2024;49(6):E315-E320. [doi: 10.1097/NNE.0000000000001635] [Medline: 38502593]
127. Koivisto JM, Multisilta J, Niemi H, Katajisto J, Eriksson E. Learning by playing: a cross-sectional descriptive study of nursing students' experiences of learning clinical reasoning. Nurse Educ Today 2016 Oct;45:22-28. [doi: 10.1016/j.nedt.2016.06.009] [Medline: 27429399]
128. Koivisto JM, Haavisto E, Niemi H, Katajisto J, Multisilta J. Elements explaining learning clinical reasoning using simulation games. Int J Serious Games 2016 Dec 21;3(4):29-43. [doi: 10.17083/ijsg.v3i4.136]
129. Havola S, Haavisto E, Mäkinen H, Engblom J, Koivisto JM. The effects of computer-based simulation game and virtual reality simulation in nursing students' self-evaluated clinical reasoning skills. Comput Inform Nurs 2021 May 04;39(11):725-735. [doi: 10.1097/CIN.0000000000000748] [Medline: 33941719]
130. Elcokany NM, Abdelhafez AI, Samuel Sharaby VM, Belal S. Use of computer-based scenarios for clinical teaching: impact on nursing students' decision-making skills. Healthcare (Basel) 2021 Sep 17;9(9):1228 [FREE Full text] [doi: 10.3390/healthcare9091228] [Medline: 34575002]

XSL•FO
RenderX

131. Albagawi B, Alsalamah Y, Alharbi M, Alrawili R, Babkair LA, Allari R, et al. The lived experiences of Saudi nursing students in digital clinical experience: a phenomenological study. Cureus 2024 Feb;16(2):e53830 [FREE Full text] [doi: 10.7759/cureus.53830] [Medline: 38465156]
132. Alharbi HF, Alsubaie A, Gharawi R, Ba Mazroo RB, Alajaleen S, Alsultan M, et al. The relationship between virtual simulation, critical thinking, and self-directed learning abilities of nursing students in Riyadh, Saudi Arabia. PeerJ 2024;12:e18150 [FREE Full text] [doi: 10.7717/peerj.18150] [Medline: 39399417]
133. Hwang GJ, Chang CY. Facilitating decision-making performances in nursing treatments: a contextual digital game-based flipped learning approach. Interact Learn Environ 2020 May 18;31(1):156-171. [doi: 10.1080/10494820.2020.1765391]
134. Yeh ML, Chen HH. Effects of an educational program with interactive videodisc systems in improving critical thinking dispositions for RN-BSN students in Taiwan. Int J Nurs Stud 2005 Mar;42(3):333-340. [doi: 10.1016/j.ijnurstu.2004.06.008] [Medline: 15708020]
135. Peng YC, Wu HS. The efficacy of high-fidelity telesimulation in preparing nursing students for emergency and critical patient care: a prospective study. J Nurs Res 2024 Oct 01;32(5):e348. [doi: 10.1097/jnr.0000000000000631] [Medline: 39360862]
136. Blanié A, Amorim MA, Benhamou D. Comparative value of a simulation by gaming and a traditional teaching method to improve clinical reasoning skills necessary to detect patient deterioration: a randomized study in nursing students. BMC Med Educ 2020 Feb 19;20(1):53 [FREE Full text] [doi: 10.1186/s12909-020-1939-6] [Medline: 32075641]
137. Fung JT, Zhang W, Yeung MN, Pang MT, Lam VS, Chan BK, et al. Evaluation of students' perceived clinical competence and learning needs following an online virtual simulation education programme with debriefing during the COVID-19 pandemic. Nurs Open 2021 Nov;8(6):3045-3054 [FREE Full text] [doi: 10.1002/nop2.1017] [Medline: 34331397]
138. Manik M, Gultom E, Sibuea R, Pailak H. Virtual simulation learning from Indonesian nursing students' perspectives. Open Access Maced J Med Sci 2022 Feb 15;10(G):112-117. [doi: 10.3889/oamjms.2022.8239]
139. Hosseini TM, Ahmady S, Edelbring S. Teaching clinical decision-making skills to undergraduate nursing students via web-based virtual patients during the COVID-19 pandemic: a new approach to the CyberPatientTM simulator. J Contemp Med Sci 2022 Feb 26;8(1):31-37. [doi: 10.22317/jcms.v8i2.1158]
140. Dubovi I. Online computer-based clinical simulations: the role of visualizations. Clin Simul Nurs 2019 Aug;33:35-41. [doi: 10.1016/j.ecns.2019.04.009]
141. Fawaz MA, Hamdan-Mansour AM. Impact of high-fidelity simulation on the development of clinical judgment and motivation among Lebanese nursing students. Nurse Educ Today 2016 Nov;46:36-42. [doi: 10.1016/j.nedt.2016.08.026] [Medline: 27591378]
142. Reierson I, Solli H, Bjørk IT. Nursing students' perspectives on telenursing in patient care after simulation. Clin Simul Nurs 2015 Apr;11(4):244-250. [doi: 10.1016/j.ecns.2015.02.003]
143. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical virtual simulation in nursing education: randomized controlled trial. J Med Internet Res 2019 Mar 18;21(3):e11529 [FREE Full text] [doi: 10.2196/11529] [Medline: 30882355]
144. Sterner A, Sköld R, Andersson H. Effects of blended simulation on nursing students' critical thinking skills: a quantitative study. SAGE Open Nurs 2023 May 18;9:23779608231177566 [FREE Full text] [doi: 10.1177/23779608231177566] [Medline: 37223219]
145. Uppor W, Klunklin A, Viseskul N, Skulphan S. Effects of experiential learning simulation-based learning program on clinical judgment among obstetric nursing students. Clin Simul Nurs 2024 Jul;92:101553. [doi: 10.1016/j.ecns.2024.101553]
146. McCaughey CS, Traynor MK. The role of simulation in nurse education. Nurse Educ Today 2010 Nov;30(8):827-832. [doi: 10.1016/j.nedt.2010.03.005] [Medline: 20483188]
147. Alharbi A, Nurfianti A, Mullen RF, McClure JD, Miller WH. The effectiveness of simulation-based learning (SBL) on students' knowledge and skills in nursing programs: a systematic review. BMC Med Educ 2024 Oct 07;24(1):1099 [FREE Full text] [doi: 10.1186/s12909-024-06080-z] [Medline: 39375684]
148. Cambon L, Minary L, Ridde V, Alla F. Transferability of interventions in health education: a review. BMC Public Health 2012 Jul 02;12(1):497 [FREE Full text] [doi: 10.1186/1471-2458-12-497] [Medline: 22747988]
149. Kalanlar B. Nursing education in the pandemic: a cross-sectional international study. Nurse Educ Today 2022 Jan;108:105213 [FREE Full text] [doi: 10.1016/j.nedt.2021.105213] [Medline: 34798476]
150. Christie AP, White TB, Martin PA, Petrovan SO, Bladon AJ, Bowkett AE, et al. Reducing publication delay to improve the efficiency and impact of conservation science. PeerJ 2021;9:e12245 [FREE Full text] [doi: 10.7717/peerj.12245] [Medline: 34721971]
151. Findley MG, Kikuta K, Denly M. External validity. Annu Rev Polit Sci 2021 May 11;24(1):365-393. [doi: 10.1146/annurev-polisci-041719-102556]
152. Lin H, Werner KM, Inzlicht M. Promises and perils of experimentation: the mutual-internal-validity problem. Perspect Psychol Sci 2021 Jul;16(4):854-863. [doi: 10.1177/1745691620974773] [Medline: 33593177]
153. Wong FM, Tang AC, Cheng WL. Factors associated with self-directed learning among undergraduate nursing students: a systematic review. Nurse Educ Today 2021 Sep;104:104998. [doi: 10.1016/j.nedt.2021.104998] [Medline: 34139583]
154. Miller C, Deckers C, Jones M, Wells-Beede E, McGee E. Healthcare simulation standards of best PracticeTM outcomes and objectives. Clin Simul Nurs 2021 Sep;58:40-44. [doi: 10.1016/j.ecns.2021.08.013]

XSL•FO
RenderX

155. Nadelson S, Nadelson L, Connor K. Asynchronous online simulation in nursing education: creating learning opportunities to meet current challenges. J Compr Nurs Res Care 2021;6(2):177. [doi: 10.33790/jcnrc1100177]
156. Suliman M, Ta'an W, Abdalrhim A, Tawalbeh L, Aljezawi M. The impact of online synchronous versus asynchronous classes on nursing students' knowledge and ability to make legal and ethical decisions. Nurse Educ Today 2022 Feb;109:105245. [doi: 10.1016/j.nedt.2021.105245] [Medline: 34952302]
157. McEnroe-Petitte D, Farris C. Using gaming as an active teaching strategy in nursing education. Teach Learn Nurs 2020 Jan;15(1):61-65. [doi: 10.1016/j.teln.2019.09.002]
158. Gold S. Design and effectiveness of a self-study pedagogical approach to using a simulation game in the classroom. Develop Simulation Bus Exp Learn 2016;43:136-140 [FREE Full text]
159. Dayagbil FT, Palompon DR, Garcia LL, Olvido MM. Teaching and learning continuity amid and beyond the pandemic. Front Educ 2021 Jul 23;6:678692. [doi: 10.3389/feduc.2021.678692]
160. Michel A, Ryan N, Mattheus D, Knopf A, Abuelezam NN, Stamp K, et al. Undergraduate nursing students' perceptions on nursing education during the 2020 COVID-19 pandemic: a national sample. Nurs Outlook 2021;69(5):903-912 [FREE Full text] [doi: 10.1016/j.outlook.2021.05.004] [Medline: 34183191]
161. Anthony B, Kamaludin A, Romli A, Raffei AF, Phon DN, Abdullah A, et al. Blended learning adoption and implementation in higher education: a theoretical and systematic review. Tech Know Learn 2020 Oct 07;27(2):531-578. [doi: 10.1007/s10758-020-09477-z]
162. Woo Y, Reeves TC. Meaningful interaction in web-based learning: a social constructivist interpretation. Internet High Educ 2007 Jan;10(1):15-25. [doi: 10.1016/j.iheduc.2006.10.005]
163. Henderson M, Ryan T, Phillips M. The challenges of feedback in higher education. Assess Eval High Educ 2019 Apr 05;44(8):1237-1252. [doi: 10.1080/02602938.2019.1599815]
164. Decker S, Alinier G, Crawford SB, Gordon RM, Jenkins D, Wilson C. Healthcare simulation standards of best PracticeTM the debriefing process. Clin Simul Nurs 2021 Sep;58:27-32. [doi: 10.1016/j.ecns.2021.08.011]
165. Handeland JA, Prinz A, Ekra EM, Fossum M. The role of manikins in nursing students' learning: a systematic review and thematic metasynthesis. Nurse Educ Today 2021 Mar;98:104661 [FREE Full text] [doi: 10.1016/j.nedt.2020.104661] [Medline: 33298327]
166. Martín-Gutiérrez J, Mora CE, Añorbe-Díaz B, González-Marrero A. Virtual technologies trends in education. Eurasia J Math Sci Tech Ed 2017 Feb;13(2):469-486. [doi: 10.12973/eurasia.2017.00626a]
167. Cypress BS, Caboral-Stevens M. "Sense of presence" in immersive virtual reality environment: an evolutionary concept analysis. Dimens Crit Care Nurs 2022;41(5):235-245. [doi: 10.1097/DCC.0000000000000538] [Medline: 35905425]
168. Dubovi I. Cognitive and emotional engagement while learning with VR: the perspective of multimodal methodology. Comput Educ 2022 Jul;183:104495. [doi: 10.1016/j.compedu.2022.104495]
169. Shorey S, Ng ED. The use of virtual reality simulation among nursing students and registered nurses: a systematic review. Nurse Educ Today 2021 Mar;98:104662. [doi: 10.1016/j.nedt.2020.104662] [Medline: 33203545]
170. Harerimana A, Mtshali NG. Using exploratory and confirmatory factor analysis to understand the role of technology in nursing education. Nurse Educ Today 2020 Sep;92:104490. [doi: 10.1016/j.nedt.2020.104490] [Medline: 32516637]
171. Sormunen M, Saaranen T, Heikkilä A, Sjögren T, Koskinen C, Mikkonen K, et al. Digital learning interventions in higher education: a scoping review. Comput Inform Nurs 2020 Dec;38(12):613-624. [doi: 10.1097/CIN.0000000000000645] [Medline: 32520782]
172. Bauce K, Kaylor MB, Staysniak G, Etcher L. Use of theory to guide integration of virtual reality technology in nursing education: a scoping study. J Prof Nurs 2023 Jan;44:1-7. [doi: 10.1016/j.profnurs.2022.10.003] [Medline: 36746594]
173. Pusic MV, Boutis K, McGaghie WC. Role of scientific theory in simulation education research. Simul Healthc 2018 Jun;13(3S Suppl 1):S7-14. [doi: 10.1097/SIH.0000000000000282] [Medline: 29369961]
174. Varpio L, Paradis E, Uijtdehaage S, Young M. The distinctions between theory, theoretical framework, and conceptual framework. Acad Med 2020 Jul;95(7):989-994. [doi: 10.1097/ACM.0000000000003075] [Medline: 31725464]

Abbreviations

CBS: computer-based simulation
D&M: DeLone and McLean
HPS: human-patient simulator
IS: information system
IVR: immersive virtual reality
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RQ: research question
SBL: simulation-based learning
TES: technology-enhanced simulation
VR: virtual reality

XSL•FO
RenderX

VS: virtual simulation

Edited by A Mavragani; submitted 23.03.24; peer-reviewed by F Oliveira, F Al Dhabbari, S Funghetto; comments to author 07.11.24; revised version received 12.12.24; accepted 16.01.25; published 18.02.25.

Please cite as:
Stenseth HV, Steindal SA, Solberg MT, Ølnes MA, Sørensen AL, Strandell-Laine C, Olaussen C, Farsjø Aure C, Pedersen I, Zlamal J, Gue Martini J, Bresolin P, Linnerud SCW, Nes AAG
Simulation-Based Learning Supported by Technology to Enhance Critical Thinking in Nursing Students: Scoping Review
J Med Internet Res 2025;27:e58744
URL: https://www.jmir.org/2025/1/e58744
doi:10.2196/58744
PMID:

©Hege Vistven Stenseth, Simen A Steindal, Marianne Trygg Solberg, Mia Alexandra Ølnes, Anne Lene Sørensen, Camilla Strandell-Laine, Camilla Olaussen, Caroline Farsjø Aure, Ingunn Pedersen, Jaroslav Zlamal, Jussara Gue Martini, Paula Bresolin, Silje Christin Wang Linnerud, Andréa Aparecida Gonçalves Nes. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Virtual Reality Enhanced Exercise Training in Upper Limb Function of Patients With Stroke: Meta-Analytic Study

Shiqi Xu[1,2*]; Yanwen Xu[1*]; Ruyi Wen[3]; Jun Wang[1]; Yuyu Qiu[2]; Chetwyn CH Chan[4*]

[1]Department of Rehabilitation Medicine, Wuxi Ninth People's Hospital Affiliated to Soochow University, Wuxi, Jiangsu, China

[2]Wuxi School of Medicine, Jiangnan University, Wuxi, Jiangsu, China

[3]School of Rehabilitation Medicine, Nanjing Medical University, Nanjing, China

[4]Department of Psychology, The Education University of Hong Kong, Hong Kong SAR, China

[*]these authors contributed equally

Corresponding Author:
Chetwyn CH Chan
Department of Psychology
The Education University of Hong Kong
Lo Ping Road, Tai Po, New Territories
Hong Kong SAR
China
Phone: 852 2948 6322
Email: cchchan@eduhk.hk

Abstract

Background: Recovery of upper limb function after stroke secondary to ischemia or hemorrhage is crucial for patients' independence in daily living and quality of life. Virtual reality (VR) is a promising computer-based technology designed to enhance the effects of rehabilitation; however, the results of VR-based interventions remain equivocal.

Objective: This study aims to review the plausible factors that may have influenced VR's therapeutic effects on improving upper limb function in patients with stroke, with the goal of synthesizing an optimal VR intervention protocol.

Methods: The databases PubMed, EMBASE, Web of Science, and Cochrane Library were queried for English-language papers published from May 2022 onward. Two reviewers independently extracted data from the included papers, and discrepancies in their findings were resolved through consensus during joint meetings. The risk of bias was assessed using the Physiotherapy Evidence Database Scale and the Methodological Index for Non-Randomized Studies. Outcome variables included the Action Research Arm Test, Box-Block Test, Functional Independence Measure, Upper Extremity Fugl-Meyer Assessment, and Wolf Motor Function Test. The plausible factors examined were age, total dosage (hours), trial length (weeks), session duration (hours/session), frequency (sessions/week), and VR content design. The Bonferroni adjustment was applied to *P* values to prevent data from being incorrectly deemed statistically significant.

Results: The final sample included 15 articles with a total of 1243 participants (age range 48.6-75.59 years). Participants in the VR therapy (VRT) group (n=455) demonstrated significantly greater improvements in upper limb function and independence in activities of daily living compared with those in the conventional therapy group (n=301). Significant factors contributing to improved outcomes in upper limb function were younger age (mean difference [MD] 5.34, 95% CI 2.18-8.5, *P*<.001; I2=0%), interventions lasting more than 15 hours (MD 9.67, 95% CI 4.19-15.15, *P*<.001; I2=0%), trial lengths exceeding 4 weeks (MD 4.02, 95% CI 1.39-6.65, *P*=.003; I2=15%), and more than 4 sessions per week (MD 3.48, 95% CI 0.87-6.09, *P*=.009; I2=0%). However, the design of the VR content, including factors such as the number of features (eg, offering exercise and functional tasks; individualized goals; activity quantification; consideration of comorbidities and baseline activity level; addressing patient needs; aligning with patient background such as education level; patient-directed goals and interests; goal setting; progressive difficulty levels; and promoting self-efficacy), did not demonstrate significant effects (MD 3.89, 95% CI –6.40 to 1.09; effect Z=1.36, *P*=.16).

Conclusions: Greater VR effects on improving upper limb function in patients with stroke were associated with higher training doses (exceeding 15 hours) delivered over 4-6 weeks, with shorter sessions (approximately 1 hour) scheduled 4 or more times per week. Additionally, younger patients appeared to benefit more from the VR protocol compared with older patients.

(J Med Internet Res 2025;27:e66802) doi:10.2196/66802

XSL•FO
RenderX

KEYWORDS

virtual reality; stroke; upper limb function; exercise training; meta-analysis

Introduction

Recovery of upper limb function after stroke is crucial for patients' independent daily living and quality of life [1-3]. Previous studies have demonstrated the positive effects of incorporating neurodevelopmental, proprioceptive neuromuscular, and motor relearning theories in designing interventions to improve patients' upper limb function [4-7]. Advances in technology have enabled virtual reality (VR) to serve as a common medium for delivering interventions aimed at enhancing upper limb function [8]. VR is a promising computer technology that allows users to interact with a simulated multisensory environment [9]. It is designed to enhance the effectiveness of rehabilitation and provide feedback on performance. However, a review of clinical studies examining the use of VR as an intervention for patients with stroke produced inconsistent findings. Compared with usual practice, VR interventions were found to be more effective in improving upper-extremity and hand functions in patients with stroke [10-12]. By contrast, 1 study reported that VR interventions were effective in improving cognitive functions but not motor functions in patients with traumatic brain injury [13], while another study found no significant differences between VR and conventional therapy for patients with stroke [14].

A review of recent studies suggested that the inconsistent results described above may stem from variations in the design, content, duration, and intensity of VR interventions [14], as well as patient-specific characteristics such as independence level and emotional state [12]. For example, VR exercise training was found to be more effective than usual practice for improving upper limb function in patients with stroke only when they participated in a high-dose program (eg, 15 hours or more) [15], whereas this effect was not observed with a low-dose program (eg, 12 hours) [16]. In the high-dose study, the VR exercise training included a variety of limb movements with varying levels of difficulty. Conversely, the low-dose study used a VR protocol consisting of 3 modes of hand and arm movements performed at different speeds. Inconsistent findings are also evident in studies evaluating the effects of VR in patients with Parkinson's disease. For example, a VR game using a 1-group pre- and posttest design demonstrated improvements in hand grip strength, dexterity, and speed among patients with stroke [17]. The VR game comprised 4 activities delivered in sequence: the reach game, the sequence game, the grab game, and the flip game. The intervention involved 30-minute individual sessions conducted 3 times per week over 6 weeks (18 sessions, totaling 9 hours). These results contrast with those of a randomized controlled trial (RCT), which found no significant effects [18]. The RCT protocol involved tracking and touching moving objects displayed on a screen while maintaining a standing balance. The schedule consisted of 40-minute individual sessions, 3 sessions per week over 6 weeks (18 sessions, totaling 12 hours). Many national clinical guidelines for stroke emphasize the importance of applying recovery principles. Beyond dosage (ie, duration per session, frequency, and overall intervention period), the design of VR training content appears to influence treatment outcomes. For instance, the 2023 edition of the National Clinical Guideline for Stroke for the United Kingdom and Ireland [19] recommends therapy for motor recovery and function that incorporates individualized content, a variety of practice modes, intensity, and the practice of functional skills. Additionally, studies have highlighted the critical role of repetitions in reshaping neuronal structures and enhancing motor system function following brain injury [20-22]. Research on nonprimates has shown that performing repeated reaching tasks can induce synaptogenesis and changes in cortical representations [23,24]. In human studies, evidence suggests that functional recovery of the upper limbs, driven by the brain's plasticity, requires at least 300 repetitions per day [25] or 100 active movements or more per treatment session [26].

The inconsistent findings regarding the effects of VR-based interventions motivated us to undertake this meta-analytic study. We aimed to examine 6 plausible factors that might influence the therapeutic effects of VR on improving upper limb function in patients with stroke. These factors included patients' age, dosage (hours), delivery schedule (session duration, frequency, and trial length), total dosage (hours), and content design. We hypothesized that all these factors would significantly impact intervention outcomes. Identifying the factors contributing to variability in intervention outcomes would enhance the content design of VR interventions and provide guidance on setting appropriate dosages and schedules to achieve better treatment results for patients.

Methods

Design

This study followed the scope and methods outlined in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and its accompanying checklist [27].

Search Strategy

The relevant materials for review and analysis were extracted on May 7, 2022, from PubMed (1966 to present), EMBASE via Ovid (1974 to present), Web of Science (1956 to present), and the Cochrane Library databases (no date restriction). The medical terms and free-text search terms were related to VR, upper limb, and rehabilitation. Detailed search strategies for each database are provided in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria set for the studies are presented in Textboxes 1 and 2.

XSL•FO
RenderX

Textbox 1. Inclusion criteria.

- Randomized controlled trial and group comparison studies.
- Results published in English in peer-reviewed journals.
- The treatment group used virtual reality–enhanced exercise alone or combined with conventional therapy, while the control group used conventional therapy alone or without treatment.
- Patients showed upper limb functional disabilities with outcome variables including the Action Research Arm Test, Box-Block Test, Functional Independence Measure, Upper Extremity Fugl-Meyer Assessment, and Wolf Motor Function Test.
- Means, SD, and effect sizes were presented; or means, SEs, *t* values, or *P* values were presented; or range or 95% CIs were presented.

Textbox 2. Exclusion criteria.

- Abstracts, case reports, or review studies.
- Non-English or nonbilingual journal articles.
- Research design without group comparison.
- Patients without stroke or measuring lower limb functions of patients with stroke.
- Missing data or additional information (means, SD, and effect sizes) cannot be obtained from the corresponding author within 1 month after initial contact.

Study Selection

All studies were imported into the EndNote referencing software (Thomson Research Soft, version X9), and duplicates were identified and removed. Two reviewers, SX and JW, co-authors of this paper, independently conducted the relevance screening. The screening process began with an evaluation of the title and abstract content of each publication based on the inclusion and exclusion criteria outlined above. Reviewers were instructed to read the full text if the abstract contained ambiguous content. Discrepancies between the 2 reviewers were resolved through consensus during joint meetings.

Data Extraction

Two research team members (SX and YX) independently extracted data from the included papers. For each paper, the researchers identified the trials and specific content, recording them in preformatted tables. The recorded information included trial design and setup, sample sizes of all groups, intervention program details, overall characteristics, and clinical outcome results. Discrepancies between the 2 reviewers were resolved through consensus during joint meetings.

Assessment of Risk of Bias

Two research team members (YX and YQ) independently assessed the risk of bias in the included trials using the Physiotherapy Evidence Database (PEDro) Scale [28] and the Methodological Index for Non-Randomized Studies (MINORS) [29]. The 11-item PEDro Scale evaluates the quality of the trials, with a score of “1” indicating the criterion is met and “0” indicating it is not. PEDro quality grades are categorized as high (6-8: good; 9-10: excellent), fair (4-5: acceptable), and poor (≤4). For nonrandomized trials, the first 8 items of the MINORS were used to assess methodological quality (risk of bias). Each MINORS item is scored as follows: “0” for not reported, “1” for reported but insufficient, and “2” for reported and sufficient. For a single-group trial, the maximum score is 16, while for a 2-group trial, it is 24. A score of 0-9 for a single-group trial indicates poor quality and 10-16 indicates high quality. In this study, the cut-off for inclusion was set at high quality or above: a score of 6 or higher on the PEDro Scale or 10 or higher on the MINORS [30]. Discrepancies in the scores assigned by the 2 reviewers were resolved through consensus during joint meetings.

The principles of therapy content (motor recovery and function) outlined in the 2023 edition of the National Clinical Guideline for Stroke for the United Kingdom and Ireland [19] were adopted as feature criteria to guide the analysis of VR content. Ten content features were identified: offering exercise and functional tasks; setting individualized goals; quantifying activity; considering comorbidities and baseline activity levels; incorporating patient needs; matching the patient’s background, such as education level; aligning with the patient’s goals and interests; involving goal setting; including progressive difficulty levels; and promoting self-efficacy. The content features of each study were collated and classified into 3 categories: “+” for few features (2-3), “++” for medium features (4-5), and “+++” for many features (≥6).

Data Synthesis and Analysis

We conducted analyses using Review Manager (RevMan) version 5.3 software. As some studies [14,31-34] reported results in terms of the median, minimum, maximum, or IQR, we converted these to means and SDs as proposed in the study by Wan et al [35]. For studies reporting only means, SEs, *t* values, or *P* values, we computed the SDs [31,32,35]. For studies reporting only ranges or 95% CIs, we calculated the means [31,32,35]. Statistical heterogeneity for each meta-analysis was assessed using the Cochran *Q* test and the I^2 statistic. Fixed-effect models were applied when heterogeneity among trials was nonsignificant (I^2<50%), while random-effect models were used for significant heterogeneity (I^2>50%). The outcome variables included the Action Research Arm Test (ARAT), Box and Block Test (BBT), Functional Independence Measure (FIM),

XSL•FO
RenderX

Upper Extremity Fugl-Meyer Assessment (FMA-UE), and Wolf Motor Function Test (WMFT). All results were reported as mean differences (MDs) with 95% CIs. The plausible factors considered were age (with the World Health Organization defining older age as 60 years and above), total dosage (hours), trial length (weeks), duration of a session (hours/session), frequency (sessions/week), and VR content design. The cut-off value was set based on the findings of previous literature [36-39] and the principle that at least two articles must be included per subgroup. As only 15 studies were available for analysis, multiple comparisons were conducted separately for each of the 6 factors to meet this criterion. Statistical significance for the meta-analysis (whole group) was set at $P<.05$. For the subgroup analyses, the potential proliferation of type I and type II errors across the 6 comparisons was controlled by applying the Bonferroni adjustment, setting statistical significance at $P<.008$ (0.05 divided by 6).

Results

Study Selection

Details of the steps taken to select the studies are presented in the PRISMA flowchart (Figure 1; also see Multimedia Appendix 2). A total of 745 studies were retrieved from 4 databases through an electronic search. Of these, 237 studies were excluded due to duplication. An additional 303 studies were excluded after reviewing their titles and abstract contents. Full-text screening of the remaining 205 studies resulted in the exclusion of 185 studies for not meeting the inclusion criteria. Among the remaining 20 studies, 5 were further excluded for being rated below the “good” cut-offs on the PEDro and MINORS scales. The final number of studies included in the analyses was 15 [11,16,33,34,40-50], consisting of 12 RCTs and 3 non-RCTs.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing 15 records entering the analyses. VR: virtual reality.

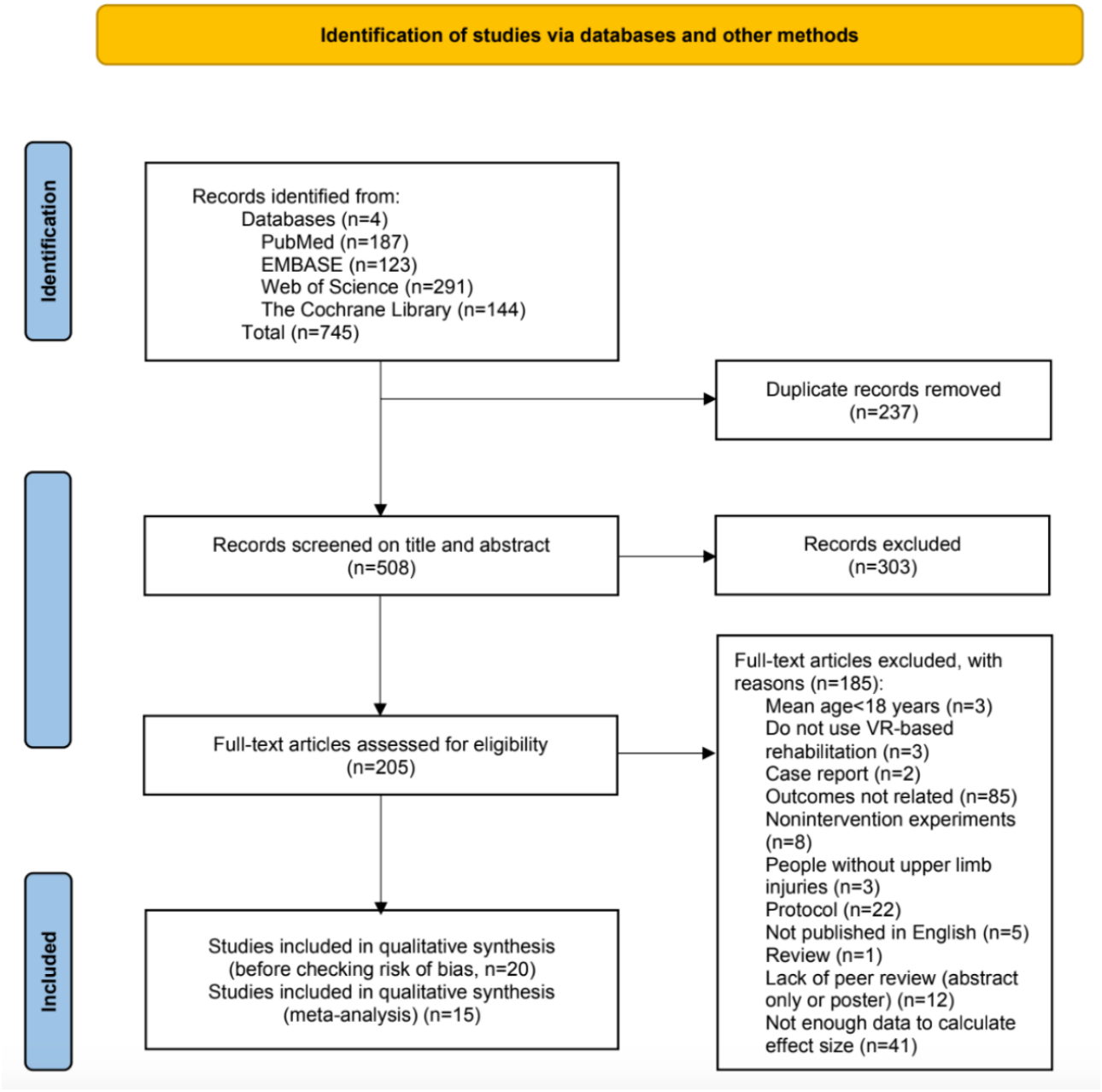

Characteristics of the Included Trials

The 15 studies included a total of 28 groups, comprising 15 VR therapy (VRT) groups and 13 conventional therapy (CON) groups, with 1203 participants in total (Table 1). All studies were published between 2013 and 2022. At the total group level (14 studies), the VRT and CON groups included 411/278 and 294/177 males/females, respectively. One study did not report gender as a grouping variable and was therefore excluded from the gender analysis at the group level. The average ages of participants in the VRT groups ranged from 49.0 to 73.0 years, while those in the CON groups ranged from 53.4 to 75.6 years. The intervention durations for the VRT groups varied from 2 weeks to 3 months. Fourteen studies [11,16,33,34,40-47,49,50] involved patients with stroke secondary to ischemia or hemorrhage, while 1 study [48] included only patients with stroke secondary to ischemia.

XSL•FO
RenderX

Table 1. Characteristics of the trials contained in 15 studies satisfying the inclusion criteria.

Study	Study design	Participants Intervention/control (no/no)	Age (years), mean Intervention/control (yes/yes)	Intervention	Control	Dosage	Total dosage (hours)	Outcomes included in this review
Brunner et al [40]	RCT[a]	62/58	62/62	YouGrabber system	Conventional training	60 minutes/day, 4 days/week, 4 weeks	16	ARAT[b], BBT[c], and FIM[d]
Abd El-Kafy et al [41]	RCT	20/20	54.13/53.42	Conventional physiotherapy and training with robot-mediated VR[e] gaming	Conventional physiotherapy	2 hours/day, 3 days/week, 12 weeks	72	ARAT and WMFT[f]
Gueye et al [42]	RCT	25/25	66.56/68.12	Armeo Spring	Conventional physiotherapy	45 minutes/day, 4 days/week, 3 weeks	9	FMA-UE[g] and FIM
Hsu et al [11]	RCT	18/17	52.9/56.9	VR-based mirror therapy and usual care	Conventional occupational therapy and usual care	50 minutes/day, 2 days/week, 9 weeks	15	FMA-UE and BBT
Kiper et al [43]	RCT	23/21	64.3/64.3	VR rehabilitation system and traditional rehabilitation	Traditional rehabilitation	2 hours/day, 5 days/week, 4 weeks	40	FMA-UE and FIM
Kiper et al [44]	RCT	68/68	62.5/66	VR rehabilitation system and traditional rehabilitation	Traditional rehabilitation	2 hours/day, 5 days/week, 4 weeks	40	FMA-UE and FIM
Rong et al [45]	RCT	20/20	56.25/62.3	Camera-based mirror visual feedback and robot-assisted training	Sham-mirror visual feedback and robot-assisted training	1.5 hours/day, 5 days/week, 4 weeks	30	FMA-UE and FIM
Schuster-Amft et al [16]	RCT	22/32	61.3/61.2	VR-based training system	Conventional therapy	45 minutes/day, 4 days/week, 4 weeks	12	BBT
Sin and Lee [46]	RCT	18/17	71.78/75.59	Xbox Kinect and conventional occupational therapy	Conventional occupational therapy	1 hour/day, 3 days/week, 6 weeks	18	FMA-UE and BBT
Taveggia et al [47]	RCT	27/27	73/68	Armeo spring and conventional treatment	Conventional treatment	1 hour/day, 5 days/week, 6 weeks	30	FIM
Yao et al [48]	RCT	20/20	63/66.2	C-tDCS[h] and VR therapy and conventional occupational and physical therapies	Sham tDCS and VR and conventional occupational and physical therapies	20 minutes/day, 5 days/week, 2 weeks	3.33	ARAT and FMA-UE
Zheng et al [49]	RCT	58/54	65.4/66.2	Low-frequency repetitive transcranial magnetic stimulation and VR and standard rehabilitation therapy for stroke	Sham repetitive transcranial magnetic stimulation and VR training and standard rehabilitation therapy for stroke	2 hours/day, 6 days/week, 4 weeks	48	WMFT
Turolla et al [50]	Non-RCT	263/113	60.2/65.4	Upper limb conventional and reinforced feedback in the virtual environment therapies	Upper limb conventional therapy	2 hours/day, 5 days/week, 4 weeks	40	FMA-UE and FIM
Borstad et al [33]	Non-RCT	16/N/A[i]	49/N/A	Constraint-induced movement therapy (recovery rapids and highly trained practitioners)	N/A	3 hours/day, 5 days/week, 2 weeks	30	ARAT and WMFT

XSL•FO
RenderX

Study	Study design	Participants	Age (years), mean	Intervention	Control	Dosage	Total dosage (hours)	Outcomes included in this review
		Intervention/control (no/no)	Intervention/control (yes/yes)					
Sebastián-Romagosa et al [34]	Non-RCT	51/N/A	60.52/N/A	Brain-computer interface systems	N/A	1 hour/day, 2 days/week, 3 months	25	FMA-UE

[a]RCT: randomized controlled trial.

[b]ARAT: Action Research Arm Test.

[c]BBT: Box-Block Test.

[d]FIM: Functional Independence Measure.

[e]VR: virtual reality.

[f]WMFT: Wolf Motor Function Test.

[g]FMA-UE: Upper Extremity Fugl-Meyer Assessment.

[h]C-tDCS: cathodal transcranial direct current stimulation.

[i]N/A: not applicable.

Risk of Bias

The mean PEDro Scale score for the 12 studies with a randomized clinical trial design was 7.2 (Table 2). All these studies reported the method of randomization, baseline comparability, blinding of outcome assessors, participant dropout rates of less than 15%, between-group outcome analysis, and point estimates with variability. Most of these studies also reported adequate assignment concealment (n=7, 58%) and intention-to-treat analyses (n=4, 33%). Among the studies with a nonrandomized clinical trial design (n=3), it is noteworthy that they demonstrated clear research objectives and prospective data collection but lacked a follow-up period and did not perform sample size calculations (Table 3).

XSL•FO
RenderX

Table 2. Physiotherapy Evidence Database Scale score for randomized clinical trials included in the review.

Studies	Total score	Methodological quality	Number of items of the Physiotherapy Evidence Database Scale										
			1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]
Ikbali Afsar et al [14]	5	Fair	N/A[l]	✓[m]	N/A	✓	N/A	N/A	✓	N/A	N/A	✓	✓
Alves et al [10]	5	Fair	N/A	✓	N/A	✓	N/A	N/A	✓	N/A	N/A	✓	✓
Brunner et al [40]	7	Good	N/A	✓	✓	✓	✓	N/A	N/A	✓	N/A	✓	✓
Abd El-Kafy et al [41]	6	Good	N/A	✓	✓	✓	N/A	N/A	✓	✓	N/A	✓	✓
Ersoy and Iyigun [51]	4	Poor	N/A	✓	✓	✓	N/A	N/A	N/A	N/A	N/A	✓	✓
Gueye et al [42]	6	Good	N/A	✓	✓	✓	N/A	N/A	N/A	✓	✓	✓	✓
Hsu et al [11]	7	Good	N/A	✓	✓	✓	N/A	N/A	✓	✓	N/A	✓	✓
Junior et al [52]	5	Fair	N/A	✓	N/A	✓	N/A	N/A	✓	N/A	N/A	X	X
Kiper et al [43]	6	Good	N/A	✓	N/A	✓	N/A	N/A	✓	✓	N/A	✓	✓
Kiper et al [44]	8	Good	N/A	✓	✓	✓	N/A	N/A	✓	✓	✓	✓	✓
Rong et al [45]	8	Good	N/A	✓	✓	✓	N/A	N/A	✓	✓	✓	✓	✓
Schuster-Amft et al [16]	8	Good	N/A	✓	✓	✓	N/A	✓	✓	✓	N/A	✓	✓
Sin and Lee [46]	7	Good	N/A	✓	N/A	✓	N/A	✓	✓	✓	N/A	✓	✓
Taveggia et al [47]	7	Good	N/A	✓	✓	✓	✓	N/A	N/A	✓	N/A	✓	✓
Yao et al [48]	8	Good	N/A	✓	✓	✓	N/A	✓	✓	✓	N/A	✓	✓
Zheng et al [49]	8	Good	N/A	✓	N/A	✓	N/A	✓	✓	✓	✓	✓	✓

[a]Item 1: Specified eligibility criteria.

[b]Item 2: Random allocation.

[c]Item 3: Concealed allocation.

[d]Item 4:Baseline comparability.

[e]Item 5: Participants were blinded.

[f]Item 6: Therapists were blinded.

[g]Item 7: Assessors were blinded.

[h]Item 8: Adequate follow-up.

[i]Item 9: Intention-to-treat analysis.

[j]Item 10: Between-group comparisons.

[k]Item 11: Point estimates and variability.

[l]N/A: not applicable.

[m]The '✓' symbol indicates that the item where it is found has been punctuated.

XSL•FO
RenderX

Table 3. Methodological Index for Non-Randomized Studies Scale score for nonrandomized clinical trials included in the review.

Studies	Total score	Methodological quality	Number of items of the Methodological Index for Non-Randomized Studies Scale[a]											
			1[b]	2[c]	3[d]	4[e]	5[f]	6[g]	7[h]	8[i]	9[j]	10[k]	11[l]	12[m]
Turolla et al [50]	15/24	Good	2	2	2	2	2	0	0	0	2	2	0	1
Borstad et al [33]	10/16	Good	2	2	2	2	1	1	0	0	N/A	N/A	N/A	N/A
Kizony et al [53]	7/16	Poor	2	1	2	2	0	0	0	0	N/A	N/A	N/A	N/A
Sebastián-Romagosa et al [34]	10/16	Good	2	2	2	2	1	1	0	0	N/A	N/A	N/A	N/A

[a]The items are scored 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate). The global ideal score is 16 for noncomparative studies and 24 for comparative studies.

[b]Item 1: A clearly stated aim.

[c]Item 2: Inclusion of consecutive patients.

[d]Item 3: Prospective collection of data.

[e]Item 4: Endpoints appropriate to the aim of the study.

[f]Item 5: Unbiased assessment of the study endpoint.

[g]Item 6: Follow-up period appropriate to the aim of the study.

[h]Item 7: Loss to follow-up less than 5%.

[i]Item 8: Prospective calculation of the study size.

[j]Item 9: An adequate control group.

[k]Item 10: Contemporary groups.

[l]Item 11: Baseline equivalence of groups.

[m]Item 12: Adequate statistical analyses.

Meta-Analysis

Whole-Group: Upper Limb Function (FMA-UE)

Participants in the VRT group (n=455) demonstrated significantly greater improvement than those in the CON group (n=301) in FMA-UE scores (MD 3.80, 95% CI 1.47-6.13, P=.001; I^2=7%; Figure 2; see also [11,16,40-50]). Within the VRT group, participants showed significant increases in FMA-UE scores following the intervention (MD 9.95, 95% CI 6.49-13.41, P<.001; I^2=56%; Figure 3; see also [11,16,33,34,40-50]).

XSL•FO
RenderX

Figure 2. Forest plot of comparisons of between-group outcomes on upper limb function. CON: conventional therapy; VRT: virtual reality therapy.

(a) FMA-UE between-group

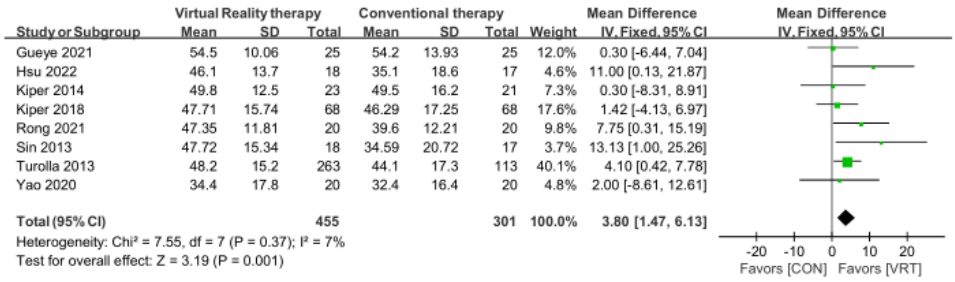

(b) ARAT between-group

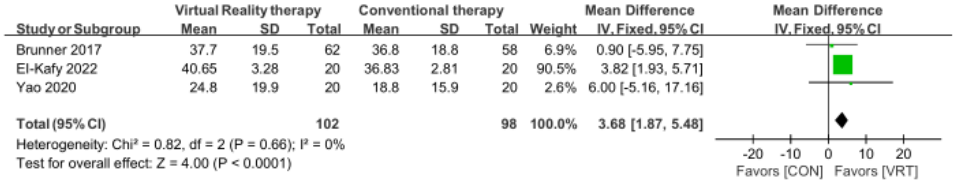

(c) FIM between-group

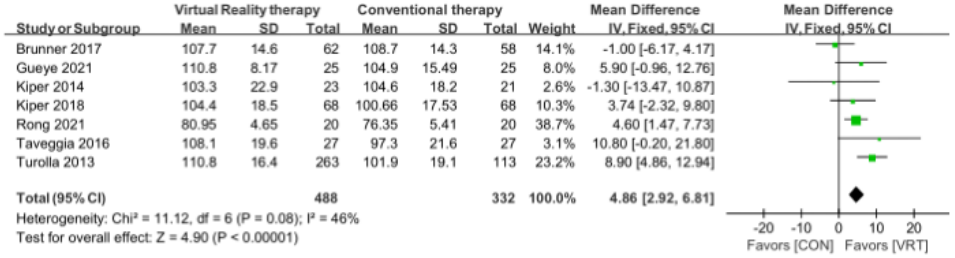

(d) WMFT between-group

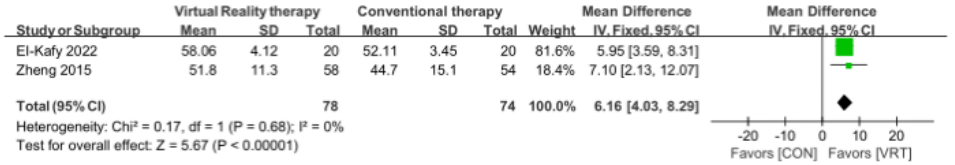

(e) BBT between-group

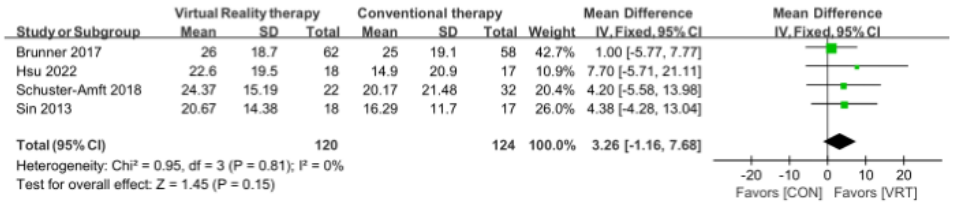

XSL•FO
RenderX

Figure 3. Forest plot of comparisons of within-group outcomes on upper limb function.

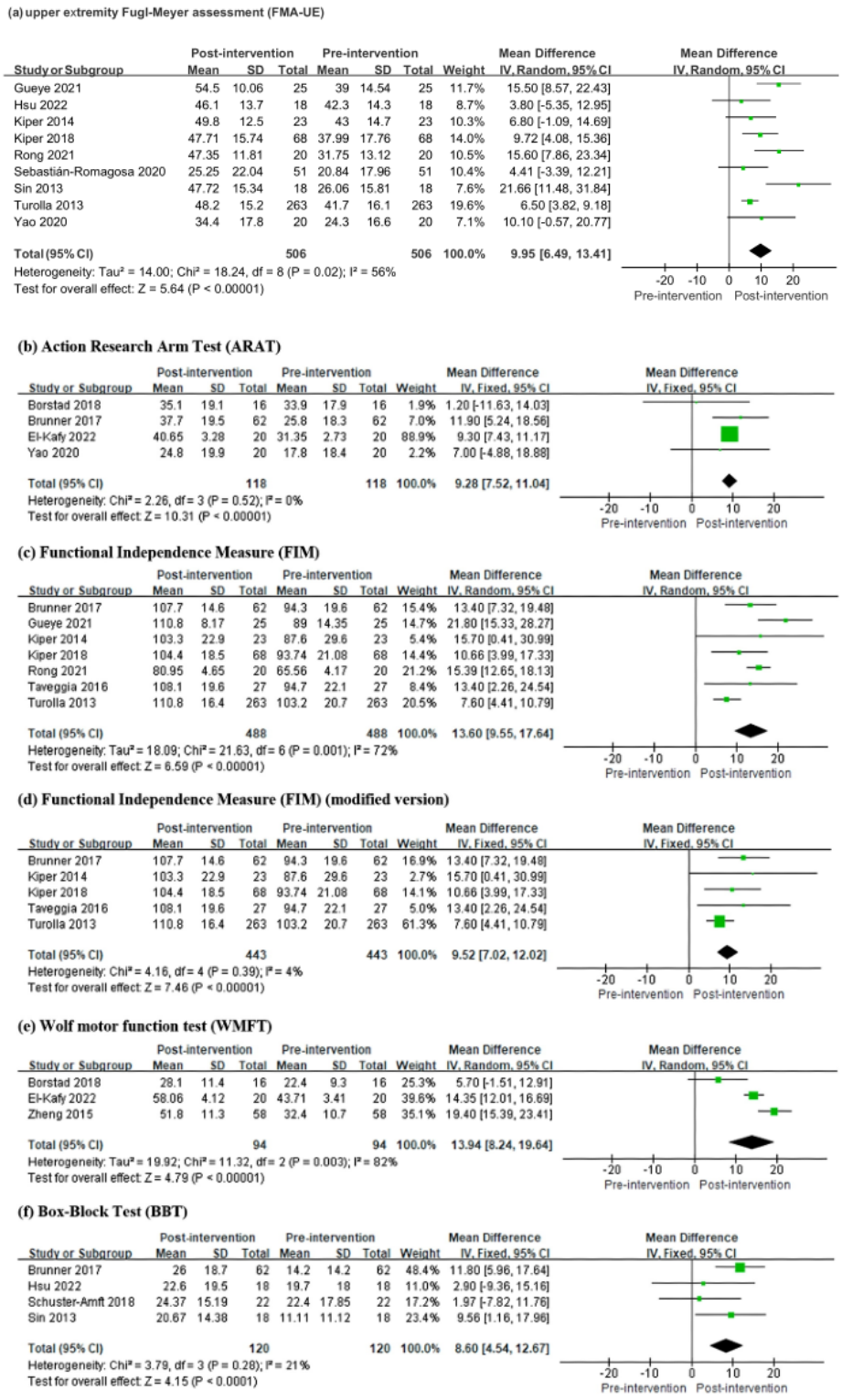

Whole Group: Upper Limb Activity Limitation (ARAT, FIM, WMFT, and BBT)

Participants in the VRT group showed significantly greater improvements than those in the CON group in upper limb function measured by the ARAT (MD 3.68, 95% CI 1.87-5.48, P<.001; I^2=0%; n [VRT]=102, n [CON]=98; Figure 2), independence in activities of daily living measured by the FIM (MD 4.86, 95% CI 2.92-6.81, P<.001; I^2=46%; n [VRT]=488, n [CON]=332; Figure 2), and upper limb functional capacity measured by the WMFT (MD 6.16, 95% CI 4.03-8.29, P<.001; I^2=0%; n [VRT]=78, n [CON]=74; Figure 2). However, there were no significant differences between the groups in upper limb finger dexterity measured by the BBT (MD 3.26, 95% CI –1.16 to 7.68, P=.15; I^2=0%; n [VRT]=120, n [CON]=124; Figure 2).

XSL•FO
RenderX

Among the VRT groups, participants demonstrated significant pre- and postintervention improvements in scores on the ARAT (MD 9.28, 95% CI 7.52-11.04, P<.001; I^2=0%), FIM (MD 13.60, 95% CI 9.55-17.64, P<.001; I^2=72%; or MD 9.52, 95% CI 7.02-12.02, P<.001; I^2=4% after the removal of outliers [21,25]), WMFT (MD 13.94, 95% CI 8.24-19.64, P<.001; I^2=82%), and BBT (MD 8.60, 95% CI 4.54-12.67, P<.001; I^2=21%; Figure 3).

Subgroup Analyses

Age Influencing the VR Effect

Compared with those in the CON group, younger patients (n=301, mean age 56.45 years) in the VRT group demonstrated significantly greater improvements in upper limb function measured by the FMA-UE (MD 5.34, 95% CI 2.18-8.5, P<.001; Table 4). No statistically significant heterogeneity was observed among older patients (SMD 1.96, 95% CI −1.50 to 5.41, P=.27; Table 4). Similarly, younger patients showed greater improvements in independence in activities of daily living measured by the FIM (MD 6.55, 95% CI 2.36-10.75, P=.002), while no statistically significant heterogeneity was observed among older patients (MD 2.68, 95% CI −0.47 to 5.83, P=.10; Table 5).

XSL•FO
RenderX

Table 4. Subgroup analyses on age, total hours, trial length, and duration based on FMA-UEa measures.

Outcomes, moderating factors, and between-group/within-group comparison	Number of trials analyzed and number of participants involved	Mean difference (95% CI)	*P* value	I^2 (%)
FMA-UE				
Age (years)				
Younger (≤60.52 years)				
Between-group	3; N (VRT[b])=301, N (CON[c])=150	5.34 (2.18 to 8.50)	.0009	0
Within-group	4; N=352	6.96 (4.64 to 9.29)	<.001	48
Older (> 60.52 years)				
Between-group	5; N (VRT)=154, N (CON)=151	1.96 (–1.50 to 5.41)	.27	0
Within-group	5; N=154	11.93 (8.54 to 15.33)	<.001	42
Total dosage (hours)				
< 15				
Between-group	2; N (VRT)=45, N (CON)=45	0.79 (–4.90 to 6.47)	.79	0
Within-group	2; N=45	13.90 (8.09 to 19.71)	<.001	0
≥15 and ≤30				
Between-group	3; N (VRT)=56, N (CON)=54	9.67 (4.19 to 15.15)	.0005	0
Within-group	4; N=107	11.14 (2.99 to 19.30)	<.001	72
> 30				
Between-group	3; N (VRT)=354, N (CON)=202	2.95 (0.06 to 5.84)	.05	0
Within-group	3; N=354	7.07 (4.76 to 9.38)	<.001	0
Trial length (weeks)				
< 4				
Between-group	2; N (VRT)=45, N (CON)=45	0.79 (–4.90 to 6.47)	.79	0
Within-group	2; N=45	13.90 (8.09 to 19.71)	<.001	0
≥4 and ≤6				
Between-group	5; N (VRT)=392, N (CON)=239	4.02 (1.39 to 6.65)	.003	15
Within-group	5; N=392	10.88 (6.07 to 15.68)	<.001	67
> 6				
Between-group	1; N (VRT)=18, N (CON)=17	11.0 (0.13 to 21.87)	.05	N/A[d]
Within-group	2; N=69	4.15 (–1.78 to 10.09)	.17	0
Duration of a session (hours)				
< 1				
Between-group	2; N (VRT)=45, N (CON)=45	0.79 (–4.90 to 6.47)	.79	0
Within-group	2; N=45	13.90 (8.09 to 19.71)	<.001	0
≥1 and<2				
Between-group	3; N (VRT)=56, N (CON)=54	9.67 (4.19 to 15.15)	.0005	0
Within-group	4; N=107	11.14 (2.99 to 19.30)	<.001	72
≥2				
Between-group	3; N (VRT)=354, N (CON)=202	2.95 (0.06 to 5.84)	.05	0
Within-group	3; N=354	7.07 (4.76 to 9.38)	<.001	0
Frequency (days/week)				
≤4				
Between-group	3; N (VRT)=61, N (CON)=59	7.01 (–1.67 to 15.68)	.11	58

XSL•FO
RenderX

Outcomes, moderating factors, and between-group/within-group comparison	Number of trials analyzed and number of participants involved	Mean difference (95% CI)	*P* value	I^2 (%)
Within-group	4; N=112	11.17 (3.14 to 19.20)	.006	73
> 4				
Between-group	5; N (VRT)=394, N (CON)=242	3.48 (0.87 to 6.09)	.009	0
Within-group	5; N=394	7.86 (5.69 to 10.03)	<.001	27

[a]FMA-UE: Upper Extremity Fugl-Meyer Assessment.

[b]VRT: virtual reality therapy.

[c]CON: conventional therapy.

[d]N/A: not applicable.

XSL•FO
RenderX

Table 5. Subgroup analyses on age, total hours, trial length, and duration based on the Functional Independence Measure.

Outcomes, moderating factors, and between-group/within-group comparison	Number of trials analyzed and number of participants involved	Mean difference (95% CI)	*P* value	I^2 (%)
Functional Independence Measure				
Age (years)				
Younger (≤60.52)				
Between-group	2; N (VRT[a])=283, N (CON[b])=133	6.55 (2.36 to 10.75)	.002	63
Within-group	2; N=283	11.54 (3.91 to 19.17)	.003	92
Older (> 60.52)				
Between-group	5; N (VRT)=205, N (CON)=199	2.68 (–0.47 to 5.83)	.10	26
Within-group	5; N=205	15.13 (11.72 to 18.55)	<.001	36
Total dosage (hours)				
<30				
Between-group	2; N (VRT)=87, N (CON)=83	2.07 (–4.65 to 8.79)	.55	60
Within-group	2; N=87	17.52 (9.29 to 25.75)	<.001	71
≥30 and <40				
Between-group	2; N (VRT)=47, N (CON)=47	5.06 (2.06 to 8.07)	.001	11
Within-group	2; N=47	15.28 (12.62 to 17.94)	<.001	0
≥40				
Between-group	3; N (VRT)=354, N (CON)=202	6.70 (3.46 to 9.94)	<.001	46
Within-group	3; N=354	8.43 (5.6 to 11.26)	<.001	0
Duration of a session (hours)				
< 1				
Between-group	1; N (VRT)=25, N (CON)=25	5.90 (–0.96 to 12.76)	.09	N/A[c]
Within-group	1; N=25	21.80 (15.33 to 28.27)	<.001	N/A
≥1 and <2				
Between-group	3; N (VRT)=109, N (CON)=105	3.54 (–1.60 to 8.68)	.18	61
Within-group	3; N=109	14.98 (12.54 to 17.41)	<.001	0
≥2				
Between-group	3; N (VRT)=354, N (CON)=202	6.70 (3.46 to 9.94)	<.001	46
Within-group	3; N=354	8.43 (5.60 to 11.26)	<.001	0
Frequency (days/week)				
≤4				
Between-group	2; N (VRT)=87, N (CON)=83	2.07 (–4.65 to 8.79)	.55	60
Within-group	2; N=87	17.52 (9.29 to 25.75)	<.001	71
>4				
Between-group	5; N (VRT)=401, N (CON)=249	5.82 (3.62 to 8.03)	<.001	26
Within-group	5; N=401	11.89 (7.31 to 16.47)	<.001	71

[a]VRT: virtual reality therapy.

[b]CON: conventional therapy.

[c]N/A: not applicable.

Dosage (Total Hours) Influencing the VR Effect

Comparisons were conducted between studies with total intervention delivery times of 15 hours or less and those exceeding 15 hours. Among studies with less than 15 hours of intervention, the VRT group did not show significant changes in upper limb function, as measured by the FMA-UE, compared with the CON group before and after the intervention (MD 0.79,

XSL•FO
RenderX

95% CI −4.90 to 6.47, *P*=.79; Table 4). By contrast, among studies with over 15 hours of intervention, the VRT group demonstrated significant improvements in upper limb function measured by the FMA-UE (MD 9.67, 95% CI 4.19-15.15, *P*<.001; Table 4).

Trial Length Influencing the VR Effect

When the trial duration was 4 weeks or longer, improvements in upper limb function measured by the FMA-UE in the VRT group were greater than those in the CON group (MD 4.02, 95% CI 1.39-6.65, *P*<.001; Table 4). For within-group comparisons, trials lasting over 4 weeks showed significant improvements in upper limb function in the VRT group measured by the ARAT (MD 9.49, 95% CI 7.69-11.29, *P*<.001; Table 6) and the BBT (MD 7.43, 95% CI 0.50-14.36, *P*=.04; Table 7). The VRT group also demonstrated improvements in upper limb function measured by the FMA-UE when the trial duration was less than 6 weeks (MD 10.88, 95% CI 6.07-15.68, *P*<.001; Table 4), but no significant improvements were observed when the trial duration was 6 weeks or longer (MD 4.15, 95% CI −1.78 to 10.09, *P*=.17; Table 4).

Table 6. Subgroup analyses on age, total hours, trial length, and duration based on Action Research Arm Test measures.

Outcomes and moderating factors	Between-group/within-group comparison	Number of trials analyzed and number of participants involved	Mean difference (95% CI)	*P* value	I^2 (%)
Action Research Arm Test					
Age (years)					
Younger (≤60.52)	Within-group	2; N=36	9.13 (7.28 to 10.98)	<.001	33
Older (>60.52)	Within-group	2; N=82	10.73 (4.92 to 16.54)	<.001	0
Total dosage (hours)					
<30	Within-group	2; N=82	10.73 (4.92 to 16.54)	<.001	0
≥30	Within-group	2; N=36	9.13 (7.28 to 10.98)	<.001	33
Trial length (weeks)					
<4	Within-group	2; N=36	4.32 (–4.39 to 13.04)	.33	0
≥4	Within-group	2; N=82	9.49 (7.69 to 11.29)	<.001	0
Duration of a session (hours)					
≤1	Within-group	2; N=82	10.73 (4.92 to 16.54)	<.001	0
>1	Within-group	2; N=36	9.13 (7.28 to 10.98)	<.001	33
Frequency (days/week)					
≤4	Within-group	2; N=82	9.49 (7.69 to 11.29)	<.001	0
>4	Within-group	2; N=36	4.32 (–4.39 to 13.04)	.33	0

XSL•FO
RenderX

Table 7. Subgroup analyses on age, total hours, trial length, and duration based on Box-Block Test measures.

Outcomes, moderating factors, and between-group/within-group comparison	Number of trials analyzed and number of participants involved	Mean difference (95% CI)	*P* value	I^2 (%)
Box-Block Test				
Age (years)				
Younger (≤60.52)				
Between-group	1; N (VRT[a])=18, N (CON[b])=17	7.70 (–5.71 to 21.11)	.26	N/A[c]
Within-group	1; N=18	2.90 (–9.36 to 15.16)	.64	N/A
Older (> 60.52)				
Between-group	3; N (VRT)=102, N (CON)=107	2.72 (–1.96 to 7.40)	.25	0
Within-group	3; N=102	9.31 (5.00 to 13.62)	<.001	30
Total dosage (hours)				
≤15				
Between-group	2; N (VRT)=40, N (CON)=49	5.42 (–2.49 to 13.32)	.18	0
Within-group	2; N=40	2.33 (–5.32 to 9.98)	.55	0
> 15				
Between-group	2; N (VRT)=80, N (CON)=75	2.28 (–3.05 to 7.62)	.4	0
Within-group	2; N=80	11.07 (6.27 to 15.87)	<.001	0
Trial length (weeks)				
≤4				
Between-group	2; N (VRT)=84, N (CON)=90	2.04 (–3.53 to 7.60)	.47	0
Within-group	2; N=84	7.70 (–1.80 to 17.20)	.11	65
> 4				
Between-group	2; N (VRT)=36, N (CON)=34	5.36 (–1.92 to 12.63)	.15	0
Within-group	2; N=36	7.43 (0.50 to 14.36)	.04	0
Duration of a session (hours)				
<1				
Between-group	2; N (VRT)=40, N (CON)=49	5.42 (–2.49 to 13.32)	.18	0
Within-group	2; N=40	2.33 (–5.32 to 9.98)	.55	0
≥1				
Between-group	2; N (VRT)=80, N (CON)=75	2.28 (–3.05 to 7.62)	.4	0
Within-group	2; N=80	11.07 (6.27 to 15.87)	<.001	0
Frequency (days/week)				
< 4				
Between-group	2; N (VRT)=36, N (CON)=34	5.36 (–1.92 to 12.63)	.15	0
Within-group	2; N=36	7.43 (0.50 to 14.36)	.04	0
≥4				
Between-group	2; N (VRT)=84, N (CON)=90	2.04 (–3.53 to 7.60)	.47	0
Within-group	2; N=84	7.70 (–1.80 to 17.20)	.11	65

[a]VRT: virtual reality therapy.

[b]CON: conventional therapy.

[c]N/A: not applicable.

XSL•FO
RenderX

Duration (Hour/Session) and Frequency (Session/Week) Influencing the VR Effect

For session durations shorter than 2 hours, participants in the VRT group (mean 1.17 hours) demonstrated significantly greater improvements in upper limb function, as measured by the FMA-UE, compared with those in the CON group (mean 1.17 hours) (MD 9.67, 95% CI 4.19-15.15, *P*<.001; Table 4). For session durations equal to or longer than 2 hours (ie, ≥2 hours/session), participants in the VRT group (mean 2.0 hours) also showed significantly greater improvements in independence in activities of daily living, as measured by the FIM, compared with those in the CON group (mean 2.0 hours; MD 6.70, 95% CI 3.46-9.94, *P*<.001; Table 5).

When the frequency exceeded 4 sessions per week, improvements in upper limb function measured by the FMA-UE (MD 3.48, 95% CI 0.87-6.09, *P*<.001) and independence in activities of daily living measured by the FIM in the VRT group were greater than those in the CON group (MD 5.82, 95% CI 3.62-8.03, *P*<.001) (Tables 4 and 5). Conversely, when the frequency was 4 or fewer sessions per week, significant changes in upper limb function measured by the ARAT were observed in the VRT group (MD 9.49, 95% CI 7.69-11.29, *P*<.001), but no significant changes were found when the frequency exceeded 4 sessions per week (MD 4.32, 95% CI –4.39 to 13.04, *P*=.33; Table 6).

Content Design Influencing the VR Effect

The content design factor was defined based on the National Clinical Guideline for Stroke for the United Kingdom and Ireland (2023 edition) [19]. The number of features in the VR interventions across all studies ranged from 2 to 10. Comparisons were made between studies with a fewer number of content features (n=7, features ≤5) and those with a greater number of features (n=8, features>5; Table 8). The common outcome variable among the included studies was upper limb function measured by the FMA-UE, and thus the comparisons were based on this measure. The MDs in FMA-UE scores did not significantly differ between the fewer and greater feature subgroups (MD 3.89, 95% CI –6.40 to 1.09; effect *Z*=1.36, *P*=.16; Figure 4; see also [11,42-46,48,50]).

XSL•FO
RenderX

Table 8. Content analysis of the design features of the VRT used in the studies.

Studies	Relative quantity of content features[a]	Content feature items									
		1[b]	2[c]	3[d]	4[e]	5[f]	6[g]	7[h]	8[i]	9[j]	10[k]
Brunner et al [40]	+	N/A[l]	✓[m]	✓	N/A	N/A	N/A	N/A	N/A	✓	N/A
Abd El-Kafy et al [41]	+++	✓	✓	✓	✓	✓	N/A	✓	N/A	N/A	N/A
Gueye et al [42]	+	N/A	✓	✓	N/A	N/A	N/A	N/A	✓	N/A	N/A
Hsu et al [11]	✓	N/A	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Kiper et al [43]	+++	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✓
Kiper et al [44]	+++	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✓
Rong et al [45]	++	✓	✓	✓	N/A	N/A	N/A	✓	N/A	N/A	N/A
Schuster-Amft et al [16]	+++	✓	✓	✓	✓	✓	N/A	✓	✓	✓	N/A
Sin and Lee [46]	++	N/A	✓	✓	✓	✓	N/A	✓	N/A	N/A	N/A
Taveggia et al [47]	+	N/A	N/A	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A
Yao et al [48]	+++	N/A	N/A	✓	✓	✓	✓	✓	N/A	N/A	✓
Zheng et al [49]	+++	✓	✓	✓	✓	✓	N/A	✓	✓	✓	N/A
Turolla et al [50]	+++	✓	✓	✓	N/A	✓	✓	✓	N/A	✓	N/A
Borstad et al [33]	+++	✓	✓	✓	✓	✓	✓	✓	✓	N/A	✓
Sebastián-Romagosa et al [34]	+	N/A	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[a]The '+', '++', and '+++' keys refer to a few (2-3), medium (4-5), and many (≥6) numbers of features, respectively.

[b]Item 1: Offering exercise and functional tasks.

[c]Item 2: Having individualized goals.

[d]Item 3: Quantifying activity.

[e]Item 4: Considering comorbidities and baseline activity level.

[f]Item 5: Incorporating patient's needs.

[g]Item 6: Matching patient's background such as education level.

[h]Item 7: Patient's goal and interest directed.

[i]Item 8: Involving goal setting.

[j]Item 9: Progressive difficulty level.

[k]Item 10: Promoting self-efficacy.

[l]N/A: not applicable.

[m]The symbol '✓' indicates that descriptions of the virtual reality intervention are found to contain the corresponding feature item.

Figure 4. Forest plot of comparisons of VRT content design between the fewer and a greater number of feature subgroups. VRT: virtual reality therapy.

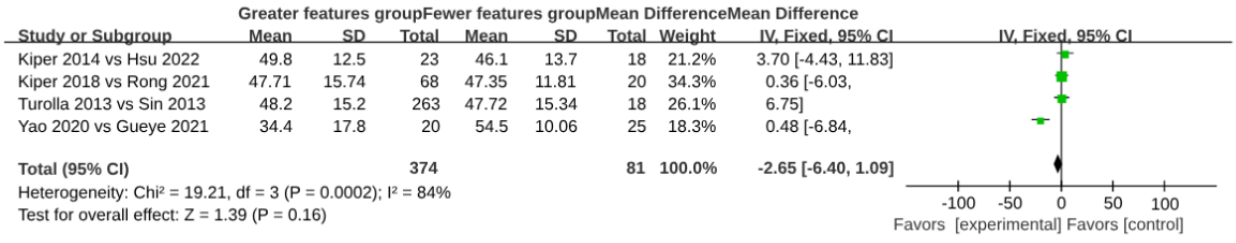

Discussion

Principal Findings

This study aimed to examine the influence of patients' age, delivery dosage and schedule, and content design on the outcomes of virtual reality interventions for patients with stroke. The meta-analytic results revealed that virtual reality training had significantly greater therapeutic effects than conventional training in improving upper limb function in patients with stroke. These effects included enhancements in upper limb function, as measured by the FMA-UE and ARAT, and improvements in activities of daily living, as measured by the FIM.

The demographic factor associated with more significant therapeutic effects was younger patient age compared with older age (mean 56.5 vs 65.9 years). Virtual reality training content

XSL•FO
RenderX

factors contributing to greater effects included higher training dosage (>15 hours) delivered over 4-6 weeks and shorter sessions (mean 1.2 vs 2.0 hours) scheduled 4 or more times per week. Notably, shorter sessions (mean 1.2 hours) were linked to better outcomes in upper limb function, whereas longer sessions (mean 2.0 hours) were associated with improved outcomes in activities of daily living.

Our findings on the therapeutic effects of virtual reality training are largely consistent with those reported in 4 previous meta-analytic studies [9,36,54,55]. The consistent results confirm that virtual reality, as a novel and potentially useful technology, is effective in improving upper limb function and activities of daily living compared with conventional therapy after stroke. However, an inconsistent result was observed in the nonsignificant effects of virtual reality training on improving patients' upper limb dexterity (measured by the BBT). This nonsignificant finding could be attributed to the smaller number of studies that used finger dexterity as an outcome variable (n=4) compared with those focusing on upper limb function or activities of daily living (n=8).

Younger patients with stroke demonstrated better outcomes than older patients in both upper limb function and activities of daily living. One plausible explanation is that younger patients may have been more motivated and actively engaged in the training, leading to better results. Previous studies have shown that increasing age is associated with lower participation in cognitively demanding activities [56]. Reduced participation inevitably results in lower gains from training. Spiteri et al [57] found that older adults (eg, 65-70 years) faced barriers such as mobility challenges, limited community accessibility, and insufficient guidance from health care professionals, which hindered their participation in physical activity (also see [58]). Our findings further emphasize the value of virtual reality training in poststroke rehabilitation. Addressing potential barriers, such as ensuring adequate supervision from health care professionals and improving accessibility, could enhance the effectiveness of VR-based training for older patients.

VR-based training offers immersive, imaginative, and interactive experiences [59,60] that facilitate functional recovery in patients with stroke [61,62]. These vivid experiences have been linked to participants engaging in more intensive cognitive and motor processes during training [63-66]. The cognitive processes involve bottom-up attention, which entails the encoding and processing of visual and auditory stimuli emitted in the virtual environment [67,68]. Top-down attention, by contrast, involves cognitive control, decision-making, and motor planning to produce motor responses [69,70]. The content design of virtual reality training did not significantly influence its therapeutic effects. In this study, we adhered to the principles outlined by the United Kingdom and Ireland for designing therapy aimed at promoting functional recovery in patients with stroke. The number of different features, whether fewer or more, based on these principles, did not emerge as a significant factor. However, 2 content features shared by most studies were setting individualized goals (item 2) and quantifying activities (item 3). Our findings suggest that counting repetitions of movements or responses performed by patients and setting treatment goals tailored to individual needs may be essential or sufficient features for designing effective virtual reality interventions.

The cognitive and motor processes involved in virtual reality training need to be repetitive to drive functional changes in patients with stroke. Our results suggest that more than 15 hours of training, scheduled over 4-6 weeks with 4 or more 1-hour sessions per week, can maximize improvements in upper limb function. By contrast, achieving significant gains in independence in activities of daily living appears to require a higher dosage, with 30 or more hours of training needed to produce meaningful outcomes. Repetitiveness is essential for initiating neural changes after a stroke. At the neurological level, posttraining changes in the motor cortices involve alterations in functional connectivity within neural networks, such as the prefrontal cortex and the basal ganglia [71,72]. Recovery of upper extremity function after a stroke has been linked to changes in functional connectivity between the bilateral primary motor cortex (M1) and the dorsolateral prefrontal cortex [73,74]. Additionally, repetitive training leads to the functional reorganization of the sensorimotor networks and the extrapyramidal system [75].

Limitation

This study has several limitations. First, the subgroup analyses did not include outcomes measured by the ARAT or the WMFT, limiting the generalizability of the results. Caution is therefore needed when interpreting the findings. Second, the therapeutic effects of VR-based training were derived from outcomes measured shortly after the completion of training. As such, the findings do not extend to short- or long-term posttraining effects. Evidence regarding the long-term benefits of VR-enhanced exercise training for improving upper limb function in patients with stroke remains limited. However, a study examining the long-term treatment effects of VR in cardiovascular rehabilitation found that combining VR-based therapy with conventional treatment was associated with sustained long-term effects on hemodynamic and autonomic outcomes. This suggests that VR could serve as a valuable new treatment modality when integrated with cardiovascular rehabilitation sessions [51-53,76]. Standalone VR treatment lacks sensory and proprioceptive inputs to the central nervous system, limiting the functionality of the central-peripheral loop. Future longitudinal studies should explore the combination of VR-based therapy with actual exercises, focusing on how to integrate these approaches effectively for patients with stroke.

Third, significant heterogeneity (I^2>50%) was observed in the meta-analysis. In the whole-group analysis, participants in the VRT group showed significantly greater improvement than those in the CON group, with I^2<50%, indicating that the results from the studies were consistent in favoring VRT. However, in the within-group pre- and postintervention analysis, we found that some studies had I^2>50%, such as the WMFT (I^2=82%). This could be attributed to clinical variability in intervention delivery, including factors such as dosage, trial length, and frequency. The heterogeneity of participants across the included trials likely reduced the effect sizes of the observed therapeutic effects, particularly for older patients and improvements in finger dexterity. Similarly, the relatively small number of studies

XSL•FO
RenderX

in the subgroup analyses may have reduced the power to identify training design content as a significant outcome factor. Future research should include more studies and patients with stroke in the analyses and investigate the long-term effects of virtual reality training.

Conclusions

Virtual reality training was effective in promoting upper limb function and independence in activities of daily living for patients with stroke. Patient age and the dosage and schedule of training delivery are factors that influence its therapeutic effects. Younger patients who participate in training for more than 15 hours tend to show better outcomes compared with older patients. Training content delivered in 1-hour sessions, with more than 4 sessions per week over 4-6 weeks, could further maximize the therapeutic effects. Additionally, the design of the VR content should be tailored to the individual needs of the patients and involve them in setting treatment goals.

Acknowledgments

This work was supported by the Wuxi Taihu Lake Talent Development Program in Medical and Health, funded by the Wuxi Municipal Office of Talent Affairs (grant THRC23074004); the Healthcare Commission Foundation of Wuxi City (grant M202123); the Wuxi Ninth People's Hospital Affiliated with Soochow University projects (grants QD202102 and YB202103); and the Wuxi Philosophy and Social Sciences Project (grant WXSK21-JY-B04).

No generative artificial intelligence was used in any portion of the manuscript writing.

Data Availability

All data generated or analyzed during this study are included in this published article. Further inquiries can be directed to the corresponding author.

Authors' Contributions

SX and YX contributed equally and are joint first authors. Both YX and CCHC are corresponding authors. YX's email address is xudeng2@163.com.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 149 KB - jmir_v27i1e66802_app1.pdf]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 95 KB - jmir_v27i1e66802_app2.pdf]

References

1. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. Bull World Health Organ 2016 Sep 01;94(9):634-634A. [doi: 10.2471/blt.16.181636]
2. GBD 2016 Lifetime Risk of Stroke Collaborators T. Global, regional, and country-specific lifetime risks of stroke, 1990 and 2016. N Engl J Med 2018 Dec 20;379(25):2429-2437. [doi: 10.1056/nejmoa1804492]
3. Bernhardt J, Urimubenshi G, Gandhi DBC, Eng JJ. Stroke rehabilitation in low-income and middle-income countries: a call to action. The Lancet 2020 Oct;396(10260):1452-1462. [doi: 10.1016/s0140-6736(20)31313-1]
4. Langhammer B, Stanghelle JK. Bobath or motor relearning programme? A comparison of two different approaches of physiotherapy in stroke rehabilitation: a randomized controlled study. Clin Rehabil 2000 Aug 01;14(4):361-369. [doi: 10.1191/0269215500cr338oa] [Medline: 10945420]
5. Logigian M, Samuels M, Falconer J, Zagar R. Clinical exercise trial for stroke patients. Arch Phys Med Rehabil 1983 Aug;64(8):364-367. [Medline: 6882175]
6. van Vliet PM, Lincoln NB, Foxall A. Comparison of Bobath based and movement science based treatment for stroke: a randomised controlled trial. J Neurol Neurosurg Psychiatry 2005 Apr;76(4):503-508 [FREE Full text] [doi: 10.1136/jnnp.2004.040436] [Medline: 15774435]
7. Dimyan MA, Cohen LG. Neuroplasticity in the context of motor rehabilitation after stroke. Nat Rev Neurol 2011 Feb 18;7(2):76-85 [FREE Full text] [doi: 10.1038/nrneurol.2010.200] [Medline: 21243015]
8. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb 10;15(2):107-117. [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]

XSL•FO
RenderX

9. Saposnik G, Levin M. Virtual reality in stroke rehabilitation. Stroke 2011 May;42(5):1380-1386. [doi: 10.1161/strokeaha.110.605451]
10. Alves SS, Ocamoto GN, de Camargo PS, Santos ATS, Terra AMSV. Effects of virtual reality and motor imagery techniques using Fugl Meyer Assessment scale in post-stroke patients. International Journal of Therapy and Rehabilitation 2018 Nov 02;25(11):587-596. [doi: 10.12968/ijtr.2018.25.11.587]
11. Hsu H, Kuo L, Lin Y, Su F, Yang T, Lin C. Effects of a virtual reality-based mirror therapy program on improving sensorimotor function of hands in chronic stroke patients: a randomized controlled trial. Neurorehabil Neural Repair 2022 Jun 28;36(6):335-345. [doi: 10.1177/15459683221081430] [Medline: 35341360]
12. Ballester BR, Nirme J, Camacho I, Duarte E, Rodríguez S, Cuxart A, et al. Domiciliary VR-based therapy for functional recovery and cortical reorganization: randomized controlled trial in participants at the chronic stage post stroke. JMIR Serious Games 2017 Aug 07;5(3):e15 [FREE Full text] [doi: 10.2196/games.6773] [Medline: 28784593]
13. Buccellato K, Nordstrom M, Murphy J, Burdea G, Polistico K, House G, et al. A randomized feasibility trial of a novel, integrative, and intensive virtual rehabilitation program for service members post-acquired brain injury. Mil Med 2020 Feb 12;185(1-2):e203-e211. [doi: 10.1093/milmed/usz150] [Medline: 31268524]
14. Ikbali Afsar S, Mirzayev I, Umit Yemisci O, Cosar Saracgil SN. Virtual reality in upper extremity rehabilitation of stroke patients: a randomized controlled trial. J Stroke Cerebrovasc Dis 2018 Dec;27(12):3473-3478. [doi: 10.1016/j.jstrokecerebrovasdis.2018.08.007] [Medline: 30193810]
15. McEwen D, Taillon-Hobson A, Bilodeau M, Sveistrup H, Finestone H. Virtual reality exercise improves mobility after stroke. Stroke 2014 Jun;45(6):1853-1855. [doi: 10.1161/strokeaha.114.005362]
16. Schuster-Amft C, Eng K, Suica Z, Thaler I, Signer S, Lehmann I, et al. Effect of a four-week virtual reality-based training versus conventional therapy on upper limb motor function after stroke: a multicenter parallel group randomized trial. PLoS One 2018 Oct 24;13(10):e0204455 [FREE Full text] [doi: 10.1371/journal.pone.0204455] [Medline: 30356229]
17. Sánchez-Herrera-Baeza P, Cano-de-la-Cuerda R, Oña-Simbaña ED, Palacios-Ceña D, Pérez-Corrales J, Cuenca-Zaldivar JN, et al. The impact of a novel immersive virtual reality technology associated with serious games in Parkinson's disease patients on upper limb rehabilitation: a mixed methods intervention study. Sensors 2020 Apr 11;20(8):2168 [FREE Full text] [doi: 10.3390/s20082168] [Medline: 32290517]
18. Pazzaglia C, Imbimbo I, Tranchita E, Minganti C, Ricciardi D, Lo Monaco R, et al. Comparison of virtual reality rehabilitation and conventional rehabilitation in Parkinson's disease: a randomised controlled trial. Physiotherapy 2020 Mar;106:36-42. [doi: 10.1016/j.physio.2019.12.007] [Medline: 32026844]
19. Intercollegiate Stroke Working Party. National Clinical Guideline for Stroke for the UK and Ireland. Intercollegiate Stroke Working Party. 2023 May 4. URL: https://www.strokeguideline.org [accessed 2025-01-28]
20. Page SJ, Schmid A, Harris JE. Optimizing terminology for stroke motor rehabilitation: recommendations from the American Congress of Rehabilitation Medicine Stroke Movement Interventions Subcommittee. Arch Phys Med Rehabil 2012 Aug 21;93(8):1395-1399 [FREE Full text] [doi: 10.1016/j.apmr.2012.03.005] [Medline: 22446292]
21. Nudo RJ, Wise BM, SiFuentes F, Milliken GW. Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. Science 1996 Jun 21;272(5269):1791-1794. [doi: 10.1126/science.272.5269.1791] [Medline: 8650578]
22. Krakauer JW, Carmichael ST, Corbett D, Wittenberg GF. Getting neurorehabilitation right: what can be learned from animal models? Neurorehabil Neural Repair 2012 Oct 30;26(8):923-931 [FREE Full text] [doi: 10.1177/1545968312440745] [Medline: 22466792]
23. Kleim JA, Barbay S, Cooper NR, Hogg TM, Reidel CN, Remple MS, et al. Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. Neurobiol Learn Mem 2002 Jan;77(1):63-77. [doi: 10.1006/nlme.2000.4004] [Medline: 11749086]
24. Luke LM, Allred RP, Jones TA. Unilateral ischemic sensorimotor cortical damage induces contralesional synaptogenesis and enhances skilled reaching with the ipsilateral forelimb in adult male rats. Synapse 2004 Dec 15;54(4):187-199. [doi: 10.1002/syn.20080] [Medline: 15472929]
25. Birkenmeier RL, Prager EM, Lang CE. Translating animal doses of task-specific training to people with chronic stroke in 1-hour therapy sessions: a proof-of-concept study. Neurorehabil Neural Repair 2010 Sep 27;24(7):620-635 [FREE Full text] [doi: 10.1177/1545968310361957] [Medline: 20424192]
26. Carey JR, Durfee WK, Bhatt E, Nagpal A, Weinstein SA, Anderson KM, et al. Comparison of finger tracking versus simple movement training via telerehabilitation to alter hand function and cortical reorganization after stroke. Neurorehabil Neural Repair 2007;21(3):216-232. [doi: 10.1177/1545968306292381] [Medline: 17351083]
27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71-112 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
28. Maher C, Sherrington C, Herbert R, Moseley A, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83(8):713-721. [doi: 10.1093/ptj/83.8.713]
29. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003 Sep 08;73(9):712-716. [doi: 10.1046/j.1445-2197.2003.02748.x] [Medline: 12956787]

XSL•FO
RenderX

30. Cashin AG, McAuley JH. Clinimetrics: Physiotherapy Evidence Database (PEDro) Scale. J Physiother 2020 Jan;66(1):59 [FREE Full text] [doi: 10.1016/j.jphys.2019.08.005] [Medline: 31521549]
31. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res 2016 Sep 27;27(6):1785-1805. [doi: 10.1177/0962280216669183] [Medline: 27683581]
32. Higgins J, Li T, Deeks J. Cochrane Handbook for Systematic Reviews of Interventions. London, UK: Cochrane; Sep 20, 2019.
33. Borstad AL, Crawfis R, Phillips K, Lowes LP, Maung D, McPherson R, et al. In-home delivery of constraint-induced movement therapy via virtual reality gaming. J Patient Cent Res Rev 2018 Jan 30;5(1):6-17. [doi: 10.17294/2330-0698.1550] [Medline: 31413992]
34. Sebastián-Romagosa M, Cho W, Ortner R, Murovec N, Von Oertzen T, Kamada K, et al. Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients-a feasibility study. Front Neurosci 2020 Oct 21;14:591435 [FREE Full text] [doi: 10.3389/fnins.2020.591435] [Medline: 33192277]
35. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014 Dec 19;14(1):135 [FREE Full text] [doi: 10.1186/1471-2288-14-135] [Medline: 25524443]
36. Laver KE, Lange B, George S, Deutsch J, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11(11):CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
37. Mekbib DB, Zhao Z, Wang J, Xu B, Zhang L, Cheng R, et al. Proactive motor functional recovery following immersive virtual reality-based limb mirroring therapy in patients with subacute stroke. Neurotherapeutics 2020 Oct;17(4):1919-1930. [doi: 10.1007/s13311-020-00882-x] [Medline: 32671578]
38. Zhang J, Fong K, Welage N, Liu K. The activation of the mirror neuron system during action observation and action execution with mirror visual feedback in stroke: a systematic review. Neural Plast 2018 Apr 24;2018:2321045 [FREE Full text] [doi: 10.1155/2018/2321045] [Medline: 29853839]
39. Chen J, Or CK, Chen T. Effectiveness of using virtual reality–supported exercise therapy for upper extremity motor rehabilitation in patients with stroke: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jun 20;24(6):e24111 [FREE Full text] [doi: 10.2196/24111] [Medline: 35723907]
40. Brunner I, Skouen JS, Hofstad H, Aßmus J, Becker F, Sanders A, et al. Virtual reality training for upper extremity in subacute stroke (VIRTUES). Neurology 2017 Dec 12;89(24):2413-2421. [doi: 10.1212/wnl.0000000000004744]
41. Abd El-Kafy EM, Alshehri MA, El-Fiky AA, Guermazi MA, Mahmoud HM. The effect of robot-mediated virtual reality gaming on upper limb spasticity poststroke: a randomized-controlled trial. Games Health J 2022 Apr 01;11(2):93-103. [doi: 10.1089/g4h.2021.0197] [Medline: 35100025]
42. Gueye T, Dedkova M, Rogalewicz V, Grunerova-Lippertova M, Angerova Y. Early post-stroke rehabilitation for upper limb motor function using virtual reality and exoskeleton: equally efficient in older patients. Neurol Neurochir Pol 2021 Feb 28;55(1):91-96. [doi: 10.5603/pjnns.a2020.0096]
43. Kiper P, Agostini M, Luque-Moreno C, Tonin P, Turolla A. Reinforced feedback in virtual environment for rehabilitation of upper extremity dysfunction after stroke: preliminary data from a randomized controlled trial. Biomed Res Int 2014 Mar 13;2014:752128 [FREE Full text] [doi: 10.1155/2014/752128] [Medline: 24745024]
44. Kiper P, Szczudlik A, Agostini M, Opara J, Nowobilski R, Ventura L, et al. Virtual reality for upper limb rehabilitation in subacute and chronic stroke: a randomized controlled trial. Arch Phys Med Rehabil 2018 May;99(5):834-842.e4. [doi: 10.1016/j.apmr.2018.01.023] [Medline: 29453980]
45. Rong J, Ding L, Xiong L, Zhang W, Wang W, Deng M, et al. Mirror visual feedback prior to robot-assisted training facilitates rehabilitation after stroke: a randomized controlled study. Front Neurol 2021 Jul 8;12:683703 [FREE Full text] [doi: 10.3389/fneur.2021.683703] [Medline: 34305792]
46. 46Sin H, Lee G. Additional virtual reality training using Xbox Kinect in stroke survivors with hemiplegia. American Journal of Physical Medicine and Rehabilitation 2013;92(10):871-880. [doi: 10.1097/phm.0b013e3182a38e40]
47. Taveggia G, Borboni A, Salvi L, Mulé C, Fogliaresi S, Villafañe J. Efficacy of robot-assisted rehabilitation for the functional recovery of the upper limb in post-stroke patients: a randomized controlled study. European Journal of Physical and Rehabilitation Medicine 2016;52(6):767-773. [doi: 10.1097/mrr.0000000000000137]
48. Yao X, Cui L, Wang J, Feng W, Bao Y, Xie Q. Effects of transcranial direct current stimulation with virtual reality on upper limb function in patients with ischemic stroke: a randomized controlled trial. J Neuroeng Rehabil 2020 Jun 15;17(1):73. [doi: 10.1186/s12984-020-00699-x] [Medline: 32539812]
49. Zheng C, Liao W, Xia W. Effect of combined low-frequency repetitive transcranial magnetic stimulation and virtual reality training on upper limb function in subacute stroke: a double-blind randomized controlled trail. J Huazhong Univ Sci Technolog Med Sci 2015 Apr 16;35(2):248-254. [doi: 10.1007/s11596-015-1419-0] [Medline: 25877360]
50. Turolla A, Dam M, Ventura L, Tonin P, Agostini M, Zucconi C, et al. Virtual reality for the rehabilitation of the upper limb motor function after stroke: a prospective controlled trial. J Neuroeng Rehabil 2013 Aug 01;10(1):85 [FREE Full text] [doi: 10.1186/1743-0003-10-85] [Medline: 23914733]

XSL•FO
RenderX

51. Ersoy C, Iyigun G. Boxing training in patients with stroke causes improvement of upper extremity, balance, and cognitive functions but should it be applied as virtual or real? Topics in Stroke Rehabilitation 2020 Jun 23;28(2):112-126. [doi: 10.1080/10749357.2020.1783918]
52. Junior VADS, Santos MDS, Ribeiro NMDS, Maldonado IL. Combining proprioceptive neuromuscular facilitation and virtual reality for improving sensorimotor function in stroke survivors: a randomized clinical trial. J Cent Nerv Syst Dis 2019 Jul 25;11:1179573519863826. [doi: 10.1177/1179573519863826] [Medline: 31384139]
53. Kizony R, Weiss PL, Harel S, Feldman Y, Obuhov A, Zeilig G, et al. Tele-rehabilitation service delivery journey from prototype to robust in-home use. Disabil Rehabil 2017 Jul 22;39(15):1532-1540. [doi: 10.1080/09638288.2016.1250827] [Medline: 28004980]
54. Lohse KR, Hilderman CGE, Cheung KL, Tatla S, Van der Loos HFM. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014 Mar 28;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
55. Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. J NeuroEngineering Rehabil 2018 Mar 27;15(1):29. [doi: 10.1186/s12984-018-0370-2]
56. Hess TM. Selective engagement of cognitive resources: motivational influences on older adults' cognitive functioning. Perspect Psychol Sci 2014 Jul 16;9(4):388-407 [FREE Full text] [doi: 10.1177/1745691614527465] [Medline: 26173272]
57. Spiteri K, Broom D, Bekhet A, de Caro JX, Laventure B, Grafton K. Barriers and motivators of physical activity participation in middle-aged and older-adults - a systematic review. J Aging Phys Act 2019 Sep 01;27(4):929-944. [doi: 10.1123/japa.2018-0343] [Medline: 31141447]
58. Gingrich N, Bosancich J, Schmidt J, Sakakibara BM. Capability, opportunity, motivation, and social participation after stroke. Top Stroke Rehabil 2023 Jul 05;30(5):423-435. [doi: 10.1080/10749357.2022.2070358] [Medline: 35510695]
59. Subramanian SK, Prasanna SS. Virtual reality and noninvasive brain stimulation in stroke: how effective is their combination for upper limb motor improvement?-a meta-analysis. PM R 2018 Nov;10(11):1261-1270. [doi: 10.1016/j.pmrj.2018.10.001] [Medline: 30503233]
60. Kim W, Cho S, Park S, Lee J, Kwon S, Paik N. A low cost kinect-based virtual rehabilitation system for inpatient rehabilitation of the upper limb in patients with subacute stroke: A randomized, double-blind, sham-controlled pilot trial. Medicine 2018 Jun;97(25):e11173. [doi: 10.1097/md.0000000000011173]
61. Pin-Barre C, Laurin J. Physical exercise as a diagnostic, rehabilitation, and preventive tool: influence on neuroplasticity and motor recovery after stroke. Neural Plast 2015 Nov 22;2015:608581. [doi: 10.1155/2015/608581] [Medline: 26682073]
62. Ribeiro de Souza F, Sales M, Rabelo Laporte L, Melo A, Manoel da Silva Ribeiro N. Body structure/function impairments and activity limitations of post-stroke that predict social participation: a systematic review. Top Stroke Rehabil 2023 Sep 05;30(6):589-602. [doi: 10.1080/10749357.2022.2095086] [Medline: 35787246]
63. Chen L, Chen Y, Fu W, Huang D, Lo W. The effect of virtual reality on motor anticipation and hand function in patients with subacute stroke: a randomized trial on movement-related potential. Neural Plast 2022 Jan 24;2022:7399995. [doi: 10.1155/2022/7399995] [Medline: 35111219]
64. Perez-Marcos D, Bieler-Aeschlimann M, Serino A. Virtual reality as a vehicle to empower motor-cognitive neurorehabilitation. Front Psychol 2018 Nov 2;9:2120 [FREE Full text] [doi: 10.3389/fpsyg.2018.02120] [Medline: 30450069]
65. Cameirão MS, Faria AL, Paulino T, Alves J, Bermúdez i Badia S. The impact of positive, negative and neutral stimuli in a virtual reality cognitive-motor rehabilitation task: a pilot study with stroke patients. J NeuroEngineering Rehabil 2016 Aug 9;13(1):70. [doi: 10.1186/s12984-016-0175-0]
66. De Luca R, Lo Buono V, Leo A, Russo M, Aragona B, Leonardi S, et al. Use of virtual reality in improving poststroke neglect: promising neuropsychological and neurophysiological findings from a case study. Appl Neuropsychol Adult 2019 Sep 22;26(1):96-100. [doi: 10.1080/23279095.2017.1363040] [Medline: 28937807]
67. Johansson B. Current trends in stroke rehabilitation. A review with focus on brain plasticity. Acta Neurol Scand 2011 Mar;123(3):147-159. [doi: 10.1111/j.1600-0404.2010.01417.x] [Medline: 20726844]
68. Ekman U, Fordell H, Eriksson J, Lenfeldt N, Wåhlin A, Eklund A, et al. Increase of frontal neuronal activity in chronic neglect after training in virtual reality. Acta Neurol Scand 2018 May 16;138(4):284-292. [doi: 10.1111/ane.12955]
69. Prochnow D, Bermúdez i Badia S, Schmidt J, Duff A, Brunheim S, Kleiser R, et al. A functional magnetic resonance imaging study of visuomotor processing in a virtual reality-based paradigm: rehabilitation gaming system. Eur J Neurosci 2013 May 17;37(9):1441-1447. [doi: 10.1111/ejn.12157] [Medline: 23414211]
70. McDermott EJ, Metsomaa J, Belardinelli P, Grosse-Wentrup M, Ziemann U, Zrenner C. Predicting motor behavior: an efficient EEG signal processing pipeline to detect brain states with potential therapeutic relevance for VR-based neurorehabilitation. Virtual Real 2023 Sep 23;27(1):347-369. [doi: 10.1007/s10055-021-00538-x] [Medline: 36915631]
71. García-Murillo DG, Alvarez-Meza A, Castellanos-Dominguez G. Single-trial kernel-based functional connectivity for enhanced feature extraction in motor-related tasks. Sensors (Basel) 2021 Apr 13;21(8):2750 [FREE Full text] [doi: 10.3390/s21082750] [Medline: 33924672]

XSL•FO
RenderX

72. Ferrazzoli D, Ortelli P, Volpe D, Cucca A, Versace V, Nardone R, et al. The ties that bind: aberrant plasticity and networks dysfunction in movement disorders-implications for rehabilitation. Brain Connect 2021 May 01;11(4):278-296. [doi: 10.1089/brain.2020.0971] [Medline: 33403893]
73. Sui Y, Kan C, Zhu S, Zhang T, Wang J, Xu S, et al. Resting-state functional connectivity for determining outcomes in upper extremity function after stroke: a functional near-infrared spectroscopy study. Front Neurol 2022 Nov 9;13:965856 [FREE Full text] [doi: 10.3389/fneur.2022.965856] [Medline: 36438935]
74. Harirchian MH, Rezvanizadeh A, Fakhri M, Oghabian MA, Ghoreishi A, Zarei M, et al. Non-invasive brain mapping of motor-related areas of four limbs in patients with clinically isolated syndrome compared to healthy normal controls. Journal of Clinical Neuroscience 2010 Jun;17(6):736-741. [doi: 10.1016/j.jocn.2009.10.010]
75. Wu Q, Yue Z, Ge Y, Ma D, Yin H, Zhao H, et al. Brain functional networks study of subacute stroke patients with upper limb dysfunction after comprehensive rehabilitation including BCI training. Front Neurol 2019 Jan 27;10:1419 [FREE Full text] [doi: 10.3389/fneur.2019.01419] [Medline: 32082238]
76. Alves da Cruz M, Laurino M, Christofaro D, Ghisi G, Vanderlei L. Long-term effects of virtual reality-based therapy in cardiovascular rehabilitation: a longitudinal study. Physiother Theory Pract 2024 Apr 25;40(4):727-735. [doi: 10.1080/09593985.2022.2160222] [Medline: 36567613]

Abbreviations

ARAT: Action Research Arm Test
BBT: Box and Block Test
CON: conventional therapy
FIM: Functional Independence Measure
FMA-UE: Upper Extremity Fugl-Meyer Assessment
MD: mean difference
MINORS: Methodological Index for Non-Randomized Studies
PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
VR: virtual reality
VRT: virtual reality therapy
WMFT: Wolf Motor Function Test

Edited by A Mavragani; submitted 23.09.24; peer-reviewed by Z Ren, FZ da Silva Areas; comments to author 04.12.24; revised version received 31.12.24; accepted 06.01.25; published 19.02.25.

Please cite as:
Xu S, Xu Y, Wen R, Wang J, Qiu Y, Chan CCH
Virtual Reality Enhanced Exercise Training in Upper Limb Function of Patients With Stroke: Meta-Analytic Study
J Med Internet Res 2025;27:e66802
URL: https://www.jmir.org/2025/1/e66802
doi:10.2196/66802
PMID:39969977

©Shiqi Xu, Yanwen Xu, Ruyi Wen, Jun Wang, Yuyu Qiu, Chetwyn CH Chan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Advancing Health Care With Digital Twins: Meta-Review of Applications and Implementation Challenges

Mickaël Ringeval[1], PhD; Faustin Armel Etindele Sosso[2], DBA; Martin Cousineau[1], PhD; Guy Paré[1], PhD

[1]HEC Montréal, Montréal, QC, Canada

[2]Centre Intégré Universitaire de Santé et de Services Sociaux du Nord-de-l'Île-de-Montréal, Montréal, QC, Canada

Corresponding Author:
Mickaël Ringeval, PhD
HEC Montréal
3000 Chemin de la Côte-Sainte-Catherine
Montréal, QC, H3T 2A7
Canada
Phone: 1 5143406000
Email: mickael.ringeval@hec.ca

Abstract

Background: Digital twins (DTs) are digital representations of real-world systems, enabling advanced simulations, predictive modeling, and real-time optimization in various fields, including health care. Despite growing interest, the integration of DTs in health care faces challenges such as fragmented applications, ethical concerns, and barriers to adoption.

Objective: This study systematically reviews the existing literature on DT applications in health care with three objectives: (1) to map primary applications, (2) to identify key challenges and limitations, and (3) to highlight gaps that can guide future research.

Methods: A meta-review was conducted in a systematic fashion, adhering to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, and included 25 literature reviews published between 2021 and 2024. The search encompassed 5 databases: PubMed, CINAHL, Web of Science, Embase, and PsycINFO. Thematic synthesis was used to categorize DT applications, stakeholders, and barriers to adoption.

Results: A total of 3 primary DT applications in health care were identified: personalized medicine, operational efficiency, and medical research. While current applications, such as predictive diagnostics, patient-specific treatment simulations, and hospital resource optimization, remain in their early stages of development, they highlight the significant potential of DTs. Challenges include data quality, ethical issues, and socioeconomic barriers. This review also identified gaps in scalability, interoperability, and clinical validation.

Conclusions: DTs hold transformative potential in health care, providing individualized care, operational optimization, and accelerated research. However, their adoption is hindered by technical, ethical, and financial barriers. Addressing these issues requires interdisciplinary collaboration, standardized protocols, and inclusive implementation strategies to ensure equitable access and meaningful impact.

(J Med Internet Res 2025;27:e69544) doi:10.2196/69544

KEYWORDS

digital twins; meta-review; health IT; applications; challenges; healthcare innovation; personalized medicine; operational efficiency

Introduction

Digital twins (DTs) are digital replicas of physical entities, processes, or systems dynamically updated with real-time data to reflect their real-world counterparts [1-4]. They enable predictive modeling, decision support, and personalized simulations in health care. DTs are often presented as 3D models or dashboards; for example, a cardiac DT might appear as a real-time, visual simulation of a patient's heart. Clinicians can interact with the model by testing different treatment scenarios, such as adjusting medication dosages or simulating surgical interventions, and observe how the heart behaves—showing changes in function, blood flow, or potential complications in real time. This technology leverages sensors, analytics, and machine learning to enable monitoring, simulation, and predictive insights. Characterized by their ability to adapt to live data inputs, DTs facilitate real-time optimization, risk

XSL•FO
RenderX

assessment, and enhanced decision-making across industries, from manufacturing to health care.

DTs offer transformative advantages over traditional tools by providing real-time insights and predictive analytics that enhance operational efficiency and decision-making. In sectors such as aerospace, DTs have reduced development times by up to 50% and significantly improved product quality [5]. As this technology advances, the global DT market is expected to grow rapidly, reaching US $110.1 billion by 2028, with an annual growth rate of 60% [6]. In health care alone, the market is valued at US $1.6 billion in 2023 and is projected to grow to US $21.1 billion by 2028, representing nearly 20% of the total DT market [7].

Health care applications of DTs offer the potential to simulate complex patient scenarios, support predictive diagnostics, and enable real-time, personalized treatment planning based on live health data [2]. Yet, as a new technology in the biomedical field, DT research in health care encompasses a wide range of clinical and operational applications at different stages of development, spanning from theoretical models to advanced experimental stages. This diversity—from individualized patient modeling to hospital workflow optimization—creates a fragmented body of research, where theoretical concepts and experimental models vary widely in sophistication. The variability in data sources, model design, and intended goals limits consensus on best practices and implementation pathways [8]. These factors highlight the need for a systematic review to synthesize the existing research, map out the various applications of DTs, and identify the gaps and challenges that remain.

The objective of this review is threefold. First, it seeks to provide a structured overview of the field, highlighting the diverse applications of DTs in health care. Second, it identifies and examines the barriers to DT adoption, along with potential strategies to address them. Third, it pinpoints gaps in existing knowledge to guide future research. The research questions addressed are as follows: (1) What are the primary applications of DTs in health care as identified in the extant literature? (2) What are the key challenges and limitations associated with implementing DTs in health care? (3) What gaps exist in current DT research in health care, and what are the implications for future research and practice?

Methods

Study Design

This meta-review was conducted and reported following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines to ensure comprehensive reporting and methodological rigor [9]. The decision to use PRISMA-ScR reflects the emergent nature of DT research in health care, which requires a scoping approach to capture the breadth and diversity of this rapidly evolving field. Despite its emergent status, the field has already produced multiple literature reviews, indicating a growing maturity in this research stream. This apparent paradox reflects the rapid pace at which new knowledge is generated in DT research, necessitating both a broad mapping of the field and a deeper synthesis of existing findings. To address this dual need, we drew inspiration from the methodological approaches of Sarrami-Foroushani et al [10], incorporating elements of scoping meta-reviews. This hybrid approach allowed us to systematically map the literature while evaluating all existing reviews, enabling a comprehensive understanding of the field without restricting our analysis to systematic reviews alone.

Search Strategy and Selection Criteria

A comprehensive literature search was conducted across multiple electronic databases without date restrictions, retrieving all available records up to May 15, 2024. The databases included PubMed, CINAHL, Web of Science, Embase, and PsycINFO. The search strategy and database selection were developed in collaboration with a professional librarian, who also executed the searches. The protocol for this review is registered on Open Science Framework [11].

Key search terms included “digital twin,” “intelligent twin,” and “mirror twin” combined with “review,” “systematic review,” and “meta-analysis.” To ensure the inclusion of all potentially relevant studies, the reference lists of selected papers were manually reviewed (snowballing). The complete search strategy for each database is provided in Multimedia Appendix 1.

We set inclusion criteria at the start of this study, allowing for potential revisions during the process. Papers were eligible for inclusion if they met the following criteria: (1) literature review paper (any type), (2) specifically about DT, (3) in the context of health care, and (4) accessible in full text in English or French. If a paper referenced DT applications in both health care and other settings, only the health care–related applications were considered.

All references were organized and deduplicated using EndNote (Clarivate), while Covidence (Veritas Health Innovation Ltd) was used for this study selection process.

Study Selection

Two reviewers (MR and FAES) independently screened all titles and abstracts, with an initial screening of 30 references conducted to ensure a shared understanding of the selection criteria. This was followed by a full-text screening of selected studies to confirm eligibility.

The meta-review initially yielded 3377 references. After removing duplicates, we screened the titles and abstracts of 2921 references, retaining 65 for full-text review. Ultimately, 25 references were included in this review, with reasons for excluding others provided in the figure in the Results section.

Data Extraction

Two authors (MR and FAES) independently extracted data from each of the included studies. Discrepancies were resolved through discussion. The following data were recorded: author, year, type of literature review (as reported), research questions, search period, included databases, final sample size, what DT models, DT application, potential or actual applications, DT provider, DT users, DT impacts, DT challenges, and main study results.

XSL•FO
RenderX

Data Analysis

A mixed methods approach was used, combining descriptive and thematic analysis to map the landscape of DT technology in health care. Descriptive analysis was used initially to summarize the general characteristics of the included studies, covering the range of review types, publication years, and scope of DT applications.

As most reviews did not specify the type of review conducted (eg, scoping review and narrative review), we classified all reviews based on predefined criteria [12,13]. This classification process was conducted by one reviewer (MR) and verified by a second reviewer (FAES). Any discrepancies were resolved through discussion, ensuring a consistent and accurate categorization of review types.

Results

General Characteristics of the Included Reviews

The included studies were published between 2021 and 2024 (Figure 1). Most of the reviews were categorized as scoping reviews (14/25, 56%) [1,2,14-25]. The remaining reviews included 6 narrative reviews (6/25, 24%) [26-31], 2 rapid reviews (2/25, 8%) [8,32], and 3 reviews for which a type could not be determined (3/25, 12%; see Multimedia Appendix 2 for the PRISMA-ScR checklist) [3,4,33].

Figure 1. Flow diagram.

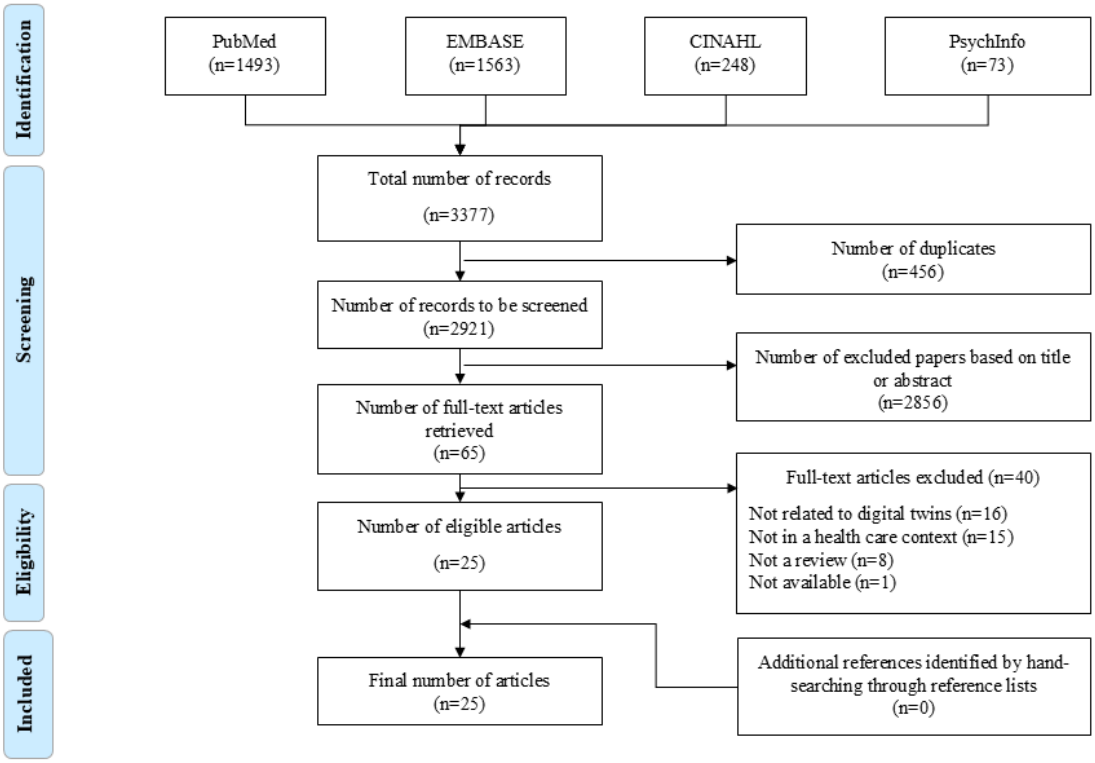

Studied Populations and Key Users of DTs

Targeted Populations

The reviews cover various populations, defined as the individuals or groups impacted by the outcomes of the DT application. For example, a DT designed for personalized medicine would mainly benefit patients, whether or not they were diagnosed with a specific disease.

Health care providers are a key target in 48% (12/25) of the reviews [2,8,19-23,25,29,30,32,33]. Similarly, patients in general are the focus of another 48% (12/25) of the reviews [2-4,8,14,19-21,27,28,30,32], underscoring the broad potential of DTs to impact patient care across diverse contexts. Beyond these general categories, 32% (8/25) of the reviews examine the general population [2,8,14,20,21,27-29], highlighting DT applications that extend to broader public health or societal benefits. Some studies further specify particular patient populations, such as those with cardiovascular diseases (7/25, 28%) [1,2,14,17,22,30,31] or multiple sclerosis (3/25, 12%) [20,24,25], suggesting targeted applications of DTs in managing chronic or complex conditions. Lastly, a small subset of reviews (2/25, 8%) does not specify any population for particular DT uses [30,32].

Users of DTs

DTs are used by various stakeholders, each benefiting uniquely from this technology. The primary users include patients [4,22-25,28,32,33], hospital management [8,14,19,21,28-30,32], and health care professionals, including clinicians [1,3,4,14,16,17,19,22-27,29-31,33], physicians [14,20,30,32], and researchers [2-4,14-17,19,20,22-25,27,29-31,33].

For patients, DTs provide tools to actively manage health conditions. For instance, patients with migraines [32] benefit from DT systems that integrate data from wearables and health records, offering real-time monitoring and predicting potential migraine triggers. This capability empowers patients to make informed decisions about lifestyle adjustments and treatment

XSL•FO
RenderX

adherence, enhancing personal control over their condition by reducing the frequency or severity of migraine episodes.

In hospital management, DTs enhance operations across four main functions [8,14,21]. First, they improve safety by simulating emergency scenarios, allowing hospitals to identify risks and refine response protocols to protect patient well-being. Second, DTs streamline information management by consolidating real-time data from various sources, which improves communication and decision-making across departments. Third, in promoting health and well-being, DTs support personalized patient care by providing tools that adapt treatment protocols to individual data and forecast patient health trends, assisting health care professionals in delivering tailored interventions and preventive care. Lastly, DTs ensure efficient operational control by optimizing resource allocation, managing patient flow, and enabling predictive maintenance for equipment, which reduces patient wait times and enhances overall hospital efficiency.

DTs also serve as valuable tools for health care professionals, especially in supporting clinical decision-making and personalized treatment strategies. For complex cases such as cardiovascular and immune-mediated diseases, DTs enable clinicians to simulate patient-specific responses by integrating clinical, genetic, and environmental data. In cardiovascular care [1,17,19,20,30,31], DTs simulate heart function and disease progression, supporting tailored interventions based on individual patient profiles. By predicting and addressing patient-specific risks in real time, DTs enhance precision in medical practice, advancing the quality and personalization of care.

In sum, DTs support distinct user groups within health care. Patients and health care professionals, especially clinicians and physicians, benefit from DTs by providing precise, individualized care. Hospital management relies on DTs to enhance resource allocation, patient flow, and adaptability to changing demands. Clinicians and researchers leverage DTs to simulate virtual trials, advancing drug development and treatment refinement. These applications correspond to 3 main areas: personalized medicine, operational efficiency and resource management, and advancements in medical research and drug development.

Main Categories of DT Applications and Potential Benefits

DTs have emerged as a transformative tool in health care, offering tailored solutions across various applications. The following sections present 3 critical categories of DT applications, each accompanied by tables outlining the inputs, processes, and outputs relevant to each application. These tables provide a structured view of how concrete DT applications address specific health care challenges, optimize resources, and drive advancements in patient care and medical research.

Category 1: Customized Treatment and Care Optimization

One of the main applications of DTs in health care is customized treatment and care optimization (Table 1). DTs span various stages of development, from proof-of-concept models [1,8,14,17,19,20,28,30] to applications under active research and clinical testing [1,4,8,14-17,19,20,22,30-32]. Companies such as Dassault Systèmes, through their Living Heart Project [1,14,17,19,20], and Siemens [14,17,19,26] are at the forefront of these advancements. Other institutions have also developed some specific DTs, such as Medtronic with 3D cardiac maps [14,17] and Oklahoma State University with targeting tumor-only locations [14]. Some studies underscore the potential for DTs in complex areas [1,8,17,19,20,23,24,26,29,30,32], underscoring the nascent stage of this innovation in the field. Given its relative novelty, it is unsurprising that some review articles focus on potential applications while others concentrate on actual implementations.

XSL•FO
RenderX

Table 1. DT[a] applications in customized treatment and care optimization.

Problem (and root causes)	Inputs	Processes	Outputs

• Variability in patient physiology
• Delayed detection
• Limited predictive capability for disease progression
	• Data:
	  • Clinical data
	  • Genomic data
	  • Imaging data
	  • Biomarkers
	  • Lifestyle data
	  • Demographic data
	• Sources:
	  • Electronic health record systems
	  • Disease registries
	  • Laboratory reports
	  • Wearable devices
	  • Surveys and questionnaires
		• Simulation: the DT can simulate a patient's disease progression, the impact of treatment options, and various clinical scenarios, such as a sudden deterioration in condition or the patient's response to a new medication. It can also model the effects of daily changes in lifestyle, nutrition, and medication adherence
		• Optimization: based on simulation results, the DT can optimize personalized treatment plans, suggesting specific drug doses, dietary adjustments, or lifestyle interventions tailored to each patient's condition and medical history
		• Real-time decision-making: the DT can immediately alert health care providers to emerging risks such as arrhythmias or abnormal glucose levels, prompting timely interventions or adjustments in treatment
			• Real-time monitoring and prognosis
			• Improved diagnosis
			• Tailored health interventions and lifestyle guidance

[a]DT: digital twin.

The customized treatment and care optimization approach addresses key challenges in health care, including variability in patient physiology, delayed detection of disease progression, and limited predictive capabilities. By integrating diverse data inputs—such as clinical records, genomic information, imaging data, biomarkers, and lifestyle and demographic factors—sourced from electronic health record (EHR) systems, disease registries, and wearable devices [1,8,14,19,20,23,32,33], DTs enable comprehensive simulations of patient health profiles.

Through these simulations, DTs model disease progression, evaluate potential treatment options, and assess various clinical scenarios, such as predicting patient responses to new medications or lifestyle changes [1,2,4,14,17,20,22,24,28-30,32]. Beyond simulation, DTs support optimization by dynamically tailoring treatment plans to a patient's evolving condition, which may include adjusting medication dosages, recommending dietary modifications, or suggesting lifestyle changes to improve outcomes [1,2,4,8,14,19,20,22-24,30,32].

Moreover, DTs support real-time decision-making by providing health care providers with alerts about emerging risks, such as abnormal glucose levels [1,14,30] or arrhythmias [31], enabling timely interventions. The outputs of this application—real-time monitoring, improved diagnosis, and personalized health interventions—underscore the potential of DTs to enhance personalized care and responsiveness in clinical practice.

Category 2: Operational Efficiency and Resource Management

Another significant category of DT applications in health care refers to operational efficiency and resource management (Table 2). DT technology serves as a valuable tool for optimizing hospital operations across various stages, from concept development [8,32] to active deployment [4,8,14,30]. For instance, GE HealthCare has developed a Command Center leveraging DTs to simulate and enhance patient flow in hospitals, including Johns Hopkins Hospital in Baltimore [8,14]. Similarly, organizations such as BioSecure and Siemens Healthineers have implemented DT-based solutions in health care settings [2,8] to optimize supply chain processes, improve response times for critical patients, and streamline workflows. While many initiatives demonstrate tangible benefits, others remain at the conceptual or exploratory stage, showcasing the potential for future DT applications in operational efficiency and resource management [21,23,32].

XSL•FO
RenderX

Table 2. DT[a] applications in operational efficiency and resource management.

Problem (and root causes)	Inputs	Processes	Outputs

• Unpredictable patient volumes
• Difficulty in dynamically adapting workflows in high-stakes situations
• Limited real-time operational insights that lead to inefficiencies and bottlenecks
	• Data:
	  • Patient data
	  • Historical trends
	• Sources:
	  • Hospital information systems
		• Simulation: the DT can simulate hospital scenarios. By modeling these patterns, DTs can analyze how variations in patient flow affect wait times, bed availability, staffing requirements, and equipment usage, giving administrators foresight into potential bottlenecks
		• Optimization: based on simulation results, the DT can recommend ideal staffing levels, identify underused or overbooked resources, and schedule maintenance during low-demand periods
		• Real-time decision-making: the DT can suggest workflow shifts, such as rerouting noncritical cases, and improve discharge planning to free up beds efficiently
			• Optimized patient flow and resource
			• Real-time adjustments in staffing and equipment usage based on patient arrivals
			• Improved discharge planning and maintenance scheduling to ensure optimal bed availability and patient throughput
			• Efficient crisis management strategies

[a]DT: digital twin.

The operational efficiency and resource management category addresses critical challenges such as unpredictable patient volumes, the difficulty of dynamically adapting workflows in critical situations, and the lack of real-time operational insights that often result in inefficiencies and bottlenecks. By leveraging inputs such as patient data and historical trends from hospital information systems [19,23], DTs can simulate various hospital scenarios. These simulations enable administrators to analyze how changes in patient flow impact key factors such as wait times, bed availability, staffing needs, and equipment use, thereby providing foresight into potential bottlenecks [2,8,14,19,23].

Beyond simulation, DTs enable optimization by recommending staffing adjustments, identifying underused or overburdened resources, and scheduling maintenance during low-demand periods [8,14,19,23,30,32]. Furthermore, DTs support real-time decision-making by suggesting workflow modifications, such as reallocating noncritical cases or expediting discharge planning to free up bed capacity [2,4,8,14,19,23]. The outputs of this category—optimized patient flow, real-time staffing and equipment adjustments, improved discharge planning, and effective crisis management strategies—highlight the significant potential of DTs to enhance resource use and operational responsiveness in health care facilities.

Category 3: Advancements in Medical Research and Drug Development

DT technology holds immense potential to transform medical research and drug development through in silico clinical trials, encompassing both prospective initiatives [3,17,19,30] and current applications (Table 3) [2-4,14,15,17,19,30]. For instance, collaborations involving Unlearn AI, Merck, Nvidia, and the University of Florida [14,19] are using DT models to simulate patient responses and design virtual clinical trials. These efforts enable the execution of single-arm studies, which are particularly valuable in scenarios where traditional randomized controlled trials may be unethical or impractical, such as rare diseases or high-risk populations. By simulating control groups, DTs help to reduce the need for placebo arms and contribute to minimizing adverse drug reactions [14,19,30]. Other applications of DTs focus on accelerating drug discovery and optimizing drug screening processes, offering promising avenues in pharmaceutical research [2-4,15,17].

XSL•FO
RenderX

Table 3. DT[a] applications in advancements of medical research and drug development.

Problem (and root causes)	Inputs	Processes	Outputs

• High costs and time demand of physical trials
• Ethical challenges in testing on human participants, especially in vulnerable populations
• Limited ability to predict individualized responses in diverse patient populations
	• Data:
	  • Clinical data
	  • Patient data
	  • Genomic data
	  • Imaging data
	  • Physiological data
	• Sources:
	  • Electronic health records
	  • Wearable and remote monitoring devices
	  • Health databases
	  • Pharmaceutical and clinical trial data
		• Simulation: the DTs are used to simulate in silico clinical trials, allowing researchers to predict drug efficacy and safety across diverse patient profiles without the need for large-scale physical trials
		• Optimization: with data from DTs, researchers can optimize trial parameters by adjusting dosages and selecting ideal patient cohorts. This approach refines drug formulations and identifies optimal therapeutic protocols, reducing unnecessary risks and costs associated with traditional clinical trials
		• Real-time decision-making: If adverse reactions are predicted, DTs can recommend alternative treatment strategies, improving patient safety and the trial's overall effectiveness
			• Synthetic control groups to minimize physical trials and enhance safety
			• Predictive modeling of drug responses and adverse reactions, allowing for more accurate and efficient trials
			• Simulations for drug efficacy testing, reducing the need for early-stage in-person trials
			• Virtualized, patient-specific trials that allow researchers to adjust dosages and predict patient outcomes

[a]DT: digital twin.

This category of DT applications addresses critical challenges in clinical trials, including the high costs and time demands of traditional physical trials, ethical concerns related to testing on human participants—particularly in vulnerable populations—and the limited capacity to predict individualized patient responses. By integrating diverse data inputs such as clinical, genomic, imaging, and physiological data from sources such as EHRs, health databases, and pharmaceutical trial data [2,3,30], DTs enable the simulation of clinical trials in digital environments.

These simulations allow researchers to evaluate early-stage drug efficacy and safety across diverse patient profiles without the need for large-scale physical trials. This approach reduces costs and enhances safety through synthetic control groups [2-4,14,17,19]. Beyond simulation, DTs enable the optimization of trial parameters by adjusting dosages, selecting ideal patient cohorts, refining drug formulations, and identifying optimal therapeutic protocols. These capabilities help minimize unnecessary risks and costs [2-4,14,15,19].

Additionally, DTs support real-time decision-making by predicting adverse reactions and suggesting alternative treatment strategies to improve both patient safety and the trial's overall success [2,3,14,17,19]. The outputs of these applications—including synthetic control groups, predictive modeling of drug responses, and virtualized patient-specific trials—demonstrate how DTs can significantly enhance the efficiency, accuracy, and personalization of clinical trials, making them more adaptable to the needs of diverse patient populations.

Types of DT

The classification of DT technologies originates from industry practices (eg, [34]), where DTs are typically categorized based on the scale and scope of the entities they represent. This framework enables a clear distinction between DTs by their purpose and level of detail, ranging from individual components to larger systems and entire processes.

Component or part DTs represent the most granular level, simulating individual components within larger systems. Examples include specific cardiac functions, such as aortic aneurysm [19], coronary anatomy [31], and coronary vessels [17]. Additionally, they encompass representations of individual cognitive profiles [25]. Asset or product DTs replicate individual physical assets, such as organs or medical equipment, to provide insights into performance and maintenance needs. This category includes organs [1,3,14,16,17,19,20,22,24,26,30,31], such as the heart [1,14,17,19,20,30,31] or the brain [14], as well as, to a lesser extent, diseases and therapeutic equipment [14,33]. System or unit DTs focus on the interactions of multiple components within a system. Examples include models of the human body (IDs 4, 21, 22, and 25), patients in various health conditions [1-3,14,17,19-21,27-30,32], and intensive care unit systems [8,14]. Lastly, process DTs simulate entire workflows to optimize health care operations. These include broader infrastructure models such as buildings [29], health care systems [8], hospitals [8,19], and physical systems [23].

DT Implementation Challenges

The implementation of DT technology in health care organizations faces several critical challenges, which can be

XSL•FO
RenderX

grouped into 3 main categories: data and model integrity; ethical, regulatory, and governance challenges; and implementation and socioeconomic disparities.

First, data and model integrity encompasses issues related to data quality, availability, and the robustness of DT models. Limited access to high-quality data, difficulties in data integration, and privacy concerns represent significant barriers to adoption. Additionally, computational power demands and scalability challenges further complicate implementation [1,2,4,8,14-21,23,24,26,27,29,32,33]. Ensuring model validation and reproducibility remains a pressing issue, as the absence of standardized methods and randomized controlled trials undermines the clinical credibility of DTs. Moreover, the need to enhance model complexity to better capture individual patient variations highlights the current technological limitations of DTs [3,16,19,22,26,28,31].

Second, ethical, regulatory, and governance challenges address ethical considerations and the regulatory frameworks surrounding the use of DTs. Persistent concerns about bias, fairness, and data ownership highlight the risk of exacerbating existing socioeconomic disparities. Ethical dilemmas arise regarding informed consent and data ownership, particularly in the context of data sharing and potential misuse [2,4,8,14,16,17,19,20,24,29]. Furthermore, regulatory frameworks for DTs remain underdeveloped, creating uncertainties in legal governance, intellectual property protection, and alignment with existing health care regulations [4,8,17-20,24].

Third, implementation and socioeconomic disparities address the challenges of integrating DTs into clinical practice and the socioeconomic barriers to widespread adoption. Health care professionals often express concerns about trust, transparency, and the risk of job displacement, contributing to resistance against DT implementation. Additionally, there is a pressing need for enhanced education and training for both health care providers and patients to bridge the knowledge gap between clinicians and data scientists [4,14,16,17,23,30,32]. On a broader scale, the socioeconomic impacts of DTs include the potential to deepen the digital divide, limit access to advanced technologies, and enable financial barriers such as cost and reimbursement issues, which hinder broader implementation [1,18-20].

Discussion

Principal Results

This review highlights the transformative potential of DT technology across 3 critical applications in health care: personalized medicine, operational efficiency, and medical research and drug development. While these advancements present significant opportunities, they also introduce challenges that require attention to fully realize the benefits of DTs.

In personalized care, DTs leverage diverse data sources—such as clinical records, genomic information, and real-time inputs from wearable devices—to create patient-specific health simulations. These simulations serve as a foundation for predictive modeling, enabling tailored interventions and optimized treatment plans, particularly in complex domains such as cardiovascular care and immune-mediated diseases. However, many applications remain at the proof-of-concept stage, underscoring the need for further research, data standardization, and validation efforts to transition from experimentation to widespread clinical adoption.

DTs enhance health care operations by simulating patient flow, resource allocation, and emergency scenarios, offering actionable insights to improve efficiency. For instance, GE HealthCare's Command Center demonstrates how DTs can reduce bottlenecks, optimize staffing, and enhance patient experiences [13]. However, adoption remains limited due to resistance to change, significant computational demands, and challenges in integrating DTs into existing workflows. Overcoming these barriers will require targeted training and interdisciplinary collaboration to facilitate the operational implementation of DTs.

DTs are transforming medical research by facilitating digital clinical trials and predictive modeling of drug responses. These innovations can significantly reduce the cost and ethical concerns associated with traditional trials while improving trial efficiency and safety. For example, collaborations between Unlearn AI and Merck are leveraging synthetic control groups to accelerate drug development [18]. However, broader implementation in this domain is contingent upon addressing critical challenges, including ensuring data quality, achieving scalability, and navigating regulatory approval processes.

Challenges and Opportunities in Advancing DT Technology in Health Care

Overview

While our findings highlight the significant promise of DTs in health care, they also reveal that most current applications remain in their infancy, with many still at the proof-of-concept or early implementation stage. These initial deployments, though limited in scope, demonstrate promising results and lay a solid foundation for future advancements. The potential applications of DTs extend well beyond these early efforts, offering innovative solutions to address some of health care's most pressing challenges. However, overcoming key barriers is essential to making DT technology more accessible, trustworthy, and practical for health care providers, thereby paving the way for wider adoption and meaningful impact.

First, challenges related to data and model integrity represent a significant concern for DT technology in health care. Integrating large datasets from wearables, EHRs, and imaging systems involves complexities in standardization, quality control, and real-time synchronization [35]. Ensuring model accuracy and minimizing artificial intelligence biases, as a key technology supporting DTs, are critical to making DTs reliable for high-stakes health care applications [36]. Addressing these issues requires standardized frameworks and rigorous validation protocols. Legal frameworks such as the European Union's General Data Protection Regulation could serve as a model for implementing robust data governance practices within DT technology. To minimize artificial intelligence biases, validated practices established in EHRs can be leveraged to enhance data

XSL•FO
RenderX

quality and ensure patient information integrity. Further, transparent validation processes, akin to clinical trials for medical devices, would foster trust and ensure DT reliability in clinical settings.

Next, ethical, regulatory, and governance challenges present significant concerns, particularly as DTs rely on sensitive health data. Issues such as data privacy, ownership, and the potential for algorithmic bias must be addressed to avoid undermining public trust and exacerbating disparities in care [37]. Adapting regulatory frameworks from virtual reality and internet of things devices could establish standards for data ownership, informed consent, and transparency, thereby mitigating risks of unauthorized data use. Furthermore, aligning these standards with international ethical guidelines could enhance public confidence in DT technology. The establishment of independent monitoring organizations—such as the European Medicines Agency [38], or ICANN (Internet Corporation for Assigned Names and Numbers) [39]—could provide consistent oversight and accountability. Such organizations would play a critical role in ensuring the ethical and responsible deployment of DTs across health care systems.

Finally, socioeconomic barriers to DT implementation raise concerns about exacerbating existing disparities in health care [40]. The high costs of advanced technologies create substantial obstacles for low- and middle-income countries to access essential tools needed for health innovation and economic modernization [41]. Moreover, the monopolization of patents and restrictive trade agreements by Western countries limits these nations' ability to develop local solutions [42]. Dependency on international loans often shifts focus toward debt repayment instead of strategic investments in health care infrastructure, further perpetuating inequality and widening the digital divide. Lessons learned from the implementation of EHRs and connected medical devices show that cost-sharing models and collaborative funding initiatives can promote more equitable access to DT technology [43]. Open-source tools and public-private partnerships involving technology companies, health care providers, and research institutions offer additional pathways to reduce costs and enhance accessibility. Additionally, targeted training programs for health care professionals can bridge the skills gap, facilitating the successful adoption of DTs across diverse health care settings [44,45].

Future Directions for DT Technology in Health Care

Expanding DT technology in health care opens several important avenues for future research. A key challenge is the lack of a universally accepted definition, particularly in health care. While most studies define the term consistently, some variability persists, reflecting broader inconsistencies in the field. This lack of consensus complicates efforts to synthesize findings and assess the maturity of DT applications. Recognized definitions, such as those by the National Academies [46] and Drummond and Gonsard [47], emphasize the dynamic, predictive, and decision-informing attributes that distinguish DTs from other digital tools, which set them apart from other digital tools. Achieving definitional clarity is foundational for advancing practical applications, as inconsistent definitions risk undermining coherence in design, validation, and implementation. Future research should prioritize refining and standardizing DT definitions to foster alignment between research and real-world needs, enabling trust among stakeholders and paving the way for innovative, targeted health care solutions.

Future research should prioritize the practical implementation of DTs through detailed case studies and real-world applications, demonstrating their feasibility and clinical impact. Transparency and explainability are equally critical for building clinicians' trust in DT recommendations and ensuring they can effectively interpret these insights. Addressing challenges related to scalability and data standardization is essential for widespread adoption, alongside rigorous validation through real-world clinical trials to establish the accuracy, safety, and seamless integration of DT systems into health care infrastructures.

Ethical considerations, including AI biases and patient privacy, require innovative approaches to algorithm development and the refinement of data ownership and informed consent models. Moreover, research into the socioeconomic impacts of DTs and cost-effective implementation strategies is vital to ensure accessibility for underserved populations. Special attention should be given to low-resource settings, where barriers such as cost, infrastructure, and training may hinder adoption. Collaboration between private and governmental sectors can ensure DT technology benefits a diverse range of patients and institutions, paving the way for equitable and sustainable adoption.

Limitations

This review has limitations that may affect the interpretation of its findings. As a meta-review, our analysis is constrained by the scope and depth of the included studies, which may not comprehensively capture all aspects of DT applications and developments in health care. Furthermore, the rapid growth of DT research in health care, particularly with significant contributions from private industry, limits the visibility of current implementation efforts. Industrial stakeholders often hesitate to publish detailed information about their proprietary DT technologies and methodologies, a challenge also observed during the COVID-19 pandemic [48]. This restricts access to real-world data, reducing insights that could deepen our understanding of DT advancements.

Conclusions

Though still in their early-stage applications, DTs hold immense potential to revolutionize health care through personalized medicine, operational efficiency, and accelerated medical research. However, realizing this potential requires addressing key challenges related to data integrity, ethical concerns, and socioeconomic disparities. Strategic investments, supportive policies, and collaborative research initiatives are crucial to ensuring equitable development. With an inclusive approach, DTs can transition from a visionary concept to a practical and transformative tool that enhances health care outcomes for diverse populations.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Jean Charbonneau, MLIS, librarian of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal Documentary Service, who validated the initial keywords and Medical Subject Headings terms of search queries and then executed the search strategy in selected databases.

Authors' Contributions

MR conceptualized this study, prepared this study's protocol, conducted the coding, drafted the initial paper, and contributed to its review and editing. FAES conceptualized this study, participated in coding, and reviewed and edited this paper. GP conceptualized and supervised this study and contributed to this paper's review and editing. MC conceptualized this study and reviewed and edited this paper. MR and FAES accessed and verified the data. All authors had full access to this study's data, including supplemental material, and approved the final version of this paper for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research queries.
[DOC File , 50 KB - jmir_v27i1e69544_app1.doc]

Multimedia Appendix 2
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 242 KB - jmir_v27i1e69544_app2.pdf]

References

1. Chu Y, Li S, Tang J, Wu H. The potential of the medical digital twin in diabetes management: a review. Front Med (Lausanne) 2023;10:1178912 [FREE Full text] [doi: 10.3389/fmed.2023.1178912] [Medline: 37547605]
2. Katsoulakis E, Wang Q, Wu H, Shahriyari L, Fletcher R, Liu J, et al. Digital twins for health: a scoping review. NPJ Digit Med 2024;7(1):77 [FREE Full text] [doi: 10.1038/s41746-024-01073-0] [Medline: 38519626]
3. Moingeon P, Chenel M, Rousseau C, Voisin E, Guedj M. Virtual patients, digital twins and causal disease models: paving the ground for in silico clinical trials. Drug Discov Today 2023;28(7):103605. [doi: 10.1016/j.drudis.2023.103605] [Medline: 37146963]
4. Venkatesh KP, Brito G, Boulos MNK. Health digital twins in life science and health care innovation. Annu Rev Pharmacol Toxicol 2024;64(1):159-170 [FREE Full text] [doi: 10.1146/annurev-pharmtox-022123-022046] [Medline: 37562495]
5. Digital twins: the key to smart product development. McKinsey & Company. URL: https://www.mckinsey.com/industries/industrials-and-electronics/our-insights/digital-twins-the-key-to-smart-product-development#/ [accessed 2025-02-04]
6. Research and Markets. URL: https://www.researchandmarkets.com/reports/5146336/digital-twin-market-by-application-predictive [accessed 2025-02-04]
7. MarketsandMarkets. URL: https://www.marketsandmarkets.com/Market-Reports/digital-twins-in-healthcare-market-74014375.html [accessed 2025-02-04]
8. Elkefi S, Asan O. Digital twins for managing health care systems: rapid literature review. J Med Internet Res 2022;24(8):e37641 [FREE Full text] [doi: 10.2196/37641] [Medline: 35972776]
9. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
10. Sarrami-Foroushani P, Travaglia J, Debono D, Clay-Williams R, Braithwaite J. Scoping meta-review: introducing a new methodology. Clin Transl Sci 2015;8(1):77-81 [FREE Full text] [doi: 10.1111/cts.12188] [Medline: 25041546]
11. van den Akker O, Peters GJ, Bakker C. Generalized Systematic Review Registration Form. MetaArXiv. Preprint posted online on September 22, 2023 [FREE Full text]
12. Paré G, Wagner G, Prester J. How to develop and frame impactful review articles: key recommendations. J Decis Syst 2023;33(4):566-582. [doi: 10.1080/12460125.2023.2197701]
13. Paré G, Kitsiou S. Handbook of eHealth evaluation: an evidence-based approach [internet]. University of Victoria. URL: https://www.ncbi.nlm.nih.gov/books/NBK481583/ [accessed 2025-02-04]
14. Armeni P, Polat I, De Rossi LM, Diaferia L, Meregalli S, Gatti A. Digital twins in healthcare: is it the beginning of a new era of evidence-based medicine? A critical review. J Pers Med 2022;12(8):1255 [FREE Full text] [doi: 10.3390/jpm12081255] [Medline: 36013204]

XSL•FO
RenderX

15. Bai J, Lu Y, Wang H, Zhao J. How synergy between mechanistic and statistical models is impacting research in atrial fibrillation. Front Physiol 2022;13:957604 [FREE Full text] [doi: 10.3389/fphys.2022.957604] [Medline: 36111152]
16. Benson M. Digital twins for predictive, preventive personalized, and participatory treatment of immune-mediated diseases. ATVB 2023;43(3):410-416. [doi: 10.1161/atvbaha.122.318331]
17. Coorey G, Figtree GA, Fletcher DF, Snelson VJ, Vernon ST, Winlaw D, et al. The health digital twin to tackle cardiovascular disease-a review of an emerging interdisciplinary field. NPJ Digit Med 2022;5(1):126 [FREE Full text] [doi: 10.1038/s41746-022-00640-7] [Medline: 36028526]
18. Drummond D, Roukema J, Pijnenburg M. Home monitoring in asthma: towards digital twins. Curr Opin Pulm Med 2023;29(4):270-276. [doi: 10.1097/MCP.0000000000000963] [Medline: 37102597]
19. Dumas M, Fay AF, Charpentier E, Matricon J. [Digital twins in healthcare: state of the art and potential use cases in a hospital setting]. Med Sci (Paris) 2023;39(12):953-957 [FREE Full text] [doi: 10.1051/medsci/2023178] [Medline: 38108726]
20. Boulos MNK, Zhang P. Digital twins: from personalised medicine to precision public health. J Pers Med 2021;11(8):745 [FREE Full text] [doi: 10.3390/jpm11080745] [Medline: 34442389]
21. Khan A, Milne-Ives M, Meinert E, Iyawa GE, Jones RB, Josephraj AN. A scoping review of digital twins in the context of the COVID-19 pandemic. Biomed Eng Comput Biol 2022;13:11795972221102115 [FREE Full text] [doi: 10.1177/11795972221102115] [Medline: 35633868]
22. Sun T, He X, Li Z. Digital twin in healthcare: recent updates and challenges. Digit Health 2023;9:20552076221149651 [FREE Full text] [doi: 10.1177/20552076221149651] [Medline: 36636729]
23. Vallée A. Digital twin for healthcare systems. Front Digit Health 2023;5:1253050 [FREE Full text] [doi: 10.3389/fdgth.2023.1253050] [Medline: 37744683]
24. Voigt I, Inojosa H, Dillenseger A, Haase R, Akgün K, Ziemssen T. Digital twins for multiple sclerosis. Front Immunol 2021;12:669811 [FREE Full text] [doi: 10.3389/fimmu.2021.669811] [Medline: 34012452]
25. Tacchino A, Podda J, Bergamaschi V, Pedullà L, Brichetto G. Cognitive rehabilitation in multiple sclerosis: three digital ingredients to address current and future priorities. Front Hum Neurosci 2023;17:1130231 [FREE Full text] [doi: 10.3389/fnhum.2023.1130231] [Medline: 36908712]
26. Wu C, Lorenzo G, Hormuth DA, Lima EABF, Slavkova KP, DiCarlo JC, et al. Integrating mechanism-based modeling with biomedical imaging to build practical digital twins for clinical oncology. Biophys Rev (Melville) 2022;3(2):021304 [FREE Full text] [doi: 10.1063/5.0086789] [Medline: 35602761]
27. Meijer C, Uh HW, El Bouhaddani S. Digital twins in healthcare: methodological challenges and opportunities. J Pers Med 2023;13(10):1522 [FREE Full text] [doi: 10.3390/jpm13101522] [Medline: 37888133]
28. Laubenbacher R, Adler F, An G, Castiglione F, Eubank S, Fonseca LL, et al. Toward mechanistic medical digital twins: some use cases in immunology. Front Digit Health 2024;6:1349595 [FREE Full text] [doi: 10.3389/fdgth.2024.1349595] [Medline: 38515550]
29. Jia P, Liu S, Yang S. Innovations in public health surveillance for emerging infections. Annu Rev Public Health 2023;44(1):55-74 [FREE Full text] [doi: 10.1146/annurev-publhealth-051920-093141] [Medline: 36626834]
30. Fischer R, Volpert A, Antonino P, Ahrens TD. Digital patient twins for personalized therapeutics and pharmaceutical manufacturing. Front Digit Health 2023;5:1302338 [FREE Full text] [doi: 10.3389/fdgth.2023.1302338] [Medline: 38250053]
31. Cluitmans MJM, Plank G, Heijman J. Digital twins for cardiac electrophysiology: state of the art and future challenges. Herzschrittmacherther Elektrophysiol 2024;35(2):118-123 [FREE Full text] [doi: 10.1007/s00399-024-01014-0] [Medline: 38607554]
32. Gazerani P. Intelligent digital twins for personalized migraine care. J Pers Med 2023;13(8):1255 [FREE Full text] [doi: 10.3390/jpm13081255] [Medline: 37623505]
33. Sigawi T, Ilan Y. Using constrained-disorder principle-based systems to improve the performance of digital twins in biological systems. Biomimetics (Basel) 2023;8(4):359 [FREE Full text] [doi: 10.3390/biomimetics8040359] [Medline: 37622964]
34. What is a digital twin? IBM. URL: https://www.ibm.com/topics/what-is-a-digital-twin [accessed 2025-02-04]
35. Martínez-García M, Hernández-Lemus E. Data integration challenges for machine learning in precision medicine. Front Med (Lausanne) 2021;8:784455 [FREE Full text] [doi: 10.3389/fmed.2021.784455] [Medline: 35145977]
36. Putrama IM, Martinek P. Heterogeneous data integration: challenges and opportunities. Data Brief 2024;56:110853 [FREE Full text] [doi: 10.1016/j.dib.2024.110853] [Medline: 39286416]
37. Alhammad N, Alajlani M, Abd-Alrazaq A, Epiphaniou G, Arvanitis T. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024;26:e50715 [FREE Full text] [doi: 10.2196/50715] [Medline: 38820572]
38. What we do. European Medicines Agency. URL: https://www.ema.europa.eu/en/about-us/what-we-do [accessed 2025-02-04]
39. ICANN for Beginners. Internet Corporation for Assigned Names and Numbers. URL: https://www.icann.org/en/beginners [accessed 2025-02-04]
40. Woods WA, Watson M, Ranaweera S, Tajuria G, Sumathipala A. Under-representation of low and middle income countries (LMIC) in the research literature: ethical issues arising from a survey of five leading medical journals: have the trends

XSL•FO
RenderX

changed? Glob Public Health 2023;18(1):2229890 [FREE Full text] [doi: 10.1080/17441692.2023.2229890] [Medline: 37401751]
41. Dinh MN, Nygate J, Tu VHM, Thwaites CL, Global Grand Challenges Event Vietnam Group. New technologies to improve healthcare in low- and middle-income countries: Global Grand Challenges satellite event, Oxford University Clinical Research Unit, Ho Chi Minh City, 17th-18th September 2019. Wellcome Open Res 2020;5:142 [FREE Full text] [doi: 10.12688/wellcomeopenres.16008.2] [Medline: 32864470]
42. Sachs JD, Karim SSA, Aknin L, Allen J, Brosbøl K, Colombo F, et al. The Lancet commission on lessons for the future from the COVID-19 pandemic. Lancet 2022;400(10359):1224-1280. [doi: 10.1016/s0140-6736(22)01585-9]
43. Tsai CH, Eghdam A, Davoody N, Wright G, Flowerday S, Koch S. Effects of electronic health record implementation and barriers to adoption and use: a scoping review and qualitative analysis of the content. Life (Basel) 2020;10(12):327 [FREE Full text] [doi: 10.3390/life10120327] [Medline: 33291615]
44. Franzen SRP, Chandler C, Lang T. Health research capacity development in low and middle income countries: reality or rhetoric? A systematic meta-narrative review of the qualitative literature. BMJ Open 2017;7(1):e012332 [FREE Full text] [doi: 10.1136/bmjopen-2016-012332] [Medline: 28131997]
45. Shumba CS, Lusambili AM. Not enough traction: barriers that aspiring researchers from low- and middle-income countries face in global health research. Journal of Global Health Economics and Policy 2021;1:e2021002. [doi: 10.52872/001c.25802]
46. Foundational Research Gaps and Future Directions for Digital Twins. National Academies. Washington DC: National Academies Press; 2024. URL: https://nap.nationalacademies.org/read/26894/chapter/1 [accessed 2025-02-04]
47. Drummond D, Gonsard A. Definitions and characteristics of patient digital twins being developed for clinical use: scoping review. J Med Internet Res 2024;26:e58504 [FREE Full text] [doi: 10.2196/58504] [Medline: 39536311]
48. Wouters OJ, Shadlen KC, Salcher-Konrad M, Pollard AJ, Larson HJ, Teerawattananon Y, et al. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. Lancet 2021;397(10278):1023-1034. [doi: 10.1016/s0140-6736(21)00306-8]

Abbreviations

DT: digital twin
EHR: electronic health record
ICANN: Internet Corporation for Assigned Names and Numbers
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by T de Azevedo Cardoso; submitted 02.12.24; peer-reviewed by D Drummond, D Madell; comments to author 10.12.24; revised version received 20.01.25; accepted 24.01.25; published 19.02.25.

Please cite as:
Ringeval M, Etindele Sosso FA, Cousineau M, Paré G
Advancing Health Care With Digital Twins: Meta-Review of Applications and Implementation Challenges
J Med Internet Res 2025;27:e69544
URL: https://www.jmir.org/2025/1/e69544
doi:10.2196/69544
PMID:39969978

©Mickaël Ringeval, Faustin Armel Etindele Sosso, Martin Cousineau, Guy Paré. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Impact of Patient Online Record Access on Documentation: Scoping Review

Eva Meier-Diedrich[1,2*], MSc; Camilla Lyckblad[3,4*], MSc; Gail Davidge[5], PhD; Maria Hägglund[6,7], PhD; Anna Kharko[6], PhD; Brian McMillan[5], PhD; Charlotte Blease[6*], PhD; Julian Schwarz[1,2*], MD

[1]Department of Psychiatry and Psychotherapy, Center for Mental Health, Immanuel Hospital Rüdersdorf, Brandenburg Medical School Theodor Fontane, Rüdersdorf, Germany

[2]Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

[3]Department of Archives, Libraries and Museums, Uppsala University, Uppsala, Sweden

[4]Centre for Medical Humanities, Uppsala University, Uppsala, Sweden

[5]Centre for Primary Care and Health Services Research, University of Manchester, Manchester, United Kingdom

[6]Participatory eHealth and Health Data Research Group, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[7]Uppsala University Hospital, Uppsala, Sweden

[*]these authors contributed equally

Corresponding Author:
Julian Schwarz, MD
Department of Psychiatry and Psychotherapy
Center for Mental Health, Immanuel Hospital Rüdersdorf
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf, 15562
Germany
Phone: 49 33638 83 501
Email: julian.schwarz@mhb-fontane.de

Abstract

Background: Online record access (ORA) is being increasingly implemented internationally. Despite reported benefits for patients, health care professionals (HCPs) have raised concerns about potential disadvantages. To date, no review has examined the empirical evidence on whether and how documentation changes following the introduction of patients' ORA.

Objective: This scoping review aimed to examine potential subjective and objective changes in HCPs' documentation after using patients' ORA.

Methods: A scoping review was conducted using a methodological framework for scoping reviews and data from 4 electronic databases. Studies examining objective and subjective changes in clinical documentation following the implementation of ORA, specifically those related to actual use experiences (rather than previous expectations), up to July 2023, were included. We used the Mixed Methods Appraisal Tool to assess the quality of the included studies. The narrative synthesis and reporting of findings were guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews).

Results: Of the 3143 papers screened, 42 (1.34%) were included in this review. The included studies mainly used qualitative methods and were predominantly published after 2016 in the United States. The included studies were conducted in different settings (inpatient and outpatient) and clinical areas (somatic, mental health, and other). In total, 8 studies analyzed clinical notes, while the remaining studies focused on the experiences of patients, HCPs, and other stakeholders with ORA. Objectively, a decrease in complexity, an increase in readability, and a change in the emotional tone of the clinical notes were observed. The length of the clinical notes was observed to change both objectively and subjectively, although the direction of this change was inconclusive. However, many HCPs also reported writing notes that were less open and more restrictive to protect sensitive or hypothetical information. While for some HCPs the implementation of ORA made the clinical notes a less efficient and valuable working tool, others perceived that ORA opened up new therapeutic opportunities through direct contact with patients.

Conclusions: The question of whether an inherently uniform clinical note can meet the diverse needs of different health care stakeholders remains unresolved, highlighting the challenges of standardizing practices in this complex sector. While ORA may

XSL•FO
RenderX

encourage HCPs to make their clinical notes more patient friendly, it may also compromise the integrity of documentation by omitting sensitive findings and expert judgment, which can put patients at risk and lead to errors that increase the risk of malpractice. Given the limitations of digital documentation in fostering trust, it is imperative to prioritize meaningful patient-HCP interactions. The use of compensatory measures, such as parallel documentation and restricted access to clinical notes, indicates systemic problems and suggests that current practices are suboptimal.

International Registered Report Identifier (IRRID): RR2-10.2196/46722

(J Med Internet Res 2025;27:e64762) doi:10.2196/64762

KEYWORDS

open notes; electronic health record; open record access; clinical notes; eHealth

Introduction

Background

The electronic health record (EHR) has evolved over time. Initially designed as a memory aid for physicians, a communication tool among clinicians, and a system for billing and reimbursement, it has more recently been made accessible to patients [1,2]. In a growing number of countries, patients are now granted partial or full access to their EHRs [3-7]. The Nordic countries and the United States have been at the forefront of this movement [8-10]. A crucial component of patients' online record access (ORA) involves accessing the clinical free-text notes written by clinicians. Granting patients access to these notes is commonly referred to as "open notes" in the literature [3]. Patient ORA reflects the zeitgeist of greater transparency in societal institutions and health care [11]. Several benefits have been cited as motivations for this practice, including empowering patients through transparency and access to information, as well as fostering innovation within the health application economy [8,12].

While ORA fulfills patient demand for transparency in care, it also necessitates a cultural shift toward higher degrees of openness among institutions and HCPs, possibly acting as a disrupter in information management behaviors among the clinicians creating them. Previous research indicates that health care professionals (HCPs) have frequently expressed skepticism toward patients' ORA [13-15]. HCPs have voiced concerns about increased workload, changed clinical routines, and ORA impacting patient safety and privacy [1,13,14,16]. Some have expressed concerns over spending more time writing notes and addressing patient inquiries and also anticipated confusion and offense among patients, particularly regarding mental health issues [3]. In terms of documentation, many HCPs expected to alter both the content and tone of their notes when patients have ORA, indicating that the practice could potentially undermine the integrity of their records [4,5,17]. For instance, a tendency to avoid technical terminology and medical jargon to enhance patient comprehension might detrimentally affect multidisciplinary communication within the team [5,18]. Other HCPs worried that they may become less detailed or candid in their documentation, feeling the need to omit information or resort to parallel documentation (a so-called shadow record) to shield patients from potentially harmful or disruptive information [13,14,19,20]. An often-overlooked risk of ORA is that individuals considered vulnerable may be pressured into revealing their records to third parties, such as relatives or abusers, potentially leading to unauthorized access to sensitive patient information without consent [13,14,19,20]. However, it is also possible that ORA could make notes more patient friendly by encouraging clinicians to use more patient-centered, more understandable, and less stigmatizing language while stimulating communication between HCPs and patients [4,21].

Studies investigating the impact of ORA on clinical documentation have predominantly concentrated on the patient experience, with limited research examining the HCPs' perspective [22]. As noted by Blease et al [1], while these studies investigate self-reports about possible documentation changes, few studies have focused on objective changes following implementation. Where such studies exist, they appear to offer inconclusive results [4,23,24] and are frequently hindered by methodological limitations. There is a growing body of qualitative research [25], along with research using natural language processing methods to examine the language used by clinicians in their records, including the potential for stigmatizing language [22]. However, it remains unclear from these studies whether patient ORA influences or indeed enhances the quality of record keeping, given the awareness that patients may read the clinician's notes [22,26].

Objectives

As highlighted earlier, HCPs are often reluctant or critical toward granting patients ORA and anticipate an additional documentation burden upon its introduction. Therefore, this scoping review focuses exclusively on studies containing postimplementation data, encompassing the experiences of various stakeholders, such as patients, HCPs, and other health care providers, while excluding preimplementation expectations. This study seeks to (1) identify, compile, and assess reported objective and subjective changes in documentation following the implementation of ORA; (2) enhance stakeholders' knowledge of the types of documentation changes that may arise because of ORA policy implementation; (3) highlight implications for documentation practices and offer recommendations for improving future clinical practice; and (4) identify knowledge gaps warranting further research.

Methods

Scoping Review

Compared to the systematic review method, which is guided by a strongly focused research question, a scoping review aims to broaden the spectrum of the available evidence in a relatively

XSL•FO
RenderX

new field of research, allowing its breadth and depth to be clearly seen [27]. We conducted a scoping review following the framework proposed by Arksey and O'Malley [27]. Their approach consists of the following five stages: (1) identifying the research question, (2) identifying the relevant studies, (3) selecting eligible studies, (4) collecting data, and (5) summarizing data and synthesizing results. The review is reported following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Multimedia Appendix 1) [28,29]. We adhered closely to the methodological approach outlined in our published review protocol [30]. Any minor deviations are comprehensively described.

Stage 1: Identifying the Research Question

Through discussions with the research team, we decided on the following research questions: (1) Does clinical documentation change after introducing ORA for patients? and (2) If so, what objective and subjective changes arise after ORA implementation? By objective, we mean the differences that can be demonstrated by a direct, quantifiable comparison of clinical notes before and after the implementation of ORA. By subjective, we refer to clinicians' perceptions of how they write their notes after ORA implementation. In the context of this scoping review, we define ORA as any channel in which patients have electronic access to their patient records (eg, through the internet and via tethered patient portals and apps).

Stage 2: Identifying Relevant Studies

First, the research team performed a rigorous manual search to obtain a basic overview of the available evidence and to refine the scope of the review as well as the search strategy, as suggested by Popay et al [31]. The literature search was then conducted in the following 4 databases on July 31, 2023: APA PsycINFO, CINAHL, PubMed, and Web of Science Core Collection. The deduplication process was then performed. The search strategy consisted of three key concepts: (1) EHRs, (2) sharing EHRs with patients, and (3) changes in documentation, which were combined with the Boolean "AND" (Textbox 1). The search terms were adapted according to different databases. The complete search string can be found in Multimedia Appendix 2.

Textbox 1. Key concepts of the search strategy.

Search string for the electronic health record

- "inpatient portal*" OR "open notes" OR opennotes OR PAEHR OR "patient portal*" OR "patient web portal*" OR "Electronic Health Records"
- "clinic notes" OR "clinical notes" OR "progress notes" OR "doctors notes" OR EHR OR "health record*" OR "health care record*" OR "medical record*" OR "mental health notes" OR "patient record*" OR "psychiatric notes" OR "psychotherapy notes" OR "visit notes"

Search string for sharing the electronic health records with patients

- "guardian access" OR "parental access" OR "parents access" OR "patient access*" OR "patients access*" OR "patient online access" OR "patients online access" OR "proxy access" OR "shared medical record*" OR "shared health record*"

Search string for documentation changes

- "Language"[Mesh] OR "Attitude"[Mesh] OR "Comprehension"[Mesh]
- accura* OR ambigu* OR characteristics OR characters OR clarity OR content* OR completeness OR comprehend* OR comprehensibl* OR comprehension* OR correctness OR dialog* OR express* OR directness OR impression* OR inaccura* OR incomplete* OR incomprehen* OR incorrectness* OR intelligib* OR interpret* OR intuitive* OR language OR length OR linguistic* OR misconception* OR misinterpret* OR misread* OR misunderstand* OR monolog* OR negative* OR pattern* OR positive* OR pronoun* OR readab* OR style* OR simplicity OR terminolog* OR transparen* OR truthful* OR unambigu* OR understand* OR untruthful* OR veracity OR wordcount* OR words OR writing
- OR attitude* OR emotion* OR experience* OR perception* OR satisfact*
- OR adopt* OR alter* OR censor* OR change* OR changing OR difference* OR introduc* OR implement* OR modif* OR postimplement*

Stage 3: Selecting Eligible Studies

Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were defined by the entire research team and were applied in the study selection process (Textbox 2). Due to the limited number of publications available on the subject, there were no restrictions on the study type. As ORA is only gradually being implemented in various countries, we refrained from any location restrictions. A wide variety of approaches exist to make clinical notes available to patients electronically [32]. We included all studies examining the actual implementation and the use of patient ORA regardless of the platform (eg, web browser or mobile apps). Studies that explored the sharing of hard copies of patients' clinical records were excluded.

During the review process, we refined the inclusion and exclusion criteria as follows: we required studies to provide empirical data on changes in clinical documentation resulting from ORA. Studies that solely focused on secure messaging between patients and clinicians were excluded.

XSL•FO
RenderX

Textbox 2. Inclusion and exclusion criteria.

Inclusion criteria

- Study design: all study types
- Publication: original, peer-reviewed work including empirical data published between January 1, 2005, and June 30, 2023, in English
- Study location: all medical disciplines, all health care settings, and no location restrictions
- Study participants: patients and health care professionals of all ages
- Studies that examine actual use by stakeholders and their experiences with patient-accessible electronic health records
- Studies that provide empirical data on documentation changes resulting from the use of online record access

Exclusion criteria

- Paper-based, disc, or USB sharing of patients' records
- Articles without empirical data (eg, comments, editorials, and news)
- Gray data (websites, tweets, and blogs)
- Studies that exclusively investigate expectations about patient-accessible electronic health records
- Studies on secure messaging

Study Selection Process

We used Rayyan Software (Rayyan Systems, Inc) for conducting a collaborative, single-blinded title and abstract screening following the predefined eligibility criteria [33]. All research team members participated in this process, and at least 2 people evaluated each record of the result set. Discrepancies were discussed, taking the full texts of the corresponding studies into account. In case of disagreements that could not be resolved, a third reviewer was involved and entrusted with the decision of including or excluding the study.

Stage 4: Collecting Data

After selecting the studies to include, metadata (eg, title, authors, and publication year) of the remaining records were exported and summarized in a Google Sheets (Google LLC) spreadsheet for further processing. To extract and organize relevant data from included studies, the spreadsheet was extended by the following parameters based on the studies' full text: country, study design, sample, characteristics of study participants (eg, gender, age, ethnicity, and type of stakeholder), treatment setting and medical specialty, study purpose, and relevant results (Multimedia Appendix 3). Data extraction was performed by EM-D and checked for correctness and completeness by CL. The quality and methodological rigor of the studies were assessed using the Mixed Methods Appraisal Tool (MMAT) [34]. Two reviewers (CL and EM-D) independently conducted the MMAT grading of all studies and reached consensus concerning the methodological quality of the studies (Multimedia Appendix 4). Two additional researchers (AK and JS) validated the MMAT grading for correctness.

Stage 5: Summarizing Data and Synthesizing Results

Study results were extracted from the full texts by the lead author (EM-D) and summarized in (1) a reduced format within a textbox, providing an overview of the findings from all included studies; and (2) a detailed version for narrative synthesis. The latter was analyzed independently by 2 researchers (CL and EM-D) using thematic analysis [35]. Objective and subjective changes in HCPs' documentation practices after the introduction of patient ORA served as guiding deductive themes, informed by the research question. However, they were open to modifications during the analytic process. As suggested by Levac et al [28], we aimed to identify patterns and relationships within and across studies to identify potential factors influencing documentation after ORA implementation. In assessing the methodological rigor of the studies, we also envisaged the potential to identify research gaps; for example, we predicted that there may be a preponderance of survey research investigating clinicians' perceptions about documentation changes rather than studies investigating objective markers of any such documentation changes. While the former studies may be useful, they may be compromised by responder biases. Results were discussed and approved by the entire research team.

Ethical Considerations

This study did not require ethics approval, as it only used publicly available data and followed the scoping review methodology.

Results

Study Selection

A total of 6036 records were identified: 1261 (20.89%) from CINAHL, 351 (5.82%) from PsycINFO, 2364 (39.17%) from PubMed, and 2060 (34.12%) from the Web of Science Core Collection. After removing duplicates, 52.07% (3143/6036) of the records remained for title, abstract, and keyword screening. At this stage, an additional 36.41% (2198/6036) of the records were eliminated, leaving 15.66% (945/6036) of the records for full-text screening to check eligibility. During the full-text screening, 903 studies were excluded, resulting in a final selection of 42 studies that met the inclusion criteria and could be included in the review. The PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram

XSL•FO
RenderX

(Figure 1), adapted from the study by Page et al [36], provides a detailed representation of the study selection process.

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram for study inclusion. ORA: online record access.

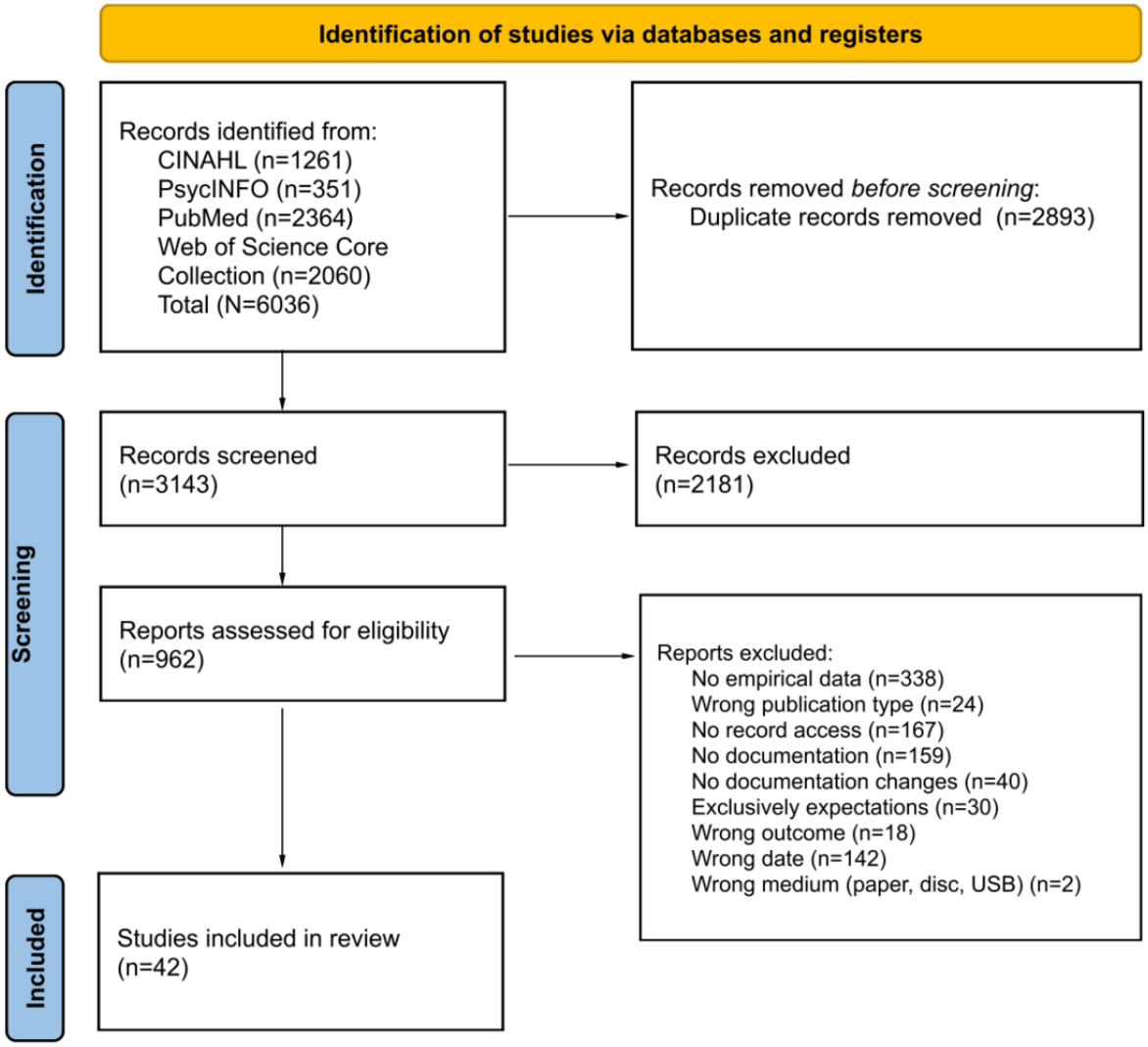

Basic Characteristics of the Body of Evidence

The included studies mainly used qualitative methods, but quantitative, descriptive, and mixed methods were also frequent. Of the 42 studies included, 23 (55%) were conducted in the United States, followed by Sweden with 9 studies (21%). Since 2017, the number of articles published on the topic of documentation changes following ORA implementation has remained relatively constant at 4 to 6 relevant articles per year. Only 19% (8/42) of the included studies analyzed clinical notes and reported on objective documentation changes, while most studies (34/42, 81%) investigated subjective documentation changes. A comprehensive overview of the basic parameters of the included studies can be found in Table 1.

XSL•FO
RenderX

Table 1. Basic characteristics of the included studies (N=42).

Parameter	Studies, n (%)	References
Study design		
Quantitative descriptive	11 (26)	[14,23,24,37-44]
Quantitative nonrandomized	5 (12)	[23,45-48]
Mixed methods	10 (24)	[4,19,49-56]
Qualitative	16 (38)	[3,13,17,20,57-68]
Publication year		
2011-2016	8 (19)	[14,37,38,44,47,48,57,63]
2017-2020	18 (43)	[3,4,39,40,43,49,51,52,55,58,60-62,64,67]
2021-2024	16 (38)	[13,17,20,23,41,42,45,46,50,53,54,56,59,65,68,69]
Country		
Canada	1 (2)	[52]
Netherlands	1 (2)	[41]
Norway	3 (7)	[51,67,68]
Sweden	9 (21)	[19,37,43,55,58,62-65]
United Kingdom	5 (12)	[13,17,20,59,61]
United States	23 (55)	[3,4,14,23,24,38-40,42,44-50,53,54,56,57,60,66,69]
Participants[a]		
Health care professionals	40 (95)	All, except for [40,47]
Patients	9 (21)	[20,38,44,45,48,54,57,59,66]
Care partners	3 (7)	[52,56,59]
Studies analyzing clinical notes	8 (19)	[4,23,24,40,45,47,49,53]
Treatment setting[a]		
Inpatient	10 (24)	[38,51,52,55,56,59,60,62,66,67]
Outpatient	29 (69)	[3,4,14,17,20,23,38,39,43-49,52,55,57,58,60-62,64-69]
Not specified	10 (24)	[19,24,37,40,42,50,53,54,63]
Clinical field[a]		
Mental health care	9 (21)	[3,14,19,51,60,62,65,66,68]
Somatic		
Primary	13 (31)	[13,20,38,44-46,53,55,57,58,65,67,69]
Oncology	8 (19)	[4,23,24,40,43,49,63,64]
General	3 (7)	[17,41,61]
Other specialities	5 (12)	[37,48,50,56,59]
Multispecialty	6 (14)	[39,42,47,51,52,63]

[a]Individual articles can be assigned to the various subparameters at the same time, which means that percentages >100% can be achieved.

Search Results

While several studies specifically examined documentation changes due to ORA implementation (11/42, 26%) [23,24,40,43,45-47,49,53,67,68], many more reported them as a secondary outcome (31/42, 74%). The results were divided into three groups: (1) objective changes, (2) subjective changes, and (3) influences on documentation practices. Both objective and subjective changes and a lack of changes were observed in the included studies. The categories identified for subjective changes were note characteristics, changes in content and functionality, and absence of subjective changes. Figure 2 provides a visual summary of the objective and subjective changes identified in the documentation. For documentation behaviors, the categories identified were influence on writing practices, secure information, and influence of sociodemographics.

XSL•FO
RenderX

Figure 2. Graphic overview of the objective and subjective changes in documentation.

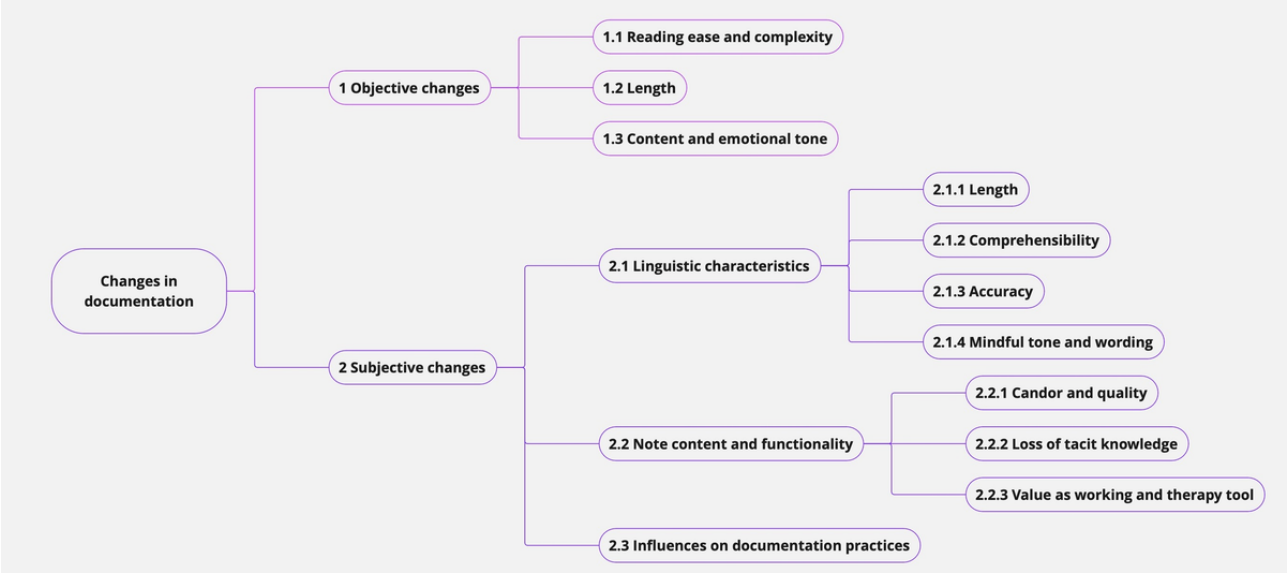

Objective Changes in Documentation

Overview

Of the 42 studies we reviewed, 8 (19%) examined clinical notes and 7 (17%) reported objective changes in documentation due to ORA. The number of notes analyzed varied notably, spanning from 200 [4] to 164,806 [53]. The studies used validated scales and tools such as Linguistic Inquiry and Word Count, Flesch-Kincaid Grade Level, Flesch Reading Ease Scores, Gunning Fog Index, Measure of Gobbledygook, Coleman-Liau Index, and Automated Readability Index as outcome measures, alongside other metrics such as the number and frequency of abbreviations, word frequencies, co-occurrences between frequent words, and note length in characters. Most studies (1/15, 7%) found changes in clinical notes after the implementation of ORA; however, there were also results where clinical notes remained unaffected, often for the same outcome measures.

Reading Ease and Complexity

In total, 3 studies examined changes in the reading ease and complexity of clinical notes [23,45,47]. Blok et al [45] and Kind et al [47] reported a decrease in reading ease and an increase in complexity, while Rahimian et al [23] found the opposite trend: a significant decrease in complexity and an increase in readability. Changes in reading ease were primarily observed in the clinical progress notes. In contrast, other note types, such as the initial notes and letters [23] or the assessment and plan sections [47], seemingly did not change.

Both reading ease and complexity are affected by the use of acronyms and clinical abbreviations [45,47]. While Blok et al [45] found no objective changes in the number and frequency of abbreviations used before and after ORA implementation, Kind et al [47] reported a significant increase in the rate of acronyms and abbreviations in clinical notes. However, the absolute rate of acronyms and abbreviations remained low at approximately 3%.

Length

When discussing the readability and complexity of the notes, studies often refer to note length as a relevant moderator. Four of the included studies [23,24,45,46] examined objective changes in note length because of ORA implementation. Two studies found a statistically significant increase in the average length of progress notes [23,45]. In contrast, Holmgren and Apathy [46] observed a brief, nonsignificant increase in note length of 27.3 characters after ORA implementation, followed by a statistically significant downward trend in note characters over the long term. These changes mainly appear to apply to unstructured visit notes and are not present in the medical history or physician's letters [23,45]. In addition, Rahimian et al [24] used n-grams to identify provider-specific variation in note length. N-grams (linguistics) are sequences of "n" items from a given sample of text or speech, where the items can be phonemes, syllables, letters, words, or base pairs, depending on the application. Rahimian et al [24] found that the number of significant n-grams remained consistent across providers.

Content and Emotional Tone

When it comes to the content of clinical notes, the objective evidence is inconclusive. Typically, only small to moderate changes in the content of clinical notes are reported when shared with patients via web-based patient portals [40,47]. Rahimian et al [24] and Jain et al [40] both use visual graphing of words used and their relatedness to analyze objective content changes in clinical notes pre- and post-ORA: Before ORA, words such as "follow," "well," and "disease" were most prominent, while after ORA, the words "old" and "well" became more prevalent [40]. According to Rahimian et al [24], the use of words such as "distress," "concerning for," and "follow-up" decreased in clinical notes post-ORA, while the word "improving" became more prevalent after ORA was implemented. According to an analysis by Alpert et al [49] using Linguistic Inquiry and Word Count, the emotional tone of the notes remained unchanged. With regard to mental health issues, an increase (pulmonology) or decrease (rheumatology) in notes mentioning mental health status was observed depending on the medical department [47].

XSL•FO
RenderX

Subjective Changes in Documentation

Overview

The included studies with qualitative survey components (ie, pure qualitative or mixed methods studies) primarily used qualitative interviews for data collection (18/26, 69%), followed by qualitative analysis of written free-text responses (3/26, 12%) to assess perceptions of changes to documentation. Some studies (3/26, 12%) conducted focus groups. In some studies (6/26, 23%), findings about documentation changes draw upon a broader dataset, while in others (20/26, 77%), they are supported only by individual quotes from practitioners. All of the studies with qualitative components (26/26, 100%) report changes in clinician-perceived note characteristics (such as length, comprehensibility, accuracy, objectivity, tone, and wording) or changes in content (candor, quality, and tacit knowledge) and functional aspects (notes as a working tool) due to ORA implementation. In contrast, in a few studies (6/26, 23%), some clinicians reported experiencing no changes in clinical documentation due to the introduction of ORA [38,49,54,55,57,61].

Linguistic Characteristics

Length

Similar to the objective changes, some clinicians reported a change in the length of their notes. In 2 studies, clinicians reported that their notes were lengthened by additional text [59] and clinically irrelevant information [42]. However, there were also reports suggesting the opposite: notes became shorter and more concise with the introduction of ORA, partly because hypothetical information was no longer included [68] (refer to the Loss of Tacit Knowledge section) and partly because of concerns about reputational risk [13] (refer to the Candor and also Influences on Documentation Practices section).

Comprehensibility

The comprehensibility of clinical notes appears to be related to their length. While a concise but accurate clinical note is easy for HCPs to understand and work with, patients often require longer explanations in clinical notes to fully understand their content. For example, Alpert et al [49] describe the challenge of composing notes that are both more comprehensible and less intimidating for patients (incorporating more paraphrases and explanations), while also limiting the number of words to ensure their usability in clinical practice. In many of the included studies, HCPs report being more aware and considerate in their writing after implementing ORA, with the goal of creating clinical notes that are more comprehensible and understandable [42,63,65,66]. To achieve this, clinicians reported writing in a clear, concise, and simple manner while avoiding ambiguous terminology [49,50,62,66-68]. Several qualitative studies reported that HCPs modified their use of medical terminology following the implementation of ORA, either by avoiding, reducing, or adapting it to make it more understandable to patients [20,42,49,62,63,68]. In addition, in 2 studies, HCPs reported a decrease in the use of abbreviations and Latin expressions [55,67]. In a recent study conducted by Keuper et al [41], 40.1% of clinicians reported adopting lay language after ORA implementation. Two studies reported a reduction in medical terminology, but HCPs expressed reluctance to completely eliminate it, recognizing its importance for communication with colleagues [66,68]. One study outlined that the described linguistic adjustments peaked shortly after the implementation of ORA but gradually decreased over time, resulting in clinicians reporting a return to their previous documentation practices [65]. Furthermore, 2 studies reported that some clinicians did not observe any effects of ORA on their writing style and continued to use medical terminology to ensure the integrity of clinical documentation [4,65].

Accuracy

Some qualitative studies (8/26, 31%) suggest that clinicians strive to document more accurately, factually, formally, precisely, and objectively when sharing clinical notes with patients [3,20,41,49,66-68], partly to mitigate potential misunderstandings with patients [20]. Other practitioners experience that the implementation of ORA leads to clinical documentation that is less accurate, less direct, and less objective, as the lack of correct medical terminology hinders the accurate communication of information to other providers while potentially introducing irrelevant clinical data [20,42,63]. Some studies (6/26, 23%) report that note accuracy after ORA implementation can be ensured by clinicians being more mindful in their documentation practices [3,20,41,49,66,68].

Mindful Tone and Wording

HCPs from 15 studies reported adjusting the tone and wording of clinical notes when patients were able to read them [3,4,14,17,39,42,49,51,52,54,55,57,58,63,66]. Many professionals reported being more careful and cautious in terms of tone and word choice when writing their clinical notes after the implementation of ORA [3,14,49,66]. They reported making efforts to write notes in a professional and respectful manner, acknowledging the patients' identity and experience [58,66]. In addition, the implementation of ORA reportedly led to perceived changes in how sensitive clinical and social information was documented [42]: some clinicians reported refraining from using language that patients might perceive as critical, provocative, or offensive to avoid upsetting or angering them [39,42,51,54,55]. For instance, clinicians reported that they avoided using subjective terms such as "troubled," "difficult," "disruptive," or "noncompliant" to describe patients' conditions [4,42]. Potentially stigmatizing or hurtful descriptions (eg, obesity and mental health issues) were often reportedly excluded from the clinical notes or paraphrased [4,42,54,55]. However, Alpert et al [57] noted that even after the implementation of ORA, derogatory terms such as "fat" were still present in the clinical notes, causing distress to the patients. In 3 studies, practitioners emphasized that the adoption of patient-friendly, sensitive documentation (after the implementation of ORA) compromised the quality of the clinical notes, as described in more detail in the following section [4,17,52].

Note Content and Functionality

Candor

In many of the included qualitative studies, clinicians reported being less candid in their documentation or omitting information

XSL•FO
RenderX

from the clinical notes. Percentages of clinicians who stated they were less candid in their documentation ranged from 15% to 52% [17,19,42,48,51,53-56,60,62,69]. Two studies showed that significantly more professionals expected to be less honest in their documentation due to ORA than they were when actually sharing their notes with patients [48,69].

Many clinicians reported becoming more selective in what they write [17,43,64], exercising greater caution in deciding what information to include [50,68], and reducing unnecessary detail in their clinical notes [14,39,60,66]. Numerous studies reported that clinicians tended to "safeguard" sensitive and potentially harmful information (eg, domestic violence) to protect patients from potential negative consequences. This may result in information being described more generally, censored, or hidden altogether from the clinical note [4,19,38,42-44,48,51,54-56,60-64,66-68]. In some cases, practitioners blocked patients' access to their clinical notes [19,38] or used parallel documentation, referred to as "shadow records." Shadow records refer to unofficial, private documentation maintained by clinicians in various formats, which is kept distinct from shared documentation [14,51]. Because of the reported changes in the level of candor, detail, and information included in the documentation, some clinicians criticized the decline in the quality and effectiveness of clinical notes [4,17,62,63].

Fewer clinicians reported that the ORA has led to more detailed and open documentation, typically to avoid patient complaints about missing information [3,65]. However, some studies suggest no change in content and candor: clinicians were already writing openly, honestly, and respectfully before ORA was implemented, perceiving that no further changes were required when records were shared with patients [3,4,58,63,68].

Loss of Tacit Knowledge

Some clinicians reported that they excluded certain information—often sensitive or not yet confirmed—from the clinical notes. Clinicians primarily avoided tentative differential diagnoses (especially in cases of suspected serious illness) [17,38,42,62,63,68]. However, they also reported withholding their own assessments and observations [17,54,62]. This included hypotheses, concerns, "gut feelings," and speculative information that might be helpful to the next clinician treating the same patient [17,20,62,68]. In doing so, clinicians may be striving to avoid causing misunderstanding and confusion [17,20,63,68] or arousing fear and offense in patients [13,42,62], especially when there is no time or opportunity to thoroughly discuss the preliminary information with the patient [20,68].

Value as a Working and Therapy Tool

The changes described in the content and structure of clinical notes resulting from the use of ORA also affect how clinical notes may be used as work and therapy tools. Some clinicians argued that the clinical notes are no longer effective as a work tool and do not adequately serve their professional purpose when shared with patients. This is attributed to the omission of preliminary or sensitive information and a shift toward more descriptive but less rigorous documentation [17,62,63]. Some studies suggested that clinicians prefer to use clinical notes primarily in the traditional sense, either as a communication tool with colleagues and other providers [4,42,67,68] or as a work tool for themselves [4,38,62]. According to several clinicians, the changes in documentation caused by ORA hinder interdisciplinary communication within the team [42,62,64] as well as the personal use of notes (eg, as a personal reminder) [4,38,62], making the clinical note a less efficient and valuable working tool [39,42,69].

Other practitioners saw the adaptation of their writing as a new opportunity to expand the function of clinical notes as a therapeutic tool. By communicating directly with the patient through clinical notes, they could better engage patients by including self-care instructions or different types of reminders [42,63], highlighting the patient's strength and progress, and reassuring the patient that their perspective is heard and understood [66]. The notes were used to emphasize important aspects of the consultation, clarify goals, and provide educational resources [66].

Influences on Documentation Practices

Two studies found that HCPs may feel uncertain or vulnerable when writing shared documentation [54,60]. Aiming to protect patients from adverse outcomes such as misunderstanding [20,49,63,66], anxiety, and confusion [17,42,63,64,68], they reported being more careful and guarded in their writing [17,49,51,63,67]. This was seemingly fueled by their desire to avoid being perceived as harsh, critical, or judgmental in their documentation [39,42,49]. In addition, clinicians were aware of the potential reputational risks, medicolegal concerns, and patient safety consequences associated with ORA, which accordingly influenced their documentation practices [13,20,60].

Finally, it should be noted that clinicians write with several target audiences in mind when ORA is available. HCPs were usually more considerate of the patient's reception of the note, while still documenting appropriately for colleagues, health insurers, billing, and other stakeholders [3,4,17,42,44,49,51,52,56,57,59-62,65,66,68]. They must also consider the possibility of the so-called secret readers, such as relatives or caregivers, who may have access to the clinical notes without the clinician's knowledge [3,42,52,60-62,64]. Therefore, the introduction of the ORA presents practitioners with the major challenge of writing a uniform note to serve multiple and sometimes conflicting needs.

Discussion

Principal Findings

This scoping review is the first to assess the potential for changes in documentation following patient ORA. Most of the included studies, which incorporated qualitative components, report changes in clinician-perceived note characteristics, such as length, comprehensibility, accuracy, objectivity, tone, and wording. In addition, these studies highlight changes in content, such as candor and quality as well as functional aspects, such as the role of notes as a working tool, following the implementation of ORA. Conversely, a minority of studies indicated that some clinicians reported no discernible changes in clinical documentation following the introduction of ORA [38,49,54,55,57,61]. Similar to the objective changes noted,

XSL•FO
RenderX

some clinicians observed alterations in the length of their notes. In 2 studies, clinicians reported that their notes were longer [42,59]. However, contrasting subjective findings indicate the opposite effect [68]. Regarding the content of free-text notes, the objective evidence remains inconclusive. Generally, only minor to moderate alterations in the content of clinical notes are reported when shared with patients via the web [40,47]. Moreover, several studies using literacy metrics investigated changes in the readability and complexity of clinical notes. While Blok et al [45] and Kind et al [47] reported an increase in complexity and a decrease in readability, Rahimian et al [23] found a significant decrease in complexity and an increase in readability.

The mixed findings reported may reflect the diverse perspectives and experiences of the different individuals involved. Just as people communicate differently, they also approach documentation in unique ways, influenced by their background, experiences, and perceptions of transparency, privacy, and integrity. Various factors, such as personal attitudes, power dynamics, and professional habits, might further contribute to the variation in documentation practices. Given this diversity, the implementation of ORA will inevitably impact health care staff's documentation practices differently. Some may view it as an opportunity to extend their caregiving beyond face-to-face interactions, using clinical notes as a therapeutic tool to better engage with patients [42,63], highlighting progress, clarifying goals, and offering reassurance through ORA [66]. Contrastingly, others grapple with ethical dilemmas regarding the inclusion or omission of certain information. Practitioners in 15 studies indicated that they modified the tone and language used in clinical notes when patients had access to them [3,4,14,17,39,42,49,51,52,54,55,57,58,63,66]. Following the introduction of ORA, numerous professionals adopted a more deliberate and cautious approach toward tone and language selection in clinical notes [3,14,49,66].

This variance in approaches underscores the need for flexibility and understanding within the health care setting. In essence, the following question arises: can a one-size-fits-all approach accommodate the diverse needs and perspectives present in health care? The answer remains uncertain, highlighting the complexities of standardizing practices across such a multifaceted sector. As the language, format, and content of clinical notes may evolve following ORA, assessing whether such changes yield benefits or pose risks is imperative.

O'Neill [70] contests the common belief that transparency and truth are inherently linked, asserting instead that they are fundamentally at odds. O'Neill [70] argues that transparency compels document writers to obscure genuine information or motives, crafting content deemed suitable for public consumption and thereby fostering deception. Worsening the argument by O'Neill [70], Nguyen [71] adds that the pressure to conform to public expectations through transparency may lead experts to compromise the integrity of their documentation, abandoning nuanced insights, tacit knowledge, and expert judgment. In several of the included qualitative studies, clinicians did express being less candid in their documentation or selectively omitting information from clinical notes due to patient access. Reported percentages of clinicians admitting to reduced candor in their documentation varied from 15% to 52% [17,19,42,48,51,53-56,60,62,69]. On the contrary, 2 studies found that more professionals expected to be less honest in their documentation due to ORA than those who actually were when sharing their notes [48,69]. Adding weight to the argument by Nguyen [71], practitioners experience that ORA implementation results in clinical documentation that is less accurate, less straightforward, and less objective; the absence of correct medical terminology hinders effective communication with other providers and possibly introduces irrelevant clinical data [20,42,63].

The increase in readability and note length following the implementation of ORA also suggests that HCPs may have started documenting in a manner they perceived to be more understandable or accessible to a broader audience, such as patients [23]. Numerous clinicians indicate increased selectiveness in their documentation [17,43,64], exercising caution in determining the information to incorporate [50,68], and trimming unnecessary details from their clinical notes [14,39,60,66]. In certain instances, practitioners went as far as to restrict patients' access to their clinical notes [19,38] or resorted to parallel documentation, also known as a "shadow record" [14,51]. Restricting patients' access to parts or all of their EHR may be justifiable in certain situations (such as suspected coercive access in the context of domestic abuse) and may therefore be acceptable clinical practice [72]. The reported changes in the level of candor, detail, and information in the documentation led some clinicians to critique a decline in the quality and effectiveness of clinical notes [4,17,62,63].

Fewer clinicians noted that ORA implementation resulted in more detailed and transparent documentation, often aiming to prevent patient complaints about missing information [3,65]. However, contrary to the theory of transparency and deception by O'Neill [70], certain studies suggest no alteration in content and candor due to ORA: clinicians had already been writing openly, honestly, and respectfully before the implementation of ORA; thus, no adjustments were deemed necessary when records were shared [3,4,58,63,68].

Numerous qualitative studies indicate that clinicians aim for greater accuracy, formality, precision, and objectivity when sharing clinical notes with patients [3,20,41,49,66-68]. In total, 6 studies indicate that clinicians can enhance note accuracy following ORA implementation by adopting more mindful documentation practices [3,20,41,49,66,68]. We can assert that ORA offers both advantages and disadvantages. They potentially enhance patient care while simultaneously posing risks regarding what information is omitted.

Extended Care Through Documentation

Patient ORA satisfies the moral argument that the information belongs to the patient. It is even argued that when patients feel in more control of their care, they will take better care of themselves [73]. While it is acknowledged that all patients, including those with mental illness, have the right to access information about their health, this raises ethical questions about the implications of transparency in health care documentation. Here lies a definite complexity in the concept of care and conducting it ethically. Can ORA completely align with the

XSL•FO
RenderX

medical "do no harm" principle? An often-overlooked risk linked to providing patients access to their medical records is the so-called secret readers—individuals other than the patients themselves who may access their medical records for both positive and negative reasons [3,42,52,59-62,64]. The possibility of secret readers may prompt physicians to self-censor in their documentation as a means of additional care and protection of patients considered vulnerable, such as children in families affected by domestic violence. Numerous studies have documented that clinicians frequently take measures to safeguard sensitive and potentially harmful information, such as instances of domestic violence, to shield patients from potential negative repercussions. Therefore, this often leads to describing information in more general terms or entirely concealing it from the clinical note [4,19,38,42-44,48,51,54-56,60-64,66-68]. Confidentiality in these fragile circumstances is paramount due to the potential escalation of abuse if a perpetrator discovers unwanted disclosure. ORA, therefore, heightens concerns about coercion and privacy breaches concerning issues of domestic violence and abuse. In established patient portals, such as the Swedish portal, safeguards have been implemented by excluding certain specifically tagged keywords, such as "risk for domestic violence," from ORA [74].

To Document or Not

When clinicians choose to exclude specific information, particularly sensitive or unconfirmed details such as provisional differential diagnoses [17,38,42,62,63,68] from shared clinical notes, it can be viewed as a form of practical care, aiming to prevent unnecessary worry or distress for the patient. Alongside these caregiving perspectives of omitting information due to ORA, a serious downside can arise when clinicians frequently refrain from including subjective assessments and observations [17,54,62]. This encompasses hypotheses, concerns, "gut feelings," and speculative information that could help the next clinician treating the same patient [17,20,62,68]. In a health care context with high staff turnover, omitting information from a shared working document, although with the best intentions of care for the patient, could inevitably put the same patient at risk. Changes to documentation that are found to cause errors and lead to patient harm could also place clinicians at increased risk of malpractice [75]. By omitting this often tacit knowledge, clinicians aim to prevent misunderstanding and confusion [17,20,63,68] as well as avoid instilling fear and offense in patients [13,42,62], particularly when there is insufficient time or opportunity to thoroughly discuss preliminary information with the patient [20,68].

Research indicates that HCPs often experience feelings of insecurity and vulnerability when composing shared documentation [54,60], striving to shield patients from potential adverse outcomes, including misunderstanding [20,63,66], anxiety, and confusion [17,42,63,68]. Consequently, they tend to be more cautious and guarded in their writing [17,51,63,67]. This caution is further fueled by their desire to avoid being perceived as overly harsh, critical, or judgmental in their documentation [39,42].

Changing Function of the EHR

The traditional function of the medical record has never been to inform, empower, or engage the patient; rather, its primary purpose has been to serve as a tool for HCPs to document clinical information. Several clinicians argue that the alterations in documentation induced by ORA hinder interdisciplinary communication within the team [42,62] and diminish the personal utility of notes (eg, as a personal reminder) [4,38,62], thereby reducing the effectiveness of clinical notes as a working tool. Certain clinicians contend that clinical notes are ineffective as a professional tool and fail to fulfill their intended purpose. This is attributed to the omission of preliminary or sensitive information and a transition toward more descriptive but less precise documentation [17,62,63]. Training on how to write clinical notes in the context of ORA could help HCPs navigate the changing role of the EHR. A recent survey of psychotherapy trainees found that 9 out of 10 believed that education on open notes should be part of the curriculum [76].

In an era of advancing digitization, ORA can be viewed as a burgeoning trend aimed at empowering patients through transparency, with medical documentation increasingly tailored for patients as one of the primary audiences. However, this shift may lead to deviations from the ethical principles that underpin patient care. It is important to recognize that not all aspects of health care can or should be shaped to be pleasing and empowering. Being a patient inherently involves vulnerability and a reliance on the expertise and trustworthiness of HCPs. Patients may feel objectified or reduced to a diagnostic label, seeking recognition of their individuality rather than being treated as a mere statistic or case study. However, in medical records, HCPs are not *describing* an individual; they are documenting a disease, which often causes confusion in a multipurpose document that is interpreted by various audiences. Research shows that clinicians opted not to use language that patients could interpret as critical, provocative, or offensive to prevent causing patients distress or frustration [39,42,51,54,55]. However, being a patient is ultimately intertwined with being vulnerable and distressed; this is essential and part of being in someone's care, which ultimately leads to the patient being forced to trust and live through that vulnerability.

While ORA can facilitate patient engagement and transparency, it also introduces new considerations regarding data privacy, accuracy, and interpretation. Consequently, it is essential to recognize the limitations of digital documentation in fostering trust and prioritize cultivating trust through meaningful patient-HCP interactions. As seen in the included research, one reason clinicians omit preliminary findings in ORAs, in particular, is the lack of time to discuss them thoroughly with the patient [20,68].

Limitations

The studies included in this scoping review have some notable limitations. The search for articles to be included in the review was conducted in the fall of 2023; therefore, the most recent evidence may not be included. Slightly less than half of the studies (19/42, 45%) were based on surveys, and it is unclear whether response bias may have affected the findings. In addition, studies reporting subjective changes in documentation

XSL•FO
RenderX

were more prevalent than objective studies. Furthermore, some studies, particularly those using qualitative interviews, had relatively small sample sizes. Both these factors may affect the generalizability of the results. However, the inclusion of qualitative studies allows a more nuanced and comprehensive understanding of the results, and the studies with small sample sizes are counterbalanced by those with larger sample sizes.

In some studies, such as those conducted by Zellmer et al [56], clinicians were able to selectively choose which notes to share with patients and which to withhold. Assuming that particularly sensitive topics might have been avoided or not shared, this may influence the extent to which the documentation changed, as such topics especially require a patient-friendly adjustment of the clinical note.

Participating clinicians volunteered to participate in the studies and thus volunteered to share their clinical notes with patients. Therefore, it can be assumed that the clinicians had a positive attitude toward ORA or were at least interested in it (self-selection bias); critical voices may be underrepresented. However, as our results show a wide variation, it can be assumed that this is not the case.

Conclusions

While there may be variations in the outcomes and attitudes among clinicians, it is evident that ORA does exert an influence on medical documentation practices. While it may not affect all health care staff uniformly, its effects are palpable for some, potentially influencing health care outcomes. Several included qualitative studies show that HCPs modify their use of medical terminology, either by avoiding, reducing, or adapting it to make it more understandable to patients [20,42,49,62,63,68]. This practical measure could indeed be seen as extended caregiving through documentation. In numerous of the included studies, professionals report being more aware and considerate in their writing after implementing ORA, with the goal of creating clinical notes that are more comprehensible and understandable [42,63,65,66]. Other patient benefits from ORA include clinicians' attempts to write notes professionally and respectfully, acknowledging the patients' identity and experience [58,66].

In contrast to the positive outcomes of documentation changes due to ORA, the presence of compensatory measures, such as parallel documentation [14,51] and restriction of patients' access to their clinical notes [19,38], indicates a systemic issue, suggesting that current practices are not yet functioning optimally. Parallel documentation in this matter refers to shadow recording including unofficial, private documentation maintained by clinicians in various formats, which is kept distinct from shared documentation. Practical compensatory measures underscore deficiencies in the current system, wherein crucial information may be omitted or obscured. If an HCP, for example, withholds information out of fear or concerns regarding reputational risks [13], this might ultimately jeopardize the safety of their colleagues and delay proper patient care. The consequences of physicians omitting information that would benefit their colleagues in differing ways, even if only practiced by a minority, can have a cascading effect on patient care. Given that patient health care journeys are collaborative efforts involving multiple professionals, the impact of 1 note being influenced by ORA extends beyond individual patients and physicians. Such practices, whether they involve avoiding gut feelings or diverting critical information to less formal channels, jeopardize patient safety.

Finally, it should be noted that clinicians must write with multiple audiences in mind when implementing ORA. Clinicians were more considerate of the patient when writing the shared note, while still documenting appropriately for colleagues, health insurers, billing, and other stakeholders [3,4,17,42,44,49,51,52,56,57,59-62,65,66,68]. Therefore, the introduction of ORA presents practitioners with the major challenge of writing a uniform note. Many clinicians expressed a desire for system-level guidance regarding optimal documentation practices to mitigate potential negative outcomes for themselves or their patients. They emphasized the importance of training in recovery-oriented and strengths-focused treatment approaches, shifting from problem-focused thinking to a collaborative relationship with clients [60].

Using ORA effectively without omitting information does present a challenge. Future research must explore practical strategies regarding how ORA can be designed to navigate the complex surroundings of such a vital and multifaceted working document, ensuring ORA does not compromise the efficiency and security of care delivery, patients, or health care staff. To this end, we expect that future scoping reviews and empirical research will focus on the use of generative artificial intelligence in documentation practices [77,78]. There is already evidence that clinicians may be using these tools to assist with writing sensitive, understandable notes that patients will read [79]. It remains unclear whether such practices effectively meet the dual requirements of preserving documentation for clinicians while making the notes more understandable and empathetic for patients [80].

Acknowledgments

The authors would like to thank Malin Barkelind, the librarian at the Uppsala University, for conducting the literature search and performing the deduplication process. BM is funded by a National Institute for Health and Care Research (NIHR) Advanced Fellowship (NIHR300887). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). All studies analyzed during this study are publicly available and can be found via the following databases: APA PsycINFO, CINAHL, PubMed, and Web of Science Core Collection.

Authors' Contributions

EM-D wrote the Methods, Results, and Limitations sections. EM-D created the figures and tables. CL wrote the Introduction and Discussion sections. CB and JS were responsible for the conception of the study and contributed to drafting the manuscript. CB, JS, EM-D, and MH were responsible for conceptualization; all the authors were responsible for data curation and validation; EM-D and CL (screening: all authors) were responsible for the formal analysis; EM-D and JS were responsible for methodology; JS, CB, and MH were responsible for supervision; EM-D was responsible for visualization and writing—review and editing; and EM-D and CL were responsible for writing—original draft. All authors participated in the screening process. All authors read, provided feedback, and approved the paper for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 178 KB - jmir_v27i1e64762_app1.pdf]

Multimedia Appendix 2
Search string.
[PDF File (Adobe PDF File), 125 KB - jmir_v27i1e64762_app2.pdf]

Multimedia Appendix 3
Relevant data from included studies.
[XLSX File (Microsoft Excel File), 47 KB - jmir_v27i1e64762_app3.xlsx]

Multimedia Appendix 4
Mixed Methods Appraisal Tool grading.
[PDF File (Adobe PDF File), 185 KB - jmir_v27i1e64762_app4.pdf]

References

1. Blease CR, O'Neill S, Walker J, Hägglund M, Torous J. Sharing notes with mental health patients: balancing risks with respect. Lancet Psychiatry 2020 Nov;7(11):924-925. [doi: 10.1016/S2215-0366(20)30032-8] [Medline: 32059796]
2. McMillan B, Eastham R, Brown B, Fitton R, Dickinson D. Primary care patient records in the United Kingdom: past, present, and future research priorities. J Med Internet Res 2018 Dec 19;20(12):e11293 [FREE Full text] [doi: 10.2196/11293] [Medline: 30567695]
3. Chimowitz H, O'Neill S, Leveille S, Welch K, Walker J. Sharing psychotherapy notes with patients: therapists' attitudes and experiences. Soc Work 2020 Apr 01;65(2):159-168. [doi: 10.1093/sw/swaa010] [Medline: 32236447]
4. Alpert JM, Morris BB, Thomson MD, Matin K, Geyer CE, Brown RF. OpenNotes in oncology: oncologists' perceptions and a baseline of the content and style of their clinician notes. Transl Behav Med 2019 Mar 01;9(2):347-356 [FREE Full text] [doi: 10.1093/tbm/iby029] [Medline: 29596633]
5. Strudwick G, Clark C, Sanches M, Strauss J. Predictors of mental health professionals' perceptions of patient portals. AMIA Annu Symp Proc 2018;2018:989-997 [FREE Full text] [Medline: 30815142]
6. Blease C, Salmi L, Rexhepi H, Hägglund M, DesRoches CM. Patients, clinicians and open notes: information blocking as a case of epistemic injustice. J Med Ethics 2021 May 14;48(10):785-793 [FREE Full text] [doi: 10.1136/medethics-2021-107275] [Medline: 33990427]
7. Blease C, McMillan B, Salmi L, Davidge G, Delbanco T. Adapting to transparent medical records: international experience with "open notes". BMJ 2022 Nov 21;379:e069861. [doi: 10.1136/bmj-2021-069861] [Medline: 36410770]
8. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
9. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy Technol 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]

XSL•FO
RenderX

10. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient use and experience with online access to electronic health records in Norway: results from an online survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
11. Blease C, Torous J, Hägglund M. Does patient access to clinical notes change documentation? Front Public Health 2020;8:577896 [FREE Full text] [doi: 10.3389/fpubh.2020.577896] [Medline: 33330320]
12. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021 Jun 02;46(2):192-204 [FREE Full text] [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
13. Davidge G, Brown L, Lyons M, Blease C, French D, van Staa T, et al. Primary care staff's views and experience of patients' online access to their electronic health record: a qualitative exploration. Br J Gen Pract 2023 Jun;73(731):e418-e426 [FREE Full text] [doi: 10.3399/BJGP.2022.0436] [Medline: 37068967]
14. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
15. Meier-Diedrich E, Neumann K, Heinze M, Schwarz J. Attitudes and expectations of psychological and medical psychotherapists towards open notes: analysis of qualitative survey responses. Psychiatr Prax 2024 Oct;51(7):376-384. [doi: 10.1055/a-2320-8929] [Medline: 38810902]
16. Schwarz J, Bärkås A, Blease C, Collins L, Hägglund M, Markham S, et al. Sharing clinical notes and electronic health records with people affected by mental health conditions: scoping review. JMIR Ment Health 2021 Dec 14;8(12):e34170 [FREE Full text] [doi: 10.2196/34170] [Medline: 34904956]
17. Blease C, Torous J, Dong Z, Davidge G, DesRoches C, Kharko A, et al. Patient online record access in English primary care: qualitative survey study of general practitioners' views. J Med Internet Res 2023 Feb 22;25:e43496 [FREE Full text] [doi: 10.2196/43496] [Medline: 36811939]
18. Louch G, Albutt A, Smyth K, O'Hara JK. What do primary care staff think about patients accessing electronic health records? A focus group study. BMC Health Serv Res 2022 Apr 29;22(1):581 [FREE Full text] [doi: 10.1186/s12913-022-07954-y] [Medline: 35488233]
19. Petersson L, Erlingsdóttir G. Open notes in Swedish psychiatric care (Part 1): survey among psychiatric care professionals. JMIR Ment Health 2018 Feb 02;5(1):e11 [FREE Full text] [doi: 10.2196/mental.9140] [Medline: 29396386]
20. Turner A, Morris R, McDonagh L, Hamilton F, Blake S, Farr M, et al. Unintended consequences of patient online access to health records: a qualitative study in UK primary care. Br J Gen Pract 2023 Jan;73(726):e67-e74 [FREE Full text] [doi: 10.3399/BJGP.2021.0720] [Medline: 36316163]
21. Schwarz J, Meier-Diedrich E, Neumann K, Heinze M, Eisenmann Y, Thoma S. Reasons for acceptance or rejection of online record access among patients affected by a severe mental illness: mixed methods study. JMIR Ment Health 2024 Feb 05;11:e51126 [FREE Full text] [doi: 10.2196/51126] [Medline: 38315523]
22. Himmelstein G, Bates D, Zhou L. Examination of stigmatizing language in the electronic health record. JAMA Netw Open 2022 Jan 04;5(1):e2144967 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.44967] [Medline: 35084481]
23. Rahimian M, Warner JL, Salmi L, Rosenbloom ST, Davis RB, Joyce RM. Open notes sounds great, but will a provider's documentation change? An exploratory study of the effect of open notes on oncology documentation. JAMIA Open 2021 Jul;4(3):ooab051 [FREE Full text] [doi: 10.1093/jamiaopen/ooab051] [Medline: 34661067]
24. Rahimian M, Warner JL, Jain SK, Davis RB, Zerillo JA, Joyce RM. Significant and distinctive n-Grams in oncology notes: a text-mining method to analyze the effect of OpenNotes on clinical documentation. JCO Clin Cancer Inform 2019 Jun;3:1-9 [FREE Full text] [doi: 10.1200/CCI.19.00012] [Medline: 31184919]
25. Fernández L, Fossa A, Dong Z, Delbanco T, Elmore J, Fitzgerald P, et al. Words matter: what do patients find judgmental or offensive in outpatient notes? J Gen Intern Med 2021 Sep;36(9):2571-2578 [FREE Full text] [doi: 10.1007/s11606-020-06432-7] [Medline: 33528782]
26. Wurster F, Fütterer G, Beckmann M, Dittmer K, Jaschke J, Köberlein-Neu J, et al. The analyzation of change in documentation due to the introduction of electronic patient records in hospitals-a systematic review. J Med Syst 2022 Jul 04;46(8):54 [FREE Full text] [doi: 10.1007/s10916-022-01840-0] [Medline: 35781136]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
29. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
30. Meier-Diedrich E, Davidge G, Hägglund M, Kharko A, Lyckblad C, McMillan B, et al. Changes in documentation due to patient access to electronic health records: protocol for a scoping review. JMIR Res Protoc 2023 Aug 28;12:e46722 [FREE Full text] [doi: 10.2196/46722] [Medline: 37639298]

XSL•FO
RenderX

31. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme. Lancaster University. URL: https://www.researchgate.net/publication/233866356_Guidance_on_the_conduct_of_narrative_synthesis_in_systematic_reviews_A_product_from_the_ESRC_Methods_Programme [accessed 2024-04-29]
32. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, Canadian Committee for Patient Accessible Health Records. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]
33. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
34. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
35. Clarke V, Braun V, Hayfield N. Thematic analysis. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods. Thousand Oaks, CA: Sage Publications; 2015:222-248.
36. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Rev Esp Cardiol (Engl Ed) 2021 Sep;74(9):790-799 [FREE Full text] [doi: 10.1016/j.rec.2021.07.010] [Medline: 34446261]
37. Ålander T, Scandurra I. Experiences of healthcare professionals to the introduction in Sweden of a public eHealth service: patients' online access to their electronic health records. Stud Health Technol Inform 2015;216:153-157. [Medline: 26262029]
38. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
39. DesRoches CM, Leveille S, Bell SK, Dong ZJ, Elmore JG, Fernandez L, et al. The views and experiences of clinicians sharing medical record notes with patients. JAMA Netw Open 2020 Mar 02;3(3):e201753 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.1753] [Medline: 32219406]
40. Jain SK, Rahimian M, Joyce RM, Zerillo JA, Warner JL. Using network graphs to visualize changing documentation styles in an oncology practice before and after OpenNotes implementation. In: Proceedings of the 2017 IEEE Workshop on Visual Analytics in Healthcare. 2017 Presented at: VAHC '17; October 1, 2017; Phoenix, AZ p. 62-68 URL: https://ieeexplore.ieee.org/document/8387502 [doi: 10.1109/vahc.2017.8387502]
41. Keuper J, Batenburg R, van Tuyl L, Verheij R. General practices' experiences with patients' web-based access to medical records: survey study. J Med Internet Res 2023 Apr 07;25:e41832 [FREE Full text] [doi: 10.2196/41832] [Medline: 37027195]
42. Leonard SM, Zackula R, Wilcher J. Attitudes and experiences of clinicians after mandated implementation of open notes by the 21st century cures act: survey study. J Med Internet Res 2023 Feb 28;25:e42021 [FREE Full text] [doi: 10.2196/42021] [Medline: 36853747]
43. Moll J, Cajander Å. Oncology health-care professionals' perceived effects of patient accessible electronic health records 6 years after launch: a survey study at a major university hospital in Sweden. Health Informatics J 2020 Jun;26(2):1392-1403 [FREE Full text] [doi: 10.1177/1460458219881007] [Medline: 31621459]
44. Walker J, Leveille SG, Ngo L, Vodicka E, Darer JD, Dhanireddy S, et al. Inviting patients to read their doctors' notes: patients and doctors look ahead: patient and physician surveys. Ann Intern Med 2011 Dec 20;155(12):811-819 [FREE Full text] [doi: 10.7326/0003-4819-155-12-201112200-00003] [Medline: 22184688]
45. Blok AC, Amante DJ, Hogan TP, Sadasivam RS, Shimada SL, Woods S, et al. Impact of patient access to online VA notes on healthcare utilization and clinician documentation: a retrospective cohort study. J Gen Intern Med 2021 Mar;36(3):592-599 [FREE Full text] [doi: 10.1007/s11606-020-06304-0] [Medline: 33443693]
46. Holmgren AJ, Apathy NC. Assessing the impact of patient access to clinical notes on clinician EHR documentation. J Am Med Inform Assoc 2022 Sep 12;29(10):1733-1736 [FREE Full text] [doi: 10.1093/jamia/ocac120] [Medline: 35831954]
47. Kind EA, Fowles JB, Craft CE, Kind AC, Richter SA. No change in physician dictation patterns when visit notes are made available online for patients. Mayo Clin Proc 2011 May;86(5):397-405 [FREE Full text] [doi: 10.4065/mcp.2010.0785] [Medline: 21531883]
48. Oster NV, Jackson SL, Dhanireddy S, Mejilla R, Ralston JD, Leveille S, et al. Patient access to online visit notes: perceptions of doctors and patients at an urban HIV/AIDS clinic. J Int Assoc Provid AIDS Care 2015;14(4):306-312 [FREE Full text] [doi: 10.1177/2325957414526783] [Medline: 24729072]
49. Alpert JM, Morris BB, Thomson MD, Matin K, Sabo RT, Brown RF. Patient access to clinical notes in oncology: a mixed method analysis of oncologists' attitudes and linguistic characteristics towards notes. Patient Educ Couns 2019 Oct;102(10):1917-1924 [FREE Full text] [doi: 10.1016/j.pec.2019.05.008] [Medline: 31109771]
50. Choi HH, Kotsenas AL, Chen JV, Bronsky C, Roth CJ, Kohli MD. Multi-institutional experience with patient image access through electronic health record patient portals. J Digit Imaging 2022 Apr;35(2):320-326 [FREE Full text] [doi: 10.1007/s10278-021-00565-9] [Medline: 35022926]

XSL•FO
RenderX

51. Johansen MA, Kummervold PE, Sørensen T, Zanaboni P. Health professionals' experience with patients accessing their electronic health records: results from an online survey. Stud Health Technol Inform 2019 Aug 21;264:504-508. [doi: 10.3233/SHTI190273] [Medline: 31437974]
52. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting families to their health record and care team: the use, utility, and impact of a client/family health portal at a children's rehabilitation hospital. J Med Internet Res 2017 Apr 06;19(4):e97 [FREE Full text] [doi: 10.2196/jmir.6811] [Medline: 28385680]
53. Richards S, Carlson K, Matthias T, Birge J. Perception versus reality: does provider documentation behavior change when clinic notes are shared electronically with patients? Int J Med Inform 2021 Jan;145:104304. [doi: 10.1016/j.ijmedinf.2020.104304] [Medline: 33129123]
54. Turvey CL, Fuhrmeister LA, Klein DM, Moeckli J, Howren MB, Chasco EE. Patient and provider experience of electronic patient portals and secure messaging in mental health treatment. Telemed J E Health 2022 Mar;28(2):189-198 [FREE Full text] [doi: 10.1089/tmj.2020.0395] [Medline: 33887164]
55. Wass S, Vimarlund V. Same, same but different: perceptions of patients' online access to electronic health records among healthcare professionals. Health Informatics J 2019 Dec;25(4):1538-1548 [FREE Full text] [doi: 10.1177/1460458218779101] [Medline: 29874962]
56. Zellmer BM, Nacht CL, Coller RJ, Hoonakker PL, Smith CA, Sklansky DJ, et al. BedsideNotes: sharing physicians' notes with parents during hospitalization. Hosp Pediatr 2021 May;11(5):503-508 [FREE Full text] [doi: 10.1542/hpeds.2020-005447] [Medline: 33795371]
57. Alpert JM, Krist AH, Aycock RA, Kreps GL. Applying multiple methods to comprehensively evaluate a patient portal's effectiveness to convey information to patients. J Med Internet Res 2016 May 17;18(5):e112 [FREE Full text] [doi: 10.2196/jmir.5451] [Medline: 27188953]
58. Cajander Å, Moll J, Englund S, Hansman A. Medical records online for patients and effects on the work environment of nurses. Stud Health Technol Inform 2018;247:271-275. [Medline: 29677965]
59. Crucefix AL, Fleming AP, Lebus CS, Slowther AM, Fritz Z. Sharing a written medical summary with patients on the post-admission ward round: a qualitative study of clinician and patient experience. J Eval Clin Pract 2021 Dec;27(6):1235-1242. [doi: 10.1111/jep.13574] [Medline: 33960593]
60. Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK. A qualitative analysis of how online access to mental health notes is changing clinician perceptions of power and the therapeutic relationship. J Med Internet Res 2017 Jun 14;19(6):e208 [FREE Full text] [doi: 10.2196/jmir.6915] [Medline: 28615152]
61. Drinkwater J, Stanley N, Szilassy E, Larkins C, Hester M, Feder G. Juggling confidentiality and safety: a qualitative study of how general practice clinicians document domestic violence in families with children. Br J Gen Pract 2017 Jun;67(659):e437-e444 [FREE Full text] [doi: 10.3399/bjgp17X689353] [Medline: 28137783]
62. Erlingsdóttir G, Petersson L, Jonnergård K. A theoretical twist on the transparency of open notes: qualitative analysis of health care professionals' free-text answers. J Med Internet Res 2019 Sep 25;21(9):e14347 [FREE Full text] [doi: 10.2196/14347] [Medline: 31573905]
63. Grünloh C, Cajander Å, Myreteg G. "The record is our work tool!"-physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167. [doi: 10.2196/jmir.5705] [Medline: 27349531]
64. Moll J, Cajander Å. On patient accessible electronic health records and the experienced effect on the work environment of nurses. Stud Health Technol Inform 2020 Jun 16;270:1021-1025. [doi: 10.3233/SHTI200316] [Medline: 32570536]
65. Muli I, Scandurra I, Cajander Å, Hägglund M. Healthcare professionals' experiences of the work environment after patients' access to their electronic health records - a qualitative study in primary care. Stud Health Technol Inform 2022 May 25;294:530-534. [doi: 10.3233/SHTI220515] [Medline: 35612136]
66. Pisciotta M, Denneson LM, Williams HB, Woods S, Tuepker A, Dobscha SK. Providing mental health care in the context of online mental health notes: advice from patients and mental health clinicians. J Ment Health 2019 Mar;28(1):64-70. [doi: 10.1080/09638237.2018.1521924] [Medline: 30468100]
67. Smaradottir BF. Patient accessible electronic health records: impacts on nursing documentation practices at a university hospital. Stud Health Technol Inform 2018;250:14-18. [Medline: 29857356]
68. Zanaboni P, Kristiansen E, Lintvedt O, Wynn R, Johansen MA, Sørensen T, et al. Impact on patient-provider relationship and documentation practices when mental health patients access their electronic health records online: a qualitative study among health professionals in an outpatient setting. BMC Psychiatry 2022 Jul 28;22(1):508 [FREE Full text] [doi: 10.1186/s12888-022-04123-7] [Medline: 35902841]
69. Ralston JD, Yu O, Penfold RB, Gundersen G, Ramaprasan A, Schartz EM. Changes in clinician attitudes toward sharing visit notes: surveys pre-and post-implementation. J Gen Intern Med 2021 Nov;36(11):3330-3336 [FREE Full text] [doi: 10.1007/s11606-021-06729-1] [Medline: 33886028]
70. O'Neill O. A Question of Trust: The BBC Reith Lectures. Cambridge, MA: Cambridge University Press; 2002.
71. Nguyen CT. Transparency is surveillance. Philos Phenomenol Res 2021 Aug 02;105(2):331-361. [doi: 10.1111/phpr.12823]
72. Safeguarding patients from harm or distress. NHS England Digital. URL: https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/guidance-on-nhs-app-features/online-access-to-gp-health-records/safeguarding-patients-from-harm-or-distress [accessed 2024-06-05]

XSL•FO
RenderX

73. Hibbard J, Gilburt H. Supporting people to manage their health. An introduction to patient activation assets. The King's Fund. URL: https://assets.kingsfund.org.uk/f/256914/x/d5fbab2178/supporting_people_manage_their_health_2014.pdf [accessed 2024-04-29]
74. Petersson L, Erlingsdóttir G. Open notes in Swedish Psychiatric Care (part 2): survey among psychiatric care professionals. JMIR Ment Health 2018 Jun 21;5(2):e10521 [FREE Full text] [doi: 10.2196/10521] [Medline: 29929946]
75. Blease C, Cohen IG, Hoffman S. Sharing clinical notes: potential medical-legal benefits and risks. JAMA 2022 Mar 22;327(8):717-718. [doi: 10.1001/jama.2021.23179] [Medline: 35119468]
76. Kharko A, Buergler S, Bärkås A, Hägglund M, Gaab J, Fagerlund AJ, et al. Open notes in psychotherapy: an exploratory mixed methods survey of psychotherapy students in Switzerland. Digit Health 2024;10:20552076241242772 [FREE Full text] [doi: 10.1177/20552076241242772] [Medline: 38559581]
77. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
78. Blease C, Torous J. ChatGPT and mental healthcare: balancing benefits with risks of harms. BMJ Ment Health 2023 Nov;26(1):51 [FREE Full text] [doi: 10.1136/bmjment-2023-300884] [Medline: 37949485]
79. Blease C, Worthen A, Torous J. Psychiatrists' experiences and opinions of generative artificial intelligence in mental healthcare: an online mixed methods survey. Psychiatry Res 2024 Mar;333:115724 [FREE Full text] [doi: 10.1016/j.psychres.2024.115724] [Medline: 38244285]
80. Blease C, Torous J, McMillan B, Hägglund M, Mandl KD. Generative language models and open notes: exploring the promise and limitations. JMIR Med Educ 2024 Jan 04;10:e51183 [FREE Full text] [doi: 10.2196/51183] [Medline: 38175688]

Abbreviations

EHR: electronic health record
HCP: health care professional
MMAT: Mixed Methods Appraisal Tool
ORA: online record access
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by A Coristine; submitted 25.07.24; peer-reviewed by F Wurster, I Mircheva; comments to author 17.09.24; revised version received 24.09.24; accepted 30.10.24; published 20.02.25.

Please cite as:
Meier-Diedrich E, Lyckblad C, Davidge G, Hägglund M, Kharko A, McMillan B, Blease C, Schwarz J
Impact of Patient Online Record Access on Documentation: Scoping Review
J Med Internet Res 2025;27:e64762
URL: https://www.jmir.org/2025/1/e64762
doi:10.2196/64762
PMID:

©Eva Meier-Diedrich, Camilla Lyckblad, Gail Davidge, Maria Hägglund, Anna Kharko, Brian McMillan, Charlotte Blease, Julian Schwarz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Interventions for Older People Experiencing Homelessness: Systematic Scoping Review

Emily Adams[1], BSc, MPH; Eddie Donaghy[2], BSc, PhD; Caroline Sanders[3], RGN, BA, MSc, PhD; Maria Klara Wolters[4], D.Phil.; Lauren Ng[2], MBBS, MA; Christa St-Jean[2], PhD; Ryan Galan[2], BSc; Stewart William Mercer[1], BSc, MSc, MBChB, PhD

[1]Advanced Care Research Centre, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[2]Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[3]Division of Population Health, Health Services Research, and Primary Care, University of Manchester, Manchester, United Kingdom

[4]Informatics, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Stewart William Mercer, BSc, MSc, MBChB, PhD
Advanced Care Research Centre
Usher Institute
University of Edinburgh
BioQuarter (Gate, 5-7)
3 Little France Rd
Edinburgh, EH16 4U
United Kingdom
Phone: 44 0131 651 7869
Email: stewart.mercer@ed.ac.uk

Abstract

Background: older people experiencing homelessness can have mental and physical indicators of aging several decades earlier than the general population and experience premature mortality due to age-related chronic conditions. Digital interventions could positively impact the health and well-being of homeless people. However, increased reliance on digital delivery may also perpetuate digital inequalities for socially excluded groups. The potential triple disadvantage of being older, homeless, and digitally excluded creates a uniquely problematic situation warranting further research. Few studies have synthesized available literature on digital interventions for older people experiencing homelessness.

Objective: This scoping review examined the use, range, and nature of digital interventions available to older people experiencing homelessness and organizations supporting them.

Methods: The scoping review followed Arksey and O'Malley's proposed methodology, PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, and recent Joanna Briggs Institute guidelines. We searched 14 databases. Gray literature sources were searched to supplement the electronic database search. A narrative synthesis approach was conducted on the included articles, and common themes were identified inductively through thematic analysis.

Results: A total of 19,915 records were identified through database and gray literature searching. We identified 10 articles reporting on digital interventions that had a clearly defined a participant age group of >50 years or a mean participant age of >50 years. A total of 9 of 10 studies were published in the United States. The study design included descriptive studies, uncontrolled pilot studies, and pilot randomized controlled trials. No studies aimed to deliver an intervention exclusively to older people experiencing homelessness or organizations that supported them. Four types of intervention were identified: telecare for people experiencing homelessness, distributing technology to enable digital inclusion, text message reminders, and interventions delivered digitally. Interventions delivered digitally included smoking cessation support, vocational training, physical activity promotion, and cognitive behavioral therapy. Overall, the included studies demonstrated evidence for the acceptability and feasibility of digital interventions for older people experiencing homelessness, and all 10 studies reported some improvements in digital inclusion or enhanced engagement among participants. However, several barriers to digital interventions were identified, particularly aspects related to digital inclusion, such as infrastructure, digital literacy, and age. Proposed facilitators for digital interventions included organizational and peer support.

XSL•FO
RenderX

Conclusions: Our findings highlight a paucity of evaluated digital interventions targeted at older people experiencing homelessness. However, the included studies demonstrated evidence of the acceptability and feasibility of digital interventions for older people experiencing homelessness. Further research on digital interventions that provide services and support older people experiencing homelessness is required. Future interventions must address the barriers older people experiencing homelessness face when accessing digital technology with the input of those with lived experience of homelessness.

Trial Registration: OSF Registries OSF.IO/7QGTY; https://doi.org/10.17605/OSF.IO/7QGTY

(J Med Internet Res 2025;27:e63898) doi:10.2196/63898

KEYWORDS

homeless; technology; digital exclusion; elderly; rough sleeping; digital intervention; older people; homelessness; systematic scoping review; aging; premature mortality; indicators; scoping review; databases; thematic analysis; telehealth; mhealth; ehealth

Introduction

Homelessness is a complex phenomenon, with different conceptualizations making it challenging to establish its prevalence and study its phenomenology and effects. The European Observatory on Homelessness proposed the European Typology of Homelessness and Housing Exclusion (ETHOS) states homelessness can include rooflessness (without a shelter of any kind or sleeping rough), houselessness (with a place to sleep but temporarily in institutions or shelter), living in insecure housing (threatened with severe exclusion due to insecure tenancies, eviction, or domestic violence), and living in inadequate housing (in caravans on illegal campsites, in unfit housing, or in extreme overcrowding) [1]. People experiencing homelessness are thought to encounter “accelerated ageing” relative to the general population [2].

An interplay of health and social deprivation leads to people experiencing homelessness with disproportionately high rates of chronic illness [3] and premature age-adjusted mortality rates [4-6]. In this study, older people experiencing homelessness are defined as people older than 50 years who have experienced chronic/episodic homelessness. Chronic homelessness is associated with accelerated aging that predisposes younger people to geriatric health conditions normally associated with older than 75 years in the general population [7]. Older people experiencing homelessness are largely invisible in research, policy, and practice despite the rapidly increasing rates of this population [8]. In the United States, currently, half of single homeless adults are aged 50 and older, compared with 11% in 1990 [9,10]. Further, forecasts from US cohorts project significant growth in aged homelessness in age groups of 50 years and older and 65 years and older, revealing that much of the impact of the postwar baby boom on the aged homeless population is already well underway [11]. Similarly, in Scotland in 2022, 16% of new homeless applications were made by persons older than 50 years [12]. Consequently, the rapidly growing population of older people experiencing homelessness is of critical public health concern and warrants further research.

Homelessness is inextricably linked to social exclusion as individuals are often marginalized from participating in economic, political, social, and cultural life [13]. Concurrently, older people experiencing homelessness are particularly marginalized in the health care system. Older people experiencing homelessness face multiple barriers to timely and effective care for multiple long-term conditions [14,15]. The recent COVID-19 pandemic resulted in a shift from traditional face-to-face health care delivery toward an expansion in using digital technology for service provision [16,17]. In this study, we define digital interventions as interventions that incorporate using and accessing a digital device.

Digital interventions offer promising opportunities to explore new ways of intervention in harm reduction, well-being enhancement, and health treatment of older people experiencing homelessness [18]. This expansion of digital health care is positive for many; however, it has raised issues of digital inequalities for socially excluded groups, including physical barriers in a lack of access to equipment and educational barriers in not being able to use the technology [19]. This is of particular concern to older people experiencing homelessness as evidence has demonstrated that people older than 50 years experiencing homelessness have a lower prevalence of smartphone and internet access than adults aged older than 65 years in the general public or low-income adults [20]. Paradoxically, digital interventions hold new opportunities for inclusion for older people experiencing homelessness while presenting significant barriers due to unaddressed inhibited accessibility.

Over the past decade, more digital interventions have been used within homeless populations [19,21-23]. There is some existing evidence that older people experiencing homelessness meaningfully engage with technology [20,24]. However, no efforts have been made to synthesize this literature on digital interventions specifically for older people experiencing homelessness. Therefore, we conducted a scoping review to synthesize existing primary data from digital health interventions for older people experiencing homelessness. Our main research question was: what is the use, range, and nature of digital interventions available to older people experiencing homelessness and organizations that support people experiencing homelessness? To answer this research question, our scoping review aims to achieve the following objectives: (1) examine current digital interventions (delivery, implementation characteristics, and context) for people experiencing homelessness and the organizations that support them, (2) examine the use of included digital interventions by older people experiencing homelessness and organizations that support them, and (3) identify the facilitators and barriers for older people experiencing homelessness to inclusion in digital interventions.

The scoping review method was chosen because it provides a systematic, rigorous, and transparent approach to mapping a

XSL•FO
RenderX

field of interest regarding existing research's volume, nature, and characteristics [23]. Given that digital interventions available to older people experiencing homelessness are a rapidly developing area of research, a scoping review is an important first step in informing future research and practice [25].

Methods

Overview

This scoping review used the guidelines of Joanna Briggs Institute's (JBI's) Methodology for Scoping Reviews [26] and follows the methodological framework proposed by Arksey and O'Malley [27] which consists of the following stages: (1) identifying the research question; (2) identifying relevant studies; (3) selecting studies; (4) charting the data; and (5) collating, summarizing, and reporting the results. Our review also complies with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [25]. We first developed a scoping review protocol, including a rationale for conducting the review, the main objectives, search strategy, inclusion and exclusion criteria, and methods for screening and data extraction, that was then piloted and discussed by the research team before finalizing. The final protocol was registered retrospectively in Open Science Framework on May 15, 2023.

Information Sources and Search Strategy

A systematic search strategy was developed in consultation with an expert librarian (RS). The search strategy also adhered to the Peer Review of Electronic Search Strategies (PRESS) guidelines [28]. We systematically searched 15 electronic databases from inception up to 28 July 2023: MEDLINE, Global Health, Cumulated Index to Nursing and Allied Health Literature (CINHAL), SCOPUS, APA PsycInfo, Embase, Academic Search Premier, International Bibliography of the Social Sciences (IBSS), Applied Social Sciences Index & Abstracts (ASSIA), Association for Computing Machinery Digital Library (ACMDL), Institute of Electrical and Electronics Engineers (IEEE), Web of Science, Educational Resources Information Centre (ERIC), and Cochrane library. Policy Commons was used to search for gray literature.

The systematic and comprehensive search strategy consisted of key search terms derived from existing search strings and bespoke for each electronic database. The search terms were as follows: homeless* OR temporary accommodation OR roofless OR unfit hous* OR inadequate hous* OR night shelter OR shelter* OR sofa surf* OR rough sleep* AND information communication technolog* OR cell phone* OR mobile app* OR mobile technolog* OR mobile health OR (m health OR e health OR mhealth or ehealth) OR online OR digital OR (telehealth OR tele health OR telemedicine OR tele medicine OR telecare OR tele care) OR social media OR internet OR (web based OR web-based) OR wearable* OR (Smartphone OR smart phone) OR Mobile phone OR Instant messag* OR (Email or electronic mail or e mail) OR (Smartwatch OR smart watch) OR (WhatsApp OR Instagram OR Facebook OR Telegram OR Signal OR Viber) (note: * indicates a wildcard). The results were combined using Boolean operations and adapted for each database (Multimedia Appendix 2). We also scanned references of the included articles for any relevant studies.

Study Selection

After the publications were retrieved and duplicates removed using Endnote (Clarivate), search results were imported into the Covidence software management system (Veritas Health Innovation) for additional deduplication and screening by multiple reviewers.

As advised by JBI guidelines for conducting scoping reviews [26], the population, concept, and context framework was used to define eligibility. Textbox 1 shows the inclusion and exclusion criteria in line with the population, concept, and context framework and contains additional study elements relevant to the eligibility criteria. The ETHOS definition of homelessness framed the inclusion of participants experiencing homelessness [29]. Organizations supporting older people experiencing homelessness were considered to be any health or social care services or third-sector organizations providing a service to people experiencing homelessness. Due to accelerated aging, those who are 50 and experiencing homelessness are defined as "older" in contemporary research [2,30]. Therefore, only studies that included participants with a mean age of 50 years and older or studies with a clearly defined group of participants older than 50 years were included. The nature and causal pathways of homelessness vary globally [31]. To acknowledge that interventions for people experiencing homelessness will diverge due to social and cultural structures, health systems, and the provision of emergency accommodation [32], the scope of this review will be refined to solely high-income countries as defined by the Organization for Economic Co-operation and Development. This allowed for interventions to be more appropriately compared in the analysis [2]. So as not to exclude any innovative interventions, the definition of "digital intervention" was kept intentionally broad to include digital, web-based, or mobile interventions used by people experiencing homelessness to improve social, health, or prospective outcomes.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria detailing the population, concept, and context framework for defining eligibility criteria for scoping reviews.

Inclusion criteria

- Population: Participants currently absolutely homeless (living in shelters or on the streets) OR Participants currently in unstable housing situations (couch surfing, transiently housed) OR Organizations (health or social care services or third sector) that support the aforementioned participants AND Publications including people experiencing homelessness aged 50 years or older (mean participant age or clearly defined participant age group older than 50 years)
- Concept: Interventions are digital in delivery (web-based, mobile, applications, training, or social media)
- Context: Population in high-income nations or the Organization for Economic Co-operation and Development countries.
- Study type: Primary data presented

Exclusion criteria

- Population: Intervention not specific to homeless populations (eg, intervention for migrants or refugees)
- Concept: No digital intervention described OR Digital interventions solely designed for children or youth experiencing homelessness (younger than 18 years)
- Study type: No primary data presented

Further to the inclusion and exclusion criteria outlined in Textbox 1, throughout the review, publications were excluded for the following: (1) no full text available and (2) no English language version available.

Level 1 screening focused on inclusion criteria based on titles and abstracts, while level 2 screening involved reviewing full-text articles. Four reviewers (EA, LN, CS, and RG) independently screened all titles and abstracts. Reviewers met throughout the screening process to discuss queries and reduce uncertainties. Two reviewers (EA and LN) completed the full-text screening independently, with disagreements resolved by discussion with the reviewers and SWM.

Charting Data and Reporting Results

The selected publications were read, annotated, and entered into a Microsoft Excel spreadsheet. EA and LN piloted the data extraction sheet with 2 of the included studies and then revised it in consultation with ED and SWM. We did not critically appraise the included studies, given that this is not typically an objective of a scoping review and the large research design heterogeneity of the studies reviewed [33].

To summarize the data, where applicable, we conducted a basic numerical analysis, for example, the proportion of participants older than 50 years in each study. Meta-analysis was not feasible due to the necessary inclusion of heterogeneous studies in answering the research question. We used a narrative synthesis approach to organize and present relevant findings. Qualitative data (eg, perceptions of digital interventions, barriers to digital interventions for older people experiencing homelessness, and facilitators to digital interventions for older people experiencing homelessness) were imported to NVivo 12 software (Lumivero) for analysis by EA. This approach is characterized by textual summaries and explanations of findings, which are first synthesized by thematic analysis to explore relationships among studies. Thematic analysis in this context involved iteratively identifying, classifying, and sorting the most important themes and concepts across studies [34]. The core research team (EA, LN, SWM, ED, MW, and CS) applied Braun and Clarke's [35,36] reflexive thematic analysis approach, which involved familiarization with the data, generating initial codes, identifying and refining preliminary themes, reviewing themes collaboratively, and ultimately constructing and defining final themes. This iterative, reflexive process—guided by our own perspectives and the review's research questions—enabled a deeper interpretation of the data, resulting in us developing the reported themes.

Results

Study Characteristics

Our searches yielded 19,915 records. After removing duplicates 18,728 records were title and abstract screened (Multimedia Appendix 3). Ten articles met the inclusion criteria [37-46]. All records included in our review were peer-reviewed studies reporting on digital interventions for older people experiencing homelessness that had a clearly defined age group of participants older than 50 years or the mean participant age was older than 50 years (Figure 1 portrays the adapted PRISMA-ScR flow diagram). This screening criteria resulted in 90.5% (n=5902) of the total participants (N=6557) being older people experiencing homelessness (older than 50 years). In 6 of the studies, participants were veterans experiencing homelessness [38,39,41,42,45,46]. A total of 9 of the 10 studies were conducted in the United States [38-46], and the other was conducted in Hungary [37]. Studies were published between 2013 and 2023, with 4 published after the 2020 lockdowns of the COVID-19 pandemic [37,41,45,46]. The study design included descriptive studies [39,43], uncontrolled pilot studies [37,38,42,44,46], and pilot randomized controlled trials [40,41,45].

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram for the identification of studies via databases [25]. OECD: Organization for Economic Co-operation and Development; PEH: people experiencing homelessness.

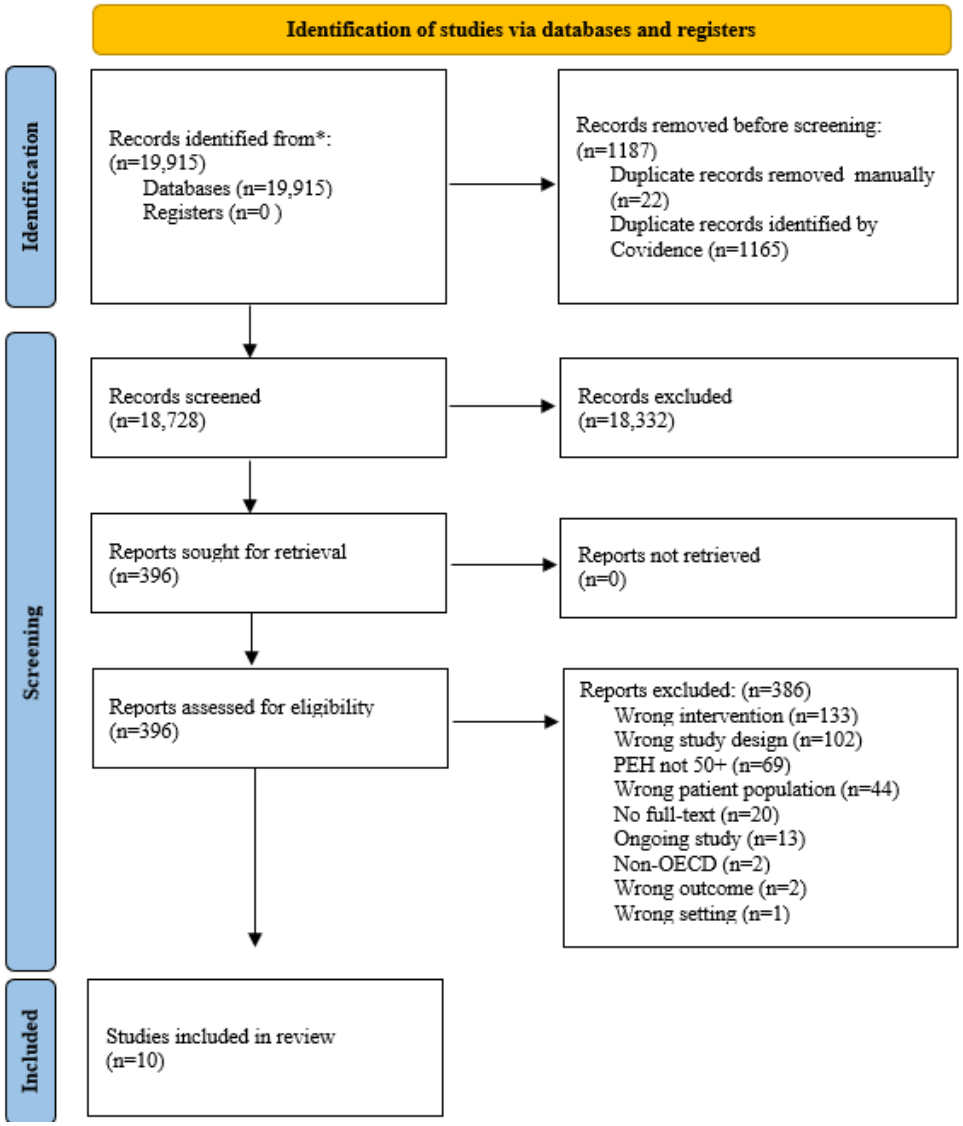

Intervention Characteristics

The included studies were heterogeneous concerning intervention content, delivery, and reported outcomes. Tables 1 and 2 show the intervention delivery and implementation. Interventions could be categorized into implementing telecare for people experiencing homelessness [37,38], distributing technology to enable digital inclusion [39,46], text message reminder interventions [40,42], and 4 interventions delivered digitally [41,43-45] (summarized in Figure 2). Interventions delivered digitally ranged from smoking cessation support [43], vocational training [41], physical activity promotion [44], and cognitive behavioral therapy [45]. We used the broad definition of telecare used by Barlow et al [47], "the use of communications technology to provide health and social care directly to the user ('patient')" where interventions self-described as "telemedicine" [38] and "telehealth" and terms were used interchangeably.

XSL•FO
RenderX

Table 1. Summary of identified study and intervention characteristics.

Author, year	Sample size (% OPEH[a], % Male)	Study design	Intervention	Intervention details
Békási et al, 2022 [37]	75 (100, 76)	Uncontrolled before and after (pre-post) pilot study	Telecare	Participants were invited to 6 web-based telecare visits bi-weekly with a focus on medical management of chronic conditions.
Gabrielian et al, 2013 [38]	14 (100, 71.4)	Uncontrolled mixed methods pilot study	Care Coordination Home Telehealth (CCHT)	CCHT used in-home messaging devices to provide health education and daily questions about clinical indicators from chronic illness care guidelines.
Garvin et al, 2021 [39]	1070 (97.1, 77.9)	Descriptive study	Tablet distribution	Veterans Association distributed tablets to access challenged veterans to be used for any clinical care that does not require physical contact (mental health therapy, medication management, primary care, palliative care, and rehabilitation care).
Kershaw et al, 2022 [40]	62 (100, 85)	Pilot randomized controlled trial (RCT)	Cell phone–based text messaging system	One-way message appointment reminders for upcoming appointments for a range of services and 2-way messages, which requested a text response, asked participants about their mood.
LePage et al, 2023 [41]	27 (81.5, 100)	Pilot developmental study and pilot RCT	Web-based vocational rehabilitation program	Manualized vocational program to aid individuals in identifying work skills, generating examples of those skills, and developing answers to typical questions one might encounter during the interview process.
McInnes et al, 2014 [42]	20 (80, 81)	Uncontrolled before and after pilot study	Text-messaging reminder intervention	Participants were sent 2 text message appointment reminders on a schedule of 5 days and 2 days before their appointment.
Reitzel et al, 2014 [43]	22 (100, 63.6)	Descriptive study	Smartphone-based Smoking cessation	The smartphone was programmed to collect latitude-longitude data via an internal GPS chip at the time the random assessment was prompted.
Rhoades et al, 2019 [44]	13 (100, 46.2)	Uncontrolled before and after pilot study	Cell phone-based physical activity intervention	Intervention to increase physical activity by encouraging walking via goal-setting and motivational text messaging, self-monitoring of walking behavior using pedometers, and providing ongoing feedback on walking performance.
Wilson et al, 2023 [45]	27 (100, 93)	RCT	Telephone-delivered cognitive behavioral therapy (CBT)	Telephone-delivered CBT, tobacco cessation pharmacotherapy, long-term incentives for abstinence-delivered counseling sessions, and optional prescribed tobacco cessation pharmacotherapy.
Wray et al, 2022 [46]	5127 (88, 87.8)	Uncontrolled before and after study	Distribution of video-enabled tablets and cell phones	Program that provided video-enabled tablets to any Veteran who was deemed to have necessary clinical services and a technological need.

[a]OPEH: older people experiencing homelessness.

XSL•FO
RenderX

Table 2. Summary of included digital intervention outcomes for older people experiencing homelessness.

Author, year	Duration (Months)	Follow-up	Outcomes measured	Main Findings
Békási et al, 2022 [37]	3	6 months post	• Feasibility • Patient experience • Medical relevance	• The study provided evidence of a feasible telecare setup in shelters for people experiencing homelessness. • Client satisfaction was high; participants reported similarly high ratings for ease of use and comfort. • Physicians reported the ability to assess the patient's condition properly and make an adequate diagnosis.
Gabrielian et al, 2013 [38]	unclear	not stated	• Program acceptability to staff and consumers • Role of peers to support illness self-management	• Participants were satisfied with CCHT[a]. • Most did not require support from peers to engage in CCHT but valued peer social assistance.
Garvin et al, 2021 [39]	6	not stated	• Tablet adoption and use	• Tablet use was more common among veterans experiencing homelessness who were younger (AOR[b]=2.77; *P*<.001); middle-aged (AOR=2.28; *P*<.001); in rural. • Use was less common among those who were Black (AOR=0.43; *P*<.001) and those with a substance use disorder (AOR=0.59; *P*<.001).
Kershaw et al, 2022 [40]	4	Immediately after (4 months)	• Number of ED encounters • Number of inpatient admissions • Appointment keeping	• No significant differences were found in ED admissions and inpatient or outpatient care between the intervention and control groups. • Appointment no-show rates were 21.0% versus 30.6% (*P*=.08).
LePage et al, 2023 [41]	7 days	6 months post	• The acceptability of the system • The impact of the system	• Veterans found the web-based program as acceptable as a hardcopy manual covering similar material. Participants randomized to the web-based system were more likely to obtain employment than people randomized to the hardcopy manual.
McInnes et al, 2014 [42]	2	Immediately after (2 months)	• Feasibility, effectiveness, and acceptability of text message reminders for people experiencing homelessness	• Participants were satisfied with the text-messaging intervention and had very few technical difficulties. • Patient canceled visits, and no-shows trended downward.
Reitzel et al, 2014 [43]	1	6 months post	• Associations between shelter proximity and real-time effects during a specific smoking quit attempt	• Closer proximity to the shelter was associated with greater negative effects only during the post-quit attempt week (*P*=.008). All participants relapsed to smoking by 1-week post-quit attempt.
Rhoades et al, 2019 [44]	1.5	12 months post	• Physical activity • Acceptability • Wellbeing	• Changes to people's physical activity levels were limited, but participants reported increased quality of life during the intervention period. Interviews revealed that the intervention was well-received and enjoyable for participants.
Wilson et al, 2023 [45]	1.5	3, 6, and 12 months post	• The effectiveness of the intervention on biochemically verified prolonged smoking abstinence	• At 6 months, participants in the mCM group were significantly more likely to meet the criteria for prolonged abstinence (AOR=3.1). Across time points, veterans in the mCM group had twice the odds of prolonged abstinence as those in the standard care group. However, by the 12-month follow-up, no statistically significant group difference in abstinence existed.
Wray et al, 2022 [46]	6	immediately (6 months post)	• Characterize device recipients • Assess in-person, telephone, and video-based engagement patterns across a variety of clinical settings	• Compared with the 6 months prior to device receipt, in the 6 months following receipt, in-person and video engagement increased by an average of 1.4 visits (8%) and 3.4 visits (125%). Tablet users had a substantially more significant increase in video-based engagement (þ3.2 visits [þ110%] vs þ0.9 [þ64%]).

[a]CCHT: Care Coordination Home Telehealth.

[b]AOR: adjusted odds ratio.

XSL•FO
RenderX

Figure 2. Summary of the 10 digital interventions included in this review. Interventions were categorized into implementing telecare for people experiencing homelessness, distributing technology to enable digital inclusion, text message reminder interventions, and 4 interventions delivered digitally. CBT: cognitive behavioral therapy.

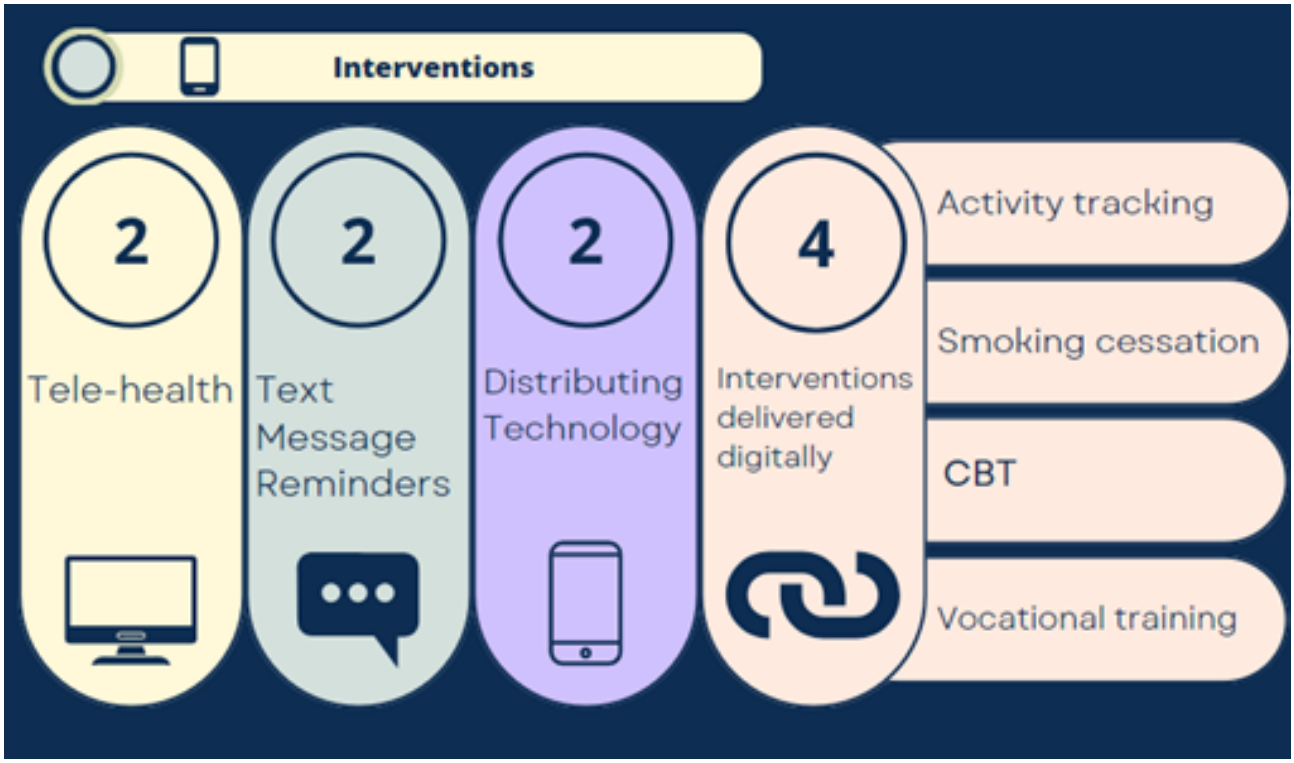

No papers included aimed to develop or deliver interventions specifically for older people experiencing homelessness participants. Additionally, no interventions were found specifically for organizations supporting people experiencing homelessness. Thus, all included studies were developed to be used by individuals experiencing homelessness without specific considerations for older age.

Outcomes From Digital Interventions

Overview

Thematic analysis of study outcomes identified 4 themes: improved digital inclusion, enhanced service engagement and care, no significant outcomes, and unintended consequences from digital interventions (summarized in Textbox 2). Individual intervention outcomes are summarized in Table 2. Barriers and facilitators for digital interventions were synthesized from the included interventions. Finally, the thematic analysis identified the overall feasibility of the included interventions.

Textbox 2. Outcomes from included digital interventions for older people experiencing homelessness and their impact on them.

Improved digital inclusion:

- More equitable access to digital tools by improving the availability of resources [37-39,44,46]
- Improved perceptions of digital intervention [37,39]

Enhancing service engagement and care:

- Improvement in appointment attendance [40,42,46]
- Improvement in medication adherence [42]
- Improvement in perceived quality of care [37,38]

Unintended outcomes:

- Disengagement with in person services [38,46]

Improved Digital Inclusion

The outcomes of 5 included studies demonstrated more equitable access to digital tools for older people experiencing homelessness [37-39,44,46]. First, 2 studies evaluating device distribution programs concluded that tablet distribution offers a model for expanding access to health-related technology and telemedicine [39,46]. In the first study, Garvin et al [39] found that nearly half (45.9%) of veterans experiencing homelessness who received a tablet went on to use the device for video health

XSL•FO
RenderX

consultation within 6 months of receipt. The most frequent tablet use was for tele-mental health support. In bivariate analyses, homeless tablet users were also less likely to have 3 or more chronic conditions (48.7% vs 56.7%; *P*=.006) or to have substance misuse disorder (47.6% vs 58.2%) [39]. The second study was a 6-month evaluation of people experiencing homelessness as digital device recipients. Wray et al [46] found engagement characteristics were similar between those who received a tablet or a cell phone, though fewer individuals with a cell phone had video encounters after receiving a device (45.3% vs 67.4%; *P*<.001), compared with those who received a tablet.

In addition to improving the availability of resources for digital inclusion, 2 studies addressed perceptions or acceptability of the intervention within the population, promoting digital inclusivity [37,39]. For example, during a 12-week telecare pilot for people experiencing homelessness in sheltered housing, Békási et al [37] demonstrated a significant difference in openness to telecare among people experiencing homelessness participants. They found that a group of previously digitally excluded homeless persons found the telecare visits helpful and valuable [37]. Similarly, Gabrielian et al [38] found support for telecare acceptability among homeless veterans with chronic conditions.

Enhancing Service Engagement and Care

Two studies used text message appointment reminders and showed improvement in appointment attendance and medication adherence for people experiencing homelessness [40,42]. In an 8-week pilot intervention period, McInnes et al [42] compared appointment attendance in pre and postintervention periods for a text messaging reminder intervention. They found that twice-weekly text message reminders led to a significant reduction of 30% in patient-cancelled appointments, and "no-shows" (missed appointments) were reduced by 19% [42]. Similarly, when they assessed the feasibility and effectiveness of text messaging to increase outpatient care engagement and medication adherence in people experiencing homelessness in Boston, Kershaw et al [40] recorded positive comments from participants overall. Qualitative findings from the follow-up interviews with intervention group participants showed that text messages functioned as social support. In addition, text messages complemented the participant's lifestyle, and appointment reminders were helpful to ensure attendance [40].

The 2 telecare interventions demonstrated that digital delivery was acceptable to people experiencing homelessness and significantly improved the perceived quality of care [37,38]. Participants in the telecare pilot by Békási et al [37] were present at more than 90% of initially planned appointments, and almost 3-quarters of recruited clients completed the whole course of 6 web-based visits. In postintervention qualitative interviews for a Care Coordination Home Telehealth intervention, participants described telecare as user-friendly and promoted illness self-management [11]. Similarly, Wray et al [46] found that the distribution of "tablets" to veterans improved participants' access to clinical services as it facilitated telecare uptake.

Three digital interventions improved the health and well-being outcomes of people experiencing homelessness [40,44,45]. In a comparative effectiveness trial of digital smoking cessation, Wilson et al [45] reported that veterans in the digitally delivered cognitive behavioral therapy group had twice the odds of prolonged tobacco abstinence compared with the control. Similarly, Rhoades et al [44] identified text messaging and the use of pedometers as a feasible and promising option for improving health and well-being among people experiencing homelessness, as slightly more than half (54%) of participants increased their weekly steps from the beginning to the end of the intervention.

Digital Interventions Demonstrated no Significant Difference

Conversely, 2 studies could not demonstrate any significant difference between control and intervention groups in prestated outcomes for some aspects of digital interventions [40,41]. Kershaw et al [40] aimed to demonstrate outcomes in the text messaging reminders that impacted inpatient care for veterans experiencing homelessness; however, no significant differences were found in an inpatient or outpatient care engagement between the intervention and control groups. Additionally, when stratified by appointment type, there were no significant differences in appointment keeping between intervention and control groups, and estimated effect sizes were small for both appointment types. However, effect sizes for completed appointments and no-shows were slightly larger for physical health appointments than for behavioral health appointments [40]. In a randomized controlled pilot test of employment outcomes, Lepage et al [41] found that the web-based vocational rehabilitation program's control and intervention groups did not differ significantly in the number of modules completed.

Unintended Outcomes of Digital Interventions

Two studies reported unintended consequences of implementing digital interventions [38,46]. Wray et al [46] observed a "substitutive effect"—where telephone-based engagement decreased while in-person and video-based engagement increased at a commensurate rate. Further, compared with those who received a cell phone, those who received a tablet had a smaller increase in in-person (1.3 visits, 8% vs 2.1 visits, 13%) visits and a greater decrease (4.6 visits, 23% vs 1.8 visits, 12%) in telephone visits. They suggest this could negatively impact patients' willingness to engage with in-person care options [46]. This aligns with the findings of Gabrielian et al [38], which reported the unintended negative consequence of participants feeling detached from the technology by the digital delivery of telehealth. In particular, participants felt digital delivery was impersonal.

Barriers and Facilitators to Digital Interventions for Older People Experiencing Homelessness

Overview

Five studies reported age and pre-existing digital exclusion as barriers to people experiencing homelessness in the studied digital interventions [37,39,40,42,46]. Table 3 summarizes the barriers and facilitators identified in the review. Three studies found that the predominant facilitators of digital interventions for people experiencing homelessness were organizational support and peer support [37,39,44].

XSL•FO
RenderX

Table 3. Summary of the reported barriers and facilitators of digital interventions of older people experiencing homelessness.

Barriers	Facilitators
Digital literacy [40,41]	Organization support/technical assistance [37,40,44,46]
Device loss and theft [40]	Charging spots [40]
Internet connection [46]	Peer support [38]
Prohibitive cost (data and minutes) [37,42]	—[a]
Age [39]	—

[a]Not applicable.

Barrier: Digital Exclusionary Factors for People Experiencing Homelessness

Three studies noted participants' difficulties in operating devices and a need for supporting digital literacy. Kershaw et al [40] highlighted digital literacy problems, with 31% (n=19) of participants reporting some technical difficulty, confusion, and usability problems when operating flip phones. Furthermore, Kershaw et al [40] reported loss and theft of mobile devices frequently during the study, requiring numerous replacements given the population is at elevated risk of experiencing theft and limited ability to store devices securely. In distributing tablets to people experiencing homelessness, Wray et al [46] highlighted challenges in maintaining connectivity to the internet and digital literacy problems as factors negatively impacting their experience of such tools. Similarly, McInnes et al [42] noted financial barriers to mobile phone use (eg, running out of minutes).

Barrier: Age

One study reported age as a barrier to the tablet adoption intervention. Following distribution, tablet use was more common among veterans experiencing homelessness who were younger (adjusted odds ratio 2.77; $P<.001$). Garvin et al [39] suggest that older veterans would benefit from simplified user interface designs and digital literacy training to increase comfort, confidence, and willingness to use.

Facilitator: Organizational Support Required

No studies identified in the review focused their intervention on organizations supporting people experiencing homelessness. However, 4 studies note that assistance from support staff or research teams was required in the set-up or duration of the intervention [37,40,44,46]. In the telecare intervention of Békási et al [37], the presence of on-site assistants served as technical support and prevented any misunderstandings regarding medication or referral issues. Rhoades et al [44] gave participants one-on-one assistance with using text messaging as needed. Wray et al [46] preconfigured devices and loaded them with videoconferencing software and mobile apps for participant ease. Kershaw et al [40] conclude that making charging more readily available where homeless persons spend their time could also help reduce theft, such as near bedsides in shelters (inside lockers) and more widely available in libraries, health clinics, food banks, and soup kitchens.

Facilitator: Peer Support

Two studies suggested implementing peer support to enhance adherence and troubleshoot utility issues [38,45]. Gabrielian et al [38] employed peer mentors to conduct visits with veterans to assess relevant psychosocial circumstances and report back to researchers on any equipment/medical concerns. They suggested that peers could significantly break down patient-level barriers to participation in telecare management. Nevertheless, institutional obstacles prevented peer contact with veterans-with 10 of 14 participants opting for adjunctive peer support [38]. Similarly, Wilson et al [45] suggest approaches that integrate peer support into smoking cessation intervention sessions/groups might be beneficial, given previous research indicating that knowing 5 quitters was associated with greater odds of achieving smoking abstinence among homeless smokers.

Potential Viability of Digital Interventions for Older People Experiencing Homelessness

Three included studies concluded that the intervention demonstrated the viability of digital delivery for older people experiencing homelessness [37,39,42]. There was evidence of a feasible telecare setup in shelters offering accommodation to people experiencing homelessness that might support the planning of future telecare services for vulnerable populations [37]. McInnes et al [42] proposed that text messages are a feasible digital intervention as they are an unobtrusive connection to patients, and mobile phones are one of the few communication tools that people experiencing homelessness can attain. Further, they concluded that appointment reminders are greatly needed for this population because they frequently lack the tools that nonhomeless take for granted: reliable mailing addresses, landline phones, wall or computerized calendars, and social supports to remind them of appointments [42].

Discussion

Principal Findings

The scoping review examined the range, nature, and use of digital interventions available to older people experiencing homelessness and organizations that support older people experiencing homelessness. We identified only concerned studies within an Organization for Economic Co-operation and Development context detailing digital interventions with participants who were older and experiencing homelessness. The nature of interventions included digitally delivered primary health care (telecare) [37,39], appointment reminders [40,42], technology distribution [39,46], and well-being interventions

XSL•FO
RenderX

delivered in a digital format [41,43-45]. Included interventions found that common barriers were existing digital exclusion factors for older people experiencing homelessness, such as digital literacy, absence of safe storage for technology, and inconsistent internet connectivity. Suggested facilitators for older people experiencing homelessness in digital interventions were organizational and peer support. The searches found no interventions designed for adoption by support services (as opposed to the older people experiencing homelessness user). Further, it should be noted there was a lack of studies reporting on other dimensions of exclusion (ethnicity, gender, etc) for older people experiencing homelessness. Finally, no included interventions were intended for sole use by a cohort of older homeless adults; subsequently, there were no specific considerations for older people in the intervention design. Therefore, this review highlights the paucity of digital interventions designed for and delivered to older people experiencing homelessness.

Older People Experiencing Homelessness and Digital Engagement

This review demonstrates evidence that digital interventions could benefit older people experiencing homelessness. Consequently, it is crucial to understand the prevalence and use of technology among older people experiencing homelessness to implement digital interventions effectively. People experiencing homelessness access to mobile devices varies greatly. One study showed that as many as 95% had a mobile phone, and 77% reported having a smartphone [48]. However, participants from an ongoing clinical trial at a homeless shelter in Texas reported lower (28.4%) access to an active cell phone. However, 88.6% of participants reported at least weekly internet use, and 77.2% used email [49]. It is well established that older adults in the general population use technological solutions at lower rates than younger adults [50]. In addition, little is known about access to and use of the internet and mobile phones by older homeless adults. Raven et al [20] make one of the only attempts to describe the access to and use of mobile phones, computers, and the internet among 350 homeless adults older than 50 years. They found that most (72.3%) participants owned or had mobile phone access. Participants used phones and the internet to communicate with medical personnel (64.6%), search for housing and employment (30.7%), and to contact their families (82.3%). They concluded that participants had a lower prevalence of smartphone and internet access than adults aged older than 65 years in the general public or low-income adults. Participants faced barriers to mobile phone and internet use, including financial barriers and functional and cognitive impairments [20].

Only one included intervention analyzed age as a variable for use, where tablet use was less likely for older participants [39]. Similarly, in a sample of homeless-experienced adults aged 50 years and older, Raven et al [20] found that almost 3-quarters of participants had current access to a mobile phone. However, participants had a lower prevalence of smartphone and internet access than adults older than 65 years in the general public or low-income adults [20]. This demonstrates that premature aging and complex social challenges that are attributed to homelessness are significant factors in digital exclusion. Concurrently, in the literature, there are significantly more evaluated digital interventions for youth experiencing homelessness (YEH) than those retrieved for this review [49,51-61]. The disparity in tailored interventions for YEH and older people experiencing homelessness further illustrates the widening digital divide for older people experiencing homelessness.

Digital Exclusion as a Barrier

Three included studies highlighted exclusionary factors associated with homelessness, causing barriers for digital interventions [40,42,46]. Technological competency, limited safe storage, and lack of internet connectivity were all referenced in this review as barriers to digital interventions for individuals experiencing homelessness [40,46]. Sieck et al [62] state that digital literacies and internet connectivity have been called the “super social determinants of health” because they address all other social determinants of health. For example, applications for employment, housing, and other assistance programs, are increasingly, and sometimes exclusively, accessible via the web [62]. In their systematic review of technology for people experiencing homelessness, Heaslip et al [19] found that older people experiencing homelessness felt further marginalized by the modern benefits system that “assumes” digital competence and confidence. As participation in most included studies was voluntary, older people experiencing homelessness with a more positive attitude and openness toward telecare might have been overrepresented in the sample, skewing the sample to those more digitally literate [37]. “Access instability” is a term used by Galperin et al [63] to describe their findings that reliable access to electrical power represents a fundamental yet understudied barrier to mobile use. Lacking a safe and reliable place to charge devices, the unstably housed must activate coping strategies that limit digital engagement and constrain use [63]. Overall, this suggests that while digital interventions have the potential to expand inclusion, existing literacy and connectivity barriers must be addressed in tandem with implementation.

Facilitators of Digital Interventions

Despite those barriers, digital interventions can also facilitate access to care. In this review, Garvin et al [39] suggest older veterans would benefit from simplified computer app designs and digital literacy training to increase comfort, confidence, and willingness to engage in their tablet adoption intervention. Concurrently, Sieck et al [62] posit that improving digital literacy skills to access valuable digital tools is key to reducing disparities. In a digital access survey for people experiencing homelessness, Sturman et al [64] found that digital literacy was positively associated with uptake in digital health interventions. Relatedly, Van den Berk-Clark and McGuire [65] argue that the issue of trust between people experiencing homelessness and support services is multifaceted and is influenced by technical competence and the degree to which the individual is made to feel welcome. A study of video consultation suggested that clinicians interacting with homeless-experienced older adults should prioritize addressing the potential skepticism of video calls. Further, it is proposed that clinicians should assess

XSL•FO
RenderX

their access to and knowledge of video conferencing technology [66].

Similarly, 4 studies found that organizational or research team support facilitated the engagement of people experiencing homelessness with digital interventions. Furthermore, the authors identified peer support as a key contributor to participants' comfort with digital interventions [37,40,44,46]. However, Gabrielian et al [38] stated that most participants did not use peer support and highlighted that fostering trust with those providing technology assistance was of primary importance. Similarly, Glover et al [54] found that YEH preferred automated mobile phone functions that avoided interaction with professionals and peers.

This review included 6 studies with interventions conducted with veterans in the United States [38,39,41,42,45,46]. The interventions benefited from the support and infrastructure of the US Department of Veterans Affairs (VA) [67]. It can be assumed that the integration of VA into the health system facilitated the implementation of the interventions. However, additional evaluative research on the context of the implementation is required to determine if internal process barriers nullify any potential facilitatory effects [68].

Implications for Policy and Practice and Future Research

This scoping review highlights several research gaps, upon which we base the following recommendations:

1. The review demonstrated a largely positive view of older people experiencing homelessness's digital interventions however, this needs to be supported by additional empirical evidence of the health and well-being benefits.
2. Additional research is needed to examine older people experiencing homelessness's access to and use of digital tools and interventions.
3. More research is needed on the digital health literacy skills of older people experiencing homelessness and their experiences of using technology to search for and access information and services.
4. Additional evaluation of the implementation infrastructure, for example, the health system deploying a telecare initiative, on the efficacy of an intervention for older people experiencing homelessness
5. Finally, future research should also focus on developing and evaluating digital interventions for older people experiencing homelessness.

The use of digital interventions is a rapidly developing area of practice for professionals with several elements to consider, including:

1. increasing access to technology
2. optimizing technology-based infrastructure,
3. providing training for community outreach and practitioners,
4. engaging service users in the co-design of diverse and contextually sensitive interventions

Due to longstanding digital barriers, implementing digital interventions without addressing older people experiencing homelessness's digital exclusion will likely further damage trust and perpetuate existing poor support service access. As such, services should:

5. systematically assess individual patients' digital literacies,
6. learn about their internet access, and
7. work to address their needs.
8. Partner with community-based organizations with expertise in digital literacy training to address comfortability.

Policy for digital inclusion of older people experiencing homelessness should prioritize free and accessible technology in public settings (eg, shelters, community centers, libraries, and harm reduction centers) and free access to mobile devices. Action is needed across government, public, private, and third-sector organizations to ensure capitalization on the potential for digital interventions to address health and well-being while minimizing the risk of exacerbating existing health inequalities.

Strengths and Limitations

This scoping review has several strengths. To the best of our knowledge, it is the first scoping review to describe the digital interventions for older people experiencing homelessness. We followed the JBI guidance for scoping reviews [33] and our database searching, handling of data, and reporting adhered to published guidelines for undertaking a robust standard scoping review [25,27].

Several limitations should be highlighted. Given the heterogeneity of study methods, we did not systematically assess the quality of studies. Similarly, only English language publications were included due to time and human resources. As there was inconsistency in how papers reported participant age, either a subgroup of participants was clearly defined as older than 50 years or the entire participant group's mean age was older than 50 years. We acknowledge that this inclusion method may generate an incomplete picture of available interventions for older people experiencing homelessness. Most participants in the included studies are veterans based in the United States; therefore, any attempt to generalize these results should be undertaken with caution.

Finally, as this study is not a formal meta-analysis, we did not use more complex statistical pooling methods or analyze the heterogeneity in outcomes reported; as such, our results should be interpreted with these considerations in mind.

Conclusion

Our findings demonstrate the paucity of bespoke digital interventions for older people experiencing homelessness. However, the included studies demonstrate some evidence for the acceptability and feasibility of digital interventions for older people experiencing homelessness. To leverage the potential benefits of digital interventions for older people experiencing homelessness, implementing such interventions will require additional consideration of the multiple exclusionary factors experienced by older people experiencing homelessness. The anticipated increase in the number of older people experiencing homelessness warrants further research on the impact of digital interventions for this vulnerable population.

XSL•FO
RenderX

Acknowledgments

We want to thank Rowena Stewart, academic support librarian at the University of Edinburgh Library, for her contribution to the literature search. We are also grateful for David Henderson's initial contribution to this scoping review, who reviewed the protocol prior to registration. Finally, thanks to Victoria Barber Fleming, who supported the pilot screening. This research was funded by the Legal & General Group (a research grant to establish the independent Advanced Care Research Centre at the University of Edinburgh). The funder had no role in the conduct of the study, interpretation, or the decision to submit for publication. The views expressed are those of the authors and not necessarily those of Legal & General.

Authors' Contributions

EA conceptualized this research, and SWM, ED, CS, and MW contributed to the study design. Search Strategy and searches were completed by EA in consultation with RS. The title and abstract screening were conducted by EA, LN, RG, and CSJ, and full-text screening and data extraction were carried out by EA and LN. The manuscript was drafted by EA. A critical review of the manuscript was undertaken by SWM, ED, CS, and MW. EA is the guarantor, accepts full responsibility for the finished work and the conduct of the study, and has access to the data. All authors approved the final manuscript**.**

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 49 KB - jmir_v27i1e63898_app1.docx]

Multimedia Appendix 2
Search Strings.
[DOCX File , 28 KB - jmir_v27i1e63898_app2.docx]

Multimedia Appendix 3
Database returned results.
[XLSX File (Microsoft Excel File), 10013 KB - jmir_v27i1e63898_app3.xlsx]

References

1. ETHOS typology on homelessness and housing exclusion. FEANTSA. 2012. URL: https://www.feantsa.org/en/toolkit/2005/04/01/ethos-typology-on-homelessness-and-housing-exclusion [accessed 2025-05-22]
2. Mantell R, Hwang YIJ, Radford K, Perkovic S, Cullen P, Withall A. Accelerated aging in people experiencing homelessness: a rapid review of frailty prevalence and determinants. Front Public Health 2023;11:1086215 [FREE Full text] [doi: 10.3389/fpubh.2023.1086215] [Medline: 37006541]
3. Brown RT, Hemati K, Riley ED, Lee CT, Ponath C, Tieu L, et al. Geriatric conditions in a population-based sample of older homeless adults. Gerontologist 2017;57(4):757-766 [FREE Full text] [doi: 10.1093/geront/gnw011] [Medline: 26920935]
4. Aldridge R, Story A, Hwang SW, Nordentoft M, Luchenski SA, Hartwell G, et al. Morbidity and mortality in homeless individuals, prisoners, sex workers, and individuals with substance use disorders in high-income countries: a systematic review and meta-analysis. Lancet 2018;391(10117):241-250 [FREE Full text] [doi: 10.1016/S0140-6736(17)31869-X] [Medline: 29137869]
5. Tweed EJ, Leyland AH, Morrison D, Katikireddi SV. Premature mortality in people affected by co-occurring homelessness, justice involvement, opioid dependence, and psychosis: a retrospective cohort study using linked administrative data. Lancet Public Health 2022;7(9):e733-e743 [FREE Full text] [doi: 10.1016/S2468-2667(22)00159-1] [Medline: 35907410]
6. Gonyea JG, Mills-Dick K, Bachman SS. The complexities of elder homelessness, a shifting political landscape and emerging community responses. J Gerontol Soc Work 2010;53(7):575-590. [doi: 10.1080/01634372.2010.510169] [Medline: 20865621]
7. Brown RT, Kiely DK, Bharel M, Mitchell SL. Factors associated with geriatric syndromes in older homeless adults. J Health Care Poor Underserved 2013;24(2):456-468 [FREE Full text] [doi: 10.1353/hpu.2013.0077] [Medline: 23728022]
8. Canham SL, Humphries J, Moore P, Burns V, Mahmood A. Shelter/housing options, supports and interventions for older people experiencing homelessness. Ageing Soc 2021;42(11):2615-2641. [doi: 10.1017/s0144686x21000234]
9. Culhane DP, Metraux S, Byrne T, Stino M, Bainbridge J. The age structure of contemporary homelessness: evidence and implications for public policy. Anal Soc Iss Public Policy 2013;13(1):228-244. [doi: 10.1111/asap.12004]

XSL•FO
RenderX

10. Hahn JA, Kushel MB, Bangsberg DR, Riley E, Moss AR. BRIEF REPORT: the aging of the homeless population: fourteen-year trends in San Francisco. J Gen Intern Med 2006;21(7):775-778 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00493.x] [Medline: 16808781]
11. Culhane D, Treglia D. The Emerging Crisis of Aged Homelessness. Philadelphia, PA: Actionable Intelligence for Social Policy at the University of Pennsylvania; 2019.
12. Homelessness in Scotland: update to 30 September 2023. Scottish Government. 2023. URL: https://www.gov.scot/publications/homelessness-in-scotland-update-to-30-september-2023/pages/characteristics/ [accessed 2024-05-20]
13. Watson J, Crawley J, Kane D. Social exclusion, health and hidden homelessness. Public Health 2016;139:96-102. [doi: 10.1016/j.puhe.2016.05.017] [Medline: 27340041]
14. Kimbler KJ, DeWees MA, Harris AN. Characteristics of the old and homeless: identifying distinct service needs. Aging Ment Health 2017;21(2):190-198. [doi: 10.1080/13607863.2015.1088512] [Medline: 26404889]
15. Chew-Graham C, O'Toole L, Taylor J, Salisbury C. 'Multimorbidity': an acceptable term for patients or time for a rebrand? Br J Gen Pract 2019;69(685):372-373 [FREE Full text] [doi: 10.3399/bjgp19X704681] [Medline: 31345795]
16. Maretzki M, Geiger R, Buxton JA. How COVID-19 has impacted access to healthcare and social resources among individuals experiencing homelessness in Canada: a scoping review. BMJ Open 2022;12(8):e058233 [FREE Full text] [doi: 10.1136/bmjopen-2021-058233] [Medline: 35914918]
17. Dawes J, May T, Fancourt D, Burton A. The impact of the COVID-19 pandemic and associated societal restrictions on people experiencing homelessness (PEH): a qualitative interview study with PEH and service providers in the UK. Int J Environ Res Public Health 2022;19(23):15526 [FREE Full text] [doi: 10.3390/ijerph192315526] [Medline: 36497601]
18. Calvo F, Carbonell X, Johnsen S. Information and communication technologies, e-Health and homelessness: a bibliometric review. Cogent Psychol 2019;6(1):1631583. [doi: 10.1080/23311908.2019.1631583]
19. Heaslip V, Richer S, Simkhada B, Dogan H, Green S. Use of technology to promote health and wellbeing of people who are homeless: a systematic review. Int J Environ Res Public Health 2021;18(13):6845 [FREE Full text] [doi: 10.3390/ijerph18136845] [Medline: 34202242]
20. Raven MC, Kaplan LM, Rosenberg M, Tieu L, Guzman D, Kushel M. Mobile phone, computer, and internet use among older homeless adults: results from the HOPE HOME cohort study. JMIR Mhealth Uhealth 2018;6(12):e10049 [FREE Full text] [doi: 10.2196/10049] [Medline: 30530464]
21. Burrows R, Mendoza A, Pedell S, Sterling L, Miller T, Lopez-Lorca A. Technology for societal change: evaluating a mobile app addressing the emotional needs of people experiencing homelessness. Health Inf J 2022;28(4):14604582221146720 [FREE Full text] [doi: 10.1177/14604582221146720] [Medline: 36548199]
22. Jiménez-Lérida C, Herrera-Espiñeira C, Granados R, Martín-Salvador A. Attending to the mental health of people who are homeless by mobile telephone follow-up: a systematic review. Healthcare (Basel) 2023;11(12):1666 [FREE Full text] [doi: 10.3390/healthcare11121666] [Medline: 37372784]
23. Polillo A, Gran-Ruaz S, Sylvestre J, Kerman N. The use of eHealth interventions among persons experiencing homelessness: a systematic review. Digit Health 2021;7:2055207620987066 [FREE Full text] [doi: 10.1177/2055207620987066] [Medline: 33598308]
24. Rhoades H, Wenzel SL, Rice E, Winetrobe H, Henwood B. No digital divide? technology use among homeless adults. J Soc Distress Homeless 2017;26(1):73-77 [FREE Full text] [doi: 10.1080/10530789.2017.1305140] [Medline: 31097900]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Peters M, Godfrey CM, Mcinerney P, Munn Z, Trico A, Khalil H. JBI manual for evidence synthesis. JBI. 2020. URL: https://synthesismanual.jbi.global [accessed 2023-03-07]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
29. Edgar B. The ETHOS definition and classification of homelessness and housing exclusion. Eur J Homelessness 2012;6(2):219-225.
30. Rogans-Watson R, Shulman C, Lewer D, Armstrong M, Hudson B. Premature frailty, geriatric conditions and multimorbidity among people experiencing homelessness: a cross-sectional observational study in a London hostel. Hous Care Support 2020;23(3/4):77-91. [doi: 10.1108/hcs-05-2020-0007]
31. Busch-Geertsema V, Culhane D, Fitzpatrick S. Developing a global framework for conceptualising and measuring homelessness. Habitat Int 2016;55:124-132. [doi: 10.1016/j.habitatint.2016.03.004]
32. Embleton L, Lee H, Gunn J, Ayuku D, Braitstein P. Causes of child and youth homelessness in developed and developing countries: a systematic review and meta-analysis. JAMA Pediatr 2016;170(5):435-444 [FREE Full text] [doi: 10.1001/jamapediatrics.2016.0156] [Medline: 27043891]

XSL•FO
RenderX

33. Peters MDJ, Marnie C, Colquhoun H, Garritty CM, Hempel S, Horsley T, et al. Scoping reviews: reinforcing and advancing the methodology and application. Syst Rev 2021;10(1):263 [FREE Full text] [doi: 10.1186/s13643-021-01821-3] [Medline: 34625095]
34. Braun V, Clarke V. Thematic analysis : a practical guide / Virginia Braun and Victoria Clarke. Los Angeles: SAGE; 2022.
35. Braun V, Clarke V, Weate P. Using thematic analysis in sport and exercise research. United Kingdom: Routledge; 2017:213-227.
36. Terry G, Hayfield N, Clarke V, Braun V. Thematic Analysis. In: Willig C, Rogers WS, editors. The SAGE Handbook of Qualitative Research in Psychology. Washington DC: SAGE Publications Ltd; 2017:17-36.
37. Békási S, Girasek E, Győrffy Z. Telemedicine in community shelters: possibilities to improve chronic care among people experiencing homelessness in Hungary. Int J Equity Health 2022;21(1):181 [FREE Full text] [doi: 10.1186/s12939-022-01803-4] [Medline: 36528777]
38. Gabrielian S, Yuan A, Andersen RM, McGuire J, Rubenstein L, Sapir N, et al. Chronic disease management for recently homeless veterans: a clinical practice improvement program to apply home telehealth technology to a vulnerable population. Med Care 2013;51(3 Suppl 1):S44-S51 [FREE Full text] [doi: 10.1097/MLR.0b013e31827808f6] [Medline: 23407011]
39. Garvin LA, Hu J, Slightam C, McInnes DK, Zulman DM. Use of video telehealth tablets to increase access for veterans experiencing homelessness. J Gen Intern Med 2021;36(8):2274-2282 [FREE Full text] [doi: 10.1007/s11606-021-06900-8] [Medline: 34027612]
40. Kershaw K, Martelly L, Stevens C, McInnes DK, Silverman A, Byrne T, et al. Text messaging to increase patient engagement in a large health care for the homeless clinic: results of a randomized pilot study. Digit Health 2022;8:20552076221129729 [FREE Full text] [doi: 10.1177/20552076221129729] [Medline: 36238754]
41. LePage JP, Walters ST, Cipher DJ, Crawford AM. Development and evaluation of an online vocational program for veterans with legal convictions and psychiatric illness. Eval Program Plann 2023;97:102254. [doi: 10.1016/j.evalprogplan.2023.102254] [Medline: 36806008]
42. McInnes DK, Petrakis BA, Gifford AL, Rao SR, Houston TK, Asch SM, et al. Retaining homeless veterans in outpatient care: a pilot study of mobile phone text message appointment reminders. Am J Public Health 2014;104(Suppl 4):S588-S594. [doi: 10.2105/AJPH.2014.302061] [Medline: 25100425]
43. Reitzel LR, Kendzor DE, Nguyen N, Regan SD, Okuyemi KS, Castro Y, et al. Shelter proximity and affect among homeless smokers making a quit attempt. Am J Health Behav 2014;38(2):161-169 [FREE Full text] [doi: 10.5993/AJHB.38.2.1] [Medline: 24629545]
44. Rhoades H, Wenzel S, Winetrobe H, Ramirez M, Wu S, Carranza A, et al. A text messaging-based intervention to increase physical activity among persons living in permanent supportive housing: feasibility and acceptability findings from a pilot study. Digit Health 2019;5:2055207619832438 [FREE Full text] [doi: 10.1177/2055207619832438] [Medline: 30834135]
45. Wilson SM, Blalock DV, Young JR, Griffin SC, Hertzberg JS, Calhoun PS, et al. Mobile health contingency management for smoking cessation among veterans experiencing homelessness: a comparative effectiveness trial. Prev Med Rep 2023;35:102311 [FREE Full text] [doi: 10.1016/j.pmedr.2023.102311] [Medline: 37455761]
46. Wray CM, Van Campen J, Hu J, Slightam C, Heyworth L, Zulman DM. Crossing the digital divide: a veteran affairs program to distribute video-enabled devices to patients in a supportive housing program. JAMIA Open 2022;5(2):ooac027 [FREE Full text] [doi: 10.1093/jamiaopen/ooac027] [Medline: 35991861]
47. Barlow J, Singh D, Bayer S, Curry R. A systematic review of the benefits of home telecare for frail elderly people and those with long-term conditions. J Telemed Telecare 2007;13(4):172-179. [doi: 10.1258/135763307780908058] [Medline: 17565772]
48. Humphry J. Homeless and Connected: Mobile Phones and the Internet in the Lives of Homeless Australians. Australia: The University of Sydney; 2015.
49. Montgomery A, Neil JM, Cannell MB, Gonzalez J, Cole A, Ra CK, et al. The prevalence and perceived utility of mobile health technology among recently incarcerated homeless adults. J Technol Behav Sci 2023;8(2):158-166 [FREE Full text] [doi: 10.1007/s41347-023-00308-3] [Medline: 36844754]
50. Wildenbos GA, Peute L, Jaspers M. Facilitators and barriers of electronic health record patient portal adoption by older adults: a literature study. Stud Health Technol Inform 2017;235:308-312. [Medline: 28423804]
51. Lal S, Halicki-Asakawa A, Fauvelle A. A scoping review on access and use of technology in youth experiencing homelessness: implications for healthcare. Front Digit Health 2021;3:782145 [FREE Full text] [doi: 10.3389/fdgth.2021.782145] [Medline: 34901927]
52. Taylor A, Grigor I, Taylor A, Moore I. HasAnswers: development of a digital tool to support young people to manage independent living. 2019 Presented at: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; 2019 May 02; Glasgow, Scotland Uk p. 1-6. [doi: 10.1145/3290607.3312840]
53. Chavez LJ, Kelleher K, Slesnick N, Holowacz E, Luthy E, Moore L, et al. Virtual reality meditation among youth experiencing homelessness: pilot randomized controlled trial of feasibility. JMIR Ment Health 2020;7(9):e18244 [FREE Full text] [doi: 10.2196/18244] [Medline: 32969834]

XSL•FO
RenderX

54. Glover AC, Schueller SM, Winiarski DA, Smith DL, Karnik NS, Zalta AK. Automated mobile phone-based mental health resource for homeless youth: pilot study assessing feasibility and acceptability. JMIR Ment Health 2019;6(10):e15144 [FREE Full text] [doi: 10.2196/15144] [Medline: 31605516]
55. Linnemayr S, Zutshi R, Shadel W, Pedersen E, DeYoreo M, Tucker J. Text messaging intervention for young smokers experiencing homelessness: lessons learned from a randomized controlled trial. JMIR Mhealth Uhealth 2021;9(4):e23989 [FREE Full text] [doi: 10.2196/23989] [Medline: 33792551]
56. Leonard NR, Casarjian B, Fletcher RR, Praia C, Sherpa D, Kelemen A, et al. Theoretically-based emotion regulation strategies using a mobile app and wearable sensor among homeless adolescent mothers: acceptability and feasibility study. JMIR Pediatr Parent 2018;1(1):e1 [FREE Full text] [doi: 10.2196/pediatrics.9037] [Medline: 30637376]
57. Jennings L, Lee N, Shore D, Strohminger N, Allison B, Conserve DF, et al. U.S. minority homeless youth's access to and use of mobile phones: implications for mhealth intervention design. J Health Commun 2016;21(7):725-733. [doi: 10.1080/10810730.2015.1103331] [Medline: 27232544]
58. Chao SH, Glover A, Rufa AK, Boley R, Schueller S, Zalta A, et al. 6.2 impact of a smartphone-based intervention on substance use amongst homeless youth. J Am Acad Child Adolesc Psychiatry 2017;56(10):S276-S277. [doi: 10.1016/j.jaac.2017.09.347]
59. Yadav A, Chan H, Jiang AX, Xu H, Rice E, Tambe M. Using social networks to aid homeless shelters: dynamic influence maximization under uncertainty. 2016 Presented at: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems; 2016 May 09; Singapore, Singapore p. 740-748.
60. Blondell AD, Robertson MJ, Brindis CD, Papanastassiou AA, Bradley SJ. Mobile Homeless Youth, Health Disparities, and Access to Care: Using Mobile Phones and Geospatial Technologies to Document Geographies of Risk and Pathways to Care. In: Freeman C, Tranter P, Skelton T, editors. Risk, Protection, Provision and Policy. Geographies of Children and Young People, vol 12. Singapore: Springer; 2017:309-343.
61. Bender K, Begun S, DePrince A, Haffejee B, Kaufmann S. Utilizing technology for longitudinal communication with homeless youth. Soc Work Health Care 2014;53(9):865-882. [doi: 10.1080/00981389.2014.925532] [Medline: 25321934]
62. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
63. Galperin H, Bar F, Nguyen H. The power divide: mobile communication in Los Angeles' skid row. Mobile Media Commun 2020;9(1):30-50. [doi: 10.1177/2050157920932608]
64. Sturman N, Ostini R, Choy M, Cyphers G. Digital health access, uptake, literacy and trust in people with experience of homelessness. Aust J Gen Pract 2024;53(4):220-226 [FREE Full text] [doi: 10.31128/AJGP-2-23-6716] [Medline: 38575543]
65. van den Berk-Clark C, McGuire J. Trust in health care providers: factors predicting trust among homeless veterans over time. J Health Care Poor Underserved 2014;25(3):1278-1290 [FREE Full text] [doi: 10.1353/hpu.2014.0115] [Medline: 25130239]
66. Zahir A, Yip D, Garcia C, Smith AN, Dhatt Z, Duke M, et al. "I Needed for You to See What I'm Talking About": experiences with telehealth among homeless-experienced older adults. Gerontol Geriatr Med 2023;9:23337214231172650 [FREE Full text] [doi: 10.1177/23337214231172650] [Medline: 37168020]
67. Tsai J, Pietrzak RH, Szymkowiak D. The problem of veteran homelessness: an update for the new decade. Am J Prev Med 2021;60(6):774-780. [doi: 10.1016/j.amepre.2020.12.012] [Medline: 33583678]
68. Cusack M, Varley AL, Montgomery AE. Applying implementation science methods to address barriers to employment services offered through the United States department of veterans affairs health care for homeless veterans program. J Veterans Stud 2022;8(1):255-265. [doi: 10.21061/jvs.v8i1.313]

Abbreviations

ACMDL: Association for Computing Machinery Digital Library
ASSIA: Applied Social Sciences Index & Abstracts
CINHAL: Cumulated Index to Nursing and Allied Health Literature
ERIC: Educational Resources Information Centre
ETHOS: European Typology of Homelessness and Housing Exclusion
IBSS: International Bibliography of the Social Sciences
IEEE: Institute of Electrical and Electronics Engineers
JBI: Joanna Briggs Institute
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
VA: Veterans Affairs
YEH: youth experiencing homelessness

XSL•FO
RenderX

Edited by N Cahill; submitted 02.07.24; peer-reviewed by T O'Toole, T Moe-Byrne, B Ward; comments to author 12.11.24; revised version received 26.11.24; accepted 26.11.24; published 21.02.25.

Please cite as:
Adams E, Donaghy E, Sanders C, Wolters MK, Ng L, St-Jean C, Galan R, Mercer SW
Digital Interventions for Older People Experiencing Homelessness: Systematic Scoping Review
J Med Internet Res 2025;27:e63898
URL: https://www.jmir.org/2025/1/e63898
doi:10.2196/63898
PMID:

©Emily Adams, Eddie Donaghy, Caroline Sanders, Maria Klara Wolters, Lauren Ng, Christa St-Jean, Ryan Galan, Stewart William Mercer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Health Interventions to Improve Mental Health in Patients With Cancer: Umbrella Review

Chuhan Zhong[1,2], MD; Xian Luo[1], PhD; Miaoqin Tan[1], MD; Jing Chi[1,2], MD; Bingqian Guo[1,2], MD; Jianyao Tang[1,2], MD; Zihan Guo[1,2], MD; Shisi Deng[1,2], MD; Yujie Zhang[1,2], MD; Yanni Wu[1], PhD

[1]Nanfang Hospital, Southern Medical University, Guangzhou, China

[2]School of Nursing, Southern Medical University, Guangzhou, China

Corresponding Author:
Yanni Wu, PhD
Nanfang Hospital
Southern Medical University
No 1838 Guangzhou Avenue North
Baiyun District, Guangdong Province
Guangzhou, 510515
China
Phone: 86 18818860076
Email: yanniwuSMU@126.com

Abstract

Background: Mental health plays a key role across the cancer care continuum, from prognosis and active treatment to survivorship and palliative care. Digital health technologies offer an appealing, cost-effective tool to address psychological needs.

Objective: This umbrella review aims to summarize and evaluate the available evidence on the efficacy of digital health interventions for improving mental health and psychosocial outcomes for populations with cancer.

Methods: Literature searches were conducted in Embase, PsycINFO, PubMed, CINAHL, the Cochrane Library, and Web of Science from their inception to February 4, 2024. Systematic reviews (with or without meta-analysis) investigating the efficacy of digital health interventions for psychosocial variables in patients with cancer were included. Quality was assessed using the Assessing the Methodological Quality of Systematic Reviews-2 tool.

Results: In total, 78 systematic reviews were included in this review. Among diverse delivery modalities and types of digital interventions, websites and smartphone apps were the most commonly used. Depression was the most frequently addressed, followed by quality of life, anxiety, fatigue, and distress. The qualities of the reviews ranged from critically low to high. Generally, despite great heterogeneity in the strength and credibility of the evidence, digital health interventions were shown to be effective for mental health in patients with cancer.

Conclusions: Taken together, digital health interventions show benefits for patients with cancer in improving mental health. Various gaps were identified, such as little research specifically focusing on older adult patients with cancer, a scarcity of reporting high-precision emotion management, and insufficient attention to other certain mood indicators. Further exploration of studies with standardized and rigorous approaches is required to inform practice.

Trial Registration: PROSPERO CRD42024565084; https://tinyurl.com/4cbxjeh9

(J Med Internet Res 2025;27:e69621) doi:10.2196/69621

KEYWORDS

digital health care services; mental health care; oncology; digital delivery modality; umbrella review; PRISMA

Introduction

Global cancer statistics for 2022 released by the International Agency for Research on Cancer of the World Health Organization indicate that there will be almost 20 million new cases of cancer and approximately 10 million cancer deaths in 2022, and the annual number of new cases of cancer will reach 35 million by 2050, a 77% increase from the 2022 level [1]. Patients of all ages may experience psychological distress at any stage of the cancer continuum, from diagnosis and active treatment to survivorship and palliative care [2-4]. The financial burden, and fear of death, together with prognostic uncertainty,

XSL•FO
RenderX

cause patients to suffer a series of negative emotional experiences [5,6]. The prevalence of psychological distress in patients with cancer (20%) was approximately twice as high as in healthy controls (10.63%) [7]. A meta-analysis of 94 interview-based studies showed that the prevalence of mood disorders in patients with cancer attained 38.2% in the first 5 years after diagnosis, with depression and anxiety being more common, affecting up to 20% and 10% of patients with cancer respectively [8]. What is worse, these symptoms persist despite recovery from cancer [9,10]. Psychological distress has been identified as the sixth vital sign in cancer care [11]. Unmet psychological support needs could adversely affect cancer prognosis, which has been negatively associated with treatment adherence, quality of life, and even impact survival rates [12-14]. However, in traditional health care practice, health care providers often focus on the progression of medical and physical symptoms, while frequently overlooking the psychological needs of patients [7,15]. Moreover, due to the constraints in time, location, and economic costs, in-person psychosocial support is difficult to consistently reach the majority of people in need [16]. Accordingly, there is an urgent need for more accessible, cost-effective, and widespread approaches to the early identification and treatment of psychological distress in patients with cancer [17].

Recently, clinicians and patients have been increasingly inclined to opt for digital delivery models [18]. Notably, the transformation of health services during the COVID-19 pandemic activated the promotion and application of digital health in cancer care, in parallel with the acceleration of the interest and investment of health systems in remote care and digital technologies [19,20]. Compared to the traditional face-to-face delivery model, digital delivery technology is perceived to offer several advantages. In addition to high scalability, low access thresholds, and availability, natural anonymity provides a more private and stigma-free environment, which helps overcome some limitations of in-person psychological support for patients with cancer [21].

In recent years, the growing interest in digital health interventions for mental health management in cancer care has led to numerous systematic reviews and meta-analyses. These studies cover a variety of interventions (ie, internet cognitive behavioral therapy [iCBT] [22,23], "web-based" mindfulness-based cognitive behavioral therapy [24,25], and music therapy [26]). Though several current systematic reviews have examined the efficacy of digital health interventions on specific indicators of mood disorders (ie, anxiety [23,27-30], depression [23,27-31], and fear of recurrence [32,33]) in patients with cancer, mental health is a complex and integrated concept, which is hard to comprehensively embrace and assess by a single systematic review or meta-analysis. Furthermore, the current evidence is extensive and scattered, inconsistent conclusions and varied interventions make it difficult to use a similar metric and methodological framework to appraise it. In this context, umbrella review has emerged as a more integrated research methodology. Nevertheless, published umbrella reviews in the field of digital health care focused on physical activity [34] or other individuals [35], lacking research on mental health aspects.

Given the earlier findings, this umbrella review aims to comprehensively summarize and appraise the available evidence on the efficacy of digital health interventions for alleviating psychological symptoms among patients with cancer.

Methods

Overview

We conducted an umbrella review, in which all currently available evidence from previously published multiple systematic reviews and meta-analyses was systematically collected and assessed, and it could provide an overall picture of the digital health care area on mental health management for patients with cancer and highlight whether the evidence base is consistent or contradictory [36-38]. It adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [39,40] and was conducted according to the recommendations for umbrella reviews to report findings [41]. The protocol was registered on PROSPERO (CRD42024565084).

Search Strategy and Inclusion or Exclusion Criteria

Six databases were searched from their inception to February 4, 2024: Embase, PsycINFO, PubMed, CINAHL, the Cochrane Library, and Web of Science. Our search strategy used the following terms: (1) neoplasm* OR cancer OR oncology* OR tumor? OR "secondary cancer" OR malignancy, (2) "web-based" OR "internet-based" OR "technology-based" OR "ehealth" OR "mhealth" OR "connected health" OR "telehealth" OR online OR digital OR mobile OR "text messag*" OR "social media" OR "internet-based cognitive behavioral therapy" OR "ICBT" OR "online mindfulness-based cognitive behavioral therapy," (3) intervention OR "self-management" OR "support care" OR program* and (4) mental health OR mental OR psycho* OR depression OR anxiety OR distress OR mood OR fatigue. The complete list of search terms is presented in Multimedia Appendix 1. The reference lists and citations of relevant studies were manually examined to identify additional publications. We did not pursue unpublished and gray literature and key journals. The eligibility criteria were structured using the Population, Intervention, Comparison, Outcome, and Study (PICOS) framework (Textbox 1) [42].

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria for this umbrella review.

Inclusion criteria

- Population: (1) The population of interest comprised patients with cancer, regardless of age, any type, and stage throughout the entire cancer continuum, from diagnosis to survivorship; and (2) focused on diverse health conditions but included cancer groups, and from which relevant data could be independently extracted.
- Intervention: Our operational definition of digital health comprises eHealth, mHealth, telehealth, virtual reality, and telemedicine and (2) any type of digital health intervention, whether psychological, physical, or supportive care interventions provided by any form of digital technology (eg, website, telephone, smartphone app, and videoconference).
- Comparison: No restrictions.
- Outcome: (1) The outcome of interest was mental health, which is defined as a state of well-being that allows individuals to cope with the normal stresses of life and function productively, with several core domains encompassing mental health literacy, self-perceptions, values, cognitive skills, emotions, self-management strategies, and quality of life [43], including but not limited to indicators of psychological well-being and any varying levels of psychological distress; and (2) other indicators and precursors of mental health, such as self-efficacy, social support, mindfulness, sleep problems, resilience, rumination, perceived stress, posttraumatic stress, or problems were also considered [44,45].
- Study design: (1) Systematic review (with or without meta-analysis); (2) published in peer-reviewed journals; and (3) written in English.

Exclusion criteria

- Population: (1) Did not exclusively consider patients with cancer and (2) all studies particularly focused on a specific area or race.
- Intervention: Did not mainly relate to digital health interventions.
- Comparison: No restrictions.
- Outcome: Did not focus on mental health.
- Study design: Other wrong study designs (such as scoping reviews, literature reviews, or primary research).

Selection and Screening Process

We imported all retrieved records into Zotero (Corporation for Digital Scholarship) for the removal of duplicates and management and for title, abstract, and full-text screening. Before the selection phases, standard training was created for each reviewer to identify the review qualification. The full text of relevant reviews was independently evaluated by 2 reviewers to finalize its eligibility (CH and JC). Disagreements were resolved by a consensus session with another reviewer (YW).

Data Extraction and Synthesis

To minimize the risk of error and bias, a predefined Microsoft Excel spreadsheet was developed and pilot-tested on 8 randomly selected reviews and then refined accordingly. Data were independently extracted by 2 authors (CH and JC; Multimedia Appendix 2).

Two reviewers (CH and JC) independently evaluated the methodological quality of the included reviews using the Assessing the Methodological Quality of Systematic Reviews-2 tool [46], a strict, validated, and reliable appraisal tool used for systematic reviews and meta-analyses on health care interventions. It consists of 16 items and rating overall confidence in the results of the review as 4 grades: high, moderate, low, or critically low. Any disagreements and conflicts were resolved through discussion by the review team until a consensus was reached.

Due to the great heterogeneity of interventions and delivery technology and the inconsistency of measured outcomes, evidence was analyzed by narrative synthesis.

Results

Study Selection

In total, 2454 records were retrieved from 6 electronic databases of which 636 studies were retained after removing 1818 duplicate records. Screened by abstracts and title, 519 records were excluded. We reviewed the full texts of the remaining 117 studies and excluded 49 studies because of the following reasons: not include mental health outcomes (n=18, 37%), unable to independently extract cancer-related results (n=14, 29%), only focused on informal caregivers or family members (n=7, 14%), protocol or not systematic (n=6, 12%), restriction of area and race (n=3, 6%), and not available full text (n=1, 2%). A hand search was conducted for references and citations, and an additional 10 records were identified for eligibility. Finally, 78 reviews were included, of which 45 (58%) reviews were meta-analyses, and 33 (42%) reviews were systematic reviews for narrative synthesis. Figure 1 illustrates the selection process flowchart according to the PRISMA guidelines.

The PRISMA checklist is found in Multimedia Appendix 3. A full list of excluded studies from the full-text review with reasons for exclusion can be found in Multimedia Appendix 4.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the systematic search and selection process.

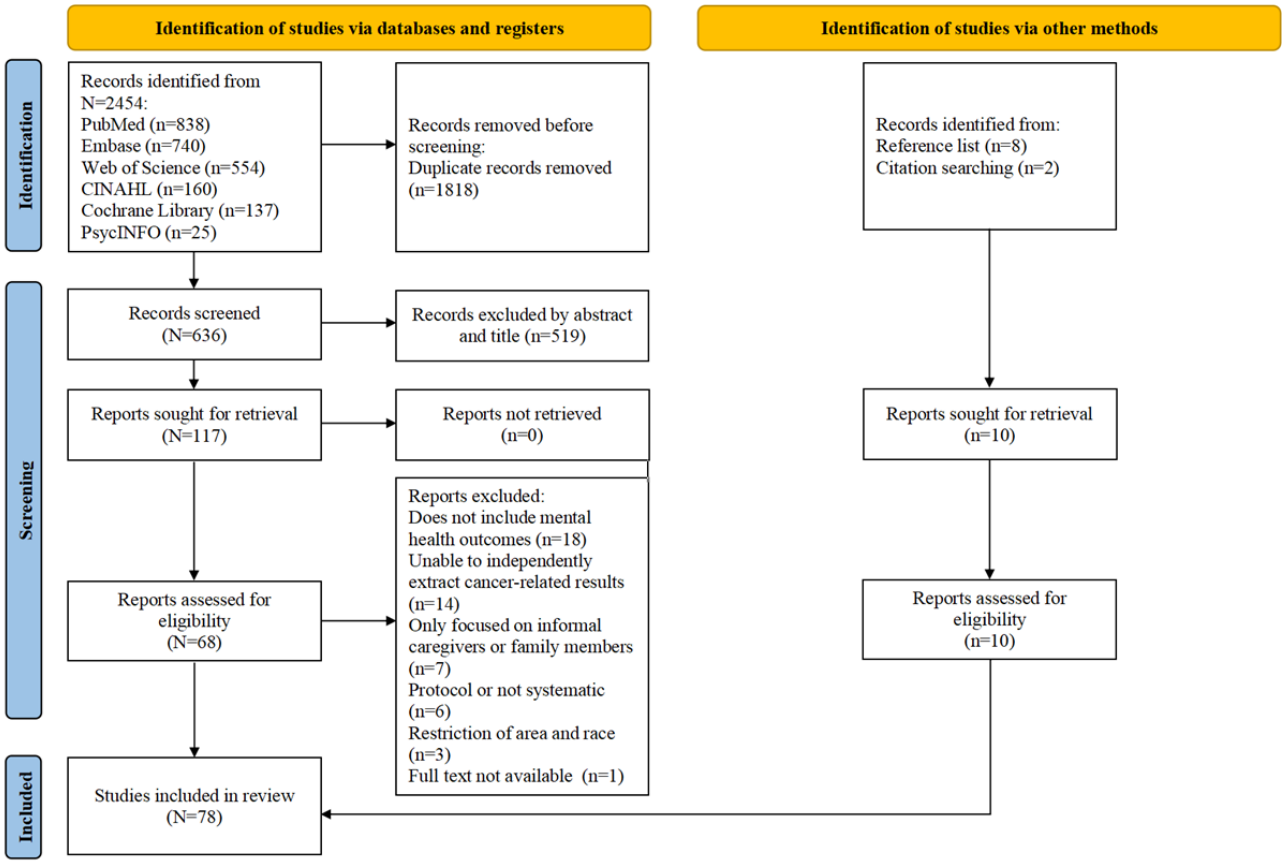

Study Characteristics

Overview

The included reviews were published between 2015 and 2024, and it is worth mentioning that 58 (74%) reviews were published after 2019. Sixty-eight (87%) reviews reported geography information to ensure diversity, with research mainly performed in the United States, Europe, and Asia Pacific, while countries in low or middle-income areas like Africa and South America were less common. The number of studies and sample size varied from 4/374 to 68/13,125. Sixty-two (80%) reviews reported the range of duration of interventions, and 36 (41%) reviews reported the follow-up period. Twenty-seven (35%) reviews included interventions that did not report explicit providers, and 29 (37%) reviews included automatic feedback or self-guided interventions, while 32 (41%) reviews included interventions that were provided by health care professionals, which consisted of mainly nurses, physicians, or psychologists, in addition to dietitians, and information engineers. A framework was developed based on this umbrella review showing how digital health interventions have been used to improve mental health in patients with cancer. Our represented framework (Figure 2) consists of 4 layers: individual, technology, involvement, and intervention, and each layer contains the corresponding aspects categories. The top layer is the main content involved in digital health interventions on mental health. The second layer is the participants required for the intervention. The third layer is the involved digital technologies and delivery platforms. The fourth layer is the target population. Extracted study information is presented in Table 1.

XSL•FO
RenderX

Figure 2. The basic framework of evidence in this umbrella review.

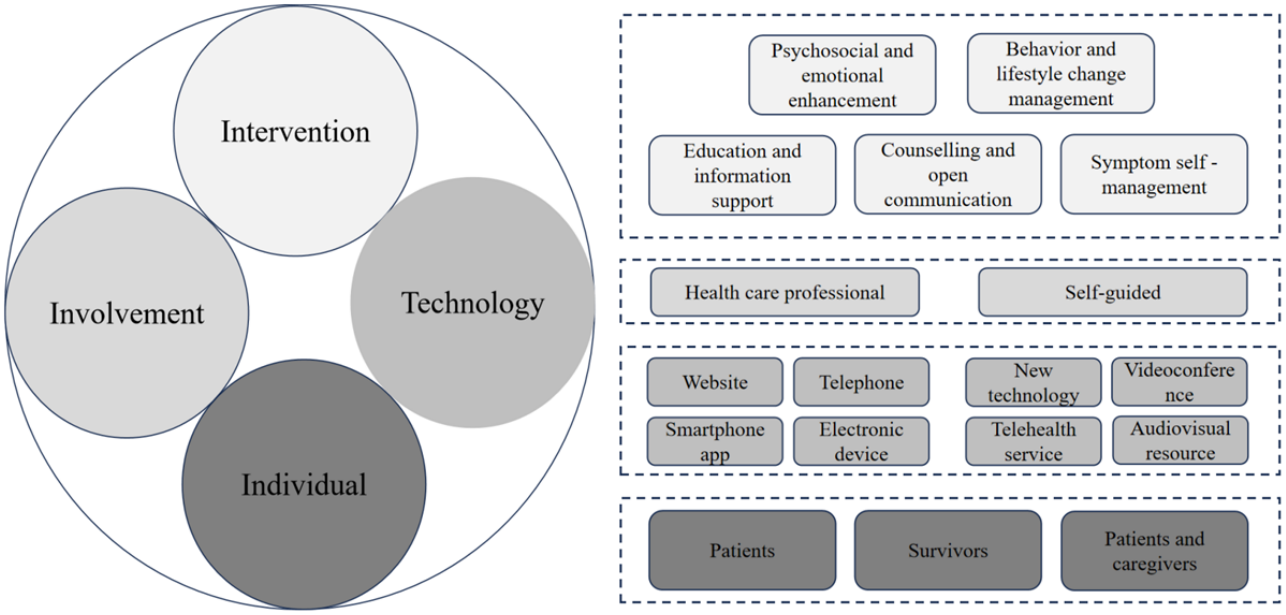

XSL•FO
RenderX

Table 1. Descriptive statistics of 78 included studies.

Characteristics	Review, n (%)
Population	
Target individuals	
Patients	39 (50)
Survivors	16 (21)
Patients and survivors	18 (23)
Patients and caregivers	5 (6)
Cancer type	
Breast	18 (23)
Head and neck	2 (3)
Colorectal	2 (3)
Lung	1 (1)
Prostate	1 (1)
Gynecological	1 (1)
Multiple	53 (68)
Intervention	
Delivery technology and channels	
Website	57 (73)
Smartphone app	45 (58)
Telephone	28 (36)
Telehealth messaging service	21 (27)
Videoconference	15 (19)
Gamification and new technology (virtual reality, artificial intelligence system, active or video game, and robot)	34 (44)
Electronic device (tablets, iPad, and personal computer)	13 (17)
Audiovisual resource (MP3 player, audio-CD, and DVD)	8 (10)
Type of intervention component	
Psychosocial and emotional enhancement	48 (62)
Behavior and lifestyle change management	23 (30)
Education and Information Support	24 (31)
Counseling and open communication	17 (22)
Symptom detection and self-management	30 (39)
Multiple or not specified	18 (23)
Provider involvement	
Health care professional	32 (41)
Fully or almost all self-guided	29 (37)
Multiple or not specified	27 (35)
Study design	
Meta-analysis	
Yes	45 (58)
No	33 (42)
Control group type (if required)	
Routine care (usual or conventional care and standard care)	8 (18)

XSL•FO
RenderX

Characteristics	Review, n (%)
Any	34 (76)
Face-to-face control	3 (4)
Publication bias (among meta-analysis)	
Yes	29 (64)
No	16 (36)
Quality or bias assessment	
Yes	78 (100)
No	0 (0)

Target Population

Of 78 included reviews, 39 (50%) mainly addressed patient-level interventions, 16 (21%) exclusively addressed survivor-level interventions, and 18 (23%) both focused on patients and survivors. The remaining 5 reviews (6%) featured patient and caregiver-level interventions [47-51], and two of them were in the form of dyads (1 review was a patient-caregiver dyad [49] and the other review was a survivor-caregiver dyad [51]), which aimed at exploring specific characteristics on web-based dyadic interventions.

25 reviews (32%) restricted participants based on cancer type, with breast cancer being the most common focus in 18 (23%) reviews, of which 6 reviews restricted inclusion criteria for only women. Two reviews each targeted colorectal cancer [52,53], and head and neck cancer [54,55], while 1 review targeted lung cancer [56], 1 review targeted prostate cancer [28], and 1 review targeted gynecological cancer [30]. The remaining 53 (67.9%) reviews included one or more specific cancer types, with 1 review exclusively focused on advanced cancer [57]. Six reviews (8%) focused on pediatric, adolescent, or young adult individuals with cancer or their caregivers. Of these, 3 reviews focused on pediatric, adolescent, or young adult patients with cancer [58-60], 2 on adolescent, or young adult patients with cancer [61,62], and 1 focused on children and adolescents and their parents with cancer [63]. No reviews mentioned older adult patients with cancer or people who have survived cancer. Because cancer type was most frequently and clearly reported across reviews, we additionally used it to stratify our synthesis results (Figure S1 in Multimedia Appendix 5).

Delivery and Type of Intervention

Rather than focus on a single intervention delivery platform, reviews were more interested in including studies of contained interventions that entirely or in part provided multiple asynchronous or synchronous delivery platforms. It should be emphasized that some in-person elements could be used as additional supplementary components (eg, printed material, educational brochures, face-to-face consultations) at the same time.

Of the multiple included delivery technologies or platforms, website (n=57, 73%) is the most frequently selected platform, smartphone app was followed by (n=45, 58%), and other common technologies are telephone (n=28, 36%), telehealth messaging services (n=21, 27%), videoconference (n=15, 19%), electronic devices (n=13, 17%), and audiovisual resource (n=8, 10%). Admittedly, the use of gamification and new technologies such as virtual reality (n=12, 15%), artificial intelligence (AI) systems (n=1), and communicative chatbots or humanoid robots (n=5, 6%) were not in the minority. As shown in Table 1, the delivery methods based on websites and smartphone apps are the most common. Therefore, we conducted a subgroup narrative synthesis for these two approaches (Figures S2 and S3 in Multimedia Appendix 5). The impact of each platform on various outcomes is represented by the number of studies conducted. Combinations of delivery platforms show a broader coverage of outcomes, indicating that multi-platform combined interventions can provide more comprehensive support.

Similarly, the included interventions were diverse. Eighteen reviews (23%) included studies that were not specifically detailed in their intervention methods or were too broad to be classified. Among the remaining 60 reviews (77%), due to a large variation in reported intervention components, we have categorized them into five dimensions: (1) psychosocial and emotional enhancement; (2) behavior and lifestyle change management; (3) education and information support; (4) counseling and ope communication; and (5) symptom detection and self-management. The first dimension is the most predominant (62%) of included reviews and commonly involves psychological interventions, psychoeducation, and social or peer support. Of the more than a dozen digital psychological interventions included, the most reported type of intervention was iCBT, which a total of 24 (31%) reviews investigated. All interventions were designed based on CBT theory, involving various elements (eg, cognitive restructuring, problem-solving strategy, coping skill training). Certainly, the structure and content of the iCBT were diverse. For instance, 3 reviews included studies that used web-based specified or tailorable CBT training modules [22,23,64], which could be completed with therapist support or with self-guidance. In general, the self-guided method is generally realized by the corresponding modules of intervention independently completed by patients. The therapist-guided method is conducted by therapists providing feedback and support, conducting intervention sessions, or monitoring symptoms via digital technology platforms (eg, email, videoconference, telephone) or internet-based interaction with groups of other patients. Four reviews reported that compared to self-guided interventions, therapist-guided interventions were more efficacious in engagement, improvement of quality of life (QoL), and

XSL•FO
RenderX

adjustment of some negative emotions [22,65,66]. In addition, 19 (24%) reviews mentioned mindfulness-based interventions, including mindfulness-based cognitive therapy (MBCT) [25,33,67,68], mindfulness-based stress reduction [25,69], mindfulness-based cancer recovery [70], and mindfulness self-compassion [32,71]. Other common psychological interventions included acceptance and commitment therapy [32,33,52,70], problem-solving therapy [24,32,65,71,72], and cognitive rehabilitation therapy [62,72]. Gratitude intervention [33], supportive expressive therapy [24], narrative therapy [24], and music therapy [26] were less commonly included. Fourteen (18%) reviews reported psychoeducation. Sixteen (20.5%) reviews reported components of social support or peer support, which adopt various formats (eg, web-based workshops, portals, or discussion forums). McCaughan et al [73] assessed the effects of online support groups on negative emotions and QoL of female patients with breast cancer.

Second, 23 reviews (30%) reported interventions related to behavior and lifestyle change, generally consisting of physical activity intervention, dietary and nutrition intervention, and exercise prescriptions. Third, 24 reviews (31%) reported information support interventions, the majority provided knowledge of cancer disease and treatment, available medical resources or health care service information, self-management strategies for emotional and physical symptoms, and so on. One review mentioned culture-related specific educational information particularly [67]. Qin et al [74] reported a pattern that based on predesigned personalized code and program, the app generates automated feedback and feeds hyper-relevant and tailored suggestions. Fourthly, 30 reviews (39%) reported symptom management interventions, including forms of skills training (eg, coping, rehabilitation), stress management, meditation adherence management, self-assessment, and symptom monitoring. Of the 3 reviews, distraction therapy was used for attention management to alleviate negative emotions in patients with cancer [24,63,75]. Finally, 17 reviews (22%) reported counseling and open communication. The main form is that patients could directly contact or consult with experts. Horn et al [54] found that nurses may be the bridge and activator to facilitate survivors to obtain more information from physicians.

Comparison

45 reviews (57.7%) included a meta-analysis of all or a subset of their included studies. In the control groups of 45 meta-analyses, 8 (18%) reviews conducted usual or conventional care and standard care to participants, and the remaining reviews were compared with any comparator (eg, waitlist control, active control, attention control, no intervention control, placebo control). In addition, 3 reviews reported that the effectiveness of digital interventions in reducing fatigue, fear of recurrence, and psychological distress was similar to face‐to‐face interventions [32,51,76]. Moreover, Chen et al [51] examined the efficacy of eHealth interventions in cancer survivorship care and found that compared with traditional face-to-face dyadic interventions, web-based dyadic interventions can break the constraints of time and space and may better address needs during post-treatment survivorship.

Implement Outcome

Neither statistical pooling of the results nor a meta‐analysis was performed because of the high heterogeneity of the included reviews. The reported targeted variable was considered positive when at least half of the studies included in the narrative synthesis showed positive results, or when the meta-analysis showed a significant effect.

Across 45 meta-analyses, the most examined outcome was depression, 29 (64.4%) of which reported positive effects of interventions relative to control, and 6 reported null findings. Followed by QoL (29 positive effects and 3 null effects), and anxiety (25 positive effects and 7 null effects). Other commonly reported outcomes include cancer-related fatigue (13 positive effects, 5 null effects), distress (10 positive effects, 3 null effects), self-efficacy (8 positive effects and 1 null effect), sleep-related problems (6 positive effects and no null effects), fear of recurrence (3 positive effects and 2 null effects), and well-being (2 positive effects, no null effects. Of the remaining 33 systematic reviews, not surprisingly, the most concerned outcomes were still QoL, depression, and anxiety, which also showed promising trends. The global positive effect of the interventions is depicted in Figure 3.

XSL•FO
RenderX

Figure 3. Summary of evidence for targeted psychological outcomes. The green indicates "positive" and purple indicates "no effect." HRQoL: health-related quality of life; QoL: quality of life.

Outcome	Meta-analyses		Narrative syntheses
	Positive	No effect	
Depression	29	6	15/19
QoL or HRQoL	29	3	16/20
Anxiety	25	7	9/17
Fatigue	13	5	9/11
Distress	10	3	5/10
Self-efficacy	8	1	5/7
Sleep disorders	6	0	3/5
Fear of recurrence	3	2	1/1
Well-being	2	0	2/4
Cognitive functions	1	0	2/2
Hope	1	0	1/1
Mindfulness	1	1	2/2
Mental health	1	0	0/1
Perceived stress	1	0	1/1
Stress	1	1	1/1
Multicomponent indicators	1	0	4/5
Anger	0	0	1/1
Emotional functioning	0	0	1/1
Mood or mood disturbance	0	1	1/1
Resilience	0	0	1/2
Rumination	0	1	0
Posttraumatic stress	0	0	1/1
Posttraumatic growth	0	2	1/1
Psychological empowerment	0	0	1/1
Social support	0	0	3/4

Meta-analyses		Narrative syntheses	
Direction of effect		n/N	
	Positive	Numerator	Number of studies with effect
	No effect	Denominator	Total studies
	No data		

Quality Assessment

The majority of included reviews had critical weaknesses and were rated as critically low (n=42, 53.8%) or low (n=19; 24.4%) confidence. Only 4 reviews were rated as high confidence (5.1%), and 13 (16.7%) were rated as moderate confidence. Common methodological weaknesses were failure to report on the sources of funding for the studies included in the review and failure to provide a full list of excluded studies (only 3.8% successfully met this criterion). Due to space limitations, the full results are found in Multimedia Appendix 6.

Discussion

Principal Findings

To our knowledge, this is the first umbrella review to extensively summarize and evaluate the scientific evidence on distinct digital health interventions in improving mental health issues for patients with cancer. It provides an overview of the current state of this domain and identifies gaps regarding delivery modalities, intervention methods, and targeted psychological indicators. Notably, the majority of literature published after 2019, emphasizes the trend to integrate technology into psychological distress management for patients with cancer, highlighting the growing importance of digital interventions in oncology care. Furthermore, the variety of delivery platforms caters to diverse patient preferences and demonstrates the adaptability of interventions for different backgrounds and populations.

XSL•FO
RenderX

The majority of reviews primarily focused on adult patients with cancers or people who have survived cancer while a minority referred to pediatric, adolescent, and young groups. However, none of them focused on older patients with cancer, which conflicts with the staggering discovery that more than 50% of new cases and nearly 70% of death cases are diagnosed with cancer aged 65 years and older [77]. Cheng et al [78] acknowledged the prevalence of spiritual needs among older adults with cancer and highlighted that psychological care should be an indispensable part of daily care. Other research showed that remote health care in older adults contributed to the QoL, level of depression, anxiety, and prognosis, as well as more favorable psychological outcomes [79-82]. Although it is often assumed that digital health interventions may not be appropriate for older adults because of several barriers such as lack of usability and perceived usefulness [83,84]. A study of patients with genitourinary cancer suggests that high engagement and interest in digital technologies were observed among older patients [85]. Moreover, older patients with cancer usually have unique preferences regarding digital health interventions [86]. In contrast to the stimulating sensory experiences and diverse games offered by complex interfaces, older adult patients with cancer prefer clear and concise user instructions, which allows those with no prior experience or nonprofessionals to also get started easily [86-88]. Future research should explore the development of personalized digital interventions for older adult patients with cancer [82,89]. Considering factors like social isolation and cognitive decline [90,91], efforts should focus on enhancing feasibility, convenience, and user engagement to achieve better outcomes [83,92,93].

Within the extensive scope of digital health technologies covered in this study, websites and smartphone apps are particularly widely used in cancer care. This widespread use is due to their advantages in accessibility and functionality compared to other delivery formats. Given the characteristics of web browsers on popularization and no need to download, web-based platforms are often easier to access, making them a preferred choice for patients who use older or insufficient memory usage devices. Even so, they may lack the immediacy and interactivity that smartphone apps provide. Whereas smartphone apps offer distinct advantages in personalized design, real-time monitoring and feedback, and the capacity for big data delivery [94,95]. By offering vivid and comprehensible materials, it bridges the gap between patients and substantial evidence-based educational resources, which can alleviate to some extent patients' anxiety due to lack of knowledge [96,97]. The capabilities of push notifications and real-time data tracking that smartphone apps have led to increased user engagement and adherence to interventions [98,99]. Likewise, tailored user interfaces and digital features can significantly increase patient satisfaction [100], which demonstrates the importance of optimizing digital health interventions to meet patient preferences and needs. In other words, the effectiveness of these interventions depends on their ability to seamlessly integrate into the user's daily life, providing personalized and timely support that meets the user's emotional and psychological needs. We agree with Kamalumpundi et al [57] that effective web-based emotion regulation interventions are far more intricate than merely offering individuals a range of app features.

The design of apps in this study mostly focuses on patient education, disease self-management, and remote monitoring of symptoms [18], but lacks highly tailored symptom management interventions for certain emotional conditions. One plausible explanation was as Krueger and Eaton [101] stated, among different diseases, there may exist shared subthreshold disorder manifestations, which could be associated with significant distress and functional impairment. This indicates that categorical diagnoses may fail to capture the underlying dimensions of mental disorders and emphasize the loss of information when complex constellations of signs and symptoms are simplified into an "either/or" binary framework. Furthermore, it helps explain why specific psychotherapeutic approaches are purportedly effective across a range of ostensibly distinct emotional states. In addition to the more concerned anxiety and depression, many other specific emotional states affect patients such as self-doubt [102], guilt [103], anger [104], and self-esteem [105], which received insufficient attention. To address this complexity, future research should focus on developing high-precision emotion recognition and decision-making applications. Advanced algorithms and machine learning techniques could be conducted to accurately identify and classify emotional states. For instance, ecological momentary assessments could be used for daily monitoring of mental health care [106]. Combined with wearable devices with social media data, real-time emotion tracking was conducted to provide tailored feedback and interventions. Furthermore, the implementation of evidence-based interventions that correspond to specific emotional categories could enhance the effectiveness of emotional management strategies [107]. Individuals experiencing anxiety may benefit from cognitive behavioral techniques and mindfulness exercises, while those with depression might require mood monitoring and behavioral activation strategies [108,109]. By integrating these tailored interventions into digital platforms, we can offer precise, individualized emotional support.

Although a few reviews included interventions with gamification (eg, active games, virtual reality) or new technologies (eg, AI systems, humanoid chatbots), which progress swiftly and as a highly promising trend to solve several limitations of current situations. The results of a scoping review by Poliani et al [110] show that gamification seems to improve QoL and reduce anxiety levels in patients with cancer, which was consistent with other findings [94,95,111,112], and broader exploration of other health-related outcomes, indicating that gamification also had a significant effect on anxiety, distress, and cognitive function. However, all gamified interventions included in the studies were ordinary commercial apps for the universal population and were not specifically developed for patients with cancer. Additionally, designing gamified interventions tailored to older patients is crucial. As the aging population grows, there is a need for age-appropriate gamification strategies that can address the unique challenges faced by older adult patients [113,114]. Simplified interfaces, cognitive training games, and physical activity-based interventions are all worthwhile options. After all, digital health interventions should be more inclusive, effectively supporting patients with cancer of all ages. Furthermore, large language models have found diverse applications in clinical practice, including supporting clinical

XSL•FO
RenderX

decision-making, intelligent question answering, generating medical documents, and assisting therapy through chatbots [115,116]. Future advances aim to explore further integration of AI technologies into mobile emotion management. This involves implementing scientific frameworks to facilitate their adoption and usage within mental health care systems, to enhance its accessibility and scalability, implementing scientific frameworks may be a proper choice for facilitating their adoption and use.

Variations in reported intervention effects across outcome measures were observed in the results [24,64,66,70,74,117]. This variability may be explained by conceptual ambiguities among studies regarding specific indicators, leading to differences in measurement tools and assessment methods. Moreover, the heterogeneity in study designs, intervention components, and outcome measurement methods, makes it hard to give a rigorous and accurate report on which of them are associated with the best efficacy.

It is mentioned that although the objective of this umbrella review was to systematically synthesize the effectiveness of digital interventions on the mental health of all patients with cancer, the majority of the patients with cancer included in the studies originate from high-income countries and regions, with limited racial representation. Consequently, it cannot draw convincing conclusions for vulnerable cancer groups such as patients from middle- and low-income backgrounds and ethnic minorities. We must acknowledge the disparity in digital health equity [118]. Typically, high-income and middle-income individuals have greater access to technology and health care facilities than the low-income group, who may also have fewer opportunities to obtain medical care [119]. A multiclinic study targeting ethnic minorities has indicated that the lack of broadband access is likely one of the significant factors affecting the adoption of telemedicine by these minority groups [120]. Furthermore, healthy digital literacy is a crucial element in being able to process complex health information and effectively absorb and use it. Generally, the health literacy levels of vulnerable groups are already low, and there are notable differences in the social, economic, or environmental contexts of health, which make it challenging for these groups to cope with and depend on complex technologies to search for information [121]. With the increasing implementation of digital health services, we must recognize the limitations of these tools and implement them in a manner that promotes optimal functionality, accessibility, and usability. Future research should consider multiple levels and perspectives to reduce health literacy barriers and enhance the acceptance of technology among specific groups.

Limitations

While this comprehensive umbrella review provides valuable insights, several limitations need consideration. First, due to heterogeneity in outcome measures, intervention components, and methods of outcome assessment, we did a narrative synthesis and were unable to check for overlap of included individual studies statistically. Moreover, it is impossible to systematically characterize the heterogeneity in the included reviews and assess the potential publication bias. Second, unclear conceptual definitions are commonly encountered. Indeed, digital health interventions constitute a relatively ambiguous concept in literature, where descriptions of intervention measures may be confusing, and there is considerable overlap between similar digital intervention definitions. Similar problems exist in the evaluation of outcome measures related to mental health. Hence, with the increasing prevalence of digital mental health services based on technologies such as the internet, big data, and AI, stricter descriptions and definitions of relevant concepts seem crucial. Third, the general demographics information collected in the included reviews most focuses on factors such as age, sex, and region, neglecting other factors like income and ethnicity, which makes it difficult to adequately consider the needs of other underrepresented groups. The priority in future research should be to fully understand these existing problems and strive to resolve them through more rigorous and cautious research.

Conclusions

In general, the review identified that various interventions delivered by digital technologies, a feasible and available approach, can facilitate subjectively assessed levels of more than a dozen emotional parameters in patients with cancer. There are also some limitations to these results. The great heterogeneity observed in the studies makes it difficult to have a quantitative synthesis of results, and several reviews have not followed the methodological requirements for reporting results. Further research is needed to develop rigorous methodological interventions allowing for scrupulous testing to determine the effective types of interventions and their exact effects. More precise and sensitive emotional measurement tools and identification systems are needed to capture subtle changes in patient's psychological needs and preferences, enhancing the targeted nature of interventions. Additionally, integrating evidence-based intervention measures into standard oncological care forms multidimensional, multilevel support strategies. This integration can provide more effective support and guidance for clinical practice and policy-making, thereby offering comprehensive and personalized psychological support and care for patients with cancer. Ultimately, this approach aims to enhance the psychosocial health and QoL of patients with cancer globally.

Acknowledgments

The authors thank all researchers who participated in this study. This study is supported by the National Natural Science Foundation of China (grant 72304131), the Outstanding Youths Development Scheme of Nanfang Hospital, Southern Medical University (2023J005), and the National Natural Science Foundation of China (72101107).

XSL•FO
RenderX

Data Availability

All the data and materials presented in this study are available within the respective articles cited in this review and from the corresponding author on reasonable request.

Authors' Contributions

YW conceptualized this umbrella review. CZ drafted the original manuscript. All authors commented on subsequent versions of the manuscript and approved the final version. All authors had access to the data and accepted responsibility for submitting it for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The complete search strategy.
[PDF File (Adobe PDF File), 96 KB - jmir_v27i1e69621_app1.pdf]

Multimedia Appendix 2
The full extraction data.
[XLSX File (Microsoft Excel File), 52 KB - jmir_v27i1e69621_app2.xlsx]

Multimedia Appendix 3
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 97 KB - jmir_v27i1e69621_app3.pdf]

Multimedia Appendix 4
A full list of excluded studies from the full-text review with reasons for exclusion.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v27i1e69621_app4.xlsx]

Multimedia Appendix 5
Subgroup narrative synthesis on cancer types, website, and mobile app for summary of evidence for targeted psychological outcomes.
[DOCX File , 689 KB - jmir_v27i1e69621_app5.docx]

Multimedia Appendix 6
The full results of the quality assessment.
[XLSX File (Microsoft Excel File), 175 KB - jmir_v27i1e69621_app6.xlsx]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Boyes AW, Girgis A, D'Este CA, Zucca AC, Lecathelinais C, Carey ML. Prevalence and predictors of the short-term trajectory of anxiety and depression in the first year after a cancer diagnosis: a population-based longitudinal study. J Clin Oncol 2013;31(21):2724-2729. [doi: 10.1200/JCO.2012.44.7540] [Medline: 23775970]
3. Osmani V, Hörner L, Klug SJ, Tanaka LF. Prevalence and risk of psychological distress, anxiety and depression in adolescent and young adult (AYA) cancer survivors: a systematic review and meta-analysis. Cancer Med 2023;12(17):18354-18367 [FREE Full text] [doi: 10.1002/cam4.6435] [Medline: 37559504]
4. Hart NH, Crawford-Williams F, Crichton M, Yee J, Smith TJ, Koczwara B, et al. Unmet supportive care needs of people with advanced cancer and their caregivers: a systematic scoping review. Crit Rev Oncol Hematol 2022;176:103728. [doi: 10.1016/j.critrevonc.2022.103728] [Medline: 35662585]
5. Wolyniec K, Sharp J, Fisher K, Tothill RW, Bowtell D, Mileshkin L, et al. Psychological distress, understanding of cancer and illness uncertainty in patients with cancer of unknown primary. Psychooncology 2022;31(11):1869-1876 [FREE Full text] [doi: 10.1002/pon.5990] [Medline: 35765251]
6. Dee EC, Nipp RD, Muralidhar V, Yu Z, Butler SS, Mahal BA, et al. Financial worry and psychological distress among cancer survivors in the United States, 2013-2018. Support Care Cancer 2021;29(9):5523-5535. [doi: 10.1007/s00520-021-06084-1] [Medline: 33725174]

XSL•FO
RenderX

7. Rao W, Yang M, Cao B, You Y, Zhang Y, Liu Y, et al. Psychological distress in cancer patients in a large Chinese cross-sectional study. J Affect Disord 2019;245:950-956. [doi: 10.1016/j.jad.2018.11.089] [Medline: 30699880]
8. Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncol 2011;12(2):160-174. [doi: 10.1016/S1470-2045(11)70002-X] [Medline: 21251875]
9. Brandenbarg D, Maass SWMC, Geerse OP, Stegmann ME, Handberg C, Schroevers MJ, et al. A systematic review on the prevalence of symptoms of depression, anxiety and distress in long-term cancer survivors: implications for primary care. Eur J Cancer Care 2019;28(3):e13086 [FREE Full text] [doi: 10.1111/ecc.13086] [Medline: 31087398]
10. Hoffman KE, McCarthy EP, Recklitis CJ, Ng AK. Psychological distress in long-term survivors of adult-onset cancer: results from a national survey. Arch Intern Med 2009;169(14):1274-1281. [doi: 10.1001/archinternmed.2009.179] [Medline: 19636028]
11. Bultz BD, Groff SL, Fitch M, Blais MC, Howes J, Levy K, et al. Implementing screening for distress, the 6th vital sign: a Canadian strategy for changing practice. Psychooncology 2011;20(5):463-469. [doi: 10.1002/pon.1932] [Medline: 21456060]
12. Wang YH, Li JQ, Shi JF, Que JY, Liu JJ, Lappin JM, et al. Depression and anxiety in relation to cancer incidence and mortality: a systematic review and meta-analysis of cohort studies. Mol Psychiatry 2020;25(7):1487-1499. [doi: 10.1038/s41380-019-0595-x] [Medline: 31745237]
13. Zebrack B, Kayser K, Sundstrom L, Savas SA, Henrickson C, Acquati C, et al. Psychosocial distress screening implementation in cancer care: an analysis of adherence, responsiveness, and acceptability. J Clin Oncol 2015;33(10):1165-1170. [doi: 10.1200/JCO.2014.57.4020] [Medline: 25713427]
14. Erdoğan Yüce G, Döner A, Muz G. Psychological distress and its association with unmet needs and symptom burden in outpatient cancer patients: a cross-sectional study. Semin Oncol Nurs 2021;37(5):151214. [doi: 10.1016/j.soncn.2021.151214] [Medline: 34483014]
15. Passik SD, Dugan W, McDonald MV, Rosenfeld B, Theobald DE, Edgerton S. Oncologists' recognition of depression in their patients with cancer. J Clin Oncol 1998;16(4):1594-1600. [doi: 10.1200/JCO.1998.16.4.1594] [Medline: 9552071]
16. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
17. Fournier V, Duprez C, Grynberg D, Antoine P, Lamore K. Are digital health interventions valuable to support patients with cancer and caregivers? An umbrella review of web-based and app-based supportive care interventions. Cancer Med 2023;12(23):21436-21451 [FREE Full text] [doi: 10.1002/cam4.6695] [Medline: 37937812]
18. Ghanbari E, Yektatalab S, Mehrabi M. Effects of psychoeducational interventions using mobile apps and mobile-based online group discussions on anxiety and self-esteem in women with breast cancer: randomized controlled trial. JMIR mHealth uHealth 2021;9(5):e19262 [FREE Full text] [doi: 10.2196/19262] [Medline: 34003138]
19. Thirugnanasundralingam K, Davies-Tuck M, Rolnik DL, Reddy M, Mol BW, Hodges R, et al. Effect of telehealth-integrated antenatal care on pregnancy outcomes in Australia: an interrupted time-series analysis. Lancet Digit Health 2023;5(11):e798-e811 [FREE Full text] [doi: 10.1016/S2589-7500(23)00151-6] [Medline: 37890903]
20. Spatz ES, Ginsburg GS, Rumsfeld JS, Turakhia MP. Wearable digital health technologies for monitoring in cardiovascular medicine. N Engl J Med 2024;390(4):346-356. [doi: 10.1056/NEJMra2301903] [Medline: 38265646]
21. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
22. Yu S, Liu Y, Cao M, Tian Q, Xu M, Yu L, et al. Effectiveness of internet-based cognitive behavioral therapy for patients with cancer: a systematic review and meta-analysis of randomized controlled trials. Cancer Nurs 2023. [doi: 10.1097/NCC.0000000000001274] [Medline: 37552227]
23. Liu T, Xu J, Cheng H, Zhang Y, Wang S, Lin L, et al. Effects of internet-based cognitive behavioral therapy on anxiety and depression symptoms in cancer patients: a meta-analysis. Gen Hosp Psychiatry 2022;79:135-145. [doi: 10.1016/j.genhosppsych.2022.09.003] [Medline: 36375342]
24. Zhang T, Wakefield CE, Ren Z, Chen W, Du X, Shi C, et al. Effects of digital psychological interventions on physical symptoms in cancer patients: a systematic review and meta-analysis. Gen Hosp Psychiatry 2023;84:47-59. [doi: 10.1016/j.genhosppsych.2023.05.016] [Medline: 37385139]
25. Fan M, Wang Y, Zheng L, Cui M, Zhou X, Liu Z. Effectiveness of online mindfulness-based interventions for cancer patients: a systematic review and meta-analysis. Jpn J Clin Oncol 2023;53(11):1068-1076. [doi: 10.1093/jjco/hyad101] [Medline: 37567587]
26. Bradt J, Dileo C, Myers-Coffman K, Biondo J. Music interventions for improving psychological and physical outcomes in people with cancer. Cochrane Database Syst Rev 2021;10(10):CD006911 [FREE Full text] [doi: 10.1002/14651858.CD006911.pub4] [Medline: 34637527]
27. Yang Y, Huang Y, Dong N, Zhang L, Zhang S. Effect of telehealth interventions on anxiety and depression in cancer patients: a systematic review and meta-analysis of randomized controlled trials. J Telemed Telecare 2024;30(7):1053-1064. [doi: 10.1177/1357633X221122727] [Medline: 36062618]

XSL•FO
RenderX

28. Qan'ir Y, Song L. Systematic review of technology-based interventions to improve anxiety, depression, and health-related quality of life among patients with prostate cancer. Psychooncology 2019;28(8):1601-1613 [FREE Full text] [doi: 10.1002/pon.5158] [Medline: 31222956]
29. Lopez-Rodriguez MM, Fernández-Millan A, Ruiz-Fernández MD, Dobarrio-Sanz I, Fernández-Medina IM. New technologies to improve pain, anxiety and depression in children and adolescents with cancer: a systematic review. Int J Environ Res Public Health 2020;17(10):3563 [FREE Full text] [doi: 10.3390/ijerph17103563] [Medline: 32438762]
30. Yen KY, Cheng JY, Li J, Toh ZA, He H. The effectiveness of digital psychosocial interventions on psychological distress, depression, anxiety, and health-related quality of life in patients with gynaecological cancer: a systematic review and meta-analysis. Support Care Cancer 2024;32(4):240. [doi: 10.1007/s00520-024-08415-4] [Medline: 38512538]
31. Agboola SO, Ju W, Elfiky A, Kvedar JC, Jethwani K. The effect of technology-based interventions on pain, depression, and quality of life in patients with cancer: a systematic review of randomized controlled trials. J Med Internet Res 2015;17(3):e65. [doi: 10.2196/jmir.4009] [Medline: 25793945]
32. Kang N, Yu ES. Is digital intervention for fear of cancer recurrence beneficial to cancer patients?: A systematic review and meta-analysis. Psychooncology 2023;32(9):1348-1358. [doi: 10.1002/pon.6199] [Medline: 37571918]
33. Lyu M, Siah RC, Lam ASL, Cheng KF. The effect of psychological interventions on fear of cancer recurrence in breast cancer survivors: a systematic review and meta-analysis. J Adv Nurs 2022;78(10):3069-3082. [doi: 10.1111/jan.15321] [Medline: 35696315]
34. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
35. Barbui C, Purgato M, Abdulmalik J, Acarturk C, Eaton J, Gastaldon C, et al. Efficacy of psychosocial interventions for mental health outcomes in low-income and middle-income countries: an umbrella review. Lancet Psychiatry 2020;7(2):162-172. [doi: 10.1016/S2215-0366(19)30511-5] [Medline: 31948935]
36. Papatheodorou S, Evangelou E. Umbrella reviews: what they are and why we need them. Methods Mol Biol 2022;2345:135-146. [doi: 10.1007/978-1-0716-1566-9_8] [Medline: 34550588]
37. Solmi M, Correll CU, Carvalho AF, Ioannidis JPA. The role of meta-analyses and umbrella reviews in assessing the harms of psychotropic medications: beyond qualitative synthesis. Epidemiol Psychiatr Sci 2018;27(6):537-542 [FREE Full text] [doi: 10.1017/S204579601800032X] [Medline: 30008278]
38. Aromataris E, Fernandez R, Godfrey C, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
40. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
41. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
42. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123(3):A12-A13. [Medline: 7582737]
43. Fusar-Poli P, Salazar de Pablo G, De Micheli A, Nieman DH, Correll CU, Kessing LV, et al. What is good mental health? A scoping review. Eur Neuropsychopharmacol 2020;31:33-46 [FREE Full text] [doi: 10.1016/j.euroneuro.2019.12.105] [Medline: 31901337]
44. Valkenburg PM, Meier A, Beyens I. Social media use and its impact on adolescent mental health: an umbrella review of the evidence. Curr Opin Psychol 2022;44:58-68 [FREE Full text] [doi: 10.1016/j.copsyc.2021.08.017] [Medline: 34563980]
45. Purgato M, Cadorin C, Prina E, Cabral Ferreira M, Del Piccolo L, Gerber M, et al. Umbrella systematic review and meta-analysis: physical activity as an effective therapeutic strategy for improving psychosocial outcomes in children and adolescents. J Am Acad Child Adolesc Psychiatry 2024;63(2):172-183. [doi: 10.1016/j.jaac.2023.04.017] [Medline: 37331468]
46. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
47. Zhang Y, Flannery M, Zhang Z, Underhill-Blazey M, Bobry M, Leblanc N, et al. Digital health psychosocial intervention in adult patients with cancer and their families: systematic review and meta-analysis. JMIR Cancer 2024;10:e46116. [doi: 10.2196/46116] [Medline: 38315546]
48. Tan JYA, Ong GYQ, Cheng LJ, Pikkarainen M, He H. Effectiveness of mHealth-based psychosocial interventions for breast cancer patients and their caregivers: a systematic review and meta-analysis. J Telemed Telecare 2025;31(2):184-197. [doi: 10.1177/1357633X231187432] [Medline: 37650270]

XSL•FO
RenderX

49. Low NJH, Leow DGW, Klainin-Yobas P. Effectiveness of technology-based psychosocial interventions on psychological outcomes among adult cancer patients and caregivers: a systematic review and meta-analysis. Semin Oncol Nurs 2024;40(1):151533. [doi: 10.1016/j.soncn.2023.151533] [Medline: 37977974]
50. Slev VN, Mistiaen P, Pasman HRW, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
51. Chen M, Gong J, Li Q. The application of eHealth in cancer survivorship care: a review of web-based dyadic interventions for post-treatment cancer survivors and caregivers. Asia Pac J Oncol Nurs 2022;9(10):100109 [FREE Full text] [doi: 10.1016/j.apjon.2022.100109] [Medline: 36118622]
52. Wan SW, Chng YJD, Lim SH, Chong C, Pikkarainen M, He H. A systematic review and meta-analysis on the effectiveness of web-based psychosocial interventions among patients with colorectal cancer. J Adv Nurs 2022;78(7):1883-1896. [doi: 10.1111/jan.15258] [Medline: 35441711]
53. Ayyoubzadeh SM, R Niakan Kalhori S, Shirkhoda M, Mohammadzadeh N, Esmaeili M. Supporting colorectal cancer survivors using eHealth: a systematic review and framework suggestion. Support Care Cancer 2020;28(8):3543-3555. [doi: 10.1007/s00520-020-05372-6] [Medline: 32152763]
54. Hulse K, Li LQ, Lowit A, Maguire R, Douglas C. Digital health in head and neck cancer: a systematic review. J Laryngol Otol 2023;137(12):1312-1325. [doi: 10.1017/S0022215123000695] [Medline: 37078528]
55. Richardson AE, Broadbent E, Morton RP. A systematic review of psychological interventions for patients with head and neck cancer. Support Care Cancer 2019;27(6):2007-2021. [doi: 10.1007/s00520-019-04768-3] [Medline: 30937599]
56. Pang L, Liu Z, Lin S, Liu Z, Liu H, Mai Z, et al. The effects of telemedicine on the quality of life of patients with lung cancer: a systematic review and meta-analysis. Ther Adv Chronic Dis 2020;11:2040622320961597. [doi: 10.1177/2040622320961597] [Medline: 33101621]
57. Kamalumpundi V, Saeidzadeh S, Chi N, Nair R, Gilbertson-White S. The efficacy of web or mobile-based interventions to alleviate emotional symptoms in people with advanced cancer: a systematic review and meta-analysis. Support Care Cancer 2022;30(4):3029-3042. [doi: 10.1007/s00520-021-06496-z] [Medline: 34735601]
58. Chandeying N, Thongseiratch T. Online interventions to improve mental health of pediatric, adolescent, and young adult cancer survivors: a systematic review and meta-analysis. Front Psychiatry 2021;12:784615 [FREE Full text] [doi: 10.3389/fpsyt.2021.784615] [Medline: 35002804]
59. Cheng L, Duan M, Mao X, Ge Y, Wang Y, Huang H. The effect of digital health technologies on managing symptoms across pediatric cancer continuum: a systematic review. Int J Nurs Sci 2021;8(1):22-29 [FREE Full text] [doi: 10.1016/j.ijnss.2020.10.002] [Medline: 33575441]
60. Zhang A, Zebrack B, Acquati C, Roth M, Levin NJ, Wang K, et al. Technology-assisted psychosocial interventions for childhood, adolescent, and young adult cancer survivors: a systematic review and meta-analysis. J Adolesc Young Adult Oncol 2022;11(1):6-16 [FREE Full text] [doi: 10.1089/jayao.2021.0012] [Medline: 33960845]
61. Viola A, Panigrahi G, Devine K. Digital interventions for adolescent and young adult cancer survivors. Curr Opin Support Palliat Care 2020;14(1):51-59. [doi: 10.1097/SPC.0000000000000480] [Medline: 31842020]
62. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
63. Ozdemir Koyu H, Kilicarslan Törüner E. The effect of technology-based interventions on child and parent outcomes in pediatric oncology: a systemic review of experimental evidence. Asia Pac J Oncol Nurs 2023;10(5):100219 [FREE Full text] [doi: 10.1016/j.apjon.2023.100219] [Medline: 37168318]
64. Horn A, Jírů-Hillmann S, Widmann J, Montellano FA, Salmen J, Pryss R, et al. Systematic review on the effectiveness of mobile health applications on mental health of breast cancer survivors. J Cancer Surviv 2023. [doi: 10.1007/s11764-023-01470-6] [Medline: 37906420]
65. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
66. Li J, Liu Y, Jiang J, Peng X, Hu X. Effect of telehealth interventions on quality of life in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]
67. Akdemir A, Smith AB, Wu VS, Rincones O, Russell H, Lyhne JD, et al. Guided versus non-guided digital psychological interventions for cancer patients: a systematic review and meta-analysis of engagement and efficacy. Psychooncology 2024;33(1):e6290. [doi: 10.1002/pon.6290] [Medline: 38282223]
68. Zou P, Huang A, Luo Y, Tchakerian N, Zhang H, Zhang C. Effects of using WeChat/WhatsApp on physical and psychosocial health outcomes among oncology patients: a systematic review. Health Informatics J 2023;29(1):14604582231164697 [FREE Full text] [doi: 10.1177/14604582231164697] [Medline: 36949645]
69. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: a systematic review. Cancer Treat Rev 2019;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]

XSL•FO
RenderX

70. Fung JYT, Lim H, Vongsirimas N, Klainin-Yobas P. Effectiveness of eHealth mindfulness-based interventions on cancer-related symptoms among cancer patients and survivors: a systematic review and meta-analysis. J Telemed Telecare 2024;30(3):451-465. [doi: 10.1177/1357633X221078490] [Medline: 35212553]
71. Ma Z, Shi Y, Yao S, Lu N, Cheng F. Effectiveness of telemedicine-based psychosocial intervention for breast cancer patients: a systematic review and meta-analysis. Support Care Cancer 2023;31(10):595. [doi: 10.1007/s00520-023-08052-3] [Medline: 37768402]
72. Akingbade O, Nguyen KT, Chow KM. Effect of mHealth interventions on psychological issues experienced by women undergoing chemotherapy for breast cancer: a systematic review and meta-analysis. J Clin Nurs 2023;32(13-14):3058-3073. [doi: 10.1111/jocn.16533] [Medline: 36168199]
73. McCaughan E, Parahoo K, Hueter I, Northouse L, Bradbury I. Online support groups for women with breast cancer. Cochrane Database Syst Rev 2017;3(3):CD011652 [FREE Full text] [doi: 10.1002/14651858.CD011652.pub2] [Medline: 28278559]
74. Qin M, Chen B, Sun S, Liu X. Effect of mobile phone app-based interventions on quality of life and psychological symptoms among adult cancer survivors: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(12):e39799 [FREE Full text] [doi: 10.2196/39799] [Medline: 36534460]
75. Kim SH, Sung JH, Yoo S, Kim S, Lee K, Oh EG, et al. Effects of digital self-management symptom interventions on symptom outcomes in adult cancer patients: a systematic review and meta-analysis. Eur J Oncol Nurs 2023;66:102404. [doi: 10.1016/j.ejon.2023.102404] [Medline: 37517339]
76. Xu A, Wang Y, Wu X. Effectiveness of e-health based self-management to improve cancer-related fatigue, self-efficacy and quality of life in cancer patients: systematic review and meta-analysis. J Adv Nurs 2019;75(12):3434-3447. [doi: 10.1111/jan.14197] [Medline: 31566769]
77. Shaffer KM, Turner KL, Siwik C, Gonzalez BD, Upasani R, Glazer JV, et al. Digital health and telehealth in cancer care: a scoping review of reviews. Lancet Digit Health 2023;5(5):e316-e327. [doi: 10.1016/S2589-7500(23)00049-3] [Medline: 37100545]
78. Cheng L, Chen H, Lin L, Li H, Zhang F. Spiritual needs of older adults with cancer: a modified concept analysis. Asia Pac J Oncol Nurs 2023;10(11):100288 [FREE Full text] [doi: 10.1016/j.apjon.2023.100288] [Medline: 38023729]
79. Panzeri A, Komici K, Cerutti P, Sacco D, Pistono M, Rossi Ferrario S. Gender differences and long-term outcome of over 75 elderlies in cardiac rehabilitation: highlighting the role of psychological and physical factors through a secondary analysis of a cohort study. Eur J Phys Rehabil Med 2021;57(2):288-297 [FREE Full text] [doi: 10.23736/S1973-9087.21.06484-4] [Medline: 33448752]
80. Balestroni G, Panzeri A, Omarini P, Cerutti P, Sacco D, Giordano A, et al. Psychophysical health of elderly inpatients in cardiac rehabilitation: a retrospective cohort study. Eur J Phys Rehabil Med 2020;56(2):197-205 [FREE Full text] [doi: 10.23736/S1973-9087.20.05970-5] [Medline: 31976637]
81. Lee K, Kim S, Kim SH, Yoo S, Sung JH, Oh EG, et al. Digital health interventions for adult patients with cancer evaluated in randomized controlled trials: scoping review. J Med Internet Res 2023;25:e38333 [FREE Full text] [doi: 10.2196/38333] [Medline: 36607712]
82. Lim CT, Rosenfeld LC, Nissen NJ, Wang PS, Patel NC, Powers BW, et al. Remote care management for older adult populations with elevated prevalence of depression or anxiety and comorbid chronic medical illness: a systematic review. J Acad Consult Liaison Psychiatry 2022;63(3):198-212 [FREE Full text] [doi: 10.1016/j.jaclp.2022.02.005] [Medline: 35189427]
83. Hasnan S, Aggarwal S, Mohammadi L, Koczwara B. Barriers and enablers of uptake and adherence to digital health interventions in older patients with cancer: a systematic review. J Geriatr Oncol 2022;13(8):1084-1091. [doi: 10.1016/j.jgo.2022.06.004] [Medline: 35752605]
84. Bolle S, Romijn G, Smets EMA, Loos EF, Kunneman M, van Weert JCM. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
85. Rodler S, Buchner A, Stief CG, Heinemann V, Staehler M, Casuscelli J. Patients' perspective on digital technologies in advanced genitourinary cancers. Clin Genitourin Cancer 2021;19(1):76-82.e6. [doi: 10.1016/j.clgc.2020.03.018] [Medline: 32527682]
86. Pang NQ, Lau J, Fong SY, Wong CYH, Tan KK. Telemedicine acceptance among older adult patients with cancer: scoping review. J Med Internet Res 2022;24(3):e28724 [FREE Full text] [doi: 10.2196/28724] [Medline: 35348462]
87. Loh KP, Ramsdale E, Culakova E, Mendler JH, Liesveld JL, O'Dwyer KM, et al. Novel mHealth app to deliver geriatric assessment-driven interventions for older adults with cancer: pilot feasibility and usability study. JMIR Cancer 2018;4(2):e10296 [FREE Full text] [doi: 10.2196/10296] [Medline: 30373733]
88. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review. BMJ Open 2022;12(3):e054561 [FREE Full text] [doi: 10.1136/bmjopen-2021-054561] [Medline: 35264349]

XSL•FO
RenderX

89. Lee ARYB, Leong I, Lau G, Tan AW, Ho RCM, Ho CSH, et al. Depression and anxiety in older adults with cancer: systematic review and meta-summary of risk, protective and exacerbating factors. Gen Hosp Psychiatry 2023;81:32-42. [doi: 10.1016/j.genhosppsych.2023.01.008] [Medline: 36724695]
90. van Deudekom FJ, Schimberg AS, Kallenberg MH, Slingerland M, van der Velden L, Mooijaart SP. Functional and cognitive impairment, social environment, frailty and adverse health outcomes in older patients with head and neck cancer, a systematic review. Oral Oncol 2017;64:27-36 [FREE Full text] [doi: 10.1016/j.oraloncology.2016.11.013] [Medline: 28024721]
91. Hoang P, King JA, Moore S, Moore K, Reich K, Sidhu H, et al. Interventions associated with reduced loneliness and social isolation in older adults: a systematic review and meta-analysis. JAMA Netw Open 2022;5(10):e2236676 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.36676] [Medline: 36251294]
92. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: systematic review. JMIR Med Inform 2020;8(8):e20359 [FREE Full text] [doi: 10.2196/20359] [Medline: 32784177]
93. Hwang M, Jiang Y. Personalization in digital health interventions for older adults with cancer: a scoping review. J Geriatr Oncol 2023;14(8):101652. [doi: 10.1016/j.jgo.2023.101652] [Medline: 37866009]
94. Tian Q, Xu M, Yu L, Yang S, Zhang W. The efficacy of virtual reality-based interventions in breast cancer-related symptom management: a systematic review and meta-analysis. Cancer Nurs 2023;46(5):E276-E287. [doi: 10.1097/NCC.0000000000001099] [Medline: 37607378]
95. Caminiti C, Annunziata MA, Di Giulio P, Isa L, Mosconi P, Nanni MG, et al. Psychosocial impact of virtual cancer care through technology: a systematic review and meta-analysis of randomized controlled trials. Cancers 2023;15(7):2090 [FREE Full text] [doi: 10.3390/cancers15072090] [Medline: 37046750]
96. Cunningham A, McPolin O, Fallis R, Coyle C, Best P, McKenna G. A systematic review of the use of virtual reality or dental smartphone applications as interventions for management of paediatric dental anxiety. BMC Oral Health 2021;21(1):244 [FREE Full text] [doi: 10.1186/s12903-021-01602-3] [Medline: 33962624]
97. Chow CHT, Van Lieshout RJ, Schmidt LA, Dobson KG, Buckley N. Systematic review: audiovisual interventions for reducing preoperative anxiety in children undergoing elective surgery. J Pediatr Psychol 2016;41(2):182-203 [FREE Full text] [doi: 10.1093/jpepsy/jsv094] [Medline: 26476281]
98. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR mHealth uHealth 2020;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
99. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
100. Liu N, Yin J, Tan S, Ngiam K, Teo H. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. J Am Med Inform Assoc 2021;28(11):2483-2501 [FREE Full text] [doi: 10.1093/jamia/ocab151] [Medline: 34472601]
101. Krueger RF, Eaton NR. Transdiagnostic factors of mental disorders. World Psychiatry 2015;14(1):27-29 [FREE Full text] [doi: 10.1002/wps.20175] [Medline: 25655146]
102. Mah K, Tran KT, Gauthier LR, Rodin G, Zimmermann C, Warr D, et al. Do correlates of pain-related stoicism and cautiousness differ in younger and older people with advanced cancer? J Pain 2018;19(3):301-316 [FREE Full text] [doi: 10.1016/j.jpain.2017.11.002] [Medline: 29155166]
103. Perez F, Hernandez M, Martinez A, Castaneda P, Ponce R, Gonzalez M, et al. Promotores' perspectives on the virtual adaptation of a hereditary breast cancer education program. J Genet Couns 2023;32(6):1226-1231 [FREE Full text] [doi: 10.1002/jgc4.1793] [Medline: 37747056]
104. Rothmund M, Pilz MJ, Egeter N, Lidington E, Piccinin C, Arraras JI, et al. Patient-reported outcome measures for emotional functioning in cancer patients: content comparison of the EORTC CAT core, FACT-G, HADS, SF-36, PRO-CTCAE, and PROMIS instruments. Psychooncology 2023;32(4):628-639. [doi: 10.1002/pon.6109] [Medline: 36707461]
105. Ringwald J, Marwedel L, Junne F, Ziser K, Schäffeler N, Gerstner L, et al. Demands and needs for psycho-oncological eHealth interventions in women with cancer: cross-sectional study. JMIR Cancer 2017;3(2):e19 [FREE Full text] [doi: 10.2196/cancer.7973] [Medline: 29175813]
106. Marciano L, Vocaj E, Bekalu MA, La Tona A, Rocchi G, Viswanath K. The use of mobile assessments for monitoring mental health in youth: umbrella review. J Med Internet Res 2023;25:e45540 [FREE Full text] [doi: 10.2196/45540] [Medline: 37725422]
107. Widnall E, Grant CE, Wang T, Cross L, Velupillai S, Roberts A, et al. User perspectives of mood-monitoring apps available to young people: qualitative content analysis. JMIR Mhealth Uhealth 2020;8(10):e18140 [FREE Full text] [doi: 10.2196/18140] [Medline: 33037875]
108. Gratzer D, Khalid-Khan F. Internet-delivered cognitive behavioural therapy in the treatment of psychiatric illness. CMAJ 2016;188(4):263-272 [FREE Full text] [doi: 10.1503/cmaj.150007] [Medline: 26527829]
109. Sequeira L, Perrotta S, LaGrassa J, Merikangas K, Kreindler D, Kundur D, et al. Mobile and wearable technology for monitoring depressive symptoms in children and adolescents: a scoping review. J Affect Disord 2020;265:314-324. [doi: 10.1016/j.jad.2019.11.156] [Medline: 32090755]

XSL•FO
RenderX

110. Poliani A, Gnecchi S, Villa G, Rosa D, Manara DF. Gamification as an educational approach for oncological patients: a systematic scoping review. Healthcare 2023;11(24):3116 [FREE Full text] [doi: 10.3390/healthcare11243116] [Medline: 38132006]
111. Ning Y, Jia Z, Zhu R, Ding Y, Wang Q, Han S. Effect and feasibility of gamification interventions for improving physical activity and health-related outcomes in cancer survivors: an early systematic review and meta-analysis. Support Care Cancer 2022;31(1):92. [doi: 10.1007/s00520-022-07550-0] [Medline: 36585478]
112. Peyrachon R, Rébillard A. Effects of active video games in patients with cancer: systematic review. JMIR Cancer 2023;9:e45037 [FREE Full text] [doi: 10.2196/45037] [Medline: 37234028]
113. Yen HY, Chiu HL. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
114. Koivisto J, Malik A. Gamification for older adults: a systematic literature review. Gerontologist 2021;61(7):e360-e372 [FREE Full text] [doi: 10.1093/geront/gnaa047] [Medline: 32530026]
115. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023;25:e48568 [FREE Full text] [doi: 10.2196/48568] [Medline: 37379067]
116. Davis J, Van Bulck L, Durieux BN, Lindvall C. The temperature feature of ChatGPT: modifying creativity for clinical research. JMIR Hum Factors 2024;11:e53559 [FREE Full text] [doi: 10.2196/53559] [Medline: 38457221]
117. Luo X, Chen Y, Chen J, Zhang Y, Li M, Xiong C, et al. Effectiveness of mobile health-based self-management interventions in breast cancer patients: a meta-analysis. Support Care Cancer 2022;30(3):2853-2876. [doi: 10.1007/s00520-021-06568-0] [Medline: 34561732]
118. Pritchett JC, Patt D, Thanarajasingam G, Schuster A, Snyder C. Patient-reported outcomes, digital health, and the quest to improve health equity. Am Soc Clin Oncol Educ Book 2023;43:e390678 [FREE Full text] [doi: 10.1200/EDBK_390678] [Medline: 37290027]
119. Marzo RR, Chen HWJ, Abid K, Chauhan S, Kaggwa MM, Essar MY, et al. Adapted digital health literacy and health information seeking behavior among lower income groups in Malaysia during the COVID-19 pandemic. Front Public Health 2022;10:998272 [FREE Full text] [doi: 10.3389/fpubh.2022.998272] [Medline: 36187682]
120. Raza MM, Venkatesh KP, Kvedar JC. Promoting racial equity in digital health: applying a cross-disciplinary equity framework. NPJ Digit Med 2023;6(1):3 [FREE Full text] [doi: 10.1038/s41746-023-00747-5] [Medline: 36631555]
121. Choukou M, Sanchez-Ramirez DC, Pol M, Uddin M, Monnin C, Syed-Abdul S. COVID-19 infodemic and digital health literacy in vulnerable populations: a scoping review. Digit Health 2022;8:20552076221076927 [FREE Full text] [doi: 10.1177/20552076221076927] [Medline: 35223076]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
iCBT: internet cognitive behavioral therapy
PICOS: Population, Intervention, Comparison, Outcome, and Study
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life

Edited by A Mavragani; submitted 04.12.24; peer-reviewed by S Kondapally, RS Gomaa Mahmoud, S Babatope, S Ajayi; comments to author 27.12.24; revised version received 15.01.25; accepted 20.01.25; published 21.02.25.

Please cite as:
Zhong C, Luo X, Tan M, Chi J, Guo B, Tang J, Guo Z, Deng S, Zhang Y, Wu Y
Digital Health Interventions to Improve Mental Health in Patients With Cancer: Umbrella Review
J Med Internet Res 2025;27:e69621
URL: https://www.jmir.org/2025/1/e69621
doi:10.2196/69621
PMID:

©Chuhan Zhong, Xian Luo, Miaoqin Tan, Jing Chi, Bingqian Guo, Jianyao Tang, Zihan Guo, Shisi Deng, Yujie Zhang, Yanni Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the

XSL•FO
RenderX

Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Reporting Quality of AI Intervention in Randomized Controlled Trials in Primary Care: Systematic Review and Meta-Epidemiological Study

Jinjia Zhong[1], BMed; Ting Zhu[1], BMed; Yafang Huang[1], PhD

School of General Practice and Continuing Education, Capital Medical University, Beijing, China

Corresponding Author:
Yafang Huang, PhD
School of General Practice and Continuing Education
Capital Medical University
4th Fl, Jieping Building, Capital Medical University
No.10 You An Men Wai Xi Tou Tiao, Fengtai district
Beijing, 100069
China
Phone: 86 18810673886
Fax: 86 1083911501
Email: yafang@ccmu.edu.cn

Abstract

Background: The surge in artificial intelligence (AI) interventions in primary care trials lacks a study on reporting quality.

Objective: This study aimed to systematically evaluate the reporting quality of both published randomized controlled trials (RCTs) and protocols for RCTs that investigated AI interventions in primary care.

Methods: PubMed, Embase, Cochrane Library, MEDLINE, Web of Science, and CINAHL databases were searched for RCTs and protocols on AI interventions in primary care until November 2024. Eligible studies were published RCTs or full protocols for RCTs exploring AI interventions in primary care. The reporting quality was assessed using CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) and SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) checklists, focusing on AI intervention–related items.

Results: A total of 11,711 records were identified. In total, 19 published RCTs and 21 RCT protocols for 35 trials were included. The overall proportion of adequately reported items was 65% (172/266; 95% CI 59%-70%) and 68% (214/315; 95% CI 62%-73%) for RCTs and protocols, respectively. The percentage of RCTs and protocols that reported a specific item ranged from 11% (2/19) to 100% (19/19) and from 10% (2/21) to 100% (21/21), respectively. The reporting of both RCTs and protocols exhibited similar characteristics and trends. They both lack transparency and completeness, which can be summarized in three aspects: without providing adequate information regarding the input data, without mentioning the methods for identifying and analyzing performance errors, and without stating whether and how the AI intervention and its code can be accessed.

Conclusions: The reporting quality could be improved in both RCTs and protocols. This study helps promote the transparent and complete reporting of trials with AI interventions in primary care.

(J Med Internet Res 2025;27:e56774) doi:10.2196/56774

KEYWORDS

artificial intelligence; randomized controlled trial; reporting quality; primary care; meta-epidemiological study

Introduction

Primary care provides a large health care delivery platform for primary care physicians to offer person-centered services for a vast patient population [1]. The practice of primary care generates substantial digital data that can be leveraged for primary care research [2,3]. The growing resources of big data, mainly consisting of electronic health records, electronic medical data, patient self-reported data, and wearable device data, have paved the way for the advancement of artificial intelligence (AI) in deep machine learning (ML) technology and natural language processing to explore AI- or ML-driven clinical tools dedicated to the improvement of clinical decision-making in the disease screening, diagnosis, prognosis, and management

XSL•FO
RenderX

[4-6]. For example, several studies in primary care have developed the prediction tool for patients presenting with acute cough, the prediction score for head and neck cancer referrals, and the prediction model for asthma exacerbation [7-9]. Nevertheless, in practice, a significant proportion of these tools have not undergone validation through a robust clinical trial, although the majority of these studies use performance indicators to showcase their superiority [10,11].

Randomized controlled trials (RCTs) are widely recognized as the gold standard for evaluating interventions [12]. Many clinical studies that investigate AI- or ML-driven clinical tools as a clinical intervention conduct a confirmatory RCT, aiming to demonstrate the clinical significance [13-15]. Transparency and explainability are essential for the widespread integration of AI systems into clinical practice, as an inaccurate prediction could result in severe consequences [16]. Given the inherent complexity of AI interventions, it is critical for RCT reporting to be comprehensive and transparent. Adhering to reporting guidelines for AI in clinical trials and transparently reporting AI interventions in RCTs are important steps toward improving research quality, fostering scientific discourse, and establishing more reliable foundations for clinical practice and decision-making [17-19].

The CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension [17] and SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) extension [18] are newly developed reporting guidelines specifically focusing on AI intervention in trials. CONSORT-AI focuses more on trial results and reporting completeness, while SPIRIT-AI emphasizes pretrial elements, such as trial design and ethical considerations. The relative items in the guidelines should be transparently and completely reported, and qualified reporting is essential for independently evaluating and replicating the trial. However, there is no systematic review and critical appraisal of RCTs for examining AI interventions in primary care, leaving the clarity of their reporting quality uncertain.

This systematic review and meta-epidemiological study aims to evaluate the reporting quality of AI interventions in published RCTs and protocols for RCTs in primary care. The assessment was based on the CONSORT-AI extension and SPIRIT-AI extension guidelines.

Methods

Study Design

This study was reported in accordance with the guidelines for reporting meta-epidemiological methodology research [20] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) reporting guidelines (Multimedia Appendix 1). The protocol for this study has been registered on PROSPERO (ID: CRD42023427694).

Search Strategy

The searches of PubMed, Embase, and Cochrane Library were conducted up to November 30, 2024. To ensure a comprehensive literature search, we further referenced a previously published study strategy [3], ensuring the completeness and accuracy of the process of collecting RCTs. This further search was applied across MEDLINE, Web of Science, and CINAHL. There were no restrictions on the year of publication or language. The detailed search strategies are listed in Multimedia Appendix 2.

Study Selection

Overview

To be included in this study, articles had to meet the following criteria: (1) RCTs or protocols for RCTs that used AI interventions or AI-assisted tools to guide a randomized intervention and (2) studies that belonged to primary care research. Based on the 1996 report of the US Institute of Medicine, primary care was defined as “the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community” [21]. Primary care research was defined as “Research done in a primary care context” [22]. Only published RCTs and published, full RCT protocols were included. We excluded (1) studies with abstracts only; (2) commentaries, letters, editorials, or reviews; and (3) animals or preclinical studies.

Stage 1: Title and Abstract Screening

The retrieved records were imported into EndNote 20 (Clarivate). Two independent reviewers (JZ and TZ) conducted the screening of titles and abstracts to identify studies potentially meeting the inclusion criteria. Any discrepancies were resolved through discussion.

Stage 2: Full-Text Screening

Using EndNote 20, the reviewers (JZ and TZ) independently assessed the full-text articles identified at stage 1 for eligibility. Any discrepancies were discussed and resolved by consulting the third researcher (YH), and consensus was reached through discussion. Studies meeting the eligibility criteria were included for data extraction.

Data Extraction Process and Data Items

Two researchers (JZ and TZ) independently extracted data from each article by using a standardized data extraction form. Discrepancies in data extraction were discussed between JZ and TZ and the remaining conflicts were resolved by YH.

Data were collected on (1) trial name, (2) trial registration identifier assigned to clinical trials registered in the database, (3) first author’s name, (4) published article name, (5) journal name, (6) publication year, (7) research topic, (8) study design (sample size, blinding, and type of comparator arm), (9) primary outcome, (10) classification of primary outcomes by result status (positive or negative), (11) type of AI model (large language model [LLM], ML, deep learning, clinical decision support system [CDSS], or risk prediction model), and (12) deployment context (clinician-assisted decision support, fully automated diagnostic systems, or patient self-management tools).

The reporting quality of included studies was assessed using the CONSORT-AI and SPIRIT-AI checklists. Each checklist consists of two components: (1) the original CONSORT 2010/SPIRIT 2013 items and (2) the AI-specific elaborations

XSL•FO
RenderX

and extension items. This study focused on evaluating the AI-related reporting quality. Only items that are separately elaborated or extended in the CONSORT-AI (14 items) and SPIRIT-AI (15 items) checklists were appraised. These items ensure that critical aspects of AI-supported interventions such as input-data setting are transparently and completely reported. Each item was assessed in all included published articles, as "Yes" or "No."

Outcomes

The primary outcomes were as follows: (1) the proportion of adequately reported items, calculated by dividing the number of adequately reported items by the total number of items in each article—high proportions would indicate high reproducibility and quality of the RCTs and protocols for RCTs—and (2) the percentage of articles adequately reporting each item, calculated by dividing the number of articles that adequately reported the item by the total number of articles evaluated for that particular item. Independent analysis was conducted for CONSORT-AI (14 items) and SPIRIT-AI (15 items) checklists based on the primary outcomes described above. Since CONSORT-AI focuses on reporting completed trials and SPIRIT-AI on trial protocol standards, analyzing them separately ensures more precise adherence to each guideline. The secondary outcomes focused on CONSORT-AI included (1) the association between primary outcome status and the reporting quality of each CONSORT-AI item, and (2) factors influencing the percentage of adequately reported items in CONSORT-AI.

Data Synthesis and Analysis

Descriptive statistical methods were used and presented as frequencies, median percentages, and IQRs. The proportion of adequately reported items and the percentage of articles adequately reporting each specific item was reported, with its corresponding 95% CI calculated using the Clopper-Person method. The analyses were performed using Microsoft Excel 2021 and OriginPro 2022 (OriginLab Corporation). The association between primary outcome result status (positive or negative) and the reporting quality of each item was examined using the Fisher exact test. The percentage of adequately reported items was compared across RCT characteristics such as type of disease using the Mann-Whitney *U* test. To assess the main effects and interaction effects between the primary outcome result status and various RCT characteristics on the percentage of adequately reported items, the Scheirer-Ray-Hare test was used. All *P* values were derived from 2-sided tests, with significance defined as $P<.05$ (R, version 4.4.2; R Foundation for Statistical Computing).

Results

Overview

The search strategies retrieved a total of 11,711 records. After screening, 40 articles finally met the inclusion criteria. These 40 full-text articles corresponded to 35 trials, including 19 RCTs and 21 protocols for RCTs (Figure 1 and Multimedia Appendix 3 [23-62]).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram of the selection procedure. RCT: randomized controlled trial.

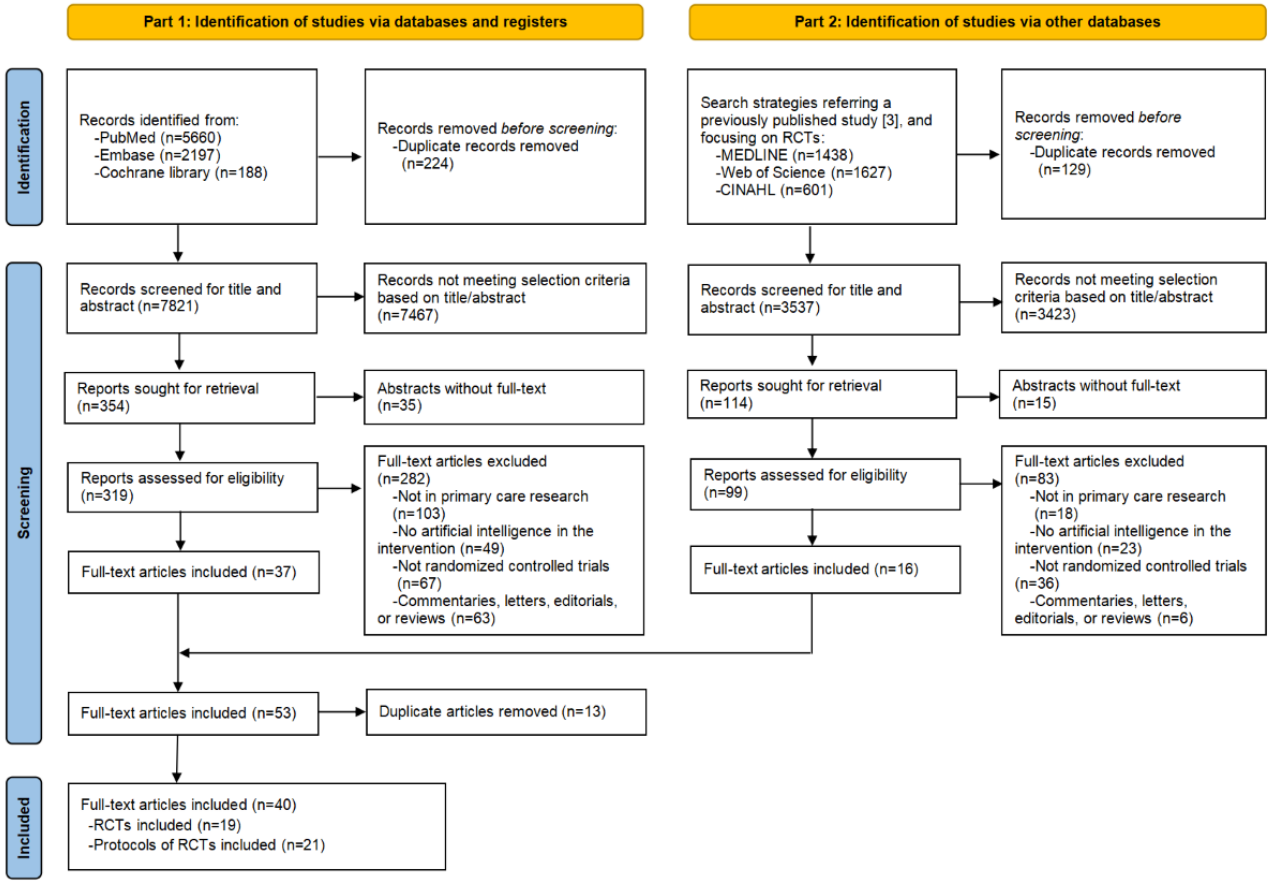

Characteristics of Included Articles

Table 1 summarizes the characteristics of the included articles. Among the 19 RCTs, 8 (42%) used LLMs, 10 (53%) were conducted in the context of clinician-assisted decision support, and 11 (58%) reported positive primary outcome results. In total, 21% (4/19) focused on cardiovascular topics. The median sample size for these RCTs was 335 (IQR 133-487) participants. Among the 21 protocols, 6 (29%) incorporated LLMs, 11 (52%) involved clinician-assisted decision support, and 6 (29%) focused on cardiovascular topics (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of included published randomized controlled trials (RCTs) and protocols for RCTs.

Characteristics	Randomized controlled trials (n=19)	Protocols (n=21)	Overall (n=40)
Reported items, median (range)	9 (7-14)	10 (7-13)	9 (7-14)
Publication year, n (%)			
2024	7 (37)	3 (14)	10 (25)
2023	3 (16)	2 (10)	5 (13)
2022	6 (32)	9 (43)	15 (38)
Pre-2021	3 (16)	7 (33)	10 (25)
Sample size, median (IQR)	335 (133-487)	—[a]	—
Blinding, n (%)			
Single blinded	6 (32)	2 (10)	8 (20)
Double-blinded	1 (5)	3 (14)	4 (10)
Open label	12 (63)	16 (76)	28 (70)
Type of comparator arm, n (%)			
Usual care	18 (95)	17 (81)	35 (88)
No treatment	1 (5)	3 (14)	4 (10)
Delay intervention	0 (0)	1 (5)	1 (3)
Type of artificial intelligence model, n (%)			
Large language model	8 (42)	6 (29)	14 (35)
Machine learning	2 (11)	4 (19)	6 (15)
Deep learning	6 (32)	6 (29)	12 (30)
Clinical decision support system	2 (11)	3 (14)	5 (13)
Risk prediction model	1 (5)	2 (10)	3 (8)
Deployment context, n (%)			
Clinician-assisted decision support	10 (53)	11 (52)	21 (53)
Fully automated diagnostic systems	3 (16)	1 (5)	4 (10)
Patient self-management tools	6 (32)	9 (43)	15 (38)
Classification of primary outcomes by result status, n (%)			
Positive	11 (58)	—	—
Negative	8 (42)	—	—
Disease or topic, n (%)			
Cardiovascular disease	4 (21)	6 (29)	10 (25)
Oncology	2 (11)	3 (14)	5 (13)
Pain management	4 (21)	1 (5)	5 (13)
Endocrine disorders	1 (5)	2 (10)	3 (8)
Respiratory diseases	2 (11)	1 (5)	3 (8)
Digestive disorders	2 (11)	0 (0)	2 (5)
Mental illness	1 (5)	2 (10)	3 (8)
Others	3 (16)[b]	6 (29)[c]	9 (23)

[a]Not applicable.

[b]Prescription management, diagnostic reasoning, and quitting smoking.

[c]Auxiliary diagnosis, prescription management, retinal disease, quit smoking, and drug overdose.

XSL•FO
RenderX

Proportion of Adequately Reported Items

In the 19 RCTs included in this review, the overall proportion of adequately reported items was 65% (172/266; 95% CI 59%-70%). Only 2 RCTs had more than 90% of items reported adequately. In the 21 protocols for RCTs, the overall proportion of adequately reported items was 68% (214/315; 95% CI 62%-73%). Two protocols of RCTs had more than 85% of items reported adequately (Multimedia Appendix 3).

Percentage of Articles Adequately Reporting Each Specific Item

The complete list of the 14 items of the CONSORT-AI extension is shown in Table 2. The percentage of RCTs that reported a specific item ranged from 11% to 100%. The best-reported sections were the title and abstract [item 1a and item b(ii)], background and objective [item 2a(i)], and participants (item 4b), all being reported in 100% (19/19) of RCTs. The poorly reported sections were intervention [item 5(i), item 5(ii), and item 5(iii)], participants [item 4a(ii)], harms (item 19), and funding (item 25). Among these poorly reported items, 4a(ii), 5(ii), and 5(iii) were related to providing the information for input data (Figure 2 and Table 2). The reporting quality increases as it approaches the outer edges of the radar chart—1a, b(i), 1a,b(ii): title; 2a(i): background and objectives; 4a(i), 4a(ii), 4b: participants; 5(i)-5(iv): intervention; 19: harms; and 25: funding.

Table 2. Adherence to CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension items.

Item	Item description	Total (n=19), n (%; 95%CI)
CONSORT-AI 1a Elaboration	Indicate that the intervention involves artificial intelligence or machine learning in the title and abstract and specify the type of model.	16 (84; 60-97)
CONSORT-AI 1b(ii) Elaboration	State the intended use of the AI[a] intervention within the trial in the title and abstract.	19 (100; 82-100)
CONSORT-AI 2a(i) Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (eg, health care professionals, patients, public).	19 (100; 82-100)
CONSORT-AI 4a(i) Elaboration	State the inclusion and exclusion criteria at the level of participants	18 (95; 74-100)
CONSORT-AI 4a(ii) Extension	State the inclusion and exclusion criteria at the level of the input data.	5 (26; 9-51)
CONSORT-AI 4b Extension	Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.	19 (100; 82-100)
CONSORT-AI 5(i) Extension	State which version of the AI algorithm was used.	12 (63; 38-84)
CONSORT-AI 5(ii) Extension	Describe how the input data were acquired and selected for the AI intervention.	13 (68; 43-87)
CONSORT-AI 5(iii) Extension	Describe how poor quality or unavailable input data were assessed and handled.	2 (11; 1-33)
CONSORT-AI 5(iv) Extension	Specify whether there was human-AI interaction in the handling of the input data, and what level of expertise was required of users.	15 (79; 54-94)
CONSORT-AI 5(v) Extension	Specify the output of the AI intervention.	15 (79; 54-94)
CONSORT-AI 5(vi) Extension	Explain how the AI intervention's outputs contributed to decision-making or other elements of clinical practice.	14 (74; 49-91)
CONSORT-AI 19 Extension	Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, justify why not.	2 (11; 1-33)
CONSORT-AI 25 Extension	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.	3 (16; 3-40)

[a]AI: artificial intelligence.

XSL•FO
RenderX

Figure 2. Radar chart to show the adequately reported percentage for each CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) item.

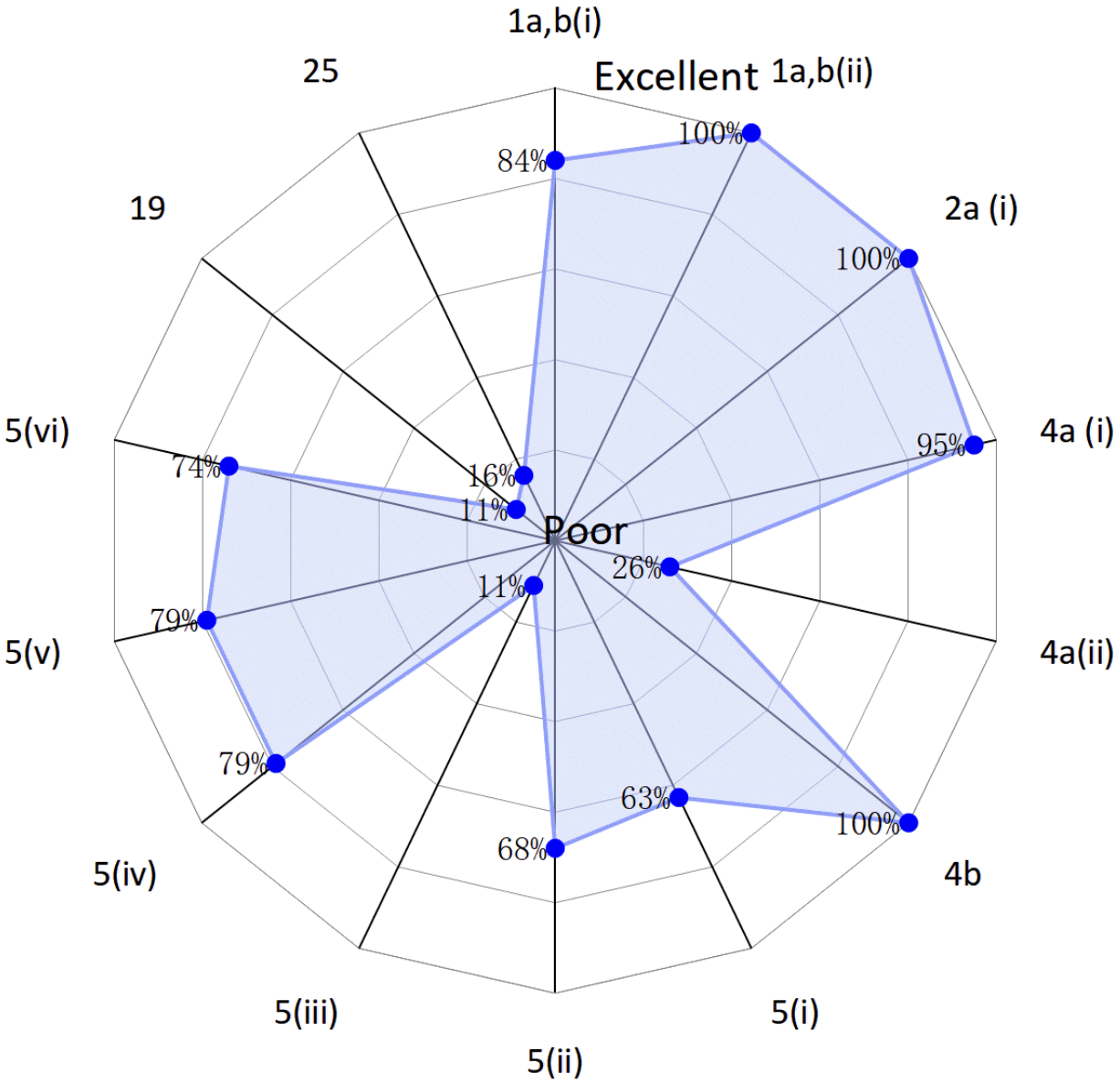

The complete list of the 15 items of the SPIRIT-AI extension is shown in Table 3. The percentage of protocols for RCTs that reported a specific item ranged from 10% to 100%. The best-reported sections were title and abstract [item 1(i) and item 1(ii)], background and rationale [item 6a(i)], and intervention [item 11a(vi)] all being reported in 100% of protocols. The poorly reported sections were intervention [item 11a(i), item 11a(iii), and item 11a(iv)], eligibility [item 10(ii)], harms (item 22), and access to data (item 29). Among these poorly reported items, 10(ii), 11a(iii), and 11a(iv) were related to providing the information for input data (Figure 3 and Table 3). The reporting quality increases as it approaches the outer edges of the radar chart—1(i), 1(ii): title; 6a(i), 6a(ii): background and rationale; 9: study setting; 10(i), 10(ii): eligibility criteria; 11a(i)-11a(vi): interventions; 22: harms; and 29: access to data.

XSL•FO
RenderX

Figure 3. Radar chart to show the adequately reported percentage for each SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) item.

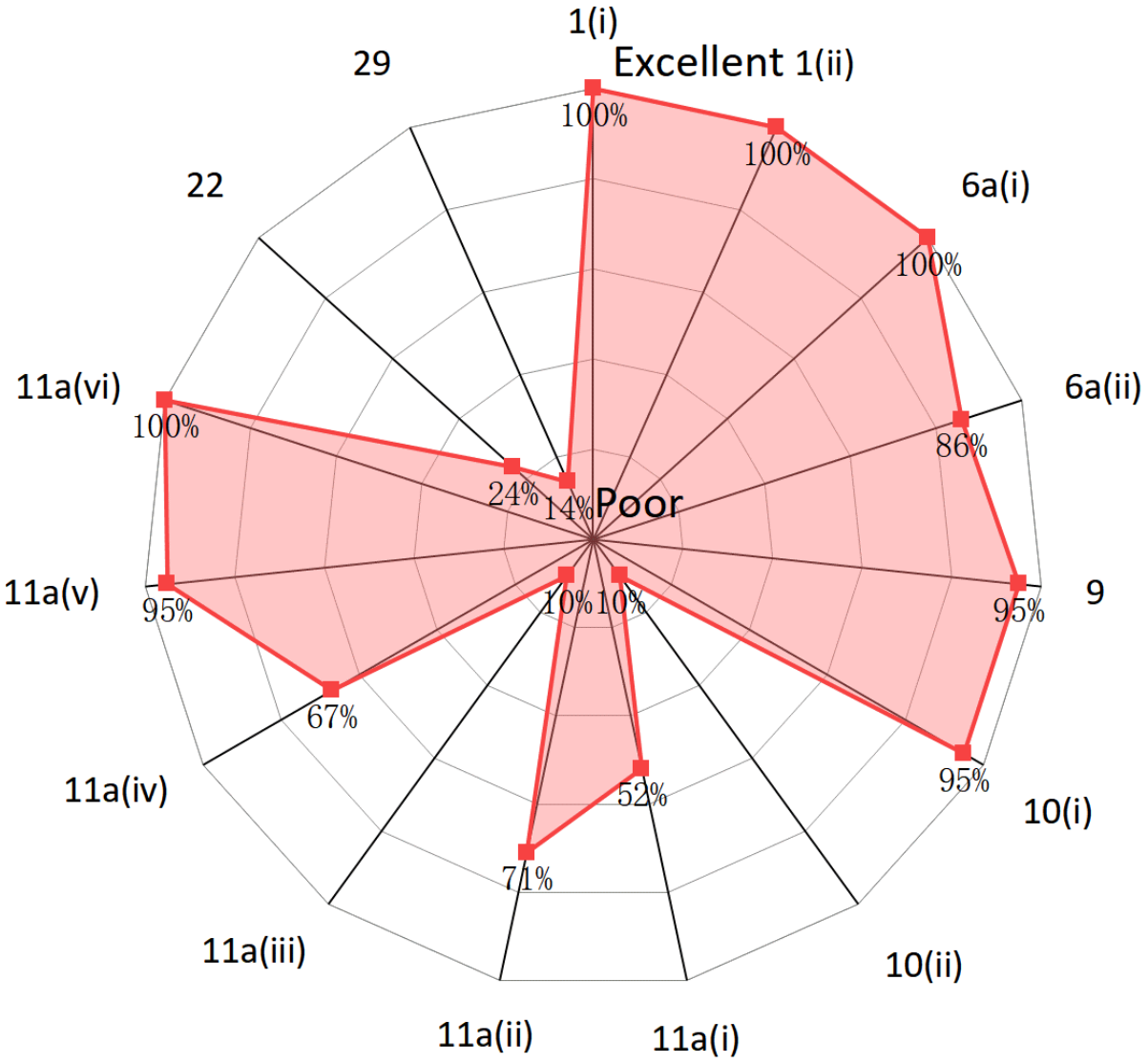

XSL•FO
RenderX

Table 3. Adherence to SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials—Artificial Intelligence) extension items.

Item	Item description	Total (n=21), n (%; 95% CI)
SPIRIT-AI 1(i) Elaboration	Indicate that the intervention involves artificial intelligence or machine learning and specify the type of model.	21 (100; 84-100)
SPIRIT-AI 1(ii) Elaboration	Specify the intended use of the AI[a] intervention.	21 (100; 84-100)
SPIRIT-AI 6a(i) Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (eg, health care professionals, patients, public).	21 (100; 84-100)
SPIRIT-AI 6a(ii) Extension	Describe any pre-existing evidence for the AI intervention.	18 (86; 64-97)
SPIRIT-AI 9 Extension	Describe the onsite and offsite requirements needed to integrate the AI intervention into the trial setting.	20 (95; 76-100)
SPIRIT-AI 10(i) Elaboration	State the inclusion and exclusion criteria at the level of participants.	20 (95; 76-100)
SPIRIT-AI 10(ii) Extension	State the inclusion and exclusion criteria at the level of the input data.	2 (10; 1-30)
SPIRIT-AI 11a(i) Extension	State which version of the AI algorithm will be used.	11 (52; 30-74)
SPIRIT-AI 11a(ii) Extension	Specify the procedure for acquiring and selecting the input data for the AI intervention.	15 (71; 48-89)
SPIRIT-AI 11a(iii) Extension	Specify the procedure for assessing and handling poor quality or unavailable input data.	2 (10; 1-30)
SPIRIT-AI 11a(iv) Extension	Specify whether there is human-AI interaction in the handling of the input data, and what level of expertise is required for users.	14 (67; 43-85)
SPIRIT-AI 11a(v) Extension	Specify the output of the AI intervention.	20 (95; 76-100)
SPIRIT-AI 11a(vi) Extension	Explain the procedure for how the AI intervention's output will contribute to decision-making or other elements of clinical practice.	21 (100; 84-100)
SPIRIT-AI 22 Extension	Specify any plans to identify and analyze performance errors. If there are no plans for this, explain why not.	5 (24; 8-47)
SPIRIT-AI 29 Extension	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.	3 (14; 3-36)

[a]AI: artificial intelligence.

Association Between Primary Outcome Result Status and the Reporting Quality of Each Item

Multimedia Appendix 4 shows the association between the primary outcome result status in RCTs and the reporting quality of each item. Fisher exact test results indicate no significant association between positive outcomes and higher reporting quality (all $P>.05$).

Factors Associated With the Percentage of Adequately Reported Items

To further explore potential differences resulting from RCT characteristics, we compared the percentage of adequately reported items by year of publication, blinding, type of AI model, deployment context, primary outcome result status, and disease type (Multimedia Appendix 5). Trials of cardiovascular disease showed a higher percentage of adequately reported items ($P=.02$). The Scheirer-Ray-Hare test did not detect statistical significance for the interaction effects between primary outcome result status and various RCT characteristics on the percentage of adequately reported items (all $P>.05$; Multimedia Appendix 5).

Discussion

Principal Results

In this study, we assessed the RCT reporting quality in published 19 RCTs and 21 RCT protocols involved in 35 trials, with 5 trials having both published RCTs and corresponding protocols. This study found the reporting quality in both RCTs and protocols for RCTs of AI intervention could be improved in 3 aspects. The first aspect concerns providing enough information for the input data, which was detailed in CONSORT-AI guidelines [items 4a(ii), 5(ii), and 5(iii)] and SPIRIT-AI guidelines [items 10(ii), 11a(iii), and 11a(iv)]. The second aspect involves the lack of information on how to identify and analyze performance errors, which was demanded by the harm section of CONSORT-AI guidelines (item 19) and SPIRIT-AI guidelines (item 22), respectively. The third aspect involves neglecting to specify whether and how the AI intervention or its code can be accessed, which was needed in CONSORT-AI guidelines (item 25) and SPIRIT-AI guidelines (item 29), respectively. Based on the evaluation of the reporting of 35 trials, our findings revealed a consistent lack of information across these three aspects across both protocols and RCTs, highlighting a widespread gap in adhering to the reporting standards recommended by these guidelines.

XSL•FO
RenderX

Strength and Limitations

This is the first study focused on the field of AI application in primary care, providing a first systematic review into the existing landscape of RCTs using AI interventions within primary care. This review not only included RCTs of AI interventions in primary care but also covered protocols, providing a more comprehensive survey in this field.

This study has 2 limitations. First, the majority of studies were conducted in the United States, which may lead to a lack of representation for other countries. Second, this study assessed the reporting quality of AI intervention–related items for RCTs. General reporting items from the original CONSORT 2010 and SPIRIT 2013 checklists were not assessed.

Comparison With Previous Work

We are aware of only one published systematic scoping review on the application of AI in community-based primary health care [3]. The authors highlighted a gap in the development and implementation of AI in primary care. However, reporting quality was not assessed previous study [3]. This systematic review is the first focusing on primary care and indicating the need for more transparent design, conduct, and analysis of AI interventions in primary care.

Implications for Research and Practice

To apply AI interventions in primary care, accurate and sufficient information is essential to guide the standardized performance of clinical treatment. The findings of this study have several implications for clinical research and practice, concerning AI safety, reliability, and reproducibility, as outlined below.

First, reporting on AI performance errors needs to be improved. As software, AI systems are likely to undergo multiple iterations and updates [10]. For example, although LLMs have successfully answered medical licensing exam questions [63], they still produce errors when stating facts or synthesizing data from medical literature [64]. Since LLMs still have many flaws, performance error reporting is essential for continuous updating. The lack of performance error reporting may lead to misdiagnosis or incorrect treatment recommendations, ultimately jeopardizing patient safety [65]. It is crucial to specify the types of performance errors, the process of identifying them, and how updated versions correct these errors during the trial.

Second, the reporting on the section of input-data setting deserves special attention. AI interventions applied to primary care should provide comprehensive information for input-data settings, including parameter selection and preprocessing before analysis by the AI system. Moreover, during the input-data setting, the inclusion and exclusion criteria should be clearly defined, along with an adequate investigation of patient characteristics relevant to the disease type. It is necessary to report these items completely to ensure the replicability of the intervention beyond the trials in real-world circumstances [19]. It also supports investigators in identifying whether input-data-handling procedures were standardized across trial sites [17]. Overall, the AI system should establish an optimal input set that includes a wide range of parameters and attribute combinations [66]. For example, ML prediction models can only accurately predict the specific outcomes after knee arthroplasty while the ability to predict more complex outcomes remains inaccuracy [67]. Thus, ML requires comprehensive patient-related indicators and the identification of specific patterns that are suitable for ML analysis. This underscores the importance of high-quality input data setting as a fundamental prerequisite for optimizing ML capabilities in clinical applications. Finally, trials should report the prespecified conditions, particularly if minimum requirements for input data are not met. Failing to transparently report how input data were acquired and selected can compromise the representativeness and generalizability of the AI model, potentially leading to critical errors in AI-driven decision-making, which may adversely affect patient safety and clinical outcomes.[17]

In particular, reporting on assessment and handling of input data needs to be improved. The reporting of AI intervention should mention the amount of poor-quality input data, as well as how this was identified and handled [17]. Poor-quality or unavailable input data can negatively impact the success of AI algorithms and the AI's performance and effectiveness. It is essential to have a plan in place for handling such scenarios, such as implementing data cleaning techniques, using data imputation methods, or seeking alternative data sources.[10] A study on cardiovascular imaging pointed out that the quality and preprocessing of input data such as echocardiography, cardiac magnetic resonance imaging, and cardiac computed tomography should be evaluated before a model can be trained using deep learning techniques [68]. Otherwise, the black-box nature of AI models may lead to a lack of trust among clinicians and sponsors [69,70].

Finally, reporting regarding access to and reuse of AI intervention and its code needs to be specified. Human-AI interaction promotes the capabilities of AI systems, as it leverages human expertise to ensure the practical application of AI models. A review on ML highlights three reasons for the necessity of human involvement: real-world problems are inherently complex, ML methods often lack explainability, and AI outputs may not always align with clinical expectations or disease judgment [71]. It is imperative that health care professionals can validate and refine AI methodologies to ensure they are clinically applicable [72,73]. Therefore, specifying whether and how the AI intervention and its underlying code can be accessed by designated health care professionals is crucial. Clear reporting on the accessibility and transparency of AI systems further helps clinicians understand their functionality, enhancing trust in their outputs and fostering the development of user-friendly interfaces, and ensuring the safe and effective use of AI in clinical practice.

Conclusion

Our study focused on AI intervention in primary care, providing a first systematic review of qualified reporting of RCTs using AI interventions. This study indicated significant gaps between the reporting guidelines and published articles and underscored the crucial items and aspects that were frequently overlooked in the reporting framework. It is said that the whole of medicine depends on the transparent reporting of clinical trials [74]. Our

XSL•FO
RenderX

findings may contribute to the enhancement of quality standards in AI research in primary care trials and help future clinical AI investigators to design, conduct, and analysis of higher-quality AI interventions for primary care.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82104133).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[PDF File (Adobe PDF File), 93 KB - jmir_v27i1e56774_app1.pdf]

Multimedia Appendix 2
Search strategies.
[DOCX File , 27 KB - jmir_v27i1e56774_app2.docx]

Multimedia Appendix 3
List and characteristics of included randomized controlled trials and protocols.
[DOCX File , 55 KB - jmir_v27i1e56774_app3.docx]

Multimedia Appendix 4
Association between primary outcome result status and the reporting quality of each item.
[DOCX File , 17 KB - jmir_v27i1e56774_app4.docx]

Multimedia Appendix 5
Association between trial characteristics and the percentage of adequately reported items.
[DOCX File , 17 KB - jmir_v27i1e56774_app5.docx]

References

1. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2016;3(3):CD006560 [FREE Full text] [doi: 10.1002/14651858.CD006560.pub4] [Medline: 33448337]
2. Rodriguez JA, Charles J, Bates DW, Lyles C, Southworth B, Samal L. Digital healthcare equity in primary care: implementing an integrated digital health navigator. J Am Med Inform Assoc 2023;30(5):965-970 [FREE Full text] [doi: 10.1093/jamia/ocad015] [Medline: 36795062]
3. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. J Med Internet Res 2021;23(9):e29839 [FREE Full text] [doi: 10.2196/29839] [Medline: 34477556]
4. Tse G, Lee Q, Chou OHI, Chung CT, Lee S, Chan JSK, et al. Healthcare big data in Hong Kong: development and implementation of artificial intelligence-enhanced predictive models for risk stratification. Curr Probl Cardiol 2024;49(1 Pt B):102168. [doi: 10.1016/j.cpcardiol.2023.102168] [Medline: 37871712]
5. Uddin Y, Nair A, Shariq S, Hannan SH. Transforming primary healthcare through natural language processing and big data analytics. BMJ 2023;381:948. [doi: 10.1136/bmj.p948] [Medline: 37137492]
6. Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, Ghadiri P, Grad R, Poenaru D, et al. Application of artificial intelligence in shared decision making: scoping review. JMIR Med Inform 2022;10(8):e36199 [FREE Full text] [doi: 10.2196/36199] [Medline: 35943793]
7. Lau K, Wilkinson J, Moorthy R. A web-based prediction score for head and neck cancer referrals. Clin Otolaryngol 2018;43(4):1043-1049. [doi: 10.1111/coa.13098] [Medline: 29543399]
8. Lisspers K, Ställberg B, Larsson K, Janson C, Müller M, Łuczko M, et al. Developing a short-term prediction model for asthma exacerbations from Swedish primary care patients' data using machine learning - based on the ARCTIC study. Respir Med 2021;185:106483 [FREE Full text] [doi: 10.1016/j.rmed.2021.106483] [Medline: 34077873]

XSL•FO
RenderX

9. Bruyndonckx R, Hens N, Verheij TJ, Aerts M, Ieven M, Butler CC, et al. Development of a prediction tool for patients presenting with acute cough in primary care: a prognostic study spanning six European countries. Br J Gen Pract 2018;68(670):e342-e350 [FREE Full text] [doi: 10.3399/bjgp18X695789] [Medline: 29632005]
10. Yang Z, Silcox C, Sendak M, Rose S, Rehkopf D, Phillips R, et al. Advancing primary care with artificial intelligence and machine learning. Healthc (Amst) 2022;10(1):100594. [doi: 10.1016/j.hjdsi.2021.100594] [Medline: 34954571]
11. Mohsen F, Al-Absi HRH, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. NPJ Digit Med 2023;6(1):197 [FREE Full text] [doi: 10.1038/s41746-023-00933-5] [Medline: 37880301]
12. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. BMJ 1996;313(7057):570-571 [FREE Full text] [doi: 10.1136/bmj.313.7057.570] [Medline: 8806240]
13. Wu L, Shang R, Sharma P, Zhou W, Liu J, Yao L, et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomised controlled trial. Lancet Gastroenterol Hepatol 2021;6(9):700-708. [doi: 10.1016/S2468-1253(21)00216-8] [Medline: 34297944]
14. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol 2019;20(12):1645-1654. [doi: 10.1016/S1470-2045(19)30637-0] [Medline: 31591062]
15. Mangas-Sanjuan C, de-Castro L, Cubiella J, Díez-Redondo P, Suárez A, Pellisé M, et al. Role of artificial intelligence in colonoscopy detection of advanced neoplasias : a randomized trial. Ann Intern Med 2023;176(9):1145-1152. [doi: 10.7326/M22-2619] [Medline: 37639723]
16. Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat Commun 2020;11(1):3852 [FREE Full text] [doi: 10.1038/s41467-020-17431-x] [Medline: 32737308]
17. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health 2020;2(10):e537-e548 [FREE Full text] [doi: 10.1016/S2589-7500(20)30218-1] [Medline: 33328048]
18. Rivera SC, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health 2020;2(10):e549-e560. [doi: 10.1016/S2589-7500(20)30219-3] [Medline: 33328049]
19. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
20. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med 2017;22(4):139-142 [FREE Full text] [doi: 10.1136/ebmed-2017-110713] [Medline: 28701372]
21. Institute of Medicine (US) Committee on the Future of Primary Care. In: Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, editors. Primary Care: America's Health in a New Era. Washington, DC: National Academies Press; 1996.
22. Starfield B. A framework for primary care research. J Fam Pract 1996;42(2):181-185. [Medline: 8606309]
23. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. JAMA Netw Open 2020;3(3):e200255 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0255] [Medline: 32119093]
24. Seol HY, Shrestha P, Muth JF, Wi CI, Sohn S, Ryu E, et al. Artificial intelligence-assisted clinical decision support for childhood asthma management: a randomized clinical trial. PLoS One 2021;16(8):e0255261 [FREE Full text] [doi: 10.1371/journal.pone.0255261] [Medline: 34339438]
25. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. Nat Med 2021;27(5):815-819. [doi: 10.1038/s41591-021-01335-4] [Medline: 33958795]
26. Itoh N, Mishima H, Yoshida Y, Yoshida M, Oka H, Matsudaira K. Evaluation of the effect of patient education and strengthening exercise therapy using a mobile messaging app on work productivity in Japanese patients with chronic low back pain: open-label, randomized, parallel-group trial. JMIR Mhealth Uhealth 2022;10(5):e35867 [FREE Full text] [doi: 10.2196/35867] [Medline: 35576560]
27. Piette JD, Newman S, Krein SL, Marinec N, Chen J, Williams DA, et al. Patient-centered pain care using artificial intelligence and mobile health tools: a randomized comparative effectiveness trial. JAMA Intern Med 2022;182(9):975-983 [FREE Full text] [doi: 10.1001/jamainternmed.2022.3178] [Medline: 35939288]
28. Rushlow DR, Croghan IT, Inselman JW, Thacher TD, Friedman PA, Yao X, et al. Clinician adoption of an artificial intelligence algorithm to detect left ventricular systolic dysfunction in primary care. Mayo Clin Proc 2022;97(11):2076-2085 [FREE Full text] [doi: 10.1016/j.mayocp.2022.04.008] [Medline: 36333015]
29. Manz CR, Zhang Y, Chen K, Long Q, Small DS, Evans CN, et al. Long-term effect of machine learning-triggered behavioral nudges on serious illness conversations and end-of-life outcomes among patients with cancer: a randomized clinical trial. JAMA Oncol 2023;9(3):414-418 [FREE Full text] [doi: 10.1001/jamaoncol.2022.6303] [Medline: 36633868]

XSL•FO
RenderX

30. Wei MT, Shankar U, Parvin R, Abbas SH, Chaudhary S, Friedlander Y, et al. Evaluation of computer-aided detection during colonoscopy in the community (AI-SEE): a multicenter randomized clinical trial. Am J Gastroenterol 2023;118(10):1841-1847. [doi: 10.14309/ajg.0000000000002239] [Medline: 36892545]
31. Yang J, Cui Z, Liao X, He X, Wang L, Wei D, et al. Effects of a feedback intervention on antibiotic prescription control in primary care institutions based on a health information system: a cluster randomized cross-over controlled trial. J Glob Antimicrob Resist 2023;33:51-60 [FREE Full text] [doi: 10.1016/j.jgar.2023.02.006] [Medline: 36828121]
32. Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med 2024;30(5):1461-1470. [doi: 10.1038/s41591-024-02961-4] [Medline: 38684860]
33. Benrimoh D, Whitmore K, Richard M, Golden G, Perlman K, Jalali S, et al. Artificial Intelligence in Depression - Medication Enhancement (AID-ME): a cluster randomized trial of a deep learning enabled clinical decision support system for personalized depression treatment selection and management. Clinical Decision Support Systems 2024 [FREE Full text]
34. Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, Matilla-Pardo B, Serrano Serrano ME, Martinez-Suberviola FJ, Dejal@ Group. Effectiveness of a conversational chatbot (Dejal@bot) for the adult population to quit smoking: pragmatic, multicenter, controlled, randomized clinical trial in primary care. JMIR Mhealth Uhealth 2022;10(6):e34273 [FREE Full text] [doi: 10.2196/34273] [Medline: 35759328]
35. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfback randomized clinical trial. JAMA Intern Med 2021;181(10):1288-1296 [FREE Full text] [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]
36. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open 2024;7(10):e2440969 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.40969] [Medline: 39466245]
37. Granviken F, Meisingset I, Bach K, Bones AF, Simpson MR, Hill JC, et al. Personalised decision support in the management of patients with musculoskeletal pain in primary physiotherapy care: a cluster randomised controlled trial (the SupportPrim project). Pain 2024. [doi: 10.1097/j.pain.0000000000003456] [Medline: 39432806]
38. Thiruvengadam NR, Solaimani P, Shrestha M, Buller S, Carson R, Reyes-Garcia B, et al. The efficacy of real-time computer-aided detection of colonic neoplasia in community practice: a pragmatic randomized controlled trial. Clin Gastroenterol Hepatol 2024;22(11):2221-2230.e15 [FREE Full text] [doi: 10.1016/j.cgh.2024.02.021] [Medline: 38437999]
39. Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of voice-based conversational artificial intelligence for basal insulin prescription management among patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023;6(12):e2340232 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.40232] [Medline: 38039007]
40. Alcoceba-Herrero I, Coco-Martín MB, Jiménez-Pérez JM, Leal-Vega L, Martín-Gutiérrez A, Dueñas-Gutiérrez C, et al. Randomized controlled trial to assess the feasibility of a novel clinical decision support system based on the automatic generation of alerts through remote patient monitoring. J Clin Med 2024;13(19):5974 [FREE Full text] [doi: 10.3390/jcm13195974] [Medline: 39408035]
41. Ortiz O, Daca-Alvarez M, Rivero-Sanchez L, Gimeno-Garcia AZ, Carrillo-Palau M, Alvarez V, TIMELY study group. An artificial intelligence-assisted system versus white light endoscopy alone for adenoma detection in individuals with Lynch syndrome (TIMELY): an international, multicentre, randomised controlled trial. Lancet Gastroenterol Hepatol 2024;9(9):802-810. [doi: 10.1016/S2468-1253(24)00187-0] [Medline: 39033774]
42. Persell SD, Karmali KN, Stein N, Li J, Peprah YA, Lipiszko D, et al. Design of a randomized controlled trial comparing a mobile phone-based hypertension health coaching application to home blood pressure monitoring alone: The Smart Hypertension Control Study. Contemp Clin Trials 2018;73:92-97. [doi: 10.1016/j.cct.2018.08.013] [Medline: 30172038]
43. Fontil V, Khoong EC, Hoskote M, Radcliffe K, Ratanawongsa N, Lyles CR, et al. Evaluation of a health information technology-enabled collective intelligence platform to improve diagnosis in primary care and urgent care settings: protocol for a pragmatic randomized controlled trial. JMIR Res Protoc 2019;8(8):e13151 [FREE Full text] [doi: 10.2196/13151] [Medline: 31389337]
44. Sandal LF, Stochkendahl MJ, Svendsen MJ, Wood K, Øverås CK, Nordstoga AL, et al. An app-delivered self-management program for people with low back pain: protocol for the selfback randomized controlled trial. JMIR Res Protoc 2019;8(12):e14720 [FREE Full text] [doi: 10.2196/14720] [Medline: 31793897]
45. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cemballi A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020;10(8):e034723 [FREE Full text] [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
46. Yao X, McCoy RG, Friedman PA, Shah ND, Barry BA, Behnken EM, et al. ECG AI-guided screening for low ejection fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. Am Heart J 2020;219:31-36. [doi: 10.1016/j.ahj.2019.10.007] [Medline: 31710842]
47. Lauffenburger JC, Yom-Tov E, Keller PA, McDonnell ME, Bessette LG, Fontanet CP, et al. Reinforcement learning to improve non-adherence for diabetes treatments by optimising response and customising engagement (REINFORCE): study

XSL•FO
RenderX

protocol of a pragmatic randomised trial. BMJ Open 2021;11(12):e052091 [FREE Full text] [doi: 10.1136/bmjopen-2021-052091] [Medline: 34862289]
48. Cheng W, Zhang Z, Hoelzer S, Tang W, Liang Y, Du Y, et al. Evaluation of a village-based digital health kiosks program: a protocol for a cluster randomized clinical trial. Digit Health 2022;8:20552076221129100 [FREE Full text] [doi: 10.1177/20552076221129100] [Medline: 36211797]
49. Chang Y, Yao Y, Cui Z, Yang G, Li D, Wang L, et al. Changing antibiotic prescribing practices in outpatient primary care settings in China: study protocol for a health information system-based cluster-randomised crossover controlled trial. PLoS One 2022;17(1):e0259065 [FREE Full text] [doi: 10.1371/journal.pone.0259065] [Medline: 34995279]
50. Han JED, Liu X, Bunce C, Douiri A, Vale L, Blandford A, et al. Teleophthalmology-enabled and artificial intelligence-ready referral pathway for community optometry referrals of retinal disease (HERMES): a cluster randomised superiority trial with a linked diagnostic accuracy study-hermes study report 1-study protocol. BMJ Open 2022;12(2):e055845 [FREE Full text] [doi: 10.1136/bmjopen-2021-055845] [Medline: 35105593]
51. Kleiman MJ, Plewes AD, Owora A, Grout RW, Dexter PR, Fowler NR, et al. Digital detection of dementia (D): a study protocol for a pragmatic cluster-randomized trial examining the application of patient-reported outcomes and passive clinical decision support systems. Trials 2022;23(1):868 [FREE Full text] [doi: 10.1186/s13063-022-06809-5] [Medline: 36221141]
52. Laranjo L, Shaw T, Trivedi R, Thomas S, Charlston E, Klimis H, et al. Coordinating health care with artificial intelligence-supported technology for patients with atrial fibrillation: protocol for a randomized controlled trial. JMIR Res Protoc 2022;11(4):e34470 [FREE Full text] [doi: 10.2196/34470] [Medline: 35416784]
53. Marshall BDL, Alexander-Scott N, Yedinak JL, Hallowell BD, Goedel WC, Allen B, et al. Preventing overdose using information and data from the environment (PROVIDENT): protocol for a randomized, population-based, community intervention trial. Addiction 2022;117(4):1152-1162 [FREE Full text] [doi: 10.1111/add.15731] [Medline: 34729851]
54. Ru X, Zhu L, Ma Y, Wang T, Pan Z. Effect of an artificial intelligence-assisted tool on non-valvular atrial fibrillation anticoagulation management in primary care: protocol for a cluster randomized controlled trial. Trials 2022;23(1):316 [FREE Full text] [doi: 10.1186/s13063-022-06250-8] [Medline: 35428315]
55. Soto-Ruiz N, Escalada-Hernández P, Martín-Rodríguez LS, Ferraz-Torres M, García-Vivar C. Web-based personalized intervention to improve quality of life and self-efficacy of long-term breast cancer survivors: study protocol for a randomized controlled trial. Int J Environ Res Public Health 2022;19(19):12240 [FREE Full text] [doi: 10.3390/ijerph191912240] [Medline: 36231542]
56. Wong CK, Hai JJ, Lau YM, Zhou M, Lui HW, Lau KK, et al. Protocol for home-based solution for remote atrial fibrillation screening to prevent recurrence stroke (HUA-TUO AF Trial): a randomised controlled trial. BMJ Open 2022;12(7):e053466 [FREE Full text] [doi: 10.1136/bmjopen-2021-053466] [Medline: 35840293]
57. Gordon BR, Qiu L, Doerksen SE, Kanski B, Lorenzo A, Truica CI, et al. Addressing metastatic individuals everyday: rationale and design of the nurse AMIE for Amazon Echo Show trial among metastatic breast cancer patients. Contemp Clin Trials Commun 2023;32:101058 [FREE Full text] [doi: 10.1016/j.conctc.2023.101058] [Medline: 36698743]
58. Heinzen EP, Wilson PM, Storlie CB, Demuth GO, Asai SW, Schaeferle GM, et al. Impact of a machine learning algorithm on time to palliative care in a primary care population: protocol for a stepped-wedge pragmatic randomized trial. BMC Palliat Care 2023;22(1):9 [FREE Full text] [doi: 10.1186/s12904-022-01113-0] [Medline: 36737744]
59. Doe G, El-Emir E, Edwards GD, Topalovic M, Evans RA, Russell R, et al. Comparing performance of primary care clinicians in the interpretation of SPIROmetry with or without artificial intelligence decision support software (SPIRO-AID): a protocol for a randomised controlled trial. BMJ Open 2024;14(6):e086736 [FREE Full text] [doi: 10.1136/bmjopen-2024-086736] [Medline: 38950987]
60. Avila-Tomas JF, Olano-Espinosa E, Minué-Lorenzo C, Martinez-Suberbiola FJ, Matilla-Pardo B, Serrano-Serrano ME, Group Dej@lo. Effectiveness of a chat-bot for the adult population to quit smoking: protocol of a pragmatic clinical trial in primary care (Dejal@). BMC Med Inform Decis Mak 2019 Dec 03;19(1):249 [FREE Full text] [doi: 10.1186/s12911-019-0972-z] [Medline: 31796061]
61. Kwan YH, Yoon S, Tai BC, Tan CS, Phang JK, Tan WB, et al. Empowering patients with comorbid diabetes and hypertension through a multi-component intervention of mobile app, health coaching and shared decision-making: protocol for an effectiveness-implementation of randomised controlled trial. PLoS One 2024;19(2):e0296338 [FREE Full text] [doi: 10.1371/journal.pone.0296338] [Medline: 38408067]
62. Zhang H, Huo X, Ren L, Lu J, Li J, Zheng X, et al. Design and rationale of the Comprehensive Intelligent Hypertension Management System (CHESS) evaluation study: a cluster randomized controlled trial for hypertension management in primary care. Am Heart J 2024;273:90-101 [FREE Full text] [doi: 10.1016/j.ahj.2024.03.018] [Medline: 38575049]
63. Kung TH, Cheatham M, Medenilla A, Sillos C, de Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for ai-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
64. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Nature 2023;614(7947):224-226. [doi: 10.1038/d41586-023-00288-7] [Medline: 36737653]

XSL•FO
RenderX

65. Moura L, Jones DT, Sheikh IS, Murphy S, Kalfin M, Kummer BR, et al. Implications of large language models for quality and efficiency of neurologic care: emerging issues in neurology. Neurology 2024;102(11):e209497. [doi: 10.1212/WNL.0000000000209497] [Medline: 38759131]
66. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion 2023;99:101805. [doi: 10.1016/j.inffus.2023.101805]
67. Hinterwimmer F, Lazic I, Suren C, Hirschmann MT, Pohlig F, Rueckert D, et al. Machine learning in knee arthroplasty: specific data are key-a systematic review. Knee Surg Sports Traumatol Arthrosc 2022;30(2):376-388 [FREE Full text] [doi: 10.1007/s00167-021-06848-6] [Medline: 35006281]
68. Wehbe RM, Katsaggelos AK, Hammond KJ, Hong H, Ahmad FS, Ouyang D, et al. Deep learning for cardiovascular imaging: a review. JAMA Cardiol 2023;8(11):1089-1098. [doi: 10.1001/jamacardio.2023.3142] [Medline: 37728933]
69. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. J Am Med Inform Assoc 2020;27(4):592-600 [FREE Full text] [doi: 10.1093/jamia/ocz229] [Medline: 32106285]
70. Adler-Milstein J, Aggarwal N, Ahmed M, Castner J, Evans BJ, Gonzalez AA, et al. Meeting the moment: addressing barriers and facilitating clinical adoption of artificial intelligence in medical diagnosis. NAM Perspect 2022;2022:10.31478/202209c [FREE Full text] [doi: 10.31478/202209c] [Medline: 36713769]
71. Maadi M, Akbarzadeh Khorshidi H, Aickelin U. A review on human-AI interaction in machine learning and insights for medical applications. Int J Environ Res Public Health 2021;18(4):2121 [FREE Full text] [doi: 10.3390/ijerph18042121] [Medline: 33671609]
72. Kim HM, Kang H, Lee C, Park JH, Chung MK, Kim M, et al. Evaluation of the clinical efficacy and trust in AI-assisted embryo ranking: survey-based prospective study. J Med Internet Res 2024;26:e52637 [FREE Full text] [doi: 10.2196/52637] [Medline: 38830209]
73. Campion JR, O'Connor DB, Lahiff C. Human-artificial intelligence interaction in gastrointestinal endoscopy. World J Gastrointest Endosc 2024;16(3):126-135 [FREE Full text] [doi: 10.4253/wjge.v16.i3.126] [Medline: 38577646]
74. Rennie D. CONSORT revised--improving the reporting of randomized trials. JAMA 2001;285(15):2006-2007. [doi: 10.1001/jama.285.15.2006] [Medline: 11308440]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
LLM: large language model
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence

Edited by T de Azevedo Cardoso; submitted 25.01.24; peer-reviewed by VT Obulareddy, X Ying, CM Chu, S Papineni; comments to author 08.11.24; revised version received 21.12.24; accepted 22.01.25; published 25.02.25.

Please cite as:
Zhong J, Zhu T, Huang Y
Reporting Quality of AI Intervention in Randomized Controlled Trials in Primary Care: Systematic Review and Meta-Epidemiological Study
J Med Internet Res 2025;27:e56774
URL: https://www.jmir.org/2025/1/e56774
doi:10.2196/56774
PMID:39998876

©Jinjia Zhong, Ting Zhu, Yafang Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effects of the Modern Digital Information Environment on Maternal Health Care Professionals, the Role of Midwives, and the People in Their Care: Scoping Review

Elisabeth Wilhelm[1], MA; Victoria Vivilaki[1], RM, PhD; Jean Calleja-Agius[2], MSc, MD, PhD; Elena Petelos[3,4], PhD; Maria Tzeli[1], RM, MSc; Paraskevi Giaxi[1], RM, PhD; Elena Triantiafyllou[1], RM, MSc; Eleni Asimaki[1], RM, MSc; Faye Alevizou[1], RM; Tina D Purnat[5], MSc

[1]Department of Midwifery, School of Healthcare and Care Sciences, University of West Attica, Athens, Greece

[2]Department of Anatomy, Faculty of Medicine & Surgery, University of Malta, Msida, Malta

[3]CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, The Netherlands

[4]Faculty of Medicine, University of Crete, Rethymno, Greece

[5]TH Chan School of Public Health, Harvard University, Boston, MA, United States

Corresponding Author:
Elisabeth Wilhelm, MA
Department of Midwifery
School of Healthcare and Care Sciences
University of West Attica
Agiou Spiridonos 28
Egaleo
Athens, 122 43
Greece
Phone: 30 21 0538 510
Email: ewilhelm@uniwa.gr

Abstract

Background: The digital information environment poses challenges for pregnant women and other people seeking care, as well as for their midwives and other health care professionals (HCPs). They can encounter questions, concerns, information gaps, and misinformation, which can influence health care decisions.

Objective: This scoping review examines how HCPs are affected by the modern digital information environment including health misinformation, its effects on the women and people they care for, and its implications for care provision.

Methods: English-language peer-reviewed literature, published from January 1, 2020, to May 31, 2024, with keywords related to midwifery, misinformation, and health equity collected and analyzed by a team of midwives and maternal care professionals and mapped onto a patient-centered conceptual model.

Results: A total of 105 studies were ultimately included. Further, 95 papers identified specific digital information environment issues that affected clients; 58 specifically highlighted digital information environment issues impacting HCPs; 91 papers identified specific topics of common questions, concerns, misinformation, information voids, or narratives; 57 papers identified patient or population vulnerability; and 75 included mentions of solutions or recommendations for addressing a digital information environment issue around clients seeking care from midwives and other HCPs. When mapped onto the Journey to Health model, the most prominent barrier was access to care and information. Individual-level issues dominate the step related to knowledge, awareness, and belief, with more social norms and wider engagement appearing at steps related to intent. Client-specific themes dominate the left-hand side of the model and provider-specific issues dominate the right-hand side of the model.

Conclusions: Misinformation, information voids, unaddressed questions and concerns, and lack of access to high-quality health information are worldwide prevalent barriers that affect both patients and HCPs. We identified individual, provider-level, health systems, and societal-level strategies that can be used to promote healthier digital information environments.

(J Med Internet Res 2025;27:e70108) doi:10.2196/70108

XSL•FO
RenderX

KEYWORDS

digital health; midwifery; misinformation; information-seeking midwifery; health information seeking; social media; medical misinformation

Introduction

Background

Pregnant women and parents-to-be are often exposed to a flood of information and advice not only from health care professionals (HCPs), but also from family, friends, and increasingly from various online sources [1]. However, not all of this online information is accurate and some might be deceptive [2], making it challenging for people to identify reliable sources of health information and to filter and act on reliable guidance regarding pregnancy, maternity, reproductive health, and other health topics [3]. The digital information environment—encompassing mobile phones, apps, social media, and websites—has profoundly influenced how people, including women and their partners in the perinatal period [4], seek and process health information. Inequities in access to health care and health information [5,6], influenced by social and commercial determinants of health [7], also affect how people's health is affected by the digital information environment [8].

Midwives and other maternal HCPs need to better understand how the people they care for seek information online, including common questions, concerns, misinformation, and narratives that may be encountered. Although information about sexuality, reproductive health, and pregnancy may come from many sources, and people may seek offline information and care from other professionals and are influenced by social and support networks [9], the digital environment is now a leading source of information [10,11]. This understanding is critical, as it can impact the HCP-client relationship [12] and is especially critical for midwives.

Midwives strive to provide woman-centered care, with respect for autonomy, choice, and control. The midwife-woman relationship has trust as its foundational basis. Evidence-based midwifery provides for deliberative exchanges for shared decision-making to establish the wishes, needs, and preferences of women, and to develop care plans that can address the needs of women, ultimately empowering them, ensuring quality care for women and newborns, and safeguarding their health and well-being [13]. The digital information environment can influence all aspects of the unique relationship between midwife and woman and can influence downstream effects of health-seeking behaviors and decision-making and how women navigate maternity care systems; in addition, it can positively or adversely affect health outcomes [14].

It is particularly important to consider that the health care relationship in the perinatal period covers the years before, during, and after pregnancy, and often multiple sequential pregnancies can result in many years of almost continuous interaction. This is a period of change as the person becomes pregnant and a parent, and develops new identities and behaviors that last a lifetime and affect all health decisions for children, mothers, and entire families [15]. Midwifery-led care, according to the International Council of Midwives [16], includes the core competency to "adapt to and adopt new and emerging technologies that have been proven to enhance midwifery practice and care." Understanding the modern digital information environment is essential to fulfilling this competency, and as the world becomes more digitally connected, its impact on people seeking care and their midwives and other maternal HCPs will only grow.

In this review, we catalog and describe how midwives and other maternal HCPs are affected by the modern digital information environment, including health misinformation, its effects on their clients, and the implications it has on care provision. The digital information environment affects both HCPs and their patients [17,18], but has differing effects on each and influences the care providers give, which is why understanding these interactions holistically is essential for improving health care. Therefore, we will use a model that encompasses the patient experience and the social and digital information landscape, as well as health care access in the review. The review also includes research that is focused on maternal HCPs beyond midwives to reflect the maturing nature of research in this field.

Based on previous literature on social media [19], digital ecosystems in health care, and the information environment in the context of health [20], we define the digital information environment as the online spaces where information—including news, social media content, and other digital material—is produced, curated, shared, discussed, and amplified across digital platforms and devices. However, access to both information and health care is not equitable for all. This review also focuses on vulnerable populations, or people who experience inequities in accessing health care, including migrants, people living in poverty, individuals with disabilities, and those with low literacy, to better understand how these groups are affected by health misinformation and the challenges within the digital information environment. Given that different HCPs have different terminology for people under their care—such as clients, patients, or parents-to-be—and this review includes papers from many medical and health disciplines, terms may be used interchangeably.

Why Focus on Midwives and Their Relationship to Their Clients?

This review will map available evidence and help inform the development of strategies for midwives to help their clients better cope with the modern digital information environment, promote inclusive communication, build trust, and more successfully encourage clients to follow health care advice, therefore improving client health outcomes. Developing a more trusting and effective midwife-client relationship can also help the clients maintain seeking health services for themselves and their children beyond the perinatal period.

The underpinning philosophy of the interaction between midwives and patients in the digital information landscape is based on trust, enabling effective collaboration in navigating

XSL•FO
RenderX

modern health care challenges [21]. Both groups operate within the same digital ecosystem, but the ways they engage with and are impacted by information and disinformation differ significantly.

Women, often the primary consumers of digital health information, increasingly rely on tools such as websites, social media, and apps to seek knowledge about their health. While this access can empower them to take a more active role in their care, it also exposes them to misinformation, conflicting advice, and volumes of iatrogenic content. The result can be confusion, anxiety, and even decisions based on unreliable sources [22].

Midwives, on the other hand, engage with the same ecosystem as intermediaries and guides. They must stay informed about the evolving digital health landscape to effectively filter, interpret, and contextualize the information women encounter [23]. Their role is to identify credible resources, dispel misinformation, and integrate digital health content into evidence-based midwifery care. This dual role of interpreter and gatekeeper reinforces trust between midwives and women, as midwives ensure that the information women rely on is accurate and applicable to their individual needs.

Moreover, the way digital information impacts midwives and women is shaped by their priorities and capabilities. For women, digital tools often serve as an entry point to health education, providing a sense of autonomy but also raising questions and concerns [24]. For midwives, these tools are professional tools [25], used to enhance shared decision-making and to address women's concerns in a way that fosters trust and collaboration. While women may experience information as fragmented or overwhelming, midwives transform the digital landscape into a resource for empowerment.

Disinformation poses a unique challenge in this ecosystem. Women may encounter intentionally misleading content that exacerbates mistrust or undermines their confidence in professional advice [26]. For midwives, disinformation adds complexity to their role, requiring them to not only correct falsehoods but also rebuild trust when it has been eroded by inaccurate digital narratives. This dynamic underscores the critical importance of midwives as trusted navigators in the health care journey.

Trust extends beyond individual interactions to midwives' broader advocacy for equitable access to credible digital resources. Systemic challenges, such as the digital divide and disparities in digital literacy, disproportionately affect women, especially those from underserved populations such as migrant mothers. Midwives address these inequities by guiding women toward accessible, understandable, and reliable health information. This advocacy, combined with evidence-based midwifery practices and critical thinking skills fostered through midwifery education, further equips midwives to support women effectively in the digital era [27].

Ultimately, the differentiated dynamics between midwives and women in the digital information ecosystem highlight the complementary roles they play. By synthesizing digital information, midwifery expertise, and trust, midwives ensure that women feel supported, informed, and confident as they navigate their health care journey [28].

Review Questions

This review aims to identify the facilitators and barriers that the modern digital information environment presents to midwives and the care they provide to their clients.

The key questions to answer in this review are: (1) How has the modern digital information environment affected midwives? (2) What common types of health misinformation are midwives asked to address? (3) How has the digital information environment affected clients' information and health-seeking behaviors?

Methods

Framework for Search and Analysis

This review adapted the UNICEF (United Nations International Children's Emergency Fund) Journey to Health model to map factors specific to HCPs and clients that surface in the review's included papers [20]. This framework is used to conceptualize the health care–seeking journey of individuals and the care provision journey of HCPs. These factors are linked to the digital information environment, including health misinformation, which may influence health information-seeking behaviors, care-seeking, and health care provision.

As shown in Figure 1, the model imagines the journey a patient or client undertakes to seek health care [29], and conversely, the HCP's own journey in providing care, with individual-level and informational attributes on the left side of the diagram and more health service and care provision aspects found on the right side. For a patient or client to fully complete a course of recommended care, such as antenatal visits, they need to complete this journey many times but may encounter barriers that can make this journey harder, or enablers that make it easier. HCPs are also situated along this journey on the right-hand side at the point of service and heavily influence the experience of care and after-service for clients. This model also allows for the digital information environment effects on the HCP and the individual to be captured in the context of larger systems that shape their interactions. Any intervention or strategy—individual, community, health system, societal, or systemic—identified to manage the digital information environment would be included and mapped on the model.

XSL•FO
RenderX

Figure 1. UNICEF's Journey to Health model, which illustrates the 6 steps a patient or client takes in seeking health care and the multilevel factors that influence this journey. This review maps data extracted from included papers onto this model. The Journey to Health is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO (CC BY-NC-SA 3.0 IGO) license. UNICEF: United Nations International Children's Emergency Fund.

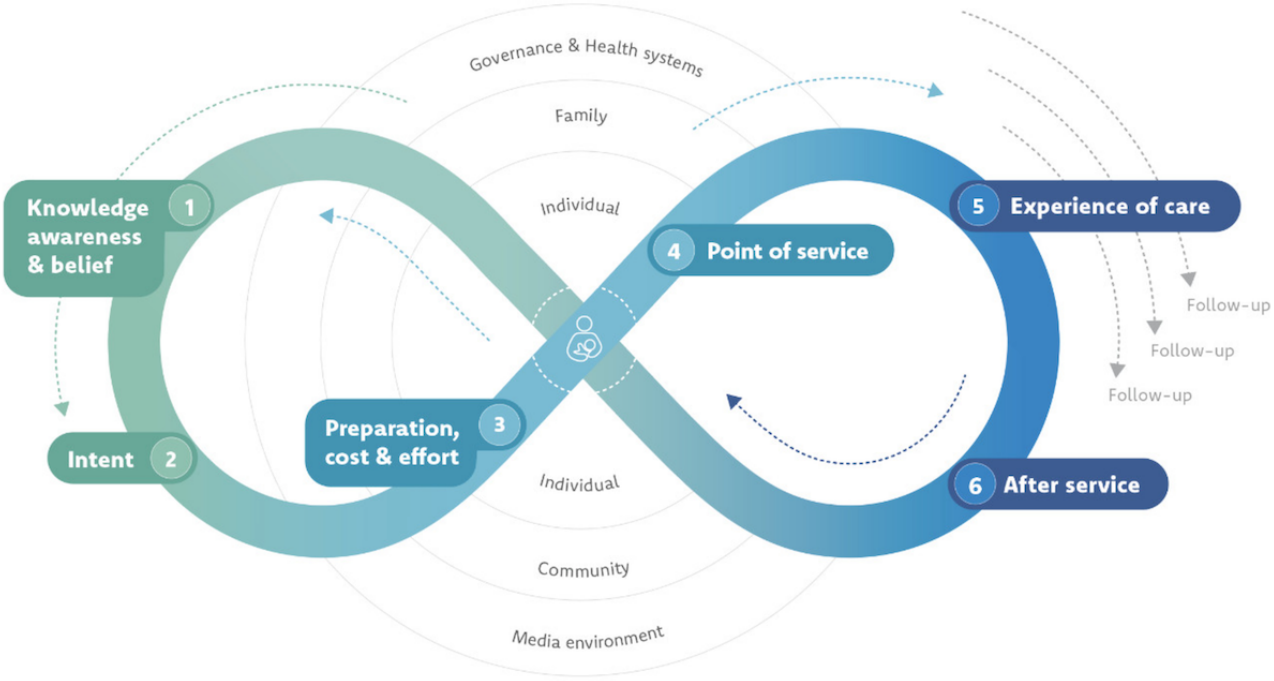

The domains of inquiry during the analysis process, aligned with this framework, include themes of digital information environment barriers and enablers to clients seeking health care, midwives, and other HCPs in providing care, and what strategies or interventions were mentioned to overcome the identified digital information environment issue.

Search Strategy

The domain being studied is digital information environment components, particularly health information, questions, concerns, information voids, narratives, health misinformation, and disinformation in online spaces [30], which can affect a person's perceptions and health care decisions. In this review, the focus was on midwives but aimed to broaden the scope also to those HCPs who provided maternal health care, people in the perinatal period, and equity considerations.

Peer-reviewed papers were identified through databases including PubMed, Scopus, Web of Science, Science Direct, Cochrane Library, Embase, and Google Scholar (specifically review papers).

The search strategy was developed by identifying keywords from relevant recent publications and identifying additional terms associated with midwives and misinformation, social media, information environment, digital health, health information, health equity, and other related Medical Subject Headings (MeSH) keywords. Search using such an approach yielded very few papers, likely due to inconsistent coding of concepts related to misinformation, information environment, and interpersonal communication challenges in a health care setting. Therefore, the keyword string was reduced to broaden the search and cast a wider net for potentially applicable papers from the information environment point of view. At the same time, terms related to midwives and pregnancy were introduced to focus the search results on midwives and their clients and reduce the number of results that were not relevant to this setting and were being returned when we used broader terms referring to HCPs. This exercise resulted in the search string provided in the following paragraph and uncovered a more diverse array of relevant papers from health care journals, social science, social media, the internet, and technology.

The final search string was modified to fit the parameters of each scientific literature database: ((pregnancy) AND (midwife) AND (misinformation) AND (communication OR question* OR concern*)) AND equity.

To be included, papers were required to be published in peer-reviewed journals in English between January 1, 2020, to May 31, 2024, in addition to other criteria (Table 1). This timeframe was chosen due to the rapidly changing digital information environment, especially during and shortly following the global COVID-19 pandemic, which saw major social media platforms weaken content moderation against health misinformation [31], TikTok grow as a leading source of health information for women [32], health care provision shift to include telehealth and increase in health information-seeking online [33,34], and misinformation identified as a growing public health concern worldwide [35,36].

XSL•FO
RenderX

Table 1. Summary of inclusion and exclusion criteria for papers in this scoping review focused on midwives, the people they care for, and the modern information environment.

	Included	Excluded
Article type	Original research, reviews, meta-analyses, or case reports published in peer-reviewed research journals	Opinions and commentaries, gray literature, preprints, duplicates, or book chapters
Methodology	Can include any methodology or type of research (qualitative and quantitative)	N/A[a]
Geography	Any country or region	N/A
Language	English	Not in English
Publication date range	January 1, 2020, to May 31, 2024	Outside of January 1, 2020, to May 31, 2024
Full text available?	Yes	No
Related to digital information environment? (eg, includes social media, apps, and internet)	Yes	No
Relevant to maternal health professionals? (eg, includes midwives or OB-GYNs[b] as health providers or pregnant women as the focus population)	Yes	No
Related to health equity? (eg, discusses disparities in information or health care access and use for different populations)	Yes	No

[a]N/A: not applicable.

[b]OB-GYN: obstetrics and gynecology.

Analysis Team Setup

The screening and analysis process was undertaken by a multidisciplinary, multinational team of health care and public health–focused academics and professionals, including several midwives. This ensured that the review of the literature to inform midwifery-led care was also led by midwives and that the extracted data and conclusions were correctly interpreted from a health care lens. A standard operating procedure (SOP) was developed and the team was trained on processes and tools to follow SOPs. The team collaborated at every stage, including screening, data extraction, analysis, and synthesis, conducting weekly check-in meetings discussing results, reviewing all entries and reconciling disparate judgments on particular papers, spot-checking one another's work, and updating SOPs and definitions to ensure consistency of screening and analysis processes.

Screening

Searches for papers for inclusion in the review were conducted in all identified peer-reviewed literature databases in English using defined date ranges and MeSH search terms. These papers (n=895) were then exported into Paperpile where some records were removed before screening, including duplicate records (n=29), non-English ones, those outside of the desired date range, or those initially flagged as not from a peer-reviewed journal (n=297). The remaining records (n=523) were screened, of which, 25 reports were not retrieved, and 498 were assessed for eligibility. Based on a full review of each paper, reports were then excluded for not being peer-reviewed (n=78), not being published within the defined date range (n=17), not being relevant to midwives (n=99), and not including an aspect of digital information environment (n=350). Ultimately, 105 papers were included in the review.

Data Extraction and Thematic Analysis

After the final list of studies to be included for review was compiled, they were loaded into a shared spreadsheet. The analysis team members then received randomized selections of the papers for further data extraction, coding, and thematic analysis. Each assigned paper was thoroughly read, reviewed, and discussed by the team. Any disagreements in judgments were resolved collaboratively before proceeding to the next stage. Papers were assessed for quality and qualitative papers were scored using a modified RATS (relevance, appropriateness, transparency, soundness) scoring framework.

In the data extraction phase, a PICO (population, intervention, comparison, and outcome) framework was used to extract and record the following information for each paper: (1) primary population of focus (HCPs), (2) stage of reproductive journey, (3) status and identities of population of focus that impact health equity, (4) country, (5) type of study, (6) link with COVID-19, (7) aspect of digital information environment investigated, (8) (if included) interventions or strategies to address digital information issue, and (9) intervention effects.

For the field (7) aspect of digital information environment investigated, the codes were developed based on a preparatory literature search and the papers analyzed: information-seeking, coping or support-seeking, community-seeking, information overload, information or news exposure, misinformation (inaccurate information on products, service, health information, or health guidance), disinformation (deliberately created inaccurate information on products, service, health information, or health guidance spread for political or financial gain),

XSL•FO
RenderX

expressing questions or concerns, digital-mediated social issues (eg, bullying, harassment, stigma, or doxxing), social media use, app use, social media content (eg, memes, hashtags, videos, or images), deceptive marketing or advertising, communication with HCP, infodemic, and other.

Coding Process Across Domains

In the coding stage, the papers were coded for specific domains, with key information from relevant sections of papers extracted and added to the spreadsheet. These domains included: how the identified digital information environment issue affected clients, how it affects HCPs, and topics of common questions, concerns, misinformation, information voids, or narratives identified. Considerations related to patient or population vulnerability and any recommendations or recommended solutions to address the infodemic environment issue were also captured and coded.

Synthesis of Results

In this phase, analysts reviewed all 105 papers according to the domains listed above. To facilitate thematic grouping and visualization, the data were transferred to Miro, a digital whiteboard, where sticky notes were created, shared, discussed, and organized into themes during videoconference calls. Identifying new themes came from grouping similar ideas together [37] and identifying visually what ideas were on their own (or outliers) and placing similar groupings of ideas close to one another or between 2 different ideas that are linked. For example, COVID-19 as a theme is often related to the vaccines theme and fertility theme. Then, after the group discussion and organization of ideas on sticky notes, a template was used to help consolidate observations for each domain. Themes were then mapped on the Journey to Health model to provide a visual perspective on where along the journey the theme fits best and discussed with the analysis team to inform implications for HCP practice.

Results

Papers by the Numbers

The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist is available in Multimedia Appendix 1. Of 105 papers (Figure 2) included in the review, 28 papers were a type of review, systematic reviews, or meta-analyses, 91 papers discussed a vulnerability or aspect of health equity, 74 papers included multiple countries, 37 were COVID-19–focused, 46 included misinformation as the major digital information environment issue, 48 had suggested interventions or solutions with no papers including intervention effects, 75 had related recommendations or solutions, and 91 included topics of common questions, concerns or misinformation identified. Only 3 papers were solely focused on midwives versus broader categories of maternal HCPs. Additional information about included papers is summarized in Table 2.

XSL•FO
RenderX

Figure 2. Identification of studies via databases included in this review focusing on midwives, their clients, and health misinformation and related digital information environment issues, adapted from the PRISMA 2020 flow diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

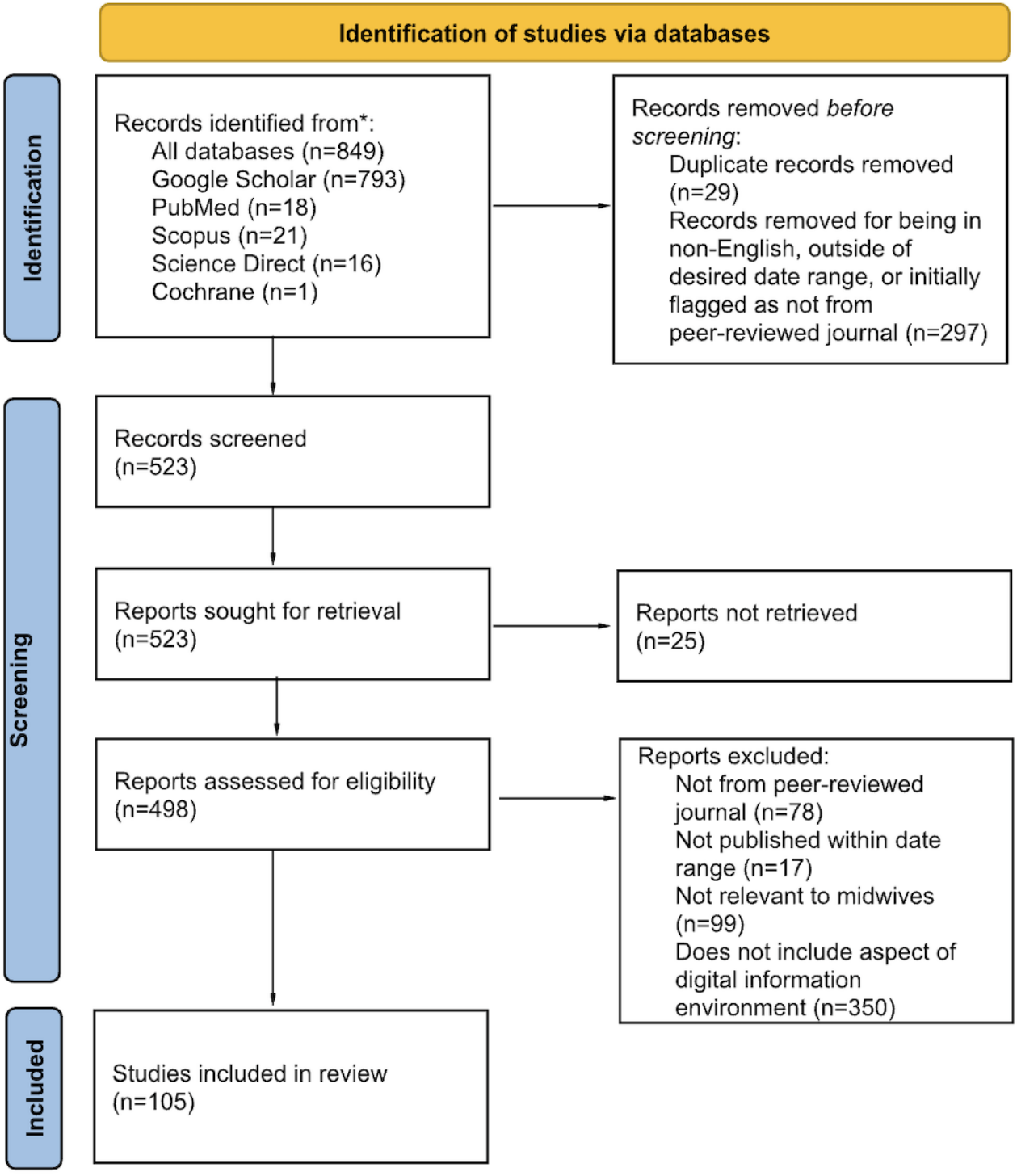

XSL•FO
RenderX

Table 2. Selected characteristics of included papers in this scoping review focused on midwives, their clients, and digital information environment challenges.

Specific characteristic	Number of papers (of 105 papers included in review), n (%)
Paper a review (eg, review, systematic review, or meta-analysis)	28 (26)
Paper related to a health vulnerability or health equity issue	91 (86)
Papers included multiple countries	74 (70)
Papers that were COVID-19–focused	37 (35)
Papers that included misinformation as a major digital information environment issue	46 (43)
Papers that included specific interventions or strategies to address digital information environment issue	48 (45)
Papers that included intervention effects	0 (0)
Papers that included recommendations or solutions to address digital information environment issue	75 (71)
Papers that included topics of common questions, concerns, or misinformation identified	91 (85)
Papers that were specific to midwives compared to other HCP[a] types or non-HCP audience	3 (2)

[a]HCP: health care professional.

Results by Domain

How Identified Digital Information Environment Issues Affected Clients

Of 105 papers, 95 papers identified specific digital information environment issues that affected clients. The most common topics were COVID-19, vaccine hesitancy, access, decision-making and care-seeking, breastfeeding, experience of care or HCP relationship, poor quality information, social norms, lack of trust, information overload, anxiety, peers, gender, and self-esteem or body image.

Several common and cross-cutting themes emerged for this domain. These included mental and cognitive barriers where clients expressed anxiety, experienced information overload, and had difficulty determining information quality and distinguishing between accurate and inaccurate information. Information voids also featured, where clients could not find the information that they were looking for from online resources or found the messaging confusing or mixed. This sometimes led to reliance on lower quality available information or misinformation. Informational barriers, such as lack of access to health information, including internet access and information available in appropriate formats, language, and cultural relevance could make people more prone to misinformation. Negative effects on behavior were also documented, where misinformation exposure can lead to attitudes and behaviors that put clients' health at risk. Finally, trust in HCPs was a barrier where poor health care encounters and stigmatizing experiences aggravated clients' mistrust in HCPs who then sought information and care elsewhere, especially from online spaces where they felt more socially connected and supported.

As shown in Figure 3, on the Journey to Health model, the clustering of themes in earlier stages of the journey highlights individual-level factors, information environment issues affecting knowledge and decision-making about health, and therefore making the rest of the journey and advancing to (3) preparation, cost, and effort and beyond more difficult. Digital information environment challenges can place informational barriers in the way of health care seeking. Access barriers further hindered the client journey, suggesting that some people who need care are not able to obtain it, and for those who do—for (5) experience of care—poor-quality interactions with HCPs by clients make it less likely they will trust them further or return.

XSL•FO
RenderX

Figure 3. Sample screenshot of Journey to Health with major themes related to client barriers and enablers related to digital information environment challenges mapped, clustering around the (1) knowledge, awareness, and belief and (2) intent stages, which can influence care seeking in later steps of the journey.

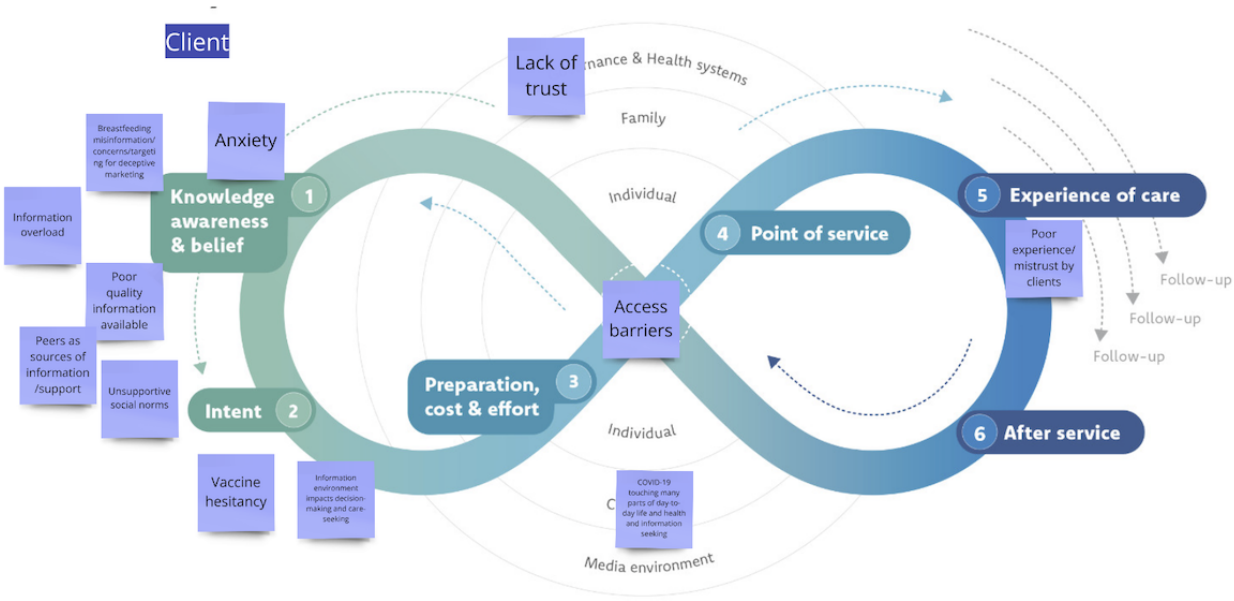

At a higher level of the socioecological diagram encompassing the journey, trust, or lack thereof, played a prominent role in influencing client attitudes toward health providers, their advice, and larger health care systems. COVID-19 also played a disruptive role, affecting the information environment and how care was offered, resulting in a chaotic information environment often seeding confusing information, providing mixed messaging, and making it difficult to communicate with HCPs.

How the Identified Digital Information Environment Issue Affected HCPs

Of the 105 papers reviewed, 58 specifically highlighted digital information environment issues impacting HCPs.

The most common topics identified included poor provision of care, stigma, the impacts of COVID-19, HCPs as both a source of misinformation and as a means of addressing it, lack of accurate guidance for HCPs, communication challenges, alternative information-seeking behaviors where clients circumvented their HCP, trust, the role of digital apps, and direct negative impacts on HCPs.

Across papers mentioning HCP-side challenges, several themes emerged. HCPs had dual-sided roles in information provision, some actively promoting misinformation and others addressing misinformation on a variety of topics, especially during the COVID-19 pandemic. COVID-19 also was the catalyst for the rapidly changing digital information environment, significantly accelerating challenges within the digital information environment for HCPs, driving greater digitization, the growth of telehealth and mobile health (mHealth), and changing client expectations around addressing misinformation. HCPs reported that misinformation was not only a challenge to providing health care but also a factor that affected their clients' attitudes and decision-making. HCPs also reported a lack of training or professional guidance on how to address misinformation and other challenges within the digital information environment, including improving their own digital literacy and evaluating mHealth apps used by clients. Misinformation contributed to less emphatic care where HCPs stigmatized or discriminated against clients based on misinformation. When clients lack access to or mistrust HCPs, they will seek and act on health information elsewhere, which can make it difficult for HPCs to understand what health information (or misinformation) influenced their clients and their resulting health decisions.

Themes were mapped onto the Journey to Health model, spanning most stages from (1) knowledge, awareness, and belief to (2) intent; (3) preparation, cost, and effort; (4) point of service; and (5) experience of care. Several themes clustered about the point of service and experience of care, indicating that digital information environment challenges such as misinformation affected health care provision. Larger social and systems forces related to the information environment and mistrust in HCPs and health systems by patients and COVID-19 as a major event disrupting health care also played a role in changing how care was provided.

Topics of Common Questions, Concerns, Misinformation, Information Voids, or Narratives Identified

Of 105 papers, 91 papers identified specific topics of common questions, concerns, misinformation, information voids, or narratives identified.

Common topics included COVID-19, vaccines, reproductive health, pregnancy, breastfeeding, family planning, LGBTQ+ (lesbian, gay, bisexual, transgender, queer, and other identities), baby, labor and birth, nutrition, obesity, and stereotypes about groups of people and mental health. These topics were not just misinformation but also reflected topics where there was confusion, lack of information, misperceptions, or questions and concerns. These are topics that patients expected their

XSL•FO
RenderX

providers to be able to address but reported feeling unsatisfied with given HCP-provided information and advice.

There were only two themes that emerged beyond specific topics, including issues related to information seeking between (1) knowledge, awareness, and belief and (2) intent and lack of access before (4) point of service on the Journey to Health. Both are about seeking information and health care, while the topical items were not specific to any stage of the journey.

Considerations Related to Patient or Population Vulnerability

Of 105 papers, 57 papers identified patient or population vulnerability.

The types of vulnerability discussed included multiple vulnerabilities, migration or refugee status, religious or ethnic minority status, adolescence, mental health vulnerabilities, disability, poverty, LGBTQ+ status, and adverse life experiences, among others. COVID-19 also highlighted the vulnerabilities of specific groups, including those with pre-existing health conditions and pregnant people.

The most common topics were access, information seeking, trust, social marginalization and stigma, literacy, and risk perception.

Emerging themes included access barriers, which included legal or policy, social, cultural, and physical access challenges to health information and care. Access to care is made more difficult if someone has a vulnerability or comes from a vulnerable population. Information seeking, shaped by literacy, identity, and access also featured as a theme. People will seek health information from sources they trust, which often includes family, friends, and larger peer groups where they share identity and community.

Themes mapped onto the Journey to Health around vulnerabilities clustered around (1) knowledge, awareness, and belief and (2) intent. Vulnerabilities can affect individual-level and social processes and barriers in earlier stages of the journey can limit access to care.

Recommendations or Solutions Are Presented to Address Digital Information Environment Issues

Of 105 papers, 75 included mentions of solutions or recommendations for addressing a digital information environment issue around clients seeking care from midwives and other HCPs. These focused on system-side interventions or solutions compared to more general societal or government-level interventions, such as technology platform regulation, content moderation, or fact-checking, which are well documented in the literature elsewhere [38,39].

Cross-cutting themes focused on addressing skills and performance gaps by HCPs. This included improving HCP communication with strategies such as health education and counseling. Solutions were often designed to improve the quality of communication, language clarity, and inclusiveness, especially for populations that have experienced stigma or discrimination, such as people with disabilities, migrants or refugees, and LGBTQ+ people. Improving knowledge and communication on related topics, such as nutrition, can also address patient concerns. Training HCPs was also a common theme, with suggested topics on how to address misinformation, communicate more clearly, and receive training to better address vaccine hesitancy. HCPs should also be trained to give more inclusive and emphatic care to LGTBQ+ people. HCPs can also address health misinformation, using techniques such as prebunking, improving resilience and health literacy in clients, and training HCPs to address misinformation successfully.

At the systems level, more work can be done to improve risk communication by improving communication by health authorities and within families, During emergencies, it is important to evolve messaging as more information becomes known, especially for pregnant women and people in the perinatal period. Several papers suggested clarifying health systems policies such as by formalizing how HCPs are trained or expected to address misinformation, encouraging more research on health information, digital environment, and patient decision-making, and ensuring that health professionals adhere to ethics regarding product marketing and sponsorship. Improving access to health care through improved access to information could be accomplished by increasing the availability of information and enhancing access to information tools and technologies (eg, mHealth apps), particularly for patients who have historically faced inequitable health care access. Strengthening digital engagement was also recommended by promoting HCPs to use digital spaces to educate and connect with clients as trusted voices and improve the quality of available health-related social media content.

Additional themes focused on community-level strategies. Increasing community outreach, participation, and cocreation could help address the lack of tailored information or services and help enhance patient communication and care. This can be carried out by improving health system–community linkages and strengthening communications beyond pregnant women, including their communities and networks as well. Mainstreaming culturally competent communication and care by HCPs could be accomplished by policies and training to improve HCP cultural competency to better care for diverse populations. This could provide more culturally appropriate health information and support midwives and HCPs in reducing stigma by providing trauma-informed care. Finally, at the individual level, health and digital literacy could be improved through health education programs for the public and via HCP in appointments with clients.

When mapped onto the Journey to Health model, recommendations and solutions predominantly cluster on the left side, associated with individual-level knowledge, awareness, and intent. However, the solutions and recommendations cut across individual-level, provider-level, societal, health systems, and government or policy levels in recognition that the complex information environment's effects may need to be addressed through multiple levels.

Putting It All Together: Mapping Across All Domains and Themes

After each domain and major theme were mapped on the Journey to Health, all themes, still color-coded, were mapped

XSL•FO
RenderX

onto the model (Figure 4). Themes were heavily clustered in steps 1-5, with access (or lack thereof) appearing consistently for all themes except for the provider. Individual-level issues dominate at step (1) knowledge, awareness, and belief, with more social norms and wider engagement appearing at step (2) intent. Client-specific themes dominate the left-hand side and provider-specific issues dominate the right-hand side.

Screenshots of the Miro whiteboard with details of the analysis of all domains are available in Multimedia Appendix 2.

Figure 4. Screenshot of Journey to Health with major themes, common topics, and vulnerabilities mapped to address digital information environment issues.

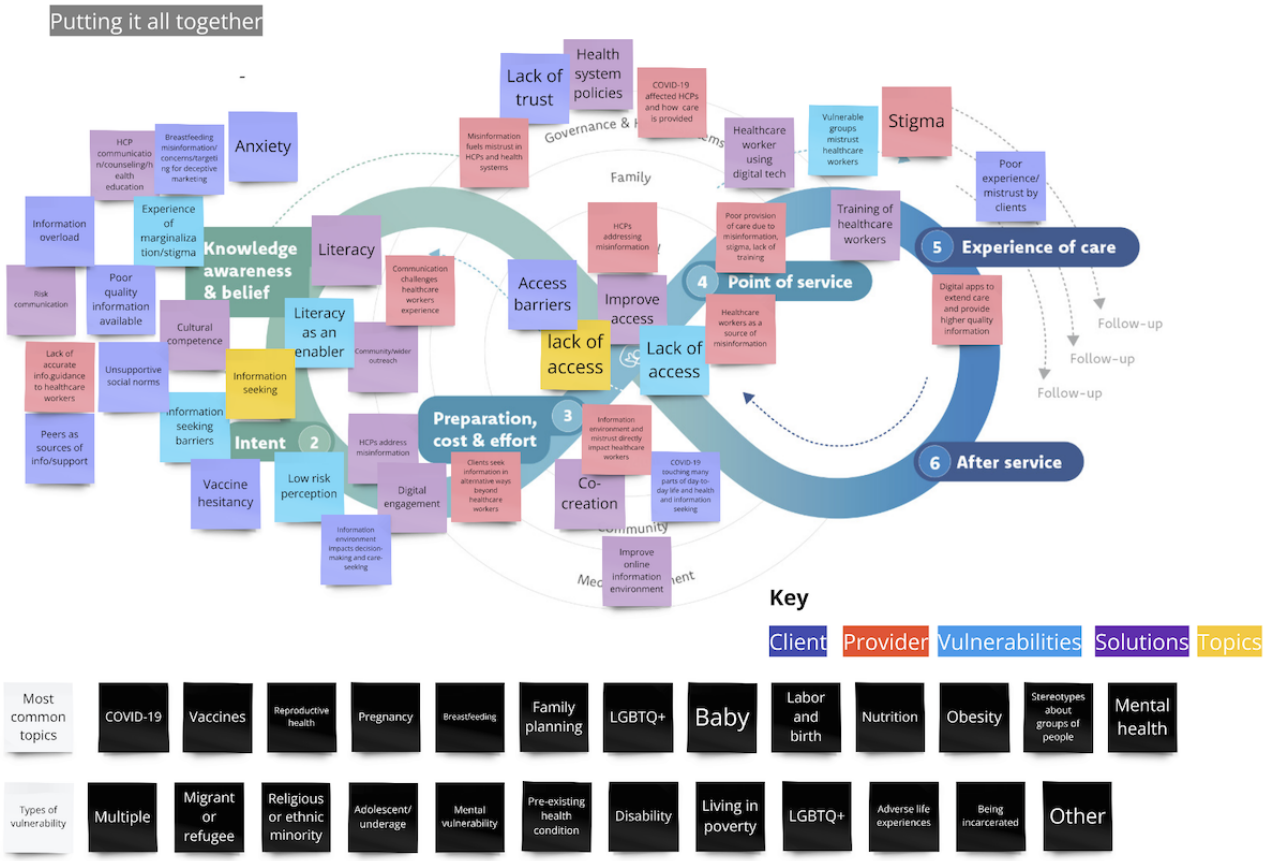

Discussion

Principal Findings

The COVID-19 pandemic, which occurred during the period of data collection for this review, featured heavily in the included papers, where 35% (37/105) were directly related to the virus and its impacts on health systems and societies. Concepts such as “infodemic” [40] emerged and gained prominence, as did the role of misinformation in affecting health care provision and health care seeking worldwide [41]. Telehealth and digital communication to provide care became more common, and in the context of mobility restrictions, people sought more information via their mobile devices. As a result, the digital divide and inequitable access to health care for some vulnerable groups became prominent. These inequities predated COVID-19 but the pandemic highlighted new concerning dimensions where inequitable access to credible, accurate health information in the digital information environment adversely impacted health.

The review illustrates that the digital information environment exacerbates existing inequalities in access for clients and patients. HCPs struggle to overcome these challenges, particularly due to a lack of training or supportive workplace policies. Access to health care is often the first barrier faced by vulnerable populations, but even when access is attained, it does not guarantee that empathetic, culturally competent care will be provided. HCPs can struggle to adequately address patient questions, concerns, or misinformation effectively.

Individuals experiencing vulnerabilities have to deal with the dual challenges of limited access to care and credible health information. This can make them more vulnerable to exposure to low-quality information or misinformation that could negatively influence their perceptions or health behaviors. The COVID-19 pandemic provided many examples, including vaccine hesitancy among pregnant people based on mixed messaging and misinformation [42], some communities such as migrants, refugees, and other minority ethnic groups not seeking health advice reflecting their information needs, language, or values [43]. HCPs did not always meet these information needs, nor did wider health systems [44].

Some HCPs are actively addressing misinformation while others spread it, reflecting a lack of consistency in standards, training, and expectations of HCPs by health systems. However, HCPs report that since the COVID-19 pandemic, misinformation is

XSL•FO
RenderX

affecting how they provide care, and that it is no longer just an inconvenient side effect of living in the modern age.

Health misinformation is widespread on a large variety of topics of interest to people in the perinatal period [45]. How people sought it was based on who they trusted and what information sources they had access to, including HCPs. In communities that did not trust HCPs or had limited access to them, people sought out information from trusted networks including family, friends, and online support groups to help them cope, and find community and information that they could not easily obtain elsewhere. For some vulnerable groups, such as LGBTQ+ people, health information was sometimes entirely missing. This information void also affects HCPs, who may provide misinformation or provide poor, nonaffirming care to their clients.

HCPs need to broaden their knowledge of the related and trending topics their clients may have questions and concerns about and not assume that clients are operating only on credible information sources on which to base their health decisions. HCPs also should recognize that they compete for their patients' attention and trust against more emotionally compelling, slickly packaged, and immediately accessible social media content.

Health literacy emerged as a meta-theme based on numerous themes about information quality, culture, accessibility, and format—ideas that surfaced for every domain. Specifically, components of literacy [46] appeared on the individual level for clients about context, skills, knowledge, and psychological factors. Importantly, lack of digital literacy skills was not simply a challenge identified among patients, but also among HCPs, where lack of formal training or unclear workplace norms exacerbate this skills gap.

However, effective communication with patients takes more than a connection and literacy skills. Dozens of papers highlight issues of racism, stigma, discrimination, poor care, or care that even aggravates pre-existing trauma that HCPs can project onto their patients. Patients who already experience multiple vulnerabilities also face much higher hurdles to arrive at an HCP's office, if misinformation and access issues have not dissuaded them from trying to seek access. Treating them with care and respectful communication is a critical factor for them to follow health care advice and seek care in the future.

Intercultural competence and understanding the digital information environment and how it affects their patients are no longer optional for HCPs. Better interventions could be cocreated with patients, communities, and providers to address digital information environment challenges effectively. Such intervention could be better keyed to the information needs of vulnerable groups and take into account the complex web of information and social networks in digital spaces that are beyond the typical scope of HCPs. Given the widespread nature of health misinformation and other digital challenges, addressing them will require community-oriented solutions. Community-based solutions to combat misinformation could include cocreated interventions involving HCPs, patients, and community leaders, aimed at developing digital health literacy programs tailored to the specific needs of vulnerable groups, ensuring culturally relevant and accessible content.

Strengths and Limitations

The strength of this scoping review is its systematic approach, providing a snapshot of current digital information environment challenges and how they relate to midwives, maternal HCPs, and clients, including during the COVID-19 pandemic. However, this review is not exhaustive and did not include languages other than English, additional scientific databases, or gray literature. In future reviews, the inclusion of non-English language studies and gray literature could enrich the analysis, offering a broader, more international perspective on how different countries and cultures are addressing digital misinformation in health care. Additionally, as MeSH terms related to misinformation are largely categorized under communication rather than medical or public health topics, there is a risk that some relevant papers may have been overlooked. In the initial development of search strings, the terms "midwife" and "pregnancy" were included to focus on the health professional type in focus and the health state in focus for this paper; however, the search terms also led to the inclusion of broader maternal health care provider categories, which led to the inclusion of papers that did not specifically target midwives but were potentially relevant to them. This synthesis acknowledges limitations in the assessment of quality and access to digital information. A scoping review approach was used to map out the current landscape, ensuring the inclusion of diverse perspectives while recognizing gaps in data. Future work should focus on addressing disparities in access to technology and digital literacy among different populations.

Conclusion and Implications for Practice

The challenges presented by the digital information environment—misinformation, information voids, unaddressed questions and concerns, and lack of access to high-quality health information—are global issues. These barriers to optimal health are particularly detrimental for individuals seeking maternal health services, especially for those facing social, economic, or health vulnerabilities. Some barriers affect both patients and HCPs in achieving a healthier digital information environment that would promote accurate, accessible health information that people trust.

Despite these challenges, HCPs, especially midwives, are well-positioned to address misinformation and other information environment challenges. This is because HCPs are trusted sources of health information, including online. They also have pre-existing relationships with clients and communities and can extend relationships and continuity of care in digital spaces. HCPs can translate health guidance and evidence through more mediums, especially when digital "work" and tools are mainstreamed into care. HCPs can also empower clients with health education and promote digital health literacy. HCPs can also provide empathetic care with enhanced intercultural competencies through updated training.

However, to enable midwives and other HCPs to effectively address digital information environment challenges, more policy and health systems–level shifts are needed. HCPs should receive improved training in digital literacy, cultural competency, and caring for populations with specific vulnerabilities. Similar to how an HCP may ask questions related to identifying risk for

XSL•FO
RenderX

intimate partner violence or about nutrition and dietary habits and tailor their advice and care accordingly, understanding how the digital information environment is affecting their patients may be a new type of risk assessment offered. All HCPs should also benefit from work environments that promote mental health, and provide needed training and additional informational support for serving diverse patient populations.

The digital information environment will continue to grow in influence and impact health care delivery and information-seeking behaviors. Therefore, HCPs should be better supported through training and inclusion in health visits to address misinformation and other related issues. Health systems must improve their communication strategies, particularly in digital spaces, to provide accurate, timely, and culturally relevant health information to diverse audiences. One promising approach is to encourage HCPs to play an active role in digital spaces, leveraging the trust patients place in them to educate the public, engage with specific communities, and address misinformation in online communities where official health communication might not reach. Furthermore, health systems must prioritize the development of digital literacy training for both patients and HCPs, and adopt policies that support culturally competent, empathetic, and inclusive care.

Although most papers included in this review focused on general groups of maternal HCPs rather than specifically midwives, there are key insights particularly relevant to midwives. Midwives have long-term relationships with their pregnant clients while also addressing the sexual and reproductive health needs of many groups of people, from adolescents to refugees to LGBTQ+ people. Midwives are well-positioned at the client level and community level to listen, educate, and engage in offline and offline spaces, and play a crucial role in addressing health-related questions, concerns, and misinformation. It is imperative to include digital literacy, social media, cultural competency, and interpersonal communication training more firmly into midwifery education and extending midwifery practices into digital spaces. This would help midwives promote the values of empowerment, respect, and collaboration in this new digital frontier and can help them better serve the needs of clients who are increasingly online and navigating complex digital health landscapes.

Acknowledgments

EW conceptualized and implemented the work, managed the data, coordinated the coding and analysis, and drafted this paper. VV, JC-A, EP, and TDP conceptualized this research and reviewed the coding, analysis, and results. MT, PG, ET, EA, and FA coded and analyzed the data. All authors discussed the results and contributed to the final paper. No large language models such as ChatGPT, Claude, or Copilot were used in the preparation and writing of this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.
[PDF File (Adobe PDF File), 95 KB - jmir_v27i1e70108_app1.pdf]

Multimedia Appendix 2
Themes identified from analysis of information environment challenges to health care professionals and their clients; each Journey to Health graphic shows the mapping of themes for health care professional, client, topic, vulnerabilities, solutions, and a consolidated overview.
[PDF File (Adobe PDF File), 186 KB - jmir_v27i1e70108_app2.pdf]

References

1. Ghiasi A. Health information needs, sources of information, and barriers to accessing health information among pregnant women: a systematic review of research. J Matern Fetal Neonatal Med 2021;34(8):1320-1330. [doi: 10.1080/14767058.2019.1634685] [Medline: 31216921]
2. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
3. John JN, Gorman S, Scales D, Gorman J. Online misleading information about women's reproductive health: a narrative review. J Gen Intern Med 2024. [doi: 10.1007/s11606-024-09118-6] [Medline: 39511120]

XSL•FO
RenderX

4. Lu Y, Barrett LA, Lin RZ, Amith M, Tao C, He Z. Understanding information needs and barriers to accessing health information across all stages of pregnancy: systematic review. JMIR Pediatr Parent 2022;5(1):e32235 [FREE Full text] [doi: 10.2196/32235] [Medline: 35188477]
5. Yee L, Simon M. The role of the social network in contraceptive decision-making among young, African American and Latina women. J Adolesc Health 2010;47(4):374-380 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.03.014] [Medline: 20864007]
6. Southwell BG, Otero Machuca J, Cherry ST, Burnside M, Barrett NJ. Health misinformation exposure and health disparities: observations and opportunities. Annu Rev Public Health 2023;44(1):113-130 [FREE Full text] [doi: 10.1146/annurev-publhealth-071321-031118] [Medline: 36207008]
7. Scales D, Gorman J. Screening for information environments: a role for health systems to address the misinformation crisis. J Prim Care Community Health 2022;13:21501319221087870 [FREE Full text] [doi: 10.1177/21501319221087870] [Medline: 35352578]
8. Nan X, Wang Y, Thier K. Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. Soc Sci Med 2022;314:115398. [doi: 10.1016/j.socscimed.2022.115398] [Medline: 36327631]
9. Kelley DE, Lewis MA, Southwell BG. Perceived support from a caregiver's social ties predicts subsequent care-recipient health. Prev Med Rep 2017;8:108-111 [FREE Full text] [doi: 10.1016/j.pmedr.2017.08.001] [Medline: 28966907]
10. Shang L, Zuo M, Ma D, Yu Q. The antecedents and consequences of health care professional-patient online interactions: systematic review. J Med Internet Res 2019;21(9):e13940 [FREE Full text] [doi: 10.2196/13940] [Medline: 31573908]
11. De Sousa Machado T, Chur-Hansen A, Due C. First-time mothers' perceptions of social support: recommendations for best practice. Health Psychol Open 2020;7(1):2055102919898611 [FREE Full text] [doi: 10.1177/2055102919898611] [Medline: 32095254]
12. Vivilaki V, Wilhelm E, Petelos E. Infodemic management challenges and evidence-based midwifery. Eur J Midwifery 2023;7:21-24 [FREE Full text] [doi: 10.18332/ejm/168728] [Medline: 37636830]
13. Brady S, Gibbons KS, Bogossian F. Defining woman-centred care: a concept analysis. Midwifery 2024;131:103954 [FREE Full text] [doi: 10.1016/j.midw.2024.103954] [Medline: 38364459]
14. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021;9(12):1740 [FREE Full text] [doi: 10.3390/healthcare9121740] [Medline: 34946466]
15. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
16. Essential Competencies for Midwifery Practice. International Confederation of Midwives. URL: https://internationalmidwives.org/resources/essential-competencies-for-midwifery-practice/ [accessed 2025-02-01]
17. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022:1-24 [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
18. Patrick M, Venkatesh RD, Stukus DR. Social media and its impact on health care. Ann Allergy Asthma Immunol 2022;128(2):139-145 [FREE Full text] [doi: 10.1016/j.anai.2021.09.014] [Medline: 34555532]
19. Ghahramani A, de Courten M, Prokofieva M. 'The potential of social media in health promotion beyond creating awareness: an integrative review'. BMC Public Health 2022;22(1):2402 [FREE Full text] [doi: 10.1186/s12889-022-14885-0] [Medline: 36544121]
20. Levin-Zamir D, Bertschi IC, McElhinney E, Rowlands G. Digital environment social media as settings for health promotion. In: Kokko S, Baybutt M, editors. Handbook of Settings-Based Health Promotion. United States: Springer International Publishing; 2022:205-224.
21. Janes G, Chesterton L, Heaslip V, Reid J, Lüdemann B, Gentil J, et al. Current nursing and midwifery contribution to leading digital health policy and practice: an integrative review. J Adv Nurs 2025;81(1):116-139 [FREE Full text] [doi: 10.1111/jan.16265] [Medline: 38946119]
22. Chaney SC, Mechael P. So many choices, how do I choose? Considerations for selecting digital health interventions to support immunization confidence and demand. J Med Internet Res 2023;25:e47713 [FREE Full text] [doi: 10.2196/47713] [Medline: 37223980]
23. Gum L, Brown A, Royals R, Matricciani L, Kelly MA. Digital professionalism in preregistration nursing and midwifery students: a scoping review to explore perceptions of professionalism when using social media. Nurse Educ Pract 2024;80:104128 [FREE Full text] [doi: 10.1016/j.nepr.2024.104128] [Medline: 39293167]
24. Lawrence V, Richardson S, Philp L. How does social media influence expectations, decision making and experiences of childbirth? British Journal of Midwifery 2023;31(4):210-219 [FREE Full text] [doi: 10.12968/bjom.2023.31.4.210]
25. Griffin G, Williams N, Bradfield Z, Hauck YL. E-professionalism and social media use amongst nurses and midwives: a cross-sectional study. Nurse Educ Pract 2021;57:103248 [FREE Full text] [doi: 10.1016/j.nepr.2021.103248] [Medline: 34775225]
26. Schneider-Kamp A, Takhar J. Interrogating the pill: rising distrust and the reshaping of health risk perceptions in the social media age. Soc Sci Med 2023;331:116081 [FREE Full text] [doi: 10.1016/j.socscimed.2023.116081] [Medline: 37441974]

XSL•FO
RenderX

27. Vivilaki V, Chronaki C, Barbounaki S, Petelos E. Accelerating the digital transformation of community midwifery during the COVID-19 pandemic. Eur J Midwifery 2021;5:44 [FREE Full text] [doi: 10.18332/ejm/142571] [Medline: 34708191]
28. Wakelin K, McAra-Couper J, Fleming T, Erlam G. Exploring the ways communication technology is used by midwives and pregnant women/people: an integrative review. N Z Coll Midwives J 2022;58:11-18 [FREE Full text] [doi: 10.12784/nzcomjnl58.2022.2.11-18]
29. Human Centered Design 4 Health. UNICEF. 2019. URL: https://www.hcd4health.org/resources [accessed 2025-02-01]
30. Purnat TD, Nguyen T, Briand S. Managing Infodemics in the 21st Century. Cham: Springer International Publishing; 2023.
31. Zenone M, Kenworthy N, Maani N. The social media industry as a commercial determinant of health. Int J Health Policy Manage 2023;12:1-4 [FREE Full text] [doi: 10.34172/ijhpm.2022.6840] [Medline: 35490262]
32. Kirkpatrick CE, Lawrie LL. TikTok as a source of health information and misinformation for young women in the United States: survey study. JMIR Infodemiology 2024;4:e54663 [FREE Full text] [doi: 10.2196/54663] [Medline: 38772020]
33. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. The Lancet Digital Health 2021;3(3):e175-e194 [FREE Full text] [doi: 10.1016/s2589-7500(20)30315-0]
34. Ming WK, Huang F, Chen Q, Liang B, Jiao A, Liu T, et al. Understanding health communication through Google Trends and news coverage for COVID-19: multinational study in eight countries. JMIR Public Health Surveill 2021;7(12):e26644 [FREE Full text] [doi: 10.2196/26644] [Medline: 34591781]
35. Kbaier D, Kane A, McJury M, Kenny I. Prevalence of health misinformation on social media-challenges and mitigation before, during, and beyond the COVID-19 pandemic: scoping literature review. J Med Internet Res 2024;26:e38786 [FREE Full text] [doi: 10.2196/38786] [Medline: 39159456]
36. Calleja N, AbdAllah A, Abad N, Ahmed N, Albarracin D, Altieri E, et al. A public health research agenda for managing infodemics: methods and results of the first WHO infodemiology conference. JMIR Infodemiology 2021;1(1):e30979 [FREE Full text] [doi: 10.2196/30979] [Medline: 34604708]
37. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Chaudhuri N, Gupta G, Bagherzadeh M, Daim T, Yalcin H. Misinformation on social platforms: a review and research agenda. Technol Soc 2024;78:102654 [FREE Full text] [doi: 10.1016/j.techsoc.2024.102654]
39. Graves L, Bélair-Gagnon V, Larsen R. From public reason to public health: professional implications of the 'Debunking Turn' in the global fact-checking field. Digital Journalism 2023;12(10):1417-1436 [FREE Full text] [doi: 10.1080/21670811.2023.2218454]
40. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
41. Caceres MMF, Sosa JP, Lawrence JA, Sestacovschi C, Tidd-Johnson A, Rasool MHU, et al. The impact of misinformation on the COVID-19 pandemic. AIMS Public Health 2022;9(2):262-277 [FREE Full text] [doi: 10.3934/publichealth.2022018] [Medline: 35634019]
42. Hsu A, Johnson T, Phillips L, Nelson TB. Sources of vaccine hesitancy: pregnancy, infertility, minority concerns, and general skepticism. Open Forum Infect Dis 2022;9(3):ofab433 [FREE Full text] [doi: 10.1093/ofid/ofab433] [Medline: 35141344]
43. Goldsmith LP, Rowland-Pomp M, Hanson K, Deal A, Crawshaw AF, Hayward SE, et al. Use of social media platforms by migrant and ethnic minority populations during the COVID-19 pandemic: a systematic review. BMJ Open 2022;12(11):e061896 [FREE Full text] [doi: 10.1136/bmjopen-2022-061896] [Medline: 36396309]
44. do Nascimento IJB, Beatriz Pizarro AB, Almeida JM, Azzopardi-Muscat N, André Gonçalves M, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022;100(9):544-561. [doi: 10.2471/blt.21.287654]
45. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
46. NNLM Health Literacy Framework. Network of the National Library of Medicine. National Library of Medicine. URL: https://allofus-dev.nnlm.gov/about/health-literacy-framework [accessed 2025-02-01]

Abbreviations

HCP: health care professional
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other identities
MeSH: Medical Subject Headings
mHealth: mobile health
PICO: population, intervention, comparison, and outcome
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RATS: relevance, appropriateness, transparency, soundness
SOP: standard operating procedure

XSL•FO
RenderX

UNICEF: United Nations International Children's Emergency Fund

Edited by A Mavragani; submitted 15.12.24; peer-reviewed by B Southwell, S Gorman; comments to author 21.12.24; revised version received 31.12.24; accepted 22.01.25; published 25.02.25.

Please cite as:
Wilhelm E, Vivilaki V, Calleja-Agius J, Petelos E, Tzeli M, Giaxi P, Triantiafyllou E, Asimaki E, Alevizou F, Purnat TD
Effects of the Modern Digital Information Environment on Maternal Health Care Professionals, the Role of Midwives, and the People in Their Care: Scoping Review
J Med Internet Res 2025;27:e70108
URL: https://www.jmir.org/2025/1/e70108
doi:10.2196/70108
PMID:39998875

©Elisabeth Wilhelm, Victoria Vivilaki, Jean Calleja-Agius, Elena Petelos, Maria Tzeli, Paraskevi Giaxi, Elena Triantiafyllou, Eleni Asimaki, Faye Alevizou, Tina D Purnat. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Features of Structured, One-to-One Videoconference Interventions That Actively Engage People in the Management of Their Chronic Conditions: Scoping Review

Yu-Ting Chen[1], BSc, MSc, PhD; Michelle Lehman[1,2], BSc, MSc, OT Reg (NS); Toni Van Denend[3], BS, MS, OTD, OTR/L; Jacqueline Kish[3,4], BA, MA, PhD, OTR/L; Yue Wu[5,6], BA, MA, MM, PhD, MT-BC; Katharine Preissner[3], BSc, MHSc, EdD; Matthew Plow[7], BA, PhD; Tanya L Packer[1,8,9], BSc, MSc, PhD, OT Reg (NS)

[1]School of Occupational Therapy, Faculty of Health, Dalhousie University, Halifax, NS, Canada

[2]Think Self-Management Inc., Halifax, NS, Canada

[3]Department of Occupational Therapy, University of Illinois Chicago, Chicago, IL, United States

[4]Center for Rehabilitation Outcomes Research, Shirley Ryan AbilityLab, Chicago, IL, United States

[5]Department of Rehabilitation Medicine, University of Minnesota-Twin Cities, Minneapolis, MN, United States

[6]Institute on Community Integration, University of Minnesota-Twin Cities, Minniapolis, MN, United States

[7]Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, United States

[8]School of Health Administration, Faculty of Health, Dalhousie University, Halifax, NS, Canada

[9]Department of Nursing, Umeå University, Umeå, Sweden

Corresponding Author:
Tanya L Packer, BSc, MSc, PhD, OT Reg (NS)
School of Health Administration
Faculty of Health
Dalhousie University
Sir Charles Tupper Medical Building, 2nd Floor
5850 College St., PO Box 15000
Halifax, NS, B3H 4R2
Canada
Phone: 1 902 494 7097
Email: tanya.packer@dal.ca

Abstract

Background: A dramatic increase in the use of videoconferencing occurred as a response to the COVID-19 pandemic, including delivery of chronic disease management programs. With this increase, clients' openness to and confidence in receiving any type of telehealth care has dramatically improved. However, the rapidity of the response was accomplished with little time to learn from existing knowledge and research.

Objective: The purpose of this scoping review was to identify features, barriers, and facilitators of synchronous videoconference interventions that actively engage clients in the management of chronic conditions.

Methods: Using scoping review methodology, MEDLINE, CINAHL, and 6 other databases were searched from 2003 onward. The included studies reported on structured, one-on-one, synchronous videoconferencing interventions that actively engaged adults in the management of their chronic conditions at home. Studies reporting assessment or routine care were excluded. Extracted text data were analyzed using thematic analysis and published taxonomies.

Results: The 33 included articles reported on 25 distinct programs. Most programs targeted people with neurological conditions (10/25, 40%) or cancer (7/25, 28%). Analysis using the Taxonomy of Every Day Self-Management Strategies and the Behavior Change Technique Taxonomy version 1 identified common program content and behavior change strategies. However, distinct differences were evident based on whether program objectives were to improve physical activity or function (7/25, 28%) or mental health (7/25, 28%). Incorporating healthy behaviors was addressed in all programs designed to improve physical activity or function, whereas only 14% (1/7) of the programs targeting mental health covered content about healthy lifestyles. Managing emotional distress and social interaction were commonly discussed in programs with objectives of improving mental health (6/25, 24% and 4/25, 16%, respectively) but not in programs aiming at physical function (2/25, 8% and 0%, respectively). In total, 13

XSL•FO
RenderX

types of behavior change strategies were identified in the 25 programs. The top 3 types of strategies applied in programs intent on improving physical activity or function were feedback and monitoring, goals and planning, and social support, in contrast to shaping knowledge, regulation, and identity in programs with the goal of improving mental health. The findings suggest that chronic condition interventions continue to neglect evidence that exercise and strong relationships improve both physical and mental health. Videoconference interventions were seen as feasible and acceptable to clients. Challenges were mostly technology related: clients' comfort, technology literacy, access to hardware and the internet, and technical breakdowns and issues. Only 15% (5/33) of the studies explicitly described compliance with health information or privacy protection regulations.

Conclusions: Videoconferencing is a feasible and acceptable delivery format to engage clients in managing their conditions at home. Future program development could reduce siloed approaches by adding less used content and behavior change strategies. Addressing client privacy and technology issues should be priorities.

(J Med Internet Res 2025;27:e58543) doi:10.2196/58543

KEYWORDS

videoconference; chronic disease management; active participation; intervention program; self-management; scoping review; Taxonomy of Every Day Self-Management Strategies; TEDSS; Behavior Change Technique Taxonomy version 1; BCTTv1; behavior change; mobile phone

Introduction

Background

With the rise in popularity and availability of technology and the restrictions imposed by the COVID-19 pandemic, many health care providers have turned to telehealth care as an alternative, enhancement, or complement to face-to-face health care delivery [1]. Telehealth care refers to any remote communication or IT between clients and health care providers (eg, phone calls, SMS text messages, emails, and videoconferencing). In the United States, the use of telehealth care increased close to 30 times between September 2019 and September 2020, with many clients becoming first-time users during the pandemic [1]. As such, clients' openness to and confidence in receiving any type of telehealth care has dramatically improved. For example, approximately 80% of clients report being satisfied with telehealth care, and 75% wish it to continue as a regular part of their care. However, the pandemic-driven shift to telehealth care was rapid and reactive to public health measures introduced to control the spread of the disease. For this reason, knowledge about effective practices, facilitators, and barriers to using telehealth care to support adults with chronic conditions is limited or incomplete. Without this knowledge, the integration of telehealth care into regular care will continue to be fragmented and reactive.

Synchronous videoconferencing has been proposed as "a viable alternative to face-to-face [health and medical] appointments" [2]. Using an internet-based platform, bidirectional audio and video signal is exchanged in real time (ie, synchronously) enabling verbal, nonverbal, and typewritten communication between health care providers and clients. A scoping review published in 2014 identified >500 studies using synchronous videoconferencing in client care between 2002 and 2012, with a steep upward increase in publications at the end of this period [3]. The review revealed that videoconferencing was used by a wide variety of health care professionals to make diagnoses, provide consultation, monitor client compliance or progress, and support clients in managing both acute and chronic physical and mental conditions. While this review provided a broad picture of how videoconferencing was being used, it did not distinguish between routine care visits and chronic condition management programs, nor did it investigate the content or active ingredients in the interactions or report the facilitators and barriers afforded by this format of service delivery. In-depth reviews in focused areas of clinical care are needed to inform optimal practice.

The Centers for Disease Control and Prevention reports that 60% of adults in the United States live with one or more chronic conditions, making chronic conditions the leading cause of death and disability and a key driver of annual health care costs [4]. Living with chronic conditions requires active engagement in ongoing medical, role, and emotional management [4,5]. This work, commonly termed self-management, includes seeking support, making decisions, and altering behaviors in the context of everyday life [6-8]. In contrast to passive receipt of educational information, active client participation is a hallmark of "informed, activated patients" described in the widely endorsed chronic care model [9,10]. The active involvement of individuals in these interventions sets them apart from many other medical interventions and warrants specific investigation.

Objectives

Preventing chronic conditions by promoting and enabling self-management is now widely regarded as critical to improving client outcomes and reducing demands on health systems [6,7]. Therefore, in-person chronic disease management and self-management programs have proliferated. Many, particularly those for older adults, who are the most likely to have chronic conditions, are condition specific and have structured content that is delivered in a 6- to 8-week period, usually consisting of weekly sessions [11]. Self-management programs are known to effectively improve health outcomes and reduce health system demands [6,7]. As such, they have become a primary focus of health service policy, redesign, and research [9,12], with delivery both in person and via telehealth care. The purpose of this scoping review was to examine how videoconferencing has been used to deliver structured, synchronous, one-to-one interventions to clients in their own homes and actively engage them in the prevention and management of their chronic conditions.

XSL•FO
RenderX

Methods

Design

A scoping review is "a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field" [13]. This methodology was selected specifically because the features of videoconferencing interventions to support adults with chronic conditions remain unclear. The 5-stage methodological framework by Arksey and O'Malley [14] was followed with augmentation consistent with recent evolutions of the scoping review methodology [15,16]. Focused on conceptual knowledge synthesis, scoping reviews do not typically assess the quality of the studies, nor are they intended to assess effectiveness. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [17] proposed by the Enhancing the Quality and Transparency of Health Research Network was used to guide reporting of study conduct and findings (Multimedia Appendix 1). The protocol was developed and registered on the Open Science Framework [18]. Ethics approval was not sought or required for this review article.

Identifying the Research Question

The research question was discussed and defined by an international research team comprising occupational therapists or researchers with expertise in chronic disease management. The agreed upon question—"What is known about the theoretical foundation, purpose and contents, active ingredients, program structure, technology, and facilitators and barriers of structured, synchronous, one-on-one videoconference interventions that actively engage clients who are living at home in the management of their chronic conditions?"—reflects the intent to help researchers and clinicians create, tailor, or transfer in-person interventions to internet-based delivery. The focus on structured, synchronous, one-on-one videoconferencing interventions was chosen because they most closely imitate in-person chronic disease interventions. Client location during the videoconference was restricted to their home.

Chronic conditions are defined broadly as "conditions that last one year or more and require ongoing medical attention or limit activities of daily living or both" [4]. In this review, we interpreted this to include populations medically at risk, such as people with obesity or frailty or requiring common interventions such as knee arthroplasty. Features such as program structure, asynchronous activities, videoconferencing technology, and program feasibility and acceptability were operationalized a priori (Textbox 1). An existing taxonomy and an existing framework were adopted to describe and categorize program content and behavior change strategies. The Taxonomy of Every Day Self-Management Strategies (TEDSS) [19] was used to operationalize program content. Active ingredients or strategies facilitating behavior change were operationalized using the Behavior Change Technique Taxonomy version 1 (BCT Taxonomy v1) hierarchical clusters [20] (see the Collating, Summarizing, and Reporting the Results section for details of each). Program purpose and theoretical foundation were not defined a priori, relying instead on the descriptions provided by each author. Similarly, barriers and facilitators were identified based on those reported in the text or quantitative reports of client satisfaction found in the study results.

Textbox 1. A priori operational definitions of key terms.

Term and operational definition or selected taxonomy
• Chronic conditions: "conditions that last one year or more and require ongoing medical attention or limit activities of daily living or both" [4]
• Program structure: intervention duration, frequency, and asynchronous activities
• Asynchronous activities: activities undertaken outside or between the videoconferencing sessions, such as homework or symptom or activity monitoring
• Videoconferencing technology: software and hardware used or supplied, compatibility and connectivity requirements, data security, and any efforts made to set up or train clients to use the technology at home
• Program content: the Taxonomy of Every Day Self-Management Strategies framework [19]
• Active ingredients and behavior change strategies: Behavior Change Techniques taxonomy version 1 hierarchical clusters [20]
• Feasibility and acceptability: retention, attrition, and attendance rates

Identifying Relevant Studies

The search strategy was developed and supported by a medical librarian with expertise in systematic and scoping reviews. Electronic databases in the field of health science and education (MEDLINE, PubMed, CINAHL, Embase, PsycINFO, OTseeker, PEDro, Cochrane, and ERIC) were searched using the following keywords—"videoconferencing," "teleconferencing," "virtual care," "skype," "teleconsultation," "video call," "telerehab," "ehealth," and "digital"—to identify videoconference interventions. These keywords reflect the wide array of terminology used to describe videoconferencing in health care interventions and client education. Because PsycINFO and ERIC often include healthy samples or populations, the keywords "patients," "illness," and "disease" were added to those searches. Keywords related to self-management were deliberately not included in the search strategies. Review of preliminary searches revealed that many interventions meeting the criteria of active engagement and self-management of chronic conditions were excluded when these keywords were added. Because their addition greatly reduced the number of studies found, selection based on active engagement was conducted during the title, abstract, and full-text screening processes. The search was limited to articles published in

XSL•FO
RenderX

English. The initial search took place on June 3, 2020. In response to the proliferation of published literature, the search was repeated on January 25, 2023, adding the period from January 2020 to January 2023. The combined searches included articles reporting on adult samples published within the 20-year period between January 2003 and January 2023. Studies published before 2003 were excluded due to the limited currency of videoconferencing technology at that time. Detailed search strategies for each database and the number of hits from each are presented in Multimedia Appendix 2.

Study Selection

Article inclusion and exclusion criteria were defined for (1) participants, (2) intervention program technology and format, and (3) intervention content and features (Table 1). Review studies and protocol papers were included. Review studies were further checked to identify missing articles fitting the selection criteria. Because protocol papers often provide rich information about program content, they were included to help map intervention structure, content, and active ingredients, recognizing that there would be no extractable data on facilitators and barriers. Routine care, defined as having regular meetings to diagnose or monitor clients' conditions; compliance with a treatment or medication regime; or treatment without a predefined structure, topics, or modules during videoconferencing were excluded.

Table 1. Inclusion and exclusion criteria for the studies—videoconference interventions for people with chronic conditions.

Category and subcategory	Inclusion criteria	Exclusion criteria
Participants		
Adults aged ≥18 years	• Adults aged ≥18 years	• Dyads of parents and children aged <18 years
Community dwelling	• Community dwelling (eg, private home, supported housing, or group homes)	• Individuals living in a hospital, long-term care facilities, or mental health facilities
One or more chronic conditions	• Having at least one chronic or medically at-risk condition requiring ongoing medical attention for >1 year; the condition must impact activities of daily living (eg, heart condition, diabetes, obesity, or knee arthroplasty)	• Caregivers as the focus of the intervention • Health professionals or trainees as the focus of the intervention • No indication of chronic conditions (eg, inactive adults, people seeking psychological consultations, smokers, and veterans)
Program technology and format		
One-on-one	• Individual meeting in which the service provider saw 1 client or 1 client with their adult companion or caregiver	• Group videoconference or one videoconference with multiple individuals at the same time (eg, educational session for multiple people)
Synchronous	• Concurrent presence of the client or client and adult companion or caregiver and service provider in real time	• Programs delivered completely asynchronously
Videoconferencing at home	• Internet-based communication through transmitted audio and video signals; the client is located at home (for RCTs[a], at least one arm had to meet this criterion)	• Communication via telephone call only or combining telephone call and videoconferencing • Videoconferencing outside the client's home (eg, community center or clinics) • Information transmission through virtual reality, robotic interventions, or platforms without interaction (eg, electronic monitoring system or automatic graphical feedback)
Intervention program content and features		
Structured program	• Interventions with a stated goal or purpose and a structure (topics, modules, frequency, and duration of sessions)	• Diagnostic and assessment-focused studies • Routine care or follow-up sessions without structured intervention modules
Active client participation	• Evidence of active client participation (eg, verbal, physical, or cognitive), including coaching and healthy behavior monitoring	• Studies that only asked participants to listen to a lecture or read study materials (passive educational approach)

[a]RCT: randomized controlled trial.

Study selection protocols for the initial search (June 2020) are described in detail. The protocols for the updated search (January 2023) were identical unless otherwise noted. At each time point, search results were imported into the Covidence platform (Veritas Health Innovation Ltd) for screening and selection. Covidence automatically removed duplicate articles.

XSL•FO
RenderX

A total of 8 researchers participated in the review process (Multimedia Appendix 3). The lead reviewers (reviewers 3 and 6) together with reviewers 2, 4, and 5 were involved throughout both the initial and updated searches. Reviewers 1 and 2 were involved in the initial search. Reviewers 7 and 8 were involved in the updated search.

Before the initial title and abstract screening, interrater reliability trials were completed to ensure that reviewers were consistent in their selection. Reviewers 1 to 6 read the same 5 articles, independently deciding whether they met the inclusion and exclusion criteria. Conflicts were discussed, and operational definitions of the eligibility criteria were refined. This process was repeated until the κ interrater agreement level reached 0.8 [13,21]. With strong interrater reliability confirmed, title and abstract searches were completed by 1 of 5 reviewers (reviewers 2 to 6). Reviewers 3, 5, and 6 were joined by reviewers 7 and 8 to complete the updated title and abstract screening. New reviewers were provided with the written operational definitions and extensively oriented before commencing their reviews.

The full texts of retained studies were downloaded or ordered through interlibrary loan and imported into Covidence. All identified articles were located and retrieved. As with the initial title and abstract search, interrater reliability testing achieved a κ agreement of at least 0.8 before beginning the final selection of articles. In total, 2 reviewers independently screened each full-text article. For both the original and updated searches, conflicts were discussed and resolved by a full professor (reviewer 6) and a postdoctoral fellow (reviewer 3); for the updated search, conflicts were resolved by the same professor and an experienced research occupational therapist (reviewer 7).

A total of 94 studies met the inclusion criteria after the first full-text screening. Close examination revealed three distinct intervention types: (1) interventions specifically designed for individuals with mental health conditions (eg, depression, anxiety, and posttraumatic stress disorder), (2) solely exercise-based or repetitive learning interventions (eg, motor learning approach, muscle-strengthening exercises, and word-finding therapy) with no evidence of active client decision-making, and (3) interventions intended to support participants to actively and deliberately manage everyday life with a chronic condition. Exercise, activity-based, or repetitive interventions that explicitly included components to manage everyday life (eg, problem-solving, information seeking, and decision-making) were included in the third group. Given that existing reviews have reported on mental health populations [22-26] and exercise-based or repetitive interventions [27-29], this manuscript reports an analysis of the third group of interventions, for which there is less evidence.

During the original search, assignment of articles to the 3 intervention types was first completed by a single, consistent reviewer (reviewer 3); 4 other reviewers each assessed one-quarter of the articles, meaning that all articles were assigned by a consistent reviewer and one other. Conflicts were resolved by the lead reviewers. When the search was updated, reviewers 3 and 6 both assigned all articles to an intervention type and resolved conflicts through discussion.

Charting the Data

Data extraction was completed in Covidence using extraction template 2.0. The extraction form was developed based on the research question and then pilot-tested by the lead reviewers, who independently extracted data from 2 of the included articles. The data extracted included general article information, research objectives, study design, participants (clients or recipients and clinicians or service providers), intervention purpose, content and features, videoconference technology and logistics, and facilitators of and barriers to delivering or receiving interventions using videoconference. A 2-hour workshop was held to familiarize team members with the data extraction process in Covidence. The 2 reviewers included for the updated search were similarly oriented. Double data extraction by 2 of 7 reviewers was then completed. The extracted data were checked, selected, or combined by 2 consistent reviewers for each search, resulting in comprehensive and rich data available for collation. Finally, the results were exported from Covidence to a CSV file and imported into Microsoft Excel for data analysis.

Collating, Summarizing, and Reporting the Results

Extracted data were managed using Microsoft Excel. Categorical data (eg, study country, type of study design, and software and hardware selected) were analyzed using descriptive analysis. Text-based data were analyzed using either a known theoretical framework or thematic analysis. Intervention content and active ingredients were analyzed using the TEDSS framework [19] and the BCT Taxonomy v1 hierarchical clusters [20], respectively. The TEDSS framework, derived from the literature and interviews with 117 individuals living with chronic neurological conditions, was chosen for its robust and comprehensive categorization of strategies commonly used and considered important by people living with chronic conditions [19,30,31]. The 7 distinctive TEDSS domains (Textbox 2) have been used to successfully identify and describe content in self-management interventions [32,33], including in scoping reviews [11,34].

XSL•FO
RenderX

Textbox 2. Domain definitions—Taxonomy of Every Day Self-Management Strategies (TEDSS).

TEDSS domain and definition

- Process strategies: strategies used to be well informed and make good decisions; often used to support use of other, nonprocess strategies
- Resource strategies: proactively seeking, pursuing, or managing needed formal or informal supports and resources
- Activity strategies: finding ways to participate in everyday activities (leisure activities, work activities, and household chores) despite problems such as fatigue, pain, memory loss, or disability
- Internal strategies: preventing and managing stress, negative emotions, and internal distress; creating inner calm
- Social interaction strategies: managing social interactions and relationships to be able to participate without exposure to negative reactions
- Health behavior strategies: maintaining a healthy lifestyle to enhance health and limit the risk of lifestyle-related illness
- Disease-controlling strategies: preventing, controlling, and limiting symptoms, complications, or disease progression

The BCT Taxonomy v1 [20] was used to categorize "observable, replicable, and irreducible component(s) of intervention[s] designed to alter or redirect causal processes that regulate behavior; that is, a technique proposed to be an 'active ingredient'" [20]. The BCT Taxonomy v1 is a well-known taxonomy used to describe behavior change interventions. In this review, we used the 16-cluster definition developed by the BCT Taxonomy v1 authors to code the strategies used to facilitate clients' behavior change [35].

All other text data (eg, theoretical foundation, intervention purpose, facilitators, and barriers) were coded and analyzed using thematic analysis. Keywords from the extracted text were first highlighted independently by 2 reviewers. Keywords were compared to generate themes before all extracted data were recoded using these themes. The extracted data (eg, intervention purpose, theoretical background, contents, and active ingredients) were cross-referenced to ensure that all the available data were included in the analysis. Data from both searches were combined before the number and percentage of programs assigned to each theme were calculated.

Results

Overview

The search of eligible articles from multiple databases generated 4067 potential articles in the initial search, with a further 1714 in the updated search. After removing duplicates, 4309 articles were screened by title and abstract, and 590 (13.69%) underwent full-text review. Of a total of 130 articles included after the full-text review, 33 (25.4%) [36-68] reported on self-management interventions for individuals with physical or cognitive impairments (ie, not solely exercise based or repetitive in nature or designed specifically for clients with a mental health diagnosis) and were included for final review (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

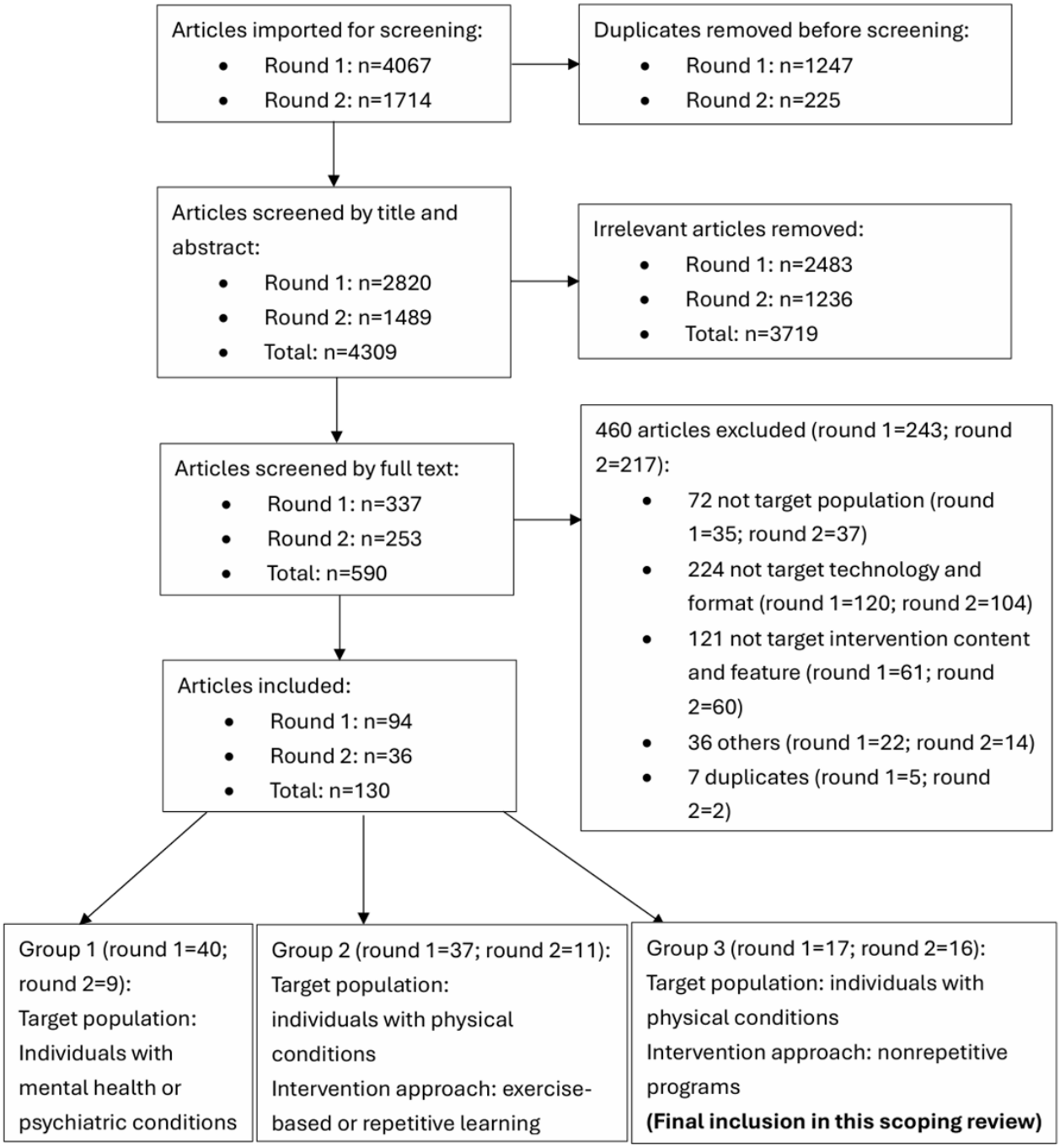

A total of 48% (16/33) of the studies were conducted or designed in the United States (Table 2). In total, 18% (6/33) of the studies were conducted in Australia and Canada each, with the remaining conducted in Israel (3/33, 9%) and the United Kingdom and Ireland (1/33, 3% each). Over half of the papers reported randomized controlled trials (12/33, 36%) or were protocol papers (8/33, 24%), and a further 18% (6/33) were pilot or feasibility studies. Other designs were less well represented (Table 2). All the included studies (33/33, 100%) were published after 2012, with 85% (28/33) published after 2018. The 33 studies reported on 25 unique intervention programs (Table 3).

XSL•FO
RenderX

Table 2. Features of the studies included in the scoping review of videoconference interventions for people with chronic conditions (N=33).

Study	Year	Country	Study design	Participants			Service providers	Intervention
				Condition	Age (y)	Sample size, N[a]		
Alencar et al [36]	2019	United States	RCT[b]	Obesity	≥18	15	Multidisciplinary	TEAM[c]
Alencar et al [37]	2020	United States	RCT	Obesity	≥18	13	Multidisciplinary	TEAM
Aubin et al [38]	2019	Canada	RCT	Cancer	18-39	30	Psychologist	Psychosocial intervention
Brunet et al [39]	2022	Canada	Protocol	Cancer	18-39	30-40	Physical activity counselor	ACCESS[d]
Chemtob et al [40]	2019	Canada	RCT	Spinal cord injury	≥18	12	Kinesiologist	LTPA[e]
Cruice et al [41]	2021	United Kingdom	Pilot or feasibility study	Stroke	37-81	29	Speech-language pathologist	Online SC[f] for participation intervention
Garland et al [42]	2021	Canada	Protocol	Cancer	≥18	124	Psychologist	CBT-I[g]
Gilboa et al [43]	2019	Israel	Protocol	Hip fracture	≥60	30	Occupational therapist	CO-OP[h]
Goren et al [44]	2022	Israel	RCT	IBD[i]	≥18	67	Social worker	COBMINDEX[j] program
Hastings et al [45]	2021	United States	RCT	Cognitive impairment or dementia	≥65	40	Nurse	Care management program
Kelleher et al [46]	2019	United States	RCT	Cancer	≥18	89	Psychologist	PCST[k]
Kline et al [47]	2019	United States	Protocol	Total knee arthroplasty	50-85	100	Multidisciplinary	Behavior change intervention
Lafaro et al [48]	2020	United States	Quasi-experimental study	Cancer	≥65	34	Multidisciplinary	Telehealth perioperative physical activity intervention
Lavelle et al [49]	2022	Ireland	Case report or case series	IBD	18-65	19	Psychologist	Brief and telehealth ACT[l]
Lawson et al [50]	2020	Australia	Pilot or feasibility study	Stroke	≥18	28	Psychologist	Modified Monash memory skills program
Lawson et al [51]	2022	Australia	Qualitative study	Stroke	≥18	34[m]	Psychologist	Modified Monash memory skills program
Lynch et al [52]	2016	United States	Protocol	Diabetes	>45	30	Nurse	TABLETS[n]
Milbury et al [53]	2020	United States	RCT	Cancer	≥18	75	Psychologist	CBM[o]
Miller et al [54]	2017	United States	Protocol	Lower limb amputation	>50	32	Physiotherapist	PABC[p]
Ng et al [55]	2013	Canada	Case report	TBI[q]	≥19	4	Occupational therapist	CO-OP
Ownsworth et al [56]	2019	Australia	Protocol	Cancer	≥18	148	Psychologist	Tele-MAST[r]
Pfammatter et al [57]	2022	United States	Protocol	Obesity	>18-70	—[s]	Health promotionist	EVO[t]
Rietdijk et al [58]	2019	Australia	Pilot or feasibility study	TBI	≥18	2	Speech-language pathologist	TBIconneCT
Rietdijk et al [59]	2020	Australia	RCT	TBI	18-70	19	Speech-language pathologist	TBIconneCT

XSL•FO
RenderX

Study	Year	Country	Study design	Participants			Service providers	Intervention
				Condition	Age (y)	Sample size, N[a]		
Somers et al [60]	2015	United States	Pilot or feasibility study	Cancer	≥19	25	Psychologist	PCST
Tanenbaum et al [61]	2021	United States	Mixed methods	Diabetes	18-50	22	Psychologist	ONBOARD[u]
Vellani et al [62]	2022	Canada	Pilot or feasibility study	Cognitive impairment or dementia	≥65	21	Nurse	VYV[v]
Winger et al [63]	2020	United States	RCT	Cancer	≥19	89	Psychologist	PCST
Winger et al [64]	2022	United States	Pilot or feasibility study	Cancer	≥18	30	Psychologist	MCPC[w]
Wood et al [65]	2022	United States	Case report or case series	Spinal cord injury	≥18	15	Dietitian	Tele–nutrition counseling program
Ymer et al [66]	2022	Australia	RCT	TBI or stroke	23-71	50[x]	Psychologist	CBT-SF[y]
Beit Yosef et al [67]	2022	Israel	RCT	TBI	≥18	8	Occupational therapist	Tele–CO-OP
Yuen [68]	2013	United States	Quasi-experimental study	Spinal cord injury	≥19	16	Occupational therapist	Telecare on oral health

[a]For protocol papers, the target sample size is reported.

[b]RCT: randomized controlled trial.

[c]TEAM: Telehealth-Enabled Approach to Multidisciplinary Care.

[d]ACCESS: Physical Activity Counseling for Young Adult Cancer Survivors.

[e]LTPA: Leisure Time Physical Activity.

[f]SC: supported conversation.

[g]CBT-I: cognitive behavioral therapy for insomnia.

[h]CO-OP: Cognitive Orientation to Daily Occupational Performance.

[i]IBD: inflammatory bowel disease.

[j]COBMINDEX: Cognitive Behavioral and Mindfulness-Based Stress Reduction With Daily Exercise.

[k]PCST: Pain Coping Skill Training.

[l]ACT: acceptance and commitment therapy.

[m]25 stroke survivors and 9 clinicians.

[n]TABLETS: Tablet-Aided Behavioral Intervention Effect on Self-Management Skills.

[o]CBM: couple-based meditation.

[p]PABC: Physical Activity Behavior Change.

[q]TBI: traumatic brain injury.

[r]Tele-MAST: Making Sense of Brain Tumour program.

[s]Not available.

[t]EVO: Elements Vital to Treat Obesity.

[u]ONBOARD: Overcoming Barriers and Obstacles to Adopting Diabetes Devices.

[v]VYV: Voice Your Values.

[w]MCPC: Meaning-Centered Pain Coping Skill Training.

[x]50 participants began the program, and 30 completed it.

[y]CBT-SF: cognitive behavioral therapy for sleep disturbance and fatigue.

XSL•FO
RenderX

Table 3. Structure, foundation, and objectives of structured videoconference intervention programs.

Intervention name	Total number of sessions	Frequency	Duration of session	Intervention objective	Theoretical foundation	Asynchronous component
Behavior change intervention [47]	≥10	Weekly	≤1 hour	• Improve physical activity or physical function	• Social cognitive theory or CBT[a] • Control theory	• Self-monitoring and tracking
PABC[b] [54]	≥10	Weekly	≤1 hour	• Improve physical activity or physical function	• Behavior change	• Self-monitoring and tracking
LTPA[c] [40]	5-9	Weekly	≤1 hour	• Improve physical activity or physical function	• Self-determination theory	• Not reported
TABLETS[d] [52]	5-9	Weekly+booster sessions	≤1 hour	• Improve physical activity or physical function • Improve quality of life • Other (diet, medication adherence, and self-monitoring behavior)	• Previously reported intervention	• Educational materials or access to website • Printable materials or manuals • Self-monitoring and tracking • Automated feedback
ACCESS[e] [39]	5-9	Not reported	≤1 hour	• Improve physical activity or physical function	• Behavior change • Self-determination theory	• Educational materials or access to website • Home practice or homework
Telehealth perioperative physical activity intervention [48]	≤4	Other than weekly	Not reported	• Improve physical activity or physical function	• Chronic care self-management model	• Printable materials or manuals • Self-monitoring and tracking • Other
Psychosocial intervention [38]	≤4	Other than weekly	≤1 hour	• Improve mental health or reduce stress • Improve quality of life	• Review of literature or clinical guidelines	• Printable materials or manuals
Tele-MAST[f] [56]	≥10	Weekly	≤1 hour	• Improve mental health or reduce stress • Improve quality of life	• Review of literature or clinical guidelines • Sense of coherence theory	• Printable materials or manuals
COBMINDEX[g] program [44]	5-9	Other than weekly	>1 hour	• Improve mental health or reduce stress	• Social cognitive theory or CBT	• Educational materials or access to website • Home practice or homework • Self-monitoring and tracking • Other
Brief and telehealth ACT[h] [49]	≤4	Weekly	>1 hour	• Improve mental health or reduce stress	• ACT	• Printable materials or manuals
CBM[i] [53]	≤4	Weekly	≤1 hour	• Improve mental health or reduce stress	• Previously reported intervention	• Printable materials or manuals • Home practice or homework • Other
TEAM[j] [36,37]	≥10	Weekly	≤1 hour	• Weight loss	• Not reported	• Self-monitoring and tracking

XSL•FO
RenderX

Intervention name	Total number of sessions	Frequency	Duration of session	Intervention objective	Theoretical foundation	Asynchronous component
Tele–nutrition counseling program [65]	5-9	Other than weekly	≤1 hour	• Weight loss • Improve quality of life	• Health belief model	• Educational materials or access to website • Home practice or homework • Self-monitoring and tracking
EVO[k] [57]	≥10	Other than weekly	≤1 hour	• Weight loss • Diabetes management	• Behavior change • Psychoeducation	• Educational materials or access to website • Home practice or homework • Self-monitoring and tracking
TBIconneCT [58,59]	≥10	Weekly	>1 hour	• Improve communication	• Previously reported intervention • Social cognitive theory or CBT	• Printable materials or manuals • Home practice or homework
Online SC[l] for participation [41]	≥10	Other than weekly	≤1 hour	• Improve communication • Improve quality of life • Improve mental health or reduce stress • Improve everyday participation or life roles	• Previously reported intervention	• Printable materials or manuals • Message board or social networking site
PCST[m] [46,60,63] and and MCPC [c] [64]	≤4	Not reported	≤1 hour	• Pain management	• Social cognitive theory or CBT	• Educational materials or access to website • Home practice or homework • Self-monitoring and tracking • Message board or social networking site
CO-OP[o] [43,55,67]	≥10	Other than weekly OR weekly[p]	≤1 hour	• Improve everyday participation or life roles	• Meichenbaum's SIT[q] or CO-OP	• Home practice or homework
Modified Monash memory skills program [50,51]	5-9	Weekly+booster sessions	>1 hour	• Improve everyday participation or life roles • Improve memory	• Previously reported intervention	• Home practice or homework
Telecare on oral health [68]	5-9	Other than weekly	≤1 hour	• Improve oral care	• Not reported	• Other
Care management program [45]	≥10	Other than weekly	Not reported	• Care management and planning • Improve physical activity or physical function	• Review of literature or clinical guidelines	• Not reported
VYV[r] [62]	≤4	Not reported	Not reported	• Care management and planning	• Behavior change • Review of literature or clinical guidelines • Representational approach to patient education	• Other
ONBOARD[s] [61]	≤4			• Diabetes management		• Not reported

XSL•FO
RenderX

Intervention name	Total number of sessions	Frequency	Duration of session	Intervention objective	Theoretical foundation	Asynchronous component
		Other than weekly	Not reported		• Review of literature or clinical guidelines • Technology acceptance model	
CBT-I[t] [42]	5-9	Weekly	≤1 hour	• Improve sleep	• Social cognitive theory or CBT	• Home practice or homework • Self-monitoring and tracking
CBT-SF[u] [66]	5-9	Weekly	≤1 hour	• Improve sleep • Improve mental health or reduce stress	• Social cognitive theory or CBT	• Not reported

[a]CBT: cognitive behavioral therapy.

[b]PABC: Physical Activity Behavior Change.

[c]LTPA: Leisure Time Physical Activity.

[d]TABLETS: Tablet-Aided Behavioral Intervention Effect on Self-Management Skills.

[e]ACCESS: Physical Activity Counseling for Young Adult Cancer Survivors.

[f]Tele-MAST: Making Sense of Brain Tumour program.

[g]COBMINDEX: Cognitive Behavioral and Mindfulness-Based Stress Reduction With Daily Exercise.

[h]ACT: acceptance and commitment therapy.

[i]CBM: couple-based meditation.

[j]TEAM: Telehealth-Enabled Approach to Multidisciplinary Care.

[k]EVO: Elements Vital to Treat Obesity.

[l]SC: supported conversation.

[m]PCST: Pain Coping Skill Training.

[n]MCPC: Meaning-Centered Pain Coping Skills Training.

[o]CO-OP: Cognitive Orientation to Daily Occupational Performance.

[p]A total of 12% (4/33) of the studies tested the same intervention and delivered it on both a weekly and nonweekly schedule.

[q]SIT: self-instructional training.

[r]VYV: Voice Your Values.

[s]ONBOARD: Overcoming Barriers and Obstacles to Adopting Diabetes Devices.

[t]CBT-I: CBT for insomnia.

[u]CBT-SF: CBT for sleep disturbance and fatigue.

Technology

Technology differed across the studies, even those reporting the same intervention program. The most common device used for videoconferencing was an iPad or tablet (12/33, 36%). Computers (6/33, 18%), any available device (3/33, 9%), and smartphones (3/33, 9%) were less common. Skype (9/33, 27%) and Zoom (9/33, 27%) were the most frequently used commercial platforms, with a variety of other commercial and custom-designed platforms reported (Multimedia Appendix 4 [36-68]).

Device source and ownership varied. Researchers provided all devices in some studies (9/33, 27%), whereas participants used their own devices in others (5/33, 15%). Finally, in some studies (4/33, 12%), researchers provided devices only if required. A total of 36% (12/33) of the studies did not report device type, and 45% (15/33) did not provide information on device ownership. Interestingly, 55% (18/33) of the studies stated that lack of access to a device, the internet, or a videoconference platform was a reason for exclusion. In total, 45% (15/33) included only participants who owned or had access to compatible devices and a reliable internet connection, excluding those without. A total of 24% (8/33) of the studies excluded participants who were unable to set up or log into the videoconference system independently or with assistance from family or friends.

In total, 15% (5/33) of the studies reported that the platform or software used was compliant with national or provincial personal and health information regulations, such as Health Insurance Portability and Accountability Act (HIPAA) or the Personal Information Protection and Electronic Documents Act (PIPEDA). A total of 27% (9/33) of the studies [36,37,39,41,45,55,62,64,68] reported strategies to enhance data security (eg, password-protected log-in procedures and encrypted data access), and 9% (3/33) [58,60,62] reported how participants were informed of data security and consented to any potential risk.

XSL•FO
RenderX

In total, 39% (13/33) of the studies confirmed participant device videoconference compatibility and internet strength. Training sessions (10/33, 30%) [37,40,41,45,48,52,55,56,65,68], securing family supports (4/33, 12%) [36,43,55,56], providing manuals (5/33, 15%) [38,45,55,58,63], home visits (2/33, 6%) [41,43], and email instructions (3/33, 9%) [45,58,68] were reported as ways to prepare and educate participants.

Intervention Programs

A total of 33 studies met the inclusion criteria and reported on 25 unique intervention programs. Therefore, program data were aggregated and are presented by program rather than by study in this section.

Programs were delivered by various health care professionals, with psychologists (9/25, 36%) being the most common, followed by multidisciplinary health care teams and nurses (3/25, 12% each) and speech-language pathologists and occupational therapists (2/25, 8% each; Table 2). Participants were people with neurological conditions (10/25, 40%); cancer (7/25, 28%); musculoskeletal conditions (3/25, 12%); or inflammatory bowel disease, obesity, and diabetes (2/25, 8% each). One intervention was delivered to participants with both neurological *and* musculoskeletal conditions. A total of 32% (8/25) of the programs required a companion to join the videoconference to practice communication skills, facilitate family support and relationships, or promote activity engagement [43,45,48,53,55,56,58,59,62,65].

The number of sessions was almost evenly divided between 3 categories. In total, 36% (9/25) of the programs were delivered in ≥10 sessions, and another 36% (9/25) spanned 5 to 9 sessions. A total of 28% (7/25) of the programs had ≤4 sessions (Table 3). Weekly meetings for ≤1 hour were the most common arrangement (9/25, 36%). Only 16% (4/25) of the programs had videoconferencing sessions longer than an hour. A total of 24% (8/33) of the studies reported participant attendance rates of 71% to 100% of sessions, with 12% (4/33) reporting 100% attendance rates. Reported attrition rates ranged between 0% and 35% (10/33, 30% of the studies; protocol papers excluded).

Almost all programs (21/25, 84%) included numerous asynchronous elements in addition to the videoconferencing sessions (Table 3). Active learning elements such as self-monitoring and tracking or between-session practice were commonly reported (10/25, 40% of the programs each), as were more passive learning elements such as printed materials or manuals and access to websites (14/25, 56% of the programs).

On the basis of explicitly stated program goals or objectives, thematic analysis identified 12 distinct program objectives across the 25 intervention programs, with 9 (36%) programs indicating more than one objective (Table 3). Even though programs designed specifically for people with mental health diagnoses and solely exercise-based programs were excluded, the 2 most common objectives were improving physical activity or function and improving mental health, each reported by 28% (7/25) of the programs. The next most frequently reported objective was improving quality of life (5/25, 20%); however, this was rarely the sole program objective. Program-specific health objectives were also found, such as weight loss or improving oral health.

No single theoretical foundation dominated program development. Thematic analysis revealed that the social cognitive theory or cognitive behavioral therapy was the most cited theoretical foundation (6/25, 24%), followed by unspecified behavior change theories (4/25, 16%; Table 3). Studies also cited reviews of the literature or clinical guidelines (5/25, 20%) or previous interventions (5/25, 20%) as program foundations.

No program included content from all TEDSS domains (Multimedia Appendix 5). However, the use of some domains was more common (Figure 2). In total, ≥50% of the programs (18/25, 72%) included content from the *process* domain (eg, finding information, problem-solving, decision-making, and action planning), the *disease control* domain (14/25, 56%; eg, medication and symptom management), the *healthy behaviors* domain (13/25, 52%; eg, diet and exercise), and the *internal* domain (13/25, 52%; eg, staying positive and reducing stress). In contrast, <25% of the programs included content from the other domains (Figure 2).

Figure 2. Intervention program content by program objective. TEDSS: Taxonomy of Every Day Self-Management Strategies.

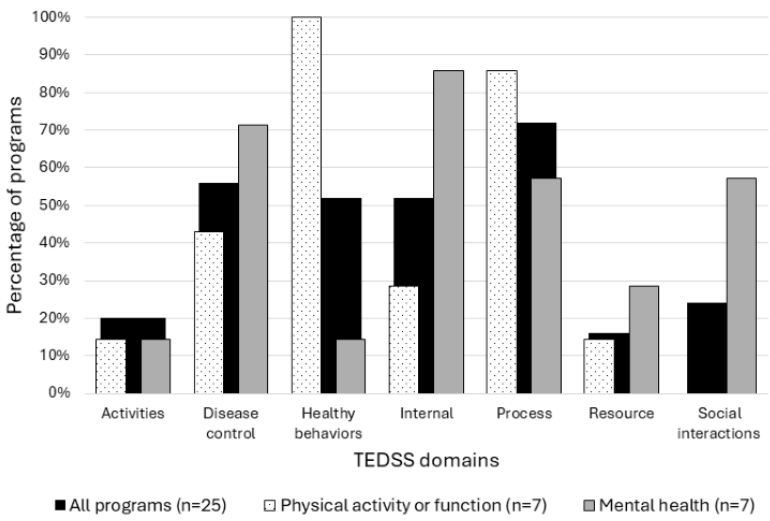

XSL•FO
RenderX

The diversity of the programs prevented a descriptive analysis based on client group or theoretical foundation; however, descriptive analysis based on 2 program objectives (improving mental health and improving physical activity or function) was possible and demonstrated unique patterns in the included content. Figure 2 illustrates markedly different content in 3 specific TEDSS domains. In total, 100% (7/7) of the interventions with the goal of improving physical activity or function included content on diet, exercise, or sleep hygiene (*healthy behavior* domain) compared to only 14% (1/7) of the interventions intended to improve mental health. The opposite trend was observed for the *internal* and *social interaction* domains—interventions with the goal of improving mental health frequently included content from these domains, whereas those focused on improving physical function did not. In fact, no programs with the goal of improving physical activity or function included content from the *social interaction* domain.

Of the 16 BCT Taxonomy v1 strategy groups, 4 (25%; goals and planning, shaping knowledge, feedback and monitoring, and repetition and substitution) were included in >50% of the intervention programs (goals and planning: 19/25, 76%; shaping knowledge: 19/25, 76%; feedback and monitoring: 16/25, 64%; repetition and substitution: 14/25, 56%), and 3 (19%) were not included at all (Figure 3). None of the programs reported using reward and threat, scheduled consequences, or covert learning techniques. As with program content, analysis based on program objective yielded different patterns of behavior change elements. In programs intent on improving physical activity or function, the feedback and monitoring, goals and planning, and social support strategies dominated. However, in programs with the goal of improving mental health, the shaping knowledge, regulation, and identity behavior change strategies were the most common.

Figure 3. Behavior change components by program objective. BCT taxonomy v1: Behavior Change Technique Taxonomy version 1.

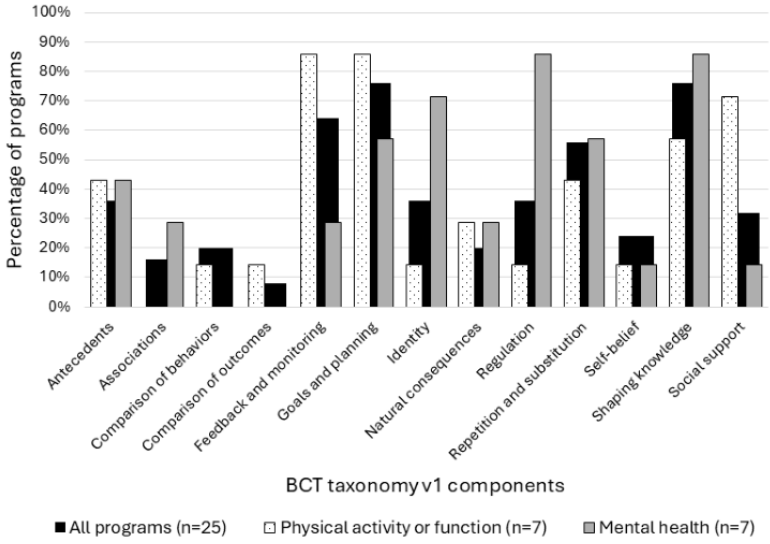

Facilitators of and Barriers to Videoconference Interventions

Text and quantitative results describing facilitators and barriers to using videoconferencing were extracted from 76% (25/33) of the articles (protocol papers were excluded). Health care provider perspectives were reported in only 6% (2/33) of the studies; therefore, only participant perspectives are reported. Five facilitators were categorized and labeled as follows:

1. Feasibility and acceptability to participants (11/25, 44%)
2. Reduced temporal and physical barriers (11/25, 44%)
3. Therapeutic benefit or therapeutic alliance (10/25, 40%)
4. Improved reach to underserviced populations or areas (6/25, 24%)
5. Allowing COVID-19 safety precautions (1/25, 4%)

Although 30% (10/33) of the papers were published in either 2021 or 2022, only 3% (1/33) of the papers [44] identified videoconferencing as a specific benefit during the period of COVID-19 restrictions. Definitions and sample quotes for each facilitator are presented in Table 4. Barriers to delivering interventions using videoconferencing were also identified and are described in Table 4:

1. Reliance on client comfort and technology literacy (8/25, 32%)
2. Limited access to technology and the internet (7/25, 28%)
3. Technical issues or breakdowns (7/25, 28%)
4. Protection of privacy and confidentiality (5/25, 20%)
5. Limited clinical observations and communication (3/25, 12%)

XSL•FO
RenderX

Table 4. Facilitators and barriers of videoconferencing interventions for people with chronic conditions (N=25).

Facilitators or barriers	Interventions, n (%)	Description	Examples
Facilitators			
Feasible and acceptable to participants [37,38,45,46,48,51,53,58,60,63,65]	11 (44)	• Low attrition rate and high completion rate • Positive client feedback—easy to follow, practical, and convenient; easy and quick to reschedule if needed	"The delivery of the TBIconneCT intervention via videoconferencing was feasible and well received by the participants." [58]
Reduced temporal and physical barriers [37,38,40,44,46,51,60,61,63-65]	11 (44)	• Saved time, cost, and energy required traveling to the clinic • Avoidance of barriers in transportation, parking, and accessibility • Particularly beneficial for individuals with greater difficulties traveling or with poorer health conditions	"The ease of completing sessions at home may have reduced treatment access barriers often reported by patients with cancer, such as difficulty with travel and parking." [63]
Therapeutic benefits from receiving the intervention at home or therapeutic alliance building [36,37,40,45,51,58,60,61,65,68]	10 (40)	• Ability to observe and provide environmental recommendations at clients' homes • Direct application and practice of skills in naturalistic settings; no need to generalize skills learned in a clinical or laboratory setting • Meaningful collaborative partnerships	"The first benefit was patient engagement and communication. Video was novel and fun for some patients, more interactive than telephone, and nurses felt they got to know video patients better...Study nurses reported the video was useful for some indications, such as viewing a drawer of medications or observing a patient doing shoulder exercises." [45]
Improved reach to underserviced populations or underserviced areas [38,51,60,62,63,65]	6 (24)	• Improvement of health care service delivery in remote areas for clients with poor health	"Participants connected from diverse locations, including rural and urban areas, spread over large geographic distances. This widespread reach would not have been possible had the intervention taken place in-person." [62]
COVID-19 [44]	1 (4)	• Allowing COVID-19 safety precautions	"In the current COVID-19 pandemic, where social distancing is an essential policy in many countries to decrease the spread of the disease, online interventions have become increasingly common." [44]
Barriers			
Reliance on patients' comfort and technology literacy [38,40,41,45,46,58,63,68].	8 (32)	• Limited confidence, comfort, and skills using technology • Reduced trustworthiness of quality health care services	"In our sample of participants with CI, not unexpectedly, we found that adopting more than one new technology at a time was especially difficult for participants (e.g., learning how to use both the iPad and the video visit software)." [45]
Limited access to compatible technology and the internet [36,40,41,46,55,62,68]	7 (28)	• Home device or software is not compatible with the requirements of videoconferencing • Insufficient internet bandwidth and upload or download speed • High cost of better devices and signal quality	"Participants needed to be familiar with using technology and have access to a computer, limiting the accessibility of this design to all members of the SCI community." [40]
Technical issues [36,45,50,55,58,62,68]	7 (28)	• Delayed or frozen audio or video signal • Technological glitches or system malfunction	"Several expressed frustration with the intermittent loss of connectivity and technical problems related to the videoconferencing during sessions." [68]
Protection of privacy and confidentiality [38,55,58,62,67]	5 (20)	• Unclear privacy protection procedures • Inability to confirm the legitimacy of the health care provider • Inconsistent client health and personal information acts across different jurisdictions	"Ensuring participants are communicating with legitimate personnel, ensuring that the delivery format does not invade privacy." [55]

XSL•FO
RenderX

Facilitators or barriers	Interventions, n (%)	Description	Examples
Limited clinical observations and communication [45,55,62]	3 (12)	• Limited environmental information beyond the scope of the camera	"The therapist could not see the complete context in which the activity was performed...this made it difficult at times for the therapist to analyze performance breakdowns in order to guide the participants most effectively." [55]

Discussion

Principal Findings

Overview

By summarizing what we know about structured, one-to-one videoconference intervention programs for people with chronic conditions, lessons can be gleaned for future development, execution, and feasibility. First, the findings of this review confirm what others have found: videoconference programs are both feasible and acceptable, often providing positive therapeutic benefits. Second, program purpose and objective may provide program developers and interventionists with important insights into program content and behavior change elements. Finally, while technology does overcome some barriers to access, it creates others.

Feasibility and Acceptability

Consistent with previous findings [69,70], this review found videoconference programs to be feasible and acceptable to clients, generating confidence for future program development. As reported by others [71,72], this review also found that spatial and temporal barriers, such as the need to travel to a hospital or clinic, were reduced or removed, and access to care for underserviced populations or areas was increased. Further evidence of participant acceptance was found in the high attendance and low attrition rates. In 67% (4/6) of the studies reporting completion rates, there was 100% completion, suggesting strong client uptake.

Reports on development of therapist-client rapport during videoconference sessions are mixed, with some indicating that providers struggle to build therapeutic alliances [26,73], whereas others indicate that videoconferencing does not hinder the therapist-client rapport or therapeutic alliance [73,74]. The studies in this review found that eye contact, facial expression, voice or tone, and other nonverbal communication were conveyed through the screen with high client satisfaction.

However, feasibility and acceptability were primarily based on qualitative data, client or health care provider impressions, or process evaluations. Future research to quantitatively assess the reduction in spatial and temporal barriers and the fidelity of therapist-client rapport are needed to confirm these findings.

Program Purpose Suggests Content and Behavior Change Elements

Analysis of program content (TEDSS domains) and behavior change strategies (BCT Taxonomy v1 groups) provides direction for health care providers and researchers developing or transferring in-person programs to internet-based format.

A total of 4 TEDSS domains and 4 BCT Taxonomy v1 groups dominated the 25 intervention programs. Over 50% of the programs (18/25, 72%; *disease and control* domain: 14/25, 56%; *healthy behavior* domain: 13/25, 52%; *internal* domain: 13/25, 52%) had content aligned with the same 4 domains (*disease control, healthy behaviors, internal, and process*) of the TEDSS framework, findings that mimic those of others using the TEDSS framework to analyze program content and program outcome measures [11,75]. However, potentially more instructive are the domains that received limited attention. Enhancing clients' role management (ie, *activities* and *social interaction* domains) and strategies to find, access, or manage formal and informal supports and resources (ie, *resources* domain) are also part of managing and living with chronic conditions [76-78]. These content areas are also the very topics that clients identify as most important and least well covered [79]. Qualitative work by Satink et al [80], Audulv et al [81], Audulv [82], and others demonstrates how self-management needs change with time, particularly once clients return home and the focus on medical management diminishes. The paucity of intervention programs addressing clients' role management has been highlighted in other reviews [11,75], as has the limited focus on finding and managing formal and informal resources [34]. On the basis of these findings, future development and adaptation of telehealth care programs could consider broadening the scope of program content and the alignment between client needs, particularly when clients are living in the community. Research to interrogate the relative importance of different content areas based on context, disease trajectory, and time since diagnosis would help identify active ingredients needed to tailor interventions.

BCT Taxonomy v1 techniques used in videoconferencing interventions appear to emulate asynchronous e-interventions for individuals with diabetes and programs for health promotion delivered using the internet or mobile phones [83-85]. Many techniques in the goals and planning and feedback and monitoring clusters, also found in this review, have had significant treatment effects in diabetes management [83-85]. Shaping knowledge and repetition and substitution (eg, practicing and rehearsing learned skills) also appeared to be common elements in the studies in this review, likely due to the inclusion criteria, which stressed active participation. Regardless, the results indicate that a wide array of behavior change techniques can be incorporated into telehealth care programs.

This review specifically excluded intervention programs based solely on repetition or intended for clients with mental health diagnoses. Despite this, the 2 most common program objectives were improving physical activity or function and improving

XSL•FO
RenderX

mental health. Comparison of TEDSS domains and BCT Taxonomy v1 elements provides valuable insights for both in-person and telehealth care program development and delivery. While the analysis was descriptive and limited to these 2 program objectives (7/33, 21% of the studies each), marked differences were found, suggesting that program objectives have, at least up to now, driven content and behavior change strategies, 2 things often considered active ingredients in chronic disease management and intervention programs. These stark differences may reflect theoretical or disciplinary differences or perhaps the siloed delivery of health care. For example, differences in content related to the *healthy behaviors*, *internal, and social interaction* domains may at first seem intuitive. Greater emphasis on the *healthy behaviors* domain in programs focused on improving physical activity or function versus a focus on the *internal* and *social interaction* intuitively understandable. However, significant research links positive mental health outcomes to healthy eating, sleep hygiene, and exercise [86,87]. Similarly, improving physical activity or function in populations living with chronic conditions is often dependent on family and friends for support and motivation [88,89]. Similar contradictions seem to appear in the chosen BCT Taxonomy v1 strategies. Why feedback and monitoring strategies are more successful in building physical activity or function or why identity and regulation strategies are more important in programs that focus on mental health could be questioned. Research to disentangle these inconsistencies and questions is needed.

Regardless of the contradictions and questions, the finding that program content and techniques and, therefore, active ingredients vary by program objective has important implications for program development, training, and effectiveness studies. For example, the development of new intervention programs can be guided by knowledge of existing active ingredients. Moreover, using the TEDSS and BCT Taxonomy v1 in the planning stages may generate innovative programs. It is also important for future systematic reviews to avoid grouping programs with different objectives to avoid findings of average effectiveness across programs with different active ingredients.

Collectively, our findings demonstrate the value of using the well-defined TEDSS and BCT Taxonomy v1 frameworks to unify how interventions are reported and compared. Using common frameworks with defined terminology can potentially address the calls to identify and isolate content or active ingredients within complex interventions and assist providers in tailoring self-management interventions to client needs.

Challenges Delivering Interventions Using Videoconferencing

Despite the feasibility and acceptability of videoconference interventions, barriers related to technology literacy, compatible devices, and quality internet access were found. Evidence of barriers was 2-fold. First, many studies (18/33, 55%) excluded participants without a stable internet connection, compatible hardware, or family or social support at home. While it is unknown how many participants were excluded for these reasons, it is likely that they were marginalized populations or lived in rural areas where Wi-Fi connection is often less stable. This technological gap must be addressed to reduce health inequity using any telehealth care approach. Second, clients' comfort level, confidence, and skills to effectively operate the necessary technology were the most common barriers reported despite efforts to provide initial training and the exclusion of participants without access. Careful planning, communication, and training are imperative to enhance comfort and confidence for service providers and clients [90]. The work by Vassilev et al [71] underscores this, finding "evidence that clients' capability in technology use increased their propensity to benefit from interventions." On the basis of the findings of this review, communicating with clients about confidence and preference of delivery format should be prioritized before the initial session.

Consistent with existing literature, technology breakdowns were a common cause of user frustration [91]. While 39% (13/33) of the studies verified the connectivity before intervention sessions, there were breakdowns or poor audio or video quality. Appointments had to be canceled or rescheduled through email or telephone. Although causes were unreported, limited bandwidth, software malfunction, and old add-on devices (eg, microphone or camera) are potential explanations. These findings recommend intentional preparation and contingency planning, such as a participant manual for troubleshooting common technology issues. Some highly qualified health care providers may also lack the technical skills or confidence to manage videoconference delivery. Adequate provider technology expertise should be assessed before service delivery via telehealth care formats [92]. Adding a technology support team member could meet a portion of this need and potentially enhance the therapeutic experience [93].

Limitations and Future Research

Although this review covered a 20-year span up to January 2023, many interventions were planned or executed before the COVID-19 pandemic and may not capture the rapid development of videoconferencing services and policies after 2020. Technology for telehealth care and videoconferencing (eg, Zoom and Microsoft Teams) has dramatically advanced with the addition of products, video technology, and health care security (HIPAA compliance). Familiarity and comfort with technologies has also increased dramatically. Upcoming articles may reveal additional recommendations. The lack of information reported on security and confidentiality issues during videoconferencing suggests the need for a dedicated and in-depth study across countries and jurisdictions.

The findings of this review are limited to information reported in manuscripts and publicly available to researchers and providers. Authors were not contacted to clarify or confirm the content or active ingredients in the interventions. Therefore, reporting of content and behavior change strategies may be incomplete. More structured and uniform reporting is needed to overcome this gap. While scoping review methodology does not include the assessment of research quality or outcome evaluation, uniformity of reporting would benefit future systematic reviews or meta-analyses, helping identify links between active ingredients and outcomes. Finally, health care providers' perspectives on videoconferencing could not be

XSL•FO
RenderX

reported due to the small number of papers or studies reporting these data. The inherent provider-client collaboration in these interventions [10] warrants the need to also understand the challenges experienced by health care providers.

Conclusions and Clinical Implications

Using videoconferencing to deliver one-to-one synchronous interventions is feasible and acceptable to clients. Health care providers can build positive therapeutic relationships via videoconference and simultaneously reduce temporal and physical barriers associated with access to care and improving reach to underserviced populations and areas. Despite this, barriers and concerns exist. Client confidence and comfort using technology, limited access to compatible and reliable hardware and internet connection, and technology breakdowns are common, highlighting the importance of technical support for providers and clients.

Structured, one-to-one videoconference interventions that actively engage people in the management of their chronic conditions remain primarily focused on medical and, to some extent, emotional management. Focused content to support integration into roles and everyday life, prioritize positive social interactions, and improve access to resources constitutes areas important to clients to strengthen interventions. The value of this content regardless of the program goal should be considered when developing future interventions. Similarly, additional behavior change techniques have potential to increase the effectiveness of interventions. With telehealth care practices expanding, further research is needed to point the way to best practice.

Moving in-person self-management interventions to internet-based formats must be intentional. While technology can reduce barriers, access to care based on device ownership and internet connectivity can be a delivery challenge. Health care teams may need to be expanded to include technical and technology support for providers and clients. Ensuring compliance with local privacy laws and protection of personal health information is required. Prepared contingency plans and a troubleshooting guide are strongly recommended in case of technology breakdowns. Existing program objectives appear to drive content and behavior change strategies, potentially limiting value to clients. Future programs could consider greater emphasis on enabling clients to find and organize formal and informal supports or manage their social interaction and daily activities, which are self-management domains that clients identify as important and poorly covered.

Acknowledgments

The authors would like to thank Kaitlin Sibald, research occupational therapist, for her contributions early in the project, including input into conceptualizing the research question and conducting early database searches. Shelley McKibbon, a medical librarian at Dalhousie University, was instrumental in refining the search strategies. Without this critical step, this review would not have been possible. ML was affiliated with Dalhousie University at the time this review was conducted and is currently affiliated with Think Self-Management Inc. JK was affiliated with the University of Illinois Chicago at the time this review was conducted and is currently affiliated with the Shirley Ryan AbilityLab. This work was partially supported through a Patient-Centered Outcomes Research Institute Project Program Award (MS-1610-37015). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute or its board of governors or the Methodology Committee.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to the complexity of manual coding and the need for related training but are available from the corresponding author on reasonable request.

Authors' Contributions

YTC: methodology, software, formal analysis, investigation, visualization, writing—original draft, writing—review and editing, and project administration.
ML: data curation, formal analysis, investigation, project administration, software, visualization, and writing—review and editing.
TVD: conceptualization, formal analysis, investigation, and writing—review and editing.
JK: conceptualization, formal analysis, investigation, and writing—review and editing.
YW: conceptualization, formal analysis, investigation, and writing—review and editing.
KP: writing—review and editing and funding acquisition.
MP: investigation, writing—review and editing, and funding acquisition.
TLP: conceptualization, methodology, software, formal analysis, visualization, writing—review and editing, supervision, project administration, and funding acquisition.

Conflicts of Interest

ML is cofounder and owner of Think Self-Management Inc, an incorporated company that disseminates evidence-based self-management interventions. TLP was an original cofounder with no financial interest. This review was planned and mostly

XSL•FO
RenderX

completed before the incorporation of this company, and no publications regarding interventions or programs distributed by Think Self-Management met the inclusion criteria for this review. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 156 KB - jmir_v27i1e58543_app1.pdf]

Multimedia Appendix 2
Search strategies and number of articles found in each database.
[PDF File (Adobe PDF File), 119 KB - jmir_v27i1e58543_app2.pdf]

Multimedia Appendix 3
Reviewer qualifications and involvement in the review process.
[PDF File (Adobe PDF File), 64 KB - jmir_v27i1e58543_app3.pdf]

Multimedia Appendix 4
Videoconferencing hardware and software used in intervention programs for people with chronic conditions.
[PDF File (Adobe PDF File), 92 KB - jmir_v27i1e58543_app4.pdf]

Multimedia Appendix 5
Intervention program content by Taxonomy of Every Day Self-Management Strategies and Behavior Change Technique Taxonomy version 1 categories.
[PDF File (Adobe PDF File), 155 KB - jmir_v27i1e58543_app5.pdf]

References

1. Gelburd R. Telehealth claim lines rise 2980% in one-year period through September 2020. AJMC. 2020 Dec 1. URL: https://www.ajmc.com/view/telehealth-claim-lines-rise-2980-in-one-year-period-through-september-2020 [accessed 2024-12-12]
2. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. PLoS One 2019 Aug 30;14(8):e0221848 [FREE Full text] [doi: 10.1371/journal.pone.0221848] [Medline: 31469865]
3. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Clinical applications of videoconferencing: a scoping review of the literature for the period 2002–2012. J Telemed Telecare 2014 Nov 16;20(7):377-383 [FREE Full text] [doi: 10.1177/1357633x14552385]
4. About chronic diseases. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/chronic-disease/about/index.html [accessed 2023-08-03]
5. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. Qual Sociol 1985;8(3):224-247 [FREE Full text] [doi: 10.1007/bf00989485]
6. Van de Velde D, De Zutter F, Satink T, Costa U, Janquart S, Senn D, et al. Delineating the concept of self-management in chronic conditions: a concept analysis. BMJ Open 2019 Jul 16;9(7):e027775 [FREE Full text] [doi: 10.1136/bmjopen-2018-027775] [Medline: 31315862]
7. Schulman-Green D, Jaser S, Martin F, Alonzo A, Grey M, McCorkle R, et al. Processes of self-management in chronic illness. J Nurs Scholarsh 2012 Jun;44(2):136-144 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01444.x] [Medline: 22551013]
8. Grey M, Knafl K, McCorkle R. A framework for the study of self- and family management of chronic conditions. Nurs Outlook 2006;54(5):278-286 [FREE Full text] [doi: 10.1016/j.outlook.2006.06.004] [Medline: 17027605]
9. Barr VJ, Robinson S, Marin-Link B, Underhill L, Dotts A, Ravensdale D, et al. The expanded Chronic Care Model: an integration of concepts and strategies from population health promotion and the Chronic Care Model. Hosp Q 2003 Nov 15;7(1):73-82. [doi: 10.12927/hcq.2003.16763] [Medline: 14674182]
10. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001 Nov;20(6):64-78. [doi: 10.1377/hlthaff.20.6.64] [Medline: 11816692]
11. Warner G, Packer TL, Kervin E, Sibbald K, Audulv Å. A systematic review examining whether community-based self-management programs for older adults with chronic conditions actively engage participants and teach them patient-oriented self-management strategies. Patient Educ Couns 2019 Dec;102(12):2162-2182. [doi: 10.1016/j.pec.2019.07.002] [Medline: 31301922]
12. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]

XSL•FO
RenderX

13. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
15. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
16. Colquhoun H. Current best practices for the conduct of scoping reviews. EQUATOR Network. 2016 May 12. URL: https://www.equator-network.org/wp-content/uploads/2016/06/Gerstein-Library-scoping-reviews_May-12.pdf [accessed 2024-12-12]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Chen YT, Lehman M, Van Denend T, Kish J, Wu Y, Preissner KL, et al. Features of structured, one-to-one videoconference interventions that actively engage people in the management of their chronic conditions: scoping review. Open Science Framework. 2024 Dec 24. URL: https://osf.io/7ks9u/?view_only=29d560b832214bd9be02af0cc8397933 [accessed 2024-12-24]
19. Audulv Å, Ghahari S, Kephart G, Warner G, Packer TL. The Taxonomy of Everyday Self-management Strategies (TEDSS): a framework derived from the literature and refined using empirical data. Patient Educ Couns 2019 Feb;102(2):367-375 [FREE Full text] [doi: 10.1016/j.pec.2018.08.034] [Medline: 30197252]
20. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
21. Dohoo IR, Martin SW, Stryhn H. Methods in Epidemiologic Research. Eagle, ID: VER Incorporated; 2012.
22. Santesteban-Echarri O, Piskulic D, Nyman RK, Addington J. Telehealth interventions for schizophrenia-spectrum disorders and clinical high-risk for psychosis individuals: a scoping review. J Telemed Telecare 2018 Aug 22;26(1-2):14-20. [doi: 10.1177/1357633x18794100]
23. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
24. Berryhill MB, Halli-Tierney A, Culmer N, Williams N, Betancourt A, King M, et al. Videoconferencing psychological therapy and anxiety: a systematic review. Fam Pract 2019 Jan 25;36(1):53-63. [doi: 10.1093/fampra/cmy072] [Medline: 30188992]
25. Muir SD, de Boer K, Nedeljkovic M, Meyer D. Barriers and facilitators of videoconferencing psychotherapy implementation in veteran mental health care environments: a systematic review. BMC Health Serv Res 2020 Nov 01;20(1):999 [FREE Full text] [doi: 10.1186/s12913-020-05858-3] [Medline: 33131495]
26. Thomas N, McDonald C, de Boer K, Brand RM, Nedeljkovic M, Seabrook L. Review of the current empirical literature on using videoconferencing to deliver individual psychotherapies to adults with mental health problems. Psychol Psychother 2021 Sep 23;94(3):854-883 [FREE Full text] [doi: 10.1111/papt.12332] [Medline: 33620133]
27. Horsley S, Schock G, Grona SL, Montieth K, Mowat B, Stasiuk K, et al. Use of real-time videoconferencing to deliver physical therapy services: a scoping review of published and emerging evidence. J Telemed Telecare 2020 Dec;26(10):581-589. [doi: 10.1177/1357633X19854647] [Medline: 31213166]
28. Grona SL, Bath B, Busch A, Rotter T, Trask C, Harrison E. Use of videoconferencing for physical therapy in people with musculoskeletal conditions: a systematic review. J Telemed Telecare 2017 Apr 12;24(5):341-355. [doi: 10.1177/1357633x17700781]
29. Velayati F, Ayatollahi H, Hemmat M. A systematic review of the effectiveness of telerehabilitation interventions for therapeutic purposes in the elderly. Methods Inf Med 2020 May 06;59(2-03):104-109. [doi: 10.1055/s-0040-1713398] [Medline: 32629502]
30. Audulv Å, Packer T, Hutchinson S, Roger KS, Kephart G. Coping, adapting or self-managing - what is the difference? A concept review based on the neurological literature. J Adv Nurs 2016 Nov 27;72(11):2629-2643. [doi: 10.1111/jan.13037] [Medline: 27272388]
31. Audulv Å, Hutchinson S, Warner G, Kephart G, Versnel J, Packer TL. Managing everyday life: self-management strategies people use to live well with neurological conditions. Patient Educ Couns 2021 Feb;104(2):413-421 [FREE Full text] [doi: 10.1016/j.pec.2020.07.025] [Medline: 32819756]
32. Cadel L, Everall AC, Packer TL, Hitzig SL, Patel T, Lofters AK, et al. Exploring the perspectives on medication self-management among persons with spinal cord injury/dysfunction and providers. Res Social Adm Pharm 2020 Dec;16(12):1775-1784. [doi: 10.1016/j.sapharm.2020.01.014] [Medline: 32057689]

XSL•FO
RenderX

33. Wright H, Turner A, Ennis S, Percy C, Loftus G, Clyne W, et al. Digital peer-supported self-management intervention codesigned by people with long COVID: mixed methods proof-of-concept study. JMIR Form Res 2022 Oct 14;6(10):e41410 [FREE Full text] [doi: 10.2196/41410] [Medline: 36166651]
34. Alizadeh N, Packer T, Chen YT, Alnasery Y. What we know about fatigue self-management programs for people living with chronic conditions: a scoping review. Patient Educ Couns 2023 Sep;114:107866 [FREE Full text] [doi: 10.1016/j.pec.2023.107866] [Medline: 37364380]
35. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. Br J Health Psychol 2015 Feb;20(1):130-150. [doi: 10.1111/bjhp.12102] [Medline: 24815766]
36. Alencar MK, Johnson K, Mullur R, Gray V, Gutierrez E, Korosteleva O. The efficacy of a telemedicine-based weight loss program with video conference health coaching support. J Telemed Telecare 2017 Dec 03;25(3):151-157. [doi: 10.1177/1357633x17745471]
37. Alencar M, Johnson K, Gray V, Mullur R, Gutierrez E, Dionico P. Telehealth-based health coaching increases m-health device adherence and rate of weight loss in obese participants. Telemed J E Health 2020 Mar;26(3):365-368 [FREE Full text] [doi: 10.1089/tmj.2019.0017] [Medline: 30994410]
38. Aubin S, Rosberger Z, Hafez N, Noory MR, Perez S, Lehmann S, et al. Cancer!? I don't have time for that: impact of a psychosocial intervention for young adults with cancer. J Adolesc Young Adult Oncol 2019 Apr 01;8(2):172-189. [doi: 10.1089/jayao.2017.0101] [Medline: 30484730]
39. Brunet J, Price J, Srikanthan A, Gillison F, Standage M, Taljaard M, et al. The physicAl aCtivity Counselling for young adult cancEr SurvivorS (ACCESS) trial: a protocol for a parallel, two-arm pilot randomized controlled trial. PLoS One 2022;17(12):e0273045 [FREE Full text] [doi: 10.1371/journal.pone.0273045] [Medline: 36584027]
40. Chemtob K, Rocchi M, Arbour-Nicitopoulos K, Kairy D, Fillion B, Sweet SN. Using tele-health to enhance motivation, leisure time physical activity, and quality of life in adults with spinal cord injury: a self-determination theory-based pilot randomized control trial. Psychol Sport Exerc 2019 Jul;43:243-252 [FREE Full text] [doi: 10.1016/j.psychsport.2019.03.008]
41. Cruice M, Woolf C, Caute A, Monnelly K, Wilson S, Marshall J. Preliminary outcomes from a pilot study of personalised online supported conversation for participation intervention for people with Aphasia. Aphasiology 2020 Jul 25;35(10):1293-1317 [FREE Full text] [doi: 10.1080/02687038.2020.1795076]
42. Garland SN, Savard J, Dalton K, Walsh NA, Seal M, Rash J, et al. Rationale and protocol for a randomized waitlist controlled trial of videoconference delivered cognitive behaviour therapy for insomnia (CBT-I) to improve perceived cognitive impairment (PCI) among cancer survivors. Contemp Clin Trials 2021 Apr;103:106322. [doi: 10.1016/j.cct.2021.106322] [Medline: 33588077]
43. Gilboa Y, Maeir T, Karni S, Eisenberg ME, Liebergall M, Schwartz I, et al. Effectiveness of a tele-rehabilitation intervention to improve performance and reduce morbidity for people post hip fracture - study protocol for a randomized controlled trial. BMC Geriatr 2019 May 20;19(1):135 [FREE Full text] [doi: 10.1186/s12877-019-1141-z] [Medline: 31109289]
44. Goren G, Schwartz D, Friger M, Banai H, Sergienko R, Regev S, et al. Randomized controlled trial of cognitive-behavioral and mindfulness-based stress reduction on the quality of life of patients with Crohn disease. Inflamm Bowel Dis 2022 Mar 02;28(3):393-408. [doi: 10.1093/ibd/izab083] [Medline: 33847758]
45. Hastings SN, Mahanna EP, Berkowitz TS, Smith VA, Choate AL, Hughes JM, et al. Video-enhanced care management for medically complex older adults with cognitive impairment. J Am Geriatr Soc 2021 Jan;69(1):77-84 [FREE Full text] [doi: 10.1111/jgs.16819] [Medline: 32966603]
46. Kelleher SA, Winger JG, Dorfman CS, Ingle KK, Moskovich AA, Abernethy AP, et al. A behavioral cancer pain intervention: a randomized noninferiority trial comparing in-person with videoconference delivery. Psychooncology 2019 Aug 19;28(8):1671-1678 [FREE Full text] [doi: 10.1002/pon.5141] [Medline: 31162756]
47. Kline PW, Melanson EL, Sullivan WJ, Blatchford PJ, Miller MJ, Stevens-Lapsley JE, et al. Improving physical activity through adjunct telerehabilitation following total knee arthroplasty: randomized controlled trial protocol. Phys Ther 2019 Jan 01;99(1):37-45 [FREE Full text] [doi: 10.1093/ptj/pzy119] [Medline: 30329126]
48. Lafaro KJ, Raz DJ, Kim JY, Hite S, Ruel N, Varatkar G, et al. Pilot study of a telehealth perioperative physical activity intervention for older adults with cancer and their caregivers. Support Care Cancer 2020 Aug 16;28(8):3867-3876 [FREE Full text] [doi: 10.1007/s00520-019-05230-0] [Medline: 31845007]
49. Lavelle J, Storan D, Eswara Murthy V, De Dominicis N, Mulcahy HE, McHugh L. Brief and telehealth acceptance and commitment therapy (ACT) interventions for stress in inflammatory bowel disease (IBD): a series of single case experimental design (SCED) studies. J Clin Med 2022 May 13;11(10):2757 [FREE Full text] [doi: 10.3390/jcm11102757] [Medline: 35628884]
50. Lawson DW, Stolwyk RJ, Ponsford JL, McKenzie DP, Downing MG, Wong D. Telehealth delivery of memory rehabilitation following stroke. J Int Neuropsychol Soc 2020 Jan 27;26(1):58-71. [doi: 10.1017/s1355617719000651]
51. Lawson DW, Stolwyk RJ, Ponsford JL, Baker KS, Tran J, Wong D. Acceptability of telehealth in post-stroke memory rehabilitation: a qualitative analysis. Neuropsychol Rehabil 2022 Jan 17;32(1):1-21. [doi: 10.1080/09602011.2020.1792318] [Medline: 32677539]

XSL•FO
RenderX

52. Lynch CP, Williams JS, J Ruggiero K, G Knapp R, Egede LE. Tablet-Aided BehavioraL intervention EffecT on Self-management skills (TABLETS) for diabetes. Trials 2016 Mar 22;17(1):157 [FREE Full text] [doi: 10.1186/s13063-016-1243-2] [Medline: 27005766]
53. Milbury K, Li Y, Durrani S, Liao Z, Tsao AS, Carmack C, et al. A mindfulness-based intervention as a supportive care strategy for patients with metastatic non-small cell lung cancer and their spouses: results of a three-arm pilot randomized controlled trial. Oncologist 2020 Nov 17;25(11):e1794-e1802 [FREE Full text] [doi: 10.1634/theoncologist.2020-0125] [Medline: 32621630]
54. Miller MJ, Stevens-Lapsley J, Fields TT, Coons D, Bray-Hall S, Sullivan W, et al. Physical activity behavior change for older veterans after dysvascular amputation. Contemp Clin Trials 2017 Apr;55:10-15 [FREE Full text] [doi: 10.1016/j.cct.2017.01.008] [Medline: 28153768]
55. Ng EM, Polatajko HJ, Marziali E, Hunt A, Dawson DR. Telerehabilitation for addressing executive dysfunction after traumatic brain injury. Brain Inj 2013 Mar 08;27(5):548-564. [doi: 10.3109/02699052.2013.766927] [Medline: 23472964]
56. Ownsworth T, Chambers S, Aitken JF, Foote M, Pinkham MB, Gordon LG, et al. Evaluation of a telehealth psychological support intervention for people with primary brain tumour and their family members: study protocol for a randomised controlled trial. Eur J Cancer Care (Engl) 2019 Jul 10;28(4):e13132. [doi: 10.1111/ecc.13132] [Medline: 31293057]
57. Pfammatter AF, Battalio SL, Olvera C, DeZelar M, Moore D, Scanlan L, et al. The EVO study protocol for a randomized controlled evaluation trial of an optimized weight management intervention. Contemp Clin Trials 2022 May;116:106750 [FREE Full text] [doi: 10.1016/j.cct.2022.106750] [Medline: 35378301]
58. Rietdijk R, Power E, Brunner M, Togher L. A single case experimental design study on improving social communication skills after traumatic brain injury using communication partner telehealth training. Brain Inj 2019 Oct 16;33(1):94-104. [doi: 10.1080/02699052.2018.1531313] [Medline: 30325220]
59. Rietdijk R, Power E, Attard M, Heard R, Togher L. Improved conversation outcomes after social communication skills training for people with traumatic brain injury and their communication partners: a clinical trial investigating in-person and telehealth delivery. J Speech Lang Hear Res 2020 Feb 26;63(2):615-632. [doi: 10.1044/2019_jslhr-19-00076]
60. Somers TJ, Abernethy AP, Edmond SN, Kelleher SA, Wren AA, Samsa GP, et al. A pilot study of a mobile health pain coping skills training protocol for patients with persistent cancer pain. J Pain Symptom Manage 2015 Oct;50(4):553-558 [FREE Full text] [doi: 10.1016/j.jpainsymman.2015.04.013] [Medline: 26025279]
61. Tanenbaum ML, Ngo J, Hanes SJ, Basina M, Buckingham BA, Hessler D, et al. ONBOARD: a feasibility study of a telehealth-based continuous glucose monitoring adoption intervention for adults with type 1 diabetes. Diabetes Technol Ther 2021 Dec 01;23(12):818-827 [FREE Full text] [doi: 10.1089/dia.2021.0198] [Medline: 34270351]
62. Vellani S, Puts M, Iaboni A, McGilton KS. Acceptability of the voice your values, an advance care planning intervention in persons living with mild dementia using videoconferencing technology. PLoS One 2022 Apr 26;17(4):e0266826 [FREE Full text] [doi: 10.1371/journal.pone.0266826] [Medline: 35472150]
63. Winger JG, Nunez C, Kelleher SA, Ingle KK, Gandhi V, Keefe FJ, et al. Predictors of intervention session completion in a randomized clinical trial of a behavioral cancer pain intervention. J Pain Symptom Manage 2020 Jun;59(6):1268-1277 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.01.020] [Medline: 32045676]
64. Winger JG, Ramos K, Kelleher SA, Somers TJ, Steinhauser KE, Porter LS, et al. Meaning-centered pain coping skills training: a pilot feasibility trial of a psychosocial pain management intervention for patients with advanced cancer. J Palliat Med 2022 Jan 01;25(1):60-69 [FREE Full text] [doi: 10.1089/jpm.2021.0081] [Medline: 34388037]
65. Wood S, Khong CM, Dirlikov B, Shem K. Nutrition counseling and monitoring via tele-nutrition for healthy diet for people with spinal cord injury: a case series analyses. J Spinal Cord Med 2022 Jul 19;45(4):547-555 [FREE Full text] [doi: 10.1080/10790268.2021.1871824] [Medline: 33606588]
66. Ymer L, McKay A, Wong D, Frencham K, Grima N, Tran J, et al. Cognitive behavioral therapy for sleep disturbance and fatigue following acquired brain injury: predictors of treatment response. J Head Trauma Rehabil 2022;37(3):E220-E230. [doi: 10.1097/HTR.0000000000000705] [Medline: 34320552]
67. Beit Yosef A, Jacobs JM, Shames J, Schwartz I, Gilboa Y. A performance-based teleintervention for adults in the chronic stage after acquired brain injury: an exploratory pilot randomized controlled crossover study. Brain Sci 2022 Feb 03;12(2):213 [FREE Full text] [doi: 10.3390/brainsci12020213] [Medline: 35203976]
68. Yuen HK. Effect of a home telecare program on oral health among adults with tetraplegia: a pilot study. Spinal Cord 2013 Jun 15;51(6):477-481 [FREE Full text] [doi: 10.1038/sc.2012.176] [Medline: 23318557]
69. Armfield NR, Bradford M, Bradford NK. The clinical use of Skype--for which patients, with which problems and in which settings? A snapshot review of the literature. Int J Med Inform 2015 Oct;84(10):737-742. [doi: 10.1016/j.ijmedinf.2015.06.006] [Medline: 26183642]
70. Mallow JA, Petitte T, Narsavage G, Barnes E, Theeke E, Mallow BK, et al. The use of video conferencing for persons with chronic conditions: a systematic review. Ehealth Telecommun Syst Netw 2016 Jun;5(2):39-56 [FREE Full text] [doi: 10.4236/etsn.2016.52005] [Medline: 29104830]
71. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015 Apr 24;10(1):59 [FREE Full text] [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]

XSL•FO
RenderX

72. Center for Substance Abuse Treatment (US). Using Technology-Based Therapeutic Tools in Behavioral Health Services. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2015.
73. Simpson S, Bell L, Knox J, Mitchell D. Therapy via videoconferencing: a route to client empowerment? Clin Psychol Psychother 2015;12(2):156-165 [FREE Full text] [doi: 10.1002/cpp.436]
74. Simpson SG, Reid CL. Therapeutic alliance in videoconferencing psychotherapy: a review. Aust J Rural Health 2014 Dec 10;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]
75. Packer TL, Fracini A, Audulv Å, Alizadeh N, van Gaal BG, Warner G, et al. What we know about the purpose, theoretical foundation, scope and dimensionality of existing self-management measurement tools: a scoping review. Patient Educ Couns 2018 Apr;101(4):579-595. [doi: 10.1016/j.pec.2017.10.014] [Medline: 29239734]
76. Glasgow RE, Huebschmann AG, Krist AH, Degruy FV. An adaptive, contextual, technology-aided support (ACTS) system for chronic illness self-management. Milbank Q 2019 Sep 19;97(3):669-691 [FREE Full text] [doi: 10.1111/1468-0009.12412] [Medline: 31424137]
77. Koetsenruijter J, van Eikelenboom N, van Lieshout J, Vassilev I, Lionis C, Todorova E, et al. Social support and self-management capabilities in diabetes patients: an international observational study. Patient Educ Couns 2016 Apr;99(4):638-643 [FREE Full text] [doi: 10.1016/j.pec.2015.10.029] [Medline: 26549171]
78. Hyman I, Shakya Y, Jembere N, Gucciardi E, Vissandjée B. Provider- and patient-related determinants of diabetes self-management among recent immigrants: implications for systemic change. Can Fam Physician 2017 Feb;63(2):e137-e144. [Medline: 28209706]
79. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133 [FREE Full text] [Medline: 25642490]
80. Satink T, Josephsson S, Zajec J, Cup EH, de Swart BJ, Nijhuis-van der Sanden MW. Self-management develops through doing of everyday activities-a longitudinal qualitative study of stroke survivors during two years post-stroke. BMC Neurol 2016 Nov 15;16(1):221 [FREE Full text] [doi: 10.1186/s12883-016-0739-4] [Medline: 27846820]
81. Audulv A, Asplund K, Norbergh KG. The integration of chronic illness self-management. Qual Health Res 2012 Mar 13;22(3):332-345. [doi: 10.1177/1049732311430497] [Medline: 22167155]
82. Audulv Å. The over time development of chronic illness self-management patterns: a longitudinal qualitative study. BMC Public Health 2013 May 07;13:452 [FREE Full text] [doi: 10.1186/1471-2458-13-452] [Medline: 23647658]
83. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020 Mar 24;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
84. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
85. El-Gayar O, Ofori M, Nawar N. On the efficacy of behavior change techniques in mHealth for self-management of diabetes: a meta-analysis. J Biomed Inform 2021 Jul;119:103839 [FREE Full text] [doi: 10.1016/j.jbi.2021.103839] [Medline: 34139330]
86. Mikkelsen K, Stojanovska L, Polenakovic M, Bosevski M, Apostolopoulos V. Exercise and mental health. Maturitas 2017 Dec;106:48-56 [FREE Full text] [doi: 10.1016/j.maturitas.2017.09.003] [Medline: 29150166]
87. Dunn AL, Jewell JS. The effect of exercise on mental health. Curr Sports Med Rep 2010;9(4):202-207. [doi: 10.1249/jsr.0b013e3181e7d9af]
88. Uchino BN. Social support and health: a review of physiological processes potentially underlying links to disease outcomes. J Behav Med 2006 Aug;29(4):377-387. [doi: 10.1007/s10865-006-9056-5] [Medline: 16758315]
89. Cohen S. Social supports and physical health: symptoms, health behaviors, and infectious disease. In: Cummings EM, Greene AL, Karraker KH, editors. Life-span Developmental Psychology: Perspectives on Stress and Coping. New York, NY: Psychology Press; 1991.
90. Brennan DM, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. Stud Health Technol Inform 2009;145:231-248. [Medline: 19592797]
91. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Technical aspects of clinical videoconferencing: a large scale review of the literature. J Telemed Telecare 2015 Feb 19;21(3):160-166. [doi: 10.1177/1357633x15571999]
92. Brennan D, Tindall L, Theodoros D, Brown J, Campbell M, Christiana D, et al. A blueprint for telerehabilitation guidelines. Int J Telerehabil 2010;2(2):31-34 [FREE Full text] [doi: 10.5195/ijt.2010.6063] [Medline: 25945175]
93. Muir SD, de Boer K, Thomas N, Seabrook E, Nedeljkovic M, Meyer D. Videoconferencing psychotherapy in the public sector: synthesis and model for implementation. JMIR Ment Health 2020 Jan 21;7(1):e14996 [FREE Full text] [doi: 10.2196/14996] [Medline: 31961334]

Abbreviations

BCT Taxonomy v1: Behavior Change Technique Taxonomy version 1
HIPAA: Health Insurance Portability and Accountability Act

XSL•FO
RenderX

PIPEDA: Personal Information Protection and Electronic Documents Act
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
TEDSS: Taxonomy of Every Day Self-Management Strategies

Edited by X Ma; submitted 19.03.24; peer-reviewed by J Lau, X Huang, SC Cheng; comments to author 16.08.24; revised version received 30.08.24; accepted 22.11.24; published 26.02.25.

Please cite as:
Chen YT, Lehman M, Van Denend T, Kish J, Wu Y, Preissner K, Plow M, Packer TL
Features of Structured, One-to-One Videoconference Interventions That Actively Engage People in the Management of Their Chronic Conditions: Scoping Review
J Med Internet Res 2025;27:e58543
URL: https://www.jmir.org/2025/1/e58543
doi:10.2196/58543
PMID:

©Yu-Ting Chen, Michelle Lehman, Toni Van Denend, Jacqueline Kish, Yue Wu, Katharine Preissner, Matthew Plow, Tanya L Packer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Using Virtual Reality to Improve Outcomes Related to Quality of Life Among Older Adults With Serious Illnesses: Systematic Review of Randomized Controlled Trials

Bhagvat Maheta[1,2], BS; Alexandra Kraft[3], BS; Nickolas Interrante[1], BS; Soraya Fereydooni[4], BS; Jeremy Bailenson[5], PhD; Brian Beams[5], MFA, MBA; Christina Keny[6,7], RN, PhD; Thomas Osborne[8,9], MD; Karleen Giannitrapani[1,10], MPH, PhD; Karl Lorenz[1,10], MSHS, MD

[1]Center for Innovation to Implementation, VA Palo Alto Health Care System, Palo Alto, CA, United States

[2]College of Medicine, California Northstate University, Menlo Park, CA, United States

[3]VA Center for Innovation to Implementation, Chapman University, Menlo Park, CA, United States

[4]School of Medicine, Yale University, New Haven, CT, United States

[5]Department of Communication, Stanford University, Stanford, CA, United States

[6]Division of Geriatrics, Department of Medicine, School of Nursing, University of San Francisco, San Francisco, CA, United States

[7]Division of Geriatrics, Department of Medicine, Veterans Affairs Medical Center, University of California, San Francisco, CA, United States

[8]Department of Radiology, School of Medicine, Stanford University, Palo Alto, CA, United States

[9]Palo Alto Health Care System, US Department of Veterans Affairs, Palo Alto, CA, United States

[10]Department of Primary Care and Population Health, School of Medicine, Stanford University, Stanford, United States

Corresponding Author:
Karl Lorenz, MSHS, MD
Center for Innovation to Implementation
VA Palo Alto Health Care System
795 Willow Rd
Palo Alto, CA, 94025
United States
Phone: 1 (650) 614 9997
Email: kalorenz@stanford.edu

Abstract

Background: Virtual reality (VR) has promise as an innovative nonpharmacologic treatment for improving a patient's quality of life. VR can be used as an adjunct or treatment for many acute and chronic conditions, including serious illnesses.

Objective: This systematic review aims to assess the current state of the literature of randomized controlled trials that use VR in patients with serious illnesses. Two secondary aims include assessing intervention components associated with improved quality of life and functional outcomes among older adults, as well as evaluating how well the randomized controlled trials adhere to consensus standards for VR research.

Methods: We searched PubMed, Embase, and CINAHL for randomized controlled studies published at any time. We screened and accepted studies that reported outcomes related to patients' quality of life, provided an immersive VR intervention, and included patients with serious illness. We narratively summarized key attributes of publications that shed light on study efficacy, generalizability, replicability, and clinical utility. All studies were assessed for study quality with the Cochrane Risk of Bias tool and for concordance with 8 recent consensus standards for VR research.

Results: From the 12,621 articles searched in May 2024, a total of 24 (0.19%) studies met the inclusion criteria, and of these, 88% (21/24) reported an improvement in at least 1 patient quality of life outcome and 67% (16/24) had a high risk of bias. In 7 (n=24, 29%) studies, VR was used to provide distraction therapy to reduce pain. In total, 5 (n=24, 21%) studies included training, supervision, and assistance in VR use, which demonstrated improvements in patient quality of life–related outcomes. Of 24 studies, 9 (38%) included patients with stroke, 9 (38%) included patients with cancer, 4 (17%) included patients with cardiovascular disease, 1 (4%) included patients with chronic obstructive pulmonary disease, and 1 (4%) included patients who reported pain in hospital. In all 9 studies that included patients with stroke, the main purpose of VR was to improve mobility and strength; these studies had higher frequency and longer durations of VR use, ranging from 2 to 9 weeks, as compared to a VR use duration of

XSL•FO
RenderX

<2 weeks for studies aiming to reduce pain or anxiety. Regarding consensus standards for VR research, 29% (7/24) of the studies adhered to all 8 criteria, and all studies (24/24, 100%) adhered to ≥5 criteria.

Conclusions: Nascent evidence suggests VR's potential in mitigating pain, anxiety, and depression and improving mobility among persons with serious illnesses. Most studies did not provide detailed information about unassisted or assisted use, suggesting that VR for older adults is currently most appropriate for observed settings with assistance available.

Trial Registration: PROSPERO CRD42022346178; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=346178

(J Med Internet Res 2025;27:e54452) doi:10.2196/54452

KEYWORDS

virtual reality; serious illness; pain; anxiety; older adults; patient outcomes; systematic review; palliative care; hospice

Introduction

Background

More than 60% of Americans will face a period of significant disability before death, incurring a diminished quality of life from the impact of serious illness, including multimorbidity [1-3]. Serious illness is defined as a health condition with a high risk of mortality that either negatively impacts daily function or quality of life or strains a person's caregivers [4]. Serious illnesses include those of indeterminate but limited prognosis and those that have a detrimental effect on the quality of life of patients and their caregivers. They encompass malignancy, organ failure, or frailty, a complex state of vulnerability that is often associated with dementia [4,5]. Serious illnesses are associated with burdensome symptoms and functional decline and include many leading causes of death, such as stroke, as well as multimorbidity [6]. Patients also face a significant financial burden because of medical treatment [7]. These conditions interfere with a person's ability to engage in activities they once enjoyed and impose emotional and physical burdens on family members and other caregivers; these conditions are of growing importance with the population aging [4].

Functional outcomes are a useful adjunct to other quality of life–relevant measures to understand the impact of serious illness on a person's life [8]. Activities of daily living (ADLs) include basic functions such as dressing, maintaining personal cleanliness, toileting, transferring, and eating. The impairment of 1 to 2 ADLs occurs in most Americans during the years before death, and ≥3 ADL impairments are associated with institutionalization. Important goals for patients include dying at home and maintaining cleanliness, both of which are abetted by functional status or require dedicated caregiving if adequate support is unavailable [9]. Evaluating functional outcomes in patients with serious illness is especially important because it directly reflects their quality of life and their ability to perform ADLs [10].

Virtual reality (VR) is one among emerging interventions (eg, smart technologies) with the potential to improve a patient's quality of life, patient-reported outcomes, psychological outcomes, and functional outcomes for persons living with serious illness [11-14]. VR offers a computer-generated, stereoscopically rendered, 3D visual environment, often with complementary sound, which responds continuously to a patient's movement. Immersive VR creates a psychological experience of treating the virtual simulation as a real experience. The sensory experience is thus distinct from nonimmersive reality, such as augmented reality that allows the participant to see computer-generated images superimposed on the real-world visual field [15,16]. Immersion and distraction therapies have been shown to be helpful in improving quality of life and functional outcomes. The use of VR in health care for serious illness care has been expanding to address pain, anxiety, and other needs in palliative care and hospice settings [17,18]. There is a need to understand the quality of research and VR's efficacy, especially in randomized controlled trials (RCTs), as it applies to older, seriously ill adults.

Objectives

Several recent reviews have explored content salient to our interest, although not in the specific context of applying VR to palliative care and serious illness [19-21]. We aim to determine the extent to which rigorous studies have focused on the rapidly growing population of older adults, the quality of that research, the extent to which it adheres to recently accepted VR research standards, and the characteristics of that research (eg, common outcomes) [22-24]. Two secondary aims include assessing the components of interventions in RCTs associated with improved quality of life and functional outcomes, as well as assessing the extent to which the current RCTs related to VR adhere to recommended consensus standards for high-quality VR research.

Methods

Protocol and Registration

Our review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocols statement and Enhancing the Quality and Transparency of Health Research guidelines [25,26]. We registered our protocol to the PROSPERO (CRD42022346178) [27].

Eligibility Criteria

Included in this systematic review are studies published in peer-reviewed journals up through May 2024 that meet the following criteria organized by population, intervention, comparator, outcomes, timing, and setting framework (Textbox 1) [28]. Functional outcomes include patient-reported outcomes relating to the quality of life and patient-demonstrated outcomes (eg, arm curl, chair stand, back scratch, chair sit, reach, walk test, overall cognition, and memory) [29,30]. Biological intermediaries (eg, functional magnetic resonance imaging scans, cortisol levels, blood test values, and balance) that do not directly demonstrate patient function are not included as

XSL•FO
RenderX

patient functional outcomes [31]. Eligibility criteria are described using the population, intervention, comparator, outcomes, timing, and setting framework (Textbox 1).

Textbox 1. Eligibility criteria using the population, intervention, comparator, outcomes, timing, and setting framework.

Population

- Inclusion criteria: studies that include adult patients (aged ≥18 years) with serious illness, such as stroke, defined by a variable but limited prognosis and an important illness impact quality of life
- Exclusion criteria: studies that include pediatric or adolescent patients (aged ≤18 years) or patients without serious illness

Intervention

- Inclusion criteria: study participants must use some form of immersive virtual reality in the intervention arm.
- Exclusion criteria
  - Intervention description insufficient to describe immersiveness
  - Augmented or extended reality (ie, nonimmersive reality)
  - 2D screens (eg, Wii and Wii Fit [Nintendo], Xbox [Microsoft Gaming], and YouGrabber)

Comparator

- Usual care

Outcomes

- Inclusion criteria: patient quality of life–related outcomes, including patient-reported (eg, pain and anxiety) or patient-demonstrated (eg, arm curl, chair stand, back scratch, chair sit, reach, walk test, overall cognition, and memory) outcomes
- Exclusion criteria: only biological intermediaries reported as results (eg, functional magnetic resonance imaging scans, cortisol levels, blood test values, and balance)

Timing: interventions with any follow-up period

Setting: any care setting (including in-patient clinics or outpatient and ambulatory care)

Selection of Sources for Evidence

Concepts included in the literature search were RCTs, VR, and serious illness as shown in Multimedia Appendix 1. We searched PubMed, Embase, and CINAHL for studies published at any time and identified 12,814 articles (n=10,799, 84.28% articles after duplicates were removed). During the title and abstract screening and full-text screening phases, 3 coauthor reviewers screened each study and were blinded to each other's decisions (BM, AK, and SF). During the title and abstract screening, we resolved conflicts predominantly by a "gold standard" reviewer (KL or KG), occasionally by majority consensus. During the full-text screening, reasons for exclusion were identified. Only the gold standard reviewer (KL or KG) resolved conflicts during full-text screening. During full-text screening, if any systematic reviews met inclusion criteria, we added their included articles to the title and abstract screening phase. We used the Covidence (Veritas Health Innovation) software to generate a PRISMA diagram to track studies at each stage of the review (Multimedia Appendix 2 [25]) [32].

Data Extraction

We built an abstraction form through an iterative process (Multimedia Appendix 3). All aspects of the included article interventions were recorded, including what the intervention entailed; the type of VR and media (eg, type of scenery or activity) used; any training, supervision, or assistance provided during VR use; the duration of sessions; frequency of sessions; and duration of intervention. The main outcomes collected were related to patient quality of life–related outcomes. Data abstraction for the included studies was done through Covidence with 2 reviewers per article. Two reviewers (BM, AK, or NI) abstracted each article independently and resolved all abstraction conflicts through a consensus discussion.

Risk of Bias and Quality Assessment

Risk assessment for bias was done using the Cochrane Risk of Bias tool for RCTs [33]. Two reviewers (SF, AK, or BM) independently assessed the risk of bias in the domains of randomization, allocation concealment, blinding, accounting of patients and outcome events, and selective outcome reporting bias. Any disagreements were discussed by each of the 2 reviewers, and a consensus was reached.

Synthesis of Results

We performed a narrative synthesis of the data abstracted due to the high degree of heterogeneity of the results, as each included article reported different forms of patient functional outcomes. We synthesized each VR-based intervention and extracted the purpose of VR within the intervention and whether there was any training, supervision, or assistance provided in the use of VR. To better understand the individual differences that may affect the efficacy of VR-based interventions, we looked at the average patient age, SD and range, type of serious illness, patient gender, patient ethnicity, and study location for each of the studies. We characterized each study's adherence

XSL•FO
RenderX

to recent consensus standards for high-quality VR research using the definitions below [22].

- Patient population: clear description of the study population (including inclusion and exclusion criteria)
- Clinical setting: appropriate setting for the VR treatment
- Control and randomization: justification for control and appropriate randomization
- Blinding and concealment of allocation: description of the method of allocation concealment
- End points: determination and justification of end points before the initiation of the study
- Study duration: determination and justification of the study duration before the initiation of the study; the study should be an a priori decision or hypothesis
- Presentation and analysis of results: clear presentation of results with appropriate statistical analysis
- Reporting the trial: reporting of the trial with a national or international registry (World Health Organization International Clinical Trials or ClinicalTrials.gov)

Results

Literature Selection

A total of 12,621 unique titles and abstracts were dual screened by coauthors as per the inclusion and exclusion criteria. A total of 152 full-text articles were retrieved and further assessed for eligibility. Of the 152 full-text articles reviewed, 128 (84.2%) studies were excluded for reasons noted in the PRISMA flow diagram (Figure 1 [34]). Finally, we included 24 total studies in our review [35-58], of which 21 (88%) reported an improvement in at least 1 patient functional outcome [35-43,45-48,51-58]. The included articles are summarized in Table 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. fMRI: functional magnetic resonance imaging; RCT: randomized controlled trial.

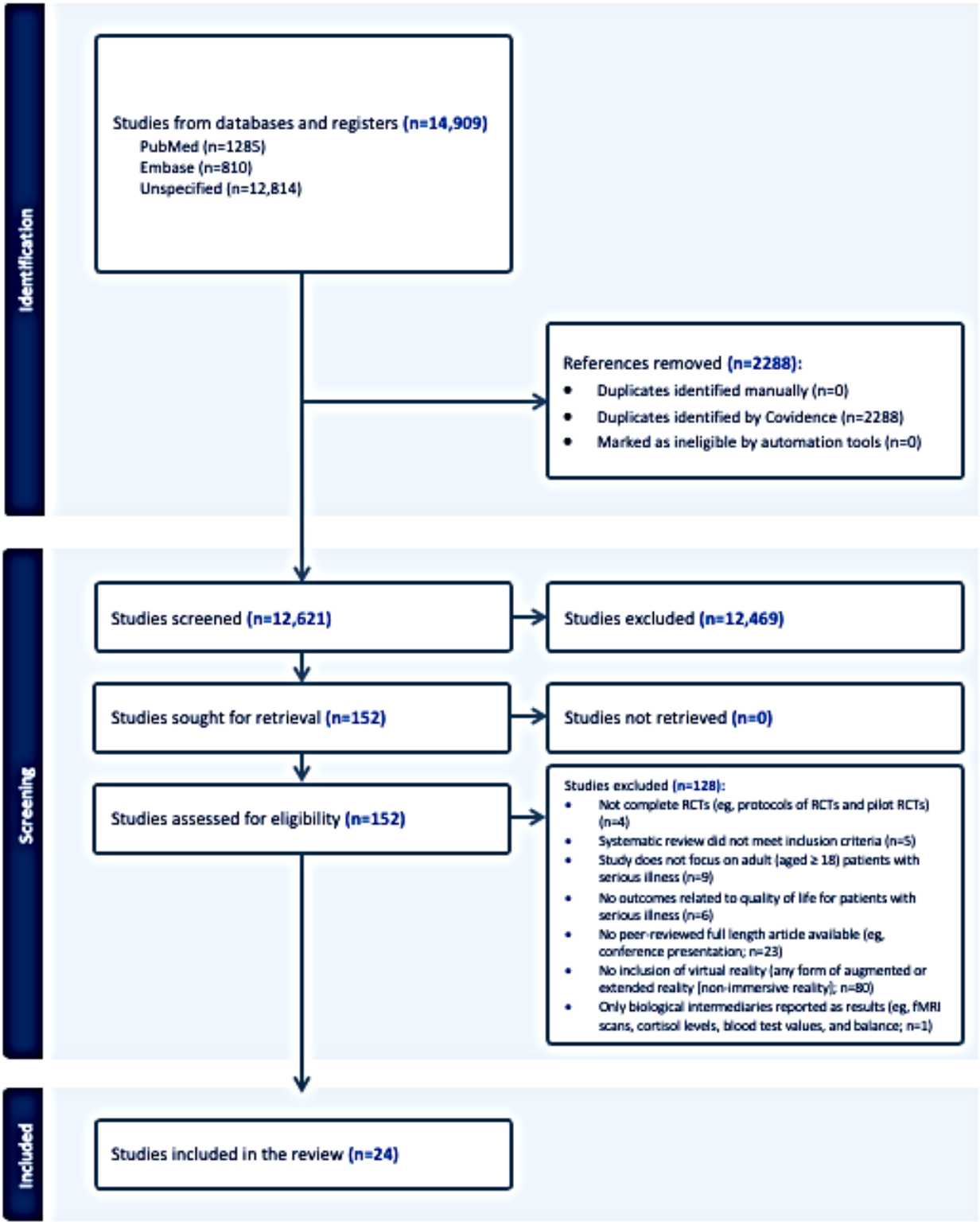

XSL•FO
RenderX

Table 1. Summary of the included studies (n=24).

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
Bani Mohammad and Ahmad [35], 2018	The patients wore a head-mounted display with headphones and chose either deep sea diving or sitting on the beach. The VR exposure session ended at the peak time of morphine efficacy.	• The patients were given morphine (same dose and timing as given to those in the intervention arm) with no additional intervention.	Distraction therapy to improve pain and anxiety in patients	The primary investigator remained near the participants during the VR session.	• Statistically significant improvement in pain scores (VASb; P<.001) and anxiety levels (SAIc; P<.001)
Burrai et al [36], 2023	Participants used the Oculus Quest 2 HMD[d] to observe specific virtual environments in a contemplative mode. The scenarios included 310 videos classified into 9 categories (eg, Africa, hills, rivers, lakes, and waterfalls). The audio content of scenarios had a background of nature sounds and soothing musical stimuli in high definition with panning stereo, and was listened to via earphones to ensure an immersive audiovisual experience with high-intensity multisensory immersion.	• One group of patients received narrative medicine, where patients shared their experience with cancer. • Another group of patients received standard care and served as the control arm.	Distraction from the stress of antiblastic therapy to reduce anxiety, fatigue, and pain	A nurse helped each participant position the VR headset and the joysticks. After immersion, the nurse removed the VR headset visor from the participant.	• Statistically significant decrease in anxiety (P=.001) and fatigue levels (P=.001) but no statistically significant change in average pain levels (P=.31)
Chatterjee et al [37], 2022	The main requirement was to allow the patients to practice ADL[e] completely or in part when placed within a suitable environment. The materials needed such as food items, kitchen gadgets, and coins were within reach of the patient when they were immersed in the environment. All the tasks were designed to be completed from a sitting position using a single handheld controller. The intervention was delivered 5 days a week for up to 2 weeks before the patient's hospital discharge. The dose of the virtue treatment varied depending on the benefit and tolerability.	• The patients received sham VR, where they were offered the same first VR session as the intervention arm and were subsequently offered the same VR session through the rest of the study.	Improve recovery and reduce the time that a patient spends in the hospital	Supervised by a therapy assistant while taking part in the VR session. Assisted in wearing the headset and ensured a comfortable and correct seating position. The virtue program was started by the therapy assistant, and the difficulty of each scenario was graded depending on their needs.	• Statistically significant improvement in attention and orientation • No statistically significant difference in the primary outcome measures at 3 months • Secondary outcomes: statistically significant reduction in anxiety scores in the mild to moderate cognitive impairment (MoCAf 15-24) treatment group
Chirico et al [38], 2020	Each chemotherapy treatment lasted for 45 to 90 minutes. Patients used the VR system for 5 to 10 minutes to become accustomed to it, and then the nurse administered the chemotherapy. In particular, participants explored an island by walking through a forest, observing different animals, climbing a mountain, and swimming in the sea.	• One group of patients received music therapy, where they were provided an MPEG audio layer 3 reader and headphones 5 minutes after starting the chemotherapy infusion. Patients listened to relaxing music pretaped by an expert music therapist for 20 minutes. • Another group of patients was allowed to choose common activities during treatment, such as conversation or reading.	Reduce the anxiety, depression, and fatigue related to chemotherapy through distraction	A research nurse, together with a trained psychologist, explained how to use the VR equipment and helped patients with putting on the headset.	• Statistically significant reduction in anxiety levels between pretest and posttest and a significant interaction effect between the time factor and the intervention factor (time×intervention) on the patient's level of anxiety. Furthermore, the analysis showed a significant main effect for the time factor. There was a significant reduction between the preintervention and the postintervention phases in the following negative mood states: tension, depression, anger, and fatigue (P<.001)

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
Groninger et al [39], 2021	The study used the Forest of Serenity (Holosphere VR), which features a 10-minute guide through a forest and waterfall with voice narration. Each assigned experience (VR or guided imagery) was administered at the bedside by the study coordinator. When the participant was ready, the coordinator left the bedside to stand immediately outside the room, started to time the experience per the protocol, and waited either for the end of the assigned experience time or for a verbal signal from the participant requesting assistance with the equipment.	• The patients received a guided imagery session on a portable tablet, which had options of guided meditation, instrumental background music, and 2D imagery of a peaceful lakeside view.	VR aimed to provide distraction therapy	The study coordinator explained the equipment and positioned it, then left the room until the end or if assistance was asked for.	• Statistically significant decrease in pain scores (Likert scale); the VR group had significantly lower scores than the guided imagery group (P<.001). • Secondary outcomes: there was a statistically significant improvement in the total FACIT-Palg 14 score and an insignificant change in the distress score
Hsu et al [40], 2022	In addition to 20 minutes of therapist-facilitated task-specific training as a usual care session, patients each received 30 minutes of VR time. The sequences of the hand exercises of VR consisted of the movements of the forearm, supination or pronation, wrist extension or flexion, finger extension or flexion, thumb opposition with the little finger, thumb extension or flexion, and tendon-gliding exercises, which involve a series of hand movements (straight hand, hook fist, straight fist, and full fist). Each movement was repeated 50 times. The participant sat in a comfortable chair in front of a desk with the VR system.	• One group of patients performed hand exercises and looked at their hand motions through a mirror. • Another group of patients received OTh in addition to physical therapy and speech therapy.	Improve upper extremity movement, quality of motor movement in the affected arm, and manual dexterity	Participants received 20 minutes of therapist-facilitated task-specific training. They were supervised and assisted by the physical therapist.	• Statistically significant improvement in only the FMA-UEi motor coordination (P=.002) and the MAL-AOUj (P=.002)
Huang et al [41], 2022	A total of 20 VR scenes from commercial games were selected for the participants. Selected scenes were chosen based on the original upper limb activities. Participants could accomplish the first-contact task with the help of their unaffected hand, and then, they were encouraged to use the affected hand or both hands at the same time. Upper limb movements in most scenes involved aiming, shooting, hitting, waving arms, punching, and throwing objects. Participants' activity performance and information were recorded, including training intensity, duration, game scores, level completed, and invalid activities. This information allowed the therapist to assign the tasks, adjust the difficulty in the VR game settings, and even design a new way to play the game for each participant according to their needs and capacities.	• The patients received OT.	VR aimed to improve upper extremity movement and range of motion	After the participant confirmed that the sight and sound were clear and comfortable, the controllers were handed to the participant, and the therapist could even design a new way to play the game for each participant according to their needs and capacities.	• Statistically significant time effect for all items of upper limb assessment (P<.05), except FMA-UE motor coordination and speed
Huang et al [42], 2024		• Patients received a 60-minute conventional rehabilitation program per day.	Improve motor impairment and ADL		• Statistically significant improvements in upper extremity motor impairment and ADL for up to at least 12 weeks postintervention, and the magnitude of these improvements was much greater than that of other intervention programs (P=.02)

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
	Participants in the imVR[k] system received the first 30 minutes of conventional rehabilitation, and in the second 30 minutes, the rehabilitation was performed in imVR systems. Participants in the imVR group were required to complete 6 programs: frying dumplings and noodles by controlling a wok handle in a virtual kitchen; popping balloons by controlling a sword in a virtual fencing hall; punching dolls by controlling a big fist in a virtual boxing arena; playing basketball in a virtual court, in which the ball is shot by a controller and the height and distance is varied over time; collecting eggs into a virtual basket by a controller; and tidying up a desk and moving objects to a designated position in a virtual office. All participants received rehabilitation training for 5 days per week for 3 weeks.			No training, supervision, or assistance was provided; however, in the early stages of rehabilitation, due to poor functioning of the hemiplegic side upper limb, participants used the help of their unaffected side limb.	
Jo et al [43], 2024	Each participant of the 360° MTG[l] performed treatment tasks using their unaffected extremities, which were recorded with a 360° camera (Insta 360° X3; Insta 360). The patients received treatment in the same place where the videos were recorded. The therapist instructed and guided patients to use the affected side following the immersive video during all tasks. The MT[m] tasks were the same as those of TMTG[n] (finger flexion and extension, wrist flexion and extension, pronation and supination, and elbow flexion and extension for 10 repetitions; the tasks instructed participants to use the affected limb while following the mirror images of the unaffected limb to induce movement re-education; and playable tasks using ring toys were also provided for 5 repetitions). The 360° MT was provided for 30 minutes per session, 3 sessions a week for 4 weeks. The 360° MTG also received additional conventional physical therapy.	• One group received traditional MT through an acrylic mirror that mirrored the unaffected side upper limb, where the affected side limb was hidden behind the mirror. • Another group received a conventional physical therapy program consisting of warm-up, circuit, and cool-down exercises.	Upper extremity rehabilitation in patients with stroke	A therapist adjusted the device to fit the eye level of each patient. The therapist instructed and guided patients to use the affected side following the immersive video during all tasks.	• FMA-UE, MFTo, and BBTp scores were significantly improved in 360° MTG and TMTG, while improvements in CGq were not significant. In group comparisons, changes in 360° MTG were significantly higher compared to changes in TMTG and CG (P<.001)
Laghlam et al [44], 2021	The VR session started during the preparative phase, at least 5 minutes before the removal of the drains, and was continued for 10 minutes after. We used a VRx helmet with a 90° field of view with head tracking. Patients had a choice between 5 different immersive environments (360° videos): a snowy mountain, a landscape in India or Camargue (France), a balloon ride, or a canoe descent.	• The patients received Kalinox (Air Liquide) 1 minute before the removal of the drains. This was delivered continuously and stopped 1 minute after removal to avoid side effects.	Reduce pain during cardiac surgery, specifically during the removal of drains	The nurse supervised and assisted patients while using VR.	• No statistically significant improvements in patient functional outcomes (Analgesia Nociception Index; P=.69)
Lee and Kang [45], 2020		• The patients received a daily routine sleep intervention.	Induce meditation and improve sleep quality		

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
	The "Relaxing Music for Meditation" program was designed to induce meditation and consists of 8 meditation videos of 30 minutes. Scenes of mountains, seas, or lakes appear on the VR screen, and background sounds and calm music allow participants to immerse themselves in the experience. The study began on the day of admission to the ICU[r].			No training, supervision, or assistance was mentioned; however, it took about 20 minutes for the experimental group to learn how to use the HMD.	• Statistically significant increase in sleep scores (Pittsburg Sleep Quality Index), sleep time, and mean sleep efficacy in the experimental group compared to CG (P=.002)
Mekbib et al [46], 2021	The VR rehabilitation training included reaching, grasping, and releasing tasks. Following the training schema, all patients in the VR group received 1-hour VR and 1-hour OT per day, 4 days per week for 2 weeks. In each VR session, the therapy modes were set by a therapist based on the patient's interest and actual motor capability. After choosing the therapy mode, the therapist randomly set 20 colored balls from the aerial view map. Then, the patient was instructed to reach, grasp, and release each ball into the basket. After completing the first VR session (20 balls), the therapist could set the next VR session and adjust the task complexity to be slightly higher or lower or keep it as it was in the previous session based on the patient's activity performance.	• The patients received OT.	Upper extremities to aid in the recovery of motor function after stroke and improve upper extremity function and the ability to perform ADL	The therapist helped train the patient. If the patient was not able to move their upper extremity toward the target ball, the therapist could help move the affected extremity to assist in the intended task.	• Statistically significant improvement in the FMA-UE scores (P<.001) and BIs scores (P=.003)
Menekli et al [47], 2022	Patients in the intervention group were educated on the use of VR glasses by the researcher. A smartphone and the various number of parks, nature, and seaside walks; submarine; and museum tours with the Matic Music were provided to patients in the intervention group. Each of these videos took approximately 3 to 10 minutes, and patients could choose the video that they wanted to watch. Patients started to use the VR after the baseline assessment (about 1 minute before the implantation) and continued to use it until the end of the implantation. Patients were instructed to use VR after implantation when they felt pain from the implantation.	• The patients received conventional care.	Pain management after implantation	Supervision and assistance were provided by a nurse, who was giving directions to patients on how to use the videos.	• Statistically significant improvement in the SAI scores (P<.001) and VAS scores (P<.001)
Ögün et al [48], 2019		• The patients received conventional upper extremity active exercises comprising the same tasks as used in the VR group.	Improve upper extremity functionality and make daily activities more accessible	None mentioned	• Statistically significant improvement in FMA-UE, ARATt, FIMu, and PASSv scores (P<.001)

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
	Patients used the VR device to play task-oriented games that focused on gripping and handling objects with arm and forearm motion and stability. A different game was used for each function, with 4 games: a cube handling game used for grip function integrated with the leap motion device to make the patients feel like they were handling a real object using their own fingers without the use of any external device to track hand motion; another leap motion–integrated game involving decorating a tree with leaves and fruits or picking up vegetables from a bowl and putting them back, which was chosen to facilitate all hand motions combined with complex motions in a task-oriented job; a kitchen experience game used for stimulating forearm supination and pronation and for combining complex arm movements; and a drumming game, selected to randomly assign each separate movement of upper extremity flexion and abduction.				
Park et al [49], 2013	The VR-based postural control program consisted of a program for the improvement of gait ability by visual feedback compared to reference motion scenes and reality motion. The program consisted of 3 stages: trunk stability and pelvic tilting in a supine position (stage 1); trunk upright control and pelvic tilting exercise in a sitting position, and a selective movement between trunk and pelvis (stage 2); and lower extremity muscle strengthening exercise and weight bearing under maintenance of trunk stability in a standing position (stage 3). The participants were provided with specific scenes on HMD showing the simultaneous output.	• The patients received targeted lower extremity muscle strengthening exercises, static and dynamic balance training, and gait training.	Improve postural control, spatiotemporal gait ability, and functional gait ability and to increase walking ability	The participant saw the prerecorded reference motion and practiced it again 3 times. No other provision of training, supervision, or assistance was mentioned.	• No statistically significant improvements in patient functional outcomes
Rousseaux et al [50], 2022	A 20-minute VR session was conducted, where participants watched a 3D graphical landscape consisting of a mountain cabin near a lake at sunrise, followed by a relaxing moment in the clouds. The display was visual and audio, with sounds of ambient nature but no voice.	• The patients received conventional care.	Lower anxiety and decrease pain	None mentioned	• No statistically significant differences in the effectiveness, either between the techniques themselves or between the techniques and no treatment
Rutkowski et al [51], 2021		• The patients received the traditional pulmonary rehabilitation program.	Reduce depression, anxiety, and stress levels and improve functional capacity	None mentioned	• Statistically significant improvement in the HADS-Ax (P<.001), HADS-Dy (P<.001) and general HADSz (P<.001)

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
	The primary aim of the software was to calm the patient down and improve his or her mood. The software features a virtual therapeutic garden and is based on the Ericksonian psychotherapy approach. The garden is a metaphor for the patient's health: at the beginning, it appears untidy and gray; however, with each session, it becomes more colorful and alive, thus symbolizing the process of recovery of energy and vigor. The 2 groups participated in the traditional pulmonary rehabilitation program. Components were performed once a day each, for 15 to 30 minutes (depending on the task), 5 times a week for 2 weeks. Exercises were performed as follows: fitness exercises while standing on the knees and lying on the side, abdomen, and back; strengthening exercises of the diaphragm with resistance; prolonged exhalation exercise; chest percussion; inhalation with a 3% sodium chloride isotonic solution administered via an ultrasonic device; and stationary cycle ergometer exercise to obtain a training heart rate according to GOLD[w] spirometric stages. The difference between the groups is in the type of relaxation training: the VR group performed 10 VR therapy sessions of 20 minutes, and the CG performed 10 Schultz autogenic training sessions of 20 minutes.				
Shin et al [52], 2023	In the virtual world, patients were greeted by their respective attending physicians and guided through the visual and audio unfolding of the RT[aa] process. Patients were guided virtually with 360° viewing of the different treatment environments and RT machines that they would encounter when undergoing RT. Throughout the physician's explanation, patients were able to witness a real example of a patient with breast cancer undergoing the RT, offering patients a concrete example of what to expect on treatment days.	• The patients received conventional care.	Alleviate anxiety related to undergoing RT	Patients were assisted with the proper positioning of the VR headset and noise-cancelation earphones provided by the hospital.	• Significant reduction in patient anxiety immediately, not only on the primary end point, APAISab, but also on the STAIac and LASAad anxiety scales. Among the 3 anxiety scales, long-term anxiety reduction was observed only when anxiety was measured by LASA
Song and Lee [53], 2021		• The patients performed exercises, such as turning on lights, arranging a chest of drawers, and arranging a kitchen as part of bilateral upper extremity training in a real environment.	Improve the range of motion, bilateral upper extremity movement, and upper limb function	There was some training, supervision, or assistance provided; however, this was not written, although figures in the study show someone guiding the participant.	• Statistically significant improvement in the proprioception test (P<.05)

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
	The VRBAT[ae] group underwent immersive VRBAT. The VR content consisted of rehabilitation tasks to improve upper limb function, the ability to perform ADL, and visual perception. The VR tasks performed included everyday activities, such as turning on lights, organizing a chest of drawers, organizing a kitchen, watering plants, and purchasing items at a convenience store. The virtual living room, kitchen, veranda, and convenience store were designed to simulate real environments. Interventions were performed for 30 minutes a day, 5 times a week, for 4 weeks, for a total of 20 sessions. All participants in the VRBAT group also underwent an hour of conventional rehabilitation per day.				
Spiegel et al [54], 2019	Participants selected one of the 21 VR experiences from an app. Patients were asked to use the headset for 10 minutes in the presence of study staff to practice with the equipment and then advised to use the headsets thrice daily, for 10 minutes per session, and later, as needed for breakthrough pain, for the subsequent 48 hours. Following these initial instructions, patients decided for themselves and in partnership with their care team whether, how frequently, and how long to use the VR equipment without direct input from study staff.	• The patients watched the "Health and Wellness Channel," which includes guided relaxation content (eg, yoga and meditation programming), discussions about health and wellness topics, and poetry readings.	Reduce pain in hospitalized patients	Participants were instructed on procedures for wearing the headset, how to select among the 21 VR experiences, and how to adjust volume and brightness. Regarding supervision, patients were initially asked to use the headset for 10 minutes with the presence and assistance of the study staff and care team.	• Significant improvement in pain scores 48 hours and 72 hours postintervention using a numeric rating scale (P=.03 and P=.04, respectively)
Torres García et al [55], 2023	The sessions recreated a chemotherapy procedure. Using VR, patients managed their anxiety and fear through relaxation, attentional focus, and mindfulness strategies. The complete intervention included 3 modules, which each patient repeated during 4 individual sessions, aiming to improve their preparation and adaptation to the cancer treatment they had to start.	• The patients received a regular psychoeducational intervention.	Improve patients' anticipatory anxiety and adherence to chemotherapy and facilitate better and faster adaptation to treatment coping mechanisms	The therapist was present throughout the procedure.	• Statistically significant differences in the decrease of anxiety and depression with the Anxiety and Depression Scale. At 3 months, a more notable decrease in the Emotional Discomfort Detection Scale was observed in the experimental group, and the fighting spirit and cognitive avoidance increased, which was assessed with the MINI-MACaf coping scale
Turrado et al [56], 2021		• The patients received conventional care.	Reduce presurgical anxiety and depression	None mentioned	• Statistically significant changes for both the HADS subscales and for those measuring status and trait in the STAI scale for patients who were exposed to the VR simulation experience (P<.001)

XSL•FO
RenderX

Study	Description of the intervention and media content	Control arm	Purpose of VR[a] in the intervention	Training, supervision, or assistance provided in the use of VR	Patient-reported outcomes or patient-demonstrated outcomes
	A realistic environment was generated in which the patient could experience the various steps of their admission to surgery, from the first interview with the surgeon to admission into the surgical ward, the operating room, and the postoperative recovery room. Patients completed both questionnaires to assess their basal anxiety levels. The patients in the intervention group had unlimited access to VR glasses and to the VR app. This group completed both questionnaires again on the day before the surgical procedure.				
Uslu and Arslan [57], 2023	The patients watched and listened to beach and nature content with VR glasses for 30 minutes. The VR glasses used were Zore G04BS VR Shinecon VR glasses, compatible with smartphones (eg, Android and IOS). The contents were determined in the YouTube library and included 360° relaxing beach and nature images.	• The patients received conventional care.	Reduce anxiety and fatigue levels	Patients were informed about using VR and viewing the content. From the content opened by the researcher, it was ensured that the patient watched and listened to the content of their choice.	• The mean postintervention anxiety scores decreased in the intervention group. In this group, the mean posttest fatigue and subscales scores decreased in all cycles compared with the mean pretest scores
Zhang et al [58], 2023	The patients watched and listened to beach and nature content with VR glasses for 30 minutes. The VR glasses used were Zore G04BS VR Shinecon VR glasses, compatible with smartphones (eg, Android and IOS). The contents were determined in the YouTube library and included 360° relaxing beach and nature images.	• The patients received conventional care.	Reducing anxiety and fatigue levels	Patients were informed about using VR and viewing the content. From the content opened by the researcher, it was ensured that the patient watched and listened to the content of their choice.	• The mean postintervention anxiety scores decreased in the intervention group. In this group, the mean posttest fatigue and subscales scores decreased in all cycles compared with the mean pretest scores

[a]VR: virtual reality.

[b]VAS: visual analog scale.

[c]SAI: State Anxiety Inventory.

[d]HMD: head-mounted display.

[e]ADL: activities of daily living.

[f]MoCA: Montreal Cognitive Assessment.

[g]FACIT-Pal: Functional Assessment of Chronic Illness Therapy–Palliative Care.

[h]OT: occupational therapy.

[i]FMA-UE: Fugl-Meyer assessment for upper extremity.

[j]MAL-AOU: motor activity log-amount of use.

[k]imVR: immersive virtual reality.

[l]MTG: mirror therapy group.

[m]MT: mirror therapy.

[n]TMTG: traditional mirror therapy group.

[o]MFT: manual function test.

[p]BBT: box and block test.

[q]CG: control group.

[r]ICU: intensive care unit.

[s]BI: Barthel Index.

XSL•FO
RenderX

[t]ARAT: action research arm test.
[u]FIM: functional independence measure.
[v]PASS: performance assessment of self-care skills.
[w]GOLD: Global Initiative for Chronic Obstructive Lung Disease.
[x]HADS-A: Hospital Anxiety and Depression Scale-Anxiety.
[y]HADS-D: Hospital Anxiety and Depression Scale-Depression.
[z]HADS: Hospital Anxiety and Depression Scale.
[aa]RT: radiation therapy.
[ab]APAIS: Amsterdam Preoperative Anxiety and Information Scale.
[ac]STAI: State-Trait Anxiety Inventory.
[ad]LASA: linear analog scale assessment.
[ae]VRBAT: VR-based bilateral arm training.
[af]Mini-MAC: mini-mental adjustment to cancer.

Participant Characteristics

In total, the studies included 1225 participants. Of the 24 included studies, 3 (12%) included patients with an average age of <50 years [43,49,58]; 8 (33%) included patients with an average age of 51 to 60 years [35,36,38-41,46,52,55,57]; 5 (21%) included patients with an average age of 61 to 65 years [42,45,48,51,52]; 3 (12%) included patients with an average age of 66 to 70 years [44,50,56]; and 1 (4%) included patients with an average age of >75 years [37]. Two studies did not report the average age of the patients, as listed in Table 2 [47,54]. No studies reported how many patients in the sample were aged >65 years. A total of 16 (67%) of the 24 included studies had a high risk of bias (Multimedia Appendix 4 [35-58]).

XSL•FO
RenderX

Table 2. Patient characteristics.

Study	Patients, n	Patient age (y), average (SD); range	Serious illness, n (%)	Sex, n (%)	Study location	Ethnicity
Bani Mohammad and Ahmad [35], 2018	80	• 52 (10.34); 30-70	Breast cancer stage: • 1: 9 (11.3) • 2: 32 (40) • 3: 29 (36.3) • 4: 10 (12.5)	• Female: 80 (100)	Jordan	• Information not provided
Burrai et al [36], 2023	74	• 59 (10.8); 33-83	• Breast cancer: 35 (47.2) • Colon cancer: 10 (13.5) • Rectum cancer: 6 (8.1) • Melanoma: 2 (2.7) • Non-Hodgkin lymphoma: 1 (1.4) • Stomach cancer: 2 (2.7) • Ovary cancer: 5 (6.8) • Pancreas cancer: 4 (5.4) • Uterine cervix cancer: 1 (1.4) • Larynx cancer: 2 (2.7) • Esophagus cancer: 1 (1.4) • Testicle cancer: 1 (1.4) • Prostate cancer: 1 (1.4) • Lung cancer: 2 (2.7) • Biliary tract cancer: 1 (1.4)	• Female: 55 (74.3) • Male: 19 (25.7)	Italy	• Italian: 24 (97.2) • Not Italian: 1 (0.8)
Chatterjee et al [37], 2022	40	• Average age: not reported (SD 17.67) • Intervention median (IQR) age: 77.5 (43-89) years • Control median (IQR) age: 63 (29-86) years	• Ischemic: 34 (85) • Hemorrhagic: 6 (15)	• Female: 19 (47.5) • Male: 21 (52.5)	United Kingdom	• Information not provided
Chirico et al [38], 2020	92	• 55 (6.00)	Breast cancer stages: • 1: 24 (26.1) • 2: 39 (42.4) • 3: 29 (31.5)	• Female: 92 (100) • Male: 0 (0)	Italy	• Information not provided
Groninger et al [39], 2021	88	• 56 (13.24)	• Heart failure: 19 (51.4) • Other cardiac diseases: 5 (13.5) • Other medical and surgical problems: 13 (35.1)	• Female: 35 (39.8) • Male: 53 (60.2)	United States	• African American: 74 (84.1) • Hispanic: 1 (1.1) • White: 13 (14.8)
Hsu et al [40], 2022	52	• 55 (12.11)	Stroke type: • Ischemic: 29 (55.8) • Hemorrhagic: 23 (44.2)	• Female: 32 (61.5) • Male: 20 (38.5)	Taiwan	• Information not provided
Huang et al [41], 2022	30	• 58 (11.78) • Intervention group range: 22-70 • Control group range: 28-71	• Infarction: 22 (73) • Hemorrhage: 8 (27)	• Female: 20 (67) • Male: 10 (33)	Taiwan	• Information not provided
Huang et al [42], 2024	40	• 64 (10.99)	Stroke type: • Ischemic: 37 (92.5) • Hemorrhagic: 3 (7.5)	• Female: 16 (40) • Male: 24 (60)	China	• Information not provided

XSL•FO
RenderX

Study	Patients, n	Patient age (y), average (SD); range	Serious illness, n (%)	Sex, n (%)	Study location	Ethnicity
Jo et al [43], 2024	45	• 50 (13.36)	Stroke paretic side: • Right: 23 (51) • Left: 22 (49)	• Female: 22 (49) • Male: 23 (51)	South Korea	• Information not provided
Laghlam et al [44], 2021	180	• 68 • Intervention group range: 61.5-75.0 • Control group range: 60.0-74.0	• Isolated coronary bypass surgery: 99 (55) • Combined coronary bypass surgery: 7 (3.9) • Isolated or combined valve surgery: 61 (33.9) • Aorta surgery: 5 (2.8) • Myxoma: 1 (0.6) • Pericardiocentesis for tamponade: 6 (3.3) • Abdominal aorta surgery: 1 (0.6)	• Female: 46 (25.56) • Male: 134 (74.4)	France	• Information not provided
Lee and Kang [45], 2020	48	• 66 (16.40)	• CHFa: 6 (12.5) • Cardiomyopathy: 2 (4.2) • MIb: 26 (54.2) • Others: 12 (25.1)	• Female: 16 (3.33) • Male: 32 (66.7)	South Korea	• Information not provided
Mekbib et al [46], 2021	23	• 56 (10.96)	Stroke type: • Ischemic: 17 (73.9) • Hemorrhagic: 6 (26.1)	• Female: 6 (26.1) • Male: 17 (73.9)	China	• Information not provided
Menekli et al [47], 2022	139	• Average age: not reported • Age by category (years), n (%): 20-30: 23 (16.55); 31-40: 34 (24.46); 42-52: 60 (43.17); 53-63: 22 (15.83)	• Breast cancer: 53 (38.1) • Lung cancer: 46 (33.1) • Stomach (gastric) cancer: 18 (12.9) • Colon cancer: 12 (8.6) • Pancreas cancer: 10 (7.2)	• Female: 87 (62.6) • Male: 52 (37.4)	Turkey	• Information not provided
Ögün et al [48], 2019	65	• 61 (9.62)	• Stroke: 65 (100)	• Female: 14 (22) • Male: 51 (78)	Turkey	• Information not provided
Park et al [49], 2013	16	• 48 (7.89)	Stroke type: • Ischemic: 9 (56.3) • Hemorrhagic: 7 (43.7)	• Female: 5 (31.25) • Male: 11 (68.75)	Korea	• Information not provided
Rousseaux et al [50], 2022	100	• 66 (11.50)	• Coronary artery bypass graft surgery: 57 (57) • Mitral valve replacement: 26 (26) • Multiple interventions: 4 (4) • Others: 4 (4)	• Female: 24 (24) • Male: 76 (76)	Belgium	• Information not provided
Rutkowski et al [51], 2021	50	• 64 (7.77)	• COPDc: 50 (100)	• Female: 41 (82) • Male: 9 (18)	Poland	• Information not provided
Shin et al [52], 2023	196	• 47.5 (7.63)		• Female: 196 (100)	South Korea	• Information not provided

XSL•FO
RenderX

Study	Patients, n	Patient age (y), average (SD); range	Serious illness, n (%)	Sex, n (%)	Study location	Ethnicity
			Breast cancer stages: • Stage 0: control group: 28 (28.3); VR group: 15 (15.5) • Stage 1A: control group: 43 (43.4); VR group: 47 (48.5) • Stage 1B: control group: 3 (3); VR group: 8 (8.3) • Stage 2A: control group: 16 (16.2); VR group: 21 (21.7) • Stage 2B: control group: 5 (5.1); VR group: 2 (2.1) • Stage 3A: control group: 2 (2); VR group: 2 (2.1) • Stage 3C: control group: 1 (1); VR group: 1 (1Stage 4: control group: 1 (1); VR group: 1 (1)			
Song and Lee [53], 2021	10	• 64 (9.18) • Intervention group range: 51-80 • Control group range: 51-80	• Hemorrhagic stroke: 3 (30) • Ischemic stroke: 7 (70)	• Female: 4 (40) • Male: 6 (60)	Korea	• Information not provided
Spiegel et al [54], 2019	120	• 60 (15.50)	Intervention: • Gastrointestinal: 13 (10.8) • Infectious disease: 15 (12.5) • Internal medicine: 25 (20.8) • Oncology: 10 (8.3) • Orthopedics: 36 (30) • Other: 21 (17.5)	• Female: 60 (50) • Male: 60 (50)	United States	• African American: 21 (34.4) • White: 38 (62.3) • Other: 2 (3.3)
Torres García et al, 2023 [55]	133	• 49 (11.6); 30-82	• Breast cancer: 133 (100)	• Female: 133 (100)	Spain	• Information not provided
Turrado et al [56], 2021	126	• 64 • Intervention group range: 41-85 • Control group range: 50-86	• Colorectal cancer: 58 (46) • Right colectomy: 21 (16.7) • Sigmoidectomy: 20 (15.9) • LARd+TaTMEe: 25 (19.8) • Transverse colon resection: 2 (1.6)	• Female: 53 (42) • Male: 73 (58)	Spain	• Information not provided
Uslu and Arslan [57], 2023	66	• 52 (8.89)	• Breast cancer patients receiving adjuvant chemotherapy: 66 (100)	• Female: 66 (100)	Turkey	• Information not provided
Zhang et al [58], 2023	60	• 34.38 (10.76)	• Acute myeloid leukemia: 47 (78.33) • Lymphoblastic leukemia: 1 (21.67)	• Female: 33 (55) • Male: 27 (45)	China	• Information not provided

[a]CHF: congestive heart failure.

[b]MI: myocardial infarction.

[c]COPD: chronic obstructive pulmonary disease.

[d]LAR: low anterior resection.

[e]TaTME: transanal total mesorectal resection.

Serious Illness

Stroke, cancer, and cardiovascular disease were the most common types of serious illnesses for which VR interventions were used to impact the outcomes related to patients' quality of life. Out of 24 included studies, 9 (38%) focused on patients who experienced stroke, which included both ischemic stroke and hemorrhagic stroke [37,40-43,46,48,49,53], and 9 (38%) included patients who were diagnosed with some type of cancer, varying in the stages of cancer in each patient [35,36,38,42,43,47,52,55-58]. In 5 (21%) of the 24 studies, only patients who were diagnosed with breast cancer were included; 1 study (4%) included patients with colorectal cancer, and 1 (4%) included patients with heterogeneous types of cancer. Four

XSL•FO
RenderX

(17%) out of 24 studies included patients who had some form of cardiovascular disease, including heart failure and other conditions that required cardiac surgery [39,44,45,50]. One study included patients with chronic obstructive pulmonary disease (COPD), and another included patients who reported pain during their hospitalization [51,54].

Purpose and Content of VR

In 9 (38%) out of 24 included studies, the main purpose of VR was to improve mobility and strength in the patient, ranging from upper extremity, lower extremity, and pelvic or abdominal region to a combination of these regions [37,40-43,46,48,49,53]. All of these mobility studies also included patients who experienced strokes. The VR media used in these studies encouraged active participation from the patients in various tasks and games. Some examples of VR media used to help improve mobility and strength in patients' poststroke treatment were daily exercises that encouraged increased range of motion, strength for specific movements, and reach and grasp through the use of balls. A total of 8 (89%) of the 9 studies that used VR for the purpose of improving mobility and strength improved patient quality of life–related outcomes [37,40-43,46,48,53].

There was some overlap in the protocols and diseases of the additional studies. Out of 24 studies, 12 (50%) used VR distraction therapy to reduce pain in patients with cancer or, cardiovascular disease or hospitalized patients [35,36,38,39,42-44,47,50,52,54-58]. A total of 9 (38%) out of 24 studies used VR to reduce anxiety in patients with cancer, cardiovascular disease, or COPD [36,38,42,43,50-52,55-58]. In studies aiming to reduce pain and anxiety, patients experienced immersive natural scenery and comforting music while comfortably sitting or lying down. In another study, to reduce anxiety, patients experienced the hospital environment and could visualize and review all the steps of an upcoming procedure [56]. In another study, patients with cardiovascular disease went through a visual guided meditation track with the VR media for patients with cardiovascular disease [45].

Training, Supervision, and Assistance

In 18 (75%) out of 24 studies, staff or physical therapists either trained patients on VR use before the intervention or supervised or assisted them through the experiment. Supervision was defined as the research team watching over the participants, ensuring their use and adherence to study guidelines. However, assistance was defined as the research team supporting patients during the intervention without constant watching throughout the experiment. A total of 5 (20.8%) out of 24 studies included VR training, supervision, and assistance, all of which reported improvements in patient quality of life–related outcomes [38,40,43,46,54]. Of 24 studies, 7 (29.2%) included VR supervision and assistance only [36,37,41,44,47,53,57], and 2 (8.3%) included VR training and assistance only [39,49]. Two (8.3%) out of 24 studies included only VR supervision while patients were using VR [35,55]. In studies that included a variation of personnel working with the patient, only 1 staff member physically accompanied the patient in the room, where they freely communicated with the patient to fulfill their role. During a typical training, the staff member would walk the patient through the VR activity and teach them how to adjust the headset. Out of 24 studies, 6 (25%) did not describe training, supervision, or assistance with the use of VR [42,45,48,50,51,56].

Frequency and Duration of Intervention on Quality of Life Outcomes

The frequency and duration of VR experience varied based on the intervention purpose; interventions aiming to improve pain or anxiety were shorter. while those aiming to improve mobility and strength were longer. Out of the 24 included studies, 9 (38%) studies with pain or anxiety reduction as a goal, involved only 1 VR session, ranging from 8 minutes to 4 hours (ie, patients could use VR at any time within a 4-hour window) [35,38,39,44,47,50,52,56]. A total of 7 (78%) of these 9 studies had improvements in pain and anxiety metrics, such as the Hospital Anxiety and Depression Scale (HADS), State Anxiety Index, or the Numeric Pain Scale (NPS). One study, which conducted a single 30-minute meditation session, showed improvement in sleep scores of patients with cardiovascular disease [45]. In 1 case where VR was used for 10 minutes per session, for 3 sessions per day for 2 days, patients experienced statistically significant improvement in the NPS (P<.04) [54]. Patients in another study who used VR for 20 minutes per session for 5 sessions per week for 2 weeks had statistically significant improvement in the HADS (P<.001) [51].

In the 9 studies aimed at improving mobility and strength in patients who had a stroke, VR was used for a longer duration than in pain and anxiety studies. Out of 9 studies, 6 (67%) lasted between 2 and 4 weeks [37,42,43,46,49,53], while 3 (33%) lasted between 5 and 9 weeks [40,41,48]. A total of 8 (89%) of these 9 studies showed an improvement in the patient's functional capacity measured via the Fugl-Meyer upper extremity assessment, proprioception, or orientation metrics.

Gold Standard Adherence

As shown in Table 3, 7 studies adhered to all the 8 gold standard VR consensus standards criteria (100%) [41,43,44,48,51,52,54], 10 adhered to 7 criteria (88%) [36,37,39,40,42,46,55-58], 6 adhered to 6 criteria (75%) [35,38,47,49,50,53], and 1 adhered to 5 criteria (62%). The reasons for not adhering to the criteria were inadequate blinding, concealment of allocation, reporting of the trial, and a priori failure to justify study duration. The studies that did not report the trial via World Health Organization International Clinical Trials or Clinicaltrials.gov originated from countries where it may not be standard practice to report the clinical trial before the study initiation. All 24 assessed studies received the institutional review board or ethics committee approval. A total of 4 (17%) of the 24 included articles were published before the 2019 study by Birckhead et al [22] that defined the gold standard VR consensus standards criteria.

XSL•FO
RenderX

Table 3. Study adherence to the "gold standard."

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Bani Mohammad and Ahmad [35], 2019								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Not completely met
Description	• A total of 80 patients (average age of 52 years) with breast cancer	Medical and surgical wards of a specialized cancer center in Jordan	Random assignment was done based on flipping a coin; if heads appeared, then the first participant was placed in the intervention group. The rest of the participants were placed in the study groups by the order of meeting the eligibility criteria.	N/A[a]	The VAS[b] was used to measure pain, and the SAI[c] was used to measure anxiety.	One session occurred for 20 to 30 minutes; VR[d] was timed with peak morphine efficacy and the goal was acute pain relief.	• There was a statistically significant improvement in pain scores and anxiety levels, and no MCIDe was mentioned.	None mentioned outside of ethics or IRB[f] approvals from individual hospitals
Burrai et al [36], 2023								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 74 patients (average age of 59 years) with cancer	Regional hospital in Italy	An independent researcher used a random number generation software with simple randomization. A total of 3 numerical codes for the 3 allocation groups were used: 1 for the VR group, 2 for the narrative medicine group, and 3 for the standard care group. To implement the random allocation sequence, sequentially numbered, opaque, sealed envelopes were used.	Owing to the nature of VR and narrative medicine interventions, it was impossible to obtain a blinding between the participants and knowledge of the group, so this study is open-label type.	Anxiety was measured with the STAI-Y1[g]; fatigue was measured with the revised Piper Fatigue Scale; and pain was measured with the VAS scored from 0 to 10, where 0 indicates no pain and 10 indicates the worst possible pain.	VR was administered just after AT[h] began and the immersion duration was 30 minutes.	• Mean anxiety significantly decreased over time for the VR group and narrative medicine group. • There was no significant difference in levels of fatigue, but overall levels of fatigue decreased for participants in the VR intervention. • There was no significant difference in pain levels.	The study was registered in the US National Institutes of Health (NCT05629507).
Chatterjee et al [37], 2022								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 40 patients (average age: not provided, median 77.5 years) with unilateral stroke	Stroke unit at the Countess of Chester Hospital NHS[i] Foundation Trust.	Participants were randomized on a 3:1 allocation basis.	This study was not effective to implement as a double-blind trial, as most patients were able to deduce whether they were a part of the treatment group.	The MoCA[j] was the primary measure used and the NEADL[k], HADS[l], and quality of life (EuroQoL[m]) were the secondary measures used.	A total of 5 days a week for up to 2 weeks before their hospital discharge time that each patient spent in a VR session was very specific to the individual. Time spent in a session would depend on their level of wellness on the day and how tired they were feeling.	• There was a statistically significant improvement in attention and orientation. • There were no statistically significant differences in the primary outcome measures at 3 months. • Secondary outcomes: there was a statistically significant reduction in anxiety scores in the mild to moderate cognitive impairment (MoCA 15-24) treatment group.	The trial was registered in the ISRCTN[n] registry (ISRCTN16608742).
Chirico et al [38], 2020								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Not completely met
Description	• A total of 58 patients (average age of 55 years) with breast cancer • Tumor stages, n (%): 1: 8 (26.6); 2: 13 (43.3); 3: 9 (30)	National Cancer Institute, IRCCS "Fonda-zione G. Pascale," Naples, Italy	Patients were randomly assigned to the VR or music therapy group and were compared with a nonconcurrently recruited CG[o].	Participants and personnel knew which groups were control and intervention groups.	SAI for adults and SV - POMS[p]	Each chemotherapy treatment lasted for 45 to 90 minutes, but the patient used VR for 20 minutes.	• There was a statistically significant reduction in anxiety levels between the pretest and posttest and a significant interaction effect between the time factor and the intervention factor (time ×intervention) on the patient's level of anxiety. Furthermore, the analysis showed a significant main effect for the time factor and significant reduction between the preintervention and the postintervention phases in the following negative mood states: tension, depression, anger, and fatigue.	None mentioned outside of ethics or IRB approvals from individual hospitals.
Groninger et al [39], 2021								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 88 patients (average age of 56 years) with ACC or AHAq stage C or D heart failure	A 912-bed academic hospital located within MWHC[r] in Washington, DC.	Participants were randomized by the study coordinator on a 1:1 basis using a computerized randomized scheme.	The nature of the study prevented participants and the study coordinator from being blinded to assigned interventions.	Self-reported pain scores and the FACIT-Pal 14[s] item scale were used.	There was 1 session for 10 minutes. In distraction therapy research, there is currently no predetermined time threshold for the effect on pain experience; 10 minutes falls within the range of time frames (2 to 15 minutes) that have demonstrated the benefit of using VR for pain management.	• There was a statistically significant improvement in pain score.	The trial was registered on ClinicalTrials.gov (identifier: NCT04572425).
Hsu et al [40], 2022								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Not completely met
Description	• A total of 52 patients (average age of 55 years) with stroke	Department of physical medicine and rehabilitation at a medical center in South Taiwan	Following eligibility screening, patients meeting the inclusion criteria were randomly allocated to conditions using opaque envelopes with computer-generated random numbers that the investigator opened upon receiving a consenting participant.	Eligible patients were randomly allocated until all the available envelopes had been exhausted, resulting in a 1:1:1 ratio in the MT[t], COT[u], or VR-MT group.	The FMA-UE[v] motor coordination score and the MAL[w] were used.	The study duration was 30 minutes twice a week for 9 weeks. VR timing was time-matched to the control group.	• There was a statistically significant improvement in only the FMA-UE motor coordination and the MAL amount of use; no MCID was mentioned.	None mentioned outside of ethics or IRB approvals from individual hospitals
Huang et al [41], 2022								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 30 patients (average age of 58 years) with stroke infarction hemorrhage and ischemic stroke	The VR equipment was installed in a room without external disturbances, and the virtual environment was set in a 6 m^2 physical space in a hospital in southern Taiwan.	A total of 30 patients with chronic stroke were randomized to the VR or COT groups	Clinical assessments were performed within 1 week before and after the interventions by another therapist not involved in training and blinded to the purpose and group allocation.	The FMA-UE and AROM[x] were used.	All participants received 16 sessions of intervention for 60 min per day, 2 to 3 days per week.	• There was a statistically significant time effect for all items of upper limb assessment except FMA-UE motor coordination and speed.	The trial was registered in the WHO[y] International Clinical Trials registry (ChiCTR2100047853).
Huang et al [42], 2024								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 40 patients (average age of 63 years) with ischemic or hemorrhagic stroke	Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, China	Each participant was randomly assigned a code based on computer-generated, permuted block randomization with a block size of 4.	Because of the nature of the intervention, participants and therapists could not be blinded to the allocated treatment. These therapists did not participate in assessments of the outcomes.	The FMA-UE [20] and the BI[z] were used.	All participants received rehabilitation training for 5 days per week for 3 weeks. The first 30 minutes was conventional rehabilitation, followed by 30 minutes of rehabilitation performed in imVR[aa] systems	• There were statistically significant improvements in FMA-UE motor impairment and BI scores.	The trial was registered on ClinicalTrials.gov (identifier: NCT03086889).
Jo et al [43], 2024								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 45 patients (average age of 50 years) with stroke	Hospital in Seoul, Republic of Korea	Randomized controlled trial with the participants randomly allocated into 3 groups using a computer-generated list of numbers at a 1:1:1 ratio: imVR-based 360° MTG[ab], TMTG[ac], and a CG	The study was conducted as an assessor-blinded and randomized controlled trial.	FMA-UE was used, and secondary outcome measurements were MFT[ad] and BBT[ae].	In addition to conventional physical therapy, traditional MT was provided for 30 minutes per session, 3 sessions a week, for 4 weeks	• There was a significant improvement in FMA-UE, MFT, and BBT in 360° MTG and TMTG and changes in 360° MTG were significantly higher compared to TMTG and CG.	The trial was registered on ClinicalTrials.gov (NCT05796843) and received approval from the Sahmyook University IRB (SYU 2023-01-009-001)
Laghlam et al [44], 2021								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 180 patients (average age of 68 years) with cardiac surgery	ICU[af] of the Centre Médico-Chirurgical Ambroise Paré in Neuilly-sur-Seine, France.	Randomization was performed using an external interactive web response system.	Patients were randomly assigned (1:1) in permuted blocks.	The NRSs[ag] for pain and anxiety were used.	One session started at least 5 minutes before the removal of the drains and continued for 10 minutes after the removal of the drains. VR was timed with morphine administration (and the goal was for acute pain relief).	• There were no statistically significant improvements in patient functional outcomes; no MCID was mentioned.	The trial was registered on ClinicalTrials.gov (NCT03956264).
Lee and Kang [45], 2020								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Did not completely meet	Completely met	Not completely met
Description	• A total of 48 patients (average age of 63 years) with cardiovascular disease	Cardiac ICU, Dong-A University Medical Center, Busan, South Korea.	Patients were randomly allocated with a 1:1 random list order using the random allocation software program (version 2.0.0). ICU nurses did not know which group the next participant would be assigned to.	Double blindness was not possible due to the interventional characteristics of the study using VR equipment.	The PSQI[ah] was used.	The study duration was 30 minutes before bedtime (9 PM to 11 PM) on the day of ICU admission. The difference in sleep quality and the effects of intervention during the ICU stay could not be investigated	• There was a statistically significant higher PTSI score in the VR group.	Nothing was mentioned other than the IRB approval.
Mekbib et al [46], 2021								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Not completely met
Description	• A total of 23 patients (average age of 56 years) with stroke	Department of Rehabilitation Medicine at Zhejiang Province People's Hospital (Hangzhou, China).	This clinical trial used a single-blind, randomized, parallel group design. Patients were randomly assigned to either the VR or OT[ai] group using random numbers generated by a computer program.	The group allocation procedure was managed by a physician who was unaware of the study protocol.	The FMA-UE and BI scores were used.	Study duration was 1 hour VR and 1 hour OT per day, 4 days per week for 2 weeks. VR timing was time-matched to the CG.	• There was a statistically significant improvement in FMA-UE and BI scores; no MCID was mentioned.	None mentioned outside of ethics or IRB approvals from individual hospitals and the Helsinki Declaration of Ethical Principles for Medical Research Involving Human Subjects
Menekli et al [47], 2022								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Not completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 139 patients (average age: not provided) with cancer	Malatya Turgut Özal University Oncology Hospital	Computer-assisted simple randomization was used to determine the groups.	There was no blinding for the patients or the researchers throughout the study.	The SAI and VAS scores were used.	The study was conducted at the liberty of the patient throughout the procedure, lasting 4 hours.	• There was a statistically significant improvement in the SAI and VAS scores; no MCID was mentioned.	The trial was registered on ClinicalTrials.gov (NCT05140707).
Ögün et al [48], 2019								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 65 patients (average age of 61 years) with ischemic stroke	Bolu Abant Izzet Baysal University, Physical Therapy and Rehabilitation Hospital	Patients were randomly divided into 2 groups, VR or control, with stratified randomization according to age, sex, and stroke onset, using an online randomization website.	Both patients and outcome assessors were masked, achieved using sham VR therapy with the CG, with the outcome assessor blinded to the groups.	FMA-UE, ARAT[aj], FIM[ak], PASS-IADL[al], and PASS-BADL[am] were used as secondary outcome measurements.	The VR group received VR rehabilitation 3 days a week, on Monday, Wednesday, and Friday, at the same time each day for 6 weeks. Each session lasted approximately 60 minutes and comprised 4 games that lasted 15 minutes each.	• There was a statistically significant improvement in FMA-UE, ARAT, FIM, and PASS scores.	The trial was registered on ClinicalTrials.gov (NCT03135418).
Park et al [49], 2013								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Not completely met
Description	• A total of 16 patients (average age of 48 years) with stroke	Stroke units of a hospital in Seoul, Korea.	The 16 participants were randomly assigned to either the experimental group (n=8) or the CG (n=8) by selection of white or black go stones 1 hour before the start of the pretest.	N/A	The patient's spatiotemporal gait ability, functional gait ability, and increased functional walking ability were measured.	The study duration was 60 minutes per day, 5 days per week for 4 weeks. VR timing was time-matched to the CG.	• There were no statistically significant improvements in patient functional outcomes; no MCID was mentioned.	None mentioned outside of ethics or IRB approvals from individual hospitals
Rousseaux et al [50], 2022								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Not completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 100 patients (average age of 66 years) who underwent cardiac surgery	Liege University Hospitals, Domaine Universitaire du Sart Tilman	Randomization was undertaken in blocks of 5 patients.	Given the nature of the techniques, neither the patients nor the investigators were blinded to the treatment assignment.	The VAS was used.	The study comprised a 20-minute session on the preoperative day (T0 and T1) and a 20-minute session after surgery (T2 and T3). No justification was written in the manuscript for the study duration.	• There was no statistically significant difference in effectiveness between hypnosis and VR	The trial was registered retrospectively on ClinicalTrials.gov (NCT03820700).
Rutkowski et al [51], 2021								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 50 patients (average age of 64 years) with COPDan	A specialist Hospital in Głuchołazy, Poland	Randomization was performed using the Research Randomizer (ratio of 1:1), a web-based service that offers instant random assignment. Sealed envelopes were used for group assignments.	An assessor-blinded parallel group	The HADS, PSQ[ao], 6MWT[ap], and lung function test with FEV1[aq] were used.	The study was conducted once a day, each session for 15 to 30 minutes (depending on the task), 5 times a week for 2 weeks. VR group performed 10 VR therapy sessions of 20 minutes.	• There was statistically significant improvement in the HADS-Aar, HADS-Das, and general HADS.	The trial was registered on ClinicalTrials.gov (NCT04601544).
Shin et al [52], 2023								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 196 patients (average age of 47.5 years) with breast cancer	Academic hospital in Seoul, South Korea.	Randomization was done using an online randomizing tool (www.randomizer. org) by an independent research coordinator.	Physicians were blinded to the study arms and did not interact with the coordinator.	Anxiety levels were measured using the APAIS[at] as the primary end point and the STAI and LASA[au] as secondary end points	The study was for a single time; on the same day of randomization, patients watched the 7- to 8-minute-long VR video.	• There was a statistically significant reduction in patient anxiety immediately and on the primary end point, APAIS, and on the STAI and LASA anxiety scales. Long-term anxiety reduction was observed only when anxiety was measured by LASA.	The protocol was registered on ClinicalTrials.gov (NCT04141943) and approved by the Severance Hospital IRB (4-2019-0795).
Song and Lee [53], 2021								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Not completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 10 patients (average age of 64 years) with stroke with chronic hemiplegia	C Rehabilitation Hospital or G Rehabilitation Hospital in Gwangju city	Randomization was achieved using the opaque sealed envelope method.	N/A	Manual function test and proprioception test	Interventions were performed for 30 minutes a day, 5 times a week, for 4 weeks, for 20 sessions.	• There was a statistically significant improvement in the proprioception test.	None mentioned outside of ethics or IRB approvals from individual hospitals
Spiegel et al [54], 2019								
Met criteria	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 120 patients (average age: not provided) with pain	Cedars-Sinai Medical Center, a large, urban, tertiary care hospital	Patients were randomized on a 1:1 basis between groups using the Microsoft Excel random number generator.	A script that used neutral language regarding both interventions was used. Investigator interactions with the study participant were minimized, relying on nonstudy nursing staff to collect pain scores.	An NRS of patient-reported pain was used.	The study duration was 10 minutes per session, 3 times a day, for 48 hours; 10 minutes was selected to reduce the risk of developing cybersickness. Longer exposure times are associated with a higher risk of cybersickness.	• There was a statistically significant improvement in pain scores 48 hours and 72 hours postintervention.	The trial was registered on ClincialTrials.gov (NCT02887989).
Torres García et al [55], 2023								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 133 patients (average age of 49 years) with breast cancer	OCCU[av] of the CIMA[aw] Campus, Milenium Iradier Medical Center, Barcelona, Spain	Randomization was carried out using a random number list obtained with the Random.org app. The order in which it was generated was unknown to the oncologist.	The oncologist was unaware of the information provided other than the order in which randomization was generated.	HADS, DDE[ax] scale, and the Mini-MAC[ay] coping scale were used.	The study comprised 4 sessions lasting approximately 30 to 45 minutes. The first 3 sessions coincided with the chemotherapy treatment prescribed by the oncologist. The fourth session was conducted 3 months after the third session to check whether the benefits of the previous interventions were maintained.	• There was a statistically significant reduction in anxiety and depression and significant improvements in emotional distress and coping mechanisms, including fighting spirit, cognitive avoidance, and anxious worry.	The study was approved by the Clinical Research Ethics Committee (CEIC) of Hospital CIMA (Barcelona, Spain), a center authorized by the Department of Health (H08621946).
Turrado et al [56], 2021								

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 126 patients (average age of 66 years) with the following condition, n (%): colorectal cancer: 58 (46); right colectomy: 21 (16.7); sigmoidectomy: 20 (15.9); LARaz+TaTMEba: 25 (19.8); transverse colon resection: 2 (1.6)	Third-level Academic Center in the Gastrointestinal Surgery Department in Barcelona, Spain	Patients were randomized using en bloc randomization with random block sizes.	Patients and health care professionals could not be blinded regarding group assignment.	HAD-D, HAD-A, STAI_A/S, and STAI_A/T scores were used.	The patients in the intervention group had unlimited access to VR glasses and to the VR app.	• There was a statistically significant improvement for both the HADS-A and HADS-D subscales and those measuring status and trait in STAI scales in patients who were exposed to the VR simulation experience.	The trial was registered on ClinicalTrials.gov (NCT04058600).
Uslu and Arslan [57], 2023								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met
Description	• A total of 66 patients (average age of 52 years) with breast cancer	Oncology outpatient department of a university hospital in Turkey	This study was conducted using the randomization block method (6:5). Combinations with randomization were enumerated using a web-based random queue generator with permuted random block allocation (block size: 33). The intervention group and CG were determined by lots.	Blinding was not possible due to the nature of the study. With VR, the patients could not be blinded during the intervention because they knew that they were in the intervention group. Randomization was prepared by a statistician, but this was hidden from the researcher performing the study.	The STAI scale and the CFS[bb] were used.	The study was conducted once every 21 days for 4 sessions (1 session per chemotherapy cycle for 4 cycles) for 30 minutes.	• There was a statistically significant decrease in mean postintervention anxiety scores and mean posttest fatigue.	The trial was registered with the US National Institutes of Health Clinical Trials Registry on September 12, 2021 (NCT05168696)
Zhang et al [58], 2023								
Met criteria	Completely met	Completely met	Completely met	Not completely met	Completely met	Completely met	Completely met	Completely met

XSL•FO
RenderX

Study	Patient population	Clinical setting	Control and randomization	Blinding and concealment of allocation	End points	Study duration	Presentation and analysis of results	Reporting the trial
Description	• A total of 60 patients (average age of 34 years) with leukemia	A public hospital in Guangzhou, China	Participants were randomly assigned to the intervention group and CG in a 1:1 ratio by generating 70 codes written on identically sized pieces of paper, which were placed in opaque, sealed, and sequentially numbered envelopes.	Owing to the nature of VR intervention, blinding the participants and the interventionist was impossible.	The SAI scale, CES-D[bc] scale, and FACT-leu[bd] questionnaire were used.	From days 1 to 14, participants experienced 14 different immersive videos. To avoid interruption from common treatment activities, every intervention occurred from 3 PM to 6 PM.	Those in the intervention group demonstrated a significantly greater reduction in anxiety (P<.05) and improvement in quality of life (P=.04). There was no significant difference in depression levels between groups (P=.09), although a decreasing trend was observed in the intervention group.	The protocol was registered retrospectively in the ISRCTN registry (ISRCTN 84842464; June 6, 2022).

[a]N/A: not applicable.
[b]VAS: Visual Analog Scale.
[c]SAI: State Anxiety Inventory.
[d]VR: virtual reality.
[e]MCID: minimal clinically important difference.
[f]IRB: institutional review board.
[g]STAI-Y1: State-Trait Anxiety Inventory-Y1.
[h]AT: antiblastic therapy.
[i]NHS: National Health Service.
[j]MoCA: Montreal Cognitive Assessment.
[k]NEADL: Nottingham extended activities of daily living.
[l]HADS: Hospital Anxiety and Depression Scale.
[m]EuroQOL: European Quality of Life.
[n]ISRCTN: International Standard Randomized Controlled Trial Number.
[o]CG: control group.
[p]SV-POMS: short version of profile of mood states.
[q]AHA or ACC: American Heart Association or American College of Cardiology.
[r]MWHC: MedStar Washington Hospital Center.
[s]FACIT-Pal 14: Functional Assessment in Chronic Illness Therapy–Palliative Care 14-item scale.
[t]MT: mirror therapy.
[u]COT: conventional occupational therapy.
[v]FMA-UE: Fugl-Meyer assessment for upper extremity.
[w]MAL: motor activity log.
[x]AROM: active range of motion.
[y]WHO: World Health Organization.
[z]BI: Barthel Index.
[aa]imVR: immersive virtual reality.
[ab]360° MTG: 360° mirror therapy group.
[ac]TMTG: traditional mirror therapy group.
[ad]MFT: manual function test.
[ae]BBT: box and block test.
[af]ICU: intensive care unit.
[ag]NRS: numeric rating scale.
[ah]PSQI: Pittsburg Sleep Quality Index.
[ai]OT: occupational therapy.
[aj]ARAT: action research arm test.
[ak]FIM: functional independence measure.
[al]PASS-IADL: performance assessment of self-care skills—instrumental activities of daily living.

XSL•FO
RenderX

[am]PASS-BADL: performance assessment of self-care skills—basic activities of daily living.
[an]COPD: chronic obstructive pulmonary disease.
[ao]PSQ: Perception of Stress Questionnaire.
[ap]6MWT: 6-minute walk test.
[aq]FEV1: forced expiratory volume for 1 second.
[ar]HADS-A: Hospital Anxiety and Depression Scale-Anxiety.
[as]HADS-D: Hospital Anxiety and Depression Scale-Depression.
[at]APAIS: Amsterdam Preoperative Anxiety and Information Scale.
[au]LASA: Linear Analog Scale Assessment.
[av]OCCU: Oncology Counseling and Care unit.
[aw]CIMA: Chartered Institute of Management Accountants.
[ax]DDE: emotional discomfort detection.
[ay]Mini-MAC: mini-mental adjustment to cancer.
[az]LAR: low anterior resection.
[ba]TaTME: transanal total mesorectal resection.
[bb]CFS: Cancer Fatigue Scale.
[bc]CES-D: Center for Epidemiological Studies-Depression Scale.
[bd]FACT-leu: functional assessment of cancer therapy-leukemia.

Discussion

Principal Findings

We found 24 RCTs of VR for older adults with serious illness that encompassed diverse conditions, such as stroke, cancer, cardiovascular disease, and COPD. Among these studies, VR has been applied to improve anxiety and pain, as measured by the HADS, State Anxiety Inventory, or NPS, or mobility and strength, as measured by the Fugl-Meyer upper extremity assessment, proprioception, or orientation metrics. In this review, short VR experiences were typically used to improve outcomes related to anxiety and pain, while longer ones were used to improve mobility and strength after stroke. Training, supervision, and assistance provided were described in 5 studies [38,40,43,46,54]. Of the included RCTs, all 24 studies addressed older adults; however, only a few focused entirely on the older adult population: 3 studies (12%) included adults with an average age of 66 to 70 years [44,50,56], and only 1 study (4%) included older adults with an average age of >75 years [37]. From the 24 RCTs included in this systematic review, 7 (29%) studies adhered to all 8 gold standard VR consensus standards criteria, and all studies (100%) adhered to at least 5 criteria.

Although there are emerging bodies of evidence on VR use for acute pain, physical therapy, and mental health, we found relatively few studies that met our criteria, focused on symptomatic and functional aspects of serious illness or their use for older adults [59-61]. Acute pain and pain associated with serious illness are distinct, as suggested by the total pain model for serious illness by Saunders; are multifactorial; and explicitly consider physical, emotional, social, and existential factors in worsening symptoms [62,63]. Function is also impaired to a greater degree in older adults, and the average American lives with 1 to 2 ADL disabilities during the last 2 years of life [64]. Comorbidity is common among older adults, and underlying conditions and treatments can interact to exacerbate symptoms [65]. Owing to this, the same interventions used to manage acute pain may not be efficacious for managing pain associated with serious illness, especially among older adults with serious illnesses.

The practicality and efficacy of VR for older patients may be limited by difficulty using the technology, cost, and limited accessibility in medical settings [66]. For example, an approximately 1- to 2-pound VR headset may be relatively heavy, which is challenging for patients with frailty, as well as those with a serious illness [67]. In many cases, older patients may tend to be unfamiliar with new technologies and, therefore, may require supervision and assistance. This role can be fulfilled with caregivers, friends, or clinicians. In addition, immersive VR can cause motion sickness, which is a concern for older adults with serious illnesses, and are costlier than nonimmersive therapies [68]. Unfortunately, 5 (21%) of the 24 assessed studies lacked descriptions of supervision, training, or assistance provided, which may limit the ability of other researchers or providers to translate the results to their settings [45,48,50,51,56]. Therefore, VR assessments in older adults should include whether and how users were trained and supported, how their clinical condition impacted their ability to use the system, and how their ability to independently use VR was verified.

The extent to which VR treatment efficacy is reliant on media content remains unclear. This is particularly important because VR audio and visual media content varies generally within and between different VR systems. Furthermore, different clinical conditions and intervention goals may lead to the use of specific VR resources, interactive functionality, interactivity, and content. As a result, there are fundamental limitations for researchers and clinicians in understanding the mechanisms and efficacy of VR due to the opaqueness or lack of content descriptions in many publications [69]. In addition, subtleties in achieving high presence or in how perceived reality is portrayed may significantly impact the clinical impact and patient satisfaction. Therefore, better-published specifications of VR media, such as soundscapes, public accessibility, and other features, are required to better evaluate and clinically apply VR to patients [11].

XSL•FO
RenderX

This evaluation is limited by the sparse availability of specific VR content that was used in the previous studies. In addition, the outcomes related to patients' quality of life in the different studies were heterogeneous, and therefore, a quantitative synthesis was not possible. These factors, combined, limited the ability to characterize the effects of specific VR features on patient outcomes. However, this may be a representation of an evolving field with limited reporting guidelines. In addition, this study focused specifically on immersive VR, and future research can further examine the effectiveness of different types of VR. One of the original aims of this study was to compare individual differences in VR effectiveness (eg, by gender, race, and ethnicity); however, this was not possible due to a lack of sufficient detail in included studies. Nonetheless, this systematic review provides researchers and clinicians with a detailed overview of the current state of the literature for older adults with serious illness, as well as an awareness and emphasis on the need to follow standard best practices to enhance VR research rigor. Future directions to build upon this systematic review include evaluating different types of VR media and frequency and duration to target improvements in specific outcomes in patients with serious illness. Future studies may also consider more structured supervision and training for VR use for older adults and determine whether this leads to increased compliance and efficacy of VR to improve outcomes in serious illness. Finally, future research can focus on making VR more accessible to everyone, including older adults [70].

Conclusions

Our systematic review identified a growing body of research evaluating VR apps for older adults with serious illness through RCTs. The literature review illustrates the potential of VR in this population and finds promising but limited extant evidence. These limitations also highlight an important gap in the development and understanding of using VR to improve health and wellness, which are crucial to address for rapidly aging populations.

Acknowledgments

This study would not have been possible without the support of Stanford University School of Medicine and the VA Center for Innovation to Implementation.

Data Availability

The datasets analyzed during this study are publicly available, as detailed in the Methods section.

Authors' Contributions

BM contributed to the project ideation, data collection, data analysis, manuscript writing, and manuscript editing. AK contributed to the project ideation, data collection, data analysis, and manuscript editing. NI contributed to the data collection, data analysis, and manuscript writing. SF, JB, BB, and KG contributed to the project ideation, data analysis, and manuscript editing. CK and TO contributed to manuscript editing. KL contributed to the project ideation, data collection, data analysis, manuscript editing, and overall project oversight. All authors read and approved the final version of this manuscript.

Conflicts of Interest

Author TO contributed to the project while working at the VA, however, is now employed at Microsoft. After his participation in the project, he also recently became an advisor for the VR enabled company RadiusXR.

Multimedia Appendix 1
Search strategy.
[DOCX File , 2920 KB - jmir_v27i1e54452_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v27i1e54452_app2.docx]

Multimedia Appendix 3
Abstraction form.
[DOCX File , 20 KB - jmir_v27i1e54452_app3.docx]

Multimedia Appendix 4
Risk of bias of included studies.
[DOCX File , 3542 KB - jmir_v27i1e54452_app4.docx]

References

XSL•FO
RenderX

1. Sanders JJ, Curtis JR, Tulsky JA. Achieving goal-concordant care: a conceptual model and approach to measuring serious illness communication and its impact. J Palliat Med 2018 Mar;21(S2):S17-S27 [FREE Full text] [doi: 10.1089/jpm.2017.0459] [Medline: 29091522]
2. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
3. Kelley AS. Epidemiology of care for patients with serious illness. J Palliat Med 2013 Jul;16(7):730-733 [FREE Full text] [doi: 10.1089/jpm.2013.9498] [Medline: 23718871]
4. Kelley AS, Bollens-Lund E. Identifying the population with serious illness: the "denominator" challenge. J Palliat Med 2018 Mar;21(S2):S7-16 [FREE Full text] [doi: 10.1089/jpm.2017.0548] [Medline: 29125784]
5. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. BMJ 2005 Apr 30;330(7498):1007-1011 [FREE Full text] [doi: 10.1136/bmj.330.7498.1007] [Medline: 15860828]
6. Ahmad FB, Anderson RN. The leading causes of death in the US for 2020. JAMA 2021 May 11;325(18):1829-1830 [FREE Full text] [doi: 10.1001/jama.2021.5469] [Medline: 33787821]
7. May P, Normand C, Cassel JB, Del Fabbro E, Fine RL, Menz R, et al. Economics of palliative care for hospitalized adults with serious illness: a meta-analysis. JAMA Intern Med 2018 Jun 01;178(6):820-829 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0750] [Medline: 29710177]
8. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas 2018;9:353-367 [FREE Full text] [doi: 10.2147/PROM.S156279] [Medline: 30464666]
9. Stone MJ. Goals of care at the end of life. Proc (Bayl Univ Med Cent) 2001 Apr 11;14(2):134-137 [FREE Full text] [doi: 10.1080/08998280.2001.11927748] [Medline: 16369602]
10. Hayes RD, Chang CK, Fernandes AC, Begum A, To D, Broadbent M, et al. Functional status and all-cause mortality in serious mental illness. PLoS One 2012 Sep 6;7(9):e44613 [FREE Full text] [doi: 10.1371/journal.pone.0044613] [Medline: 22970266]
11. Javaid M, Haleem A. Virtual reality applications toward medical field. Clin Epidemiol Global Health 2020 Jun;8(2):600-605. [doi: 10.1016/j.cegh.2019.12.010]
12. Park B, Kim Y, Park J, Choi H, Kim SE, Ryu H, et al. Integrating biomarkers from virtual reality and magnetic resonance imaging for the early detection of mild cognitive impairment using a multimodal learning approach: validation study. J Med Internet Res 2024 Apr 17;26:e54538 [FREE Full text] [doi: 10.2196/54538] [Medline: 38631021]
13. Elavsky S, Knapova L, Janiš K, Cimler R, Kuhnova J, Cernicky T. Multiple perspectives on the adoption of SMART technologies for improving care of older people: mixed methods study. J Med Internet Res 2024 Feb 07;26:e45492 [FREE Full text] [doi: 10.2196/45492] [Medline: 38324345]
14. Sheng Y, Bond R, Jaiswal R, Dinsmore J, Doyle J. Augmenting k-means clustering with qualitative data to discover the engagement patterns of older adults with multimorbidity when using digital health technologies: proof-of-concept trial. J Med Internet Res 2024 Mar 28;26:e46287 [FREE Full text] [doi: 10.2196/46287] [Medline: 38546724]
15. Kardong-Edgren S, Farra SL, Alinier G, Young HM. A call to unify definitions of virtual reality. Clin Simul Nurs 2019 Jun;31:28-34. [doi: 10.1016/j.ecns.2019.02.006]
16. Osborne TF, Arreola DM, Veigulis ZP, Morley C, Choudhry O, Lan W, et al. Augmented reality demonstration survey results from a veteran affairs medical center. Fed Pract 2023 Mar;40(3):90-97 [FREE Full text] [doi: 10.12788/fp.0364] [Medline: 37228426]
17. Rawlins CR, Veigulis Z, Hebert C, Curtin C, Osborne TF. Effect of immersive virtual reality on pain and anxiety at a veterans affairs health care facility. Front Virtual Real 2021 Oct 5;2:719681. [doi: 10.3389/frvir.2021.719681]
18. McCullough M, Osborne TF, Rawlins C, Reitz RJ3, Fox PM, Curtin C. The impact of virtual reality on the patients and providers experience in wide-awake, local-only hand surgery. J Hand Surg Glob Online 2023 May;5(3):290-293 [FREE Full text] [doi: 10.1016/j.jhsg.2023.01.014] [Medline: 37323984]
19. Corregidor-Sánchez AI, Segura-Fragoso A, Rodríguez-Hernández M, Jiménez-Rojas C, Polonio-López B, Criado-Álvarez JJ. Effectiveness of virtual reality technology on functional mobility of older adults: systematic review and meta-analysis. Age Ageing 2021 Feb 26;50(2):370-379. [doi: 10.1093/ageing/afaa197] [Medline: 33068106]
20. Mo J, Vickerstaff V, Minton O, Tavabie S, Taubert M, Stone P, et al. How effective is virtual reality technology in palliative care? A systematic review and meta-analysis. Palliat Med 2022 Jul 30;36(7):1047-1058 [FREE Full text] [doi: 10.1177/02692163221099584] [Medline: 35635018]
21. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 01;22(6):e17331 [FREE Full text] [doi: 10.2196/17331] [Medline: 32478662]
22. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]

XSL•FO
RenderX

23. Välimäki M, Hätönen HM, Lahti ME, Kurki M, Hottinen A, Metsäranta K, et al. Virtual reality for treatment compliance for people with serious mental illness. Cochrane Database Syst Rev 2014 Oct 08;2014(10):CD009928 [FREE Full text] [doi: 10.1002/14651858.CD009928.pub2] [Medline: 25300174]
24. Ose SO, Færevik H, Kaasbøll J, Lindgren M, Thaulow K, Antonsen S, et al. Exploring the potential for use of virtual reality technology in the treatment of severe mental illness among adults in mid-Norway: collaborative research between clinicians and researchers. JMIR Form Res 2019 Jun 10;3(2):e13633 [FREE Full text] [doi: 10.2196/13633] [Medline: 31199315]
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
27. Bailenson J, Fereydooni S, Giannitrapani K, Kraft A, Lorenz K, Maheta B, et al. Using virtual reality to improve quality of life among seriously ill adults: a systematic review of randomized control trials. PROSPERO 2022 CRD42022346178. 2022. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022346178 [accessed 2025-01-05]
28. Hartmann KE, Matchar DB, Chang SM. Chapter 6: Assessing applicability of medical test studies in systematic reviews. In: Chang SM, Matchar DB, Smetana GW, Umscheid CA, editors. Methods Guide for Medical Test Reviews [Internet]. Rockville, MD: Agency for Healthcare Research and Quality (US); Jun 2012.
29. Deyo RA. Measuring functional outcomes in therapeutic trials for chronic disease. Control Clin Trials 1984 Sep;5(3):223-240. [doi: 10.1016/0197-2456(84)90026-6] [Medline: 6488807]
30. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144 [FREE Full text] [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
31. Fung CH, Hays RD. Prospects and challenges in using patient-reported outcomes in clinical practice. Qual Life Res 2008 Dec 18;17(10):1297-1302 [FREE Full text] [doi: 10.1007/s11136-008-9379-5] [Medline: 18709564]
32. Van der Mierden S, Tsaioun K, Bleich A, Leenaars CH. Software tools for literature screening in systematic reviews in biomedical research. ALTEX 2019;36(3):508-517 [FREE Full text] [doi: 10.14573/altex.1902131] [Medline: 31113000]
33. Lundh A, Gøtzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008 Apr 21;8:22 [FREE Full text] [doi: 10.1186/1471-2288-8-22] [Medline: 18426565]
34. Henderson LK, Craig JC, Willis NS, Tovey D, Webster AC. How to write a Cochrane systematic review. Nephrology (Carlton) 2010 Sep;15(6):617-624. [doi: 10.1111/j.1440-1797.2010.01380.x] [Medline: 20883282]
35. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. Palliat Support Care 2018 Sep 10;17(1):29-34. [doi: 10.1017/s1478951518000639]
36. Burrai F, Ortu S, Marinucci M, De Marinis MG, Piredda M. Effectiveness of immersive virtual reality in people with cancer undergoing antiblastic therapy: a randomized controlled trial. Semin Oncol Nurs 2023 Aug;39(4):151470 [FREE Full text] [doi: 10.1016/j.soncn.2023.151470] [Medline: 37455151]
37. Chatterjee K, Buchanan A, Cottrell K, Hughes S, Day TW, John NW. Immersive virtual reality for the cognitive rehabilitation of stroke survivors. IEEE Trans Neural Syst Rehabil Eng 2022;30:719-728. [doi: 10.1109/tnsre.2022.3158731]
38. Chirico A, Maiorano P, Indovina P, Milanese C, Giordano GG, Alivernini F, et al. Virtual reality and music therapy as distraction interventions to alleviate anxiety and improve mood states in breast cancer patients during chemotherapy. J Cell Physiol 2020 Jun 20;235(6):5353-5362. [doi: 10.1002/jcp.29422] [Medline: 31957873]
39. Groninger H, Stewart D, Fisher JM, Tefera E, Cowgill J, Mete M. Virtual reality for pain management in advanced heart failure: a randomized controlled study. Palliat Med 2021 Dec 25;35(10):2008-2016. [doi: 10.1177/02692163211041273] [Medline: 34431398]
40. Hsu HY, Kuo LC, Lin YC, Su FC, Yang TH, Lin CW. Effects of a virtual reality-based mirror therapy program on improving sensorimotor function of hands in chronic stroke patients: a randomized controlled trial. Neurorehabil Neural Repair 2022 Jun 28;36(6):335-345. [doi: 10.1177/15459683221081430] [Medline: 35341360]
41. Huang CY, Chiang WC, Yeh YC, Fan SC, Yang WH, Kuo HC, et al. Effects of virtual reality-based motor control training on inflammation, oxidative stress, neuroplasticity and upper limb motor function in patients with chronic stroke: a randomized controlled trial. BMC Neurol 2022 Jan 11;22(1):21 [FREE Full text] [doi: 10.1186/s12883-021-02547-4] [Medline: 35016629]
42. Huang Q, Jiang X, Jin Y, Wu B, Vigotsky AD, Fan L, et al. Immersive virtual reality-based rehabilitation for subacute stroke: a randomized controlled trial. J Neurol 2024 Mar 10;271(3):1256-1266 [FREE Full text] [doi: 10.1007/s00415-023-12060-y] [Medline: 37947856]
43. Jo S, Jang H, Kim H, Song C. 360° immersive virtual reality-based mirror therapy for upper extremity function and satisfaction among stroke patients: a randomized controlled trial. Eur J Phys Rehabil Med 2024 Apr;60(2):207-215. [doi: 10.23736/s1973-9087.24.08275-3]
44. Laghlam D, Naudin C, Coroyer L, Aidan V, Malvy J, Rahoual G, et al. Virtual reality vs. Kalinox® for management of pain in intensive care unit after cardiac surgery: a randomized study. Ann Intensive Care 2021 May 13;11(1):74 [FREE Full text] [doi: 10.1186/s13613-021-00866-w] [Medline: 33983498]

XSL•FO
RenderX

45. Lee SY, Kang J. Effect of virtual reality meditation on sleep quality of intensive care unit patients: a randomised controlled trial. Intensive Crit Care Nurs 2020 Aug;59:102849. [doi: 10.1016/j.iccn.2020.102849] [Medline: 32241625]
46. Mekbib DB, Debeli DK, Zhang L, Fang S, Shao Y, Yang W, et al. A novel fully immersive virtual reality environment for upper extremity rehabilitation in patients with stroke. Ann N Y Acad Sci 2021 Jun 14;1493(1):75-89. [doi: 10.1111/nyas.14554] [Medline: 33442915]
47. Menekli T, Yaprak B, Doğan R. The effect of virtual reality distraction intervention on pain, anxiety, and vital signs of oncology patients undergoing port catheter implantation: a randomized controlled study. Pain Manag Nurs 2022 Oct;23(5):585-590. [doi: 10.1016/j.pmn.2022.03.004] [Medline: 35367144]
48. Ögün MN, Kurul R, Yaşar MF, Turkoglu SA, Avci S, Yildiz N. Effect of leap motion-based 3D immersive virtual reality usage on upper extremity function in ischemic stroke patients. Arq Neuro Psiquiatr 2019 Oct;77(10):681-688. [doi: 10.1590/0004-282x20190129]
49. Park YH, Lee CH, Lee BH. Clinical usefulness of the virtual reality-based postural control training on the gait ability in patients with stroke. J Exerc Rehabil 2013 Oct 31;9(5):489-494 [FREE Full text] [doi: 10.12965/jer.130066] [Medline: 24282810]
50. Rousseaux F, Dardenne N, Massion PB, Ledoux D, Bicego A, Donneau AF, et al. Virtual reality and hypnosis for anxiety and pain management in intensive care units: a prospective randomised trial among cardiac surgery patients. Eur J Anaesthesiol 2022 Jan 01;39(1):58-66 [FREE Full text] [doi: 10.1097/EJA.0000000000001633] [Medline: 34783683]
51. Rutkowski S, Szczegielniak J, Szczepańska-Gieracha J. Evaluation of the efficacy of immersive virtual reality therapy as a method supporting pulmonary rehabilitation: a randomized controlled trial. J Clin Med 2021 Jan 18;10(2):352 [FREE Full text] [doi: 10.3390/jcm10020352] [Medline: 33477733]
52. Shin J, Chang JS, Kim JS, An JY, Chung SY, Yoon SY, et al. An investigation of the effect of virtual reality on alleviating anxiety in patients with breast cancer undergoing radiation therapy: a randomized controlled trial. Int J Radiat Oncol Biol Phys 2023 Dec 01;117(5):1191-1199. [doi: 10.1016/j.ijrobp.2023.06.275] [Medline: 37451473]
53. Song YH, Lee HM. Effect of immersive virtual reality-based bilateral arm training in patients with chronic stroke. Brain Sci 2021 Aug 03;11(8):1032 [FREE Full text] [doi: 10.3390/brainsci11081032] [Medline: 34439651]
54. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: a randomized comparative effectiveness trial. PLoS One 2019 Aug 14;14(8):e0219115 [FREE Full text] [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]
55. Torres García A, Morcillo Serra C, Argilés Huguet M, González Gardó L, Abad Esteve A, Ramos Quiroga JA. Efficacy of a virtual reality intervention for reducing anxiety, depression, and increasing disease coping in patients with breast cancer before their first chemotherapy dose. Cogn Ther Res 2023 Oct 27;48(3):451-465. [doi: 10.1007/s10608-023-10440-2]
56. Turrado V, Guzmán Y, Jiménez-Lillo J, Villegas E, de Lacy FB, Blanch J, et al. Exposure to virtual reality as a tool to reduce peri-operative anxiety in patients undergoing colorectal cancer surgery: a single-center prospective randomized clinical trial. Surg Endosc 2021 Jul;35(7):4042-4047. [doi: 10.1007/s00464-021-08407-z] [Medline: 33683433]
57. Uslu A, Arslan S. The effect of using virtual reality glasses on anxiety and fatigue in women with breast cancer receiving adjuvant chemotherapy: a pretest-posttest randomized controlled study. Semin Oncol Nurs 2023 Oct;39(5):151503. [doi: 10.1016/j.soncn.2023.151503] [Medline: 37741768]
58. Zhang B, Jin X, Kuang X, Shen B, Qiu D, Peng J, et al. Effects of a virtual reality-based meditation intervention on anxiety and depression among patients with acute leukemia during induction chemotherapy: a randomized controlled trial. Cancer Nurs 2023 Jan 23;47(3):E159-E167. [doi: 10.1097/NCC.0000000000001206] [Medline: 36693237]
59. Feng H, Li C, Liu J, Wang L, Ma J, Li G, et al. Virtual reality rehabilitation versus conventional physical therapy for improving balance and gait in Parkinson's disease patients: a randomized controlled trial. Med Sci Monit 2019 Jun 05;25:4186-4192. [doi: 10.12659/msm.916455]
60. Smith V, Warty RR, Sursas JA, Payne O, Nair A, Krishnan S, et al. The effectiveness of virtual reality in managing acute pain and anxiety for medical inpatients: systematic review. J Med Internet Res 2020 Nov 02;22(11):e17980 [FREE Full text] [doi: 10.2196/17980] [Medline: 33136055]
61. Baghaei N, Chitale V, Hlasnik A, Stemmet L, Liang HN, Porter R. Virtual reality for supporting the treatment of depression and anxiety: scoping review. JMIR Ment Health 2021 Sep 23;8(9):e29681 [FREE Full text] [doi: 10.2196/29681] [Medline: 34554097]
62. Ong CK, Forbes D. Embracing Cicely Saunders's concept of total pain. BMJ 2005 Sep 10;331(7516):576-577 [FREE Full text] [doi: 10.1136/bmj.331.7516.576-d] [Medline: 16150775]
63. Clark D. 'Total pain', disciplinary power and the body in the work of Cicely Saunders, 1958-1967. Soc Sci Med 1999 Sep;49(6):727-736. [doi: 10.1016/s0277-9536(99)00098-2] [Medline: 10459885]
64. Smith AK, Walter LC, Miao Y, Boscardin WJ, Covinsky KE. Disability during the last two years of life. JAMA Intern Med 2013 Sep 09;173(16):1506-1513 [FREE Full text] [doi: 10.1001/jamainternmed.2013.8738] [Medline: 23836120]
65. Desbiens NA, Wu AW, Broste SK, Wenger NS, Connors AFJ, Lynn J, et al. Pain and satisfaction with pain control in seriously ill hospitalized adults: findings from the SUPPORT research investigations. For the SUPPORT investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment. Crit Care Med 1996 Dec;24(12):1953-1961. [doi: 10.1097/00003246-199612000-00005] [Medline: 8968261]

XSL•FO
RenderX

66. Baker S, Waycott J, Robertson E, Carrasco R, Neves BB, Hampson R, et al. Evaluating the use of interactive virtual reality technology with older adults living in residential aged care. Inf Process Manag 2020 May;57(3):102105. [doi: 10.1016/j.ipm.2019.102105]
67. Lee LN, Kim MJ, Hwang WJ. Potential of augmented reality and virtual reality technologies to promote wellbeing in older adults. Appl Sci 2019 Aug 30;9(17):3556-3847 [FREE Full text] [doi: 10.3390/app9173556]
68. Lee J, Kim M, Kim J. A study on immersion and VR sickness in walking interaction for immersive virtual reality applications. Symmetry 2017 May 22;9(5):78. [doi: 10.3390/sym9050078]
69. Mazurek J, Kiper P, Cieślik B, Rutkowski S, Mehlich K, Turolla A, et al. Virtual reality in medicine: a brief overview and future research directions. Hum Mov 2019 May 27;20(3):16-22. [doi: 10.5114/hm.2019.83529]
70. Peck TC, McMullen KA, Quarles J. DiVRsify: break the cycle and develop VR for everyone. IEEE Comput Graph Appl 2021 Nov 1;41(6):133-142. [doi: 10.1109/mcg.2021.3113455]

Abbreviations

ADL: activity of daily living
COPD: chronic obstructive pulmonary disease
HADS: Hospital Anxiety and Depression Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
VR: virtual reality

Edited by T de Azevedo Cardoso; submitted 09.11.23; peer-reviewed by A AL-Asadi, SQ Yoong, A Turco; comments to author 30.04.24; revised version received 09.07.24; accepted 25.11.24; published 26.02.25.

Please cite as:
Maheta B, Kraft A, Interrante N, Fereydooni S, Bailenson J, Beams B, Keny , Osborne T, Giannitrapani K, Lorenz K
Using Virtual Reality to Improve Outcomes Related to Quality of Life Among Older Adults With Serious Illnesses: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2025;27:e54452
URL: https://www.jmir.org/2025/1/e54452
doi:10.2196/54452
PMID:

©Bhagvat Maheta, Alexandra Kraft, Nickolas Interrante, Soraya Fereydooni, Jeremy Bailenson, Brian Beams, Christina Keny, Thomas Osborne, Karleen Giannitrapani, Karl Lorenz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Digital Diabetes Management Technology on Blood Glucose in Patients With Type 2 Diabetes at Home: Systematic Review and Meta-Analysis

Yuping Xiao[1], MA; Zhenzhen Wang[1], MA; Lintao Zhang[2], MA; Nina Xie[1], MA; Fangyao Chen[3], PhD; Zihao Song[4], MA; Sha Zhao[5], MA

[1]School of Nursing, Shaanxi University of Chinese Medicine, Xianyang, China

[2]Acupuncture and Moxibustion Department, Affiliated Hospital of Shaanxi University of Chinese Medicine, Xianyang, China

[3]School of Public Health, Xi'an Jiaotong University, Xi'an, China

[4]Department of Clinical Medicine of Traditional Chinese and Western Medicine, First School of Clinical Medicine, Shaanxi University of Chinese Medicine, Xianyang, China

[5]Xiangya School of Nursing, Central South University, Changsha, China

Corresponding Author:
Zhenzhen Wang, MA
School of Nursing
Shaanxi University of Chinese Medicine
1, Middle Section of Century Avenue
Xianyang, 712046
China
Phone: 86 15291011163
Email: wangzz2005@126.com

Abstract

Background: Patients with type 2 diabetes mellitus (T2DM) face elevated morbidity, mortality, and care costs. Digital self-monitoring of blood glucose (SMBG) can automatically upload data to apps, share the data with health care providers, reduce errors, and aid long-term diabetes management.

Objective: This study aimed to assess the effectiveness of digital diabetes management techniques based on digital SMBG on blood glucose in patients with T2DM at home.

Methods: A systematic search was conducted in PubMed, Embase, Web of Science, China National Knowledge Infrastructure, Wanfang, China Biomedical Literature Database, and Cochrane Library for articles published from the establishment of each database to December 25, 2023. Data were extracted independently by 2 researchers (YX and NX), and the risk of bias in individual trials was rated using the Cochrane risk-of-bias tool. A meta-analysis was conducted using RevMan 5.3 (Cochrane).

Results: Twelve studies were included, involving 1669 participants. The meta-analysis found that in the digital diabetes management group, hemoglobin A_{1c} (mean difference [MD] –0.52%, 95% CI –0.63% to –0.42%; *P*<.001), fasting blood sugar (MD –0.42, 95% CI –0.65 to –0.19 mmol/L; *P*<.001), 2-hour postprandial blood sugar (MD –0.64, 95% CI –0.97 to –0.32 mmol/L; *P*<.001), and BMI (MD –1.55, 95% CI –2.92 to –0.17 kg/m^2; *P*=.03) were each improved compared to the control group.

Conclusions: Digital diabetes management has been shown to effectively improve blood glucose levels and BMI in individuals with T2DM in home settings. A key feature of successful digital health interventions is the frequent SMBG by patients, supported by dedicated health care professionals who provide timely, personalized, and responsive guidance.

Trial Registration: PROSPERO CRD42024560431; https://tinyurl.com/yfam3nms

(J Med Internet Res 2025;27:e66441) doi:10.2196/66441

KEYWORDS

digital diabetes management; type 2 diabetes; patients; home monitoring; blood glucose management; meta-analysis; PRISMA

XSL•FO
RenderX

Introduction

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder primarily characterized by persistent hyperglycemia due to insulin resistance or inadequate insulin secretion [1]. The International Diabetes Federation [2] reported that the global prevalence of diabetes reached 533 million cases in 2022, with projections indicating a surge to 783 million by 2045, of which over 90% are expected to be T2DM cases [3]. This condition often manifests insidiously, progressing over an extended period, with chronic hyperglycemia as its hallmark clinical feature [4]. Poorly controlled blood glucose levels over time can inflict severe damage on vital organs and tissues, including the heart, brain, and kidneys [5], leading to debilitating complications that significantly compromise patients' health and quality of life [6].

T2DM is a lifelong condition, and short-term treatment is insufficient to achieve effective blood sugar control. Patients require lifelong management and a long-term, standardized approach to home care [7,8]. International scholars universally agree on the importance of home management for individuals with T2DM [9]. Research by Yan et al [10] has highlighted the high demand among patients with diabetes for home blood glucose monitoring, a finding consistent with the observations by Simsek et al [11] regarding chronic disease management in older adult populations. An increasing number of patients now seek personalized, home-based guidance from health care professionals (HCPs), underscoring the pivotal role of the home environment in diabetes rehabilitation [12,13]. Moreover, studies demonstrate that home blood glucose self-monitoring effectively improves glycemic control and reduces the risk of long-term complications [14].

Traditionally, home blood glucose management has relied on clinician-led, face-to-face consultations. However, such offline consultations are time-consuming for HCPs and impose economic burdens on patients. In addition, these short-term interactions often fail to adequately support patients' long-term self-management efforts, rendering them less effective [15].

Mobile internet technology has brought about a revolution in diabetes management by introducing digital platforms designed for home-based patients with T2DM. Digital diabetes management entails using advanced tools to gather self-monitoring of blood glucose (SMBG) data, which is subsequently sent to HCPs for analysis and feedback. This technology simplifies blood glucose monitoring and boosts patient-provider communication [16]. Research indicates that it allows HCPs to assess data in real time, reducing the necessity for outpatient visits and easing the burden on both health care facilities and patients [17].

Despite its benefits, the manual entry of SMBG data during home care remains time-intensive, particularly for older adult patients, and is prone to errors, including transcription inaccuracies or intentional misreporting [18,19]. In addition, HCPs often face challenges in interpreting SMBG data to identify lifestyle or treatment issues. Recent advancements in SMBG devices now allow automatic uploads of blood glucose measurements to secure digital health applications accessible to HCPs [20-22]. Real-time monitoring of this data enables timely interventions when abnormal readings are detected, thereby facilitating more accurate management of hypoglycemia or hyperglycemia. Furthermore, this technology supports long-term diabetes care by informing adjustments to insulin, diet, and exercise regimens [23].

Currently, no meta-analysis has systematically assessed the effectiveness of digital SMBG-based diabetes management technologies. Evaluating the impact of these technologies on blood glucose control in home settings is essential for understanding their utility in managing T2DM. This study aims to synthesize global research findings through meta-analysis, examining the effects of digital SMBG-based diabetes management on glycemic control in home-based patients with T2DM. The results are intended to provide evidence to support the broader application of digital SMBG technologies in home care environments.

Methods

Overview

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines published in 2020 [24], and a search protocol retrospectively registered on July 5, 2024 on PROSPERO (CRD42024560431) formed the basis of the review process.

Search Strategy

We searched PubMed, Web of Science, Embase, Cochrane Library, China National Knowledge Infrastructure, Wanfang, and the Chinese Biomedical Literature Database using the terms "Diabetes Mellitus, Type 2," "Internet," and "RCT" from inception until December 25, 2023. The search strategy used Boolean logic to combine medical subject headings and text word searches. We also screened references from relevant previous systematic reviews as additional sources, and the languages were limited to Chinese and English. The complete search strategy can be found in Multimedia Appendix 1.

Eligibility Criteria

Inclusion criteria were formulated based on the populations, interventions, comparisons, outcomes, and study design (PICOS) framework:

1. Populations: Participants with T2DM
2. Interventions: Patients used a digital blood glucose management system that allowed for the automatic upload of SMBG data from home to a secure platform via Bluetooth or other wireless technologies, and HCPs provided timely and responsive personalized guidance
3. Comparisons: Traditional home blood glucose monitoring methods (including routine outpatient follow-up, blood glucose monitoring logs, etc)
4. Outcomes: Hemoglobin A_{1c} (HbA_{1c}), fasting blood glucose (FBG), 2-hour postprandial blood glucose (PBG), and BMI. HbA_{1c} is the gold standard for long-term blood glucose control, reflecting the average blood glucose levels over the past 2 to 3 months. FBG can present information on blood glucose in the fasting state, which helps to understand

XSL•FO
RenderX

basic glucose metabolism. Measuring the PBG can evaluate the body's ability to handle the glucose load after a meal and provide insights into the PBG regulation. BMI is a practical tool for assessing overall body fat, which is associated with health risks caused by obesity and metabolic disorders, and the included studies reported at least one of these findings.

5. Study design: Randomized controlled trials

Exclusion criteria included (1) studies involving patients with gestational diabetes or type 1 diabetes, (2) review or discussion papers, and (3) studies with insufficient data.

Study Selection

Two reviewers (YX and NX) independently performed the screening and selection of studies. Disagreements were resolved through discussion or consultation with a third reviewer (ZW). First, we imported the retrieved literature into the NoteExpress software (Beijing Aiqinhai Software Company) and used it to identify and remove duplicate literature. Then, we evaluated whether the titles and abstracts were eligible. In addition, clearly incompatible studies were excluded after screening the titles and abstracts. Finally, full texts of qualified studies that strictly adhered to the inclusion and exclusion criteria [25] were screened to determine their final eligibility for the review.

Data Extraction

Data extraction from the studies included was performed by 2 independent reviewers. Characteristics of the studies (author, year, country, and sample size) and intervention characteristics (equipment used by the intervention, intervening measures, control measures, duration, and outcomes) were included.

Risk-of-Bias Assessment

The quality of the included studies was assessed using the Cochrane risk-of-bias tool, administered by 2 independent assessors. This tool evaluates the quality of study methodology across seven core dimensions: (1) the process of randomization sequence generation, (2) the concealment of allocation, (3) blinding of participants and researchers, (4) blinding during outcome assessment, (5) completeness of outcome data and handling of missing data, (6) selective reporting of study results, and (7) the potential for other sources of bias. The risk of bias for each domain was categorized into 3 levels: low, unclear, or high. Disagreements between the assessors were resolved through discussion, or, if necessary, consultation with a senior reviewer (ZW).

Data Analysis

The meta-analysis was conducted using RevMan 5.3 [26], a desktop version of Review Manager software used for Cochrane intervention and flexible reviews. For continuous outcomes, we calculated the mean difference (MD) as a metric; while for dichotomous variables, the risk ratio was used and visualized by forest plots. In each analysis, I^2 was used to measure statistical heterogeneity among studies. According to the values of P and I^2, the fixed-effect model ($P>.10$, $I^2<50\%$) or random-effects model ($0<P<.10$, $I^2\geq50\%$) were selected [27].

Results

Results of Literature Search

A total of 5177 papers were retrieved from 7 databases. After removing 3460 duplicate records, 1360 articles were deleted due to unqualified titles and abstracts. A more detailed full-text screening was conducted on the remaining 357 articles, and 345 of them were excluded for not meeting the established criteria. As a result, the final systematic review included 12 meta-analysis studies [28-39]. The detailed research screening process and results are shown in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram: papers included and excluded in this review. RCT: randomized controlled trial; T2DM: type 2 diabetes mellitus.

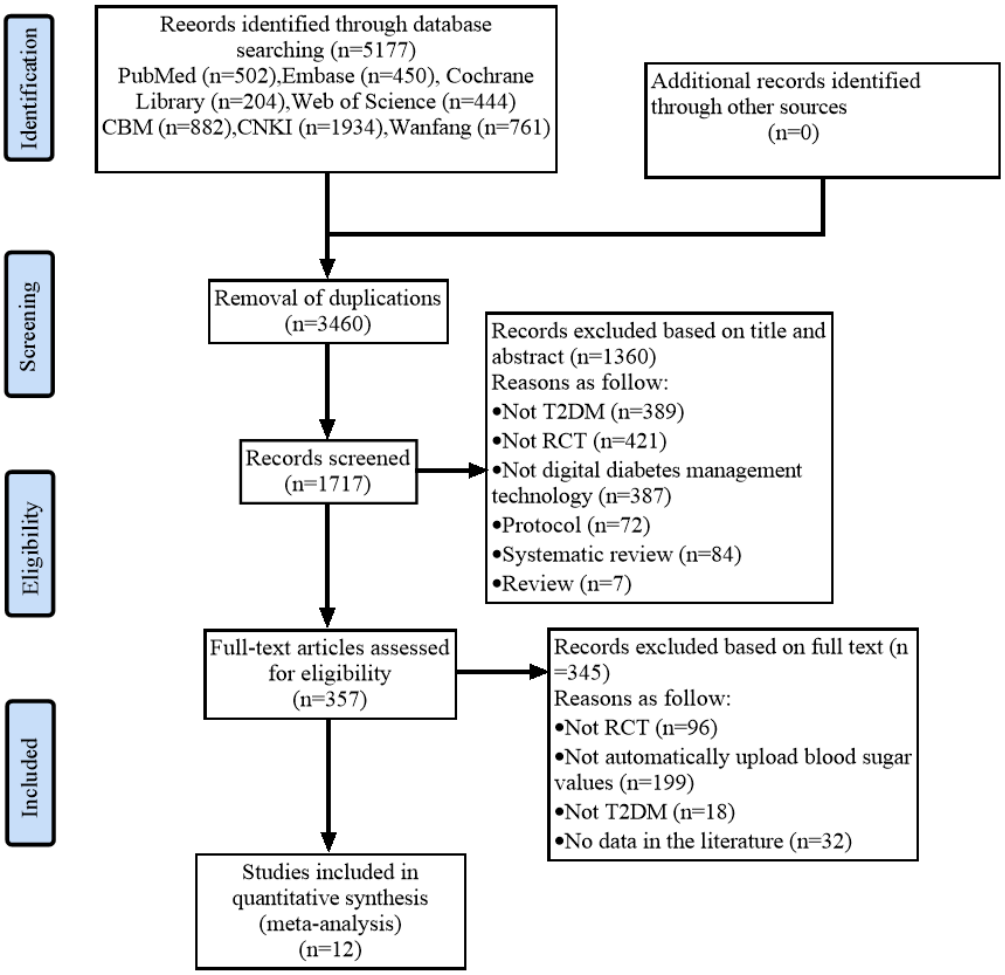

Study Characteristics

The research includes trials conducted in 5 countries: China [28-34,36], the United States [35], India [37], the United Kingdom [38], and Italy [39]. These trials were published between 2015 and 2023. The sample size ranged from 64 to 285 participants, and the intervention duration varied from 1 to 12 months. Among the 12 studies, 6 used Bluetooth technology to achieve automatic upload of blood glucose values [28,29,32,34,38,39], one study adopted an implantable blood glucose sensor for automatic data transmission [33], and another 5 studies used local area network technology to automatically upload blood glucose measurement values [30,31,35-37]. Multimedia Appendix 2 [28-39] presents the author, year, country, sample size, equipment used by the intervention, intervening measure, control measure, duration, and outcomes.

Risk of Bias

Based on the Cochrane criteria, a risk-of-bias assessment is presented in Figures 2 and 3 [28-39]. A total of 12 studies were included, all of which were assigned a literature quality grade of B [28-39]. Randomized methods were reported in detail in 9 (75%) of the 12 studies [28-30,34-39]. The remaining 3 studies did not provide detailed randomization methods. In terms of allocation concealment, 5 studies explicitly mentioned the methods used [30,32,35-37]. In all studies, participants were not blinded due to the specificity of the intervention method.

XSL•FO
RenderX

Figure 2. Risk-of-bias analysis of each included study.

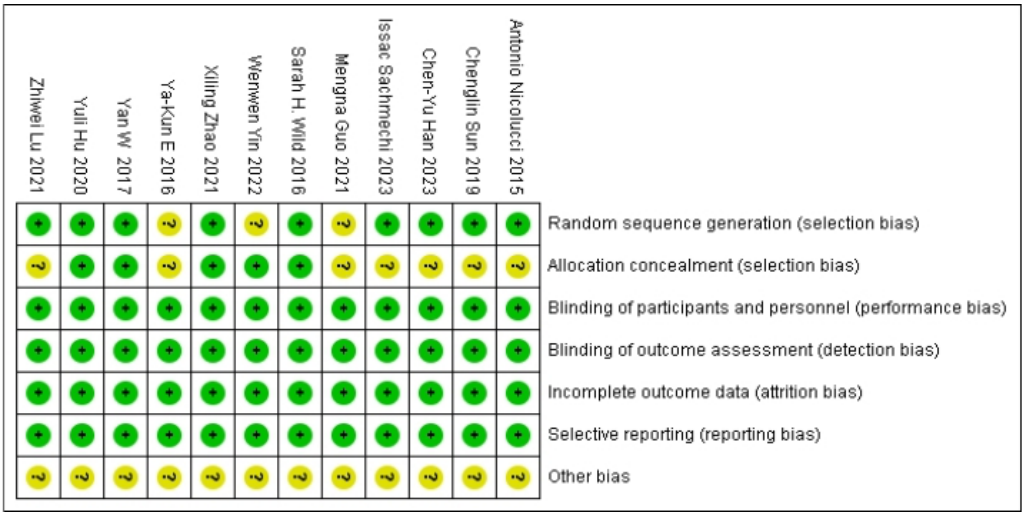

Figure 3. Overall risk-of-bias analysis of included studies.

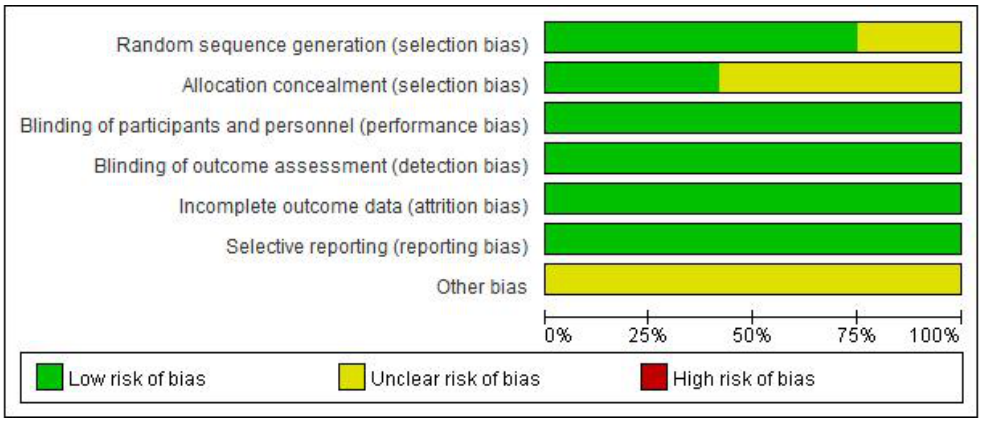

Results of the Meta-Analysis

Effects on HbA_{1c}

Twelve studies reported the effectiveness of digital diabetes management in improving HbA_{1c} [28-39]. The meta-analysis results showed a significant improvement in HbA_{1c} (n=1669; MD −0.52%, 95% CI −0.63 to −0.42; *z* value 9.67, *P*<.001; I^2=34%; fixed-effects model) compared to the control group (refer to Figure 4 [28-39]).

Figure 4. Forest plot of hemoglobin A_{1c} (HbA_{1c}).

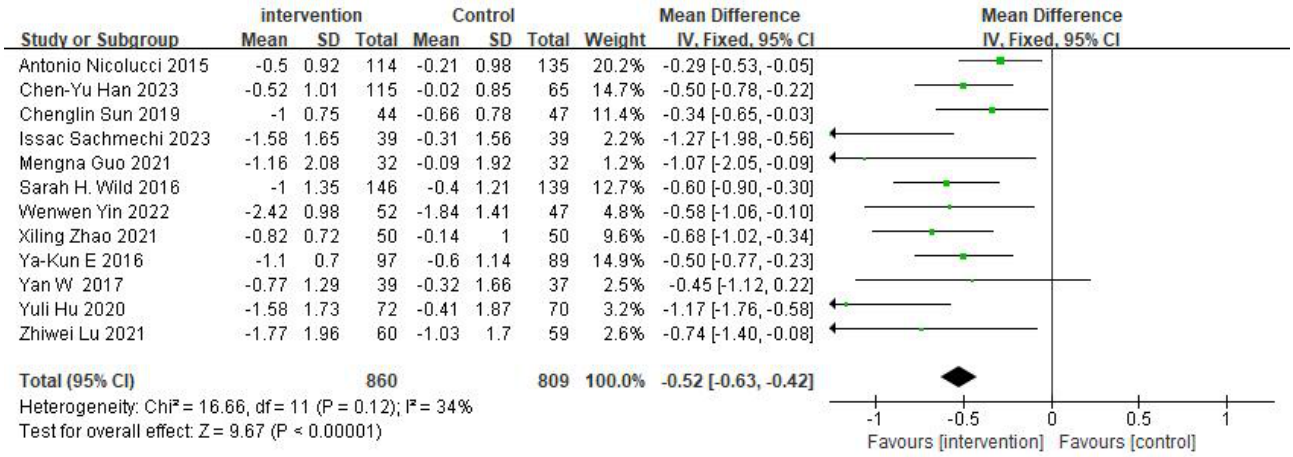

Effects on FBG

Seven studies were reported on FBG [28-34]. The results indicated that participants in the intervention group experienced a significantly greater reduction in FBG compared to the control group (n=839; MD −0.42, 95% CI −0.65 to −0.19 mmol/L; *z* value 3.59; *P*<.001; I^2=5%; fixed-effects model; refer to Figure 5 [28-34]).

XSL•FO
RenderX

Figure 5. Forest plot of fasting blood sugar.

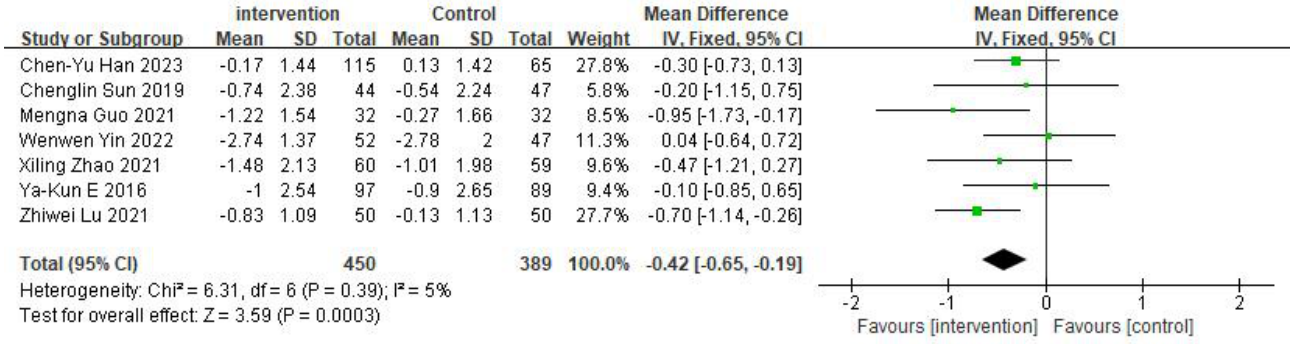

Effects on PBG

Five studies reported on PBG [29-33]. The results showed that participants in the intervention group had a significantly greater reduction in PBG compared to the control group (n=540; MD −0.64, 95% CI −0.97 to −0.32 mmol/L; *z* value=3.92; *P*<.001; I^2=46%; fixed-effects model; refer to Figure 6 [29-33]).

Figure 6. Forest plot of 2-hour postprandial blood glucose.

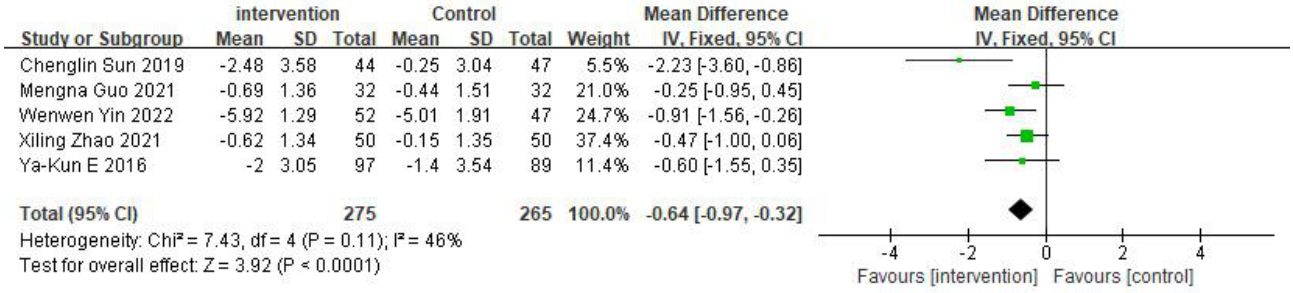

Effects on BMI

Six studies reported on the effectiveness of digital diabetes management in improving BMI [28-33]. The meta-analysis results showed a significant improvement in BMI (n=720; MD −1.55, 95% CI −2.92 to −0.17 kg/m^2; *z* value=2.20; *P*=.03, I^2=93%; random-effects model) compared to the control group (refer to Figure 7 [28-33]). Due to high heterogeneity, 3 studies were identified as contributing to the variability, as revealed by sensitivity analysis [29-31]. After removing these 3 studies, the heterogeneity was reduced to 0% (MD −2.16, 95% CI −2.55 to −1.76 kg/m^2; *z* value=10.70; *P*<.001).

Figure 7. Forest plot of BMI.

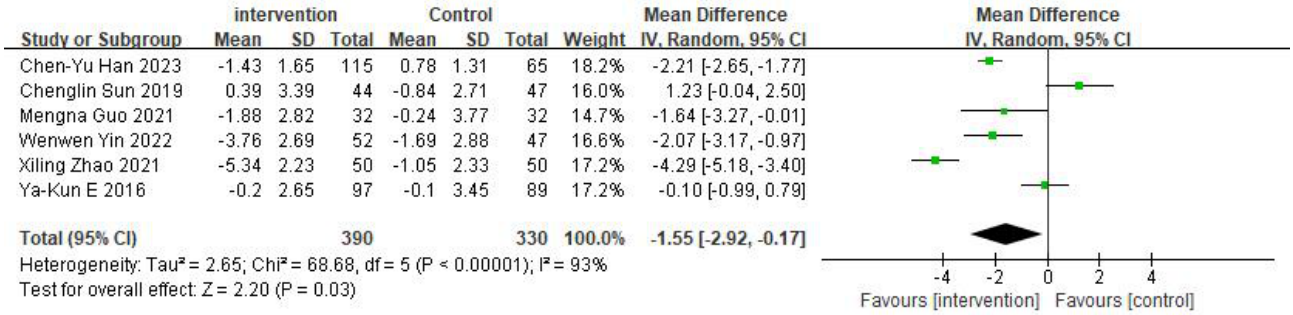

Publication Bias and Sensitivity Analyses

Most of the points in the distribution of HbA_{1c}, an outcome measure, were clustered within the funnel plot; however, the distribution was not entirely symmetrical, suggesting the possibility of publication bias in the results for HbA_{1c} in digital diabetes management (refer to Figure 8). We also performed a sensitivity analysis to assess how the results varied when individual studies were excluded and to evaluate the stability of the findings. When we removed a study from the model and recalculated the pooled estimates for the remaining studies, heterogeneity either disappeared or decreased, and the pooled results remained consistent, indicating the robustness of our findings.

XSL•FO
RenderX

Figure 8. Funnel plot. MD: mean difference.

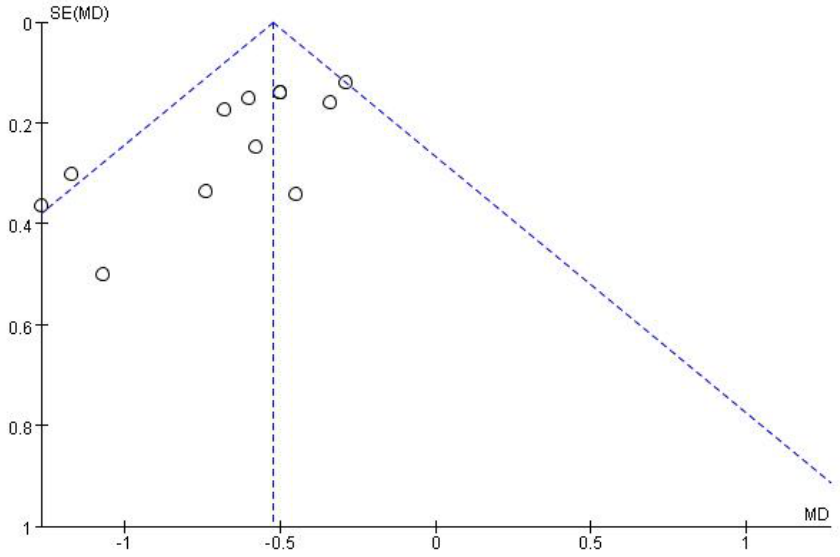

Subgroup Analysis

Subgroup analysis of BMI based on the length of the intervention revealed that digital diabetes management for 3 months or less resulted in more significant improvements in BMI (refer to Table 1).

Table 1. Subgroup analysis.

Outcomes	Studies, n	Participants, n	Heterogeneity		Effect model	Meta-Analysis	
			P value	I^2 value (%)		Effect size (95% CI)	*P* value
BMI	6	720	<.001	93	Random	–1.55 (–2.92 to –0.17)	.03
≤3 months	2	164	.005	87	Random	–3.06 (–5.65 to –0.47)	.02
6 months	4	556	<.001	92	Random	–0.84 (–2.37 to 0.69)	.28

Discussion

Principal Findings

This meta-analysis evaluated the effectiveness of digital diabetes management technology based on digital SMBG for home-based patients with T2DM. The findings demonstrated significant improvements in HbA_{1c}, FBG, PBG, and BMI. Key features of effective digital health interventions include frequent patient engagement in SMBG and timely, personalized guidance provided by specialized HCPs.

Studies have shown that while the compliance rate for blood glucose control among patients with T2DM during hospitalization can reach as high as 80%, this rate significantly decreases by 25% to 30% once they are discharged and transition to home self-management [40,41]. Consequently, there is an urgent need to support home-based patients with T2DM in managing blood glucose fluctuations [42]. This study demonstrates that compared with traditional home blood glucose management, digital diabetes management technologies can significantly improve blood glucose levels in home-based patients with T2DM. Twelve studies [28-39] reported improvements in HbA_{1c} levels among home-based patients with T2DM following the use of digital diabetes management technologies, 7 studies reported that digital diabetes management technologies effectively improve FBG levels [28-34], and 5 studies examined the impact on PBG levels [29-33]. In conclusion, digital diabetes management technologies play a pivotal role in blood glucose management for home-based patients with T2DM and should be widely implemented in clinical practice. Ongoing focus on blood glucose monitoring is essential for maintaining stable blood sugar levels in these patients [43].

At the same time, the meta-analysis results of this study suggest that digital diabetes management technologies can reduce the BMI of home-based patients with T2DM, consistent with the findings of a systematic review by Marcolino et al [44]. However, the heterogeneity in this study is relatively high. Sensitivity analysis, conducted by excluding studies one at a time, indicated that the variability may be due to differences in intervention durations. Subgroup analysis based on intervention duration showed that when the intervention lasted less than 3 months, it effectively improved BMI. In contrast, when the intervention duration exceeded 6 months, no similar improvement was observed. Given the relatively small sample size, future large-scale, double-blind randomized controlled trials are needed to further validate the impact of digital diabetes management technologies on BMI.

With continuous innovation in sensor, microelectronics, and data analysis technologies, researchers have been actively investing in research and development. Meanwhile, the market's

XSL•FO
RenderX

strong demand for convenient and efficient blood glucose monitoring has spurred the rapid development of continuous glucose monitoring (CGM) technology. Medical staff assist patients in accessing blood glucose data via supporting mobile apps to formulate treatment plans. CGM technology offers continuous, dynamic blood glucose data, enabling a comprehensive grasp of blood glucose fluctuations, allowing timely detection of asymptomatic hypoglycemia and hyperglycemia, and contributing to optimized treatment plans and prevention of diabetes complications.

Nonetheless, CGM technology has limitations. Its equipment is costly, making long-term use unaffordable for some patients. In addition, frequent blood glucose checks may cause psychological stress in some, negatively affecting blood glucose control. Therefore, SMBG continues to play an indispensable role in diabetes management. SMBG offers patients immediate blood glucose readings at specific time points, enabling prompt adjustments to diet, exercise, or insulin dosage to maintain glucose levels within a safe range and prevent hypoglycemia or hyperglycemia. Furthermore, compared to CGM, SMBG devices are simpler, more affordable, and thus more accessible to a broader patient population, especially those with limited financial resources or relatively stable glucose fluctuations. SMBG also promotes active patient participation, increasing their awareness of disease management and enabling them to better understand the impact of daily factors on blood glucose levels. This process fosters improved lifestyle adjustments, enhances confidence in glucose control, and strengthens compliance with self-management regimens.

Among the 12 included randomized controlled trials, only 2 studies evaluated the impact of SMBG frequency on blood glucose levels [28,36], so a meta-analysis was not conducted. Both studies demonstrated that more frequent home blood glucose monitoring improved blood glucose control in home-based patients with T2DM. We observed that the frequency of SMBG in the intervention group was significantly higher than in the control group, and the digital diabetes management technology notably enhanced blood glucose control. Tomah et al [45] reported that higher SMBG frequency correlated with greater improvements in HbA_{1c}, suggesting that the frequency of measurements may directly influence clinically relevant outcomes, with 4 to 8 measurements per day being the most effective. Similarly, Mannucci et al [46] found that SMBG was more effective when a structured monitoring plan was implemented, and HCPs used the collected data. None of the 12 randomized controlled trials included in this study considered the frequency of contact with HCPs as an outcome measure to assess improvements in blood glucose levels. As a result, a meta-analysis could not be conducted. However, an analysis of the intervention strategies in these studies revealed that the frequency of contact with HCPs was higher in the intervention group than in the control group. In addition, a systematic review found that SMBG without support from HCPs did not yield significant clinical benefits [47]. This lack of benefit may be attributed to treatment inertia, where poor adherence to lifestyle changes and prescribed medications is common among patients with T2DM [48,49]. When feedback from HCPs is minimal, patients may feel anxious and uncertain when faced with poor SMBG control. In conclusion, digital diabetes management technologies, along with SMBG supported by HCPs, can significantly improve blood glucose control in home-based patients with T2DM.

This study advocates for the promotion of digital diabetes management technologies based on digital SMBG for home-based patients with T2DM. Such technologies can help eliminate transcription errors associated with manual data entry and offer increased convenience for patients, especially older adult individuals with T2DM, who may struggle with inputting blood glucose values due to poor vision or limited dexterity. This technology provides an age-friendly solution for blood glucose management. In addition, for older adult patients who live alone or have limited mobility, HCPs can remotely monitor blood glucose data, offer timely care, and quickly detect and address hypoglycemia or hyperglycemia. This improves the continuity and responsiveness of diabetes management. Furthermore, the use of digital devices in clinical practice can reduce the workload of clinical nurses and enhance overall work efficiency. Thus, it is recommended that this technology be widely implemented in both home and clinical settings to improve the safety and efficiency of diabetes management.

Comparison to Previous Work

The results of this meta-analysis align with those of previous studies on the effectiveness of digital diabetes management technology for improving blood glucose levels in home-based patients with T2DM. However, in many previous studies, SMBG devices required manual data entry for glucose measurements, whereas this study focuses on digital SMBG technologies capable of automatic data uploading [50-56]. To date, no other meta-analysis has specifically evaluated the application of digital diabetes management technology based on automatic SMBG for home-based patients with T2DM.

In addition, the results of this study revealed that digital diabetes management technology can significantly reduce BMI levels in home-based patients with T2DM, consistent with the findings of a systematic review by Marcolino et al [44]. However, several other systematic reviews suggest that digital diabetes management technology has no significant effect on BMI improvement [57-59]. This discrepancy highlights the need for future high-quality, large-sample studies to further investigate the impact of digital diabetes management technology on BMI. Such studies should explore potential differences in results due to variations in patient populations, intervention measures, and follow-up durations. Moreover, standardizing research methodologies and evaluation criteria will enhance the comparability and reliability of findings, providing a stronger evidence base for determining whether this technology can effectively improve BMI in patients with T2DM.

Strengths and Limitations

This systematic review and meta-analysis has several notable strengths. First, to the best of our knowledge, it is the first meta-analysis to evaluate digital diabetes management techniques based on digital SMBG for improving blood glucose levels and BMI in home-based patients with T2DM. Second, a comprehensive search of multiple databases was conducted,

XSL•FO
RenderX

with strict adherence to established methodological tools to ensure the reliability and transparency of the research process. Third, this study included a detailed subgroup analysis of BMI outcomes, confirming that the duration of intervention significantly influences the effectiveness of digital diabetes management technologies in home-based patients with T2DM. Finally, by analyzing multiple outcome indicators, this review provides a more comprehensive understanding of the role of digital health in diabetes management, offering valuable references for clinical practice.

We also acknowledge that this meta-analysis has some limitations. First, confined to Chinese and English literature, it excluded non–English and Chinese materials, failing to reflect global research. Future reviews should cover more languages. Second, most selected studies lack blinding details, risking measurement biases. Future research must focus on blinded outcome evaluation to minimize performance biases. Third, the HbA_{1c} funnel plot asymmetry indicates publication bias. Future studies should use multicenter, large-sample, randomized controlled trials for reliable results. Fourth, done over a year ago, the study may miss recent research due to the field's rapid development, and lacking timeliness. Future research needs constant updates. Fifth, 8 (67%) of the 12 studies were from China, with regional limitations as China's medical resources and telemedicine acceptance differ from elsewhere. Future studies should include more diverse regions to boost conclusion generality.

Future Directions

Overall, future digital health research can fill the existing gaps and improve the approach in 5 key areas. First, for the diabetes management technology based on SMBG, continuous exploration should be carried out on its advantages in blood glucose control. In particular, the impact of SMBG supported by HCPs on blood glucose should be assessed to tap more potential. Second, the intervention measures should be elaborated in detail so as to achieve the maximum degree of reproducibility and universality in different clinical contexts. Third, a comprehensive evaluation system will be constructed, covering aspects such as cost-effectiveness, patient satisfaction, and technology availability. This system is used to measure the value of intervention measures for the sustainable development of digital health in diabetes management. Fourth, researchers should be encouraged to study the differences in the efficacy of different population subgroups or provide patient-level data, enabling future reviews to identify the patient characteristics related to treatment outcomes. Fifth, researchers should actively collect and report patients' views on the intervention measures to improve the design basis.

Conclusion

This meta-analysis demonstrates that digital diabetes management technologies can effectively regulate blood glucose levels, improve glucose metabolism, and reduce BMI in patients with T2DM. However, due to the variation in intervention strategies, frequencies, and content across the included studies, future research should prioritize multicenter, large-sample, high-quality randomized controlled trials to refine and standardize these interventions. Notably, effective digital health interventions share a common characteristic: frequent patient engagement in SMBG, supported by HCPs who provide timely and personalized guidance.

Acknowledgments

This research was funded by the Innovation Training Project of College Students' Innovation and Entrepreneurship Training Program (S202210716065), the Project of Shaanxi Sports Bureau (20240063), and the Key Discipline Construction Project of High-level Traditional Chinese Medicine in Nursing of Shaanxi University of Chinese Medicine (2024XKZD16). In addition, we declare that we did not use generative artificial intelligence in any part of writing the manuscript.

Data Availability

All data generated or analyzed during this study are included in this published paper and its supplementary information files.

Authors' Contributions

YX contributed to conceptualization, data curation, formal analysis, methodology, software, and writing the original draft. LZ performed formal analysis, software, and writing the original draft. NX handled conceptualization, methodology, writing review, and editing. FC managed methodology, project administration, writing review, and editing. ZS contributed to conceptualization, methodology, and project administration. SZ handled project administration, writing, review, and editing. ZW managed conceptualization, funding acquisition, methodology, project administration, writing review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 49 KB - jmir_v27i1e66441_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Characteristics of the 12 included studies.
[DOCX File , 23 KB - jmir_v27i1e66441_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Checklist.
[DOCX File , 270 KB - jmir_v27i1e66441_app3.docx]

References

1. Ahmad E, Lim S, Lamptey R, Webb DR, Davies MJ. Type 2 diabetes. Lancet 2022 Nov 19;400(10365):1803-1820. [doi: 10.1016/S0140-6736(22)01655-5] [Medline: 36332637]
2. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119 [FREE Full text] [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
3. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, 9 edition. Diabetes Res Clin Pract 2019;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
4. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol 2011 Nov 08;8(4):228-236. [doi: 10.1038/nrendo.2011.183] [Medline: 22064493]
5. Wong TY, Cheung CMG, Larsen M, Sharma S, Simó R. Diabetic retinopathy. Nat Rev Dis Primers 2016;2:16012. [doi: 10.1038/nrdp.2016.12] [Medline: 27159554]
6. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018;14(2):88-98. [doi: 10.1038/nrendo.2017.151] [Medline: 29219149]
7. Heflin C, Hodges L, Ojinnaka CO, Arteaga I. Hypertension, diabetes and medication adherence among the older supplemental nutritional assistance program population. J Appl Gerontol 2022;41(3):780-787 [FREE Full text] [doi: 10.1177/07334648211022493] [Medline: 34137290]
8. Wang Z, Wu Y, Wu J, Wang M, Wang X, Wang J, et al. Trends in prevalence and incidence of type 2 diabetes among adults in Beijing, China, from 2008 to 2017. Diabet Med 2021;38(9):e14487. [doi: 10.1111/dme.14487] [Medline: 33278034]
9. Samson SL, Vellanki P, Blonde L, Christofides EA, Galindo RJ, Hirsch IB, et al. American association of clinical endocrinology consensus statement: comprehensive type 2 diabetes management algorithm - 2023 update. Endocr Pract 2023;29(5):305-340. [doi: 10.1016/j.eprac.2023.02.001] [Medline: 37150579]
10. Yan M, Yu Y, Li S, Zhang P, Yu J. Effectiveness of king's theory of goal attainment in blood glucose management for newly diagnosed patients with type 2 diabetes: randomized controlled trial. J Med Internet Res 2024;26:e59142 [FREE Full text] [doi: 10.2196/59142] [Medline: 39481094]
11. Simsek H, Erkoyun E, Akoz A, Ergor A, Ucku R. Unmet health and social care needs and associated factors among older people aged ≥ 80 years in Izmir, Turkey. East Mediterr Health J 2021;27(8):772-781 [FREE Full text] [doi: 10.26719/emhj.21.009] [Medline: 34486713]
12. Faridani L, Abazari P, Heidarpour M, Melali H, Akbari M. The effect of home care on readmission and mortality rate in patients with diabetes who underwent general surgeries. J Educ Health Promot 2021;10:418 [FREE Full text] [doi: 10.4103/jehp.jehp_81_21] [Medline: 35071624]
13. Smith JM, Jarrín OF, Lin H, Tsui J, Dharamdasani T, Thomas-Hawkins C. Racial disparities in post-acute home health care referral and utilization among older adults with diabetes. Int J Environ Res Public Health 2021;18(6):3196 [FREE Full text] [doi: 10.3390/ijerph18063196] [Medline: 33808769]
14. Koo DJ, Moon SJ, Moon S, Park SE, Rhee E, Lee WY, et al. Long-term glycemic control improvement after the home and self-care program for patients with type 1 diabetes: real-world-based cohort study. J Med Internet Res 2024;26:e60023 [FREE Full text] [doi: 10.2196/60023] [Medline: 39259960]
15. Simon D. Therapeutic inertia in type 2 diabetes: insights from the PANORAMA study in France. Diabetes Metab 2012;38 Suppl 3:S47-S52. [doi: 10.1016/S1262-3636(12)71534-3] [Medline: 22541602]
16. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018;18(1):495 [FREE Full text] [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]
17. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024;26:e55757 [FREE Full text] [doi: 10.2196/55757] [Medline: 39037772]
18. Blackwell M, Tomlinson PA, Rayns J, Hunter J, Sjoeholm A, Wheeler BJ. Exploring the motivations behind misreporting self-measured blood glucose in adolescents with type 1 diabetes - a qualitative study. J Diabetes Metab Disord 2015;15:16 [FREE Full text] [doi: 10.1186/s40200-016-0238-6] [Medline: 27274982]

XSL•FO
RenderX

19. Blackwell M, Wheeler BJ. Clinical review: the misreporting of logbook, download, and verbal self-measured blood glucose in adults and children with type I diabetes. Acta Diabetol 2017;54(1):1-8. [doi: 10.1007/s00592-016-0907-4] [Medline: 27605000]
20. Bode B, King A, Russell-Jones D, Billings LK. Leveraging advances in diabetes technologies in primary care: a narrative review. Ann Med 2021;53(1):805-816 [FREE Full text] [doi: 10.1080/07853890.2021.1931427] [Medline: 34184589]
21. Cappon G, Vettoretti M, Sparacino G, Facchinetti A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 2019;43(4):383-397 [FREE Full text] [doi: 10.4093/dmj.2019.0121] [Medline: 31441246]
22. Mihai DA, Stefan DS, Stegaru D, Bernea GE, Vacaroiu IA, Papacocea T, et al. Continuous glucose monitoring devices: a brief presentation (Review). Exp Ther Med 2022;23(2):174 [FREE Full text] [doi: 10.3892/etm.2021.11097] [Medline: 35069855]
23. Handa T, Onoue T, Kobayashi T, Maeda R, Mizutani K, Yamagami A, et al. Effects of digitization of self-monitoring of blood glucose records using a mobile app and the cloud system on outpatient management of diabetes: single-armed prospective study. JMIR Diabetes 2024;9:e48019 [FREE Full text] [doi: 10.2196/48019] [Medline: 38241065]
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
25. Cochrane Handbook for Systematic Reviews of Interventions. URL: https://training.cochrane.org/handbook/current/chapter-04 [accessed 2024-03-14]
26. Cochrane training. RevMan 5. URL: http://community.cochrane.org/tools/review-production-tools/revman-5 [accessed 2024-02-26]
27. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
28. Han CY, Zhang J, Ye XM, Lu JP, Jin HY, Xu WW, et al. Telemedicine-assisted structured self-monitoring of blood glucose in management of T2DM results of a randomized clinical trial. BMC Med Inform Decis Mak 2023;23(1):182 [FREE Full text] [doi: 10.1186/s12911-023-02283-4] [Medline: 37710226]
29. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone-based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
30. Zhao XL. Impact of blood glucose information management system on body composition, blood glucose, blood lipids and self-management ability of patients with type 2 diabetes mellitus. J Chronic Dis 2021;22(06):871-874. [doi: 10.16440/j.cnki.1674-8166.2021.06.014]
31. Ya-Kun E, Yu-Jia L, Xiao-Kun G. The application of telemedicine in the integrated management of type 2 diabetes mellitus. Chin J Diabetes 2016;24(05):439-442. [doi: 10.3969/j.issn.1006-6187.2016.05.013]
32. Yin W, Liu Y, Hu H, Sun J, Liu Y, Wang Z. Telemedicine management of type 2 diabetes mellitus in obese and overweight young and middle-aged patients during COVID-19 outbreak: a single-center, prospective, randomized control study. PLoS One 2022;17(9):e0275251 [FREE Full text] [doi: 10.1371/journal.pone.0275251] [Medline: 36174028]
33. Guo M, Meng F, Guo Q, Bai T, Hong Y, Song F, et al. Effectiveness of mHealth management with an implantable glucose sensor and a mobile application among Chinese adults with type 2 diabetes. J Telemed Telecare 2023;29(8):632-640. [doi: 10.1177/1357633X211020261] [Medline: 34152238]
34. Lu Z, Li Y, He Y, Zhai Y, Wu J, Wang J, et al. Internet-based medication management services improve glycated hemoglobin levels in patients with type 2 diabetes. Telemed J E Health 2021;27(6):686-693. [doi: 10.1089/tmj.2020.0123] [Medline: 32907521]
35. Sachmechi I, Salam S, Amini M, Khan R, Spitznogle A, Belen T. Frequent monitoring of blood glucose levels via a remote patient monitoring system helps improve glycemic control. Endocr Pract 2023;29(6):441-447. [doi: 10.1016/j.eprac.2023.03.270] [Medline: 36965657]
36. Yan W, Shanshan Y, Jiaping N, Xiaojing L. Application effect of blood glucose information management system on the follow-up among patients with type 2 diabetes. Chin Nurs Manage 2017;17(09):1245-1248. [doi: 10.3969/j.issn.1672-1756.2017.09.021]
37. Hu Y, Wen X, Ni L, Wang F, Hu S, Fang F. Effects of telemedicine intervention on the management of diabetic complications in type 2 diabetes. Int J Diabetes Dev Ctries 2020;41(2):322-328 [FREE Full text] [doi: 10.1007/s13410-020-00893-6]
38. Wild SH, Hanley J, Lewis SC, McKnight JA, McCloughan LB, Padfield PL, et al. Supported telemonitoring and glycemic control in people with type 2 diabetes: the telescot diabetes pragmatic multicenter randomized controlled trial. PLoS Med 2016;13(7):e1002098 [FREE Full text] [doi: 10.1371/journal.pmed.1002098] [Medline: 27458809]
39. Nicolucci A, Cercone S, Chiriatti A, Muscas F, Gensini G. A randomized trial on home telemonitoring for the management of metabolic and cardiovascular risk in patients with Type 2 diabetes. Diabetes Technol Ther 2015;17(8):563-570. [doi: 10.1089/dia.2014.0355] [Medline: 26154338]
40. Chuanjiao L, Liaofang W, Chang Z, Ying F, Yuxiang Z. Analysis of point-of-care bedside glucose data of 896 non-critically ill patients with diabetes. Chin J Diabetes 2021;29(02):91-94. [doi: 10.3969/j.issn.1006-6187.2021.02.003]

XSL•FO
RenderX

41. Liqun Z, Xianfeng H, Ling R, Lafang W, Guoyue Y. The study of self-management status in elderly diabetic mllitus. Clin Med Eng 2011;18(07):1135-1137. [doi: 10.3969/j.issn.1674-4659.2011.07.1135]
42. Allemann S, Houriet C, Diem P, Stettler C. Self-monitoring of blood glucose in non-insulin treated patients with type 2 diabetes: a systematic review and meta-analysis. Curr Med Res Opin 2009;25(12):2903-2913. [doi: 10.1185/03007990903364665] [Medline: 19827909]
43. Feng Y, Zhao Y, Mao L, Gu M, Yuan H, Lu J, et al. The effectiveness of an eHealth family-based intervention program in patients with uncontrolled type 2 diabetes Mellitus (T2DM) in the community via WeChat: randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e40420 [FREE Full text] [doi: 10.2196/40420] [Medline: 36939825]
44. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
45. Tomah S, Mahmoud N, Mottalib A, Pober DM, Tasabehji MW, Ashrafzadeh S, et al. Frequency of self-monitoring of blood glucose in relation to weight loss and A1C during intensive multidisciplinary weight management in patients with type 2 diabetes and obesity. BMJ Open Diabetes Res Care 2019;7(1):e000659 [FREE Full text] [doi: 10.1136/bmjdrc-2019-000659] [Medline: 31413841]
46. Mannucci E, Antenore A, Giorgino F, Scavini M. Effects of structured versus unstructured self-monitoring of blood glucose on glucose control in patients with non-insulin-treated type 2 diabetes: a meta-analysis of randomized controlled trials. J Diabetes Sci Technol 2018;12(1):183-189 [FREE Full text] [doi: 10.1177/1932296817719290] [Medline: 28697625]
47. Farmer AJ, Perera R, Ward A, Heneghan C, Oke J, Barnett AH, et al. Meta-analysis of individual patient data in randomised trials of self monitoring of blood glucose in people with non-insulin treated type 2 diabetes. BMJ 2012;344:e486. [doi: 10.1136/bmj.e486] [Medline: 22371867]
48. Egede LE. Lifestyle modification to improve blood pressure control in individuals with diabetes: is physician advice effective? Diabetes Care 2003;26(3):602-607. [doi: 10.2337/diacare.26.3.602] [Medline: 12610008]
49. Donnan PT, MacDonald TM, Morris AD. Adherence to prescribed oral hypoglycaemic medication in a population of patients with Type 2 diabetes: a retrospective cohort study. Diabet Med 2002;19(4):279-284. [doi: 10.1046/j.1464-5491.2002.00689.x] [Medline: 11942998]
50. Eberle C, Stichling S. Effect of telemetric interventions on glycated hemoglobin A1c and management of type 2 diabetes mellitus: systematic meta-review. J Med Internet Res 2021;23(2):e23252 [FREE Full text] [doi: 10.2196/23252] [Medline: 33595447]
51. Kerr D, Ahn D, Waki K, Wang J, Breznen B, Klonoff DC. Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis. J Med Internet Res 2024;26:e55757 [FREE Full text] [doi: 10.2196/55757] [Medline: 39037772]
52. Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. J Med Internet Res 2021;23(2):e23244 [FREE Full text] [doi: 10.2196/23244] [Medline: 33605889]
53. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
54. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
55. Alharbi NS, Alsubki N, Jones S, Khunti K, Munro N, de Lusignan S. Impact of information technology-based interventions for type 2 diabetes mellitus on glycemic control: a systematic review and meta-analysis. J Med Internet Res 2016;18(11):e310 [FREE Full text] [doi: 10.2196/jmir.5778] [Medline: 27888169]
56. Shen Y, Wang F, Zhang X, Zhu X, Sun Q, Fisher E, et al. Effectiveness of internet-based interventions on glycemic control in patients with Type 2 diabetes: meta-analysis of randomized controlled trials. J Med Internet Res 2018;20(5):e172 [FREE Full text] [doi: 10.2196/jmir.9133] [Medline: 29735475]
57. Kim JE, Park TS, Kim KJ. The Clinical Effects of Type 2 Diabetes Patient Management Using Digital Healthcare Technology: A Systematic Review and Meta-Analysis. Healthcare (Basel) 2022;10(3):522 [FREE Full text] [doi: 10.3390/healthcare10030522] [Medline: 35327000]
58. Wu C, Wu Z, Yang L, Zhu W, Zhang M, Zhu Q, et al. Evaluation of the clinical outcomes of telehealth for managing diabetes: a PRISMA-compliant meta-analysis. Medicine (Baltimore) 2018;97(43):e12962 [FREE Full text] [doi: 10.1097/MD.0000000000012962] [Medline: 30412116]
59. Alfarwan N, Hodkinson A, Panagioti M, Hassan L, Kontopantelis E. Clinical and cost-effectiveness of telemedicine among patients with type 2 diabetes in primary care: a systematic review and meta-analysis. Diabet Med 2024;41(8):e15343. [doi: 10.1111/dme.15343] [Medline: 38780107]

XSL•FO
RenderX

Abbreviations

CGM: continuous glucose monitoring
FBG: fasting blood glucose
HbA_{1c}: hemoglobin A_{1c}
HCP: health care professional
MD: mean difference
PBG: postprandial blood glucose
PICOS: populations, interventions, comparisons, outcomes, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SMBG: self-monitoring of blood glucose
T2DM: type 2 diabetes mellitus

Edited by N Cahill; submitted 12.09.24; peer-reviewed by M Grady, AN Ali; comments to author 29.11.24; revised version received 19.12.24; accepted 03.02.25; published 03.03.25.

Please cite as:
Xiao Y, Wang Z, Zhang L, Xie N, Chen F, Song Z, Zhao S
Effectiveness of Digital Diabetes Management Technology on Blood Glucose in Patients With Type 2 Diabetes at Home: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e66441
URL: https://www.jmir.org/2025/1/e66441
doi:10.2196/66441
PMID:40053775

©Yuping Xiao, Zhenzhen Wang, Lintao Zhang, Nina Xie, Fangyao Chen, Zihao Song, Sha Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Poststroke eHealth Technologies–Based Rehabilitation for Upper Limb Recovery: Systematic Review

Margherita Rampioni[1], PhD; Sara Leonzi[1], MSc; Luca Antognoli[1], PhD; Anna Mura[2], PhD; Vera Stara[1], PhD

[1]Innovative Models for Aging Care and Technology, IRCCS-INRCA, Istituto Nazionale di Riposo e Cura per Anziani, Ancona, Italy

[2]Neuroscience Institute, University Miguel Hernández, Elche, Spain

Corresponding Author:
Margherita Rampioni, PhD
Innovative Models for Aging Care and Technology
IRCCS-INRCA, Istituto Nazionale di Riposo e Cura per Anziani
via Santa Margherita 5
Ancona, 60124
Italy
Phone: 39 +39 071 8004614
Email: m.rampioni@inrca.it

Abstract

Background: Stroke is one of the most common cerebral vascular diseases, usually affecting people aged 60 years and older. It leads to a variety of disabilities requiring motor and cognitive rehabilitation. Poststroke rehabilitation is critical for recovery, particularly for upper limb impairments, which affect approximately 80% of stroke survivors. Conventional rehabilitation often faces barriers such as cost, accessibility, and patient adherence. In contrast, eHealth technologies offer a promising alternative by providing accessible, cost-effective, and engaging rehabilitation solutions.

Objective: While numerous systematic reviews have explored various aspects of technology-based rehabilitation for poststroke upper limb recovery, there is a notable lack of comprehensive synthesis of these findings. This gap presents challenges, primarily due to the focus on specific technologies, which complicates understanding the overall effectiveness of these interventions. Consequently, clinicians and researchers may find it difficult to assess the field holistically, potentially hindering informed decision-making in clinical practice. This review synthesizes evidence from systematic reviews evaluating the effectiveness of eHealth technology–based interventions for upper limb recovery in poststroke individuals. Two main questions are examined: (1) Are eHealth technology–based therapies more or equally effective than conventional therapies for stroke rehabilitation? (2) What are the main clinical considerations for low-cost eHealth technology–based rehabilitation?

Methods: Comprehensive literature searches were conducted in PubMed, Web of Science, Scopus, Embase, and Google Scholar using predefined inclusion criteria based on the Population, Intervention, Comparison, Outcome, and Study Design (PICOS) framework. Systematic reviews published in English without date restrictions were included. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flowchart guided study selection. Methodological quality was assessed using the Assessment of Multiple Systematic Reviews (AMSTAR 2) criteria.

Results: A total of 1792 records were screened, resulting in 7 systematic reviews published between 2019 and 2023 being included. These reviews encompassed 95 studies involving 2995 participants with a mean age of 58.8 years across acute, subacute, and chronic stroke phases. Interventions included telerehabilitation, mobile health (mHealth) apps, augmented reality (AR), virtual reality (VR), wearable devices, and exergames. While AR and VR demonstrated potential benefits when combined with conventional therapies (eg, AR showing significant improvements in upper limb function with a standardized mean difference 0.657; $P<.001$), evidence for stand-alone effectiveness remained inconclusive due to heterogeneity in study designs, intervention protocols, and outcome measures. Most reviews were rated as critically low quality due to methodological limitations.

Conclusions: eHealth technologies hold promise for enhancing upper limb rehabilitation post stroke by addressing barriers such as cost and accessibility while providing engaging interventions. However, the field remains fragmented with insufficient evidence to establish clear efficacy. Future research should focus on standardizing protocols, optimizing neurorehabilitation principles such as dosage and task specificity, and improving methodological rigor to evaluate these interventions' long-term impact better.

(J Med Internet Res 2025;27:e57957) doi:10.2196/57957

XSL•FO
RenderX

KEYWORDS

stroke; rehabilitation; technology-based interventions; upper limb; technologies-based rehabilitation; limb; systematic review; cerebral vascular diseases; patient; effectiveness; database; therapy; conventional therapy; mobile phone

Introduction

Background

A stroke event refers to the alteration in brain functions after a sudden disruption of brain blood flow. It has an incidence of 15 million people a year. Stroke is one of the most common cerebral vascular diseases, usually affecting the population aged 60 years and older, leading to a variety of disabilities for about 5 million surviving patients who require motor and cognitive rehabilitation [1,2].

This massive incidence puts tremendous pressure on health care systems to satisfy the need for effective and sustainable solutions for rehabilitation poststroke after hospital discharge. Indeed, the consequences of stroke encompass a spectrum of disabilities, spanning physical impairments to cognitive challenges, including issues with language, social interactions, and emotional well-being [3]. In about 80% of cases, patients with stroke experience motor impairments of the upper limbs [4], which are functionally complex and difficult to recover, making activities of daily living very difficult for the patients. Poststroke disability negatively impacts the quality of life and health, and only 25% of patients recover with only minor impairments [5].

Poststroke rehabilitation has a lead role in the recovery of the patients, and it is never too late to start [6]. The level of recovery varies from person to person, but with time, effort, and support, many patients with stroke can significantly improve their function and quality of life [7]. For this, stroke rehabilitation should be planned and implemented in a structured and coordinated approach to help patients regain their physical, cognitive, and functional abilities. For instance, the increasing understanding of brain plasticity is relevant to the rehabilitation strategy and outcome after brain damage, and the principles of experience-dependent plasticity are valuable for treatment [8,9]. Indeed, rehabilitation requires multiple and varied therapeutic approaches at any stage of disease (acute and chronic phases) [10]. In addition, the quantity and quality of treatment are important for an effective recovery: rehabilitation after a stroke should be task-oriented, offered in large doses, and with an active learning component for providing intentional and more effective training [11]. Training several times daily is necessary to exploit the neuronal plasticity that ensures effective neurorehabilitation.

Nevertheless, high-intensity treatments are very expensive for the health care system, and hospital stays often must be reduced. This could mean that patients could still have functional deficits on the day of hospital discharge [4], especially in the upper limb [11,12].

In addition, outpatient therapy is not feasible for many patients because of the high costs of individual specialized therapies and the logistical difficulties of transportation from or to the hospital.

Home rehabilitation is a good option to continue recovery from chronic diseases such as stroke and a good alternative to manage long-term rehabilitation problems. Therefore, home rehabilitation is necessary to provide such a large amount of training [13] because traditional rehabilitation programs do not usually comply with all the abovementioned requisites. Thus, to maintain progress in the recovery process, the conventional rehabilitation pathway must include self-administered home exercises, a home rehabilitation solution, and outpatient therapy.

However, even within these strategies, certain limitations exist. Patients often lack motivation or lose interest in performing exercises independently, for example, they may consider them too difficult or too easy or monotonous. Here, technology can play a crucial role. The absence of external feedback can be overcome with interactive systems that provide real-time guidance and encouragement, stimulating patients to continue their rehabilitation [14].

Technology plays a crucial role in bridging the gap between home rehabilitation and remote therapy. By providing tools and platforms, it enables patients with stroke to receive therapy, guidance, and support without the need for a physical presence in a health care facility. This not only improves long-term recovery but also allows hospitals to reduce the length of outpatient treatment and its associated costs, while maximizing the treatment capacity of therapists.

Technology That Impacts Poststroke Rehabilitation

Attempting to reduce and overcome all the mentioned logistical and economic barriers of long-term rehabilitation, technological solutions offer a beneficial and effective alternative to conventional therapy, making rehabilitation more accessible to everyone. Rehabilitation technologies can significantly enhance the effectiveness and accessibility of stroke rehabilitation programs when incorporating principles of experience-dependent plasticity [8,9,15]. These advanced and simple technologies are more commonly explored and used in developed countries, mainly due to their readiness and availability, but their potential for accessibility is universal, making them a promising solution for poststroke rehabilitation worldwide [16].

Advanced technological interventions for poststroke rehabilitation include robotics, transcranial magnetic stimulation, transcranial direct current stimulation, brain-computer interface, and functional electrical stimulation. These therapies, while effective, require a certain level of expertise and resources that may not be readily available in a home setting. Therefore, they are typically conducted in a clinical or rehabilitation center under the supervision of trained professionals. Other current less invasive technologies are virtual reality (VR), augmented reality (AR), and activity trackers (such as accelerometers, gyroscopes, pedometers, breath sensing, heart rate monitors, and calorie trackers) [17]. As a cutting-edge and computer-generated simulation technology, VR can create an enriched environment, facilitate task-specific training, and provide multimodal feedback to augment functional recovery

XSL•FO
RenderX

[18]. Users can use their hands or movement sensors, such as gloves or joystick, to interact with virtual objects. The development of low-cost sensors has allowed this technology to spread at the consumer level: the Sony PlayStation 3, Microsoft Xbox 360, and Nintendo Wii are examples of VR-based consoles [15]. AR technology combines real and virtual objects to provide an interactive real-time experience in a common environment [19]. Users can interact with AR through devices such as smartphones, visors, display, or active mirrors. These devices use cameras, sensors, and software to overlay digital information onto the real world. In contrast, functional electrical stimulation or neuromuscular electrical stimulators have been used predominantly for stimulating lower and upper extremity functions [20].

Telerehabilitation (TR), mobile health (mHealth) apps, assistive technologies, and electromechanical gait training (with variable automated speed and sensing treadmill) are simple technological interventions for poststroke rehabilitation [17]. Home-based TR is a branch of telemedicine that consists of the use of a variety of telecommunication platforms (such as telephone visits, mobile apps, serious games, web-based self-care programs, web-based videoconferencing, and sensor-based telemonitoring) by health care professionals to provide necessary patient care and remote evaluation, supervision, and support for persons with disabilities living at home [11,21]. Home-based TR allows to meet the rehabilitation needs of stroke survivors living in rural areas with limited health services, especially in westernized countries, where the stroke burden is rapidly increasing [22]. Home-based TR implies access at any time and from any place to rehabilitation services to address stroke aspects: (1) *motor function*: upper and lower extremities, balance, and gait; (2) *cognitive function*: spatial neglect, cognition, and memory; and (3) *language*: aphasia [23]. TR can be delivered synchronously or asynchronously, depending on patients' needs, medical conditions, and treatment plans [24]. The choice between synchronous and asynchronous TR depends on various factors, including the patient's needs, the nature of the therapy, and the available technology. Synchronous sessions are often used for live consultations, real-time feedback, and interactive exercises. Asynchronous sessions are more flexible and beneficial when patients need rehabilitation in their daily routines. Some TR programs use a combination of both approaches to offer a well-rounded service that combines real-time interaction with flexibility and convenience.

Also, mHealth apps are defined as health and well-being mobile services for medical care delivered using a mobile app or other wireless technology. These are interesting for their mobility, multifunctional skills such as reminders and videos, and ability to support specific rehabilitation goals and promote self-management [25]. mHealth apps for stroke rehabilitation can target different aspects of the disease, and there are mobile apps designed as games to improve finger dexterity and programs to increase adherence to home rehabilitation exercises, for example, for upper limb rehabilitation [26,27].

Objectives

This study identifies and appraises published systematic reviews. The aim is to describe the quality, summarize and compare the conclusions, and discuss the strengths and weaknesses in the effectiveness of low-cost eHealth technology–based rehabilitation for the recovery of the upper limb in poststroke individuals.

While numerous systematic reviews have examined various aspects of technology-based rehabilitation for poststroke upper limb recovery, there is a lack of comprehensive synthesis of these findings. This gap presents several challenges. The first challenge concerns the focus on specific technologies, which makes it difficult for clinicians and researchers to understand the overall effectiveness of technology-based interventions and the field holistically. Without this comprehensive understanding, health care providers may struggle to make informed decisions about implementing these interventions in clinical practice. We aim to address these gaps by reviewing systematic reviews and providing a comprehensive, high-level synthesis of the current evidence.

This systematic review focuses on technology-based rehabilitation with interventions supported by synchronous (real-time care) or asynchronous (not real-time care) TR, mHealth apps, or eHealth portable devices equipped with VR or AR apps. Hence, in this systematic review, we assessed the low-cost eHealth systems to determine their impact on improving upper limb functional recovery among patients with stroke by examining 2 main questions: (1) Are eHealth technology-based therapies more or equally effective than conventional therapies for stroke rehabilitation? (2) What are the main clinical considerations for low-cost eHealth technology–based rehabilitation?

Methods

Databases, Criteria, and Search Strategy

Separate literature searches were conducted in PubMed, Web of Science, Scopus and Embase databases, and Google Scholar. The Population, Intervention, Comparison, Outcome, and Study Design (PICOS) framework was used to define the following inclusion criteria: (P) the target population was composed of persons with poststroke, (I) we considered technological interventions for upper limb rehabilitation in stroke survivors (ie, eHealth rehabilitation, at home TR, and smartphone-based rehabilitation), (C) studies were selected with or with no control group comparison, (O) we considered upper limb function recovery outcomes, and (S) we search for systematic reviews written in English that evaluated the effectiveness of technology-based intervention for the rehabilitation of the upper limb in poststroke patients. There was no restriction on publication dates. The combination of key terms reported in Multimedia Appendix 1 was used for the search in each database. The searches were finalized in May 2023. Papers were excluded if they were not systematic review papers or were not written in English.

According to the predefined criteria, the screening phase was based on analyzing titles and then abstracts. Later, full-paper articles of those titles or abstracts of screened publications were reviewed independently by MR and SL. VS was involved in reaching a consensus in cases of disagreement. Studies that met

XSL•FO
RenderX

the inclusion criteria were included, and the results of the searches were summarized. We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [28] flowchart in the retrieval and selection process. Each included systematic review was summarized by extracting essential data to answer the 2 research questions. The themes analyzed are (1) types of interventions and technology used, (2) effectiveness of eHealth technology–based interventions for recovering the upper limb, and (3) methodological quality assessments using the Assessment of Multiple Systematic Reviews (AMSTAR 2) criteria. The summaries were synthesized to provide a cohesive narrative highlighting similarities and differences across studies. This process involved comparing intervention types (eg, VR, AR, and mHealth apps), their reported effectiveness, and any noted limitations or gaps in evidence.

Study Quality Assessment

Three authors independently and anonymously appraised the final papers' methodological quality using the AMSTAR 2 [29]. AMSTAR 2 is designed to appraise systematic reviews of health care interventions and rate their overall confidence. The tool can appraise various aspects of systematic reviews of randomized controlled trials (RCTs), nonrandomized studies of interventions, or both.

It is composed of 16 items evaluated either with "yes" or "no" (items 1, 3, 5, 6, 10, 13, 14, and 16); with "yes," "partial yes," or "no" (items 2, 4, 7, 8, and 9); or with "yes," "no," or "no meta-analysis conducted" (items 11, 12, and 15). This tool evaluates the overall quality based on performance in critical and noncritical domains, which are assigned different weights in the rating rules. A "yes" answer means the item is fulfilled and is considered a positive result. Based on the performance in these 16 domains with different weights, overall ratings were generated, and the quality was determined to be "high," "moderate," "low," or "critically low." Namely, high confidence was defined as no or 1 noncritical weakness. Moderate confidence was defined as more than 1 noncritical weakness. Low confidence was defined as 1 critical flaw with or with no noncritical weaknesses. Critically low confidence was defined as more than 1 critical flaw with or with no noncritical weaknesses. Each review was appraised by 3 authors (MR, SL, and VR). In case of disagreement, it was planned that a fourth author (LA) would have solved them, but this was not the case. The analysis is reported in Multimedia Appendix 2.

Results

Characteristics of the Selected Reviews

As reported in the PRISMA flowchart (Figure 1), 7 systematic reviews were included. A summary of the studies and their findings is reported in Multimedia Appendix 3. Initially**,** the search process identified 1450 records from the databases and an additional 342 by Google Scholar. After duplicates (n=453) were removed, 1339 papers remained for initial screening by title. From this process, other 1297 papers were excluded because (1) studies involved patients for general rehabilitation (n=18), (2) the intervention was not specific for upper limb recovery functions (n=195), (3) the intervention was performed using other technologies (n=402), and (4) studies dealt with other topics or disorders (n=682). This process resulted in 42 potentially eligible abstracts. The authors analyzed the retained abstracts to obtain the final list of full-text papers to be reviewed. After examining the abstracts, 24 were excluded, as they did not fit the established criteria of the target population and the specific technology-based intervention. A second screening step was performed for those full-text papers that matched all the criteria (n=18). The other 11 were excluded from this process because the type of studies were not systematic reviews. The 7 studies included were published between 2019 and 2023.

XSL•FO
RenderX

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flowchart.

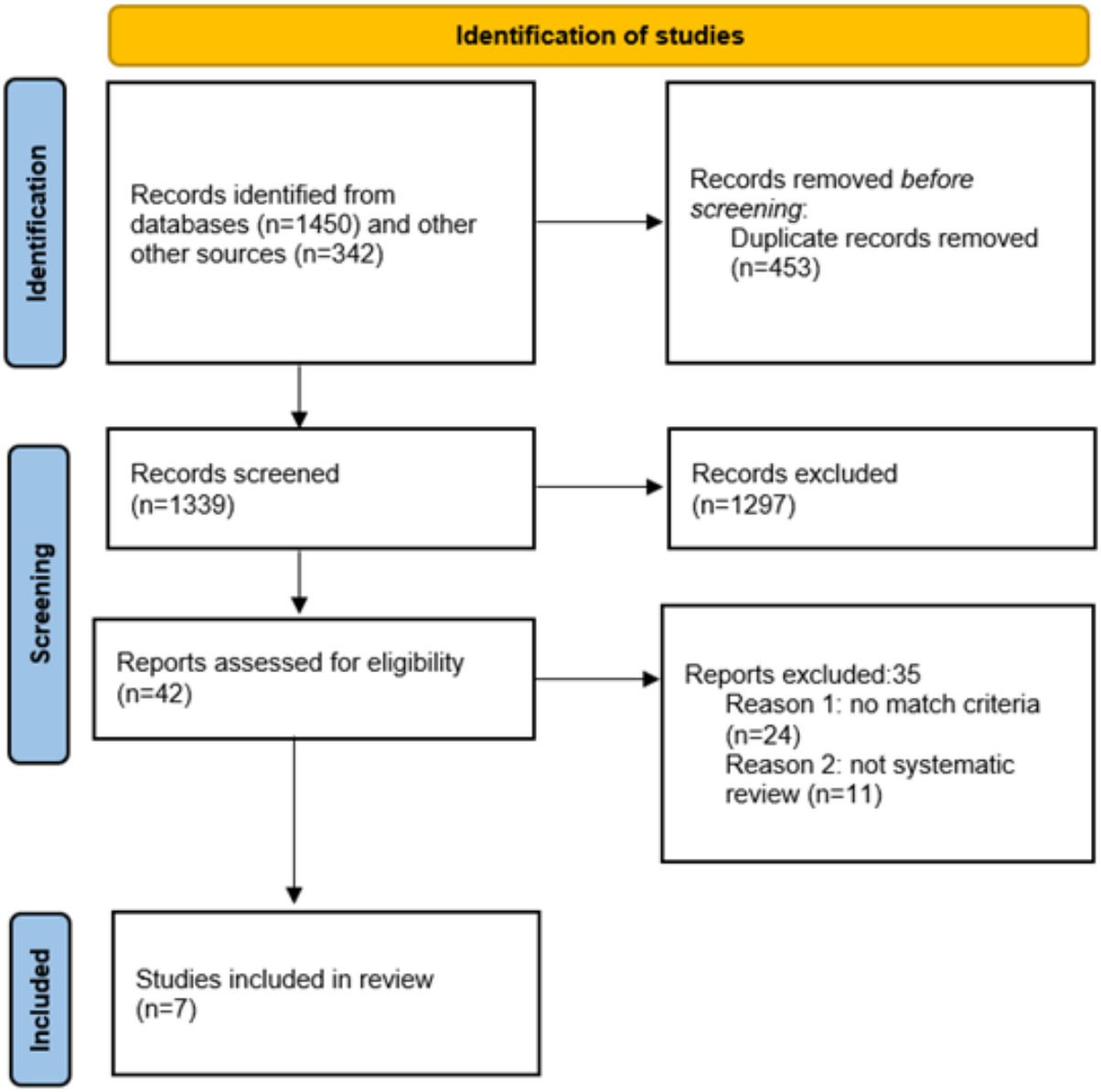

Five of the selected reviews defined the study's design using the PICOS framework [30-34] whereas the other 2 defined the inclusion criteria [35,36]. Four reviews [30,31,35,36] were registered in the research protocol on the international database of prospectively registered systematic reviews in health and social care (PROSPERO). A total of 95 studies were included in the 7 reviews that overall mapped not only RCTs [30,32,33,35,36] and RCTs and non-RCT studies [31,34] but also observational studies [32] and uncontrolled clinical trials [34]. A total of 2995 patients were enrolled with a mean age of 58.78 years (SD 0.15 years). The average duration of stroke onset among the participants ranged from 7 days to more than 5 years, as each of the 7 reviews included patients in the acute (7 days to 3 months), subacute (3-6 months), and chronic (>6 months) phases. Only 1 paper also included patients without specifying stroke duration [35].

The chronic patients enrolled have had a stroke from 6 months or more [33,36], 21 months [35], 3 years [30,34] to 5 years [31], and more than 5 years [32].

Four of the selected reviews [30-32,34] investigated the effectiveness of technology-based rehabilitation interventions on physical functioning compared with a combination of traditional treatments in patients with stroke [30], upper limb wearable technology for improving physical activity and social participation in adult stroke survivors [31], AR for the upper and lower limb functional recovery after stroke [32], and mHealth apps containing a physical training in stroke rehabilitation [34]. Two reviews [33,35] examined the effects of home-based exergaming interventions on upper limb activity after stroke, compared with conventional therapy, in postintervention and follow-up [33]; the use of mobile apps for stroke rehabilitation on stroke-related impairments (motor paresis, aphasia, and neglect); and functional outcomes

XSL•FO
RenderX

(adherence to exercise, activities of daily living, quality of life, secondary stroke prevention, and depression and anxiety) [35]. Only 1 review [36] gathered evidence on VR-based TR for patients after stroke and compared it with conventional in-person rehabilitation.

As reported in Table 1, all 7 reviews analyzed the upper limb function as an outcome. The categorization of the most used outcome measures of upper limb is reported in Multimedia Appendix 4.

Table 1. Outcomes.

Outcomes (upper limb function)	Authors (years)						
	Rintala et al (2019) [30]	Parker et al (2020) [31]	Phan et al (2022) [32]	Gelineau et al (2022) [33]	Rintala et al (2023) [34]	Szeto et al (2023) [35]	Hao et al (2023) [36]
Spasticity	N/A[a]	√	N/A	N/A	N/A	N/A	√
Pain	√	√	√	N/A	√	N/A	√
Strength	√	√	√	N/A	√	√	√
Difficulty perceived	√	√	√	√	N/A	N/A	√
Upper extremity function	√	√	√	√	√	√	√
Motor impairment	√	√	√	√	√	√	√
Fine manual dexterity	√	√	√	√	√	√	√
Gross manual dexterity	√	√	√	√	√	√	√

[a]N/A: not applicable.

Five reviews also assessed lower limb function and walking [30,32,34-36], balance [30,32,34,36] and physical activity, and function [30,34]. Participation was included as an outcome in 3 of the selected reviews [30-32], while the other 3 [34-36] evaluated the quality of life. Three studies assessed the activity of upper limb function [31-33], lower limb function [32], and participation [31,32] after stroke, according to the World Health Organization's International Classification of Functioning, Disability and Health framework [37].

Types of Interventions and Technology Used

Different low-cost eHealth technology–based interventions were reported in the 7 reviews analyzed. Web-based video monitoring and physical home exercises use smartphones, tablets, and web-based programs for real-time therapist-patient interaction and feedback. mHealth apps aim to improve limb movement and dexterity through user interaction with the app, allowing therapists to preselect exercises and assist caregivers in monitoring patient progress. Wearable technologies such as sensors and accelerometers capture actions or measurements and relay information for analysis and are connected to video games that require active body movements. AR projects virtual objects in real environments, allowing patients to interact with them and receive feedback. VR creates immersive simulations that require patient interaction, often with remote or in-person therapist supervision.

Web-Based Video

Web-based video monitoring and techniques for monitoring physical home exercises, goal settings or overall treatment, gamification, and accelerometers are mapped in 4 of the selected reviews [30,33-35]. The tools used varied from smartphones, phones, tablets, DVDs and programs via the web [30], and video game environments that required active body movements to control the game [33], to mHealth apps [34,35].

Looking at the details of the selected reviews, web-based video monitoring, phone calls, and messaging are the most common technologies used till now, as mentioned by Rintala et al [30]. They normally ensure a real-time therapist-patient interaction, with a "call" frequency ranging from 3 to 5 times per week to 1 per month. The therapist can also provide feedback through the web and, when necessary, by scheduling virtual training (eg, exercise videos) in advance [30].

Another option for remote technology–based rehabilitation, analyzed by Gelineau et al [33], is to provide exergames requiring a physical interaction (active body movements) to play the game. In that case, nonspecific video game systems (eg, Nintendo Wii, Xbox Kinect, etc) need to be combined with specific rehabilitation systems (eg, Rehabilitation Gaming System, virtual gloves, etc) or specific rehabilitation devices (eg, Hand Mentor Pro, Polhemus 3, etc). Patients with a prescription, usually ranging from 3 to 7 days per week, can use exergames systems independently, producing self-reported or observational measures (eg, technological sensors and therapist telehealth visits) [33].

mHealth Apps

Leveraging everyday technologies, such as smartphones and tablets or PCs, has become increasingly prominent in rehabilitation. This is made possible through implementing mHealth apps well investigated in the study by Gelineau et al [33]. These apps incorporate physical training components and offer a customizable approach tailored to individual patient requirements. This customization encompasses specific goals, desired difficulty levels, and the duration of app usage.

Moreover, these apps serve various purposes, including interactive gaming (eg, FINDEX and ARMStrokes), prescription of exercise routines (eg, CARE4STROKE), and progress monitoring (eg, STARFISH). They can also be integrated with

XSL•FO
RenderX

devices such as inertial measurement unit sensors and pedometers to enhance their effectiveness. This innovative use of technology complements traditional physiotherapy and widens the spectrum of rehabilitation possibilities.

Generally, the "gaming apps" require limb movement and interaction, aiming, for example, to improve upper limb and finger dexterity. In the "exercise prescription" app a therapist preselects a set of standardized exercises shown in the app, allowing the caregiver assistance, if necessary. "Monitoring apps" supervise and check patients' physical behavior (eg, number of steps per day, walking or sitting time, walking distance or speed, etc).

For all the types of apps, the prescription is maximally 30 minutes per session, from 1 to 7 times per week. The therapist's constant presence is not necessary, thanks to auditory and vibration feedback provided directly by the apps [34].

The mobile apps can be on any operating system (iOS, Android, and Windows) and on any aspect of stroke impairment rehabilitation (motor paresis, aphasia, and neglect). Focusing more on their possible goals, we could have therapy apps (with users' active device interaction to complete activities) or rehabilitation videos (exercises mobile guide), education apps (to learn about stroke), reminders (messaging to encourage compliance), or even a combination of them [35].

Wearable Technologies

Parker et al [31] described 2 types of wearable devices worn by patients. The first device operated independently and functioned as a central connector for other devices, whereas the second one captured specific action or executed a measurement and then sent the information to a primary wearable device for analysis.

The "wearable technologies" are something portable to wear externally on the body and to be used independently of a therapist. They can be used in both clinical and nonclinical settings to facilitate recovery, provide formative and real-time feedback, and measure intervention outcomes over long periods of time. They span from microelectromechanical systems (with accelerometers, gyroscopes, etc) to electromyographic biofeedback to robotics. This type of intervention can last 3-12 weeks, with various intensities [31].

AR- and VR-Based Rehabilitation

AR- and VR-based rehabilitation systems providing games were reviewed by Phan et al [32] and Hao et al [36]. Such systems included using head-mounted displays, Leap Motion, HD webcam, Microsoft Kinect v2 sensors connected to a laptop, or AR mirrors with visual tracking methods. AR technology is a real-time projection of virtual objects or scenarios in a common and real environment or place. During the game, the patients will interact with these virtual objects and receive automatic instructions, score, and visual and audio feedback via tracking devices (mouse, arm skate, HD Webcam, etc). The AR systems usually allow the creation of a personal training program with various training intensities, and the treatment duration is between 30 minutes and 1 hour per session [32]. VR is a computer-generated simulation technology that requires patient interaction, providing multimodal feedback. VR systems are often installed at patients' houses and may include the remote supervision of a therapist (eg, virtual-based TR group, interfaces for patient, and therapist remote communications). It is also used at hospitals or clinics, with a therapist's in-person supervision or instruction [36].

Effectiveness of eHealth Technology–Based Interventions for Recovering the Upper Limb

The selected reviews were analyzed regarding the effectiveness of the low-cost eHealth interventions. An effective intervention is based on statistically significant differences in 1 or more outcomes from the baseline, compared with the control, or if there is no control comparator, compared with another SD. Although all technology-based interventions included in these studies are economically favorable rehabilitation delivery models, no clear or direct evidence of impact on function recovery has been underlined.

Parker et al [31] highlighted little evidence to support the use of wearable technologies to improve activity and participation in the recovery of the upper limb.

Even if all the other reviews [30,32-36] reported similar effectiveness to conventional treatments, results are interpreted and discussed cautiously. For example, Rintala et al [30,34], Gelineau et al [33], and Szeto et al [35] discussed that rehabilitation technology, including mHealth apps, may have benefits as an additional treatment, but a lack of robust evidence does not consent authors to determine apparent effects.

Also, in the case of AR-based [32] and VR-based [36] TR, patients might still have a similar subjective experience of rehabilitation with therapists' supervision as in-person rehabilitation. According to Phan et al [32], the use of AR significantly influenced the upper limb function (standardized mean difference 0.657, 95% CI 0.287-1.026; $P<.001$). AR-based applications could offer options for increasing treatment intensity and promoting motor recovery after a stroke used with conventional rehabilitation methods. Also, for Hao et al [36], VR-based TR is a promising avenue for patients with stroke that can potentially overcome the barriers of traditional in-person rehabilitation. VR-based TR achieved comparable outcomes in the upper extremity function and equivalent effects on balance ability compared with in-person rehabilitation.

One aspect that holds significant clinical relevance is the duration of the intervention, which is measured in terms of both the quantity and the intensity of the dosage. Studies suggested a positive correlation between the time allocated for therapy and the therapy outcomes [38]. In the 7 reviews, the duration of the trials primarily ranged from 2 to 12 weeks.

According to Parker et al [31], improvements were observed across some studies for the control and intervention groups. The increase in the amount of rehabilitation has led to improvements in all the studies analyzed. This observation could suggest that a fundamental mechanism for improvement is the augmentation of the amount of rehabilitation administered. This concept has been acknowledged and incorporated into the national clinical guidelines for stroke. Szeto et al [35] concluded that the dosage and duration should be customized to address individual problems.

XSL•FO
RenderX

Some studies included in the study by Gelineau et al [33] reported follow-ups. However, in most of these cases, the duration of the follow-up is too short (4-24 weeks), leading to inconsistent results.

The selected reviews did not mention or report rigorous application of the experience-dependent plasticity principles, which are not always made explicit or clear [8,9].

In conclusion, further research is necessary to differentiate between the mechanisms of dosage and intensity. This will aid in understanding the impact of the volume of rehabilitation activity and how it compares with the intensity, which is the amount of rehabilitation administered over a specific time of period.

Quality Assessment

Among the 7 reviews assessed, 6 were rated as critically low quality [31-36], and 1 as low quality according to AMSTAR-2 criteria (Multimedia Appendix 2). Key issues affecting the validity of these reviews included inadequate protocol registration, inconsistent risk of bias assessments due to poor reporting on blinding and allocation concealment, and substantial heterogeneity across studies.

Only 1 review [36] lacked an explicit statement about protocol registration before the study began. All reviews demonstrated partial adequacy in their literature searches, with written protocols or guides that included review questions, search strategies, inclusion or exclusion criteria, and risk of bias assessments. None justified the exclusion of individual studies.

For risk of bias assessment, only 1 review [36] failed to use a satisfactory technique, while another [35] partially excluded nonrandomized studies of interventions. Three reviews [31,34,35] did not conduct a meta-analysis and 1 omitted it [32]. Five reviews [30,31,33-35] considered the risk of bias in their discussions. Only 2 reviews [30,32] adequately investigated publication bias and its potential impact.

In noncritical domains, 2 reviews [35,36] did not include all PICO components. All reviews explained their study design selections and performed study selection in duplicate; however, only 3 [30,31,33] duplicated data extraction. All reviews adequately described included studies; 4 [30,33,34,36] met all required criteria. Only 1 review reported funding sources for included studies. While 3 reviews [31,34,35] did not conduct a meta-analysis, only 2 [30,33] assessed the impact of bias on meta-analysis results. Five reviews [30,32-34,36] satisfactorily explained observed heterogeneity. Only 1 review failed to report potential conflicts of interest [30].

Overall, the critically low and low-quality assessments of the 7 reviews reflect the methodological shortcomings reported across these studies as low to moderate due to methodological limitations such as small sample sizes, lack of blinding, and heterogeneity in interventions.

Discussion

Principal Findings

Low-cost eHealth technology–based interventions should provide affordable and scalable science-based TR to cope with the pressure on health care systems to satisfy the need for effective and sustainable solutions for rehabilitation post stroke after hospital discharge.

In this study, 7 systematic reviews were selected to map the actual scenario of the use of low-cost eHealth technology–based interventions by examining two main questions: (1) Are eHealth technology–based therapies more or equally effective than conventional therapies for stroke rehabilitation? (2) What are the main clinical considerations for low-cost eHealth technology–based rehabilitation?

Most of these studies used advanced technologies with AR- and VR-based technologies and simple technologies with TR, mHealth apps, assistive technologies, and gait training.

Despite the differences among the reviews, all reached a unanimous conclusion, highlighting a cautious interpretation of data due to unclear evidence of the effectiveness of the examined technologies in restoring lost upper limb function [30-36]. Indeed, the variability in interventions, study designs, participant demographics, and measured outcomes contributed to heterogeneity across the studies, underscoring the need for future research to establish rehabilitation intervention principles that can inform the development of targeted innovative technologies [30-36]. Studies indicated that AR, VR, wearable devices, and exergames can improve upper limb motor function when combined with conventional rehabilitation methods, but their effectiveness alone is often comparable with traditional therapy [30,32-34,36].

Different clinical considerations favoring low-cost eHealth technology–based rehabilitation can be derived from the analyzed reviews. A central consideration is the potential of technology-driven interventions to enhance at-home poststroke rehabilitation for motivating survivors through interactive, user-friendly, engaging, and cost-effective tools. Indeed, eHealth systems could be seen as a way to reduce and overcome logistical and economic barriers of long-term conventional rehabilitation approaches by providing engaging and motivating rehabilitation interventions at home. Even if cognitive impairments or a lack of familiarity with technology can hinder the adoption of eHealth solutions, systems must be user-friendly to accommodate stroke survivors with varying levels of ability [30]. Continued research into the usability and effectiveness of this kind of system can guide the designers in developing a more user-friendly and engaging virtual ambient. Moreover, health care institutions should invest in patient education and engagement by introducing innovative systems into clinical practice.

Recent literature suggests that stroke affects the entire brain and its network properties, making it a network disease. Therefore, stroke-related neurological deficits and their recovery depend on neural network interaction patterns and follow principles of network plasticity [39]. In addition, the observation that network

XSL•FO
RenderX

interactions are correlated with current and future neurological function directly leads to whether their modulation through therapy might be feasible and clinically useful [40]. This evidence could open new frontiers for developing technology-based rehabilitation in poststroke individuals.

The feasibility of modulating brain networks through technology-based therapy may depend on several factors inherently linked to neurorehabilitation principles. According to these principles, exposure to specific training experiences in a recovery pathway improves impairment precisely because of the activation of these mechanisms of neuronal plasticity and remodeling [8]. The following neurorehabilitation principles should guide the selection of training experiences and rehabilitation technologies to optimize effectiveness [15]: massed practice, spaced practice, dosage, task-specific practice, goal-oriented practice, variable practice, increasing difficulty, multisensory stimulation, rhythmic cueing, explicit feedback or knowledge of results, implicit feedback or knowledge of performance, modulate effector selection, action observation or embodied practice, motor imagery, and social interaction. On the contrary, the selected reviews did not consistently adhere to all the principles of experience-dependent plasticity [8]. For example, interventions involving web-based video monitoring and phone calls often lack clear dosage specifications [30]. The AR system analyzed by Phan et al [32] used specific sensors [41] and viewers [42] to deliver exercises through games in both clinical and home settings. This system incorporated the neurorehabilitation principles of feedback (eg, game scores and social interaction), variable practice, and training intensity. However, AR systems face limitations: they are not always user-friendly, portable, or low-cost. While they offer a viable rehabilitation option, they may not be the most convenient solution for home-based upper limb rehabilitation.

Gelineau et al [33] highlighted the heterogeneity in exergame-based rehabilitation regarding the supervision and treatment dosage. Exergames are promising technologies that support stroke recovery and TR but still need to be explored more from clinical and neurorehabilitative perspectives. mHealth apps studies [34,35] often did not specify dosage prescriptions, indicating a limited application of neuroscience principles.

In VR-based TR [31,36], some systems offer synchronous patient supervision, while others provide only asynchronous settings; both ensure activity feedback and patient supervision. Nevertheless, the dosage prescriptions for these TR systems require further exploration. Low-cost eHealth technologies offer a promising avenue for stroke rehabilitation. They reduce barriers to rehabilitation by enabling remote monitoring and therapy delivery. This is particularly beneficial for patients in rural or underserved areas who face challenges accessing traditional in-person care [30,36]. Technologies such as wearable sensors and off-the-shelf gaming systems (eg, Wii) require poor infrastructure, making them suitable for home use [31,33]. Moreover, wearable devices and sensors allow continuous monitoring of movement quality, intensity, and frequency, providing valuable data for therapists to refine treatment plans remotely [31].

In conclusion, the integration of low-cost eHealth technology–based rehabilitation into clinical practice requires addressing usability challenges, standardizing protocols, rigorous adherence to all the principles of experience-dependent plasticity [8], and ensuring long-term effectiveness through further research. Combining these technologies with conventional therapies appears to yield the most beneficial outcomes for patients with stroke [43,44].

Limitations

Data sources were drawn only from PubMed, Web of Science, Scopus and Embase databases, and Google Scholar, excluding other data sources. Moreover, it is possible to miss relevant studies, especially if they are published in nonindexed journals, unavailable in electronic databases, or written in languages not included in the search criteria. Even if the PICOS framework was used to define the inclusion criteria, the combination of key terms to target the population and the specific technology-based intervention could have omitted some results from the search. Another significant limitation is the search strategy, which relied on a simple keyword search in titles and abstracts across databases. This approach may have inadvertently excluded relevant studies due to the lack of use of MeSH (Medical Subject Headings) and other controlled vocabulary terms that could have enhanced the comprehensiveness of the search. Medical Subject Headings terms are designed to capture all relevant papers on a topic, regardless of variations in terminology used by different authors. Future reviews should consider incorporating comprehensive search strategies, including keywords and controlled vocabulary terms across multiple fields within each database, to enhance the validity and reliability of conclusions drawn.

The heterogeneity and the different quality levels of the included studies are the significant limitations claimed by almost all the selected reviews [30,31,33-36]. Even if each review assessed quality using Physiotherapy Evidence Database Scale [30,33,34,36], Cochrane Risk of Bias for RCTs [31,35], Downs and Black Instrument for non-RCTs [31,34], and QualSyst [32], given the differences in assessment quality, our reviews' results should be interpreted and generalized with caution.

Moreover, we restricted the search strategy to only low-cost eHealth technology–based interventions for the upper limb recovery. This decision could have been the reason for excluding other relevant studies [45,46].

Implications for Clinical Practice and Future Research

This review indicated that the field of technology-based rehabilitation is still fragmented due to poor evidence achieved in terms of efficacy. This is probably due to the high heterogeneity of the experimental studies. We mapped various study designs, outcomes, and quality levels that demonstrate only the potential to assist with stroke recovery and augment face-to-face rehabilitation. Future research is called to frame a more robust methodology with a larger sample size [31-36] and valid measurement tools [31,34], deeper investigation on different gender and age groups [31,32], and other impacts depending on the stroke stage [33,35]. Also, the trial length and the study follow-up time were mentioned [35,36]. Regarding

XSL•FO
RenderX

technology readiness level, different issues were underlined to overcome the drawbacks of the reported technology-based rehabilitation interventions, for example, the integration and interoperability with intelligent infrastructure and the design of an attractive, user-friendly, portable, and low-cost system [32,34,35]. So, when developing a technology-based rehabilitation program, it is crucial to carefully plan and link all the relevant actors, user-driven design guidelines, and principles of neuroscience.

With this potential, the preferred option is to better understand the impact of technology interventions on varying types of stroke deficits and related outcomes, both alone and in combination with traditional rehabilitation. As Gelineau et al [33] have already highlighted the need for a gold standard, this field of research could benefit from standardized protocols provided to patients, enabling comparison and interpretation to discover evidence currently missing [15].

Conclusions

This review pooled the findings of 7 systematic reviews to map the actual scenario of the use of low-cost eHealth technology–based interventions and determine whether (1) eHealth technology–based therapies are more or equally effective than conventional therapies for stroke rehabilitation and (2) what the main clinical considerations for low-cost eHealth technology-based rehabilitation are. It found heterogeneity among interventions and measures, but the commonality of no clear or direct evidence of impact on recovery of function has been underlined. Advanced and simple technologies used for stroke rehabilitation allow overcoming financial, physical, and attitudinal barriers while providing engaging, specific, and low-cost exercises with constant feedback and supervision. These technologies serve as a valuable enhancement to traditional rehabilitation methods, particularly for the upper limb [47]. Unfortunately, demonstrating the efficacy of such interventions for restoring function after a stroke is still a challenge. Combining these technologies with interventions that follow neurorehabilitation principles may be more effective in promoting the recovery and retention of motor and cognitive functions after a stroke. Indeed, the integration of low-cost eHealth technology–based rehabilitation into clinical practice requires addressing usability challenges, standardizing protocols, rigorous adherence to all the principles of experience-dependent plasticity [8], and ensuring long-term effectiveness through further research. Combining these technologies with conventional therapies appears to yield the most beneficial outcomes for patients with stroke [48,49].

Acknowledgments

All authors provided intellectual contributions and critical feedback and reviewed the final manuscript. This study is cofunded by the EU Active and Assisted Living Program (AAL-2020-7-227-CP) and partially supported by Ricerca Corrente funding from the Italian Ministry of Health.

Authors' Contributions

VS designed the study; VS, SL, and MR conceptualized and drafted the manuscript and analyzed the data; and AM and LA reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Strings used for the search in PubMed, Web of Science, Scopus and Embase databases, and Google Scholar.
[DOCX File , 13 KB - jmir_v27i1e57957_app1.docx]

Multimedia Appendix 2
Amstar-2 .
[DOCX File , 23 KB - jmir_v27i1e57957_app2.docx]

Multimedia Appendix 3
Characteristics of the reviewed studies.
[DOCX File , 26 KB - jmir_v27i1e57957_app3.docx]

Multimedia Appendix 4
Outcome measures.
[DOCX File , 34 KB - jmir_v27i1e57957_app4.docx]

Multimedia Appendix 5
PRISMA checklist.

XSL•FO
RenderX

[DOCX File , 23 KB - jmir_v27i1e57957_app5.docx]

References

1. Wu D, Zhang H, Leng Y, Li K, Li S, Lo WLA. A bibliometric analysis of telerehabilitation services for patients with stroke. Front Neurol 2022;13:1026867 [FREE Full text] [doi: 10.3389/fneur.2022.1026867] [Medline: 36698904]
2. Moradi V, Babaee T, Esfandiari E, Lim SB, Kordi R. Telework and telerehabilitation programs for workers with a stroke during the COVID-19 pandemic: a commentary. Work 2021;68(1):77-80. [doi: 10.3233/WOR-203356] [Medline: 33427710]
3. National Institute of Neurological Disorders and Stroke. Post-stroke rehabilitation. NINDS Publications. 2020. URL: https://catalog.ninds.nih.gov/publications/post-stroke-rehabilitation [accessed 2023-06-16]
4. Rozevink SG, van der Sluis CK, Garzo A, Keller T, Hijmans JM. HoMEcare aRm rehabiLItatioN (MERLIN): telerehabilitation using an unactuated device based on serious games improves the upper limb function in chronic stroke. J Neuroeng Rehabil 2021;18(1):48 [FREE Full text] [doi: 10.1186/s12984-021-00841-3] [Medline: 33726801]
5. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. Circulation 2020;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
6. Zeiler SR. Should we care about early post-stroke rehabilitation? Not yet, but soon. Curr Neurol Neurosci Rep 2019;19(3):13. [doi: 10.1007/s11910-019-0927-x] [Medline: 30788609]
7. Ostrowska PM, Śliwiński M, Studnicki R, Hansdorfer-Korzon R. Telerehabilitation of post-stroke patients as a therapeutic solution in the era of the Covid-19 pandemic. Healthcare (Basel) 2021;9(6):654 [FREE Full text] [doi: 10.3390/healthcare9060654] [Medline: 34072939]
8. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
9. Maier M, Ballester BR, Verschure PFMJ. Principles of neurorehabilitation after stroke based on motor learning and brain plasticity mechanisms. Front Syst Neurosci 2019;13:74 [FREE Full text] [doi: 10.3389/fnsys.2019.00074] [Medline: 31920570]
10. Platz T. Clinical Pathways in Stroke Rehabilitation: Evidence-Based Clinical Practice Recommendations. Greifswald, Germany: Springer; 2021.
11. Cramer SC, Dodakian L, Le V, See J, Augsburger R, McKenzie A, National Institutes of Health StrokeNet Telerehab Investigators. Efficacy of home-based telerehabilitation vs In-Clinic therapy for adults after stroke: a randomized clinical trial. JAMA Neurol 2019;76(9):1079-1087 [FREE Full text] [doi: 10.1001/jamaneurol.2019.1604] [Medline: 31233135]
12. Rajsic S, Gothe H, Borba HH, Sroczynski G, Vujicic J, Toell T, et al. Economic burden of stroke: a systematic review on post-stroke care. Eur J Health Econ 2019;20(1):107-134. [doi: 10.1007/s10198-018-0984-0] [Medline: 29909569]
13. Forman CR, Nielsen JB, Lorentzen J. Neuroplasticity at home: improving home-based motor learning through technological solutions. a review. Front Rehabil Sci 2021;2:789165 [FREE Full text] [doi: 10.3389/fresc.2021.789165] [Medline: 36188793]
14. Neibling BA, Jackson SM, Hayward KS, Barker RN. Perseverance with technology-facilitated home-based upper limb practice after stroke: a systematic mixed studies review. J Neuroeng Rehabil 2021;18(1):43 [FREE Full text] [doi: 10.1186/s12984-021-00819-1] [Medline: 33627126]
15. Maier M, Rubio Ballester B, Duff A, Duarte Oller E, Verschure PFMJ. Effect of specific over nonspecific VR-based rehabilitation on poststroke motor recovery: a systematic meta-analysis. Neurorehabil Neural Repair 2019;33(2):112-129 [FREE Full text] [doi: 10.1177/1545968318820169] [Medline: 30700224]
16. Ferreira B, Menezes P. Gamifying motor rehabilitation therapies: challenges and opportunities of immersive technologies. Information 2020;11(2):88. [doi: 10.3390/info11020088]
17. Selamat SNS, Che Me R, Ahmad Ainuddin H, Salim MSF, Ramli HR, Romli MH. The application of technological intervention for stroke rehabilitation in southeast Asia: a scoping review with stakeholders' consultation. Front Public Health 2021;9:783565 [FREE Full text] [doi: 10.3389/fpubh.2021.783565] [Medline: 35198531]
18. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017;11(11):CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
19. Gorman C, Gustafsson L. The use of augmented reality for rehabilitation after stroke: a narrative review. Disabil Rehabil Assist Technol 2022;17(4):409-417. [doi: 10.1080/17483107.2020.1791264] [Medline: 32663112]
20. Sousa ASP, Moreira J, Silva C, Mesquita I, Macedo R, Silva A, et al. Usability of functional electrical stimulation in upper limb rehabilitation in post-stroke patients: a narrative review. Sensors (Basel) 2022;22(4):1409 [FREE Full text] [doi: 10.3390/s22041409] [Medline: 35214311]
21. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: review of the state-of-the-art and areas of application. JMIR Rehabil Assist Technol 2017;4(2):e7 [FREE Full text] [doi: 10.2196/rehab.7511] [Medline: 28733271]
22. Chen J, Sun D, Zhang S, Shi Y, Qiao F, Zhou Y, et al. Effects of home-based telerehabilitation in patients with stroke: a randomized controlled trial. Neurology 2020;95(17):e2318-e2330. [doi: 10.1212/WNL.0000000000010821] [Medline: 32999058]

XSL•FO
RenderX

23. Nikolaev VA, Nikolaev AA. Recent trends in telerehabilitation of stroke patients: a narrative review. NeuroRehabilitation 2022;51(1):1-22. [doi: 10.3233/NRE-210330] [Medline: 35527574]
24. Morse H, Biggart L, Pomeroy V, Rossit S. Exploring perspectives from stroke survivors, carers and clinicians on virtual reality as a precursor to using telerehabilitation for spatial neglect post-stroke. Neuropsychol Rehabil 2022;32(5):707-731. [doi: 10.1080/09602011.2020.1819827] [Medline: 32942950]
25. Zhou X, Du M, Zhou L. Use of mobile applications in post-stroke rehabilitation: a systematic review. Top Stroke Rehabil 2018:1-11. [doi: 10.1080/10749357.2018.1482446] [Medline: 30209991]
26. Chung BPH, Chiang WKH, Lau H, Lau TFO, Lai CWK, Sit CSY, et al. Pilot study on comparisons between the effectiveness of mobile video-guided and paper-based home exercise programs on improving exercise adherence, self-efficacy for exercise and functional outcomes of patients with stroke with 3-month follow-up: a single-blind randomized controlled trial. Hong Kong Physiother J 2020;40(1):63-73 [FREE Full text] [doi: 10.1142/S1013702520500079] [Medline: 32489241]
27. Grau-Pellicer M, Lalanza J, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Top Stroke Rehabil 2020;27(5):354-368. [doi: 10.1080/10749357.2019.1691816] [Medline: 31790639]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
29. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
30. Rintala A, Päivärinne V, Hakala S, Paltamaa J, Heinonen A, Karvanen J, et al. Effectiveness of technology-based distance physical rehabilitation interventions for improving physical functioning in stroke: a systematic review and meta-analysis of randomized controlled trials. Arch Phys Med Rehabil 2019;100(7):1339-1358. [doi: 10.1016/j.apmr.2018.11.007] [Medline: 30529323]
31. Parker J, Powell L, Mawson S. Effectiveness of upper limb wearable technology for improving activity and participation in adult stroke survivors: systematic review. J Med Internet Res 2020;22(1):e15981 [FREE Full text] [doi: 10.2196/15981] [Medline: 31913131]
32. Phan HL, Le TH, Lim JM, Hwang CH, Koo K. Effectiveness of augmented reality in stroke rehabilitation: a meta-analysis. Appl Sci 2022;12(4):1848. [doi: 10.3390/app12041848]
33. Gelineau A, Perrochon A, Robin L, Daviet J, Mandigout S. Measured and perceived effects of upper limb home-based exergaming interventions on activity after stroke: a systematic review and meta-analysis. Int J Environ Res Public Health 2022;19(15):9112 [FREE Full text] [doi: 10.3390/ijerph19159112] [Medline: 35897472]
34. Rintala A, Kossi O, Bonnechère B, Evers L, Printemps E, Feys P. Mobile health applications for improving physical function, physical activity, and quality of life in stroke survivors: a systematic review. Disabil Rehabil 2023;45(24):4001-4015 [FREE Full text] [doi: 10.1080/09638288.2022.2140844] [Medline: 36325613]
35. Szeto SG, Wan H, Alavinia M, Dukelow S, MacNeill H. Effect of mobile application types on stroke rehabilitation: a systematic review. J Neuroeng Rehabil 2023;20(1):12 [FREE Full text] [doi: 10.1186/s12984-023-01124-9] [Medline: 36694257]
36. Hao J, Pu Y, Chen Z, Siu K. Effects of virtual reality-based telerehabilitation for stroke patients: a systematic review and meta-analysis of randomized controlled trials. J Stroke Cerebrovasc Dis 2023;32(3):106960. [doi: 10.1016/j.jstrokecerebrovasdis.2022.106960] [Medline: 36586244]
37. Raghavendra P, Bornman J, Granlund M, Björck-Akesson E. The World Health Organization's International Classification of Functioning, Disability and Health: implications for clinical and research practice in the field of augmentative and alternative communication. Augment Altern Commun 2007;23(4):349-361. [doi: 10.1080/07434610701535905] [Medline: 19925317]
38. Daly JJ, McCabe JP, Holcomb J, Monkiewicz M, Gansen J, Pundik S. Long-dose intensive therapy is necessary for strong, clinically significant, upper limb functional gains and retained gains in severe/moderate chronic stroke. Neurorehabil Neural Repair 2019;33(7):523-537 [FREE Full text] [doi: 10.1177/1545968319846120] [Medline: 31131743]
39. Guggisberg AG, Koch PJ, Hummel FC, Buetefisch CM. Brain networks and their relevance for stroke rehabilitation. Clin Neurophysiol 2019;130(7):1098-1124 [FREE Full text] [doi: 10.1016/j.clinph.2019.04.004] [Medline: 31082786]
40. Coscia M, Wessel M, Chaudary U, Millán JDR, Micera S, Guggisberg A, et al. Neurotechnology-aided interventions for upper limb motor rehabilitation in severe chronic stroke. Brain 2019;142(8):2182-2197 [FREE Full text] [doi: 10.1093/brain/awz181] [Medline: 31257411]
41. Assis GAD, Corrêa AGD, Martins MBR, Pedrozo WG, Lopes RDD. An augmented reality system for upper-limb post-stroke motor rehabilitation: a feasibility study. Disabil Rehabil Assist Technol 2016;11(6):521-528. [doi: 10.3109/17483107.2014.979330] [Medline: 25367103]
42. Bank PJM, Cidota MA, Ouwehand PEW, Lukosch SG. Patient-tailored augmented reality games for assessing upper extremity motor impairments in Parkinson's disease and stroke. J Med Syst 2018;42(12):246 [FREE Full text] [doi: 10.1007/s10916-018-1100-9] [Medline: 30374695]

XSL•FO
RenderX

43. Ballester BR, Nirme J, Camacho I, Duarte E, Rodríguez S, Cuxart A, et al. Domiciliary VR-based therapy for functional recovery and cortical reorganization: randomized controlled trial in participants at the chronic stage post stroke. JMIR Serious Games 2017;5(3):e15 [FREE Full text] [doi: 10.2196/games.6773] [Medline: 28784593]
44. Nijenhuis SM, Prange-Lasonder GB, Stienen AH, Rietman JS, Buurke JH. Effects of training with a passive hand orthosis and games at home in chronic stroke: a pilot randomised controlled trial. Clin Rehabil 2017;31(2):207-216. [doi: 10.1177/0269215516629722] [Medline: 26869596]
45. Chen Y, Abel KT, Janecek JT, Chen Y, Zheng K, Cramer SC. Home-based technologies for stroke rehabilitation: a systematic review. Int J Med Inform 2019;123:11-22 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.12.001] [Medline: 30654899]
46. Toh SFM, Chia PF, Fong KNK. Effectiveness of home-based upper limb rehabilitation in stroke survivors: a systematic review and meta-analysis. Front Neurol 2022;13:964196 [FREE Full text] [doi: 10.3389/fneur.2022.964196] [Medline: 36188398]
47. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015;96(3):418-425.e2. [doi: 10.1016/j.apmr.2014.10.019] [Medline: 25448245]
48. Keser Z, Ikramuddin S, Shekhar S, Feng W. Neuromodulation for post-stroke motor recovery: a narrative review of invasive and non-invasive tools. Curr Neurol Neurosci Rep 2023;23(12):893-906. [doi: 10.1007/s11910-023-01319-6] [Medline: 38015351]
49. Adeyemo BO, Simis M, Macea DD, Fregni F. Systematic review of parameters of stimulation, clinical trial design characteristics, and motor outcomes in non-invasive brain stimulation in stroke. Front Psychiatry 2012;3:88 [FREE Full text] [doi: 10.3389/fpsyt.2012.00088] [Medline: 23162477]

Abbreviations

AMSTAR 2: Assessment of Multiple Systematic Reviews
AR: augmented reality
MeSH: Medical Subject Headings
mHealth: mobile health
PICOS: Population, Intervention, Comparison, Outcome, and Study Design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
TR: telerehabilitation
VR: virtual reality

Edited by A Coristine; submitted 01.03.24; peer-reviewed by SFM Toh, D Bonaiuti; comments to author 07.05.24; revised version received 29.07.24; accepted 18.12.24; published 04.03.25.

Please cite as:
Rampioni M, Leonzi S, Antognoli L, Mura A, Stara V
Poststroke eHealth Technologies–Based Rehabilitation for Upper Limb Recovery: Systematic Review
J Med Internet Res 2025;27:e57957
URL: https://www.jmir.org/2025/1/e57957
doi:10.2196/57957
PMID:40053744

©Margherita Rampioni, Sara Leonzi, Luca Antognoli, Anna Mura, Vera Stara. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Enabling Telemedicine From the System-Level Perspective: Scoping Review

Xuezhu Li[1*], BA; Lifeng Huang[1*], BA; Hui Zhang[1], PhD; Zhanming Liang[2], PhD

[1]School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]College of Public Health, Medical and Veterinary Science, James Cook University, Townsville, Queensland, Australia

[*]these authors contributed equally

Corresponding Author:
Hui Zhang, PhD
School of Public Health
Sun Yat-sen University
No. 74 Zhongshan Road 2, Yuexiu District
Guangzhou, 510080
China
Phone: 86 13922252312
Email: zhanghui3@mail.sysu.edu.cn

Abstract

Background: Telemedicine is a strategy for providing health care services remotely that improves service accessibility. Telemedicine has attracted growing research interest in the past 10 years, including systematic reviews that synthesize evidence to share experiences and enhance knowledge. However, most of the published systematic reviews have focused on synthesizing evidence from studies on telemedicine at the organizational level. A collected understanding of factors on the system level that influence the successful implementation and adoption of telemedicine needs to be developed, especially in regional and rural areas.

Objective: This scoping review aims to explore key success factors and challenges that influence the implementation and adoption of telemedicine at the system level, particularly in regional and rural areas.

Methods: This scoping review was conducted in accordance with the framework by Arksey and O'Malley and reported using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews). A total of 5 databases (CINAHL, Cochrane, Medline, Ovid, and Scopus) were searched for research articles published in English between January 2010 and 2023, using the established inclusion criteria.

Results: Of the 10,691 papers identified, 89 were included in this review, including 16 (17.98%) studies conducted in regional and rural areas and 13 (14.61%) in metropolitan areas. Another 13 (14.61%) studies were conducted in both metropolitan areas and regional and rural areas. Overall, 6 categories with more than 70 key success factors, including system-level requirements (n=13, 18.40%), economic considerations and funding (n=6, 8.70%), technological requirements (n=6, 8.70%), organizational requirements (n=19, 27.54%), understanding and supporting clinicians (n=12, 17.39%), and understanding and improving patients' perceptions (n=13, 18.84%), were identified. Additionally, 5 categories containing over 50 challenges, including those related to system levels (n=11, 23.91%), technological requirements (n=6, 13.04%), organizational requirements (n=13, 28.26%), clinicians (n=10, 21.74%), and patients (n=6, 13.04%), were identified. Among the identified factors, 11 (9.57%) were specific to regional and rural areas.

Conclusions: This scoping review confirms that the successful implementation of telemedicine requires collective efforts at both the system and organizational levels, including coordination and collaboration across different regions and organizations. It underscores the importance of establishing a national network that enhances public awareness of telemedicine and clarity in payment and benefit distribution models and strengthens data security protection measures. The review also highlights the necessity of addressing infrastructural deficiencies, including internet connectivity in regional and rural areas, and suggests the implementation of targeted incentives and support measures. The required collective efforts are detailed in the proposed framework that promotes popularizing telemedicine, enhancing the overall quality and efficiency of health care services, and achieving broader health equity.

(J Med Internet Res 2025;27:e65932) doi:10.2196/65932

XSL•FO
RenderX

KEYWORDS

telemedicine; telehealth; digital health; success factors; challenges

Introduction

Background

Telemedicine enables medical professionals to provide diagnosis and treatment to patients from different locations by using phone or video calls. It is regarded as a strategy for addressing many challenges facing health care service provision, particularly for improving the accessibility of services provided remotely [1]. The World Health Organization defines telemedicine as "the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities" [2]. This definition has been commonly applied in the 10 years since it was first published [3-6]. Telemedicine is one of the main tools in the current development of health systems, with significant potential for diagnosing, treating, and preventing diseases [7]. Different from the broad definition of telehealth, which covers both clinical and nonclinical services including training and continuing medical education for practitioners, telemedicine solely refers to remote clinical services [8]. However, both telemedicine and telehealth are core components of eHealth, a concept that includes all health care practices supported by electronic processes such as information and communication technologies (ICTs) [9,10].

Telemedicine can be implemented in the following 3 models: televisit, telemonitoring, and tele-expertise [11]. Televisits occur between doctors and patients through phone calls, video conferencing software, or secure email systems [11]. Telemonitoring involves using personal health technologies to collect, transmit, evaluate, and communicate patients' personal health data from outside hospitals or clinical departments (eg, patients' homes) and relay them to health care providers [12]. Tele-expertise allows health care professionals to seek medical opinions and guidance from, and share patient data with, experts via phone, video conferencing software, or secure email [13]. According to the care delivery modalities, telemedicine can be divided into synchronous and asynchronous telemedicine [14]. Synchronous telemedicine requires a real-time 2-way interaction between the doctor and patient [1] or between doctors [15]. In asynchronous telemedicine, clinical data elements, such as medical reports, images, and video recordings, are stored and transmitted for later evaluation [16]. Patients send photos of lesions, radiological scans, and lab results to doctors, who respond later in the same manner [17]. Based on these implementation and delivery methods, telemedicine is widely used in medical consultations, patient follow-ups, specialist consultations, drug prescription and delivery, teletriage and screening, and telerehabilitation, among other services [18].

The existing evidence indicates that telemedicine can bring multiple benefits. For patients, the use of telemedicine increases access to care [19], improves care outcomes [20], reduces costs [21], and leads to high patient satisfaction [22,23]. For health care providers, telemedicine saves a significant amount of time and costs [24] and reduces travel time [25]. Health care providers also show high satisfaction with telemedicine [26]. Given the numerous benefits of telemedicine, it is important to understand and identify its key enabling factors and the specific challenges in implementing it.

Key Success Factors

The existing research has focused on factors that enable the implementation and adoption of telemedicine services, including individual, organizational, technological, and economic factors. One individual enabling factor is doctor training and skill-building, which can mitigate the knowledge and skill gaps among health care personnel, enabling them to use telemedicine more easily [27]. Additionally, patients' satisfaction with telemedicine [28] and their positive opinion of digital devices [29] are also key individual factors for effective telemedicine implementation. At the organizational level, coordination between different health care levels and integrating telemedicine into clinicians' workflows [27] can enable the successful implementation of telemedicine services. At the technological level, the availability of technical support, good network coverage, user-friendly applications, and privacy protection are important [30]. Furthermore, in low- and middle-income countries, economic factors may include funding telemedicine projects to purchase equipment and software and pay for mobile services [30]. In high-income countries and regions, establishing a financial framework for telemedicine, financial benefits, and cost savings is a key economic factor [27].

Challenges Facing Telemedicine Implementation and Adoption

On the other hand, barriers to the adoption and implementation of telemedicine have been identified in the literature, including technology and equipment, patient privacy, user attitudes, and reimbursement policies. Common technology and equipment–related barriers include the lack of telemedicine equipment, inability to connect to the internet [31], and lack of access to or comfort with the necessary technology for a video visit [32]. User attitudes toward telemedicine could result in resistance to using telemedicine among both health care providers [33] and patients [34]. Evidence also indicates that the lack of reimbursement for telemedicine services [35] hinders the adoption and implementation of telemedicine.

Efforts in Synthesizing Evidence on Telemedicine

To further understand the current state of the research on telemedicine, we systematically retrieved 305 scoping reviews, systematic reviews, and meta-analysis papers on telemedicine published between 2010 and 2023 from the Scopus database. The search and review of the 305 review articles confirmed that existing studies mainly focused on the following four aspects: (1) the application of telemedicine before and after COVID-19 (n=63, 20.66%), particularly in specific diseases or populations;

XSL•FO
RenderX

(2) evaluating the effectiveness of telemedicine use (n=72, 23.61%), such as reducing patient mortality and improving patient quality of life; (3) the economic evaluation of telemedicine (n=109, 35.74%), such as cost-effectiveness and cost-utility analyses; and (4) examining factors influencing the implementation and adoption of telemedicine (n=43, 14.10%).

Among these scoping reviews, systematic reviews, and meta-analysis papers, the majority focused on synthesizing evidence from studies on telemedicine at the organizational level. Only a few papers synthesized information from studies focusing on understanding factors on the societal level or funding, policy, and regulation (system-level) factors influencing the successful implementation and adoption of telemedicine. System-level factors, beyond the boundaries of individual organizations, specifically refer to the national structures and policy environments, funding, and investment in infrastructure required to deliver telemedicine services, a process that involves collaborative efforts across multiple organizations [36]. Furthermore, a thorough understanding of the challenges facing telemedicine implementation and adoption in regional and rural areas is limited. Therefore, a scoping review was conducted to answer the following questions:

(1) What are the key factors that can influence the implementation and adoption of telemedicine at the system level, particularly in regional and rural areas?

(2) What are the key challenges facing the implementation and adoption of telemedicine at the system level, particularly in regional and rural areas?

Methods

Overview

The systematic review was conducted following the 5-step process described by Arksey and O'Malley [37], chosen for the exploratory and descriptive nature of our study objectives. The keywords and abstract screening, conducted by the authors XL and LH, only included studies in the health context. Authors XL and LH worked on the full-text review and data extraction together, in consideration of their lack of experience in conducting the systematic literature review and data extraction. Author ZL reviewed abstracts that received contradictory assessment outcomes from XL and LH. ZL also performed full-text reviews on the articles when required. A thematic analysis of the factors that influence the implementation of telemedicine was performed by contracted research officer LK. The Covidence systematic literature review platform [38] was used to perform both the abstract and full-text screening.

On December 16, 2023, a literature search was performed in the following 5 databases: CINAHL, Cochrane, Medline, Ovid, and Scopus. Table 1 details the 4 key concepts with associating keywords that were used to conduct the keyword search. Each search strategy includes keywords from concept 1 and either concept 2, 3, or 4. The key concepts were connected using the conjunction "AND," and the keywords under the same key concept were connected using the logical operator "OR." Search terms used were not restricted to the title only. They were found within the title, abstract, or keywords.

Table 1. Keywords and key concepts.

Concept 1	Concept 2	Concept 3	Concept 4
Telemedicine	Factors	Design	Location
• Telehealth • Tele-health • Tele-care • Telemedicine • Tele-medicine • Remote clinical service • Tele-critical care • (same concept use OR)	• Barriers • Challenge • Difficulty • Facilitator • Factors • Failure • Obstacles • Success	• Application • Execution • Financing • Funding • Implementation • Model • Operation • Regulation • Strategy	• District • Nonmetropolitan • Regional • Remote • Rural

Inclusion and Exclusion Criteria

Papers that met all the following criteria were included in the data extraction: (1) published in English in or after 2010; (2) published in a peer-reviewed journal; and (3) research findings that are relevant to 1 or 2 of the research questions for the review. Papers that did not meet all the aforementioned criteria were excluded from the final data extraction.

Data extraction

An Excel (Microsoft Corp) spreadsheet was set up for data extraction with the following categories: (1) year of publication; (2) country where the study was conducted; (3) purpose of the paper; (4) study design; (5) methods for data collection; (6) study population and sector; (7) target population; (8) sample size; (9) response rate; (10) metro or regional/rural location; and (11) factors that influence the adoption or implementation of telemedicine.

Data Synthesis

The data synthesis focused on key factors and challenges and involved a descriptive summary of the included studies. Guided by Braun and Clarke's thematic analysis [39], key themes were identified by examining data patterns and grouping similar concepts. A comparative analysis was performed to explore differences in findings between regional/rural and metropolitan areas. The results were synthesized to identify commonalities and discrepancies in the literature, addressing identified challenges that guided the development of a framework for implementing and adopting telemedicine. To ensure the accuracy of the analytical process, a codebook with descriptive meanings

XSL•FO
RenderX

of each theme was generated and discussed between the authors to refine codes and enhance the depth of the analysis.

Critical Appraisal

In line with the scoping review framework, we did not conduct a critical appraisal of the papers [37].

Results

Study Selection

This study was reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [40]. The completed PRISMA-ScR checklist is available in Multimedia Appendix 1. The PRISMA flowchart (Figure 1) details the review process and search results. The initial key concept and keyword search found 10,691 potentially relevant articles from the 4 databases. Title screening reduced the number to 530 papers, which were uploaded to Covidence [38] for abstract screening. After the abstract screening, 212 articles were included in the full-text review. The full-text review confirmed the inclusion of 89 articles for data extraction. Multimedia Appendix 2 contains a description of the 89 articles.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

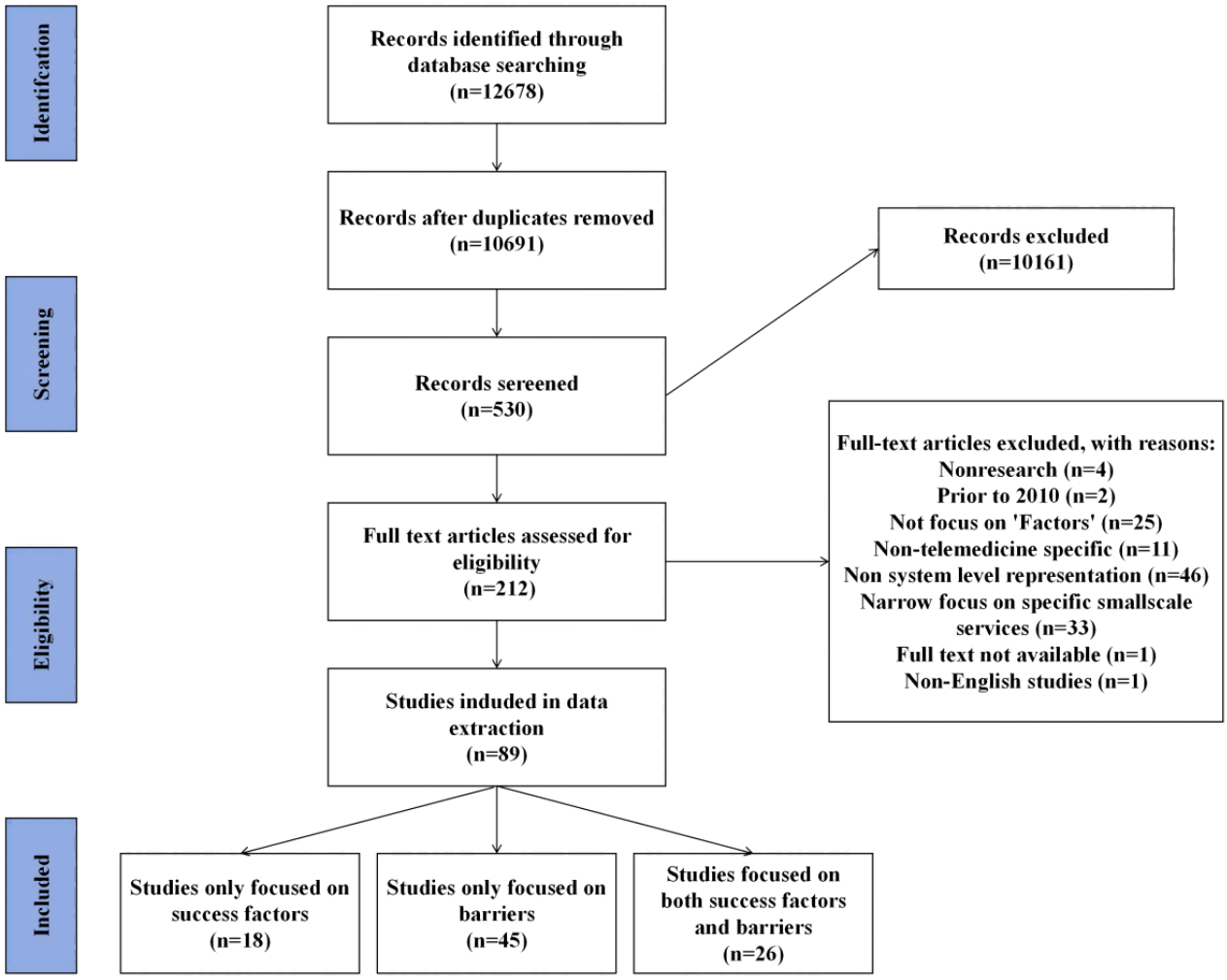

Study Location and Year of Publication

Publication dates ranged from 2010 to 2023 (Multimedia Appendix 3). A total of 66 (74.16%) studies were published between 2020 and 2023. Among them, 4 (4.49%) studies were conducted in multiple countries in the following continents: the Middle East, South America and Europe, Sub-Saharan Africa, and North America. A total of 28 (31.46%) studies were conducted in the United States, 10 (11.24%) in Australia, 5 (5.62%) in China, 3 (3.37%) in Canada, 3 (3.37%) in Ethiopia, 3 (3.37%) in Germany, 3 (3.37%) in the United Kingdom, 2 (2.25%) in Egypt, and 2 (2.25%) in Uganda. Only 1 (1.12%) study was conducted on a global scale, which included 75 countries. Another 25 (28.10%) studies were conducted in 25 different countries. Of the 89 studies, 13 (14.61%) were conducted in metropolitan areas and 16 (17.98%) in regional and rural areas. Another 13 (14.61%) studies were conducted in both metropolitan areas and regional and rural areas.

Sectors and Target Population

A total of 21 (23.60%) studies were conducted in multiple sectors (hospitals, primary health care, nursing homes, aged

XSL•FO
RenderX

care, health authorities, and so on). On the other hand, 55 (61.80%) studies were single-sector studies, including 37 (41.57%) studies in hospitals, 14 (15.73) in primary health care, 2 (2.25%) in community health services, and 2 (2.25%) in mental health. Another 13 (14.61%) studies did not specify the sectors in which they were conducted.

Of the 89 studies, 31 (34.83%) focused on multiple target groups including patients, clinicians, managers, and service providers. The other 58 (65.17%) studies focused on single target groups, including 27 (30.34%) on clinicians, 18 (20.22%) on service providers, 6 (6.74%) on other users (eg, technicians, decision-makers, and medical students), 5 (5.62%) on patients, and 2 (2.25%) on managers.

Methods Used to Identify Factors

A total of 46 (51.69%) and 35 (39.33%) studies adopted quantitative and qualitative approaches, respectively. Among the studies, 10 (11.24%) adopted a mixed methods approach combining both quantitative and qualitative methods. The most common methods for data collection were surveys (n=43, 48.31%), interviews (n=24, 26.97%), and focus group discussions (n=7, 7.87%).

Key Success Factors

Among the identified articles, 44 (49.44%) discussed factors important to the implementation and adoption of telemedicine. More than 70 factors were mentioned in different papers. Based on the similarities and differences, these factors were organized into 6 different categories, as detailed in Textbox 1. The factors were placed into 4 categories—individual, organizational, technological, and economic. Individual factors were subcategorized based on their relevance to clinicians and patients. A category of system-level requirements was also created to accommodate the health system–level factors identified in this review. In total, the review identified 16 (17.98%) studies that focused on regional, rural, and/or remote areas. Factors identified in these studies were compared with the factors associated with metropolitan areas. For each included source of evidence, the detailed data that were charted are displayed in Multimedia Appendix 4.

XSL•FO
RenderX

Textbox 1. Factors important for the success of telemedicine (listed in alphabetical order). The * symbol indicates factors that are common to both metropolitan and nonmetropolitan areas. Factors that are specific to nonmetropolitan areas are italicized. ICT: information and communications technology

System-level requirements

- Established performance requirements and monitoring mechanism
- Improved security and reliability*
- Improvement of digital and health literacy
- National Health ICT index
- Policy and guidelines for the use of telemedicine
- Promotion of telemedicine (to both the workforce and the community)
- Provision of incentives
- Remuneration for service providers
- Robust telehealth program design*
- Telehealth equipment and software applications
- Workforce development
- Additional support and infrastructure investment in rural and remote areas
- Addressing unmet health care needs

Economic considerations and funding

- Cost and benefits sharing between providers and stakeholders
- Financial support and remuneration
- Health insurance coverage and arrangements
- Prior consideration of economic feasibility
- Provision of equipment grants
- Sustained funding allocation (beyond the grant period)

Technological requirements

- Data security and protection
- Efficient and functional network and information-sharing
- Improved ICT infrastructure and support
- Telehealth equipment and software applications
- Up-to-date computers, software, and telecommunication devices
- User friend software and system

Organizational requirements

- Engaging staff in designing the telemedicine goals and processes
- Established performance indicators that measure the benefits of telemedicine to patients, clinicians, and organizations
- Governance and policy
- ICT infrastructure and support
- Improved awareness of the benefits of telemedicine across organizations
- Leadership and management support
- Needs assessment and preparation
- Ongoing monitoring and modification of staff workload
- Patient education and support
- Promotion of telemedicine to both patients and key stakeholders
- Provision of incentives
- Provision of technical support

XSL•FO
RenderX

- Resources and support
- Simplified, standardized, and transparent processes
- Staff capability and competence
- Stakeholder engagement support
- Utilization of existing expertise
- Well-established quality improvement and monitoring processes
- Visible staff champions

Understanding and supporting clinicians

- Being flexible and adaptive to change*
- Digital skills*
- Feeling empowered and efforts are recognized
- Manageable workload and appropriate transition
- Perceived benefits of telemedicine for themselves, patients, and the organization *
- Positive perception and attitudes toward telemedicine*
- Positive teamwork*
- Receiving regular information and updates*
- After-hours staff support
- Encouraged safer clinical practices
- Increased access to relevant resources
- Perceived patient benefits and quality of care

Understanding and improving patients' perceptions

- Access to required telecommunication devices*
- Established confidence and trust in clinicians and the organization
- Increased awareness of telemedicine and its benefits to themselves
- Positive perception and attitudes
- Preexisting relationship with clinicians and organization
- Prior exposure to telemedicine
- Support in improving digital and health literacy
- Timely and accessible support and troubleshooting
- Transparent feedback system*
- Good internet access
- Improved patient-provider communication
- Improved patient experiences and outcomes
- Previous positive experience with telemedicine

Major Challenges Facing Telemedicine Implementation and Adoption

A total of 71 (79.78%) articles identified articles discussed challenges facing the implementation and adoption of telemedicine. More than 50 challenges were mentioned in different papers. These challenges (Textbox 2) were placed into the same categories as those created for the success factors.

XSL•FO
RenderX

Textbox 2. Challenges facing telemedicine (listed in alphabetical order). ICT: information and communications technology.

System-level

- Absence of nationwide integrated telemedicine framework
- Absence of regulatory framework
- Absence of coordination of telemedicine at the national level
- Inadequate efforts in addressing network and data security concerns
- Lack of funding to support organization readiness
- Lack of insurance coverage
- Lack of performance monitoring
- Lack of standardized and specific guidelines for the protection of patient privacy and confidentiality
- Social and cultural concerns
- Unclear reimbursement mechanism
- Unclear or confusing financial arrangements and funding and insurance model

Technological requirements

- High cost of equipment
- Lack of ICT infrastructure
- Lack of integration with other digital/online platforms
- Not user-friendly
- Poor internet access
- Unclear data processing procedure and techniques

Organizational requirements

- Concerns about meeting legal requirements
- High staff turnover and burnout
- Human resources issues, such as high staff turnover and increased training demands
- Inadequate information and data sharing
- Inadequate budgeting and funding to meet increased costs, such as investment in technology upgrades
- Increased administration and human resource support requirements
- Lack of investment in training and upskilling prior to the implementation of telemedicine
- Legal concerns (eg, potential lawsuits because of patient complaints)
- Negative effects on quality of care (eg, long wait time, poor clinician-patient interaction, reduction of patient safety, and lack of physical examination)
- Patient scheduling
- Staff workload issues
- Unclear governance and policy direction
- Unclear performance and outcome measures

Clinicians

- Concerns about patient confidentiality
- Inadequate digital literacy
- Lack of confidence and trust from patients
- Lack of confidence in patient consent and choices
- Lack of familiarity with new equipment, software, and devices
- Patient scheduling
- Resistance from patients

XSL•FO
RenderX

- Resistance from team members
- Responsibility for data security and protection
- Unmanageable workload

Patients

- Affordability of digital devices
- Increased isolation due to the reduced interaction with clinicians and health care organizations
- Language barriers
- Patient resistance due to various factors that were not addressed, such as reduced contact time with clinicians and unclear consultation mechanisms and feedback systems
- Perceived decreased access to care
- Poor internet access

Discussion

Principal Findings

This study fills a knowledge gap by confirming key system-level factors that are critical to the successful implementation and adoption of telemedicine, as well as factors specific to nonmetropolitan areas. Particularly, the study reinforced the importance of 4 success factors that have not been discussed in published systematic review papers. The first factor is the development of a national health ICT index, a comprehensive measure of a country's ICT capabilities, which is crucial for the large-scale implementation of telemedicine [41]. The index includes national ICT infrastructure and access, ICT usage levels, and ICT skills [41]. The second factor is the provision of additional support and investment in building necessary infrastructure in rural and remote areas to enhance telemedicine adoption in these regions [42-44]. The third factor is the need for targeted efforts to raise awareness and promote the benefits of telemedicine among both health care professionals and patients, addressing misconceptions and facilitating adoption [45,46]. The fourth factor is investment in innovation and robust design in telemedicine programs, which can enhance efficiency and sustainability [47-50]. This study also suggests the necessity of developing a nationwide integrated telemedicine framework to improve coordination across national, regional, and local telemedicine initiatives and ensure that national and local policies and strategies are aligned [51]. The factors identified in this scoping review are common to health care organizations, rather than problems specific to individual organizations, and hence may need to be addressed at the system level.

Framework for the Implementation and Adoption of Telemedicine

This scoping review confirms that a successful implementation of telemedicine requires collective efforts at both the system and organizational levels. At the national level, policies and regulations to set the direction, quality expectation, and boundary of telemedicine service provision are necessary. Additionally, investments in improving infrastructure and system-level readiness and addressing potential barriers that may undermine the benefits of telemedicine are required [41]. These efforts can be categorized into 7 core areas, as illustrated in Figure 2—a framework that guides the implementation and adoption of telemedicine proposed by our study. The framework highlights the changes and ongoing efforts required within health systems and health care organizations to help realize the benefits of telemedicine. Continuous improvements to enhance the maturity of health care organizations provide a foundation for innovation and future system and organizational transformations. It is also critical to achieve strategic alignment across the system, between the system and organizations, and across the individual organization internally. The success of telemedicine is a system matter rather than a silo organization business [51].

XSL•FO
RenderX

Figure 2. Framework for the successful implementation of telemedicine. ICT: information and communication technology.

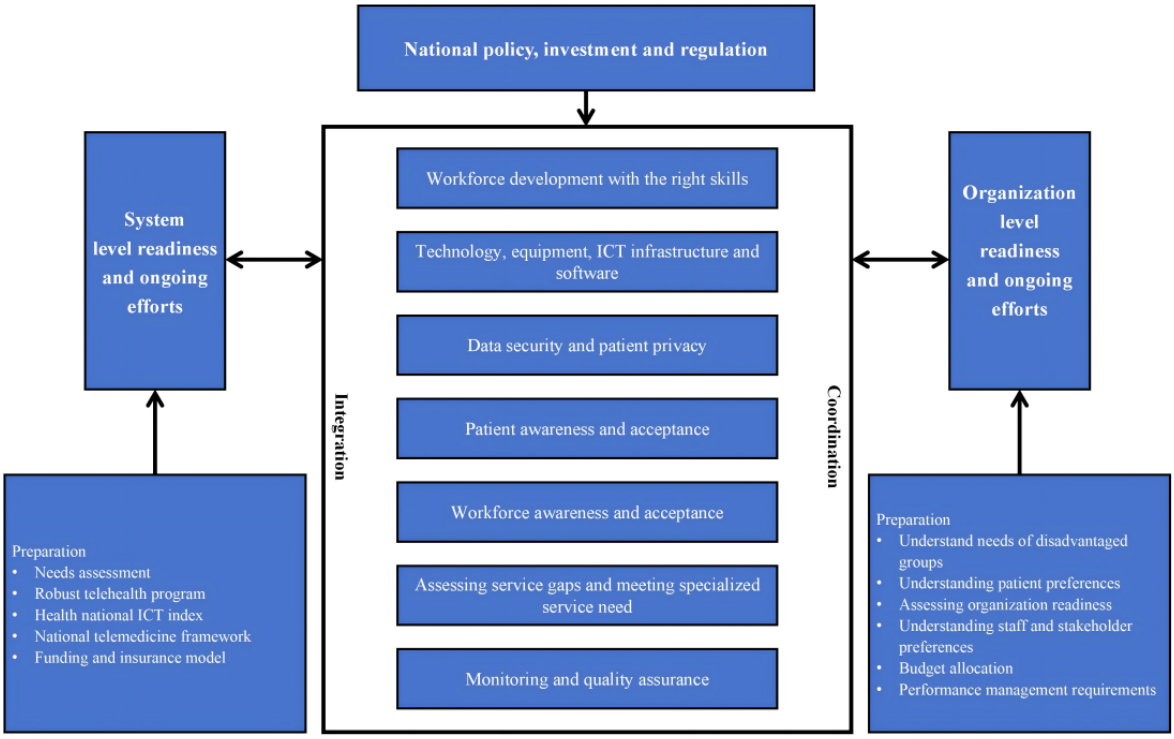

Specifically, the successful implementation of telemedicine necessitates assessment at both system and organizational levels of their respective readiness, as well as the efforts required [36,52]. For a system as complex as health care, different sectors and organizations are interconnected; hence, the implementation of telemedicine is a system-level matter rather than an individual health care organization's sole business [53]. Therefore, the successful implementation of telemedicine requires the adoption of a system integration approach. Organizational requirements; directions; and system-side funding, policy, and regulation must be aligned to enable cooperation and coordination across organizations [36]. Coordination implies the necessity for intraorganizational and interorganizational collaboration and coordination to ensure that key factors that facilitate the successful implementation of telemedicine are addressed or attended to [54].

Furthermore, common factors that impact the implementation of telemedicine cannot be considered in isolation due to the dynamic, bidirectional relationship between factors at the system and organizational levels. Enhancing common factors may influence the readiness and maturity of systems and organizations, providing a foundation for innovation and thereby altering the preparation and effort required for future system-level and organization-level transformations.

As mentioned in the Introduction, there are factors localized to health care organizations that impact the implementation of telemedicine, such as ICT infrastructure, internet connectivity, technical support systems [55], reimbursement, and data security [56]. This scoping review confirms that we must consider these factors at the system level, as they cannot be resolved by a single organization. Hence, support and efforts at both the organization and system level are required. The successful implementation of telemedicine typically requires strong team leadership, suitable training, flexibility, the use of cost-effective and straightforward systems, and implementation within a defined legal and regulatory framework [57]. Addressing organization-level barriers at the system level enables organizations to implement telemedicine successfully.

Improving Awareness, Understanding, and Readiness

One of the key factors emerging in recent studies is resistance to the utilization of telemedicine by both health care professionals and patients, which is common across health care organizations and countries [58-61]. This is partly attributed to health professionals and patients' lack of understanding of telemedicine's benefits [62], patients' fear of losing face-to-face interaction with doctors [63], and the difficulties encountered in utilizing telemedicine [64]. Efforts are required to improve the health workforce and patients' and carers' understanding of the important role and inevitable trend of telemedicine in improving the quality of care. It requires broader government actions to raise public awareness [45] and improve health and digital literacy [44]. Evaluating health systems and organizations' readiness for implementing telemedicine is a critical step that requires adequate preparation, planning, support, and coordination across the health system. This may include developing policies that broadly guide telemedicine design, implementation, and utilization [45,65,66]. Studies also suggest the necessity of offering incentives and financial support to health care organizations in improving infrastructure and enhancing training for health care workers and patients [47,67].

XSL•FO
RenderX

This systemic approach ensures that the unique circumstances of each organization are considered, facilitating a more effective and widespread adoption of telemedicine.

Financial Arrangements, Funding, and Insurance Model

Health insurance coverage, financial arrangements, funding mechanisms, and health insurance models were discussed in 23 (25.84%) identified papers. Additionally, concerns about lower reimbursement rates provided for telemedicine services compared to face-to-face visits, and inappropriate reimbursement strategies were raised [68]. Although some of the factors were discussed in a previous review paper that focused on the organizational level [69], this scoping review explores these factors from a system-level perspective. Unclear or confusing financial arrangements and payment models were identified as key obstacles to the successful adoption of telemedicine [59], as the development of payment models has not kept pace with the rapid expansion of telemedicine.

Internationally, the traditional fee-for-service payment system is still predominantly used to reimburse or pay for telemedicine services [11]. As a result, clinical providers may alter the way services are provided to ensure they receive adequate reimbursement. For instance, some previously free medical services offered by health professionals were modified for new names such as audio-only telemedicine visits or e-visits, to be billed [70]. This non–user-friendly service model discourages patients from accessing services, hindering the development and expansion of telemedicine. Systematically designed payment policies and payment models that support and adequately reimburse service providers are crucial for the continuous development and adoption of telemedicine [71]. The examined studies suggested exploring hybrid models that combine capitation payment and fee-for-service payment [11], bundled payment models [72,73], and other models that are superior to the traditional fee-for-service model to promote telemedicine adoption.

To design a feasible and equitable funding model for telemedicine, mechanisms that support cost and benefit sharing among stakeholders must first be established. For example, a study in Norway [51] shows that telemedicine is primarily supported by internal organizational funds or funded through national or regional research and innovation projects, lacking sustainable long-term financing channels, which affects the continuous development and scaling of telemedicine projects. A feasible and equitable cost-sharing mechanism requires partnership between governments, health care institutions, and other stakeholders, making telemedicine broadly accessible and sustainable. Additionally, mechanisms that enable fair benefit distribution and provide incentives to stakeholders are critical. For example, in the United States, specialized services are often provided by more advanced and well-equipped medical centers or hospitals to patients initially treated and referred by primary care providers. Under the current funding model, the primary care advanced medical centers or hospitals may receive greater economic returns. Medicare only pays a fixed telemedicine service facility fee to the initiating party—primary care institutions—while paying the providing party 100% of the equivalent offline service fees [74]. Additionally, general practitioners may experience increased workloads without proportional financial gain due to inadequate compensation mechanisms [75]. Ultimately, primary care providers, being the least favorable party for benefit distribution, lack the motivation to support telemedicine. The distribution of benefits in telemedicine is thus a complex issue that requires comprehensive consideration and the gradual development of mechanisms to promote equitable redistribution among all stakeholders [51].

Network and Data Security

Previously published systematic reviews confirmed that the stability and security of the telemedicine network, as well as patient and service data, are among the key concerns of telemedicine adoption at health care organizations [56,57]. This review, based on 16 (17.98%) identified papers, confirmed that it is a key systemic factor that hinders the success of telemedicine and must be promptly addressed in and across health care organizations and the health system. Network concerns are related to not only problems of internet availability and connectivity [43,76,77] but also the absence of an integrated nationwide telemedicine network that supports the adoption of telemedicine [51]. Internet availability and connectivity concerns refer to internet availability [43], network latency [76], and poor network connections [77]. These issues pose significant technical challenges to the adoption of telemedicine. On the other hand, the lack of coordination of telemedicine initiatives at the national, regional, and local levels has resulted in the fragmentation of the health care sector, hindering the construction of a nationwide integrated telemedicine network [51].

To build a nationwide integrated telemedicine network, institutions must coordinate and align their strategic plans to national strategies and policies, while ensuring patient quality and safety [51]. Concerns about data security were identified among both health professionals and patients [78]. These concerns are related to low digital literacy levels, measures adapted to enable the secure use of information, and adopted communication technologies [79]. Patient concerns about data security can lead to resistance to using telemedicine, highlighting the need for clinical standards, certifications for service providers and institutions offering telemedicine services, and robust legal frameworks to protect user and institutional data security [7].

Telemedicine in Regional and Rural Areas

Considering health disparities, geographical constraints, and limitations on infrastructure in regional and rural areas, telemedicine is an important initiative that should address the health care needs of the local population and improve service provision by clinicians, including addressing the challenges of workforce shortages [80,81]. Telemedicine presents an opportunity to alleviate the imbalance between health care demand and supply in nonmetropolitan areas by meeting patients' urgent needs for high-quality health care services and addressing disparities in health care resource allocation, diagnostic and treatment capabilities, and quality between urban and rural areas [82]. For patients living in rural and remote areas, telemedicine is becoming critical to enhancing access to

XSL•FO
RenderX

health care specialists [83] and high-quality health care services, such as timely e-consultations and tailored treatment plans [84] from their homes, through telemedicine platforms. On the other hand, telemedicine allows health care providers to receive support and expert advice online, thus directly improving the capacity of health professionals to deliver quality patient care [81]. To fully leverage the potential of telemedicine, comprehensive strategies and substantial investments in infrastructure, along with training for health professionals who provide or receive services via telemedicine platforms, are required [85].

One significant challenge in regional and rural areas is the lack of infrastructure enabling the smooth adoption and implementation of telemedicine. Poor access and unstable internet service have limited the coverage and effectiveness of telemedicine initiatives. Therefore, prioritizing functional broadband access with a sufficient download speed that meets the requirements of telemedicine in rural areas is an essential government/system-level investment [85,86]. Creating a community broadband network may be a good example [86]. It is worth noting that even when broadband is available, the affordability of broadband usage can be a barrier for rural populations [86]. Therefore, government/system-level investments must include relevant policy support and incentives that enable affordable broadband service provision, such as subsidies for broadband deployment and usage, in underserved rural areas [87].

Telemedicine is a core strategy for improving health service availability and accessibility in regional, rural, and remote areas. To maximize its intended benefits, telemedicine must be designed to address the specific health care needs of the servicing areas [43,81] and enhance the support to clinical staff for safer and more effective clinical practices [42], such as remote expert consultation and after-hours staff support [80]. Studies also confirmed the importance of addressing the skepticism and fear of how losing face-to-face interactions between patients and clinicians may affect the quality of care received [88]. Hence, specific strategies should be developed to enhance patient-provider communication in the process of telemedicine implementation and adoption [81].

Limitations

First, this study only searched for English articles, and studies in other languages were excluded, which may be a potential limitation. Second, due to the inability to access and analyze unpublished articles, some important factors might not have been included. Third, only studies about telemedicine related to clinical services were included in this review, and some studies may have been overlooked.

Conclusions

This study reveals the key system-level factors for the successful implementation of telemedicine in health care systems and provides an in-depth insight into factors specific to regional and rural areas. The results show that the widespread implementation of telemedicine is constrained by multifaceted factors, including economy, technology, organization, and individual factors. This study underscores the importance of establishing a national network enhancing public awareness of telemedicine, ensuring clarity in payment and benefit distribution models, and strengthening data security protection measures. In addition, for regional and rural areas, this study highlights the necessity for addressing infrastructural deficiencies, particularly in terms of internet connectivity, and suggests the implementation of targeted incentives and support measures. In summary, this study provides valuable insights for policy makers, health care administrators, and health care providers by highlighting the importance of coordination and collaboration across different regions and organizations. Through these efforts as detailed in the proposed framework, it is possible to promote the popularization of telemedicine, enhance the overall quality and efficiency of health care services, and achieve broader health equity.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 72274225), Guangdong Basic and Applied Basic Research Foundation (grant 2023A1515011725), and China Medical Board Open Competition Project (grant 19-339). The authors would like to acknowledge the work of Lee Koh.

Data Availability

The data used and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 31 KB - jmir_v27i1e65932_app1.docx]

Multimedia Appendix 2
Description of the 89 articles.
[DOCX File , 39 KB - jmir_v27i1e65932_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Year of publication and number of studies.
[DOCX File , 17 KB - jmir_v27i1e65932_app3.docx]

Multimedia Appendix 4
Data extraction.
[XLSX File (Microsoft Excel File), 191 KB - jmir_v27i1e65932_app4.xlsx]

References

1. Rogers DG, Santamaria K, Seng EK, Grinberg AS. Behavioral health, telemedicine, and opportunities for improving access. Curr Pain Headache Rep 2022 Dec 23;26(12):919-926 [FREE Full text] [doi: 10.1007/s11916-022-01096-w] [Medline: 36418847]
2. World Health Organization. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2010.
3. Valk-Draad MP, Bohnet-Joschko S. Nursing home-sensitive hospitalizations and the relevance of telemedicine: a scoping review. Int J Environ Res Public Health 2022 Oct 10;19(19):12944 [FREE Full text] [doi: 10.3390/ijerph191912944] [Medline: 36232255]
4. Kaur KN, Niazi F, Thakur R, Saeed S, Rana S, Singh H. Patient satisfaction for telemedicine health services in the era of COVID-19 pandemic: A systematic review. Front Public Health 2022 Dec 16;10:1031867 [FREE Full text] [doi: 10.3389/fpubh.2022.1031867] [Medline: 36589954]
5. Pengput A, Schwartz DG. Telemedicine in Southeast Asia: a systematic review. Telemed J E Health 2022 Dec 01;28(12):1711-1733. [doi: 10.1089/tmj.2021.0516] [Medline: 35417250]
6. Kruse CS, Molina-Nava A, Kapoor Y, Anerobi C, Maddukuri H. Analyzing the effect of telemedicine on domains of quality through facilitators and barriers to adoption: systematic review. J Med Internet Res 2023 Jan 05;25:e43601 [FREE Full text] [doi: 10.2196/43601] [Medline: 36602844]
7. Cunha AS, Pedro AR, Cordeiro JV. Facilitators of and barriers to accessing hospital medical specialty telemedicine consultations during the COVID-19 pandemic: systematic review. J Med Internet Res 2023 Jul 10;25:e44188 [FREE Full text] [doi: 10.2196/44188] [Medline: 37262124]
8. What is telehealth? How is telehealth different from telemedicine? Office of the National Coordinator for Health Information Technology. URL: https://www.sbmabenefits.com/what-is-the-difference-between-mhealth-ehealth-telehealth-and-telemedicine/ [accessed 2024-12-29]
9. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
10. Bitar H, Alismail S. The role of eHealth, telehealth, and telemedicine for chronic disease patients during COVID-19 pandemic: A rapid systematic review. Digit Health 2021 Apr 19;7:20552076211009396 [FREE Full text] [doi: 10.1177/20552076211009396] [Medline: 33959378]
11. Raes S, Trybou J, Annemans L. How to pay for telemedicine: a comparison of ten health systems. Health Syst Reform 2022 Jan 01;8(1):2116088 [FREE Full text] [doi: 10.1080/23288604.2022.2116088] [Medline: 36084277]
12. Gijsbers H, Feenstra TM, Eminovic N, van Dam D, Nurmohamed SA, van de Belt T, et al. Enablers and barriers in upscaling telemonitoring across geographic boundaries: a scoping review. BMJ Open 2022 Apr 20;12(4):e057494 [FREE Full text] [doi: 10.1136/bmjopen-2021-057494] [Medline: 35443957]
13. Ohannessian R, Yaghobian S, Duong TA, Medeiros de Bustos E, Le Douarin Y, Moulin T, et al. France Is the first country to reimburse tele-expertise at a national level to all medical doctors. Telemed J E Health 2021 Apr 01;27(4):378-381. [doi: 10.1089/tmj.2020.0083] [Medline: 32552392]
14. Siddiqui J, Herchline T, Kahlon S, Moyer KJ, Scott JD, Wood BR, et al. Infectious Diseases Society of America position statement on telehealth and telemedicine as applied to the practice of infectious diseases. Clin Infect Dis 2017 Feb 01;64(3):237-242. [doi: 10.1093/cid/ciw773] [Medline: 28096274]
15. Esterle L, Mathieu-Fritz A. Teleconsultation in geriatrics: impact on professional practice. Int J Med Inform 2013 Aug;82(8):684-695. [doi: 10.1016/j.ijmedinf.2013.04.006] [Medline: 23746716]
16. Sirintrapun SJ, Lopez AM. Telemedicine in cancer care. Am Soc Clin Oncol Educ Book 2018 May 23;38:540-545 [FREE Full text] [doi: 10.1200/EDBK_200141] [Medline: 30231354]
17. Al-Hazmi AM, Sheerah HA, Arafa A. Perspectives on telemedicine during the era of COVID-19; what can Saudi Arabia do? Int J Environ Res Public Health 2021 Oct 11;18(20):10617 [FREE Full text] [doi: 10.3390/ijerph182010617] [Medline: 34682363]
18. Mahmoud K, Jaramillo C, Barteit S. Telemedicine in low- and middle-income countries during the COVID-19 pandemic: a scoping review. Front Public Health 2022;10:914423 [FREE Full text] [doi: 10.3389/fpubh.2022.914423] [Medline: 35812479]

XSL•FO
RenderX

19. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72 [FREE Full text] [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
20. Zhang W, Cheng B, Zhu W, Huang X, Shen C. Effect of telemedicine on quality of care in patients with coexisting hypertension and diabetes: a systematic review and meta-analysis. Telemed J E Health 2021 Jun 01;27(6):603-614. [doi: 10.1089/tmj.2020.0122] [Medline: 32976084]
21. Hayward K, Han S, Simko A, James H, Aldana P. Socioeconomic patient benefits of a pediatric neurosurgery telemedicine clinic. J Neurosurg Pediatr 2019 Oct 25;25(2):204-208. [doi: 10.3171/2019.8.peds1925] [Medline: 31653803]
22. Saeed S, Singhal M, Kaur K, Shannawaz M, Koul A, Arora K, et al. Acceptability and satisfaction of patients and providers with telemedicine during the COVID-19 pandemic: a systematic review. Cureus 2024 Mar;16(3):e56308 [FREE Full text] [doi: 10.7759/cureus.56308] [Medline: 38628988]
23. Williams AM, Bhatti UF, Alam HB, Nikolian VC. The role of telemedicine in postoperative care. Mhealth 2018 May;4:11-11 [FREE Full text] [doi: 10.21037/mhealth.2018.04.03] [Medline: 29963556]
24. Parikh NR, Chang EM, Kishan AU, Kaprealian TB, Steinberg ML, Raldow AC. Time-driven activity-based costing analysis of telemedicine services in radiation oncology. Int J Radiat Oncol Biol Phys 2020 Oct 01;108(2):430-434 [FREE Full text] [doi: 10.1016/j.ijrobp.2020.06.053] [Medline: 32890526]
25. Hendra K, Neemuchwala F, Chan M, Ly NP, Gibb ER. Patient and provider experience with cystic fibrosis telemedicine clinic. Front Pediatr 2021 Nov 24;9:784692 [FREE Full text] [doi: 10.3389/fped.2021.784692] [Medline: 34900879]
26. Speller-Brown B, Carhuas C, Stone A, Connolly M, Martin B, Zhang A, et al. Telemedicine in sickle cell disease: patient, parent, and provider perspectives. Pediatr Blood Cancer 2023 Mar 14;70(3):e30125. [doi: 10.1002/pbc.30125] [Medline: 36518026]
27. Saigí-Rubió F, Borges do Nascimento IJ, Robles N, Ivanovska K, Katz C, Azzopardi-Muscat N, et al. The current status of telemedicine technology use across the World Health Organization European Region: an overview of systematic reviews. J Med Internet Res 2022 Oct 27;24(10):e40877 [FREE Full text] [doi: 10.2196/40877] [Medline: 36301602]
28. Labiris G, Panagiotopoulou EK, Kozobolis VP. A systematic review of teleophthalmological studies in Europe. Int J Ophthalmol 2018 Feb 18;11(2):314-325 [FREE Full text] [doi: 10.18240/ijo.2018.02.22] [Medline: 29487825]
29. Raja M, Bjerkan J, Kymre IG, Galvin KT, Uhrenfeldt L. Telehealth and digital developments in society that persons 75 years and older in European countries have been part of: a scoping review. BMC Health Serv Res 2021 Oct 26;21(1):1157 [FREE Full text] [doi: 10.1186/s12913-021-07154-0] [Medline: 34696789]
30. Dhyani VS, Krishnan JB, Mathias EG, Hossain MM, Price C, Gudi N, et al. Barriers and facilitators for the adoption of telemedicine services in low-income and middle-income countries: a rapid overview of reviews. BMJ Innov 2023 May 11;9(4):215-225. [doi: 10.1136/bmjinnov-2022-001062]
31. Albon D, Van Citters AD, Ong T, Dieni O, Dowd C, Willis A, et al. Telehealth use in cystic fibrosis during COVID-19: Association with race, ethnicity, and socioeconomic factors. J Cyst Fibros 2021 Dec;20 Suppl 3:49-54 [FREE Full text] [doi: 10.1016/j.jcf.2021.09.006] [Medline: 34930543]
32. Haynes SC, Kompala T, Neinstein A, Rosenthal J, Crossen S. Disparities in telemedicine use for subspecialty diabetes care during COVID-19 shelter-in-place orders. J Diabetes Sci Technol 2021 Sep 14;15(5):986-992 [FREE Full text] [doi: 10.1177/1932296821997851] [Medline: 33719622]
33. Xue Y, Liang H, Mbarika V, Hauser R, Schwager P, Kassa Getahun M. Investigating the resistance to telemedicine in Ethiopia. Int J Med Inform 2015 Aug;84(8):537-547. [doi: 10.1016/j.ijmedinf.2015.04.005] [Medline: 25991059]
34. Keptner KM, Heath M. The impact of resistance on telemedicine use for people with disabilities. J Telemed Telecare 2023 Nov 28. [doi: 10.1177/1357633x231213412]
35. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
36. World Health Organization. Framework for the Implementation of a Telemedicine Service. Washington, DC: World Health Organization; 2016.
37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
38. Covidence. URL: https://www.covidence.org/ [accessed 2024-08-29]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
41. Lee S, Begley CE, Morgan R, Chan W, Kim S. m-Health policy readiness and enabling factors: comparisons of Sub-Saharan Africa and Organization for Economic Cooperation and Development Countries. Telemed J E Health 2018 Nov;24(11):908-921 [FREE Full text] [doi: 10.1089/tmj.2017.0278] [Medline: 29432073]
42. Allan J, Webster E, Chambers B, Nott S. "This is streets ahead of what we used to do": staff perceptions of virtual clinical pharmacy services in rural and remote Australian hospitals. BMC Health Serv Res 2021 Dec 04;21(1):1306 [FREE Full text] [doi: 10.1186/s12913-021-07328-w] [Medline: 34863164]

XSL•FO
RenderX

43. Vakkalanka JP, Nataliansyah MM, Merchant KA, Mack LJ, Parsons S, Mohr NM, et al. Telepsychiatry services across an emergency department network: A mixed methods study of the implementation process. Am J Emerg Med 2022 Sep;59:79-84. [doi: 10.1016/j.ajem.2022.06.053] [Medline: 35810736]
44. Wubante SM, Nigatu AM, Jemere AT. Health professionals' readiness and its associated factors to implement Telemedicine system at private hospitals in Amhara region, Ethiopia 2021. PLoS One 2022 Sep 29;17(9):e0275133 [FREE Full text] [doi: 10.1371/journal.pone.0275133] [Medline: 36174016]
45. Ashley C, Williams A, Dennis S, McInnes S, Zwar NA, Morgan M, et al. Telehealth's future in Australian primary health care: a qualitative study exploring lessons learnt from the COVID-19 pandemic. BJGP Open 2023 Feb 07;7(2):BJGPO.2022.0117. [doi: 10.3399/bjgpo.2022.0117]
46. Moroz I, Archibald D, Breton M, Cote-Boileau E, Crowe L, Horsley T, et al. Key factors for national spread and scale-up of an eConsult innovation. Health Res Policy Syst 2020 Jun 03;18(1):57 [FREE Full text] [doi: 10.1186/s12961-020-00574-0] [Medline: 32493357]
47. Banbury A, Smith AC, Mehrotra A, Page M, Caffery LJ. A comparison study between metropolitan and rural hospital-based telehealth activity to inform adoption and expansion. J Telemed Telecare 2021 Mar 26;29(7):540-551. [doi: 10.1177/1357633x21998201]
48. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e18426 [FREE Full text] [doi: 10.2196/18426] [Medline: 33095175]
49. Berry CA, Kwok L, Gofine M, Kaufman M, Williams DA, Terlizzi K, et al. Utilization and staff perspectives on an on-demand telemedicine model for people with intellectual and developmental disabilities who reside in certified group residences. Telemed Rep 2023 Aug 01;4(1):204-214 [FREE Full text] [doi: 10.1089/tmr.2023.0024] [Medline: 37529771]
50. Zailani S, Gilani MS, Nikbin D, Iranmanesh M. Determinants of telemedicine acceptance in selected public hospitals in Malaysia: clinical perspective. J Med Syst 2014 Sep 20;38(9):111. [doi: 10.1007/s10916-014-0111-4] [Medline: 25038891]
51. Alami H, Gagnon MP, Wootton R, Fortin JP, Zanaboni P. Exploring factors associated with the uneven utilization of telemedicine in Norway: a mixed methods study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):180 [FREE Full text] [doi: 10.1186/s12911-017-0576-4] [Medline: 29282048]
52. Jennett P, Yeo M, Pauls M, Graham J. Organizational readiness for telemedicine: implications for success and failure. J Telemed Telecare 2003 Dec 01;9 Suppl 2(2_suppl):S27-S30. [doi: 10.1258/135763303322596183] [Medline: 14728753]
53. Khodadad-Saryazdi A. Exploring the telemedicine implementation challenges through the process innovation approach: A case study research in the French healthcare sector. Technovation 2021:A [FREE Full text] [doi: 10.1016/j.technovation.2021.102273]
54. Broens THF, Huis IVRMHA, Vollenbroek-Hutten MMR, Hermens HJ, van HAT, Nieuwenhuis LJM. Determinants of successful telemedicine implementations: a literature study. J Telemed Telecare 2007 Jun;13(6):303-309. [doi: 10.1258/135763307781644951] [Medline: 17785027]
55. Sagaro GG, Battineni G, Amenta F. Barriers to sustainable telemedicine implementation in Ethiopia: a systematic review. Telemed Rep 2020 Nov 01;1(1):8-15 [FREE Full text] [doi: 10.1089/tmr.2020.0002] [Medline: 35722252]
56. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
57. Saliba V, Legido-Quigley H, Hallik R, Aaviksoo A, Car J, McKee M. Telemedicine across borders: a systematic review of factors that hinder or support implementation. Int J Med Inform 2012 Dec;81(12):793-809. [doi: 10.1016/j.ijmedinf.2012.08.003] [Medline: 22975018]
58. Elkefi S, Layeb S. Telemedicine's future in the post-Covid-19 era, benefits, and challenges: a mixed-method cross-sectional study. Behav Inf Technol 2022 Oct 26;42(15):2639-2653. [doi: 10.1080/0144929x.2022.2137060]
59. Onsongo S, Kamotho C, Rinke de Wit TF, Lowrie K. Experiences on the utility and barriers of telemedicine in healthcare delivery in Kenya. Int J Telemed Appl 2023 May 3;2023:1487245-1487210 [FREE Full text] [doi: 10.1155/2023/1487245] [Medline: 37180825]
60. Le Bras A, Zarca K, Mimouni M, Durand-Zaleski I. Implementing technologies: assessment of telemedicine experiments in the Paris region: reasons for success or failure of the evaluations and of the deployment of the projects. Int J Environ Res Public Health 2023 Feb 09;20(4):3031 [FREE Full text] [doi: 10.3390/ijerph20043031] [Medline: 36833723]
61. Nduka SO, Nwaodu MA, Nduka IJ. Telepharmacy services in a developing country: Nigerian community pharmacists' and patients' perspectives on the clinical benefits, cost, and challenges. Telemed J E Health 2023 Aug 01;29(8):1238-1251. [doi: 10.1089/tmj.2022.0385] [Medline: 36594935]
62. Khalid A, Dong Q, Chuluunbaatar E, Haldane V, Durrani H, Wei X. Implementation science perspectives on implementing telemedicine interventions for hypertension or diabetes management: scoping review. J Med Internet Res 2023 Mar 14;25:e42134 [FREE Full text] [doi: 10.2196/42134] [Medline: 36917174]
63. Li W, Liu W, Liu S, Li J, Wang W, Li K. Perceptions of patients with chronic obstructive pulmonary disease towards telemedicine: A qualitative systematic review. Heart Lung 2021 Sep;50(5):675-684. [doi: 10.1016/j.hrtlng.2021.03.081] [Medline: 34107391]

XSL•FO
RenderX

64. Tabaeeian RA, Hajrahimi B, Khoshfetrat A. A systematic review of telemedicine systems use barriers: primary health care providers' perspective. J Sci Technol Policy Manag 2022 Dec 22;15(3):610-635. [doi: 10.1108/jstpm-07-2021-0106]
65. Luciano E, Mahmood MA, Mansouri Rad P. Telemedicine adoption issues in the United States and Brazil: Perception of healthcare professionals. Health Informatics J 2020 Dec 19;26(4):2344-2361 [FREE Full text] [doi: 10.1177/1460458220902957] [Medline: 32072843]
66. Ray KN, Felmet KA, Hamilton MF, Kuza CC, Saladino RA, Schultz BR, et al. Clinician attitudes toward adoption of pediatric emergency telemedicine in rural hospitals. Pediatr Emer Care 2017;33(4):250-257. [doi: 10.1097/pec.0000000000000583]
67. Casillas A, Valdovinos C, Wang E, Abhat A, Mendez C, Gutierrez G, et al. Perspectives from leadership and frontline staff on telehealth transitions in the Los Angeles safety net during the COVID-19 pandemic and beyond. Front Digit Health 2022 Aug 9;4:944860 [FREE Full text] [doi: 10.3389/fdgth.2022.944860] [Medline: 36016601]
68. Theis RP, Dorbu JI, Mavrodieva ME, Guerrero RA, Wright SE, Donahoo WT, et al. Telehealth implementation response to COVID-19 in the OneFlorida+ clinical research network: perspectives of clinicians and health systems leaders. Telemed J E Health 2024 Jan 26;30(1):268-277. [doi: 10.1089/tmj.2023.0029] [Medline: 37358611]
69. Harst L, Timpel P, Otto L, Richter P, Wollschlaeger B, Winkler K, et al. Identifying barriers in telemedicine-supported integrated care research: scoping reviews and qualitative content analysis. J Public Health (Berl) 2019 Apr 09;28(5):583-594. [doi: 10.1007/s10389-019-01065-5]
70. Adler-Milstein J, Mehrotra A. Paying for digital health care — problems with the fee-for-service system. N Engl J Med 2021 Sep 02;385(10):871-873. [doi: 10.1056/nejmp2107879]
71. Bajowala SS, Milosch J, Bansal C. Telemedicine pays: billing and coding update. Curr Allergy Asthma Rep 2020 Jul 27;20(10):60 [FREE Full text] [doi: 10.1007/s11882-020-00956-y] [Medline: 32715353]
72. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/nejmsr1503323]
73. Zhao M, Hamadi H, Haley DR, Xu J, White-Williams C, Park S. Telehealth: advances in alternative payment models. Telemed J E Health 2020 Dec;26(12):1492-1499. [doi: 10.1089/tmj.2019.0294] [Medline: 32101100]
74. Xiao G, Xu H. Analysis and enlightenment of Medicare payment Internet+medical service policy in America. Soft Sci Health 2022;36(8):90-96 This is a Chinese language article. [doi: 10.3969/j.issn.1003-2800.2022.08.018]
75. Moffatt J, Eley DS. Barriers to the up-take of telemedicine in Australia - a view from providers. Rural Remote Health 2011 Feb 10. [doi: 10.22605/rrh1581]
76. Hansen RN, Saour BM, Serafini B, Hannaford B, Kim L, Kohno T, et al. Opportunities and barriers to rural telerobotic surgical health care in 2021: report and research agenda from a stakeholder workshop. Telemed J E Health 2022 Jul 01;28(7):1050-1057 [FREE Full text] [doi: 10.1089/tmj.2021.0378] [Medline: 34797741]
77. Crotty BH, Hyun N, Polovneff A, Dong Y, Decker MC, Mortensen N, et al. Analysis of clinician and patient factors and completion of telemedicine appointments using video. JAMA Netw Open 2021 Nov 01;4(11):e2132917 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.32917] [Medline: 34735013]
78. Wade VA, Eliott JA, Hiller JE. A qualitative study of ethical, medico-legal and clinical governance matters in Australian telehealth services. J Telemed Telecare 2012 Mar 20;18(2):109-114. [doi: 10.1258/jtt.2011.110808] [Medline: 22267306]
79. Těšinová JK, Dobiášová K, Dušek Z, Tobiášová A. Development of telemedicine in the Czech Republic from patients' and other key stakeholders' perspective. Front Public Health 2023 Oct 23;11:1202182 [FREE Full text] [doi: 10.3389/fpubh.2023.1202182] [Medline: 37937075]
80. Gutierrez J, Moeckli J, McAdams N, Kaboli PJ. Perceptions of telehospitalist services to address staffing needs in rural and low complexity hospitals in the Veterans Health Administration. J Rural Health 2020 Jun 15;36(3):355-359. [doi: 10.1111/jrh.12403] [Medline: 31840307]
81. Nyoni T, Evers EC, Pérez M, Jeffe DB, Fritz SA, Colditz GA, et al. Perceived barriers and facilitators to the adoption of telemedicine infectious diseases consultations in southeastern Missouri hospitals. J Telemed Telecare 2023 Jan 19;30(9):1462-1474. [doi: 10.1177/1357633x221149461]
82. Gao J, Fan C, Chen B, Fan Z, Li L, Wang L, et al. Telemedicine is becoming an increasingly popular way to resolve the unequal distribution of healthcare resources: evidence from China. Front Public Health 2022 Jul 6;10:916303 [FREE Full text] [doi: 10.3389/fpubh.2022.916303] [Medline: 35874991]
83. Parmar P, Mackie D, Varghese S, Cooper C. Use of telemedicine technologies in the management of infectious diseases: a review. Clin Infect Dis 2015 Apr 01;60(7):1084-1094. [doi: 10.1093/cid/ciu1143] [Medline: 25516192]
84. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021 Aug 26;23(8):e29575 [FREE Full text] [doi: 10.2196/29575] [Medline: 34435965]
85. Hirko K, Kerver J, Ford S, Szafranski C, Beckett J, Kitchen C, et al. Telehealth in response to the COVID-19 pandemic: Implications for rural health disparities. J Am Med Inform Assoc 2020 Nov 01;27(11):1816-1818 [FREE Full text] [doi: 10.1093/jamia/ocaa156] [Medline: 32589735]
86. Graves J, Abshire D, Amiri S, Mackelprang J. Disparities in technology and broadband internet access across rurality. Fam Community Health 2021;44(4):257-265. [doi: 10.1097/fch.0000000000000306] [Medline: 34269696]

XSL•FO
RenderX

87. Prieger JE. The broadband digital divide and the economic benefits of mobile broadband for rural areas. Telecommun Policy 2013 Jul;37(6-7):483-502. [doi: 10.1016/j.telpol.2012.11.003]
88. Clure C, Sheeder J, Teal S, Cohen R. Telemedicine to improve reproductive health care for rural Coloradans: Perceptions of interest and access. J Rural Health 2023 Jan 13;39(1):172-178. [doi: 10.1111/jrh.12697] [Medline: 35829619]

Abbreviations

ICT: information and communication technology

Edited by A Mavragani; submitted 30.08.24; peer-reviewed by K Baudin, C McCabe; comments to author 19.12.24; revised version received 07.01.25; accepted 04.02.25; published 05.03.25.

Please cite as:
Li X, Huang L, Zhang H, Liang Z
Enabling Telemedicine From the System-Level Perspective: Scoping Review
J Med Internet Res 2025;27:e65932
URL: https://www.jmir.org/2025/1/e65932
doi:10.2196/65932
PMID:40053725

©Xuezhu Li, Lifeng Huang, Hui Zhang, Zhanming Liang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Natural Language Processing Technologies for Public Health in Africa: Scoping Review

Songbo Hu[1*], MPhil; Abigail Oppong[1*], BSc; Ebele Mogo[2], DrPH; Charlotte Collins[1], PhD; Giulia Occhini[1], PhD; Anna Barford[1], PhD; Anna Korhonen[1], PhD

[1]Language Technology Lab, University of Cambridge, Cambridge, United Kingdom

[2]Cambridge Centre for Human Inspired Artificial Intelligence, University of Cambridge, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Songbo Hu, MPhil
Language Technology Lab
University of Cambridge
English Faculty Building
9 West Road
Cambridge, CB3 9DA
United Kingdom
Phone: 44 1223 335010
Email: sh2091@cam.ac.uk

Abstract

Background: Natural language processing (NLP) has the potential to promote public health. However, applying these technologies in African health systems faces challenges, including limited digital and computational resources to support the continent's diverse languages and needs.

Objective: This scoping review maps the evidence on NLP technologies for public health in Africa, addressing the following research questions: (1) What public health needs are being addressed by NLP technologies in Africa, and what unmet needs remain? (2) What factors influence the availability of public health NLP technologies across African countries and languages? (3) What stages of deployment have these technologies reached, and to what extent have they been integrated into health systems? (4) What measurable impact has these technologies had on public health outcomes, where such data are available? (5) What recommendations have been proposed to enhance the quality, cost, and accessibility of health-related NLP technologies in Africa?

Methods: This scoping review includes academic studies published between January 1, 2013, and October 3, 2024. A systematic search was conducted across databases, including MEDLINE via PubMed, ACL Anthology, Scopus, IEEE Xplore, and ACM Digital Library, supplemented by gray literature searches. Data were extracted and the NLP technology functions were mapped to the World Health Organization's list of essential public health functions and the United Nations' sustainable development goals (SDGs). The extracted data were analyzed to identify trends, gaps, and areas for future research. This scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) reporting guidelines, and its protocol is publicly available.

Results: Of 2186 citations screened, 54 studies were included. While existing NLP technologies support a subset of essential public health functions and SDGs, language coverage remains uneven, with limited support for widely spoken African languages, such as Kiswahili, Yoruba, Igbo, and Zulu, and no support for most of Africa's >2000 languages. Most technologies are in prototyping phases, with only one fully deployed chatbot addressing vaccine hesitancy. Evidence of measurable impact is limited, with 15% (8/54) studies attempting health-related evaluations and 4% (2/54) demonstrating positive public health outcomes, including improved participants' mood and increased vaccine intentions. Recommendations include expanding language coverage, targeting local health needs, enhancing trust, integrating solutions into health systems, and adopting participatory design approaches. The gray literature reveals industry- and nongovernmental organizations–led projects focused on deployable NLP applications. However, these projects tend to support only a few major languages and specific use cases, indicating a narrower scope than academic research.

XSL•FO
RenderX

Conclusions: Despite growth in NLP research for public health, major gaps remain in deployment, linguistic inclusivity, and health outcome evaluation. Future research should prioritize cross-sectoral and needs-based approaches that engage local communities, align with African health systems, and incorporate rigorous evaluations to enhance public health outcomes.

International Registered Report Identifier (IRRID): RR2-doi:10.1101/2024.07.02.24309815

(J Med Internet Res 2025;27:e68720) doi:10.2196/68720

KEYWORDS

public health; global health; health promotion; essential public health functions; Africa; natural language processing; artificial intelligence; machine learning; technology; mobile phone

Introduction

Public Health Needs in Africa

Most African countries face major challenges in meeting the sustainable development goal (SDG) 3 targets for good health and well-being [1-3]. Key public health challenges include high rates of infectious diseases, maternal and child health inequities, and a growing burden of noncommunicable diseases, alongside the critical need for resilient emergency response systems [4,5].

Some of these challenges stem from acute shortages in the health workforce and weak public health surveillance systems, among other weaknesses in public health systems [6,7]. For instance, Africa has only 1400 epidemiologists, despite an estimated need for 6000 [4]. These issues are further amplified by structural weaknesses in health systems and insufficient multisectoral coordination for health [5,8], which are particularly exposed by public health emergencies [9], such as the COVID-19 pandemic and the mpox outbreaks. During the COVID-19 pandemic, while several African countries were able to rapidly leverage their past experiences with outbreaks to respond to COVID-19, they also faced challenges, such as inadequate adherence to infection control, insufficient personal protective equipment, poor contact tracing, supply chain shortages, and a lack of training for key personnel [10].

To systematically strengthen public health capacities, the World Health Organization (WHO) has outlined 12 essential public health functions (EPHFs) [11]. These functions include a broad range of activities, from disease surveillance and health promotion to emergency preparedness and equitable access to health care services. However, many countries across the region, especially those with lower income levels (note that in 2024, Africa comprises upper-middle–income, lower-middle–income, and low-income countries [12]), face substantial challenges in fully implementing these functions, mainly because of limited financial, infrastructural, and health care workforce resources [5]. In resource-constrained settings, innovative technologies, such as artificial intelligence (AI) technologies, could play a crucial role in supporting the implementation of EPHFs [5,13], thereby improving public health outcomes and advancing progress toward achieving health-related SDGs.

Natural Language Processing Technologies for Public Health

Natural language processing (NLP) is a vibrant interdisciplinary field within AI research, known by various terms in different disciplines, such as NLP in computer science, computational linguistics in linguistics, speech recognition in engineering, computational psycholinguistics in psychology, and language technologies in public discourse [14]. Despite the diversity in terminologies and research focuses within these disciplines, they share the common goal of enabling computers to interpret, understand, and generate human language [14,15]. NLP allows computers to perform a wide range of language-based tasks, including facilitating human-machine communication; improving human-to-human interactions; and processing text and speech data for practical NLP applications across different sectors, including public health.

NLP holds significant potential for advancing public health in Africa by addressing the ongoing challenges faced by many countries. By appropriately leveraging NLP, countries can improve health communication, enhance disease surveillance, support workforce training, and optimize limited resources [2,16,17], all of which are crucial for achieving SDG 3 targets. NLP technologies can be used to process and analyze large volumes of health data from diverse sources, including social media, medical records, and public health reports, to identify emerging health threats and track disease patterns in real time. This capability is especially valuable in regions with limited health workforce and surveillance infrastructure, as it enables faster, data-driven responses to public health emergencies.

In practice, NLP-driven tools have already shown promise in Africa. For instance, during the COVID-19 pandemic, WhatsApp chatbots in South Africa, Rwanda, and Senegal were used to disseminate reliable information and facilitate rapid COVID-19 testing, while a Telegram-based chatbot in Ghana was developed to combat misinformation and provide accurate data to the public [18]. Such tools can bridge communication gaps by delivering health information in local languages, empowering communities to recognize symptoms, prevent disease transmission, and respond more effectively. These innovations could play a transformative role in strengthening health systems across the continent, making them more resilient and responsive to both everyday health needs and unexpected crises.

Africa has made progress toward achieving some of its innovation and technology targets [19]. Specifically, the continent is making strides in mobile network coverage, with approximately 89% of the total African population now having access to mobile networks. Countries like Mali, Namibia, and Guinea-Bissau have achieved 100% 2G mobile network coverage [20]. This expanding network coverage creates new opportunities for cloud-based NLP applications in public health.

XSL•FO
RenderX

Cloud computing, which uses remote servers to store, manage, and process data, allows African countries to access computing power that was previously unattainable [13]. This scalability is crucial for deploying NLP-based health solutions in resource-constrained settings where local infrastructure may be insufficient or absent. The synergy of cloud technology and increasing network accessibility opens the door to the expansion of NLP technologies in Africa, providing promising opportunities to improve public health outcomes across the continent.

However, a primary obstacle to the development of NLP applications is a lack of essential digital datasets for the >2000 languages spoken on the continent [21,22]. The development of modern NLP-based health applications for African language communities requires large-scale datasets to fully unlock the capabilities of deep learning models; however, there is a scarcity of digitized, in-language (ie, datasets collected in the specific languages spoken by the target user of the NLP technology), and in-domain (ie, datasets tailored to a specific use case or application, such as health education or disease surveillance, rather than general-purpose language) data. This scarcity is particularly profound in the health sector, where data tailored to specific African languages and contexts are often completely absent. This conjunction of linguistic diversity and data scarcity creates significant obstacles to developing effective NLP technologies tailored to Africa's specific public health needs. Moreover, even when NLP technologies are developed, their successful deployment, validation, and integration into existing health systems are critical for achieving a meaningful positive impact. In resource-constrained environments, lessons learned from previous experience suggest that NLP technologies should be integrated into existing systems and institutions, rather than aiming to replace them [23]. This requires overcoming various obstacles, including the development of a nuanced understanding of local public health needs, the creation of sustainable and scalable solutions, and ensuring equitable access for all users [13,23-25].

Research Gaps

Most previous reviews related to NLP technologies in public health have focused on global health [2,16,26] or low- and middle-income countries [13] as a whole, examined AI applications without focusing specifically on NLP [2,13,16,27], or focused on a single type of NLP application, such as chatbots [18]. Unlike other AI technologies, NLP applications are heavily influenced by the languages and cultures they are designed to serve [28,29]. Given Africa's vast linguistic diversity and the complex spectrum of public health challenges faced by countries in the region, an Africa-focused review is critical for a more nuanced understanding of how NLP can be tailored to meet the diverse health needs across the continent. This approach aligns with pan-African initiatives, such as the African Union's Agenda 2063 [30], which seeks to address health challenges and promote collaboration across borders. At the same time, modern NLP technologies often share similar development paradigms, meaning that advancements in one type of application can provide valuable insights and sometimes resources to others. These benefits extend beyond the experience gained during application development and include the shared use of digital resources across applications, often improving performance through NLP's *transfer learning* techniques [31]. Therefore, a broader review of NLP technologies, compared to one focused on a specific application, provides researchers and developers with a more comprehensive set of evidence to guide future development.

To the best of our knowledge, this is the first scoping review to comprehensively examine the application of NLP technologies to public health in Africa. By mapping the current evidence, this review aims to provide insights into the key barriers and opportunities for the development and deployment of these technologies. Specifically, the review aims to answer five main research questions: (1) *Needs and availability:* What public health needs are being addressed by NLP technologies in Africa, and what unmet needs remain? (2) *Prevalence and distribution:* What factors influence the availability of public health NLP technologies across African countries and languages? (3) *Deployment and integration:* What stages of deployment have these technologies reached, and to what extent have they been integrated into health systems? (4) *Public health impact:* What measurable impact has these technologies had on public health outcomes, where such data are available? (5) *Outlook:* What recommendations have been proposed to enhance the quality, cost, and accessibility of health-related NLP technologies in Africa? By answering these questions, the review aims to provide actionable recommendations for future research and development.

Methods

Overview

The paradigms of developing NLP technologies have evolved significantly since the origin of NLP in the 1940s. The early rule-based systems, such as ELIZA [32], were followed by a shift toward machine learning–based methods in the 1990s. This new approach leveraged large datasets, reducing reliance on manually crafted rules. In 2013, the introduction of Word2Vec marked a major milestone for NLP [33], by representing words as vectors. This approach formed the foundation for neural language models. Subsequently, pretrained language models, such as BERT (Bidirectional Encoder Representations from Transformers) [34] and GPT-2 (Generative Pre-trained Transformer 2) [31], have become the backbone for NLP development, allowing systems to be fine-tuned and developed using datasets of thousands of examples.

Recent advances in large language models [35,36] have further simplified NLP development, allowing systems to achieve optimal performance after learning on just a handful of task-specific examples.

In the context of public health in Africa, the development of NLP technologies will likely use a mixture of paradigms, depending on the availability of task-specific datasets and computational resources. For the purposes of this scoping review, we define *NLP technologies* broadly to include *any computational systems that process natural language, either as input or output*. This inclusive definition ensures that the review captures a wide range of applications in public health across

XSL•FO
RenderX

Africa. In Multimedia Appendix 1 [18,37-108], we provide examples of technologies that fall within or outside the scope of this review. The scoping review maps the current evidence on NLP technologies within the framework of EPHFs. The review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines [109] (see the PRISMA-ScR checklist in Multimedia Appendix 2), and its review protocol is available on *medRxiv* [110].

Search Strategy

A systematic search of *academic literature* was conducted on May 13, 2024, and updated on October 3, 2024, using the following five electronic bibliographic databases:

1. *MEDLINE via PubMed:* Medical and public health literature
2. *ACL Anthology:* NLP and language science literature
3. *Scopus:* Broad interdisciplinary scope, including medical research
4. *IEEE Xplore:* Engineering literature, particularly in NLP and health informatics
5. *ACM Digital Library:* Computing literature, including NLP and health informatics

The search included studies published from January 1, 2013, to October 3, 2024, with the aim being to capture recent developments in NLP, particularly after the introduction of neural language models in 2013 [33]. No language restrictions were applied, although the search terms were in English.

Search terms were developed around three key areas:

1. *Africa:* The names of all 55 African Union member countries and African languages with >1 million native speakers
2. *Public health:* On the basis of the 12 EPHFs outlined by the WHO
3. *NLP:* As suggested by a team of experts in the field

These terms were combined with general phrases and Medical Subject Headings. The search strategy for each database was tailored with database-specific features to enhance the retrieval of relevant studies. The complete search strategy for MEDLINE (PubMed) and the full list of search terms are detailed in Multimedia Appendix 1. Reference chaining of relevant articles was also conducted.

Discussion of NLP technologies for public health in Africa occurs beyond the academic literature, spanning a diverse array of contributors, formats, and outlets. As such, sparse academic literature on this topic does not necessarily indicate a lack of progress [13]. Many promising technologies addressing public health challenges in Africa are introduced through media outlets, as well as by the individuals, companies, governments, and nongovernmental organizations (NGOs) that develop and use them. These contributions are often presented on the web or shared at events, such as conferences. Therefore, this scoping review also mapped evidence from a broad gray literature. In addition to structured academic databases, the following gray literature sources were included: (1) preprints, non–peer-reviewed studies, and reports; (2) media articles and blog posts; (3) commercial products from startups and established companies; (4) initiatives led by NGOs; and (5) proceedings and presentations from events and conferences. The complete search strategy for gray literature is available in Multimedia Appendix 1.

Screening and NLP Technology Selection Criteria

This review includes NLP technologies designed to support public health in Africa. In addition, we consider digital and computational resources essential for the development and deployment of these technologies, such as digital datasets, hardware, and software toolkits.

The selection of sources for this scoping review followed a systematic 2-step screening process. Initially, titles and abstracts were reviewed by one reviewer (AO) to exclude studies meeting any of the predefined exclusion criteria (Textbox 1), such as those not involving NLP technologies, unrelated to public health, or lacking a focus on Africa. Studies that passed this initial screening were then subjected to full-text screening, where studies meeting all the inclusion criteria (Textbox 1) were included. The full-text screening was conducted by the same reviewer (AO), and the reasons for exclusion at this stage were documented.

Before the formal screening process, pilot screenings were conducted to refine our screening guidelines (Multimedia Appendix 1) and to ensure consistency and accuracy in study selection. In these pilots, 10% (179/1791 for title and abstract screening and 36/361 for full-text screening) of candidate papers were randomly selected and independently reviewed by 2 trained reviewers (AO and CC) following the predefined screening guidelines. Interreviewer agreement was assessed using exact match rate and Cohen κ to ensure reliability. The screening process and guidelines were iteratively refined by a review coordinator (SH), and the pilot screening was repeated until both an exact match rate of 0.9 and a Cohen κ score of 0.8, indicating almost perfect agreement, were achieved. Following the pilot, formal screening was conducted by a single reviewer (AO), with any concerns resolved by a review coordinator (SH), a subject matter expert in NLP.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- All types of scientific publications aimed at an academic audience (eg, peer-reviewed articles, conference proceedings, and book chapters); for gray literature search, other web-based publications (eg, blog posts and media outlets)
- Studies focusing on the development, evaluation, or adaptation of natural language processing (NLP) technologies specifically promoting public health
- Studies demonstrating direct or indirect relevance to the population in the continent of Africa
- Studies published between January 1, 2013, and October 3, 2024
- Studies published in any language

Exclusion criteria

- Articles without full-text availability; for articles not available through Cambridge University libraries, full text was requested by emailing the authors
- Studies unrelated to NLP technologies or their application to public health; for example, non-NLP applications, where no language technologies were involved, and the technology was used to perform tasks, such as predicting outcomes solely from structured datasets or images
- Studies focused on non-African contexts, except where such studies offer comparative insights relevant to African NLP technologies
- Studies published before January 1, 2013, or after October 3, 2024
- No language requirement specified

Data Extraction and Synthesis

Data were extracted for each included study based on a predefined data extraction template (Multimedia Appendix 1), capturing key information on study descriptions, the characterization of NLP technologies (Textbox 2), and their contributions to EPHFs, SDGs, and SDG 3 targets specifically. In addition, where such data were available, any public health outcomes measured and recommendations for future development were documented. One reviewer (AO) completed the data extraction for all included studies, with any concerns resolved through team discussions. Due to the heterogeneity of study goals, methodologies, evaluation methods, and outcomes, a formal meta-analysis was not attempted. Instead, a narrative synthesis of the results was conducted, with introduced NLP technologies characterized according to the categories outlined in the data extraction template. The extracted data were analyzed to identify trends, gaps, and areas for future research. In addition, the authors' recommendations for future development were documented and summarized to provide guidance for advancing NLP technologies in public health.

A similar pipelined approach of screening, data extraction, and synthesis was applied to the gray literature. Given that our gray literature search covered sources beyond academic publications, we omitted undisclosed data extraction items, as commercial products often lack full disclosure of their design and implementation. A detailed description of our approach to identifying and synthesizing the gray literature is available in Multimedia Appendix 1.

XSL•FO
RenderX

Textbox 2. Selected characterization of natural language processing technologies in public health.

- Natural language processing (NLP) applications: The NLP application each system performs, such as conversational assistant, language translation, or automated diagnosis
- Modality: Type of data processed by the NLP application (eg, text, audio, and image)
- Supported languages: The set of languages supported by the NLP technology; languages are documented using ISO (International Organization for Standardization) 639-2 codes
- Target countries: Countries or regions where the introduced NLP technology is applied or intended to be used; countries are documented using the Alpha-3 code from the ISO 3166 standard
- Evaluation method (Adapted from Laranjo et al [111])
  - Technical performance: Intrinsic evaluation measures such as accuracy, precision, recall, and F1 score
  - User experience: Results on usability testing, user satisfaction surveys, and qualitative feedback from health care providers
  - Health-related measure: Extrinsic evaluation measures such as patient engagement rates, reduction in diagnostic errors, or improvements in treatment outcomes
- Domain coverage
  - General domain: Data concerning general language processing outside specialized contexts
  - Research domain: Research articles and professional materials for expert audiences
  - Clinical domain: Clinical notes, patient interactions, and other health care–specific communications
- Target users
  - Health care providers: Direct care providers including physicians, nurses, practitioners, community health workers, and other health care professionals
  - Public health officials and policy makers: Individuals involved in public health policy, administration, and epidemiology
  - Researchers and data scientists: Academics and professionals focused on public health research and data analysis
  - Specific equity-seeking groups: Populations grouped by protected demographic characteristics, such as people with disabilities, children, LGBTQ+ (lesbian, gay, bisexual, trans, queer) individuals, and older adults, who advocate for health equity within and beyond their group
  - General public: The broader community, especially those at higher risk or in need of specific health interventions
  - Others: Any target users that do not fit into the above categories
- Deployment stage
  - Conceptualization: This initial stage is when the need for an NLP application is identified, and its feasibility is considered
  - Design and prototyping: Development of initial prototypes; these prototypes are usually evaluated based on their technical performance
  - Validation: Rigorous testing of the system with public health outcomes to validate its effectiveness and efficiency in real-world settings
  - Deployment and operational: Deployment of the NLP technology in actual public health settings, where it is actively used
  - Not applicable: The study does not introduce or use any new NLP technologies
- Level of accessibility
  - Open-source: Publicly accessible datasets and tools that are open-source for future research and analysis
  - Publicly available: Datasets and NLP applications that are accessible to the general public via web or mobile but not necessarily open-source
  - Limited access: Datasets and NLP applications available only to certain users or under specific conditions
  - Closed access: Datasets and applications that are not openly accessible outside the group of authors but may be available upon request or through collaboration
- Available platform
  - Mobile apps: Technologies accessible via mobile apps
  - Web-based applications: Technologies accessible via web applications or web-based platforms
  - Web service: Technologies accessible via web-based application programming interfaces without user interfaces
  - Dataset: Specific datasets published in the study
  - NLP tool and library: Specific NLP tools and libraries; these tools usually require installations on each deployed computer, which require expertise in computer science

XSL•FO
RenderX

Results

Overview

The initial database search on May 13, 2024, retrieved 1791 citations, and the final updated search on October 3, 2024, retrieved an additional 404 citations (Figure 1). The updated search retrieved 289 additional articles from PubMed, 70 from ACL Anthology, 2 from Scopus, 36 from IEEE Xplore, and 7 from ACM Digital Library, resulting in 6 additional papers being identified for full-text eligibility assessment. After removing 9 duplicate citations, 2186 unique records were screened. During the title and abstract screening, 1825 articles were excluded. Full-text reviews were conducted for the remaining 361 articles, which included 6 articles identified through the final updated search. Following the full-text screening, 311 articles were excluded, resulting in the inclusion of 2.29% (50/2186) studies. An additional 4 studies were identified through reference chaining of the included studies. Before the formal screening process, 3 rounds of pilot screenings, covering 10% (179/1791 for title and abstract screening and 36/361 for full-text screening) of the citations, were conducted to ensure consistency and reliability. The final round achieved interreviewer agreement scores of 0.97 for accuracy and 0.89 for Cohen κ in title and abstract screening, and perfect agreement (ie, 1.0 for both measures) in full-text screening.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the search and study selection process for 54 included studies. NLP: natural language processing.

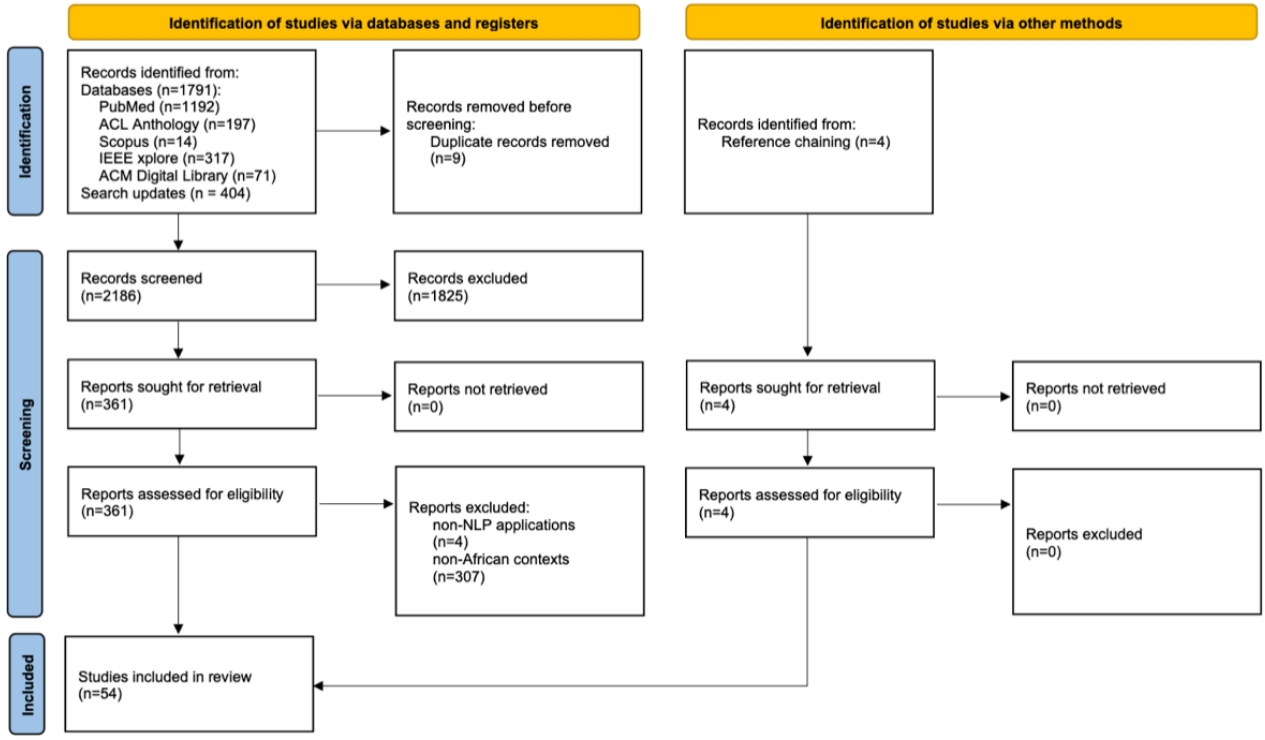

In this section, we provide an overview of the academic literature on NLP technologies for public health in Africa and present our findings in response to the 5 aforementioned research questions. In addition, we separately summarize relevant gray literature, which provides complementary perspectives to the academic literature. By combining these 2 sources of evidence, we aim to provide a comprehensive and up-to-date analysis of the landscape, while adhering to the rigorous methodological standards of this scoping review.

Description of Academic Literature

Over the past decade, there has been a rapid increase in the number of publications on NLP for public health in Africa, with a notable spike in 2022, where 6 (43%) out of the 14 papers published that year were in response to the COVID-19 pandemic (Textbox 3). The number of academic papers from authors affiliated with African and non-African institutions has been similar (Textbox 3). Of the 54 included citations, 38 (70%) papers were contributed by authors affiliated with African institutions [18,37-73], while 35 (65%) papers [38,40-44,46,48,50-52,59,61,62,64,66,69,71,73-89] were authored by researchers affiliated with institutions outside Africa. For readability, we do not provide a full list of in-text citations for all our categorizations throughout this review, instead highlighting specific papers where necessary. A complete table with all the categorizations and their corresponding references is available in Multimedia Appendix 1. Notably, 19 (35%) papers [38,40-44,46,48,50-52,59,61,62,64,66,69,71,73] stem from collaborations between African and non-African institutions, highlighting the prevalence and importance of cross-border and cross-continental collaborations at the intersection of NLP and public health research. Among these 19 papers, 16 (84%) involved coauthorship between researchers from North America and Africa, 6 (32%) involved coauthorship between Asia and Africa, 4 (21%) involved coauthorship between Europe and

XSL•FO
RenderX

Africa, and 1 (5%) involved coauthorship between Oceania and Africa. Researchers affiliated with institutions in the United States and South Africa (18/54, 33% papers each) emerged as the leading contributors to NLP research for public health in Africa.

Textbox 3. Number of papers by publication years, author institutional affiliations by country, languages supported, and African countries and regions supported in the 54 included studies.

Publication years

- 2015: n=1; 2016: n=2; 2017: n=2; 2018; n=1; 2019: n=5; 2020: n=9; 2021: n=8; 2022: n=14; 2023: n=9; 2024: n=3
- Note that data gathering ended on October 3, 2024, providing only 9 months of data for 2024, which means the 2024 data are not directly comparable to the other years.

Author institutional affiliations by country

- South Africa: n=18; the United States: n=18; Kenya: n=8; Canada: n=6; India: n=5; Germany: n=3; Iran: n=3; Rwanda: n=3; Saudi Arabia: n=3; the United Kingdom: n=3; others: n=23
- Countries with fewer than 2 papers are grouped under others, including Belgium, Brazil, Cameroon, Egypt, Eswatini, Ethiopia, France, Hungary, Indonesia, Italy, Lebanon, Lesotho, Morocco, Netherlands, New Zealand, Nigeria, Qatar, Senegal, Sierra Leone, Spain, Switzerland, Tanzania, and Uganda.
- Note that a single paper can have authors from multiple countries, so the total number of country affiliations exceeds the total number of papers reviewed.

Language supported

- English: n=40; Arabic: n=8; Kiswahili: n=7; French: n=4; Zulu: n=4; Amharic: n=3; Hausa: n=3; Hindi: n=3; Northern Sotho: n=3; Xhosa: n=3; others: n=64
- Languages supported by <3 technologies are grouped under others, which includes Afrikaans (n=2), Bengali (n=2), Chinese (n=2), Gujarati (n=2), Igbo (n=2), Indonesian (n=2), Japanese (n=2), Korean (n=2), Marathi (n=2), North Ndebele (n=2), Portuguese (n=2), Sinhala (n=2), Sotho (n=2), Spanish (n=2), Urdu (n=2), Assamese (n=1), Central Atlas Tamazight (n=1), Czech (n=1), Dutch (n=1), German (n=1), Iloko (n=1), Italian (n=1), Kannada (n=1), Kikuyu (n=1), Kinyarwanda (n=1), Luo (n=1), Malay (n=1), Malayalam (n=1), Nepali (n=1), Nyankole (n=1), Panjabi (n=1), Persian (n=1), Polish (n=1), Pushto (n=1), Russian (n=1), Shona (n=1), Somali (n=1), Swati (n=1), Tagalog (n=1), Tamil (n=1), Telugu (n=1), Thai (n=1), Tigrinya (n=1), Tsonga (n=1), Tswana (n=1), Turkish (n=1), Uighur (n=1), Venda (n=1), and Yoruba (n=1).
- Languages are identified using the ISO 639-2 code, and different dialectal variants of a language (eg, Arabic) are not distinguished.
- A single technology may support multiple languages.

African countries and regions supported

- South Africa: n=25; Kenya: n=14; Nigeria: n=9; Rwanda: n=7; Egypt: n=5; Ethiopia: n=5; Uganda: n=5; Zimbabwe: n=4; Cameroon: n=3; Eritrea: n=3; Morocco: n=3; Somalia: n=3; Tunisia: n=3; others: n=55
- Countries and regions supported by <3 technologies are grouped under others, which includes Algeria (n=2), Botswana (n=2), Democratic Republic of the Congo (n=2), Eswatini (n=2), Lesotho (n=2), Malawi (n=2), Mozambique (n=2), Namibia (n=2), Niger (n=2), Senegal (n=2), South Sudan (n=2), Sudan (n=2), Tanzania (n=2), Angola (n=1), Benin (n=1), Burkina Faso (n=1), Burundi (n=1), Cabo Verde (n=1), Central African Republic (n=1), Chad (n=1), Comoros (n=1), Congo (n=1), Côte d'Ivoire (n=1), Djibouti (n=1), Equatorial Guinea (n=1), Gabon (n=1), Gambia (n=1), Ghana (n=1), Guinea (n=1), Guinea-Bissau (n=1), Liberia (n=1), Libya (n=1), Madagascar (n=1), Mali (n=1), Mauritania (n=1), Mauritius (n=1), Sao Tome and Principe (n=1), Seychelles (n=1), Sierra Leone (n=1), Togo (n=1), Western Sahara (n=1), and Zambia (n=1).
- Countries and regions are identified using the ISO 3166 code.

Approximately 48% (26/54) of the papers included in this review did not disclose their source of funding. Of the 28 (52%) papers that did report funding, some of them received funding from more than one source, with the vast majority of papers (26/28, 93%) supported by public entities. This included government grants, NGOs, and research councils. Only 2 (4%) papers were funded by industry actors [38,88]. Geographically, with international funding sources determined by their headquarters' location, most of the papers that disclosed funding were financially supported by institutions in North America (14/54, 26%) and Europe (9/54, 17%). Reflecting the global nature of these research contributions, funding was also sourced from other continents, including Africa and Asia, demonstrating the wide range of financial support for these studies. Notably, no funding was recorded from Oceania and South America.

The data used to develop NLP technologies for public health in Africa is generally up-to-date. Among the 36 (67%) papers that reported the year of data collection, most studies (30/54, 56%) used data collected either in the same year or within 1 year before publication. Specifically, 13 (24%) papers used data collected in the same year, while 17 (31%) used data with a 1-year delay. In this review, for papers introducing a dataset, unless otherwise specified, assume the year of data collection is the same as the year of publication. For papers that introduce NLP applications or perform an analysis using a data set from other sources, the year of data collection is the year when the data was originally published.

XSL•FO
RenderX

Most of the data (36/54, 67% papers) fall within the general domain, primarily produced and consumed by the general public, such as social media data. In addition, 15 (28%) studies covered clinical domains, including clinical notes, patient interactions, and other health care–specific communications, while 8 (15%) papers focused on the research domain, covering research articles and professional materials aimed at expert audiences. It should be noted that one study can cover multiple domains simultaneously, as seen in 5 (9%) studies. Regarding data modality, most of the data (53/54, 98% studies) used to develop these technologies were text-based, with minimal use of other modalities. Only 4 (7%) papers used audio [47,63] or image [44,48] data, highlighting a limited exploration of non–text-based data in NLP applications for public health in Africa.

Needs and Availability

Most of the reviewed papers focused on conversational assistants (17/54, 31%) and sentiment analysis (15/54, 28%). Additional applications included machine translation (3/54, 6%), thematic analysis (3/54, 6%), information extraction (3/54, 6%), and outbreak detection (2/54, 4%). Fewer papers addressed tasks, such as infection detection, misinformation detection, disease prediction, optical character recognition, question-answering, hate speech detection, medical report generation, and speech recognition, with each of these applications covered by only 1 study. A smaller subset of studies focused on fundamental NLP challenges in the context of public health, such as syntax parsing (1/54, 2%), word embedding (1/54, 2%), and lexical processing (1/54, 2%), rather than user-facing applications.

Most available NLP technologies for public health in Africa were designed to serve expert users, such as researchers (45/54, 83%), policy makers (38/54, 70%), and health care providers (30/54, 56%). Fewer than half of the systems were public-facing (25/54, 46%) and targeted toward equity-seeking groups (8/54, 15%). This focus on expert-driven systems suggests an opportunity to develop more public-facing NLP technologies that engage and empower communities to proactively manage their health. After mapping the currently available NLP technologies into the WHO's EPHF framework, Figure 2 shows that 9 (75%) out of 12 EPHFs were addressed by existing NLP technologies in Africa, with EPHF 3 (ie, public health stewardship), EPHF 8 (ie, community engagement and social participation), and EPHF 12 (ie, access to and utilization of health products, supplies, equipment, and technologies) remaining unaddressed. Most studies predominantly focused on 4 EPHFs: EPHF 7 (ie, health promotion; 31/54, 57%), EPHF 11 (ie, public health research, evaluation, and knowledge; 25/54, 46%), EPHF 10 (ie, health service quality and equity; 24/54, 44%), and EPHF 1 (ie, public health surveillance and monitoring; 23/54, 43%). Work on other EPHFs remains relatively sparse, with only a handful of papers addressing them. When each paper was assigned one primary EPHF, only 6 EPHFs were the main focus of these studies, leaving 6 EPHFs unaddressed, including EPHF 2 (ie, public health emergency management), EPHF 3, EPHF 4 (ie, multisectoral planning, financing, and management for public health), EPHF 5 (ie, health protection), EPHF 8, and EPHF 12.

When mapping the included NLP technologies to the United Nations' SDGs, all the reviewed NLP technologies contributed to SDG 3 (good health and well-being). The interconnected nature of the SDGs means that contributing to one SDG often supports progress in others. For example, 15/54 (28%) studies contributed to SDG 10 (reduced inequality), 10/54 (19%) studies to SDG 9 (industry, innovation, and infrastructure), 6/54 (11%) studies to SDG 8 (decent work and economic growth), 5/54 (9%) studies to SDG 4 (quality education), 4/54 (7%) studies to SDG 5 (gender equality), 1 (2%) study to SDG 15 (life on land), and 1 (2%) study to SDG 16 (peace, justice, and strong institutions). Thus, even with a primary focus on health, these projects may have a far-reaching impact on other SDGs [112]. When zooming in on the targets of SDG 3, currently available NLP technologies in Africa only cover 6 of the 13 specific targets (Figure 3). Of the 4 means of implementation listed for SDG 3, available technologies engage 3 (ie, tobacco control, access to vaccines and medicines, and health financing).

XSL•FO
RenderX

Figure 2. Distribution of all the essential public health functions (EPHFs) to which the 54 included papers contribute. The blue line represents EPHF coverage, where each paper may contribute to multiple EPHFs. The orange line indicates the primary EPHF, where each paper is assigned a single, most relevant EPHF focus.

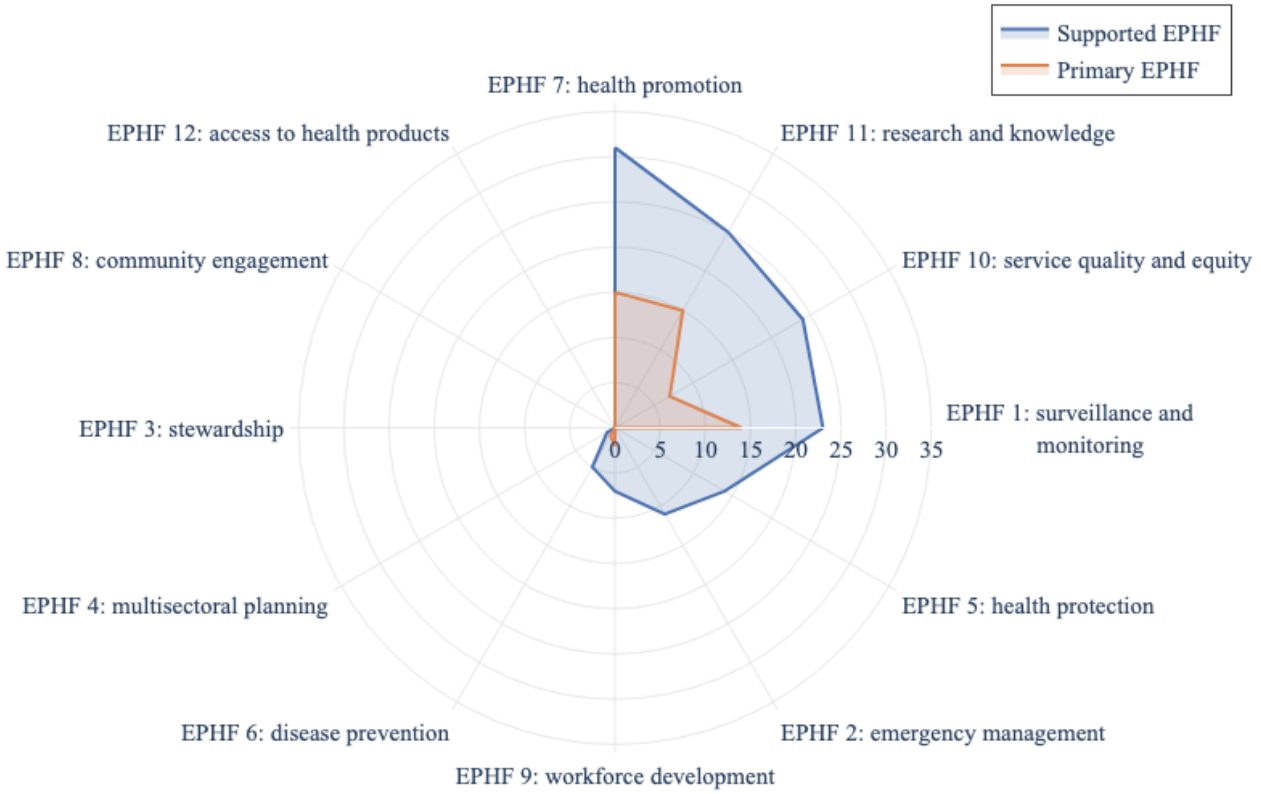

Figure 3. Distribution of specific targets (3.1-3.9) and means of implementation (3.a-3.d) under Sustainable Development Goal 3 (good health and well-being) to which the 54 included papers contribute. NCD: noncommunicable disease; WHO: World Health Organization.

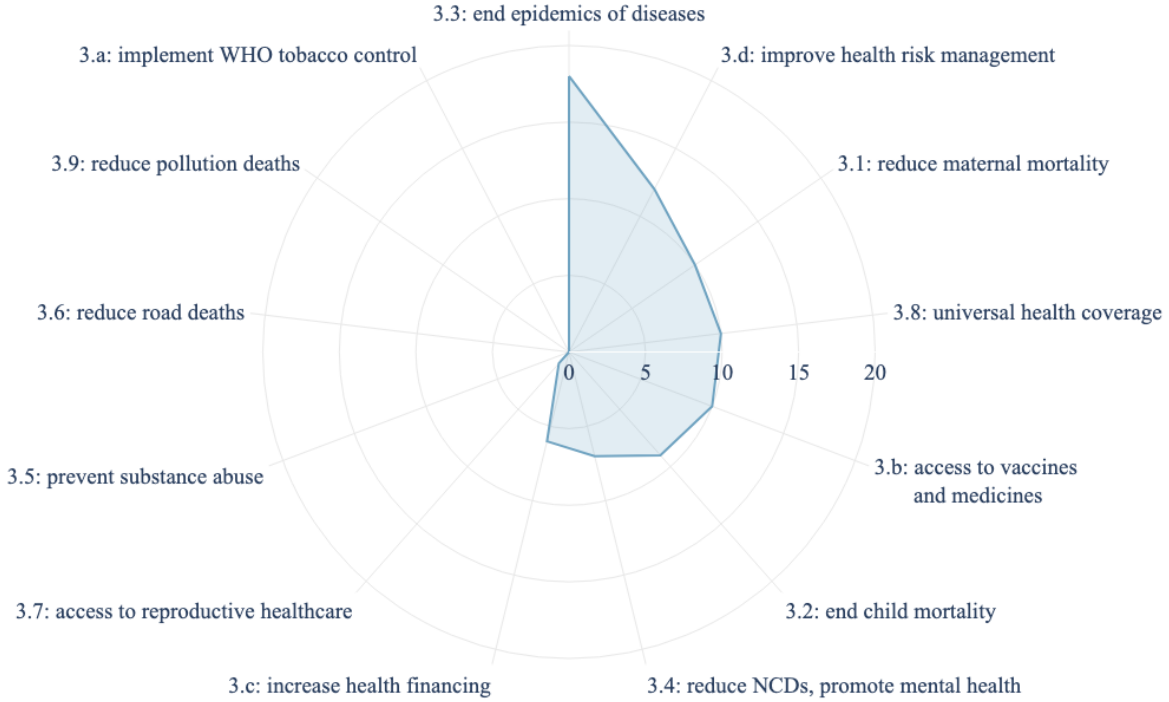

Prevalence and Distribution

The availability of NLP technologies for public health in Africa is strongly influenced by the languages these technologies support. Most NLP technologies predominantly serve widely spoken high-resource languages (Textbox 3), such as English (40/54, 74%), Arabic (8/54, 15%), and French (4/54, 7%), reflecting their status as official languages in academic, governmental, and professional sectors across the continent. In contrast, support for indigenous African languages is

XSL•FO
RenderX

significantly limited. While some widely spoken African languages, such as Kiswahili (7/54, 13%) and Zulu (4/54, 7%), are represented, many other indigenous languages remain underrepresented or entirely absent from these technologies. Overall, 59 languages were supported by the 54 studies included in this review, a number that falls far short of covering Africa's linguistic diversity.

The availability of NLP technologies for public health in Africa varies significantly across countries and regions. As shown in Textbox 3, South Africa is the primary target country for these technologies, with 25 (46%) out of 54 studies targeting this country, followed by Kenya (14/54, 26%) and Nigeria (9/54, 17%). In contrast, 29 African countries and regions, including Angola, Benin, and Burkina Faso, are supported by only 1 technology each, highlighting uneven distributions of NLP technologies across the continent.

Our review reveals a geographic concentration of available NLP technologies in certain countries, especially South Africa, Kenya, and Nigeria, suggesting a need for future efforts to expand NLP technology development to underserved regions (Figure 4). The results further highlight a major gap in linguistic inclusivity within the existing NLP technologies across the continent, where languages spoken in these better-supported regions receive more attention compared to those in other areas (Figure 5).

Figure 4. A map illustrating the distribution of target countries across Africa based on the 54 studies included in this review. Darker shades indicate a higher number of studies per country, as outlined in Textbox 3. Note that small African island states are missing from this map.

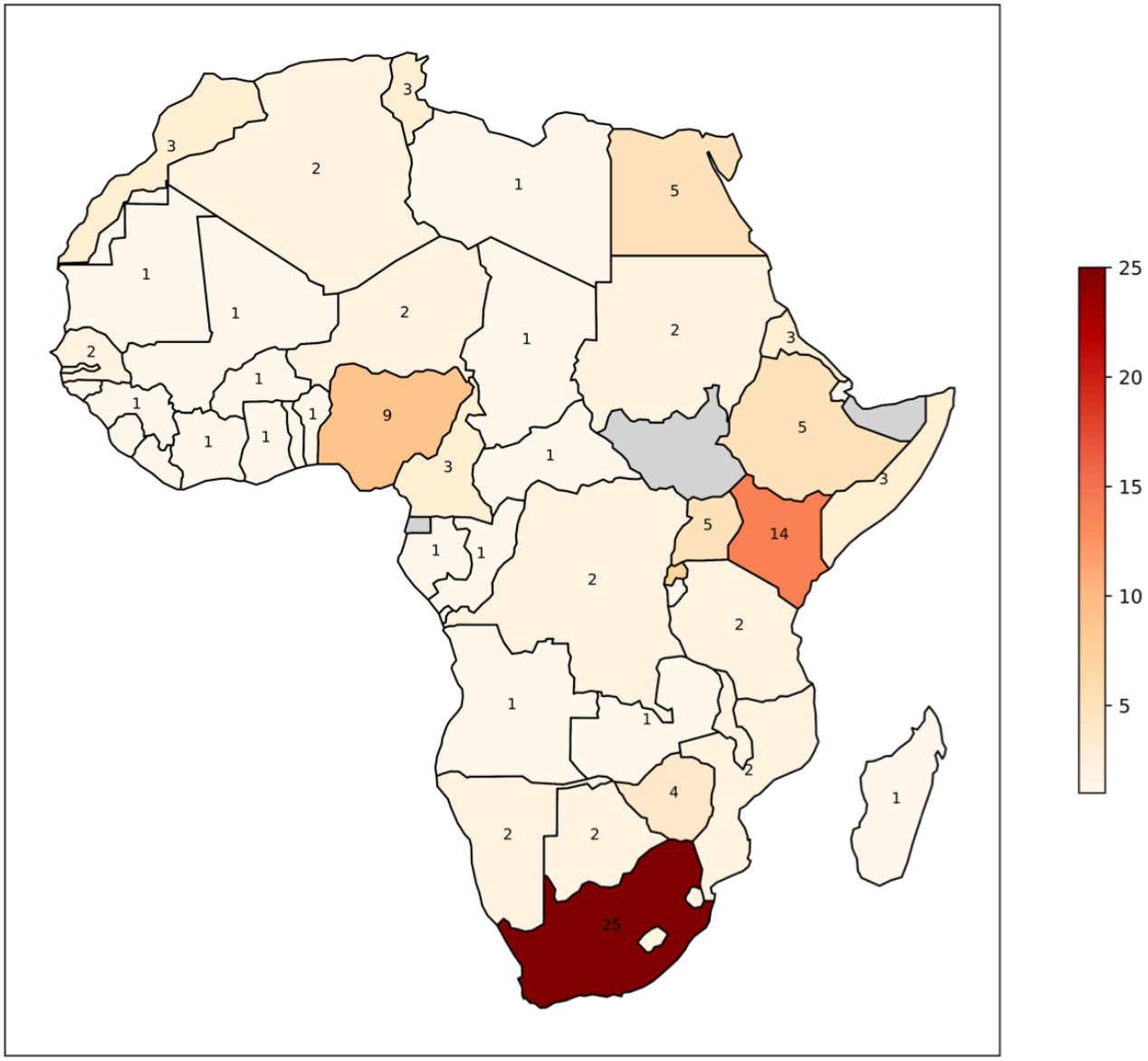

XSL•FO
RenderX

Figure 5. A map illustrating the distribution of supported languages across Africa based on the 54 studies included in this review. Each name is positioned near the geographic center of its primary speaking population, with a larger font size indicating a greater number of studies available. The data on supported languages is based on Textbox 3. Note that our study used English language search terms and small African island states are missing from this map.



Deployment and Integration

Among the 54 included studies, 4 (7%) were reviews of existing NLP technologies related to health chatbots [18], HIV prevention in Africa [84], HIV prevention in Malawi specifically [85], and chatbots for HIV prevention [59]. These reviews do not introduce new NLP technologies but rather summarize findings from other research that has introduced new technologies and reported primary results. As such, the concept of deployment is less relevant to these review studies. Therefore, this subsection focuses on the 50 studies that directly introduce new NLP applications.

Most NLP applications for public health in Africa are still in the early stages of development, with only 1 (2%) out of the 50 studies fully deployed and operational. This deployed system is a Facebook messenger chatbot designed to address vaccine hesitancy in Kenya and Nigeria, collecting real-time data on vaccine hesitancy trends from user interactions [89]. Most studies (44/50, 88%) are in the *design and prototyping* phase, where they are evaluated only based on their technical performance in controlled, lab-based environments. Meanwhile, 5 (10%) studies have reached the *validation* stage, where their effectiveness has been tested in real-world settings through methods, such as expert reviews [37,74] and user testing [39,44,45]. Specifically, 2 (4%) studies [37,74] involved health

XSL•FO
RenderX

care professionals who reviewed system performance in practical scenarios. The remaining 3 (6%) studies [39,44,45] conducted evaluations with small samples of target users to test the developed NLP technologies. However, these systems were accessible only to a limited number of users and have not yet achieved full deployment.

Regarding accessibility (as defined in Textbox 2), >half (29/50, 58%) of the NLP applications are publicly accessible, allowing general use without significant restrictions. Of these, 12 (24%) are *open-source*, enabling researchers and developers to build new NLP applications based on their published systems. By contrast, a significant number of systems are categorized as having *limited access* (18/50, 36%) or are *closed access* (3), likely due to the sensitive nature of health-related data, raising concerns around privacy and data security.

In terms of platform support, most NLP technologies for public health in Africa are offered as tools and libraries (29/50, 58%), datasets (5/50, 10%), or web services (4/50, 8%), all of which require a certain level of technical expertise in computer science to exploit effectively. A smaller proportion of technologies are provided as mobile apps (11/50, 22%) and web-based applications (9/50, 18%), offering more user-friendly interfaces that can be accessed by a broader range of users, including public health practitioners and the general public. This distribution suggests an opportunity to develop NLP technologies with more accessible interfaces to promote wider adoption and usability of these technologies for public health across Africa.

Of the 50 reviewed NLP technologies, 40 (80%) indicated an intent to integrate their solutions into existing public health systems, and 41 (82%) were designed to be interoperable with various health infrastructures. This reflects a clear recognition among researchers and developers of the importance of ensuring that these technologies function seamlessly within current health frameworks. However, despite this intent, only 1 (2%) study has reached the stage of deployment, highlighting the need to move these technologies from development into operational use.

Scope and Public Health Impact

Out of the 54 studies reviewed, 30 (56%) aimed to develop NLP applications or inform public policies to improve public health outcomes in Africa. Of these, 22 (41%) specifically focused on addressing public health challenges within African countries, directly targeting the health issues faced by local communities. The remaining 8 (15%) studies adopted a broader global health perspective. While they do not exclusively tailor their approaches to Africa, these studies aim to promote public health on a global scale, with intended outcomes that also benefit Africa. The other 24 (44%) studies were divided into 18 (33%) studies that advance NLP technologies for public health using African data (eg, social media, health records, or public health data) without directly targeting specific health challenges, and 6 (11%) studies that contribute to global health discussions, with Africa serving as a case study or example.

In terms of evaluation, nearly all studies (51/54, 94%) reported technical performance using a variety of automatic evaluation metrics. Classification metrics were the most commonly used, such as accuracy (used in 22/54, 41% studies), precision (13/54, 24%), F_1-score (15/54, 28%), and recall (11/54, 20%), making these the 4 most frequently applied automatic metrics. However, because of the wide variation in evaluation approaches, direct comparisons between the studies were impractical and not attempted. In contrast, only 11 (20%) studies reported evaluation results based on user experiences, such as usability testing, user satisfaction surveys, or qualitative feedback from health care providers. Furthermore, only 8 (15%) studies attempted to evaluate these technologies using health-related measures. Among these, 2 (4%) studies confirmed a positive impact on public health outcomes, with NLP-based interventions shown to improve participants' mood [71] and increase vaccine intentions and willingness [89].

Outlook and Ethical Consideration

Among the 54 included papers, 20 (37%) provided recommendations for the future development of NLP technologies for public health in Africa. A thematic analysis of these recommendations identified 6 key areas for future research: addressing specific public health challenges with NLP (11/54, 20%), expanding data coverage for underrepresented languages (8/54, 15% studies), contextualizing solutions to local health needs (6/54, 11% studies), enhancing trust and ethical standards (5/54, 9% studies), integrating NLP technologies with existing health systems (5/54, 9% studies), and incorporating participatory design with domain expert input (3/54, 6% studies).

In terms of ethical considerations, 46 (85%) out of 54 studies attempted to engage stakeholders during the study design and implementation, while 38 (70%) studies explicitly addressed data privacy compliance. Approximately half of the studies (26/54, 48%) involved the local community in their research. However, only 16 (30%) studies reported receiving explicit ethics approval from an independent review board, and 10 (19%) studies mentioned obtaining informed consent from human participants. It is important to note that not all types of studies involved human participants or required explicit ethics approval in advance. In addition to these considerations, 45 (83%) papers highlighted other ethical concerns, including bias and fairness (11/54, 20%), cultural relevance and appropriateness (11/54, 20%), avoiding miscommunication by NLP technologies (9/54, 17%), preventing misuse of NLP technologies (9/54, 17%), data sharing and accessibility (7/54, 13%), adherence to regulatory standards (2/54, 4%), data representativeness (1/54, 2%), and fair compensation for participants (1/54, 2%).

Description of Gray Literature

Our gray literature review covered two types of sources: (1) academic literature, including unpublished preprints and peer-reviewed articles not indexed in the 5 structured databases, and (2) nonacademic sources, such as online articles, blog posts, products from startups and established companies, initiatives from NGOs, and proceedings from events and conferences. Full results of the gray literature search are detailed in Multimedia Appendix 1, with key findings highlighted below.

Within the academic gray literature, we identified 11 relevant articles from the first 100 Google Scholar results, with 9 (9%)

XSL•FO
RenderX

peer-reviewed articles and 2 (2%) preprints. These studies generally aligned with the patterns observed in the aforementioned structured database search. Each study involved researchers affiliated with at least one African institution, with contributions from South Africa (6/11, 55% studies) [90-95], Nigeria (4/11, 36%) [95-98], Guinea (1/11, 9%) [99], and Rwanda (1/11, 9%) [100]. In addition, 4 (36%) studies involved collaborations with international researchers from institutions based in the United States (3/11, 27% studies), Canada (3/11, 27%), Germany (1/11, 9%), and Mexico (1/11, 9%). Funding was disclosed in 6 (55%) of the 11 studies, all supported by public entities.

The primary NLP applications developed in these studies were conversational assistants (4/11, 36% studies) and sentiment analysis tools (3/11, 27% studies). These studies primarily supported EPHF 7 (ie, health promotion, 8/11, 73% studies). Regarding language coverage, nearly all studies (11/11, 100%) reported support for English. A smaller number addressed African languages, including Ndebele (2/11, 18% studies), Sotho (2/11, 18%), Kiswahili (2/11, 18%), Swati (2/11, 18%), Venda (2/11, 18%), Xhosa (2/11, 18%), Zulu (2/11 18%), with one study each for Afrikaans, Hausa, Kinyarwanda, Northern Sotho, Shona, Tsonga, and Tswana. For target countries, Nigeria and South Africa were the primary focus, each covered in 5 (45%) studies. Notably, none of the studies provided performance evaluations based on health-related measures or reported reaching the stage of actual deployment.

Outside academia, commercial products and NGO-led initiatives have focused on creating practical NLP solutions aimed at real-world public health impact. On the basis of our search results, 4 NLP technologies were developed as commercial products by companies [101-104], and another 4 were created as part of initiatives led by NGOs [105-108]. These projects were often in partnership with charitable organizations like the Bill and Melinda Gates Foundation, international bodies, such as the WHO, and industry partners like Google or Meta, frequently collaborating with telecom providers to reach populations with lower literacy levels and limited access to public health resources. The primary focus of these NLP technologies was on disseminating public health information through conversational assistants, with applications supporting EPHF 7 (ie, health promotion) and SDG 3 (good health and well-being). Most tools were designed in English with the limited inclusion of widely spoken African languages like Hausa, Kiswahili, and Zulu. In contrast to academic literature, NLP technologies from these nonacademic sources typically disclosed only limited details about their design and implementation.

Furthermore, our review of events and conferences did not introduce additional evidence of NLP technologies tailored to African public health challenges. A lack of standardized protocols for reporting NLP technologies, such as established reporting standards or controlled vocabularies for indexing, may explain why no relevant NLP technologies were retrieved during our search, likely due to limited keyword overlap.

Discussion

Principal Findings

Research into NLP technologies for public health in Africa is an emerging field, with significant growth since 2019. Current studies primarily focus on 2 applications: conversational agents for public health information dissemination and sentiment analysis tools that track public health attitudes on social media. Most studies target high-resource languages like English, Arabic, and French, with limited support for widely spoken African languages, such as Kiswahili and Zulu, and no support for most of Africa's >2000 languages.

Most NLP applications remain in the prototype stage, with evaluations often limited to technical performance metrics in controlled settings. Only a handful of studies have validated their systems in real-world contexts, and just 1 has reached full deployment. Until now, most systems have been developed as technical NLP tools rather than targeted health interventions, with limited evaluation of their impact on public health outcomes through rigorous study designs and implementation research approaches.

While current research highlights the potential of NLP to address public health needs in Africa, this potential remains largely unrealized in terms of measurable public health outcomes. The following discussion explores pathways for public health and NLP researchers to contribute to the development and deployment of NLP technologies toward achieving positive health impacts in Africa. In addition, we reviewed the strengths and limitations of our review approach, providing context for readers to critically evaluate the subsequent discussion.

Bridging Technical NLP Performance With Health-Related Outcomes

The review of 54 studies highlights the growing effort to leverage NLP technology for health improvement in Africa. However, it identifies a significant gap in evaluating real-world health outcomes or the behavioral antecedents of these outcomes. Most studies (51/54, 94%) emphasized technical performance, using metrics, such as accuracy, precision, F_1-score, and recall. In comparison, only 11 (20%) studies incorporated user-centered evaluations, such as usability testing or health care provider feedback. While some studies [45,47,54] assessed user outcomes like the accuracy of health communications and improvements in health care interactions, only 2 (4%) studies [71,89] measured explicit health-related impacts. Specifically, 1 (2%) paper [71] demonstrated improvements in participants' mood through an automated intervention targeting maternal mental health in Kenya, while another paper [89] showed increased vaccine willingness via a chatbot addressing individual concerns. These examples illustrate the potential of NLP interventions to influence public health, while their rarity highlights the need for more research focused on evaluating health impacts.

An overreliance on technical NLP metrics limits our understanding of whether these technologies effectively address real-world health challenges. To ensure NLP solutions meet their intended public health goals, future research should

XSL•FO
RenderX

incorporate tools to evaluate health-related measures and behavioral outcomes of NLP solutions alongside technical performance. Tools and frameworks already exist to guide the evaluation of health interventions, such as the WHO's "Monitoring and Evaluating Digital Health Interventions" framework [113], which provides standardized guidelines for assessing the impact of digital health technologies on health outcomes and behaviors. Despite the availability of such resources, they remain underutilized in the evaluation of NLP technologies. To fully realize the potential of NLP for public health, it is essential that future studies adopt these established frameworks to rigorously measure both health outcomes and behavioral changes. Integrating these tools will strengthen evidence on the real-world effectiveness of NLP interventions and support more impactful, data-driven public health strategies.

Deployment, Integration, and Cross-Sectoral Development

For NLP technologies to be deployed in an impactful way, they must be integrated into African health systems and broader public health infrastructure, ensuring accessibility to diverse groups of users. The results of our review of the academic literature have shown the nascent nature of NLP deployment in Africa, with only 1 technology, a Facebook messenger chatbot collecting data on vaccine hesitancy [89], having reached full deployment. Other described technologies (40/50, 80%) were designed with the potential for integration into public health systems, and most apps under development are available without significant restrictions (ie, either open-source or publicly available). However, many of these apps require substantial expertise in computer science for installation and use, limiting their accessibility. For effective integration, these technologies need to be accessible to their intended users, such as health care workers, patients, and nonspecialists. Approximately 20 apps in this review were designed to be delivered via mobile- or web-based interfaces, increasing their potential usability.

In contrast, industry-led commercial products and NGO-driven initiatives have generally progressed further, often yielding immediate, tangible impacts for African communities. These initiatives commonly partner with organizations [114] like the Bill and Melinda Gates Foundation [115], the WHO, and companies [116], such as Google or Meta, and frequently collaborate with telecom providers to enhance accessibility for populations with limited resources and lower literacy levels. Unlike academic studies, which typically prioritize proof-of-concept and feasibility testing, these projects aim for direct public health impact, real-world validation, and, at times, profitability. However, as highlighted in this review, nonacademic projects tend to focus on narrower applications, primarily conversational assistants, offer limited language support, serve smaller populations, and address a more focused range of public health challenges compared to the diverse objectives often seen in academic research.

Moving forward, bridging the gap between NLP research and accessible, real-world applications will be essential for delivering positive public health impacts. The narrower focus of nonacademic projects highlights a need for extended collaboration between academic and nonacademic researchers, combining priorities, expertise, and resources to enhance NLP's potential in addressing Africa's public health needs. Cross-sectoral partnerships offer a promising model for advancing academic NLP technologies from proof-of-concept to impactful public health solutions across the continent.

Toward Needs-Based Approaches

As we move toward the SDGs' 2030 deadline, it is sobering to note that "current progress falls far short of what is required to meet the SDGs" [117]. Within this, the world is off track to achieve SDG 3 [117]. The SDG dashboard map offers a country-by-country breakdown of each of the SDG 3 indicators [118]. Progress toward SDG 3 in all but one mainland African country (ie, Tunisia) is described as "major challenges remain" (ie, the most concerning category), while Tunisia and the island nations of Cabo Verde, Mauritius, and the Seychelles are in the less severe category of "significant challenges remain." In terms of progress, no African countries are currently considered to be "on track" or "decreasing" their progress; instead, they are all described as having major challenges in their progress toward the SDG 3 (ie, good health and well-being) targets [3].

In response to this somewhat bleak outlook, the United Nations prescribes that "changing course requires prioritizing the achievement of universal health coverage, strengthening health systems, investing in disease prevention and treatment, and addressing disparities in access to care and services, especially for vulnerable populations" [117]. Furthermore, it should be recognized that poverty and inequality constrain the possibilities for health gains [119], highlighting the need for a paradigm focused not only on treatment but on prevention, equity, and intersectional, multisectoral approaches to health promotion.

There is also a need to address technological and infrastructural limitations which still exist. Globally, a third of people remain offline—that is 30% of men and 35% of women [117]. In 2015, 15.6% of people in sub-Saharan Africa had internet access, rising to 37% by 2023 (*ibid*). Furthermore, a study of 15 countries (of which 7 were African countries) demonstrated how access to this technology often varies, with lower phone ownership in rural compared to urban areas, and varied ownership levels between poorer and wealthier income groups [120].

A review of the successes and limitations of telemedicine deployment in Africa during the COVID-19 pandemic [121] demonstrates what this means in practice. The study found the following technologies were used "videos, telephones, smart wearable digital devices, messaging mobile apps, virtual programs, online health education modules, SMSs, live audio–visual communication, and other digital platforms." Among these, phones were the most widely used. Some of the difficulties faced included an array of digital challenges ranging from low connectivity and high data costs to the inaccessibility of smartphones, nondelivery of messages, and insufficient digital skills. This was in a broader context characterized by a lack of telemedicine frameworks and policies to support a roll out; some patients and health care personnel preferred not to use these technologies, and there was an underlying shortage of health care personnel [121].

XSL•FO
RenderX

For NLP technologies to address real-world health challenges, they should be viewed not just as technical solutions but as tools shaped by and responsive to the local context. Developing effective NLP applications will require a community-centered approach [122,123], grounded in local needs, ethical principles, infrastructures, and capacities to ensure these tools are truly accessible and impactful. Engaging people in research, including coresearchers, can facilitate a closer understanding of local needs and suitable ways to address these [124].

The Need for Culturally and Linguistically Inclusive NLP Applications in Africa

Africa has exceptional linguistic diversity, with >2000 languages spoken across the continent [21,22]. This includes widely used official languages, such as Arabic, English, and French, alongside popular indigenous languages, such as Zulu, as well as a large majority of underrepresented languages spoken by smaller communities. Kiswahili spans several African countries, uniting East Africa as the shared language of politics, trade, music, literary tradition, and religion (both Islam and Christianity) [125]. Nigeria is the most linguistically diverse, with >500 indigenous languages [126]. While official languages tend to have relatively sufficient digital data to support NLP development, most indigenous African languages fall into the category of being low-resource, extremely low-resource, or even no-resource, often lacking any digital data essential for NLP technologies. The scarcity of digital language resources forms significant performance disparities in NLP systems [127]. These disparities, including higher error rates for underrepresented languages (ie, error rate disparities [128]), contribute to broader inequities, limiting access to advancements in NLP technology and impeding speakers of underrepresented languages from fully benefiting from progress in NLP technology.

To develop inclusive NLP applications that equitably serve African populations, strategically expanding digital datasets for underserved languages is essential. This is particularly the case for languages with limited online representation [129]. Concurrently, advancements in multilingual NLP and cross-lingual transfer learning provide promising opportunities [130-132]. These approaches allow neural language models, the backbones of most modern NLP applications, to leverage knowledge from high-resource languages to perform well in low-resource contexts, even with minimal in-language data. By combining efforts in data collection with advancements in NLP research, these technologies can better support Africa's linguistic diversity, contributing to public health solutions that promote, rather than hinder, health equity.

In addition to linguistic inclusivity, cultural relevance is essential for NLP technologies [29,133] to meet the diverse needs of African communities effectively. Distinct cultural practices, health beliefs, and communication styles influence how groups perceive and interact with technology.

Embedding cultural practices into the design and implementation of NLP technologies is essential for ensuring their relevance and successful integration into Africa's health systems and institutions. For example, a health education dialogue system advising users to visit a *general practitioner*, a term commonly associated with primary care doctors in the United Kingdom's National Health Service, may be irrelevant or confusing to many users in Africa. Furthermore, Loveys et al [134] highlighted how expressions of depression in text, such as the ratio of positive to negative emotions, vary across cultures. Similarly, the extensive use of traditional medicine in many African countries [135] highlights the need for NLP technologies to incorporate culturally recognized terms and references to these practices. Another good example is the recent work by Olatunji et al [136], which introduces a geo-culturally diverse dataset of clinically diverse questions and answers annotated by health experts. This dataset enables the development of question-answering systems serving African patients.

Respecting and integrating these cultural nuances into NLP design and implementation can enhance trust, ensure alignment with the expectations and needs of each community, and ultimately promote equitable public health outcomes.

Integrative Development for NLP in Public Health

Our analysis shows that NLP technologies focusing on health in Africa intersect most frequently with SDG 10 (ie, reduced inequality) and SDG 9 (ie, industry, innovation, and infrastructure). This study did not identify any intersections with other SDGs, even those that are particularly relevant to public health, such as SDG 1 (ie, end poverty), SDG 2 (ie, zero hunger), SDG 6 (ie, clean water and sanitation), SDG 7 (ie, affordable and clean energy), SDG 11 (ie, sustainable cities and communities), SDG 12 (ie, responsible consumption and production), SDG 13 (ie, climate action), and SDG 17 (ie, partnership for the goals) are absent. This highlights an opportunity for more cross-cutting intersections between NLP applications for health and broader sustainable economic development efforts, as well as the importance of a more integrated approach to achieving the SDGs [112,137].

The current coverage of the SDG 3 targets and means of implementation by existing NLP technologies (Figure 3) highlights the need for more investment in the WHO framework convention on tobacco control (SDG 3.a), substance abuse (SDG 3.5), road traffic (SDG 3.6), and environmental health (SDG 3.9) for sustained and long-term health impacts. While the goals of ending epidemics (SDG 3.3), reducing maternal mortality (SDG 3.1), and achieving universal health coverage (SDG 3.8) are strongly represented in the literature, there remains significant potential to invest more in cross-cutting activities that have long-term impacts on health systems.

The limited attention given to 6 key EPHFs (ie, highlighted by the orange line in Figure 2) also highlights a significant gap in the research landscape, with critical public health functions, such as emergency management, stewardship, and multisectoral planning being overlooked. These underaddressed EPHFs and SDG 3 targets play a fundamental role in building resilient health systems, especially in low-resource settings. Therefore, there is a need for more balanced research efforts to ensure all aspects of public health are adequately supported.

Developing NLP technologies to address these underresearched EPHFs and SDG 3 targets requires a deep understanding of local contexts and the integration of NLP technologies within

XSL•FO
RenderX

them. For instance, addressing EPHF 12 (ie, access to health products) requires integrating NLP technologies into existing logistics and supply chain systems, which often vary significantly between countries. Such highly specialized and localized contexts introduce additional challenges and requirements for system development.

Future studies should aim to bridge these gaps by aligning NLP development with more integrative, cross-disciplinary collaborative approaches that promote system-wide impact. By embedding NLP technologies within broader health system goals and leveraging collaborative input across fields, these solutions can become more effective, resilient, and responsive to public health challenges. Such an approach will help guide interventions and inform policy development, ensuring that public health improvements are enduring and responsive to Africa's evolving health needs.

Strengths and Limitations of the Review

This review has several strengths. We developed and followed a protocol guided by the PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analyses—Protocols) guidelines [138], ensuring a transparent and systematic approach throughout the study. Our literature search was comprehensive and multidisciplinary, integrating both academic and gray literature sources to capture a broad range of perspectives and to provide a current reflection of the scope of research. The inclusion of gray literature, such as reports from industry and NGOs, enabled a wider view of NLP applications in public health that extends beyond the academic literature. Screening and data extraction were conducted following predefined guidelines created by a team of domain experts, enhancing both the consistency and relevance of the data collected. The extracted data is publicly available in a machine-readable format to facilitate future research (Multimedia Appendix 3).

However, there are some limitations to consider. Our review relied primarily on English search terms, which may have inadvertently excluded studies published in other languages without English annotations. Although we included 5 representative databases spanning public health, NLP, computing, and engineering, some relevant studies may have been missed because of the highly interdisciplinary nature of the field. In addition, NLP research is frequently disseminated as preprints, blog posts (eg, OpenAI's introduction of ChatGPT), and in peer-reviewed machine learning conference proceedings (such as International Conference on Learning Representations and Conference on Neural Information Processing Systems) not indexed by our selected databases. Consequently, certain emerging studies may not have been captured, especially given the rapid development of NLP applications in public health. However, the results of our gray literature search indicate that most academic papers, including preprints and conference proceedings, have been covered in this review.

Although this review included a broad range of gray literature, the search for nonacademic sources was far from exhaustive, and some relevant NLP technologies may have been overlooked. The primary focus of this review was on peer-reviewed academic literature, with gray literature providing supplementary insights. This work could be further expanded through a series of expert interviews, such as with representatives from the WHO or the SDG by 2030 committees, to interrogate the factors shaping the development, or lack thereof, of NLP applications for public health in Africa [139,140]. In addition, the heterogeneity of NLP methodologies, applications, and evaluation metrics hindered our ability to conduct a formal meta-analysis, resulting in a reliance on narrative synthesis. Despite efforts to maintain consistency during data extraction, inconsistencies in reporting approaches for included studies presented challenges, particularly in evaluating public health focus and outcomes, which impacts comparability across different technologies.

These challenges reveal a broader misalignment between the rigorous methodological standards of systematic reviews and the agile, fast-evolving, and highly interdisciplinary nature of NLP research. Future research should explore methodological adaptations that better align with the field of NLP, especially for applications in health contexts.

Conclusions

The application of NLP technologies to public health in Africa is a promising and rapidly evolving field, with the potential to enhance health care accessibility, equity, and efficiency across the continent. However, significant gaps persist in real-world deployment, language inclusivity, and the rigorous evaluation of health outcomes. The identification of, and responses to, such gaps would be greatly enabled by the establishment of reporting standards for NLP technologies.

Future research should adopt a needs-based and cross-sectoral approach, engaging expertise from diverse fields, including the expertise of local communities on their own needs and possible solutions, and using existing frameworks for evaluating their public health impacts and outcomes. This approach will help build a deeper understanding of needs and support the tailored design of NLP technologies to effectively address public health challenges where this technology can be useful and is wanted. Furthermore, qualitative research, such as expert interviews, can contribute to better understand the dynamics and demand for progress in this area. By bridging existing gaps in meaningful local engagement, NLP research can better support resilient, culturally relevant, and equitable public health systems in Africa.

Acknowledgments

This work has been supported by the UK Research and Innovation Frontier Research Grant EP/Y031350/1 under the UK government's funding guarantee for European Research Council Advanced Grants. SH has been supported by the Cambridge International Scholarship.

XSL•FO
RenderX

Authors' Contributions

EM and AK conceived the idea for this scoping review and provided project supervision. SH developed the research protocol and conducted the academic literature search. AO, CC, and SH participated in the screening of papers, with AO as the primary reviewer, CC as the secondary reviewer, and SH as the coordinator. SH designed the data extraction template, while AO carried out the data extraction. Data analysis and visualization were conducted by SH, AO, and GO. SH led the drafting of the manuscript, with substantial contributions from SH, AO, EM, CC, GO, AB, and AK in writing and refining the paper. All authors engaged in discussions throughout the project. SH and AO contributed equally as shared first authors. ChatGPT was used exclusively for editorial assistance, such as improving language clarity and identifying grammatical errors, alongside editing by the authors. The authors have reviewed all content carefully and assume full responsibility for the manuscript's final form.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition of natural language processing, search strategy, screening guidelines, data extraction template, results, and supplementary materials.
[PDF File (Adobe PDF File), 1524 KB - jmir_v27i1e68720_app1.pdf]

Multimedia Appendix 2
The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist for the scoping review.
[DOCX File , 109 KB - jmir_v27i1e68720_app2.docx]

Multimedia Appendix 3
Extracted data from included studies in JSON format.
[ZIP File (Zip Archive), 23 KB - jmir_v27i1e68720_app3.zip]

References

1. Goal 3: ensure healthy lives and promote well-being for all at all ages. United Nations. URL: https://sdgs.un.org/goals/goal3 [accessed 2024-11-10]
2. The Lancet. Artificial intelligence in global health: a brave new world. Lancet 2019 Apr 13;393(10180):1478. [doi: 10.1016/S0140-6736(19)30814-1] [Medline: 30983572]
3. Sachs JD, Lafortune G, Fuller G. The SDGs and the UN summit of the future: sustainable development report 2024. Dublin University Press. URL: https://dashboards.sdgindex.org/chapters [accessed 2024-11-10]
4. Feldscher K. Africa CDC head is driving a new public health agenda on the continent. Harvard T. H. Chan School of Public Health. URL: https://www.hsph.harvard.edu/news/features/africa-cdc-head-is-driving-a-new-public-health-agenda-on-the-continent/ [accessed 2024-11-10]
5. Health in Africa over the next 50 years. African Development Bank. 2013. URL: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/Economic_Brief_-_Health_in_Africa_Over_the_Next_50_Years.pdf [accessed 2024-11-10]
6. GBD 2017 SDG Collaborators. Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related Sustainable Development Goals for 195 countries and territories: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):2091-2138. [doi: 10.1016/S0140-6736(18)32281-5] [Medline: 30496107]
7. Bartolomeos KK. The case for investing in public health surveillance in low- and middle-income countries. Afr J Emerg Med 2018 Dec;8(4):127-128. [doi: 10.1016/j.afjem.2018.07.005] [Medline: 30534514]
8. Dorling D, Newman M, Barford A. The Atlas of the Real World: Mapping The Way We Live. London, UK: Thames & Hudson; 2008.
9. Gebremeskel AT, Otu A, Abimbola S, Yaya S. Building resilient health systems in Africa beyond the COVID-19 pandemic response. BMJ Glob Health 2021 Jun 28;6(6):e006108 [FREE Full text] [doi: 10.1136/bmjgh-2021-006108] [Medline: 34183330]
10. Talisuna A, Iwu C, Okeibunor J, Stephen M, Musa EO, Herring BL, et al. Assessment of COVID-19 pandemic responses in African countries: thematic synthesis of WHO intra-action review reports. BMJ Open 2022 May 02;12(5):e056896 [FREE Full text] [doi: 10.1136/bmjopen-2021-056896] [Medline: 35501083]
11. Application of the essential public health functions: an integrated and comprehensive approach to public health. World Health Organization. URL: https://www.who.int/publications/i/item/9789240088306 [accessed 2024-11-10]
12. World Bank country classifications by income level for 2024-2025. World Bank. URL: https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025 [accessed 2024-11-10]

XSL•FO
RenderX

13. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018;3(4):e000798 [FREE Full text] [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
14. Jurafsky D, Martin JH. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Upper Saddle River, NJ: Prentice Hall; 2008.
15. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. Trends Anaesth Crit Care 2021 Jun;38:4-9. [doi: 10.1016/j.tacc.2021.02.007]
16. Artificial intelligence in global health: defining a collective path forward. United States Agency for International Development. 2019. URL: https://2017-2020.usaid.gov/cii/ai-in-global-health [accessed 2024-11-10]
17. Mogo E. Dr. AI will see you now. JSTOR Daily. URL: https://daily.jstor.org/dr-ai-will-see-you-now [accessed 2024-11-10]
18. Phiri M, Munoriyarwa A. Health chatbots in Africa: scoping review. J Med Internet Res 2023 Jun 14;25:e35573. [doi: 10.2196/35573] [Medline: 35584083]
19. Africa's SDGs progress uneven, accelerated efforts required to meet 2030 deadline: new report. United Nations Development Programme. URL: https://www.undp.org/africa/press-releases/africas-sdgs-progress-uneven-accelerated-efforts-required-meet-2030-deadline-new-report [accessed 2024-11-10]
20. 2023 Africa sustainable development report. United Nations Development Programme. URL: https://www.undp.org/africa/publications/2023-africa-sustainable-development-report[accessed [accessed 2024-11-10]
21. Heine B, Nurse D. African Languages: An Introduction. Cambridge, MA: Cambridge University Press; 2000.
22. Lodhi AY. The language situation in Africa today. Nord J Afr Stud 1993;2(1):11.
23. Weber JS, Toyama K. Remembering the past for meaningful AI-D. In: Proceedings of the 2010 Conference on Artificial Intelligence for Development, Papers from the 2010 AAAI Spring Symposium, Technical Report. 2010 Presented at: AAAI/SSS '10; March 22-24, 2010; Stanford, CA URL: https://www.researchgate.net/publication/221250994_Remembering_the_Past_for_Meaningful_AI-D
24. Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-11-10]
25. Practical guidance for scaling up health service innovations. World Health Organization. 2009. URL: https://www.who.int/publications/i/item/9789241598521 [accessed 2024-11-10]
26. Baclic O, Tunis M, Young K, Doan C, Swerdfeger H, Schonfeld J. Challenges and opportunities for public health made possible by advances in natural language processing. Can Commun Dis Rep 2020 Jun 04;46(6):161-168. [doi: 10.14745/ccdr.v46i06a02] [Medline: 32673380]
27. Tshimula JM, Kalengayi M, Makenga D, Lilonge D, Asumani M, Madiya D, et al. Artificial intelligence for public health surveillance in Africa: applications and opportunities. arXiv. Preprint posted online August 5, 2024 2024 [FREE Full text] [doi: 10.48550/arXiv.2408.02575]
28. Adebara I, Abdul-Mageed M. Towards Afrocentric NLP for African languages: where we are and where we can go. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022 Presented at: ACL '22; May 22-27, 2022; Dublin, Ireland p. 3814-3841 URL: https://aclanthology.org/2022.acl-long.265.pdf [doi: 10.18653/v1/2022.acl-long.265]
29. Liu CC, Gurevych I, Korhonen A. Culturally aware and adapted NLP: a taxonomy and a survey of the state of the art. arXiv. Preprint posted online June 6, 2024 2024 [FREE Full text] [doi: 10.48550/arXiv.2406.03930]
30. Agenda 2063: overview. African Union. URL: https://au.int/en/agenda2063/overview [accessed 2024-11-10]
31. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI Blog. 2019. URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf [accessed 2024-11-10]
32. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]
33. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. 2013 Presented at: NIPS '13; December 5-10, 2013; Lake Tahoe, NV URL: https://dl.acm.org/doi/10.5555/2999792.2999959 [doi: 10.5040/9781474284974.00399]
34. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL '19; June 2-7, 2019; Minneapolis, MN p. 4171-4186 URL: https://aclanthology.org/N19-1423.pdf [doi: 10.18653/v1/n18-2]
35. GPT-4 technical report. arXiv. Preprint posted online March 15, 2023. 2023. URL: https://arxiv.org/abs/2303.08774 [accessed 2024-11-10]
36. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. Llama: open and efficient foundation language models. arXiv. Preprint posted online February 27, 2023 2023 [FREE Full text] [doi: 10.48550/arXiv.2302.13971]
37. Ogamba M, Gitonga J, Muriithi B, Olukuru J, Sevilla J. Wellness Buddy: an AI mental health chatbot for Kenyan university students. In: Proceedings of the 2023 First International Conference on the Advancements of Artificial Intelligence in

XSL•FO
RenderX

African Context. 2023 Presented at: AAIAC '23; November 15-16, 2023; Arusha, Tanzania p. 1-6 URL: https://ieeexplore.ieee.org/document/10465291 [doi: 10.1109/aaiac60008.2023.10465291]
38. Mondal I, Ahuja K, Jain M, O'Neill J, Bali K, Choudhury M. Global readiness of language technology for healthcare: what would it take to combat the next pandemic? In: Proceedings of the 29th International Conference on Computational Linguistics. 2022 Presented at: COLING '22; October 12-17, 2022; Gyeongju, Republic of Korea p. 4320-4335 URL: https://aclanthology.org/2022.coling-1.381.pdf
39. Gangavarapu A. LLMs: a promising new tool for improving healthcare in low-resource nations. In: Proceedings of the 2023 IEEE Global Humanitarian Technology Conference. 2023 Presented at: GHTC '23; October 12-15, 2023; Radnor, PA p. 252-255 URL: https://ieeexplore.ieee.org/document/10354650 [doi: 10.1109/ghtc56179.2023.10354650]
40. Green EP, Pearson N, Rajasekharan S, Rauws M, Joerin A, Kwobah E, et al. Expanding access to depression treatment in Kenya through automated psychological support: protocol for a single-case experimental design pilot study. JMIR Res Protoc 2019 Apr 29;8(4):e11800. [doi: 10.2196/11800] [Medline: 31033448]
41. Schmude M, Salim N, Azadzoy H, Bane M, Millen E, O'Donnell L, et al. Investigating the potential for clinical decision support in sub-Saharan Africa with AFYA (artificial intelligence-based assessment of health symptoms in Tanzania): protocol for a prospective, observational pilot study. JMIR Res Protoc 2022 Jun 07;11(6):e34298 [FREE Full text] [doi: 10.2196/34298] [Medline: 35671073]
42. Ogbuokiri B, Ahmadi A, Bragazzi NL, Movahedi Nia Z, Mellado B, Wu J, et al. Public sentiments toward COVID-19 vaccines in South African cities: an analysis of Twitter posts. Front Public Health 2022;10:987376. [doi: 10.3389/fpubh.2022.987376] [Medline: 36033735]
43. Tresner-Kirsch D, Mikkelson AA, Yinka-Banjo C, Akinyemi M, Goyal S. Intent recognition on low-resource language messages in a health marketplace chatbot. In: Proceedings of the 2023 IEEE 11th International Conference on Healthcare Informatics. 2023 Presented at: ICHI '23; Houston, TX; June 26-29, 2023 p. 457-459 URL: https://www.computer.org/csdl/proceedings-article/ichi/2023/026300a457/1SN7zPjRwl2 [doi: 10.1109/ichi57859.2023.00066]
44. Fletcher RR, Olubeko O, Sonthalia H, Kateera F, Nkurunziza T, Ashby JL, et al. Application of machine learning to prediction of surgical site infection. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:2234-2237. [doi: 10.1109/EMBC.2019.8857942] [Medline: 31946345]
45. Batani J, Mbunge E, Leokana L. A deep learning-based chatbot to enhance maternal health education. In: Proceedings of the 2024 Conference on Information Communications Technology and Society. 2024 Presented at: ICTAS '24; March 7-8, 2024; Durban, South Africa p. 7-11 URL: https://ieeexplore.ieee.org/document/10507149 [doi: 10.1109/ictas59620.2024.10507149]
46. Ogbuokiri B, Ahmadi A, Nia Z, Mellado B, Wu J, Orbinski J, et al. Vaccine hesitancy hotspots in Africa: an insight from geotagged Twitter posts. IEEE Trans Comput Soc Syst 2024 Feb;11(1):1325-1338. [doi: 10.1109/tcss.2023.3236368]
47. Marais L, Louw JA, Badenhorst J, Calteaux K, Wilken I, van Niekerk N, et al. AwezaMed: a multilingual, multimodal speech-to-speech translation application for maternal health care. In: Proceedings of the 2020 IEEE 23rd International Conference on Information Fusion. 2020 Presented at: FUSION '20; July 6-9, 2020; Rustenburg, South Africa p. 1-8 URL: https://ieeexplore.ieee.org/document/9190240 [doi: 10.23919/fusion45008.2020.9190240]
48. Annapareddy N, Fallin K, Folks R, Jarrard W, Durieux M, Moradinasab N, et al. Handwritten text and digit classification on Rwandan perioperative flowsheets via YOLOv5. In: Proceedings of the 2022 Systems and Information Engineering Design Symposium. 2022 Presented at: SIEDS '22; April 28-29, 2022; Charlottesville, VA p. 270-275 URL: https://ieeexplore.ieee.org/document/9799426 [doi: 10.1109/sieds55548.2022.9799426]
49. Hassib M, Hossam N, Sameh J, Torki M. AraDepSu: detecting depression and suicidal ideation in Arabic tweets using transformers. In: Proceedings of the 7th Arabic Natural Language Processing Workshop. 2022 Presented at: WANLP '22; December 8, 2022; Abu Dhabi, United Arab Emirates p. 302-311. [doi: 10.18653/v1/2022.wanlp-1.28]
50. Mondal I, Bali K, Jain M, Choudhury M, O'Neill J, Ochieng M, et al. Language patterns and behaviour of the peer supporters in multilingual healthcare conversational forums. In: Proceedings of the 13th Language Resources and Evaluation Conference. 2022 Presented at: LREC '22; June 20-25, 2022; Marseille, France p. 963-975 URL: https://aclanthology.org/2022.lrec-1.102.pdf [doi: 10.1007/S10579-005-2692-5]
51. Ngao N, Wang Z, Nderu L, Mwalili T, August T, Ronen K. Detecting urgency in multilingual medical SMS in Kenya. In: Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: Student Research Workshop. 2022 Presented at: AACL/IJCNLP '22; November 20, 2022; Virtual Event p. 68-75 URL: https://aclanthology.org/2022.aacl-srw.10.pdf [doi: 10.18653/v1/2022.aacl-srw.10]
52. Mondal I, Bali K, Jain M, Choudhury M, Sharma A, Gitau E, et al. A linguistic annotation framework to study interactions in multilingual healthcare conversational forums. In: Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop. 2021 Presented at: LAW/DMR '21; November 3-7, 2021; Punta Cana, Dominican Republic p. 66-77 URL: https://aclanthology.org/2021.law-1.7.pdf [doi: 10.18653/v1/2021.law-1.7]
53. Daniel J, Brink W, Eloff R, Copley C. Towards automating healthcare question answering in a noisy multilingual low-resource setting. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at:

XSL•FO
RenderX

ACL '19; July 28-August 2, 2019; Florence, Italy p. 948-953 URL: https://aclanthology.org/P19-1090.pdf [doi: 10.18653/v1/p19-1090]
54. Byamugisha J, Keet CM, DeRenzi B. Evaluation of a Runyankore grammar engine for healthcare messages. In: Proceedings of the 10th International Conference on Natural Language Generation. 2017 Presented at: SIGGEN '17; September 4-7, 2017; Santiago de Compostela, Spain p. 105-113 URL: https://aclanthology.org/W17-3514.pdf [doi: 10.18653/v1/w17-3514]
55. Mojapelo ML. Semantics of body parts in African WordNet: a case of Northern Sotho. In: Proceedings of the 8th Global WordNet Conference. 2016 Presented at: GWC '16; January 27-30, 2016; Bucharest, Romania p. 234-242 URL: https://aclanthology.org/2016.gwc-1.34.pdf
56. Vorster C, Kritzinger A, Coetser LE, van der Linden J. Preliminary reliability of South African adaptation and Northern Sotho translation of the modified checklist for autism in toddlers, revised with follow-up. S Afr J Commun Disord 2021 Jul 22;68(1):e1-e7. [doi: 10.4102/sajcd.v68i1.831] [Medline: 34342485]
57. Ghanem A, Asaad C, Hafidi H, Moukafih Y, Guermah B, Sbihi N, et al. Real-time infoveillance of Moroccan social media users' sentiments towards the COVID-19 pandemic and its management. Int J Environ Res Public Health 2021 Nov 19;18(22):12172 [FREE Full text] [doi: 10.3390/ijerph182212172] [Medline: 34831927]
58. Nsoesie EO, Oladeji O, Abah AS, Ndeffo-Mbah ML. Forecasting influenza-like illness trends in Cameroon using Google search data. Sci Rep 2021 Mar 24;11(1):6713 [FREE Full text] [doi: 10.1038/s41598-021-85987-9] [Medline: 33762599]
59. van Heerden A, Bosman S, Swendeman D, Comulada WS. Chatbots for HIV prevention and care: a narrative review. Curr HIV/AIDS Rep 2023 Dec;20(6):481-486 [FREE Full text] [doi: 10.1007/s11904-023-00681-x] [Medline: 38010467]
60. Cassim N, Mapundu M, Olago V, Celik T, George JA, Glencross DK. Using text mining techniques to extract prostate cancer predictive information (Gleason score) from semi-structured narrative laboratory reports in the Gauteng province, South Africa. BMC Med Inform Decis Mak 2021 Nov 25;21(1):330 [FREE Full text] [doi: 10.1186/s12911-021-01697-2] [Medline: 34823522]
61. Ntinga X, Musiello F, Keter AK, Barnabas R, van Heerden A. The feasibility and acceptability of an mHealth conversational agent designed to support HIV self-testing in South Africa: cross-sectional study. J Med Internet Res 2022 Dec 12;24(12):e39816 [FREE Full text] [doi: 10.2196/39816] [Medline: 36508248]
62. Greyling T, Rossouw S. Positive attitudes towards COVID-19 vaccines: a cross-country analysis. PLoS One 2022;17(3):e0264994 [FREE Full text] [doi: 10.1371/journal.pone.0264994] [Medline: 35271637]
63. du Toit J, Hattingh R, Pitcher R. The accuracy of radiology speech recognition reports in a multilingual South African teaching hospital. BMC Med Imaging 2015 Mar 04;15:8 [FREE Full text] [doi: 10.1186/s12880-015-0048-1] [Medline: 25879906]
64. Movahedi Nia Z, Bragazzi NL, Ahamadi A, Asgary A, Mellado B, Orbinski J, et al. Off-label drug use during the COVID-19 pandemic in Africa: topic modelling and sentiment analysis of ivermectin in South Africa and Nigeria as a case study. J R Soc Interface 2023 Sep;20(206):20230200 [FREE Full text] [doi: 10.1098/rsif.2023.0200] [Medline: 37700708]
65. Achilonu OJ, Singh E, Nimako G, Eijkemans RM, Musenge E. Rule-based information extraction from free-text pathology reports reveals trends in South African female breast cancer molecular subtypes and Ki67 expression. Biomed Res Int 2022;2022:6157861 [FREE Full text] [doi: 10.1155/2022/6157861] [Medline: 35355821]
66. Nia ZM, Asgary A, Bragazzi N, Mellado B, Orbinski J, Wu J, et al. Nowcasting unemployment rate during the COVID-19 pandemic using Twitter data: the case of South Africa. Front Public Health 2022;10:952363 [FREE Full text] [doi: 10.3389/fpubh.2022.952363] [Medline: 36530702]
67. Babington-Ashaye A, de Moerloose P, Diop S, Geissbuhler A. Design, development and usability of an educational AI chatbot for people with Haemophilia in Senegal. Haemophilia 2023 Jul;29(4):1063-1073. [doi: 10.1111/hae.14815] [Medline: 37347648]
68. Cassim N, Mapundu M, Olago V, George JA, Glencross DK. Using big data techniques to improve prostate cancer reporting in the Gauteng province, South Africa. Stud Health Technol Inform 2019 Aug 21;264:1437-1438. [doi: 10.3233/SHTI190472] [Medline: 31438169]
69. Movassagh M, Bebell LM, Burgoine K, Hehnly C, Zhang L, Moran K, et al. Vaginal microbiome topic modeling of laboring Ugandan women with and without fever. NPJ Biofilms Microbiomes 2021 Sep 10;7(1):75 [FREE Full text] [doi: 10.1038/s41522-021-00244-1] [Medline: 34508087]
70. Mishra A. Chatbot in IsiXhosa for remote pre/post-natal care. engrXiv. Preprint posted online July 18, 2019 2019 [FREE Full text] [doi: 10.31224/osf.io/jm54z]
71. Green EP, Lai Y, Pearson N, Rajasekharan S, Rauws M, Joerin A, et al. Expanding access to perinatal depression treatment in Kenya through automated psychological support: development and usability study. JMIR Form Res 2020 Oct 05;4(10):e17895 [FREE Full text] [doi: 10.2196/17895] [Medline: 33016883]
72. Mugoye K, Okoyo H, Mcoyowo S. Smart-bot technology: conversational agents role in maternal healthcare support. In: Proceedings of the 2019 IST-Africa Week Conference. 2019 Presented at: IST-Africa '19; May 8-10, 2019; Nairobi, Kenya p. 1-7 URL: https://ieeexplore.ieee.org/document/8764817 [doi: 10.23919/istafrica.2019.8764817]
73. Sewunetie W, Beza A, Abebe H, Abuhay TM, Admass W, Hassen H, et al. Large language models for sexual, reproductive, and maternal health rights. In: Proceedings of the 2024 IEEE 12th International Conference on Healthcare Informatics.

XSL•FO
RenderX

2024 Presented at: ICHI '24; June 3-6, 2024; Orlando, FL p. 568-573 URL: https://www.computer.org/csdl/proceedings-article/ichi/2024/837300a568/1ZCgQ4gMcwg [doi: 10.1109/ichi61247.2024.00091]
74. Arsevska E, Roche M, Falala S. Monitoring disease outbreak events on the web using text-mining approach and domain expert knowledge. In: Proceedings of the 10th International Conference on Language Resources and Evaluation. 2016 Presented at: LREC '16; May 8-13, 2016; Portorož, Slovenia p. 3407-3411 URL: https://aclanthology.org/L16-1543.pdf
75. Anastasopoulos A, Cattelan A, Dou ZY, Federico M, Federmann C, Genzel D, et al. TICO-19: the translation initiative for COVID-19. In: Proceedings of the 1st Workshop on NLP for COVID-19. 2020 Presented at: EMNLP '20; July 1-5, 2020; Virtual Event URL: https://aclanthology.org/2020.nlpcovid19-2.5.pdf [doi: 10.18653/v1/2020.nlpcovid19-2.5]
76. Kim J, Bak BR, Agrawal A, Wu J, Wirtz V, Hong T, et al. COVID-19 vaccine misinformation in middle income countries. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023 Presented at: EMNLP '23; December 6-10, 2023; Singapore, Singapore p. 3903-3915 URL: https://aclanthology.org/2023.emnlp-main.237.pdf [doi: 10.18653/v1/2023.emnlp-main.237]
77. Al-Thubaity A, Alkhereyf S, Bahanshal AO. AraNPCC: the Arabic newspaper COVID-19 corpus. In: Proceedinsg of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection. 2022 Presented at: OSACT '22; June 12-15, 2022; Marseille, France p. 32-40 URL: https://aclanthology.org/2022.osact-1.4.pdf [doi: 10.18653/v1/2022.wanlp-1.1]
78. Oyebode O, Orji R. Detecting factors responsible for diabetes prevalence in Nigeria using social media and machine learning. In: Proceedings of the 15th International Conference on Network and Service Management. 2019 Presented at: CNSM '19; October 21-25, 2019; Halifax, NS p. 1-4 URL: https://ieeexplore.ieee.org/document/9012679 [doi: 10.23919/cnsm46954.2019.9012679]
79. Krishnamoorthy M, Hameed MS, Kopinski T, Schwung A. Disease prediction based on individual's medical history using CNN. In: Proceedings of the 20th IEEE International Conference on Machine Learning and Applications. 2021 Presented at: ICMLA '21; December 13-16, 2021; Pasadena, CA p. 89-94 URL: https://ieeexplore.ieee.org/document/9680243 [doi: 10.1109/icmla52953.2021.00022]
80. Mubarak H, Hassan S, Chowdhury SA, Alam F. ArCovidVac: analyzing Arabic tweets about COVID-19 vaccination. In: Proceedings of the 13th Language Resources and Evaluation Conference. 2022 Presented at: LREC '22; June 20-25, 2022; Marseille, France p. 3220-3230 URL: https://aclanthology.org/2022.lrec-1.344.pdf
81. Yuan M, Zhang M, van Durme B, Findlater L, Boyd-Graber J. Interactive refinement of cross-lingual word embeddings. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020 Presented at: EMNLP '20; November 16-20, 2020; Virtual Event p. 5984-5996 URL: https://aclanthology.org/2020.emnlp-main.482.pdf [doi: 10.18653/v1/2020.emnlp-main.482]
82. Yadav A, Vishwakarma DK. A language-independent network to analyze the impact of COVID-19 on the world via sentiment analysis. ACM Trans Internet Technol 2021 Sep 14;22(1):1-30. [doi: 10.1145/3475867]
83. Verma M, Moudgil N, Goel G, Pardeshi P, Joseph J, Kumar N, et al. People's perceptions on COVID-19 vaccination: an analysis of twitter discourse from four countries. Sci Rep 2023 Aug 31;13(1):14281 [FREE Full text] [doi: 10.1038/s41598-023-41478-7] [Medline: 37653001]
84. Marcus JL, Sewell WC, Balzer LB, Krakower DS. Artificial intelligence and machine learning for HIV prevention: emerging approaches to ending the epidemic. Curr HIV/AIDS Rep 2020 Jun;17(3):171-179 [FREE Full text] [doi: 10.1007/s11904-020-00490-6] [Medline: 32347446]
85. Thiabaud A, Triulzi I, Orel E, Tal K, Keiser O. Social, behavioral, and cultural factors of HIV in Malawi: semi-automated systematic review. J Med Internet Res 2020 Aug 14;22(8):e18747 [FREE Full text] [doi: 10.2196/18747] [Medline: 32795992]
86. Kiyasseh D, Zhu T, Clifton D. The promise of clinical decision support systems targetting low-resource settings. IEEE Rev Biomed Eng 2022;15:354-371. [doi: 10.1109/rbme.2020.3017868]
87. Alshalan R, Al-Khalifa H, Alsaeed D, Al-Baity H, Alshalan S. Detection of hate speech in COVID-19-related tweets in the Arab region: deep learning and topic modeling approach. J Med Internet Res 2020 Dec 08;22(12):e22609 [FREE Full text] [doi: 10.2196/22609] [Medline: 33207310]
88. Martin S, Kilich E, Dada S, Kummervold PE, Denny C, Paterson P, et al. "Vaccines for pregnant women…?! absurd" - mapping maternal vaccination discourse and stance on social media over six months. Vaccine 2020 Sep 29;38(42):6627-6637. [doi: 10.1016/j.vaccine.2020.07.072] [Medline: 32788136]
89. Rosenzweig LR, Offer-Westort M. Testing interventions to address vaccine hesitancy on Facebook in East and West Africa. Open Science Framework. 2022 Jan 07. URL: https://osf.io/m4e6d/ [accessed 2024-11-10]
90. Keet CM. Natural language generation requirements for social robots in sub-Saharan Africa. In: Proceedings of the 2021 IST-Africa Conference. 2021 Presented at: IST-Africa '21; May 10-14, 2021; Cape Town, South Africa p. 1-8 URL: https://ieeexplore.ieee.org/document/9576993
91. Daniel JE. Applications of natural language processing for low-resource languages in the healthcare domain. Faculty of Science at Stellenbosch University. 2020 Mar. URL: https://scholar.sun.ac.za/server/api/core/bitstreams/7a1beb3b-4859-43d2-b1c8-54fc80811e7d/content [accessed 2024-11-10]

XSL•FO
RenderX

92. Manaka T, Zyl TV, Kar D, Wade A. Multi-step transfer learning in natural language processing for the health domain. Neural Process Lett 2024 May 20;56(3):177. [doi: 10.1007/S11063-024-11526-Y]
93. Gbashi S, Adebo OA, Doorsamy W, Njobeh PB. Systematic delineation of media polarity on COVID-19 vaccines in Africa: computational linguistic modeling study. JMIR Med Inform 2021 Mar 16;9(3):e22916 [FREE Full text] [doi: 10.2196/22916] [Medline: 33667172]
94. Stone K, Ojo S, Tu C. Natural language processing of South African online question-answering health forum HIV and AIDS dataset using topic modeling. In: Nagar AK, Jat DS, Mishra DK, Joshi A, editors. Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 2. Cham, Switzerland: Springer; 2022:387-395.
95. Olatunji T, Afonja T, Yadavalli A, Emezue CC, Singh S, Dossou BF, et al. AfriSpeech-200: pan-African accented speech dataset for clinical and general domain ASR. Trans Assoc Comput Linguist 2023;11:1669-1685 [FREE Full text] [doi: 10.1162/tacl_a_00627]
96. Tresner-Kirsch D, Yinka-Banjo C, Akinyemi M, Ajayi O. Auditing natural language processing for gender equality in sub-Saharan African healthcare systems: framework development and evaluation. JMIRP. Preprint posted online August 18, 2023 2023 Aug [FREE Full text] [doi: 10.2196/preprints.51954]
97. Oyebode O, Orji R. Likita: a medical chatbot to improve healthcare delivery in Africa. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC URL: https://www.hcixb.org/papers_2018/hcixb18-final-22.pdf
98. Olaleye TO. Opinion mining analytics for spotting omicron fear-stimuli using REPTree classifier and natural language processing. IJRASET 2022 Jan 31;10(1):995-1005. [doi: 10.22214/ijraset.2022.39903]
99. Magassouba AS, Sylla Y, Toure AT, Leno NN, Diallo DB, Camara G, et al. A mixed cross-sectional study with natural language processing analysis on computer literacy and access among healthcare workers in Guinea. J Health Inform Dev Ctries 2023;17(02):1-14 [FREE Full text] [doi: 10.21203/rs.3.rs-122687/v1]
100. Lester RT, Manson M, Semakula M, Jang H, Mugabo H, Magzari A, et al. Natural language processing to evaluate texting conversations between patients and healthcare providers during COVID-19 home-based care in Rwanda at scale. PLOS Digit Health 2025 Jan;4(1):e0000625. [doi: 10.1371/journal.pdig.0000625] [Medline: 39813181]
101. Disrupt A. Kenya's SophieBot selected for Merck Nairobi accelerator. Disrupt Africa. 2017. URL: https://disruptafrica.com/2017/09/22/kenyas-sophiebot-selected-for-merck-nairobi-accelerator/ [accessed 2024-11-06]
102. We use AI to support and optimize health decisions. ElsaHealth. URL: https://www.elsa.health/ [accessed 2024-11-06]
103. Mbaza AI chatbot. Digital Umuganda. URL: https://digitalumuganda.com/chatbot/ [accessed 2024-11-06]
104. Jacaranda launches open-source LLM in five African languages. Jacaranda Health. URL: https://jacarandahealth.org/jacaranda-launches-open-source-llm-in-five-african-languages/ [accessed 2024-11-06]
105. Aurum Institute Ghana receives grand challenges grant for catalyzing equitable artificial intelligence (AI) use. Aurum Institute. URL: https://www.auruminstitute.org/component/content/article/45-press-releases/772-aurum-institute-ghana-receives-grand-challenges-grant-for-catalyzing-equitable-artificial-intelligence-ai-use?Itemid=101 [accessed 2024-11-06]
106. The largest study on LLMs in African healthcare. AfriMedQA. URL: https://afrimedqa.com/ [accessed 2024-11-06]
107. Ada supporting mothers in South Africa. Ada Health GmbH. 2023. URL: https://ada.com/press/230905-ada-supporting-mothers-in-southafrica/ [accessed 2024-11-06]
108. World's first AI health guidance app in Swahili. Ada Health GmbH. URL: https://ada.com/press/191119-worlds-first-ai-health-guidance-app-in-swahili/ [accessed 2024-11-06]
109. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
110. Hu S, Oppong A, Mogo E, Barford A, Occhini G, Collins C, et al. Review protocol: a scoping review of natural language processing technologies for public health in Africa. medRxiv. Preprint posted online July 3, 2024 2024 [FREE Full text] [doi: 10.1101/2024.07.02.24309815]
111. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
112. Pradhan P, Costa L, Rybski D, Lucht W, Kropp JP. A systematic study of sustainable development goal (SDG) interactions. Earth's Future 2017 Nov 30;5(11):1169-1179. [doi: 10.1002/2017EF000632]
113. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2024-11-10]
114. Bishop M, Green M. Philanthrocapitalism Rising. Soc 2015 Nov 5;52(6):541-548. [doi: 10.1007/S12115-015-9945-8]
115. McCoy D, Kembhavi G, Patel J, Luintel A. The Bill and Melinda Gates Foundation's grant-making programme for global health. Lancet 2009 May 09;373(9675):1645-1653. [doi: 10.1016/S0140-6736(09)60571-7] [Medline: 19427959]
116. Economic impact of big tech platforms on the viability of local broadcast news. BIA Advisory Services & National Association of Broadcasters. 2021. URL: https://www.nab.org/documents/newsRoom/pdfs/Economic_Impact_Tech_Platforms_Broadcast_News.pdf[accessed [accessed 2024-11-10]

XSL•FO
RenderX

117. The sustainable development goals report 2024. United Nations. 2024. URL: https://unstats.un.org/sdgs/report/2024/ [accessed 2024-11-10]
118. Sustainable development report 2024: interactive map. Sustainable Development Solutions Network. URL: https://dashboards.sdgindex.org/map [accessed 2024-11-10]
119. Cerf ME. The Sustainable development goals: contextualizing Africa's economic and health landscape. Glob Chall 2018 Aug;2(8):1800014 [FREE Full text] [doi: 10.1002/gch2.201800014] [Medline: 31565342]
120. LeFevre AE, Shah N, Bashingwa JJ, George AS, Mohan D. Does women's mobile phone ownership matter for health? Evidence from 15 countries. BMJ Glob Health 2020 May;5(5):e002524 [FREE Full text] [doi: 10.1136/bmjgh-2020-002524] [Medline: 32424014]
121. Chitungo I, Mhango M, Mbunge E, Dzobo M, Musuka G, Dzinamarira T. Utility of telemedicine in sub-Saharan Africa during the COVID-19 pandemic. A rapid review. Hum Behav Emerg Technol 2021 Dec;3(5):843-853 [FREE Full text] [doi: 10.1002/hbe2.297] [Medline: 34901772]
122. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. Am J Public Health 2010 Apr;100(S1):S40-S46. [doi: 10.2105/ajph.2009.184036]
123. Amugongo LM, Bidwell NJ, Corrigan CC. Invigorating ubuntu ethics in AI for healthcare: enabling equitable care. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2023 Presented at: FAccT '23; June 12-15, 2023; Chicago, IL p. 583-592 URL: https://dl.acm.org/doi/10.1145/3593013.3594024 [doi: 10.1145/3593013.3594024]
124. Proefke R, Barford A. Creating spaces for co-research. J Br Acad 2023 Nov 2;11:19-42. [doi: 10.5871/jba/011s3.019]
125. Habwe J. The role of Kiswahili in the integration of East Africa. J Pan Afr Stud 2009:2-10 [FREE Full text]
126. Emeka-Nwobia NU. The place of indigenous Nigerian languages in national development. Res Humanit Soc Sci 2015;5(12):112-116. [doi: 10.7575/aiac.ijalel.v.2n.3p.29]
127. Hu S, Zhou H, Yuan M, Gritta M, Zhang G, Iacobacci I, et al. A systematic study of performance disparities in multilingual task-oriented dialogue systems. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023 Presented at: EMNLP '23; December 6-10, 2023; Singapore, Singapore p. 6825-6851 URL: https://aclanthology.org/2023.emnlp-main.422.pdf [doi: 10.18653/v1/2023.emnlp-main.422]
128. Barocas S, Hardt M, Narayanan A. Fairness and Machine Learning: Limitations and Opportunities. Mahwah, NJ: MIT Press; 2023.
129. Orife I, Kreutzer J, Sibanda B, Whitenack D, Siminyu K, Martinus L, et al. Masakhane -- machine translation for Africa. arXiv. Preprint posted online March 13, 2020 2020 [FREE Full text] [doi: 10.48550/arXiv.2003.11529]
130. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, et al. Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 Presented at: ACL '20; July 5-10, 2020; Virtual Event p. 8440-8451 URL: https://aclanthology.org/2020.acl-main.747.pdf [doi: 10.18653/v1/2020.acl-main.747]
131. Conneau A, Rinott R, Lample G, Williams A, Bowman S, Schwenk H, et al. XNLI: evaluating cross-lingual sentence representations. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018 Presented at: EMNLP '18; October 31-November 4, 2018; Brussels, Belgium p. 2475-2485 URL: https://aclanthology.org/D18-1269.pdf [doi: 10.18653/v1/d18-1269]
132. Hu J, Ruder S, Siddhant A, Neubig G, Firat O, Johnson M. XTREME: a massively multilingual multi-task benchmark for evaluating cross-lingual generalization. In: Proceedings of the 37th International Conference on Machine Learning. 2020 Presented at: ICML '20; July 13-18, 2020; Virtual Event p. 4411-4421 URL: https://dl.acm.org/doi/pdf/10.5555/3524938.3525348
133. Hovy D, Yang D. The importance of modeling social factors of language: theory and practice. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021 Presented at: NAACL '21; June 6-11, 2021; Virtual Event p. 588-602 URL: https://aclanthology.org/2021.naacl-main.49.pdf [doi: 10.18653/v1/2021.naacl-main.49]
134. Loveys K, Torrez J, Fine A, Moriarty G, Coppersmith G. Cross-cultural differences in language markers of depression online. In: Proceedings of the 5th Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: CLPsych '18; June 5, 2018; New Orleans, LA p. 78-87 URL: https://aclanthology.org/W18-0608.pdf [doi: 10.18653/v1/w18-0608]
135. Traditional medicine: questions and answers. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/traditional-medicine [accessed 2024-11-10]
136. Olatunji T, Nimo C, Owodunni A, Abdullahi T, Ayodele E, Sanni M, et al. AfriMed-QA: a Pan-African, multi-specialty, medical question-answering benchmark dataset. arXiv. Preprint posted online November 23, 2024 2024 [FREE Full text] [doi: 10.48550/arXiv.2411.15640]
137. Conforti C, Hirmer S, Morgan D, Basaldella M, Ben OY. Natural language processing for achieving sustainable development: the case of neural labelling to enhance community profiling. In: Proceedings of the 2020 Conference on Empirical Methods

XSL•FO
RenderX

in Natural Language Processing. 2020 Presented at: EMNLP '20; November 16-20, 2020; Virtual Event p. 8427-8444 URL: https://aclanthology.org/2020.emnlp-main.677.pdf [doi: 10.18653/v1/2020.emnlp-main.677]
138. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
139. Barford A, Ahmad SR. Levers for a corporate transition to a plastics circular economy. Bus Strat Env 2022 Jun 29;32(4):1203-1217. [doi: 10.1002/bse.3182]
140. Barford A, Beales A, Zorila M. An expert study of systemic influences on progress towards living wages: a key to unlock the sustainable development goals. Bus Strat Dev 2025 Jan 20;8(1):55. [doi: 10.1002/bsd2.70048]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations
EPHF: essential public health function
GPT-2: Generative Pre-trained Transformer 2
NGO: nongovernmental organization
NLP: natural language processing
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analyses—Protocols
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SDG: sustainable development goal
WHO: World Health Organization

Edited by A Mavragani; submitted 12.11.24; peer-reviewed by A Hidki, G Xiao; comments to author 20.12.24; revised version received 09.01.25; accepted 24.01.25; published 05.03.25.

Please cite as:
Hu S, Oppong A, Mogo E, Collins C, Occhini G, Barford A, Korhonen A
Natural Language Processing Technologies for Public Health in Africa: Scoping Review
J Med Internet Res 2025;27:e68720
URL: https://www.jmir.org/2025/1/e68720
doi:10.2196/68720
PMID:40053738

©Songbo Hu, Abigail Oppong, Ebele Mogo, Charlotte Collins, Giulia Occhini, Anna Barford, Anna Korhonen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Facilitators and Barriers to the Implementation of Digital Health Technologies in Hospital Settings in Lower- and Middle-Income Countries Since the Onset of the COVID-19 Pandemic: Scoping Review

Sheng Qian Yew[1], MBBS, BMedSci, MSc, MPH, PhD; Daksha Trivedi[2], BSc, MSc, MPhil, PhD; Nurul Iman Hafizah Adanan[3], BSc, MSc; Boon How Chew[4], MD, MMED, PhD

[1]Department of Public Health Medicine, Faculty of Medicine, National University of Malaysia, Cheras, Malaysia

[2]Centre for Research in Public Health and Community Care, University of Hertfordshire, Hertforshire, United Kingdom

[3]Clinical Research Unit, Hospital Sultan Abdul Aziz Shah, Serdang, Malaysia

[4]Faculty of Medicine and Health Sciences, Department of Family Medicine, Serdang, Malaysia

Corresponding Author:
Daksha Trivedi, BSc, MSc, MPhil, PhD
Centre for Research in Public Health and Community Care
University of Hertfordshire
College Lane, Hatfield
Hertforshire, AL10 9AB
United Kingdom
Phone: 44 01707286389
Email: d.trivedi@herts.ac.uk

Abstract

Background: Although the implementation process of digital health technologies (DHTs) has been extensively documented in high-income countries, the factors that facilitate and prevent their implementation in lower- and middle-income countries (LMICs) may differ for various reasons.

Objective: To address this gap in research, this scoping review aims to determine the facilitators and barriers to implementing DHTs in LMIC hospital settings following the onset of the COVID-19 pandemic. Additionally, the review outlined the types of DHTs that have been implemented in LMICs' hospitals during this pandemic and finally developed a classification framework to categorize the landscape of DHTs.

Methods: Systematic searches were conducted on PubMed, Scopus, Web of Science, and Google Scholar for studies published from March 2020 to December 2023. We extracted data on authors, publication years, study objectives, study countries, disease conditions, types of DHTs, fields of clinical medicine where the DHTs are applied, study designs, sample sizes, characteristics of the study population, study location, and data collection methods of the included studies. Both quantitative and qualitative data were utilized to conduct a thematic analysis, using a deductive method based on the Practical, Robust Implementation and Sustainability Model (PRISM), to identify facilitators and barriers to DHT implementation. Finally, all accessible DHTs were identified and organized to create a novel classification framework.

Results: Twelve studies were included from 292 retrieved articles. Telemedicine (n=5) was the most commonly used DHT in LMICs' hospitals, followed by hospital information systems (n=4), electronic medical records (n=2), and mobile health (n=1). These 4 DHTs, among the other existing DHTs, allowed us to develop a novel classification framework for DHTs. The included studies used qualitative methods (n=4), which included interviews and focus groups, quantitative methods (n=5), or a combination of both (n=2). Among the 64 facilitators of DHT implementation, the availability of continuous on-the-job training (n=3), the ability of DHTs to prevent cross-infection (n=2), and positive previous experiences using DHTs (n=2) were the top 3 reported facilitators. However, of the 44 barriers to DHT implementation, patients with poor digital literacy and skills in DHTs (n=3), inadequate awareness regarding DHTs among health care professionals and stakeholders (n=2), and concerns regarding the accuracy of disease diagnosis and treatment through DHTs (n=2) were commonly reported.

Conclusions: In the postpandemic era, telemedicine, along with other DHTs, has seen increased implementation in hospitals within LMICs. All facilitators and barriers can be categorized into 6 themes, namely, (1) Aspects of the Health Care System; (2)

XSL•FO
RenderX

Perspectives of Patients; (3) External Environment; (4) Implementation of Sustainable Infrastructure; (5) Characteristics of Health Care Organization; and (6) Characteristics of Patients.

(J Med Internet Res 2025;27:e63482) doi:10.2196/63482

KEYWORDS

digital health implementation; facilitators; barriers; digital health classification framework; lower- and middle-income countries

Introduction

Populations residing in lower- and middle-income countries (LMICs) face numerous unmet health care needs due to various factors [1], including an aging population [2], escalating health care costs [3], widening income disparities [4], increased child morbidity and mortality [5], the emergence of new epidemics and pandemics [6], and growing racial discrimination in health care access [7]. The COVID-19 pandemic has further exacerbated the existing disparities and limitations of health care systems in LMICs, further exposing the issues of understaffing, underfunding, inadequate infrastructure, limited access to testing and treatment, and vulnerability to health emergencies [8]. In response to this unprecedented pandemic, many hospitals in LMICs have attempted to leverage digital health technologies (DHTs) as an innovative approach to curb the spread of the SARS-CoV-2, improve health care provision, and strengthen pandemic response efforts [9].

Generally, DHTs are defined as a set of information and communications technologies utilized in medicine and health care to manage illnesses and promote wellness [10]. These technologies have expanded as a transformative force in health care since the onset of the COVID-19 pandemic era, thereby offering a myriad of benefits that revolutionize health care delivery and improve patient outcomes. These technologies encompass a wide range of digital tools, including mobile health (mHealth) [11], telemedicine [12], wearable technologies [13], electronic medical records [14], big data analytics [15], Internet of Medical Things [16], blockchain in health care [17], metaverse [18], software as a medical device [19], augmented reality [20], and virtual reality [21]. With increasing computing power and appreciation of artificial intelligence [22] and machine learning [23] in health and medicine, many such smart tools are an making appearance in different aspects of hospital care [24,25].

Despite the benefits of DHTs, such as improved access to health care, enhanced patient engagement and empowerment, efficient health care delivery, timely and personalized care, remote monitoring, and data-driven decision-making, the implementation process of DHTs in the hospital settings, including its facilitators and barriers, is mostly described in the context of high-income countries [26,27]. Given the differences in resources, infrastructure, health care systems, socioeconomic status, level of digital divide, and regulatory frameworks in the LMICs compared with the high-income countries, the evidence on facilitators and barriers to implementing DHTs reported in previous literature may not apply to the LMICs [28]. To address these research gaps, this scoping review aimed to:

- Provide an overview of the facilitators and barriers in implementing DHTs within hospital settings in LMICs since the onset of the COVID-19 pandemic.
- Identify and describe the types of DHTs that have been put into practice in hospitals within LMICs since the onset of the COVID-19 pandemic.
- Develop a classification framework to better define the landscape of DHTs, providing a more comprehensive and practical understanding.

Methods

Design

The scoping review was conducted using the methodological framework developed by Arksey and O'Malley [29]. The protocol has been registered in the Open Science Framework and has been previously published [30]. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [27,31] was used to conduct and report our findings.

Identifying Relevant Literature

To comprehensively identify the literature relevant to DHTs, a broad, sensitive, and specific search strategy was applied to capture all DHT-related literature. With the assistance of an information technologist, a comprehensive list of literature relevant to DHTs was identified according to the criteria below. Textboxes 1 and 2 describe the inclusion and exclusion criteria adopted for this scoping review. The focus was on the period between March 2020 and December 2023, as the World Health Organization (WHO) officially declared COVID-19 as a pandemic in March 2020.

XSL•FO
RenderX

Textbox 1. Inclusion criteria.

- Studies that were conducted in lower- and middle-income countries (LMICs). (The World Bank classifies countries by income each year, covering all nations with a population over 30,000. In 2023, countries are divided into 4 income categories based on their gross national income (GNI) per capita. Low-income countries are defined as those with a GNI per capita of US $1145 or less. LMICs have a GNI per capita between US $1146 and US $4515, while upper-middle-income countries fall between US $4516 and US $14,005. High-income countries are those with a GNI per capita exceeding US $14,005. Based on this classification, LMICs in the current review are defined as countries with GNI below US $14,005 per capita.)
- Studies that reported the implementation of digital health technologies in hospital settings (for both acute and chronic conditions).
- Studies that were reported between March 2020 and December 2023.
- Studies that were reported in the English language.
- Qualitative studies (phenomenology, ethnography, grounded theory, case study, etc), quantitative studies (case-control, cohort study, cross-sectional, randomized controlled trials, etc), mixed methods studies, and reviews (narrative review, scoping review, systematic review, meta-analysis, etc).
- Relevant gray literature (eg, Google Scholar).

Textbox 2. Exclusion criteria.

- Studies that were nondigital-based (ie, studies that did not investigate the effectiveness of digital interventions, such as paper-based studies and postage surveys).
- Studies that implemented digital health technologies (DHTs) in primary care or community settings alone. However, studies that concurrently reported on the implementation of DHTs in hospital settings will still be included.
- Studies that used DHTs in dentistry and nonclinical medicine area (eg, dentistry, basic sciences, medical education, medical engineering, nutrition, dietetics, veterinary science, laboratory experimentations, and medical anthropology).

Systematic searches were conducted in PubMed, Scopus, and Web of Science databases. Potentially relevant gray literature was searched through targeted searches of Google Scholar. Lateral searching included screening reference lists in identified studies or reviews for relevant publications. Articles published in English between March 2020 and December 2023 were retrieved. Two investigators (SQY and NIHA) independently performed literature searches in the aforementioned electronic databases.

The search strategy was developed based on the "Population-Concept-Context" (PCC) framework as recommended by the Joanna Briggs Institute for Scoping Reviews [32] (Table 1). It aimed to identify the intersection between the "Population," "Concept," and "Context." Based on the PCC framework, the search strategy was "Population combined" AND "Concept" AND "Context." The details of the search strategy and search terms are tabulated in Multimedia Appendix 1.

Table 1. The PCC[a] framework used to generate search terms.

Framework	Search terms
Population	• Population 1: hospital settings (search terms as in Multimedia Appendix 1). • Population 2: LMICsb (search terms as in Multimedia Appendix 1). • Population combined: Population 1 AND Population 2.
*C*oncept	• Digital health technologies (search terms as in Multimedia Appendix 1).
*C*ontext	• Facilitators, barriers, and implementation.

[a]PCC: Population-Concept-Context.

[b]LMICs: lower- and middle-income countries

Study Selection

Records were downloaded in Rayyan software (an artificial intelligence–assisted article screening software) [24] and after deduplication, all titles and abstracts were screened independently against the inclusion criteria by 2 reviewers (SQY and NIHA). Among all titles and abstracts found, 70 out of 279 (25%) were randomly selected and screened to establish interreviewer reliability. The interreviewer reliability (κ) was 0.63, indicating good reliability. Full-text articles of potentially relevant papers identified were screened independently by SQY and NIHA. Disagreements were resolved via discussion, with consultation from a third reviewer (BHC) if needed to reach a consensus. We included studies that met the review criteria and reported on barriers and facilitators of DHT implementation. For this review, DHTs were defined as a set of information and communications technologies used in medicine and health care to manage illnesses and to promote wellness [10].

XSL•FO
RenderX

Charting the Data

Two authors (SQY and NIHA) independently extracted the following data from the included papers: authors, publication years, study objectives, study countries, disease conditions, types of DHTs, fields of clinical medicine where the DHTs are applied, study designs, sample sizes, characteristics of the study population, study location, and data collection methods of the intervention or program. We categorized the DHTs according to the types of equipment and method of operation. Any disagreements were resolved by discussion and consultation with a third author (BHC).

Collating, Summarizing, and Reporting the Results

Descriptions of the included studies, such as study countries, types of DHTs, fields of clinical medicine where the DHTs are applied, and study designs were reported using descriptive statistics (eg, frequency distribution). Facilitators and barriers to DHT implementation, which were in the form of quantitative and qualitative data, were thematically analyzed using a deductive approach. To guide the thematic analysis, facilitators and barriers to DHT implementation were organized using the Practical, Robust Implementation, and Sustainability Model (PRISM) [33].

During the coding process, 2 authors (SQY and NIHA) identified the key components of the PRISM relevant to DHT implementation. They mapped the facilitators and barriers reported in the included studies to the predefined categories and subsequently grouped similar categories under overarching themes. By organizing the facilitators and barriers into themes and categories, patterns, trends, and relationships among the factors influencing DHT implementation can be identified, providing a structured framework for understanding the complexities of implementing DHTs in LMICs' hospital settings [34].

Consultation With Stakeholders

Clinicians and information technology (IT) experts from a local teaching hospital were invited to help interpret and contextualize the findings. Through interactive discussions and collaborative sessions, stakeholders provided valuable input on the categorization of facilitators and barriers as well as the implications of implementation of DHTs in their respective contexts. Moving forward, the authors plan to continue engagement with these stakeholders to disseminate the findings of the scoping review through presentations, policy briefs, and peer-reviewed publications.

Results

Literature Search

A systematic search yielded 295 titles and abstracts. After removing 16 duplicates, 279 unique articles remained. Among these, 157 were excluded based on title or abstract review. We thoroughly evaluated 122 full-text articles, and ultimately, 12 met the eligibility criteria. The PRISMA flow diagram was used to illustrate the search decision process of the scoping review [35] (Figure 1 and Multimedia Appendix 2).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. LMIC: lower- and middle-income country.

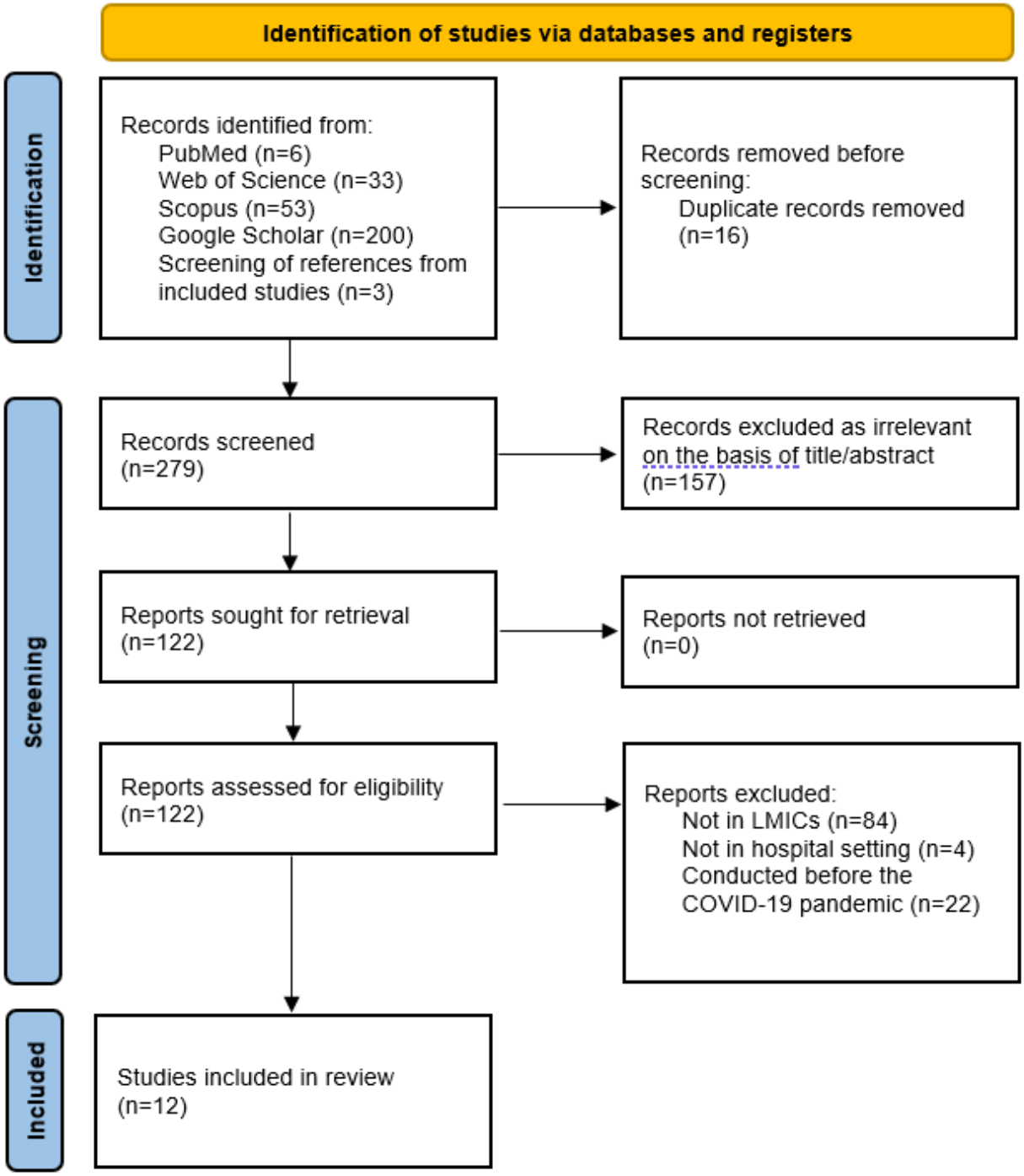

Characteristics of the Included Studies

Of the 12 articles, 6 (50%) were conducted in Asia [36-41], 4 (33%) in Africa [42-45], 1 (8%) in the Middle East [46], and 1 (8%) in South America [47]. Most of the studies (10/12, 83%) [36,37,40-47] did not specify the specific field of clinical medicine where DHTs were applied. However, some reported implementation in respiratory medicine (1/12, 8%) [38] and cancer (1/12, 8%) [39]. Telemedicine (5/12, 42%) [37,38,40,41,46] was the most commonly reported technology, followed by hospital information systems (4/12, 33%) [36,42,45,47], electronic medical records (2/12, 17%) [43,44], and mHealth (1/12, 8%) [39]. In terms of study design, there were 5 (42%) cross-sectional studies [37,41,42,44,46], 4 (33%) qualitative studies [38-40,43], 2 (17%) mixed method studies [45,47], and 1 (8%) case study [36]. The participant numbers ranged from 12 to 3386 individuals (Table 2).

XSL•FO
RenderX

Table 2. Overview of the characteristics of the included studies.

Authors and publication years	Study details								
	Objective(s)	Country	Disease condition	Types of DHTs[a]	Field of clinical medicine	Design	Sample sizes and characteristics of the population	Location	Data collection method
Abdulai et al [36]	To investigate the attitudes, opportunities, and challenges in using health information system.	India	General	Health information system	General	Case study	20 health care providers and managers	Apollo Hospital and Medanta Hospital	Self-administered questionnaire
Alboraie et al [37]	To assess the usefulness of telemedicine and the different barriers hindering its utilization.	Egypt	General	Telemedicine	General medicine, surgery, and radiology	Cross-sectional study	642 health care providers	All hospitals across Egypt	Online self-administered questionnaire
Baradwan and Al-Hanawi [46]	To gain a holistic understanding of the perceptions and barriers of the end users (participants) toward the utility of telemedicine.	Saudi Arabia	General	Telemedicine	General	Cross-sectional study	1024 individuals	Nationwide	Self-administered questionnaire
Jiang et al [38]	To explore the perceptions and experiences of older patients and health care providers in the application of telehealth and online health information to chronic disease management of chronic obstructive pulmonary disease.	China	Chronic obstructive pulmonary disease	Telemedicine	Respiratory medicine	Qualitative	31 older patients with chronic obstructive pulmonary disease and 23 health care providers	A community hospital in Jiangnan, China	In-depth interviews
Mekuria et al [42]	To assess the level of health information system utilization among health professionals in public health facilities.	Ethiopia	General	Health information system	General	Cross-sectional study	378 health care providers	Dire Dawa Administration in eastern Ethiopia	Self-administered questionnaire
Mussi et al [47]	To understand the implementation of a hospital information system in university hospitals.	Brazil	General	Health information system	General	Exploratory mixed methods	770 health managers (24 in in-depth interviews, 10 in focus group discussions, and 736 in cross-sectional study)	5 university hospitals in Brazil	Interviews, focus group discussions, questionnaires, and documentary research

XSL•FO
RenderX

Authors and publication years	Study details								
	Objective(s)	Country	Disease condition	Types of DHTs[a]	Field of clinical medicine	Design	Sample sizes and characteristics of the population	Location	Data collection method
Ngugi et al [43]	To explore end users' perceptions and experiences on factors facilitating and hindering electronic medical record use in health care facilities.	Kenya	General	Electronic medical record	General	Qualitative	20 health care providers	20 health care facilities in Kenya	Focus group discussion
Ning et al [39]	To explore the needs and perceptions of patients with head and neck cancer regarding mobile health–based physical activity programs.	China	Head and neck cancer	Mobile health	ENT[b]	Qualitative	17 patients diagnosed with head and neck cancer	First Hospital of Shanxi Medical University, Taiyuan, China	In-depth interviews
Shardha et al [40]	To explore the view of health care professionals regarding the benefits, challenges, and prospects of telemedicine to address the gap that hinders its effective use in the rural areas.	Pakistan	General	Telemedicine	Nonsurgical disciplines	Qualitative (narrative)	12 health care providers	Two tertiary hospitals in Sindh, Pakistan	In-depth interviews
Tesfa et al [44]	To assess the level of electronic health record utilization and associated factors among health care professionals at teaching hospitals.	Ethiopia	General	Electronic medical record	General	Cross-sectional study	383 health care providers	University of Gondar Specialized Teaching Hospital and Tibebe Ghion Specialized Teaching Hospital	Self-administered questionnaire
Tilahun et al [45]	To evaluate the outcomes and share experiences of working with universities to strengthen the national health information system.	Ethiopia	General	Health information system	General	Mixed methods study	23 health care providers	47 health care organizations in Ethiopia	In-depth interviews

XSL•FO
RenderX

Authors and publication years	Study details								
	Objective(s)	Country	Disease condition	Types of DHTs[a]	Field of clinical medicine	Design	Sample sizes and characteristics of the population	Location	Data collection method
Yu-Tong et al [41]	To assess telehealth readiness among clinical nurses and explore the factors that affect their telehealth readiness.	China	General	Telemedicine	General	Cross-sectional study	3386 nurses	19 hospitals in China	Self-administered questionnaire

[a]DHT: digital health technology.

[b]ENT: ear, nose, and throat.

Facilitators and Barriers to DHT Implementation

Among the 12 studies, 9 explored both the facilitators and barriers to implementing DHTs in hospital settings in LMICs since the onset of COVID-19 [36-40,43-45,47]. Two studies focused solely on facilitators [41,42], while 1 solely addressed barriers [46]. A total of 63 facilitators and 44 barriers were identified. Subsequently, these facilitators and barriers were systematically organized into 6 themes according to the PRISM, which include the following: (1) Aspects of the Health Care System; (2) Perspectives of Patients; (3) External Environment; (4) Implementation of Sustainable Infrastructure; (5) Characteristics of Health Care Organization; and (6) Characteristics of Patients. These are outlined in Table 3.

XSL•FO
RenderX

Table 3. Facilitators and barriers to DHT[a] implementation in hospital settings in lower- and middle-income countries after the COVID-19 pandemic.

Themes and categories	Codes for facilitators	Codes for barriers
Aspects of the Health Care System (n^b=11)		
Readiness (n_f[c]=4; n_b[d]=3)	• Understand the basics of information technology [37] • Knowledge in protecting data confidentiality [37] • High level of self-efficacy [42] • Good attitude toward DHTs [42] • High awareness about the implementation of DHTs [47] • High personal interest and motivation [47] • The high willingness of health care providers to provide DHTs for patients [41]	• Inadequate awareness regarding DHTs [40,44] • Feeling that DHTs are only complementary to traditional health care [38] • Preference of some health care providers for printed materials [44]
Strength of the evidence base (n_f=1; n_b=2)	• The ability of DHTs to offer statistical analysis and reports for decision-making [36] • The ability of DHTs to offer timely and up-to-date patient information [36]	• Concerns regarding accuracy in disease diagnosis and treatment [40,46]
Addresses barriers to frontline staff (n_f=1; n_b=2)	• Easy access to computers in the working area [44]	• Fear of inappropriate data protection might breach patient's privacy [37] • Inadequate computers at the workplace and service point local area network at the workplace [43]
Coordination across departments and specialties (n_f=1; n_b=1)	• Use of single information technology policy and standard hospital information system [47]	• Lack of interoperability standards between systems [47]
Burden (complexity and cost; n_f=4; n_b=1)	• Perceive that operations of DHTs are not complex [42] • The availability of information technology budget [47] • Regular upgrades on DHT hardware [43] • Funding support focusing on research and learning opportunities on DHTs [45]	• The need for retrospective data entry into systems [43]
Usability and adaptability (n_f=4; n_b=3)	• The ability of DHTs to prevent cross-infection [38,40] • Ability to provide health care in remote areas [40] • The functionalities of DHTs align with hospital's practices [47] • User-friendliness of DHTs [43]	• Difficulty in performing physical examination with DHTs [40] • Lack of compliance among health care professionals and patients [40] • DHTs might reduce the quality of medical services by increasing the probability of medical mistakes [37] • Perceive DHTs as research agenda only [45] • Some DHTs are not user-friendly [36]
Trialability and reversibility (n_f=1; n_b=0)	• Conduct beta tests and pilot tests before implementation of DHTs [47]	• N/Ae
Ability to observe results (n_f=3; n_b=0)	• Positive previous experiences using DHTs [41,47] • The ability of DHTs to increase administrative and health care efficiency [36] • The ability of DHTs to reduce medication errors [36]	• N/A
Perspectives of Patients (n=5)		

XSL•FO
RenderX

Themes and categories	Codes for facilitators	Codes for barriers
Patient centeredness (n_f=1; n_b=1)	• Using technology gives a superior feeling [38] • Repeated motivational or praising words increase patients' usage of DHTs [38]	• The feeling of the distance between the patient and health care providers during remote interaction [38] • The feeling of being forced to use DHTs [38]
Addresses patient barriers (n_f=1; n_b=1)	• Involvement of patients' children could improve the efficiency and effectiveness of DHTs [38]	• Patients' resistance to change practice [46]
Seamlessness of transition between program elements (n_f=1; n_b=0)	• Unification of hospitals in a single network and centralized management [47]	• N/A
Service and access (n_f=1; n_b=0)	• Easy access to online health information and applying the knowledge [38]	• N/A
Burden (complexity and cost; n_f=2; n_b=0)	• The ability of DHTs to reduce stress among patients [40] • DHTs are cost-effective [38]	• N/A
Feedback of results (n_f=2; n_b=0)	• DHTs are easy and convenient to use [38,39] • DHTs are able to improve the comfort level of patients [39]	• N/A
External Environment (n=3)		
Competition (n_f=0; n_b=2)	• N/A	• Competition for resources between information technology equipment and other health care equipment [45,47]
Regulatory environment (n_f=1; n_b=2)	• Good communication process between the regulatory authority and the hospitals [47]	• Political and economic instabilities [45,47]
Community resources (n_f=1; n_b=0)	• Good internet access among users of DHTs [44]	• N/A
Implementation of Sustainable Infrastructure (n=5)		
Adopter training and support (n_f=1; n_b=0)	• Familiarity with handling DHT tools [40]	• N/A
Relationship and communication with adopters (n_f=0; n_b=1)	• N/A	• The communication gap between supervisors and supervisees [45]
Adaptable protocols and procedures (n_f=1; n_b=0)	• Remote health service can be carried out easily among clinicians [37]	• N/A
Facilitation of sharing of best practices (n_f=1; n_b=0)	• Experience sharing between advanced and beginner hospitals [47]	• N/A
Plan for sustainability (n_f=1; n_b=1)	• Health care facilities prioritize the implementation and sustainability of DHTs [47]	• The end of the COVID-19 pandemic reduced the demand for DHTs [46]
Characteristics of the Health Care Organization (n=11)		

XSL•FO
RenderX

Themes and categories	Codes for facilitators	Codes for barriers
Organizational health and culture (n_f=3; n_b=1)	• The availability of fair-to-good organizational support [42] • Commitment from the top management or managers [47] • Top management mandates the adoption of DHTs [47] • Promote experience and knowledge sharing among hospital networks [47] • A workplace culture that is receptive to change [47] • Understand the policies made by the government [41]	• Frequent changes in the hospital management team [47] • Lack of medium- and long-term information technology policy planning [47]
Management support and communication (n_f=2; n_b=1)	• Presence of supervision at the workplace [42] • The availability of fair-to-good technical support [42] • Provision of support to groups with little knowledge of technology [47]	• Lack of planning and implementation of system design [47]
Shared goals and cooperation (n_f=2; n_b=1)	• The information technology department of the hospital is directly linked to local information technology sectors [47] • Create a multidisciplinary management committee [47] • Partnership with universities to develop DHTs [45]	• Health care providers' resistance to change practice [46]
Clinical leadership (n_f=1; n_b=1)	• High level of decision-making autonomy in the workplace [42]	• Weak leadership and poor commitment at the hospital level [45]
Systems and training (n_f=3; n_b=5)	• The workplace has sufficient facilities [37] • The availability of continuous on-the-job training [37,43,47] • Regular system upgrades [43] • The availability of prompt technical assistance from desk support [43]	• Lack of training in advanced technologies [36,40] • Delays in system development, implementation, and updates [43,47] • Frequent power blackouts [43] • Poor electrical power and backup facilities [36] • Lack of skill among health care providers in accessing and using DHTs [44]
Data and decision support (n_f=2; n_b=2)	• The availability of 24-hour technical support services at the workplace [37] • Centralization of system development and technical support [47]	• Lack of contingency plan for DHTs [47] • Delayed information technology support [43]
Staffing and incentives (n_f=3; n_b=3)	• Workplace has sufficient trained personnel in DHTs [37] • Improved financing of human and material resources [47] • Sufficient human resource allocation [41]	• Lack of incentives for health care providers to use DHTs [38] • High staff turnover rate, work overload, and reduced staffing [43,47] • Health care providers who are old aged [43]
Expectation of sustainability(n_f=0; n_b=3)	• N/A	• Health care providers do not have sufficient time and energy to participate and offer timely feedback while using DHTs [38,44] • Unrealistic expectations toward DHTs [47]
Characteristics of patients (n=4)		

XSL•FO
RenderX

Themes and categories	Codes for facilitators	Codes for barriers
Demographics (n_f=0; n_b=2)	• N/A	• Older patients contemplated using and have low confidence in DHTs [38] • Personal inertia and the conservative mentality of the older generation [38] • Costs of accessing the internet beyond affordability for patients from lower socioeconomic status [39] • Poor cellular network connectivity and internet access in some geographical areas [39]
Knowledge and beliefs (n_f=2; n_b=4)	• Confidence to learn, engage, and sustainably participate in DHTs [38] • High eHealth and computer literacy among patients [38,44] • High information-searching skills among patients [44]	• Patients have poor digital literacy and skills in DHTs [36,38,39] • Patients have skepticism about the accuracy of remote diagnosis [38] • Social prejudice against DHTs [38]

[a]DHT: digital health technology.

[b]n: number of studies in general.

[c]n_f: number of studies that reported on facilitators.

[d]n_b: number of studies that reported on barriers.

[e]N/A: not applicable.

Specifically, the theme of Aspects of the Health Care System included 23 facilitators and 13 barriers (n=11 studies). Among the prominent facilitators were the ability of DHTs to prevent cross-infection (n=2) and positive previous experiences using DHTs (n=2). By contrast, inadequate awareness regarding DHTs (n=2) and concerns regarding accuracy in disease diagnosis and treatment (n=2) were the leading barriers highlighted.

Regarding the theme of Perspectives of Patients, 9 facilitators and 3 barriers were reported (n=5 studies). Among these, the convenience and easiness of using DHTs (n=2) were noted as the primary facilitator. However, the feeling of distance between the patient and health care providers during remote interaction (n=1), the feeling of being forced to use DHTs (n=1), and patients' resistance to change practice could pose challenges to the implementation of DHTs in hospital settings within LMICs.

The successfulness of implementing DHTs also relies on the theme of External Environment (n=3 studies). Within this context, 2 facilitators are aiding the implementation, which include a good communication process between the regulatory authority and the hospitals (n=1) and good internet access among users of DHTs (n=1). Among the 2 barriers identified in this context, competition for resources between IT equipment and other health care equipment (n=2) stands out as the more prevalent barrier.

The theme of Implementation of Sustainable Infrastructure comprises 4 facilitators and 2 barriers (n=5 studies). Notably, familiarity with handling DHTs tools (n=1), the ease of conducting remote health service among clinicians (n=1), and experience sharing between advanced and beginner hospitals (n=1) emerge as facilitators for DHT implementation. Conversely, barriers in this theme include the communication gap between stakeholders (n=1) and the end of the COVID-19 pandemic reducing the demand for DHTs (n=1).

The Characteristics of Health Care Organization theme consists of 22 facilitators and 17 barriers to DHT implementation (n=11 studies). Previous studies have highlighted the availability of continuous on-the-job training (n=3) as the most frequent facilitator of DHT implementation in the health care organization. Conversely, barriers such as lack of trainings in advanced technologies (n=2); delays in system development, implementation, and updates (n=2); high staff turnover rate, work overload, and reduced staffing (n=2); as well as insufficient time and energy among health care providers to participate and offer timely feedback while using DHTs (n=2) prevent the implementation of DHTs.

The Characteristics of Patients was the final theme that could influence the successfulness of DHT implementation in LMICs' hospital settings. It comprised 3 facilitators and 7 barriers (n=4 studies). Notably, the most frequently mentioned facilitator in this theme is the high eHealth and computer literacy among patients (n=2). Not surprisingly, the exact opposite, that is, patients with poor digital literacy and skills in DHTs (n=3), is reported as the commonest barrier.

The relationship between codes, categories, and themes of facilitators and barriers to DHT implementation is summarized in Figure 2, in which the codes and categories are arranged from the most frequent (top) to the least frequent (bottom).

Figure 3 illustrates the network analysis of the facilitators of DHT implementation. Of note, facilitators such as "Perspectives of Patients" and "Aspect of the Health Care System" are central to the network. This suggests their broad impact across multiple domains, making them key intervention points for DHT implementation. Although no obvious clusters were observed, the interconnectedness suggests that addressing certain facilitators (eg, "Perspectives of Patients," "Aspect of the Health Care System," and "Characteristics of the Health Care Organization") can have ripple effects on others. Figure 4 illustrates the network analysis of the barriers to DHT

XSL•FO
RenderX

implementation. It was noted that some barriers are more central, such as the "Characteristics of Health Care Organizations," "Aspect of the Health Care System," "Implementation of Sustainable Infrastructure," and "Characteristics of Patients" barriers, highlighting their pivotal roles in the network. These central barriers act as key influencers and bridge multiple barriers. Meanwhile, the "External Environment" and "Perspectives of Patients" barriers are more toward the periphery and, hence, have less impact on the network. A cluster was observable in the network analysis, involving the "Characteristics of Health Care Organization" and "Aspect of the Health Care System" barriers. Many barriers are interconnected, which indicate that addressing one (eg, "Characteristics of Health Care Organizations") may alleviate others. However, it should be noted that some nodes, such as "Readiness" and "Usability and Adaptability," can be both facilitators of and barriers to DHT implementation, and hence, can appear in both network analyses.

Since the onset of the COVID-19 pandemic, the most commonly utilized DHTs in LMICs' hospitals were telemedicine, followed by health information systems, electronic medical records, and mHealth, as depicted in Figure 5.

Figure 2. Summary of the codes, categories, and themes of facilitators and barriers of digital health technology (DHT) implementation. IT: information technology; LAN: local area network.

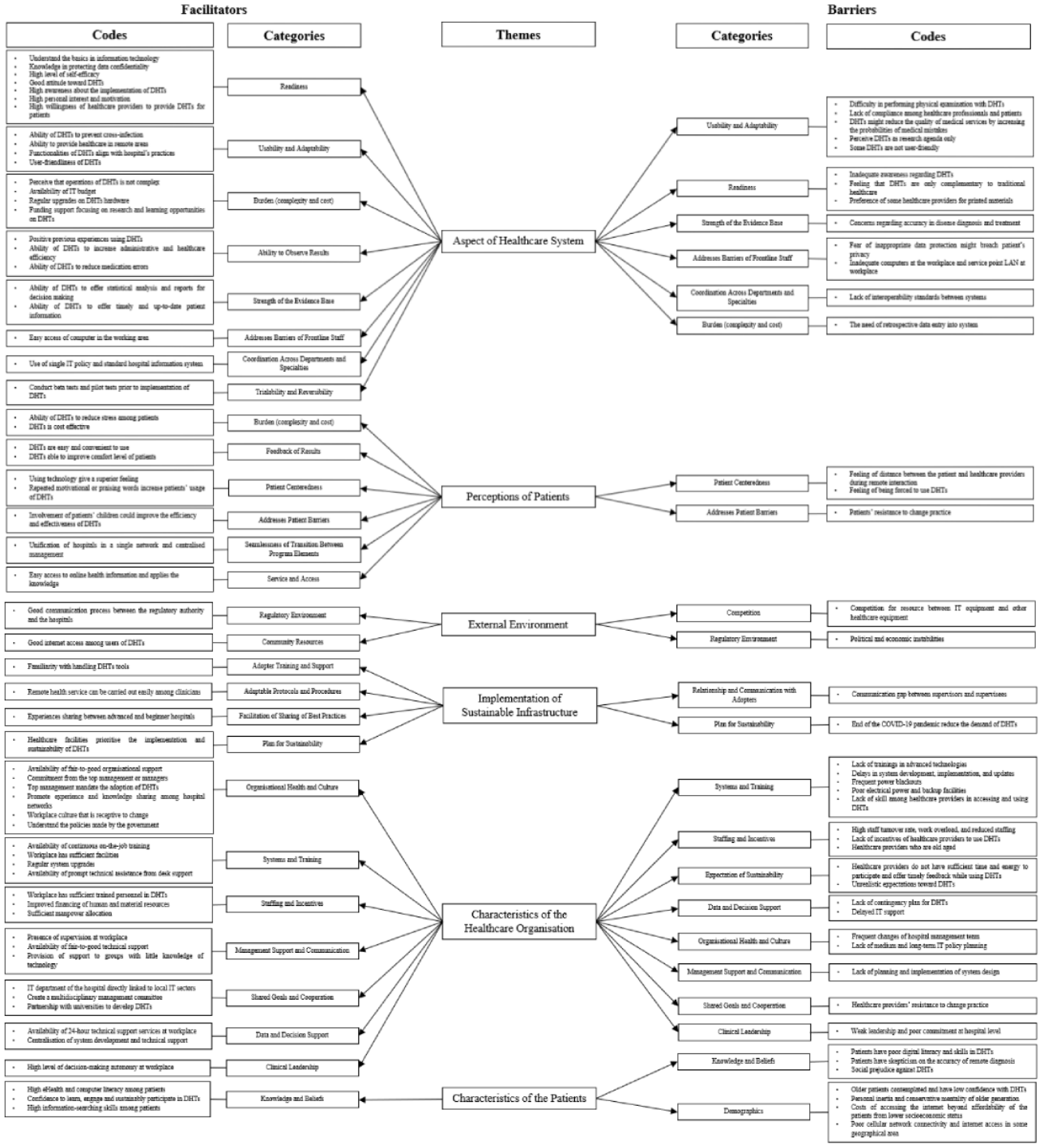

XSL•FO
RenderX

Figure 3. Network analysis to illustrate the relationship between the codes of facilitators of digital health technologies (DHTs) implementation in lower- and middle-income countries' (LMICs) hospital settings since the onset of the COVID-19 pandemic.

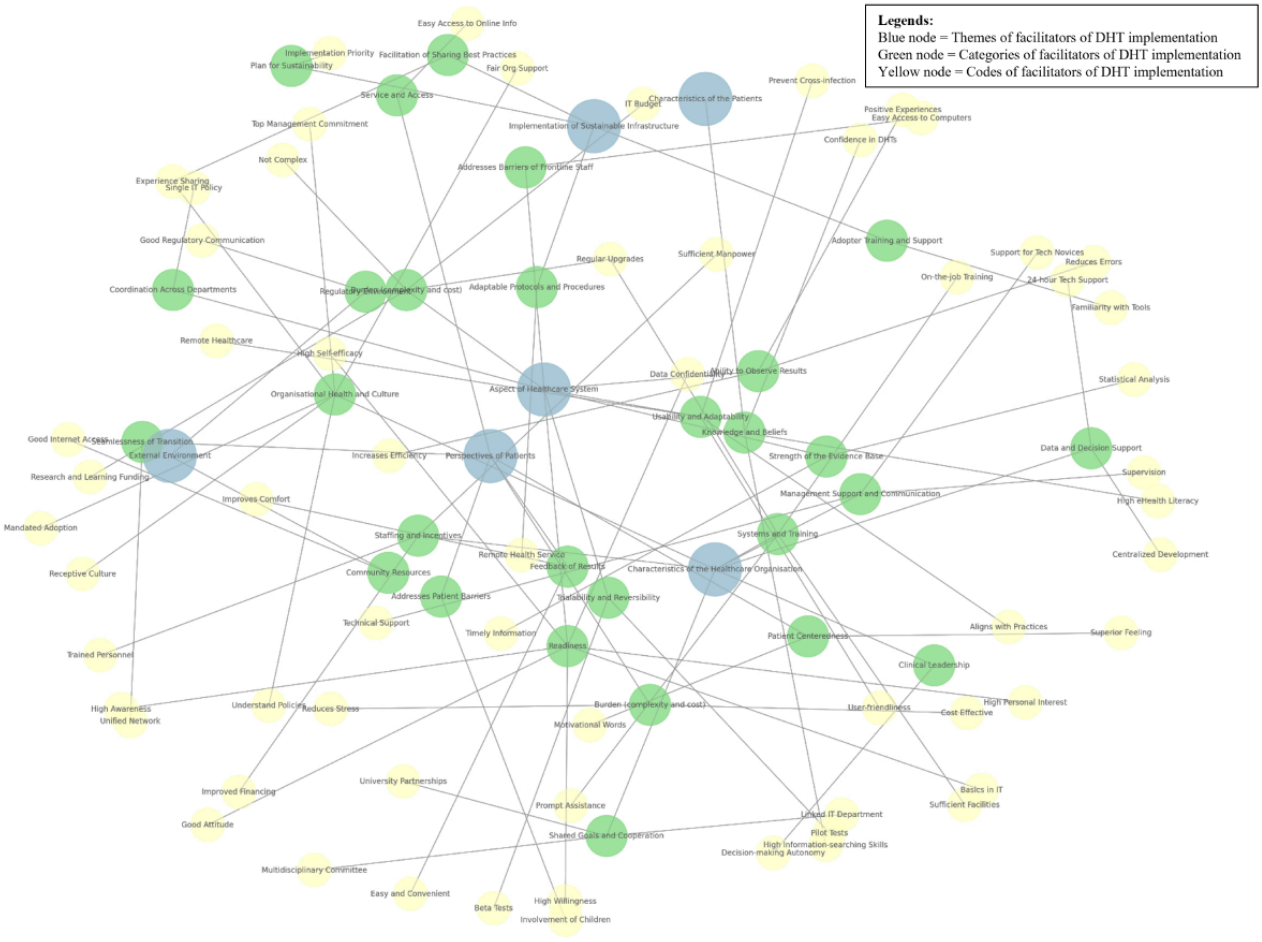

XSL•FO
RenderX

Figure 4. Network analysis to illustrate the relationship between the codes of barriers of digital health technologies (DHTs) implementation in lower- and middle-income countries' (LMICs') hospital settings in the post-COVID-19 era. IT: information technology; LAN: local area network.

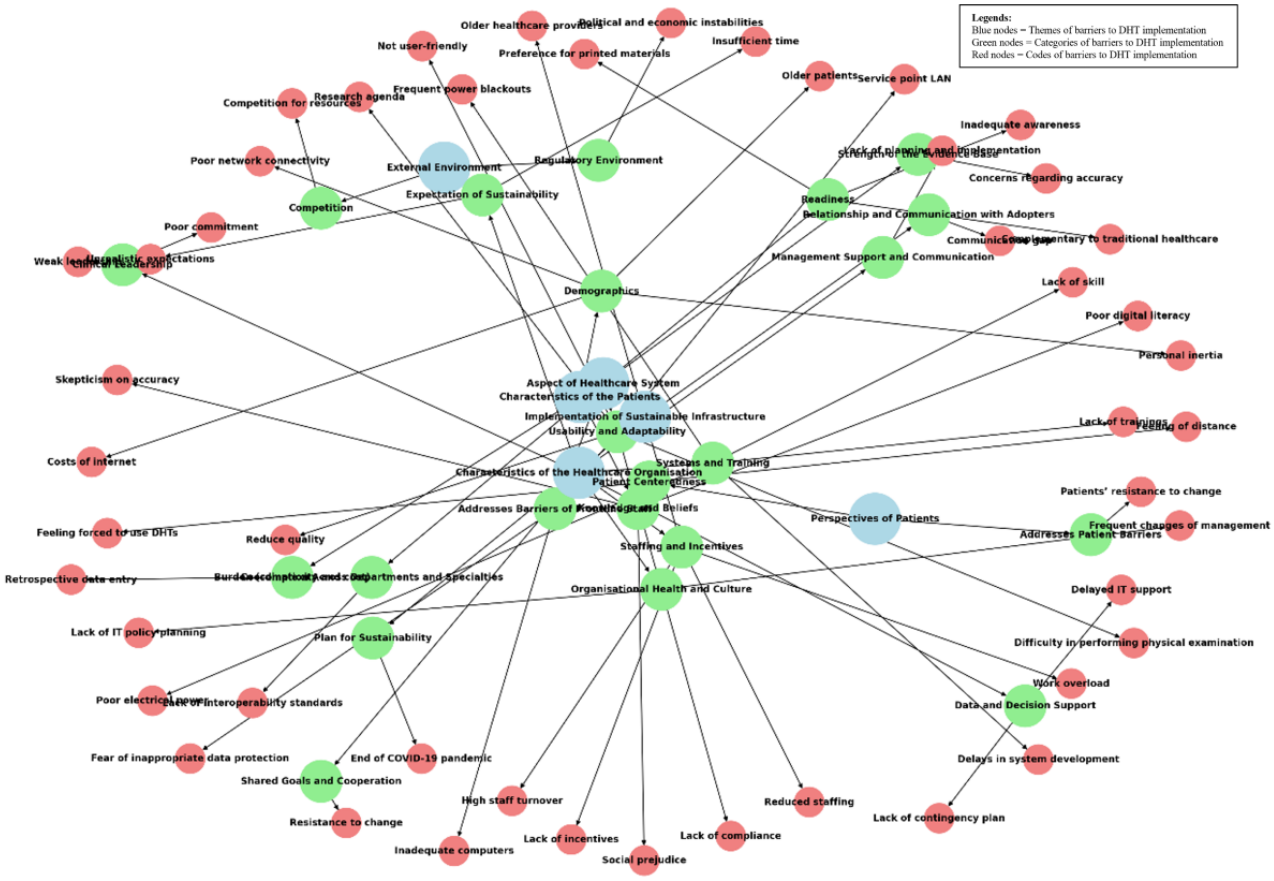

Figure 5. Digital health technologies (DHTs) in the lower- and middle-income countries (LMICs) since the onset of the COVID-19 pandemic.

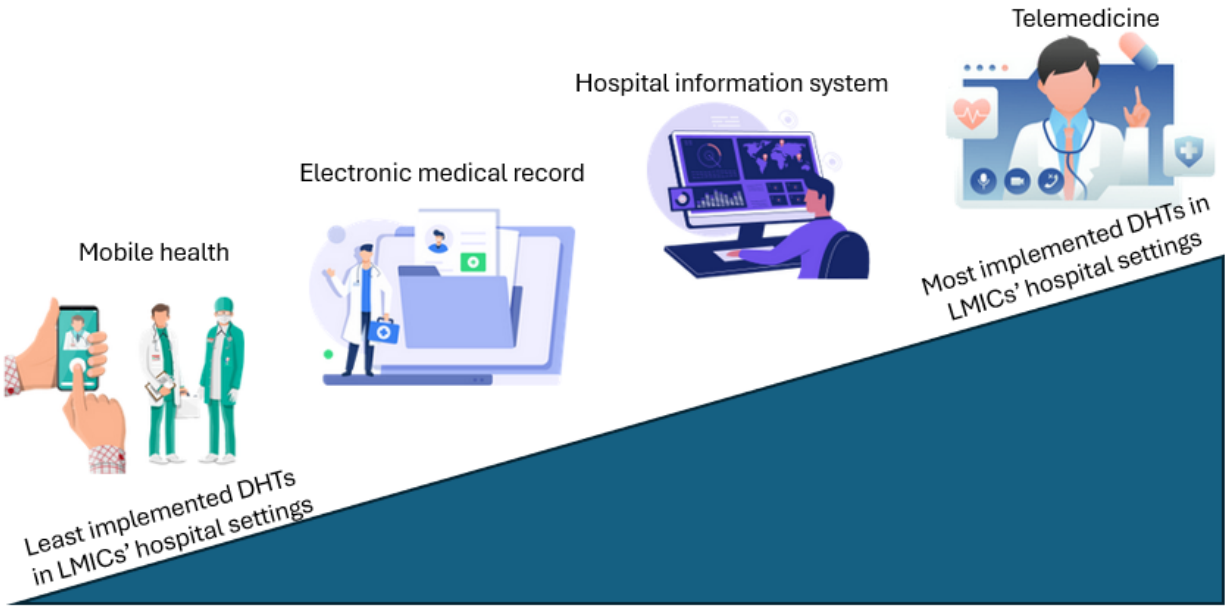

XSL•FO
RenderX

Discussion

Facilitators of and Barriers to DHT Implementation in LMICs

The most common facilitators of DHT implementation include the convenience and easiness of using DHTs, the ability of DHTs to prevent cross-infection, positive previous experiences using DHTs, the availability of continuous on-the-job training, and high eHealth and computer literacy among patients. By contrast, we identified that the most frequently reported barriers are inadequate awareness regarding DHTs; concerns regarding accuracy in disease diagnosis and treatment; competition for resources between IT equipment and other health care equipment; lack of trainings in advanced technologies; delays in system development, implementation, and updates; high staff turnover rate, work overload, and reduced staffing; insufficient time and energy among health care providers to use DHTs; and patients with poor digital literacy and skills in DHTs.

To our knowledge, this scoping review is the first to examine the facilitators and barriers to implementing DHTs in hospitals within LMICs since the onset of the COVID-19 pandemic. It is important to note the scarcity of published literature on this topic in LMICs. Even existing reviews often focused on studies from high-income countries or concentrated on specific diseases [27,48,49]. For example, a systematic review by Whitelaw et al [27], comprising 29 studies, investigated DHT implementation in cardiovascular diseases predominantly in high-income countries. As expected, the facilitators and barriers identified in this review significantly differed from our findings. Whitelaw et al highlighted facilitators such as enhanced communication with clinicians, personalized DHT components, user-friendly interfaces, institutional and organizational support, increased efficiency, and perceived usefulness of DHTs. The reported barriers were complex technology, technological apprehension, increased workload, unreliable technology, and lack of integration with electronic medical records. These discrepancies may stem from various factors such as differences in health care infrastructure [50], socioeconomic contexts [51], health care financing [52], as well as costs of DHT development and implementation [53] between high-income and LMIC settings. Additionally, the focus on specific diseases in previous literature might not fully capture the diverse facilitators and barriers relevant to DHT implementation across various health care contexts.

DHTs That Have Been Implemented

To our surprise, telemedicine was the most commonly reported technology implemented in hospital settings since the onset of the COVID-19 pandemic, exceeding other DHTs such as hospital information systems, electronic medical records, and mHealth. This can be understood by the fact that telemedicine is well-suited for health care providers and patients who are self-isolating, as it effectively reduces the risk of COVID-19 transmission. It eliminates the need for direct physical contact, ensures ongoing care for the community, and ultimately lowers the rates of illness and death during the COVID-19 pandemic [54].

Classification Framework for DHT Implementation in LMICs

Despite the emergence of various innovative technologies such as wearable technologies, metaverse, Internet of Medical Things, blockchain in health care, big data analytics, artificial intelligence, machine learning, software as medical devices, augmented reality, and virtual reality, the implementation of DHTs in hospitals within LMICs remains rudimentary compared with high-income countries. The current utilization of DHTs in LMICs, such as telemedicine, hospital information systems, electronic medical records, and mHealth, has informed the key components of a new classification framework that emphasizes LMICs. Based on recent literature [13,15,17-23,55] and the specific facilitators and barriers to utilizing these DHTs in LMICs, we propose a classification framework of DHTs, as illustrated in Figure 6.

XSL•FO
RenderX

Figure 6. Classification framework of digital health technologies (DHTs). LMIC: lower- and middle-income country.

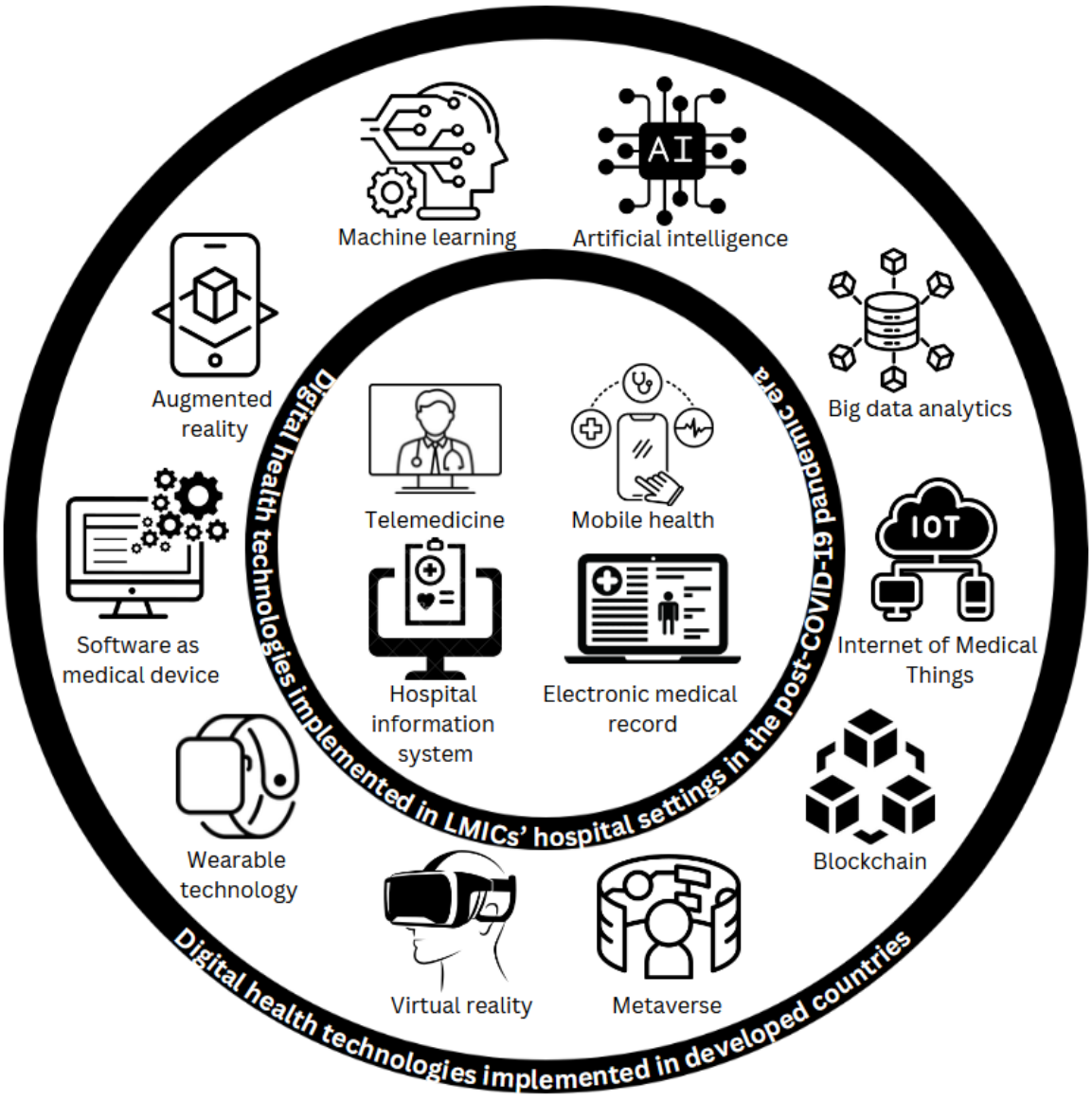

This classification framework is justified by the fact that simpler DHTs, such as telemedicine, hospital information systems, electronic medical records, and mHealth, are easier to implement in LMICs due to lower costs and infrastructure requirements, adaptability to limited resources, and minimal training needs. These technologies rely mainly on basic internet connectivity, mobile devices, and computers, making them more accessible and feasible given the broader availability of mobile networks in LMICs. They also integrate easily into existing health care workflows, requiring only short training sessions and basic digital literacy. Furthermore, because LMICs prioritize addressing immediate health care needs such as infectious diseases, these accessible technologies receive more support and funding. By contrast, advanced technologies such as artificial intelligence, virtual reality, and blockchain demand specialized infrastructure, technical expertise, and substantial financial resources, which are often scarce in LMICs.

This proposed classification framework also consists of other DHTs that were implemented in high-income countries. This, hopefully, will aid health care stakeholders in LMICs to delineate the scope of DHTs in LMICs and help them to prioritize the right DHTs that could be implemented in the scarcity of resources. Consequently, policy makers can allocate health care funding (eg, the Ministry of Health can determine the DHT classes that are eligible for subsidies), and insurers can reimburse patients who utilized DHTs (ie, identifying which DHT classes can be covered by insurance policies) [56,57]. Moreover, clear and effective communication among the public, health care providers, and technology experts can be facilitated [57,58]. Legally, a defined classification framework enables authorities to regulate and govern DHT usage [58,59]. Finally, a clearly outlined classification framework for DHTs may streamline implementation, especially in resource-scarce environments, as stakeholders can prioritize which DHT classes should be implemented [56].

Strengths and Limitations

This scoping review possesses several notable strengths. Initially, we comprehensively gathered data from a diverse array of databases, including gray literature sources such as Google

XSL•FO
RenderX

Scholar. This deliberate choice was made to ensure the inclusivity and robustness of the studies included. Additionally, during the planning phase of the review, we actively engaged relevant stakeholders, including information technologists, seeking and incorporating their feedback into the review protocol. Furthermore, in structuring and categorizing the identified barriers and facilitators, we utilized a conceptual model developed by Feldstein and Glasgow [33], thereby enhancing the analytical rigor of the review. Moreover, our review offers a wealth of qualitative and quantitative insights into the implementation of DHTs, providing comprehensive guidance for stakeholders interested in implementing DHTs within their respective settings.

Despite its strengths, our review does have certain limitations. Specifically, in focusing on identifying barriers and facilitators of DHT implementation in hospital settings within LMICs, we restricted our inclusion criteria to studies conducted solely within hospitals. Consequently, many studies providing insights from primary care and community perspectives were excluded. Additionally, studies that did not explicitly address facilitators and barriers were omitted, potentially overlooking valuable information on other beneficial DHTs for patient care. Furthermore, as our primary aim was to offer an overview of existing literature on DHTs, we did not apply any quality control measures to the included studies. The variation in analytical methods, interview questions, and study populations across the included papers presents a potential inconsistency in the findings and makes quality assessment difficult. Additionally, with 128 LMICs at the time of the literature search, listing each country by name in the search string was impractical, which could result in some LMICs not being captured in the search results.

Implications and Recommendations

Understanding the facilitators and barriers to DHTs in hospital settings of LMICs has profound implications for both clinical practice and research. Identifying facilitators of DHT implementation enables the effective integration of DHTs in hospitals, leading to improved diagnostic accuracy, better patient monitoring, improved patient records management, and the availability of remote health care services. Conversely, recognizing the barriers to DHT implementation can prompt various stakeholders in LMICs to devise strategies that address issues in terms of financial resources, technological infrastructure, and knowledge among health care providers. In terms of research, this review is hoped to promote international collaborations, governmental support, and investment in digital infrastructure, which in turn boost the capacity for data collection, analysis, and sharing. These advancements can enhance research capabilities, allowing for more comprehensive epidemiological studies, real-time disease monitoring, and the development of context-specific health interventions.

To maximize the benefits of DHTs in hospital settings in LMICs, several recommendations can be made to leverage facilitators and mitigate barriers. First, the convenience and ease of using DHTs should be emphasized through user-friendly interfaces and intuitive designs, ensuring that health care providers can quickly and efficiently adopt these technologies in their daily routines. Second, the role of DHTs in preventing cross-infection should be highlighted in training programs and awareness campaigns, particularly in regions where infectious diseases are prevalent. Positive previous experiences with DHTs should be shared widely to build trust and demonstrate the tangible benefits of these technologies. Continuous on-the-job training must be provided to keep health care workers up-to-date with the latest advancements and best practices, while also enhancing their technical proficiency. Promoting high eHealth and computer literacy among patients through community outreach and education programs will further support the effective use of DHTs.

Addressing barriers requires a multifaceted approach. To tackle inadequate awareness and concerns regarding the accuracy of DHTs, it is crucial to implement comprehensive educational initiatives that highlight the reliability and clinical efficacy of these technologies. Clear, evidence-based information should be disseminated to health care providers and patients to build confidence in DHTs. Competition for resources between IT equipment and other health care tools can be mitigated by advocating for balanced budget allocations and seeking external funding or partnerships to supplement resources. Training programs in advanced technologies must be expanded to equip health care providers with the necessary skills to use DHTs effectively. To address delays in system development and implementation, streamlined processes and timelines should be established, with dedicated project management teams overseeing these initiatives. High staff turnover rates and work overload can be managed by improving working conditions, offering competitive salaries, and providing career development opportunities. Finally, enhancing digital literacy among patients through targeted educational programs will ensure that they can engage with DHTs effectively, ultimately leading to better health outcomes. By implementing these recommendations, LMICs can overcome barriers and fully harness the potential of DHTs to improve clinical practice and advance health care research.

Conclusions

The scoping review on DHT implementation in LMICs since the onset of the COVID-19 pandemic underscores significant policy, clinical, and research implications. Policy makers can utilize the insights to craft targeted strategies for DHT adoption, while the developed classification framework aids in prioritizing DHTs. In clinical settings, understanding DHT types, facilitators, and barriers enhances decision-making for improved patient care and resource optimization. Additionally, investing in capacity building and training programs for health care professionals is crucial, with a focus on enhancing digital literacy and technical skills. Moreover, identifying research priorities and aligning funding with key challenges such as improving digital literacy and addressing concerns about DHT accuracy will drive advancements in DHT implementation and health care delivery in LMICs.

In the postpandemic era, telemedicine has been widely used in hospital settings in LMICs, among other DHTs. Our findings also reveal numerous facilitators and barriers to DHT implementation in LMIC hospital settings. These factors can be organized into 6 themes, namely, (1) Aspects of the Health Care System; (2) Perspectives of Patients; (3) External

XSL•FO
RenderX

Environment; (4) Implementation of Sustainable Infrastructure; (5) Characteristics of Health Care Organization; and (6) Characteristics of Patients.

Acknowledgments

We acknowledge the contributions of the information technologist at the library of Universiti Putra Malaysia for the provision of technical and intellectual assistance during the literature search.

Data Availability

All data retrieved and analyzed are included in this scoping review.

Authors' Contributions

SQY and BHC developed the concept for the scoping review. SQY and DT were involved in designing the research methodology. NIHA and SQY carried out an extensive literature review to identify relevant frameworks and methodological approaches for the scoping review. SQY prepared the initial draft of the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 16 KB - jmir_v27i1e63482_app1.docx]

Multimedia Appendix 2
The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 115 KB - jmir_v27i1e63482_app2.pdf]

References

1. World Health Organization (WHO). Tracking universal health coverage: 2023 global monitoring report. WHO. Geneva, Switzerland: WHO; 2023 Sep 18. URL: https://www.who.int/publications/i/item/9789240080379 [accessed 2024-02-10]
2. Shah SA, Safian N, Ahmad S, Nurumal SR, Mohammad Z, Mansor J, et al. Unmet healthcare needs among elderly Malaysians. JMDH 2021 Oct;Volume 14:2931-2940. [doi: 10.2147/jmdh.s326209]
3. Njagi P, Arsenijevic J, Groot W. Cost-related unmet need for healthcare services in Kenya. BMC Health Serv Res 2020 Apr 17;20(1):322 [FREE Full text] [doi: 10.1186/s12913-020-05189-3] [Medline: 32303244]
4. Sohn M, Che X, Park HJ. Unmet healthcare needs, catastrophic health expenditure, and health in South Korea's universal healthcare system: progression towards improving equity by NHI type and income level. Healthcare (Basel) 2020 Oct 16;8(4):408 [FREE Full text] [doi: 10.3390/healthcare8040408] [Medline: 33081357]
5. Baek S, Choi EH, Lee J. Unmet healthcare needs of children in vulnerable families in South Korea: finding from the Community Child Center Child Panel Survey. Int J Environ Res Public Health 2020 Nov 07;17(21):8241 [FREE Full text] [doi: 10.3390/ijerph17218241] [Medline: 33171882]
6. Kim J, You M, Shon C. Impact of the COVID-19 pandemic on unmet healthcare needs in Seoul, South Korea: a cross-sectional study. BMJ Open 2021 Aug 26;11(8):e045845 [FREE Full text] [doi: 10.1136/bmjopen-2020-045845] [Medline: 34446482]
7. Harris RB, Cormack DM, Stanley J. Experience of racism and associations with unmet need and healthcare satisfaction: the 2011/12 adult New Zealand Health Survey. Aust N Z J Public Health 2019 Feb;43(1):75-80 [FREE Full text] [doi: 10.1111/1753-6405.12835] [Medline: 30296819]
8. Rosenthal A, Waitzberg R. The challenges brought by the COVID-19 pandemic to health systems exposed pre-existing gaps. Health Policy Open 2023 Dec;4:100088. [doi: 10.1016/j.hpopen.2022.100088] [Medline: 36536931]
9. Mitgang EA, Blaya JA, Chopra M. Digital health in response to COVID-19 in low- and middle-income countries: opportunities and challenges. Glob Policy 2021 Jul;12(Suppl 6):107-109. [doi: 10.1111/1758-5899.12880] [Medline: 34230840]
10. Ronquillo Y, Meyers A, Korvek S. Digital Health. StatPearls. Treasure Island (FL): StatPearls Publishing; Jan 2025.
11. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747 [FREE Full text] [doi: 10.3390/ijerph19073747] [Medline: 35409431]

XSL•FO
RenderX

12. Roy J, Levy DR, Senathirajah Y. Defining telehealth for research, implementation, and equity. J Med Internet Res 2022 Apr 13;24(4):e35037 [FREE Full text] [doi: 10.2196/35037] [Medline: 35416778]
13. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
14. Ibrahim AA, Ahmad Zamzuri M'I, Ismail R, Ariffin AH, Ismail A, Muhamad Hasani MH, et al. The role of electronic medical records in improving health care quality: a quasi-experimental study. Medicine (Baltimore) 2022 Jul 29;101(30):e29627 [FREE Full text] [doi: 10.1097/MD.0000000000029627] [Medline: 35905245]
15. Belle A, Thiagarajan R, Soroushmehr SMR, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. Biomed Res Int 2015;2015:370194 [FREE Full text] [doi: 10.1155/2015/370194] [Medline: 26229957]
16. Dwivedi R, Mehrotra D, Chandra S. Potential of Internet of Medical Things (IoMT) applications in building a smart healthcare system: a systematic review. J Oral Biol Craniofac Res 2022 Mar;12(2):302-318 [FREE Full text] [doi: 10.1016/j.jobcr.2021.11.010] [Medline: 34926140]
17. Yoon HJ. Blockchain technology and healthcare. Healthc Inform Res 2019 Apr;25(2):59-60 [FREE Full text] [doi: 10.4258/hir.2019.25.2.59] [Medline: 31131139]
18. Petrigna L, Musumeci G. The metaverse: a new challenge for the healthcare system: a scoping review. J Funct Morphol Kinesiol 2022 Aug 30;7(3):63 [FREE Full text] [doi: 10.3390/jfmk7030063] [Medline: 36135421]
19. Yu J, Zhang J, Sengoku S. Innovation process and industrial system of US Food and Drug Administration–approved software as a medical device: review and content analysis. J Med Internet Res 2023 Nov 24;25:e47505 [FREE Full text] [doi: 10.2196/47505] [Medline: 37999948]
20. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]
21. Halbig A, Babu S, Gatter S, Latoschik M, Brukamp K, von Mammen S. Opportunities and challenges of virtual reality in healthcare – a domain experts inquiry. Front Virtual Real 2022 Mar 23;3:3. [doi: 10.3389/frvir.2022.837616]
22. Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: a powerful paradigm for scientific research. Innovation (Camb) 2021 Nov 28;2(4):100179 [FREE Full text] [doi: 10.1016/j.xinn.2021.100179] [Medline: 34877560]
23. Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2021 Mar;2(3):160 [FREE Full text] [doi: 10.1007/s42979-021-00592-x] [Medline: 33778771]
24. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. J Biomed Inform 2019 Dec;100:103311 [FREE Full text] [doi: 10.1016/j.jbi.2019.103311] [Medline: 31629922]
25. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021 Apr 22;23(4):e25759 [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
26. Iyamu I, Gómez-Ramírez O, Xu AX, Chang HJ, Watt S, Mckee G, et al. Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review. Digit Health 2022;8:20552076221102255. [doi: 10.1177/20552076221102255] [Medline: 35656283]
27. Whitelaw S, Pellegrini D, Mamas M, Cowie M, Van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar 04;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
28. Jacob KS. Public health in low- and middle-income countries and the clash of cultures. J Epidemiol Community Health 2009 Jul;63(7):509. [doi: 10.1136/jech.2009.086934] [Medline: 19525248]
29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
30. Yew SQ, Trivedi D, Adanan NIH, Chew BH. Facilitators and barriers of digital health technologies implementation in hospital settings in lower-income and middle-income countries since the COVID-19 pandemic: a scoping review protocol. BMJ Open 2024 Jan 31;14(1):e078508 [FREE Full text] [doi: 10.1136/bmjopen-2023-078508] [Medline: 38296272]
31. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
32. Jbi. JBI Manual for Evidence Synthesis. JBI Global Wiki. 2024. URL: https://jbi-global-wiki.refined.site/space/MANUAL [accessed 2024-08-04]
33. Feldstein AC, Glasgow RE. A Practical, Robust Implementation and Sustainability Model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf 2008 Apr;34(4):228-243. [doi: 10.1016/s1553-7250(08)34030-6] [Medline: 18468362]
34. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. International Journal of Qualitative Methods 2023 Nov 08;22:22. [doi: 10.1177/16094069231205789]
35. PRISMA Flow Diagram. PRISMA Statement. 2020. URL: https://www.prisma-statement.org/prisma-2020-flow-diagram [accessed 2024-08-04]

XSL•FO
RenderX

36. Abdulai M, Aldheleai H, Bokhari M. Health care professionals' use of health information systems (HIS) in Indian hospitals. IJCA 2021 Apr 19;174(28):5-12. [doi: 10.5120/ijca2021921185]
37. Alboraie M, Abdalgaber M, Youssef N, Moaz I, Abdeen N, Abosheaishaa H, et al. Healthcare providers' perspective about the use of telemedicine in Egypt: a national survey. Int J Telemed Appl 2022 Mar 12;2022:3811068-3811068 [FREE Full text] [doi: 10.1155/2022/3811068] [Medline: 35313723]
38. Jiang Y, Sun P, Chen Z, Guo J, Wang S, Liu F, et al. Patients' and healthcare providers' perceptions and experiences of telehealth use and online health information use in chronic disease management for older patients with chronic obstructive pulmonary disease: a qualitative study. BMC Geriatr 2022 Jan 03;22(1):9 [FREE Full text] [doi: 10.1186/s12877-021-02702-z] [Medline: 34979967]
39. Ning Y, Dong Z, Jia Z, Zhao W, Ding Y, Wang Q, et al. Development of mobile health-based interventions to promote physical activity in patients with head and neck cancer: a qualitative study. Front Public Health 2023;11:1260804. [doi: 10.3389/fpubh.2023.1260804] [Medline: 38074711]
40. Shardha H, Kumar G, Sagar N, Kumar R, Qazi M, Munir S, et al. Perceptions of telemedicine among healthcare professionals in rural tertiary care hospitals of rural Sindh, Pakistan: a qualitative study. Ann Med Surg (Lond) 2024 Feb;86(2):726-733. [doi: 10.1097/MS9.0000000000001688] [Medline: 38333274]
41. Yu-Tong T, Yan Z, Zhen L, Bing X, Qing-Yun C. Telehealth readiness and its influencing factors among Chinese clinical nurses: a cross-sectional study. Nurse Educ Pract 2022 Jan;58:103278 [FREE Full text] [doi: 10.1016/j.nepr.2021.103278] [Medline: 34954659]
42. Mekuria S, Adem HA, Ayele BH, Musa I, Enyew DB. Routine health information system utilization and associated factors among health professionals in public health facilities in Dire Dawa, eastern Ethiopia: a cross-sectional study. Digit Health 2023;9:20552076231203914 [FREE Full text] [doi: 10.1177/20552076231203914] [Medline: 37808236]
43. Ngugi PN, Were MC, Babic A. Users' perception on factors contributing to electronic medical records systems use: a focus group discussion study in healthcare facilities setting in Kenya. BMC Med Inform Decis Mak 2021 Dec 26;21(1):362. [doi: 10.1186/s12911-021-01737-x] [Medline: 34955098]
44. Tesfa GA, Kalayou MH, Zemene W. Electronic health-information resource utilization and its associated factors among health professionals in Amhara regional state teaching hospitals, Ethiopia. Adv Med Educ Pract 2021;12:195-202. [doi: 10.2147/AMEP.S289212] [Medline: 33688292]
45. Tilahun B, Gashu KD, Mekonnen ZA, Endehabtu BF, Asressie M, Minyihun A, et al. Strengthening the national health information system through a capacity-building and mentorship partnership (CBMP) programme: a health system and university partnership initiative in Ethiopia. Health Res Policy Syst 2021 Dec 09;19(1):141 [FREE Full text] [doi: 10.1186/s12961-021-00787-x] [Medline: 34886865]
46. Baradwan S, Al-Hanawi M. Perceived Knowledge, Attitudes, and Barriers Toward the Adoption of Telemedicine Services in the Kingdom of Saudi Arabia: Cross-Sectional Study. JMIR formative research. 2023 :7. [doi: 10.2196/46446]
47. Mussi CC, Luz R, Damázio DDR, Santos EMD, Sun V, Porto BSDS, et al. The large-scale implementation of a health information system in Brazilian university hospitals: process and outcomes. Int J Environ Res Public Health 2023 Oct 25;20(21):6971. [doi: 10.3390/ijerph20216971] [Medline: 37947529]
48. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, Martinez EZ, Zucoloto ML, Østengaard L, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161-128 [FREE Full text] [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
49. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024 Feb 26;24(1):243 [FREE Full text] [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
50. Mumtaz H, Riaz MH, Wajid H, Saqib M, Zeeshan MH, Khan SE, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023 Sep 28;5:1203945 [FREE Full text] [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
51. Chidambaram S, Jain B, Jain U, Mwavu R, Baru R, Thomas B, et al. An introduction to digital determinants of health. PLOS Digit Health 2024 Jan;3(1):e0000346 [FREE Full text] [doi: 10.1371/journal.pdig.0000346] [Medline: 38175828]
52. Sambo L, Simões J, do Rosario O, Martins M. In: Firth JD, Conlon CP, Cox TM, Firth J, Conlon C, Cox T, editors. Financing Health Care In Low-Income Developing Countries: A Challenge For Equity In Health. Oxford, UK: Oxford University Press; Jan 2020:168-176.
53. Sureshkumar K, Bindu M, John S, Kamarul IM. Editorialvidence on low-cost technologies for neurological rehabilitation in low and middle-income countries. Frontiers in Neurology. 2023 :14.
54. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193. [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
55. Huang C, Wang J, Wang S, Zhang Y. Internet of medical things: A systematic review. Neurocomputing. 2023 :557. [doi: 10.1016/j.neucom.2023.126719]
56. National Institute for Health and Care Excellence (NICE). Evidence standards framework for digital health technologies. NICE. 2019 Mar. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2025-02-11]

XSL•FO
RenderX

57. Functional classification, according to their intended use, of digital solutions used in the context of medical and paramedical care. Haute Autorité de Santé. 2021. URL: https://www.has-sante.fr/upload/docs/application/pdf/2021-03/dm_eval_270_proposition_classification.pdf [accessed 2024-08-04]
58. World Health Organization (WHO). Classification of digital health interventions v1. WHO. 2018. URL: https://iris.who.int/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf?sequence=1 [accessed 2024-08-04]
59. Food and Drug Administration (FDA). Digital Health Center of Excellence. FDA. 2024. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence [accessed 2024-08-04]

Abbreviations

DHT: digital health technology
GNI: gross national income
IT: information technology
LMIC: lower- and middle-income country
mHealth: mobile health
PCC: Population-Concept-Context
PRISM: Practical, Robust Implementation and Sustainability Model
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 21.06.24; peer-reviewed by D Linde, E Baker; comments to author 29.10.24; revised version received 01.11.24; accepted 09.12.24; published 06.03.25.

Please cite as:
Yew SQ, Trivedi D, Adanan NIH, Chew BH
Facilitators and Barriers to the Implementation of Digital Health Technologies in Hospital Settings in Lower- and Middle-Income Countries Since the Onset of the COVID-19 Pandemic: Scoping Review
J Med Internet Res 2025;27:e63482
URL: https://www.jmir.org/2025/1/e63482
doi:10.2196/63482
PMID:40053793

©Sheng Qian Yew, Daksha Trivedi, Nurul Iman Hafizah Adanan, Boon How Chew. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Diagnostic Performance of Artificial Intelligence–Based Methods for Tuberculosis Detection: Systematic Review

Seng Hansun[1,2], MCS; Ahmadreza Argha[3,4,5], PhD; Ivan Bakhshayeshi[3,6], MRES; Arya Wicaksana[7], MEngSc; Hamid Alinejad-Rokny[4,5,6], PhD; Greg J Fox[8], PhD; Siaw-Teng Liaw[9], PhD; Branko G Celler[10], PhD; Guy B Marks[1,2,11], PhD

[1]School of Clinical Medicine, South West Sydney, UNSW Medicine & Health, UNSW Sydney, Sydney, Australia
[2]Woolcock Vietnam Research Group, Woolcock Institute of Medical Research, Sydney, Australia
[3]Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, Australia
[4]Tyree Institute of Health Engineering, UNSW Sydney, Sydney, Australia
[5]Ageing Future Institute, UNSW Sydney, Sydney, Australia
[6]BioMedical Machine Learning Lab, Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, Australia
[7]Informatics Department, Universitas Multimedia Nusantara, Tangerang, Indonesia
[8]NHMRC Clinical Trials Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[9]School of Population Health and School of Clinical Medicine, UNSW Sydney, Sydney, Australia
[10]Biomedical Systems Research Laboratory, School of Electrical Engineering and Telecommunications, UNSW Sydney, Sydney, Australia
[11]Burnet Institute, Melbourne, Australia

Corresponding Author:
Seng Hansun, MCS
School of Clinical Medicine, South West Sydney
UNSW Medicine & Health
UNSW Sydney
High Street, Kensington, NSW
Sydney, 2052
Australia
Phone: 61 456541224
Email: s.hansun@unsw.edu.au

Abstract

Background: Tuberculosis (TB) remains a significant health concern, contributing to the highest mortality among infectious diseases worldwide. However, none of the various TB diagnostic tools introduced is deemed sufficient on its own for the diagnostic pathway, so various artificial intelligence (AI)–based methods have been developed to address this issue.

Objective: We aimed to provide a comprehensive evaluation of AI-based algorithms for TB detection across various data modalities.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) 2020 guidelines, we conducted a systematic review to synthesize current knowledge on this topic. Our search across 3 major databases (Scopus, PubMed, Association for Computing Machinery [ACM] Digital Library) yielded 1146 records, of which we included 152 (13.3%) studies in our analysis. QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies version 2) was performed for the risk-of-bias assessment of all included studies.

Results: Radiographic biomarkers (n=129, 84.9%) and deep learning (DL; n=122, 80.3%) approaches were predominantly used, with convolutional neural networks (CNNs) using Visual Geometry Group (VGG)-16 (n=37, 24.3%), ResNet-50 (n=33, 21.7%), and DenseNet-121 (n=19, 12.5%) architectures being the most common DL approach. The majority of studies focused on model development (n=143, 94.1%) and used a single modality approach (n=141, 92.8%). AI methods demonstrated good performance in all studies: mean accuracy=91.93% (SD 8.10%, 95% CI 90.52%-93.33%; median 93.59%, IQR 88.33%-98.32%), mean area under the curve (AUC)=93.48% (SD 7.51%, 95% CI 91.90%-95.06%; median 95.28%, IQR 91%-99%), mean sensitivity=92.77% (SD 7.48%, 95% CI 91.38%-94.15%; median 94.05% IQR 89%-98.87%), and mean specificity=92.39% (SD 9.4%, 95% CI 90.30%-94.49%; median 95.38%, IQR 89.42%-99.19%). AI performance across different biomarker types showed mean accuracies of 92.45% (SD 7.83%), 89.03% (SD 8.49%), and 84.21% (SD 0%); mean AUCs of 94.47% (SD 7.32%), 88.45% (SD 8.33%), and 88.61% (SD 5.9%); mean sensitivities of 93.8% (SD 6.27%), 88.41% (SD 10.24%), and 93% (SD 0%); and

XSL•FO
RenderX

mean specificities of 94.2% (SD 6.63%), 85.89% (SD 14.66%), and 95% (SD 0%) for radiographic, molecular/biochemical, and physiological types, respectively. AI performance across various reference standards showed mean accuracies of 91.44% (SD 7.3%), 93.16% (SD 6.44%), and 88.98% (SD 9.77%); mean AUCs of 90.95% (SD 7.58%), 94.89% (SD 5.18%), and 92.61% (SD 6.01%); mean sensitivities of 91.76% (SD 7.02%), 93.73% (SD 6.67%), and 91.34% (SD 7.71%); and mean specificities of 86.56% (SD 12.8%), 93.69% (SD 8.45%), and 92.7% (SD 6.54%) for bacteriological, human reader, and combined reference standards, respectively. The transfer learning (TL) approach showed increasing popularity (n=89, 58.6%). Notably, only 1 (0.7%) study conducted domain-shift analysis for TB detection.

Conclusions: Findings from this review underscore the considerable promise of AI-based methods in the realm of TB detection. Future research endeavors should prioritize conducting domain-shift analyses to better simulate real-world scenarios in TB detection.

Trial Registration: PROSPERO CRD42023453611; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023453611

(J Med Internet Res 2025;27:e69068) doi:10.2196/69068

KEYWORDS

AI; artificial intelligence; deep learning; diagnostic performance; machine learning; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analysis; QUADAS-2; Quality Assessment of Diagnostic Accuracy Studies version 2; systematic literature review; tuberculosis detection

Introduction

Tuberculosis (TB) stood as the dominant infectious disease threat worldwide, affecting over 10.5 million individuals and claiming the lives of 1.3 million people in 2022 [1]. A pivotal strategy in halting this worldwide epidemic revolves around breaking the transmission chain by identifying and treating all individuals with infectious forms of TB [2]. Various diagnostic tools have been introduced, including chest radiography, tuberculin skin tests, interferon-gamma release assays, sputum smear microscopy, sputum mycobacterial culture, and an array of nucleic acid amplification tests [3]. Nonetheless, each of these diagnostic modalities faces significant implementation challenges, and none alone is deemed adequate for the diagnostic pathway. Consequently, researchers have endeavored to integrate artificial intelligence (AI)–based algorithms to augment the detection of TB.

AI encompasses a spectrum of techniques enabling computer programs to tackle intricate problems by emulating human cognitive processes [4]. Presently, 2 widely used terms denote AI techniques that necessitate minimal-to-no human intervention: machine learning (ML) and deep learning (DL). ML possesses the capability to discern meaningful patterns within datasets autonomously, without explicit programming [5]. Conversely, DL emerges as a subdomain within ML, leveraging deep and intricate neural networks to extract features or patterns from datasets for subsequent analysis [5,6].

In this study, we aimed to comprehensively evaluate the performance of AI-based algorithms, particularly ML and DL, for TB detection. This aligns closely with the first pillar of the World Health Organization (WHO) End TB Strategy, which includes systematic screening for TB in high-risk groups [7]. In fact, a novel recommendation was issued by WHO in 2021: the approval of AI tools to analyze chest X-rays (CXRs) for TB detection in place of human readers [8].

Several notable review papers have emerged. Jimmy et al [9] conducted a meticulous systematic review, delving into various DL methods for TB detection based on CXR. Similarly, Sharma et al [10] contributed a narrative survey spotlighting DL-based convolutional neural networks (CNNs) tailored for TB diagnosis from CXR. Zeyu et al [11] embarked on a narrative review aimed at dissecting diverse DL-based approaches for TB diagnosis. Oloko-Oba and Viriri [12] scrutinized DL techniques for TB detection via chest radiographs. Santosh et al [13] conducted a systematic review, albeit with a narrower time frame spanning 5 years (2016-2021), focusing on DL for TB screening using CXR. Carvalho et al [14] contributed a systematic review on DL techniques used for classifying TB bacilli in microscopic Ziehl-Neelsen (ZN) sputum smear images. Da Silva Barros et al [15] conducted a systematic review on ML models geared toward predicting TB treatment outcomes. Distinguished from these endeavors, our current review took a broader perspective, encompassing not only DL but also ML as integral components of AI-based methods for TB detection.

Siddiqui and Garg [16] examined recent studies of intelligent techniques (ML and DL) for diagnosing pulmonary TB. Singh et al [17] scrutinized the drawbacks of conventional TB diagnostics, while exploring the utility of ML and DL methods for TB diagnosis. Additionally, they highlighted several commercial computer-aided detection (CAD) tools as promising AI-driven instruments. Notably, both reviews were reported as narrative reviews, rather than using the systematic review methodology we adopted here.

Harris et al [18] meticulously conducted a systematic review focusing on the diagnostic accuracy of AI-based tools in TB detection using CXR. Hansun et al [19] recently published a systematic review examining the effectiveness of ML and DL methods for TB detection using CXR. Both studies aligned closely with our review but were limited to CXR. In contrast, our review encompassed a broader spectrum of diagnostic data, including radiographic, biochemical, physiological, and other clinical data. This broader scope enabled a more comprehensive evaluation of AI-based methods across diverse data modalities for TB detection.

XSL•FO
RenderX

Methods

Design, Registration, and Information Sources

In conducting this systematic review, we adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Multimedia Appendix 1) [20]. Our review protocol, in accordance with the PRISMA-Protocol 2015 statement [21,22], was registered on PROSPERO (Prospective Register of Systematic Reviews; ID CRD42023453611) [23,24].

The systematic review drew upon 3 primary academic search systems and online databases: Scopus, PubMed, and the Association for Computing Machinery (ACM) Digital Library. These platforms are widely recognized as pivotal sources for comprehensive literature reviews [25]. The searches encompassed all published literature up to July 25, 2023.

Ethical Considerations

As this systematic literature review focused on retrospective studies, no ethical approval was required.

Search Strategy

For our search strategy, we used 3 primary keywords: "artificial intelligence," "tuberculosis," and "detect*." Additionally, we incorporated several alternative terms for each main keyword to ensure comprehensive coverage during the search process, resulting in the findings outlined in Table 1. Initially, we identified a total of 1146 records; however, certain records were not available from the respective databases (n=58, 5.1%). The full search queries can be seen in Multimedia Appendix 2.

Table 1. Sample search queries and results.

Search query	Scopus, n/N (%)	PubMed, n/N (%)	ACM[a], n/N (%)
(TITLE-ABS-KEY ("Machine Learning" OR "Predictive Analytics" OR "Statistical Learning" OR "Deep Learning" OR "Artificial Intelligence" OR "AI") AND TITLE-ABS-KEY ("Tuberculosis" OR "TB") AND TITLE-ABS-KEY ("Early detection" OR "detect*" OR "Early diagnosis" OR "diagnosis")) AND DOCTYPE (ar OR cp) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p"))	857/1146 (74.8)	284/1146 (24.8)	5/1146 (0.4)
Not available for download (n=58)	53/58 (91.4)	5/58 (8.6)	0
Available for download (n=1088)	804/1088 (73.9)	279/1088 (25.6)	5/1088 (0.5)

[a]ACM: Association for Computing Machinery.

Inclusion and Exclusion Criteria, Data Extraction, and Storage

All downloaded records underwent scrutiny against the inclusion and exclusion criteria outlined for this review study. There were 4 criteria that all included studies had to meet: (1) full-text papers reporting original data published in peer-reviewed journals or proceedings, (2) cross-sectional diagnostic test evaluations, (3) papers focused on TB detection using AI, and (4) papers written in English.

The information extracted from each included study comprised the title, authors, publication year, journal title, objectives, and outcome measures. Dataset characteristics, including their source and total number of data items, were documented, along with details regarding the modality and type of data used. The AI methods applied, evaluation techniques used, performance metrics, best results obtained, and any comparisons with other studies were recorded. Additionally, outcome types, citation counts, and information regarding sponsors or funding sources, were also gathered as part of the comprehensive extraction process.

Each review author was actively engaged in the conceptualization and execution of all phases outlined in the PRISMA 2020 guidelines. At least 2 review authors were involved in each phase, including identification, screening, and assessment of eligibility for inclusion. To ensure accuracy and transparency, we documented the results of each phase using a standardized spreadsheet.

Outcomes Assessed

The first outcome from this review was the compilation of diverse AI methods used for TB detection. The second outcome encompassed diagnostic performance exhibited by various AI methods in TB detection. This summary encompassed key metrics, such as accuracy, area under the receiver operating characteristic (ROC) curve (AUROC), sensitivity, and specificity, providing valuable insights into the efficacy and reliability of these methods in clinical practice.

Strategy for Data Analysis and Synthesis

We adopted a narrative synthesis approach, integrating information extracted from the included studies within the text. The narrative synthesis method facilitated an in-depth exploration of the relationships and insights both within and across included studies, stratified by various attributes, such as modality, biomarker types, and AI methods. Descriptive statistics, primarily using box-whisker and violin plots, alongside tables and figures, presented the quantitative outcomes of this systematic review. Given the pronounced heterogeneity observed in study designs, comparators, biomarker types, evaluation techniques, and AI methods encompassed in this review study, a meta-analysis was deemed impractical.

Risk-of-Bias Assessment

In assessing the risk of bias within the included studies, we used the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies version 2) tool. This tool is widely recommended for systematic reviews as it enables the evaluation of bias and applicability in diagnostic accuracy studies [26]. It comprises

XSL•FO
RenderX

4 key domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. To tailor the assessment to the specific focus of this review on evaluating AI-based methods' performance in TB detection, adjustments were made to the questions in the QUADAS-2 tool. Two review authors (SH and AW) assessed each included study, and a third opinion was sought from the author AA or the author GBM in the case of disagreements.

Results

Overview

Figure 1 delineates all the processes conducted throughout the systematic review. We adhered to the PRISMA 2020 flowchart, including the incorporation of a process timeline and the inclusion of a registered protocol status in PROSPERO.

Figure 1. PRISMA 2020 flowchart with timeline and PROSPERO information. ACM: Association for Computing Machinery; AI: artificial intelligence; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis; TB: tuberculosis.

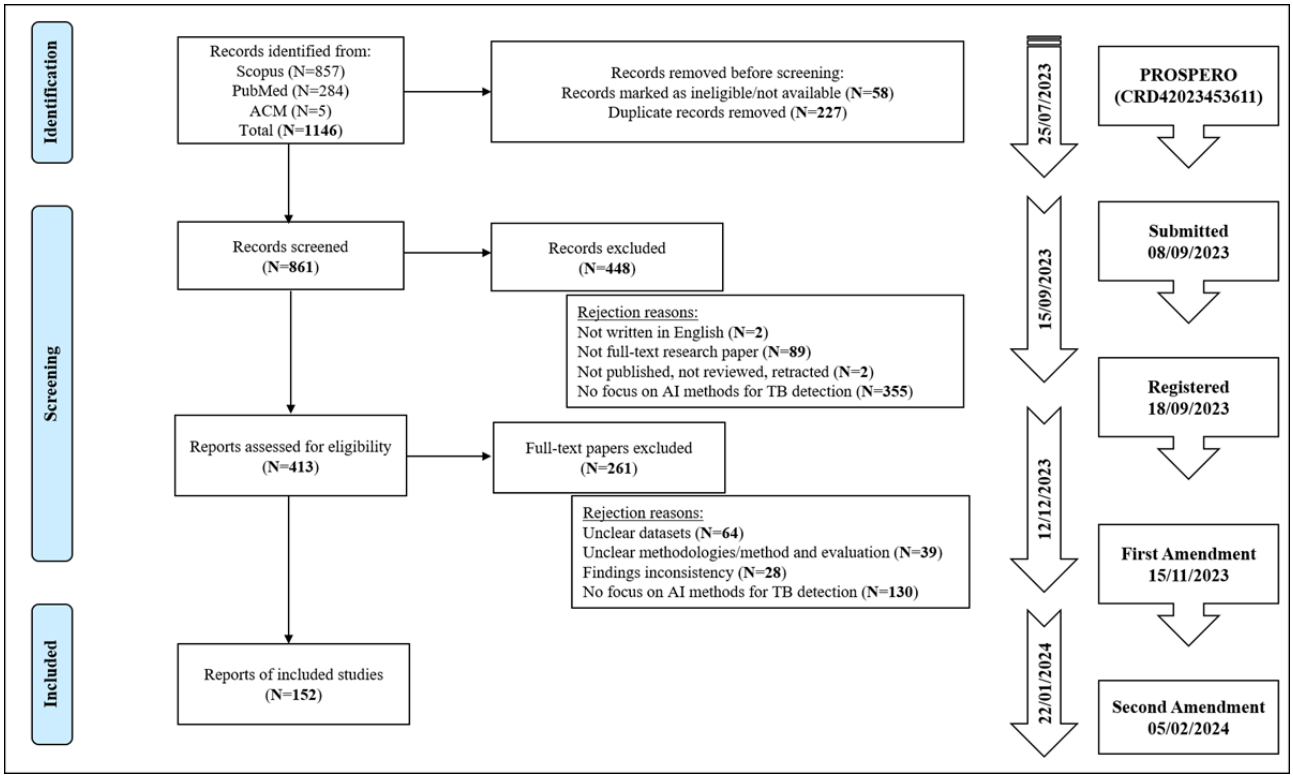

Of the 1146 records initially identified from Scopus (n=857, 74.8%), PubMed (n=284, 24.8%), and ACM Digital Library (n=5, 0.4%), 58 (5.1%) were found to be unavailable. Upon checking and removing duplicates (n=227, 20.9%), we were left with 861 (79.1%) unique titles to be passed to the screening phase. Subsequently, titles, abstracts, and keywords underwent examination based on 4 primary selection criteria, where 413 (48%) records remained for eligibility assessment.

All 413 records underwent thorough assessment by examining their full-text content. From this pool, 261 (63.2%) papers were excluded, leaving us with a total of 152 (36.8%) papers being included as part of this systematic review. Subsequently, the data extraction process was carried out for all included studies.

The entirety of the review process spanned approximately 6 months (July 2023-January 2024).

General Characteristics of Included Studies

A summary of the general characteristics of the 152 studies included in this systematic review is included in Tables 2 and 3. The majority of the included studies opted for a single modality (n=141, 92.8%), with a smaller subset using a multimodal approach (n=11, 7.2%). Among the various data types used, radiographic data emerged as the most prevalent data type (n=129, 84.9%), followed by biochemical (n=21, 13.8%) and physiological (n=16, 10.5%) data types. Several studies using a multimodal approach integrated more than 1 data type, such as a combination of radiographic and physiological data types.

XSL•FO
RenderX

Table 2. Summary of mutually exclusive general characteristics of the included studies (N=152).

Characteristics	Studies, n (%)
Modality	
Single	141 (92.8)
Multimodal	11 (7.2)
Evaluation technique	
Holdout	95 (62.5)
k-Fold CV[a]	48 (31.6)
External data	9 (5.9)
Outcome type	
Model	143 (94.1)
Application/prototype	8 (5.3)
Clinical evaluation	1 (0.7)
TL[b]	89 (58.6)
Comparison with other studies	68 (44.7)
Funding/sponsor	78 (51.3)

[a]CV: cross-validation.

[b]TL: transfer learning.

Table 3. Summary of nonmutually exclusive general characteristics of the included studies (N=152).

Characteristics	Studies, n (%)
AI[a] type	
ML[b]	49 (32.2)
DL[c]	122 (80.3)
Reference standard	
Human reader	103 (67.8)
Bacteriological	43 (28.3)
Not available	24 (15.8)
Type of data	
Radiographic	129 (84.9)
Biochemical	21 (13.8)
Physiological and clinical	16 (10.5)

[a]AI: artificial intelligence.

[b]ML: machine learning.

[c]DL: deep learning.

In terms of evaluation techniques, the majority of studies (n=95, 62.5%) used the holdout evaluation technique, which involves dividing the entire dataset into training and test sets. In total, 48 (31.6%) studies used the k-fold cross-validation (CV) technique, dividing the dataset into k partitions and conducting the evaluation k times. Only 9 (5.9%) studies incorporated external datasets during the evaluation phase. Of the 152 included studies, 143 (94.1%) focused on AI model development, while 8 (5.3%) studies concentrated on application or prototype creation, and 1 (0.7%) study examined the clinical evaluation of AI-based CAD tools.

Regarding the AI approaches, a large number of studies (n=122, 80.3%) used various DL methods, while 49 (32.2%) studies used ML methods. We also collected information regarding various reference standards used in the included studies. We classified them into 2 groups, namely the human reader and bacteriological standards. Of the 152 included studies, 103 (67.8%) used human reader reference standards, 43 (28.3%) used bacteriological reference standards, and 24 (15.8%) did not report any reference standard. Similar to the data-types group, both AI-type and reference standard groups were not mutually exclusive.

XSL•FO
RenderX

In addition, 89 (58.6%) studies used the TL approach, a relatively recent and validated strategy for addressing domain-specific challenges [27], especially prevalent in studies using DL methods. Moreover, 68 (44.7%) studies conducted comparisons with multiple prior studies. Notably, an equal proportion of studies received funding or sponsorship from various entities (n=78, 51.3%) compared to those that did not receive any funding (n=74, 48.7%). All included studies were published between 2011 and 2023.

A total of 141 distinct data sources were identified across the included studies, comprising 2,072,457 records. Among these, the most frequently used data sources were Shenzhen (SZ; n=51, 36.2%), Montgomery County (MC; n=46, 32.6%), Kaggle TB CXR (n=6, 4.3%), and TBX11K (n=5, 3.5%), all of which provided radiographic data. MC and SZ were the predominant data sources, both publicly available CXR datasets from the US National Library of Medicine. MC contains 138 images, while SZ offers 662 images [28]. The Kaggle TB CXR dataset contains 7000 images, although only 4200 (60%) are publicly accessible [29]. Conversely, TBX11K, introduced in 2020, stands out as a newer and more extensive CXR dataset for TB-related research, featuring a total of 11,200 images [30]. For further insights into the extracted data from all included studies, detailed information is available in Multimedia Appendix 3.

AI Techniques for TB Detection

Among the 152 included studies, the majority were published within the past 8 years, with only 3 (2%) studies predating 2016 [31-33]. As of January 18, 2024, these studies collectively garnered 6070 citations on Google Scholar. Remarkably, the top 3 (2%) cited publications were identified as Lakhani and Sundaram [34] with 1695 citations, Pasa et al [35] with 312 citations, and Lopes and Valiati [36] with 260 citations. This underscored the burgeoning interest among researchers in integrating AI-based methods into TB detection studies.

The included studies implemented various AI-based algorithms, which could be broadly categorized into 2 major approaches: ML and DL. Specifically, 102 (67.1%) studies used DL methods [30,34,35,37-135], while 30 (19.7%) studies used ML techniques [31-33,136-162]. Additionally, 19 (12.5%) studies used a combination of both ML and DL approaches [36,163-180]. Furthermore, 1 (0.7%) study [181] focused on the implementation of a DL-based CAD tool, namely Lunit INSIGHT v4.7.2, specifically designed for systematic TB screening in a low-prevalence setting.

Figure 2 illustrates the diverse array of ML methods used in the included studies. Notably, support vector machines (SVMs; n=33, 21.7%), random forests (RFs; n=17, 11.2%), and logistic regression (LR; n=15, 9.9%) emerged as the 3 most prevalent ML techniques.

Figure 2. Different ML methods identified from the included studies. ANN: artificial neural network; BMO: bird mating optimizer; DT: decision tree; DTE-SVM: deep transferred EfficientNet with support vector machine; ELM: extreme learning machine; ERT: extremely randomized tree; GBM: Generalized Boosting Machine; GENFIS: genetic-neuro-fuzzy inferential system; GMM: Gaussian mixture model; kNN: k nearest neighbor; Lasso: least absolute shrinkage and selection operator; LBP: local binary pattern; LDA: linear discriminant analysis; LR: logistic regression; MBO: monarch butterfly optimization; ML: machine learning; MLP: multilayer perceptron; MLSM: multilevel similarity measure; NB: naïve Bayes; OSELM: online sequential extreme learning machine. PLS-DA: partial least squares–discriminant analysis; PCA: principal component analysis; QD: quadratic discriminant; RF: random forest; SLDT: stacked loopy decision tree; SLFN: single hidden layer feedforward neural network; SVM: support vector machine; WSO: water strider optimization; XGBoost: extreme gradient boosting.

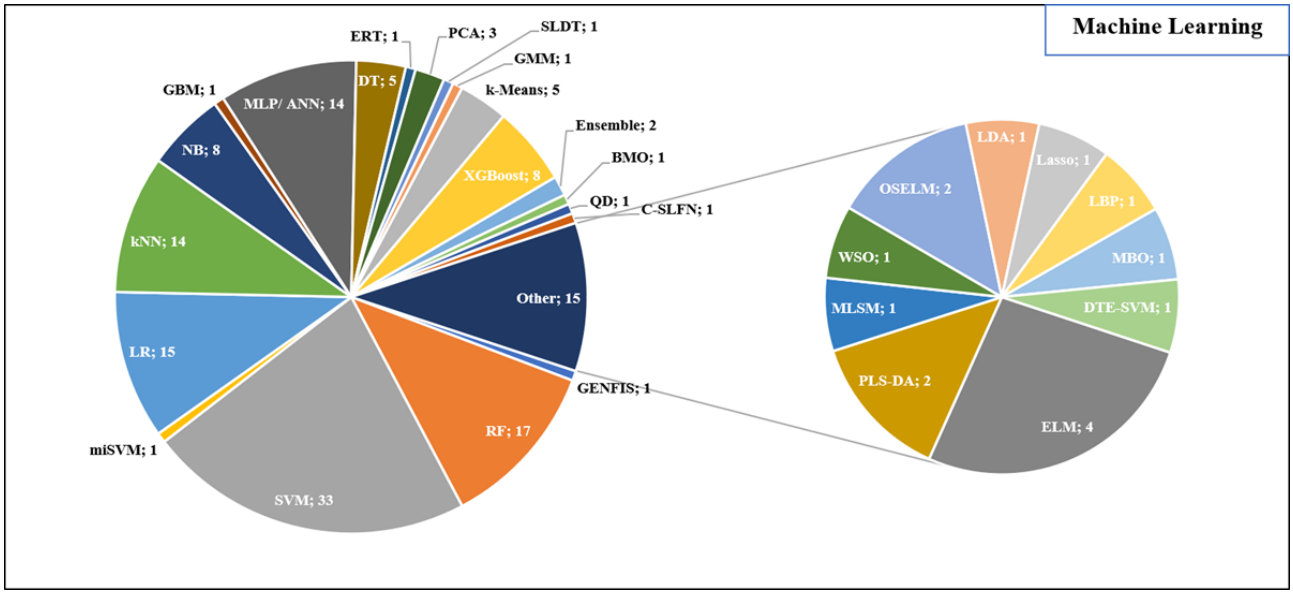

CNNs were the predominant DL method among the included studies. Alongside custom CNNs (n=39, 25.7%), many researchers have leveraged established DL architectures, notably including Visual Geometry Group (VGG)-16 (n=37, 24.3%), ResNet-50 (n=33, 21.7%), and DenseNet-121 (n=19, 12.5%). Figure 3 provides an overview of the various DL methods identified in the included studies.

XSL•FO
RenderX

Figure 3. DL methods identified from the included studies. AFC-CNN: adaptive fractional crow convolutional neural network; BCNN: Bayesian-based convolutional neural network; bi-LSTM: bidirectional long short-term memory; BRANN: Bayesian regularization artificial neural network; CBAM: convolutional block attention module; CNN: convolutional neural network; DenseNet-SPP: DenseNet with spatial pyramid pooling; DL: deep learning; FC-SVNN: fractional crow search–based deep convolutional neural network; FCOS: fully convolutional one stage; GAF-CNN: Gramian angular field convolutional neural network; GAN: generative adversarial network; GCN: global complex network; GIN: graph isomorphism network; GRU: gated recurrent unit; IEViT: image enhanced vision transformer; LCN: local complex network; LSTM: long short-term memory; MSI-PTDM: multistream integration–pulmonary tuberculosis diagnosis model; MT-MinCutPool: multivariate time series with MinCutPool; RCNN: region-based convolutional neural network; SSD: single-shot multibox detector; VGG: Visual Geometry Group; ViT: vision transformer; YOLO: you only look once.

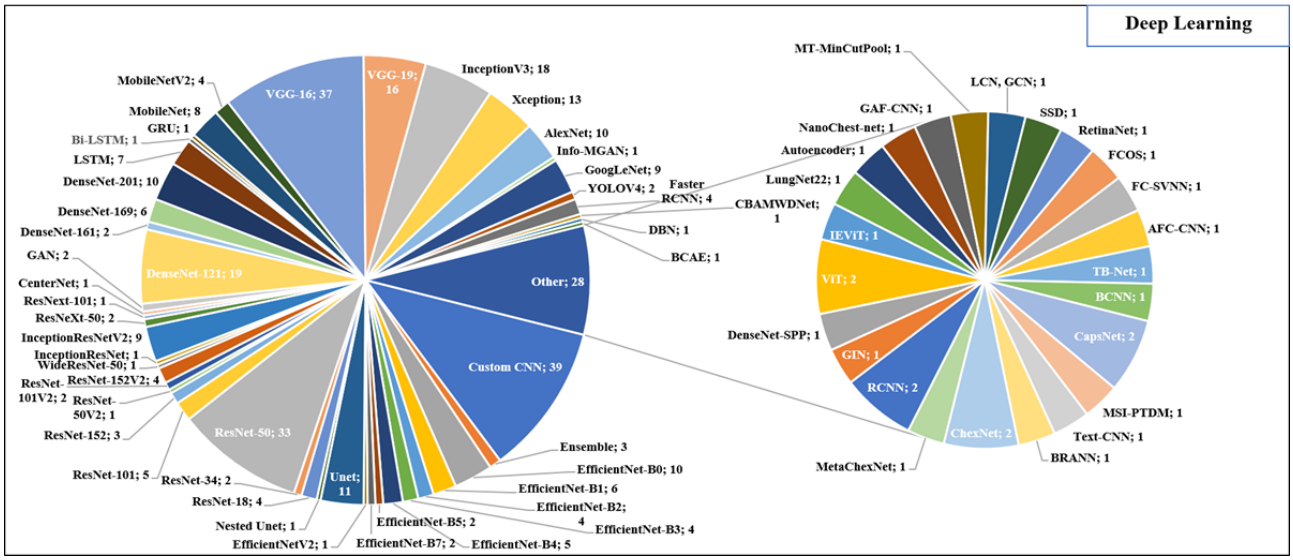

Regarding the diagnostic performance of AI methods used in the literature, the most frequently used evaluation metrics encompassed accuracy (n=128, 84.2%), sensitivity (n=112, 73.7%), the AUC (n=87, 57.2%), and specificity (n=77, 50.7%). Accuracy denotes the proportion of correctly predicted cases, encompassing true positives and true negatives among all cases [182]. Sensitivity measures a model's ability to accurately identify individuals with a condition [183,184]. The AUC, derived from the ROC curve, quantitatively assesses a model's performance [182]. Specificity gauges a model's accuracy in identifying individuals without a condition [183,184]. Table 4 presents the distribution of various performance metrics identified in the included studies. For an in-depth exploration of the performance metrics used across the included studies, refer to the detailed description provided in Multimedia Appendix 4.

XSL•FO
RenderX

Table 4. Different performance evaluation metrics in the included studies (N=152).

Metric	Studies, n (%)
Accuracy	128 (84.2)
AUROC[a]	87 (57.2)
Sensitivity/recall/true-positive rate	112 (73.7)
Specificity/true-negative rate	77 (50.7)
Precision/positive predictive value	56 (36.8)
Negative predictive value	9 (5.9)
F_1-score	55 (36.2)
Average recall	1 (0.7)
Mean average precision	2 (1.3)
Matthews correlation coefficient	7 (4.6)
True detection rate	1 (0.7)
Area under the alternative free-response ROC[b] curve	1 (0.7)
Error loss	1 (0.7)
Cohen kappa	3 (2.0)
Root mean square error	2 (1.3)
False positive rate	4 (2.6)
False negative rate	1 (0.7)
Squared error	1 (0.7)
Balanced accuracy	1 (0.7)
Fowlkes-Mallows index	1 (0.7)

[a]AUROC: area under the receiver operating characteristic curve.

[b]ROC: receiver operating characteristic.

AI Diagnostic Performance

Figure 4 provides a comprehensive overview of the overall performance results across all included studies. From the depicted box plot, it is evident that accuracy ranged from 46.48% [59] to 100% [118], with a mean value of 91.93% (SD 8.10%, 95% CI 90.52%-93.33%) and a median of 93.59% (IQR 88.33%-98.32%). Similarly, the AUC ranged from 45.96% [59] to 100% [72,78,96,178], with a mean value of 93.48% (SD 7.51%, 95% CI 91.90%-95.06%) and a median of 95.28% (IQR 91%-99%). Sensitivity spanned from 60% [156] to 100% [85,92,100,118,128,132,141,146,163,178,181], with a mean value of 92.77% (SD 7.48%, 95% CI 91.38%-94.15%) and a median of 94.05% (IQR 89%-98.87%). Meanwhile, specificity ranged from 53.2% [129] to 100% [113,162,174,177], with a mean value of 92.39% (SD 9.4%, 95% CI 90.30%-94.49%) and a median of 95.38% (IQR 89.42%-99.19%). Several outliers were detected for each metric, including 4 for accuracy, at 46.48% [59], 63% [126], 63.9% [129], and 70% [156]; 2 for the AUC, at 45.96% [59] and 69.2% [129]; 3 for sensitivity, at 60% [156], 65.52% [155], and 69.07% [143]; and 5 for specificity, at 53.2% [129], 60% [146], 60.8% [168], 69.44% [67], and 73% [152].

XSL•FO
RenderX

Figure 4. Overall accuracy, AUC, sensitivity, and specificity of the included studies. AUC: area under the curve.

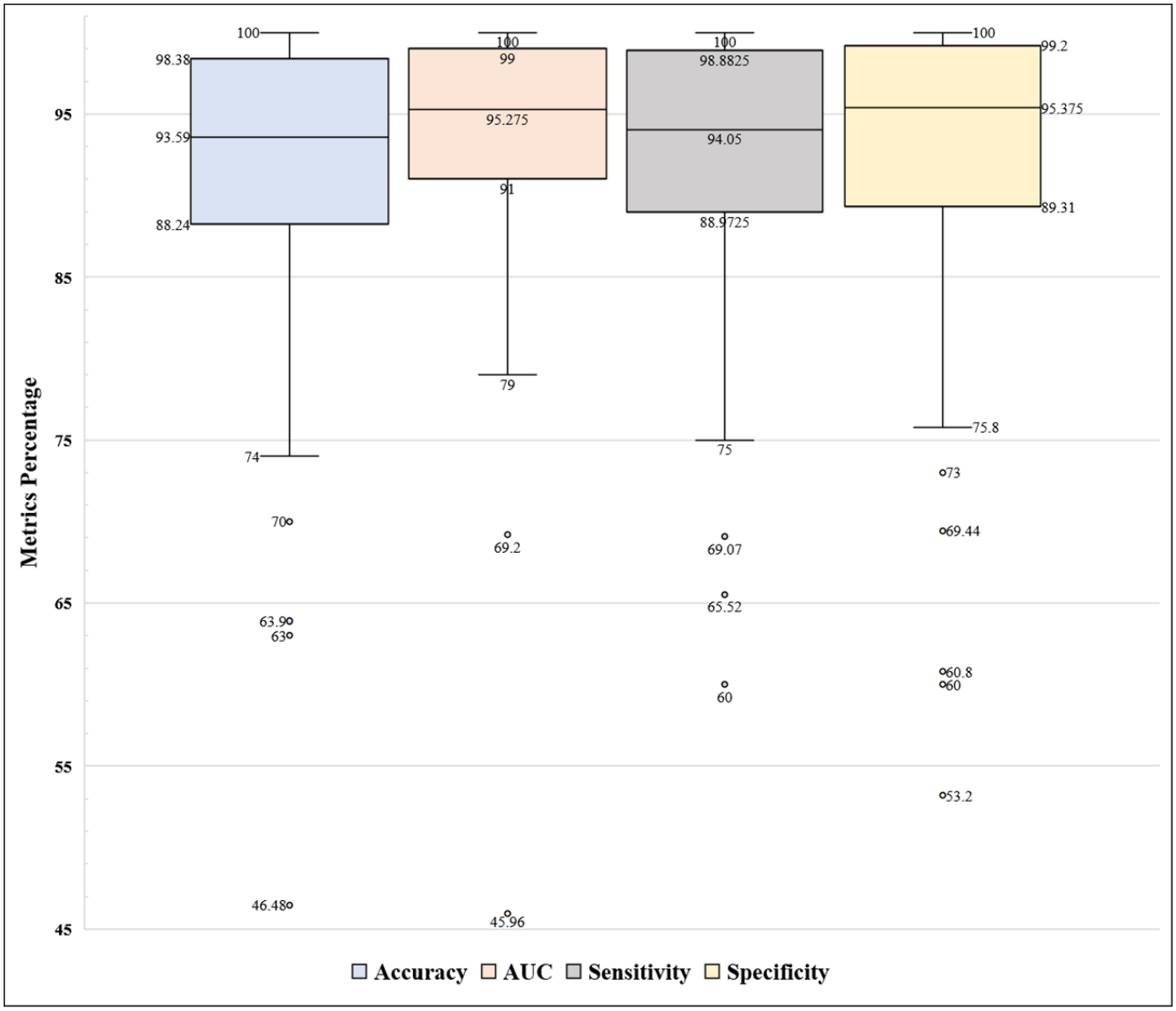

We conducted further performance comparisons based on several criteria, including whether a study used the TL method and whether it adopted a single (S) or multimodal (M) approach. Figure 5 presents split-grouped violin plots illustrating the comparative performance results of the included studies based on TL (yes [Y], no [N]) and the modality used.

Figure 5. Comparative performance based on TL (left) and modality (right). AUC: area under the curve; M: multimodal; N: no; S: single; TL: transfer learning; Y: yes.

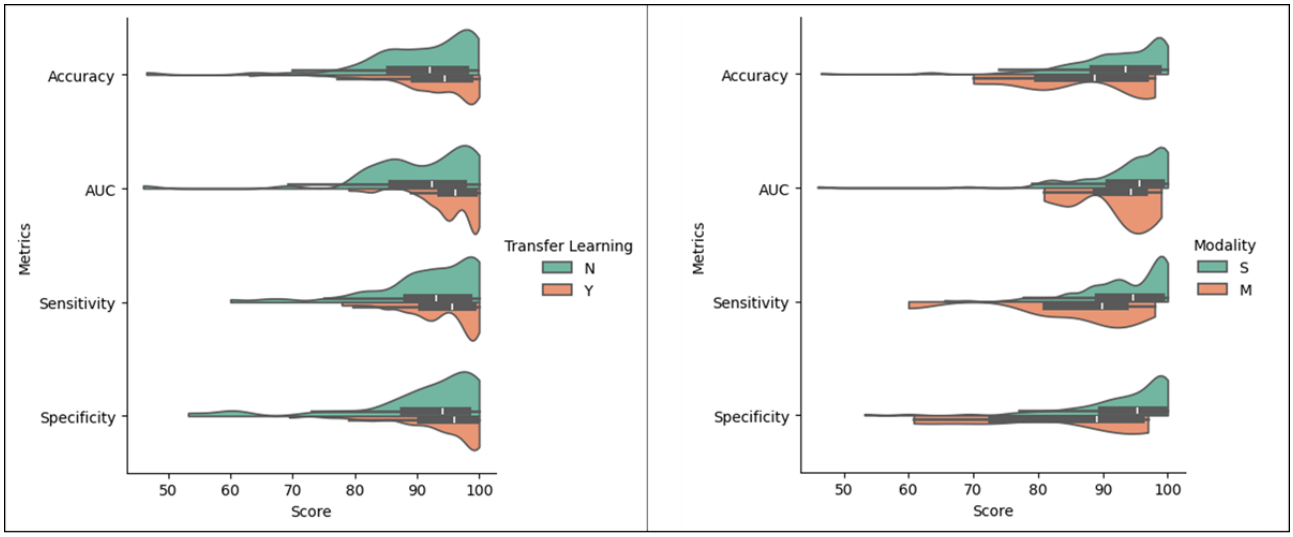

XSL•FO
RenderX

We also examined the performance of AI methods used in all the included studies based on the AI techniques applied (ML, DL, or both [ML/DL]), the types of data or biomarkers used (radiographic [R], molecular [M], physiological [P]), and the reference standard used (human [H], bacteriological [B], or both [H/B]). Figure 6 showcases the box plots, offering a comparative analysis of the performance results of the included studies based on AI methods, biomarker types, and reference standards.

Figure 6. Comparative performance based on AI methods (left), biomarker types (middle), and reference standards (right). AI: artificial intelligence; AUC: area under the curve; B: bacteriological; DL: deep learning; H: human; H/B: both human and bacteriological; M: molecular; ML: machine learning; ML/DL: both machine learning and deep learning; P: physiological; R: radiographic.

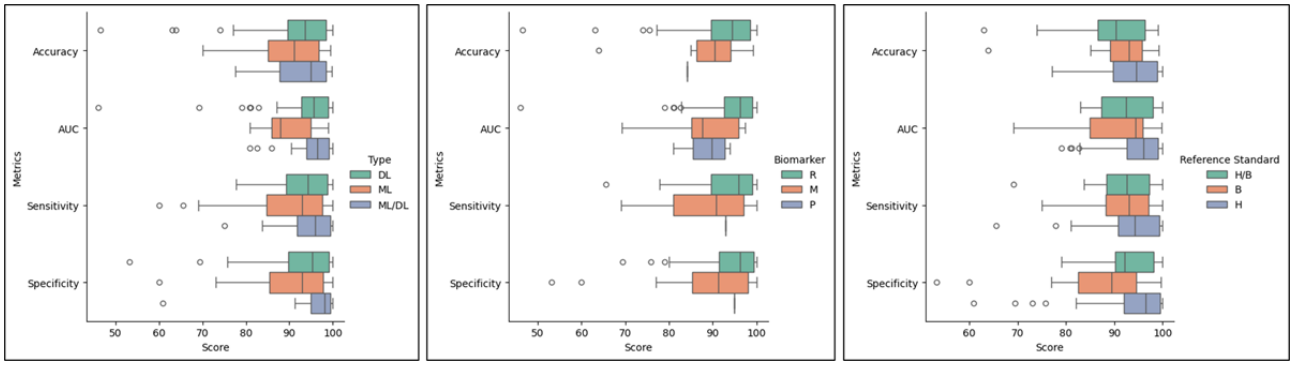

Specifically, the AI performance across different biomarker types showed mean accuracies of 92.45% (SD 7.83%), 89.03% (SD 8.49%), and 84.21% (SD 0%); mean AUCs of 94.47% (SD 7.32%), 88.45% (SD 8.33%), and 88.61% (SD 5.9%); mean sensitivities of 93.8% (SD 6.27%), 88.41% (SD 10.24%), and 93% (SD 0%); and mean specificities of 94.2% (SD 6.63%), 85.89% (SD 14.66%), and 95% (SD 0%) for radiographic, molecular/biochemical, and physiological biomarkers, respectively. Meanwhile, AI performance across various reference standards show mean accuracies of (SD 7.3%), 93.16% (SD 6.44%), 88.98% (SD 9.77%); mean AUCs of 90.95% (SD 7.58%), 94.89% (SD 5.18%), 92.61% (SD 6.01%); mean sensitivities of 91.76% (SD 7.02%), 93.73% (SD 6.67%), 91.34% (SD 7.71%); and mean specificities of 86.56% (SD 12.8%), 93.69% (SD 8.45%), 92.7% (SD 6.54%) for bacteriological, human reader, and combined bacteriological and human reader reference standards, respectively.

Risk-of-Bias Assessment Result

Figure 7 presents the QUADAS-2 outcomes across all included studies. Among the assessed studies, 18 (11.8%) [32,46-48,57,70,95,97,111,124,127,136,140,144,153,169,170,178] were flagged for a high risk of bias, with 1 (0.7%) study [141] marked as having an unclear risk concerning patient selection, largely attributable to incomplete or absent information in the data selection process. Moreover, 4 (2.6%) studies [41,101,108,174] were identified as having a high risk regarding the index test, attributed to the absence of crucial details on model architecture and parameters. Additionally, 3 (2%) studies [57,156,158] were classified as having a high risk and 8 (5.3%) studies [32,63,64,97,98,111,140,161] as having an unclear risk concerning reference standards, due to incomplete or missing information regarding the reference standards used. Concerning flow and timing, 2 (1.3%) studies [88,95] were deemed to have a high risk of bias, while 8 (5.3%) studies [53,65,67,87,102,141,156,179] were assigned an unclear risk, primarily due to unclear or absent information regarding the time interval and intervention provided in those studies.

Figure 7. Summary of QUADAS-2 results of the included studies. QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies version 2.

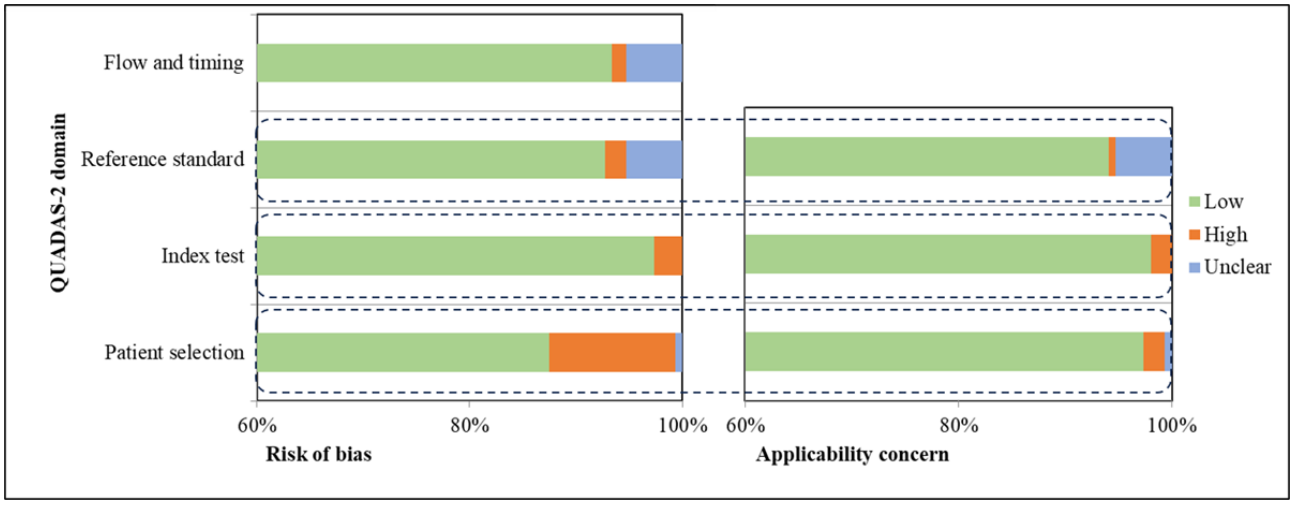

The majority of the included studies exhibited a low level of concern about applicability. Only 3 (2%) studies [48,88,144] were identified with a high risk and 1 (0.7%) study [141] with an unclear risk in terms of patient selection. Additionally, 3

XSL•FO
RenderX

(2%) studies [41,108,151] were flagged with a high risk concerning the index test, while 1 (0.7%) study [158] showed a high risk, and 8 (5.3%) studies [32,63,64,97,98,111,140,161] had an unclear risk concerning reference standards. For comprehensive QUADAS-2 results, please refer to Multimedia Appendix 5.

Discussion

Principal Findings

This review study examined the diagnostic performance of various AI-based algorithms, including both ML and DL, for TB detection across different techniques, data modalities, and reference standards used in the included studies. Approximately 84.9% (n=129) of the included studies used radiographic data types, particularly CXRs. Notably, the SZ and MC datasets emerged as the most frequently used, with 51 and 46 studies, respectively, making use of them. However, it is noteworthy that the availability of data in SZ and MC, totaling 800 images, is relatively limited, posing challenges, especially for studies using DL methods where larger datasets are often required for optimal performance. Consequently, numerous techniques have been introduced to augment data quantity [185,186], including conventional methods, as well as DL-based augmentation techniques.

Few recent studies have used other data types, such as exhaled breath particles [146,150] and cough sounds [63,64,175]. Although not recommended as a standard tool for TB detection, results from these noninvasive approaches could be a supplement to the standard TB detection tool using CXRs. Hence, further studies that analyze CXRs together with other data types are encouraged.

Concerning modality, the majority of the included studies adopted a single modality (n=141, 92.8%) rather than using a multimodal approach (n=11, 7.2%). Studies using a single modality tended to exhibit higher median values and narrower distribution ranges compared to those using a multimodal approach (Figure 5, right). This suggests that the single-modality approach may outperform the multimodal approach. However, more outliers were observed in all performance metrics for the single modality compared to the multimodal approach. Additionally, these outcomes were not obtained under identical scenario settings. When applied under similar scenario settings, the multimodal approach has demonstrated the potential to enhance performance results compared to the single modality [9,187]. This is supported by findings from several included studies that have directly compared single-modality and multimodal approaches [79,86,95].

DL emerged as a cornerstone in the majority of the included studies (n=122, 80.3%), with CNNs taking the forefront. Among these, architectures such as VGG-16 [188], ResNet-50 [189,190], and DenseNet-121 [191] have gained notable traction. More recent DL architectures, including EfficientNets [192], Vision Transformer (ViT) [193], and ConvNeXts [194], have gained increasing attention. These emerging models, with more efficient architectures that require fewer computational resources, show potential for application in medical image processing, particularly in TB detection tasks.

Some of the included studies explored various explainable AI (XAI) methods to bolster the interpretability and trustworthiness of their AI models' outcomes. Originating from the need to strike a balance between interpretability and accuracy in ML and DL models [195], XAI has garnered considerable attention. Among the XAI techniques used in the included studies were gradient-weighted class activation mapping (Grad-CAM), local interpretable model-agnostic explanation (LIME), and Shapley additive explanations (SHAP).

TL is a powerful technique that capitalizes on knowledge acquired from a source dataset and adapts it to a target domain, relaxing the assumption that training data must be independent and identically distributed with test data [196]. TL proves especially beneficial in scenarios with a limited data volume. Indeed, TL has been extensively used in the studies encompassed in this systematic review (n=89, 58.6%), particularly those using DL methods. Notably, as illustrated in Figure 5 (left), investigations integrating TL tend to exhibit a higher median accuracy compared to counterparts. Additionally, they showcase a narrower distribution range, suggesting that TL-based approaches consistently deliver superior and more consistent results.

The majority of the included studies (n=143, 94.1%) focused primarily on the development of AI models for TB detection. Only 8 studies [44,73,92,106,111,137,153,176] progressed beyond model development to build prototypes or systems aimed at supporting real-world applications. Among these, 3 studies implemented their proposed models in real-world settings: eRx, a mobile health system for TB diagnosis deployed in Lima, Peru [111]; the multistream integration–pulmonary tuberculosis diagnosis model (MSI-PTDM), multistream integration for a TB diagnosis model implemented in China [176]; and AIChest4All, an automated CXR-screening system used in Thailand [44]. We excluded studies evaluating commercial TB detection tools, such as Lunit INSIGHT, CAD4TB, and qXR, as they typically do not provide detailed discussions of the underlying AI models. For further insights into commercial TB detection tools, interested readers are referred to the narrative review by Singh et al [17].

In terms of overall performance metrics, the AI methods displayed remarkable achievements. Further analysis based on the AI methods used in the included studies revealed interesting insights. As illustrated in Figure 6 (left), DL exhibits higher accuracy with less variation compared to ML and ML/DL. However, regarding the AUC, sensitivity, and specificity, studies integrating ML/DL tend to yield superior results, followed by DL and then ML. This suggests that the fusion of ML and DL methods may enhance model performance, albeit few studies have explored this avenue.

We also examined the performance results concerning the types of data used. Radiographic data-type studies showcased superior median scores across all performance metrics (Figure 6, middle). Notably, they demonstrated a higher level of consistency, evidenced by their shorter IQRs in all performance metrics. Regarding the reference standards used in the included studies,

XSL•FO
RenderX

studies that used the human reader exceled compared to those that used bacteriological and both reference standards, as indicated by the higher median and mean scores on all performance metrics in Figure 6 (right). Despite accuracy, the IQRs were also shorter for the AUC, sensitivity, and specificity. This finding might be rooted in the large number of included studies that used DL methods on radiographic data types with human reader reference standards.

Finally, it is worth noting that the majority of the included studies overlooked the evaluation of proposed solutions regarding the domain-shift problem. Remarkably, only 1 study [126] addressed domain-shift analysis in the context of TB detection, as highlighted by Hansun et al [197]. Although our review confirmed the high performance of AI-based methods across various data types in TB detection tasks, these evaluations were predominantly conducted on in-domain datasets. In other words, the datasets used for validation and testing shared the same distribution and characteristics as the training data. However, many AI methods, particularly DL, struggle when applied to real-world settings [198]. This discrepancy often arises due to differences in data distribution between the training data and real-world datasets. As such, it is imperative to broaden our focus to include not only in-domain evaluations but also domain-shift analysis in future research endeavors.

Limitations and Strengths

This systematic review has several limitations that warrant acknowledgment. First, our search was limited to the literature published in English from 3 primary academic databases: Scopus, PubMed, and ACM Digital Library. Although these sources are comprehensive, it is possible that relevant studies might exist in other databases, potentially leading to a partial representation of the literature.

Second, we did not conduct further investigations into aspects with high or unclear risks from the bias assessment. Given the prevalence of radiographic datasets across most included studies, we assumed uniformity across these aspects. However, this assumption may overlook potential variations that could impact the overall assessment.

Lastly, in this review study, we opted not to perform a meta-analysis. Although a meta-analysis could provide valuable insights, the diverse nature of the included studies made it challenging to ensure meaningful comparability across studies, potentially affecting the validity and reliability of the synthesized findings. It is also important to note that the variations in data distribution and experimental settings across studies may lead to different results. Although the descriptive statistics reported in this study have been grouped based on several approaches, they should be seen as general performance results of AI for TB detection.

Despite the aforementioned limitations, our review study is among the few to systematically review the evidence available in the literature regarding the efficacy of AI-based methods for TB detection. Although most review studies focus only on a specific biomarker type for TB detection, we performed a more comprehensive review of AI-based methods across diverse biomarker types. We further assessed the AI methods' performance based on several dominant approaches used in the included studies, including TL, multimodalities, reference standards, and ML/DL fusion methods for TB detection, which have never been explored before.

Key Takeaways

AI-based approaches, particularly DL, have been extensively used for TB detection with high-accuracy results. A single-modality approach with chest radiographs has been dominantly used. Further studies that analyze chest radiographs together with other data types (multimodality) are encouraged. AI models trained and tested on radiographic data tend to achieve higher performance compared to other data types. Emerging DL models, such as EfficientNets, ViT, and ConvNeXts, show the potential to enhance TB detection results. TL has shown a great advantage in handling a limited data volume and consistently delivering superior results. Most of the included studies evaluated their proposed solutions using in-domain datasets. Future research endeavors should prioritize conducting domain-shift analyses to better simulate real-world scenarios in TB detection.

Conclusion

This systematic review underscores the considerable promise of AI-based approaches in TB detection. Among the array of AI methods, DL emerges as the predominant choice. This preference for DL is attributed to its consistently robust performance, likely bolstered by the prevalence of studies using radiographic data. A notable observation across many studies is the use of relatively small datasets. Despite achieving commendable results, the potential of DL models could be further enhanced with larger and more diverse datasets.

Finally, although AI models demonstrate impressive performance on in-domain datasets, there is a notable gap in evaluating their robustness to domain shifts. Future research endeavors should prioritize conducting domain-shift analyses to better simulate real-world scenarios in TB detection. This approach would provide invaluable insights into the generalizability and applicability of AI-based methods beyond controlled settings.

Acknowledgments

We used the generative artificial intelligence (AI) tool Chat Generative Pretrained Transformer (ChatGPT) by OpenAI during the editing process of the initial draft, which was subsequently reviewed and revised by the review authors. The original ChatGPT transcripts are available in Multimedia Appendix 6.

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in the published paper and its Supplementary and Figure/Appendix Files.

Authors' Contributions

SH handled conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft, visualization, and project administration; AA and GBM, conceptualization, methodology, validation, formal analysis, investigation, writing—review and editing, and supervision; IB and AW, investigation, resources, data curation, and writing—review and editing; HAR, software, formal analysis, resources, and writing—review and editing; GF, methodology, formal analysis, investigation, and writing—review and editing; and STL and BGC, methodology, investigation, writing—review and editing, and supervision. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[PDF File (Adobe PDF File), 612 KB - jmir_v27i1e69068_app1.pdf]

Multimedia Appendix 2
Full search queries.
[PDF File (Adobe PDF File), 88 KB - jmir_v27i1e69068_app2.pdf]

Multimedia Appendix 3
Full characteristics of the included studies.
[XLSX File (Microsoft Excel File), 113 KB - jmir_v27i1e69068_app3.xlsx]

Multimedia Appendix 4
Complete data summary.
[XLSX File (Microsoft Excel File), 146 KB - jmir_v27i1e69068_app4.xlsx]

Multimedia Appendix 5
QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies version 2) summary.
[XLSX File (Microsoft Excel File), 68 KB - jmir_v27i1e69068_app5.xlsx]

Multimedia Appendix 6
ChatGPT (Chat Generative Pretrained Transformer) transcript.
[PDF File (Adobe PDF File), 148 KB - jmir_v27i1e69068_app6.pdf]

References

1. Global tuberculosis report 2023. World Health Organization. 2023. URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023 [accessed 2025-02-20]
2. Lung T, Marks GB, Nhung NV, Anh NT, Hoa NLP, Anh LTN, et al. Household contact investigation for the detection of tuberculosis in Vietnam: economic evaluation of a cluster-randomised trial. Lancet Global Health 2019 Mar;7(3):e376-e384. [doi: 10.1016/s2214-109x(18)30520-5]
3. Dande P, Samant P. Acquaintance to artificial neural networks and use of artificial intelligence as a diagnostic tool for tuberculosis: a review. Tuberculosis (Edinb) 2018 Jan;108:1-9. [doi: 10.1016/j.tube.2017.09.006] [Medline: 29523307]
4. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Markets 2021 Apr 08;31(3):685-695. [doi: 10.1007/s12525-021-00475-2]
5. Woschank M, Rauch E, Zsifkovits H. A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics. Sustainability 2020 May 06;12(9):3760. [doi: 10.3390/su12093760]
6. Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, et al. Machine learning and artificial intelligence: definitions, applications, and future directions. Curr Rev Musculoskelet Med 2020 Feb 25;13(1):69-76 [FREE Full text] [doi: 10.1007/s12178-020-09600-8] [Medline: 31983042]
7. World Health Organization. WHO Operational Handbook on Tuberculosis. Module 2: Screening - Systematic Screening for Tuberculosis Disease. Geneva: World Health Organization; 2021.

XSL•FO
RenderX

8. David P, Onno J, Keshavjee S, Ahmad Khan F. Conditions required for the artificial-intelligence-based computer-aided detection of tuberculosis to attain its global health potential. Lancet Digital Health 2022 Oct;4(10):e702-e704. [doi: 10.1016/s2589-7500(22)00172-8]
9. Jimmy, Cenggoro T, Pardamean B. Systematic literature review: an intelligent pulmonary TB detection from chest X-rays. 2021 Presented at: 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI); October 28, 2021; Jakarta, Indonesia p. 136-141. [doi: 10.1109/iccsai53272.2021.9609717]
10. Sharma R, Gupta P, Kaur H. A deep learning approach for tuberculosis diagnosis from chest X-rays: a survey. 2021 Presented at: International Conference on Emerging Technologies: AI, IoT, and CPS for Science & Technology Applications; September 6-7, 2021; Chandigarh, India.
11. Zeyu D, Yaakob R, Azman A. A review of deep learning-based detection methods for tuberculosis. 2022 Presented at: 2022 IEEE International Conference on Computing (ICOCO); November 14-16, 2022; Kota Kinabalu, Malaysia p. 68-73. [doi: 10.1109/icoco56118.2022.10031813]
12. Oloko-Oba M, Viriri S. A systematic review of deep learning techniques for tuberculosis detection from chest radiograph. Front Med (Lausanne) 2022;9:830515 [FREE Full text] [doi: 10.3389/fmed.2022.830515] [Medline: 35355598]
13. Santosh K, Allu S, Rajaraman S, Antani S. Advances in deep learning for tuberculosis screening using chest X-rays: the last 5 years review. J Med Syst 2022 Oct 15;46(11):82 [FREE Full text] [doi: 10.1007/s10916-022-01870-8] [Medline: 36241922]
14. Mota Carvalho TF, Santos VLA, Silva JCF, Figueredo LJDA, de Miranda SS, Duarte RDO, et al. A systematic review and repeatability study on the use of deep learning for classifying and detecting tuberculosis bacilli in microscopic images. Prog Biophys Mol Biol 2023 Jul;180-181:1-18. [doi: 10.1016/j.pbiomolbio.2023.03.002] [Medline: 37023799]
15. da Silva Barros MHLF, da Silva Neto SR, de Almeida Rodrigues MG, de Souza Sampaio V, Endo P. How machine learning can help the classification of treatment outcomes of tuberculosis: a systematic review. 2023 Presented at: 56th Hawaii International Conference on System Sciences; January 5-6, 2023; Maui, Hawaii p. 1386-1395. [doi: 10.24251/hicss.2023.173]
16. Siddiqui A, Garg V. Diagnosis of pulmonary tuberculosis through intelligent techniques: a review. 2021 Presented at: 2021 International Conference on Computing Sciences (ICCS); December 4-5, 2021; Phagwara, India p. 189-193. [doi: 10.1109/iccs54944.2021.00045]
17. Singh M, Pujar GV, Kumar SA, Bhagyalalitha M, Akshatha HS, Abuhaija B, et al. Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. Electronics 2022 Aug 23;11(17):2634. [doi: 10.3390/electronics11172634]
18. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest X-rays for pulmonary tuberculosis. PLoS One 2019 Sep 3;14(9):e0221339 [FREE Full text] [doi: 10.1371/journal.pone.0221339] [Medline: 31479448]
19. Hansun S, Argha A, Liaw S, Celler BG, Marks GB. Machine and deep learning for tuberculosis detection on chest X-rays: systematic literature review. J Med Internet Res 2023 Jul 03;25:e43154 [FREE Full text] [doi: 10.2196/43154] [Medline: 37399055]
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
21. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
22. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
23. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. Syst Rev 2018 Feb 20;7(1):32 [FREE Full text] [doi: 10.1186/s13643-018-0699-4] [Medline: 29463298]
24. Hansun S, Argha A, Bakhshayeshi I, Wicaksana A, Liaw S, Celler B, et al. Artificial intelligence-based tuberculosis detection: a systematic review. PROSPERO. 2023. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023453611 [accessed 2024-01-25]
25. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020 Mar 28;11(2):181-217 [FREE Full text] [doi: 10.1002/jrsm.1378] [Medline: 31614060]
26. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
27. Hansun S, Argha A, Alinejad-Rokny H, Liaw S, Celler B, Marks G. Revisiting transfer learning method for tuberculosis diagnosis. 2023 Presented at: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 24-27, 2023; Sydney, Australia p. 1-4. [doi: 10.1109/embc40787.2023.10340441]

XSL•FO
RenderX

28. Jaeger S, Candemir S, Antani S, Wáng YXJ, Lu P, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg 2014 Dec;4(6):475-477 [FREE Full text] [doi: 10.3978/j.issn.2223-4292.2014.11.20] [Medline: 25525580]
29. Rahman T, Khandakar A, Kadir MA, Islam KR, Islam KF, Mazhar R, et al. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. IEEE Access 2020;8:191586-191601. [doi: 10.1109/access.2020.3031384]
30. Liu Y, Wu Y, Ban Y, Wang H, Cheng MM. Rethinking computer-aided tuberculosis diagnosis. 2020 Presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 13-19, 2020; Seattle, WA p. 2643-2652. [doi: 10.1109/cvpr42600.2020.00272]
31. Melendez J, van Ginneken B, Maduskar P, Philipsen RHHM, Reither K, Breuninger M, et al. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays. IEEE Trans. Med. Imaging 2015 Jan;34(1):179-192. [doi: 10.1109/tmi.2014.2350539]
32. Osman M, Mohd NM, Mashor M, Jaafar H. Compact single hidden layer feedforward network for Mycobacterium tuberculosis detection. 2011 Presented at: 2011 IEEE International Conference on Control System, Computing and Engineering; November 25-27, 2011; Penang, Malaysia p. 432-436. [doi: 10.1109/iccsce.2011.6190565]
33. Osman M, Mashor M, Jaafar H. Online sequential extreme learning machine for classification of Mycobacterium tuberculosis in Ziehl-Neelsen stained tissue. 2012 Presented at: 2012 International Conference on Biomedical Engineering (ICoBE); February 27-28, 2012; Penang, Malaysia p. 139-143. [doi: 10.1109/icobe.2012.6178971]
34. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017 Aug;284(2):574-582. [doi: 10.1148/radiol.2017162326] [Medline: 28436741]
35. Pasa F, Golkov V, Pfeiffer F, Cremers D, Pfeiffer D. Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. Sci Rep 2019 Apr 18;9(1):6268 [FREE Full text] [doi: 10.1038/s41598-019-42557-4] [Medline: 31000728]
36. Lopes U, Valiati J. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput Biol Med 2017 Oct 01;89:135-143 [FREE Full text] [doi: 10.1016/j.compbiomed.2017.08.001] [Medline: 28800442]
37. Tasci E, Uluturk C, Ugur A. A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. Neural Comput Appl 2021 Jun 07;33(22):15541-15555 [FREE Full text] [doi: 10.1007/s00521-021-06177-2] [Medline: 34121816]
38. Ul Abideen Z, Ghafoor M, Munir K, Saqib M, Ullah A, Zia T, et al. Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks. IEEE Access 2020;8:22812-22825. [doi: 10.1109/access.2020.2970023]
39. Morís DI, de Moura J, Novo J, Ortega M. Unsupervised contrastive unpaired image generation approach for improving tuberculosis screening using chest X-ray images. Pattern Recogn Lett 2022 Dec;164:60-66. [doi: 10.1016/j.patrec.2022.10.026]
40. Mizan M, Hasan M, Hassan S. A comparative study of tuberculosis detection using deep convolutional neural network. 2020 Presented at: 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT); November 28-29, 2020; Dhaka, Bangladesh p. 157-161. [doi: 10.1109/icaict51780.2020.9333464]
41. Elashmawy A, Elamvazuthi I, Ali S, Natarajan E, Paramasivam S. A hybridized pre-processing method for detecting tuberculosis using deep learning. 2020 Presented at: 2020 8th International Conference on Intelligent and Advanced Systems (ICIAS); July 13-15, 2021; Kuching, Malaysia p. 1-5. [doi: 10.1109/icias49414.2021.9642622]
42. Malik H, Anees T, Chaudhry MU, Gono R, Jasiński M, Leonowicz Z, et al. A novel fusion model of hand-crafted features with deep convolutional neural networks for classification of several chest diseases using X-ray images. IEEE Access 2023;11:39243-39268. [doi: 10.1109/access.2023.3267492]
43. Prabakaran J, Selvaraj P. Advance IoT intelligent healthcare system for lung disease classification using ensemble techniques. Comput Syst Sci Eng 2023;46(2):2141-2157. [doi: 10.32604/csse.2023.034210]
44. Thammarach P, Khaengthanyakan S, Vongsurakrai S, Phienphanich P, Pooprasert P, Yaemsuk A, et al. AI chest 4 all. 2020 Presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, QC p. 1229-1233. [doi: 10.1109/embc44109.2020.9175862]
45. Kim K, Lee JH, Je Oh S, Chung MJ. AI-based computer-aided diagnostic system of chest digital tomography synthesis: demonstrating comparative advantage with X-ray-based AI systems. Comput Methods Programs Biomed 2023 Oct;240:107643 [FREE Full text] [doi: 10.1016/j.cmpb.2023.107643] [Medline: 37348439]
46. Abdollahi J, Mahmoudi L. An artificial intelligence system for detecting the types of the epidemic from X-rays. 2022 Presented at: 2022 27th International Computer Conference, Computer Society of Iran (CSICC); February 23-24, 2022; Tehran, Iran p. 1-6. [doi: 10.1109/csicc55295.2022.9780523]
47. Nguyen NH, Hoang V, Bui T, Truong S, Minh T, Nguyen VD, et al. An efficient approach for tuberculosis diagnosis on chest X-ray. 2022 Presented at: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI); March 28-31, 2022; Kolkata, India. [doi: 10.1109/isbi52829.2022.9761426]
48. Saha P. An ensemble CNN-Dempster Shafer based tuberculosis detection from chest X-ray images. 2022 Presented at: 2022 IEEE Calcutta Conference (CALCON); December 10-11, 2022; Kolkata, India p. 228-232. [doi: 10.1109/calcon56258.2022.10060463]

XSL•FO
RenderX

49. Bouali Y, Boucetta I, Bekkouch I, Bouache M, Mazouzi S. An image dataset for lung disease detection and classification. 2021 Presented at: 2021 International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS); December 15-16, 2021; Skikda, Algeriaz p. 1-6. [doi: 10.1109/ictaacs53298.2021.9715227]
50. Huy VTQ, Lin C. An improved DenseNet deep neural network model for tuberculosis detection using chest X-ray images. IEEE Access 2023;11:42839-42849. [doi: 10.1109/access.2023.3270774]
51. Roopa NK, Mamatha GS. An add-on CNN based model for the detection of tuberculosis using chest X-ray images. IJACSA 2023;14(3):113-123. [doi: 10.14569/IJACSA.2023.0140313]
52. Hooda R, Mittal A, Sofat S. A novel ensemble method for PTB classification in CXRs. Wireless Pers Commun 2020 Jan 16;112(2):809-826. [doi: 10.1007/s11277-020-07075-x]
53. Ahmad I, Shin S. A perceptual encryption-based image communication system for deep learning-based tuberculosis diagnosis using healthcare cloud services. Electronics 2022 Aug 11;11(16):2514. [doi: 10.3390/electronics11162514]
54. Nijiati M, Ma J, Hu C, Tuersun A, Abulizi A, Kelimu A, et al. Artificial intelligence assisting the early detection of active pulmonary tuberculosis from chest X-rays: a population-based study. Front Mol Biosci 2022 Apr 8;9:874475 [FREE Full text] [doi: 10.3389/fmolb.2022.874475] [Medline: 35463963]
55. Sharma A, Rani S, Gupta D. Artificial intelligence-based classification of chest X-ray images into COVID-19 and other infectious diseases. Int J Biomed Imaging 2020 Oct 6;2020:8889023-8889010 [FREE Full text] [doi: 10.1155/2020/8889023] [Medline: 33061946]
56. Oltu B, Guney S, Dengiz B, Agildere M. Automated tuberculosis detection using pre-trained CNN and SVM. 2021 Presented at: 2021 44th International Conference on Telecommunications and Signal Processing (TSP); July 26-28, 2021; Brno, Czech Republic p. 92-95. [doi: 10.1109/tsp52935.2021.9522644]
57. Mithra KS, Sam Emmanuel WR. Automated identification of Mycobacterium bacillus from sputum images for tuberculosis diagnosis. SIViP 2019 Jun 5;13(8):1585-1592. [doi: 10.1007/s11760-019-01509-1]
58. Hooda R, Mittal A, Sofat S. Automated TB classification using ensemble of deep architectures. Multimed Tools Appl 2019 Jul 22;78(22):31515-31532. [doi: 10.1007/s11042-019-07984-5]
59. Silva JF, Silva JM, Pinho E, Costa C. 3D-CNN in drug resistance detection and tuberculosis classification. 2017 Presented at: CLEF 2017: Conference and Labs of the Evaluation Forum; September 11-14, 2017; Dublin, Ireland p. 1-10.
60. Xiong Y, Ba X, Hou A, Zhang K, Chen L, Li T. Automatic detection of Mycobacterium tuberculosis using artificial intelligence. J Thorac Dis 2018 Mar;10(3):1936-1940 [FREE Full text] [doi: 10.21037/jtd.2018.01.91] [Medline: 29707349]
61. Panicker RO, Kalmady KS, Rajan J, Sabu M. Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. Biocybernet Biomed Eng 2018;38(3):691-699. [doi: 10.1016/j.bbe.2018.05.007]
62. Hu M, Liu Y, Zhang Y, Guan T, He Y. Automatic detection of tuberculosis bacilli in sputum smear scans based on subgraph classification. 2019 Presented at: 2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE); November 22-24, 2019; Shenzhen, China p. 1-7. [doi: 10.1109/icmipe47306.2019.9098210]
63. Pahar M, Klopper M, Reeve B, Warren R, Theron G, Diacon A, et al. Automatic tuberculosis and COVID-19 cough classification using deep learning. 2022 Presented at: 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET); July 20-22, 2022; Prague, Czech Republic p. 1-9. [doi: 10.1109/icecet55527.2022.9873469]
64. Pahar M, Theron G, Niesler T. Automatic tuberculosis detection in cough patterns using NLP-style cough embeddings. 2022 Presented at: 2022 International Conference on Engineering and Emerging Technologies (ICEET); October 27-28, 2022; Kuala Lumpur, Malaysia p. 1-6. [doi: 10.1109/iceet56468.2022.10007261]
65. Raju M, Aswath A, Kadam A, Pagidimarri V. Automatic detection of tuberculosis using deep learning methods. In: Laha A, editor. Advances in Analytics and Applications. Springer Proceedings in Business and Economics. Singapore: Springer; 2019:119-129.
66. Manivannan K, Sathiamoorthy S. Bird mating optimizer with deep learning-based tuberculosis detection using chest radiographs. IJETT 2023 Feb 28;71(2):341-348. [doi: 10.14445/22315381/ijett-v71i2p236]
67. Karnkawinpong T, Limpiyakorn Y. Chest X-ray analysis of tuberculosis by convolutional neural networks with affine transforms. 2018 Presented at: CSAI '18: Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence; December 8-10, 2018; Shenzhen, China p. 90-93. [doi: 10.1145/3297156.3297251]
68. Rajaraman S, Zamzmi G, Folio L, Alderson P, Antani S. Chest X-ray bone suppression for improving classification of tuberculosis-consistent findings. Diagnostics (Basel) 2021 May 07;11(5):840 [FREE Full text] [doi: 10.3390/diagnostics11050840] [Medline: 34067034]
69. Shelke A, Inamdar M, Shah V, Tiwari A, Hussain A, Chafekar T, et al. Chest X-ray classification using deep learning for automated COVID-19 screening. SN Comput Sci 2021 May 26;2(4):300 [FREE Full text] [doi: 10.1007/s42979-021-00695-5] [Medline: 34075355]
70. Thakur K, Kaur M, Kumar Y. A comprehensive analysis of deep learning-based approaches for prediction and prognosis of infectious diseases. Arch Comput Methods Eng 2023 Jun 08;30(7):1-21 [FREE Full text] [doi: 10.1007/s11831-023-09952-7] [Medline: 37359745]
71. Karnkawinpong T, Limpiyakorn Y. Classification of pulmonary tuberculosis lesion with convolutional neural networks. 2018 Presented at: 2018 11th International Conference on Computer and Electrical Engineering; October 12–14, 2018; Tokyo, Japan. [doi: 10.1088/1742-6596/1195/1/012007]

XSL•FO
RenderX

72. Karaca B, Guney S, Dengiz B, Agildere M. Comparative study for tuberculosis detection by using deep learning. 2021 Presented at: 2021 44th International Conference on Telecommunications and Signal Processing (TSP); July 26-28, 2021; Brno, Czech Republic p. 88-91. [doi: 10.1109/tsp52935.2021.9522634]
73. Owais M, Arsalan M, Mahmood T, Kim YH, Park KR. Comprehensive computer-aided decision support framework to diagnose tuberculosis from chest X-ray images: data mining study. JMIR Med Inform 2020 Dec 07;8(12):e21790 [FREE Full text] [doi: 10.2196/21790] [Medline: 33284119]
74. Jose E, Soto-Diaz R, Madera N, Soto C, Burgos-Florez F, Rodriguez AF, et al. Computer-aided diagnosis for tuberculosis classification with water strider optimization algorithm. Comput Syst Sci Eng 2023;46(2):1337-1353. [doi: 10.32604/csse.2023.035253]
75. Visuña L, Yang D, Garcia-Blas J, Carretero J. Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning. BMC Med Imaging 2022 Oct 15;22(1):178 [FREE Full text] [doi: 10.1186/s12880-022-00904-4] [Medline: 36243705]
76. Xie Y, Wu Z, Han X, Wang H, Wu Y, Cui L, et al. Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs. J Healthc Eng 2020 Aug 24;2020:9205082-9205012 [FREE Full text] [doi: 10.1155/2020/9205082] [Medline: 32908660]
77. Xu T, Yuan Z. Convolution neural network with coordinate attention for the automatic detection of pulmonary tuberculosis images on chest X-rays. IEEE Access 2022;10:86710-86717. [doi: 10.1109/access.2022.3199419]
78. Mahbub MK, Biswas M, Gaur L, Alenezi F, Santosh K. Deep features to detect pulmonary abnormalities in chest X-rays due to infectious diseaseX: COVID-19, pneumonia, and tuberculosis. Inf Sci (N Y) 2022 May;592:389-401 [FREE Full text] [doi: 10.1016/j.ins.2022.01.062] [Medline: 36532848]
79. Heo S, Kim Y, Yun S, Lim S, Kim J, Nam C, et al. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. Int J Environ Res Public Health 2019 Jan 16;16(2):250 [FREE Full text] [doi: 10.3390/ijerph16020250] [Medline: 30654560]
80. Nijiati M, Zhou R, Damaola M, Hu C, Li L, Qian B, et al. Deep learning based CT images automatic analysis model for active/non-active pulmonary tuberculosis differential diagnosis. Front Mol Biosci 2022 Dec 5;9:1086047 [FREE Full text] [doi: 10.3389/fmolb.2022.1086047] [Medline: 36545511]
81. Liu C, Tsai CC, Kuo L, Kuo P, Lee M, Wang J, et al. A deep learning model using chest X-ray for identifying TB and NTM-LD patients: a cross-sectional study. Insights Imaging 2023 Apr 15;14(1):67 [FREE Full text] [doi: 10.1186/s13244-023-01395-9] [Medline: 37060419]
82. Kazemzadeh S, Yu J, Jamshy S, Pilgrim R, Nabulsi Z, Chen C, et al. Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists. Radiology 2023 Jan;306(1):124-137. [doi: 10.1148/radiol.212213] [Medline: 36066366]
83. Showkatian E, Salehi M, Ghaffari H, Reiazi R, Sadighi N. Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. Pol J Radiol 2022 Feb 28;87:e118-e124. [doi: 10.5114/pjr.2022.113435] [Medline: 35280947]
84. Zhou W, Cheng G, Zhang Z, Zhu L, Jaeger S, Lure FYM, et al. Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing. Quant Imaging Med Surg 2022 Apr;12(4):2344-2355 [FREE Full text] [doi: 10.21037/qims-21-676] [Medline: 35371946]
85. Rahman M, Cao Y, Sun X, Li B, Hao Y. Deep pre-trained networks as a feature extractor with XGBoost to detect tuberculosis from chest X-ray. Comput Electr Eng 2021 Jul;93:107252. [doi: 10.1016/j.compeleceng.2021.107252]
86. Gozes O, Greenspan H. Deep feature learning from a hospital-scale chest X-ray dataset with application to TB detection on a small-scale dataset. 2019 Jul Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 4076-4079. [doi: 10.1109/EMBC.2019.8856729]
87. Nguyen Q, Nguyen B, Dao S, Unnikrishnan B, Dhingra R, Ravichandran S, et al. Deep learning models for tuberculosis detection from chest X-ray images. 2019 Presented at: 2019 26th International Conference on Telecommunications (ICT); April 8-10, 2019; Hanoi, Vietnam p. 381-385. [doi: 10.1109/ict.2019.8798798]
88. Mohammad M, Reddy M. Deep transfer learning approach for classification of chest infections in radiographic X-ray images. 2022 Presented at: 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA); September 8, 2022; Gunupur, India p. 1-6. [doi: 10.1109/iccsea54677.2022.9936426]
89. Hooda R, Sofat S, Kaur S, Mittal A, Meriaudeau F. Deep learning: a potential method for tuberculosis detection using chest radiography. 2017 Presented at: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA); September 12-14, 2017; Kuching, Malaysia p. 497-502. [doi: 10.1109/icsipa.2017.8120663]
90. Ullah N, Marzougui M, Ahmad I, Chelloug SA. DeepLungNet: an effective DL-based approach for lung disease classification using CRIs. Electronics 2023 Apr 14;12(8):1860. [doi: 10.3390/electronics12081860]
91. Uçar M. Deep neural network model with Bayesian optimization for tuberculosis detection from X-ray images. Multimed Tools Appl 2023 Apr 05;82(24):36951-36972. [doi: 10.1007/s11042-023-15212-4]
92. Yan C, Wang L, Lin J, Xu J, Zhang T, Qi J, et al. A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. Eur Radiol 2022 Apr 29;32(4):2188-2199 [FREE Full text] [doi: 10.1007/s00330-021-08365-z] [Medline: 34842959]

XSL•FO
RenderX

93. Singh D, Kumar V, Kaur M. Densely connected convolutional networks-based COVID-19 screening model. Appl Intell (Dordr) 2021 Feb 07;51(5):3044-3051 [FREE Full text] [doi: 10.1007/s10489-020-02149-6] [Medline: 34764584]
94. Manh TH, Thanh HL, Duc VN, Nang TD. Detecting tuberculosis from Vietnamese X-ray imaging using transfer learning approach. Comput Mater Contin 2023;74(3):5001-5016. [doi: 10.32604/cmc.2023.033429]
95. Lay J, Cenggoro Binus University TW, Pardamean B, Gozali J, Tanumihardja D. Detection of pulmonary tuberculosis from chest X-ray images using multimodal ensemble method. Commun Math Biol Neurosci 2022;126:1. [doi: 10.28919/cmbn/7776]
96. Duong LT, Le NH, Tran TB, Ngo VM, Nguyen PT. Detection of tuberculosis from chest X-ray images: boosting the performance with vision transformer and transfer learning. Expert Syst Appl 2021 Dec;184:115519. [doi: 10.1016/j.eswa.2021.115519]
97. Setiawan A, Rusydi M. Detection of Mycobacterium tuberculosis using residual neural network. 2022 Presented at: 2022 International Seminar on Intelligent Technology and Its Applications (ISITIA); July 20-21, 2022; Surabaya, Indonesia p. 7-11. [doi: 10.1109/isitia56226.2022.9855300]
98. Tiwari M, Patankar M, Chaurasia V, Shandilya M, Kumar A, Potnis A. Detection of tuberculosis bacilli using deep learning. 2023 Presented at: 2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP); March 4-5, 2023; Bhopal, India p. 492-496. [doi: 10.1109/ihcsp56702.2023.10127220]
99. Meraj SS, Yaakob R, Azman A, Rum SNM, Nazri ASA, Zakaria NF. Detection of pulmonary tuberculosis manifestation in chest X-rays using different convolutional neural network (CNN) models. IJEAT 2019 Oct 30;9(1):2270-2275. [doi: 10.35940/ijeat.a2632.109119]
100. Han D, Chen Y, Li X, Li W, Zhang X, He T, et al. Development and validation of a 3D-convolutional neural network model based on chest CT for differentiating active pulmonary tuberculosis from community-acquired pneumonia. Radiol Med 2023 Jan 27;128(1):68-80 [FREE Full text] [doi: 10.1007/s11547-022-01580-8] [Medline: 36574111]
101. Hwang EJ, Park S, Jin K, Kim JI, Choi SY, Lee JH, et al. Development and validation of a deep learning–based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. Clin Infect Dis 2019 Aug 16;69(5):739-747 [FREE Full text] [doi: 10.1093/cid/ciy967] [Medline: 30418527]
102. Nguyen H, Tran N, Le B. Diagnosing tuberculosis using graph neural network. 2022 Presented at: 2022 14th International Conference on Knowledge and Systems Engineering (KSE); October 19-21, 2022; Nha Trang, Vietnam p. 1-6. [doi: 10.1109/kse56063.2022.9953751]
103. Rasheed J, Alsubai S. A hybrid deep fused learning approach to segregate infectious diseases. Comput Mater Contin 2023;74(2):4239-4259. [doi: 10.32604/cmc.2023.031969]
104. Han D, He T, Yu Y, Guo Y, Chen Y, Duan H, et al. Diagnosis of active pulmonary tuberculosis and community acquired pneumonia using convolution neural network based on transfer learning. Acad Radiol 2022 Oct;29(10):1486-1492. [doi: 10.1016/j.acra.2021.12.025] [Medline: 35063352]
105. Park M, Lee Y, Kim S, Kim Y, Kim SY, Kim Y, et al. Distinguishing nontuberculous mycobacterial lung disease and Mycobacterium tuberculosis lung disease on X-ray images using deep transfer learning. BMC Infect Dis 2023 Jan 19;23(1):32 [FREE Full text] [doi: 10.1186/s12879-023-07996-5] [Medline: 36658559]
106. Prasitpuriprecha C, Jantama SS, Preeprem T, Pitakaso R, Srichok T, Khonjun S, et al. Drug-resistant tuberculosis treatment recommendation, and multi-class tuberculosis detection and classification using ensemble deep learning-based system. Pharmaceuticals (Basel) 2022 Dec 22;16(1):13 [FREE Full text] [doi: 10.3390/ph16010013] [Medline: 36678508]
107. Ayaz M, Shaukat F, Raja G. Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. Phys Eng Sci Med 2021 Mar 18;44(1):183-194 [FREE Full text] [doi: 10.1007/s13246-020-00966-0] [Medline: 33459996]
108. Kotei E, Thirunavukarasu R. Ensemble technique coupled with deep transfer learning framework for automatic detection of tuberculosis from chest X-ray radiographs. Healthcare (Basel) 2022 Nov 21;10(11):2335 [FREE Full text] [doi: 10.3390/healthcare10112335] [Medline: 36421659]
109. Ahmad Hijazi MH, Qi Yang L, Alfred R, Mahdin H, Yaakob R. Ensemble deep learning for tuberculosis detection. IJEECS 2020 Feb 01;17(2):1014. [doi: 10.11591/ijeecs.v17.i2.pp1014-1020]
110. Hijazi MHA, Kieu Tao Hwa S, Bade A, Yaakob R, Saffree Jeffree M. Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images. IJ-AI 2019 Dec 01;8(4):429. [doi: 10.11591/ijai.v8.i4.pp429-435]
111. Alcantara MF, Cao Y, Liu B, Liu C, Zhang N, Zhang P, et al. eRx – a technological advance to speed-up TB diagnostics. Smart Health 2020 May;16:100117. [doi: 10.1016/j.smhl.2020.100117]
112. Griffin T, Chen Q, Sun X, Wang D, Brunette M, Cao Y, et al. eRxNet: a pipeline of convolutional neural networks for tuberculosis screening. 2021 Presented at: 2021 Third International Conference on Transdisciplinary AI (TransAI); September 20-22, 2021; Laguna Hills, CA p. 47-56. [doi: 10.1109/transai51903.2021.00017]
113. Ameen Z, Saleh MA, Altrjman C, Alturjman S, Abdulkadir R. Explainable residual network for tuberculosis classification in the IoT era. 2021 Presented at: 2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT); December 27-28, 2021; Nicosia, Turkey p. 9-12. [doi: 10.1109/fones-aiot54873.2021.00012]
114. Ravi V, Acharya V, Alazab M. A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. Cluster Comput 2023 Jul 19;26(2):1181-1203 [FREE Full text] [doi: 10.1007/s10586-022-03664-6] [Medline: 35874187]

XSL•FO
RenderX

115. Bhandari M, Shahi TB, Siku B, Neupane A. Explanatory classification of CXR images into COVID-19, pneumonia and tuberculosis using deep learning and XAI. Comput Biol Med 2022 Nov;150:106156 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.106156] [Medline: 36228463]
116. Ignatius JLP, Selvakumar S, Paul KGJL, Kailash AB, Keertivaas S, Akarvin Raja Prajan SAJ. Histogram matched chest X-rays based tuberculosis detection using CNN. Comput Syst Sci Eng 2023;44(1):81-97. [doi: 10.32604/csse.2023.025195]
117. Rashid R, Khawaja S, Akram M, Khan A. Hybrid RID network for efficient diagnosis of tuberculosis from chest X-rays. 2018 Presented at: 2018 9th Cairo International Biomedical Engineering Conference (CIBEC); December 20-22, 2018; Cairo, Egypt p. 167-170. [doi: 10.1109/cibec.2018.8641816]
118. Okolo GI, Katsigiannis S, Ramzan N. IEViT: an enhanced vision transformer architecture for chest X-ray image classification. Comput Methods Programs Biomed 2022 Nov;226:107141 [FREE Full text] [doi: 10.1016/j.cmpb.2022.107141] [Medline: 36162246]
119. Munadi K, Muchtar K, Maulina N, Pradhan B. Image enhancement for tuberculosis detection using deep learning. IEEE Access 2020;8:217897-217907. [doi: 10.1109/access.2020.3041867]
120. Zainab YZS, Jameel A, Usman AM. Impact of transfer learning on chest X-ray (CXR) images for tuberculosis classification. 2021 Presented at: 2021 International Conference on Robotics and Automation in Industry (ICRAI); October 26-27, 2021; Rawalpindi, Pakistan p. 1-8. [doi: 10.1109/icrai54018.2021.9651373]
121. Alcantara MF, Cao Y, Liu C, Liu B, Brunette M, Zhang N, et al. Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Perú. Smart Health 2017 Jun;1-2:66-76. [doi: 10.1016/j.smhl.2017.04.003]
122. Abbas A, Abdelsamea M. Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network. 2018 Presented at: 2018 13th International Conference on Computer Engineering and Systems (ICCES); December 18-19, 2018; Cairo, Egypt p. 122-126. [doi: 10.1109/icces.2018.8639200]
123. Jasmine Pemeena Priyadarsini M, Kotecha K, Rajini GK, Hariharan K, Utkarsh Raj K, Bhargav Ram K, et al. Lung diseases detection using various deep learning algorithms. J Healthc Eng 2023 Feb 03;2023:1-13. [doi: 10.1155/2023/3563696] [Medline: 36776955]
124. Shamrat FMJM, Azam S, Karim A, Islam R, Tasnim Z, Ghosh P, et al. LungNet22: a fine-tuned model for multiclass classification and prediction of lung disease using X-ray images. J Pers Med 2022 Apr 24;12(5):680 [FREE Full text] [doi: 10.3390/jpm12050680] [Medline: 35629103]
125. Hamida S, El Gannour O, Cherradi B, Raihani A, Moujahid H, Ouajji H. A novel COVID-19 diagnosis support system using the stacking approach and transfer learning technique on chest X-ray images. J Healthc Eng 2021 Nov 5;2021:1-17 [FREE Full text] [doi: 10.1155/2021/9437538] [Medline: 34777739]
126. Ravin N, Saha S, Schweitzer A, Elahi A, Dako F, Mollura D, et al. Mitigating domain shift in AI-based TB screening with unsupervised domain adaptation. IEEE Access 2022;10:45997-46013. [doi: 10.1109/access.2022.3168680]
127. Harsha Veena RN, Sreeja AN, Reddy KH, Hasmitha V, Lavanya R. Multiclass deep model for diagnosis of COVID-19 using chest X-ray. 2022 Presented at: 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI); April 28-30, 2022; Tirunelveli, India p. 1767-1771. [doi: 10.1109/icoei53556.2022.9777107]
128. Luján-García JE, Villuendas-Rey Y, López-Yáñez I, Camacho-Nieto O, Yáñez-Márquez C. NanoChest-Net: a simple convolutional network for radiological studies classification. Diagnostics (Basel) 2021 Apr 26;11(5):775 [FREE Full text] [doi: 10.3390/diagnostics11050775] [Medline: 33925844]
129. Liu C, Cohen I, Vishinkin R, Haick H. Nanomaterial-based sensor array signal processing and tuberculosis classification using machine learning. JLPEA 2023 May 29;13(2):39. [doi: 10.3390/jlpea13020039]
130. Pattanasuwan C, Chongstitvatana P. Screening TB using deep transfer learning. 2021 Presented at: 2021 25th International Computer Science and Engineering Conference (ICSEC); November 18-20, 2021; Chiang Rai, Thailand p. 330-333. [doi: 10.1109/icsec53205.2021.9684617]
131. Park S, Kim G, Oh Y, Seo JB, Lee SM, Kim JH, et al. Self-evolving vision transformer for chest X-ray diagnosis through knowledge distillation. Nat Commun 2022 Jul 04;13(1):3848 [FREE Full text] [doi: 10.1038/s41467-022-31514-x] [Medline: 35789159]
132. Wong A, Lee JRH, Rahmat-Khah H, Sabri A, Alaref A, Liu H. TB-Net: a tailored, self-attention deep convolutional neural network design for detection of tuberculosis cases from chest X-ray images. Front Artif Intell 2022 Apr 7;5:827299 [FREE Full text] [doi: 10.3389/frai.2022.827299] [Medline: 35464996]
133. Nafisah SI, Muhammad G. Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. Neural Comput Appl 2022 Apr 19:1-21 [FREE Full text] [doi: 10.1007/s00521-022-07258-6] [Medline: 35462630]
134. Tsai A, Zhou Z, Ou Y, Wang I. Tuberculosis detection based on multiple model ensemble in chest X-ray image. 2022 Presented at: 2022 10th International Conference on Orange Technology (ICOT); November 10-11, 2022; Shanghai, China p. 1-4. [doi: 10.1109/icot56925.2022.10008144]
135. Tao Hwa SK, Bade A, Hijazi MHA, Saffree Jeffree M. Tuberculosis detection using deep learning and contrast-enhanced canny edge detected X-ray images. IJ-AI 2020 Dec 01;9(4):713. [doi: 10.11591/ijai.v9.i4.pp713-720]

XSL•FO
RenderX

136. Huang Y, Darr CM, Gangopadhyay K, Gangopadhyay S, Bok S, Chakraborty S. Applications of machine learning tools for ultra-sensitive detection of lipoarabinomannan with plasmonic grating biosensors in clinical samples of tuberculosis. PLoS One 2022 Oct 25;17(10):e0275658 [FREE Full text] [doi: 10.1371/journal.pone.0275658] [Medline: 36282804]
137. Wu J, Bai J, Wang W, Xi L, Zhang P, Lan J, et al. ATBdiscrimination: an in silico tool for identification of active tuberculosis disease based on routine blood test and T-SPOT.TB detection results. J Chem Inf Model 2019 Oct 14;59(11):4561-4568. [doi: 10.1021/acs.jcim.9b00678]
138. Chandra TB, Verma K, Singh BK, Jain D, Netam SS. Automatic detection of tuberculosis related abnormalities in chest X-ray images using hierarchical feature extraction scheme. Expert Syst Appl 2020 Nov;158:113514. [doi: 10.1016/j.eswa.2020.113514]
139. Wu Y, Wang H, Wu F. Automatic classification of pulmonary tuberculosis and sarcoidosis based on random forest. 2017 Presented at: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); October 14-16, 2017; Shanghai, China p. 1-5. [doi: 10.1109/cisp-bmei.2017.8302280]
140. AbuHassan K, Bakhori N, Kusnin N, Azmi U, Tania M, Evans B, et al. Automatic diagnosis of tuberculosis disease based on plasmonic ELISA and color-based image classification. 2017 Presented at: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 11-15, 2017; Jeju, South Korea p. 4512-4515. [doi: 10.1109/embc.2017.8037859]
141. Tania M, Lwin K, Shabut A, Hossain M. Clustering and classification of a qualitative colorimetric test. 2018 Presented at: 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE); August 16-17, 2018; Southend, UK p. 7-11. [doi: 10.1109/iccecome.2018.8658480]
142. Hu X, Wang J, Ju Y, Zhang X, Qimanguli W, Li C, et al. Combining metabolome and clinical indicators with machine learning provides some promising diagnostic markers to precisely detect smear-positive/negative pulmonary tuberculosis. BMC Infect Dis 2022 Aug 25;22(1):707 [FREE Full text] [doi: 10.1186/s12879-022-07694-8] [Medline: 36008772]
143. Bobak CA, Titus AJ, Hill JE. Comparison of common machine learning models for classification of tuberculosis using transcriptional biomarkers from integrated datasets. Appl Soft Comput 2019 Jan;74:264-273. [doi: 10.1016/j.asoc.2018.10.005]
144. Pholo MD, Hamam Y, Khalaf AB, Du C. Differentiating between COVID-19 and tuberculosis using machine learning and natural language processing. RIA 2022 Apr 30;36(2):313-318. [doi: 10.18280/ria.360216]
145. Ullah R, Khan S, Ali Z, Ali H, Ahmad A, Ahmed I. Evaluating the performance of multilayer perceptron algorithm for tuberculosis disease Raman data. Photodiagn Photodyn Ther 2022 Sep;39:102924. [doi: 10.1016/j.pdpdt.2022.102924] [Medline: 35609805]
146. Beccaria M, Bobak C, Maitshotlo B, Mellors TR, Purcaro G, Franchina FA, et al. Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. J Breath Res 2018 Nov 05;13(1):016005 [FREE Full text] [doi: 10.1088/1752-7163/aae80e] [Medline: 30394364]
147. Govindarajan S, Swaminathan R. Extreme learning machine based differentiation of pulmonary tuberculosis in chest radiographs using integrated local feature descriptors. Comput Methods Programs Biomed 2021 Jun;204:106058. [doi: 10.1016/j.cmpb.2021.106058] [Medline: 33789212]
148. Hu X, Liao S, Bai H, Wu L, Wang M, Wu Q, et al. Integrating exosomal microRNAs and electronic health data improved tuberculosis diagnosis. EBioMedicine 2019 Feb;40:564-573 [FREE Full text] [doi: 10.1016/j.ebiom.2019.01.023] [Medline: 30745169]
149. Xing Z, Ding W, Zhang S, Zhong L, Wang L, Wang J, et al. Machine learning-based differentiation of nontuberculous mycobacteria lung disease and pulmonary tuberculosis using CT images. Biomed Res Int 2020 Sep 30;2020:1-10 [FREE Full text] [doi: 10.1155/2020/6287545] [Medline: 33062689]
150. Ketchanji Mougang YC, Endale Mangamba L, Capuano R, Ciccacci F, Catini A, Paolesse R, et al. On-field test of tuberculosis diagnosis through exhaled breath analysis with a gas sensor array. Biosensors (Basel) 2023 May 22;13(5):570 [FREE Full text] [doi: 10.3390/bios13050570] [Medline: 37232931]
151. Rashidi HH, Khan IH, Dang LT, Albahra S, Ratan U, Chadderwala N, et al. Prediction of tuberculosis using an automated machine learning platform for models trained on synthetic data. J Pathol Inform 2022;13:10 [FREE Full text] [doi: 10.4103/jpi.jpi_75_21] [Medline: 35136677]
152. Chen J, Wu L, Lv Y, Liu T, Guo W, Song J, et al. Screening of long non-coding RNAs biomarkers for the diagnosis of tuberculosis and preliminary construction of a clinical diagnosis model. Front Microbiol 2022 Mar 3;13:774663 [FREE Full text] [doi: 10.3389/fmicb.2022.774663] [Medline: 35308365]
153. Maulana R, Kurniawan D, Ichsan M. Smart devices for self-diagnosing of lung condition based on body temperature and fingernail color. 2020 Presented at: 2020 2nd World Symposium on Artificial Intelligence (WSAI); June 27-29, 2020; Guangzhou, China p. 113-117. [doi: 10.1109/wsai49636.2020.9143283]
154. Urooj S, Suchitra S, Krishnasamy L, Sharma N, Pathak N. Stochastic learning-based artificial neural network model for an automatic tuberculosis detection system using chest X-ray images. IEEE Access 2022;10:103632-103643. [doi: 10.1109/access.2022.3208882]
155. Fonseca A, Rocha B, Nogueira E, Vieira G, Fernandes D, Lima J, et al. Tuberculosis detection in chest radiography: a combined approach of local binary pattern features and monarch butterfly optimization algorithm. 2022 Presented at: 2022

XSL•FO
RenderX

IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC); June 27-July 1, 2022; Los Alamitos, CA p. 1408-1413. [doi: 10.1109/compsac54236.2022.00223]
156. Omisore MO, Samuel OW, Atajeromavwo EJ. A genetic-neuro-fuzzy inferential model for diagnosis of tuberculosis. Appl Comput Inform 2017 Jan;13(1):27-37. [doi: 10.1016/j.aci.2015.06.001]
157. Sharma A, Sharma A, Malhotra R, Singh P, Chakrabortty RK, Mahajan S, et al. An accurate artificial intelligence system for the detection of pulmonary and extra pulmonary tuberculosis. Tuberculosis (Edinb) 2021 Dec;131:102143. [doi: 10.1016/j.tube.2021.102143] [Medline: 34794086]
158. Vijayaraj M, Abhinand PA, Venkatesan P, Ragunath PK. An ANN model for the differential diagnosis of tuberculosis and sarcoidosis. Bioinformation 2020 Jul 31;16(7):539-546 [FREE Full text] [doi: 10.6026/97320630016539] [Medline: 32994679]
159. Melendez J, Sánchez CI, Philipsen RHHM, Maduskar P, Dawson R, Theron G, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. Sci Rep 2016 Apr 29;6(1):25265 [FREE Full text] [doi: 10.1038/srep25265] [Medline: 27126741]
160. Khan S, Ullah R, Shahzad S, Anbreen N, Bilal M, Khan A. Analysis of tuberculosis disease through Raman spectroscopy and machine learning. Photodiagnosis Photodyn Ther 2018 Dec;24:286-291. [doi: 10.1016/j.pdpdt.2018.10.014] [Medline: 30359757]
161. Rosli FA, Mashor MY, Noor SSM. An automated intelligent identification and counting system procedure for tuberculosis. J Phys Conf Series 2019 Nov;1372(1):012079. [doi: 10.1088/1742-6596/1372/1/012079]
162. Inbaraj XA, Villavicencio C, Macrohon JJ, Jeng J, Hsieh J. A novel machine learning approach for tuberculosis segmentation and prediction using chest-X-ray (CXR) images. Appl Sci 2021 Sep 28;11(19):9057. [doi: 10.3390/app11199057]
163. Xu S, Yuan H. A three-methylation-driven gene-based deep learning model for tuberculosis diagnosis in patients with and without human immunodeficiency virus co-infection. Microbiol Immunol 2022 Jun 27;66(6):317-323 [FREE Full text] [doi: 10.1111/1348-0421.12983] [Medline: 35510555]
164. Saif AFM, Imtiaz T, Shahnaz C, Zhu W, Ahmad MO. Exploiting cascaded ensemble of features for the detection of tuberculosis using chest radiographs. IEEE Access 2021;9:112388-112399. [doi: 10.1109/access.2021.3102077]
165. Wang L, Zhang X, Tang J, Ma Z, Usman M, Liu Q, et al. Machine learning analysis of SERS fingerprinting for the rapid determination of Mycobacterium tuberculosis infection and drug resistance. Comput Struct Biotechnol J 2022;20:5364-5377 [FREE Full text] [doi: 10.1016/j.csbj.2022.09.031] [Medline: 36212533]
166. Khatibi T, Shahsavari A, Farahani A. Proposing a novel multi-instance learning model for tuberculosis recognition from chest X-ray images based on CNNs, complex networks and stacked ensemble. Phys Eng Sci Med 2021 Mar 22;44(1):291-311. [doi: 10.1007/s13246-021-00980-w] [Medline: 33616887]
167. Chithra RS, Jagatheeswari P. Severity detection and infection level identification of tuberculosis using deep learning. Int J Imaging Syst Tech 2020 Apr 06;30(4):994-1011. [doi: 10.1002/ima.22427]
168. Feng B, Chen X, Chen Y, Lu S, Liu K, Li K, et al. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. Eur Radiol 2020 Dec 27;30(12):6497-6507. [doi: 10.1007/s00330-020-07024-z] [Medline: 32594210]
169. Kant S, Srivastava M. Towards automated tuberculosis detection using deep learning. 2018 Presented at: 2018 IEEE Symposium Series on Computational Intelligence (SSCI); November 18-21, 2018; Bangalore, India p. 1250-1253. [doi: 10.1109/ssci.2018.8628800]
170. Laeli A, Rustam Z, Pandelaki J. Tuberculosis detection based on chest X-rays using ensemble method with CNN feature extraction. 2021 Presented at: 2021 International Conference on Decision Aid Sciences and Application (DASA); December 7-8, 2021; Sakheer, Bahrain p. 682-686. [doi: 10.1109/dasa53625.2021.9682237]
171. Huang C, Wang W, Zhang X, Wang S, Zhang Y. Tuberculosis diagnosis using deep transferred EfficientNet. IEEE/ACM Trans Comput Biol Bioinf 2023 Sep 1;20(5):2639-2646. [doi: 10.1109/tcbb.2022.3199572]
172. Chang R, Chiu Y, Lin J. Two-stage classification of tuberculosis culture diagnosis using convolutional neural network with transfer learning. J Supercomput 2020 Jan 16;76(11):8641-8656. [doi: 10.1007/s11227-020-03152-x]
173. Rajaraman S, Candemir S, Xue Z, Alderson PO, Kohli M, Abuya J, et al. A novel stacked generalization of models for improved TB detection in chest radiographs. 2018 Presented at: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 18-21, 2018; Honolulu, HI p. 718-721. [doi: 10.1109/EMBC.2018.8512337]
174. Rashidi HH, Dang LT, Albahra S, Ravindran R, Khan IH. Automated machine learning for endemic active tuberculosis prediction from multiplex serological data. Sci Rep 2021 Sep 09;11(1):17900 [FREE Full text] [doi: 10.1038/s41598-021-97453-7] [Medline: 34504228]
175. Pahar M, Klopper M, Reeve B, Warren R, Theron G, Niesler T. Automatic cough classification for tuberculosis screening in a real-world environment. Physiol Meas 2021 Nov 26;42(10):105014 [FREE Full text] [doi: 10.1088/1361-6579/ac2fb8] [Medline: 34649231]
176. Wang M, Lee C, Wei Z, Ji H, Yang Y, Yang C. Clinical assistant decision-making model of tuberculosis based on electronic health records. BioData Min 2023 Mar 16;16(1):11 [FREE Full text] [doi: 10.1186/s13040-023-00328-y] [Medline: 36927471]

XSL•FO
RenderX

177. Fati SM, Senan EM, ElHakim N. Deep and hybrid learning technique for early detection of tuberculosis based on X-ray images using feature fusion. Appl Sci 2022 Jul 14;12(14):7092. [doi: 10.3390/app12147092]
178. Nahiduzzaman M, Faruq Goni MO, Robiul Islam M, Sayeed A, Shamim Anower M, Ahsan M, et al. Detection of various lung diseases including COVID-19 using extreme learning machine algorithm based on the features extracted from a lightweight CNN architecture. Biocybern Biomed Eng 2023 Jun 26;43(3):528-550 [FREE Full text] [doi: 10.1016/j.bbe.2023.06.003] [Medline: 38620111]
179. Patel M, Das A, Pant V. Detection of tuberculosis in radiographs using deep learning-based ensemble methods. 2021 Presented at: 2021 Smart Technologies, Communication and Robotics (STCR); October 9-10, 2021; Sathyamangalam, India p. 1-7. [doi: 10.1109/stcr51658.2021.9588936]
180. Mehrrotraa R, Ansari MA, Agrawal R, Tripathi P, Bin Heyat MB, Al-Sarem M, et al. Ensembling of efficient deep convolutional networks and machine learning algorithms for resource effective detection of tuberculosis using thoracic (chest) radiography. IEEE Access 2022;10:85442-85458. [doi: 10.1109/access.2022.3194152]
181. Lee JH, Park S, Hwang EJ, Goo JM, Lee WY, Lee S, et al. Deep learning-based automated detection algorithm for active pulmonary tuberculosis on chest radiographs: diagnostic performance in systematic screening of asymptomatic individuals. Eur Radiol 2021 Feb 28;31(2):1069-1080. [doi: 10.1007/s00330-020-07219-4] [Medline: 32857202]
182. Zeng G. On the confusion matrix in credit scoring and its analytical properties. Commun Stat - Theory Methods 2019 Feb 07;49(9):2080-2093. [doi: 10.1080/03610926.2019.1568485]
183. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017 Nov 20;5:307 [FREE Full text] [doi: 10.3389/fpubh.2017.00307] [Medline: 29209603]
184. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Contin Educ Anaesth Crit Care Pain 2008 Dec 12;8(6):221-223. [doi: 10.1093/bjaceaccp/mkn041]
185. Islam Z, Abdel-Aty M, Cai Q, Yuan J. Crash data augmentation using variational autoencoder. Accid Anal Prev 2021 Mar;151:105950. [doi: 10.1016/j.aap.2020.105950] [Medline: 33370603]
186. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019 Jul 6;6(1):60. [doi: 10.1186/s40537-019-0197-0]
187. Lin N, Guo S, Xie L. TB-Transformer: integrating mouse trace with object bounding-box for image caption. 2022 Presented at: 2022 IEEE 24th International Conference on High Performance Computing & Communications; 8th International Conference on Data Science & Systems; 20th International Conference on Smart City; 8th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys); December 18-20, 2022; Hainan, China p. 810-814. [doi: 10.1109/hpcc-dss-smartcity-dependsys57074.2022.00133]
188. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Preprint posted online 2014. [doi: 10.48550/arxiv.1409.1556]
189. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV p. 770-778. [doi: 10.1109/cvpr.2016.90]
190. Colombo FM, Mello GR, Andrade BF, de OL, Kritski A, Koenigkam SM, et al. Preliminary results on pulmonary tuberculosis detection in chest X-ray using convolutional neural networks. 2020 Presented at: ICCS 2020: 20th International Conference on Computational Science; June 3–5, 2020; Amsterdam, The Netherlands p. 563-576. [doi: 10.1007/978-3-030-50423-6_42]
191. Huang G, Liu Z, Van DML, Weinberger K. Densely connected convolutional networks. 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI p. 2261-2269. [doi: 10.1109/cvpr.2017.243]
192. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. arXiv Preprint posted online 2019 [FREE Full text] [doi: 10.1007/978-1-4842-6168-2_10]
193. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. OpenReview. 2021 Jan 12. URL: https://openreview.net/forum?id=YicbFdNTTy [accessed 2025-02-22]
194. Liu Z, Mao H, Wu C, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. 2022 Presented at: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 18-24, 2022; New Orleans, LA p. 11966-11976. [doi: 10.1109/cvpr52688.2022.01167]
195. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion 2023 Nov;99:101805. [doi: 10.1016/j.inffus.2023.101805]
196. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Kůrková V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I, editors. Lecture Notes in Computer Science (Volume 11141). Cham: Springer; 2018:270-279.
197. Hansun S, Argha A, Alinejad-Rokny H, Alizadehsani R, Gorriz JM, Liaw S, et al. A new ensemble transfer learning approach with rejection mechanism for tuberculosis disease detection. IEEE Trans Radiat Plasma Med Sci 2024:1. [doi: 10.1109/trpms.2024.3474708]
198. Cohen JP, Hashir M, Brooks R, Bertrand H. On the limits of cross-domain generalization in automated X-ray prediction. 2020 Presented at: Third Conference on Medical Imaging with Deep Learning; July 6-8, 2020; Montreal, QC p. 136-155.

XSL•FO
RenderX

Abbreviations

ACM: Association for Computing Machinery
AI: artificial intelligence
AUC: area under the curve
AUROC: area under the receiver operating characteristic curve
CAD: computer-aided detection
ChatGPT: Chat Generative Pretrained Transformer
CNN: convolutional neural network
CV: cross-validation
CXR: chest X-ray
DL: deep learning
MC: Montgomery County
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PROSPERO: Prospective Register of Systematic Reviews
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies version 2
ROC: receiver operating characteristic
SZ: Shenzhen
TB: tuberculosis
TL: transfer learning
VGG: Visual Geometry Group
ViT: Vision Transformer
WHO: World Health Organization
XAI: explainable AI

Edited by A Mavragani; submitted 21.11.24; peer-reviewed by Z Zhang, C Ordun; comments to author 24.12.24; revised version received 10.01.25; accepted 07.02.25; published 07.03.25.

Please cite as:
Hansun S, Argha A, Bakhshayeshi I, Wicaksana A, Alinejad-Rokny H, Fox GJ, Liaw ST, Celler BG, Marks GB
Diagnostic Performance of Artificial Intelligence–Based Methods for Tuberculosis Detection: Systematic Review
J Med Internet Res 2025;27:e69068
URL: https://www.jmir.org/2025/1/e69068
doi:10.2196/69068
PMID:40053773

©Seng Hansun, Ahmadreza Argha, Ivan Bakhshayeshi, Arya Wicaksana, Hamid Alinejad-Rokny, Greg J Fox, Siaw-Teng Liaw, Branko G Celler, Guy B Marks. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Application of Machine Learning for Patients With Cardiac Arrest: Systematic Review and Meta-Analysis

Shengfeng Wei[1*], BD; Xiangjian Guo[1*], MM; Shilin He[1*], BD; Chunhua Zhang[1], BD; Zhizhuan Chen[1], BD; Jianmei Chen[1], MM; Yanmei Huang[1], MM; Fan Zhang[1*], BD; Qiangqiang Liu[1*], MM

Department of Emergency Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Qiangqiang Liu, MM
Department of Emergency Medicine
The First Affiliated Hospital, Sun Yat-sen University
58 Zhongshan 2nd Road, Yuexiu District
Guangzhou, 510000
China
Phone: 86 18928825921
Email: liuqq9@mail.sysu.edu.cn

Abstract

Background: Currently, there is a lack of effective early assessment tools for predicting the onset and development of cardiac arrest (CA). With the increasing attention of clinical researchers on machine learning (ML), some researchers have developed ML models for predicting the occurrence and prognosis of CA, with certain models appearing to outperform traditional scoring tools. However, these models still lack systematic evidence to substantiate their efficacy.

Objective: This systematic review and meta-analysis was conducted to evaluate the prediction value of ML in CA for occurrence, good neurological prognosis, mortality, and the return of spontaneous circulation (ROSC), thereby providing evidence-based support for the development and refinement of applicable clinical tools.

Methods: PubMed, Embase, the Cochrane Library, and Web of Science were systematically searched from their establishment until May 17, 2024. The risk of bias in all prediction models was assessed using the Prediction Model Risk of Bias Assessment Tool.

Results: In total, 93 studies were selected, encompassing 5,729,721 in-hospital and out-of-hospital patients. The meta-analysis revealed that, for predicting CA, the pooled C-index, sensitivity, and specificity derived from the imbalanced validation dataset were 0.90 (95% CI 0.87-0.93), 0.83 (95% CI 0.79-0.87), and 0.93 (95% CI 0.88-0.96), respectively. On the basis of the balanced validation dataset, the pooled C-index, sensitivity, and specificity were 0.88 (95% CI 0.86-0.90), 0.72 (95% CI 0.49-0.95), and 0.79 (95% CI 0.68-0.91), respectively. For predicting the good cerebral performance category score 1 to 2, the pooled C-index, sensitivity, and specificity based on the validation dataset were 0.86 (95% CI 0.85-0.87), 0.72 (95% CI 0.61-0.81), and 0.79 (95% CI 0.66-0.88), respectively. For predicting CA mortality, the pooled C-index, sensitivity, and specificity based on the validation dataset were 0.85 (95% CI 0.82-0.87), 0.83 (95% CI 0.79-0.87), and 0.79 (95% CI 0.74-0.83), respectively. For predicting ROSC, the pooled C-index, sensitivity, and specificity based on the validation dataset were 0.77 (95% CI 0.74-0.80), 0.53 (95% CI 0.31-0.74), and 0.88 (95% CI 0.71-0.96), respectively. In predicting CA, the most significant modeling variables were respiratory rate, blood pressure, age, and temperature. In predicting a good cerebral performance category score 1 to 2, the most significant modeling variables in the in-hospital CA group were rhythm (shockable or nonshockable), age, medication use, and gender; the most significant modeling variables in the out-of-hospital CA group were age, rhythm (shockable or nonshockable), medication use, and ROSC.

Conclusions: ML represents a currently promising approach for predicting the occurrence and outcomes of CA. Therefore, in future research on CA, we may attempt to systematically update traditional scoring tools based on the superior performance of ML in specific outcomes, achieving artificial intelligence–driven enhancements.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42024518949; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=518949

XSL•FO
RenderX

(J Med Internet Res 2025;27:e67871) doi:10.2196/67871

KEYWORDS

cardiac arrest; machine learning; prognosis; systematic review; artificial intelligence; AI

Introduction

Background

Cardiac arrest (CA) remains a critical challenge in contemporary medicine, characterized by a dismally low survival rate and poor prognosis, and, therefore, has garnered global attention [1]. CA can be classified by the occurrence location into in-hospital CA (IHCA) and out-of-hospital CA (OHCA). Despite advancements in cardiopulmonary resuscitation techniques, global registry data indicate that the incidence and survival rates of CA have not significantly improved. The incidence of IHCA in the United States increased to 900 to 1000 per 100,000 hospitalized patients between 2008 and 2017, compared to 160 per 100,000 in the United Kingdom from 2011 to 2013 and 840 per 100,000 in China as of 2020 [2-4]. Meanwhile, the estimated averages of incidence of OHCA under emergency medical services (EMS) in North America, Asia, and Europe from 2010 to 2020 were 47.3, 45.9, and 40.6 per 100,000 people, respectively. The estimated averages of the survival rates of IHCA from 2010 to 2020 were 25% in the United States, 18% in the United Kingdom, and only 9.4% in China. For OHCA, the estimated averages of the survival rates during this same period were 10% to 12% in the United States, 8% in Europe, and just 3.6% in Asia [2,5]. These low survival rates also impose significant economic burdens on nations. According to relevant reviews, the cost-effectiveness threshold for CA ranged from US $20,000 to US $150,000 per quality-adjusted life year, with each life saved potentially reducing costs by US $19,000 to US $71,000 per case [6].

Although efforts to establish CA centers independently began in various regions of the United States as early as 2000 to 2010 [7] and Germany initiated CA center certification throughout the country [8] in August 2019 aiming to provide evidence-based, bundled care to improve CA survival rates, CA remains a formidable clinical challenge. If resuscitation is not timely, the patient may lose consciousness within approximately 10 seconds, with irreversible hypoxic-ischemic brain injury occurring within 4 minutes [9], and if resuscitation is delayed beyond 10 minutes, survival is practically impossible [10,11]. Thus, early prediction and identification of CA are critical factors in preventing death and poor outcomes and represent a major challenge that requires urgent clinical attention.

Objectives

However, there is a scarcity of efficient, internationally recognized, and universally accepted assessment tools for early prediction and identification of CA risk and adverse outcomes. In recent years, with the rapid advancement of artificial intelligence (AI), many researchers have used machine learning (ML) to address clinical challenges. Commonly used ML approaches can be broadly categorized into supervised and unsupervised learning. In the context of supervised ML, clinical predictors can be incorporated into various models. In these models, their parameters are adjusted based on outcome variables to generate predictions regarding the probability of positive event occurrence [12]. It is now common to see ML being used to predict disease progression and even to diagnose and treat complex diseases effectively. For instance, in 2019, several authors, including Hatib et al [13] and Wijnberge et al [14], successfully predicted intraoperative hypotensive events using ML, leading to the clinical translation of these models into products that significantly enhanced patient safety during surgical anesthesia [15]. By 2023, some researchers had similarly affirmed the substantial potential of ML models in cancer detection, prognosis, and treatment, recognizing their exciting discoveries and contributions to advancing medical practice [16]. The aforementioned studies were based on supervised ML and the extensive use of interpretable clinical features to construct predictive models and simultaneously demonstrate the promising predictive performance of ML in clinical events across various fields. In this context, some researchers have also developed different ML models for risk prediction in CA. Recent reviews by Sem et al [17] and Chen et al [18] indicate that ML appears to exhibit high accuracy in both the management and risk prediction of CA. However, these reviews do not quantitatively synthesize the results of ML models, which significantly limits our ability to interpret the specific value of various ML models in CA applications and the selection of appropriate models. Therefore, we conducted this systematic review and meta-analysis to review the predictive performance of ML for the occurrence of CA, good neurological prognosis after CA, mortality, and the return of spontaneous circulation (ROSC) after CA to provide evidence-based guidance for the development and updating of simple prediction tools with high accuracy and direct access to results.

Methods

Study Registration

This study was conducted in adherence to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and prospectively registered with PROSPERO (International Prospective Register of Systematic Reviews; ID CRD42024518949). The detailed PRISMA checklist is presented in Multimedia Appendix 1.

Eligibility Criteria

Detailed inclusion and exclusion criteria were defined to screen the original studies relevant to our systematic review from the retrieved literature (Textbox 1).

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria for original studies.

Inclusion criteria

- Study type: the included studies must be case-control, cohort, nested case-control, case-cohort, or cross-sectional studies.
- Model construction: although some studies, due to limited sample sizes, lacked independent external validation, we could not dismiss their contributions. In our analysis, it was necessary to synthesize results from the training and validation sets to assess the presence of severe overfitting. Therefore, those with no external validation were also included.
- Outcomes: studies that comprehensively constructed machine learning (ML) models for cardiac arrest (CA) occurrence prediction or clinical outcomes following CA were selected.
- Language: we included original studies in English.

Exclusion criteria

- Study type: studies categorized as meta-analyses, reviews, guidelines, expert opinions, or conference abstracts and not fully peer reviewed and published were removed.
- Model construction: studies with only risk factor analysis but no construction of a complete ML model were excluded, those with a limited number of samples (<20) were not included, and those only focusing on the accuracy of univariate predictors were removed.
- Outcomes: in existing ML studies, model performance was assessed using the receiver operating characteristic curve, C-statistic, sensitivity, specificity, accuracy, recall, precision, confusion matrix, or F_1-score. However, a few original studies that lacked at least one of these metrics and, therefore, did not evaluate model performance adequately were excluded.
- Language: non–English-language original studies were excluded.

Data Sources and Search Strategy

A systematic search of the PubMed, Embase, Cochrane Library, and Web of Science databases was carried out from their inception to May 17, 2024. The search strategy involved controlled vocabulary and free-text terms, with no restrictions on geographical location or publication year. The detailed search strategy is presented in Multimedia Appendices 2-5.

Study Selection and Data Extraction

The retrieved studies were imported into EndNote X9 (Clarivate Analytics). Their titles and abstracts were reviewed. After the exclusion of duplicates, the preliminary eligible original studies were selected and their full texts downloaded for determining the final inclusion. An electronic spreadsheet was prepared to extract the following information: first author, publication year, author's country, study type, patient source, prediction events, data balance, location of CA occurrence, number of cases with study events, total number of cases, number of cases in the training and validation sets, method of validation set generation, missing data–handling methods, and types of models used. Study selection and data extraction were independently conducted by 2 researchers. Disagreements were discussed and resolved with a third author.

Risk of Bias in the Studies

The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias in all the included original studies. PROBAST comprises several questions across 4 domains—participant, predictor, outcome, and statistical analysis—which reflect the overall risk of bias and applicability. These domains consist of 2, 3, 6, and 9 questions, respectively, each with 3 possible answers (*Yes* or *Probablyyes*, *No* or *Probably no*, and *No information*). A domain was classified as high risk if any question was answered with *No* or *Probably no*. Conversely, a domain was regarded as low risk only if every question was answered with *Yes* or *Probably yes*. The overall risk of bias was assessed as low when all domains were deemed to be low risk. When at least one domain was high risk, the overall risk of bias was rated as high. In total, 2 authors independently assessed the risk of bias using PROBAST and cross-checked their findings. Any discrepancies were addressed by consulting with a third author to reach agreement.

Outcomes

The primary outcome measure was the C-index, which reflects the predictive ability of ML models for IHCA and OHCA. However, we found that the C-index might not have accurately described the predictive performance of ML for positive events, particularly in models built on severely imbalanced data, as these original studies often suffered from such imbalance. This limitation was evident in predicting the occurrence of IHCA and OHCA, the good cerebral performance category score 1 to 2 (CPC 1-2), mortality, and ROSC. Therefore, in addition to the C-index, our primary outcome measures encompassed sensitivity and specificity. Our secondary outcome was the frequency of variables used in the ML models.

Synthesis Methods

A meta-analysis of the C-index, a measure of the general accuracy of ML models, was carried out. When the 95% CI and SE for the C-index were missing in some studies, the SE was estimated based on the study by Debray et al [19]. Due to the differences in the included variables and inconsistent parameters among the ML models, random-effects models were prioritized in the meta-analysis of the C-index. In addition, a meta-analysis on sensitivity and specificity was conducted through a bivariate mixed-effects model based on diagnostic 2 × 2 tables. However, most original studies did not report these tables. In such cases, we used the following methods to calculate the 2 × 2 tables: (1) calculation based on sensitivity, specificity, precision, and case numbers; and (2) calculation based on the best Youden index to extract sensitivity and specificity, followed by case number

XSL•FO
RenderX

integration. Nevertheless, this method allowed for meta-analysis only when there were ≥4 models. For <4 models, we presented the range of sensitivity and specificity. Our meta-analysis was conducted in R (version 4.2.0; R Foundation for Statistical Computing).

Results

Study Selection

A total of 1270 articles were obtained from databases, with 599 (47.17%) being duplicates. Among these 599 duplicates, 471 (78.6%) were found to be duplicates via software, and 128 (21.4%) were manually identified as duplicates. After the elimination of duplicates, 671 articles were screened by title and abstract, with 169 (25.2%) being selected for full-text review. After the exclusion of conference abstracts published in full text without peer review (19/169, 11.2%), studies with risk factor analyses but no complete ML models (22/169, 13%), studies lacking outcome indicators (28/169, 16.6%), and studies with severe statistical errors (7/169, 4.1%), a total of 93 articles were included finally. The detailed process is illustrated in Figure 1.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for study selection.

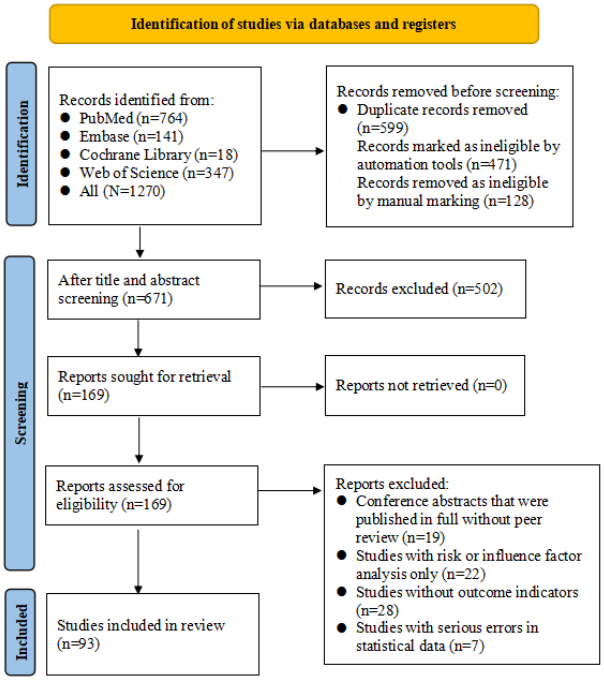

Study Characteristics

The 93 selected studies were published between 2011 and 2024, covering 14 countries, primarily South Korea, China, Japan, the United States, and Singapore. Among the 93 studies, there were 23 (25%) prospective cohort studies and 3 (3%) case-control studies, with the remainder (67/93, 72%) being retrospective cohort studies. Data for 26% (24/93) of the studies were sourced from multiple centers, whereas 37% (34/93) of the studies used data from registry databases and the rest (35/93, 38%) were single-center studies. In 30% (28/93) of the studies, the predicted outcome was the occurrence of CA. In 42 studies, the predicted outcome was the neurological prognosis of patients with CA, with 10 (24%) studies focused on patients with IHCA and the remainder (n=32, 76%) focused on patients with OHCA. In 27% (25/93) of the studies, the predicted outcome was CA mortality, and in 12% (11/93) of the studies, the predicted outcome was ROSC in patients with CA. The 93 studies collectively encompassed a total of 5,729,721 cases, including 1,737,085 OHCA cases and 3,992,636 IHCA cases. Regarding the predictive models constructed, 81% (75/93) of the studies had independent validation sets, but only 27% (25/93) used external validation, primarily using k-fold cross-validation and random-sampling internal validation methods. A total of 34% (32/93) of the studies described methods to prevent data overfitting, mainly through cross-validation. In total, 17 types of ML models were involved, with logistic regression (LR), random forest (RF), deep learning, and decision trees (DTs) being the most prominent. In addition, these studies validated several previously established scoring tools, including the Cardiac Arrest Neurological Prognosis score, distance scoring system, Emergency Department In-Hospital Cardiac Arrest Score, FACTOR score, Modified Early Warning Score, National Early Warning Score, National Early Warning Score 2, OHCA score, proposed scoring system, Rapid Emergency Medicine Score, Simplified Acute Physiology Score II, Cardiac Arrest Hospital Prognosis score, and ROSC after CA score. The details of the included studies are shown in Tables 1 and 2.

XSL•FO
RenderX

Table 1. Characteristics of the included studies.

Study	Year of publication	Country of first author	Study type (case-control, cohort study [retrospective or prospective], nested cohort study, or case-cohort study)	Patient sources (single center, multicenter, or registration database)	Predictive events
Wang et al [20]	2024	Taiwan, China	Retrospective cohort study	Multicenter	Cardiac arrest
Raheem et al [21]	2024	Pakistan	Retrospective cohort study	Single center	Cardiac arrest
Amacher et al [22]	2024	Switzerland	Prospective cohort study	Single center	In-hospital mortality and CPC 3-5[a]
Cho et al [23]	2024	Republic of Korea	Retrospective cohort study	Single center	Cardiac arrest
Shin et al [24]	2024	Republic of Korea	Retrospective cohort study	Single center	Cardiac arrest
Ding et al [25]	2024	China	Prospective cohort study	Single center	CPC 1-2[b] and in-hospital mortality
Nishioka et al [26]	2024	Japan	Retrospective cohort study	Multicenter	CPC 3-5
Kajino et al [27]	2024	Japan	Retrospective cohort study	Registration database	CPC 1-2
Pham et al [28]	2024	United States	Prospective cohort study	Multicenter	Cardiac arrest
Rahadian et al [29]	2024	Japan	Prospective cohort study	Registration database	VF[c] or VT[d]
Wang et al [30]	2024	Taiwan, China	Retrospective cohort study	Registration database	Cardiac arrest
Tsai et al [31]	2024	Taiwan, China	Retrospective cohort study	Single center	CPC 1-2
Schweiger et al [32]	2024	Switzerland	Prospective cohort study	Single center	In-hospital mortality
Caputo et al [33]	2024	Switzerland	Prospective cohort study	Registration database	ROSC[e]
Lu et al [34]	2023	Taiwan, China	Retrospective cohort study	Single center	Cardiac arrest
Dünser et al [35]	2023	Austria	Retrospective cohort study	Single center	NROSC[f] and CPC 3-5
Bang et al [36]	2023	Republic of Korea	Retrospective cohort study	Multicenter	In-hospital mortality
Zhang et al [37]	2023	China	Retrospective cohort study	Multicenter	In-hospital mortality
Li and Xing [38]	2023	China	Retrospective cohort study	Single center	CPC 3-5 and NROSC
Ding et al [39]	2023	China	Retrospective cohort study	Single center	Cardiac arrest
Uehara et al [40]	2023	Japan	Prospective cohort study	Multicenter	CPC 1-2
Shin et al [41]	2023	Republic of Korea	Retrospective cohort study	Registration database	CPC 1-2 and ROSC
Kawai et al [42]	2023	Japan	Retrospective cohort study	Single center	CPC 3-5
Imamura et al [43]	2023	Japan	Retrospective cohort study	Multicenter	30-day mortality
Hessulf et al [44]	2023	Sweden	Retrospective cohort study	Multicenter	30-day survival
Yoon et al [45]	2023	Republic of Korea	Retrospective cohort study	Single center	CPC 3-5
Chang et al [46]	2023	Republic of Korea	Retrospective cohort study	Registration database	ROSC, survival to discharge, and CPC 1-2
Wang et al [47]	2023	China	Retrospective cohort study	Multicenter	ROSC
Shinada et al [48]	2023	Japan	Retrospective cohort study	Registration database	CPC 1-2
Xu et al [49]	2022	China	Case-control study	Single center	Cardiac arrest
Tsai et al [50]	2022	Taiwan, China	Retrospective cohort study	Registration database	Cardiac arrest
Tang et al [51]	2022	China	Retrospective cohort study	Registration database	Cardiac arrest
Kim et al [52]	2022	Republic of Korea	Retrospective cohort study	Registration database	Cardiac arrest
Chae et al [53]	2022	Republic of Korea	Retrospective cohort study	Single center	Cardiac arrest
Sun et al [54]	2022	Taiwan, China	Retrospective cohort study	Single center	Cardiac arrest
Wong et al [55]	2022	Singapore	Prospective cohort study	Multicenter	Survival to discharge
Tran et al [56]	2022	United States	Prospective cohort study	Registration database	In-hospital mortality

XSL•FO
RenderX

Study	Year of publication	Country of first author	Study type (case-control, cohort study [retrospective or prospective], nested cohort study, or case-cohort study)	Patient sources (single center, multicenter, or registration database)	Predictive events
Rajendram et al [57]	2022	Singapore	Retrospective cohort study	Multicenter	Survival to discharge and CPC 1-2
Rafi et al [58]	2022	France	Retrospective cohort study	Single center	Cardiac arrest
Liu et al [59]	2022	Singapore	Retrospective cohort study	Registration database	ROSC
Lin et al [60]	2022	Taiwan, China	Retrospective cohort study	Registration database	CPC 1-2
Kawai et al [61]	2022	Japan	Retrospective cohort study	Multicenter	CPC 1-2
Itagaki et al [62]	2022	Japan	Retrospective cohort study	Single center	Brain death
Harris et al [63]	2022	United States	Retrospective cohort study	Registration database	Prehospital ROSC in pediatric OHCA[g]
Harford et al [64]	2022	United States	Retrospective cohort study	Registration database	CPC 1-2
Harford et al [65]	2022	United States	Retrospective cohort study	Registration database	CPC 1-2
Chung et al [66]	2021	Taiwan, China	Retrospective cohort study	Single center	CPC 1-2
Chi et al [67]	2021	Taiwan, China	Retrospective cohort study	Registration database	In-hospital mortality
Wang et al [68]	2021	China	Retrospective cohort study	Multicenter	CPC 1-2
Bae et al [69]	2021	Republic of Korea	Retrospective cohort study	Single center	CPC 3-5
Mueller et al [70]	2021	Austria	Prospective cohort study	Single center	CPC 1-2
Lee et al [71]	2021	Republic of Korea	Retrospective cohort study	Multicenter	Cardiac arrest
Lim et al [72]	2021	Republic of Korea	Prospective cohort study	Registration database	CPC 1-2
Lo and Siu [73]	2021	Hong Kong, China	Retrospective cohort study	Registration database	ROSC
Lonsain et al [74]	2021	Belgium	Retrospective cohort study	Single center	24-hour survival
Nishioka et al [75]	2021	Japan	Prospective cohort study	Registration database	CPC 1-2
Beom et al [76]	2021	Republic of Korea	Prospective cohort study	Multicenter	Survival to discharge and CPC 1-2
Cheng et al [77]	2021	Taiwan, China	Retrospective cohort study	Single center	CPC 1-2
Kim et al [78]	2021	Republic of Korea	Retrospective cohort study	Registration database	Survival to discharge and CPC 1-2
Seo et al [79]	2021	Republic of Korea	Prospective cohort study	Registration database	CPC 1-2
Song et al [80]	2021	Republic of Korea	Retrospective cohort study	Single center	CPC 3-5
Sun et al [81]	2021	Hong Kong, China	Retrospective cohort study	Multicenter	ROSC
Youn et al [82]	2021	Republic of Korea	Prospective cohort study	Multicenter	Significant coronary artery disease among survivors of OHCA without STE[h]
Heo et al [83]	2021	Republic of Korea	Prospective cohort study	Multicenter	CPC 3-5
Wang et al [84]	2020	China	Retrospective cohort study	Registration database	CPC 1-2
Hong et al [85]	2020	Republic of Korea	Retrospective cohort study	Single center	Cardiac arrest
Cho et al [86]	2020	Republic of Korea	Retrospective cohort study	Single center	Cardiac arrest
Hirano et al [87]	2020	Japan	Retrospective cohort study	Registration database	Death at 1 month or survival with poor neurological function (CPC 3-5) and 30-day mortality
Okada et al [88]	2020	Japan	Prospective cohort study	Registration database	CPC 1-2
Liu et al [89]	2020	Singapore	Retrospective cohort study	Registration database	ROSC
Elola et al [90]	2020	Spain	Retrospective cohort study	Registration database	Cardiac arrest
Hsieh et al [91]	2020	Taiwan, China	Retrospective cohort study	Registration database	Cardiac arrest

XSL•FO
RenderX

Study	Year of publication	Country of first author	Study type (case-control, cohort study [retrospective or prospective], nested cohort study, or case-cohort study)	Patient sources (single center, multicenter, or registration database)	Predictive events
Baldi et al [92]	2020	Italy	Prospective cohort study	Multicenter	Survival to hospital admission
Li et al [93]	2019	China	Case-control study	Multicenter	Cardiac arrest
Srivilaithon et al [94]	2019	Thailand	Case-control study	Single center	Cardiac arrest
Lee et al [95]	2019	Republic of Korea	Retrospective cohort study	Single center	CPC 3-5
Liu et al [96]	2019	Taiwan, China	Retrospective cohort study	Single center	Cardiac arrest
Jang et al [97]	2019	Republic of Korea	Retrospective cohort study	Single center	Cardiac arrest
Seki et al [98]	2019	Japan	Prospective cohort study	Multicenter	1-year survival
Park et al [99]	2019	Republic of Korea	Retrospective cohort study	Registration database	CPC 1-2
Kwon et al [100]	2019	Republic of Korea	Retrospective cohort study	Registration database	CPC 1-2 and survival to discharge
Kong et al [101]	2019	Republic of Korea	Prospective cohort study	Multicenter	CPC 1-2 and survival to discharge
Harford et al [102]	2019	United States	Retrospective cohort study	Registration database	CPC 1-2
Kwon et al [103]	2018	Republic of Korea	Retrospective cohort study	Multicenter	Cardiac arrest and in-hospital mortality
Chang et al [104]	2018	Taiwan, China	Retrospective cohort study	Single center	Cardiac arrest
Shin et al [105]	2018	Republic of Korea	Retrospective cohort study	Multicenter	CPC 1-2
Lee et al [106]	2017	Republic of Korea	Retrospective cohort study	Single center	Survival to hospital discharge
Liu et al [107]	2015	Singapore	Retrospective cohort study	Single center	Cardiac arrest
Goto et al [108]	2014	Japan	Retrospective cohort study	Registration database	CPC 1-2 and 30-day survival
Goto et al [109]	2013	Japan	Retrospective cohort study	Registration database	CPC 1-2 and 30-day survival
Hock Ong et al [110]	2012	Singapore	Prospective cohort study	Single center	Cardiac arrest and in-hospital mortality
Hayakawa et al [111]	2011	Japan	Prospective cohort study	Registration database	CPC 1-2

[a]CPC 3-5: poor cerebral performance category score 3 to 5.

[b]CPC 1-2: good cerebral performance category score 1 to 2.

[c]VF: ventricular fibrillation.

[d]VT: ventricular tachycardia.

[e]ROSC: return of spontaneous circulation.

[f]NROSC: non-ROSC.

[g]OHCA: out-of-hospital cardiac arrest.

[h]STE: ST segment elevation.

XSL•FO
RenderX

Table 2. Analytical characteristics of the included studies.

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Wang et al [20]	Unbalanced	In hospital	474	224,413	182,716	External validation	41,697	Deletion	Logistic regression, National Early Warning Score, and Modified Early Warning Score
Raheem et al [21]	Unbalanced	In hospital	5483	97,353	77,886	Internal validation	19,467	Deletion	Artificial neural network, random forest, and logistic regression
Amacher et al [22]	Balanced	In hospital and out of hospital	IHM[b]: 309; CPC 3-5[c]: 309	713	—[d]	—	713	No processing	Out-of-hospital CA score, the Cardiac Arrest Hospital Prognosis score, and logistic regression
Cho et al [23]	Unbalanced	In hospital	228	95,607	—	External validation	95,607	Deletion	Deep learning, Modified Early Warning Score, and National Early Warning Score
Shin et al [24]	Unbalanced	In hospital	198	1995	970	External validation	1025	Deletion	Deep learning, logistic regression, random forest, and National Early Warning Score
Ding et al [25]	Balanced	In hospital	CPC 1-2[e]: 20; IHM: 30	53	—	Internal validation	53	Deletion	Logistic regression and Cox regression
Nishioka et al [26]	Balanced	Out of hospital	6486	7587	3337	External validation	4250	Supplement	Logistic regression
Kajino et al [27]	Unbalanced	Out of hospital	11,411	302,799	149,425	Internal validation	153,374	Deletion	Deep learning
Pham et al [28]	Balanced	Out of hospital	210	434	231	External validation (multicenter)	203	—	Logistic regression
Rahadian et al [29]	Unbalanced	Out of hospital	860	20,713	17,162	—	3551	Imputation	Logistic regression, LASSO[f], and random forest
Wang et al [30]	Unbalanced	Out of hospital	84	48,371	32,244	—	16,127	—	Logistic regression
Tsai et al [31]	Balanced	Out of hospital	CPC 1-2: 127	443	265	Internal validation	178	—	Logistic regression
Schweiger et al [32]	Balanced	Out of hospital	120	291	138	Internal validation	153	Supplement	FACTOR score
Caputo et al [33]	Unbalanced	Out of hospital	2719	12,577	—	Internal validation	12,577	—	Logistic regression

XSL•FO
RenderX

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Lu et al [34]	Unbalanced	Emergency department	636	316,465	237,349	Random sampling	79,116	Supplement	Logistic regression, random forest, National Early Warning Score 2, and XGBoost[g]
Dünser et al [35]	Balanced	Operating rooms and departments outside the ICU[h]	NROSC[i]: 390; CPC 3-5: 559	630	—	Internal validation	630	No processing	Random forest
Bang et al [36]	Balanced	In hospital	411	1133	754	Random sampling	379	Deletion	Logistic regression
Zhang et al [37]	Balanced	In hospital	495	561	561	—	—	Deletion	Logistic regression
Li and Xing [38]	Balanced	In hospital	NROSC: 564; CPC 3-5: 229	851	851	Internal validation (bootstrap)	—	Deletion	Logistic regression
Ding et al [39]	Balanced	In hospital	1769	3592	2873	Internal validation	719	Deletion	Support vector machine, random forest, XGBoost, decision tree, and logistic regression
Uehara et al [40]	Unbalanced	Out of hospital	71	8422	4239	Random sampling (1:1)	4183	Deletion	Logistic regression
Shin et al [41]	Unbalanced	Out of hospital	ROSC[j]: 3095; CPC 1-2: 990	16,992	—	Random sampling	—	Deletion	K-nearest neighbor, decision tree, random forest, support vector machine, logistic regression, and deep learning
Kawai et al [42]	Balanced	Out of hospital	254	321	257	Random sampling (8:2)	64	Deletion	Deep learning
Imamura et al [43]	Balanced	Out of hospital	172	274	194	External validation (multicenter)	80	Deletion	Logistic regression
Hessulf et al [44]	Unbalanced	Out of hospital	6191	55,615	44,492	Random sampling	11,123	Algorithm	XGBoost
Yoon et al [45]	Balanced	Out of hospital	74	131	—	External validation	131	Deletion	Logistic regression
Chang et al [46]	Unbalanced	Out of hospital	ROSC: 11,996; STD[k]: 11,833; 30-day survival: 7760; CPC 1-2: 3673	157,654	157,654	Internal validation	—	Deletion	LightGBM[l]
Wang et al [47]	Unbalanced	Out of hospital	156	2685	2685	Internal validation	—	Deletion	Logistic regression
Shinada et al [48]	Unbalanced	Out of hospital	1128	5340	4286	Internal validation	1054	Deletion	Naïve Bayes

XSL•FO
RenderX

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Xu et al [49]	Balanced	Emergency department and out of hospital	150	600	600	—	—	Deletion	Logistic regression
Tsai et al [50]	Unbalanced	Emergency department	623	325,502	325,502	—	—	No processing	Logistic regression, Modified Early Warning Score, and National Early Warning Score
Tang et al [51]	Balanced	ICU	107	486	—	Internal validation	486	Algorithm	National Early Warning Score, random forest, artificial neural network, and deep learning
Kim et al [52]	Unbalanced	Emergency department	5431	1,350,693	1,080,554	Random sampling	270,139	Deletion	Logistic regression, XGBoost, artificial neural network, and logistic regression
Chae et al [53]	Unbalanced	In hospital	573	34,452	—	Random sampling	34,452	Supplement	Decision tree, random forest, logistic regression, and artificial neural network
Sun et al [54]	Unbalanced	Emergency department	240	145,557	—	External validation	145,557	Deletion	Emergency department, in-hospital CA score, Modified Early Warning Score, and Rapid Emergency Medicine Score
Wong et al [55]	Unbalanced	Out of hospital	855	4776	3582	Random sampling	1194	Deletion	Random forest
Tran et al [56]	Balanced	Out of hospital	996	2999	2999	Internal validation	—	Deletion	Logistic regression
Rajendram et al [57]	Unbalanced	Out of hospital	STD: 3549; CPC 1-2: 1754	24,897	—	External validation	24,897	Deletion	Random forest
Rafi et al [58]	Balanced	Out of hospital	410	820	—	Internal validation	820	Supplement (algorithm)	Logistic regression, random forest, and artificial neural network
Liu et al [59]	Unbalanced	Out of hospital	12,729	153,611	119,477	External validation (multicenter)	34,134	Deletion	Random forest
Lin et al [60]	Unbalanced	Out of hospital	160	3520	2816	Random sampling (8:2)	704	Deletion	Decision tree and random forest
Kawai et al [61]	Unbalanced	Out of hospital	286	8274	6620	Internal validation (cross-validation)	1654	Deletion	Artificial neural network

XSL•FO
RenderX

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Itagaki et al [62]	Balanced	Out of hospital	BD[m]: 77	419	—	Internal validation (bootstrap)	419	Deletion	Logistic regression
Harris et al [63]	Unbalanced	Out of hospital	399	1726	1381	Random sampling	345	Supplement (algorithm)	Logistic regression, random forest, and LightGBM
Harford et al [64]	Unbalanced	Out of hospital	379	1798	957	Internal validation (cross-validation)	241 and 600	Deletion	LightGBM, XGBoost, decision tree, random forest, k-nearest neighbor, logistic regression, and deep learning
Harford et al [65]	Unbalanced	Out of hospital	670	9595	5750	Random sampling	1445 and 2400	Deletion	Deep learning
Chung et al [66]	Unbalanced	In hospital	94	796	637	Random sampling	159	No processing	Artificial neural network
Chi et al [67]	Balanced	In hospital	87,311	168,693	168,693	—	—	Deletion	HVec[n]
Wang et al [68]	Unbalanced	In hospital	46	159	80	External validation (multicenter)	79	Deletion	CANP[o] score
Bae et al [69]	Balanced	In hospital	643	982	671	External validation (prospective)	311	Deletion	Logistic regression
Mueller et al [70]	Balanced	In hospital	223	475	475	—	—	Deletion	Logistic regression
Lee et al [71]	Unbalanced	In hospital	425	332,371	173,368	External validation (multicenter)	159,003	Algorithm	Deep learning and Modified Early Warning Score
Lim et al [72]	Unbalanced	—	892	8240	4712	External validation (prospective)	3528	Deletion	Logistic regression
Lo and Siu [73]	Unbalanced	Out of hospital	2787	8157	6525	Random sampling	1632	Deletion	Logistic regression, random forest, and artificial neural network
Lonsain et al [74]	Balanced	Out of hospital	168	192	192	Internal validation	—	—	Logistic regression
Mueller et al [70]	Balanced	Out of hospital	761	1874	—	Internal validation	1874	—	Logistic regression
Nishioka et al [75]	Balanced	Out of hospital	382	2354	1329	External validation (prospective)	1025	Supplement (algorithm)	Logistic regression
Beom et al [76]	Balanced	Out of hospital	STD: 475; CPC 1-2: 315	1432	496 (survival prognosis validation group) and 489 (neurological prognosis validation group)	Random sampling (7:3)	STD: 227; CPC 1-2: 220	Deletion	Logistic regression

XSL•FO
RenderX

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Cheng et al [77]	Unbalanced	Out of hospital	86	1071	—	Random sampling (9:1)	1071	Deletion	Logistic regression, XGBoost, and support vector machine
Kim et al [78]	Unbalanced	Out of hospital	1986	39,602	39,602	Internal validation	—	Deletion	Random forest, LightGBM, and artificial neural network
Seo et al [79]	Balanced	Out of hospital	105	5739	5739	Internal validation	—	Supplement (algorithm)	Random forest, XGBoost, and logistic regression
Song et al [80]	Balanced	Out of hospital	61	106	—	External validation	106	Deletion	Out-of-hospital CA score
Sun et al [81]	Unbalanced	Out of hospital	148	447	447	Internal validation	—	Deletion	Logistic regression
Youn et al [82]	Unbalanced	Out of hospital	127	331	—	Internal validation	331	Deletion	Random forest, CatBoost, and logistic regression
Heo et al [83]	Balanced	Out of hospital	704	903	631	External validation (prospective)	158 and 114	Mean	Ensemble learning and logistic regression
Wang et al [84]	Balanced	In hospital and out of hospital	114	262	262	Internal validation	—	No processing	Logistic regression
Hong et al [85]	Unbalanced	Emergency department	993	214,307	168,488	Random sampling	45,819	Supplement	Modified Early Warning Score, logistic regression, artificial neural network, and random forest
Cho et al [86]	Unbalanced	Inpatient ward	11	8039	—	External validation	8039	No processing	Modified Early Warning Score and deep learning
Hirano et al [87]	Balanced	Out of hospital	30-day mortality: 13,329	30,049	23,668	Internal validation (10-fold cross-validation)	6381	Deletion	Logistic regression, support vector machine, random forest, artificial neural network, and multilayer perceptron
Okada et al [88]	Unbalanced	Out of hospital	114	916	458	Internal validation	458	—	Logistic regression
Liu et al [89]	Unbalanced	Out of hospital	5190	63,059	44,141	Internal validation	18,918	—	ROSC after CA score and random forest
Elola et al [90]	Unbalanced	Out of hospital	55	162	96	Internal validation (5-fold cross-validation)	66	—	Random forest
Hsieh et al [91]	Unbalanced	Out of hospital	660	252,771	168,522	Internal validation	84,249	Deletion	Logistic regression

XSL•FO
RenderX

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Baldi et al [92]	Unbalanced	Out of hospital	625	2709	1962	Internal validation	747	—	Logistic regression
Li et al [93]	Unbalanced	Emergency department	164	656	656	Random sampling	—	Supplement (algorithm)	Decision tree
Srivilaithon et al [94]	Unbalanced	Emergency department	250	1250	—	External validation	1250	Deletion	National Early Warning Score
Lee et al [95]	Balanced	In hospital	367	580	580	—	—	Deletion	Logistic regression
Liu et al [96]	Unbalanced	Emergency department	124	43,569	43,569	Internal validation	—	No processing	AdaBoost[p], random forest, naïve Bayes, decision tree, logistic regression, artificial neural network, and deep learning
Jang et al [97]	Unbalanced	Emergency department	1568	523,852	261,926	—	261,926	Deletion	Artificial neural network, Modified Early Warning Score, logistic regression, and random forest
Seki et al [98]	Unbalanced	Out of hospital	432	7326	5718	External validation (prospective)	1608	Imputation	Random forest
Park et al [99]	Unbalanced	Out of hospital	2805	19,832	15,860	Random sampling (8:2)	3972	Deletion	Logistic regression, XGBoost, support vector machine, random forest, and artificial neural network
Kwon et al [100]	Unbalanced	Out of hospital	CPC 1-2: 3812; STD: 6435	36,190	28,045	—	8145	—	Deep learning, logistic regression, random forest, and support vector machine
Kong et al [101]	Unbalanced	Out of hospital	CPC 1-2: 156; STD: 251	737	524	External validation	213	—	Logistic regression
Harford et al [102]	Unbalanced	Out of hospital	250	2244	1584	Internal validation	660	Supplement (algorithm)	Deep learning
Kwon et al [103]	Unbalanced	In hospital	CA: 415; IHM: 795	50,359	46,725	External validation (multicenter)	3634	Supplement (median)	Deep learning, Modified Early Warning Score, logistic regression, and random forest

XSL•FO
RenderX

Study	Balance of data (balanced or unbalanced)	Location of CA[a]	Number of cases of studied events	Total number of cases	Number of cases in the training set	Generation of validation set	Number of cases in the validation set	Handling method for missing values	Model type
Chang et al [104]	Unbalanced	Emergency department	124	43,569	—	Internal validation	43,569	Supplement (mean)	Logistic regression, decision tree, random forest, and XGBoost
Shin et al [105]	Unbalanced	Out of hospital	CPC 1-2: 86	456	228	Internal validation	228	Deletion	Decision tree
Lee et al [106]	Unbalanced	Emergency department	21	111	—	Internal validation (bootstrap)	111	Deletion	Logistic regression and Simplified Acute Physiology Score II
Liu et al [107]	Unbalanced	Emergency department	52	1025	—	Internal validation (cross-validation)	1025	Deletion	Proposed scoring system and distance scoring system
Goto et al [108]	Unbalanced	Out of hospital	CPC 1-2: 205; 30-day survival: 581	5379	3693	External validation (prospective)	1686	Deletion	Decision tree
Goto et al [109]	Unbalanced	Out of hospital	CPC 1-2: 7769; 30-day survival: 16,332	390,226	307,896	Internal validation	82,330	Deletion	Decision tree
Hock Ong et al [110]	Unbalanced	Emergency department	CA: 43; IHM: 86	925	—	External validation	925	Deletion	Modified Early Warning Score and support vector machine
Hayakawa et al [111]	Unbalanced	Out of hospital	244	1497	862	External validation (prospective)	635	Deletion	Logistic regression

[a]CA: cardiac arrest.

[b]IHM: in-hospital mortality.

[c]CPC 3-5: poor cerebral performance category score 3 to 5.

[d]Not provided.

[e]CPC 1-2: good cerebral performance category score 1 to 2.

[f]LASSO: least absolute shrinkage and selection operator.

[g]XGBoost: Extreme Gradient Boosting.

[h]ICU: intensive care unit.

[i]NROSC: nonreturn of spontaneous circulation.

[j]ROSC: return of spontaneous circulation.

[k]STD: survival to discharge.

[l]LightGBM: Light Gradient-Boosting Machine.

[m]BD: brain death.

[n]HVec: hierarchical vectorizer.

[o]CANP: Cardiac Arrest Neurological Prognosis.

[p]AdaBoost: Adaptive Boosting.

Risk of Bias in the Studies

After our exclusion of previously established scoring tools, a quality assessment of 208 ML models, involving 17 types, was conducted. In total, 24% (50/208) of these models originated from case-control studies, which introduced a high risk of bias in study participant selection. Regarding predictive factors, 1% (2/208) of the models were linked to a high risk of bias owing to the use of outcome information. Regarding outcome assessment, as both CA and prognosis outcomes were clearly defined using standard definitions, no additional predictive factors were required, resulting in a low risk of bias in outcome assessment. The included ML models were primarily derived from large-sample statistical analyses; however, 9.1% (19/208)

XSL•FO
RenderX

of the models were based on a very small number of cases, with an event per variable value of <10. In addition, inappropriate deletion methods were applied to address missing data in 65.4% (136/208) of the models, and only univariate analysis was used to screen for predictive factors in 22.6% (47/208) of the models, ultimately resulting in a high risk of bias for 76.9% (160/208) of the models in the domain of statistical analysis, as detailed in Figure 2.

Figure 2. Assessment results for the risk of bias in the included models.

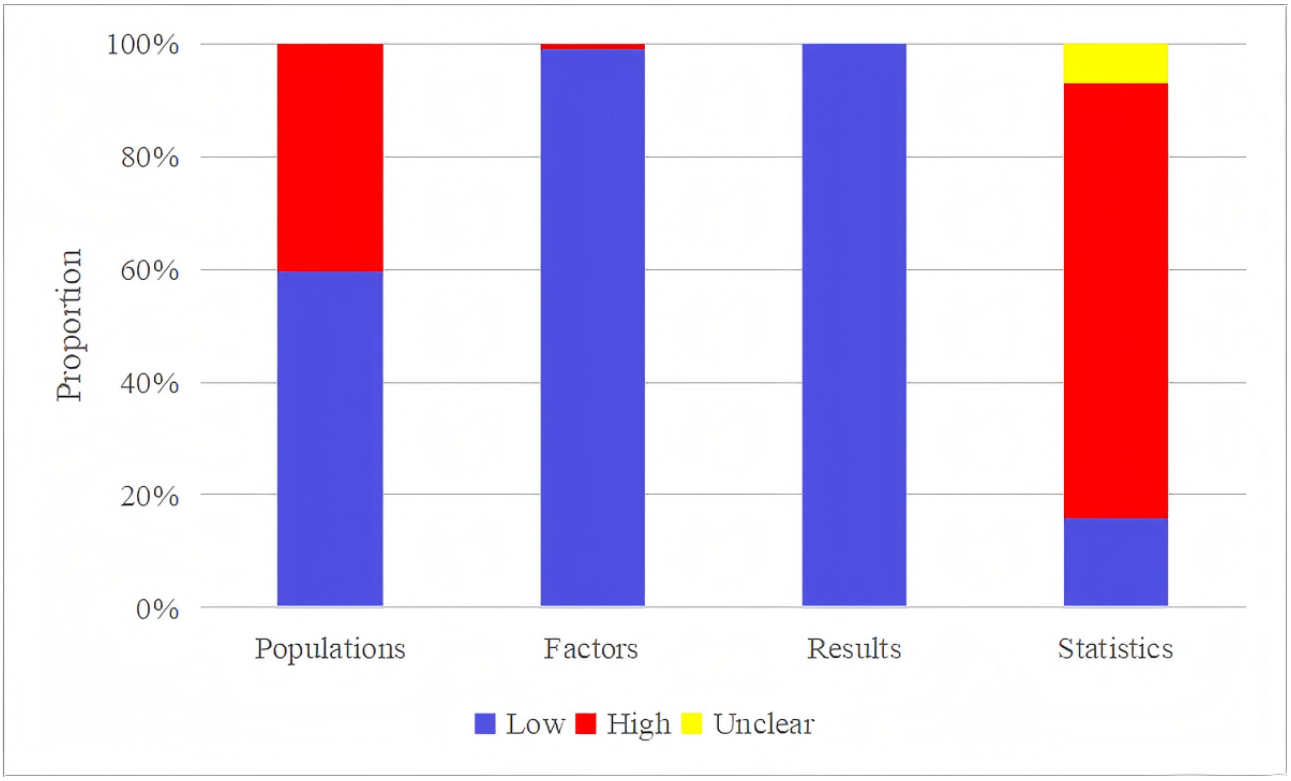

Meta-Analysis

CA Occurrence

A meta-analysis of ML models for predicting CA occurrence in the training set was conducted through a random-effects model. The analysis revealed a C-index of 0.84 (95% CI 0.82-0.86; 38/208, 18.3% of the models), sensitivity of 0.78 (95% CI 0.70-0.84; 34/208, 16.3% of the models), and specificity of 0.84 (95% CI 0.80-0.88; 34/208, 16.3% of the models). Similarly, a meta-analysis of ML models for predicting CA occurrence in the validation set was conducted, yielding a C-index of 0.89 (95% CI 0.87-0.91; 52/208, 25% of the models), sensitivity of 0.83 (95% CI 0.78-0.87; 43/208, 20.7% of the models), and specificity of 0.93 (95% CI 0.88-0.96; 43/208, 20.7% of the models; Multimedia Appendices 6-13).

Because of the diverse sources of modeling data from both balanced and imbalanced datasets and the variety of models, a subgroup analysis was conducted based on the data model type. The detailed results are presented in Multimedia Appendices 14-18.

Favorable Neurological Outcomes (CPC 1-2)

A meta-analysis of ML models for predicting CPC 1-2 in the training set was conducted using a random-effects model. The results indicated a C-index of 0.90 (95% CI 0.89-0.92; 21/208, 10.1% of the models), sensitivity of 0.72 (95% CI 0.47-0.98; 15/208, 7.2% of the models), and specificity of 0.85 (95% CI 0.79-0.90; 15/208, 7.2% of the models). Similarly, the meta-analysis of ML models for predicting CPC 1-2 in the validation set revealed a C-index of 0.86 (95% CI 0.85-0.87; 69/208, 33.2% of the models), sensitivity of 0.72 (95% CI 0.61-0.81; 44/208, 21.2% of the models), and specificity of 0.79 (95% CI 0.66-0.88; 44/208, 21.2% of the models; Multimedia Appendix 19, and Table S1 and Figures S1-S3 in Multimedia Appendix 20).

It was hypothesized that there might have been differences in CPC 1-2 between patients experiencing IHCA and OHCA. As the real-world data closely resembled balanced datasets, a subgroup analysis was conducted exclusively on the IHCA and OHCA populations. The detailed results are shown in Multimedia Appendix 19, and Table S1 and subgroup analysis report S1 in Multimedia Appendix 20.

CA Mortality

A random-effects model was used for the meta-analysis of ML models for predicting CA mortality in the training set. The analysis indicated a C-index of 0.80 (95% CI 0.76-0.84; 14/208, 6.7% of the models), sensitivity of 0.82 (95% CI 0.58-0.94; 7/208, 3.4% of the models), and specificity of 0.76 (95% CI 0.51-0.91; 7/208, 3.4% of the models). Similarly, a meta-analysis of ML models for predicting CA mortality in the validation set revealed a C-index of 0.85 (95% CI 0.82-0.87; 28/208, 13.5% of the models), sensitivity of 0.83 (95% CI 0.79-0.87; 23/208, 11.1% of the models), and specificity of 0.79 (95% CI 0.74-0.83; 23/208, 11.1% of the models; Tables S2-S3 and Figures S4-S6 in Multimedia Appendix 20).

XSL•FO
RenderX

It was thought that there might have been differences in mortality rates between patients experiencing IHCA and OHCA. Given that the real-world data closely resembled balanced datasets, a subgroup analysis was conducted exclusively on the IHCA and OHCA populations. The detailed analysis results are provided in Tables S2-S3 and subgroup analysis report S2 in Multimedia Appendix 20.

ROSC Analysis

A meta-analysis of ML models for predicting ROSC following CA in the training set was conducted using a random-effects model. The analysis yielded a C-index of 0.83 (95% CI 0.79-0.88; 10/208, 4.8% of the models), sensitivity of 0.52 (95% CI 0.31-0.73; 8/208, 3.8% of the models), and specificity of 0.91 (95% CI 0.88-0.93; 8/208, 3.8% of the models). Similarly, a meta-analysis of ML models for predicting ROSC in the validation set revealed a C-index of 0.77 (95% CI 0.74-0.80; 13/208, 6.3% of the models), sensitivity of 0.53 (95% CI 0.31-0.74; 6/208, 2.9% of the models), and specificity of 0.88 (95% CI 0.71-0.96; 6/208, 2.9% of the models; Tables S4-S5 and Figures S7-S9 in Multimedia Appendix 20).

It was postulated that there may have been differences in ROSC between patients who experienced IHCA and OHCA. As the real-world data closely resembled balanced datasets, a subgroup analysis was performed solely on the IHCA and OHCA populations. The comprehensive analysis results are provided in Tables S4-S5 and subgroup analysis report S3 in Multimedia Appendix 20.

Modeling Variables

Modeling variables were extracted and weighted for analysis from the 93 studies on ML models for predicting CA and CPC 1-2. Among the 28 studies on predicting CA, the variables with the highest weights were respiratory rate (n=22, 79%), blood pressure (n=20, 71%), age (n=19, 68%), temperature (n=19, 68%), oxygen saturation (n=15, 54%), and airway (n=9, 32%). Among the 42 studies on predicting CPC 1-2, the results showed that the modeling variables with the highest weights in the IHCA group were rhythm (shockable or nonshockable; 8/10, 80%), age (7/10, 70%), medication use (6/10, 60%), gender (5/10, 50%), and Glasgow Coma Scale (GCS; 5/10, 50%). The modeling variables with the highest weights in the OHCA group were age (25/32, 78%), rhythm (shockable or nonshockable; 24/32, 75%), medication use (18/32, 56%), ROSC (14/32, 44%), gender (12/32, 38%), no-flow time (resuscitation duration; 12/32, 38%), EMS transport (scene interval, arrival time, and response time; 12/32, 38%), defibrillation (11/32, 34%), and GCS (6/32, 19%). The detailed results of the modeling variables and weight analysis are presented in Tables S6 and S7 in Multimedia Appendix 20.

Discussion

Summary of the Principal Findings

It was observed that ML has garnered widespread attention among numerous researchers in the management of CA, particularly focusing on early CA risk prediction in both in-hospital and out-of-hospital populations. Our systematic review and meta-analysis demonstrated a relatively favorable predictive value of ML in the validation set for forecasting CA risk, with a C-index of 0.89. Similarly, ML also appeared to exhibit a relatively favorable predictive value for neurological outcomes (CPC 1-2) and mortality in patients who had already experienced CA, with pooled C-indexes of 0.86 and 0.85, respectively. However, in predicting ROSC following CA, ML seemed to display a predictive value comparable to that of traditional scoring tools, with a pooled C-index of 0.77.

Comparison With Previous Reviews

Currently, in clinical practice, classic early warning scoring tools, including the Cardiac Arrest Risk Triage (CART) score, Modified Early Warning Score, and VitalPAC Early Warning Score, are commonly used for predicting the occurrence of CA. A previous review by Churpek et al [112] found that, among these tools, the CART had the highest accuracy in predicting CA compared to the others. However, the CART had certain limitations, with an area under the curve of 0.83, sensitivity calculated at 0.61 based on the optimal Youden index, and a specificity of 0.84. Moreover, the CART has not been externally validated, and the included population is limited to ward inpatients. Therefore, whether the CART can dynamically monitor the occurrence of CA in real-time clinical events, improve rescue success rates, and enhance patient outcomes requires prospective validation using high-quality, large-sample external data. Our summarized results of the ML models reveal that ML has certain clinical predictive value in forecasting the occurrence of CA and demonstrates relatively favorable accuracy, with an overall C-index of 0.89, sensitivity of 0.83, and specificity of 0.93. Comparatively, this is superior to previous scoring tools, providing a certain clinical basis for the future establishment of more reliable early warning scoring systems for predicting CA.

In a recent review by Carrick et al [113], the accuracy of scoring tools for predicting survival or neurological outcomes following CA, such as the OHCA score, Cardiac Arrest Hospital Prognosis score, and Good Outcome Following Attempted Resuscitation score, was summarized. These 3 tools, which have undergone rigorous clinical validation, exhibited relatively high accuracy, with C-indexes of 0.79, 0.83, and 0.76, respectively. However, our summarized results of ML models suggested that ML seems to exhibit more favorable accuracy, with an overall C-index of 0.86, sensitivity of 0.72, and specificity of 0.79 for predicting favorable neurological outcomes. For predicting CA mortality, ML achieved an overall C-index of 0.85, sensitivity of 0.83, and specificity of 0.79.

Among various ROSC prediction models for CA that have been developed in the current clinical field, the ROSC after CA score developed by Gräsner et al [114] using data from 5471 patients with OHCA from the German Resuscitation Registry has attracted the most attention. It has been externally validated in several European and Asian countries, demonstrating relatively good accuracy, with an area under the curve of 0.736 in a recent large-scale external validation study [115]. However, our summarized results of ML models indicated that the overall C-index of ML was 0.77, with a sensitivity of 0.53 and specificity of 0.88. Comparatively, these ML models did not

XSL•FO
RenderX

seem to significantly outperform traditional scoring tools in predicting ROSC outcomes for patients with CA.

Modeling Variables in ML

In our review, the modeling variables of the discussed models primarily originated from common clinical features. It was found that variables such as respiratory rate, blood pressure, age, temperature, oxygen saturation, and airway were key predictors in existing ML models, and they also constitute critical variables in traditional scoring models [116,117]. Therefore, the impact of these variables on predicting the occurrence of CA is well established. However, these predictors differ to some extent from the findings of recent studies, such as the review by Andersen et al [3], which identified CA risk factors. The review by Andersen et al [3] suggested that age was more associated with post-CA prognosis and reduced survival rates, whereas a history of cardiac diseases such as myocardial infarction, arrhythmias, and heart failure was recognized as the most common risk factor for CA occurrence. Other potential risk factors included the use of certain medications, such as those that prolong the QT interval, opioids, and sedatives. Nonetheless, the review concurred that respiratory function and body temperature also had predictive significance for CA, with early interventions targeting these factors being crucial for achieving reversible outcomes [118].

In the models we reviewed that aimed to predict CPC 1-2 outcomes in patients with IHCA and OHCA, the modeling variables with the highest weight were age, rhythm (shockable or nonshockable), medication use, ROSC, gender, no-flow time (resuscitation duration), defibrillation, EMS transport (scene interval, arrival time, and response time), and GCS. When compared to the review by Sandroni et al [119], which highlighted the predictive value of GCS, biological markers (eg, neuron-specific enolase), and electrophysiological indicators (eg, somatosensory evoked potential) for favorable neurological outcomes, our findings show some differences. In addition, complex variables such as medical imaging might need consideration in clinical practice. In recent years, AI methods have been widely used in medical imaging for identifying disease progression and prognosis, demonstrating superior accuracy and cost-effectiveness compared to traditional clinical feature–based predictive models [120]. Therefore, in the prediction of CA occurrence and prognosis, the high-value variables identified in recent studies, such as electrocardiography [121] and ultrasound [122], are not reflected in traditional scoring tools. This raises the question of whether it is worth further validating these more complex variables and attempting to identify more efficient predictive factors to develop or update risk-scoring tools in the field of CA.

Clinical Applications of ML

Our study reveals that ML methods appear to outperform traditional scoring tools in predicting the occurrence and progression of CA. Therefore, the development of simple auxiliary tools based on ML theory is recommended to facilitate rapid risk screening of CA for both in-hospital and out-of-hospital patients, enabling timely formulation of appropriate treatment strategies. These ML-based CA prediction models would be particularly beneficial for emergency departments and out-of-hospital response teams. Under the current circumstances, in which emergency departments worldwide are facing challenges of overcrowding, resource limitations, and a high influx of patients who are critically ill [123-125], relying solely on human assessment of CA risk based on various clinical data could pose significant challenges to the efficiency of CA treatment. Furthermore, the complexity and volume of clinical data, including patient demographics, laboratory results, imaging data, and textual notes from health care providers, are continuously increasing. Thus, using ML to analyze large datasets and handle complex variables, such as clinical images, seems to be a more feasible approach [126]. The development of simplified ML prediction tools or intelligent reading systems has the potential to mitigate risks such as treatment delays and poor prognoses in patients with CA in emergency departments. These tools could also enhance health care service quality, reduce human resource costs, and support the formulation of targeted therapeutic strategies. Similarly, ML models that incorporate real-time input of variables such as vital signs, electrocardiograms, and response times for out-of-hospital rescue scenarios can accurately predict positive CA events. This capability aids response teams in avoiding repeated and frequent evaluations, enabling timely decisions on whether to implement preventive therapeutic interventions to avert CA or, in cases of CA occurrence, whether to initiate extracorporeal membrane oxygenation cardiopulmonary resuscitation to improve survival rates [127,128].

In addition, our findings indicate that the balance of data significantly impacts the outcomes of ML model construction in CA-related studies. This effect is particularly pronounced in scenarios in which the outcome metrics exhibit severe imbalance. For instance, in predicting the occurrence of CA in our study, the rationale for the selected predictive factors remained challenging, and the accuracy of the constructed model was often influenced by the overwhelming proportion of negative events [129]. In such cases, the C-index hardly represented the actual outcome prediction accuracy of the model. Therefore, in our study, the sensitivity and specificity of ML models were also summarized [130]. The data balance in the studies we included was primarily addressed using oversampling, but these studies rarely considered validating models constructed from balanced data against imbalanced data. This raises certain doubts regarding the accuracy of existing models constructed based on balanced data when applied to real-world cases of CA. Our study reveals that CA events in the real world are often inherently imbalanced. Therefore, we recommended prioritizing the use of models constructed from real-world data or validating models constructed from balanced data against real-world data to ascertain their true effectiveness [131].

Ethical Considerations and Model Selection

Although ML models demonstrated relatively satisfactory accuracy in predicting the occurrence and progression of CA in our study, several common challenges inherent to ML modeling should be acknowledged. For instance, compared to traditional scoring tools, ML models rely on general algorithms to generate desired outputs in response to specific input data, a process characterized by less explicit rules [132]. In addition,

XSL•FO
RenderX

algorithmic biases may result in unrepresentative datasets [133], and the reliability of model validation remains a concern [134]. These issues underscore the ethical challenges associated with the application of AI in medicine, including result interpretability, algorithmic transparency, predictive fairness, and data privacy [135,136]. These potential ethical concerns are specifically reflected in a patient survey study on the prevention of CA occurrence and development conducted by Maris et al [137]. The study results indicate that, while AI-driven CA treatment decisions offer objective data, the absence of patient involvement and informed consent, along with the interpretability of the model, suggest that the overuse of AI technology may ultimately undermine patient trust in physicians. This, in turn, poses challenges to the current high-quality health care goal of patient-centered care [138] in the field of cardiovascular disease treatment, which is built on shared decision-making [139], respect for patient autonomy, and mutual trust. Therefore, in the high-risk and critical treatment of CA, it seems that physicians should continue to make final decisions in collaboration with ML models based on evidence-based clinical experience and the values of the patients.

Therefore, based on ethical considerations, the choice of different models during research remains challenging as model interpretability and accuracy are factors that need to be considered comprehensively during model construction [140]. Selecting models with higher interpretability, such as LR, Cox regression, or DTs, can facilitate better communication, interaction, and trust between health care providers and patients. However, these models may have limited predictive value for certain outcome events [141]. On the other hand, models whose interpretability is poorer, including neural networks, support vector machines (SVMs), and Extreme Gradient Boosting, often perform exceptionally well in predicting outcomes [142]. At this point, it may become necessary to grant patients and their families greater rights to information and autonomy, enabling their active participation in medical decision-making. In our research, LR was the most frequently used model type as it facilitated the development of predictive nomograms, which are simple and easily applicable tools.

Among the 17 ML models that we included, artificial neural networks, RF, and LR appeared to demonstrate relatively ideal predictive value in forecasting the occurrence of CA and were more frequently used by clinical researchers. In predicting neurological outcomes, our study found that LR remained the model most commonly selected by clinicians, followed by DT, RF, SVM, and others. If we aim to develop a simplified predictive scoring scale to assist in clinical practice, priority may be given to using LR for its development and subsequent updates. This preference arises because, according to our research findings, LR demonstrates relatively satisfactory accuracy and facilitates the construction of straightforward and practical predictive nomograms [143,144]. Furthermore, considering the interpretability of models is essential in real-world clinical practice. However, if the objective is to develop auxiliary applications for disease surveillance and prediction in clinical settings, alternative, more complex models may be considered. For example, models such as neural networks, SVM, and Extreme Gradient Boosting, which demonstrated higher accuracy in our study, could be appropriate choices. On the other hand, when dealing with image-based features, such as medical imaging or electrocardiograms, it may be necessary to use models with lower interpretability, such as those based on deep learning [145], rather than confining the analysis to commonly used clinical features with stronger interpretability.

Prospects

In addition, we observed a minimal number of studies that constructed models based on artificial neural networks and ensemble learning, which exhibited highly favorable results, indicating that further validation of these models may be required in future research. Currently, in clinical practice, there is an increasing preference for using simple scoring tools based on interpretable clinical features. While opting for such tools may reduce the ethical dilemmas encountered in clinical settings, relying solely on traditional methods and highly interpretable clinical indicators, such as the Delphi method, during the development of these scoring tools seems to introduce significant bias into the constructed models. Therefore, we considered using multicenter real-world big data, using ML approaches, and incorporating a broader range of cases and clinical features to construct interpretable scoring tools and promote their application. Regarding the processing of clinical image features, our expectation lies in the development of intelligent reading tools based on deep learning methods. Nonetheless, in our study, there was limited research on deep learning based on medical imaging and ultrasounds, particularly in the field of CA, where such research remains relatively underexplored. Therefore, future research on CA should actively explore the integration of medical imaging and ultrasonography. In selecting datasets and algorithms for the development of AI prediction models, it is crucial to rigorously investigate and address the ethical shortcomings of AI applications in health care. Efforts should be made to minimize the influence of individual characteristics such as gender, race, skin color, and socioeconomic status, ensuring that population representation and sample size are carefully considered. Sufficient numbers and the quality of representative populations should be selected from diverse regions, ethnicities, and age groups to establish standardized big data models, thereby maximizing the potential of AI technologies [146,147]. It is essential to adopt a multifaceted, interdisciplinary approach; strengthen data protection systems to prevent the leakage of patient information; and conduct extensive reviews to avoid biases [148], ultimately preventing unfairness toward individuals or patient populations in the development of intelligent diagnostic or predictive tools for CA [149,150].

Advantages and Limitations of This Study

Our study has 3 strengths. First, it represents the first attempt to summarize evidence comparing ML models and scoring tools in predicting the occurrence and prognosis of CA, thereby providing an evidence-based foundation for the subsequent clinical update and development of new scoring tools or AI early warning systems in the field of CA. Second, our study encompassed 93 original studies with large sample sizes, covering 14 countries and involving 5,729,721 patients,

XSL•FO
RenderX

significantly enhancing the strength of our evidence. Third, we conducted a detailed discussion of the accuracy of different models on balanced and imbalanced data. However, this study also had the following limitations. First, most of the original studies on the prediction of CA occurrence (28/93, 30%) constructed models based solely on imbalanced data without validating them on balanced data. Second, many model validation processes primarily used internal validation through random sampling, lacking external multicenter validation to examine their generalizability. Third, due to potential differences in the predictive performance of different models for outcome events, despite our in-depth discussion of various ML models and datasets, the limited number of studies on certain ML models restricted our ability to interpret the results of ML applications in CA more comprehensively. Fourth, due to the small number of included studies, we did not strictly distinguish between IHCA and OHCA populations when summarizing the predictors of CA. Fifth, as this review only included English-language original studies, there may be potential language bias.

Conclusions

Current traditional scoring tools have demonstrated relatively ideal efficacy in predicting the occurrence and prognosis of CA. On the basis of this review, ML appeared to offer greater advantages in predicting the occurrence of CA, neurological functional prognosis, and mortality outcomes. However, for predicting outcomes associated with ROSC after CA, ML models did not seem to significantly outperform traditional models. Therefore, in future studies on CA, researchers may explore the systematic updating of traditional scoring tools based on the superior performance of ML in specific outcomes. This approach would enable the implementation of AI-driven enhancements within complex and diverse clinical data, thereby assisting clinicians in monitoring and providing early warnings for multiple predictive factors. For outcomes that are still unpredictable, multicenter large-sample studies are warranted.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

SW, XG, and SH contributed to original draft preparation and writing. SW, QL, and FZ conducted the initial study investigation and design. SW, CZ, and ZC critically reviewed and edited the manuscript. SW, YH, and QL extracted the data. FZ and JC performed the formal analysis. QL, YH, and FZ assessed the risk of bias of the included studies. QL and FZ collaborated on project administration and supervision and serve as co–corresponding authors. All authors commented on previous versions of the manuscript and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 98 KB - jmir_v27i1e67871_app1.pdf]

Multimedia Appendix 2
Literature search strategy in PubMed.
[DOCX File , 12 KB - jmir_v27i1e67871_app2.docx]

Multimedia Appendix 3
Literature search strategy in the Cochrane Library.
[DOCX File , 12 KB - jmir_v27i1e67871_app3.docx]

Multimedia Appendix 4
Literature search strategy in Embase.
[DOCX File , 12 KB - jmir_v27i1e67871_app4.docx]

Multimedia Appendix 5
Literature search strategy in Web of Science.
[DOCX File , 11 KB - jmir_v27i1e67871_app5.docx]

XSL•FO
RenderX

Multimedia Appendix 6
Meta-analysis results for the C-index of predictive models of in-hospital cardiac arrest risk.
[DOCX File , 16 KB - jmir_v27i1e67871_app6.docx]

Multimedia Appendix 7
Meta-analysis results for the sensitivity and specificity of predictive models for in-hospital cardiac arrest risk.
[DOCX File , 16 KB - jmir_v27i1e67871_app7.docx]

Multimedia Appendix 8
Forest plot of the C-index meta-analysis of prediction models of in-hospital cardiac arrest risk in the training set.
[DOCX File , 46 KB - jmir_v27i1e67871_app8.docx]

Multimedia Appendix 9
Forest plot of the C-index meta-analysis of prediction models of in-hospital cardiac arrest risk in the validation set.
[DOCX File , 43 KB - jmir_v27i1e67871_app9.docx]

Multimedia Appendix 10
Forest plot of the sensitivity meta-analysis of predictive models of in-hospital cardiac arrest risk in the training set.
[DOCX File , 43 KB - jmir_v27i1e67871_app10.docx]

Multimedia Appendix 11
Forest plot of the sensitivity meta-analysis of predictive models of in-hospital cardiac arrest risk in the validation set.
[DOCX File , 41 KB - jmir_v27i1e67871_app11.docx]

Multimedia Appendix 12
Forest plot of the specificity meta-analysis of predictive models of in-hospital cardiac arrest risk in the training set.
[DOCX File , 42 KB - jmir_v27i1e67871_app12.docx]

Multimedia Appendix 13
Forest plot of the specificity meta-analysis of predictive models of in-hospital cardiac arrest risk in the validation set.
[DOCX File , 41 KB - jmir_v27i1e67871_app13.docx]

Multimedia Appendix 14
Subgroup analysis for predicting cardiac arrest occurrence.
[DOCX File , 11 KB - jmir_v27i1e67871_app14.docx]

Multimedia Appendix 15
Meta-analysis results for the C-index of predictive models of in-hospital cardiac arrest risk in balanced datasets.
[DOCX File , 15 KB - jmir_v27i1e67871_app15.docx]

Multimedia Appendix 16
Meta-analysis results for the sensitivity and specificity of risk prediction models for in-hospital cardiac arrest in balanced datasets.
[DOCX File , 14 KB - jmir_v27i1e67871_app16.docx]

Multimedia Appendix 17
Meta-analysis results for the C-index of prediction models for in-hospital cardiac arrest risk in imbalanced datasets.
[DOCX File , 17 KB - jmir_v27i1e67871_app17.docx]

Multimedia Appendix 18
Meta-analysis results for the sensitivity and specificity of risk prediction models for in-hospital cardiac arrest in imbalanced datasets.
[DOCX File , 15 KB - jmir_v27i1e67871_app18.docx]

Multimedia Appendix 19

XSL•FO
RenderX

Meta-analysis results for the C-index of predictive models of favorable neurological function (good cerebral performance category score 1-2).
[DOCX File , 16 KB - jmir_v27i1e67871_app19.docx]

Multimedia Appendix 20
Tables/figures/subgroup analysis results.
[DOCX File , 331 KB - jmir_v27i1e67871_app20.docx]

References

1. Hirsch KG, Abella BS, Amorim E, Bader MK, Barletta JF, Berg K, et al. Critical care management of patients after cardiac arrest: a scientific statement from the American Heart Association and Neurocritical Care Society. Circulation 2024 Jan 09;149(2):e168-e200 [FREE Full text] [doi: 10.1161/CIR.0000000000001163] [Medline: 38014539]
2. Gräsner JT, Herlitz J, Tjelmeland IB, Wnent J, Masterson S, Lilja G, et al. European Resuscitation Council Guidelines 2021: epidemiology of cardiac arrest in Europe. Resuscitation 2021 Apr;161:61-79. [doi: 10.1016/j.resuscitation.2021.02.007] [Medline: 33773833]
3. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-hospital cardiac arrest: a review. JAMA 2019 Mar 26;321(12):1200-1210 [FREE Full text] [doi: 10.1001/jama.2019.1696] [Medline: 30912843]
4. Feng X, Yuguo C. Summary of the report on cardiac arrest and cardiopulmonary resuscitation in China (2022 edition). Chin Circul J 2023;38(10). [doi: 10.3969/j.issn.1000-3614.2023.10.002]
5. Zheng J, Lv C, Zheng W, Zhang G, Tan H, Ma Y, et al. Incidence, process of care, and outcomes of out-of-hospital cardiac arrest in China: a prospective study of the BASIC-OHCA registry. The Lancet Public Health 2023 Dec;8(12):e923-e932. [doi: 10.1016/s2468-2667(23)00173-1]
6. Stanger DE, Fordyce CB. The cost of care for cardiac arrest. Resuscitation 2018 Oct;131:A7-A8. [doi: 10.1016/j.resuscitation.2018.08.010] [Medline: 30099120]
7. Donnino MW, Rittenberger JC, Gaieski D, Cocchi MN, Giberson B, Peberdy MA, et al. The development and implementation of cardiac arrest centers. Resuscitation 2011 Aug;82(8):974-978. [doi: 10.1016/j.resuscitation.2011.03.021] [Medline: 21570761]
8. Scholz KH, Andresen D, Böttiger BW, Busch HJ, Fischer M, Frey N, et al. [Quality indicators and structural requirements for Cardiac Arrest Centers-German Resuscitation Council (GRC)]. Anaesthesist 2017 May 4;66(5):360-362. [doi: 10.1007/s00101-017-0311-7] [Medline: 28474242]
9. Drennan IR, Lin S, Thorpe KE, Morrison LJ. The effect of time to defibrillation and targeted temperature management on functional survival after out-of-hospital cardiac arrest. Resuscitation 2014 Nov;85(11):1623-1628. [doi: 10.1016/j.resuscitation.2014.07.010] [Medline: 25107539]
10. Kleinman ME, Brennan EE, Goldberger ZD, Swor RA, Terry M, Bobrow BJ, et al. Part 5: adult basic life support and cardiopulmonary resuscitation quality. Circulation 2015 Nov 03;132(18_suppl_2):1623-1628. [doi: 10.1161/cir.0000000000000259]
11. Panchal AR, Bartos JA, Cabañas JG, Donnino MW, Drennan IR, Hirsch KG, et al. Part 3: adult basic and advanced life support: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S366-S468 [FREE Full text] [doi: 10.1161/CIR.0000000000000916] [Medline: 33081529]
12. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 2020 Mar 27;9(2):14 [FREE Full text] [doi: 10.1167/tvst.9.2.14] [Medline: 32704420]
13. Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. Anesthesiology 2018 Oct;129(4):663-674. [doi: 10.1097/ALN.0000000000002300] [Medline: 29894315]
14. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. JAMA 2020 Mar 17;323(11):1052-1060 [FREE Full text] [doi: 10.1001/jama.2020.0592] [Medline: 32065827]
15. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/nejmra1814259]
16. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell 2023 Apr 13;186(8):1772-1791 [FREE Full text] [doi: 10.1016/j.cell.2023.01.035] [Medline: 36905928]
17. Sem M, Mastrangelo E, Lightfoot D, Aves T, Lin S, Mohindra R. The ability of machine learning algorithms to predict defibrillation success during cardiac arrest: a systematic review. Resuscitation 2023 Apr;185:109755. [doi: 10.1016/j.resuscitation.2023.109755] [Medline: 36842672]

XSL•FO
RenderX

18. Chen CC, Massey SL, Kirschen MP, Yuan I, Padiyath A, Simpao AF, et al. Electroencephalogram-based machine learning models to predict neurologic outcome after cardiac arrest: a systematic review. Resuscitation 2024 Jan;194:110049 [FREE Full text] [doi: 10.1016/j.resuscitation.2023.110049] [Medline: 37972682]
19. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2019 Sep;28(9):2768-2786 [FREE Full text] [doi: 10.1177/0962280218785504] [Medline: 30032705]
20. Wang YM, Chiu IM, Chuang YP, Cheng CY, Lin CF, Cheng FJ, et al. RAPID-ED: a predictive model for risk assessment of patient's early in-hospital deterioration from emergency department. Resusc Plus 2024 Mar;17:100570 [FREE Full text] [doi: 10.1016/j.resplu.2024.100570] [Medline: 38357677]
21. Raheem A, Waheed S, Karim M, Khan NU, Jawed R. Prediction of major adverse cardiac events in the emergency department using an artificial neural network with a systematic grid search. Int J Emerg Med 2024 Jan 04;17(1):4 [FREE Full text] [doi: 10.1186/s12245-023-00573-2] [Medline: 38178007]
22. Amacher SA, Arpagaus A, Sahmer C, Becker C, Gross S, Urben T, et al. Prediction of outcomes after cardiac arrest by a generative artificial intelligence model. Resusc Plus 2024 Jun;18:100587 [FREE Full text] [doi: 10.1016/j.resplu.2024.100587] [Medline: 38433764]
23. Cho KJ, Kim KH, Choi J, Yoo D, Kim J. External validation of deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest in patients admitted to general wards based on rapid response system operating and nonoperating periods: a single-center study. Crit Care Med 2024 Mar 01;52(3):e110-e120 [FREE Full text] [doi: 10.1097/CCM.0000000000006137] [Medline: 38381018]
24. Shin Y, Cho KJ, Chang M, Youk H, Kim YJ, Park JY, et al. The development and validation of a novel deep-learning algorithm to predict in-hospital cardiac arrest in ED-ICU (emergency department-based intensive care units): a single center retrospective cohort study. Signa Vitae 2024;20(4):83-98. [doi: 10.22514/sv.2024.045]
25. Ding G, Kuang A, Zhou Z, Lin Y, Chen Y. Development of prognostic models for predicting 90-day neurological function and mortality after cardiac arrest. Am J Emerg Med 2024 May;79:172-182. [doi: 10.1016/j.ajem.2024.02.022] [Medline: 38457952]
26. Nishioka N, Yamada T, Nakao S, Yoshiya K, Park C, Nishimura T, et al. External validation of updated prediction models for neurological outcomes at 90 days in patients with out-of-hospital cardiac arrest. J Am Heart Assoc 2024 May 07;13(9):e033824 [FREE Full text] [doi: 10.1161/JAHA.123.033824] [Medline: 38700024]
27. Kajino K, Daya MR, Onoe A, Nakamura F, Nakajima M, Sakuramoto K, et al. Development and validation of a prehospital termination of resuscitation (TOR) rule for out - of hospital cardiac arrest (OHCA) cases using general purpose artificial intelligence (AI). Resuscitation 2024 Apr;197:110165. [doi: 10.1016/j.resuscitation.2024.110165] [Medline: 38452995]
28. Pham HN, Holmstrom L, Chugh H, Uy-Evanado A, Nakamura K, Zhang Z, et al. Dynamic electrocardiogram changes are a novel risk marker for sudden cardiac death. Eur Heart J 2024 Mar 07;45(10):809-819 [FREE Full text] [doi: 10.1093/eurheartj/ehad770] [Medline: 37956651]
29. Rahadian RE, Okada Y, Shahidah N, Hong D, Ng YY, Chia MY, et al. Machine learning prediction of refractory ventricular fibrillation in out-of-hospital cardiac arrest using features available to EMS. Resusc Plus 2024 Jun;18:100606 [FREE Full text] [doi: 10.1016/j.resplu.2024.100606] [Medline: 38533482]
30. Wang SA, Chang CJ, Do Shin S, Chu SE, Huang CY, Hsu LM, et al. Development of a prediction model for emergency medical service witnessed traumatic out-of-hospital cardiac arrest: a multicenter cohort study. J Formos Med Assoc 2024 Jan;123(1):23-35 [FREE Full text] [doi: 10.1016/j.jfma.2023.07.011] [Medline: 37573159]
31. Tsai H, Chi CY, Wang LW, Su YJ, Chen YF, Tsai MS, et al. Outcome prediction of cardiac arrest with automatically computed gray-white matter ratio on computed tomography images. Crit Care 2024 Apr 09;28(1):118 [FREE Full text] [doi: 10.1186/s13054-024-04895-2] [Medline: 38594772]
32. Schweiger V, Hiller P, Utters R, Fenice A, Cammann VL, Di Vece D, et al. A novel score to predict in-hospital mortality for patients with acute coronary syndrome and out-of-hospital cardiac arrest: the FACTOR study. Clin Res Cardiol 2024 Apr 08;113(4):591-601 [FREE Full text] [doi: 10.1007/s00392-023-02367-1] [Medline: 38329513]
33. Caputo ML, Baldi E, Burkart R, Wilmes A, Cresta R, Benvenuti C, et al. Validation of Utstein-Based score to predict return of spontaneous circulation (UB-ROSC) in patients with out-of-hospital cardiac arrest. Resuscitation 2024 Apr;197:110113 [FREE Full text] [doi: 10.1016/j.resuscitation.2024.110113] [Medline: 38218400]
34. Lu TC, Wang CH, Chou FY, Sun JT, Chou EH, Huang EP, et al. Machine learning to predict in-hospital cardiac arrest from patients presenting to the emergency department. Intern Emerg Med 2023 Mar;18(2):595-605. [doi: 10.1007/s11739-022-03143-1] [Medline: 36335518]
35. Dünser MW, Hirschl D, Weh B, Meier J, Tschoellitsch T. The value of a machine learning algorithm to predict adverse short-term outcome during resuscitation of patients with in-hospital cardiac arrest: a retrospective study. Eur J Emerg Med 2023 Aug 01;30(4):252-259. [doi: 10.1097/MEJ.0000000000001031] [Medline: 37115946]
36. Bang HJ, Oh SH, Jeong WJ, Cha K, Park KN, Youn CS, et al. A novel cardiac arrest severity score for the early prediction of hypoxic-ischemic brain injury and in-hospital death. Am J Emerg Med 2023 Apr;66:22-30. [doi: 10.1016/j.ajem.2023.01.003] [Medline: 36669440]

XSL•FO
RenderX

37. Zhang Y, Rao C, Ran X, Hu H, Jing L, Peng S, et al. How to predict the death risk after an in-hospital cardiac arrest (IHCA) in intensive care unit? A retrospective double-centre cohort study from a tertiary hospital in China. BMJ Open 2023 Oct 05;13(10):e074214 [FREE Full text] [doi: 10.1136/bmjopen-2023-074214] [Medline: 37798030]
38. Li Z, Xing J. A model for predicting return of spontaneous circulation and neurological outcomes in adults after in-hospital cardiac arrest: development and evaluation. Front Neurol 2023;14:1323721 [FREE Full text] [doi: 10.3389/fneur.2023.1323721] [Medline: 38046585]
39. Ding X, Wang Y, Ma W, Peng Y, Huang J, Wang M, et al. Development of early prediction model of in-hospital cardiac arrest based on laboratory parameters. Biomed Eng Online 2023 Dec 06;22(1):116 [FREE Full text] [doi: 10.1186/s12938-023-01178-9] [Medline: 38057823]
40. Uehara K, Tagami T, Hyodo H, Ohara T, Sakurai A, Kitamura N, et al. Prehospital ABC (Age, Bystander and Cardiogram) scoring system to predict neurological outcomes of cardiopulmonary arrest on arrival: post hoc analysis of a multicentre prospective observational study. Emerg Med J 2023 Jan;40(1):42-47. [doi: 10.1136/emermed-2020-210864] [Medline: 35667823]
41. Shin SJ, Bae HS, Moon HJ, Kim GW, Cho YS, Lee DW, et al. Evaluation of optimal scene time interval for out-of-hospital cardiac arrest using a deep neural network. Am J Emerg Med 2023 Jan;63:29-37. [doi: 10.1016/j.ajem.2022.10.011] [Medline: 36544293]
42. Kawai Y, Kogeichi Y, Yamamoto K, Miyazaki K, Asai H, Fukushima H. Explainable artificial intelligence-based prediction of poor neurological outcome from head computed tomography in the immediate post-resuscitation phase. Sci Rep 2023 Apr 08;13(1):5759 [FREE Full text] [doi: 10.1038/s41598-023-32899-5] [Medline: 37031248]
43. Imamura S, Miyata M, Tagata K, Yokomine T, Ohmure K, Kawasoe M, et al. Prognostic predictors in patients with cardiopulmonary arrest: a novel equation for evaluating the 30-day mortality. J Cardiol 2023 Aug;82(2):146-152 [FREE Full text] [doi: 10.1016/j.jjcc.2023.01.006] [Medline: 36682713]
44. Hessulf F, Bhatt DL, Engdahl J, Lundgren P, Omerovic E, Rawshani A, et al. Predicting survival and neurological outcome in out-of-hospital cardiac arrest using machine learning: the SCARS model. EBioMedicine 2023 Mar;89:104464 [FREE Full text] [doi: 10.1016/j.ebiom.2023.104464] [Medline: 36773348]
45. Yoon JA, Kang C, Park JS, You Y, Min JH, In YN, et al. Quantitative analysis of early apparent diffusion coefficient values from MRIs for predicting neurological prognosis in survivors of out-of-hospital cardiac arrest: an observational study. Crit Care 2023 Oct 25;27(1):407 [FREE Full text] [doi: 10.1186/s13054-023-04696-z] [Medline: 37880777]
46. Chang H, Kim JW, Jung W, Heo S, Lee SU, Kim T, et al. Machine learning pre-hospital real-time cardiac arrest outcome prediction (PReCAP) using time-adaptive cohort model based on the Pan-Asian Resuscitation Outcome Study. Sci Rep 2023 Nov 21;13(1):20344 [FREE Full text] [doi: 10.1038/s41598-023-45767-z] [Medline: 37990066]
47. Wang JJ, Zhou Q, Huang ZH, Han Y, Qin CZ, Chen ZQ, et al. Establishment of a prediction model for prehospital return of spontaneous circulation in out-of-hospital patients with cardiac arrest. World J Cardiol 2023 Oct 26;15(10):508-517 [FREE Full text] [doi: 10.4330/wjc.v15.i10.508] [Medline: 37900904]
48. Shinada K, Matsuoka A, Koami H, Sakamoto Y. Bayesian network predicted variables for good neurological outcomes in patients with out-of-hospital cardiac arrest. PLoS One 2023 Sep 28;18(9):e0291258 [FREE Full text] [doi: 10.1371/journal.pone.0291258] [Medline: 37768915]
49. Xu Y, Zhang H, Zhao Z, Wen K, Tian C, Zhai Q, et al. A retrospective study: quick scoring of symptoms to estimate the risk of cardiac arrest in the emergency department. Emerg Med Int 2022 Nov 18;2022:6889237 [FREE Full text] [doi: 10.1155/2022/6889237] [Medline: 36438861]
50. Tsai CL, Lu TC, Fang CC, Wang CH, Lin JY, Chen WJ, et al. Development and validation of a novel triage tool for predicting cardiac arrest in the emergency department. West J Emerg Med 2022 Mar 23;23(2):258-267 [FREE Full text] [doi: 10.5811/westjem.2021.8.53063] [Medline: 35302462]
51. Tang Q, Cen X, Pan C. Explainable and efficient deep early warning system for cardiac arrest prediction from electronic health records. Math Biosci Eng 2022 Jul 08;19(10):9825-9841 [FREE Full text] [doi: 10.3934/mbe.2022457] [Medline: 36031970]
52. Kim JH, Choi A, Kim MJ, Hyun H, Kim S, Chang HJ. Development of a machine-learning algorithm to predict in-hospital cardiac arrest for emergency department patients using a nationwide database. Sci Rep 2022 Dec 16;12(1):21797 [FREE Full text] [doi: 10.1038/s41598-022-26167-1] [Medline: 36526686]
53. Chae M, Gil HW, Cho NJ, Lee H. Machine learning-based cardiac arrest prediction for early warning system. Mathematics 2022 Jun 13;10(12):2049. [doi: 10.3390/math10122049]
54. Sun JT, Chang CC, Lu TC, Lin JC, Wang CH, Fang CC, et al. External validation of a triage tool for predicting cardiac arrest in the emergency department. Sci Rep 2022 May 24;12(1):8779 [FREE Full text] [doi: 10.1038/s41598-022-12781-6] [Medline: 35610350]
55. Wong XY, Ang YK, Li K, Chin YH, Lam SS, Tan KB, et al. Development and validation of the SARICA score to predict survival after return of spontaneous circulation in out of hospital cardiac arrest using an interpretable machine learning framework. Resuscitation 2022 Jan;170:126-133. [doi: 10.1016/j.resuscitation.2021.11.029] [Medline: 34843878]

XSL•FO
RenderX

56. Tran AT, Hart AJ, Spertus JA, Jones PG, McNally BF, Malik AO, et al. A risk-adjustment model for patients presenting to hospitals with out-of-hospital cardiac arrest and ST-elevation myocardial infarction. Resuscitation 2022 Mar;171:41-47 [FREE Full text] [doi: 10.1016/j.resuscitation.2021.12.021] [Medline: 34968532]
57. Rajendram MF, Zarisfi F, Xie F, Shahidah N, Pek PP, Yeo JW, et al. External validation of the Survival After ROSC in Cardiac Arrest (SARICA) score for predicting survival after return of spontaneous circulation using multinational pan-Asian cohorts. Front Med (Lausanne) 2022 Sep 8;9:930226 [FREE Full text] [doi: 10.3389/fmed.2022.930226] [Medline: 36160129]
58. Rafi S, Gangloff C, Paulhet E, Grimault O, Soulat L, Bouzillé G, et al. Out-of-hospital cardiac arrest detection by machine learning based on the phonetic characteristics of the caller's voice. Stud Health Technol Inform 2022 May 25;294:445-449. [doi: 10.3233/SHTI220498] [Medline: 35612119]
59. Liu N, Liu M, Chen X, Ning Y, Lee JW, Siddiqui FJ, et al. Development and validation of an interpretable prehospital return of spontaneous circulation (P-ROSC) score for patients with out-of-hospital cardiac arrest using machine learning: a retrospective study. EClinicalMedicine 2022 Jun;48:101422 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101422] [Medline: 35706500]
60. Lin WC, Huang CH, Chien LT, Tseng HJ, Ng CJ, Hsu KH, et al. Tree-based algorithms and association rule mining for predicting patients' neurological outcomes after first-aid treatment for an out-of-hospital cardiac arrest during COVID-19 pandemic: application of data mining. Int J Gen Med 2022 Sep;Volume 15:7395-7405. [doi: 10.2147/ijgm.s384959]
61. Kawai Y, Okuda H, Kinoshita A, Yamamoto K, Miyazaki K, Takano K, et al. Visual assessment of interactions among resuscitation activity factors in out-of-hospital cardiopulmonary arrest using a machine learning model. PLoS One 2022;17(9):e0273787 [FREE Full text] [doi: 10.1371/journal.pone.0273787] [Medline: 36067174]
62. Itagaki Y, Hayakawa M, Maekawa K, Kodate A, Moriki K, Takahashi Y, et al. Early prediction model of brain death in out-of-hospital cardiac arrest patients: a single-center retrospective and internal validation analysis. BMC Emerg Med 2022 Nov 04;22(1):177 [FREE Full text] [doi: 10.1186/s12873-022-00734-1] [Medline: 36333656]
63. Harris M, Crowe RP, Anders J, D'Acunto S, Adelgais KM, Fishe JN. Identification of factors associated with return of spontaneous circulation after pediatric out-of-hospital cardiac arrest using natural language processing. Prehosp Emerg Care 2023;27(5):687-694. [doi: 10.1080/10903127.2022.2074180] [Medline: 35510881]
64. Harford S, Del Rios M, Heinert S, Weber J, Markul E, Tataris K, et al. A machine learning approach for modeling decisions in the out of hospital cardiac arrest care workflow. BMC Med Inform Decis Mak 2022 Jan 25;22(1):21 [FREE Full text] [doi: 10.1186/s12911-021-01730-4] [Medline: 35078470]
65. Harford S, Darabi H, Heinert S, Weber J, Campbell T, Kotini-Shah P, et al. Utilizing community level factors to improve prediction of out of hospital cardiac arrest outcome using machine learning. Resuscitation 2022 Sep;178:78-84 [FREE Full text] [doi: 10.1016/j.resuscitation.2022.07.006] [Medline: 35817268]
66. Chung CC, Chiu WT, Huang YH, Chan L, Hong CT, Chiu HW. Identifying prognostic factors and developing accurate outcome predictions for in-hospital cardiac arrest by using artificial neural networks. J Neurol Sci 2021 Jun 15;425:117445. [doi: 10.1016/j.jns.2021.117445] [Medline: 33878655]
67. Chi CY, Ao S, Winkler A, Fu KC, Xu J, Ho YL, et al. Predicting the mortality and readmission of in-hospital cardiac arrest patients with electronic health records: a machine learning approach. J Med Internet Res 2021 Sep 13;23(9):e27798 [FREE Full text] [doi: 10.2196/27798] [Medline: 34515639]
68. Wang G, Zhang Z, Xu X, Sun Q, Yang H, Zhang J. Derivation and validation of the CANP scoring model for predicting the neurological outcome in post-cardiac arrest patients. Neurosciences (Riyadh) 2021 Oct;26(4):372-378 [FREE Full text] [doi: 10.17712/nsj.2021.4.20210056] [Medline: 34663710]
69. Bae DH, Lee HY, Jung YH, Jeung KW, Lee BK, Youn CS, et al. PROLOGUE (PROgnostication using LOGistic regression model for Unselected adult cardiac arrest patients in the Early stages): development and validation of a scoring system for early prognostication in unselected adult cardiac arrest patients. Resuscitation 2021 Mar;159:60-68. [doi: 10.1016/j.resuscitation.2020.12.022] [Medline: 33388366]
70. Mueller M, Grafeneder J, Schoergenhofer C, Schwameis M, Schriefl C, Poppe M, et al. Initial blood pH, lactate and base deficit add no value to peri-arrest factors in prognostication of neurological outcome after out-of-hospital cardiac arrest. Front Med (Lausanne) 2021;8:697906 [FREE Full text] [doi: 10.3389/fmed.2021.697906] [Medline: 34604252]
71. Lee YJ, Cho KJ, Kwon O, Park H, Lee Y, Kwon JM, et al. A multicentre validation study of the deep learning-based early warning score for predicting in-hospital cardiac arrest in patients admitted to general wards. Resuscitation 2021 Apr 22;163:78-85 [FREE Full text] [doi: 10.1016/j.resuscitation.2021.04.013] [Medline: 33895236]
72. Lim HJ, Ro YS, Kim KH, Park JH, Hong KJ, Song KJ, et al. The ED-PLANN score: a simple risk stratification tool for out-of-hospital cardiac arrests derived from emergency departments in Korea. J Clin Med 2021 Dec 29;11(1):174 [FREE Full text] [doi: 10.3390/jcm11010174] [Medline: 35011915]
73. Lo YH, Siu YC. Predicting survived events in nontraumatic out-of-hospital cardiac arrest: a comparison study on machine learning and regression models. J Emerg Med 2021 Dec;61(6):683-694. [doi: 10.1016/j.jemermed.2021.07.058] [Medline: 34548227]

XSL•FO
RenderX

74. Lonsain WS, De Lausnay L, Wauters L, Desruelles D, Dewolf P. The prognostic value of early lactate clearance for survival after out-of-hospital cardiac arrest. Am J Emerg Med 2021 Aug;46:56-62. [doi: 10.1016/j.ajem.2021.03.013] [Medline: 33721591]
75. Nishioka N, Kobayashi D, Kiguchi T, Irisawa T, Yamada T, Yoshiya K, et al. Development and validation of early prediction for neurological outcome at 90 days after return of spontaneous circulation in out-of-hospital cardiac arrest. Resuscitation 2021 Nov;168:142-150. [doi: 10.1016/j.resuscitation.2021.09.027] [Medline: 34619295]
76. Beom JH, Park I, You JS, Roh YH, Kim MJ, Park YS. Predictive model of good clinical outcomes in patients undergoing coronary angiography after out-of-hospital cardiac arrest: a prospective, multicenter observational study conducted by the Korean Cardiac Arrest Research Consortium. J Clin Med 2021 Aug 20;10(16):3695 [FREE Full text] [doi: 10.3390/jcm10163695] [Medline: 34441996]
77. Cheng CY, Chiu IM, Zeng WH, Tsai CM, Lin CH. Machine learning models for survival and neurological outcome prediction of out-of-hospital cardiac arrest patients. Biomed Res Int 2021 Sep 17;2021:9590131 [FREE Full text] [doi: 10.1155/2021/9590131] [Medline: 34589553]
78. Kim JW, Ha J, Kim T, Yoon H, Hwang SY, Jo IJ, et al. Developing a time-adaptive prediction model for out-of-hospital cardiac arrest: nationwide cohort study in Korea. J Med Internet Res 2021 Jul 05;23(7):e28361 [FREE Full text] [doi: 10.2196/28361] [Medline: 36260382]
79. Seo DW, Yi H, Bae HJ, Kim YJ, Sohn CH, Ahn S, et al. Prediction of neurologically intact survival in cardiac arrest patients without pre-hospital return of spontaneous circulation: machine learning approach. J Clin Med 2021 Mar 05;10(5):1089 [FREE Full text] [doi: 10.3390/jcm10051089] [Medline: 33807882]
80. Song HG, Park JS, You Y, Ahn HJ, Yoo I, Kim SW, et al. Using out-of-hospital cardiac arrest (OHCA) and cardiac arrest hospital prognosis (CAHP) scores with modified objective data to improve neurological prognostic performance for out-of-hospital cardiac arrest survivors. J Clin Med 2021 Apr 22;10(9):1825 [FREE Full text] [doi: 10.3390/jcm10091825] [Medline: 33922191]
81. Sun KF, Poon KM, Lui CT, Tsui KL. Clinical prediction rule of termination of resuscitation for out-of-hospital cardiac arrest patient with pre-hospital defibrillation given. Am J Emerg Med 2021 Dec;50:733-738. [doi: 10.1016/j.ajem.2021.09.058] [Medline: 34879495]
82. Youn CS, Yi H, Kim YJ, Song H, Kim N, Kim WY. Early identification of resuscitated patients with a significant coronary disease in out-of-hospital cardiac arrest survivors without ST-segment elevation. J Clin Med 2021 Dec 02;10(23):5688 [FREE Full text] [doi: 10.3390/jcm10235688] [Medline: 34884390]
83. Heo JH, Kim T, Shin J, Suh GJ, Kim J, Jung YS, et al. Prediction of neurological outcomes in out-of-hospital cardiac arrest survivors immediately after return of spontaneous circulation: ensemble technique with four machine learning models. J Korean Med Sci 2021 Jul 19;36(28):e187 [FREE Full text] [doi: 10.3346/jkms.2021.36.e187] [Medline: 34282605]
84. Wang H, Tang L, Zhang L, Zhang ZL, Pei HH. Development a clinical prediction model of the neurological outcome for patients with coma and survived 24 hours after cardiopulmonary resuscitation. Clin Cardiol 2020 Sep;43(9):1024-1031 [FREE Full text] [doi: 10.1002/clc.23403] [Medline: 32573817]
85. Hong S, Lee S, Lee J, Cha WC, Kim K. Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: model development and retrospective clinical validation study. JMIR Med Inform 2020 Aug 04;8(8):e15932 [FREE Full text] [doi: 10.2196/15932] [Medline: 32749227]
86. Cho KJ, Kwon O, Kwon JM, Lee Y, Park H, Jeon KH, et al. Detecting patient deterioration using artificial intelligence in a rapid response system. Crit Care Med 2020 Apr;48(4):e285-e289. [doi: 10.1097/CCM.0000000000004236] [Medline: 32205618]
87. Hirano Y, Kondo Y, Sueyoshi K, Okamoto K, Tanaka H. Early outcome prediction for out-of-hospital cardiac arrest with initial shockable rhythm using machine learning models. Resuscitation 2021 Jan;158:49-56 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.11.020] [Medline: 33227398]
88. Okada Y, Kiguchi T, Irisawa T, Yamada T, Yoshiya K, Park C, et al. Development and validation of a clinical score to predict neurological outcomes in patients with out-of-hospital cardiac arrest treated with extracorporeal cardiopulmonary resuscitation. JAMA Netw Open 2020 Nov 02;3(11):e2022920. [doi: 10.1001/jamanetworkopen.2020.22920] [Medline: 33231635]
89. Liu N, Ho AF, Pek PP, Lu TC, Khruekarnchana P, Song KJ, et al. Prediction of ROSC after cardiac arrest using machine learning. Stud Health Technol Inform 2020 Jun 16;270:1357-1358. [doi: 10.3233/SHTI200440] [Medline: 32570657]
90. Elola A, Aramendi E, Rueda E, Irusta U, Wang H, Idris A. Towards the prediction of rearrest during out-of-hospital cardiac arrest. Entropy (Basel) 2020 Jul 09;22(7):758 [FREE Full text] [doi: 10.3390/e22070758] [Medline: 33286529]
91. Hsieh MJ, Chiang WC, Sun JT, Chang WT, Chien YC, Wang YC, et al. A prediction model for patients with emergency medical service witnessed out-of-hospital cardiac arrest. J Formos Med Assoc 2021 May;120(5):1229-1236 [FREE Full text] [doi: 10.1016/j.jfma.2020.09.017] [Medline: 33067067]
92. Baldi E, Caputo ML, Savastano S, Burkart R, Klersy C, Benvenuti C, et al. An Utstein-based model score to predict survival to hospital admission: the UB-ROSC score. Int J Cardiol 2020 Jun 01;308:84-89 [FREE Full text] [doi: 10.1016/j.ijcard.2020.01.032] [Medline: 31980268]

XSL•FO
RenderX

93. Li H, Wu TT, Yang DL, Guo YS, Liu PC, Chen Y, et al. Decision tree model for predicting in-hospital cardiac arrest among patients admitted with acute coronary syndrome. Clin Cardiol 2019 Nov;42(11):1087-1093 [FREE Full text] [doi: 10.1002/clc.23255] [Medline: 31509271]
94. Srivilaithon W, Amnuaypattanapon K, Limjindaporn C, Imsuwan I, Daorattanachai K, Dasanadeba I, et al. Predictors of in-hospital cardiac arrest within 24 h after emergency department triage: a case-control study in urban Thailand. Emerg Med Australas 2019 Oct;31(5):843-850. [doi: 10.1111/1742-6723.13267] [Medline: 30887710]
95. Lee HY, Jung YH, Jeung KW, Lee BK, Youn CS, Mamadjonov N, et al. Ion shift index as a promising prognostic indicator in adult patients resuscitated from cardiac arrest. Resuscitation 2019 Apr;137:116-123. [doi: 10.1016/j.resuscitation.2019.02.020] [Medline: 30807816]
96. Liu JH, Chang HK, Wu CT, Lim WS, Wang HC, Jang JS. Machine learning based early detection system of cardiac arrest. In: Proceedings of the International Conference on Technologies and Applications of Artificial Intelligence. 2019 Presented at: TAAI 2019; November 21-23, 2019; Kaohsiung, Taiwan. [doi: 10.1109/taai48200.2019.8959922]
97. Jang DH, Kim J, Jo YH, Lee JH, Hwang JE, Park SM, et al. Developing neural network models for early detection of cardiac arrest in emergency department. Am J Emerg Med 2020 Jan;38(1):43-49. [doi: 10.1016/j.ajem.2019.04.006] [Medline: 30982559]
98. Seki T, Tamura T, Suzuki M, SOS-KANTO 2012 Study Group. Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique. Resuscitation 2019 Aug;141:128-135. [doi: 10.1016/j.resuscitation.2019.06.006] [Medline: 31220514]
99. Park JH, Shin SD, Song KJ, Hong KJ, Ro YS, Choi JW, et al. Prediction of good neurological recovery after out-of-hospital cardiac arrest: a machine learning analysis. Resuscitation 2019 Sep;142:127-135. [doi: 10.1016/j.resuscitation.2019.07.020] [Medline: 31362082]
100. Kwon JM, Jeon KH, Kim HM, Kim MJ, Lim S, Kim KH, et al. Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. Resuscitation 2019 Jun;139:84-91. [doi: 10.1016/j.resuscitation.2019.04.007] [Medline: 30978378]
101. Kong T, Chung SP, Lee HS, Kim S, Lee J, Hwang SO, et al. The prognostic usefulness of the lactate/albumin ratio for predicting clinical outcomes in out-of-hospital cardiac arrest: a prospective, multicenter observational study (koCARC) study. Shock 2020 Apr;53(4):442-451. [doi: 10.1097/SHK.0000000000001405] [Medline: 31306348]
102. Harford S, Darabi H, Del Rios M, Majumdar S, Karim F, Vanden Hoek T, et al. A machine learning based model for out of hospital cardiac arrest outcome classification and sensitivity analysis. Resuscitation 2019 May;138:134-140. [doi: 10.1016/j.resuscitation.2019.03.012] [Medline: 30885826]
103. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018 Jun 26;7(13):e008678 [FREE Full text] [doi: 10.1161/JAHA.118.008678] [Medline: 29945914]
104. Chang HK, Wu CT, Liu JH, Jang JS. Using machine learning algorithms in medication for cardiac arrest early warning system construction and forecasting. In: Proceedings of the Conference on Technologies and Applications of Artificial Intelligence. 2018 Presented at: TAAI 2018; November 30-December 2, 2018; Taichung, Taiwan. [doi: 10.1109/taai.2018.00010]
105. Shin SM, Kim KS, Suh GJ, Kim K, Kwon WY, Shin J, et al. Prediction of neurological outcomes following the return of spontaneous circulation in patients with out-of-hospital cardiac arrest: retrospective fast-and-frugal tree analysis. Resuscitation 2018 Dec;133:65-70. [doi: 10.1016/j.resuscitation.2018.10.002] [Medline: 30292802]
106. Lee SW, Han KS, Park JS, Lee JS, Kim SJ. Prognostic indicators of survival and survival prediction model following extracorporeal cardiopulmonary resuscitation in patients with sudden refractory cardiac arrest. Ann Intensive Care 2017 Aug 30;7(1):87 [FREE Full text] [doi: 10.1186/s13613-017-0309-y] [Medline: 28856660]
107. Liu T, Lin Z, Ong ME, Koh ZX, Pek PP, Yeo YK, et al. Manifold ranking based scoring system with its application to cardiac arrest prediction: a retrospective study in emergency department patients. Comput Biol Med 2015 Dec 01;67:74-82. [doi: 10.1016/j.compbiomed.2015.10.001] [Medline: 26498047]
108. Goto Y, Maeda T, Nakatsu-Goto Y. Decision tree model for predicting long-term outcomes in children with out-of-hospital cardiac arrest: a nationwide, population-based observational study. Crit Care 2014 Jul 27;18(3):R133. [doi: 10.1186/cc13951] [Medline: 24972847]
109. Goto Y, Maeda T, Goto Y. Decision-tree model for predicting outcomes after out-of-hospital cardiac arrest in the emergency department. Crit Care 2013 Jul 11;17(4):R133 [FREE Full text] [doi: 10.1186/cc12812] [Medline: 23844724]
110. Hock Ong ME, Lee Ng CH, Goh K, Liu N, Koh ZX, Shahidah N, et al. Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score. Crit Care 2012 Jun 21;16(3):R108 [FREE Full text] [doi: 10.1186/cc11396] [Medline: 22715923]
111. Hayakawa K, Tasaki O, Hamasaki T, Sakai T, Shiozaki T, Nakagawa Y, et al. Prognostic indicators and outcome prediction model for patients with return of spontaneous circulation from cardiopulmonary arrest: the Utstein Osaka Project. Resuscitation 2011 Jul;82(7):874-880. [doi: 10.1016/j.resuscitation.2011.02.045] [Medline: 21482007]
112. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. Chest 2013 Jul;143(6):1758-1765 [FREE Full text] [doi: 10.1378/chest.12-1605] [Medline: 23732586]

XSL•FO
RenderX

113. Carrick RT, Park JG, McGinnes HL, Lundquist C, Brown KD, Janes WA, et al. Clinical predictive models of sudden cardiac arrest: a survey of the current science and analysis of model performances. JAHA 2020 Aug 18;9(16). [doi: 10.1161/jaha.119.017625]
114. Gräsner JT, Meybohm P, Lefering R, Wnent J, Bahr J, Messelken M, et al. ROSC after cardiac arrest--the RACA score to predict outcome after out-of-hospital cardiac arrest. Eur Heart J 2011 Jul;32(13):1649-1656. [doi: 10.1093/eurheartj/ehr107] [Medline: 21515626]
115. Liu N, Ong ME, Ho AF, Pek PP, Lu TC, Khruekarnchana P, et al. Validation of the ROSC after cardiac arrest (RACA) score in Pan-Asian out-of-hospital cardiac arrest patients. Resuscitation 2020 May;149:53-59. [doi: 10.1016/j.resuscitation.2020.01.029] [Medline: 32035177]
116. Bartkowiak B, Snyder AM, Benjamin A, Schneider A, Twu NM, Churpek MM, et al. Validating the electronic cardiac arrest risk triage (eCART) score for risk stratification of surgical inpatients in the postoperative setting: retrospective cohort study. Ann Surg 2019 Jul;269(6):1059-1063 [FREE Full text] [doi: 10.1097/SLA.0000000000002665] [Medline: 31082902]
117. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the between the flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. Resuscitation 2018 Mar;123:86-91 [FREE Full text] [doi: 10.1016/j.resuscitation.2017.10.028] [Medline: 29169912]
118. Lascarrou JB, Merdji H, Le Gouge A, Colin G, Grillet G, Girardie PF, et al. Targeted temperature management for cardiac arrest with nonshockable rhythm. N Engl J Med 2019 Dec 12;381(24):2327-2337. [doi: 10.1056/NEJMoa1906661] [Medline: 31577396]
119. Sandroni C, Nolan J, Cavallaro F, Antonelli M. In-hospital cardiac arrest: incidence, prognosis and possible measures to improve survival. Intensive Care Med 2007 Mar;33(2):237-245. [doi: 10.1007/s00134-006-0326-z] [Medline: 17019558]
120. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. J Am Coll Cardiol 2019 Mar 26;73(11):1317-1335 [FREE Full text] [doi: 10.1016/j.jacc.2018.12.054] [Medline: 30898208]
121. Thorén A, Rawshani A, Herlitz J, Engdahl J, Kahan T, Gustafsson L, et al. ECG-monitoring of in-hospital cardiac arrest and factors associated with survival. Resuscitation 2020 May;150:130-138. [doi: 10.1016/j.resuscitation.2020.03.002] [Medline: 32209378]
122. Teran F, Prats MI, Nelson BP, Kessler R, Blaivas M, Peberdy MA, et al. Focused transesophageal echocardiography during cardiac arrest resuscitation: JACC review topic of the week. J Am Coll Cardiol 2020 Aug 11;76(6):745-754 [FREE Full text] [doi: 10.1016/j.jacc.2020.05.074] [Medline: 32762909]
123. Marzano L, Darwich AS, Jayanth R, Sven L, Falk N, Bodeby P, et al. Diagnosing an overcrowded emergency department from its electronic health records. Sci Rep 2024 May 30;14(1):9955 [FREE Full text] [doi: 10.1038/s41598-024-60888-9] [Medline: 38688997]
124. Gross TK, Lane NE, Timm NL, Committee on Pediatric Emergency Medicine. Crowding in the emergency department: challenges and best practices for the care of children. Pediatrics 2023 Mar 01;151(3):e2022060972. [doi: 10.1542/peds.2022-060972] [Medline: 36806666]
125. Austin EE, Blakely B, Tufanaru C, Selwood A, Braithwaite J, Clay-Williams R. Strategies to measure and improve emergency department performance: a scoping review. Scand J Trauma Resusc Emerg Med 2020 Jul 15;28(1):55 [FREE Full text] [doi: 10.1186/s13049-020-00749-2] [Medline: 32539739]
126. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]
127. Richardson SA, Anderson D, Burrell AJ, Byrne T, Coull J, Diehl A, et al. Pre-hospital ECPR in an Australian metropolitan setting: a single-arm feasibility assessment-the CPR, pre-hospital ECPR and early reperfusion (CHEER3) study. Scand J Trauma Resusc Emerg Med 2023 Dec 13;31(1):100 [FREE Full text] [doi: 10.1186/s13049-023-01163-0] [Medline: 38093335]
128. Low CJ, Ling RR, Ramanathan K, Chen Y, Rochwerg B, Kitamura T, et al. Extracorporeal cardiopulmonary resuscitation versus conventional CPR in cardiac arrest: an updated meta-analysis and trial sequential analysis. Crit Care 2024 Mar 21;28(1):57 [FREE Full text] [doi: 10.1186/s13054-024-04830-5] [Medline: 38383506]
129. Banerjee P, Dehnbostel FO, Preissner R. Prediction is a balancing act: importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. Front Chem 2018;6:362 [FREE Full text] [doi: 10.3389/fchem.2018.00362] [Medline: 30271769]
130. Luu J, Borisenko E, Przekop V, Patil A, Forrester JD, Choi J. Practical guide to building machine learning-based clinical prediction models using imbalanced datasets. Trauma Surg Acute Care Open 2024;9(1):e001222 [FREE Full text] [doi: 10.1136/tsaco-2023-001222] [Medline: 38881829]
131. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. BMJ 2024 Oct 03;386:e078276 [FREE Full text] [doi: 10.1136/bmj-2023-078276] [Medline: 39227063]
132. Hinton G. Deep learning-a technology with the potential to transform health care. JAMA 2018 Oct 18;320(11):1101-1102. [doi: 10.1001/jama.2018.11100] [Medline: 30178065]

XSL•FO
RenderX

133. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018 Dec 18;169(12):866-872 [FREE Full text] [doi: 10.7326/M18-1990] [Medline: 30508424]
134. Staffa SJ, Zurakowski D. Statistical development and validation of clinical prediction models. Anesthesiology 2021 Oct 01;135(3):396-405. [doi: 10.1097/ALN.0000000000003871] [Medline: 34330146]
135. Luxton DD. Recommendations for the ethical use and design of artificial intelligent care providers. Artif Intell Med 2014 Oct;62(1):1-10. [doi: 10.1016/j.artmed.2014.06.004] [Medline: 25059820]
136. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021 Mar 15;22(1):14 [FREE Full text] [doi: 10.1186/s12910-021-00577-8] [Medline: 33588803]
137. Maris MT, Koçar A, Willems DL, Pols J, Tan HL, Lindinger GL, et al. Ethical use of artificial intelligence to prevent sudden cardiac death: an interview study of patient perspectives. BMC Med Ethics 2024 May 04;25(1):42 [FREE Full text] [doi: 10.1186/s12910-024-01042-y] [Medline: 38575931]
138. Goldfarb MJ, Saylor MA, Bozkurt B, Code J, Di Palo KE, Durante A, et al. Patient-centered adult cardiovascular care: a scientific statement from the American Heart Association. Circulation 2024 May 14;149(20):e1176-e1188 [FREE Full text] [doi: 10.1161/CIR.0000000000001233] [Medline: 38602110]
139. Dennison Himmelfarb CR, Beckie TM, Allen LA, Commodore-Mensah Y, Davidson PM, Lin G, et al. Shared decision-making and cardiovascular health: a scientific statement from the American Heart Association. Circulation 2023 Oct 12;148(11):912-931. [doi: 10.1161/CIR.0000000000001162] [Medline: 37577791]
140. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst 2021 Dec;32(11):4793-4813. [doi: 10.1109/TNNLS.2020.3027314] [Medline: 33079674]
141. Martin SA, Townend FJ, Barkhof F, Cole JH. Interpretable machine learning for dementia: a systematic review. Alzheimers Dement 2023 May;19(5):2135-2149 [FREE Full text] [doi: 10.1002/alz.12948] [Medline: 36735865]
142. Saqib K, Khan AF, Butt ZA. Machine learning methods for predicting postpartum depression: scoping review. JMIR Ment Health 2021 Dec 24;8(11):e29838 [FREE Full text] [doi: 10.2196/29838] [Medline: 34822337]
143. Ferreira-Santos D, Amorim P, Silva Martins T, Monteiro-Soares M, Pereira Rodrigues P. Enabling early obstructive sleep apnea diagnosis with machine learning: systematic review. J Med Internet Res 2022 Oct 30;24(9):e39452 [FREE Full text] [doi: 10.2196/39452] [Medline: 36178720]
144. International Bladder Cancer Nomogram Consortium, Bochner BH, Kattan MW, Vora KC. Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. J Clin Oncol 2006 Aug 20;24(24):3967-3972. [doi: 10.1200/JCO.2005.05.3884] [Medline: 16864855]
145. Papadimitroulas P, Brocki L, Christopher Chung N, Marchadour W, Vermet F, Gaubert L, et al. Artificial intelligence: deep learning in oncological radiomics and challenges of interpretability and data harmonization. Phys Med 2021 Mar;83:108-121 [FREE Full text] [doi: 10.1016/j.ejmp.2021.03.009] [Medline: 33765601]
146. Chen RJ, Wang JJ, Williamson DF, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng 2023 Jul 28;7(6):719-742 [FREE Full text] [doi: 10.1038/s41551-023-01056-8] [Medline: 37380750]
147. Huang J, Galal G, Etemadi M, Vaidyanathan M. Evaluation and mitigation of racial bias in clinical machine learning models: scoping review. JMIR Med Inform 2022 May 31;10(5):e36388 [FREE Full text] [doi: 10.2196/36388] [Medline: 35639450]
148. Vorisek CN, Stellmach C, Mayer PJ, Klopfenstein SA, Bures DM, Diehl A, et al. Artificial intelligence bias in health care: web-based survey. J Med Internet Res 2023 Jul 22;25:e41089 [FREE Full text] [doi: 10.2196/41089] [Medline: 37347528]
149. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. Lancet Digit Health 2020 May;2(5):e221-e223 [FREE Full text] [doi: 10.1016/S2589-7500(20)30065-0] [Medline: 33328054]
150. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med 2018 Mar 15;378(11):981-983 [FREE Full text] [doi: 10.1056/NEJMp1714229] [Medline: 29539284]

Abbreviations

AI: artificial intelligence
CA: cardiac arrest
CART: Cardiac Arrest Risk Triage
CPC 1-2: good cerebral performance category score 1 to 2
DT: decision tree
EMS: emergency medical services
GCS: Glasgow Coma Scale
IHCA: in-hospital cardiac arrest
LR: logistic regression
ML: machine learning

XSL•FO
RenderX

OHCA: out-of-hospital cardiac arrest
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
RF: random forest
ROSC: return of spontaneous circulation
SVM: support vector machine

Edited by A Mavragani; submitted 23.10.24; peer-reviewed by S Netherton, M Bak; comments to author 02.12.24; revised version received 19.12.24; accepted 16.01.25; published 10.03.25.

Please cite as:
Wei S, Guo X, He S, Zhang C, Chen Z, Chen J, Huang Y, Zhang F, Liu Q
Application of Machine Learning for Patients With Cardiac Arrest: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e67871
URL: https://www.jmir.org/2025/1/e67871
doi:10.2196/67871
PMID:40063076

©Shengfeng Wei, Xiangjian Guo, Shilin He, Chunhua Zhang, Zhizhuan Chen, Jianmei Chen, Yanmei Huang, Fan Zhang, Qiangqiang Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Factors Influencing eHealth Literacy Worldwide: Systematic Review and Meta-Analysis

Zhong Hua[1*], MSc, PhD; Song Yuqing[2,3*], PhD, MD; Liu Qianwen[1], BA, MA; Chen Hong[1], MD, PhD

[1]Department of Nursing, West China School of Nursing, West China Hospital, Sichuan University, Chengdu, China

[2]Department of Gynecological Nursing, West China Second University Hospital, Sichuan University, Chengdu, China

[3]Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Chen Hong, MD, PhD
Department of Nursing
West China School of Nursing, West China Hospital
Sichuan University
No 37, Guoxuexiang, Wuhou District, Sichuan
Chengdu, 610041
China
Phone: 86 18980601733
Email: 1366109878@qq.com

Abstract

Background: eHealth literacy has increasingly emerged as a critical determinant of health, highlighting the importance of identifying its influencing factors; however, these factors remain unclear. Numerous studies have explored this concept across various populations, presenting an opportunity for a systematic review and synthesis of the existing evidence to better understand eHealth literacy and its key determinants.

Objective: This study aimed to provide a systematic review of factors influencing eHealth literacy and to examine their impact across different populations.

Methods: We conducted a comprehensive search of papers from PubMed, CNKI, Embase, Web of Science, Cochrane Library, CINAHL, and MEDLINE databases from inception to April 11, 2023. We included all those studies that reported the eHealth literacy status measured with the eHealth Literacy Scale (eHEALS). Methodological validity was assessed with the standardized Joanna Briggs Institute (JBI) critical appraisal tool prepared for cross-sectional studies. Meta-analytic techniques were used to calculate the pooled standardized β coefficient with 95% CIs, while heterogeneity was assessed using I2, the Q test, and τ2. Meta-regressions were used to explore the effect of potential moderators, including participants' characteristics, internet use measured by time or frequency, and country development status. Predictors of eHealth literacy were integrated according to the Literacy and Health Conceptual Framework and the Technology Acceptance Model (TAM).

Results: In total, 17 studies met the inclusion criteria for the meta-analysis. Key factors influencing higher eHealth literacy were identified and classified into 3 themes: (1) actions (internet usage: β=0.14, 95% CI 0.102-0.182, I2=80.4%), (2) determinants (age: β=–0.042, 95% CI –0.071 to –0.020, I2=80.3%; ethnicity: β=–2.613, 95% CI –4.114 to –1.112, I2=80.2%; income: β=0.206, 95% CI 0.059-0.354, I2=64.6%; employment status: β=–1.629, 95% CI –2.323 to –0.953, I2=99.7%; education: β=0.154, 95% CI 0.101-0.208, I2=58.2%; perceived usefulness: β=0.832, 95% CI 0.131-1.522, I2=68.3%; and self-efficacy: β=0.239, 95% CI 0.129-0.349, I2=0.0%), and (3) health status factor (disease: β=–0.177, 95% CI –0.298 to –0.055, I2=26.9%).

Conclusions: This systematic review, guided by the Literacy and Health Conceptual Framework model, identified key factors influencing eHealth literacy across 3 dimensions: actions (internet usage), determinants (age, ethnicity, income, employment status, education, perceived usefulness, and self-efficacy), and health status (disease). These findings provide valuable guidance for designing interventions to enhance eHealth literacy.

Trial Registration: PROSPERO CRD42022383384; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022383384

(J Med Internet Res 2025;27:e50313) doi:10.2196/50313

XSL•FO
RenderX

KEYWORDS

meta-analysis; eHealth literacy; eHealth; Technology Acceptance Model; Literacy and Health Conceptual Framework; social determinants of health; digital health; consumer

Introduction

The links between literacy and health have attracted increasing international attention over the past decade [1]. Numerous studies have demonstrated that health outcomes, such as health behaviors and disease risks, are influenced by various forms of literacy, including health literacy, eHealth literacy, and other related literacy [2]. Especially in an era characterized by advanced information technology, the importance of eHealth literacy cannot be more pronounced due to the widespread adoption of digital health applications. eHealth literacy is defined as the capacity to search for, access, and analyze health information from electronic resources to solve health problems [3]. According to the Lily model of eHealth literacy, the 6 core literacies of eHealth literacy encompass essential aspects of health literacy and digital information literacy, including traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy [3]. Therefore, eHealth literacy not only directly influences the adoption of eHealth by users [3] but also is founded as a new determinant of health [4].

Studies have also found that eHealth literacy is positively associated with different aspects of health promotion and care. It influences an individual's acquisition, judgment, and application of health knowledge [5]. It also shapes their perception and attitude toward health and disease [6,7], especially in an era of advanced social media. eHealth literacy also plays a facilitating role in promoting the positive adoption of health-promoting behaviors [8-10]. Ultimately, this has a positive impact on health-related physical, behavioral, psychosocial, and cognitive outcomes [2,11,12]. Thus, understanding the determinants of eHealth literacy to identify intervention targets is of significant research value. Such efforts can enhance patients' eHealth literacy levels, ultimately contributing to improved health outcomes.

It is crucial for health care professionals to assess an individual's eHealth literacy before developing and conducting target interventions. Some tools have been developed to measure eHealth literacy. These include the eHealth Literacy Scale (eHEALS), the eHealth Literacy Questionnaire (eHLQ) [13], and the Digital Health Literacy Instrument (DHLI) [14]. eHEALS, developed by Norman and Skinner [15], is the most widely used scale to assess eHealth literacy. It comprises 8 items and encompasses 3 dimensions (application, evaluation, and decision-making capabilities related to online health information and services). eHEALS has been translated into different languages and validated as exhibiting robust reliability and validity. Additionally, numerous studies have evaluated individuals' eHealth literacy. However, the findings have not been systematically summarized. Therefore, this review aimed to synthesize the findings of studies that have used eHEALS to assess eHealth literacy. Although many studies have explored eHealth literacy among various populations, such as students, older adults, and the general population, the results have been inconsistent. Additionally, the influencing factors of eHealth literacy are diverse, including sociodemographic factors (ie, age, gender, and educational level), health-related behaviors, and internet usage. Despite the large number of cross-sectional surveys conducted to explore these factors, complex influences from both micro and macro levels have been identified, with results varying significantly [16]. The Literacy and Health Conceptual Framework, developed by the Literacy and Health Research Program in Canada, outlines 4 key themes: actions, determinants, literacy, and effects of literacy. Actions (eg, policy, training) and determinants (eg, sociodemographic status) can influence literacy (eg, general, health, and eHealth literacy). The Technology Acceptance Model (TAM) is a widely used theoretical framework. TAM posits that perceived usefulness and perceived ease of use directly influence an individual's attitude toward using a technology, in turn, affecting their intention to use the technology. TAM has been widely used to explore the user acceptance of various information systems and technologies. Thus, the Literacy and Health Conceptual Framework [1] and TAM can serve as theoretical frameworks to summarize and integrate the influencing factors of eHealth literacy [17,18].

Consequently, we systematically reviewed studies on eHealth literacy among different populations and synthesized their findings through a meta-analysis to understand the eHealth literacy status and examine its influencing factors.

Methods

Overview

This systematic literature review was conducted in accordance with the Cochrane Collaboration [19] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20] (Multimedia Appendix 1). The study was registered in PROSPERO (International Prospective Register of Systematic Reviews; ID CRD42022383384).

Search Strategy

We searched PubMed, CNKI, Embase, Web of Science, Cochrane Library, CINAHL, and MEDLINE databases from inception to April 11, 2023, to obtain a preliminary list of relevant studies. To develop search strategies, we checked 2 main sets of keywords: (1) eHealth literacy and corresponding synonyms (digital health literacy, electronic health literacy, e-health literacy, online health literacy, etc) and (2) influencing factors and corresponding synonyms. For both sets of keywords, we also reviewed the thesaurus and Medical Subject Headings (MeSH) terms. Next, we developed a search strategy for each database. An expert in information retrieval was consulted during this process. After formulating the initial search strategy, we conducted a preliminary search for each database. The search terms and results for each database were then sent to the expert for confirmation and refinement, ensuring the accuracy and comprehensiveness of the search strategy across all databases.

XSL•FO
RenderX

The detailed strategies for the different databases are presented in Multimedia Appendix 2. These search terms were used to search titles and abstracts in all the selected databases without applying any filters or limits.

Eligibility Criteria and Study Selection

We included all studies that were written in English or Chinese and reported the eHealth literacy status measured with eHEALS. For influencing factor analysis, we excluded (1) studies that did not examine the relationship between influencing factors and eHealth literacy or included only univariate analysis; (2) studies that did not provide effect size indicator values in the results or values that could not be converted to β, SE, and 95% CI values; and (3) reviews, case studies, poster presentations, and conference presentations. However, we examined their references to identify additional relevant papers for inclusion. We also manually searched the reference lists of studies in the final sample for additional relevant papers.

Data Extraction and Management

Two reviewers (authors HZ and QL) independently screened each identified study in duplicate for inclusion, initially reviewing the title and abstract, followed by a full-text review, as necessary. All discrepancies were resolved through consensus. We extracted the following data from each study: country, year conducted, sample size, questionnaire used to measure eHealth literacy, status of eHealth literacy, influencing factors with details, and effect size values. Where relevant data were missing or incomplete, we contacted the authors for clarification and verification.

Critical Appraisal of Methodological Quality

The included studies were critically appraised for methodological quality using the standardized Joanna Briggs Institute (JBI) critical appraisal tool prepared for cross-sectional studies and cohort studies [21] (Multimedia Appendix 3). Each study was evaluated by 2 independent assessors.

Data Analysis

Descriptive Statistics and Narrative Synthesis of the Studies in the Final Sample

Descriptive statistics were used to summarize the characteristics of the included studies. Narrative synthesis was used to synthesize the findings related to eHealth literacy in the studies, for which means (SDs) were reported. The year, country, sample size, study population, mean age (SD) of the sample, funding, mean score (SD) of eHEALS, and influencing factors were examined.

Random-Effects Meta-Analyses and Fixed-Effects Meta-Analyses to Examine Factors Affecting eHealth Literacy

Meta-analyses were conducted separately for eHealth literacy and each influencing factor using StataMP 17 software. Pooling was considered when at least 2 studies assessed eHealth literacy and its influencing factors. Standardized β coefficient estimates and their 95% CIs were calculated using random- or fixed-effects meta-analysis based on I^2 and the number of studies [22,23]. If $I^2 \leq 50\%$, the study was considered homogeneous and a fixed-effects model was used for meta-analysis. If $I^2 > 50\%$, the random-effects model was selected. If $I^2 > 85\%$ and the source of heterogeneity could not be determined, meta-analysis was not performed and descriptive analysis was used. In addition, the fixed effects model was considered if the number of studies that reported eHealth literacy's influencing factors was small (<5) [23,24]. Forest plots were created to show the results, and all comparisons were 2-tailed using a threshold of $P \leq .05$.

The heterogeneity of effect sizes was determined using the following statistics: (1) Q test of heterogeneity, (2) τ^2 estimate of true between-study variance, and (3) I^2 statistic of the proportion of true variation in observed effects. The Egger test was used to assess the possibility of publication bias [25].

Meta-Regression Analysis to Examine the Cofounders Affecting eHealth Literacy

The effect of moderators (participants' characteristics, internet use measured by time or frequency, country development status) was measured via random-effects univariate meta-regressions using restricted maximum-likelihood estimation. For the analysis, participants were categorized into 3 groups: students (coded as 1), the general population (coded as 2), and older adults (coded as 3). Regarding the country variable, studies conducted in low- to middle-income countries were coded as 1, while those conducted in high-income countries were coded as 2. The development status of a country was queried on the United Nations Big Data Global Platform [26]. To conduct cofounder analyses and obtain robust coefficient estimates, we followed the recommendations of Fu et al [27] and conducted moderator analysis only when at least 4 studies per group were available.

Assessment of the Quality of Evidence

The quality of evidence for the results of the meta-analyses was evaluated by authors QL and HZ using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) framework [28]. For each factor influencing eHealth literacy, the initial quality of evidence was considered high and was downgraded by 1 level for every significant concern identified within the domains of risk of bias, inconsistency, indirectness, imprecision, and publication bias. The risk of bias was evaluated by examining potential biases in participant selection, measurement instruments, data collection, and analysis. The inconsistency domain was assessed using I^2 values, and the GRADE quality was downgraded when I^2 was 50% or higher. Indirectness was evaluated based on whether the studies directly assessed the influence of various factors on eHealth literacy. Imprecision was determined by the width of the CIs of the estimates and the sample size. Publication bias was assessed using the Egger test, with GRADE quality being downgraded for statistically significant results ($P<.05$) in this test.

XSL•FO
RenderX

Theoretical Support for the Integration Framework of Influencing Factors

This study combined the Literacy and Health Conceptual Framework [1] and TAM [17,18] to synthesize and present factors that influence eHealth literacy, as identified in the literature review. The Literacy and Health Conceptual Framework (Figure 1), developed through the Literacy and Health Research Program sponsored by the Social Sciences and Humanities Research Council of Canada, comprises 4 key themes: actions, determinants, literacy, and effects of literacy. Actions and determinants can impact literacy and the effects of literacy. Each theme is further subdivided into various conceptual elements that have been delineated in greater detail. Arrows and line segments within the framework illustrate the interconnections and relationships among these elements. Actions include policy, training, community development, and communication. Determinants comprise sociodemographic status, living and working status, education, and personal ability. Literacy comprises health literacy, general literacy, and other literacy. In our study, we used the Literacy and Health Conceptual Framework as the primary structure to organize all influencing factors (Figure 2). We identified literacy as eHealth literacy and summarized its influencing factors into actions and determinants.

Figure 1. The Literacy and Health Conceptual Framework.

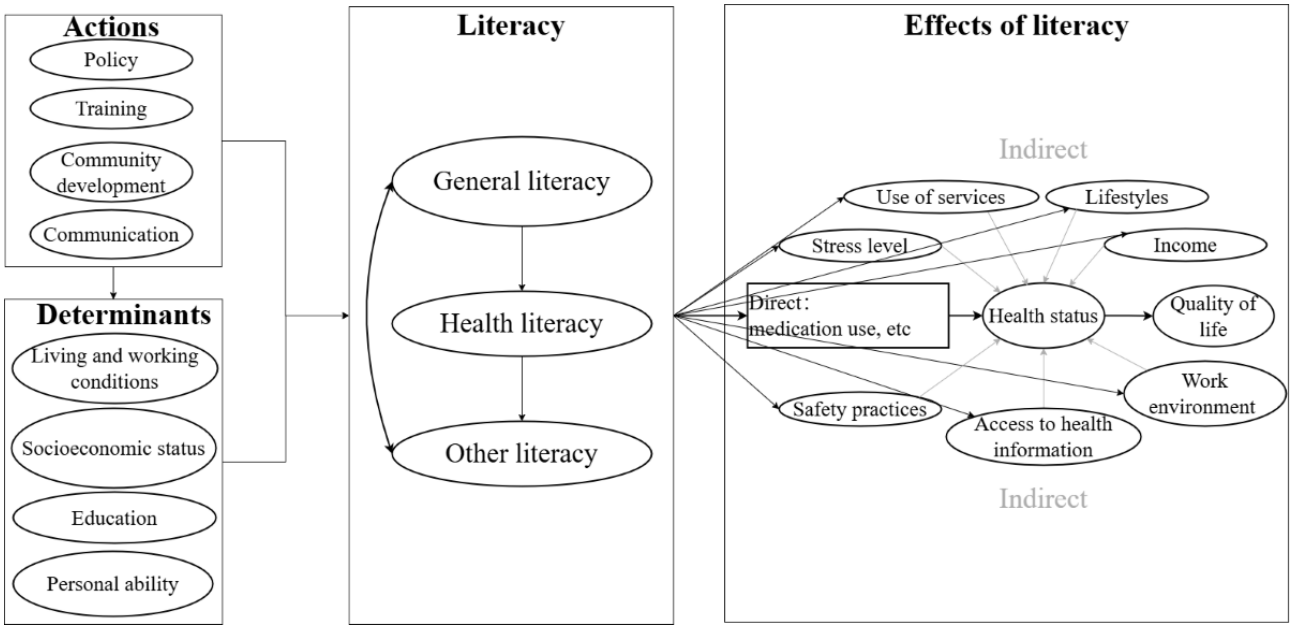

TAM, a theoretical framework developed by Venkatesh and coworkers [17,18], predicts individuals' adoption of new information technologies and has been extensively applied in eHealth domains. TAM posits that perceived usefulness and perceived ease of use mediate individuals' intention to use technology by shaping their attitudes toward it. The third iteration of this model, TAM3, expands on the original by integrating additional components, such as computer self-efficacy, which reflects an individual's confidence in their ability to perform tasks using a computer. Studies have found that perceived importance is strongly associated with both perceived usefulness and eHealth literacy [29-31]. In our study, we categorized key aspects of TAM3, including perceived usefulness, perceived importance, and self-efficacy, under personal capabilities. This emphasized that technology acceptance abilities are part of an individual's overall capabilities (Figure 1).

XSL•FO
RenderX

Figure 2. Factors influencing eHealth literacy in our analysis.

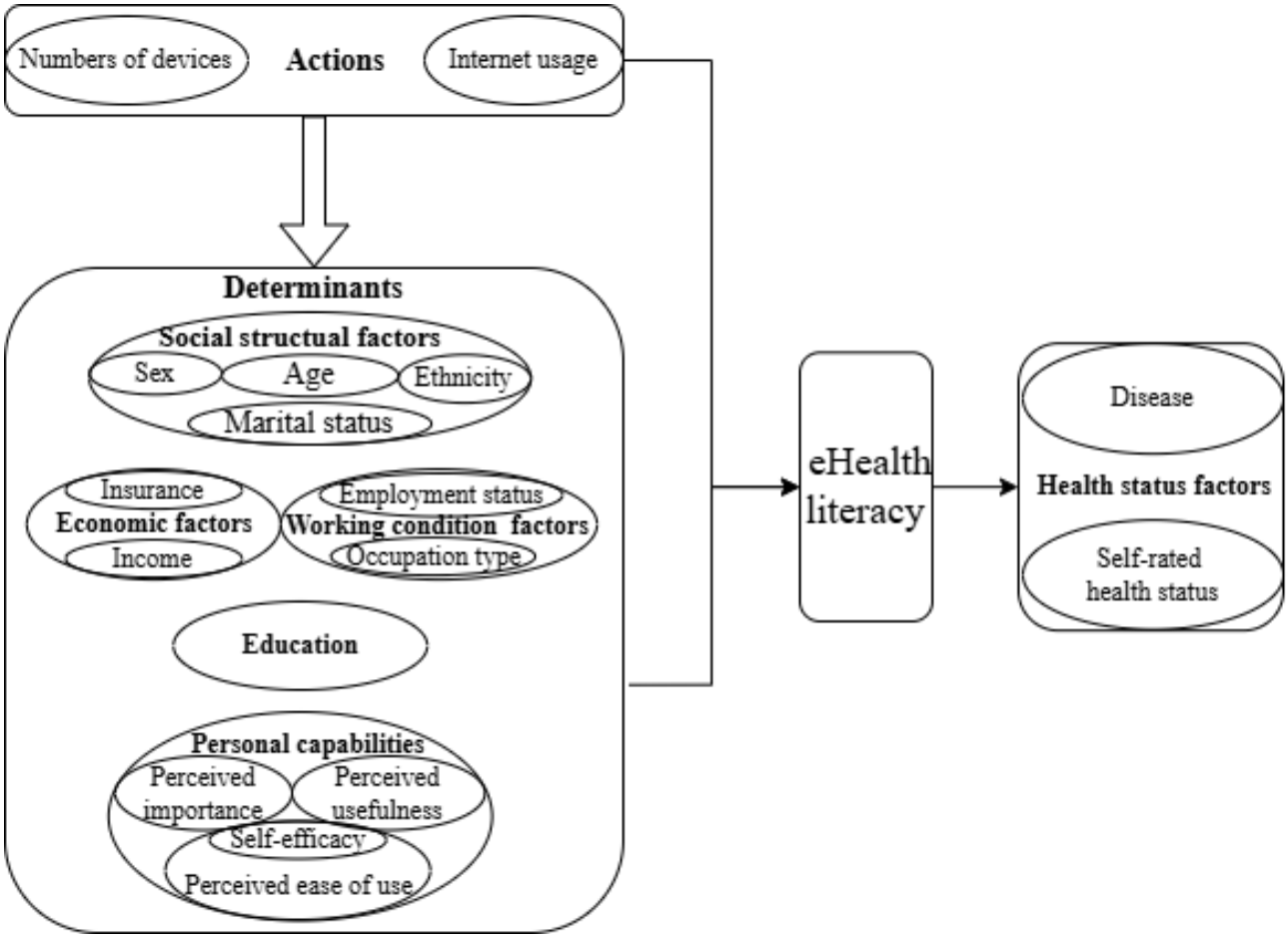

Results

Literature Selection Process and Characteristics of the Final Studies

The process of literature search and selection is shown in Figure 3. In total, 1295 potential papers were identified, but only 17 (1.3%) were identified as eligible and included in the final review to analyze the relationship between eHealth literacy and influencing factors (Figure 2). All the studies were cross-sectional in design (Table 1 and Table S1 in Multimedia Appendix 4). They were conducted in Turkey (n=1, 6%), China (n=1, 6%), Italy (n=1, 6%), South Korea (n=2, 12%), Serbia (n=1, 6%), Sweden (n=1, 6%), Australia (n=1, 6%), Vietnam (n=1, 6%), Kuwait (n=1, 6%), Saudi Arabia (n=1, 6%), Thailand (n=2, 12%), the United States (n=1, 6%), Malaysia (n=1, 6%), the United Kingdom (n=1, 6%), and Portugal (n=1, 6%). The study samples were the general population (n=3, 17.6%), older adults (n=3, 17.6%), nursing students (n=2, 12%), adult residents of urban and rural communities (n=1, 6%), undergraduate and postgraduate students (n=1, 6%), only undergraduate students (n=1, 6%), middle school students (n=1, 6%), adolescents (n=1, 6%), adults (n=1, 6%), medical students (n=2, 12%), and army personnel (n=1, 6%). Only 10 (59%) studies reported the age of participants. The mean age of students ranged from 20.13 (SD 2.16) to 24 (SD 1.48) years, and the mean age of adults ranged from 47 (SD not reported) to 78 (SD 2.4) years. All the studies were published in 2015 and later, and the majority (n=14, 82%) were published between 2018 and 2023. In addition, 15 (88%) studies had sample sizes greater than 100.

XSL•FO
RenderX

Figure 3. PRISMA flow diagram of the study selection process. eHEALS: eHealth Literacy Scale; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

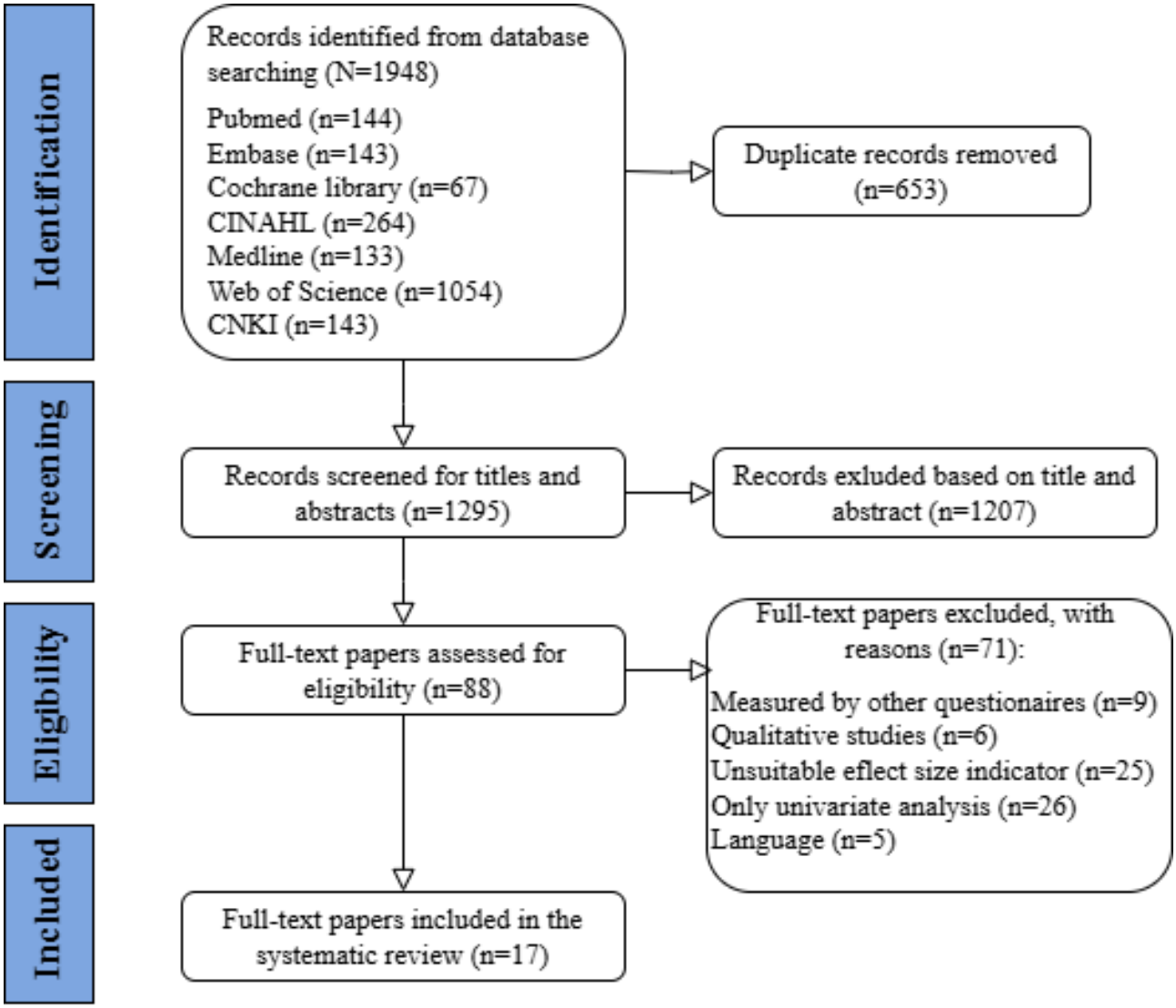

XSL•FO
RenderX

Table 1. Details of the 17 studies included in this meta-analysis.

Study	Country	Study population	Age (years), mean (SD)	eHealth literacy status[a], mean (SD)
Tennant et al [9]	United States	Older adults	67.46 (9.98)	29.05 (5.75)
Alhuwail and Abdulsalam [29]	Kuwait	General population	NR[b]	28.63 (5.6)
Sinan et al [31]	Turkey	Nursing students at nursing departments of 2 state universities in Ankara, Turkey	21.14 (1.62)	29.28 (4.73)
Zuo et al [32]	China	Adult residents of urban and rural communities	NR	32 (24.40)[c], 28 (21.34)[d]
Del Giudice et al [33]	Italy	Undergraduate and postgraduate students	NR	28.2 (6.2)
Park [34]	South Korea	Middle school students	NR	28.72 (5.68)
Gazibara et al [35]	Serbia	Adolescents (14-19 years)	NR	26.0 (10.0)
Ghazi et al [36]	Sweden	Older adult participants of the Swedish National Study on Aging and Care, Blekinge (SNAC-B)	78 (2.4)	27.5 (10.8)
Richtering et al [37]	Australia	Adults with moderate-to-high risk for a cardiovascular disease event	67 (8.0)	27.2 (6.59)
Hoang Nguyen and Bich Thi Le [38]	Vietnam	Medical students	NR	27.03 (3.54)
Almoajel et al [39]	Saudi Arabia	General population	24 (1.48)	28.79 (6.75)
Wongjinda and Taneepanichsakul [40]	Thailand	Royal Thai Army (RTA) personnel	NR	31.6 (NR)
Lee et al [41]	Malaysia	General population	47 (NR)	27.38 (6.59)
Kim and Jeon [42]	South Korea	Nursing students	21.69 (2.77)	29.68 (NR)
Holch and Marwood [43]	United Kingdom	Undergraduate students	20.13 (2.16)	29.46 (4.91)
Tanasombatkul et al [44]	Thailand	Medical students	22 (NR)	33.45 (3.28)
Martins et al [45]	Portugal	Older adults	67.36 (7.23)	NR

[a]Measured with the eHealth Literacy Scale (eHEALS).

[b]NR: not reported.

[c]Median (IQR) of eHEALS for rural community residents.

[d]Median (IQR) of eHEALS for urban community residents.

Quality of Included Studies

Based on the JBI critical appraisal tool, the overall quality of the included studies was high. Of the 17 studies, 6 (35%) achieved a perfect score of 100%, 7 (41%) scored 87.5%, 3 (18%) scored 75%, and only 1 (5%) scored 62.5% (Multimedia Appendix 3). This result demonstrated that the majority of the included studies (n=13, 76%) were rated as high-quality studies, with scores of 75% or higher [46].

Influencing Factors

According to the Literacy and Health Conceptual Framework [1], the factors influencing eHealth literacy were categorized into 3 themes: actions, determinants, and health status (Figure 2). Perceived importance, perceived usefulness, and self-efficacy, as defined by TAM3 [17], were integrated under the broader concept of personal capabilities. All the outcomes of the meta-analyses are presented in Table 2.

XSL•FO
RenderX

Table 2. Results of the analysis of influencing factors.

Theme and influencing factors	Studies (N=17), n (%); sample size, N	β (95% CI)	I^2 (%)	Egger test *P* value	GRADE[a] quality of evidence
Actions					
Number of devices	3 (18); 781	0.172 (–0.460 to 0.805)	0	.35	Moderate
Internet usage	10 (59); 4243	0.141 (0.102 to 0.182)[b]	80.4	.56	Low
Determinants: social determinants of health					
Age	13 (76); 4330	–0.042 (–0.071 to –0.020)[b]	80.3	.18	Low
Ethnicity	2 (12); 559	–2.613 (–4.114 to –1.112)[b]	80.2	N/A[c]	Very low
Marital status	4 (24); 1535	–0.143 (–0.354 to 0.069)	27.4	.84	Moderate
Sex	8 (47); 2812	–0.134(–0.561 to 0.304)	64.6	.30	Moderate
Determinants: economic and working conditions					
Income	6 (35); 1997	0.206 (0.059 to 0.354)[b]	64.6	.30	Moderate
Insurance	2 (12); 789	–0.007 (–0.199 to 0.184)	0	N/A	Moderate
Employment status	2 (12); 612	–1.629 (–2.323 to –0.953)[b]	99.7	N/A	Very low
Occupation type	2 (12); 916	–0.242 (–0.644 to 0.160)	42.7	N/A	Moderate
Determinants: personal capabilities					
Perceived importance	3 (18); 1745	0.563 (–0.198 to 1.325)	58.1	.14	Moderate
Perceived usefulness	2 (12); 686	0.832 (0.131 to 1.522)[b]	68.3	N/A	Low
Self-efficacy	2 (12); 393	0.239 (0.129 to 0.349)[b]	0	N/A	Moderate
Determinants: education	16 (94); 6331	0.154 (0.101 to 0.208)[b]	58.2	.86	Moderate
Health status					
Self-rated health status	3 (18); 1451	0.564 (–0.198 to 1.323)	17.3	.47	Moderate
Disease	7 (41); 2817	–0.177 (–0.298 to –0.055)[b]	26.9	.29	High

[a]GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

[b]The factor was significant in the meta-analysis.

[c]N/A: not applicable (as <3 experiments were analyzed).

Influencing Factors in the Actions Theme

The influencing factors in actions included the number of devices and internet usage. Three studies [9,38,44] identified an association between eHealth literacy and the number of devices (Figure 4). These studies examined whether the number of electronic devices an individual owns is correlated with their eHealth literacy. However, the results of our analysis did not yield any statistically significant findings. Eight studies [29,31,33,37,40,42,44,45] demonstrated a positive correlation between eHealth literacy and internet usage (β=0.14, 95% CI 0.10-0.18).

Figure 4. Forest plot of the meta-analytic association between the number of devices and eHealth literacy. IV: inverse variance weighted.

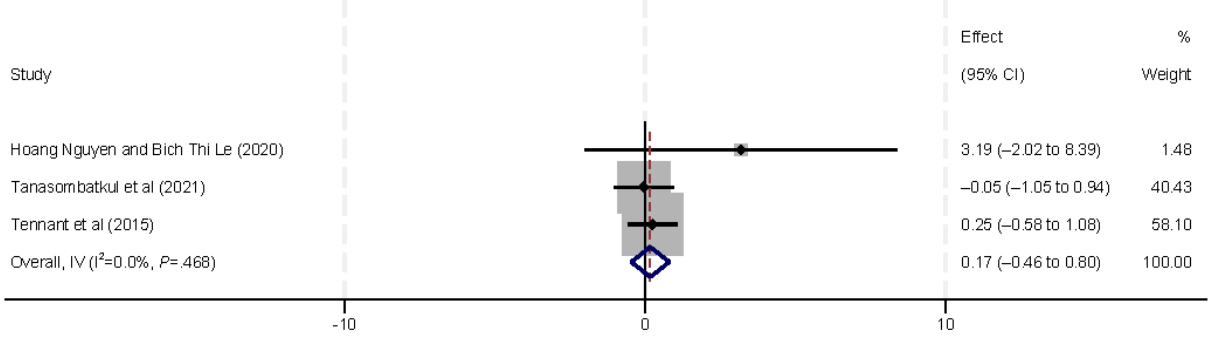

XSL•FO
RenderX

Furthermore, we conducted subgroup analyses to examine internet use measured by duration or frequency (Figure 5). The results from both subgroups showed a positive correlation between internet usage and eHealth literacy, indicating that higher levels of internet use are associated with increased eHealth literacy.

Figure 5. Forest plot of the meta-analytic association between internet usage and eHealth literacy. IV: inverse variance weighted.

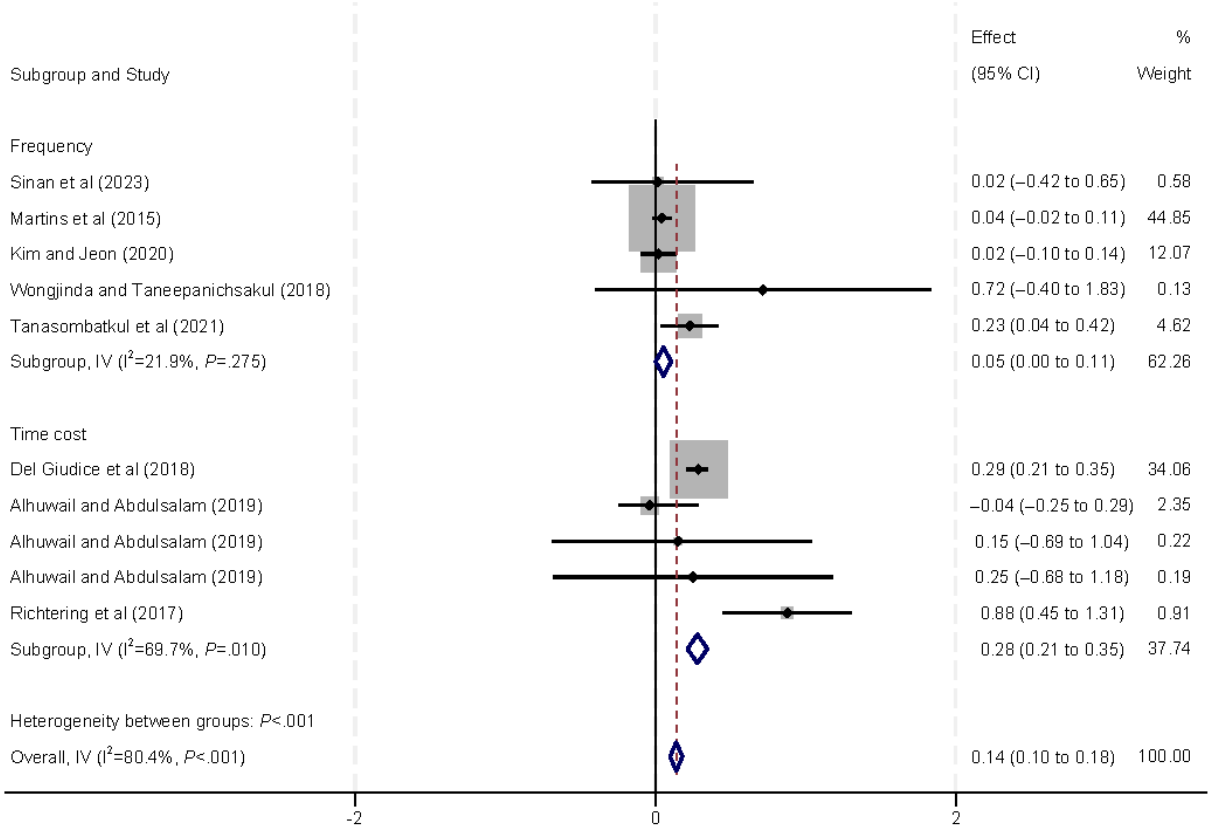

Influencing Factors in the Determinants Theme

Social determinants of health, economic and working conditions, education, and personal capabilities were included under the determinants theme.

Social Determinants of Health

The following social determinants of health were analyzed to assess their relationship with eHealth literacy: age, ethnicity, marital status, and sex. The age of participants was significantly associated with eHealth literacy (Figure 6). Only 1 (13%) [33] of the 8 (47%) studies assessing eHealth literacy found a positive correlation with age, with the other 9 (53%) studies [9,29,37-42,44] reporting a negative correlation. Subgroup analyses based on age revealed that studies conducted among students showed a positive correlation between eHealth literacy and age (β=0.10, 95% CI 0.04 to –0.17). Conversely, in the general population (β=–0.05, 95% CI –0.08 to –0.02) and among older adults (β=–0.21, 95% CI –0.29 to –0.14), eHealth literacy was found to be negatively correlated with age.

XSL•FO
RenderX

Figure 6. Forest plot of the meta-analytic association between age and eHealth literacy. IV: inverse variance weighted.

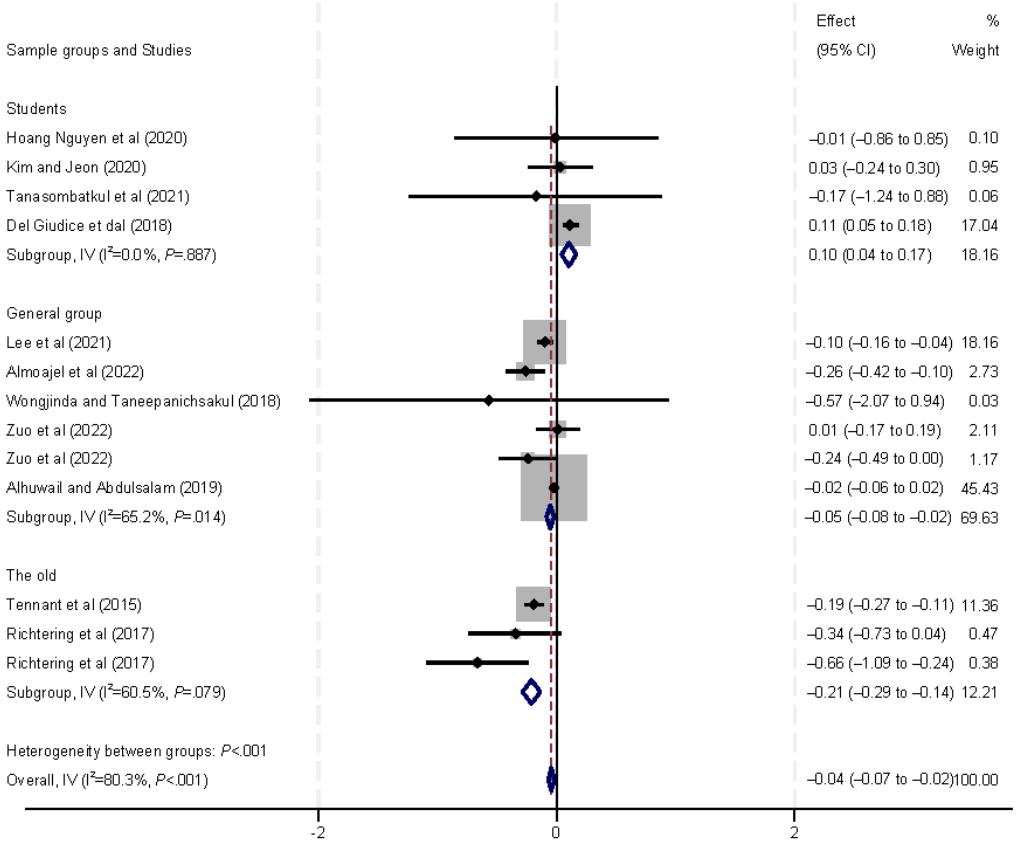

Ethnicity was also found to affect eHealth literacy (β=–2.61, 95% CI –4.11 to –1.11) [9,41]. Lee et al [41] found that lower eHealth literacy is associated with patients of minority ethnicity (Malaysian Chinese). However, Tennant et al [9] found that ethnicity is not significantly associated with eHealth literacy (Figure 7). Marital status was also included in the analysis (Figure 8). However, no statistically significant association was found, consistent with the results of the 3 (18%) studies [9,32,39].

Figure 7. Forest plot of the meta-analytic association between ethnicity and eHealth literacy. IV: inverse variance weighted.

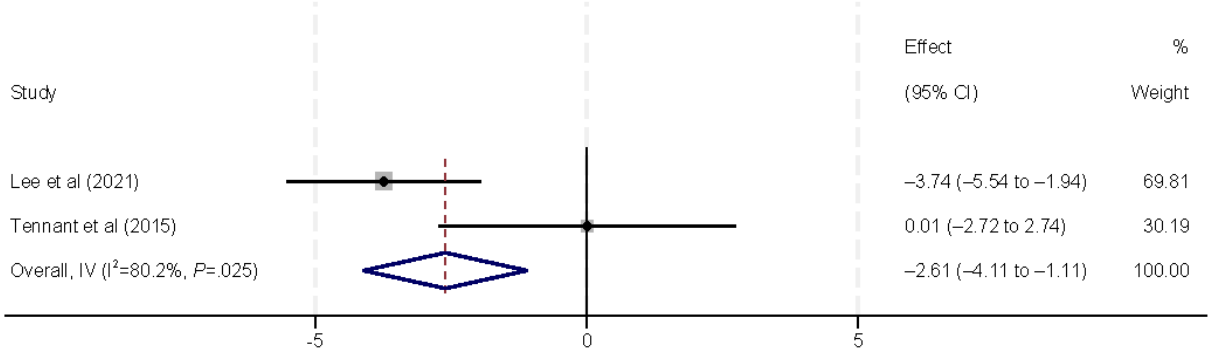

XSL•FO
RenderX

Figure 8. Forest plot of the meta-analytic association between marital status and eHealth literacy. IV: inverse variance weighted.

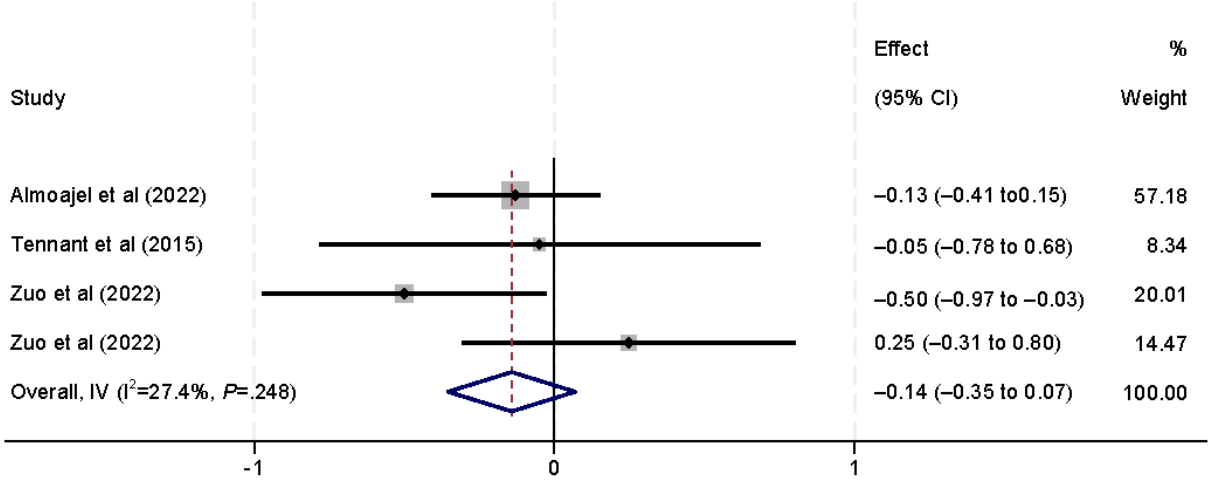

Seven studies [9,29,32,37,38,41,44] assessed the relationship between sex and eHealth literacy using multivariate analysis (Figure 9). Of these, 6 (15%) studies [9,29,32,37,41,44] found that women are at a higher risk of low eHealth literacy than men. However, the final meta-analysis results showed no statistically significant association.

Figure 9. Forest plot of the meta-analytic association between sex and eHealth literacy. IV: inverse variance weighted.

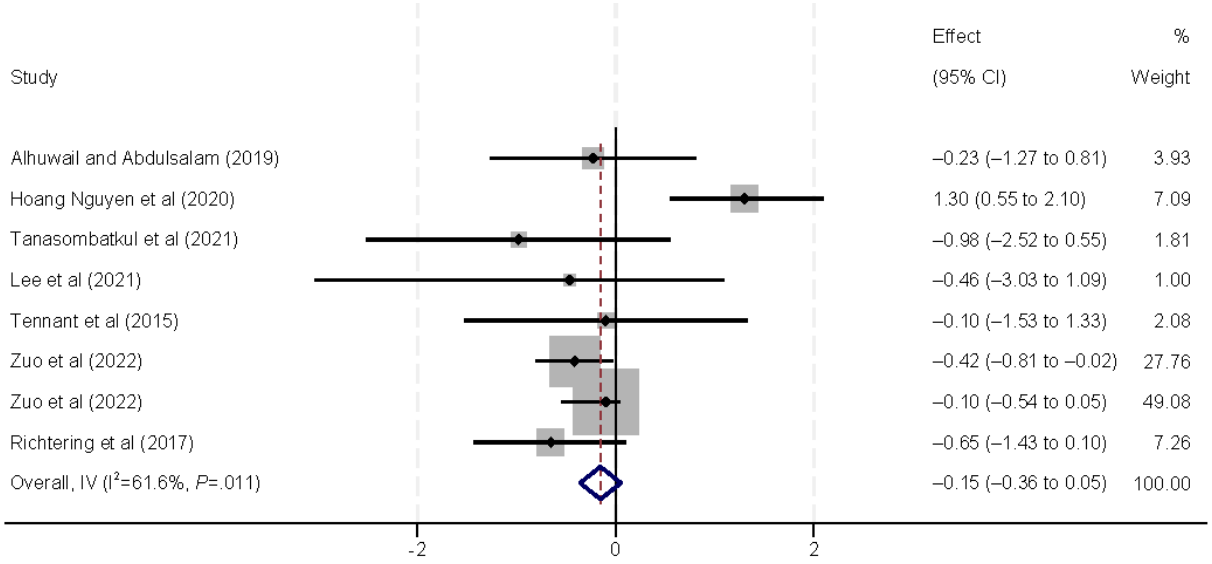

Economic and Working Conditions

Regarding economic and working conditions, we analyzed eHealth literacy in relation to economic aspects (eg, income and insurance coverage), as well as working conditions (eg, employment status and occupation type). All 5 (29%) included studies [9,32,38,40,44] found a positive correlation between income and eHealth literacy (Figure 10), which was also confirmed by the final meta-analysis (β=0.21, 95% CI 0.06-0.35). However, in subgroup analyses, the correlation between income and eHealth literacy was only confirmed in the general population (β=0.63, 95% CI 0.36-0.91). Private insurance is also considered an economic factor. Neither of the 2 (12%) included studies [37,39] found a correlation between the presence or absence of private insurance and the level of eHealth literacy, which aligns with the findings of our meta-analysis (Figure 11). Unemployment status was identified as a risk factor for eHealth literacy levels by Almoajel et al [39] and Lee et al [41]. However, our meta-analysis revealed substantial heterogeneity, which prevented reliable statistical inferences (Figure 12). Zuo et al [32] observed variations in eHealth literacy according to occupation type, but these findings were not supported by the final results of our meta-analysis (Figure 13).

XSL•FO
RenderX

Figure 10. Forest plot of the meta-analytic association between income and eHealth literacy. IV: inverse variance weighted.

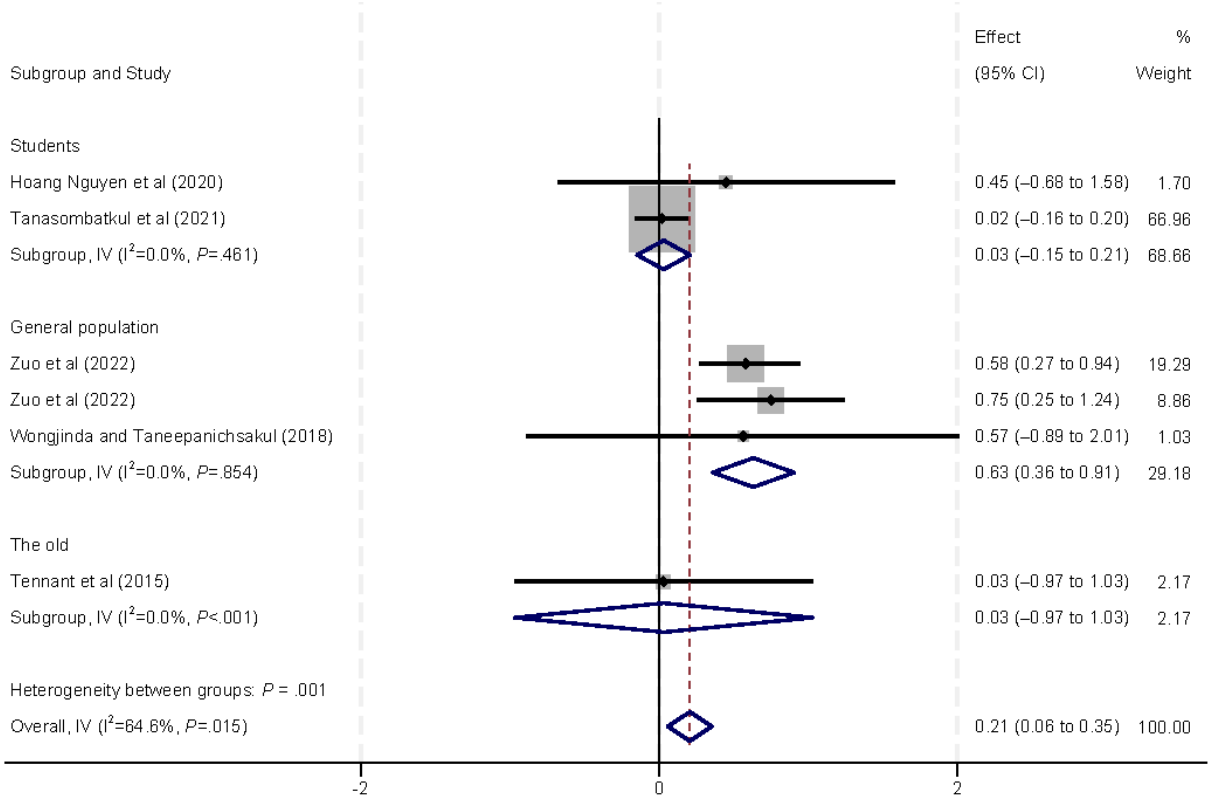

Figure 11. Forest plot of the meta-analytic association between insurance and eHealth literacy. IV: inverse variance weighted.

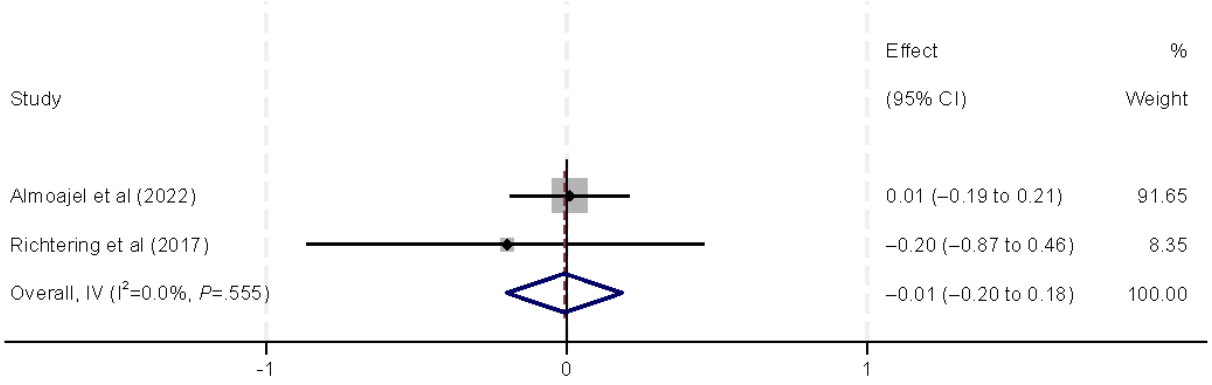

Figure 12. Forest plot of the meta-analytic association between employment status and eHealth literacy. DL: DerSimonian and Laird.

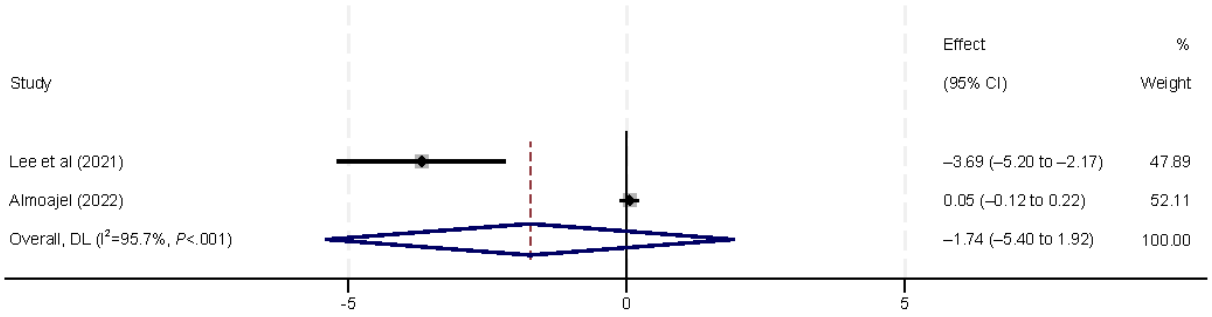

XSL•FO
RenderX

Figure 13. Forest plot of the meta-analytic association between occupation type and eHealth literacy. IV: inverse variance weighted.

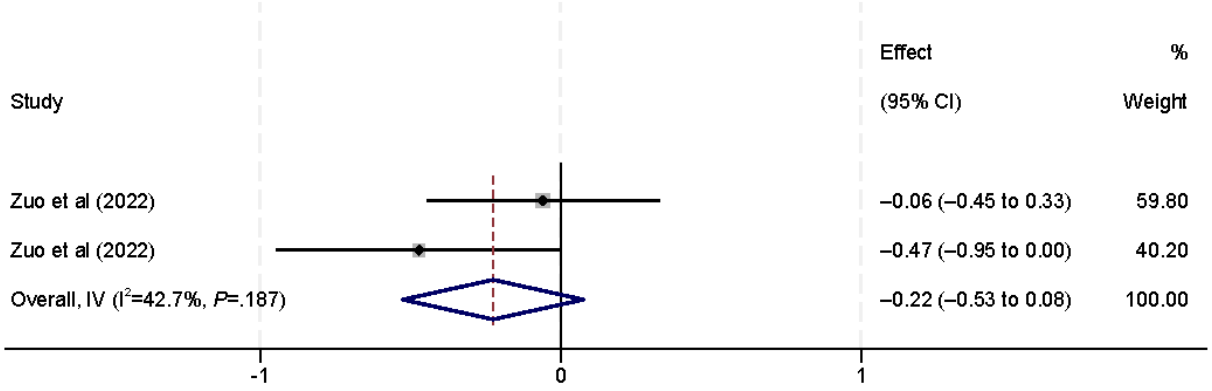

Education

As shown in Figure 14, 10 (59%) studies [9,29,31-33,37,39,40,42,45] identified lower educational status or a lack of formal education as being associated with lower eHealth literacy (β=0.15, 95% CI 0.10-0.21).

Figure 14. Forest plot of the meta-analytic association between education and eHealth literacy. DL: DerSimonian and Laird.

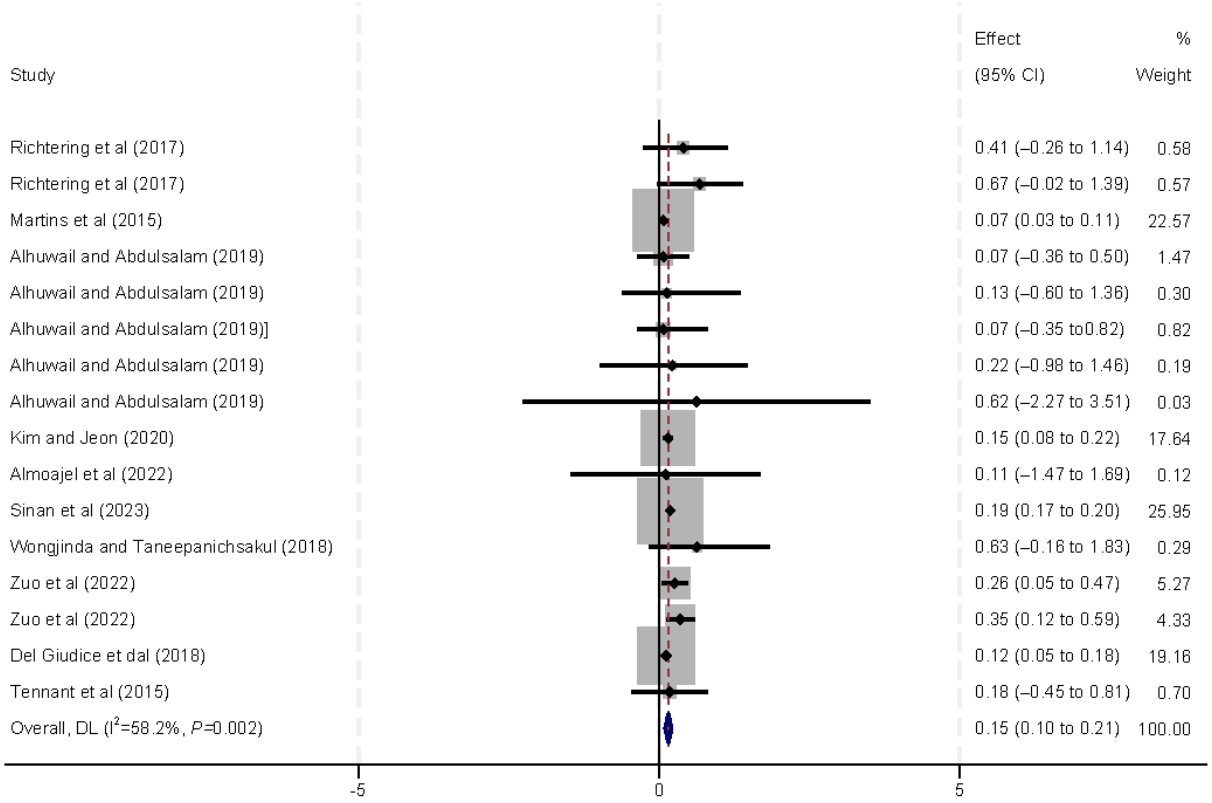

Personal Capabilities

The personal capabilities of perceived importance, perceived usefulness, and self-efficacy were found to significantly influence eHealth literacy (Figures 15-17). Perceived importance [29,31,40], perceived usefulness (β=0.83, 95% CI 0.13-1.52) [29,40], and self-efficacy (β=0.24, 95% CI 0.13-0.35) [42,43] were positively correlated with eHealth literacy levels in all included studies.

XSL•FO
RenderX

Figure 15. Forest plot of the meta-analytic association between perceived importance and eHealth literacy. DL: DerSimonian and Laird.

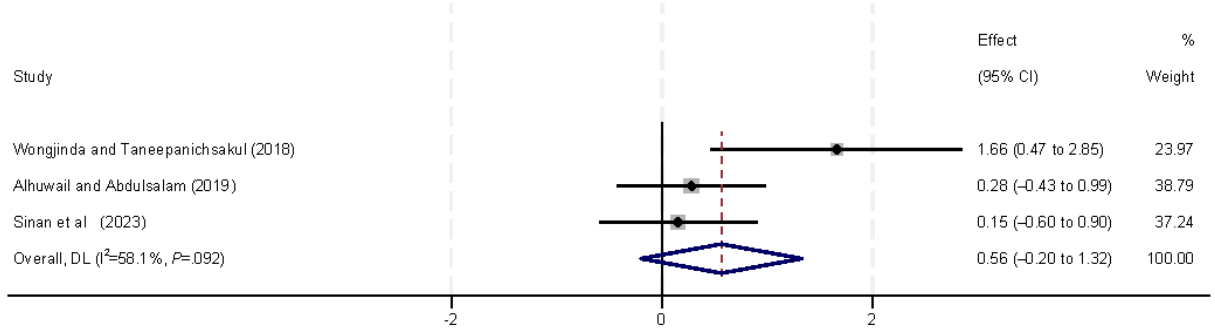

Figure 16. Forest plot of the meta-analytic association between perceived usefulness and eHealth literacy. IV: inverse variance weighted.

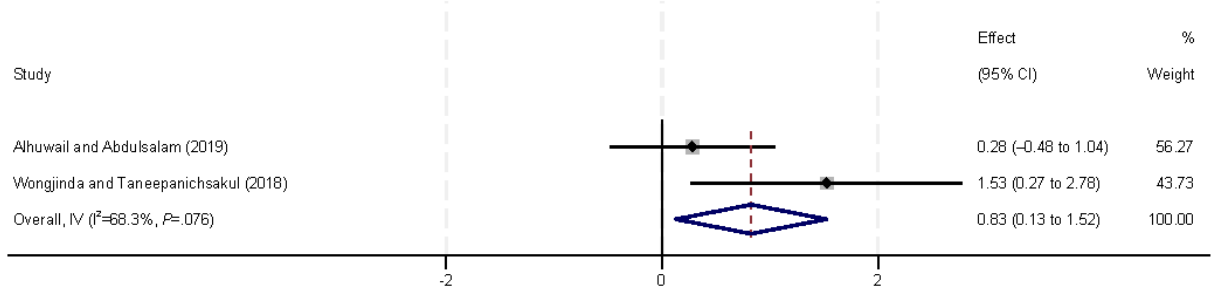

Figure 17. Forest plot of the meta-analytic association between self-efficacy and eHealth literacy. IV: inverse variance weighted.

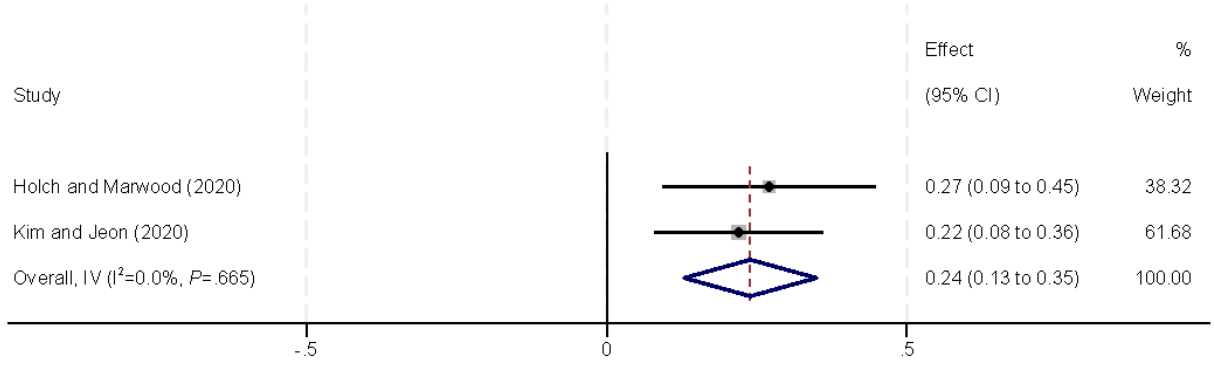

Influencing Factors in the Health Status Theme

Health status, specifically self-rated health and disease status, was found to significantly influence eHealth literacy (Figures 18 and 19). According to Del Giudice et al [33], Tennant et al [9], and Wongjinda and Taneepanichsakul [40], individuals with better self-rated health exhibit higher levels of eHealth literacy. The 6 (35%) included studies [32,34,37,40,41,44] found that participants with existing disease or at a higher risk for disease have lower levels of eHealth literacy (β=–0.15, 95% CI –0.8 to –0.16).

Figure 18. Forest plot of the meta-analytic association between self-rated health and eHealth literacy. IV: inverse variance weighted.

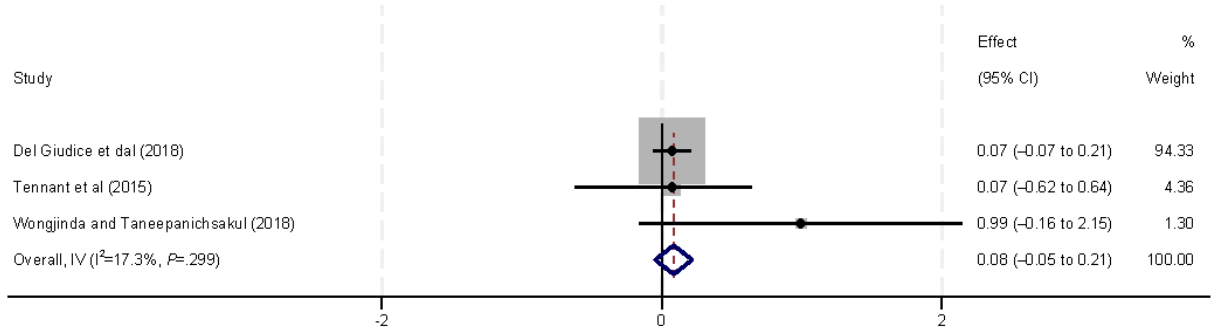

XSL•FO
RenderX

Figure 19. Forest plot of the meta-analytic association between disease and eHealth literacy. IV: inverse variance weighted.

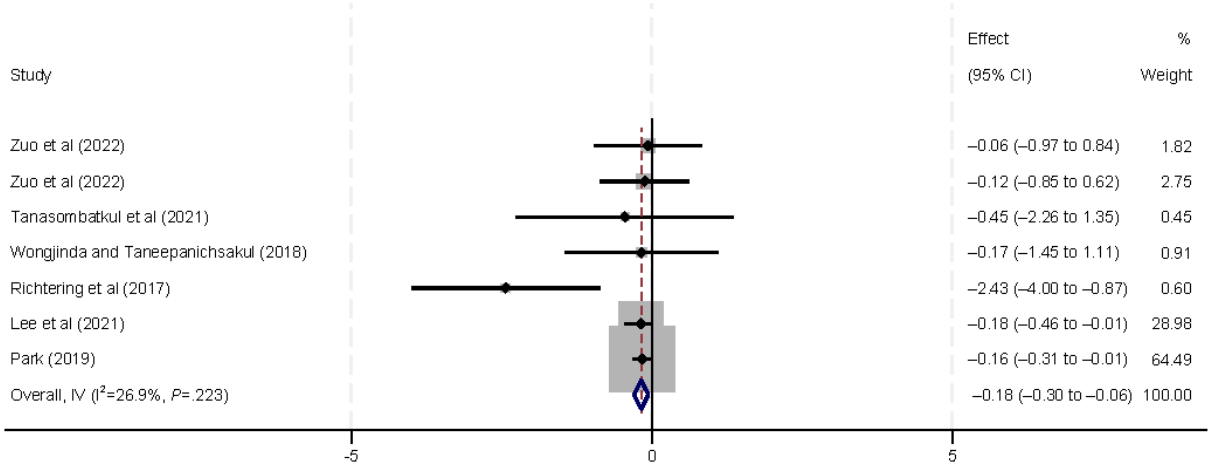

Meta-Regression Analyses to Examine Cofounders Affecting eHealth Literacy

In the meta-regression analysis of confounding factors, only age grouping significantly moderated the relationship between income and eHealth literacy (Table 3). For other influencing factors, neither age nor the developmental level of the country where the study was conducted was identified as a significant confounding factor.

Table 3. Confounder analysis results.

Confounder and influencing factors	β coefficient (SE)	*t* test (*df*)	*P* value	I^2 (%)	τ^2
Age group					
Internet usage[a]	0.06 (0.13)	0.49 (7)	.64	67.0	0.072
Sex	–0.323 (–0.78)	0.42 (6)	.69	70.2	0.397
Income[a]	0.60 (0.17)	3.63 (4)	.04[b]	100.0	0
Marital status	0.09 (0.51)	0.19 (2)	.87	50.8	0.0481
Education[a]	–0.018 (0.019)	–0.98 (14)	.35	0	0.001
Disease[a]	–0.91 (0.37)	–2.48 (5)	.07	0	0
Country					
Internet usage[a]	0.08 (0.22)	0.37 (7)	.73	67.0	0.072
Sex	0.18 (1.53)	0.12 (6)	.91	70.2	0.397
Income[a]	–1.20 (–0.59)	2.05 (4)	.13	100.0	0
Marital status	0.09 (0.51)	0.19 (2)	.87	50.8	0.04814
Education[a]	–0.07 (0.03)	–2.10 (14)	.06	0	0.001
Disease[a]	–0.89 (0.37)	–2.35 (5)	.08	0	0

[a]Significant in meta-analyses.

[b]Significant *P* value.

Discussion

Principal Findings

We summarized and analyzed the findings from relevant scientific studies, highlighting significant variability in the reported factors influencing eHealth literacy across different study populations. Our meta-analysis indicated that the majority of influencing factors are demographic variables (eg, age, ethnicity) and social determinants of health (eg, income, education). Additionally, we identified factors amenable to clinical intervention, such as internet usage, self-efficacy, and the perceived usefulness of eHealth. However, most of these factors did not achieve a high-level grade according to the GRADE framework.

Influencing Factors of eHealth Literacy

Numerous cross-sectional studies have identified various influencing factors of eHealth literacy. This study categorized the influencing factors into 3 main themes based on the Literacy

XSL•FO
RenderX

and Health Conceptual Framework and TAM. In the *actions* theme, this study revealed that internet usage is positively associated with eHealth literacy. A possible explanation is that eHealth literacy encompasses an individual's comprehensive ability to seek, discover, understand, evaluate, and apply health information obtained through electronic channels to address health-related issues. According to the eHealth literacy Lily model, individuals who use the internet more frequently tend to exhibit higher levels of information literacy, media literacy, and computer literacy [15]. This enables them to find and access health information through electronic channels more easily, thereby enhancing their eHealth literacy. Furthermore, some studies have identified a significant negative correlation between computer use stress and eHealth literacy among users [47]. Additionally, users with higher internet usage frequency tend to exhibit stronger self-efficacy in using eHealth technologies, which may reduce their computer use stress and, consequently, enhance their ability to access and use internet health information and services [48]. Recent studies have also confirmed that eHealth literacy can be improved by internet use interventions [49]. Therefore, it is beneficial for health care professionals to provide health education that enhances internet skills, such as training in searching for and evaluating health information.

In the *determinants* theme, various social determinants of health and other complex demographic factors are included. Factors such as age, ethnicity, economic status, and education are closely associated with eHealth literacy [9,14]. The effect of age varies across different population subgroups. Among student groups, eHealth literacy is positively correlated with age, whereas in the general population and among older adults, the correlation is negative. According to socioemotional selectivity theory [50], individual needs change with age. Young people tend to engage in growth-related activities, such as accumulating knowledge and advancing their careers [51]. Consequently, students are typically more active in digital environments, due to both their educational settings and their personal interests, which increasingly leverage information technology for learning and communication [52]. This constant interaction likely enhances their eHealth literacy as they progress through their academic careers. As people grow older, their values shift toward meaning and social needs [51]. When users are not yet aware of their health-related needs, they are less likely to proactively seek health information, hindering actions that could enhance their eHealth literacy. Additionally, research has shown that with aging, especially in the older adult phase, cognitive and physiological decline makes it more challenging to perceive and manage eHealth-related technologies [53]. This reduced ease of use can result in lower adoption of eHealth technologies, making it challenging to improve eHealth literacy levels. Other social determinants of health, such as ethnicity, economic status, and education, contribute to this issue in complex and often unmodifiable ways, closely intertwined with the digital divide [9]. A survey conducted by Choi and Dinitto [6] revealed that low-income and older populations exhibit significantly lower levels of eHealth literacy and internet usage compared to the general population in the United States. This disparity can be attributed to factors such as limited access to computers and internet technologies, financial constraints that prevent the acquisition of personal devices, and limitations caused by medical conditions and disabilities [6]. Modifiable factors that can be targeted for intervention include self-efficacy [54] and the perceived usefulness of eHealth. Self-efficacy reflects an individual's confidence in the ability [55], and eHealth literacy is defined as an ability [3]. It is reasonable to hypothesize a positive correlation between self-efficacy and eHealth literacy. Moreover, self-efficacy explains the eHealth-adopting motivation, behaviors, and awareness of appraisal and application skills [56]. Perceived usefulness is significantly correlated with eHealth literacy, even as a supplemental question for eHEALS [29,30,37]. This relationship can be explained by TAM [18]. Perceived usefulness is one of the core beliefs that determines individuals' intention to use IT, defined as the extent to which a person believes that using IT will enhance their performance [17]. Some researchers have proved that individuals with high perceived usefulness may have a more active attitude toward eHealth use [57,58]. In summary, the association between eHealth literacy and background attributes is difficult to modify. However, meso- and macrolevel approaches can be used. At the macro level, these include the level of government support and policy direction [16], for instance, providing technology subsidies or allowances to low-income individuals. Government agencies and nongovernmental organizations (NGOs) can play a significant role in recycling and refurbishing unused or discarded computers for older adults and low-income individuals. These initiatives can be cost-effective, enabling individuals to live more independently, reducing their reliance on both formal and informal support systems, and enhancing their quality of life [6]. The meso level encompasses various characteristics related to the planning, design, and development of eHealth services [16]. Emphasizing user-friendliness is essential. For example, recommendations including simplifying operational interfaces and steps optimize workflows, create more visually prominent interfaces, provide clearer system prompts, and replace technical terms with plain language and illustrations in designs tailored for older adults [57]. From an individual's micro perspective, interventions can focus on enhancing self-efficacy and perceived usefulness. Older adults who are hesitant to use the internet due to low self-efficacy in technology can be encouraged through demonstrations and educational interventions. Younger individuals with strong computer and internet skills can assist by volunteering or being employed to teach older adults how to engage with social media, browse the web, and participate in health-related activities. Particularly for low-income, homebound older adults who have limited exposure to computers and internet technology but substantial needs, it is important to emphasize the multiple benefits of internet use and to provide resources, such as equipment and training [6].

In the *health status* theme, disease was identified as a negative influencing factor for eHealth literacy [37,40,41,44]. Previous research indicates that individuals with higher eHealth literacy often achieve better health outcomes [12,59]. However, the relationship between eHealth literacy and health outcomes is complex and not strictly linear. Our study suggests a substantial overlap between populations with low eHealth literacy and those experiencing poor health conditions. Both groups are influenced by social determinants of health and are more likely to be older,

XSL•FO
RenderX

less educated, and of a lower socioeconomic status. Additionally, various studies have identified mediating factors in the relationship between health outcomes and eHealth literacy. For example, a cross-sectional study by Li et al [60] involving 802 Chinese residents revealed that social media use mediates the relationship between eHealth literacy and health-promoting behaviors. Similarly, Rabenbauer and Mevenkamp's [61] study of 224 Australian patients with chronic back pain identified self-efficacy as a mediator in the relationship between eHealth literacy and healthy habits. Therefore, further research is needed to clarify the relationship between eHealth literacy and health outcomes, particularly by exploring this connection after controlling for confounding factors. Future research could also apply health- and literacy-related theories to investigate whether additional factors moderate this relationship.

Limitations

This study has several limitations that should be acknowledged. First, only studies using the eHEALS measurement tool were included in this review. Insufficient data from other research instruments, such as the eHLQ and the Digital Health Literacy Questionnaire (DHLQ), prevented a comprehensive synthesis of findings across different tools. Second, papers written in languages other than English and Chinese were excluded from this review, potentially introducing language and publication bias. Third, due to the limited number of randomized controlled trials (RCTs) and the inability to combine their effect sizes, RCTs were not included in the meta-analyses. This exclusion is a significant limitation, as RCTs are considered the gold standard for establishing causal relationships. Their exclusion means that the meta-analysis relied more heavily on observational studies, which are more susceptible to confounding factors and bias. This reliance can limit the ability to draw firm conclusions about the causal relationship between eHealth literacy and health outcomes. Fourth, the standardized beta coefficient (β) was used as the effect size measure for data integration in this study, allowing for the inference of correlation relationships but not supporting causal inference. Moreover, the majority of studies included in the meta-analysis focused on student and adolescent populations, which limits the generalizability of the findings to broader populations. Additionally, some of the results reported I^2 exceeding 50%, indicating substantial heterogeneity. This high level of heterogeneity is expected, as eHealth literacy is significantly influenced by social and demographic characteristics, as well as variability in the measurement of influencing factors. This heterogeneity limits the generalizability of the identified factors influencing eHealth literacy. Therefore, the results of the meta-analysis and meta-regression should be interpreted with caution. Finally, all studies included in this research categorized gender using binary terms. This approach overlooks the experiences and outcomes of nonbinary and transgender individuals, potentially leading to a less comprehensive understanding of how eHealth literacy impacts various gender identities. We encourage future research to explore whether differences exist.

Conclusion

Based on the Literacy and Health Conceptual Framework, we explored the impact of various factors on users' eHealth literacy across 3 dimensions: actions, determinants, and health status. We found that users' eHealth literacy is influenced by various factors, including internet usage, age, ethnicity, income, employment status, education, perceived usefulness, self-efficacy, and disease. These findings provide valuable guidance for designing interventions to enhance eHealth literacy.

Acknowledgments

This work was funded by the Science & Technology Department of the Sichuan Province Program (2023YFS0045).

Authors' Contributions

ZH was responsible for writing the manuscript and data processing, SY for writing and editing the manuscript, LQ for data selection, and CH for draft touch-ups and guidance on writing. All authors contributed to the paper and have approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISIMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 33 KB - jmir_v27i1e50313_app1.docx]

Multimedia Appendix 2
Detailed search strategies for different databases.
[DOCX File , 15 KB - jmir_v27i1e50313_app2.docx]

Multimedia Appendix 3
Critical appraisal of methodological quality.
[DOCX File , 27 KB - jmir_v27i1e50313_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Details of studies included for meta-analysis (N=17).
[DOCX File , 43 KB - jmir_v27i1e50313_app4.docx]

References

1. Rootman I, Gordon-El-Bihbety D, Frankish J, Hemming H, Kaszap M, Langille L, et al. National literacy and health research program needs assessment and environmental scan. Academia.edu. 2000. URL: https://tinyurl.com/3nu3s2hm [accessed 2025-02-18]
2. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions: a systematic review. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
3. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
4. Arias López MDP, Ong BA, Borrat Frigola X, Fernández AL, Hicklent RS, Obeles AJT, et al. Digital literacy as a new determinant of health: a scoping review. PLOS Digit Health 2023 Oct;2(10):e0000279 [FREE Full text] [doi: 10.1371/journal.pdig.0000279] [Medline: 37824584]
5. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
6. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
7. Jeon J, Kim S. The mediating effects of digital literacy and self-efficacy on the relationship between learning attitudes and ehealth literacy in nursing students: a cross-sectional study. Nurse Educ Today 2022 Jun;113:105378. [doi: 10.1016/j.nedt.2022.105378] [Medline: 35504068]
8. Islam MM, Touray M, Yang H, Poly TN, Nguyen P, Li YJ, et al. E-health literacy and health information seeking behavior among university students in Bangladesh. Stud Health Technol Inform 2017;245:122-125. [Medline: 29295065]
9. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
10. Xu RH, Zhou L, Wong EL, Wang D. The association between patients' eHealth literacy and satisfaction with shared decision-making and well-being: multicenter cross-sectional study. J Med Internet Res 2021 Sep 24;23(9):e26721 [FREE Full text] [doi: 10.2196/26721] [Medline: 34559062]
11. Ojelabi AO, Graham Y, Haighton C, Ling J. A systematic review of the application of Wilson and Cleary health-related quality of life model in chronic diseases. Health Qual Life Outcomes 2017 Dec 11;15(1):241 [FREE Full text] [doi: 10.1186/s12955-017-0818-2] [Medline: 29228977]
12. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PK. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022 Apr;157:106997. [doi: 10.1016/j.ypmed.2022.106997] [Medline: 35189203]
13. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
14. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
15. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
16. Norgaard O. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowledge Manag E-Learn: Int J 2015 Dec 15:522-540. [doi: 10.34105/j.kmel.2015.07.035]
17. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decis Sci 2008 May 09;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
18. Venkatesh V, Davis FD. A theoretical extension of the Technology Acceptance Model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
19. Julian H, James T. Cochrane Handbook for Systematic Reviews of Interventions, Version 63. 2022. URL: https://training.cochrane.org/handbook/current [accessed 2023-06-12]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
21. Ma L, Wang Y, Yang Z, Huang D, Weng H, Zeng X. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res 2020 Feb 29;7(1):7 [FREE Full text] [doi: 10.1186/s40779-020-00238-8] [Medline: 32111253]

XSL•FO
RenderX

22. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016 Jul 12;6(7):e010247 [FREE Full text] [doi: 10.1136/bmjopen-2015-010247] [Medline: 27406637]
23. Tufanaru C, Munn Z, Stephenson M, Aromataris E. Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. Int J Evid Based Healthc 2015 Sep;13(3):196-207. [doi: 10.1097/XEB.0000000000000065] [Medline: 26355603]
24. Dettori JR, Norvell DC, Chapman JR. Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. Global Spine J 2022 Sep;12(7):1624-1626 [FREE Full text] [doi: 10.1177/21925682221110527] [Medline: 35723546]
25. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
26. Data query. United Nations Statistics Big Data and Global Center for Data Science. URL: https://unbdc.stats.gov.cn/zt/gjtjsj/sjcx/ [accessed 2025-02-18]
27. Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt T, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol 2011 Nov;64(11):1187-1197. [doi: 10.1016/j.jclinepi.2010.08.010] [Medline: 21477993]
28. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
29. Alhuwail D, Abdulsalam Y. Assessing electronic health literacy in the state of Kuwait: survey of internet users from an Arab state. J Med Internet Res 2019 May 24;21(5):e11174 [FREE Full text] [doi: 10.2196/11174] [Medline: 31127723]
30. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): reliability and validity testing in younger adults recruited online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
31. Sinan O, Ayaz-Alkaya S, Akca A. Predictors of eHealth literacy levels among nursing students: a descriptive and correlational study. Nurse Educ Pract 2023 Mar;68:103592. [doi: 10.1016/j.nepr.2023.103592] [Medline: 36889169]
32. Zuo Q, Cheng J, Peng W, Zhu L, Jiang X. Differences analysis of community residents' e-Health literacy level and influencing factors between urban and rural. Chin Nurs Res 2022 Feb;36(4):587-593.
33. Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, et al. Correlation between eHealth literacy and health literacy using the eHealth Literacy Scale and real-life experiences in the health sector as a proxy measure of functional health literacy: cross-sectional web-based survey. J Med Internet Res 2018 Oct 31;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
34. Park BK. Factors influencing eHealth literacy of middle school students in Korea: a descriptive cross-sectional study. Healthc Inform Res 2019 Jul;25(3):221-229 [FREE Full text] [doi: 10.4258/hir.2019.25.3.221] [Medline: 31406614]
35. Gazibara T, Cakic M, Cakic J, Grgurevic A, Pekmezovic T. Familiarity with the internet and health apps, and specific topic needs are amongst the factors that influence how online health information is used for health decisions amongst adolescents. Health Info Libr J 2024 Sep;41(3):283-297. [doi: 10.1111/hir.12440] [Medline: 35652454]
36. Ghazi SN, Berner J, Anderberg P, Sanmartin Berglund J. The prevalence of eHealth literacy and its relationship with perceived health status and psychological distress during Covid-19: a cross-sectional study of older adults in Blekinge, Sweden. BMC Geriatr 2023 Jan 04;23(1):5 [FREE Full text] [doi: 10.1186/s12877-022-03723-y] [Medline: 36597040]
37. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
38. Hoang Nguyen L, Bich Thi Le T. E-health literacy of medical students at a university in Central Vietnam. Indian J Public Health Res Dev 2020 Feb 1;11(2):1299. [doi: 10.37506/v11/i2/2020/ijphrd/195001]
39. Almoajel A, Alshamrani S, Alyabsi M. The relationship between e-Health literacy and breast cancer literacy among Saudi women. Front Public Health 2022;10:841102 [FREE Full text] [doi: 10.3389/fpubh.2022.841102] [Medline: 35462821]
40. Wongjinda L, Taneepanichsakul S. Determinants of eHealth literacy level among Royal Thai Army personnel: a case study. Southeast Asian J Trop Med Public Health 2018;49(1):175-182.
41. Lee WL, Lim ZJ, Tang LY, Yahya NA, Varathan KD, Ludin SM. Patients' readiness and eHealth literacy: implications for adoption and deployment of eHealth in the COVID-19 era and beyond. Comput Inform Nurs 2021 Nov 02;40(4):244-250. [doi: 10.1097/CIN.0000000000000854] [Medline: 34740221]
42. Kim S, Jeon J. Factors influencing eHealth literacy among Korean nursing students: a cross-sectional study. Nurs Health Sci 2020 Sep;22(3):667-674. [doi: 10.1111/nhs.12711] [Medline: 32154981]
43. Holch P, Marwood JR. EHealth literacy in UK teenagers and young adults: exploration of predictors and factor structure of the eHealth Literacy Scale (eHEALS). JMIR Form Res 2020 Sep 08;4(9):e14450 [FREE Full text] [doi: 10.2196/14450] [Medline: 32897230]
44. Tanasombatkul K, Pinyopornpanish K, Angkurawaranon C, Buawangpong N, Rojanasumapong A, Jiraporncharoen W. Is electronic health literacy associated with learning outcomes among medical students in the first clinical year?: a cross-sectional study. Eur J Investig Health Psychol Educ 2021 Aug 19;11(3):923-932 [FREE Full text] [doi: 10.3390/ejihpe11030068] [Medline: 34563081]

XSL•FO
RenderX

45. Martins A, Andrade I, Pocinho R, Belo P. e-Health literacy in ageing. 2015 Presented at: TEEM '15: 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality; October 7-9, 2015; Porto, Portugal p. 53-57. [doi: 10.1145/2808580.2808589]
46. Al-Qahtani S, Jalal Z, Paudyal V, Mahmood S, Mason J. The role of pharmacists in providing pharmaceutical care in primary and secondary prevention of stroke: a systematic review and meta-analysis. Healthcare (Basel) 2022 Nov 18;10(11):2315 [FREE Full text] [doi: 10.3390/healthcare10112315] [Medline: 36421639]
47. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020 Feb 24;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
48. Shaojie L. Study on the Status and Influencing Factors of eHealth Literacy Among Older Adults. Hunan, China: Zhongnan University; 2024.
49. Xie B. Experimenting on the impact of learning methods and information presentation channels on older adults' e-health literacy. J Am Soc Inf Sci 2011 Jun 06;62(9):1797-1807. [doi: 10.1002/asi.21575]
50. Carstensen LL. Evidence for a life-span theory of socioemotional selectivity. Curr Dir Psychol Sci 1995 Oct 01;4(5):151-156. [doi: 10.1111/1467-8721.ep11512261]
51. Hauk N, Hüffmeier J, Krumm S. Ready to be a silver surfer? A meta-analysis on the relationship between chronological age and technology acceptance. Comput Hum Behav 2018 Jul;84:304-319. [doi: 10.1016/j.chb.2018.01.020]
52. Gikas J, Grant MM. Internet High Educ 2013 Oct;19:18-26. [doi: 10.1016/j.iheduc.2013.06.002]
53. Salthouse T. Major Issues in Cognitive Aging. Oxford, UK: Oxford University Press; 2009. [doi: 10.1093/acprof:oso/9780195372151.001.0001]
54. Schrauben SJ, Appel L, Rivera E, Lora CM, Lash JP, Chen J, CRIC Study Investigators. Mobile health (mHealth) technology: assessment of availability, acceptability, and use in CKD. Am J Kidney Dis 2021 Jun;77(6):941-950.e1 [FREE Full text] [doi: 10.1053/j.ajkd.2020.10.013] [Medline: 33309860]
55. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: W H Freeman; 1997.
56. Taba M, Allen TB, Caldwell PHY, Skinner SR, Kang M, McCaffery K, et al. Adolescents' self-efficacy and digital health literacy: a cross-sectional mixed methods study. BMC Public Health 2022 Jun 20;22(1):1223 [FREE Full text] [doi: 10.1186/s12889-022-13599-7] [Medline: 35725389]
57. Liu S, Zhao H, Fu J, Kong D, Zhong Z, Hong Y, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in Southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996 [FREE Full text] [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
58. Yang E, Chang SJ, Ryu H, Kim HJ, Jang SJ. Comparing factors associated with eHealth literacy between young and older adults. J Gerontol Nurs 2020 Aug 01;46(8):46-56. [doi: 10.3928/00989134-20200707-02] [Medline: 32936926]
59. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
60. Li S, Cui G, Yin Y, Wang S, Liu X, Chen L. Health-promoting behaviors mediate the relationship between eHealth literacy and health-related quality of life among Chinese older adults: a cross-sectional study. Qual Life Res 2021 Aug 04;30(8):2235-2243 [FREE Full text] [doi: 10.1007/s11136-021-02797-2] [Medline: 33661455]
61. Rabenbauer LM, Mevenkamp N. Factors in the effectiveness of e-health interventions for chronic back pain: how self-efficacy mediates e-health literacy and healthy habits. Telemed J E Health 2021 Feb 01;27(2):184-192. [doi: 10.1089/tmj.2019.0301] [Medline: 32397853]

Abbreviations

DL: DerSimonian and Laird
eHEALS: eHealth Literacy Scale
eHLQ: eHealth Literacy Questionnaire
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
IV: inverse variance weighted
JBI: Joanna Briggs Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
TAM: Technology Acceptance Model

XSL•FO
RenderX

Edited by N Cahill; submitted 26.06.23; peer-reviewed by A Sheon, HJT van Mens; comments to author 14.02.24; revised version received 30.04.24; accepted 29.01.25; published 10.03.25.

Please cite as:
Hua Z, Yuqing S, Qianwen L, Hong C
Factors Influencing eHealth Literacy Worldwide: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e50313
URL: https://www.jmir.org/2025/1/e50313
doi:10.2196/50313
PMID:

©Zhong Hua, Song Yuqing, Liu Qianwen, Chen Hong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The State of the Art of eHealth Self-Management Interventions for People With Chronic Obstructive Pulmonary Disease: Scoping Review

Eline te Braake[1,2], MSc; Roswita Vaseur[2], MSc; Christiane Grünloh[1,2], PhD; Monique Tabak[2], Prof Dr, IR

[1]Roessingh Research and Development, Enschede, The Netherlands

[2]Faculty of Electrical Engineering, Mathematics, and Computer Science, Biomedical Signals and Systems Group, University of Twente, Enschede, The Netherlands

Corresponding Author:
Eline te Braake, MSc
Roessingh Research and Development
Roessinghbleekseweg 33b
Enschede, 7522 AH
The Netherlands
Phone: 31 0880875734
Email: e.tebraake@rrd.nl

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a common chronic incurable disease. Treatment of COPD often focuses on symptom management and progression prevention using pharmacological and nonpharmacological therapies (eg, medication, inhaler use, and smoking cessation). Self-management is an important aspect of managing COPD. Self-management interventions are increasingly delivered through eHealth, which may help people with COPD engage in self-management. However, little is known about the actual content of these eHealth interventions.

Objective: This literature review aimed to investigate the state-of-the-art eHealth self-management technologies for COPD. More specifically, we aimed to investigate the functionality, modality, technology readiness level, underlying theories of the technology, the positive health dimensions addressed, the target population characteristics (ie, the intended population, the included population, and the actual population), the self-management processes, and behavior change techniques.

Methods: A scoping review was performed to answer the proposed research questions. The databases PubMed, Scopus, PsycINFO (via EBSCO), and Wiley were searched for relevant articles. We identified articles published between January 1, 2012, and June 1, 2022, that described eHealth self-management interventions for COPD. Identified articles were screened for eligibility using the web-based software Rayyan.ai. Eligible articles were identified, assessed, and categorized by the reviewers, either directly or through a combination of methods, using Atlas.ti version 9.1.7.0. Thereafter, data were charted accordingly and presented with the purpose of giving an overview of currently available literature while highlighting existing gaps.

Results: A total of 101 eligible articles were included. This review found that most eHealth technologies (91/101, 90.1%) enable patients to self-monitor their symptoms using (smart) measuring devices (39/91, 43%), smartphones (27/91, 30%), or tablets (25/91, 27%). The self-management process of "taking ownership of health needs" (94/101, 93.1%), the behavior change technique of "feedback and monitoring" (88/101, 87%), and the positive health dimension of "bodily functioning" (101/101, 100%) were most often addressed. The inclusion criteria of studies and the actual populations reached show that a subset of people with COPD participate in eHealth studies.

Conclusions: The current body of literature related to eHealth interventions has a strong tendency toward managing the physical aspect of COPD self-management. The necessity to specify inclusion criteria to control variables, combined with the practical challenges of recruiting diverse participants, leads to people with COPD being included in eHealth studies that only represent a subgroup of the whole population. Therefore, future research should be aware of this unintentional blind spot, make efforts to reach the underrepresented population, and address multiple dimensions of the positive health paradigm.

(J Med Internet Res 2025;27:e57649) doi:10.2196/57649

KEYWORDS

eHealth; self-management; interventions; chronic obstructive pulmonary disease; COPD; review

XSL•FO
RenderX

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is a common disabling lung condition characterized by chronic respiratory symptoms that cause persistent, mostly progressive airflow limitations [1]. It is one of the major issues of public health, and its prevalence, mortality, and morbidity are increasing [1-3]. COPD was listed as the third leading cause of death worldwide in 2019 [4], and it is estimated that by 2040, deaths from COPD will rise to 4.4 million per year [5]. In addition, people with a lower socioeconomic status are at increased risk of developing COPD [6]. Although COPD may, in some cases, be the result of a genetic risk factor, it is, in most cases, caused by exposure to tobacco smoking and the inhalation of toxic particles and gases from indoor and outdoor air pollution [7,8]. People with COPD often experience symptoms such as dyspnea, fatigue, chest tightness, activity limitation, and cough that may be accompanied by sputum production [7]. Furthermore, they may experience acute events, known as exacerbations, which can lead to hospitalization. Although COPD is chronic and thus not curable, it is, however, treatable, and disease progression is preventable [3,9]. Therefore, the treatment of COPD often focuses on reducing symptoms and future risks of exacerbation with the use of pharmacological and nonpharmacological therapies (eg, inhaler use, vaccinations, smoking cessation, and self-management) [7]. Given its chronic nature and the impact of the disease on all facets of one's life, an important aspect of treating COPD and secondary prevention is chronic disease management.

An essential component of chronic disease management is self-management [10]. Owing to the variation in the literature regarding the definition of self-management in COPD, this paper defines self-management as follows: "The ability of an individual to manage one's symptoms, treatment, physical, social, and emotional consequences, and lifestyle changes. It includes means of empowerment, educating oneself, being autonomous, learning and adapting to new behaviors, acceptance, and adapting to a new balance in life." It requires patients to take an involved and responsible role in their health, with the aim of becoming active participants [10,11]. Self-management interventions or programs are developed to help patients engage in self-management, and their effectiveness is investigated in research.

Self-management interventions or programs are shown to have positive effects, for example, in supporting patients to develop and improve their self-management skills and disease knowledge [12-14]. Camus-García et al [15] found that self-management interventions may improve clinical outcomes in COPD (eg, improvements in health-related quality of life) and lower the probability of hospital admissions. The actual content of such self-management intervention programs for COPD is diverse [12], and it remains unclear which specific elements are essential for designing a successful program. In the following sections, some elements that can be considered when designing an intervention program will be briefly described: content for COPD self-management, processes of self-management, and behavior change techniques (BCTs).

The diversity of content may be explained by the numerous objectives and end points of self-management intervention programs [16]. Interventions focus on acute exacerbation management and admission avoidance by incorporating exacerbation action plans [17] and often also include education, exercise training, and breathing strategies [18]. However, research suggests that intervention programs that only include education or action plans alone may not result in behavioral change, increased patient confidence, or the acquisition of new skills that patients learn or practice [16,19].

Besides the content of self-management intervention programs, the design of the intervention program should also reflect that self-management consists of different processes. Schulman-Green et al [20] identified different self-management processes for chronic illnesses, such as "learning,'" "taking ownership of health needs," and "performing health promotion activities." All processes are divided into specific self-management tasks (eg, "learning about condition and health needs" and "changing behavior to minimize health impact") and skills (eg, "acquiring information" and "reducing stress") [20]. Schulman-Green et al [20] concluded that the identification of such processes may help support and guide future self-management intervention programs. They also demonstrated that these various processes should be viewed within the broader context, as their significance to patients may vary depending on where they are in their patient journey [20]. Therefore, more knowledge about such self-management processes within self-management eHealth intervention programs is needed to support the development of such interventions.

Self-management interventions may also aim to change a certain behavior of the patient, so the incorporation of BCTs can be beneficial to designing a successful intervention program. A BCT is "a specific observable, replicable, and irreducible component of an intervention program designed to alter or redirect causal processes that regulate behavior" [21] and can be included in the design of any type of self-management intervention program. By adding these "active ingredients" (eg, "feedback" and "self-monitoring"), chances for achieving behavioral change may be increased [21]. Thus, combining self-management processes and BCTs in intervention programs may lead to positive results for one's self-management. However, to the best of our knowledge, no research is dedicated to investigating the presence of BCTs and self-management processes in current self-management interventions for COPD.

One way to support people with COPD in engaging in self-management is through the use of eHealth interventions. eHealth interventions can be defined as "An eHealth technology specifically focused on intervening in an existing context by changing behaviors and/or cognitions" [22]. eHealth interventions to support self-management may help people with chronic diseases become more independent and empowered by, for example, gaining knowledge about their disease, monitoring and reporting daily symptoms, and learning specific self-management skills [23-25]. Therefore, the use of eHealth interventions in COPD care represents a promising way of

XSL•FO
RenderX

delivering health services, such as support in self-management [26].

In the current literature, a diverse range of eHealth interventions aim to support patients in their self-management, and these are increasingly provided to support patients in health communication, self-monitoring, and their medical treatment [10,26]. Available literature revealed that current eHealth interventions for COPD mainly focused on COPD care, education, smoking cessation, medication adherence, exercise, diet, and symptom management [27,28]. This indicates a tendency toward managing the physical aspect of COPD in self-management eHealth interventions. However, the physical aspect of one's disease is only 1 dimension of the positive health paradigm. As conceptualized by Huber et al [29], "Health includes the ability to adapt, and self manage in the face of social, physical, and emotional challenges," also referred to as "positive health." Huber et al [30] stated that positive health as a concept has several important health indicators, categorized into 6 dimensions: "bodily functions," "mental well-being," "meaningfulness," "quality of life," "social participation," and "daily functioning" [30]. They stressed the fact that paying attention to these indicators could support shared decision-making and bridge the gap between health care and the social context. Therefore, these dimensions are all important to consider when self-managing one's disease. However, no research is available regarding the extent of positive health dimensions addressed in current self-management eHealth interventions for COPD.

Furthermore, using eHealth to support people with COPD might entail some challenges, as low health literacy is prevalent among people with COPD [31,32]. In addition, moderate levels of self-reported eHealth literacy are common among people with COPD [33]. Some studies revealed that people with COPD experienced technical barriers when using eHealth interventions for self-management [34]. However, Williams et al [35] indicated a few technical issues experienced by people with COPD when using eHealth to support self-management, leading to uncertainty about whether such eHealth technologies are suitable for the whole COPD population. Although some information about eHealth use for this population is available [27-28,33,35-36], little research is dedicated to investigating whether current eHealth interventions account for the wider population of people with COPD (such as those with eHealth literacy). Therefore, it should be investigated whether there is a difference between the intended population that eHealth interventions aim to target and the eventual included population in those studies. As knowledge and new insights derived from those studies often serve as a starting point for future work, it can be very valuable to look into the representation of the COPD population within studies.

Objectives

To summarize, little is known about the actual content and design of self-management eHealth interventions for people with COPD. Therefore, this scoping review aimed to investigate the current state-of-the-art eHealth interventions for COPD self-management and identify potential gaps in the literature, which may provide insight into or serve as inspiration for the development of future eHealth self-management interventions. "State of the art" within the context of this study can be defined as follows: "The collection of all underlying components that form the basis for the eHealth self-management interventions for people with COPD." Figure 1 shows how the different parts of this review contribute to an overall picture of the current literature and highlights the specific aspects explored in this review. More specifically, we aimed to unravel the state of the art of eHealth self-management interventions by using the following subquestions:

- What is the "e" in eHealth self-management?
- What is the "health" in eHealth self-management?
- Who is the "self" in self-management?
- What is the "management" in eHealth self-management?

XSL•FO
RenderX

Figure 1. Flowchart of the research questions. COPD: chronic obstructive pulmonary disease; TRL: technology readiness level.

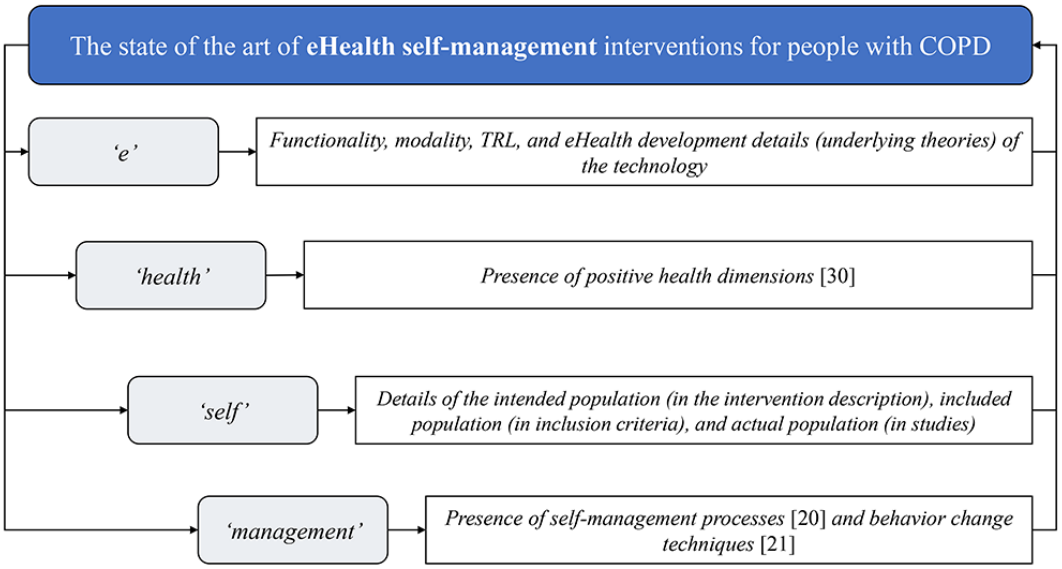

Methods

Overview

A scoping review was performed to investigate the currently available literature. According to Munn et al [37], a scoping review is an ideal tool for providing, for example, an overview of existing literature, identifying key characteristics, and highlighting knowledge gaps, among others. Hence, this was deemed the most suitable method for answering the proposed research questions. Parts of the PRISMA-ScR (Preferred Reporting Items For Systematic Reviews and Meta-Analyses extension for Scoping Reviews) protocol (items 1-7, 9-11, and 13-21), as proposed by Tricco et al [38], were followed and tailored to this study to ensure a systematic approach for answering the research questions. The PRISMA-ScR checklist is provided in Multimedia Appendix 1 [38]. The protocol for this review was not published.

Search Strategy

The first reviewer (EtB) was responsible for identifying relevant articles in the databases PubMed, Scopus, PsycINFO, and Wiley. Combinations of the search terms "self-management," "COPD," and "eHealth" were used to generate the search string. The search string that was used for this review is provided in Multimedia Appendix 2.

Study Selection

Studies were considered eligible if they were original research and portrayed an eHealth intervention supporting the self-management of COPD. The eHealth self-management intervention should actively involve and engage individuals with COPD, ensuring they experience personal benefits from their self-management efforts, supported and encouraged by the intervention. Per definition, self-management aims for patients to become active participants in their care, which means that patient involvement within self-management eHealth interventions is essential to be able to fulfill an active role. Therefore, we did not consider that eHealth interventions actually support self-management when patients themselves are not involved (eg, if the interaction with the eHealth service is limited to collecting data). Furthermore, articles needed to be published between January 1, 2012, and June 1, 2022. As of 2012, eHealth technologies were upcoming, and their relevance for future health care appeared to be promising. For example, national efforts to implement eHealth in current care were presented in 2012 in the Netherlands (eg, [39]). The complete list of assessment and eligibility criteria is presented in Textbox 1.

XSL•FO
RenderX

Textbox 1. Assessment and eligibility criteria for studies.

Concept

- Studies describing an eHealth intervention supporting the self-management of chronic obstructive pulmonary disease (COPD) were included.
- Studies not fulfilling the inclusion criteria were excluded.

Population

- Studies involving adults aged ≥18 years diagnosed with COPD (and other chronic conditions provided that the eHealth technology has a dedicated part toward COPD) were included.
- Studies using general terms such as "older adults," "rural patients," or "communities" or referencing unspecified multimorbidity or chronic conditions without any clarification about the population were excluded.

eHealth technology

- Studies where eHealth technologies were used to support people with COPD in engaging in self-management, involving patients in their intervention were included if they used at least 1 self-management process, as defined by Schulman-Green et al [20], and in the case of sole monitoring, where patients were able to see their data.
- Studies collecting data solely for research purposes to train machine learning or artificial intelligence algorithms without any further patient engagement were excluded.

Study design

- Original research studies were included.
- Reviews, protocols, abstracts, letters, conference proceedings, commentaries, notes, short surveys, and erratum were excluded.

Language

- Studies in English were included.
- Studies not fulfilling the inclusion criteria were excluded.

Year of publication

- Studies published between January 1, 2012, and June 1, 2022, were included.
- Studies not fulfilling the inclusion criteria were excluded.

Procedure

The screening was performed on June 3, 2022, with screening supported by the web-based software Rayyan.ai (Rayyan Systems Inc) [40]. To screen articles for title and abstract, both reviewers (EtB and RV) adhered to the eligibility criteria that were discussed before the start of the screening (Textbox 1). One reviewer (EtB) screened all articles for title and abstract. The second reviewer (RV) screened 20% (118/588) of the titles and abstracts of those studies. After this first screening, a discussion took place to compare discrepancies and come to a consensus between reviewers (EtB and RV). Both reviewers had previous experience with performing a systematic review. After this first screening, it was necessary to revise and clarify some of the inclusion criteria to arrive at a satisfactory level of agreement between reviewers. This means that with the use of the revised inclusion criteria, clear and substantiated decisions could be made on whether to include a certain article in this review. For the full-text screening, the same process was applied. In this screening, the level of agreement between reviewers was satisfactory. Reasons for excluding articles during the full-text screening were recorded. Before extracting the data, a data extraction form was developed, discussed, and agreed upon with 3 authors (EtB, CG, and MT). This form was piloted after the full-text screening to reduce errors during data extraction. Data extraction was performed by the first author (EtB) using Atlas.ti (version 9.1.7.0; Lumivero) [41], based on the data extraction form. Some of the data to answer the subquestions were directly extracted (eg, type of study and year of study), some data were the result of an assessment or categorization of the reviewers (eg, positive health dimensions), and some data were a combination of both (eg, BCTs). After the extraction, data were clustered and charted in various ways (eg, bar charts, tables, and descriptive presentations). Finally, charted data were scrutinized and synthesized by 1 reviewer (EtB) before discussing the results with 2 authors (CG and MT). Thereafter, results were written down to answer the proposed research questions. An overview of how the articles were extracted and charted is provided in Multimedia Appendix 3.

Results

Search Results

Figure 2 shows the detailed flowchart of the inclusion of studies. A total of 893 articles were identified during the initial search, of which 305 (35.1%) duplicates were removed, and the remaining 588 (%) articles were screened on title and abstract; this screening phase resulted in 189 (32.1%) articles that could be assessed for full text. After full-text screening of 189 articles,

XSL•FO
RenderX

88 articles (46.6%) were excluded, resulting in 101 articles (53.4%) being included in this scoping review.

Figure 2. Flowchart of the inclusion of studies. COPD: chronic obstructive pulmonary disease.

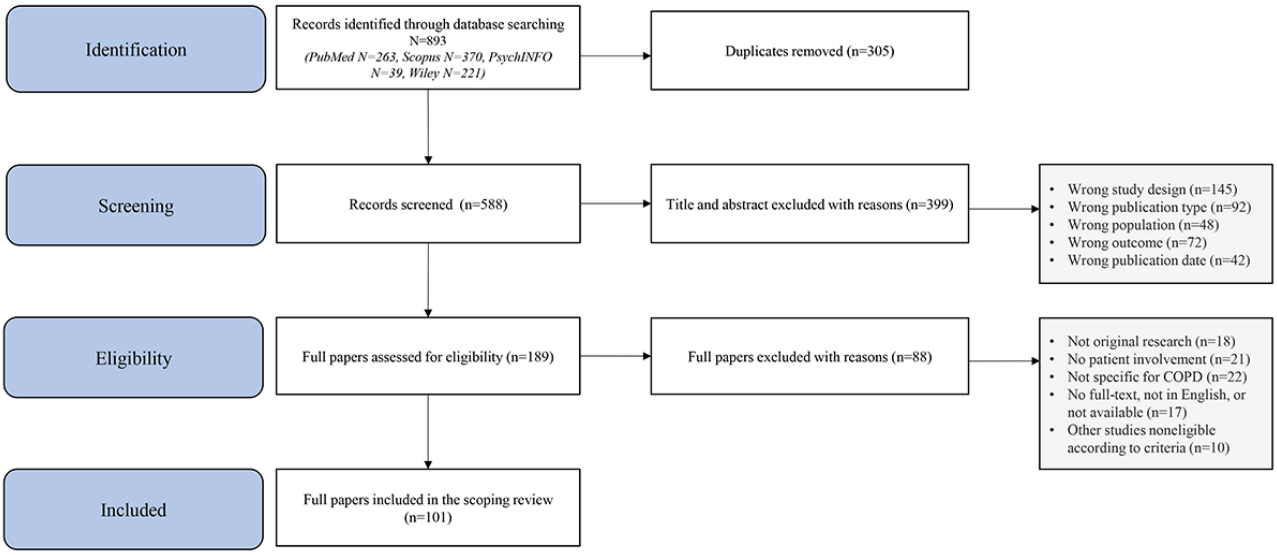

Study Characteristics

The included papers represented 101 unique studies. Most articles (18/101, 17.8%) were published in 2021, followed by 2020 (15/101, 14.8%) and 2017 (13/101, 12.9%). As shown in Table 1, the most common study types were either randomized controlled trials (18/101, 17.8%) or (prospective) pilot studies (16/101, 15.8%).

Table 1. Type of studies (n=101).

Type of studies	Frequency, n (%)
Randomized controlled trials	18 (17.8)
Prospective pilot study	17 (16.8)
Evaluation study	10 (9.9)
Feasibility study	9 (8.9)
Qualitative study	9 (8.9)
Retrospective or secondary analysis	7 (6.9)
Observational study	6 (5.9)
Design or development study	4 (4)
Implementation study	3 (3)
Mixed methods study	3 (3)
Exploratory study	3 (3)
Others	12 (11.9)

The e in eHealth

This section focuses on the functionality, modality, technology readiness level (TRL) [42], and eHealth development details of the used technologies. Details about the functionality and modality of the eHealth interventions are provided in Multimedia Appendix 4 [25-27,32,36,43-139]. Of the 101 included studies, 76 (75.2%) mentioned the name of their eHealth technologies. Some articles reported on studies using the same eHealth technologies (eg, “EDGE” [35,51-54], “It’s Life!” [43,55,56], “MasterYourBreath” [57-60], and “COMET” [61,62]). In some cases, studies using the same eHealth technology together portrayed the process of developing, testing, and evaluating an eHealth intervention. A total of 50 unique eHealth technologies were found in this review.

Most articles (91/101, 90.1%) included self-monitoring (eg, monitoring of symptoms) as a function of their technology. Of the 101 articles, 69 (68.3%) included the function of educating or informing (eg, education on COPD) and 27 (26.7%) supported communication (eg, eConsults with HCPs and peer-to-peer support chats). Most articles (68/101, 67.3%) included >1 function within their technology.

Table 2 shows that a (smart) measurement device (eg, wearable or monitoring system) was the most common (39/101, 38.6%) modality used in the studies, followed by a smartphone (27/101, 26.7%) and tablet (25/101, 24.7%). If studies used >1 device,

XSL•FO
RenderX

the most common combination was a (smart) measurement device with a tablet (19/101, 18.8%) or smartphone (8/101, 7.9%).

Table 2. Overview of modalities of the eHealth technologies. Some articles had used >1 device, resulting in 131 modalities within 101 articles.

Modalities of the eHealth technology	Frequency, n (%)
Smart measurement and monitoring devices	45 (34.4)
Smartphone	26 (19.8)
Normal phone or landline	16 (12.2)
Computer, laptop, or whiteboard	16 (12.2)
Tablet	15 (11.5)
Not mentioned	9 (6.9)
Television	3 (2.3)
Other	1 (0.8)

This review found no article that explicitly stated their TRL. According to our assessment and categorization, 47 eHealth technologies in the articles were assessed to be in the development phase (TRL 4 to TRL 6), 53 in the deployment phase (TRL 7 to TRL 9), and 0 in the research phase (TRL 1 to TRL 3).

Details about the eHealth development process showed that only 14 (13.9%) out of 101 studies explicitly mentioned using either a user-centered design, participatory design, scenario-based methods, reflective lifeworld research, or action research approach. Furthermore, of the 101 studies, 18 (17.8%) reported details about the theories on which their self-management intervention was based. Some of these were targeted toward BCTs independent of technology use, while others were technology related and more targeted toward technological adoption or persuasive design. Table 3 lists the various theories identified as underlying the eHealth interventions. Theories that were present could be divided into the following categories: behavioral change, technological adoption or persuasive design, and unspecified. This review found 11 different behavior change theories, 3 different technological adoption or persuasive design theories, and 3 unspecified theories. Of all the different theories within the different categories, the social cognitive theory was most often used (5/11, 45%).

XSL•FO
RenderX

Table 3. Theories used within the eHealth self-management interventions.

Categories and theories	References
Behavioral change	
HBM[a]	[60,63]
Social cognitive theory	[60,64-67]
Self-care theory	[68]
Transtheoretical model	[60,64]
Five A's model	[56]
ASE[b]	[60]
Self-efficacy theory	[69]
I-Change model	[57-60]
Self-determination model	[70]
TGG[c]	[71]
Theory of planned behavior	[60]
Technological adoption or persuasive design	
TAM[d]	[72]
UTAUT[e]	[73]
eHealth-based PCC[f]	[74]
Unspecified	
Goal setting theories	[60]
Implementation theory	[60]
Health promotion	[69]

[a]HBM: health belief model.

[b]ASE: attitude-social influence-self-efficacy model.

[c]TGG: Tech to Goal.

[d]TAM: technology acceptance model.

[e]UTAUT: unified theory of acceptance and use of technology.

[f]PCC: person-centered care.

The Health in eHealth Technologies for Self-Management

Table 4 shows how many eHealth technologies used in the studies addressed the different positive health dimensions. All the included articles (N=101, 100%) addressed (at least) the dimension of bodily functioning, 45 (44.6%) addressed daily functioning, 13 (12.9%) addressed participation, and 12 (11.9%) addressed mental well-being. We were not able to identify any indications that the dimensions of meaningfulness and quality of life were explicitly addressed in any of the eHealth technologies supporting self-management. Details about the positive health dimensions are provided in Multimedia Appendix 5.

Most studies (48/101, 47.5%) focused on 1 specific dimension, namely, bodily functioning. Other articles (42/101, 41.6%) focused on 2 dimensions, 11 (10.9%) on 3 dimensions, and only 3 (3%) on 4 dimensions within their eHealth technology. The combination of the dimensions of bodily functioning and daily functioning was the most common (33/101, 32.7%), followed by the combinations of bodily functioning, daily functioning, and mental well-being (5/101, 5%); bodily functioning and participation (4/101, 4%); bodily functioning, daily functioning, and participation (3/101, 3%); bodily functioning, mental well-being, and participation (3/101, 3%); body functioning, mental well-being, participation, and daily functioning (3/101, 3%); and bodily functioning and mental well-being (1/101, 1%).

XSL•FO
RenderX

Table 4. Distribution of the positive health dimensions. Some technologies addressed >1 dimension, which are counted separately in this table, resulting in 172 dimensions found within 101 articles.

Positive health dimension	Frequency, n (%)
Bodily functioning	101 (58.7)
Daily functioning	46 (26.7)
Participation	13 (7.6)
Mental well-being	12 (7)
Meaningfulness	0 (0)
Quality of life	0 (0)

To investigate whether there may be an increase or decrease in certain positive health dimensions over time, we compared the presence of certain dimensions with the years of the studies (Figure 3). Such information can be useful to see whether the target dimensions of eHealth interventions are changing with time. When comparing the presence of the dimensions with the years of the studies, we found that in the years 2013 to 2015 and 2017 to 2018, the dimension of bodily functioning is dominantly present, followed by daily functioning. From 2017 to 2021, a small increase in the presence of the dimension of mental well-being could be seen over the years. In the years 2020 and 2021, the presence of the dimensions of daily functioning and participation was almost equal compared to bodily functioning.

Figure 3. Distribution of the positive health dimensions in articles over time. Dimensions of "quality of life" and "meaningfulness" were not displayed, as no included articles explicitly addressed these.

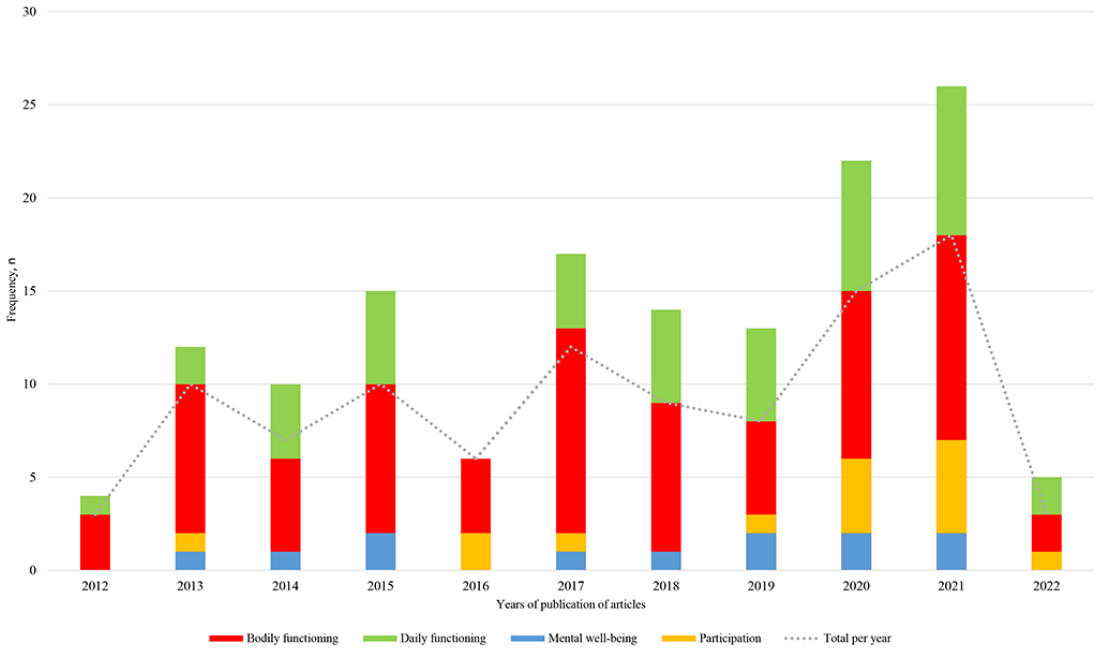

The Self in Self-Management

Overview

All 101 included papers (partly) described the intended population for the intervention (as stated in the intervention description), the included population (as stated in the inclusion and exclusion criteria), and the final actual study population (as stated in the demographics of study participants). In some studies, certain inclusion criteria were required to participate, thereby restricting the group of eligible participants (ie, the actual population). This scoping review extracted the following inclusion criteria: disease-specific (needing to have a certain severity of COPD), capability-related (needing to be cognitively capable, able to read and write, understand certain language, willing or able to provide consent), age-related (needing to have a minimum or maximum age), smoking history–related (being a current or former smoker), and technology-related (needing to have digital skills, internet access, own a certain device). More details about the concept of the self in self-management are provided in Multimedia Appendix 6.

Intended Population

There was some variation in the specific intended populations targeted in the articles. As shown in Figure 4, most studies (59/101, 58.4%) were targeted at persons with COPD in general, with 23 (22.8%) focusing on ≥1 specific COPD severities and 19 (18.8%) focusing on COPD in combination with other chronic conditions. Some articles included >1 comorbidity.

XSL•FO
RenderX

Figure 4. Intended population. CHD: chronic heart disease; CHF: chronic heart failure; COPD: chronic obstructive pulmonary disease.

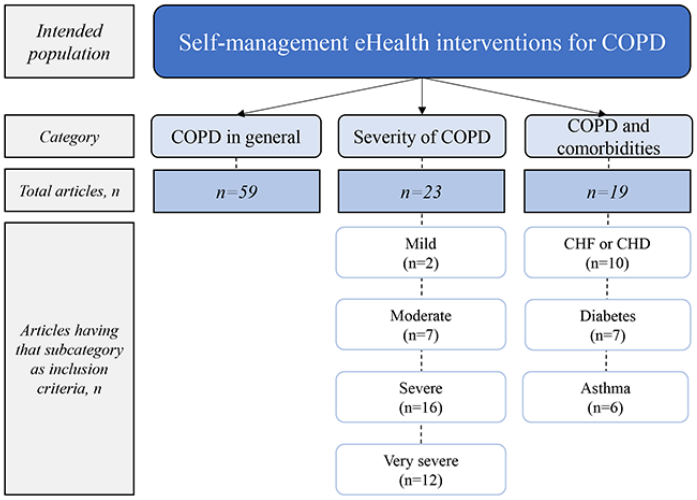

Included Population

Figure 5 presents an overview of the identified included population. More studies (50/101, 49.5%) than outlined in the Intended Population section (23/101, 22.8%) had disease-specific inclusion criteria (focusing on ≥1 COPD severities). Of the 101 articles, 50 (49.5%) had capability-related inclusion criteria, requiring participants to, for example, be cognitively capable or able to write and read to be eligible for participation. Furthermore, in 37.6% (38/101) of the articles, participants needed to have a certain minimum age, with 40 years being the most common. In 7.9% (8/101) of the articles, the age needed to be below a certain maximum. The maximum age of 70 years was the most commonly mentioned inclusion criterion, cited 4 times. A total of 12 (11.9%) out of 101 articles had inclusion criteria regarding smoking (history) in which participants needed to be, for example, a former smoker. Finally, 38.6% (39/101) of the articles had technology-related inclusion criteria. Participants needed, for example, to own a smartphone or tablet and have digital skills to participate. Only 1 study explicitly mentioned having no exclusion criteria based on age, comorbidities, and previous participation in pulmonary rehabilitation. Furthermore, in the same study, participants did not need to have previous experience using digital technology.

Figure 5. Included population. COPD: chronic obstructive pulmonary disease.

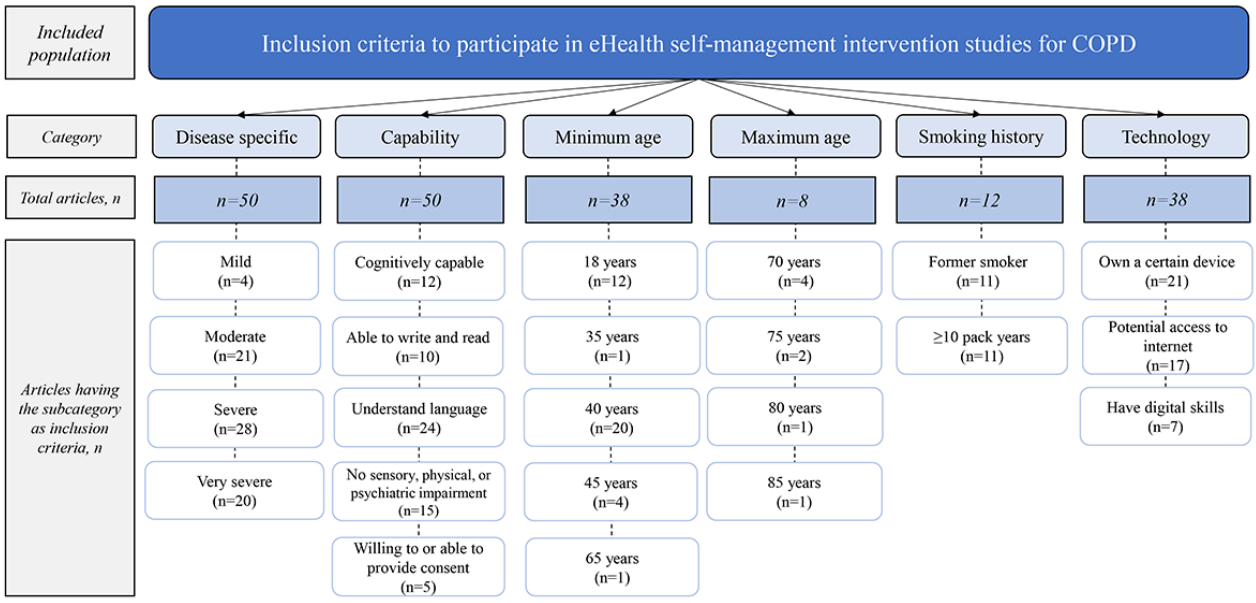

Actual Population

Figure 6 shows the actual population included in the studies. In 24.7% (25/101) of the articles that mentioned the severity of their participants, most participants had moderate or severe COPD. Out of the 101 articles, only 21 (20.8%) shared a clear description of the education level of their participants, which were then categorized for this study. The educational level of participants could be categorized as low, medium, and high, which were almost equally distributed. Of the 70.3% (71/101) of the articles that shared the mean age of their participants, we calculated the combined mean age, which resulted in 64.85 years. The gender of participants was clearly mentioned in 88 (87.1%) out of 101 articles and was almost equally distributed. In 29.7% (30/101) of the articles that shared the smoking history of their participants, almost half of the participants (51%) were

XSL•FO
RenderX

reported as current or former smokers. In 10.9% (11/101) of the articles that described technology-related experience, 89% of the participants had experience with technology.

Figure 6. Actual population. COPD: chronic obstructive pulmonary disease.

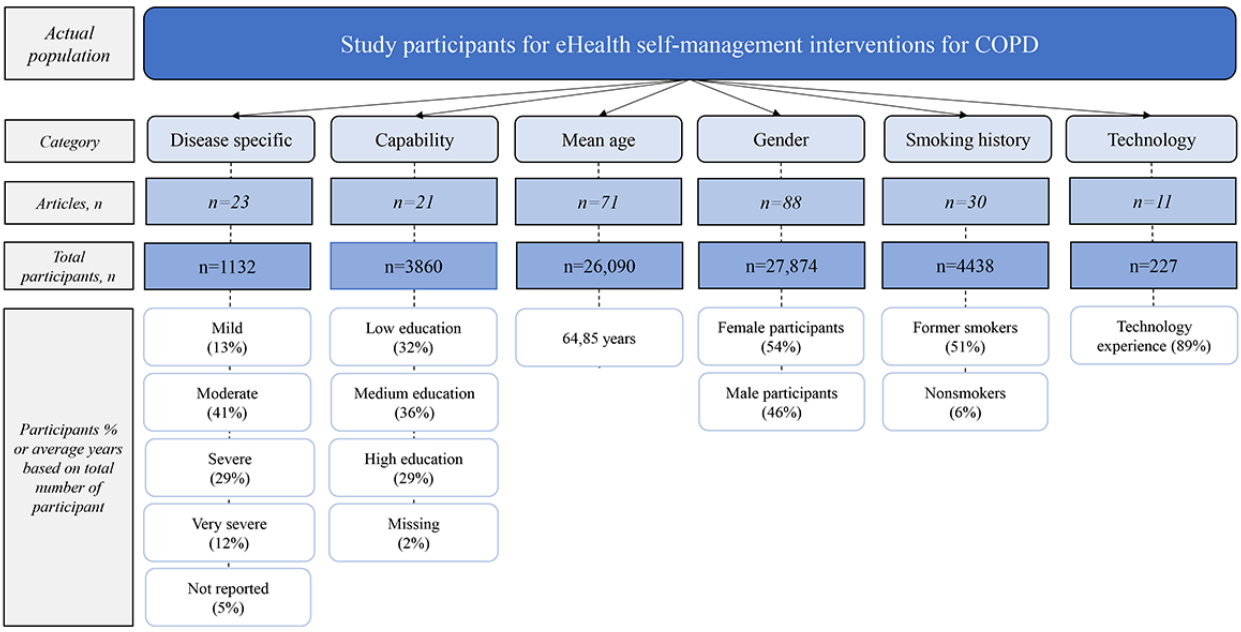

The Management in Self-Management

This section describes which self-management processes and BCTs were found within the different eHealth technologies. Details about this section are provided in Multimedia Appendix 7 (overview of self-management processes) and Multimedia Appendix 8 (overview of BCTs).

Self-Management Processes

Table 5 shows the self-management processes found in the articles. No article explicitly described which self-management processes were reflected in the intervention design. When analyzing how self-management processes were supported within the different studies, we identified that most studies (94/101, 93.1%) addressed the process of taking ownership toward health needs (eg, by including self-monitoring of symptoms or setting goals). Of the 101 included studies, 71 (70.3%) focused on the process of learning (eg, by including education within their technology), 27 (26.7%) on health care resources (eg, by enabling communication with health care professionals within the technology), 23 (22.8%) on performing health promotion activities (eg, by performing exercise or skill training), 17 (16.8%) on social resources (eg, by involving caregiver and family or peer-to-peer support), 1 (1%) on adjusting (eg, ways to cope), and 1 (1%) on integrating illness into daily life (eg, alternating daily lives to conserve energy). We found no eHealth technologies specifically focusing on the self-management processes: meaning making, spiritual resources, psychological resources, processing emotions, or community resources.

Table 5. Self-management processes within eHealth technologies. Each process is counted separately in this figure resulting in 234 self-management processes within 101 articles. “Meaning making,” “spiritual resources,” “psychological resources,” “processing emotions,” and “community resources” were not displayed, as these processes were not included.

Self-management process	Frequency, n (%)
Taking ownership of health needs	94 (40.2)
Learning	71 (30.3)
Health care resources	27 (11.5)
Performing health promotion activities	23 (9.8)
Social resources	17 (7.3)
Adjusting	1 (0.4)
Integrating illness into daily life	1 (0.4)

BCTs Used

Table 6 shows the BCTs extracted in this study. Only 2 (2%) out of 101 studies explicitly stated which BCTs were used. When analyzing the descriptions in the studies, we identified that feedback and monitoring were mostly used in the different articles (88/101, 87.1%; eg, monitoring activity status). This was followed by shaping knowledge (66/101, 65.3%; eg, receiving education), goals and planning (38/101, 37.6%; eg, action planning), associations (23/101, 22.8%; eg, receiving

XSL•FO
RenderX

status updates), social support (14/101, 13.7%; eg, communication with other people with COPD), regulation (11/101, 10.9%; eg, addressing medication adherence), repetition and substitution (10/101, 9.9%; eg, habit formation), rewards and threat (6/101, 5.9%; eg, receiving visual rewards), natural consequences (5/101, 4.9%; eg, information about health consequences), self-belief (5/101, 4.9%; eg, increasing self-efficacy), comparison of behavior (5/101, 4.9%; eg, follow along exercise video), comparison of outcomes (2/101, 2%; eg, information about the effect of physical activity), antecedents (1/101, 1%; eg, adding objects to the environment), and identity (1/101, 1%; eg, prompt identification as a role model). The BCTs of covert learning and scheduled consequences were not observed in the studies.

Table 6. Behavior change techniques within eHealth technologies. Each technique was counted separately in this figure. Resulting in 275 behavior change techniques within 101 articles. "Scheduled consequences" and "covert learning" were not displayed, as these processes were not included.

Behavior change technique	Frequency, n (%)
Feedback and monitoring	88 (32)
Shaping knowledge	66 (24)
Goals and planning	38 (13.8)
Associations	23 (8.4)
Social support	14 (5.1)
Regulation	11 (4)
Repetition and substitution	10 (3.6)
Rewards and threat	6 (2.2)
Natural consequences	5 (1.8)
Comparison of behavior	5 (1.8)
Self-belief	5 (1.8)
Comparison of outcomes	2 (0.7)
Identity	1 (0.4)
Antecedents	1 (0.4)

Discussion

Principal Findings

This scoping review outlines the state of the art of eHealth self-management interventions for COPD. In the current literature, most eHealth technologies for COPD self-management focus on the physical aspect of self-management. eHealth technologies that include other aspects (eg, the social or mental aspects) are currently underrepresented in the literature. Moreover, it appeared that although eHealth interventions often aimed to target the whole COPD population, mostly only a subgroup of the COPD population was represented within the eHealth technology studies.

Self-Management of COPD

Underlying Theories, Techniques, and Processes

Only a few studies (32/101, 31.7%) reported on using underlying theories and specific BCTs supporting their self-management eHealth interventions. No article explicitly mentioned focusing on certain self-management processes. This is surprising, given the fact that all studies aim to improve self-management and thus aim to achieve some sort of behavioral change. As the concept of self-management varies in the literature, reporting on the use of such processes, techniques, and theories may be beneficial for understanding underlying structures and processes that will initiate behavior change to improve self-management. Building on these processes, techniques, and theories and providing more detailed reports on what was perceived as useful, beneficial, and desirable for this target population can help advance the field and contribute to the existing body of work. This may simultaneously be valuable for informing future eHealth self-management initiatives as they can take into account these theories, techniques, and processes in their developments. When looking at the literature, the lack of reporting on underlying theories was prevalent in the review of Heimer et al [140], in which only 3 of the included studies reported specific BCTs. Furthermore, other studies encountered the problem of low reporting on BCTs; Hardeman et al [141] and Lorencatto et al [142] concluded that fewer than half of the planned BCTs were specified in the final published articles. In addition, a review from de Bruin et al [143] revealed that reporting about the active content of behavioral interventions varies considerably between studies. This limits the readers' ability to compare, interpret, and generalize the effects of these studies [143]. Thus, including such theories in eHealth interventions and transparency in later reporting may lead to opportunities for achieving sustainable behavioral change.

The Physical Aspects of Self-Management

As we found in this review, there is a tendency toward managing the physical aspect of one's disease in current eHealth technologies for COPD self-management. This is reflected throughout the different findings of this review. First, the functionality "self-monitoring" and the BCT "feedback and

XSL•FO
RenderX

monitoring" were most often addressed. Although self-monitoring is very valuable and exacerbations may be detected at an early stage, this, nonetheless, demonstrates that the main focus lies on what happens with or inside the body. Second, the dominant physical aspect also manifests itself when looking at the self-management processes that are supported by the different technologies: "psychological resources," "spiritual resources," and "community resources" were not found to be included. The self-management process of "taking ownership of health needs" was mostly present, followed by "learning." However, other processes, such as "integrating illness into daily life" and "adjusting," were only observed once, although the target group had to deal with these aspects every single day [144]. We believe that not addressing these processes is a missed opportunity, given that supporting people with COPD during their day-to-day activities might lead to even better improved outcomes of self-management. Finally, the dimensions of bodily functioning and daily functioning were most frequently used. This illustrates the current underrepresentation of other dimensions within current eHealth technologies for COPD self-management. Other dimensions (eg, participation and mental well-being) were not as dominantly represented or not observed at all (eg, meaningfulness and quality of life). As we observed a small increase in dimensions over the past few years, we might notice a small change of focus. However, this is not as fundamental and still leaves a lot of room for improvement on this matter. When examining other chronic diseases (eg, rheumatoid arthritis), a review by Seppen et al [145] identified 4 different types of eHealth interventions used in the included articles. Although not explicitly stated, interventions were all related to the physical aspect (ie, medication adherence, activity plan, information, disease monitoring, and activity monitoring) [145]. Thus, the tendency of the physical domain may not only be limited to COPD. Therefore, future studies should investigate whether this view is also present in other chronic diseases.

Inclusiveness and Representation of People With COPD

As all eHealth technologies target people with COPD as end users, only 14 (13.9%) out of 101 articles reported involving the patient perspective in their design or development process. This raised the question of whether and how the needs and perspectives of patients were taken into account. As including the perspective of end users leads to a better fit and increases the chances of successful adoption and sustained use [146], researchers should consider using such design principles when developing future eHealth technologies. This may provide many opportunities for improvements in self-management eHealth technologies for COPD.

Furthermore, it appeared that although articles outlined the target group to be the general population of people with COPD, they often recruited a specific subset of people with COPD. Certain inclusion criteria are made within the studies (eg, needing to own a smartphone and needing to have a certain disease severity). The consequence of such inclusion criteria leads to a restriction, in that only a selected group of individuals are included in the studies. While this may be due to practicalities (eg, the complexity of COPD as a progressive lung disease), the question remains whether the intervention is generalizable or applicable to the wider population of people with COPD, especially if they were not part of the studies in the first place. This is particularly relevant as there is no golden standard to determine, for example, when someone is considered too old to participate, and opinions on such matters are likely to vary widely among researchers. Even when the restricting the patient group may be justified for the study purposes, it still affects the generalizability of study results. Therefore, awareness and transparency should be provided regarding these potential restrictions related to the patient group within such studies.

Given that the most often used device is a (smart) measurement device (in combination with), a smartphone, or tablet, the group of people eligible to use these technologies in daily life is further limited. Effectively, this means that certain groups of people (eg, those who lack resources to buy these devices) are not included in the research and, therefore, the intervention might not be tested on people who are unfamiliar with smart devices or have low digital literacy. Previous studies showed that moderate levels of eHealth literacy and low levels of health literacy are prevalent among the COPD population [32,33]. Thus, we cannot assume that this population has access to eHealth technologies and has mastered the skills to manage such interventions without any support. Therefore, guidance should be made available to help those who need support in using these eHealth technologies. Furthermore, although there might be some practical reasoning behind the inclusion criteria (eg, lack of budget to provide devices for all participants), it may simultaneously widen the gap between people included in eHealth technology studies and the group of people who might need the support the most. This, in turn, could make health care accessible only to those with high (digital) literacy and those already equipped with the necessary resources to improve their health. This is most likely an unintentional and undesirable direction to head, but without proper awareness, it may easily become the blind spot in current and future eHealth intervention studies. Future studies must, therefore, be aware of possible subgroups, make efforts to include the underprivileged population, and be transparent in their research regarding the population reached. While we acknowledge the challenge of successfully recruiting participants who are representative of the population as a whole and recognize that this is extremely difficult to achieve, we nevertheless recommend that future studies strive to reach those people who are underrepresented and difficult to reach.

Strengths and Limitations

This review provides a very first overview and a diverse insight into different underlying components of self-management eHealth interventions for COPD. It highlights the existing gaps in the literature and uncovers opportunities for the development of eHealth self-management interventions. To the best of our knowledge, no research has yet examined all these various aspects.

However, this review also has its limitations. First, data extraction and categorization were challenging due to the style of writing in articles (which comes with certain formats and word count limitations), the lack of explicit reporting on certain aspects, and the overlap between some processes and

XSL•FO
RenderX

dimensions. It might be the case that through incomplete reporting in articles, certain self-management processes, BCTs, health dimensions, aspects of the technology, or details about the "self" could not be extracted. However, albeit being a scoping review, we followed a very systematic approach to mitigate this limitation as much as possible. Therefore, we believe that we were able to give a complete picture of the current literature about eHealth technologies for COPD self-management.

Second, some dimensions of positive health and self-management processes are closely related, intertwined with, or support each other. For example, the dimension of "quality of life" was not observed to be explicitly addressed within the eHealth interventions. However, studies may have an overarching goal of increasing the quality of life of people through the use of their intervention. One should be aware of this when interpreting these results, as the quality of life of people with COPD might still be affected by the use of eHealth intervention.

Third, most articles (470/588, 80%) were screened by 1 reviewer, leaving room for potential subjective judgment. However, by implementing several measures to ensure alignment (eg, extensive discussion of eligibility criteria, 20% of the screening conducted by 2 reviewers, and discussions to resolve discrepancies), we minimized possible implicit biases.

Finally, this scoping review investigated a broad range of aspects to grasp the state of the art, but not all. For example, this paper did not assess the effectiveness or impact of the interventions, and it also did not evaluate the quality or strength of the evidence. As this is a relatively new area of research, we should map the existing literature first. Furthermore, this review included only English articles, used COPD as a search term while it included a broad spectrum of lung diseases (such as emphysema and chronic bronchitis), and had a search date limitation. Consequently, some studies might have been missed, or some aspects may not have been investigated. However, to the best of our ability, we tried to provide a first overview while leaving opportunities for future research to focus on the aspects that were not covered within this scoping review. As such, this state-of-the-art overview could serve as a starting point for future systematic reviews and original research that will dive into more specific research areas.

Conclusions

This scoping review provides an overview of the state-of-the-art eHealth technologies for COPD self-management interventions. We showed that current eHealth technologies tend to address the physical aspect of COPD self-management. These findings reveal a gap in the available literature, as many dimensions of the positive health paradigm and self-management processes are not addressed in current eHealth interventions for COPD self-management. However, as COPD is a chronic disease and exerts its impact on all aspects of one's life, the underrepresented dimensions and processes might be very important to include. This might give people with COPD the tools needed to be able to adapt toward a new balance in life, and this would consider the person as a whole instead of only the bodily representation in the context of a disease. Our review also showcases another gap, namely, the effect of inclusion criteria that leads to a subgroup of people with COPD included in eHealth technology studies. Therefore, one should be cautious when interpreting results, as this may give a distorted view of the COPD population within these studies. These gaps demonstrate the need for more inclusive research and design of eHealth self-management interventions for people with COPD, focusing on multiple dimensions of the health paradigm. Future work should, therefore, go beyond the physical dimension and focus on including individuals in research who could benefit most from eHealth self-management interventions.

Acknowledgments

The research is supported by the European project RE-SAMPLE, which is funded by the European Union's Horizon 2020 research and innovation program (965315).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

All authors made significant contributions to the work, including in the conception, study design, data collection, screening, data extraction, analysis, and interpretation or in all of these areas. They also participated in drafting, revising, or critically reviewing the article and gave final approval for the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items For Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 81 KB - jmir_v27i1e57649_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Search string.
[DOCX File , 24 KB - jmir_v27i1e57649_app2.docx]

Multimedia Appendix 3
Overview of the extraction and charting details.
[DOCX File , 28 KB - jmir_v27i1e57649_app3.docx]

Multimedia Appendix 4
Overview of the functionality and modality of eHealth interventions.
[DOCX File , 75 KB - jmir_v27i1e57649_app4.docx]

Multimedia Appendix 5
Overview of the positive health dimensions.
[DOCX File , 52 KB - jmir_v27i1e57649_app5.docx]

Multimedia Appendix 6
Overview of the self.
[DOCX File , 89 KB - jmir_v27i1e57649_app6.docx]

Multimedia Appendix 7
Overview of the self-management processes.
[DOCX File , 62 KB - jmir_v27i1e57649_app7.docx]

Multimedia Appendix 8
Overview of the behavior change techniques.
[DOCX File , 67 KB - jmir_v27i1e57649_app8.docx]

References

1. Bednarek M, Maciejewski J, Wozniak M, Kuca P, Zielinski J. Prevalence, severity and underdiagnosis of COPD in the primary care setting. Thorax 2008 May 01;63(5):402-407. [doi: 10.1136/thx.2007.085456] [Medline: 18234906]
2. Jankowska-Polańska B, Kasprzyk M, Chudiak A, Uchmanowicz I. Relation between illness acceptance and quality of life in patients with Chronic Obstructive Pulmonary Disease (COPD). Adv Respir Med 2015 Dec 21;84(1):3-10. [doi: 10.5603/piap.a2015.0079]
3. Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. The Lancet 2007 Sep;370(9589):765-773. [doi: 10.1016/s0140-6736(07)61380-4]
4. Chronic obstructive pulmonary disease (COPD). World Health Organization. 2024 Nov 6. URL: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) [accessed 2025-01-27]
5. Foreman KJ, Marquez N, Dolgert A, Fukutaki K, Fullman N, McGaughey M, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories. Lancet 2018 Nov 10;392(10159):2052-2090 [FREE Full text] [doi: 10.1016/S0140-6736(18)31694-5] [Medline: 30340847]
6. Gershon AS, Dolmage TE, Stephenson A, Jackson B. Chronic obstructive pulmonary disease and socioeconomic status: a systematic review. COPD 2012 Jun 12;9(3):216-226 [FREE Full text] [doi: 10.3109/15412555.2011.648030] [Medline: 22497534]
7. Global strategy for prevention, diagnosis and management of COPD: 2024 report. Global Initiative for Chronic Obstructive Lung Disease. 2024. URL: https://goldcopd.org/2024-gold-report/ [accessed 2024-10-28]
8. Sin DD, Doiron D, Agusti A, Anzueto A, Barnes PJ, Celli BR, et al. Air pollution and COPD: GOLD 2023 committee report. Eur Respir J 2023 May 23;61(5):2202469 [FREE Full text] [doi: 10.1183/13993003.02469-2022] [Medline: 36958741]
9. Valipour A, Aisanov Z, Avdeev S, Koblizek V, Kocan I, Kopitovic I, et al. Recommendations for COPD management in Central and Eastern Europe. Expert Rev Respir Med 2022 Feb 03;16(2):221-234 [FREE Full text] [doi: 10.1080/17476348.2021.2023498] [Medline: 35001780]
10. Huygens MW, Vermeulen J, Swinkels IC, Friele RD, van Schayck OC, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 08;16(1):232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]

XSL•FO
RenderX

11. Audulv A, Asplund K, Norbergh K. Who's in charge? The role of responsibility attribution in self-management among people with chronic illness. Patient Educ Couns 2010 Oct;81(1):94-100. [doi: 10.1016/j.pec.2009.12.007] [Medline: 20060256]
12. Effing TW, Bourbeau J, Vercoulen J, Apter AJ, Coultas D, Meek P, et al. Self-management programmes for COPD: moving forward. Chron Respir Dis 2012 Feb;9(1):27-35 [FREE Full text] [doi: 10.1177/1479972311433574] [Medline: 22308551]
13. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: an overview of reviews. Chron Respir Dis 2017 Aug 24;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
14. Hosseinzadeh H, Shnaigat M. Effectiveness of chronic obstructive pulmonary disease self-management interventions in primary care settings: a systematic review. Aust J Prim Health 2019;25(3):195. [doi: 10.1071/py18181]
15. Camus-García E, González-González AI, Heijmans M, Niño de Guzmán E, Valli C, Beltran J, et al. Self-management interventions for adults living with Chronic Obstructive Pulmonary Disease (COPD): the development of a Core Outcome Set for COMPAR-EU project. PLoS One 2021 Mar 1;16(3):e0247522 [FREE Full text] [doi: 10.1371/journal.pone.0247522] [Medline: 33647039]
16. Bourbeau J, van der Palen J. Promoting effective self-management programmes to improve COPD. Eur Respir J 2009 Mar 27;33(3):461-463 [FREE Full text] [doi: 10.1183/09031936.00001309] [Medline: 19251792]
17. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 04;8(8):CD011682 [FREE Full text] [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
18. Bourbeau J, Nault D, Dang-Tan T. Self-management and behaviour modification in COPD. Patient Educ Couns 2004 Mar;52(3):271-277 [FREE Full text] [doi: 10.1016/s0738-3991(03)00102-2]
19. Russell S, Ogunbayo OJ, Newham JJ, Heslop-Marshall K, Netts P, Hanratty B, et al. Qualitative systematic review of barriers and facilitators to self-management of chronic obstructive pulmonary disease: views of patients and healthcare professionals. NPJ Prim Care Respir Med 2018 Jan 17;28(1):2 [FREE Full text] [doi: 10.1038/s41533-017-0069-z] [Medline: 29343739]
20. Schulman-Green D, Jaser S, Martin F, Alonzo A, Grey M, McCorkle R, et al. Processes of self-management in chronic illness. J Nurs Scholarsh 2012 Jun;44(2):136-144 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01444.x] [Medline: 22551013]
21. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. van Gemert-Pijnen L, Kip H, Kelders SM, Sanderman R. Introducing eHealth. In: Kip H, Beerlage-de Jong N, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development. London, UK: Routledge; 2018.
23. Effing TW, Vercoulen JH, Bourbeau J, Trappenburg J, Lenferink A, Cafarella P, et al. Definition of a COPD self-management intervention: International Expert Group consensus. Eur Respir J 2016 Jul;48(1):46-54 [FREE Full text] [doi: 10.1183/13993003.00025-2016] [Medline: 27076595]
24. Bourne C, Houchen-Wolloff L, Patel P, Bankart J, Singh S. Self-management programme of activity coping and education-SPACE for COPD(C)-in primary care: a pragmatic randomised trial. BMJ Open Respir Res 2022 Oct 17;9(1):e001443 [FREE Full text] [doi: 10.1136/bmjresp-2022-001443] [Medline: 36253020]
25. Lundell S, Modig M, Holmner Å, Wadell K. Perceptions of home telemonitoring use among patients with chronic obstructive pulmonary disease: qualitative study. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16343 [FREE Full text] [doi: 10.2196/16343] [Medline: 32490844]
26. Marklund S, Tistad M, Lundell S, Östrand L, Sörlin A, Boström C, et al. Experiences and factors affecting usage of an eHealth tool for self-management among people with chronic obstructive pulmonary disease: qualitative study. J Med Internet Res 2021 Apr 30;23(4):e25672 [FREE Full text] [doi: 10.2196/25672] [Medline: 33929327]
27. Hallensleben C, van Luenen S, Rolink E, Ossebaard HC, Chavannes NH. eHealth for people with COPD in the Netherlands: a scoping review. Int J Chron Obstruct Pulmon Dis 2019 Jul;Volume 14:1681-1690. [doi: 10.2147/copd.s207187]
28. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 May 23;5(5):CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
29. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ 2011 Jul 26;343(jul26 2):d4163. [doi: 10.1136/bmj.d4163] [Medline: 21791490]
30. Huber M, van Vliet M, Giezenberg M, Winkens B, Heerkens Y, Dagnelie PC, et al. Towards a 'patient-centred' operationalisation of the new dynamic concept of health: a mixed methods study. BMJ Open 2016 Jan 12;6(1):e010091 [FREE Full text] [doi: 10.1136/bmjopen-2015-010091] [Medline: 26758267]

XSL•FO
RenderX

31. Kale MS, Federman AD, Krauskopf K, Wolf M, O'Conor R, Martynenko M, et al. The association of health literacy with illness and medication beliefs among patients with chronic obstructive pulmonary disease. PLoS One 2015;10(4):e0123937 [FREE Full text] [doi: 10.1371/journal.pone.0123937] [Medline: 25915420]
32. Roberts NJ, Ghiassi R, Partridge MR. Health literacy in COPD. Int J Chron Obstruct Pulmon Dis 2008 Aug;3(4):499-507 [FREE Full text] [doi: 10.2147/copd.s1088] [Medline: 19281068]
33. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based health information seeking and eHealth literacy among patients living with Chronic Obstructive Pulmonary Disease (COPD). Health Commun 2018 Dec 05;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
34. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008 Apr 16;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
35. Williams V, Price J, Hardinge M, Tarassenko L, Farmer A. Using a mobile health application to support self-management in COPD: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e392-e400 [FREE Full text] [doi: 10.3399/bjgp14X680473] [Medline: 24982491]
36. Sloots J, Bakker M, van der Palen J, Eijsvogel M, van der Valk P, Linssen G, et al. Adherence to an eHealth self-management intervention for patients with both COPD and heart failure: results of a pilot study. COPD 2021 Jul;Volume 16:2089-2103. [doi: 10.2147/copd.s299598]
37. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
38. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
39. Nationale implementatie agenda (NIA) eHEALTH. Eerste Kamer der Staten-Generaal. 2012 Jun 7. URL: https://eerstekamer.nl/9370000/1/j4nvjlhjvvt9eu4_j9vvkfvj6b325az/vj0bfx0w1pxt [accessed 2013-02-25]
40. Rayyan homepage. Rayyan. URL: https://www.rayyan.ai/ [accessed 2023-10-19]
41. ATLAS.ti scientific software development GmbH, version 9.1.7. ATLAS.ti. 2021 Nov 11. URL: https://atlasti.com/updates [accessed 2025-01-27]
42. Jansen-Kosterink S, Broekhuis M, van Velsen L. Time to act mature-gearing eHealth evaluations towards technology readiness levels. Digit Health 2022 Jul 11;8:20552076221113396 [FREE Full text] [doi: 10.1177/20552076221113396] [Medline: 35847525]
43. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 02;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
44. Ali L, Wallström S, Fors A, Barenfeld E, Fredholm E, Fu M, et al. Effects of person-centered care using a digital platform and structured telephone support for people with chronic obstructive pulmonary disease and chronic heart failure: randomized controlled trial. J Med Internet Res 2021 Dec 13;23(12):e26794 [FREE Full text] [doi: 10.2196/26794] [Medline: 34898447]
45. Choi JY, George M, Yun SY. Development of a smartphone application for Korean patients with chronic obstructive pulmonary disease: self-monitoring based action plans. Appl Nurs Res 2021 Oct;61:151475. [doi: 10.1016/j.apnr.2021.151475] [Medline: 34544569]
46. Korpershoek YJ, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JC. User-centered design of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease (Copilot): mixed methods study. J Med Internet Res 2020 Jun 15;22(6):e15449 [FREE Full text] [doi: 10.2196/15449] [Medline: 32538793]
47. Korpershoek YJ, Holtrop T, Vervoort SC, Schoonhoven L, Schuurmans MJ, Trappenburg JC. Early-stage feasibility of a mobile health intervention (Copilot) to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease: multimethods approach. JMIR Form Res 2020 Nov 19;4(11):e21577 [FREE Full text] [doi: 10.2196/21577] [Medline: 33211013]
48. Wang L, Guo Y, Wang M, Zhao Y. A mobile health application to support self-management in patients with chronic obstructive pulmonary disease: a randomised controlled trial. Clin Rehabil 2021 Jan;35(1):90-101. [doi: 10.1177/0269215520946931] [Medline: 32907384]
49. Doyle J, Murphy E, Gavin S, Pascale A, Deparis S, Tommasi P, et al. A digital platform to support self-management of multiple chronic conditions (ProACT): findings in relation to engagement during a one-year proof-of-concept trial. J Med Internet Res 2021 Dec 15;23(12):e22672 [FREE Full text] [doi: 10.2196/22672] [Medline: 34914612]
50. Cameron-Tucker HL, Wood-Baker R, Joseph L, Walters JA, Schüz N, Walters EH. A randomized controlled trial of telephone-mentoring with home-based walking preceding rehabilitation in COPD. Int J Chron Obstruct Pulmon Dis 2016;11:1991-2000 [FREE Full text] [doi: 10.2147/COPD.S109820] [Medline: 27601892]

XSL•FO
RenderX

51. Farmer A, Williams V, Velardo C, Shah SA, Yu LM, Rutter H, et al. Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2017 May 03;19(5):e144 [FREE Full text] [doi: 10.2196/jmir.7116] [Medline: 28468749]
52. Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. BMC Med Inform Decis Mak 2015 Jun 18;15:46 [FREE Full text] [doi: 10.1186/s12911-015-0171-5] [Medline: 26084626]
53. Velardo C, Shah SA, Gibson O, Clifford G, Heneghan C, Rutter H, et al. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):19 [FREE Full text] [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]
54. Whelan ME, Velardo C, Rutter H, Tarassenko L, Farmer AJ. Mood monitoring over one year for people with chronic obstructive pulmonary disease using a mobile health system: retrospective analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 22;7(11):e14946 [FREE Full text] [doi: 10.2196/14946] [Medline: 31755872]
55. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and web-based monitoring and feedback tool embedded in primary care increases physical activity: a cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
56. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. A pilot study of a tool to stimulate physical activity in patients with COPD or type 2 diabetes in primary care. J Telemed Telecare 2014 Jan 10;20(1):29-34. [doi: 10.1177/1357633x13519057]
57. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013 Jan 16;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
58. Voncken-Brewster V, Tange H, de Vries H, Nagykaldi Z, Winkens B, van der Weijden T. A randomized controlled trial evaluating the effectiveness of a web-based, computer-tailored self-management intervention for people with or at risk for COPD. Int J Chron Obstruct Pulmon Dis 2015 Jun 08;10:1061-1073 [FREE Full text] [doi: 10.2147/COPD.S81295] [Medline: 26089656]
59. Voncken-Brewster V, Tange H, Moser A, Nagykaldi Z, de Vries H, van der Weijden T. Integrating a tailored e-health self-management application for chronic obstructive pulmonary disease patients into primary care: a pilot study. BMC Fam Pract 2014 Jan 08;15:4 [FREE Full text] [doi: 10.1186/1471-2296-15-4] [Medline: 24400676]
60. Voncken-Brewster V, Amoureus M, de Vries H, Nagykaldi Z, Winkens B, van der Weijden T, et al. The impact of participant characteristics on use and satisfaction of a web-based computer-tailored chronic obstructive pulmonary disease self-management intervention: a process evaluation. JMIR Form Res 2017 Aug 21;1(1):e1 [FREE Full text] [doi: 10.2196//formative.6585] [Medline: 30684399]
61. Bourbeau J, Casan P, Tognella S, Haidl P, Texereau JB, Kessler R. An international randomized study of a home-based self-management program for severe COPD: the COMET. Int J Chron Obstruct Pulmon Dis 2016 Jun 28;11:1447-1451 [FREE Full text] [doi: 10.2147/COPD.S107151] [Medline: 27418817]
62. Kessler R, Casan-Clara P, Koehler D, Tognella S, Viejo JL, Dal Negro RW, et al. COMET: a multicomponent home-based disease-management programme versus routine care in severe COPD. Eur Respir J 2018 Jan 11;51(1):1701612 [FREE Full text] [doi: 10.1183/13993003.01612-2017] [Medline: 29326333]
63. Alharbey R, Chatterjee S. An mHealth assistive system "MyLung" to empower patients with chronic obstructive pulmonary disease: design science research. JMIR Form Res 2019 Mar 19;3(1):e12489 [FREE Full text] [doi: 10.2196/12489] [Medline: 30888329]
64. Coultas DB, Jackson BE, Russo R, Peoples J, Singh KP, Sloan J, et al. Home-based physical activity coaching, physical activity, and health care utilization in chronic obstructive pulmonary disease. Chronic obstructive pulmonary disease self-management activation research trial secondary outcomes. Ann Am Thorac Soc 2018 Apr;15(4):470-478 [FREE Full text] [doi: 10.1513/AnnalsATS.201704-308OC] [Medline: 29283670]
65. Coventry PA, Blakemore A, Baker E, Sidhu M, Fitzmaurice D, Jolly K. The push and pull of self-managing mild COPD: an evaluation of participant experiences of a nurse-led telephone health coaching intervention. Qual Health Res 2018 Dec 01;29(5):658-671. [doi: 10.1177/1049732318809679]
66. Jolly K, Sidhu MS, Hewitt CA, Coventry PA, Daley A, Jordan R, et al. Self management of patients with mild COPD in primary care: randomised controlled trial. BMJ 2018 Jun 13;361:k2241 [FREE Full text] [doi: 10.1136/bmj.k2241] [Medline: 29899047]
67. Park SK, Bang CH, Lee SH. Evaluating the effect of a smartphone app-based self-management program for people with COPD: a randomized controlled trial. Appl Nurs Res 2020 Apr;52:151231. [doi: 10.1016/j.apnr.2020.151231] [Medline: 31955942]
68. Bugajski A, Frazier SK, Cousin L, Rechenberg K, Brown J, Lengerich AJ, et al. Effects of a digital self-care intervention in adults with COPD: a pilot study. West J Nurs Res 2019 Dec 19;42(9):736-746. [doi: 10.1177/0193945919892282]
69. Hoaas H, Andreassen HK, Lien LA, Hjalmarsen A, Zanaboni P. Adherence and factors affecting satisfaction in long-term telerehabilitation for patients with chronic obstructive pulmonary disease: a mixed methods study. BMC Med Inform Decis Mak 2016 Feb 25;16:26 [FREE Full text] [doi: 10.1186/s12911-016-0264-9] [Medline: 26911326]

XSL•FO
RenderX

70. Miller S, Teufel R2, Nichols M, Davenport P, Mueller M, Silverman E, et al. Feasibility of RESP-FIT: technology-enhanced self-management intervention for adults with COPD. Int J Chron Obstruct Pulmon Dis 2021 Dec 03;16:3263-3273 [FREE Full text] [doi: 10.2147/COPD.S326675] [Medline: 34887656]
71. Thomas RM, Locke ER, Woo DM, Nguyen EH, Press VG, Layouni TA, et al. Inhaler training delivered by internet-based home videoconferencing improves technique and quality of life. Respir Care 2017 Nov;62(11):1412-1422. [doi: 10.4187/respcare.05445]
72. Bentley CL, Powell L, Potter S, Parker J, Mountain GA, Bartlett YK, et al. The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16203 [FREE Full text] [doi: 10.2196/16203] [Medline: 32490838]
73. Robinson SA, Wan ES, Shimada SL, Richardson CR, Moy ML. Age and attitudes towards an internet-mediated, pedometer-based physical activity intervention for chronic obstructive pulmonary disease: secondary analysis. JMIR Aging 2020 Sep 09;3(2):e19527. [doi: 10.2196/19527] [Medline: 32902390]
74. Barenfeld E, Ali L, Wallström S, Fors A, Ekman I. Becoming more of an insider: a grounded theory study on patients' experience of a person-centred e-health intervention. PLoS One 2020 Nov 23;15(11):e0241801 [FREE Full text] [doi: 10.1371/journal.pone.0241801] [Medline: 33226986]
75. Burkow TM, Vognild LK, Johnsen E, Bratvold A, Risberg MJ. Promoting exercise training and physical activity in daily life: a feasibility study of a virtual group intervention for behaviour change in COPD. BMC Med Inform Decis Mak 2018 Dec 18;18(1):136 [FREE Full text] [doi: 10.1186/s12911-018-0721-8] [Medline: 30563507]
76. Alcazar B, de Lucas P, Soriano JB, Fernández-Nistal A, Fuster A, González-Moro JM, et al. The evaluation of a remote support program on quality of life and evolution of disease in COPD patients with frequent exacerbations. BMC Pulm Med 2016 Nov 08;16(1):140 [FREE Full text] [doi: 10.1186/s12890-016-0304-3] [Medline: 27821164]
77. Au DH, Macaulay DS, Jarvis JL, Desai US, Birnbaum HG. Impact of a telehealth and care management program for patients with chronic obstructive pulmonary disease. Ann Am Thorac Soc 2015 Mar;12(3):323-331. [doi: 10.1513/AnnalsATS.201501-042OC] [Medline: 25642649]
78. Barenfeld E, Fuller JM, Wallström S, Fors A, Ali L, Ekman I. Meaningful use of a digital platform and structured telephone support to facilitate remote person-centred care – a mixed-method study on patient perspectives. BMC Health Serv Res 2022 Apr 04;22:442. [doi: 10.1186/s12913-022-07831-8]
79. Barken TL, Thygesen E, Söderhamn U. Unlocking the limitations: living with chronic obstructive pulmonary disease and receiving care through telemedicine-a phenomenological study. J Clin Nurs 2018 Jan;27(1-2):132-142. [doi: 10.1111/jocn.13857] [Medline: 28425194]
80. Benzo RP, Kramer KM, Hoult JP, Anderson PM, Begue IM, Seifert SJ. Development and feasibility of a home pulmonary rehabilitation program with health coaching. Respir Care 2018 Feb;63(2):131-140 [FREE Full text] [doi: 10.4187/respcare.05690] [Medline: 29066590]
81. Billington J, Coster S, Murrells T, Norman I. Evaluation of a nurse-led educational telephone intervention to support self-management of patients with chronic obstructive pulmonary disease: a randomized feasibility study. COPD 2015 Aug;12(4):395-403 [FREE Full text] [doi: 10.3109/15412555.2014.974735] [Medline: 25474080]
82. Bødker M, Juul Nielsen A. Providing rehabilitation online - invisible work and diagnostic agents. J Health Organ Manag 2015;29(7):948-964 [FREE Full text] [doi: 10.1108/JHOM-06-2014-0091] [Medline: 26556161]
83. Boer LM, van der Heijden M, van Kuijk NM, Lucas PJ, Vercoulen JH, Assendelft WJ, et al. Validation of ACCESS: an automated tool to support self-management of COPD exacerbations. Int J Chron Obstruct Pulmon Dis 2018 Oct 10;13:3255-3267 [FREE Full text] [doi: 10.2147/COPD.S167272] [Medline: 30349231]
84. Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, et al. A smart mobile health tool versus a paper action plan to support self-management of chronic obstructive pulmonary disease exacerbations: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 09;7(10):e14408 [FREE Full text] [doi: 10.2196/14408] [Medline: 31599729]
85. Burkow TM, Vognild LK, Østengen G, Johnsen E, Risberg MJ, Bratvold A, et al. Internet-enabled pulmonary rehabilitation and diabetes education in group settings at home: a preliminary study of patient acceptability. BMC Med Inform Decis Mak 2013 Mar 05;13:33 [FREE Full text] [doi: 10.1186/1472-6947-13-33] [Medline: 23496829]
86. Burkow TM, Vognild LK, Johnsen E, Risberg MJ, Bratvold A, Breivik E, et al. Comprehensive pulmonary rehabilitation in home-based online groups: a mixed method pilot study in COPD. BMC Res Notes 2015 Dec 10;8:766 [FREE Full text] [doi: 10.1186/s13104-015-1713-8] [Medline: 26651831]
87. Cooper R, Giangreco A, Duffy M, Finlayson E, Hamilton S, Swanson M, et al. Evaluation of myCOPD digital self-management technology in a remote and rural population: real-world feasibility study. JMIR Mhealth Uhealth 2022 Feb 07;10(2):e30782 [FREE Full text] [doi: 10.2196/30782] [Medline: 35129453]
88. Criner GJ, Cole T, Hahn KA, Kastango K, Eudicone JM, Gilbert I. Use of a digital chronic obstructive pulmonary disease respiratory tracker in a primary care setting: a feasibility study. Pulm Ther 2021 Dec;7(2):533-547 [FREE Full text] [doi: 10.1007/s41030-021-00168-3] [Medline: 34463947]

XSL•FO
RenderX

89. De San Miguel K, Smith J, Lewin G. Telehealth remote monitoring for community-dwelling older adults with chronic obstructive pulmonary disease. Telemed J E Health 2013 Sep;19(9):652-657. [doi: 10.1089/tmj.2012.0244] [Medline: 23808885]
90. Deng N, Chen J, Liu Y, Wei S, Sheng L, Lu R, et al. Using mobile health technology to deliver a community-based closed-loop management system for chronic obstructive pulmonary disease patients in remote areas of China: development and prospective observational study. JMIR Mhealth Uhealth 2020 Nov 25;8(11):e15978 [FREE Full text] [doi: 10.2196/15978] [Medline: 33237036]
91. Dhadge N, Shevade M, Kale N, Narke G, Pathak D, Barne M, et al. Monitoring of inhaler use at home with a smartphone video application in a pilot study. NPJ Prim Care Respir Med 2020 Oct 16;30(1):46 [FREE Full text] [doi: 10.1038/s41533-020-00203-x] [Medline: 33067469]
92. Early F, Young JS, Robinshaw E, Mi EZ, Mi EZ, Fuld JP. A case series of an off-the-shelf online health resource with integrated nurse coaching to support self-management in COPD. Int J Chron Obstruct Pulmon Dis 2017 Oct 09;12:2955-2967 [FREE Full text] [doi: 10.2147/COPD.S139532] [Medline: 29070947]
93. Farias R, Sedeno M, Beaucage D, Drouin I, Ouellet I, Joubert A, et al. Innovating the treatment of COPD exacerbations: a phone interactive telesystem to increase COPD Action Plan adherence. BMJ Open Respir Res 2019 Apr 09;6(1):e000379 [FREE Full text] [doi: 10.1136/bmjresp-2018-000379] [Medline: 31178998]
94. Fitzsimmons DA, Thompson J, Bentley CL, Mountain GA. Comparison of patient perceptions of telehealth-supported and specialist nursing interventions for early stage COPD: a qualitative study. BMC Health Serv Res 2016 Aug 22;16(1):420 [FREE Full text] [doi: 10.1186/s12913-016-1623-z] [Medline: 27549751]
95. Haesum LK, Soerensen N, Dinesen B, Nielsen C, Grann O, Hejlesen O, et al. Cost-utility analysis of a telerehabilitation program: a case study of COPD patients. Telemed J E Health 2012 Nov;18(9):688-692. [doi: 10.1089/tmj.2011.0250] [Medline: 23020647]
96. Hoaas H, Morseth B, Holland AE, Zanaboni P. Are physical activity and benefits maintained after long-term telerehabilitation in COPD? Int J Telerehabil 2016 Dec 15;8(2):39-48 [FREE Full text] [doi: 10.5195/ijt.2016.6200] [Medline: 28775800]
97. Houchen-Wolloff L, Orme M, Barradell A, Clinch L, Chaplin E, Gardiner N, et al. Web-based self-management program (SPACE for COPD) for individuals hospitalized with an acute exacerbation of chronic obstructive pulmonary disease: nonrandomized feasibility trial of acceptability. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e21728 [FREE Full text] [doi: 10.2196/21728] [Medline: 34114960]
98. Huniche L, Dinesen B, Nielsen C, Grann O, Toft E. Patients' use of self-monitored readings for managing everyday life with COPD: a qualitative study. Telemed J E Health 2013 May;19(5):396-402. [doi: 10.1089/tmj.2012.0135] [Medline: 23531094]
99. Kargiannakis M, Fitzsimmons DA, Bentley CL, Mountain GA. Does telehealth monitoring identify exacerbations of chronic obstructive pulmonary disease and reduce hospitalisations? An analysis of system data. JMIR Med Inform 2017 Mar 22;5(1):e8 [FREE Full text] [doi: 10.2196/medinform.6359] [Medline: 28330829]
100. Kaye L, Gondalia R, Thompson A, Stempel DA, Barrett MA. The relationship between objective app engagement and medication adherence in asthma and COPD: a retrospective analysis. Sci Rep 2021 Dec 21;11(1):24343 [FREE Full text] [doi: 10.1038/s41598-021-03827-2] [Medline: 34934164]
101. Kjellsdotter A, Andersson S, Berglund M. Together for the future - development of a digital website to support chronic obstructive pulmonary disease self-management: a qualitative study. J Multidiscip Healthc 2021 Apr 06;14:757-766 [FREE Full text] [doi: 10.2147/JMDH.S302013] [Medline: 33854327]
102. Knox L, Gemine R, Rees S, Bowen S, Groom P, Taylor D, et al. Assessing the uptake, engagement, and safety of a self-management app, COPD.Pal®, for chronic obstructive pulmonary disease: a pilot study. Health Technol 2021 Feb 24;11:557-562. [doi: 10.1007/s12553-021-00534-w]
103. Koff PB, Min SJ, Freitag TJ, Diaz DL, James SS, Voelkel NF, et al. Impact of proactive integrated care on chronic obstructive pulmonary disease. Chronic Obstr Pulm Dis 2021 Jan;8(1):100-116 [FREE Full text] [doi: 10.15326/jcopdf.2020.0139] [Medline: 33238087]
104. Kooij L, Vos PJ, Dijkstra A, van Harten WH. Effectiveness of a mobile health and self-management app for high-risk patients with chronic obstructive pulmonary disease in daily clinical practice: mixed methods evaluation study. JMIR Mhealth Uhealth 2021 Feb 04;9(2):e21977 [FREE Full text] [doi: 10.2196/21977] [Medline: 33538699]
105. Lee AC, Oliver S, Fletcher K, Robinson J. The impact of telehealth support for patients with diabetes or chronic obstructive pulmonary disease on unscheduled secondary care utilisation: a service evaluation. Inform Prim Care 2012;20(4):263-270 [FREE Full text] [doi: 10.14236/jhi.v20i4.16] [Medline: 23890338]
106. Lilholt PH, Hæsum LK, Hejlesen OK. Exploring user experience of a telehealth system for the Danish TeleCare North trial. Stud Health Technol Inform 2015;210:301-305. [Medline: 25991154]
107. Maathuis I, Jones VM, Oudshoorn N. Telecare and self-management: a guideline for anticipating future care in scenario-based design. Interact Des Archit(s) J 2014 Dec 20;23(1):142-159. [doi: 10.55612/s-5002-023-009]
108. Mark DD, Ikehara C, Matsuura C, Hara K, Li D. Validating the impact of teaching pursed-lips breathing with Skype: a pilot study. J Hosp Palliat Nurs 2013 Dec;15(8):424-432. [doi: 10.1097/NJH.0000000000000015]

XSL•FO
RenderX

109. Marquis N, Larivée P, Dubois MF, Tousignant M. Are improvements maintained after in-home pulmonary telerehabilitation for patients with chronic obstructive pulmonary disease? Int J Telerehabil 2015 Jan 29;6(2):21-30 [FREE Full text] [doi: 10.5195/ijt.2014.6156] [Medline: 25945226]
110. Mathar H, Fastholm P, Sandholm N. A qualitative study of televideo consultations for COPD patients. Br J Nurs 2015;24(4):205-209. [doi: 10.12968/bjon.2015.24.4.205] [Medline: 25723262]
111. Mierdel S, Owen K. Telehomecare reduces ER use and hospitalizations at William Osler Health System. Stud Health Technol Inform 2015;209:102-108. [Medline: 25980711]
112. Nield M, Hoo GW. Real-time telehealth for COPD self-management using Skype™. COPD 2012 Dec 04;9(6):611-619 [FREE Full text] [doi: 10.3109/15412555.2012.708067] [Medline: 22946768]
113. North M, Bourne S, Green B, Chauhan AJ, Brown T, Winter J, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: the RESCUE trial. NPJ Digit Med 2020 Oct 30;3(1):145 [FREE Full text] [doi: 10.1038/s41746-020-00347-7] [Medline: 33145441]
114. Nyberg A, Tistad M, Wadell K. Can the COPD web be used to promote self-management in patients with COPD in Swedish primary care: a controlled pragmatic pilot trial with 3 month- and 12 month follow-up. Scand J Prim Health Care 2019 Mar 31;37(1):69-82 [FREE Full text] [doi: 10.1080/02813432.2019.1569415] [Medline: 30700230]
115. Orme MW, Weedon AE, Saukko PM, Esliger DW, Morgan MD, Steiner MC, et al. Findings of the chronic obstructive pulmonary disease-sitting and exacerbations trial (COPD-SEAT) in reducing sedentary time using wearable and mobile technologies with educational support: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e84 [FREE Full text] [doi: 10.2196/mhealth.9398] [Medline: 29643055]
116. Patel N, Kinmond K, Jones P, Birks P, Spiteri MA. Validation of COPDPredict™: unique combination of remote monitoring and exacerbation prediction to support preventative management of COPD exacerbations. Int J Chron Obstruct Pulmon Dis 2021 Jun;Volume 16:1887-1899. [doi: 10.2147/copd.s309372]
117. Rassouli F, Boutellier D, Duss J, Huber S, Brutsche MH. Digitalizing multidisciplinary pulmonary rehabilitation in COPD with a smartphone application: an international observational pilot study. Int J Chron Obstruct Pulmon Dis 2018 Nov;Volume 13:3831-3836. [doi: 10.2147/copd.s182880]
118. Rixon L, Hirani SP, Cartwright M, Beynon M, Doll H, Steventon A, et al. A RCT of telehealth for COPD patient's quality of life: the whole system demonstrator evaluation. Clin Respir J 2017 Jul 17;11(4):459-469. [doi: 10.1111/crj.12359] [Medline: 26260325]
119. Rodriguez Hermosa JL, Fuster Gomila A, Puente Maestu L, Amado Diago CA, Callejas González FJ, Malo De Molina Ruiz R, et al. Compliance and utility of a smartphone app for the detection of exacerbations in patients with chronic obstructive pulmonary disease: cohort study. JMIR Mhealth Uhealth 2020 Mar 19;8(3):e15699 [FREE Full text] [doi: 10.2196/15699] [Medline: 32191213]
120. Schnoor K, Versluis A, Bakema R, van Luenen S, Kooij MJ, van den Heuvel JM, et al. A pharmacy-based ehealth intervention promoting correct use of medication in patients with asthma and COPD: nonrandomized pre-post study. J Med Internet Res 2022 Jun 08;24(6):e32396 [FREE Full text] [doi: 10.2196/32396] [Medline: 35675120]
121. Sheridan A, Jennings A, Keane S, Power A, Kavanagh P. "A breath of fresh air" for tackling chronic disease in Ireland? An evaluation of a self-management support service for people with chronic respiratory diseases. Ir J Med Sci 2020 May 28;189(2):551-556. [doi: 10.1007/s11845-019-02081-w] [Medline: 31463893]
122. Sieverink F, Kelders S, Braakman-Jansen A, van Gemert-Pijnen J. Evaluating the implementation of a personal health record for chronic primary and secondary care: a mixed methods approach. BMC Med Inform Decis Mak 2019 Nov 27;19(1):241 [FREE Full text] [doi: 10.1186/s12911-019-0969-7] [Medline: 31775734]
123. Stamenova V, Liang K, Yang R, Engel K, van Lieshout F, Lalingo E, et al. Technology-enabled self-management of chronic obstructive pulmonary disease with or without asynchronous remote monitoring: randomized controlled trial. J Med Internet Res 2020 Jul 30;22(7):e18598 [FREE Full text] [doi: 10.2196/18598] [Medline: 32729843]
124. Steventon A, Tunkel S, Blunt I, Bardsley M. Effect of telephone health coaching (Birmingham OwnHealth) on hospital use and associated costs: cohort study with matched controls. BMJ 2013 Aug 06;347(aug06 2):f4585 [FREE Full text] [doi: 10.1136/bmj.f4585] [Medline: 23920348]
125. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Looijmans-van den Akker I, et al. The effect of integration of self-management web platforms on health status in chronic obstructive pulmonary disease management in primary care (e-Vita study): interrupted time series design. J Med Internet Res 2017 Aug 16;19(8):e291 [FREE Full text] [doi: 10.2196/jmir.8262] [Medline: 28814380]
126. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Numans ME, et al. High level of integration in integrated disease management leads to higher usage in the e-Vita study: self-management of chronic obstructive pulmonary disease with web-based platforms in a parallel cohort design. J Med Internet Res 2017 May 31;19(5):e185 [FREE Full text] [doi: 10.2196/jmir.7037] [Medline: 28566268]
127. Talboom-Kamp EP, Holstege MS, Chavannes NH, Kasteleyn MJ. Effects of use of an eHealth platform e-Vita for COPD patients on disease specific quality of life domains. Respir Res 2019 Jul 10;20(1):146 [FREE Full text] [doi: 10.1186/s12931-019-1110-2] [Medline: 31291945]

XSL•FO
RenderX

128. Tabak M, VanderValk P, Hermens H, Vollenbroek-Hutten M, Brusse-Keizer M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2014 Sep;9(1):935-944. [doi: 10.2147/copd.s60179]
129. Ter Stal S, Sloots J, Ramlal A, Op den Akker H, Lenferink A, Tabak M. An embodied conversational agent in an eHealth self-management intervention for chronic obstructive pulmonary disease and chronic heart failure: exploratory study in a real-life setting. JMIR Hum Factors 2021 Nov 04;8(4):e24110 [FREE Full text] [doi: 10.2196/24110] [Medline: 34734824]
130. van Buul AR, Wildschut TS, Bonten TN, Kasteleyn MJ, Slats AM, Chavannes NH, et al. A systematic diagnostic evaluation combined with an internet-based self-management support system for patients with asthma or COPD. Int J Chron Obstruct Pulmon Dis 2018;13:3297-3306 [FREE Full text] [doi: 10.2147/COPD.S175361] [Medline: 30349234]
131. van Buul AR, Derksen C, Hoedemaker O, van Dijk O, Chavannes NH, Kasteleyn MJ. eHealth program to reduce hospitalizations due to acute exacerbation of chronic obstructive pulmonary disease: retrospective study. JMIR Form Res 2021 Mar 18;5(3):e24726 [FREE Full text] [doi: 10.2196/24726] [Medline: 33734091]
132. van der Heijden M, Lucas PJ, Lijnse B, Heijdra YF, Schermer TR. An autonomous mobile system for the management of COPD. J Biomed Inform 2013 Jun;46(3):458-469 [FREE Full text] [doi: 10.1016/j.jbi.2013.03.003] [Medline: 23500485]
133. van Lieshout F, Yang R, Stamenova V, Agarwal P, Cornejo Palma D, Sidhu A, et al. Evaluating the implementation of a remote-monitoring program for chronic obstructive pulmonary disease: qualitative methods from a service design perspective. J Med Internet Res 2020 Oct 09;22(10):e18148 [FREE Full text] [doi: 10.2196/18148] [Medline: 33034565]
134. van Zelst CM, Kasteleyn MJ, van Noort EM, Rutten-van Molken MP, Braunstahl GJ, Chavannes NH, et al. The impact of the involvement of a healthcare professional on the usage of an eHealth platform: a retrospective observational COPD study. Respir Res 2021 Mar 21;22(1):88 [FREE Full text] [doi: 10.1186/s12931-021-01685-0] [Medline: 33743686]
135. Vatnøy TK, Thygesen E, Dale B. Telemedicine to support coping resources in home-living patients diagnosed with chronic obstructive pulmonary disease: patients' experiences. J Telemed Telecare 2016 Jul 08;23(1):126-132. [doi: 10.1177/1357633x15626854]
136. Vorrink S, Huisman C, Kort H, Troosters T, Lammers JW. Perceptions of patients with chronic obstructive pulmonary disease and their physiotherapists regarding the use of an eHealth intervention. JMIR Hum Factors 2017 Sep 19;4(3):e20 [FREE Full text] [doi: 10.2196/humanfactors.7196] [Medline: 28928110]
137. Walters J, Cameron-Tucker H, Wills K, Schüz N, Scott J, Robinson A, et al. Effects of telephone health mentoring in community-recruited chronic obstructive pulmonary disease on self-management capacity, quality of life and psychological morbidity: a randomised controlled trial. BMJ Open 2013 Sep 06;3(9):e003097 [FREE Full text] [doi: 10.1136/bmjopen-2013-003097] [Medline: 24014482]
138. Zanaboni P, Lien LA, Hjalmarsen A, Wootton R. Long-term telerehabilitation of COPD patients in their homes: interim results from a pilot study in Northern Norway. J Telemed Telecare 2013 Oct;19(7):425-429. [doi: 10.1177/1357633X13506514] [Medline: 24218358]
139. Zanaboni P, Hoaas H, Aarøen Lien L, Hjalmarsen A, Wootton R. Long-term exercise maintenance in COPD via telerehabilitation: a two-year pilot study. J Telemed Telecare 2016 Jul 09;23(1):74-82. [doi: 10.1177/1357633x15625545]
140. Heimer M, Schmitz S, Teschler M, Schäfer H, Douma ER, Habibovic M, et al. eHealth for maintenance cardiovascular rehabilitation: a systematic review and meta-analysis. Eur J Prev Cardiol 2023 Oct 26;30(15):1634-1651. [doi: 10.1093/eurjpc/zwad145] [Medline: 37154363]
141. Hardeman W, Michie S, Fanshawe T, Prevost AT, Mcloughlin K, Kinmonth AL. Fidelity of delivery of a physical activity intervention: predictors and consequences. Psychol Health 2008 Dec 14;23(1):11-24. [doi: 10.1080/08870440701615948] [Medline: 25159904]
142. Lorencatto F, West R, Stavri Z, Michie S. How well is intervention content described in published reports of smoking cessation interventions? Nicotine Tob Res 2013 Jul;15(7):1273-1282. [doi: 10.1093/ntr/nts266] [Medline: 23262584]
143. de Bruin M, Black N, Javornik N, Viechtbauer W, Eisma MC, Hartman-Boyce J, et al. Underreporting of the active content of behavioural interventions: a systematic review and meta-analysis of randomised trials of smoking cessation interventions. Health Psychol Rev 2021 Jun 13;15(2):195-213 [FREE Full text] [doi: 10.1080/17437199.2019.1709098] [Medline: 31906781]
144. Kaptain RJ, Helle T, Kottorp A, Patomella AH. Juggling the management of everyday life activities in persons living with chronic obstructive pulmonary disease. Disabil Rehabil 2022 Jul 20;44(14):3410-3421 [FREE Full text] [doi: 10.1080/09638288.2020.1862314] [Medline: 33470850]
145. Seppen BF, den Boer P, Wiegel J, Ter Wee MM, van der Leeden M, de Vries R, et al. Asynchronous mHealth interventions in rheumatoid arthritis: systematic scoping review. JMIR Mhealth Uhealth 2020 Nov 05;8(11):e19260 [FREE Full text] [doi: 10.2196/19260] [Medline: 33151161]
146. Kip H, van Gemert-Pijnen L. Holistic development of eHealth technology. In: Kip H, Beerlage-de Jong N, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London, UK: Routledge; 2018.

XSL•FO
RenderX

Abbreviations

BCT: behavior change technique
COPD: chronic obstructive pulmonary disease
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
TRL: technology readiness level

Edited by N Cahill; submitted 22.02.24; peer-reviewed by M Vervloet, A Sanchez-Vazquez; comments to author 06.06.24; revised version received 03.07.24; accepted 19.12.24; published 10.03.25.

Please cite as:
te Braake E, Vaseur R, Grünloh C, Tabak M
The State of the Art of eHealth Self-Management Interventions for People With Chronic Obstructive Pulmonary Disease: Scoping Review
J Med Internet Res 2025;27:e57649
URL: https://www.jmir.org/2025/1/e57649
doi:10.2196/57649
PMID:

©Eline te Braake, Roswita Vaseur, Christiane Grünloh, Monique Tabak. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Intelligent Robot Interventions for People With Dementia: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Wenqi Fan[1], MSN; Rui Zhao[1], MSN; Xiaoxia Liu[1], MSN; Lina Ge[1], MSN

Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China

Corresponding Author:
Lina Ge, MSN
Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University
Heping District/Sanhao Street, 36th
Shenyang, 110004
China
Phone: 86 18940251669
Email: geln@sj-hospital.org

Abstract

Background: The application of intelligent robots in therapy is becoming more and more important for people with dementia. More extensive research is still needed to evaluate its impact on behavioral and psychological dementia symptoms, as well as quality of life in different care settings.

Objective: The purpose of this research is to methodically assess how well intelligence robot interventions work for patients with dementia.

Methods: In accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, a comprehensive search was conducted on PubMed, CINAHL, the Cochrane Library, Embase, and Web of Science from the time of their founding to February 2024, to identify relevant randomized controlled trials on the use of intelligent robots in people with dementia. Two authors (WF and RZ) independently applied the Cochrane Collaboration bias assessment tool to assess the included studies' quality. The intervention effect of intelligent robots on patients with dementia was summarized using a fixed-effect model or a random-effects model with Stata software (version 16.0; StataCorp). Subgroup analysis was performed according to the intelligent robot type and the intervention duration. Publication bias was tested using funnel plots, Egger tests, and the trim-and-fill method.

Results: In total, 15 studies were finally included for systematic review, encompassing 705 participants, of which 12 studies were subjected to meta-analysis. The meta-analysis found that compared with the control group, intelligent robot intervention significantly reduced the levels of agitation (standardized mean difference –0.36, 95% CI –0.56 to –0.17; $P<.001$) and anxiety (weighted mean difference –1.93, 95% CI –3.13 to –0.72; $P=.002$) in patients with dementia. However, the intervention of intelligent robots had no significant effect on the following (all $P>.05$): cognitive function, neuropsychiatric symptoms, depression, quality of life, step count during the day, and the hours of lying down during the night of patients with dementia. Subgroup analysis revealed that the improvement of depression was related to the duration of the intervention (≤12 vs 12 weeks: 0.08, 95% CI –0.20 to 0.37 vs –0.68, 95% CI –1.00 to –0.37; $P=.26$) and was independent of the type of intelligent robots (animal robots vs humanoid robots: –0.30, 95% CI –0.75 to 0.15 vs 0.07, 95% CI –0.21 to –0.34; $P=.26$).

Conclusions: This study shows that intelligent robot intervention can help improve the agitation and anxiety levels of people with dementia. The intervention may be more effective the longer it is implemented. The appearance of the intelligent robot has no effect on the intervention effect. Further research is needed to help collect physiological data, such as physical activity in people with dementia; explore the impact of other intelligent robot design features on the intervention effect; and provide a reference for improving intelligent robots and intervention programs.

Trial Registration: PROSPERO CRD42024523007; https://tinyurl.com/mwscn985

(J Med Internet Res 2025;27:e59892) doi:10.2196/59892

KEYWORDS

intelligent robot; artificial intelligence; dementia; agitation; anxiety; meta-analysis

XSL•FO
RenderX

Introduction

Background

As the global aging situation becomes more serious, dementia has become a highly prevalent disease in the older population [1]. The characteristic of dementia is the degradation of memory, cognition, behavior, and daily activity abilities [2]. Symptoms such as agitation and depression increase the risk of secondary problems like fractures and falls, which can seriously impair patients' quality of life. In addition, it puts more strain on those who provide care, raising the expense of nursing and medical care. According to the August 2020 update of the "Guidelines for Dementia Prevention, Intervention, and Care" by the Lancet Committee, psychotropic medications typically have little effect on neurological or mental symptoms and can have substantial side effects [3]. As a first line of treatment for behavioral and psychological symptoms of dementia, nonpharmacological therapies are advised due to the drawbacks and safety concerns associated with medication therapy [4].

Multiple studies have shown that music therapy and multisensory stimulation intervention have significant effects in improving behavioral and psychological symptoms of dementia [5,6]. The senses enable patients to interact effectively with the environment and communicate better with others. The effects of pet therapy applied in the geriatric population are mainly in the form of improved social interaction, improved emotional state, enhanced cognitive level, and improved problems. Studies on patients with dementia have demonstrated the beneficial effects of pet therapy [7]. Based on the above evidence, using intelligent robots for intervention in patients with dementia may have the following advantages:

1. Multiplicity: Intelligent robots can comprehensively provide tactile, visual, and auditory multisensory stimulation interventions, music interventions, and pet interventions; they can also assist in social activities, accompany patients with dementia, provide social-emotional support, etc.
2. Interactivity: Intelligent robots are capable of interacting with patients, and the interaction modes mainly include touch interaction; voice interaction; and somatosensory interaction, which is a two-way active form rather than one-way passive form and is conducive to improving patient participation and enriching the emotional experience.
3. Telemedicine: The therapist can use remote control software installed in the device to control the activation and progress of the smart robot scripts, enabling telemedicine.
4. Less restriction: The use of intelligent robots to accompany patients is less restricted by place and population compared to animal interventions; for example, patients with hair allergies are not suitable to receive canine companionship but can interact with pet robots.

In recent years, intelligence robot interventions have caught the interest of scientists studying dementia. A variety of intelligent robots, including pet robots, companion robots, social assistance robots, and humanoid robots, are being used for neuropsychiatric symptoms, cognitive function, and pain in patients with dementia [8,9]. The results of intelligence robot intervention vary, nevertheless. For instance, a mixed methods study revealed that following a 6-week intervention, the robotic companion dog or cat group's feelings of melancholy and loneliness considerably dropped as compared to the control group [10]. The study's qualitative findings revealed that the robotic companion pet experience was viewed positively by the participants, their families, and professional caregivers, who felt that the robot enhanced communication and offered companionship [10]. Conversely, another study indicated that a social robot intervention did not significantly improve positive affective states [11]. There is currently conflict over the efficacy of interventions on agitation and quality of life, as well as a lack of data for objective indicators like physical activity and sleep duration, despite the publication of a few meta-analyses of intelligence robot intervention [8,9,11-13]. Subgroup analyses of intervention settings, such as the kind and purpose of intelligent robots and the length, the location, and the format of the intervention, have not been done in previous studies. New trials have been published in the last two years; thus, an updated review of this material is still necessary.

Objectives

Given this, this study aimed to examine how intelligence robots affect both subjective and objective markers in patients with dementia by a thorough assessment of the literature and meta-analysis. Furthermore, the type of intelligent robots, as well as subgroup analysis based on intervention times, are discussed. Overall, the findings will offer support for further relevant studies and therapeutic applications.

Methods

Overview

We conducted this systematic review and meta-analysis by following the PRISMA (Preferred Reporting Items in the Systematic Review and Meta-Analyses) guidelines (Multimedia Appendix 1).

Literature Search Strategy

WF and RZ conducted a comprehensive search in February 2024 across the following databases: Web of Science, CINAHL, PubMed, Embase, Cochrane Library, and CINAHL. We found additional relevant publications by searching the bibliographies of the included papers and previous relevant systematic reviews. Details about the customized search approach for each database can be found in Tables S1-S6 in Multimedia Appendix 2. The following search terms were used: "dementia," "Alzheimer's disease," "AD," "robotics," "robotic," "robot," and "robot-assisted."

Criteria for Inclusion and Exclusion

The included studies satisfied the following criteria: (1) adult patients were diagnosed with dementia; (2) the study was a randomized controlled trial (RCT); (3) the intervention group received intelligence robot intervention, with no restrictions on the kinds of intelligent robots; (4) the control group was given standard treatment; and (5) the study was published in English. The exclusion criteria are as follows: (1) absence of data for the network meta-analysis; (2) duplicate data; and (3) qualitative

XSL•FO
RenderX

research, books, review studies, conference abstracts, or study protocols.

Study Selection and Data Extraction

EndNote 20.4 (Clarivate Analytics) was used to eliminate duplicates before deciding if the searched records were eligible. Initial screening of literature by two independent authors (WF and RZ). WF and RZ discussed and settled any differences between them until they came to an agreement. A third author (XL) was consulted if a consensus could not be achieved. The first author, the country, the year of publication, the type of study, the analyzed sample size, the average age of the participants, the specifics of the intervention (methods, frequency, and duration), and assessment time points were all extracted using a sheet by two independent authors, WF and XL.

Risk of Bias Assessment

Two authors (WF and RZ) independently applied the Cochrane Collaboration risk of bias tool [14] to assess the quality of the included studies.

Statistical Analysis

The meta-analysis was conducted using Stata software (version 16.0; StataCorp). To address potential bias caused by differences in indicator levels between the intervention and control groups at baseline, we calculated the mean difference and SD between postintervention and baseline values to reflect within-group changes.

The effect size, which quantifies the difference in the magnitude of change between groups, was expressed as either the standardized mean difference (SMD) or the weighted mean difference (WMD). The WMD was used when studies measured outcomes on the same scale; otherwise, the SMD was applied.

Heterogeneity was considered significant when the *P* value of the Cochran *Q* test was less than 0.10 and the Higgins I^2 statistic was greater than 50%, and a random effects model was used to account for heterogeneity between studies; otherwise, a fixed effects model was used [15]. A forest plot and a funnel plot were generated for the analysis. The calculation of Egger regression tests was done to assess publication bias.

Forest plots were generated, and sensitivity analyses were performed to further assess the stability of the study results. The calculation of Egger regression tests, visualization of the funnel plot, and trim-and-fill method were used to evaluate publication bias [16].

Subgroup analyses were conducted to explore the influence of study design factors (intervention duration) and intervention conditions (types of intelligent robots) on the effect of the intervention and to assess possible sources of heterogeneity.

Results

Literature Screening

The PRISMA flowchart of study identification and screening is outlined in Figure 1. A total of 1319 records were extracted from the electronic database search, and 3 records were identified through hand searching of reference. Following the elimination of duplicates, we went over the abstracts and titles of 797 (60.29%) publications to weed out any that were not relevant. In the end, after verifying compliance with the inclusion criteria and exclusion criteria, a total of 15 (1.13%) studies were included in the systematic review, and 12 (0.91%) of these studies were included in the meta-analysis.

Figure 1. PRISMA flow diagram of the study selection process. PRISMA: Preferred Reporting Items in the Systematic Review and Meta-Analyses; RCT: randomized controlled trial.

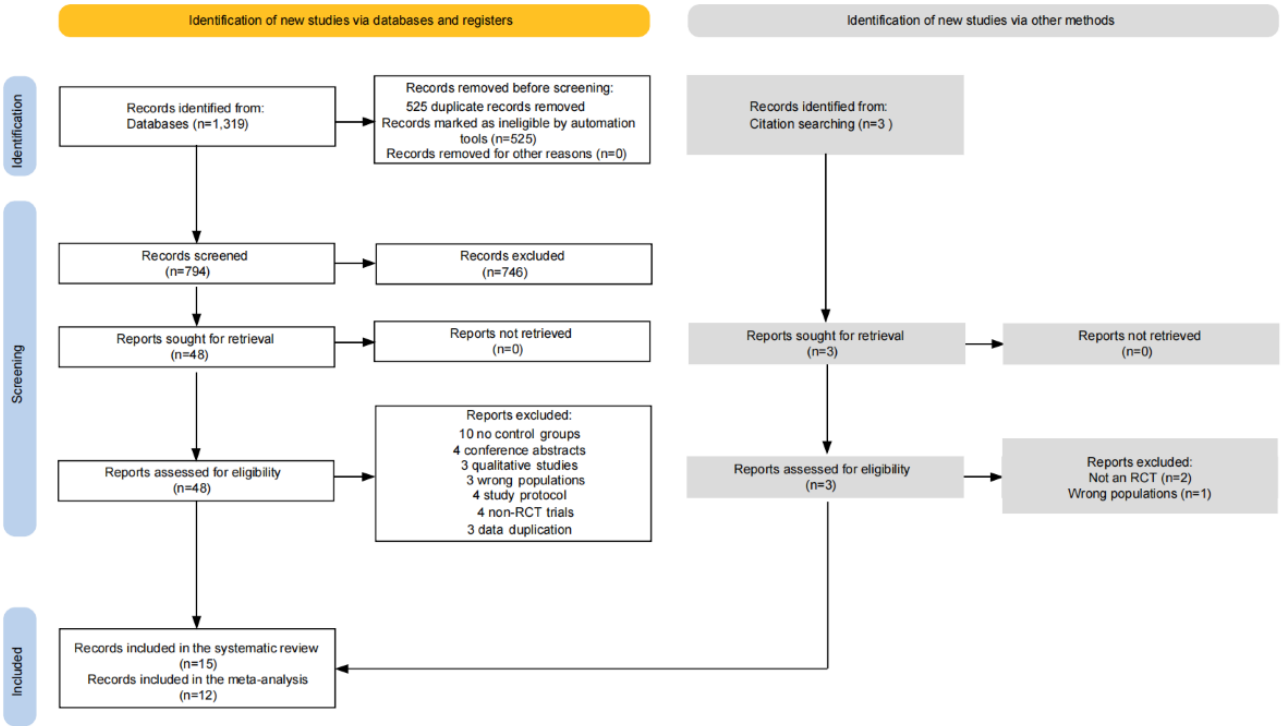

XSL•FO
RenderX

Study Characteristics

The characteristics of the included studies in the systematic review are shown in Table 1. A total of 15 studies were conducted in 8 countries, including Australia (5/15, 33%) [17-21], Norway (3/15, 20%) [22-24], China (2/15, 13%) [25,26], Spain (1/15, 7%) [27], New Zealand (1/15, 7%) [28], Japan (1/15, 7%) [29], the United Kingdom (1/15, 7%) [30], and the United States (1/15, 7%) [31]. All studies were published after 2015, and the number of participants in each study ranged from 22 to 175, with intervention periods ranging from 6 weeks to 3 months. Participants were mainly female, with the proportion ranging from 64% to 88%, and the average age of participants was reported to be from 83.4 to 89.0 years in 13 (87%) [17-26,29-31] studies. The venues where the interventions were implemented included long-term care facilities, nursing homes, and other venues. Only 2 (13%) [27,29] studies were single-center studies, while the remaining 13 (87%) [17-26,28,30,31] studies were multicenter studies.

XSL•FO
RenderX

Table 1. Characteristics of included studies (n=15).

Study (year), country	Study type	Analyzed sample size (IG[a] and CG[b])	Women (%)	Age (years)	Intervention methods	Frequency or duration of intervention	Follow-up duration	Assessment time points
Sugiyama and Nakamura [29] (2022), Japan	RCT[c]	• 14 and 8	• 86.3%	• Mean 89.0 (SD 4.3)	Playing the shiritori game with a dialogue interactive robot	• 10 minutes/session • 3 sessions/week • Total: 33 sessions over 11 weeks	12 weeks	• T0 • T1: 11 weeks
Bradwell et al [30] (2022), United Kingdom	Stratified cluster RCT	• 26 and 37	• 77.8%	• Mean 87.2 (SD 7.4)	Interacting with pet robots	• Total: 516.3 hours over 4 months	4 months	• T0 • T1: 4 months
Pu et al [17] (2021), Australia; Pu et al [18] (2020), Australia	RCT	• 21 and 20 • 21 and 22	• 70.7% • 69.8%	• Mean 86.0 (SD 7.5) • Mean 86.0 (SD 7.4)	Social robot intervention	• 30 minutes/session • 5 sessions/week • Total: 30 sessions over 6 weeks	6 weeks	• T0 • T1: 1 week • T2: 6 weeks • T3: After
Jøranson et al [22] (2021), Norway; Jøranson et al [23] (2016), Norway; Jøranson et al [24] (2015), Norway	Cluster RCT	• 27 and 26	• 66.7%	• Mean 84.6 (SD 6.9)	Intervention for robot-assisted activity	• 30 minutes/session • 2 sessions/week • Total: 14 sessions over 12 weeks	12 weeks	• T0 • T1: 12 weeks
Ke et al [25] (2020), China; Chen et al [26] (2020), China	RCT with an ABAB[d] withdrawal design	• 52 and 51	• 79.6%	• Mean 87.2 (SD 7.4)	Social robot intervention	• N/Ae over 16 weeks	32 weeks	• T0 • T1: 8 weeks • T2: 16 weeks • T3: 24 weeks • T4: 32 weeks
Moyle et al [19] (2018), Australia; Mervin et al [20] (2018), Australia; Moyle et al [21] (2017), Australia	Cluster RCT	• 67 and 53	• 73.3%	• Mean 84.4 (SD 8.0)	Paro robot intervention	• 15 minutes/session • 3 sessions/week • Total: 30 sessions over 10 weeks	15 weeks	• T0 • T1: 5 weeks • T2: 10 weeks • T3: 15 weeks
Petersen et al [31] (2017), United States	RCT	• 35 and 26	• 77%	• Mean 83.4 (SD 5.9)	Paro robot intervention	• 20 minutes/session • 3 sessions/week • Total: 36 sessions over 12 weeks	3 months	• T0 • T1: 3 months
Liang et al [28] (2017), New Zealand	RCT	• 13 and 11	• 64%	• Range 67-98	Paro robot intervention	• 30 minutes/session 2-3 sessions/week • Total: 12-18 sessions over 6 weeks	12 weeks	• T0 • T1: 6 weeks • T2: 12 weeks

XSL•FO
RenderX

Study (year), country	Study type	Analyzed sample size (IG[a] and CG[b])	Women (%)	Age (years)	Intervention methods	Frequency or duration of intervention	Follow-up duration	Assessment time points
Soler et al [27] (2015), Spain	RCT	• Phase 1: IG-NAO: 30 ; CG-PARO: 33; CG: 38 • Phase 2: IG-PARO: 42; CG: 32	• Phase 1: 88% • Phase 2: 90%	• Phase 1: mean 84.7 (range 68-87) • Phase 2: mean 84.7 (range 69-87)	Paro/Nao robot intervention	• 30-40 minutes/session • 2 sessions/week • Total: 24 sessions over 12 weeks	12 weeks	• T0 • T1: 12 weeks

[a]IG: intervention group.

[b]CG: control group.

[c]RCT: randomized controlled trial.

[d]ABAB design, also known as the reversal design, has four phases: a baseline measurement, a measurement under test conditions, a return to the baseline measurement, and a remeasurement under test conditions.

[e]N/A: not applicable.

Risk of Bias Assessment

Each included RCTs' bias risk is displayed in Table S7 in Multimedia Appendix 2. A computer-generated random list was used to generate the randomization sequence in 6 of the RCTs [17,19,23,25,27,28,30], while 1 RCT [31] included coin tossing. The risk of selection bias was deemed questionable since the randomization procedure was not reported in detail in 2 other RCTs [28,29]. However, because it was challenging to blind staff members and participants in the intelligent robot interventions, only 1 RCT [19] was deemed to have a low risk of performance bias.

Analysis of Overall Effects

Cognitive Function

Cognitive function was measured in 5 studies [26-29,31], and 1 RCT [27] reported humanoid and pet robot intervention with different scales. The combined data showed that there was no discernible variation in cognitive function (SMD 0.09, 95% CI –0.09 to 0.26; *P*=.46). Furthermore, the results showed no heterogeneity (I^2=0%). In Figure 2 [26-29,31], the forest diagram is displayed.

Figure 2. Forest plot of cognitive function. M: Minimental state examination scale; N: Nao robot; P: Paro robot (Pet robot); S: Severe minimental state examination scale. SMD: standardized mean difference.

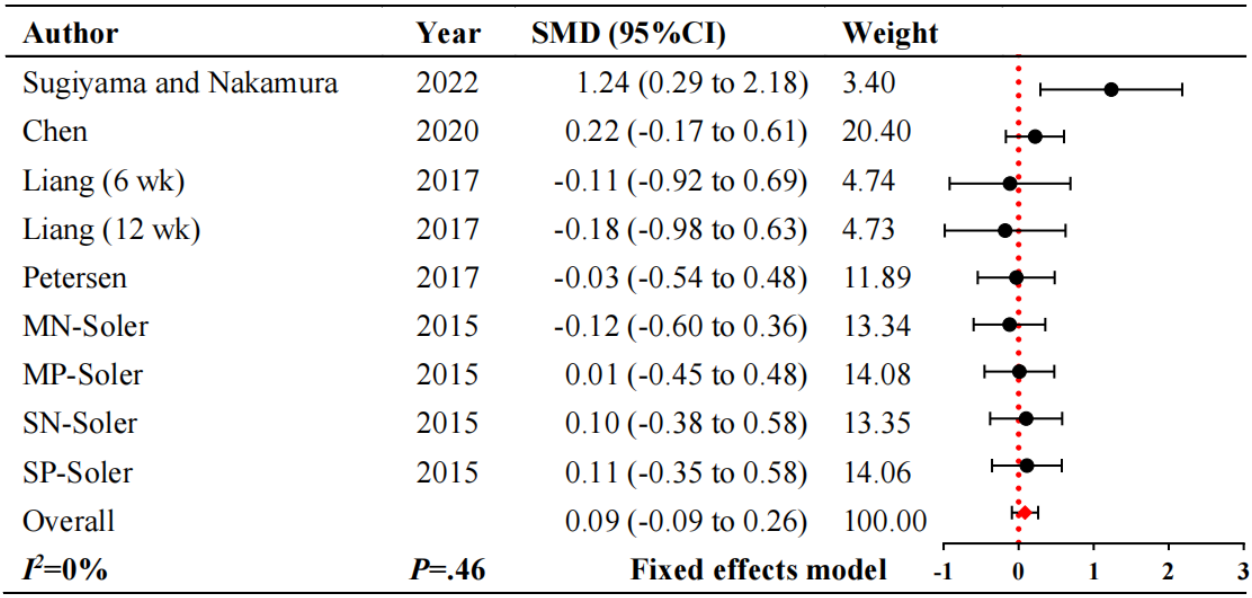

XSL•FO
RenderX

Neuropsychiatric Symptoms

Regarding neuropsychiatric symptoms, 5 studies [26-30] involving 552 participants reported changes in neuropsychiatric symptoms. Figure 3 [26-30] illustrates that intelligent robot interventions cannot significantly improve neuropsychiatric symptoms in patients with dementia compared to usual care (SMD –0.09, 95% CI –0.35 to 0.16; *P*=.46). These studies showed considerable heterogeneity (I^2=50.2%; *P*=.07).

Figure 3. Forest plot of neuropsychiatric symptoms. N: Nao robot (humanoid robot); P: Paro robot (Pet robot); SMD: standardized mean difference.

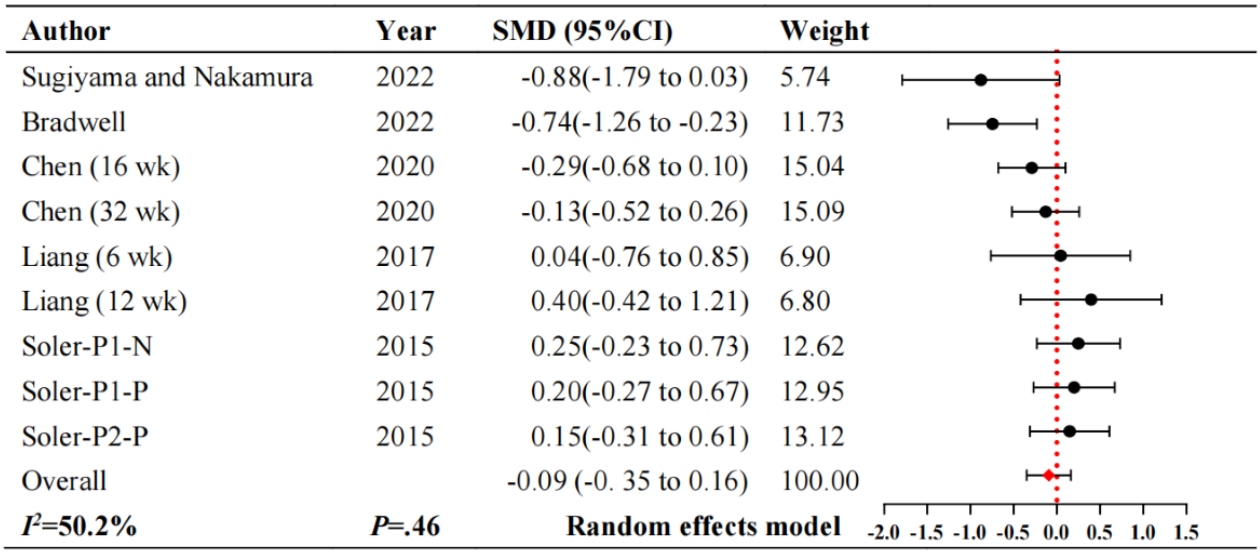

Agitation

Agitation was measured in 4 studies [18,20,24,28]. Figure 4 [18,20,24,28] displays a substantial decline in agitation levels (SMD –0.36, 95% CI –0.56 to –0.17; *P*<.001). These studies showed no heterogeneity (I^2=0).

Figure 4. Forest plot of agitation. SMD: standardized mean difference.

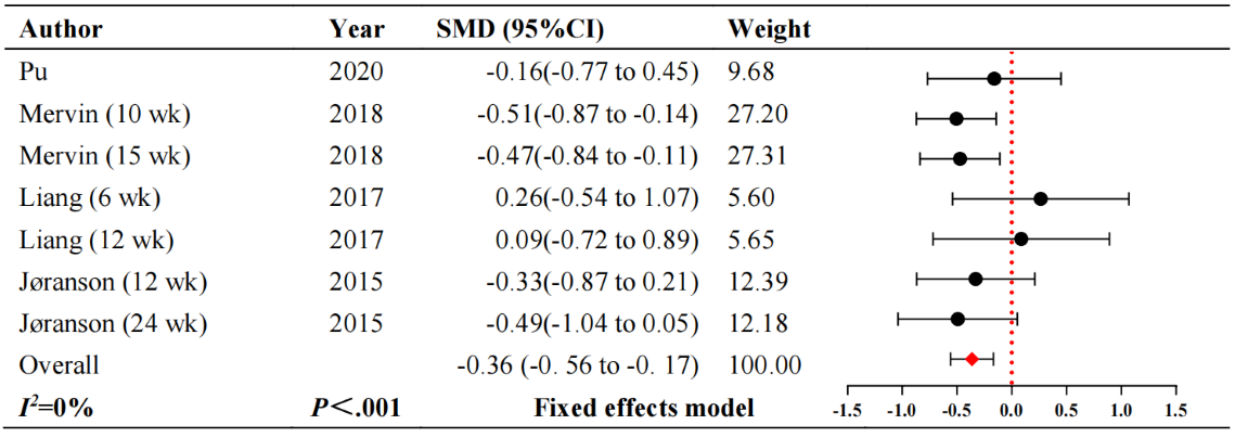

Anxiety

For anxiety, 2 studies [18,31] involving 102 participants reported changes in anxiety. Figure 5 [18,31] illustrates that intelligent robot interventions significantly reduce anxiety scores in patients with dementia compared to usual care (WMD –1.93, 95% CI –3.13 to –0.72; *P*=.002). These studies showed no heterogeneity (I^2=0).

XSL•FO
RenderX

Figure 5. Forest plot of anxiety. WMD: weighted mean difference.

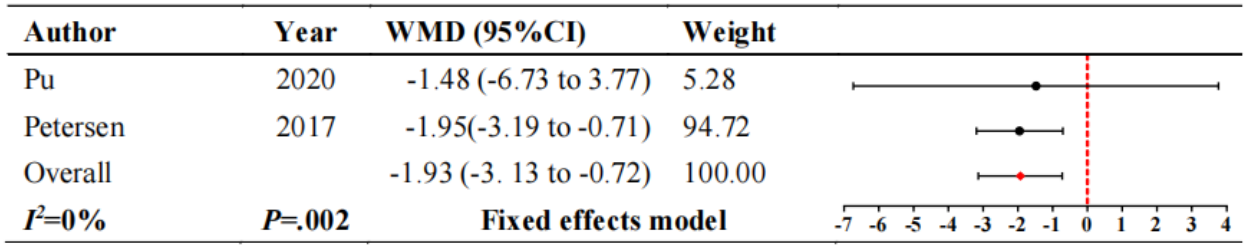

Depression

Regarding depression, 5 studies [18,24,26,28,31] involved 222 participants in the control group and 240 participants in the robot intervention group. Figure 6 [18,24,26,28,31] shows that the intelligent robot interventions cannot relieve depression symptoms in patients with dementia compared to usual care (SMD –0.20, 95% CI –0.54 to 0.15; *P*=.26). These studies showed considerable heterogeneity (I^2=67.8%; *P*=.003).

Figure 6. Forest plot of depression. SMD: standardized mean difference.

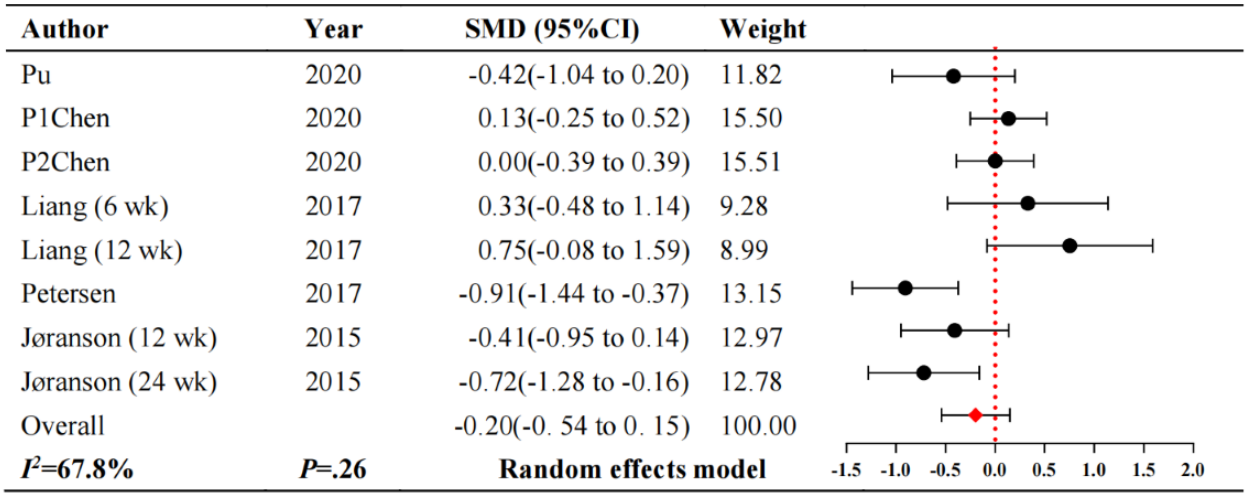

Quality of Life

For quality of life, 3 studies [23,26,27] were measured. The quality of life scores displayed in Figure 7 [23,26,27] shows that the intelligent robot interventions cannot improve the quality of life in patients with dementia compared to usual care (SMD 0.05, 95% CI –0.30 to 0.41; *P*=.77). These studies showed considerable heterogeneity (I^2=66.1%; *P*=.02).

Figure 7. Forest plot of quality of life. SMD: standardized mean difference.

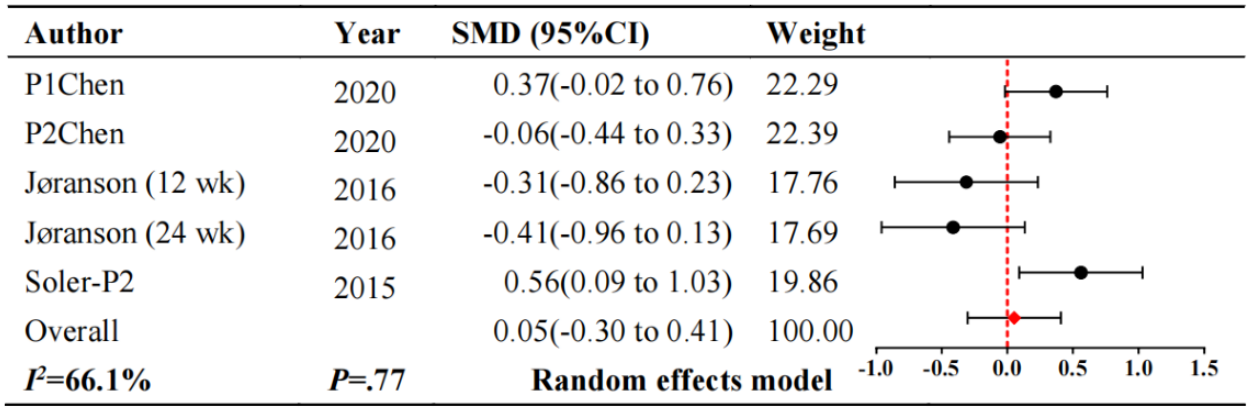

Data of SenseWear Armband

Sleep and activity data were collected using the SenseWear armband from 2 RCTs [17,19]. The combined data revealed a significant decrease in the amount of time spent lying down during the day (WMD –0.48, 95%CI –0.90 to –0.07; *P*=.02; I^2=0) and reported nonsignificant impacts on length of awake during the daytime or nighttime, length of time lying down during the nighttime, and step count or length of physical activity during the daytime or nighttime (all *P*>.05; Table 2).

XSL•FO
RenderX

Table 2. Meta-analysis of sleep and activity data of SenseWear armband.

	WMD[a]	95% CI	*P* value	I^2 (%)
Daytime				
Lying down (hours)	–0.48	–0.90 to –0.07	.02	0.00
Awake (hours)	0.22	–0.12 to 0.57	.20	2.10
Step count	–31.71	–117.80 to 54.38	.47	43.80
Physical activity (hours)[b]	–0.06	–0.36 to 0.24	.69	0.00
Nighttime				
Lying down (hours)	–0.40	–0.92 to 0.11	.12	0.00
Awake (hours)	0.45	–0.07 to 0.97	.09	0.00
Step count	–10.54	–45.48 to 24.40	.55	41.90
Physical activity (hours)[a]	–0.10	–0.23 to 0.40	.17	0.00

[a]WMD: weighted mean difference.

[b]Time spent in at least light physical activity (>1.5 metabolic equivalent of task).

Subgroup Analysis

Different Durations of Interventions

The Effects on Neuropsychiatric Symptoms

Both short-duration (<12 weeks; SMD –0.18, 95% CI –0.46 to 0.10; *P*=.21; I^2=18%) and long-duration (≥12 weeks; SMD –0.03, 95% CI –0.47 to 0.42; *P*=.91; I^2=70.3%) intelligent robot interventions were not found to enhance neuropsychiatric symptoms significantly (Figure S1 in Multimedia Appendix 2).

The Effects on Agitation

The subgroup analysis results showed that different intervention durations significantly improved agitation in people with dementia. Still, the longer the intervention continued, the better the effect (12 weeks; SMD –0.41, 95% CI –0.80 to –0.03; *P*=.04; I^2=0%; Figure S2 in Multimedia Appendix 2).

The Effects on Depression

A significant effect favoring the more extended intervention on relieving depression (12 weeks; SMD –0.68, 95% CI –1.00 to –0.37; *P*<.001; I^2=0; Figure S3 in Multimedia Appendix 2).

The Effects on Quality of Life

There was no significant difference between different duration intelligent robot interventions and usual care in improving quality (Figure S4 in Multimedia Appendix 2).

Different Types of Intelligent Robots

There was no significant difference between different types of intelligent robot interventions and usual care in improving cognitive function, neuropsychiatric symptoms, depression, and quality of life (Figures S5-S8 in Multimedia Appendix 2).

Sensitivity Analysis

The research findings of the five outcome measures, namely cognitive function, neuropsychiatric symptoms, agitation, depression, and quality of life, indicate that the point estimates of the combined effect sizes were obtained after removing 1 study at a time and analyzing the remaining studies. The results showed that the exclusion of a particular study does not significantly alter the overall results, indicating that an evaluation of the 5 outcomes had considerable stability (Figures S9-S13 in Multimedia Appendix 2).

Publication Bias

Funnel plots and Egger tests were performed to evaluate publication bias of 5 outcome measures, namely cognitive function, neuropsychiatric symptoms, agitation, depression, and quality of life. The results showed no publication bias in cognitive function, neuropsychiatric symptoms, depression, and quality of life (Figures S14-S17 in Multimedia Appendix 2). For agitation, the results of the Egger test showed possible publication bias. There was no significant difference between the effect size estimates obtained using the trim-and-fill method and those obtained by meta-analysis, indicating the results were reliable (Figures S18 and S19 in Multimedia Appendix 2).

Description of Outcomes Not Suitable for Meta-Analysis

A total of 3 studies [21-22,25] were left out of the meta-analysis due to the inability to combine the inclusion of several datasets. Jøranson et al [22] looked into how the intelligent robot Paro affected the sleep habits of patients with dementia living in a nursing home. This study used wrist actigraphy to assess objectively sleep-wake patterns and showed that an intelligent robotic intervention significantly increased the percentage of sleep efficiency and reduced the frequency of nocturnal awakenings in patients with Alzheimer disease. Ke et al [25] assessed the effects of the humanoid social robot (Kabochan) on technology acceptance among older individuals with dementia; they found that being exposed to intelligent robots may enhance their perceptions of the technology's utility and attitudes toward it. Moyle et al [21] showed that participants in the PARO group were more verbally and visually engaged than participants in plush toy.

XSL•FO
RenderX

Discussion

Principal Findings

This study indicates that intelligence robot interventions effectively influenced agitation, anxiety, and length of time lying down during the daytime. The current meta-analysis did discover, however, that intelligence robot interventions had no discernible effects on depression, quality of life, cognitive function, neuropsychiatric symptoms, sleep duration, wakefulness during the day or night, step count, or frequency of physical activity.

In this study, intelligent robot intervention significantly reduces levels of anxiety in patients with dementia, which is consistent with meta-analytic reviews of Saragih et al [8]. Anxiety is highly prevalent across dementia stages, with an overall pooled prevalence of 39% [32]. In recent years, various clinical applications of intelligent robots have been used to provide high-quality emotional support and companionship [33]. Intelligent robots can provide emotional support and companionship to patients with dementia, easing anxiety by interacting with them and sharing stories, music, or simple games. In future studies, the effects of different design elements in intelligent robot interventions on the anxiety levels of patients with dementia can be further explored to understand the preferences of patients with dementia for the appearance characteristics and interaction contents of intelligent robots, so as to improve the references for the design of intelligent robots and clinical application programs, and to give full play to their effectiveness in alleviating the anxiety levels of patients with dementia.

Agitation improved in the intelligent robot intervention group. One study has shown that providing patients with dementia with sensory stimulation, mainly through sight, sound, touch, taste, and smell, can help to improve agitation [34]. Intelligent robots integrate the application of visual, tactile, and auditory multiple sensing technologies, such as the social robot (Paro) covered with artificial fur, capable of moving and emitting sounds to achieve multisensory stimulation, which can put patients with dementia in a multisensory enriched environment and alleviate their agitated state. It has been shown that agitation prolongs the hospital stay of people with dementia and imposes physical, psychological, and financial burdens on caregivers [35]. More research is needed in the future to evaluate whether these interventions also benefit shorter hospital stays, better caregiver psychological health, and lower costs.

The intelligent robot intervention was not significant for overall depression levels in patients with dementia, which is consistent with the findings of Abbott et al [36]. Subgroup analyses based on the length of the intervention were carried out, and the findings indicated that when the intervention lasted for 12 weeks, the intelligent robot intervention helped patients with dementia experience lower levels of depression; however, no such beneficial effect was observed when the intervention lasted for shorter than 12 weeks. Depression in people with dementia often involves deeper emotional issues, and the provision of emotional support and cognitive stimulation by intelligent robots may take longer to produce significant improvements [12]. Subsequent investigations ought to delve deeper into the impact of intervention duration on the efficacy of intelligent robot interventions. The humanoid and animal robots did not differ from one another in subgroup analyses depending on the type of intelligent robot, and neither had a significant impact on the depression levels of patients with dementia. The usefulness of humanoid and animal robot treatments has not been the subject of as many studies, and this study's sample size was limited. In order to provide some guidance for the appearance design and function setting of intelligent robots, the sample size should be increased in subsequent research, and various types of intelligent robots should be thoroughly examined.

Cognitive function and neuropsychiatric symptoms in patients with dementia did not significantly improve with intelligent robot interventions. The cognitive deterioration involved in dementia is usually associated with structural changes in the brain [37], and although intelligent robots can provide social interaction and behavioral stimulation, these may not be sufficient to reverse or significantly improve the cognitive impairment caused by neurodegenerative diseases. Intelligent robots lack the human caregiver's capacity for emotional depth, empathy, and complex decision processing, and their role in neuropsychiatric symptoms in patients with dementia may be limited.

The relevant data for this study came from wearable devices, and the included studies reported difficulties in wearing wearable devices in patients with dementia. The results did not suggest a significant positive effect on sleep and physical activity. Given that individuals with dementia were reticent to wear armbands, particularly at night [17], this conclusion should be interpreted cautiously. To increase patient happiness and participation, more appropriate techniques for tracking sleep and physical activity in patients with dementia should be investigated in the future.

In the group receiving intervention, the quality of life did not improve. Studies have shown that one of the main factors influencing the quality of life of patients with dementia is their capacity to carry out activities of daily living [38]. Patients with dementia also tend to be less capable of taking care of themselves, and some even experience difficulties walking. Although intelligent robots can provide a certain degree of life guidance and support, they cannot completely replace human care and assistance. The cognitive abilities, neuropsychiatric symptoms, the stress of being the primary caregiver, and other factors all affect the quality of life of patients with dementia [39].

Subgroup analysis according to the length of intervention showed that intelligent robot intervention for 12 weeks or more had a more positive effect on the depression level and agitation of patients with dementia, and there was no significant difference in neuropsychiatric symptoms and quality of life. In patients with dementia receiving 12 weeks or more of intervention, symptoms such as depression and agitation may change more significantly due to improvements in emotional regulation and behavioral interventions. However, neuropsychiatric symptoms and quality of life are affected by a combination of more complex physical, psychological, and

XSL•FO
RenderX

social factors, and the effect of the intervention may not be significantly different in these dimensions. This suggests that although long-term interventions have a positive impact on mood and behavior, longer or more comprehensive interventions may be needed to improve neurodegenerative lesions and quality of life. A subgroup analysis was performed according to whether the intelligent robot was a humanoid robot or a pet robot. The results showed that there was no significant difference in the improvement of cognitive function, neuropsychiatric symptoms, depression, and quality of life in patients with dementia between the two types of robots. Although humanoid robots and pet robots differ in appearance and behavior, their core role in providing intervention is often similar, that is, to improve the mood and behavior of patients through interaction, emotional companionship, cognitive training, and other means. This may be the reason for the similar intervention effects of the two.

Strengths and Limitations

This study has several practical implications for practice. First, interventions based on intelligent robots can be considered for inclusion in the long-term care of people with dementia. Second, the shape characteristics of intelligent robots may have no effect on the intervention effect on people with dementia, and other design characteristics need to be further explored. Third, when using wearable devices to assess sleep, physical activity, etc, in people with dementia, the wearing comfort and compliance of people with dementia should be considered.

The following limitations of this study need to be taken into account when interpreting the results. First, the sample size included in this study was small, and further analysis of the impact of factors such as the age, gender, and geographical location of participants on the effectiveness of the intervention was lacking. Second, only original studies published in English were included, and there is a possibility that relevant studies published in other languages were missed. In addition, not all studies were assessed by professional blinders, which may introduce bias into the study.

Despite these limitations, this review aims to comprehensively evaluate the impact of intelligent robot interventions on people with dementia. Seven outcome indicators, including objective indicators of sleep and activity, were meta-analyzed. The impact of the intervention duration and robot type on smart robot interventions was further explored to provide a reference for the future optimization of smart robot development and intervention program design.

Conclusions

This meta-analysis comprehensively assessed the impact of intelligent robot interventions on people with dementia and showed that intelligent robot interventions can reduce agitation and anxiety, with a limited role in improving depression, anxiety, cognitive functioning, neuropsychiatric symptoms, and quality of life, and that sleep- and physical activity-related outcomes need to be analyzed with caution. Subgroup analyses showed better results for the longer the duration of the intelligent robot intervention. The types of intelligent robots included humanoid robots and animal robots, and subgroup analyses showed no difference in their effectiveness. The use of intelligent robots in nursing is still in its infancy, although a number of nations have implemented intelligent robotic interventions for patients with dementia. These interventions have primarily involved women and have small sample sizes and have been concentrated in high-income nations, which may be related to the epidemiological features of dementia. Subsequent research endeavors ought to enhance the conceptual framework of sentient robots and delve deeper into the ways in which the nature of the sentient robot and the length of the intervention affect its efficacy. When using wearable technology to gather physiological data, researchers should take patient compliance into account.

Acknowledgments

The authors would like to thank all authors of the included studies.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

WF, RZ, and XL contributed to the conceptualization and design of the study. WF and RZ collected the data. WF, RZ, and XL analyzed the data. WF, RZ, and XL drafted the manuscript. WF, RZ, XL, and LG extensively revised the manuscript. All authors read and approved the final manuscript submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[OCX File , 27 KB - jmir_v27i1e59892_app1.ocx]

Multimedia Appendix 2

XSL•FO
RenderX

Supplementary tables and figures.
[DOC File , 216 KB - jmir_v27i1e59892_app2.doc]

References

1. Nandi A, Counts N, Chen S, Seligman B, Tortorice D, Vigo D, et al. Global and regional projections of the economic burden of Alzheimer's disease and related dementias from 2019 to 2050: a value of statistical life approach. EClinicalMedicine 2022 Sep;51:101580-101589 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101580] [Medline: 35898316]
2. Asl AM, Ulate MM, Martin MF, van der Roest H. Methodologies used to study the feasibility, usability, efficacy, and effectiveness of social robots for elderly adults: scoping review. J Med Internet Res 2022 Aug 01;24(8):e37434 [FREE Full text] [doi: 10.2196/37434] [Medline: 35916695]
3. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 08;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
4. Bessey LJ, Walaszek A. Management of behavioral and psychological symptoms of dementia. Curr Psychiatry Rep 2019 Jul 01;21(8):66. [doi: 10.1007/s11920-019-1049-5] [Medline: 31264056]
5. Sánchez A, Maseda A, Marante-Moar MP, de Labra C, Lorenzo-López L, Millán-Calenti JC. Comparing the effects of multisensory stimulation and individualized music sessions on elderly people with severe dementia: a randomized controlled trial. J Alzheimers Dis 2016 Mar 08;52(1):303-315. [doi: 10.3233/JAD-151150] [Medline: 27060958]
6. Dimitriou TD, Tsolaki M. Evaluation of the efficacy of randomized controlled trials of sensory stimulation interventions for sleeping disturbances in patients with dementia: a systematic review. Clin Interv Aging 2017;12:543-548 [FREE Full text] [doi: 10.2147/CIA.S115397] [Medline: 28360513]
7. Zafra-Tanaka JH, Pacheco-Barrios K, Tellez WA, Taype-Rondan A. Effects of dog-assisted therapy in adults with dementia: a systematic review and meta-analysis. BMC Psychiatry 2019 Jan 24;19(1):41-50 [FREE Full text] [doi: 10.1186/s12888-018-2009-z] [Medline: 30678665]
8. Saragih ID, Tonapa SI, Sun TL, Chia-Ju L, Lee BO. Effects of robotic care interventions for dementia care: a systematic review and meta-analysis randomised controlled trials. J Clin Nurs 2021 Nov;30(21-22):3139-3152. [doi: 10.1111/jocn.15856] [Medline: 34041803]
9. Lu LC, Lan SH, Hsieh YP, Lin LY, Lan SJ, Chen JC. Effectiveness of companion robot care for dementia: a systematic review and meta-analysis. Innovation Aging 2021;5(2):igab013 [FREE Full text] [doi: 10.1093/geroni/igab013] [Medline: 34316517]
10. Fogelson DM, Rutledge C, Zimbro KS. The impact of robotic companion pets on depression and loneliness for older adults with dementia during the COVID-19 pandemic. J Holist Nurs 2022 Dec;40(4):397-409. [doi: 10.1177/08980101211064605] [Medline: 34889146]
11. Feng Y, Barakova EI, Yu S, Hu J, Rauterberg GWM. Effects of the level of interactivity of a social robot and the response of the augmented reality display in contextual interactions of people with dementia. Sensors (Basel) 2020 Jul 05;20(13):3771-3782 [FREE Full text] [doi: 10.3390/s20133771] [Medline: 32635640]
12. Hsieh CJ, Li PS, Wang CH, Lin SL, Hsu TC, Tsai CT. Socially assistive robots for people living with dementia in long-term facilities: a systematic review and meta-analysis of randomized controlled trials. Gerontology 2023;69(8):1027-1042 [FREE Full text] [doi: 10.1159/000529849] [Medline: 36871553]
13. Leng M, Liu P, Zhang P, Hu M, Zhou H, Li G, et al. Pet robot intervention for people with dementia: a systematic review and meta-analysis of randomized controlled trials. Psychiatry Res 2019 Jan;271:516-525. [doi: 10.1016/j.psychres.2018.12.032] [Medline: 30553098]
14. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
15. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
16. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
17. Pu L, Moyle W, Jones C, Todorovic M. The effect of a social robot intervention on sleep and motor activity of people living with dementia and chronic pain: a pilot randomized controlled trial. Maturitas 2021 Feb;144:16-22. [doi: 10.1016/j.maturitas.2020.09.003] [Medline: 33358203]
18. Pu L, Moyle W, Jones C, Todorovic M. The effect of using PARO for people living with dementia and chronic pain: a pilot randomized controlled trial. J Am Med Dir Assoc 2020 Aug;21(8):1079-1085. [doi: 10.1016/j.jamda.2020.01.014] [Medline: 32122797]
19. Moyle W, Jones C, Murfield J, Thalib L, Beattie E, Shum D, et al. Effect of a robotic seal on the motor activity and sleep patterns of older people with dementia, as measured by wearable technology: a cluster-randomised controlled trial. Maturitas 2018 Apr;110:10-17 [FREE Full text] [doi: 10.1016/j.maturitas.2018.01.007] [Medline: 29563027]

XSL•FO
RenderX

20. Mervin MC, Moyle W, Jones C, Murfield J, Draper B, Beattie E, et al. The cost-effectiveness of using PARO, a therapeutic robotic seal, to reduce agitation and medication use in dementia: findings from a cluster-randomized controlled trial. J Am Med Dir Assoc 2018 Jul;19(7):619-622.e1. [doi: 10.1016/j.jamda.2017.10.008] [Medline: 29325922]
21. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ERA, Shum DKH, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Dir Assoc 2017 Sep 01;18(9):766-773 [FREE Full text] [doi: 10.1016/j.jamda.2017.03.018] [Medline: 28780395]
22. Jøranson N, Olsen C, Calogiuri G, Ihlebæk C, Pedersen I. Effects on sleep from group activity with a robotic seal for nursing home residents with dementia: a cluster randomized controlled trial. Int Psychogeriatr 2021 Oct;33(10):1045-1056 [FREE Full text] [doi: 10.1017/S1041610220001787] [Medline: 32985396]
23. Jøranson N, Pedersen I, Rokstad AMM, Ihlebaek C. Change in quality of life in older people with dementia participating in Paro-activity: a cluster-randomized controlled trial. J Adv Nurs 2016 Dec;72(12):3020-3033. [doi: 10.1111/jan.13076] [Medline: 27434512]
24. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
25. Ke C, Lou VW, Tan KC, Wai MY, Chan LL. Changes in technology acceptance among older people with dementia: the role of social robot engagement. Int J Med Inform 2020 Sep;141:104241-104249. [doi: 10.1016/j.ijmedinf.2020.104241] [Medline: 32739611]
26. Chen K, Lou VW, Tan KC, Wai MY, Chan LL. Effects of a humanoid companion robot on dementia symptoms and caregiver distress for residents in long-term care. J Am Med Dir Assoc 2020 Nov;21(11):1724-1728.e3. [doi: 10.1016/j.jamda.2020.05.036] [Medline: 32713772]
27. Soler MV, Agüera-Ortiz L, Rodríguez JO, Rebolledo CM, Muñoz AP, Pérez IR, et al. Social robots in advanced dementia. Front Aging Neurosci 2015;7:133-144 [FREE Full text] [doi: 10.3389/fnagi.2015.00133] [Medline: 26388764]
28. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc 2017;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]
29. Sugiyama H, Nakamura K. Temporary improvement of cognitive and behavioral scales for dementia elderly by shiritori word game with a dialogue robot: a pilot study. Front Robot AI 2022;9:941056-941072 [FREE Full text] [doi: 10.3389/frobt.2022.941056] [Medline: 36530499]
30. Bradwell H, Edwards KJ, Winnington R, Thill S, Allgar V, Jones RB. Implementing affordable socially assistive pet robots in care homes before and during the COVID-19 pandemic: stratified cluster randomized controlled trial and mixed methods study. JMIR Aging 2022 Aug 24;5(3):e38864 [FREE Full text] [doi: 10.2196/38864] [Medline: 35830959]
31. Petersen S, Houston S, Qin H, Tague C, Studley J. The utilization of robotic pets in dementia care. J Alzheimers Dis 2017;55(2):569-574 [FREE Full text] [doi: 10.3233/JAD-160703] [Medline: 27716673]
32. Leung DKY, Chan WC, Spector A, Wong GHY. Prevalence of depression, anxiety, and apathy symptoms across dementia stages: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2021 Sep;36(9):1330-1344. [doi: 10.1002/gps.5556] [Medline: 33905138]
33. Otaka E, Osawa A, Kato K, Obayashi Y, Uehara S, Kamiya M, et al. Positive emotional responses to socially assistive robots in people with dementia: pilot study. JMIR Aging 2024 Apr 11;7:e52443 [FREE Full text] [doi: 10.2196/52443] [Medline: 38623717]
34. Mitchell AM, Chiappetta L, Boucek L, Cain M, Patterson G, Owens K, et al. Nonpharmacological therapeutic techniques to decrease agitation in geriatric psychiatric patients with dementia. J Gerontol Nurs 2015;41(2):53-59. [doi: 10.3928/00989134-20141014-02] [Medline: 25347866]
35. Sano M, Cummings J, Auer S, Bergh S, Fischer CE, Gerritsen D, et al. Agitation in cognitive disorders: progress in the international psychogeriatric association consensus clinical and research definition. Int Psychogeriatr 2024;36(4):238-250. [doi: 10.1017/s1041610222001041]
36. Abbott R, Orr N, McGill P, Whear R, Bethel A, Garside R, et al. How do "robopets" impact the health and well-being of residents in care homes? A systematic review of qualitative and quantitative evidence. Int J Older People Nurs 2019;14(3):e12239 [FREE Full text] [doi: 10.1111/opn.12239] [Medline: 31070870]
37. Gan J, Shi Z, Zuo C, Zhao X, Liu S, Chen Y, et al. Analysis of positron emission tomography hypometabolic patterns and neuropsychiatric symptoms in patients with dementia syndromes. CNS Neurosci Ther 2023 Aug;29(8):2193-2205 [FREE Full text] [doi: 10.1111/cns.14169] [Medline: 36924296]
38. Andersen CK, Wittrup-Jensen KU, Lolk A, Andersen K, Kragh-Sørensen P. Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia. Health Qual Life Outcomes 2004 Sep 21;2:52 [FREE Full text] [doi: 10.1186/1477-7525-2-52] [Medline: 15383148]
39. Clarke C, Woods B, Moniz-Cook E, Mountain G, Øksnebjerg L, Chattat R, et al. Measuring the well-being of people with dementia: a conceptual scoping review. Health Qual Life Outcomes 2020 Jul 24;18(1):249 [FREE Full text] [doi: 10.1186/s12955-020-01440-x] [Medline: 32709238]

XSL•FO
RenderX

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference
WMD: weighted mean difference

Edited by A Mavragani; submitted 24.04.24; peer-reviewed by S Su, R Saifi; comments to author 22.10.24; revised version received 10.12.24; accepted 02.01.25; published 10.03.25.

Please cite as:
Fan W, Zhao R, Liu X, Ge L
Intelligent Robot Interventions for People With Dementia: Systematic Review and Meta-Analysis of Randomized Controlled Trials
J Med Internet Res 2025;27:e59892
URL: https://www.jmir.org/2025/1/e59892
doi:10.2196/59892
PMID:

©Wenqi Fan, Rui Zhao, Xiaoxia Liu, Lina Ge. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Examining Challenges to Co-Design Digital Health Interventions With End Users: Systematic Review

Anthony Duffy[1], BA, MDM; Nazanin Boroumandzad[1], BSc; Alfredo Lopez Sherman[1], BA; Gregory Christie[2], BSc, MSc, PhD; Indira Riadi[3], BSc, PhD; Sylvain Moreno[1], BSc, MSc, PhD

[1]School of Interactive Arts & Technology, Simon Fraser University, Surrey, BC, Canada

[2]Meta, Seattle, WA, United States

[3]School of Gerontology, Simon Fraser Universiity, Vancouver, BC, Canada

Corresponding Author:
Sylvain Moreno, BSc, MSc, PhD
School of Interactive Arts & Technology
Simon Fraser University
13450 102 Ave #250
Surrey, BC, V3T 0A3
Canada
Phone: 1 (778) 782 9742
Email: sylvain_moreno@sfu.ca

Abstract

Background: Digital health interventions (DHIs) are changing the dynamic of health care by providing personalized, private, and instantaneous solutions to end users. However, the explosion of digital health has been fraught with challenges. The approach to co-design with end users varies across a diverse domain of stakeholders, often resulting in siloed approaches with no clear consensus. The concept of validating user experiences contrasts greatly between digital stakeholders (ie, user experience and retention) and health stakeholders (ie, safety and efficacy). Several methodologies and frameworks are being implemented to address this challenge to varying degrees of success.

Objective: We aimed to broadly examine the advancements and challenges to co-design DHIs with end users over the last decade. This task was undertaken to identify the key problem areas at the domain level, with the ultimate goal of creating recommendations for better approaches to co-design DHIs with end users.

Methods: We conducted a systematic search of key databases for co-design studies involving end users in DHIs. Searches were divided into 3 relevant streams: health behavior, user experience, and digital methodologies and frameworks. The eligibility criteria were guided by the PerSPEcTiF framework and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist. In line with this framework, studies were included in this review that (1) address research on DHIs; (2) focus on interaction and co-design with end users; (3) explain results such that uptake, effectiveness, satisfaction, and health outcomes are discernible, positively or negatively; and (4) describe actionable procedures for better DHI design. The search was conducted in a diverse group of 6 bibliographical databases from January 2015 to May 2024: PsycINFO, PubMed (MEDLINE), Web of Science, CINAHL, Institute of Electrical and Electronics Engineers Xplore, and Scopus. From the 13,961 studies initially screened for titles and abstracts, 489 (3.6%) were eligible for a full-text screening, of which 171 (1.2%) studies matched the inclusion criteria and were included in a qualitative synthesis.

Results: Of the 171 studies analyzed across 52 journals, we found 5 different research approaches, spanning 8 different digital health solution types and 5 different design methodologies. These studies identified several core themes when co-designing with end users: advancements, which included participatory co-design; challenges, which included participatory co-design, environment and context, testing, and cost and scale; and gaps, which included a pragmatic hybridized framework and industry implementability.

Conclusions: This research supports a pragmatic shift toward using mixed methods approaches at scale, methods that are primed to take advantage of the emerging big data era of digital health co-design. This organic outlook should blend the vision of digital health co-designers with the pragmatism of Agile design methodology and the rigor of health care metrics.

Trial Registration: PROSPERO CRD42021238164; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021238164

International Registered Report Identifier (IRRID): RR2-10.2196/28083

XSL•FO
RenderX

(J Med Internet Res 2025;27:e50178) doi:10.2196/50178

KEYWORDS

digital health; end users; user experience; health behavior; intervention; co-design; mobile health; mHealth; digital health citizen

Introduction

Background

Digital health interventions (DHIs) are revolutionizing health care delivery by providing personalized, private, and instantaneous solutions to end users. An umbrella term, a DHI is a mobile- or web-based digital solution that facilitates a diverse array of health interventions—be they emotional, decisional, or behavioral. A DHI is designed for a complex set of end users—those who use the intervention, including patients, caregivers, and health care professionals [1]. With the ever-increasing ubiquity of mobile health (mHealth) technologies, cost-effective scalable DHIs are emerging almost daily [2]. These solutions range from self-monitoring to diagnostic, educational, and behavior change–based interventions. At the heart of a DHI is the contextual adaptation [3] around the needs of end users, those who use the technology directly to impact their health [4], such as a patient with diabetes monitoring blood sugar on an mHealth app. This is an exciting yet challenging shift in health care design. It promises more autonomy to the end user through a bottom-up [5] design approach; however, it ruptures the top-down [6] traditional health care approach that positions experts as the central designers of interventions. This evolution marks a shift from visiting health facilities for health services toward the ubiquitous digital participation of end users and health experts. This has produced a new era of health literacy built around diagnostics, services, and self-help. This reciprocal, empowering approach is creating a more mutual understanding [7], evolving the dynamic [8] among health care professionals, engineers, designers, and end users in DHI co-design, ultimately shifting the balance of power toward the patient. Nonetheless, the promise of DHIs is tempered by the fact that 26% of apps are discarded after first use [9]. To address this concern, numerous design approaches such as user-centered and human-centered design, among others, are being used to varying degrees of success [10]. Each approach seeks to incorporate end users, from the primary stages of ideation to the final testing stages of production, gathered around multidisciplinary teams of digital and health care professionals. It is within this focus that we have examined the prominent plus points (ie, advancements), pain points (ie, challenges), and system-wide gaps to incorporate end users in co-design, issues that cut across the entire range of design approaches used in digital health at present.

The Complexity of Hybridizing Digital and Health Approaches to Incorporate End Users

Despite today's digital sphere increasingly being one of digital natives, the path toward a fluid, organic digital health design space has been fraught with challenges. From a digital perspective, rapid Agile development methods evolve to garner end-user satisfaction in an evolving form of iterative qualitative evaluation. From a health care perspective, slow, safe expert-led studies produce large quantitative datasets toward validating patient involvement [11]. Digital health, in its nascent stages, sits somewhat uncomfortably between these two domains. From the macrolevel of policy down to the microlevel of co-design workshops, the very *language of collaboration* may differ. Hybridizing these 2 contrarian methodologies is challenging. The concept of building validation around the user or patient varies, depending on the given perspective. These challenges, along with others, have led to a multiplicity of design approaches, a disparity in conceptualizing the end user, and competing agendas [6,7,12,13] among DHI stakeholders that muddy the waters of digital health design. In this regard, we were inspired to conduct this systematic review on the basis of the initial analysis of our previous study titled, "The Challenges Toward Real-world Implementation of Digital Health Design Approaches" [10]. In our primary study, we found that irrespective of the domain, design methodology, or user definition (Multimedia Appendix 1), there are common pain points (ie, repeated points of conflict or displeasure resulting in a poor user experience [UX]) emerging that impact digital health at the domain level.

In addition, numerous studies [3,6,11,14,15] have noted the challenge to design frameworks that effectively define and implement "co-design" or "co-leadership" around end-user involvement in DHI design. On one hand, health stakeholders may lament the lack of alignment of UX design to health care delivery [9]; on the other hand, digital stakeholders may insist that user-centered design (UCD) is essential toward acceptability [16]. In both cases, each stakeholder is viewing *efficacy* through a different lens. This cuts to the heart of the matter of co-designing DHIs with end users; the overarching philosophy and methods used are starkly different. Therefore, there is a need to better understand the common pain points in this collaboration, which in turn may lead to a better perspective on why, when, and how to incorporate end users in the design of DHIs.

Objective of This Study

This study seeks to compile the core challenges to incorporate end users in the design of DHIs over the last decade. We sought to thematically analyze current approaches, to analyze evidence of pain points that presented a challenge or conflict in the incorporation of end users in the co-design process, and to explore gaps and opportunities toward a more organic co-design approach in digital health. This includes a broad stroke of studies, including quantitative, qualitative, mixed methods, and research through design. It features the perspectives of health care, psychology, engineering, development, and design in the production of mHealth, eHealth, and extended reality DHIs. This intentionally broad approach is reflective of the nascency of digital health as a field and the open-mindedness of this study, such that the divergence and convergence of methodologies and tools is viewed as a prosperous and necessary growing pain toward a more organic digital health approach to end users.

XSL•FO
RenderX

It is hoped that through critical analysis of these advancements, challenges, and gaps, synthesized findings from the following questions lead to an improved approach toward incorporating end users in DHI co-design:

1. How are end users currently incorporated into iterative DHI design, and what is the effectiveness of current methods?
2. What are the most common pain points (ie, challenges) encountered while incorporating end users in DHI design activities?
3. What are the current gaps in end-user research at the domain level, and how can these be addressed to better integrate end users into the iterative design process in digital health, aiming to improve efficacy and uptake?

Methods

Study Design

This systematic review followed an a priori published protocol with detailed methods registered with PROSPERO (CRD42021238164) [17]. The review followed the Cochrane Handbook for Systematic Reviews [18] and was reported using the PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols) guidelines [19].

We followed six stages in this systematic review: (1) literature search, (2) article selection, (3) data extraction, (4) quality appraisal, (5) data analysis, and (6) data synthesis.

Eligibility Criteria

Considering the complexity of DHIs, we used the PerSPEcTiF [20] guidelines for intervention (Table 1). We selected it due to its suitability for qualitative synthesis in the health care domain.

In line with this framework, studies were included in this review that (1) address research on DHIs; (2) focus on interaction and co-design with end users; (3) explain results such that uptake, effectiveness, satisfaction, and health outcomes are discernible, positively or negatively; and (4) describe actionable procedures for better DHI co-design.

Table 1. Details of the PerSPEcTiF question formulation framework as applied to this review.

Initials	Definition	Detail
Per	Perspective	From the perspective of end users
S	Setting	In the setting of digital health
P	Phenomenon of interest or problem	What are the most prominent pain points
E	Environment	Within an environment of designing DHIs[a]
C	Comparison (optional)	—[b]
Ti	Time	During ideation and co-designing
F	Findings	In relation to understanding the challenges of successfully incorporating end users in the successful design of DHIs

[a]DHI: digital health intervention.

[b]Not applicable.

Search Strategy

We systematically searched 6 electronic databases: PsycINFO, PubMed (MEDLINE), Web of Science, CINAHL, Institute of Electrical and Electronics Engineers Xplore, and Scopus. We selected these databases according to preliminary searches and consultations with experts and librarians in this field. Keywords related to DHIs from preliminary searches were compiled and refined. We adapted the search strategy as needed to return a breadth of papers without retrieving an unmanageably large number of irrelevant articles.

A list of the search terms used in this review were grouped into 3 clusters (Table 2). In addition, we have included all search strings for replicability (Multimedia Appendix 2).

Table 2. Clusters of categorized search keywords.

Categories	Keywords
DHIs[a]	"digital health" or "mHealth" or "eHealth" and "end user*" or "end user*"
User experience	"user experience" or "UX" or "user centred" or "user centered" or "human centred" or "human centered" or "patient centred" or "patient centered" or "person based" or "person centered" or "person centred" or "participatory design" or "involvement" and "digital health" or "mHealth" or "eHealth"
Methodologies and frameworks	"agile" or "scrum" or "kanban" and "digital health" or "mHealth" or "eHealth"

[a]DHI: digital health intervention.

Cluster 1 includes DHIs (owing to behavior change in health care resulting from DHIs). Cluster 2 includes UX (owing to engineering, development, and design in digital health). Cluster 3 includes methodologies and frameworks (owing to digital project facilitation and to health care intervention design and policy).

XSL•FO
RenderX

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were based on the peer-reviewed studies outlined in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Qualitative studies
- Case studies
- Observational studies, including cross-sectional surveys
- Nonrandomized studies (eg, before and after studies and interrupted time series studies)
- Cohort studies

Exclusion criteria

- Noncritical analysis of end-user involvement in digital health design (ie, no results or discussion on end-user incorporation)
- Single person studies
- Nonprimary research (eg, reviews and commentaries)
- Randomized controlled trials
- Nonhuman studies

Due to the rapid innovation of digital health, only studies in the English language from January 1, 2015, to May 31, 2024, were included. Broadly put, the inclusion criteria included any population group that involved the critical analysis of end-user incorporation (ie, inclusive of terms such as patient, person, and human) in the co-design of a DHI. Studies that did not meet this criteria were excluded, including those that included end users but did not provide any critical analysis of the process; reviews and commentaries due to them being nonobservational or nonexperimental; single-user studies due to the inability to generalize; randomized controlled trials (RCTs) due to a lack of focus on the end-user integration in the design of a DHI; and nonhuman studies, which are impertinent to the design of a DHI.

Screening and Article Selection

All articles identified and selected from database searches were stored in the reference management software Paperpile (Paperpile LLC), which was used to eliminate duplicates and to tag and organize the research findings. The initial screening process involved 2 independent reviewers (AD and either IR or NB) who screened titles and abstracts of all studies for possible inclusion using an Excel (Microsoft Corp) spreadsheet (Multimedia Appendices 3 and 4) to ensure no studies met the exclusion criteria. After filtering results on the basis of the possible inclusion criteria, the second stage involved a full-text review of the remaining articles, which was conducted by 3 independent reviewers (AD and IR or AS). Upon scrutiny of the full-text selections, reviewers (AD and IR or AS) removed items that did not meet the inclusion criteria. In the third and final stage, reviewers (AD and IR or AS) met to cross-reference eligibility of the final inclusion list, as well as to settle any conflicts. This cross-reference involved the individual review of all items to white list them per the inclusion criteria. For items in conflict, reviewers (AD and IR or AS) simultaneously scanned items side by side to determine whether the items met the inclusion criteria. Since this was the third round of filtering, the majority of items met the inclusion criteria, so the determining factor in conflicts was to collectively review each item to see whether the studies actively involved end users in the design of the DHI and whether critical analysis of this was present in the results or discussion sections of the respective study. This final step removed any subjectivity and resolved conflict.

Data Extraction, Analysis, and Synthesis

A data extraction template was created by one of the authors (AD) that provided a record of the study characteristics. The categorization of themes and subthemes emerged organically from the findings (Multimedia Appendix 5). The collection, extraction, and subsequent analysis and synthesis of data is outlined in Table 3.

Raw content (ie, quotes and reflective analysis) was extracted and collected in a research document (Multimedia Appendix 6). Data were extracted and analyzed relative to the goals of the research questions, following guidance from the Cochrane Handbook for Systematic Reviews of Interventions [18]. These data are synthesized and presented in narrative form in the Results section. Full characteristics of all included end-user studies are available in Multimedia Appendix 7.

XSL•FO
RenderX

Table 3. Data extraction characteristics and the identified themes and subthemes.

Category	Description
Data extraction	
Summary of characteristics	Full citation, author, PMID[a], year, and journal
Study type	Case study, quantitative study, qualitative study, research article, and mixed methods study
Study details	Target population, end users, duration (h), and solution
Design methodology	User-centered design, human-centered design, patient-centered design, person-based approach, patient-led design, and general participatory design.
Project facilitation method	Kanban, Scrumban, Waterfall, XP[b], Agile inspired, design thinking, and unspecified
Data analysis and synthesis (themes and subthemes)	
Key plus points (ie, positives)	
Participatory co-design	Collaboration, ideation, improved mode of delivery, behavior modification, and in situ environment
Key pain points (ie, negatives)	
Participatory co-design	Collaboration, the language of digital health, co-design, competing interests, hierarchical versus democratic design, and methodological misalignment
Environment and context	Untested in situ and multiplicity of user types
Testing	Lack of participants, quality of participants, condensed time frame, lack of comprehension, technology, and cultural sensitivity
Cost and scale	
Challenges (ie, gaps in the research)	Organic co-design approaches, fluidity of digital health efficacy evaluation, Agile digital health methods, industry implementability, supply chain, medical device versus consumer app, and research and development limitations

[a]PMID: PubMed identifier.

[b]XP: extreme programming.

Results

Search Results

The search strategy identified 13,961 studies. After removing duplicates (n=794), a total of 13,167 (94%) studies were screened for title and abstract. After the initial title and abstract screening phase, 12,678 (91%) studies were excluded, leaving 489 (4%) potentially relevant papers. In the second full-text screening phase, 318 (2%) studies were excluded. This resulted in the inclusion of 171 (1%) relevant studies in this systematic review. Reasons for exclusion included a lack of critical analysis of end-user involvement in digital health co-design and nonprimary research (eg, reviews and op-eds, RCTs, and single-user case studies). A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1) was used to record the details of the screening and selection process so that the study can be reproduced.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study inclusion and exclusion process.

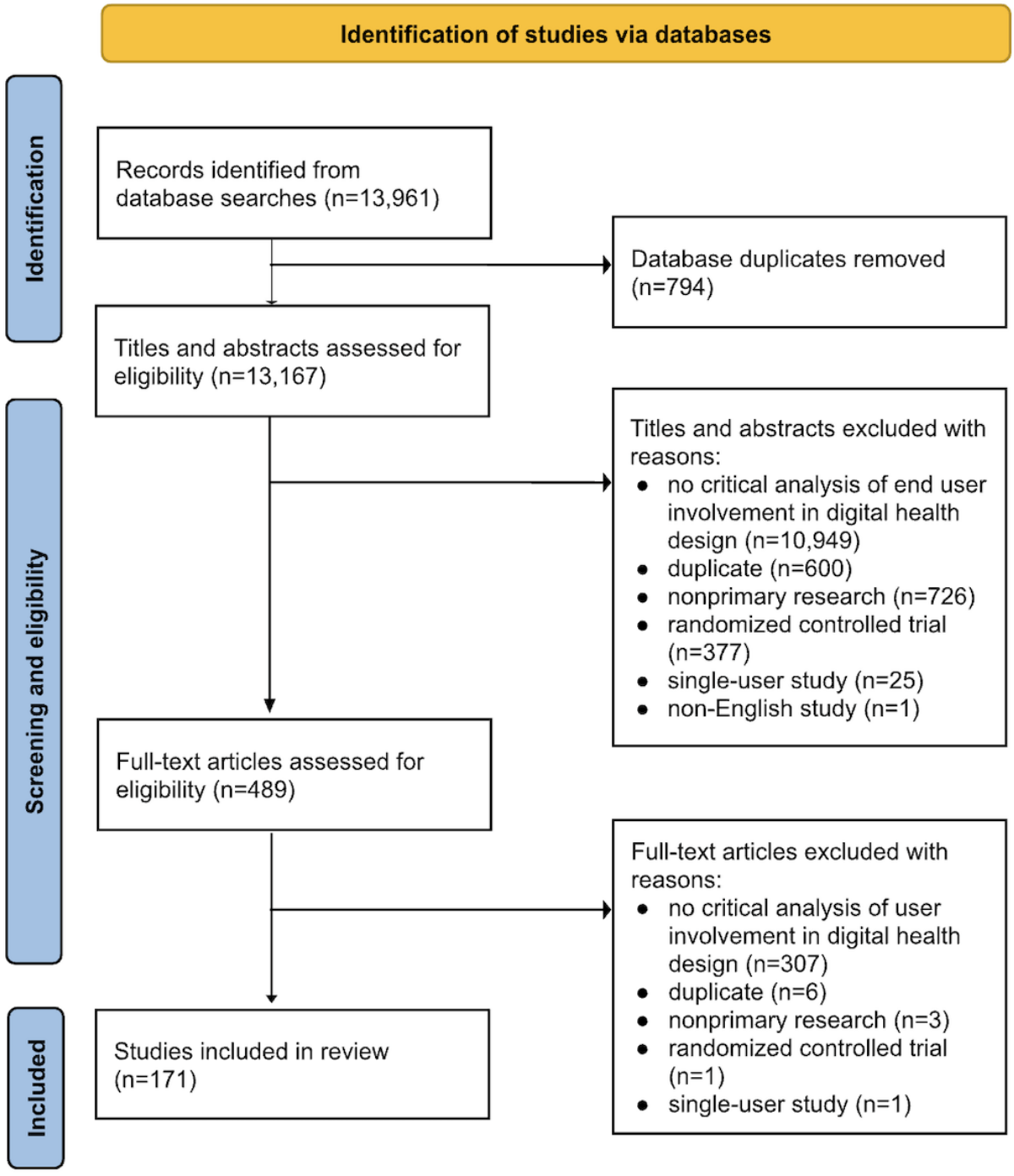

Study Characteristics

Before addressing our research questions, we present a record of characteristics detailing the diversity of digital health studies retrieved (Table 4).

XSL•FO
RenderX

Table 4. Analysis of study characteristics (N=171).

Study characteristics	Studies, n (%)
Year	
2015	4 (2.3)
2016	7 (4.1)
2017	14 (8.2)
2018	18 (10.5)
2019	25 (14.6)
2020	38 (22.2)
2021	10 (5.8)
2022	26 (15.2)
2023	24 (14)
2024	5 (2.9)
Type and solution	
Qualitative	107 (62.6)
Quantitative	6 (3.5)
Mixed methods	43 (25.1)
Case study	12 (7)
Research article	4 (2.3)
Methodology and methods (project facilitation[a])	
Agile	24 (14)
Scrum	9 (5.3)
Design thinking	10 (5.8)
Unspecified	133 (77.8)
Target population and end users	
Patients	91 (53.2)
Older adults	23 (13.4)
Health professionals	15 (8.8)
Young people	11 (6.4)
Adults	9 (5.3)
Patients and health professionals	5 (2.9)
Other	17 (9.9)
Journal type	
JMIR	94 (55)
BMC[b]	11 (6.4)
IEEE[c]	6 (3.5)
Other	60 (35.1)
Digital solution	
mHealth[d]	97 (56.7)
Web based	35 (20.5)
eHealth	22 (12.9)
Device	7 (4.1)
Serious gaming	6 (3.5)

XSL•FO
RenderX

Study characteristics	Studies, n (%)
Other	4 (2.4)
Design methodology[e]	
User-centered design	84 (49.1)
Participatory design	58 (33.9)
Human-centered design	18 (10.5)
Person-centered design	9 (5.3)
Person-based approach	4 (2.3)
Person-led design	1 (0.6)
Number of end users involved	
1-5	21 (12.3)
6-10	44 (25.7)
11-15	28 (16.4)
16-20	15 (8.8)
21-30	25 (14.6)
31-40	7 (4.1)
41-50	6 (3.5)
≥51	9 (5.3)
Unspecified	16 (9.4)

[a]Three studies identified with Scrum, which is a subset of Agile.

[b]BMC: Biomed Central.

[c]IEEE: Institute of Electrical and Electronics Engineers.

[d]mHealth: mobile health.

[e]Three studies identified more than one methodology.

Study Types

Owing to the diversity of digital health, the 171 studies identified span 5 different research approaches. Most approaches were qualitative (107/171, 62.6%), followed by mixed methods (43/171, 25.1%), indicative of the social and emotional influence on DHIs. These studies were collected from 52 different journals across multiple research domains, with JMIR accounting for the majority (94/171, 55%).

Study Details

The DHIs spanned 8 different solution types with mHealth (97/171, 56.7%) representing the majority. This development is in line with the growing dominance of mobile digital devices with up to 92% of internet traffic at present being mobile based [21].

These solutions were used in 15 different populations with patient-focused populations being the most prominent (91/171, 53.2%). There was considerable variation in the number of end users involved in a DHI co-design with 6 to10 (44/171, 25.7%) end users being the most prominent, followed by 11 to15 (28/171, 16.4%), 21 to 30 (25/171, 14.6%), and 1 to 5 (21/171, 12.3%) end users. Only 9 studies (5.3%) involved >50 end users.

Study Design Methodology

As a guiding design methodology, UCD (84/171, 49.1%) was most prominent, followed by the broader approach of participatory design (58/171, 33.9%), then human-centered design (18/171, 10.5%). Other approaches such as patient-centered design, patient-based approach, and patient-led design were used sparingly.

Study Project Facilitation Method

Finally, in addition to tracking the design methodology, we tracked the project facilitation method used. The vast majority of studies did not specify a project facilitation method (133/171, 77.8%) that was used. Of those that did, Agile was used the most (24/171, 14%), followed by design thinking (10/171, 5.8%). Scrum, a subset of Agile, was next (9/171, 5.3%), while none specified other common industry methods such as Kanban, Extreme Programming, or Waterfall. This is particularly interesting when comparing how digital health prototypes are facilitated in academic research environments versus typical industry methods.

We present the aforementioned study characteristics to illustrate the diversity of investigating the incorporation of end users in co-design. As discussed, this rigorous approach sought to incorporate as many vantage points as possible to identify common advancements, challenges, and opportunities for improvement in the domain. A thematic analysis of the plus points (ie, advantages), pain points (ie, disadvantages), and challenges to incorporate end users in the co-design of DHIs

XSL•FO
RenderX

(Figure 2) as addressed in the 3 research questions is presented later.

A brief definition of the resultant themes and subthemes is presented in Table 5. This is followed by an explanation of the results relevant to each of the 3 research questions in this study.

Figure 2. Flowchart of the thematic analysis of plus points (ie, advantages), pain points (ie, disadvantages), and challenges to co-design digital health interventions (DHIs) with end users.

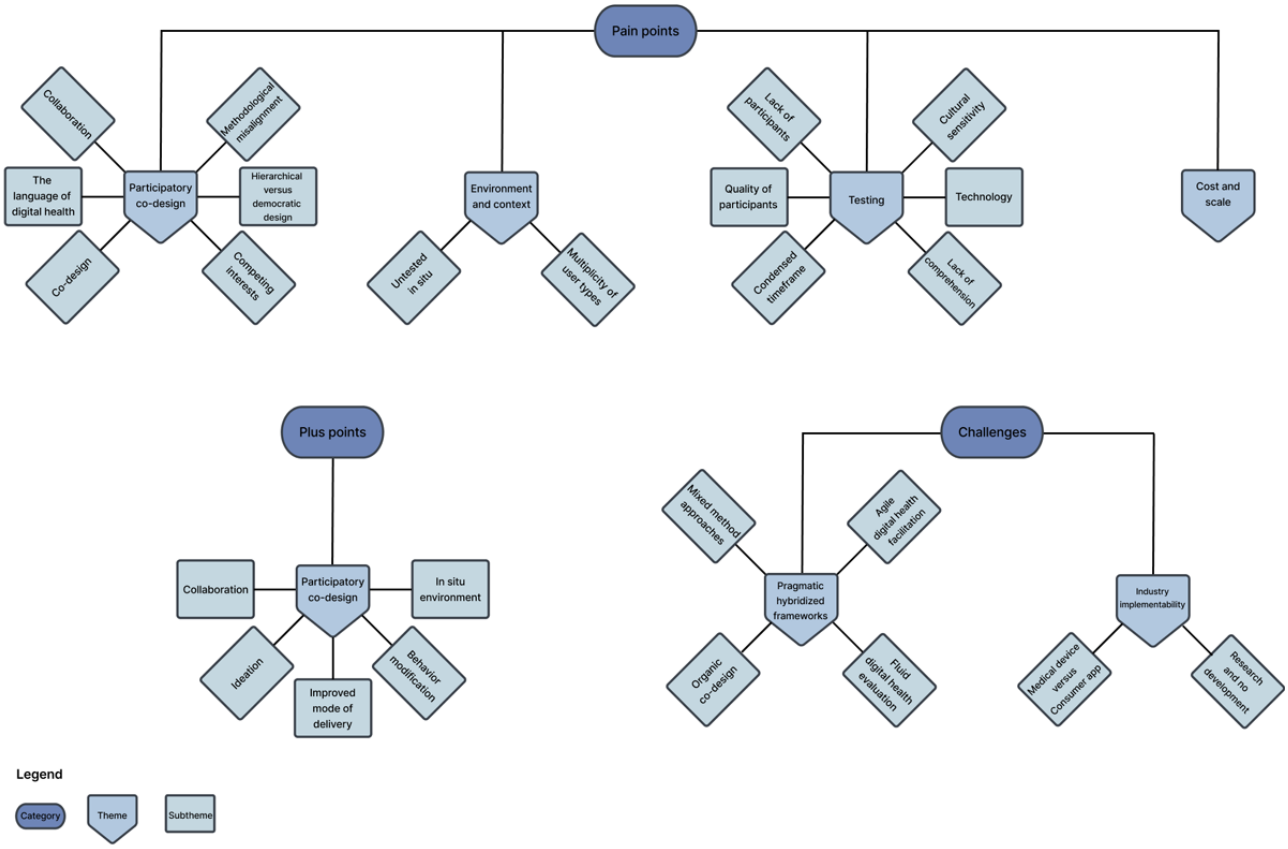

XSL•FO
RenderX

Table 5. Description of the themes and subthemes identified in the data synthesis.

Section and theme	Description	Studies, n (%)
Plus points		
Theme: Participatory co-design	Positive advancements from the active participation of end users in the co-design workshop phases to ideate and test DHIs[a]	95 (55.5)
Subthemes		
Collaboration	Collaboration among stakeholders (eg, developers, health professionals, and end users) in the co-design of a DHI	69 (72.6)
Ideation	Creative process to construct ideas, requirements, and design concepts in the nascent stages of co-design with end users	55 (57.9)
Immediate behavior modification	The benefits of instant behavior changes obtained by end users through the knowledge enhancing collaboration of DHI design workshops	16 (16.8)
In situ environment	Ideating and testing with end users in the actual environment the intervention is designed for (ie, hospital, care homes, among others)	7 (7.4)
Pain points		
Theme: Participatory co-design	Failures when incorporating end users in the active participation of co-design workshop phases to ideate and test DHIs	70 (40.9)
Subthemes		
Collaboration challenges	Failures to curate and implement collaboration approaches among stakeholders (eg, developers, health professionals and end users) in the co-design of a DHI	14 (20)
The language of digital health	Difficulties communicating a common language of co-design considering the various semantics that stem from differences in collaborator perspectives (ie, digital, health care, and end users)	14 (20)
Cultural sensitivity	Understanding the nuances of culturally appropriate approaches that are used to describe the content and functionality of DHIs with end users	13 (18.6)
Competing interests	Contrasting objectives defined by health and digital goals that create friction in the process of designing around the end users	18 (25.7)
Methodological misalignment	Differences in digital health intervention design stemming from methodological misalignment (ie, digital, health or behavioral approaches to co-design)	13 (18.6)
Hierarchical versus democratic design	The challenging contrast between the slow, steady, and data-driven approach to health care design and the laissez-faire rapid approach to digital design innovation, making processes and evaluation difficult in digital health	17 (24.3)
Theme: Environment and context	The collaboration and testing environment of the end user and the contextual understanding of who the end user is	38 (22.2)
Subthemes		
Untested in situ	The lack of co-designing or testing in the contextual environment of the user	23 (60.5)
Multiplicity of user types	Lack of understanding of the complexity of user types (ie, patient, person, clinician, therapist, and caregiver)	16 (42.1)
Theme: testing	Problems incurred during the testing phase with end-user collaborators that led to questions on the validity, reliability, and generalizability of results	113 (66.1)
Subthemes		
Lack of participants	In reference to a lack of end-user testers leading to questions on the validity and reliability of results and determinations	75 (66.4)
Quality of participants	In reference to a lack of broader representation of user types (ie, patient user, caregiver user, and clinician user), including demographic concerns (ie, ethnic diversity), thus weakening the generalizability of results	58 (51.3)

XSL•FO
RenderX

Section and theme	Description	Studies, n (%)
Condensed timeframe and limited participation	The condensed time frame leading to limited participation of end users in collaboration workshops and testing, raising questions on rigor and applicability of results broadly	25 (22.1)
Lack of technological comprehension	Inclusive of knowledge of technical terminology and digital device variance (ie, iPhone vs Android)	14 (12.4)
Challenges		
Theme: Pragmatic hybridized frameworks	The need for new frameworks that are friction reducing, pragmatic (ie, fit for purpose), and hybridized (ie, a collective fluid combination of digital, health and behavioral objectives), leading toward more transdisciplinary approaches in DHI design	60 (35.1)
Subthemes		
The value of mixed method approaches to end-user data	Despite the advantages that mixed methods approaches bring to DHIs (ie, blending emotional and functional feedback for better design decisions), the uptake in mixed methods approaches to data collection is low and inhibits more pragmatic approaches to DHI evaluation	24 (40)
Organic co-design approaches	The challenge to cocreate approaches to end-user involvement in co-design that are not a sum-of-its-parts approach; siloing digital (ie, usability) and health care (ie, safety) elevation as unrelated metrics	25 (41.7)
Fluid digital health efficacy evaluation	The challenge to amalgamate the *fail fast* approach of digital design and the *first do no harm* approach of health care design around end-user design involvement	39 (65)
Agile digital health project facilitation	The difficulty to implement Agile industry-styled co-design practices with the rigor and requirements of health care safety and efficacy	18 (30)
Theme: Industry implementability	The lack of consideration when co-designing with end users for downstream industry implementation within the complexity of the health care system (ie, government policy and industry constraints and regulations that impact design)	13 (7.6)
Subthemes		
Medical device vs consumer app	The challenge to navigate consumer (ie, end user) preferences that collide with medical device regulation	5 (38.5)
Supply chain	Health care delivery involves the collaboration of industry suppliers and government regulatory bodies that bring DHIs to the end users. The lack of health care business analysts in the co-design process with end users negates business modeling factors that could reduce downstream deployment challenges.	6 (46.2)
Research and no development	The lack of recognition and resolution of the aforementioned failures and challenges often leads to prototypes that never reach the development or deployment stages	5 (38.5)

[a]DHI: digital health intervention.

Research Question 1: How Are End Users Currently Incorporated Into Iterative DHI Co-Design, and What Is the Effectiveness of Current Methods?

Theme: Participatory Co-Design

The most positive and influential approach to the incorporation of end users in the co-design of DHIs can broadly be termed as participatory co-design (ie, an approach wherein end users participate in the co-design process). In total, 95 of the 171 (55.5%) included studies discussed the advantages and advancements of participatory co-design in the digital health domain, irrespective of the domain approach (ie, engineering and health among others). This overarching methodological approach culminated in subthemes such as collaboration, ideation, environment, and improved behavior modification.

Subtheme: Collaboration

Collaboration (69/171, 40.4%) [6,7,9,12,22-87] was identified as a key benefit of participatory co-design. Studies explained that collaboration in DHI design allows complex teams to co-create dynamic solutions [30] that uniquely combine a variety of digital and health approaches around the end user [23,88,89]. Collaboration may take the form of interviews, design workshops, and crowdsourcing (ie, web-based surveys and mass testing). This constructive, creative, iterative [90], and engaging process reflects the diverse mosaic of digital health stakeholders, while increasing the understanding of end-user needs due to the variety of vantage points [91]. Increasingly, studies showed that collaborating with a diverse stakeholder group around the end user, from ideation to production, increased end-user design participation [8,28,36], improving evidence-based approaches

XSL•FO
RenderX

[92]. This development is increasingly important, since earlier attempts to incorporate end users in late-production testing have been widely found to be ineffective [22]. Collaboration, particularly through the early low-fidelity stages of prototype design, increased adaptability of the solution in a low-cost phase [36]. It also created mutual understanding among all stakeholders regarding their roles and how they relate to the DHI as a sum of its parts rather than as siloed experts [37,93]. This approach is increasing sensitivity to the cultural context and personal preferences of end users, which in turn is improving the common language of communication in digital health solutions [32]. In short, collaboration is the element of participatory co-design that is closing the gap between the patient, clinician, and designer [39].

Subtheme: Ideation

Ideation (55/171, 32.2%) was identified as a key benefit in participatory co-design. Loosely defined as the ability to generate new ideas [30], studies used it as a hands-on approach to incorporate users within digital health teams in workshops [29] to cocreate and prioritize end-user preferences [30,94] and subsequent features in a DHI. This process was used early in the DHI design process, often encapsulated in design sprints or health game jams [95] that empathetically seek to understand end-user needs in a process of divergence and convergence [14,38], whereby finding differences is seen as a necessary building block on the path to agreeing on a robust co-designed solution.

Subtheme: Immediate Behavior Modification

Immediate behavior modification (16/171, 9%) [15,24,27,39,42,51,64,68,69,74,76,96-98] in participating end users was identified as a beneficial yet perhaps unexpected advantage of in situ collaboration and ideation. Studies observed instant behavior modification from incorporating end users in DHI workshops. The multidisciplinary nature of rapid iteration that organically collects design, development, health care, and end-user experience data in a decentralized experimental environment allowed for persuasive design and behavior change in real time. This reciprocal process allowed digital and health stakeholders to educate end users on how a DHI can improve their experience, while end users were able to educate other stakeholders on how and what makes a positive UX for them.

Subtheme: In Situ Environment

Co-designing in situ (7/171, 4%) [35,53,72,89,91,99,100] such that on-site collaboration with end users took place in hospitals, clinical environments, or homes [89] increased the realism of stakeholder participation, leading to a more natural evaluation of the intervention. In such studies, this led to a better ability to co-conceptualize solutions [101]. This shift is in response to studies showing limitations on the quality of data and the potential effectiveness of DHIs that were commonly not tested in situ (eg, university laboratory environments), which challenges external validity [35,100]. In situ co-design increased familiarity of surroundings, enhanced comfort, and provided a more genuine, meaningful interaction with end users [91].

Research Question 2: What Are the Most Common Pain Points (That Is, Challenges) Encountered While Incorporating End Users in DHI Design Activities?

Theme: Participatory Co-Design

Participatory co-design was simultaneously a core theme of the positive advancements as well as common failures to incorporate end users in the design of DHIs (70/171, 40.9%). This dichotomy is not unexpected considering the nascency of the burgeoning digital health industry. These challenges are related to collaboration, communication (ie, the language of digital health), competing interests, hierarchical versus democratic design approaches, and methodological misalignment. Broadly speaking, these challenges are growing pains toward blending digital and health care approaches in co-design.

Subtheme: Collaboration Challenges

Collaboration challenges (14/171, 8%) [5,7,30,41,53,65,68,71,88,89,102-105] in participatory co-design ranged from the challenge to curate collaboration approaches based on the given use-case scenario, to variance in stakeholders' knowledge bases (ie, technology, health care, and UX), to miscalculating the need for onboarding co-designers in the workshop process, to mediating disagreements [102], and misallocating stakeholder types (ie, too many clinicians, not enough end users, or vice versa) [103]. These factors led to disagreements, increased timelines, and overbudgeting that challenged the effectiveness of the co-design around the end user. These collaboration challenges are resultant of a clash of cultures (eg, digital and health care) and are defined by the following subthemes.

Subtheme: The Language of Digital Health Co-Design

Challenges to form a common language in co-design (14/171, 8%) [42,60,62,64,65,71,74-76,103,106-109] identified the difficulty of blending the semantics of health care, digital design, and the colloquialisms of end users. Design "sprints" [108] sometimes felt rushed to health care stakeholders and end users if unexplained. While safety may indicate imminent danger to end users, in health care, safety is more indicative of preventing harm from happening [107]. Conceptualizing what a prototype was or explaining what a minimum viable product represents in nascent design stages was not a given for nondigital stakeholders. Usability (ie, technical), UX (ie, emotional), and patient efficacy (ie, medical) are all by-products of the co-design process, but the language, interpretation, and implementation in the workshop meant different things to different stakeholders leading to conflicting understandings.

Subtheme: Cultural Sensitivity

Stemming from the common language issue, studies identified the challenge to collaborate with end users in a way that is culturally sensitive and appropriate (13/171, 7.6%) [38,39,42,44,47,57,68,88,110-115]). Cultural sensitivity has been observed in 4 main categories: content, functionality, technology, and user interface [3]. Studies cited the challenge of using inclusive digital media in workshops that is culturally befitting, as well as addressing sensory needs (eg, vision and hearing deficiencies) of patient end users [88]. For example, the use of gamified incentives and rewards in mHealth apps

XSL•FO
RenderX

[38] for communities that suffer from gambling addictions may be a critical oversight. Studies that incorporated cultural frameworks built with end users showed promise toward negating downstream challenges in the design cycle, particularly concerning appropriate content, features, and imagery in the interface design [3,39].

Subtheme: Competing Interests

Competing interests (18/171, 10.5%) [4,7,9,14,16,43,64,78,88,98,101-103,106,116-119] were found to be a prominent challenge within participatory co-design with end users. This can be summarized as the focus on health systems and services versus the needs of matching digital technologies to user preferences [14]. These competing lenses [101] led to contrasting perspectives on how to incorporate end users in DHI co-design. Broadly, it speaks to the clash of cultures as expert-led health care approaches and innovative user-driven approaches collide. The concept of "shared ownership" proved uncomfortable in numerous studies [9,98,101,118]. Although there was a shared vision on improving quality and safety around end users, exactly *how* this was achieved is a point of conflict [7]. Disease management, improved well-being, and UX are all facets of an intelligent design, yet they speak to different objectives [95]. Stakeholders often felt internal pressure to deliver aspects of the solution that reflect their domain (ie, digital, health care, and policy), which challenged the co- in co-design. Where a patient stakeholder may desire anonymous data tracking (eg, for privacy), a health care professional may desire mapped data tracking (eg, for safety and electronic records), and a digital stakeholder may need analytics (eg, to drive sales) [7]. The inability to corral these competing interests led to disengagement or abandonment of the collaboration in some cases [119].

Subtheme: Methodological Misalignment

Ironically, a paradox exists in that the emergence of diverse stakeholder groups seeking to co-design with end users has led to difficulties in communicating and co-designing [120]. Methodological misalignment (13/171, 7.6%) [29,35,47,58,60,64,65,69,91,120-122] in the approach to evaluating safety, usability, and efficacy differed vastly between health care and digital experts, respectively. The goals of these contrasting stakeholders were quite divergent [35]. In this regard, digital health is a domain struggling to formulate its own unique design culture around end users. Contextual barriers [91,121] exist such that the qualitative nature of rapid design sprints does not quantify the risks, safety, or efficacy of a longitudinal study, which spans large datasets (ie, RCTs). Simply put, it can be difficult to gain buy-in due to the vast variation in research approaches.

Subtheme: Hierarchical Versus Democratic Design

Methodological misalignment surfaced most prominently as a conflict between hierarchical and democratic concepts of designing DHIs (17/171, 9.94%) [3,6,9,12,15,35,38,69,90,93,113,123-128]. Traditionally, a more paternalistic relationship [3] exists between health care experts and patients (ie, end users). This hierarchical construct often relies on expert-driven "hard data" decisions. This is contrary to the more democratic, inclusive approach to digital design workshops, where the end user and other stakeholders are often seen as equal. The concept of *leadership and followership*, such that design workshops often see different stakeholders take the lead based on a given exercise [93], is a new and rupturing approach to health care interventions. Evidence-based and experience-based approaches accentuated the juggling act of balancing perspectives and soothing egos. On the design side, certain biases related to the pragmatic mode of using user-driven exercises conflicted with evidence-based research in health care, especially in the arena of safety and clinical efficacy. Design workshop approaches to user personas and user journeys tell a story about a user and their preferences, but they do not account for clinical perspectives on harm prevention and safety. Where digital experts may insist "the solution must occur organically" around the end user, this democratic approach may be viewed as oversimplification by health experts who may insist that a "first do no harm" approach must involve quantitative data that exemplifies safety and efficacy. In this context, the aforementioned challenges culminate in the battle between innovation and clinical reality [15].

Theme: Environment and Contextualization

After understanding the various challenges that participatory co-design poses to stakeholders working with end users in the design of a DHI, an emergent theme that developed was the impact of the collaboration environment and the contextualization of the end user (38/171, 22.2%).

Subtheme: Untested In Situ

Numerous studies observed the challenging impact of not designing a DHI with end users in situ (23/171, 13.4%) [3,7,15,30,68,73,74,77,99,106,110,113,116,129-138]. In situ here refers to co-designing or testing in the contextual environment of the user. For example, a triage digital tablet in a waiting room or a sleep diary app in a user's bedroom. This limited generalizability, challenged the value [110] of design workshops, and weakened the ability to evaluate the true context of use [3,116,129-131,134-137]. Without testing DHIs in their natural environment, the context of the data collected was sometimes considered incomplete. Furthermore, it does not observe how end users would interact with the data and experiences they are creating, *in real-world settings* [30,106,133]. For many mHealth solutions, app use was found to vary by time and location, with various environmental factors impacting how end users may interact with the intervention [7,132]. This made it increasingly difficult to generalize both design considerations and testing conclusions from a controlled laboratory environment. For digital stakeholders, this may render UX evaluation speculative. For health stakeholders, clinical efficacy evaluation may be incomplete.

Subtheme: Multiplicity of User Types

Another factor impacting the contextualization of the DHI was the multiplicity of user types (16/171, 9.4%) [6,47,54,58,72-74,77,80,85,130,137,139-142]. End users were viewed as patients, persons, clinicians, therapists, and passive users (ie, an observer such as a caregiver), among others. Numerous DHIs had more than one end user, each using the solution for a different objective. Unlike traditional collaborative design workshops that involve 1 end-user type or a potential

XSL•FO
RenderX

customer, mHealth solutions often provide services to numerous types of end users. For instance, in heart medication apps, a patient may objectify a knowledge base with push notifications and, as a user goal, a therapist may objectify behavior change as a user goal and a physician may objectify metrics to track medication history as a user goal—all 3 stakeholders may be end users of the same DHI [137]. Therefore, positioning the patient as the sole end user may insufficiently reflect the scope of end-user incorporation in the design process.

Theme: Testing

Foundational challenges in participatory co-design, inclusive of environmental and contextual issues, unsurprisingly culminated with pain points downstream in testing with end users (113/171, 66.1%). Testing subthemes included a lack of participants, the quality of participants, a lack of testing comprehension, and condensed time frames.

Subtheme: Lack of Participants

A lack of participants (ie, end users) challenged the generalizability of results (75/171, 43.8%) [3,7,11,15,25,27,28,32,34,36,41,42,45-49,51-57,60,61,63-66,69,71,72,74,75,80,84,90,94,98,99,104,105,108,110,119,127,129,131-134,136,140,142-163]. This limitation in the number of participants is a common challenge for co-design workshops [28]. The reliance on 5 test participants [11,131,157] is a common practice in qualitative research, where findings from 5 end users are often assumed to reflect the experiences of 50 or 500 end users. This approach is foreign to traditionally quantifiable health research, and is admittedly speculative and potentially biased, as observed in many of the study limitations. Therefore, despite the richness in emotional feedback derived from and around end users in co-design workshops, the lack of quantifiable evidence limits and weakens the broad applicability of results [129]. Terms such as "sufficient" [8] or "generally provable" were used alongside admissions of lack of generalizability. The rapid, cost-effective, human-centered approach [15] to design workshops concedes larger sample sizes, edge cases, and an overall breadth of data obtained from large-scale patient trials in health care. Ironically, while aiming to use methods that address "the needs of end users," [32,34,147,156,157,164,165] these same methods inherently reduce the scope of end user needs, conceding large-scale feedback that could further advocate for design workshops emphasizing "acceptability" and "usability," with large-scale, quantifiable validation from a broader pool of end users.

Subtheme: Lack of Diversity of Participants

The lack of diversity among participants (58/171, 33.9%) [12,14,30,35,36,38,41,42,45-49,51-57,61,63-66,88,95,98-100,104,105,111,124,130,139,145,148,152,155,159,161,166-169] was evident in a third of studies. Studies with small sample sizes were not always able to account for various cultural and educational user differences within a demographic [14,30,88,95,100,111,127,145,155]. The lack of representability, both in variety of end-user types (eg, patient, clinician, caregiver, and others) and skillset (ie, tech savvy superusers [124] vs standard users) contributed to consistent validity and generalizability concerns. With the clinician (ie, patient advisor), patient, and caregiver (ie, observer) often being co-users of the DHI, excluding one end-user group or more from the co-design or testing phases was seen as a risk of biasing results [29,139,168]. The user test of a patient and that of a clinician may vary drastically due to contrary perspectives and user needs, despite them both being end users of the same DHI [35,36,148].

Subtheme: Condensed Time Frame and Limited Participation

In some cases, the quantity and quality of participants was directly impacted by the rapid, condensed time frame (25/171, 14.6%) [14,30,35,37,53,60,64-66,71,80,83,92,99,100,108,116,119,146,148,149,152,170-172] of co-design workshops built around end users. Some studies recognized that despite the overtly flexible approach to iterating with end users, involving them alongside clinicians and digital stakeholders, from conception to validation, is a time-constrained process affected by availability and budget [14,30,100,108,148,149,152,171]. The challenge to coordinate health stakeholders' involvement in end-user co-design was more apparent, considering that design workshops are not a part of the day-to-day workflow of health professionals [170]. In addition, patient health often dictated availability and digital teams often floated between multiple projects simultaneously [116]. The short time frames often resulted in only low-fidelity prototypes with a limited scope of input, often making it difficult to speculate whether higher fidelity finished products would garner the same feedback [119].

Subtheme: Lack of Technological Comprehension

The previously mentioned pain point on the language of digital health co-design translated to further challenges downstream when testing ideas and prototypes due to a lack of technological comprehension (14/171, 8.2%) [14,23,27,32,36,71,74-76,116,144,145,173] among some stakeholders. Assumptions connecting personal experiences and digital health solutions were sometimes unvalidated, creating a disconnect among stakeholders [14,173]. In addition, familiarity with testing devices (eg, iPhone vs Android or extended reality interfaces) [144] was not ubiquitous. The lack of perception and clarity of objectives sometimes intimidated end users [36] and increased the difficulty in blending clinical, usability, and UX feedback in a constrained timeframe.

In summary, participatory co-design with DHI end users has highlighted challenges on the difficulty of amalgamating common interests, common language, and the collision of systematic and organic design approaches, culminating in methodological misalignments that produce varying testing approaches and hoped outcomes. This challenges the broader goal of defining a digital health framework around end-user involvement in co-design.

Research Question 3: What Are the Current Gaps in End-User Research at the Domain Level, and How Can These Be Addressed to Better Integrate End Users Into the Iterative Design Process in Digital Health, Aiming to Improve Efficacy and Uptake?

After assessing plus points and pain points in the first 2 research questions, our final question sought to understand the limitations

XSL•FO
RenderX

and gaps that are challenging the advancement of end-user incorporation in the design of DHIs.

Theme: Pragmatic Hybridized Framework

Despite positive advancements in incorporating end users in digital health co-design, the various aforementioned flash points led to advocating for a more pragmatic hybridized digital health framework (60/171, 35%). Subthemes centered around the need for a more mixed method, adaptable, and contextual approach to incorporating end users; one that is a fluid, whole-of-system digital health approach.

Subtheme: The Value of Mixed-Method Approaches to End-User Data

Numerous studies identified taking a more mixed-method approach (2/171, 14%) [26,36,40,48,50,52,57,66,72,80,81,96,104,138,141,154,155,161,174-178] as a cornerstone toward hybridizing digital and health approaches to incorporate end users. Unsurprisingly, mixed methods studies suffered less limitations in the form of validity and generalizability. Notably, studies found that there were discrepancies between what end users verbalized versus what was being observed from them [36], indicative of a mismatch between end-user desires and end-user actions. This observation strengthened the need for other forms of quantitative assessment for correlation. Another study found that pooling larger datasets through the System Usability Scale [26,141] and weighing them against smaller qualitative datasets was seen as a more prudent method of testing assumptions and weighing feedback. A study by He et al [176] found that end-user preferences for a tutorial (documented in qualitative interviews) was implemented, but they were subsequently found to be ignored by end users during quantitative data use analysis of the next iteration. Gaining metrics from hypothesis testing [37] was seen as an emotionally detached compliment to qualitative workshop approaches to gain personal insights from end-user preferences and behavior. An additional benefit to this approach was comparing data between experiment and control groups [149]. In some cases, the "hard data" metrics from mHealth app use were used to support or reject findings from interviews and observations. This was strongly exemplified in a study by Fico et al [37] in which a user-centered workshop design (ie, a qualitative approach) was complemented by bio-signal sensors to implement "a reliable information workflow." In this regard, valuable emotional feedback from end users was cross-referenced with sensory data analytics for stronger conclusions [148,177]. In essence, the pragmatic, synthesized coevaluation of quantitative and qualitative data may only serve to strengthen the reliability of studies by blending quantifiable statistics with emotional insights from end-user analysis.

Subtheme: Organic Co-Design Approaches

Building upon the value of richer mixed method approaches to evaluating end-user co-design involvement, the need for more organic co-design approaches (25/171, 14.6%) [3,9,16,38,40,46,47,50,52,57-59,61,64,65,68,81,82,89,106,116,126,141,177] was discussed in a number of studies. Organic, as in a holistic co-design approach, one that combines digital, health, and behavioral metrics into a more unified process. This is a response to the problem of misaligning usability and health care design objectives and the recognition that "end users" are not one single homogenous group [9]. Co-design workshops were found to have benefited from a funneling of methodologies and methods, whether qualitative or quantitative, that reflect digital (ie, technical and usability) and health care (ie, medical and psychoanalytical) around end-user involvement. Those studies that found this to be successful noted that it is only possible where diverse stakeholder involvement around the end users is consistent throughout the iterative design process [9,38,89]. Values such as impartiality [116] and interdisciplinarity [127,177] were identified as key elements toward hybridizing approaches. These values recognized that the needs of stakeholders and end users may be highly disparate [16]. Therefore, if acceptability [127] is a ubiquitous term that satisfies digital goals of usability, health care goals of efficacy, and end-user goals of usefulness, then refining a co-design approach that is organically suited to the needs of the digital health ecosystem is paramount.

Subtheme: Fluid Digital Health Efficacy Evaluation

The call for a more organic, digital health co-design approach was backed by calls to integrate digital and health efficacy metrics into a singular framework for evaluation (39/171, 22.8%) [9,12,14,16,34,35,39,46,47,50,52,55,56,58,61,64-66,69,71,78,80-82,92,96,99,104,106,107,118,142,146,157,170,171,175,179-181]. Various studies found a siloed approach that demarcated digital and health stakeholders was ineffective toward curating interventions. Studies recognized that design workshops did not always account for technical usability, behavior changes or patient engagement approaches that are all part of the digital health evaluation ecosystem [9,92,106,126,149,157,170,180]. Determining user motivation and confidence did not account for health system constraints [180], and determining health safety did not account for user engagement. Essentially, studies found a need to bridge [14,39,106,118,170] digital concepts such as increased usability, with health care concepts, such as reduced risk, in a workshop environment that fluidly and holistically facilitates both needs around the end user.

Subtheme: Agile Digital Health Project Facilitation

Numerous studies (18/171, 10.5%) [3,11,35,50,53,62,64,66,67,82,98,117,121,123,152,155,171,182] discussed the challenge to funnel digital and health co-design approaches into an adaptation of Agile methodology. Agile has been used in DHIs as a rapid design framework in response to legacy methods such as RCTs, which span years [11]. Agile uses short-cycle iterations (ie, sprints) with small sample sizes (ie, teams) [35] producing mainly qualitative results. However, efforts to port Agile "out of the box" into digital health have been met with challenges. While Agile presents exciting methods to incorporate end users and produce rapid emotional responses, studies have found that this qualitative analysis struggles to incorporate the rigor of health care development [121]. For example, some studies found that in Agile design sprints, patient end users were unsure of what a health solution should look like [152]. This exemplified the difficulty of entirely democratizing Agile digital health workshops. This, coupled with the small sample size of Agile teams, raised questions on validity and generalizability [98]. Findings exemplified the need

XSL•FO
RenderX

for having project leads, such as Scrum masters and product owners, who are familiar with health care contexts [182] and can effectively blend digital health design sprints that better satisfy the needs of all stakeholders. In this context, evidence-based and theory driven approaches can meet around user needs and preferences. As Schwartz et al [98] summarized in their study, "experimenting with agile adaptation in digital health will require a transparency of the development process, collecting shared difficulties and solutions toward a more standardized implementation suited to mHealth interventions."

Theme: Industry Implementability

Industry implementability (13/171, 7.6%) [3,13,16,24,31,35,36,91,106,146,153,171,177] represented downstream awareness of the need to develop a more fluid digital health hybridized design framework around end users. This is being fueled by the emerging policy shift toward the digital empowerment of patient end users. This new environment is further integrating government policy and administration, eHealth records, and encrypted end-to-end access through various collaborations between government bodies, industries, and end users (ie, citizens). Understanding the challenge to co-design for medical devices versus consumer devices was highlighted by studies that found conventional UCD processes hitting the wall of medical device regulation [31]. Nonetheless, medical regulation is part of the digital health ecosystem and needs to be considered in the ideation "design features" [36]. The needs of distributed health services were found to differ considerably with end-user preferences [16]. This reality addressed the complexity of propelling nascent stage DHI co-design toward the market. From industry suppliers to government regulatory bodies to end users, potential blockers in postdevelopment stages loomed large due to misalignment with clinical workflow and health policy [91,106]. Cooperation between public and private sectors to deliver DHIs added layers of bureaucracy [3]. In some cases, conceptual prototypes were not implementable in a given health care information technology ecosystem [153] due to security or privacy policies. This challenge exposed a gap in the business model to identify critical success factors [171] that are often missing in digital health stakeholder collaborations around end users.

Discussion

Principal Findings

This review examined the advancements and challenges of co-designing DHIs with end users, highlighting key gaps and limitations that cut through a wide range of design methodologies. The burgeoning growth of the digital health industry is a major advancement in health care delivery, one that is seeing exciting developments and major investment. Participatory co-design approaches that promote collaboration and ideation around the end users, often in situ, are promoting positive behavior change by seeking to absorb a diverse stakeholder group into a more fluid digital health ethos. These iterative design approaches are often centered around UCD methodologies and other similar approaches (eg, human-, person-, and patient- centered approaches) that emphasizes a more holistic empathetic approach, positioning the end users as coequal in the design phase (Multimedia Appendix 1). Co-design is an exploratory process that experiences divergence and convergence. Through this disruptive process, the collective brain of a diverse stakeholder group is sought out in order to resolve conflicts and to amalgamate design needs. It is hoped that through these approaches, a common language and improved cultural context may form, leading to a better co-design outcome, one that provides both better efficacy and UX.

These promising developments are tempered by the sheer complexity of combining digital and health approaches into one funneled, organic design around the end user. The rapidity of co-design workshops, with their limited cohort size, has raised many questions about the validity and reliability of findings. While an RCT is the gold standard in health care validation, its timeframe often outlives the development of a digital solution [11]. In response to this, researchers are experimenting within the nascent stages of digital health design, porting a variety of research methodologies (ie, observational and experimental, whether qualitative or quantitative) and design methods (ie, Agile and behavior change, among others), mixing and matching concepts seeking out how best to validate DHIs with end users.

Exactly *how* this incorporation occurs is the flash point of conflict. Digital design perspectives that are heavily borrowed from the mainstream digital industry are rooted in UX approaches. Health design perspectives focus on long-term safety and efficacy (Multimedia Appendix 8). While digital approaches are suited to rapid innovation, pivoting changes in real-time, health approaches are rooted in rigor and safety, implementing change slowly with longitudinal data. The variance between qualitative and quantitative approaches, UX and behavior change, and efficacy and usability accentuates the downstream challenges that occur in ideation, testing, and implementation of DHIs. This culture clash calls for pragmatism. Hybridizing digital and health approaches such that the demarcation between them eventually dissolves is the apex of DHI co-design. To better achieve this, the digital health domain must become comfortable with the uncomfortable, allowing the tensions of hierarchical health approaches to collide with flat-management digital approaches. Finding common ground between complaints of digital UX approaches that oversimplify health concerns and health approaches that are bureaucratic and time intensive will be paramount in creating an equilibrium that blends competing agendas.

Understanding the Advancements, Failures, and Challenges to Incorporate End Users in DHI Co-Design

Advancements

In response to the call for better contextual design [3] to curate better health outcomes [4] around end users, various digital, health, and behavioral approaches are being implemented in digital health. From a digital perspective, collaboration and ideation with end users is a response to traditional top-down waterfall approaches that backload testing and feedback late in the development process, when it is often too late to properly incorporate fundamental changes [15]. From a health care perspective, there is a push toward digital health citizens [183];

XSL•FO
RenderX

persons who take more control of their own health care, whether observing metrics, building their knowledge base, or proactively engaging health care professionals digitally. Our findings highlight that these 2 motivating factors are driving improvements in collaborative ideation and testing with end users in their natural environments, leading to improved behavior modification. These positive findings in participatory co-design are tunneling digital health toward a more organic design process with end users, one that is seeking to balance health care and purposeful design more fluidly. These positive developments are exemplifying theoretical concepts that advocate for more reciprocal, self-empowering approaches to co-design [184] with end users, as digital health began to emerge a decade ago. Nonetheless, this is still very much a nascent exploratory stage for digital health, one that through the process of discovery continues to uncover new challenges.

Failures and Challenges

The findings in this study indicate that participatory co-design with DHI end users has interconnected challenges that underscore the difficulty of amalgamating common interests, common language among stakeholders, and both systematic and innovative design approaches. These challenges culminate in methodological misalignments that produce variance in ideating and testing, resulting in divergent goals. These themes complicate the broader objective of defining a digital health framework around end-user involvement in co-design. Optimistically, some studies focused on the core theme of pragmatism, seeking to implement digital health solutions around end users that are fit for purpose, albeit more theoretically than experimentally. To resolve the key challenges identified in the Introduction section, namely the multiplicity of design approaches, resolving disparate understandings of the end user, and dissolving competing agendas [6,7,12,13], DHIs require frameworks that are truly representative of all aspects of digital health delivery, breaking through the common silo disconnects. From the messiness of brainstorming ideation to the effectiveness of user testing right through the product implementation and system-wide distribution, an evolution is needed, one that provides broader, richer data points to better understand end-user incorporation and to provide better validation of iterative design decisions. The need for more data points and better testing methods is supported by the fact that 112 of 171 (65.5%) studies in this review encountered failures or challenges in testing phases citing either a lack of end users or the limited diversity among them. This speaks to the lack of data impacting digital health design decisions and its overarching philosophies.

The complexity of bringing DHIs to market cannot be accomplished solely with small, often speculative user studies. Downstream implementation and policy concerns can render years of research fruitless. New frameworks must emerge that are holistic and consider a global outlook on digital health distribution. Doing so may require evolving the digital health ecosystem by dissolving the tribalism of qualitative versus quantitative approaches, instead mixing methods pragmatically to better account for the sheer breadth of digital health success factors. Theoretically, hybridizing the 3 pillars—digital (eg, experience and usability), health (eg, medical and psychoanalytical), and business (eg, critical success factors)—may be crucial for better collectivizing design insights, improving industry implementability, and reducing downstream failures in product market fit. Practically, there is a need for co-designing frameworks and platforms that provide rapidly accessible, contextual, and demographic data to better justify iterative design decisions around end users. The end user is a person, a patient, a user, and a consumer. Their incorporation spans many personas and multiple user journeys that beget numerous types of data points relative to the evaluation of a DHI, from preconception to postdelivery; a “whole of system” co-design approach around health end users will be paramount to the future.

On the basis our findings (Figure 3), we propose the core recommendations illustrated in subsequent sections.

XSL•FO
RenderX

Figure 3. Mind map of key challenges and proposed solutions. DHI: digital health intervention.

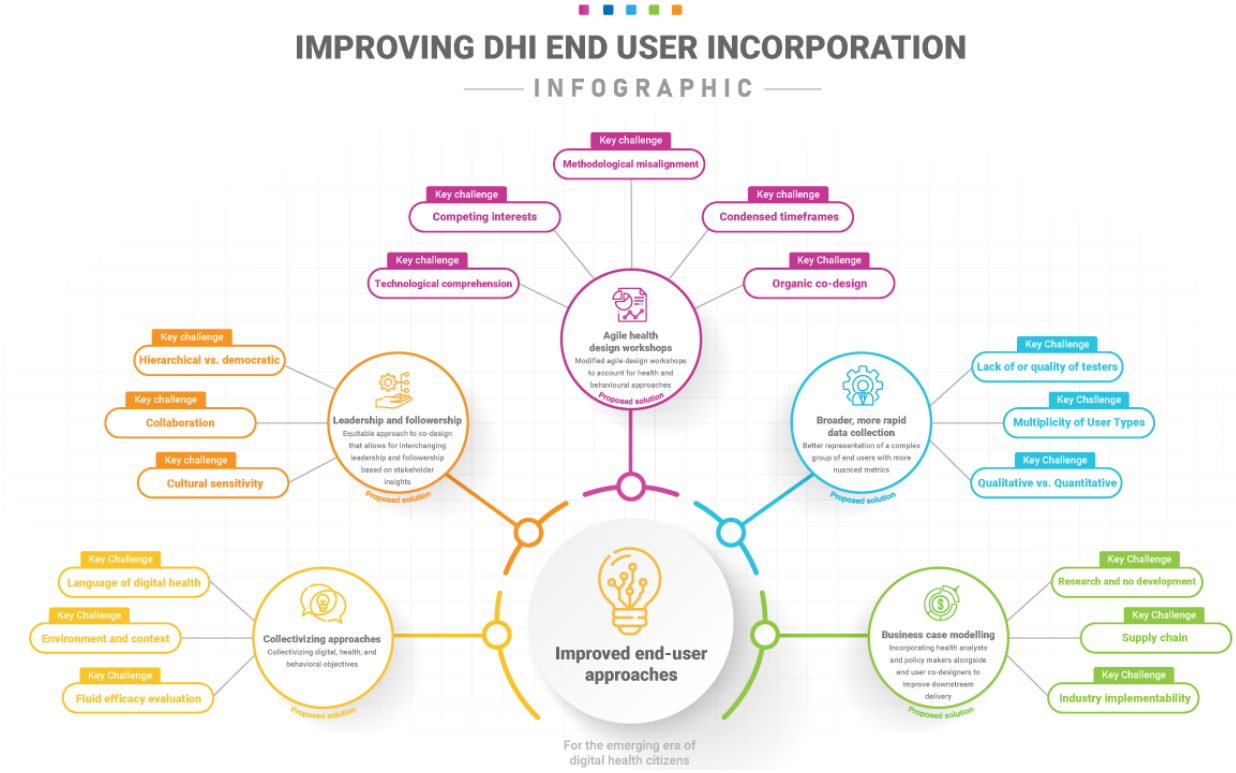

Recommendation 1: Agile Health Design Workshops

Considering that Agile methodology is a project facilitation approach conceptualized outside of health care, unsurprisingly it has faced challenges when ported out of the box. To combine lightweight, rapid, qualitative Agile methods with more rigorous quantitative health methods, we recommend Agile workshops that triangulate emotional and observational feedback with statistical biosensor and analytical feedback that may soothe concerns about validity, generalizability, and attrition. This topic in itself could spawn a study on a mixed methods Agile framework design for digital health. Building on the calls for more Agile efficient approaches to digital health co-design suggested by Peiris et al [185] and Wilson et al [186], we propose the recommendations outlined in Textbox 2.

Textbox 2. Agile efficient approaches to digital health co-design.

- Involving digital health agilists who have an overarching and empathetic view of the needs of digital, health, and end-user stakeholders. Coordinators of digital health projects (ie, Scrum masters and product owners) need to align the collective goals that are unique to digital health, ones that may not be solely reflected in traditional user experience design workshops.
- A preiteration sprint (ie, sprint 0) that involves creating shared digital health milestones and backlog grooming that combines the goals of health stakeholders (eg, clinical assessment, safety, and efficacy), digital stakeholders (eg, usability and technical validation), and end users (eg, user experience and acceptability). This preliminary approach may also account for discrepancies in technical ability and comprehension, as well as cultural sensitivities.
- From this, the selection of co-design activities that seek to combine these goals into an organic digital health workshop process may occur [171]. These activities should acquire metrics, whether qualitative or quantitative, that pragmatically fit the co-design objectives of a digital health team.
- Subsequent sprint retrospectives and sprint planning that continuously engages and reevaluates the collective goals of the digital health objectives and milestones, merging user experience validators with health and behavioral assessments.

Using the analogy of the toolbox, much like a carpenter must be pragmatic when selecting the given tools for the task at hand, digital health workshops must use a similar logic by hybridizing the methodologies, methods, and metrics that are fit for purpose [121]. Standardizing this process has proven difficult [98]. Shoehorning standard Agile practices into digital health continues to prove challenging, most notably in the struggle to validate studies with small participant groups in health care [104]. Therefore, there is a need to produce rapid quantifiable methods within the Agile design cycle, which points toward crowdsourcing ideation and testing data at scale. Charting a path that is reflective of collective co-designed digital health goals may reduce previously observed methodological misalignment, improving the flow of the co-design and building toward a more organic set of collective hoped outcomes. We do not suggest this as a magic elixir or standardized framework but as a flexible Agile health toolkit that seeks to amalgamate digital health approaches around the end user in the Agile workflow. This process must be tailored, balancing what participants say, do, imagine, test, and evolve, without being exhaustive [127]. As Schwartz et al [98] summarized in their study, experimenting with Agile adaptation in digital health

XSL•FO
RenderX

will require a transparency of the development process, collecting shared difficulties and solutions toward a more standardized implementation suited to mHealth interventions. As a result, this may reduce downstream challenges of divergent objectives, hoped outcomes and failed implementation.

Recommendation 2: Account for the Multiplicity of End-User Types With Broader, More Rapid Data Collection

While traditional UCD workshops hone in on the end user as a single entity, DHIs are often more complex, with multiple types of end users. With 112 of 171 studies (65.5%) in this review citing challenges in either the number or variance of end users involved, recent digital health studies highlight the drawbacks of traditional UCD workshops, particularly in health care, where the "user (ie, patient) does not always have answers" [65] to clinical interventions [56]. The oversimplification of what a digital health end user is may deeply affect the integrity of the co-design. Considering that an mHealth app for heart medication may be used for tracking medication use; analytical data may serve multiple end users (eg, patient, caregiver, pharmacist, and clinician). Each type of end user has different goals, user experiences, and hoped outcomes. This reality was summarized well by Honary et al [139] where they keenly recognized that relatives of mHealth end users have different roles and responsibilities in the DHI. This growing pain in digital health is coming to the forefront as a study by Mak et al [48] advocated for mixed methods approaches that can better substantiate findings in diverse populations. To address this challenge and to account for the complexity of user types, broader pools of co-design data are needed; data that drill down on end-user demographics to create personas from rapid database-driven results. In this regard, a study by Tyo and Desroches [96] has experimented with the quantification of user personas to analyze clusters to better inform design requirements. These emerging studies speak to the need for quantifiable digital co-design platforms that can be deployed at scale.

Recommendation 3: Implement a Leadership and Followership Approach to Co-Design Workshops

A key flash point in the co-design process is the concept of democratic versus hierarchical approaches. The bottom-up (ie, shared and innovative) approach of design workshops is diametrically opposed to the top-down (ie, systematic and preplanned) approach of health care. In our findings, the high value of "equal coleadership" [6,14,127] was often championed as a guiding light in qualitative workshops. This has advantages and disadvantages, such that the democratic process brings many voices to the table, but this same process can eliminate evidence-based research from the solution [58]. Where criticism may exist of clinician-led studies being narrow minded, entirely democratic qualitative workshops can also be challenged for simplemindedness. For example, when a patient insists on a design preference [157] that conflicts with a safety issue backed up by years of quantifiable empirical data collected by clinicians. As Mirkovic et al [12] found, there is a need for equity over equality when it comes to workshop design with experts and end users. Undoubtedly, the value of a diverse array of stakeholders is of high importance; however, considering them all coequal through all stages of the design process does not serve to strengthen the design around the end user, rather it may weaken it if popular is preferred over practical, or entertainment over effectiveness. The plain democratization of the design process may lead to safety and efficacy concerns being voted off the island so to speak. Similarly, patient concerns with UX could be demoted since they are not central to behavior change or health policy goals.

The process of divergence and convergence in design workshops can instead benefit from a leadership and followership approach, such that design stakeholders lead the facilitation of UX objectives, health stakeholders lead the facilitation of safety and efficacy objectives, and patient stakeholders may lead discussion on their experiences and lifestyles [67,93,177]. In each example, the collaborative approach is central; however, there is identifiable leadership and followership such that a given expertise is provided enough ownership [101] to guide the workshop forward more effectively. This workshop approach may shift the perspective from *coequals* to *coleaders*, jointly seeking to learn from one another in a collaborative space. The goal of this approach would be to better balance person-centeredness, clinical acceptability, and UX [6] without eliminating expertise through a democratic design process.

Recommendation 4: Collectivizing Digital, Health, and Behavioral Objectives

A more pragmatic flexible framework that better understands the complexity of end-user types, one that seeks to blend the hoped outcomes of a disparate group of stakeholders, by necessity may benefit from collectivizing digital, health, and behavioral metrics and goals. Of the 171 studies, 70 (40.9%) studies listed various co-design challenges in collaboration and alignment. As the study by Ledel et al [7] found, both evidence-based (ie, health) and experience-based (ie, digital) approaches have merit for the end user, but friction exists when aligning these objectives. To reduce analysis in a vacuum and to align goals collectively, we suggest workshop co-design activities that blend objectives, such as safety (ie, health based) and usability (ie, digital based) can be blended into safe usability; behavior modification (ie, health based) and UX (ie, digital based) into user behavior; and efficacy (ie, health based) and uptake (ie, digital based) into effective uptake.

If the digital health end product is a fusion of digitally delivered health objectives, then the co-design process should suitably reflect this both philosophically and practically, aiming to reduce friction and combine goals and milestones. As Donovan et al [58] summarized in their study, "DHIs do not exist in isolation; they are integrated into existing clinical pathways." Considering that stakeholder disengagement and disparity in hoped outcomes is a frequent challenge, the aim of designing interviews, observations, and testing, whether qualitative or quantitative, that blend approaches would be to increase the fluidity of the co-design process with end users. This fluidity may reduce the challenge of balancing the innovation equilibrium; letting designers innovate and health stakeholders evaluate, with patient experiences provide the link between the two. This environment can better accommodate clinical reality, incorporating biomedical knowledge, clinical workflows, and overarching

XSL•FO
RenderX

health organizational requirements [35] around Agile practices that improve user collaboration, ideation, and rapid testing. This is very much an experimental process, but it begins with a change in thinking that can prevent implementation challenges downstream.

Recommendation 5: Include Business Case Modeling in the Co-Design Process for Downstream Implementability

Our final recommendation looks forward toward the downstream challenges of industry implementation and scalability. Rarely, in any of the studies observed, did we find co-design workshops that included health care policy or business stakeholders who addressed the clinical workflow challenges of delivering DHIs at scale to end users. Only 13 of 171 (7.6%) studies described the challenge of industry implementability, yet presumably every DHI project seeks to bring a viable product to market. In 2018, Hetrick et al [16] found that despite the success of their mHealth youth co-design, barriers exist toward the implementation of the app in clinical care settings that factor various service and governance factors. This identified the need to include health care analysts and policy makers in the co-design phase. Elements of public and private sector collaboration are interwoven into the fabric of health care service delivery that have their own set of needs that uniquely affect the constraints of a DHI. This factor needs to be incorporated into the co-design process to prevent blue-sky prototypes being discarded. A study by Vial et al [61] called for a transdisciplinary approach that includes business model definition coaching as a complement to the co-design team [61]. This accountability [15] recognizes the final leg of a DHI, the health care delivery platform. The inability to recognize this prevents scalability [12] and does not factor the ever-increasing integration of DHIs with government objectives to deliver ubiquitous services at scale.

Preparing for the Era of Digital Health Citizens

These 5 recommendations are responses to the challenges identified in this review, with an eye toward the emerging era of digital health citizenry, in which techno-sociality and bio-sociality are undemarcating the boundaries of digital health [187]. This “health citizen” agenda envisions digital health as more of a public good that provides tools, opportunities, responsibilities, and rights [183]. This transformation is built upon the prosumerist nature of today’s technologies, such that any given DHI may be an end-user experience of patients, clinicians, caregivers, therapists, and passive observers collectively, facilitated through algorithmically driven big data solutions (ie, artificial intelligence and machine learning) that are intuitive, customizable, and adaptable. The delivery of health care is increasingly blending government (eg, health records and policy), medical (eg, public and private sector solutions), and citizens (eg, end users) into one fluid infrastructure. This relationship is becoming increasingly reciprocal, more productive than facilitative in nature [187]. The potential for crowdsourcing data in tomorrow’s era of digital health seems boundless.

Needless to say, this vision needs to be brought to the forefront of the messier, creative stages of co-design workshops. The reliance on small qualitative datasets does not leverage the vastness of crowdsourcing end-user feedback that is equally capable of collecting quantifiable, demographically contextual statistics, while garnering emotional qualitative data in the form of asynchronous digital co-design platforms. If there is one word to define the evolution of digital health moving forward, it may be pragmatism. Leveraging the right tools to incorporate end users in the design of DHIs should mean removing any boundaries to the type of data collected. This means removing any allegiances to particular schools of thought on qualitative or quantitative, observational, or statistical, instead focusing on a pragmatic “patient activation” [187], approach to garner positive involvement from end users at scale. There is no need for competing interests methodologically; a combination of metrics that increases the understanding of digital health users only serves to strengthen the confidence of design solutions, reducing attrition rates and increasing validity and generalizability. The digital health citizen is envisioned not as a user, an expert, or a novice, but as an active participant in the health care ecosystem, one who constantly learns and adapts to their health care needs. It is fluid and proactive rather than static and receptive. Therefore, there is a clear need to move the incorporation of end users in DHI co-design from small participant workshops objective to a dynamic socio-digital platform of ideating, testing, and evaluating in real time, at scale. Looking forward, this approach involves a new outlook that leverages the vision of the digital health citizen methodology, with the pragmatism of Agile tools, and the rigors of health care metrics. Experimenting with processes and tools that expand the range of data collection may better identify end-user types, better triangulate preferences and challenges, and provide stronger integration with health policies and implementation objectives downstream. Refocusing co-design approaches within this lens serves to better use the emerging potential of big data, potentially increasing the user perspectives of all stakeholders across the digital health spectrum, without sacrificing the rapidity and innovation of the design environment.

Strengths and Limitations

This systematic review drew from a large swath of studies, using various methodologies and techniques to incorporate end users in DHIs. The strength of this approach was identifying a large number of studies for analysis and highlighting key global issues that digital health stakeholders face, including points of failure and challenges. We used the PerSPEcTiF framework due to its suitability to qualitative studies in health care. However, many of these studies were limited in the number of participants involved, bringing the validity and reliability of findings from these studies into question. The largely qualitative approach to studies selected, with vast variance in data collection approaches, did not allow for sensitivity analysis on the basis of measurement criteria (ie, setting, duration, and demographics).

A limitation also exists in the generalizability of findings toward specific scenarios (ie, mental health and patient aids) in which domain-specific usages weigh heavily on the context of end-user incorporation. We recommend further research to investigate these themes as subsets within specific digital health

XSL•FO
RenderX

communities (ie, mental health and serious gaming) In addition, in many cases, studies did not provide perspectives on risk of harm when using approaches to incorporate end users. This issue requires critical analysis itself. With digital health being a hybrid domain, there is no standardized approach to quality assessment, leaving the potential for risk of bias. These factors complicate data extraction and interpretation. This may be improved by using a health-centric qualitative framework in such studies, such as Confidence in the Evidence from Reviews of Qualitative Research [188].

Conclusions

The challenge to incorporate end users in the co-design of DHIs is complicated by a variety of methodological influences on problem definition, testing approaches, and hoped outcomes. The porting of legacy Agile, behavioral, and analytical methods has often resulted in conflicts that weaken conclusions that can be drawn from the co-design. In this study, we sought to draw a line through the common pain points and gaps in the literature that are omnipresent regardless of the design approach. The subsequent recommendations stemming from this systematic review serve as guidelines that may help soothe the path toward the big data environment that is emerging through a socio-technical digital health citizen approach into the mid-21st century.

Acknowledgments

SM received funding from Simon Fraser University Community Trust Endowment Fund and AGE-WELL, Canada's technology and aging network. AD received funding from the Graduate Dean's Entrance Scholarship at the School of Interactive Arts and Technology at Simon Fraser University. The authors wish to thank Ms Adena Brons, liaison librarian: interactive arts and technology, Simon Fraser University Library, for guiding our search strategy.

Data Availability

All data generated or analyzed during this study are included in this published manuscript and its supplementary information files.

Authors' Contributions

The study protocol was conceived by AD, GC, and SM. AD prepared and drafted the manuscript. All authors provided input into the design, edits, and revision of the manuscript. All authors read and approved the final manuscript. AD is the guarantor of the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A Venn diagram of the interdisciplinary intersection of digital health interventions.
[PDF File (Adobe PDF File), 826 KB - jmir_v27i1e50178_app1.pdf]

Multimedia Appendix 2
End users in design of digital health interventions: search string queries.
[XLSX File (Microsoft Excel File), 69 KB - jmir_v27i1e50178_app2.xlsx]

Multimedia Appendix 3
End users in design of digital health interventions: screening template (2015 to 2021).
[XLSX File (Microsoft Excel File), 973 KB - jmir_v27i1e50178_app3.xlsx]

Multimedia Appendix 4
End users in design of digital health interventions: screening template (2022 to May 31, 2024).
[XLSX File (Microsoft Excel File), 7325 KB - jmir_v27i1e50178_app4.xlsx]

Multimedia Appendix 5
End users in design of digital health interventions: data extraction template for categorization of themes and codes.
[XLSX File (Microsoft Excel File), 225 KB - jmir_v27i1e50178_app5.xlsx]

Multimedia Appendix 6

XSL•FO
RenderX

End users in design of digital health interventions: raw data content (ie, quotes and reflective analysis) extracted from eligible studies in the review.
[PDF File (Adobe PDF File), 1075 KB - jmir_v27i1e50178_app6.pdf]

Multimedia Appendix 7
Characteristics of included end-user studies.
[XLSX File (Microsoft Excel File), 67 KB - jmir_v27i1e50178_app7.xlsx]

Multimedia Appendix 8
Comparison of evaluation in digital and health-based perspectives.
[PDF File (Adobe PDF File), 705 KB - jmir_v27i1e50178_app8.pdf]

Multimedia Appendix 9
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 118 KB - jmir_v27i1e50178_app9.pdf]

References

1. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
2. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
3. Ospina-Pinillos L, Davenport TA, Navarro-Mancilla AA, Cheng VW, Cardozo Alarcón AC, Rangel AM, et al. Involving end users in adapting a Spanish version of a web-based mental health clinic for young people in Colombia: exploratory study using participatory design methodologies. JMIR Ment Health 2020 Feb 08;7(2):e15914 [FREE Full text] [doi: 10.2196/15914] [Medline: 32027313]
4. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
5. Ector GI, Westerweel PE, Hermens RP, Braspenning KA, Heeren BC, Vinck OM, et al. The development of a web-based, patient-centered intervention for patients with chronic myeloid leukemia (CMyLife): design thinking development approach. J Med Internet Res 2020 May 15;22(5):e15895 [FREE Full text] [doi: 10.2196/15895] [Medline: 32412424]
6. Kerkhof Y, Pelgrum-Keurhorst M, Mangiaracina F, Bergsma A, Vrauwdeunt G, Graff M, et al. User-participatory development of FindMyApps; a tool to help people with mild dementia find supportive apps for self-management and meaningful activities. Digit Health 2019 Mar 26;5:2055207618822942 [FREE Full text] [doi: 10.1177/2055207618822942] [Medline: 30944726]
7. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KM, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of EPIO. J Med Internet Res 2020 Jan 21;22(1):e15889 [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]
8. Arvidsson S, Gilljam BM, Nygren J, Ruland CM, Nordby-Bøe T, Svedberg P. Redesign and validation of Sisom, an interactive assessment and communication tool for children with cancer. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e76 [FREE Full text] [doi: 10.2196/mhealth.5715] [Medline: 27343004]
9. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
10. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022 Sep 09;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
11. van Kessel K, Babbage DR, Kersten P, Drown J, Sezier A, Thomas PW, et al. Design considerations for a multiple sclerosis fatigue mobile app MS Energize: a pragmatic iterative approach using usability testing and resonance checks. Internet Interv 2021 Apr;24:100371 [FREE Full text] [doi: 10.1016/j.invent.2021.100371] [Medline: 33614414]
12. Mirkovic J, Jessen S, Kristjansdottir OB, Krogseth T, Koricho AT, Ruland CM. Developing technology to mobilize personal strengths in people with chronic illness: positive codesign approach. JMIR Form Res 2018 Jun 05;2(1):e10774 [FREE Full text] [doi: 10.2196/10774] [Medline: 30684404]
13. Ravn Jakobsen P, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. Int J Environ Res Public Health 2018 Feb 13;15(2):330 [FREE Full text] [doi: 10.3390/ijerph15020330] [Medline: 29438343]

XSL•FO
RenderX

14. Latulippe K, Hamel C, Giroux D. Co-design to support the development of inclusive eHealth tools for caregivers of functionally dependent older persons: social justice design. J Med Internet Res 2020 Nov 09;22(11):e18399 [FREE Full text] [doi: 10.2196/18399] [Medline: 33164905]
15. Patel D, Sarlati S, Martin-Tuite P, Feler J, Chehab L, Texada M, et al. Designing an information and communications technology tool with and for victims of violence and their case managers in San Francisco: human-centered design study. JMIR Mhealth Uhealth 2020 Aug 24;8(8):e15866 [FREE Full text] [doi: 10.2196/15866] [Medline: 32831179]
16. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
17. Duffy A, Christie G, Moreno S. Examining challenges to the incorporation of end users in the design of digital health interventions: protocol for a systematic review. JMIR Res Protoc 2021 Jul 26;10(7):e28083 [FREE Full text] [doi: 10.2196/28083] [Medline: 34309578]
18. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd edition. Hoboken, NJ: John Wiley & Sons; 2019.
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
20. Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. BMJ Glob Health 2019 Jan 25;4(Suppl 1):e001107 [FREE Full text] [doi: 10.1136/bmjgh-2018-001107] [Medline: 30775019]
21. Moon D. Network traffic characteristics and analysis in recent mobile games. Appl Sci 2024 Feb 08;14(4):1397 [FREE Full text] [doi: 10.3390/app14041397]
22. Boyd AD, Moores K, Shah V, Sadhu E, Shroff A, Groo V, et al. My interventional drug-eluting stent educational app (MyIDEA): patient-centered design methodology. JMIR Mhealth Uhealth 2015 Jul 02;3(3):e74 [FREE Full text] [doi: 10.2196/mhealth.4021] [Medline: 26139587]
23. Abujarad F, Alfano S, Bright TJ, Kannoth S, Grant N, Gueble M, et al. Building an informed consent tool starting with the patient: the patient-centered virtual multimedia interactive informed consent (VIC). AMIA Annu Symp Proc 2017;2017:374-383 [FREE Full text] [Medline: 29854101]
24. Brox E, Konstantinidis ST, Evertsen G. User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]
25. Gabrielli S, Dianti M, Maimone R, Betta M, Filippi L, Ghezzi M, et al. Design of a mobile app for nutrition education (TreC-LifeStyle) and formative evaluation with families of overweight children. JMIR Mhealth Uhealth 2017 Apr 13;5(4):e48 [FREE Full text] [doi: 10.2196/mhealth.7080] [Medline: 28408361]
26. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. JMIR Mhealth Uhealth 2017 May 30;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
27. Allin S, Shepherd J, Tomasone J, Munce S, Linassi G, Hossain SN, et al. Participatory design of an online self-management tool for users with spinal cord injury: qualitative study. JMIR Rehabil Assist Technol 2018 Mar 21;5(1):e6 [FREE Full text] [doi: 10.2196/rehab.8158] [Medline: 29563075]
28. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018 Sep 06;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
29. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
30. Ahmed R, Toscos T, Rohani Ghahari R, Holden RJ, Martin E, Wagner S, et al. Visualization of cardiac implantable electronic device data for older adults using participatory design. Appl Clin Inform 2019 Aug 18;10(4):707-718 [FREE Full text] [doi: 10.1055/s-0039-1695794] [Medline: 31533172]
31. Birnie KA, Campbell F, Nguyen C, Lalloo C, Tsimicalis A, Matava C, et al. iCanCope PostOp: user-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents. JMIR Form Res 2019 Apr 22;3(2):e12028 [FREE Full text] [doi: 10.2196/12028] [Medline: 31008704]
32. Gonsalves PP, Hodgson ES, Kumar A, Aurora T, Chandak Y, Sharma R, et al. Design and development of the " POD adventures" smartphone game: a blended problem-solving intervention for adolescent mental health in india. Front Public Health 2019 Aug 23;7:238 [FREE Full text] [doi: 10.3389/fpubh.2019.00238] [Medline: 31508404]
33. Jefferson UT, Zachary I, Majee W. Employing a user-centered design to engage mothers in the development of a mHealth breastfeeding application. Comput Inform Nurs 2019 Oct;37(10):522-531. [doi: 10.1097/CIN.0000000000000549] [Medline: 31414995]

XSL•FO
RenderX

34. Aufegger L, Bùi KH, Bicknell C, Darzi A. Designing a paediatric hospital information tool with children, parents, and healthcare staff: a UX study. BMC Pediatr 2020 Oct 08;20(1):469 [FREE Full text] [doi: 10.1186/s12887-020-02361-w] [Medline: 33032549]
35. Cornet VP, Toscos T, Bolchini D, Rohani Ghahari R, Ahmed R, Daley C, et al. Untold stories in user-centered design of mobile health: practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17703 [FREE Full text] [doi: 10.2196/17703] [Medline: 32706745]
36. Davies A, Mueller J, Hennings J, Caress A, Jay C. Recommendations for developing support tools with people suffering from chronic obstructive pulmonary disease: co-design and pilot testing of a mobile health prototype. JMIR Hum Factors 2020 May 15;7(2):e16289 [FREE Full text] [doi: 10.2196/16289] [Medline: 32410730]
37. Fico G, Martinez-Millana A, Leuteritz JP, Fioravanti A, Beltrán-Jaunsarás ME, Traver V, et al. User centered design to improve information exchange in diabetes care through eHealth : results from a small scale exploratory study. J Med Syst 2019 Nov 18;44(1):2 [FREE Full text] [doi: 10.1007/s10916-019-1472-5] [Medline: 31741069]
38. Jessen S, Mirkovic J, Nes LS. MyStrengths, a strengths-focused mobile health tool: participatory design and development. JMIR Form Res 2020 Jul 24;4(7):e18049 [FREE Full text] [doi: 10.2196/18049] [Medline: 32706651]
39. Slater H, Stinson JN, Jordan JE, Chua J, Low B, Lalloo C, et al. Evaluation of digital technologies tailored to support young people's self-management of musculoskeletal pain: mixed methods study. J Med Internet Res 2020 Jun 05;22(6):e18315 [FREE Full text] [doi: 10.2196/18315] [Medline: 32442143]
40. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Development and evaluation of the usefulness, usability, and feasibility of iNNOV breast cancer: mixed methods study. JMIR Cancer 2022 Feb 15;8(1):e33550 [FREE Full text] [doi: 10.2196/33550] [Medline: 35166682]
41. Liu S, La H, Willms A, Rhodes RE. A "No-Code" app design platform for mobile health research: development and usability study. JMIR Form Res 2022 Aug 18;6(8):e38737 [FREE Full text] [doi: 10.2196/38737] [Medline: 35980740]
42. Adler RF, Morales P, Sotelo J, Magasi S. Developing an mHealth app for empowering cancer survivors with disabilities: co-design study. JMIR Form Res 2022 Jul 26;6(7):e37706 [FREE Full text] [doi: 10.2196/37706] [Medline: 35881439]
43. Sunjaya AP, Martin A, Jenkins C. A design thinking approach to developing a clinical decision support system for breathlessness in primary care. Stud Health Technol Inform 2022 Jun 06;290:839-843. [doi: 10.3233/SHTI220197] [Medline: 35673136]
44. Khowaja K, Waheeda Syed W, Singh M, Taheri S, Chagoury O, Al-Thani D, et al. A participatory design approach to develop visualization of wearable actigraphy data for health care professionals: case study in Qatar. JMIR Hum Factors 2022 Apr 08;9(2):e25880 [FREE Full text] [doi: 10.2196/25880] [Medline: 35394442]
45. Preuhs K, van Keulen H, Andree R, Wins S, van Empelen P. A tailored web-based video intervention (ParentCoach) to support parents with children with sleeping problems: user-centered design approach. JMIR Form Res 2022 Apr 19;6(4):e33416 [FREE Full text] [doi: 10.2196/33416] [Medline: 35438640]
46. Moltrecht B, Patalay P, Bear HA, Deighton J, Edbrooke-Childs J. A transdiagnostic, emotion regulation app (Eda) for children: design, development, and lessons learned. JMIR Form Res 2022 Jan 19;6(1):e28300 [FREE Full text] [doi: 10.2196/28300] [Medline: 35044312]
47. Morse B, Soares A, Kwan BM, Allen M, Lee RS, Desanto K, et al. A transgender health information resource: participatory design study. JMIR Hum Factors 2023 Jun 15;10:e42382 [FREE Full text] [doi: 10.2196/42382] [Medline: 37318836]
48. Mak WW, Ng SM, Leung FH. A web-based stratified stepped care platform for mental well-being (TourHeart+): user-centered research and design. JMIR Form Res 2023 Mar 22;7:e38504 [FREE Full text] [doi: 10.2196/38504] [Medline: 36947112]
49. Brown M, Lord E, John A. Adaptation of ACTivate your wellbeing, a digital health and well-being program for young persons: co-design approach. JMIR Form Res 2023 Apr 13;7:e39913 [FREE Full text] [doi: 10.2196/39913] [Medline: 37052994]
50. Yusupov I, Vandermorris S, Plunkett C, Astell A, Rich JB, Troyer AK. An agile development cycle of an online memory program for healthy older adults. Can J Aging 2022 Dec;41(4):647-656. [doi: 10.1017/S0714980821000763] [Medline: 35256025]
51. Mueller EL, Cochrane AR, Campbell ME, Nikkhah S, Miller AD. An mHealth app to support caregivers in the medical management of their child with cancer: co-design and user testing study. JMIR Cancer 2022 Mar 16;8(1):e33152 [FREE Full text] [doi: 10.2196/33152] [Medline: 35293867]
52. Lawrence K, Cho J, Torres C, Alfaro-Arias V. Building virtual health training tools for residents: a design thinking approach. Front Digit Health 2022 Jun 13;4:861579 [FREE Full text] [doi: 10.3389/fdgth.2022.861579] [Medline: 35770138]
53. Fox S, Brown LJ, Antrobus S, Brough D, Drake RJ, Jury F, et al. Co-design of a smartphone app for people living with dementia by applying agile, iterative co-design principles: development and usability study. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e24483 [FREE Full text] [doi: 10.2196/24483] [Medline: 35029539]
54. O'Connor A, Tai A, Brinn M, Thuc Hien Hoang N, Cataldi D, Carson-Chahhoud K. Co-design of an augmented reality asthma inhaler educational intervention for children: development and usability study. JMIR Pediatr Parent 2023 Jul 25;6:e40219 [FREE Full text] [doi: 10.2196/40219] [Medline: 37490325]

XSL•FO
RenderX

55. Taxter A, Johnson L, Tabussi D, Kimura Y, Donaldson B, Lawson E, et al. Co-design of an electronic dashboard to support the coproduction of care in pediatric rheumatic disease: human-centered design and usability testing. J Particip Med 2022 Apr 22;14(1):e34735 [FREE Full text] [doi: 10.2196/34735] [Medline: 35133283]
56. Liang Z, Melcer E, Khotchasing K, Hoang NH. Co-design personal sleep health technology for and with university students. Front Digit Health 2024 Apr 9;6:1371808 [FREE Full text] [doi: 10.3389/fdgth.2024.1371808] [Medline: 38655450]
57. Ekstedt M, Kirsebom M, Lindqvist G, Kneck Å, Frykholm O, Flink M, et al. Design and development of an eHealth service for collaborative self-management among older adults with chronic diseases: a theory-driven user-centered approach. Int J Environ Res Public Health 2021 Dec 30;19(1):391 [FREE Full text] [doi: 10.3390/ijerph19010391] [Medline: 35010652]
58. Donovan G, Hall N, Smith F, Ling J, Wilkes S. Two-way automated text messaging support from community pharmacies for medication taking in multiple long-term conditions: human-centered design with nominal group technique development study. JMIR Form Res 2022 Dec 21;6(12):e41735 [FREE Full text] [doi: 10.2196/41735] [Medline: 36542458]
59. Pfeuffer N, Beyer A, Penndorf P, Leiz M, Radicke F, Hoffmann W, et al. Evaluation of a health information exchange system for geriatric health care in rural areas: development and technical acceptance study. JMIR Hum Factors 2022 Sep 15;9(3):e34568 [FREE Full text] [doi: 10.2196/34568] [Medline: 36107474]
60. Lundell S, Toots A, Sönnerfors P, Halvarsson A, Wadell K. Participatory methods in a digital setting: experiences from the co-creation of an eHealth tool for people with chronic obstructive pulmonary disease. BMC Med Inform Decis Mak 2022 Mar 18;22(1):68 [FREE Full text] [doi: 10.1186/s12911-022-01806-9] [Medline: 35303895]
61. Vial S, Boudhraâ S, Dumont M, Tremblay M, Riendeau S. Developing a mobile app with a human-centered design lens to improve access to mental health care (mentallys project): protocol for an initial co-design process. JMIR Res Protoc 2023 Aug 22;12:e47220 [FREE Full text] [doi: 10.2196/47220] [Medline: 37606978]
62. Tessarolo F, Petsani D, Conotter V, Nollo G, Conti G, Nikolaidou M, et al. Developing ambient assisted living technologies exploiting potential of user-centred co-creation and agile methodology: the CAPTAIN project experience. J Ambient Intell Human Comput 2022 Apr 09;15(5):2661-2676 [FREE Full text] [doi: 10.1007/S12652-021-03649-0]
63. McCann L, Hewitt C, McMillan KA. Developing an e-prehabilitation system of care for young adults diagnosed with cancer: user-centered design study. JMIR Cancer 2023 Mar 30;9:e41441 [FREE Full text] [doi: 10.2196/41441] [Medline: 36995740]
64. Arigo D, Lobo AF, Ainsworth MC, Baga K, Pasko K. Development and initial testing of a personalized, adaptive, and socially focused web tool to support physical activity among women in midlife: multidisciplinary and user-centered design approach. JMIR Form Res 2022 Jul 26;6(7):e36280 [FREE Full text] [doi: 10.2196/36280] [Medline: 35881431]
65. Schwartz-Lasfargues C, Roux-Gendron C, Edomskis P, Marque I, Bayon Y, Lange JF, et al. Development of a connected sensor system in colorectal surgery: user-centered design case study. JMIR Hum Factors 2022 Jul 08;9(3):e31529 [FREE Full text] [doi: 10.2196/31529] [Medline: 35802406]
66. Turesson C, Liedberg G, Björk M. Development of a digital support application with evidence-based content for sustainable return to work for persons with chronic pain and their employers: user-centered agile design approach. JMIR Hum Factors 2022 Mar 14;9(1):e33571 [FREE Full text] [doi: 10.2196/33571] [Medline: 35285814]
67. Østervang C, Lassen A, Schmidt T, Coyne E, Dieperink KB, Jensen CM. Development of a health information system to promote emergency care pathways: a participatory design study. Digit Health 2022 Dec 26;8:20552076221145856 [FREE Full text] [doi: 10.1177/20552076221145856] [Medline: 36601282]
68. Ghorayeb A, Comber R, Gooberman-Hill R. Development of a smart home interface with older adults: multi-method co-design study. JMIR Aging 2023 Jun 16;6:e44439 [FREE Full text] [doi: 10.2196/44439] [Medline: 37327037]
69. Johansen SK, Kanstrup AM, Thomsen JL, Christensen MN, Rathleff MS. Exploring the barriers and facilitators for supporting adolescents with knee pains adherence to mobile health apps: a think-aloud study. Digit Health 2023 Oct 19;9:20552076231205750 [FREE Full text] [doi: 10.1177/20552076231205750] [Medline: 37868153]
70. Weber P, Mahmood F, Ahmadi M, Von Jan V, Ludwig T, Wieching R. Fridolin: participatory design and evaluation of a nutrition chatbot for older adults. i-com 2023;22(1):33-51. [doi: 10.1515/icom-2022-0042]
71. Holmqvist M, Ros A, Lindenfalk B, Thor J, Johansson L. How older persons and health care professionals co-designed a medication plan prototype remotely to promote patient safety: case study. JMIR Aging 2023 Apr 07;6:e41950 [FREE Full text] [doi: 10.2196/41950] [Medline: 37027205]
72. Woodward M, Dixon-Woods M, Randall W, Walker C, Hughes C, Blackwell S, Thiscovery Authorship Group, et al. How to co-design a prototype of a clinical practice tool: a framework with practical guidance and a case study. BMJ Qual Saf 2024 Mar 25;33(4):258-270 [FREE Full text] [doi: 10.1136/bmjqs-2023-016196] [Medline: 38124136]
73. Isangula K, Pallangyo ES, Ndirangu-Mugo E. Interventions co-designed by healthcare providers and clients for improving therapeutic relationships in maternal and child healthcare: a pilot study using human centered design in rural Tanzania. BMC Nurs 2023 Sep 14;22(1):314 [FREE Full text] [doi: 10.1186/s12912-023-01472-w] [Medline: 37704982]
74. Tran-Nguyen K, Berger C, Bennett R, Wall M, Morin SN, Rajabiyazdi F. Mobile app prototype in older adults for postfracture acute pain management: user-centered design approach. JMIR Aging 2022 Oct 17;5(4):e37772 [FREE Full text] [doi: 10.2196/37772] [Medline: 36251348]

XSL•FO
RenderX

75. Sönnerfors P, Skavberg Roaldsen K, Lundell S, Toots A, Wadell K, Halvarsson A. Preferences for an eHealth tool to support physical activity and exercise training in COPD: a qualitative study from the viewpoint of prospective users. BMC Pulm Med 2023 Feb 13;23(1):65 [FREE Full text] [doi: 10.1186/s12890-023-02353-3] [Medline: 36782155]
76. Young SR, Lattie EG, Berry AB, Bui L, Byrne GJ, Yoshino Benavente JN, et al. Remote cognitive screening of healthy older adults for primary care with the MyCog mobile app: iterative design and usability evaluation. JMIR Form Res 2023 Jan 10;7:e42416 [FREE Full text] [doi: 10.2196/42416] [Medline: 36626223]
77. De Sutter E, Geerts D, Yskout K, Verreydt S, Borry P, Barbier L, et al. Testing and practical implementation of a user-friendly personalized and long-term electronic informed consent prototype in clinical research: mixed methods study. J Med Internet Res 2023 Dec 19;25:e46306 [FREE Full text] [doi: 10.2196/46306] [Medline: 38113088]
78. Deady M, Collins D, Gayed A, Harvey SB, Bryant R. The development of a smartphone app to enhance post-traumatic stress disorder treatment in high-risk workers. Digit Health 2023 Feb 22;9:20552076231155680 [FREE Full text] [doi: 10.1177/20552076231155680] [Medline: 36845080]
79. Webb MM, Bridges P, Aruparayil N, Chugh C, Beacon T, Singh T, et al. The RAIS device for global surgery: using a participatory design approach to navigate the translational pathway to clinical use. IEEE J Transl Eng Health Med 2022;10:3700212-3700212 [FREE Full text] [doi: 10.1109/JTEHM.2022.3177313] [Medline: 35865752]
80. Yau KW, Tang TS, Görges M, Pinkney S, Amed S. Using human-centered design and cocreation to create the live 5-2-1-0 mobile app to promote healthy behaviors in children: app design and development. JMIR Pediatr Parent 2023 May 17;6:e44792 [FREE Full text] [doi: 10.2196/44792] [Medline: 37195754]
81. Ludlow K, Russell JK, Ryan B, Brown RL, Joynt T, Uhlmann LR, et al. Co-designing a digital mental health platform, "Momentum", with young people aged 7-17: a qualitative study. Digit Health 2023 Nov 28;9:20552076231216410 [FREE Full text] [doi: 10.1177/20552076231216410] [Medline: 38033517]
82. van de Riet L, Aris AM, Verouden NW, van Rooij T, van Woensel JB, van Karnebeek CD, et al. Designing eHealth interventions for children with complex care needs requires continuous stakeholder collaboration and co-creation. PEC Innov 2024 Dec;4:100280 [FREE Full text] [doi: 10.1016/j.pecinn.2024.100280] [Medline: 38596601]
83. Dahildahil RO, Isip-Tan IT, Marcelo PG, Grepo LC. User-centered design in time and resource-limited settings: enhancing the usability of 'hearing for life' (HeLe) device. Acta Med Philipp 2023 Sep 27;57(9):39-53 [FREE Full text] [doi: 10.47895/amp.v57i9.1594] [Medline: 39483800]
84. Bonet-Olivencia S, Carrillo-Leal J, Rao A, Sasangohar F. User-centered design of a diabetes self-management tool for underserved populations. J Diabetes Sci Technol 2024 Jan 18;18(1):22-29. [doi: 10.1177/19322968231212220] [Medline: 37978811]
85. Reimer LM, Nissen L, von Scheidt M, Perl B, Wiehler J, Najem SA, et al. User-centered development of an mHealth app for cardiovascular prevention. Digit Health 2024;10:20552076241249269 [FREE Full text] [doi: 10.1177/20552076241249269] [Medline: 38774157]
86. Appleton KM, Bray J, Price S, Liebchen G, Jiang N, Mavridis I, et al. A mobile phone app for the provision of personalized food-based information in an eating-out situation: development and initial evaluation. JMIR Form Res 2019 Nov 04;3(4):e12966 [FREE Full text] [doi: 10.2196/12966] [Medline: 31682575]
87. van den Berg LN, Hallensleben C, Chavannes NH, Versluis A. Developing a smartphone application that promotes responsible short-acting Beta2-agonist use in People with asthma: a participatory design. Int J Environ Res Public Health 2022 Jul 12;19(14):63 [FREE Full text] [doi: 10.3390/ijerph19148496] [Medline: 35886348]
88. Donald M, Beanlands H, Straus SE, Smekal M, Gil S, Elliott MJ, et al. A web-based self-management support prototype for adults with chronic kidney disease (my kidneys my health): co-design and usability testing. JMIR Form Res 2021 Feb 09;5(2):e22220 [FREE Full text] [doi: 10.2196/22220] [Medline: 33560245]
89. Ehrler F, Lovis C, Blondon K. A mobile phone app for bedside nursing care: design and development using an adapted software development life cycle model. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12551 [FREE Full text] [doi: 10.2196/12551] [Medline: 30973339]
90. Humphries SM, Rondung E, Norlund F, Sundin Ö, Tornvall P, Held C, et al. Designing a web-based psychological intervention for patients with myocardial infarction with nonobstructive coronary arteries: user-centered design approach. J Med Internet Res 2020 Sep 17;22(9):e19066 [FREE Full text] [doi: 10.2196/19066] [Medline: 32940615]
91. Andersson SR, Hassanen S, Momanyi AM, Onyango DK, Gatwechi DK, Lutukai MN, et al. Using human-centered design to adapt supply chains and digital solutions for community health volunteers in nomadic communities of Northern Kenya. Glob Health Sci Pract 2021 Mar 15;9(Suppl 1):S151-S167 [FREE Full text] [doi: 10.9745/GHSP-D-20-00378] [Medline: 33727327]
92. Fleisher L, Bass SB, Shwarz M, Washington A, Nicholson A, Alhajji M, et al. Using theory and user-centered design in digital health: the development of the mychoice communication tool to prepare patients and improve informed decision making regarding clinical trial participation. Psychooncology 2020 Jan 02;29(1):114-122. [doi: 10.1002/pon.5254] [Medline: 31654442]
93. Saparamadu AA, Fernando P, Zeng P, Teo H, Goh A, Lee JM, et al. User-centered design process of an mHealth app for health professionals: case study. JMIR Mhealth Uhealth 2021 Mar 26;9(3):e18079 [FREE Full text] [doi: 10.2196/18079] [Medline: 33769297]

XSL•FO
RenderX

94. Ahtinen A, Andrejeff E, Vuolle M, Väänänen K. Walk as you work: user study and design implications for mobile walking meetings. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction. 2016 Presented at: NordiCHI '16; October 23-27, 2016; Gothenburg, Sweden p. 1-10 URL: https://dl.acm.org/doi/10.1145/2971485.2971510 [doi: 10.1145/2971485.2971510]
95. Balli F. Developing digital games to address airway clearance therapy in children with cystic fibrosis: participatory design process. JMIR Serious Games 2018 Nov 21;6(4):e18 [FREE Full text] [doi: 10.2196/games.8964] [Medline: 30463835]
96. Tyo MB, Desroches ML. Content analysis of qualitative interviews for user-centered design of a prototype mobile health app for direct support professionals' resilience. Disabil Health J 2024 Apr;17(2):101544 [FREE Full text] [doi: 10.1016/j.dhjo.2023.101544] [Medline: 37981491]
97. Shanmuganathan C, Anish TP, Thamizharasi M, Joseph M. An iterative constraint spectral model for ophthalmic disease detection using transfer learning. In: Proceedings of the 2022 International Conference on Data Science, Agents & Artificial Intelligence. 2022 Presented at: ICDSAAI '22; December 8-10, 2022; Chennai, India p. 1-7 URL: https://ieeexplore.ieee.org/document/10028807 [doi: 10.1109/icdsaai55433.2022.10028807]
98. Schwartz LA, Psihogios AM, Henry-Moss D, Daniel LC, Ver Hoeve ES, Velazquez-Martin B, et al. Iterative development of a tailored mHealth intervention for adolescent and young adult survivors of childhood cancer. Clin Pract Pediatr Psychol 2019 Mar;7(1):31-43 [FREE Full text] [doi: 10.1037/cpp0000272] [Medline: 38882593]
99. Ortet C, Vale Costa L. "Listen to your immune system when it's calling for you": monitoring autoimmune diseases using the iShU app. Sensors (Basel) 2022 May 18;22(10):3834 [FREE Full text] [doi: 10.3390/s22103834] [Medline: 35632243]
100. Ramsey AT, Bray M, Acayo Laker P, Bourdon JL, Dorsey A, Zalik M, et al. Participatory design of a personalized genetic risk tool to promote behavioral health. Cancer Prev Res (Phila) 2020 Jul;13(7):583-592 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-20-0029] [Medline: 32209550]
101. Nota I, Drossaert CH, Melissant HC, Taal E, Vonkeman HE, Haagsma CJ, et al. Development of a web-based patient decision aid for initiating disease modifying anti-rheumatic drugs using user-centred design methods. BMC Med Inform Decis Mak 2017 Apr 26;17(1):51 [FREE Full text] [doi: 10.1186/s12911-017-0433-5] [Medline: 28441950]
102. Stratton E, Choi I, Peters D, Calvo RA, Harvey SB, Glozier N. Co-designing a web-based decision aid tool for employees disclosure of mental health conditions: a participatory study design using employee and organizational preferences. JMIR Form Res 2020 Nov 06;4(11):e23337 [FREE Full text] [doi: 10.2196/23337] [Medline: 33155982]
103. van Leersum CM, Moser A, van Steenkiste B, Reinartz M, Stoffers E, Wolf JR, et al. What matters to me - a web-based preference elicitation tool for clients in long-term care: a user-centred design. BMC Med Inform Decis Mak 2020 Mar 17;20(1):57 [FREE Full text] [doi: 10.1186/s12911-020-1067-6] [Medline: 32183786]
104. Poncette AS, Mosch LK, Stablo L, Spies C, Schieler M, Weber-Carstens S, et al. A remote patient-monitoring system for intensive care medicine: mixed methods human-centered design and usability evaluation. JMIR Hum Factors 2022 Mar 11;9(1):e30655 [FREE Full text] [doi: 10.2196/30655] [Medline: 35275071]
105. Engdahl P, Svedberg P, Lexén A, Tjörnstrand C, Strid C, Bejerholm U. Co-design process of a digital return-to-work solution for people with common mental disorders: stakeholder perception study. JMIR Form Res 2023 Jan 18;7:e39422 [FREE Full text] [doi: 10.2196/39422] [Medline: 36652285]
106. Verhoeven F, Cremers A, Schoone M, Van Dijk J. Mobiles for mobility: participatory design of a ‘happy walker’ that stimulates mobility among older people. Gerontechnology 2016;15(1):32-44 [FREE Full text] [doi: 10.4017/GT.2016.15.1.008.00]
107. Couture B, Lilley E, Chang F, DeBord Smith A, Cleveland J, Ergai A, et al. Applying user-centered design methods to the development of an mHealth application for use in the hospital setting by patients and care partners. Appl Clin Inform 2018 Apr 09;9(2):302-312 [FREE Full text] [doi: 10.1055/s-0038-1645888] [Medline: 29742756]
108. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia: development and usability study of thinkability. JMIR Aging 2020 Nov 16;3(2):e17105 [FREE Full text] [doi: 10.2196/17105] [Medline: 33196451]
109. Schiro J, Pelayo S, Martinot A, Dubos F, Beuscart-Zéphir MC, Marcilly R. Applying a human-centered design to develop a patient prioritization tool for a pediatric emergency department: detailed case study of first iterations. JMIR Hum Factors 2020 Sep 04;7(3):e18427 [FREE Full text] [doi: 10.2196/18427] [Medline: 32886071]
110. Abubakar Kawu A, Kahuhu Koruhama A, Arinze Ikwunne T, Pretorius C. Gamifying the “unspoken”: designing to resolve myths on contraceptives among the Himba people of Namibia. In: Proceedings of the 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services. 2020 Presented at: MobileHCI '20; October 5-8, 2020; Oldenburg, Germany p. 1-4 URL: https://dl.acm.org/doi/10.1145/3406324.3424592 [doi: 10.1145/3406324.3424592]
111. Tobias G, Spanier AB. Developing a mobile app (iGAM) to promote gingival health by professional monitoring of dental selfies: user-centered design approach. JMIR Mhealth Uhealth 2020 Aug 14;8(8):e19433 [FREE Full text] [doi: 10.2196/19433] [Medline: 32795985]
112. Tonkin E, Jeffs L, Wycherley TP, Maher C, Smith R, Hart J, et al. A smartphone app to reduce sugar-sweetened beverage consumption among young adults in Australian remote indigenous communities: design, formative evaluation and user-testing. JMIR Mhealth Uhealth 2017 Dec 12;5(12):e192 [FREE Full text] [doi: 10.2196/mhealth.8651] [Medline: 29233803]

XSL•FO
RenderX

113. Ospina-Pinillos L, Davenport T, Mendoza Diaz A, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eClinic: qualitative study. J Med Internet Res 2019 Aug 02;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
114. Hou M, Hindriks K, Eiben AE, Baraka K. A process-oriented framework for robot imitation learning in human-centered interactive tasks. In: Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication. 2023 Presented at: RO-MAN '23; August 28-31, 2023; Busan, Korea p. 1745-1752 URL: https://ieeexplore.ieee.org/document/10309326 [doi: 10.1109/ro-man57019.2023.10309326]
115. Wannheden C, Revenäs Å. How people with Parkinson's disease and health care professionals wish to partner in care using eHealth: co-design study. J Med Internet Res 2020 Sep 21;22(9):e19195 [FREE Full text] [doi: 10.2196/19195] [Medline: 32955448]
116. Giles SJ, Reynolds C, Heyhoe J, Armitage G. Developing a patient-led electronic feedback system for quality and safety within Renal PatientView. J Ren Care 2017 Mar 19;43(1):37-49 [FREE Full text] [doi: 10.1111/jorc.12186] [Medline: 27990782]
117. Bevan Jones R, Thapar A, Rice F, Beeching H, Cichosz R, Mars B, et al. A web-based psychoeducational intervention for adolescent depression: design and development of MoodHwb. JMIR Ment Health 2018 Feb 15;5(1):e13 [FREE Full text] [doi: 10.2196/mental.8894] [Medline: 29449202]
118. Bradway M, Morris RL, Giordanengo A, Årsand E. How mHealth can facilitate collaboration in diabetes care: qualitative analysis of co-design workshops. BMC Health Serv Res 2020 Nov 30;20(1):1104 [FREE Full text] [doi: 10.1186/s12913-020-05955-3] [Medline: 33256732]
119. Naeemabadi M, Søndergaard JH, Klastrup A, Schlünsen AP, Lauritsen RE, Hansen J, et al. Development of an individualized asynchronous sensor-based telerehabilitation program for patients undergoing total knee replacement: participatory design. Health Informatics J 2020 Dec;26(4):2492-2511 [FREE Full text] [doi: 10.1177/1460458220909779] [Medline: 32175788]
120. Kim JE, Lee J, Kwon H, Bessho M, Sakamura K. Towards supporting childhood atopic dermatitis management: a user-centered design approach. In: Proceedings of the 33rd International Symposium on Computer-Based Medical Systems. 2020 Presented at: CBMS '20; July 28-30, 2020; Rochester, MN p. 368-373 URL: https://ieeexplore.ieee.org/document/9183058 [doi: 10.1109/cbms49503.2020.00075]
121. Allsop MJ, Johnson O, Taylor S, Hackett J, Allen P, Bennett MI, et al. Multidisciplinary software design for the routine monitoring and assessment of pain in palliative care services: the development of PainCheck. JCO Clin Cancer Inform 2019 Sep;3:1-17 [FREE Full text] [doi: 10.1200/CCI.18.00120] [Medline: 31577449]
122. Schleimer E, Pearce J, Barnecut A, Rowles W, Lizee A, Klein A, et al. A precision medicine tool for patients with multiple sclerosis (the Open MS BioScreen): human-centered design and development. J Med Internet Res 2020 Jul 06;22(7):e15605 [FREE Full text] [doi: 10.2196/15605] [Medline: 32628124]
123. Terp M, Laursen BS, Jørgensen R, Mainz J, Bjørnes CD. A room for design: through participatory design young adults with schizophrenia become strong collaborators. Int J Ment Health Nurs 2016 Dec;25(6):496-506. [doi: 10.1111/inm.12231] [Medline: 27293176]
124. Carr EC, Babione JN, Marshall D. Translating research into practice through user-centered design: an application for osteoarthritis healthcare planning. Int J Med Inform 2017 Aug;104:31-37 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.05.007] [Medline: 28599814]
125. Hesseldal L, Christensen JR, Olesen TB, Olsen MH, Jakobsen PR, Laursen DH, et al. Long-term weight loss in a primary care-anchored eHealth lifestyle coaching program: randomized controlled trial. J Med Internet Res 2022 Sep 23;24(9):e39741 [FREE Full text] [doi: 10.2196/39741] [Medline: 36149735]
126. Marko-Holguin M, Cordel SL, Van Voorhees BW, Fogel J, Sykes E, Fitzgibbon M, et al. A two-way interactive text messaging application for low-income patients with chronic medical conditions: design-thinking development approach. JMIR Mhealth Uhealth 2019 May 01;7(5):e11833 [FREE Full text] [doi: 10.2196/11833] [Medline: 31042152]
127. Guay M, Labbé M, Séguin-Tremblay N, Auger C, Goyer G, Veloza E, et al. Adapting a person's home in 3D using a mobile app (MapIt): participatory design framework investigating the app's acceptability. JMIR Rehabil Assist Technol 2021 May 11;8(2):e24669 [FREE Full text] [doi: 10.2196/24669] [Medline: 33973867]
128. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]
129. Smaradottir B, Holen-Rabbersvik E, Thygesen E, Fensli R, Martinez S. User-centred design of the user interface of a collaborative information system for inter-municipal dementia team. In: Proceedings of the 2015 International Conference on Health Informatics. 2015 Presented at: HEALTHINF '15; January 12-15, 2015; Lisbon, Portugal p. 22-31 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0005222704460453 [doi: 10.5220/0005222704460453]
130. Smaradottir B, Gerdes M, Martinez S, Fensli R. The EU-project United4Health: user-centred design of an information system for a Norwegian telemedicine service. J Telemed Telecare 2016 Oct 09;22(7):422-429 [FREE Full text] [doi: 10.1177/1357633X15615048] [Medline: 26541347]

XSL•FO
RenderX

131. Uddin AA, Morita PP, Tallevi K, Armour K, Li J, Nolan RP, et al. Development of a wearable cardiac monitoring system for behavioral neurocardiac training: a usability study. JMIR Mhealth Uhealth 2016 Apr 22;4(2):e45 [FREE Full text] [doi: 10.2196/mhealth.5288] [Medline: 27106171]
132. Jibb LA, Cafazzo JA, Nathan PC, Seto E, Stevens BJ, Nguyen C, et al. Development of a mHealth real-time pain self-management app for adolescents with cancer: an iterative usability testing study [formula: see text]. J Pediatr Oncol Nurs 2017;34(4):283-294. [doi: 10.1177/1043454217697022] [Medline: 28376666]
133. Rodrigues AM, Sniehotta FF, Birch-Machin MA, Olivier P, Araújo-Soares V. Systematic and iterative development of a smartphone app to promote sun-protection among holidaymakers: design of a prototype and results of usability and acceptability testing. JMIR Res Protoc 2017 Jun 12;6(6):e112 [FREE Full text] [doi: 10.2196/resprot.7172] [Medline: 28606892]
134. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
135. Afrizal SH, Hidayanto AN, Hakiem N, Sartono A, Priyambodo S, Eryando T. Design of mHealth application for integrating antenatal care service in primary health care: a user-centered approach. In: Proceedings of the 4th International Conference on Informatics and Computing. 2019 Presented at: ICIC '19; October 16-17, 2019; Semarang, Indonesia p. 1-6 URL: https://ieeexplore.ieee.org/document/8985911 [doi: 10.1109/icic47613.2019.8985911]
136. Sedlmayr B, Schöffler J, Prokosch HU, Sedlmayr M. User-centered design of a mobile medication management. Inform Health Soc Care 2019 Mar 05;44(2):152-163. [doi: 10.1080/17538157.2018.1437042] [Medline: 29504838]
137. Laine TH, Normark J, Lindvall H, Lindqvist AK, Rutberg S. A distributed multiplayer game to promote active transport at workplaces: user-centered design, implementation, and lessons learned. IEEE Trans Games 2020 Dec;12(4):386-397. [doi: 10.1109/TG.2020.3021728]
138. Tonga E, Williamson E, Srikesavan C, Özen T, Sarıtaş F, Lamb SE. A hand exercise mobile app for people with rheumatoid arthritis in Turkey: design, development and usability study. Rheumatol Int 2021 Jun 18;41(6):1151-1160. [doi: 10.1007/s00296-021-04860-0] [Medline: 33870452]
139. Honary M, Fisher NR, McNaney R, Lobban F. A web-based intervention for relatives of people experiencing psychosis or bipolar disorder: design study using a user-centered approach. JMIR Ment Health 2018 Dec 07;5(4):e11473 [FREE Full text] [doi: 10.2196/11473] [Medline: 30530457]
140. Span M, Hettinga M, Groen-van de Ven L, Jukema J, Janssen R, Vernooij-Dassen M, et al. Involving people with dementia in developing an interactive web tool for shared decision-making: experiences with a participatory design approach. Disabil Rehabil 2018 Jun;40(12):1410-1420. [doi: 10.1080/09638288.2017.1298162] [Medline: 28286969]
141. Schild S, Sedlmayr B, Schumacher AK, Sedlmayr M, Prokosch H, St Pierre M, German Cognitive Aid Working Group. A digital cognitive aid for anesthesia to support intraoperative crisis management: results of the user-centered design process. JMIR Mhealth Uhealth 2019 Apr 29;7(4):e13226 [FREE Full text] [doi: 10.2196/13226] [Medline: 31033445]
142. Buck B, Wingerson M, Whiting E, Snyder J, Monroe-DeVita M, Ben-Zeev D. User-centered development of bolster, an mHealth intervention for early psychosis caregivers: needs assessment, prototyping, and field trial. JMIR Ment Health 2023 Nov 30;10:e50522 [FREE Full text] [doi: 10.2196/50522] [Medline: 38032692]
143. Ahufinger S, Pequeno E, Herrero P. Mob-EEG: a user-centered design for wireless electroencephalogram in intensive care units. In: Proceedings of the 32nd International Symposium on Computer-Based Medical Systems. 2019 Presented at: CBMS '19; June 5-7, 2019; Cordoba, Spain p. 204-207 URL: https://ieeexplore.ieee.org/abstract/document/8787513 [doi: 10.1109/cbms.2019.00049]
144. Alberts NM, Badawy SM, Hodges J, Estepp JH, Nwosu C, Khan H, et al. Development of the InCharge health mobile app to improve adherence to hydroxyurea in patients with sickle cell disease: user-centered design approach. JMIR Mhealth Uhealth 2020 May 08;8(5):e14884 [FREE Full text] [doi: 10.2196/14884] [Medline: 32383683]
145. Corazza F, Snijders D, Arpone M, Stritoni V, Martinolli F, Daverio M, et al. Development and usability of a novel interactive tablet app (PediAppRREST) to support the management of pediatric cardiac arrest: pilot high-fidelity simulation-based study. JMIR Mhealth Uhealth 2020 Oct 01;8(10):e19070 [FREE Full text] [doi: 10.2196/19070] [Medline: 32788142]
146. Erguera XA, Johnson MO, Neilands TB, Ruel T, Berrean B, Thomas S, et al. WYZ: a pilot study protocol for designing and developing a mobile health application for engagement in HIV care and medication adherence in youth and young adults living with HIV. BMJ Open 2019 May 05;9(5):e030473 [FREE Full text] [doi: 10.1136/bmjopen-2019-030473] [Medline: 31061063]
147. Kerkhof YJ, den Ouden ME, Soeteman S, Scholten A, Ben Allouch S, Willems CG. Development of a memory application for structuring and supporting daily activities of clients with intellectual disabilities. Technol Disabil 2017 Jul 07;29(1-2):77-89. [doi: 10.3233/tad-160164]
148. Marien S, Legrand D, Ramdoyal R, Nsenga J, Ospina G, Ramon V, et al. A User-Centered design and usability testing of a web-based medication reconciliation application integrated in an eHealth network. Int J Med Inform 2019 Jun;126:138-146. [doi: 10.1016/j.ijmedinf.2019.03.013] [Medline: 31029255]

XSL•FO
RenderX

149. Martin S, Armstrong E, Thomson E, Vargiu E, Solà M, Dauwalder S, et al. A qualitative study adopting a user-centered approach to design and validate a brain computer interface for cognitive rehabilitation for people with brain injury. Assist Technol 2018 Jul 14;30(5):233-241. [doi: 10.1080/10400435.2017.1317675] [Medline: 28708963]
150. Navarro-Alamán J, Lacuesta R, García-Magariño I, Gallardo J, Ibarz E, Lloret J. Close2U: an app for monitoring cancer patients with enriched information from interaction patterns. J Healthc Eng 2020 Jul 15;2020:3057032-3057013 [FREE Full text] [doi: 10.1155/2020/3057032] [Medline: 32733661]
151. Newton A, Bagnell A, Rosychuk R, Duguay J, Wozney L, Huguet A, et al. A mobile phone-based app for use during cognitive behavioral therapy for adolescents with anxiety (MindClimb): user-centered design and usability study. JMIR Mhealth Uhealth 2020 Dec 08;8(12):e18439 [FREE Full text] [doi: 10.2196/18439] [Medline: 33289671]
152. Noordman J, Driesenaar JA, van Bruinessen IR, van Dulmen S. ListeningTime; participatory development of a web-based preparatory communication tool for elderly cancer patients and their healthcare providers. Internet Interv 2017 Sep;9:51-56 [FREE Full text] [doi: 10.1016/j.invent.2017.05.002] [Medline: 30135837]
153. Petersen M, Hempler NF. Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):91 [FREE Full text] [doi: 10.1186/s12911-017-0493-6] [Medline: 28651639]
154. Pham Q, Graham G, Lalloo C, Morita PP, Seto E, Stinson JN, et al. An analytics platform to evaluate effective engagement with pediatric mobile health apps: design, development, and formative evaluation. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11447 [FREE Full text] [doi: 10.2196/11447] [Medline: 30578179]
155. Sin J, Henderson C, Woodham LA, Sesé Hernández A, Gillard S. A multicomponent eHealth intervention for family carers for people affected by psychosis: a coproduced design and build study. J Med Internet Res 2019 Aug 06;21(8):e14374 [FREE Full text] [doi: 10.2196/14374] [Medline: 31389333]
156. Stevenson JE, Oscarsson M. User-centred iterative design to develop an evidence-based communication application for maternity care. Health Informatics J 2021 May 24;27(2):14604582211014579 [FREE Full text] [doi: 10.1177/14604582211014579] [Medline: 34030521]
157. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
158. Willard S, Cremers G, Man YP, van Rossum E, Spreeuwenberg M, de Witte L. Development and testing of an online community care platform for frail older adults in the Netherlands: a user-centred design. BMC Geriatr 2018 Apr 07;18(1):87 [FREE Full text] [doi: 10.1186/s12877-018-0774-7] [Medline: 29625562]
159. Pienkowska A, Ang CS, Mammadova M, Mahadzir MD, Car J. A diabetes education app for people living with type 2 diabetes: co-design study. JMIR Form Res 2023 Sep 18;7:e45490 [FREE Full text] [doi: 10.2196/45490] [Medline: 37721799]
160. Alvarez SJ, Pencheva B, Westfall E, Mwalija C, Parsell M, Greenleaf M, et al. A novel mobile health application to support cancer surveillance needs of patients and families with cancer predisposition syndromes. Pediatr Blood Cancer 2023 Oct 06;70(10):e30537 [FREE Full text] [doi: 10.1002/pbc.30537] [Medline: 37415085]
161. Shanmugavel A, Shakya PR, Shrestha A, Nepal J, Shrestha A, Daneault J, et al. Designing and developing a mobile app for management and treatment of gestational diabetes in Nepal: user-centered design study. JMIR Form Res 2024 Jan 17;8:e50823 [FREE Full text] [doi: 10.2196/50823] [Medline: 38231562]
162. Dannenberg MD, Bienvenida JC, Bruce ML, Nguyen T, Hinn M, Matthews J, et al. End-user views of an electronic encounter decision aid linked to routine depression screening. Patient Educ Couns 2019 Mar;102(3):555-563. [doi: 10.1016/j.pec.2018.10.002] [Medline: 30497800]
163. Rao N, Perdomo S, Jonassaint C. A novel method for digital pain assessment using abstract animations: human-centered design approach. JMIR Hum Factors 2022 Jan 07;9(1):e27689 [FREE Full text] [doi: 10.2196/27689] [Medline: 34994697]
164. Zhang M, Heng S, Song G, Fung DS, Smith HE. Co-designing a mobile gamified attention bias modification intervention for substance use disorders: participatory research study. JMIR Mhealth Uhealth 2019 Oct 03;7(10):e15871 [FREE Full text] [doi: 10.2196/15871] [Medline: 31584003]
165. Calvillo-Arbizu J, Roa-Romero LM, Estudillo-Valderrama MA, Salgueira-Lazo M, Aresté-Fosalba N, Del-Castillo-Rodríguez NL, et al. User-centred design for developing e-Health system for renal patients at home (AppNephro). Int J Med Inform 2019 May;125:47-54. [doi: 10.1016/j.ijmedinf.2019.02.007] [Medline: 30914180]
166. van Velsen L, van Weering MD, Luub F, Neven A. Travelling with my SOULMATE: participatory design of an mHealth travel companion for older adults. In: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health. 2019 Presented at: ICT4AWE '19; May 2-4, 2019; Crete, Greece p. 38-47 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0007680200380047 [doi: 10.5220/0007680200380047]
167. Stawarz K, Preist C, Tallon D, Wiles N, Kessler D, Turner K, et al. Design considerations for the integrated delivery of cognitive behavioral therapy for depression: user-centered design study. JMIR Ment Health 2020 Sep 03;7(9):e15972 [FREE Full text] [doi: 10.2196/15972] [Medline: 32880580]

XSL•FO
RenderX

168. Stoyanov SR, Zelenko O, Staneva A, Kavanagh DJ, Smith C, Sade G, et al. Development of the Niggle app for supporting young people on their dynamic journey to well-being: co-design and qualitative research study. JMIR Mhealth Uhealth 2021 Apr 20;9(4):e21085 [FREE Full text] [doi: 10.2196/21085] [Medline: 33877050]
169. Hoffman AS, Bateman DR, Ganoe C, Punjasthitkul S, Das AK, Hoffman DB, et al. Development and field testing of a long-term care decision aid website for older adults: engaging patients and caregivers in user-centered design. Gerontologist 2020 Jul 15;60(5):935-946 [FREE Full text] [doi: 10.1093/geront/gnz141] [Medline: 31773140]
170. Backman C, Harley A, Kuziemsky C, Mercer J, Peyton L. MyPath to home web-based application for the geriatric rehabilitation program at Bruyère continuing care: user-centered design and feasibility testing study. JMIR Form Res 2020 Sep 14;4(9):e18169 [FREE Full text] [doi: 10.2196/18169] [Medline: 32924953]
171. Korpershoek YJ, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JC. User-centered design of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease (Copilot): mixed methods study. J Med Internet Res 2020 Jun 15;22(6):e15449 [FREE Full text] [doi: 10.2196/15449] [Medline: 32538793]
172. Da Silva Júnior JL, Biduski D, Bellei EA, Becker OH, Daroit L, Pasqualotti A, et al. A bowling exergame to improve functional capacity in older adults: co-design, development, and testing to compare the progress of playing alone versus playing with peers. JMIR Serious Games 2021 Jan 29;9(1):e23423 [FREE Full text] [doi: 10.2196/23423] [Medline: 33512319]
173. Dirin M, Dirin A, Laine TH. User-centered design of a context-aware nurse assistant (CANA) at Finnish elderly houses. In: Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication.: Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. IMCOM ?15. Association for Computing Machinery; 2015 Presented at: IMCOM '15; January 8-10, 2015; Bali, Indonesia p. 1-8 URL: https://dl.acm.org/doi/10.1145/2701126.2701225 [doi: 10.1145/2701126.2701225]
174. Lyles CR, Altschuler A, Chawla N, Kowalski C, McQuillan D, Bayliss E, et al. User-centered design of a tablet waiting room tool for complex patients to prioritize discussion topics for primary care visits. JMIR Mhealth Uhealth 2016 Sep 14;4(3):e108 [FREE Full text] [doi: 10.2196/mhealth.6187] [Medline: 27627965]
175. Marien S, Legrand D, Ramdoyal R, Nsenga J, Ospina G, Ramon V, et al. A web application to involve patients in the medication reconciliation process: a user-centered usability and usefulness study. J Am Med Inform Assoc 2018 Nov 01;25(11):1488-1500 [FREE Full text] [doi: 10.1093/jamia/ocy107] [Medline: 30137331]
176. He X, Zhang R, Rizvi R, Vasilakes J, Yang X, Guo Y, et al. ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design. BMC Med Inform Decis Mak 2019 Aug 08;19(Suppl 4):150 [FREE Full text] [doi: 10.1186/s12911-019-0857-1] [Medline: 31391091]
177. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
178. Mathenjwa T, Adeagbo O, Zuma T, Dikgale K, Zeitlin A, Matthews P, et al. Development and acceptability of a tablet-based app to support men to link to HIV care: mixed methods approach. JMIR Mhealth Uhealth 2020 Nov 24;8(11):e17549 [FREE Full text] [doi: 10.2196/17549] [Medline: 33231558]
179. van Bruinessen I, van Klooster JW, Boessen A, Heide L, Vollenbroek-Hutten M, van Dulmen S. Involving end-users in the development and implementation of a web-based, physical activity platform for elderly. In: Proceedings of the 2017 International Conference on e-Health. 2017 Presented at: eHealth '17; October 12-15, 2017; Dalian, China p. 145-149 URL: https://www.iadisportal.org/digital-library/involving-end-users-in-the-development-and-implementation-of-a-web-based-physical-activity-platform-for-elderly
180. Chan NH, Merali HS, Mistry N, Kealey R, Campbell DM, Morris SK, et al. Development of a novel mobile application, HBB prompt, with human factors and user-centred design for helping babies breathe skills retention in Uganda. BMC Med Inform Decis Mak 2021 Feb 04;21(1):39 [FREE Full text] [doi: 10.1186/s12911-021-01406-z] [Medline: 33541340]
181. Balatsoukas P, Porat T, Sassoon I, Essers K, Kokciyan N, Chapman M, et al. User involvement in the design of a data-driven self-management decision support tool for stroke survivors. In: Proceedings of the 18th International Conference on Smart Technologies. 2019 Presented at: IEEE EUROCON '19; July 1-4, 2019; Novi Sad, Serbia p. 1-6 URL: https://ieeexplore.ieee.org/document/8861812 [doi: 10.1109/EUROCON.2019.8861812]
182. Woods L, Cummings E, Duff J, Walker K. Conceptual design and iterative development of a mHealth app by clinicians, patients and their families. Stud Health Technol Inform 2018;252:170-175. [Medline: 30040701]
183. Powell J, Newhouse N, Boylan AM, Williams V. Digital health citizens and the future of the NHS. Digit Health 2016 Oct 10;2:2055207616672033 [FREE Full text] [doi: 10.1177/2055207616672033] [Medline: 29942569]
184. Kress HG, Aldington D, Alon E, Coaccioli S, Collett B, Coluzzi F, et al. A holistic approach to chronic pain management that involves all stakeholders: change is needed. Curr Med Res Opin 2015;31(9):1743-1754. [doi: 10.1185/03007995.2015.1072088] [Medline: 26172982]
185. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018;3(1):e000676 [FREE Full text] [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]

XSL•FO
RenderX

186. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018 Sep 13;1(1):46 [FREE Full text] [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
187. Petrakaki D, Hilberg E, Waring J. The cultivation of digital health citizenship. Soc Sci Med 2021 Feb;270:113675 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113675] [Medline: 33434718]
188. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci 2018 Jan 25;13(Suppl 1):2 [FREE Full text] [doi: 10.1186/s13012-017-0688-3] [Medline: 29384079]

Abbreviations

DHI: digital health intervention
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols
RCT: randomized controlled trial
UCD: user-centered design
UX: user experience

Edited by A Mavragani; submitted 21.06.23; peer-reviewed by E Baker, E Rosenfeld; comments to author 19.09.23; revised version received 06.12.23; accepted 07.12.24; published 14.03.25.

Please cite as:
Duffy A, Boroumandzad N, Sherman AL, Christie G, Riadi I, Moreno S
Examining Challenges to Co-Design Digital Health Interventions With End Users: Systematic Review
J Med Internet Res 2025;27:e50178
URL: https://www.jmir.org/2025/1/e50178
doi:10.2196/50178
PMID:

©Anthony Duffy, Nazanin Boroumandzad, Alfredo Lopez Sherman, Gregory Christie, Indira Riadi, Sylvain Moreno. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Association of Digital Health Interventions With Maternal and Neonatal Outcomes: Systematic Review and Meta-Analysis

Jianing Wang[1], MSc; Nu Tang[2], MSc; Congcong Jin[3], BSc; Jianxue Yang[4], MSc; Xiangpeng Zheng[4], MSc; Qiujing Jiang[5], MSc; Shengping Li[6], MSc; Nian Xiao[7], BSc; Xiaojun Zhou[1], MPH

[1]Department of Maternal and Child Information Management, Women and Children's Hospital of Chongqing Medical University, Chonqing, China
[2]Community Health Care Office, Women and Children's Hospital of Chongqing Medical University, Chongqing, China
[3]Pregnancy Health Center, Women and Children's Hospital of Chongqing Medical University, Chongqing, China
[4]Department of Maternal and Child Health, Chongqing Municipal Health Commission, Chongqing, China
[5]Department of Health Education, Women and Children's Hospital of Chongqing Medical University, Chongqing, China
[6]Department of Child Healthcare, Women and Children's Hospital of Chongqing Medical University, Chongqing, China
[7]Department of Equipment Management, Women and Children's Hospital of Chongqing Medical University, Chongqing, China

Corresponding Author:
Xiaojun Zhou, MPH
Department of Maternal and Child Information Management
Women and Children's Hospital of Chongqing Medical University
No.120 Longshan Road, Yubei District
Chonqing, 401147
China
Phone: 86 23 6723 2442
Fax: 86 23 6384 0868
Email: 1186440689@qq.com

Abstract

Background: Gestational weight gain (GWG) is crucial to maternal and neonatal health, yet many women fail to meet recommended guidelines, increasing the risk of complications. Digital health interventions offer promising solutions, but their effectiveness remains uncertain. This study evaluates the impact of such interventions on GWG and other maternal and neonatal outcomes.

Objective: This study aimed to investigate the effect of digital health interventions among pregnant women and newborns.

Methods: A total of 2 independent researchers performed electronic literature searches in the PubMed, Embase, Web of Science, and Cochrane Library databases to identify eligible studies published from their inception until February 2024; an updated search was conducted in August 2024. The studies included randomized controlled trials (RCTs) related to maternal and neonatal clinical outcomes. The Revised Cochrane risk-of-bias tool for randomized trials was used to examine the risk of publication bias. Stata (version 15.1; StataCorp) was used to analyze the data.

Results: We incorporated 42 pertinent RCTs involving 148,866 participants. In comparison to the routine care group, GWG was markedly reduced in the intervention group (standardized mean difference–0.19, 95% CI –0.25 to –0.13; $P<.001$). A significant reduction was observed in the proportion of women with excessive weight gain (odds ratio [OR] 0.79, 95% CI 0.69-0.91; $P=.001$), along with an increase in the proportion of women with adequate weight gain (OR 1.33, 95% CI 1.10-1.64; $P=.003$). Although no significant difference was reported for the proportion of individuals below standardized weight gain, there is a significant reduction in the risk of miscarriage (OR 0.66, 95% CI 0.46-0.95; $P=.03$), preterm birth (OR 0.8, 95% CI 0.75-0.86; $P<.001$), as well as complex neonatal outcomes (OR 0.93, 95% CI 0.87-0.99; $P=.02$). Other maternal and fetal outcomes were not significantly different between the 2 groups (all $P>.05$).

Conclusions: The findings corroborate our hypothesis that digitally facilitated health care can enhance certain facets of maternal and neonatal outcomes, particularly by mitigating excessive weight and maintaining individuals within a reasonable weight gain range. Therefore, encouraging women to join the digital health team sounds feasible and helpful.

Trial Registration: PROSPERO CRD42024564331; https://tinyurl.com/5n6bshjt

(J Med Internet Res 2025;27:e66580) doi:10.2196/66580

XSL•FO
RenderX

KEYWORDS

digital health; telemedicine; telehealth; mobile health; mHealth; mobile phone; intervention; meta-analysis; pregnant women; systematic review

Introduction

Pregnancy is a unique physiological phase marked by significant physical, psychological, and behavioral changes that impact maternal and neonatal outcomes [1]. A key aspect of pregnancy is gestational weight gain (GWG), which plays a crucial role in maternal and infant health. However, studies indicate many women fail to meet the recommended GWG guidelines. In 2018, only 28%, 31%, and 32% of women in the United States, Europe, and Asia achieved the recommended weight gain during pregnancy [2]. This issue is even more prevalent in low- and middle-income countries. A 2023 study across 24 countries found that 55% (65,505/118,207) of participants experienced inadequate GWG, 23% (26,746/118,207) gained excessive weight, and only 22% (25,956/118,207) adhered to the recommended guidelines [3].

Maternal weight gain has a profound impact on pregnancy outcomes, including gestational complications, infant mortality, and long-term health for both mother and child [4]. Excessive GWG is linked to higher risks of complications such as large for gestational age (LGA), macrosomia, cesarean delivery, and postpartum weight retention [5-7]. On the other hand, insufficient GWG is associated with increased risks of miscarriage, infants who are small for gestational age (SGA), low birth weight, and preterm birth [8-10]. Therefore, promoting healthy gestational weight gain is crucial in reducing pregnancy complications and minimizing the risks of maternal and neonatal morbidity and mortality.

Digital health interventions, including applications, websites, digital programs, and other smart devices, have gained significant attention for their potential to enhance physical and mental well-being, particularly in low-resource settings such as Africa and South Asia [11]. Telemedicine involves using telecommunications technology to deliver clinical health care remotely, enabling health care providers to diagnose, treat, and monitor patients from a distance. Telehealth is a broader concept encompassing telemedicine and additional services, such as health education, disease prevention, and remote monitoring. Mobile health (mHealth) refers explicitly to using mobile devices like smartphones and tablets to deliver health care services, track health conditions, and promote healthy behaviors [12]. While these technologies overlap, each serves a distinct purpose, and together, they form key components of modern health care interventions.

Evidence suggests that technology-mediated interventions can be as effective or superior to routine care in improving maternal and neonatal health outcomes [13]. For example, a meta-analysis found that digital health interventions for mothers with gestational diabetes improved self-care, leading to better weight and glycemic control and lower rates of macrosomia and cesarean deliveries [14]. Another meta-analysis, combining data from 21 randomized controlled trials (RCTs) and controlled clinical trials, reported that web-based interventions significantly increased the likelihood of vaginal delivery while reducing emergency cesarean sections and neonatal complications. However, no improvements in glucose profiles were observed [15]. Furthermore, a study by He et al [16] demonstrated that mHealth interventions significantly decreased the incidence of gestational diabetes, preterm births, and macrosomia in pregnant women with overweight or obesity. In addition, participants in the intervention group gained 1.12 kg less than those in the routine care group [16]. However, some studies have contradicted these findings, reporting no significant impact of telemedicine on maternal or neonatal outcomes [17-19].

The contradictory findings underscore the need for further research into the effectiveness of digital health interventions, especially to evaluate their impact on pregnant women with varying risk profiles and refine strategies to improve maternal and neonatal health outcomes. To our knowledge, this is the first meta-analysis to broaden the participant scope, including not only high-risk groups, such as those with gestational diabetes, overweight, or obesity, but also low-risk or nonspecific pregnant women. This review aims to systematically evaluate studies investigating the impact of digital health interventions on maternal health outcomes, including gestational weight management and neonatal health outcomes in pregnant women.

Methods

Overview

The study protocol was preregistered in PROSPERO (CRD42024564331). The manuscript was structured following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [20], and the checklist is presented in Multimedia Appendix 1.

Search Strategy

The initial comprehensive literature search for this meta-analysis was conducted in February 2024 and updated in August 2024 to capture any newly published studies. The search spanned 4 major English-language databases: PubMed, Embase, Cochrane Library, and Web of Science. Keywords derived from relevant articles were used, including terms such as telemetry, digital health, e-consultation, telemonitoring, smartphone technology, online communication, and digital health technology. Details of the search strategy are provided in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

Eligible studies were required to be RCTs published in English, focusing on pregnant women aged 18 years and older. Digital health interventions, such as phone calls, text messages, and interactive apps (eg, YouTube, Twitter, and WeChat), were implemented in the intervention group, while the control group received standard care. The studies evaluated either maternal outcomes, such as GWG and pregnancy complications, or neonatal outcomes, such as preterm birth and SGA, defined as a birth weight at or below the 10th percentile for gestational age.

XSL•FO
RenderX

Studies were excluded if they did not include a control group, involved both intervention and control groups receiving digitally mediated treatments, failed to report the desired outcomes, measured them only postpartum, or lacked accessible full-text articles or usable data.

Study Selection and Data Extraction

Following the removal of duplicates, 2 researchers (JW and NT) independently screened the remaining articles by evaluating their titles and abstracts, excluding those that were irrelevant. The full texts of studies identified as potentially relevant were retrieved and further assessed to determine their eligibility for inclusion. In cases where discrepancies arose between the two researchers, these were resolved through discussion or, if necessary, by consulting a third researcher to achieve consensus.

For articles selected for further analysis, JW and NT used a standardized data extraction worksheet developed in Microsoft Excel 2016. The extracted data encompassed key study characteristics, including the year of publication, authors, country, sample size, inclusion and exclusion criteria, maternal and neonatal health outcomes, gestational age at enrollment, prepregnancy maternal BMI, type of digital applications, specific interventions, control measures, duration of intervention, high-risk factors of participants, and the effects of digital care on maternal and neonatal health. Any disagreements during the data extraction process were resolved through iterative discussions between the authors until a consensus was reached.

Evaluation of the Methodological Quality of the Studies

The bias of the RCTs included in this meta-analysis was evaluated using the Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [21]. JW and NT independently assessed the risk of bias for each included study. Discrepancies between the two reviewers were resolved through iterative discussions, revisiting study details, and assessment criteria to reach a consensus. If disagreements persisted, a third reviewer provided an independent evaluation, with the final decision based on majority agreement.

Data Analysis

Data analysis for the meta-analysis was performed using Stata (version 15.1; StataCorp). Effect sizes were calculated and presented as forest plots to facilitate quantitative synthesis. Standardized mean differences (SMDs) were used for continuous variables, while odds ratios (ORs) were applied for dichotomous outcomes. The choice between fixed-effect and random-effects models was determined by the level of heterogeneity, with I^2 values above 50% indicating substantial heterogeneity; a random-effects model was used for I^2>50%, while I^2≤50% warranted a fixed-effect model. A P value of <.05 was considered statistically significant. To assess the robustness of the synthesized results, sensitivity analyses were conducted by sequentially excluding each study and reanalyzing the data, as well as by restricting the analysis to studies with a low risk of bias. To assess a potential publication bias, funnel plots were used, and the Egger regression test was used to calculate the publication bias (Multimedia Appendix 3).

Results

Search Results

A total of 19,936 studies were retrieved from the 4 databases, and 1 additional article was identified through a manual search. After removing duplicates, 11,606 studies remained for further evaluation. Titles and abstracts were screened, and 69 studies were deemed relevant. Following a detailed review of the full texts, 42 RCTs met the inclusion criteria and were incorporated into the meta-analysis. Figure 1 shows the complete screening process.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the literature screening and selection process.

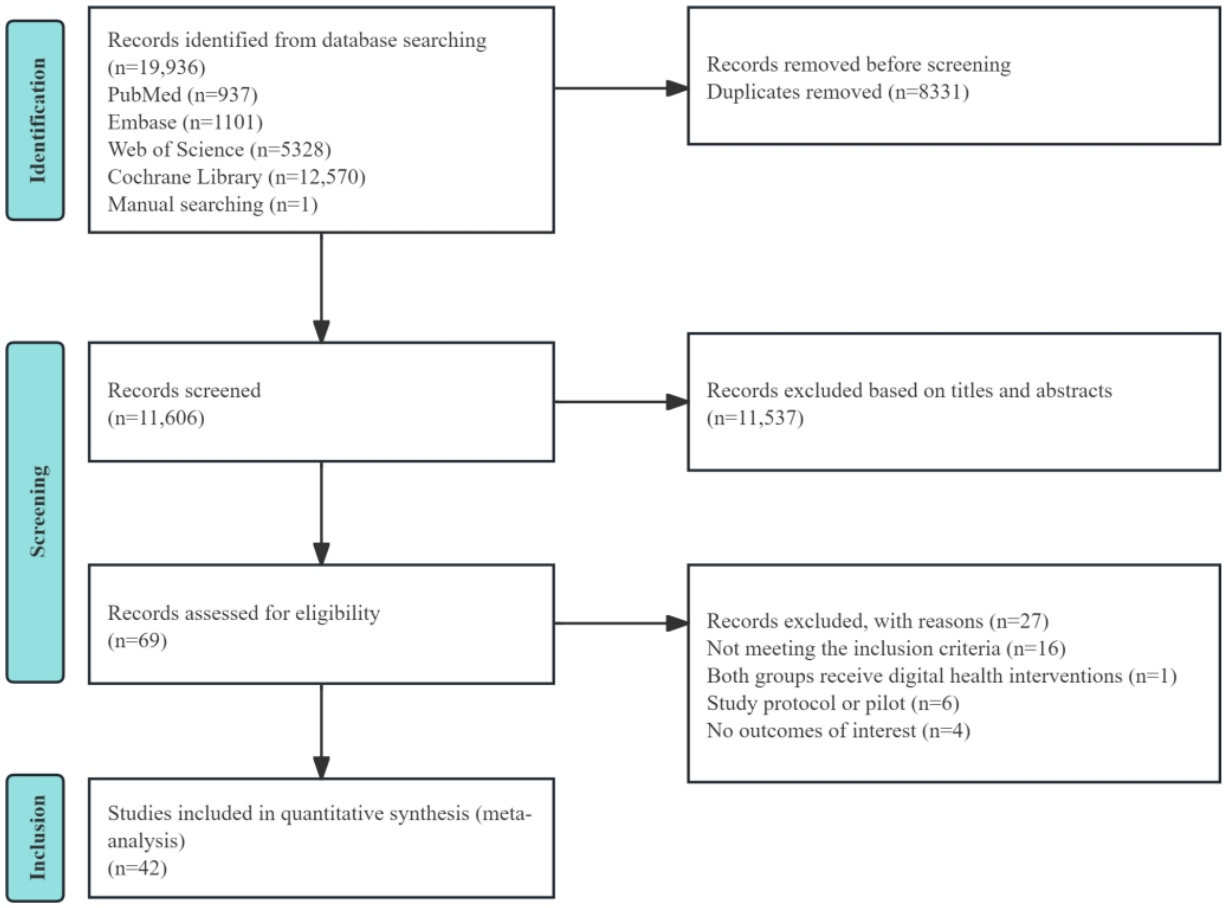

Study Characteristics

This meta-analysis includes data from 42 RCTs involving a total of 148,866 participants. Its primary focus is to evaluate the impact of digital health interventions on maternal and neonatal outcomes, particularly among populations with gestational diabetes mellitus (GDM) and other high-risk pregnancy factors. The studies, conducted between 2007 and 2024, were geographically distributed across Asia (13/42 studies, 31%), Europe (13/42 studies, 31%), North America (9/42 studies, 21%), Australasia (6/42 studies, 14%), and Africa (1/42 study, 3%). The majority of studies were published in the United States (9/42 studies, 21%), Australia (6/42 studies, 14%), and China (6/42 studies, 14%).

Digital health interventions used three primary delivery modalities: (1) mobile devices, including smartphones and tablets; (2) website-based platforms; and (3) mobile apps incorporating social software (eg, Facebook [Meta Platforms], Zoom [Zoom Communications, and WeChat [Tencent Holdings Limited]) and other digital health tools. Table 1 presents the detailed characteristics of the 42 RCTs [22-63].

XSL•FO
RenderX

Table 1. The characteristics of included studies.

Author, year	Country	Type of digital health	Duration of intervention	Sample size, n		High-risk factors
				IG[a]	CG[b]	
Homko et al [22], 2007	United States	Internet	To birth	32	25	GDM[c]
Pérez-Ferre et al [23], 2010	Spain	Mobile phone	To birth	49	48	GDM
Phelan et al [24], 2011	United States	Telephone	To birth	201	200	—[d]
Homko et al [25], 2012	United States	Internet	To birth	40	40	GDM
Given et al [26], 2015	United Kingdom	Internet	To birth	24	26	GDM
Durnwald et al [27], 2016	United States	Telephone	To birth	49	52	GDM
Herring et al [28], 2016	United States	SMS text message and telephone	To birth	33	33	Overweight or obesity
Smith et al [29], 2016	United States	Internet	To birth	24	21	—
Zairina et al [30], 2016	Australia	Mobile phone	To birth	36	36	Asthma
Willcox et al [31], 2017	Australia	Telephone and internet	To birth	45	46	—
Sagedal et al [32], 2017	Norway	Mobile phone	To 36 weeks gestation	296	295	—
Kennedy et al [33], 2018	Ireland	Internet	To birth	125	125	—
Kennelly et al [34], 2018	Ireland and the Netherlands	Mobile phone	To birth	278	287	GDM
Mackillop et al [35], 2018	United Kingdom	Mobile phone	To birth	101	102	GDM
Miremberg et al [36], 2018	Israel	Mobile phone	To birth	60	60	GDM
Patel et al [37], 2018	India	SMS text message and telephone	6 months after delivery	519	518	—
Rasekaba et al [38], 2018	Australia	Internet	To birth	61	34	GDM
Al-Ofi et al [39], 2019	Saudi Arabia	Telemonitoring device and SMS text message	6 weeks after delivery	27	30	GDM
Borgen et al [40], 2019	Norway	Mobile phone	To birth	115	123	GDM
Carolan-Olah and Sayakhot [41], 2019	Australia	Internet	To birth	52	58	GDM
Guo et al [42], 2019	China	Mobile phone	To birth	64	60	GDM
Sung et al [43], 2019	South Korea	Mobile phone	To birth	11	10	GDM
Butler Tobah et al [44], 2019	United States	Telemonitoring device and telephone	To birth	150	150	—
Ferrara et al [45], 2020	United States	Telephone	To 38 weeks gestation	199	195	Overweight or obesity
Huang et al [46], 2020	Australia	Internet	12 weeks	30	27	—
Tomyabatra [47], 2020	Thailand	Mobile phone	To birth	432	400	—
Ding et al [48], 2021	China	Mobile phone	To birth	104	111	Overweight or obesity
LeBlanc et al [49], 2021	United States	Telephone and internet	To birth	89	80	Overweight or obesity
Sandborg et al [50], 2021	Sweden	Mobile phone	6 months	152	153	—
Su et al [51], 2021	China	Internet	6 months	56	56	GDM
Sun and Lingying [52], 2021	China	Mobile phone	To birth	40	40	GDM
Tian et al [53], 2021	China	Mobile phone	To birth	133	136	GDM
Yew et al [54], 2021	Singapore	Telemedicine device and telephone	To birth	170	170	GDM
Gonzalez-Plaza et al [55], 2022	Spain	Mobile phone	To birth	78	72	Obesity
Uria-Minguito et al [56], 2022	Spain	Internet	To 38-39 weeks gestation	102	101	—

XSL•FO
RenderX

Author, year	Country	Type of digital health	Duration of intervention	Sample size, n		High-risk factors
				IG[a]	CG[b]	
Bekker et al [57], 2023	Netherlands	Telemonitoring device and telephone	To birth	100	100	—
Munda et al [58], 2023	Slovenia	Telemedicine device and video conferencing system	To birth	53	52	GDM
Sharma et al [59], 2023	India	Mobile phone	To birth	65	66	—
Skalecki et al [60], 2023	Australia	Telemonitoring device	To birth	13,771	12,628	—
Wakwoya et al [61], 2023	Ethiopia	SMS text message and telephone	To birth	163	163	—
Téoule et al [62], 2024	Germany	Mobile phone	To birth	49	48	—
Wang et al [63], 2024	China	Mobile phone	To birth	29	29	—

[a]IG: intervention group.

[b]CG: control group.

[c]GDM: gestational diabetes mellitus.

[d]Not applicable.

Quality Assessment Results of the Studies

A total of 25 RCTs included in this review, focusing on GWG, were evaluated using the RoB 2 tool. Among these, 11 studies were determined to have a low risk of bias, 3 were identified as high risk, and 11 presented some concerns regarding potential bias. Studies that focused on secondary outcomes, such as neonatal health, were not included in this assessment, as the RoB 2 evaluation was specifically applied to studies addressing the primary outcome of GWG. Figure 2 [22-25,27-29,31,32,34,42-46,48,50,52,54-56,58,61-63] shows the risk-of-bias assessment.

XSL•FO
RenderX

Figure 2. Risk-of-bias domains. ROB-2. RoB 2: Revised Cochrane risk-of-bias tool for randomized trials.

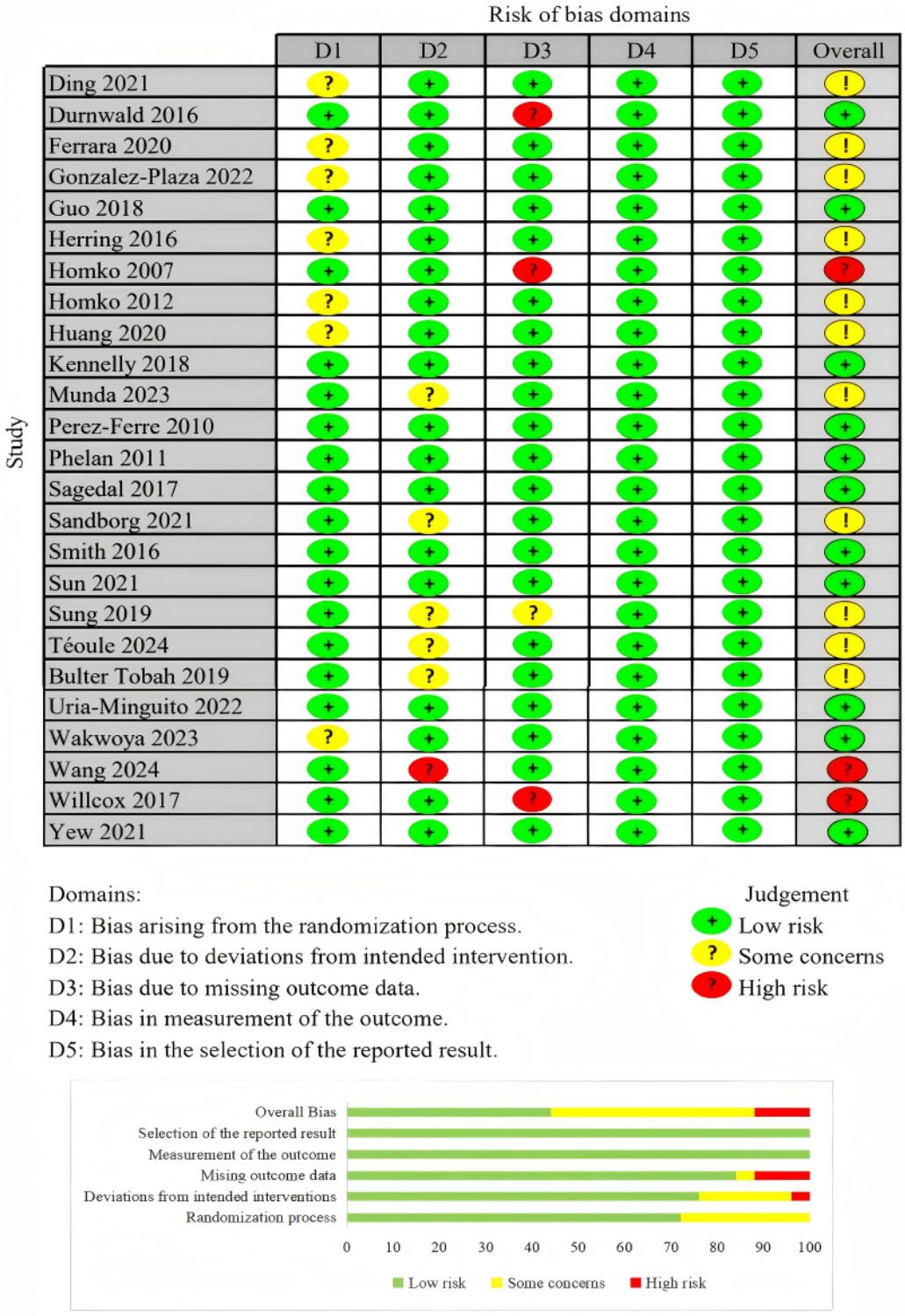

Meta-Analysis Results

Effect on GWG

GWG was analyzed in 25 studies involving 4315 participants. A pooled analysis using a random-effects model showed that digital health interventions effectively controlled GWG compared with routine care (I^2=54.2%; SMD –0.19, 95% CI –0.25 to –0.13; P<.001; Figure 3 [22-25,27-29,31,32,34,42-46,48,50,52,54-56,58,62,63]). Among these, 14 studies with 2675 participants reported the proportion of individuals exceeding the Institute of Medicine (IOM) recommendations for total weight gain during pregnancy, which are based on prepregnancy BMI categories for women: 12.5-18 kg for underweight (BMI<18.5 kg/m^2), 11.5-16 kg for normal weight (BMI 18.5-24.9 kg/m^2), 7-11.5 kg for overweight (BMI 25-29.9 kg/m^2), and 5-9 kg for obesity (BMI≥30 kg/m^2). In comparison, 10 studies with 1630 participants examined the proportion of women achieving sufficient weight gain according

XSL•FO
RenderX

to these recommendations. Both analyses showed no heterogeneity (I^2=0%, P=.45; I^2=0%, P=.85, respectively).

Figure 3. Effect on gestational weight gain. SMD: standardized mean difference.

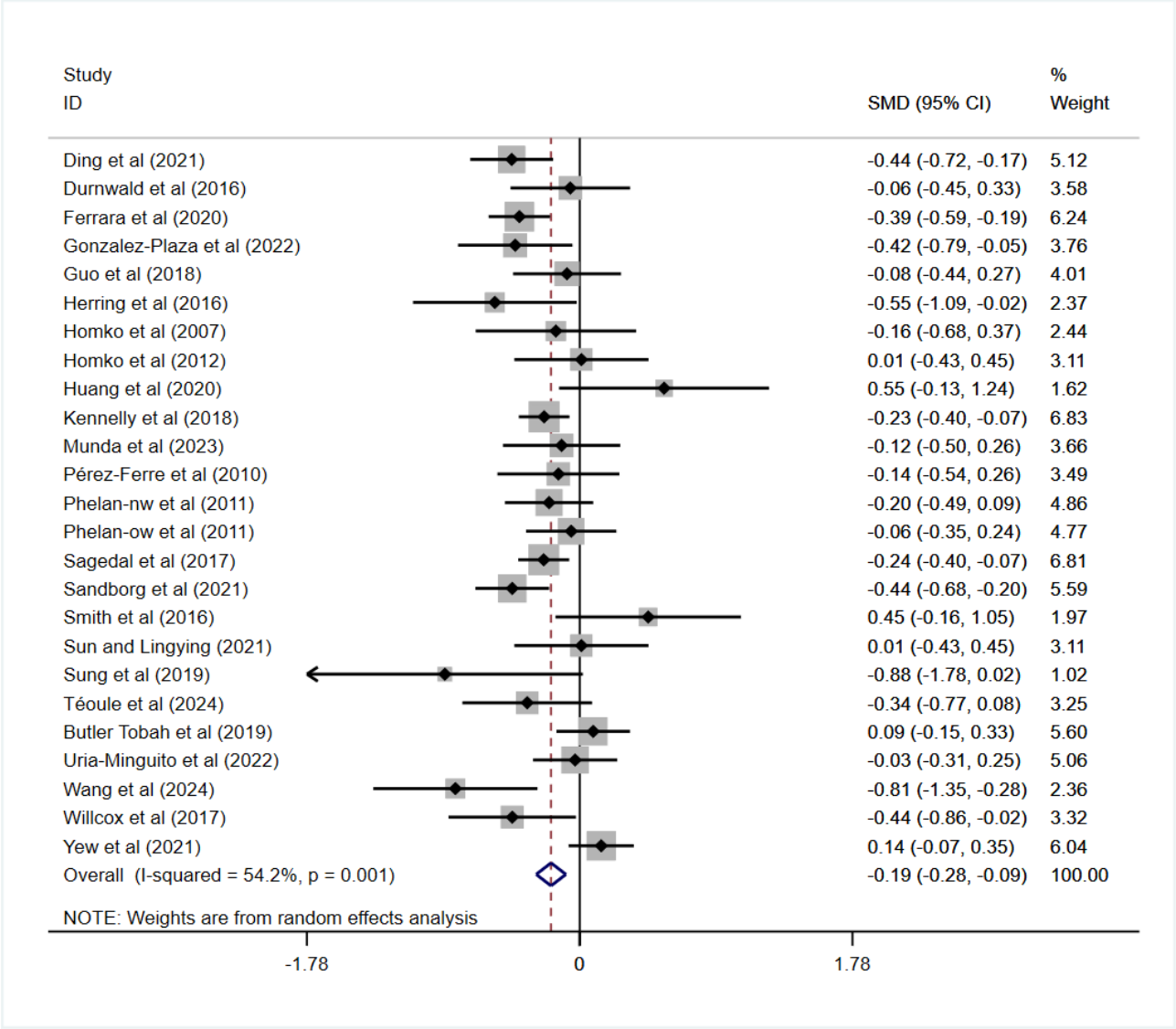

A fixed-effects model revealed a significant reduction in the proportion of women exceeding recommended GWG (OR 0.79, 95% CI 0.69-0.91; P=.001; Figure 4 [24,28,29,31,32,45,46,48-50,55,56,58,62]) and a significant increase in those meeting IOM GWG guidelines (OR 1.34, 95% CI 1.10-1.64; P=.003; Figure 5 [24,28,29,45,46,48,50,55,62,63]). However, no significant difference was observed in the proportion of participants falling below the IOM GWG guidelines.

XSL•FO
RenderX

Figure 4. Effect on excessive Institute of Medicine (IOM) total weight gain. OR: odds ratio.

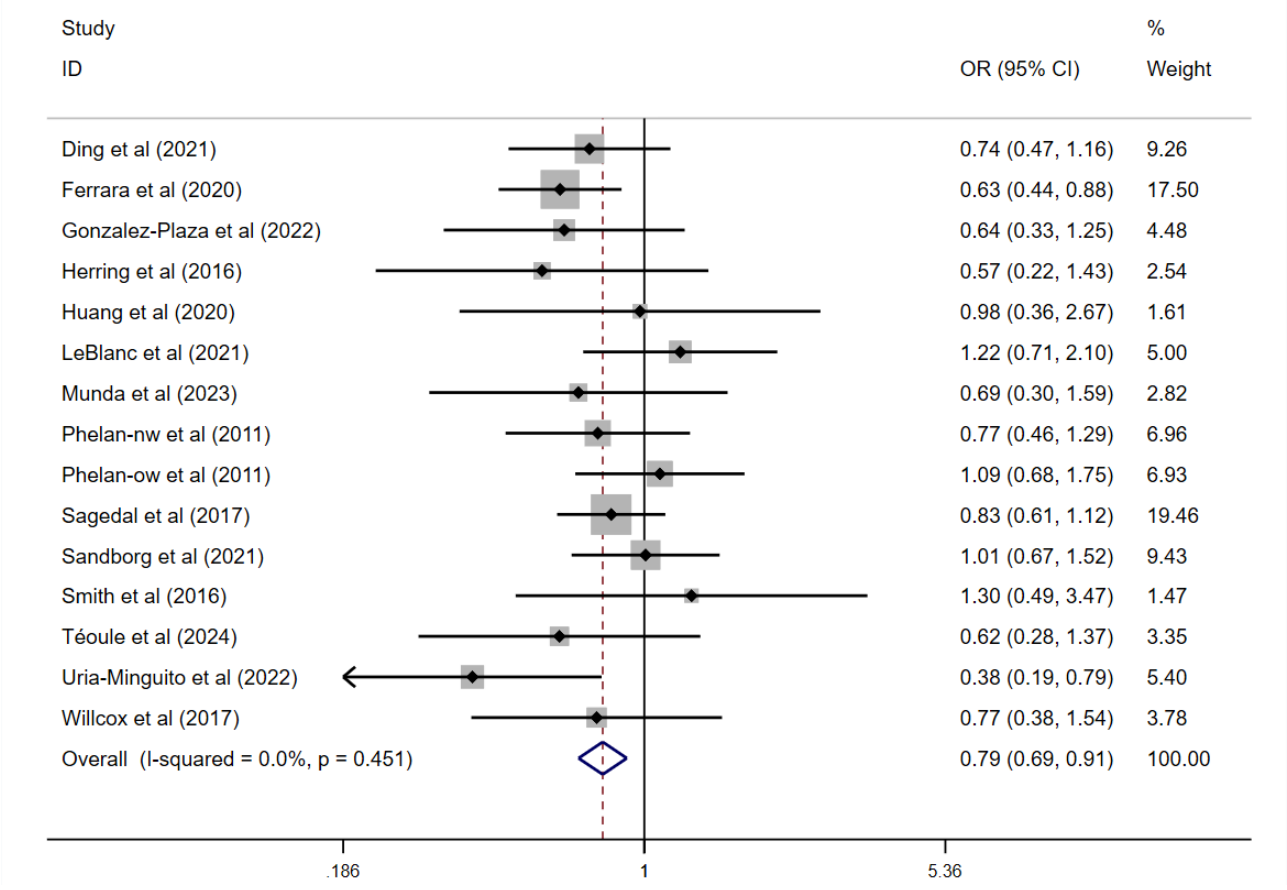

Figure 5. Effect on adequate Institute of Medicine (IOM) total weight gain. OR: odds ratio.

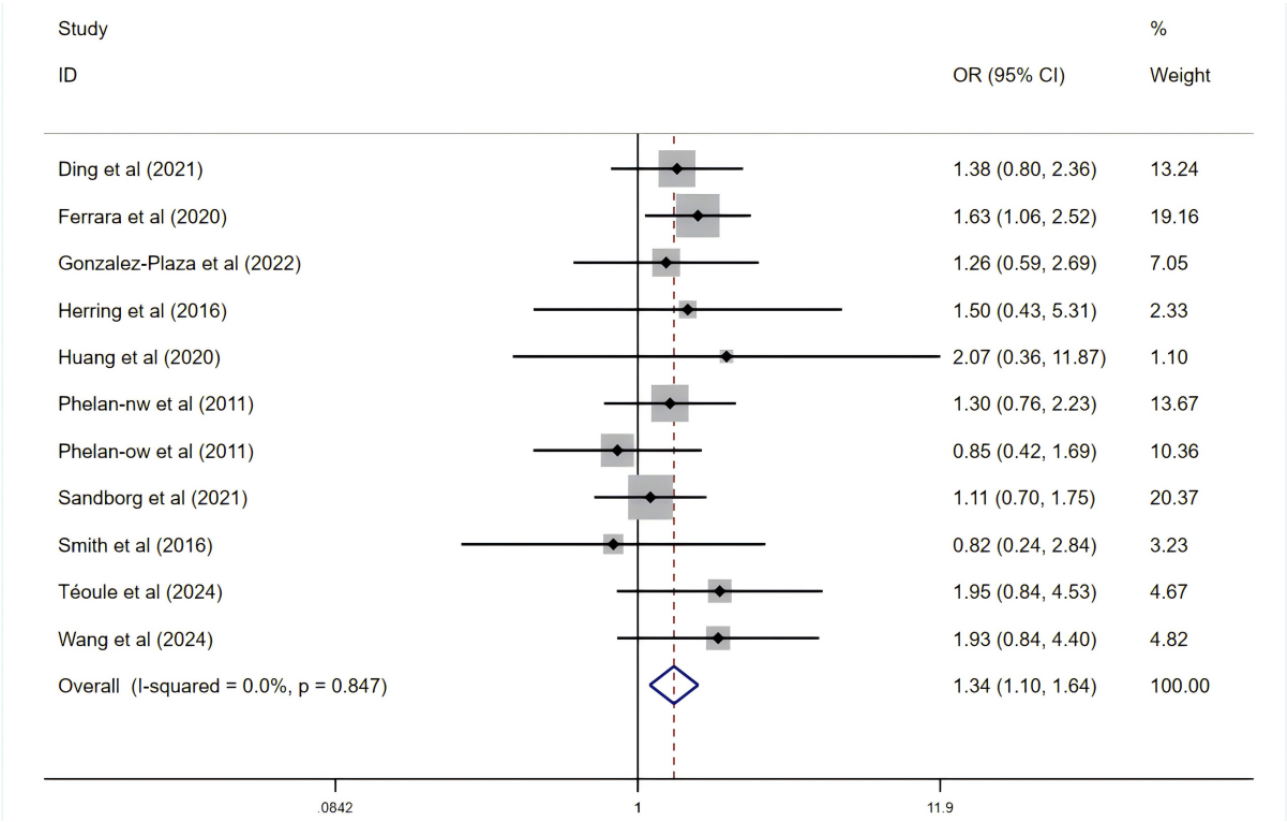

Subgroup analysis of 6 studies revealed that overweight or obese participants experienced a weight gain reduction of 0.348 kg compared with the control group (I^2=0%; SMD –0.35, 95% CI –0.45 to –0.24; P <.001). However, no significant difference

XSL•FO
RenderX

was observed between participants with GDM and those without high-risk conditions.

Effect on Delivery Mode

A pooled analysis of 34 studies (n=147,382) found no statistically significant difference in cesarean section rates between the intervention and control groups (OR 1.03, 95% CI 0.99-1.06; *P*=.12), with no heterogeneity detected. Similarly, data from 13 studies (n=4450) showed no significant impact on vaginal delivery rates (OR 1.05, 95% CI 0.95-1.15; *P*=.37), with no evidence of heterogeneity in these findings.

Effect on Gestational Age

Analysis of gestational week at delivery across 23 studies (n=5330) revealed high heterogeneity (I^2=94.8%; *P*<.001). Using a random-effects model, no statistically significant difference was observed between the intervention and control groups (SMD –0.004, 95% CI –0.27 to 0.26; *P*=.97).

Effect on Other Maternal Outcomes

Miscarriage was reported in 6 studies involving 142,385 participants. These studies detected no heterogeneity (I^2=0%; *P*=.66). A fixed-effects model revealed a statistically significant difference in miscarriage rates between the intervention and control groups (OR 0.66, 95% CI 0.46-0.95; *P*=.03; Figure 6 [32,37,44,49,55,60]). However, no significant differences were observed in the risk of shoulder dystocia, based on 4 studies (OR 0.35, 95% CI 0.12-1.02; *P*=.06), or in fasting blood glucose levels, analyzed in 11 studies (OR –0.16, 95% CI –0.32 to 0.01; *P*=.07).

Figure 6. Effect on miscarriages. OR: odds ratio.

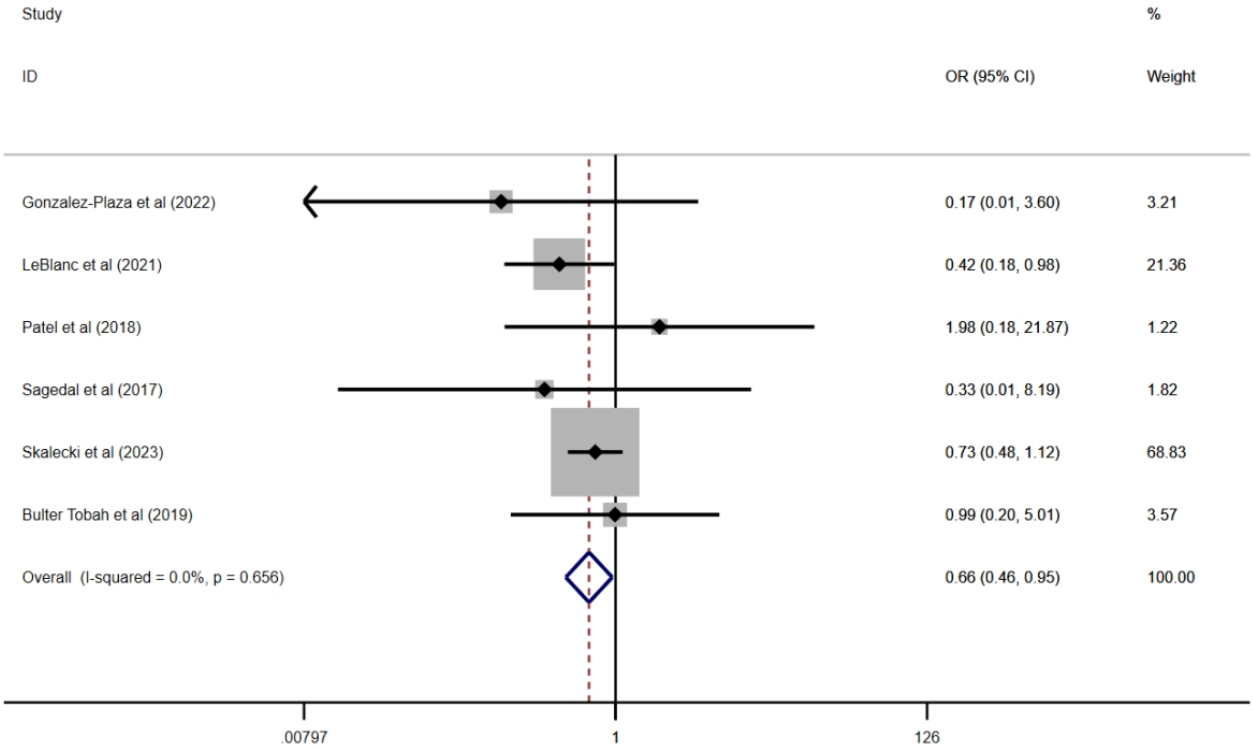

A total of 12 studies (n=2769) assessed the prevalence of gestational diabetes, while 20 studies (n=3398) evaluated the incidence of gestational hypertension or preeclampsia. Both outcomes demonstrated low heterogeneity (I^2=16.9%; *P*=.27 and I^2=7.2%; *P*=.34, respectively). Using the Mantel-Haenszel fixed-effects model, no significant reduction was found in the prevalence of gestational diabetes (OR 0.87, 95% CI 0.70-1.08; *P*=.20) or gestational hypertension or preeclampsia (OR 0.88, 95% CI 0.70-1.11; *P*=.27).

Effects on Preterm Birth

A pooled analysis of 22 studies involving 144,695 participants revealed a significantly lower prevalence of preterm births (before 37 weeks) among neonates in the intervention group compared to the control group (OR 0.80, 95% CI 0.75-0.86; *P*<.001; Figure 7 [22-26,30,32,35,44,45,47-49,51-55,57,58,60,62]). No heterogeneity was detected (I^2=0%; *P*=.90).

XSL•FO
RenderX

Figure 7. Effect on the prevalence of preterm birth. OR: odds ratio.

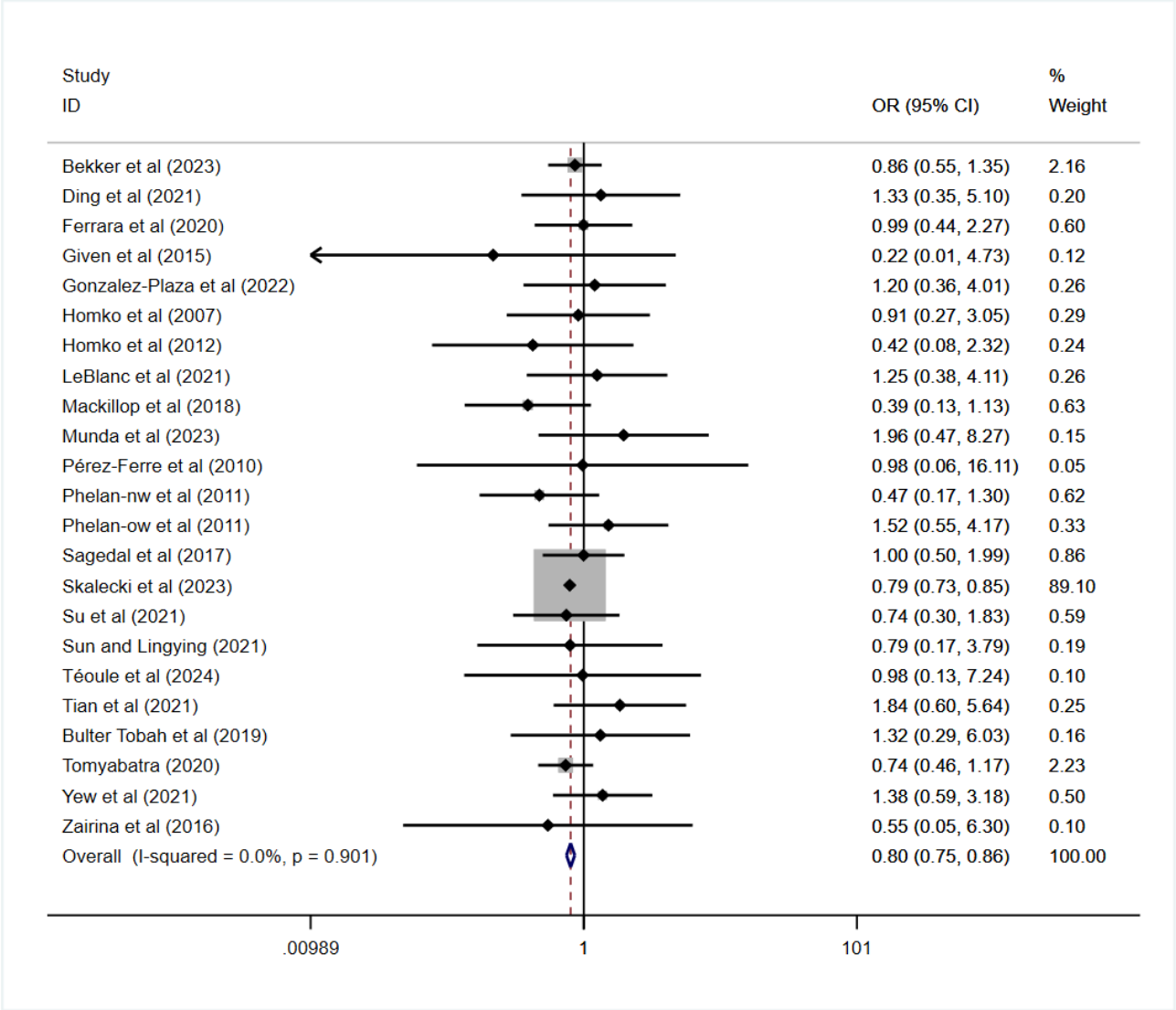

Effects on Infant Circumstance

A total of 26 studies assessed neonatal birth weight (I^2=39.6%; *P*=.02), and 6 examined birth length (I^2=30.6%; *P*=.21). Compared with the control group, computer-based health interventions showed no statistically significant differences in birth weight (SMD 0.02, 95% CI –0.06 to 0.09, *P*=.71) or birth length (SMD –0.06, 95% CI –0.21 to 0.10; *P*=.48). Furthermore, 5 studies evaluated infant head circumference, but no significant difference was observed between the groups (SMD 0.02, 95% CI –0.10 to 0.14; *P*=.74).

Effects on SGA and LGA

A total of 11 RCTs involving 142,303 participants evaluated the incidence of SGA and showed no heterogeneity among the studies (I^2=0%; *P*=.90). A fixed-effects model revealed a significant difference in SGA incidence between the intervention and control groups (OR 1.18, 95% CI 1.14-1.22; *P*<.001; Figure 8 [28,30,32,34,43,45,49,55,57,58,60]). In contrast, 13 studies (n=2403) assessed the incidence of LGA, also without heterogeneity, but the overall effect for LGA was not statistically significant (OR 0.91, 95% CI 0.69-1.19; *P*=.48).

XSL•FO
RenderX

Figure 8. Effect on the prevalence of small for gestational age (SGA). OR: odds ratio.

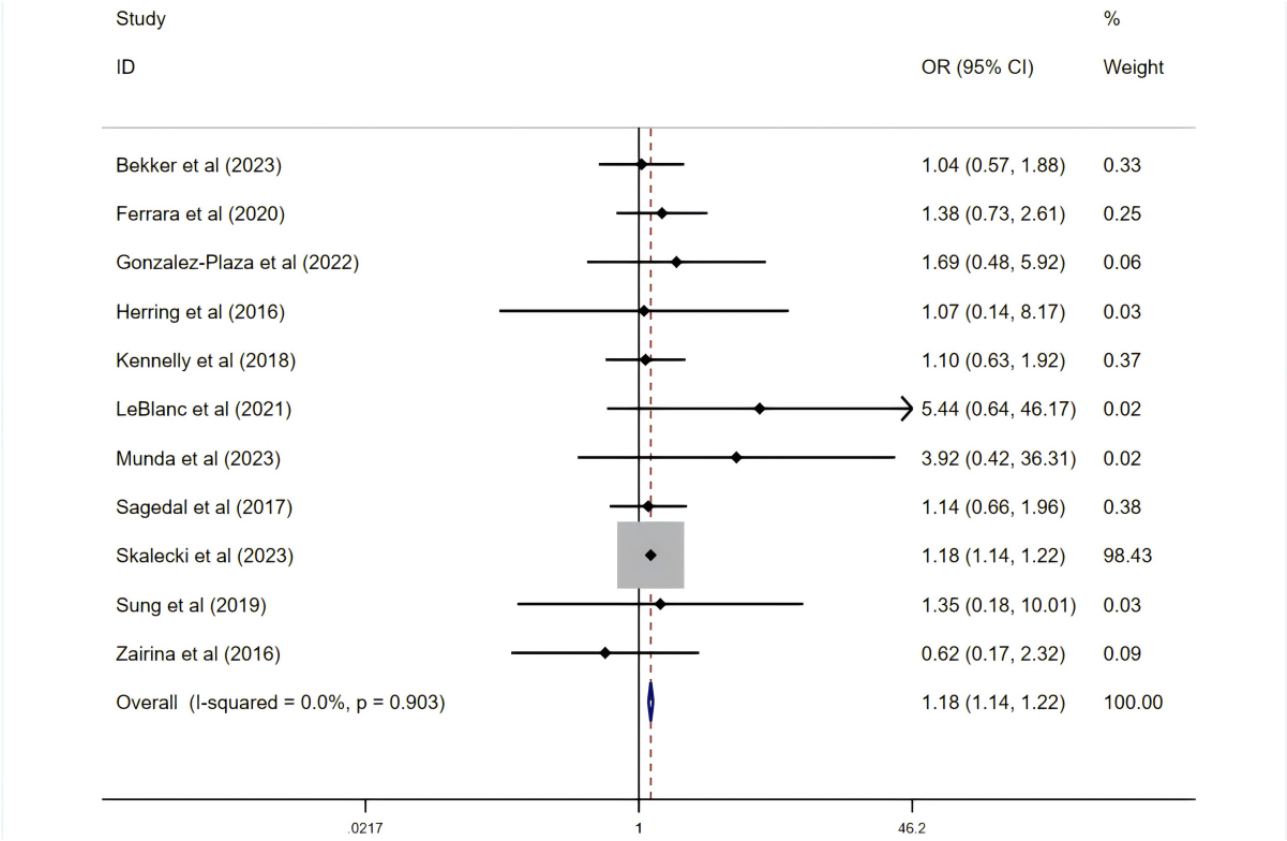

Effects on Neonatal Complications

In relation to neonatal complications, 12 studies addressed the issue of neonatal hypoglycemia (I^2=0%; P=.94), 16 studies reviewed ICU admissions (I^2=15.6%; P=.28), 8 studies mentioned jaundice or hyperbilirubinemia (I^2=0%; P=.94), and 8 studies assessed respiratory distress syndrome (RDS; I^2=4.6%; P=.39). For each condition, the intervention group exhibited no statistically significant decrease in incidence. However, pooled results from 6 studies involving 140,762 participants indicated that the digital health group experienced a significant decrease in combined complications (OR 0.93, 95% CI 0.87-0.98; P=.02; Figure 9 [22,36,54,58,60,62]), with no heterogeneity detected (I^2=0%; P=.76).

XSL•FO
RenderX

Figure 9. Effect on the prevalence of composite neonatal complications. OR: odds ratio.

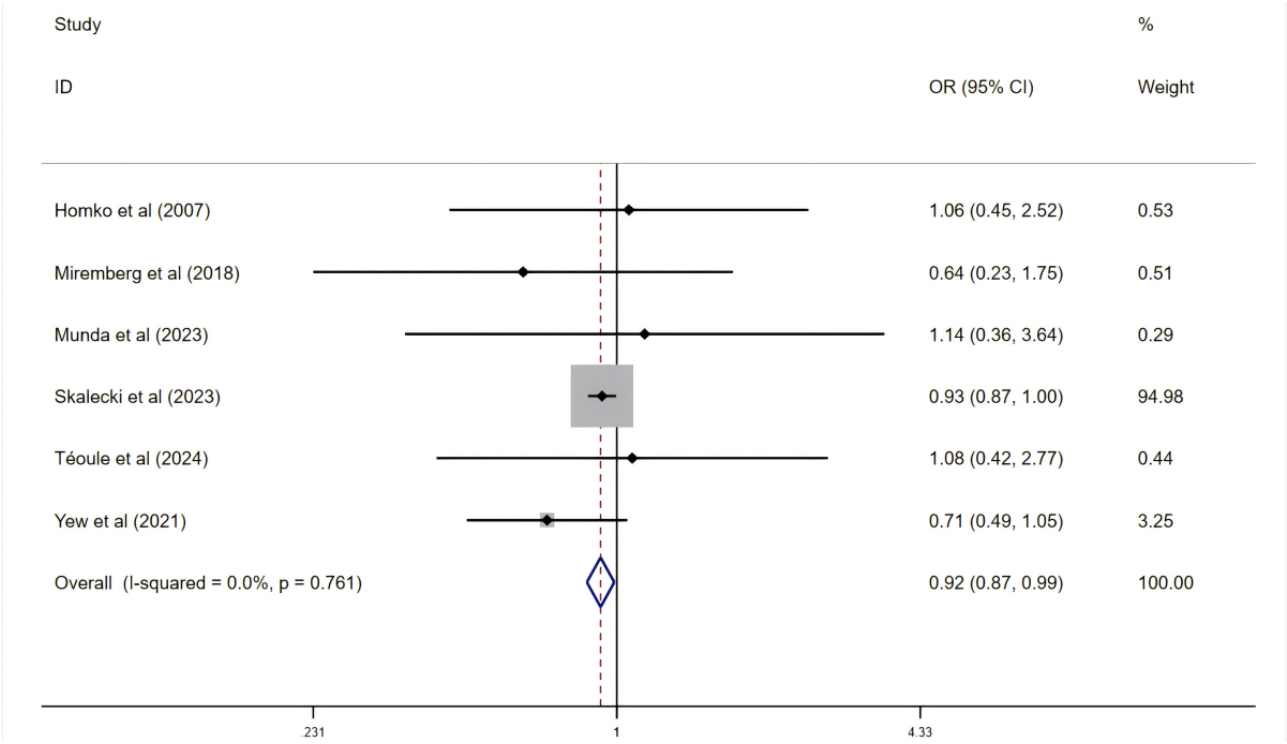

Effects on Other Neonatal Outcomes

Based on 5 studies involving 481 newborns, the Apgar score analysis revealed no statistically significant difference between the intervention and control groups (SMD –0.11, 95% CI –0.29 to 0.07; *P*=.25). Similarly, 13 studies and 11 studies examined whether mHealth interventions reduced the risk of macrosomia (birth weight≥4000 g) and low birth weight (<2500 g), respectively, with no heterogeneity detected. Compared with routine care, the intervention group showed no significant improvement in the incidence of macrosomia (OR 0.90, 95% CI 0.72-1.14; *P*=.39) or low birth weight (OR 1.00; 95% CI 0.80-1.24; *P*=.97).

The results of the meta-analysis for 26 outcomes are summarized in Table 2.

XSL•FO
RenderX

Table 2. Meta-analysis results of 26 outcomes.

Outcomes	Studies, n	Participants, n		Heterogeneity		Statistical method	Effect estimate
		IG[a]	CG[b]	I^2 (%)	*P* value		
GWG[c]	25	2155	2160	54.2	<.001	SMD[d] (inverse variance, random, 95% CI)	–0.19 (–0.25 to –0.13)
Proportion of exceeding IOM[e] GWG	14	1340	1335	0	.45	OR[f] (Mantel-Haenszel, fixed, 95% CI)	0.79 (0.69-0.91)
Proportion of meeting IOM GWG	10	812	818	0	.85	OR (Mantel-Haenszel, fixed, 95% CI)	1.34 (1.10-1.64)
Proportion of below IOM GWG	9	786	797	19.7	.26	OR (Mantel-Haenszel, fixed, 95% CI)	1.19 (0.91-1.55)
Gestational diabetes	12	1380	1389	16.9	.27	OR (Mantel-Haenszel, fixed, 95% CI)	0.87 (0.70-1.08)
Gestational hypertension or preeclampsia	20	1704	1694	7.2	.37	OR (Mantel-Haenszel, fixed, 95% CI)	0.88 (0.70-1.11)
Miscarriage	6	14,947	127,438	0	.66	OR (Mantel-Haenszel, fixed, 95% CI)	0.66 (0.46-0.95)
Cesarean delivery	34	17,467	129,915	0	.71	OR (Mantel-Haenszel, fixed, 95% CI)	1.03 (0.99-1.06)
Vaginal delivery	14	2247	2203	0	.95	OR (Mantel-Haenszel, fixed, 95% CI)	1.05 (0.95-1.15)
Shoulder dystocia	4	469	471	20.1	.29	OR (Mantel-Haenszel, fixed, 95% CI)	0.35 (0.12-1.02)
Gestational age at delivery	23	2696	2634	94.8	<.001	SMD (inverse variance, random, 95% CI)	–0.004 (–0.27 to 0.26)
Fasting blood glucose	11	788	773	57.4	.009	SMD (inverse variance, random, 95% CI)	–0.16 (–0.32 to 0.01)
Preterm birth	22	16,115	128,580	0	.90	OR (Mantel-Haenszel, fixed, 95% CI)	0.80 (0.75-0.86)
Infant birth weight	26	2681	2618	39.6	.02	SMD (inverse variance, fixed, 95% CI)	0.02 (–0.06 to 0.09)
Infant birth length	6	602	608	30.6	.21	SMD (inverse variance, fixed, 95% CI)	–0.06 (–0.21 to 0.10)
Infant head circumference	5	563	563	0	.64	SMD (inverse variance, fixed, 95% CI)	0.02 (–0.10 to 0.14)
Apgar score	5	241	240	0	.51	SMD (inverse variance, fixed, 95% CI)	–0.11 (–0.29 to 0.07)
SGA[g] (≤10%)	11	14,903	127,400	0	.90	OR (Mantel-Haenszel, fixed, 95% CI)	1.18 (1.14-1.22)
LGA[h] (≥90%)	13	1208	1195	0	.73	OR (Mantel-Haenszel, fixed, 95% CI)	0.91 (0.69-1.19)
Macrosomia	17	1707	1692	0	.76	OR (Mantel-Haenszel, fixed, 95% CI)	0.90 (0.72-1.14)
Birth weight ＜2500 g	10	2021	1965	0	.56	OR (Mantel-Haenszel, fixed, 95% CI)	1.00 (0.80-1.24)
Neonatal hypoglycemia	12	784	755	0	.94	OR (Mantel-Haenszel, fixed, 95% CI)	0.87 (0.65-1.15)
ICU[i] admission	16	15,464	127,924	15.6	.28	OR (Mantel-Haenszel, fixed, 95% CI)	1.00 (0.93-1.08)
Jaundice or hyperbilirubinemia	8	505	499	0	.94	OR (Mantel-Haenszel, fixed, 95% CI)	0.89 (0.61-1.30)
RDS[j]	8	873	828	4.6	.39	OR (Mantel-Haenszel, fixed, 95% CI)	0.71 (0.49-1.03)
Neonatal composite outcome	6	14,132	126,630	0	.76	OR (Mantel-Haenszel, fixed, 95% CI)	0.93 (0.87-0.99)

[a]IG: intervention group.
[b]CG: control group
[c]GWG: gestational weight gain.
[d]SMD: standardized mean difference.
[e]IOM: Institute of Medicine.
[f]OR: odds ratio.
[g]SGA: small for gestational age.
[h]LGA: large for gestational age.
[i]ICU: intensive care unit.
[j]RDS: respiratory distress syndrome.

XSL•FO
RenderX

Discussion

Principal Findings

This meta-analysis revealed that digital health interventions significantly improved excessive GWG, reduced miscarriage and preterm birth incidence, and enhanced neonatal outcomes. However, the benefits for women with insufficient weight gain were limited, and an increased rate of infants who are SGA was observed.

Comparison With Previous Work

This review evaluated the impact of digital health interventions on maternal and neonatal outcomes, with a particular focus on GWG. The findings revealed that digital interventions significantly improved GWG management among pregnant women. Compared to the control group, the intervention group showed a notable decrease in women exceeding recommended weight gain and increased adherence to the IOM GWG guidelines (P<.05). These results align with previous meta-analyses by Islam et al [64] and He et al [16]. However, the reduction in weight gain observed in our study (–0.145 kg) was less pronounced than the reductions reported by Islam et al [64] (–1.07 kg) and Antoun et al [65] (–1.99 kg). This discrepancy may be attributed to differences in the types and intensity of digital interventions and participant characteristics.

Subgroup analysis further revealed that significant GWG reductions were predominantly observed in participants who were overweight or obese rather than those with GDM. This may be explained by the intensive medical management typically provided to patients with GDM, which may diminish the additional benefits of digital interventions. These findings underscore the potential value of digital health care for individuals with prepregnancy overweight or obesity, as effective weight control in these populations is crucial. Excess weight not only increases the risk of pregnancy complications but also poses significant long-term health risks for their children [66].

The findings did not align with our hypothesis that digital health interventions would benefit participants experiencing insufficient GWG. This discrepancy may stem from the complex causes of inadequate weight gain during pregnancy, which include factors such as prepregnancy anemia, gestational diabetes, unhealthy lifestyle behaviors (eg, substance abuse), parity, and crowded living conditions [67]. Addressing these issues requires more than education on healthy diets, self-monitoring, or food supplementation alone. Additionally, 4 of the 9 studies on insufficient GWG focused on women with overweight and obese, with the digital health interventions targeting weight loss rather than promoting weight gain. This lack of individualized weight management strategies may have limited the effectiveness of the interventions in supporting appropriate weight gain for women below the recommended range.

In our study, while digital health interventions did not address insufficient GWG, they were associated with a reduction in miscarriage and preterm birth rates, as well as an increase in the prevalence of infants who are SGA. The reduction in miscarriage aligns with findings from Victa et al [68], which demonstrated that telemedicine monitoring during pregnancy significantly lowers miscarriage risk by enabling early detection of potential complications and timely medical intervention. In addition, the decreased risk of preterm birth and neonatal composite outcomes (P<.05) is consistent with the study by Guo et al [15], which highlighted the positive impact of digital health interventions on neonatal health.

Although insufficient GWG is known to increase the risk of miscarriage and preterm birth, our findings suggest that digital interventions may mitigate these risks through mechanisms beyond weight control. These interventions likely promote comprehensive health management by emphasizing nutritional intake, lifestyle improvements, and self-management efficacy, thereby improving maternal health and reducing risks associated with miscarriage and preterm labor, even without significantly addressing inadequate weight gain [69].

Further supporting this perspective, our analysis revealed a significant reduction in composite neonatal complications (P<.05), including conditions such as hypoglycemia, jaundice, and acute respiratory distress syndrome. By focusing on combined outcomes, the analysis provides a holistic view of the intervention's benefits, which aligns more closely with real-world clinical scenarios and underscores the broader potential of digital health interventions to improve maternal and neonatal health.

Pooling data revealed an unexpected correlation between digital health interventions and a higher incidence of infants who are SGA. This outcome may be attributed to the characteristics of the study population, as 9 out of the 11 studies on SGA involved participants with high-risk factors—5 focused on overweight and obesity, 2 on GDM, and 2 on other conditions. Interventions targeting weight control likely resulted in overly strict weight gain or dietary restrictions, increasing the energy gap [70] and potentially impairing placental development [71]. In addition, including participants already at substantial risk for SGA-related conditions may have amplified the negative effects of these interventions. Moreover, the inclusion of participants already at substantial risk for SGA-related conditions may have amplified the observed effects, complicating the interpretation of the actual impact of digital health interventions on SGA. While subgroup analyses could provide more nuanced insights, the variability and limited availability of stratified BMI data across studies hindered such analyses. Future studies should prioritize detailed BMI stratification and reporting to better elucidate the differential impacts of digital health interventions on SGA incidence.

Despite this, digital health interventions offer significant benefits for women and newborns through multiple mechanisms. They provide evidence-based, tailored educational resources that address women's specific needs [72,73] while encouraging active involvement in health monitoring. These tools enable women to track food intake, physical activity, and physical parameters such as weight changes and blood pressure [74], with positive feedback reinforcing their ability to manage their health effectively. Moreover, some intervention platforms facilitate online communication with skilled clinical and nursing personnel, reducing pregnancy-related stress and enhancing

XSL•FO
RenderX

health literacy [75]. These multifaceted approaches contribute to improved maternal and neonatal outcomes, even when specific challenges persist.

Strengths and Limitations

The strengths of this study encompass its inclusive population, as it considered all pregnant women rather than concentrating exclusively on specific subgroups such as those with diabetes or obesity, thereby enhancing the generalizability of the findings. In addition, we used a wide range of outcome measures, providing a comprehensive understanding of the impact of digital health interventions on maternal and neonatal health. Furthermore, our study incorporated the most recent literature, ensuring relevance to current clinical practice.

However, the study also has limitations. First, we expanded the search terms to include a broader range of eligible studies, thus enabling a more comprehensive investigation. Nevertheless, the quantity of published RCTs was restricted, indicating that the findings of our study necessitate careful interpretation and require further validation through supplementary research. Second, the follow-up period for the majority of the included studies concluded at birth, indicating a need for further investigation into the long-term effects on women and children to understand the implications of our findings fully. Third, the large sample size in the study by Skalecki et al [60], which focused on telemonitoring fetal movement, may influence the overall results. While a larger sample size is often beneficial for increasing the statistical power of a study, it may also introduce a bias toward the generalizability of the findings, especially in the context of the intervention types and outcome measures. A more specific analysis of how large sample sizes in studies with different intervention types could affect the overall results and the generalizability of the findings would provide a more comprehensive perspective. Fourth, the included studies encompassed a heterogeneous population of pregnant women with varying BMI categories (eg, underweight, normal weight, overweight, and obesity). However, some studies did not stratify participants by prepregnancy BMI or provide detailed subgroup analyses. This lack of consistent BMI reporting limits our ability to interpret the differential effects of digital health interventions across BMI subgroups. Future studies with standardized BMI stratification and detailed subgroup reporting are essential to delineate these effects better. Fifth, while the reduction in GWG observed in our study has general clinical significance due to the known risks of excessive weight gain, the lack of specific data on other contextual factors, such as dietary habits, physical activity levels, and socioeconomic status, may affect the interpretation of the results. These factors could mediate the relationship between digital health interventions and GWG, highlighting the need for more comprehensive datasets in future research.

Further Research

Further studies should investigate the potential differential impacts of these interventions across various populations, considering factors such as age, ethnicity, and health conditions, and determine the most effective digital intervention tailored to participants' backgrounds.

Conclusion

Digital health interventions facilitate maternal weight gain, which later influences neonatal health outcomes by decreasing complications such as miscarriages, preterm birth, and combined neonatal complications.

Acknowledgments

The authors acknowledge the support of the Medical Research Project of the Chongqing Health Commission (2024WSJK107), the Chongqing Education Commission Humanities and Social Sciences Research Project (23SKGH041), and the Chongqing Health Center for Women and Children in funding this research. The authors used ChatGPT (OpenAI) [76] to improve the language of this manuscript and acknowledge its use as a reference tool. However, after using this tool, authors meticulously reviewed and edited content as necessary and fully took responsibility for the content of the publication.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

NX and XZ have contributed equally to this work and share cocorresponding authorship. NX and XZ initiated the study and secured funding. JW and NT conducted the systematic literature search as the first and second reviewers, screened studies for eligibility, extracted relevant data, and assessed the risk of bias in studies. SL served as the third reviewer to resolve discrepancies between JW and NT through discussion and consensus. CJ and QJ managed data entry and provided statistical expertise. JW conducted data analysis and drafted the initial manuscript. JY and XZ critically reviewed and revised the manuscript. All authors contributed significantly to the conception and design of the study, participated in the data interpretation, and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 78 KB - jmir_v27i1e66580_app1.doc]

Multimedia Appendix 2
Details of the search.
[DOC File , 130 KB - jmir_v27i1e66580_app2.doc]

Multimedia Appendix 3
Publication bias.
[DOC File , 312 KB - jmir_v27i1e66580_app3.doc]

References

1. Lee VV, Vijayakumar S, Ng WY, Lau NY, Leong QY, Ooi DSQ, et al. Personalization and localization as key expectations of digital health intervention in women pre- to post-pregnancy. NPJ Digit Med 2023 Sep 30;6(1):183 [FREE Full text] [doi: 10.1038/s41746-023-00924-6] [Medline: 37775533]
2. Goldstein RF, Abell SK, Ranasinha S, Misso ML, Boyle JA, Harrison CL, et al. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med 2018 Aug 31;16(1):153 [FREE Full text] [doi: 10.1186/s12916-018-1128-1] [Medline: 30165842]
3. Perumal N, Wang D, Darling AM, Liu E, Wang M, Ahmed T, et al. Suboptimal gestational weight gain and neonatal outcomes in low and middle income countries: individual participant data meta-analysis. BMJ 2023 Sep 21;382:e072249 [FREE Full text] [doi: 10.1136/bmj-2022-072249] [Medline: 37734757]
4. Champion ML, Harper LM. Gestational weight gain: update on outcomes and interventions. Curr Diab Rep 2020 Feb 27;20(3):11 [FREE Full text] [doi: 10.1007/s11892-020-1296-1] [Medline: 32108283]
5. Darling AM, Wang D, Perumal N, Liu E, Wang M, Ahmed T, et al. Risk factors for inadequate and excessive gestational weight gain in 25 low- and middle-income countries: An individual-level participant meta-analysis. PLoS Med 2023 Jul;20(7):e1004236 [FREE Full text] [doi: 10.1371/journal.pmed.1004236] [Medline: 37486938]
6. Yin B, Hu L, Wu K, Sun Y, Meng X, Zheng W, et al. Maternal gestational weight gain and adverse pregnancy outcomes in non-diabetic women. J Obstet Gynaecol 2023 Dec;43(2):2255010 [FREE Full text] [doi: 10.1080/01443615.2023.2255010] [Medline: 37670680]
7. Hutchins F, Krafty R, El Khoudary SR, Catov J, Colvin A, Barinas-Mitchell E, et al. Gestational weight gain and long-term maternal obesity risk: a multiple-bias analysis. Epidemiology 2021 Mar 01;32(2):248-258 [FREE Full text] [doi: 10.1097/EDE.0000000000001310] [Medline: 33284167]
8. Chen Y, Liu M, Zhang Y, Chen Z, Mei H, Liu Y, et al. Gestational weight gain and neonatal outcomes in different zygosity twins: a cohort study in Wuhan, China. BMJ Open 2023 Jan 10;13(1):e056581 [FREE Full text] [doi: 10.1136/bmjopen-2021-056581] [Medline: 36627159]
9. Perumal N, Wang D, Darling AM, Wang M, Liu E, Urassa W, et al. Associations between gestational weight gain adequacy and neonatal outcomes in Tanzania. Ann Nutr Metab 2022;78(3):156-165 [FREE Full text] [doi: 10.1159/000522197] [Medline: 35124672]
10. Eick SM, Welton M, Claridy MD, Velasquez SG, Mallis N, Cordero JF. Associations between gestational weight gain and preterm birth in Puerto Rico. BMC Pregnancy Childbirth 2020 Oct 07;20(1):599 [FREE Full text] [doi: 10.1186/s12884-020-03292-1] [Medline: 33028249]
11. Western MJ, Armstrong MEG, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 09;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
12. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187 [FREE Full text] [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
13. Snoswell CL, Chelberg G, de Guzman KR, Haydon HH, Thomas EE, Caffery LJ, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare 2023 Oct;29(9):669-684 [FREE Full text] [doi: 10.1177/1357633X211022907] [Medline: 34184580]
14. Leblalta B, Kebaili H, Sim R, Lee SWH. Digital health interventions for gestational diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials. PLOS Digit Health 2022 Feb;1(2):e0000015 [FREE Full text] [doi: 10.1371/journal.pdig.0000015] [Medline: 36812531]
15. Guo P, Chen D, Xu P, Wang X, Zhang W, Mao M, et al. Web-based interventions for pregnant women with gestational diabetes mellitus: systematic review and meta-analysis. J Med Internet Res 2023 Jan 19;25:e36922 [FREE Full text] [doi: 10.2196/36922] [Medline: 36656629]

XSL•FO
RenderX

16. He Y, Huang C, He Q, Liao S, Luo B. Effects of mHealth-based lifestyle interventions on gestational diabetes mellitus in pregnant women with overweight and obesity: systematic review and meta-analysis. JMIR Mhealth Uhealth 2024 Jan 17;12:e49373 [FREE Full text] [doi: 10.2196/49373] [Medline: 38231555]
17. Kim SH, Park JH, Jung SY, de Gagne JCJJP. Internet-based interventions for preventing premature birth among pregnant women: systematic review. JMIR Pediatr Parent 2024 Apr 02;7:e54788 [FREE Full text] [doi: 10.2196/54788] [Medline: 38564247]
18. Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim KJD. Telemedicine interventions for gestational diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2015 Oct;110(1):1-9 [FREE Full text] [doi: 10.1016/j.diabres.2015.07.007] [Medline: 26264410]
19. Milne-Ives M, Lam C, de Cock C, van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
20. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015 Jun 02;162(11):777-784 [FREE Full text] [doi: 10.7326/M14-2385] [Medline: 26030634]
21. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
22. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306 [FREE Full text] [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
23. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
24. Phelan S, Phipps MG, Abrams B, Darroch F, Schaffner A, Wing RR. Randomized trial of a behavioral intervention to prevent excessive gestational weight gain: the Fit for Delivery Study. Am J Clin Nutr 2011 Apr;93(4):772-779 [FREE Full text] [doi: 10.3945/ajcn.110.005306] [Medline: 21310836]
25. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
26. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-Mum: a feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888 [FREE Full text] [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
27. Durnwald C, Kallan M, Allison K, Sammel M, Wisch S, Elovitz M, et al. A randomized clinical trial of an intensive behavior education program in gestational diabetes mellitus women designed to improve glucose levels on the 2-hour oral glucose tolerance test. Am J Perinatol 2016 Oct;33(12):1145-1151 [FREE Full text] [doi: 10.1055/s-0036-1585085] [Medline: 27398697]
28. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: a randomized clinical trial. Obesity (Silver Spring) 2016 Jan;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
29. Smith K, Lanningham-Foster L, Welch A, Campbell C. Web-based behavioral intervention increases maternal exercise but does not prevent excessive gestational weight gain in previously sedentary women. J Phys Act Health 2016 Jun;13(6):587-593 [FREE Full text] [doi: 10.1123/jpah.2015-0219] [Medline: 26594820]
30. Zairina E, Abramson MJ, McDonald CF, Li J, Dharmasiri T, Stewart K, et al. Telehealth to improve asthma control in pregnancy: a randomized controlled trial. Respirology 2016 Jul;21(5):867-874. [doi: 10.1111/resp.12773] [Medline: 27037722]
31. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
32. Sagedal LR, Øverby NC, Bere E, Torstveit MK, Lohne-Seiler H, Småstuen M, et al. Lifestyle intervention to limit gestational weight gain: the Norwegian Fit for Delivery randomised controlled trial. BJOG 2017 Jan;124(1):97-109. [doi: 10.1111/1471-0528.13862] [Medline: 26768233]
33. Kennedy RAK, Reynolds CME, Cawley S, O'Malley E, McCartney DM, Turner MJ. A web-based dietary intervention in early pregnancy and neonatal outcomes: a randomized controlled trial. J Public Health (Oxf) 2019 Jun 01;41(2):371-378. [doi: 10.1093/pubmed/fdy117] [Medline: 30010835]
34. Kennelly MA, Ainscough K, Lindsay KL, O'Sullivan E, Gibney ER, McCarthy M, et al. Pregnancy exercise and nutrition with smartphone application support: a randomized controlled trial. Obstet Gynecol 2018 May;131(5):818-826. [doi: 10.1097/AOG.0000000000002582] [Medline: 29630009]

XSL•FO
RenderX

35. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
36. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
37. Patel A, Kuhite P, Puranik A, Khan SS, Borkar J, Dhande L. Effectiveness of weekly cell phone counselling calls and daily text messages to improve breastfeeding indicators. BMC Pediatr 2018 Oct 30;18(1):337 [FREE Full text] [doi: 10.1186/s12887-018-1308-3] [Medline: 30376823]
38. Rasekaba TM, Furler J, Young D, Liew D, Gray K, Blackberry I, et al. Using technology to support care in gestational diabetes mellitus: Quantitative outcomes of an exploratory randomised control trial of adjunct telemedicine for gestational diabetes mellitus (TeleGDM). Diabetes Res Clin Pract 2018 Aug;142:276-285. [doi: 10.1016/j.diabres.2018.05.049] [Medline: 29885390]
39. Al-Ofi EA, Mosli HH, Ghamri KA, Ghazali SM. Management of postprandial hyperglycaemia and weight gain in women with gestational diabetes mellitus using a novel telemonitoring system. J Int Med Res 2019 Feb;47(2):754-764 [FREE Full text] [doi: 10.1177/0300060518809872] [Medline: 30442052]
40. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
41. Carolan-Olah M, Sayakhot P. A randomized controlled trial of a web-based education intervention for women with gestational diabetes mellitus. Midwifery 2019 Jan;68:39-47. [doi: 10.1016/j.midw.2018.08.019] [Medline: 30343264]
42. Guo H, Zhang Y, Li P, Zhou P, Chen LM, Li SY. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
43. Sung JH, Lee DY, Min KP, Park CY. Peripartum management of gestational diabetes using a digital health care service: a pilot, randomized controlled study. Clin Ther 2019 Nov;41(11):2426-2434. [doi: 10.1016/j.clinthera.2019.09.005] [Medline: 31587813]
44. Butler Tobah YS, LeBlanc A, Branda ME, Inselman JW, Morris MA, Ridgeway JL, et al. Randomized comparison of a reduced-visit prenatal care model enhanced with remote monitoring. Am J Obstet Gynecol 2019 Dec;221(6):638.e1-638.e8. [doi: 10.1016/j.ajog.2019.06.034] [Medline: 31228414]
45. Ferrara A, Hedderson MM, Brown SD, Ehrlich SF, Tsai AL, Feng J, et al. A telehealth lifestyle intervention to reduce excess gestational weight gain in pregnant women with overweight or obesity (GLOW): a randomised, parallel-group, controlled trial. Lancet Diabetes Endocrinol 2020 Jun;8(6):490-500 [FREE Full text] [doi: 10.1016/S2213-8587(20)30107-8] [Medline: 32445736]
46. Huang RC, Silva D, Beilin L, Neppe C, Mackie KE, Roffey E, et al. Feasibility of conducting an early pregnancy diet and lifestyle e-health intervention: the Pregnancy Lifestyle Activity Nutrition (PLAN) project. J Dev Orig Health Dis 2020 Feb;11(1):58-70. [doi: 10.1017/S2040174419000400] [Medline: 31391133]
47. Tomyabatra K. Antenatal care health education intervened by social networking on mobile phone compared with usual care to improve maternal and neonatal outcomes: randomized controlled trial. J Med Assoc Thai 2020 Jun 15;103(6):529-535 [FREE Full text] [Medline: 242275450]
48. Ding B, Gou B, Guan H, Wang J, Bi Y, Hong Z. WeChat-assisted dietary and exercise intervention for prevention of gestational diabetes mellitus in overweight/obese pregnant women: a two-arm randomized clinical trial. Arch Gynecol Obstet 2021 Sep;304(3):609-618. [doi: 10.1007/s00404-021-05984-1] [Medline: 33570656]
49. LeBlanc ES, Smith NX, Vesco KK, Paul IM, Stevens VJ. Weight loss prior to pregnancy and subsequent gestational weight gain: Prepare, a randomized clinical trial. Am J Obstet Gynecol 2021 Jan;224(1):99.e1-99.e14. [doi: 10.1016/j.ajog.2020.07.027] [Medline: 32687819]
50. Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen MH, et al. Effectiveness of a smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): randomized controlled trial. JMIR Mhealth Uhealth 2021 Mar 11;9(3):e26091 [FREE Full text] [doi: 10.2196/26091] [Medline: 33704075]
51. Su MC, Chao AS, Chang MY, Chang YL, Chen CL, Sun JC. Effectiveness of a nurse-led web-based health management in preventing women with gestational diabetes from developing metabolic syndrome. J Nurs Res 2021 Dec 01;29(6):e176. [doi: 10.1097/jnr.0000000000000456] [Medline: 34570053]
52. Sun Y, Lingying O. Impact of telemedicine system automatic reminder on outcomes in women with gestational glycosuria. Ethiopian Journal of Health Development 2021;35(3) [FREE Full text] [doi: 10.4314/ejhd.v35i3.1]
53. Tian Y, Zhang S, Huang F, Ma L. Comparing the efficacies of telemedicine and standard prenatal care on blood glucose control in women with gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 25;9(5):e22881 [FREE Full text] [doi: 10.2196/22881] [Medline: 33783365]

XSL•FO
RenderX

54. Yew TW, Chi C, Chan SY, van Dam RM, Whitton C, Lim CS, et al. A randomized controlled trial to evaluate the effects of a smartphone application-based lifestyle coaching program on gestational weight gain, glycemic control, and maternal and neonatal outcomes in women with gestational diabetes mellitus: the SMART-GDM study. Diabetes Care 2021 Feb;44(2):456-463 [FREE Full text] [doi: 10.2337/dc20-1216] [Medline: 33184151]
55. Gonzalez-Plaza E, Bellart J, Arranz Á, Luján-Barroso L, Crespo Mirasol E, Seguranyes G. Effectiveness of a step counter smartband and midwife counseling intervention on gestational weight gain and physical activity in pregnant women with obesity (Pas and Pes Study): randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 15;10(2):e28886 [FREE Full text] [doi: 10.2196/28886] [Medline: 35166684]
56. Uria-Minguito A, Silva-José C, Sánchez-Polán M, Díaz-Blanco Á, García-Benasach F, Carrero Martínez V, et al. The effect of online supervised exercise throughout pregnancy on the prevention of gestational diabetes in healthy pregnant women during COVID-19 pandemic: a randomized clinical trial. Int J Environ Res Public Health 2022 Oct 28;19(21):14104 [FREE Full text] [doi: 10.3390/ijerph192114104] [Medline: 36360995]
57. Bekker MN, Koster MPH, Keusters WR, Ganzevoort W, de Haan-Jebbink JM, Deurloo KL, et al. Home telemonitoring versus hospital care in complicated pregnancies in the Netherlands: a randomised, controlled non-inferiority trial (HoTeL). Lancet Digit Health 2023 Mar;5(3):e116-e124 [FREE Full text] [doi: 10.1016/S2589-7500(22)00231-X] [Medline: 36828605]
58. Munda A, Mlinaric Z, Jakin PA, Lunder M, Pongrac Barlovic D. Effectiveness of a comprehensive telemedicine intervention replacing standard care in gestational diabetes: a randomized controlled trial. Acta Diabetol 2023 Aug;60(8):1037-1044 [FREE Full text] [doi: 10.1007/s00592-023-02099-8] [Medline: 37185903]
59. Sharma S, Soni S, Kaushik S, Kalaivani M, Dadhwal V, Sharma KA, et al. SwasthGarbh: a smartphone app for improving the quality of antenatal care and ameliorating maternal-fetal health. IEEE J Biomed Health Inform 2023 Jun;27(6):2729-2738 [FREE Full text] [doi: 10.1109/JBHI.2022.3211426] [Medline: 36191117]
60. Skalecki S, Lawford H, Gardener G, Coory M, Bradford B, Warrilow K, et al. My Baby's Movements: an assessment of the effectiveness of the My Baby's Movements phone program in reducing late-gestation stillbirth rates. Aust N Z J Obstet Gynaecol 2023 Jun;63(3):378-383 [FREE Full text] [doi: 10.1111/ajo.13647] [Medline: 36717966]
61. Wakwoya EB, Belachew T, Girma T. Effect of intensive nutrition education and counseling on hemoglobin level of pregnant women in East Shoa zone, Ethiopia: randomized controlled trial. BMC Pregnancy Childbirth 2023 Sep 19;23(1):676 [FREE Full text] [doi: 10.1186/s12884-023-05992-w] [Medline: 37726668]
62. Téoule J, Woll C, Ray J, Sütterlin M, Filsinger B. The effectiveness of integrated online health-coaching on physical activity and excessive gestational weight gain: a prospective randomized-controlled trial. Arch Gynecol Obstet 2024 Jul;310(1):307-314 [FREE Full text] [doi: 10.1007/s00404-023-07296-y] [Medline: 38217763]
63. Wang L, Zou L, Yi H, Li T, Zhou R, Yang J, et al. The implementation of online and offline hybrid weight management approach for pregnant women based on the Fogg behavior model in Hainan, China: a pilot randomized controlled trial. BMC Pregnancy Childbirth 2024 Jul 30;24(1):516 [FREE Full text] [doi: 10.1186/s12884-024-06699-2] [Medline: 39080659]
64. Islam MM, Poly TN, Walther BA, Jack Li YCJJM. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]
65. Antoun J, Itani H, Alarab N, Elsehmawy AJJM. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 08;10(4):e35479 [FREE Full text] [doi: 10.2196/35479] [Medline: 35394443]
66. Santos S, Voerman E, Amiano P, Barros H, Beilin LJ, Bergström A, et al. Impact of maternal body mass index and gestational weight gain on pregnancy complications: an individual participant data meta-analysis of European, North American and Australian cohorts. BJOG 2019 Jul;126(8):984-995 [FREE Full text] [doi: 10.1111/1471-0528.15661] [Medline: 30786138]
67. Victor A, de França da Silva Teles L, de Carvalho LF, Biagio LD, Argentato PP, Luzia LA, et al. Predictors of inadequate gestational weight gain according to iom recommendations and intergrowth-21st standards: the araraquara cohort study. BMC Pregnancy Childbirth 2024 Sep 03;24(1):579 [FREE Full text] [doi: 10.1186/s12884-024-06749-9] [Medline: 39227805]
68. Victa AGLBD, McCallum C, Menezes GJSED. Atenas Program: pioneer Brazilian service of outpatient care by telemedicine for women with miscarriage or incomplete abortion. Saúde debate 2024 Jun;48(141):e8414 [FREE Full text] [doi: 10.1590/2358-289820241418414i]
69. Beckie TM, Sengupta A, Dey AK, Dutta K, Ji M, Chellappan S. A mobile health behavior change intervention for women with coronary heart disease: a randomized controlled pilot study. J Cardiopulm Rehabil Prev 2024 Jan 01;44(1):40-48 [FREE Full text] [doi: 10.1097/HCR.0000000000000804] [Medline: 37285601]
70. Walker R, Bennett C, Blumfield M, Gwini S, Ma J, Wang F, et al. Attenuating pregnancy weight gain-what works and why: a systematic review and meta-analysis. Nutrients 2018 Jul 22;10(7):944 [FREE Full text] [doi: 10.3390/nu10070944] [Medline: 30037126]
71. Hasegawa J, Nakamura M, Hamada S, Okuyama A, Matsuoka R, Ichizuka K, et al. Gestational weight loss has adverse effects on placental development. J Matern Fetal Neonatal Med 2012 Oct;25(10):1909-1912 [FREE Full text] [doi: 10.3109/14767058.2012.664666] [Medline: 22348351]

XSL•FO
RenderX

72. Zafman KB, Riegel ML, Levine LD, Hamm RF. An interactive childbirth education platform to improve pregnancy-related anxiety: a randomized trial. Am J Obstet Gynecol 2023 Jul;229(1):67.e1-67.e9 [FREE Full text] [doi: 10.1016/j.ajog.2023.04.007] [Medline: 37054807]
73. Chaudhary K, Nepal J, Shrestha K, Karmacharya M, Khadka D, Shrestha A, et al. Effect of a social media-based health education program on postnatal care (PNC) knowledge among pregnant women using smartphones in Dhulikhel hospital: a randomized controlled trial. PLoS One 2023;18(1):e0280622 [FREE Full text] [doi: 10.1371/journal.pone.0280622] [Medline: 36662821]
74. van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020 May 15;22(5):e15773 [FREE Full text] [doi: 10.2196/15773] [Medline: 32412417]
75. Chen HH, Lee CF, Huang JP, Hsiung Y, Chi LK. Effectiveness of a nurse-led mHealth app to prevent excessive gestational weight gain among overweight and obese women: a randomized controlled trial. J Nurs Scholarsh 2023 Jan;55(1):304-318 [FREE Full text] [doi: 10.1111/jnu.12813] [Medline: 36121127]
76. ChatGPT. OpenAI. URL: https://chatgpt.com/ [accessed 2025-02-26]

Abbreviations

GDM: gestational diabetes mellitus
GWG: gestational weight gain
IOM: Institute of Medicine
LGA: large for gestational age
mHealth: mobile health
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: Revised Cochrane risk-of-bias tool for randomized trials
SGA: small for gestational age
SMD: standardized mean difference

Edited by T de Azevedo Cardoso; submitted 18.09.24; peer-reviewed by C Liu, L Peng; comments to author 29.11.24; revised version received 06.01.25; accepted 10.02.25; published 14.03.25.

Please cite as:
Wang J, Tang N, Jin C, Yang J, Zheng X, Jiang Q, Li S, Xiao N, Zhou X
Association of Digital Health Interventions With Maternal and Neonatal Outcomes: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e66580
URL: https://www.jmir.org/2025/1/e66580
doi:10.2196/66580
PMID:

©Jianing Wang, Nu Tang, Congcong Jin, Jianxue Yang, Xiangpeng Zheng, Qiujing Jiang, Shengping Li, Nian Xiao, Xiaojun Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Wellness Programs in the Workplace: Meta-Review

Saeed Amirabdolahian[1*], MS; Guy Pare[2*], PhD; Stefan Tams[1*], PhD

[1]Information Technology Department, HEC Montréal, Montreal, QC, Canada

[2]Research Chair in Digital Health, HEC Montréal, Montreal, QC, Canada

[*]all authors contributed equally

Corresponding Author:
Guy Pare, PhD
Research Chair in Digital Health
HEC Montréal
3000 Cote-Ste-Catherine Road
Montreal, QC, H3T 2A7
Canada
Phone: 1 514 340 6812
Email: guy.pare@hec.ca

Abstract

Background: Corporate wellness programs are increasingly using digital technologies to promote employee health. Digital wellness programs (DWPs) refer to initiatives that deliver health interventions through digital tools. Despite a growing body of evidence on DWPs, the literature remains fragmented across multiple health domains.

Objective: This study aims to provide a comprehensive synthesis of existing research on the efficacy (eg, impact on employee's physical health, mental well-being, behavioral changes, and absenteeism) and acceptability (eg, engagement, perceived usefulness, and adoption) of employer-provided DWPs. Specifically, we aim to map the extent, range, and nature of research on this topic; summarize key findings; identify gaps; and facilitate knowledge dissemination.

Methods: We conducted a meta-review of studies published between 2000 and 2023. We adopted a database-driven search approach, including the MEDLINE, PsycINFO, ProQuest Central, and Web of Science Core Collection databases. The inclusion criteria consisted of (1) review articles; (2) publications in English, French, or German; (3) studies reporting on digital health interventions implemented in organizations; (4) studies reporting on nonclinical or preclinical employee populations; and (5) studies assessing the efficacy and acceptability of employer-provided DWPs. We performed a descriptive numerical summary and thematic analysis of the included studies.

Results: Out of 593 nonduplicate studies screened, 29 met the inclusion criteria. The most investigated health domains included mental health (n=19), physical activity (n=8), weight management (n=6), unhealthy behavior change (n=4), and sleep management (n=2). In total, 24 reviews focused on the efficacy of DWPs, primarily in relation to health-related outcomes (eg, stress and weight), while fewer reviews addressed organization-related outcomes (eg, burnout and absenteeism). Four reviews explored the mechanisms of action, and 3 assessed the acceptability of DWPs using various measures. Overall, the findings support the efficacy and acceptability of DWPs, although significant gaps persist, particularly regarding the durability of outcomes, the role of technology, and the causal mechanisms underlying behavioral change.

Conclusions: While DWPs show promise across a variety of health domains, several aspects of their effectiveness remain underexplored. Practitioners should capitalize on existing evidence of successful DWPs while acknowledging the limitations in the literature.

(J Med Internet Res 2025;27:e70982) doi:10.2196/70982

KEYWORDS

digital wellness programs; corporate wellness; health interventions; efficacy; acceptability; meta-review; mHealth; eHealth

Introduction

Employee health is closely intertwined with stakeholders' financial interests, and technology can play a pivotal role in advancing both. A recent large-scale survey identified a clear link between employee well-being and firm performance [1]. Beyond this alignment, employee wellness has become an organizational expectation [2], with surveys indicating that most

XSL•FO
RenderX

employees believe their organizations are responsible for supporting their health and well-being [3,4]. To promote employee health, many companies invest in corporate wellness programs, with more than half of the organizations offering at least one such initiative [5]. These programs have evolved into a strategic priority, not only as an attractive job benefit but also as a means to reduce absenteeism and lower health care costs [6].

A meta-analysis found that for every one US dollar invested in corporate wellness programs, organizations save US $3.27 in health care costs and US $2.73 through reduced absenteeism [7]. Beyond financial benefits, these programs foster a sense of organizational support, enhancing employee commitment and prosocial behavior in the workplace [8-10]. In addition, the growing emphasis on corporate social responsibility has drawn practitioners' attention to wellness programs as a means to fulfill these expectations [11]. Given the significant amount of time employees spend at work, the workplace provides a unique environment for implementing wellness initiatives. Organizations can effectively scale these programs by identifying key risk factors and integrating targeted wellness initiatives into their routines and culture.

The growing integration of digital technologies into wellness programs has significantly improved their accessibility and scalability. However, despite the expanding body of research on digital wellness programs (DWPs), the literature remains highly fragmented across multiple domains and disciplines. This fragmentation poses challenges for organizational practitioners, making it difficult to adopt evidence-based best practices effectively.

In this study, we define DWPs as initiatives that deliver intervention content to employees through digital tools, either partially or fully. DWPs are not simply digital updates of traditional programs; they exhibit unique features that merit focused research for several reasons. First, participants in DWPs play a more active role than in traditional wellness programs. They can creatively adapt and personalize digital tools, which can influence the programs' outcomes [12]. Second, digital technologies add complexity to the design of wellness programs. Previous studies have highlighted factors such as interaction modularity [13], the design of digital nudges [14], and engagement with technology features [15] as critical determinants of the efficacy of DWPs. Third, digital technologies enable novel ways of interacting with employees, creating innovative wellness solutions previously unattainable. For example, automated digital interventions can provide health feedback without direct human supervision. These considerations have spurred a growing interest in research on DWPs.

Despite the considerable body of research, knowledge on DWPs is still fragmented across multiple disciplines. This fragmentation is reflected in the numerous reviews that examine the same topic yet reach varying conclusions. The effectiveness of DWPs often depends on how their boundaries are defined. For instance, some reviews report highly promising outcomes (eg, [16,17]), while others highlight conflicting evidence, raising doubts about their overall efficacy (eg, [18,19]). This poses a challenge, as review studies play a critical role in guiding practitioners' decision-making [20,21]. In addition, research on DWPs is expanding rapidly across various health domains (eg, mental health, physical activity, and weight management) yet often develops in silos, limiting the generalizability of findings beyond specific areas. This fragmentation is particularly concerning because, regardless of the health domain, DWPs share key similarities in their behavioral mechanisms, technological components, and target populations. To address this gap, this study provides a comprehensive overview of the literature on DWPs, aiming to navigate the scattered research landscape and facilitate cross-pollination of ideas to inform future research.

Given the abundance of studies and reviews on DWPs, we have adopted a meta-review methodology to cover the broad scope of the relevant literature. This approach ensures that we address all compelling evidence and identify research gaps in the field. This meta-review began with the broad research questions (RQs): "What do we know?" and "What remains unanswered?" During the full-text screening process, it became clear that the central focus of the literature revolves around the efficacy and acceptability of DWPs. Thus, the initial questions were refined into the following specific research inquiries:

1. What technologies and health domains have been investigated in prior research on DWPs?
2. What do we know about the efficacy and acceptability of DWPs?

This meta-review answers these questions by categorizing the existing literature into 6 health domains identified inductively. This classification helps to uncover avenues for future research within each domain, identify points of agreement and disagreement, compare findings across domains, and encourage the cross-pollination of ideas.

In short, the primary objective of this study is to synthesize existing research on the efficacy and acceptability of DWPs in workplace settings. Specifically, we aim to (1) assess the effectiveness of DWPs in improving employees' physical and mental well-being, (2) examine the acceptability and engagement levels of these interventions, and (3) identify gaps in the literature to inform future research and practice.

Methods

Research Design

This study uses a meta-review method, which is particularly useful in interdisciplinary research when an extensive body of literature exists, and a broad exploration of the topic is necessary [22]. This approach uses various types of literature reviews as input and adopts a comprehensive search strategy and considers diverse literature to address broad RQs [23,24].

The meta-review method aligns well with our research objectives, considering that an initial exploratory search revealed that synthesizing existing reviews would be valuable. Much of the relevant knowledge is dispersed across various review studies on closely related topics, yet many of these reviews were conducted independently, without referencing findings from

XSL•FO
RenderX

other reviews. This lack of integration limits the potential to consolidate insights across similar reviews.

As protocol registration is not a standard requirement for meta-reviews, this study was not prospectively registered. However, to ensure transparency and systematicity [25,26], we adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework (Multimedia Appendix 1 [27]) and followed the guidelines set forth by Paré et al [25].

Search Strategy

The search process followed 3 key steps, as recommended by Gusenbauer and Haddaway [28]. The first step involved a preliminary lookup search on Google Scholar to identify existing reviews on DWPs. This step allowed us to assess the feasibility of conducting a scoping meta-review and helped pinpoint key pieces of evidence on the topic.

The second step was an exploratory search to determine the approximate scope of the topic. The authors carried out the search with the assistance of a professional librarian and discussed initial findings to develop a comprehensive list of search terms and refine the search queries accordingly.

Finally, the third step involved a systematic search of relevant studies published between January 2000 and December 2023. The search was initially carried out in July 2023 and updated in January 2024. The search timeframe (2000-2023) was chosen to capture the evolution of DWPs in the workplace. Digital components of wellness programs began gaining traction in the early 2000s, and review articles synthesizing research in this area started appearing around this time. This timeframe ensures comprehensive coverage of relevant literature while maintaining a focused and manageable scope. We adopted a database-driven search approach [29], using 4 multidisciplinary research databases: MEDLINE, PsycINFO, ProQuest Central and the Web of Science Core Collection. This multistep process ensured that the search terms and databases used captured the breadth of this multidisciplinary topic. In addition, we conducted a forward snowballing search to complement the systematic search.

Textbox 1 presents the finalized search terms and the logical structure of the search query. Since we aimed to capture all ranges of intervention outcomes and comparisons, the search query did not include them. The search string used in each database can be found in Multimedia Appendix 2.

Textbox 1. Search terms and the logical combinations of conducted queries.

(systematic review OR scoping review OR narrative review OR descriptive review OR meta-analysis OR mapping review OR realist review OR meta-synthesis OR comparative review OR conceptual review OR literature review OR integrative review OR meta-ethnography OR meta-synthesis OR meta-review OR narrative review OR narrative synthesis OR qualitative review)

AND

(workplace OR work place OR worker OR staff OR workforce OR employee OR occupational setting OR organizational setting OR organisational setting)

AND

(m-health OR mhealth OR mobile health OR e-health OR ehealth OR digital health OR digital wellness OR virtual care OR self-tracking OR self-quantification OR quantified-self OR wearable OR iCBT) OR ((physical health OR physical activity OR mental health OR sleep health OR psychological health OR well-being OR wellbeing OR prescriptions OR preventive health OR sleep health OR sleep disorders) AND (app OR information technology OR information system OR digital OR platform OR self-quantification OR quantified self OR smartphone OR internet-based OR mobile phone OR cell phone OR virtual OR teleconsultation))

Screening

A set of inclusion and exclusion criteria was developed to clearly define the scope of this review (Textbox 2). We included all reviews that reported on DWPs aimed at nonclinical or preclinical employee populations. Consistent with our definition of DWPs, interventions were considered digital if their content was delivered to employees fully or partially through digital technologies. The studies had to be available in English, French, or German. We included studies published between January 2000 and the end of 2023, which is a broad timeframe for the subject matter. However, we did not impose a date limit on the primary studies included in the review articles.

When screening literature reviews for inclusion in a meta-review, it is possible that some studies only partially meet the eligibility criteria. For instance, a review might include both digital (eligible) and nondigital (ineligible) interventions. In such cases, we only retained those reviews where the eligible part of the study was analyzed separately.

The identified studies were systematically screened in 2 steps using Covidence, a web-based platform designed for managing and screening citations collaboratively. In the first step, one reviewer screened the abstracts of the identified studies and excluded only those that clearly did not meet the inclusion criteria. Where there was uncertainty, the paper was flagged for further discussion. Flagged papers were then reviewed in working sessions where the authors reached a consensus before making any exclusion decisions. Papers not excluded at this stage were retained for full-text screening. In the second step, 3 independent reviewers conducted the full-text screening. Each reviewer read and assessed the papers independently. In 2 subsequent working sessions, any disagreements were resolved through discussion, ensuring a final consensus on the inclusion of each paper.

XSL•FO
RenderX

Textbox 2. List of inclusion and exclusion criteria.

Inclusion criteria

- Population
  - Any nonclinical or preclinical working population
  - Studies that focused on working and nonworking populations simultaneously were included only if they present their findings regarding the working population separately
- Intervention and exposure
  - Any digital health interventions used in workplace settings. It includes interventions aimed at employees' physical health; mental health; well-being; prescriptions; or preventive health, employee assistance programs, substance or alcohol misuse interventions, sexual health interventions, sleep disorders, etc
  - Studies that have focused on digital and nondigital health interventions were included only if they have presented their findings regarding the working population separately
- Outcomes
  - Efficacy or acceptability outcomes
- Study design
  - Stand-alone review articles in peer-reviewed outlets
  - Review reports from governments, industry leaders, and prominent consulting firms
- Time
  - Published between January 2000 and December 2023

Exclusion criteria

- Population
  - Any population outside of a workplace setting (eg, students)
  - Special occupational groups without typical freedoms (eg, soldiers) or with informal employment (eg, sex workers)
- Intervention and exposure
  - Digital health interventions not supported or funded by workplaces, transition-to-retirement interventions, and interventions aimed at clinical populations (eg, patients with cancer)
  - Interventions mainly based on outdated digital technologies such as CD-ROMs and PDAs were excluded
  - Interventions specifically implemented during the pandemic
  - Long-term management of mental illnesses
- Outcomes
  - No exclusion
- Study design
  - Primary studies
  - Review protocols
  - Conference abstract
  - Theses and dissertations
- Time
  - No exclusion

Data Charting

In this stage, we charted key information from the included studies. We adapted an existing data extraction tool designed for meta-review studies [30]. This tool includes 17 data items: title, author or authors, year of publication, place of publication, research objectives, review type, information sources searched,

XSL•FO
RenderX

inclusion years range, inclusion and exclusion criteria, search terms used, corpus size, aim of interventions, reported work and health outcomes, geographic dispersion of the included studies, organizational settings, intervention characteristics, intervention durations, and key findings. Microsoft Excel was used to chart the data for analysis. As recommended by review methodology guidelines, this process was highly iterative [31]. One author initially extracted the data items for all included studies. Subsequently, all authors participated in 4 working sessions to discuss and validate the data extraction table. The analysis of the extracted data is presented in the following section, and the full dataset is available upon request from the first author.

Data Analysis

The data analysis comprised a descriptive numerical summary and a thematic analysis [31]. The descriptive numerical summary involved a straightforward process of characterizing the included studies. However, the thematic analysis used a highly iterative process. Specifically, we applied conventional qualitative content analysis, which is appropriate when no prior theoretical framework exists and researchers aim to avoid using preconceived categories [32]. Although there are several behavior change theories for various health interventions, no overarching framework or theory is available for organizing the diverse interventions within DWPs across different health domains.

Following the approach outlined by Hsieh and Shannon [32], themes were allowed to emerge naturally from the data through an iterative process. This involved carefully reading the data items, deriving codes, making notes on the emerging codes, revisiting the texts to contextualize the codes, and refining the coding scheme to develop themes. This process unfolded over the course of 6 working sessions, during which the emerging themes were discussed and refined. The result of this step was the categorization of health domains, RQs, and digital technologies and the separation of evidence on efficacy outcomes from mechanisms of action. The Results and Discussion sections are structured according to these categories.

Results

Summary of the Search and Screening Process

The PRISMA flowchart shown in Figure 1 outlines the search and screening process. Initially, the search yielded 967 records, from which 374 duplicates were removed. This left 593 studies, which were screened by abstract. After this step, 42 studies were identified as eligible for full-text screening. The authors ultimately agreed to include 25 studies in the corpus. To ensure thoroughness, a snowballing search was conducted, retrieving a list of 813 papers from Google Scholar using web-scraping tools. One author screened the titles and abstracts for potentially relevant papers which were manually retrieved for full-text review. This process added 7 papers for full-text screening, of which 4 were included in the final corpus. As a result, the search and screening process concluded with a total of 29 studies included in the database.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

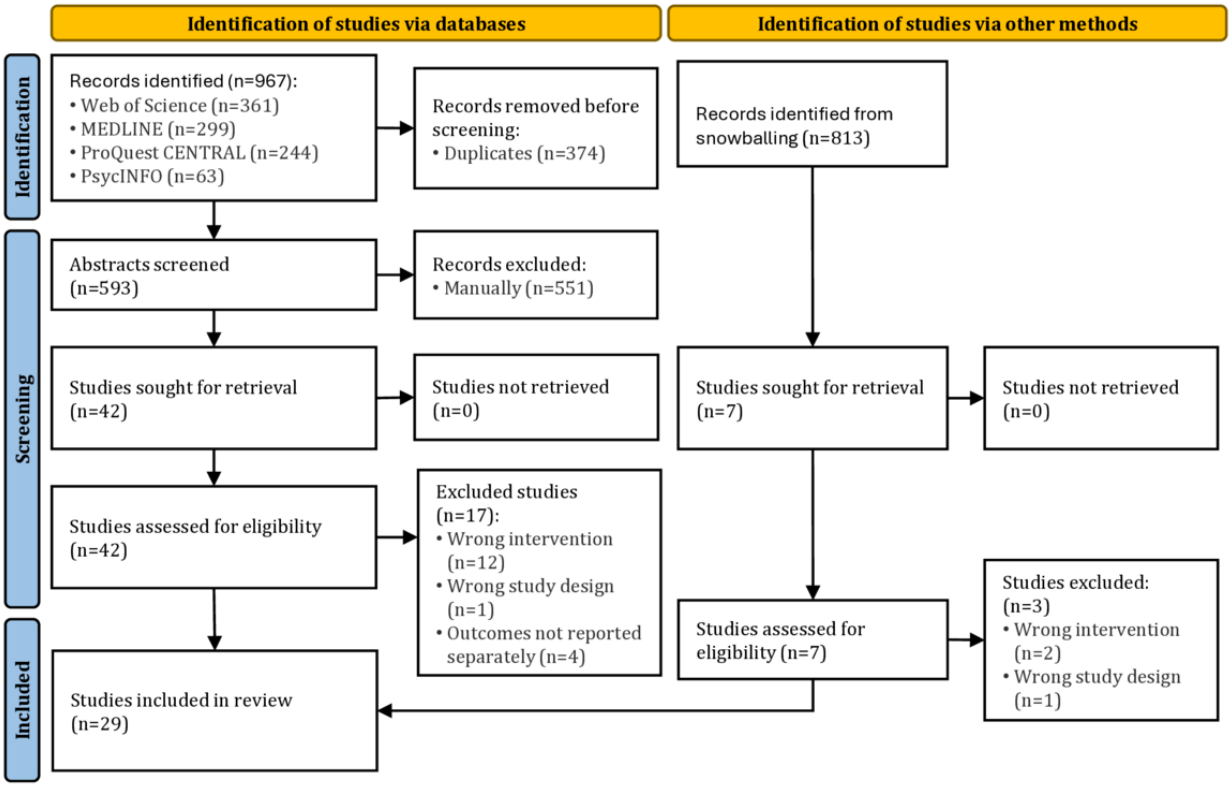

XSL•FO
RenderX

Characteristics of the Included Studies

This section outlines the key characteristics of the included studies.

Date of Publication

Although our inclusion criteria limited the selection to studies published since 2000, the earliest identified review article appeared in 2014, confirming the appropriateness of our chosen publication period. Notably, there was a significant surge in publications in 2023, with 9 out of the 11 studies from that year focusing specifically on DWPs aimed at improving employees' mental health. Unsurprisingly, all these studies cited the recent COVID-19 pandemic as a key motivation.

Health Domains

As shown in Table 1, DWP interventions concern 6 health domains. Mental health (n=19) emerged as the most frequently investigated domain, encompassing interventions aimed at addressing employees' stress, anxiety, burnout, and overall quality of life. The next most explored domains are physical activities (n=8), weight management (n=6), unhealthy behavior change (n=4), and sleep management (n=2). Four studies were categorized as "other." These include programs aimed at screening and reducing varied health risk factors that relate to diabetes and other chronic diseases.

XSL•FO
RenderX

Table 1. Profile of included studies.

Study	Search range	Review type	Corpus size	Type of research question	Health domains[a-h]					
					Mental health[a]	Physical activity[b]	Weight management[c]	Unhealthy behavior change[d]	Sleep management[e]	Others[f]
Aneni et al [18]	Up to and including 2012	Systematic review (narrative analysis)	29	Efficacy		✓	✓	✓		
Armaou et al [33]	1990-2019	Systematic review (narrative analysis)	51	Efficacy	✓					
Bégin et al [34]	Up to and including 2021	Scoping review	56	Efficacy	✓					
Buckingham et al [35]	2007-2017	Systematic review (narrative analysis)	25	Efficacy, mechanisms of action, acceptability		✓				
Bullock et al [19]	2009-2021	Scoping review	2	Efficacy					✓	
Carolan et al [36]	2000-2016	Meta-analysis	21	Efficacy, acceptability	✓					
Damen et al [37]	2009-2019	Scoping review	45	Mechanisms of action		✓				
Freak-Poli et al [38]	Up to and including 2016	Meta-analysis	14	Efficacy		✓				
Jones et al [39]	2014-2019	Systematic review (narrative analysis)	12	Efficacy	✓					
Jung and Cho [40]	Up to and including 2020	Meta-analysis	8	Efficacy		✓	✓			
Kuster et al [41]	Up to and including 2017	Systematic review (narrative analysis)	2	Efficacy						
Lee et al [42]	Up to and including 2021	Systematic review (narrative analysis)	11	Efficacy			✓			
López-Del-Hoyo et al [43]	Up to and including 2022	Systematic review (narrative analysis)	27	Efficacy	✓					
Moe-Byrne et al [44]	2000-2022	Systematic review (narrative analysis)	7	Efficacy	✓					
Paganin and Simbula [45]	2009-2019	Systematic review (narrative analysis)	31	Efficacy	✓		✓			✓
Park et al [46]	Up to and including 2021	Systematic review (narrative analysis)	7	Efficacy	✓					
Phillips et al [16]	Up to and including 2018	Meta-analysis	50	Efficacy	✓			✓		
Scheutzow et al [47]	2005-2019	Systematic review (narrative analysis)	28	Acceptability	✓					
Sevic et al [48]	1990-2021	Systematic review (narrative analysis)	17	Efficacy, mechanisms of action		✓		✓		✓
Stratton et al [49]	1975-2016	Meta-analysis	23	Efficacy	✓					
Stratton et al [50]	1991-2019	Meta-analysis	28	Efficacy	✓					
Stratton et al [17]	2004-2020	Meta-analysis	75	Efficacy	✓					
Sundstrom et al [51]	2000-2021	Scoping review	20	Efficacy				✓		

XSL•FO
RenderX

Study	Search range	Review type	Corpus size	Type of research question	Health domains[a-h]					
					Mental health[a]	Physical activity[b]	Weight management[c]	Unhealthy behavior change[d]	Sleep management[e]	Others[f]
Thai et al [52][g]	2010-2021	Scoping review	24	Efficacy	✓	✓	✓		✓	✓
Tung et al [53]	2006-2020	Systematic review (narrative analysis)	14	Mechanisms of action		✓	✓			✓
Velana and Rinkenauer [54]	2000-2020	Systematic review (narrative analysis)	6[h]	Efficacy	✓					
Vertola et al [55]	2012-2022	Systematic review (narrative analysis)	11	Efficacy	✓					
Webster et al [56]	2010-2018	Scoping review	11	Efficacy	✓					
Xiong et al [57]	Up to and including 2021	Meta-analysis	19	Efficacy	✓					

[a]n=19.

[b]n=8.

[c]n=6.

[d]n=4.

[e]n=2.

[f]n=4.

[g]The inclusion criteria are limited to digital wellness programs in low- and middle-income countries.

[h]The corpus size does not include studies on non-IT interventions.

Type of RQs

In line with the objective of this meta-review, we categorized the literature reviews in our corpus into 2 groups based on the RQs they addressed. As shown in Textbox 2, the vast majority of reviews (27/29, 93%) focused on questions related to the efficacy of DWPs. In contrast, only 3 reviews addressed questions concerning the acceptability of these programs, examining criteria such as uptake, continued use, and compliance with the interventions. While efficacy and acceptability were predefined codes, a third category of RQs emerged inductively during the data charting process. Specifically, we identified 4 reviews exploring “how” DWPs influence intervention outcomes. These RQs were coded under interventions' mechanisms of action. Studies in this category primarily examined the behavior change techniques used in interventions and the theoretical explanations of how these interventions drive outcomes.

Types of Reviews

We categorized the identified studies according to review types as specified by their authors. In those cases where review types were not explicitly mentioned in a study (eg, [39]), we coded the paper according to the guidelines for review typologies [58].

As shown in Table 1, we identified 3 types of literature reviews in our corpus. The most frequent type was the systematic review, comprising 15 studies. These reviews focus on narrow RQs related to the efficacy of DWPs and typically use narrative synthesis methods rather than statistical techniques to analyze their findings. Similar to meta-analyses, systematic reviews limit their inclusion criteria to experimental and quasi-experimental studies. The second frequent group consisted of 8 meta-analyses, which apply statistical methods to address focused RQs about the efficacy of DWPs. These reviews include experimental and quasi-experimental studies while excluding other empirical research designs. Finally, the least frequent type was the scoping review, with 6 studies. Scoping reviews investigate broader RQs, such as exploring outcomes, theories, and behavior change techniques related to specific groups of DWPs. These reviews aim to provide a comprehensive overview of the existing literature by summarizing a wide range of available evidence.

Digital Technologies

The reviews discussing DWPs include various technological components ranging from messaging apps to website portals and telecommunication solutions. We categorized these technologies into five groups: (1) websites and online telecommunication, (2) apps and computer software, (3) email, (4) text messaging services—SMS and messaging apps—and (5) wearable devices. The types of technologies used in each health domain are presented in the mapping tables corresponding to each domain.

Main Findings

In this section, we present the findings organized according to the 6 health domains discussed earlier.

XSL•FO
RenderX

Mental Health

Overview

As mentioned earlier, mental health is the most frequently discussed domain in the literature on DWPs. The World Health Organization defines mental health as a state of mental well-being that enables people to cope with the stresses of life, realize their abilities, learn well and work well, and contribute to their community [59]. Anxiety, depression, and stress-related conditions are the most common mental health issues reported in the workforce. Many organizations are interested in mental health as it is the leading cause of disabilities, sick leave requests, and long-term work incapacities [60]. In addition, research shows that improving organizational mental wellness programs can improve productivity and presenteeism at work [61,62]. The identified reviews on mental health DWPs discuss the efficacy and the acceptability of interventions. However, no review was found that examines the mechanisms of action used in this health domain.

Table 2 maps the reviews on mental health DWPs according to the type of RQs asked, the types of digital technologies included, and 12 categories of outcomes reported in the literature. Most reviews in this domain (16/19, 84%) investigate the efficacy of DWPs and include a wide range of digital technologies and intervention outcomes. Except for 4 reviews that are limited to health care occupations [43,46,54,56], all the included reviews have no limitation on the inclusion of working populations.

XSL•FO
RenderX

Table 2. Mapping of reviews on mental health digital wellness programs.

Study	Type of RQ[a]	Digital technologies					Outcomes										
		Wb[b]	Ap[c]	Email	Ms[d]	Wearables	Stress	Anxiety	Depression	Sc[e]	Pw[f]	Insomnia	Resilience	En[g]	Absenteeism	Presenteeism	Productivity	Burnout
Stratton et al [49]	Efficacy	✓	✓				✓	✓	✓	✓								
Stratton et al [50]	Efficacy	✓	✓											✓	✓	✓	✓	
Stratton et al [17]	Efficacy	✓	✓				✓	✓	✓									
Armaou et al [33]	Efficacy	✓	✓			✓	✓	✓	✓	✓								
Carolan et al [36]	Efficacy and acceptability	✓	✓	✓	✓						✓						✓	
Xiong et al [57]	Efficacy	✓	✓						✓									
Phillips et al [16]	Efficacy	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓						✓
Kuster et al [41]	Efficacy	✓	✓				✓											
Begin et al [34]	Efficacy	✓	✓			✓	✓	✓	✓	✓			✓					
Jones et al [39]	Efficacy		✓				✓	✓			✓							
Vertola et al [55]	Efficacy		✓				✓				✓							✓
Paganin and Simbula [45]	Efficacy		✓				✓				✓	✓						
Thai et al [52]	Efficacy	✓		✓			✓	✓	✓									
López-Del-Hoyo et al [43]	Efficacy	✓	✓				✓	✓	✓	✓			✓					✓
Moe-Byrne et al [44]	Efficacy	✓	✓	✓			✓	✓	✓			✓	✓			✓		

XSL•FO
RenderX

Study	Type of RQ[a]	Digital technologies					Outcomes											
		Wb[b]	Ap[c]	Email	Ms[d]	Wearables	Stress	Anxiety	Depression	Sc[e]	Pw[f]	Insomnia	Resilience	En[g]	Absenteeism	Presenteeism	Productivity	Burnout
Park et al [46]	Efficacy	✓	✓				✓	✓	✓				✓					✓
Velana and Rinkenauer [54]	Efficacy	✓	✓				✓	✓	✓									
Webster et al [56]	Efficacy	✓	✓		✓		✓	✓	✓									✓
Scheutzow et al [47]	Acceptability	✓	✓															

[a]RQ: research question.

[b]Wb: websites and online telecommunication.

[c]Ap: apps and software.

[d]Ms: messaging services.

[e]Sc: self-compassion and mindfulness.

[f]Pw: psychological well-being.

[g]En: employee engagement.

Efficacy

In the domain of mental health, 17 literature reviews assessed the efficacy of DWPs. Table 3 summarizes the findings from these reviews, grouping the outcomes into 12 different categories. The wide range of efficacy outcomes in this health domain reflects the diversity of programs such as those that aim at reducing stress, anxiety, and depression as well as those that aim at enhancing overall mental wellness among employees. The following paragraphs summarize the findings according to the health outcomes of DWPs.

In some cases, the findings were pooled together, making it difficult to categorize them according to specific mental health outcomes. For instance, a meta-analysis [49] combined outcomes related to stress, anxiety, depression, and mindfulness under the broader term “mental health conditions.” This meta-analysis found a small effect size (g=0.24, 95% CI 0.13-0.35, k=23, N=1328, I^2=67%). Subsequent studies have evaluated these results as promising but noted that further clarification is needed regarding specific outcomes [16,17]. In addition, 4 literature reviews [33,34,45,56] did not synthesize or aggregate their findings around specific outcome measures. These reviews primarily conducted scoping surveys of the variability in outcomes and intervention characteristics. The following paragraphs summarize the results of both meta-analyses and non–meta-analyses according to each outcome category.

Stress is the most frequently reported outcome, with 10 reviews addressing its measurement. Two meta-analyses [16,17] evaluated the effectiveness of DWPs on stress. Phillips et al [16] reported a medium-sized effect of DWPs on stress reduction (g=0.54, 95% CI 0.35-0.72, k=22, N=not available, I^2=84%). However, Stratton et al [17], using a much larger corpus, reported a smaller effect size for DWPs on stress (g=0.25, 95% CI 0.17-0.34, k=57, N=10,160, I^2=76%). Three non–meta-analyses [52,54,55] found unanimous evidence supporting the significant effect of DWPs in reducing stress. However, 5 non–meta-analyses [39,41,43,44,46] reported conflicting findings.

Anxiety is the second most frequently discussed outcome, addressed in 9 reviews. Two meta-analyses [16,17] found small but significant effects of DWPs on anxiety reduction. Stratton et al [17] reported an effect size of g=0.26 (95% CI 0.13-0.39, k=29, N=4961, I^2=77%), while Phillips et al [16] found a slightly larger effect size of g=0.34 (95% CI 0.18-0.50, k=15, N=not available, I^2=71%). Among non–meta-analyses, 3 reviews [39,52,54] provided unanimous evidence supporting the efficacy of DWPs in reducing anxiety. However, López-Del-Hoyo et al [43] found conflicting evidence, and Park et al [46] reported no significant effect of DWPs on anxiety in the literature.

XSL•FO
RenderX

Table 3. Summary of mental health digital wellness program outcomes.

Study	Outcomes											
	Stress	Anxiety	Depression	Sc[a]	Pw[b]	Insomnia	Resilience	En[c]	Absenteeism	Presenteeism	Productivity	Burnout
Stratton et al [49][d]	s+[e]	s+	s+	s+	—[f]	—	—	—	—	—	—	—
Stratton et al [50]	—	—	—	—	—	—	—	s+	ns[g]	ns	s+	—
Stratton et al [17]	s+	s+	s+	—	—	—	—	—	—	—	—	—
Carolan et al [36]	—	—	—	—	s+	—	—	—	—	—	s+	—
Xiong et al [57]	—	—	s+	—	—	—	—	—	—	—	—	—
Phillips et al [16]	m+[h]	s+	s+	s+	s+	m+	—	—	—	—	—	m+
Kuster et al [41]	~[i] (1/2)[j]	—	—	—	—	—	—	—	—	—	—	—
Vertola et al [55]	+[k] (9/9)	—	—	—	~ (5/6)	—	—	—	—	—	—	+ (3/3)
Thai et al [52]	+ (3/3)	+ (2/2)	+ (2/2)	—	—	—	—	—	—	—	—	—
López-Del-Hoyo et al [43]	~ (13/22)	~ (5/7)	~ (3/7)	+ (4/4)	—	—	~ (5/6)	—	—	—	—	~ (5/7)
Park et al [46]	~ (3/5)	-[l] (0/1)	+ (1/1)	—	—	—	+ (1/1)	—	—	—	—	+ (1/1)
Velana et al [54]	+ (5/5)	+ (1/1)	+ (2/2)	+ (1/1)	—	—	—	—	—	—	—	—
Moe-Byrne et al [44][m]	~ (3/4)	~ (2/6)	~ (2/6)	—	—	~ (1/2)	~ (2/3)	—	—	~ (3/4)	—	—
Jones et al [39]	~ (unclear)	+ (2/2)	—	—	~ (2/3)	—	—	—	—	—	—	—

[a]Sc: self-compassion and mindfulness.

[b]Pw: psychological well-being.

[c]En: employee engagement.

[d]In this study, the results for stress, anxiety, depression, and self-compassion are pooled together.

[e]In meta-analysis studies, s+ stands for small effect size (g<0.3).

[f]Not applicable.

[g]In meta-analysis studies, ns stands for nonsignificant results.

[h]In meta-analysis studies, m+ stands for medium effect size (0.3<g<0.5).

[i]In non–meta-analysis reviews, ~ stands for found conflicting evidence.

[j](#/#) = (number of studies with significant evidence for this outcome) / (total number of studies reporting this outcome).

[k]In non–meta-analysis reviews, + stands for found unanimous evidence in favor of the effectiveness.

[l]In non-meta-analysis reviews, - stands for found no significant evidence.

[m]In this study, the results for anxiety and depression are reported together.

Depression was discussed in 7 reviews. Three meta-analyses [16,17,57] reported small but significant effects of DWPs on depression. Stratton et al [17] found an effect size of g=0.26 (95% CI 0.19-0.34, k=46, N=10,155, I^2=67%); Phillips et al [16] reported an effect size of g=0.30 (95% CI 0.18-0.42, k=17, N=not available, I^2=61%); and Xiong et al [57], specifically examining internet-based cognitive behavioral therapy programs, found an effect size of g=0.31 (95% CI 0.17-0.44, k=19, N=5898, I^2=83%). In addition, 3 non–meta-analyses [46,52,54] reported significant findings regarding the impact of DWPs on depression. However, López-Del-Hoyo et al [43] found conflicting evidence on the effects of DWPs in this area.

Mindfulness and self-compassion outcomes were discussed in 1 meta-analysis by Phillips et al [16], which found a small effect

XSL•FO
RenderX

size (g=0.42, 95% CI 0.24-0.60, k=5, N=not available, I^2=0). In addition, 2 non–meta-analyses [43,54] reported significant evidence supporting the efficacy of DWPs in these outcomes.

Various secondary outcomes were also identified in the literature, including perceived well-being, psychological resilience, and sleep-related symptoms such as insomnia. For perceived well-being, 2 meta-analyses [16,36] reported small effect sizes. Carolan et al [36] found an effect size of g=0.37 (95% CI 0.23-0.50, k=21, N=2438, I^2=81%), and Phillips et al [16] found a similar effect size of g=0.35 (95% CI 0.25-0.46, k=7, N=not available, I^2=0). However, 2 non–meta-analyses [39,55] found conflicting evidence. For psychological resilience, 1 non–meta-analysis [44] found conflicting evidence in the literature. Regarding insomnia, 1 meta-analysis [16] reported a medium effect size (g=0.52, 95% CI 0.39-0.65, k=6, N=not available, I^2=0%) after removing an outlier. In addition, 1 non–meta-analysis [44] found conflicting evidence on insomnia outcomes.

Interestingly, several reviews reported on job-related outcomes. One meta-analysis [50] focused exclusively on workplace-related outcomes, finding a small effect on employee engagement (g=0.19, 95% CI 0.10-0.28, k=10, N=2497, I^2=16%) but no significant effect on absenteeism (g=0.28, 95% CI −0.02 to 0.57, k=12, N=3325, I^2=94%) or presenteeism (g=0.18, 95% CI −0.05 to 0.41, k=8, N=2102, I^2=85%). Regarding presenteeism, 1 non–meta-analysis [44] found conflicting evidence related to tailored DWPs. For employee productivity, 2 meta-analyses found small effect sizes: Stratton et al [50] reported g=0.16 (95% CI 0.03-0.29, k=16, N=3053, I^2=62%), and Carolan et al [36] found g=0.25 (95% CI 0.09-0.41, k=13, N=1295, I^2=75%). In terms of job burnout, Phillips et al [16] found a medium effect size (g=0.51, 95% CI 0.26-0.75, k=8, N=not available, I^2=79%). Two non–meta-analyses [46,55] reported unanimous evidence supporting the efficacy of DWPs in reducing burnout, while López-Del-Hoyo et al [43] found conflicting evidence.

Acceptability

Two non–meta-analyses [36,47] investigated the acceptability of DWPs related to mental health (Table 4). According to these reviews, uptake rates—the proportion of targeted employees who start the program—are an underreported measure of acceptability in the existing literature. The average uptake rate is 11%, indicating that slightly more than 1 in 10 targeted employees initiate the program. Scheutzow et al [47] also identified several factors contributing to lower uptake rates, including employees' fear of workplace stigmatization and their preference to keep health matters separate from workplace experiences. These barriers emphasize the need for organizations to address concerns about privacy and stigma to improve participation in DWPs.

Table 4. Acceptability of mental health digital wellness programs.

Study	Acceptability measures							
	Uptake	Adherence rate	Attrition rate	Dropout rate	Compliance rate	Satisfaction score	Int[a]	PU[b]
Scheutzow et al [47]	11% (n=2)	54% (n=2)	32% (n=3)	51% (range 15%-68%; n=7)	68% (n=2)	82% (n=10)	55% positive (n=9)	85%
Carolan et al [36]	—[c]	—	23% (range 3%-54%; n=20)	—	45% (range 3%-95%; n=19)	—	—	—

[a]Int: interest or willingness to use.

[b]PU: perceived usability or usefulness.

[c]Not applicable.

Regarding adherence rates, 2 empirical studies reported an average rate of 54%, as indicated by the study by Scheutzow et al [47]. However, the definition of adherence varies across these studies. One study defines adherence as continued participation after the first day of the intervention, while the other measures it as completing a specified number of activities within a given time frame, akin to a compliance rate. As a result, these findings should be interpreted with caution.

Attrition, dropout, and completion rates—which are similar measures—vary significantly across empirical studies. According to the literature reviews, the average rates for mental health DWPs are 32% for attrition, 51% for dropout, and 68% for completion according to the results in Table 4. The most frequently cited reasons for employees dropping out of these programs include concerns about privacy, lack of time, technical difficulties, lack of motivation, confidence in their ability to manage stress independently, dissatisfaction with the interventions, and high stress levels at the start of the program. In addition, the data suggest that younger employees are more likely to drop out of these interventions [47].

Higher completion rates are observed in DWPs delivered over shorter periods, those employing self-monitoring or tailored interventions, and those using secondary modalities such as email and SMS text messaging to enhance user engagement [36]. Finally, available evidence indicates that employees generally express high levels of satisfaction with DWPs, rating them as useful and showing a strong willingness to continue using them. However, Scheutzow et al [47] note that these positive satisfaction ratings often contrast with the unfavorable attrition and completion rates observed in many primary studies.

Physical Activity

Overview

The second most frequently discussed health domain is physical activity. According to the World Health Organization, physical activity refers to any movement of the human body, whether it

XSL•FO
RenderX

involves performing daily tasks or altering one's physical position [63]. Insufficient physical activity has long been associated with a range of health complications, including cancer, anxiety, depression, cardiovascular disease, musculoskeletal disorders, and premature death [64,65]. Organizations implement interventions in both sedentary workplaces and physically demanding environments to reduce health risk factors [66]. While sedentary behavior and physical activity interventions are sometimes discussed separately in the literature, we examine them together because both aim to increase active behavior. Table 5 summarizes the literature reviews within this domain. All the reviews in this health domain assess the efficacy of the programs. While most (7/9, 77%) examine the changes in intervention outcomes, 4 out of 9 (44%) examine the mechanisms of action used in these interventions. In addition, 1 review investigates the acceptability of these programs.

Table 5. Mapping of physical activity digital wellness programs.

Study	Type of RQ[a]	Digital technologies					Outcomes				
		Wb[b]	Ap[c]	Email	Ms[d]	Wearables	Am[e]	Sb[f]	An[g]	Presenteeism	Pw[h]
Jung and Cho [40]	Efficacy		✓		✓	✓	✓				
Aneni et al [18]	Efficacy	✓		✓			✓				
Freak-Poli et al [38]	Efficacy					✓	✓	✓	✓		✓
Thai et al [52]	Efficacy	✓	✓	✓		✓	✓				
Paganin and Simbula [45]	Efficacy		✓				✓	✓			
Sevic et al [48]	Efficacy and acceptability	✓	✓	✓		✓	✓	✓	✓	✓	
Buckingham et al [35]	Efficacy, mechanisms of action, and acceptability		✓			✓	✓	✓	✓	✓	✓
Tung et al [53]	Mechanisms of action	✓	✓			✓					
Damen et al [37]	Mechanisms of action	✓	✓	✓		✓					

[a]RQ: research question.

[b]Wb: websites and online telecommunication.

[c]Ap: apps and software.

[d]Ms: messaging services.

[e]Am: amount of physical activity (steps, duration, metabolic equivalent of task, etc).

[f]Sb: sedentary behavior (time, frequency of transitions, etc).

[g]An: anthropometric (weight, waist circumference, etc) and blood test indicators.

[h]Pw: psychological well-being.

Efficacy

The interventions in this health domain are less varied than those in the mental health domain, and the evidence on the efficacy of these interventions can be summarized into primary and secondary outcomes. As shown in Table 6, the primary outcomes of interest are measures such as the number of steps taken and the duration of physical activity, which we refer to as the amount of physical activity. A meta-analysis conducted by Jung and Cho [40] found that DWPs have a small but significant effect on the amount of physical activity (g=0.22, 95% CI 0.03-0.41, k=8, N=not available, I^2=78%). However, 5 non–meta-analyses [18,35,38,48,52] reported conflicting evidence, with roughly half of the empirical studies finding significant changes in the amount of physical activity.

The secondary outcomes include measures of sedentary behavior (eg, sitting time and frequency of transitions to standing), anthropometric measures (eg, weight and waist circumference), physical readiness (eg, maximal oxygen uptake and aerobic capacity tests), and psychological well-being. These outcomes have been less-frequently discussed in prior studies. However, some non–meta-analyses provide evidence that physical activity DWPs can be effective in improving these outcomes.

Regarding sedentary behavior, 3 non–meta-analyses [35,38,48] reported conflicting evidence, with approximately half of the empirical studies showing significant improvements. For anthropometric measures, 2 non–meta-analyses [35,38] reported conflicting evidence, whereas Sevic et al [48] found only nonsignificant results. In terms of physical readiness, Buckingham et al [35] identified 2 interventions with conflicting results, whereas Sevic et al [48] found only 1 study with nonsignificant results. Regarding psychological well-being, Freak-Poli et al [38] reported conflicting evidence, whereas Buckingham et al [35] found significant positive effects.

XSL•FO
RenderX

Table 6. Summary of physical activity digital wellness program outcomes.

Study[a]	Outcomes				
	Am[b]	Sb[c]	An[d]	PR[e]	PW[f]
Jung and Cho [40]	s+[g]	—[h]	—	—	—
Aneni et al [18]	~[i] (3/11)[j]	—	—	—	—
Freak-Poli et al [38]	~ (4/10)	~ (1/2)	~ (4/14)	—	~ (1/2)
Thai et al [52]	~ (3/4)	—	—	—	—
Sevic et al [48]	~ (3/7)	~ (2/3)	-[k] (0/3)	- (0/1)	—
Buckingham et al [35]	~ (14/25)	~ (4/10)	~ (3/5)	~ (1/2)	+[l] (2/2)

[a]We excluded the study by Paganin and Simbula [45] because results were not presented in line with our categories.

[b]Am: amount of physical activity (steps, duration, metabolic equivalent of task, etc).

[c]Sb: sedentary behavior (time, frequency of transitions, etc).

[d]An: anthropometric measures (weight, waist circumference, etc).

[e]PR: physical readiness (maximal oxygen uptake and aerobic readiness).

[f]PW: psychological well-being.

[g]In meta-analysis studies, s+ denotes small effect size (g<0.3).

[h]Not available.

[i]In non–meta-analysis studies, ~ denotes conflicting evidence.

[j](#/#) = (number of studies with significant evidence for this outcome) / (total number of studies reporting this outcome).

[k]In non-meta-analysis reviews, - stands for found no significant evidence.

[l]In non–meta-analysis reviews, + stands for found unanimous evidence in favor of the effectiveness.

Mechanisms of Action

Four non–meta-analyses [35,37,48,53] examined mechanisms of action. These studies mainly conducted a scoping survey of the literature on DWPs with two objectives: (1) categorizing the intervention into existing taxonomies and (2) summarizing the theories used in the development of DWPs. These non–meta-analysis reviews are summarized in subsequent paragraphs.

Three reviews [37,48,53] categorized physical activity DWPs into the established taxonomies of behavior change techniques and intervention strategies [67,68]. According to these taxonomies, behavior change techniques are specific activities used in interventions to induce a change in one's behavior, whereas intervention strategies (also known as intervention functions) are high-level approaches to design interventions. Sevic et al [48] and Tung et al [53] examined the scope of behavior change techniques and intervention strategies used in DWPs. According to these reviews, enablement and education are the most frequent intervention strategies. Enablement refers to those intervention strategies that enhance capabilities and opportunities to change behavior. It can be implemented through various behavior change techniques such as self-monitoring, feedback provision, and goal setting and action planning. For its part, education refers to the techniques that increase one's knowledge of health-related topics. It can be implemented through various behavior change techniques such as providing instructions for healthy behavior and informing of the consequences of behavior. These reviews state that they could not clarify which behavioral techniques and strategies are more successful than others due to the heterogeneous characteristics of interventions compared with the limited number of empirical studies in the literature.

The fourth non–meta-analysis review [35] presents the theories used in the development of physical activity DWPs. In this regard, the most frequently cited theories are socioecologic theory, self-determination theory, social cognitive theory, and the theory of reasoned action. Similarly, this review does not discuss which theories are more successful in designing DWPs. Three reviews in our sample [37,45,53] agree that most empirical studies lack a theoretical basis for developing DWPs in this domain.

Acceptability

The acceptability of physical health DWPs is discussed in one literature review, which focuses on mHealth technologies [35]. Three main acceptability outcomes are reported in the surveyed empirical studies. The most commonly reported outcome is the attrition rate of interventions, which refers to the number of participants who fail to provide data at the final follow-up. The average attrition rate is 18%, which is considered acceptable for health promotion interventions. However, some interventions report attrition rates as high as 74%, indicating significant variability in the acceptability of the programs. This variability is attributed to several factors including employee demographics (with higher attrition rates among females and younger employees) and the duration of the interventions (with higher attrition in interventions lasting more than 12 weeks). Buckingham et al [35] also cite possible reasons for the lack of engagement after initial use, including privacy concerns, diminished interest, and lack of trust in the accuracy of the technologies. Despite the wide range of attrition rates,

XSL•FO
RenderX

Buckingham et al [35] report that employees who continued using the programs until the follow-up generally expressed high levels of satisfaction. However, due to the high heterogeneity of interventions, the review was unable to provide a clear explanation for the variability in attrition rates or the reasons for the decline in the use of technological components.

Weight Management

Overview

Weight management interventions aim to help individuals achieve and maintain a healthy body weight, which can reduce the risk of chronic diseases and improve overall health and quality of life. Compared with employees with normal weight, employees with overweight and obesity incur higher direct costs (eg, insurance claims) and indirect costs (eg, absenteeism) for their companies. It is estimated that organizational costs associated with employees with overweight and obesity are 11% and 45% higher, respectively, than those for employees with normal weight [69]. Weight management interventions typically involve a combination of strategies such as modifying dietary intake and increasing physical activity [70]. We identified 5 literature reviews on weight management DWPs (Table 7). These DWPs adopt either a universal approach targeting all employees or a targeted approach focusing specifically on employees with obesity. Only 1 review [42] focuses solely on the targeted approach.

Table 7. Mapping of reviews on weight management digital wellness programs.

Study	Type of RQ[a]	Digital technologies					Outcomes			
		Wb[b]	Ap[c]	Email	Ms[d]	Wearables	An[e]	Am[f]	D[g]	PR[h]
Jung and Cho [40]	Efficacy				✓		✓			
Aneni et al [18]	Efficacy	✓		✓	✓		✓	✓	✓	✓
Lee et al [42]	Efficacy	✓	✓	✓	✓	✓	✓			
Thai et al [52]	Efficacy	✓		✓	✓	✓	✓			
Tung et al [53]	Mechanisms of action	✓	✓		✓					

[a]RQ: research question.

[b]Wb: websites and online telecommunication.

[c]Ap: apps and software.

[d]Ms: messaging services.

[e]An: anthropometric measures (BMI, weight, waist circumference, etc).

[f]Am: amount of physical activity (steps, duration, metabolic equivalent of task, etc).

[g]D: dietary intake.

[h]PR: physical readiness (maximal oxygen uptake and aerobic readiness).

Efficacy

Table 8 summarizes the outcomes reported for weight management DWPs. As expected, the most frequently discussed outcomes are related to anthropometric measures such as BMI, weight, and waist circumference. The only meta-analysis [16] conducted in this domain failed to establish significant support for the efficacy of DWPs on these measures (g=0.02, 95% CI −0.07 to 0.10, k=4, N=not available, I^2=0%). However, several empirical studies reported in non–meta-analyses show a significant effect of DWPs on anthropometric outcomes. Specifically, 2 non–meta-analyses [18,52] found evidence from high-quality studies indicating small to modest reductions in weight and waist circumference.

Furthermore, 1 non–meta-analysis review [42] found unanimous evidence supporting the effectiveness of weight management DWPs that adopt a targeted approach. These interventions used various strategies such as teleconsultation, health data monitoring, personalized advice, and the provision of online resources. This review highlights the notably high success rates of targeted weight management DWPs, a valuable finding given that obese employees tend to have higher absenteeism, greater health risks, and increased health care costs [69,71].

Secondary outcomes of weight management DWPs include the amount of physical activity, dietary intake, and physical readiness (eg, maximal oxygen uptake and aerobic capacity). Only 1 non–meta-analysis [18] included these secondary outcomes but did not find significant changes. This lack of evidence has been attributed to the limited number of empirical studies reporting on these secondary outcomes.

XSL•FO
RenderX

Table 8. Summary of weight management digital wellness program outcomes.

Study	Outcomes			
	An[a]	Am[b]	D[c]	PR[d]
Jung and Cho [40]	ns[e]	—[f]	—	—
Aneni et al [18]	~[g] (2/6)[h]	-[i] (0/3)	- (0/2)	- (0/1)
Lee et al [42]	+[j] (11/11)	—	—	—
Thai et al [52]	~ (3/4)	—	—	—

[a]An: anthropometric measures (BMI, weight, waist circumference, etc) and blood test indicators (cholesterol, blood pressure, etc).

[b]Am: amount of physical activity (steps, duration, metabolic equivalent of task, etc).

[c]D: dietary intake.

[d]PR: physical readiness (maximal oxygen uptake and aerobic readiness).

[e]In meta-analysis studies, ns denotes nonsignificant results.

[f]Not available.

[g]In non–meta-analysis studies, ~ denotes found conflicting evidence.

[h](#/#) = (number of studies with significant evidence for this outcome) / (total number of studies reporting this outcome).

[i]In non-meta-analysis reviews, - stands for found no significant evidence.

[j]In non–meta-analysis studies, + denotes found unanimous evidence in favor of the effectiveness.

Mechanisms of Action

One non–meta-analysis review [53] examines the mechanisms of action in weight management programs. This review highlights that digital technologies can support various behavioral change strategies commonly used in these programs, such as enablement and education. The enablement strategy includes behavior change techniques such as self-monitoring of sitting time, self-monitoring of dietary behavior, and telemonitoring of physical activity data by a remote coach. The education strategy involves techniques such as delivering instructive content about nutrition, health risks, and physical activity through various online channels. Although this literature review summarizes the range and frequency of these behavior change techniques, it is unable to identify which specific techniques are most effective due to the heterogeneity of the intervention characteristics.

Unhealthy Behavior Change

Overview

The DWPs in this category are designed to modify behavioral patterns associated with higher health risks, such as smoking, poor eating habits, and excessive alcohol consumption. These modifiable unhealthy behaviors are linked to the earlier onset of chronic diseases and can lead to avoidable costs for organizations, including increased absenteeism and health care expenses [72,73]. As shown in Table 9, we identified 4 non–meta-analysis reviews related to unhealthy behavior change; however, only one of these reviews [51] exclusively focuses on unhealthy behavior change as its primary topic of interest.

Table 9. Mapping of reviews on unhealthy behavioral change digital wellness programs.

Study	Type of RQ[a]	Digital technologies					Outcomes				Unhealthy behavior addressed
		Wb[b]	Ap[c]	Email	Ms[d]	Wearables	BC[e]	An[f]	MH[g]	WO[h]	
Aneni et al [18]	Efficacy	✓		✓			✓	✓			Diet and smoking
Phillips et al [16]	Efficacy	✓					✓				Alcohol
Sevic et al [48]	Mechanisms of action	✓					✓				Diet and alcohol
Sundstrom et al [51]	Efficacy	✓		✓			✓		✓	✓	Alcohol (exclusively)

[a]RQ: research question.

[b]Wb: websites and online telecommunication.

[c]Ap: apps and software.

[d]Ms: messaging services.

[e]BC: behavioral change (eg, change in dietary choices and frequency of drinking and smoking).

[f]An: anthropometric measures (BMI, weight, waist circumference, etc).

[g]MH: mental health (eg, stress, anxiety, and depression).

[h]WO: work-related outcomes (eg, time management on the job and rumination in the work).

XSL•FO
RenderX

Efficacy

Table 10 summarizes the findings for the 3 types of unhealthy behaviors targeted by DWPs: smoking, poor diet, and excessive alcohol consumption. The most frequently cited outcomes in these studies relate to changes in the occurrence of the targeted behaviors, such as the frequency of alcohol consumption and the amount of healthy food intake. All included non–meta-analyses [16,18,48,51] reported conflicting evidence regarding the efficacy of DWPs in reducing smoking, improving diet, and curbing alcohol consumption. It is important to note that all the diet interventions included in these reviews were a part of multiple-component health interventions that also incorporated physical activity or weight management modules. For their part, 2 non–meta-analyses evaluated secondary outcomes of behavioral change DWPs. Aneni et al [18] found conflicting evidence regarding the effectiveness of diet-focused DWPs, which showed effectiveness in reducing weight and blood pressure. Sundstrom et al [51] also found conflicting evidence on alcohol-related DWPs, specifically concerning their impact on employees' mental health and work-life balance.

Table 10. Summary of unhealthy behavioral change digital wellness program outcomes.

Study	Outcomes		
	BC[a]	An[b]	MW[c]
Diet			
Aneni et al [18]	~[d] (6/7)[e]	~ (5/6)	—[f]
Sevic et al [48][g]	~ (3/4)	—	—
Smoking			
Aneni et al [18]	~ (3/4)	—	—
Alcohol			
Sevic et al [48]	+[h] (2/2)	—	—
Phillips et al [16]	~ (2/5)	—	—
Sundstrom et al [51]	~ (4/7)	—	~ (4/5)

[a]BC: behavioral change (eg, change in dietary choices, frequency of drinking and smoking).

[b]An: anthropometrics (weight and blood pressure).

[c]MW: mental well-being (eg, stress, anxiety, and depression) and work-life balance (eg, work strain and time management).

[d]In non–meta-analysis studies, + denotes found unanimous evidence in favor of the effectiveness.

[e](#/#) = (number of studies with significant evidence for this outcome) / (total number of studies reporting this outcome).

[f]Not available.

[g]The diet interventions included in this review were in a multicomponent health intervention.

[h]In non–meta-analysis studies, ~ denotes found conflicting evidence.

Mechanisms of Action

One non–meta-analysis review [48] outlines the behavior change techniques used in DWPs aimed at dietary improvements and alcohol reduction. According to this review, digital technologies facilitate several behavior change techniques, including prompting employees, providing feedback, enhancing self-belief, offering social support, giving instructions to support behavior change, introducing nudging objects to employees' immediate environment, and raising awareness about the consequences of unhealthy behaviors.

Sleep Management

Overview

The DWPs in this health domain aim to improve sleep quality and, consequently, enhance daytime alertness and performance at work. Poor sleep quality among employees is associated with productivity losses, an increased risk of chronic conditions, slower and less-accurate cognitive responses, and a higher likelihood of fatal occupational accidents [74]. Sleep management is especially crucial for shift workers in stressful or high-risk occupations. As shown in Table 11, we identified 2 non–meta-analysis reviews [19,52] that focus on the efficacy of sleep management DWPs. Bullock et al [19] examine sleep management programs delivered via mobile apps specifically designed to help shift workers. Thai et al [52] focus on sleep management interventions implemented in organizations located in low- and middle-income countries.

XSL•FO
RenderX

Table 11. Mapping of reviews on sleep management digital wellness programs.

Study	Type of RQ[a]	Digital technologies					Outcomes	
		Wb[b]	Ap[c]	Email	Ms[d]	Wearables	SM[e]	SS[f]
Bullock et al [19][g]	Efficacy		✓				✓	✓
Thai et al [52]	Efficacy		✓		✓		✓	✓

[a]RQ: research question.

[b]Wb: websites and online telecommunication.

[c]Ap: apps and software.

[d]Ms: messaging services.

[e]SM: sleep-related measures (sleep duration, sleep debt, sleep difficulties, etc).

[f]SS: symptoms of poor sleep (daytime fatigue and daytime sleepiness).

[g]This study is only limited to shift workers.

Efficacy

As shown in Table 12, the outcomes of sleep management DWPs are categorized into 2 main areas. The first category involves measures of sleep quantity and quality. Two non–meta-analysis reviews [19,52] found unanimous evidence supporting the effectiveness of DWPs in improving sleep-related outcomes. The second category pertains to symptoms of poor sleep, such as daytime fatigue and sleepiness. Similarly, both reviews found consistent evidence supporting the effectiveness of sleep management DWPs in alleviating these symptoms. Thai et al [52] suggest that sleep interventions that use digital technologies to provide tailored advice to employees are more effective than those offering on-demand content. However, both reviews caution that the body of evidence is still too small to draw definitive conclusions, and thus, the findings should be interpreted with care.

Table 12. Summary of sleep management digital wellness program outcomes.

Study	Outcomes	
	SM[a]	SS[b]
Bullock et al [19]	+[c] (1/1)[d]	+ (1/1)
Thai et al [52]	+ (1/1)	+ (2/2)

[a]SM: sleep-related measures (eg, sleep duration, sleep debt, sleep difficulties).

[b]SS: symptoms of poor sleep (eg, daytime fatigue and daytime sleepiness)

[c]In non–meta-analysis studies, + denotes found unanimous evidence in favor of the effectiveness.

[d](#/#) = (number of studies with significant evidence for this outcome) / (total number of studies reporting this outcome).

Other Health Domains

The final category includes interventions that do not fit into the previously discussed domains. Three of the identified reviews [45,48,52] only touch on these DWPs tangentially. Interventions in this category address diverse issues, such as health risk assessments, reducing risk factors for chronic diseases, improving workplace ergonomics, and lowering blood glucose levels in at-risk employees. Empirical studies referencing these interventions report significant outcomes. However, due to the heterogeneity of the health issues addressed and the limited empirical evidence available, we were unable to synthesize or aggregate the findings. As a result, we acknowledge their presence within this body of research without attempting further analysis.

Discussion

In this section, we synthesize and critically analyze the findings on DWPs with a focus on understanding efficacy, durability, mechanisms of action, acceptability, and targeting approaches across the identified health domains. While the findings reveal promising outcomes in several areas, they also underscore significant limits and gaps in the literature that necessitate further investigation to fully harness the potential of these programs.

Principal Findings

The findings on DWPs indicate that navigating this diverse body of evidence requires a detailed examination of various aspects of these programs, including their targeted health domain, their reported outcomes, and their recruitment strategies. Principal findings have been summarized according to these aspects.

First, this meta-review shows mixed or conflicting evidence regarding the efficacy of DWPs for mental health and physical activity, the two most studied domains. In the mental health domain, while some reviews highlight small to medium positive effects on outcomes such as stress, anxiety, and depression, others report conflicting or inconclusive findings. For example, while some interventions lead to meaningful reductions in stress and anxiety [16,17], these effects are not consistently observed

XSL•FO
RenderX

across all reviews [43,46]. Similarly, for physical activity, the efficacy of DWPs varies significantly. Although several studies indicate increases in the amount of physical activity mainly through goal-setting and self-monitoring features (eg, [35,40]), others show limited impact. This suggests that factors such as intervention design and participant engagement may influence outcomes. Consequently, while DWPs hold potential for improving mental health and physical activity, the inconsistent evidence highlights the need for deeper exploration of the factors that make certain interventions more successful than others.

Second, in the domain of weight management, non–meta-analyses [18,42,52] report significant reductions in BMI and body weight, especially among individuals with obesity participating in DWPs [42]. These results suggest that targeted interventions aimed at high-risk populations are more effective than generic programs. Although non–meta-analyses report many interventions with significant improvements in BMI and body weight, the only meta-analysis study on weight management DWPs [40] failed to show a significant change. The lack of convergence between the non–meta-analyses and the meta-analysis study in this health domain suggests that factors such as health risk perception may influence intervention efficacy.

Third, the evidence for unhealthy behavioral change, such as smoking cessation and alcohol reduction, is promising but limited. While the available non–meta-analysis reviews [16,18,48,51] indicate significant reductions in unhealthy behaviors, there is a lack of sufficient studies to provide comparative insights. Thus, while DWPs appear effective for behavior change, the relative effectiveness of different interventions remains unclear due to the limited evidence base.

Fourth, non–meta-analysis reviews in the sleep management domain show promising effects of DWPs on sleep outcomes, such as duration, latency, and quality [19,52]. The agreement between studies highlights the efficacy of DWPs in improving sleep health, but the body of evidence is limited to a few studies.

Fifth, DWPs are generally well-accepted by participants, particularly when they are personalized, user friendly, and easily accessible. However, acceptability measures are not free from bias. For example, satisfaction surveys typically gather feedback only from participants who complete the program, making these results susceptible to survival bias. In other words, while high satisfaction scores may suggest the program's success, they can inflate acceptability measures by excluding the opinions of dissatisfied users who dropped out. Similarly, compliance rates, often gauged through log-in data, may understate the program's success if users disengage after achieving their desired outcomes. It is crucial for researchers to recognize and account for these limitations in their studies.

In addition, the findings of this meta-review emphasize that the effectiveness and acceptability of DWPs are influenced by how user friendly and personalized they are. This aligns with prior research [75] that indicates the effective use of self-monitoring technologies depends on users' attitude toward the system and their perceived role in managing their conditions. In this respect, providing more guidance to reliant users (those who assume a passive role in managing their health) and greater autonomy to engaged users (those who assume an active role in managing their health) could enhance program success. This suggests customizing DWPs to individual user profiles can maximize benefits.

Finally, this meta-review highlights that the recruitment strategy of DWPs plays a crucial role in determining their efficacy. DWPs that use a targeted approach, recruiting participants based on risk factors or health conditions, tend to produce larger effect sizes than those with a universal approach. Review studies focusing on mental health, weight management, and unhealthy behavioral change demonstrate that targeted programs tend to yield better outcomes. In contrast, DWPs with a universal approach may produce smaller effect sizes, likely due to the dilution of outcomes across a more diverse participant pool. For example, individuals with moderate alcohol consumption may show minimal change, reducing the overall impact measured in the program [51]. Therefore, it is important to interpret the results of DWPs in light of their recruitment strategy and acknowledge that universal programs may not achieve the same level of effectiveness as targeted interventions.

Gaps and Promising Avenues for Research

This study exposed critical gaps and highlighted areas where further research is needed. The meta-review approach allowed us to survey a rich body of research that focuses on the same phenomenon but progresses in parallel with their focus on different health domains. We identified several gaps that were shared across various health domains. However, some gaps were identified by borrowing ideas from the research in other health domains. For example, a lack of focus on mechanisms of action in the mental health domain became apparent after comparing reviews in this domain with the ones in other health domains. In what follows, we highlight areas with significant gaps and provide recommendations to guide future research. The most important area concerns the inconsistent findings that were observed for the two most studied domains: mental health and physical activity.

Regarding mental health, the inconsistent findings are linked primarily to outcomes such as stress, anxiety, and depression. To resolve this inconsistency by explaining when DWPs have beneficial effects, moderation analysis is useful. This type of analysis examines whether the effects of DWPs, such as their stress-reducing effects depend on certain moderating factors [76]. Since technology is at the core of DWPs, the literature on technostress provides several moderating factors that are worth exploring. In particular, self-efficacy is a relevant factor because it assesses the extent to which persons believe that they have the ability to use technologies such as DWPs successfully (eg, [77,78]). Usually, people with higher levels of self-efficacy are more motivated to use these technologies effectively and make a greater effort to learn about them, whereas individuals with lower levels of self-efficacy give up more easily when the benefits of using the digital technology do not accrue immediately. This suggests that most people with higher levels of self-efficacy will use DWPs more effectively and will derive greater mental health benefits from using them than those with lower levels of it. However, this might not always be the case as the success of a DWP likely depends on its design elements

XSL•FO
RenderX

and content. Even expert users with high levels of self-efficacy might become frustrated or disengaged when DWPs are not well-designed or are not sufficiently user friendly. For example, poor navigation features, a clunky interface, or overly simplistic features might all hinder engagement with DWPs and thus frustrate users. This means that not all users with high levels of self-efficacy will derive greater mental health benefits from DWPs. Notwithstanding this fact, self-efficacy holds the potential to resolve some of the conflicting findings in the literature (eg, [79]).

Other moderating factors from the technostress literature are also relevant. Worker age is an important factor because older workers experience a greater mental workload when they engage with technologies such as DWPs [80-83]. Since this increased mental workload leads to stress [83], it can undermine the stress-reducing potential of DWPs. Other relevant moderating factors include differing levels of motivation to improve health-related outcomes, differing levels of IT support available to use DWPs, and differing levels of experience with such digital tools as messaging services, online telecommunication, and videoconferencing apps [78,83-86]. These factors can impact how effectively workers engage with DWPs and thus affect the degree to which the potential benefits of these technologies are realized. DWPs can be expected to lead to greater benefits when users are more motivated to improve their health when they receive more IT support, and when they have more experience with relevant tools. This means that these factors can help explain some of the conflicting findings in the extant literature.

However, the big 5 personality traits are likely the most promising moderating factors because they have been shown to moderate various relationships between engagement with technology and stress-related variables (eg, [87-89]). They have also been shown to affect the stress process more broadly [90], and they include 5 constructs: neuroticism, openness, agreeableness, extraversion, and conscientiousness. Regarding the moderating impact of personality traits on the stress-reducing potential of DWPs, individuals high in neuroticism are likely to report lower mental health benefits from engaging with DWPs. This is because they tend to experience greater dissatisfaction, frequently complain, and perceive even minor frustrations as overwhelming, making them less likely to find DWPs beneficial. In contrast, individuals high in openness are typically curious and willing to explore new experiences, which can lead to greater satisfaction with DWPs. Their openness to new approaches may result in more positive evaluations of the benefits of engaging with these systems, making them more likely to perceive DWPs as beneficial.

Regarding agreeableness, individuals high in this trait tend to be cooperative and are more likely to follow the advice of nurses, doctors, and artificial intelligence chatbots [87-89]. As a result, they are likely to derive greater benefits from their engagement with DWPs compared with those with lower agreeableness. Similarly, extraverted individuals are typically social, active, and outgoing. They often use technology strategically to maintain a positive social image, which suggests that they may view DWPs as a tool for sustaining well-being, job performance, and influence within the organization. Consequently, they are more likely to engage effectively with DWPs and perceive their engagement as beneficial. Finally, conscientious individuals tend to be detail oriented, dependable, and persistent. Even if they experience initial frustration, their intrinsic motivation and strong work ethic make them more likely to persist in using DWPs. Their commitment to efficiency and effectiveness further enhances their likelihood of successfully integrating DWPs into their routines.

Regarding physical activity, our study highlights that goal setting is a widely used behavior change technique. Goal-setting theory suggests several important moderators that influence its effectiveness. Specifically, for goal setting to be successful, individuals must possess both the ability and the situational resources necessary to achieve the goal [91]. In addition, goal importance and commitment to goal attainment serve as key moderators of effectiveness. However, one particularly noteworthy moderator is feedback, which plays a crucial role in goal progress. Regular feedback enables individuals to adjust their effort and refine their strategies, ensuring they stay on track toward goal attainment [91]. This suggests that the 2 primary drivers of increased physical activity in DWPs, goal-setting and self-monitoring features, may interact. Future research could explore whether the impact of goal setting on physical activity is enhanced when combined with self-monitoring features that provide a continuous stream of relevant feedback.

This meta-review also sheds light on methodological and structural attributes of the literature that future research should focus on. One such attribute across all health domains is the scarcity of long-term studies. Although many DWPs demonstrate initial effectiveness, the sustainability of these effects remains largely unexplored. For example, prior literature reviews on mental health and physical activity [17,35,38] have reported a decline in positive outcomes after a few months. However, the specific conditions under which these reductions occur, and whether they are more pronounced in DWPs compared with nondigital wellness programs, are still unclear. A recent study provides evidence of a temporary digital placebo effect in smartphone-based mental health interventions [92]. The digital placebo occurs when users experience benefits from digital technologies due to their beliefs in the tools rather than from their actual functionalities [93]. Initial expectancy and credibility of digital tools can taper off and potentially lead to reduced effectiveness in outcomes [92]. Yet, it is not clear who is more prone to this effect, what factors in intervention designs exacerbate it, and what practices can mitigate its impact. Hence, future research must prioritize longitudinal studies that answer these questions regarding DWPs and potentially compare them with their nondigital counterparts.

Another significant gap in the literature is the limited understanding of the mechanisms that drive the success of DWPs. Only 4 reviews in our sample have delved into the specific behavior change techniques that are used in these programs [35,37,48,53]. While these reviews offer valuable insights, many questions remain unanswered regarding the comparative advantage of various techniques. Surprisingly, despite the numerous reviews conducted on mental health DWPs, we did not find any review that examines the mechanisms of action used in this domain. Although some

XSL•FO
RenderX

reviews [17,57] compare effect sizes of various therapeutic approaches such as cognitive behavioral therapy and positive psychology, it remains unclear whether their success is attributable to the therapeutic approaches or the effective activation of causal mechanisms of action by digital interventions. Psychotherapy scholars have recently called for the integration of these mechanisms into the research on mental health interventions [94-96]. This allows researchers to develop treatment strategies that focus on the most important mechanisms [96] and to examine moderating factors that influence active mechanisms of change [95].

Moreover, recent research underscores the importance of studying behavior change techniques in light of the causal mechanisms of change [97,98]. For example, Carey et al [98] provide an extensive list of causal mechanisms of action used in behavioral interventions. Future research on DWPs can leverage this list to examine how IT-related factors (eg, design and personal innovativeness with IT) and organization-related factors (eg, organizational culture and perceived organizational support) can weaken or strengthen the mechanisms of action applied in DWPs.

In a similar vein, prior reviews [35,48] call attention to the lack of strong theoretical foundations in prior research. We echo this concern and highlight 2 potential consequences that have created a gap in our understanding of DWPs. First, an atheoretical approach to designing health interventions limits our ability to generalize findings. For example, combining behavior change techniques can "lead to combinations of counteracting mechanisms in mHealth designs that obfuscate the techniques' effects" [99]. Leveraging health and behavior change theories, researchers can articulate the boundary conditions of compatible interventions. Second, an atheoretical approach limits our ability to understand why and under what conditions DWPs influence organization-related outcomes such as burnout. For example, the theory of job demands-resources [100] provides a useful perspective in studying the organization-related outcomes of mental health DWPs. However, it should be further elaborated to explain how the characteristics of DWPs can contribute to job resources. To advance the field, it is essential that future studies are grounded in relevant theoretical frameworks, elucidate the processes by which DWPs influence health- and organization-related outcomes, and ultimately create value in organizations.

This meta-review also highlights limitations in the scope of interventions examined in the literature. While existing studies primarily focus on individual-level outcomes, such as personal health and productivity, wellness interventions can also be analyzed at higher organizational levels, including work groups and entire organizations [101]. Research on nondigital wellness initiatives has demonstrated organization-level outcomes, such as insurance claims and cost-benefit analyses of wellness programs [102]. However, similar outcomes have not yet been examined in the context of digital wellness initiatives. To address this critical research gap, future studies could leverage archival data, such as insurance claims and unit performance reports, to assess the broader organizational impact of DWPs.

As mentioned earlier, the variability in outcome measures across studies presents a significant barrier to synthesizing findings and drawing generalizable conclusions. Particularly in the domains of physical activity and mental health, there is a need for standardized outcome measures that can be consistently applied across studies. This would allow for more robust comparisons and a clearer understanding of the true impact of DWPs. Moreover, most studies focus on health-related outcomes neglecting the organizational implications of these programs. Therefore, there is a clear need for more research that links DWPs to positive organizational outcomes such as productivity, presenteeism, retention, and job satisfaction. Understanding these links is crucial for demonstrating the value of DWPs to employers and other stakeholders.

Finally, the role of technology in DWPs is often treated superficially, with scant attention given to how specific technological features influence the efficacy and acceptability of these programs. For example, engagement is an understudied factor in understanding the efficacy of DWPs beyond common measures of acceptability, such as frequency of use. Engagement in digital interventions is a multidimensional construct that involves physical, cognitive, and affective investment in the stimulus (eg, app) [103]. Research shows that using more features of wearable trackers (ie, a form of physical engagement) is associated with higher health-related outcomes [15]. Future research should demonstrate how and under what conditions each dimension of engagement contributes to intervention outcomes. Moreover, despite the large body of research on engagement with technology, there is limited understanding of how to enhance engagement in the technological stimulus in DWPs. For example, Nahum-Shani et al [103] suggest that environmental dynamics and participants' information overload are important factors in the context of information-rich digital interventions. In this respect, future research should elucidate how employee-related factors (eg, cognitive and emotional resources) and organizational environment (eg, supportive climate) contribute to the engagement and eventually to the success of DWPs.

In addition, abstracting the technology components of DWPs into a device or mobile app without considering the underlying function of technology downplays its contribution to the success of DWPs. For example, a mobile app can have various functions such as delivering nudges and facilitating self-monitoring. The reviews in our corpus did not present their findings in this aspect, suggesting that accumulated evidence is lacking regarding the role of technology in DWPs. Future research should delve deeper into the technological aspects of DWPs by examining how different features and components of digital tools contribute to or detract from the success of digital interventions.

Implications for Practice

The findings of this meta-review hold significant implications for organizations, particularly for human resource (HR) managers tasked with improving employee health and well-being. While DWPs show potential in promoting employee wellness, their effectiveness is not consistent across all health domains. The mixed evidence on mental health and physical

XSL•FO
RenderX

activity—the two most studied areas—suggests that HR managers should have measured expectations regarding the immediate impact of these programs. Although DWPs can positively influence outcomes such as stress, anxiety, and physical activity, their success depends largely on program design, employee engagement, and the specific context of implementation. Therefore, HR managers should focus on tailoring DWPs to address the diverse needs of their workforce, ensuring interventions are personalized and adaptable. Conducting employee wellness assessments (eg, anonymous surveys, biometric screenings, or focus groups) can help organizations identify key health risks, understand employee preferences, and gauge engagement levels, allowing them to tailor interventions accordingly. By leveraging data-driven insights, HR managers can design more personalized and adaptable programs, ensuring that DWPs are aligned with employees' specific needs rather than taking a one-size-fits-all approach. In addition, these assessments provide valuable feedback loops, enabling organizations to refine and optimize their wellness strategies over time, ultimately improving program effectiveness and employee well-being.

Given the variability in outcomes, regular monitoring and assessment are essential to ensure that DWPs remain relevant and impactful. HR managers should evaluate the programs over time by seeking employee feedback (eg, satisfaction) and monitoring various acceptability measures (eg, uptake rate, completion rate, and engagement) and effectiveness measures (eg, individual and organization-related outcomes). Relying solely on any one of these measures can provide a narrow perspective on the success of the programs. Implementing regular monitoring mechanisms, including A/B testing of intervention approaches, can help refine these initiatives for greater impact.

While DWPs have shown promise in facilitating behavioral changes such as smoking cessation and alcohol reduction, the evidence remains limited. HR managers should not expect immediate or significant results in these areas. To enhance the impact of DWPs on behavior change, these programs should be integrated with broader organizational strategies such as health coaching, peer support groups, and incentives for healthy habits. Creating a supportive culture that encourages ongoing healthy lifestyle choices can help sustain these behavioral changes over time. Organizations should embed digital wellness check-ins into daily workflows and encourage leadership to model healthy behaviors, such as setting boundaries for after-hours emails.

Technology plays a crucial role in the success of DWPs. Wearables, mobile apps, and online platforms have proven effective in promoting the health of employees in various domains. They enable many opportunities for implementing organizational interventions that can easily be scaled up. However, as reflected in the acceptability measures discussed in the literature, HR managers should acknowledge that there is a great deal of variation in terms of adoption, continuous use, and the extent of engagement with technological initiatives among employees. Examining the compatibility of technologies with the target population before implementing DWPs can contribute to the success of such programs. For example, if the target population of a weight reduction DWP is employees with obesity, the organization can assess their self-management style [75] before investing in those programs that highly depend on self-monitoring devices. Personalization of wellness plans, offering a mix of mobile apps, virtual coaching, and wearable devices, can further improve engagement by catering to diverse employee preferences.

For DWPs to be truly impactful, they must be part of a larger organizational culture that prioritizes employee well-being. HR managers play a pivotal role in promoting this culture by embedding wellness as a core organizational value. This involves not only offering DWPs but also aligning organizational policies with initiatives that support work-life balance, stress management, and mental health. A culture of shared responsibility for wellness can significantly enhance the effectiveness of DWPs, leading to more positive outcomes for both employees and the organization.

In addition to improving employee health, HR managers can consider the return on investment for the organization. Evidence from this meta-review suggests that DWPs can potentially impact organizational outcomes such as engagement, absenteeism, and productivity, particularly when addressing prevalent issues such as mental health, sedentarism, and weight management. Highlighting potential cost savings and productivity gains can help HR managers present DWPs as both a health initiative and a strategic investment, securing the necessary resources and support for successful implementation.

In conclusion, DWPs offer organizations a valuable opportunity to improve employee health, enhance productivity, and reduce health care costs. However, the success of these programs depends on thoughtful implementation, continuous evaluation, and a supportive organizational culture. HR managers are at the forefront of this effort, playing a critical role in designing, promoting, and sustaining effective DWPs that meet the diverse needs of their workforce. By leveraging cross-disciplinary expertise and integrating digital health interventions within existing workplace structures, organizations can maximize the potential of DWPs for long-term employee well-being and organizational success.

Study Limitations and Future Research

This meta-review has several limitations. First, the initial abstract screening was conducted by a single reviewer, which may have introduced some selection bias. However, to minimize this risk, the reviewer applied a conservative approach, excluding studies only when it was certain they did not meet the inclusion criteria. Any study with uncertainty was flagged and discussed in consensus meetings among the authors before a final decision was made. While this approach helped mitigate bias, future meta-reviews may benefit from a dual-independent screening process to further enhance reliability. Second, the quality and scope of the included reviews varied, which could have affected the generalizability of our findings. Finally, while our review methodology was efficient, it may have missed recent studies due to publication delays. Future research should address these limitations by incorporating a broader range of studies and tackling more specific RQs inspired by this meta-review. In addition, there is a need for more rigorous methodological

XSL•FO
RenderX

approaches, particularly in examining long-term efficacy, mechanisms of action, and the role of technology in DWPs. Researchers should also prioritize the development of standardized outcome measures and further explore the organizational implications of DWPs, which remain underrepresented in the literature.

Conclusions

This meta-review provides a comprehensive synthesis of DWPs across various health domains, offering valuable insights into their efficacy and acceptability. The findings highlight promising outcomes, particularly in the areas of mental health, physical activity, and weight management, where interventions typically led to small to medium improvements. However, some inconsistencies persist in the extant literature that future studies must address. This study also reveals variability in efficacy depending on recruitment strategies, with targeted DWPs generally producing more substantial effects compared with universal programs. Significant gaps also remain, especially the scarcity of long-term studies and the limited understanding of the mechanisms driving DWP success. It also emphasizes the need for theory-driven research to better explain how and why certain DWPs are effective. For HR managers and organizations, the findings suggest that implementing tailored DWPs can enhance employee well-being and organizational outcomes. However, ongoing evaluation and adaptation are crucial to ensure these programs achieve their intended goals. In summary, while DWPs show great potential for improving employees' wellness, further research is necessary to optimize their design and implementation to ensure they deliver sustainable benefits in diverse workplace settings.

Authors' Contributions

SA conceptualized this study, conducted the coding, drafted the initial paper, and contributed to its review and editing. GP conceptualized this study, participated in coding, and reviewed and edited this paper. ST conceptualized this study, participated in coding, and reviewed this paper. All authors had full access to this study's data, including supplemental material, and approved the final version of this paper for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) checklist.
[DOCX File , 85 KB - jmir_v27i1e70982_app1.docx]

Multimedia Appendix 2
Full search strings.
[DOCX File , 18 KB - jmir_v27i1e70982_app2.docx]

References

1. De Neve JE, Kaats M, Ward G. Workplace wellbeing and firm performance. Wellbeing Research Centre. 2024. URL: https://wellbeing.hmc.ox.ac.uk/wp-content/uploads/2023/11/2304-WP-Workplace-Wellbeing-and-Firm-Performance-DOI-2024.pdf [accessed 2024-04-29]
2. Wolf J, Prüss-Ustün A, Ivanov I, Mugdal S, Corvalán C, Bos R, et al. Preventing disease through a healthier and safer workplace. World Health Organization. URL: https://iris.who.int/bitstream/handle/10665/272980/9789241513777-eng.pdf [accessed 2025-02-01]
3. O'Boyle E. 4 things gen Z and millennials expect from their workplace. Gallup. 2021 Mar 30. URL: https://web.archive.org/web/20250201024432/https://www.gallup.com/workplace/336275/things-gen-millennials-expect-workplace.aspx [accessed 2024-04-29]
4. Gordon D. 82% of workers expect employers to support well-being, yet 1/4 feel bad at work, new survey shows. Forbes. URL: https://www.forbes.com/sites/debgordon/2023/09/30/82-of-workers-expect-employers-to-support-well-being-yet-14-feel-bad-at-work-new-survey-shows/ [accessed 2024-04-29]
5. Claxton G, Rae M, Young G, McDermott D. 2023 employer health benefits survey. Kaiser Family Foundation. URL: https://files.kff.org/attachment/Employer-Health-Benefits-Survey-2023-Annual-Survey.pdf [accessed 2025-11-17]
6. Berry LL, Mirabito AM, Baun WB. What's the hard return on employee wellness programs? Harvard Business Review. 2010. URL: https://hbr.org/2010/12/whats-the-hard-return-on-employee-wellness-programs [accessed 2023-06-18]
7. Baicker K, Cutler D, Song Z. Workplace wellness programs can generate savings. Health Aff (Millwood) 2010 Feb;29(2):304-311. [doi: 10.1377/hlthaff.2009.0626] [Medline: 20075081]
8. Medina E, Prieto L. Moderating effects of work-life balance programs' perceived value on relationships between organizational support and employee outcomes. J Manag Issues 2022;34(1):61-81.

XSL•FO
RenderX

9. Muñoz RM, Andrade SM, Peña I, Donate MJ. Wellness programs in times of COVID-19, perceived organizational support and affective commitment: effects on employee innovative behavior. Eur J Innov Manag 2022 Jun 13;26(7):23-44. [doi: 10.1108/ejim-02-2022-0072]
10. Peña I, Andrade SM, María Muñoz R, Barba-Sánchez V. Wellness programs, perceived organizational support, and their influence on organizational performance: an analysis within the framework of sustainable human resource management. Sage Open 2024 Feb 14;14(1):6520. [doi: 10.1177/21582440241229358]
11. Alonso-Nuez MJ, Cañete-Lairla M, García-Madurga M, Gil-Lacruz AI, Gil-Lacruz M, Rosell-Martínez J, et al. Corporate social responsibility and workplace health promotion: a systematic review. Front Psychol 2022;13:1011879 [FREE Full text] [doi: 10.3389/fpsyg.2022.1011879] [Medline: 36312126]
12. Bruni EA, Andrei F, Tirabeni L. Engaging the body, appropriating a corporate wellness programme. Qual Res Organ Manag 2022 Aug 12;17(5):88-107. [doi: 10.1108/qrom-02-2022-2296]
13. Ghose A, Guo X, Li B, Dang Y. Empowering patients using smart mobile health platforms: evidence of a randomized field experiment. arXiv Preprint posted online on February 10, 2021 [FREE Full text] [doi: 10.25300/misq/2022/16201]
14. Babar Y, Curley S, Ke Z, Liu D, Sheffler Z. The effects of digitally delivered nudges in a corporate wellness program. J Assoc Inf Syst 2023;24(1):136-160. [doi: 10.17705/1jais.00783]
15. Giddens L, Leidner D, Gonzalez E. The role of Fitbits in corporate wellness programs: does step count matter? In: Proceedings of the 50th Hawaii International Conference on System Sciences. 2017 Presented at: HICSS '17; January 4-7, 2017; Hilton Waikoloa Village, HI URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/022b4c5b-084e-4110-988c-8c9bd62cad8c/content [doi: 10.24251/hicss.2017.438]
16. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 01;45(6):560-576 [FREE Full text] [doi: 10.5271/sjweh.3839] [Medline: 31184758]
17. Stratton E, Lampit A, Choi I, Malmberg Gavelin H, Aji M, Taylor J, et al. Trends in effectiveness of organizational eHealth interventions in addressing employee mental health: systematic review and meta-analysis. J Med Internet Res 2022 Sep 27;24(9):e37776 [FREE Full text] [doi: 10.2196/37776] [Medline: 36166285]
18. Aneni EC, Roberson LL, Maziak W, Agatston AS, Feldman T, Rouseff M, et al. A systematic review of internet-based worksite wellness approaches for cardiovascular disease risk management: outcomes, challenges and opportunities. PLoS One 2014;9(1):e83594 [FREE Full text] [doi: 10.1371/journal.pone.0083594] [Medline: 24421894]
19. Bullock B, Learmonth C, Davis H, Al Mahmud A. Mobile phone sleep self-management applications for early start shift workers: a scoping review of the literature. Front Public Health 2022;10:936736 [FREE Full text] [doi: 10.3389/fpubh.2022.936736] [Medline: 36033744]
20. Paré G, Kitsiou S. Methods for literature reviews. In: Lau F, Kuziemsky C, editors. Handbook of eHealth Evaluation: An Evidence-Based Approach. Victoria, BC: University of Victoria; 2017:157-180.
21. Cowell JM. The advantage of literature reviews for evidence-based practice. J Sch Nurs 2015 Feb;31(1):5. [doi: 10.1177/1059840514564387] [Medline: 25631692]
22. Sarrami-Foroushani P, Travaglia J, Debono D, Clay-Williams R, Braithwaite J. Scoping meta-review: introducing a new methodology. Clin Transl Sci 2015 Feb;8(1):77-81 [FREE Full text] [doi: 10.1111/cts.12188] [Medline: 25041546]
23. Ringeval M, Etindele Sosso FA, Cousineau M, Paré G. Advancing health care with digital twins: meta-review of applications and implementation challenges. J Med Internet Res 2025 Feb 19;27:e69544 [FREE Full text] [doi: 10.2196/69544] [Medline: 39969978]
24. Hennessy EA, Johnson BT, Keenan C. Best practice guidelines and essential methodological steps to conduct rigorous and systematic meta-reviews. Appl Psychol Health Well Being 2019 Nov;11(3):353-381 [FREE Full text] [doi: 10.1111/aphw.12169] [Medline: 31290288]
25. Paré G, Tate M, Johnstone D, Kitsiou S. Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. Eur J Inf Syst 2017 Dec 19;25(6):493-508. [doi: 10.1057/S41303-016-0020-3]
26. Templier M, Paré G. Transparency in literature reviews: an assessment of reporting practices across review types and genres in top IS journals. Eur J Inf Syst 2017 Nov 30;27(5):503-550. [doi: 10.1080/0960085x.2017.1398880]
27. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Gusenbauer M, Haddaway NR. What every researcher should know about searching - clarified concepts, search advice, and an agenda to improve finding in academia. Res Synth Methods 2021 Mar 08;12(2):136-147 [FREE Full text] [doi: 10.1002/jrsm.1457] [Medline: 33031639]
29. Hiebl MR. Sample selection in systematic literature reviews of management research. Organ Res Methods 2021 Jan 29;26(2):229-261. [doi: 10.1177/1094428120986851]
30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
31. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

XSL•FO
RenderX

32. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
33. Armaou M, Araviaki E, Dutta S, Konstantinidis S, Blake H. Effectiveness of digital interventions for deficit-oriented and asset-oriented psychological outcomes in the workplace: a systematic review and narrative synthesis. Eur J Investig Health Psychol Educ 2022 Oct 03;12(10):1471-1497 [FREE Full text] [doi: 10.3390/ejihpe12100102] [Medline: 36286087]
34. Bégin C, Berthod J, Martinez LZ, Truchon M. Use of mobile apps and online programs of mindfulness and self-compassion training in workers: a scoping review. J Technol Behav Sci 2022 Sep 06;7(4):477-515 [FREE Full text] [doi: 10.1007/s41347-022-00267-1] [Medline: 36091081]
35. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
36. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
37. Damen I, Brombacher H, Lallemand C, Brankaert R, Brombacher A, van Wesemael P, et al. A scoping review of digital tools to reduce sedentary behavior or increase physical activity in knowledge workers. Int J Environ Res Public Health 2020 Jan 13;17(2):499 [FREE Full text] [doi: 10.3390/ijerph17020499] [Medline: 31941096]
38. Freak-Poli RL, Cumpston M, Albarqouni L, Clemes SA, Peeters A. Workplace pedometer interventions for increasing physical activity. Cochrane Database Syst Rev 2020 Jul 21;7(7):CD009209 [FREE Full text] [doi: 10.1002/14651858.CD009209.pub3] [Medline: 32700325]
39. Jones NM, Johnson M, Sathappan AV, Torous J. Benefits and limitations of implementing mental health apps among the working population. Psychiatr Ann 2021 Feb;51(2):76-83. [doi: 10.3928/00485713-20210112-01]
40. Jung J, Cho I. Promoting physical activity and weight loss with mHealth interventions among workers: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e30682 [FREE Full text] [doi: 10.2196/30682] [Medline: 35060913]
41. Kuster A, Dalsbø TK, Luong Thanh BY, Agarwal A, Durand-Moreau QV, Kirkehei I. Computer-based versus in-person interventions for preventing and reducing stress in workers. Cochrane Database Syst Rev 2017 Aug 30;8(8):CD011899 [FREE Full text] [doi: 10.1002/14651858.CD011899.pub2] [Medline: 28853146]
42. Lee Y, Lee N, Lim H, Sung S. Weight reduction interventions using digital health for employees with obesity: a systematic review. Diabetes Metab Syndr Obes 2022;15:3121-3131 [FREE Full text] [doi: 10.2147/DMSO.S384450] [Medline: 36246516]
43. López-Del-Hoyo Y, Fernández-Martínez S, Pérez-Aranda A, Barceló-Soler A, Bani M, Russo S, et al. Effects of eHealth interventions on stress reduction and mental health promotion in healthcare professionals: a systematic review. J Clin Nurs 2023 Sep;32(17-18):5514-5533 [FREE Full text] [doi: 10.1111/jocn.16634] [Medline: 36703266]
44. Moe-Byrne T, Shepherd J, Merecz-Kot D, Sinokki M, Naumanen P, Hakkaart-van Roijen L, et al. Effectiveness of tailored digital health interventions for mental health at the workplace: a systematic review of randomised controlled trials. PLOS Digit Health 2022 Oct 21;1(10):e0000123 [FREE Full text] [doi: 10.1371/journal.pdig.0000123] [Medline: 36812547]
45. Paganin G, Simbula S. Smartphone-based interventions for employees' well-being promotion: a systematic review. Electron J Appl Stat Anal 2020;13(3):682-712. [doi: 10.1285/i20705948v13n3p682]
46. Park J, Jung S, Ha D, Lee B, Kim M, Sim K, et al. The effectiveness of e-healthcare interventions for mental health of nurses: a PRISMA-compliant systematic review of randomized controlled trials. Medicine (Baltimore) 2022 Jun 24;101(25):e29125 [FREE Full text] [doi: 10.1097/MD.0000000000029125] [Medline: 35758346]
47. Scheutzow J, Attoe C, Harwood J. Acceptability of web-based mental health interventions in the workplace: systematic review. JMIR Ment Health 2022 May 11;9(5):e34655 [FREE Full text] [doi: 10.2196/34655] [Medline: 35544305]
48. Sevic A, Hashemi NS, Thørrisen MM, Strømstad K, Skarpaas LS, Storm M, et al. Effectiveness of eHealth interventions targeting employee health behaviors: systematic review. J Med Internet Res 2023 Apr 20;25:e38307 [FREE Full text] [doi: 10.2196/38307] [Medline: 37079369]
49. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017 Dec 21;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
50. Stratton E, Jones N, Peters S, Torous J, Glozier N. Digital mHealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021 Aug 01;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
51. Sundström C, Forsström D, Berman A, Khadjesari Z, Sundqvist K. What do we know about alcohol internet interventions aimed at employees?-A scoping review. Front Public Health 2023;11:929782 [FREE Full text] [doi: 10.3389/fpubh.2023.929782] [Medline: 36778564]
52. Thai YC, Sim D, McCaffrey TA, Ramadas A, Malini H, Watterson JL. A scoping review of digital workplace wellness interventions in low- and middle-income countries. PLoS One 2023 Feb 28;18(2):e0282118 [FREE Full text] [doi: 10.1371/journal.pone.0282118] [Medline: 36854022]

XSL•FO
RenderX

53. Tung C, Chang C, Jian J, Du Y, Wu C. Studying wearable health technology in the workplace using the Behavior Change Wheel: a systematic literature review and content analysis. Inform Health Soc Care 2022 Oct 02;47(4):434-443. [doi: 10.1080/17538157.2022.2042303] [Medline: 35272548]
54. Velana M, Rinkenauer G. Individual-level interventions for decreasing job-related stress and enhancing coping strategies among nurses: a systematic review. Front Psychol 2021 Jul 19;12:708696 [FREE Full text] [doi: 10.3389/fpsyg.2021.708696] [Medline: 34349711]
55. Vertola G, Marcello A, Bottone M, Sperandeo R, Muzii B, Scandurra C. Use and effectiveness of mobile health applications for stress management and emotional self-regulation in adult workers: a systematic review. In: Proceedings of the 13th IEEE International Conference on Cognitive Infocommunications. 2022 Presented at: CogInfoCom '22; September 21-23, 2022; Budapest, Hungary URL: https://ieeexplore.ieee.org/document/10081617 [doi: 10.1109/coginfocom55841.2022.10081617]
56. Webster N, Oyebode J, Jenkins C, Bicknell S, Smythe A. Using technology to support the emotional and social well-being of nurses: a scoping review. J Adv Nurs 2020 Jan;76(1):109-120. [doi: 10.1111/jan.14232] [Medline: 31599997]
57. Xiong J, Wen JL, Pei GS, Han X, He DQ. Effectiveness of internet-based cognitive behavioural therapy for employees with depression: a systematic review and meta-analysis. Int J Occup Saf Ergon 2023 Mar;29(1):268-281. [doi: 10.1080/10803548.2022.2043647] [Medline: 35172706]
58. Paré G, Trudel MC, Jaana M, Kitsiou S. Synthesizing information systems knowledge: a typology of literature reviews. Inf Manag 2015 Mar;52(2):183-199. [doi: 10.1016/j.im.2014.08.008]
59. World mental health report: transforming mental health for all. World Health Organization. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2024-10-02]
60. Harvey SB, Modini M, Joyce S, Milligan-Saville JS, Tan L, Mykletun A, et al. Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. Occup Environ Med 2017 Mar;74(4):301-310. [doi: 10.1136/oemed-2016-104015] [Medline: 28108676]
61. Cancelliere C, Cassidy JD, Ammendolia C, Côté P. Are workplace health promotion programs effective at improving presenteeism in workers? A systematic review and best evidence synthesis of the literature. BMC Public Health 2011 May 26;11:395 [FREE Full text] [doi: 10.1186/1471-2458-11-395] [Medline: 21615940]
62. Lerner D, Adler DA, Rogers WH, Chang H, Greenhill A, Cymerman E, et al. A randomized clinical trial of a telephone depression intervention to reduce employee presenteeism and absenteeism. Psychiatr Serv 2015 Jun 01;66(6):570-577 [FREE Full text] [doi: 10.1176/appi.ps.201400350] [Medline: 25726984]
63. Physical activity. World Health Organization. 2024 Jun 26. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2024-07-07]
64. McTiernan A, Friedenreich CM, Katzmarzyk PT, Powell KE, Macko R, Buchner D, et al. Physical activity in cancer prevention and survival: a systematic review. Med Sci Sports Exerc 2019 Jun;51(6):1252-1261 [FREE Full text] [doi: 10.1249/MSS.0000000000001937] [Medline: 31095082]
65. Warburton DE, Bredin SS. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
66. Viester L, Verhagen EA, Bongers PM, van der Beek AJ. Effectiveness of a worksite intervention for male construction workers on dietary and physical activity behaviors, body mass index, and health outcomes: results of a randomized controlled trial. Am J Health Promot 2018 Mar 06;32(3):795-805 [FREE Full text] [doi: 10.1177/0890117117694450] [Medline: 28730841]
67. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
68. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
69. Van Nuys K, Globe D, Ng-Mak D, Cheung H, Sullivan J, Goldman D. The association between employee obesity and employer costs: evidence from a panel of U.S. employers. Am J Health Promot 2014 May 01;28(5):277-285. [doi: 10.4278/ajhp.120905-quan-428]
70. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. JAMA 2013 Jan 02;309(1):71-82 [FREE Full text] [doi: 10.1001/jama.2012.113905] [Medline: 23280227]
71. Jans M, van den Heuvel SG, Hildebrandt VH, Bongers PM. Overweight and obesity as predictors of absenteeism in the working population of the Netherlands. J Occup Environ Med 2007 Sep;49(9):975-980. [doi: 10.1097/JOM.0b013e31814b2eb7] [Medline: 17848853]
72. Bertera RL. The effects of behavioral risks on absenteeism and health-care costs in the workplace. J Occup Med 1991 Nov;33(11):1119-1124. [doi: 10.1097/00043764-199111000-00006] [Medline: 1765851]
73. Ng R, Sutradhar R, Yao Z, Wodchis W, Rosella L. Smoking, drinking, diet and physical activity-modifiable lifestyle risk factors and their associations with age to first chronic disease. Int J Epidemiol 2020 Feb 01;49(1):113-130 [FREE Full text] [doi: 10.1093/ije/dyz078] [Medline: 31329872]

XSL•FO
RenderX

74. Robbins R, Jackson CL, Underwood P, Vieira D, Jean-Louis G, Buxton OM. Employee sleep and workplace health promotion: a systematic review. Am J Health Promot 2019 Sep;33(7):1009-1019 [FREE Full text] [doi: 10.1177/0890117119841407] [Medline: 30957509]
75. Savoli A, Barki H, Pare G. Examining how chronically ill patients' reactions to and effective use of information technology can influence how well they self-manage their illness. MIS Q 2020 Jan 01;44(1):351-389. [doi: 10.25300/misq/2020/15103]
76. MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. Health Psychol 2008 Mar;27(2S):S99-100 [FREE Full text] [doi: 10.1037/0278-6133.27.2(Suppl.).S99] [Medline: 18377161]
77. Tams S, Thatcher JB, Craig K. How and why trust matters in post-adoptive usage: the mediating roles of internal and external self-efficacy. J Strateg Inf Syst 2018 Jun;27(2):170-190. [doi: 10.1016/j.jsis.2017.07.004]
78. Salo M, Makkonen M, Hekkala R. The interplay of IT users' coping strategies: uncovering momentary emotional load, routes, and sequences. MIS Q 2020;44(3):1143-1175 [FREE Full text] [doi: 10.25300/misq/2020/15610]
79. Pirkkalainen H, Salo M, Tarafdar M, Makkonen M. Deliberate or instinctive? Proactive and reactive coping for technostress. J Manag Inf Syst 2019 Oct 09;36(4):1179-1212. [doi: 10.1080/07421222.2019.1661092]
80. Baham C, Kalgotra P, Nasirpouri Shadbad F, Sharda R. Generational differences in handling technology interruptions: a qualitative study. Eur J Inf Syst 2022 May 10;32(5):858-878. [doi: 10.1080/0960085X.2022.2070557]
81. Tams S, Grover V, Thatcher J. Modern information technology in an old workforce: toward a strategic research agenda. J Strateg Inf Syst 2014 Dec;23(4):284-304. [doi: 10.1016/j.jsis.2014.10.001]
82. Tams S, Grover V, Thatcher J, Ahuja M. Grappling with modern technology: interruptions mediated by mobile devices impact older workers disproportionately. Inf Syst E-Bus Manage 2021 May 12;20(4):635-655. [doi: 10.1007/S10257-021-00526-3]
83. Tams S, Thatcher JB, Grover V. Concentration, competence, confidence, and capture: an experimental study of age, interruption-based technostress, and task performance. J Assoc Inf Syst 2018;19:857-908. [doi: 10.17705/1jais.00511]
84. Riedl R. On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. Electron Mark 2022;32(1):153-177 [FREE Full text] [doi: 10.1007/s12525-021-00501-3] [Medline: 35600914]
85. Riedl R, Kostoglou K, Wriessnegger SC, Müller-Putz GR. Videoconference fatigue from a neurophysiological perspective: experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). Sci Rep 2023 Oct 26;13(1):18371 [FREE Full text] [doi: 10.1038/s41598-023-45374-y] [Medline: 37884593]
86. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
87. Kayiş AR, Satici SA, Yilmaz MF, Şimşek D, Ceyhan E, Bakioğlu F. Big five-personality trait and internet addiction: a meta-analytic review. Comput Human Behav 2016 Oct;63:35-40. [doi: 10.1016/j.chb.2016.05.012]
88. Pflügner K, Maier C, Mattke J, Weitzel T. Personality profiles that put users at risk of perceiving technostress: a qualitative comparative analysis with the big five personality traits. Bus Inf Syst Eng 2020 Nov 17;63(4):389-402. [doi: 10.1007/S12599-020-00668-7]
89. Srivastava SC, Chandra S, Shirish A. Technostress creators and job outcomes: theorising the moderating influence of personality traits. Inf Syst J 2015 May 15;25(4):355-401. [doi: 10.1111/isj.12067]
90. Penley JA, Tomaka J. Associations among the Big Five, emotional responses, and coping with acute stress. Pers Individ Dif 2002 May;32(7):1215-1228. [doi: 10.1016/S0191-8869(01)00087-3]
91. Latham GP. Goal-setting theory: causal relationships, mediators, and moderators. In: Braddick O, editor. Oxford Research Encyclopedias: Psychology. Oxford, UK: Oxford Academic Press; 2016.
92. Stalujanis E, Neufeld J, Glaus Stalder M, Belardi A, Tegethoff M, Meinlschmidt G. Induction of efficacy expectancies in an ambulatory smartphone-based digital placebo mental health intervention: randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 17;9(2):e20329 [FREE Full text] [doi: 10.2196/20329] [Medline: 33594991]
93. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
94. Hesser H. Identifying causal mechanisms in psychotherapy: what can we learn from causal mediation analysis? Clin Psychol Psychother 2022 May;29(3):1050-1058 [FREE Full text] [doi: 10.1002/cpp.2687] [Medline: 34768315]
95. Cuijpers P, Reijnders M, Huibers MJ. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019 May 07;15(1):207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
96. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
97. Kok G, Gottlieb NH, Peters GJ, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
98. Carey RN, Connell LE, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019 Jul 17;53(8):693-707 [FREE Full text] [doi: 10.1093/abm/kay078] [Medline: 30304386]
99. Spohrer K, Fallon M, Hoehle H, Heinzl A. Designing effective mobile health apps: does combining behavior change techniques really create synergies? J Manag Inf Syst 2021 Aug 06;38(2):517-545. [doi: 10.1080/07421222.2021.1912936]

XSL•FO
RenderX

100. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001;86(3):499-512. [doi: 10.1037//0021-9010.86.3.499]
101. Kozlowski SW, Klein KJ. A multilevel approach to theory and research in organizations: contextual, temporal, and emergent processes. In: Klein KJ, Kozlowski SW, editors. Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions. Hoboken, NJ: John Wiley & Sons; 2000:3-90.
102. Chapman LS. Meta-evaluation of worksite health promotion economic return studies: 2012 update. Am J Health Promot 2012 Mar 01;26(4):TAHP1-TAH12. [doi: 10.4278/ajhp.26.4.tahp] [Medline: 22375583]
103. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]

Abbreviations

DWP: digital wellness program
HR: human resource
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question

Edited by T de Azevedo Cardoso; submitted 07.01.25; peer-reviewed by M Armaou, A Sevic; comments to author 28.01.25; revised version received 11.02.25; accepted 24.02.25; published 14.03.25.

Please cite as:
Amirabdolahian S, Pare G, Tams S
Digital Wellness Programs in the Workplace: Meta-Review
J Med Internet Res 2025;27:e70982
URL: https://www.jmir.org/2025/1/e70982
doi:10.2196/70982
PMID:

©Saeed Amirabdolahian, Guy Pare, Stefan Tams. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Physical Activity and Sedentary Behavior Interventions for Community-Living Adults: Umbrella Review

Eilidh Russell[1], BSc, MRES; Alison Kirk[2], BSc, PhD; Mark D Dunlop[1], BSc, PhD; William Hodgson[2], BSc, MRES; Mhairi Patience[2], BSc, MSc; Kieren Egan[1], BSc, PhD

[1]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, United Kingdom

[2]Department of Physical Activity and Health, University of Strathclyde, Glasgow, United Kingdom

Corresponding Author:
Eilidh Russell, BSc, MRES
Department of Computer and Information Sciences
University of Strathclyde
26 Richmond Street
Glasgow, G1 1XH
United Kingdom
Phone: 44 01415524400
Fax: 44 01415524400
Email: eilidh.russell.2017@uni.strath.ac.uk

Abstract

Background: Digital interventions hold significant potential for improving physical activity (PA) and reducing sedentary behavior (SB) in adults. Despite increasing interest, there remain surprising gaps in the current knowledge of how best to deliver these interventions, including incorporating appropriate theoretical frameworks and behavior change techniques. Following numerous systematic reviews, there is now significant potential for umbrella reviews to provide an overview of the current evidence.

Objective: This umbrella review aimed to explore digital PA and SB interventions for community-living adults across effectiveness, key components, and methodological quality.

Methods: This review followed the Joanna Briggs Institute framework for umbrella reviews. Key search terms were developed iteratively, incorporating physical and sedentary activity alongside digital interventions. We searched 7 online databases (Web of Science Core Collection, CINAHL, APA PsycINFO, Inspec, the Cochrane Library, MEDLINE [Ovid], and PROSPERO) alongside gray literature databases. Information was extracted and tabulated from each included article on intervention effectiveness, key components, and content acknowledging both the digital and human elements. The study quality was appraised using A Measurement Tool to Assess systematic Reviews 2 (AMSTAR 2). The corrected covered area method was used to assess the overlap of primary studies included in the systematic reviews. All relevant research findings were extracted and reported.

Results: Search terms identified 330 articles, of which 5 (1.5%) met the inclusion criteria. The most common PA outcomes identified were daily steps, moderate-to-vigorous PA, total PA, and PA change. Reviews with meta-analysis reported that digital interventions improved multiple PA outcomes (daily steps, moderate-to-vigorous PA time, and total PA time). However, findings from the remaining systematic reviews were mixed. Similarly, the findings for SB were contrasting. Regarding intervention components, monitor- and sensor-only intervention delivery methods were most frequently implemented. Eleven theoretical frameworks were identified, with social cognitive theory being the most prominent theory. In total, 28 different behavior change techniques were reported, with goal setting, self-monitoring, feedback, and social support being the most frequently used. All 5 systematic reviews were of low or critically low quality, each incorporating unique primary studies (corrected covered area=0%).

Conclusions: This umbrella review highlights the potential of digital interventions to increase PA and reduce SB among community-living adults. However, the disparate nature of current academic knowledge means potentially efficacious research may not realistically translate to real work impact. Our review identified a lack of consensus around outcomes and components at both individual (eg, difficult to collate and compare findings) and multiple study (poor reported quality of systematic reviews) levels. Collective, concerted action is required to standardize outcomes and improve systematic review reporting to optimize future learning around digital interventions to increase PA and reduce SB in community-living adults, including traditionally overlooked populations, like informal carers.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42023450773; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023450773

(J Med Internet Res 2025;27:e66294) doi:10.2196/66294

KEYWORDS

physical activity; sedentary behavior; digital health interventions; behavior change; theoretical frameworks; umbrella review; mobile phone; community-living adults

Introduction

Background

Physical activity (PA) is defined as "any body movement generated by the contraction of the skeletal muscles that raises energy expenditure above resting metabolic rate and is characterized by its modality, frequency, intensity, duration, and context of practice" [1]. Engaging in regular PA can reduce the risk of cardiovascular disease, type 2 diabetes, cancer, weight gain and obesity, and Alzheimer disease and significantly improve mental health and quality of life [2-4]. Alongside these individual health benefits, PA also provides benefits for societies, environments, and economies [5]. Sedentary behavior (SB) is described as "any waking behaviors characterized by an energy expenditure ≤1.5 metabolic equivalents while in a sitting, reclining, or lying position" [1]. Prolonged SB has been found to increase all-cause mortality and is also associated with poor disease outcomes, such as increased risk of developing cardiovascular disease, diabetes mellitus, hypertension, and certain types of cancer [6-8].

Despite the well-documented evidence for improving PA levels and reducing SB, many adults do not engage in sufficient levels of activity and have high levels of SB. Data from the Lancet assessing global trends in insufficient PA among adults from 2000 to 2022 using population-based surveys found that nearly one-third (31.3%) of the adults worldwide were not meeting the World Health Organization recommendations for moderate-to-vigorous aerobic activity in 2022 [9]. Findings from this work also suggest that if trends from 2010 to 2022 continue, the global prevalence of insufficient PA will reach 34.7% by 2030. Therefore, there is a need for effective interventions to increase PA levels and reduce SB.

Digital health interventions (DHIs) can be used to promote PA and reduce SB. World Health Organization describes digital health as the use of digital, mobile, and wireless technologies to aid the achievement of health-related objectives and includes both mobile health and eHealth [10]. DHIs can include smartphone apps, wearable activity trackers, online platforms, and web-based interventions. A favorable feature of DHIs is their ability to be scaled for large populations while also being able to be tailored for specific populations [10]. The use of DHIs has also been recommended by the global action plan on PA 2018 to 2030 to promote and support participation in PA due to their success and benefits of use [11].

As PA is described to be a complex behavior [12,13], it requires more time and commitment to change habits than other health behaviors. It has been recommended that PA and SB interventions be informed by theoretical frameworks and include appropriate behavior change techniques (BCTs) as they can be important for increasing effectiveness and adherence toward the behaviors [14,15]. Despite such previous work, there is little consensus as to what the best theoretical frameworks and BCTs for changing PA and SB are. For example, research suggests that combined PA and dietary interventions based on specific theories, such as cognitive theory and theory of planned behavior, are no more effective in achieving changes to PA outcomes than interventions with no stated theoretical grounding [16]. However, meta-regression of this work and an additional study [17] suggest that the use of a cluster of "self-regulatory" intervention techniques had a positive relationship with increasing PA, alongside additional outcomes, such as weight loss and dietary changes. Similarly, the available evidence to support the use of BCTs to improve PA is contrasting, with positive changes to PA behavior reported, but not all findings were statistically significant. Therefore, digital interventions with a theoretical grounding and BCTs could encourage improvements in PA and SB; however, more investigation into these aspects of intervention is required.

Moreover, many groups in society face significant additional barriers to PA either in an acute or chronic perspective, including individuals who face mental and physical health conditions, those facing higher levels of economic deprivation, and individuals in later years of life. An emerging and sizable group who are facing systemic barriers to PA and reducing SB includes informal carers. Informal carers could particularly benefit from digital PA and SB interventions as Horne et al [18] found that carers often prioritize the needs of those they care for over their own health, which could result in lower levels of PA and higher levels of SB. Caring duties can also act as a barrier for where, when, and how long they can participate in PA programs. Allowing this group a digital intervention that considers their own unique circumstances, preferences, and barriers could significantly improve adherence to higher PA levels and reduce SB. There is little published evidence to understand the use of these digital interventions for improving these behaviors within this population. In contrast, there are numerous primary studies and systematic reviews exploring the use of digital interventions to improve PA and SB levels in adults. Although having numerous systematic reviews on a certain topic can be beneficial, it has been reported that too many findings on a specific topic or research area can lead to replication issues by amplifying biases, inconsistencies, and misinterpretations from the available literature [19]. To eliminate the issues caused by multiple systematic reviews, umbrella reviews have been recommended.

Umbrella reviews are used to summarize the evidence from multiple research syntheses. This type of review provides an overview of the available evidence on a chosen topic and can be used to compare or contrast findings from published

XSL•FO
RenderX

systematic reviews [20]. An umbrella review is particularly beneficial for decision makers as it provides "user-friendly" summaries of the available knowledge and literature without the need to read and understand the findings of multiple systematic reviews themselves [21]. Umbrella reviews evaluating digital interventions currently do exist [22-24]. However, only a few specifically focus on groups within society who face additional barriers to participating in PA.

This Study

The overall scope of this research is to support PA and reduce SB in informal carers via digital devices. However, the lack of published primary studies and systematic reviews exploring digital PA and SB interventions for informal carers means that this paper will focus on digital PA and SB interventions for the community-living adult population, as there are numerous published systematic reviews in this topic area. We will address, in the Discussion section, the potential implications for the caring population.

The research questions for this umbrella review were as follows:

1. How effective are digital interventions in changing PA and SBs in community-living adults?
2. What components are included in digital PA and SB interventions, including delivery methods (technologies and human elements used) and content (theoretical frameworks and behavior change strategies used)?
3. What is the methodological quality of the included systematic reviews?

Methods

The methodology for this umbrella review was guided by the Joanna Briggs Institute Manual for Evidence Synthesis [25]. A protocol for this review was submitted to PROSPERO (CRD42023450773).

Inclusion Criteria

The inclusion criteria were generated using the population, intervention, context, outcomes, and study (PICOS) method. The inclusion and exclusion criteria for this umbrella review are shown in Textboxes 1 and 2.

Textbox 1. Inclusion criteria.

Population

- Community-living adults aged ≥18 years.

Interventions

- Digital interventions that were designed to increase physical activity (PA) levels and reduce sedentary behavior (SB) in adults who are physically capable of living at home.
- Interventions could be multimodal in nature (eg, they could include a digital and in-person element) if the digital element was the main component of the intervention.

Context and outcomes

- The main outcomes of this review were the effectiveness of interventions in changing PA and SB.
- The components and content included in digital PA and SB interventions, including the technological and human components, behavior change strategies, and theoretical frameworks, were also the main outcomes of this review.
- The overlap of included primary studies and the quality of the systematic reviews were secondary outcomes of this umbrella review.

Study type

- Systematic review.

Publication date

- Published in or after 2013.

Language

- English.

XSL•FO
RenderX

Textbox 2. Exclusion criteria.

Population

- Children or young adults aged ≤18 years.

Adults who were not capable of living independently or not community living.

Interventions

- Digital interventions that were not designed to increase PA levels and reduce SB.

Interventions where the digital element was not the main component of the intervention.

Digital interventions targeting multiple behaviors (eg, diet and PA).

Context and outcomes

- Systematic reviews of digital PA and SB interventions that do not report on the included outcomes (eg, systematic reviews assessing the effectiveness of PA and SB interventions on reducing blood glucose or mental health outcomes).

Study type

- Primary studies or any other type of review.

Publication date

- Published before 2013.

Language

- Other than English.

Search Strategy

Searches were conducted in July 2023 in the following 7 chosen databases: Web of Science Core Collection, CINAHL, APA PsycINFO, Inspec, Cochrane Library, MEDLINE (Ovid), and PROSPERO. These databases were used as they contain papers from health, social care, psychology, and computing and therefore encompass the topic area of this umbrella review. The gray literature databases, such as arXiv, Core, OpenGrey, and OpenDissertations were, also searched alongside the reference list of included systematic reviews. The rationale for focusing on systematic reviews from 2013 onward allowed the research team to maximize the availability of knowledge around digital interventions for improving PA and SB. The publication period up to 2013 marked the period when digital tools became more commercial, sophisticated, and widespread. Before 2013, digital interventions often relied on less accessible or less integrated technology. However, from 2013 the advancement of sensors and integration within digital interventions became more mainstream and transformed the landscape of these interventions. The inclusion criteria allowed this review to capture technologies most relevant to the present day and the future. In addition, by including systematic reviews from the past 10 or more years, primary studies 10 years before 2013 were able to be captured.

With guidance from subject librarians, a comprehensive keyword search strategy was created. These keywords in relation to adults, digital, PA, SB, theoretical models, BCTs, and systematic reviews were all initially searched individually to help refine the search criteria and formulate a more concise and comprehensive search strategy. The final search combined all keywords to identify systematic reviews on digital PA and SB interventions for community-living adults that incorporated elements of BCTs and theoretical frameworks (Multimedia Appendix 1).

Study Selection

Results from the database and gray literature searches (excluding Core) were imported into the EndNote Online reference manager (Clarivate). Duplicates were then removed in EndNote. Core search results were unable to be exported into this system; therefore, screening of these articles remained in the CORE database.

After this removal process, the remaining articles were exported from EndNote to the Rayyan (Rayyan Systems, Inc) website [26] for further screening. Title and abstract screening was conducted in Rayyan independently by the main reviewer (ER) and a second reviewer (WH). Title and abstract screening was blinded to reviewers until the process had been completed to lessen the potential for bias to occur. Full-text screening was also conducted in Rayyan with a supplementary Microsoft Excel document used by the main reviewer for tracking progress, dividing the workload between additional reviewers, and for additional comments and reasoning for inclusion or exclusion. Full-text screening was conducted independently by the main reviewer alongside the second reviewer (WH) and an additional reviewer (MP). The full-text screening process was also blinded. Conflicts were resolved through discussion between the reviewers.

Corrected Covered Area

The overlap of the primary studies included in the chosen systematic reviews was also assessed. As systematic reviews with similar research questions may include some of the same studies within their reviews, it is important to measure the

XSL•FO
RenderX

overlap of these studies as it can disguise a lack of current research or give a false impression of new evidence.

The corrected covered area (CCA) was calculated by creating a citation matrix where primary publications were listed in rows and the included systematic reviews of the umbrella review were represented in columns. The number of times a given study was cited in the systematic reviews was then calculated using the following formula [27]:

$$CCA = N - r / rc - r$$

where N is the total number of times primary publications appeared in reviews (inclusive of double counting), r is the number of unique primary publications, and c is the number of systematic reviews included in the umbrella review. The degree of overlap was then classified, where 0% to 5% was considered "slight overlap," 6% to 10% was considered "moderate overlap," 11% to 15% was considered "high overlap," and >15% was considered "very high overlap" [27].

Data Extraction

In order to answer research questions 1 and 2 of this umbrella review, a Microsoft Excel spreadsheet for data extraction was created using the Joanna Briggs Institute Manual for Evidence Synthesis data extraction guidelines. Extracted information included article and publication information (author, date, and digital object identifier), study information (number of included studies, date range, study design, location, and inclusion of meta-analysis), population demographics (sample size, age, and sex), intervention details, measures of effectiveness, and methodological quality rating. Data extraction was also quality checked for agreement between members of the review team.

Information on changes in PA and SB was extracted to answer the first research question. The most common PA outcomes (daily step count, moderate-to-vigorous PA [MVPA] min, total daily or weekly PA time, and PA change) and SB outcomes (SB min or change in SB) were used to report the effectiveness of these behaviors. To report on the components included within digital PA and SB interventions, the extracted information consisted of the following: whether the intervention focused on PA or SB only or if it targeted both behaviors, whether the intervention was shaped by a theoretical framework and what the frameworks were, and if the intervention included BCTs and what they were. Data extracted from the systematic reviews also included information on the delivery of the intervention (eg, smartphone, web based, wearable activity tracker–only interventions, or mixed delivery methods), human and technological components included, and the duration of the interventions.

Methodological Quality

The quality of the included systematic reviews was assessed using A Measurement Tool to Assess Systematic Reviews-2 (AMSTAR-2). This tool comprises 16 questions that are used to assess the confidence of the results of the review. The AMSTAR-2 rating of a review can range from high (when a review paper has <1 critical weakness) to moderate (with >1 noncritical weakness), low (with 1 critical weakness), and critically low (with >1 critical weakness) [28]. It is recommended that the ratings of the individual items should not be combined to create an overall rating, instead users should consider the impact of each individual item to determine an overall confidence rating.

All 16 questions within the AMSTAR-2 tool are of importance; however, there are 7 critical domains that can affect the validity of a review and the conclusions that can be drawn [28]. These critical domains relate to the review methodology, literature search strategy, justification for excluded studies, techniques for assessing the risk of bias, methodology for the statistical combination of results if meta-analysis was used, accounting for the risk of bias when discussing the results, and investigating publication bias and discussing the impact on results of the review, if quantitative synthesis was performed.

The methodological quality of each included systematic review was assessed by the main reviewer. The second reviewer assessed the methodological quality of 1 included review individually to cross-check the quality assessment process. Any disagreements or conflicts were resolved through discussion.

Presentation of Findings

The results of this umbrella review provided a summary of systematic reviews assessing digital PA and SB interventions for community-living adults. These results were reported as a summary of the synthesized results. Findings were summarized to identify if interventions have been beneficial or effective, if no differences between interventions and control groups have been identified or mixed results, or to show if interventions were less effective than the comparator or not effective. These summaries can be used to easily understand if a statistically significant change in PA or SB was achieved.

A quantitative summary of the included systematic reviews in relation to their delivery methods and their content (inclusion of theoretical frameworks and BCTs) was also created. The technological delivery methods were split into 6 columns consisting of web-only delivery, smartphone- or tablet-only delivery, video game–only delivery, virtual reality–only delivery, monitor- or sensor-only delivery (including wearable activity monitors, pedometers, accelerometers, and heart rate sensors), and a mixed technological delivery column. The tally of these technological delivery components matched the number in the included studies column of the table. If a systematic review mentioned the use of theoretical frameworks and BCTs, they were marked "yes" with the number of mentioned frameworks and BCTs included in brackets.

Results

Study Inclusion

Academic database and gray literature searches identified 361 and 73 articles for screening, respectively (refer to Multimedia Appendix 1 for full search strategy and outputs). These 434 articles were exported into Endnote Online for initial screening, where 104 (24%) duplicates were removed. After this process, the remaining articles (n=330, 76%) were exported to the screening website Rayyan for title and abstract screening.

Title and abstract screening removed 263 (79.7%) of the 330 articles. The reasons for removal are displayed in Figure 1. With

XSL•FO
RenderX

an agreement of 84%, the interrater reliability of title and abstract screening indicated substantial agreement between the main reviewer and second reviewer (WH; Cohen κ=0.69; *P*<.001). Conflicts (41/330, 12.4%) were resolved through discussions and clarification of the protocol.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of the selection process. PA: physical activity; SB: sedentary behavior.

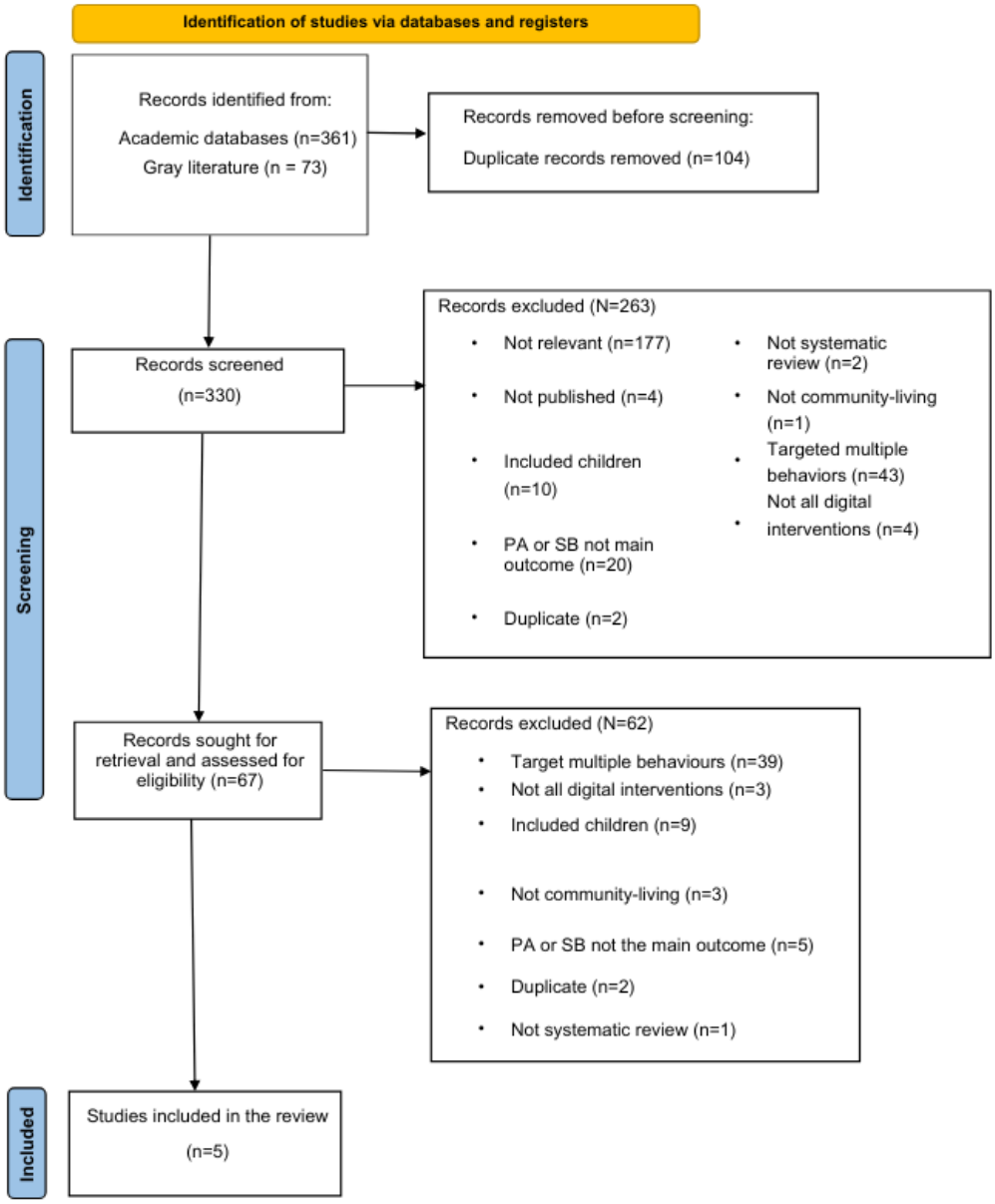

After title and abstract screening, 67 articles were eligible for full-text screening. Screening of these articles was conducted in Rayyan with an additional Microsoft Excel spreadsheet used to track decisions. A total of 62 (93%) articles were removed during the screening process (Figure 1). There was 69% agreement between the main reviewer and additional reviewers for full-text screening, corresponding to a fair agreement (Cohen κ=0.25; *P*=.02). All conflicts (21/67, 31%) were resolved through discussions. The final full-text screening resulted in 5 systematic reviews to be included within this umbrella review [29-33].

CCA Scores

The measured overlap of primary studies included in the eligible systematic reviews was measured to be 0% (Multimedia Appendix 2 [29-33]). This indicates that all included systematic reviews consisted of different primary studies.

Characteristics of the Included Studies

A summary of the characteristics of the included systematic reviews is displayed in Table 1. The total number of primary studies included in the 5 systematic reviews represented 96 original research papers conducted across 24 different countries. The included systematic reviews' publication dates ranged from 2018 [30] to 2023 [31,33]. The research designs varied within each systematic review and included randomized controlled trials, developmental studies, pre- and postintervention studies, and implementation studies. In total, 2 (40%) of the 5 systematic reviews included meta-analysis [31,33].

XSL•FO
RenderX

Table 1. Characteristics of the included systematic reviews.

Study	Countries of the studies	Papers reviewed and publication date range	Population	Sample size	The type of study included
Lee et al [29], 2023	Australia, Iran, Ireland, Japan, Spain, and the United States	18 (2010-2020)	Adults with obesity or adults who were overweight	Range 26-1755	RCT[a] and pre-post study without a control group
Berry et al [30], 2018	Australia, Canada, the Netherlands, the United Kingdom, and the United States	9 (2006-2017)	Individuals with "arthritis" and those with chronic conditions	Range 20-968	RCT, pre-post test, 2-group pre-post test, and implementation study
Wang and Kassavou [31], 2023	Australia, Canada, China, Germany, Israel, Japan, Korea, the Republic of Slovenia, Spain, Taiwan, Thailand, Turkey, the United Kingdom, and the United States	16 (2008-2021)	Adults with a history or explicit diagnosis of stroke	799	RCT
Daryabeygi-Khotbehsara et al [32], 2021	Italy, Netherlands, New Zealand, and the United States	8 (2015-2020)	Insufficiently active and sedentary healthy adults, healthy and highly educated young adults, overweight and sedentary adults, adults with chronic low back pain, and students and staff from a university setting	Range 10-104	Pre-post intervention; RCT; 3-arm quasi-experimental, single-group micro randomized trial; and developmental study
Wu et al [33], 2023	Australia, Belgium, Canada, Chile, China, Denmark, Finland, Germany, Japan, Netherlands, Portugal, Spain, the United Kingdom, and the United States	45 (2003-2022)	Individuals living with chronic diseases and other diseases	Range 15-1023	RCT

[a]RCT: randomized controlled trial.

Participant Characteristics

The sample sizes within the systematic reviews ranged from a single-group microintervention study consisting of 10 participants to a randomized controlled trial with 7144 participants. A total of 4 (80%) of the 5 systematic reviews described participants' ages, which ranged from 19 to 89 years. Most systematic reviews (4/5, 80%) reported mixed sexes of participants, but there was a greater proportion of female participants in 2 (40%) of the systematic reviews.

Intervention Characteristics

Overview

Of the 5 included systematic reviews, 2 (40%) reviews included interventions that only focused on improving PA [29,30], while the remaining systematic reviews included PA and SB interventions. The duration of interventions ranged from 2 weeks to 12 months.

Research Question 1: Effectiveness of Interventions

All (5/5, 100%) systematic reviews reported on changes in PA where there was a mixture of approaches, and 2 (40%) systematic reviews reported statistical significance of results. Both systematic reviews that included meta-analysis reported statistically significant changes in PA. Wu et al [33] reported significant increases in daily steps from 38 analyzed studies (standardized mean difference [SMD]=0.59; 95% CI 0.44-0.75; P<.001), total daily PA from 4 studies analyzed (SMD=0.21; 95% CI 0.01-0.40; P=.04), and time spent in MVPA from 18 analyzed studies (SMD=0.54; 95% CI 0.36-0.72; P<.001). Wang and Kassavou [31] found significant improvements from 10 studies analyzed, which included 326 participants, to support the use of digital interventions for increasing PA (SMD=0.39; 95% CI 0.17-0.61; P<.001; Table 2).

XSL•FO
RenderX

Table 2. Effectiveness of interventions.

Study	Meta-analysis	PA[a] or SB[b] only	PA or SB outcome assessment	Common PA outcomes				Common SB outcomes
				Steps[c]	MVPA[d]	Total PA[e]	PA change[f]	SB minutes or change
Lee et al [29], 2023	No	PA only	Self-report and device based	—[g]	—	—	Mixed results	N/A[h]
Berry et al [30], 2018	No	PA only	Self-report and device based	Mixed results	Mixed results	Mixed results	—	N/A
Wang and Kassavou [31], 2023	Yes	PA and SB	Self-report and objective measurement	—	—	—	Statistically significant improvement	Statistically nonsignificant change
Daryabeygi-Khotbehsara et al [32], 2021	No	PA and SB	Unclear	Mixed results	Statistically nonsignificant change	Statistically nonsignificant change	—	Statistically nonsignificant change
Wu et al [33], 2023	Yes	PA and SB	Objective	Statistically significant improvement	Statistically significant improvement	Statistically significant improvement	—	Statistically significant improvement

[a]PA: physical activity.

[b]SB: sedentary behavior.

[c]Steps included accelerometer, mean number of steps per day, step counts, average daily steps, daily walking time, daily step count, or daily steps.

[d]Moderate-to-vigorous PA (MVPA) outcomes included moderate PA minutes per week, MVPA time, average daily minutes MVPA, or MVPA minutes.

[e]Total PA outcomes included total PA minutes per week, total PA time, or total daily PA.

[f]PA change outcomes included PA or PA behavior.

[g]Not available.

[h]N/A: not applicable.

For 3 (60%) of the 5 systematic reviews that did not include meta-analysis, the effectiveness of digital interventions on PA was mixed (Table 2). Lee et al [29] reported 14 of the 18 interventions had a positive effect on PA; however, 4 interventions were reported to have "no effect" on PA. Similarly, Berry et al [30] found 6 studies reported significant increases in PA, but 3 studies reported nonsignificant improvements. Finally, Daryabeygi-Khotbehsar et al [32] reported significant changes in light-intensity PA, average daily steps, mean daily step time, and daily walking time. However, this systematic review also reported nonsignificant changes in daily minutes of MVPA and daily step counts.

Of the 5 included reviews, findings for the effectiveness of digital interventions to improve SB were also mixed across the 3 (60%) reviews reporting this behavior. The meta-analysis conducted by Wu et al [33] included 11 studies analyzing the effect of digital interventions on improving SB and found a small but favorable effect size for improving sedentary time by reducing time spent sedentary using digital interventions (SMD=–0.10; 95% CI –0.19 to –0.01; *P*=.03). However, 1 (20%) systematic review with meta-analysis, analyzing 11 studies, and 1 (20%) systematic review without meta-analysis, reported Statistically nonsignificant changes in SB (SMD=–0.13, 95% CI –0.31 to 0.05, N=473, *P*=.53 [31] and adjusted mean difference –9.5 min, SE 7.5; 95% CI 19.98-1.05; *P*=.08 [32]).

In answer to the first research question, digital PA interventions have the potential to improve PA outcomes. Of the 5 included systematic reviews used in this umbrella review, 2 (40%) undertook meta-analysis, which found statistically significant improvements in daily steps, minutes spent in MVPA, total PA time, and PA behavior. However, these findings were not universally true across the included systematic reviews of this umbrella review as the remaining included reviews, which did not include meta-analysis, reported mixed or statistically nonsignificant results across these outcomes. Findings from this umbrella review highlighted that the effectiveness of digital SB interventions was also contrasting. In addition, 1 (20%) systematic review with meta-analysis reported statistically significant improvements in SB through reduced sedentary time (min), but the other systematic review with meta-analysis and one other systematic review found statistically nonsignificant improvements in SB.

XSL•FO
RenderX

Research Question 2: Intervention Components and Content

Intervention Delivery

Both systematic reviews that included PA-only interventions used a range of technological delivery methods, including web-only methods (such as virtual avatars), smartphone- or tablet-only methods (including smartphone apps), monitor- or sensor-based methods (such as commercial grade accelerometers), and a mixed technological delivery method of a smartphone app linked to a wearable activity tracker (Table 3). The human components used in these interventions included feedback provided by trained moderators, private messaging, and video and audio communication between participants and moderators.

Table 3. Quantitative summary of components and content of included reviews.

Study	Included studies, (n=96), n (100%)	Web only, (n=15), n (16%)	Smartphone or tablet only, (n=10), n (11%)	Video game only, (n=5), n (5%)	Virtual reality only, (n=4), n (4%)	Monitor or sensor only, (n=36), n (39%)	Mixed technological intervention, (n=22), n (24%)	Theoretical framework (number identified)	Behavior change technique (number identified)
Lee et al [29], 2023	18 (19)	11 (61)	3 (17)	0 (0)	0 (0)	4 (22)	0 (0)	Yes (5)	Yes (11)
Berry et al [30], 2018[a]	9 (9)	4 (80)	—[b]	0 (0)	0 (0)	—	1 (20)	Yes (1)	Yes (15)
Wang and Kassavou [31], 2023	16 (17)	0 (0)	3 (19)	5 (31)	4 (25)	3 (19)	1 (6)	No (N/A[c])	No (N/A)
Daryabeygi-Khotbehsara et al [32], 2021	8 (8)	—	4 (50)	0 (0)	0 (0)	—	4 (50)	Yes (8)	Yes (7)
Wu et al [33], 2023	45 (47)	—	—	0 (0)	0 (0)	29 (64)	16 (36)	No (N/A)	Yes (5)

[a]Berry et al [30] included 9 studies evaluating 5 different interventions; therefore, the tally of intervention delivery components will equal 5.

[b]Not available.

[c]N/A: not applicable.

For the 3 (60%) of the 5 systematic reviews that targeted both PA and SB, the technological components consisted of a variety of smartphone- and tablet-only interventions, video game–based interventions, virtual reality interventions, monitor- and sensor-based interventions, and a range of mixed technological delivery methods. Wang and Kassavou [31] reported 1 mixed intervention of a mobile phone and tablet intervention with video game components. Daryabeygi-Khotbehsara et al [32] reported 4 mixed interventions, 2 of which were smartphone apps with PA monitors or sensors (heart rate wristbands and Fitbit One [Google LLC]), 1 was a smartphone app with Fitbit and web-based mobile questionnaire, and 1 was a mixed intervention that did not mention the operating system but consisted of text messages and activPAL 3 (PAL Technologies Ltd) activity tracker. Finally, Wu et al [33] reported 16 mixed interventions, 4 of which were web based and monitor or sensor interventions and 12 which were smartphone and tablet and monitor or sensor mixed interventions (Table 3). The human components used in these interventions included manual logging of activity, PA consultations, nurse consultations, group education, training courses, and activity books. Only 1 (20%) systematic review [31] did not report on the human components used within the interventions.

Theoretical Frameworks

Of the 5 included systematic reviews, 2 (40%) reviews that included only PA interventions [29,30] and 1 (20%) systematic review that included PA and SB interventions [32] reported the use of theoretical frameworks to guide the interventions (Table 3). However, 1 (20%) of these systematic reviews [30] only reported on 1 framework for guiding the interventions and did not explicitly report on the use of multiple theoretical frameworks or gave details on other frameworks used.

Of the 5 systematic reviews, 3 (60%) that reported the use of theoretical models within interventions reported that social cognitive theory was the most prominent theory and guided most interventions. The frameworks reported by Lee et al [29] and Daryabeygi-Khotbehsara et al [32], included self-regulation theory (n=2), computational agent model (n=1), control theory (n=1), exploit-explore strategy (n=1), Fogg behavior model (n=1), health action process approach (n=1), health belief model (n=1), learning theory (n=1), self-efficacy theory (n=1), and transtheoretical model (n=1).

Behavior Change Techniques

A total of 4 (80%) of the 5 included systematic reviews [29,30,32,33] reported on the use of BCTs to support interventions (Table 3), and 3 (75%) of the 4 systematic reviews [29,30,32] reported the most frequently used BCTs to be goal setting, self-monitoring, feedback, and social support. Aside from the most frequently used BCTs, the additional BCTs identified from the 4 systematic reviews include information on health consequences (2/28, 7%), action planning, body changes (1/28, 4%), the Coventry, Aberdeen, and London–Refined taxonomy of BCTs (1/28, 4%), culturally

XSL•FO
RenderX

relevant information (1/28, 4%), discussion about incompatible beliefs (1/28, 4%), distraction (1/28, 4%), emotional support (1/28, 4%), framing and reframing (1/28, 4%), habit formation (1/28, 4%), information on how to perform PA (1/28, 4%), past activity performances (1/28, 4%), PA advice (1/28, 4%), problem-solving (1/28, 4%), prompts and cues (1/28, 4%), review of goals (1/28, 4%), reviewing behavior (1/28, 4%), reward (1/28, 4%), self-efficacy (1/28, 4%), self-identity (1/28, 4%), self-talk (1/28, 4%), social modeling (1/28, 4%), theory-based strategies (1/28, 4%), and watching videos (1/28, 4%).

Only 1 (20%) of the 5 systematic reviews did not report on the inclusion of BCTs used within PA and SB interventions [31]. However, it did report on the effectiveness of self-monitoring and behavior change. For changes in both PA and SB, interventions underpinned by self-monitoring elements were more likely to explain effectiveness. In relation to PA, self-monitoring interventions had a moderate to large impact on improving behavior (SMD=0.56; 95% CI 0.21-0.91; *P*=.002). Similarly, SB self-monitoring interventions significantly explained the overall effectiveness of reducing sedentary time compared to those not including such components (SMD=–0.34; 95% CI –0.65 to –0.03; *P*=.03).

In answer to the second research question, of the 5 included systematic reviews used in this umbrella review, 4 (80%) reported the use of human elements within interventions [29,30,32,33], all 5 (100%) reported the digital elements used to deliver the interventions, 3 (60%) reported the use of theoretical frameworks to guide interventions, and 4 (80%) reported the use of BCTs to support the interventions. Monitor- and sensor-only intervention delivery methods were used most frequently followed by mixed technological delivery methods. In total, 11 different theoretical frameworks were identified from the systematic reviews, with social cognitive theory reportedly being the most prominent theory used to guide interventions. In addition, 28 different BCTs were reported in the systematic reviews, with goal setting, self-monitoring, feedback, and social support being reported as the most frequently used BCTs. Moreover, 1 (20%) systematic review also reported that self-monitoring was more likely to explain improvements in both PA and SB compared to interventions that did not include this BCT.

Methodological Quality

The methodological quality of all included systematic reviews was assessed by the main reviewer, and the quality assessment process of the main reviewer was checked by the second reviewer. This quality assessment of the process resulted in moderate agreement (69%; Cohen κ=0.43; *P*=.04).

The use of the AMSTAR-2 tool resulted in 20% (1/5) low-quality systematic reviews [32] and 80% (4/5) critically low–quality systematic reviews [29-31,33] (Table 4). The common methodological strengths of the included reviews (common in ≥3 systematic reviews) were the inclusion of Population, Intervention, Comparison, and Outcome components, reporting that the review methods were established before the conduct of the review, the explanation of study design, descriptions of the included studies, using a satisfactory technique for assessing the risk of bias in individual studies, accounting for the risk of bias when interpreting results, and reporting any potential conflicts of interest. The common areas of weakness (present in ≥3 systematic reviews) in noncritical domains were not performing study selection in duplicate, not performing data extraction in duplicate, not reporting the sources of funding for the primary studies included in the review, and not providing a satisfactory explanation and discussion of any heterogeneity observed in the results of the review. The common areas of weakness in critical domains were not using a comprehensive literature search strategy and not providing a list of excluded studies and the justification for the exclusions.

Table 4. Methodological quality of included systematic reviews.

Study	Noncritical domain strengths, n	Critical domain strengths, n	Noncritical domain weaknesses, n	Critical domain weaknesses, n	AMSTAR-2[a] rating
Lee et al [29], 2022	6	3	2	2	Critically low
Berry et al [30], 2018	4	1	4	4	Critically low
Wang and Kassavou [31], 2023	5	3	4	4	Critically low
Daryabeygi-Khotbehsara et al [32], 2021	3	4	5	1	Low
Wu et al [33], 2023	4	5	5	2	Critically low

[a]AMSTAR: A Measurement Tool to Assess Systematic Reviews-2.

Discussion

Principal Findings

This umbrella review synthesized evidence from 5 systematic reviews exploring digital PA and SB interventions for community-living adults. Social cognitive theory, goal setting, self-monitoring, feedback, and social support were found to be the most prominent theoretical frameworks and BCTs used in digital interventions. Overall, the systematic reviews reported contrasting findings for digital interventions, with some reviews reporting statistically significant improvements in PA and others reporting mixed effects for improvements in PA. Similarly, evidence regarding the effectiveness of digital interventions for SB was mixed, as only 1 (20%) systematic review reported statistically significant reductions in sedentary time, while 2 (40%) systematic reviews reported no significant changes in

XSL•FO
RenderX

SB. Finally, the included systematic reviews had 0% CCA overlap, but they were of low and critically low quality.

Effectiveness to Improve PA and SB

Digital interventions had the potential to significantly improve PA outcomes; however, findings from this umbrella review were mixed. A total of 3 (60%) of the 5 systematic reviews comprised mixed results to support the effectiveness of interventions to significantly improve PA. The 2 (40%) systematic reviews that reported statistically significant results for changes in PA consisted of smartphone or tablet, video game, virtual reality, monitor- or sensor-based interventions, and mixed technological delivery methods. Surprisingly, these systematic reviews did not report on the theoretical basis of the interventions, and only 1 (20%) systematic review discussed the use of BCTs within the interventions.

The effectiveness of interventions targeting SB varied across the included systematic reviews. While one systematic review suggested the use of digital interventions and had small but significant reductions in sedentary time, other systematic reviews found nonsignificant effects on SB. There were some differences between the intervention delivery methods of the significant and nonsignificant systematic reviews, which could explain the differences in success for changing SB, for example, of the 5 included reviews, 2 (40%) nonsignificant systematic reviews had a lower proportion of monitor- or sensor-based and mixed technological delivery methods. However, these are observations. It could be argued that the challenge in measuring PA and SB outcomes is heavily dependent on the metrics used. Previous work aiming to compare movement behavior features processed by commonly used accelerometer metrics among adults found statistically significant differences in 24-hour movement behavior across 4 different metrics used to measure these behaviors [34]. Therefore, suggesting the success of intervention may be overestimated or underestimated depending on the metrics used to measure these PA and SB outcomes, thus resulting in different results and conclusions.

In addition, although most changes to SB were nonsignificant, reductions in SB were reported. Research has found that breaking up SB with short bursts of light-intensity PA can improve metabolic health and self-reported fatigue and reduce all-cause mortality [35]. Therefore, these interventions can still be beneficial for health, but the mixed evidence suggests that further research is needed to determine the most effective strategies to include within digital interventions to reduce SB within community-living adults.

Traditionally, public health research has focused on the importance of increasing PA to improve health and, as a result, there are more interventions to improve PA. However, as worldwide levels of SBs have risen in recent years [36] and more information about the health consequences of this has been recognized, the interest in SB has grown, but the development of interventions and subsequent research exploring effectiveness is still emerging. The substantial variation in how SB is defined across accelerometer brands and studies underscores the evolving nature of SB research, interventions, and commercial products [37]. The differences in the effectiveness of SB interventions within this umbrella review may be due to the systematic reviews and primary studies within them using different technologies and terminologies to understand SB. Once again, this highlights the difficulty in interpreting whether the interventions were effective or not.

Although findings are inconclusive regarding the effect of digital interventions on SB, this research has revealed that the inclusion of self-monitoring is more likely to explain changes in behavior in favor of the intervention. Of the 5 included systematic reviews, 1 (20%) reported that interventions consisting of self-monitoring components had a moderate-to-large impact on positively changing PA behavior, and although the change in SB was reported as not statistically significant, self-monitoring interventions were more effective in changing SB than those without this component.

Previously published systematic reviews have described self-monitoring to be a valuable and effective instrument for encouraging positive changes in PA behaviors [38,39]. Accurate self-monitoring of PA and SB has the potential to increase individuals' awareness of their behavior, which can encourage them to change their habits toward PA and SB. Self-monitoring through digital devices can also provide detailed data on individuals' movement behaviors and help detect patterns and triggers of these behaviors. Allowing tailored feedback based on these data can also enhance the effectiveness of this type of intervention [40].

Although research suggests the use of self-monitoring for effective behavior change, evidence of effectiveness for this BCT has been found to vary significantly across different populations and health behaviors [41]. To achieve more effective changes in PA and SB outcomes, self-monitoring needs to be used consistently [42]. Therefore, this highlights the need for research into the use of this BCT element in different populations, including informal carers, and for different health behaviors for it to be implemented effectively to create successful interventions, including digital interventions, to improve PA and reduce SB.

Theoretical Frameworks and BCTs

Findings from this umbrella review regarding the most prominent theoretical framework and implemented BCTs have been supported by previous literature. A systematic review and meta-analysis aimed to identify the most effective BCTs for increasing PA in digital and face-to-face interventions in adults with obesity or adults who are overweight reported that 45 included intervention studies were grounded in one or multiple theoretical frameworks [43]. Social cognitive theory and self-regulation theory, alone or in combination, were used most often in digital interventions aimed at increasing PA [43]. A scoping review of user models for personalized PA interventions also revealed prominent theoretical frameworks used in interventions, including Fogg behavior model, social cognitive theory, transtheoretical model, and health action process approach [44], all of which were identified within this umbrella review. Finally, scoping reviews have also reported that the most frequently implemented BCTs in digital PA interventions for adults include goal setting, self-monitoring, and feedback on behaviors [45,46].

XSL•FO
RenderX

Although there are similarities between the findings of this umbrella review and previous work, other theoretical frameworks and BCTs have been reported in these studies that were not reported in this umbrella review. Ghanvatkar et al [44] mentioned that the theory of planned behavior, the I-change model, and the behavior change wheel were also found to frequently guide digital PA interventions [44]. Similarly, De Santis et al [45] and Meinhart et al [46] found that providing educational and motivational content, feedback on exercise, and integrated tailored responses were all present in digital interventions. In addition, a systematic review exploring the effectiveness of BCTs at increasing older adults' self-efficacy and PA reported that using barrier identification and problem-solving, providing rewards contingent on successful behavior, and demonstrating the behavior were significantly associated with higher levels of PA [47]. It also reported that the use of providing normative information about others' behavior, providing information on where and when to perform the behavior, and planning social support or social change resulted in significantly lower PA behavior effect sizes for older adults [47].

While there is overlap between the findings of this umbrella review and previous research regarding the theoretical frameworks and BCTs used in digital PA interventions, there are additional frameworks and BCTs not reported in this umbrella review, highlighting the diversity of these interventions for promoting and improving PA and SB levels. The reason for these differences may be due to the different populations targeted and the different interventions used. However, it is also important to consider that not all interventions will use the same frameworks or BCTs nor will they recognize and report on all frameworks or BCTs used. Intervention descriptions may not have been sufficiently detailed and precise enough to identify the theoretical frameworks and BCTs present within primary studies or systematic reviews. With regard to this umbrella review, the identification of the implemented theoretical frameworks and BCTs was limited to their presentation within the systematic reviews used for this research. Therefore, our findings are subject to the parameters of the systematic reviews and not the primary studies reported within them.

It is also important to note that just because certain frameworks and BCTs were not reported, it does not mean they are not used or worthwhile being used; it is simply that they were not reported to be present in the systematic reviews included in this umbrella review. There are many other theoretical frameworks and BCTs that could potentially be of use, specifically for different populations, as demonstrated by the other reviews discussed. However, more research is needed to understand the effectiveness of these theories and BCTs, especially within the informal caring population.

Quality of Included Reviews

The quality assessment process, which showed moderate agreement between reviewers, highlights the credibility of the quality assessment process. However, the methodological quality of the included systematic reviews was low (20%) and critically low (80%).

While there were some common strengths of the systematic reviews regarding methodological quality, for example, the inclusion of Population, Intervention, Comparison, and Outcome components, reporting that the review methods were established before the conduct of the review and the explanation of study design can strengthen the methodological rigor of these systematic reviews. The methodological weaknesses, such as a lack of a comprehensive literature search strategy and a lack of justification for excluded studies, could reduce the confidence in the validity of the included systematic reviews.

The low confidence ratings of these systematic reviews could be because of the high standards set by the AMSTAR-2 tool. Although AMSTAR-2 has been recognized as a useful appraisal tool to understand the quality of published systematic reviews [48], there are some challenges in using this method for quality assessment. It has been discussed in previous work and reported in other umbrella reviews that AMSTAR-2 produces a generally high proportion of critically low confidence ratings among systematic reviews of health interventions, and there is ongoing debate whether this is due to poor discriminatory powers of the tool or the low quality of systematic reviews of health care interventions [49]. Individual items of AMSTAR-2 have been discussed in work by De Santis et al [49] to highlight difficulties of the tool and potential reasons for lower quality ratings of systematic reviews, for example, the "all-or-none" ratings of items that do not allow to differentiate between missing information and incomplete reporting of information for some items. The agreement of 69% between reviewers for approaching AMSTAR-2 in this umbrella review highlights the subjectiveness of this methodological tool, which has also been criticized by De Santis et al [49]. Due to the rapid increase in systematic reviews for health care interventions, it has been recommended that the items of AMSTAR-2 be empirically tested to improve the appraisal process and increase user consensus.

As this umbrella review explores digital interventions for improving PA and SB for community-living adults, its findings may have practical implications for public health policies and future digital interventions. However, as findings are from low and critically low–quality systematic reviews, there is a need for cautious interpretation. Stakeholders and policy makers must be aware of the methodological limitations of these systematic reviews when considering the findings presented within this umbrella review. Furthermore, these findings suggest the importance and need for more high-quality research within this field to provide methodologically sound evidence for future policy makers and decision makers.

Strengths

There are multiple strengths of this umbrella review. First, this umbrella review followed the Joanna Briggs Institute framework, ensuring a rigorous and systematic approach. Adherence to this framework also promotes transparency and replicability, making the results more robust and credible. The incorporation of a large number of primary publications from the included systematic reviews allowed a broad and inclusive scope of this review to be conducted. All primary publications were unique with a CCA score of 0%, which indicates no

XSL•FO
RenderX

overlap and strengthens the robustness of the review's conclusions. Furthermore, the geographic reach of the reviews included and the primary studies within them were extensive, with studies conducted in 24 different countries. This wide geographic representation adds to the generalizability of the findings across diverse populations and different groups of community-living adults. Another strength of this review was the identification of a vast range of theoretical frameworks and BCTs, which provides a comprehensive understanding of the strategies used in digital PA and SB interventions for community-living adults. Finally, the search strategy of this review highlights a critical gap in the existing literature as very few systematic reviews focused purely on PA and SB and theoretical frameworks and BCTs, as numerous systematic reviews had to be excluded due to their inclusion of multiple behaviors, such as diet and sleep.

Limitations

The lack of high-quality systematic reviews within this umbrella review restricts the potential to use these findings to aid future decision-making. Therefore, there is a need for more high-quality systematic reviews in the digital PA and SB intervention research field that include justifications for the search strategy, provide lists of excluded studies, report the sources of funding for the included studies, and provide an explanation for heterogeneity. As most (3/5, 60%) included systematic reviews lacked meta-analysis and reported multiple different PA outcomes, it is difficult to synthesize quantitative data and nonquantitative data to provide a clear answer to understand the effectiveness of these interventions in improving PA and SB. Finally, it is possible that some relevant studies were missed due to differences in the search strategies, inclusion criteria, and the databases used. Furthermore, restricting the review to English-language publications may have excluded important research published in other languages.

Future Work

Future work should include more primary studies and systematic reviews to be published specifically focusing on PA- and SB-only interventions instead of interventions that combine these behaviors with other health behaviors, such as diet and sleep. There is also a need for more primary studies to use consistent and accurate measures of PA and SB, and standardized outcomes for measuring these behaviors should be implemented.

In addition, further investigation is also warranted for the development and refinement of intervention strategies specifically for reducing SB among community-living adults to achieve more robust outcomes of the effectiveness of digital interventions to reduce this behavior. Recommendations for future work also include better reporting of theoretical frameworks and BCTs used in digital interventions for PA and SB and meta-analysis of currently published systematic reviews to allow for further evaluation to be conducted into the effectiveness of these components to change behaviors. Finally, more meta-analyses of systematic reviews assessing the effectiveness of digital interventions on PA and SB are needed to provide a more comprehensive and accurate picture of the current evidence.

Conclusions

This umbrella review aimed to evaluate the effectiveness of digital interventions for changing PA and SBs among community-living adults, identify the components and content of these interventions, and assess the methodological quality of the included systematic reviews.

Overall, this umbrella review indicates that digital interventions have the potential to improve PA and reduce SB. Meta-analyses suggest these interventions can be effective, particularly those incorporating self-monitoring and BCTs. However, more work is needed to understand the effectiveness and impact of specific theoretical frameworks and BCTs on changing PA and SBs. The mixed findings and methodological weaknesses highlight the need for more comprehensive reporting and analyses to be conducted to understand the effectiveness of digital interventions for community-living adults.

In relation to informal carers, this umbrella review highlights that digital PA and SB interventions offer a valuable opportunity to support behavior change in this population due to their scalability and numerous delivery methods. These digital interventions have been found to consist of a theoretical grounding and can incorporate numerous BCTs, all of which could improve the effectiveness of these interventions to support and encourage PA and reduce SB in informal carers. Despite the potential benefits of these interventions for informal carers, more coordinated efforts and research are needed to understand the potential and effectiveness of digital interventions for improving PA and reducing SB in this group.

Acknowledgments

The authors would like to thank the University of Strathclyde graduate Leslie Stretch, whose philanthropy made this research possible (the funder of the study had no role in any aspect of the study design, analysis, or decision to publish this work). They would also like to acknowledge Bethan Naunton Morgan for her guidance and knowledge surrounding the umbrella review process, which she generously shared before conducting this review.

Authors' Contributions

ER designed and conducted the research; screened the studies; extracted, analyzed, and interpreted the data; and wrote this paper. KE, AK, and MDD made substantial contributions to the study conception and design of the research. WH screened the studies and assessed the methodology quality assessment process. MP screened the studies. All authors have read and agreed to the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy, keywords, and outputs.
[DOC File , 54 KB - jmir_v27i1e66294_app1.doc]

Multimedia Appendix 2
Overlap of reviews.
[XLSX File (Microsoft Excel File), 26 KB - jmir_v27i1e66294_app2.xlsx]

References

1. Tremblay M, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, et al. Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/S12966-017-0525-8]
2. Wahid A, Manek N, Nichols M, Kelly P, Foster C, Webster P, et al. Quantifying the association between physical activity and cardiovascular disease and diabetes: a systematic review and meta‐analysis. JAHA 2016 Sep;5(9). [doi: 10.1161/jaha.115.002495] [Medline: 27628572]
3. Albini A, La Vecchia C, Magnoni F, Garrone O, Morelli D, Janssens JP, et al. Physical activity and exercise health benefits: cancer prevention, interception, and survival. Eur J Cancer Prev 2025 Jan 01;34(1):24-39. [doi: 10.1097/CEJ.0000000000000898] [Medline: 38920329]
4. Posadzki P, Pieper D, Bajpai R, Makaruk H, Könsgen N, Neuhaus AL, et al. Exercise/physical activity and health outcomes: an overview of Cochrane systematic reviews. BMC Public Health 2020 Nov 16;20(1):1724 [FREE Full text] [doi: 10.1186/s12889-020-09855-3] [Medline: 33198717]
5. Global status report on physical activity 2022. World Health Organization. 2022 Oct 19. URL: https://www.who.int/publications/i/item/9789240059153 [accessed 2024-05-05]
6. Patterson R, McNamara E, Tainio M, de Sá TH, Smith AD, Sharp SJ, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol 2018 Sep;33(9):811-829 [FREE Full text] [doi: 10.1007/s10654-018-0380-1] [Medline: 29589226]
7. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc 2009 May;41(5):998-1005. [doi: 10.1249/MSS.0b013e3181930355] [Medline: 19346988]
8. Dunstan DW, Barr EL, Healy GN, Salmon J, Shaw JE, Balkau B, et al. Television viewing time and mortality: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Circulation 2010 Jan 26;121(3):384-391. [doi: 10.1161/CIRCULATIONAHA.109.894824] [Medline: 20065160]
9. Strain T, Flaxman S, Guthold R, Semenova E, Cowan M, Riley LM, et al. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5·7 million participants. Lancet Glob Health 2024 Aug;12(8):e1232-e1243 [FREE Full text] [doi: 10.1016/S2214-109X(24)00150-5] [Medline: 38942042]
10. Monitoring and evaluating digital health interventions. World Health Organization. 2016 Dec 12. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2024-05-20]
11. Global action plan on physical activity 2018–2030: more active people for a healthier world. World Health Organization. 2018 Jun 1. URL: http://who.int/publications/i/item/9789241514187 [accessed 2024-05-20]
12. Dishman RK, Buckworth J. Adherence to physical activity. In: Morgan WP, editor. Physical Activity and Mental Health. Milton Park, UK: Taylor & Francis; 1997:63-80.
13. Rhodes RE, Nigg CR. Advancing physical activity theory: a review and future directions. Exerc Sport Sci Rev 2011 Jul;39(3):113-119. [doi: 10.1097/JES.0b013e31821b94c8] [Medline: 21705861]
14. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
15. Greaves C, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
16. Dombrowski SU, Sniehotta FF, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychol Rev 2012 Mar;6(1):7-32 [FREE Full text] [doi: 10.1080/17437199.2010.513298]
17. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701 [FREE Full text] [doi: 10.1037/a0016136] [Medline: 19916637]

XSL•FO
RenderX

18. Horne J, Kentzer N, Smith L, Trott M, Vseteckova J. A systematic review on the prevalence of physical activity, and barriers and facilitators to physical activity, in informal carers in the United Kingdom. J Phys Act Health 2021 Feb 01;18(2):212-218. [doi: 10.1123/jpah.2020-0526] [Medline: 33485270]
19. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. Evid Based Ment Health 2018 Aug;21(3):95-100 [FREE Full text] [doi: 10.1136/ebmental-2018-300014] [Medline: 30006442]
20. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015 Sep;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
21. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. PLoS One 2012;7(11):e49667 [FREE Full text] [doi: 10.1371/journal.pone.0049667] [Medline: 23166744]
22. Naunton Morgan B, Windle G, Sharp R, Lamers C. eHealth and web-based interventions for informal carers of people with dementia in the community: umbrella review. J Med Internet Res 2022 Jul 22;24(7):e36727 [FREE Full text] [doi: 10.2196/36727] [Medline: 35867388]
23. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
24. Longhini J, Marzaro C, Bargeri S, Palese A, Dell'Isola A, Turolla A, et al. Wearable devices to improve physical activity and reduce sedentary behaviour: an umbrella review. Sports Med Open 2024 Jan 14;10(1):9 [FREE Full text] [doi: 10.1186/s40798-024-00678-9] [Medline: 38219269]
25. Aromataris EC, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom PT. Chapter 10: Umbrella reviews. In: JBI Manual for Evidence Synthesis. Adelaide, South Australia: Joanna Briggs Institute; 2020.
26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
27. Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol 2014 Apr;67(4):368-375. [doi: 10.1016/j.jclinepi.2013.11.007] [Medline: 24581293]
28. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
29. Lee S, Patel P, Myers ND, Pfeiffer KA, Smith AL, Kelly KS. A systematic review of eHealth interventions to promote physical activity in adults with obesity or overweight. Behav Med 2023;49(3):213-230. [doi: 10.1080/08964289.2022.2065239] [Medline: 35465851]
30. Berry A, McCabe CS, Muir S, Walsh N. Digital behaviour change interventions to facilitate physical activity in osteoarthritis: a systematic review. Phys Ther Revs 2018 Jul 02;23(3):197-206 [FREE Full text] [doi: 10.1080/10833196.2018.1470747]
31. Wang SC, Kassavou A. Digital health behavioural interventions to support physical activity and sedentary behaviour in adults after stroke: a systematic literature review with meta-analysis of controlled trials. Behav Sci (Basel) 2023 Jan 10;13(1):62 [FREE Full text] [doi: 10.3390/bs13010062] [Medline: 36661634]
32. Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan D, McVicar J, Abdelrazek M, Maddison R. Smartphone-based interventions to reduce sedentary behavior and promote physical activity using integrated dynamic models: systematic review. J Med Internet Res 2021 Sep 13;23(9):e26315 [FREE Full text] [doi: 10.2196/26315] [Medline: 34515637]
33. Wu S, Li G, Du L, Chen S, Zhang X, He Q. The effectiveness of wearable activity trackers for increasing physical activity and reducing sedentary time in older adults: a systematic review and meta-analysis. Digit Health 2023;9:20552076231176705 [FREE Full text] [doi: 10.1177/20552076231176705] [Medline: 37252261]
34. Willems I, Verbestel V, Dumuid D, Calders P, Lapauw B, De Craemer M. A comparative analysis of 24-hour movement behaviors features using different accelerometer metrics in adults: implications for guideline compliance and associations with cardiometabolic health. PLoS One 2024;19(9):e0309931 [FREE Full text] [doi: 10.1371/journal.pone.0309931] [Medline: 39288135]
35. Vincent GE, Jay SM, Vandelanotte C, Ferguson SA. Breaking up sitting with light-intensity physical activity: implications for shift-workers. Int J Environ Res Public Health 2017 Oct 16;14(10):1233 [FREE Full text] [doi: 10.3390/ijerph14101233] [Medline: 29035315]
36. Katzmarzyk PT, Friedenreich C, Shiroma EJ, Lee IM. Physical inactivity and non-communicable disease burden in low-income, middle-income and high-income countries. Br J Sports Med 2022 Jan;56(2):101-106 [FREE Full text] [doi: 10.1136/bjsports-2020-103640] [Medline: 33782046]
37. Sievänen H, Kujala UM. Accelerometry-simple, but challenging. Scand J Med Sci Sports 2017 Jun;27(6):574-578. [doi: 10.1111/sms.12887] [Medline: 28466474]
38. Knittle K, Nurmi J, Crutzen R, Hankonen N, Beattie M, Dombrowski SU. How can interventions increase motivation for physical activity? A systematic review and meta-analysis. Health Psychol Rev 2018 Sep;12(3):211-230. [doi: 10.1080/17437199.2018.1435299] [Medline: 29385950]

XSL•FO
RenderX

39. Olander EK, Fletcher H, Williams S, Atkinson L, Turner A, French DP. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013 Mar 03;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]
40. Page EJ, Massey AS, Prado-Romero PN, Albadawi S. The use of self-monitoring and technology to increase physical activity: a review of the literature. Perspect Behav Sci 2020 Sep;43(3):501-514 [FREE Full text] [doi: 10.1007/s40614-020-00260-0] [Medline: 33029577]
41. Hennessy EA, Johnson BT, Acabchuk RL, McCloskey K, Stewart-James J. Self-regulation mechanisms in health behavior change: a systematic meta-review of meta-analyses, 2006-2017. Health Psychol Rev 2020 Mar;14(1):6-42 [FREE Full text] [doi: 10.1080/17437199.2019.1679654] [Medline: 31662031]
42. Carels RA, Darby LA, Rydin S, Douglass OM, Cacciapaglia HM, O'Brien WH. The relationship between self-monitoring, outcome expectancies, difficulties with eating and exercise, and physical activity and weight loss treatment outcomes. Ann Behav Med 2005 Dec;30(3):182-190. [doi: 10.1207/s15324796abm3003_2] [Medline: 16336069]
43. Carraça E, Encantado J, Battista F, Beaulieu K, Blundell J, Busetto L, et al. Effective behavior change techniques to promote physical activity in adults with overweight or obesity: a systematic review and meta-analysis. Obes Rev 2021 Jul;22 Suppl 4(Suppl 4):e13258 [FREE Full text] [doi: 10.1111/obr.13258] [Medline: 33949778]
44. Ghanvatkar S, Kankanhalli A, Rajan V. User models for personalized physical activity interventions: scoping review. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11098 [FREE Full text] [doi: 10.2196/11098] [Medline: 30664474]
45. De Santis KK, Jahnel T, Matthias K, Mergenthal L, Al Khayyal H, Zeeb H. Evaluation of digital interventions for physical activity promotion: scoping review. JMIR Public Health Surveill 2022 May 23;8(5):e37820 [FREE Full text] [doi: 10.2196/37820] [Medline: 35604757]
46. Meinhart F, Stütz T, Sareban M, Kulnik ST, Niebauer J. Mobile technologies to promote physical activity during cardiac rehabilitation: a scoping review. Sensors (Basel) 2020 Dec 24;21(1):65 [FREE Full text] [doi: 10.3390/s21010065] [Medline: 33374322]
47. French DP, Olander EK, Chisholm A, Mc Sharry J. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
48. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
49. De Santis KK, Pieper D, Lorenz RC, Wegewitz U, Siemens W, Matthias K. User experience of applying AMSTAR 2 to appraise systematic reviews of healthcare interventions: a commentary. BMC Med Res Methodol 2023 Mar 16;23(1):63 [FREE Full text] [doi: 10.1186/s12874-023-01879-8] [Medline: 36927334]

Abbreviations

AMSTAR-2: A Measurement Tool to Assess Systematic Reviews-2
BCT: behavior change technique
CCA: corrected covered area
DHI: digital health intervention
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PICOS: population, intervention, context, outcomes, and study
SB: sedentary behavior
SMD: standardized mean difference

Edited by A Mavragani; submitted 10.09.24; peer-reviewed by J Longhini, H Sievanen; comments to author 03.12.24; revised version received 04.02.25; accepted 06.02.25; published 18.03.25.

Please cite as:
Russell E, Kirk A, Dunlop MD, Hodgson W, Patience M, Egan K
Digital Physical Activity and Sedentary Behavior Interventions for Community-Living Adults: Umbrella Review
J Med Internet Res 2025;27:e66294
URL: https://www.jmir.org/2025/1/e66294
doi:10.2196/66294
PMID:

XSL•FO
RenderX

©Eilidh Russell, Alison Kirk, Mark D Dunlop, William Hodgson, Mhairi Patience, Kieren Egan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Enhancing Patient Outcome Prediction Through Deep Learning With Sequential Diagnosis Codes From Structured Electronic Health Record Data: Systematic Review

Tuankasfee Hama[1], MD; Mohanad M Alsaleh[1,2], MSc; Freya Allery[1], MSc; Jung Won Choi[1], PhD; Christopher Tomlinson[1], MBBS; Honghan Wu[1], Prof Dr; Alvina Lai[1], PhD; Nikolas Pontikos[3], PhD; Johan H Thygesen[1], PhD

[1]Institute of Health Informatics, University College London, London, United Kingdom

[2]Department of Health Informatics, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia

[3]UCL Institute of Ophthalmology, University College London, London, United Kingdom

Corresponding Author:
Tuankasfee Hama, MD
Institute of Health Informatics
University College London
222 Euston Road
London, NW1 2DA
United Kingdom
Phone: 44 0207679200
Email: tuankasfee.hama.21@ucl.ac.uk

Abstract

Background: The use of structured electronic health records in health care systems has grown rapidly. These systems collect huge amounts of patient information, including diagnosis codes representing temporal medical history. Sequential diagnostic information has proven valuable for predicting patient outcomes. However, the extent to which these types of data have been incorporated into deep learning (DL) models has not been examined.

Objective: This systematic review aims to describe the use of sequential diagnostic data in DL models, specifically to understand how these data are integrated, whether sample size improves performance, and whether the identified models are generalizable.

Methods: Relevant studies published up to May 15, 2023, were identified using 4 databases: PubMed, Embase, IEEE Xplore, and Web of Science. We included all studies using DL algorithms trained on sequential diagnosis codes to predict patient outcomes. We excluded review articles and non–peer-reviewed papers. We evaluated the following aspects in the included papers: DL techniques, characteristics of the dataset, prediction tasks, performance evaluation, generalizability, and explainability. We also assessed the risk of bias and applicability of the studies using the Prediction Model Study Risk of Bias Assessment Tool (PROBAST). We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist to report our findings.

Results: Of the 740 identified papers, 84 (11.4%) met the eligibility criteria. Publications in this area increased yearly. Recurrent neural networks (and their derivatives; 47/84, 56%) and transformers (22/84, 26%) were the most commonly used architectures in DL-based models. Most studies (45/84, 54%) presented their input features as sequences of visit embeddings. Medications (38/84, 45%) were the most common additional feature. Of the 128 predictive outcome tasks, the most frequent was next-visit diagnosis (n=30, 23%), followed by heart failure (n=18, 14%) and mortality (n=17, 13%). Only 7 (8%) of the 84 studies evaluated their models in terms of generalizability. A positive correlation was observed between training sample size and model performance (area under the receiver operating characteristic curve; *P*=.02). However, 59 (70%) of the 84 studies had a high risk of bias.

Conclusions: The application of DL for advanced modeling of sequential medical codes has demonstrated remarkable promise in predicting patient outcomes. The main limitation of this study was the heterogeneity of methods and outcomes. However, our analysis found that using multiple types of features, integrating time intervals, and including larger sample sizes were generally related to an improved predictive performance. This review also highlights that very few studies (7/84, 8%) reported on challenges related to generalizability and less than half (38/84, 45%) of the studies reported on challenges related to explainability. Addressing these shortcomings will be instrumental in unlocking the full potential of DL for enhancing health care outcomes and patient care.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42018112161; https://tinyurl.com/yc6h9rwu

(J Med Internet Res 2025;27:e57358) doi:10.2196/57358

KEYWORDS

deep learning; electronic health records; EHR; diagnosis codes; prediction; patient outcomes; systematic review

Introduction

Background

In recent decades, there has been a rapid growth in the use of electronic health records (EHRs) in health care systems, making them an important tool for health care workers and allowing for secondary use for research purposes. Structured EHRs contain temporal records of patient visits, incorporating various clinical data such as diagnosis codes, procedures, and laboratory test results, all of which may help researchers in predicting patient outcomes. Patient timelines can be organized based on diagnosis codes and their corresponding visit times, allowing deep learning (DL) algorithms to model and understand disease progression, with the time between visits representing the speed of disease progression. Using sequential diagnostic data from EHRs in this manner is a promising avenue for DL-based studies, but the degree to which this information has been used in published studies and its benefits has not yet been explored in the context of a systematic review, which is what this study sets out to do.

Classical machine learning (ML) techniques that require feature selection can be applied to include diagnosis codes as a binary feature for outcome prediction; for example, a recent study applied ML-based algorithms (logistic regression, extreme gradient boosting, and random forest) to identify cardiomyopathy [1]. However, traditional ML approaches cannot take full advantage of structured EHR data due to four key challenges:

1. Feature selection—manual feature selection, which requires medical knowledge from professional health care workers, is a time-consuming task and an expensive process.
2. High dimensionality—models suffer from a high-dimensional input representation due to the vast number of medical codes available (eg, Medical Information Mart for Intensive Care [MIMIC]-IV includes >15,000 unique *International Classification of Diseases, Tenth Revision* [ICD-10], codes that appear in the patient records) [2].
3. Hierarchy—the hierarchical structure of diagnosis codes may represent relationships between similar disease categories, but this information is ignored by traditional ML-based approaches.
4. Temporality—the majority of traditional ML techniques struggle to effectively capture information contained in the temporal chronological sequence of patients' medical history, where the time between consecutive visits may vary in length from a few days to numerous months. Significant predictive insights may be hidden in the temporal intervals and sequence of diagnosis codes in a patient's evolving medical history because deterioration or improvement in outcomes may follow specific patterns and frequencies of interactions with the health care system [3-5].

To comprehensively uncover and understand the impact of these intricate temporal and sequential relationships within the data, advanced DL methods are essential.

DL approaches have been applied previously in the health care domain. Systematic reviews show a good progression in DL-based algorithms for various medical data types, such as clinical notes [6], medical images [7], and physiological signals [8]. DL emerges as a solution to overcome the aforementioned limitations of traditional ML for the following reasons: (1) DL functions as an end-to-end system that can automatically uncover an association between input and output with minimal need for feature engineering or domain expertise; (2) DL models can generate an effective embedding space to cope with the high-dimensional problem (eg, a study demonstrated the effectiveness of an autoencoder in transforming RNA sequence data with approximately 20,000 features into a low-dimensional representation with approximately 1000 features, achieving high classification performance [9]); (3) some DL techniques, such as graph neural networks (GNNs), have been shown to give a good representation of hierarchical data [10]; and (4) to deal with temporal information, long short-term memory (LSTM) and temporal convolutional neural network (CNN) models have been adapted widely for complex sequential information in health care (eg, an LSTM model has been shown to be able to achieve a good performance in analyzing information from high-volume regular sequences such as intensive care unit [ICU] monitoring data [11]).

Although many DL techniques are well suited for hierarchical and time-series data, they face challenges in handling sequential diagnosis codes. In EHRs, diagnosis codes occur at irregular time intervals, reflecting the varying times between medical events, that is, some patients may have multiple visits within the same week, while for others, there may be months or years between visits. This irregularity complicates analysis but is also a source of information because it may capture the rapid or slow progression of conditions. Moreover, diagnosis codes require an embedding layer before being processed by a DL model, unlike continuous values from ICU monitors, which can be directly processed. Other challenges with DL are the need for extensive datasets and concerns about explainability [12]. The performance of existing DL techniques depends on the volume and quality of the training dataset, which, in the field of health data science, may be problematic because large-scale datasets may not be available due to privacy concerns. This can also contribute to the generalizability issue because models may perform well on internal training and test datasets but perform poorly on independent external data sources. Moreover, DL model predictions can sometimes be unclear to clinicians, and there is a need to explain the main factors contributing to the

XSL•FO
RenderX

model's output. Therefore, it becomes a significant challenge for researchers to use special DL techniques for outcome prediction by using sequential diagnosis codes. This systematic review will comprehensively explore these challenges and the approaches used to deal with them in the published literature.

To date, several reviews have analyzed DL methods trained on EHR data [13-16]. Various kinds of EHR data for DL-based algorithms have been surveyed: (1) structured data (diagnosis codes, medication codes, procedure codes, laboratory test results, and vital signs) and (2) unstructured data (clinical notes, medical images, and physiological signals). However, none of the reviews primarily focused on the use of sequential diagnostic data in DL for outcome prediction. Moreover, none of them reported on the inclusion of external validation, which can be problematic because models are applied to different data distributions. Many questions remain unanswered, such as common DL techniques, types of diagnosis codes, additional features (eg, time between visits, demographic data, and medications), dataset characteristics, prediction tasks, generalizability, and explainability.

Objectives

We conducted a systematic review to answer these questions and investigate the current state of DL in the context of outcome prediction using sequential diagnostic information. By summarizing the research in this area, our review can help guide future DL-based prediction studies by identifying current research gaps and challenges. The main objective of this systematic review was to identify and summarize existing DL studies that use sequential diagnosis codes as key predictors of patient outcomes. In addition, this study investigates the challenges of generalizability and explainability in these predictive models.

Methods

Definition

In this systematic review, we defined sequential diagnosis codes as medical codes (eg, Systemized Nomenclature of Medicine–Clinical Terms; *International Classification of Diseases, Ninth Revision*; and *ICD-10* codes) assigned to patients to represent their visits within the health care system. This review examined various categories of DL algorithms, including recurrent neural networks (RNNs), LSTM models, CNNs, transformer-based models, and GNNs, in addition to some techniques such as time-awareness and attention mechanisms. No restrictions were placed on study outcomes, which included mortality, hospitalization status, and onset of disease (eg, hypertension, diabetes, heart attack, stroke, and cancer).

Search Strategy

As our review combines knowledge from both health care and engineering, we sought to identify all relevant studies in both domains using 4 databases: PubMed, Embase, IEEE Xplore, and Web of Science. In addition, we conducted a manual search of the reference lists of the included studies to identify additional relevant articles. We searched the databases up to May 15, 2023. To promote transparency and prevent duplication, the study protocol was registered in PROSPERO (CRD42023434032).

We used 4 main groups of keywords centered around DL techniques, EHRs, sequence, and prediction. The literature search included the following search terms: ("deep learning" OR "RNN" OR "LSTM" OR "CNN" OR "transformer" OR "BERT" OR "time attention" OR "attention based" OR "graph neural network") AND ("electronic health records" OR "EHRs" OR "electronic health record" OR "EHR" OR "electronic medical record" OR "EMR" OR "electronic medical records" OR "EMRs") AND ("longitudinal" OR "visit" OR "sequential" OR "sequence" OR "temporal") AND ("risk" OR "predictions" OR "prediction" OR "patient outcome" OR "prognosis"). The search query returned the same set of results as those obtained using the built-in search functionality of the literature databases.

Inclusion and Exclusion Criteria

This systematic review followed the Population, Intervention, Comparison, and Outcomes framework to identify and select articles in the databases [17]. The population included patients of all ages in EHR databases, the intervention involved DL-based methods for sequential diagnosis codes, the comparison was between different algorithms, and the outcome was model performance.

We included all studies using DL algorithms to predict patient outcomes by training the models on sequential or longitudinal diagnosis codes, as defined in the aforementioned search terms. We excluded review articles and non–peer-reviewed papers. In addition, we excluded papers that primarily dealt with other nondiagnostic EHR data types, including physiological signals, clinical notes, and medical images. To reduce bias, 2 reviewers independently screened all studies. Any discrepancies between the reviewers were resolved through discussion to reach a consensus. The level of agreement between the reviewers was assessed using the Cohen κ coefficient.

Extraction and Analysis

We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [18] (Multimedia Appendix 1) to report our findings. We evaluated the following aspects in the included papers: DL techniques, characteristics of the datasets, prediction tasks, and performance evaluation. For each study, we selected either the novel, proposed technique or the best-performing technique as the main DL model architecture. The findings are presented using plots, figures, and tables.

This review followed the suggestions of a previous study [19] to assess the generalizability of the applied models. Generalizability was evaluated based on the potential applicability of outcome predictive models beyond their original development context, focusing on 3 key aspects:

1. Demographic validation, which investigates the model's adaptability to distinct clinical contexts, including disparities related to sex or ethnicity and variations in age groups
2. Temporal validation, which focuses on assessing the model's performance over time within its original development environment

XSL•FO
RenderX

3. Geographic validation, which explores the model's capacity to extend its utility beyond its original development setting to different locations, institutions, or geographic contexts

In addition, we evaluated the explainability of the model in each included study.

Risk-of-Bias and Quality Assessment

The Prediction Model Study Risk of Bias Assessment Tool (PROBAST) [20] was used to evaluate both the risk of bias (ROB) and the applicability of the best-performing DL models in the included studies. ROB was evaluated based on a set of 20 questions categorized into 4 domains: participants, predictors, outcome, and analysis. Applicability was evaluated via a main question for each of the following 3 domains: participants, predictors, and outcomes. Each domain was rated as having low, unclear, or high ROB. If multiple models were reported in a study, only the model with the highest area under the receiver operating characteristic curve (AUROC) and F_1-score was evaluated. One reviewer conducted the PROBAST assessment for all included studies.

Results

Study Selection

Figure 1 presents the PRISMA diagram of the search and screening results. Initially, our search identified 740 records, of which 377 (50.9%) duplicates were removed. The screening process consisted of 2 stages: title and abstract screening, followed by full-text screening. During the title and abstract screening, we assessed the study aim, objectives, and methods to determine whether each paper fell within the scope of our review. Ultimately, 84 (11.4%) of the initially identified 740 articles were included in the final analysis. The agreement between reviewers had a Cohen κ coefficient of 0.65, indicative of a moderate agreement [21]. All included studies are listed in Table 1 and Multimedia Appendix 2 [22-105].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the search and screening results.

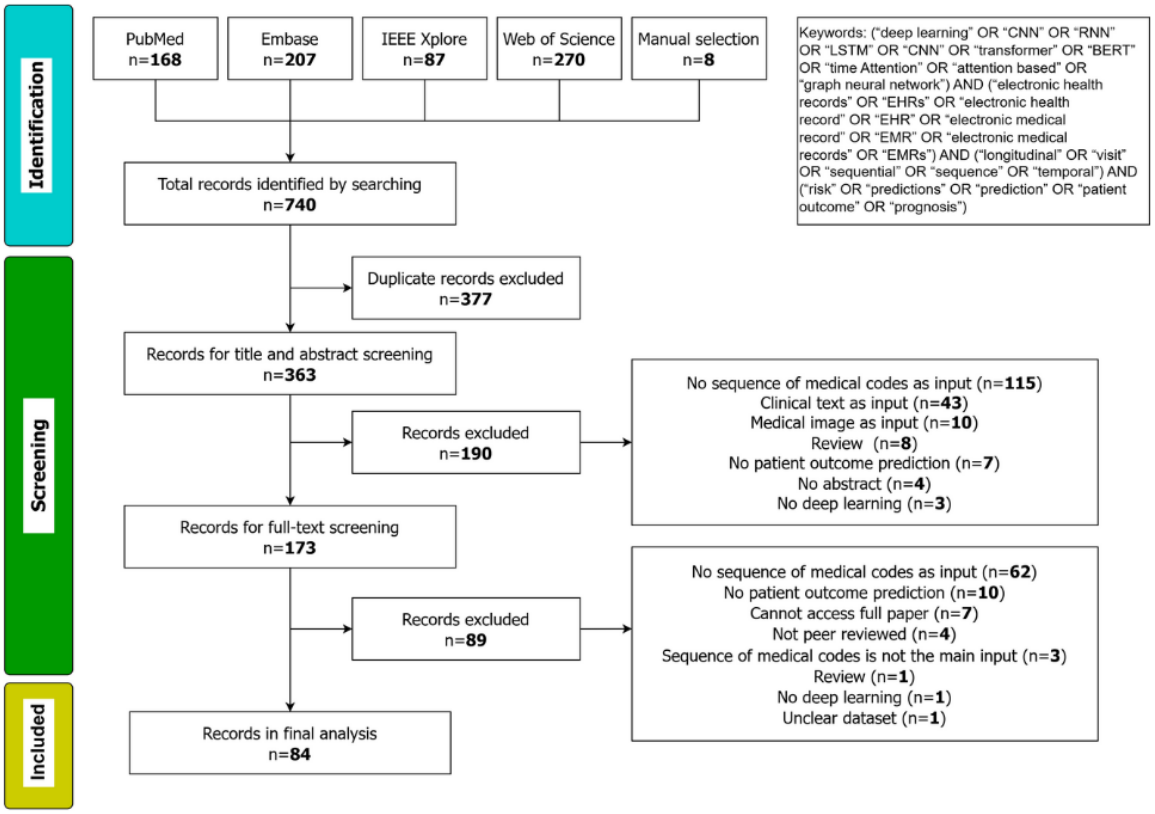

XSL•FO
RenderX

Table 1. Summary list of included studies, highlighting deep learning (DL) approaches, prediction tasks, additional features, included time, performance evaluation, risk of bias (ROB), and concern of applicability (COA).

Study; year	DL approach	Prediction task	Additional features	Included time	Performance evaluation	ROB	COA
Miotto et al [22]; 2016	Autoencoders	New onset of disease	Demographic data, medications, procedures, laboratory test results, and clinical text	No	Accuracy, AUROC[a], and F_1-score	Unclear	Low
Nguyen et al [23]; 2016	CNN[b]	Readmission	No	Yes (as month)	Accuracy	High	Unclear
Choi et al [24]; 2016	GRU[c]	Next-visit diagnosis	Medications and procedures	Yes	Recall	High	Unclear
Pham et al [25]; 2016	LSTM[d]	Readmission	Procedures, medications, and admission type	Yes (as day)	F_1-score	High	Unclear
Choi et al [26]; 2016	RNN[e]	Heart failure	Medications and procedures	Yes	AUROC	High	Low
Ma et al [27]; 2017	GRU	Next-visit diagnosis	Procedures	No	Accuracy	High	Unclear
Choi et al [28]; 2017	RNN	Next-visit diagnosis and heart failure	No	No	Accuracy and AUROC	High	Low
Sha and Wang [29]; 2017	GRU	Mortality	No	No	AUROC, F_1-score, and MCC[f]	High	Unclear
Choi et al [30]; 2017	GRU	Heart failure	Medications and procedures	Yes	AUROC	High	Low
Suo et al [31]; 2017	CNN	Diabetes, obesity, and COPD[g]	No	No	Accuracy	High	Unclear
Amirkhan et al [32]; 2017	LSTM	Colorectal cancer	Medications and laboratory test results	Yes (as day)	AUROC	Low	Low
Lei et al [33]; 2018	RNN	Mortality and comorbidity	Demographic data, laboratory test results, medications, and procedures	No	Accuracy, AUC[h], and F_1-score	High	Low
Park et al [34]; 2018	RNN	CVDs[i]	Medications	Yes	Sensitivity, specificity, positive predictive value, F_1-score, and AUROC	High	Unclear
Bai et al [35]; 2018	RNN	Next-visit diagnosis	Procedures	Yes	Accuracy and F_1-score	High	High
Choi et al [36]; 2018	GRU and CNN	Heart failure and next-visit diagnosis	Medications and procedures	No	AUROC and AUPRC[j]	High	Low
Qiao et al [37]; 2018	RNN	Next-visit diagnosis	Procedures	Yes	Recall and AUROC	High	Unclear
Wang et al [38]; 2018	GRU	Next-visit diagnosis	Medications and demographic data	No	Precision and F_1-score	High	Unclear
Ma et al [39]; 2018	LSTM	Next-visit diagnosis	No	Yes (as week)	Accuracy and precision	High	Unclear
Zhang et al [40]; 2018	GRU	Hospitalization	Demographic data	No	AUROC, sensitivity, specificity, and F_2-score	Unclear	Low
Jin et al [41]; 2018	LSTM	Heart failure	No	No	AUROC, AUPRC, and F_1-score	High	Unclear

XSL•FO
RenderX

Study; year	DL approach	Prediction task	Additional features	Included time	Performance evaluation	ROB	COA
Guo et al [42]; 2019	LSTM	Next-visit diagnosis	Medications	No	Accuracy, recall, precision, and F_1-score	High	Unclear
Lin et al [43]; 2019	LSTM	Readmission	Vital signs and demographic data	Yes (as hour)	AUC and recall	Unclear	Unclear
Wang et al [44]; 2019	RNN	Next-visit diagnosis	Physical symptoms and medications	No	Precision	High	Unclear
Gao et al [45]; 2019	GRU	Next-visit diagnosis	Demographic data	No	Recall and precision	High	Unclear
Ma et al [46]; 2019	GNN[k]	Next-visit diagnosis	No	No	Precision and accuracy	High	Unclear
AlSaad et al [47]; 2019	LSTM	Asthma	No	No	AUROC	High	Low
Zhang et al [48]; 2019	LSTM and CNN	MCI[l], Alzheimer disease, and Parkinson disease	No	No	AUROC and F_1-score	High	Unclear
Ashfaq et al [49]; 2019	LSTM	Readmission	Human-derived features, procedures, medications, and laboratory test results	No	AUROC and F_1-score	High	Unclear
Ruan et al [50]; 2019	RNN-DAE[m]	Mortality	Medications, laboratory test results, and demographic data	No	AUROC	High	Low
Huang et al [51]; 2019	LSTM	Mortality	Laboratory test results	No	Accuracy, AUROC, and AUPRC	High	Low
Xiang et al [52]; 2019	LSTM	Heart failure	Medications and procedures	Yes	AUROC	High	Unclear
Gupta et al [53]; 2019	LSTM	Obesity	Demographic data, conditions, procedures, medications, and measurement	Yes (as month)	AUROC	High	Low
Shi et al [54]; 2020	LSTM	Mortality	Demographic data	No	Accuracy, recall, and F_1-score	High	Low
Li et al [55]; 2020	Transformers	Next-visit diagnosis and new onset of disease	Age	No	AUROC and precision	Low	Low
Peng et al [56]; 2020	Transformers	Readmission and next-visit diagnosis	Procedures	Yes (as day)	AUPRC and precision	Unclear	Low
Almog et al [57]; 2020	LSTM	Fracture	Demographic data	No	AUROC, recall, specificity, precision, and AUPRC	High	Unclear
Luo et al [58]; 2020	Transformers	COPD, heart failure, and kidney disease	No	Yes	Accuracy, precision, recall, F_1-score, and AUROC	High	Unclear
Zhang et al [59]; 2020	Transformers	Next-visit diagnosis, heart failure, diabetes, and chronic kidney disease	No	No	Precision, accuracy, and AUROC	High	Unclear
Rongali et al [60]; 2020	LSTM	Mortality	Procedures, laboratory test results, medications, clinical events, and demographic data	No	AUROC	Unclear	Unclear

XSL•FO
RenderX

Study; year	DL approach	Prediction task	Additional features	Included time	Performance evaluation	ROB	COA
Ye et al [61]; 2020	Transformers and CNN	Heart failure, kidney disease, and dementia	No	No	AUROC, precision, recall, and F_1-score	Unclear	Unclear
Zeng et al [62]; 2020	Transformers	Next-visit diagnosis	Procedures and medications	Yes	Recall	High	Low
An et al [63]; 2020	LSTM	Next-visit diagnosis	Medications and procedures	No	Jaccard similarity score, AUPRC, recall, and F_1-score	High	Low
Kabeshova et al [64]; 2020	LSTM and GRU	Relapse of urinary problems	Medications, procedures, and length of stay	Yes (as day)	Precision, AUROC, and AUPRC	Low	Low
Darabi et al [65]; 2020	Transformers	Readmission, mortality, length of stay, and next-visit diagnosis	Demographic data and clinical text	No	AUROC and AUPRC	Unclear	Unclear
An et al [66]; 2021	LSTM	Risk of CVDs	Demographic data, patient type, hospital visit times, and surgery history	No	Recall, precision, F_1-score, and AUROC	Low	Low
Meng et al [67]; 2021	Transformers	Depression	Demographic data and visit	No	AUROC and AUPRC	High	Low
Rasmy et al [68]; 2021	Transformers	Heart failure and cancers	No	No	AUROC	Low	Low
Ju et al [69]; 2021	CNN	Diabetes and heart failure	Vital signs, demographic data, medications, allergies, and smoking status	Yes (as day)	Accuracy, and AUROC	Unclear	Unclear
Harerimana et al [70]; 2021	GRU	Length of stay and mortality	Demographic data, free-text diagnosis, procedures	No	Accuracy, AUROC, AUPRC, F_1-score, and linear weighted κ	High	Unclear
Ningrum et al [71]; 2021	CNN	Risk of OA^n knee	Demographic data and medications	Yes (as week)	AUROC, sensitivity, specificity, and precision	High	Low
Florez et al [72]; 2021	Transformers	Next-visit diagnosis	No	No	Recall, precision, and AUC	High	Unclear
Pham et al [73]; 2021	Transformers	Cardiac complication risk	Demographic data	Yes (as day)	AUROC	High	Low
Boursalie et al [74]; 2021	Transformers	Next-visit diagnosis	Demographic data, medicine, and treatment	Yes	Precision and recall	High	Unclear
Dong et al [75]; 2021	LSTM	Opioid use disorder	Procedures, laboratory test results, medications, clinical events, and demographic data	No	AUROC, precision, recall, and F_1-score	Low	Low
Kwak et al [76]; 2021	GRU	CVDs	Medication and demographic data	No	AUROC and AUPRC	Low	Low
Sun et al [77]; 2021	GRU	Mortality, readmission, sepsis, and heart failure	No	Yes	AUROC, AUPRC, and accuracy	High	Low
Men et al [78]; 2021	LSTM	Next-visit diagnosis	Disease types and demographic data	Yes	AUROC, precision, recall, and F_1-score	High	Low
Shi et al [79]; 2021	CNN	Next-visit diagnosis	Medications	No	Accuracy	High	Unclear

XSL•FO
RenderX

Study; year	DL approach	Prediction task	Additional features	Included time	Performance evaluation	ROB	COA
Lu et al [80]; 2021	Multilayer perceptron	Next-visit diagnosis and heart failure	No	No	AUROC and F_1-score	High	Low
Peng et al [81]; 2021	Transformers	Next-visit diagnosis	No	Yes	Accuracy	High	Unclear
An et al [82]; 2021	Bi-LSTM-CNN[o]	CVDs	Medications, laboratory test results, and examination	Yes	Recall, precision, F_1-score, and AUROC	High	Unclear
Poulain et al [83]; 2021	Transformers	Risk of CVDs	Demographic data	Yes (as age)	MSE[p]	High	Unclear
Pang et al [84]; 2021	Transformers	Heart failure, mortality, diabetes, and hospitalization	Medications, procedures, and age	Yes (as month)	AUROC and AUPRC	Unclear	Low
Rao et al [85]; 2022	Transformers	Heart failure	Medications, age, and calendar year	Yes (as year)	AUROC and AUPRC	Low	Low
Du et al [86]; 2022	LSTM	Mortality	No	No	Accuracy, AUROC, and F_1-score	High	Low
De Barros and Rodrigues [87]; 2022	LSTM	Next-visit diagnosis	No	No	Recall, precision, AUROC, and F_1-score	High	Unclear
Liu et al [88]; 2022	Transformers	Next-visit diagnosis and mortality	Medications, laboratory test results, clinical events, and demographic data	Yes	AUROC, AUPRC, and recall	High	Low
Yang et al [89]; 2022	LSTM	Mortality	Admission type	Yes (as day)	AUC, precision, recall, and F_1-score	High	Low
Chen et al [90]; 2022	Transformers	Next-visit diagnosis and new onset of disease	Procedures	Yes	Precision and recall	Low	Low
Sun et al [91]; 2022	GRU	Next-visit diagnosis	No	No	F_1-score and recall	High	Unclear
Yu et al [92]; 2022	Logical perception	Next-visit diagnosis and mortality	Medications and procedures	No	Accuracy, precision, recall, and F_1-score	High	Low
Niu et al [93]; 2022	GRU	Mortality	Demographic data, procedures, medications, and vital signs	No	AUROC, F_1-score, precision, sensitivity, and specificity	Unclear	Unclear
AlSaad et al [94]; 2022	RNN	Emergency visit	No	No	AUROC, AUPRC, and F_1-score	Low	Low
Gerrard et al [95]; 2022	Transformers	Next-visit diagnosis and readmission	No	No	AUROC and F_1-score	High	Low
AlSaad et al [96]; 2022	RNN	Preterm birth	Procedures, medications, and laboratory test results	No	AUROC, AUPRC, sensitivity, and specificity	Unclear	Low
Ramchand et al [97]; 2022	RNN	Hospitalization for COVID-19 infection	Demographic data	Yes	AUROC, F_1-score, sensitivity, and specificity	High	Low
Andjelkovic et al [98]; 2022 [98]	LSTM and GRU	Lung cancer, breast cancer, cervix uteri cancer, and liver cancer	No	No	Accuracy, AUROC, recall, specificity, precision, and F_1-score	High	Low
Yu et al [99]; 2022	LSTM	Mortality	Medications and procedures	No	Precision, recall, AUROC, accuracy, and F_1-score	High	Unclear

XSL•FO
RenderX

Study; year	DL approach	Prediction task	Additional features	Included time	Performance evaluation	ROB	COA
Li et al [100]; 2022	Transformers	Heart failure, stroke, and coronary heart disease	Age	Yes (as age)	AUROC and precision	Low	Low
Li et al [101]; 2023	Transformers	Heart failure, diabetes, chronic kidney disease, and stroke	Medications, procedures, laboratory test results, blood pressure, drinking status, smoking status, and BMI	No	AUROC and AUPRC	Low	Low
Dong et al [102]; 2023	LSTM	Opioid overdose	Medications, laboratory test results, clinical events, and demographic data	No	precision, recall, F_1-score, and AUROC	Low	Low
Guo et al [103]; 2023	Transformers	Mortality, long length of stay, readmission, and ICU[q] admission	Demographic data, laboratory test results, procedures, and medications	Yes	AUROC and AUPRC	Unclear	Unclear
Liang and Guo [104]; 2023	Transformers	Heart failure	Demographic data, laboratory test results, procedures, and medications	No	Accuracy, AUROC, and F_1-score	High	Low
Lee et al [105]; 2023	CNN	Psoriatic arthritis	Medications	Yes	AUROC, sensitivity, specificity, PPV[r], and NPV[s]	High	Low

[a]AUROC: area under the receiver operating characteristic curve.

[b]CNN: convolutional neural network.

[c]GRU: gated recurrent unit.

[d]LSTM: long short-term memory.

[e]RNN: recurrent neural network.

[f]MCC: Matthews correlation coefficient.

[g]COPD: chronic obstructive pulmonary disease.

[h]AUC: area under the curve.

[i]CVD: cardiovascular disease.

[j]AUPRC: area under the precision-recall curve.

[k]GNN: graph neural network.

[l]MCI: mild cognitive impairment.

[m]RNN-DAE: recurrent neural network–based denoising autoencoder.

[n]OA: osteoarthritis.

[o]Bi-LSTM-CNN: bidirectional long short-term memory–convolutional neural network.

[p]MSE: mean squared error.

[q]ICU: intensive care unit.

[r]PPV: positive predictive value.

[s]NPV: negative predictive value.

DL Techniques

We analyzed 84 DL models from the included studies. Among these 84 models, the most commonly applied DL technique for learning sequential diagnosis codes was RNNs and their derivatives (n=47, 56%), followed by transformers (n=22, 26%), which have been regularly applied in studies since their introduction in 2017 (Figure 2). Among the 38 studies that used embedding techniques to represent diagnostic data, the most frequently used embedding method was Word2Vec (n=15, 39%), followed by GNNs (n=9, 24%) and transformers (n=3, 8%). More than half of the 84 studies presented their input feature as a sequence of visit embeddings (n=45, 54%), followed by a sequence of diagnosis codes (n=24, 29%), a sparse matrix (n=10, 12%), and a mixed representation (n=5, 6%).

XSL•FO
RenderX

Figure 2. The publication pattern in terms of (A) the number of deep learning models, (B) embedding techniques, and (C) explainability. The thin bar representing 2023 reflects the partial year of data because the search only extended to May 2023. CNN: convolutional neural network; GRU: gated recurrent unit; PV-DBOW: paragraph vector–distributed bag of words; RNN: recurrent neural network.

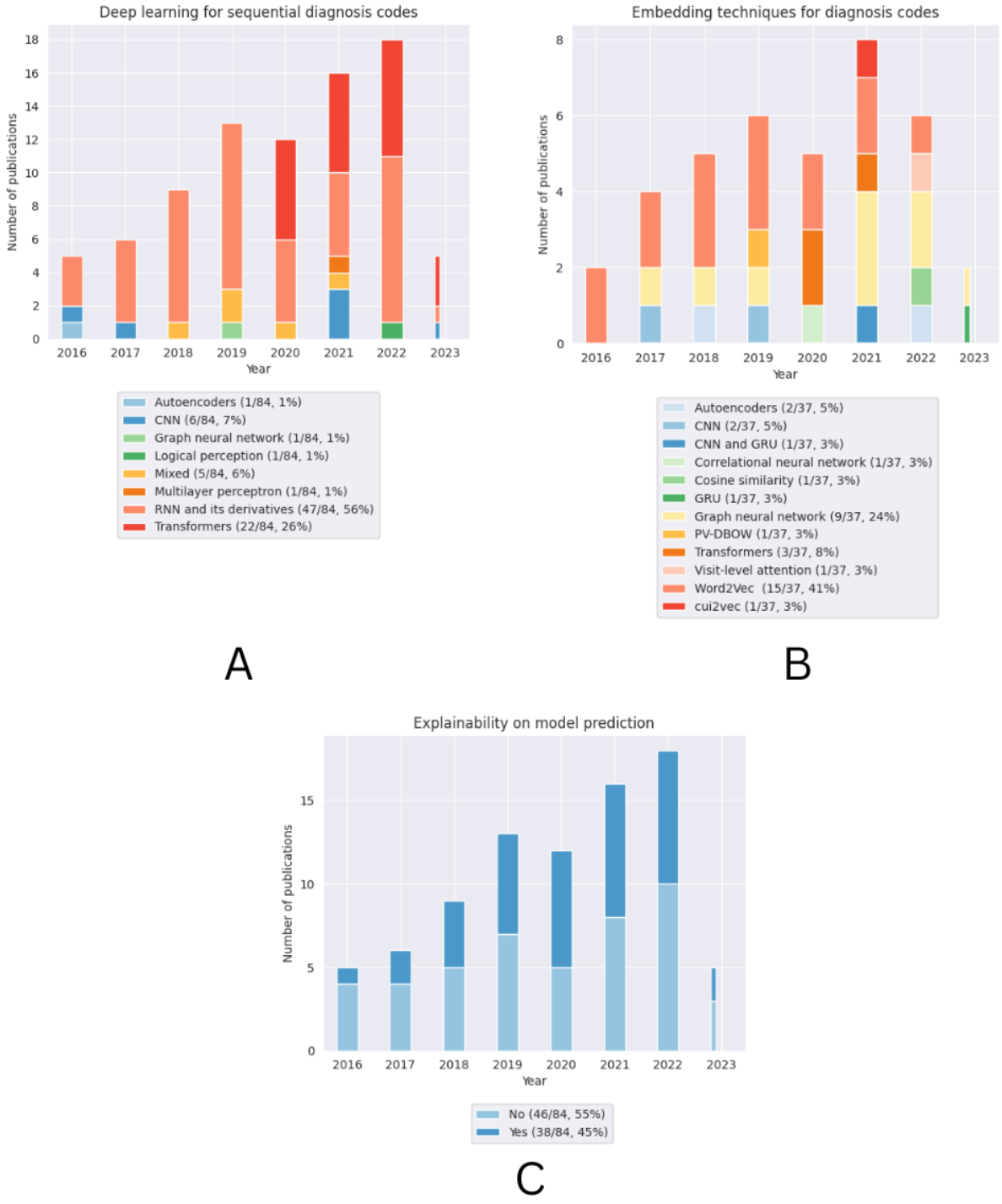

Dataset Characteristics

Across 125 training datasets, the median sample size was 29,256 (IQR 92,572; range 1095-5,231,614) patients. The most frequently used datasets (82/125, 65.6%) originated from the United States. Moreover, there is an increase in sample size for training models (Figure S1 in Multimedia Appendix 3). The publicly available MIMIC-III dataset (31/125, 24.8%) was the most popular, followed by the Clinical Practice Research Datalink (11/125, 8.8%), Cerner Health Facts (8/125, 6.4%), Sutter Health (5/125, 4%), and MIMIC-IV (3/125, 2.4%; Figure S2 in Multimedia Appendix 3). The most frequently used coding system was *International Classification of Diseases, Ninth Revision* (82/125, 65.6%), followed by *ICD-10* (30/125, 24%). The most frequently incorporated additional feature in the 84 studies was medications (n=38, 45%), followed by demographic data (n=33, 39%), procedures (n=29, 35%), laboratory test results (n=15, 18%), and clinical events (n=4, 5%; Figure S3 in Multimedia Appendix 3). Some studies (35/84, 42%)

XSL•FO
RenderX

integrated time information into their DL models for understanding patient prognosis.

Prediction Tasks

The highest frequency of predicted outcome variables were patient trajectory (n=51, 39.8%), cardiovascular disease and risks (n=33, 25.8%), admission (n=15, 11.7%), neurological diseases (n=6, 4.7%), and malignancy (n=6; 4.7%). The main subgroups of prediction tasks were next-visit diagnosis (n=30, 23.4%), heart failure (n=18, 14.1%), and mortality (n=17, 13.3%). Figure 3 presents the chord diagram illustrating the relationships between features and outcome predictions. The widest band is the connection between medications and next-visit diagnosis (10/263, 3.8%).

Figure 3. Chord diagram showing the relationship between features (red) and predictions (blue), derived from the "Additional features" and "Prediction task" columns in Table 1, respectively. In the chord diagram, "No" represents the absence of any features; "Visit" refers to the number of times a patient visited a health care provider; "Human-derived features" refers to features extracted from human input, such as manually recorded clinical observations or patient-reported outcomes; "Measurement" represents objective quantifications, such as laboratory test results, vital signs, and other instrument-based evaluations; "Medications" refers to prescribed drugs; "Medicine/treatment" is a broader term that includes medications and surgical procedures; "Calendar year" denotes the year of a patient's clinic visit; and "Conditions" refers to medical conditions.

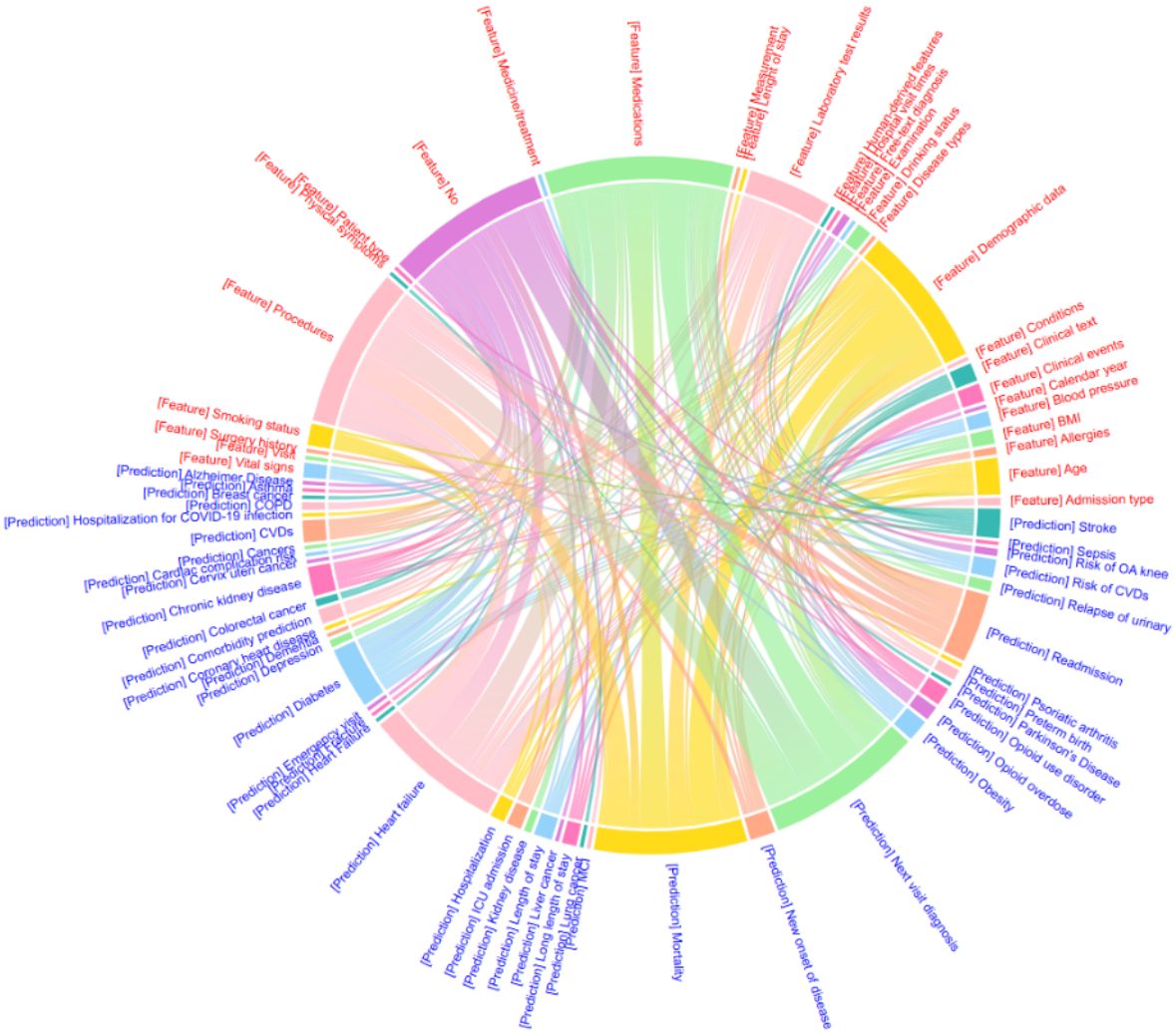

Performance Evaluation

A variety of model performance metrics were reported across the 84 included studies. The best DL model performance in each study was reported using AUROC (41/84, 49%), F_1-score (25/84, 30%), area under the precision-recall curve (13/84, 15%), precision (16/84, 19%), and recall (14/84, 17%). The relationship between sample size, the number of features, and AUROC was examined (Figure S4 in Multimedia Appendix 3). A statistically significant relationship was found between sample size and AUROC (P=.02), indicating that changes in sample size have a notable impact on AUROC. However, there was no statistically significant relationship between the number of features and AUROC.

Generalizability and Explainability

An assessment of generalizability with external validation was uncommon among the included studies. Overall, only 7 (8%) of the 84 studies evaluated generalizability across ≥1 of the following categories: demographic validation (n=3, 43%)

XSL•FO
RenderX

[39,48,55], temporal validation (n=2, 29%) [100,103], and geographic validation (n=4, 57%) [24,39,68,100]. Regarding explainability, less than half of the studies (38/84, 45%) incorporated a mechanism to interpret their predictions. The publication pattern in terms of explainability is shown in Figure 2.

ROB and Concern of Applicability

Overall, the included studies had a high ROB (59/84, 70%), which was mainly driven by high ROB in the analysis domain (53/84, 63%). The main reason for high ROB in the analysis domain was the imbalance between the number of patients and their outcomes. Our assessment found low ROB in the participant (66/84, 79%) and predictor (81/84, 96%) domains due to broad inclusion criteria in general and similar predictor definitions across all participants, respectively. Within the applicability assessment, 3 domains were evaluated: participants, predictors, and outcomes. Overall, 45% (38/84) of the included studies showed either unclear or high concern of applicability because the characteristics of most included studies were not clearly mentioned. The full results of the ROB and applicability assessment are shown in Figure 4.

Figure 4. The Prediction Model Study Risk of Bias Assessment Tool (PROBAST) results: (A) risk of bias (ROB) and (B) concern of applicability (COA).

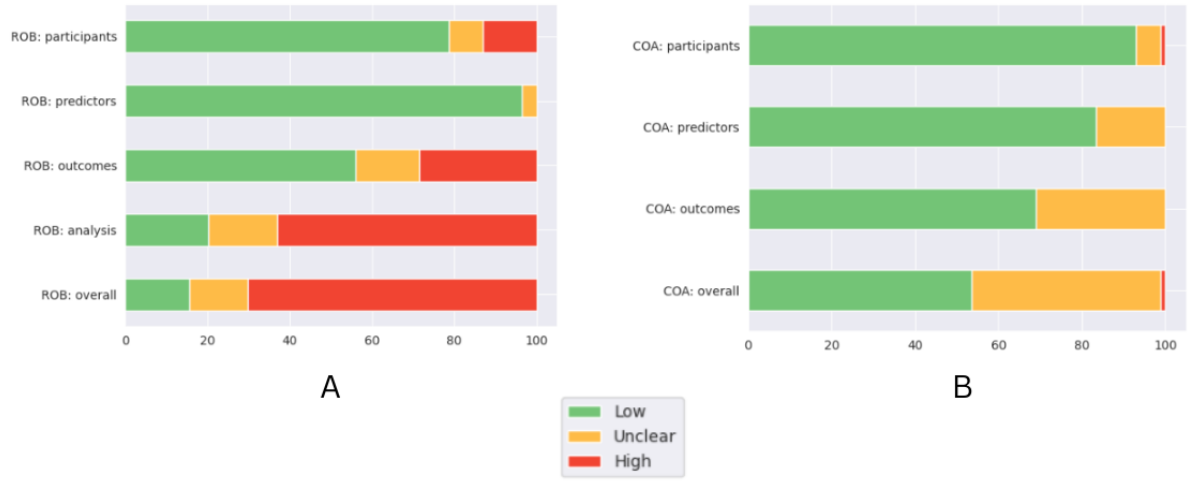

Discussion

Model Architecture

This systematic review offers insights into the contemporary DL approaches used to model patients' diagnostic history to predict outcomes. We explored 84 studies that met our inclusion criteria in this research area. The application of DL for advanced modeling of sequential medical codes is a rapidly growing research area. This is evident from the increasing number of publications found each year in this review. Considering the escalating interest in DL, there have been obvious publication patterns since 2016. After the emergence of bidirectional encoder representations from transformers (BERT) at the end of 2018 [106], transformer models have been increasingly used for sequential diagnosis codes, as reflected in the literature. More recently, there has been a prominent and highly successful showcase of transformers, with OpenAI's ChatGPT with GPT-3.5 and ChatGPT with GPT-4 [107] serving as prime examples of their capabilities.

Diagnosis code and language share some similar aspects. We can consider diagnosis codes as words and code sequences in a medical history as sentences. Typically, DL models designed for diagnosis codes aim to capture relationships between diagnosis codes within patient visits and across patient visits, which is similar to natural language processing (NLP) approaches that learn connections between words within sentences and across different sentences. Numerous studies in our review applied recent NLP techniques, such as RNNs and transformers, to sequential diagnosis codes. The main difference between diagnosis codes within patient visits and words within sentences is the irregular time interval between medical events. To accommodate this difference, many studies adjust the original DL algorithms to make them suitable for sequential diagnosis codes. Some examples include RETAIN [26], Dipole [27], EHAN [34], Timeline [35], Patient2Vec [40], DeepRisk [108], COAM [42], CLOUT [60], HAN [70], IoHAN [86], AttentionHCare [87], DeepMPM [89], and tBNA-PR [104], all of which use attention mechanisms and show superior performance compared to models without attention mechanisms.

Although RNNs and their derivatives made up the majority of the models (47/84, 56%) used for patient outcome prediction, several studies (15/84, 18%) showed that transformers have superior performance [55,56,59,62,67,72,73,81,83-85,88,90,95,104]. The transformer model consists of a multihead self-attention unit that computes in parallel. A crucial part of this architecture is the positional encoder, which enables transformers to understand the order and adjacency of information, in a similar way to CNNs and RNNs, respectively. Evidence from 2 (2%) of the 84 studies included in this review shows that positional embedding can improve model performance [55,64]. Moreover, models that combine transformers and RNNs have been introduced; for example, the medical BERT model combined with bidirectional gated recurrent unit improved the AUROC of the base models

XSL•FO
RenderX

by 1.62% to 6.14% [68]. Pretrained transformers models such as BERT for EHRs [55] and medical BERT [68] demonstrate impressive performance; however, both require a huge amount of pretraining patient data—approximately 1.6 million and 28 million samples, respectively.

Encoding diagnosis codes is a challenging task. Word2Vec is the most common method used for code-level embedding. Originally, Word2Vec was designed to capture the semantic relationships between words from large text corpora through unsupervised learning. Similarly, Word2Vec for diagnosis codes enables the encoding of meaningful relationships between medical conditions. Some models use separate training for the input data, with methods such as Word2Vec, GNNs, or transformers, while others that skip this step still require some form of embedding for codes during the training process. Consequently, it is challenging to make direct comparisons between DL models with embeddings and those without. Some studies show that learning hierarchical information through code-level embeddings can enhance the power of model prediction; for example, code-level embedding using Word2Vec can improve model performance [24]. GNNs can be a powerful tool for a hierarchical encoder with the potential to capture relationships between diagnosis codes. Many studies have demonstrated that using GNNs to embed diagnosis codes can provide effective predictions, including models such as Graph-Based Attention Model, knowledge-based attention model, Co-Attention Memory networks for diagnosis Prediction, multirelational EHR representation learning method, self-supervised graph learning framework with hyperbolic embeddings for temporal health event prediction, Sequential Diagnosis Prediction with Transformer and Ontological Representation, hypergraph-based disease prediction model using EHRs, Sequential visits and Medical Ontology, and integrated deep learning model combining long short term memory and graph neural networks. Interestingly, using multilevel representations, combining visit level and variable level, for a patient is better than single-level representation based on visit or variable alone [89].

Large language models (LLMs) have revolutionized NLP. LLMs are trained on huge amounts of text data from the internet, books, and other sources; and they can perform a wide range of language-related tasks. Recently, researchers have explored the ability of LLMs to understand medical codes; for example, a study evaluated several LLMs, including ChatGPT with GPT-3.5, ChatGPT with GPT-4, Gemini Pro, and Llama2-70b Chat, for their ability to generate correct medical codes based on code descriptions [109]. However, the study found that LLMs frequently generated incorrect codes. The findings suggest that LLMs lack an understanding of the meaning of medical codes. Therefore, it is still a challenge to use LLMs for clinical codes.

Characteristics and Features

Generally, the performance of DL models depends on the setting of the training dataset, such as inpatient or outpatient populations. Most of the included studies (31/125, 24.8%) used models trained on the MIMIC-III dataset, which focuses on ICU admissions in the United States, with a short interval between diagnosis codes and patients considered high risk [110]. Patients in the MIMIC dataset will have more severe illnesses than a general hospital population, which would be a mix of inpatients and outpatients. The type of clinical data in the training dataset is very important for robust disease prediction. The dataset should not be focused only on specific clinical settings, such as hospital admissions or ICUs, due to data availability. However, more than half of the included studies (52/92, 57%) did not clearly report the setting of their training dataset. Moreover, a study showed that patient vital signs had a greater influence on mortality prediction in ICU settings than diagnosis codes [93]; yet, vital signs are not available in every dataset. Therefore, we suggest that it is important to provide the details of the clinical setting of the training dataset.

There may be inherent biases in how diagnosis codes are recorded in EHR data, influenced by their primary billing sources. In addition, for a single appointment, acute conditions are more likely to be recorded than chronic diseases. Any biases in the recording of diagnosis codes may transfer to the learning process of a DL algorithm. This potential transfer of bias should be carefully considered when interpreting the results. Most of the included studies (42/125, 33.6%) used datasets, such as MIMIC-III and Clinical Practice Research Datalink, that provide long-term follow-up data for large patient populations. These datasets typically include diagnosis codes for a wide range of acute and chronic medical conditions. DL studies can benefit from the availability of longer sequences of these codes, reflecting longer periods of the patient's medical history and therefore more likely to capture both conditions.

The performance of the DL model in predicting patient outcomes depends not only on diagnosis codes but also on other features, such as demographic data [43], treatments [40,56,71], the number of visits [56], and the interval between visits [64,88]. Several studies have shown that diagnosis codes alone cannot provide the best predictive performance compared to models incorporating multiple features, such as medications, procedures, laboratory test results, demographic data, and so on [82,88,101,108]. Moreover, a study that applied BERT for EHRs has highlighted that combining diagnosis codes with factors such as age, segment, and position lead to improved precision scores compared to relying solely on diagnosis codes [55]. We propose that integrating a wide array of supplementary features, including demographic data, medication records, medical procedures, and laboratory test results, can potentially improve the performance of DL models when analyzing sequential diagnosis codes.

Many of the studies (33/84, 39%) included basic demographic data, such as gender and age, as model features. Ethnicity is another important demographic factor because it is a significant risk factor for many diseases, such as heart disease [111]. However, only a few studies (4/84, 5%) specifically described ethnicity as an input feature of their models. Most of the datasets reviewed (63/84, 75%) originated from the United States and European countries. In addition, research has shown that race and ethnicity data in EHRs are often incomplete and inaccurate, especially for minority populations [112]. These limitations can make it challenging to generalize the developed models to external settings with different patient populations and health care systems.

XSL•FO
RenderX

As diagnosis codes occur at irregular time intervals, adding time intervals as a feature can help a model to understand disease progression. Hypothetically, patients with shorter follow-up times are more likely to have severe conditions. Ablation studies, which assess model performance by removing some features, have demonstrated that integrating time intervals between patient visits can enhance predictive power. This improvement has been demonstrated across multiple models, including BiteNet [56], CATNet [88], CEHR-BERT [84], Deepr [23], DL-BERT [90], EHAN [34], and SETOR [81]. Moreover, research applying HiTANet has shown that integrating time intervals between the last visit and current visit can improve model performance [58]. A study that examined the impact of integrating time intervals as days, weeks, and months into model predictions found that using weeks as the unit yielded the best prediction performance, corresponding to the weekly follow-up pattern in real clinical practice [78]. We believe that the time interval between visits can serve as an indicator of disease progression.

Next-visit prediction is a widely applied task for evaluating DL performance in sequential diagnosis codes. However, this task carries an ROB; for example, while DL-BERT performs well in predicting the next disease based on a patient's existing history, its precision drops—by approximately 50%—when predicting a new-onset disease that has not previously occurred in the patient's history. This indicates that the model relies heavily on prior diagnoses rather than capturing underlying disease progression [90]. Another challenge for next-visit prediction is the issue of missing data. Sometimes, patients seek care at a different location or with a different provider, and this information may not be captured in the available data. This is a major limitation for most studies, which can be mitigated through better data linkage and information sharing across health providers.

Evaluation, Generalizability, and Explainability

Model evaluation in health care is often complicated by class imbalance due to the nature of the medical domain, where the number of individuals with a disease is usually lower than the number of individuals without it. AUROC and F_1-score are commonly used evaluation metrics in publications for health care research area and were reported by 82% (69/84) of the studies included in this review. However, F_1-score focuses on positive prediction and avoid the value of true negatives. Alternatively, a study suggested using the Matthews correlation coefficient for model evaluation because it provides a more balanced assessment of positive and negative predictions [113].

Generalizability is an important issue for outcome prediction in health care. When a model is trained on a specific population, it will only perform well on patients with similar characteristics. The evaluation should include both internal and external validation. Assessing performance on different data distributions is crucial for outcome prediction [19,114]. Predictive performance degrades when models are tested on out-of-distribution years [103]; however, only a small proportion of the included studies (7/84, 8%) validated their models using data outside of their training distribution. Validating models across different settings—demographic, temporal, and geographic—is very important for real-world applications.

While DL models can yield a good performance in outcome prediction, they are frequently regarded as black box models, lacking the explainability needed to understand their internal processes and the main contributing factor for a given prediction. This lack of explainability—when the underlying reasoning is unclear—is a significant concern because it may reduce trust in predictions among health care workers and complicate the explanation of clinical decisions to patients. In recent years, researchers have focused on explainability. Shapley additive explanations (SHAP) was introduced to understand model predictions [115]. SHAP can be used to explain the decisions of LSTM models and transformers; for instance, SHAP values can explain meaningful medical codes to predict the risk of opioid overdose with 11 codes related to medications and 2 codes related to mental disorders [102]. A study introduced a tool to visualize multihead self-attention in transformer models, which can show the relationship between 2 sentences [116]. Some studies used this tool to show disease relationship between 2 sequences of diagnosis codes [55,67]. We propose that explainable artificial intelligence should be included in research studies in this area.

Limitations

This review has several limitations. First, while our search was effective in capturing a significant proportion of the papers in this research area, it is important to note that variations in search terms may have led to the unintentional exclusion of other relevant studies, which is why we conducted a manual search to identify additional relevant articles. Second, the included studies exhibited heterogeneity, which made it difficult to compare the studies and conduct a meta-analysis. There was variability in input features and prediction tasks. Moreover, model evaluation metrics were reported differently. Although many of the included studies (57/84, 68%) introduced single, novel DL-based models, some studies (27/84, 32%) applied multiple standard models for outcome prediction. In such cases, we considered only the best-performing DL model in each study. Finally, it is important to note that PROBAST was not originally designed for evaluating DL models, and certain questions in the analysis domain are not suitable for these models.

Recommendations and Future Work

On the basis of the findings of this review, our main recommendation for future studies using DL alongside sequential diagnostic patient information is to ensure that the generalizability of the developed models is tested on independent datasets. This will significantly reduce the ROB in key findings. Given the aforementioned limitations, we also think that studies should apply a more specific ROB assessment tool to DL models because not all questions in the current PROBAST assessment framework are suitable.

Conclusions

The application of DL for advanced modeling of sequential medical codes has demonstrated remarkable promise in predicting patient outcomes. The main limitation of this study was the heterogeneity of methods and outcomes. However, our

XSL•FO
RenderX

analysis found that using multiple types of features, integrating time intervals, and including larger sample sizes were generally related to an improved predictive performance in the included studies. This review also highlights that very few studies (7/84, 8%) reported on the challenges related to generalizability and almost half (38/84, 45%) of the studies reported challenges related to explainability of DL models. Addressing these shortcomings will be instrumental in unlocking the full potential of DL for enhancing health care outcomes and patient care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 129 KB - jmir_v27i1e57358_app1.pdf]

Multimedia Appendix 2
Additional details of included studies.
[XLSX File (Microsoft Excel File), 82 KB - jmir_v27i1e57358_app2.xlsx]

Multimedia Appendix 3
Additional figures.
[DOCX File , 149 KB - jmir_v27i1e57358_app3.docx]

References

1. Huda A, Castaño A, Niyogi A, Schumacher J, Stewart M, Bruno M, et al. A machine learning model for identifying patients at risk for wild-type transthyretin amyloid cardiomyopathy. Nat Commun 2021 May 11;12(1):2725 [FREE Full text] [doi: 10.1038/s41467-021-22876-9] [Medline: 33976166]
2. Johnson AE, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 03;10(1):1 [FREE Full text] [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
3. Pivovarov R, Albers DJ, Sepulveda JL, Elhadad N. Identifying and mitigating biases in EHR laboratory tests. J Biomed Inform 2014 Oct;51:24-34 [FREE Full text] [doi: 10.1016/j.jbi.2014.03.016] [Medline: 24727481]
4. Hripcsak G, Albers DJ, Perotte A. Parameterizing time in electronic health record studies. J Am Med Inform Assoc 2015 Jul;22(4):794-804 [FREE Full text] [doi: 10.1093/jamia/ocu051] [Medline: 25725004]
5. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ 2018 Apr 30;361:k1479 [FREE Full text] [doi: 10.1136/bmj.k1479] [Medline: 29712648]
6. Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, et al. Deep learning in clinical natural language processing: a methodical review. J Am Med Inform Assoc 2020 Mar 01;27(3):457-470 [FREE Full text] [doi: 10.1093/jamia/ocz200] [Medline: 31794016]
7. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, et al. A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. Proc IEEE Inst Electr Electron Eng 2021 May;109(5):820-838 [FREE Full text] [doi: 10.1109/JPROC.2021.3054390] [Medline: 37786449]
8. Rim B, Sung NJ, Min S, Hong M. Deep learning in physiological signal data: a survey. Sensors (Basel) 2020 Feb 11;20(4):969 [FREE Full text] [doi: 10.3390/s20040969] [Medline: 32054042]
9. Xiao Y, Wu J, Lin Z, Zhao X. A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data. Comput Methods Programs Biomed 2018 Nov;166:99-105. [doi: 10.1016/j.cmpb.2018.10.004] [Medline: 30415723]
10. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: a review of methods and applications. AI Open 2020;1:57-81. [doi: 10.1016/j.aiopen.2021.01.001]
11. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit Health 2020 Apr;2(4):e179-e191 [FREE Full text] [doi: 10.1016/S2589-7500(20)30018-2] [Medline: 33328078]
12. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2021;2(6):420 [FREE Full text] [doi: 10.1007/s42979-021-00815-1] [Medline: 34426802]
13. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]

XSL•FO
RenderX

14. Si Y, Du J, Li Z, Jiang X, Miller T, Wang F, et al. Deep representation learning of patient data from Electronic Health Records (EHR): a systematic review. J Biomed Inform 2021 Mar;115:103671 [FREE Full text] [doi: 10.1016/j.jbi.2020.103671] [Medline: 33387683]
15. Xie F, Yuan H, Ning Y, Ong ME, Feng M, Hsu W, et al. Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies. J Biomed Inform 2022 Feb;126:103980 [FREE Full text] [doi: 10.1016/j.jbi.2021.103980] [Medline: 34974189]
16. Carrasco-Ribelles LA, Llanes-Jurado J, Gallego-Moll C, Cabrera-Bean M, Monteagudo-Zaragoza M, Violán C, et al. Prediction models using artificial intelligence and longitudinal data from electronic health records: a systematic methodological review. J Am Med Inform Assoc 2023 Nov 17;30(12):2072-2082 [FREE Full text] [doi: 10.1093/jamia/ocad168] [Medline: 37659105]
17. Miller SA, Forrest JL. Enhancing your practice through evidence-based decision making: PICO, learning how to ask good questions. J Evid Based Dent Pract 2001 Oct;1(2):136-141. [doi: 10.1016/s1532-3382(01)70024-3]
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. de Hond AA, Shah VB, Kant IM, Van Calster B, Steyerberg EW, Hernandez-Boussard T. Perspectives on validation of clinical predictive algorithms. NPJ Digit Med 2023 May 06;6(1):86 [FREE Full text] [doi: 10.1038/s41746-023-00832-9] [Medline: 37149704]
20. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan 01;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
22. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
23. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: a convolutional net for medical records. arXiv Preprint posted online on July 26, 2016 [FREE Full text] [doi: 10.1109/jbhi.2016.2633963]
24. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
25. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: a deep dynamic memory model for predictive medicine. arXiv Preprint posted online on February 1, 2016 [FREE Full text] [doi: 10.1007/978-3-319-31750-2_3]
26. Choi E, Bahadori MT, Kulas JA, Schuetz A, Stewart WF, Sun J. RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. arXiv Preprint posted online on August 19, 2016 [FREE Full text]
27. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. arXiv Preprint posted online on June 19, 2017 [FREE Full text] [doi: 10.1145/3097983.3098088]
28. Choi E, Bahadori MT, Song L, Stewart WF, Sun J. GRAM: graph-based attention model for healthcare representation learning. KDD 2017 Aug;2017:787-795 [FREE Full text] [doi: 10.1145/3097983.3098126] [Medline: 33717639]
29. Sha Y, Wang MD. Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. ACM BCB 2017 Aug;2017:233-240. [doi: 10.1145/3107411.3107445] [Medline: 32577628]
30. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2017 Mar 01;24(2):361-370 [FREE Full text] [doi: 10.1093/jamia/ocw112] [Medline: 27521897]
31. Suo Q, Ma F, Yuan Y, Huai M, Zhong W, Zhang A, et al. Personalized disease prediction using a CNN-based similarity learning method. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2017 Presented at: BIBM 2017; November 13-16, 2017; Kansas City, MO. [doi: 10.1109/bibm.2017.8217759]
32. Amirkhan R, Hoogendoorn M, Numans ME, Moons L. Using recurrent neural networks to predict colorectal cancer among patients. In: Proceedings of the IEEE Symposium Series on Computational Intelligence. 2017 Presented at: SSCI 2017; November 27-December 1, 2017; Honolulu, HI. [doi: 10.1109/ssci.2017.8280826]
33. Lei L, Zhou Y, Zhai J, Zhang L, Fang Z, He P, et al. An effective patient representation learning for time-series prediction tasks based on EHRs. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2018 Presented at: BIBM 2018; December 3-6, 2018; Madrid, Spain. [doi: 10.1109/bibm.2018.8621542]
34. Park S, Kim YJ, Kim JW, Park JJ, Ryu B, Ha JW. Interpretable prediction of vascular diseases from electronic health records via deep attention networks. In: Proceedings of the IEEE 18th International Conference on Bioinformatics and Bioengineering. 2018 Presented at: BIBE 2018; October 29-31, 2018; Taichung, Taiwan. [doi: 10.1109/bibe.2018.00028]
35. Bai T, Egleston BL, Zhang S, Vucetic S. Interpretable representation learning for healthcare via capturing disease progression through time. KDD 2018 Aug;2018:43-51 [FREE Full text] [doi: 10.1145/3219819.3219904] [Medline: 31037221]
36. Choi E, Xiao C, Stewart WF, Sun J. MiME: multilevel medical embedding of electronic health records for predictive healthcare. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018 Presented at: NIPS'18; December 3-8, 2018; Montreal, QC URL: https://proceedings.neurips.cc/paper/2018/hash/934b535800b1cba8f96a5d72f72f1611-Abstract.html

XSL•FO
RenderX

37. Qiao Z, Zhao S, Xiao C, Li X, Qin Y, Wang F. Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018 Presented at: IJCAI'18; July 13-19, 2018; Stockholm, Sweden. [doi: 10.24963/ijcai.2018/489]
38. Wang WW, Li H, Cui L, Hong X, Yan Z. Predicting clinical visits using recurrent neural networks and demographic information. In: Proceedings of the IEEE 22nd International Conference on Computer Supported Cooperative Work in Design. 2018 Presented at: CSCWD 2018; May 9-11, 2018; Nanjing, China. [doi: 10.1109/cscwd.2018.8465194]
39. Ma F, You Q, Xiao H, Chitta R, Zhou J, Gao J. KAME: knowledge-based attention model for diagnosis prediction in healthcare. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018 Presented at: CIKM '18; October 22-26, 2018; Torino, Italy. [doi: 10.1145/3269206.3271701]
40. Zhang J, Kowsari K, Harrison JH, Lobo JM, Barnes LE. Patient2Vec: a personalized interpretable deep representation of the longitudinal electronic health record. IEEE Access 2018;6:65333-65346. [doi: 10.1109/access.2018.2875677]
41. Jin B, Che C, Liu Z, Zhang S, Yin X, Wei X. Predicting the risk of heart failure with EHR sequential data modeling. IEEE Access 2018;6:9256-9261. [doi: 10.1109/access.2017.2789324]
42. Guo W, Ge W, Cui L, Li H, Kong L. An interpretable disease onset predictive model using crossover attention mechanism from electronic health records. IEEE Access 2019;7:134236-134244. [doi: 10.1109/access.2019.2928579]
43. Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One 2019;14(7):e0218942 [FREE Full text] [doi: 10.1371/journal.pone.0218942] [Medline: 31283759]
44. Wang W, Guo C, Xu J, Liu A. Bi-dimensional representation of patients for diagnosis prediction. In: Proceedings of the IEEE 43rd Annual Computer Software and Applications Conference. 2019 Presented at: COMPSAC 2019; July 15-19, 2019; Milwaukee, WI. [doi: 10.1109/compsac.2019.10235]
45. Gao J, Wang X, Wang Y, Yang Z, Gao J, Wang J, et al. CAMP: co-attention memory networks for diagnosis prediction in healthcare. In: Proceedings of the 2019 IEEE International Conference on Data Mining. 2019 Presented at: ICDM 2019; November 08-11, 2019; Beijing, China. [doi: 10.1109/icdm.2019.00120]
46. Ma F, Wang Y, Xiao H, Yuan Y, Chitta R, Zhou J, et al. Incorporating medical code descriptions for diagnosis prediction in healthcare. BMC Med Inform Decis Mak 2019 Dec 19;19(Suppl 6):267 [FREE Full text] [doi: 10.1186/s12911-019-0961-2] [Medline: 31856806]
47. AlSaad R, Malluhi Q, Janahi I, Boughorbel S. Interpreting patient-specific risk prediction using contextual decomposition of BiLSTMs: application to children with asthma. BMC Med Inform Decis Mak 2019 Nov 08;19(1):214 [FREE Full text] [doi: 10.1186/s12911-019-0951-4] [Medline: 31703676]
48. Zhang XS, Tang F, Dodge HH, Zhou J, Wang F. MetaPred: meta-learning for clinical risk prediction with limited patient electronic health records. KDD 2019 Aug;2019:2487-2495 [FREE Full text] [doi: 10.1145/3292500.3330779] [Medline: 33859865]
49. Ashfaq A, Sant'Anna A, Lingman M, Nowaczyk S. Readmission prediction using deep learning on electronic health records. J Biomed Inform 2019 Sep;97:103256 [FREE Full text] [doi: 10.1016/j.jbi.2019.103256] [Medline: 31351136]
50. Ruan T, Lei L, Zhou Y, Zhai J, Zhang L, He P, et al. Representation learning for clinical time series prediction tasks in electronic health records. BMC Med Inform Decis Mak 2019 Dec 17;19(Suppl 8):259 [FREE Full text] [doi: 10.1186/s12911-019-0985-7] [Medline: 31842854]
51. Huang Y, Yang X, Xu C. Time-guided high-order attention model of longitudinal heterogeneous healthcare data. arXiv Preprint posted online on November 28, 2019 [FREE Full text] [doi: 10.1007/978-3-030-29908-8_5]
52. Xiang Y, Xu J, Si Y, Li Z, Rasmy L, Zhou Y, et al. Time-sensitive clinical concept embeddings learned from large electronic health records. BMC Med Inform Decis Mak 2019 Apr 09;19(Suppl 2):58 [FREE Full text] [doi: 10.1186/s12911-019-0766-3] [Medline: 30961579]
53. Gupta M, Phan TL, Bunnell T, Beheshti R. Obesity prediction with EHR data: a deep learning approach with interpretable elements. arXiv Preprint posted online on December 5, 2019 [FREE Full text] [doi: 10.1145/3506719]
54. Shi P, Hou F, Zheng X, Yuan F. Analysis of electronic health records based on long short‐term memory. Concurr Comput 2020 Jul 25;32(14):e5684 [FREE Full text] [doi: 10.1002/cpe.5684]
55. Li Y, Rao S, Solares JR, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: transformer for electronic health records. Sci Rep 2020 Apr 28;10(1):7155 [FREE Full text] [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
56. Peng X, Long G, Shen T, Wang S, Jiang J, Zhang C. BiteNet: bidirectional temporal encoder network to predict medical outcomes. In: Proceedings of the 2020 IEEE International Conference on Data Mining. 2020 Presented at: ICDM 2020; November 17-20, 2020; Sorrento, Italy. [doi: 10.1109/icdm50108.2020.00050]
57. Almog YA, Rai A, Zhang P, Moulaison A, Powell R, Mishra A, et al. Deep learning with electronic health records for short-term fracture risk identification: crystal bone algorithm development and validation. J Med Internet Res 2020 Oct 16;22(10):e22550 [FREE Full text] [doi: 10.2196/22550] [Medline: 32956069]
58. Luo J, Ye M, Xiao C, Ma F. HiTANet: hierarchical time-aware attention networks for risk prediction on electronic health records. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020 Presented at: KDD '20; July 6-10, 2020; Virtual Event, CA. [doi: 10.1145/3394486.3403107]

XSL•FO
RenderX

59. Zhang X, Qian B, Cao S, Li Y, Chen H, Zheng Y, et al. INPREM: an interpretable and trustworthy predictive model for healthcare. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020 Presented at: KDD '20; July 6-10, 2020; Virtual Event. [doi: 10.1145/3394486.3403087]
60. Rongali S, Rose AJ, McManus DD, Bajracharya AS, Kapoor A, Granillo E, et al. Learning latent space representations to predict patient outcomes: model development and validation. J Med Internet Res 2020 Mar 23;22(3):e16374 [FREE Full text] [doi: 10.2196/16374] [Medline: 32202503]
61. Ye M, Luo J, Xiao C, Ma F. LSAN: modeling long-term dependencies and short-term correlations with hierarchical attention for risk prediction. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020 Presented at: CIKM '20; October 19-23, 2020; Virtual Event, Ireland. [doi: 10.1145/3340531.3411864]
62. Zeng X, Feng Y, Moosavinasab S, Lin D, Lin S, Liu C. Multilevel self-attention model and its use on medical risk prediction. Pac Symp Biocomput 2020;25:115-126 [FREE Full text] [Medline: 31797591]
63. An Y, Mao Y, Zhang L, Jin B, Xiao K, Wei X, et al. RAHM: relation augmented hierarchical multi-task learning framework for reasonable medication stocking. J Biomed Inform 2020 Aug;108:103502 [FREE Full text] [doi: 10.1016/j.jbi.2020.103502] [Medline: 32673789]
64. Kabeshova A, Yu Y, Lukacs B, Bacry E, Gaïffas S. ZiMM: a deep learning model for long term and blurry relapses with non-clinical claims data. J Biomed Inform 2020 Oct;110:103531 [FREE Full text] [doi: 10.1016/j.jbi.2020.103531] [Medline: 32818667]
65. Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. TAPER: Time-Aware Patient EHR Representation. IEEE J Biomed Health Inform 2020 Nov;24(11):3268-3275. [doi: 10.1109/JBHI.2020.2984931] [Medline: 32287023]
66. An Y, Huang N, Chen X, Wu F, Wang J. High-risk prediction of cardiovascular diseases via attention-based deep neural networks. IEEE/ACM Trans Comput Biol Bioinform 2021;18(3):1093-1105. [doi: 10.1109/TCBB.2019.2935059] [Medline: 31425047]
67. Meng Y, Speier W, Ong MK, Arnold CW. Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J Biomed Health Inform 2021 Aug;25(8):3121-3129 [FREE Full text] [doi: 10.1109/JBHI.2021.3063721] [Medline: 33661740]
68. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med 2021 May 20;4(1):86 [FREE Full text] [doi: 10.1038/s41746-021-00455-y] [Medline: 34017034]
69. Ju R, Zhou P, Wen S, Wei W, Xue Y, Huang X, et al. 3D-CNN-SPP: a patient risk prediction system from electronic health records via 3D CNN and spatial pyramid pooling. IEEE Trans Emerg Top Comput Intell 2021 Apr;5(2):247-261. [doi: 10.1109/tetci.2019.2960474]
70. Harerimana G, Kim JW, Jang B. A deep attention model to forecast the Length Of Stay and the in-hospital mortality right on admission from ICD codes and demographic data. J Biomed Inform 2021 Jun;118:103778 [FREE Full text] [doi: 10.1016/j.jbi.2021.103778] [Medline: 33872817]
71. Ningrum DN, Kung WM, Tzeng IS, Yuan SP, Wu CC, Huang CY, et al. A deep learning model to predict knee osteoarthritis based on nonimage longitudinal medical record. J Multidiscip Healthc 2021;14:2477-2485 [FREE Full text] [doi: 10.2147/JMDH.S325179] [Medline: 34539180]
72. Florez AY, Scabora L, Eler DM, Rodrigues JF. APEHR: automated Prognosis in electronic health records using multi-head self-attention. In: Proceedings of the IEEE 34th International Symposium on Computer-Based Medical Systems. 2021 Presented at: CBMS 2021; June 07-09, 2021; Aveiro, Portugal. [doi: 10.1109/cbms52027.2021.00077]
73. Pham TH, Yin C, Mehta L, Zhang X, Zhang P. Cardiac complication risk profiling for cancer survivors via multi-view multi-task learning. In: Proceedings of the IEEE International Conference on Data Mining. 2021 Presented at: ICDM 2021; December 07-10, 2021; Auckland, New Zealand. [doi: 10.1109/icdm51629.2021.00061]
74. Boursalie O, Samavi R, Doyle TE. Decoder transformer for temporally-embedded health outcome predictions. In: Proceedings of the 20th IEEE International Conference on Machine Learning and Applications. 2021 Presented at: ICMLA 2021; December 13-16, 2021; Pasadena, CA. [doi: 10.1109/icmla52953.2021.00235]
75. Dong X, Deng J, Rashidian S, Abell-Hart K, Hou W, Rosenthal RN, et al. Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. J Am Med Inform Assoc 2021 Jul 30;28(8):1683-1693 [FREE Full text] [doi: 10.1093/jamia/ocab043] [Medline: 33930132]
76. Kwak H, Chang J, Choe B, Park S, Jung K. Interpretable disease prediction using heterogeneous patient records with self-attentive fusion encoder. J Am Med Inform Assoc 2021 Sep 18;28(10):2155-2164 [FREE Full text] [doi: 10.1093/jamia/ocab109] [Medline: 34198329]
77. Sun C, Dui H, Li H. Interpretable time-aware and co-occurrence-aware network for medical prediction. BMC Med Inform Decis Mak 2021 Nov 02;21(1):305 [FREE Full text] [doi: 10.1186/s12911-021-01662-z] [Medline: 34727940]
78. Men L, Ilk N, Tang X, Liu Y. Multi-disease prediction using LSTM recurrent neural networks. Expert Syst Appl 2021 Sep;177:114905. [doi: 10.1016/j.eswa.2021.114905]
79. Shi Y, Guo Y, Wu H, Li J, Li X. Multi-relational EHR representation learning with infusing information of diagnosis and medication. In: Proceedings of the IEEE 45th Annual Computers, Software, and Applications Conference. 2021 Presented at: COMPSAC 2021; July 12-16, 2021; Madrid, Spain. [doi: 10.1109/compsac51774.2021.00241]

XSL•FO
RenderX

80. Lu C, Reddy CK, Ning Y. Self-supervised graph learning with hyperbolic embedding for temporal health event prediction. arXiv Preprint posted online on June 9, 2021 [FREE Full text] [doi: 10.1109/tcyb.2021.3109881]
81. Peng X, Long G, Shen T, Wang S, Jiang J. Sequential diagnosis prediction with transformer and ontological representation. In: Proceedings of the IEEE International Conference on Data Mining. 2021 Presented at: ICDM 2021; December 07-10, 2021; Auckland, New Zealand. [doi: 10.1109/icdm51629.2021.00060]
82. An Y, Tang K, Wang J. Time-aware multi-type data fusion representation learning framework for risk prediction of cardiovascular diseases. IEEE/ACM Trans Comput Biol Bioinform 2021 Oct 07;PP. [doi: 10.1109/TCBB.2021.3118418] [Medline: 34618675]
83. Poulain R, Gupta M, Foraker R, Beheshti R. Transformer-based multi-target regression on electronic health records for primordial prevention of cardiovascular disease. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2021 Dec;2021:726-731 [FREE Full text] [doi: 10.1109/bibm52615.2021.9669441] [Medline: 36684475]
84. Pang C, Jiang X, Kalluri KS, Spotnitz M, Chen R, Perotte A, et al. CEHR-BERT: incorporating temporal information from structured EHR data to improve prediction tasks. arXiv Preprint posted online on November 10, 2021. [doi: 10.48550/arXiv.2111.08585]
85. Rao S, Li Y, Ramakrishnan R, Hassaine A, Canoy D, Cleland J, et al. An explainable transformer-based deep learning model for the prediction of incident heart failure. IEEE J Biomed Health Inform 2022 Jul;26(7):3362-3372 [FREE Full text] [doi: 10.1109/JBHI.2022.3148820] [Medline: 35130176]
86. Du J, Zeng D, Li Z, Liu J, Lv M, Chen L, et al. An interpretable outcome prediction model based on electronic health records and hierarchical attention. Int J Intell Syst 2021 Oct 11;37(6):3460-3479. [doi: 10.1002/int.22697]
87. De Barros PH, Rodrigues JF. AttentionHCare: advances on computer-aided medical prognosis using attention-based neural networks. In: Proceedings of the International Joint Conference on Neural Networks. 2022 Presented at: IJCNN 2022; July 18-23, 2022; Padua, Italy. [doi: 10.1109/ijcnn55064.2022.9892642]
88. Liu S, Wang X, Xiang Y, Xu H, Wang H, Tang B. CATNet: cross-event attention-based time-aware network for medical event prediction. Artif Intell Med 2022 Dec;134:102440. [doi: 10.1016/j.artmed.2022.102440] [Medline: 36462902]
89. Yang F, Zhang J, Chen W, Lai Y, Wang Y, Zou Q. DeepMPM: a mortality risk prediction model using longitudinal EHR data. BMC Bioinformatics 2022 Oct 14;23(1):423 [FREE Full text] [doi: 10.1186/s12859-022-04975-6] [Medline: 36241976]
90. Chen X, Lin J, An Y. DL-BERT: a time-aware double-level BERT-style model with pre-training for disease prediction. In: Proceedings of the IEEE International Conference on Big Data. 2022 Presented at: Big Data 2022; December 17-20, 2022; Osaka, Japan. [doi: 10.1109/BigData55660.2022.10020513]
91. Sun Z, Yang X, Feng Z, Xu T, Fan X, Tian J. EHR2HG: modeling of EHRs data based on hypergraphs for disease prediction. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2022 Presented at: BIBM 2022; December 06-08, 2022; Las Vegas, NV. [doi: 10.1109/bibm55620.2022.9995204]
92. Yu F, Cui L, Cao Y, Zhu F, Xu Y, Liu N. Feature-guided logical perception network for health risk prediction. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2022 Presented at: BIBM 2022; December 06-08, 2022; Las Vegas, NV. [doi: 10.1109/bibm55620.2022.9995625]
93. Niu K, Lu Y, Peng X, Zeng J. Fusion of sequential visits and medical ontology for mortality prediction. J Biomed Inform 2022 Mar;127:104012 [FREE Full text] [doi: 10.1016/j.jbi.2022.104012] [Medline: 35144001]
94. AlSaad R, Malluhi Q, Janahi I, Boughorbel S. Predicting emergency department utilization among children with asthma using deep learning models. Healthc Anal 2022 Nov;2:100050. [doi: 10.1016/j.health.2022.100050]
95. Gerrard L, Peng X, Clarke A, Schlegel C, Jiang J. Predicting outcomes for cancer patients with transformer-based multi-task learning. In: Proceedings of the Advances in Artificial Intelligence. 2021 Presented at: AI 2021; February 2-4, 2022; Sydney, Australia. [doi: 10.1007/978-3-030-97546-3_31]
96. AlSaad R, Malluhi Q, Boughorbel S. PredictPTB: an interpretable preterm birth prediction model using attention-based recurrent neural networks. BioData Min 2022 Feb 14;15(1):6 [FREE Full text] [doi: 10.1186/s13040-022-00289-8] [Medline: 35164820]
97. Ramchand S, Tsang G, Cole D, Xie X. RetainEXT: enhancing rare event detection and improving interpretability of health records using temporal neural networks. In: Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics. 2022 Presented at: BHI 2022; September 27-30, 2022; Ioannina, Greece. [doi: 10.1109/bhi56158.2022.9926906]
98. Andjelkovic J, Ljubic B, Hai AA, Stanojevic M, Pavlovski M, Diaz W, et al. Sequential machine learning in prediction of common cancers. Informatics Med Unlocked 2022;30:100928. [doi: 10.1016/j.imu.2022.100928]
99. Yu F, Cui L, Cao Y, Liu N, Huang W, Xu Y. Similarity-aware collaborative learning for patient outcome prediction. In: Proceedings of the Database Systems for Advanced Applications. 2022 Presented at: DASFAA 2022; April 11-14, 2022; Virtual Event. [doi: 10.1007/978-3-031-00126-0_31]
100. Li Y, Salimi-Khorshidi G, Rao S, Canoy D, Hassaine A, Lukasiewicz T, et al. Validation of risk prediction models applied to longitudinal electronic health record data for the prediction of major cardiovascular events in the presence of data shifts. Eur Heart J Digit Health 2022 Dec;3(4):535-547 [FREE Full text] [doi: 10.1093/ehjdh/ztac061] [Medline: 36710898]

XSL•FO
RenderX

101. Li Y, Mamouei M, Salimi-Khorshidi G, Rao S, Hassaine A, Canoy D, et al. Hi-BEHRT: hierarchical transformer-based model for accurate prediction of clinical events using multimodal longitudinal electronic health records. IEEE J Biomed Health Inform 2023 Feb;27(2):1106-1117 [FREE Full text] [doi: 10.1109/JBHI.2022.3224727] [Medline: 36427286]
102. Dong X, Wong R, Lyu W, Abell-Hart K, Deng J, Liu Y, et al. An integrated LSTM-HeteroRGNN model for interpretable opioid overdose risk prediction. Artif Intell Med 2023 Jan;135:102439 [FREE Full text] [doi: 10.1016/j.artmed.2022.102439] [Medline: 36628797]
103. Guo LL, Steinberg E, Fleming SL, Posada J, Lemmon J, Pfohl SR, et al. EHR foundation models improve robustness in the presence of temporal distribution shift. Sci Rep 2023 Mar 07;13(1):3767 [FREE Full text] [doi: 10.1038/s41598-023-30820-8] [Medline: 36882576]
104. Liang Y, Guo C. Heart failure disease prediction and stratification with temporal electronic health records data using patient representation. Biocybern Biomed Eng 2023 Jan;43(1):124-141. [doi: 10.1016/j.bbe.2022.12.008]
105. Lee LT, Yang HC, Nguyen PA, Muhtar MS, Li YC. Machine learning approaches for predicting psoriatic arthritis risk using electronic medical records: population-based study. J Med Internet Res 2023 Mar 28;25:e39972 [FREE Full text] [doi: 10.2196/39972] [Medline: 36976633]
106. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online on October 11, 2018 [FREE Full text]
107. ChatGPT homepage. ChatGPT. URL: https://chat.openai.com/ [accessed 2023-10-30]
108. Zhang Y, Wang H, Zhang D, Wang D. DeepRisk: a deep transfer learning approach to migratable traffic risk estimation in intelligent transportation using social sensing. In: Proceedings of the 2019 15th International Conference on Distributed Computing in Sensor Systems. 2019 Presented at: DCOSS 2019; May 29-31, 2019; Santorini, Greece. [doi: 10.1109/dcoss.2019.00039]
109. Soroush A, Glicksberg BS, Zimlichman E, Barash Y, Freeman R, Charney AW, et al. Large language models are poor medical coders — benchmarking of medical code querying. NEJM AI 2024 Apr 25;1(5) [FREE Full text] [doi: 10.1056/aidbp2300040]
110. Johnson AE, Pollard TJ, Shen LW, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
111. Post WS, Watson KE, Hansen S, Folsom AR, Szklo M, Shea S, et al. Racial and ethnic differences in all-cause and cardiovascular disease mortality: the MESA study. Circulation 2022 Jul 19;146(3):229-239 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.122.059174] [Medline: 35861763]
112. Yemane L, Mateo CM, Desai AN. Race and ethnicity data in electronic health records-striving for clarity. JAMA Netw Open 2024 Mar 04;7(3):e240522. [doi: 10.1001/jamanetworkopen.2024.0522] [Medline: 38466312]
113. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020 Jan 02;21(1):6 [FREE Full text] [doi: 10.1186/s12864-019-6413-7] [Medline: 31898477]
114. Guo LL, Pfohl SR, Fries J, Johnson AE, Posada J, Aftandilian C, et al. Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. Sci Rep 2022 Feb 17;12(1):2726 [FREE Full text] [doi: 10.1038/s41598-022-06484-1] [Medline: 35177653]
115. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv Preprint posted online on May 22, 2017 [FREE Full text]
116. Vig J. Visualizing attention in transformer-based language representation models. arXiv Preprint posted online on April 4, 2019 [FREE Full text] [doi: 10.1090/mbk/121/79]

Abbreviations

AUROC: area under the receiver operating characteristic curve
BERT: bidirectional encoder representations from transformers
CNN: convolutional neural network
DL: deep learning
EHR: electronic health record
ICD-10: International Classification of Diseases, Tenth Revision
ICU: intensive care unit
LLM: large language model
LSTM: long short-term memory
MIMIC: Medical Information Mart for Intensive Care
ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Study Risk of Bias Assessment Tool
RNN: recurrent neural network

XSL•FO
RenderX

ROB: risk of bias
SHAP: Shapley additive explanations

Edited by A Coristine; submitted 19.02.24; peer-reviewed by R Perotte, A Ahmadipour, Y Xin; comments to author 07.10.24; revised version received 14.12.24; accepted 18.02.25; published 18.03.25.

Please cite as:
Hama T, Alsaleh MM, Allery F, Choi JW, Tomlinson C, Wu H, Lai A, Pontikos N, Thygesen JH
Enhancing Patient Outcome Prediction Through Deep Learning With Sequential Diagnosis Codes From Structured Electronic Health Record Data: Systematic Review
J Med Internet Res 2025;27:e57358
URL: https://www.jmir.org/2025/1/e57358
doi:10.2196/57358
PMID:

©Tuankasfee Hama, Mohanad M Alsaleh, Freya Allery, Jung Won Choi, Christopher Tomlinson, Honghan Wu, Alvina Lai, Nikolas Pontikos, Johan H Thygesen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Patients' Experiences of Digital Health Interventions for the Self-Management of Chronic Pain: Systematic Review and Thematic Synthesis

Ashleigh Main[1], BSc, MSc; Haruno McCartney[1], BSc; Maryam Ibrar[1], MA, MSc; Fiona Muirhead[2], BSc, MSc, PhD; Alexandra Mavroeidi[3], BSc, MSc, PhD; Harleen Kaur Rai[1], BSc, MSc, PhD; Roma Maguire[1], BSc, MSc, PhD

[1]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, United Kingdom

[2]Physical Activity for Health, School of Psychological Sciences and Health, University of Strathclyde, Glasgow, United Kingdom

[3]Department of Occupational Therapy, Human Nutrition and Dietetics, School of Health & Life Sciences, Glasgow Caledonian University, Glasgow, United Kingdom

Corresponding Author:
Ashleigh Main, BSc, MSc
Department of Computer and Information Sciences
University of Strathclyde
Livingston Tower
26 Richmond Street
Glasgow, G1 1XH
United Kingdom
Phone: 44 0141 552 4400
Email: ashleigh.main@strath.ac.uk

Abstract

Background: Research regarding the effectiveness of digital health interventions (DHIs) for people living with chronic pain is widely documented, although it is often measured against changes in clinical outcomes. To gain a comprehensive understanding of the full impact of DHIs, it is vital to understand the experience of individuals who are using them. An exploration of qualitative data regarding the experience of people living with chronic pain engaging with DHIs could provide a more in-depth account of how individuals interact and engage with such tools, uncovering the overall impact DHIs can have on the lives of people living with chronic pain.

Objective: This qualitative systematic review and thematic synthesis aimed to appraise and synthesize relevant qualitative evidence on patients' experiences of engaging with DHIs for the self-management of chronic pain symptoms.

Methods: A systematic literature search of qualitative and mixed methods studies published between 2013 and 2023 was conducted across 6 databases: MEDLINE, PubMed, Embase, CINAHL, PsycINFO, and Scopus. Eligible studies included adult patients aged ≥18 years with a chronic pain diagnosis (ie, >12 weeks) reporting on the experience of engaging in a DHI for the self-management of chronic pain. The Critical Appraisal Skills Program checklist for qualitative research was used to appraise each study. Following a 3-step inductive thematic synthesis approach, the researcher performed line-by-line coding of each eligible article to identify descriptive themes. Through iterative evaluation of the descriptive themes, analytical themes that facilitated a deeper understanding of the data were derived.

Results: In total, 37 qualitative and mixed methods studies were included in the review. Thematic synthesis revealed three overarching themes encompassing five subthemes: (1) personal growth, with 2 subthemes (*gaining new insights* and *renewed mindset*); (2) active involvement, with 3 subthemes (*motivation*, *improved access*, and *health care decision-making*); and (3) connectedness and support.

Conclusions: A positive experience with DHIs among people living with chronic pain is achieved through an improved understanding of their condition, greater self-awareness of how symptoms impact their lives, and an increase in motivation to play an active role in their health care. DHIs promote the confidence and independence of people living with chronic pain, as well as facilitate a sense of ongoing support between routine appointments. However, DHIs may disempower people living with chronic pain by placing too much focus on their pain and should be used as an adjunct to existing care as opposed to replacing in-person appointments. To appropriately meet the needs of people living with chronic pain, the content and features of DHIs

XSL•FO
RenderX

should be personalized. Development of future DHIs should use a co-design approach involving key stakeholders to ensure the needs of people living with chronic pain are met.

Trial Registration: PROSPERO CRD42023445100; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023445100

International Registered Report Identifier (IRRID): RR2-10.2196/52469

(J Med Internet Res 2025;27:e69100) doi:10.2196/69100

KEYWORDS

chronic pain; digital health; digital tool; digital health intervention; mobile health; mHealth; eHealth; self-management; pain management; person-centered; patient experiences; systematic review; thematic synthesis

Introduction

Background

Chronic pain, defined as pain that persists past the normal healing time of 3 months [1], is one of the leading causes of disability worldwide [2]. The intricate and diverse nature of chronic pain makes it a complex condition to fully understand, as pain can occur without a clear underlying cause and cannot be attributed to injury or disease. This is described as *chronic primary pain* and includes diagnoses such as fibromyalgia, chronic primary musculoskeletal pain, and chronic primary neuropathic pain. Pain that occurs as the result of an initial injury or underlying illness is known as *chronic secondary pain*, and this includes conditions such as osteoarthritis, rheumatoid arthritis, and cancer related chronic pain [3]. A systematic review assessing prevalence in the United Kingdom estimated that approximately 43.5% of the population experience chronic pain, with 12% of people living with chronic pain reporting pain that is moderately to severely disabling [4].

Treatment

It is well known that chronic pain is multifaceted, and this can present a substantial burden to the lives of those affected [5]. Physical symptoms are often accompanied by distressing psychological, social, and economic factors that can impact every facet of an individual's life [2]. To reflect its multifaceted nature, it is crucial that treatment addresses the diverse biopsychosocial aspects of chronic pain. Evidence indicates that treatment that uses a combination of approaches is the most effective for improving pain outcomes [6,7]; however, access to adequate multidisciplinary treatment is limited [8]. Lengthy waiting times for treatment are reported around the world [9-12], leaving patients with no other option than to rely on their general practitioner for advice [8]. Self-management, defined as "the individual's ability to manage the symptoms, treatment, physical, psychosocial consequences, and lifestyle changes inherent in living with a long-term disorder" [13], poses a possible solution to the issue of limited access to treatment. Self-management techniques encompass a range of disciplines, integrating approaches such as tracking symptoms, engaging in exercise, relaxation techniques, and cognitive behavioral therapy (CBT) to address chronic pain symptoms [14]. Such strategies enable individuals to play an active role in their health care by treating and managing their pain on a day-to-day basis without relying on input from health services or health care professionals (HCPs). A growing body of evidence supports that self-management interventions are effective in assisting people living with chronic pain manage their symptoms [14-17].

Digital Health

A promising mode of delivering self-management interventions is through the use of digital health technologies. The need for digitization of health care delivery became evident following the COVID-19 pandemic, as digital technologies offer an accessible, affordable, and flexible approach to managing chronic pain symptoms [18]. Digital health interventions (DHIs) use various modes of delivery, including smartphone apps, websites, wearable activity trackers, and virtual reality [19], and the literature highlights promising effects of DHIs on clinical outcomes. Pfeifer et al [20] evaluated the efficacy of app-based interventions on a variety of chronic pain conditions. An analysis of 22 studies found that apps are significantly more effective in reducing pain compared to control groups. Another systematic review that analyzed the impact of web-based and mobile-based self-management interventions on patients with chronic low back pain indicated that there was a clinically important effect both at immediate and short-term follow-up on pain relief and improved disability [21]. Furthermore, Moreno-Ligero et al [22] reviewed 22 randomized controlled trials to evaluate the effectiveness of mobile health interventions on a range of chronic pain conditions. The results suggested that the use of mobile health interventions can have beneficial effects on pain intensity, functional disability, and quality of life (QoL).

Despite an abundance of the literature outlining the benefits of DHIs for people living with chronic pain, the findings primarily focus on clinical outcomes. Research regarding clinical outcomes plays a vital role in supporting evidence-based practice, although these studies are not able to provide a comprehensive understanding of how users interact and engage with DHIs. The literature shows clear evidence that DHIs are effective in improving symptoms of chronic pain, although the full potential of DHIs is limited as engagement and retention rates within these interventions are low. A review evaluating the efficacy of web-based CBT interventions for chronic pain found that as many as 26.6% of users dropped out before completing the intervention [23]. Similarly, a more recent review exploring the DHIs for the self-management of osteoarthritis found that 6 out of the 7 studies evaluated reported less than 70% adherence to the intervention [24]. For people living with chronic pain to obtain maximum results from self-management DHIs, users need to engage with and continue to use these tools. This strongly depends on a user's intention to use the tool [25],

XSL•FO
RenderX

although the intentions behind why a user engages are poorly understood. Research regarding the barriers and facilitators to the use of DHIs among people living with chronic pain reported that too much information, a lack of tailored or personalized content, and a lack of motivation negatively affect engagement [26,27].

To overcome these barriers, it is vital that DHIs meet the actual needs of the people who are using them. To do this, developers of such interventions require a comprehensive understanding of how users experience DHIs as this could unveil aspects that will encourage engagement and retention. A qualitative exploration of how people living with chronic pain experience the use of DHIs could provide insights into how users benefit from DHIs as well as highlight the associated challenges. These findings could inform the design and development of future DHIs, resulting in more effective tools that meet the needs of people living with chronic pain.

A recent review exploring experiences of patients with chronic pain with DHIs has found that by engaging with DHIs, participants learned new skills as to how to manage their pain and gained increased confidence and reassurance with regards to managing their condition. In addition, the review stated that users were able to improve the perception of their condition, thereby obtaining a more optimistic outlook on their life [28]. The findings from this review highlighted some key recommendations for future DHIs, including providing tailored content, structured goal setting facilitated by prompts, and prioritizing features that change patients' attitudes and behaviors. This review aims to build on this evidence with a particular focus on self-management interventions. It will also include qualitative data from mixed methods studies, which dominate this domain of research and could therefore offer valuable insights concerning the patient experience of engaging with DHIs that may have been previously overlooked.

By exploring qualitative evidence regarding patients' experiences of engaging with specifically self-management DHIs, the findings from this review could improve the development of future DHIs by providing an in-depth explanation of how self-management DHIs impact the lives of people living with chronic pain, and which features should be included or omitted in future DHIs to enhance engagement and retention. In turn, this will uncover the full potential of how self-management DHIs can benefit people living with chronic pain.

Aim and Objectives

The aim of this qualitative systematic review and thematic synthesis is to identify, appraise, and synthesize relevant qualitative evidence on patients' experiences of engaging with DHIs for the self-management of chronic pain symptoms.

The objectives of this synthesis are as follows:

1. To explore how people living with chronic pain experience DHIs for the self-management of chronic pain
2. To identify factors that could improve the design and development of future DHIs

Methods

Overview

This systematic review and thematic synthesis was conducted and reported in accordance with the Enhancing Transparency in Reporting the Synthesis of Qualitative Research (ENTREQ) guidelines [29]; a completed checklist is presented in Multimedia Appendix 1. The protocol has been published [30] and is registered on PROSPERO (CRD42023445100).

Search Strategy

The search strategy was adapted and applied to 6 electronic databases as follows: Embase, MEDLINE, PubMed, CINAHL, PsycINFO, and Scopus. The full search strategy for each database can be found in Multimedia Appendix 2. To retrieve relevant articles not identified on the electronic databases, forward and backward citation searches were performed on the included studies. The sample, phenomenon of interest, design, evaluation, and research (SPIDER) framework [31], which describes the sample (S), phenomenon of interest (PI), design (D), evaluation (E), and research type (R), was used to structure the search terms and develop a comprehensive search strategy relevant to the research question. The search terms that comply with each of the key characteristics of the SPIDER framework are outlined in Table 1 [31].

Each electronic database was searched on July 4, 2023, and updated on January 29, 2024. Each search was restricted to papers published in the English language; gray literature, other reviews, protocols, and unpublished studies were excluded from the review. Due to major advancements in technology, only articles published in the last 10 years (ie, 2013-2023) were included, those published before 2013 were excluded due to the potential risk of no longer being relevant. Search results were imported to EndNote (version 20; Clarivate), and duplicates were removed.

XSL•FO
RenderX

Table 1. Search terms for each electronic database.

Categories	Search terms
Sample	“chronic pain” OR “chronic pain” OR “persistent pain” OR “long-term pain” OR fibromyalgia OR “rheumatoid arthritis” OR osteoarthritis OR “neuro* pain” OR “musculoskeletal pain*” OR “orofacial pain” OR “visceral pain” OR endometriosis OR headache OR “irritable bowel syndrome” OR “back pain*” OR “low* back pain” OR “neck pain” OR “joint pain*” OR sciatica OR “cancer pain” OR “post surg* pain” OR “post trauma* pain” OR “complex regional pain” OR “chronic primary headache” OR “chronic primary visceral pain” OR “chronic musculoskeletal pain”
Phenomenon of interest	“telehealth” OR “digital health” OR mhealth OR ehealth OR telenursing OR telerehabilitation OR “digital intervention” OR “remote consultation” OR “electronic health” OR “internet health” OR “digital health” OR “health technolog*“ OR ”health management app*“ OR ”digital health programme“ OR ”digital health app*“
Design	”focus group*“ OR interview*
Evaluation	experience* OR perspective* OR perception* OR accept* OR satisf* OR view* OR attitude*
Research type	qualitative OR ”mixed method*“ OR “grounded theory” OR ethnography OR “thematic analysis”

Eligibility Criteria

The inclusion and exclusion criteria were based on the key characteristics of the SPIDER framework outlined earlier [31]. Textbox 1 describes the inclusion and exclusion criteria, and a comprehensive account of the inclusion and exclusion criteria is detailed in the study protocol [30].

Textbox 1. Inclusion and exclusion criteria based on the sample, phenomenon of interest, design, evaluation, and research (SPIDER) framework.

Inclusion criteria

- Articles reporting on adults (ie, aged ≥18 years) who were diagnosed with any chronic pain condition exceeding 12 weeks
- Articles reporting qualitative data, including mixed methods studies with a qualitative component, on patients’ experiences of engaging with a digital health intervention (DHI) for the self-management of chronic pain
- Articles reporting on DHIs, regardless of modality, to deliver self-management strategies for managing chronic pain symptoms
- Articles published from January 1, 2013, to December 31, 2023
- Articles published in the English language

Exclusion criteria

- Articles reporting on participants aged <18 years who have not received a diagnosis exceeding 12 weeks
- Gray literature, reviews, protocols, dissertations, and articles reporting on quantitative data from mixed methods studies
- Articles describing participant expectations and perceptions rather than the experience of engaging in a DHI
- Articles describing the experience of carers or health care professionals
- Articles describing DHIs that deliver management strategies that involve regular live health care professional contact (eg, through videoconferencing or Skype)
- Articles published before January 1, 2013
- Articles published in languages other than the English language

Screening and Data Extraction

Articles retrieved from the electronic database were uploaded to EndNote and duplicates were removed. Articles were then imported to Rayyan (Rayyan Systems, Inc), a software that facilitates screening in the systematic review process. After thoroughly discussing the eligibility criteria to ensure that reviewers had a comprehensive understanding of which studies to include and exclude, all of the titles and abstracts were screened by 2 independent reviewers (AM and HM). There was a high rate of agreement, as 89% (33/37) of the papers were agreed upon. Where disagreement between reviewers did occur, this was resolved through discussion between 2 reviewers through a thorough investigation of each paper against the inclusion and exclusion criteria (AM and MI). In the following phase, full-text articles of the included titles and abstracts were obtained, and the same 2 reviewers (AM and HM) assessed the full-text articles for eligibility. During this round, there was a

XSL•FO
RenderX

96% agreement rate, and again, the reviewers resolved any disagreement through discussion. In one case, resolution between the reviewers could not be reached and a third reviewer was consulted (FM).

A data extraction table was devised by the research team, and 2 reviewers (AM and HM) independently extracted data from 20% (7/37) of the included articles. Data extraction forms were then compared between reviewers to ensure that the key characteristics of each study were extracted in a systematic and standardized way. The first author (AM) then continued to extract data from the remaining included articles.

Quality Assessment and Critical Appraisal

Due to the subjective nature of qualitative research, assessing the methodological quality of included studies is a crucial step to ensure the validity of results. The Critical Appraisal Skills Program (CASP) qualitative research checklist was chosen to appraise each article as the checklist is endorsed by Cochrane's Supplemental Guidance for Inclusion of Qualitative Research [32]. Assessment consists of scoring 10 questions regarding methodological quality as "yes," "can't tell," or "no." The questions concerned issues such as research design, recruitment strategy, data collection, ethics, data analysis, and presentation and interpretation of findings.

Two independent reviewers (AM and HM) assessed the CASP checklist criteria before commencing the quality assessment phase. This was to fully understand the criteria before applying each paper against the checklist. The 2 reviewers appraised 20% (7/37) of the included articles. After comparing scores of the subsample, which reached a high degree of agreement (6/7, 85%), discrepancies were resolved through discussion. The first author then continued to appraise the remaining 30 articles. There are currently no standardized rules for the exclusion of studies based on their appraisal score; therefore, all studies were included in the synthesis. The average quality appraisal score of the included articles was high (ie, 8.8/10).

Thematic Synthesis

A 3-step inductive thematic synthesis approach as described by Thomas and Harden [33] was applied to the results and discussion sections of each included article. This methodology was selected as it is recommended by the *Cochrane Handbook for Systematic Reviews of Interventions* [32]. The 3-step methodology provides an opportunity to extend beyond the findings of the primary studies by deriving analytical themes from descriptive themes, which in turn facilitates a deeper conceptual understanding of the data. Qualitative evidence syntheses have faced criticism for decontextualizing the results of individual studies [33]. Therefore, before the 3-step thematic synthesis approach began, each paper was read in its entirety to grasp a full understanding.

In step 1, the first author (AM) imported the primary studies to NVivo (Lumivero), where line-by-line coding of the results and discussion sections of each primary study was performed to identify concepts within the data. In step 2, items coded in step 1 were grouped together to create descriptive themes, closely aligning with the themes identified in the primary research. Finally in stage 3, by continually examining and interpreting the descriptive themes, analytical themes that reflected a deeper understanding of the data were derived.

Results

Study Selection

The search yielded 1709 records. After duplicates were removed, the titles and abstracts of 868 (51%) records were screened, and 85 (5%) records were retrieved for full-text screening. A total of 48 (57%) records were excluded; reasons for exclusion can be found in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 1; Multimedia Appendix 3). A total of 37 (43%) records met all eligibility criteria and were included in the review (Multimedia Appendix 4 [34-70]).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study selection flowchart. DHI: digital health intervention.

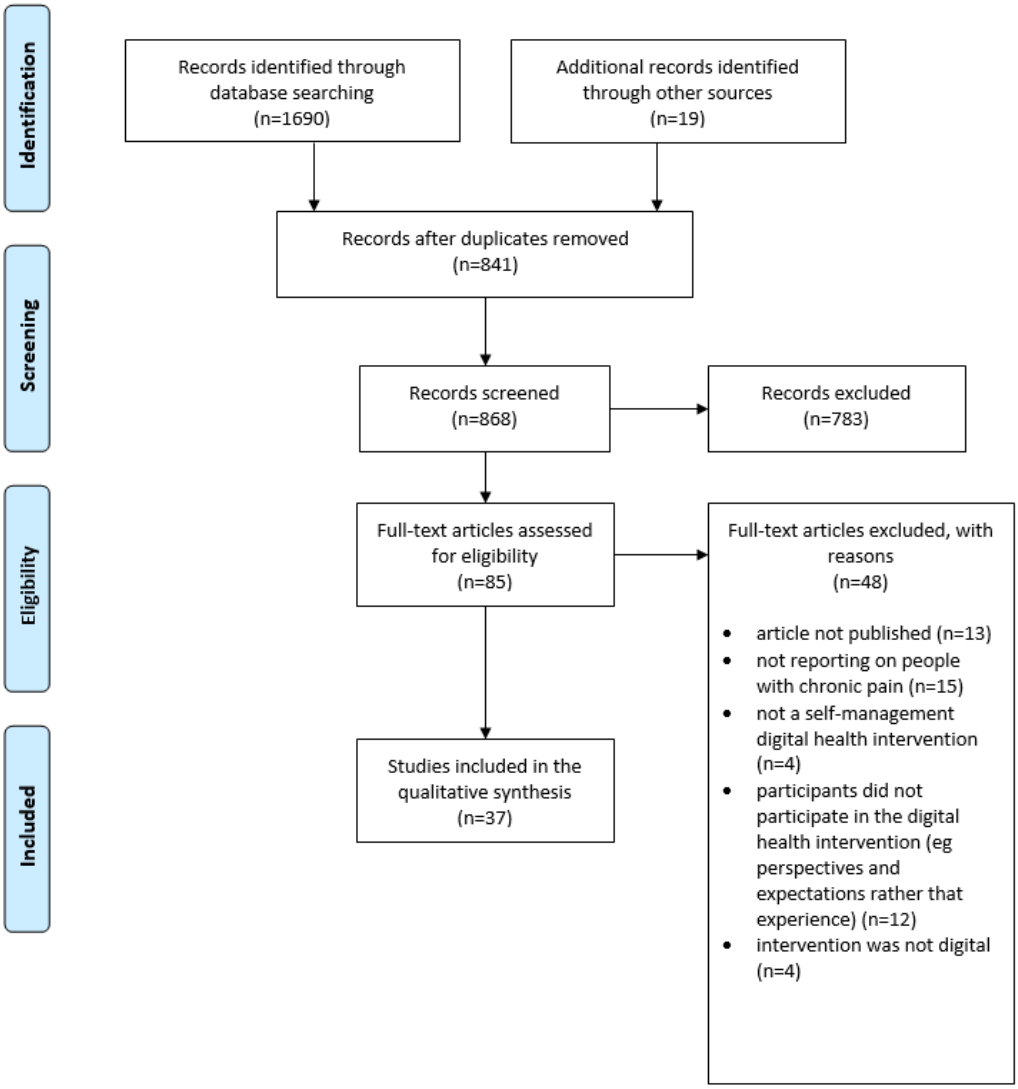

Study Characteristics

The 37 included studies used a range of qualitative methods to collect data, including interviews [34-62], focus groups [63-67], questionnaires [68], and a combination of interviews and questionnaires [69,70]. In total, 18 (49%) studies conducted interviews remotely [34,38,41-43,45-49,52,54-56,58,62,69,70] and 5 (14%) studies had the option of both in-person or remote interviews [37,39,53,59,61]. A total of 2 (5%) studies did not detail how the interviews took place [36,60]. Focus groups were conducted in person in 2 (5%) studies [63,64], on the web in 2 (5%) studies [65,66], and 1 (3%) study used a combination of both in-person and online methods [67].

The sample sizes ranged from 7 to 36, with a total of 707 participants. A total of 4 of 37 (11%) studies included only female participants [37,45,59,65], and the remaining 33 (89%) studies included both male and female participants [34-36,38-44,46,48-58,60-64,66-70]. The interventions were investigated in the United States [42,44,45,51,54,60,61], the Netherlands [35,46,59,62,66,69,70], Canada [40,41,53,67], the United Kingdom [34,36,37,58], Australia [43,47,49], Sweden [39,50,55,63], Germany [48,52], New Zealand [64], France [68], Belgium [65], Norway [38], and Turkey [57], and 1 study was conducted in both Sweden and Denmark [56].

In terms of chronic pain conditions, 6 of 37 (16%) studies did not differentiate between chronic pain diagnoses [38,40,42,44,47,55], while the remaining 31 (84%) studies focused on specific chronic pain conditions: knee or hip osteoarthritis [34,39,41,43,49,53,54,63,64], rheumatoid arthritis [36,48,57,62,69,70], low back pain [35,52,56,61,68], cancer pain [45,51,65,67], musculoskeletal pain [46,50,60], chronic pelvic pain [37], irritable bowel syndrome [58], and temporomandibular disorder [59], and 1 study evaluated rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis [66].

A range of digital tools were used to deliver the telehealth interventions; smartphone apps [35-38,42,44,48,53,56,57,62,64,66,68,70], web-based interventions [34,39,41,43,45,46,49-52,54,58-61,65,69], virtual reality headsets [40,67], social media [47], and wearable activity trackers [63], and 1 intervention offered both a smartphone app and web-based platform [55]. The purpose of the interventions varied, with most studies using a multitude of different intervention components [34,39,41-43,45,46,50,

XSL•FO
RenderX

51,53,55,56,59,62,65,69]. While the remaining interventions also incorporated a range of components, the main focus of the intervention was on CBT or mindfulness and meditation strategies [37,38,40,58,67], self-monitoring and symptom tracking [36,48,60,64,66,70], physical activity [35,49,57,63,68], virtual auricular point acupressure [44,61], social support [47], and shared decision-making [52,54].

The study characteristics pertaining to each included article can be found in Multimedia Appendix 4, and a table of intervention characteristics can be found in Multimedia Appendix 5 [34-70].

Quality Assessment

Overview

Of the 37 studies, 35 (95%) studies included in the review were rated as "high" quality by scoring >7 out of 10 [34-46,48-68,70]. A total of 2 (5%) studies were rated as "low" quality [47,69], failing to describe multiple key aspects, such as the research design, recruitment strategy, data collection, and data analysis process; 2 (5%) studies did not clearly state the aims of the research [44,51]; 4 (11%) studies did not provide thorough detail as to how the data were collected [36,47,60,69]; and 3 (8%) studies failed to give an in-depth description of the data analysis process [47,69,70]. The item that was scored most poorly throughout the included studies was the consideration of researcher bias, with 26 (70%) studies lacking sufficient detail on the relationship between the researcher and the participants [34-37,40-43,46-48,51,53-57,60-62,65-70]. The full quality appraisal for each study can be found in Multimedia Appendix 6.

Thematic Synthesis Findings

The data revealed 3 analytical themes that encapsulated patients' experiences of DHIs for chronic pain management as follows: (1) personal growth, with 2 subthemes (*gaining new insights* and *renewed mindset*); (2) active involvement with 3 subthemes *motivation*, *improved access*, and *health care decision-making*; and (3) connectedness and support. The subsequent sections provide a description of each theme.

Theme 1: Personal Growth ("I Could See the Change That Was Happening")

Overview

A dominant theme throughout the studies was that participants experienced personal changes that made them grow and evolve as individuals, facilitating personal growth both physically and emotionally (34/37, 92%). Improved understanding of their condition led participants to adopt greater self-awareness of their symptoms. Participants obtained an optimistic outlook on life through enhanced confidence and independence, ultimately accomplishing a sense of acceptance of their diagnosis. This analytical theme encompasses 2 subthemes: gaining new insights and a renewed mindset.

Gaining New Insights: "I Understand How My Brain Controls the Pain"

Most studies reported that self-management DHIs improved participants' self-awareness, facilitating a deeper understanding of their condition [34]. Participants were able to recognize patterns and triggers that exacerbate pain by tracking their symptoms within web-based and app-based interventions [36,54,58,64]: "The one thing it did make me do was look at my diet more and made me associate different reactions and triggers for my IBS" [58]. This sense of enhanced self-awareness allowed participants to understand the relationship between different variables and how this impacted their pain, enabling them to make practical changes to their lifestyles resulting in better symptom management.

Participants were able to identify thoughts and emotions that had a negative impact on their pain using CBT [58]. Mindfulness and meditation sessions equipped participants with various skills to reduce stress and anxiety [38]. Some participants were unaware that CBT could be useful for pain management before participating in the intervention [50]. Many of the interventions encouraged participants to self-reflect, which enhanced their overall understanding of the different aspects of their pain [38,43,46,50,58,60,64].

Interventions that supported physical activity gave patients insight into what exercises they can do to alleviate pain [41,43,50,56,57,62,63]. This made participants aware of how their overall health can improve:

> *Because I'm doing a bit of exercise in combination with the diet. So, I can see if I continue this part, not only my knee will benefit, [but] also my health.* [41]

Conversely, some studies found that enhanced insights into pain can also have a negative effect [37,38,40,46,62,67,70]. Participants occasionally found that too much focus on their symptoms made them aware of how poorly they were coping, "Every time I got those registrations...you have to start to notice, and feel, that you are not actually doing so well after all" [38]. For some, this increased awareness was perceived as "heavy to encounter" [46], particularly for patients who are trying to forget about their pain. It was also highlighted that regularly tracking and monitoring symptoms can risk placing too much emphasis on the negative aspects of an individual's health [55,60,62]: "It just triggered me to start looking at my overall health" [60]. One study reported that some elements of CBT were "too confronting for patients" [62], and another study that delivered a virtual reality computer-simulated environment through a binocular headset found that focusing on pain resulted in participants experiencing more pain [67].

Renewed Mindset: "It's Brought a Positive Into My Lifestyle"

The studies reviewed suggested that DHIs instilled confidence and reassurance in many of the participants. Information within the DHIs validated some participants' beliefs and reassured them that they were not alone [58]. Collecting data on pain and other symptoms increased participants' confidence when it came to making decisions about treatment [54]. Participants felt reassured that self-management strategies were effective based on data collected [35], particularly with the help of graphical summaries that displayed patterns of their pain [48]:

> *I felt that my back was very painful this week, but actually my pain score after doing the exercises is decreasing. That is, for me, a reminder I'm going in*

XSL•FO
RenderX

> *the right direction, and I find that very reassuring.* [35]

Video instructions incorporated into exercise-based interventions promoted a sense of confidence by reassuring participants that they were completing the exercises correctly [59]. One study noted that exercises for managing knee pain promoted participants confidence to maintain their lifestyle and not give up because of their pain: "It's helped me not only with my mobility but my self-confidence to be able to go, yeah, I can get up there all right and come down there" [49].

Participants developed an optimistic attitude about their pain throughout CBT sessions [38,43,58], which offered participants alternative coping strategies so that they were confident enough to not depend heavily on medication: "I got to the stage where I wouldn't leave the house without making sure the medication was in my handbag...while I was doing the trial I never thought like that at all" [58]. Another article reported that app-based CBT was effective in changing participants' perspectives by providing hope that they can live a good QoL despite their pain [38]. In one study, family members noticed an increase in independence because of web-based pain-coping skills training [51]. Educational sessions were also found to be effective in promoting acceptance in some participants by changing the way they think about their pain [46,56]. Some studies reported that participants could use the DHIs to self-manage their symptoms instead of relying on input from HCPs [52,56,65]: "When you understand everything in the program and know how to apply this information into daily life, a conversation with a health care provider wouldn't be necessary anymore" [65].

By changing their perceptions of pain, many participants achieved a degree of acceptance which resulted in them coming to terms with their diagnosis [38,43,46,51,52,56,69]: "I'm just trying to learn to live with it [OA], and I know it's a cliché, but just accept the new normal, and not be living in the past and remembering how when I could just walk all day and not worry about it" [43].

On the other hand, some studies noted the adverse experiences of engaging in the DHIs on participants' psychological well-being [37,38,40,48,49,63,65,67]. Monitoring disease activity on a smartphone app evoked negative thoughts for one participant:

> *If you dwell too often on your own disease activity, it can be associated with negative thoughts. You may be more likely to get into such a negative vortex.* [48]

In an intervention that used a virtual reality immersive experience, some participants experienced anxiety, depression, and stress that aggravated their pain [67].

Theme 2: Active Involvement ("I Would Never Have Done It If It Hadn't Been Digital")

Overview

This theme illustrates how DHIs can promote patients to play an active role in the management of their chronic pain. As opposed to being a "passive" recipient of care, accessible self-management resources motivated participants to act and engage in conversations about their health care. This analytical theme is underpinned by 3 descriptive themes (motivation, improved access, and health care decision-making).

Motivation: "I Dare Do More Things!"

The studies continually described how participants were motivated to engage in behaviors that positively impacted their pain, and therefore promoted a greater sense of control [39,43,44,65,68]: "Thanks to the program, I got the feeling that I had more control on my pain" [65]. Participants expressed a sense of motivation that went beyond the goal of the intervention, such as inspiring them to lose weight [49] or to change their occupation to something that was less physically demanding and better suited to their needs [43]. The results from the exercised-based programs suggested that participants were motivated to exercise more as a direct result of the DHI [35,41,52,56-58,62,63,69]: "Thanks to the app, I could see what exactly it was I was supposed to do...That definitely increased how often I exercised" [35].

Notifications either as alerts or emails were noted to have a motivating effect on participant engagement throughout the interventions [34,35,39,41,43,49,57,59,63]: "Reminders and notifications are my favorite, I love that some notifications motivate me by saying 'how about moving? have a nice day!'" [57]. Nevertheless, some of the studies found that notifications and prompts had a negative impact on participants feeling of empowerment [49,63]. In a study that used wearable activity trackers to encourage physical activity, notifications prompting participants to reach their step goal were perceived as distressing when participants were in too much pain to meet their target [63].

Although most participants experienced enhanced motivation, some advised that they felt less motivated to engage with the DHI [39,41,45,49,58,62,63]. Participants struggled to stay motivated when completing a home-based yoga program due to outside distractions and other commitments [45]. Once pain had improved, some participants did not have the same motivation to continue with the program: "The pain was generally reduced...and then I did not feel as motivated as before so I quit doing the exercises" [39]. When participants felt demotivated and had not completed the recommended exercises, SMS text messaging reminders evoked feelings of shame or guilt in one study [49], and another described how the weekly content delivered to participants through a web-based self-care intervention felt like "unfinished homework" [41].

Improved Access: "It's Awfully Good to Have the Tools at Home"

Participants were motivated to play an active role in their health care because of convenient and flexible access to self-management resources. Participants in the app-based interventions commented on how having quick, convenient, and flexible access to self-management strategies encouraged them to take action [35,38,59,64]:

> *You've got your phone with you, and you can just [get it] ...done. Whereas, with a piece of paper, you've got to go and sit down somewhere, and you've got to write it out.* [64]

XSL•FO
RenderX

The flexibility of being able to choose when and where to engage in the various self-management techniques enabled participants to take control over their symptoms at a time that suited them, stimulating engagement with the DHIs [38,39,50,51,59,65]: "It's awfully good to have the tools at home to use when you can find the time" [51]. Participants advised how they were able to access the self-management strategies while on a waiting list for specialized treatment:

> *Yes, it was also like, there was such a long waiting list for primary care. So I thought...this...this I can start doing right away.* [39]

The affordability of the interventions was also valued, as participants who might not be able to afford a gym membership now have access to a range of exercises to relieve pain [41].

Health Care Decision-Making: "This Made Me Feel Safe to Share Those Thoughts"

Various features of the DHIs improved patient-physician communication during consultations, enabling them to play an active role in making decisions about their health care [35,36,41,46,48,50,52,54,60,63,66,70]. In a web-based intervention, participants noted how the CBT modules empowered them to become more involved in their rehabilitation as they had the tools to have meaningful discussions with their HCP [50]. Symptom tracking proved valuable in accurately portraying trends, highlighting exactly when pain improved or declined, which may provide insights that may have otherwise been overlooked [36,48,54]: "I don't think in that half hour that you've got that consultation doesn't always show the bigger picture of what you're actually dealing with" [36]. Some participants even advised their HCP that they wanted to change their treatment plan because of the intervention: "I said I didn't want the knee replacement yet because I found these exercises—I told him I was doing these exercises and I thought, 'they feel great.' I can walk more. I can do quite a few different things" [41].

Theme 3: Connectedness and Support (an "In-the-Pocket Friend")

This theme articulates how DHIs are valued as a source of ongoing support outside of routine health care appointments, with some participants even describing the intervention as a "friend when in need" [38]. The DHIs were found to be useful in fostering a sense of ongoing support once treatment had ended [55,65]. Others found the interventions helpful in between consultations as they could rely on the digital tool for assistance:

> *After listening to the therapist, I would come home and still have questions or forgot what the therapist said. Then, I had something to fall back on, and that was very pleasant.* [35]

Of the 37 studies, 6 (16%) studies described how the digital format of the intervention made it feel as though they had a digital "friend" or friendly motivating supporter [38,39,41,47,56,65]. In an app-based pain management CBT intervention, participants indicated that they perceived the app as a potential "friend when in need" that could offer social support to help participants feel less lonely: "If you are in a lot of pain, then you are...quite alone...I have felt that you can withdraw and listen to the soothing voice...it has given me a break..., and it has provided relaxation" [38]. Participants appreciated the conversational style of an online pain management program [65]. Another study described the app format as "an in-the-pocket friend" that encouraged participants to keep going:

> *Yes, a bit like an exercise partner who asks 'hey, shouldn't we go work out today?' and then you actually might get going. Where on the other hand, if you were on your own, you might forget about it or take the easy way out.* [56]

Sharing experiences with others was described as "emotionally cathartic" [47] in a social media–based intervention which encouraged participants to interact on various social media platforms.

However, several participants wished to be able to interact with others throughout the DHI, but the intervention lacked the features to do so [62]. Many studies noted how the DHIs lacked HCP communication [39,41,49,56,58,60,62,63,66]. Some participants preferred being able to interact with their HCP as opposed to a virtual agent:

> *For me, this [program] couldn't replace a session with a therapist. Despite that the program was very valuable, I prefer a face-to-face conversation with a HCP, where you can interact with a real person.* [65]

Despite providing a source of ongoing support and social connectedness for many participants, it was highlighted that DHIs should be used as an adjunct to usual care and should not replace routine face-to-face appointments [35,37,42,52,56,60,63,65,66]: "The app is good progress, but it's not yet a replacement of the physical therapist" [35].

A list of studies adhering to each subtheme can be found in Tables 2-4. Multimedia Appendix 7 provides additional participant quotations from the studies illustrating each theme.

XSL•FO
RenderX

Table 2. Themes identified in each study: personal growth.

Studies including theme 1 (personal growth)	Presence of the subthemes	
	Gaining new insights	Renewed mindset
Algeo et al [34], 2017	✓[a]	
Arensman et al [35], 2022		✓
Austin et al [36], 2020	✓	✓
Ball et al [37], 2020	✓	✓
Bostrom et al [38], 2022	✓	✓
Cronstrom et al [39], 2019	✓	
De Groef et al [65], 2023	✓	✓
Garrett et al [40], 2017	✓	
Garrett et al [67], 2020	✓	✓
Godziuk et al [41], 2023	✓	✓
Grolier et al [68], 2023		
Hogan et al [42], 2022	✓	
Hoving et al [69], 2014	✓	✓
Jeon et al [43], 2019	✓	✓
Kawi et al [44], 2022		
Knoerl et al [45], 2022		
Lamper et al [46], 2021	✓	✓
Merolli et al [47], 2016	✓	
Muehlensiepen et al [48], 2023	✓	✓
Nelligan et al [49], 2020	✓	✓
Nordin et al [50], 2017	✓	✓
Östlind et al [63], 2022	✓	✓
Overton et al [64], 2023	✓	
Rini et al [51], 2018	✓	✓
Schlett et al [52], 2022	✓	✓
Seppen et al [70], 2020	✓	✓
Seppen et al [66], 2023	✓	
Shewchuk et al [53], 2021	✓	
Stern et al [54], 2022	✓	✓
Svanholm et al [55], 2023	✓	✓
Svendsen et al [56], 2022	✓	✓
Tonga et al [57], 2021	✓	
Tonkin-Crine et al [58], 2013	✓	✓
van der Meer et al [59], 2022	✓	✓
Whitney et al [60], 2018	✓	✓
Yeh et al [61], 2022	✓	
Zuidema et al [62], 2019	✓	✓

[a]Subtheme was identified within the article.

XSL•FO
RenderX

Table 3. Themes identified in each study: active involvement.

Studies including theme 2 (active involvement)	Presence of the subthemes		
	Improved access	Motivation	Health care decision-making
Algeo et al [34], 2017		✓[a]	
Arensman et al [35], 2022	✓	✓	✓
Austin et al [36], 2020	✓		✓
Ball et al [37], 2020		✓	
Bostrom et al [38], 2022	✓	✓	
Cronstrom et al [39], 2019	✓	✓	
De Groef et al [65], 2023	✓	✓	
Garrett et al [40], 2017	✓		
Garrett et al [67], 2020			
Godziuk et al [41], 2023	✓	✓	✓
Grolier et al [68], 2023		✓	
Hogan et al [42], 2022			
Hoving et al [69], 2014		✓	
Jeon et al [43], 2019		✓	
Kawi et al [44], 2022	✓	✓	
Knoerl et al [45], 2022	✓	✓	
Lamper et al [46], 2021		✓	✓
Merolli et al [47], 2016	✓	✓	
Muehlensiepen et al [48], 2023		✓	✓
Nelligan et al [49], 2020	✓	✓	
Nordin et al [50], 2017	✓	✓	✓
Östlind et al [63], 2022		✓	✓
Overton et al [64], 2023	✓		
Rini et al [51], 2018	✓	✓	
Schlett et al [52], 2022	✓	✓	✓
Seppen et al [70], 2020	✓	✓	✓
Seppen et al [66], 2023	✓	✓	✓
Shewchuk et al [53], 2021		✓	✓
Stern et al [54], 2022		✓	✓
Svanholm et al [55], 2023		✓	✓
Svendsen et al [56], 2022	✓	✓	
Tonga et al [57], 2021	✓	✓	
Tonkin-Crine et al [58], 2013		✓	
van der Meer et al [59], 2022	✓	✓	
Whitney et al [60], 2018		✓	✓
Yeh et al [61], 2022			
Zuidema et al [62], 2019		✓	

[a]Subtheme was identified within the article.

XSL•FO
RenderX

Table 4. Themes identified in each study: connectedness and support.

Studies including theme 3 (connectedness and support)	Presence of theme
Algeo et al [34], 2017	
Arensman et al [35], 2022	✓[a]
Austin et al [36], 2020	
Ball et al [37], 2020	✓
Bostrom et al [38], 2022	✓
Cronstrom et al [39], 2019	✓
De Groef et al [65], 2023	✓
Garrett et al [40], 2017	
Garrett et al [67], 2020	
Godziuk et al [41], 2023	✓
Grolier et al [68], 2023	
Hogan et al [42], 2022	✓
Hoving et al [69], 2014	✓
Jeon et al [43], 2019	
Kawi et al [44], 2022	
Knoerl et al [45], 2022	✓
Lamper et al [46], 2021	✓
Merolli et al [47], 2016	✓
Muehlensiepen et al [48], 2023	✓
Nelligan et al [49], 2020	✓
Nordin et al [50], 2017	
Östlind et al [63], 2022	✓
Overton et al [64], 2023	
Rini et al [51], 2018	✓
Schlett et al [52], 2022	✓
Seppen et al [70], 2020	
Seppen et al [66], 2023	✓
Shewchuk et al [53], 2021	
Stern et al [54], 2022	✓
Svanholm et al [55], 2023	✓
Svendsen et al [56], 2022	✓
Tonga et al [57], 2021	
Tonkin-Crine et al [58], 2013	✓
van der Meer et al [59], 2022	
Whitney et al [60], 2018	✓
Yeh et al [61], 2022	
Zuidema et al [62], 2019	✓

[a]Subtheme was identified within the article.

Summary of Synthesis

The analytical themes that have been derived from the qualitative synthesis suggest that while many patients report positive experiences of participating in DHIs, the content needs to be tailored to fit the patients' needs otherwise DHIs can have a negative effect on patient outcomes or can lead to poor engagement. This suggests that "a one-size-fits-all" approach

XSL•FO
RenderX

is not always effective in achieving the aims of patients wishing to self-manage their chronic pain symptoms.

DHIs can support the acquisition of knowledge and understanding through improved self-awareness, while encouraging independence and confidence in patients, ultimately stimulating self-efficacy. Accessibility and features that foster motivation, confidence, and promote a sense of ongoing support outside of appointments with HCPs should be prioritized in the future development of DHIs as this is likely to facilitate engagement and improve health outcomes for people living with chronic pain.

Discussion

Overview

This qualitative systematic review aimed to identify, appraise, and synthesize qualitative evidence on the experience of engaging in self-management DHIs among patients with a varying range of chronic pain conditions. The findings revealed that most participants reported a positive experience with DHIs through improved self-awareness, confidence, and independence; through enhanced motivation; and by playing an active role in the management of their chronic pain condition. However, contradictory perspectives exist within each theme, creating a discrepancy as to what features optimize the acceptability of DHIs. These contradictions suggest that for patients to engage with and benefit from digital self-management interventions for chronic pain, it is crucial that content is tailored to the needs of the targeted patient group. The involvement of patients throughout the design and development of DHIs for the self-management of chronic pain could ensure that these needs are met. In the subsequent sections, each theme will be discussed in relation to existing literature and will outline how the results can inform the development and delivery of future DHIs.

Principal Findings

The first analytical theme "personal growth" relates to the personal attributes developed by participating in the DHIs. Patients became more self-aware of how pain impacted other aspects of their life, ultimately leading to a better understanding of their condition and improved pain management [41,58,60,64]. These findings echo the results of a review by Strain et al [28] that explored patients' experience with a range of digital pain management interventions and reported that DHIs enabled people living with chronic pain to become aware of how patterns of behavior could impact their pain symptoms, creating meaningful behavior change. Another review of qualitative studies on user experience with DHIs within psychosis found similar results, suggesting that participants' increased insight allowed them to understand links between their symptoms, mood, and behavior, which equipped them with the tools to better manage their symptoms [71]. On the other hand, many participants did not desire an enhanced level of self-awareness, coinciding with the results of an umbrella review of experiences of DHIs across a range of chronic conditions which found that too much information on a participant's health can be a risk without professional interpretation [72]. This suggests that although features that facilitate self-awareness such as symptom tracking and CBT should be prioritized in the design of future DHIs, caution should be taken during the design phase so that DHIs do not place too much emphasis on negative aspects of health or produce summaries of self-reported symptoms that could be misinterpreted.

Further to the first analytical theme of "personal growth," participant's attitude about their pain improved as patients reported feeling more confident, reassured, and independent. Confidence to communicate pain with others is a crucial factor for people living with chronic pain, who commonly suffer in silence because of barriers to treatment and the associated stigma of a diagnosis [73]. Taylor et al [72] reported similar findings, stating that participants experienced feelings of confidence and reassurance from DHIs across various chronic conditions, as participants could use their logged data to legitimize their reason for contacting an HCP. Patients also became empowered to be more independent [43,51,52,55,56,58,60,65], reporting that they did not have to rely on medication [65] or advice from HCPs [58] as they were able to use self-management strategies to manage their pain. Due to the demands and challenges of living with chronic pain, it is common for patients to lose their sense of "self," as many individuals are unable to upkeep their lifestyle and continue with activities they previously enjoyed [74]. By promoting confidence and enabling participants to become more independent, DHIs can inspire patients to have a more optimistic outlook on life, encouraging them to maintain their usual activities without fear of exacerbating their pain which will ultimately improve their QoL.

Another notable experience was that participants reported gaining a degree of acceptance regarding their diagnosis [38,43,46,51,52,56,69]. Research supports that pain acceptance and optimism are crucial steps toward self-management [75]. A systematic review of the concept of "self" among patients with chronic pain reported that negative evaluations of oneself can lead to more complications with daily functioning, and when an individual views their sense of "self" as more positive, they are more likely to experience better functioning [74]. This idea is further supported by research that states improved self-efficacy and pain acceptance are effective strategies in reducing future pain disability and protecting individuals against chronic pain [76,77]. The findings from this review suggest that patients with chronic pain can benefit from improved pain self-efficacy and acceptance of their diagnosis when engaging with self-management DHIs, and this could carry massive potential in reducing the future burden of pain.

Second, "active involvement" emerged as an analytical theme. This theme highlighted how participants were motivated to engage in positive health behaviors through increased access to self-management resources, as they were able to actively manage their pain outside of a health care setting. Strain et al [28] also reported that digital pain management interventions encouraged participants to play a more active role in their care, strengthening the idea that digital health can be an effective means of alleviating the strain on the health care system by encouraging people living with chronic pain to take control into their own hands. The importance of taking action to reduce chronic pain symptoms was demonstrated in a population-based study, which showed that patients who relied on passive coping

XSL•FO
RenderX

strategies (such as pain medication and rest) had as many as 3 times the amount of health care appointments and double the level of disability compared to those who adopted active coping strategies such as exercise [78]. In particular, exercise-based interventions that encouraged participants to hit daily exercise targets were found to be motivating [35,41,52,56-58,62,63,69]. In many of the studies, notifications motivated participants to engage with the digital tool [34,35,39,41,43,49,57,59,63], a finding consistent with other reviews on patients' experiences with DHIs [71,72]. On the other hand, notifications caused unnecessary distress for some, particularly when targets had not been reached [41,49]. By stimulating motivation to engage in health behaviors that could improve levels of pain, DHIs could potentially play a major role in improving pain-related outcomes in people living with chronic pain. However, it is important to set realistic and achievable goals, as unobtainable targets run the risk of creating an adverse effect on motivation [63]. These findings suggest that participants should be given the option to customize the frequency of notifications to align with their treatment goals, as this will avoid notifications having a negative effect and hindering engagement.

Access to specialized pain services poses a barrier to effective treatment, with many patients reporting low satisfaction with the quality of pain management care they receive because of long waiting times and lack of interest from HCPs [8]. The flexibility of being able to choose when and where to engage in the various self-management techniques motivated participants to take control of their symptoms at a time that suited them [38,39,50,51,59,65].

Another way that the participants felt that they were playing a more active role in their health care was by contributing more during consultations with HCPs. Equipping patients with tools that enable them to actively participate in health care discussions will ensure that their treatment is personalized to their specific needs, which has the potential to improve various health outcomes.

The third analytical theme "connectedness and support" outlines how DHIs were found to be useful in fostering a sense of ongoing support between health care appointments and at the end of treatment [35,55,65]. Social support is crucial in the treatment of chronic pain as research indicates that higher levels of social support are associated with lower levels of pain [79]. A systematic review exploring factors affecting engagement suggested that technology supporting human interaction with both peers and HCPs could promote engagement and enrollment as this allows users to obtain social support quickly and effectively [80]. The findings from this review demonstrate that DHIs are effective tools for facilitating social support, although social support in the included articles was perceived through time spent with family and friends and communication with virtual avatars embedded into the interventions, which many participants referred to as a virtual "friend." Most studies included in the review lacked features that allowed participants to directly interact with other patients within the intervention. In the one study that facilitated sharing experiences with others, participants described being connected with others as "emotionally cathartic" [47], although it should be highlighted that this study scored poorly on the CASP appraisal tool; therefore, these findings should be interpreted with caution. This is in line with the findings of a review exploring social media uses for various health purposes, which reported that online communities can provide invaluable advice by offering emotional support as well as sharing information such as medication, diagnosis, treatment options, and self-care activities [81].

Chronic pain is often regarded as an "invisible illness," with many patients having faced skepticism and suspicion of exaggeration or even accused of lying about their condition [82-85]. Therefore, it is important that patients with chronic pain obtain support from others in a similar position as only those with lived experience can truly comprehend the extent to which chronic pain interferes with daily life. Strain et al [28] reported similar findings, stating that interactions with other users or with HCPs have the potential to reduce anxiety and loneliness, although the anonymity offered through DHIs was also valued; therefore, it could be that users have the option to stay anonymous. Features enabling communication with HCPs might facilitate patient engagement and continued use as many patients reported an impersonal feel to DHIs as they were unable to receive feedback and did not experience accountability. Others mentioned a preference for being able to interact with an HCP as opposed to a virtual agent [65]. The inclusion of interactive features could stimulate the sense of ongoing support; however, DHIs should enhance traditional face-to-face care as opposed to serving as a replacement. Future DHIs could incorporate an element of interaction with HCPs that could be continued into routine appointments allowing for more effective consultations.

Limitations

This review should be interpreted with caution, as several limitations exist. First, this synthesis only included studies from 2013 onward in an attempt to capture the most relevant and up-to-date DHI systems that have been developed over the last decade. Consequently, the findings reported in this review may not reflect the experiences of patients with chronic pain who engaged with digital self-management interventions before 2013, which could be dramatically different due to advancements in digital technology. Second, as with all qualitative syntheses, this review relies on secondary data that were analyzed through a varying range of theoretical underpinnings and methods of data analysis [86]. Another limitation of this review is that the diversity of the samples, which included a range of chronic pain conditions, made it challenging to identify which DHI would be the most beneficial for each individual patient group. Furthermore, the DHIs encompassed multiple intervention components, which were often intertwined and difficult to separate in the results. Therefore, this review could not always draw conclusions about which intervention components elicited which experiences, reducing the richness of the findings. Another limitation was that many of the studies did not report the frequency and duration of the intervention, nor how long after the intervention the qualitative data were collected. Thus, it is unknown whether the timing or frequency of the DHI influenced the patient experience. Of the 33 included studies that did detail the intervention length, only 1 (3%) study was conducted for 12

XSL•FO
RenderX

months, with the remaining studies lasting ≤6 months; therefore, the long-term experience of DHIs for the self-management of chronic pain could not be synthesized and remains not fully understood. Future trials should explore the experience of DHIs for the self-management of chronic pain over a period of 2 years to fully grasp the long-term experience of these interventions. The largest sample size was 36 participants, with 86% (32/37) of studies obtaining samples of <30 participants. It is important to consider sampling bias that may have occurred among the studies with small samples, and that the results may not be representative of all patients with chronic pain. Future studies should involve larger samples so that the results can be generalized to the larger population. Another limitation to consider is that most of the studies were conducted in high-income countries, as the implementation of DHIs in low- and middle-income countries presents several challenges, such as resource limitations and costs [87].

Conclusions

The findings from this systematic review and thematic synthesis suggest that DHIs for the self-management of chronic pain elicit mostly positive experiences among people living with chronic pain; however, opinions on which features are optimal for improving pain outcomes are diverse, highlighting that a "one-size-fits-all" approach is not effective. Personalized content should be incorporated to ensure that DHIs address the needs of this patient group, and content should be customizable to avoid negative effects on adoption and engagement. This could be achieved by involving patients in the early stages of intervention design and development; therefore, a co-design approach for the creation of future DHIs for the self-management of chronic pain is advised. The results of this review can be used to inform the development of future DHIs, suggesting that such DHIs should be accessible, convenient, and flexible and be easily integrated into patients' lives without exacerbating the burden on those already coping with a range of challenging symptoms. Intervention components, including CBT, symptom tracking, exercise programs, and educational content should be prioritized. These features have demonstrated improvements in patients' self-awareness, confidence, independence, and acceptance, all while fostering motivation to engage in positive health behaviors, decisions about treatment and care with HCPs, and the feeling of being supported outside of routine health care appointments. The findings from this review point to a lack of social networking functionalities with both peers and HCPs in DHIs for the self-management of chronic pain, suggesting a potential gap in the research. Further investigation surrounding the availability of such social features should be conducted, as social networking functionalities could provide a means of support that could relieve the burden on both HCPs and services.

Acknowledgments

The authors would like to acknowledge the University of Strathclyde librarians Elaine Blair from the Faculty of Science and Sarah Kevill from Humanities and Social Sciences for their invaluable assistance in developing the search strategies for this review. This systematic review is part of an ongoing PhD studentship funded by the Dr David Stirling Health Research Fund aimed at supporting research into chronic pain.

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). Additional material can be obtained from the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Enhancing Transparency in Reporting the Synthesis of Qualitative Research (ENTREQ) checklist.
[DOC File , 46 KB - jmir_v27i1e69100_app1.doc]

Multimedia Appendix 2
Search strategy for each database.
[DOC File , 40 KB - jmir_v27i1e69100_app2.doc]

Multimedia Appendix 3
PRISMA checklist.
[DOCX File , 32 KB - jmir_v27i1e69100_app3.docx]

Multimedia Appendix 4
Summary of study characteristics from each qualitative or mixed methods study included in the review (N=37).
[DOC File , 177 KB - jmir_v27i1e69100_app4.doc]

XSL•FO
RenderX

Multimedia Appendix 5
Summary of intervention characteristics from each qualitative or mixed methods study included in the review (N=37).
[DOC File , 104 KB - jmir_v27i1e69100_app5.doc]

Multimedia Appendix 6
The Critical Appraisal Skills Program (CASP) qualitative research checklist.
[DOC File , 92 KB - jmir_v27i1e69100_app6.doc]

Multimedia Appendix 7
A table of quotes from the included studies illustrating each theme and subtheme identified within the data.
[DOC File , 40 KB - jmir_v27i1e69100_app7.doc]

References

1. International Classification of Diseases, 11th Revision. World Health Organization. URL: https://icd.who.int/en [accessed 2024-04-29]
2. Mills SE, Nicolson KP, Smith BH. Chronic pain: a review of its epidemiology and associated factors in population-based studies. Br J Anaesth 2019 Aug;123(2):e273-e283 [FREE Full text] [doi: 10.1016/j.bja.2019.03.023] [Medline: 31079836]
3. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng193 [accessed 2024-04-29]
4. Fayaz A, Croft P, Langford RM, Donaldson LJ, Jones GT. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. BMJ Open 2016 Jun 20;6(6):e010364 [FREE Full text] [doi: 10.1136/bmjopen-2015-010364] [Medline: 27324708]
5. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021 May 29;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]
6. Hooten WM. Chronic pain and mental health disorders: shared neural mechanisms, epidemiology, and treatment. Mayo Clin Proc 2016 Jul;91(7):955-970. [doi: 10.1016/j.mayocp.2016.04.029] [Medline: 27344405]
7. Staudt MD. The multidisciplinary team in pain management. Neurosurg Clin N Am 2022 Jul;33(3):241-249. [doi: 10.1016/j.nec.2022.02.002] [Medline: 35718393]
8. Hadi MA, Alldred DP, Briggs M, Marczewski K, Closs SJ. 'Treated as a number, not treated as a person': a qualitative exploration of the perceived barriers to effective pain management of patients with chronic pain. BMJ Open 2017 Jun 12;7(6):e016454 [FREE Full text] [doi: 10.1136/bmjopen-2017-016454] [Medline: 28606909]
9. Lynch ME, Campbell F, Clark AJ, Dunbar MJ, Goldstein D, Peng P, et al. A systematic review of the effect of waiting for treatment for chronic pain. Pain 2008 May;136(1-2):97-116. [doi: 10.1016/j.pain.2007.06.018] [Medline: 17707589]
10. Hogg MN, Gibson S, Helou A, DeGabriele J, Farrell MJ. Waiting in pain: a systematic investigation into the provision of persistent pain services in Australia. Med J Aust 2012 Apr 02;196(6):386-390. [doi: 10.5694/mja12.10140] [Medline: 22471539]
11. Burke AL, Mathias JL, Denson LA. Waiting for multidisciplinary chronic pain services: a prospective study over 2.5 years. J Health Psychol 2020 Aug;25(9):1198-1212. [doi: 10.1177/1359105317752828] [Medline: 29322830]
12. Fashler SR, Cooper LK, Oosenbrug ED, Burns LC, Razavi S, Goldberg L, et al. Systematic review of multidisciplinary chronic pain treatment facilities. Pain Res Manag 2016;2016:5960987-5960919 [FREE Full text] [doi: 10.1155/2016/5960987] [Medline: 27445618]
13. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
14. Mann EG, Lefort S, Vandenkerkhof EG. Self-management interventions for chronic pain. Pain Manag 2013 May;3(3):211-222 [FREE Full text] [doi: 10.2217/pmt.13.9] [Medline: 24654764]
15. Hestmann R, Bratås O, Grønning K. Chronic pain self-management interventions in primary care - does it make any difference? A qualitative study. BMC Health Serv Res 2023 May 24;23(1):537 [FREE Full text] [doi: 10.1186/s12913-023-09548-8] [Medline: 37226178]
16. Geraghty AW, Maund E, Newell D, Santer M, Everitt H, Price C, et al. Self-management for chronic widespread pain including fibromyalgia: a systematic review and meta-analysis. PLoS One 2021;16(7):e0254642 [FREE Full text] [doi: 10.1371/journal.pone.0254642] [Medline: 34270606]
17. Huber S, Priebe JA, Baumann KM, Plidschun A, Schiessl C, Tölle TR. Treatment of low back pain with a digital multidisciplinary pain treatment app: short-term results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]
18. Hussain A, Haroon H, Ahmed A, Gilani SA. Digital technologies in management of chronic pain - a systematic review. J Pak Med Assoc 2022 Jun;72(6):1158-1165 [FREE Full text] [doi: 10.47391/JPMA.3885] [Medline: 35751328]
19. Rejula V, Anitha J, Belfin RV, Peter JD. Chronic pain treatment and digital health era-an opinion. Front Public Health 2021 Dec 10;9:779328 [FREE Full text] [doi: 10.3389/fpubh.2021.779328] [Medline: 34957031]

XSL•FO
RenderX

20. Pfeifer AC, Uddin R, Schröder-Pfeifer P, Holl F, Swoboda W, Schiltenwolf M. Mobile application-based interventions for chronic pain patients: a systematic review and meta-analysis of effectiveness. J Clin Med 2020 Nov 05;9(11):3557 [FREE Full text] [doi: 10.3390/jcm9113557] [Medline: 33167300]
21. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
22. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023 Feb 02;11:e40844 [FREE Full text] [doi: 10.2196/40844] [Medline: 36729570]
23. Macea DD, Gajos K, Daglia Calil YA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929 [FREE Full text] [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
24. Patten RK, Tacey A, Pile R, Parker A, De Gori M, Tran P, et al. Digital self-management interventions for osteoarthritis: a systematic scoping review of intervention characteristics, adherence and attrition. Arch Public Health 2022 Mar 31;80(1):103 [FREE Full text] [doi: 10.1186/s13690-022-00854-x] [Medline: 35361270]
25. Cranen K, Groothuis-Oudshoorn CG, Vollenbroek-Hutten MM, IJzerman MJ. Toward patient-centered telerehabilitation design: understanding chronic pain patients' preferences for web-based exercise telerehabilitation using a discrete choice experiment. J Med Internet Res 2017 Jan 20;19(1):e26 [FREE Full text] [doi: 10.2196/jmir.5951] [Medline: 28108429]
26. Fernandes LG, Devan H, Fioratti I, Kamper SJ, Williams CM, Saragiotto BT. At my own pace, space, and place: a systematic review of qualitative studies of enablers and barriers to telehealth interventions for people with chronic pain. Pain 2022 Feb 01;163(2):e165-e181. [doi: 10.1097/j.pain.0000000000002364] [Medline: 34433776]
27. Svendsen MJ, Wood KW, Kyle J, Cooper K, Rasmussen CD, Sandal LF, et al. Barriers and facilitators to patient uptake and utilisation of digital interventions for the self-management of low back pain: a systematic review of qualitative studies. BMJ Open 2020 Dec 12;10(12):e038800 [FREE Full text] [doi: 10.1136/bmjopen-2020-038800] [Medline: 33310794]
28. Strain JD, Welch L, Sadler E. Systematic review and narrative synthesis of the experiences of individuals with chronic pain participating in digital pain management interventions. PLoS One 2024;19(7):e0306455 [FREE Full text] [doi: 10.1371/journal.pone.0306455] [Medline: 38976713]
29. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
30. Main A, McCartney H, Ibrar M, Rai HK, Muirhead F, Mavroeidi A, et al. Patients' experiences of digital health interventions for the self-management of chronic pain: protocol for a systematic review and thematic synthesis. JMIR Res Protoc 2024 Mar 07;13:e52469 [FREE Full text] [doi: 10.2196/52469] [Medline: 38451694]
31. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
32. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Glob Health 2019;4(Suppl 1):e000893 [FREE Full text] [doi: 10.1136/bmjgh-2018-000893] [Medline: 30775016]
33. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
34. Algeo N, Hunter D, Cahill A, Dickson C, Adams J. Usability of a digital self-management website for people with osteoarthritis: a UK patient and public involvement study. Int J Ther Rehabil 2017 Feb 02;24(2):78-82. [doi: 10.12968/ijtr.2017.24.2.78]
35. Arensman R, Kloek C, Pisters M, Koppenaal T, Ostelo R, Veenhof C. Patient perspectives on using a smartphone app to support home-based exercise during physical therapy treatment: qualitative study. JMIR Hum Factors 2022 Sep 13;9(3):e35316 [FREE Full text] [doi: 10.2196/35316] [Medline: 36098993]
36. Austin L, Sharp CA, van der Veer SN, Machin M, Humphreys J, Mellor P, et al. Providing 'the bigger picture': benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. Rheumatology (Oxford) 2020 Feb 01;59(2):367-378 [FREE Full text] [doi: 10.1093/rheumatology/kez207] [Medline: 31335942]
37. Ball E, Newton S, Rohricht F, Steed L, Birch J, Dodds J, et al. mHealth: providing a mindfulness app for women with chronic pelvic pain in gynaecology outpatient clinics: qualitative data analysis of user experience and lessons learnt. BMJ Open 2020 Mar 12;10(3):e030711 [FREE Full text] [doi: 10.1136/bmjopen-2019-030711] [Medline: 32165550]
38. Bostrøm K, Varsi C, Eide H, Børøsund E, Kristjansdottir Ó, Schreurs KM, et al. Engaging with EPIO, a digital pain self-management program: a qualitative study. BMC Health Serv Res 2022 Apr 29;22(1):577 [FREE Full text] [doi: 10.1186/s12913-022-07963-x] [Medline: 35488295]
39. Cronström A, Dahlberg LE, Nero H, Ericson J, Hammarlund CS. 'I would never have done it if it hadn't been digital': a qualitative study on patients' experiences of a digital management programme for hip and knee osteoarthritis in Sweden. BMJ Open 2019 May 24;9(5):e028388 [FREE Full text] [doi: 10.1136/bmjopen-2018-028388] [Medline: 31129601]
40. Garrett B, Taverner T, McDade P. Virtual reality as an adjunct home therapy in chronic pain management: an exploratory study. JMIR Med Inform 2017 May 11;5(2):e11 [FREE Full text] [doi: 10.2196/medinform.7271] [Medline: 28495661]

XSL•FO
RenderX

41. Godziuk K, Prado CM, Quintanilha M, Forhan M. Acceptability and preliminary effectiveness of a single-arm 12-week digital behavioral health intervention in patients with knee osteoarthritis. BMC Musculoskelet Disord 2023 Feb 17;24(1):129 [FREE Full text] [doi: 10.1186/s12891-023-06238-8] [Medline: 36797720]
42. Hogan TP, Etingen B, McMahon N, Bixler FR, Am L, Wacks RE, et al. Understanding adoption and preliminary effectiveness of a mobile app for chronic pain management among US military veterans: pre-post mixed methods evaluation. JMIR Form Res 2022 Jan 20;6(1):e33716 [FREE Full text] [doi: 10.2196/33716] [Medline: 35049515]
43. Jeon YH, Flaherty I, Urban H, Wortley S, Dickson C, Salkeld G, et al. Qualitative evaluation of evidence-based online decision aid and resources for osteoarthritis management: understanding patient perspectives. Arthritis Care Res (Hoboken) 2019 Jan;71(1):46-55. [doi: 10.1002/acr.23572] [Medline: 29609208]
44. Kawi J, Yeh CH, Lukkahatai N, Hardwicke RL, Murphy T, Christo PJ. Exploring the feasibility of virtually delivered auricular point acupressure in self-managing chronic pain: qualitative study. Evid Based Complement Alternat Med 2022;2022:8079691 [FREE Full text] [doi: 10.1155/2022/8079691] [Medline: 36072397]
45. Knoerl R, Bockhoff J, Fox E, Giobbie-Hurder A, Berry DL, Berfield J, et al. Cancer survivors' perspectives of virtual yoga for chronic chemotherapy-induced peripheral neuropathy pain during the COVID-19 pandemic. Comput Inform Nurs 2022 Sep 01;40(9):641-647 [FREE Full text] [doi: 10.1097/CIN.0000000000000937] [Medline: 35930413]
46. Lamper C, Huijnen I, de Mooij M, Köke A, Verbunt J, Kroese M. An eCoach-pain for patients with chronic musculoskeletal pain in interdisciplinary primary care: a feasibility study. Int J Environ Res Public Health 2021 Nov 06;18(21):11661 [FREE Full text] [doi: 10.3390/ijerph182111661] [Medline: 34770177]
47. Merolli M, Gray K, Martin-Sanchez F. Patient participation in chronic pain management through social media: a clinical study. Stud Health Technol Inform 2016;225:577-581. [Medline: 27332267]
48. Muehlensiepen F, May S, Hadaschik K, Vuillerme N, Heinze M, Grahammer M, et al. Digitally supported shared decision-making and treat-to-target in rheumatology: a qualitative study embedded in a multicenter randomized controlled trial. Rheumatol Int 2023 Apr;43(4):695-703 [FREE Full text] [doi: 10.1007/s00296-022-05224-y] [Medline: 36229671]
49. Nelligan RK, Hinman RS, Teo PL, Bennell KL. Exploring attitudes and experiences of people with knee osteoarthritis toward a self-directed eHealth intervention to support exercise: qualitative study. JMIR Rehabil Assist Technol 2020 Nov 26;7(2):e18860 [FREE Full text] [doi: 10.2196/18860] [Medline: 33242021]
50. Nordin C, Michaelson P, Eriksson MK, Gard G. It's about me: patients' experiences of patient participation in the web behavior change program for activity in combination with multimodal pain rehabilitation. J Med Internet Res 2017 Jan 18;19(1):e22. [doi: 10.2196/jmir.5970] [Medline: 28100440]
51. Rini C, Vu MB, Lerner H, Bloom C, Carda-Auten J, Wood WA, et al. A qualitative study of patient and provider perspectives on using web-based pain coping skills training to treat persistent cancer pain. Palliat Support Care 2018 Apr;16(2):155-169. [doi: 10.1017/S1478951517000086] [Medline: 28264744]
52. Schlett C, Röttele N, van der Keylen P, Schöpf-Lazzarino AC, Klimmek M, Körner M, et al. The acceptance, usability, and utility of a web portal for back pain as recommended by primary care physicians: qualitative interview study with patients. JMIR Form Res 2022 Dec 29;6(12):e38748 [FREE Full text] [doi: 10.2196/38748] [Medline: 36580365]
53. Shewchuk B, Green LA, Barber T, Miller J, Teare S, Campbell-Scherer D, et al. Patients' use of mobile health for self-management of knee osteoarthritis: results of a 6-week pilot study. JMIR Form Res 2021 Nov 25;5(11):e30495 [FREE Full text] [doi: 10.2196/30495] [Medline: 34842526]
54. Stern BZ, Pila S, Joseph LI, Rothrock NE, Franklin PD. Patients' perspectives on the benefits of feedback on patient-reported outcome measures in a web-based personalized decision report for hip and knee osteoarthritis. BMC Musculoskelet Disord 2022 Aug 23;23(1):806 [FREE Full text] [doi: 10.1186/s12891-022-05764-1] [Medline: 35999585]
55. Svanholm F, Turesson C, Löfgren M, Björk M. Acceptability of the eHealth intervention sustainable worker digital support for persons with chronic pain and their employers (SWEPPE): questionnaire and interview study. JMIR Hum Factors 2023 Sep 28;10:e46878 [FREE Full text] [doi: 10.2196/46878] [Medline: 37768708]
56. Svendsen MJ, Nicholl BI, Mair FS, Wood K, Rasmussen CD, Stochkendahl MJ. One size does not fit all: participants' experiences of the selfBACK app to support self-management of low back pain-a qualitative interview study. Chiropr Man Therap 2022 Oct 03;30(1):41 [FREE Full text] [doi: 10.1186/s12998-022-00452-2] [Medline: 36192724]
57. Tonga E, Williamson E, Srikesavan C, Özen T, Sarıtaş F, Lamb SE. A hand exercise mobile app for people with rheumatoid arthritis in Turkey: design, development and usability study. Rheumatol Int 2021 Jun 18;41(6):1151-1160. [doi: 10.1007/s00296-021-04860-0] [Medline: 33870452]
58. Tonkin-Crine S, Bishop FL, Ellis M, Moss-Morris R, Everitt H. Exploring patients' views of a cognitive behavioral therapy-based website for the self-management of irritable bowel syndrome symptoms. J Med Internet Res 2013 Sep 03;15(9):e190 [FREE Full text] [doi: 10.2196/jmir.2672] [Medline: 24001787]
59. van der Meer HA, de Pijper L, van Bruxvoort T, Visscher CM, Nijhuis-van der Sanden MW, Engelbert RH, et al. Using e-Health in the physical therapeutic care process for patients with temporomandibular disorders: a qualitative study on the perspective of physical therapists and patients. Disabil Rehabil 2022 Feb 16;44(4):617-624 [FREE Full text] [doi: 10.1080/09638288.2020.1775900] [Medline: 32543903]

XSL•FO
RenderX

60. Whitney RL, Ward DH, Marois MT, Schmid CH, Sim I, Kravitz RL. Patient perceptions of their own data in mHealth technology-enabled N-of-1 trials for chronic pain: qualitative study. JMIR Mhealth Uhealth 2018 Oct 11;6(10):e10291 [FREE Full text] [doi: 10.2196/10291] [Medline: 30309834]
61. Yeh CH, Kawi J, Ni A, Christo P. Evaluating auricular point acupressure for chronic low back pain self-management using technology: a feasibility study. Pain Manag Nurs 2022 Jun;23(3):301-310. [doi: 10.1016/j.pmn.2021.11.007] [Medline: 34961729]
62. Zuidema RM, Van Dulmen S, Nijhuis-van der Sanden MW, Fransen J, Van Gaal BG. qLessons learned from patients with access to an online self-management enhancing program for RA patients: Qualitative analysis of interviews alongside a randomized clinical trial. Patient Educ Couns 2019 Jun;102(6):1170-1177. [doi: 10.1016/j.pec.2019.01.005] [Medline: 30638903]
63. Östlind E, Ekvall Hansson E, Eek F, Stigmar K. Experiences of activity monitoring and perceptions of digital support among working individuals with hip and knee osteoarthritis - a focus group study. BMC Public Health 2022 Aug 30;22(1):1641 [FREE Full text] [doi: 10.1186/s12889-022-14065-0] [Medline: 36042425]
64. Overton M, Swain N, Falling C, Gwynne-Jones D, Fillingim R, Mani R. Experiences and perceptions of using smartphone ecological momentary assessment for reporting knee osteoarthritis pain and symptoms. Clin J Pain 2023 Sep 01;39(9):442-451. [doi: 10.1097/AJP.0000000000001138] [Medline: 37335088]
65. De Groef A, Evenepoel M, Van Dijck S, Dams L, Haenen V, Wiles L, et al. Feasibility and pilot testing of a personalized eHealth intervention for pain science education and self-management for breast cancer survivors with persistent pain: a mixed-method study. Support Care Cancer 2023 Jan 16;31(2):119. [doi: 10.1007/s00520-022-07557-7] [Medline: 36645526]
66. Seppen BF, Wiegel J, Nurmohamed MT, Bos WH, Ter Wee MM. Facilitators and barriers to adhere to monitoring disease activity with ePROs: a focus group study in patients with inflammatory arthritis. Rheumatol Int 2023 Apr;43(4):677-685 [FREE Full text] [doi: 10.1007/s00296-022-05263-5] [Medline: 36627396]
67. Garrett BM, Tao G, Taverner T, Cordingley E, Sun C. Patients perceptions of virtual reality therapy in the management of chronic cancer pain. Heliyon 2020 May;6(5):e03916 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03916] [Medline: 32426540]
68. Grolier M, Arefyev A, Pereira B, Tavares Figueiredo I, Gerbaud L, Coudeyre E. Refining the design of a smartphone application for people with chronic low back pain using mixed quantitative and qualitative approaches. Disabil Rehabil Assist Technol 2023 Feb;18(2):145-150. [doi: 10.1080/17483107.2020.1839575] [Medline: 33151760]
69. Hoving JL, Zoer I, van der Meer M, van der Straaten Y, Logtenberg-Rutten C, Kraak-Put S, et al. E-health to improve work functioning in employees with rheumatoid arthritis in rheumatology practice: a feasibility study. Scand J Rheumatol 2014 Sep 02;43(6):481-487. [doi: 10.3109/03009742.2014.914568] [Medline: 25178316]
70. Seppen BF, Wiegel J, L'ami MJ, Duarte Dos Santos Rico S, Catarinella FS, Turkstra F, et al. Feasibility of self-monitoring rheumatoid arthritis with a smartphone app: results of two mixed-methods pilot studies. JMIR Form Res 2020 Sep 21;4(9):e20165 [FREE Full text] [doi: 10.2196/20165] [Medline: 32955447]
71. Dennard S, Patel R, Garety P, Edwards C, Gumley A. A systematic review of users experiences of using digital interventions within psychosis: a thematic synthesis of qualitative research. Soc Psychiatry Psychiatr Epidemiol (Forthcoming) 2024 May 27. [doi: 10.1007/s00127-024-02692-4] [Medline: 38802509]
72. Taylor ML, Thomas EE, Vitangcol K, Marx W, Campbell KL, Caffery LJ, et al. Digital health experiences reported in chronic disease management: an umbrella review of qualitative studies. J Telemed Telecare 2022 Dec;28(10):705-717. [doi: 10.1177/1357633X221119620] [Medline: 36346938]
73. Katzman JG, Gallagher RM. Pain: the silent public health epidemic. J Prim Care Community Health 2024;15:21501319241253547 [FREE Full text] [doi: 10.1177/21501319241253547] [Medline: 38742616]
74. Yu L, Norton S, Harrison A, McCracken LM. In search of the person in pain: a systematic review of conceptualization, assessment methods, and evidence for self and identity in chronic pain. J Contextual Behav Sci 2015 Oct;4(4):246-262. [doi: 10.1016/j.jcbs.2015.10.003]
75. Wright MA, Wren AA, Somers TJ, Goetz MC, Fras AM, Huh BK, et al. Pain acceptance, hope, and optimism: relationships to pain and adjustment in patients with chronic musculoskeletal pain. J Pain 2011 Nov;12(11):1155-1162 [FREE Full text] [doi: 10.1016/j.jpain.2011.06.002] [Medline: 21820969]
76. Karasawa Y, Yamada K, Iseki M, Yamaguchi M, Murakami Y, Tamagawa T, et al. Association between change in self-efficacy and reduction in disability among patients with chronic pain. PLoS One 2019 Apr 16;14(4):e0215404 [FREE Full text] [doi: 10.1371/journal.pone.0215404] [Medline: 30990842]
77. Bérubé M, Choinière M, Laflamme YG, Gélinas C. Acute to chronic pain transition in extremity trauma: a narrative review for future preventive interventions (part 2). Int J Orthop Trauma Nurs 2017 Feb;24:59-67. [doi: 10.1016/j.ijotn.2016.04.001] [Medline: 27527536]
78. Blyth FM, March LM, Nicholas MK, Cousins MJ. Self-management of chronic pain: a population-based study. Pain 2005 Feb;113(3):285-292. [doi: 10.1016/j.pain.2004.12.004] [Medline: 15661435]
79. Macdonald G, Leary MR. Why does social exclusion hurt? The relationship between social and physical pain. Psychol Bull 2005 Mar;131(2):202-223. [doi: 10.1037/0033-2909.131.2.202] [Medline: 15740417]

XSL•FO
RenderX

80. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
81. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
82. Koesling D, Bozzaro C. Chronic pain patients' need for recognition and their current struggle. Med Health Care Philos 2021 Dec 14;24(4):563-572 [FREE Full text] [doi: 10.1007/s11019-021-10040-5] [Medline: 34259964]
83. Goldberg DS. Job and the stigmatization of chronic pain. Perspect Biol Med 2010;53(3):425-438. [doi: 10.1353/pbm.0.0166] [Medline: 20639609]
84. Wakefield EO, Belamkar V, Litt MD, Puhl RM, Zempsky WT. "There's nothing wrong with you": pain-related stigma in adolescents with chronic pain. J Pediatr Psychol 2022 Apr 08;47(4):456-468 [FREE Full text] [doi: 10.1093/jpepsy/jsab122] [Medline: 34871426]
85. García-Rodríguez MI, Biedma-Velázquez L, Serrano-Del-Rosal R. The legitimacy of pain according to sufferers. PLoS One 2023;18(11):e0291112 [FREE Full text] [doi: 10.1371/journal.pone.0291112] [Medline: 37967079]
86. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodol 2009 Aug 11;9:59 [FREE Full text] [doi: 10.1186/1471-2288-9-59] [Medline: 19671152]
87. Duggal M, El Ayadi A, Duggal B, Reynolds N, Bascaran C. Editorial: challenges in implementing digital health in public health settings in low and middle income countries. Front Public Health 2022 Jan 10;10:1090303 [FREE Full text] [doi: 10.3389/fpubh.2022.1090303] [Medline: 36703825]

Abbreviations

CASP: Critical Appraisal Skills Program
CBT: cognitive behavioral therapy
DHI: digital health intervention
ENTREQ: Enhancing Transparency in Reporting the Synthesis of Qualitative Research
HCP: health care professional
QoL: quality of life
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SPIDER: sample, phenomenon of interest, design, evaluation, and research

Edited by A Mavragani; submitted 27.11.24; peer-reviewed by E Osborn, L Welch; comments to author 10.01.25; revised version received 22.01.25; accepted 24.01.25; published 18.03.25.

Please cite as:
Main A, McCartney H, Ibrar M, Muirhead F, Mavroeidi A, Rai HK, Maguire R
Patients' Experiences of Digital Health Interventions for the Self-Management of Chronic Pain: Systematic Review and Thematic Synthesis
J Med Internet Res 2025;27:e69100
URL: https://www.jmir.org/2025/1/e69100
doi:10.2196/69100
PMID:

©Ashleigh Main, Haruno McCartney, Maryam Ibrar, Fiona Muirhead, Alexandra Mavroeidi, Harleen Kaur Rai, Roma Maguire. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Health Education for Chronic Lung Disease: Scoping Review

Chao Sun[1,2], MPH; Huohuo Dai[1,2], MSc; Rianne M.J.J. van der Kleij[1,2], PhD; Rong Li[3], BSc; Hengchang Wu[4], MPH; Cynthia Hallensleben[1,2], BSc; Sofie H Willems[1,2], MD, PhD; Niels H Chavannes[1,2], MD, PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

[2]National eHealth Living Lab, Leiden, The Netherlands

[3]School of Pharmacy, Hunan University of Chinese Medicine, Changsha, China

[4]School of Public Health, Wuhan University, Wuhan, China

Corresponding Author:
Chao Sun, MPH
Department of Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Leiden, 2300 RC
The Netherlands
Phone: 31 0623826614
Email: c.sun@lumc.nl

Abstract

Background: Chronic lung disease (CLD) is one of the most prevalent noncommunicable diseases globally, significantly burdening patients and increasing health care expenditures. Digital health education (DHE) is increasingly important in chronic disease prevention and management. However, DHE characteristics and impacts in CLD are rarely reported.

Objective: This study aimed to provide an overview of the existing literature on DHE for CLD, with a focus on exploring the DHE mediums, content, mechanisms, and reported outcomes in patients with CLD.

Methods: We searched PubMed, Web of Science, Embase, PsycINFO, and The Cochrane Library with the assistance of a librarian specialist. Articles were screened by the reviewer team with ASReview (Utrecht University) and EndNote X9 (Clarivate Analytics) based on predefined inclusion and exclusion criteria and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. Quality assessment was conducted with the Critical Appraisal Skills Program tool. A descriptive analysis was used to summarize the study characteristics, DHE characteristics, and outcomes.

Results: A total of 22 studies were included in this review with medium or high quality. They were published between 2000 and 2022, showing an increasing publication trend with the year, mostly in developed countries (16/22, 73%). Websites and mobile apps (10/22, 45%) were the most widely used DHE medium. Education on self-management skills of CLD was the primary topic (21/22, 95%), 4/22 (18%) of which mentioned DHE mechanisms. The majority of studies reported positive changes in CLD awareness (14/16, 88%), clinical outcomes (3/6, 50%), DHE feasibility, acceptability, and satisfaction (6/8, 75%), lifestyle outcomes (3/3, 100%), and psychosocial outcomes (7/8, 88%). Only 2 studies reported cost-effectiveness (2/22, 9%).

Conclusions: Despite the heterogeneity of the study situation, some aspects can be concluded. DHE can improve disease awareness and clinical outcomes in patients with chronic lung disease, with good feasibility, acceptability, and satisfaction through different mediums and learning content. There is still relatively little research among people in low- and middle-income countries. Future research should consider the impact on cost-effectiveness, duration, frequency, and theoretical mechanisms of the DHE to maximize the potential impact. It should also be conducted in the context of health services research to better reflect the real world.

(J Med Internet Res 2025;27:e53142) doi:10.2196/53142

KEYWORDS

digital health education; digital health interventions; chronic lung disease; eHealth; scoping review

XSL•FO
RenderX

Introduction

Chronic lung disease (CLD) is one of the most prevalent noncommunicable diseases globally, placing a large burden on patients and increasing health care expenditures, primarily including chronic obstructive pulmonary disease (COPD) and asthma. The total number of deaths due to CLD increased by 18% during the past 3 decades, with the highest mortality and disability-adjusted life years in regions with a low sociodemographic index, mainly in Africa, Middle East, Central Asia, Southeast Asia, and South America, also namely low- and middle-income countries (LMICs) [1-3]. Smoking, air pollution, and occupational exposure were the most significant risk factors for CLD [4]. Considering the acute exacerbations [5], complicated comorbidities, and increasing costs of CLD [6], raising disease awareness and promoting healthy lifestyles through health education and management is crucial to improving people's health conditions among patients with CLD and the general population.

Digital health education (DHE) offers knowledge or skills about health and health care in a timely and cost-effective manner by using all kinds of information and communication technologies (computer-assisted learning, mobile learning, and digital simulation-based learning), which is increasingly important in disease prevention and management [7]. With the accessibility of the internet and the popularity of advanced electronic or digital technology, diverse DHE mediums have emerged, such as websites, multimedia, and mobile apps, providing learners with health information and instant feedback [8-11]. In addition, many studies reported a potential positive impact of digital health interventions on various diseases and mental health status, including improving people's knowledge, skills, adherence, and other health-related outcomes [12-15]. However, despite the benefits of digital health intervention, the evidence is scarce regarding the characteristics and impact of DHE, especially in patients with CLD.

Currently, suboptimal disease control in patients with CLD is noticed due to their lower medication adherence and incorrect use of inhaler devices, which further contribute to increased exacerbation and hospitalizations [16]. This highlights the importance of DHE to deliver continuous and efficient educational content that helps them better understand their condition and treatment regardless of time and space. However, to the best of our knowledge, no reviews have focused on the characteristics and impact of DHE on CLD. It is essential to conduct a review to understand what aspects of DHE have been explored that could facilitate or hinder their adherence. Furthermore, previous studies on digital health interventions in CLD focused on multidimensional interventions combining disease education with monitoring, coaching, and self-management [17-19]. A review is needed to know if DHE could also lead to significant changes, making a rationale for not giving all patients a multidimensional and, thereby, costly intervention.

Therefore, for this study, we aim to review the existing literature on DHE for CLD, with a focus on exploring the DHE mediums, content, mechanisms, and reported outcomes in patients with CLD. Specifically, we intend to address the following research questions: (1) What are the reported characteristics of the DHE for patients with CLD, in terms of type of DHE mediums, content, duration, frequency, and mechanisms? (2) What outcomes are reported in terms of the impact of DHE on patients' knowledge of CLD, treatment, management, and health-related outcomes? (3) What is known about the feasibility, acceptability, and patient satisfaction of DHE for patients with CLD?

Methods

Study Selection

Search strategies were developed with a medical information specialist from Leiden University Medical Center. Searches were conducted in PubMed, Web of Science, Embase, PsycINFO, and The Cochrane Library combining key search terms "chronic lung disease," "digital technology," "health education," and related terms. To ensure a comprehensive scope by covering all relevant studies on DHE for CLD and capturing a full breadth of the topic, studies were searched from inception to the day we conducted the search (February 26, 2023) without starting time limits. The complete search strategy is available in Multimedia Appendix 1. Literature screening was performed with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [20] checklist (Multimedia Appendix 2). Inclusion and exclusion were performed according to the Participants, Intervention, Comparison, and Outcomes (PICO) strategy [21], shown in Textbox 1.

After removing duplicates, one reviewer (CS) used the open-source machine learning tool ASReview Lab19 [22] to screen all the titles and abstracts. Another reviewer (RL) used Endnote X9 to double-screen a random proportion of all articles independently. The choice of 15% was made by the reviewer team to ensure the efficiency and reliability of screening based on ASReview developers' recommendations [23] and previously published articles that screened the same way [24]. If the agreement rate between the 2 reviewers was less than 80%, an additional random 15% of all titles and abstracts were screened by RL. Title and abstract screening were stopped only when the agreement rate exceeded 80%. Then, the included articles were selected by screening the full text by CS and HD. CS then screened all references of included articles to identify if extra articles could be included. Any discrepancies during the screening process were solved by achieving consensus or discussing with the review team.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion

- Participants: people diagnosed with chronic lung disease by a clinical professional.
- Intervention: digital health education focused on chronic lung disease (providing knowledge or skills about health and healthcare by using information and communication technologies).
- Comparison: no restriction.
- Outcome: all outcomes.
- Study design: randomized or nonrandomized controlled trials, qualitative and quantitative studies, mixed methods studies, observational studies.

Exclusion

- Not digital health education intervention or blended intervention (eg, interventions that combine education and skills training) that include less than 50% digital health education.
- Studies with subjects <18 years of age.
- Commentaries, reviews, letters, dissertations, editorials, conference proceedings, books, and protocols.
- Not published in English.
- No full text is available.
- Not peer-reviewed articles.

Justification

- To ensure a comprehensive review, different study designs were considered: qualitative studies can be used for exploring patients' experiences, opinions, and attitudes. Quantitative studies can be used to explore the impact of digital health education on chronic lung disease.
- We excluded participants aged less than 18 because there are fewer digital health education interventions directly targeting children with chronic lung disease, but more at their parents, teachers, or pediatric-care professionals.
- Studies with no full text cannot provide complete information about their methods, results, and conclusions.
- A systematic search (query of experts, professional societies, service providers, etc) of non–peer-reviewed publications could not be carried out for reasons of research economics.

Data Extraction

Reviewer teams (CS, HD, and HW) extracted data with a predesigned standardized data extraction form (Multimedia Appendix 3). Study characteristics (author, publication year, country, profession, study design, sample size, and population levels) were extracted to understand the specific context of each study and assess the relevance, research development, and quality. DHE characteristics (type of DHE medium, content, intervention mechanisms, duration, and frequency) were extracted to understand what aspect of educational platform, content, and behavioral theories were used and how. Outcomes, main findings, and conclusions of included studies were extracted to analyze the impact of DHE on patients with CLD.

Data Synthesis

Study characteristics and DHE characteristics were summarized using means and percentages where appropriate. Outcome indicators were summarized in 6 categories. CLD awareness evaluated patients' ability to perceive or to be conscious of CLD or health states (eg, health literacy). Clinical outcomes included physiological and biochemical indicators measured by clinical testing and laboratory equipment (eg, blood pressure, forced expiratory volume [FEV]). Feasibility, acceptability, and satisfaction [25] focused on the design or the development of DHE, evaluating the state or degree of being easily or conveniently used, being subject to acceptance, and the expectations, needs, and pleasure of patients. Lifestyle outcomes included diets, sleep, medication adherence, and physical activity. Cost-effectiveness evaluated the cost expenditure of patients or health care systems on the whole process of disease prevention, treatment, and management. Psychosocial outcomes were observed effects on patients' emotions, well-being, quality of life (QoL), behavior or social interactions [26]. The impacts of DHE were categorized as positive, negative, or no significant difference based on the extracted results and conclusions. Descriptive analysis was used to summarize all the coded data.

Quality Assessment

The Critical Appraisal Skills Program (CASP) checklist was used and adapted for appraising the evidence of different study designs [27]. Following CASP tool recommendations, we classified the evidence quality as high (more than 2/3 of questions are answered by "yes"), moderate (more than 1/3 of questions are answered by "yes"), or low (lower than 1/3 of questions are answered by "yes").

Results

Study Characteristics

A total of 22 articles were included in this scoping review (Figure 1), consisting of 19 interventional studies (9 randomized controlled trials [RCTs], 7 quasi-experimental studies, and 3 prospective interventional studies) and 3 mixed methods studies.

XSL•FO
RenderX

These studies were published between 2000 and 2022 in health care or related professions, showing a distinct increasing trend with the highest number after 2020 (n=11). Details of included articles and the trend of publication year were shown in Multimedia Appendix 3. Among these studies, only 27% (6/22) were conducted in LMICs, with China, India, Indonesia, Lebanon, and Iran 1 or 2, respectively (Figure 2). United States accounted for 32% (7/22) of these studies among developed countries. The sample size of the studies varied from 15 to 9452 participants, of which the population level ranged from hospitals, and clinics to general practices, communities, and recruited volunteers. By following the CASP checklist, all studies were medium (n=4) to high quality (n=18). Further details can be found in Table 1 [28-49] and the quality assessment process was described in Multimedia Appendix 4.

Figure 1. PRISMA-ScR diagram of the review. CLD: chronic lung disease; DHE: digital health education.

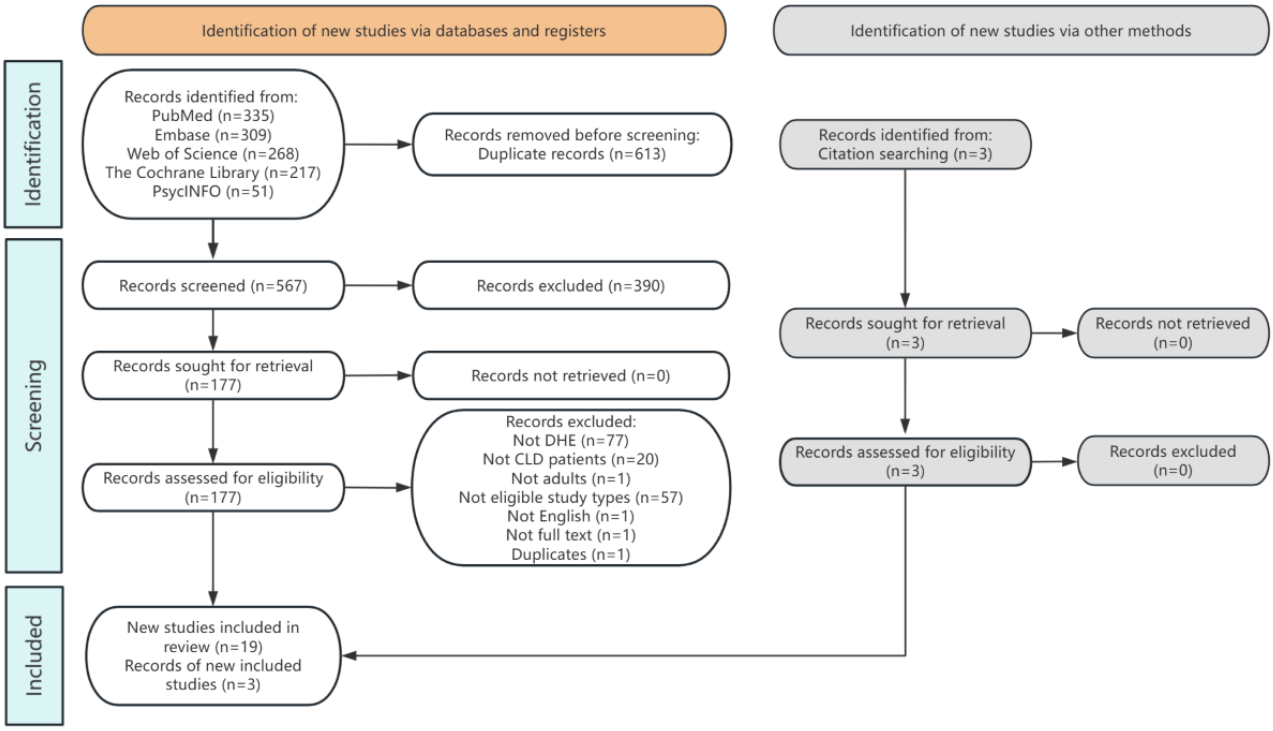

Figure 2. Frequency of publication countries.

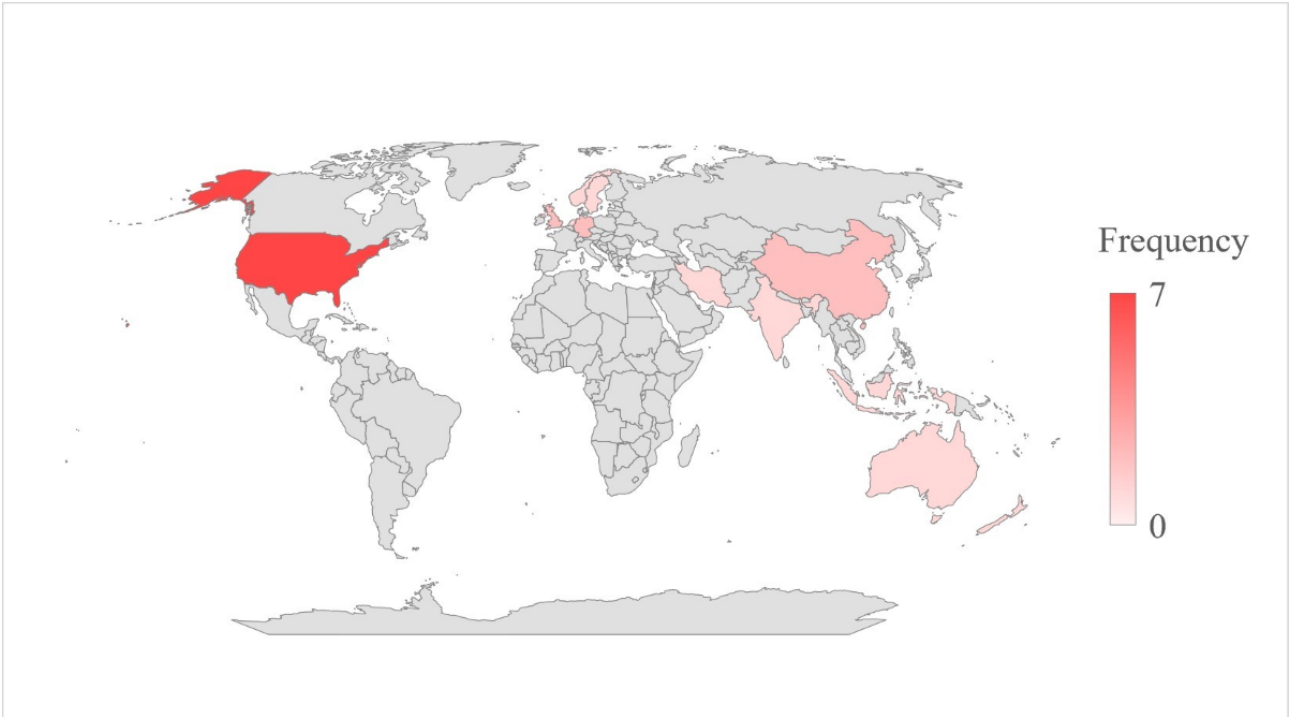

XSL•FO
RenderX

Table 1. Study characteristics of included articles.

Author	Year	Country	Profession	Study design	Sample size	Population levels	Quality assessment
Nyberg et al [28]	2019	Sweden	Community medicine	Quasi-experiment	83 patients with COPD[a]	Primary care centers	High
Mongiardo et al [29]	2021	United States	Pulmonary care	Prospective interventional study	72 patients with COPD	Outpatient pulmonary clinics	High
Atherton [30]	2000	United States	Analytical services	Quasi-experiment	437 patients with asthma	MyAsthma users	High
Liu et al [31]	2013	China	Nursing	RCT[b]	60 patients with COPD	Volunteer recruitment	Medium
Burns et al [32]	2013	Australia	Health initiatives	Quasi-experiment	51 patients with asthma	Web-based and offline recruitment	High
Wu et al [33]	2017	United States	Public health	Mixed methods study	48 patients with CLD[c]	General medicine service	High
Farzandipour et al [34]	2019	Iran	Health information	Mixed methods study	30 patients with asthma, 10 specialists	Clinic	High
Nohra et al [35]	2022	Lebanon	Public health	Quasi-experiment	15 patients with CLD	Empirical convenience sample	Medium
Newhouse et al [36]	2016	United Kingdom	Primary care	RCT	148 patients with asthma	Primary care practices	Medium
Korsbakke Emtekaer Haesum et al [37]	2016	Norway	Health science and technology	RCT	116 patients with COPD	Telephone recruitment	Medium
Wan et al [38]	2017	United States	Pulmonary care	RCT	109 patients with COPD	General pulmonary clinics	High
Ghozali et al [39]	2022	Indonesia	Pharmacy	RCT	140 patients with asthma	Community	High
Kellerer et al [40]	2022	Germany	General practice	Quasi-experiment	72 patients with asthma	General practices	High
Hsia et al [41]	2020	United States	Medicine	Prospective interventional study	40 patients with asthma	The Montefiore Asthma Center	High
Zhao et al [42]	2020	China	Respiratory medicine	RCT	713 patients with CLD	Hospital	High
Hsia et al [43]	2020	United States	Immunology	Prospective interventional study	60 patients with CLD	Outpatient multispecialty	High
Schnoor et al [44]	2022	The Netherlands	Public health	Quasi-experiment	9452 patients with CLD	Community pharmacies	High
Petrie et al [45]	2012	New Zealand	Psychology	RCT	216 patients with CLD	Flyers and emails recruitment	High
Chan et al [46]	2023	India	Immunology	Mixed methods study	57 patients with CLD	Clinic	High
Moore et al [47]	2009	United Kingdom	Pulmonary rehabilitation	RCT	20 patients with COPD	A program	High
Sobel et al [48]	2009	United States	Internal medicine	Quasi-experiment	130 participants with/without asthma	Organization, education center, clinic	High
Kohler et al [49]	2020	Germany	General practice	RCT	82 patients with asthma	Pulmonary rehabilitation clinic	High

[a]COPD: chronic obstructive pulmonary disease.

[b]RCT: randomized controlled trial.

[c]CLD: chronic lung disease.

XSL•FO
RenderX

DHE Mediums, Intervention Mechanisms, and Contents

Regarding DHE mediums among the 22 studies, 10 (45%) of them were websites or web-based platforms developed by researchers, along with 6 (27%) studies delivered by mobile applications or tablets, 5 (23%) by video or text messaging, and 1 (5%) by MP3 (Table 2 [28-49]). No more than half (9/22, 41%) of studies conducted DHE for over 3 months. Only 7 studies reported intervention frequency and 4 provided DHE based on the "teach-to-go" strategy, Theory of Planned Behavior, and self-regulation.

Table 2. Digital health education characteristics of included articles.

Author	Digital health education medium	Frequency	Duration	Theory
Nyberg et al [28]	An interactive web page	N/A[a]	6 months	N/A
Mongiardo et al [29]	A web-based platform	Daily or weekly	12 months	N/A
Atherton [30]	www.MyAsthma.com	N/A	6 months	N/A
Liu et al [31]	A web-based animated diagram and video	N/A	4 months	N/A
Burns et al [32]	A website	N/A	2 months	Theory of Planned Behavior
Wu et al [33]	Virtual learning platform	N/A	N/A	"teach-to-goal" strategy
Farzandipour et al [34]	A smartphone app	N/A	1 month	N/A
Nohra et al [35]	PowerPoint presentations and videos	Weekly	6 weeks	N/A
Newhouse et al [36]	A multimedia website	N/A	2 weeks	N/A
Korsbakke Emtekaer Haesum et al [37]	A tablet	N/A	4 weeks	N/A
Wan et al [38]	A website	Weekly	3 months	Self-regulation
Ghozali et al [39]	AsmaDroid app	N/A	4 weeks	N/A
Kellerer et al [40]	A web page	N/A	6 months	N/A
Hsia et al [41]	ASTHMAXcel mobile app	One time use	6 months	N/A
Zhao et al [42]	MP3 player	N/A	6 months	N/A
Hsia et al [43]	ASTHMAXcel mobile app	N/A	12 months	N/A
Schnoor et al [44]	A web-based platform SARA[b]	N/A	12 months	N/A
Petrie et al [45]	Text messaging	Irregularly	18 weeks	N/A
Chan et al [46]	Adapted ASTHMAXcel mobile application	N/A	After completing 8 chapters	N/A
Moore et al [47]	Video	4 times a week	6 weeks	N/A
Sobel et al [48]	Multimedia curriculum	Once	Instantly	"teach-to-goal" strategy
Kohler et al [49]	A website	N/A	3 weeks	N/A

[a]N/A: not applicable.

[b]SARA: Service Apothecary Respiratory Advice.

Different contents of DHE in CLD were summarized and divided into several categories (Multimedia Appendix 5). Self-management skills of CLD (21/22, 95%) were the most common educational content, including managing general CLD conditions (20/22, 91%), and exacerbations or asthma attacks (7/22, 32%). Treatment of CLD, disease characteristics, motivational messages, or stories were imparted in 15, 15, and 5 studies, respectively. Specifically, medicine or inhaler education (14/22, 64%) was the most frequently mentioned topic about CLD treatment, followed by breathing exercises or physical activity (5/22, 23%) and psychosocial therapy (3/22, 14%).

DHE Outcomes for Chronic Lung Disease

After classification and synthesizing, we found that 16/22 (73%) studies explored participants' CLD awareness. Clinical outcomes (6/22, 27%), DHE feasibility, acceptability, and satisfaction (8/22, 36%), lifestyle (3/22, 14%), cost-effectiveness (2/22, 9%), and psychosocial outcomes (8/22, 36%) were also reported in different studies.

CLD Awareness

Fourteen out of 16 studies reported positive changes in CLD awareness. Among them, 9 interventional studies noticed a statistically significant increase ($P<.05$) when evaluating CLD knowledge, literacy, and skills in the intervention group compared with the control group [28,32,35,39,40,42,45,48,49].

XSL•FO
RenderX

In a cohort study, Mongiardo et al [29] found a significant improvement in knowledge about exercise, vaccination, inhaled bronchodilators, and inhaled steroids (P<.01) at 12 months among patients with COPD who used a COPD web-based platform that delivers education as part of a physical activity intervention. Similar results were reported in the study by Hsia et al [41,43] by measuring Asthma Knowledge Questionnaire scores among patients with asthma who used ASTHMAXcel—a novel, guidelines-based smartphone app designed to improve patient education and outpatient management of asthma. Chan et al [46] also evaluated an adapted ASTHMAXcel and got the same result using Asthma Knowledge Questionnaire scores among patients with asthma in India. The mean score of patients' self-management knowledge of patients with asthma was significantly increased to 4.3 (SD 0.56; P<.001) after using a smartphone-based app, according to the mixed methods study by Farzandipour et al [34]. When evaluating knowledge and skills after a 6-week home-based educational and telemonitoring program, most of the participants answered correctly [35]. However, Korsbakke Emtekaer Haesum et al [37] reported no statistically significant difference in functional health literacy between the intervention and control group when they focused on the effect of the Telekit—a tele-homecare intervention combining a tablet and software that provided information and instructions on managing COPD. No significant differences in COPD knowledge were noted between the pedometer plus website group and the control group in the RCT by Wan et al [38].

Clinical Outcomes

In total, 3 out of 6 studies reported positive clinical outcomes of DHE. These clinical outcomes mainly included changes in the impact of COPD in daily life evaluated by the COPD Assessment Test [28], FEV, forced expiratory volume in 1 second (FEV1)/forced vital capacity (FVC), peak expiratory volume, 6-minute walking distance [31], clinical parameters (systolic and diastolic blood pressure, pulse, FEV1, FVC) [37], dyspnea [38], exacerbation rates, antimycotic treatment difference [44], change in walking capacity [47]. In an RCT, patients with COPD who received a web-based animated diagram and video-guided instruction showed significant improvements in FEV, FEV1/FVC, and peak expiratory volume compared with patients who received conventional education (P<.05) [31]. In addition, a significantly lower increase in exacerbation rates over time was found compared to the control participants in participants with asthma but not in individuals with COPD [44]. The walking capacity also showed statistically significant improvements (P<.05) within the intervention group measured by the Incremental Shuttle Walk Test after a home exercise video program among patients with moderate to severe COPD [47]. According to Andre et al [28], however, the mean scores of COPD Assessment Test were not highly reduced in patients with mainly moderate COPD, and the difference is statistically insignificant between the baseline and 3-, and 12-month time point. There were also RCTs providing information on managing COPD by a tablet and a website with no statistically significant difference in clinical outcomes, although the intervention duration lasted for several weeks and even 3 months [37,38].

DHE Feasibility, Acceptability, and Satisfaction

In total, 6 out of 8 studies indicated a positive effect on DHE feasibility, acceptability, and satisfaction. In the study by Nyberg et al [28], the feasibility analysis of a COPD web page on self-management skills combining texts, pictures, videos, and interactive components was primarily focused on the number of users and the total time they spent on the site during the initial 3 months, showing that 77% (33/43) of patients having access to the COPD Web were considered users and the average total time on the site was 45 minutes, which was quite low. Another feasibility study of a novel experience-based internet intervention to support self-management in chronic asthma was also relatively low, reporting that the median number of logins during 2 weeks was 2, and the median amount of time spent on the site was 17 minutes [36]. In addition, the e-Health Impact Questionnaire scores they used to measure the respondents' attitudes toward the website were quite similar between groups. Satisfaction, ease-of-use, and usefulness were assessed when analyzing the acceptability of an education website AsthmaWise for older asthma adults, of which 69% (35/51) had decided to visit it to learn more about how to manage their asthma, 94% (48/51) found it useful and almost all participants would recommend it to others [32]. Wu et al [33] developed a virtual teach-to-goal inhaler technique learning module and found that over 80% (39/48) reported that they were likely to use virtual teach-to-goal inhaler technique learning module to gain skills to improve their health. Furthermore, they identified several domains from focus groups for optimal patient use: access, functionality, and quality. In a quasi-experiment, all 15 participants considered the educational and telemonitoring program adequate and acceptable, declaring they had no difficulty explaining it to others and would suggest it to others [35]. The satisfaction rate regarding an audio-educational method (82%, 291/357 vs 66%, 235/356; P<.05) and effective feedback (83%, 68/82 vs 53%, 42/80; P<.001) were significantly higher in the experimental group than in the control group for patients from respiratory medicine wards [42]. Patients were highly satisfied in the ASTHMAXcel group, scoring on average 27.9 out of 30 points on the satisfaction survey, but no difference was found in either intervention [43]. An average of 50 out of 54 maximum points was noted based on the questionnaire for user interface satisfaction about participants' postinterventional median satisfaction for the adapted ASTHMAXcel [46].

Lifestyle

All (3/3) studies reported positive changes in patients' lifestyle. When evaluating the effect of a COPD web in promoting self-management of patients, the amount of daily physical activity and time spent in training exercises in the intervention group at 3 months was increased compared with the usual care group, but not statistically different at 12 months [28]. The daily step count in the pedometer plus website group also significantly improved from baseline beginning and maintained to week 13 based on the theory of self-regulation [38]. When comparing medication adherence in the year after the implementation of an eHealth intervention, Service Apothecary Respiratory Advice, both study groups showed an increase in mean medication adherence over time. Especially, it was significantly higher for

XSL•FO
RenderX

Service Apothecary Respiratory Advice participants (P=.01), according to Schnoor et al [44].

Cost-Effectiveness Outcomes

Only 2 studies evaluated the cost-effectiveness of DHE. A study conducted in Lebanon estimated the overall cost of the educational and telemonitoring program was US $200 per patient (educational sessions only) and US $244 per patient (including telemonitoring services), which consisted of the general cost of materials or patients, time nurses spent on patients, hospitalizations and emergency department visits [35]. However, it was difficult for them to evaluate this intervention's relative cost given the complicated economic situation in Lebanon. The other study reported that all health care utilization endpoints, including asthma-related emergency department visits, hospitalizations, and oral prednisone use, decreased at 4 months and 6 months based on survey results [41].

Psychosocial Outcomes

Improved psychosocial outcomes were reported in 7 out of 8 studies. Notably, QoL, usually evaluated by self-reported questionnaire, was the most frequent outcome in these studies. In total, 7 out of 8 studies reported improvement in the intervention group or at the follow-up point [30,31,35,36,38,41,47]. A total of 5 of 15 patients showed lower anxiety and depression scores, and 6 had higher QoL scores compared to baseline measures, according to Rita's study [35]. Minimal differences of change from baseline between groups were noticed regarding how well patients with asthma manage their condition and self-efficacy of being able to undertake certain activities [36]. Wan et al [38] also compared health-related QoL, exercise self-efficacy, depression, social support, motivation, and confidence to exercise, but no significant differences were noted between groups. All participants showed an increase in health state from baseline based on the physical and mental component of the 36-Item Short Form Survey (including physical and mental dimensions), with a more significant improvement in the intervention group that provided a multimedia website illustrating asthma-related information [36]. No other psychosocial outcomes were analyzed in these studies.

Discussion

Principal Findings

Current evidence presents comprehensive characteristics and the potentially positive impact of DHE on CLD. A total of 22 published studies with medium to high quality from 2000 to 2022 were analyzed in this review, including 9 (41%) RCTs, 7 (32%) quasi-experimental studies, 3 (14%) mixed methods studies, and 3 (14%) prospective interventional studies. Most studies (16/22, 73%) were conducted in developed countries. Websites and mobile applications (10/22, 45%) were the most widely used medium to conduct DHE among patients with CLD, followed by mobile apps or tablets, video or text messaging, and MP3. Participants were mainly patients with CLD recruited from hospitals and general practices. No more than half of these studies (9/22, 41%) implemented DHE interventions for more than 3 months, and few (3/22, 14%) were based on theoretical mechanisms when educating the target population. The DHE content varied from a general introduction of CLD characteristics (definition and etiology), and treatment to self-management skills of CLD, motivational messages, and stories for patients and their relatives. Positive impacts were reported among most of these studies, including increased CLD awareness (14/16, 88%), clinical (3/6, 50%), lifestyle (3/3, 100%), psychosocial outcomes (7/8, 88%) and impressive feasibility, satisfaction, and acceptability (6/8, 75%). Only 2 studies reported cost-effectiveness outcomes.

Comparison With Previous Work

The growing number of publications pertaining to DHE in CLD and the widespread use of websites and mobile apps shed light on an increasingly important role of digital technology in health and health care. As of January 2023, there were 5.16 billion internet users worldwide, over 60% of the global population currently connected to the world wide web with an estimated 192 minutes per day per user spent on the web [50]. In addition, integrating digital technology in health care is revolutionary by providing easy access to health services, improving patient engagement, supporting remote supervision of physical activity regimes at home, delivering timely detection of deteriorating health status, and serving as a cost-saving alternative health service [51]. Although not mentioned in included studies, the integration of artificial technology (AI) in health care, which is another promising DHE medium, holds a significant potential to deliver more efficient and personalized educational content through advanced algorithms and predictive analytics [52,53]. Therefore, users could access a diverse range of health information with better accessibility, efficiency, and interactivity. The increasing trend of exploring the impacts of DHE intervention will continue in future research.

More attention should be paid to implementing and scaling DHE to LMICs and the general population with limited access to digital technology. As reported, the internet penetration rate of China was relatively low at 76% compared with South Korea and Japan, which ranged over the 90% mark, although it is home to the largest digital community in the world, with about 1.07 billion netizens as of 2022 [54]. The same pattern is apparent in Southeast Asian countries with little internet accessibility at home and low self-perceived digital literacy among Southeast Asian youth, thereby reducing health care engagement and hindering health information dissemination for the local people [55]. Larger gains in health can be achieved in these areas due to the huge population underserved, which could also contribute to global health in the future. Furthermore, new opportunities and challenges are presented when scaling up DHE with the increasing aging population and worldwide health problems. Digital technologies play an increasingly important role not only in health education and promotion but also in addressing issues of health equality [56].

Future studies should further elucidate DHE characteristics and mechanisms to ensure long-standing effectiveness and verifiable impacts. First, the predominance of DHE contents focusing on self-management skills of CLD highlights the importance of empowering individuals to take an active role in their own health, which corresponds to the current research trends of

XSL•FO
RenderX

chronic diseases [57,58]. Second, successful health education depends on repeatedly using a few messages of proven benefit in many forums [59]. As such, the DHE content and the authenticity of contents should be warranted, of which the latter was rarely mentioned. As an essential health education component, DHE frequency, duration, and intervention theories are closely connected with prospective health outcomes. Several existing models, such as the Unified Theory of Acceptance and Use of Technology, Technology Acceptance Model, Social Cognitive Theory, and Theory of Planned Behavior, have been identified as facilitators of adopting digital health interventions [60]. Specifically, interventions based on the theory of planned behavior tended to have more substantial effects on behavior compared to interventions incorporating fewer techniques [61]. In this review, however, few studies illuminated the frequency (7/22, 32%) and intervention theory (4/22, 18%), indicating a potential gap in understanding how DHE interventions work and the underlying mechanisms that impact patients with CLD and related outcomes.

The effects reported on awareness or knowledge are commonly acknowledged [62-64] because digital technologies can support a wide range of self-directed learning activities, providing learners with diverse resources, information, and network resources that meet their learning needs [9]. Improvement in clinical outcomes suggests the potential of DHE to empower patients to make informed decisions about their health, adhere to treatment regimens, and adopt healthier lifestyles after enhancing their understanding of disease and health conditions. Similar positive effects in QoL and health status were also reported in other reviews, but they mainly focus on digital health interventions [65,66]. Insignificant effects in several studies might be related to the intervention duration or inclusion of participants [37,38,44]. The reported impacts on DHE feasibility, acceptability, and satisfaction shed light on users' active engagement and willingness with these interventions. These positive findings align with other studies [67,68], indicating that DHE is generally well-received by individuals with CLD and implies promising practicality and scalability in real-world settings.

A limited number of studies reported impacts on lifestyle (n=3) and cost-effectiveness outcomes (n=2) in this review, which needs more evidence to illustrate in the future. The lifestyle outcomes extracted from the included studies were limited to physical activity and medication adherence. According to previous studies, however, a sedentary and unhealthy lifestyle, including smoking, having an unhealthy diet, and insufficient exercise could fuel chronic disease progression by changing interstitial cell behavior [69]. In addition, we noticed that only one study applied behavioral theories but implemented DHE for only 2 months when exploring DHE impacts on lifestyle [32]. However, evidence has shown that a successful digital intervention can influence lifestyle management [70], and the theory of behavior change maintenance can guide the development and evaluation of interventions, thereby promoting sustained change in health behavior and lifestyles [71]. As such, a deeper exploration of specific lifestyles, such as the dietary pattern, types, frequencies, and intensities of exercise and other health-related outcomes should be mapped, with validated and effective theories. Currently, few studies in this review compare the cost or health care utilization after DHE intervention, which guided future cost-effectiveness studies with more accurate and complete data to fill in the gap.

Recommendation on Future Research

Based on the current review, we recommend that future research and development in DHE for people with CLD should focus on elucidating the characteristics and mechanisms of the DHE intervention, prioritizing less developed areas, investigating more lifestyle or behavioral outcomes, evaluating cost-effectiveness among different DHE mediums and ensuring long-term sustainability, to ultimately improve the effectiveness, scalability of DHE, and the health status of patients with CLD. With the integration of AI in health care, we hope to see more personalized and efficient education content delivered and look forward to studies reporting its potential and effectiveness. Given that several studies were of medium quality based on the CASP tool, we recommend future researchers consider incorporating blinding among participants, investigators, and analysts, as well as addressing confounding factors in both study design and analysis, which will enhance the robustness and reliability.

Strengths and Limitations

This is the first review to explore the characteristics and impact of DHE on CLD. Diverse studies in this review were synthesized and analyzed under the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guideline, all showing medium or high-quality evidence with a practical quality assessment tool. The detailed descriptions of DHE and its impact on different outcomes provide an important theoretical basis and new thoughts for further research in DHE. However, this review included studies focusing more on mixed digital health intervention. Despite the potential benefits, whether it is the independent or compounding influence of DHE remains to be discussed further due to the inability of subgroup analysis. The incomplete screening of titles and abstracts by the second reviewer may have introduced potential bias into the review process. The exclusion of non-English articles and studies without full-text access could result in publication bias to some extent. In addition, it is worth mentioning the caution of extrapolating the impact to other DHE mediums, despite recognizing the reliability of health information in these studies. Variations in researchers' profession might also contribute to heterogeneity in study implementation and reporting.

Conclusion

Despite the heterogeneity of the study situation, some aspects can be concluded. DHE can improve disease awareness and clinical outcomes among patients with chronic lung disease with good feasibility, acceptability, and satisfaction through different mediums and educational content. There is still relatively limited research among people in low- and middle-income countries. Future research should consider the impacts on cost-effectiveness outcomes, duration, frequency, and theoretical mechanisms of the DHE intervention to maximize the potential effects. It should also be conducted in

XSL•FO
RenderX

the context of health services research to better reflect the real-world settings.

Acknowledgments

We would like to acknowledge Claudia Pees (medical information specialist from Leiden University Medical Center) for helping with search strategies and the developers of ASReview for providing technical assistance. This study is funded by the China Scholarship Council. No generative artificial intelligence was used in any portion of the manuscript writing.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1-5.

Authors' Contributions

CS led the study design, and data analysis and is the main author of manuscript writing; HD, RL, and HW contributed to the literature screening and extracting; RK, SW, CH, and NC contributed to the study design and conception, editing, and proofreading. All authors read and approved of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 21 KB - jmir_v27i1e53142_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 84 KB - jmir_v27i1e53142_app2.docx]

Multimedia Appendix 3
Data extraction form and included articles.
[XLSX File (Microsoft Excel File), 24 KB - jmir_v27i1e53142_app3.xlsx]

Multimedia Appendix 4
Quality assessment of included studies.
[XLSX File (Microsoft Excel File), 13 KB - jmir_v27i1e53142_app4.xlsx]

Multimedia Appendix 5
Overview of digital health education (DHE) content in chronic lung disease (CLD).
[PNG File , 24 KB - jmir_v27i1e53142_app5.png]

References

1. Chronic respiratory diseases. World Health Organization. 2023. URL: https://www.who.int/health-topics/chronic-respiratory-diseases [accessed 2023-05-06]
2. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med 2020;8(6):585-596. [doi: 10.1016/S2213-2600(20)30105-3] [Medline: 32526187]
3. Li X, Cao X, Guo M, Xie M, Liu X. Trends and risk factors of mortality and disability adjusted life years for chronic respiratory diseases from 1990 to 2017: systematic analysis for the global burden of disease study 2017. BMJ 2020;368:m234 [FREE Full text] [doi: 10.1136/bmj.m234] [Medline: 32075787]
4. Safiri S, Carson-Chahhoud K, Noori M, Nejadghaderi SA, Sullman MJM, Ahmadian Heris J, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the global burden of disease study 2019. BMJ 2022;378:e069679 [FREE Full text] [doi: 10.1136/bmj-2021-069679] [Medline: 35896191]
5. Rabe KF, Watz H. Chronic obstructive pulmonary disease. Lancet 2017;389(10082):1931-1940. [doi: 10.1016/S0140-6736(17)31222-9] [Medline: 28513453]

XSL•FO
RenderX

6. Iheanacho I, Zhang S, King D, Rizzo M, Ismaila AS. Economic burden of chronic obstructive pulmonary disease (COPD): a systematic literature review. Int J Chron Obstruct Pulmon Dis 2020;15:439-460 [FREE Full text] [doi: 10.2147/COPD.S234942] [Medline: 32161455]
7. Bajpai S, Semwal M, Bajpai R, Car J, Ho AHY. Health professions' digital education: review of learning theories in randomized controlled trials by the digital health education collaboration. J Med Internet Res 2019;21(3):e12912 [FREE Full text] [doi: 10.2196/12912] [Medline: 30860483]
8. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022;4(3):e200-e206 [FREE Full text] [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
9. Curran V, Matthews L, Fleet L, Simmons K, Gustafson DL, Wetsch L. A review of digital, social, and mobile technologies in health professional education. J Contin Educ Health Prof 2017;37(3):195-206. [doi: 10.1097/CEH.0000000000000168] [Medline: 28834849]
10. Baumgart DC. Smartphones in clinical practice, medical education, and research. Arch Intern Med 2011;171(14):1294-1296. [doi: 10.1001/archinternmed.2011.320] [Medline: 21788549]
11. Han ER, Yeo S, Kim MJ, Lee YH, Park KH, Roh H. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC Med Educ 2019;19(1):460 [FREE Full text] [doi: 10.1186/s12909-019-1891-5] [Medline: 31829208]
12. Kaihara T, Scherrenberg M, Intan-Goey V, Falter M, Kindermans H, Frederix I, et al. Efficacy of digital health interventions on depression and anxiety in patients with cardiac disease: a systematic review and meta-analysis. Eur Heart J Digit Health 2022;3(3):445-454 [FREE Full text] [doi: 10.1093/ehjdh/ztac037] [Medline: 36712158]
13. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018;4 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
14. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III randomized controlled trial of eRAPID: eHealth intervention during chemotherapy. J Clin Oncol 2021;39(7):734-747 [FREE Full text] [doi: 10.1200/JCO.20.02015] [Medline: 33417506]
15. Wagner LI, Tooze JA, Hall DL, Levine BJ, Beaumont J, Duffecy J, et al. Targeted eHealth intervention to reduce breast cancer survivors' fear of recurrence: results from the foRtitude randomized trial. J Natl Cancer Inst 2021;113(11):1495-1505 [FREE Full text] [doi: 10.1093/jnci/djab100] [Medline: 34057469]
16. Adisa R, Ufuah UF, Ige OM. Impact of pharmacist-led intervention in medication adherence and inhaler usage on asthma and chronic obstructive pulmonary disease control: a quasi-experimental study. BMC Health Serv Res 2024;24(1):1199 [FREE Full text] [doi: 10.1186/s12913-024-11683-9] [Medline: 39379970]
17. Rasulnia M, Burton BS, Ginter RP, Wang TY, Pleasants RA, Green CL, et al. Assessing the impact of a remote digital coaching engagement program on patient-reported outcomes in asthma. J Asthma 2018;55(7):795-800. [doi: 10.1080/02770903.2017.1362430] [Medline: 28800273]
18. Arnaert A, Ahmad H, Mohamed S, Hudson E, Craciunas S, Girard A, et al. Experiences of patients with chronic obstructive pulmonary disease receiving integrated telehealth nursing services during COVID-19 lockdown. BMC Nurs 2022;21(1):205 [FREE Full text] [doi: 10.1186/s12912-022-00967-2] [Medline: 35915510]
19. Kuhn L, Reeves K, Taylor Y, Tapp H, McWilliams A, Gunter A, et al. Planning for action: the impact of an asthma action plan decision support tool integrated into an electronic health record (EHR) at a large health care system. J Am Board Fam Med 2015;28(3):382-393 [FREE Full text] [doi: 10.3122/jabfm.2015.03.140248] [Medline: 25957371]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. da Costa Santos CM, de Mattos Pimenta CA, Nobre MRC. The PICO strategy for the research question construction and evidence search. Rev Lat Am Enfermagem 2007;15(3):508-511 [FREE Full text] [doi: 10.1590/s0104-11692007000300023] [Medline: 17653438]
22. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell 2021;3(2):125-133. [doi: 10.1038/s42256-020-00287-7]
23. Penfornis KM, Van Vliet MHM, Meijer E, Gebhardt WA. Mapping the evidence on identity processes and identity-related interventions in the smoking and physical activity domains: a scoping review protocol. BMJ Open 2022;12(7):e058405 [FREE Full text] [doi: 10.1136/bmjopen-2021-058405] [Medline: 35831054]
24. Westendorp J, Geerse OP, van der Lee ML, Schoones JW, van Vliet MHM, Wit T, et al. Harmful communication behaviors in cancer care: a systematic review of patients and family caregivers perspectives. Psychooncology 2023;32(12):1827-1838. [doi: 10.1002/pon.6247] [Medline: 37957777]
25. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]

XSL•FO
RenderX

26. Bourne P, Camic PM, Crutch SJ. Psychosocial outcomes of dyadic arts interventions for people with a dementia and their informal caregivers: a systematic review. Health Soc Care Community 2021;29(6):1632-1649. [doi: 10.1111/hsc.13267] [Medline: 33368794]
27. Critical Appraisal Skills Programme. 2022. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2023-05-10]
28. Nyberg A, Tistad M, Wadell K. Can the COPD web be used to promote self-management in patients with COPD in swedish primary care: a controlled pragmatic pilot trial with 3 month- and 12 month follow-up. Scand J Prim Health Care 2019;37(1):69-82 [FREE Full text] [doi: 10.1080/02813432.2019.1569415] [Medline: 30700230]
29. Mongiardo MA, Robinson SA, Finer EB, Cruz Rivera PN, Goldstein RL, Moy ML. The Effect of a web-based physical activity intervention on COPD knowledge: a secondary cohort study. Respir Med 2021;190:106677 [FREE Full text] [doi: 10.1016/j.rmed.2021.106677] [Medline: 34775350]
30. Atherton M. Outcome measures of efficacy associated with a web-enabled asthma self-management programme: findings from a quasi-experiment. Disease Management and Health Outcomes 2000;8(4):233-242. [Medline: 30825713]
31. Liu F, Cai H, Tang Q, Zou Y, Wang H, Xu Z, et al. Effects of an animated diagram and video-based online breathing program for dyspnea in patients with stable COPD. Patient Prefer Adherence 2013;7:905-913 [FREE Full text] [doi: 10.2147/PPA.S43305] [Medline: 24049441]
32. Burns P, Jones SC, Iverson D, Caputi P. AsthmaWise - a field of dreams? The results of an online education program targeting older adults with asthma. J Asthma 2013;50(7):737-744. [doi: 10.3109/02770903.2013.799688] [Medline: 23745621]
33. Wu M, Woodrick NM, Arora VM, Farnan JM, Press VG. Developing a virtual teach-to-goal inhaler technique learning module: a mixed methods approach. J Allergy Clin Immunol Pract 2017;5(6):1728-1736 [FREE Full text] [doi: 10.1016/j.jaip.2017.04.032] [Medline: 28600133]
34. Farzandipour M, Nabovati E, Arani MH, Akbari H, Sharif R, Anvari S. Enhancing asthma patients' self-management through smartphone-based application: design, usability evaluation, and educational intervention. Appl Clin Inform 2019;10(5):870-878 [FREE Full text] [doi: 10.1055/s-0039-1700866] [Medline: 31724144]
35. Nohra RG, Chaaban T, Sacre H, Salameh P, Aoun Bacha Z, Le Bon Chami B, et al. Evaluating the feasibility and pretesting the impact of an educational and telemonitoring program for COPD patients in Lebanon. Int J Chron Obstruct Pulmon Dis 2022;17:949-965 [FREE Full text] [doi: 10.2147/COPD.S339592] [Medline: 35510164]
36. Newhouse N, Martin A, Jawad S, Yu L, Davoudianfar M, Locock L, et al. Randomised feasibility study of a novel experience-based internet intervention to support self-management in chronic asthma. BMJ Open 2016;6(12):e013401 [FREE Full text] [doi: 10.1136/bmjopen-2016-013401] [Medline: 28031210]
37. Korsbakke Emtekaer Haesum L, Ehlers L, Hejlesen OK. Interaction between functional health literacy and telehomecare: short-term effects from a randomized trial. Nurs Health Sci 2016;18(3):328-333. [doi: 10.1111/nhs.12272] [Medline: 26856258]
38. Wan ES, Kantorowski A, Homsy D, Teylan M, Kadri R, Richardson CR, et al. Promoting physical activity in COPD: insights from a randomized trial of a web-based intervention and pedometer use. Respir Med 2017;130:102-110 [FREE Full text] [doi: 10.1016/j.rmed.2017.07.057] [Medline: 29206627]
39. Ghozali MT, Satibi S, Ikawati Z, Lazuardi L. The efficient use of smartphone apps to improve the level of asthma knowledge. J Med Life 2022;15(5):625-630 [FREE Full text] [doi: 10.25122/jml-2021-0367] [Medline: 35815086]
40. Kellerer C, Schultz K, Linde K, Kohler B, Hapfelmeier A, Schneider A. Implementation of an internet-based asthma patient education program in primary care: results of a single-arm trial. J Asthma 2022;60(1):195-202. [doi: 10.1080/02770903.2022.2040027] [Medline: 35129399]
41. Hsia BC, Wu S, Mowrey WB, Jariwala SP. Evaluating the ASTHMAXcel mobile application regarding asthma knowledge and clinical outcomes. Respir Care 2020;65(8):1112-1119. [doi: 10.4187/respcare.07550] [Medline: 32487751]
42. Zhao Y, Liu M, Fan Z, Li J, Shi L, Zhang Y, et al. Application of audio education in respiratory medicine wards. Clin Nurs Res 2020;29(6):392-397. [doi: 10.1177/1054773819829622] [Medline: 30773048]
43. Hsia B, Mowrey W, Keskin T, Wu S, Aita R, Kwak L, et al. Developing and pilot testing ASTHMAXcel, a mobile app for adults with asthma. J Asthma 2020;58(6):834-847. [doi: 10.1080/02770903.2020.1728770] [Medline: 32046564]
44. Schnoor K, Versluis A, Bakema R, van Luenen S, Kooij MJ, van den Heuvel JM, et al. A pharmacy-based eHealth intervention promoting correct use of medication in patients with asthma and COPD: nonrandomized pre-post study. J Med Internet Res 2022;24(6):e32396 [FREE Full text] [doi: 10.2196/32396] [Medline: 35675120]
45. Petrie KJ, Perry K, Broadbent E, Weinman J. A text message programme designed to modify patients' illness and treatment beliefs improves self-reported adherence to asthma preventer medication. Br J Health Psychol 2012;17(1):74-84. [doi: 10.1111/j.2044-8287.2011.02033.x] [Medline: 22107110]
46. Chan A, Kodali S, Lee GY, Gadhave S, Feldman JM, Arora S, et al. Evaluating the effect and user satisfaction of an adapted and translated mobile health application ASTHMAXcel© among adults with asthma in Pune, India. J Asthma 2023;60(8):1513-1523. [doi: 10.1080/02770903.2022.2155188] [Medline: 36511602]
47. Moore J, Fiddler H, Seymour J, Grant A, Jolley C, Johnson L, et al. Effect of a home exercise video programme in patients with chronic obstructive pulmonary disease. J Rehabil Med 2009;41(3):195-200 [FREE Full text] [doi: 10.2340/16501977-0308] [Medline: 19229454]

XSL•FO
RenderX

48. Sobel RM, Paasche-Orlow MK, Waite KR, Rittner SS, Wilson EAH, Wolf MS. Asthma 1-2-3: a low literacy multimedia tool to educate African American adults about asthma. J Community Health 2009;34(4):321-327. [doi: 10.1007/s10900-009-9153-9] [Medline: 19353250]
49. Kohler B, Kellerer C, Schultz K, Wittmann M, Atmann O, Linde K, et al. An internet-based asthma self-management program increases knowledge about asthma. Dtsch Arztebl Int 2020;117(5):64-71 [FREE Full text] [doi: 10.3238/arztebl.2020.0064] [Medline: 32070472]
50. Number of internet and social media users worldwide as of January. 2023. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2023-05-13]
51. Ramachandran HJ, Oh JL, Cheong YK, Jiang Y, Teo JYC, Seah CWA, et al. Barriers and facilitators to the adoption of digital health interventions for COPD management: a scoping review. Heart Lung 2023;59:117-127. [doi: 10.1016/j.hrtlng.2023.02.004] [Medline: 36801546]
52. Koo TH, Zakaria AD, Ng JK, Leong XB. Systematic review of the application of Artificial intelligence in healthcare and nursing care. Malays J Med Sci 2024;31(5):135-142. [doi: 10.21315/mjms2024.31.5.9] [Medline: 39416729]
53. Issa WB, Shorbagi A, Al-Sharman A, Rababa M, Al-Majeed K, Radwan H, et al. Shaping the future: perspectives on the integration of artificial intelligence in health profession education: a multi-country survey. BMC Med Educ 2024;24(1):1166 [FREE Full text] [doi: 10.1186/s12909-024-06076-9] [Medline: 39425151]
54. Thomala LL. Internet usage in China - statistics & facts. Statista. URL: https://www.statista.com/topics/1179/internet-usage-in-china/#topicOverview [accessed 2023-05-13]
55. Kameke LV. Internet usage in Southeast Asia - statistics & facts. 2023. URL: https://www.statista.com/topics/9093/internet-usage-in-southeast-asia/#topicOverview [accessed 2023-05-13]
56. Thompson M. The geographies of digital health - digital therapeutic landscapes and mobilities. Health Place 2021;70:102610. [doi: 10.1016/j.healthplace.2021.102610] [Medline: 34174771]
57. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: an overview of reviews. Chron Respir Dis 2017;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
58. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019;40:127-146 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
59. Loevinsohn BP. Health education interventions in developing countries: a methodological review of published articles. Int J Epidemiol 1990;19(4):788-794. [doi: 10.1093/ije/19.4.788] [Medline: 2084004]
60. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
61. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
62. Martinengo L, Yeo NJY, Markandran KD, Olsson M, Kyaw BM, Car LT. Digital health professions education on chronic wound management: a systematic review. Int J Nurs Stud 2020;104:103512. [doi: 10.1016/j.ijnurstu.2019.103512] [Medline: 32086027]
63. Cueva M, Kuhnley R, Revels L, Schoenberg NE, Dignan M. Digital storytelling: a tool for health promotion and cancer awareness in rural alaskan communities. Int J Circumpolar Health 2015;74:28781 [FREE Full text] [doi: 10.3402/ijch.v74.28781] [Medline: 26343881]
64. Soloe C, Burrus O, Subramanian S. The effectiveness of mHealth and eHealth tools in improving provider knowledge, confidence, and behaviors related to cancer detection, treatment, and survivorship care: a systematic review. J Cancer Educ 2021;36(6):1134-1146 [FREE Full text] [doi: 10.1007/s13187-021-01961-z] [Medline: 33598832]
65. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017;5(5):CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
66. Yin AL, Hachuel D, Pollak JP, Scherl EJ, Estrin D. Digital health apps in the clinical care of inflammatory bowel disease: scoping review. J Med Internet Res 2019;21(8):e14630 [FREE Full text] [doi: 10.2196/14630] [Medline: 31429410]
67. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022;24(12):e43086 [FREE Full text] [doi: 10.2196/43086] [Medline: 36548034]
68. Skogberg M, Mackiewicz K, Mänd K, Tuuling L, Urdzina-Merca I, Salanterä S, et al. Promoting the health and wellbeing of children: a feasibility study of a digital tool among professionals. PLoS One 2022;17(3):e0265355 [FREE Full text] [doi: 10.1371/journal.pone.0265355] [Medline: 35316279]
69. Huston P. A sedentary and unhealthy lifestyle fuels chronic disease progression by changing interstitial cell behaviour: a network analysis. Front Physiol 2022;13:904107 [FREE Full text] [doi: 10.3389/fphys.2022.904107] [Medline: 35874511]
70. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]

XSL•FO
RenderX

71. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]

Abbreviations

AI: artificial intelligence
CASP: The Critical Appraisal Skills Program
CLD: chronic lung disease
COPD: chronic obstructive pulmonary disease
DHE: digital health education
FEV: forced expiratory volume
FEV1: forced expiratory volume in one second
FVC: forced vital capacity
LMIC: low- and middle-income country
PICO: Participants, Intervention, Comparison, and Outcomes
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
QoL: quality of life
RCT: randomized controlled trial

Edited by T de Azevedo Cardoso; submitted 31.10.23; peer-reviewed by M Redaelli; comments to author 02.10.24; revised version received 22.11.24; accepted 20.12.24; published 18.03.25.

Please cite as:
Sun C, Dai H, M.J.J. van der Kleij R, Li R, Wu H, Hallensleben C, Willems SH, Chavannes NH
Digital Health Education for Chronic Lung Disease: Scoping Review
J Med Internet Res 2025;27:e53142
URL: https://www.jmir.org/2025/1/e53142
doi:10.2196/53142
PMID:

©Chao Sun, Huohuo Dai, Rianne M.J.J. van der Kleij, Rong Li, Hengchang Wu, Cynthia Hallensleben, Sofie H Willems, Niels H Chavannes. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Behavioral Therapy–Based Digital Interventions for Treating Osteoarthritis: Systematic Review and Meta-Analysis

Beiyao Zhu[1*], MD; Dian Zhu[2*], PhD; Xiao'ao Xue[3], MD; Hongyi Yang[2], PhD; Shurong Zhang[3], MD

[1]Shanghai Jiao Tong University, The Ninth People's Hospital, Shanghai, China

[2]School of Design, Shanghai Jiao Tong University, Shanghai, China

[3]Department of Sports Medicine, Huashan Hospital, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Shurong Zhang, MD
Department of Sports Medicine
Huashan Hospital
Fudan University
No 12 Urumqi Middle Road
Jing'an District
Shanghai, 200040
China
Phone: 86 18901626266
Email: zhangshurong@huashan.org.cn

Abstract

Background: Osteoarthritis (OA) is characterized by pain, functional impairments, muscle weakness, and joint stiffness. Since OA heightens reliance on heath care resources and exacerbates socioeconomic burden, remote OA rehabilitation using digital technologies is rapidly evolving.

Objective: The aim of this study was to analyze the efficacy of behavioral therapy–based digital interventions for patients with OA.

Methods: This study is a systematic review of randomized controlled trials (RCTs) that assessed the effects of behavioral therapy–based digital intervention tools for OA. These RCTs were searched from inception to June 2023 in the Web of Science, Embase, Cochrane Library, Ovid, and PubMed databases.

Results: Ten eligible RCTs comprising 1895 patients with OA were included. Digital tools based on either cognitive behavioral therapy (CBT) or behavior change technique (BCT) were investigated. All studies demonstrated low-to-moderate effects on pain reduction in the short term (standardized mean difference [SMD] –0.20, 95% CI –0.35 to –0.05). Six studies reported improvement in physical function (SMD –0.20, 95% CI –0.41 to 0.00), and 5 confirmed increased pain self-efficacy (SMD 0.22, 95% CI 0.02-0.42). In subgroup analysis, compared with CBT, BCT-based digital interventions demonstrated their effects on pain reduction (SMD –0.25, 95% CI –0.49 to 0.00) and physical function (SMD –0.26, 95% CI –0.54 to –0.01) in the short term. In addition, physiotherapist involvement in treatment had a positive effect on pain control (SMD –0.14, 95% CI –0.27 to –0.02). Furthermore, web-based digital tools improved physical function in the short term (SMD –0.28, 95% CI –0.54 to –0.01).

Conclusions: Moderate- and low-quality evidence supported that behavioral therapy–based digital tools improved pain intensity, physical function, and self-efficacy in the short term. However, affective interactions between patients and professionals may affect the clinical outcomes.

Trial Registration: PROSPERO CRD42023430716; https://tinyurl.com/yc49vzyy

(J Med Internet Res 2025;27:e56227) doi:10.2196/56227

KEYWORDS

osteoarthritis; digital intervention; behavioral therapy; treatment; systematic review; meta-analysis; pain; impairment; quality of life; socioeconomic burden; psychotherapy-based digital intervention; patient; pain reduction

XSL•FO
RenderX

Introduction

Osteoarthritis (OA) is a prevalent joint disease affecting 29.7% of the population older than 40 years [1]. OA is characterized by pain, functional impairments, muscle weakness, joint stiffness, and reduced health-related quality of life (HRQoL) [2,3]. This causes heightened reliance on heath care resources and exacerbates socioeconomic burden [4,5]. Recent patient-centered guideline updates advocate for the implementation of rehabilitation interventions, including exercise therapy and educational programs as recommended nonpharmacological and nonsurgical treatment options [6,7]. These guidelines place a strong emphasis on the role of physical exercise, self-efficacy, and self-management strategies in the rehabilitation process for individuals with OA [8,9].

Psychological evidence–based auxiliary therapies have been integrated into the rehabilitation process to enhance rehabilitation outcomes [10]. These therapies provide a potent means of reinforcing patient education and behavior change. Behavioral theories, rooted in the social psychology of behavior change, are implemented as psychological evidence–based auxiliary therapies aimed at optimizing the positive components of interventions targeting patient health behaviors [11]. Many therapists incorporate physical therapy with behavioral therapy into their treatment protocols to alter observable maladaptive behavior patterns by eliciting new responses to given stimuli [12]. Behavioral science focuses on predicting, explaining, and altering behavior, encompassing approaches such as cognitive behavioral therapy (CBT) and behavior change techniques (BCTs) [13]. CBT, derived from the social cognitive theory, is a problem-focused approach that helps individuals identify and modify dysfunctional beliefs, thoughts, and behavior patterns contributing to their issues [14,15]. BCTs are considered active ingredients and include techniques such as feedback, self-monitoring, and behavior reinforcement. These techniques can be employed individually or in combination and in various forms [11,16]. Both CBT and BCT can enhance patients' ability to self-manage chronic diseases, particularly among those experiencing anxiety, depression, and chronic pain [17-19]. However, study outcomes have been heterogeneous. Some have confirmed that CBT is beneficial for pain improvement [20-22] and that BCT could enhance physical activity compliance in patients with lower extremity OA [23-25], while others have argued that CBT has no effect on patients who catastrophize about pain or has only a small positive effect on pain [12,26].

The increased prevalence of OA has resulted in increasing demand for therapists [27]. To this end, remote OA rehabilitation using digital technologies (eg, telephones, websites, mobile apps) is rapidly evolving to alleviate the socioeconomic burden [28-30]. Digitalized rehabilitation is comparable to physical therapist supervision in terms of exercise quality, physical training supervision, and sport-specific self-efficacy [31,32]. Accumulating evidence has demonstrated long-term improvement in pain and physical function with digital-based rehabilitation in patients with OA [33,34]. Moreover, some studies have reported that internet-based CBT and BCT have led to significant improvements in physical activity and exercise behavior compared with traditional treatments [24,35,36].

Although several studies have investigated the impact of digital behavior change interventions based on the social cognitive theory on physical activity in patients with OA [34,37], several shortcomings remain. First, different digital tools differ in terms of the digital format and underlying behavioral theories, making it challenging to provide evidence-based recommendations for their development process. Second, existing studies have not fully explored broader outcome measures. In order to fully assess the rehabilitation outcomes of patients with OA, a wider range of outcome metrics should be considered to fully evaluate the effectiveness of digital tools based on behavioral therapies. Finally, considering that relevant studies were conducted 3-5 years ago and new digital interventions for knee OA have emerged since then, there is an urgent need for an up-to-date review and evaluation of digital intervention tools incorporating behavioral therapies. Therefore, the aim of this review was to assess the effectiveness of digital interventions based on behavioral therapies for patients with OA in terms of pain, physical function, disability, and HRQoL. In addition, subgroup analyses of digital tool formats, therapist involvement, and underlying behavioral theories were conducted.

Methods

Study Design

This review was performed in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Multimedia Appendix 1) [38] and guidelines published in the Cochrane Handbook of Systematic Evaluation [39]. This study protocol is registered with PROSPERO (International Register of Prospective Systematic Reviews: CRD42023430716).

Eligibility Criteria

Studies fulfilling the following criteria were included in this review and meta-analysis: randomized controlled trials (RCTs); participants with knee, hip, or ankle OA; comparing CBT- or BCT-based digital interventions with other treatments; and those addressing at least one component of pain, physical function, HRQoL, pain self-efficacy, or physical activity. Only peer-reviewed studies involving participants older than 18 years were eligible for inclusion. Studies published in any language other than English and those without sufficient data were excluded. In this review, the term "digital intervention" refers to any solution or technology that delivers health information from health care providers to patients over a distance.

Literature Search

The literature search was performed using Web of Science, Embase, Cochrane Library, Ovid, and PubMed databases, from date of inception to June 27, 2023 (Table S1 of Multimedia Appendix 2). In addition, the reference lists of relevant reviews and selected papers were manually examined for potentially relevant/eligible trials.

Study Selection

All duplicate references were removed. Subsequently, titles and abstracts were manually screened for potentially eligible studies by 2 reviewers (DZ and HY), and relevant RCTs were identified.

XSL•FO
RenderX

The full text was then retrieved by both reviewers to assess eligibility for inclusion. Disagreements between the researchers were resolved through discussion or consultation with a third reviewer (BZ). The data were cross-validated by a third researcher (BZ) using EndNote 20 (Clarivate Analytics).

Data Extraction

Two independent reviewers (DZ and HY) collected the data. For continuous outcomes, the following data were extracted: mean (SD) and sample size at baseline and follow-up. For dichotomous outcomes, the number of cases and the total sample size were extracted. The dataset comprised study information, participant characteristics, type of behavioral therapy, type of digital tool, study duration, and outcome measures, including pain and function scores. If a study used multiple pain scales, the scale with the highest sensitivity to changes was used [40]. The function subscales of the Western Ontario and McMaster Universities Arthritis Index and the Knee Injury and Osteoarthritis Outcome Score/Hip dysfunction and Osteoarthritis Outcome Score Physical Function Shortform were used to assess functional improvement. Harmonized physical function was used, with higher scores indicating more severe physical dysfunction. The authors of studies with missing data were contacted. When the authors were unavailable, data were estimated using recommendations from the Cochrane Handbook (eg, estimation of SD from SEs) [39]. In trials in which SD was not reported, missing data were imputed from 95% CIs, SEs, *P* values, baseline changes, graphical representations, median (IQR), or SDs from baseline [41]. Trials in which imputations were not possible were excluded from the quantitative analysis.

The short-term effect was considered as follow-up period ≤6 months after randomization, and the long-term effect was considered as follow-up period >6 months after randomization. When >1 timepoint was available during the same follow-up period, the point closer to the end of the intervention was considered.

Risk of Bias Assessment

Risk of bias was assessed using the risk-of-bias tool in the Cochrane Handbook of Systematic Evaluation [42]. Seven domains, including random sequence generation, allocation concealment, blinding of participants and personnel, incomplete outcome data, selective reporting, and other biases, were used to evaluate quality of evidence [43]. Each domain was assigned a judgement of low, high, or unclear risk of bias. Furthermore, the quality level of this meta-analysis was evaluated according to GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach [44,45]. The quality of evidence was classified as high, moderate, low, or very low.

Data Analysis

A random-effects meta-analysis was performed using Review Manager version 5.4 (Cochrane Collaboration). Standardized mean differences (SMDs) were calculated to standardize the results to a uniform scale when studies assessed the same outcomes using different instruments. The SMDs for pain, physical function, HRQoL, pain self-efficacy, and physical activity were calculated by comparing interventions using behavioral therapy–based digital tools with other conventional methods. Therefore, a negative SMD value for pain and physical function and a positive SMD value for HRQoL, pain self-efficacy, and physical activity favor behavioral therapy–based digital tools. The magnitude of SMD was interpreted in accordance with the guidelines reported by Cohen [46], as follows: SMD <0.2 (small), 0.2-0.8 (medium), and >0.8 (large). For clinical interpretation, the mean differences were calculated.

Sensitivity and subgroup analyses were performed to assess the potential impact of the sources of heterogeneity. To investigate the potential impact of methodological quality on the estimates, we performed a sensitivity analysis by removing 1 study at a time, and trials with poor methodological quality were removed. Subgroups were defined in terms of the type of behavioral therapy (ie, CBT or BCT), type of digital tool (apps, wearable devices, and phones), and therapist involvement. The heterogeneity of the pooled studies was examined using the chi-square test and the I^2 statistic, with I^2>50% indicating substantial heterogeneity [47]. Publication bias was assessed by visual inspection of funnel plots and Egger test for meta-analyses ≥10 trials. All analyses were performed using Review Manager version 5.4.

Results

Study Selection and Characteristics

The initial search retrieved 1628 papers after removing duplicates, of which 1557 that did not fulfill the inclusion criteria were excluded. Two additional records were found: one through cross-referencing of bibliographies and the other by contacting the corresponding authors. Subsequently, 73 eligible full-text papers were reviewed, of which 10 RCTs were ultimately included in the quantitative analysis (Figure 1, Table 1) [24,35,36,48-53].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the studies selected in this review. OA: osteoarthritis; RCT: randomized controlled trial.

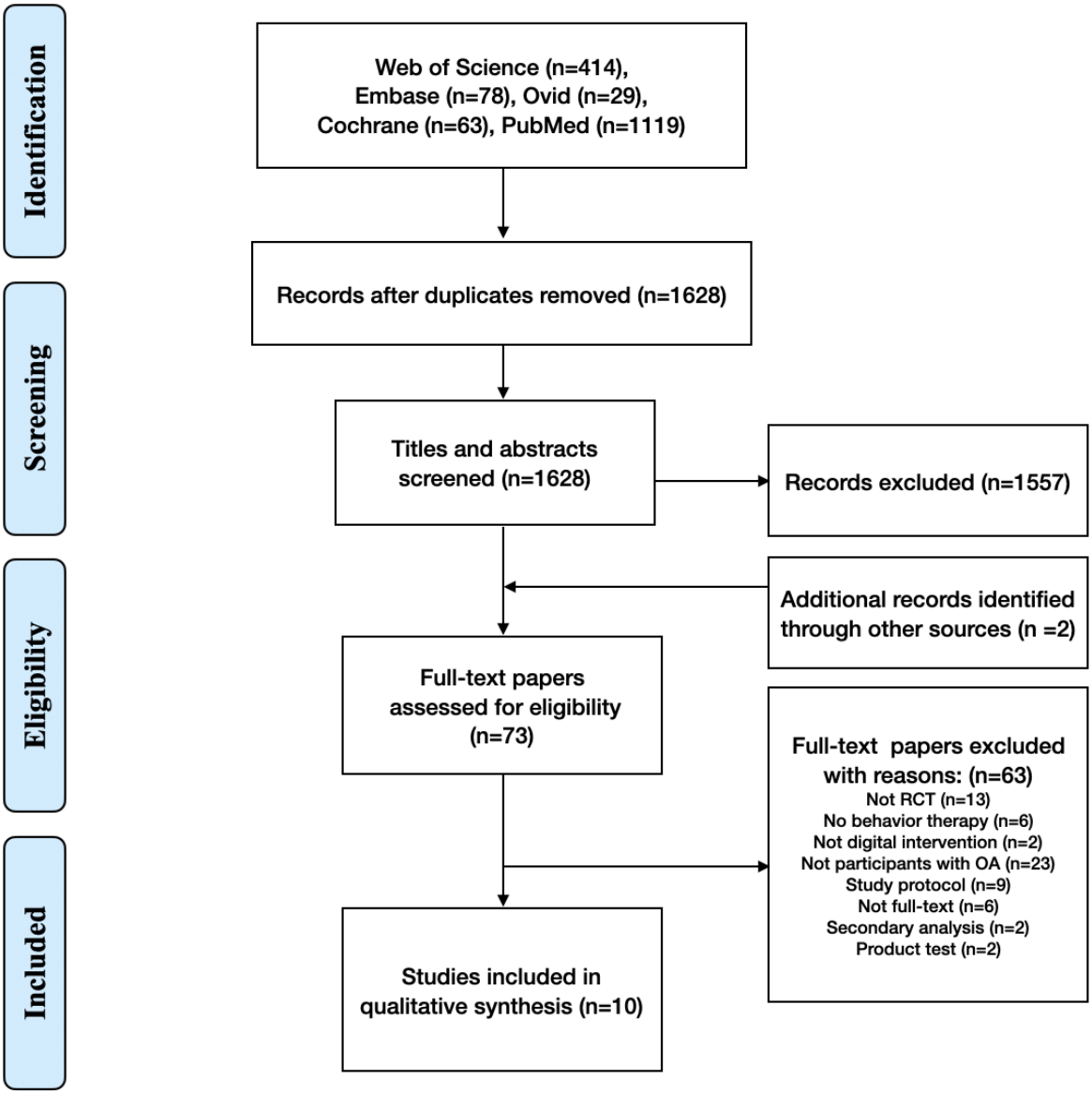

XSL•FO
RenderX

Table 1. Characteristics of the included studies.

Study author(s)	Nationality	Participants		Outcome (primary or secondary)	Follow-up (months)	Postintervention attrition (%)	
		Intervention group	Control group			Intervention group	Control group
Bennell et al [36]	Australia	SMS text message (n=56)	Non-SMS text message (n=54)	Exercise Adherence Rating Scale, days of the week to exercise (0-3), pain (Knee Injury and Osteoarthritis Outcome Score), overall mean knee pain (Numeric Rating Scale), symptoms (Knee Injury and Osteoarthritis Outcome Score), function (Knee Injury and Osteoarthritis Outcome Score), knee-related quality of life (Knee Injury and Osteoarthritis Outcome Score), health-related quality of life, self-efficacy: pain (Arthritis Self-Efficacy Scale), self-efficacy: function (Arthritis Self-Efficacy Scale), self-efficacy: other (Arthritis Self-Efficacy Scale), Behavioral Factors in Osteoarthritis Management Scale, Physical Activity Scale for the Elderly, Pain Catastrophizing Scale	6	14.29	5.55
Rini et al [48]	United States	Digital application (n=58)	Nonintervention (n=55)	Pain, self-efficacy in pain management, pain-related anxiety, pain-related functional interference, positive and negative emotions	3	1.72	5.45
Hinman et al [49]	Australia	Campaign advice and support (n=87)	Available services (n=88)	Overall mean knee pain (Numeric Rating Scale), Pain with Activities of Daily Living (Western Ontario and McMaster Universities Arthritis Index), mean pain while walking (Numeric Rating Scale), Pain Self-Efficacy (Arthritis Self-Efficacy Scale), Functional Self-Efficacy (Arthritis Self-Efficacy Scale), Fear of Movement, Physical Activity Scale for the Elderly, barriers to physical activity, benefits of physical activity, health-related quality of life	6	5.74	13.63
McCurry et al [50]	United States	Cognitive behavioral therapy-I (n=163)	Education (n=164)	Insomnia Severity Index, Flinders Fatigue Scale, Arthritis Pain Intensity and Interference with Activity, Brief Pain Inventory-short form, Patient Health Questionnaire-8 items, Measure of Depression	12	14.72	6.09
Pelle et al [26]	Netherlands	Dr Bart (n=214)	Conventional physiotherapy group (n=213)	Incidence rate ratio, pain, symptoms and dysfunction (Knee Injury and Osteoarthritis Outcome Score or Hip dysfunction and Osteoarthritis Outcome Score), quality of life (EQ-5D-3L), total hours of physical activity, Patient Activation Measure-13 items, knowledge skills confidence, Illness Perception Questionnaire cognitive and emotional perceptions	6	39.25	22.07
Mecklenburg et al [51]	United States	Hinge Health (n=101)	Education (n=61)	Pain, Physical Function Scale (Knee Injury and Osteoarthritis Outcome Score), Visual Analog Scale Pain, Visual Analog Scale Hardness Score, surgical intent	3	33.33	40.98
Nelligan et al [24]	Australia	Automatic Behavior Change SMS Support (n=103)	Web-based (n=103)	Physical Function (Western Ontario and McMaster Universities Osteoarthritis Index), Overall Mean Knee Pain (Numeric Rating Scale), Pain, Knee Quality of Life, Sport and Recreation (Knee Injury and Osteoarthritis Outcome Score), Assessment of quality of life-6D Physical Activity Scale, Arthritis Self-Efficacy Scale, Exercise Self-Efficacy Scale	6	11.65	10.68
Bennell et al [35]	Australia	Telephone counseling (n=84)	Nontelephone counseling (n=84)	Pain, physical function (Western Ontario and McMaster Universities Osteoarthritis Index), Numeric Rating Scale walking pain (range 0-10), Assessment of quality of life II, Physical Activity Scale for the Elderly, Athletic Activity Scale total activity time, step time (hours/day)	18	14.29	16.67
Kloek et al [52]	Netherlands	Web-based (n=109)	Conventional physiotherapy group (n=99)	Physical functioning, functions of daily living, functional limitations (Hip dysfunction and Osteoarthritis Outcome Score or Knee Injury and Osteoarthritis Outcome Score), self-perceived effects, arthritis self-efficacy scale, pain, and fatigue	12	18.30	12.10
Li et al [53]	Canada	Fitbit (immediate group) (n=26)	Education (delay group) (n=25)	Mean daily moderate to vigorous physical activity, Knee Injury and Osteoarthritis Outcome Score, Partners in Health Scale, Theory of Planned Behavior Questionnaire, Patient Health Questionnaire-9 items, Self-Reporting Habits Index	9.75	7.70	4

XSL•FO
RenderX

Study Characteristics

In this qualitative analysis, 9 RCTs and 1 cluster RCT, comprising 1895 patients, were included (Tables S2-S4 [24,26,35,36,48-53] of Multimedia Appendix 2). The trials were conducted in Europe (2/10, 20%) [26,52], Oceania (4/10, 40%) [24,35,36,49], and North America (4/10, 40%) [48,50,51,53]. The sample sizes of the included trials ranged from 51 to 427; the mean age of the patients was 62.5 (SD 8.1) years, and females were predominant in the pooled population (1231/1895, 64.9%), which is consistent with the global prevalence of OA. Six studies recruited participants with knee OA only [24,35,36,49,51,53], whereas 4 recruited participants with both knee and hip OA [26,48,50,52]. CBT-based (4 studies) [35,48,50,51] and BCT-based (6 studies) [24,26,36,49,52,53] therapies were the most used behavioral therapies blended into the digital intervention for patients with OA. Three types of digital tools were used: apps or websites in 9 studies [24,26,35,36,48,49,51-53], wearable devices in 3 studies [51-53], and SMS text messages in 1 study [50]. Four studies included face-to-face communication with physical therapists or fitness instructors [36,51-53]. All trials reported short-term effects (up to 6 months after randomization) [24,26,35,36,48-53] and 6 reported long-term effects (>6 months after randomization) [24,26,35,50,52,53]. Pain intensity, physical function, HRQoL, pain self-efficacy, and physical activity were evaluated in 10, 6, 6, 5, and 5 trials, respectively.

Risk of Bias in the Included Trials

In general, the most frequent risks of bias for RCTs were incomplete outcomes (6/10, 60%) and blinding of participants/personnel (6/10, 60%). Other biases and selective reporting accounted for the second most frequent risks of bias (2/10, 20%). The overall confidence in the cumulative evidence varied from very low to moderate, with low confidence being the most commonly identified (Figure 2 [24,26,35,36,48-53], Table 2).

XSL•FO
RenderX

Figure 2. Risk of bias. Red indicates high risk, green indicates low risk, and yellow indicates unclear risk.

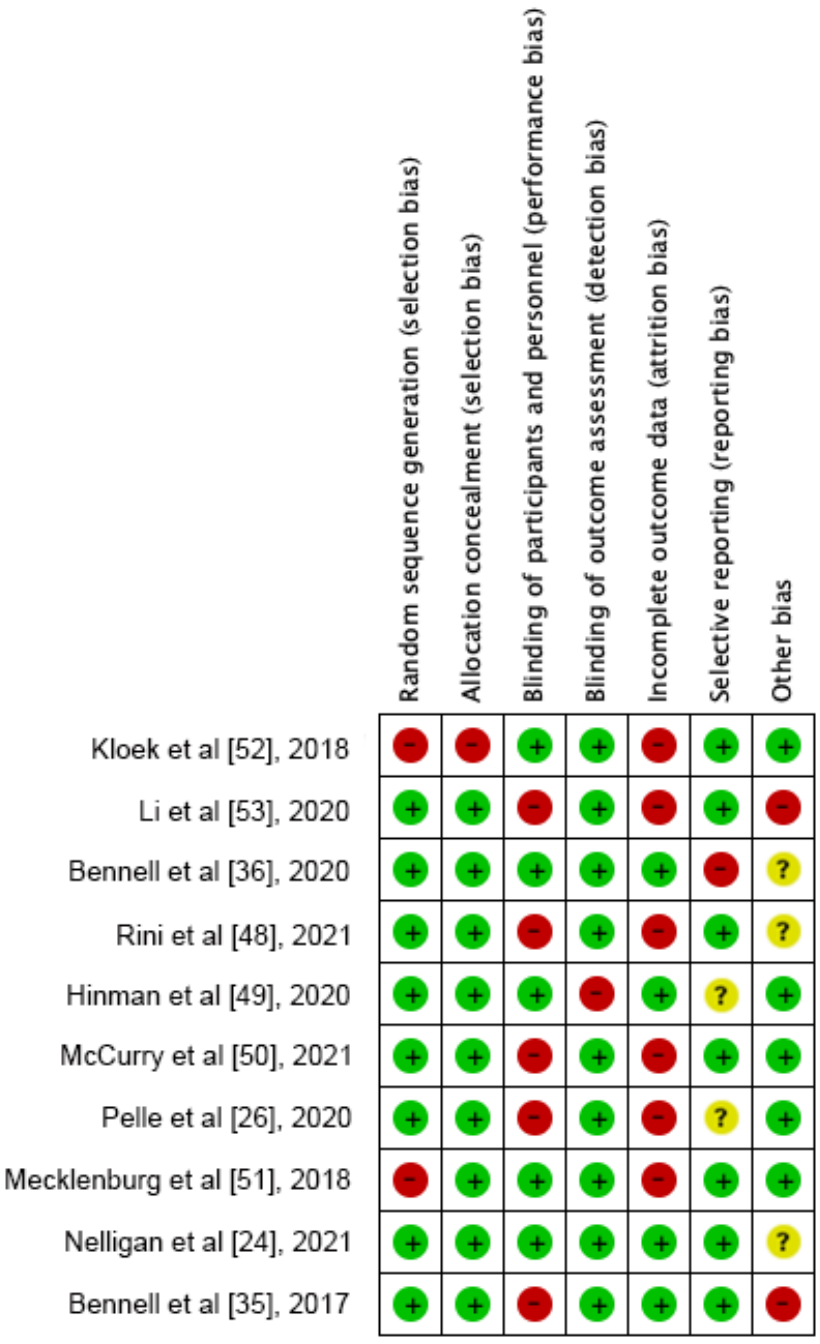

XSL•FO
RenderX

Table 2. Summary of the findings for efficacy outcomes.[a]

Outcome, time	Studies (patients), n	SMD[b] (95% CI)	I^2 (%)	Quality of evidence assessment					
				Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Rating[c]
Pain (short-term)	10 (1895)	–0.20 (–0.35 to 0.05)	65	_[d]	–	+[e]	+	+	Low
Pain (long-term)	6 (1271)	–0.01 (–0.12 to 0.10)	0	–	+	+	+	+	Moderate
Physical function (short-term)	6 (974)	–0.20 (–0.41 to 0.00)	62	–	–	+	+	+	Low
Physical function (long-term)	3 (519)	–0.13 (–0.32 to 0.06)	0	–	+	+	+	+	Moderate
Health-related quality of life (short-term)	6 (1137)	0.28 (–0.05 to 0.61)	86	–	–[f]	+	+	+	Very low
Pain self-efficacy (long-term)	5 (812)	0.22 (0.02 to 0.42)	10	–	+	+	–	+	Low
Physical activity (short-term)	5 (1086)	0.05 (–0.07 to 0.17)	0	–	+	+	+	+	Moderate
Subgroup outcome									
Behavior therapy pain outcomes (short-term)									
BCT[g]	6 (1137)	–0.25 (–0.49 to 0.00)	74	–	–	+	+	+	Low
CBT[h]	4 (758)	–0.10 (–0.24 to 0.05)	0	–	+	+	+	–	Low
Behavior therapy pain outcomes (long-term)									
BCT	4 (781)	–0.01 (–0.15 to 013)	0	–	+	+	+	+	Moderate
CBT	2 (490)	0.00 (–0.18 to 0.17)	0	–	+	+	+	+	Moderate
Physical function outcomes (short-term)									
BCT	5 (659)	–0.28 (–0.54 to –0.01)	67	–	–	+	+	+	Low
CBT	2 (363)	–0.05 (–0.26 to 0.16)	0	–	+	+	+	–	Low
Physical function outcomes (long-term)									
BCT	2 (311)	–0.21 (–0.44 to 0.02)	6	–	+	+	+	–	Low
CBT	1 (208)	0.00 (–0.28.0.27)	N/A[i]	–	+	+	+	–	Low
Pain outcomes with therapist involvement (short-term)									
With	6 (1039)	–0.14 (–0.27 to –0.02)	0	–	+	+	+	+	Moderate
Without	4 (856)	–0.23 (–0.60 to 0.14)	85	–	–[f]	+	+	+	Very low
Pain outcomes with therapist involvement (long-term)									
With	5 (844)	–0.04 (–0.17 to 0.10)	0	–	+	+	+	+	Moderate
Without	1 (427)	0.05 (–0.14 to 0.24)	N/A	–	+	+	+	–	Low
Physical function outcomes with therapists (short-term)									

XSL•FO
RenderX

Outcome, time	Studies (patients), n	SMD[b] (95% CI)	I^2 (%)	Quality of evidence assessment					
				Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Rating[c]
With	4 (706)	–0.17 (–0.40 to 0.07)	60	–	–	+	+	+	Low
Without	2 (316)	–0.27 (–0.75 to 0.20)	76	–	–	+	+	+	Low
Pain outcomes with digital tools (short-term)									
Telephone	1 (282)	–0.09 (–0.33 to 0.14)	N/A	–	+	+	+	–	Low
Web-based	6 (1199)	–0.23 (–0.47 to 0.01)	75	–	–[–2]	+	+	+	Very low
Wearable	3 (414)	–0.12 (–0.31 to 0.08)	0	–	+	+	+	–	Low
Pain outcomes with digital tools (long-term)									
Telephone	1 (282)	–0.03 (–0.27 to 0.20)	N/A	–	+	+	+	–	Low
Web-based	3 (738)	–0.02 (–0.16 to 0.12)	0	–	+	+	+	+	Moderate
Wearable	2 (251)	0.05 (–0.20 to 0.30)	0	–	+	+	+	–	Low
Physical function outcomes with digital tools (short-term)									
Web-based	4 (659)	–0.28 (–0.54 to –0.01)	67	–	–	+	+	+	Low
Wearable	2 (363)	–0.05 (–0.26 to 0.16)	0	–	+	+	+	–	Low
Physical function outcomes with digital tools (long-term)									
Web-based	2 (311)	–0.21 (–0.44 to 0.02)	6	–	+	+	+	–	Low
Wearable	1 (208)	0.00 (–0.28 to 0.27)	N/A	–	+	+	+	–	Low

[a]A funnel plot and Egger test was used to judge publication bias (Figure S1 of Multimedia Appendix 2).

[b]SMD: standardized mean difference.

[c]GRADE (Grading of Recommendations Assessment, Development, and Evaluation) Working Group grades of evidence. High: further research is very unlikely to change our confidence in the estimate of effect; moderate: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; low: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; very low: we are very uncertain about the estimate.

[d]– indicates low-quality evidence.

[e]+ indicates high-quality evidence.

[f]–[–2] indicates a moderate to high degree of inconsistency.

[g]BCT: behavior change technique.

[h]CBT: cognitive behavioral therapy.

[i]N/A: not applicable. Since only 1 study was included, there was no heterogeneity in the analysis.

Main Outcomes

For pain reduction, 10 trials with 1895 patients reported a moderate effect of behavioral therapy–based digital tools in reducing pain in the short term (4-24 weeks) (SMD –0.20, 95% CI –0.35 to –0.05; P=.008; I^2=58%), with low-quality evidence [24,26,35,36,48-53]. However, in the long-term follow-up (≥24 weeks), this effect was ambiguous when 6 trials with 1271 patients were included (SMD –0.01, 95% CI –0.12 to 0.10; P=.88; I^2=0%), with moderate-quality evidence [26,35,49,50,52,53]. For physical function, low certainty evidence supported significant improvement in the short term (SMD –0.20, 95% CI –0.41 to 0.00; P=.05; I^2=62%) from 6 trials with 974 patients [26,35,36,49,50,52]. However, its effect showed no significance in 3 trials with 519 patients in the long term (SMD –0.13, 95% CI –0.32 to 0.06; P=.18; I^2=17%), with moderate-quality evidence [35,49,52]. Sensitivity analysis revealed that the results of physical function were unstable;

XSL•FO
RenderX

therefore, these results should be interpreted with caution. For pain self-efficacy, low-quality evidence from 5 trials with 812 patients reported a positive effect (SMD 0.22, 95% CI 0.02-0.42; P=.03; I^2=10%) [24,36,48,49,52]. However, for HRQoL, very low-quality evidence from 6 trials with 1137 patients demonstrated an uncertain effect (SMD 0.28, 95% CI –0.05 to 0.61; P=.10; I^2=86%) [24,26,35,36,49,53]. Further, moderate quality evidence demonstrated an uncertain effect (SMD 0.05, 95% CI –0.07 to 0.17; P=.45; I^2=0%) for physical activity [24,26,35,36,49]. All results are presented in Figure 3 [24,26,35,36,48-53].

Figure 3. Main results of this study. HQoL: health-related quality of life.

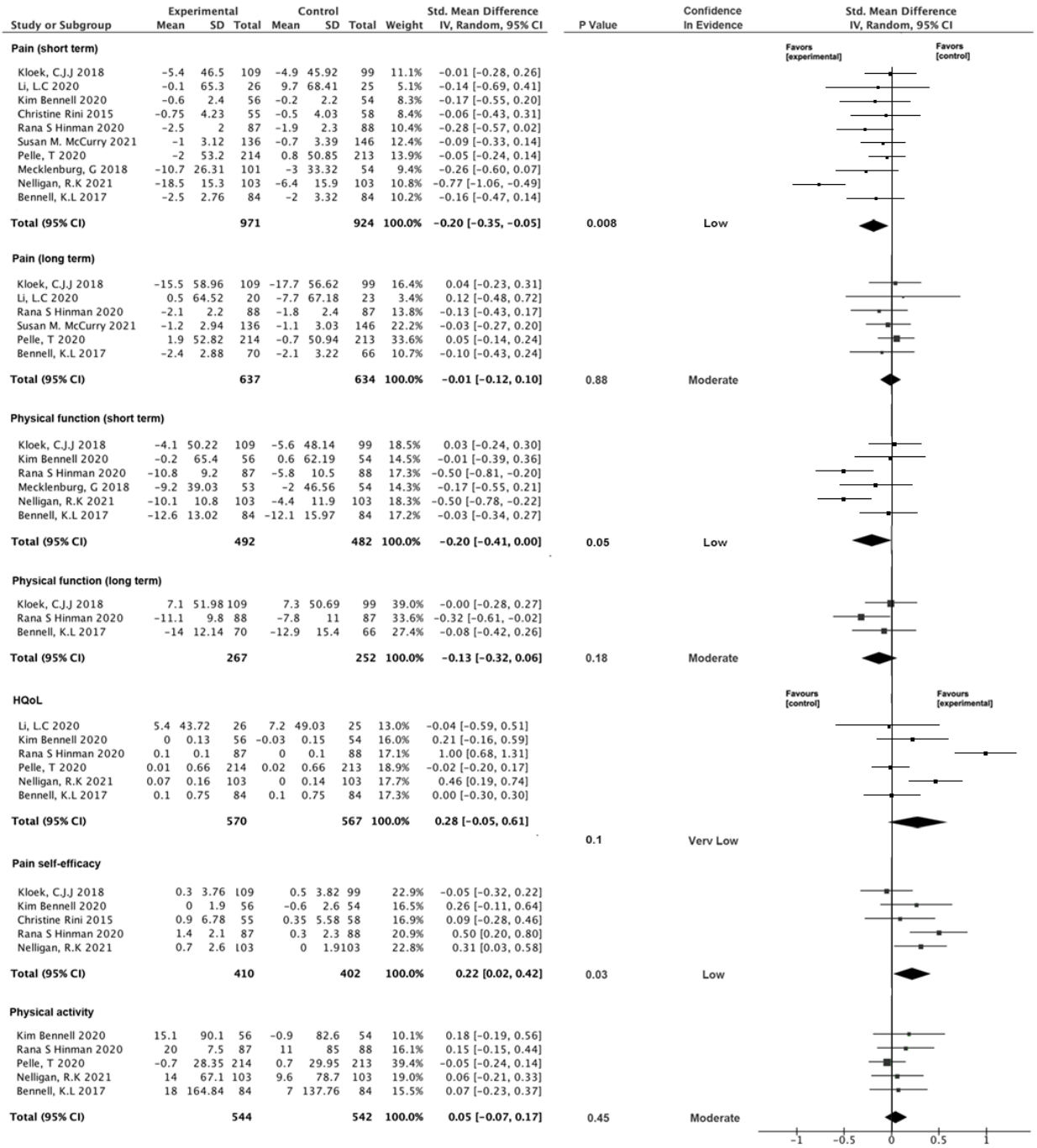

Subgroup Analyses

Subgroup analyses of behavioral therapy, digital tools, and therapist involvement were performed, and significant improvements were found in the short-term intervention.

Behavioral Therapy

BCTs (6 studies, 1177 participants) demonstrated a significant impact on pain reduction (SMD –0.25, 95% CI –0.49 to 0.00; P=.05; I^2=74%) [24,26,35,36,49,53]. However, our findings indicate that CBT (n=4 studies, n=758 participants) did not have

XSL•FO
RenderX

a significant effect on pain (SMD –0.10, 95% CI –0.24 to 0.05; P=.19; I^2=0%) [48,50-52] (Figure 4 [24,26,35,36,48-53]).

Figure 4. Pain outcomes in different behavioral therapy models.

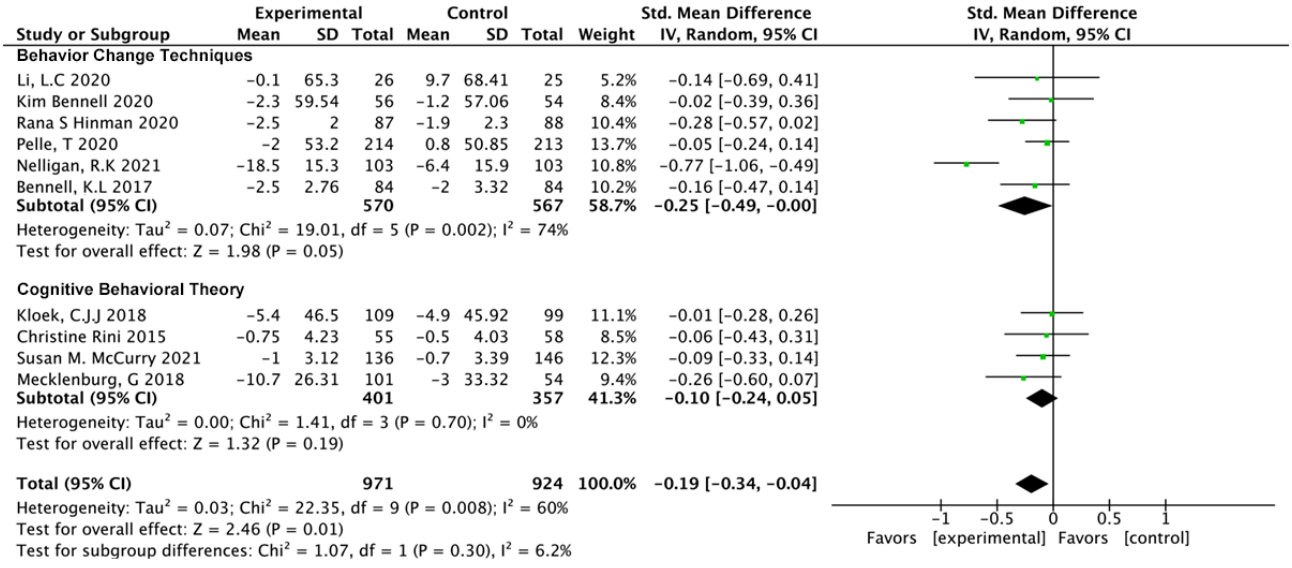

BCTs (4 studies, 659 participants) were found to significantly reduce physical functional impairment (SMD –0.28, 95% CI –0.54 to –0.01; P=.05; I^2=67%) [24,35,36,49]. In contrast, our findings show that CBT (n=2 studies, n=363 participants) did not significantly affect pain (SMD –0.20, 95% CI –0.26 to 0.16; P=.63; I^2=0%) [51,52] (Figure 5 [24,35,36,49,51,52]).

Figure 5. Physical function outcomes in different behavioral therapy models.

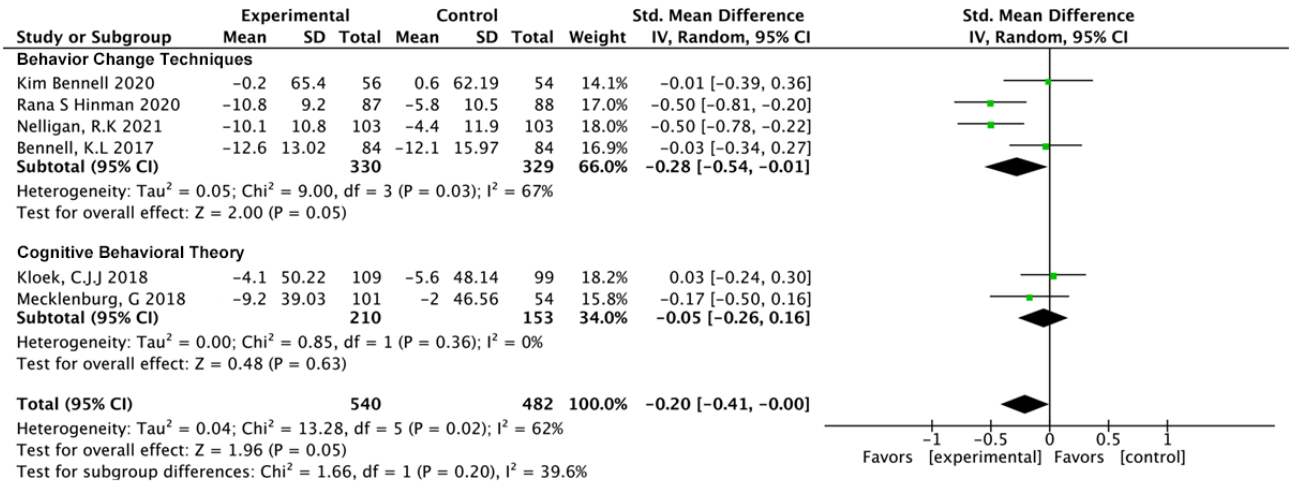

XSL•FO
RenderX

Therapist Involvement

Six studies involving 1039 participants revealed significant effects favoring the intervention in pain reduction when therapist involvement was present (SMD –0.14, 95% CI –0.27 to –0.02; P=.02; I^2=0%) [35,49-53]. Conversely, 4 studies involving 856 participants revealed no significant effects in pain reduction for interventions without therapist involvement (SMD –0.23, 95% CI –0.60 to 0.14; P=.22; I^2=85%) [24,26,36,48] (Figure 6 [24,26,35,36,48-53]).

Figure 6. Pain outcomes related to therapist involvement.

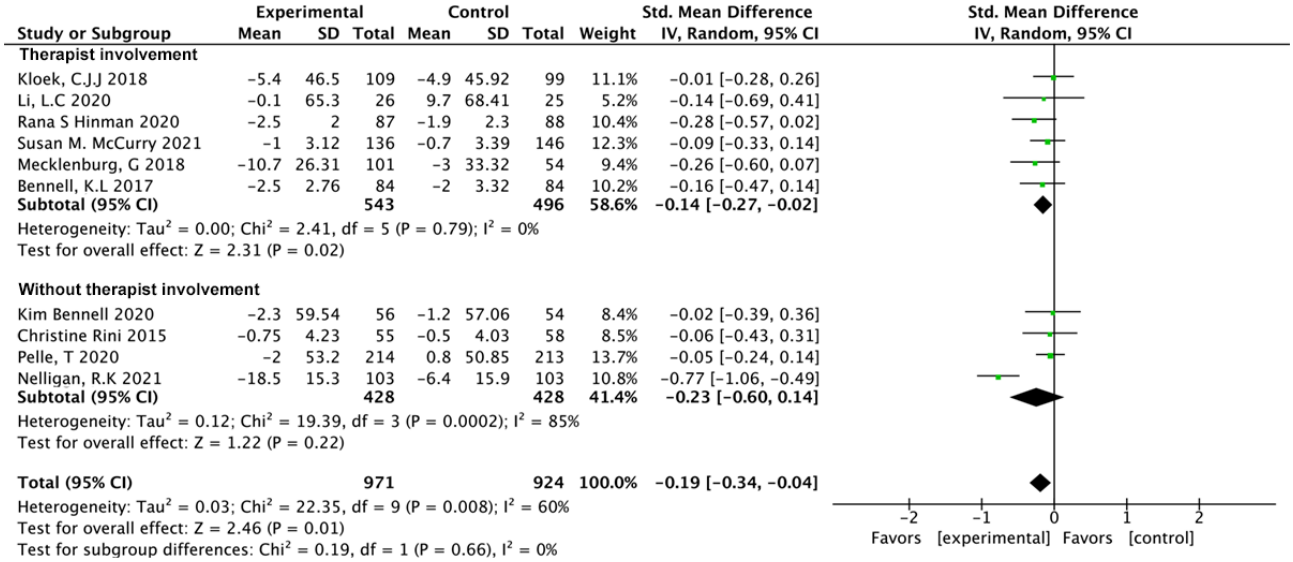

Four studies involving 706 participants (SMD –0.17, 95% CI –0.40 to 0.07; P=.17; I^2=60%) [35,49,51,52] and 2 studies involving 316 participants [24,36] (SMD –0.22, 95% CI –0.37 to –0.06; P=.26; I^2=76%) revealed that therapist involvement did not result in significant improvements in physical functional impairment (Figure 7 [24,35,36,49,51,52]).

XSL•FO
RenderX

Figure 7. Physical function outcomes related to therapist involvement.

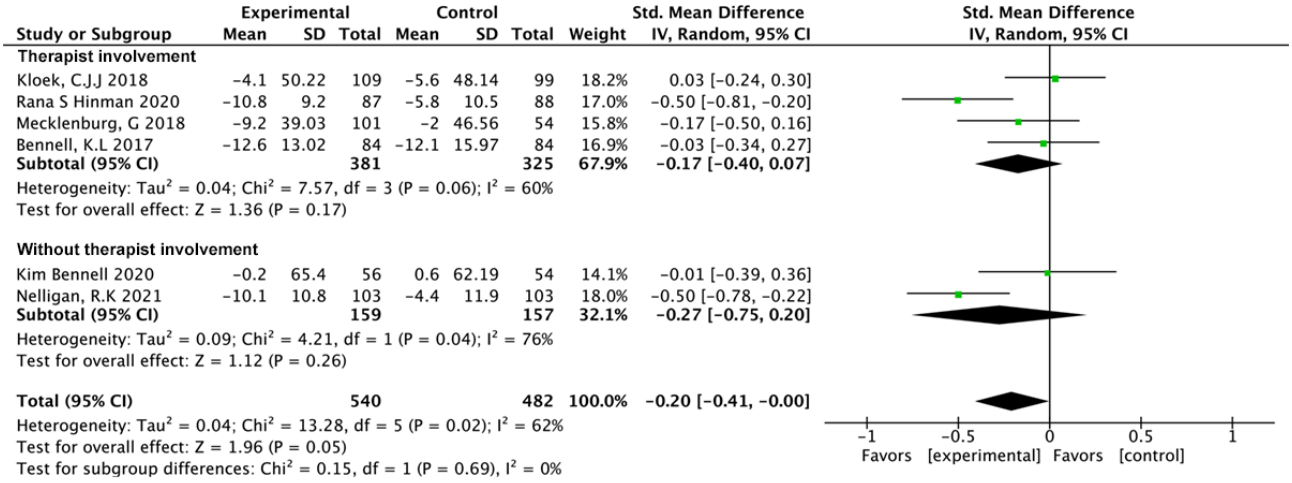

Digital Tools

Neither the use of a mobile app/website (SMD –0.23, 95% CI –0.47 to 0.01; *P*=.06; I^2=75%) [50], telephone communication (SMD –0.09, 95% CI –0.33 to 0.14; *P*=.44) [24,26,35,36,48,49] nor the use of a wearable device as a monitoring device (SMD –0.12, 95% CI –0.31 to 0.08; *P*=.24; I^2=0%) [51-53] were associated with the effect in favor of the intervention (Figure 8 [24,26,35,36,48-53]).

Figure 8. Pain outcomes with different digital tools.

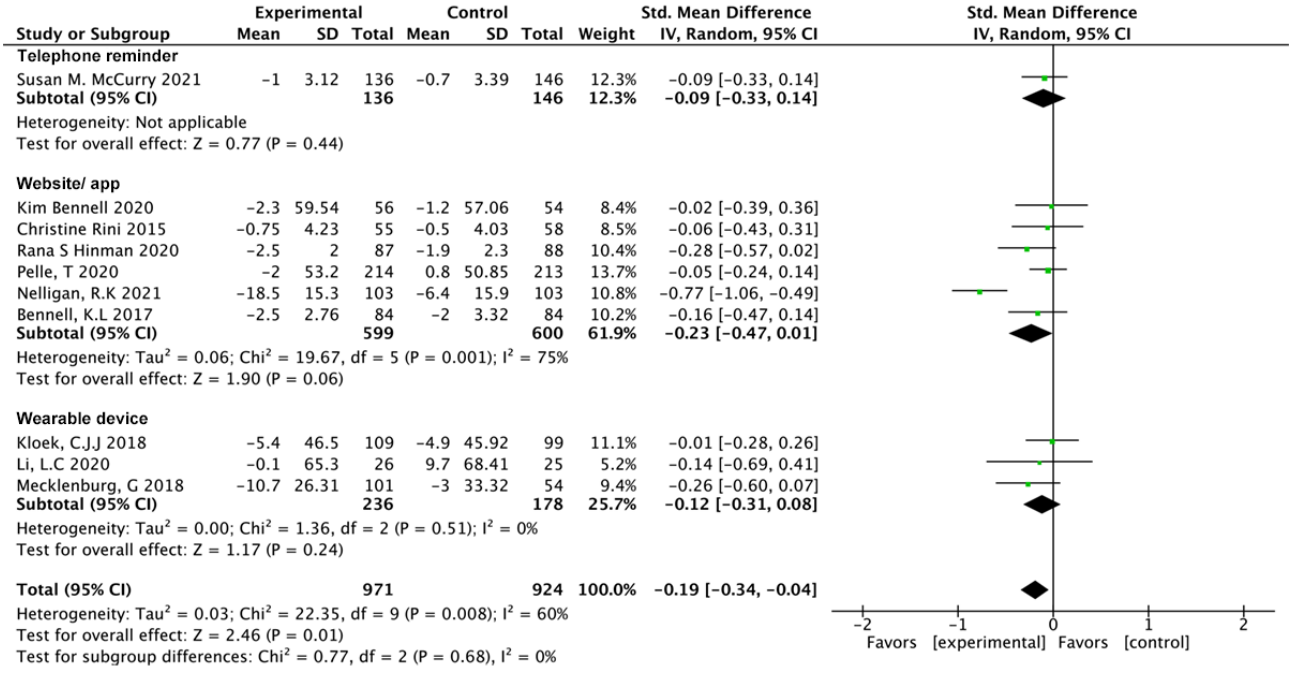

Four studies involved 659 participants who used a mobile app/website and observed a significant effect in favor of the intervention in terms of physical dysfunction (SMD –0.28, 95% CI –0.54 to –0.01; *P*=.05; I^2=67%) [24,35,36,49]. However,

XSL•FO
RenderX

there was no difference for only 2 items involving wearables (SMD –0.05, 95% CI –0.26 to 0.16; P=.63; I^2=0%) [51,52] (Figure 9 [24,35,36,49,51,52]).

Figure 9. Physical function outcomes with different digital tools.

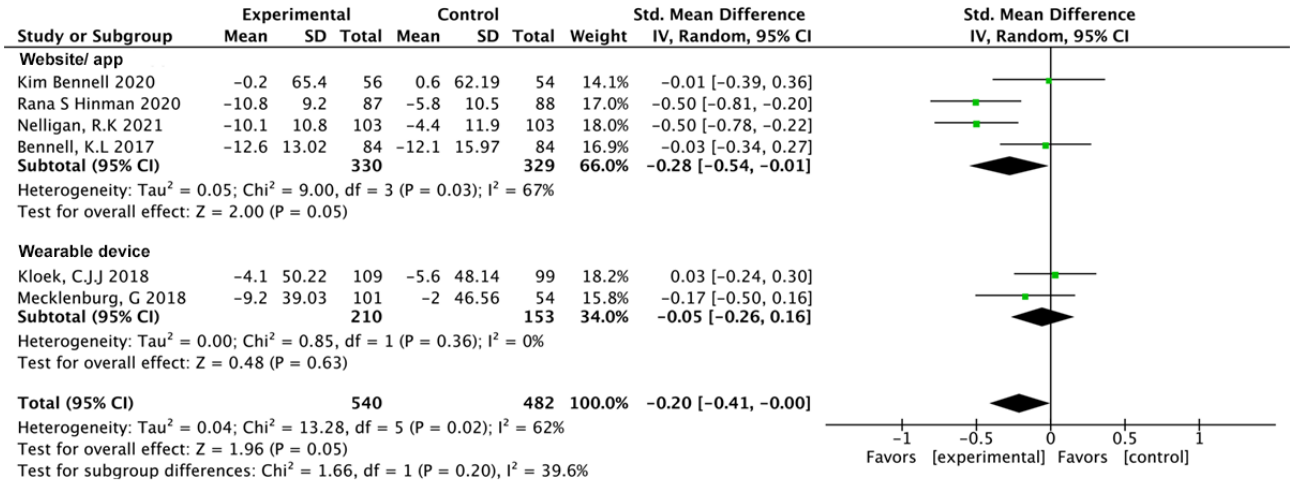

Risks of Bias Sensitivity Analysis

After excluding studies at high risk of bias, the results regarding randomization, allocation concealment, incomplete outcome reporting, selective reporting, and other biases demonstrated stability. However, instability was observed in the results when studies with unclear patient blinding were excluded. For outcomes related to physical function, studies excluding those at high risk of bias showed stability in allocation concealment and incomplete outcome reporting. Nevertheless, exclusion of trials with biases related to randomization, blinding, selective reporting, and other factors resulted in unstable findings. Therefore, these results warrant cautious interpretation (Table 3). Additionally, due to the inclusion of predominantly fewer than 10 studies, the potential impact of small-study effects was explored using only funnel plots. The inspection of funnel plots (Figure S1 of Multimedia Appendix 2) suggested small study effects/publication bias for the overall results in terms of pain in the short and medium terms.

XSL•FO
RenderX

Table 3. Sensitivity of the studies.

Variable	Number of studies (patients)	Standardized mean difference (95% CI)	I^2 (%)	Overall effect *P* value	Subgroup differences *P* value
Pain					
Randomization					.52
High risk and unclear risk	2 (363)	–0.12 (–0.37 to 0.13)	26	.34	
Low risk	8 (1532)	–0.22 (–0.40 to –0.04)	65	.02	
Allocation concealment					.18
High risk and unclear risk	1 (208)	–0.01 (–0.28 to 0.26)	0	.94	
Low risk	9 (1687)	–0.22 (–0.38 to –0.06)	60	.006	
Blinding					.30
High risk and unclear risk	6 (1216)	–0.11 (–0.23 to 0.00)	0	.05	
Low risk	4 (679)	–0.31 (–0.66 to 0.05)	81	.09	
Incomplete outcome					.14
High risk and unclear risk	5 (1123)	–0.09 (–0.20 to 0.03)	0	.15	
Low risk	5 (772)	–0.30 (–0.57 to –0.04)	70	.03	
Selective reporting					.37
High risk and unclear risk	4 (825)	–0.12 (–0.25 to 0.02)	0	.09	
Low risk	6 (1070)	–0.25 (–0.49 to 0.00)	73	.05	
Other bias					.22
High risk and unclear risk	4 (535)	–0.34 (–0.68 to 0.01)	73	.06	
Low risk	6 (1360)	–0.11 (–0.21 to –0.01)	0	.05	
Physical function					
Randomization					.20
High risk and unclear risk	4 (659)	–0.28 (–0.54 to –0.01)	67	.05	
Low risk	2 (363)	–0.05 (–0.26 to 0.16)	0	.63	
Allocation concealment					.10
High risk and unclear risk	1 (208)	0.03 (–0.24 to 0.30)	N/A[a]	.83	
Low risk	5 (814)	–0.26 (–0.47 to –0.04)	58	.02	
Blinding					.71
High risk and unclear risk	3 (343)	–0.27 (–0.73 to 0.19)	79	.25	
Low risk	3 (679)	–0.17 (–0.43 to 0.09)	63	.19	
Incomplete outcome					.20
High risk and unclear risk	2 (363)	–0.05 (–0.26 to 0.16)	0	.63	
Low risk	4 (659)	–0.28 (–0.54 to –0.01)	67	.05	
Selective reporting					.71
High risk and unclear risk	2 (285)	–0.27 (–0.75 to 0.21)	75	.27	
Low risk	4 (737)	–0.17 (–0.41 to 0.07)	64	.17	
Other bias					.94
High risk and unclear risk	3 (484)	–0.20 (–0.52 to 0.13)	69	.25	
Low risk	3 (538)	–0.21 (–0.53 to 0.11)	70	.19	

[a]N/A: not applicable. Since only 1 study was included, there was no heterogeneity in the analysis.

XSL•FO
RenderX

Discussion

Main Findings

This systematic review and meta-analysis report provides moderate- to low-quality evidence regarding the therapeutic effects of 3 digital tools based on 2 behavioral therapies for patients with OA. Digital tools focusing on behavior change have shown short-term efficacy, including reductions in pain intensity and physical function impairment along with improvements in pain self-efficacy. However, our study did not find significant differences in sports and recreational activities or overall physical activity. Given the risk of bias in the included studies, the overall quality of evidence is moderate to low. Furthermore, the evidence quality regarding HRQoL outcomes is deemed very low; so, the findings should be interpreted with caution.

Our study compares the short-term and long-term effects of behavior theory–based digital interventions on pain. However, contrary to Safari et al's [34] previous findings, we found no significant long-term effects of behavior theory–based digital interventions. This discrepancy may stem from the incomplete success of digital interventions in promoting adherence to prescribed exercise regimens among patients with musculoskeletal disease [30,54]. A recent review indicated that digital interventions have not been shown to enhance compliance with therapeutic exercise among patients with chronic musculoskeletal disease [30]. Therefore, further high-quality research is needed to conclusively determine the efficacy of these interventions for long-term therapeutic effects.

Among the 10 studies included, 9 observed positive changes in patient pain [24,26,35,36,48-51,53]. Current research integrates behavior therapy principles into digital interventions aimed at enhancing patients' ability to modify their decision-making frameworks and guide behavior change in patients with OA [55]. The educational components of digital tools include pathophysiological and etiological information on OA, guideline-based therapies, exercise for OA management, and pain management strategies, and these are closely aligned with the social cognitive theories of health behavior [19]. These address motivational and volitional determinants of exercise behavior [31,56]. Our study suggests that such semisupervised, freely accessible interventions may be an effective option for reducing pain [24], as pain intensity correlates positively with the use of exercise, courses, and health self-management messaging functionalities [57]. Additionally, improvements in physical function impairment were observed in 6 studies. Research has demonstrated the impact of digital health interventions on physical function through exercise and aerobic training programs [58,59]. This is because the digital tools included in the studies mostly incorporate multicomponent exercises recommended by clinical guidelines, which can enhance balance and mobility [60,61].

Low-quality evidence revealed that a behavioral therapy–based digital tool had a positive effect on pain self-efficacy, which is consistent with previous research that found greater self-confidence activation in digital self-monitoring programs than in traditional management programs [33,62,63]. Self-efficacy refers to an individual's perception of being capable of making positive changes in their lives can help them to be more likely to initiate and maintain positive behavior changes [64,65]. Previous studies have reported significant indirect effects of digital tools on health behaviors, and we suggest that self-efficacy could be used as an indirect mediator in clinical practice [65,66].

This review indicates that behavior therapy–based digital interventions have no significant impact on HRQoL, which is consistent with that reported in a previous systematic review [67]. However, the mechanisms underlying patients' perceptions of quality of life remain unclear, necessitating validated behavior change measures such as patient activation measures to effectively assess such systems. Additionally, we found that behavior therapy–based digital interventions do not significantly influence physical activity levels [55,68]. This may be attributed to the lack of face-to-face supervision. Nonetheless, despite these findings, such interventions could potentially offer substantial public health benefits, despite the modest effect size.

Subgroup Findings

The effects of behavioral therapy were independently assessed in a subanalysis. However, our results were not consistent with those of previous studies in that only BCT-based digital tools had an impact on outcomes in patients with OA [16,20,33]. In previous studies, it was difficult to trace patient participation on a digital platform, and multiple variations could be introduced into the outcomes [69,70]. The underlying mechanism by which behavior change–based digital tools can reduce pain and physical function in patients with OA is that such interventions can elicit patient intentions and modify their decision-making structures [55]. The educational components of the digital tools include information regarding the pathology and etiology of OA, treatment according to guidelines, exercise for OA, and pain and symptom relief, which are closely related to the social cognitive theories of health behavior [19]. These factors examine the influences of motivation and willpower on exercise behavior [31,56]. Our research indicated that this semisupervised, freely accessible intervention may be effective for pain reduction [24].

Another interesting finding in our study was that physical therapist involvement was shown to be effective in pain reduction and physical function, although with low-quality evidence. This is consistent with previous evidence suggesting that communication and interaction with professionals can increase patient motivation [67,71]. Digital tools provide the opportunity to assume the role of an aid to assist patients in engaging in less complex exercises as part of independent home training [72,73]. However, it sometimes cannot replace physical therapists who achieve similar effects [31,74,75]. Typical digital intervention tools rely on self-management and lack timely reminders and professional communication [67,71]. Studies have shown that digital tools without therapist involvement do not offer significant advantages over conventional care interventions [26]. Therefore, future digital intervention tools should aim to blend traditional OA care with digital OA care solutions. Heath care should provide solutions for personalized, comprehensive, simple, reliable, and continuous mixed heath care [76,77].

XSL•FO
RenderX

Another finding is that digital tools based on website or application development show significantly better short-term improvements in physical function compared to wearable devices. This may be due to the complex wearing process of wearable devices, especially among older adults who may have resistance to wearing such devices and find them potentially hindering daily activities. Hence, future application development should fully consider patient characteristics and needs rather than solely relying on hospital staff knowledge and experience [78,79].

Clinical and Research Implications

This study confirms that digital tools combined with behavioral therapy, especially BCTs, have moderate and clinically meaningful treatment effects in patients with hip and knee OA in the short and long term. These findings echo the current guidelines that recommend rehabilitation, emphasizing exercise, self-efficacy, and self-management for patients with OA. As digital technologies rapidly evolve, the affective interaction between patients and professionals remains a crucial factor and should be considered in future studies.

Strengths and Limitations

This study has several limitations. First, we were unable to fully explore the reasons for heterogeneity because many covariates for behavioral therapy–based digital tools in OA were not usually reported in the trials. Second, subgroup analyses to explore the potential impact of a high risk of bias and characteristics of the population were limited by the small number of included trials or because the data were poorly reported. Third, despite the use of extensive search techniques, the evaluation may have been hampered by language bias because we only included studies published in English in the 5 databases. Therefore, we cannot guarantee that all potentially eligible studies were included. Fourth, this meta-analysis included only 10 studies, with 9 RCTs and 1 RCT cluster. Owing to this rather low number of studies, the Q-static was reduced in power, and I^2 could have been biased. Finally, this study was prone to publication bias and other risks of bias and was limited only to the information reported in the studies.

Conclusions

We found moderate- and low-quality evidence supporting behavioral therapy–based digital tools improving pain intensity, physical function, and self-efficacy in the short term. Low- and very-low-quality evidence demonstrated uncertain effects of physical activity on HRQoL or its long-term effects. However, affective interactions between patients and professionals may affect clinical outcomes. Our findings should be evaluated by clinicians, stakeholders, and researchers, considering that most digital tools currently have low or very low certainty of evidence. Our findings highlight the need to conduct large-scale trials with high methodological quality.

Acknowledgments

This work was supported by the Shanghai Municipal Education Commission (2024AIYB014), Huashan Hospital, Fudan University Clinical Research Project (2023YN007), and Shanghai Soft Science Project (24692108200). We thank the members of our steering committee for their oversight and expert guidance. We also thank the stakeholders who contributed to our project management group and people with lived experience of receiving or providing exercise therapy for patients with osteoarthritis who took part in the stakeholder workshops and web-based surveys for their useful feedback.

Data Availability

Data are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BZ, DZ
Data curation: BZ
Formal analysis: BZ, XX
Funding acquisition: SZ
Investigation: HY, DZ
Methodology: BZ
Project administration: SZ
Resources: XX
Validation: DZ
Visualization: HY
Writing – original draft: BZ, DZ
Writing – review & editing: SZ, XX

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

PRISMA checklist.
[DOCX File , 32 KB - jmir_v27i1e56227_app1.docx]

Multimedia Appendix 2
Supplementary data, including search strategy, demographic data of the study participants, mechanisms for implementing the research intervention, functions of digital tools, and funnel plot for pain.
[DOCX File , 51 KB - jmir_v27i1e56227_app2.docx]

References

1. Arden N, Nevitt MC. Osteoarthritis: epidemiology. Best Pract Res Clin Rheumatol 2006 Feb;20(1):3-25. [doi: 10.1016/j.berh.2005.09.007] [Medline: 16483904]
2. Allen K, Thoma L, Golightly Y. Epidemiology of osteoarthritis. Osteoarthritis Cartilage 2022 Feb;30(2):184-195 [FREE Full text] [doi: 10.1016/j.joca.2021.04.020] [Medline: 34534661]
3. Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and beyond: a Lancet Commission. Lancet 2020 Nov 28;396(10264):1711-1712. [doi: 10.1016/S0140-6736(20)32230-3] [Medline: 33159851]
4. Quicke J, Conaghan P, Corp N, Peat G. Osteoarthritis Cartilage 2022 Feb;30(2):196-206 [FREE Full text] [doi: 10.1016/j.joca.2021.10.003] [Medline: 34695571]
5. Abramoff B, Caldera FE. Osteoarthritis: pathology, diagnosis, and treatment options. Med Clin North Am 2020 Mar;104(2):293-311. [doi: 10.1016/j.mcna.2019.10.007] [Medline: 32035570]
6. Bannuru R, Osani M, Vaysbrot E, Arden N, Bennell K, Bierma-Zeinstra S, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589 [FREE Full text] [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
7. Rostron ZPJ, Green RA, Kingsley M, Zacharias A. Efficacy of exercise-based rehabilitation programs for improving muscle function and size in people with hip osteoarthritis: a systematic review with meta-analysis. Biology (Basel) 2021 Nov 30;10(12):1251 [FREE Full text] [doi: 10.3390/biology10121251] [Medline: 34943166]
8. Whittaker J, Truong L, Dhiman K, Beck C. Osteoarthritis year in review 2020: rehabilitation and outcomes. Osteoarthritis Cartilage 2021 Feb;29(2):190-207 [FREE Full text] [doi: 10.1016/j.joca.2020.10.005] [Medline: 33242604]
9. Fan I, Govil D, Semciw A. The effectiveness of exercise based digital health interventions (requiring internet) in management of hip and knee osteoarthritis: a systematic review and meta-analysis. Osteoarthritis and Cartilage 2022 Apr;30:S402. [doi: 10.1016/j.joca.2022.02.540]
10. Zhang L, Fu T, Zhang Q, Yin R, Zhu L, He Y, et al. Effects of psychological interventions for patients with osteoarthritis: a systematic review and meta-analysis. Psychol Health Med 2018 Jan;23(1):1-17. [doi: 10.1080/13548506.2017.1282160] [Medline: 28140653]
11. Haynes SN, O'Brien WH. Functional analysis in behavior therapy. Clinical Psychology Review 1990 Jan;10(6):649-668. [doi: 10.1016/0272-7358(90)90074-k]
12. Sutton S. The contribution of behavioural science to primary care research: development and evaluation of behaviour change interventions. Prim Health Care Res Dev 2011 Jun 22;12(04):284-292. [doi: 10.1017/s1463423611000168]
13. O'Leary A, Shoor S, Lorig K, Holman HR. A cognitive-behavioral treatment for rheumatoid arthritis. Health Psychol 1988;7(6):527-544. [doi: 10.1037//0278-6133.7.6.527] [Medline: 3063517]
14. Early BP, Grady MD. Embracing the contribution of both behavioral and cognitive theories to cognitive behavioral therapy: Maximizing the richness. Clin Soc Work J 2016 Jun 14;45(1):39-48. [doi: 10.1007/s10615-016-0590-5]
15. Longmore RJ, Worrell M. Do we need to challenge thoughts in cognitive behavior therapy? Clin Psychol Rev 2007 Mar;27(2):173-187. [doi: 10.1016/j.cpr.2006.08.001] [Medline: 17157970]
16. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
17. Barlow J. How to use education as an intervention in osteoarthritis. Best Pract Res Clin Rheumatol 2001 Oct;15(4):545-558. [doi: 10.1053/berh.2001.0172] [Medline: 11567538]
18. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
19. Jönsson T, Dell'Isola A, Lohmander LS, Wagner P, Cronström A. Comparison of face-to-face vs digital delivery of an osteoarthritis treatment program for hip or knee osteoarthritis. JAMA Netw Open 2022 Nov 01;5(11):e2240126 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.40126] [Medline: 36326763]
20. Lindberg MF, Aamodt A, Badawy M, Bergvad IB, Borchgrevink P, Furnes O, et al. The effectiveness of exercise therapy and education plus cognitive behavioral therapy, alone or in combination with total knee arthroplasty in patients with knee osteoarthritis - study protocol for the MultiKnee trial. BMC Musculoskelet Disord 2021 Dec 20;22(1):1054 [FREE Full text] [doi: 10.1186/s12891-021-04924-z] [Medline: 34930194]

XSL•FO
RenderX

21. Larkin L, Gallagher S, Cramp F, Brand C, Fraser A, Kennedy N. Behaviour change interventions to promote physical activity in rheumatoid arthritis: a systematic review. Rheumatol Int 2015 Oct;35(10):1631-1640. [doi: 10.1007/s00296-015-3292-3] [Medline: 25994094]
22. Birch S, Stilling M, Mechlenburg I, Hansen TB. No effect of cognitive behavioral patient education for patients with pain catastrophizing before total knee arthroplasty: a randomized controlled trial. Acta Orthop 2020 Feb;91(1):98-103 [FREE Full text] [doi: 10.1080/17453674.2019.1694312] [Medline: 31762342]
23. Willett M, Duda J, Fenton S, Gautrey C, Greig C, Rushton A. Effectiveness of behaviour change techniques in physiotherapy interventions to promote physical activity adherence in lower limb osteoarthritis patients: A systematic review. PLoS One 2019;14(7):e0219482 [FREE Full text] [doi: 10.1371/journal.pone.0219482] [Medline: 31291326]
24. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021 Jun 01;181(6):776-785 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
25. Pearson J, Walsh N, Carter D, Koskela S, Hurley M. Developing a web-based version of an exercise-based rehabilitation program for people with chronic knee and hip pain: a mixed methods study. JMIR Res Protoc 2016 May 19;5(2):e67 [FREE Full text] [doi: 10.2196/resprot.5446] [Medline: 27197702]
26. Pelle T, Bevers K, van der Palen J, van den Hoogen F, van den Ende C. Effect of the dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthritis Cartilage 2020 Apr;28(4):418-427 [FREE Full text] [doi: 10.1016/j.joca.2020.02.831] [Medline: 32119972]
27. Dantas LO, Salvini TDF, McAlindon TE. Knee osteoarthritis: key treatments and implications for physical therapy. Braz J Phys Ther 2021;25(2):135-146 [FREE Full text] [doi: 10.1016/j.bjpt.2020.08.004] [Medline: 33262080]
28. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 03;20(1):506 [FREE Full text] [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]
29. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
30. Bunting J, Withers T, Heneghan N, Greaves C. Digital interventions for promoting exercise adherence in chronic musculoskeletal pain: a systematic review and meta-analysis. Physiotherapy 2021 Jun;111:23-30. [doi: 10.1016/j.physio.2020.08.001] [Medline: 33316865]
31. Durst J, Roesel I, Sudeck G, Sassenberg K, Krauss I. Effectiveness of human versus computer-based instructions for exercise on physical activity-related health competence in patients with hip osteoarthritis: randomized noninferiority crossover trial. J Med Internet Res 2020 Sep 28;22(9):e18233 [FREE Full text] [doi: 10.2196/18233] [Medline: 32985991]
32. Allen K, Arbeeva L, Callahan L, Golightly Y, Goode A, Heiderscheit B, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
33. Shah N, Costello K, Mehta A, Kumar D. Applications of digital health technologies in knee osteoarthritis: narrative review. JMIR Rehabil Assist Technol 2022 Jun 08;9(2):e33489 [FREE Full text] [doi: 10.2196/33489] [Medline: 35675102]
34. Safari R, Jackson J, Sheffield D. Digital self-management interventions for people with osteoarthritis: systematic review with meta-analysis. J Med Internet Res 2020 Jul 20;22(7):e15365 [FREE Full text] [doi: 10.2196/15365] [Medline: 32706657]
35. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94. [doi: 10.1002/acr.22915] [Medline: 27111441]
36. Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJ, et al. Behavior change text messages for home exercise adherence in knee osteoarthritis: randomized trial. J Med Internet Res 2020 Sep 28;22(9):e21749 [FREE Full text] [doi: 10.2196/21749] [Medline: 32985994]
37. Berry A, McCabe CS, Muir S, Walsh N. Digital behaviour change interventions to facilitate physical activity in osteoarthritis: a systematic review. Physical Therapy Reviews 2018 Jul 02;23(3):197-206. [doi: 10.1080/10833196.2018.1470747]
38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
39. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, West Sussex, England: John Wiley & Sons Ltd, The Atrium; Mar 1961.
40. Jüni P, Reichenbach S, Dieppe P. Osteoarthritis: rational approach to treating the individual. Best Pract Res Clin Rheumatol 2006 Aug;20(4):721-740. [doi: 10.1016/j.berh.2006.05.002] [Medline: 16979535]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]

XSL•FO
RenderX

42. Wientzek AV, Vigl M, Steindorf K, Brühmann B, Bergmann MM, Harttig U, et al. The improved physical activity index for measuring physical activity in EPIC Germany. PLoS One 2014;9(3):e92005 [FREE Full text] [doi: 10.1371/journal.pone.0092005] [Medline: 24642812]
43. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 24;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.ad]
44. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. PLoS One 2014;9(7):e99682 [FREE Full text] [doi: 10.1371/journal.pone.0099682] [Medline: 24992266]
45. No authors listed. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2015 Jun 17;350:h3326. [doi: 10.1136/bmj.h3326] [Medline: 26085374]
46. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 2013.
47. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 2011 Dec;64(12):1294-1302. [doi: 10.1016/j.jclinepi.2011.03.017] [Medline: 21803546]
48. Rini C, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
49. Hinman RS, Campbell PK, Lawford BJ, Briggs AM, Gale J, Bills C, et al. Does telephone-delivered exercise advice and support by physiotherapists improve pain and/or function in people with knee osteoarthritis? Telecare randomised controlled trial. Br J Sports Med 2020 Jul;54(13):790-797. [doi: 10.1136/bjsports-2019-101183] [Medline: 31748198]
50. McCurry SM, Zhu W, Von Korff M, Wellman R, Morin CM, Thakral M, et al. Effect of telephone cognitive behavioral therapy for insomnia in older adults with osteoarthritis pain: a randomized clinical trial. JAMA Intern Med 2021 Apr 01;181(4):530-538 [FREE Full text] [doi: 10.1001/jamainternmed.2020.9049] [Medline: 33616613]
51. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
52. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: a cluster-randomized controlled trial. Phys Ther 2018 Jul 01;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
53. Li LC, Feehan LM, Xie H, Lu N, Shaw CD, Gromala D, et al. Effects of a 12-week multifaceted wearable-based program for people with knee osteoarthritis: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19116 [FREE Full text] [doi: 10.2196/19116] [Medline: 32618578]
54. Chen H, Chuang T, Lin P, Lin Y, Chuang Y. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
55. Tack C. A model of integrated remote monitoring and behaviour change for osteoarthritis. BMC Musculoskelet Disord 2021 Aug 09;22(1):669 [FREE Full text] [doi: 10.1186/s12891-021-04555-4] [Medline: 34372822]
56. Xie S, Wang Q, Wang L, Wang L, Song K, He C. Effect of internet-based rehabilitation programs on improvement of pain and physical function in patients with knee osteoarthritis: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Jan 5;23(1):e21542. [doi: 10.2196/21542]
57. Cenamor J. Use of health self-management platform features: The case of a specialist ehealth app. Technological Forecasting and Social Change 2022 Dec;185:122066. [doi: 10.1016/j.techfore.2022.122066]
58. Batrakoulis A, Jamurtas AZ, Metsios GS, Perivoliotis K, Liguori G, Feito Y, et al. Comparative efficacy of 5 exercise types on cardiometabolic health in overweight and obese adults: a systematic review and network meta-analysis of 81 randomized controlled trials. Circ: Cardiovascular Quality and Outcomes 2022 Jun;15(6):e008243. [doi: 10.1161/circoutcomes.121.008243]
59. Solis-Navarro L, Gismero A, Fernández-Jané C, Torres-Castro R, Solá-Madurell M, Bergé C, et al. Effectiveness of home-based exercise delivered by digital health in older adults: a systematic review and meta-analysis. Age Ageing 2022 Nov 02;51(11):afac243 [FREE Full text] [doi: 10.1093/ageing/afac243] [Medline: 36346736]
60. Zhang M, Wang W, Li M, Sheng H, Zhai Y. Efficacy of mobile health applications to improve physical activity and sedentary behavior: a systematic review and meta-analysis for physically inactive individuals. Int J Environ Res Public Health 2022 Apr 18;19(8):4905 [FREE Full text] [doi: 10.3390/ijerph19084905] [Medline: 35457775]
61. Yadav L, Haldar A, Jasper U, Taylor A, Visvanathan R, Chehade M, et al. Utilising digital health technology to support patient-heath care provider communication in fragility fracture recovery: systematic review and meta-analysis. Int J Environ Res Public Health 2019 Oct 22;16(20):4047 [FREE Full text] [doi: 10.3390/ijerph16204047] [Medline: 31652597]
62. Seppen BF, den Boer P, Wiegel J, Ter Wee MM, van der Leeden M, de Vries R, et al. Asynchronous mHealth interventions in rheumatoid arthritis: systematic scoping review. JMIR Mhealth Uhealth 2020 Nov 05;8(11):e19260 [FREE Full text] [doi: 10.2196/19260] [Medline: 33151161]

XSL•FO
RenderX

63. Baumeister H, Kraft R, Baumel A, Pryss R, Eva-Maria M. Persuasive e-Health design for behavior change. In: Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Heidelberg: Cham: Springer International Publishing; 2022:347-364.
64. Resnick B. Self-efficacy. Middle Range Theory for Nursing 2008:1-14. [doi: 10.1891/9780826159922.0010]
65. Nash VR, Ponto J, Townsend C, Nelson P, Bretz MN. Cognitive behavioral therapy, self-efficacy, and depression in persons with chronic pain. Pain Manag Nurs 2013 Dec;14(4):e236-e243. [doi: 10.1016/j.pmn.2012.02.006] [Medline: 23273826]
66. Rabenbauer LM, Mevenkamp N. Factors in the effectiveness of e-health interventions for chronic back pain: how self-efficacy mediates e-health literacy and healthy habits. Telemed J E Health 2021 Feb;27(2):184-192. [doi: 10.1089/tmj.2019.0301] [Medline: 32397853]
67. Valentijn PP, Tymchenko L, Jacobson T, Kromann J, Biermann CW, AlMoslemany MA, et al. Digital health interventions for musculoskeletal pain conditions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Sep 06;24(9):e37869 [FREE Full text] [doi: 10.2196/37869] [Medline: 36066943]
68. Webb J, Stewart D. Let's Move with Leon. A randomised controlled trial of a UK digital intervention to improve physical activity in people with a musculoskeletal condition. Public Health 2023 Apr;217:125-132 [FREE Full text] [doi: 10.1016/j.puhe.2023.01.030] [Medline: 36878121]
69. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
70. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
71. Kroon FPB, van der Burg LRA, Buchbinder R, Osborne RH, Johnston RV, Pitt V. Self-management education programmes for osteoarthritis. Cochrane Database Syst Rev 2014 Jan 15;2014(1):CD008963 [FREE Full text] [doi: 10.1002/14651858.CD008963.pub2] [Medline: 24425500]
72. Clarkson P, Vassilev I, Rogers A, Brooks C, Wilson N, Lawson J, et al. Integrating a Web-Based Self-Management Tool (Managing Joint Pain on the Web and Through Resources) for People With Osteoarthritis-Related Joint Pain With a Web-Based Social Network Support Tool (Generating Engagement in Network Involvement): Design, Development, and Early Evaluation. JMIR Form Res 2020 Nov 26;4(11):e18565 [FREE Full text] [doi: 10.2196/18565] [Medline: 33242011]
73. Hurley DA, Keogh A, Mc Ardle D, Hall AM, Richmond H, Guerin S, et al. Evaluation of an e-learning training program to support implementation of a group-based, theory-driven, self-management intervention for osteoarthritis and low-back pain: pre-post study. J Med Internet Res 2019 Mar 07;21(3):e11123 [FREE Full text] [doi: 10.2196/11123] [Medline: 30843863]
74. Slater H, Dear BF, Merolli MA, Li LC, Briggs AM. Use of eHealth technologies to enable the implementation of musculoskeletal Models of Care: Evidence and practice. Best Pract Res Clin Rheumatol 2016 Jun;30(3):483-502 [FREE Full text] [doi: 10.1016/j.berh.2016.08.006] [Medline: 27886943]
75. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the Internet? A systematic review of randomized controlled trials. Pain® 2011;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012]
76. Östlind E, Ekvall Hansson E, Eek F, Stigmar K. Experiences of activity monitoring and perceptions of digital support among working individuals with hip and knee osteoarthritis - a focus group study. BMC Public Health 2022 Aug 30;22(1):1641 [FREE Full text] [doi: 10.1186/s12889-022-14065-0] [Medline: 36042425]
77. Stuhlreyer J, Roder C, Krug F, Zöllner C, Flor H, Klinger R. A digital application and augmented physician rounds reduce postoperative pain and opioid consumption after primary total knee replacement (TKR): a randomized clinical trial. BMC Med 2022 Dec 05;20(1):469 [FREE Full text] [doi: 10.1186/s12916-022-02638-0] [Medline: 36464680]
78. Timmers T, Janssen L, van der Weegen W, Das D, Marijnissen W, Hannink G, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323 [FREE Full text] [doi: 10.2196/15323] [Medline: 31638594]
79. Mrklas KJ, Barber T, Campbell-Scherer D, Green LA, Li LC, Marlett N, et al. Co-Design in the development of a mobile health app for the management of knee osteoarthritis by patients and physicians: qualitative study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17893 [FREE Full text] [doi: 10.2196/17893] [Medline: 32673245]

Abbreviations

BCT: behavior change technique
CBT: cognitive behavioral therapy
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HRQoL: health-related quality of life
OA: osteoarthritis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PROSPERO: International Register of Prospective Systematic Reviews

XSL•FO
RenderX

RCT: randomized controlled trial
SMD: standardized mean difference

Edited by A Mavragani; submitted 12.01.24; peer-reviewed by F Chang, DS Fonseca; comments to author 31.05.24; revised version received 09.07.24; accepted 17.10.24; published 19.03.25.

Please cite as:
Zhu B, Zhu D, Xue X, Yang H, Zhang S
Behavioral Therapy–Based Digital Interventions for Treating Osteoarthritis: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e56227
URL: https://www.jmir.org/2025/1/e56227
doi:10.2196/56227
PMID:

©Beiyao Zhu, Dian Zhu, Xiao'ao Xue, Hongyi Yang, Shurong Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Digital Lifestyle Interventions on Depression, Anxiety, Stress, and Well-Being: Systematic Review and Meta-Analysis

Jacinta Brinsley[1], PhD; Edward J O'Connor[2], PhD; Ben Singh[1], PhD; Grace McKeon[3], PhD; Rachel Curtis[1], PhD; Ty Ferguson[1], PhD; Georgia Gosse[4], BPT; Iris Willems[5,6], MSc; Pieter-Jan Marent[5,7], MSc; Kimberley Szeto[1], BHSc; Joseph Firth[8,9], PhD; Carol Maher[1], PhD

[1]Alliance for Research in Exercise, Nutrition and Activity, University of South Australia, Adelaide, Australia

[2]Health & Biosecurity, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Adelaide, Australia

[3]Discipline of Psychiatry and Mental Health, School of Clinical Medicine, Nutrition, Exercise and Social Equity (NExuS), University of New South Wales, Sydney, Australia

[4]IIMPACT in Health, University of South Australia, Adelaide, Australia

[5]Department of Movement and Sports Sciences, Research Centre for Aging Young, Ghent University, Ghent, Belgium

[6]Research Foundation Flanders, Brussels, Belgium

[7]Department of Movement Sciences, KU Leuven, Leuven, Belgium

[8]Division of Psychology and Mental Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom

[9]Manchester Academic Health Science Centre, Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom

Corresponding Author:
Jacinta Brinsley, PhD
Alliance for Research in Exercise, Nutrition and Activity
University of South Australia
108 North Terrace
Adelaide, 5000
Australia
Phone: 61 8 8302 6558
Email: jacinta.brinsley@unisa.edu.au

Abstract

Background: There is a growing body of robust evidence to show that lifestyle behaviors influence mental health outcomes. Technology offers an accessible and cost-effective implementation method for interventions, yet the study of the effectiveness of interventions to date has been specific to the mode of delivery, population, or behavior.

Objective: The primary aim of this review was to comprehensively evaluate the effectiveness of digital lifestyle interventions for improving symptoms of depression, anxiety, stress, and well-being as coprimary outcomes in adults. The secondary aim was to explore the technological, methodological, intervention-specific, and population-specific characteristics that were associated with major changes in mental health outcomes.

Methods: A systematic search was conducted across the MEDLINE, CINAHL, Embase, Emcare, PsycINFO, and Scopus databases to identify studies published between January 2013 and January 2023. Randomized controlled trials of lifestyle interventions (physical activity, sleep, and diet) that were delivered digitally; reported changes in symptoms of depression, anxiety, stress, or well-being in adults (aged ≥18 years); and were published in English were included. Multiple authors independently extracted data, which was evaluated using the 2011 Levels of Evidence from the Oxford Centre for Evidence-Based Medicine. Inverse-variance random-effects meta-analyses were used for data analysis. The primary outcome was the change in symptoms of depression, anxiety, stress, and well-being as measured by validated self-report of clinician-administered outcomes from pre- to postintervention. Subgroup analyses were conducted to determine whether results differed based on the target lifestyle behavior, delivery method, digital features, design features, or population characteristics.

Results: Of the 14,356 studies identified, 61 (0.42%) were included. Digital lifestyle interventions had a significant small-to-medium effect on depression (standardized mean difference [SMD] −0.37; $P<.001$), a small effect on anxiety (SMD −0.29; $P<.001$) and stress (SMD −0.17; $P=.04$), and no effect on well-being (SMD 0.14; $P=.15$). Subgroup analyses generally

XSL•FO
RenderX

suggested that effects were similar regardless of the delivery method or features used, the duration and frequency of the intervention, the population, or the lifestyle behavior targeted.

Conclusions: Overall, these results indicate that delivering lifestyle interventions via a range of digital methods can have significant positive effects on depression (P<.001), anxiety (P<.001), and stress (P=.04) for a broad range of populations, while effects on well-being are inconclusive. Future research should explore how these interventions can be effectively implemented and embedded within health care with a concerted focus on addressing digital health equity.

Trial Registration: PROSPERO CRD42023428908; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023428908

(J Med Internet Res 2025;27:e56975) doi:10.2196/56975

KEYWORDS

depression; anxiety; stress; well-being; mental health; lifestyle intervention; physical activity; sleep; diet; digital health; mobile phone

Introduction

Background

Mental disorders are among the leading causes of global disease burden [1] and a significant risk factor for premature mortality [2]. Recent estimates suggest that between 5% and 19% of global disability-adjusted life-years can be attributed to mental disorders [1,3], causing annual economic losses of approximately US $4.7 trillion [4]. These mental health difficulties cause enormous psychological impact, have widespread deleterious effects on health behaviors [5], and vastly increase the odds of experiencing physical health morbidities [6].

There is increasing recognition that poor lifestyle behaviors, such as diet, physical activity, and sleep, contribute to both poor physical and mental health. Concurrently, there is increasing evidence for the role of lifestyle interventions in preventing, managing, and treating mental illness [7-9]. Numerous clinical guidelines and health policy documents now call for lifestyle behaviors to be addressed alongside pharmacotherapies and psychotherapy, as part of mental health care [10].

Digital technology offers a promising avenue for revolutionizing mental health care delivery on a global scale. Digital interventions can also reduce the stigma associated with seeking help [11], while data-driven insights allow for personalized and efficient interventions tailored to individual needs.

In recent years, an increasing number of studies have demonstrated the effectiveness of digital interventions in delivering mental health treatments [12], including compelling outcomes for smartphone-delivered psychological interventions in reducing depression and anxiety symptoms [13,14]. Concurrently, research related to digital lifestyle interventions is increasing rapidly [15]. A growing body of evidence shows that these interventions can be effective for improving health behaviors such as physical activity [16], sleep quality [17], weight loss [18], and healthy diet [19]. However, there is little evidence specifically focused on the effect of digital lifestyle interventions on mental health and well-being outcomes in the general population. The few reviews in this area have focused on singular lifestyle behaviors such as physical activity [20], specific clinical populations [21], or a single delivery method [22] or have reported lifestyle behavior changes but not changes in mental health [19,23]. Moreover, only 1 review has synthesized data from >10 randomized controlled trials (RCTs) [19]. This fragmented approach limits our ability to rigorously discern the effectiveness of the various intervention components and combinations on mental health across the general population.

Objectives

This systematic review and meta-analysis takes a broad approach, aiming to provide a thorough synthesis of the evidence on digitally delivered lifestyle interventions for depression, anxiety, stress, and well-being outcomes in adults. Recognizing that mental health is more than the absence of disorders and distress [24], we also evaluated the effectiveness of digital lifestyle interventions on psychological well-being (ie, well-being), as operationalized by van Agteren et al [25]. Here, well-being refers to positive aspects of mental health, including positive affect and life satisfaction (ie, subjective well-being), as well as meaning, purpose, and related concepts (ie, psychological well-being).

Given the breadth of the review, the secondary aim was to explore the technology-specific intervention features that are associated with greater mental health outcomes. We further considered the impact of various methodological (ie, study quality), intervention-specific (ie, type, delivery method, and delivery features), and population-specific (ie, general population, those with a mental illness, and those with a physical illness) characteristics. Finally, this review aimed to shed light on the overall quality of the evidence provided in the meta-analyses and discuss the implications of the evidence for future research and lifestyle intervention delivery.

Methods

This systematic review and meta-analysis adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [26] and is registered with PROSPERO (CRD42023428908). A completed PRISMA checklist for this study is available in Multimedia Appendix 1.

Search Strategy and Selection of Studies

The following electronic databases were searched from January 2013 to January 10, 2023: MEDLINE (Ovid), CINAHL (EBSCOhost), Embase (Ovid), Emcare (Ovid), PsycINFO (Ovid), and Scopus. Subject heading, keyword, and Medical Subject Headings term searches relating to technology,

XSL•FO
RenderX

modifiable lifestyle behaviors (ie, physical activity, diet, and sleep), mental health, and study design were used. Full search queries are detailed in Multimedia Appendix 2. Additional searches were conducted using Google Scholar and by scanning the reference lists of included papers and review articles to identify any additional studies. Studies were exported to Covidence (Veritas Health Innovation Ltd), and duplicates were removed. Studies were screened in duplicate with conflicts resolved between the 2 reviewers or a third independent reviewer.

Inclusion and Exclusion Criteria

To be included in this systematic review and meta-analysis, studies needed to meet the following criteria: (1) involve adults (aged ≥18 years) of any health status and (2) include a digitally delivered lifestyle intervention targeting physical activity, diet, sleep, or any combination thereof. We defined digital lifestyle interventions as interventions delivered via digital technologies (such as apps, websites, and wearable devices) that were self-guided and did not require real-time clinician delivery (eg, behavior change and education) and aimed to improve lifestyle behaviors [27]. Considering the nature of the research, we included interventions that were supplemented with contact for follow-up or adherence and technology support (eg, orientation session or phone call to create a user profile or a peer-driven social forum moderated by a clinician). Lifestyle interventions were eligible if at least 50% of the intervention was focused on physical activity, diet, and sleep. Physical activity was considered in the broadest sense, including overall physical activity levels, structured exercise interventions, and studies examining reducing physical inactivity (ie, sedentary time). Diet was considered as any intervention targeting food intake (ie, amount or type). Studies focusing on supplementation or specific nutrient treatments were excluded. Sleep interventions were considered as anything targeting sleep; therefore, psychological therapies targeting sleep (eg, cognitive behavioral therapy for insomnia) were eligible for inclusion; (3) studies that had an adequate control condition of no treatment, treatment as usual, waitlist control, or attention (ie, sham) control were included, while control groups that received an alternative intervention comprising behavior change techniques or an in-person version of the digital intervention were excluded, as they would limit our ability to evaluate the effectiveness of the intervention; (4) studies that reported changes in symptoms of depression, anxiety, psychological stress, or well-being using validated tools were included, and due to the broad and varied definitions of well-being [28], we considered well-being measures listed as per the review by van Agteren et al [25]. Studies that focused on momentary affective states were considered outside the scope of this review and were excluded; (5) RCTs were included. No restriction was placed on the setting or context of the included studies.

Data Synthesis and Analysis

In total, 2 reviewers independently extracted data using a predefined data extraction sheet, cross-checked the data, assessed study quality, and resolved disagreements by discussion or referral to a third reviewer (JB or CM). The authors of the included studies were contacted to provide additional data for inclusion in the meta-analysis if required.

Intervention effect sizes (ie, differences between intervention and control groups) for outcome data were calculated across all studies and standardized to Hedges *g* [29] along with the SE. Studies were grouped by outcomes (ie, depression, anxiety, stress, and well-being) for analyses. Random-effects meta-analyses were conducted due to expected heterogeneity. All analyses used an inverse variance method with restricted maximum-likelihood estimator for τ^2 and Hartung-Knapp adjustment for the random effects model. Standardized mean differences (SMDs) were used as the effect measure for meta-analyses to allow comparison of data from different scales. If means and SDs were not reported in a study, authors were contacted twice before means and SDs were calculated based on available data using recommended formulas (eg, using sample size, median, and range) [30]. Meta-analyses were conducted only when at least 5 studies were included. Effect sizes were categorized as small (0.2), medium (0.5), or large (≥0.8) [31], with a significance level set at .05 [32]. Cochrane *Q* test was used to assess statistical heterogeneity, and the I^2 statistic was used to quantify the proportion of the overall outcome effect attributed to heterogeneity. The following cutoff values for the I^2 statistic were used: 0%-29%=no heterogeneity, 30%-49%=moderate heterogeneity, 50%-74%=substantial heterogeneity, and 75%-100%=considerable heterogeneity [33]. Publication bias was assessed with visual inspection of funnel plots and tested using the Begg-Mazumdar [34] and Egger regression method [35], with a *P* value <.05 suggesting the presence of bias. Where significant bias was detected, a Duval and Tweedie [36] trim-and-fill analysis was applied. Heterogeneity was explored with subgroup and sensitivity analyses. Sensitivity analyses were conducted by removing outliers, studies with samples <100, studies of poor to fair quality, and those with attrition >25% to investigate changes in effect for each outcome. All analyses were conducted in R software (version 4.2.0, “Mountain Hydrangea”; R Foundation for Statistical Computing).

Quality Assessment

The risk of bias of each study was assessed using the Physiotherapy Evidence Database Scale (PEDro) [37]. The PEDro scale comprises 10 questions, each answered as yes or no. A study was deemed to be high quality with a score of 6 to 10, fair quality with a score of 4 to 5, and low quality with a score <4.

The overall level of evidence was graded using the Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence [38] as follows: grade A—consistent level 1 studies (ie, individual RCTs with narrow CIs), grade B—consistent level 2 (ie, individual cohort studies or low-quality RCTs) or level 3 studies (ie, individual case-control studies) or extrapolations from level 1 studies, grade C—level 4 studies or extrapolations from level 2 or 3 studies, or grade D—level 5 (ie, expert opinion without explicit critical appraisal) evidence or inconsistent or inconclusive studies of any level.

XSL•FO
RenderX

Subgroup Analyses

The central aim of this review was to examine the effect of various methodological, intervention-specific, and population-specific factors on the effectiveness of digital lifestyle interventions for mental health outcomes [39]. Intervention-specific factors included the lifestyle behavior; the intervention targeted (ie, physical activity alone, diet alone, sleep alone, herein referred to as the respective singular behavior, or any combination of these behaviors, which is referred to herein as multibehavioral); delivery method (ie, web based, app, email or text, and wearable device); features of delivery format (ie, use of a chatbot or gamification); frequency and duration of the intervention; whether the intervention was co-designed, individualized, or publicly available; and whether it contained a theoretical underpinning. The frequency of intervention delivery was split into three categories: (1) intended daily use; (2) 1 to 4 times per week; and (3) "other," which included fortnightly, monthly, varied, once-off, and self-paced engagement. Methodological factors were control group types and study quality. Study quality was based on the risk of bias, where studies were categorized into poor, fair, good, or excellent as per the PEDro scoring. Control groups were divided into no treatment, including waitlist; treatment as usual; or attention control. To determine which populations might benefit most from digital lifestyle interventions, we conducted subgroup analyses for each outcome for apparently healthy adults; those with a physical health condition; those diagnosed with a mental health condition; and, due to the number of sleep-specific studies, people who met the criteria for insomnia.

Deviation From the Registered Protocol

We originally planned to include interventions targeting alcohol, smoking, and substance use as behaviors, considering their association with poor mental health outcomes. However, due to the unfeasible volume of search results and their distinct focus on addiction, these lifestyle behaviors were excluded from this review.

Results

Included Studies

Searching of databases yielded 14,357 results. Following the removal of duplicates, 9727 potentially eligible studies remained for which abstracts were screened. At full-text stage, 230 studies were reviewed, and 169 were removed because they failed to meet the inclusion criteria (Multimedia Appendix 3). The remaining 61 studies were included in the review, and 59 studies were included in the quantitative synthesis (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA flowchart of study selection.

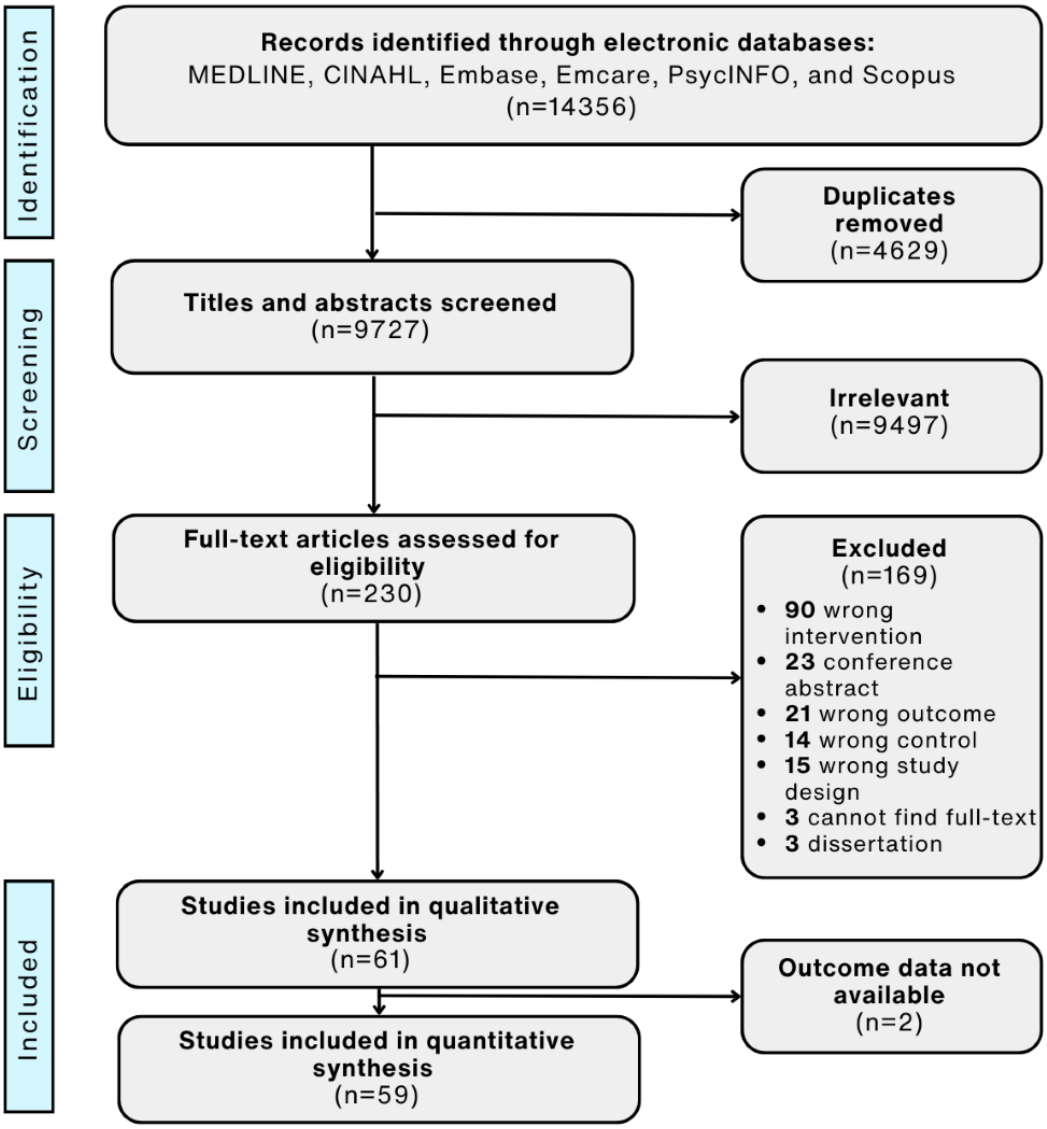

Study Characteristics

In total, 22,483 participants across 61 studies were included in the review. Sample sizes ranged between 20 and 3755 participants, and the mean age ranged between 19 and 68 years. Interventions lasted between 2 and 52 weeks. Studies originated from Europe, America, Canada, Australia, New Zealand, the United Kingdom, Iran, and Asia. Full details of each study are outlined in Table 1.

XSL•FO
RenderX

Table 1. Descriptive characteristics of the included studies.

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Abbott et al [40], 2021	United States	Individuals experiencing major depressive disorder; 71; 86%	Efficacy for depressive symptoms	PA[c], diet, and sleep	Web based	Waitlist	6	No	PHQ-9[d] (secondary)	No	6
Abedi et al [41], 2015	Iran	Postmenopausal women; 106; 100%	Effectiveness for depression, anxiety, and insomnia	PA	Wearable	No treatment	12	Yes	BDI[e] (primary)	No	6
Bade et al [42], 2021	United States	Individuals with advanced-stage lung cancer; 40; 75%	Feasibility, effectiveness for PA, QoL[f], depression	PA	Wearable, app, and SMS text messages	Treatment as usual	12	Yes	PHQ-9 (secondary)	No	7
Bailey et al [43], 2020	United Kingdom	Individuals with diabetes; 20; 50%	Feasibility	PA and sleep	App	Treatment as usual	8	Yes	WEMWBS[g] (secondary)	No	7
Bennion et al [44], 2020	United States	Low-income postpartum women; 370; 100%	Effectiveness for weight loss	PA and diet	Web based and SMS text messages	Treatment as usual	52	Yes	EPDS[h] (secondary)	No	7
Bossen et al [45], 2013	Netherlands	Individuals with hip and knee OA[i]; 199; 65%	Effectiveness for PA	PA	Web based and emails	Waitlist	9	Yes	HADS[j] (secondary)	Yes; Join2Move	7
Brindal et al [46], 2019	Australia	Individuals who had 5% weight loss in past 2 years; 88; 75%	Effectiveness for weight, food, exercise, mood, and stress	PA and diet	App	Attention control	12	Yes	DASS[k]-21 and SLS[l]-5 (primary)	Yes: MotiMate	7
Carli et al [47], 2022	Italy and Portugal	Individuals with somatic disorders; 425; 44%	Effectiveness for depressive symptoms	PA, diet, and sleep	App and sensorized shirt	Treatment as usual	12	Yes	BDI-II (primary)	Yes; NEVERMIND	8
Chee et al [48], 2016	United States	Asian American women with depressive symptoms; 33; 100%	Preliminary efficacy for depressive symptoms	PA	Web based	No treatment	12	Yes	CES-D[m] (primary)	No	5
Cheng et al [49], 2019	United States	Individuals experiencing insomnia; 1358; 38%	Efficacy for depression	Sleep	Web based and emails	Attention control	12	Yes	CES-D (primary)	Yes; Sleepio	9

XSL•FO
RenderX

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Christensen et al [50], 2016	Australia	Individuals experiencing insomnia and depressive symptoms; 1149; 74%	Effectiveness for depressive symptoms	Sleep	App and emails	Attention control	6	Yes	PHQ-9 (primary)	Yes; Sleepio	9
Conner et al [51], 2017	New Zealand	Young adults; 115; 67%	Effectiveness for fruit or vegetable consumption	Diet	SMS text messages	No treatment	2	No	CES-D and HADS (primary)	No	7
Devi et al [52], 2014	United Kingdom	Individuals with stable angina; 95; 26%	Effectiveness for cardiac rehabilitation	PA and diet	Website	Treatment as usual	6	Yes	HADS (secondary)	Yes; ActivateYourHeart	8
Duan et al [53], 2017	China	University students; 493; 71%	Effectiveness for PA and fruit or vegetable consumption	PA and diet	Website	No intervention	8	Yes	CES-D (secondary)	No	4
Duan et al [54], 2018	China	Individuals with coronary heart disease; 114; 57%	Efficacy for PA and fruit or vegetable consumption	PA and diet	Website	Treatment as usual	8	Yes	CES-D (secondary)	No	5
Eberts et al [55], 2015	Germany	Teachers with insomnia symptoms; 128; 74%	Efficacy for work-related stress and sleep	Sleep	Website	Waitlist	8	Yes	CES-D (secondary)	No	7
Eckert et al [56], 2022	United States	Individuals with allogenic bone marrow transplant; 72; 55%	Feasibility	PA	Website	Attention control	12	No	PROMIS[n] (exploratory)	Yes; Udaya	7
Edney et al [57], 2020	Australia	Community; 284; 74%	Effectiveness for PA	PA	App	Waitlist	12	Yes	DASS-21 (secondary)	No	6
Espie et al [58], 2019	United Kingdom	Individuals experiencing insomnia; 1712; 78%	Effectiveness for health and well-being	Sleep	App	Treatment as usual	12	Yes	WEMWBS (primary), PHQ-9, and GAD-7[o] (secondary)	Yes; Sleepio	8
Falk et al [59], 2022	United States	Sedentary employees; 47; 74%	Associations with well-being, stress, mood, and fatigue	PA	Website	Waitlist	12	Yes	PSS[p] (exploratory)	No	7

XSL•FO
RenderX

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Felder et al [60], 2020	United States	Pregnant women; 208; 100%	Efficacy for insomnia	Sleep	App	Treatment as usual	10	Yes	EPDS and GAD-7 (secondary)	Yes; Sleepio	7
Freeman et al [61], 2017	United Kingdom	University students with insomnia; 3755; 71%	Effectiveness for insomnia, paranoia, and hallucinations	Sleep	App	Treatment as usual	10	Yes	PHQ-9, GAD-7, and WEMWBS (secondary)	Yes; Sleepio	10
Glozier et al [62], 2019	Australia	Older men experiencing depression; 87; 0%	Effectiveness for depression and insomnia	Sleep	App	Attention control	12	Yes	CES-D (primary)	Yes; SHUTi	9
Golsteijn et al [63], 2018	Netherlands	Patients with cancer; 478; 13%	Efficacy for PA	PA	Web based and wearable	Treatment as usual	12	Yes	HADS (secondary)	No	7
Hershner and O'Brien [64], 2018	United States	College students; 551; 57%	Effectiveness for sleep behavior	Sleep	Web based	No intervention	8	No	PHQ-9 (secondary)	No	4
Hilmarsdottir et al [65], 2020	Iceland	Individuals with type 2 diabetes; 37; 51%	Effectiveness for metabolic health	PA and diet	App	Not reported	52	Yes	HADS and PAID[q] (exploratory)	Yes; Sidekick-Health	7
Horsch et al [66], 2017	Netherlands	Individuals experiencing insomnia; 151; 63%	Efficacy for insomnia	Sleep	App	Waitlist	7	Yes	HADS (secondary)	Yes; Sleepcare	8
Houchen-Wolloff et al [67], 2018	United Kingdom	Individuals with coronary heart disease; 60; 10%	Feasibility of web-based cardiac rehab	PA	Website	Treatment as usual	8	Yes	HADS (exploratory)	Yes; ActivateYourHeart	8
Huberty et al [68], 2019	United States	Individuals with myoloproliferative neoplasm; 62; 94%	Feasibility and preliminary effectiveness for QoL	PA	Website	Waitlist	12	No	PROMIS (exploratory)	No	8
Ifejika et al [69], 2020	United States	Individuals who experienced stroke; 36; 44%	Feasibility and preliminary effectiveness for monitoring diet	Diet	App	Attention control	26	Yes	PHQ-9 (secondary)	Yes; Lose it!	6
Joutsenniemi et al [70], 2014[r]	Finland	General population; 3274; 83%	Effectiveness of PA versus positive psychology on happiness	PA	Email	Attention control	16	No	BDI (secondary)	No	7

XSL•FO
RenderX

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Katz et al [71], 2018	United States	Individuals with rheumatoid arthritis; 96; 88%	Effectiveness for PA and fatigue	PA	Wearable	Treatment as usual	21	Yes	PHQ-8 (secondary)	No	7
Kuhn et al [72], 2022	United States	Veterans; 50; 42%	Feasibility, acceptability, and preliminary efficacy for insomnia	Sleep	App	Waitlist	6	Yes	PHQ-8 and GAD-7 (secondary)	Yes; insomnia coach	9
Lambert et al [73], 2018	United Kingdom	Individuals who experienced depression; 62; 84%	Feasibility and acceptability in community setting	PA	Web based	Treatment as usual	8	Yes	PHQ-8 and GAD-7 (secondary)	No	6
Lambert et al [74], 2022	Canada	Individuals with prostate cancer; 49; 0%	Feasibility, acceptability, and clinical significance for anxiety and QoL	PA	Web based	Waitlist	10	Yes	HADS (secondary)	No	7
Lopez et al [75], 2019	France	Individuals experiencing insomnia; 46; 74%	Efficacy for insomnia	Sleep	Web based	Treatment as usual	12	Yes	BDI (secondary)	No	7
Lorenz et al [76], 2019	Switzerland	Individuals experiencing insomnia; 56; 69%	Efficacy for insomnia	Sleep	Web based	Waitlist	6	Yes	BDI and APSQ[s] (secondary)	Yes; Mementor somnium	9
McGrath et al [77], 2017	United Kingdom	Individuals with elevated blood pressure; 134; 61%	Efficacy for reducing blood pressure	Sleep	Web based and text or email	Treatment as usual	8	Yes	BDI and BAI[t] (secondary)	Yes; Sleepio	9
Mensorio [78], 2019	Spain	Individuals with obesity and hypertension; 106; 80%	Efficacy for promoting lifestyle changes	PA and diet	Web based	Treatment as usual	12	Yes	DASS-21 (secondary)	No	7
Mueller et al [79], 2022	United Kingdom	Individuals who are overweight or obese; 388; 78%	Effectiveness for bodyweight, eating behavior, PA, and well-being	PA and diet	Web based and email	Treatment as usual	12	Yes	PHQ-9 and GAD-7 (secondary)	No	8
Murawski et al [80], 2019	Australia	Individuals who are inactive and experience poor sleep; 160; 80%	Efficacy for sleep and PA	PA and sleep	App, wearable, and email or text	Waitlist	12	Yes	DASS-21 (secondary)	Yes; balanced	6

XSL•FO
RenderX

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Nystrom et al [81], 2017	Sweden	Individuals experiencing depression; 312; 76%	Evaluate or compare PA and behavior activation	PA	App and text or email	Waitlist	12	Yes	PHQ-9 and GAD-7 (secondary)	No	7
Philippot et al [82], 2022	Belgium	Higher education students; 30; 92%	Feasibility for psychological symptoms	PA	Web based	No intervention	4	No	DASS-21 (secondary)	No	7
Przybylko et al [83], 2021	Australia	General population; 508; 70%	Effectiveness for mental health and well-being	PA and diet	Web based	Waitlist	10	Yes	DASS-21 (primary)	Yes; The Lift Project	6
Puig-Ribera et al [84], 2017	Spain	Office workers; 264; 65%	Effectiveness for presenteeism, well-being, and performance	PA	Web based and wearable	Attention control	19	Yes	WEMWBS (secondary)	No	6
Puterman et al [85], 2022	Canada	Inactive general population; 334; 87%	Effectiveness for depressive symptoms	PA	App	Waitlist	6	No	CES-D (primary)	No	6
Ritterband et al [86], 2012	United States	Patients with cancer experiencing insomnia; 28; 86%	Effectiveness for insomnia	Sleep	Web based and email	Waitlist	9	No	HADS (secondary)	Yes; SHUT-i	8
Savard et al [87], 2014[r]	Canada	Individuals with breast cancer; 242; 100%	Efficacy of video-based cognitive behavioral therapy in breast cancer	Sleep	DVD	No intervention	6	Yes	HADS (secondary)	No	10
Serrat et al [88], 2022	Spain	Individuals with fibromyalgia; 330; 97%	Effectiveness for fibromyalgia	PA	Web based	Treatment as usual	12	No	HADS (secondary)	No	7
Shaffer et al [89], 2022	United States	Older adults; 311; 69%	Effectiveness for depression and anxiety	Sleep	Web based and email	Attention control	10	Yes	HADS (primary)	Yes; SHUT-i	3
Spanhel et al [90], 2022	Germany	Refugees; 66; 27%	Feasibility for insomnia and sleep	Sleep	Web based	Waitlist	4	Yes	PHQ-9 (secondary)	No	8
Stiglbauer et al [91], 2019	Austria	University students; 105; 63%	Efficacy of self-tracking device on health	PA	App and wearable	Waitlist	2	Yes	PERMA[u]-Profiler (secondary)	No	5

XSL•FO
RenderX

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Teychenne et al [92], 2021	Australia	Postpartum women; 62; 100%	Feasibility and acceptability for health behaviors	PA	Web based	No intervention	12	Yes	EPDS and GAD-7 (secondary)	No	7
Vandelanotte et al [93], 2022	Australia	General population; 501; 72%	Efficacy on depression, anxiety, stress, and QoL	PA	Web based	Treatment as usual	12	Yes	DASS-21 (primary)	No	6
Wan et al [94], 2017	United States	Individuals with coronary obstructive pulmonary disease; 118; 2%	Effectiveness for increasing step count	PA	Web based and wearable	Attention control	12	Yes	BDI (secondary)	No	8
Wang et al [95], 2022	China	Individuals with PCOS[v]; 122; 100%	Effectiveness for PCOS	PA and diet	App	Treatment as usual	52	Yes	SDS[w] and SAS[x] (secondary)	No	7
Wang F and Boros S [96], 2020	Hungary	Young adults; 54; 53%	Effectiveness for sleep quality, stress, and life satisfaction	PA	Wearable	No intervention	4	Yes	PSS-4 and SWLS[y] (secondary)	No	5
Wong et al [97], 2021	China	Individuals with depressive symptoms; 79; 85%	Efficacy for depressive symptoms	PA, diet, and sleep	App	Waitlist	8	Yes	PHQ-9 and GAD-7 (primary)	No	7
Young et al [98], 2021	Australia	Individuals who are overweight or obese; 125; 0%	Effectiveness for weight loss and depression	PA, diet, and sleep	Web based and wearable	No intervention	12	Yes	PHQ-9 (primary) and GAD-7 (secondary)	Yes; SHED-IT	7
Xuto et al [99], 2022	Thailand	Pregnant women; 66; 100%	Effectiveness for health behaviors and anxiety	PA and diet	SMS text messages	Treatment as usual	28	No	STAI[z] (secondary)	No	8

XSL•FO
RenderX

Study	Country	Sample population; N; female (%)	Intervention						Mental health outcome	Publicly available	Risk of bias, PEDro[a] score
			Study aim	Behaviors targeted	Delivery format	Comparator	Duration (wk)	Included BCTs[b]			
Yudi et al [100], 2017	Australia	Individuals with acute coronary syndrome; 2016; 16%	Efficacy for exercise capacity and cardiac risk factor	PA	App	Treatment as usual	8	Yes	CDS[aa] and HADS-A (secondary)	No	8

[a]PEDro: Physiotherapy Evidence Database Scale.
[b]BCT: behavior change technique.
[c]PA: physical activity.
[d]PHQ-9: 9-item Patient Health Questionnaire.
[e]BDI: Beck Depression Inventory.
[f]QoL: Quality of life.
[g]WEMWBS: Warwick-Edinburgh Mental Well-Being Scale.
[h]EPDS: Edinburgh postnatal depression scale.
[i]OA: osteoarthritis.
[j]HADS: Hospital Anxiety and Depression Scale.
[k]DASS: Depression Anxiety and Stress Scale.
[l]SLS: selective laser sintering.
[m]CES-D: Centre for Epidemiological Studies Depression.
[n]PROMIS: Patient-Reported Outcomes Measurement Information System.
[o]GAD-7: Generalized Anxiety Disorder-7.
[p]PSS: Perceived Stress Scale.
[q]PAID: problem areas in diabetes.
[r]Not included in quantitative analyses.
[s]APSQ: Perceived Stress Scale.
[t]BAI: bay-annulated indigo.
[u]PERMA: Positive Emotions, Engagement, Relationships, Meaning, Accomplishment.
[v]PCOS: Polycystic Ovarian Syndrome.
[w]SDS: Self-Rating Depression Scale.
[x]SAS: Self-Rating Anxiety Scale.
[y]SWLS: Satisfaction with Life Scale.
[z]STAI: State Trait Anxiety Inventory.
[aa]CDS: Cardiac Depression Scale.

A total of 55 studies assessed depressive symptoms [40-42,44-58,60-83,85-90,92-95,97,98,100], 35 assessed anxiety symptoms [45,46,50-52,57,58,60,61,63,65-68,72-74,76-83,86,88,89,92,93,95,97-100], 11 assessed stress [44,46,57,59,65,78,80,82,83,93,96], and 6 assessed well-being [43,46,61,84,91,96]. Overall, 24 studies included apparently healthy adults [44,46,48,50,51,53,55,57,59,64,70,76,80,82-85,89-93,99]; 23 included adults with a physical health condition [43,45,47,52,54,56,63,65,67-69,71,74,77-79,86,88,94,95,98,100]; 6 included adults with a mental health condition, all of which were depressive disorders [40,41,62,73,81,97]; and a further 8 included adults with insomnia [49,58,60,61,66,72,75,87]. Interventions were digitally delivered via the web (38/61, 62%) [40,44,45,48-50,52-56,58-64,67,68,73-79,82-84,86,88-90,92-94,98], app (17/61, 28%) [42,43,46,47,57,60,65,66,69,72,80,81,85,91,95,97,100], email, or SMS text messages (15/61, 25%) [44,46,51,58,61,63,70,77,79-81,86,89,99,100], and 16% (10/61) of the interventions included a wearable activity tracker [41,42,57,63,71,80,91,94,96,98]. A total of 15 interventions used >1 delivery method [42,44,57,58,61,63,77,79-81,86,89,91,94,98]. Of the 61 interventions, 26 (43%) targeted physical activity [41,42,45,48,56,57,59,63,67,68,70,71,73,74,81,82,84,85,88,91-94,96,100], 17 (29%) targeted sleep [49,50,55,58,60-62,64,66,72,75-77,86,87,89,90], 2 (3%) targeted diet [51,69], and 16 (27%) targeted multiple lifestyle behaviors [42,44,57,58,61,63,79-81,86,89,91,94,98]. Of the 17 studies targeting multiple behaviors, 4 (24%) targeted physical activity, diet, and sleep [40,47,97,98]; 11 (65%) targeted physical activity and diet [44,46,52-54,65,78,79,83,95,99]; and 2 (12%) targeted physical activity and sleep [43,80].

XSL•FO
RenderX

Mental health (ie, depression, anxiety, stress, and well-being) outcomes were measured as a primary outcome in 14 studies [41,46-51,62,83,85,89,93,97,98]. The risk of bias of the included studies is detailed in Multimedia Appendix 4.

Meta-Analyses

For 2 of the included studies, the outcome data of interest was not presented in the manuscript and was unavailable after contacting the corresponding author [70,87].

Publication Bias

Visual inspection of funnel plots (Multimedia Appendix 5) suggests that there was no publication bias for depression (Kendall τ=−0.093, P=.33; Egger regression=0.19, P=.85), anxiety (Kendall τ=−0.106, P=.38; Egger regression=0.90, P=.37), or stress (Kendall τ=−0.055, P=.88; Egger regression=−0.11, P=.92).

Heterogeneity

There was significant heterogeneity across study data for depression (Q=166; P<.001; I^2=69%) and anxiety (Q=55; P=.01; I^2=39%). Stress and well-being data had moderate, nonsignificant heterogeneity (Q=17; P=.07; I^2=42% and Q=8; P=.15; I^2=39%, respectively).

Overall Effects of Digital Lifestyle Interventions on Symptoms of Depression

Pooled effects from 53 studies (ie, 12,385 participants) showed a small-to-medium significant effect of digital lifestyle interventions for reducing depressive symptoms in comparison to all control conditions (SMD=−0.37; 95% CI −0.46 to −0.27; P<.001; Figure 2).

XSL•FO
RenderX

Figure 2. Meta-analysis of the effects of digital lifestyle interventions on symptoms of depression [40-42,44-58,60-69,71-83,85,86,88-90,92-95,97,98,100]. SMD: standardized mean difference.

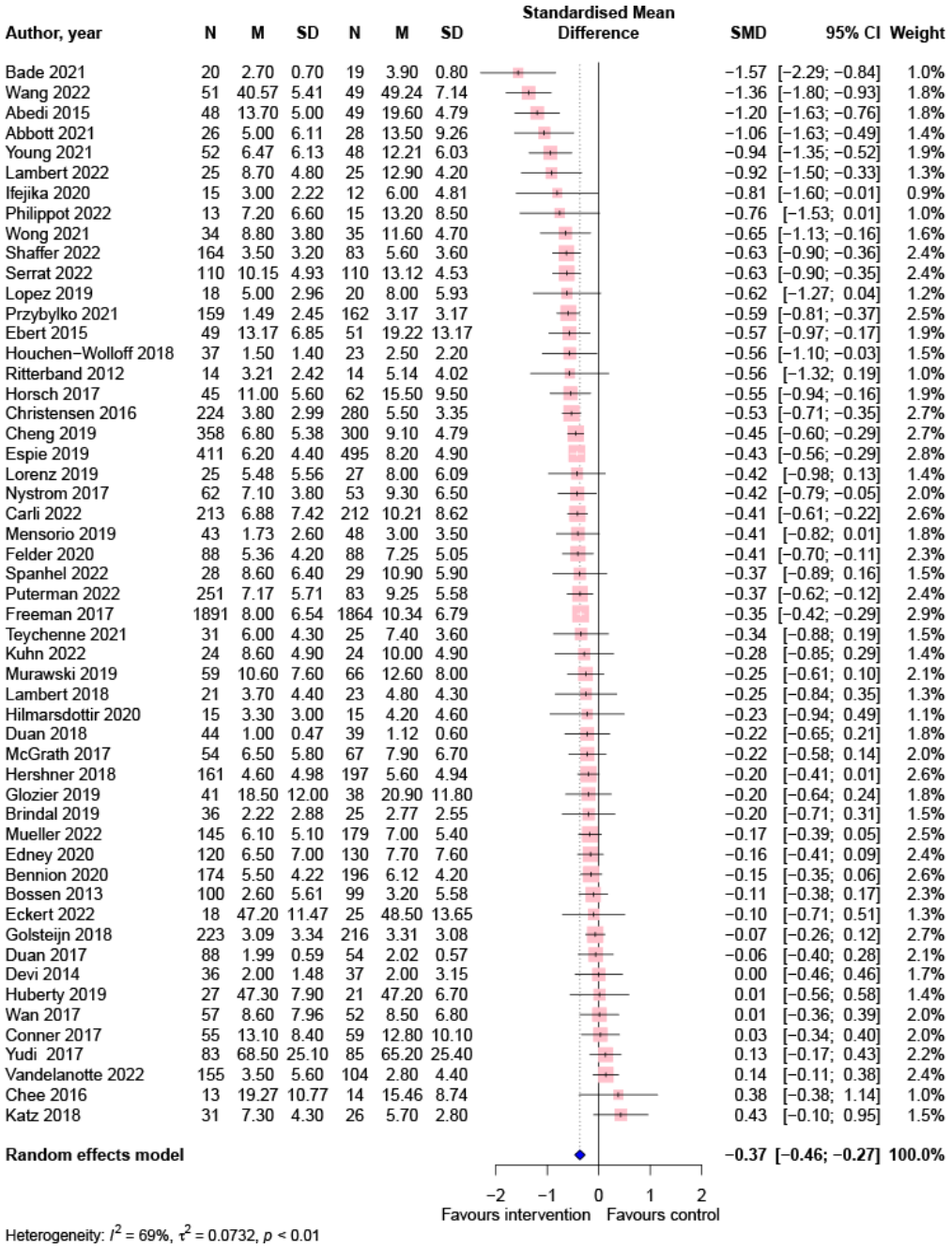

Overall Effects of Digital Lifestyle Interventions on Symptoms of Anxiety

Figure 3 displays the pooled effect size from 35 studies (9383/22,483, 41.73%), showing a small-to-medium positive effect of digital lifestyle interventions compared to all control conditions on symptoms of anxiety (SMD=−0.29; 95% CI −0.36 to −0.21; *P*<.001).

XSL•FO
RenderX

Figure 3. Meta-analysis of the effects of digital lifestyle interventions on symptoms of anxiety [45,46,50-52,57,58,60,61,63,65-68,72,73,75-83,86,88,89,92,93,95,97-100]. SMD: standardized mean difference.

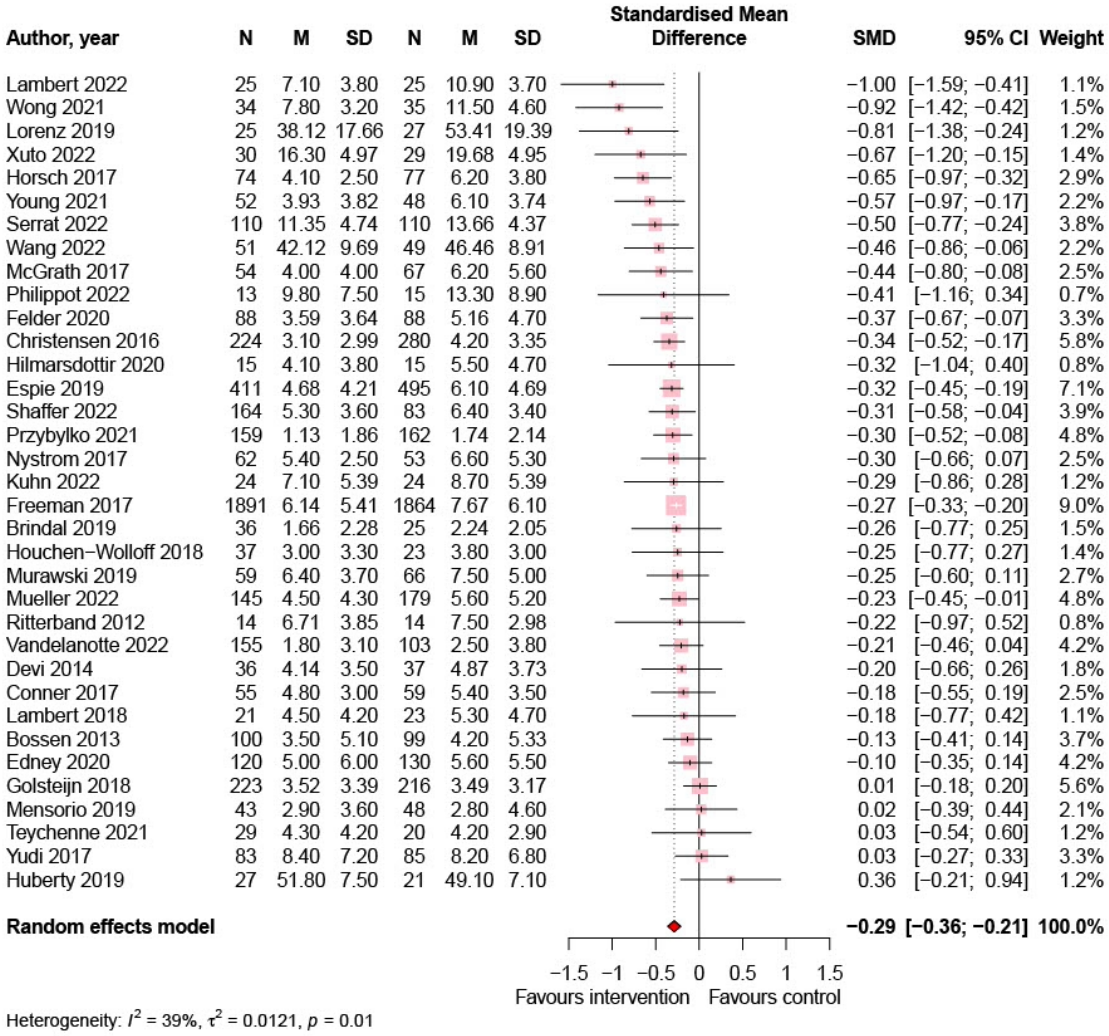

Overall Effects of Digital Lifestyle Interventions on Stress

Pooled effects from 11 studies (n=1608) showed a small positive effect of digital lifestyle interventions compared to all control conditions on stress (SMD=−0.17; 95% CI −0.33 to −0.01; *P*=.04; Figure 4).

XSL•FO
RenderX

Figure 4. Meta-analysis of the effects of digital lifestyle interventions on stress [44,46,57,59,65,78,80,82,83,93,96]. SMD: standardized mean difference.

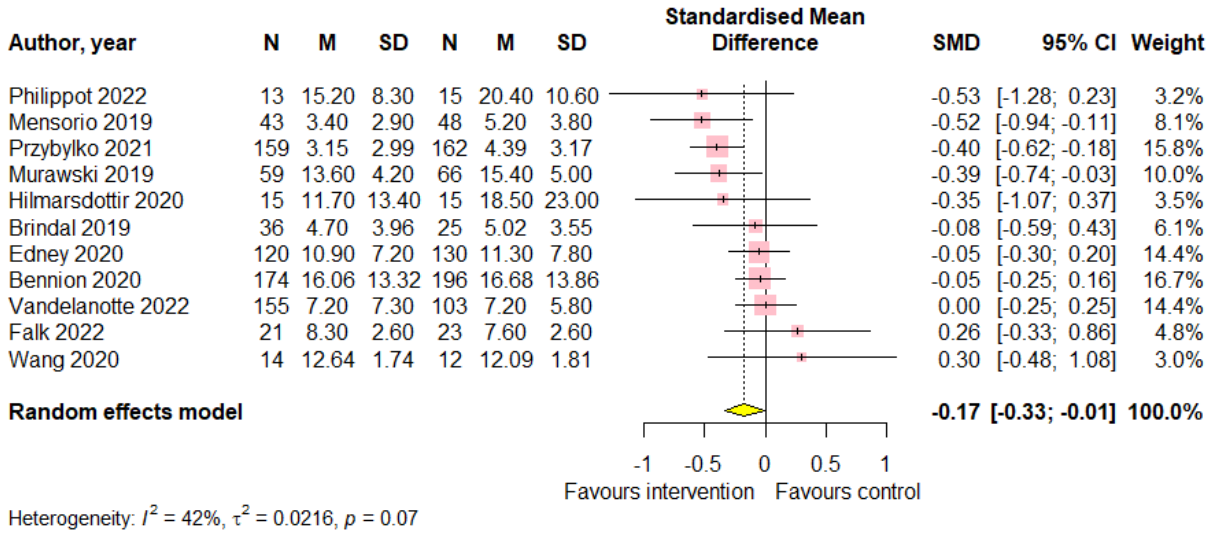

Overall Effects of Digital Lifestyle Interventions on Well-Being

Pooled effects from 6 studies (4204/22,483, 18.7%) suggest no significant effect of digital lifestyle interventions compared to all control conditions on well-being (SMD=0.14; 95% CI −0.07 to 0.37; *P*=.15; Figure 5), although the data show a nonsignificant (*P*=.15) trend in the direction that favors digital lifestyle interventions.

Figure 5. Meta-analysis of the effects of digital lifestyle interventions on well-being [43,46,61,84,91,96]. SMD: standardized mean difference.

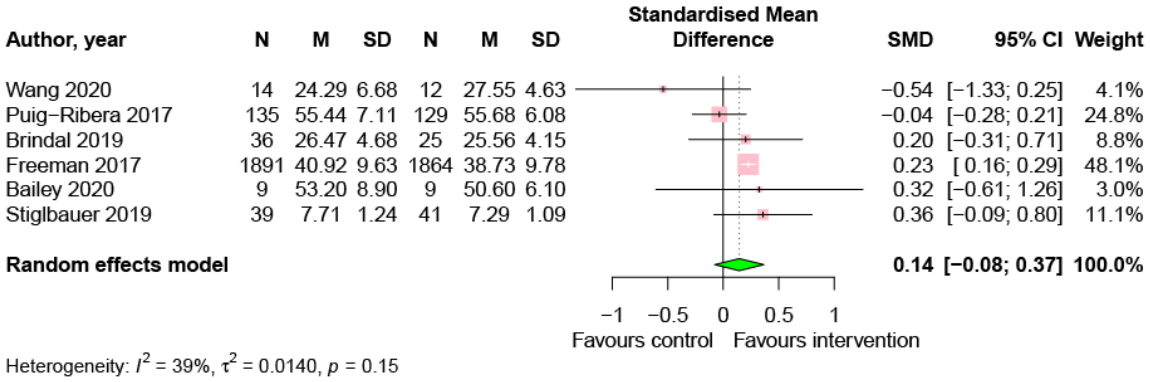

Subgroup Analyses

Overview

To understand which aspects of digital lifestyle interventions contribute to their effectiveness in improving mental health and well-being, and for whom they are most beneficial, we conducted a series of comparative subgroup analyses. These analyses were based on technology-specific features, methodological approaches, intervention characteristics, and population characteristics. Due to the limited number of studies for stress and well-being, these analyses were conducted for depressive and anxiety outcomes only. The results are outlined in Tables 2 and 3.

XSL•FO
RenderX

Table 2. Subgroup analyses of depression outcomes.

	Number of studies, k	Intervention, n	Control, n	Meta-analysis		Heterogeneity		Subgroup test, *P* value
				Standardized mean difference (95% CI)	*P* value	*Q*	I^2 (%)	
Overall	53	6285	6100	−0.37 (−0.4641 to −0.2673)	<.0001	166.32	68.7	—[a]
Population characteristics								
General population	17	1792	1586	−0.30 (−0.4252 to −0.1753)	<.001	43.41	63.1	.46
Mental diagnosis	7	288	288	−0.51 (−0.9005 to −0.1131)	.02	21.72	67.8	—
Physical diagnosis	21	1370	1373	−0.36 (−0.5828 to −0.1455)	<.001	87.88	77.2	—
Insomnia	7	2835	2853	−0.39 (−0.4428 to −0.3350)	<.001	3.40	3.4	—
Intervention characteristics								
Behavior								
Diet	2	70	71	−0.31 (−5.5627 to 4.9394)	.59	3.55	71.8	.75
Physical activity	20	1445	1197	−0.29 (−0.5138 to −0.0708)	.01	79.58	76.1	—
Sleep	16	3595	3639	−0.41 (−0.4662 to −0.3437)	<.001	14.58	0	—
Multiple behavior	15	1175	1193	−0.43 (−0.6452 to −0.2173)	<.001	52.71	73.4	—
Frequency								
Daily	17	1336	1368	−0.32 (−0.5658 to −0.0695)	.02	69.12	76.9	.40
1-4 times per wk	26	1880	1598	−0.34 (−0.4473 to −0.2368)	<.001	53.52	53.3	—
Other[b]	10	3069	3134	−0.52 (−0.8015 to −0.2339)	<.001	42.91	79	—
Duration								
≤6 wk	9	682	582	−0.38 (−0.6239 to −0.1410)	.006	16.47	51.4	.95
7-16 wk	37	5244	5172	−0.35 (−0.4639 to −0.2454)	<.001	113.61	68.3	—
≥4 mo	7	359	346	−0.40 (−0.9378 to 0.1295)	.11	34.87	82.8	—
Individualized								
No	16	1003	1054	−0.40 (−0.6365 to −0.1716)	<.001	64.61	76.8	.64
Yes	37	5282	5046	−0.35 (−0.4538 to −0.2400)	<.001	101.5	64.5	—
Theoretical underpinning								
No	31	4715	4603	−0.39 (−0.4850 to −0.2945)	<.001	60.33	50.3	.69
Yes	22	1570	1497	−0.35 (−0.5497 to −0.1444)	<.001	69.35	78.2	—
Co-designed								

XSL•FO
RenderX

	Number of studies, k	Intervention, n	Control, n	Meta-analysis		Heterogeneity		Subgroup test, *P* value
				Standardized mean difference (95% CI)	*P* value	*Q*	I^2 (%)	
No	47	5927	5707	−0.38 (−0.4864 to −0.2714)	<.001	154.81	70.3	.09
Yes	6	358	393	−0.20 (−0.4351 to 0.0271)	0.07	7.47	33.1	—
Publicly available								
No	32	2330	2147	−0.35 (−0.5054 to −0.1878)	<.001	123.4	74.9	.42
Yes	21	3955	3953	−0.42 (−0.4961 to −0.3381)	<.001	29.88	33.1	—
Delivery method								
Wearable								
No	45	5675	5494	−0.36 (−0.4473 to −0.2698)	<.001	110.04	60	.74
Yes	8	610	606	−0.44 (−0.9897 to 0.1122)	.10	53.74	87	—
App								
No	38	5169	5142	−0.33 (−0.4450 to −0.2248)	<.001	117.66	68.6	.31
Yes	15	1116	958	−0.46 (−0.6866 to −0.2263)	<.001	47.98	70.8	—
Web based								
No	17	1162	1004	−0.44 (−0.6971 to −0.1767)	<.001	75.3	78.8	.46
Yes	36	5123	5096	−0.34 (−0.4329 to −0.2451)	<.001	90.78	61.4	—
Text or email								
No	40	2930	2704	−0.39 (−0.5114 to −0.2767)	<.001	118.39	67.1	.31
Yes	13	3355	3396	−0.29 (−0.4770 to −0.0945)	<.001	44.25	72.9	—
Multiple methods								
No	39	2839	2618	−0.37 (−0.4907 to−0.2499)	<.001	119.97	68.3	.82
Yes	14	3446	3482	−0.35 (−0.5333 to −0.1598)	<.001	45.17	71.2	—
Delivery features								
Video								
No	36	2494	2537	−0.41 (−0.5435 to −0.2760)	<.001	112.77	69	.24
Yes	17	3791	3563	−0.30 (−0.4402 to −0.1596)	<.001	53.10	69.9	—
Chatbot								
No	47	5134	4848	−0.35 (−0.4665 to −0.2397)	<.001	156.18	70.5	.08
Yes	6	1151	1252	−0.46 (−0.5112 to −0.4055)	<.001	1.22	0	—
Gamification or interaction								

XSL•FO
RenderX

	Number of studies, k	Intervention, n	Control, n	Meta-analysis		Heterogeneity		Subgroup test, *P* value
				Standardized mean difference (95% CI)	*P* value	*Q*	I^2 (%)	
No	40	3193	3031	−0.36 (−0.4785 to −0.2362)	<.001	129.65	69.9	.69
Yes	13	3092	3069	−0.40 (−0.5797 to −0.2176)	<.001	34.28	65	—
Notifications								
No	32	4102	3898	−0.33 (−0.4756 to −0.1915)	<.001	109.38	71.7	.45
Yes	21	2183	2202	−0.41 (−0.5399 to −0.2705)	<.001	54.74	63.5	—
Social features								
No	37	5149	4961	−0.36 (−0.4726 to −0.2525)	<.001	105.75	66	.96
Yes	16	1136	1139	−0.37 (−0.5950 to −0.1426)	<.001	60.26	75.1	—
Methodological characteristics								
Control group								
Waitlist control or none	36	4135	3988	−0.44 (−0.5629 to −0.3111)	<.001	101.68	65.6	.11
Treatment as usual	11	1626	1522	−0.22 (−0.4330 to −0.0081)	.04	47.08	78.8	—
Attention	6	524	590	−0.27 (−0.5498 to 0.0089)	.06	13.32	62.5	—
Outcome measure								
Primary	13	1751	1444	−0.39 (−0.6066 to −0.1666)	.002	50.31	76.1	.64
Secondary	36	4437	4572	−0.37 (−0.4934 to −0.2438)	<.001	108.51	67.7	—
Exploratory	4	97	84	−0.23 (−0.6696 to 0.1958)	0.18	2.40	0	—
Study quality								
Poor to fair	5	470	387	−0.23 (−0.6161 to 0.1573)	.17	11.44	65.0	.53
Good	41	3198	3113	−0.39 (−0.5169 to −0.2660)	<.001	146.70	72.7	—
Excellent	7	2617	2600	−0.39 (−0.4768 to −0.3020)	<.001	5.79	0	—
Publication year								
2014 or earlier	3	150	150	−0.12 (−0.5666 to 0.3230)	.36	1.59	0	.01
2015-2017	12	2669	2690	−0.29 (−0.5450 to −0.0515)	.02	73.81	77.0	—
2018-2020	20	1947	2005	−0.27 (−0.3636 to −0.1753)	<.001	45.06	75.6	—
2021or later	18	1519	1255	−0.57 (−0.7795 to −0.3635)	<.001	30.47	37.7	—

[a]Not applicable.

[b]Other frequency includes fortnightly, monthly, once-off, and self-paced.

XSL•FO
RenderX

Table 3. Subgroup analyses of anxiety outcomes.

	Number of studies, k	Intervention, n	Control, n	Meta-analysis		Heterogeneity		Subgroup test, *P* value
				Standardized mean difference (95% CI)	*P* value	*Q*	I^2 (%)	
Overall	35	4689	4694	−0.29 (−0.3620 to −0.2116)	<.001	55.42	38.7	__[a]
Population								
General population	11	905	916	−0.28 (−0.3829 to −0.1731)	<.001	9.84	0	.94
Mental diagnosis	4	160	159	−0.33 (−0.9751 to 0.3087)	.19	8.52	64.8	—
Physical diagnosis	14	972	988	−0.26 (−0.4282 to −0.0940)	<.001	29.52	56.0	—
Insomnia	6	2652	2631	−0.30 (−0.3947 to −0.2113)	<.001	5.60	10.7	—
Intervention characteristics								
Behavior								
Diet	1	55	59	−0.18 (−0.5505 to 0.1858)	.33	0	0	.47
Physical activity	14	1035	952	−0.20 (−0.3696 to −0.0388)	.01	27.28	52.4	—
Sleep	10	2969	3019	−0.33 (−0.4024 to −0.2486)	<.001	9.68	7.0	—
Multi	10	630	664	−0.31 (−0.4536 to −0.1672)	<.001	11.41	21.1	—
Frequency								
Daily	8	631	695	−0.26 (−0.4352 to −0.0776)	.01	11.44	38.8	.65
Weekly	19	1236	1204	−0.35 (−0.4289 to −0.2175)	<.001	25.36	29.0	—
Other[b]	8	2822	2795	−0.25 (−0.4396 to −0.0653)	.02	16.81	58.3	—
Duration								
≤6 wk	6	377	442	−0.34 (−0.4925 to −0.1771)	.003	3.76	0	.31
7-16 wk	24	4143	4111	−0.27 (−0.3705 to −0.1700)	<.001	47.57	51.7	—
≥4 mo	5	169	141	−0.41 (−0.6219 to −0.1924)	.006	1.78	0	—
Individualized								
No	11	726	742	−0.28 (−0.5126 to −0.0560)	.02	26.41	62.1	.99
Yes	24	3963	3952	−0.28 (−0.3544 to −0.2140)	<.001	29.01	20.7	—
Theoretical underpinning								
No	21	3594	3657	−0.29 (−0.3499 to −0.2390)	<.001	23.45	14.7	.73
Yes	14	1095	1037	−0.26 (−0.4315 to −0.1010)	.004	29.12	55.4	—
Co-designed								

XSL•FO
RenderX

	Number of studies, k	Intervention, n	Control, n	Meta-analysis		Heterogeneity		Subgroup test, *P* value
				Standardized mean difference (95% CI)	*P* value	*Q*	I^2 (%)	
No	31	4383	4354	−0.29 (−0.3689 to −0.2100)	<.001	48.41	38.0	.93
Yes	4	306	340	−0.31 (−0.8339 to 0.2240)	0.16	6.91	56.6	—
Publicly available								
No	19	1350	1313	−0.25 (−0.3946 to −0.1114)	.001	39.57	54.5	0.51
Yes	16	3339	3381	−0.31 (−0.3519 to −0.2486)	<.001	12.86	0	—
Delivery methods								
Wearable								
No	31	4235	4234	−0.29 (−0.3450 to −0.2374)	<.001	42.14	28.8	.36
Yes	4	454	460	−0.18 (−0.5553 to 0.1952)	0.22	7.18	58.2	—
App								
No	24	4043	4047	−0.27 (−0.3562 to −0.1905)	<.001	36.93	37.7	.52
Yes	11	646	647	−0.33 (−0.5017 to −0.1572)	.002	18.19	45.0	—
Web based								
No	12	643	647	−0.33 (−0.5043 to −0.1623)	.001	20.29	45.8	.49
Yes	23	4046	4047	−0.27 (−0.3548 to −0.1922)	<.001	34.91	37.0	—
Text or email								
No	23	1498	1484	−0.33 (−0.4464 to −0.2169)	<.001	36.98	40.5	.17
Yes	12	3191	3210	−0.24 (−0.3309 to −0.1421)	<.001	15.81	30.4	—
Multiple methods used								
No	24	1494	1479	−0.32 (−0.4375 to −0.2033)	<.001	39.94	42.4	.29
Yes	11	3195	3215	−0.25 (−0.3375 to −0.1562)	<.001	13.68	26.9	—
Delivery features								
Video								
No	25	1810	1840	−0.30 (−0.3823 to −0.2206)	<.001	28.87	16.9	.63
Yes	10	2879	2854	−0.25 (−0.4528 to −0.0569)	.02	25.61	64.9	—
Chatbot								
No	30	3867	3727	−0.25 (−0.3373 to −0.1717)	<.001	44.6	35.0	.10
Yes	5	822	967	−0.37 (−0.5304 to −0.2107)	.003	5.78	30.7	—
Gamification or interaction								

XSL•FO
RenderX

	Number of studies, k	Intervention, n	Control, n	Meta-analysis		Heterogeneity		Subgroup test, *P* value
				Standardized mean difference (95% CI)	*P* value	*Q*	I^2 (%)	
No	23	1610	1639	−0.27 (−0.3771 to −0.1705)	<.001	36.56	39.8	.91
Yes	12	3079	3055	−0.28 (−0.3546 to −0.2070)	<.001	18.64	41.0	—
Notifications								
No	19	3072	3034	−0.23 (−0.3652 to −0.1042)	.001	37.21	51.6	.15
Yes	16	1617	1660	−0.34 (−0.4087 to −0.2658)	<.001	13.33	0	—
Social features								
No	25	3957	3943	−0.29 (−0.4038 to −0.1912)	<.001	48.83	50.8	.70
Yes	10	732	751	−0.27 (−0.3727 to −0.1698)	<.001	6.59	0	—
Methodological characteristics								
Control group								
Waitlist control or none	25	3221	3223	−0.31 (−0.4068 to −0.2061)	<.001	37.88	36.6	.46
Treatment as usual	8	1208	1166	−0.24 (−0.4126 to −0.0741)	.01	16.01	56.3	—
Attention	2	260	305	−0.34 (−0.6574 to −0.0126)	.05	0.09	0	—
Outcome measure								
Primary	7	827	747	−0.32 (−0.4533 to −0.1792)	.001	7.09	15.3	.50
Secondary	25	3783	3888	−0.29 (−0.3877 to −0.1973)	<.001	42.9	44.1	—
Exploratory	3	79	59	−0.05 (−0.9823 to 0.8771)	.83	3.06	34.6	—
Study quality								
Poor	1	164	83	−0.31 (−0.5757 to −0.0448)	.02	0	0	.87
Good	29	2307	2349	−0.27 (−0.3684 to −0.1805)	<.001	50.24	44.3	—
Excellent	5	2218	2262	−0.31 (−0.4406 to −0.1751)	.003	4.79	16.5	—
Publication year								
2014 or earlier	3	150	150	−0.16 (−0.2635 to −0.0525)	.02	0.09	0	.003
2015-2017	7	2443	2485	−0.30 (−0.4578 to −0.1345)	.004	10.55	43.1	—
2018-2020	12	1105	1177	−0.19 (−0.3372 to −0.0500)	.01	19.17	42.6	—
2021 or later	13	991	882	−0.39 (−0.5282 to −0.2492)	<.001	18.21	34.1	—

[a]Not applicable.

[b]Other frequency includes fortnightly, monthly, once-off, and self-paced.

XSL•FO
RenderX

Delivery Method and Features

The effectiveness of interventions for improving depression and anxiety did not significantly differ based on the digital delivery method used or their technological features (eg, video, gamification, notifications, interactive design, and social features). There was a nonsignificant trend for interventions using chatbots to be slightly more effective for reducing depressive symptoms than interventions not using chatbots (*P*=.08).

Intervention Characteristics

Interventions had similar effects on depression and anxiety, regardless of whether this was achieved through physical activity, sleep, diet, or multiple lifestyle behaviors. Diet-only interventions demonstrated nonsignificant effects on depression and anxiety, likely due to the paucity of studies.

Intervention effectiveness did not differ based on the frequency or duration of the intervention; however, long-term interventions (≥4 months) demonstrated a nonsignificant effect on depressive symptoms. Intervention effectiveness did not differ based on the intervention being personalized, based on a theoretical framework, whether they had been co-designed with consumers, or whether they were publicly available.

Methodological Characteristics

The effectiveness of interventions did not differ when comparing waitlist or no treatment controls, treatment-as-usual controls, and attention or sham controls. Similarly, effectiveness did not differ based on study quality.

Considering we included studies that did not specifically target mental health, but captured it as an outcome, we compared interventions with depression or anxiety as a primary outcome to those capturing these as secondary outcomes or exploratory outcomes. Intervention effectiveness did not differ based on whether the outcome was primary or secondary; however, the few studies that captured depression and anxiety as exploratory outcomes demonstrated nonsignificant intervention effects.

Population Characteristics

Digital lifestyle interventions demonstrated comparable effectiveness for improving depression and anxiety across all populations: apparently healthy adults, people with a mental disorder, people with a physical health condition, and people with insomnia. For depressive symptoms, the largest effect was seen in people with a mental disorder, whereas for anxiety symptoms, a nonsignificant effect was observed for those with a mental disorder.

Publication Year

The effectiveness of interventions on depression and anxiety differed significantly based on the year of publication, with more recently published interventions demonstrating the greatest effects.

Mixed-effects metaregression analysis showed no significant effect of intervention duration, study quality, or publication year on the effect of digital lifestyle interventions for depression, anxiety, stress, or well-being outcomes. A detailed overview of the metaregression analyses is provided in Multimedia Appendix 6.

Sensitivity Analyses

Sensitivity analyses for all meta-analyses are detailed in Multimedia Appendix 7. For depression, when small studies were omitted from the analysis, the effect size shrunk but remained significant (SMD=−0.28; 95% CI −0.38 to −0.19; *P*<.001). For anxiety, the effect size was slightly larger when studies with >25% attrition were removed (SMD=−0.31; 95% CI −0.43 to −0.18; *P*<.001). For stress, omitting small studies or studies with >25% attrition no longer demonstrated a significant effect (*P*=.13 and *P*=.22, respectively). The overall effect on well-being remained nonsignificant when fair to poor quality studies, small studies, and studies with >25% attrition were removed (all *P* values >.10).

Level of Evidence

Overall, the level of evidence for digital lifestyle interventions for improving symptoms of depression, anxiety, and stress is grade A: consistent level 1 studies. The grade of recommendation for digital lifestyle interventions for improving symptoms of well-being was grade D: level 5 evidence.

Discussion

Principal Findings

This systematic review aimed to comprehensively examine the effectiveness of digital lifestyle interventions for improving symptoms of depression, anxiety, stress, and well-being in adults. We identified 61 RCTs involving 22,483 participants. The findings suggest that digital lifestyle interventions had small-to-medium favorable effects on depression, anxiety, and stress symptoms across a broad range of populations and via a broad range of digital delivery methods. These effects are similar in magnitude to those observed for established mental health treatments, such as antidepressant medication [101] and psychotherapy [102]. No significant effect was observed for well-being, although in comparison to mental ill health, fewer studies (n=6) of lower average quality measured well-being outcomes. These studies operationalized well-being broadly using either global measures that span various affective and functional dimensions [43,61,84,91] or scales of overall life satisfaction [46,96]. These are broad constructs that tend to be relatively stable [103]. As a result, the available evidence offers limited precision for elucidating the effects of digital lifestyle interventions on specific well-being domains. Two-thirds of the included studies (41/61, 67%) were published within the last 5 years, reflecting the increased interest in using digital technology for health behavior change for mental health, with greater reductions in depressive and anxiety symptoms observed for more recently published interventions. This likely reflects the volume of studies published recently but may also be driven by improved quality of digital interventions as technology, particularly artificial intelligence, is better used within health promotion [104].

To better understand what factors drive the effects of digital lifestyle interventions on depressive and anxiety symptoms, we conducted a range of comparative subgroup analyses on specific

XSL•FO
RenderX

features relating to the intervention, methodology, technology, and population. These analyses found that interventions demonstrated comparable effectiveness for reducing symptoms of depression and anxiety in apparently healthy adults or adults with a mental disorder, a physical health condition, or insomnia. Equivalent effectiveness was found regardless of the lifestyle behaviors targeted, method of digital delivery (eg, app and website), technological features (ie, chatbot and gamification), intervention duration, or intended frequency of engagement. Overall, our subgroup analyses indicated that there are no specific features that drive the effectiveness of digital lifestyle interventions for improving symptoms of depression or anxiety, except for publication year.

Although differences among population groups were not significant, it appears that digital lifestyle interventions are effective for improving depressive symptoms in people with a depressive disorder but are potentially less effective at improving symptoms of anxiety in people with a depressive disorder. Interestingly, depressive disorders were the only diagnosed mental disorder in the included studies, highlighting a clear gap. Therefore, the effectiveness of these interventions as a self-management approach for other mental disorders is unclear and should be explored in future studies.

We also observed that interventions using chatbots trended toward being more effective than those without chatbots, though these results were not statistically significant. These findings have been supported with recent research,. Emerging evidence suggests chatbots can help improve lifestyle behaviors [105], although it appears their effectiveness for delivering psychological interventions (ie, therapy) is less conclusive [106]. Co-designed interventions showed a trend for being less effective. This is likely due to the small number of studies included in the subgroup analysis and the high variation of co-design approaches used in the design of health interventions, which are often poorly described and rarely evaluated [107].

A key challenge of digital health interventions is maintaining user engagement over time [13]. It is possible that the lack of significant intervention effects at 4 months was due to drop off in engagement. Alternatively, it is possible that symptoms may recur in some individuals even if they remain engaged with the intervention, reflecting the recurring nature of mental health symptoms. Digital lifestyle interventions may offer an appropriate short-term strategy to engage people in mental health–promoting behaviors and initiate positive shifts in mental health, whereas more sustainable mental health benefits may require other mental health treatment approaches. These findings reinforce the existing literature that the clinical implications of lifestyle interventions, even when delivered digitally, are compelling as a self-management tool for targeting both the risk and protective factors for mental health and should be given consideration as part of standard mental health care. This review did not directly compare digital to face-to-face delivery of these interventions; therefore, the impact that delivery method has on effectiveness should be examined directly in future research. Given the role of patient-practitioner rapport and other psychosocial benefits of in-person care, it is pivotal to understand how digital interventions may best fit and integrate within the broader health care setting, including alongside gold standard treatments, to optimize patient outcomes.

In line with previous research [108], we found interventions targeting multiple or single behaviors to have comparable effectiveness. Interventions that aim to change multiple health behaviors concurrently seem logical, given the cooccurrence of and bidirectional relationship between unhealthy lifestyle behaviors and poor mental health. In particular, some health behaviors may have synergistic impacts (eg, people who improve their sleep may have more energy for physical activity and, in turn, people who are more physically active may sleep better); therefore, future research should explore how behavior change techniques can be implemented most effectively to address multiple behaviors and aim to identify moderators of effective interventions and whether delivering lifestyle behavior change within or alongside traditional psychological interventions has synergistic effects on outcomes or improves adherence.

This review also focused on adults, although adolescence is a period when mental health issues commonly present and offers an opportune time for preventive intervention. Digital interventions would seem to be appealing and accessible for this population, given their familiarity with technology. A recent review by Raeside et al [109] included 11 interventions and found small nonsignificant effects for digital lifestyle interventions on mental health outcomes in adolescents. Considering the shared protective and risk factors for mental health and chronic diseases, future research efforts should focus on if and how digital tools can be best harnessed to address the interplay between lifestyle behaviors, mental ill health, and well-being and engage young people in the design and development of interventions through participatory research methods, such as co-design.

Strengths and Limitations

This is the largest review to evaluate the effects of digital interventions targeting lifestyle behaviors on depression, anxiety, stress, and well-being in adults. We adhered to rigorous methodological approaches for the conduct and reporting of systematic reviews and meta-analyses and conducted comprehensive database searches. In total, 87% (53/61) of the included RCTs received good or excellent quality ratings, increasing confidence in the findings of this review. Given the very large number of studies included in the meta-analyses, the dataset was sufficient to support numerous subgroup analyses, facilitating closer interrogation of the factors influencing intervention effectiveness. Our review included digital lifestyle interventions that have been evaluated using an RCT design and published in a peer-reviewed academic journal. Many commercial digital products claim to support lifestyle changes and mental health but often lack rigorous scientific assessment. The conclusions from this study should not be generalized to such unevaluated products.

Most of the study's limitations arose from the limitations of the included studies. First, only 2 studies reported on diet-only interventions, and few studies measured stress and well-being, resulting in uncertain and underpowered results for these meta-analyses and precluding subgroup analyses with stress

XSL•FO
RenderX

and well-being as outcomes. Second, all but 2 studies [41,99] were conducted in high-income countries, limiting the generalizability of our results. Third, it is important to acknowledge that our subgroup analyses, such as the comparison between single- and multibehavioral interventions, were conducted using available data from studies not specifically designed to test these differences. Consequently, these findings should be interpreted cautiously and should not be used as definitive evidence for or against any particular intervention approach. Finally, we acknowledge significant heterogeneity as a limitation, and we conducted numerous subgroup and sensitivity analyses to mitigate its impact and assess the robustness of our findings, which remained consistent with the main analyses.

Future Directions

This review highlights some key areas for future research. First, there were limited studies focused on addressing diet as a stand-alone intervention; however, a nonsignificant effect in the direction of lifestyle interventions for reducing depressive symptoms highlights that further research is needed to better determine if, and how, effective these interventions may be in addressing mental health symptoms. Second, digital interventions may be less effective over longer periods, emphasizing a need to address longer-term strategies or the potential to integrate other mental health treatment approaches, which may demonstrate greater effects over time. Research also needs to address how best to integrate these evidence-based interventions into service delivery to support health care capacity. Third, for a more complete understanding of how digital lifestyle interventions influence mental health, future research should ensure that positive aspects of mental well-being are measured alongside symptoms of mental ill health. A growing body of literature suggests that these are distinct continua, with both overlapping and unique antecedents, that should be measured in tandem [110]. Finally, as the field continues to mature and research elucidates optimal intervention designs, other factors that are imperative to consider are the influence of both health literacy and digital literacy as well as the impact of health status and socioeconomic status [111]. Furthermore, future research should consider the digital determinants of health and the direct and indirect impacts on health equity when designing health promotion interventions [112].

Conclusions

Overall, the small-to-medium effects indicate that digital lifestyle interventions may provide an effective short-term self-management strategy for mental health support by encouraging positive shifts in lifestyle behaviors. All digital delivery platforms and features demonstrated comparable effectiveness for depression and anxiety outcomes. The effectiveness on well-being is inconclusive, with few studies capturing this positive dimension of mental health. Digitally delivered lifestyle interventions support self-management and target both risk and protective factors for mental health and should be given consideration as part of mental health care and support. This is especially relevant in situations where access to in-person treatment is limited, such as due to long waitlists, financial or geographical barriers, or stigma associated with traditional treatment options.

Future research should explore how best to implement these findings and integrate such interventions with existing health care services, with concerted efforts to improve health equity by considering the digital determinants of health and impacts of digital and health literacy in the design and implementation of these interventions.

Acknowledgments

This study received no funding. CM received an Investigator Grant from the Medical Research Future Fund (1193862). JF is supported by a UK Research and Innovation Future Leaders Fellowship (MR/T021780/1). JF has received honoraria and consultancy fees from Atheneum, Informa, Gillian Kenny Associates, Bayer, Big Health, Hedonia, Strive Coaching, Wood For Trees, Nutritional Medicine Institute, Angelini, ParachuteBH, Richmond Foundation, and Nirakara, independent of this work, and declares no competing interests.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JB and CM were responsible for conceptualization. JB completed database searching. JB, EO, BS, GM, RC, TF, GG, IW, PJM, and KS completed article screening, data extraction, and critical appraisal. JB was responsible for data curation and analysis. JB, CM, and JF contributed to data interpretation. JB, EO, BS, GM, TF, GG, IW, PJM, and KS wrote the original draft. CM and JF provided advice. All authors contributed to reviewing and editing the final manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

PRISMA checklist.
[PDF File (Adobe PDF File), 66 KB - jmir_v27i1e56975_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 16 KB - jmir_v27i1e56975_app2.docx]

Multimedia Appendix 3
Reasons for exclusion at full-text screening.
[DOCX File , 91 KB - jmir_v27i1e56975_app3.docx]

Multimedia Appendix 4
Physiotherapy Evidence Database Scale risk of bias for included studies.
[DOCX File , 65 KB - jmir_v27i1e56975_app4.docx]

Multimedia Appendix 5
Funnel plots for meta-analyses.
[DOCX File , 142 KB - jmir_v27i1e56975_app5.docx]

Multimedia Appendix 6
Meta-regression analyses for all outcomes.
[DOCX File , 15 KB - jmir_v27i1e56975_app6.docx]

Multimedia Appendix 7
Sensitivity analyses for all meta-analyses.
[DOCX File , 15 KB - jmir_v27i1e56975_app7.docx]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. O'Connor RC, Worthman CM, Abanga M, Athanassopoulou N, Boyce N, Chan LF, et al. Gone too soon: priorities for action to prevent premature mortality associated with mental illness and mental distress. Lancet Psychiatry 2023 Jun;10(6):452-464 [FREE Full text] [doi: 10.1016/s2215-0366(23)00058-5]
3. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
4. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine 2022 Dec;54:101675 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]
5. Hoang D, Kristoffersen I, Li IW. All in the mind? Estimating the effect of mental health on health behaviours. Soc Sci Med 2019 Mar;225:69-84. [doi: 10.1016/j.socscimed.2019.02.017] [Medline: 30818089]
6. Stubbs B, Vancampfort D, Veronese N, Kahl KG, Mitchell AJ, Lin P, et al. Depression and physical health multimorbidity: primary data and country-wide meta-analysis of population data from 190 593 people across 43 low- and middle-income countries. Psychol Med 2017 Apr 04;47(12):2107-2117. [doi: 10.1017/s0033291717000551]
7. Firth J, Marx W, Dash S, Carney R, Teasdale SB, Solmi M, et al. The effects of dietary improvement on symptoms of depression and anxiety: a meta-analysis of randomized controlled trials. Psychosom Med 2019 Apr;81(3):265-280 [FREE Full text] [doi: 10.1097/PSY.0000000000000673] [Medline: 30720698]
8. Singh B, Olds T, Curtis R, Dumuid D, Virgara R, Watson A, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. Br J Sports Med 2023 Sep;57(18):1203-1209. [doi: 10.1136/bjsports-2022-106195] [Medline: 36796860]
9. Alvaro PK, Roberts RM, Harris JK. A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. Sleep 2013 Jul 01;36(7):1059-1068 [FREE Full text] [doi: 10.5665/sleep.2810] [Medline: 23814343]
10. Malhi GS, Bassett D, Boyce P, Bryant R, Fitzgerald PB, Fritz K, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry 2015 Dec;49(12):1087-1206. [doi: 10.1177/0004867415617657] [Medline: 26643054]
11. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]

XSL•FO
RenderX

12. Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159 [FREE Full text] [doi: 10.2196/35159] [Medline: 35551058]
13. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
14. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
15. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
16. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health 2017 Nov;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]
17. Shin JC, Kim J, Grigsby-Toussaint D. Mobile phone interventions for sleep disorders and sleep quality: systematic review. JMIR Mhealth Uhealth 2017 Sep 07;5(9):e131 [FREE Full text] [doi: 10.2196/mhealth.7244] [Medline: 28882808]
18. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MD, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
19. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021 Feb 22;23(2):e23786 [FREE Full text] [doi: 10.2196/23786] [Medline: 33616534]
20. Carneiro L, Rosenbaum S, Ward PB, Clemente FM, Ramirez-Campillo R, Monteiro-Júnior RS, et al. Web-based exercise interventions for patients with depressive and anxiety disorders: a systematic review of randomized controlled trials. Braz J Psychiatry 2022;44(3):331-341 [FREE Full text] [doi: 10.1590/1516-4446-2021-2026] [Medline: 34852034]
21. Sawyer C, McKeon G, Hassan L, Onyweaka H, Martinez Agulleiro L, Guinart D, et al. Digital health behaviour change interventions in severe mental illness: a systematic review. Psychol Med 2023 Sep 28;53(15):6965-7005. [doi: 10.1017/s0033291723002064]
22. Young CL, Trapani K, Dawson S, O'Neil A, Kay-Lambkin F, Berk M, et al. Efficacy of online lifestyle interventions targeting lifestyle behaviour change in depressed populations: a systematic review. Aust N Z J Psychiatry 2018 Sep 27;52(9):834-846. [doi: 10.1177/0004867418788659] [Medline: 30052063]
23. McKeon G, Papadopoulos E, Firth J, Joshi R, Teasdale S, Newby J, et al. Social media interventions targeting exercise and diet behaviours in people with noncommunicable diseases (NCDs): a systematic review. Internet Interv 2022 Mar;27:100497 [FREE Full text] [doi: 10.1016/j.invent.2022.100497] [Medline: 35198409]
24. Keyes CL. Mental illness and/or mental health? Investigating axioms of the complete state model of health. J Consult Clin Psychol 2005 Jun;73(3):539-548 [FREE Full text] [doi: 10.1037/0022-006X.73.3.539] [Medline: 15982151]
25. van Agteren J, Iasiello M, Lo L, Bartholomaeus J, Kopsaftis Z, Carey M, et al. A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. Nat Hum Behav 2021 May 19;5(5):631-652. [doi: 10.1038/s41562-021-01093-w] [Medline: 33875837]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906 [FREE Full text] [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
27. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
28. Dodge R, Daly A, Huyton J, Sanders L. The challenge of defining wellbeing. Int J Wellbeing 2012 Aug 28;2(3):222-235 [FREE Full text] [doi: 10.5502/ijw.v2i3.4]
29. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
30. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005 Apr 20;5:13 [FREE Full text] [doi: 10.1186/1471-2288-5-13] [Medline: 15840177]
31. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
32. Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. BMC Med Res Methodol 2014 Nov 21;14:120 [FREE Full text] [doi: 10.1186/1471-2288-14-120] [Medline: 25416419]
33. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 6.4, 2023. London, UK: Cochrane Collaboration; 2023.

XSL•FO
RenderX

34. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
35. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634. [doi: 10.1136/bmj.315.7109.629]
36. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun 24;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
37. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003 Aug;83(8):713-721. [Medline: 12882612]
38. The Oxford levels of evidence 2. Oxford Centre for Evidence-Based Medicine. 2011. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence [accessed 2025-02-06]
39. Tanniou J, van der Tweel I, Teerenstra S, Roes KC. Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. BMC Med Res Methodol 2016 Feb 18;16(1):20 [FREE Full text] [doi: 10.1186/s12874-016-0122-6] [Medline: 26891992]
40. Abbott RD, Sherwin K, Klopf H, Mattingly HJ, Brogan K. Efficacy of a multimodal online lifestyle intervention for depressive symptoms and quality of life in individuals with a history of major depressive disorder. Cureus 2020 Jul 08;12(7):e9061 [FREE Full text] [doi: 10.7759/cureus.9061] [Medline: 32656047]
41. Abedi P, Nikkhah P, Najar S. Effect of pedometer-based walking on depression, anxiety and insomnia among postmenopausal women. Climacteric 2015 Sep 23;18(6):841-845. [doi: 10.3109/13697137.2015.1065246] [Medline: 26100101]
42. Bade BC, Gan G, Li F, Lu L, Tanoue L, Silvestri GA, et al. "Randomized trial of physical activity on quality of life and lung cancer biomarkers in patients with advanced stage lung cancer: a pilot study". BMC Cancer 2021 Apr 01;21(1):352 [FREE Full text] [doi: 10.1186/s12885-021-08084-0] [Medline: 33794808]
43. Bailey DP, Mugridge LH, Dong F, Zhang X, Chater AM. Randomised controlled feasibility study of the MyHealthAvatar-diabetes smartphone app for reducing prolonged sitting time in type 2 diabetes mellitus. Int J Environ Res Public Health 2020 Jun 19;17(12):4414 [FREE Full text] [doi: 10.3390/ijerph17124414] [Medline: 32575482]
44. Bennion KA, Tate D, Muñoz-Christian K, Phelan S. Impact of an internet-based lifestyle intervention on behavioral and psychosocial factors during postpartum weight loss. Obesity (Silver Spring) 2020 Oct 04;28(10):1860-1867 [FREE Full text] [doi: 10.1002/oby.22921] [Medline: 32888250]
45. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
46. Brindal E, Hendrie GA, Freyne J, Noakes M. A mobile phone app designed to support weight loss maintenance and well-being (MotiMate): randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 04;7(9):e12882 [FREE Full text] [doi: 10.2196/12882] [Medline: 31486407]
47. Carli V, Petros NG, Hadlaczky G, Vitcheva T, Berchialla P, Bianchi S, et al. The NEVERMIND e-health system in the treatment of depressive symptoms among patients with severe somatic conditions: a multicentre, pragmatic randomised controlled trial. EClinicalMedicine 2022 Jun;48:101423 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101423] [Medline: 35706482]
48. Chee W, Kim S, Ji X, Park S, Chee E, Tsai HM, et al. The effect of a culturally tailored web-based physical activity promotion program on Asian American midlife women's depressive symptoms. Asian Pac Isl Nurs J 2016 Dec 28;1(4):162-173. [doi: 10.9741/23736658.1044]
49. Cheng P, Kalmbach DA, Tallent G, Joseph CL, Espie CA, Drake CL. Depression prevention via digital cognitive behavioral therapy for insomnia: a randomized controlled trial. Sleep 2019 Oct 09;42(10):zsz150 [FREE Full text] [doi: 10.1093/sleep/zsz150] [Medline: 31535688]
50. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psychiatry 2016 Apr;3(4):333-341. [doi: 10.1016/S2215-0366(15)00536-2] [Medline: 26827250]
51. Conner TS, Brookie KL, Carr AC, Mainvil LA, Vissers MC. Let them eat fruit! The effect of fruit and vegetable consumption on psychological well-being in young adults: a randomized controlled trial. PLoS One 2017;12(2):e0171206 [FREE Full text] [doi: 10.1371/journal.pone.0171206] [Medline: 28158239]
52. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
53. Duan YP, Wienert J, Hu C, Si GY, Lippke S. Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e106 [FREE Full text] [doi: 10.2196/jmir.7152] [Medline: 28396306]
54. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a web-based intervention for multiple health behavior changes in patients with coronary heart disease in home-based rehabilitation: pilot randomized controlled trial. J Med Internet Res 2018 Nov 19;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]

XSL•FO
RenderX

55. Ebert DD, Berking M, Thiart H, Riper H, Laferton JA, Cuijpers P, et al. Restoring depleted resources: efficacy and mechanisms of change of an internet-based unguided recovery training for better sleep and psychological detachment from work. Health Psychol 2015 Dec;34S:1240-1251. [doi: 10.1037/hea0000277] [Medline: 26651465]
56. Eckert R, Huberty J, Kurka J, Laird B, Mesa R, Palmer J. A randomized pilot study of online hatha yoga for physical and psychological symptoms among survivors of allogenic bone marrow transplant. Int J Yoga Therap 2022 Jan 01;32(2022):Article 12. [doi: 10.17761/2022-D-21-00047] [Medline: 36305845]
57. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
58. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
59. Falk GE, Mailey EL, Okut H, Rosenkranz SK, Rosenkranz RR, Montney JL, et al. Effects of sedentary behavior interventions on mental well-being and work performance while working from home during the COVID-19 pandemic: a pilot randomized controlled trial. Int J Environ Res Public Health 2022 May 24;19(11):6401 [FREE Full text] [doi: 10.3390/ijerph19116401] [Medline: 35681986]
60. Felder JN, Epel ES, Neuhaus J, Krystal AD, Prather AA. Efficacy of digital cognitive behavioral therapy for the treatment of insomnia symptoms among pregnant women: a randomized clinical trial. JAMA Psychiatry 2020 May 01;77(5):484-492 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.4491] [Medline: 31968068]
61. Freeman D, Sheaves B, Goodwin GM, Yu LM, Nickless A, Harrison PJ, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. Lancet Psychiatry 2017 Oct;4(10):749-758 [FREE Full text] [doi: 10.1016/S2215-0366(17)30328-0] [Medline: 28888927]
62. Glozier N, Christensen H, Griffiths KM, Hickie IB, Naismith SL, Biddle D, et al. Adjunctive internet-delivered cognitive behavioural therapy for insomnia in men with depression: a randomised controlled trial. Aust N Z J Psychiatry 2019 Apr 07;53(4):350-360. [doi: 10.1177/0004867418797432] [Medline: 30191722]
63. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106 [FREE Full text] [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
64. Hershner S, O'Brien LM. The impact of a randomized sleep education intervention for college students. J Clin Sleep Med 2018 Mar 15;14(3):337-347 [FREE Full text] [doi: 10.5664/jcsm.6974] [Medline: 29510791]
65. Hilmarsdóttir E, Sigurðardóttir Á, Arnardóttir RH. A digital lifestyle program in outpatient treatment of type 2 diabetes: a randomized controlled study. J Diabetes Sci Technol 2021 Sep 17;15(5):1134-1141 [FREE Full text] [doi: 10.1177/1932296820942286] [Medline: 32680441]
66. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
67. Houchen-Wolloff L, Gardiner N, Devi R, Robertson N, Jolly K, Marshall T, et al. Web-based cardiac REhabilitatioN alternative for those declining or dropping out of conventional rehabilitation: results of the WREN feasibility randomised controlled trial. Open Heart 2018;5(2):e000860 [FREE Full text] [doi: 10.1136/openhrt-2018-000860] [Medline: 30364405]
68. Huberty J, Eckert R, Dueck A, Kosiorek H, Larkey L, Gowin K, et al. Online yoga in myeloproliferative neoplasm patients: results of a randomized pilot trial to inform future research. BMC Complement Altern Med 2019 Jun 07;19(1):121 [FREE Full text] [doi: 10.1186/s12906-019-2530-8] [Medline: 31174535]
69. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e17816 [FREE Full text] [doi: 10.2196/17816] [Medline: 32319963]
70. Joutsenniemi K, Kaattari C, Härkänen T, Pankakoski M, Langinvainio H, Lönnqvist J, et al. E-mail-based exercises in happiness, physical activity and readings: a randomized trial on 3274 Finns. Afr J Psychiatry 2014;17(5):1-13 [FREE Full text] [doi: 10.4172/psychiatry.1000140]
71. Katz P, Margaretten M, Gregorich S, Trupin L. Physical activity to reduce fatigue in rheumatoid arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018 Jan;70(1):1-10. [doi: 10.1002/acr.23230] [Medline: 28378441]
72. Kuhn E, Miller KE, Puran D, Wielgosz J, YorkWilliams SL, Owen JE, et al. A pilot randomized controlled trial of the insomnia coach mobile app to assess its feasibility, acceptability, and potential efficacy. Behav Ther 2022 May;53(3):440-457. [doi: 10.1016/j.beth.2021.11.003] [Medline: 35473648]
73. Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-based intervention using behavioral activation and physical activity for adults with depression (The eMotion Study): pilot randomized controlled trial. J Med Internet Res 2018 Jul 16;20(7):e10112 [FREE Full text] [doi: 10.2196/10112] [Medline: 30012547]
74. Lambert SD, Duncan LR, Culos-Reed SN, Hallward L, Higano CS, Loban E, et al. Feasibility, acceptability, and clinical significance of a dyadic, web-based, psychosocial and physical activity self-management program (TEMPO) tailored to

XSL•FO
RenderX

the needs of men with prostate cancer and their caregivers: a multi-center randomized pilot trial. Curr Oncol 2022 Feb 01;29(2):785-804 [FREE Full text] [doi: 10.3390/curroncol29020067] [Medline: 35200566]
75. Lopez R, Evangelista E, Barateau L, Chenini S, Bosco A, Billiard M, et al. French language online cognitive behavioral therapy for insomnia disorder: a randomized controlled trial. Front Neurol 2019;10:1273 [FREE Full text] [doi: 10.3389/fneur.2019.01273] [Medline: 31866929]
76. Lorenz N, Heim E, Roetger A, Birrer E, Maercker A. Randomized controlled trial to test the efficacy of an unguided online intervention with automated feedback for the treatment of insomnia. Behav Cogn Psychother 2019 May;47(3):287-302 [FREE Full text] [doi: 10.1017/S1352465818000486] [Medline: 30185239]
77. McGrath ER, Espie CA, Power A, Murphy AW, Newell J, Kelly C, et al. Sleep to lower elevated blood pressure: a randomized controlled trial (SLEPT). Am J Hypertens 2017 Mar 01;30(3):319-327. [doi: 10.1093/ajh/hpw132] [Medline: 28391289]
78. Mensorio MS, Cebolla-Martí A, Rodilla E, Palomar G, Lisón JF, Botella C, et al. Analysis of the efficacy of an internet-based self-administered intervention ("Living Better") to promote healthy habits in a population with obesity and hypertension: an exploratory randomized controlled trial. Int J Med Inform 2019 Apr;124:13-23. [doi: 10.1016/j.ijmedinf.2018.12.007] [Medline: 30784422]
79. Mueller J, Richards R, Jones RA, Whittle F, Woolston J, Stubbings M, et al. Supporting weight management during COVID-19: a randomized controlled trial of a web-based, ACT-based, guided self-help intervention. Obes Facts 2022;15(4):550-559 [FREE Full text] [doi: 10.1159/000524031] [Medline: 35417910]
80. Murawski B, Plotnikoff RC, Rayward AT, Oldmeadow C, Vandelanotte C, Brown WJ, et al. Efficacy of an m-health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
81. Nyström MB, Stenling A, Sjöström E, Neely G, Lindner P, Hassmén P, et al. Behavioral activation versus physical activity via the internet: a randomized controlled trial. J Affect Disord 2017 Jun;215:85-93 [FREE Full text] [doi: 10.1016/j.jad.2017.03.018] [Medline: 28319696]
82. Philippot A, Moulin P, Charon MH, Balestra C, Dubois V, de Timary P, et al. Feasibility of online high-intensity interval training (HIIT) on psychological symptoms in students in lockdown during the COVID-19 pandemic: a randomized controlled trial. Front Psychiatry 2022;13:904283 [FREE Full text] [doi: 10.3389/fpsyt.2022.904283] [Medline: 35800026]
83. Przybylko G, Morton D, Kent L, Morton J, Hinze J, Beamish P, et al. The effectiveness of an online interdisciplinary intervention for mental health promotion: a randomized controlled trial. BMC Psychol 2021 May 11;9(1):77 [FREE Full text] [doi: 10.1186/s40359-021-00577-8] [Medline: 33975645]
84. Puig-Ribera A, Bort-Roig J, Giné-Garriga M, González-Suárez AM, Martínez-Lemos I, Fortuño J, et al. Impact of a workplace 'sit less, move more' program on efficiency-related outcomes of office employees. BMC Public Health 2017 May 16;17(1):455 [FREE Full text] [doi: 10.1186/s12889-017-4367-8] [Medline: 28511642]
85. Puterman E, Hives B, Mazara N, Grishin N, Webster J, Hutton S, et al. COVID-19 Pandemic and Exercise (COPE) trial: a multigroup pragmatic randomised controlled trial examining effects of app-based at-home exercise programs on depressive symptoms. Br J Sports Med 2022 May;56(10):546-552 [FREE Full text] [doi: 10.1136/bjsports-2021-104379] [Medline: 34580067]
86. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an internet intervention to improve the sleep of cancer survivors with insomnia. Psychooncology 2012 Jul;21(7):695-705 [FREE Full text] [doi: 10.1002/pon.1969] [Medline: 21538678]
87. Savard J, Ivers H, Savard MH, Morin CM. Is a video-based cognitive behavioral therapy for insomnia as efficacious as a professionally administered treatment in breast cancer? Results of a randomized controlled trial. Sleep 2014 Aug 01;37(8):1305-1314 [FREE Full text] [doi: 10.5665/sleep.3918] [Medline: 25083010]
88. Serrat M, Albajes K, Navarrete J, Almirall M, Lluch Girbés E, Neblett R, et al. Effectiveness of two video-based multicomponent treatments for fibromyalgia: the added value of cognitive restructuring and mindfulness in a three-arm randomised controlled trial. Behav Res Ther 2022 Nov;158:104188 [FREE Full text] [doi: 10.1016/j.brat.2022.104188] [Medline: 36116229]
89. Shaffer KM, Smith JG, Glazer JV, Camacho F, Chow PI, Mattos M, et al. Effects of an internet-delivered insomnia intervention for older adults: a secondary analysis on symptoms of depression and anxiety. J Behav Med 2022 Oct;45(5):728-738 [FREE Full text] [doi: 10.1007/s10865-022-00353-y] [Medline: 35932397]
90. Spanhel K, Hovestadt E, Lehr D, Spiegelhalder K, Baumeister H, Bengel J, et al. Engaging refugees with a culturally adapted digital intervention to improve sleep: a randomized controlled pilot trial. Front Psychiatry 2022 Feb 23;13:832196 [FREE Full text] [doi: 10.3389/fpsyt.2022.832196] [Medline: 35280163]
91. Stiglbauer B, Weber S, Batinic B. Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial. Comput Hum Behav 2019 May;94:131-139. [doi: 10.1016/j.chb.2019.01.018]
92. Teychenne M, Abbott G, Stephens LD, Opie RS, Olander EK, Brennan L, et al. Mums on the move: a pilot randomised controlled trial of a home-based physical activity intervention for mothers at risk of postnatal depression. Midwifery 2021 Feb;93:102898. [doi: 10.1016/j.midw.2020.102898] [Medline: 33290891]

XSL•FO
RenderX

93. Vandelanotte C, Duncan MJ, Plotnikoff RC, Rebar A, Alley S, Schoeppe S, et al. Impact of a web-based personally tailored physical activity intervention on depression, anxiety, stress and quality of life: secondary outcomes from a randomized controlled trial. Ment Health Phys Act 2022 Oct;23:100477. [doi: 10.1016/j.mhpa.2022.100477]
94. Wan ES, Kantorowski A, Homsy D, Teylan M, Kadri R, Richardson CR, et al. Promoting physical activity in COPD: insights from a randomized trial of a web-based intervention and pedometer use. Respir Med 2017 Sep;130:102-110 [FREE Full text] [doi: 10.1016/j.rmed.2017.07.057] [Medline: 29206627]
95. Wang L, Liu Y, Tan H, Huang S. Transtheoretical model-based mobile health application for PCOS. Reprod Health 2022 May 12;19(1):117 [FREE Full text] [doi: 10.1186/s12978-022-01422-w] [Medline: 35549736]
96. Wang F, Boros S. Effects of a pedometer-based walking intervention on young adults' sleep quality, stress and life satisfaction: randomized controlled trial. J Bodyw Mov Ther 2020 Oct;24(4):286-292 [FREE Full text] [doi: 10.1016/j.jbmt.2020.07.011] [Medline: 33218524]
97. Wong VW, Ho FY, Shi NK, Tong JT, Chung KF, Yeung WF, et al. Smartphone-delivered multicomponent lifestyle medicine intervention for depressive symptoms: a randomized controlled trial. J Consult Clin Psychol 2021 Dec;89(12):970-984. [doi: 10.1037/ccp0000695] [Medline: 35025538]
98. Young MD, Drew RJ, Kay-Lambkin F, Collins CE, Callister R, Kelly BJ, et al. Impact of a self-guided, eHealth program targeting weight loss and depression in men: a randomized trial. J Consult Clin Psychol 2021 Aug;89(8):682-694. [doi: 10.1037/ccp0000671] [Medline: 34472895]
99. Xuto P, Toyohiko K, Prasitwattanaseree P, Sriarporn P. Effect of receiving text messages on health care behavior and state anxiety of Thai pregnant women: a randomized controlled trial. Int J Community Based Nurs Midwifery 2022 Jan;10(1):18-29 [FREE Full text] [doi: 10.30476/IJCBNM.2021.89364.1604] [Medline: 35005038]
100. Yudi MB, Clark DJ, Tsang D, Jelinek M, Kalten K, Joshi SB, et al. SMARTphone-based, early cardiac REHABilitation in patients with acute coronary syndromes: a randomized controlled trial. Coron Artery Dis 2021 Aug 01;32(5):432-440. [doi: 10.1097/MCA.0000000000000938] [Medline: 32868661]
101. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018 Apr 07;391(10128):1357-1366 [FREE Full text] [doi: 10.1016/S0140-6736(17)32802-7] [Medline: 29477251]
102. Linde K, Sigterman K, Kriston L, Rücker G, Jamil S, Meissner K, et al. Effectiveness of psychological treatments for depressive disorders in primary care: systematic review and meta-analysis. Ann Fam Med 2015 Jan 12;13(1):56-68 [FREE Full text] [doi: 10.1370/afm.1719] [Medline: 25583894]
103. Hudson NW, Anusic I, Lucas RE, Donnellan MB. Comparing the reliability and validity of global self-report measures of subjective well-being with experiential day reconstruction measures. Assessment 2020 Jan;27(1):102-116 [FREE Full text] [doi: 10.1177/1073191117744660] [Medline: 29254354]
104. Ayenigbara IO. The evolving nature of artificial intelligence: role in public health and health promotion. J Public Health (Oxf) 2024 May 29;46(2):e322-e323. [doi: 10.1093/pubmed/fdad240] [Medline: 37973395]
105. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023 Jun 23;6(1):118 [FREE Full text] [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]
106. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
107. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
108. Gómez-Gómez I, Bellón J, Resurrección DM, Cuijpers P, Moreno-Peral P, Rigabert A, et al. Effectiveness of universal multiple-risk lifestyle interventions in reducing depressive symptoms: systematic review and meta-analysis. Prev Med 2020 May;134:106067 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106067] [Medline: 32194097]
109. Raeside R, Jia S, Todd A, Hyun K, Singleton A, Gardner L, et al. Are digital health interventions that target lifestyle risk behaviors effective for improving mental health and wellbeing in adolescents? A systematic review with meta-analyses. Adolescent Res Rev 2023 Aug 12;9(2):193-226. [doi: 10.1007/s40894-023-00224-w]
110. Iasiello M, van Agteren J, Ali K, Fassnacht DB. J Posit Psychol 2023 Feb 26;19(2):337-341. [doi: 10.1080/17439760.2023.2179935]
111. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st century Cures Act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
112. Kickbusch I, Holly L. Addressing the digital determinants of health: health promotion must lead the charge. Health Promot Int 2023 Jun 01;38(3):daad059. [doi: 10.1093/heapro/daad059] [Medline: 37264549]

Abbreviations

PEDro: Physiotherapy Evidence Database Scale

XSL•FO
RenderX

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by A Coristine; submitted 31.01.24; peer-reviewed by RT Villarino, M Pikkarainen, M Schneider; comments to author 03.05.24; revised version received 27.05.24; accepted 27.09.24; published 20.03.25.

Please cite as:
Brinsley J, O'Connor EJ, Singh B, McKeon G, Curtis R, Ferguson T, Gosse G, Willems I, Marent PJ, Szeto K, Firth J, Maher C Effectiveness of Digital Lifestyle Interventions on Depression, Anxiety, Stress, and Well-Being: Systematic Review and Meta-Analysis J Med Internet Res 2025;27:e56975
URL: https://www.jmir.org/2025/1/e56975
doi:10.2196/56975
PMID:

©Jacinta Brinsley, Edward J O'Connor, Ben Singh, Grace McKeon, Rachel Curtis, Ty Ferguson, Georgia Gosse, Iris Willems, Pieter-Jan Marent, Kimberley Szeto, Joseph Firth, Carol Maher. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

AI-Derived Blood Biomarkers for Ovarian Cancer Diagnosis: Systematic Review and Meta-Analysis

He-Li Xu[1,2,3*], MPH; Xiao-Ying Li[1,2,3*], PhD; Ming-Qian Jia[1,2,3,4*], MPH; Qi-Peng Ma[5*], MSc; Ying-Hua Zhang[6], BM; Fang-Hua Liu[1,2,3], MPH; Ying Qin[1,2,3,4], MPH; Yu-Han Chen[4,5], MPH; Yu Li[4,5], MPH; Xi-Yang Chen[1,2,3,4], MPH; Yi-Lin Xu[1,2,3], MPH; Dong-Run Li[1,2,3], MPH; Dong-Dong Wang[1,2,3,4], MPH; Dong-Hui Huang[1,2,3], PhD; Qian Xiao[1,5], PhD; Yu-Hong Zhao[1,2,3], Prof Dr Med; Song Gao[5], Prof Dr Med; Xue Qin[5], PhD; Tao Tao[5], PhD; Ting-Ting Gong[5], Prof Dr Med; Qi-Jun Wu[1,2,3,4,5,7], Prof Dr Med

[1]Department of Clinical Epidemiology, Shengjing Hospital of China Medical University, ShenYang, China

[2]Clinical Research Center, Shengjing Hospital of China Medical University, ShenYang, China

[3]Liaoning Key Laboratory of Precision Medical Research on Major Chronic Disease, Shengjing Hospital of China Medical University, ShenYang, China

[4]Department of Epidemiology, School of Public Health, China Medical University, ShenYang, China

[5]Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, ShenYang, China

[6]Department of Undergraduate, Shengjing Hospital of China Medical University, ShenYang, China

[7]NHC Key Laboratory of Advanced Reproductive Medicine and Fertility (China Medical University), National Health Commission, ShenYang, China

[*]these authors contributed equally

Corresponding Author:
Qi-Jun Wu, Prof Dr Med
Department of Clinical Epidemiology
Shengjing Hospital of China Medical University
No. 36, San Hao Street
ShenYang, 110004
China
Phone: 86 024 96615 13652
Email: wuqj@sj-hospital.org

Abstract

Background: Emerging evidence underscores the potential application of artificial intelligence (AI) in discovering noninvasive blood biomarkers. However, the diagnostic value of AI-derived blood biomarkers for ovarian cancer (OC) remains inconsistent.

Objective: We aimed to evaluate the research quality and the validity of AI-based blood biomarkers in OC diagnosis.

Methods: A systematic search was performed in the MEDLINE, Embase, IEEE Xplore, PubMed, Web of Science, and the Cochrane Library databases. Studies examining the diagnostic accuracy of AI in discovering OC blood biomarkers were identified. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies–AI tool. Pooled sensitivity, specificity, and area under the curve (AUC) were estimated using a bivariate model for the diagnostic meta-analysis.

Results: A total of 40 studies were ultimately included. Most (n=31, 78%) included studies were evaluated as low risk of bias. Overall, the pooled sensitivity, specificity, and AUC were 85% (95% CI 83%-87%), 91% (95% CI 90%-92%), and 0.95 (95% CI 0.92-0.96), respectively. For contingency tables with the highest accuracy, the pooled sensitivity, specificity, and AUC were 95% (95% CI 90%-97%), 97% (95% CI 95%-98%), and 0.99 (95% CI 0.98-1.00), respectively. Stratification by AI algorithms revealed higher sensitivity and specificity in studies using machine learning (sensitivity=85% and specificity=92%) compared to those using deep learning (sensitivity=77% and specificity=85%). In addition, studies using serum reported substantially higher sensitivity (94%) and specificity (96%) than those using plasma (sensitivity=83% and specificity=91%). Stratification by external validation demonstrated significantly higher specificity in studies with external validation (specificity=94%) compared to those without external validation (specificity=89%), while the reverse was observed for sensitivity (74% vs 90%). No publication bias was detected in this meta-analysis.

Conclusions: AI algorithms demonstrate satisfactory performance in the diagnosis of OC using blood biomarkers and are anticipated to become an effective diagnostic modality in the future, potentially avoiding unnecessary surgeries. Future research

XSL•FO
RenderX

is warranted to incorporate external validation into AI diagnostic models, as well as to prioritize the adoption of deep learning methodologies.

Trial Registration: PROSPERO CRD42023481232; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023481232

(J Med Internet Res 2025;27:e67922) doi:10.2196/67922

KEYWORDS

artificial intelligence; AI; blood biomarker; ovarian cancer; diagnosis; PRISMA

Introduction

Ovarian Cancer Diagnosis: Status and Demands

Ovarian cancer (OC) is the deadliest gynecologic malignancy, characterized by nonspecific symptoms that often remain undetected until the disease has progressed [1,2]. The conventional diagnosis of OC principally depends on imaging techniques (encompassing ultrasound, computed tomography, and magnetic resonance imaging); serum biomarkers (such as cancer antigen 125, carcinoembryonic antigen, and human epididymis protein 4); along with the invasive procedure (histological biopsy) [3,4]. However, the sensitivity and specificity of imaging techniques and biomarkers are restricted [5]. Furthermore, the histopathological test is inherently invasive [3]. Therefore, there is an urgent need for more accurate, noninvasive, and reliable diagnostic methods.

Potential of Noninvasive Blood Markers for OC

Minimally invasive diagnostic procedures, particularly the use of blood samples, are among the foremost common methods of detection [6]. Moreover, patients are generally more willing to undergo blood tests, leading to higher compliance rates [7]. Blood contains a rich repertoire of biomolecules, including proteins, nucleic acids, and metabolites, which can potentially serve as indicators of OC diagnosis [8-11]. The development of omics has opened new doors for biomarker discovery. The genomics, proteomics, and metabolomics of blood samples can provide a wealth of information about the molecular changes in cancer [12]. For instance, Ke et al [13] systematically investigated OC metabolism through the metabolic profiling of 448 plasma samples. The analysis of dysregulated metabolic pathways extends our current understanding of OC metabolism. Similarly, Dhar et al [8] applied glycoproteomics to serum of women with OC or benign pelvic masses and healthy controls and analyzed glycosylation patterns in serum markers and supported the hypothesis that blood glycoproteomic profiling can be used for OC diagnosis and staging. Notably, the exponential growth of multiomics data has presented a major challenge that surpasses traditional analytic capabilities [14,15]. Fortunately, artificial intelligence (AI) algorithms offer a solution [16,17]. AI can manage complex datasets and spot hidden patterns and potential biomarkers, enabling more accurate OC diagnosis and personalized treatment.

Application of AI in OC Blood Markers

AI, particularly machine learning (ML) and deep learning (DL), has attracted increasing attention in medical research due to its capability to analyze large biomedical datasets [18-23]. AI-driven models have emerged as a promising tool for developing predictive models for OC by analyzing complex and multidimensional datasets to uncover biomarkers. For instance, a multicenter retrospective study screened 52 features from laboratory tests in blood samples to build an ML model. Integrating 20 base AI models, it performed well internally and externally, outperforming the CA125 and HE4 biomarkers in identifying OC [24]. In addition, a blood-based metabolite panel demonstrated independent predictive ability and complemented the risk of ovarian malignancy algorithm for distinguishing early-stage OC from benign disease to better inform clinical decision-making [25]. Despite such encouraging results, the related results are scattered. Whether the application of AI in OC blood biomarker research can significantly improve diagnostic accuracy remains controversial.

Purpose of This Study (AI and OC Blood Markers)

Therefore, there is a crucial need for a high-quality synthesis of the available evidence. The purpose of this study is to provide a systematic overview of the diagnostic accuracy of AI techniques in identifying OC blood markers as well as to elucidate their applicability, potential, and limitations.

Methods

Protocol Registration and Study Design

This meta-analysis was conducted in adherence to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist (Multimedia Appendix 1) and MOOSE (Meta-Analysis of Observational Studies in Epidemiology) reporting guidelines [26,27]. The protocol was prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42023481232).

Literature Search and Eligibility Criteria

A comprehensive search of the MEDLINE, Embase, IEEE Xplore, PubMed, Web of Science, and the Cochrane Library databases was carried out from inception to January 16, 2024. Detailed search strategies are summarized in Multimedia Appendix 2. Two independent investigators (MQJ and HLX) assessed the records after removing the duplicates at the title and abstract level, and finally at the full-text level, according to the inclusion and exclusion criteria (Textbox 1). Two investigators (MQJ and XYL) independently appraised the articles for eligibility. An inconsistency in selection was reconciled through discussion with a third independent investigator (HLX).

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Population: adults (aged ≥18 years) with potential ovarian cancer lesions
- Intervention: artificial intelligence–assisted blood test
- Comparison: histopathology or other reliable clinical diagnosis
- Outcomes: diagnostic performance (ie, sensitivity and specificity or detailed information that could extract or construct 2×2 contingency tables)
- Studies: original articles (ie, observational studies or randomized controlled trials)
- Language: English

Exclusion criteria

- Population: nonhuman samples or other diseases
- Intervention: nonblood samples or no artificial intelligence algorithms
- Comparison: no control group
- Outcomes: no diagnostic performance data (ie, 2×2 contingency tables cannot be extracted or constructed from the provided data)
- Studies: records, such as letters, conference abstracts, case reports, or review articles
- Language: non-English

Data Extraction

Data were independently extracted by 2 investigators (QPM and XYL) using a predefined data extraction sheet, with any discrepancies resolved through the adjudication of a third investigator (HLX). Necessary data of 2×2 contingency tables that included true positives (TP), false positives, true negatives (TN), and false negatives were extracted. The details of the data are presented in Multimedia Appendix 3 [8,24,25,28-64]. In cases where these values were not explicitly reported, values were calculated using descriptive statistics available in the study. For studies presenting multiple contingency tables, either for identical or disparate AI algorithms, each table was treated as an independent result [65].

Study Quality Assessment

The risk of bias and concerns about the applicability of all included studies were assessed by 2 independent investigators (MQJ and HLX), using the Quality Assessment of Diagnostic Accuracy Studies–Artificial Intelligence (QUADAS-AI) criteria [66]. Conflicts were discussed and solved with a third investigator (XYL). The risk of bias assessment included 4 domains: patient selection, index test, reference standard, and flow and timing. For assessing clinical applicability, only the first 3 domains were evaluated [66,67]. In addition, the median was used as the threshold for determining the risk level of bias, with studies classified as low risk if ≥2 domains were deemed as low risk and high risk if <2 domains were considered low risk [68].

Data Analysis

We used the bivariate diagnostic random effects model to compute the summary receiver operating characteristics to determine summary estimates of the sensitivity, specificity, and area under the curve (AUC) with their respective 95% CIs [69]. Sensitivity was defined as the probability of a person with OC having a positive test result, indicating the capacity of the index test to identify patients, considered by the equation: sensitivity = TP / (TP + false negatives). Specificity was the probability of a woman without OC having a negative test result, reflecting the test's ability to correctly identify OC-free individuals calculated by the equation: specificity = TN / (false positives + TN) [70,71]. The performance of the test can also be assessed using the AUC. This area may be interpreted as the probability that a random woman with OC has a higher value of the measurement than a random person without OC. In general, an AUC of >0.80 is considered good [72]. A perfect test would have an AUC of 1 and a useless test would have an AUC of 0.5 [73].

Study heterogeneity was determined using the I^2 statistic and the Q test [74]. When the same or different AI models were tested within the same article, the proposed model with the best accuracy was used for further meta-analysis [65]. Subgroup and regression analyses were performed to explore potential sources of heterogeneity. Sensitivity analysis was implemented via a qualitative systematic review of studies from which concatenated tables could not be extracted or constructed. Publication bias was assessed using funnel plot asymmetry test by Deeks et al [75].

Subgroup analyses were performed according to the following: (1) AI algorithms (ML or DL), (2) external validation (yes or no), (3) levels of risk of bias (low or high), (4) year of publication (after or before 2022), (5) geographical distribution (Asia, North America, or Europe), (6) sample size (>300 or ≤300 as median), (7) blood sample type (serum or plasma), (8) biomarkers type (protein or mixed), and (9) number of modeling biomarkers (>8 or ≤8 as median).

The variability in sensitivity and specificity estimates was graphically represented through a cross-hairs plot, generated using R software (version 4.2.1; R Foundation for Statistical Computing) [76]. All other statistical analyses were conducted

XSL•FO
RenderX

in Stata (version 17.0; StataCorp). The statistical significance was defined as $P<.05$.

Results

Study Selection and Characteristics of Eligible Studies

The database search identified 1566 records from which 604 duplicates were removed. We then performed the title and abstract screening of 962 records and subsequently a full-text evaluation of 55 records. Following the exclusion of 15 articles, as detailed in Multimedia Appendix 4 [77-91], a total of 40 studies were included in this meta-analysis (Figure 1).

Most of the studies were performed with retrospectively (21/40, 52%) and prospectively (18/40, 45%) collected data, and one study collected both retrospective and prospective data (Table 1). In total, 12% (5/40) of studies sourced their data from public databases. In terms of AI algorithm types, a total of 36 (90%) studies were classified as ML, whereas 10% (4/40) of studies were classified as DL. Most (36/40, 90%) of the studies validated their algorithms, while only some (7/40, 18%) studies carried out an external validation (Table 2). The blood samples were mainly serum (27/40, 68%), and the type of blood biomarkers was mainly protein (25/40, 62%; Table 3).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for study selection and processing in the meta-analysis.

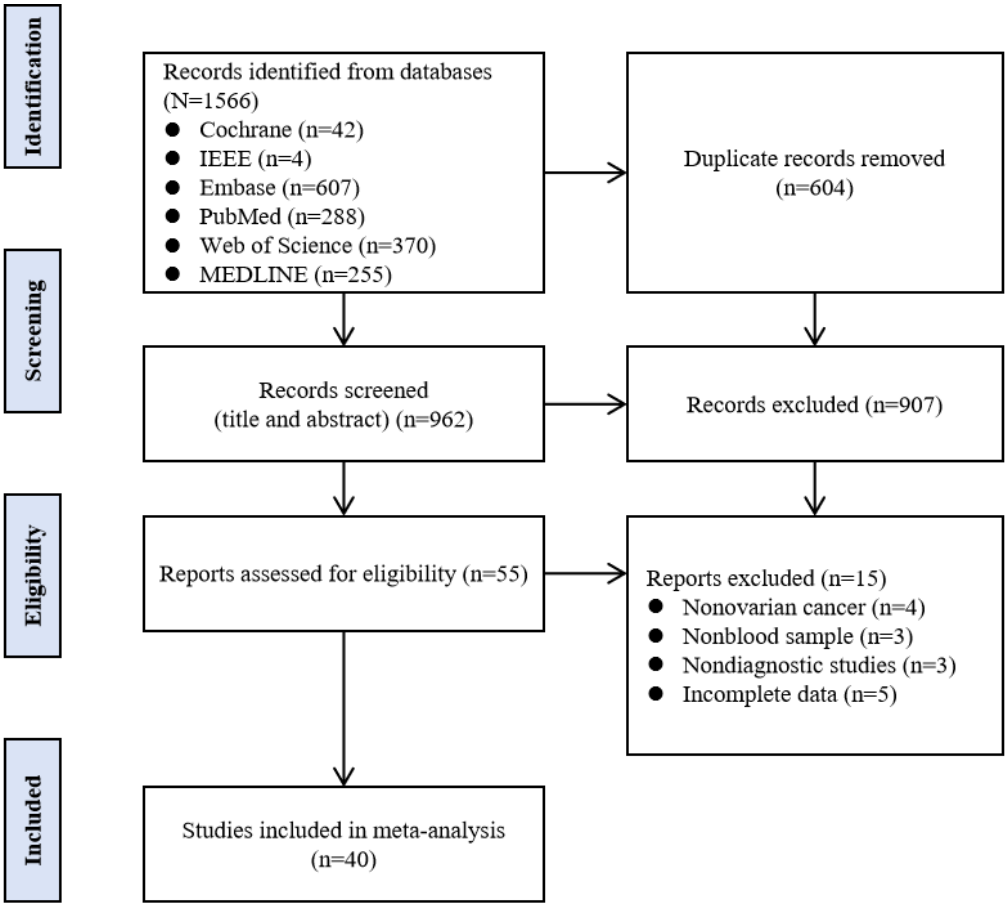

XSL•FO
RenderX

Table 1. Baseline characteristics (study design, data source, selection criteria, time frame, age, and sample size) of 40 included studies on artificial intelligence–based ovarian cancer diagnosis with blood samples.

Study	Study design	Data source	Selection criteria	Time frame	Age (y), mean or median	Sample size, n
Cai et al [24], 2024	Retrospective	Data from Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, central China; Women's Hospital, School of Medicine, Zhejiang University, eastern China; and Sun Yat-Sen University Cancer Center, southern China	Patients with a history of other malignant cancers or precancers, pregnancy in the last 6 mo, or affected by HIV; not newly diagnosed in any of the 3 hospitals; individuals without any available laboratory tests were excluded	From January 2012 to April 2021	Cancer: (53/51/56)[a]; control: (34/34/48)[a]	10,992 (3007/5641/2344)[a]
Abuzinadah et al [28], 2023	Retrospective	Data from the Third Affiliated Hospital of Soochow University	All patients underwent postoperative case diagnosis, and none of them had received preoperative radiotherapy or chemotherapy	From July 2011 to July 2018	NR[b]	349 (244/105)[c]
Bifarin and Fernandez [29], 2024	Retrospective	Data from a serum lipidomic analysis of ovarian cancer patients of Korean descent	NR	NR	NR	325 (227/98)[c]
Cameron et al [30], 2023	Retrospective	Data from the Welcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, the Emergency Medicine Research Group at the Edinburgh Royal Infirmary, the Beatson West of Scotland Cancer Centre in Glasgow, the University of Swansea, Royal Preston Hospital, and Manchester Cancer Research Centre	NR	NR	Ovary: 61; NCS[d] female individuals only: 56	385 (NR)
Chen et al [31], 2023	Retrospective	Data from the Department of Gynecology of Harbin Medical University Cancer Hospital, Gene Expression Omnibus database, UCSC Xena	Patients with a primary radiological diagnosis of ovarian tumor; newly diagnosed patients without any significant comorbidities or history of previous malignancies; willingness to participate in the study and provision of written informed consent	From December 2020 to July 2021	NR	44 (44)[e]
Dhar et al [8], 2023	Retrospective	Data from Indivumed (Hamburg, Germany)	NR	NR	NR	351 (237/114)[c]
Hamidi et al [32], 2023	Retrospective	Data from the Gene Expression Omnibus database, a Japanese nationwide research project, and patients with cancer who were referred or admitted to the National Cancer Center Hospital	The serum samples for noncancer controls who had no history of cancer and no hospitalization during the previous 3 months were collected; patients with cancer who were treated with preoperative chemotherapy and radiotherapy before serum collection were excluded	NR	Internal set: 52	3411 (2156/3079/92/240)[f]

XSL•FO
RenderX

Study	Study design	Data source	Selection criteria	Time frame	Age (y), mean or median	Sample size, n
Lai et al [33], 2023	Retrospective	Data from clinical laboratory examination	NR	From January 2013 to October 2022	46.4	778 (545/233)[g]
Li et al [34], 2024	Retrospective	Data from the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital)	No patients with ovarian cancer received chemotherapy, radiotherapy, or surgery, no healthy donors had a history of cancer before sample collection	NR	Stage 1: 53.6; stage 2: 55.8; stage 3 and 4: 54.9; control: 34.1	69 (NR)
Reilly et al [35], 2023	Retrospective and prospective	Data from multiple studies spanning multiple centers	Patient age ≥18 years; informed consent provided by the patient to participate in research; patient agreeable to phlebotomy; patient had a documented adnexal mass	NR	47.5	2186 (NR)
Zhang et al [36], 2023	Retrospective	Data from patients who underwent physical examination at Chinese People's Armed Police Force and First Medicine Center, People's Liberation Army General Hospital	Discharge diagnosis confirmed by clinical signs, imaging, and pathology for patients with gynecologic tumors and benign gynecologic diseases; if no histopathological examination was available, it was consistently confirmed by ≥2 types of imaging evidence, availability of laboratory test data at the time of first diagnosis, and blood collection before treatment	From January 2010 to June 2019	Ovarian cancer: (52.11/52.69/55.81)[h]; NOMGT[i]: (54.66/52.66/53.95)[h]; BGD[j]: (44.30/46.33/39.21)[h]; Healthy control: (49.96/47.26/49.67)[h]	1633(600/301)[g]
Ahamad et al [37], 2022	Retrospective	Data from Third Affiliated Hospital of Schow University	NR	From July 2017 to July 2018	NR	349 (85/21)[c]
Bahado-Singh et al [38], 2022	Prospective	Data from Oakland University William Beaumont School of Medicine	NR	NR	Cases: 66.2; control: 67.8	17 (NR)
Gupta et al [39], 2022	Retrospective	Data from 3 separate commercial biobanks: Dx Biosamples (San Diego, CA), Reprocell USA Inc (Beltsville, MD), and Fidelis Research AD (Sofa, Bulgaria)	NR	NR	NR	1243 (681)[k]
Hinestrosa et al [40], 2022	Retrospective	Data from a commercial biorepository (ProteoGenex, Inglewood, CA, United States)	The control group has no known history of cancer, autoimmune diseases, or neurodegenerative disorders, nor any presence of diabetes mellitus	From January 2014 to September 2020	60	323 (216/107)[c]
Irajizad et al [25], 2022	Prospective	Data from Anderson Cancer Center and at the Fred Hutchinson Cancer Research Center	NR	NR	NR	409 (108/118)[c]
Kim et al [41], 2022	Retrospective	Data from Oakland University William Beaumont School of Medicine	NR	NR	NR	269 (215)[e]

XSL•FO
RenderX

Study	Study design	Data source	Selection criteria	Time frame	Age (y), mean or median	Sample size, n
Li et al [42], 2022	Prospective	Data from 3 institutions	Women diagnosed with benign, borderline, and malignant ovarian tumors	From December 2018 to January 2020	NR	362 (178/184)[c]
Pais et al [43], 2022	Retrospective	Data from a commercially stored biological sample biobank (Invent diagnostica, Berlin, Germany)	NR	NR	NR	181 (NR)
Jeong et al [44], 2021	Retrospective	Data from Kangnam Sacred Heart Hospital	NR	From June 2014 to December 2020	Cancer: 54; control: 49	730 (511/219)[c]
Lu et al [45], 2020	Prospective	Data from the Third Affiliated Hospital of Soochow University	None of the patients with ovarian cancer received preoperative chemotherapy or radiotherapy	From July 2011 to July 2018	NR	349(235/114)[c]
Banaei et al [46], 2019	Prospective	Data from the UMass Memorial Medical Center Chemotherapy Infusion Center and Gastroenterology Clinics and Innovative Research	NR	NR	NR	20 (40/160)[c]
Whitwell et al [47], 2018	Retrospective	Data from a synthetic dataset modeled from the United Kingdom Collaborative Trial of Ovarian Cancer Screening	Trial participants at enrollment were postmenopausal women aged 50-74 y who had no family history of ovarian cancer	NR	NR	89 (NR)
Ivanova et al [48], 2016	Retrospective	Data from the clinical diagnostic laboratory of the LLC LYTECH, the Blokhin Cancer Research Center of Russian Academy of Medical Sciences, the National Research Center of Coloproctology, the Moscow Dermatovenerologic Dispensary, clinical hospitals of Peoples' Friendship University of Russia	NR	NR	Cancer: 52; control: 49	67 (NR)
Jiang et al [49], 2013	Prospective	Data from the affiliated hospital, Sun Yat-Sen University	NR	NR	61.7	87 (51/36)[l]
Yang et al [50], 2013	Prospective	Data from the Peking University Third Hospital	NR	From January 2003 to December 2009	Stage I/II: 54.8; stage III: 57.3; stage IV: 58.2; normal: 52.8; carcinoid: 51.6	246 (NR)
Shan et al [51], 2012	Prospective	Data from the Tampa, Florida metropolitan area	Women with a prior unilateral or bilateral oophorectomy were ineligible, as were women with a previous history of cancer. All patients underwent preoperative radiologic imaging, either by pelvic ultrasound, CT, or MRI. Only patients who underwent surgery based on clinical suspicion of ovarian cancer were eligible	NR	NR	423 (NR)

XSL•FO
RenderX

Study	Study design	Data source	Selection criteria	Time frame	Age (y), mean or median	Sample size, n
Thakur et al [52], 2011	Prospective	Data from the FDA-NCI clinical proteomics program databank	NR	NR	NR	216 (173/43)[c]
Donach et al [53], 2010	Prospective	Data from Padua Hospital (now the Veneto Oncology Institute)	NR	From 1999 to 2005	48	201 (NR)
Ziganshin et al [54], 2008	Prospective	Data from Byelorussian Oncology Center with patients with ovarian cancer and the Clinical Diagnostic Laboratory with clinically healthy women	NR	NR	Cancer: 51;control: 49	118 (NR)
Liu et al [55], 2007	Prospective	Data from Northwestern University, Johns Hopkins University in Baltimore, MD, and the University of Innsbruck, Austria	Normal samples were from patients who had 4-year follow-up examinations to ensure that they did not have cancer at the time the samples were taken	From 1999 to 2002	NR	563 (315/78/170)[m]
Zhang et al [56], 2007	Prospective	Data from the Duke University Medical Center, Durham, NC, St Bartholomew's Hospital, London, United Kingdom, and the Groningen University Hospital, Groningen, Netherlands	NR	NR	NR	468 (200/150)[c]
Chatterjee et al [57], 2006	Retrospective	Data from the Barbara Ann Karmanos Cancer Institute, the MD Anderson Cancer Center, Weill Medical College of Cornell University, Northwestern University Robert H Lurie Comprehensive Cancer Center, and the Gynecologic Oncology Group Tissue Bank	NR	NR	NR	129 (85/44)[c]
Lin et al [58], 2006	Prospective	Data from the Tri-Service General Hospital, Taiwan, and Republic of China	Patients with any history of cancer, operations that had removed body organ, or current chronic or acute diseases were excluded	NR	NR	65 (65)[e]
Liu [59], 2006	Prospective	Data from Clinical Proteomic Program Databank	NR	NR	NR	253 (NR)
Wu et al [60], 2006	Prospective	Data from the Tri-Service General Hospital, Taiwan	No history of gynecologic tumors and had a normal pelvic examination and pelvic sonography	NR	NR	65 (NR)
Li and Ramamohanrao [61],2004	Prospective	Data from a public website	NR	From November 2003	NR	253 (215/112)[c]
Li et al [62], 2004	Retrospective	Data from a public website	NR	From February 2002	NR	469 (100/116)[c]
Zhang et al [63], 1999	Retrospective	Data from an existing data set of clinically diagnosed with pelvic masses and University of Texas MD Anderson Cancer Center	NR	NR	NR	625 (174/255)[c]

XSL•FO
RenderX

Study	Study design	Data source	Selection criteria	Time frame	Age (y), mean or median	Sample size, n
Wilding et al [64], 1994	Prospective	Data from the Hospital of the University of Pennsylvania	Patients with carcinoma in situ were excluded	NR	NR	98 (NR)

[a]Training/external validation 1/external validation 2.

[b]NR: not reported.

[c]Training/testing.

[d]NCS: noncancer symptomatic.

[e]Training.

[f]Training/internal validation/external validation 1/external validation 2.

[g]Training/internal validation.

[h]Training/internal validation/external validation.

[i]NOMGT: nonovarian malignant gynecologic tumor.

[j]BGD: benign gynecologic disease.

[k]Testing.

[l]Training/prediction.

[m]Training/testing/internal validation.

XSL•FO
RenderX

Table 2. Artificial intelligence algorithm features (reference standard, algorithm type, type of internal validation, and external validation) of 40 included studies on artificial intelligence–based ovarian cancer diagnosis with blood samples.

Study	Reference standard	Algorithms	ML[a] or DL[b]	Type of internal validation	External validation
Cai et al [24], 2024	Histopathology	MCF[c], XGB[d], LGBM[e], Cat-Boost, GBM[f], RF[g], NB[h], LR[i]	ML	5-fold cross-validation	Yes
Abuzinadah et al [28], 2023	Histopathology	RF, KNN[j], SGD[k], ETC[l], XGB, GBM	ML	K-fold cross-validation	No
Bifarin and Fernandez [29], 2024	Histopathology	AutoML[m], RF, SVM[n], KNN	ML	5-fold cross-validation	No
Cameron et al [30], 2023	Histopathology	NR	ML	A nested cross-validation	No
Chen et al [31], 2023	Histopathology	CBS[o], GISTIC[p]	ML	3-fold cross-validation	Yes
Dhar et al [8], 2023	Histopathology	LR	ML	10-fold cross-validation	No
Hamidi et al [32], 2023	NR	LR, DT[q], RF, ANN[r], XGB	ML	5-fold cross-validation	Yes
Lai et al [33], 2023	Histopathology	SVM	ML	A validation (unclear)	No
Li et al [34], 2024	Histopathology	LDA[s], RF, NN[t], SVM	ML	A validation (unclear)	No
Reilly et al [35], 2023	Histopathology	MIA3G[u]	DL	NR[v]	No
Zhang et al [36], 2023	Histopathology	LR, FLD[w], SVM, RF, ANN	ML	Cross-validation	Yes
Ahamad et al [37], 2022	Histopathology	RF, SVM, DT, XGBM, LR, GBM, LGBM	ML	5-fold cross-validation	No
Bahado-Singh et al [38], 2022	NR	RF, SVM, LDA, PAM[x], GLM[y], DL	ML	10-fold cross-validation	No
Gupta et al [39], 2022	NR	OVR[z]	ML	NR	No
Hinestrosa et al [40], 2022	Histopathology	RFE[aa]	ML	5-fold cross-validation	No
Irajizad et al [25], 2022	NR	DL, RF, EL[ab], GBM	DL	5-fold cross-validation	Yes
Kim et al [41], 2022	Histopathology	DT, LR, ANN, RF, SVM	ML	10-fold cross-validation	Yes
Li et al [42], 2022	Histopathology	SVM	ML	5-fold cross-validation	No
Pais et al [43], 2022	Histopathology	EvA-3[ac], OSC[ad]	ML	A validation (unclear)	No
Jeong et al [44], 2021	NR	ROMA[ae]	ML	3-fold cross-validation	No
Lu et al [45], 2020	Histopathology	ROMA, DT, LR	ML	10-fold cross-validation	No
Banaei et al [46], 2019	NR	CT[af], KNN	ML	5-fold cross-validation	No
Whitwell et al [47], 2018	Histopathology	Parenclitic networks, LR, RDLG[ag]	ML	Monte Carlo cross-validation	No
Ivanova et al [48], 2016	Histopathology	GA[ah], SNN[ai]	ML	Leave one out cross-validations	No
Jiang et al [49], 2013	NR	ANN, CT, Split-point score analysis	ML	One cross-validation	No
Yang et al [50], 2013	Histopathology	ANN	ML	Blind test validation	No
Shan et al [51], 2012	Histopathology	HH-SVM[aj]	ML	5-fold cross-validation	No
Thakur et al [52], 2011	NR	ANNs, LDA	ML	Cross-validation	No
Donach et al [53], 2010	NR	ANN	ML	NR	No
Ziganshin et al [54], 2008	NR	GA, SNN	ML	Cross-validation	No
Liu et al [55], 2007	NR	PLS[ak], SVM, DT C5.0	ML	10-fold cross-validation	No
Zhang et al [56], 2007	Histopathology	ANN	ML	Cross-validation	Yes
Chatterjee et al [57], 2006	NR	Feed-forward NN	DL	NR	No

XSL•FO
RenderX

Study	Reference standard	Algorithms	ML[a] or DL[b]	Type of internal validation	External validation
Lin et al [58], 2006	NR	DT	ML	Cross-validation	No
Liu [59], 2006	NR	SVM	ML	10-fold cross-validation	No
Wu et al [60], 2006	Histopathology	CT	ML	Cross-validation	No
Li and Ramamohanrao [61], 2004	NR	SVM, NB, KNN, DT, CS4[al]	ML	10-fold cross-validation	No
Li et al [62], 2004	NR	SVM	ML	One out cross-validation	No
Zhang et al [63], 1999	Histopathology	ANN	ML	Cross-validation	No
Wilding et al [64], 1994	Histopathology	Backpropagation NN	DL	A validation (unclear)	No

[a]ML: machine learning.
[b]DL: deep learning.
[c]MCF: multi-criteria decision making-based classification fusion.
[d]XGB: extreme gradient boosting.
[e]LGBM: light gradient boosting machine.
[f]GBM: gradient boosting machine.
[g]RF: random forest.
[h]NB: naive Bayes.
[i]LR: logistic regression.
[j]KNN: k-nearest neighbor.
[k]SGD: stochastic gradient descent.
[l]ETC: extra-trees classifier.
[m]AutoML: automated machine learning.
[n]SVM: support vector machine.
[o]CBS: circular binary segmentation algorithm.
[p]GISTIC: the genomic identification of significant targets in cancer 2.0 algorithm.
[q]DT: decision tree.
[r]ANN: artificial neural network.
[s]LDA: Linear Discriminant Analysis.
[t]NN: neural network.
[u]MIA3G: multivariate index assay 3G.
[v]NR: not reported.
[w]FLD: Fisher linear discriminant.
[x]PAM: prediction analysis for microarrays.
[y]GLM: generalized linear model.
[z]OVR: *One Versus Rest* classifier multiclass classification model.
[aa]RFE: Recursive Feature Elimination.
[ab]EL: ensemble learning.
[ac]Eva-3: evolutionary algorithm 3
[ad]OSC: compose the classification algorithm.
[ae]ROMA: risk of ovarian malignancy algorithm.
[af]CT: classification tree.
[ag]RDLG: raw data logistic regression.
[ah]GA: genetic algorithm.
[ai]SNN: supervised neural network.
[aj]HH-SVM: hybrid huberized support vector machine.
[ak]PLS: partial least-square regression.
[al]CS4: cascading-and-sharing for ensembles of decision trees.

XSL•FO
RenderX

Table 3. Biomarker characteristics (blood sample type, detection method, biomarker type, number of modeling, and detection biomarkers) of 40 included studies on artificial intelligence–based ovarian cancer diagnosis with blood samples.

Study	Blood sample type	Device or method	Biomarker type	Number of modeling biomarkers	Number of detection marker
Cai et al [24], 2024	Blood	NR[a]	Protein, mixed	52	NR
Abuzinadah et al [28], 2023	Blood	General chemical tests, blood routine tests	Mixed	49	49
Bifarin and Fernandez [29], 2024	Serum	NR	Mixed	17	17
Cameron et al [30], 2023	Serum	Spectrum	Mixed	5	5
Chen et al [31], 2023	Plasma	LC-WGS[b]	DNA	NR	NR
Dhar et al [8], 2023	Serum	LCMS[c]	Protein	27	571
Hamidi et al [32], 2023	Serum	miRNA labeling kit and miRNA Oligo Chip	RNA	10	2568
Lai et al [33], 2023	Serum	MS[d]	Protein	7	95
Li et al [34], 2024	Plasma	Nanoflow cytometry, SEC[e], CA125 ELISA[f] kit, and HE4 ELISA kit	Protein	7	7
Reilly et al [35], 2023	Serum	Roche cobas 6000 clinical analyzer	Protein	7	7
Zhang et al [36], 2023	Blood	NR	Mixed	25	25
Ahamad et al [37], 2022	Blood and serum	NR	Mixed	47	47
Bahado-Singh et al [38], 2022	Plasma	Illumina Infnium MethylationEPIC BeadChip arrays or methylation analysis	DNA	25	179,238
Gupta et al [39], 2022	Serum	UHPLC-MS[g]	Mixed	25	6336
Hinestrosa et al [40], 2022	Plasma	ACE[h]	Protein	34	42
Irajizad et al [25], 2022	Plasma	LCMS analysis	Mixed	7	475
Kim et al [41], 2022	Serum	Immunoassay, C8000 analyzer, diazo reagent	Mixed	NR	NR
Li et al [42], 2022	Plasma	QIAamp Circulating Nucleic Acid kit	DNA	5	1272
Pais et al [43], 2022	Serum	MALDINR-TOF MS[i]	Protein	CHCA[j]:26-57; SA[k]:12-113	CHCA:8500; SA:8500
Jeong et al [44], 2021	Serum	2-step chemiluminescent microparticle immunoassay	Protein	16	3
Lu et al [45], 2020	Serum	Roche Cobas 8000 modular analyzer series	Protein	2	2
Banaei et al [46], 2019	Serum	A microfluidic SERS[l]-based immunoassay method	Protein	5	5
Whitwell et al [47], 2018	Serum	Olink's multiplex immunoassay Oncology II panel	Protein	92	92
Ivanova et al [48], 2016	Serum	MALDI-TOF MS	Protein	7	200-400
Jiang et al [49], 2013	Serum	ELISA	Protein	5	174
Yang et al [50], 2013	Serum	SELDI-TOF MS[m]	Protein	184	184
Shan et al [51], 2012	Serum	Liquid chromatography electrospray tandem mass spectrometry	Mixed	18	18

XSL•FO
RenderX

Study	Blood sample type	Device or method	Biomarker type	Number of modeling biomarkers	Number of detection marker
Thakur et al [52], 2011	Serum	SELDI-TOF MS	Protein	NR	NR
Donach et al [53], 2010	Serum	Radioimmunoassay kits	Protein	4	6
Ziganshin et al [54], 2008	Serum	MALDI-TOF MS or Ultraflex TOF mass spectrometer	Protein	MB-IMAC Cu[n]:13; MB-WCX:11	MB-HIC8[o]:135;MB-HIC18[p]:137; MB-IMAC Cu:115; MB-WCX[q]:96
Liu et al [55], 2007	Serum	prOTOF MS	Protein	NR	96
Zhang et al [56], 2007	Serum	Radioimmunoassay kits	Protein	4	4
Chatterjee et al [57], 2006	Serum	ELISA	Protein	65	65
Lin et al [58], 2006	Plasma	SELDI analysis,WCX2 chip analysis,SAX2 chip analysis	Protein	3	4
Liu [59], 2006	Serum	MS	Protein	NR	NR
Wu et al [60], 2006	Plasma	SELDI-TOF MS	Protein	5	NR
Li and Ramamohanrao [61], 2004	Serum	MS	Protein	72	15,154
Li et al [62], 2004	Serum	SELDI-TOF MS	Protein	10	15,155
Zhang et al [63], 1999	Serum	Radioimmunoassay kits, a spectrophotometric method with a kit	Mixed	4	4
Wilding et al [64], 1994	Serum and plasma	Radioimmunoassay	Mixed	8	8

[a]NR: not reported.

[b]LC-WGS: low-coverage whole genome sequencing.

[c]LC-MS: liquid chromatography-mass spectrometry.

[d]MS: mass spectrometry.

[e]SEC: size exclusion chromatography.

[f]ELISA: enzyme-linked immunosorbent assay.

[g]UHPLC-MS: ultra-high performance liquid chromatography-mass spectrometry

[h]ACE: alternating current electrokinetics.

[i]MALDINR-TOF MS: matrix-assisted laser desorption/Ionization neutral reflector time-of-flight mass spectrometry.

[j]CHCA: α-Cyano-4-hydroxycinnamic acid matrix.

[k]SA: sinapinic acid matrix.

[l]SERS: surface-enhanced Raman spectroscopy.

[m]SELDI -TOF MS: surface-enhanced laser desorption and ionization mass spectrometry.

[n]MB-IMAC Cu: magnetic beads MB-IMAC Cu.

[o]MB-HIC8: magnetic beads MB-HIC 8.

[p]MB-HIC18: magnetic beads MB-HIC 18.

[q]MB-WCX: magnetic beads MB-WCX.

Quality Assessment

The quality of the included studies was appraised using the QUADAS-AI (Figures S1 and S2 in Multimedia Appendix 5). In detail, most of the studies were rated as having a high or unclear risk of bias based on patient selection (22/40, 55%) and index test (33/40, 82%) domains. These assessments might be attributed to the absence of explicit delineation of included patients, such as previous testing history and clinical setting, as well as to deficiencies in rigorous external validation of the AI models.

Pooled Performance of AI Algorithms

The summary receiver operating characteristics curves for the 40 included studies with 342 contingency tables are shown in Figure 2. The pooled sensitivity and specificity were 85% (95% CI 83%-87%) and 91% (95% CI 90%-92%), respectively, with an AUC of 0.95 (95 % CI 0.92-0.96) for all AI algorithms. Notably, when contingency tables with the highest accuracy were extracted from each study, the pooled sensitivity and specificity were 95% (95% CI 90%-97%) and 97% (95% CI 95%-98%), respectively, with an AUC of 0.99 (95% CI 0.98-1.00). Reported point estimates and CIs of all included studies are shown in a cross-hairs plot (Figure 3).

XSL•FO
RenderX

Figure 2. Summary receiver operating characteristic (SROC) curves of all studies included in the meta-analysis (n=40). (A) SROC curves of all studies included in the meta-analysis (40 studies with 342 tables). (B) SROC curves of studies when selecting contingency tables reporting the highest accuracy (40 studies with 40 tables). AUC: area under the curve; SENS: summary sensitivity; SPEC: summary specificity.

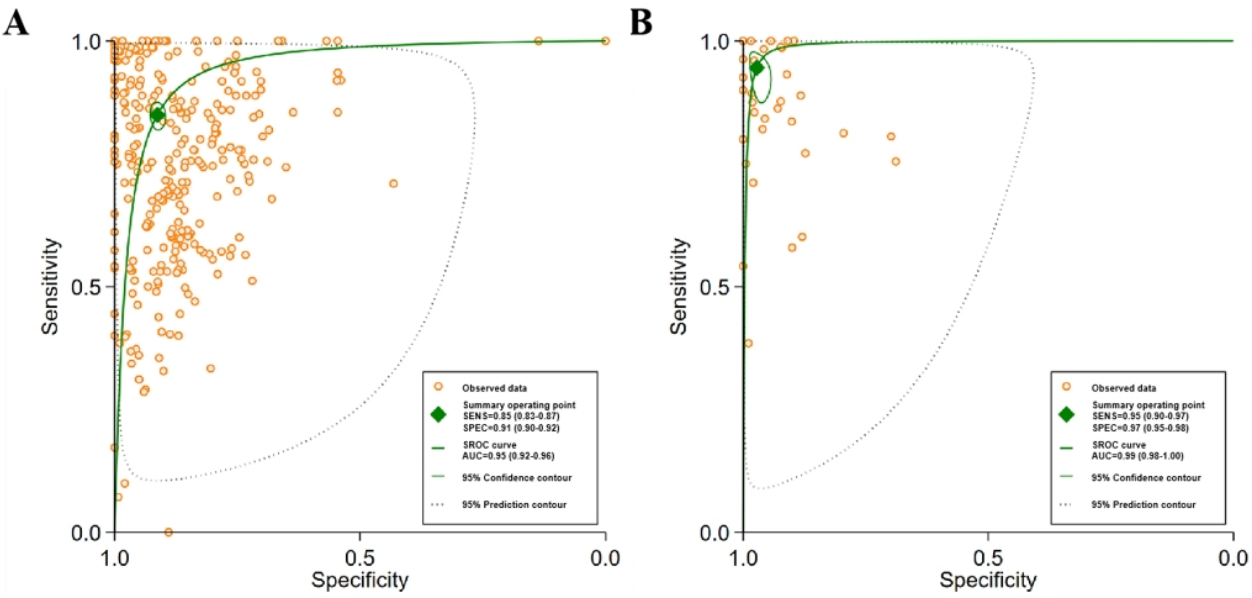

Figure 3. Cross-hair plot of all studies included in the meta-analysis (n=40).

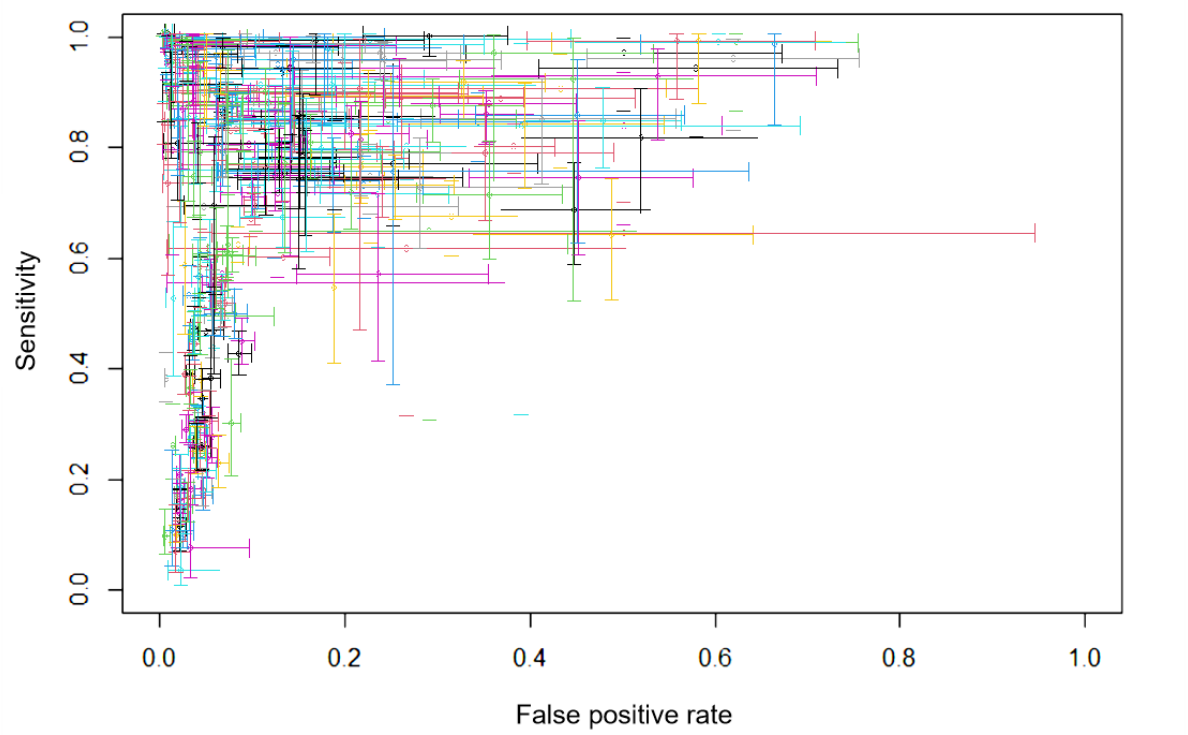

Subgroup Analyses and Meta-Regression Analysis

Results of the subgroup analyses revealed that acceptable diagnostic performance was observed in all subgroups, ranging from 74% to 98% for sensitivity and 85% to 96% for specificity. Detailed results are shown in Table 4, and the corresponding plots are presented in Figures S3-S21 in Multimedia Appendix 5. We divided the studies into subgroups according to the modalities of algorithms (Figures S3 and S13 in Multimedia Appendix 5), existence of external validation (Figures S4 and S14 in Multimedia Appendix 5), levels of risk of bias (Figures S5 and S15 in Multimedia Appendix 5), year of publication (Figures S6 and S16 in Multimedia Appendix 5), geographical distribution (Figures S7 and S17 in Multimedia Appendix 5), sample size (Figures S8 and S18 in Multimedia Appendix 5), blood sample type (Figures S9 and S19 in Multimedia Appendix 5), biomarkers type (Figures S10 and S20 in Multimedia Appendix 5), and number of modeling biomarkers (Figures S11 and S21 in Multimedia Appendix 5).

XSL•FO
RenderX

Table 4. Summary estimate of pooled performance of artificial intelligence–derived blood biomarkers for the diagnosis of ovarian cancer.

	Number of studies, n (%)	Sensitivity			*P* value[a]	Specificity			*P* value[a]
		Sensitivity (95% CI)	I^2 (95% CI)	*P* value[b]		Specificity (95% CI)	I^2 (95% CI)	*P* value[b]	
Overall	40 (100)	0.85 (0.83-0.87)	97.54 (97.41-97.66)	<.001	—[c]	0.91 (0.90-0.92)	99.15 (99.12-99.18)	<.001	—
Artificial intelligence algorithms[d]					.07				.10
Machine learning	36 (90)	0.86 (0.83-0.88)	97.79 (97.67-97.90)	<.001	—	0.92 (0.90-0.93)	99.27 (99.24-99.30)	<.001	—
Deep learning	4 (10)	0.77 (0.70-0.82)	85.81 (81.19-90.42)	<.001	—	0.85 (0.83-0.87)	68.55 (55.84-81.26)	<.001	—
External validation					<.001				<.001
Yes	7 (18)	0.74 (0.69-0.79)	98.35 (98.23-98.46)	<.001	—	0.94 (0.93-0.95)	99.54 (99.52-99.57)	<.001	—
No	33 (82)	0.90 (0.88-0.92)	94.91 (94.48-95.34)	<.001	—	0.89 (0.86-0.91)	94.17 (93.65-94.68)	<.001	—
Levels of risk of bias[e]					<.001				.28
Low	31 (78)	0.81 (0.79-0.84)	97.40 (97.25-97.55)	<.001	—	0.91 (0.89-0.92)	99.07 (99.03-99.11)	<.001	—
High	9 (22)	0.98 (0.96-0.99)	95.51 (94.74-96.27)	<.001	—	0.94 (0.89-0.96)	96.85 (96.37-97.33)	<.001	—
Year of publication					<.001				<.001
After 2022	19 (48)	0.81 (0.78-0.83)	97.44 (97.28-97.59)	<.001	—	0.90 (0.88-0.91)	99.04 (98.99-99.08)	<.001	—
Before 2022	21 (52)	0.95 (0.91-0.97)	96.83 (96.47-97.19)	<.001	—	0.96 (0.93-0.97)	97.64 (97.39-97.88)	<.001	—
Geographical distribution					0.01				.92
Asia	18 (45)	0.82 (0.78-0.84)	97.78 (97.65-97.91)	<.001	—	0.91 (0.89-0.92)	99.23 (99.20-99.26)	<.001	—
North America	15 (38)	0.92 (0.88-0.95)	94.47 (93.70-95.24)	<.001	—	0.91 (0.87-0.93)	95.88 (95.35-96.40)	<.001	—
Europe	7 (18)	0.91 (0.81-0.96)	96.54 (95.69-97.38)	<.001	—	0.96 (0.90-0.98)	97.02 (96.33-97.71)	<.001	—
Sample size					<.001				.06
>300	21 (52)	0.83 (0.80-0.85)	97.78 (97.65-97.91)	<.001	—	0.91 (0.89-0.92)	99.24 (99.21-99.27)	<.001	—
≤300	19 (48)	0.92 (0.88-0.95)	93.93 (93.07-94.79)	<.001	—	0.94 (0.91-0.96)	92.88 (91.82-93.94)	<.001	—
Blood sample type[f]					<.02				.12
Serum	27 (68)	0.94 (0.92-0.96)	98.13 (97.98-98.29)	<.001	—	0.96 (0.95-0.98)	98.57 (98.46-98.68)	<.001	—
Plasma	8 (20)	0.83 (0.78-0.87)	83.32 (79.03-87.61)	<.001	—	0.91 (0.88-0.94)	70.15 (61.13-79.17)	<.001	—
Marker type[g]					.12				<.001
Protein	25 (62)	0.87 (0.82-0.91)	98.65 (98.56-98.75)	<.001	—	0.95 (0.93-0.96)	99.65 (99.63-99.66)	<.001	—
Mixed	12 (30)	0.79 (0.77-0.82)	94.74 (94.27-95.21)	<.001	—	0.86 (0.84-0.88)	98.56 (98.47-98.64)	<.001	—
Number of modeling marker[h]					.54				.31

XSL•FO
RenderX

	Number of studies, n (%)	Sensitivity			*P* value[a]	Specificity			*P* value[a]
		Sensitivity (95% CI)	I^2 (95% CI)	*P* value[b]		Specificity (95% CI)	I^2 (95% CI)	*P* value[b]	
>8	15 (38)	0.82 (0.79-0.85)	97.73 (97.59-97.87)	<.001	—	0.90 (0.88-0.92)	99.20 (99.16-99.23)	<.001	—
≤8	14 (35)	0.79 (0.74-0.83)	88.11 (85.74-90.49)	<.001	—	0.90 (0.88-0.92)	79.83 (75.13-84.52)	<.001	—

[a]*P* value for heterogeneity between subgroups with meta-regression analysis.

[b]*P* value for heterogeneity within each subgroup.

[c]Not applicable.

[d]Artificial intelligence algorithms include machine learning and deep learning.

[e]Low: ≥2 domain low risk; high: <2 domain low risk.

[f]Five articles that used incomplete information on blood sample type were excluded from this subgroup analysis.

[g]Three DNA studies that used DNA and 1 study that used RNA were excluded for this subgroup analyses.

[h]In total, 11 articles used incomplete information on the number of model markers were excluded for this subgroup analysis.

The meta-analysis uncovered substantial heterogeneity among studies, as evidenced by an I^2 of 97.54% (*P*<.05) for sensitivity and 99.15% (*P*<.05) for specificity (Figure S12 in Multimedia Appendix 5). To further explore the causes of study heterogeneity, a meta-regression analysis was conducted (Table 4). The results showed that both external validation and year of publication were significant factors that influenced study heterogeneity with regard to sensitivity and specificity. Subgroups based on levels of risk of bias, geographical distribution, sample size, and blood sample type showed intergroup heterogeneity in the sensitivity of prediction (*P*<.05). In terms of marker type, specificity presented significant heterogeneity between groups (*P*<.05).

Sensitivity Analyses and Publication Bias

A qualitative systematic review was performed for studies lacking directly or indirectly available contingency tables. The findings from this review were in alignment with the main analysis (Tables S1-S3 in Multimedia Appendix 6 [77-81]). In addition, the analysis did not reveal any publication bias in this meta-analysis (*P*=.72; Multimedia Appendix 7).

Discussion

Principal Findings

The burgeoning evolution of AI within the medical field has captured the attention of an increasing cadre of researchers, particularly in its applicability to disease diagnosis [92,93]. To the best of our knowledge, this meta-analysis is a pioneering effort specifically exploring the efficacy of AI in OC diagnosis via blood biomarkers. AI algorithms exhibited exceptional diagnostic capabilities for OC, boasting a pooled sensitivity of 85% (95% CI 83%-87%) and specificity of 91% (95% CI 90%-92%). Moreover, we identified substantial heterogeneity among the selected studies and determined the potential contributing factors through subgroup and meta-regression analyses. Overall, these results should be interpreted with caution as described by the constraints mentioned in subsequent sections.

Heterogeneity

Heterogeneity is an inevitable problem in meta-analyses [94]. Significant interstudy heterogeneity was noted in terms of sensitivity (I^2=97.54%) and specificity (I^2=99.15%) in this study. External validation emerged as a crucial variable influencing study heterogeneity. Studies without external validation might yield results that were hard to generalize owing to factors such as sample selection bias and the characteristics of the research setting [95]. To address this, future research should focus on standardizing and applying the validation procedures, thus getting closer to the goal of providing more accurate and reliable diagnostic tools for clinical practice [96]. Besides, several factors contributed to the heterogeneity observed in sensitivity in this study. Studies with a pronounced risk of bias were predisposed to introduce uncertainties. Such biases could stem from flaws in study design, improper data collection, or inappropriate statistical analyses, all of which might distort the true relationship between the biomarker and the condition under investigation [67,97]. On the other hand, geographic disparities might be attributed to a complex interplay of genetic polymorphisms and environmental factors that differentially modulate biomarker expression levels, which could result in significant variability in biomarker sensitivity [98,99]. In addition, larger sample sizes generally offer enhanced statistical power and precision, enabling more reliable estimations of biomarker performance [100]. Understanding and accounting for these factors comprehensively will help to further reduce heterogeneity and enhance the validity and clinical relevance of meta-analysis results, ultimately leading to more precise and useful diagnostic tools for clinical application.

Implication of Blood Sample Types

Blood specimens are relatively stable and can be easily accessed. Therefore, blood-based biomarkers have been regarded as a minimally invasive method with great value for disease diagnosis [101,102]. Plasma and serum are rich sources of information regarding an individual's health state and are the focus of this study's investigation [103]. Particularly noteworthy is that serum samples showed higher sensitivity than plasma ones in this study, which can potentially be ascribed to multiple

XSL•FO
RenderX

factors. First, the distinct sample preparation procedures for serum and plasma may lead to variations in the concentration and availability of biomarkers. Serum is obtained after blood clotting, during which certain intracellular components can be released, potentially augmenting the repertoire of biomarkers [104]. In contrast, the anticoagulants used in plasma collection may impede the integrity or accessibility of biomarkers [105]. Second, the microenvironment within serum and plasma varies. Serum contains a more intricate milieu of proteins, enzymes, and other biomolecules that can interact with cancer biomarkers in a manner that heightens their detectability [102]. For example, the presence of specific binding proteins or proteases in serum may modify the conformation of biomarkers, rendering them more amenable to detection by the analytic methods used [106]. Furthermore, the centrifugation processes involved in separating serum and plasma can differentially partition the biomarkers. The speed, duration, and temperature of centrifugation may cause certain biomarkers to be preferentially retained in the serum fraction, contributing to the observed higher sensitivity [107]. In addition, the limited number of included studies for plasma may contribute to this phenomenon to some extent. Therefore, it is of paramount importance to judiciously select the sample type in the context of developing and implementing blood-based biomarker assays for OC.

Implication of Algorithm Types

In subgroup analysis, ML surpassed DL in both sensitivity and specificity. This phenomenon warrants an in-depth exploration of the underlying reasons and improvement directions. The edge of ML likely stems from its algorithmic traits. For structured and well-defined data, traditional algorithms (eg, logistic regression and support vector machine) can adeptly capture biomarker-disease associations via mathematical and statistical tenets, yielding high diagnostic accuracy [108,109]. DL, empowered by its strong automatic feature extraction and complex architecture, can theoretically handle large data and extract deep patterns [110]. However, in this study, the number of studies included for DL was only 4, compared with 36 for ML, presumably constraining the exertion of DL's advantages. The scant number of DL studies gives rise to data that are circumscribed in both sample variety and the expanse of feature distribution. To break through this dilemma, several optimization strategies can be considered. First, data augmentation, such as random rotations, scaling, flipping, and noise addition to the data can enhance dataset diversity, facilitating the DL model to learn more extensive features and patterns for better generalization [110-112]. Second, transfer learning is applicable. Using pretrained models from related medical or bioanalysis fields and fine-tuning with OC data, the DL model can draw on prior knowledge, accelerating training convergence and potentially improving performance [113,114]. In addition, model compression techniques, including pruning (ie, eliminating less important connections or neurons to maintain performance while reducing complexity) and quantization (ie, lowering parameter precision for faster inference and less memory use) can be used [115]. While these strategies hold promise, their implementation and efficacy in the context of OC diagnosis warrant further investigation and optimization.

Implication of External Validation

At present, the data amassed for AI applications in OC diagnosis is circumscribed by the paucity of diverse external validation. Many studies rely on a single dataset for discovery, with cross-validation to estimate algorithm performance. Given the generalizable issues to unseen data, accuracy drops when tested on other research datasets, and substantially when tested on clinical data [116,117]. To address these challenges, several strategies can be implemented. First, multicenter collaborations should be actively pursued. Combining data from different medical institutions and regions to build a heterogeneous and comprehensive dataset and exposing algorithms to wider patient characteristics and biomarker profiles will enhance generalizability [118-120]. Second, standardized data collection and annotation protocols are crucial. They ensure data consistency and comparability among studies, minimizing variability from inconsistent methods [121]. This allows algorithms to be trained on more reliable and reproducible data, strengthening the foundation for AI application. Moreover, continuous evaluation and improvement of algorithms in clinical settings are essential. Prospective studies integrating the algorithms into routine practice and monitoring their performance can offer valuable feedback [122]. This iterative testing and refinement process helps algorithms adapt to clinical complexity, leading to more accurate OC diagnostic tools. Despite persisting challenges, we anticipate that these efforts will incrementally enhance diagnostic accuracy. Sustained refinement and collaboration are essential to exploiting the full potential of AI in OC diagnosis.

Future Directions

Beyond the previous mentions, the existing literature in the field exhibits certain areas of improvement to reduce the gap between research and deployment. First, one of the richest data sources of patient health and clinical history is embedded in the electronic health records of a patient but remains hugely underutilized. AI's ability to integrate blood biomarkers with other clinically relevant nonblood biomarkers, such as age, cancer history, and family history of cancer, could potentially outpace current practices if trained on sufficiently extensive datasets [123]. Future research could explore the synergistic integration of AI tools with clinical expertise, echoing a more realistic clinical scenario. Second, the problem of explainability is the subject of intensive research and various initiatives. Although symbolic AI or simple ML models, such as decision trees or linear regression, are still fully understood by people, understanding becomes increasingly difficult with more advanced techniques and is now impossible with many DL models; this situation can lead to unexpected results and nondeterministic behavior [124,125]. Third, data privacy and patient consent are critical concerns that need to be addressed before adopting the use of AI in clinical practice [126]. The integration of AI into clinical workflows requires careful consideration of ethical, legal, and regulatory aspects [127,128]. Transparent guidelines and regulations should be established to govern the use of AI in health care and ensure its responsible and ethical implementation.

XSL•FO
RenderX

Strengths and Limitations

Our meta-analysis has several strengths. First, to the best of our knowledge, it represents a novel effort as the first systematic review and meta-analysis dedicated to evaluating the diagnostic performance of blood biomarker-based AI for OC. Our findings illuminate the considerable potential AI holds in this domain, while also highlighting the advantages of blood tests, such as their noninvasive nature, better patient compliance, and cost-effectiveness. Second, our comprehensive investigation included multiple subgroup analyses, all of which yielded acceptable diagnostic performance for the AI model. Third, the stringent quality assessment of all included studies was conducted using QUADAS-AI tool. In addition, the robustness of our meta-analysis was reinforced through a sensitivity analysis underpinned by a qualitative systematic review.

The results of our meta-analysis are likely to be overestimated or underestimated for some reasons. One limitation lies in the high heterogeneity of the studies included. Nonetheless, we thoroughly explored potential sources of between-study heterogeneity through meta-regression and subgroup analyses. Another limitation is that the contingency tables of 5 studies included in our systematic review were not directly or indirectly available. These studies provided only indicators, such as AUC, accuracy, and F_1-score, which did not allow for the construction of contingency tables [77-81]. Nevertheless, we conducted a qualitative systematic review of these 5 studies and discovered that the findings aligned with the main analysis. Moreover, although no publication bias was noticed, it is still highly likely that there is unpublished material for this topic from the ever-growing nature of the framework and the likelihood of undisclosed research for commercial development [129]. Moreover, available AI research tends to be skewed toward the publication of positive results, indicating a potential publication bias.

Conclusions

The findings of this study indicated that the use of AI for the analysis of noninvasive blood biomarkers in OC diagnostics holds substantial potential for achieving satisfactory predictive outcomes. Among the analyzed studies, those that used DL were notably fewer in number than those that used ML. This underscores a critical need for future research to prioritize the incorporation of DL methodologies. Furthermore, pursuing external validation datasets was a necessary avenue to optimize the performance and applicability of AI in this field.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (QJW: 2022YFC2704205), the Natural Science Foundation of China (QJW: 82073647 and 82373674; TTG: 82103914), Outstanding Scientific Fund of Shengjing Hospital (QJW), Liaoning Province Science and Technology Plan (QJW: 2023JH2/20200019), and Scientific Research Project of Education Department of Liaoning Province (TTG: LJKMZ20221137).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HLX, TTG, and QJW contributed to the study design. HLX, XYL, MQJ, QPM, YHZ, FHL, YQ, YHC, YL, XYC, YLX, DRL, and DDW contributed to the collection and analysis of data. HLX, XYL, MQJ, QPM, DHH, QX, YHZ, SG, XQ, TT, TTG, and QJW wrote the first draft of the manuscript and edited the manuscript. All authors read and approved the final manuscript. HLX, XYL, MQJ, and QPM contributed equally to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 216 KB - jmir_v27i1e67922_app1.pdf]

Multimedia Appendix 2
Search terms and search strategy.
[DOCX File , 11 KB - jmir_v27i1e67922_app2.docx]

Multimedia Appendix 3
Contingency tables extracted from included studies.
[DOCX File , 102 KB - jmir_v27i1e67922_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

The list of the excluded records during the process of full-text review.
[DOCX File , 16 KB - jmir_v27i1e67922_app4.docx]

Multimedia Appendix 5
The results of Quality Assessment of Diagnostic Accuracy Studies–Artificial Intelligence and subgroup analyses.
[DOCX File , 39633 KB - jmir_v27i1e67922_app5.docx]

Multimedia Appendix 6
The characteristics for the 5 studies without sufficient data.
[DOCX File , 19 KB - jmir_v27i1e67922_app6.docx]

Multimedia Appendix 7
Publication bias.
[DOCX File , 93 KB - jmir_v27i1e67922_app7.docx]

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49 [FREE Full text] [doi: 10.3322/caac.21820] [Medline: 38230766]
2. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin 2018 Jul;68(4):284-296 [FREE Full text] [doi: 10.3322/caac.21456] [Medline: 29809280]
3. Li G, Zhang Y, Li K, Liu X, Lu Y, Zhang Z, et al. Transformer-based AI technology improves early ovarian cancer diagnosis using cfDNA methylation markers. Cell Rep Med 2024 Aug 20;5(8):101666 [FREE Full text] [doi: 10.1016/j.xcrm.2024.101666] [Medline: 39094578]
4. Zeng S, Wang XL, Yang H. Radiomics and radiogenomics: extracting more information from medical images for the diagnosis and prognostic prediction of ovarian cancer. Mil Med Res 2024 Dec 14;11(1):77 [FREE Full text] [doi: 10.1186/s40779-024-00580-1] [Medline: 39673071]
5. Dochez V, Caillon H, Vaucel E, Dimet J, Winer N, Ducarme G. Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. J Ovarian Res 2019 Mar 27;12(1):28 [FREE Full text] [doi: 10.1186/s13048-019-0503-7] [Medline: 30917847]
6. Perakis S, Speicher MR. Emerging concepts in liquid biopsies. BMC Med 2017 Apr 06;15(1):75 [FREE Full text] [doi: 10.1186/s12916-017-0840-6] [Medline: 28381299]
7. Qian L, Sun R, Xue Z, Guo T. Mass spectrometry-based proteomics of epithelial ovarian cancers: a clinical perspective. Mol Cell Proteomics 2023 Jul;22(7):100578 [FREE Full text] [doi: 10.1016/j.mcpro.2023.100578] [Medline: 37209814]
8. Dhar C, Ramachandran P, Xu G, Pickering C, Čaval T, Wong M, et al. Diagnosing and staging epithelial ovarian cancer by serum glycoproteomic profiling. Br J Cancer 2024 Jun;130(10):1716-1724. [doi: 10.1038/s41416-024-02644-4] [Medline: 38658783]
9. Terp SK, Stoico MP, Dybkær K, Pedersen IS. Early diagnosis of ovarian cancer based on methylation profiles in peripheral blood cell-free DNA: a systematic review. Clin Epigenetics 2023 Feb 14;15(1):24 [FREE Full text] [doi: 10.1186/s13148-023-01440-w] [Medline: 36788585]
10. Singh A, Shannon CP, Kim YW, Yang CX, Balshaw R, Cohen Freue GV, et al. Novel blood-based transcriptional biomarker panels predict the late-phase asthmatic response. Am J Respir Crit Care Med 2018 Feb 15;197(4):450-462. [doi: 10.1164/rccm.201701-0110OC] [Medline: 29087730]
11. O'Neill S, Bohl M, Gregersen S, Hermansen K, O'Driscoll L. Blood-based biomarkers for metabolic syndrome. Trends Endocrinol Metab 2016 Jun;27(6):363-374. [doi: 10.1016/j.tem.2016.03.012] [Medline: 27150849]
12. Gao Y, Cao D, Li M, Zhao F, Wang P, Mei S, et al. Integration of multiomics features for blood-based early detection of colorectal cancer. Mol Cancer 2024 Aug 22;23(1):173. [doi: 10.1186/s12943-024-01959-3] [Medline: 39175001]
13. Ke C, Hou Y, Zhang H, Fan L, Ge T, Guo B, et al. Large-scale profiling of metabolic dysregulation in ovarian cancer. Int J Cancer 2015 Feb 01;136(3):516-526. [doi: 10.1002/ijc.29010] [Medline: 24895217]
14. Xie J, Xu P, Lin Y, Zheng M, Jia J, Tan X, et al. LncRNA-miRNA interactions prediction based on meta-path similarity and Gaussian kernel similarity. J Cell Mol Med 2024 Oct;28(19):e18590 [FREE Full text] [doi: 10.1111/jcmm.18590] [Medline: 39347925]
15. Yin S, Xu P, Jiang Y, Yang X, Lin Y, Zheng M, et al. Predicting the potential associations between circRNA and drug sensitivity using a multisource feature-based approach. J Cell Mol Med 2024 Oct;28(19):e18591 [FREE Full text] [doi: 10.1111/jcmm.18591] [Medline: 39347936]
16. Wang T, Sun J, Zhao Q. Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism. Comput Biol Med 2023 Feb;153:106464. [doi: 10.1016/j.compbiomed.2022.106464] [Medline: 36584603]

XSL•FO
RenderX

17. Wang J, Zhang L, Sun J, Yang X, Wu W, Chen W, et al. Predicting drug-induced liver injury using graph attention mechanism and molecular fingerprints. Methods 2024 Jan;221:18-26. [doi: 10.1016/j.ymeth.2023.11.014] [Medline: 38040204]
18. Mohsen F, Al-Absi HR, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. NPJ Digit Med 2023 Oct 25;6(1):197 [FREE Full text] [doi: 10.1038/s41746-023-00933-5] [Medline: 37880301]
19. Chen Z, Zhang L, Sun J, Meng R, Yin S, Zhao Q. DCAMCP: a deep learning model based on capsule network and attention mechanism for molecular carcinogenicity prediction. J Cell Mol Med 2023 Oct;27(20):3117-3126 [FREE Full text] [doi: 10.1111/jcmm.17889] [Medline: 37525507]
20. Wang W, Zhang L, Sun J, Zhao Q, Shuai J. Predicting the potential human lncRNA-miRNA interactions based on graph convolution network with conditional random field. Brief Bioinform 2022 Nov 19;23(6):bbac463. [doi: 10.1093/bib/bbac463] [Medline: 36305458]
21. Zhu F, Niu Q, Li X, Zhao Q, Su H, Shuai J. FM-FCN: a neural network with filtering modules for accurate vital signs extraction. Research (Wash D C) 2024;7:0361 [FREE Full text] [doi: 10.34133/research.0361] [Medline: 38737196]
22. Liu L, Wei Y, Zhang Q, Zhao Q. SSCRB: predicting circRNA-RBP interaction sites using a sequence and structural feature-based attention model. IEEE J Biomed Health Inform 2024 Mar;28(3):1762-1772. [doi: 10.1109/JBHI.2024.3354121] [Medline: 38224504]
23. Yang X, Sun J, Jin B, Lu Y, Cheng J, Jiang J, et al. Multi-task aquatic toxicity prediction model based on multi-level features fusion. J Adv Res 2025 Feb;68:477-489 [FREE Full text] [doi: 10.1016/j.jare.2024.06.002] [Medline: 38844122]
24. Cai G, Huang F, Gao Y, Li X, Chi J, Xie J, et al. Artificial intelligence-based models enabling accurate diagnosis of ovarian cancer using laboratory tests in China: a multicentre, retrospective cohort study. Lancet Digit Health 2024 Mar;6(3):e176-e186 [FREE Full text] [doi: 10.1016/S2589-7500(23)00245-5] [Medline: 38212232]
25. Irajizad E, Han CY, Celestino J, Wu R, Murage E, Spencer R, et al. A blood-based metabolite panel for distinguishing ovarian cancer from benign pelvic masses. Clin Cancer Res 2022 Nov 01;28(21):4669-4676 [FREE Full text] [doi: 10.1158/1078-0432.CCR-22-1113] [Medline: 36037307]
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
27. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [doi: 10.1001/jama.283.15.2008] [Medline: 10789670]
28. Abuzinadah N, Kumar Posa S, Alarfaj AA, Alabdulqader EA, Umer M, Kim TH, et al. Improved prediction of ovarian cancer using ensemble classifier and shaply explainable AI. Cancers (Basel) 2023 Dec 11;15(24):5793 [FREE Full text] [doi: 10.3390/cancers15245793] [Medline: 38136346]
29. Bifarin OO, Fernández FM. Automated machine learning and explainable AI (AutoML-XAI) for metabolomics: improving cancer diagnostics. J Am Soc Mass Spectrom 2024 Jun 05;35(6):1089-1100 [FREE Full text] [doi: 10.1021/jasms.3c00403] [Medline: 38690775]
30. Cameron JM, Sala A, Antoniou G, Brennan PM, Butler HJ, Conn JJ, et al. A spectroscopic liquid biopsy for the earlier detection of multiple cancer types. Br J Cancer 2023 Nov;129(10):1658-1666 [FREE Full text] [doi: 10.1038/s41416-023-02423-7] [Medline: 37717120]
31. Chen L, Ma R, Luo C, Xie Q, Ning X, Sun K, et al. Noninvasive early differential diagnosis and progression monitoring of ovarian cancer using the copy number alterations of plasma cell-free DNA. Transl Res 2023 Dec;262:12-24. [doi: 10.1016/j.trsl.2023.07.005] [Medline: 37499745]
32. Hamidi F, Gilani N, Arabi Belaghi R, Yaghoobi H, Babaei E, Sarbakhsh P, et al. Identifying potential circulating miRNA biomarkers for the diagnosis and prediction of ovarian cancer using machine-learning approach: application of Boruta. Front Digit Health 2023;5:1187578 [FREE Full text] [doi: 10.3389/fdgth.2023.1187578] [Medline: 37621964]
33. Lai Z, Wang Z, Yuan Z, Zhang J, Zhou J, Li D, et al. Disease-specific haptoglobin N-glycosylation in inflammatory disorders between cancers and benign diseases of 3 types of female internal genital organs. Clin Chim Acta 2023 Jul 01;547:117420. [doi: 10.1016/j.cca.2023.117420] [Medline: 37285951]
34. Li J, Li Y, Li Q, Sun L, Tan Q, Zheng L, et al. An aptamer-based nanoflow cytometry method for the molecular detection and classification of ovarian cancers through profiling of tumor markers on small extracellular vesicles. Angew Chem Int Ed Engl 2024 Jan 22;63(4):e202314262. [doi: 10.1002/anie.202314262] [Medline: 38012811]
35. Reilly GP, Dunton CJ, Bullock RG, Ure DR, Fritsche H, Ghosh S, et al. Validation of a deep neural network-based algorithm supporting clinical management of adnexal mass. Front Med (Lausanne) 2023;10:1102437 [FREE Full text] [doi: 10.3389/fmed.2023.1102437] [Medline: 36756174]
36. Zhang T, Pang A, Lyu J, Ren H, Song J, Zhu F, et al. Application of nonlinear models combined with conventional laboratory indicators for the diagnosis and differential diagnosis of ovarian cancer. J Clin Med 2023 Jan 20;12(3):844 [FREE Full text] [doi: 10.3390/jcm12030844] [Medline: 36769493]

XSL•FO
RenderX

37. Ahamad MM, Aktar S, Uddin MJ, Rahman T, Alyami SA, Al-Ashhab S, et al. Early-stage detection of ovarian cancer based on clinical data using machine learning approaches. J Pers Med 2022 Jul 25;12(8):1211 [FREE Full text] [doi: 10.3390/jpm12081211] [Medline: 35893305]
38. Bahado-Singh RO, Ibrahim A, Al-Wahab Z, Aydas B, Radhakrishna U, Yilmaz A, et al. Precision gynecologic oncology: circulating cell free DNA epigenomic analysis, artificial intelligence and the accurate detection of ovarian cancer. Sci Rep 2022 Nov 03;12(1):18625 [FREE Full text] [doi: 10.1038/s41598-022-23149-1] [Medline: 36329159]
39. Gupta A, Sagar G, Siddiqui Z, Rao KV, Nayak S, Saquib N, et al. A non-invasive method for concurrent detection of early-stage women-specific cancers. Sci Rep 2022 Feb 10;12(1):2301 [FREE Full text] [doi: 10.1038/s41598-022-06274-9] [Medline: 35145183]
40. Hinestrosa JP, Kurzrock R, Lewis JM, Schork NJ, Schroeder G, Kamat AM, et al. Early-stage multi-cancer detection using an extracellular vesicle protein-based blood test. Commun Med (Lond) 2022;2:29 [FREE Full text] [doi: 10.1038/s43856-022-00088-6] [Medline: 35603292]
41. Kim M, Chen C, Wang P, Mulvey JJ, Yang Y, Wun C, et al. Detection of ovarian cancer via the spectral fingerprinting of quantum-defect-modified carbon nanotubes in serum by machine learning. Nat Biomed Eng 2022 Mar;6(3):267-275 [FREE Full text] [doi: 10.1038/s41551-022-00860-y] [Medline: 35301449]
42. Li N, Zhu X, Nian W, Li Y, Sun Y, Yuan G, et al. Blood-based DNA methylation profiling for the detection of ovarian cancer. Gynecol Oncol 2022 Nov;167(2):295-305. [doi: 10.1016/j.ygyno.2022.07.008] [Medline: 36096974]
43. Pais RJ, Zmuidinaite R, Lacey JC, Jardine CS, Iles RK. A rapid and affordable screening tool for early-stage ovarian cancer detection based on MALDI-ToF MS of blood serum. Appl Sci 2022 Mar 16;12(6):3030. [doi: 10.3390/app12063030]
44. Jeong S, Son DS, Cho M, Lee N, Song W, Shin S, et al. Evaluation of combined cancer markers with lactate dehydrogenase and application of machine learning algorithms for differentiating benign disease from malignant ovarian cancer. Cancer Control 2021;28:10732748211033401 [FREE Full text] [doi: 10.1177/10732748211033401] [Medline: 34923833]
45. Lu M, Fan Z, Xu B, Chen L, Zheng X, Li J, et al. Using machine learning to predict ovarian cancer. Int J Med Inform 2020 Sep;141:104195. [doi: 10.1016/j.ijmedinf.2020.104195] [Medline: 32485554]
46. Banaei N, Moshfegh J, Mohseni-Kabir A, Houghton JM, Sun Y, Kim B. Machine learning algorithms enhance the specificity of cancer biomarker detection using SERS-based immunoassays in microfluidic chips. RSC Adv 2019 Jan 14;9(4):1859-1868 [FREE Full text] [doi: 10.1039/c8ra08930b] [Medline: 35516124]
47. Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. Oncotarget 2018 Apr 27;9(32):22717-22726 [FREE Full text] [doi: 10.18632/oncotarget.25216] [Medline: 29854310]
48. Ivanova OM, Ziganshin RH, Arapidi GP, Kovalchuk SI, Azarkin IV, Sorokina AV, et al. Scope and limitations of MALDI-TOF MS blood serum peptide profiling in cancer diagnostics. Russ J Bioorg Chem 2016 Sep 27;42(5):497-505. [doi: 10.1134/S1068162016050071]
49. Jiang W, Huang R, Duan C, Fu L, Xi Y, Yang Y, et al. Identification of five serum protein markers for detection of ovarian cancer by antibody arrays. PLoS One 2013;8(10):e76795 [FREE Full text] [doi: 10.1371/journal.pone.0076795] [Medline: 24116163]
50. Yang J, Zhu Y, Guo H, Wang X, Gao R, Zhang L, et al. Identifying serum biomarkers for ovarian cancer by screening with surface-enhanced laser desorption/ionization mass spectrometry and the artificial neural network. Int J Gynecol Cancer 2013 May;23(4):667-672. [doi: 10.1097/IGC.0b013e31827e1989] [Medline: 23574802]
51. Shan L, Chen YA, Davis L, Han G, Zhu W, Molina AD, et al. Measurement of phospholipids may improve diagnostic accuracy in ovarian cancer. PLoS One 2012;7(10):e46846 [FREE Full text] [doi: 10.1371/journal.pone.0046846] [Medline: 23082132]
52. Thakur A, Mishra V, Jain SK. Feed forward artificial neural network: tool for early detection of ovarian cancer. Sci Pharm 2011;79(3):493-505 [FREE Full text] [doi: 10.3797/scipharm.1105-11] [Medline: 21886899]
53. Donach M, Yu Y, Artioli G, Banna G, Feng W, Bast RC, et al. Combined use of biomarkers for detection of ovarian cancer in high-risk women. Tumour Biol 2010 Jun;31(3):209-215 [FREE Full text] [doi: 10.1007/s13277-010-0032-x] [Medline: 20393825]
54. Ziganshin RK, Alekseev DG, Arapidi GP, Ivanov VT, Moshkovskiĭ SA, Govorun VM. Serum proteome profiling for ovarion cancer diagnosis using ClinProt magnetic bead technique and MALDI-TOF-mass-spectrometry [Article in Russian]. Biomed Khim 2008;54(4):408-419. [Medline: 18988457]
55. Liu C, Shea N, Rucker S, Harvey L, Russo P, Saul R, et al. Proteomic patterns for classification of ovarian cancer and CTCL serum samples utilizing peak pairs indicative of post-translational modifications. Proteomics 2007 Nov;7(22):4045-4052. [doi: 10.1002/pmic.200601044] [Medline: 17952875]
56. Zhang Z, Yu Y, Xu F, Berchuck A, van Haaften-Day C, Havrilesky LJ, et al. Combining multiple serum tumor markers improves detection of stage I epithelial ovarian cancer. Gynecol Oncol 2007 Dec;107(3):526-531 [FREE Full text] [doi: 10.1016/j.ygyno.2007.08.009] [Medline: 17920110]
57. Chatterjee M, Mohapatra S, Ionan A, Bawa G, Ali-Fehmi R, Wang X, et al. Diagnostic markers of ovarian cancer by high-throughput antigen cloning and detection on arrays. Cancer Res 2006 Jan 15;66(2):1181-1190 [FREE Full text] [doi: 10.1158/0008-5472.CAN-04-2962] [Medline: 16424057]

XSL•FO
RenderX

58. Lin YW, Lin CY, Lai HC, Chiou JY, Chang CC, Yu MH, et al. Plasma proteomic pattern as biomarkers for ovarian cancer. Int J Gynecol Cancer 2006;16 Suppl 1:139-146. [doi: 10.1111/j.1525-1438.2006.00475.x] [Medline: 16515582]
59. Liu Y. Serum proteomic pattern analysis for early cancer detection. Technol Cancer Res Treat 2006 Feb;5(1):61-66 [FREE Full text] [doi: 10.1177/153303460600500108] [Medline: 16417403]
60. Wu SP, Lin YW, Lai HC, Chu TY, Kuo YL, Liu HS. SELDI-TOF MS profiling of plasma proteins in ovarian cancer. Taiwan J Obstet Gynecol 2006 Mar;45(1):26-32 [FREE Full text] [doi: 10.1016/S1028-4559(09)60186-8] [Medline: 17272204]
61. Li J, Ramamohanarao K. A tree-based approach to the discovery of diagnostic biomarkers for ovarian cancer. In: Proceedings of the 8th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. 2004 Presented at: PAKDD '04; May 26-28, 2004; Sydney, Australia p. 682-691 URL: https://link.springer.com/chapter/10.1007/978-3-540-24775-3_80 [doi: 10.1007/978-3-540-24775-3_80]
62. Li L, Tang H, Wu Z, Gong J, Gruidl M, Zou J, et al. Data mining techniques for cancer detection using serum proteomic profiling. Artif Intell Med 2004 Oct;32(2):71-83. [doi: 10.1016/j.artmed.2004.03.006] [Medline: 15364092]
63. Zhang Z, Barnhill SD, Zhang H, Xu F, Yu Y, Jacobs I, et al. Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecol Oncol 1999 Apr;73(1):56-61. [doi: 10.1006/gyno.1999.5320] [Medline: 10094881]
64. Wilding P, Morgan MA, Grygotis AE, Shoffner MA, Rosato EF. Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. Cancer Lett 1994 Mar 15;77(2-3):145-153. [doi: 10.1016/0304-3835(94)90097-3] [Medline: 8168061]
65. Xue P, Wang J, Qin D, Yan H, Qu Y, Seery S, et al. Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis. NPJ Digit Med 2022 Feb 15;5(1):19 [FREE Full text] [doi: 10.1038/s41746-022-00559-z] [Medline: 35169217]
66. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. Nat Med 2021 Oct;27(10):1663-1665. [doi: 10.1038/s41591-021-01517-0] [Medline: 34635854]
67. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
68. Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou Y, et al. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. EClinicalMedicine 2022 Nov;53:101662 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101662] [Medline: 36147628]
69. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 2005 Jan;79(1):16-20. [doi: 10.1016/j.athoracsur.2004.09.040] [Medline: 15620907]
70. Khatami F, Saatchi M, Zadeh SS, Aghamir ZS, Shabestari AN, Reis LO, et al. A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis. Sci Rep 2020 Dec 28;10(1):22402 [FREE Full text] [doi: 10.1038/s41598-020-80061-2] [Medline: 33372194]
71. Jones CM, Ashrafian H, Skapinakis P, Arora S, Darzi A, Dimopoulos K, et al. Diagnostic accuracy meta-analysis: a review of the basic principles of interpretation and application. Int J Cardiol 2010 Apr 15;140(2):138-144. [doi: 10.1016/j.ijcard.2009.05.063] [Medline: 19592119]
72. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. Cancer Discov 2021 Apr;11(4):900-915 [FREE Full text] [doi: 10.1158/2159-8290.CD-21-0090] [Medline: 33811123]
73. Schlattmann P. Tutorial: statistical methods for the meta-analysis of diagnostic test accuracy studies. Clin Chem Lab Med 2023 Apr 25;61(5):777-794 [FREE Full text] [doi: 10.1515/cclm-2022-1256] [Medline: 36656998]
74. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006 Jun;11(2):193-206. [doi: 10.1037/1082-989X.11.2.193] [Medline: 16784338]
75. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005 Sep;58(9):882-893. [doi: 10.1016/j.jclinepi.2005.01.016] [Medline: 16085191]
76. Phillips B, Stewart LA, Sutton AJ. 'Cross hairs' plots for diagnostic meta-analysis. Res Synth Methods 2010 Jul;1(3-4):308-315. [doi: 10.1002/jrsm.26] [Medline: 26061474]
77. Jing B, Chen G, Yang M, Zhang Z, Zhang Y, Zhang J, et al. Development of prediction model to estimate future risk of ovarian lesions: a multi-center retrospective study. Prev Med Rep 2023 Oct;35:102296 [FREE Full text] [doi: 10.1016/j.pmedr.2023.102296] [Medline: 37455762]
78. Lu H, Liu Y, Wang J, Fu S, Wang L, Huang C, et al. Detection of ovarian cancer using plasma cell-free DNA methylomes. Clin Epigenetics 2022 Jun 09;14(1):74 [FREE Full text] [doi: 10.1186/s13148-022-01285-9] [Medline: 35681212]
79. Kawakami E, Tabata J, Yanaihara N, Ishikawa T, Koseki K, Iida Y, et al. Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. Clin Cancer Res 2019 May 15;25(10):3006-3015. [doi: 10.1158/1078-0432.CCR-18-3378] [Medline: 30979733]

XSL•FO
RenderX

80. Song HJ, Yang ES, Kim JD, Park CY, Kyung MS, Kim YS. Best serum biomarker combination for ovarian cancer classification. Biomed Eng Online 2018 Nov 06;17(Suppl 2):152 [FREE Full text] [doi: 10.1186/s12938-018-0581-6] [Medline: 30396341]
81. Elias KM, Fendler W, Stawiski K, Fiascone SJ, Vitonis AF, Berkowitz RS, et al. Diagnostic potential for a serum miRNA neural network for detection of ovarian cancer. Elife 2017 Oct 31;6:e28932 [FREE Full text] [doi: 10.7554/eLife.28932] [Medline: 29087294]
82. González G, Lakatos K, Hoballah J, Fritz-Klaus R, Al-Johani L, Brooker J, et al. Characterization of cell-bound CA125 on immune cell subtypes of ovarian cancer patients using a novel imaging platform. Cancers (Basel) 2021 Apr 25;13(9):2072 [FREE Full text] [doi: 10.3390/cancers13092072] [Medline: 33922973]
83. Gao Q, Lin YP, Li BS, Wang GQ, Dong LQ, Shen BY, et al. Unintrusive multi-cancer detection by circulating cell-free DNA methylation sequencing (THUNDER): development and independent validation studies. Ann Oncol 2023 May;34(5):486-495 [FREE Full text] [doi: 10.1016/j.annonc.2023.02.010] [Medline: 36849097]
84. Ramamoorthy S, Sundaramoorthy S, Gupta A, Siddiqui Z, Sagar G, Rao KV, et al. Beta-validation of a non-invasive method for simultaneous detection of early-stage female-specific cancers. medRxiv Preprint posted online October 6, 2023 [FREE Full text] [doi: 10.1101/2023.10.06.23296638]
85. Luan Y, Zhong G, Li S, Wu W, Liu X, Zhu D, et al. A panel of seven protein tumour markers for effective and affordable multi-cancer early detection by artificial intelligence: a large-scale and multicentre case-control study. EClinicalMedicine 2023 Jul;61:102041 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102041] [Medline: 37387788]
86. Marcišauskas S, Ulfenborg B, Kristjansdottir B, Waldemarson S, Sundfeldt K. Univariate and classification analysis reveals potential diagnostic biomarkers for early stage ovarian cancer Type 1 and Type 2. J Proteomics 2019 Mar 30;196:57-68. [doi: 10.1016/j.jprot.2019.01.017] [Medline: 30710757]
87. Chiappa V, Interlenghi M, Bogani G, Salvatore C, Bertolina F, Sarpietro G, et al. A decision support system based on radiomics and machine learning to predict the risk of malignancy of ovarian masses from transvaginal ultrasonography and serum CA-125. Eur Radiol Exp 2021 Jul 26;5(1):28 [FREE Full text] [doi: 10.1186/s41747-021-00226-0] [Medline: 34308487]
88. Hwangbo S, Kim SI, Kim JH, Eoh KJ, Lee C, Kim YT, et al. Development of machine learning models to predict platinum sensitivity of high-grade serous ovarian carcinoma. Cancers (Basel) 2021 Apr 14;13(8):1875 [FREE Full text] [doi: 10.3390/cancers13081875] [Medline: 33919797]
89. Akazawa M, Hashimoto K. Artificial intelligence in ovarian cancer diagnosis. Anticancer Res 2020 Aug;40(8):4795-4800. [doi: 10.21873/anticanres.14482] [Medline: 32727807]
90. Wu M, Zhao Y, Dong X, Jin Y, Cheng S, Zhang N, et al. Artificial intelligence-based preoperative prediction system for diagnosis and prognosis in epithelial ovarian cancer: a multicenter study. Front Oncol 2022;12:975703 [FREE Full text] [doi: 10.3389/fonc.2022.975703] [Medline: 36212430]
91. Yue Z, Sun C, Chen F, Zhang Y, Xu W, Shabbir S, et al. Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis. Biomed Opt Express 2021 May 01;12(5):2559-2574 [FREE Full text] [doi: 10.1364/BOE.421961] [Medline: 34123488]
92. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
93. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med 2018 Dec;284(6):603-619 [FREE Full text] [doi: 10.1111/joim.12822] [Medline: 30102808]
94. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect 2014 Feb;20(2):105-113 [FREE Full text] [doi: 10.1111/1469-0691.12474] [Medline: 24274632]
95. Wang SC, Nickel G, Venkatesh KP, Raza MM, Kvedar JC. AI-based diabetes care: risk prediction models and implementation concerns. NPJ Digit Med 2024 Feb 15;7(1):36 [FREE Full text] [doi: 10.1038/s41746-024-01034-7] [Medline: 38361152]
96. Hsu W, Hippe DS, Nakhaei N, Wang PC, Zhu B, Siu N, et al. External validation of an ensemble model for automated mammography interpretation by artificial intelligence. JAMA Netw Open 2022 Nov 01;5(11):e2242343 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.42343] [Medline: 36409497]
97. Tomlinson E, Cooper C, Davenport C, Rutjes AW, Leeflang M, Mallett S, et al. Common challenges and suggestions for risk of bias tool development: a systematic review of methodological studies. J Clin Epidemiol 2024 Jul;171:111370 [FREE Full text] [doi: 10.1016/j.jclinepi.2024.111370] [Medline: 38670243]
98. Wu D, Dou J, Chai X, Bellis C, Wilm A, Shih CC, et al. Large-scale whole-genome sequencing of three diverse Asian populations in Singapore. Cell 2019 Oct 17;179(3):736-49.e15 [FREE Full text] [doi: 10.1016/j.cell.2019.09.019] [Medline: 31626772]
99. Padilla CM, Kihal-Talantikit W, Perez S, Deguen S. Use of geographic indicators of healthcare, environment and socioeconomic factors to characterize environmental health disparities. Environ Health 2016 Jul 22;15(1):79 [FREE Full text] [doi: 10.1186/s12940-016-0163-7] [Medline: 27449640]
100. Gibson E, Hu Y, Huisman HJ, Barratt DC. Designing image segmentation studies: statistical power, sample size and reference standard quality. Med Image Anal 2017 Dec;42:44-59 [FREE Full text] [doi: 10.1016/j.media.2017.07.004] [Medline: 28772163]

XSL•FO
RenderX

101. Gao F, Dai L, Wang Q, Liu C, Deng K, Cheng Z, et al. Blood-based biomarkers for Alzheimer's disease: a multicenter-based cross-sectional and longitudinal study in China. Sci Bull (Beijing) 2023 Aug 30;68(16):1800-1808 [FREE Full text] [doi: 10.1016/j.scib.2023.07.009] [Medline: 37500404]
102. Zhu Y. Plasma/serum proteomics based on mass spectrometry. Protein Pept Lett 2024;31(3):192-208 [FREE Full text] [doi: 10.2174/0109298665286952240212053723] [Medline: 38869039]
103. Geyer PE, Voytik E, Treit PV, Doll S, Kleinhempel A, Niu L, et al. Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. EMBO Mol Med 2019 Nov 07;11(11):e10427 [FREE Full text] [doi: 10.15252/emmm.201910427] [Medline: 31566909]
104. Zhang X, Takeuchi T, Takeda A, Mochizuki H, Nagai Y. Comparison of serum and plasma as a source of blood extracellular vesicles: increased levels of platelet-derived particles in serum extracellular vesicle fractions alter content profiles from plasma extracellular vesicle fractions. PLoS One 2022;17(6):e0270634 [FREE Full text] [doi: 10.1371/journal.pone.0270634] [Medline: 35749554]
105. Lan J, Núñez Galindo A, Doecke J, Fowler C, Martins RN, Rainey-Smith SR, et al. Systematic evaluation of the use of human plasma and serum for mass-spectrometry-based shotgun proteomics. J Proteome Res 2018 Apr 06;17(4):1426-1435. [doi: 10.1021/acs.jproteome.7b00788] [Medline: 29451788]
106. Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C, et al. Differences between human plasma and serum metabolite profiles. PLoS One 2011;6(7):e21230 [FREE Full text] [doi: 10.1371/journal.pone.0021230] [Medline: 21760889]
107. Malm L, Tybring G, Moritz T, Landin B, Galli J. Metabolomic quality assessment of EDTA plasma and serum samples. Biopreserv Biobank 2016 Oct;14(5):416-423. [doi: 10.1089/bio.2015.0092] [Medline: 27348730]
108. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]
109. Parvatikar PP, Patil S, Khaparkhuntikar K, Patil S, Singh PK, Sahana R, et al. Artificial intelligence: machine learning approach for screening large database and drug discovery. Antiviral Res 2023 Dec;220:105740. [doi: 10.1016/j.antiviral.2023.105740] [Medline: 37935248]
110. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun (Lond) 2020 Apr;40(4):154-166 [FREE Full text] [doi: 10.1002/cac2.12012] [Medline: 32277744]
111. Wu M, Yan C, Liu H, Liu Q. Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. Biosci Rep 2018 Jun 29;38(3):BSR20180289 [FREE Full text] [doi: 10.1042/BSR20180289] [Medline: 29572387]
112. Teramoto A, Tsukamoto T, Kiriyama Y, Fujita H. Automated classification of lung cancer types from cytological images using deep convolutional neural networks. Biomed Res Int 2017;2017:4067832 [FREE Full text] [doi: 10.1155/2017/4067832] [Medline: 28884120]
113. Mathivanan SK, Sonaimuthu S, Murugesan S, Rajadurai H, Shivahare BD, Shah MA. Employing deep learning and transfer learning for accurate brain tumor detection. Sci Rep 2024 Mar 27;14(1):7232 [FREE Full text] [doi: 10.1038/s41598-024-57970-7] [Medline: 38538708]
114. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT, Alinejad-Rokny H. A review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues. J Biomed Inform 2021 Jan;113:103627 [FREE Full text] [doi: 10.1016/j.jbi.2020.103627] [Medline: 33259944]
115. Tung F, Mori G. Deep neural network compression by in-parallel pruning-quantization. IEEE Trans Pattern Anal Mach Intell 2020 Mar;42(3):568-579. [doi: 10.1109/TPAMI.2018.2886192] [Medline: 30561340]
116. Winchester LM, Harshfield EL, Shi L, Badhwar A, Khleifat AA, Clarke N, et al. Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia. Alzheimers Dement 2023 Dec;19(12):5860-5871. [doi: 10.1002/alz.13390] [Medline: 37654029]
117. Mårtensson G, Ferreira D, Granberg T, Cavallin L, Oppedal K, Padovani A, et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. Med Image Anal 2020 Dec;66:101714 [FREE Full text] [doi: 10.1016/j.media.2020.101714] [Medline: 33007638]
118. Wang X, Chou K, Zhang G, Zuo Z, Zhang T, Zhou Y, et al. Breast cancer pre-clinical screening using infrared thermography and artificial intelligence: a prospective, multicentre, diagnostic accuracy cohort study. Int J Surg 2023 Oct 01;109(10):3021-3031 [FREE Full text] [doi: 10.1097/JS9.0000000000000594] [Medline: 37678284]
119. Zeng Q, Klein C, Caruso S, Maille P, Allende DS, Mínguez B, et al. Artificial intelligence-based pathology as a biomarker of sensitivity to atezolizumab-bevacizumab in patients with hepatocellular carcinoma: a multicentre retrospective study. Lancet Oncol 2023 Dec;24(12):1411-1422. [doi: 10.1016/S1470-2045(23)00468-0] [Medline: 37951222]
120. Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: current status and future challenges - a systematic review. Artif Intell Med 2021 Oct;120:102164. [doi: 10.1016/j.artmed.2021.102164] [Medline: 34629152]
121. Parimbelli E, Wilk S, Cornet R, Sniatala P, Sniatala K, Glaser SL, et al. A review of AI and data science support for cancer management. Artif Intell Med 2021 Jul;117:102111 [FREE Full text] [doi: 10.1016/j.artmed.2021.102111] [Medline: 34127240]

XSL•FO
RenderX

122. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. JAMA Cardiol 2021 Jun 01;6(6):624-632 [FREE Full text] [doi: 10.1001/jamacardio.2021.0185] [Medline: 33599681]
123. Xiang H, Xiao Y, Li F, Li C, Liu L, Deng T, et al. Development and validation of an interpretable model integrating multimodal information for improving ovarian cancer diagnosis. Nat Commun 2024 Mar 27;15(1):2681 [FREE Full text] [doi: 10.1038/s41467-024-46700-2] [Medline: 38538600]
124. Nensa F, Demircioglu A, Rischpler C. Artificial intelligence in nuclear medicine. J Nucl Med 2019 Sep;60(Suppl 2):29S-37S [FREE Full text] [doi: 10.2967/jnumed.118.220590] [Medline: 31481587]
125. Nussberger AM, Luo L, Celis LE, Crockett MJ. Public attitudes value interpretability but prioritize accuracy in artificial intelligence. Nat Commun 2022 Oct 03;13(1):5821 [FREE Full text] [doi: 10.1038/s41467-022-33417-3] [Medline: 36192416]
126. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey. J Med Internet Res 2021 Aug 26;23(8):e26162 [FREE Full text] [doi: 10.2196/26162] [Medline: 34236994]
127. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics 2021 Sep 15;22(1):122 [FREE Full text] [doi: 10.1186/s12910-021-00687-3] [Medline: 34525993]
128. Larson DB, Magnus DC, Lungren MP, Shah NH, Langlotz CP. Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. Radiology 2020 Jun;295(3):675-682. [doi: 10.1148/radiol.2020192536] [Medline: 32208097]
129. Sarayar R, Lestari YD, Setio AA, Sitompul R. Accuracy of artificial intelligence model for infectious keratitis classification: a systematic review and meta-analysis. Front Public Health 2023;11:1239231 [FREE Full text] [doi: 10.3389/fpubh.2023.1239231] [Medline: 38074720]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
DL: deep learning
ML: machine learning
MOOSE: Meta-Analysis of Observational Studies in Epidemiology
OC: ovarian cancer
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-AI: Quality Assessment of Diagnostic Accuracy Studies–Artificial Intelligence
TN: true negatives
TP: true positives

Edited by A Mavragani; submitted 24.10.24; peer-reviewed by Q Zhao, Z Jiang; comments to author 18.12.24; revised version received 06.01.25; accepted 22.01.25; published 24.03.25.

Please cite as:
Xu HL, Li XY, Jia MQ, Ma QP, Zhang YH, Liu FH, Qin Y, Chen YH, Li Y, Chen XY, Xu YL, Li DR, Wang DD, Huang DH, Xiao Q, Zhao YH, Gao S, Qin X, Tao T, Gong TT, Wu QJ
AI-Derived Blood Biomarkers for Ovarian Cancer Diagnosis: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e67922
URL: https://www.jmir.org/2025/1/e67922
doi:10.2196/67922
PMID:40126546

©He-Li Xu, Xiao-Ying Li, Ming-Qian Jia, Qi-Peng Ma, Ying-Hua Zhang, Fang-Hua Liu, Ying Qin, Yu-Han Chen, Yu Li, Xi-Yang Chen, Yi-Lin Xu, Dong-Run Li, Dong-Dong Wang, Dong-Hui Huang, Qian Xiao, Yu-Hong Zhao, Song Gao, Xue Qin, Tao Tao, Ting-Ting Gong, Qi-Jun Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Facilitators and Barriers to Digital Mental Health Interventions for Depression, Anxiety, and Stress in Adolescents and Young Adults: Scoping Review

Shimin Zhu[1,2*], PhD; Yongyi Wang[1*], MSc; Yuxi Hu[1], MA

[1]Department of Applied Social Sciences, Hong Kong Polytechnic University, Hong Kong SAR, China (Hong Kong)

[2]Mental Health Research Centre, The Hong Kong Polytechnic University, Hong Kong SAR, China, Hong Kong SAR, China (Hong Kong)

[*]these authors contributed equally

Corresponding Author:
Shimin Zhu, PhD
Department of Applied Social Sciences
Hong Kong Polytechnic University
11 Yuk Choi Road, Hung Hom, Kowloon
Hong Kong SAR, 999077
China (Hong Kong)
Phone: 852 27665787
Email: jasmine.zhu@polyu.edu.hk

Abstract

Background: Digital mental health interventions (DMHIs) offer unique strengths as emerging services with practical applications for adolescents and young adults (AYAs) experiencing depression, anxiety, and stress. Although promising, acceptance and participation in DMHIs vary across interventions, participants, and contexts. It is essential to delineate and synthesize the factors that promote or hinder DMHI use.

Objective: This review aims to assess and synthesize the facilitators and barriers to accessing DMHIs for depression, anxiety, and stress in AYAs through a scoping review.

Methods: A comprehensive search was conducted across multiple databases, including PubMed, Web of Science, PsycINFO, CNKI, OpenGrey, and APA PsycExtra, up to October 31, 2023. Articles examining facilitators and barriers to DMHIs among AYAs with disorders or symptoms of depression, anxiety, and stress were included. Data synthesis and analysis involved quality assessment, thematic analysis, and relative frequency meta-analysis.

Results: A total of 27 records met the eligibility criteria, and 14 facilitators and 13 barriers were identified across the external, intervention, and individual levels. The relative frequency meta-analysis indicated that factors influencing AYAs' use of DMHIs varied based on delivery modes. Among these factors, "quality and effect" emerged as the predominant theme—high quality and effect served as the primary facilitator, while low quality and effect acted as a barrier across both portable and nonportable devices, as well as single and multiple platforms.

Conclusions: The uptake of DMHIs among AYAs is influenced by a complex interplay of facilitators and barriers, particularly those related to quality and effect. Our syntheses provide crucial guidance for intervention designers, emphasizing the importance of user-centered approaches that balance scientific rigor with engaging and adaptive features. Enhancing the alignment of DMHIs with adolescent needs can improve both adoption and real-world mental health impact.

Trial Registration: PROSPERO CRD42023479880; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023479880

(J Med Internet Res 2025;27:e62870) doi:10.2196/62870

KEYWORDS

digital mental health interventions; adolescents; young adults; common mental disorders; thematic analysis; relative frequency of occurrence

XSL•FO
RenderX

Introduction

Digital mental health interventions (DMHIs) have emerged and gained popularity as new tools and approaches for mental health services, driven by the development of the internet and mobile devices [1]. Nearly half of the global population with mental health needs lacks access to treatment or services, whereas more than half has access to smartphones and the internet. This contrast underscores the growing shift of mental health services toward mobile and digital health solutions [2]. DMHIs refer to intervention processes and programs that deliver mental health services through the web, technology, and mobile platforms [3,4]. Various forms of DMHIs have been found effective in promoting mental health, including smart digital apps, such as innovative communication apps designed to enhance communication and support young people experiencing low mood and suicidal thoughts [5]; web-based programs, such as a psychoeducational multimedia program for young people suffering from or at a high risk of depression, as well as their families, caregivers, friends, and professionals [6,7]; interactive games [8]; and email and text message communication [9,10].

DMHIs offer several advantages and have strong potential to become widely adopted service options, particularly for adolescents and young adults (AYAs). First, DMHIs align with the digital era and are well-suited to the new generation, who are familiar with, comfortable using, and frequent users of the web, digital media, and screen devices. According to the International Telecommunication Union's Facts and Figures 2023, 79% of people aged 15-24 years use the internet globally [11]. Second, compared with face-to-face help-seeking or treatment, DMHIs can alleviate AYAs' feelings of shame and effectively address concerns about stigmatization [12]. This was confirmed by young people's feedback on a DMHI app, which suggested that discreet and easy-to-conceal apps helped mitigate the stigma associated with mental health problems [13]. Third, DMHIs are user-friendly. Some services adopt youth-friendly language and expressions while adjusting their layout to enhance usability [12]. Additionally, DMHIs offer flexibility, as they are not restricted by time, allowing users to engage with them according to their schedules [14]. Fourth, DMHIs provide high accessibility. Unlike traditional psychotherapy approaches, such as counseling, which require significant time for queuing and waiting, DMHIs have greatly improved AYAs' access to mental health support through digital and online platforms [15]. At the same time, DMHIs can help mitigate the inaccessibility of mental health services caused by geographical remoteness or exceptional circumstances, such as lockdowns during infectious disease outbreaks [16,17]. Lastly, DMHIs offer scalability, as they have the potential to reach a wider audience due to their anonymity, accessibility, cost-effectiveness, and ability to provide timely feedback [4].

AYAs' health, including mental health, plays a crucial role throughout the life course [18]. Health during this period is fundamental to their development, shaping long-term well-being and influencing the foundation of a healthy life for the next generation [18]. However, common mental disorders (CMDs) among adolescents are highly prevalent and contribute significantly to the burden of noncommunicable diseases [19-21]. CMDs encompass distress states characterized by anxiety, depression, and unexplained somatic symptoms [22]. According to data from the World Health Organization (WHO) and various regions, 1 in 7 individuals aged 10-19 years had a mental disorder in 2021, accounting for 13% of the global burden of disease in this age group [19]. Moreover, a 2023 survey in the United States found that nearly 90% of youth faced mental health challenges [23]. In the postpandemic era, CMDs present a growing challenge, highlighting the need for more accessible mental health services [24].

CMDs in AYAs require early and timely intervention; however, the uptake of mental health services remains inadequate [25-27]. The incidence of mental disorders has been reported to increase significantly after the age of 14 years [28]. Without timely intervention, these disorders can persist into adulthood, potentially impairing both physical and mental health and limiting opportunities for a fulfilling life [19]. Although mental health services and resources are currently available [18], a large proportion of AYAs do not access them [26]. Reasons for low service utilization are stigma, limited knowledge, low trust in the therapeutic relationship, high costs, accessibility issues, and other barriers [27], preventing those in need from receiving early interventions. DMHIs may serve as a viable alternative for youth requiring mental health services.

Although DMHIs offer many advantages [8,10,17], AYAs' intention to use them and their usage patterns vary. For effectiveness, a web-based mental health intervention program for adolescents found that participants with high adherence (using the site for 30 minutes or more per week) reported significantly lower depression and stress levels, along with improved well-being [7]. Another study examining the effectiveness of a new computerized cognitive behavioral therapy program found that it led to meaningful improvements in participants' depression levels [8]. Nevertheless, the usage, intention, engagement, and adherence to DMHIs remain relatively low [29]. For instance, a study in the United Kingdom examining young people's attitudes toward computerized therapy found that only 25% expressed interest, another 25% were not interested, and half were unsure [30]. Additionally, several intervention studies have reported inadequate engagement and completion rates, with low user adherence (participants completing less than half of the intervention components) and high attrition rates (over 20%) [31]. Given the high demand for mental health services for CMDs among AYAs, it is essential to explore users' perspectives on the factors that facilitate or hinder their use of DMHIs. Identifying these facilitators and barriers is crucial for the effective development and promotion of DMHIs.

Previous studies have rarely focused on the experiences, attitudes, or perceptions of AYAs regarding DMHIs as a primary research objective [32,33]. Much of the existing research has collected participants' feedback only after evaluating a specific DMHI [17,33]. As a result, current findings on the facilitators and barriers to DMHI utilization are largely indirect and fragmented, highlighting the need for a more cohesive and comprehensive synthesis.

XSL•FO
RenderX

This study aims to conduct a scoping review to examine the existing literature on the facilitators and barriers to DMHI use among AYAs experiencing depression, anxiety, and stress. In this study, facilitators are defined as factors that enhance access to, usage of, or intention to use DMHIs, while barriers are factors that hinder use or reduce the intention to use them [34]. Synthesizing these facilitators and barriers will provide critical insights for promoting DMHIs to better address mental health needs.

Methods

Scoping Review Framework and Registration

This scoping review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) checklist to identify and map key characteristics, facilitators, barriers, and related themes in AYAs' use of DMHIs (Multimedia Appendix 1) [35]. The review protocol was registered on PROSPERO (CRD42023479880).

Search Strategy

The following major electronic databases were searched as of October 31, 2023: PubMed, Web of Science, PsycINFO, and CNKI. Although ScienceDirect was initially listed in our registered protocol, it was excluded from the final search due to significant content overlap with other databases. However, we ensured comprehensive coverage by searching additional relevant databases. Given that DMHIs are emerging technologies, no start date was set for the literature search to capture a broader range of publications. To supplement the search strategy, we manually retrieved bibliographies from relevant studies and included gray literature sources (OpenGrey and APA PsycExtra). The searches were restricted to English and Chinese. Further details are provided in Multimedia Appendix 2.

Eligibility Criteria

This research included studies examining facilitators and barriers to DMHIs among AYAs with disorders or symptoms of depression, anxiety, and stress. Reviews, recommendations, comments, newspapers, letters, conference abstracts, and research from other stakeholders' perspectives were used to enhance the understanding of the topic but were excluded from the analysis. The following exclusion criteria were applied: (1) studies involving participants with serious illnesses (eg, heart failure or trauma), as these conditions could impose greater limitations on their thoughts and behaviors; (2) studies with participants younger than 10 or older than 26 years; (3) studies on digital health services unrelated to mental health; (4) studies assessing only the effectiveness of DMHIs, participants' attitudes, or willingness—without discussing influencing factors; and (5) studies with unavailable full texts due to access restrictions or incomplete records were excluded to ensure transparency and replicability of the review.

Study Selection and Data Extraction

The literature screening process was conducted using EndNote 20 (Clarivate Plc), and Microsoft Excel 2020 was used to record the extracted data. Two independent researchers (YW and YH) reviewed the literature and extracted data. First, YW identified and removed duplicate records, which were then manually checked for accuracy by YH. Second, the researchers independently screened studies based on titles and abstracts, followed by a full-text review to determine eligibility. Third, data extraction was conducted independently by both researchers. Any discrepancies at each stage were discussed and resolved by 3 researchers (YW, YH, and SZ). The extracted data from each included study comprised 4 dimensions: (1) metadata and study context (eg, authors, year, study design, and sample size); (2) characteristics of the study population (eg, country or region, race, sample type, and basic demographics); (3) characteristics of DMHIs (eg, type, delivery mode, number of sessions, and inclusion of self-help or self-directed tools); and (4) facilitators and barriers influencing AYAs with CMDs in accessing DMHIs.

Data Synthesis and Analysis

In line with the WHO's 3-level digital health outcomes assessment framework [36], which includes the health system, provider, and client perspectives, this review adopted a thematic analysis framework comprising 3 levels: (1) external level, (2) intervention level, and (3) individual level.

Thematic analysis in this study followed a 6-phase process (Textbox 1) [37].

XSL•FO
RenderX

Textbox 1. Thematic analysis.

1. **Familiarization with data**

   All researchers (SZ, YW, and YH) thoroughly read the literature, annotating key insights and recording analytical ideas for discussion.

2. **Initial coding**

   YW and YH independently conducted manual coding, identifying and categorizing relevant data segments on facilitators and barriers. Their coded results were compared and collated to ensure accuracy and comprehensiveness.

3. **Theme development**

   SZ, YW, and YH discussed and organized the codes into themes, subthemes, and levels, refining their relationships. YW then supplemented and structured the original data for clarity.

4. **Review of themes**

   YW reexamined the data extraction, subthemes, and thematic framework, engaging in discussions with SZ and YH until a consensus was reached.

5. **Definition and refinement**

   The essence of each theme and subtheme was defined. YW drafted the initial narrative, which was then reviewed and refined by SZ and YH for coherence, consistency, and minimal overlap.

6. **Final reporting**

   A structured and logical report was generated, presenting the thematic findings.

After identifying themes and subthemes, a relative frequency meta-analysis was conducted to assess the occurrence of each facilitator and barrier across different delivery modes: completely nonportable devices, portable devices, single-platform, and multiple-platform interventions. Studies that did not specify the delivery model were excluded. The analysis used the relative frequency of occurrence (RFO) with 95% CIs, performed using the metaprop function in R (R Foundation) [38].

Quality Assessment

The quality of the included literature was assessed using the Mixed Methods Appraisal Tool (MMAT) [39]. This study applied the 2018 version, which was developed through a literature review of critical appraisal tools, user interviews, and electronic Delphi consultations with international experts. The MMAT evaluates 5 research categories: (1) qualitative studies, (2) randomized controlled trials, (3) nonrandomized studies, (4) quantitative descriptive studies, and (5) mixed methods studies. In addition to 2 general screening questions applicable to all study types, 5 specific criteria were established for each research category—qualitative, quantitative randomized controlled trials, quantitative nonrandomized studies, quantitative descriptive studies, and mixed methods studies [39].

Results

Search Selection

A total of 6063 records were retrieved from electronic databases and gray literature sources. After removing duplicates, 2498 records remained. Title and abstract screening identified 131 studies for full-text review, of which 27 met the inclusion and exclusion criteria and were included in data extraction, synthesis, and analysis (Figure 1; also see Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram. AYA: adolescents and young adult; DMHI: digital mental health intervention.

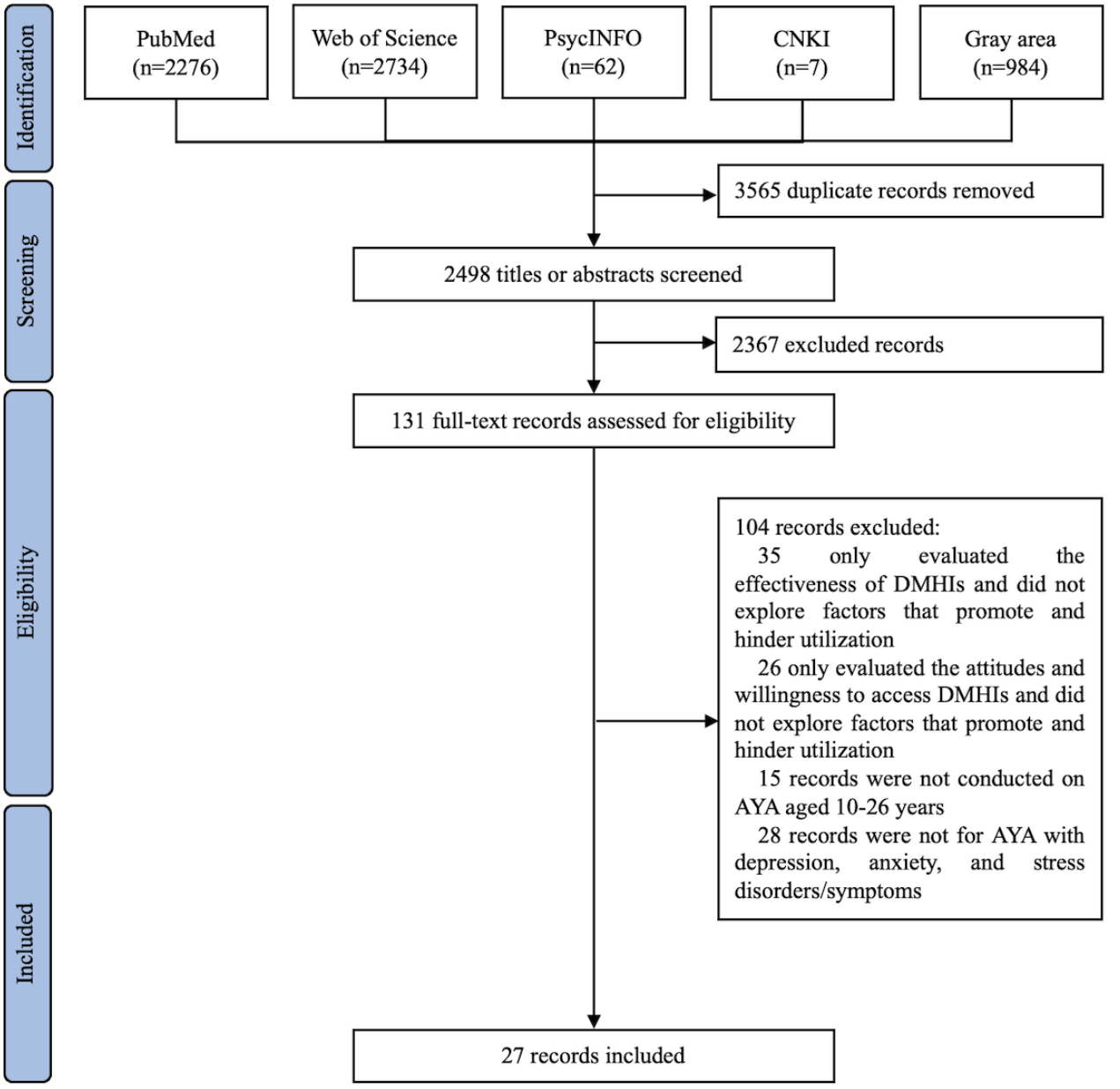

Description of Included Studies

Table 1 presents the characteristics of the included studies. Among the 27 studies, 7 (26%) were quantitative, 6 (22%) were qualitative, and 14 (52%) applied mixed methods. Geographically, 4 (15%) studies were conducted in England, 5 (19%) in the United States, 3 (11%) in New Zealand, 7 (26%) in Australia, 2 (7%) in Ireland, and 2 (7%) in China. Additionally, 1 (4%) study was conducted in each of Canada, Sweden, and South Africa, while 1 (4%) spanned both Chile and Colombia. Regarding participant recruitment channels, 5 (19%) studies recruited participants from health institutions, 4 (15%) from the community, 6 (22%) from schools, and 2 (7%) from youth centers and a survey, respectively. Additionally, 10 (37%) studies recruited participants from multiple sources, including health institutions, schools, communities, teams, and existing research studies. The DMHIs examined in the included studies can be broadly categorized into web/internet-based, computer-based, app-based, game-based, and other formats. Specifically, 13 (48%) studies delivered DMHIs via web/internet-based programs, 6 (22%) were app-based, and 2 (7%) were game-based. One study (4%) utilized a text messaging tool, while 1 (4%) implemented a chatbot. Several interventions employed hybrid delivery methods, such as text messaging combined with web-based programs (1/27, 4%), computer-based combined with web-based programs (2/27, 7%), and web-based combined with app-based interventions (1/27, 4%). Regarding conflicts-of-interest disclosures, 19 studies included a conflicts-of-interest component, with 7 studies [5,8,10,16,17,40,41] explicitly declaring potential or actual conflicts of interest.

Regarding the mental health symptoms and therapeutic approaches covered by the DMHIs in the included studies, 22 out of 27 (81%) aimed to improve depressive symptoms. Additionally, 14 studies addressed other mental health

XSL•FO
RenderX

challenges, including anxiety, stress, worry, anger, and suicidal ideation. Notably, 3 studies [12,32,42] specifically targeted anxiety disorders. Among the 20 studies that explicitly stated a theoretical foundation for intervention development, 14 incorporated cognitive behavioral therapy, with 3 [8,15,30] utilizing computerized cognitive behavioral therapy. Furthermore, 2 studies [17,43] were based on peer support, 1 on behavioral activation [44], 1 on behavior change and resiliency [40], 1 on positive psychology [7], and 1 on mindfulness [45].

XSL•FO
RenderX

Table 1. Characteristics of articles included in this scoping review.

Study	Study design	Location	Race[a]	Sample type	Number of participants[a]	Participants' characteristics[a]	Type and delivery mode	Sessions	Self-help/self-directed tool
Horgan et al [43]	Mixed methods	Ireland	98.3% White and 1.7% Asian or Asian Irish	School	118	Age: 18-24 years, mean 20.6 years, 64.4% males	A website; no specific mode of delivery	Not specific	No, peer support
Giovanelli et al [16]	Mixed methods	United States	64% White, 14% Asian, 14% Black, and 7% mixed race	Community and school	14 (interviews: 5)	Age: 15-18 years, mean 16 years	Appa Health, a smartphone app	Video sessions weekly	Not specific
Van Voorhees Benjamin et al [40]	Quantitative	United States	23% African American, 5% Hispanic, 6% Asian, and 4% other	Health institution	83	Age: 14-21 years, mean (SD) 17.4 (2.14) years, 56% females	An internet-based depression prevention program; no specific mode of delivery	Not specific	Not specific
Suffoletto et al [10]	Quantitative	United States	92.30% White, 1.92% Black, and 5.77% more than one	Health institution (primary care and mental health clinic)	52 at baseline (45 completed follow-ups)	Age (of the control group): mean (SD): 18.7 (0.48) years, 100% females; age (of the intervention group): mean (SD): 18.7 (0.42) years, 79% females	MoST-MH, an automated mobile support tool, delivered via mobile phones	Not specific	Yes
Goodyear-Smith et al [51]	Mixed methods	New Zealand	Not specific	School and community	30	Age: <25 years, 93% females	YouthCHAT, questions delivered on an e-tablet	Not specific	Yes
Gericke et al [33]	Qualitative	South Africa	77.78% White and 22.22% Black	School	9	Age: 17-20 years, mean (SD) 18.9 (1.2) years, 66.67% females	ICare, a transdiagnostic semiguided iCBT[b] intervention; no specific mode of delivery	7	Not specific
Sweeney et al [14]	Quantitative	Australia	Not specific	Community	217	Age: 13-18 years, mean 16.98 years, 71.9% females	Online therapy; no specific mode of delivery	Not specific	Not specific
Sit et al [44]	Mixed methods	China	Not specific	School	38 (interviews: 6)	Age: 18-25 years	Step-by-Step, a mobile app, delivered via smartphones or laptops	5	Yes
Thabrew et al [5]	Mixed methods	New Zealand	15% Māori, 65% New Zealand European, 15% Asian, and 4% Middle Eastern, Latin American, and African	Health institution and community	Quantitative: 26 young people; qualitative: 13 young people	Age (of young people): 16-25 years, mean: 17.7 years, 65% females	"Village," a digital communication app, delivered via smartphones	Not specific	Not sure
Lilja et al [32]	Mixed methods	Sweden	Not specific	Health institution	14	Age: 13-18 years, 93% females	An iCBT program (Anxiety Help for Adolescents, a guided internet-delivered self-help treatment program); no specific mode of delivery	Not specific	Yes

XSL•FO
RenderX

Study	Study design	Location	Race[a]	Sample type	Number of participants[a]	Participants' characteristics[a]	Type and delivery mode	Sessions	Self-help/self-directed tool
Monshat et al [45]	Qualitative	Australia	Not specific	Community	13	Age: 16-26 years, mean: 22 years, 60% females	Online mindfulness training program; no specific mode of delivery	Not specific	Not specific
Kruzan et al [47]	Qualitative	United States	56% White, 10% Asian, 8% Black or African American, 2% American Indian or Alaskan Native, 12% more than one race, and 14% not reported	Health institution	50	Age: 18-25 years, 76% females	Online screening, delivered via mobile phones	The first asynchronous remote community group: 6; the second asynchronous remote community group: 8	Yes
Kuosmanen et al [15]	Mixed methods	Ireland	Not specific	Youth centers	40	Age: 15-20 years	SPARX-R, a computerized mental health game	Not specific	Yes
Dingwall et al [49]	Mixed methods	Australia	Not specific	Community	33 at baseline (30 completed the 4-week follow-ups)	Age (of 30 young people): 12-18 years, mean (SD): 14.0 (1.55) years, 43.33% females	AIMhi-Y app; smartphone based	4	Not specific
Clark et al [42]	Qualitative	Australia	Not specific	Health institution, research institution, community, and school	29	Age: 12-18 years, mean: 15.17 years	Computerized mental health help‐seeking	Not specific	Not specific
Wozney et al [12]	Mixed methods	Canada	Not specific	Community	Cycle 1: 4 young people; cycle 2: 4 young people	Age (of young people): <20 (the age range for selecting participants was 15-24 years old), 50% females	Breathe, an internet-based cognitive behavioral therapy program; delivered via computers, phones, and emails	2	Yes
Birrell et al [17]	Mixed methods	Australia	Not specific	School	166	Age, mean (SD): 15.3 (0.41) years	Mind your Mate, a mobile app; delivered via smartphones	Not specific	Yes
Stallard et al [30]	Mixed methods	Australia	Not specific	Health institution	37	Age: 8-17 years, mean 14.5 years	Computerized therapy	Not specific	Not specific
Kornfield et al [50]	Qualitative	United States	54.84% White, 16.13% more than 1 race, 12.90% Black or African American, 9.68% Asian, and 6.45% not reported	Survey	Discussion group: 22; co-design workshops: 9	Age: 18-25 years	Automated text messaging tool; delivered via mobile phones	Not specific	Yes

XSL•FO
RenderX

Study	Study design	Location	Race[a]	Sample type	Number of participants[a]	Participants' characteristics[a]	Type and delivery mode	Sessions	Self-help/self-directed tool
Grist et al [53]	Quantitative	England	Not specific	School	775	Age: 11-16 years	Internet and smartphone/tablet apps	Not specific	Yes
Bevan Jones et al [6]	Qualitative	England	Not specific	Health institution and study	Interviews: 4 young people; focus groups: 29 young people in 3 groups	Age (of young people interviewed): 13-18 years, 75% females; age (of young people in focus groups): 13-19 years, 68.97% females	MoodHwb, a web-based program; delivered via tablets or laptops	Not specific	Yes
Bevan Jones et al [46]	Mixed methods	England	95% White (young people interviewed)	Health institution, school, team, and study	Quantitative: 43 young people at baseline (36 completed the follow-ups); qualitative: 19 young people	Age (of young people at baseline): 13-23 years, mean (SD): 16.3 (2.36) years, 79% females; age (of young people interviewed): 14-19 years, mean (SD): 16.5 (1.78) years, 74% females	MoodHwb, a multiplatform program	Not specific	Yes
Merry et al [8]	Quantitative	New Zealand	58.5%-60.2% New Zealand, 22.6%-25.5% European Māori, 7.5%-8.5% Pacific people, 4.3%-8.6% Asian, and 1.1%-3.2% other	Health institution and school	Baseline: 187; postintervention: 170; 3-month follow-up: 168	Age: 12-19 years	Computerized cognitive behavioral therapy intervention (SPARX); a game	7 modules	Yes
Martínez et al [41]	Mixed methods	Chile and Colombia	Not specific	School	199	Age, mean (SD): 14.8 (1.0) years, 53.27% females	Take Care of Your Mood, an internet-based program for prevention and early intervention; delivered via computers or smartphones	Not specific	Not specific
Manicavasagar et al [7]	Mixed methods	Australia	Not specific	School and community	235	Age: 12-18 years	A web-based positive psychology program; delivered via computers	Not specific	Yes
Sawrikar et al [48]	Quantitative	England	Not specific	Community and school	248	Age: 17-25 years, mean (SD): 23.31 (1.91) years, 40.7% females	Digital mental health interventions; delivered via the internet or on a smartphone	Not specific	Not specific

XSL•FO
RenderX

Study	Study design	Location	Race[a]	Sample type	Number of participants[a]	Participants' characteristics[a]	Type and delivery mode	Sessions	Self-help/self-directed tool
He et al [52]	Mixed methods	China	92.57% Han	Community and school	148	Age, mean (SD): 18.78 (0.88) years	Chatbot, a software program with artificial intelligence; delivered via WeChat platform, e-book, etc	25.54 sessions on average	Yes

[a]The "Race," "Number of participants," and "Participants' characteristics" columns only present data on adolescents and young adults and exclude relevant data from other stakeholders (eg, parents, clinicians, and school staff).

[b]iCBT: internet-based cognitive behavioral therapy.

Quality Assessment

The quality assessment (Table 2) indicated that qualitative studies and quantitative randomized controlled trials were generally of high quality. However, all quantitative descriptive studies exhibited varying degrees of sample representativeness.

XSL•FO
RenderX

Table 2. Quality assessment of articles included by the Mixed Methods Appraisal Tool, 2018 version.

	Qualitative					Quantitative randomized controlled trials					Quantitative nonrandomized					Quantitative descriptive					Mixed methods				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Horgan et al [43]	Y[a]	Y	N[b]	C[c]	Y	—[d]	—	—	—	—	—	—	—	—	—	Y	N	Y	N	C	Y	Y	Y	C	N
Giovanelli et al [16]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	Y	N	Y	Y	Y	Y	Y	Y	Y	N
Van Voorhees Benjamin et al [40]	—	—	—	—	—	Y	C	Y	Y	N	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Suffoletto et al [10]	—	—	—	—	—	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Goodyear-Smith et al [51]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	C	N	Y	C	Y	Y	Y	Y	Y	N
Gericke et al [33]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Sweeney et al [14]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	N	Y	C	Y	—	—	—	—	—
Sit et al [44]	Y	Y	Y	C	Y	—	—	—	—	—	N	Y	Y	C	Y	—	—	—	—	—	Y	Y	Y	Y	C
Thabrew et al [5]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	C	C	Y	N	Y	Y	Y	Y	Y	N
Lilja et al [32]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	C	N	Y	Y	Y	Y	Y	Y	Y	N
Monshat et al [45]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Kruzan et al [47]	Y	Y	C	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Kuosmanen et al [15]	Y	Y	C	Y	Y	—	—	—	—	—	—	—	—	—	—	C	N	Y	C	Y	Y	Y	Y	Y	N
Dingwall et al [49]	Y	C	Y	Y	Y	—	—	—	—	—	N	Y	Y	Y	Y	—	—	—	—	—	Y	Y	Y	C	N
Clark et al [42]	Y	Y	Y	C	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Wozney et al [12]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	Y	N	Y	Y	Y	Y	Y	Y	Y	N
Birrell et al [17]	Y	Y	C	C	Y	Y	N	Y	N	N	—	—	—	—	—	—	—	—	—	—	Y	Y	Y	Y	N
Stallard et al [30]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	C	N	Y	Y	Y	Y	Y	Y	Y	N
Kornfield et al [50]	Y	Y	Y	C	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Grist et al [53]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	N	Y	C	Y	—	—	—	—	—
Bevan Jones et al [6]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Bevan Jones et al [46]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	Y	C	Y	C	Y	Y	Y	Y	Y	C
Merry et al [8]	—	—	—	—	—	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Martínez et al [41]	Y	Y	Y	C	Y	—	—	—	—	—	—	—	—	—	—	Y	C	Y	Y	Y	Y	Y	Y	Y	C
Manicavasagar et al [7]	Y	C	C	Y	Y	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	Y	Y	Y	Y	C
Sawrikar et al [48]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Y	N	Y	C	Y	—	—	—	—	—
He et al [52]	Y	Y	C	C	Y	Y	N	Y	Y	N	—	—	—	—	—	—	—	—	—	—	Y	Y	Y	Y	N

[a]Y: Yes.

[b]N: No.

[c]C: Cannot tell.

[d]—: Not applicable.

Thematic Synthesis

Overview

The thematic analysis identified facilitators and barriers across 3 levels: external, intervention, and individual. Figure 2 presents the themes, while Multimedia Appendix 3 provides an overview of the subthemes and corresponding examples. A detailed breakdown of all themes, subthemes, and supporting extracts from the original studies is available in Multimedia Appendix 4.

XSL•FO
RenderX

Figure 2. Themes of facilitators and barriers under the three-level framework.

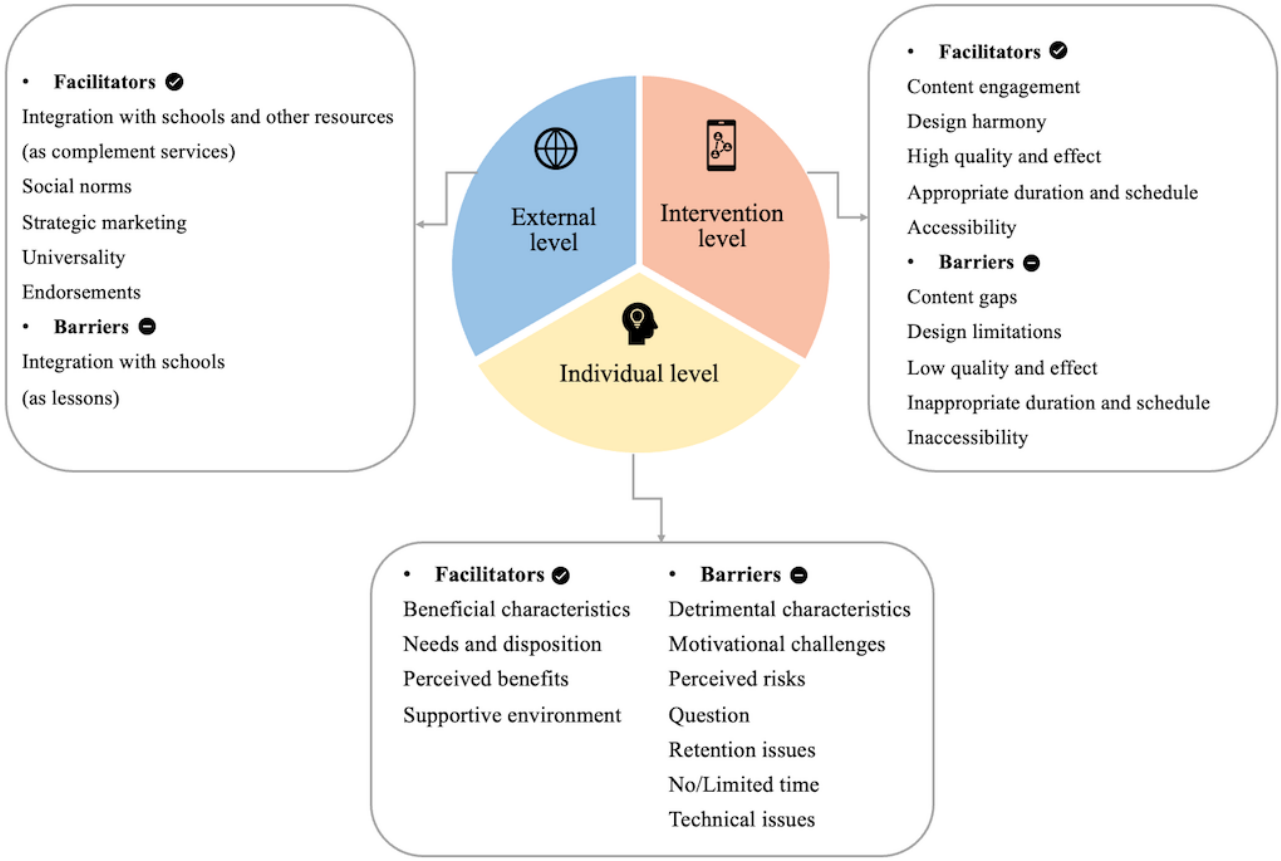

Facilitators: External Level

It is recommended that the adoption and implementation of DMHIs be integrated with other services [46,47], with schools serving as a key environment for AYAs [6]. Social norms, particularly subjective norms, exert pressure based on the expectations of AYAs [48]. Additionally, the strategic marketing and broad accessibility of DMHIs play a crucial role in shaping user motivation [6,45,47], stigma, and social isolation [15]. Endorsements from friends, peers, care providers, professionals, and reputable programs can further enhance assessment and engagement [10,47].

Facilitators: Intervention Level

Literature suggests that key factors promoting widespread acceptance of DMHIs are closely tied to program features, including diverse information types, personalization, comprehensive support, and communication sessions for feedback and sharing [6,14,16,32,33,40,41,44-47,49,50]. Testimonials and entertainment elements were recommended and well-received, alongside retention strategies such as notifications, messages, and incentives [15,41,44,46,49]. Evidence indicates that multimedia delivery, a professional and polished design, co-design, character inclusion, personalization, diverse presentation formats, and appropriate language are critical for enhancing product quality and user appeal [6,10,12,15,16,32,33,41,42,44-47,49,50]. Many participants were surprised by the high quality of the programs and appreciated their overall design and feel [44,49]. Concrete benefits, such as the ability to track progress, were highly praised [14]. Factors contributing to sustained acceptance included an engaging and enjoyable experience, positive emotional responses, ease of interaction and use, high relevance, and appealing visual design [5-7,10,12,14-16,40,41,43-46,48,49,51,52]. The importance of appropriate durations and flexible schedules was also emphasized, ensuring convenience and adaptability to users' needs [6,12,16,41,44-46]. High accessibility further expanded user reach, allowing engagement at any time and from remote locations [6,14,16,41]. Additionally, many participants identified free or low-cost access as a crucial factor in promoting DMHI use [14,46,47].

Facilitators: Individual Level

Female participants reported significantly higher perceived helpfulness of the intervention [14]. Individuals with more severe psychological symptoms, greater knowledge, and prior experience with online therapies demonstrated stronger motivation to seek help and enroll [14,42,48]. AYAs were more likely to engage with DMHIs if they had genuine mental health needs [6,33,43,50,53], preferred solitude or home-based interventions, valued anonymity and autonomy [5,30,33,45], and held positive attitudes toward mental health issues and technology [14,48]. Frequently cited perceived benefits of DMHI use included support during difficult periods, perceived usefulness, privacy, and improved time management [33,49,51]. Additionally, a supportive environment—encompassing both technical infrastructure and interpersonal influences—played a crucial role in facilitating engagement [6,47].

XSL•FO
RenderX

Barriers: External Level

Several studies presented an opposing perspective, suggesting that integrating DMHIs with schools could reduce their appeal [46] and lower participation rates [40,54]. Students might find the association with school tasks frustrating [46], and school policies restricting the use of electronic mobile devices could further limit engagement [54].

Barriers: Intervention Level

Some content components faced criticism. Religious overtones, particularly elements related to meditation and spirituality, were not well received [45]. Additionally, the absence of therapist support and direct human interaction led to frustration and disappointment among users [14,33,44]. Issues such as robotic responses, inappropriate multimedia elements, and poor language choices in design were perceived as barriers, causing confusion and disengagement [5,45,46]. Opinions on personalization varied: while excessive customization [15,44] could create confusion, difficulty, and cognitive burden [5,15], insufficient personalization resulted in overly generalized content [14]. Regarding quality and effectiveness, unattractive design, lack of relevance, and unsatisfactory user experiences were significant hindering factors [7,14,15,51]. Repetitiveness was particularly noted for diminishing the initial appeal of a DMHI website over time [7]. Inappropriate durations and scheduling further contributed to negative experiences [15,44,45,51]. Additionally, inaccessibility—primarily due to technical issues and financial constraints—was identified as an objective barrier to engagement [5,16,47,49,52].

Barriers: Individual Level

Physical illness was identified as a potential reason preventing participants from attending to or completing DMHIs on time [8]. Additional barriers included a lack of confidence and necessary connections to engage with internet-based therapies [17,32]. A lack of motivation was cited as a reason for not even downloading the app [17]. Participants who preferred face-to-face support or had reservations about human-like messaging systems expressed hesitation toward DMHIs [30,50,53]. Perceived risks were particularly emphasized by many AYAs, including concerns about privacy, security, and credibility, as well as fears of stigma and cyberbullying [14,42]. Nonuse or noncompletion of DMHIs was often attributed to doubts regarding their helpfulness, validity, and usefulness, as well as low priority, low interest, and a lack of persistence [5,8,14,17,33,43,45,49,53]. Additionally, limited time and technical issues remained significant barriers to engagement [7,8,14,17,33,49].

Relative Frequency Meta-Analysis

Based on the generated themes, the predominant facilitators and barriers varied across different delivery modes, as shown in Figures 3 and 4. For completely nonportable devices, quality and effect emerged as the most relevant facilitator (RFO=53%, 95% CI 0.24-0.81) and barrier (RFO=42%, 95% CI 0.01-0.91) to DMHI use. By contrast, for portable devices, the primary facilitators included high quality and effect (RFO=21%, 95% CI 0.12-0.31), design harmony (RFO=20%, 95% CI 0.11-0.28), and content engagement (RFO=13%, 95% CI 0.04-0.23), while the dominant barrier was low quality and effect (RFO=15%, 95% CI 0.02-0.35). For single-platform DMHIs, usage was primarily associated with 2 key facilitators: high quality and effect (RFO=32%, 95% CI 0.13-0.50) and perceived benefits (RFO=18%, 95% CI 0.07-0.31). The predominant barrier was low quality and effect (RFO=30%, 95% CI 0.05-0.60). Regarding DMHIs available across multiple platforms, the most influential facilitators were design harmony (RFO=25%, 95% CI 0.12-0.38), high quality and effect (RFO=23%, 95% CI 0.10-0.37), and content engagement (RFO=10%, 95% CI 0.01-0.26). The primary barriers were perceived risks (RFO=17%, 95% CI 0-0.54) and low quality and effect (RFO=13%, 95% CI 0-0.36). Further details are provided in Multimedia Appendix 5.

XSL•FO
RenderX

Figure 3. Relative frequency of occurrence of facilitators in the 4 delivery modes.

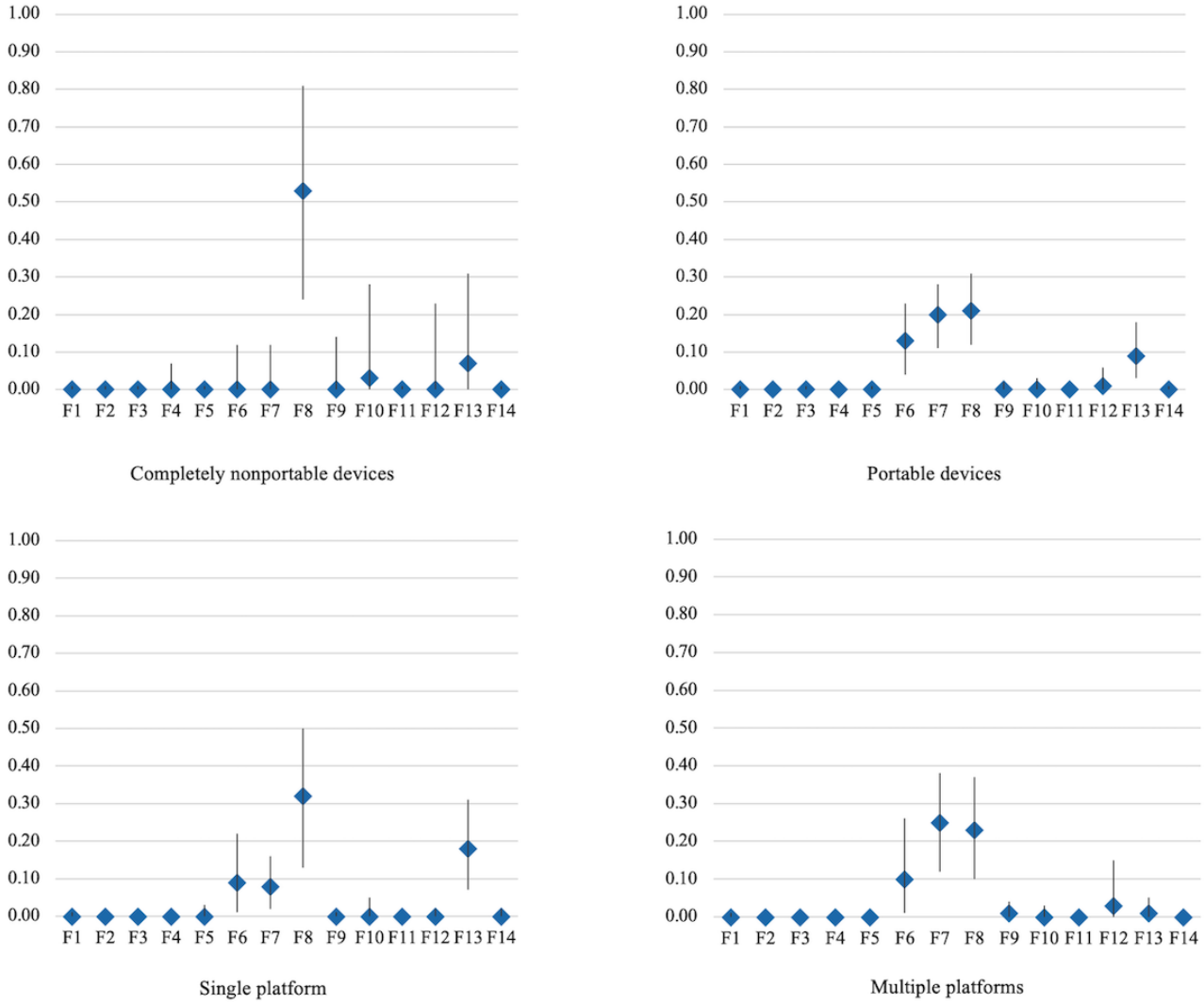

XSL•FO
RenderX

Figure 4. Relative frequency of occurrence of barriers in the 4 delivery modes.

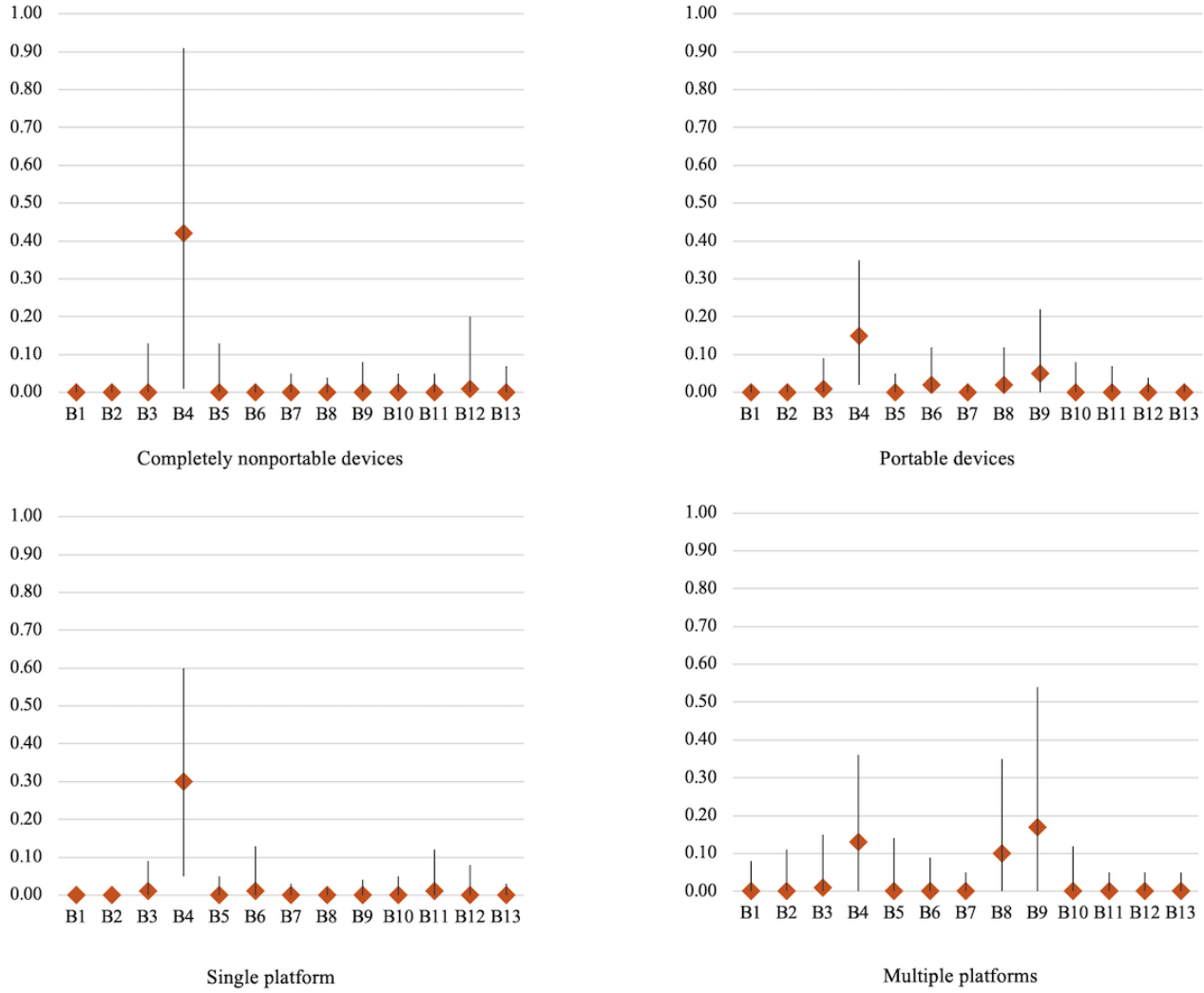

Discussion

Principal Findings

This review systematically identified facilitators and barriers to AYAs' access to DMHIs by analyzing quantitative, qualitative, and mixed methods literature using a 6-step thematic analysis approach. Following WHO guidelines, a 3-level theoretical framework was used to categorize these influential factors. Fourteen facilitators were identified, including integration with schools and other resources, social norms, and strategic marketing. Conversely, 13 barriers were recognized, such as integration with schools, content gaps, and design limitations.

The DMHIs examined in our review primarily targeted depression and were largely based on cognitive behavioral therapy. However, their development and application vary widely. Some studies have shown that DMHIs incorporating mindfulness training and peer support can effectively alleviate psychological challenges, including anxiety, stress, and resilience [43,45]. These findings underscore the potential of diverse DMHI approaches in mental health support and contribute to the growing digital landscape of psychological interventions.

During the RFO analysis, we focused specifically on digital modes of intervention, excluding studies that did not specify their delivery modes. Research supports the notion that different platform types may lead to variations in functionality, usage, and performance [55]. In our study, the theme of "quality and effect" emerged as the predominant facilitator and barrier across all DMHI delivery modes. This finding suggests that the perceived quality and effectiveness of DMHIs were central concerns for both participants and researchers. High-quality interventions that demonstrate positive outcomes are more likely to enhance user engagement, support sustained use, and encourage recommendations to others [51]. Conversely, interventions that fail to meet user expectations or inadequately address their needs may result in disengagement or abandonment [7].

When examining the subthemes, the most common and consistent points across the 3 levels were summarized. Participants showed greater willingness to engage when DMHIs were integrated with other resources [46,47], possibly due to increased accessibility and a reduced risk of judgment or embarrassment [56]. Previous research and professionals particularly valued the benefits of combining DMHIs with traditional medical resources and incorporating them into daily

XSL•FO
RenderX

life [54,57,58]. The use of multimedia, also referred to as "aesthetics" or "visual assets," was widely accepted and recommended for its ability to attract attention, generate interest, facilitate understanding, and enhance usability and satisfaction [12,59,60]. Perceived helpfulness or usefulness aligns with the Technology Acceptance Model, as it influences participants' attitudes toward use, thereby shaping behavioral intentions and actual usage behaviors [61]. Negative emotional experiences led to reduced engagement among AYAs [15], with certain emotions, such as frustration and nervousness, contributing to physical and mental stress [62]. Privacy, security, and credibility concerns remained common barriers in DMHIs, similar to traditional psychotherapy. Despite technological advances and digitization, these remote services pose unique and heightened privacy risks to clients [54,63].

Furthermore, 2 factors—DMHI integration with schools and personalization—can serve as both facilitators and barriers. On the one hand, integrating DMHIs into school settings may significantly enhance accessibility to psychological support by streamlining delivery to students and potentially addressing gaps in resources and teacher training [46]. Embedding mental health interventions within the school environment allows students to access support without requiring external referrals or additional logistical arrangements. On the other hand, integrating DMHIs with schools may make them less appealing to students and could evoke negative feelings [45,46]. Additionally, strict school regulations on smart device usage may limit students' ability to engage freely with DMHIs [54], potentially reducing their effectiveness by restricting flexible and convenient access to these digital tools. Regarding personalization, participants frequently highlighted its advantages, such as greater flexibility, tailored monitoring experiences, feedback, and additional benefits [44,45]. However, some participants viewed personalization negatively, as it often required additional actions and steps, adding to their burden and inconvenience [15]. Balancing the positive and negative aspects of these 2 factors requires strategies that maximize benefits while minimizing drawbacks. Suggested approaches include offering opt-in participation and flexible access options, allowing students to decide whether to engage and choose between on-campus or off-campus access to school-integrated DMHIs to enhance their sense of control and reduce stress. Additionally, making personalization optional by preconfiguring basic settings can improve ease of use, enabling users to engage with personalized features only to the extent they require.

Although many randomized controlled trials indicated that DMHIs were satisfying [8,10], they were not considered a substitute for traditional mental health services but rather an augmentation, potentially expanding accessibility and enhancing engagement [56,64,65]. From users' perspectives, the most intuitive feedback was their preference for face-to-face communication, along with concerns and dissatisfaction regarding privacy issues and the mechanical, fixed responses of DMHIs [5,14,66]. However, this does not imply that users are inclined to reject DMHIs; rather, they support DMHIs as a supplementary form of care alongside traditional mental health services [65]. From a scientific research and development perspective, a hybrid model integrating DMHIs with traditional mental health services can maximize the advantages of both, such as addressing low retention rates in DMHIs and assisting mental health professionals in providing follow-up care [67,68]. Consequently, the notion that DMHIs can serve as a "digital glue" to enhance user engagement in mental health services is more widely accepted, as it facilitates a seamless transition between digital and nondigital services [56]. Therefore, supporting hybrid digital and traditional mental health services should be the central focus for the future development and implementation of DMHIs.

An effective design process is crucial for the success of DMHIs in the mental health field, with a particular emphasis on human-centered or user-centered design [69,70]. Developers and designers must adopt this approach to refine and enhance DMHIs, ensuring they align with users' needs. Future research should focus on identifying the key components among the myriad and complex facilitators and barriers, sharpening the design process, and exploring which types of DMHIs are best suited for different mental health conditions.

Additionally, basic user characteristics may serve as unique influencing factors. Current findings suggest that women tend to spend more time using mental health apps [71], whereas men show greater interest in simulation game-based DMHIs [72]. Limited research has explored differences in DMHI usage based on race and ethnicity, though some evidence indicates that people of color may face barriers to accessing culturally responsive care in DMHIs [3]. Further systematic exploration and synthesis are needed to identify specific cultural factors that facilitate or hinder DMHI uptake among AYAs.

This scoping review summarized the facilitators and barriers to DMHIs for AYAs with depression, anxiety, and stress, categorizing them in a structured manner. By synthesizing the literature, the review offers insights for future intervention service designers and therapists, supporting the translation of DMHIs from research to practice—an essential step for advancing mental and public health. However, our review has some limitations. First, due to our inclusion and exclusion criteria, the quality of the included studies varies, with some not meeting high-quality standards. Second, the participants' characteristics and age ranges in the reviewed studies were constrained by our exclusion criteria. Factors that facilitate or hinder adults or older adults with CMDs from engaging with DMHIs may similarly affect AYAs. Third, we excluded studies published in languages other than English and Chinese, which may have led to the omission of culturally specific factors. Another potential limitation is the search strategy. While we used keyword searches across multiple databases, this approach may have been less comprehensive than using built-in search structures (eg, MeSH terms in PubMed), potentially resulting in the omission of relevant studies. Finally, we acknowledge that the frequency of an individual factor does not necessarily indicate its significance, underscoring the need for follow-up studies to clarify its importance.

Conclusion

This scoping review systematically searched, screened, and synthesized the literature on facilitators and barriers to DMHIs for AYAs with depression, anxiety, and stress. Through thematic

XSL•FO
RenderX

synthesis, we identified a series of themes and subthemes at the external, intervention, and individual levels, highlighting key factors influencing the use and adoption of DMHIs. By consolidating these factors, this review provides insights that can inform the design and implementation of more effective DMHIs for AYAs.

Acknowledgments

We appreciate the contributions of previous researchers to this research topic. We also thank the editors and reviewers for their valuable comments, which have significantly improved this manuscript. SZ received funding from the General Research Fund of the Hong Kong Research Grants Council (grant 15608322). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

SZ conceptualized the study, developed the methodology, validated the results, supervised the research activity planning and execution, and revised and edited the manuscript. YW developed the methodology; selected the studies; extracted, processed, and analyzed the data; wrote the first draft of the manuscript; and reviewed and edited the manuscript. YH selected the studies; extracted, processed, and analyzed the data; and reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File , 108 KB - jmir_v27i1e62870_app1.docx]

Multimedia Appendix 2
Literature search strategy.
[DOCX File , 22 KB - jmir_v27i1e62870_app2.docx]

Multimedia Appendix 3
Themes, subthemes definitions, and related examples.
[DOCX File , 31 KB - jmir_v27i1e62870_app3.docx]

Multimedia Appendix 4
Facilitators and barriers.
[DOCX File , 385 KB - jmir_v27i1e62870_app4.docx]

Multimedia Appendix 5
Metaprop results.
[DOCX File , 15215 KB - jmir_v27i1e62870_app5.docx]

References

1. Wangelin B, Szafranski D, Gros D. Telehealth technologies in evidence-based psychotherapy. In: Computer-Assisted and Web-Based Innovations in Psychology, Special Education, and Health. New York, NY: Academic Press; 2016:119-140.
2. Digital mental health 101: what clinicians need to know when getting started. American Psychiatric Association. URL: https://www.psychiatry.org/getmedia/8042631d-ab6f-4e9e-8aa6-3052c3644f50/APA-Digital-Mental-Health-101-Part-1.pdf?ref=marco.health [accessed 2023-04-03]
3. Kodish T, Schueller SM, Lau AS. Barriers and strategies to improve digital mental health intervention uptake among college students of color: A modified Delphi study. Journal of Behavioral and Cognitive Therapy 2023 Mar;33(1):10-23. [doi: 10.1016/j.jbct.2022.12.002]
4. Liverpool S, Mota CP, Sales CMD, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]

XSL•FO
RenderX

5. Thabrew H, Kumar H, Steadman E. Acceptability and feasibility of "village," a digital communication app for young people experiencing low mood, thoughts of self-harm, and suicidal ideation to obtain support from family and friends: mixed methods pilot open trial. JMIR Form Res 2023 Mar 13;7:e41273 [FREE Full text] [doi: 10.2196/41273] [Medline: 36912882]
6. Bevan Jones R, Thapar A, Rice F, Beeching H, Cichosz R, Mars B, et al. A web-based psychoeducational intervention for adolescent depression: design and development of MoodHwb. JMIR Ment Health 2018 Feb 15;5(1):e13 [FREE Full text] [doi: 10.2196/mental.8894] [Medline: 29449202]
7. Manicavasagar V, Horswood D, Burckhardt R, Lum A, Hadzi-Pavlovic D, Parker G. Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. J Med Internet Res 2014 Jun 04;16(6):e140 [FREE Full text] [doi: 10.2196/jmir.3176] [Medline: 24901900]
8. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
9. Campagnola M, Burlibasa A, Gutman LM. Barriers and enablers to the delivery of email communication for a helpline service for young people. PEC Innov 2022 Dec;1:100059 [FREE Full text] [doi: 10.1016/j.pecinn.2022.100059] [Medline: 37213768]
10. Suffoletto B, Goldstein T, Gotkiewicz D, Gotkiewicz E, George B, Brent D. Acceptability, engagement, and effects of a mobile digital intervention to support mental health for young adults transitioning to college: pilot randomized controlled trial. JMIR Form Res 2021 Oct 14;5(10):e32271 [FREE Full text] [doi: 10.2196/32271] [Medline: 34647893]
11. Measuring Digital Development–Facts and Figures 2023. International Telecommunication Union. 2023. URL: https://www.itu.int/itu-d/reports/statistics/facts-figures-2023/ [accessed 2023-10-10]
12. Wozney L, Baxter P, Newton AS. Usability evaluation with mental health professionals and young people to develop an internet-based cognitive-behaviour therapy program for adolescents with anxiety disorders. BMC Pediatr 2015 Dec 16;15:213 [FREE Full text] [doi: 10.1186/s12887-015-0534-1] [Medline: 26675420]
13. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
14. Sweeney GM, Donovan CL, March S, Forbes Y. Logging into therapy: adolescent perceptions of online therapies for mental health problems. Internet Interv 2019 Mar;15:93-99 [FREE Full text] [doi: 10.1016/j.invent.2016.12.001] [Medline: 30792959]
15. Kuosmanen T, Fleming TM, Barry MM. The implementation of SPARX-R computerized mental health program in alternative education exploring the factors contributing to engagement and dropout. Children and Youth Services Review 2018;84:176-184. [doi: 10.1016/j.childyouth.2017.11.032]
16. Giovanelli A, Sanchez Karver T, Roundfield KD, Woodruff S, Wierzba C, Wolny J, et al. The Appa health app for youth mental health: development and usability study. JMIR Form Res 2023 Oct 04;7:e49998 [FREE Full text] [doi: 10.2196/49998] [Medline: 37792468]
17. Birrell L, Debenham J, Furneaux-Bate A, Prior K, Spallek S, Thornton L, et al. Evaluating a peer-support mobile app for mental health and substance use among adolescents over 12 months during the COVID-19 pandemic: randomized controlled trial. J Med Internet Res 2023 Sep 27;25:e45216 [FREE Full text] [doi: 10.2196/45216] [Medline: 37756116]
18. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478 [FREE Full text] [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
19. Mental health of adolescents. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2024-10-10]
20. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. JAMA Pediatr 2019 Apr 01;173(4):389-391 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5399] [Medline: 30742204]
21. Mangal A, Thakur A, Nimavat KA, Dabar D, Yadav SB. Screening for common mental health problems and their determinants among school-going adolescent girls in Gujarat, India. J Family Med Prim Care 2020 Jan;9(1):264-270 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_732_19] [Medline: 32110602]
22. Risal A. Common mental disorders. Kathmandu Univ Med J (KUMJ) 2011;9(35):213-217. [doi: 10.3126/kumj.v9i3.6308] [Medline: 22946143]
23. Main mental health challenges faced by youth U.S. 2023. Statista. 2023. URL: https://www.statista.com/statistics/1412704/mental-health-challenges-among-us-youth-by-type/ [accessed 2023-11-29]
24. Magklara K, Kyriakopoulos M. The impact of the COVID-19 pandemic on children and young people. Psychiatriki 2023 Dec 29;34(4):265-268 [FREE Full text] [doi: 10.22365/jpsych.2023.024] [Medline: 37982248]
25. Liddle SK, Robinson L, Vella SA, Deane FP. Profiles of mental health help seeking among Australian adolescent males. J Adolesc 2021 Oct;92:34-45. [doi: 10.1016/j.adolescence.2021.08.010] [Medline: 34416479]
26. Lipari R, Hedden S, Blau G, Rubenstein L. Adolescent mental health service use and reasons for using services in specialty, educational, and general medical settings. In: The CBHSQ Report. Rockville MD: Substance Abuse and Mental Health Services Administration; May 5, 2016.

XSL•FO
RenderX

27. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
28. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
29. Hamilton J, Siegel D, Carper M. Digital media interventions for adolescent mental health. In: Handbook of Adolescent Digital Media Use and Mental Health. Cambridge, MA: Cambridge University Press; Jun 30, 2022:389-416.
30. Stallard P, Velleman S, Richardson T. Computer use and attitudes towards computerised therapy amongst young people and parents attending child and adolescent mental health services. Child Adolesc Ment Health 2010 May;15(2):80-84. [doi: 10.1111/j.1475-3588.2009.00540.x] [Medline: 32847246]
31. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019;10:759 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
32. Lilja JL, Rupcic Ljustina M, Nissling L, Larsson AC, Weineland S. Youths' and parents' experiences and perceived effects of internet-based cognitive behavioral therapy for anxiety disorders in primary care: mixed methods study. JMIR Pediatr Parent 2021 Nov 01;4(4):e26842 [FREE Full text] [doi: 10.2196/26842] [Medline: 34723830]
33. Gericke F, Ebert DD, Breet E, Auerbach RP, Bantjes J. A qualitative study of university students? experience of Internet-based CBT for depression. Counselling and Psychotherapy Research 2021;21(4):792-804. [doi: 10.1002/capr.12465]
34. Zaman N, Mujahid K, Ahmed F, Mahmud S, Naeem H, Riaz U, et al. What are the barriers and facilitators to seeking help for mental health in NHS doctors: a systematic review and qualitative study. BMC Psychiatry 2022 Sep 07;22(1):595 [FREE Full text] [doi: 10.1186/s12888-022-04202-9] [Medline: 36071392]
35. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
36. Monitoring and evaluating digital health interventions. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2016-12-12]
37. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006/01/01 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Wang N. Conducting meta-analyses of proportions in R. Journal of Behavioral Data Science 2023;3(2):64-126 [FREE Full text] [doi: 10.35566/jbds/v3n2/wang]
39. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Mixed Methods Appraisal Tool (MMAT) Version 2018 User Guide. Mixed Methods Appraisal Tool (MMAT). 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2018-08-01]
40. Van Voorhees BW, Fogel J, Pomper BE, Marko M, Reid N, Watson N, et al. Adolescent dose and ratings of an internet-based depression prevention program: a randomized trial of primary care physician brief advice versus a motivational interview. J Cogn Behav Psychother 2009;9(1):1-19 [FREE Full text] [Medline: 20694059]
41. Martínez V, Espinosa-Duque D, Jiménez-Molina Á, Rojas G, Vöhringer PA, Fernández-Arcila M, et al. Feasibility and acceptability of "Cuida tu Ánimo" (Take Care of Your Mood): an internet-based program for prevention and early intervention of adolescent depression in Chile and Colombia. Int J Environ Res Public Health 2021 Sep 13;18(18):18189628 [FREE Full text] [doi: 10.3390/ijerph18189628] [Medline: 34574553]
42. Clark LH, Hudson JL, Dunstan DA, Clark GI. Capturing the attitudes of adolescent males? Towards computerised mental health help-seeking. Australian Psychologist 2020;53(5):416-426. [doi: 10.1111/ap.12341]
43. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. Arch Psychiatr Nurs 2013 Apr;27(2):84-89 [FREE Full text] [doi: 10.1016/j.apnu.2012.12.005] [Medline: 23540518]
44. Sit HF, Hong IW, Burchert S, Sou EKL, Wong M, Chen W, et al. A feasibility study of the WHO digital mental health intervention step-by-step to address depression among Chinese young adults. Front Psychiatry 2021;12:812667 [FREE Full text] [doi: 10.3389/fpsyt.2021.812667] [Medline: 35069297]
45. Monshat K, Vella-Brodrick D, Burns J, Herrman H. Mental health promotion in the internet age: a consultation with Australian young people to inform the design of an online mindfulness training programme. Health Promot Int 2012 Jun;27(2):177-186. [doi: 10.1093/heapro/dar017] [Medline: 21398335]
46. Bevan Jones R, Thapar A, Rice F, Mars B, Agha SS, Smith D, et al. A digital intervention for adolescent depression (MoodHwb): mixed methods feasibility evaluation. JMIR Ment Health 2020 Jul 17;7(7):e14536 [FREE Full text] [doi: 10.2196/14536] [Medline: 32384053]
47. Kruzan KP, Meyerhoff J, Nguyen T, Mohr DC, Reddy M, Kornfield R. I Wanted to See How Bad it Was: Online Self-screening as a Critical Transition Point Among Young Adults with Common Mental Health Conditions. Proc SIGCHI Conf Hum Factor Comput Syst 2022 Apr;2022 [FREE Full text] [doi: 10.1145/3491102.3501976] [Medline: 35531062]

XSL•FO
RenderX

48. Sawrikar V, Mote K. Technology acceptance and trust: overlooked considerations in young people's use of digital mental health interventions. Health Policy and Technology 2022;11(4):100686. [doi: 10.1016/j.hlpt.2022.100686]
49. Dingwall KM, Povey J, Sweet M, Friel J, Shand F, Titov N, et al. Feasibility and acceptability of the Aboriginal and Islander Mental Health Initiative for youth app: nonrandomized pilot with first nations young people. JMIR Hum Factors 2023 Jun 07;10:e40111 [FREE Full text] [doi: 10.2196/40111] [Medline: 37285184]
50. Kornfield R, Meyerhoff J, Studd H, Bhattacharjee A, Williams JJ, Reddy M, et al. Meeting users where they are: user-centered design of an automated text messaging tool to support the mental health of young adults. Proc SIGCHI Conf Hum Factor Comput Syst 2022 Apr;2022:3502046 [FREE Full text] [doi: 10.1145/3491102.3502046] [Medline: 35574512]
51. Goodyear-Smith F, Corter A, Suh H. Electronic screening for lifestyle issues and mental health in youth: a community-based participatory research approach. BMC Med Inform Decis Mak 2016 Nov 08;16(1):140 [FREE Full text] [doi: 10.1186/s12911-016-0379-z] [Medline: 27821128]
52. He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, et al. Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: single-blind, three-arm randomized controlled trial. J Med Internet Res 2022 Nov 21;24(11):e40719 [FREE Full text] [doi: 10.2196/40719] [Medline: 36355633]
53. Grist R, Cliffe B, Denne M, Croker A, Stallard P. An online survey of young adolescent girls' use of the internet and smartphone apps for mental health support. BJPsych Open 2018 Jul;4(4):302-306 [FREE Full text] [doi: 10.1192/bjo.2018.43] [Medline: 30083383]
54. Knapp AA, Cohen K, Nicholas J, Mohr DC, Carlo AD, Skerl JJ, et al. Integration of digital tools into community mental health care settings that serve young people: focus group study. JMIR Ment Health 2021 Aug 19;8(8):e27379 [FREE Full text] [doi: 10.2196/27379] [Medline: 34420928]
55. Balcombe L, De Leo D. Evaluation of the use of digital mental health platforms and interventions: scoping review. Int J Environ Res Public Health 2022 Dec 26;20(1):20010362 [FREE Full text] [doi: 10.3390/ijerph20010362] [Medline: 36612685]
56. Bond RR, Mulvenna MD, Potts C, O'Neill S, Ennis E, Torous J. Digital transformation of mental health services. Npj Ment Health Res 2023 Aug 22;2(1):13 [FREE Full text] [doi: 10.1038/s44184-023-00033-y] [Medline: 38609479]
57. Patel V, Saxena S, Frankish H, Boyce N. Sustainable development and global mental health--a Lancet Commission. Lancet 2016 Mar 19;387(10024):1143-1145. [doi: 10.1016/S0140-6736(16)00208-7] [Medline: 27025320]
58. Balcombe L, De Leo D. An integrated blueprint for digital mental health services amidst COVID-19. JMIR Ment Health 2020 Jul 22;7(7):e21718 [FREE Full text] [doi: 10.2196/21718] [Medline: 32668402]
59. Palmer BW, Lanouette NM, Jeste DV. Effectiveness of multimedia aids to enhance comprehension of research consent information: a systematic review. IRB 2012;34(6):1-15 [FREE Full text] [Medline: 23342734]
60. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
61. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
62. Ghahramani F, Wang J. Intention to adopt mHealth apps among informal caregivers: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 17;9(3):e24755 [FREE Full text] [doi: 10.2196/24755] [Medline: 33729166]
63. Lustgarten SD, Garrison YL, Sinnard MT, Flynn AW. Digital privacy in mental healthcare: current issues and recommendations for technology use. Curr Opin Psychol 2020 Dec;36:25-31 [FREE Full text] [doi: 10.1016/j.copsyc.2020.03.012] [Medline: 32361651]
64. Cohen KA, Stiles-Shields C, Winquist N, Lattie EG. Traditional and nontraditional mental healthcare services: usage and preferences among adolescents and younger adults. J Behav Health Serv Res 2021 Oct;48(4):537-553 [FREE Full text] [doi: 10.1007/s11414-020-09746-w] [Medline: 33474642]
65. Younes N, Chollet A, Menard E, Melchior M. E-mental health care among young adults and help-seeking behaviors: a transversal study in a community sample. J Med Internet Res 2015 May 15;17(5):e123 [FREE Full text] [doi: 10.2196/jmir.4254] [Medline: 25979680]
66. Mazurka R, Vallis EH, Chen L, Freeman K, Langley R, Ross B, et al. Preferences for virtual versus in-person mental and physical healthcare in Canada: a descriptive study from a cohort of youth and their parents enriched for severe mental illness. BMJ Paediatr Open 2024 Jan 08;8(1):e002197. [doi: 10.1136/bmjpo-2023-002197] [Medline: 38191204]
67. Lim CT, Fuchs C, Torous J. Integrated digital mental health care: a vision for addressing population mental health needs. Int J Gen Med 2024;17:359-365. [doi: 10.2147/IJGM.S449474] [Medline: 38318335]
68. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJEWC. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
69. Thimbleby H, Lewis A, Williams J. Making healthcare safer by understanding, designing and buying better IT. Clin Med (Lond) 2015 Jun;15(3):258-262 [FREE Full text] [doi: 10.7861/clinmedicine.15-3-258] [Medline: 26031976]

XSL•FO
RenderX

70. Vial S, Boudhraâ S, Dumont M. Human-centered design approaches in digital mental health interventions: exploratory mapping review. JMIR Ment Health 2022 Jun 07;9(6):e35591 [FREE Full text] [doi: 10.2196/35591] [Medline: 35671081]
71. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017 Jan 05;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
72. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]

Abbreviations

AYA: adolescents and young adult
CMDs: common mental disorders
DMHI: digital mental health intervention
MMAT: Mixed Methods Appraisal Tool
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RFO: relative frequency of occurrence
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 03.06.24; peer-reviewed by TM Li, F Kong; comments to author 29.11.24; revised version received 03.01.25; accepted 26.02.25; published 24.03.25.

Please cite as:
Zhu S, Wang Y, Hu Y
Facilitators and Barriers to Digital Mental Health Interventions for Depression, Anxiety, and Stress in Adolescents and Young Adults: Scoping Review
J Med Internet Res 2025;27:e62870
URL: https://www.jmir.org/2025/1/e62870
doi:10.2196/62870
PMID:

©Shimin Zhu, Yongyi Wang, Yuxi Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effects of Digital Intelligent Interventions on Self-Management of Patients With Diabetic Foot: Systematic Review

Jinyan Zhou[1], MM; Shanni Ding[1], MM; Yihong Xu[1], MM; Hongying Pan[1], MM

Nursing Department, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Hongying Pan, MM
Nursing Department
Sir Run Run Shaw Hospital
Zhejiang University School of Medicine
3 Qingchun East Road
Hangzhou, 310016
China
Phone: 86 13857188922
Email: 3191016@zju.edu.cn

Abstract

Background: Diabetic foot (DF) is one of the most common and serious complications of diabetes. Effective self-management by patients can delay disease progression and improve quality of life. Digital intelligent technologies have emerged as advantageous in assisting patients with chronic diseases in self-management. However, the impact of digital intelligent technologies on self-management of patients with DF remains unclear.

Objective: This systematic review aimed to determine the effects of digital intelligent interventions on self-management in patients with DF.

Methods: A systematic literature search was conducted across PubMed, Web of Science, Embase, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, Cochrane Central Register of Controlled Trial, ProQuest, China National Knowledge Internet, WanFang, China Science and Technology Journal Database, and SinoMed up to February 6, 2025, to identify eligible articles. Randomized controlled trials (RCTs) that assessed the effects of digital intelligent interventions on self-management of patients with DF were included. In total, 2 researchers independently conducted literature screening, quality assessment, and data extraction. The Cochrane Risk of Bias 2.0 tool (revised version 2019) for RCTs was used to assess the quality of the studies. A qualitative synthesis was conducted on the extracted data.

Results: In total, 1079 articles were retrieved, and 18 RCTs were included. All studies were rated as having a high risk of bias. The digital intelligent interventions in the included studies varied in forms, components, and durations. The intervention forms included WeChat (Tencent Holdings Limited; 7/18, 39%), apps (4/18, 22%), electronic platforms (3/18, 17%), mixed interventions (3/18, 17%), and smartphone thermography (1/18, 6%). The intervention components included self-management education (17/18, 94%), blood glucose and foot condition monitoring (8/18, 44%), self-management supervision and follow-up (6/18, 33%), and other components like foot risk assessment, foot care reminders, visit reminders, and remote consultations. Intervention durations ranged from 5 weeks to 12 months, with the majority (10/18, 56%) lasting 6 months. Among the 18 included studies, 17 studies (17/18, 94%) indicated that, compared with routine care, digital intelligent interventions significantly improved the self-management behaviors of patients with DF, including diabetes control, foot care behaviors, and blood glucose monitoring. Only 1 study (1/18, 6%) showed that the effects of digital intelligent interventions were not significantly different from those of routine care.

Conclusions: In this systematic review, evidence suggests that digital intelligent interventions can improve self-management behaviors and capabilities in patients with DF. However, due to the overall low quality of the included studies, current evidence should be interpreted and applied with caution. This field is still in the exploratory stage, with significant heterogeneity among different studies and a lack of consensus on intervention strategies, necessitating further exploration tailored to different populations. Future RCTs with large sample sizes and rigorous design are needed to develop high-quality evidence.

Trial Registration: PROSPERO CRD42024524473; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024524473

(J Med Internet Res 2025;27:e64400) doi:10.2196/64400

XSL•FO
RenderX

KEYWORDS

diabetic foot; self-management; digital intelligent intervention; systematic review; mobile phone

Introduction

Diabetic foot (DF) refers to foot conditions in patients with diagnosed or past diabetes mellitus (DM), manifesting as one or more of the following: peripheral neuropathy, peripheral arterial disease, infection, ulcers, Charcot neuroarthropathy, gangrene, or amputation [1]. DF is one of the most common and severe chronic complications of DM. Data from the International Working Group on the Diabetic Foot (IWGDF) indicated that 85% of patients with DF were at risk of amputation, with a mortality rate of 50% to 68% within 5 years post amputation, and a 10-year survival rate of only 24% [2]. Compared with patients with typical DM, those with DF face higher treatment costs, more frequent clinic visits, longer treatment journeys, greater physical trauma and psychological burdens, and higher mortality rates [2]. The treatment journey for DF is lengthy, as the disease persists and presents both long-term and intricate aspects, requiring patients to engage in prolonged, continuous, and steadfast self-management. In addition to regular DM self-management activities such as monitoring blood sugar, maintaining a balanced diet, and exercising, the self-management for patients with DF also includes foot monitoring and care such as daily inspection of both feet (temperature, blisters, wounds, etc), wearing well-fitting shoes, wearing seamless socks and changing them daily, washing feet daily and carefully drying them, avoiding the use of heating pads or warmers for the feet, using moisturizers for skin care, but not between the toes, and trimming nails in a straight line [3]. Studies showed that proficient self-management notably enhanced treatment adherence and slowed disease progression in patients with DF, thereby diminishing the likelihood of adverse outcomes like amputations [4]. However, due to factors such as lack of knowledge, inadequate social support, and low self-efficacy, the self-management status of patients with DF is often poor [5].

For self-management of patients with DF, traditional intervention methods often involve health education during hospitalization, followed by follow-ups through phone calls or home visits after discharge, which are time-consuming and costly, and have lower communication efficiency [6]. Digital intelligence is the integration of digitalization and intelligence, including core technology clusters such as big data, cloud computing, artificial intelligence, the internet, the Internet of Things, and 5G [7]. Digital intelligent interventions are developed based on digital intelligence, using technologies driven by digital intelligence, such as mobile apps, machine learning algorithms, virtual reality, wearable devices, and brain-computer interfaces, to provide patients with more convenient and real-time health management solutions, characterized by real-time capability, intelligence, quantifiability, visualizability, and optimizability [8]. Compared with traditional intervention methods, digital intelligent interventions can overcome the limitations of time and space, providing more efficient and personalized self-management support for patients through text, pictures, audio, and video anytime and anywhere [8]. In light of the current poor self-management levels of patients with DF, digital intelligent interventions may offer a potentially promising path forward.

At present, the applications of digital intelligent interventions in the self-management of patients with DF mainly include health monitoring, health education, and follow-up, with forms such as apps, electronic platforms, and wearable devices [9-12]. These technologies provide convenient guidance and support for health care professionals to assist patients in self-management, showing certain advantages in enhancing patient satisfaction. However, applying digital intelligent technologies in the self-management of patients with DF is still in the exploratory stage, with significant differences in intervention forms, intervention durations, and specific strategies among various studies [9-12]. Furthermore, due to methodological challenges such as small sample sizes and insufficient representativeness of the study population [9,10], the effectiveness of these intervention measures in the self-management of patients with DF remains unclear. To address this knowledge gap, there is an urgent need for a systematic review of digital intelligent interventions' impact on self-management among patients with DF. To the best of our knowledge, this is the first systematic review focusing on the impact of digital intelligent interventions on self-management in patients with DF. Such a study would help clarify the role of digital intelligent technologies in DF management and provide valuable insights for improving self-management strategies tailored to the unique needs of patients with DF.

Methods

Overview

A systematic review was carried out following the guidelines outlined in the Cochrane Handbook for Systematic Reviews of Interventions, and the reporting standards specified in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement 2020 [13]. Based on the registered protocol, we had initially intended to use RevMan 5.3 software (Cochrane) for data synthesis and meta-analysis. However, due to the great heterogeneity of the included studies and the limited availability of comparable data, a quantitative analysis could not be performed. As a result, we opted for a qualitative analysis to summarize the findings. This deviation from the original protocol is reported here to maintain transparency.

Eligibility Criteria

Eligibility criteria were meticulously devised by the PICOS (P: population, I: intervention, C: comparison, O: outcome, S: study design) framework [13]. This framework entails defining the Population (P), Intervention (I), Comparison (C), Outcome (O), and Study design (S) characteristics to ensure clarity and consistency in study selection.

XSL•FO
RenderX

Population

Patients with a confirmed diagnosis of DF. This study used the definition of DF according to the IWGDF [1], which refers to a foot condition in individuals with current or previously diagnosed DM characterized by one or more of the following: peripheral neuropathy, peripheral artery disease, infection, ulcer, neuro-osteoarthropathy, gangrene, or amputation.

Intervention

Digital intelligent interventions. As mentioned in the introduction, digital intelligent interventions refer to health management solutions that integrate both digital technologies and intelligent features. These interventions rely not only on digital tools but also on intelligent capabilities, enabling real-time, dynamic, and personalized health management through data analysis, real-time feedback, and personalized adjustments. Intervention forms include apps, electronic platforms, wearable devices, and so on. It is important to emphasize that simple digital tools, such as static and noninteractive health education videos, although possessing certain digital characteristics, do not fall within the scope of digital intelligent interventions due to the lack of intelligent analysis and personalized adjustments. Furthermore, the digital intelligent intervention was the core intervention of the intervention group.

Comparison

Nondigital intelligent interventions refer to health management approaches that either lack digital technology entirely or use digital tools without intelligent features. These interventions include conventional, nondigital methods such as face-to-face health education, printed materials, or standard counseling. They may also involve digital tools that are not equipped with intelligent functionalities, such as static, noninteractive health education videos, which provide information but do not offer real-time feedback, personalized adjustments, or data-driven insights. Essentially, these interventions lack the dynamic, real-time, and personalized capabilities that characterize digital intelligent interventions.

Outcome

Patients' self-management behavior or self-management ability. For patients with DF, self-management activities include monitoring blood sugar, a healthy diet, regular exercise, medication adherence, foot monitoring, and care. Commonly used scales to assess the self-management behavior or ability of patients with DF include the Summary of Diabetes Self-Care Activities (SDSCA) [14], Type 2 Diabetes Self-Care Scale (2-DSCS) [15], and Diabetic Foot Self-Care Questionnaire of the University of Malaga (DFSQ-UMA) [16].

Study Design

This review exclusively included randomized controlled trials (RCTs) with full text that were published in either English or Chinese.

Based on the above PICOS, there were certain inclusion and exclusion criteria for this study (Textbox 1).

Textbox 1. The inclusion and exclusion criteria for screening studies.

Inclusion criteria:

- Population: patients with a confirmed diagnosis of diabetic foot.
- Intervention: the intervention group received digital intelligent interventions.
- Comparison: the control group received nondigital intelligent interventions.
- Outcome: patients' self-management behavior or self-management ability.
- Study design: randomized controlled trial.

Exclusion criteria:

- Duplicate published articles.
- No full text available.
- Not in Chinese or English.

Search Strategy

A systematic literature search was conducted across PubMed, Web of Science, Embase, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, Cochrane Central Register of Controlled Trials, ProQuest, China National Knowledge Internet, WanFang, China Science and Technology Journal Database, and SinoMed up to February 6, 2025. A combination of subject terms and free words was used with logical operators for searching. In addition, the reference lists of included studies were scrutinized, and when necessary, studies were manually retrieved to identify other potential studies meeting the inclusion criteria. The detailed search strategies for each database are provided in Table S1 in Multimedia Appendix 1. The detailed information on the PRISMA checklist can be found in Multimedia Appendix 2.

Study Selection

References from databases underwent deduplication using NoteExpress (Beijing Aiqinhai Software Company). In total, 2 reviewers (JZ and SD) independently assessed the titles and abstracts of studies for initial screening. Subsequently, a full-text–level assessment was conducted by 2 independent reviewers (JZ and SD) to select the eligible articles for inclusion. Any disagreements were resolved through discussion or arbitration involving a third researcher (YX).

XSL•FO
RenderX

Quality Assessment

Two researchers (JZ and SD) independently conducted quality assessments using the Cochrane Risk of Bias 2.0 tool (revised version 2019, developed by Sterne et al [17]) for RCTs, which evaluates five bias domains, that are (1) randomization process, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of outcome, and (5) selection of reported results. Each domain and the overall bias were assessed as either low risk of bias, some concerns, or high risk of bias [13]. Any disagreements were resolved by seeking input from a third senior expert (YX).

Data Extraction

The data extraction process adhered to the "checklist of items for data collection form" outlined in the Cochrane Handbook and aligned with the specific PICOS principle of this review. In total, 2 reviewers (JZ and SD) independently extracted the following data from each study: first author, year published, country, included centers, participant, intervention content, intervention duration, use of intention-to-treat (ITT), outcome measures, and results of interest. Any disparities were addressed through discussion, leading to a consensus on all extracted data. In cases where additional data were needed, the authors of the included studies were contacted for further information.

Data Synthesis

The differences in the intervention methods, intervention durations, outcome assessment methods, and reporting formats across the included studies led to significant heterogeneity among the studies, hindering the quantitative synthesis of the data. Therefore, this study focused on qualitative synthesis. Tables 1 and 2 summarize the basic characteristics of the included studies. Table 3 summarizes the intervention details and the intervention outcomes of the included studies. We narratively synthesized the extracted data in terms of the population of intervention, digital intelligent intervention forms, components and durations, control group interventions, evaluation indicators and tools, and the effects of digital intelligent interventions on the self-management of patients with DF.

XSL•FO
RenderX

Table 1. The basic characteristics of the participants in the included studies.

Authors, year	Country	Participant					
		Type[a]	Number (I[b]/C[c])	Age (years), mean (SD)		Male, n (%)	
				I	C	I	C
Qin et al, 2023 [12]	Indonesia	Diabetic patients with a history of DFU[d] (Wagner grade 0)	60/60	61.7 (10.78)	59.7 (10.61)	29 (48)	29 (48)
Marques et al, 2023 [10]	Brazil	T2DM[e] medical diagnosis and foot at risk (Wagner grade 0)	20/22	—[f]	—	—	—
Firdaus et al, 2023 [9]	Malaysia	Last 2 readings of A1c[g] more than 6.5, has at least stage 1 or above risk of foot ulcer development (Wagner grade 0)	29/29	50.03 (8.03)	57.59 (7.18)	9 (31)	11 (38)
Xu et al, 2023 [18]	China	Patients with DF[h] discharged after unilateral transverse tibial bone transport, Wagner grade was 3-4	21/21	64.5 (8.6)	65.8 (8.4)	18 (86)	19 (90)
Zhang, 2022 [6]	China	Patients with DF with Wagner grade 0	46/46	65.69 (3.78)	65.69 (3.78)	25 (54)	26 (57)
Zhang et al, 2021 [19]	China	Patients with DF with Wagner grade 1-2	43/43	64.42 (4.20)	66.31 (4.56)	27 (63)	24 (56)
Chu et al, 2021 [20]	China	Patients with DF	60/60	51.49 (12.73)	52.42 (11.40)	37 (62)	33 (55)
Li et al, 2021 [21]	China	Patients at high risk of DF (Wagner grade 0)	58/58	—	—	31 (53)	35 (60)
Ye and Yu, 2021 [22]	China	Patients with DF	29/29	62.9 (9.4)	63.2 (8.4)	18 (62)	15 (52)
Qu, 2020 [23]	China	Patients with type 2 diabetic foot with Wagner grade 1-4	40/40	64.5 (5.3)	64.0 (4.5)	28 (70)	30 (75)
Xie et al, 2020 [24]	China	Patients with DF with Wagner grade 0	100/95	61.1 (7.3)	61.9 (6.8)	48 (48)	48 (51)
Xu, 2019 [25]	China	Patients with DF	46/46	54.86 (6.07)	53.15 (5.16)	24 (52)	25 (54)
Liu, 2019 [26]	China	Patients with DF (Wagner grade 0-3)	26/32	64.73 (11.85)	65.09 (11.54)	14 (44)	18 (69)
Liu et al, 2018 [27]	China	Patients with DF with Wagner grade 0	90/90	59.8 (9.6)	60.2 (9.3)	62 (62)	64 (64)
Qin et al, 2020 [28]	China	Patients with DF with Wagner grade 0	67/67	53.19 (7.85)	53.84 (8.02)	39 (58)	41 (61)
Hu, 2018 [29]	China	Patients with DF with Wagner grade 0	50/47	56.16 (10.24)	58.34 (8.97)	31 (62)	29 (62)
Bai et al, 2022 [30]	China	Patients with DF with Wagner grade 0-2	67/67	63.94 (9.03)	63.07 (8.52)	41 (61)	37 (55)
Wang et al, 2023 [31]	China	Patients with DF with Wagner grade 0-2	64/64	63.07 (9.12)	62.12 (9.05)	36 (56)	34 (53)

[a]This study applied the Wagner grading system to classify patients with diabetic foot, which is currently the most widely used grading system in clinical practice and research. Grade 0 indicates risk factors for foot ulcers but currently no ulcers are present, while grades 1 and above indicate varying degrees of ulcers or necrosis. A higher grade indicates a greater severity of the disease. The corresponding grade was extracted from the original literature or graded according to its description (the grade in parentheses indicates our grading according to the description), if the original literature had neither grade nor sufficient description to judge, the grade was not recorded.

[b]I: intervention group.

[c]C: control group.

[d]DFU: diabetic foot ulcer.

[e]T2DM: type 2 diabetes mellitus.

[f]Not available.

[g]A_{1C}: glycated hemoglobin.

[h]DF: diabetic foot.

XSL•FO
RenderX

Table 2. Interventions and results of the included studies.

Authors, year	Digital intelligent intervention, content	Control intervention, content	Results of interest (I[a] vs C[b])	*P* value
Qin et al, 2023 [12]	Smartphone thermography evaluation+personalized foot care and education	Usual care and education using a booklet	LSM[c] 67.5 (SE 1.20) versus LSM 51.9 (SE 1.52)	<.001
Marques et al, 2023 [10]	Standard nursing consultations+application use	Standard nursing consultations	Mean 4.52 (SD 0.78 versus mean 4.51 (SD 0.93)	.99
Firdaus et al, 2023 [9]	Diabetic Care Self-Management Mobile Health Application Program	Standard care	After controlling for age, monthly income, and family history, the foot care behavior of the intervention group was significantly better than that of the control group; F_1=30.374	<.001
Xu et al, 2023 [18]	"Internet+" follow-up nursing intervention (WeChat group, WeChat official account, and blood glucose monitoring app)	Health guidance was given before discharge and routine telephone follow-up	The scores of self-care skills, self-care responsibility, self-concept, and health knowledge in the intervention group were higher than those in the control group	<.01; <.01; .02; .03, respectively
Zhang, 2022 [6]	Transitional care based on WeChat platform (WeChat patient group, WeChat official account, educational knowledge in the form of text, pictures, cartoons, audio, and video)	Telephone follow-up was conducted after discharge	Mean 65.11 (SD 6.15) versus mean 54.26 (SD 7.94)	<.001
Zhang et al, 2021 [19]	Goal-oriented hospital-community linkage continuous nursing care (WeChat official account, WeChat patient group)	Health guidance was given before discharge and regular telephone follow-up after discharge	Mean 58.65 (SD 5.46) versus mean 53.44 (SD 5.81)	<.001
Chu et al, 2021 [20]	WeChat platform follow-up and continuous care (WeChat official account, WeChat patient group)	Routine telephone follow-up and outpatient education	The scores of self-management ability in all dimensions (prevention and treatment of high and low blood glucose, foot care, blood glucose monitoring, reasonable diet, medical compliance, and regular exercise) in the intervention group were higher than those in the control group	<.05
Li et al, 2021 [21]	The diabetic foot app was used for remote management after discharge	Telephone follow-up	Self-monitoring of blood glucose: n=36 (62%) versus n=18 (31%); Regular return visit: n=49 (84%) versus n=33 (57%)	<.05; <.05, respectively
Ye and Yu, 2021 [22]	Transitional care based on WeChat platform (WeChat official account, WeChat patient group)	Routine nursing follow-up	The scores of medication, exercise, diet, blood glucose measurement, and foot care in the intervention group were significantly higher than those in the control group	<.001
Qu, 2020 [23]	Routine nursing+continuing care by WeChat	Routine nursing	The proportion of self-management ability indicators in the intervention group was higher than that in the control group	<.001
Xie et al, 2020 [24]	Internet+blood glucose management (blood glucose management platform+blood glucose manager app)	Conventional blood glucose management	The scores of all dimensions (healthy diet, exercise, self-monitoring of blood glucose, monitoring of blood glucose according to medical advice, self-examination of feet, and medication) in the intervention group were significantly higher than those in the control group	<.001; <.001; <.001; <.001; .01; <.001, respectively
Xu, 2019 [25]	Extended care group combined with "Internet+" mode (WeChat, SMS, QQ [Shenzhen Tencent Computer System Co Ltd] and other internet communication tools, internet extended nursing platform)	Routine nursing	Mean 87.36 (SD 9.27) versus mean 71.68 (SD 9.68)	<.05

XSL•FO
RenderX

Authors, year	Digital intelligent intervention, content	Control intervention, content	Results of interest (I[a] vs C[b])	*P* value
Liu, 2019 [26]	Telephone follow-up after discharge+DF app management	Telephone follow-up after discharge	Blood glucose monitoring: n=15 (58%) versus n=10 (31%); Regular follow-up visits: n=16 (62%) versus n=12 (38%); Regular dressing changes: n=22 (85%) versus n=18 (56%)	.04; .07; .02, respectively
Liu et al, 2018 [27]	Transitional care based on WeChat platform (WeChat group, educational knowledge in the form of text, pictures, cartoons, and video)	Telephone follow-up	The scores of all dimensions (reasonable diet, regular exercise, compliance with doctor's advice, blood glucose monitoring, foot care, prevention of high and low blood glucose) in the intervention group were significantly higher than those in the control group	<.01; <.001; <.001; <.001; <.001; <.001, respectively
Qin et al, 2020 [28]	Transitional care based on WeChat platform (WeChat group, educational knowledge in the form of text, pictures, cartoons, and video)	Routine nursing and telephone follow-up	Mean 74.75 (SD 7.45) versus mean 54.34 (SD 6.13)	<.05
Hu, 2018 [29]	Routine education+telephone follow-up education+platform management	Routine education+telephone follow-up education	Mean 32.19 (SD 4.31) versus mean 28.14 (SD 6.12)	<.001
Bai et al, 2022 [30]	Hospital-community integrated management of hierarchical medical system by medical alliance comanagement	Routine management	The scores of all dimensions (personal care, foot care, and wearing shoes and socks) in the intervention group were significantly higher than those in the control group	<.001; <.001; .001, respectively
Wang et al, 2023 [31]	Medical alliances co-manage hierarchical medical system (Medical alliance information platform, remote consultation)	Routine nursing	The scores of all dimensions (personal care, foot care, and wearing shoes and socks) in the intervention group were significantly higher than those in the control group	.007; .003; .001, respectively

[a]I: intervention group.

[b]C: control group.

[c]LSM: least squares mean.

XSL•FO
RenderX

Table 3. Intervention details and outcomes of the included studies.

Authors, year	Participant		Digital intelligent intervention form	Digital intelligent intervention component	Control group intervention	Intervention duration	Evaluation indicator	Evaluation tool	Whether the digital intelligent intervention is effective compared with conventional intervention
	Wagner grade	Number (I[a]/C[b])							
Qin et al, 2023 [12]	0	60/60	Smartphone thermography	Foot risk assessment	Routine risk assessment	6 m	Foot care behavior	DFCB[c] questionnaire	Yes
Marques et al, 2023 [10]	0	20/22	App	Foot care education	Face-to-face education	3 m	Self-care activities with diabetes	QAD[d]	No
Firdaus et al, 2023 [9]	0	29/29	App	Foot care education, foot care reminders, and foot care behavior supervision	Face-to-face education	5 w	Foot care behavior	DFSBS[e]	Yes
Xu et al, 2023 [18]	3-4	21/21	WeChat+app	Self-management education, blood glucose monitoring, self-management behavior follow-up	Face-to-face education and telephone follow-up	6 m	General self-management ability	Self-care Agency Scale	Yes
Zhang, 2022 [6]	0	46/46	WeChat	Self-management education, self-management behavior follow-up	Telephone follow-up and education	__[f]	Diabetes self-management ability	Diabetes Self-management Behavior Scale	Yes
Zhang et al, 2021 [19]	1-2	43/43	WeChat	Self-management education, self-management behavior follow-up	Face-to-face education and telephone follow-up	6 m	Diabetes self-management ability	SDSCA[g]	Yes
Chu et al, 2021 [20]	Ungraded	60/60	WeChat	Self-management education	Face-to-face education and telephone follow-up	12 w	Diabetes self-management ability	2-DSCS[h]	Yes
Li et al, 2021 [21]	0	58/58	App	Foot care education, blood glucose, and foot monitoring	Telephone follow-up	6 m	Specific behaviors	Self-monitoring of blood glucose and regular return visit	Yes
Ye and Yu, 2021 [22]	Ungraded	29/29	WeChat	Self-management education	Telephone follow-up	6 m	Diabetes self-management ability	2-DSCS	Yes

XSL•FO
RenderX

Authors, year	Participant		Digital intelligent intervention form	Digital intelligent intervention component	Control group intervention	Intervention duration	Evaluation indicator	Evaluation tool	Whether the digital intelligent intervention is effective compared with conventional intervention
	Wagner grade	Number (I^a/C^b)							
Qu, 2020 [23]	1-4	40/40	WeChat	Self-management education, blood glucose monitoring	Face-to-face education	—	Specific behaviors	Listen to a lecture on diabetes, discuss treatment options with the doctor, self-monitoring of blood glucose, remeasuring blood glucose at the hospital, master the method of foot examination, master the method of insulin injection	Yes
Xie et al, 2020 [24]	0	100/95	Electronic platform+app	Blood glucose monitoring and blood glucose management education	Face-to-face education and telephone follow-up	12 m	Diabetes self-management ability	SDSCA	Yes
Xu, 2019 [25]	Ungraded	46/46	WeChat+electronic platform	Self-management education, disease monitoring	Face-to-face education using a booklet, telephone follow-up	3 m	Foot care behaviors	Self-management Model Questionnaire	Yes
Liu, 2019 [26]	0-3	26/32	App	Foot care education, reminders for subsequent visits	Telephone follow-up	3 m	Specific behaviors	Blood glucose monitoring, regular follow-up visits, regular dressing changes	Partly yes
Liu et al, 2018 [27]	0	90/90	WeChat	Foot care education, blood glucose, and foot monitoring	Telephone follow-up	6 m	Diabetes self-management ability	2-DSCS	Yes
Qin et al, 2020 [28]	0	67/67	WeChat	Foot care education, blood glucose, and foot monitoring	Face-to-face education and telephone follow-up	6 m	Diabetes self-management ability	Diabetes Self-management Behavior Scale	Yes
Hu, 2018 [29]	0	50/47	Electronic platform	Foot care education, blood glucose, and foot monitoring	Face-to-face education and telephone follow-up	6 m	Diabetes self-management ability	SDSCA	Yes

XSL•FO
RenderX

Authors, year	Participant		Digital intelligent intervention form	Digital intelligent intervention component	Control group intervention	Intervention duration	Evaluation indicator	Evaluation tool	Whether the digital intelligent intervention is effective compared with conventional intervention
	Wagner grade	Number (I[a]/C[b])							
Bai et al, 2022 [30]	0-2	67/67	Electronic platform	Self-management education, self-management behavior follow-up, hospital-community remote consultation and two-way referral	Face-to-face education and telephone follow-up	6 m	Foot care behaviors	DFSQ-UMA[i]	Yes
Wang et al, 2023 [31]	0-2	64/64	Electronic platform	Self-management education, self-management behavior follow-up, hospital-community remote consultation and two-way referral	Face-to-face education and telephone follow-up	6 m	Foot care behaviors	DFSQ-UMA	Yes

[a]I: intervention group.

[b]C: control group.

[c]DFCB: diabetic foot care behavior.

[d]QAD: Diabetes Self-care Activities Questionnaire.

[e]DFSBS: Diabetic Foot Selfcare Behavior Scale.

[f]Not available.

[g]SDSCA: Summary of Diabetes Self-Care Activities.

[h]2-DSCS: Type 2 Diabetes Self-care Scale.

[i]DFSQ-UMA: diabetic foot self-care questionnaire of the University of Malaga.

Results

Search Results and Study Selection

The flowchart of the literature screening process and reasons for exclusion is illustrated in Figure 1. A total of 1079 records were identified through databases and registration systems. After removing duplicates and screening the titles and abstracts, we conducted a full-text review of 82 articles. Finally, 18 RCTs were included.

XSL•FO
RenderX

Figure 1. PRISMA flow chart for study selection. RCT: randomized controlled trial.

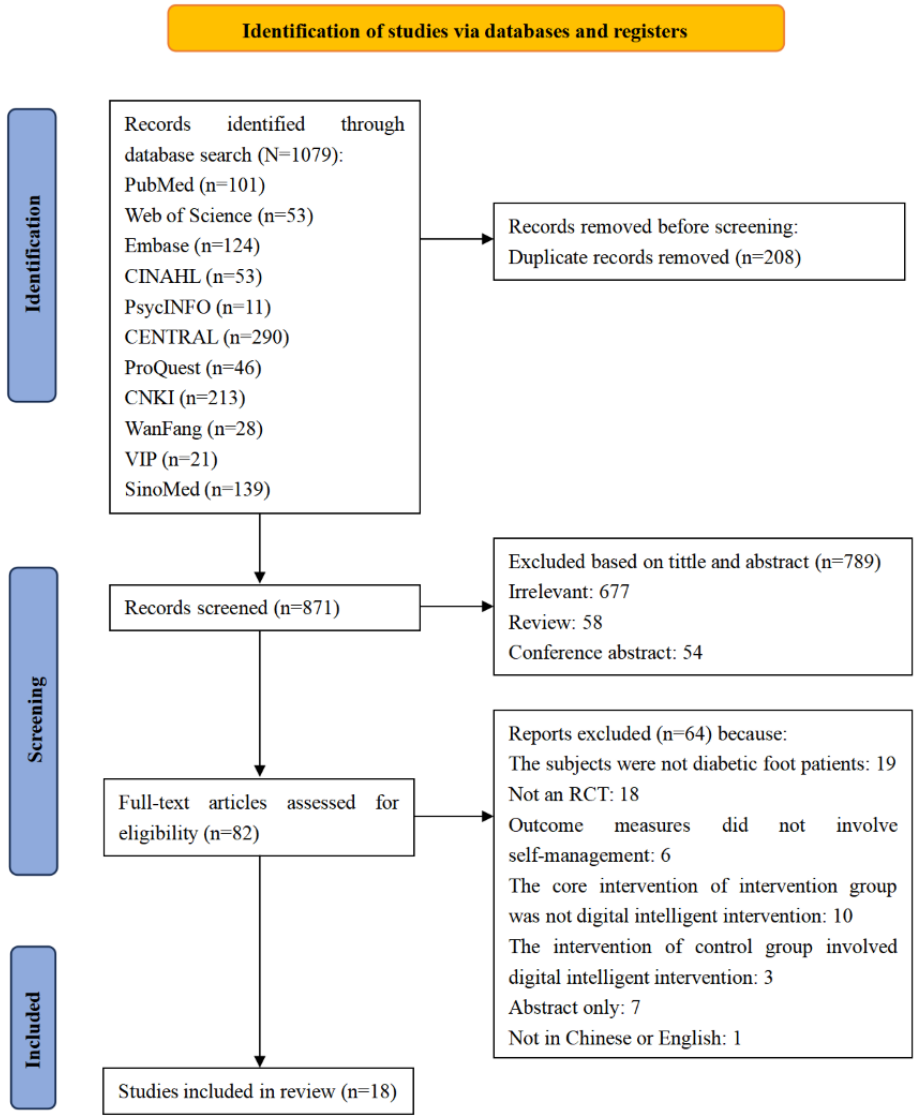

Study Characteristics

Table 1 summarizes the basic characteristics of the participants in the included studies. Table 2 summarizes the interventions and results of the included studies. Details of the included studies are provided in Table S2 in Multimedia Appendix 1. The publication years were from 2018 to 2023 [6,9,10,12,18-31], and 78% (14/18) of the studies were published in the past 5 years (2020 and beyond) [6,9,10,12,18-24,28,30,31]. Except for 1 study [12] conducted at 2 centers, all other studies [6,9,10,18-31] were single-center studies. Studies were conducted in Indonesia [12], Brazil [10], Malaysia [9], and China [6,18-31], with sample sizes ranging from 42 to 195. The age of participants ranged from 50.03 to 66.31, and the proportion of males ranged from 31% to 90%. The intervention forms for the intervention group included WeChat (Tencent Holdings Limited) [6,19,20,22,23,27,28], app [9,10,21,26], electronic platform [29-31], smartphone thermography [12], and mixed interventions (WeChat+app [18], electronic platform+app [24], and WeChat+electronic platform [25]). The control group primarily received standard or usual care interventions focused on diabetes management and foot care. The interventions included health education, health monitoring and follow-up through paper leaflets [12,25], face-to-face communication or examination [9,10,12,18-20,23-25,28-31], and phone calls [6,18-22,24-31]. The intervention duration ranged from 12 months [24] and 6 months [12,18,19,21,22,27-31] to 3 months [10,20,25,26] and 5 weeks [9]. The measurement tools for outcome indicators included Self-care Agency Scale [18], SDSCA [6,19,24,28,29], 2-DSCS [20,22,27], Diabetes Self-care Activities Questionnaire (QAD) [10], Diabetes Self-management Behavior Scale [6,28], DFSQ-UMA [30,31], diabetic foot care behavior (DFCB) [12], Diabetic Foot Selfcare Behavior Scale (DFSBS) [9], and Self-management Model Questionnaire [25]. In addition, some studies [21,23,26] evaluated the self-management behavior of patients by observing their behaviors such as blood glucose monitoring and regular return visits. In terms of the reporting form of the results, some studies [6,10,12,19,25,28,29] compared the total score of the scale between the intervention group and the control group, some [9,18,20,22,24,27,30,31] compared the scores of each dimension of the scale between the 2 groups, and others [21,23,26] compared the number and proportion of people who adopted self-management behaviors between the 2 groups.

XSL•FO
RenderX

Study Quality

All of the 18 studies were assessed as having an overall high risk of bias (Figure 2 [6,9,10,12,18-31]). In the randomization process domain, 2 studies [6,26] lacked details on randomization, and the results of baseline between-group comparisons in 1 study [9] suggested potential differences, raising some concerns. One study [21] grouped participants based on odd and even discharge orders, deemed high risk due to inadequate concealment of allocation sequence before grouping. In the domain of deviations from the intended interventions, only 1 study [10] implemented double-blinding, 1 study [29] did not blind participants, 1 study [12] did not blind participants or researchers, and information on blinding was missing in the remaining studies [6,9,18-31]. All studies lacked sufficient information to determine whether intervention deviations from the expected were due to the trial environment. 14 studies [6,12,18-26,28,30,31] conducted ITT analyses, while four studies [9,10,27,29] excluded some randomized participants in their analyses. In the domain of missing outcome data, 5 studies [9,10,12,27,29] had a considerable amount of missing outcome data. The reasons for participant attrition in 2 studies [9,12] were health-related, in 2 studies [10,29] were unrelated to health, and in 1 study [27], the reasons for attrition could not be determined due to insufficient information. In the domain of measurement of the outcome, 3 studies [21,23,26] did not use standardized measurement scales, and 2 studies [6,23] did not specify the measurement timing. In 2 studies [12,29], self-reported scales were used for measurement without blinding participants, so the assessors were aware of the interventions received by the participants. For the 12 studies [6,9,18-20,22,24,25,27,28,30,31], the lack of information made it impossible to determine whether assessors were aware of the interventions received by the participants. Since self-reported scales were used for measurement in the studies [6,9,12,18-20,22,24,25,27-31], which inherently carried subjectivity, knowing the interventions received by the participants could potentially influence the outcomes. In the selection of the reported result domain, 7 studies [18,20,22,24,27,30,31] were rated as high risk due to not reporting the total scale score and total score comparison results between groups.

Figure 2. Risk of bias summary: review authors' judgments about the risk of bias item for each included study [6,9,10,12,18-31].

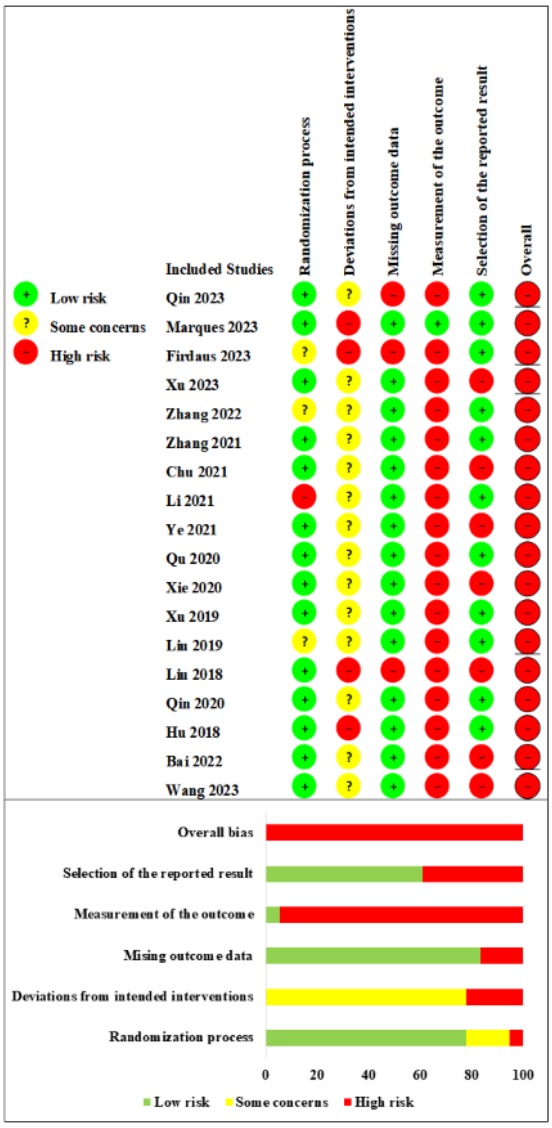

XSL•FO
RenderX

Population of Intervention

The intervention details and the intervention outcomes of the included studies are shown in Table 3. The intervention population consisted of patients with DF. In the included studies, most of the participants were classified as Wagner grade 0 [6,9,10,12,21,24,27-29] (9 out of 18 studies, 50%), while some studies included patients with Wagner grade 0-2 [30,31] or Wagner grade 0-3 [26], and others included patients with Wagner grade 1-2 [19], Wagner grade 1-4 [23], or Wagner grade 3-4 [18]. Wagner grade 0 indicates the absence of foot ulcers but the presence of risk factors for ulcer development (such as a history of foot ulcers, peripheral neuropathy, or peripheral arterial disease). Wagner grade 1 and above indicates the presence of foot ulcers or necrosis of varying severity. Some studies [20,22,25] did not apply the Wagner classification system to assess the severity of foot disease in the patients (3 out of 18 studies, 17%).

Digital Intelligent Intervention Forms, Components, and Durations

Digital intelligent intervention forms, components, and durations of the included studies (N=18) are shown in Multimedia Appendix 3. The forms of digital intelligent interventions included WeChat (7 studies [6,19,20,22,23,27,28], 39%), app (4 studies [9,10,21,26], 22%), electronic platforms (3 studies [29-31], 17%), mixed interventions (WeChat+app [18], WeChat+electronic platform [25], app+electronic platform [24], 3 studies, 17%), and smartphone thermography (1 study [12], 6%). The components of digital intelligent interventions included self-management education (17 studies [6,9,10,18-31], 94%), blood glucose and foot condition monitoring (8 studies [18,21,23-25,27-29], 44%), self-management supervision and follow-up (6 studies [6,9,18,19,30,31], 33%), foot risk assessment (1 study [12], 6%), foot care reminders (1 study [9], 6%), reminders for subsequent visits (1 study [26], 6%), and hospital-community remote consultation and 2-way referral (2 studies [30,31], 11%). The duration of the interventions ranged from 12 months (1 study [24], 6%), 6 months (10 studies [12,18,19,21,22,27-31], 56%), 3 months (4 studies [10,20,25,26], 22%) to 5 weeks (1 study [9], 6%), with 2 studies [6,23] (11%) not specifying the duration of the intervention.

Control Group Interventions

The control group interventions in the included studies consisted of routine care, including routine risk assessment (1 study [12]), face-to-face education (12 studies [9,10,18-20,23-25,28-31]), and telephone follow-up (14 studies [6,18-22,24-31]). The intervention content primarily involved face-to-face self-management health education provided by nurses at discharge, as well as follow-up and self-management guidance via telephone by nurses after discharge.

Evaluation Indicators and Tools

The evaluation indicators in the included studies encompassed four categories: (1) foot care behavior (5 studies [9,12,25,30,31], 28%), diabetes self-management ability (9 studies [6,10,19,20,22,24,27-29], 50%), general self-management ability (1 study [18], 6%), and specific behaviors (3 studies [21,23,26], 17%). The tools used to assess foot care behavior included the DFCB questionnaire [12], DFSBS [9], Self-management Model Questionnaire [25], and DFSQ-UMA [30,31]. The assessment content focused on whether patients regularly checked their feet, the condition of their foot care, and whether they selected appropriate footwear. The tools used to evaluate diabetes self-management ability included the QAD [10], Diabetes Self-management Behavior Scale [6,28], SDSCA [19,24,29], and 2-DSCS [20,22,27]. The assessment content primarily covered blood glucose monitoring, proper diet, regular exercise, foot care, and medication adherence. The tool used to assess self-management ability was the Self-care Agency Scale [18], which evaluated a patient's general disease self-management ability, including 4 dimensions, that is, self-care skills, self-care responsibility, self-concept, and health knowledge level. Specific behaviors reflecting the patient's self-management status, consisted of blood glucose monitoring, foot examination, and regular follow-up visits in the included studies [21,23,26].

Effects of Digital Intelligent Interventions on Self-Management of Patients With DF

Among the 18 included studies, 17 studies [6,9,12,18-31] (94%) indicated that, compared with routine care, digital intelligent interventions significantly improved the self-management levels of patients with DF. However, 1 study [10] (6%) showed that the effects of digital intelligent interventions were not significantly different from those of routine care. Based on the results of the included studies, when using WeChat for intervention, a 6-month intervention for patients with Wagner grade 0 [27,28] or 1-2 [19], or a 3-month [20] or 6-month [22] intervention for patients without classification, can significantly improve patients' self-management ability in diabetes. When using an electronic platform for intervention, a 6-month intervention for patients with Wagner grade 0 [20] or 0-2 [30,31] can significantly enhance patients' diabetes self-management or foot care levels. When using an app for intervention, a 5-week [9] or 6-month [21] intervention for patients with Wagner grade 0 can improve patients' foot care behaviors, self-monitoring of blood glucose, and regular return visits; a 3-month [26] intervention for patients with Wagner grade 0-3 can improve behaviors of blood glucose monitoring and regular dressing changes, but had no significant effect on regular follow-up visits; while a 3-month [10] intervention for patients with Wagner grade 0 cannot significantly improve their diabetes self-management ability. Using smartphone thermography [10] for a 6-month intervention for patients with Wagner grade 0 can significantly improve foot care behaviors. Using WeChat combined with app [18] for a 6-month intervention for patients with Wagner grades 3-4 can significantly improve general self-management ability. Using an electronic platform combined with app [24] for a 12-month intervention for patients with Wagner grade 0 can significantly improve their diabetes self-management ability. Using WeChat combined with electronic platform [25] for a 3-month intervention for unclassified patients can significantly improve foot care behaviors.

XSL•FO
RenderX

Discussion

Principal Findings

A total of 18 RCTs were included in this systematic review, all of which were assessed as high-risk studies. Methodological deficiencies in the included studies primarily included unclear details of randomization, lack of allocation concealment before grouping, absence of blinding, failure to conduct ITT analysis, high rates of missing data and loss to follow-up, absence of standardized assessment tools, unclear assessment timing, potential awareness of intervention status among assessment personnel, and failure to report total scores of scales. Due to the significant heterogeneity among studies, we conducted a qualitative synthesis of the included studies. The results showed that overall, digital intelligent interventions improved self-management ability among patients with DF. However, the target populations, intervention forms, intervention durations, and assessment tools vary across different studies, and there is currently no standardized intervention paradigm in this field.

The most noteworthy point is that, among the 18 studies, only 1 study [10] did not show a significant effect of digital intelligent intervention, and the reasons for this phenomenon need to be analyzed in depth. In that study [10], the participants were patients with Wagner grade 0, the intervention form was an app (used for foot care education), the intervention duration was 3 months, and QAD was used to assess the patients' diabetes self-management ability. In another study [9], where the app intervention (used for foot care education, foot care reminders, and foot care behavior supervision) also targeted patients with Wagner grade 0, but the duration was 5 weeks, the DFSBS assessment showed significant improvements in patients' foot care behaviors. This seems to illustrate the importance of the intervention components, suggesting that even with a shorter intervention duration, focusing on more aspects of self-management (not just education, but also reminders and supervision) can lead to better outcomes. In 1 study [21], a 6-month intervention with the app (used for foot care education, blood glucose, and foot monitoring) in patients with Wagner grade 0, improvements in self-monitoring of blood glucose and regular visits were observed. Similarly, for app interventions targeting patients with Wagner grade 0, the outcomes are better when the intervention included more components and lasted longer. When we broaden our perspective beyond just app interventions and patients with Wagner grade 0, and compare this study [10] with others, we find that regardless of the intervention form (WeChat, App, electronic platforms, or mixed interventions), the intervention duration (shorter or longer), and target population (Wagner grade 0 or Wagner grade ≥1), it seems that as long as the intervention is not limited to health education alone (but also includes reminders, supervision, or follow-ups, etc), the significant effects of digital intelligent interventions can be observed.

The deeper reasons for these results need to be discussed. Health education does have an effect on self-management behaviors of patients with diabetes, as existing studies [32,33] have confirmed. However, our results suggest that, compared with conventional face-to-face health education, health education delivered solely through digital intelligent means (such as apps) does not show more significant effects than traditional interventions. A possible reason is that, although apps can overcome the boundaries of time and space, allowing patients to receive health education anytime and anywhere, such interventions do not guarantee that patients will actively engage with the app for health education. In such cases, the intervention may be less effective than the conventional 1-time "forced" face-to-face health education during discharge. In other studies, the inclusion of intervention components, such as reminders, supervision, and follow-up, helped to enhance patients' adherence to using digital intelligent tools for health education. At the same time, these interventions directly reminded and supervised patients in implementing self-management behaviors, enabling the real-time, dynamic, and personalized features of digital intelligent interventions to be fully used, resulting in significantly better outcomes compared with conventional interventions.

This provides us with an important insight that it is not only the functions of digital intelligent tools themselves that are important, but it seems even more crucial to focus on how to attract users to continuously use these tools in order to fully leverage the advantages of digital intelligent technology. In the included studies, the methods used to remind and encourage patients to use the tools included setting up learning contracts [9] and real-time reminders in WeChat groups [18,21,27,28]. While these measures achieved good results, there was a lack of (or no report on) efforts made to enhance the attractiveness of the digital intelligent tools themselves. Factors that influence people's use of digital intelligent tools have been explored in previous studies. One study [34] explored the technology adoption intention of users of smart wearable health technologies and identified the key factors affecting user technology adoption. The research [34] found that the more the user's experience of smart health technology aligns with personal needs, the more positive the user's emotional response, which in turn promotes higher user engagement. In addition, the operability and interaction design of the technology (such as the intuitiveness and ease of use of the interface) are also important factors in determining user engagement [34]. Based on this, the following strategies can be adopted to improve user engagement with digital intelligent tools. First, the users' experience should be personalized. Personalization can strengthen users' emotional attachment to the technology because they feel that the tool is more closely aligned with their lives, which in turn increases engagement and ultimately promotes usage [35]. Specific measures include (1) offering customized features, that is, developing features that cater to specific health goals and needs; for example, allow users to customize exercise plans, track specific health indicators (eg, blood sugar levels), or set personalized reminders (eg, foot check-ups). (2) Developing adaptive interfaces [36], that is, providing adaptive interfaces that change based on user preferences, usage patterns, and health progress. This can also include adjusting the complexity of information based on the user's knowledge level and health literacy. Second, the focus should be on intuitive and user-friendly design. A user-friendly design increases the likelihood of long-term engagement because users feel more confident and comfortable when using the device, encouraging

XSL•FO
RenderX

continued usage [37]. Specific measures include (1) simple and clear interface (ie, ensure the interface is simple, navigation is easy, and information is clear and straightforward; make sure buttons, menus, and options are simple and intuitive, reducing the complexity of setup procedures); (2) user testing and feedback (ie, continuously improve the design through user feedback; conduct usability tests to identify potential barriers in the interface and simplify them). Third, emotional engagement and motivational incentives are also crucial. Emotional engagement, driven by motivation and rewards, can strengthen the bond between users and the device, leading to higher usage frequency and adherence [38]. Specific measures include (1) gamification and reward mechanisms [39], that is, introduce gamification elements such as achievement badges, points, or milestones to recognize users' health progress. These rewards can be linked to self-management behaviors (eg, monitoring blood sugar and performing foot care) or improvements in health indicators (eg, blood sugar levels and foot condition). (2) Social sharing features [40], that is, allow users to share their progress with friends or on social media. Social support can significantly increase motivation and promote engagement. Community-driven challenges (eg, foot care goals) can also enhance a sense of belonging and competition. Finally, continuous support and updates for the tool are essential. Ensure that the device is regularly updated, adding new features and improvements based on user feedback. Software updates should not only address issues but also enhance overall functionality, improving the value of the user experience [40]. Ongoing support and updates increase the reliability of the technology and the commitment to the users, thus encouraging long-term adoption.

The final aspect worth discussing is the regional distribution of the studies. The 18 studies we included were all conducted in limited-income countries (Indonesia, Brazil, Malaysia, and China), with 15 of them being conducted in China. One of the reasons for the high number of studies from China could be the inclusion of 4 Chinese language databases in our search. However, this is not the only or primary reason, as we also conducted systematic searches across 7 authoritative English language databases. In addition, after carefully reviewing the search terms, we consider our search to be reasonably comprehensive. Therefore, it is more likely that the regional distribution of the studies reflects the current global distribution of the application of digital intelligent technologies in self-management of patients with DF. Although digital intelligent technologies have been rapidly developed in high-income countries, this study focuses specifically on their application in the self-management of patients with DF. Compared with high-income countries, limited-income countries, due to relatively lower economic levels and uneven distribution of medical resources, face significant challenges in diabetes management and early prevention of DF, resulting in a higher number of patients with DF and consequently, a greater focus on self-management research within this patient group [41]. Similarly, the possible reason for the large number of studies from China is that it has the world's largest population and the highest number of patients with diabetes [42]. In addition, due to the vast territory of China and the uneven distribution of health care resources [43], managing diabetes and preventing DF pose significant challenges, resulting in a higher number of patients with DF [44] and, consequently, more research on their self-management.

Limitations

First, the overall quality of the included studies is relatively low, with several methodological flaws that may affect the reliability of the results. Second, due to significant heterogeneity among the included studies, a meta-analysis could not be conducted. Although most included studies indicated that digital intelligent interventions were effective, the lack of quantitatively synthesized results hinders the determination of the optimal intervention strategy, such as the best form of intervention or the ideal duration of the intervention. Finally, although we made efforts to ensure the comprehensiveness of the search terms (keywords and their synonyms) and databases (4 Chinese databases and 7 English databases), the majority of the included studies were conducted in China, which may introduce cultural and regional biases and limit the generalizability of the findings.

Conclusions

In this systematic review, evidence suggested that digital intelligent interventions can improve self-management behaviors and capabilities in patients with DF. However, due to the overall low quality of the included studies, current evidence should be interpreted and applied with caution. At present, this field is still in the exploratory stage, with significant heterogeneity among different studies and a lack of consensus on intervention strategies, necessitating further exploration tailored to different populations. Future RCTs with large sample sizes and rigorous design are needed to develop high-quality evidence and determine the optimal intervention strategies. In addition, in order to fully leverage the advantages of digital intelligent technology, the human-centered design of digital intelligent tools, including personalization, ease of use, and intuitiveness, should also be given attention, alongside their health management functions.

Acknowledgments

This research was funded by Zhejiang University "Double First-Class" Construction Nursing Advantage and Characteristic Discipline Scientific Research Cultivation Fund Project (HL2024010), Zhejiang University School of Medicine Affiliated Sir Run Run Shaw Hospital Nursing Clinical Research Project (2024HLKY05), Zhejiang Province Major Social Welfare Program Project (2023C03191 and 2022C03134), Science and Technology Innovation 2030- "New Generation of Artificial Intelligence" Major Project (2022ZD0160703), and Zhejiang Province Medical and Health Technology Program Project (2022KY836 and 2024KY1142).

XSL•FO
RenderX

Data Availability

Datasets generated and analyzed during the study are available from the corresponding author upon reasonable request.

Authors' Contributions

JZ and HP conceptualized the study. JZ, SD, and YX contributed to data collection and analyses. JZ contributed to writing the first draft of the manuscript. All authors contributed to original data collection and execution of studies, data interpretation, and editing and approval of the final manuscript. HP is the guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy of each database and details of the included studies.
[DOCX File , 50 KB - jmir_v27i1e64400_app1.docx]

Multimedia Appendix 2
PRISMA checklist.
[PDF File (Adobe PDF File), 88 KB - jmir_v27i1e64400_app2.pdf]

Multimedia Appendix 3
Digital intelligent intervention forms, components, and durations of the included studies.
[PNG File , 94 KB - jmir_v27i1e64400_app3.png]

References

1. van Netten JJ, Bus SA, Apelqvist J, Chen P, Chuter V, Fitridge R, International Working Group on the Diabetic Foot. Definitions and criteria for diabetes-related foot disease (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3654. [doi: 10.1002/dmrr.3654] [Medline: 37186781]
2. Bakker K, Apelqvist J, Lipsky BA, Van Netten JJ, International Working Group on the Diabetic Foot. The 2015 IWGDF guidance documents on prevention and management of foot problems in diabetes: development of an evidence-based global consensus. Diabetes Metab Res Rev 2016 Jan;32 Suppl 1(S1):2-6. [doi: 10.1002/dmrr.2694] [Medline: 26409930]
3. Schaper NC, van Netten JJ, Apelqvist J, Bus SA, Fitridge R, Game F, IWGDF Editorial Board. Practical guidelines on the prevention and management of diabetes-related foot disease (IWGDF 2023 update). Diabetes Metab Res Rev 2024 Mar;40(3):e3657 [FREE Full text] [doi: 10.1002/dmrr.3657] [Medline: 37243927]
4. Schaper NC, Van Netten JJ, Apelqvist J, Lipsky BA, Bakker K, International Working Group on the Diabetic Foot (IWGDF). Prevention and management of foot problems in diabetes: a summary guidance for daily practice 2015, based on the IWGDF guidance documents. Diabetes Res Clin Pract 2017 Feb;124:84-92. [doi: 10.1016/j.diabres.2016.12.007] [Medline: 28119194]
5. Costa IG, Tregunno D, Camargo-Plazas P. I cannot afford off-loading boots: perceptions of socioeconomic factors influencing engagement in self-management of diabetic foot ulcer. ANS Adv Nurs Sci 2020;43(4):322-337. [doi: 10.1097/ANS.0000000000000328] [Medline: 32956088]
6. Zhang F. Analysis of the effects of continuous nursing based on WeChat platform on self-management behavior and self-efficacy of patients with diabetic foot. Harbin Medical Journal 2022;42,:102-103.
7. Zhang W, Zhang C, Cao L, Liang F, Xie W, Tao L, et al. Application of digital-intelligence technology in the processing of Chinese materia medica. Front Pharmacol 2023;14:1208055 [FREE Full text] [doi: 10.3389/fphar.2023.1208055] [Medline: 37693890]
8. Wang Y, Xiong X, Liu M, Yang Y, Liu S, Chen H, et al. Application progress of digital and intelligent intervention technology in the mental disorders of patients with coronary heart disease. Chinese Journal of Nursing, Review 2024;59(16):1956-1961 [FREE Full text]
9. Firdaus MKZH, Jittanoon P, Boonyasopun U, Che Hasan MK. The effect of mHealth program on behavior modification and health outcomes among patients with diabetes: a randomized controlled trial study. Belitung Nurs J 2023;9(5):437-447 [FREE Full text] [doi: 10.33546/bnj.2664] [Medline: 37901368]
10. Marques ADB, Moreira TMM, Mourão LF, Florêncio RS, Cestari VRF, Garces TS, et al. Mobile application for adhering to diabetic foot self-care: randomized controlled clinical trial. Comput Inform Nurs 2023 Nov 01;41(11):877-883. [doi: 10.1097/CIN.0000000000001024] [Medline: 37235486]
11. Moulaei K, Malek M, Sheikhtaheri A. A smart wearable device for monitoring and self-management of diabetic foot: a proof of concept study. Int J Med Inform 2021 Feb;146:104343. [doi: 10.1016/j.ijmedinf.2020.104343] [Medline: 33260090]

XSL•FO
RenderX

12. Qin Q, Oe M, Nakagami G, Kashiwabara K, Sugama J, Sanada H, et al. The effectiveness of a thermography-driven preventive foot care protocol on the recurrence of diabetic foot ulcers in low-medical resource settings: An open-labeled randomized controlled trial. Int J Nurs Stud 2023 Oct;146:104571. [doi: 10.1016/j.ijnurstu.2023.104571] [Medline: 37586286]
13. Higgins JPT, Chandler TJ, Cumpston J, Li T, Page MJ, Welch VA. Cochrane Handbook for Systematic Reviews of Interventions version 6.4. UK: John Wiley & Sons; 2023.
14. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
15. Lee NP, Fisher WP. Evaluation of the diabetes self care scale: an illustration of the Rasch model of measurement. J Nurs Meas 2002;10(3):171-187. [doi: 10.1891/jnum.10.3.171.52558] [Medline: 12885144]
16. Navarro-Flores E, Morales-Asencio JM, Cervera-Marín JA, Labajos-Manzanares MT, Gijon-Nogueron G. Development, validation and psychometric analysis of the diabetic foot self-care questionnaire of the University of Malaga, Spain (DFSQ-UMA). J Tissue Viability 2015 Feb;24(1):24-34. [doi: 10.1016/j.jtv.2014.11.001] [Medline: 25523014]
17. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
18. Xu G, Liu M, Ju Y, Kong Y, Miao Y. Application of Internet + follow-up management in patients with diabetic foot after tibial transverse bone transport. Tianjin Journal of Nursing 2023;31:699-702.
19. Zhang A, Xueqin H, Chunhong Y, Qiong W. The application effect of goal-oriented hospital-community linkage continuous nursing in patients with diabetic foot. Proceeding of Clinical Medicine 2021;30:945-948. [doi: 10.16047/j.cnki.cn14-1300/r.2021.12.020]
20. Chu, L, Wang J, You Q, Fang Z, Dou L. Continuous nursing pathway for diabetes foot based on WeChat platform. Journal of Clinical Medicine in Practice 2021;25:90-93 [FREE Full text]
21. Li K, Feng Y, Sun M, Cui Y, Gong X, Duan W, et al. The self-influence of telemedicine system in the management of diabetic high risk foot. Electronic Journal of Foot and Ankle Surgery 2021;8:27-31.
22. Ye Q, Yu Y. Study on the application effect of continuous care on WeChat platform in diabetic foot patients' self-management behavior. China & Foreign Medical Treatment 2021;40:172-174. [doi: 10.16662/j.cnki.1674-0742.2021.04.172]
23. Qu Y. Effects of transitional care based on WeChat APP on self-management and self-efficacy of patients with diabetic foot. The Medical Forum 2020;24:5236-5237.
24. Xie X, Wang X, Zhao D, Dong H, Zheng Z, Liu S, et al. Application of internet plus blood glucose management mode in the education and management of diabetic foot patients. China Medical Herald 2020;17:157-160.
25. Xu, J. Effect of extended care group combined with internet + on hospital discharge nursing for diabetic foot. Journal of Clinical Medicine in Practice 2019;23:111-117.
26. Liu Y. Study on the advantages of managing patients with diabetic foot based on telemedicine system. University of Jinan 2019.
27. Liu J, Liu, X, Meng G, Tang J. Study on self management behavior of diabetes foot in rural middle aged and elderly based on WeChat's continuous nursing. Medical Education Research and Practice 2018;26:109-112. [doi: 10.13555/j.cnki.c.m.e.2018.01.029]
28. Tan X, Liu J, Wu J, Hong H, Liu C. Effects of continuous nursing based on WeChat platform on self-management behavior and self-efficacy of patients with diabetic foot. International Journal of Nursing 2020;39:954-957. [doi: 10.3760/cma.j.issn.1673-4351.2020.05.061]
29. Hu S. Study on the effect of intelligent hierarchical management platform on self-management ability of patients with high-risk factors of diabetic foot. Ji'nan University 2018.
30. Bai J, Di R, Chen J. Application of medical alliance co-management hierarchical diagnosis and treatment model in hospital-community comprehensive management of diabetic foot patients. Chinese Journal of Modern Nursing 2022;28:1138-1142. [doi: 10.3760/cma.j.cn115682-20210926-04387]
31. Wang Y, Zhang M, Lyu F, Zheng X, Hu H, Lin T. Observation on the effect of medical alliance co-management hierarchical diagnosis and treatment model applied to diabetic foot patients. Medical Innovation of China 2023;20:175-179.
32. Diriba DC, Leung DYP, Suen LKP. Nurse-led self-management education and support programme on self-management behaviour and quality of life among adults with type 2 diabetes: A pilot randomized controlled trial. Int J Nurs Pract 2024 Apr;30(2):e13245. [doi: 10.1111/ijn.13245] [Medline: 38351899]
33. Othman N, Al-Otaibi T, Halim MA, Said T, Elserwy N, Mahmoud F, et al. Effect of repeated structured diabetes education on lifestyle knowledge and self-care diabetes management in kidney transplant patients with posttransplant diabetes. Exp Clin Transplant 2024 Jan;22(Suppl 1):128-140 [FREE Full text] [doi: 10.6002/ect.MESOT2023.O31] [Medline: 38385386]
34. Liu X. The influence mechanism of smart wearable health technologies user technology adoption intention: based on the mediating role of user engagement. 2024 Presented at: IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAI); 2024 July 05-07; Beijing, China p. 296-301. [doi: 10.1109/bdai62182.2024.10692759]
35. Rana A, Sanner S, Bouadjenek MR, Di Carlantonio R, Farmaner G. User experience and the role of personalization in critiquing-based conversational recommendation. ACM Trans. Web 2024 Oct 08;18(4):1-21. [doi: 10.1145/3597499]

XSL•FO
RenderX

36. Wang W, Khalajzadeh H, Grundy J, Madugalla A, McIntosh J, Obie HO. Adaptive user interfaces in systems targeting chronic disease: a systematic literature review. User Model User-Adap Inter 2023 Dec 18;34(3):853-920. [doi: 10.1007/s11257-023-09384-9]
37. Najafi A, Letchumanan V. Design and implementation of surepkay; a user friendly multimedia content management system for digital signage. 2011 Presented at: 3rd International Conference on Computer Technology and Development (ICCTD 2011); 2011 Nov 25-27; Chengdu, China p. 805-810. [doi: 10.1115/1.859919.paper132]
38. Esmaeilzadeh P. The influence of gamification and information technology identity on postadoption behaviors of health and fitness app users: empirical study in the United States. JMIR Serious Games 2021 Jul 05;9(3):e28282 [FREE Full text] [doi: 10.2196/28282] [Medline: 34812736]
39. Bitrián P, Buil I, Catalán S. Enhancing user engagement: The role of gamification in mobile apps. Journal of Business Research 2021 Aug;132:170-185. [doi: 10.1016/j.jbusres.2021.04.028]
40. Jia C, Qi H. Users' health information sharing behavior in social media: an integrated model. Humanit Soc Sci Commun 2024 Dec 04;11(1):1647. [doi: 10.1057/s41599-024-04188-7]
41. Swaminathan N, Awuah WA, Bharadwaj HR, Roy S, Ferreira T, Adebusoye FT, et al. Early intervention and care for diabetic foot ulcers in low and middle income countries: addressing challenges and exploring future strategies: a narrative review. Health Sci Rep 2024 May;7(5):e2075 [FREE Full text] [doi: 10.1002/hsr2.2075] [Medline: 38690005]
42. Xu Y, Lu J, Li M, Wang T, Wang K, Cao Q, et al. Diabetes in China part 1: epidemiology and risk factors. The Lancet Public Health 2024 Dec;9(12):e1089-e1097. [doi: 10.1016/s2468-2667(24)00250-0]
43. Zhang T, Xu Y, Ren J, Sun L, Liu C. Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. Int J Equity Health 2017 Mar 03;16(1):42 [FREE Full text] [doi: 10.1186/s12939-017-0543-9] [Medline: 28253876]
44. Jiang YF, Wang X, Xia L, Fu X, Xu Z, Ran X, et al. A cohort study of diabetic patients and diabetic foot ulceration patients in China. Wound Repair Regen 2015;23(2):222-230. [doi: 10.1111/wrr.12263] [Medline: 25682850]

Abbreviations

DF: diabetic foot
DFCB: diabetic foot care behavior
DFSBS: Diabetic Foot Selfcare Behavior Scale
DFSQ-UMA: Diabetic Foot Self-Care Questionnaire of the University of Malaga
DM: diabetes mellitus
ITT: intention-to-treat
IWGDF: International Working Group on the Diabetic Foot
PICOS: P: population, I: intervention, C: comparison, O: outcome, S: study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QAD: Diabetes Self-care Activities Questionnaire
RCT: randomized controlled trial
SDSCA: Summary of Diabetes Self-Care Activities
2-DSCS: Type 2 Diabetes Self-Care Scale

Edited by A Coristine; submitted 16.07.24; peer-reviewed by M Peeples, X Jiang, L Hess; comments to author 15.11.24; revised version received 02.01.25; accepted 27.02.25; published 25.03.25.

Please cite as:
Zhou J, Ding S, Xu Y, Pan H
Effects of Digital Intelligent Interventions on Self-Management of Patients With Diabetic Foot: Systematic Review
J Med Internet Res 2025;27:e64400
URL: https://www.jmir.org/2025/1/e64400
doi:10.2196/64400
PMID:

©Jinyan Zhou, Shanni Ding, Yihong Xu, Hongying Pan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Physiological Sensors Equipped in Wearable Devices for Management of Long COVID Persisting Symptoms: Scoping Review

Shikha Kukreti[1], BDS, MPH, PhD; Meng-Ting Lu[1,2], MPH; Chun-Yin Yeh[3], RN, PhD; Nai-Ying Ko[1,4], RN, PhD

[1]Department of Nursing, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[2]Deparment of Public Health, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[3]Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan City, Taiwan

[4]International Doctoral Program in Nursing, Deaprtment of Nursing, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

Corresponding Author:
Nai-Ying Ko, RN, PhD
Department of Nursing, College of Medicine
National Cheng Kung University
No.1, Dasyue Rd, East District
Tainan City, 701
Taiwan
Phone: 886 062353535
Email: konaiying@gmail.com

Abstract

Background: Wearable technology has evolved in managing COVID-19, offering early monitoring of key physiological parameters. However, the role of wearables in tracking and managing long COVID is less understood and requires further exploration of their potential.

Objective: This review assessed the application and effectiveness of wearable devices in managing long COVID symptoms, focusing on commonly used sensors and their potential for improving long-term patient care.

Methods: A literature search was conducted across databases including PubMed, Embase, Web of Science, and Cochrane Central, adhering to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines. The search was updated regularly throughout 2024. Abstract and full-text screening and selection were facilitated using Rayyan software developed by Qatar Computing Research Institute. Quality appraisal was conducted using the Joanna Briggs Institute (JBI) critical appraisal tool to ensure the methodological rigor of the included studies. Data were extracted on study characteristics, wearable devices, sensors used, and monitored physiological parameters, and the results were synthesized in a narrative format.

Results: A total of 1186 articles were identified, and after duplicate removal and screening, 15 studies were initially included, with 11 studies meeting the criteria for final data synthesis. The included studies varied in design, ranging from observational to interventional trials, and involved sample sizes from 3 to 17,667 participants across different countries. In total, 10 different wearable devices were used to monitor long COVID symptoms, capturing key metrics such as heart rate variability, body temperature, sleep, and physical activity. Smartwatches were the most used wearable devices and fitness trackers with electrocardiography and photoplethysmography sensors were used to monitor heart rate, oxygen saturation, and respiratory rate. Of the 10 devices, 4 were Food and Drug Administration–approved, emphasizing the reliability and validation of the physiological data collected. Studies were primarily conducted in the United States and Europe, reflecting significant regional research interest in wearable technology for long COVID management.

Conclusions: This review highlights the potential of wearable technology in providing continuous and personalized monitoring for long COVID patients. Although wearables show promise in tracking persistent symptoms, further research is needed to improve usability, validate long-term efficacy, and enhance patient engagement.

(J Med Internet Res 2025;27:e69506) doi:10.2196/69506

XSL•FO
RenderX

KEYWORDS

wearable devices; long COVID; physiological sensors; review; COVID; COVID-19

Introduction

Wearable devices have emerged as essential tools in the detection and management of COVID-19 due to their ability to continuously monitor key physiological metrics such as heart rate, oxygen saturation, and body temperature in real time [1,2]. These devices provide a proactive approach to identifying early signs of COVID-19, enabling timely interventions even before severe symptoms appear. Unlike acute COVID-19, where symptoms typically manifest and resolve within a few weeks, long COVID involves persistent and varied symptoms that can last for months or even years after the initial infection [3]. This complexity requires not only the ability to detect acute symptoms but also to monitor and manage chronic and fluctuating conditions over extended periods. Long COVID is a multifaceted condition that requires not only a deep understanding of its complex pathophysiology but also the integration of advanced digital technologies like wearables to effectively monitor, analyze, and manage its persistent and fluctuating symptoms over time. For patients with long COVID, wearable technology offers invaluable insight into persistent and fluctuating symptoms, helping to track and manage the condition over extended periods [4]. Despite their potential, there is still a need to refine and enhance the effectiveness of wearable technology, ensuring consistency in how different devices capture and analyze data for long COVID's complex health effects.

Although a few studies have investigated the use of wearable devices for long COVID detecting symptoms, findings have varied due to differences in device types, sensor configurations, and analytical methods [5,6]. This variation underscores the need for a comprehensive review to synthesize existing knowledge, evaluate the effectiveness of different wearable technologies, and identify best practices for their application in long COVID management. A thorough review can help bridge the gaps in this research and guide future innovations in wearable technology for more effective monitoring and management of long COVID.

Wearable devices equipped with sensors for monitoring heart rate, respiratory rate, oxygen saturation, and body temperature have shown promise in detecting anomalies associated with COVID-19 [7]. Previous studies have explored the role of wearable devices in the detection of COVID-19 symptoms, demonstrating their potential to provide early warnings and facilitate timely interventions [7-9]. Research has demonstrated that changes in these metrics can indicate the onset of infection before severe symptoms manifest [10]. However, while these devices have been effective in detecting acute COVID-19 symptoms, managing long COVID presents a more complex challenge, as it involves tracking persistent, fluctuating symptoms over extended periods, requiring continuous and adaptable monitoring strategies.

Understanding physiological sensors equipped in wearable devices is crucial because it enables continuous and precise monitoring of post-COVID complex symptoms, providing insights into their patterns and severity over time. Effective management of long COVID relies on identifying which sensors are most used and how they perform in capturing relevant physiological data. Current literature reveals a variety of sensors used in wearable devices, such as those for monitoring heart rate variability, oxygen saturation, and respiratory rate [11]. Given the urgency to address long COVID effectively, a comprehensive scoping review is essential to consolidate existing research, evaluate the efficacy of different sensors, and inform the development of more targeted and effective monitoring solutions. The primary objective of this scoping review is to evaluate and summarize existing research on the application of wearable technology for monitoring and managing persistent long COVID symptoms. This review seeks to identify the various wearable devices and sensors used, the specific physiological parameters they monitor, and the impact of this monitoring on the management of long COVID.

Methods

Overview

This scoping review was conducted in adherence with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines [12]. Ethical approval was not required for this Scoping review. Refer to Multimedia Appendix 1 for the PRISMA checklist.

Search Strategy and Selection Criteria

A comprehensive literature search was conducted across electronic databases, including PubMed (MEDLINE), Embase, Cochrane Central Register of Controlled Trials (known as CENTRAL), and the international clinical trials registry platform. An additional search for datasets was conducted by SK using Google Scholar. The reference lists of included articles and reviews were manually searched to identify additional relevant studies for full-text screening. To ensure as current a review as possible, we repeated the above searches on June 15, 2024, and July 10, 2024, and again on November 15, 2024, before final analysis. The search strategy combined keywords related to long COVID (eg, "Long COVID," "Post-Acute Sequelae of SARS-CoV-2 Infection," "PASC"), wearable technology (eg, "wearable devices," "smartwatches," "fitness trackers"), and physiological monitoring (eg, "heart rate variability," "blood oxygen saturation," "respiratory rate;" Multimedia Appendix 2).

Selection Process and Data Extraction

This scoping review focused on studies that involved the use of wearable technology for monitoring or managing long COVID symptoms, specifically those published in peer-reviewed journals. It included research that reported on specific physiological parameters monitored by wearable devices, such as heart rate, respiratory rate, and oxygen saturation. The review excluded studies that did not specifically focus on long COVID,

XSL•FO
RenderX

those that did not involve wearable technology, and those that were not available in full text, including review articles, conference abstracts, or incomplete articles. The initial abstract screening was conducted using Rayyan [13], a web-based tool for systematic reviews. The title and abstracts were independently reviewed by 2 reviewers (SK and MTL), and a third reviewer (NYK or CYL) acted as an arbitrator in case of disagreements between the 2 other reviewers.

Data were extracted from the included studies using a JBI standardized form. Information on study characteristics (eg, author, year, country), type of wearable device used, sensors used, physiological parameters monitored, and the outcomes related to long COVID management were recorded. A narrative synthesis was conducted to summarize the findings, with a focus on the types of wearable devices and sensors used, the physiological parameters monitored, and the impact on long COVID symptom management.

Results

Overview

A total of 1186 articles were identified from 3 databases, 1 registry, and Google Scholar. Following the removal of duplicates and the initial eligibility screening, 16 studies were included in the review (Figure 1). These 16 studies underwent a JBI quality assessment by the 2 authors, and out of these, only 11 met the criteria for data synthesis. Therefore, a total of 11 studies were included in the final review.

Figure 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 flow diagram showing identification of studies from search databases.

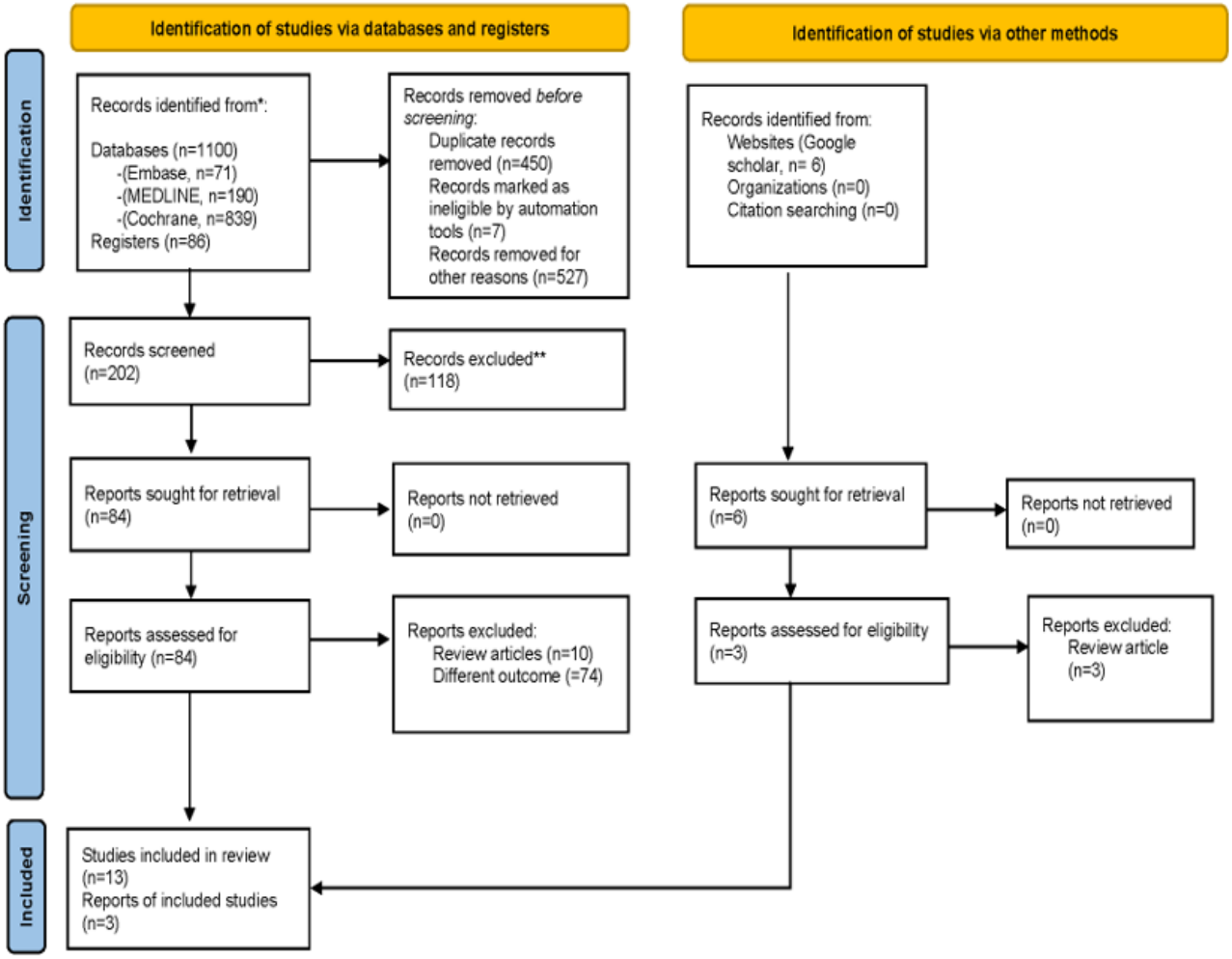

Study Characteristics

A total of 11 studies met the inclusion criteria and were included in the review [5,6,14-22]. These studies varied in design, ranging from observational studies to interventional trials, and were conducted across different countries (Table 1). The studies' sample sizes ranged widely from as few as 3 participants [20] to as many as 17,667 participants [21].

XSL•FO
RenderX

Table 1. Characteristics of the included 11 studies.

Number	Author and year	Country and region	Study design	Sample size	Wearable technology	Regulatory status
1	Mekhael et al (2022) [5]	New Orleans, United States	Prospective observational study	710	Biostrap	FDA-approved[a]
2	Lonini et al (2021) [14]	Illinois, United States	Pilot study	28	Suprasternal notch device	N/A[b]
3	Mekhael et al (2024) [19]	United States	Observational	122	Biostrap	FDA-approved
4	Xue et al (2022) [15]	Ireland	Cross-sectional	108	Shimmer device, Finapres New Observation and Verification Apparatus	Published study by Schutte et al, 2004
5	Laguarta-Val et al (2024) [16]	Madrid, Spain	A nonrandomized parallel controlled trial	29	Polar Ignite 2 device	N/A
6	Corrado et al (2024) [17]	United Kingdom	Pilot	13	Polar H10 chest Strap or Fitbit Charge 5 smartwatch	FDA-approved (Fitbit)
7	Corrêa et al (2023) [18]	Brazil	Cross-sectional	73	Samsung mobile phone	N/A
8	Romaszko-Wojtowicz et al (2022) [20]	Poland	Case study	3	Aidmed	American National Standard for Ambulatory Electrocardiographs
9	Kerling et al (2024) [6]	Germany	Prospective, randomized, parallel-group	62	Garmin	FDA-approved
10	Stewart et al (2024) [21]	United Kingdom	Longitudinal case-control	17,667	Fitbit	FDA-approved
11	Radin et al (2024) [22]	United States	Prospective observational study	553	Fitbit	FDA-approved

[a]FDA: Food and Drug Administration.

[b]Not applicable.

Wearable Devices and Sensors

The wearable devices identified in the studies included a mix of commercially available devices (eg, smartwatches, fitness trackers) and research-specific devices designed for health monitoring. In total, 10 different wearable devices were used to monitor long COVID symptoms, capturing key metrics such as heart rate variability (HRV), body temperature, sleep, and physical activity, providing a comprehensive understanding of the physiological effects of the condition. The most used devices were smartwatches (eg, Biostrap, Fitbit, Garmin) and fitness trackers, which are widely accessible and user-friendly.

In the included studies, seven different types of sensors were used, including photoplethysmography (PPG), electrocardiography (ECG), accelerometers, temperature sensors, gyroscopes, near-infrared spectroscopy, and electromyography (Table 2). These sensors were used across multiple devices to monitor key physiological metrics such as heart rate, heart rate variability, peripheral capillary oxygen saturation (SpO_2), respiratory rate, temperature, physical activity, and sleep-related movements [5,6,14-22]. Sensors integrated into these devices were primarily focused on monitoring cardiovascular and respiratory parameters.

XSL•FO
RenderX

Table 2. Study wearable device details of included 11 studies.

Number	Author	Sensor	Physiological characteristics
1	Mekhael et al (2022) [5]	• PPG[a] • Accelerometer	• Heart rate • Heart rate variability • Respiratory rate • SpO2[b] • Movement
2	Lonini et al (2021) [14]	• Adam sensor	• Body motions • Vibrations induced by sounds produced by heartbeats, coughing, or breathing
3	Mekhael et al (2024) [19]	• PPG • Accelerometer	• Heart rate • Heart rate variability • Respiratory rate • SpO2 • Movement
4	Xue et al (2022) [15]	• Digital artery photoplethysmography • Near-infrared spectroscopy • Electromyography	• Heart rate • Heart rate variability • SpO2 • Fatigue
5	Laguarta-Val et al (2024) [16]	• ECG[c] • temperature • accelerometer	• Heart rate • Temperature • Physical activity
6	Corrado et al (2024) [17]	• ECG • PPG • Accelerometer or gyroscope	• Heart rate variability • Heart rate • Sleep rotation movement
7	Corrêa et al (2023) [18]	• Accelerometers • Gyroscopes (inertia sensors)	• Physical activity or mobility
8	Romaszko-Wojtowicz et al (2022) [20]	• ECG • Temperature • PPG	• Heart rate • Temperature • Respiratory rate
9	Kerling et al (2024) [6]	• ECG • Accelerometer	• Physical activity • Heart rate • Sleep
10	Stewart et al (2024) [21]	• ECG • Accelerometers	• Heart rate • Heart rate variability • Physical activity and step count
11	Radin et al (2024) [22]	• ECG • Accelerometers	• Resting heart rate • Physical activity and step count • Sleep

[a]PPG: photoplethysmography.

[b]SpO_2: peripheral capillary oxygen saturation.

[c]ECG: electrocardiography.

ECG was used in 6 out of the 10 studies [14,16,22] and was used to monitor heart rate, and heart rate variability, and to detect arrhythmias and other cardiac abnormalities.

PPG was used in 3 studies [5,17,19] and was also used to measure heart rate, SpO_2 levels, and respiratory rate.

Accelerometers were integrated in 6 studies [6,17] and used to monitor physical activity, sleep patterns, and mobility. They are critical for measuring movements and providing insights into daily activity levels and sleep quality.

Some studies reported the use of more advanced sensors, such as Adam sensors [14] which are used to monitor body motion as it works on vibrations.

XSL•FO
RenderX

Food and Drug Administration Approval and Device Validation

Devices like Fitbit [17,21,22], Polar Ignite [16,17], Garmin [17], and Biostrap [5,19] got Food and Drug Administration (FDA) clearance for ECG functionalities, ensuring regulatory oversight and validation of the accuracy, safety, and reliability of the data collected (Table 1). In addition, the medical device Aidmed is approved by the American National Standard for Ambulatory Electrocardiographs.

Physiological Parameters Monitored

The physiological parameters monitored by these wearable devices included heart rate, heart rate variability, blood oxygen saturation, respiratory rate, skin temperature, and physical activity levels (Figure S1 in Multimedia Appendix 2). These parameters were chosen for their relevance to the symptoms experienced by individuals with long COVID, such as fatigue, autonomic dysfunction, and respiratory distress.

Country of Origin and Regional Focus

The studies were predominantly conducted in the United States and Europe.

Five studies originated in the United States, reflecting a significant research interest in wearable technology for long COVID in this region. The studies by Mekhael et al [5], Radin et al [22], and Lonini et al [14] demonstrated extensive use of wearable devices like Biostrap (Sontakey), Fitbit (Google), and the Suprasternal notch device (Sonica Health**)** [19].

European studies (United Kingdom, Spain, Poland, Germany, and Ireland) were also included, indicating a broad European engagement in the field. The study by Xue et al [15] in Ireland used the Shimmer device (Shimmer Sensing) and Finapres NOVA (Finapres Medical Systems), while the United Kingdom–based study by Corrado et al [17] used the Polar H10 chest strap (manufactured by Polar Electro) and Fitbit Charge 5 smartwatch. Similarly, the study by Kerling et al [6] in Germany used a Garmin device (Garmin Ltd).

Measurement Approaches in Long COVID Studies

Devices like Biostrap and Fitbit tracked sleep stages, HRV, and physical activity, helping to identify sleep disruptions and reduced physical endurance in patients (Multimedia Appendix 3). Tools such as the Suprasternal notch device and Shimmer captured critical data on altered respiration, cough frequency, and neurocardiovascular responses, shedding light on the multisystem impacts of long COVID. Interventions using devices like the Polar H10 and Garmin provided insights into HRV biofeedback and energy expenditure, aiding rehabilitation and symptom relief. Overall, these devices offer valuable tools for understanding and managing long COVID symptoms, enabling tailored care and long-term monitoring for affected individuals.

Discussion

Principal Findings

This scoping review offers a comprehensive evaluation of the existing literature on the use of wearable health devices for monitoring and managing long COVID symptoms. The findings highlight the potential of wearable devices to revolutionize the management of this complex and persistent of long COVID symptoms by continuous physiological monitoring. This review identifies a wide range of wearable devices used in published studies, including commercially available smartwatches, fitness trackers, and research-specific medical devices. The use of 10 different wearable devices in detecting long COVID symptoms highlights the potential of these technologies in early symptom identification [5,6,14-22]. Sensors such as PPG, ECG, and accelerometers were commonly used to monitor cardiovascular and respiratory functions, reflecting the prevalence of these symptoms in long COVID patients. These devices continuously track key physiological parameters like heart rate, HRV, SpO_2, and sleep patterns, allowing for the early detection of subtle health changes and proactive management of symptoms. The widespread use of accelerometers also emphasizes the growing importance of physical activity and sleep monitoring in managing chronic conditions like long COVID.

One notable finding in this review is that only a limited number of wearable devices have received FDA approval for specific health monitoring features. For instance, Fitbit's ECG app, cleared by the FDA in 2020, is accessible on models such as the Sense, Sense 2, and Charge 5 [23]. Garmin has also secured FDA clearance for its ECG app, now available on compatible models like the Venu 2 Plus and Epix Pro series [24]. The inclusion of FDA-approved devices brings a level of regulatory oversight, which is critical in clinical settings where wearable data may influence patient care decisions. FDA-approved wearables are validated for safety and effectiveness, offering health care providers more confidence in the data collected, especially when managing complex conditions like long COVID. In contrast, studies that do not emphasize the use of FDA-approved devices may face challenges in translating their findings into clinical practice. Comparing these studies underscores the need for standardized protocols and further validation studies to assess the long-term efficacy of wearable devices, ensuring their integration into routine health care for monitoring and managing long COVID symptoms.

This review also highlights issues related to user adherence and engagement. The success of wearable technology in managing chronic conditions like long COVID depends not only on the accuracy of the data collected but also on the willingness of users to consistently wear the devices and engage with the data. A few studies reported issues with user adherence, which could be attributed to factors such as device comfort, battery life, and the user interface of the accompanying apps [19,25]. Despite advancements in wearable technology, a significant gap exists in ensuring the seamless, secure, and standardized handling of the vast amounts of personal health data collected by these devices. In agreement with Smith et al [26], data interoperability and secure sharing across platforms, underscore the need for improved systems for data management and privacy to effectively integrate wearable technologies into clinical practice [26]. These issues highlight the need for more user-centered design approaches in the development of wearable devices, ensuring that they are not only functional but also comfortable and easy to use over extended periods.

XSL•FO
RenderX

In terms of the geographic distribution of the studies, the majority were conducted in the United States [5,14,19] and Europe [15-17,20,21], reflecting a significant research interest in these regions. This regional focus may be influenced by the availability of resources and funding for research on wearable technology, as well as the higher prevalence of long COVID in these areas due to the earlier and more widespread outbreaks of COVID-19. However, there is a need for more research in other regions, particularly in low- and middle-income countries, where the burden of long COVID may be significant but less well-documented. Furthermore, this review also found that while wearable devices offer valuable insights, the complexity and vast symptomatology of long COVID remain underexplored. Included studies focused on specific symptoms, with limited attention to common issues like fatigue, which is less frequently measured by these devices. This highlights the limitations of wearable technology in addressing the full spectrum of long COVID symptoms, underscoring the need for more comprehensive research.

This review provides a distinct perspective from typical COVID-19 wearable device studies by focusing on the long-term management of long COVID, rather than acute symptom detection. The key difference lies in the chronic nature of long COVID, where persistent and fluctuating symptoms, such as fatigue, cardiovascular issues, and sleep disturbances, require continuous, long-term monitoring. In contrast, COVID-19 reviews often emphasize short-term monitoring for acute symptoms, such as changes in respiratory rate and oxygen saturation, during the active phase of infection. Moreover, this review addresses unique challenges like user adherence and device validation for chronic use, which are less of a concern in acute COVID-19 monitoring.

There are several limitations to this review that must be acknowledged. One key limitation is the reliance on small sample sizes in some studies, which may reduce the statistical power and the ability to draw broader conclusions. In addition, the use of self-reported data in certain studies could introduce bias and affect the accuracy of the findings. Another limitation is the study design, with most studies lacking a longitudinal approach, which is crucial for assessing the long-term effectiveness of wearable technology in tracking long COVID symptoms. Furthermore, the number of studies available for review remains limited, highlighting the urgent need for more high-quality research to address the evolving challenges of long COVID management through wearable technology.

Conclusion

This review highlights the significant potential of wearable technology in managing long COVID by leveraging sensors like ECG, PPG, and accelerometers to monitor crucial physiological metrics such as heart rate, oxygen saturation, and physical activity. Wearable devices provide real-time insights into persistent and fluctuating symptoms, offering a more personalized approach to patient care. The advancements in device accuracy and the increasing availability of FDA-approved options further reinforce their importance in clinical use. Furthermore, enhancing the usability of these devices to improve user engagement will enable broader adoption and sustained monitoring over extended periods. Ultimately, wearable technology stands out as a promising tool to support individualized care and improve outcomes for long COVID patients, shaping the future of health care monitoring.

Future Implications and Requirements

For wearable technology to fully realize its potential in managing long COVID, several future implications are required. Standardization of protocols for sensor use and data interpretation is critical. This will allow for greater consistency across studies and ensure that health care providers can rely on the data generated by wearable devices. Wearability plays a crucial role in adherence, and future innovations should focus on compact, lightweight designs and hypoallergenic materials to enhance comfort and minimize user burden. Equally important is ensuring that sensors are accurate, durable, and capable of sustained performance in diverse real-world conditions. User-centered design is essential for improving adherence to wearable devices, ensuring that they are comfortable, easy to use, and integrated seamlessly into patients' lives. Longitudinal studies are also needed to assess the sustained benefits of wearable technology in managing long COVID symptoms and improving health outcomes over time.

Building on the current research, the future applications of wearable health devices extend beyond long COVID management to other infectious diseases, such as influenza or emerging infectious diseases with similar symptomatology. Wearable devices that continuously monitor vital signs such as body temperature, heart rate, and respiratory functions could facilitate early illness detection and proactive management by identifying subtle physiological changes indicative of infection. Such applications are particularly relevant for influenza, where symptoms frequently overlap with long COVID, including fatigue, respiratory issues, and fever. In addition, integrating wearable data into public health systems could provide real-time insights into disease trends and hotspots, supporting more responsive and informed health interventions. Future efforts should also explore the integration of advanced data analytics and artificial intelligence to interpret sensor data more effectively, paving the way for personalized health care solutions.

This review illustrates that wearables have the potential to reshape not only individual patient care but also population-level disease management, positioning them as vital tools in addressing both chronic and acute public health challenges. Overall, to fully integrate wearable devices into long COVID management, a concerted effort is required to address these challenges. By developing standardized methods, improving device usability, and expanding research across diverse populations, wearable technology could become a cornerstone in the effective long-term management of long COVID.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA extension for scoping reviews checklist (PRISMA-ScR).
[DOCX File , 86 KB - jmir_v27i1e69506_app1.docx]

Multimedia Appendix 2
Literature search for conducting scoping review: documentation form.
[DOCX File , 1013 KB - jmir_v27i1e69506_app2.docx]

Multimedia Appendix 3
Study outcomes and results.
[DOCX File , 23 KB - jmir_v27i1e69506_app3.docx]

References

1. Strain T, Wijndaele K, Dempsey PC, Sharp SJ, Pearce M, Jeon J, et al. Wearable-device-measured physical activity and future health risk. Nat Med 2020;26(9):1385-1391 [FREE Full text] [doi: 10.1038/s41591-020-1012-3] [Medline: 32807930]
2. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR mHealth uHealth 2020;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
3. Ely EW, Brown LM, Fineberg HV, National Academies of Sciences, Engineering, Medicine Committee on Examining the Working Definition for Long Covid. Long covid defined. N Engl J Med 2024 Nov 07;391(18):1746-1753. [doi: 10.1056/NEJMsb2408466] [Medline: 39083764]
4. Babu M, Lautman Z, Lin X, Sobota MH, Snyder MP. Wearable devices: implications for precision medicine and the future of health care. Annu Rev Med 2024;75:401-415 [FREE Full text] [doi: 10.1146/annurev-med-052422-020437] [Medline: 37983384]
5. Mekhael M, Lim CH, El Hajjar AH, Noujaim C, Pottle C, Makan N, et al. Studying the effect of long COVID-19 infection on sleep quality using wearable health devices: observational study. J Med Internet Res 2022;24(7):e38000 [FREE Full text] [doi: 10.2196/38000] [Medline: 35731968]
6. Kerling A, Beyer S, Dirks M, Scharbau M, Hennemann A, Dopfer-Jablonka A, et al. Effects of a randomized-controlled and online-supported physical activity intervention on exercise capacity, fatigue and health related quality of life in patients with post-COVID-19 syndrome. BMC Sports Sci Med Rehabil 2024;16(1):33 [FREE Full text] [doi: 10.1186/s13102-024-00817-5] [Medline: 38308307]
7. Channa A, Popescu N, Skibinska J, Burget R. The rise of wearable devices during the COVID-19 pandemic: a systematic review. Sensors (Basel) 2021;21(17):5787 [FREE Full text] [doi: 10.3390/s21175787] [Medline: 34502679]
8. Cheong SHR, Ng YJX, Lau Y, Lau ST. Wearable technology for early detection of COVID-19: a systematic scoping review. Prev Med 2022;162:107170 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107170] [Medline: 35878707]
9. Weizman Y, Tan AM, Fuss FK. Use of wearable technology to enhance response to the coronavirus (COVID-19) pandemic. Public Health 2020;185:221-222 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.048] [Medline: 32668337]
10. Miller DJ, Capodilupo JV, Lastella M, Sargent C, Roach GD, Lee VH, et al. Analyzing changes in respiratory rate to predict the risk of COVID-19 infection. PLoS One 2020;15(12):e0243693 [FREE Full text] [doi: 10.1371/journal.pone.0243693] [Medline: 33301493]
11. Syversen A, Dosis A, Jayne D, Zhang Z. Wearable sensors as a preoperative assessment tool: a review. Sensors (Basel) 2024;24(2):482 [FREE Full text] [doi: 10.3390/s24020482] [Medline: 38257579]
12. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
13. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
14. Lonini L, Shawen N, Botonis O, Fanton M, Jayaraman C, Mummidisetty CK, et al. Rapid screening of physiological changes associated with COVID-19 using soft-wearables and structured activities: a pilot study. IEEE J Transl Eng Health Med 2021;9:4900311 [FREE Full text] [doi: 10.1109/JTEHM.2021.3058841] [Medline: 33665044]
15. Xue F, Monaghan A, Jennings G, Byrne L, Foran T, Duggan E, et al. A novel methodology for the synchronous collection and multimodal visualization of continuous neurocardiovascular and neuromuscular physiological data in adults with long COVID. Sensors (Basel) 2022;22(5):1758 [FREE Full text] [doi: 10.3390/s22051758] [Medline: 35270905]
16. Laguarta-Val S, Varillas-Delgado D, Lizcano-Álvarez Á, Molero-Sánchez A, Melian-Ortiz A, Cano-de-la-Cuerda R, et al. Effects of aerobic exercise therapy through nordic walking program in lactate concentrations, fatigue and quality-of-life in

XSL•FO
RenderX

patients with long-COVID syndrome: a non-randomized parallel controlled trial. J Clin Med 2024;13(4):1035 [FREE Full text] [doi: 10.3390/jcm13041035] [Medline: 38398348]
17. Corrado J, Iftekhar N, Halpin S, Li M, Tarrant R, Grimaldi J. HEART rate variability biofeedback for LOng COVID dysautonomia (HEARTLOC): results of a feasibility study. Adv Rehabil Sci Pract. 2024 2024:13. [doi: 10.1101/2023.09.09.23295208]
18. Corrêa BDC, Santos EGR, Belgamo A, Pinto GHL, Xavier SS, Silva CC, et al. Smartphone-based evaluation of static balance and mobility in long-lasting COVID-19 patients. Front Neurol 2023;14:1277408 [FREE Full text] [doi: 10.3389/fneur.2023.1277408] [Medline: 38148981]
19. Mekhael M, Ho C, Noujaim C, Assaf A, Younes H, El Hajjar AH, et al. Compliance challenges in a longitudinal COVID-19 cohort using wearables for continuous monitoring: observational study. J Med Internet Res 2023;25:e43134 [FREE Full text] [doi: 10.2196/43134] [Medline: 36763647]
20. Romaszko-Wojtowicz A, Maksymowicz S, Jarynowski A, Jaśkiewicz Ł, Czekaj Ł, Doboszyńska A. Telemonitoring in long-COVID patients-preliminary findings. Int J Environ Res Public Health 2022;19(9):5268 [FREE Full text] [doi: 10.3390/ijerph19095268] [Medline: 35564663]
21. Stewart C, Ranjan Y, Conde P, Sun S, Zhang Y, Rashid Z, et al. Physiological presentation and risk factors of long COVID in the UK using smartphones and wearable devices: a longitudinal, citizen science, case–control study. The Lancet Digital Health 2024;6(9):e640-e650. [doi: 10.1016/s2589-7500(24)00140-7]
22. Radin JM, Vogel JM, Delgado F, Coughlin E, Gadaleta M, Pandit JA, et al. Long-term changes in wearable sensor data in people with and without Long Covid. NPJ Digit Med 2024;7(1):246 [FREE Full text] [doi: 10.1038/s41746-024-01238-x] [Medline: 39271927]
23. FDA-cleared wearables and their health benefits. MUO. URL: https://www.makeuseof.com/fda-cleared-wearables-health-benefits/ [accessed 2024-12-02]
24. Understand your body better with the new ECG app from Garmin. Garmin. URL: https://www.garmin.com/en-US/newsroom/press-release/sports-fitness/understand-your-body-better-with-the-new-ecg-app-from-garmin/ [accessed 2024-12-02]
25. Chan HY. Wearable devices for Long COVID: prospects, challenges and options. Asian Bioeth Rev 2024;16(4):757-769. [doi: 10.1007/s41649-023-00272-3] [Medline: 39398460]
26. Smith AA, Li R, Tse ZTH. Reshaping healthcare with wearable biosensors. Sci Rep 2023;13(1):4998 [FREE Full text] [doi: 10.1038/s41598-022-26951-z] [Medline: 36973262]

Abbreviations

ECG: electrocardiography
FDA: Food and Drug Administration
HRV: heart rate variability
JBI: Joanna Briggs Institute
PPG: photoplethysmography
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SpO2: peripheral capillary oxygen saturation

Edited by A Mavragani; submitted 02.12.24; peer-reviewed by E Popa, S Sharma, R Singh; comments to author 26.12.24; revised version received 03.01.25; accepted 07.01.25; published 26.03.25.

Please cite as:
Kukreti S, Lu MT, Yeh CY, Ko NY
Physiological Sensors Equipped in Wearable Devices for Management of Long COVID Persisting Symptoms: Scoping Review
J Med Internet Res 2025;27:e69506
URL: https://www.jmir.org/2025/1/e69506
doi:10.2196/69506
PMID:40137051

©Shikha Kukreti, Meng-Ting Lu, Chun-Yin Yeh, Nai-Ying Ko. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Efficacy of eHealth Interventions for Hemodialysis Patients: Systematic Review and Meta-Analysis

Xu-Hua Zhou[1,2], MSc; Hui Chen[1,2], MSN; Weiwei Yang[1,2], BSc; Li Wang[3], BSc; Lin Chen[1,2], BSc; Ying Zhu[1,2], BSc; Yingjun Zhang[1,2], MSN; Mei Shi[1,2], BSc; Qin Zhang[1,2], BSc

[1]Hemodialysis Center, Department of Nephrology, West China Hospital, Sichuan University, Chengdu, China

[2]West China School of Nursing, Sichuan University, Chengdu, China

[3]Department of Gastroenterology, The First Affiliated Hospital of Chengdu Medical College, Chengdu, China

Corresponding Author:
Lin Chen, BSc
Hemodialysis Center
Department of Nephrology
West China Hospital, Sichuan University
37 Guo-xue-xiang Lane, Wuhou district, Chengdu, Sichuan, PR China
Chengdu, 610041
China
Phone: 86 18980601205
Email: clhxxuetou@163.com

Abstract

Background: Within hemodialysis patient populations, eHealth interventions have been considered as an alternative and complementary option to routine care services. However, the efficacy of eHealth interventions for hemodialysis patients remains poorly understood owing to a lack of rigorous quantitative evidence synthesis.

Objective: This meta-analysis aimed to evaluate the efficacy of eHealth interventions in improving quality of life, treatment adherence, and psychological outcomes (anxiety and depression) among hemodialysis patients. In addition, the study sought to identify specific intervention components and methodological quality associated with enhanced quality of life and health outcomes in this population.

Methods: A comprehensive search was performed across PubMed, Web of Science, Embase, CINAHL, Cochrane Library, PsycINFO, China National Knowledge Infrastructure, WanFang, China Science and Technology Journal Database, and China BioMedical Literature Database databases from their inception to September 7, 2024. Randomized controlled trials on eHealth interventions for hemodialysis patients published in English or Chinese were included. Critical appraisal was carried out independently by 2 reviewers to assess the bias risk of the studies included. Quantitative synthesis of the outcomes of interest was conducted using a random-effects model. The quality of evidence for the outcomes was evaluated following the Grading of Recommendations, Assessment, Development, and Evaluation approach.

Results: A total of 17 randomized controlled trials involving 1728 participants were included in this meta-analysis out of 5741 articles identified in the initial database search and additional search references. In the 17 studies, 8 kinds of eHealth intervention delivery formats were used, including text messages, telephone sessions, video, network platforms, social media, computers, websites, and mobile apps. The majority of research studies used a single form of eHealth intervention, and 7 studies adopted a combined approach of 2 or more eHealth technologies. The duration of eHealth interventions demonstrated substantial variability across studies, spanning from 4 weeks to 12 months, of which 3 months was the most common. A total of 14 (82%) studies were considered to have “some concern” about selection bias. In addition, 15 (88%) trials were classified as having a “high risk” of performance and detection bias, and all trials were judged to be at “low risk” of attrition and reporting bias. The pooled results revealed a significant difference between the eHealth interventions and control groups on quality of life (standardized mean difference [SMD]=0.87, 95 % CI 0.38 to 1.37, low certainty evidence), treatment adherence (SMD=1.11, 95 % CI 0.30 to 1.91, moderate certainty evidence), anxiety (SMD=–2.11, 95 % CI –3.25 to –0.97, moderate certainty evidence), and depression (SMD=–2.46, 95 % CI –3.68 to –1.25, moderate certainty evidence).

XSL•FO
RenderX

Conclusions: eHealth interventions could be a beneficial approach for improving quality of life and treatment adherence and reducing anxiety and depression among hemodialysis patients. However, future high-quality randomized controlled trials are essential to draw more reliable conclusions.

Trial Registration: PROSPERO CRD42024589799; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024589799

(J Med Internet Res 2025;27:e67246) doi:10.2196/67246

KEYWORDS

hemodialysis; eHealth interventions; quality of life; treatment adherence; anxiety; depression; meta-analysis; kidney; systematic review; kidney diseases; kidney function; chronic diseases

Introduction

End-stage renal disease (ESRD), which arises as a consequence of the progression of chronic kidney disease (CKD), is a condition that affects millions globally and is recognized as one of the most widespread and significant chronic diseases [1-3]. Hemodialysis (HD), the primary modality of renal replacement therapy for advanced kidney failures, plays a crucial role in mitigating the otherwise inevitable progressive decline in kidney function [4]. As of 2020, it is estimated that 88.07% of patients with ESRD in the United States were undergoing HD, compared with 46% to 98% of patients in Europe [5,6]. HD functions as a partial replacement for kidney activity and plays a critical role in maintaining water-electrolyte homeostasis and, as a result, significantly contributes to the prolongation of patient survival [7,8]. However, while undergoing long-term HD treatment, patients are often accompanied by adverse effects such as anxiety, depression, and fatigue, which significantly impair their compliance with treatment and quality of life [9,10]. Furthermore, the financial burden on HD patients, their families, and society is typically substantial due to the prohibitive cost of dialysis and loss of productivity [11]. Therefore, effective, accessible, and cost-effective interventions aimed at alleviating the negative emotions experienced by HD patients, while improving their adherence and quality of life, are urgently sought.

With the continual refinement of dialysis technology, the life expectancy of HD patients has increased markedly [12]. As a result, their focus has shifted from merely extending survival to enhancing quality of life [13]. To achieve this, HD patients must comply with a strict and well-structured treatment protocol that comprises regular medication intake, diet management, fluid control, and physical activity [14]. However, given the poor accessibility of medical resources and inadequate health literacy, they frequently find it difficult to obtain appropriate self-management support, including health guidance and supervision, psychological counseling, and scheduled follow-up visits [15,16]. These barriers tend to precipitate adverse emotions, such as anxiety and depression, which diminish treatment adherence and ultimately impede the enhancement of the quality of life [17,18]. Furthermore, HD patients primarily receive care services in outpatient clinics and spend the remainder at home or in the community. In contrast, in low- and middle-income countries, underdeveloped community care networks further limit access to timely care and health monitoring, and the effectiveness of their treatments for this condition is usually significantly compromised [19,20].

Benefiting from the rapid development of electronic information technology, eHealth interventions have been recognized as a powerful approach for disease prevention, health behavior maintenance, and improved health outcomes owing to their accessibility and cost-effectiveness [21,22]. In the field of health care, eHealth interventions mainly deliver continuous and personalized services to patients, primarily through digital platforms including the internet, mobile apps, websites, and text messaging [23,24]. Multiple studies have revealed the substantial potential of eHealth interventions, not only in enhancing the quality of life but also in enhancing self-efficacy and significantly reducing anxiety and depression in patients with chronic diseases, thereby underscoring their effectiveness as a multifaceted therapeutic approach capable of addressing both physical and psychological dimensions of patient health [25,26]. However, while a growing body of original research has evaluated the effects of eHealth on quality of life, treatment adherence, anxiety, and depression in HD patients, substantial variability remains in terms of intervention formats, duration, control conditions, and outcome measures [20,27,28]. For instance, a Thailand-based study demonstrated that a tele-home health care model (including video visits, telephone counseling, web-based education, and monitoring) significantly improved the quality of life of patients with HD at 3 months [27]. Similarly, another research conducted in Iran reported that a 1-month period of nurse-delivered tele-nursing significantly reduced the levels of anxiety and depression in HD patients [29]. However, a recent study evaluating the difference between an internet-based self-help intervention and usual care in improving anxiety, depression, and quality of life among HD patients revealed no significant difference in the clinical efficacy of both interventions [28]. Variations in these results may stem from a variety of factors such as different cultural backgrounds, patient characteristics, intervention format, and duration of the intervention. In other words, the mechanisms through which eHealth interventions improve health outcomes in HD patients remain unclear. In addition, the complexity of eHealth interventions also poses serious challenges for health care providers in their implementation [30]. Last but not least, most of the previous research on the application of eHealth in HD patients was conducted with small samples, feasibility trials, or qualitative methods, making it difficult to draw reliable conclusions. To our knowledge, no meta-analyses have been conducted to date that comprehensively investigate the efficacy of eHealth interventions specifically targeting HD patients, highlighting a significant gap in the literature and underscoring the need for comprehensive evidence synthesis in this area. Therefore, a meta-analysis is warranted to clarify the

XSL•FO
RenderX

effectiveness of eHealth interventions in HD patients and to identify the optimal form and duration of these interventions for standardized clinical applications in the future.

Methods

Design

This meta-analysis was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Multimedia Appendix 1) [31]. The study protocol was registered with PROSPERO (CRD42024589799).

Literature Search

A literature search was conducted in PubMed, Web of Science, Embase, CINAHL, Cochrane Library, PsycINFO, China National Knowledge Infrastructure (CNKI), WanFang, VIP, and CBM from the inception to September 7, 2024. We used the following search terms: dialysis, hemodialysis, hemodialysis, maintenance hemodialysis, eHealth, website, internet, text messaging, email, digital health, telephone, smartphone, phone, mobile phone, mobile device, mobile health, mHealth, app, application, video, computer, RCT, and randomized controlled trial. The search strategy was designed using Boolean operators to combine relevant terms. For example, in PubMed, the following search string was used: ("dialysis" OR "hemodialysis" OR "haemodialysis" OR "maintenance hemodialysis" OR "maintenance haemodialysis" OR "MHD") AND ("telemedicine" OR "internet-based intervention" OR "telerehabilitation" OR "web" OR "website" OR "internet" OR "text messaging" OR "email" OR "digital health" OR "eHealth" OR "e-health" OR "telephone" OR "smartphone" OR "phone" OR "mobile phone" OR "mobile device" OR "technology" OR "mobile health" OR "mHealth" OR "m-health" OR "online" OR "app" OR "application" OR "video" OR "computer") AND ("RCT" OR "randomized clinical trial" OR "randomized controlled trial" OR "randomized trial" OR "randomised controlled trial" OR "randomised trial"). In addition, we complemented the search with a list of references that were incorporated into the literature or relevant definitive reviews. The specific details of all search strategies are contained in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

The eligibility criteria were determined based on the PICOS (Population, Intervention, Comparison, Outcomes, and Study design) acronym (Textbox 1).

Textbox 1. Eligibility criteria.

Inclusion criteria:

- Population: adult patients (aged 18 years or older) received hemodialysis treatment.
- Intervention: the interventions were administered through a variety of eHealth technologies, including websites, the internet, social media platforms, telephone communications, video consultations, and other digital health tools.
- Comparison: the control groups were assigned to usual care (routine health education and counseling, periodic follow-up, psychological care, and standard physical examinations) without placebo and eHealth technology.
- Outcomes: the study results were focused on the outcomes of quality of life, treatment adherence, and anxiety or depression.
- Study design: adhered to a randomized controlled trial format. In addition, we only included studies published in English or Chinese.

Exclusion criteria:

- Duplicate publications.
- Conference abstracts, study protocols, letters, case reports, and reviews.
- Without sufficient data for meta-analysis.

Study Selection and Data Extraction

All citations were imported into Endnote X20 for data management. Following the import of all references, duplicate entries were removed, after which 2 reviewers independently screened the remaining records by reading the title, abstract, and full text using the predefined inclusion criteria. To ensure methodological rigor in data extraction, 2 independent reviewers (XHZ and HC) retrieved data and recorded them in a structured spreadsheet, which was subsequently cross-verified. From each study, the following data were meticulously extracted: author, publication year, country, age, sample size, details of the experimental and control conditions, duration of the intervention, and measures. In cases where discrepancies arose between the 2 reviewers, a third reviewer (WWY) was consulted to achieve consensus.

Risk of Bias Assessment

Two reviewers (XHZ and HC) independently assessed the methodological quality of the included studies using the Cochrane Risk of Bias Tool, and any disagreements between them were clarified by discussion with a third researcher (WWY). In cases where consensus could not be reached, the final decision was made by arbitration, with the third reviewer casting the deciding vote. The tool comprises 7 components, including random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, missing outcome data, selective reporting, and other biases [32]. Each domain within the assessment was categorized as presenting either a "low risk," "unclear risk," or "high risk" of bias, and for the overall risk of bias evaluation, a study was deemed to carry a "low risk of bias" only if all individual items within the assessment were consistently identified as

XSL•FO
RenderX

representing low risk. A study was rated as having "some concerns" when it raised concerns in at least 1 area. It was deemed to have a "high risk of bias" if at least 1 domain exhibited a high risk or if multiple domains indicated concerns.

Quality of Evidence Assessment

Two researchers separately appraised the quality of evidence for outcomes such as quality of life, treatment adherence, anxiety, and depression, which were evaluated using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework. Discussion with a third researcher was conducted to clarify disagreements. The framework categorized evidence into 4 levels, "high," "moderate," "low," and "very low" based on limitations in the dimensions of risk of bias, inconsistency, directness, imprecision, and publication bias [33]. In addition, if the effect size was substantial or the dose-response ratio strong, the evidence grade for the outcome could be upgraded by 1 level.

Statistical Analysis

This study identified quality of life and treatment adherence as primary outcomes and anxiety and depression as secondary outcomes. The meta-analysis and heterogeneity test were conducted with RevMan (version 5.3; Cochrane) and Stata (version 17.0; StataCorp). As the trials measured the outcomes of interest in various tools, the standardized mean difference (SMD) with 95% CI was used to estimate the pooled intervention effect [34]. The SMD magnitudes were interpreted as small (<.5), moderate (.5-.8), and large (>.8) [35]. The I^2 statistic and P values were used to evaluate heterogeneity. A fixed-effect model would be applied to collapsed data if $I^2 \leq 50\%$ and $P>.1$, otherwise a random-effect model would be performed to provide more reliable estimates [34]. In addition, we conducted a subgroup analysis stratified the duration and format of interventions to identify potential contributors to heterogeneity [36]. Given that over 10 studies reported quality of life as an outcome, publication bias was evaluated both visually using a funnel plot and quantitatively through the Egger linear regression method. The trim and fill analysis was performed to adjust for any publication bias and to estimate its effect on effect sizes. Finally, a sensitivity analysis was undertaken to assess the stability of the pooled results by applying a leave-one-out approach.

Results

Study Selection

The initial database search resulted in the retrieval of 5738 articles, with an additional 3 studies identified through a manual review of reference lists. After removing duplicates, 3581 articles were reviewed based on their titles and abstracts. As a result, 142 articles met the initial validation criteria. Subsequently, following a full-text examination, 17 papers were included in this review. The study selection and literature screening process are detailed in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA flowchart of study selection and literature screening process. RCT: randomized controlled trial.

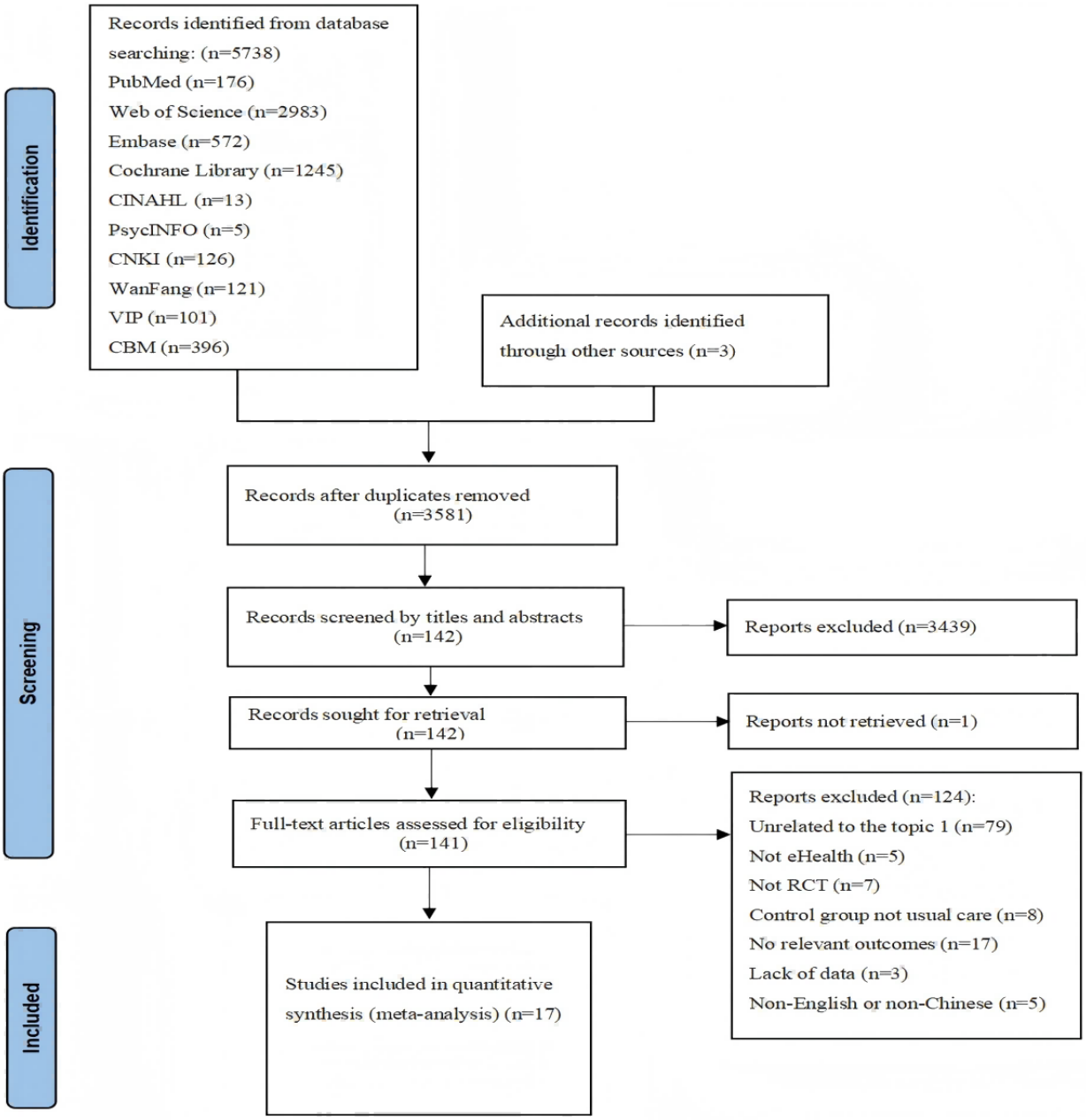

Characteristics of the Included Studies

A total of 1728 HD patients were recruited in 17 trials [10,17,20,27,28,37-48], of which 879 were assigned to the experimental group based on eHealth technologies, while 849 were in the control group and received usual care. Of the 17 randomized controlled trials, 2 were clustered trials, 1 was a three-arm trial, and 2 were feasibility trials. The studies were conducted in China (n=8), Iran (n=4), Australia (n=1), Turkey (n=1), Oman (n=1), the Netherlands (n=1), Thailand (n=1), with the publication years spanning between 2015 and 2024. The mean ages of all participants ranged from 27 (SD 11.5) to 69.13 (SD 11.82) years old. Nevertheless, three studies failed to provide the age of the participants [20,37,38]. Multimedia Appendix 3 details the characteristics of the included studies.

Main Features of Interventions

In the 17 studies, 8 kinds of eHealth intervention delivery formats were used, including text messages, telephone sessions, video, network platforms, social media, computers, websites, and mobile apps. The majority of research used a single form of eHealth intervention, and 7 studies adopted a combined approach of 2 or more eHealth technologies [17,20,28,37-40]. The duration of eHealth interventions demonstrated substantial variability across studies, spanning from 4 weeks to 12 months, of which 3 months was the most common.

Main Features of Controls

All HD patients in the control group were assigned to receive usual care during the intervention, including health assessment, advice and counseling, laboratory and physical examinations, and follow-up care. Nevertheless, there were large gaps between studies in terms of descriptions of the components of usual care,

XSL•FO
RenderX

with 6 of the studies [10,20,27,39,41,42] not detailing the elements of the intervention in the control group.

Outcome Measures

In total, 7 validated scales were used to measure the quality of life among HD patients: the EuroQol 5-Dimension (EQ-5D) [41], the 36-Item Short Form Health Survey (SF-36) [10,37,43-45], the 12-Item Short Form Health Survey (SF-12) [28], the Kidney Disease Quality of Life–Short Form (KDQOL–SF) [39,42], the 9-item Thai Health Status Assessment Instrument [29], the World Health Organization Quality Of Life-BREF (WHOQOL-BREF) [40], and World Health Organization Quality Of Life-SF (WHOQOL-SF) [20].

Five studies that reported on treatment adherence in HD patients applied 2 various instruments to measure this indicator, including the End-Stage Renal Disease Adherence Questionnaire (ESRD-AQ) [17,20] and the Treatment adherence scale for maintenance hemodialysis patients with end-stage renal disease [44,47].

Three scales were administered to determine HD participants' anxiety: the Depression Anxiety and Stress Scale (DASS) [27], the Self-Rating Anxiety Scale (SAS) [45-47], and the Beck Anxiety Inventory (BAI) [28,48].

Four solid instruments were used to assess the level of depression among HD patients: DASS [27], the Self-rating Depression Scale (SDS) [45-47], the Beck Depression Inventory-II (BDI-II) [28,48], and the Beck Depression Inventory-Short Form (BDI-SF) [20].

Risk of Bias

A total of 14 studies were considered to have "some concern" about selection bias due to not reporting the specific methods used for random sequence generation or the details of allocation concealment [10,20,27,28,37,39,40,42-48]. In addition, 15 studies were considered to be at "high risk" of performance and detection bias as they did not apply blinding to participants, personnel, or outcome assessors [10,17,20,27-29,37,40,41,43-48]. Ultimately, all trials were judged as having a low risk of attrition and reporting bias because of their low rate of missing outcome data and explicit reasons for missing data, as well as the reported results were consistent with those described in the published study protocols or in the methods sections of the articles. A summary of the risk of bias assessment for the included studies is detailed in Figure 2.

XSL•FO
RenderX

Figure 2. Risk of bias assessment summary of included studies.

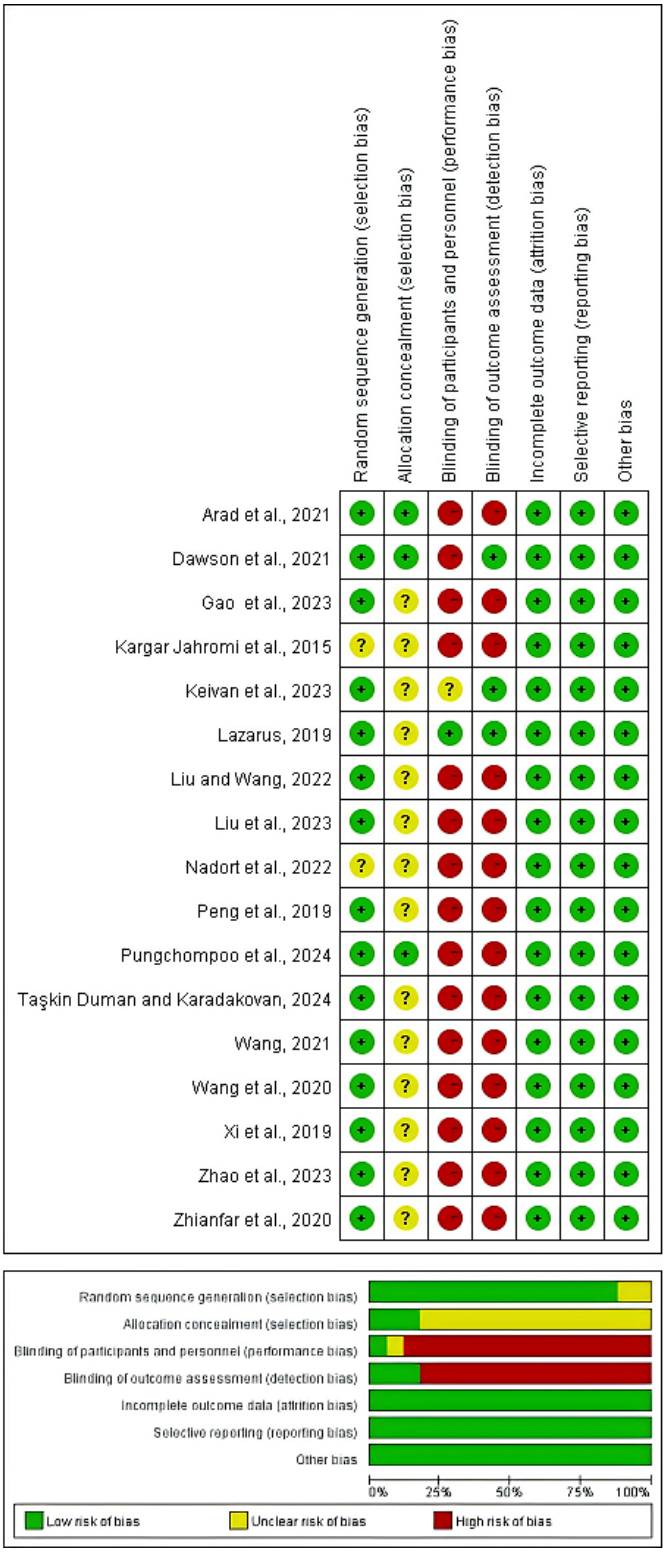

Main Results of Meta-Analysis

Quality of Life

Twelve trials investigated the effect of eHealth interventions on the quality of life of HD patients [10,20,28,37-45]. The pooled result demonstrated a statistically significant efficacy of eHealth interventions in enhancing quality of life, with a substantial effect size (SMD 0.87, 95% CI 0.38-1.37; Multimedia Appendix 4). The sensitivity analysis indicated that no single trial was sufficient to alter the overall result. The sensitivity analysis confirmed the robustness of the meta-analysis findings, as the exclusion of any individual study did not alter the overall results (Figure S2a in Multimedia Appendix 5).

In the analysis categorized by intervention duration, eHealth interventions of short duration (≤3 months) were effective in improving quality of life among HD patients, with a substantial

XSL•FO
RenderX

reduction in heterogeneity (SMD 0.5, 95% CI 0.28-0.73; I^2=61%; Multimedia Appendix 4). However, it showed a significantly higher positive impact on quality of life than long-term (>3 months) eHealth interventions (SMD 2.25, 95% CI 0.29-4.21; I^2=98%; Multimedia Appendix 4). Subgroup analysis of the different intervention formats revealed that a single eHealth intervention (SMD 1.5, 95% CI 0.45-2.55; I^2=97%; Multimedia Appendix 4) had a greater positive impact on quality of life than a combination of eHealth interventions.

Treatment Adherence

Five trials evaluated the impact of eHealth interventions on adherence among HD patients [17,20,44,47,48]. The pooled results demonstrated that eHealth interventions significantly enhanced HD patients' treatment adherence (SMD 1.11, 95% CI 0.3-1.91; Multimedia Appendix 4). The sensitivity analysis demonstrated the robustness of the results (Figure S2b in Multimedia Appendix 5).

In the analysis classified by intervention duration, the short-term (≤3 months) eHealth interventions were effective in terms of treatment adherence among HD patients (SMD 1.34, 95% CI 0.31-2.37; I^2=95%), whereas long-term (>3 months) eHealth interventions failed to demonstrate a significant impact (SMD 0.25, 95% CI −0.11 to 0.61; Multimedia Appendix 4). When comparing the relative effects between subgroups of different intervention formats in terms of treatment adherence, it was found that neither single (SMD 0.64, 95% CI −0.05 to 1.33; I^2=91%) nor combined eHealth interventions (SMD 1.89, 95% CI −0.69 to 4.48; I^2=97%) had a significant effect on treatment adherence (Multimedia Appendix 4).

Anxiety

Six trials provided sufficient data about the impact of eHealth interventions on anxiety among HD patients [27,28,45-48]. The pooled results revealed that eHealth interventions led to a significant reduction in anxiety levels among HD patients (SMD −2.11, 95% CI −3.25 to −0.97; Multimedia Appendix 4). Despite the very high heterogeneity of the combined results, sensitivity analysis confirmed the stability of the result (Figure S2c in Multimedia Appendix 5). When comparing the effects between subgroups of different intervention duration on anxiety of HD patients, short-term (≤3 months) eHealth interventions (SMD −1.95, 95% CI −3.84 to −0.07; I^2=98%) had significantly lower pooled effects than long-term (≤3 months) eHealth interventions (SMD −2.29, 95% CI −3.74 to −0.83; I^2=95%; Multimedia Appendix 4). In addition, the subgroup analyses revealed that single eHealth interventions significantly reduced anxiety in HD patients (SMD −2.54, 95% CI −3.71 to −1.39; I^2=96%), while mixed eHealth interventions showed no statistically significant impact (SMD 0.08, 95% CI −0.28 to 0.43; Multimedia Appendix 4).

Depression

Seven trials reported adequate data on the effect of eHealth interventions on anxiety among HD patients [20,27,28,45-48]. The pooled results showed that eHealth interventions had a statistically significant impact on reducing depression, with a large effect size (SMD −2.46, 95% CI −3.6 to −1.25; Multimedia Appendix 4). The sensitivity analysis revealed that the results remained consistent, with no significant changes after excluding any individual studies (Figure S2d in Multimedia Appendix 3).

When comparing the effects between subgroups of different intervention duration on depression of HD patients, short-term (≤3 months) eHealth interventions (SMD −1.69, 95% CI −3.03 to −0.35; I^2=97%) had significantly lower pooled effects than long-term (≤3 months) eHealth interventions (SMD −3.53, 95% CI −6.50 to −0.57; I^2=98%; Multimedia Appendix 4). Furthermore, a subgroup analysis stratified by intervention formats revealed that a single eHealth intervention was more effective in reducing anxiety in HD patients (SMD −3.40, 95% CI −5.07 to −1.72; I^2=95%; Multimedia Appendix 4).

Publication Bias

We used the Egger test and a funnel plot to evaluate potential publication bias in studies related to quality of life. The asymmetry observed in the funnel plot indicated the presence of publication bias among the included studies (Figure S3 in Multimedia Appendix 6). Furthermore, the results of the Egger test indicated a significant bias, with a regression coefficient (beta1) of 13.92 (SE 2.084). The corresponding z value was 6.68, and the P value was <.001, suggesting that small-study effects are present in the analysis. This result (P<.001) indicates that the observed effect size may be influenced by small sample studies, which could lead to potential publication bias. Thus, the trim-and-fill analysis was performed to counteract the bias. However, there were no missing studies during the analysis and the corrected effect sizes remained unchanged, indicating the stability of the meta-analysis results.

Quality of Evidence

Following the guidance in the GRADE standard, the certainty of the evidence for the quality of life, treatment adherence, anxiety, and depression were classified as low, moderate, moderate, and moderate, respectively. The justification for the ratings and other details are listed in Table 1.

XSL•FO
RenderX

Table 1. Grading of Recommendations, Assessment, Development, and Evaluation summary of the quality of the evidence for the outcomes.

Outcome	Quality assessment					Number of participants (studies)	Effect size, SMD (95% CI)	Quality of the evidence (GRADE[a])
	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Quality of life	Serious[b]	Serious[c]	Not serious[d]	Not serious[e]	Publication bias strongly suspected[f] large effect size[g]	1243 (12)	0.87 (0.38 to 1.37)	Low
Treatment adherence	Serious[b]	Serious[c]	No serious[d]	No serious[e]	Large effect size[g]	508 (5)	1.11 (0.3 to 1.91)	Moderate
Anxiety	Serious[b]	Serious[c]	No serious[d]	No serious[e]	Large effect size[g]	536 (6)	−2.11 (−3.25 to −0.97)	Moderate
Depression	Serious[b]	Serious[c]	No serious[d]	No serious[e]	Large effect size[g]	634 (7)	−2.46 (−3.68 to −1.25)	Moderate

[a]GRADE: Grading of Recommendations, Assessment, Development, and Evaluation.

[b]Most information is from studies rated as high risk of bias.

[c]I^2>50%; heterogeneity could not be explained by the form and duration of intervention.

[d]The various eHealth interventions are compared directly with usual care among cancer patients.

[e]The 95% CI excludes pooled effect sizes that are not clinically significant.

[f]The results of the Egger test and funnel plot revealed the evidence for publication bias.

[g]The pooled effect size >0.8.

Discussion

Principal Findings

As far as we know, this study was the first meta-analysis of randomized controlled trials to systematically assess the efficacy of eHealth interventions on quality of life, treatment adherence, anxiety, and depression among HD patients. A total of 17 trials involving 1728 HD patients were included in this meta-analysis. We found varied quality of evidence for the effects of eHealth interventions on quality of life, treatment adherence, anxiety, and depression in people with HD. Specifically, low-quality evidence demonstrated that eHealth interventions have significant clinical effects in enhancing the quality of life among HD patients compared with the control groups. In addition, moderate quality evidence indicated that eHealth interventions can significantly enhance treatment adherence and reduce anxiety and depression in HD patients. However, given the worrying overall methodological quality, more high-quality clinical trials remain essential to draw reliable conclusions in the future. Specifically, the majority of studies included in the meta-analysis had selection, performance, and detection biases, and these potential biases may limit the generalizability of the findings to a wider population, and the lack of blinding may have led to exaggerated estimates of the efficacy of the eHealth interventions. In addition, the subjectivity of the outcome measures combined with unblinded assessments may have affected the results. Furthermore, although the sensitivity analysis results indicated the robustness of the estimated outcomes, significant heterogeneity was still observed that was difficult to explain through subgroup analysis. First, differences in the methodological quality of the included studies, particularly inconsistencies in randomization and blinding, may have contributed to the variability in the results. Second, differences in sample characteristics, such as age, gender, and disease severity, could also be an important source of heterogeneity. Finally, different measurement tools and evaluation standards may lead to the same variable being recorded and assessed in various ways, thereby increasing the inconsistency of the results.

Our findings demonstrated a large effect of eHealth interventions on enhancing the quality of life among HD patients, consistent with a previous meta-analysis evaluating the impact of eHealth intervention on the quality of life of breast cancer patients [49]. Currently, the Internet, telephones, and mobile apps are essential components of people's daily lives, and they have laid a good foundation for the application of eHealth interventions in HD patients [29]. According to the biopsychosocial model, the individual health status is the consequence of the interaction of biological, psychological, and social factors [50]. The timeliness, rapidity, and accuracy of eHealth interventions enable HD patients to receive continuous care, based on which they can constantly monitor their physiological status and reduce complications [27,28]. Furthermore, through channels such as email and mobile apps, HD patients are able to receive psychological support from family and society as well as psychological counseling from health care professionals, thus enhancing mental health [46]. Finally, over the internet, social media, or apps, HD patients could seek help from health care professionals (health counseling, psychosocial support, and visit coordination), thus improving social adaptation [48]. These positive aspects of improvement in physical, psychological, and social factors ultimately boost the overall health of the patient and achieve an overarching quality of life [37,41]. However, the estimated effect size of eHealth interventions on quality of life exhibited substantial heterogeneity. When the study by Zhao et al [45] was excluded, a significant reduction in the overall effect size was observed, with I^2 decreasing from 93.87% to 79.9%. This suggests that a substantial impact was made on the overall result of this study, which may have been a significant

XSL•FO
RenderX

source of the detected heterogeneity. The potential overestimation of the intervention effect in the study of Zhao et al [45] may be attributed to the lack of rigorous randomization and blinding procedures. Therefore, the effectiveness of eHealth interventions on the quality of life of HD patients should be carefully considered. In addition, it is noteworthy that the subgroup analyses indicated that a long-term and single form of eHealth intervention appeared to be more efficacy in enhancing HD patients' quality of life. This could be because HD patients improve their disease management knowledge and skills over time as the intervention progresses, which helps in effectively preventing and reducing complications. However, if the intervention is overly complex and time-consuming, such as involving multiple eHealth technologies that require extensive eHealth literacy, it may lead to poor patient adherence and potentially increase the overall burden on the patient [49]. Therefore, when implementing eHealth interventions for HD patients, careful consideration should be given to optimizing the duration and format of the intervention to avoid placing unnecessary burdens on patients.

In line with findings on quality of life, our findings demonstrated that eHealth interventions exerted a favorable effect on treatment adherence among HD patients. Treatment adherence among HD patients primarily encompasses dialysis schedule, prescribed session time, medication regimens, and dietary restrictions (namely salt and water intake) [14,47]. In recent years, eHealth interventions have emerged as an effective approach for assessing, monitoring, and managing self-care activities in HD patients [51,52]. The Health Belief Model holds that an individual's health behaviors depend on his or her perception of the threat of disease and assessment of the benefits of and barriers to behavioral modification [53]. Through the vehicle of eHealth technology (mobile health apps and remote monitoring), HD patients can maintain closer contact with health care professionals, facilitating access to personalized health information and reminders, thus increasing awareness of the severity of the disease and the imperative for treatment [17,20]. In addition, eHealth platforms make it possible to systematically collect health data from HD patients, and with real-time feedback and visualization of data (eg, dialysis outcomes and lab metrics), it is easier for patients to perceive the benefits of their treatment, thus increasing their insights into treatment adherence [48]. In subgroup analyses, short-term eHealth interventions seem to be more effective in enhancing treatment adherence among HD patients. Furthermore, no statistically significant difference was observed in the impact on treatment adherence among HD patients, regardless of whether they received a single form of eHealth intervention or a combination of multiple forms. This underscores the ongoing need to further investigate the optimal timing and format of eHealth interventions to enhance treatment adherence among HD patients in the future.

Due to the severe burden of the disease, HD patients generally exhibit high levels of anxiety [54]. Although routine nursing interventions may temporarily reduce anxiety, their effectiveness commonly declines over time due to a lack of continuity, leading to decreased compliance and affecting patients' quality of life [55]. This review identified a significant impact of eHealth interventions in alleviating anxiety among HD patients. This finding might be explained through self-efficacy theory. Self-efficacy refers to the individual's belief in his or her ability to successfully perform the necessary behaviors to achieve a desired outcome [56]. Monitoring, reminding and feedback were the most common eHealth intervention components in the included studies, and these features provide HD patients with the tools and resources to actively manage their disease. When patients perceive tangible gains in their health status, their self-efficacy increases, resulting in a reduction in anxiety associated with disease management [46]. Subgroup analyses indicated that anxiety reductions were more pronounced when patients received eHealth interventions for over 3 months, compared with shorter-term interventions. In addition, a single form of eHealth intervention showed statistically significant effects in reducing anxiety among HD patients, whereas the impact of combined interventions was not significant. However, the effectiveness of eHealth interventions in alleviating anxiety among HD patients requires further validation through more rigorous clinical trials, given the limited number of studies using combined intervention formats, the heterogeneity of interventions, and the moderate quality of the current evidence.

Similar to the pooled results of anxiety, our findings suggested that eHealth interventions had a significant efficacy in decreasing depression among HD patients. Consistent with our findings, a previous study reported that providing psychological counseling services, along with sharing health knowledge and communication skills through a WeChat chatbot, can effectively alleviate depression in HD patients [45]. However, another study found no statistically significant effect of web-based problem-solving therapy on depression levels among HD patients [28]. This inconsistency may be explained by variations in the duration and format of eHealth interventions across different studies. According to the Stress and Coping Theory, equipping patients with effective coping strategies, such as emotional support and problem-solving skills, to help them manage stress could reduce the psychological burden associated with chronic illness [57]. In the included studies, psychological counseling was frequently a crucial element of eHealth intervention, and with the online platform, health care professionals were able to support patients in venting their negative emotions promptly, thereby reducing depression levels. The subgroup analyses revealed that reductions in depressive symptoms were more pronounced when participants received eHealth interventions lasting longer than 3 months or delivered in a single format. The prevalence of depression in HD patients has been estimated at 56.8% [58], which highlights the urgency of targeted interventions. As a chronic mental condition, depression typically requires sustained and consistent interventions to achieve significant improvements [59]. eHealth interventions shorter than 3 months may be insufficient for alleviating depression in HD patients, and combined interventions may present challenges in implementation. Therefore, extending the duration and increasing the frequency of interventions are recommended in future studies to facilitate the demission of depression.

XSL•FO
RenderX

Limitations

Some limitations should be considered. First, this study only included randomized controlled trials published in English and Chinese, with no incorporation of gray literature such as unpublished studies, conference abstracts, and theses. This exclusion may have contributed to publication bias, particularly in the evaluation of quality of life, as gray literature often includes negative or null results that are less likely to be published in peer-reviewed journals. Consequently, the omission of such studies may have led to an overestimation of the intervention effect. Furthermore, the lack of grey literature may have reduced the diversity of the evidence base, potentially underestimating the heterogeneity among studies and limiting the generalizability of the findings to broader populations or settings. Second, the substantial heterogeneity observed in the pooled outcomes of interest, which resulted from the considerable differences in methodological quality and clinical characteristics of the included patients, could not be clearly explained by subgroup analysis, as no definitive source of heterogeneity was identified. Finally, as the majority of studies included in the meta-analysis were conducted in low- and middle-income countries, the applicability of the findings to high-income countries may be limited.

Implications

eHealth interventions have attracted growing attention worldwide, with numerous studies highlighting their significant potential in the health care of chronically ill patients [23,56]. This meta-analysis provided the first quantitative synthesis of research findings on the administration of eHealth interventions in HD patients and found that these interventions were effective in improving quality of life and treatment adherence, as well as reducing anxiety and depression levels. These findings emphasize the value of integrating eHealth interventions into clinical nursing practice as a valuable and effective strategy. Despite significant heterogeneity in the included studies, this study consolidated the findings of the impact of eHealth interventions on different health-related outcomes among HD patients and would constitute valuable evidence for health care practitioners and researchers to understand the range of areas where eHealth interventions could make a difference, and the magnitude of the impacts that they would have made. While additional research is required to validate these findings, eHealth interventions continue to show great promise in enhancing quality of life, improving treatment adherence, and alleviating anxiety and depression in HD patients. In addition, we also observed that the duration and format of eHealth interventions had a remarkable effect on their efficacy, prompting the notion that we might have to determine the optimal balance to foster patient health outcomes. Notably, the methodological appraisal revealed a high risk of bias in nearly all included studies, underscoring the urgent need for improvements in study design and intervention delivery. Therefore, when designing and implementing eHealth interventions in the future, it is crucial to incorporate appropriate strategies to enhance patient motivation and adherence, thereby ensuring the effectiveness of the interventions.

Conclusions

In conclusion, this meta-analysis highlighted that eHealth interventions may serve as an effective strategy for enhancing quality of life, improving treatment adherence, and alleviating anxiety and depression in HD patients. Thus, health care providers should consider eHealth interventions as a pivotal strategy for facilitating HD patient health outcomes. Furthermore, the efficacy of eHealth interventions for HD patients varied significantly depending on the duration and format of the intervention. Therefore, further in-depth research on eHealth intervention strategies for this population is essential to enhance intervention effectiveness and address the growing health care needs of HD patients. Last but not least, it is essential that future high-quality studies be conducted to provide more robust evidence regarding the clinical benefits of eHealth interventions for HD patients.

Acknowledgments

The authors thank all those who contributed so much to the literature search, data extraction, and analysis.

Data Availability

The datasets used in this study are publicly accessible, with detailed information provided in the Methods section.

Authors' Contributions

XHZ contributed to the conceptualization, data curation, formal analysis, methodology, software, validation, visualization, and writing of the original draft. HC contributed to the data curation, validation, and writing – review and editing. WWY contributed to the data curation, formal analysis, methodology, validation, and writing – review and editing. LW contributed to the data curation, methodology, supervision, and writing – review and editing. LC contributed to the conceptualization, methodology, supervision, and writing – review and editing. YZ contributed to the conceptualization and data curation. YJZ contributed to conceptualization, data curation, and methodology. MS contributed to the conceptualization, data curation, methodology, supervision, validation, and writing – review and editing. QZ contributed to conceptualization, data curation, and validation. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 33 KB - jmir_v27i1e67246_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 30 KB - jmir_v27i1e67246_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies.
[DOCX File , 22 KB - jmir_v27i1e67246_app3.docx]

Multimedia Appendix 4
Results of meta-analysis: forest plots for the comparison of eHealth interventions against usual care.
[DOCX File , 1410 KB - jmir_v27i1e67246_app4.docx]

Multimedia Appendix 5
Sensitivity analysis.
[DOC File , 493 KB - jmir_v27i1e67246_app5.doc]

Multimedia Appendix 6
Publication bias of included studies.
[DOCX File , 65 KB - jmir_v27i1e67246_app6.docx]

References

1. Perez-Dominguez B, Suso-Marti L, Dominguez-Navarro F, Perpiña-Martinez S, Calatayud J, Casaña J. Effects of resistance training on patients with end-stage renal disease: an umbrella review with meta-analysis of the pooled findings. J Nephrol 2023;36(7):1805-1839. [doi: 10.1007/s40620-023-01635-7] [Medline: 37318646]
2. Prabu P, Acree L, Waller JL, Linder DF, Bollag WB, Mohammed A, et al. Sleep apnea in end-stage renal disease patients: risk factors and mortality. J Investig Med 2023;71(5):465-470. [doi: 10.1177/10815589231162541] [Medline: 36945195]
3. Xie E, Ye Z, Wu Y, Zhao X, Li Y, Shen N, et al. The triglyceride-glucose index predicts 1-year major adverse cardiovascular events in end-stage renal disease patients with coronary artery disease. Cardiovasc Diabetol 2023;22(1):292 [FREE Full text] [doi: 10.1186/s12933-023-02028-7] [Medline: 37891651]
4. Hargrove N, El Tobgy N, Zhou O, Pinder M, Plant B, Askin N, et al. Effect of aerobic exercise on dialysis-related symptoms in individuals undergoing maintenance hemodialysis: a systematic review and meta-analysis of clinical trials. Clin J Am Soc Nephrol 2021;16(4):560-574 [FREE Full text] [doi: 10.2215/CJN.15080920] [Medline: 33766925]
5. Kidney Disease: Improving Global Outcomes (KDIGO) Hepatitis C Work Group. KDIGO 2022 clinical practice guideline for the prevention, diagnosis, evaluation, and treatment of hepatitis C in chronic kidney disease. Kidney Int 2022;102(6S):S129-S205 [FREE Full text] [doi: 10.1016/j.kint.2022.07.013] [Medline: 36410841]
6. Kelly DM, Rothwell PM. Prevention and treatment of stroke in patients with chronic kidney disease: an overview of evidence and current guidelines. Kidney Int 2020;97(2):266-278 [FREE Full text] [doi: 10.1016/j.kint.2019.09.024] [Medline: 31866114]
7. Eckert K, Motemaden L, Alves M. Effect of hemodialysis compared with conservative management on quality of life in older adults with end-stage renal disease: systematic review. J Hosp Palliat Nurs 2018;20(3):279-285. [doi: 10.1097/NJH.0000000000000444] [Medline: 30063679]
8. Liyanage T, Ninomiya T, Jha V. Worldwide access to treatment for end-stage kidney disease: a systematic review. J. Vasc. Surg 2015;62(4):1089. [doi: 10.1016/j.jvs.2015.08.003]
9. Abeywickrama HM, Wimalasiri S, Koyama Y, Uchiyama M, Shimizu U, Kakihara N, et al. Quality of life and symptom burden among Chronic Kidney Disease of Uncertain Etiology (CKDu) patients in Girandurukotte, Sri Lanka. Int J Environ Res Public Health 2020;17(11):4041 [FREE Full text] [doi: 10.3390/ijerph17114041] [Medline: 32517110]
10. Taşkin Duman H, Karadakovan A. The effect of video training on symptom burden, comfort level, and quality of life in hemodialysis patients: clustered randomized controlled trial. Patient Educ Couns 2024;126:108314. [doi: 10.1016/j.pec.2024.108314] [Medline: 38761675]
11. Jha V, Al-Ghamdi SMG, Li G, Wu M, Stafylas P, Retat L, et al. Global economic burden associated with chronic kidney disease: a pragmatic review of medical costs for the inside CKD research programme. Adv Ther 2023;40(10):4405-4420 [FREE Full text] [doi: 10.1007/s12325-023-02608-9] [Medline: 37493856]

XSL•FO
RenderX

12. Schold JD, Flechner SM, Poggio ED, Augustine JJ, Goldfarb DA, Sedor JR, et al. Residential area life expectancy: association with outcomes and processes of care for patients with ESRD in the United States. Am J Kidney Dis 2018;72(1):19-29. [doi: 10.1053/j.ajkd.2017.12.014] [Medline: 29525324]
13. Hall RK, Cary MP, Washington TR, Colón-Emeric CS. Quality of life in older adults receiving hemodialysis: a qualitative study. Qual Life Res 2020;29(3):655-663 [FREE Full text] [doi: 10.1007/s11136-019-02349-9] [Medline: 31691203]
14. Alemayehu YH, Seylania K, Bahramnezhad F. The relationship between health literacy and quality of life among hemodialysis patients: an integrative review. Hum Antibodies 2020;28(1):75-81. [doi: 10.3233/HAB-190394] [Medline: 31524153]
15. Chen C, Zheng J, Liu X, Liu J, You L. Role of health literacy profiles in fluid management of individuals receiving haemodialysis: a cross-sectional study. J Adv Nurs 2024;80(6):2325-2339. [doi: 10.1111/jan.15973] [Medline: 38012855]
16. Yao X, Wang Y, Liu Z, Li Y, Fan X. Influencing factors of resource utilisation in haemodialysis patients-Based on socioecological pyramid model. J Clin Nurs 2022;31(5-6):770-780 [FREE Full text] [doi: 10.1111/jocn.15935] [Medline: 34396613]
17. Arad M, Goli R, Parizad N, Vahabzadeh D, Baghaei R. Do the patient education program and nurse-led telephone follow-up improve treatment adherence in hemodialysis patients? A randomized controlled trial. BMC Nephrol 2021;22(1):119 [FREE Full text] [doi: 10.1186/s12882-021-02319-9] [Medline: 33827478]
18. Oquendo LG, Asencio JMM, de Las Nieves CB. Contributing factors for therapeutic diet adherence in patients receiving haemodialysis treatment: an integrative review. J Clin Nurs 2017;26(23-24):3893-3905. [doi: 10.1111/jocn.13804] [Medline: 28295744]
19. Katafuchi R, Tanaka S, Matsuo T, Tamai O, Yoshimine K, Yano K, Kasuya Chronic Kidney Disease Committee. The effect of the Kasuya CKD network on prevention of the progression of chronic kidney disease: successful collaboration of a public health service, primary care physicians and nephrologists-community based cohort study. Clin Exp Nephrol 2023;27(1):32-43. [doi: 10.1007/s10157-022-02267-0] [Medline: 36205816]
20. Zhianfar L, Nadrian H, Asghari Jafarabadi M, Espahbodi F, Shaghaghi A. Effectiveness of a multifaceted educational intervention to enhance therapeutic regimen adherence and quality of life amongst Iranian hemodialysis patients: a randomized controlled trial (MEITRA study). J Multidiscip Healthc 2020;13:361-372 [FREE Full text] [doi: 10.2147/JMDH.S247128] [Medline: 32341649]
21. Ahmed Kamal M, Ismail Z, Shehata IM, Djirar S, Talbot NC, Ahmadzadeh S, et al. Telemedicine, e-health, and multi-agent systems for chronic pain management. Clin Pract 2023;13(2):470-482. [doi: 10.3390/clinpract13020042] [Medline: 36961067]
22. Li Y, Li J, Zhang Y, Ding Y, Hu X. The effectiveness of e-Health interventions on caregiver burden, depression, and quality of life in informal caregivers of patients with cancer: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2022;127:104179. [doi: 10.1016/j.ijnurstu.2022.104179] [Medline: 35124473]
23. Davies SR. In children and adolescents with long-term physical conditions and depression and/or anxiety, do psychological therapies improve outcomes? Cochrane Clinical Answers 2019.
24. Zhang L, Ni Z, Liu Y, Chen H. The effectiveness of e-health on reducing stigma, improving social support and quality of life among people living with HIV: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2023;148:104606. [doi: 10.1016/j.ijnurstu.2023.104606] [Medline: 37801937]
25. Fernández-Rodríguez R, Zhao L, Bizzozero-Peroni B, Martínez-Vizcaíno V, Mesas AE, Wittert G, et al. Are e-Health interventions effective in reducing diabetes-related distress and depression in patients with type 2 diabetes? A systematic review with meta-analysis. Telemed J E Health 2024;30(4):919-939. [doi: 10.1089/tmj.2023.0374] [Medline: 38010739]
26. Zhou L, Zhou Y, Yu P, Meng F, Xu Y, Jiang Y. Effects of e-health interventions on health outcomes in patients with rheumatoid arthritis: a systematic review and meta-analysis. J Clin Nurs 2023;32(9-10):1748-1759. [doi: 10.1111/jocn.16236] [Medline: 35102628]
27. Kargar Jahromi M, Javadpour S, Taheri L, Poorgholami F. Effect of nurse-led telephone follow ups (tele-nursing) on depression, anxiety and stress in hemodialysis patients. Glob J Health Sci 2015;8(3):168-173 [FREE Full text] [doi: 10.5539/gjhs.v8n3p168] [Medline: 26493429]
28. Nadort E, Schouten RW, Boeschoten RE, Smets Y, Chandie Shaw P, Vleming LJ, et al. Internet-based treatment for depressive symptoms in hemodialysis patients: a cluster randomized controlled trial. Gen Hosp Psychiatry 2022;75:46-53 [FREE Full text] [doi: 10.1016/j.genhosppsych.2022.01.008] [Medline: 35134703]
29. Pungchompoo W, Parinyachitta S, Pungchompoo S, Udomkhwamsuk W, Suwan P. The feasibility of integrating a home telehealth model for older persons living with hemodialysis. BMC Geriatr 2024;24(1):378 [FREE Full text] [doi: 10.1186/s12877-024-04981-8] [Medline: 38671357]
30. Wikström L, Schildmeijer K, Nylander EM, Eriksson K. Patients' and providers' perspectives on e-health applications designed for self-care in association with surgery - a scoping review. BMC Health Serv Res 2022;22(1):386 [FREE Full text] [doi: 10.1186/s12913-022-07718-8] [Medline: 35321707]
31. Brennan SE, Munn Z. PRISMA 2020: a reporting guideline for the next generation of systematic reviews. JBI Evid Synth 2021;19(5):906-908. [doi: 10.11124/JBIES-21-00112] [Medline: 33989266]
32. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]

XSL•FO
RenderX

33. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
34. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions. Great Britain: Cochrane Book Series; 2008:243-296.
35. Schwarzer G, Carpenter JR, Rücker G. Statistical methods for sensitivity analysis in meta-analysis. Package 'metasens'. URL: https://cran.r-project.org/web/packages/metasens/metasens.pdf [accessed 2025-03-11]
36. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
37. Gao M, Kang Y, Qian HY, Li J, Qian ZD. Application of internet + continuity of care in maintenance hemodialysis patients. Tianjin Nursing 2023:573-576.
38. Pungchompoo W, Parinyachitta S, Pungchompoo S, Udomkhwamsuk W, Suwan P. The feasibility of integrating a home telehealth model for older persons living with hemodialysis. BMC Geriatr 2024;24(1):378 [FREE Full text] [doi: 10.1186/s12877-024-04981-8] [Medline: 38671357]
39. Keivan S, Shariati A, Miladinia M, Haghighizadeh MH. Role of self-management program based on 5A nursing model in quality of life among patients undergoing hemodialysis: a randomized clinical trial. BMC Nephrol 2023;24(1):58 [FREE Full text] [doi: 10.1186/s12882-023-03108-2] [Medline: 36922765]
40. Liu XG, Wang DF. Application of self-management oriented 5A nursing mode based on internet + in patients with maintenance hemodialysis. Clinical Med Eng 2022:1749-1750.
41. Dawson J, Campbell KL, Craig JC, Tong A, Teixeira-Pinto A, Brown MA, et al. A text messaging intervention for dietary behaviors for people receiving maintenance hemodialysis: a feasibility study of KIDNEYTEXT. Am J Kidney Dis 2021;78(1):85-95.e1. [doi: 10.1053/j.ajkd.2020.11.015] [Medline: 33421456]
42. Lazarus ER. Effectiveness of education and exercise on quality of life among patients undergoing hemodialysis. Clin Epidemiol Glob Health 2019;7(3):402-408. [doi: 10.1016/j.cegh.2018.07.003]
43. Peng YL, Xu HD, Liao Y. The effect of APP-based weight management on the weight of patients with maintenance hemodialysis during the interdialysis. Chinese Clinical Nursing 2019:193-196. [doi: 10.3969/j.issn.1674-3768.2019.03.003]
44. Wang P, Li WN, Li GJ, Guo Y, Wang XC. Effect of remote extended nursing model on treatment compliance, dialysis adequacy, and life quality in patients with chronic kidney disease. Nanning City, Guangxi: Journal of Guangxi Medical University; 2020:771-777. [doi: 10.16190/j.cnki.45-1211/r.2020.04.037]
45. Zhao CL, Wang L, SI DX. A nursing study on improving negative psychology of hemodialysis patients with therapeutic communication system based on network platform. Psychology Monthly 2023:193-195. [doi: 10.19738/j.cnki.psy.2023.19.059]
46. XI W, Luo JX, Liu L. Effects of a short video APP on anxiety and depression in young patients on maintenance hemodialysis. Journal of Qiqihar Medical College 2020:2506-2507. [doi: 10.3969/j.issn.1002-1256.2019.19.050]
47. Liu J, Zhao S, Zhang Q. Influence of WeChat chat robot intervention on the psychological status, treatment compliance and complications of patients with hemodialysis arteriovenous fistula. Journal of Bengbu Medical College 2023:1121-1125. [doi: 10.13898/j.cnki.issn.1000-2200.2023.08.023]
48. Wang YF. Application effect of internet-centric mindfulness decompression nursing in maintenance hemodialysis patients. Medical Higher Vocational Education and Modern Nursing 2021:149-152. [doi: 10.3969/j.issn.2096-501X.2021.02.014]
49. Singleton AC, Raeside R, Hyun KK, Partridge SR, Di Tanna GL, Hafiz N, et al. Electronic health interventions for patients with breast cancer: systematic review and meta-analyses. J Clin Oncol 2022;40(20):2257-2270 [FREE Full text] [doi: 10.1200/JCO.21.01171] [Medline: 35500200]
50. Almeida VC, Pereira LCD, Machado SDC, Maciel LYDS, de Farias Neto JP, de Santana Filho VJ. The use of a biopsychosocial model in the treatment of patients with chronic. Patient Educ Couns 2024;121:108117. [doi: 10.1016/j.pec.2023.108117] [Medline: 38183922]
51. Monteiro A, Luciana M, Scofano R. Evaluation of the experience with the use of telemedicine in a home dialysis program - a qualitative and quantitative study. Research Square. URL: http://dx.doi.org/10.21203/rs.3.rs-1523752/v1 [accessed 2025-03-15]
52. Torabikhah M, Farsi Z, Sajadi SA. Comparing the effects of mHealth app use and face-to-face training on the clinical and laboratory parameters of dietary and fluid intake adherence in hemodialysis patients: a randomized clinical trial. BMC Nephrol 2023;24(1):194 [FREE Full text] [doi: 10.1186/s12882-023-03246-7] [Medline: 37386428]
53. Sazali MF, Rahim SSSA, Mohammad AH, Kadir F, Payus AO, Avoi R, et al. Improving tuberculosis medication adherence: the potential of integrating digital technology and health belief model. Tuberc Respir Dis (Seoul) 2023;86(2):82-93 [FREE Full text] [doi: 10.4046/trd.2022.0148] [Medline: 36597583]
54. Huang CW, Wee PH, Low LL, Koong YLA, Htay H, Fan Q, et al. Prevalence and risk factors for elevated anxiety symptoms and anxiety disorders in chronic kidney disease: a systematic review and meta-analysis. Gen Hosp Psychiatry 2021;69:27-40. [doi: 10.1016/j.genhosppsych.2020.12.003] [Medline: 33516963]
55. Lin MY, Liu MF, Hsu LF, Tsai PS. Effects of self-management on chronic kidney disease: a meta-analysis. Int J Nurs Stud 2017;74:128-137. [doi: 10.1016/j.ijnurstu.2017.06.008] [Medline: 28689160]

XSL•FO
RenderX

56. Zhang S, Li J, Hu X. Peer support interventions on quality of life, depression, anxiety, and self-efficacy among patients with cancer: a systematic review and meta-analysis. Patient Educ Couns 2022;105(11):3213-3224. [doi: 10.1016/j.pec.2022.07.008] [Medline: 35858869]
57. Kiyak S. The relationship of depression, anxiety, and stress with pregnancy symptoms and coping styles in pregnant women: a multi-group structural equation modeling analysis. Midwifery 2024;136:104103. [doi: 10.1016/j.midw.2024.104103] [Medline: 38986391]
58. Abdi A, Dalvand S, Vahedian-azimi A, Ghanei Gheshlagh R, Kurdi A. Prevalence of depression among Iranian patients under hemodialysis: a systematic review and meta-analysis. J Nephropathol 2018;7(4):224-232. [doi: 10.15171/jnp.2018.47]
59. Palmer S, Vecchio M, Craig JC, Tonelli M, Johnson DW, Nicolucci A, et al. Prevalence of depression in chronic kidney disease: systematic review and meta-analysis of observational studies. Kidney Int 2013;84(1):179-191 [FREE Full text] [doi: 10.1038/ki.2013.77] [Medline: 23486521]

Abbreviations

BAI: Beck Anxiety Inventory
BDI-II: Beck Depression Inventory-II
BDI-SF: Beck Depression Inventory-Short Form
CKD: chronic kidney disease
DASS: Depression Anxiety and Stress Scale
EQ-5D: EuroQol 5-Dimension
ESRD: end-stage renal disease
ESRD-AQ: End-Stage Renal Disease Adherence Questionnaire
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
HD: hemodialysis
KDQOL-SF: Kidney Disease Quality of Life–Short Form
PICOS: Population, Intervention, Comparison, Outcomes, and Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SAS: Self-Rating Anxiety Scale
SDS: Self-rating Depression Scale
SF-12: 12-Item Short Form Health Survey
SF-36: 36-Item Short Form Health Survey
WHOQOL-BREF: World Health Organization Quality of Life-BREF
WHOQOL-SF: World Health Organization Quality of Life-SF

Edited by J Sarvestan, T Leung; submitted 07.10.24; peer-reviewed by X Liu, X Fu, G Rostoker; comments to author 20.02.25; revised version received 22.02.25; accepted 03.03.25; published 26.03.25.

Please cite as:
Zhou XH, Chen H, Yang W, Wang L, Chen L, Zhu Y, Zhang Y, Shi M, Zhang Q
Efficacy of eHealth Interventions for Hemodialysis Patients: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e67246
URL: https://www.jmir.org/2025/1/e67246
doi:10.2196/67246
PMID:

©Xu-Hua Zhou, Hui Chen, Weiwei Yang, Li Wang, Lin Chen, Ying Zhu, Yingjun Zhang, Mei Shi, Qin Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Social Media and eHealth Literacy Among Older Adults: Systematic Literature Review

Chenglin Zhang[1], MSocSci; Emma Mohamad[1,2], PhD; Arina Anis Azlan[1,2], PhD; Anqi Wu[1], MSocSci; Yilian Ma[1], MSocSci; Yihan Qi[1], MSocSci

[1]Centre for Research in Media and Communication, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[2]Komunikasi Kesihatan (Healthcomm)-Universiti Kebangsaan Malaysia Research Group, Universiti Kebangsaan Malaysia, Selangor, Malaysia

Corresponding Author:
Emma Mohamad, PhD
Centre for Research in Media and Communication
Faculty of Social Sciences and Humanities
Universiti Kebangsaan Malaysia
Bangi
Selangor, 43600
Malaysia
Phone: 60 389215457
Fax: 60 389215457
Email: emmamohamad@ukm.edu.my

Abstract

Background: The advent of social media has significantly transformed health communication and the health-related actions of older adults, offering both obstacles and prospects for this generation to embrace eHealth developments.

Objective: We aimed to investigate the correlation between social media and eHealth literacy in older individuals and answer four research questions: (1) What are the specific social media behaviors (including general use behaviors and health behaviors) of older adults on social media? (2) How do these behaviors impact their eHealth literacy? (3) How does eHealth literacy influence older adults' social media behaviors? and (4) What factors influence older adults' use of social media for health-related purposes?

Methods: Using predetermined keywords and inclusion criteria, we searched Scopus, Web of Science, and PubMed databases for English-language journal articles published from 2000 to 2024, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) principles. The initial search identified 1591 publications, and after removing duplicates, 48.21% (767/1591) of publications remained. Ultimately, 1% (16/1591) of studies met the inclusion criteria. A research question–driven manual qualitative thematic analysis was conducted, guided by the categorization of social media use behaviors, the definition of eHealth literacy, and the social-ecological model to provide direction for coding and thematic analysis. In addition, attention was given to identifying unanticipated behaviors or phenomena during the coding process, and these were subsequently incorporated into the analytical framework.

Results: The results indicated that older adults' general social media use behaviors are primarily characterized by social media preferences, with 2 subthemes identified. Their social media health behaviors revealed 5 main themes and 14 subthemes. Among the primary themes, health information behavior appeared most frequently (12/16, 75%), followed by self-management (8/16, 50%). Other themes included health decision-making (4/16, 25%), telemedicine (3/16, 19%), and health interventions (2/16, 13%). Cross-thematic analysis confirmed that older adults' social media use behaviors and their eHealth literacy had a reciprocal relationship. Finally, the study revealed that the use of social media to improve eHealth literacy among older adults was influenced by individual, interpersonal, institutional or organizational, and social factors.

Conclusions: The reciprocal relationship between older adults' social media use and eHealth literacy highlights the importance of establishing a long-term positive mechanism that mutually reinforces social media health practices and eHealth literacy. Based on the findings, this review proposes key directions for efforts to achieve this goal: (1) leveraging postpandemic momentum to enhance eHealth literacy among older adults through social media, (2) reconsidering the dimensions of eHealth literacy among older adults in the context of Web 2.0, (3) actively developing age-friendly integrated social media health service platforms, (4) optimizing social media for engaging and reliable health information for older adults, and (5) integrating social support systems to foster lifelong eHealth learning for older adults.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e66058) doi:10.2196/66058

KEYWORDS

eHealth literacy; digital health literacy; older adults; social media; health information; systematic review

Introduction

Background

The continuous proliferation and extensive use of social media have deeply revolutionized the dissemination of information, transformed the health care industry, and advanced the progress of eHealth. Social media platforms refer to online platforms that enable individuals to engage in interactions and present themselves selectively, either in real time or with a delay, to both wide and specific audiences [1], which have had robust expansion in recent years. Presently, the global monthly active users of the 2 prominent social media platforms, Facebook and Twitter, have attained sums of 3.06 billion and 2.5 million accordingly [2]. Especially during the COVID-19 pandemic, the adoption of social media in health areas was maximized. In the presence of lockdown and social distancing measures, social media emerged as the most effective platform for individuals to obtain health-related information [3,4]. The introduction of social media signaled a fundamental change in the field of eHealth communication. Its widespread use and robustness have established social media as a potential platform for the public to obtain health information [5,6], improve their health behaviors [7], promote social support [8-10], manage diseases, and choose health care institutions [11].

For older adults, who are often marginalized in digital innovation, social media offers a potential solution to bridge the health communication gap and promote active aging. According to the Pew Internet report [12] released in 2021, older adults are among the fastest-growing groups using social media. Moreover, social media use among older adults has grown further with the outbreak of the COVID-19 pandemic [13]. With the increasing use of social media among older adults, their opportunities to engage in eHealth are also expanding. Unlike younger individuals, who prefer a variety of software options, older adults tend to favor a single platform [14]. Therefore, compared to other specialized digital health platforms, social media serves as a more familiar and user-friendly integrated platform for health information, social support, and health care services for older adults. This effectively lowers the barriers for older adults to adopt eHealth technologies. Recently, some research has indicated that social media can effectively enable older individuals to access health information and knowledge, thereby improving their health management [15,16]. Moreover, the experience of social media has been found to be associated with older adults' willingness to adopt telemedicine [17].

However, it does not mean that there is no threshold for older adults to truly benefit from eHealth technology via social media. eHealth literacy, which is a crucial determinant of an individual's use of health technology [18], is an essential competency that older adults need to possess to successfully integrate into the eHealth trend. The notion of eHealth literacy, initially introduced by Norman and Skinner [19], refers to the "ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." However, the definition was presented in the context of Web 1.0, ignoring the internet-based nature of technology in Web 2.0. The definition has subsequently been updated by several scholars [20-26]. The interactivity of eHealth systems is an important element that most scholars focus on in the new definition. In particular, Paige et al [26] developed the concept and content of eHealth literacy based on the transactional model of communication, further emphasizing the interactivity factor. They define eHealth literacy as "...the ability to locate, understand, exchange, and evaluate health information from online environments in the presence of dynamic contextual factors and to apply the knowledge gained across ecological levels for the purposes of maintaining or improving health" [26]. Their conceptualization not only preserves the fundamental elements of conventional eHealth literacy but also underscores that *communication* refers to the capacity to establish connections and identities with other internet users by exchanging health-related information, which sheds light on our understanding of eHealth literacy based on the setting of Web 2.0.

eHealth literacy profoundly influences the effectiveness of an individual's use of social media to engage in eHealth. Because the credibility of information on social media is lower than that of professional websites, eHealth literacy, particularly critical health literacy, is essential to help evaluate the reliability of the information [27,28]. In addition, high eHealth literacy enables individuals to be better informed and equipped to access, share, and discuss health-related information and issues [29] and gain better satisfaction with the health information [30]. Therefore, eHealth literacy is a crucial skill that older adults must possess to successfully integrate into the eHealth trend through social media.

However, a harsh reality is that eHealth literacy among older people is generally found to be low compared to the younger generation [31-33]. Research has found that lower eHealth literacy makes it challenging for older adults to participate in eHealth, widening the digital divide in eHealth [34-36]. Older adults are perceived to be targets of misinformation and fraud on social media platforms and are more vulnerable to online scams [37]. They are very concerned about their privacy and security when sharing information on social media, and they consider this conduct to be highly risky [37,38]. Moreover, older adults tend to lack the perceived usefulness and ease of use of new technologies, making them slower to adopt these technologies compared to younger individuals [39], and thus they are more likely to encounter technical difficulties. These challenges not only prevent older people from truly benefiting from eHealth but may also lead to a digital divide in eHealth.

Another trend worth noting is the significant potential of social media itself to enhance eHealth literacy. Previous studies have

XSL•FO
RenderX

found a positive correlation among individuals' proficiency in using social media, their level of activity, and the number of social media platforms they use with their eHealth literacy [9,10]. In particular, the frequent use of social media to search for and engage with health information holds great potential for improving individual eHealth literacy [40]. Furthermore, the social support provided by social media can effectively enhance individuals' understanding of health information and their ability to make informed health decisions, thereby improving eHealth literacy [9]. Social media is also an essential tool for eHealth literacy interventions. For example, Jiang [41] found that health interventions implemented through community-based, proactive health management apps on social media not only increased patients' engagement in health management but also improved their eHealth literacy. Similar findings have been observed in studies involving older adults. Research indicates a significant association between older adults' health-related behaviors on social media and their eHealth literacy levels [42]. A notable characteristic is the heightened use of social media and enhanced eHealth awareness among older individuals following the COVID-19 pandemic [13,43,44].

Therefore, within the context of Web 2.0, a highly promising approach to effectively integrating older adults into the eHealth landscape is to establish a virtuous cycle where their effective use of social media for health-related behaviors reinforces their eHealth literacy. On the one hand, older adults can use familiar social media platforms to access health information and engage in health management or health interventions, gradually enhancing their eHealth literacy through practice. On the other hand, as their eHealth literacy improves, their ability to effectively use social media for eHealth engagement and their satisfaction with these interactions will also increase, ultimately bridging the digital divide in the eHealth domain.

To achieve this goal, it becomes urgent to explore the relationship between social media and eHealth literacy among older adults. Understanding older adults' social media use patterns—such as frequency of use and platform preferences—along with the impact of health information–seeking behaviors, health management, and health interventions conducted via social media on their eHealth literacy, is essential for identifying both the potential and barriers of social media in promoting health and enhancing eHealth literacy for this population. Simultaneously, examining the influence of eHealth literacy on older adults' ability to effectively use social media for health-related engagement can help identify current gaps in their eHealth literacy and propose targeted interventions. By deeply grasping the relationship between social media and eHealth literacy, older adults can comprehensively and systematically build a social media–based eHealth integration platform that meets their level of eHealth literacy. Ultimately, a virtuous cycle of mutual promotion between social media and eHealth literacy of older individuals will be formed.

Existing systematic literature studies on eHealth literacy in older individuals have mostly concentrated on examining the determinants that impact eHealth literacy in this population [45] or the intervention [32,46]. Some studies on social media and older adults' health primarily focus on the impact of social media on their mental health [47,48]. However, few studies systematically and comprehensively review the behaviors of older adults using social media for health purposes, nor do they summarize the correlations between these behaviors and health outcomes. In particular, research on behavioral patterns and preferences of older adults engaging with health-related content on social media; the role of platform design in promoting health literacy; how older adults' eHealth literacy influences their health-related behaviors on social media; and how these behaviors, in turn, affect their eHealth literacy is limited. Without a clear understanding of these dynamics, the significance and potential of social media as a tool for enhancing older adults' eHealth literacy and bridging the digital divide in eHealth remain insufficiently elucidated. This gap may ultimately turn social media from a promising tool for improving older adults' health into a potential weapon that harms this vulnerable population. Furthermore, with the outbreak and end of the COVID-19 pandemic, the user landscape of social media and the developmental landscape of eHealth technology have both changed, necessitating a reassessment of social media use patterns and older individuals' eHealth literacy.

Research Questions

This systematic literature review addressed these gaps by focusing on the correlation between older adults' social media behavior, including general social media use behavior (eg, social media use patterns or habits) and social media health behaviors (eg, health information behaviors, self-management, or interventions), and their level of eHealth literacy. Specifically, this study aimed to answer the following research questions (RQs):

RQ1: What are the specific social media behaviors (including general use behavior and health behaviors) of older adults on social media?

RQ2: How do these behaviors impact their eHealth literacy?

RQ3: How does eHealth literacy influence older adult's social media behaviors?

RQ4: What factors influence older adults' use of social media for health-related purposes?

Methods

Overview

Through a comprehensive literature review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria [49], this review investigated the relationship between social media and older individuals' eHealth literacy and associated characteristics. PRISMA checklist is provided in Multimedia Appendix 1. We did not register the procedure for this review to a systematic review database.

Data Resources and Search Strategy

Scopus, Web of Science, and PubMed were selected as databases for this review. The online database search was conducted on May 17, 2024. Although the notion of eHealth literacy was initially presented in 2006 [19], there is some previous literature that may use other vocabularies, such as computer health literacy

XSL•FO
RenderX

and internet health literacy, to refer to eHealth literacy. To ensure the thoroughness of the assessment, the literature search was conducted within the period framework of 2000 to 2024. Search terms were generated around the following research themes: "social media," "eHealth literacy," and "older people." Because eHealth literacy is a broad concept, this study not only covers the synonyms of eHealth literacy but also expands on the definition provided by Norman [19] and Paige [26] by adding related content about what eHealth literacy encompasses, such as health literacy, digital literacy, information literacy, and media literacy. Furthermore, the objective of this study was to investigate the correlation between social media and the eHealth literacy of older adults, as well as associated elements. Consequently, the primary focus was given to social networking platforms that facilitate communication and interaction with the public regarding health matters, health information, and health management. Therefore, the definition provided by Carr and Hayes [1] was adopted, and the latest global user activity rankings were used to filter the top-ranked social media as search extensions [2]. All keywords were searched in the database in combination with Boolean logic. A comprehensive list of search queries is presented in Textbox 1. A separate search string for each database is provided in Multimedia Appendix 2.

Textbox 1. Keywords search strings (the asterisk is a wildcard used to include multiple word variations in the search query).

Older adults

- "old*" OR "elder*" OR "senior*" OR "aging*" OR "ageing*" OR "aged*" OR "geriatric*" OR "gerontolog*" OR "retiree*" OR "baby boomer" OR "pensioner*" OR "silver generation"

AND

eHealth literacy

- "eHealth literacy" OR "eHealth literacy" OR "telehealth literacy" OR "mHealth literacy" OR "health literacy" OR "medic* literacy" OR "digital literacy" OR "information literacy" OR "media literacy" OR "internet literacy" OR "computer literacy" OR "virtual literacy" OR "intelligent literacy" OR "technolog* literacy" OR "online literacy" OR "web based literacy" OR "Web 2.0 literacy" OR "ICT literacy" OR "Information and Communication Technology literacy"

AND

Social media

- "social media" OR "social network*" OR "social platform*" OR "social web*" OR "web 2.0" OR "online social*" OR "digital social*" OR "virtual social*" OR "social channels" OR "social apps" OR "community media*" OR "community platform" OR Facebook OR "YouTube" OR WhatsApp OR Instagram OR WeChat OR "TikTok" OR Douyin OR Telegram OR Snapchat OR Kuaishou OR "Sina Weibo" OR QQ OR Twitter OR Pinterest

Study Selection Criteria

The selected papers for review had to adhere to the pre-established inclusion criteria. First, the included papers should have been primary research articles written in English and published between January 2000 and May 2024. Other document types, such as reviews, editorials, conference papers, letters, notes, commentaries, book chapters, books, or study protocols, were excluded from the considered literature. Second, the participants under study must have been older individuals. Given the absence of a universally defined age for older individuals, the United Nations generally designates individuals aged ≥60 years as older individuals [50]. However, some articles use national retirement ages as criteria for the research population [42] or define older people as those aged ≥50 years from the perspective of cognitive aging [51]. Considering the diversity in defining older people and the current minimum retirement age of 50 worldwide [52], this review included participants aged ≥50 years to ensure a broader spectrum of articles. Furthermore, it is anticipated that the papers will pertain to eHealth literacy among older persons who are active on social media and investigate the elements that impact their use of social media to enhance their eHealth literacy. Considering this criterion, the following articles were omitted: (1) discussions concerning phone calls, SMS text messages, email, web portals, and telemedicine or mobile health (mHealth) without social media engagement; (2) articles that were not based on health contexts, such as internet literacy, media literacy, digital literacy, and information literacy, for older adults; and (3) articles that were based on internet or social media behaviors of older adults unrelated to eHealth literacy. Furthermore, this paper identified a correlation between eHealth literacy and the social media activities of older people. Details on the particular inclusion and exclusion criteria can be found in Textbox 2.

XSL•FO
RenderX

Textbox 2. Study selection criteria.

Inclusion criteria

- Research topic: eHealth literacy and social media use in the older population
- Research content: articles that investigate the influence and associated variables of social media on eHealth literacy in the older population; articles that examine the effect of older individuals' eHealth literacy on their social media activities
- Target group: various community groups, including marginalized migrant populations, retirees, and those aged ≥50 years
- Article type: primary research journal articles
- Language: English
- Timeline: between 2000 and 2024

Exclusion criteria

- Research topic: topics such as internet literacy, media literacy, digital literacy, and information literacy among older adults in non–health-related contexts; health behaviors of older people based on nonsocial platforms, such as phone calls, SMS text messages, email, web portals, and telemedicine or mobile health without social media engagement; internet or social media behaviors of older adults unrelated to eHealth literacy
- Research content: articles in which the correlation between eHealth literacy and social media use among older people was not explicitly referenced
- Target group: children, adolescents, middle-aged group, and other young groups aged <50 years; study population without a clear age range
- Article type: reviews, editorials, conference papers, letters, notes, commentaries, book chapters, books, or study protocols
- Language: non-English
- Timeline: before 2000

Literature Screening

There were 3 steps in literature screening. The first step was removing duplicates. Articles identified in the database were filtered into Microsoft Excel and duplicates were removed. Following this was the process of screening the title and abstract. A preliminary screening was conducted by 2 reviewers (CZ and AW) who independently identified important themes in the title and abstract. The age range of the study population was first identified, and studies where the population was not aged ≥50 years were excluded. If the age range could not be identified from the title and abstract, the article was reviewed for relevance to the topic, and theme-related studies were included for full-text review; otherwise, they were excluded. For example, the study by Willis [53] was included for full-text review because the abstract mentioned interviews with online health community members without specifying their ages. However, the full-text review revealed that participants were aged 37 to 85 years, resulting in this study's exclusion.

The last step was full-text screening. Following the initial screening, the selected articles were reviewed in full text. During this process, 2 independent researchers carefully and comprehensively reviewed the full texts of the selected studies. This stage prioritized the exclusion of literature that did not fit the language, type of article, and age of the study population. Among them, the study conducted by Aung et al [54] was omitted from consideration because it was only a research plan.

Subsequent full-text reviews primarily excluded literature that did not address social media or was not explicitly about social media, as well as literature from nonhealth contexts. For example, some (39/153, 25.5%) articles, such as Mao et al [55] and Evans et al [56] discussed the association between telemedicine or mHealth and eHealth literacy in older adults, but these technologies did not involve social media platforms and were, therefore, excluded. Moreover, a few (4/153, 2.6%) articles discussed internet behaviors and eHealth literacy among older individuals, but it was unclear whether internet behavior related to social media was excluded [57-60]. In addition, some (3/153, 2%) articles, while exploring important components of eHealth literacy in older adults, such as digital literacy, information literacy, or media literacy, were not based on a health context and were, therefore, excluded [61-63]. Notably, several (3/153, 2%) papers examined the influence of social media on alleviating loneliness among older individuals, which did not address eHealth literacy and were excluded [64-66]. Finally, some (3/153, 2%) articles, while dealing with social media and eHealth literacy, did not allow the interaction between the 2 to be judged and were, therefore, excluded [67-69].

For articles that met the inclusion criteria in the full-text review, we annotated the study population, the research methodology, and the relevant factors in preparation for subsequent data extraction. To ensure the reliability of the study, a cross-review was conducted, and YM was introduced to resolve any discrepancies.

Quality Appraisal

This review used the quality assessment (QA) tool developed by Sirriyeh et al [69] to evaluate the quality of the included studies. The QA tool establishes 6 QAs and assigns scores on a 3-point scale from 0 to 1 (0=no, 0.5=partly, and 1=yes). Studies with an overall score of >3 points are recognized as being of sufficient quality. Two researchers (CZ and AW) independently scored the included studies, and the sum of the scores obtained was used to assess the study quality. Inconsistent results were resolved by third-party (YM) intervention.

XSL•FO
RenderX

Data Extraction and Analysis

To capture the basic information and trends of the studies included, we developed a standardized table summarizing the fundamental characteristics of the research. This table extracted details, such as the authors, publication year, country, participants' age, and research methods, from the included literature. Furthermore, to clarify the roles played by different types of social media, we extracted and categorized the social media platforms mentioned in the articles. In addition, as no restrictions were imposed on the research methods of the included studies, which covered qualitative, quantitative, and mixed methods approaches, thematic analysis [70] was used to describe and analyze the literature. According to Braun and Clarke [70], thematic analysis is a flexible research tool with a high degree of theoretical freedom, making it applicable within various theoretical frameworks and capable of providing rich and detailed data descriptions.

We conducted a manual, qualitative content analysis guided by the RQs, integrating classifications of social media health behaviors, definitions of eHealth literacy [26], and the social-ecological model (SEM) [71] to inform the coding and thematic analysis. At the same time, we paid attention to identifying new behaviors or phenomena not initially anticipated and incorporated them into the analytical framework. The qualitative thematic analysis framework is illustrated in Figure 1.

Figure 1. Qualitative thematic analysis framework map. RQ: research question.

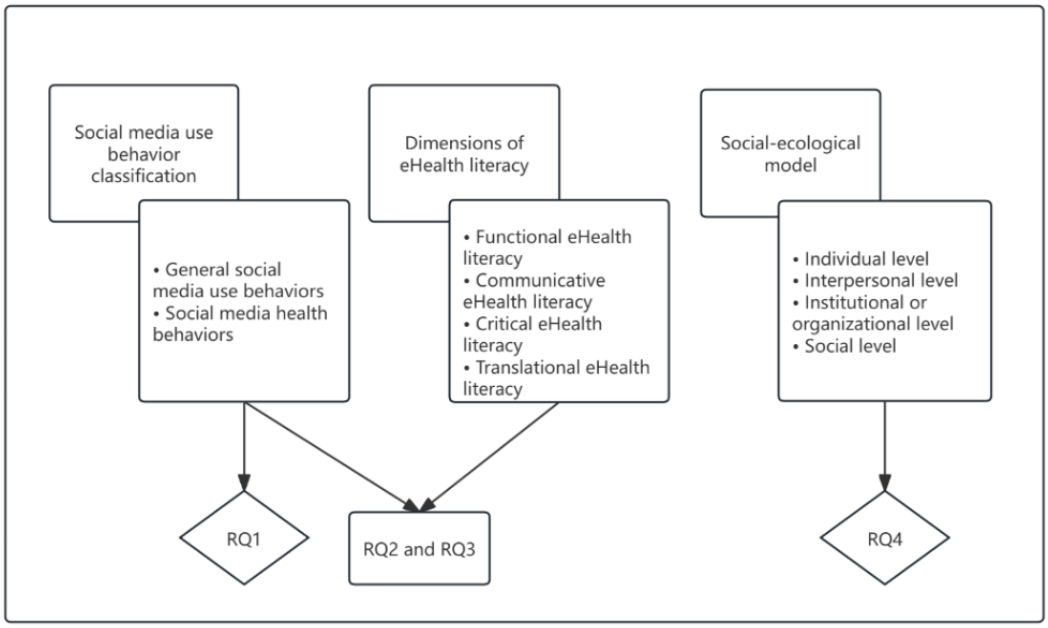

Specifically, guided by RQ1, in addition to focusing on older adults' general social media behaviors (including use habits, such as frequency of use and platform preferences), we reviewed the literature on the application of social media in the health domain. This allowed us to extract consumer-related behavioral manifestations resulting from the use of social media in health contexts. As both health consumers and professionals can act as providers or recipients of support during interactions on health platforms [72], we examined health behaviors not only from the perspective of consumers or patients but also from the perspectives of health care professionals or other organizations interacting with consumers or patients. Our analysis identified key consumer-related applications of social media in health, including health information (searching, sharing, and evaluating); telemedicine; health management; health education; health promotion; and health intervention [73-76]. These applications served as the foundation for our coding and thematic extraction. To ensure rigor, we used the "conceptual matrix" proposed by Webster and Watson [77] (Multimedia Appendix 3 [13,14,42,44,51,69,78-88]) to examine themes and systematically record emerging themes during the analysis process.

RQ2 and RQ3 explored the correlation between older adults' health behaviors on social media and their eHealth literacy. We adopted the definition of eHealth literacy provided by Paige et al [26] as the guiding framework for thematic analysis. Paige et al [26] delineated eHealth literacy into four dimensions:

1. Functional eHealth literacy: basic skills in reading and writing (typing) about health to effectively function on the internet
2. Communicative eHealth literacy: the ability to collaborate, adapt, and manage communication about health in social online environments with multimedia
3. Critical eHealth literacy: the ability to evaluate the credibility, relevance, and risks of sharing and receiving health information on the internet
4. Translational eHealth literacy: the ability to apply health knowledge gained from the internet across diverse ecologic contexts

We extracted descriptions from the literature relevant to these 4 dimensions and performed theoretical mapping to determine the eHealth literacy dimensions they pertained to. For studies that only referred to eHealth literacy as a general concept, we assumed that they encompassed all 4 dimensions.

XSL•FO
RenderX

Finally, we conducted a cross-analysis between older adults' social media use behaviors—including general and health-related behaviors (as identified in the themes extracted based on RQ1)—and the 4 dimensions of eHealth literacy. This analysis aimed to uncover the relationships between the 2, focusing on whether older adults' social media use behaviors enhance or hinder their eHealth literacy and whether their eHealth literacy strengthens or diminishes their health-related behaviors on social media. Given the diversity of research methods in the included studies, our qualitative content analysis not only focused on explicit conclusions from studies that directly examined these relationships but also considered implicit or related findings within the research.

Guided by RQ4, we examined the factors influencing older adults' use of social media to access health information or health services in each article. Because these factors included not only barriers and facilitators but also neutral factors, we adopted the SEM to enable a more comprehensive extraction and integration of these influencing factors. SEM, developed by Bronfenbrenner [71], provides a systematic framework for understanding the impact of various environmental elements on individual behavior and development. According to SEM, the influencing factors are categorized into 4 dimensions: individual, interpersonal, institutional or organizational, and social levels.

CZ and AW analyzed all the articles and extracted the relevant themes. AAA and YQ reviewed the full text of the articles to validate the extracted themes. All conflicts were resolved by EM. Finally, multiple meetings were held to discuss the results of the thematic analysis and reach a consensus.

Results

Search Results

Search results and data collection process are shown in the PRISMA flow diagram in Figure 2. Initially, 1591 articles were identified (Scopus, 840/1591, 52.79%; Web of Science, 501/1591, 31.49%; and PubMed, 250/1591, 15.71%). A total of 48.21% (767/1591) of articles were screened by excluding certain articles based on the exclusion criteria, such as article type, language, and date, and after duplicates were removed. Subsequently, 38.6% (614/1591) of articles were eliminated following the screening of the title and abstract, while 8.6% (137/1591) of articles were eliminated after the assessment of the entire text. Publications were eliminated throughout the full-text screening process for the following reasons: (1) non-English (3/153, 2%), (2) note (1/153, 0.6%), (3) commentary (1/153, 0.7%), (4) study protocol (1/153, 0.7%), (5) younger population (65/153, 42.5%), (6) non–social media (39/153, 25.5%), (7) outside health context (24/153, 15.7%), and (8) not related to the correlation between eHealth literacy and older population (3/153, 2%). Although a considerable amount of research was identified at the outset, many articles used social media only as a tool for the recruitment of research participants or for distributing questionnaires, without discussing social media, and were, therefore, excluded.

Figure 2. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram depicting the process of study selection.

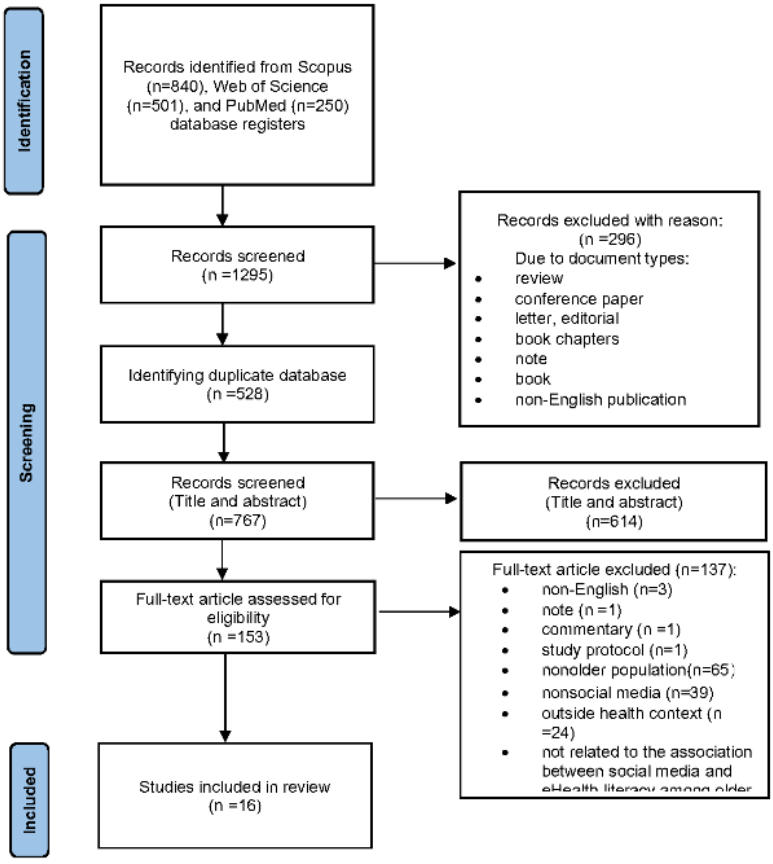

XSL•FO
RenderX

Description of Studies

Comprehensive information regarding the included research is provided in Table 1.

Generally, most (15/16, 94%) articles were published between 2020 and 2024, namely, after the COVID-19 pandemic. Only 1 (6%) study was conducted in 2015 [78]. Nearly two-thirds (11/16, 69%) of the studies were from Asian countries, including China (6/16, 38%) [14,42,44,79-81], Korea (2/16, 13%) [82,83], Thailand (2/16, 13%) [13,84], Singapore (1/16, 6%) [59]. In total, 13% (2/16) of studies were from the United States [85,86], and 1 (6%) study was from Italy [87]. Concerning the age definition of older individuals, except for 2 (13%) papers that did not provide a specific age criterion [87,88], the age definition ranges of the studies mainly used ≥50 years (5/16, 31%) [14,79,80,84,86] or ≥60 years (5/16, 31%) [13,44,80,83,84], and a few studies used ≥65 years (2/16, 13%) [82,85] or ≥55 years (1/16, 6%) [81]. Moreover, 1 (6%) Chinese study used the retirement age as the range of the study population, that is, ≥55 years for women and ≥60 years for men, respectively [42]. Most (12/16, 75%) studies were quantitative (Textbox 2), including surveys (10/16, 62%) [13,42,78-80,82,84,86-88] and experiments (2/16, 13%) [51,83], while 2 (13%) studies used mixed methods [44,85]. Only 2 (13%) studies were qualitative, using interviews [14,81].

Table 1. Study characteristics.

Study	Country	Population age (y)	Research method
Matchanova et al [51], 2024	United States	• 50	Experiment
Ye [80], 2024	China	• >60	Survey
Oh et al [86], 2023	United States	• >50	Survey
Vitolo et al [87], 2023	Italy	• 66 to 84	Survey
Jang et al [82], 2023	Korea	• >65	Survey
Kachentawa [13], 2023	Thailand	• >60	Survey
Lee and Ryu [83], 2023	Korea	• >60	Experiment
Wang and Zhang [42], 2023	China	• Men: >60 • Women: >55	Survey
Liu et al [44], 2022	China	• >60	Mixed methods (survey and interview)
Chai [81], 2022	China	• >55	Interview
Tan et al [88], 2022	Singapore	• 56 to 75	Survey
Chen et al [85], 2021	United States	• >65	Mixed methods (experiment and interview)
Wu and Yu [14], 2021	China	• >50	Interview
Ubolwan et al [84], 2020	Thailand	• >60	Survey
Shang and Zuo [79], 2020	China	• >50	Survey
Tennant et al [78], 2015	United States	• >50	Survey

Regarding features related to social media, most (11/16, 69%) articles discussed a wide range of social media platforms [13,44,78-82,84,86-88], and a few (5/16, 31%) articles explored the role of individual social media only [14,42,51,83,85], covering WeChat (2/16, 13%) [83], Facebook (1/16, 6%) [51], The BAND app (NAVER Corp) (1/16, 6%) [83], and Virtual Online Communities for Aging-Life Experiences (1/16, 6%) [85]. Moreover, social media was mostly (11/16, 69%) regarded as both a medium for health information and health-related relationships [13,42,44,51,78,79,81,82,84,86,88]. Furthermore, 1 (6%) study examined the dual role of social media as a medium for health-related interactions and information [80]. Social media was discussed as a health intervention or health education tool (2/16, 13%) [83,85], and a health management and communication tool (1/16, 6%) [14]. Only 1 (6%) article did not discuss the role of social media [87].

To further capture the characteristics of older adults in using social media platforms, this review extracted and categorized the social media platforms mentioned in the studies included (Table 2). Kordzadeh [72] categorized social media platforms applied in health care into 2 types: online social networks and virtual health communities (VHCs). The former includes large-scale, general-purpose internet platforms, such as

XSL•FO
RenderX

Facebook, Twitter, Instagram, and YouTube, which are designed for broad collaboration among internet users. In contrast, the latter consists of social media sites specifically designed for individuals to discuss health-related topics.

On the basis of the included studies, most (11/16, 69%) explored the role of general social media in the health of older adults without specifying the particular platforms. Among the studies that addressed online social network platforms, the most commonly mentioned were WeChat and Facebook. These platforms are widely popular among older adults in their respective regions and are frequently used for accessing and sharing health-related information. Conversely, VHC platforms exhibited a more distinct health-focused orientation. Only 2 (13%) studies in the included literature specifically examined VHC platforms, namely BAND and Virtual Online Communities for Aging-Life Experiences, highlighting their value as tools for online support groups and health interventions [83,85].

Table 2. Social media platforms and categorization included in literature (N=16).

Social media	Social media types		Studies	Studies, n (%)
	OSN[a]	VHC[b]		
Unspecified	✓	✓	Ye [80], 2024; Oh et al [86], 2023; Vitolo et al [87], 2023; Jang et al [82], 2023; Kachentawa [13], 2023; Liu et al [44], 2022; Chai [81], 2022; Tan et al [88], 2022; Ubolwan et al [84], 2020; Shang and Zuo et al [79], 2020; and Tennant et al [78], 2015	11 (69)
WeChat	✓	—[c]	Wang and Zhang [42], 2023 and Wu and Yu [14], 2021	2 (12)
Facebook	✓	—	Matchanova et al [51], 2024	1 (6)
BAND	—	✓	Lee and Ryu [83], 2023	1 (6)
VOCALE[d]	—	✓	Chen et al [85], 2021	1 (6)

[a]OSN: online social network.

[b]VHC: virtual health community.

[c]Not applicable.

[d]VOCALE: Virtual Online Communities for Aging-Life Experiences.

Quality of the Included Studies

According to the results of the quality appraisal (Multimedia Appendix 4 [13,14,42,44,51,69,78-88]), the scores of the included studies ranged from 3.5 to 5.5, with all of them exceeding 3, which is considered to be of sufficient quality. Moreover, none of the studies obtained a score of 0 on the individual evaluation items. This suggests that the included studies effectively elucidated the purpose, interest, utility, methodological process, approach concept, comparison and measurement with other similar studies, and limitations of the study on a high level of clarity.

Social Media Use Behaviors of Older Adults

As this study focused on the relationship between social media and eHealth literacy among older adults, an analysis of their social media use behaviors was conducted. This thematic analysis included both the general use behaviors of older adults and their health-related behaviors stemming from social media use. Table 3 presents the results of the thematic analysis, identifying 2 themes related to general social media use behaviors and 5 themes associated with health-related behaviors.

XSL•FO
RenderX

Table 3. Social media use behaviors of older adults (N=16).

Theme, description, and subtheme	Sample quotes	Studies	Studies, n (%)
Social media use habit			
The frequency of older adults' social media use, their platform preferences, and so on			
Frequency of social media use	"Frail and pre-frail patients used digital tools less frequently and accessed the Internet less frequently compared to robust patients." [87]	Vitolo et al [87], 2023; Ubolwan et al [84], 2020; and Tennant et al [78], 2015	3 (19)
Social media platform preference	"The most widely adopted function for senior citizens was instant messaging... Compared with the diversity of software applications that the younger generation use, older adults were more interested in a limited range of software affordance which benefited them most." [14]	Wu and Yu [14], 2021	1 (6)
Health information			
The fundamental behaviors of older adults related to health information on social media, including searching, sharing, learning, and evaluating health information			
Health information acquisition	"...participants obtained information about COVID19 and general health from various resources...." [44]	Ye [80], 2024; Oh et al [86], 2023; Jang et al [82], 2023; Kachentawa [13], 2023; Wang and Zhang [42], 2023; Liu et al [44], 2022; Chai [81], 2022; Tan et al [88], 2022; and Tennant et al [78], 2015	9 (56)
Health information sharing	"Among older adults, a higher likelihood of sharing false COVID-19 headlines was associated with lower verbal IQ and numeracy skills." [51]	Ye [80], 2024; Matchanova et al [51], 2023; Kachentawa [13], 2023; Lee and Ryu [83], 2023; and Wang and Zhang [42], 2023	5 (31)
Health information evaluation	"...interviewees were able to evaluate pandemic rumors as false information, but they may not understand the underpinning scientific logic." [81]	Matchanova et al [51], 2023; Liu et al [44], 2022; Chai [81], 2022; Tan et al [88], 2022; and Tennant et al [78], 2015	5 (31)
Health knowledge learning	"Abundant health knowledge resources are available on social media to facilitate technology-enhanced knowledge learning among older adults." [79]	Shang and Zuo [79], 2020	1 (6)
Health information resource trust	"...trust in formal sources of information on its own positively predicts vaccination status..." [88]	Tan et al [88], 2022	1 (6)
Frequency of health information use	"...the frequency of using online health information also has a positive influence on the three levels of e-health literacy of older adults." [80]	Ye [80], 2024	1 (6)
Self-management			
Older adults are engaged in social media–based behaviors to manage symptoms, treatments, and lifestyles			
Health prevention	"All the older adults consciously followed the prevention and control strategies they had learned..." [44]	Jang et al [82], 2023; Kachentawa [13], 2023; Chai [81], 2022; Liu et al [44], 2022; Chen et al [85], 2021; and Wu and Yu [14], 2021	6 (38)
Lifestyle management	"In this study, participants expressed greater appreciation of physical activity and the need to be more proactive about self-management, which could lead to a healthier lifestyle." [85]	Ye [80], 2024; Chen et al [85], 2021; Chai [81], 2022; and Liu et al [44], 2022	4 (25)
Medical management	"During the lockdown, citizens managed their living needs and medical requirements by the 'grid managed' online community facilitated by SNS chat groups..." [14]	Wu and Yu [14], 2021 and Oh et al [86], 2023	2 (12)
Telemedicine			
Older adults access health care through social media			

XSL•FO
RenderX

Theme, description, and subtheme	Sample quotes	Studies	Studies, n (%)
Communicating with health care providers	"Among the health communication factors, social media use, face-to-face PPC, electronic PPC, and use of electronic communication through electronic patient platforms were facilitating communication factors of CRC screening among NHWs." [86]	Oh et al [86], 2023	1 (6)
Access to health care	"A face-to-face visit with health care providers was limited due to home quarantine and isolation, so older adults also learned to seek medical care through the Internet." [44]	Liu et al [44], 2022 and Wu and Yu [14], 2021	2 (12)
Health decision-making			
Older adults integrate relevant information from social media with their personal knowledge and beliefs to make health decisions that meet their needs			
Medical decision	"...those who placed greater trust in social media being less likely to have received at least 1 dose of the vaccine." [88]	Tan et al [88], 2022 and Kachentawa [13], 2023	2 (12)
Addressing general health issues	"With regard to problem-solving, participants mentioned effects they experienced, including that the discussion affected participants' behavior, either in terms of provoking a more active approach to addressing health issues, or considering the use of the skills in their own lives." [85]	Ye [80], 2024 and Chen et al [85], 2021	2 (12)
Health intervention			
Systematic behavioral changes to improve health outcomes among older adults using social media			
Online group interventions	"...a gamified walking program, called the MSG walking program (mHealth devices and Social media apps using Gamified walking program), was designed to enable socializing in conjunction with mHealth apps and devices to operate an online walking program for retired older adults during COVID-19." [83]	Lee and Ryu [83], 2023 and Chen et al [85], 2021	2 (12)

The general social media use behaviors of older adults describe their use patterns, including frequency of use, platform preferences, and other related habits. Among the included studies, 19% (3/16) [78,84,87] investigated the time spent or frequency of social media use by older adults. Wu and Yu [14] explored platform preferences among older adults, noting their inclination toward instant text messaging features on social media. Unlike younger users, older adults tended to prefer platforms with limited features and familiar interfaces, rather than overly complex or feature-rich social media platforms.

For the thematic analysis of older adults' health behaviors on social media, we adopted a combined approach of using predefined themes and identifying emergent themes. Drawing from previous research on the application of social media in health communication, certain themes were established in advance. These themes were further refined and expanded during the full-text review stage. Ultimately, 5 primary themes emerged: health information, health interventions, health decision-making, self-management, and telemedicine—all of which aligned with our predefined themes.

However, in the extraction of subthemes, we uncovered more nuanced information. For instance, while the predefined theme of health information behavior initially encompassed information-seeking, sharing, and evaluation behaviors, the coding process revealed additional discussions by researchers on older adults' learning of health knowledge, trust in health information sources, and frequency of health information use. For example, the study by Tan et al [88] examined how trust in different information sources influenced older adults' decisions regarding COVID-19 vaccination [88].

Notably, health information behavior was the most frequently discussed theme, appearing in 75% (12/16) of the studies. Within this category, particular emphasis was placed on information-seeking (9/16, 56%), sharing (5/16, 31%), and evaluation behaviors (5/16, 31%). Furthermore, social media information also impacted older adults' health decision-making. These decisions were not only limited to medical choices, such as vaccination [13,88], but also extended to addressing daily health-related challenges [80,85].

Self-management emerged as the second most frequently discussed theme (8/16, 50%), referring to older adults' health management behaviors related to their diseases and lifestyles through social media. Among the aspects of self-management, health prevention was the most extensively explored subtheme (5/16, 31%), primarily due to the prevalence of the COVID-19 pandemic as the context in many studies. In addition to health prevention, older adults use social media for managing their lifestyles [81,85] and medical needs [14,86].

Finally, social media also serves as an important platform for older adults to engage in telemedicine and for implementing health interventions targeted at this population. Health interventions refer to planned efforts to modify older adults' health behaviors via social media to improve their health outcomes. For instance, Lee and Ryu [83] and Chen et al [85] conducted online group health interventions for older adults using VHC platforms, demonstrating the potential of such platforms in facilitating health-related behavioral changes.

XSL•FO
RenderX

Correlation Between Social Media Use and the eHealth Literacy in Older Individuals

Overview

Figure 3 [13,14,42,44,51,78,80-88] presents the comprehensive correlation between social media use and the eHealth literacy in older adults. The included studies revealed that social media and older adults' eHealth literacy are mutually influencing and interacting. In total, 56% (9/16) of the articles examined the influence of social media on the eHealth literacy of older individuals [13,14,78,82-87], 13% (2/16) only discussed the value of older adults' eHealth literacy on their social media use [42,51], and 31% (5/16) covered the interplay between the two [14,44,80,81,88].

Figure 3. The correlation between social media use and eHealth literacy (eHL) in older individuals. Reference sources for these data include (9/16, 56%) [13,14,78,82-87]; (5/16, 31%) [14,44,80,81,88]; and (2/16, 13%) [42,51].

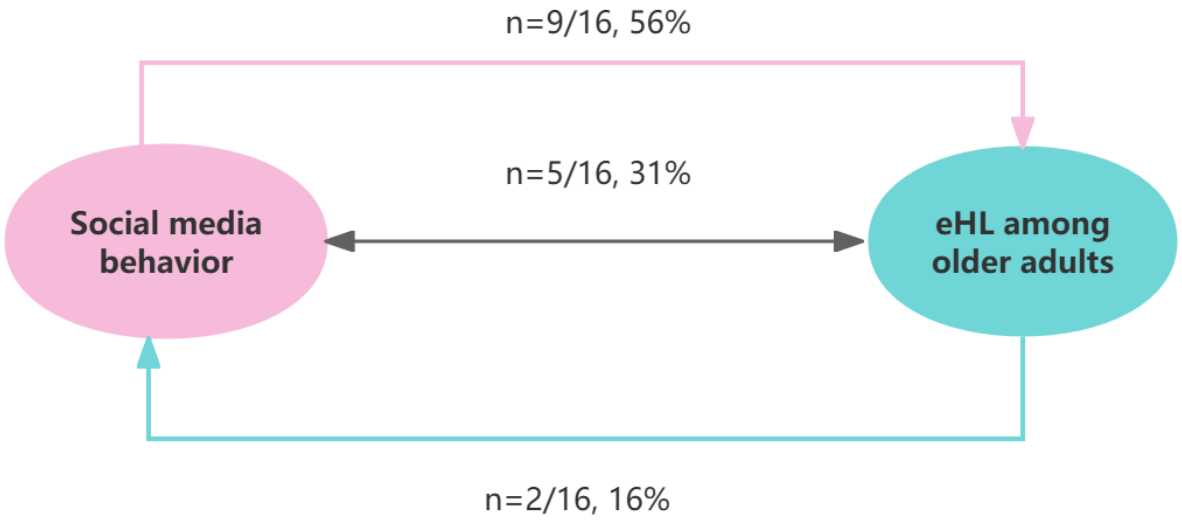

Impact of Social Media on eHealth Literacy of Older Individuals

Overview

The overwhelming majority (13/16, 81%) of the research included confirmed the impact of social media on the eHealth literacy of older individuals (Table 4), and the pathways to impact included social media use habits (4/16, 25%), social media health information behaviors (7/16, 44%), social media–based health interventions (2/16, 13%), telemedicine (2/16, 13%), and health decision-making (1/16, 6%).

Table 5 categorizes and counts the behavioral themes.

XSL•FO
RenderX

Table 4. The impact of social media on eHealth literacy (eHL) of older individuals.

Study and social media behavioral dimensions that affect eHL	Theoretical mapping: dimensions of affected eHL	Mechanism of association and key findings
Ye [80], 2024		
• Frequency of health information use • Health information acquisition • Health information sharing • Self-management (lifestyle management)	• Translational eHL: application ability, decision-making ability • Critical eHL: critical judgment ability	The social media health information behavior of older adults (including frequency of use, acquisition, and sharing of health information) and self-management behavior through online health information significantly affect their translational eHL and critical eHL.
Oh et al [86], 2023		
• Telemedicine (communicating with health care providers)	• Communicative eHL: health communication	Older adults' use of social media to communicate with health care providers enhances their communicative eHL.
Vitolo et al [87], 2023		
• Frequency of social media use	• eHL: digital health literacy	A lower frequency of social media use is correlated with lower eHL in older persons.
Jang et al [82], 2023		
• Health information acquisition	• Translational eHL: application of digital resources and health prevention	Older adults with limited sources of health information (unable to diversify their use of digital information resources, including social media, and traditional resources) exhibit lower digital resource use and health prevention behaviors, hindering their eHL to some extent, especially translational eHL.
Kachentawa [13], 2023		
• Health information evaluation • Health information sharing • Self-management (health prevention) • Health decision-making (medical decision)	• Critical eHL: understanding and evaluation of health information • Functional eHL: ability to learn new things on social media, basic technical skills for using social media, ability to solve basic problems on social media, ability to understand how family explains using social media • Translational eHL: ability to apply health information	Literacies related to social media, including media and information literacy, as well as computer and social media literacy (corresponding to critical eHL, functional eHL, and translational eHL), should be included as indicators of older adults' eHL.
Lee and Ryu [83], 2023		
• Health intervention (online group intervention)	• eHL	Social media–based, online group health interventions can provide social support for older adults and enhance their eHL.
Liu et al [44], 2022		
• Health information acquisition • Telemedicine (accessing health care services) • Self-management (health prevention)	• eHL	During the COVID-19 pandemic, older adults' use of social media for health information acquisition, telemedicine, and self-management contributed to their higher eHL regarding the COVID-19 pandemic.
Chai [81], 2022		
• Health information acquisition • Self-management (lifestyle management)	• eHL: health literacy • Translational eHL: ability to apply health information	Older adults' access to health information through social media (both directly from social media platforms and from social connections maintained by social media platforms) and self-management using this information promotes their eHL.
Tan et al [88], 2022		

XSL•FO
RenderX

Study and social media behavioral dimensions that affect eHL	Theoretical mapping: dimensions of affected eHL	Mechanism of association and key findings
• Health information trust	• Translational eHL: decision-making and applied behaviors based on health information	Older adults who are more trusting of social media and have limited access to other sources of information are likely to make poor health decisions, suggesting that inaccurate information on social media may hinder translational eHL among older adults.
Chen et al [85], 2021		
• Health intervention (online group interventions)	• eHL: health literacy and health self-efficacy • Translational eHL: applying health literacy for problem-solving and self-management	Social media–based, online group health interventions can improve eHealth literacy among older adults, as evidenced by increased health literacy, health self-efficacy, and translational eHL.
Wu and Yu [14], 2021		
• Self-management (medical and lifestyle management) • Health prevention • Platform preference	• Functional eHL and Translational eHL: ability to use online communities for health purposes • Functional eHL: perceived ease of use	Social media promotes the ability of older adults to familiarize themselves with digital technology and use digital tools for health prevention and self-management by linking their online social connections; that is, it promotes their functional eHL and translational eHL. At the same time, older people prefer the perceived ease of use of digital technologies and are more inclined to adopt familiar social media platforms, which lowers their threshold for digital technology adoption, that is, increases their functional literacy.
Ubolwan et al [84], 2020		
• Frequency of social media use	• eHL	The more frequently older adults use social media, the higher their level of eHL.
Tennant et al [78], 2015		
• Frequency of social media use • Health information acquisition • Social support (general) • Health information evaluation	• eHL • Functional eHL: computer literacy • Critical literacy: media literacy	The use of social media, the social support provided by social media, and social media–based health information behaviors can improve eHealth literacy among older adults, especially finding and evaluating health information can improve their functional literacy and critical literacy.

XSL•FO
RenderX

Table 5. Statistics on social media use behaviors affecting eHealth literacy among older adults (N=16).

Social media use behaviors	Studies	Studies, n (%)
Social media use habits		4 (25)
Frequency of social media use	Vitolo et al [87], 2023; Ubolwan et al [84], 2020; and Tennant et al [78], 2015	3 (19)
Platform preference	Wu and Yu [14], 2021	1 (6)
Health information behaviors		7 (44)
Health information acquisition	Ye [80], 2024; Jang et al [82], 2023; Liu et al [44], 2022; Chai [81], 2022; and Tennant et al [78], 2015	5 (31)
Health information sharing	Ye [80], 2024 and Kachentawa [13], 2023	2 (12)
Health information evaluation	Kachentawa [13], 2023 and Tennant et al [78], 2015	2 (12)
Health information trust	Tan et al [88], 2022	1 (6)
Frequency of health information use	Ye [80], 2024	1 (6)
Self-management		5 (31)
Health prevention	Kachentawa [13], 2023; Chai [81], 2022; Liu et al [44], 2022; and Wu and Yu [14], 2021	4 (25)
Lifestyle management	Ye [80], 2024 and Chai [81], 2022	2 (12)
Medical management	Wu and Yu [14], 2021	1 (6)
Health interventions		2 (12)
Online group interventions	Chen et al [81], 2021 and Lee and Ryu [83], 2023	2 (12)
Telemedicine		2 (12)
Communicating with health care providers	Oh et al [86], 2023	1 (6)
Accessing health care services	Liu et al [44], 2022	1 (6)
Health decision-making		1 (6)
Medical decision	Kachentawa [13], 2023	1 (6)

Social Media Use Habits

Social media use habits are themes extracted based on older adults' general social media use behaviors, which mainly concern older adults' social media use frequency and platform preferences. Of the 16 included literature, 4 (25%) papers discussed the social media use habits of older adults, 3 (19%) focused on the frequency of social media use by older adults, and 1 (6%) focused on older adults' preferences for social media platforms. The findings indicated a substantially positive correlation between social media use patterns and eHealth literacy.

The most prominent manifestation was that increased use of social media increased the eHealth literacy of older individuals. Ubolwan et al [84] included social media use in assessing the level of eHealth literacy and its correlates in the older population in Thailand. This research revealed a statistically significant positive correlation between the duration of internet or online social media use among older people and their eHealth literacy status. Similarly, Tennant et al [78] discovered that, compared to those who seldom use social media, individuals who used social media regularly had better eHealth literacy. Vitolo et al [87] discussed the adverse effects of reduced frequency of social media use on eHealth literacy in older adults from the opposite perspective, that is, frail older patients had limited use of digital technology and the internet and poorer digital health literacy.

In addition, the study by Wu and Yu [14] was the only study that specifically examined older adults' preferences for social media platforms. Their findings revealed that the most commonly used social media feature among older adults was instant text messaging. Unlike younger generations, who tended to use a wide variety of apps, older adults preferred a limited selection of platforms, often opting for those they were already familiar with. The study also highlighted that older adults place greater emphasis on the perceived ease of use of eHealth technologies rather than their practicality.

Social Media Health Information Behaviors

Health information behavior on social media was the most frequently discussed social media activity influencing older adults' eHealth literacy. This behavior encompasses various aspects, including health information–seeking, sharing, evaluation, trust, and use frequency. Studies consistently suggest that older adults' health information behaviors on social media positively impact their eHealth literacy, particularly functional eHealth literacy, critical eHealth literacy, and translational eHealth literacy.

In particular, Ye [80] thoroughly assessed the influence of older individuals' behaviors of using health information on social media. The study examined the correlation between 4 aspects of older persons' online health information use habits, including social media (frequency of using online health information, forwarding and sharing online health information, using online

XSL•FO
RenderX

health information to regulate one's behavior, and positive attitudes toward online health information), and 3 dimensions of eHealth literacy (the ability to apply eHealth literacy, the ability to make critical judgments and decision-making ability) [80]. A notable positive association was observed among the 7 variables, namely, stronger habits of internet use of health information among older persons corresponding to better eHealth literacy. Among these practices, the dissemination and interchange of internet health information among older people had the most significant influence on their eHealth literacy. Moreover, the habits of older individuals' internet health information use had a significant impact on older people's health decision-making ability. The study further affirmed that older individuals' adoption of social media to browse health content was an effective path to enhance their eHealth literacy because they would improve their understanding of health information in various ways during browsing to regulate their health behaviors.

Kachentawa [13] further recognized the important positive impacts of social media use on older individuals' eHealth literacy and suggested that 2 important literacies related to social media should be included in the system of indicators of older people's eHealth literacy in Thailand within the setting of the new normal lifestyle. These 2 indicators include media and information literacy and computer and social media literacy. The former elucidates the capacity of older Thai individuals to assess the reliability of the COVID-19 pandemic information shared on social media platforms and to comprehend tactics and demonstrations of technology associated with social media. The latter focuses on the capacity of older individuals to acquire new knowledge on social media, including the ability to use relevant apps, communicate with others, obtain accurate information regarding the COVID-19 pandemic, and resolve fundamental challenges encountered on social media. It can be concluded that these 2 indicators encompassed older adults' behaviors related to understanding, evaluating, sharing, and seeking health information on social media. This further underscores the significant impact of health information behaviors on social media in shaping eHealth literacy among older adults.

The increased social media–based health information behaviors among older adults were also found to help them improve their eHealth literacy. Tennant et al [78] analyzed social media health information–seeking behaviors and noted that using social media to obtain and evaluate eHealth information contributed to their eHealth literacy, especially for functional eHealth literacy and critical eHealth literacy. The study conducted by Liu et al [44] investigated older individuals' eHealth literacy during the COVID-19 epidemic. The findings demonstrated that more than half of the older adults had better eHealth literacy during this period. The study also identified several key factors, such as advancements in IT and efforts in epidemic prevention and control, that enabled older adults to effectively use digital technologies, including social media.

In addition, social media impacts the eHealth literacy of older adults by expanding their sources of health information. The study by Chai [81] indirectly demonstrated the impact of social media on improving older individuals' eHealth literacy by exposing how social media facilitated their access to health information. This study pointed out that social media enhanced health literacy among older adults by providing them with direct health information, on the one hand, and helping them to access health content and health promotion resources from social networks by maintaining social connections, on the other hand. Jang et al [82] supported this view from the opposite perspective. Their study found that older adults who had less access to digital resources, among others, had restricted health sources and exhibited lower preventive behaviors that were needed to enhance their eHealth literacy.

Finally, it is worth noting that older adults' trust in the sources of health information on social media can significantly influence their translational eHealth literacy. Tan et al [88] highlighted that older adults who exhibited greater trust in social media and had limited access to alternative information sources were more likely to make suboptimal health decisions. This finding suggests that the presence of inaccurate information on social media may impede the development of translational eHealth literacy among older adults.

Self-Management

Older adults' self-health management behaviors on social media, including health prevention, medical management, and lifestyle management, had a positive impact on their eHealth literacy, particularly their translational eHealth literacy. Given that some (12/16, 75%) studies were conducted in the context of the COVID-19 pandemic, they specifically highlighted how older adults' use of health information for health prevention influenced their eHealth literacy [13,44,81].

Kachentawa [13] proposed a media and information literacy indicator within the framework of eHealth literacy for older adults in the "new normal." This indicator highlighted the ability of older adults in Thailand to effectively use health information from social media to protect themselves from COVID-19 infection. Similarly, focusing on older adults during the COVID-19 pandemic, Wu and Yu [14] explored the use of social media as a grid-based management platform during this period. This platform reconnected older adults to an online model of a familiar social network, enabling them to become accustomed to digital technologies and access eHealth services. It not only supported health prevention efforts but also facilitated the management of their medical needs.

Liu et al [44] noted that the normalization of the pandemic prevention and control measures encouraged older adults to adopt electronic devices and ITs for health management and prevention, thereby enhancing their eHealth literacy. Chai [81] focused on how older adults applied health information obtained from social media to their daily practices during the COVID-19 pandemic, improving their health behaviors and outcomes. This process, in essence, represents an enhancement in their translational eHealth literacy.

Furthermore, Ye [80] explored the mechanisms through which older adults enhanced their eHealth literacy via self-management on social media. The study highlighted that, through the process of acquiring health information on social media platforms, older adults gained increased health knowledge, improved their health

XSL•FO
RenderX

awareness, and developed stronger problem-solving abilities related to health issues. These advancements enabled them to effectively use online health knowledge for self-management, such as regulating their own health behaviors. In addition, this process fostered a heightened sense of health self-efficacy, ultimately leading to improved eHealth literacy.

Health Interventions

In total, 13% (2/16) of health intervention studies conducted on social platforms had validated the efficacy of social media in augmenting eHealth literacy among older individuals [83,85]. Lee and Ryu [83] incorporated social media into a walking program for older adults, which included mobile medical devices and gamification elements. Findings showed that participants' eHealth literacy improved, and the positive impact of social media was generally appreciated by older adults. Similarly, the study by Chen et al [85], while not directly measuring changes in older adults' eHealth literacy levels after intervening in older adults' health through social media, quantified substantial improvements in health literacy and overall health self-efficacy while also noting a rise in older persons' understanding of the significance of health, as well as their capacity to address health issues.

Telemedicine

Social media platforms hold significant potential as telemedicine platforms, facilitating communication between older adults and health care providers while helping them access online health services. Only 13% (2/16) of studies had discussed the role of social media as a telemedicine platform in enhancing eHealth literacy among older adults [44,86]. Oh et al [86] revealed that Asian Americans improved their confidence and capabilities by communicating with health care providers through online platforms, such as social media, thereby enhancing their health information literacy. This finding reflects the potential of social media communication with health care providers to promote communication literacy among older adults. Furthermore, during the COVID-19 pandemic, the role of social media in telemedicine was effectively demonstrated. Liu et al [44] found that home isolation measures during the COVID-19 pandemic, which restricted in-person medical consultations, encouraged older adults to adopt telemedicine practices. These online health behaviors played a critical role in improving their eHealth literacy during the COVID-19 pandemic.

Health Decision-Making

Health decision-making emphasizes the specific decision-making behaviors of older adults based on health information obtained from social media, fundamentally reflecting the application of health information. This corresponds to the dimension of translational eHealth literacy. Kachentawa [13] identified media and information literacy as a key indicator of older adults' eHealth literacy in the new normal. This indicator included their health decision-making abilities, particularly their capacity to use health information from social media to make treatment decisions after contracting COVID-19. This, to some extent, reflects the translational dimension of eHealth literacy.

Impact of eHealth Literacy on Older Individuals' Social Media Health Behaviors

According to the results, eHealth literacy, in turn, affects older adults' performance of health behaviors on social media. Of the 16 studies, 8 (50%) were involved, and the details are provided in Table 6. The most discussed dimensions of eHealth literacy were critical eHealth literacy (4/16, 25%) and functional literacy (4/16, 25%).

XSL•FO
RenderX

Table 6. The impact of eHealth literacy (eHL) on older adults' social media behaviors.

Study	Theoretical mapping: dimensions of eHL	Influenced social media health behaviors	Mechanism of association and key conclusions
Ye [80], 2024	• eHL	• Health information acquisition	The level of eHealth literacy among older adults is directly proportional to the diversity of their access to information.
Matchanova et al [51], 2023	• Functional eHL and critical eHL: verbal IQ, numeracy skills, knowledge	• Health information sharing • Health information evaluation	Older adults who lack functional literacy and critical eHL inaccurately assess health information on social media and are more likely to share incorrect health information.
Wang and Zhang [42], 2023	• eHL: digital literacy • Communicative eHL: interactive health literacy • Critical health literacy	• Health information acquisition	eHL was positively associated with older adults' social media health information acquisition behaviors, as evidenced by communicative eHL predicting older adults' passive and long-term health information gathering behaviors and critical eHL predicting older adults' active information acquisition behaviors.
Liu et al [44], 2022	• Functional eHL: barriers to the use of digital technology • Critical eHL: distinguish health information	• Health information acquisition • Telemedicine (accessing health care services) • Health information evaluation • Health decision-making	Older adults who lack functional eHL have difficulty in accessing health care services through social media, and the lack of critical eHL makes it difficult for them to properly evaluate health information and make sound health decisions.
Chai [81], 2022	• Critical eHL: understanding and evaluation of health information • Functional eHL: know little about new digital technologies	• Health information evaluation • Frequency of social media use	Older adults who lack critical literacy struggle to understand the scientific logic of health rumors on social media and tend to reject new communication technologies, such as social media, due to a lack of digital literacy.
Tan et al [88], 2022	• Critical eHL: trust in health information sources	• Health decision-making (medical decisions)	Lack of critical eHL diminishes the ability of older adults to make sound health decisions.
Wu and Yu [14], 2021	• Functional eHL: unfamiliarity with the new digital technology	• Telemedicine (accessing health care services)	Lack of functional eHL affects attitudes and willingness to adopt telemedicine among older adults.
Shang and Zuo [79], 2020	• eHL	• Health knowledge learning	Self-efficacy and perceived benefits partially moderate the relationship between eHealth literacy and willingness to learn about health.

First, critical eHealth literacy directly affects older adults' assessment of social media health messages. Matchanova et al [51] discovered that persons with little knowledge and fundamental numeracy abilities might have had impaired accuracy in assessing the truthfulness of news headlines about the COVID-19 pandemic on social media and were more prone to disseminating false information on social networks. Chai [81] found inconsistencies in older adults' understanding and evaluation of the COVID-19 pandemic–related rumors. While some older adults were able to identify misinformation about the pandemic, they often lacked an understanding of the underlying scientific logic. This highlights a gap in critical eHealth literacy among older adults.

Second, a lack of critical eHealth literacy can also affect older adults' ability to make sound health decisions. Tan et al [88] noted that older adults' trust in social media disinformation affected their hesitancy to take a new coronary vaccine. Although the findings showed that social media only slightly predicted vaccination intention among older adults, the study also found that trusting only formal sources and distrusting informal sources was associated with higher vaccination intentions. Therefore, improving older individuals' ability to discern health information on social media can eliminate their trust in false health information, thereby shaping their health behaviors.

Similarly, Liu et al [44] observed that older persons predominantly relied on information to avoid COVID-19 infection and to enhance their well-being. They also highlighted that their limited capacity to understand and assess health-related information hindered their ability to distinguish between reliable information on social media, thus impacting their capacity to make informed health choices [44]. Similarly, critical eHealth literacy is essential for older people to effectively access health information on social media. Insufficient skills can hinder

XSL•FO
RenderX

accurate information assessment, increase vulnerability to misinformation, and complicate the health decision-making process.

Furthermore, research has revealed a correlation between functional literacy and the inclination of older populations to use social media platforms (3/16, 19%) [14,80,81]. Ye [80] demonstrated that older people who relied heavily on online health information exhibited a more favorable and hopeful outlook toward their health. Moreover, they were more inclined to use internet-based platforms, such as social media, to obtain health-related information. In contrast, Chai [81] argued from the opposite perspective that older adults who lacked functional eHealth literacy showed rejection of new communication technologies, such as social media, as most of them have had unpleasant experiences with the adoption of new technologies. Similarly, Wu and Yu [14] found that the lack of functional eHealth literacy affected attitudes and willingness to adopt telemedicine through social media among older adults.

Finally, eHealth literacy directly affects older adults' health information behaviors [42,51,79]. Matchanova et al [51] revealed that eHealth literacy also affected health information sharing behaviors. The study conducted by Shang and Zuo [79] primarily aimed to investigate the inclination of older individuals to acquire knowledge about health through the use of social media platforms. Using self-efficacy and perceived benefits as mediating variables, they examined the effect of eHealth literacy as a cognitive state on older people's willingness to acquire health knowledge through social media. Research indicated that older individuals who possessed greater eHealth literacy tended to view themselves as having greater self-efficacy and perceived benefits in their health-related knowledge. Self-efficacy and perceived benefits exerted a minor moderating effect on the association between eHealth literacy and the inclination of older individuals to acquire knowledge about health via social media.

Wang and Zhang [42] found that older adults usually browsed health information passively, while those with higher digital literacy were more likely to actively seek health information online. In addition, positive correlations were observed between interactive literacy and functional literacy in health literacy and the health information acquisition habits of older persons. Thus, higher eHealth literacy among older people led to more proactivity and improved performance in using social media to access eHealth information.

Factors Influencing Older Adults' Use of Social Media to Improve eHealth Literacy

Overview

Previous articles dealing with the factors affecting eHealth literacy in older adults have tended to list points based on empirical results and lacked theory-guided categorization. The study by Shi et al [45] adopted a SEM to classify the factors affecting eHealth literacy among older adults, which provides inspiration for this review. In this review, the elements influencing the use of social media by an older adult to enhance eHealth literacy were classified into four main areas: (1) individual factors, including older adults' inadequate digital skills (7/16, 44%) [13,14,42,44,51,79,87] and age (3/16, 19%) [14,51,87]; (2) interpersonal factors, involving social support (6/16, 38%) [14,78,80,81,83,85]; (3) institutional or organizational factors, including misinformation on social media (7/16, 44%) [13,42,44,51,78,82,88] and privacy and security (1/16, 6%) [14]; and (4) social factors, including social media penetration (8/16, 50%) [13,14,42,44,79,82,84,86] and cultural norms and values (1/16, 6%) [14]. Figure 4 [13,14,42,44,51,78-88] presents the framework of these factors.

Figure 4. Framework diagram of the influencing factors on older persons' use of social media to enhance their eHealth literacy. This figure categorizes key factors affecting eHealth literacy, with reference sources as follows: (1) social factors: social media penetration [13,14,42,44,79,82,84,86] and cultural norms and values [14], (2) institutional and organizational factors: misinformation on social media: [13,42,44,51,78,82,88] and privacy and security [14], (3) interpersonal factors: social support [14,78,80,81,83,85], and (4) individual factors: inadequate digital skills [13,14,42,44,51,79,87] and age [14,51,87].

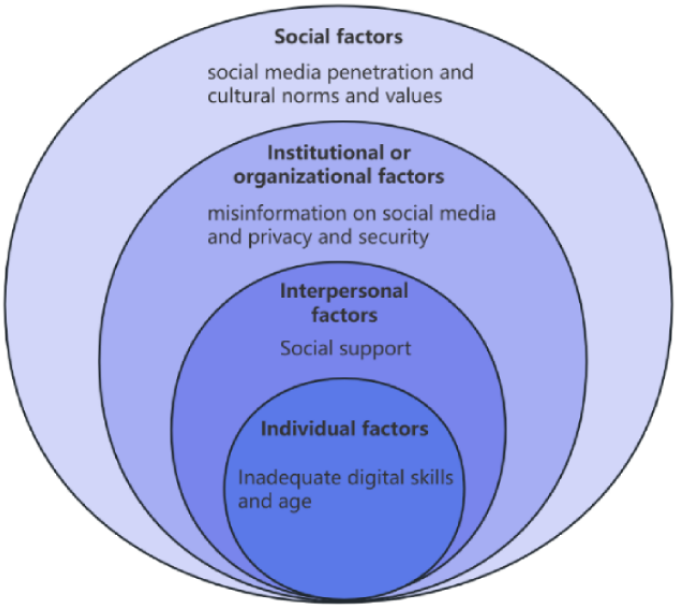

XSL•FO
RenderX

Individual Factors

In the context of SEM, the individual level refers to individual characteristics [89,90], including knowledge and skills, attitudes and beliefs, demographic factors, and self-efficacy [90-92]. Inadequate digital skills among older adults and age are identified as individual factors in this review.

Inadequate Digital Skills Among Older Adults

The prevalence of inadequate knowledge and proficiency in using digital devices among older people poses a significant obstacle to their eHealth literacy. The first obstacle is the reluctance of older individuals to use social media to acquire health information and services. Ubolwan et al [84] contended that the limited social media abilities of older persons might result in hesitancy to use social media for eHealth participation. This, in turn, could negatively influence older people's eHealth literacy, as there exists a direct relationship between the frequency of social media use and eHealth literacy. Wu and Yu [14] pointed out that older people are deprived of lifelong learning environments and have relatively low confidence in digital technology after retirement. A prominent manifestation is that, unlike the younger generation, who care about the practicality of digital health care, older people are more concerned with the perceived ease of use of digital technology [14]. Moreover, the lack of digital experience leaves older adults with high health management needs facing low technology adoption. Second, it hinders older adults' effective access to health resources. Chai [81] noted that older adults' lack of skills in using digital technology can make them face challenges in health information retrieval. Similarly, Wang and Zhang [42] argued that older individuals with little digital literacy are less able to access health information via the internet or technological devices.

Furthermore, the lack of proficiency in digital skills among older individuals can result in the digital divide among them extending to eHealth, therefore impeding the achievement of health equity [13,42,78]. Kachentawa [13] argued that perceived self-efficacy and digital competence affected older adults' adoption of digital technology. While perceived self-efficacy is primarily related to older people's enjoyment of using digital technology, the core problem remains in older adults' lack of digital skills. This issue is reflected in the indicators developed by the author related to the eHealth literacy of older persons, namely computer and social media literacy [13]. It suggests that in the new normal, older adults must not only have the knowledge and skills to use digital devices, such as computers and social media, but also have the ability to solve basic problems with social media. Tennant et al [78] highly emphasized the important role of digital literacy (computer literacy) in the older adult population, arguing that lower digital literacy among older adults prevents them from truly benefiting from health care innovations and may reopen the digital divide, exacerbating current health disparities.

Age

Another individual factor is age. However, the results of the studies were not consistent. While the study by Wu and Yu [14] identified aging as a significant determinant of older people's digital technology use and their attitudes toward telemedicine, Vitolo et al [87] found that digital literacy was only partially influenced by age. The role of age was further weakened in the study by Matchanova et al [51], where the findings did not demonstrate a substantial impact of age on the probability of sharing false or accurate information online. Nevertheless, it was observed that this phenomenon could be attributed to the age of the study participants (aged 50 years), but most of the significant effects on susceptibility to misinformation and sharing were centered at the age of 65 years. Thus, there may be differences in perceived susceptibility to social media misinformation about health by age.

Interpersonal Factors

The interpersonal dimension of SEM emphasizes the influence of social networks and social support systems on individual behaviors [89-91]. In this review, social support plays a crucial role in enabling older people to use social media to enhance their eHealth literacy.

The social supportive role of social media is manifested first and foremost in increasing the health information resources available to older adults by linking their relationships on online platforms. Social relationships, the traditional way for older individuals to obtain health information or resources, have achieved rapid dissemination on online platforms with social media connections. This model played an important role during the COVID-19 pandemic. Wu and Yu [14] noted that during the COVID-19 pandemic, social media platforms, such as WeChat, expanded offline community management to online, creating a new model of online acquaintance relationships and, therefore, continued maintaining health support through social networks. Chai [81] also noted that older adults accessed health information through online social relationships during the COVID-19 pandemic.

Furthermore, the effect of social support is reflected in the interactive advantages of social media in facilitating health communication and health information sharing behaviors among older adults. Tennant et al [78] argued that the greater social support that older adults derive from social media is based on stronger Web 2.0-connected support networks and the multifunctionality factor of social media. Older adults can enhance their understanding of their health status and connect with online friends with similar experiences through social support venues built on interactive social media. Ye [80] further clarified that older adults' online health information sharing and exchange behaviors had the greatest impact on their eHealth literacy, outweighing the effect of the other dimensions (frequency of use of online health information, use of online health information to regulate their behaviors, and the impact of favorable attitudes toward online health information) on eHealth literacy.

Finally, the facilitating role of social support provided by social media is reflected in the strength of social media as a medium for group-based health interventions. Of the 16 studies, 2 (13%) social media–based health intervention studies found that social media was effective in motivating older adults to participate in health intervention programs [83,85]. Lee and Ryu [83] found that mHealth apps incorporating social media allowed older adults who were socially isolated during the COVID-19

XSL•FO
RenderX

pandemic to feel socially connected again, as they shared and interacted with their peers and encouraged them to use these apps. The rationale for the finding was explained in more detail in the study by Chen et al [85]. They pointed out that interacting with others is a key incentive for older adults to engage in health interventions. Older adults enjoy sharing experiences and opinions in online discussions on social media and, over time, they develop a sense of socialization that enhances feelings of connectedness and engagement.

Institutional or Organizational Factors

Overview

On the basis of the SEM, institutional or organizational factors refer to the influence of institutional or organizational settings individuals are embedded in on personal behaviors [89,90]. In the health context, organizational structures and processes, as well as the environment within the organization, including safety, have an impact on people's health behaviors [90]. This review revealed that misinformation on social media and privacy and security concerns associated with social media pose possible obstacles that hinder older people from enhancing their eHealth literacy using social media.

Misinformation on Social Media

The issue of information credibility on social media has attracted a great deal of attention from researchers, who have argued that the proliferation of false information on social media is a serious threat to effective access to health information by older adults [13,42,78,79]. This was especially true during the COVID-19 pandemic, when false news on social media was rampant and caused great harm to older adults [13] and affected their health decision-making abilities and health behavioral norms. For example, studies by Jang et al [82] and Tan et al [88] found that trust in misinformation on social media increased new coronary vaccine hesitancy in older adults. Social media also contributes to the dissemination of misinformation about health among older adults. Matchanova et al [51] noted that the social media sharing function does not require the involvement of subjective agency other than motivation, which may influence the reflective accuracy of older adults to spread misinformation about health.

Part of the literature identifies the lack of older adults' ability to assess social media information, such as media literacy or information literacy, that is, the ability to assess the trustworthiness of social media. Kachentawa [13] included this competency as one of the indicators of eHealth literacy among the new normal older adults and emphasized the importance of assessing the trustworthiness of the information in the context of social media to improve eHealth literacy among older adults. Similarly, Tennant et al [78] noted a real lack of confidence among older adults in assessing the quality of social media health information. Furthermore, Liu et al [44] found that older adults had the lowest information judgment scores. Nonetheless, Wang and Zhang [42] revisited the relationship between key health literacies related to the ability to assess health information and older adults' health information acquisition behaviors from the perspective of the motivation of older adults' health information acquisition behaviors. The study yielded a different and interesting finding that critical health literacy was not significantly correlated with health information acquisition behavior; that is, even if older adults are unable to critically analyze health information, they are able to facilitate their health information access behaviors as long as they have basic literacy and communication skills.

Privacy and Security

Wu and Yu [14] identified financial security as a barrier that hinders older persons from the adoption of digital services. Older adults are anxious about the potential loss of money through the use of online payments.

Social or Community Factors

Overview

The social or community level in SEM refers to societal influence and cultural or subcultural norms and values [89]. Regarding social influence, the development of social media technology can affect personal health behaviors. In this review, this influential factor is identified as social media penetration. Meanwhile, the influence of cultural norms was reflected in a study in China [14].

Social Media Penetration

The high penetration of social media among older adults is a notable trend that has been the focus of most studies, and this trend's development establishes a crucial basis for the function of social media in enhancing eHealth literacy among the older population. Given the growing population of older individuals using social media, it has become a crucial medium for them to obtain and exchange health-related information [13,44,82] and creates novel avenues for older individuals to acquire health advice [79]. This high penetration is mainly achieved through 2 aspects: one is driven by IT (5/16, 31%) [42,44,79,84,85], and the other is catalyzed by the COVID-19 pandemic (2/16, 13%) [13,44]. One consequence of the swift advancement of IT is the establishment of a technological framework that enables older people to use social media. Various studies have indicated an increase in the frequency of social media use among older adults in recent years [44,79,84,85]. Moreover, this IT push is inadvertently creating social pressure on older people to learn to use digital technology tools [14]. Furthermore, the outbreak and normalized management of the COVID-19 pandemic further increased the adoption of social media by older adults. During the COVID-19 pandemic, older adults faced physical health risks and social isolation situations, and social media became not only a tool for them to maintain social relationships but also an important platform for them to obtain health-related information and health services [13,44].

The high penetration of social media is also evident in the significant expansion of channels for accessing health information, which has transformed the ways they obtain it. Ye [80] pointed out that social media can bring more health channels to older adults and increase their access to health information. Furthermore, Oh et al [86] argued that social media can serve as an effective channel to supplement older patients' health information and knowledge when they have poor communication with their health providers, thus improving their communicative eHealth literacy.

XSL•FO
RenderX

Cultural Norms and Values

Wu and Yu [14], in the study of older adults in Wuhan, China, found that China's social culture of acquaintance profoundly influenced older adults' adoption of digital health care. Chinese older adults were more likely to adopt programs based on acquaintance relationships and traditional health care processes than those in other countries. The acquaintance relationship in Chinese society is a distinct cultural phenomenon characterized by the dichotomous connection between acquaintances and the line between insiders and strangers [14]. During the COVID-19 pandemic, the "grid management" model facilitated through social media chat groups linked a new type of online acquaintance relationship that could help manage the health care needs of older adults by providing them with a sense of trust and setting common rules. To this end, their study indicates the need to take into account the cultural environment while encouraging the acceptance of digital health among older individuals.

Discussion

Principal Findings

This study is the initial systematic literature review that investigates the correlation between social media use behaviors and eHealth literacy in the older adult population. The increasing global aging population, the proliferation of digital technologies, and the COVID-19 pandemic have significantly immersed older adults in the social media and eHealth phenomenon. This situation offers them both prospects and obstacles to fully benefit from digital health technologies. On the one hand, the advantages of social media have expanded the channels for older adults to engage in eHealth and, on the other hand, older adults' eHealth literacy greatly influences their willingness to adopt social media, as well as their performance of health behaviors based on social media platforms.

To further integrate older adults into the eHealth trend and achieve active aging, it is essential to establish a positive feedback loop between social media behaviors and eHealth literacy. This involves enabling older adults to gradually enhance their eHealth literacy through health-related practices on social media, which, in turn, improves their social media health behaviors and experiences, ultimately bridging the digital divide. Therefore, this study aimed to synthesize and analyze recent research on this relationship, uncovering patterns and influencing factors to identify both the potential and barriers of using social media to enhance eHealth literacy among older adults. The findings are expected to inform targeted eHealth literacy interventions for this population, fostering a sustainable cycle of mutual reinforcement between health practices and health skills.

We extracted themes related to older adults' social media behaviors and the dimensions of their eHealth literacy for cross-analysis. The findings reveal that older adults' social media behaviors and eHealth literacy are mutually influential and interactive. Specifically, this review provides comprehensive answers to the RQs.

Regarding the first RQ, we identified that older adults' social media behaviors encompass general use behaviors and health-related behaviors. General use behaviors include habits such as use frequency and platform preferences, while health-related behaviors involve a range of activities aimed at health purposes, such as health information behaviors, health decision-making, self-management, participation in health interventions, and engagement in telemedicine. Among these, health information behaviors are the most common and widely discussed health activities.

For the second RQ, most (14/16, 88%) of the reviewed studies confirmed the positive impact of older adults' social media behaviors on enhancing their eHealth literacy. First, regarding general use behaviors, the frequency of social media use significantly improves older adults' eHealth literacy, aligning with previous research that observed a positive correlation between the frequency of internet technology use and eHealth literacy [93,94]. In addition, familiar social media platforms enhance older adults' perceived ease of use of digital technologies, thereby increasing their willingness to engage in health-related activities on social media [14].

Second, health-related behaviors on social media are the most effective and direct practices for improving older adults' eHealth literacy. These practices include health information behaviors, health management, health interventions, telemedicine, and health decision-making, with health information behaviors being the most frequently (7/16, 44%) discussed. Older adults' self-management and health decision-making activities are often built upon their understanding and application of health information obtained via social media. Furthermore, social media serves as a highly effective platform for group-based, online health interventions, where its interactive features enhance older adults' participation in interventions and their willingness to adopt digital technologies [83,85]. The platform's potential as a telemedicine tool, especially during the COVID-19 pandemic, has been highlighted as older adults began using social media for telemedicine and, through these practices, improved their eHealth literacy.

For the third RQ, the findings revealed that older adults' eHealth literacy reciprocally influences their social media behaviors. This includes their willingness to adopt social media and other digital technologies [14,44,81], their ability to understand and evaluate health information on social media [44,51,88], their willingness to learn health knowledge through social media [79], and their proactivity in acquiring health information [42]. Among these, functional and critical eHealth literacy have the most significant impacts. A lack of functional eHealth literacy not only hinders older adults' effective use of social media and other digital technologies but also reduces their willingness to adopt them. A deficiency in critical eHealth literacy affects their ability to understand and evaluate information, which further impacts their health decision-making and self-management behaviors.

By integrating the insights from RQs 2 and 3, we conclude that older adults' eHealth literacy and their social media behaviors exhibit a reciprocal relationship, highlighting the need for

XSL•FO
RenderX

interventions that leverage this dynamic to enhance both dimensions.

For the fourth RQ, older adults' use of social media for health participation is influenced by individual, interpersonal, institutional or organizational, and social factors. At the individual level, inadequate digital skills [13,14,42,44,51,79,87] and age-related challenges [14,51,87] limit their ability to access and use social media effectively, contributing to the digital divide. Interpersonal factors, such as social support [14,78,80,81,83,85], enhanced health information acquisition or sharing and participation in health interventions, which are vital for improving eHealth literacy. Institutional barriers, including misinformation [13,42,44,51,78,82,88] and privacy concerns [14], further hinder their engagement, emphasizing the need for critical eHealth literacy. Moreover, the high penetration of social media, driven by technological advancements and the COVID-19 pandemic, has expanded access to health information and services, offering new opportunities to bridge these gaps [13,14,42,44,79,82,84,86]. Finally, promoting the use of eHealth technologies among older adults also requires consideration of cultural contextual factors at the societal level [14].

Considerations for Establishing a Reciprocal Mechanism Between Social Media Health Practices and Older Adults' eHealth Literacy

Overview

The findings reveal the immense potential of social media to enhance older adults' eHealth literacy. To further promote their active use of social media for health interactions, information acquisition, and service use, it is essential to enable older adults to continually improve their eHealth literacy through social media–based health practices. This improvement, in turn, can lead to positive feedback in their social media use experiences, ultimately establishing a reciprocal reinforcement system between social media use and eHealth literacy. To this end, theoretical and practical efforts can be made in the following areas.

Leveraging Postpandemic Momentum to Enhance eHealth Literacy Among Older Adults Through Social Media

The high penetration of social media and the catalyst of the COVID-19 pandemic have facilitated changes in older adults' technology's use behaviors and health behaviors, laying the groundwork for improving older adults' eHealth literacy. The continuous development and popularization of social media have revolutionized the traditional pattern of health information dissemination [95,96] and changed the habits of older adults in accessing eHealth information or services [44,97]. Especially with the COVID-19 outbreak, this pattern has deepened further. On the one hand, the social blocking measures forced older adults to rely on social media for information and lifestyle services, as well as social support, among others [98]. On the other hand, the epidemic posed multiple risks to older adults [99-101], who were forced to step out of their comfort zones and start using social media and engaging in eHealth activity practices. The COVID-19 pandemic is arguably a major reason for the rise in social media use among older adults [13]. Moreover, during this period, older adults showed unprecedented enthusiasm for learning about eHealth tools [102]. Interest is an important reason for improving eHealth literacy [103]. Therefore, the present moment, that is, after the end of the COVID-19 pandemic and in the context of older people's heightened experience of and interest in eHealth technologies, as well as social media, is an optimal time to improve eHealth literacy among older people. Countries around the world should take this opportunity to lay and plan to encourage older adults to use social media more to participate in eHealth activities and to accelerate the pace of active integration of older adults into the trend of eHealth technology.

Reconsidering the Dimensions of eHealth Literacy Among Older Adults in the Context of Web 2.0

Findings regarding the impact of eHealth literacy on older adults' social media health behaviors bring more serious practical considerations to the achievement of this goal. According to the review, eHealth literacy among older adults can, in turn, affect older adults' social media use behaviors and performance, including influencing older adults' understanding and evaluation of social media health information [44,51,88], willingness to use social media to access or learn health information [79-81], and access to health information initiative [42]. More importantly, the lack of digital skills and misinformation on social media, the 2 prominent factors affecting older adults' use of social media to improve eHealth literacy, are closely related to 2 important components of eHealth literacy: digital literacy and critical health literacy, respectively. The lack of these 2 literacies not only prevents older adults from truly benefiting from eHealth but also has the potential to exacerbate health disparities by extending the digital divide into the eHealth domain [42,79]. Kachentawa [13] innovatively included social media literacy in the indicators of eHealth literacy for older adults in the new normal. In his opinion, social media literacy among older adults included not only the ability to assess information but also the skills in understanding, using, and dealing with problems in social media. However, there is still no consensus on the concept of social media literacy. Some scholars define the concept of social media literacy as encompassing the technical and practical skills required for using social media [104,105], while others prefer to use the concept of media literacy as a substitute for social media literacy [106,107]. As the use of social media among older adults increases, social media literacy should be further emphasized as a component of eHealth literacy among older adults. In addition, relevant interventions should be centered on how to adapt functional and critical eHealth literacy for older adults.

Actively Developing Age-Friendly Integrated Social Media Health Service Platforms

Self-efficacy has been confirmed by previous research to be positively associated with older adults' willingness to use digital health technology [108]. In turn, self-efficacy in older adults was positively associated with the perceived ease of use of eHealth technology [109]. Thus, digital literacy among older adults can be achieved by lowering the threshold of eHealth

XSL•FO
RenderX

technologies. In addition to further strengthening IT infrastructure, efforts should be made to actively develop social media health service features and digital tools tailored to the needs of older adults.

Existing research on intelligent health and medical services for older adults has extensively investigated their needs, with a primary focus on developing diverse mHealth apps. For instance, Zhou et al [110] developed a home health management app tailored to the needs of older users, which promoted self-health management among this population. Similarly, Blinka et al [111] designed a sensor-based mobile app to help older adults assess frailty. However, this review reveals that older adults' digital platform preferences significantly influence their willingness to adopt technology. Wu and Yu [14] highlighted that older adults prefer limited and familiar digital platforms, which enhance their perceived ease of use and, consequently, their willingness to engage in health activities via social media. In addition, among the 16 included studies, 2 (13%) articles focusing on health interventions that used specialized health platforms. Although the intervention results showed improvements in older adults' eHealth literacy, there remains a potential risk of low adoption willingness if these interventions are to be widely implemented among the older population.

Older adults who have experienced the COVID-19 pandemic are already familiar with social media platforms, and further investigation into their preferences for these platforms is warranted. Aging impacts cognitive abilities, leading to slower reactions and increased cognitive vulnerability [112,113]. The proliferation of mobile apps can negatively affect their perceived ease of use. Therefore, effectively leveraging the social media platforms that older adults are already familiar with to create an integrated mHealth platform—incorporating features such as health information dissemination, online consultations, health monitoring, and interactive support—can enhance their willingness to adopt and use these technologies. For example, Samad-Soltani et al [114] proposed integrating Telegram-based smart bots into health care, which not only minimizes the need for app installations but also supports a variety of health-related tasks while fostering interaction with contacts. Another example is the hospital WeChat official account in China, which is a platform for embedding the hospital's eHealth system into WeChat, China's most popular social media, to facilitate older adults' access to health care information and services [79,115]. In the future, the integration of mHealth and social media is expected to become a new generation of eHealth trends, and related studies can pay more attention to this trend.

Last but not least, the development of a social media health integration platform should take into account other factors that influence the adoption of technology by older adults. Although the influences of privacy and security, age, and culture were not widely discussed in the included literature, they still play an important role in previous work on factors affecting eHealth literacy among older adults [45,116]. Therefore, future research should focus on the influence of these factors on older adults' use of social media for eHealth engagement. Previous studies have found that issues of privacy and security have been widely identified as a barrier to older adults' use of digital health technologies [109,117], and social media amplified this challenge [38]. Accordingly, the resolution of privacy and security issues will potentially have a direct impact on older adults' willingness to use social media to engage in eHealth. In addition, age, as well as cultural factors, may affect older adults' attitudes toward using social media to enhance eHealth literacy. Relevant technology development should also fully consider the characteristics and needs of older adults at different age stages and in different cultural contexts.

Optimizing Social Media for Engaging and Reliable Health Information for Older Adults

The findings indicate that health information behaviors on social media are the most common and direct form of health practice for enhancing eHealth literacy among older adults. Behaviors, such as self-management and health decision-making, are often built upon their understanding and evaluation of the health information or knowledge they acquire. Previous studies have highlighted that the type and source of information are crucial contextual factors influencing older adults' online health information–seeking behaviors [118,119]. However, in this review, as most studies were published during the COVID-19 pandemic, there is a predominant focus on older adults' acquisition and application of COVID-19 pandemic–related information, with limited attention to other types of health information. Future research should place greater emphasis on older adults' preferences for other health information types. In addition, the review reveals that trust in health information sources significantly influences older adults' health decisions [88]. Moreover, due to a lack of critical eHealth literacy, older adults are more susceptible to the negative impact of misinformation on social media [44,51]. Therefore, establishing engaging and trustworthy social media health information channels tailored to older adults is imperative.

Currently, machine learning algorithms have been introduced into health management and monitoring for older adults, demonstrating positive effects in promoting self-management and preventing health risks [120,121]. For instance, the research by Zhao et al [122] used high-order hybrid clustering algorithms that integrated physical activity and medical perspectives, advancing personalized health care and rehabilitation for older adults. However, studies on optimizing social media algorithms for health information targeted at older adults remain relatively limited. Future research could enhance social media algorithms by prioritizing accurate, clear, and relevant health content tailored to older adults. Personalized recommendations could encourage older users to use reliable medical resources and participate in health promotion activities, thereby reducing their exposure to misinformation.

In addition, older adults are considered more inclined to trust health information disseminated through authoritative channels [81]. While social media provides an abundance of health information and convenience for older adults, the prevalence of misinformation often undermines their judgment. Therefore, it is crucial to encourage authoritative health care institutions to play a more active role in creating health content on social media. These institutions should regularly disseminate reliable health information through social media platforms while leveraging the social features of these platforms to offer

XSL•FO
RenderX

knowledge-based question and answer sessions and interactive opportunities, such as expert online consultations.

Furthermore, the study by Nassetta and Gross [123] found that labels on social media can help mitigate the impact of misinformation, but this requires users to pay attention to the source of information. Thus, governments and health care institutions could collaborate to introduce reliable information verification mechanisms on social media platforms. Examples include implementing "trusted source" labels or certification marks for health information verified by authoritative organizations. These labels should be prominently displayed to enhance older adults' trust in the credibility of information sources.

Integrating Social Support Systems to Foster Lifelong eHealth Learning for Older Adults

The findings of this review indicate that the potential of social media to enhance older adults' eHealth literacy primarily lies in the social support it provides. First, social support serves as an important supplementary channel for older adults to access health resources. In this review, health information behaviors emerged as the most frequently discussed theme influencing older adults' eHealth literacy. In addition to the health information directly obtained by older adults from social media, many health resources are accessed through the online social networks maintained by social media platforms [14,81].

Second, the social support older adults gain from the interactive advantages of social media can promote their health-sharing and communication behaviors, which have a significant positive impact on their eHealth literacy [80]. Moreover, as a crucial platform for online, group-based health interventions, the greatest advantage of social media lies in leveraging social support to encourage older adults to consistently use digital health technologies [83,85].

Although social support through social media has been found to have a positive effect on older adults' mental health, health management, and health education in previous studies [15,124,125], it is still controversial whether social support is directly related to older adults' eHealth literacy. Arcury et al [93] found no direct association between social support and eHealth literacy, whereas Kim and Sung [126] found that peer support was significantly related to the eHealth literacy of older individuals. This review did not find a direct association between the 2 but observed that social support provided by social media has the potential to promote the continued use of eHealth devices among older adults; however, this relationship has not been rigorously confirmed in the included literature [14,78,80,81,83,85]. Moreover, studies do find that peer support increases internet use among older adults [98]. Thus, the correlation between social support provided by social media and eHealth literacy among older adults may have been moderated through ongoing eHealth device use behaviors, and the relationship between the 3 could be explored in depth in the future.

While the direct relationship between social media and older adults' eHealth literacy remains unclear, social media does reshape older adults' health behaviors by linking their social networks and social support on online platforms. Studies suggested that the perceived positive effects of social support on health outcomes are not only limited to face-to-face interactions but are also applicable in the social media context [127,128]. However, a relatively small portion of the studies in this review (6/16, 38%) focused on the positive impact of social support on eHealth literacy among older adults. Most studies (10/16, 63%) lacked an exploration of older adults' health-related interactions on social media, such as peer-to-peer interactions and interactions with health care professionals. To effectively leverage social media in building a sustainable mechanism to enhance older adults' eHealth literacy, further exploration of their interactive behaviors on social media is needed, with an emphasis on the critical role of social support in encouraging older adults to actively engage in eHealth practices.

To this end, communities or health care institutions could use social media to establish virtual support groups focusing on different health topics, where volunteers regularly share health knowledge and encourage older adults to exchange their health experiences. Previous studies have called on hospitals and clinics worldwide to develop social media–supported medical apps, establish support groups, and encourage users to actively engage in social interactions with peers and health care providers [72].

In addition to the social support offered by online social media, offline social support remains indispensable. Despite the increased use of digital technologies among older adults driven by IT and the COVID-19 pandemic, research indicates that they still prefer traditional face-to-face interactions [129]. The development of technology should not be detached from the temperature of human beings, and the enhancement of eHealth literacy among older adults requires the intervention of "real people." For example, older adults who lack functional digital skills often rely heavily on family members for technical assistance and advice [129]. However, as the caregiver burden on families increases, it is essential to involve community resources to help older adults establish a long-term mechanism for learning and practicing eHealth technologies.

Communities could host social media health education workshops to teach older adults how to use social media for accessing health information, using telemedicine services, and identifying misinformation. In addition, the role of social work volunteers should be emphasized to provide ongoing online technical support and guidance for older adults. As older adults gain more positive experiences from using social media to participate in health practices, they will likely engage more actively in the eHealth trend. This engagement will, in turn, enhance their eHealth literacy, allowing them to refine their behaviors and ultimately achieve a long-term positive feedback loop between social media health practices and eHealth literacy.

Limitations

This review has several limitations. First, we only searched for keywords in the abstracts and titles, which, on the one hand, led to a large number of studies that used social media to recruit respondents unrelated to the topic appearing in the first search results. On the other hand, it could have resulted in the omission of articles in the primary content of the text dealing with the

XSL•FO
RenderX

correlation between social media and eHealth literacy among older individuals. Second, the sample size for this review was relatively small due to the specificity of the research topic. Third, the thematic analysis conducted in this review was qualitative, and some (6/16, 38%) of the included studies did not directly examine the relationship between social media and eHealth literacy. Instead, they provided indirect evidence or theoretical suggestions that social media might enhance eHealth literacy in older adults. Therefore, the conclusions drawn in this review are primarily based on correlation rather than causation. Finally, the scarcity of existing research investigating this specific correlation highlights the need for more empirical studies that directly explore how social media impacts older adults' eHealth literacy. Future research should address these gaps by using diverse methodologies and larger sample sizes to establish more comprehensive and definitive conclusions.

Conclusions

Given the significant potential of social media in enhancing eHealth literacy among older individuals and the need to improve eHealth literacy among this population, this review sought to investigate the correlation between the social media use habits of older adults and their eHealth literacy. The review analyzed studies conducted from January 2000 to May 2024. The results found that social media and older adults' eHealth literacy influence and interact with each other. Older adults' social media use behaviors, including social media use habits and health behaviors, have a significant impact on their eHealth literacy, suggesting that social media has significant potential to enhance eHealth literacy among older adults. In turn, older adults' eHealth literacy affects their willingness to adopt digital technologies, such as social media, as well as their performance of social media behaviors for health purposes. Promoting a virtuous cycle of interaction between the 2 fosters active aging and helps establish a long-term mechanism to enhance eHealth literacy among older adults. Achieving this goal requires a comprehensive consideration of factors that influence older adults' use of social media to improve eHealth literacy, including social media penetration, insufficient digital skills, misinformation on social media, age, privacy and safety, and cultural norms and values. On the basis of the results of the review, this study proposes several future considerations for improving eHealth literacy among older adults through social media.

Acknowledgments

The authors would like to thank the anonymous reviewers and editors for their helpful comments on the manuscript. This study did not receive external funding.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

CZ, EM, and AAA were responsible for the conceptualization, research design, and search strategy of this review. YQ was responsible for all database searches and result collation, and AW was responsible for paper screening, data extraction, and critical appraisal. YM was involved in conflict resolution during the screening process. EM, AAA, and YQ were involved in the data analysis and data interpretation and drafted the manuscript. All authors were involved in reviewing and editing the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 53 KB - jmir_v27i1e66058_app1.pdf]

Multimedia Appendix 2
Searching strategies.
[DOCX File , 16 KB - jmir_v27i1e66058_app2.docx]

Multimedia Appendix 3
Conceptual matrix for thematic analysis of social media behavior.
[DOCX File , 23 KB - jmir_v27i1e66058_app3.docx]

Multimedia Appendix 4
Quality assessment for the studies reviewed.
[DOCX File , 21 KB - jmir_v27i1e66058_app4.docx]

XSL•FO
RenderX

References

1. Carr CT, Hayes RA. Social media: defining, developing, and divining. Atl J Commun 2015 Feb 06;23(1):46-65. [doi: 10.1080/15456870.2015.972282]
2. Dixon SJ. Most used social networks 2024, by number of users. Statista. 2024 Jul 10. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2024-08-02]
3. Ameri F, Dastani M, Sabahi A, Hooshangi F, Rahimkarimi M, Rajabi E, et al. The role of e-health literacy in preventive behaviors for COVID-19: a systematic review. J Health Lit 2022 Jan;6(4):88-97. [doi: 10.22038/jhl.2021.61581.1241]
4. Choukou MA, Sanchez-Ramirez DC, Pol M, Uddin M, Monnin C, Syed-Abdul S. COVID-19 infodemic and digital health literacy in vulnerable populations: a scoping review. Digit Health 2022 Feb 10;8:20552076221076927 [FREE Full text] [doi: 10.1177/20552076221076927] [Medline: 35223076]
5. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating engaging health promotion campaigns on social media: observations and lessons from Fitbit and Garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]
6. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: 10.17061/phrp2521517] [Medline: 25848735]
7. Wu T, He Z, Zhang D. Impact of communicating with doctors via social media on consumers' e-health literacy and healthy behaviors in China. Inquiry 2020;57:46958020971188 [FREE Full text] [doi: 10.1177/0046958020971188] [Medline: 33238788]
8. Lin CA, Hsu YM, Chen H. Creating a virtuous circle during a pandemic threat: netizens' resilience via information-induced psychological distance and social media approach. Online Inf Rev 2023 Oct 13;48(3):562-580. [doi: 10.1108/oir-09-2022-0494]
9. Mutiara VS, Pamungkasari EP, Padmawati RS, Probandari A. Empowering women through a path analysis: identification of effective factors based on social support and community participation on e-health literacy in women's virtual health society. J Health Lit 2024 Jun;9(2):65-77 [FREE Full text] [doi: 10.22038/JHL.2024.77087.1517]
10. Braun LA, Zomorodbakhsch B, Keinki C, Huebner J. Information needs, communication and usage of social media by cancer patients and their relatives. J Cancer Res Clin Oncol 2019 Jul;145(7):1865-1875. [doi: 10.1007/s00432-019-02929-9] [Medline: 31123824]
11. Sabahi A, Ahmadian L, Mirzaee M, Khajouei R. Patients' preferences for receiving laboratory test results. Am J Manag Care 2017 Apr 01;23(4):e113-e119 [FREE Full text] [Medline: 28554212]
12. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022 Jan 13. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2024-08-02]
13. Kachentawa K. The development of E-health literacy indicators for older Thai adults in adjusting to live with the new normal lifestyle. J Behav Sci 2023;18(1):49-64 [FREE Full text]
14. Wu Z, Yu P. The role of digital devices in chronic disease management during the Covid-19 pandemic—a study of senior citizens in Wuhan. Asian J Commun 2021 Sep 26;31(5):373-390. [doi: 10.1080/01292986.2021.1931895]
15. Fu L, Xie Y. The effects of social media use on the health of older adults: an empirical analysis based on 2017 Chinese general social survey. Healthcare (Basel) 2021 Sep 01;9(9):1143 [FREE Full text] [doi: 10.3390/healthcare9091143] [Medline: 34574917]
16. Han M, Tan XY, Lee R, Lee JK, Mahendran R. Impact of social media on health-related outcomes among older adults in Singapore: qualitative study. JMIR Aging 2021 Feb 17;4(1):e23826 [FREE Full text] [doi: 10.2196/23826] [Medline: 33595437]
17. Choi SL, Hites L, Bolland AC, Lee J, Payne-Foster P, Bissell K. Telehealth uptake among middle-aged and older Americans during COVID-19: chronic conditions, social media communication, and race/ethnicity. Aging Ment Health 2024;28(1):160-168. [doi: 10.1080/13607863.2022.2149696] [Medline: 36450359]
18. Manafò E, Wong S. Promoting eHealth literacy in older adults: key informant perspectives. Can J Diet Pract Res 2013;74(1):37-41. [doi: 10.3148/74.1.2013.37] [Medline: 23449214]
19. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
20. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov 17;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
21. Klecun E, Lichtner V, Cornford T. e-Literacy in health care. Stud Health Technol Inform 2014;205:843-847. [Medline: 25160306]
22. Gilstad H. Toward a comprehensive model of eHealth literacy. In: Proceedings of the Practical Aspects of Health Informatics. 2014 Presented at: PAHI 2014; May 19-20 , 2014; Trondheim, Norway. [doi: 10.13140/2.1.4569.0247]
23. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]

XSL•FO
RenderX

24. Bautista JR. From solving a health problem to achieving quality of life: redefining eHealth literacy. J Lit Technol 2015 Dec;16(2):33-54.
25. Griebel L, Enwald H, Gilstad H, Pohl AL, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
26. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
27. Tan A, Gutierrez D, Milam EC, Neimann AL, Zampella J. Patient health-seeking behavior on WeChat: social media and dermatology. JAAD Case Rep 2020 Jan 27;6(2):136-138 [FREE Full text] [doi: 10.1016/j.jdcr.2019.11.015] [Medline: 32072002]
28. Taba M, Allen TB, Caldwell PH, Skinner SR, Kang M, McCaffery K, et al. Adolescents' self-efficacy and digital health literacy: a cross-sectional mixed methods study. BMC Public Health 2022 Jun 20;22(1):1223 [FREE Full text] [doi: 10.1186/s12889-022-13599-7] [Medline: 35725389]
29. Sun H, Qian L, Xue M, Zhou T, Qu J, Zhou J, et al. The relationship between eHealth literacy, social media self-efficacy and health communication intention among Chinese nursing undergraduates: a cross-sectional study. Front Public Health 2022 Oct 31;10:1030887 [FREE Full text] [doi: 10.3389/fpubh.2022.1030887] [Medline: 36388375]
30. Kor PP, Leung AY, Parial LL, Wong EM, Dadaczynski K, Okan O, et al. Are people with chronic diseases satisfied with the online health information related to COVID-19 during the pandemic? J Nurs Scholarsh 2021 Jan;53(1):75-86. [doi: 10.1111/jnu.12616] [Medline: 33316121]
31. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
32. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
33. Liu S, Zhao H, Fu J, Kong D, Zhong Z, Hong Y, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996 [FREE Full text] [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
34. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]
35. Hoogland AI, Mansfield J, Lafranchise EA, Bulls HW, Johnstone PA, Jim HS. eHealth literacy in older adults with cancer. J Geriatr Oncol 2020 Jul;11(6):1020-1022 [FREE Full text] [doi: 10.1016/j.jgo.2019.12.015] [Medline: 31917114]
36. Lozoya SV, Guirado MA, Zapata Gonzalez A, Lopez AB. Use of technologies and self-efficacy in older adults. IEEE Rev Iberoam Tecnol Aprendiz 2022 May;17(2):125-130. [doi: 10.1109/RITA.2022.3166870]
37. Christiansen L, Lindberg C, Sanmartin Berglund J, Anderberg P, Skär L. Using mobile health and the impact on health-related quality of life: perceptions of older adults with cognitive impairment. Int J Environ Res Public Health 2020 Apr 13;17(8):2650 [FREE Full text] [doi: 10.3390/ijerph17082650] [Medline: 32294911]
38. Trigo JD, Rubio Ó, Martínez-Espronceda M, Alesanco Á, García J, Serrano-Arriezu L. Building standardized and secure mobile health services based on social media. Electronics 2020 Dec 21;9(12):2208. [doi: 10.3390/electronics9122208]
39. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
40. Jiao W, Chang A, Ho M, Lu Q, Liu MT, Schulz PJ. Predicting and empowering health for generation Z by comparing health information seeking and digital health literacy: cross-sectional questionnaire study. J Med Internet Res 2023 Oct 30;25:e47595 [FREE Full text] [doi: 10.2196/47595] [Medline: 37902832]
41. Jiang Y. Impact of community proactive health management application on electronic health literacy and self-management of hypertensive patients. Public Health Nurs 2024;41(6):1436-1445. [doi: 10.1111/phn.13424] [Medline: 39297673]
42. Wang W, Zhang H. Behavior patterns and influencing factors: health information acquisition behavior of Chinese senior adults on WeChat. Heliyon 2023 May 29;9(6):e16431 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e16431] [Medline: 37303534]
43. Ghazi SN, Berner J, Anderberg P, Sanmartin Berglund J. The prevalence of eHealth literacy and its relationship with perceived health status and psychological distress during Covid-19: a cross-sectional study of older adults in Blekinge, Sweden. BMC Geriatr 2023 Jan 04;23(1):5 [FREE Full text] [doi: 10.1186/s12877-022-03723-y] [Medline: 36597040]
44. Liu S, Wang XQ, Yang BX, Luo D, Liu Y, Fang XJ, et al. Electronic health literacy among older adults in the context of the COVID-19 pandemic: a mixed-methods study. J Nurs Manag 2022 Sep;30(6):1949-1959 [FREE Full text] [doi: 10.1111/jonm.13664] [Medline: 35509174]
45. Shi Y, Ma D, Zhang J, Chen B. In the digital age: a systematic literature review of the e-health literacy and influencing factors among Chinese older adults. Z Gesundh Wiss 2023;31(5):679-687 [FREE Full text] [doi: 10.1007/s10389-021-01604-z] [Medline: 34104627]

XSL•FO
RenderX

46. Pourrazavi S, Kouzekanani K, Bazargan-Hejazi S, Shaghaghi A, Hashemiparast M, Fathifar Z, et al. Theory-based e-health literacy interventions in older adults: a systematic review. Arch Public Health 2020 Aug 10;78:72 [FREE Full text] [doi: 10.1186/s13690-020-00455-6] [Medline: 32793345]
47. Wiwatkunupakarn N, Pateekhum C, Aramrat C, Jirapornchaoren W, Pinyopornpanish K, Angkurawaranon C. Social networking site usage: a systematic review of its relationship with social isolation, loneliness, and depression among older adults. Aging Ment Health 2022 Jul;26(7):1318-1326. [doi: 10.1080/13607863.2021.1966745] [Medline: 34427132]
48. Lei X, Matovic D, Leung WY, Viju A, Wuthrich VM. The relationship between social media use and psychosocial outcomes in older adults: a systematic review. Int Psychogeriatr 2024 Sep;36(9):714-746 [FREE Full text] [doi: 10.1017/S1041610223004519] [Medline: 38287785]
49. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
50. World population prospects 2017 – data booklet. United Nations, Department of Economic and Social Affairs, Population Division. 2017. URL: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2017_world_population_prospects-2017_revision_databooklet.pdf [accessed 2025-03-17]
51. Matchanova A, Woods SP, Neighbors C, Medina LD, Podell K, Beltran-Najera I, et al. Are accuracy discernment and sharing of COVID-19 misinformation associated with older age and lower neurocognitive functioning? Curr Psychol 2023 Mar 08:1-13 (forthcoming) [FREE Full text] [doi: 10.1007/s12144-023-04464-w] [Medline: 37359606]
52. Retirement age. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=Retirement_age&oldid=1236301617 [accessed 2024-08-03]
53. Willis E. Applying the health belief model to medication adherence: the role of online health communities and peer reviews. J Health Commun 2018;23(8):743-750. [doi: 10.1080/10810730.2018.1523260] [Medline: 30280998]
54. Aung MN, Koyanagi Y, Nagamine Y, Nam EW, Mulati N, Kyaw MY, et al. Digitally Inclusive, Healthy Aging Communities (DIHAC): a cross-cultural study in Japan, Republic of Korea, Singapore, and Thailand. Int J Environ Res Public Health 2022 Jun 07;19(12):6976 [FREE Full text] [doi: 10.3390/ijerph19126976] [Medline: 35742230]
55. Mao L, Mohan G, Normand C. Use of information communication technologies by older people and telemedicine adoption during COVID-19: a longitudinal study. J Am Med Inform Assoc 2023 Nov 17;30(12):2012-2020 [FREE Full text] [doi: 10.1093/jamia/ocad165] [Medline: 37572310]
56. Evans HE, Forbes CC, Galvão DA, Vandelanotte C, Newton RU, Wittert G, et al. Usability, acceptability, and safety analysis of a computer-tailored web-based exercise intervention (ExerciseGuide) for individuals with metastatic prostate cancer: multi-methods laboratory-based study. JMIR Cancer 2021 Jul 28;7(3):e28370 [FREE Full text] [doi: 10.2196/28370] [Medline: 34318759]
57. Yang E, Lee KH. The moderating effects of disability on mobile internet use among older adults: population-based cross-sectional study. J Med Internet Res 2022 Apr 04;24(4):e37127 [FREE Full text] [doi: 10.2196/37127] [Medline: 35377329]
58. Thompson J, Tod A, Bissell P, Bond M. Understanding food vulnerability and health literacy in older bereaved men: a qualitative study. Health Expect 2017 Dec;20(6):1342-1349 [FREE Full text] [doi: 10.1111/hex.12574] [Medline: 28544135]
59. Tan QY, Roberts HC, Fraser SD, Amar K, Ibrahim K. What are the modifiable factors of treatment burden and capacity among people with Parkinson's disease and their caregivers: a qualitative study. PLoS One 2023 Mar 30;18(3):e0283713 [FREE Full text] [doi: 10.1371/journal.pone.0283713] [Medline: 36996125]
60. Turnbull ML, Jin Y, Yau AH, Lai MS, Cheung MY, Kwan WY, et al. mHealth in hyper-connected Hong Kong: examining attitudes and access to mobile devices and health information among older Chinese residents. Mhealth 2021 Jul 20;7:43 [FREE Full text] [doi: 10.21037/mhealth-20-123] [Medline: 34345620]
61. Munger K, Gopal I, Nagler J, Tucker JA. Accessibility and generalizability: are social media effects moderated by age or digital literacy? Res Politics 2021 Jun 09;8(2). [doi: 10.1177/20531680211016968]
62. Tirado-Morueta R, Aguaded-Gómez JI, Hernando-Gómez Á. The socio-demographic divide in internet usage moderated by digital literacy support. Technol Soc 2018 Nov;55:47-55. [doi: 10.1016/j.techsoc.2018.06.001]
63. Delello JA, McWhorter RR. Reducing the digital divide: connecting older adults to iPad technology. J Appl Gerontol 2017 Jan;36(1):3-28. [doi: 10.1177/0733464815589985] [Medline: 26084479]
64. Dhakal U, Koumoutzis A, Vivoda JM. Better together: social contact and loneliness among U.S. older adults during COVID-19. J Gerontol B Psychol Sci Soc Sci 2023 Feb 19;78(2):359-369 [FREE Full text] [doi: 10.1093/geronb/gbac136] [Medline: 36112389]
65. Cipolletta S, Gris F. Older people's lived perspectives of social isolation during the first wave of the COVID-19 pandemic in Italy. Int J Environ Res Public Health 2021 Nov 11;18(22):11832 [FREE Full text] [doi: 10.3390/ijerph182211832] [Medline: 34831586]
66. Hošnjak AM, Pavlović A. Older adults knowledge about using smart technology during the Covid-19 crisis-a qualitative pilot study. IFAC Pap OnLine 2021;54(13):675-679 [FREE Full text] [doi: 10.1016/j.ifacol.2021.10.529] [Medline: 38620852]

XSL•FO
RenderX

67. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: cross-sectional survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
68. Karkhah S, Maroufizadeh S, Hakimi E, Ghazanfari MJ, Osuji J, Javadi-Pashaki N. Information seeking behavior on COVID-19 among older adults: a cross-sectional study in northern Iran. Gerontol Geriatr Med 2022 Sep 06;8:23337214221120746 [FREE Full text] [doi: 10.1177/23337214221120746] [Medline: 36090317]
69. Sirriyeh R, Lawton R, Gardner P, Armitage G. Reviewing studies with diverse designs: the development and evaluation of a new tool. J Eval Clin Pract 2012 Aug;18(4):746-752. [doi: 10.1111/j.1365-2753.2011.01662.x] [Medline: 21410846]
70. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
71. Bronfenbrenner U. The Ecology of Human Development: Experiments by Nature and Design. Cambridge, MA: Harvard University Press; 1979.
72. Kordzadeh N. Social media in health care. In: Wickramasinghe N, Troshani I, Tan J, editors. Contemporary Consumer Health Informatics. Cham, Switzerland: Springer; Mar 22, 2016.
73. Farsi D, Martinez-Menchaca HR, Ahmed M, Farsi N. Social media and health care (part II): narrative review of social media use by patients. J Med Internet Res 2022 Jan 07;24(1):e30379 [FREE Full text] [doi: 10.2196/30379] [Medline: 34994706]
74. Jane M, Hagger M, Foster J, Ho S, Pal S. Social media for health promotion and weight management: a critical debate. BMC Public Health 2018 Jul 28;18(1):932 [FREE Full text] [doi: 10.1186/s12889-018-5837-3] [Medline: 30055592]
75. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]
76. Paton C, Bamidis PD, Eysenbach G, Hansen M, Cabrer M. Experience in the use of social media in medical and health education. Yearb Med Inform 2011;20(01):21-29. [doi: 10.1055/s-0038-1638732]
77. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002 Jun;26(2):xiii-xxiii. [doi: 10.2307/4132319]
78. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
79. Shang L, Zuo M. Investigating older adults' intention to learn health knowledge on social media. Educ Gerontol 2020 May 03;46(6):350-363. [doi: 10.1080/03601277.2020.1759188]
80. Ye W. The impact of internet health information usage habits on older adults' e-health literacy. Digit Health 2024 May 07;10:20552076241253473 [FREE Full text] [doi: 10.1177/20552076241253473] [Medline: 38726215]
81. Chai X. How has the nationwide public health emergency of the COVID-19 pandemic affected older Chinese adults' health literacy, health behaviors and practices, and social connectedness? Qualitative evidence from urban China. Front Public Health 2021 Mar 10;9:774675 [FREE Full text] [doi: 10.3389/fpubh.2021.774675] [Medline: 35356089]
82. Jang Y, Jung J, Park NS, Kim MT, Chung S. COVID-19 health information sources and their associations with preventive behaviors: a typological study with older residents in Seoul, South Korea. PLoS One 2023 Aug 31;18(8):e0290824 [FREE Full text] [doi: 10.1371/journal.pone.0290824] [Medline: 37651388]
83. Lee J, Ryu H. Usability of a new digital walking program for older adults: a pilot study. BMC Geriatr 2023 Apr 01;23(1):193 [FREE Full text] [doi: 10.1186/s12877-023-03739-y] [Medline: 37003966]
84. Ubolwan K, Kheokao J, Yingrengreong S, Chuaintha E. Evaluation of factors associated with eHealth literacy among older adult social media users in Thailand. Southeast Asian J Trop Med Public Health 2020;51(5):762.
85. Chen AT, Chu F, Teng AK, Han S, Lin SY, Demiris G, et al. Promoting problem solving about health management: a mixed-methods pilot evaluation of a digital health intervention for older adults with pre-frailty and frailty. Gerontol Geriatr Med 2021 Jan 09;7:2333721420985684 [FREE Full text] [doi: 10.1177/2333721420985684] [Medline: 33457461]
86. Oh KM, An K, Lee M, Shin C, Steves SL. Colorectal cancer screening disparities in Asian Americans: the influences of patient-provider communication and social media use. Cancer Causes Control 2023 Sep;34(9):813-827 [FREE Full text] [doi: 10.1007/s10552-023-01720-z] [Medline: 37266764]
87. Vitolo M, Ziveri V, Gozzi G, Busi C, Imberti JF, Bonini N, et al. DIGItal health literacy after COVID-19 outbreak among frail and non-frail cardiology patients: the DIGI-COVID study. J Pers Med 2022 Dec 31;13(1):99 [FREE Full text] [doi: 10.3390/jpm13010099] [Medline: 36675760]
88. Tan M, Straughan PT, Cheong G. Information trust and COVID-19 vaccine hesitancy amongst middle-aged and older adults in Singapore: a latent class analysis approach. Soc Sci Med 2022 Mar;296:114767 [FREE Full text] [doi: 10.1016/j.socscimed.2022.114767] [Medline: 35144226]
89. Bronfenbrenner U. Making Human Beings Human: Bioecological Perspectives on Human Development. Thousand Oaks, CA: SAGE Publications; 2005.
90. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]

XSL•FO
RenderX

91. Feinberg I, Tighe EL, Greenberg D, Mavreles M. Health literacy and adults with low basic skills. Adult Educ Q 2018 Jun 20;68(4):297-315. [doi: 10.1177/0741713618783487]
92. Fenta ET, Kidie AA, Tiruneh MG, Anagaw TF, Bogale EK, Dessie AA, et al. Exploring barriers of health literacy on non-communicable disease prevention and care among patients in north Wollo zone public hospitals; Northeast, Ethiopia, 2023: application of socio-ecological model. BMC Public Health 2024 Apr 06;24(1):971 [FREE Full text] [doi: 10.1186/s12889-024-18524-8] [Medline: 38581006]
93. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
94. Kumar A, Tandon P, Meratwal G. Electronic health literacy and demand for digital health technology among the elderly in Rajasthan. J Surg Spec Rural Pract 2023;4(2):57-62. [doi: 10.4103/jssrp.jssrp_5_23]
95. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
96. Ma X, Lu J, Liu W. Influencing factors on health information to improve public health literacy in the official WeChat account of Guangzhou CDC. Front Public Health 2021 Aug 3;9:657082 [FREE Full text] [doi: 10.3389/fpubh.2021.657082] [Medline: 34414152]
97. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in China: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
98. Sun X, Yan W, Zhou H, Wang Z, Zhang X, Huang S, et al. Internet use and need for digital health technology among the elderly: a cross-sectional survey in China. BMC Public Health 2020 Sep 11;20(1):1386 [FREE Full text] [doi: 10.1186/s12889-020-09448-0] [Medline: 32917171]
99. Banerjee D. 'Age and ageism in COVID-19': elderly mental health-care vulnerabilities and needs. Asian J Psychiatr 2020 Jun;51:102154 [FREE Full text] [doi: 10.1016/j.ajp.2020.102154] [Medline: 32403024]
100. Petretto DR, Pili R. Ageing and COVID-19: what is the role for elderly people? Geriatrics (Basel) 2020 Apr 26;5(2):25 [FREE Full text] [doi: 10.3390/geriatrics5020025] [Medline: 32357582]
101. Zhang Q, Song W. The challenges of the COVID-19 pandemic: approaches for the elderly and those with Alzheimer's disease. MedComm (2020) 2020 Jun 15;1(1):69-73 [FREE Full text] [doi: 10.1002/mco2.4] [Medline: 34172985]
102. CSM releases survey report on user media consumption and usage expectations during the epidemic. CSM Media Research. 2020 Feb 25. URL: https://www.csm.com.cn/Content/2020/02-25/1015314370.html [accessed 2024-07-11]
103. Lwin MO, Panchapakesan C, Sheldenkar A, Calvert GA, Lim LK, Lu J. Determinants of eHealth literacy among adults in China. J Health Commun 2020 May 03;25(5):385-393. [doi: 10.1080/10810730.2020.1776422] [Medline: 32552607]
104. Schreurs L, Vandenbosch L. Introducing the Social Media Literacy (SMILE) model with the case of the positivity bias on social media. J Child Media 2020 Aug 16;15(3):320-337. [doi: 10.1080/17482798.2020.1809481]
105. Munusamy S, Syasyila K, Shaari AA, Pitchan MA, Kamaluddin MR, Jatnika R. Psychological factors contributing to the creation and dissemination of fake news among social media users: a systematic review. BMC Psychol 2024 Nov 18;12(1):673 [FREE Full text] [doi: 10.1186/s40359-024-02129-2] [Medline: 39558439]
106. Li Y, Mohamad E, Azlan AA. The impact of social media dependency on Chinese netizens' willingness to forward health rumors: the moderating role of social media literacy. Stud Media Commun 2025;13(1):154-162. [doi: 10.11114/smc.v13i1.7393]
107. Lim JH, Ahmad MB, Prihadi KD. New media literacy, self-acceptance, and psychological flexibility in enhancing Gen Z wellbeing. Psychol Soc Educ 2024 Nov 30;16(3):10-19. [doi: 10.21071/pse.v16i3.17235]
108. Jokisch MR, Schmidt LI, Doh M. Acceptance of digital health services among older adults: findings on perceived usefulness, self-efficacy, privacy concerns, ICT knowledge, and support seeking. Front Public Health 2022 Dec 13;10:1073756 [FREE Full text] [doi: 10.3389/fpubh.2022.1073756] [Medline: 36582385]
109. Pywell J, Vijaykumar S, Dodd A, Coventry L. Barriers to older adults' uptake of mobile-based mental health interventions. Digit Health 2020 Feb 11;6:2055207620905422 [FREE Full text] [doi: 10.1177/2055207620905422] [Medline: 32110429]
110. Zhou C, Guo J, Zhang Y, Huang T, Kaner J. A home health management app designed in the post-epidemic era using empirical evidence based on the demand collection of elderly users. Humanit Soc Sci Commun 2024 May 13;11:619. [doi: 10.1057/s41599-024-03042-0]
111. Blinka MD, Buta B, Bader KD, Hanley C, Schoenborn NL, McNabney M, et al. Developing a sensor-based mobile application for in-home frailty assessment: a qualitative study. BMC Geriatr 2021 Feb 04;21(1):101 [FREE Full text] [doi: 10.1186/s12877-021-02041-z] [Medline: 33541276]
112. Malek Rivan NF, Shahar S, Singh DK, Che Din N, Mahadzir H, You YX, et al. Development of cognitive frailty screening tool among community-dwelling older adults. Heliyon 2024 Jul 06;10(14):e34223 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e34223] [Medline: 39104490]
113. Marzuki AA, Wong KY, Chan JK, Na SY, Thanaraju A, Phon-Amnuaisuk P, et al. Mapping computational cognitive profiles of aging to dissociable brain and sociodemographic factors. NPJ Aging 2024 Oct 31;10(1):50. [doi: 10.1038/s41514-024-00171-3] [Medline: 39482289]

XSL•FO
RenderX

114. Samad-Soltani T, Rezaei-Hachesu P, Ghazisaeedi M. Pervasive decision support systems in healthcare using intelligent robots in social media. Iran J Public Health 2017 Jan;46(1):148-149 [FREE Full text] [Medline: 28451546]
115. Shen L, Wang S, Chen W, Fu Q, Evans R, Lan F, et al. Understanding the function constitution and influence factors on communication for the WeChat official account of top tertiary hospitals in China: cross-sectional study. J Med Internet Res 2019 Dec 09;21(12):e13025 [FREE Full text] [doi: 10.2196/13025] [Medline: 31815674]
116. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
117. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. 'Having to learn this so late in our lives…' Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar;37(1):41-52 [FREE Full text] [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
118. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 02;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
119. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
120. Woodman RJ, Mangoni AA. A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. Aging Clin Exp Res 2023 Nov;35(11):2363-2397 [FREE Full text] [doi: 10.1007/s40520-023-02552-2] [Medline: 37682491]
121. Qian K, Zhang Z, Yamamoto Y, Schuller BW. Artificial intelligence internet of things for the elderly: from assisted living to health-care monitoring. IEEE Signal Process Mag 2021 Jul;38(4):78-88. [doi: 10.1109/msp.2021.3057298]
122. Zhao N, Zhao W, Tang X, Jiao C, Zhang Z. Design of health information management model for elderly care using an advanced higher-order hybrid clustering algorithm from the perspective of sports and medicine integration. PLoS One 2024 May 17;19(5):e0302741 [FREE Full text] [doi: 10.1371/journal.pone.0302741] [Medline: 38758774]
123. Nassetta J, Gross K. State media warning labels can counteract the effects of foreign misinformation. Harv Kennedy Sch Misinformation Rev 2020 Oct 30 [FREE Full text] [doi: 10.37016/MR-2020-45]
124. Arthur EK, Pisegna J, Oliveri JM, Aker H, Krok-Schoen JL. Older cancer survivors' perspectives and use of telehealth in their cancer survivorship care in the United States: a ResearchMatch® sample. J Geriatr Oncol 2022 Nov;13(8):1223-1229. [doi: 10.1016/j.jgo.2022.08.004] [Medline: 35985929]
125. Gao Z, Ryu S, Zhou W, Adams K, Hassan M, Zhang R, et al. Effects of personalized exercise prescriptions and social media delivered through mobile health on cancer survivors' physical activity and quality of life. J Sport Health Sci 2023 Nov;12(6):705-714 [FREE Full text] [doi: 10.1016/j.jshs.2023.07.002] [Medline: 37467931]
126. Kim HS, Sung JH. The influence of digital informatization level, self-efficacy, and social support on digital health literacy in the elderly with cancer. Asian Oncol Nurs 2022 Dec 31;22(4):255-263. [doi: 10.5388/aon.2022.22.4.255]
127. Zhou T. Understanding online health community users' information adoption intention: an elaboration likelihood model perspective. Online Inf Rev 2021 Jun 01;46(1):134-146. [doi: 10.1108/oir-09-2020-0412]
128. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43:115-135. [doi: 10.1007/s11747-014-0403-8]
129. Yuan S, Hussain SA, Hales KD, Cotten SR. What do they like? Communication preferences and patterns of older adults in the United States: the role of technology. Educ Gerontol 2015 Dec 31;42(3):163-174. [doi: 10.1080/03601277.2015.1083392]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QA: quality assessment
RQ: research question
SEM: social-ecological model
VHC: virtual health community

XSL•FO
RenderX

Edited by A Schwartz; submitted 03.09.24; peer-reviewed by T Huang, T Akindahunsi, K Elechi; comments to author 26.11.24; revised version received 15.01.25; accepted 19.02.25; published 26.03.25.

Please cite as:
Zhang C, Mohamad E, Azlan AA, Wu A, Ma Y, Qi Y
Social Media and eHealth Literacy Among Older Adults: Systematic Literature Review
J Med Internet Res 2025;27:e66058
URL: https://www.jmir.org/2025/1/e66058
doi:10.2196/66058
PMID:

©Chenglin Zhang, Emma Mohamad, Arina Anis Azlan, Anqi Wu, Yilian Ma, Yihan Qi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Use of Mukbang in Health Promotion: Scoping Review

Xiao Wang[1,2], MSc; Yuxue Xiao[1], MSc; Sujin Nam[3], PhD; Ting Zhong[4], BSN; Dongyan Tang[4], BSN; William Ho Cheung Li[5], PhD; Peige Song[6], MSc, MD, PhD; Wei Xia[1], MPhil, PhD

[1]School of Nursing, Sun Yat-Sen University, Guangzhou, China

[2]Peking University Sixth Hospital, Beijing, China

[3]College of Nursing, Yonsei University, Seoul, Republic of Korea

[4]Sun Yat-sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China

[5]Nethersole School of Nursing, The Chinese University of Hong Kong, Hong Kong, China

[6]School of Public Health, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Wei Xia, MPhil, PhD
School of Nursing
Sun Yat-Sen University
North Campus, No 74, 2nd Yat-Sen Road
Yuexiu District
Guangzhou, 510080
China
Phone: 86 02087334851
Fax: 86 020 873330
Email: xiaw23@mail.sysu.edu.cn

Abstract

Background: Mukbang is a recent internet phenomenon in which anchors publicly record and show their eating through short video platforms. Researchers reported a tangible impact of mukbang on the psychological and physical health, appetite, and eating behavior of the public, it is critical to obtain clear and comprehensive insights concerning the use of mukbang to promote the viewers' appetite, eating behaviors, and health to identify directions for future work.

Objective: This scoping review aims to comprehensively outline the current evidence regarding the impact of mukbang consumption on dietary behaviors, appetite regulation, flavor perception, and physical and psychological well-being. Specifically, we conducted an analysis of public perceptions and attitudes toward mukbang while summarizing the reciprocal influence it has on health promotion.

Methods: This study was conducted as a scoping review following the Joanna Briggs Institute guideline and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. We comprehensively searched 8 electronic databases in Chinese, English, and Korean languages. We also searched gray literature sources like Google Scholar and ProQuest. We used a data extraction chart to extract information relevant to the impact of mukbang on health. The extracted data were qualitatively analyzed to form different themes related to health, categorizing and integrating the results based on the type of study (qualitative, observational, and experimental).

Results: This scoping review finally included 53 studies; the annual distribution exhibited a consistent upward trend across all categories since their initial publication in 2017. Based on the results of the analysis, we have summarized 4 themes, which showed that mukbang may have positive effects on viewers' appetite, food choices, and weight control; it can also meet the psychological needs of viewers and provide digital companionship and happiness. However, excessive viewing may also be harmful to viewer's health, which has also caused health concerns for some viewers.

Conclusions: This study conducted a comprehensive search, screening, and synthesis of existing studies focusing on mukbang and health across various languages and varying levels of quality, which has presented the analytical evidence of the relationship between mukbang and dietary behaviors, appetite, flavor perception, and health. According to the results, future research could consider analyzing the beneficial and harmful factors of mukbang, thereby further optimizing the existing mukbang videos accordingly to explore the potential of using mukbang for health intervention or promotion, so as to improve or customize the content of mukbang based on this scoping review, maximize the appetite and health promotion effects of mukbang videos.

Trial Registration: INPLASY INPLASY2022120109; https://inplasy.com/inplasy-2022-12-0109/

XSL•FO
RenderX

(J Med Internet Res 2025;27:e56147) doi:10.2196/56147

KEYWORDS

mukbang; health promotion; eating behaviors; appetite; scoping review

Introduction

Background

Mukbang (or meokbang), a portmanteau of the Korean words "eat" (*meokneun*) and "broadcast" (*bangsong*), also known as "eating broadcast," is an emergent new media phenomenon that first appeared in 2008 [1]; in this phenomenon, mukbang hosts sit in front of the camera for a few hours and live stream their consumption of vast quantities of food while interacting with viewers [2]. Mukbang became popular in South Korea in 2010 and was broadcast through a video streaming platform AfreecaTV (now called "SOOP") [3]. Due to the advancement of the internet and smartphones, short video platforms, such as YouTube and TikTok are gradually gaining worldwide popularity. Short video platform development has enabled the evolution of the presentation form of mukbang from live broadcasts to prerecorded videos by facilitating the editing and release of these videos to a larger audience [4], enhancing the popularity of mukbang [5], and making mukbang a global trend in 2015 [6]. Currently, mukbang enjoys enormous web-based engagement. Popular mukbang channels have more than 5 million YouTube subscribers, and a single video often attracts hundreds of millions of views [7]. An investigation of the characteristics of mukbang audiences showed that mukbang is more attractive to adolescents aged between 18-28 years [8], and the male-to-female viewer ratio is approximately 1:2 [8]. The YouTube keyword tool indicates the average monthly numbers of searches of the term mukbang over the previous 12 months of data for the United States of America and South Korea to be 365,000 and 9600, respectively [9]. This suggests that mukbang is ubiquitous; it spreads into our living environments and imperceptibly affects people's lives and daily interactions, particularly among young people. In this study, according to the definition and origin of mukbang, we classify videos with "eating" behavior as their content as mukbang; in other words, mukbang is no longer limited to the display of eating in a live broadcast but covers all web-based videos having eating-related contents.

Earlier studies have extensively investigated and analyzed the components and contents of mukbang. According to these studies, mukbang comprises 2 essential elements: auditory and visual sensory stimulation [10] and web-based interaction [11,12]. In sensory stimulation, distinct chomping, chewing sounds made by hosts can evoke a sensation known as an autonomous sensory meridian response (ASMR) [3,13-15], which refers to experiencing a pleasantly warm and tingling sensation starting at the crown of one's head and spreading down the body [16]. In addition, the display of food in mukbang also stimulates viewers' desire for the specific food shown in the videos [17-19]. Another element of mukbang is web-based interaction, which is established by bullet chatting, giving gifts, or liking posts [10,20]. Web-based interaction enables viewers to create emotional relationships with other viewers and the host [21,22]. Further, mukbang's 2 essential elements can both affect viewers' emotions and viewing motivations [23]; closely linked to their different psychological needs; and bring them happiness, peace, and vicarious satisfaction [20].

The current trend of mukbang viewing suggests a potential correlation between watching mukbang and specific physiological or psychological needs. Many studies adopt an empirical approach to examine people's motivation to watch mukbang. Such studies divide the main motivations for watching mukbang into pursuing the thrills of hunger, vicarious eating, emotional establishment, alleviating loneliness, and relieving anorexia [24]. Due to the fact that most of the young viewers are highly educated and experiencing high social pressure and social isolation [25], alleviating the feelings of loneliness associated with eating alone is the most frequently mentioned reason for watching mukbang [21]. Mukbang hosts share fascinating aspects of their lives with their audience while eating, which alleviates the viewers' feelings of social loneliness. This was particularly relevant during the COVID-19 pandemic [26,27]. Another common motivation to watch mukbang during the pandemic period was to obtain information about food, such as its taste, nutritional value, and cooking procedures [27,28]. Today, social demands to maintain body shape are on the rise; hence, an increasing number of people watch mukbang to experience vicarious satiation, satiety, and enjoyment through visual and auditory stimuli [29]. Furthermore, some viewers watch mukbang to avoid the problems and negative feelings caused by real-life experiences [30]. Therefore, mukbang is often considered to relieve loneliness, reduce negative emotions, and provide psychological pleasure and satisfaction, all of which attract scholars' attention to the relationship between mukbang and physical and psychological health. Compared with the studies on mukbang, research on food-related videos is more comprehensive. Further, scholars use experimental designs to examine the effects of food-related videos on viewers' psychology, appetite, and eating behavior from an objective perspective [15,17,31]. According to the results of 4 separate meta-analyses on neural responses to food advertising conducted by Arrona-Cardoza et al [32], van der Laan et al [33], Yeung [34], and Yang et al [35], food images or commercials can cause stronger brain responses than nonfood images or commercials in different areas of the brain. Furthermore, a comparison between high- and low-calorie foods revealed higher neural activation on viewing high-calorie foods than low-calorie foods, particularly among individuals with overweight or obesity [32,33,35,36]. This confirmed that, by affecting neurofeedback in the brain's food-reward system, food-related videos significantly influence viewers' appetite, food intake, and eating behaviors [37-39]. Further, Boyland et al [40] and Kidd and Loxton [41] designed a randomized controlled trial (RCT) that suggested that watching food advertisements increases viewers' food intake. On this basis, some studies indicate that watching mukbang can increase food consumption by augmenting taste sensations [26]. Further,

XSL•FO
RenderX

mukbang increases flavor perceptions, help individuals suppress their desire to substitute healthy food with unhealthy ones, and, thereby, improves individuals' appetite [14].

There is a growing body of research and informational studies on mukbang that encompasses various topics such as audience psychology, food culture, communication effects, food waste, physical and mental health, and the exploration of new mukbang trends. However, the divergent research on mukbang and health yields contradictory findings and conclusions. Researchers consider mukbang a useful tool to improve appetite and food intake [13,42], while in other studies, viewers believe that mukbang has a negative impact on health and dietary behaviors [31]. Furthermore, studies on mukbang's health effects are scattered across academic fields and fail to provide a general summary or description of the topic; the quality of these studies also varies considerably; thus, due to the complexity of information, readers may struggle to obtain effective insights into the impact of mukbang on physical and mental health. They might overlook the benefits of mukbang in stimulating appetite and increasing food intake or excessively rely on it for weight loss, leading to severe consequences such as overeating. Consequently, there is an urgent need for a comprehensive review that integrates key findings from both academic and nonacademic studies on mukbang's influence on health based on a thorough screening of existing research. After conducting a comprehensive search of academic and nonacademic studies on both mukbang and health, we integrate the main aspects of each paper to identify and discuss the impact of mukbang on viewers' physical and psychological health. Further, this study aims to systematically map the research on mukbang and health promotion, identify any research gaps, and reveal the impact of watching mukbang on viewers' physical and psychological health by combining, comparing, and integrating relevant literature to lay a foundation for the application of mukbang to promote the physical and psychological health of viewers.

Research Goals

Research reveals that mukbang may affect the physical and psychological health of mukbang hosts and viewers. This study systematically maps the positive aspects of and gaps in research on mukbang's impact on health to help prevent overweight and obesity and improve the eating behaviors and nutritional status of the public. This scoping analysis addresses the following research questions (RQs).

- RQ1: What are the specific effects of mukbang watching on physical and psychological health?

We reviewed and summarized studies exploring the relationship between mukbang and health or the effects of mukbang on health. Further, we identified the potential effects of mukbang on physical and psychological health promotion and clarify different mukbang-related ideas and health-related issues.

- RQ2: What are viewers' opinions regarding the impact of mukbang on their physical and psychological health?

We examined mukbang user groups and their perceptions of the relationship between mukbang and health to identify the specific requirements that should be satisfied or improved by mukbang.

- RQ3: What are the future applications of mukbang in the field of health?

Finally, based on the arrangement of the content on the health effects of mukbang in existing studies, we discussed the future prospects of mukbang in promoting appetite, improving dietary behaviors, and enhancing mental health to lay the foundation for its application in medical and health fields.

The findings of this study are intended to benefit health care professionals, researchers, technology providers, and all (both current and potential) viewers in designing, conducting, and evaluating mukbang videos.

Methods

Study Design

This scoping review is conducted and reported following the Joanna Briggs Institute (JBI) guideline and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [43]. Further, this scoping review protocol was prospectively registered (INPLASY2022120109) to limit reporting bias. The initial search was conducted in March 2023, and the literature screening and data extraction processes were completed in July 2023.

Search Strategy

To conduct a comprehensive review, we considered both academic and nonacademic studies on mukbang and health issues. Textbox 1 presents the inclusion and exclusion criteria for this study. We used the keywords mukbang, meokbang, "eating broadcast," "culinary videos," "online eating," "eating show," "food media," "health," "health promotion," "eating disorder," "eating behaviors," "disordered eating," "binge eating," and "obesity" to search studies that were published before June 30, 2023, in academic databases. The same keywords in Chinese and Korean were also searched in Chinese and Korean databases, respectively, using similar strategies adjusted based on the database. For gray literature sources, we used simple terms mukbang, "eating broadcast," "online eating," ASMR, and health for more comprehensive results. The academic databases were PubMed, Embase, Web of Science, Researching Information Sharing Service (Korean), DBpia Scholarly Database (Korean), China National Knowledge Infrastructure (Chinese), China Science and Technology Journal Database (VIP Database, Chinese), and Wan Fang Data (Chinese). The following study types were retrieved from these databases using optimized searches: review, meta-analysis, observational quantitative, qualitative, mixed, and interventional studies on mukbang and health issues. Gray literature searches were conducted of databases Open Access theses and dissertations, ProQuest, OpenGrey, and Google Scholar, including commentaries, magazine papers, news, blogs, books or book chapters, and ongoing studies, to map the complete academic scope of the impact of mukbang watching on viewers' health and eating behaviors. We used "Title/Abstract" or "All Fields" as retrieval fields and a Boolean strategy to connect each identified keyword to an appropriate search strategy. We used the conjunction "or" to connect synonyms, while

XSL•FO
RenderX

parentheses were used for prioritized searches, and all the keywords with distinct meanings were linked by "and" (Multimedia Appendix 2). The bibliographies of all relevant retrieved studies were also examined to identify further relevant studies. The searched studies were written in English, Chinese, or Korean languages. The decision to consider these languages was motivated by the following consensus since mukbang originated in South Korea, the literature on mukbang in Korean is informative and, hence, should be searched. In addition, since the researchers, who are from China, were interested in identifying possible studies (in Chinese) from China and mukbang has extensive dissemination and audience in China, we decided to include studies from Chinese and Korean databases.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Studies describing the relationship between mukbang watching and dietary behaviors, disordered eating, or physical and psychological health.
- Studies demonstrating the impact of mukbang on health (eg, physical health, dietary behaviors, psychological health).
- Studies exploring the reasons for mukbang watching from the perspective of psychological or personal needs, while paying attention to the impact of mukbang on public health.
- Studies focusing on the design, development, or availability of the interventions using mukbang to demonstrate the usefulness of mukbang in the health field.
- Studies focusing on the challenges and obstacles of integrating mukbang videos into clinical practice.
- Food advertisements and food-related videos must include images of celebrities eating.

Exclusion criteria

- Studies only involve the basic concepts and situations such as the definition, origin, development, and epidemic situation of mukbang.
- Studies only cover the impact of shooting mukbang videos on the host's own physical and psychological health but do not mention the effect on the viewers.
- Studies only mention digital media, social software, or social media, but do not clearly indicate mukbang in the full text. Studies only contain food pictures (static representations of food).
- The main field of study was communication, literature, or aesthetics, rather than medicine, psychology, or health.

Study Selection

In this phase, we evaluated the retrieved studies' relevance to the review according to the independent inclusion and exclusion criteria presented in Textbox 1. All identified records from the academic databases and gray literature websites were imported into Rayyan (Qatar Computing Research Institute) and EndNote 20 (version 4.1; Clarivate) for further analysis and screening. The Rayyan platform enables multiple collaborators to independently and synchronously screen a shared set of studies, ensuring the confidentiality of each collaborator's results until the final stage of each step of the screening process. After importing all the data, duplicate results were removed. Two researchers independently screened the titles and abstracts of the retrieved studies and excluded the studies that did not contain mukbang or health elements. The studies that definitely and potentially satisfied the inclusion criteria were tagged "included" and "maybe" using the screening and classification function of Rayyan, respectively. Subsequently, an independent assessment of full-text studies was performed. The full texts of the studies that did not match the search theme or review's scope were excluded from the study. All the studies whose full texts were unavailable were excluded, as well. Disagreements were solved through discussions and consensus with a third researcher. After the full-text selection, JBI critical appraisal tools were used to evaluate the quality of the included cross-sectional studies, RCTs, quasiexperimental studies, systematic reviews, and qualitative research. We did not conduct a quality appraisal for commentaries since there is no unified writing format or standard requirements.

Data Extraction and Charting

According to a framework of the JBI template data extraction instrument [44], researchers developed and modified a data extraction chart based on the inclusion criteria, evidence source details and characteristics, information on research methods or tools, mukbang characteristics, and health-related elements (Multimedia Appendix 3). The data extraction chart was used as a research tool to manually extract and integrate useful information from the included studies. Search results were described and narratively synthesized to clarify the effects of mukbang watching on viewers' physical and psychological health and the comments of the viewers on the relationship between mukbang and health. Further, in line with the purpose of this review, the retrieved studies were sorted, analyzed, and synthesized to clarify the effects of mukbang watching on viewers' eating behaviors, appetite, food preferences, body weight, and psychological health. Accordingly, we highlighted the numbers and proportions of different types of publications and studies and the target audience, language, country of origin, publication year, disciplinary areas, specific research methods, and key findings of all the included studies (Multimedia Appendix 4). To present the results clearly, we used the year filter of each database to classify the retrieved studies by publication year and counted the numbers and types of studies published in different years. We then plotted the data of the included studies using Excel to graphically show the types of

XSL•FO
RenderX

included studies and their trends over the years. Our goal was to collect the contents of the studies describing the health effects of mukbang watching, as well as clarifying the public's views and attitudes regarding mukbang videos. Through inductive analysis, we characterized the influence of mukbang watching on viewers' physical and psychological health and described the different attitudes of viewers toward the health-related benefits of mukbang watching.

Results

Search Results

Our study search found 1465 publications in databases; 307 additional records were identified through gray literature resources, as well. Following the removal of duplicates, 1002 titles and abstracts were screened independently by 2 reviewers, and 278 ambiguous studies were discussed by reviewers together after the screening. In total, 876 studies were excluded since they did not examine mukbang watching or health; further, 72 studies were included because their titles or abstracts explicitly mentioned mukbang watching and health. In addition, 52 publications were allocated to the maybe list since these studies' eligibility criteria could not be determined completely from their titles or abstracts. Among them, 124 full-text studies were screened, and 66 publications were included in the review for result extraction and analysis based on the eligibility criteria. Among these 66 studies, since 13 were not available due to limitations such as incomplete full-text coverage, restricted access rights, or inability to establish contact with the original author, 53 publications were finally included in the review. According to the evaluation criteria of JBI [45], the quality evaluation studies basically conform to the writing format and standards of the corresponding research types, the quality evaluation of the included studies is presented in Multimedia Appendix 5 [13,15,17,18,22,24,25,27-30,46-70]. Given that the scoping review aims to ensure the most comprehensive understanding and characterization of existing research by searching for studies relevant to the topic to the greatest extent and including as many studies as possible that meet the inclusion criteria, quality evaluation results have not been used as exclusion criteria. The search and screening procedure is depicted in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1), and some important details of the included studies are presented in Multimedia Appendix 6 [2,13-15,17,18,22,24-31,46,47-69,71-83].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for studies screening.

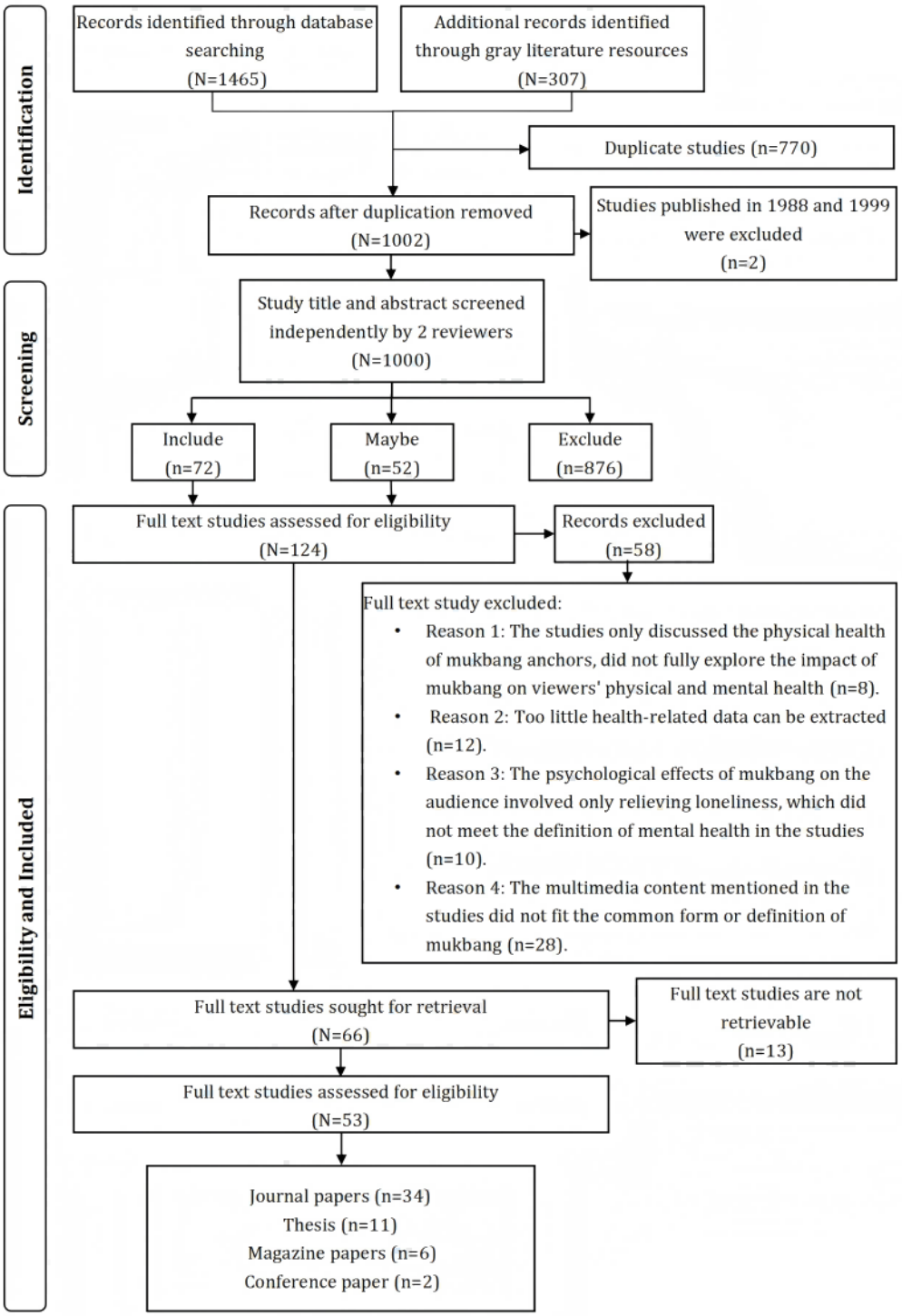

Study Characteristics

Overall, our search results indicated that studies on mukbang first appeared in 2015, whereas those on mukbang and health first appeared in 2017. Table 1 indicates the number of included studies classified based on different research types. In general, the types of studies on the relationship between mukbang and health mainly include cross-sectional studies, qualitative studies, reviews, commentaries, and intervention studies, which indicates that research focuses on the relationship between mukbang and health becomes more comprehensive, scholars are attempting to explore mukbang's effects on health using intervention studies with higher levels of evidence. Simultaneously, however, the commentary also accounted for a high proportion of the included studies (n=10, 19%), suggesting that despite the growing academic interest in studying mukbang's effects on health, many scholars also favor commentary as a genre that allows for greater expression of opinions and insights.

XSL•FO
RenderX

Table 1. The number of the included studies, grouped by different research types.

Research types	Studies, n (%)
Cross-sectional study	20 (38)
Interventional study	10 (19)
Commentary	10 (19)
Qualitative study	9 (17)
Review	4 (7)

This review considered studies published in scientific journals (n=34, 64.2%), master's theses (n=12, 22.6%), magazine papers (n=4, 7.5%), conference papers (n=1, 1.9%), book chapters (n=1, 1.9%), and study registrations (n=1, 1.9%). Figure 2 presents combined information on the characteristics and publication years of the included studies focusing on mukbang and health and retrieved studies. Since the earliest mukbang appeared in 2010, we excluded the studies published before 2010 (a total of 10 studies) from the search results. Consequently, the number of "all retrieved studies" amounts to 992. The stacked bar chart below represents the characteristics of all the retrieved resources; The stacked bar chart above represents the characteristics of the included studies. Moreover, to enhance the clarity of the comparison between all retrieved studies and included studies in Figure 2A, we have consolidated newspaper, book, and website studies under a category labeled "Other types of studies." According to Figure 2B, most studies examining the effects of mukbang on health and nutrition were journal papers and theses, which indicated that the content of mukbang related to dietary habits and health gradually attracted scholarly attention. The number of journal papers and theses on mukbang, which has been increasing since 2015, peaked in 2020 and 2021. South Korea was the first country to publish a mukbang-related study, probably because mukbang originated in Korea and became popular in Korean society earlier than in other countries. Korea is followed by China, which has been publishing a large number of studies every year since 2016. Most Chinese studies on mukbang were published in communication and journalism journals, whereas most English-language studies were published in journals on psychology, nutrition, and health. In this review, only the information relevant to mukbang's impact on physical and psychological health was extracted from the included studies.

XSL•FO
RenderX

Figure 2. The number of mukbang-related studies per year grouped by different characteristics: (A) all the included studies and (B) all the retrieved studies.

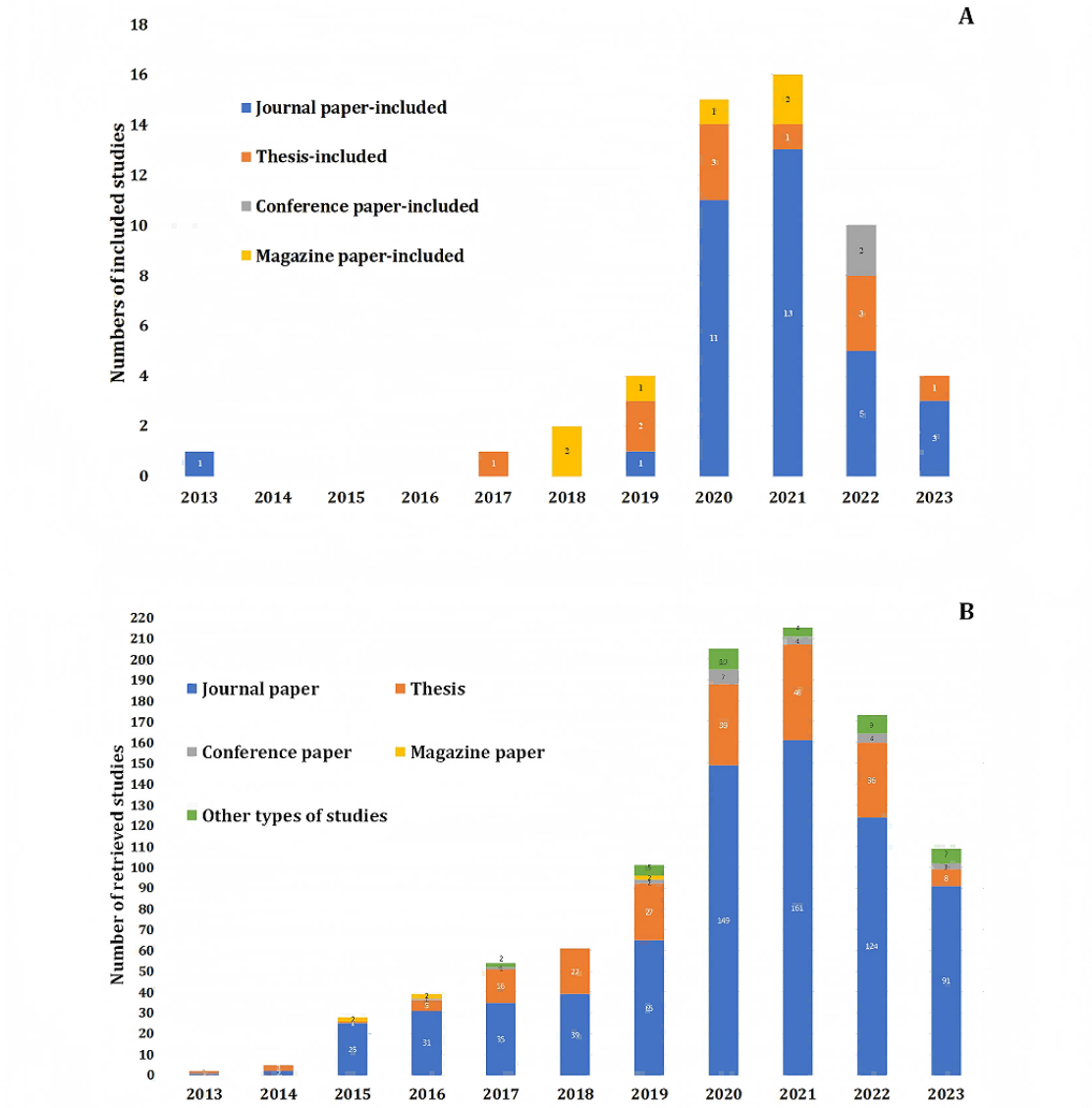

Mukbang's Impact on Physical Health

Studies found that mukbang has the potential to provide health-related information to the viewer, and may have positive effects on viewers' appetite [71], food choices [47], and weight control [48,72]. Based on the included cross-sectional studies, food-related programs such as mukbang and cooking shows, viewing frequency and duration [49] are commonly regarded having a positive impact on food choices, food intake, dietary behaviors [50], and overall health of the viewers [47,51], especially during the COVID-19 pandemic that many people decided to watch mukbang and cooking videos to entertain themselves and cook by themselves to diversify their meals and maintain a healthy lifestyle [27]. A cross-sectional survey also tentatively explored the potential for experiencing vicarious satisfaction through watching mukbang, which indicated that nearly a quarter of the audience watched mukbang when feeling hungry to experience satiation, satisfy their desire for food, and enjoy web-based food spuriously without gaining weight [52]. To explore the aforementioned findings further, some scholars have carried out qualitative and mixed studies. The studies have turned out that the nutritional content of food-related videos may have a positive effect on viewers' healthy eating behaviors, viewers tend to obtain interesting information about healthy eating from food-related videos [25]. Many observational studies indicated certain impacts of mukbang consumption on viewers' appetite, food intake, food choice, and dietary preference, some scholars subsequently conducted interventional studies to validate these causal relations. Kim [71] pointed out that watching mukbang with normal emotion may increase viewers' appetite and food intake. RCTs have been conducted to compare the effects of mukbang and nonfood videos, mukbang and cooking shows, and mukbang with different contents respectively on satiation, disgust, and desire for food, which found that mukbang reduced the viewer's sense of satiety and promoted their appetite to varying degrees, but it also increased the dieters' disgust to stabilize their food intake in a relatively balanced state. Interestingly, the study revealed that regardless of the food presented, watching mukbang invariably increased viewers' desire to eat fruits and vegetables [48]. The findings are further supported by other similar studies. Review and commentary studies suggested that mukbang can help people

XSL•FO
RenderX

with eating disorder symptoms experience a sense of vicarious satiation to control their appetite and, to some extent, stop endless eating [73]. In contrast, for individuals who engage in excessive dieting, watching mukbang could help them relax their strict dietary requirements and reduce their anxiety regarding dieting to avoid malnutrition [2]. These results confirm that mukbang can increase satiety, control food intake, and avoid overeating and, thereby, help viewers lose weight [2,53].

Nonetheless, mukbang was reported to licit negative effects on certain viewers in some cases. Preliminary investigations of cross-sectional studies indicated that mukbang contains large amounts of high-calorie food may cause concern that it can increase high-calorie food intake and impair the dietary behaviors of mukbang viewers [27,53,54], especially for those watching mukbang frequently [49,55]. von Ash et al [56] investigated the relationship between mukbang viewing and disordered eating behaviors using the Mukbang Addiction Scale and the eating disorders examination questionnaire, which showed a positive relationship between binge eating or purging tendencies and problematic mukbang viewing; moreover, mukbang viewing, and additional characteristics of mukbang viewing (eg, frequency of watching, average watch time, eating while watching), also be found that may be associated with eating disorder symptoms. On this basis, qualitative studies using content analysis and textual analysis to further explore the possible negative impact of mukbang on viewers found that about 83.5% of mukbang presented substantial content showing harmful eating habits like overeating, excessive exposure to it may change the original dietary preferences and eating habits of viewers and increasing their desire to consume amounts of unhealthy foods [18,57]. While some famous mukbang hosts demonstrate excessive food consumption, they are still energetic and keep in shape (particularly in Big Eater Mukbang videos), this contrast may mislead viewers to believe that overeating does not pose any risks to their health and physique [24]. These perspectives above are supported by some commentary studies that look upon mukbang in an unfavorable manner [72,74,75]. Intervention studies shed insights on the impact of food-related videos (like mukbang) and nonfood-related videos on viewers' eating behavior, food choices, and food intake. They found that mukbang watching increased the consumption of unhealthy food and negatively affected the eating behaviors and physical health of viewers, particularly unsuccessful restrained dieters [17,31,48]. Over time, such unhealthy eating habits can injure the viewers' stomachs and make them overweight and obese, which seriously affects their health and leads to binge eating, bulimia nervosa, and anorexia nervosa [76].

Mukbang's Impact on Psychological Health

Numerous studies have examined mukbang's impact on mental health. Cross-sectional studies demonstrated that mukbang viewers often choose to watch videos when they feel lonely and desire companionship to satisfy their spiritual needs [58]. Based on the cross-sectional study, the researchers conducted qualitative research by analyzing comments and web-based posts of mukbang videos and conducting semistructured interviews. The findings not only supported that mukbang can release pressure and loneliness and provide enjoyable entertainment, but also revealed that mukbang can digitally enhance viewers' immersive experience and taste sensations, and help them experience a sense of vicarious satisfaction [30,59]. Researchers also conducted several interventional studies to explore mukbang's potential effect on the viewer's taste sensations and found that mukbang has the potential to improve viewers' perceptions of the taste of food and make the food taste delicious [22,26,60,61]. One of the interventional studies has investigated how different video types (including mukbang) affect the taste sensations of viewers eating plain white rice. The trial's results revealed that both mukbang and cooking videos can enhance viewers' taste sensation of spiciness and help them find the food is mouthwatering, satisfy their cravings, and attain psychological satiation [14]. Further, reviews and commentary studies both discussed how mukbang meets the psychological needs of its viewers. The integrated results showed that mukbang watching can provide digital companionship to viewers, and relieve their loneliness and daily life pressure, making them feel relaxed and happy [2,62,63,77-80]. In addition to relieving negative emotions, the distinct chomping sounds made by hosts can help relax viewers' brains and make them fall asleep easily [64,81], the sound made by ASMR mukbang hosts, the attractiveness of mukbang hosts and the presentation of food elicit visual and auditory contentment has been found can make the viewers feel involved in the warm dining atmosphere shown in the video, which further enhances the food's taste and viewers' appetites to some extent [62,72,81,82].

Public Perceptions of the Role of Mukbang in Health Promotion

By using the bullet chatting and commenting functions of short video platforms, people can interact with mukbang hosts and express their thoughts and opinions on the relationship between mukbang and health. Scholars have conducted interviews and text analyses to clarify viewers' attitudes toward mukbang [13,24,25,46]. An et al [83] used big data and text analysis methods to examine viewers' comments on popular mukbang videos on YouTube; results indicated that mukbang contents stimulating unhealthy eating behaviors led to negative reactions from viewers, whereas contents featuring healthy eating gained more viewers' favor and sympathy. The results of some studies on the comments for mukbang content reveal that solitary viewers generally consider mukbang a dining companion who creates a warm dining atmosphere and relieves their feelings of loneliness and social isolation [62]. Other viewers consider mukbang a tool that provides psychological gratification to help relieve immediate hunger and help them control their diet and lose weight [24]. However, regarding Big Eater Mukbang, some viewers felt that the excessive amount of food consumed at one time by mukbang hosts caused food wastage, was extremely harmful to the body, and promoted extremely unhealthy behaviors [13,24,46]. Moreover, they questioned whether the mukbang hosts ate all the food shown in the video since their slim figures suggested otherwise [46]. Along with being concerned about the hosts' health, the viewers warn themselves and appeal to others not to blindly follow the unhealthy eating behavior shown in the videos. Further, the results of qualitative interviews suggest that most viewers watch mukbang mainly for entertainment and to gain positive emotional value [65].

XSL•FO
RenderX

They were curious about the taste of the food presented in the videos, developed cravings for these specific foods after watching mukbang, and expressed their doubts regarding the very high food intake of mukbang hosts [46].

Future Applications of Mukbang

Scholars have started analyzing and discussing the advantages and disadvantages of these videos from different perspectives and indicating the future direction of mukbang research by analyzing the shortcomings of their own research in their studies. Some scholars opine that future research should consider various aspects of the mukbang-watching experience, particularly the mechanisms of mukbang-induced overeating and problematic eating [66]. Other scholars clarify that the different dietary statuses and watching motivations of mukbang viewers strongly affect health. James et al [14] and Yoo et al [55] found that extreme and failed dieters are more likely than other dieters to damage their physical health and lead to eating disorders after watching mukbang videos. They indicate that future research can further examine the factors and elements that can help promote healthy behaviors in mukbang videos, spread the concept of a healthy diet through these videos, and incorporate such concepts in medical and health care fields [14,55]. Scholars should focus on the creation of enjoyable and useful content that encourages people to continue watching mukbang and making dietary changes [67]. In addition, mukbang can be considered for future clinical use to address the nutritional problems caused by loss of appetite and picky eating [26].

Discussion

Overview

This scoping review addressed and resolved all its research questions, including examining the specific effects (RQ1) and viewers' opinions (RQ2) of mukbang watching on physical and psychological health, and examined the future applications of mukbang in health promotion (RQ3). In summary, this review addressed and resolved all its research questions. This scoping review synthesized the evidence of the impact of mukbang watching on viewers' dietary behavior, eating disorders (if any), and overall health. Mukbang is both beneficial and harmful to the viewer's health and eating behavior, which mainly depends on the viewer's own attitude toward food and weight loss, as well as their attention to their own body. Additionally, it identifies the viewers' attitudes toward and concerns about mukbang watching, which facilitates the regulation and improvement of mukbang video content and lays a foundation for the videos' application in health promotion.

Principal Findings

Many studies cited in this review agree that watching mukbang can positively affect viewers' dietary behaviors, appetite, food preferences, and physical and psychological health. This may be due to the fact that mukbang can elicit a communal dining experience that contributes to the viewer's relaxation during consumption; and the visual, auditory, and other sensory stimuli may stimulate the viewer's brain to synthesize hunger-inducing hormones, consequently reducing satiety and promoting appetite [15,64,68,81]. This result suggests the possibility of using mukbang positively to develop healthy eating habits and appetites by enhancing taste experiences and reducing negative emotions [15,30,81], which might be beneficial for people attempting to improve the taste of a bland diet without indulging in overeating or using unhealthy food additives [13]. The impact of mukbang watching on dietary behaviors and health promotion primarily may depend on the characteristics of mukbang viewers and the content of mukbang videos [46]. Several studies indicate that the main viewers of mukbang are young individuals, who often develop food cravings, dissatisfaction with their body shape, and an obsession with weight loss or dieting behaviors [68]. This may be due to the dietary and sociocultural characteristics of teenagers because they tend to prioritize taste over nutritional value when making food choices, thus they can be easily influenced by mukbang content [69]. For individuals suffering from anorexia nervosa, mukbang can be an effective physical therapy method to increase food intake by stimulating appetite [30]. In addition to the characteristics of mukbang viewers, the content of mukbang videos plays an important role in shaping eating habits, appetite levels, mood states, and overall health [46]. Hence, it is important to change the types of food shown in mukbang videos and gradually incorporate nutritional information to enhance the nutritional value of displayed food and the dietary behaviors of viewers. Additionally, the eating behaviors exhibited by hosts potentially influence viewer behavior. Extreme practices, such as binge eating, should be avoided or corrected to display clean and harmonious dining scenes so that mukbang videos positively influence the audience's eating behavior [59].

However, the impact of mukbang on its viewers is not uniformly positive. Some mukbang videos continue to propagate unhealthy concepts and behaviors, such as excessive consumption of high-calorie foods, misleading viewers to blindly imitate and lead to obesity and gastrointestinal disorders [28,42,72]; further, the overreliance on psychological gratification can lead to mukbang video addiction and endless dieting [29,53]. This means that individuals who achieve weight loss through dieting frequently may choose mukbang as a substitute for conventional eating. However, long-term compensation can engender binge eating, which may culminate in guilt and even depression. This pattern can initiate a cycle of weight loss attempts and mukbang indulgence, ultimately culminating in excessive dietary intake. Thus, mukbang is a double-edged sword that may cause both positive and negative influences on viewers' dietary habits and mental health. To elucidate this phenomenon, several scholars have presented their own perspectives in the research, a potential rationale for this may be posited as the ambivalence of young individuals toward the pursuit of a desirable physique and the indulgence in delicious food [59]. However, there is a lack of studies investigating the mechanisms underlying the influence of mukbang on eating behavior from an objective perspective, thus making it unclear whether watching mukbang has positive or negative effects on viewers. Based on the findings of this study, these 2 opposite effects of mukbang may be attributed to factors such as content variation, exposure levels, viewer's self-discipline, and dieting status. Future studies can further explore these influencing factors that contribute to the dual nature of mukbang, helping viewers watch mukbang in a

XSL•FO
RenderX

healthier way [48]. Thus, mukbang must be used cautiously in health promotion.

Implications for Future Research

According to the results of our study, the majority of studies on mukbang and its impact on health primarily focus on the perspective of mukbang anchors or clarify the deleterious effect of performing mukbang on anchors' health, and only a few studies examine the negative effects of mukbang on viewers' eating behavior and health by using questionnaires and interviews. Future research can further investigate the variations in psychological states among individuals of diverse age groups and genders when exposed to mukbang, as well as examine the impact of mukbang on their psychological well-being [68]. Besides, to date, no scholar has thoroughly investigated whether the negative effects of mukbang on health are real, or clarified which aspects of mukbang videos are responsible for these negative effects; existing studies also do not provide accurate conclusions on the relationship between mukbang and health, researchers have not yet conducted experimental studies from a physiological perspective to examine the objective effects of mukbang videos on viewers' eating habits, food preferences, and physical and mental health and clarify the mechanisms by which the videos affect viewers' health. Future studies can focus on exploring the effects of mukbang watching on the physical health of the viewers, or the relationship between disordered eating psychopathology and different mukbang viewing characteristics (eg, frequency of watching, average watch time, eating while watching), using laboratory methods or other rigorous experimental designs with the aid of objective instruments or tools [56]. As excessive problematic food intake is increasingly viewed as an addictive disorder, mukbang is thought to have a potential risk of inducing binge eating, the relationship between mukbang watching and addiction, as well as the influencing factors both worth further investigations to keep mukbang viewers away from addictive behaviors [29,30]. Meanwhile, excessive or problematic mukbang viewing may serve as an indicative signal of potential eating disorders, which can guide clinicians to identify and diagnose eating disorders more effectively. Therefore, mukbang content that may cause the development of eating disorders could be explored and establish criteria for evaluation, so as to help mukbang viewers stay away from those unhealthy mukbang and maintain healthy eating habits while enjoying the pleasant dining atmosphere presented by mukbang [56].

Future research can provide insights into the effect of different mukbang content on different taste perceptions, which can help clarify which mukbang content enhances a particular taste sensation among viewers [14,26]. Additionally, given the specific effect of mukbang on food preference, scholars can further test whether some particular types of mukbang might be useful for dieting and whether mukbang with fast food is the most harmful given these likely increase unhealthy food intake; more work is also needed to address the impact of the feature of web-based communication between anchors and viewers, the figure of the host and the food stimuli in mukbang [48]. In addition, according to Yoon [50], future research can focus on the development of mukbang that integrates nutrition and medicine and assess its impact and satisfaction levels. Simultaneously, studies have found that ASMR mukbang has the potential to reduce viewer's stress levels, and improve their sleep quality, future research may consider exploring the mechanisms involved or comparing the differences in the physical or mental health of viewers between ASMR videos and ASMR mukbang.

Strengths and Limitations

This is the first scoping review that systematizes studies on the relationship between mukbang and health and is, therefore, the first study to expand the research possibilities in this field. Studies relevant to the review were identified from both academic databases and gray literature websites in different languages. Since mukbang originated in Korea, Korean databases were specifically searched to obtain a more comprehensive perspective on mukbang-related studies, which might have contributed to the inclusion of studies with different contexts and origins. The term mukbang is not yet a descriptor indexed in databases; however, this was not an impediment to identifying the main studies related to the research objective. Our study comprehensively concluded the potential impact of mukbang on viewers' health and synthesized the possible causes of mukbang's negative effects on health by integrating and analyzing the included studies; however, we also found that there is a lack of rigorous investigations into the influencing factors of these negative impacts or potential relationships between positive and negative effects of mukbang on health. Meanwhile, this review identified methodological limitations in existing studies, suggesting that future studies could consider using clinical trials and other objective indicators to further explore the actual implications of mukbang on viewers. Furthermore, this review clarifies the role of mukbang in health promotion and, hopefully, eliminates some viewers' prejudice against mukbang videos and guides the country and media to supervise and manage unhealthy content in mukbang videos.

However, this scoping review has some limitations. First, the major limitation of this study is the uneven quality of the included studies. Considering the potential possibility of using mukbang as a future health intervention, existing high-quality studies on the health effects of mukbang are insufficient to support the conduct of systematic reviews or meta-analyses, thus, we tried to use scoping review to provide direction for future research. Second, in addition to the limitations of the number and quality of mukbang-related studies, this review included several studies focusing on food-related videos (or advertisements), which might have affected the accuracy and specificity of the review's conclusions. This is because a clear definition of differences between food-related and mukbang videos was not available in the early stages of research. Therefore, we redefined mukbang and clearly distinguished it from food-related videos (or advertisements) to ensure the relevance of our conclusions to the field of study. Third, the categories and associated keywords (eg, health, eating behavior, and eating disorder) might have limited the search results. Hence, we adopted a method to search the results and simultaneously performed manual secondary screening to capture more results pertaining to health promotion. However, we might have missed relevant materials that use other terms or do not explicitly use the searched keywords. Nevertheless,

XSL•FO
RenderX

we could present a representative overview of the existing studies on the subject in question.

Conclusions

This scoping review comprehensively searched databases and other sources for studies on mukbang and health and extracted and summarized the effects of mukbang on viewers' dietary behaviors, appetite, flavor perception, food preference, and health promotion, as well as their perceptions of and attitudes toward mukbang. The results of this study indicate that mukbang has both positive and negative implications for health. Factors such as the content of mukbang videos and the characteristics and physical states of viewers play crucial roles in influencing health outcomes and dietary behaviors. These findings provide valuable guidance to mukbang hosts and health care providers and promote their collaboration in developing healthy mukbang content and assisting viewers in adopting appropriate viewing habits. Finally, this scoping review underscores the potential benefits of using mukbang videos to promote healthy eating habits and emphasizes the need for future research, which will contribute to our understanding of these videos' impact on human health.

Acknowledgments

The authors would like to thank Shengjun Sun for the contributions he made in the selection of Korean studies. The authors would also like to thank Duoduo Yang for her great contributions in study screening and data extraction. The study was funded by Guangdong Provincial Philosophy and Social Science Planning Project (WX, GD22YSH06), Natural Science Foundation of Guangdong Province (WX, 2024A1515012147), Guangzhou Basic Research Program Basic and Applied Basic research Projects-General Projects (WX, 202201011311), Youth Talent Support Programme of Guangdong Provincial Association for Science and Technology (WX, SKXRC202314), Fundamental Research Funds for the Central Universities, Sun Yat-sen University (WX, 23qnpy111), and Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation (XW, pdjh2023b0020). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

All data generated or analyzed during this study are included in this published paper (and Multimedia Appendices 1-6).

Authors' Contributions

WX conceived the idea for the review, supervised the study, and provided methodological guidance.
TZ executed the draft search strategy and provided data for Multimedia Appendix 2.
DT established the inclusion and exclusion criteria and provided data for Textbox 1.
SN provided assistance in the search for Korean literature.
XW drafted and wrote the initial protocol and manuscript, summarized the main information and evaluated the quality evaluation of the included studies, and provided data for Multimedia Appendix 5 and Multimedia Appendix 6.
XW and YX completed the study screening and data extraction and provided data for Multimedia Appendix 4.
WX and XW critically revised this scoping review. WHCL conducted the data extraction chart and provided Multimedia Appendix 3.
PS conceptualized the study and provided feedback on it.
All authors reviewed and provided input on drafts, subsequently revised it for publication, and read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) fillable checklist.
[PDF File (Adobe PDF File), 139 KB - jmir_v27i1e56147_app1.pdf]

Multimedia Appendix 2
Proposed sample search strategies.
[DOCX File , 33 KB - jmir_v27i1e56147_app2.docx]

Multimedia Appendix 3
Data extraction chart.
[DOCX File , 22 KB - jmir_v27i1e56147_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Data extraction chart for all included studies.
[DOCX File , 405 KB - jmir_v27i1e56147_app4.docx]

Multimedia Appendix 5
Quality evaluation of part of the included articles.
[ZIP File (Zip Archive), 902 KB - jmir_v27i1e56147_app5.zip]

Multimedia Appendix 6
Main information of the included studies.
[DOCX File , 46 KB - jmir_v27i1e56147_app6.docx]

References

1. Styawan Z, Buwana DS. Watching attitude factors in delivering mukbang shows. J Humanit Soc Sci Bus 2023;2(2):442-454. [doi: 10.55047/jhssb.v2i2.540]
2. Donnar G. 'Food porn' or intimate sociality: committed celebrity and cultural performances of overeating in meokbang. Celebr Stud 2017;8(1):122-127. [doi: 10.1080/19392397.2016.1272857]
3. Jenging R, Mohamad F. Mukbang and me: implications on cognition and physical well-being among undergraduates. J Cogn Sci Hum Dev 2022;8(2):77-93 [FREE Full text] [doi: 10.33736/jcshd.4901.2022]
4. Wu Y. Research on Governance of Network Mukbang Phenomenon [dissertation]. Nanchang, China: Nanchang University; 2022.
5. Simth J. What is mukbang? And why is it so popular? fenced.ai. URL: http://fenced.ai/blogs/what-is-mukbang/ [accessed 2022-12-27]
6. Mukbang. Wikipedia. 2023. URL: https://en.wikipedia.org/w/index.php?title=Mukbang&oldid=1172046753 [accessed 2023-09-02]
7. Harris M. 'I don't like to eat alone': Inside the world of 'mukbangs,' extreme-eating videos that are making YouTubers rich. Insider. 2020. URL: https://www.insider.com/mukbang-video-the-extreme-eating-trend-making-youtube-rich-2020-2 [accessed 2023-09-24]
8. Cheng M. Research on Audience Use and Satisfaction in "Mukbang" Programs [dissertation]. Harbin, China: Heilongjiang University; 2020.
9. YouTube keyword tool. Ahrefs. 2023. URL: https://ahrefs.com/youtube-keyword-tool [accessed 2023-09-16]
10. Reddy NV, Mohabbat AB. Autonomous sensory meridian response: your patients already know, do you? Cleve Clin J Med 2020;87(12):751-754 [FREE Full text] [doi: 10.3949/ccjm.87a.20005] [Medline: 33229391]
11. Bruno AL, Chung S. Mŏkpang: pay me and I'll show you how much I can eat for your pleasure. J Jpn Korean Cine 2017;9(2):155-171. [doi: 10.1080/17564905.2017.1368150]
12. Song H. The making of microcelebrity: AfreecaTV and the younger generation in neoliberal South Korea. Soc Media Soc 2018;4(4):205630511881490. [doi: 10.1177/2056305118814906]
13. Strand M, Gustafsson SA. Mukbang and disordered eating: a netnographic analysis of online eating broadcasts. Cult Med Psychiatry 2020;44(4):586-609 [FREE Full text] [doi: 10.1007/s11013-020-09674-6] [Medline: 32277331]
14. James M, Ranasinghe N, Tang A, Oehlberg L. Watch your flavors: augmenting people's flavor perceptions and associated emotions based on videos watched while eating. 2022 Presented at: CHI EA '22: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems; April 29 2022-May 5, 2022; New Orleans, LA p. 429. [doi: 10.1145/3491101.3519846]
15. Han Y. Study on emotional and physiological changes according to food content types and scenes—ASMR mukbang vs nomal mukbang. J Image Cult Contents 2022(26):259-297. [doi: 10.24174/jicc.2022.06.26.259]
16. Poerio GL, Blakey E, Hostler TJ, Veltri T. More than a feeling: autonomous sensory meridian response (ASMR) is characterized by reliable changes in affect and physiology. PLoS One 2018;13(6):e0196645 [FREE Full text] [doi: 10.1371/journal.pone.0196645] [Medline: 29924796]
17. Bodenlos JS, Wormuth BM. Watching a food-related television show and caloric intake. A laboratory study. Appetite 2013;61(1):8-12. [doi: 10.1016/j.appet.2012.10.027] [Medline: 23154219]
18. Castelló-Martínez A, Tur-Viñes V. Obesity and food-related content aimed at children on YouTube. Clin Obes 2020;10(5):e12389. [doi: 10.1111/cob.12389] [Medline: 32726886]
19. Halford JC, Boyland EJ, Hughes GM, Stacey L, McKean S, Dovey TM. Beyond-brand effect of television food advertisements on food choice in children: the effects of weight status. Public Health Nutr 2008;11(9):897-904. [doi: 10.1017/S1368980007001231] [Medline: 18005487]
20. Kircaburun K, Stavropoulos V, Harris A, Calado F, Emirtekin E, Griffiths MD. Development and validation of the mkbang addiction scale. Int J Ment Health Addiction 2020;19(4):1031-1044. [doi: 10.1007/s11469-019-00210-1]

XSL•FO
RenderX

21. Chen R. Study on the use of eating webcast' and its impact on mood loneliness—taking XX university students as an example [dissertation]. Anhui University. 2019. URL: https://tinyurl.com/yv4vmf5y [accessed 2023-08-01]
22. Wang C, Peng Y, Qiu L, Wan X. Cloud-based commensality: enjoy the company of co-diners without social facilitation of eating. Front Psychol 2021;12:758966 [FREE Full text] [doi: 10.3389/fpsyg.2021.758966] [Medline: 34867658]
23. Frankel A. Mukbang is changing digital communications. Anthropology News. URL: https://www.anthropology-news.org/articles/mukbang-is-changing-digital-communications/ [accessed 2022-12-03]
24. Wang S. A Study on Potential Health Issues Behind the Popularity of "Mukbang" in China [dissertation]. Murfreesboro, Tennessee: Middle Tennessee State University; 2020.
25. Teng S, Khong KW, Pahlevan Sharif S, Ahmed A. YouTube video comments on healthy eating: descriptive and predictive analysis. JMIR Public Health Surveill 2020;6(4):e19618 [FREE Full text] [doi: 10.2196/19618] [Medline: 33001036]
26. James MN, Ranasinghe N, Tang A, Oehlberg L. Flavor-videos: Enhancing the flavor perception of food while eating with videos. : ACM; 2022 Presented at: IMX '22: Proceedings of the 2022 ACM International Conference on Interactive Media Experiences; June 22-24, 2022; Aveiro, Portugal p. 33-46. [doi: 10.1145/3505284.3529967]
27. Kang H, Yun S, Lee H. Dietary life and mukbang- and cookbang-watching status of university students majoring in food and nutrition before and after COVID-19 outbreak. J Nutr Health 2021;54(1):104. [doi: 10.4163/jnh.2021.54.1.104]
28. Yun S, Kang H, Lee H. Mukbang- and cookbang-watching status and dietary life of university students who are not food and nutrition majors. Nutr Res Pract 2020;14(3):276-285 [FREE Full text] [doi: 10.4162/nrp.2020.14.3.276] [Medline: 32528634]
29. Kircaburun K, Savcı M, Emirtekin E, Griffiths MD. Uses and gratifications of problematic mukbang watching—the role of eating and social gratification: a pilot study. J Psychiatr Res 2022;146:28-30. [doi: 10.1016/j.jpsychires.2021.12.036] [Medline: 34952300]
30. Kircaburun K, Balta S, Emirtekin E, Tosuntas, Demetrovics Z, Griffiths MD. Compensatory usage of the internet: the case of mukbang watching on youTube. Psychiatry Investig 2021;18(4):269-276 [FREE Full text] [doi: 10.30773/pi.2019.0340] [Medline: 33849247]
31. Alblas MC, Mollen S, Fransen ML, van den Putte B. See the cake and have it too? Investigating the effect of watching a TV cooking show on unhealthy food choices. Physiol Behav 2021;236:113409 [FREE Full text] [doi: 10.1016/j.physbeh.2021.113409] [Medline: 33789166]
32. Arrona-Cardoza P, Labonté K, Cisneros-Franco JM, Nielsen DE. The effects of food advertisements on food intake and neural activity: a systematic review and meta-analysis of recent experimental studies. Adv Nutr 2023;14(2):339-351 [FREE Full text] [doi: 10.1016/j.advnut.2022.12.003] [Medline: 36914293]
33. van der Laan LN, de Ridder DTD, Viergever MA, Smeets PAM. The first taste is always with the eyes: a meta-analysis on the neural correlates of processing visual food cues. Neuroimage 2011;55(1):296-303. [doi: 10.1016/j.neuroimage.2010.11.055] [Medline: 21111829]
34. Yeung AWK. Brain responses to watching food commercials compared with nonfood commercials: a meta-analysis on neuroimaging studies. Public Health Nutr 2021;24(8):2153-2160 [FREE Full text] [doi: 10.1017/S1368980020003122] [Medline: 32883385]
35. Yang Y, Wu Q, Morys F. Brain responses to high-calorie visual food cues in individuals with normal-weight or obesity: an activation likelihood estimation meta-analysis. Brain Sci 2021;11(12):1587 [FREE Full text] [doi: 10.3390/brainsci11121587] [Medline: 34942889]
36. Belfort-DeAguiar R, Seo D. Food cues and obesity: overpowering hormones and energy balance regulation. Curr Obes Rep 2018;7(2):122-129 [FREE Full text] [doi: 10.1007/s13679-018-0303-1] [Medline: 29619632]
37. Bruce AS, Pruitt SW, Ha O, Cherry JBC, Smith TR, Bruce JM, et al. The influence of televised food commercials on children's food choices: evidence from ventromedial prefrontal cortex activations. J Pediatr 2016;177:27-32.e1 [FREE Full text] [doi: 10.1016/j.jpeds.2016.06.067] [Medline: 27526621]
38. Gearhardt AN, Yokum S, Harris JL, Epstein LH, Lumeng JC. Neural response to fast food commercials in adolescents predicts intake. Am J Clin Nutr 2020;111(3):493-502 [FREE Full text] [doi: 10.1093/ajcn/nqz305] [Medline: 31940031]
39. Masterson TD, Stein WM, Beidler E, Bermudez M, English LK, Keller KL. Brain response to food brands correlates with increased intake from branded meals in children: an fMRI study. Brain Imaging Behav 2019;13(4):1035-1048 [FREE Full text] [doi: 10.1007/s11682-018-9919-8] [Medline: 29971684]
40. Boyland EJ, Burgon RH, Hardman CA. Reactivity to television food commercials in overweight and lean adults: physiological, cognitive and behavioural responses. Physiol Behav 2017;177:182-188. [doi: 10.1016/j.physbeh.2017.05.005] [Medline: 28478032]
41. Kidd C, Loxton NJ. Junk food advertising moderates the indirect effect of reward sensitivity and food consumption via the urge to eat. Physiol Behav 2018;188:276-282. [doi: 10.1016/j.physbeh.2018.02.030] [Medline: 29458116]
42. Sanskriti S, Guglani I, Joshi S, Anjankar A. The spectrum of motivations behind watching mukbang videos and its health effects on its viewers: a review. Cureus 2023;15(8):e44392 [FREE Full text] [doi: 10.7759/cureus.44392] [Medline: 37786568]

XSL•FO
RenderX

43. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Annals of Internal Medicine 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
44. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping Reviews (2020). In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: JBI; 2024.
45. Critical appraisal tools. JBI. URL: https://jbi.global/critical-appraisal-tools [accessed 2024-06-29]
46. Sultana SFS, Das P. Content analysis of mukbang videos: preferences, attitudes and concerns. J Posit Sch Psychol 2022;6(9):4811-4822.
47. Shin K. A Study on Food and Nutrition-Related Media Consumption and its Influence on Dietary Habits of Adolescents and Adults in Daegu and Gyeongbuk Regio [dissertation]. Daegu, Korea: Kyungpook National University; 2021.
48. Xu W. Does Watching Mukbangs Help You Diet? The Effect of the Mukbang on the Desire to Eat [dissertation]. Ithaca, NY: Cornell University; 2019.
49. Kang D. Some Elementary School Students in Ulsan Mukbang Viewing Status and Eating Behavior [dissertation]. Daegu, Korea: Kyungpook National University; 2022.
50. Yoon S. A Study on the Influence of Information Characteristics of TV Cooking Program on its Information Acceptance and Dietary Change [dissertation]. Seoul, South Korea: Sejong Cyber University; 2017.
51. Bang SY. A Study on the Eating Habits and Health Behavior of People in Their 20s and 30s According to the Use of Social Media Food Content [dissertation]. Gwangju, South Korea: Chonnam National University; 2022.
52. Di Y. A Study on the Motivation and Influence of Watching Network Mukbang Programs [dissertation]. Sichuan, China: Chengdu Sport University; 2022.
53. Kircaburun K, Yurdagül C, Kuss D, Emirtekin E, Griffiths MD. Problematic mukbang watching and its relationship to disordered eating and internet addiction: a pilot study among emerging adult mukbang watchers. Int J Ment Health Addiction 2020;19(6):2160-2169. [doi: 10.1007/s11469-020-00309-w]
54. Park S. A Study on the Use of YouTube Food Content and the Actual Consumption of Delivery Food by University Students in Gwangju [dissertation]. Kwangju, South Korea: Chosun University; 2022.
55. Yoo S, Shin G, Kim S. Does mukbang watching really affect obesity?: Focusing on the factors related to health and mukbang watching. Korean J Journal Commun Stud 2021;65(2):205-240. [doi: 10.20879/kjjcs.2021.65.2.006]
56. von Ash T, Huynh R, Deng C, White MA. Associations between mukbang viewing and disordered eating behaviors. Int J Eat Disord 2023;56(6):1188-1198. [doi: 10.1002/eat.23915] [Medline: 36807921]
57. Kang E, Lee J, Kim KH, Yun YH. The popularity of eating broadcast: content analysis of "mukbang" YouTube videos, media coverage, and the health impact of "mukbang" on public. Health Informatics J 2020;26(3):2237-2248 [FREE Full text] [doi: 10.1177/1460458220901360] [Medline: 31992111]
58. Li B. Correlation analysis of watching motivation of mukbang and the influence of mukbang on audience. PR Mag 2020(4):13-14. [doi: 10.16645/j.cnki.cn11-5281/c.2020.04.008]
59. Zhang L, Cui L. [What satisfaction can you get from watching others eat?—A qualitative study based on the audience of Bilibili]. J News Res 2020;11(5):45-49 [FREE Full text] [doi: 10.3969/j.issn.1674-8883.2020.05.025]
60. Kawai N, Guo Z, Nakata R. Watching a remote-video confederate eating facilitates perceived taste and consumption of food. Physiol Behav 2021;238:113469. [doi: 10.1016/j.physbeh.2021.113469] [Medline: 34058217]
61. Ngqangashe Y, Backer CJSD. The differential effects of viewing short-form online culinary videos of fruits and vegetables versus sweet snacks on adolescents' appetites. Appetite 2021;166:105436. [doi: 10.1016/j.appet.2021.105436] [Medline: 34119561]
62. Ma M, Yang J. [The rise of digital table: a review of foreign researches on mukbang]. News Res 2021(9):62-65. [doi: 10.3969/j.issn.1003-3629.2021.09.012]
63. Han C, Yin X. Psychological exploration of "Mukbang" audience from the perspective of structuralism. J Hebei Univ Econ Bus 2018;18(3):21-25. [doi: 10.14178/j.cnki.issn1673-1573.2018.03.004]
64. Zhou J. Research on the relieving effect of ASMR chewing sounds on anxiety in food video. Int J Educ Humanit 2023;8(3):89-91. [doi: 10.54097/ijeh.v8i3.8391]
65. Ngqangashe Y, Maldoy K, De BC, Vandebosch H. Exploring adolescents' motives for food media consumption using the theory of uses and gratifications. Communications 2022;47(1):73-92. [doi: 10.1515/commun-2019-0164]
66. Kim J, Choi S, Kim H, An S. Binge drinking and obesity-related eating: the moderating roles of the eating broadcast viewing experience among Korean adults. Int J Environ Res Public Health 2021;18(15):8066 [FREE Full text] [doi: 10.3390/ijerph18158066] [Medline: 34360359]
67. Jo S, Choi H. [A study on the cookbang YouTube program use and dietary change based on the technology acceptance model]. Culinary Sci Hosp Res 2021(135):163-171. [doi: 10.20878/cshr.2021.27.10.016]
68. Lee D. A Research on the Causal Factor to Binge Eating That Affects Dieting Deriving From One Person Media Viewership—Focused on Females Between the Ages of 20-30 [dissertation]. Seoul, South Korea: Chung-Ang University Graduate School of Arts; 2019.
69. Lee S, Lee SH. Actual status of mukbang viewing and food habits of university students in Wonju area. Korean J Community Living Sci 2022;33(4):631-643. [doi: 10.7856/kjcls.2022.33.4.631]

XSL•FO
RenderX

70. Alblas MC. Consuming Media, Consuming Food? Reactivity to Palatable Food Cues in Television Content. Amsterdam, Netherlands: Amsterdam School of Communication Research; 2021.
71. Kim SY. The Effect of Mukbang Show and Induced Anxiety on Eating Behavior [dissertation]. Gyeongbuk, Korea: Daegu University; 2020.
72. Li Y. [Get out of the emotional hunger—psychological and behavioral analysis of the popularity of mukbang]. Psychol Health 2020(12):10-11 [FREE Full text]
73. Pan M. An analysis of mukbang watching and watching behaviors of restricted eating groups: based on vicarious gratification theory. Radio TV J 2020(10):161-162. [doi: 10.19395/j.cnki.1674-246x.2020.10.083]
74. Wei J. [The reflect on the negative impact of "mukbang fever" on teenagers]. Youth Soc 2018;35:40-40 [FREE Full text] [doi: 10.17104/9783406720376-91]
75. Gao Z. The impacts and countermeasure of the novelty seeking live videos chaos on teenagers. Sci Technol Inf 2021;19(13):242-244. [doi: 10.16661/j.cnki.1672-3791.2101-5042-9490]
76. Nam NH. A Study on the Prevalence of Watching Mukbang and Factors Related Food Behaviors in Adults [dissertation]. Gwangju, South Korea: School of Education Chonnam National University; 2020.
77. Kircaburun K, Harris A, Calado F, Griffiths MD. The psychology of mukbang watching: a scoping review of the academic and non-academic literature. Int J Ment Health Addiction 2020;19(4):1190-1213. [doi: 10.1007/s11469-019-00211-0]
78. Pereira-Castro MR, Pinto AG, Caixeta TR, Monteiro RA, Bermúdez XPD, Mendonça AVM. Digital forms of commensality in the 21st century: a scoping review. Int J Environ Res Public Health 2022;19(24):16734 [FREE Full text] [doi: 10.3390/ijerph192416734] [Medline: 36554616]
79. Fangxiu L. [Research on audience psychology and profit model of mukbang from the perspective of communication]. News Dissem 2018(9):57-58 [FREE Full text]
80. Tian R. [Psychological research of mukbang audience from the perspective of use and satisfaction theory]. West Broadcast Telev 2020(6):71-72 [FREE Full text] [doi: 10.3969/j.issn.1006-5628.2020.06.034]
81. Zhong X. [The study of ASMR mukbang on network platforms]. Satellite TV IP Multimedia 2021;18:151-152 [FREE Full text] [doi: 10.12246/j.issn.1673-0348.2021.18.074]
82. Wu M. [The reason of "Mukbang Heat" phenomenon from the perspective of communication science]. West Broadcast Telev 2019(21):62-63. [doi: 10.3969/j.issn.1006-5628.2019.21.029]
83. An S, Lim Y, Lee H. A study of viewers' comments on online mukbang videos: a big-data analysis of perceptions toward eating behavior. Korean J Journal Commun Stud 2020;64(2):269-310. [doi: 10.20879/kjjcs.2020.64.2.007]

Abbreviations

ASMR: autonomous sensory meridian response
JBI: Joanna Briggs InstitutePRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RQ: research question
RCT: randomized controlled trial

Edited by T Leung; submitted 08.01.24; peer-reviewed by U Okpalauwaekwe, E Lyons; comments to author 04.05.24; revised version received 29.06.24; accepted 24.02.25; published 27.03.25.

Please cite as:
Wang X, Xiao Y, Nam S, Zhong T, Tang D, Li WHC, Song P, Xia W
Use of Mukbang in Health Promotion: Scoping Review
J Med Internet Res 2025;27:e56147
URL: https://www.jmir.org/2025/1/e56147
doi:10.2196/56147
PMID:

©Xiao Wang, Yuxue Xiao, Sujin Nam, Ting Zhong, Dongyan Tang, William Ho Cheung Li, Peige Song, Wei Xia. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

How TikTok Influencers Disclose Food and Beverage Brand Partnerships: Descriptive Study

Roxanne Dupuis[1,2], PhD; Aviva A Musicus[3,4], SCD; Brittany Edghill[2,5], MPA; Emma Keteku[6], MD; Marie A Bragg[1,2,7], PhD

[1]Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States
[2]NYU Food Environment and Policy Research Coalition, New York, NY, United States
[3]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[4]Center for Science in the Public Interest, Washington, DC, United States
[5]Office of Science and Research, NYU Grossman School of Medicine, New York, NY, United States
[6]New York Presbyterian Hospital-Weill Cornell, New York, NY, United States
[7]Marketing Department, Stern School of Business, New York University, New York, NY, United States

Corresponding Author:
Roxanne Dupuis, PhD
Department of Population Health
NYU Grossman School of Medicine
180 Madison Avenue
New York, NY, 10016
United States
Phone: 1 646 470 4505
Email: roxanne.dupuis@nyulangone.org

Abstract

Background: Food and beverage marketing is an important influence on the health and diets of adolescents. Food and beverage companies spend billions of dollars annually on advertisements to promote their products and are increasingly focusing on social media influencers. Influencer product endorsements blur the line between entertainment and marketing.

Objective: This study aimed to quantify how often TikTok influencers promote products from food and beverage brands and document the range of ways they disclose brand relationships in their content.

Methods: We collected up to 100 videos posted on or before July 1, 2022, from each of the top 100 influencers on TikTok in the United States and recorded information about the influencer (eg, number of followers) and video (eg, number of views and likes). For each video that contained food or beverage products, we identified the main product featured. A team of research assistants then coded each video for how the product was featured (ie, in the video, audio, or caption) and, for branded products, whether the video was accompanied by any disclosures of brand relationships. Average pairwise percentage agreement among coders was 92%, and average pairwise Cohen κ was 0.82.

Results: Among the 8871 videos from 97 influencers that made up the final analytical sample, we identified 1360 videos (15.3%) that featured at least one food or beverage product. These 1360 videos were viewed >9 million times and received >1 million likes each. Nearly half (n=648, 47.6%) of the videos featured a branded product. Most videos featuring a branded product did not contain a brand relationship disclosure (n=449, 69.3%). Among videos that disclosed a brand relationship, influencers used 10 different types of disclosures. Tagging a brand in the video's caption was the most common disclosure method (n=182, 28.1%). Six types of caption hashtags were used to disclose brand relationships, including #[brandname] (n=63, 9.7%) and #ad (n=30, 4.6%). Only 1 video (0.2%) made use of TikTok's official disclosure label and only 1 video (0.2%) verbally mentioned a contractual agreement with a brand.

Conclusions: Among the food and beverage videos with disclosures we identified, the most frequently used mechanism—tagging the brand—did not clearly differentiate between sponsored content and the influencer trying to attract a brand or followers who may like that brand. Social media users, particularly adolescents, need clearer, more robust disclosures from influencers to protect against the undue influence of food marketing. These findings may also inform calls for the Children's Food and Beverage Advertising Initiative—the largest self-regulatory pledge to reduce unhealthy food marketing—to include older adolescents, who are heavily targeted by food and beverage companies on social media.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e60891) doi:10.2196/60891

KEYWORDS

social media; social media marketing; social media influencer; food and beverage marketing; adolescent health

Introduction

Food and beverage companies spend billions of dollars annually on advertisements to promote their products [1]. Food marketing (including beverages) is an important influence on the health and diets of children and adolescents. Advertisements affect brand preferences and eating behaviors [2]. Experimental evidence shows that children and adolescents consume significantly more calories when viewing television advertisements featuring unhealthy foods compared to nonfood products [3-5]. In 2006, the Institute of Medicine recognized food marketing toward children as "out of balance" with healthy eating, putting children and adolescents' health at risk [1].

Most food advertisement studies have focused on television [1,3-6], but adolescents are shifting toward digital entertainment and social media [7]. Half (51%) of adolescents in the US spend an average of 4.8 hours daily on social media [8]. One popular social media platform among adolescents is the short video–sharing app TikTok. The social media company's internal documents indicate that, in 2020, one-third of their 49 million US daily users were younger than 14 years [9]. Since then, the app's popularity has grown exponentially, with 150 million users in the United States, excluding children under the age of 13 years [10]. TikTok's growing popularity makes it poised to overtake YouTube as the most-used app among adolescents [11], but to date there is little research on the effect of food advertisements on this platform [12-14].

Following this shift in media consumption by adolescents, food and beverage companies are increasingly focusing on social media marketing [15]. A study in Australia found that adolescents were exposed to roughly 17 food advertisements for every hour they spent on the internet or on social media and video-sharing apps on their phones [16]. One food company advertising strategy is to pay "influencers," who are "regular, everyday people" who become online celebrities with large followings, with posts that generate billions of views, to promote products either discreetly or overtly on their social media accounts [14,17]. Adolescents are especially susceptible to social media advertisements from influencers [18] because they are likely to evade adolescents' media literacy skills. The processing of commercialized media content model [19] purports that when children recognize what they are seeing as a commercial, they are able to use their defenses (eg, skepticism, recognition of persuasive intent) to reduce an advertisement's influence. But influencers blur the boundary between entertainment and marketing, which may reduce adolescents' ability to recognize and defend against the persuasive intent associated with their messaging. Because social media influencers are seen as regular, everyday people, their product endorsements may be experienced as credible recommendations rather than commercial advertisements [20].

The Federal Trade Commission (FTC), the government agency that regulates marketing on social media platforms in the United States, requires influencers to disclose sponsored posts, but these disclosures may not operate as intended. The FTC permits very subtle disclosures, such as adding the words "#ad" or "#sponsored" at the bottom of posts [21], so viewers may not even notice them among the other text surrounding the image.

Little is known about how TikTok influencers disclose food brand partnerships. Our goal in this study was to quantify how often TikTok influencers promote products from food brands and document the range of ways they disclose brand relationships in their content.

Methods

Data Collection

We collected up to 100 videos posted on or before July 1, 2022, from each of the top 100 influencers on TikTok in the United States. Each video was collected by capturing the full video and captions on a research assistant's phone. Research assistants noted the date the video was posted and the numbers of views, likes, comments, and "saves" or bookmarks. They also collected profile information, such as whether the influencer was "verified" with a blue checkmark and the total number of likes and followers.

Ethical Considerations

This study was determined to be non–human subjects research and used preexisting data that included publicly available information and therefore does not require institutional review board review per Federal Regulations for the Protection of Human Research Subjects (45 CFR 46.104(d)(4)) [22]. TikTok allows users to select whether their accounts are public or private. All collected data (ie, videos) were publicly available to anyone on TikTok. No user (ie, follower) data were collected. Only deidentified and aggregated data are presented.

Content Analysis

Using a 3-step approach, we first identified whether a food product was featured. Second, we identified the main product, based on whether it was (1) the only featured food product; (2) a sponsored product; or (3) the most prominent or dominant product (eg, the first to appear or to grab the viewer's eye). For each main product, the research team then noted whether it was branded (eg, Coke bottle vs banana). Third, a team of research assistants viewed each food product video to capture information about how the product was featured (ie, in the video, caption, or audio) and, for any branded product, where the brand or logo was featured and whether the video was accompanied by any brand disclosures.

Statistical Analysis

Using ReCal3 [23], we calculated interrater reliability among coders in the third step using a small sample of "practice" videos

XSL•FO
RenderX

(n=6) that reflected the range of disclosures used [24]. Average pairwise percentage agreement was 92%, and average pairwise Cohen κ was 0.82. We deemed these to be acceptable given the descriptive and exploratory nature of this work. We report descriptive statistics.

Results

From the initial potential pool of 10,000 videos across 100 influencers, we had a final analytical sample of 8871 videos from 97 influencers (Multimedia Appendix 1, Figure S1). We identified 1360 videos (15.3%) that featured at least one food product. These 1360 videos originated from 89 unique influencers (median 9, IQR 4-21 videos per influencer), each with >17 million followers and >526 million profile likes. The videos themselves were viewed >9 million times and received >1 million likes (Table 1).

Table 1. Follower engagement with videos featuring food or beverage products, overall and for branded products only.

Engagement	Featuring all food or beverage products (n=1360), mean (SD)	Featuring branded food or beverage products (n=648), mean (SD)
Views	9,249,401 (14,963,961)	8,924,298 (15,584,292)
Likes	1,062,903 (1,777,784)	1,099,153 (1,826,587)
Comments	7524 (14,437)	7189 (13,530)
Bookmarks or "saves"	30,144 (59,915)	33,542 (67,666)

Nearly half (n=648, 47.6%) of the videos with food products featured a branded product. Product brands or logos were most often featured in the video itself (n=606, 93.5%), followed by the video caption (n=224, 34.6%), and/or the accompanying audio (n=126, 19.4%).

Most videos featuring a branded food or beverage product did not contain a brand relationship disclosure (n=449, 69.3%). Among videos that disclosed a brand relationship, influencers used 10 different types of disclosures (Table 2). Tagging a brand in the video's captions was the most common disclosure method (n=182, 28.1%). Six different types of caption hashtags were used to disclose brand relationships, including #[brandname] (n=63, 9.7%) and #ad (n=30, 4.6%).

Table 2. The 10 different types of brand relationship disclosures used in videos featuring branded food and beverage products[a] (n=648).

Types of disclosure	Disclosures, n (%)
No disclosure	449 (69.3)
Caption hashtags (#)	
#[brandname]	63 (9.7)
#ad	30 (4.6)
#[brand]partner	22 (3.4)
#[brand]ambassador	3 (0.5)
#sponsored or #sponsorship	1 (0.2)
Other hashtag (eg, brand slogan)	39 (6.0)
Brand tags (@)	
Brand tagged in video	4 (0.6)
Brand tagged in caption	182 (28.1)
Official TikTok disclosure label	1 (0.2)
Contractual agreement verbally disclosed	1 (0.2)

[a]Not all disclosures represent actual brand relationships; disclosures are not mutually exclusive.

Discussion

Principal Results

Nearly half of the TikTok posts by top influencers that included food and beverage products featured a branded product. The majority (69.3%) of these posts, however, did not include any brand relationship disclosures, suggesting either that no relationship existed or that they purposely or accidentally omitted its disclosure. By comparison, a recent study of food-related Instagram posts by celebrities (athletes, actors, and musicians) found that less than 5% of the identified posts that were sponsored by a food company were identified by hashtags (eg, "#ad") or related terms (eg, "paid advertisement") below the post [25].

Among the food videos with disclosures we identified, the most frequently used mechanism—tagging the brand—did not clearly

XSL•FO
RenderX

differentiate between sponsored content and the influencer trying to attract a brand or followers who may like that brand. These findings have implications for how social media marketing should be regulated. The ambiguity around brand relationship disclosures makes it difficult for social media users, particularly adolescents, to make informed decisions about the products they see [18]. This concern has been recognized by the FTC in the most recent iteration (July 2023) of its endorsement guides [26], which emphasize that marketing practices geared toward adults may not be appropriate for children. In fact, bigger and clearer disclosures would likely work better among children and adolescents. A qualitative study among participants aged 12-16 years reported that they did not mind if influencers were paid, but they did not like seeing explicit disclosures that were "...too pushy, like, this is advertising, because then you will keep this in mind all the time" [27].

Limitations

Given the dearth of information on influencer marketing, our study set out to systematically collect and record information on how social media influencers disclose relationships with food brands. However, our study has limitations, including (1) the use of a convenience sample of influencers and posts, (2) the focus on only the most prominent food product in each video, (3) the inability to infer intentionality in whether a product was meant to be featured in a video (eg, when the only food product featured was in the background), and (4) the inability to distinguish whether the presence or lack of a disclosure implied the presence or lack of a brand relationship.

Conclusions

Social media users, particularly adolescents, need clearer, more robust disclosures from influencers to protect against the undue influence of food marketing. These findings may inform calls for the Children's Food and Beverage Advertising Initiative—the largest self-regulatory pledge designed to reduce unhealthy food marketing—to extend its protections to older adolescents, who are heavily targeted by food companies [28,29].

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA248441. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the NYU Food Environment and Policy Research Coalition research assistants who worked on this project.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

RD: conceptualization, data curation, formal analysis, investigation, methodology, supervision, visualization, writing (original draft)
AAM: conceptualization, methodology, writing (review and editing)
BE: project administration, supervision, writing (review and editing)
EK: conceptualization, project administration, resources, writing (review and editing)
MAB: conceptualization, funding acquisition, methodology, supervision, validation, writing (review and editing)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection and cleaning flow chart.
[PDF File (Adobe PDF File), 10 KB - jmir_v27i1e60891_app1.pdf]

References

1. Institute of Medicine. Food Marketing to Children and Youth: Threat or Opportunity?. Washington, DC: National Academies Press; 2006.
2. Boyland E, McGale L, Maden M, Hounsome J, Boland A, Angus K, et al. Association of food and nonalcoholic beverage marketing with children and adolescents' eating behaviors and health: a systematic review and meta-analysis. JAMA Pediatr 2022 Jul 01;176(7):e221037 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.1037] [Medline: 35499839]
3. Halford JC, Boyland EJ, Hughes GM, Stacey L, McKean S, Dovey TM. Beyond-brand effect of television food advertisements on food choice in children: the effects of weight status. Public Health Nutr 2008 Sep;11(9):897-904. [doi: 10.1017/S1368980007001231] [Medline: 18005487]
4. Halford JC, Gillespie J, Brown V, Pontin EE, Dovey TM. Effect of television advertisements for foods on food consumption in children. Appetite 2004 Apr;42(2):221-225. [doi: 10.1016/j.appet.2003.11.006] [Medline: 15010186]

XSL•FO
RenderX

5. Harris JL, Bargh JA, Brownell KD. Priming effects of television food advertising on eating behavior. Health Psychol 2009 Jul;28(4):404-413 [FREE Full text] [doi: 10.1037/a0014399] [Medline: 19594263]
6. Cairns G, Angus K, Hastings G, Caraher M. Systematic reviews of the evidence on the nature, extent and effects of food marketing to children. A retrospective summary. Appetite 2013 Mar;62:209-215 [FREE Full text] [doi: 10.1016/j.appet.2012.04.017] [Medline: 22561190]
7. Rideout V. The common sense census: media use by tweens and teens. 2015. URL: https://www.commonsensemedia.org/sites/default/files/research/report/census_researchreport.pdf [accessed 2025-02-14]
8. Rothwell J. Teens spend average of 4.8 hours on social media per day. Gallup. URL: https://news.gallup.com/poll/512576/teens-spend-average-hours-social-media-per-day.aspx [accessed 2024-10-18]
9. Zhong R, Frenkel S. A third of TikTok's U.S. users may be 14 or under, raising safety questions. New York Times. 2020 Aug 14. URL: https://www.nytimes.com/2020/08/14/technology/tiktok-underage-users-ftc.html [accessed 2025-02-14]
10. Celebrating our thriving community of 150 million Americans. TikTok. URL: https://newsroom.tiktok.com/en-us/150-m-us-users [accessed 2025-02-14]
11. Newton C. TikTok is having a monster 2020. The Verge. 2020 Jun 10. URL: https://www.theverge.com/interface/2020/6/10/21285309/tiktok-2020-user-numbers-revenue-smash-hit-mea-culpa [accessed 2022-01-21]
12. Brooks R, Christidis R, Carah N, Kelly B, Martino F, Backholer K. Turning users into 'unofficial brand ambassadors': marketing of unhealthy food and non-alcoholic beverages on TikTok. BMJ Glob Health 2022 Jun;7(6):e009112 [FREE Full text] [doi: 10.1136/bmjgh-2022-009112] [Medline: 35879104]
13. Potvin Kent M, Bagnato M, Amson A, Remedios L, Pritchard M, Sabir S, et al. #junkfluenced: the marketing of unhealthy food and beverages by social media influencers popular with Canadian children on YouTube, Instagram and TikTok. Int J Behav Nutr Phys Act 2024 Apr 11;21(1):37 [FREE Full text] [doi: 10.1186/s12966-024-01589-4] [Medline: 38605322]
14. De Veirman M, Hudders L, Nelson MR. What is influencer marketing and how does it target children? A review and direction for future research. Front Psychol 2019;10:2685 [FREE Full text] [doi: 10.3389/fpsyg.2019.02685] [Medline: 31849783]
15. Raffoul A, Ward ZJ, Santoso M, Kavanaugh JR, Austin SB. Social media platforms generate billions of dollars in revenue from U.S. youth: findings from a simulated revenue model. PLoS One 2023;18(12):e0295337 [FREE Full text] [doi: 10.1371/journal.pone.0295337] [Medline: 38150418]
16. Kelly B, Bosward R, Freeman B. Australian children's exposure to, and engagement with, web-based marketing of food and drink brands: cross-sectional observational study. J Med Internet Res 2021 Jul 12;23(7):e28144 [FREE Full text] [doi: 10.2196/28144] [Medline: 34255675]
17. Arnold A. Why YouTube stars influence millennials more than traditional celebrities. Forbes. 2017 Jun 20. URL: https://www.forbes.com/sites/under30network/2017/06/20/why-youtube-stars-influence-millennials-more-than-traditional-celebrities/ [accessed 2024-12-18]
18. Radesky J, Chassiakos YR, Ameenuddin N, Navsaria D, Council on Communication and Media. Digital advertising to children. Pediatrics 2020 Jul;146(1):e20201681. [doi: 10.1542/peds.2020-1681] [Medline: 32571990]
19. Buijzen M, Van RE, Owen L. Introducing the PCMC model: an investigative framework for young people's processing of commercialized media content. Commun Theory 2010;20(4):427-450. [doi: 10.1111/j.1468-2885.2010.01370.x]
20. Schouten AP, Janssen L, Verspaget M. Celebrity vs. influencer endorsements in advertising: the role of identification, credibility, and product-endorser fit. Int J Advert 2019 Jul 02;39(2):258-281. [doi: 10.1080/02650487.2019.1634898]
21. Disclosures 101 for social media influencers. Federal Trade Commission. 2019 Nov. URL: https://www.degruyter.com/document/doi/10.7312/schw92626-015/html [accessed 2025-02-14]
22. Code of Federal Regulations: 46.104 Exempt Research. U.S. Department of Health and Human Services. 2018 Jun 19. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-d/index.html#46.404
23. Freelon D. ReCal: intercoder reliability calculation as a web service. Int J Internet Sci 2010;5(1):20-33. [doi: 10.4135/9781526421036905572]
24. Lombard M, Snyder-Duch J, Bracken CC. Content analysis in mass communication: assessment and reporting of intercoder reliability. Human Comm Res 2002 Oct;28(4):587-604. [doi: 10.1111/j.1468-2958.2002.tb00826.x]
25. Turnwald BP, Anderson KG, Markus HR, Crum AJ. Nutritional analysis of foods and beverages posted in social media accounts of highly followed celebrities. JAMA Netw Open 2022 Jan 04;5(1):e2143087 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.43087] [Medline: 35019982]
26. 16 CFR Part 255 -- guides concerning use of endorsements and testimonials in advertising. Federal Trade Commission. 2023. URL: https://www.ecfr.gov/current/title-16/part-255 [accessed 2024-02-27]
27. Van Dam S, Van Reijmersdal E. Insights in adolescents' advertising literacy, perceptions and responses regarding sponsored influencer videos and disclosures. Cyberpsychology 2019 May 23;13(2):2. [doi: 10.5817/cp2019-2-2]
28. Harris J, Heard A, Schwartz M. Older but still vulnerable: all children need protection from unhealthy food marketing. Yale Rudd Center. URL: https://media.ruddcenter.uconn.edu/wp-content/uploads/sites/2909/2024/06/Protecting_Older_Children_3_14.pdf [accessed 2025-02-14]
29. Recommendations for responsible food marketing to children. Healthy Eating Research. 2015. URL: https://healthyeatingresearch.org/research/recommendations-for-responsible-food-marketing-to-children/ [accessed 2025-02-14]

XSL•FO
RenderX

Abbreviations

FTC: Federal Trade Commission

Edited by A Mavragani; submitted 31.05.24; peer-reviewed by A Cardoso, L Hagen, CH Chan; comments to author 03.12.24; revised version received 21.01.25; accepted 06.02.25; published 28.02.25.

Please cite as:
Dupuis R, Musicus AA, Edghill B, Keteku E, Bragg MA
How TikTok Influencers Disclose Food and Beverage Brand Partnerships: Descriptive Study
J Med Internet Res 2025;27:e60891
URL: https://www.jmir.org/2025/1/e60891
doi:10.2196/60891
PMID:40053812

©Roxanne Dupuis, Aviva A Musicus, Brittany Edghill, Emma Keteku, Marie A Bragg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Unpacking Performance Factors of Innovation Systems and Studying Germany's Attempt to Foster the Role of the Patient Through a Market Access Pathway for Digital Health Applications (DiGAs): Exploratory Mixed Methods Study

Sara Gehder[1], MSc; Moritz Goeldner[1], Prof Dr

Working Group for Data-Driven Innovation, Hamburg University of Technology, Hamburg, Germany

Corresponding Author:
Moritz Goeldner, Prof Dr
Working Group for Data-Driven Innovation
Hamburg University of Technology
Am Schwarzenberg-Campus 4
Hamburg, 21073
Germany
Phone: 49 40 428784777
Email: moritz.goeldner@tuhh.de

Abstract

Background: Health care innovation faces significant challenges, including system inertia and diverse stakeholders, making regulated market access pathways essential for facilitating the adoption of new technologies. The German Digital Healthcare Act, introduced in 2019, offers a model by enabling digital health applications (DiGAs) to be reimbursed by statutory health insurance, improving market access and patient empowerment. However, the factors influencing the success of these pathways in driving innovation remain unclear.

Objective: This study aims to identify the key performance factors of the innovation system shaped by the *patient-relevant structural and procedural improvement* (pSVV) pathway within the DiGA model. By examining how this pathway supports the entry of innovative digital health technologies, we seek to uncover the systemic dynamics that influence its effectiveness in fostering patient-centered digital health solutions.

Methods: This study, conducted from May 2023 to November 2024, used a mixed methods approach. A descriptive analysis assessed how DiGA manufacturers use positive health care effects, giving a market overview of the pSVV technology. A qualitative analysis using grounded theory and Gioia methodology provided insights into stakeholder perspectives, focusing on manufacturers and regulatory bodies. A functional-structural analysis examined how components of the innovation system, such as actors, institutions, interactions, and infrastructure, interact and impact the effectiveness of the pathway.

Results: The descriptive analysis showed that only 11 (20%) of the 56 DiGAs available in Germany used the pSVV pathway, with only 1 (2%) provisionally listed DiGA using pSVV as a primary end point; 6 of 9 (67%) pSVV key areas were used. The qualitative analysis revealed that manufacturers prioritize demonstrating medical benefits over pSVV due to evidence requirements and uncertainties around pSVV acceptance. Operational barriers hindered the adoption of pSVV, despite a positive reception among stakeholders. The systemic analysis identified key issues, including a lack of entrepreneurial focus on pSVV, limited regulatory experience, inadequate measurement methods, and entrenched practices prioritizing medical benefits, that hinder market formation and legitimacy.

Conclusions: This study identifies key factors for effectively implementing innovation systems through regulated market access pathways, including content and format security, clearer framework specification, active innovation process management, and market formation stimulation. Addressing these factors can reduce uncertainties and promote wider adoption of digital health technologies. The findings highlight the need for future research on patient empowerment and the development of methodologies beyond traditional therapeutic outcomes.

(J Med Internet Res 2025;27:e66356) doi:10.2196/66356

XSL•FO
RenderX

KEYWORDS

regulatory market access pathways; digital health application; DiGA; patient-relevant structural and procedural improvement; pSVV pathway analysis; qualitative and systemic analysis; policy and stakeholder insights; innovation system analysis

Introduction

Background

Innovation in health care is well known for being challenging. The complexity is driven by factors such as the involvement of diverse stakeholders in highly regulated markets [1,2], the need to address "wicked problems," complex issues without straightforward solutions [3], and substantial "system inertia," meaning resistance to change, which is especially pronounced in health care systems [4]. Policy approaches have been shown to act as inertial forces that can significantly shape the development of innovative technologies in health care [5], particularly digital health technologies, where patient access is greatly limited without broad and scalable market access pathways [6]. Mobile health apps are becoming increasingly available to consumers worldwide [7-9]; however, significant barriers, including regulatory complexity, insufficient evidence of clinical benefit, and data privacy concerns, still prevent their full integration into health care systems, leaving them underused in clinical practice [10-14].

To address these challenges, health policy approaches are being adopted globally to standardize regulatory processes and speed up market access, particularly in the rapidly evolving field of digital health [15-18]. Notable examples of health policy approaches include regulated market access pathways, such as *the fast-track process for digital health applications* pathway in Germany [19] and the *Prise en Charge Anticipée Numérique* (PECAN) pathway in France [20], which aim to facilitate the entry of digital health applications (DiGAs; German: Digitale Gesundheitsanwendung) into their respective markets. These new approaches are gaining significant international attention as pioneering models in the field [6,21].

To illustrate, the German Digital Healthcare Act (DVG), introduced in 2019, established the DiGA pathway to foster the integration of so-called DiGAs into the health care system. This pathway allows DiGAs to be reimbursed by statutory health insurance when prescribed by health care professionals or provided by health insurers. With approximately 90% of the German population covered by statutory health insurance [22], these "apps on prescription" are widely accessible at no cost to patients. Eligible DiGAs are included in the national DiGA Directory [23], which is managed by the Federal Institute for Drugs and Medical Devices (BfArM; German: Bundesinstitut für Arzneimittel und Medizinprodukte) to ensure transparency and oversee the approval process [24,25]. To qualify, a DiGA must be Conformité Européenne (CE)–marked (classes I and IIa or IIb), be based primarily on digital technology intended for direct patient use, and support the detection, monitoring, treatment, or alleviation of diseases, injuries, or disabilities [26]. Following a recent regulatory update in April 2024, digital medical devices classified as class IIb have been added to be eligible for recognition as a DiGA. However, this inclusion comes with specific stipulations; these devices are required to demonstrate medical benefit only, and they are eligible solely for permanent listing, excluding the option for preliminary listing. In addition, a DiGA must meet stringent requirements for safety, quality, functionality, privacy, and data security as defined in the respective DiGA ordinance (DiGAV; German: Digitiale Gesundheitsanwendungen-Verordnung) [24]. To enhance transparency around the assessment criteria and specific requirements established by BfArM, a DiGA guide [27] for manufacturers, service providers, and users was published. This guide provides a concise overview of the applicable ordinances and regulations, detailing how BfArM interprets the normative requirements set forth in the DVG and DiGAV.

Despite some emerging research examining the market dynamics following the implementation of regulated market access pathways, such as the number of products entering the market, time to market, pricing, and study designs used to demonstrate efficacy [28,29], as well as factors contributing to the success of specific DiGA products [30], broader factors influencing successful technology integration remain underexplored. This study addresses these gaps to better inform future policy and innovation strategies. To identify potential success factors, it is crucial to consider not only the regulated market access pathway itself but also the surrounding elements that influence its effectiveness. The broader network of actors, institutions, and resources involved in the creation, diffusion, and use of new technologies is defined as the technology innovation system [31]. A health policy approach, such as the creation of a regulated market access pathway that "brings about processes of change and stimulates technological innovation," can be characterized as a systemic instrument [32]. Systemic instruments are mechanisms designed to shape the conditions under which technological innovation systems operate. Understanding the dynamics that a systemic instrument brings to a technological innovation system requires the consideration of various functional and structural components, all of which must be in place to effectively implement the targeted technological innovation [33]. These components, as outlined by Wieczorek and Hekkert [34], include the structural components, such as actors, institutions, interactions, and infrastructure, as well as the functional components, such as entrepreneurial activities (F1), knowledge development (F2), knowledge diffusion (F3), guidance of the search (F4), market formation (F5), mobilization of resources (F6), and creation of legitimacy (F7) [35].

This exploratory study aimed to identify factors contributing to the success of an innovation system shaped by a regulated market access pathway in the digital health sector. Through an exemplary case study, we sought insights to support ongoing improvements in digital health policy, focusing on optimizing market access pathways for digital health technologies. The central research question is as follows: What are the performance factors of an innovation system created by a regulated market access pathway that supports the market entry of innovative digital health technologies?

XSL•FO
RenderX

To address this question, we selected a specific case from the 2019 German policy innovation DVG, particularly the "patient-relevant structural and procedural improvements" (pSVV; German: Patientenrelevante Struktur- und Verfahrensverbesserung) pathway. The pSVV pathway is a targeted segment within the DiGA pathway, specifically modifying the element of product value assessment, 1 of the 4 "full-stack" elements (regulatory authorization, product value assessment, pricing and reimbursement, and patient access) that define a comprehensive market access pathway [6], promoting innovations aimed to specifically elevate DiGAs that enhance patient–health care provider interactions and support patient health behaviors.

Unlike traditional pharmaceuticals, which require evidence of "therapeutic efficacy" through clinical studies for coverage eligibility [36], DiGAs are allowed to demonstrate their effectiveness through "positive health care effects" defined as either "medical benefit" or pSVV. While the medical benefit criterion focuses on outcomes, such as improved health status, reduced disease duration, prolonged survival, or improved quality of life, the pSVV criterion introduces 9 additional areas to demonstrate the value that is equally eligible as defined in the DiGA guide. These areas include improving care coordination, adherence to guidelines and standards, enhancing patient adherence, facilitating access to care, increasing patient safety, and promoting health literacy and autonomy. This expansion of the traditional product value assessment beyond the traditional therapeutic benefits is further complemented in the DiGA guide by explicitly broadening the acceptable methods of evidence beyond the traditional randomized controlled trials. We selected this case due to the maturity of the evolving DiGA ecosystem [28] in Germany, which offers greater transparency and availability of data on use and uptake compared to other countries. Within this context, the pSVV pathway was chosen as a focal point because it represents a well-defined component within the DiGA pathway, involving a manageable number of stakeholders while offering sufficient accumulated experience to provide valuable insights into the factors influencing digital health innovation.

In addition, the introduction of the pSVV pathway is groundbreaking as it extends beyond traditional treatment parameters, embracing a holistic approach to health care that prioritizes patient empowerment and streamlined care processes [19,37,38]. This aligns with the global trend toward patient empowerment, emphasizing the importance of patient agency and involvement in health care processes [16,39,40].

Our research used a mixed methods approach to provide a comprehensive exploration of the factors at play. We began with a descriptive analysis of the innovation success of the targeted pSVV technology, which helps contextualize the case and assess the effectiveness of the policy intervention [41]. Following this, a qualitative analysis of the perception and use of the market access pathway was conducted to develop theoretical insights into the factors that influence the adoption of specific innovation pathways [42,43]. In addition, a functional-structural analysis of the technological innovation system was performed to uncover broader factors beyond the pathway itself that are essential for successful technological innovation [34].

Contribution

This study contributes to the field by providing a comprehensive analysis of factors that support the successful integration of digital health innovations. Our findings will provide policy makers and stakeholders worldwide with practical insights to improve regulated market access pathways and drive innovation strategies that better meet the needs of health care systems and patient empowerment.

Methods

Setting and Study Design

This exploratory study used a mixed methods approach to investigate factors influencing the success of a regulated market access pathway for digital health technologies. We conducted a descriptive analysis to assess the adoption of pSVV technology, followed by a qualitative analysis to explore stakeholder perspectives on the pSVV pathway, using grounded theory and Gioia methodology to ensure neutrality and objectivity. Finally, we performed a functional-structural analysis to identify additional success factors within the broader technology innovation system. Each method is described in detail subsequently.

This study was carried out between May 2023 and November 2024 with interviews being conducted between May 2023 and November 2023.

First, a descriptive analysis of the use of positive health care effects by DiGA manufacturers was conducted to establish an understanding of the current developmental status of the pSVV technology in the market.

Second, a qualitative analysis was performed to explore the success factors of a regulated market access pathway by subsequently analyzing the perception and use of the pSVV pathway of key stakeholders. Since the pSVV pathway was only introduced in 2019 and has not been studied yet, a qualitative research approach was adopted. This approach allows for deep exploration of complex and less understood topics [44]. Unlike quantitative methods, which rely on larger sample sizes and predefined variables, qualitative methods provide flexibility for exploring nuanced, context-specific insights, particularly valuable given the relatively small number of relevant stakeholders (35 DiGA manufacturers as of early 2023). By using the inductive grounded theory method [45] alongside the Gioia methodology [42,43], we objectively analyzed the views of market participants on the pSVV pathway, avoiding any imposition of preconceived theories and ensuring neutrality in the analysis. For the qualitative analysis, we followed the Consolidated Criteria for Reporting Qualitative Research-32 (COREQ-32) checklist [46].

Building on the dataset from the qualitative analysis, we conducted a functional-structural analysis of the technology innovation system established by the pSVV pathway to identify additional success factors related to regulated market access pathways. This analysis followed the systemic policy framework

XSL•FO
RenderX

established by Wieczorek and Hekkert [34], enabling a comprehensive examination of the innovation system's functional dynamics and structural components within the current policy context.

While this study provided valuable insights into potential strategies to improve the pSVV pathway in particular, the international focus of this study precluded detailed insights into Germany-specific factors. To address this, we conducted a subsequent analysis focusing specifically on the question, "What factors facilitate or hinder the implementation of pSVV within the approval process for DiGA?" and "What specific requirements and expectations of relevant stakeholders must be addressed to establish the concept of pSVV in the market?" The results of this specific analysis were published in a German-language journal to inform stakeholders in this context [47].

Sampling and Data Collection

For the execution of our descriptive analysis regarding the state of adoption of pSVV, we evaluated publicly available data on the positive health care effects of DiGA from the DiGA directory [23]. The data were retrieved via an application programming interface provided by BfArM for scientific use.

In our qualitative study, we used a hybrid approach of criterion sampling and snowball sampling for participant recruitment. We aimed to engage with stakeholders who have a direct influence on which digital health products enter the market and, if entry occurs through the DiGA pathway, determine the associated positive health care effects. The inclusion criteria are defined subsequently.

The first inclusion criterion was organizations that are directly involved with the DiGA ordinances and regulations, whether developing it, ensuring or supporting its implementation, or actively using it.

The second criterion was individuals, within these organizations, whose professional roles directly involve the development, implementation, or use of the DiGAVs and regulations.

The third inclusion criterion was additionally individuals who have actively decided against using the DiGA pathway, provided they have extensively engaged with the ordinances and regulations and are potentially qualified to use it (eg, manufacturers of DiGAs not listed as DiGA).

Individuals without direct and active influence over the selection of products entering the DiGA market or the choice of the positive health care effect were excluded.

Consequently, the exclusion criteria led to the exclusion of stakeholder groups in the DiGA market, such as DiGA users, prescribers, and researchers in the field, while the inclusion criteria led to the identification of the stakeholder groups presented in Textbox 1.

Textbox 1. Stakeholder groups included in qualitative analysis.

- Digital health application (DiGA) manufacturers with patient-relevant structural and procedural improvement (pSVV; DiGA with pSVV): This category includes decision makers from manufacturers of DiGAs that are listed in the DiGA directory and have implemented pSVV as a positive health care effect.
- DiGA manufacturers without pSVV (DiGA without pSVV): This category comprises manufacturers who are listed in the DiGA directory but have not used pSVV as a positive health care effect.
- Manufacturers of mobile health applications not listed in the DiGA directory (non-DiGA): This group consists of manufacturers whose CE–certified mobile health apps are not, or are not yet, listed in the DiGA directory.
- Consultants: This group refers to professionals from advisory firms who specialize in assisting manufacturers in the process of getting listed in the DiGA directory.
- Regulatory bodies: This group encompasses public institutions that are deeply involved in the development and management of the DiGA pathway and the evolution of the guiding ordinances and regulations (Digital Healthcare Act [DVG], DiGA ordinance [DiGAV], and DiGA guide).

Due to the very small size of stakeholder groups 1 and 5, yet their significant relevance to our research questions, we contacted all organizations in these groups in accordance with our inclusion criteria. For groups 2 to 4, given the novelty and niche nature of the topic, we opted for snowball sampling [48] to achieve the highest possible participation rate. Individuals were excluded if someone from the same organization had already participated in the study, to avoid multiple responses from the same context. This combined approach resulted in a 66% positive response rate to our participation invitations. Participant recruitment began with initial email contacts, followed by referrals from existing participants in line with snowball sampling. Efforts were focused on ensuring a diverse representation of the identified stakeholder groups.

Our outreach targeted 29 individuals, with 19 (66%) agreeing to participate. Reasons for nonparticipation included lack of response (n=9, 31%) and time constraints (n=1, 4%). The distribution of speakers across groups and their organizational roles are detailed in Table S1 in Multimedia Appendix 1.

For the interviews, a semistructured approach was used with decision makers who had consented to participate and endorsed a data protection declaration. Efforts were made to ensure neutrality, with no prior relationships between interviewers and participants. An interview guide with 7 open-ended questions was prepared to facilitate conversation while ensuring neutrality and objectivity in the analysis. The themes of the questions were developed after preliminary data gathering, which included limited literature research and initial conversations with 2 experts in the field (a managing director from a manufacturer's association and a member of an organization that has 1 DiGA listed in the DiGA directory), adhering to the standards of the Gioia methodology [42]. These experts were not part of the subsequent interviews, ensuring that the development of the interview guide was informed yet independent of the interview

XSL•FO
RenderX

process itself (Multimedia Appendix 1). In total, 2 central themes emerged from this process: the general approval process for DiGA and the specifics of the pSVV pathway. For each of these themes, we formulated specific questions, which were refined during the initial interviews to ensure clarity and neutrality. Interviews were scheduled at the convenience of the participants and conducted via Zoom (Zoom Video Communications, Inc), with only the participant and the primary interviewer (SG) present, though 3 selected interviews had a second researcher (MG) present for quality assurance. No interviews were repeated.

Conducted in German, interviews lasted 25 to 35 minutes, audio tracks were recorded, transcribed using the software Whisper Transcription (Good Snooze), and anonymized. Data saturation was reached when no new first-order concepts or second-order themes emerged during analysis, as reflected in field notes and initial coding. This indicated that the data structure, developed through the Gioia methodology, was comprehensive, and additional data collection would not yield further insights. Following the analysis, data triangulation was enhanced through additional discussions with 4 experts from regulatory and digital health sectors to refine the results. This included a conversation with leading representatives from BfArM, which, due to confidentiality agreements, cannot be directly cited in this study. Insights from these discussions were used to cross-verify emerging themes and validate our interpretations, ensuring robustness in our conclusions.

Ethical Considerations

Informed consent was obtained from all participants who were interviewed strictly as representatives of their respective employers. Participants were selected based solely on their relevance to the research question and their experience in the field, with no consideration of personal or socially sensitive attributes. No personal data or socially relevant information were collected, and all data were fully anonymized to ensure privacy and confidentiality. No compensation was provided for participation in the study.

Furthermore, the study did not involve vulnerable populations, health care professionals, medical procedures, health-related data, or any form of deception, manipulation, or intervention that could impact participants' well-being. In addition, there was no risk of physical, psychological, social, or economic harm to participants. Following consultation with the head of the Institute for Ethics in Technology at Hamburg University of Technology and a review by the German Association for Experimental Economic Research eV (certificate number kNjYR7Ag), it was confirmed that specific ethics approval was not required. This approach fully aligns with best practices for ensuring ethical compliance in socioeconomic research.

Descriptive Analysis

For this analysis, we included all DiGA listed in the DiGA directory at BfArM as of November 8, 2024. This study specifically aimed to compare how often medical benefits are used versus pSVV, as published in the DiGA directory [23]. To further refine our understanding of the adoption levels of pSVV, we also examined the use rates across the 9 key areas of pSVV detailed in the DiGA guide [27].

Qualitative Analysis

Transcripts were methodically analyzed in a 3-step procedure using the grounded theory and the well-established Gioia methodology [42,49] to ensure qualitative rigor in inductive research. After familiarization with the primary data, a codebook was developed by generating codes in an iterative process using the software MaxQDA (VERBI Software). SG coded all (19/19, 100%) interviews, and 2 (11%) randomly selected interviews were cocoded by MG. In addition, SG and MG regularly discussed and reflected on the data collection process and preliminary findings throughout all phases of the research. A consensus on all codes was eventually reached by the research team.

After agreeing on a coding scheme that best captured the diversity of the material, the transcripts and codes were analyzed again to identify second-order categories. First-order concepts included concepts, such as "lack of methods is the biggest hurdle for proof of pSVV" or "pSVV good ideas but difficult to operationalize." First-order concepts were combined into second-order themes, such as "study design for pSVV difficult due to lack of measurement instruments." Ultimately, second-order themes were merged hierarchically into five aggregate dimensions that best summarize the theoretical contribution in the data: (1) "key aspects for manufacturers within the regulated market access pathway," (2) "key aspects for study design," (3) "reasons for considering possibilities within new pathway," (4) "hurdles for considering possibilities within new pathway," and (5) "targeted innovative technology important for the health care system" (refer to Figure S1 and Table S3 in Multimedia Appendix 1).

Following the analysis, a triangulation of the results was conducted with 4 individuals from institutions involved in the process: 1 interview participant and 3 additional experts with extensive subject matter expertise, who had not participated in the initial interview study. All 4 experts were selected from the regulatory field, as many of the derived implications pertained to regulatory issues.

Functional-Structural Analysis

The first-order concepts, originally derived from the previous qualitative analysis as elucidated earlier, were subject to an additional analysis using the functional-structural systemic problem analyses framework [34], a methodology specifically tailored for the exploration of technology innovation systems [50,51]. Our analytical focus centered on the innovative pSVV technology, concentrating on the first-order concepts that laid the foundation for dimensions 2 to 5. Notably, the first dimension, "Key aspects for manufacturers within the regulated market access pathway," was excluded from this analysis, as it did not provide insights specific to the pSVV aspects of the DiGA pathway, but rather addressed the pathway in general.

In step 1, to facilitate a comprehensive analysis, all first-order concepts corresponding to dimensions 2 to 5 were systematically organized into a 4×7 matrix. This matrix integrates the 4 structural elements and 7 functional dimensions of an innovation

XSL•FO
RenderX

system [34,50,52]. Following the analysis framework, the objective was to elucidate overarching "systemic problems" hindering the intended innovation within the innovation system for pSVV. Refer to Table S4 in Multimedia Appendix 1 for the comprehensive assignment of concepts to the functional and structural dimensions of the innovation system and the derivation of "systemic problems." The sorting of concepts, derivation of instruments, and clustering into strategic factors were conducted by SG and subsequently agreed upon by the research team.

In step 2, to address the identified "systemic problems," we first identified "systemic instruments" within the functional-structural framework of Wieczorek and Hekkert [34], tailored to specific systemic issues.

In step 3, we organized the "systemic instruments" identified in step 2 into 3 strategic factors to form actionable clusters. A detailed overview of the derivation process is available in Table S5 in Multimedia Appendix 1.

Results

Current State of pSVV Adoption in the DiGA Market: Descriptive Analysis

As of November 8, 2024, out of the 56 DiGAs available in Germany, only 1 (2%) provisionally listed DiGA uses pSVV as its primary end point, with none in the permanently listed category. While 10 (19%) DiGAs incorporate pSVV as a secondary end point, the predominant focus remains on traditional medical benefits. The adoption of pSVV in comparison to medical benefits stands at 11 versus 45, meaning only 20% (11/56) of all currently listed DiGAs use pSVV. Moreover, only 1 (2%) of 56 DiGA exclusively focuses on pSVV as a primary end point [23].

The 11 (20%) of the 56 DiGA currently listed that use pSVV as a primary or secondary end point have collectively used 12 pSVV end points (Figure 1). Among these, 6 (67%) of the 9 possible key areas outlined in the DiGA guide have been used. In total, 4 (44%) of these areas have been used more than once, namely "patient autonomy," "health literacy," "coping with disease-related difficulties in everyday life," and "alignment of treatment with guidelines and recognized standards."

The 11 (20%) of the 56 DiGA currently listed that use pSVV are owned by 10 organizations, with 1 organization using pSVV end points in clinical trials to prove the positive care effect for 2 of their DiGA products. We estimated the timing of the adoption of the pSVV pathway by examining the time frame of clinical studies and found that 2 DiGA demonstrating a pSVV had already been conducting their clinical studies before the pathway was introduced in 2019 (velibra and vorvida, both developed by GAIA AG).

Figure 1. The number of digital health applications (DiGAs) listed in the DiGA directory in Germany based on positive health care effects and status of listing (as of November 8, 2024), including the breakdown of specific key areas used within patient-relevant structural and procedural improvements (pSVV). *Eleven DiGAs using pSVV use 12 pSVV end points with 1 DiGA using 2 pSVV key areas.

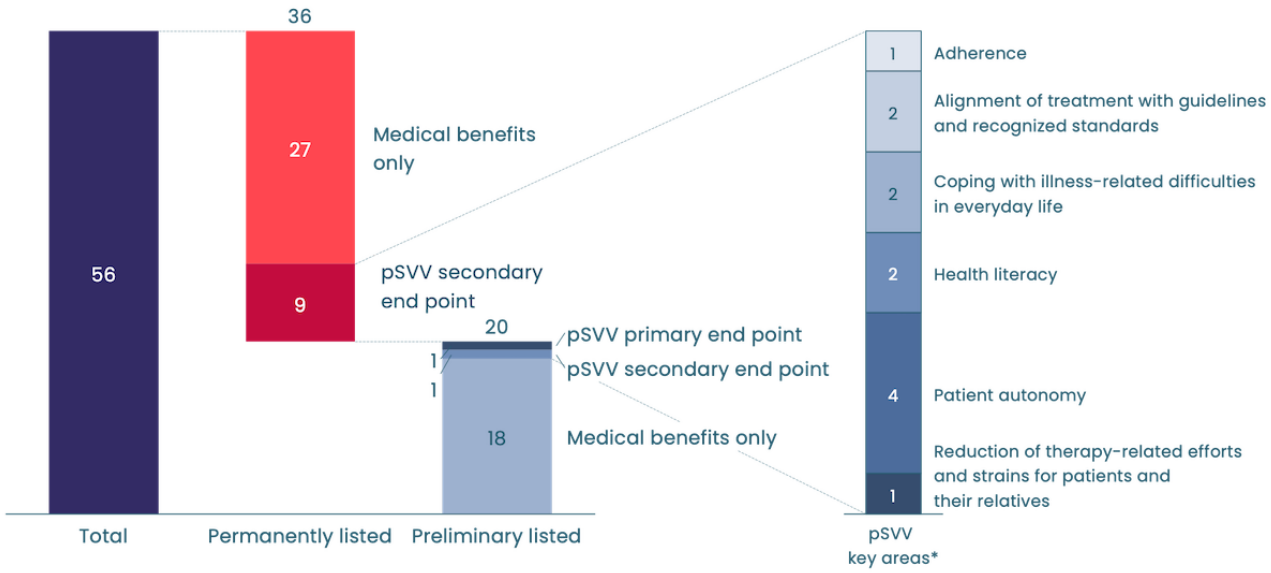

Analyzing Perception and Use of the pSVV Pathway by Key Stakeholders: Qualitative Analysis

Overview

The results are presented based on the data structure developed through the Gioia methodology (Figure S1 in Multimedia Appendix 1), which organizes findings into first-order concepts, second-order themes, and overarching dimensions. In the subsequent sections, findings are systematically presented separating the data structure into 3 segments of aggregated dimensions and further detailed by their corresponding second-order themes. To illustrate our approach, Figures 2-4 provide a visual representation of the respective parts of the data structure. In addition, Table S3 in Multimedia Appendix 1 includes selected codes that were instrumental in forming the data structure.

XSL•FO
RenderX

Figure 2. Graphical illustration of selected parts of the data structure derived from Gioia methodology illustrating key aspects for manufacturers within the regulated market access pathway as well as the key aspects of choosing the health care effect. BfArM: Federal Institute for Drugs and Medical Devices; pVE: positive health care effects.

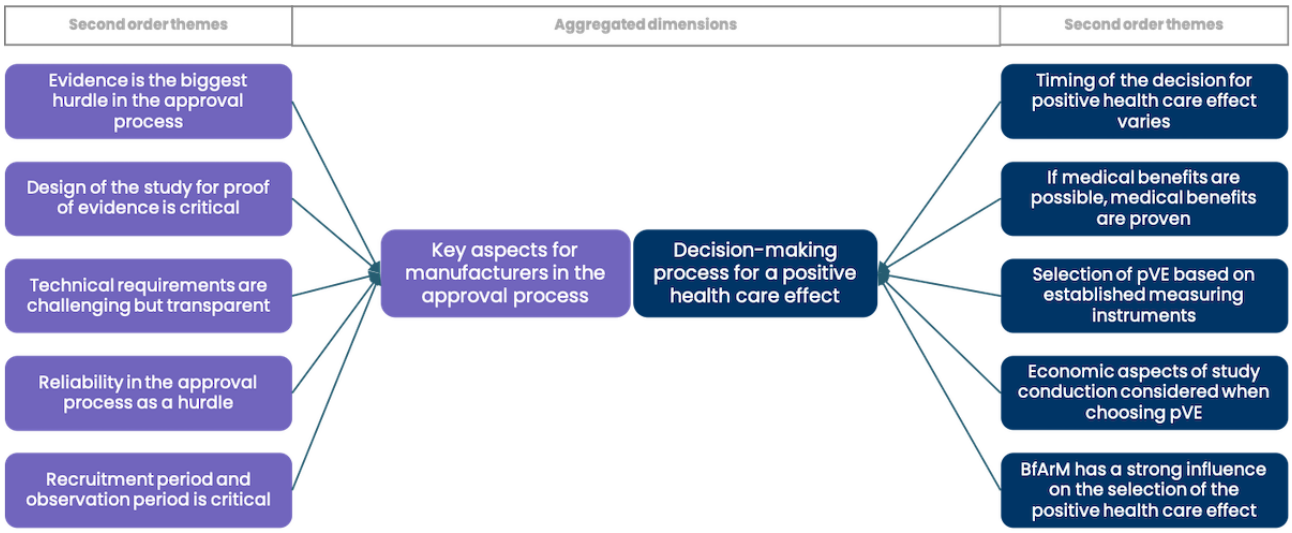

Key Considerations in the Approval Process for DiGAs: Evidence as the Key Aspect

Being asked about the key considerations a DiGA manufacturer needs to face during the approval process, most (14/16, 88%) of the surveyed manufacturers and consultants concurred on a central theme—evidence takes the forefront, serving as the primary concern for manufacturers as "...(evidence) is the crucial point, which often determines the inclusion or non-inclusion (in the DiGA Directory)" (Consultant_5). In conjunction with this, the DiGA study design for evidence generation is deemed critical. The execution of the study is recognized as one of the primary cost drivers for manufacturers and the approving authority (BfArM) has a strong influence on the final design, as described by a participant, "And one can clearly see that an evidence level is crystallizing, which, within the framework that is possible, is already on the strictest side" (DiGA without pSVV_4).

While technical requirements, such as data privacy, interoperability, and accessibility were also mentioned, discussions in these domains with the BfArM were described as negligible due to the "clear and transparent" (Consultant_2) nature of the requirements. One respondent encapsulated this sentiment by stating the following:

> *The key issue is undoubtedly the matter of evidence. Less so is data privacy and security, as these are issues that can be adequately addressed.* [DiGA without pSVV_3]

Nevertheless, a certain level of planning uncertainty was identified in the process, stemming from updates to the DiGA guide or alterations in statements during consultations with the BfArM, as articulated by a manufacturer:

> *One of the most challenging aspects is the disparity between what is theoretically granted regarding research design and the actual decision practices of the BfArM.* [DiGA without pSVV_2]

In addition, insights into the procedural aspects revealed that both recruitment and observation periods are critically constrained.

Key Aspects of Study Design: Strong Tendency Toward Traditional End Points

Regarding the decision on which positive health care effect a future DiGA will be evaluated for and therefore which end point is used in the study design, significant timing variations are observed. DiGA manufacturers, especially those who began developing their product after the DVG was introduced in 2019, typically determine the positive health care effect early in the development phase, often before initiating the coding process:

> *Ideally, I make the decision before the development begins, as the entire development process needs to be built around this choice.* [DiGA with pSVV_1]

In contrast, manufacturers with established products leveraging the new DiGA approval pathway as a market entry point did not have this advantage.

In the selection of the clinical end point, manufacturers strongly lean toward medical benefits:

> *The medical aspects always take precedence.* [DiGA with pSVV_2]

Demonstrating a positive health care effect without proof of medical benefit is perceived as challenging in 2 ways. One way is to find the appropriate instrument for evidence generation for proving pSVV, "If I look at the list of pSVV, then I wonder for 80% of the things, how am I supposed to prove that?" (DiGA without pSVV_1), while the other lies in finding an agreement with the BfArM on a study design aiming to prove pSVV, "We have to advise relying on the medical benefit...it is not fundamentally that they (BfArM) reject it (pSVV), but we also experience that they...are not so certain about it" (Consultant_5). In addition, it is perceived that other stakeholders in the system are better equipped to evaluate medical benefits compared to pSVV, which could be relevant for price negotiations:

XSL•FO
RenderX

> *If you believe you can demonstrate medical benefits, that is probably the better path towards price negotiations because it's something the system understands and can perhaps evaluate better.* [Consultant_6]

The perceived challenges conclude in a perception of a stronger acceptance of evidence that substantiates a medical benefit, highlighting a potential current bias toward well-known, medically oriented outcomes.

Manufacturers and consultants choose the end points, if possible, based on the example of other DiGA that are already listed and where validated measurement instruments are available:

> *[The selection criteria are] essentially hard facts in the sense that it has worked before, and there are validated measurement instruments.* [Consultant_3]

In addition, both listed DiGA and unlisted health applications prioritize economic considerations:

> *Ultimately, we paid attention to...economic aspects, like the speed at which the study is conducted.* [DiGA without pSVV_2]

At present, pivotal guidance in the process emanates from consultations with the BfArM. The influence of the conversations is underscored by the following statement:

> *So as a manufacturer, you would never do anything where the BfArM advice said otherwise.* [DiGA without pSVV_5]

Reasons for Considering a New Pathway: Additional Evidence

Both, consultants and DiGA manufacturers, currently view the use of pSVV primarily as an opportunity for generating additional evidence alongside medical benefits. This perspective was succinctly expressed by one participant, who mentioned the following:

> *pSVV is included as a secondary endpoint in such a study. If possible, you confirm it, but you would always prefer to demonstrate medical benefits.* [Consultant_3]

DiGA manufacturers and consultants have pinpointed 3 substantial benefits associated with the implementation of the possibilities given within the new pSVV pathway. Potential differentiation from providers of similar DiGA was mentioned:

> *But in the case of [indication X], it could well be that with our DiGA, other DiGA are also approved or provisionally approved. In such instances, I could envision that a pSVV might serve as a consideration for or against a DiGA.* [DiGA without pSVV_4]

Overall positive impacts on sales were also mentioned:

> *Distribution is also a crucial task we undertake when engaging with physicians, ensuring a clear delineation of what is effective beyond standard practices. Particularly with DiGA, the argument hinges on the fact that the pSVV dimension precisely provides the benefit.* [DiGA without pSVV_3]

Moreover, there is an expectation of a positive impact on the official price negotiations occurring later in the process between manufacturers and payers. It is highlighted that "The more evidence you have, the greater the likelihood of maintaining certain price points" (DiGA with pSVV_1) and "Possibly, pSVV can indeed be an enhancing criterion (during the price negotiation process) for the respective product" (Regulatory body_2).

Hurdles for Considering a New Pathway: Operational Insecurities

When asked about the reasons why manufacturers tend to not use the new possibilities within the pSVV pathway, many (5/7, 71%) manufacturers and consultants referred to a lack of clarity on operationalizing pSVV as outlined in the DiGA guide, particularly concerning the methods applicable for assessing the impact:

> *But when I look at them [the pSVV key areas], I wonder with about 80% of them, how am I supposed to prove that? And yes, how do I actually incorporate that into studies?* [DiGA without pSVV_1]

In addition, there is skepticism among DiGA manufacturers and consultants regarding the economic impact of demonstrating pSVV. One participant succinctly expressed this viewpoint by stating the following:

> *Purely from the procedural and financial perspective—if you were to ask our boss, who is responsible for finances—there is a clear answer: this is an absolute disaster.* [DiGA with pSVV_1]

At the same time, introducing and establishing new measurement tools like the use of real-world data is deemed too expensive and risky for DiGA manufacturers:

> *If you had to essentially innovate methods and also deal with the uncertainty...in the worst-case scenario, you would end up with three studies.... That alone costs you a million, just for these studies. No one is willing to foot that bill. And then, if it goes wrong...* [Consultant_3]

In addition, manufacturers and consultants observed a general uncertainty within the system regarding the concept of pSVV:

> *Not that they (BfArM) fundamentally reject it, but we experience that they, how should I say, are not so certain about it.* [Consultant_5]

Moreover, they observed a general lack of knowledge about pSVV that leads to the need for explanation:

> *There is a significant need for explanation, actually, among the different stakeholders. Both with patients and with doctors. What is it exactly? Why is it good? What is in it for me?* [DiGA with pSVV_1]

Consultancies sometimes actively discourage the choice of pSVV in the clinical trial:

> *If you believe you can demonstrate medical benefits, that is probably the better path towards price negotiations because it's something the system*

XSL•FO
RenderX

> *understands and can perhaps evaluate better.* [Consultant_6]

Relevance Perception: Targeted Innovative Technology Is Important for the Health Care System

Looking at the general perception of pSVV as a concept, stakeholders highlighted the systemic benefits, emphasizing the impact on the overall health care system efficiency and supply effects:

> *It...makes sense from the overall health care system perspective.* [Non-DiGA_2]

Patient-centric advantages emerged, revealing the influence of DiGA beyond conventional medical outcomes. A consultant illustrated this by pointing out the positive impact of patient empowerment on intervention success:

> *Individuals with higher health literacy typically respond more effectively to interventions, as they have a better understanding and feel a greater sense of control.* [Consultant_1]

DiGA might also address practical challenges like long waiting times, presenting substantial value beyond traditional medical considerations:

> *But if we now say we would compare ourselves to a traditional intervention...then the pSVV...plays an even greater role. Because when I say, I have something here with a positive effect, and it is more accessible than the therapy, that is already a value where I say, it makes a difference.* [DiGA with pSVV_2]

However, the unique benefits of DiGA, such as the improvement of patient adherence and reduction of therapy-related efforts, often eluded recognition in conventional benefit assessments:

> *pSVV is crucial for the system, but the evidence must focus on it. Thus, the usability of the product must be ensured and tailored to that. The evidence must reflect the particular added value of these applications.* [DiGA without pSVV_5]

Regulatory bodies acknowledged this oversight, emphasizing the need to broaden evaluation criteria to encompass these critical aspects:

> *DiGA can have medical benefits, but they are not solely limited to medical utility.... And these are points that get lost in traditional benefit assessments.* [Regulatory body_1]

In addition, the political and collaborative nature of the development process that led to the pSVV concept was highlighted: “We had many stakeholders involved there. On the one hand, we had colleagues from self-administration, but we also had the manufacturers [stretching the collective effort] in understanding potential, addressing challenges, and bridging the gap between traditional assessment methods and innovative logics” (Regulatory body_1).

Analyzing the pSVV Technology Innovation System: 5 Systemic Problems Identified

Actors’ Presence Problem in “Entrepreneurial Activities” and “Market Formation”

Examining the participants presently engaged in the pSVV innovation system reveals a discernible pattern. Only one of the DiGA entrepreneurs within this emergent system has developed a product with a primary emphasis on pSVV. Notably, some manufacturers have established market longevity before the introduction of the DVG and thus pSVV; therefore, orienting their product toward medical benefits. Moreover, some actors are concurrently navigating evidence-generation endeavors while exploring alternative pathways for market access in Germany, such as individual “selective contracts” with statutory health insurance under §140a SGB V (Sozialgesetzbuch 5; a German law that covers all regulation concerning statutory health insurance and therefore also covers DiGA and selective contracts), which lacks the concept of pSVV. The absence of *entrepreneurial activities* singularly centered on pSVV emerges as a contributory factor hindering the *formation of a distinct market* for dedicated pSVV-related regulated digital medical devices.

Actors’ Capability Problem in “Knowledge Dissemination”

The supervising governmental institution, in its role as advisor for manufacturers before and during the admission process, has limited experience with pSVV but rather a strong focus on medical benefits. This emphasis can be attributed to its dual role, as the BfArM in Germany oversees admissions not only for DiGAs but also for traditional pharmaceuticals. The lack of experience with pSVV in this key role as a knowledge multiplicator and also in this specific innovation system hampers *knowledge dissemination*. In addition, practitioners prescribing DiGA as well as end users of the products have limited experience with products focusing on pSVV, which leads to a need for significant efforts to explain the concept while the demand from potential users is unclear.

Infrastructure Presence Problem in “Knowledge Development” and “Guidance of Search”

The suitability of existing measurement methods designed for therapeutic effects in assessing pSVV is often limited, contributing to its restricted adoption. Simultaneously, the lack of positive examples for products in the pSVV category hampers *knowledge development*. In addition, the absence of positive price examples not only further constrains manufacturers’ support for pSVV but also contradicts the *guidance of search*.

Institution Intensity Problem in “Knowledge Development” and “Resources Mobilization”

In the *institutional* dimension, it was noted in particular that the existing *instructions* (DiGA guide) pose challenges to the operationalization of pSVV. The *established practices* and formed *expectations* regarding pSVV are perceived as counterproductive, particularly for cost-sensitive manufacturers who cannot afford to allocate their limited resources toward pSVV despite existing barriers. The limitations of the DiGA

XSL•FO
RenderX

guide concerning pSVV, namely lack of examples or standards and possible end points for evidence demonstration, are identified as a challenge in *knowledge development*, while the *established practices*, inclined toward medical benefits, hinder *resource allocation* for the advancement of pSVV.

Institution Intensity and Interaction Intensity Problem in "Guidance of the Search," "Market Formation," and "Creation of Legitimacy"

Established practices in the approval process that make it difficult to enter the market without proven medical benefits diverge from the introduced laws and regulations, creating a misalignment. Interactions within the approval process, especially with the official advisory body, tend to emphasize medical benefits. Notably, consultancies actively guide manufacturers toward prioritizing medical benefits over pSVV, thus steering the guidance of the search away from pSVV and hindering market formation.

Positive expectations on the value of using pSVV that would support the *creation of legitimacy* remain mainly hypothetical. The economic effect, specifically, was formulated in the subjunctive by interviewees. Coupled with the pervasive *established practices*, this poses a significant obstacle to *creating legitimacy* for pSVV.

Discussion

Principal Findings

Our research explored the factors that need to be considered when bringing a specific digital health technology innovation to market through the implementation of a regulated market access pathway as a health policy approach.

For this example case study, the pSVV pathway, the descriptive analysis revealed that only a few DiGAs currently in the market have demonstrated a pSVV. Of the 11 currently listed DiGAs that have entered the market with a pSVV as one element of a positive health care effect since the introduction of the pathway in 2019, only 1 DiGA has a primary focus on pSVV (ie, pSVV as the primary end point of the positive health care effect). Notably, 2 of the products (developed by the same DiGA manufacturer) that demonstrate pSVV as a secondary end point conducted their clinical studies before 2019. This indicates that few manufacturers have taken the initiative to develop a product with a special focus on improving structures and procedures for patients, even though the introduction of the pSVV pathway was intended to encourage this. To date, pSVV is mostly used as an additional end point to medical benefits. To provide a basis for interpreting the results in further analyses on the performance factors of innovation systems shaped by regulated market access pathways, our initial conclusion is that the innovation in DiGAs targeting pSVV, which the pSVV pathway was designed to foster, has not yet been fully realized within the system. This is particularly daunting, as previous studies emphasize the need to focus more on products that address structural and procedural improvements for patients [53-55] and consider the concept of pSVV to be a "fundamental mechanism of action for DiGAs" [56].

Our qualitative analysis of stakeholders' perspectives during market access reveals that evidence generation is the most critical component of the DiGA approval process. While multiple factors are considered in determining a positive health care effect, our findings suggest that, at the operational level, manufacturers of DiGAs primarily focus on demonstrating medical benefits as evidence for positive health care outcomes (Figure 2). Economic considerations, in light of the lack of established methodologies for measuring the defined key areas and uncertainty about the acceptability of the proposed study design, including pSVV, lead to more cautious use. Bearing the cost for methodological innovation is not considered attractive among manufacturers, particularly as DiGAs with pSVV so far do not negotiate substantially higher prices than DiGAs without pSVV [28]. Consistent with the findings of the descriptive analysis, the use of pSVV is often considered "incidentally" as an addition to medical benefit, only when a methodology for demonstrating one of the pSVV key areas exists for a specific product (Figure 3). It can be concluded that the specific challenges associated with proving the efficacy of pSVV, particularly the lack of established methodologies, currently outweigh its potential benefits. Previous studies have highlighted content and format security [57] as a critical factor, particularly for pioneers in medical technology, which also appears to hold true for the market entry of innovative products in DiGAs. The fact that only one company has attempted to pioneer a product specifically targeting benefits under pSVV may be attributed to the innovators' need for content and format security, in this case specifically the lack of standards to generate evidence for the targeted technology innovation. Without such security, the costs for pioneers become prohibitively high.

However, the implementation of the pSVV concept is recognized as vital for the health care system's evolution itself (Figure 4).

Our qualitative analysis highlights a common gap in health care innovation between the perceived importance of new technologies and their actual implementation. This gap is often influenced by systemic factors, such as regulatory conditions and inherent system inertia, alongside a general openness to innovation among stakeholders [58]. For the pSVV pathway, while there is strong stakeholder support for introducing products that demonstrate pSVV, significant operational hurdles, particularly concerning content and format security, have delayed its adoption. These barriers suggest that, despite a willingness to embrace this innovation, the pathway's practical challenges currently limit its broader impact (Figure 5).

XSL•FO
RenderX

Figure 3. Graphical illustration of selected parts of the data structure derived from Gioia methodology illustrating hurdles of DiGA manufacturers for considering possibilities within the new pathway as well as reasons for doing so. DiGA: digital health application; pSVV: patient-relevant structural and procedural improvement.

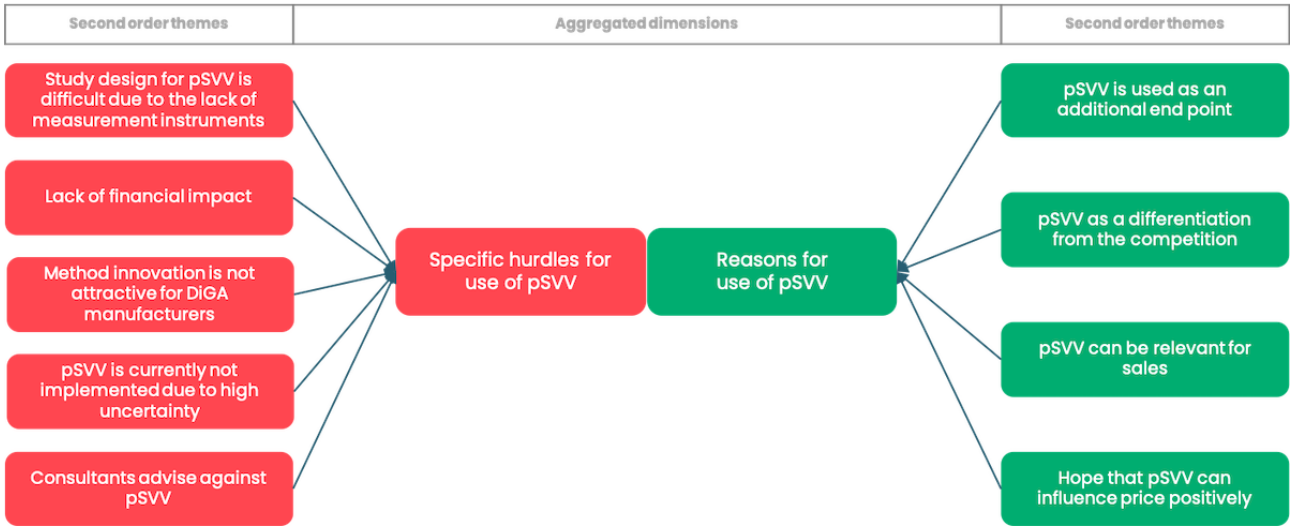

Figure 4. Graphical illustration of a selected part of the data structure derived from Gioia methodology illustrating the perception of the importance of patient-relevant structural and procedural improvements (pSVVs) in the health care system. DiGA: digital health application.

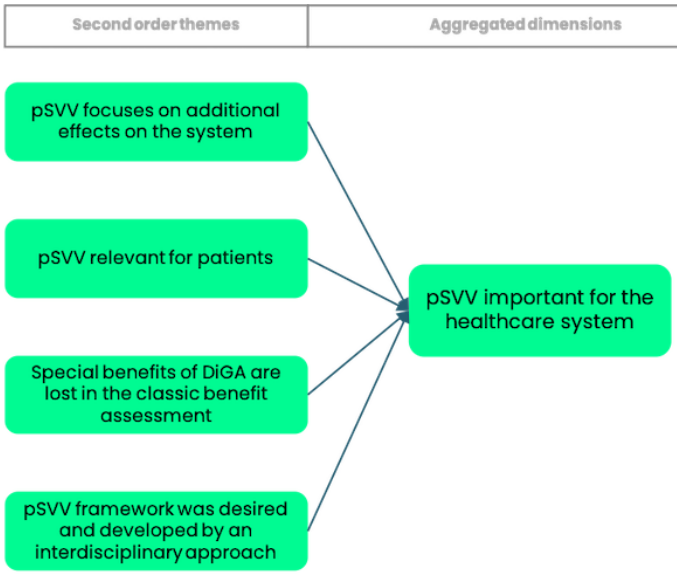

To further explore performance factors in the innovation system shaped by the pSVV pathway, we conducted an analysis of the innovation system's maturity level in relation to pSVV. Our findings indicate 3 central strategic factors of an innovation system in which we can expedite the adoption of innovative technology: *framework specification*, *active innovation processmanagement*, and *market formation stimulation*

While previous work has highlighted the general need for frameworks [16,59], our emphasis is on the degree of specificity required for the successful implementation of a market access pathway for DiGAs. In our example, the specification of the pSVV pathway, defined in the DiGA guide, requires 2D adjustments. First, the creation of a knowledge infrastructure would enable entrepreneurs to navigate the DiGA market, including pSVV, better if the development of exemplary models for pSVV products and especially congruent methodologies for demonstrating the evidence base of pSVV would exist. In addition, a multitude of studies highlight the ongoing challenges associated with identifying appropriate study designs that are capable of measuring effects solely attributable to the use of digital products [60-62]. We conclude that the presence of a specific level of knowledge infrastructure, which underpins the innovation driven by the introduction of a regulated market access pathway, can affect how widely it is adopted [6,63].

XSL•FO
RenderX

Figure 5. Underlying conflict leading to the underuse of patient-relevant structural and procedural improvements (pSVVs).

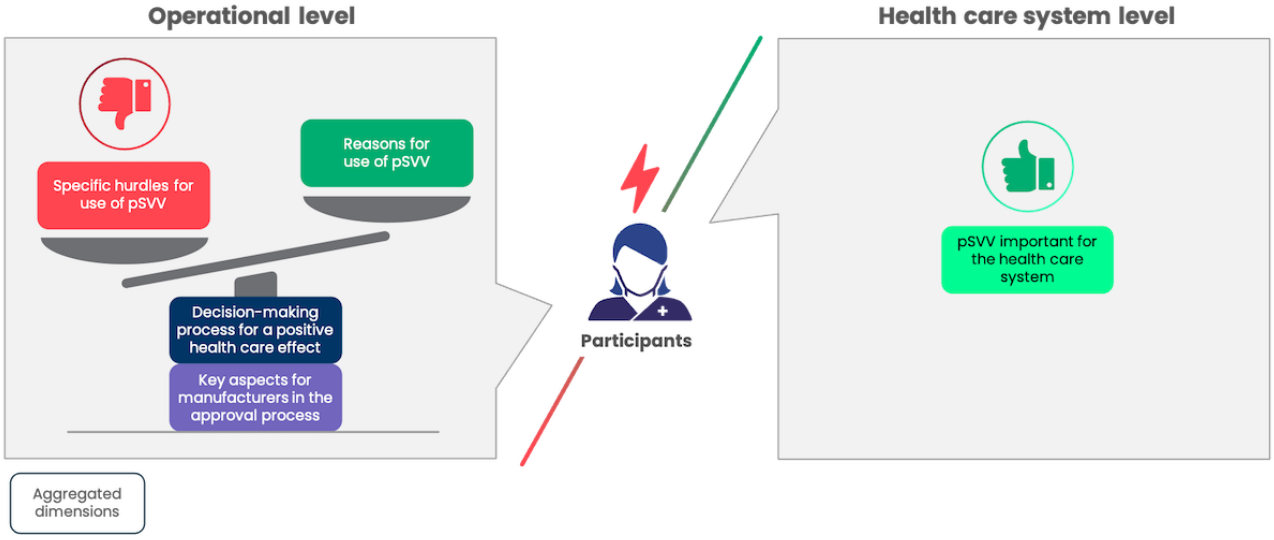

Second, studies in the field of regulated medical products have indicated that the uncertainty about the accepted content and format during the approval of new products can make the pathway into the market unattractive for first movers [57]. Our analysis has shown that the DiGA guide, which serves as the most relevant instruction at the operational level in this system, requires timely updates to keep pace with advancing knowledge development and ensuring up-to-date format and content security. Simultaneously, for the successful implementation of a pathway, *active innovation process management* is crucial. Providing training to key actors in the defined process, such as the BfArM or consultants, has been shown to be essential for empowering stakeholders to support and advance pSVV as an innovation. In addition, as previous work has highlighted the importance of education and awareness as well as patient and health care provider support while training for physicians on digital health solutions is limited [25,64], the inclusion of user groups and prescribing physicians can be focused to sharpen the objectives of pSVV. Conversely, it is crucial to ensure that involved actors are not overly familiar with traditional processes, as this can lead to inflexible adherence to these methods. The involvement of key actors with an excess of experience in medical benefits (ie, experienced regulatory assessor previously involved in assessing pharmaceutics) could potentially impede innovation speed if their participation is overly predominant.

Furthermore, targeted *market formation stimulation* can aim to introduce manufacturers into the DiGA market, focusing primarily on addressing pSVV key areas [65]. Creating positive examples can accelerate innovation at multiple levels. In addition, stimulating constructive exchange regarding the pSVV concept can expedite knowledge generation and distribution, contributing to faster market formation.

In summary, our study revealed that the lack of established methodologies for proving pSVV has led to significant uncertainty for manufacturers regarding the acceptable formats and evidence required, contributing to slower adoption. To complement these findings on the specific pathway level, the functional-structural analysis identified 3 key factors critical to enhancing the total innovation system: clearer framework specification to provide precise guidelines, active innovation process management to support stakeholders in navigating the regulatory process, and market formation stimulation to foster positive examples and increase engagement with the pSVV pathway. These factors are essential for overcoming both the operational and systemic barriers to successfully introducing innovative digital health technologies into a market.

As initiators of the systemic innovation of pSVV, the regulatory body can implement additional measures to promote its further establishment. This includes stimulating the development of a knowledge infrastructure by involving all stakeholders in the innovation system. Academia can actively contribute to advancing the definition of the pSVV and particularly developing suitable methodologies for evidence generation [66-68]. Existing mechanisms to promote scientific progress, such as dedicated research funds (eg, in Germany by the Federal Ministry of Education and Research or the Innovation Fund of the Federal Joint Committee), could be set up for this objective [69].

Initiating a public debate on pSVV can engage relevant stakeholders, such as professional societies and associations in a discourse to further develop the pSVV framework [13,66]. Economic measures, such as granting financial support for conducting clinical trials assessing pSVV or even tax incentives for implementation, can provide additional incentives for manufacturers driving innovation and initiating the pSVV market until these products are established in the standard of care. Furthermore, spaces can be created for key stakeholders in the system to educate themselves about pSVV. This could include offering policy laboratories or training programs for knowledge multipliers.

Consultancies and the BfArM, as influential players in the system, can provide room for their employees to further educate themselves about the possibilities of pSVV, to prevent disproportionately biased inclinations toward medical benefits. Ensuring a balanced wealth of experience regarding the use of medical benefits within the workforce can promote openness to innovative processes. It is critical to ensure that the experience of the staff responsible for evaluating the proof of benefit of

XSL•FO
RenderX

DiGAs is balanced in assessing medical quality [70]. Previous work has shown that the Food and Drug Administration might actively aid in similar innovation processes [71]. This assistance is rendered through the administration's in-depth knowledge to improve study designs and by championing innovative methods. Although the breadth of authority and scope of the regulated processes vary, BfArM or other regulatory agencies involved in the approval process and specifically in evaluating coverage eligibility can also contribute their expertise. For manufacturers seeking to use the potential of pSVV in their DiGAs, it is important to prioritize evidence of positive health care effects early in the development process. Identifying the "active ingredient" and its mechanism is crucial for crafting studies effectively assessing DiGA [60].

Besides proposing solutions for individual stakeholder groups within the health care system (Figure 6), the collaboration and coordination among all involved parties should be underscored as crucial for success. As highlighted on an international scale, effective collaboration between regulatory agencies, academia, app developers, payers, and health care providers is essential to truly leverage advances in science and technology and to translate the vision into reality, guided by the right incentives and guidelines [69,71,72].

Figure 6. Central strategic factors for progress exploration in the innovation system derived from an exemplary analysis of the patient-relevant structural and procedural improvement pathway.

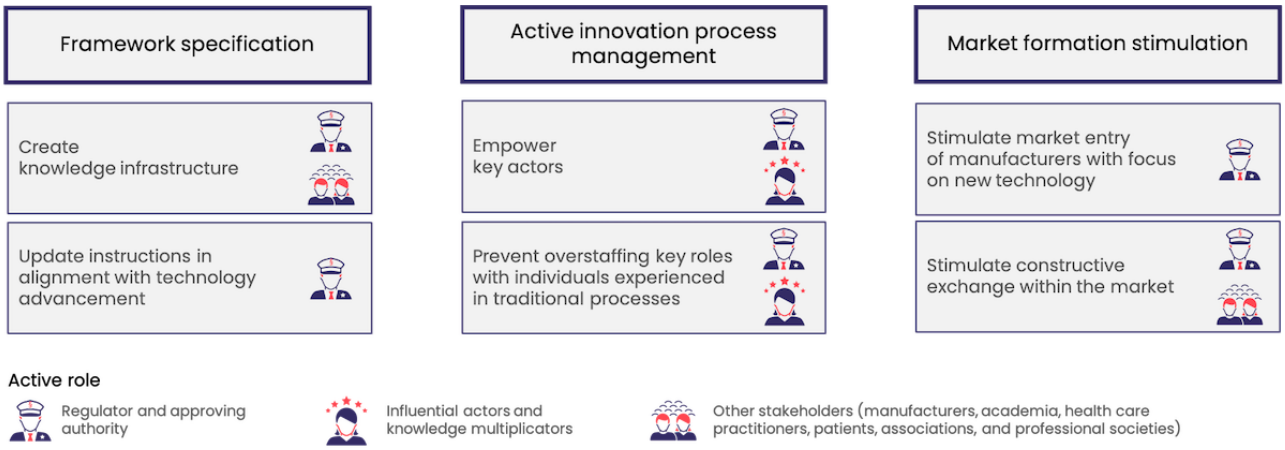

Our analysis suggests that the strategic factors identified are valuable considerations for the establishment of a regulatory market access pathway, to ensure its widespread and successful adoption. Drawing on previous research, the strategic factor of content and format security [57] is highlighted, alongside the adoption of recognized and accepted methods for measuring proposed parameters [59,61,68,73-76]. This is particularly relevant in digital health technology aimed at patient empowerment, where significant progress is still needed [68,74]. Furthermore, it is advantageous to strike an optimal balance between leveraging established processes and frameworks allowing room for innovation [77]. As indicated by the strategic factor of innovation process management, it is crucial to carefully select and train key players and individuals in pivotal roles, equipping them with the specific skills and mindset needed to drive innovation. Finally, the strategic factor of market formation stimulation points to the critical need to consider the practicality of a framework in enabling a viable business model, particularly within the context of systemic innovation [21,60,78].

Limitations

Acknowledging the limitations of this study, a conclusive assessment of the economic value of pSVV is precluded, as no product with pSVV as a primary end point has undergone final price negotiation yet (Table S2 in Multimedia Appendix 1). In addition, it could be questioned whether pSVV can ever be the primary end point of a DiGA, as all DiGAs must also be CE–certified medical devices with the corresponding intended purpose. We have not explored this discussion in detail, as it can certainly be conducted more effectively with selected examples of products focusing on pSVV, once these are available in the market. Furthermore, our qualitative analysis, conducted through semistructured interviews, did not encompass all functional-structural dimensions of innovation systems. Entities, such as academia or associations were not mentioned by any interviewees and were thus not part of the evaluation. In addition, our sampling strategy did not include patients or physicians, as the study concentrated on identifying reasons for the scarce number of products targeting the pSVV key focus areas. This decision acknowledges the limited familiarity of end users and prescribers with the pSVV concept at this point and their marginal impact on market entry decisions shortly after the introduction of a new framework designed to facilitate market access.

Conclusions

This exploratory study examined critical factors for bringing digital health technology innovation to market through a regulated market access pathway, with a focus on the pSVV pathway as an example case. Key success factors identified included content and format security, the need for precise framework specification, active innovation process management, and market formation stimulation. These elements collectively help reduce uncertainties for manufacturers and promote broader adoption of digital health innovations. Our findings revealed that although the pSVV pathway was intended to encourage a focus on pSVVs, the current use remains limited, with most manufacturers prioritizing medical benefits as evidence of positive health care outcomes.

XSL•FO
RenderX

The mixed methods approach enriched the analysis by allowing a deep exploration of stakeholder perspectives through qualitative analysis, complemented by the functional-structural analysis of the technology innovation system. The combination of grounded theory and Gioia methodology enabled an objective and detailed examination of how market participants view and use the pSVV pathway, while the systemic analysis highlighted additional success factors in the broader innovation system context.

Our findings contribute valuable insights for stakeholders in the digital health sector, providing recommendations for policy adjustments that can strengthen innovation pathways. With this work, we contribute by identifying critical areas for improvement, including the need for clearer guidelines, the development of a knowledge infrastructure, and enhanced training for stakeholders to help overcome operational barriers. In addition, fostering public discourse on pSVV and introducing economic incentives, such as financial support for clinical trials, may further drive innovation. These recommendations are valuable for health care systems and regulatory bodies seeking to support similar pathways and advance patient-centered digital health solutions.

Furthermore, the impartial perspectives gathered from stakeholders, combined with a holistic analysis of the innovation system, contribute to identifying clear fields for action and proposing solutions relevant beyond this specific case. Our findings provide transferable insights for health care systems embarking on similar systemic innovations, as seen in countries like France, Belgium, and Austria, where comparable concepts to pSVV are being introduced. The implications of our research extend to other systemic innovation initiatives in health care systems facing significant adoption challenges that could otherwise lead to initiative failure. Ongoing efforts from national and international scholars are critical in supporting the success of these initiatives. In Germany, for example, a reimbursement pathway for digital nursing care applications (DiPA), a “regulatory sibling” to DiGA, has been established. While speculative, similar challenges may emerge for DiPA manufacturers, particularly concerning the nursing benefit that aligns with both medical benefit and pSVV, possibly contributing to the current absence of DiPA in the marketplace and highlighting gaps in this systemic innovation approach.

Acknowledgments

The authors are grateful to the 2 anonymous reviewers whose insightful comments and suggestions contributed substantially to the quality of this manuscript. Furthermore, the authors extend their heartfelt thanks to all the participants of this study, whose time, effort, and invaluable contributions made this research possible.

The study was part of a doctorate program at our university. Publishing fees were supported by the Funding Programme Open Access Publishing of the Hamburg University of Technology.

The authors used generative artificial intelligence (ChatGPT, version 4o; OpenAI) to incorporate minor improvements in sentence structure and wording throughout the publication; the contents and conclusions have purely been developed by the authors [79].

Data Availability

The datasets generated and analyzed during this study are not publicly available due to data protection reasons but are available from the corresponding author on reasonable request.

Authors' Contributions

SG conceived and designed the study, interviewed the participants coded the interviews, analyzed the data, and drafted the manuscript. MG transcribed the interviews using Whisper Transcription (Good Snooze). MG cocoded on randomly selected interviews and took part in reflexive discussion and consensus building together with SG. MG supervised the project. SG and MG interpreted the results and finalized the manuscript. All authors approved the final manuscript.

Conflicts of Interest

SG is employed at a statutory health insurance company in Germany. She declares that her work at the statutory health insurance is unrelated to this topic. MG has previously worked as a consultant for digital health application (DiGA) manufacturers. He declares that his previous work is unrelated to the results of the conducted work.

Multimedia Appendix 1

Semistructured interview questionnaires, overview of interview participants, full data structure based on Gioia methodology, negotiated digital health application prices and health care effects, selected interview codes generated using grounded theory, functional-structural analysis of patient-relevant structural and procedural improvement as a systemic innovation, and derivation of strategic fields using the functional-structural framework by Wieczorek and Hekkert [34].

[DOCX File , 651 KB - jmir_v27i1e66356_app1.docx]

XSL•FO
RenderX

References

1. Haring M, Freigang F, Amelung V, Gersch M. What can healthcare systems learn from looking at tensions in innovation processes? A systematic literature review. BMC Health Serv Res 2022 Oct 28;22(1):1299 [FREE Full text] [doi: 10.1186/s12913-022-08626-7] [Medline: 36307839]
2. Siopis G, Moschonis G, Eweka E, Jung J, Kwasnicka D, Asare BY, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with hypertension: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023;5(3):e144-e159 [FREE Full text] [doi: 10.1016/S2589-7500(23)00002-X] [Medline: 36828607]
3. Hunter DJ, Bengoa R. Meeting the challenge of health system transformation in European countries. Policy Soc 2023 Mar;42(1):14-27 [FREE Full text] [doi: 10.1093/polsoc/puac022]
4. Coiera E. Why system inertia makes health reform so difficult. BMJ 2011 Jun 23;342:d3693. [doi: 10.1136/bmj.d3693] [Medline: 21700652]
5. Gilbert S, Anderson S, Daumer M, Li P, Melvin T, Williams R. Learning from experience and finding the right balance in the governance of artificial intelligence and digital health technologies. J Med Internet Res 2023 Apr 14;25:e43682 [FREE Full text] [doi: 10.2196/43682] [Medline: 37058329]
6. Coder M, McBride L, McClenahan S. Core elements of national policy for digital health technology evidence and access. NPJ Digit Med 2024 Aug 13;7(1):212 [FREE Full text] [doi: 10.1038/s41746-024-01209-2] [Medline: 39138261]
7. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
8. Kotecha D, Chua WW, Fabritz L, Hendriks J, Casadei B, Schotten U, et al. European Society of Cardiology smartphone and tablet applications for patients with atrial fibrillation and their health care providers. Europace 2018 Feb 01;20(2):225-233 [FREE Full text] [doi: 10.1093/europace/eux299] [Medline: 29040548]
9. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023 Mar 10;6(1):38 [FREE Full text] [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
10. Weik L, Fehring L, Mortsiefer A, Meister S. Understanding inherent influencing factors to digital health adoption in general practices through a mixed-methods analysis. NPJ Digit Med 2024 Feb 27;7(1):47 [FREE Full text] [doi: 10.1038/s41746-024-01049-0] [Medline: 38413767]
11. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
12. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
13. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
14. Sedhom R, McShea MJ, Cohen AB, Webster JA, Mathews SC. Mobile app validation: a digital health scorecard approach. NPJ Digit Med 2021 Jul 15;4(1):111 [FREE Full text] [doi: 10.1038/s41746-021-00476-7] [Medline: 34267296]
15. Enforcement policy for digital health devices for treating psychiatric disorders during the coronavirus disease 2019 (COVID-19) public health emergency. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Office of Product Evaluation and Quality. 2020 Apr. URL: https://downloads.regulations.gov/FDA-2020-D-1138-0068/attachment_1.pdf [accessed 2024-12-18]
16. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31 [FREE Full text] [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
17. Ju JH, Sim B, Lee J, Lee JY. Reimbursement of digital therapeutics: future perspectives in Korea. Korean Circ J 2022 Apr;52(4):265-279 [FREE Full text] [doi: 10.4070/kcj.2022.0014] [Medline: 35388995]
18. Nomura A. Digital health, digital medicine, and digital therapeutics in cardiology: current evidence and future perspective in Japan. Hypertens Res 2023 Sep;46(9):2126-2134 [FREE Full text] [doi: 10.1038/s41440-023-01317-8] [Medline: 37258624]
19. Ludewig G, Klose C, Hunze L, Matenaar S. Digital health applications: statutory introduction of patient-centred digital innovations into healthcare. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64(10):1198-1206 [FREE Full text] [doi: 10.1007/s00103-021-03407-9] [Medline: 34529096]
20. Martin T, Hervias A, Armoiry X, Martelli N. Early access programs for medical devices in France: Overview of recent reforms and outcomes (2015-2022). Health Policy 2024 Oct;148:105146 [FREE Full text] [doi: 10.1016/j.healthpol.2024.105146] [Medline: 39154454]
21. Stern AD, Matthies H, Hagen J, Brönnecke JB, Debatin JF. Want to see the future of digital health tools? Look to Germany. Harvard Business Review. 2020 Dec 2. URL: https://hbr.org/2020/12/want-to-see-the-future-of-digital-health-tools-look-to-germany [accessed 2024-12-18]

XSL•FO
RenderX

22. Hullegie P, Klein TJ. The effect of private health insurance on medical care utilization and self-assessed health in Germany. Health Econ 2010 Sep;19(9):1048-1062. [doi: 10.1002/hec.1642] [Medline: 20665706]
23. DiGA-Verzeichnis. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2024-11-08]
24. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3(1):94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
25. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
26. Lauer W, Löbker W, Höfgen B. Digital health applications (DiGA): assessment of reimbursability by means of the "DiGA Fast Track" procedure at the Federal Institute for Drugs and Medical Devices (BfArM). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1232-1240 [FREE Full text] [doi: 10.1007/s00103-021-03409-7] [Medline: 34529095]
27. The fast-track process for digital health applications (DiGA) according to section 139e SGB V. A guide for manufacturers, service providers and users. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.html [accessed 2024-12-18]
28. Goeldner M, Gehder S. Digital health applications (DiGAs) on a fast track: insights from a data-driven analysis of prescribable digital therapeutics in Germany from 2020 to mid-2024. J Med Internet Res 2024 Aug 29;26:e59013. [doi: 10.2196/59013] [Medline: 39208415]
29. Schmidt L, Pawlitzki M, Renard BY, Meuth SG, Masanneck L. The three-year evolution of Germany's Digital Therapeutics reimbursement program and its path forward. NPJ Digit Med 2024 May 24;7(1):139 [FREE Full text] [doi: 10.1038/s41746-024-01137-1] [Medline: 38789620]
30. Schramm L, Carbon CC. Critical success factors for creating sustainable digital health applications: a systematic review of the German case. Digital Health 2024;10:20552076241249604 [FREE Full text] [doi: 10.1177/20552076241249604] [Medline: 38665883]
31. Woiceshyn J, Eriksson P. How innovation systems in Finland and Alberta work: lessons for policy and practice. Innovation 2014;16(1):19-31. [doi: 10.5172/impp.2014.16.1.19]
32. Smits R, Kuhlmann S. The rise of systemic instruments in innovation policy. Int J Foresight Innov Policy 2004;1(1/2):4. [doi: 10.1504/ijfip.2004.004621]
33. van Mierlo B, Leeuwis C, Smits R, Woolthuis RK. Learning towards system innovation: evaluating a systemic instrument. Technol Forecast Soc Change 2010 Feb;77(2):318-334. [doi: 10.1016/j.techfore.2009.08.004]
34. Wieczorek AJ, Hekkert MP. Systemic instruments for systemic innovation problems: a framework for policy makers and innovation scholars. Sci Public Policy 2012 Feb 06;39(1):74-87. [doi: 10.1093/scipol/scr008]
35. Hekkert MP, Suurs RA, Negro SO, Kuhlmann S, Smits RE. Functions of innovation systems: a new approach for analysing technological change. Technol Forecast Soc Change 2007 May;74(4):413-432. [doi: 10.1016/j.techfore.2006.03.002]
36. General methods: version 6.1 of 24 January 2022. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. 2022. URL: https://www.iqwig.de/methoden/general-methods_version-6-1.pdf [accessed 2024-12-18]
37. Dadgar M, Joshi KD. The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. J Assoc Inf Syst 2018 Feb;19(2). [doi: 10.17705/jais1.00485]
38. Fürstenau D, Gersch M, Schreiter S. Digital therapeutics (DTx). Bus Inf Syst Eng 2023 Apr 06;65:349-360. [doi: 10.1007/s12599-023-00804-z]
39. Goodday SM, Geddes JR, Friend SH. Disrupting the power balance between doctors and patients in the digital era. Lancet Digit Health 2021 Mar;3(3):e142-e143 [FREE Full text] [doi: 10.1016/S2589-7500(21)00004-2] [Medline: 33516653]
40. Ryll B. Digital health applications-patient-centric care with disruptive potential. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1207-1212 [FREE Full text] [doi: 10.1007/s00103-021-03421-x] [Medline: 34477883]
41. Loeb S, Dynarski S, McFarland D, Morris P, Reardon S, Reber S. Descriptive analysis in education: a guide for researchers. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. 2017. URL: https://files.eric.ed.gov/fulltext/ED573325.pdf [accessed 2024-12-18]
42. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research: notes on the Gioia Methodology. Org Res Methods 2012 Jul 24;16(1):15-31. [doi: 10.1177/1094428112452151]
43. Magnani G, Gioia D. Using the Gioia Methodology in international business and entrepreneurship research. Int Bus Rev 2023 Apr;32(2):102097. [doi: 10.1016/j.ibusrev.2022.102097]
44. Tracy SJ. Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact. Hoboken, NJ: John Wiley & Sons; 2019.
45. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications; 1990.
46. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Quality Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

XSL•FO
RenderX

47. Gehder S, Goeldner M. Structural benefits on prescription? Exploring potentials and challenges of the pSVV concept in the approval process of digital health applications in Germany. Z Evid Fortbild Qual Gesundhwes 2024 Dec 16 (forthcoming) [FREE Full text] [doi: 10.1016/j.zefq.2024.11.003] [Medline: 39690098]
48. Gabler S. Schneeballverfahren und verwandte Stichprobendesigns. ZUMA Nachrichten 1992;16(31):47-69. [doi: 10.1007/978-3-322-89527-1_4]
49. Qureshi HA, Ünlü Z. Beyond the paradigm conflicts: a four-step coding instrument for grounded theory. Int J Qual Methods 2020 Jun 02;19. [doi: 10.1177/1609406920928188]
50. Hekkert M, Suurs R, Negro S, Kuhlmann S, Smits R. Functions of innovation systems: a new approach for analysing technological change. In: Foxon T, Köhler J, Oughton C, editors. Technological Forecasting and Social Change. Cheltenham, UK: Edward Elgar Publishing; May 2007:413-432.
51. Bergek A, Jacobsson S, Carlsson B, Lindmark S, Rickne A. Analyzing the functional dynamics of technological innovation systems: a scheme of analysis. Res Policy 2008 Apr;37(3):407-429. [doi: 10.1016/j.respol.2007.12.003]
52. Hekkert MP, Janssen MJ, Wesseling JH, Negro SO. Mission-oriented innovation systems. Environ Innov Soc Transitions 2020 Mar;34:76-79. [doi: 10.1016/j.eist.2019.11.011]
53. Wangler J, Jansky M. What potential and added value do DiGA offer for primary care?-Results of a survey of general practitioners in Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022 Dec;65(12):1334-1343 [FREE Full text] [doi: 10.1007/s00103-022-03608-w] [Medline: 36269336]
54. Mäder M, Timpel P, Schönfelder T, Militzer-Horstmann C, Scheibe S, Heinrich R, et al. Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: an analysis. BMC Health Serv Res 2023 Apr 17;23(1):369 [FREE Full text] [doi: 10.1186/s12913-023-09287-w] [Medline: 37069592]
55. Lantzsch H, Eckhardt H, Campione A, Busse R, Henschke C. Digital health applications and the fast-track pathway to public health coverage in Germany: challenges and opportunities based on first results. BMC Health Serv Res 2022 Sep 21;22(1):1182 [FREE Full text] [doi: 10.1186/s12913-022-08500-6] [Medline: 36131288]
56. Schlieter H, Kählig M, Hickmann E, Fürstenau D, Sunyaev A, Richter P, et al. Digital health applications (DiGA) in the area of tension between progress and criticism : discussion paper from the "digital health" specialist group of the German Informatics Society. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Jan;67(1):107-114 [FREE Full text] [doi: 10.1007/s00103-023-03804-2] [Medline: 38086924]
57. Stern AD. Innovation under regulatory uncertainty: evidence from medical technology. J Public Econ 2017 Jan;145:181-200 [FREE Full text] [doi: 10.1016/j.jpubeco.2016.11.010] [Medline: 28652646]
58. Omachonu VK, Einspruch NG. Innovation in healthcare delivery systems: a conceptual framework. Innov J 2010;15(1).
59. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
60. Torous J, Stern AD, Bourgeois FT. Regulatory considerations to keep pace with innovation in digital health products. NPJ Digit Med 2022 Aug 19;5(1):121 [FREE Full text] [doi: 10.1038/s41746-022-00668-9] [Medline: 35986056]
61. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020 Aug 27;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
62. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54 [FREE Full text] [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
63. Bertha A, Alaj R, Bousnina I, Doyle MK, Friend D, Kalamegham R, et al. Incorporating digitally derived endpoints within clinical development programs by leveraging prior work. NPJ Digit Med 2023 Aug 10;6(1):139 [FREE Full text] [doi: 10.1038/s41746-023-00886-9] [Medline: 37563201]
64. Giunti G, Guisado-Fernandez E, Belani H, Lacalle-Remigio JR. Mapping the access of future doctors to health information technologies training in the European Union: cross-sectional descriptive study. J Med Internet Res 2019 Aug 12;21(8):e14086 [FREE Full text] [doi: 10.2196/14086] [Medline: 31407668]
65. Wilts H, Bringezu S, Bleischwitz R, Lucas R, Wittmer D. Challenges of metal recycling and an international covenant as possible instrument of a globally extended producer responsibility. Waste Manage Res 2011 Jul 19;29(9):902-910. [doi: 10.1177/0734242x11415311]
66. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digit Health 2020 Apr;2(4):e163-e165 [FREE Full text] [doi: 10.1016/S2589-7500(20)30057-1] [Medline: 33328077]
67. Facey K, Lo Scalzo A, Ørtenblad L. EUnetHTA HTA core model: patient and social aspects. EUnetHTA HTA Core Model v3.0. Maastricht, The Netherlands: European Union; 2016. URL: https://www.eunethta.eu/hta-core-model-3-0/ [accessed 2024-03-01]
68. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206 [FREE Full text] [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
69. Brückner S, Brightwell C, Gilbert S. FDA launches health care at home initiative to drive equity in digital medical care. NPJ Digit Med 2024 Aug 21;7(1):204 [FREE Full text] [doi: 10.1038/s41746-024-01198-2] [Medline: 39169161]

XSL•FO
RenderX

70. Donabedian A. The quality of care: how can it be assessed? JAMA 1988 Sep 23;260(12):1743-1748 [FREE Full text] [doi: 10.1001/jama.1988.03410120089033]
71. Hamburg MA. Innovation, regulation, and the FDA. New Engl J Med 2010 Dec 02;363(23):2228-2232. [doi: 10.1056/nejmsa1007467]
72. Diao JA, Venkatesh KP, Raza MM, Kvedar JC. Multinational landscape of health app policy: toward regulatory consensus on digital health. NPJ Digit Med 2022 May 11;5(1):61 [FREE Full text] [doi: 10.1038/s41746-022-00604-x] [Medline: 35545663]
73. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
74. Brönneke JB, Herr A, Reif S, Stern AD. Dynamic HTA for digital health solutions: opportunities and challenges for patient-centered evaluation. Int J Technol Assess Health Care 2023 Nov 17;39(1):e72. [doi: 10.1017/s0266462323002726]
75. Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of Digital Health Technologies in neurology trials. NPJ Digit Med 2023 Feb 10;6(1):23 [FREE Full text] [doi: 10.1038/s41746-023-00767-1] [Medline: 36765123]
76. Gordon WJ, Coravos AR, Stern AD. Ushering in safe, effective, secure, and ethical medicine in the digital era. NPJ Digit Med 2021 Mar 25;4(1):56 [FREE Full text] [doi: 10.1038/s41746-021-00424-5] [Medline: 33767377]
77. Rouleau G, Wu K, Ramamoorthi K, Boxall C, Liu RH, Maloney S, et al. Mapping theories, models, and frameworks to evaluate digital health interventions: scoping review. J Med Internet Res 2024 Feb 05;26:e51098 [FREE Full text] [doi: 10.2196/51098] [Medline: 38315515]
78. Kvedar JC, Mittermaier M, Pritzker J. The industry impact of the American Medical Association's Digital Medicine Payment Advisory Group (DMPAG). NPJ Digit Med 2022 Dec 24;5(1):193 [FREE Full text] [doi: 10.1038/s41746-022-00743-1] [Medline: 36566288]
79. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion paper: “so what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manage 2023 Aug;71:102642. [doi: 10.1016/j.ijinfomgt.2023.102642]

Abbreviations

BfArM: Federal Institute for Drugs and Medical Devices (German: Bundesinstitut für Arzneimittel und Medizinprodukte)
CE: Conformité Européenne
COREQ-32: Consolidated Criteria for Reporting Qualitative Research-32
DiGA: digital health application (German: Digitale Gesundheitsanwendung)
DiGAV: digital health application ordinance (German: Digitiale Gesundheitsanwendungen-Verordnung)
DiPA: digital nursing care application
DVG: Digital Healthcare Act
PECAN: Prise en Charge Anticipée Numérique
pSVV: patient-relevant structural and procedural improvement (German: Patientenrelevante Struktur- und Verfahrensverbesserung)

Edited by A Mavragani; submitted 10.09.24; peer-reviewed by S Greene, K Mouloudj; comments to author 04.11.24; revised version received 08.11.24; accepted 13.11.24; published 06.01.25.

Please cite as:
Gehder S, Goeldner M
Unpacking Performance Factors of Innovation Systems and Studying Germany’s Attempt to Foster the Role of the Patient Through a Market Access Pathway for Digital Health Applications (DiGAs): Exploratory Mixed Methods Study
J Med Internet Res 2025;27:e66356
URL: https://www.jmir.org/2025/1/e66356
doi:10.2196/66356
PMID:

©Sara Gehder, Moritz Goeldner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine

Kuo Zhang[1*], MD, PhD; Xiangbin Meng[2*], MD, PhD; Xiangyu Yan[3*], PhD; Jiaming Ji[4], PhD; Jingqian Liu[5], PhD; Hua Xu[6], PhD; Heng Zhang[7], MSc; Da Liu[8], MD, PhD; Jingjia Wang[9], MD, PhD; Xuliang Wang[9], MD, PhD; Jun Gao[9], MD, PhD; Yuan-geng-shuo Wang[9], MD, PhD; Chunli Shao[9], MD, PhD; Wenyao Wang[9], MD, PhD; Jiarong Li[10], PhD; Ming-Qi Zheng[8*], MD, PhD; Yaodong Yang[4*], PhD; Yi-Da Tang[9], MD, PhD

[1]Department of Cardiology, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]Pengcheng Laboratory, Shenzhen, Guangdong, China

[3]School of Disaster and Emergency Medicine, Tianjin University, Tianjin, China

[4]Institute for Artificial Intelligence, Peking University, Beijing, China

[5]China Telecom, Beijing, China

[6]Division of Emerging Interdisciplinary Areas, Hong Kong University of Science and Technology, Hong Kong, China (Hong Kong)

[7]Institute for Artificial Intelligence, Hefei University of Technology, Hefei, Anhui, China

[8]Department of Cardiology, the First Hospital of Hebei Medical University, Graduate School of Hebei Medical University, Shijiazhuang, Hebei, China

[9]Department of Cardiology and Institute of Vascular Medicine, Key Laboratory of Molecular Cardiovascular Science, Ministry of Education, Peking University Third Hospital, Beijing, China

[10]Henley Business School, University of Reading, RG6 6UD, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Yi-Da Tang, MD, PhD
Department of Cardiology and Institute of Vascular Medicine
Key Laboratory of Molecular Cardiovascular Science, Ministry of Education
Peking University Third Hospital
49 North Garden Road
Beijing, 100191
China
Phone: 86 88396171
Email: tangyida@bjmu.edu.cn

Abstract

Large language models (LLMs) are rapidly advancing medical artificial intelligence, offering revolutionary changes in health care. These models excel in natural language processing (NLP), enhancing clinical support, diagnosis, treatment, and medical research. Breakthroughs, like GPT-4 and BERT (Bidirectional Encoder Representations from Transformer), demonstrate LLMs' evolution through improved computing power and data. However, their high hardware requirements are being addressed through technological advancements. LLMs are unique in processing multimodal data, thereby improving emergency, elder care, and digital medical procedures. Challenges include ensuring their empirical reliability, addressing ethical and societal implications, especially data privacy, and mitigating biases while maintaining privacy and accountability. The paper emphasizes the need for human-centric, bias-free LLMs for personalized medicine and advocates for equitable development and access. LLMs hold promise for transformative impacts in health care.

(J Med Internet Res 2025;27:e59069) doi:10.2196/59069

KEYWORDS

large language models; LLMs; digital health; medical diagnosis; treatment; multimodal data integration; technological fairness; artificial intelligence; AI; natural language processing; NLP

XSL•FO
RenderX

Introduction

Recent advancements in artificial intelligence (AI) have catalyzed the development and significant breakthroughs of large language models (LLMs), placing them at the forefront of AI research [1-4]. LLMs are deep learning models that generate human-like text by predicting the next word in a sequence based on statistical patterns learned from vast text data. These models leverage deep learning algorithms to interpret and generate natural language, using extensive corpus data to enhance pretrained language models, a cornerstone of natural language processing (NLP) [5,6]. Characterized by their immense scale, these models often consist of hundreds of millions to billions of parameters and are trained on vast textual datasets [7,8]. Their ability to efficiently process natural language data with minimal human intervention, capturing intricate grammatical structures, lexical nuances, and semantic contexts, is noteworthy. Globally recognized LLMs include the ChatGPT series, BERT (Bidirectional Encoder Representations from Transformer), PaLM, LaMDA, and Meta's Llama series, with China contributing models such as Baidu's "Wenxin Yiyan," 360's LLM, Alibaba's "Tongyi Qianwen," and SenseTime's LLM [9]. The evolution of LLMs represents over 7 years of relentless technological innovation and research, marking a significant milestone in AI development since the inception of the Turing machine.

LLMs primarily function to comprehend, generate, and interact through language. In NLP tasks, such as text classification, named entity recognition, and sentiment analysis, their proficiency is unparalleled [10-12]. Beyond these applications, LLMs are expanding their influence. In mathematics, they assist in solving complex problems and contributing to mathematical proofs [13]. In software development, their capabilities include automatic code generation, debugging assistance, and complex algorithm explanation [14]. Intriguingly, LLMs are venturing into artistic creation, exhibiting talent in generating poetry, stories, and music [15,16].

In the medical domain, LLMs are poised to revolutionize clinical decision support. They can assist health care professionals in diagnosing diseases with enhanced accuracy and speed, provide treatment recommendations, and facilitate the analysis of medical records by processing large volumes of medical data [17-20]. They are instrumental in swiftly navigating vast medical literature, providing health care professionals with essential research, guidelines, and information, thus saving time and grounding medical treatments in current knowledge [21-25]. Additionally, LLMs can interact directly with patients, offer medical consultations, and handle document processing efficiently [26-28]. For example, health care professionals use LLMs to assist in diagnosing diseases by quickly processing and interpreting large volumes of patient data such as electronic health records and imaging results. Clinicians also leverage LLMs for treatment planning, where the models suggest potential treatment options based on the latest medical guidelines and patient-specific data. Moreover, LLMs are used in streamlining administrative tasks, such as generating and managing medical documentation, allowing clinicians to spend more time with their patients. Their role in drug research and development is also emerging, aiding in new drug discoveries through detailed analysis of chemical and biological data [29,30]. As such, LLMs are reshaping research methodologies and applications across various fields, particularly in medicine, equipping doctors with advanced tools for more accurate and efficient diagnosis and treatment, while offering patients more convenient and effective medical services. The potential for broader applications of LLMs in the medical field is vast, and there is a strong rationale to expect their significant impact on future health care advancements (Figure 1).

XSL•FO
RenderX

Figure 1. Timeline of mainstream LLMs commercially available to the public. The technological evolution of LLMs, highlighting several key technologies and models. It includes RNNs and LSTMs from the 1990s, Google's Transformer model introduced in 2017, Google's BERT model released in 2018, and the GPT series by OpenAI. Specific emphasis is placed on three major milestones: the first open-source LLM—GPT-2, and the first widely acclaimed LLM—GPT-3. These developments signify major advancements in LLMs within the field of natural language processing. BERT: Bidirectional Encoder Representations from Transformers; LLM: large language model; LSTM: long short-term memory network; RNN: recurrent neural network.

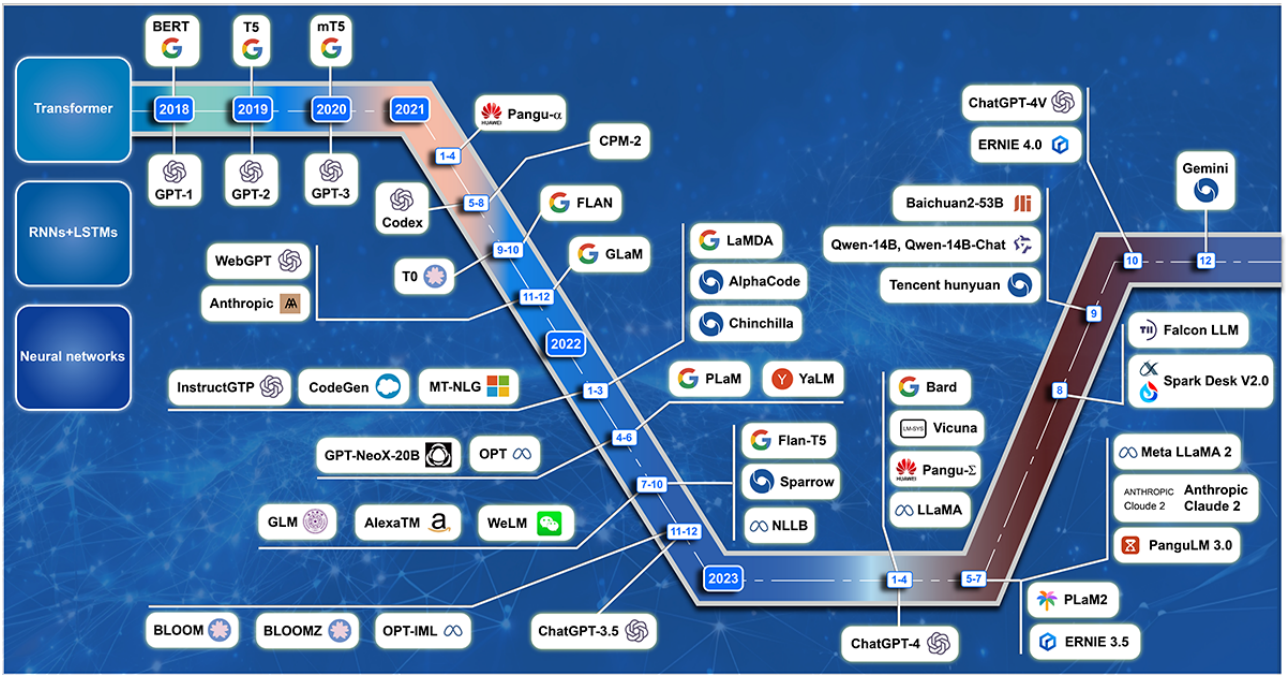

LLM Technical Background and Hardware Infrastructure

The evolution of LLMs, like OpenAI's GPT-3 and Google's BERT, has been monumental, driven by advancements in AI chip computing power and large, high-quality datasets [31]. The Transformer model, introduced by Google in 2016, underpins this progress, predicting words in sentences based on statistical correlations [32,33]. Notably, GPT-3 in 2020 showcased the significance of model size and data quality.

The operation and training of LLMs, such as ChatGPT, require substantial hardware infrastructure [34]. This includes graphics processing units (GPUs) or tensor processing units (TPUs) with thousands of cores, extensive RAM (several terabytes), over 48 GB of VRAM on GPUs, high-performance solid-state drives, and fast, low-latency networks (10 to 100 Gbps) [35,36]. Effective cooling systems and reliable power supplies are also essential. Compatibility with software frameworks, like TensorFlow and PyTorch, is necessary for optimizing training and deployment. The training of GPT-3, for instance, costs around US $1.4 million, and operational costs for models, like ChatGPT, can reach up to US $700,000 daily, with significant energy consumption.

Future technology advancements are expected to reduce the costs and improve the efficiency of LLMs. Progress in GPU and TPU technologies, along with hardware tailored for LLM training, will drive efficiency. Compact model structures through knowledge distillation, model pruning, transfer learning, energy-efficient practices, distributed training, and edge computing are anticipated. Semisupervised and self-supervised learning methods will also play a role in training models with fewer labeled datasets [37,38]. ChatGPT's recent updates showcase improvements in response speed, handling complex queries, multimodal functionality, global language support, and enhanced privacy and security measures [39].

In health care, deploying large-scale medical models faces unique challenges due to data security and privacy concerns. Hospitals typically have CPUs for general computing, with limited access to GPUs. Medical LLMs, generally smaller than general-purpose LLMs, still require substantial investment in operational hardware [40,41]. For instance, a model with 13 billion parameters might cost under US $138,000 while larger models for entire hospitals may require advanced GPU solutions costing around RMB 10 million. Effective deployment demands careful consideration of model scale, computational resources, data security, and cost control (Figure 2).

XSL•FO
RenderX

Figure 2. The architectural designs of LLMs: a study of self-attention mechanisms and structural variations. The image depicts the hardware infrastructure for LLMs and their implementation in the BERT and GPT models. On the left, there is a network diagram showing servers and computing devices needed to run these models, labeled with hardware such as TPU and GPU. On the right, the structure of BERT and GPT is compared in detail, including positional encoding, self-attention mechanisms, feed-forward networks, addition and normalization layers, and the computation of output probabilities. Although these models have different approaches to processing text, both are large neural network models based on deep learning and self-attention mechanisms. BERT: Bidirectional Encoder Representations from Transformers. GPU: graphics processing unit; LLM: large language model. TPU: tensor processing unit.

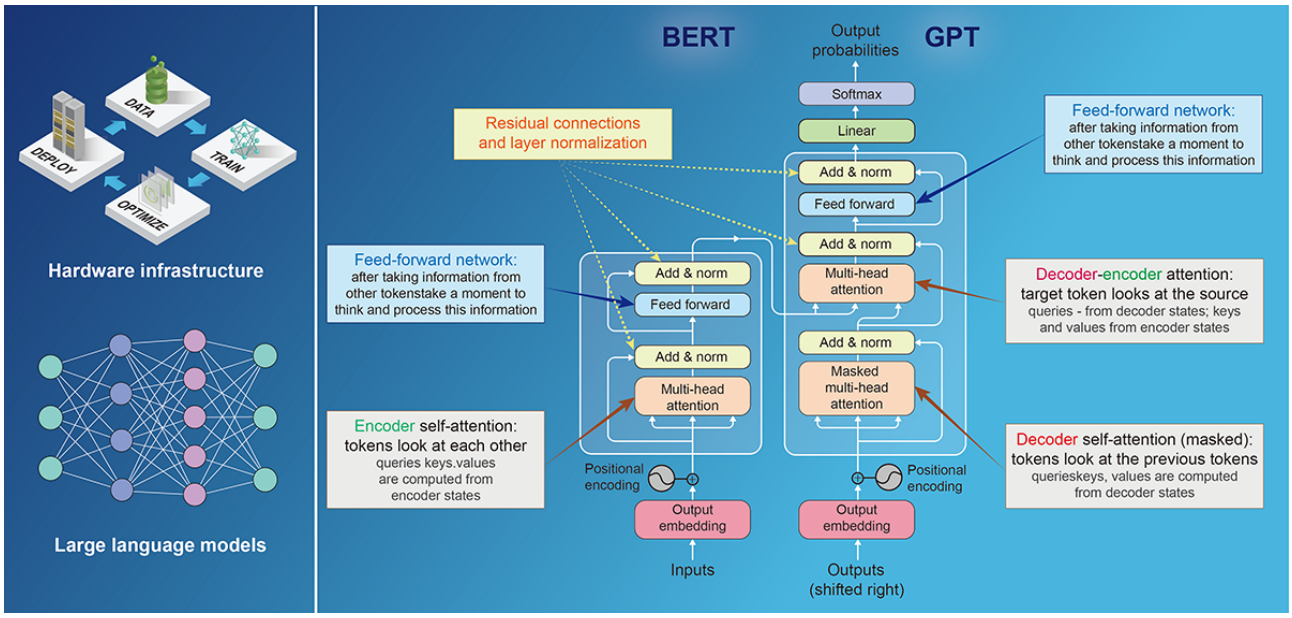

Advancing the Integration of LLMs in Health Care: The Imperative for Evidence-Based Research and Collaborative Evaluation

Overview

In the contemporary health care landscape, the paradigm of evidence-based medicine is instrumental in shaping medical decision-making processes. This methodology integrates top-tier research evidence with clinical expertise and aligns it with patient values and expectations, thereby informing patient care decisions. Evidence-based medicine ensures that medical interventions are grounded in scientific evidence rather than solely relying on a physician's experience or intuition, enhancing patient safety and the efficacy of treatments [42-45].

The integration of LLMs into the medical field introduces a significant challenge: the current scarcity of evidence-based medical research concerning the application of LLMs in health care settings [46]. Although LLMs have shown remarkable efficacy in various sectors, the unique context of medicine, with its direct implications for human life and health, necessitates a cautious approach to the introduction of untested technologies or methods into clinical practice [47]. Despite their robust data processing capabilities, LLMs present a potential risk for prediction errors in clinical environments. The medical domain, with its complex interplay of biology, physiology, and pathology, might be challenging for machine learning models to fully encapsulate, especially considering the intricacies and variability inherent in medical data [48]. Furthermore, the realm of medical decision-making often requires a high level of expertise and experience, aspects that may not be entirely replicable by LLMs. The consequences of medical decisions far surpass those in other sectors, where a misdiagnosis or incorrect treatment recommendation could directly jeopardize a patient's life. Hence, it is imperative to back any new technological innovation, including LLMs, with solid scientific evidence before they are implemented in medical practice.

Currently, empirical studies examining the application of LLMs in the medical field are limited. This scarcity of research implies an inability to definitively assess the accuracy, reliability, and safety of LLMs within a health care context. Model reliability refers to the consistency and dependability of a model's outputs across different datasets or under varying conditions. In medical applications, the reliability of LLMs is critical, as it directly affects the accuracy of diagnoses and treatment recommendations, where any inconsistency could have serious consequences for patient care. To comprehensively understand the potential benefits and risks associated with LLMs in medicine, a more robust body of clinical research is required. This research should encompass randomized controlled trials, observational studies, and extensive collaborative research, which are critical to evaluating the clinical utility of LLMs accurately [49].

To accelerate the empirical evaluation of LLMs in the medical field, fostering collaboration between medical institutions, research organizations, and technology companies is essential. This interdisciplinary collaboration ensures the comprehensiveness and quality of the research, facilitating the rapid advancement and application of LLM technologies. To enhance the transparency, trustworthiness, and ethical application of LLMs in health care, it is crucial to address the societal implications, particularly in terms of data privacy. Publicizing research findings and fostering interdisciplinary

XSL•FO
RenderX

collaboration among doctors, researchers, and ethicists will be key to ensuring that LLMs are used responsibly and equitably. Furthermore, the integration of robust data privacy measures and adherence to ethical standards must be a priority to prevent potential misuse or unintended consequences that could undermine public trust. Such an approach ensures that LLMs' application in the medical field is underpinned by scientific rigor, is safe, and genuinely benefits both patients and the health care system.

Integrated Application of LLMs in Medical System

As we witness ongoing advancements in medical technology, the integration of LLMs with other tools and platforms within health care systems becomes increasingly crucial [50]. This fusion provides health care professionals with powerful tools to process, analyze, and effectively use vast amounts of health care data [23,51-54]. The integration of LLMs, such as ChatGPT, into medical systems has the potential to drive transformative progress in health care delivery. First, LLMs can potentially enhance diagnostic accuracy and clinical decision-making by analyzing comprehensive medical data to identify relevant information and suggest potential diagnoses based on presented symptoms [55-57]. Second, their proficiency in text processing and generation assists medical professionals in efficiently summarizing medical literature, facilitating research, and improving communication between health care providers and patients [58-61]. The rapid adoption of readily available LLMs, such as ChatGPT, within the medical community, signifies recognition of their potential to transform health care delivery [62-66].

However, the application of LLMs in clinical settings is not without challenges [67]. A primary concern is the generalizability of these models. Although LLMs have shown outstanding performance in numerous standard tasks, the complexity and diversity of the medical field suggest that these models may be susceptible to prediction errors in real clinical scenarios. Such errors can have serious implications, particularly when they influence critical health and life decisions. Additionally, the medical field encompasses a vast array of domain-specific knowledge that might exceed the training scope of LLMs, potentially leading to misunderstandings in complex medical scenarios.

Despite these challenges, the potential benefits and impact of LLMs in health care are considerable. LLMs can notably enhance the efficiency of medical workflows by automating routine processes such as appointment scheduling, diagnosis, and report generation [68]. Their data-driven recommendations provide powerful decision support to doctors, assisting them in making more accurate and timely decisions. Current digital health workflows often burden physicians with extensive data entry, querying, and management tasks, leading to information overload and fatigue. LLMs can alleviate these burdens by automating these tasks, thereby saving valuable time for health care providers. Moreover, by analyzing and integrating patients' medical data, LLMs can offer tailored diagnoses and treatment recommendations, improving the overall quality of health care delivery. LLMs also play a crucial role in enhancing doctor-patient interactions. Leveraging NLP technology, they can better comprehend patients' needs and concerns, offering more personalized medical advice [69]. This not only boosts patient satisfaction but also enhances the overall effectiveness of medical services. The potential of LLMs to optimize digital health care workflows is undeniable. With further technological advancements and empirical research, LLMs are expected to play an increasingly significant role in the future of health care (Figure 3).

Figure 3. Integration of LLMs in health care systems across different scales. LLMs can assist in monitoring and analyzing patient health records, treatment plans, and laboratory results at the individual bed level while managing care schedules and facilitating doctor-patient communication. At the hospital level, LLM helps manage patient data, operational logistics, staff scheduling, and resource allocation, while analyzing epidemic trends and hospital infection rates. At the community level, LLM can be used to predict public health crises, manage vaccination campaigns, coordinate community health initiatives, and analyze population health data to improve health policy. LLM: large language model.

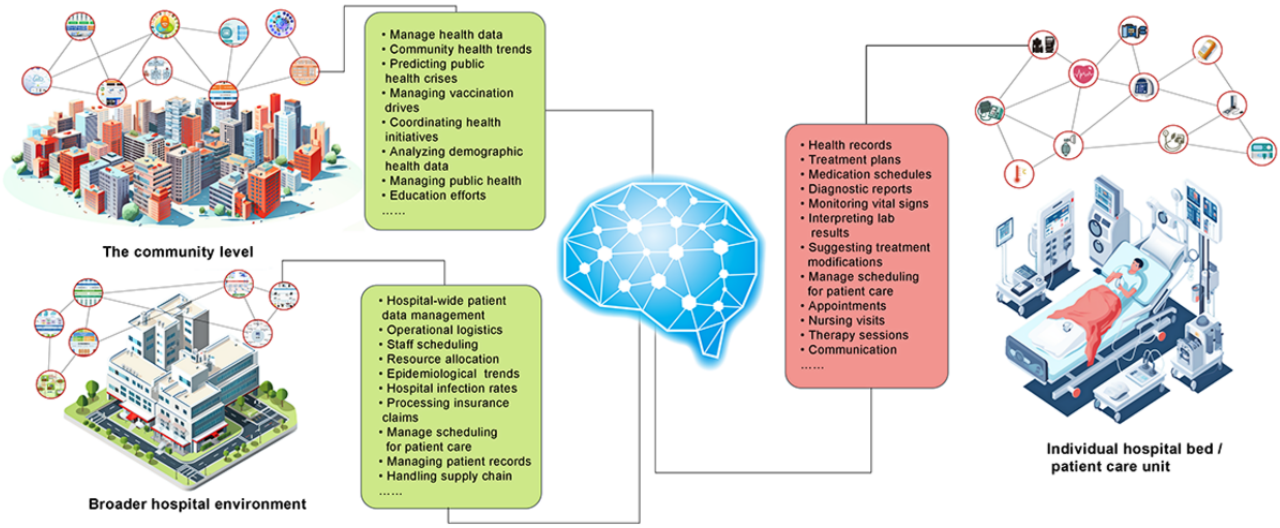

XSL•FO
RenderX

Multimodal LLMs in Real-World Medical Scenarios

The advent of multimodal LLMs is bringing about a paradigm shift in the medical field by offering the capability to process and generate diverse data types such as text, images, sounds, and videos. This integration of multiple data types enables LLMs to provide more comprehensive and accurate predictions, thereby unlocking unprecedented potential [70-73]. To understand their role, it is essential to define what multimodal LLMs entail. Multimodal LLMs excel in processing, interpreting, and generating a wide array of data types, which significantly enhances their predictive capabilities. For instance, in the medical field, combining textual data from patient records with imaging data from magnetic resonance imaging (MRI), computed tomography scans, and x-rays allows these models to provide more nuanced and precise diagnoses. Additionally, integrating audio data from patient interviews or video data from medical procedures can further enrich the model's understanding, leading to more accurate and personalized treatment recommendations. By leveraging the strengths of various data types, multimodal LLMs can offer a holistic view of a patient's condition, which is often crucial for complex diagnoses and treatment planning.

The utility of LLMs is increasingly becoming a focal point in medical imaging [74-76]. For instance, when a patient undergoes an MRI or computed tomography scan, an LLM can swiftly analyze and integrate the image data with the patient's textual medical records, thereby providing more comprehensive and detailed diagnostic insights. Additionally, LLMs have the capability to automatically identify and highlight crucial areas in medical images, thus providing clinicians with clear references that aid in identifying potential issues [77]. Moreover, LLMs can generate automated image reports, offering initial interpretations and treatment suggestions based on the analyzed image data, significantly boosting the efficiency and accuracy of medical diagnoses and treatments.

Multimodal LLMs are revolutionizing the field of telemedicine, transforming the dynamics of doctor-patient interactions [55,78]. For instance, LLMs have been successfully integrated into MRI analysis, where they can rapidly interpret imaging data and provide diagnostic recommendations. This has significantly reduced the time required for diagnosis and improved accuracy. However, the use of LLMs is not without its challenges. A notable example is Google BARD, which recently demonstrated racial bias in patient diagnosis, disproportionately affecting minority groups. This case highlights the dual-edged nature of LLMs in health care—they offer substantial benefits in efficiency and accuracy, yet they also pose significant risks if not properly validated and monitored for biases. Furthermore, the integration of LLMs with smart sensors and devices enables the continuous monitoring of patient's physiological data, such as heart rate and blood pressure, facilitating early detection and intervention for any health anomalies, thus significantly bolstering patient health management.

In summary, multimodal LLMs offer a novel and efficacious approach to diagnosis, treatment, and health care management. Their robust capabilities in data processing and integration allow medical professionals to deliver more precise and efficient services to patients. At the same time, these models enable patients to access medical advice and care with greater convenience. As these technologies continue to evolve and improve, their significance and impact in the medical field are expected to grow exponentially (Figure 4).

XSL•FO
RenderX

Figure 4. The importance of multimodal large language models in medical applications. The central heart represents the cardiac health status of the human body. The surrounding circular icons depict various cardiac conditions including coronary artery disease, hypertension, arrhythmia, heart failure, valvular heart disease, cardiomyopathy, and congenital heart defects. These conditions are detected and analyzed through different medical imaging and diagnostic technologies such as electrocardiography, heart sounds, echocardiogram, coronary angiography, cardiac MRI, and nuclear cardiology. The results from these diagnostics are processed by an AI system to determine the type and severity of cardiac disease, assisting physicians in formulating treatment plans. MRI: magnetic resonance imaging.

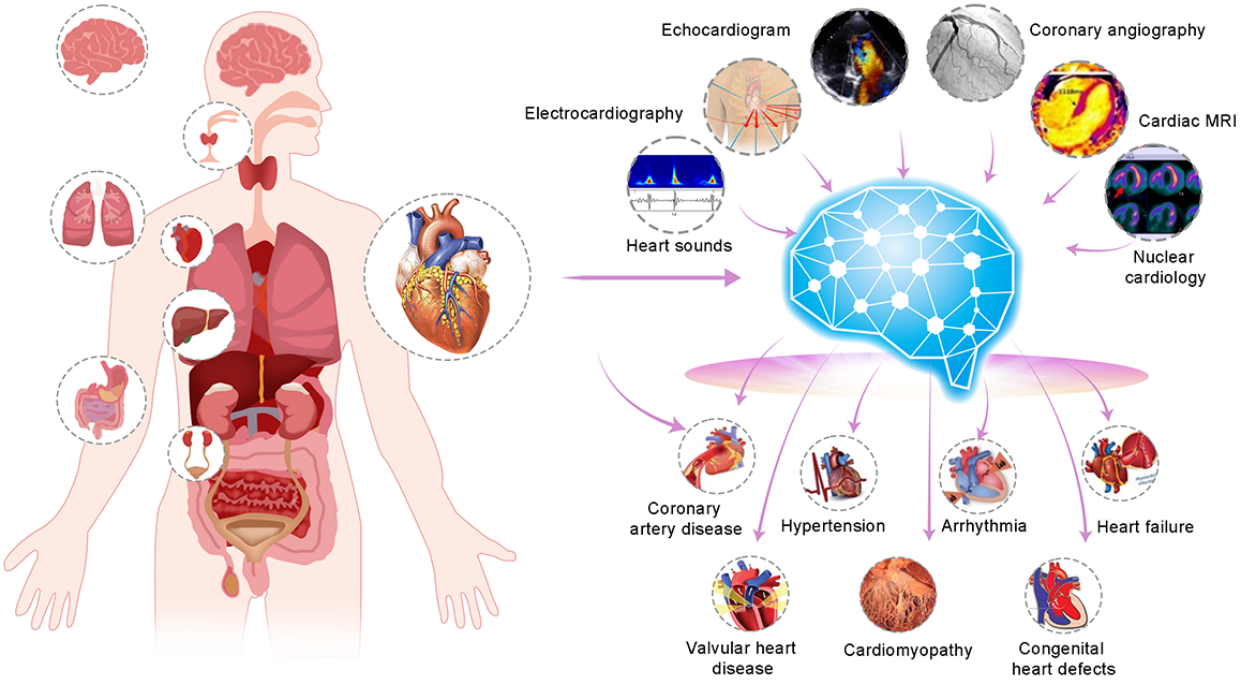

The Key Role of LLMs in Medical Research

In the field of fundamental medical research, the capabilities of LLMs in AI are being increasingly recognized [79-82]. LLMs can swiftly retrieve and organize crucial information from vast biomedical literature, providing researchers with an efficient tool to access and synthesize the latest research findings on specific drugs, diseases, or genes [83]. In drug discovery, LLMs can predict the activity, toxicity, and pharmacokinetic properties of new compounds, facilitating early-stage drug screening [84]. These predictions not only save time but also facilitate the early-stage screening of potential drug molecules. LLMs can use existing literature and databases to predict the potential functions of newly discovered genes, a crucial aspect of genomic research, given the daily discovery and study of new genes. While protein structure prediction depends primarily on specialized models, such as AlphaFold, LLMs can enhance these models by supplying pertinent information from literature, thereby increasing prediction accuracy. In epidemiological research, LLMs can aid researchers in tracking and predicting disease spread by analyzing social media and other web-based text data, offering data support for public health decision-making. Finally, in bioinformatics applications, LLMs can assist researchers in predicting patterns, functional domains, and similarities to known biological sequences. Despite their extensive applications in biomedicine, LLMs cannot entirely replace laboratory experiments or in-depth biomedical expertise. Instead, they should be considered powerful supplementary tools, rather than replacements.

LLMs play a pivotal role in clinical research. They aid doctors and researchers by extracting essential information from medical records, and by organizing and categorizing data for easier analysis and application. For instance, they can expedite the selection of suitable patients for enrolment, thereby enhancing the design and implementation of clinical trials. In the role of a clinical research coordinator, these models assist with data entry, verification, and analysis, thereby accelerating the clinical research process. Through automated data processing and real-time analysis, LLMs can ensure data accuracy and completeness, while also reducing the workload of clinical research coordinators. This, in turn, speeds up the clinical research process and enhances research quality.

Although LLMs have revolutionized biomedicine by simplifying literature searches, aiding drug discovery, annotating gene functions, and supporting epidemiological studies, they experience certain drawbacks. Their ability to swiftly parse large datasets and make predictions may be counterbalanced by potential limitations in real-world validation [85]. For example, while they can predict a drug molecule's properties, the actual biological response may vary. Similarly, despite gene function predictions being well grounded, they may not fully encapsulate the breadth of gene interactions. Moreover, using LLMs to analyze epidemiological trends without correlating them to underlying data could misdirect public health interventions. Therefore, while LLMs are undeniably beneficial to biomedicine, it is essential to adopt a balanced approach, combining their computational prowess with rigorous experimental validation and expert review, to fully harness their potential without sacrificing scientific rigor (Figure 5).

XSL•FO
RenderX

Figure 5. The crucial role of LLMs in medical science: bridging basic research and clinical trials. This illustration highlights the versatile roles of LLMs in medical research. LLMs analyze medical texts to uncover trends and inform research directions, facilitate hypothesis generation, and enhance clinical trial designs. They personalize medicine through data-driven treatment plans and use predictive modeling to inform clinical trial outcomes. LLMs also streamline research by integrating data and maintaining regulatory compliance. They assist in medical communication and education and evaluate the societal impact of clinical research. LLM: large language model.

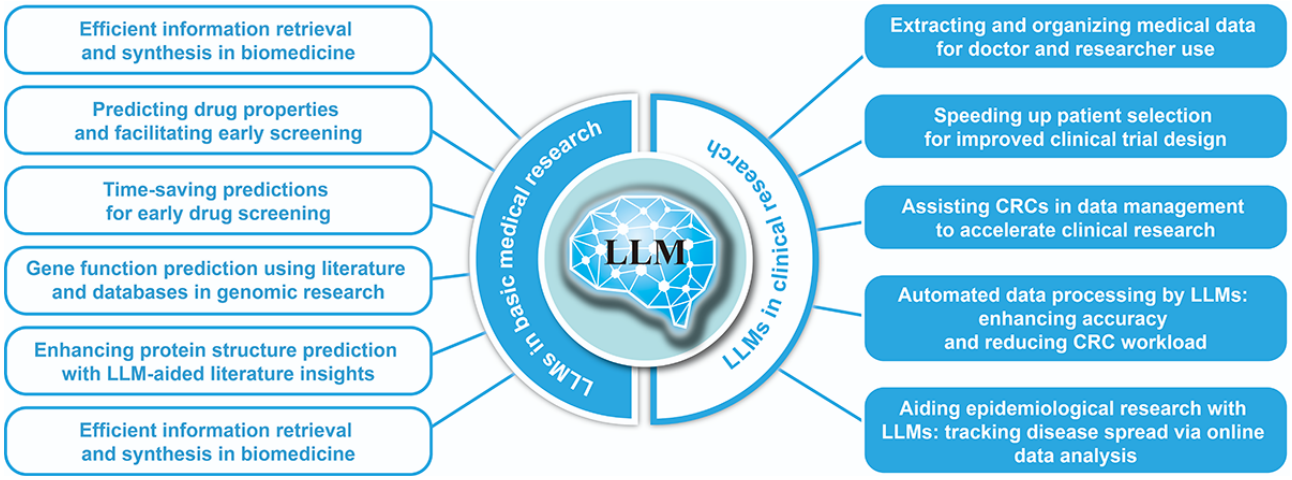

Great Challenges of LLMs in Medical Scenarios and Feasible Roadmap

The integration of technology in health care invariably brings a mix of anticipation and challenges, particularly given its direct impact on human life and health. As a leading exemplar of current AI technology, LLMs present a complex array of opportunities and challenges in the medical field, warranting thorough exploration and discussion [86-88].

Handling medical data, some of the most private and sensitive information about individuals, is a significant challenge for LLMs. As LLMs are increasingly integrated into health care, ethical considerations surrounding data privacy and societal impact must be prioritized during their development and deployment. The key lies not only in using this data to enhance medical efficiency but also in implementing robust data protection frameworks to prevent misuse, leakage, and unauthorized access. Furthermore, addressing these ethical challenges requires ongoing dialogue among technologists, health care professionals, policy makers, and the public to ensure that LLM deployment aligns with societal values and legal standards [80,89]. A potential technical solution involves anonymizing patient data, ensuring that neither processing nor transmission stages can be linked to specific individuals. Concurrently, medical organizations and technology providers must establish robust data management and access protocols, ensuring clear authorization and purpose for each data access.

Interpretive challenges loom large with LLMs in medicine. Medical decision-making is distinct from other fields due to its complexity and direct implications for patients' lives and health. When LLMs provide diagnostic or treatment suggestions, it is vital that the rationale behind these recommendations is transparent and comprehensible [90-92]. This brings us to the concept of interpretability in machine learning, which refers to the ability to understand and explain how a model makes its decisions. In the context of health care, interpretability is a significant challenge because clinicians must trust and validate the outputs of LLMs, especially when these models influence critical medical decisions [93]. Developing mental models can aid LLMs in presenting their decision-making logic in a manner that is more accessible to human users. Leveraging deep learning and other machine learning technologies, LLMs can extract disease pathophysiological mechanisms from a vast corpus of medical literature and data, providing a scientific basis for their outputs. To further enhance interpretability, LLMs could use visual tools, like graphics and animations, to clarify the logic and evidence underpinning their decisions for both physicians and patients [94,95].

The issue of technical bias and the possibility of generating misleading information or "hallucinations" are inherent challenges in LLMs. In this context, hallucinations refer to instances where LLMs produce outputs that are factually incorrect or misleading, often because the model attempts to generate an answer despite lacking sufficient context or knowledge. These hallucinations can be especially problematic in medical scenarios, where inaccurate information can have severe consequences. The data sources for these models, often anonymized consultation data and digital materials, are not uniform and vary in quality, sometimes containing erroneous samples. Fine-tuning LLMs based on such data may lead to biased or skewed medical recommendations [96,97]. Addressing this requires rigorous data auditing and the establishment of continuous bias-correction mechanisms. To mitigate the risk of hallucinations, knowledge enhancement methods, such as integrating a knowledge retrieval library or search enhancement tools, can be beneficial. The LLM's responses can be cross-referenced with retrieved data to filter out inconsistencies with reality. Another approach involves reinforcement learning based on human feedback, where high-quality feedback is provided to fine-tune and correct model outputs in collaboration with medical experts [98,99].

The potential of AI to create "information cocoons" through personalized content, potentially reinforcing biases, is another critical aspect that needs to be addressed, especially in the medical domain [100]. AI technologies, including LLMs, in medicine require stringent scrutiny and continuous evaluation

XSL•FO
RenderX

to align with the field's unique characteristics and ethical standards. Ensuring privacy protection, eliminating biases and discrimination, and establishing clear accountability are essential. The use of LLMs should be guided by respect for life, aiming to enhance patient well-being and treatment outcomes, without compromising individual interests. A continuous monitoring and evaluation system is crucial for assessing the effectiveness of LLMs and managing potential risks. Regulations should be regularly updated to keep pace with AI advancements, ensuring medical safety and patient rights. By prioritizing safety, fairness, and effectiveness, we can fully leverage LLMs and other AI technologies to facilitate a transformative revolution in medicine, while upholding human values and rights.

In the era of information and intelligence within the medical field, the application of LLMs harbors immense potential [101]. However, the accompanying challenges are equally noteworthy and merit careful consideration. The ongoing discourse should emphasize not only the deeper integration of LLMs into medical practice but also their alignment with both the professional needs of health care providers and the experiential needs of patients [102,103].

Incorporating the theory of mind into LLMs can significantly enhance their utility in the medical field. This concept, which involves understanding others' thoughts, feelings, and intentions, is crucial for fostering trust and empathy within health care interactions. Medicine is not solely a science; it is also an art, deeply influenced by each patient's unique emotional, value-based, and experiential landscape. An AI system endowed with the capability to appreciate and respond to these individual differences can offer more personalized and compassionate medical advice [104,105]. By using the theory of mind, LLMs can gain deeper insights into patients' inherent needs and respond with more attentive and empathetic advice [106-108]. When LLMs can emulate the thoughts and feelings of both doctors and patients, their outputs transcend mere data; they become imbued with empathy and human care, enhancing the patient's treatment experience and fostering stronger trust and communication between doctors and patients. For example, in interactions with terminal patients, LLMs could suggest more compassionate communication strategies, aiding both doctors and patients in navigating these sensitive and complex situations.

LLMs can be synergistically combined with other advanced technologies, such as virtual reality and augmented reality, to transform medical consultations into more immersive and informative experiences. This integration can provide patients with a deeper understanding of their health conditions, empowering them to make more informed decisions regarding their treatment. The evolution of LLMs is also contingent upon the development of efficient and precise algorithms capable of adeptly handling complex medical data, which is essential for accurate and timely medical decision-making. As technology progresses, the use of LLMs in the medical field is expected to become increasingly intelligent, efficient, and personalized, thereby enhancing not only the quality of medical services but also the overall patient experience and driving the evolution and transformation of the health care industry.

In our pursuit of technological progress, we must adhere to a fundamental principle: ensuring that technology is accessible to all. This is particularly pertinent in the context of LLM adoption, where it is crucial not to overlook those who may be marginalized by the technology gap [109,110]. Whether addressing the needs of rural farmers or urban older adults, every individual should have the opportunity to benefit from LLMs. This broad adoption must span various geographical regions and encompass diverse languages and cultural contexts, catering to users speaking English, Chinese, or local dialects [111,112]. Achieving this objective is not solely a technological challenge but also a social imperative. We must ensure that the design and application of LLMs overcome language and cultural barriers, truly reaching and benefiting a diverse global populace. Additionally, addressing technology accessibility issues is vital. For individuals in technologically underserved areas or older adults unfamiliar with new technologies, simpler access methods and more user-friendly interfaces are needed to facilitate effortless use of LLMs.

While the potential of LLMs in health care is significant, realizing this potential requires ongoing research, innovation, and dedication. Continuous efforts are necessary to refine LLM technology continually and ensure its broad adoption across all sectors of society. We firmly believe that with sustained commitment, LLMs will catalyze transformative changes in health care, benefiting society at large. By championing technological inclusivity, we can not only enhance the quality and efficiency of medical services but also promote overall societal health and well-being.

Economic Considerations in the Deployment of LLMs in Health Care

LLMs require significant computational resources for training and maintenance, which translates to substantial financial costs. In the medical domain, these costs can be particularly prohibitive due to the need for specialized data, high levels of accuracy, and continuous updates to ensure model relevance and safety.

Training a state-of-the-art LLM, such as GPT-3, requires access to extensive hardware infrastructure, including thousands of GPUs or TPUs, large amounts of RAM, and high-speed data storage solutions [113,114]. According to estimates, the training cost of models, like GPT-3, can reach up to US $1.4 million, with operational costs amounting to several hundred thousand dollars per day when deployed at scale. In a medical context, where accuracy and reliability are paramount, these costs are even higher due to the additional requirements for data security, privacy, and compliance with health care regulations.

Several studies have documented the economic challenges associated with deploying LLMs in health care [115,116]. For instance, the cost of implementing LLMs in hospital settings, including the necessary infrastructure upgrades, staff training, and ongoing maintenance, has been reported to be a major barrier to widespread adoption. Moreover, the need for regular updates to the models, which involves retraining them with new medical data, adds to the operational expenses [1].

As technology advances, it is expected that the costs associated with LLMs will decrease, making them more accessible to a

XSL•FO
RenderX

broader range of health care providers. The development of more energy-efficient hardware, combined with advances in machine learning techniques, is likely to contribute to this trend. However, until these cost reductions are realized, careful planning and resource allocation will be essential for any institution looking to implement LLMs in their health care practice.

Conclusions

The era of digitalization and informatization underscores the transformative potential of LLMs in medicine. The evolution of this technology signifies a paradigm shift in medical services, offering unique opportunities and challenges to the medical community. LLMs, with their advanced NLP capabilities, have a wide range of applications including emergency triage, older people care, and the enhancement of digital medical workflows. As the diversity of medical data expands, LLMs' ability to process multimodal data will play a crucial role in enabling more precise, personalized medical diagnoses and treatments.

Despite the promising trajectory of LLMs in the medical field, ensuring their safety and effectiveness in clinical practice remains a critical challenge. Currently, the regulation of LLMs in health care is still in its early stages, with several frameworks being developed to address the unique risks and challenges they pose. Regulatory bodies, such as the US Food and Drug Administration, European Medicines Agency, and China's National Medical Products Administration, have begun to formulate guidelines that apply to AI-driven medical devices including LLMs. These guidelines typically focus on the validation of the models through rigorous clinical trials, ensuring that they meet specific safety, efficacy, and ethical standards before they can be deployed in clinical settings. However, the growth potential of LLMs in the medical arena is significant. They can enhance patient experiences through the integration of virtual reality and augmented reality, offer comprehensive medical advice through multimodal research, and humanize doctor-patient interactions using the theory of mind. With ongoing advancements in algorithms and computational power, we anticipate considerable improvements in LLMs' processing speed and accuracy.

However, the path to technological advancement is not always linear. To ensure the benefits of LLMs are accessible to all, it is imperative to promote equitable development and address the digital divide, particularly for economically and technologically disadvantaged regions and groups. This goal requires the collective efforts of health care professionals, computer science experts, government regulatory bodies, patients, and their families. Such a collaborative approach will ensure that the application of LLM technology in the medical field genuinely contributes to the betterment of humanity, significantly enhancing health and well-being.

Acknowledgments

This study was funded by the National Key R&D Program of China (2020YFC2004705), the National Natural Science Foundation of China (81825003, 81900272, 91957123, 82270376, 623B2003), and the Beijing Nova Program (Z201100006820002) from Beijing Municipal Science & Technology Commission. This study was also supported by the S&T Program of Hebei (22377719D and 22377771D), the Key Science and Technology Research Program of Hebei Provincial Health Commission (20230991), the Industry University Research Cooperation Special Project (CXY2024020), the Hebei Province Finance Department Project (LS202214, ZF2024226), the Hebei Provincial Health Commission Project (20190448), the Key Discipline Construction Project of Shanghai Pudong New Area Health Commission (PWZxk2022-20), the Science and Technology Plan Project of Jiangxi Provincial Health Commission ([SK] P220227143), Tianjin University Science and Technology Innovation Leading Talent Cultivation Project (2024XQM-0024), and Tianjin Natural Science Foundation (23JCQNJC01430). This study was also supported by the China Hebei International Joint Research Center for Structural Heart Disease, and the Hebei Province Finance Department Project (LS202101).

Conflicts of Interest

None declared.

References

1. Arora A, Arora A. The promise of large language models in health care. Lancet 2023;401(10377):641. [doi: 10.1016/S0140-6736(23)00216-7] [Medline: 36841609]
2. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating artificial intelligence responses to public health questions. JAMA Netw Open 2023;6(6):e2317517 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.17517] [Medline: 37285160]
3. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
4. Li R, Kumar A, Chen JH. How chatbots and large language model artificial intelligence systems will reshape modern medicine: fountain of creativity or pandora's box? JAMA Intern Med 2023;183(6):596-597. [doi: 10.1001/jamainternmed.2023.1835] [Medline: 37115531]
5. Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. JAMA 2023;330(4):315-316. [doi: 10.1001/jama.2023.9651] [Medline: 37410482]

XSL•FO
RenderX

6. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 2023;330(1):78-80 [FREE Full text] [doi: 10.1001/jama.2023.8288] [Medline: 37318797]
7. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models. ArXiv Preprint posted online on July 12, 2023. [doi: 10.48550/arXiv.2307.06435]
8. Ayoub NF, Lee YJ, Grimm D, Balakrishnan K. Comparison between ChatGPT and Google search as sources of postoperative patient instructions. JAMA Otolaryngol Head Neck Surg 2023;149(6):556-558 [FREE Full text] [doi: 10.1001/jamaoto.2023.0704] [Medline: 37103921]
9. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. ArXiv Preprint posted online on March 31, 2023. [doi: 10.48550/arXiv.2303.18223]
10. Chang K. Natural language processing: recent development and applications. Appl Sci 2023;13(20):11395. [doi: 10.3390/app132011395]
11. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digital Health 2023;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
12. Garcia MB. Using AI tools in writing peer review reports: should academic journals embrace the use of ChatGPT? Ann Biomed Eng 2024;52(2):139-140. [doi: 10.1007/s10439-023-03299-7] [Medline: 37368125]
13. Azerbayev Z, Schoelkopf H, Paster K, Santos MD, McAleer S, Jiang AQ, et al. Llemma: an open language model for mathematics. ArXiv Preprint posted online on October 16, 2023. [doi: 10.48550/arXiv.2310.10631]
14. Xiong W, Guo Y, Chen H. The program testing ability of large language models for code. ArXiv Preprint posted online on October 09, 2023. [doi: 10.48550/arXiv.2310.05727]
15. Yuan A, Coenen A, Reif E, Ippolito D, Assoc CM. Wordcraft: story writing with large language models. 2022 Presented at: Wordcraft Story Writing With Large Language Models. 27th Annual International Conference on Intelligent User Interfaces (ACM IUI); 2022 2022 Mar ; Univ Helsinki, ELECTR NETWORK; 2022; March 22, 2022; Helsinki, Finland p. 841-852. [doi: 10.1145/3490099.3511105]
16. Chu Y, Liu P. Public aversion against ChatGPT in creative fields? Innovation (Camb) 2023;4(4):100449 [FREE Full text] [doi: 10.1016/j.xinn.2023.100449] [Medline: 37485080]
17. Voelker R. The promise and pitfalls of AI in the complex world of diagnosis, treatment, and disease management. JAMA 2023;330(15):1416-1419. [doi: 10.1001/jama.2023.19180] [Medline: 37755919]
18. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
19. He S, Yang F, Zuo JP, Lin ZM. ChatGPT for scientific paper writing-promises and perils. Innovation (Camb) 2023;4(6):100524 [FREE Full text] [doi: 10.1016/j.xinn.2023.100524] [Medline: 38028132]
20. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
21. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023;9:e48291 [FREE Full text] [doi: 10.2196/48291] [Medline: 37261894]
22. Cooper A, Rodman A. AI and medical education—a 21st-century pandora's box. N Engl J Med 2023;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
23. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. JAMA Ophthalmol 2023;141(6):589-597 [FREE Full text] [doi: 10.1001/jamaophthalmol.2023.1144] [Medline: 37103928]
24. Mehrabanian M, Zariat Y. ChatGPT passes anatomy exam. Br Dent J 2023;235(5):295. [doi: 10.1038/s41415-023-6286-7] [Medline: 37684439]
25. Baker S. AI could be an opportunity for research managers. Nature 2023. [doi: 10.1038/d41586-023-03277-y] [Medline: 37857880]
26. Ray PP. Broadening the horizon: a call for extensive exploration of ChatGPT's potential in obstetrics and gynecology. Am J Obstet Gynecol 2023;229(6):706. [doi: 10.1016/j.ajog.2023.07.016] [Medline: 37454961]
27. Kovoor JG, Gupta AK, Bacchi S. ChatGPT: effective writing is succinct. BMJ 2023;381:1125. [doi: 10.1136/bmj.p1125] [Medline: 37220943]
28. Decker H, Trang K, Ramirez J, Colley A, Pierce L, Coleman M, et al. Large language model-based chatbot vs surgeon-generated informed consent documentation for common procedures. JAMA Netw Open 2023;6(10):e2336997 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36997] [Medline: 37812419]
29. Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, et al. Large language models generate functional protein sequences across diverse families. Nat Biotechnol 2023;41(8):1099-1106 [FREE Full text] [doi: 10.1038/s41587-022-01618-2] [Medline: 36702895]
30. Askr H, Elgeldawi E, Ella HA, Elshaier YAMM, Gomaa MM, Hassanien AE. Deep learning in drug discovery: an integrative review and future challenges. Artif Intell Rev 2023;56(7):5975-6037 [FREE Full text] [doi: 10.1007/s10462-022-10306-1] [Medline: 36415536]

XSL•FO
RenderX

31. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digital Med 2022;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
32. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P. Language models are few-shot learners. 2020 Presented at: Proceedings of the 34th International Conference on Neural Information Processing Systems; 2020 December 06; Red Hook, NY, United States p. 1877-1901.
33. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with clip latents. ArXiv Preprint posted online on October 13, 2022. [doi: 10.48550/arXiv.2204.06125]
34. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Technol 2020;21(1):5485-5551 [FREE Full text]
35. Carpenter KA, Altman RB. Using GPT-3 to build a lexicon of drugs of abuse synonyms for social media pharmacovigilance. Biomolecules 2023;13(2):387 [FREE Full text] [doi: 10.3390/biom13020387] [Medline: 36830756]
36. Ali R, Tang OY, Connolly ID, Fridley JS, Shin JH, Sullivan PLZ, et al. Performance of ChatGPT, GPT-4, and Google bard on a neurosurgery oral boards preparation question bank. Neurosurgery 2023;93(5):1090-1098. [doi: 10.1227/neu.0000000000002551] [Medline: 37306460]
37. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. Attention is all you need. In: Advances in Neural Information Processing Systems. United States: Curran Associates, Inc; 2017.
38. Manolitsis I, Feretzakis G, Tzelves L, Kalles D, Katsimperis S, Angelopoulos P, et al. Training ChatGPT models in assisting urologists in daily practice. Stud Health Technol Inform 2023;305:576-579. [doi: 10.3233/SHTI230562] [Medline: 37387096]
39. An J, Ding W, Lin C. ChatGPT: tackle the growing carbon footprint of generative AI. Nature 2023;615(7953):586. [doi: 10.1038/d41586-023-00843-2] [Medline: 36944743]
40. Sharir O, Peleg B, Shoham Y. The cost of training nlp models: a concise overview. ArXiv Preprint posted online on April 19, 2020. [doi: 10.48550/arXiv.2004.08900]
41. Srinivasan V, Gandhi D, Thakker U, Prabhakar R. Training large language models efficiently with sparsity and dataflow. ArXiv Preprint posted online on April 11, 2023. [doi: 10.48550/arXiv.2304.05511]
42. Zhou Z, Wang X, Li X, Liao L. Is ChatGPT an evidence-based doctor? Eur Urol 2023;84(3):355-356. [doi: 10.1016/j.eururo.2023.03.037] [Medline: 37061445]
43. Blum J, Menta AK, Zhao X, Yang VB, Gouda MA, Subbiah V. Pearls and pitfalls of ChatGPT in medical oncology. Trends Cancer 2023;9(10):788-790. [doi: 10.1016/j.trecan.2023.06.007] [Medline: 37407364]
44. Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. JAMA 2023;330(4):313-314. [doi: 10.1001/jama.2023.9630] [Medline: 37410497]
45. Ji J, Qiu T, Chen B, Zhang B, Lou H, Wang K, et al. AI alignment: a comprehensive survey. ArXiv Preprint posted online on October 30, 2023. [doi: 10.48550/arXiv.2310.19852]
46. Ward E, Gross C. Evolving methods to assess chatbot performance in health sciences research. JAMA Intern Med 2023;183(9):1030-1031. [doi: 10.1001/jamainternmed.2023.2567] [Medline: 37459112]
47. Butte AJ. Artificial intelligence-from starting pilots to scalable privilege. JAMA Oncol 2023;9(10):1341-1342. [doi: 10.1001/jamaoncol.2023.2867] [Medline: 37615950]
48. Hu ZY, Han FJ, Yu L, Jiang Y, Cai G. AI-link omnipotent pathological robot: bridging medical meta-universe to real-world diagnosis and therapy. Innovation (Camb) 2023;4(5):100494 [FREE Full text] [doi: 10.1016/j.xinn.2023.100494] [Medline: 37663933]
49. Thirunavukarasu AJ. Large language models will not replace healthcare professionals: curbing popular fears and hype. J R Soc Med 2023;116(5):181-182 [FREE Full text] [doi: 10.1177/01410768231173123] [Medline: 37199678]
50. The Lancet Regional Health-Europe. Embracing generative AI in health care. Lancet Reg Health Eur 2023;30:100677 [FREE Full text] [doi: 10.1016/j.lanepe.2023.100677] [Medline: 37465322]
51. NA. Will ChatGPT transform healthcare? Nat Med 2023;29(3):505-506. [doi: 10.1038/s41591-023-02289-5] [Medline: 36918736]
52. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, et al. Health system-scale language models are all-purpose prediction engines. Nature 2023;619(7969):357-362 [FREE Full text] [doi: 10.1038/s41586-023-06160-y] [Medline: 37286606]
53. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. JAMA 2023;330(9):866-869. [doi: 10.1001/jama.2023.14217] [Medline: 37548965]
54. Kluger N. Potential applications of ChatGPT in dermatology. J Eur Acad Dermatol Venereol 2023;37(7):e941-e942. [doi: 10.1111/jdv.19152] [Medline: 37102458]
55. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis 2023;23(4):405-406. [doi: 10.1016/S1473-3099(23)00113-5] [Medline: 36822213]
56. Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. JAMA 2023;330(4):317-318. [doi: 10.1001/jama.2023.11440] [Medline: 37410477]
57. Jiang K, Zhu M, Bernard G. Few-shot learning for identification of COVID-19 symptoms using generative pre-trained transformer language models. In: Koprinska I, editor. Machine Learning and Principles and Practice of Knowledge Discovery

XSL•FO
RenderX

in Databases. ECML PKDD 2022. Communications in Computer and Information Science, vol 1753. Switzerland: Springer, Cham; 2023.
58. Miller K, Gunn E, Cochran A, Burstein H, Friedberg JW, Wheeler S, et al. Use of large language models and artificial intelligence tools in works submitted to journal of clinical oncology. J Clin Oncol 2023;41(19):3480-3481. [doi: 10.1200/JCO.23.00819] [Medline: 37200609]
59. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
60. Hua HU, Kaakour AH, Rachitskaya A, Srivastava S, Sharma S, Mammo DA. Evaluation and comparison of ophthalmic scientific abstracts and references by current artificial intelligence chatbots. JAMA Ophthalmol 2023;141(9):819-824. [doi: 10.1001/jamaophthalmol.2023.3119] [Medline: 37498609]
61. Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, et al. Use of artificial intelligence chatbots for cancer treatment information. JAMA Oncol 2023;9(10):1459-1462 [FREE Full text] [doi: 10.1001/jamaoncol.2023.2954] [Medline: 37615976]
62. Marchandot B, Matsushita K, Carmona A, Trimaille A, Morel O. ChatGPT: the next frontier in academic writing for cardiologists or a pandora's box of ethical dilemmas. Eur Heart J Open 2023;3(2):oead007 [FREE Full text] [doi: 10.1093/ehjopen/oead007] [Medline: 36915398]
63. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(25):2400. [doi: 10.1056/NEJMc2305286] [Medline: 37342941]
64. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
65. Kataoka Y, So R. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(25):2399. [doi: 10.1056/NEJMc2305286] [Medline: 37342939]
66. Fernandes AC, Souto MEVC. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(25):2399-2400. [doi: 10.1056/NEJMc2305286] [Medline: 37342940]
67. Gottlieb S, Silvis L. How to safely integrate large language models into health care. JAMA Health Forum 2023;4(9):e233909 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.3909] [Medline: 37733359]
68. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA Oncol 2023;9(10):1437-1440. [doi: 10.1001/jamaoncol.2023.2947] [Medline: 37615960]
69. Bernstein IA, Zhang YV, Govil D, Majid I, Chang RT, Sun Y, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. JAMA Netw Open 2023;6(8):e2330320 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.30320] [Medline: 37606922]
70. Kiros R, Salakhutdinov R, Zemel R. Multimodal neural language models. 2014 Presented at: Proceedings of the 31st International Conference on International Conference on Machine Learning—Volume 32; June 21, 2014; Beijing China URL: https://proceedings.mlr.press/v32/kiros14.html
71. Driess D, Xia F, Sajjadi MSM, Lynch C, Chowdhery A, Ichter B, et al. Palm-e: an embodied multimodal language model. ArXiv Preprint posted online on March 06, 2023. [doi: 10.48550/arXiv.2303.03378]
72. Zhang T, Liu N, Xu J, Liu Z, Zhou Y, Yang Y, et al. Flexible electronics for cardiovascular healthcare monitoring. Innovation (Camb) 2023;4(5):100485 [FREE Full text] [doi: 10.1016/j.xinn.2023.100485] [Medline: 37609559]
73. Volpe NJ, Mirza RG. Chatbots, artificial intelligence, and the future of scientific reporting. JAMA Ophthalmol 2023;141(9):824-825. [doi: 10.1001/jamaophthalmol.2023.3344] [Medline: 37498621]
74. Li S. ChatGPT has made the field of surgery full of opportunities and challenges. Int J Surg 2023;109(8):2537-2538 [FREE Full text] [doi: 10.1097/JS9.0000000000000454] [Medline: 37195807]
75. O'Hern K, Yang E, Vidal NY. ChatGPT underperforms in triaging appropriate use of Mohs surgery for cutaneous neoplasms. JAAD Int 2023;12:168-170 [FREE Full text] [doi: 10.1016/j.jdin.2023.06.002] [Medline: 37404248]
76. Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou Y, et al. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. EClinicalMedicine 2022;53:101662 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101662] [Medline: 36147628]
77. Zhu M, Chen Z, Yuan Y. DSI-Net: deep synergistic interaction network for joint classification and segmentation with endoscope images. IEEE Trans Med Imaging 2021;40(12):3315-3325. [doi: 10.1109/TMI.2021.3083586] [Medline: 34033538]
78. Cheng K, Wu C, Gu S, Lu Y, Wu H, Li C. WHO declares the end of the COVID-19 global health emergency: lessons and recommendations from the perspective of ChatGPT/GPT-4. Int J Surg 2023;109(9):2859-2862 [FREE Full text] [doi: 10.1097/JS9.0000000000000521] [Medline: 37246993]
79. Madden MG, McNicholas BA, Laffey JG. Assessing the usefulness of a large language model to query and summarize unstructured medical notes in intensive care. Intensive Care Med 2023;49(8):1018-1020. [doi: 10.1007/s00134-023-07128-2] [Medline: 37338549]

XSL•FO
RenderX

80. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digital Health 2023;5(6):e333-e335 [FREE Full text] [doi: 10.1016/S2589-7500(23)00083-3] [Medline: 37120418]
81. van Heerden AC, Pozuelo JR, Kohrt BA. Global mental health services and the impact of artificial intelligence-powered large language models. JAMA Psychiatry 2023;80(7):662-664. [doi: 10.1001/jamapsychiatry.2023.1253] [Medline: 37195694]
82. Kwok K, Wei W, Tsoi M, Tang A, Chan M, Ip M, et al. How can we transform travel medicine by leveraging on AI-powered search engines? J Travel Med 2023;30(4):taad058. [doi: 10.1093/jtm/taad058] [Medline: 37098161]
83. NA. AI-powered structure-based drug design inspired by the lock-and-key model. Nat Comput Sci 2023;3(10):827-828. [doi: 10.1038/s43588-023-00552-w] [Medline: 38177772]
84. NA. Upswing in AI drug-discovery deals. Nat Biotechnol 2023;41(10):1361. [doi: 10.1038/s41587-023-02002-4] [Medline: 37828287]
85. Xu Z, editor. Using large pre-trained language model to assist FDA in premarket medical device classification. In: SoutheastCon. Orlando, FL, USA: IEEE; 2023.
86. Wu C, Lei J, Zheng Q, Zhao W, Lin W, Zhang X. Can GPT-4V (ision) Serve Medical Applications? Case Studies on GPT-4V for Multimodal Medical Diagnosis. ArXiv Preprint posted online on October 15, 2023. [doi: 10.48550/arXiv.2310.09909]
87. He K, Mao R, Lin Q, Ruan Y, Lan X, Feng M. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. ArXiv Preprint posted online on October 09, 2023. [doi: 10.48550/arXiv.2310.05694]
88. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023;388(13):1201-1208. [doi: 10.1056/NEJMra2302038] [Medline: 36988595]
89. Marks M, Haupt CE. AI chatbots, health privacy, and challenges to HIPAA compliance. JAMA 2023;330(4):309-310. [doi: 10.1001/jama.2023.9458] [Medline: 37410450]
90. Mello MM, Guha N. ChatGPT and physicians' malpractice risk. JAMA Health Forum 2023;4(5):e231938 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.1938] [Medline: 37200013]
91. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA 2023;330(4):311-312. [doi: 10.1001/jama.2023.9618] [Medline: 37410449]
92. NA. ChatGPT is a black box: how AI research can break it open. Nature 2023;619(7971):671-672. [doi: 10.1038/d41586-023-02366-2] [Medline: 37491394]
93. Tang YD, Dong ED, Gao W. LLMs in medicine: the need for advanced evaluation systems for disruptive technologies. Innovation (Camb) 2024;5(3):100622 [FREE Full text] [doi: 10.1016/j.xinn.2024.100622] [Medline: 38699776]
94. Grigorian A, Shipley J, Nahmias J, Nguyen N, Schwed AC, Petrie BA, et al. Implications of using chatbots for future surgical education. JAMA Surg 2023;158(11):1220-1222. [doi: 10.1001/jamasurg.2023.3875] [Medline: 37728910]
95. Kozlov M, Biever C. AI 'breakthrough': neural net has human-like ability to generalize language. Nature 2023;623(7985):16-17. [doi: 10.1038/d41586-023-03272-3] [Medline: 37880533]
96. Ferryman K, Mackintosh M, Ghassemi M. Considering biased data as informative artifacts in AI-assisted health care. N Engl J Med 2023;389(9):833-838. [doi: 10.1056/NEJMra2214964] [Medline: 37646680]
97. Tan TF, Teo ZL, Ting DSW. Artificial intelligence bias and ethics in retinal imaging. JAMA Ophthalmol 2023;141(6):552-553. [doi: 10.1001/jamaophthalmol.2023.1490] [Medline: 37140916]
98. Harris E. Large language models answer medical questions accurately, but can't match clinicians' knowledge. JAMA 2023;330(9):792-794. [doi: 10.1001/jama.2023.14311] [Medline: 37548971]
99. Kim J, Cai ZR, Chen ML, Simard JF, Linos E. Assessing biases in medical decisions via clinician and AI chatbot responses to patient vignettes. JAMA Netw Open 2023;6(10):e2338050 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.38050] [Medline: 37847506]
100. Piao J, Liu J, Zhang F, Su J, Li Y. Human–AI adaptive dynamics drives the emergence of information cocoons. Nat Mach Intell 2023;5(11):1214-1224. [doi: 10.1038/s42256-023-00731-4]
101. Nordling L. How ChatGPT is transforming the postdoc experience. Nature 2023;622(7983):655-657. [doi: 10.1038/d41586-023-03235-8] [Medline: 37845528]
102. The Lancet. AI in medicine: creating a safe and equitable future. Lancet 2023;402(10401):503. [doi: 10.1016/S0140-6736(23)01668-9] [Medline: 37573071]
103. The Lancet. AI in medicine: creating a safe and equitable future. Lancet 2023;402(10401):503. [doi: 10.1016/S0140-6736(23)01668-9] [Medline: 37573071]
104. Topol EJ. Machines and empathy in medicine. Lancet 2023;402(10411):1411. [doi: 10.1016/S0140-6736(23)02292-4] [Medline: 37865458]
105. Hagendorff T, Fabi S, Kosinski M. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. Nat Comput Sci 2023;3(10):833-838 [FREE Full text] [doi: 10.1038/s43588-023-00527-x] [Medline: 38177754]

XSL•FO
RenderX

106. Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DSW, Liu N. Large language models in health care: development, applications, and challenges. Health Care Sci 2023;2(4):255-263 [FREE Full text] [doi: 10.1002/hcs2.61] [Medline: 38939520]
107. Wu Y, Si Z, Gong H, Zhu SC. Learning active basis model for object detection and recognition. Int J Comput Vis 2009;90(2):198-235. [doi: 10.1007/s11263-009-0287-0]
108. Horgan J. The consciousness conundrum. IEEE Spectr 2008;45(6):36-41. [doi: 10.1109/mspec.2008.4531459]
109. Lam K. ChatGPT for low- and middle-income countries: a Greek gift? Lancet Reg Health West Pac 2023;41:100906 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100906] [Medline: 37745974]
110. Wang X, Sanders HM, Liu Y, Seang K, Tran BX, Atanasov AG, et al. ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Health West Pac 2023;41:100905 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100905] [Medline: 37731897]
111. Hswen Y, Voelker R. New AI tools must have health equity in their DNA. JAMA 2023;330(17):1604-1607. [doi: 10.1001/jama.2023.19293] [Medline: 37819644]
112. Seghier ML. ChatGPT: not all languages are equal. Nature 2023;615(7951):216. [doi: 10.1038/d41586-023-00680-3] [Medline: 36882613]
113. Tan B, Zhu Y, Liu L, Wang H, Zhuang Y, Chen J. RedCoast: a lightweight tool to automate distributed training of LLMs on any GPU/TPUs. 2024 Presented at: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: System Demonstrations); June 01, 2024; Mexico City, Mexico p. 137-147. [doi: 10.18653/v1/2024.naacl-demo.14]
114. Narayanan D, Shoeybi M, Casper J, LeGresley P, Patwary M, Korthikanti V, et al. Efficient large-scale language model training on GPU clusters using megatron-LM. 2021 Presented at: SC21: International Conference for High Performance Computing, Networking, Storage and Analysis; November 13, 2021; New York, NY, United States p. 1-15. [doi: 10.1145/3458817.3476209]
115. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
116. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. iScience 2024;27(5):109713 [FREE Full text] [doi: 10.1016/j.isci.2024.109713] [Medline: 38746668]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
GPU: graphics processing unit
LLM: large language model
MRI: magnetic resonance imaging
NLP: natural language processing
TPU: tensor processing unit

Edited by A Schwartz; submitted 01.04.24; peer-reviewed by Y Khan, K Singh, S Mao, D Patel, D Ghosh, P Sarmadi; comments to author 13.08.24; revised version received 26.08.24; accepted 10.09.24; published 07.01.25.

Please cite as:
Zhang K, Meng X, Yan X, Ji J, Liu J, Xu H, Zhang H, Liu D, Wang J, Wang X, Gao J, Wang YGS, Shao C, Wang W, Li J, Zheng MQ, Yang Y, Tang YD
Revolutionizing Health Care: The Transformative Impact of Large Language Models in Medicine
J Med Internet Res 2025;27:e59069
URL: https://www.jmir.org/2025/1/e59069
doi:10.2196/59069
PMID:

©Kuo Zhang, Xiangbin Meng, Xiangyu Yan, Jiaming Ji, Jingqian Liu, Hua Xu, Heng Zhang, Da Liu, Jingjia Wang, Xuliang Wang, Jun Gao, Yuan-geng-shuo Wang, Chunli Shao, Wenyao Wang, Jiarong Li, Ming-Qi Zheng, Yaodong Yang, Yi-Da Tang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The CeHRes Roadmap 2.0: Update of a Holistic Framework for Development, Implementation, and Evaluation of eHealth Technologies

Hanneke Kip[1,2], PhD; Nienke Beerlage-de Jong[1], PhD; Lisette J E W C van Gemert-Pijnen[1], Prof Dr; Saskia M Kelders[1,3], PhD

[1]Section of Psychology, Health & Technology, Centre for eHealth and Wellbeing, University of Twente, Enschede, Netherlands

[2]Department of Research, Transfore, Deventer, Netherlands

[3]Optentia Research Unit, North-West University, Potchefstroom, South Africa

Corresponding Author:
Hanneke Kip, PhD
Section of Psychology, Health & Technology
Centre for eHealth and Wellbeing
University of Twente
Drienerlolaan 5
Enschede, 7522
Netherlands
Phone: 31 534896536
Email: h.kip@utwente.nl

Abstract

To ensure that an eHealth technology fits with its intended users, other stakeholders, and the context within which it will be used, thorough development, implementation, and evaluation processes are necessary. The CeHRes (Centre for eHealth and Wellbeing Research) Roadmap is a framework that can help shape these processes. While it has been successfully used in research and practice, new developments and insights have arisen since the Roadmap's first publication in 2011, not only within the domain of eHealth but also within the different disciplines in which the Roadmap is grounded. Because of these new developments and insights, a revision of the Roadmap was imperative. This paper aims to present the updated pillars and phases of the CeHRes Roadmap 2.0. The Roadmap was updated based on four types of sources: (1) experiences with its application in research; (2) literature reviews on eHealth development, implementation, and evaluation; (3) discussions with eHealth researchers; and (4) new insights and updates from relevant frameworks and theories. The updated pillars state that eHealth development, implementation, and evaluation (1) are ongoing and intertwined processes; (2) have a holistic approach in which context, people, and technology are intertwined; (3) consist of continuous evaluation cycles; (4) require active stakeholder involvement from the start; and (5) are based on interdisciplinary collaboration. The CeHRes Roadmap 2.0 consists of 5 interrelated phases, of which the first is the contextual inquiry, in which an overview of the involved stakeholders, the current situation, and points of improvement is created. The findings from the contextual inquiry are specified in the value specification, in which the foundation for the to-be-developed eHealth technology is created by formulating values and requirements, preliminarily selecting behavior change techniques and persuasive features, and initiating a business model. In the Design phase, the requirements are translated into several lo-fi and hi-fi prototypes that are iteratively tested with end users and other stakeholders. A version of the technology is rolled out in the Operationalization phase, using the business model and an implementation plan. In the Summative Evaluation phase, the impact, uptake, and working mechanisms are evaluated using a multimethod approach. All phases are interrelated by continuous formative evaluation cycles that ensure coherence between outcomes of phases and alignment with stakeholder needs. While the CeHRes Roadmap 2.0 consists of the same phases as the first version, the objectives and pillars have been updated and adapted, reflecting the increased emphasis on behavior change, implementation, and evaluation as a process. There is a need for more empirical studies that apply and reflect on the CeHRes Roadmap 2.0 to provide points of improvement because just as with any eHealth technology, the Roadmap has to be constantly improved based on the input of its users.

(J Med Internet Res 2025;27:e59601) doi:10.2196/59601

XSL•FO
RenderX

KEYWORDS

eHealth development; eHealth implementation; CeHRes Roadmap; participatory development; human-centered design; persuasive design; eHealth framework

Introduction

eHealth refers to the use of information and communication technology to support health, well-being, and health care [1]. If used well, eHealth technologies such as internet-based interventions, virtual reality, wearables, and mobile apps can be used to improve the quality and efficiency of care [2-6]. By achieving the intended effects, they can be used to target global challenges for health care, for example, limited access to care, an increase in the number of people living with (chronic) diseases, and shortages of staff and funding [7,8]. Over the past years, much attention has been paid to improving the impact and uptake of eHealth by academia, health care, and commercial companies. Yet, there still is room for improvement as expected benefits are often not (fully) achieved in practice [9,10].

An explanation for eHealth interventions not reaching their full potential is a suboptimal fit between the design of the eHealth technology, the characteristics and needs of its users, and the demands of the context in which it is used [11]. Therefore, thorough development, implementation, and evaluation processes are recommended to increase the fit between technology, people, and context and thus to realize the full potential of eHealth [12]. Unfortunately, this does not always happen. For example, despite the importance of participatory development for creating eHealth technologies, end users and other stakeholders are often not involved at all or not in the most optimal way [9,13]. Furthermore, during eHealth development, sufficient attention is not paid to behavior changes by integrating evidence-based behavior change theories [12,13]. In addition, the implementation of eHealth is a major challenge but often receives insufficient attention in research and practice [14,15]. Implementation is often seen as a single, postdesign activity, as opposed to an ongoing, iterative process that involves complex ecosystems and starts with the development of a technology. In addition, because eHealth technologies are quite new, there is an urgent need for more insight into if, why, how, when, and for whom they work, using innovative research methods that are suitable for answering these types of complex questions [16,17]. Finally, in research and practice, development, implementation, and evaluation projects are often conducted separately from each other, overlooking their interrelationships. Consequently, bolstering the impact of eHealth requires the use of comprehensive frameworks that guide interrelated development, implementation, and evaluation processes [12,18].

The CeHRes (Centre for eHealth and Wellbeing Research) Roadmap 2.0 is one such holistic framework and is specifically aimed at developing, implementing, and evaluating eHealth technologies that fit the needs and characteristics of people and their contexts. The original CeHRes Roadmap was first published in 2011 and is based on a comprehensive systematic literature search, resulting in 60 papers and 16 frameworks (van Gemert-Pijnen et al [11] provide a complete overview of the search process and outcomes). It included frameworks focused on, for example, user-centered design of teleconsulting systems [19], continuous evaluation of eHealth [20], usability engineering [21], interdisciplinary design and evaluation [22], and implementation and behavioral science [23]. These frameworks' target groups, goals, foundations, and strategies and principles can be found in the study by van Gemert-Pijnen et al [11]. An analysis of the strategies and principles was used to create the original CeHRes Roadmap and its underlying pillars. The Roadmap consists of 5 interrelated phases: contextual inquiry, value specification, design, operationalization, and summative evaluation [11,24,25]. These phases are connected using formative evaluation cycles. The original Roadmap is supported by pillars that highlight the importance of participatory development, continuous evaluation, and intertwining development and implementation, accounting for changes in the organization of health care because of eHealth, the importance of persuasive design, and the use of advanced methods to assess impact. These pillars are illustrative of the interdisciplinary nature of the Roadmap, which draws from domains such as psychology, persuasive technology, engineering, human-centered design, and business modeling. Over the past years, the original CeHRes Roadmap has been applied to a broad range of contexts and has proven to be a useful tool for researchers and practitioners in the field of eHealth (refer to the study by Kip et al [13] for a partial overview). The broad and extensive use of the Roadmap has led to multiple lessons learned and suggestions for improvements by researchers [26,27]. In addition, since 2011, there have been developments within the interdisciplinary domain of eHealth. For example, participatory development has become more common; there is an increase of well-substantiated design principles and guidelines; knowledge on innovative implementation and evaluation approaches has increased; there is more interdisciplinary collaboration and knowledge; and collaborations between researchers, commercial companies, and health care organizations have resulted in new methods and frameworks [14]. In addition, in the last decade, the terminology has changed and there is a new classification of digital interventions, services, and applications [28]. These new developments and insights call for an update of the CeHRes Roadmap, embracing novel perspectives and advancements. It is important to note that, for consistency reasons, we have decided to use the term "eHealth" instead of digital health (interventions).

In this paper, we present an updated version of the phases and pillars of the CeHRes Roadmap. This update is based on 4 different types of resources. First, we used experiences, lessons learned, and points of improvement that were reported in studies that used parts of the original Roadmap [26,27,29-42]. Second, literature reviews on eHealth development, implementation, and evaluation in general were used [13,14,16,43]. Third, we gathered input from discussions with eHealth researchers from different domains (eg, psychiatry, informal caregiving, planetary health, health psychology, and health sciences) on their

XSL•FO
RenderX

experiences with and viewpoints on development, implementation, and evaluation. Fourth, we updated and added new frameworks, models, and methods from the different disciplines that are relevant to the CeHRes Roadmap [11]. More specifically, chapters on psychology and behavior change, technology development models, human-centered design, business modeling, value-based design and requirement engineering, implementation science, and evaluation methods grounded in behavioral sciences were used as foundations for the revisions and updates of the Roadmap [24].

The main objective of this paper is to present the updated version of the CeHRes Roadmap: the CeHRes Roadmap 2.0. The CeHRes Roadmap 2.0 should not be viewed as a new framework or a major revision but as a fine-tuned, improved version of the original Roadmap. In this paper, first, the revised pillars will be presented and explained, including references to the studies on which they are based. Second, the updated phases of the CeHRes Roadmap 2.0 are described, starting with a short explanation of their main objectives, followed by examples of relevant methods and frameworks from different disciplines.

The CeHRes Roadmap 2.0

In Figure 1, the new visualization of the CeHRes Roadmap 2.0 is provided. In this section, the revised underlying pillars will first be discussed, after which the 5 phases and formative evaluation cycles will be explained.

Figure 1. The CeHRes Roadmap 2.0. CeHRes: Centre for eHealth and Wellbeing Research.

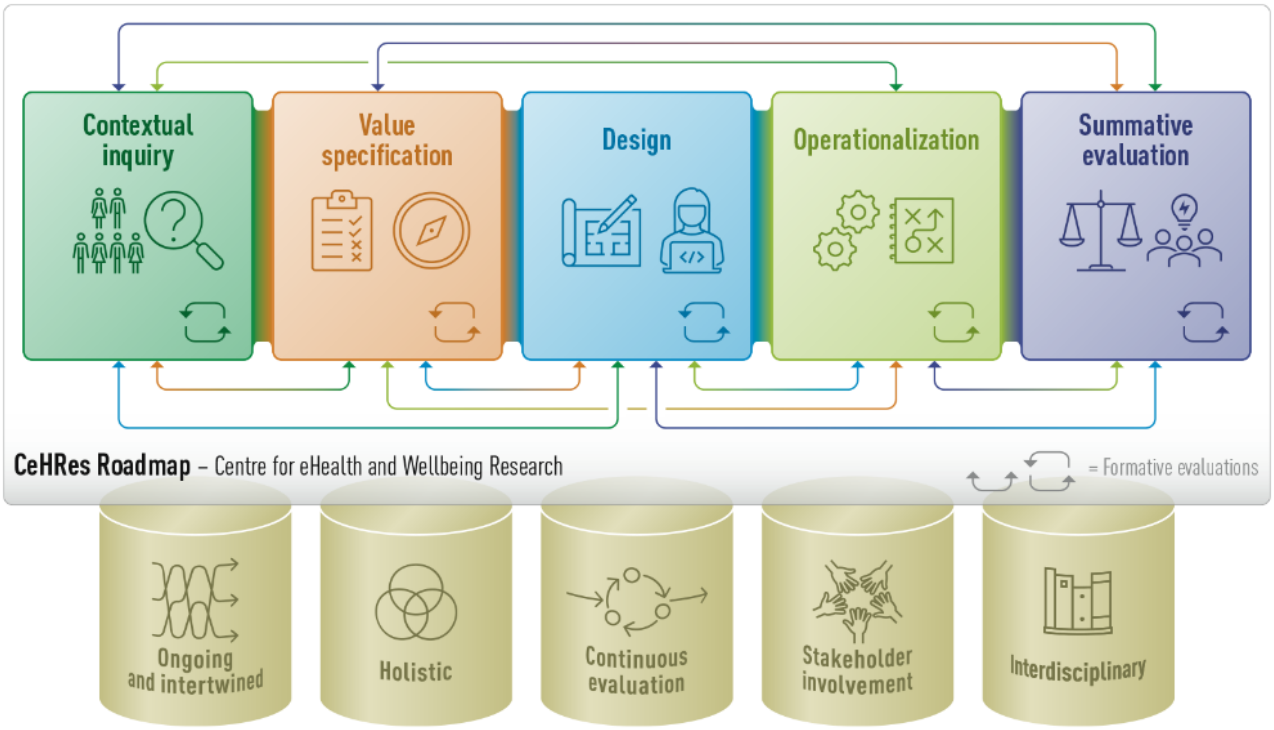

The Pillars of the CeHRes Roadmap 2.0

The Roadmap's pillars provide the foundation upon which it is built. They are the underlying assumptions that are kept in mind throughout all phases of the CeHRes Roadmap 2.0.

eHealth Development, Implementation, and Evaluation Processes Are Ongoing and Intertwined Rather Than Phases Containing Separate and Sequential Activities

While it might sound like development, implementation, and evaluation are separate and consecutive stages, this is not the case; they are highly interrelated activities that are all relevant from the start [44]. This also applies to the visualization of many development frameworks; while, for overview purposes, processes or phases are visualized as separate activities, in reality, they are all intertwined [11,24,44,45]. In particular, implementation is seen, still too often, as a postdesign activity, while it plays a role from the start, for example, by identifying potential implementation issues as early as possible to ensure that they can be accounted for during development. In line with this, eHealth development, implementation, and evaluation processes are never really completed; there will always be room for improvement and new questions will keep on arising. However, in practice, this is often difficult to realize because of limited resources, such as time or money.

eHealth Development, Implementation, and Evaluation Processes Require a Holistic Approach, Focused on the Fit Between Technology, People, and Context

Regardless of the type or goal of an eHealth intervention, there are interrelationships between the design and content of a technology, the needs and preferences of the people involved, and the context in which it is used [46]. These interrelationships warrant a holistic approach, in which the different elements are all interdependent and part of one whole instead of separate constructs [47]. First, technology and the behavior of people are interrelated. The introduction and sustained use of technology ideally influence the way end users behave, resulting in, for example, better health outcomes or increased well-being. To enable this behavior change, adherence is an important

XSL•FO
RenderX

concept, which means that eHealth should be used as intended, which can be viewed as use behavior. Concepts such as engagement—the extent to which someone is involved or occupied with an eHealth technology from an emotional, behavioral, and cognitive perspective—are important to consider as well [48]. Second, the use of eHealth is never isolated, but influences and is influenced by the context in which it is used. The introduction of eHealth creates novel structures and processes for health care delivery, and as a result, an adapted or new ecosystem for health care emerges [49]. Furthermore, the content of a technology can be influenced by contextual factors, such as new treatment protocols.

eHealth Development, Implementation, and Evaluation Processes Require a Multimethod, Iterative Approach With Continuous Evaluation Cycles

eHealth development, implementation, and evaluation are not linear processes with consecutive steps; they are iterative, flexible, and dynamic, with constant changes to the process and products [50,51]. This is in line with an agile approach, which is characterized by the division of large tasks into rapid, shorter phases and constant adaptations of plans based on the outcomes of evaluations [52]. To realize this, a multimethod approach is key. Ideally, researchers need to be able to draw from a broad toolkit of qualitative and quantitative research methods and products and select the method that best fits their research objective, the practical demands of the context, and the characteristics, needs, and wishes of the participants [13]. To ensure a coherent approach and to prevent the project team from "getting lost," there is a need for continuous formative evaluations in which outcomes of activities are critically analyzed, evaluated, and adapted with stakeholders [12]. Such an iterative, agile approach might be challenging and complex and can seem very different from a rigorous scientific approach in which all research activities are meticulously planned. However, such an iterative, agile approach is not a synonym for unstructured or messy; it is a way to shape systematic, structured yet dynamic high-quality research processes that are able to adapt to changes and new insights [52]. Such new insights may lead to the development of new technologies, but this approach is also applicable when new insights lead to re-evaluation and redesign of existing technologies. Consequently, while a broad research plan with potentially suitable methods can definitely be developed, it might be necessary to deviate from this plan based on outcomes at any point during the process or new insights regarding the needs of stakeholders or possibilities of a technology.

eHealth Development, Implementation, and Evaluation Processes Require Constant Active Involvement of Stakeholders

To create eHealth that meets the needs and wishes of users and other stakeholders, a participatory approach is recommended, in which stakeholders are actively involved throughout the process, from development to evaluation [30,45,53,54]. Emerging approaches, such as action research and design science highlight that the active involvement of stakeholders goes beyond merely involving end users; stakeholders such as managers, health care insurers, and technology developers are also involved to ensure a holistic approach [40,55,56]. In participatory development, stakeholders can take on different roles, ranging from an informant who mostly provides input on products when invited by researchers to a cocreator who is actively involved in creating ideas and products [57]. While ideally, stakeholders have an active role, in some cases a more passive role might be more suitable for a specific objective. How a stakeholder is involved is not fixed and will differ depending on the phase of the process and the type of activity that is conducted. It is important to ensure that stakeholders are only involved when it truly is of added value. Merely involving end users for the sake of involving them does not benefit anyone involved and takes up the valuable time of participants and researchers [58]. Ideally, stakeholders should be part of an interdisciplinary project team that coordinates the entire eHealth process. Finally, participatory development does not always have to be about creating new technologies, because existing technologies can be redesigned and reused in different contexts as well, which also requires active stakeholder involvement.

eHealth Development, Implementation, and Evaluation Processes Are Based on an Interdisciplinary Approach

To capture the complexity of eHealth, an interdisciplinary approach toward research and development is required [12,14]. In such an approach, theories, methods, and models from different disciplines are combined and ideally merged and integrated, resulting in new frameworks, concepts, and models [59]. While there are multiple domains that can be relevant for eHealth, there are some main paradigms that underpin the CeHRes Roadmap 2.0, which are psychology and behavioral sciences, human-centered design, engineering, persuasive technology, implementation science, and business modeling [11,24]. Furthermore, domain-specific theories that are relevant to a specific context or target group often play an important role and should be integrated as well [29]. Finally, the interdisciplinary nature of eHealth is important when composing the project team that coordinates the development, implementation, and evaluation processes. Putting together a team with members (and stakeholders) from different disciplines is deemed essential to ensure that all relevant perspectives are actively involved throughout to prevent tunnel vision [29,60,61].

Changes Compared to the Original CeHRes Roadmap

Compared to the first version of the Roadmap, several changes are made to the pillars while retaining the main essence of the original Roadmap. An important change is that the pillars were broadened to better reflect the vision that development, implementation, and evaluation aspects of eHealth are interrelated. For example, in the initial pillars, the term "eHealth development" was used, but in the new version, we broadened this to "eHealth development, implementation, and evaluation processes." Furthermore, the terminology was adapted to better fit the iterative, interdisciplinary nature of eHealth. For example, the term participatory development was replaced by the term "active involvement of stakeholders." The phrasing and changes of the updated pillars better capture the essence of the CeHRes Roadmap 2.0.

XSL•FO
RenderX

The Phases of the CeHRes Roadmap 2.0

Overview

Before diving into the phases of the CeHRes Roadmap 2.0, it is important to explain how the Roadmap should be used. It is not a step-by-step guideline but more of a set of underlying assumptions that are relevant to any eHealth process. Strictly speaking, the first 3 phases of the Roadmap 2.0 are related to the development of eHealth, operationalization fits best with implementation, and summative and formative evaluation are most aligned with the Evaluation phase. This distinction into 5 phases is made for visualization and overview purposes, in practice, the first phases will also be relevant for implementation and evaluation efforts and vice versa. In line with this, the visualization of the phases as separate blocks does not in any way imply that the process is sequential; this is also done for overview purposes. In reality, phases overlap and project teams will go back and forth between the different phases, as represented by the arrows in Figure 1. Therefore, it is important to note that the description of the phases of the CeHRes Roadmap 2.0 is not a concrete guideline for creating impactful eHealth interventions. It is meant to serve as a structured overview of principles and methods that can be used by researchers and practitioners to shape their own and unique development, implementation, and evaluation processes. Consequently, methods or frameworks that are not described in this paper can most definitely be used too. A key feature is that they fit within the process and are aligned with relevant research questions. To conclude, the information mentioned subsequently is aimed to support eHealth researchers through inspiration, guidance, and concrete input for guiding eHealth development, implementation, and evaluation processes.

Contextual Inquiry

Overview

The first step in any eHealth-related project is a thorough investigation of the context in which it will be used. A contextual inquiry is required to paint a picture of the (groups of) people that will come in contact with the technology and their context and to identify the main points of improvement for which an eHealth technology can be of added value. A thorough contextual inquiry assists researchers and developers in keeping a focus on the people and their context from the start of the development process [30,62]. This increases the chances of developing an eHealth technology that provides a solution for relevant issues, fits within a specific context, and is accepted by relevant stakeholders [12,63]. Consequently, the outcomes of a contextual inquiry provide the foundation for all following development, implementation, and evaluation activities. The objectives of the contextual inquiry are described in subsequent sections, including several illustrative examples of suitable methods, frameworks, and concepts that might be used within this phase.

To Identify and Gain Insights Into the Roles and Activities of Relevant Stakeholders

Due to the importance of actively involving stakeholders in every phase of the CeHRes Roadmap 2.0, a clear overview has to be provided from the start of the development process. Stakeholder identification can be conducted, using methods such as expert recommendations on stakeholders, snowball sampling with stakeholders, and literature studies [40]. Once an initial overview is created, the roles and tasks of stakeholders have to be analyzed. An example of a useful approach is the stakeholder salience model, in which stakeholders are mapped based on their power, legitimacy, and urgency [64]. On the basis of analyses, key stakeholders who should be involved more intensively throughout the process can be selected.

To Analyze the Current Situation Regarding the Involved People, Organizations, Wider Context, and Technologies

A thorough understanding of the current situation is essential to ensure that the project team knows for whom and for what context they are creating eHealth and to specify the focus of the project. Using multiple methods, an overview can be created of the involved people (eg, their daily lives, attitudes, behavior and determinants of behavior), the organizations in which eHealth will be implemented (eg, organizational characteristics, protocols, and domain-specific theories), the wider context (eg, national policies, legislation, and funding possibilities), and information on relevant technologies and technological infrastructures that are or might be used in the intended or other contexts (eg, working mechanisms, experiences with similar technologies, and technological possibilities) [13]. It is important to truly understand the context before going into specific solutions to prevent tunnel vision and to ensure a good fit between the eHealth technology and the involved stakeholders. To achieve this, a context should not be viewed as a single entity but as a complex ecosystem of interrelated platforms, technologies people, organizations, and other stakeholders. Finally, it is important to pay attention to (potential) ethical points of attention from the start of the development process, such as privacy, accessibility, or data ownership.

To Identify the Main Points of Improvement for Which an eHealth Technology Could Be a Solution

Merely describing a context is not sufficient for developing eHealth; the project team needs to gain insights into what the points of improvements are in the current situation that can be addressed by a to-be-developed eHealth technology. These points can be related to, for example, efficiency, effectiveness, patient satisfaction, safety, or quality improvement. An important aspect of this activity is to describe the behavior that needs to be changed to improve the situation and identify accompanying determinants that can be targeted by eHealth technology [45]. When identifying and describing points of improvement and accompanying behavior, newly gathered information from stakeholders is combined with scientific sources to increase the robustness of the findings. By integrating methods such as desk research, (systematic) literature reviews, focus groups, interviews, and observations, a complete, multifaceted picture of what has to be changed can be painted [13].

Value Specification

Overview

In the Value Specification phase, the topics that were identified in the contextual inquiry are narrowed down and translated into

XSL•FO
RenderX

values and specific requirements for a to-be-developed eHealth technology. This is again done in close cooperation with the identified stakeholders; however, other stakeholders might also pop up in this phase. Information from the literature is required to be able to incorporate evidence-based behavior change techniques (BCTs) into a technology that facilitates behavior change. Moreover, in this phase, the development team needs to take the first concrete steps toward implementation, among other things, by initiating the development of a business model. While the value specification can be quite complex and time consuming, it is important to put effort into laying the groundwork for the content and design of a technology before diving into its actual creation. This again ensures that the technology is in line with the preferences and characteristics of stakeholders, incorporates evidence-based working mechanisms, and remains in line with the identified points of improvement. The objectives of this phase and examples of accompanying methods are described in subsequent sections.

To Formulate and Prioritize Values From All Identified Key Stakeholders and Set Concrete Goals Based on These Values

While there are many definitions of values, in relation to eHealth development, values can be viewed as what stakeholders find important regarding the technology from their personal or professional perspective, also related to, for example, ethics and morality. In other words, values are abstract concepts that represent the ideals or interests of stakeholders and summarize what they find important on an abstract level [38,65]. Values can be related to social, behavioral, cognitive, emotional, economic, or health care issues. Examples are saving costs, increasing access to health care, or improving autonomy [39,65]. Values can be uncovered through (ideally) a combination of multiple methods such as interviews, focus groups, or questionnaires [65]. When collecting data, it is pivotal to carefully consider the intended target audience when finding a fitting way to discuss values. For some target audiences (eg, clients in forensic psychiatry), discussing values in an abstract way (eg, “increase self-control”) can be very complex [66]. In that case, it can help to show and discuss concrete examples of existing or potential technologies [29]. Because different stakeholders are involved, conflicting values are expected to arise, for example, the manager’s perspective might differ from that of the patient. This is not problematic but requires deliberate decision-making and prioritizing by the interdisciplinary project team. To ensure that these more abstract values are translated into a concrete technology, more specific requirements are formulated. Furthermore, to guarantee that values are incorporated throughout the further implementation and evaluation activities, it is useful to formulate accompanying objectives that are formulated in a specific, measurable, achievable, realistic, and timely (SMART) way. These objectives can serve as the basis for evaluation and implementation studies.

To Select Initial BCTs and Persuasive Features That Fit With the Values and Needs of the Key Stakeholders

In previous activities, the behavior that needs to be changed was analyzed and the values of the key stakeholders that play a role in doing so were identified. On the basis of that information, fitting behavior change methods can now be selected to increase the chances of actually reaching the intended behavior change and to improve adherence and engagement to the eHealth technology. This is important because changing behavior is highly complicated. To increase the chances of an eHealth technology’s success in actually changing behavior, the chosen strategy should fit the users and what they are trying to achieve. BCTs are derived from behavior change theories from psychology and can be defined as a general technique to influence or create changes in the predictors of specific behavior [67]. Examples of BCT categories are goals and planning, social support, feedback and monitoring, and reward and threat. In addition, persuasive features from the Persuasive Systems Design (PSD) model can be selected [68]. The persuasive features are divided into 4 categories: primary task support, dialogue support, credibility support, and social support. BCTs partly overlap with persuasive techniques because they both target behavior change, but the main difference is that the PSD model is specifically aimed at technology, while BCTs are applicable to any kind of behavior change intervention. Finally, depending on the project, domain-specific theories can also be used to support behavior change. Identifying and integrating behavior change methods in technology is a complex and highly iterative process that will remain relevant throughout the process, but the foundation is laid in the Value Specification phase.

To Translate the Values Into Specific Requirements to Describe the to-Be-Developed eHealth Technology

Requirements describe what an eHealth technology should do, what data it should store or retrieve, what content it should display, and what kind of user experience it should provide [38]. There are different types of requirements related to functionality and modality, service, organization, content, usability, and user experience. Particularly due to the increasing importance and the use of (big) data, attention should be paid to data-related topics, such as data ownership, the way data are collected, and how and where to save generated data. Requirements should be related to values; an abstract value is often related to multiple more specific requirements. When formulating requirements, data that were used to formulate values can be used, but new data are often collected as well, for example, through focus groups, interviews, card sorting, and literature reviews [13]. It is also important to integrate the selected potential behavior change mechanisms into requirements to ensure that they are integrated into the technology. A specific product that can help in designing for the end users and facilitating requirement formulation is a persona: a hypothetical archetype of a user presented by means of a short biography with a photo, sociodemographic information, and their values [69,70]. Scenarios can be used to link personas to the technology, because they describe the potential, envisioned use of an eHealth technology by the end users. Scenarios can contain information about how the end user uses a technology in relation to their characteristics and values, what activities are supported by a technology, which steps a user follows, in which context this happens, and what the technology’s main functionalities are [71].

XSL•FO
RenderX

To Create the First Version of a Business Model for the eHealth Technology

To prevent a technology from not being used after it has been introduced in practice or not even being introduced at all, it is important to start developing a business model as soon as possible. A business model can be defined as "the rationale of how an organization creates, delivers, and captures value" [72]. A frequently used method to create a business model is the business model canvas, which consists of 9 blocks. These blocks represent the value proposition (eg, the eHealth technology), the organizational activities that are necessary to deliver the eHealth technology to practice, the main customers and users, and financial aspects and cost-benefit ratios. A business model can be filled in using multiple methods, such as focus groups, desk research, and outcomes of earlier development activities [32]. While the initiation of the business model takes place in the value specification, it is adapted throughout the entire process of eHealth development, implementation, and evaluation.

Design

Overview

While this is not a "hard" transition, the actual eHealth technology is created in the Design phase based on the output of the contextual inquiry and the value specification. This is an extremely dynamic, iterative, creative, and collaborative phase during which there is an active collaboration between end users, other stakeholders, researchers, designers, behavior change experts, content experts, and perhaps even funders. While there are many ways to guide this process, the Double Diamond Model (or framework for innovation) can be used to shape the cycles of iteration between design and testing that are relevant for this phase [73,74]. It is important to note that the final technology is not developed all at once. If that is done, important problems may be missed and arise only after the technology has been introduced in practice, thus wasting a lot of time and effort. Therefore, multiple prototypes, visual representations of the to-be-developed technology, are developed based on the requirements. These prototypes are evaluated by users and experts and are constantly updated in an iterative process. It must be stressed that throughout the prototyping process, outcomes of previous phases (eg, requirements, BCTs, and stakeholder lists) can be adapted or complemented because of new insights, related to, for example, usability, new organizational opportunities, or technological possibilities or impossibilities. To support this process, design principles for digital developments can be used [75]. Furthermore, in the Design phase, insights into barriers that might arise during implementation are gathered, illustrating the interrelatedness between development and implementation.

Both Low-Fidelity and High-Fidelity Prototypes of the eHealth Technology Are Developed

Prototypes are visual representations of a technology. They can be crude and more sketch like (low fidelity) or can resemble the to-be-developed technology more closely (high fidelity), but they are in any case based on previously formulated requirements [76]. The first step in the prototyping process is to come up with initial design ideas by means of ideation. These ideas are increasingly specified based on input from stakeholders in formative evaluations throughout the prototyping process. There are several methods to create prototypes, such as paper-based prototyping (eg, sketching), computer-based digital prototyping, storyboards that show how the technology can be used, or 3D prototyping with materials such as cardboard, clay, or 3D printing. Content for the prototypes can be initiated via, for example, literature studies, desk research on existing protocols, card sorting, Delphi studies with experts, and further refined during the design process [13].

Persuasive Features, BCTs, and Domain-Specific Theories Are Operationalized and Integrated Into the Design of the eHealth Technology

While an initial overview of potentially useful BCTs and persuasive features was made in the value specification, this set of techniques will most likely be adapted throughout (and even after) the prototyping process, in line with the complexity of behavior change. Furthermore, the techniques have to be operationalized and integrated into prototypes of the design [33]. Integrating persuasive features and BCTs in the design of an eHealth technology requires an iterative and creative approach, for example, using brainstorming and cocreation sessions with stakeholders, involving experts on design and behavior change, and investigating techniques and features in similar technologies. Once these techniques and features are integrated into prototypes, they are evaluated with stakeholders and experts to investigate if they are potentially effective and fit the skills and preferences of end users. While doing so, it is important to also consider checking if the BCTs and persuasive features are integrated successfully into the design according to experts and stakeholders. An example of this is to analyze perceived persuasiveness according to stakeholders [77].

Usability Tests of the Prototypes Are Conducted With End Users, Experts, and Other Stakeholders

Prototypes are evaluated with stakeholders to gain insights into their opinions and thus iterate toward the most suitable version of an eHealth technology for a specific target group in a specific context. This can be done by means of usability tests, which can be used to study, for example, how someone interacts with the system, to test the ease of use and user-friendliness of the technology, or to assess whether requirements are correctly translated into the design. There are multiple methods for usability testing [78]. The think-aloud method lets intended end users navigate through a prototype guided by a scenario, a task that is to be dealt with by using the prototype, while verbalizing their thoughts, actions, and ideas. In a heuristic evaluation, design experts assess usability by comparing it to a predefined set of principles such as "recognition rather than recall" or "minimalist design" [79]. In a cognitive walk-through, experts are asked to execute tasks that a user would want to perform [80]. These are just a couple of examples of the multitude of available methods for usability testing.

Operationalization

Overview

Operationalization refers to the planning and actions for the introduction, dissemination, adoption, and internalization of (a

XSL•FO
RenderX

first functioning version of) the technology in the intended context. In this phase, the technology is launched, marketing is set into motion, and organizational working procedures are put into practice. In this phase, the business model is completed and applied to practice, using input from earlier phases as well as by collecting new data. In addition, previously collected data and new research are combined into a concrete implementation plan with specific activities, ideally structured by means of an implementation framework.

Complete the Business Model With Input of the Stakeholders and Put It Into Practice

The business model, which was initiated in the Value Specification phase, is made as complete as possible and is rolled out in this phase. Concrete plans on how to act upon the content of the 9 blocks of the business model canvas are created, in close cooperation with stakeholders, using methods such as focus groups, desk research, or interviews [32]. While the business model is made as complete as possible, this does not mean that the model is fixed: chances are that it will be updated constantly, in line with, for example, new insights, policy changes, new customer groups, or new financing structures. Finally, it is important to carefully consider matters related to data in a business model, who owns the data that are collected by technology, and what is the potential value of these data [81].

Create an Overview of Implementation Barriers and Facilitators Using Input From the Previous Phases, New Data, and Implementation Frameworks

Because implementation is not a postdesign activity, information on, for example, implementation barriers and facilitators has already been collected throughout the previous activities. In this phase, previous outcomes are combined with newly collected data on implementation factors and strategies to provide an overview of factors to account for when planning, executing, and evaluating implementation processes [43]. The process of planning for implementation can be structured by means of implementation frameworks. A much-used framework is the Consolidated Framework for Implementation Research (CFIR), which incorporates a broad range of factors from existing theories and divides them into 5 domains: innovation, outer setting, inner setting, individuals, and implementation process [82]. Another approach, which is focused on value-based technology in health care, is the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework, consisting of 7 domains: condition, technology, value proposition, adopters, organization, wider system, and embedding and adaptation over time [83]. While there are different implementation frameworks, they all highlight the multilevel nature of implementation and emphasize that it is important to pay attention to a broad range of stakeholders and organizations.

Develop a Concrete Implementation Plan Based on the Previously Identified Implementation Factors and Apply It to Practice

On the basis of the previously identified implementation barriers and facilitators, a concrete plan for implementation is comprised, including implementation outcomes and strategies. While there is no step-by-step approach to do this, specific implementation outcomes can be determined to steer and evaluate the process, related to, for example, acceptability, cost, and sustainability [84]. To achieve the objectives, specific implementation strategies are identified, which can be linked to the previously identified barriers and facilitators (ie, factors) [36]. Overviews of implementation strategies can be used to guide this process [85]. Active stakeholder involvement in identifying suitable activities for implementation is important, partly because there is not much evidence-based information on which strategies can be effectively used for overcoming implementation barriers [43,85]. Examples include training health care providers, development of technology-enhanced treatment protocols, or practical support of clinicians in using the intervention [15]. The accompanying materials, such as flyers or training sessions, are ideally cocreated with stakeholders.

Summative Evaluation

Overview

While summative evaluation is visualized as the final part of the CeHRes Roadmap 2.0, it is not a postimplementation and development activity but an iterative, multimethod process that is highly intertwined with and based on activities and outcomes from development and implementation processes [16]. In evaluation processes, researchers investigate to what extent the eHealth technology succeeds in addressing the objectives formulated in previous phases. For overview purposes, a rough distinction can be made between 3 aspects of eHealth evaluation, although they overlap. The first aspect is the evaluation of the *impact* of eHealth on its users, stakeholders, and their context, mostly aimed at assessing its benefits and effects on different types of outcomes (eg, clinical, organizational, and behavioral). The second aspect is related to assessing the *uptake* of eHealth to gain insight into how, by whom, and when a technology was used. The third element is related to the evaluation of *working mechanisms* to gain insight into why an eHealth technology was effective or not and why it was used. By combining these types of evaluations, insights can be gained into if, why, how, for whom, and when eHealth works, and points of improvement for its design and implementation can be identified. Several frameworks can be used to shape these complex, multifaceted evaluation processes. Examples are the Multiphase Optimization Strategy (MOST) framework, which consists of 3 phases: preparation, optimization, and evaluation [86]; Process evaluation, in which attention is paid to the relationship between context, implementation, and mechanisms of impact, and their influence on the outcomes of an intervention [87]; or realist evaluations, in which information is collected to build Context-Mechanism-Outcome configurations to explain how a specific context triggers specific mechanisms that lead to particular outcomes [88]. These approaches can all be used for the evaluation of complex interventions and account for the role of the context, the technology, and people when evaluating an eHealth technology.

Determine the Impact of the eHealth Technology on the Users, Other Stakeholders, and Their Context

To assess the impact of eHealth, it is important to translate the previously formulated values and accompanying objectives into research questions. These can relate to, for example, clinical

XSL•FO
RenderX

values, lifestyle behaviors, quality of care, efficiency, or other types of benefits. To paint a comprehensive picture of the influence of eHealth on the users and contexts, a multi- or mixed methods approach can be used. The choice of a method depends on the research questions and characteristics of the target group and context. Some examples of methods that can be used are randomized controlled trials (RCTs), interviews or focus groups, single-case experimental designs, or health technology assessments. When choosing what method to use, it is important to thoroughly consider the context in which a technology is used and keep an eye on the limitations of a specific method [89].

Analyze the Uptake of the eHealth Technology in Terms of Adoption or Use of the Technology by Its Intended Users

Evaluations of uptake are used to gain insight into by whom, how often, when, and how an eHealth technology is used. This type of evaluation can be used to answer questions about, for example, how often users logged in, how often a technology was used by different organizations, and what features of it were used or not and in what order. While there are multiple ways to answer these questions, log data analysis is a commonly used method [37,90]. By gaining more insights into the use of technology and by connecting it to outcomes, the black box of eHealth can be made more transparent. To illustrate, if an RCT shows no significant effects for an intervention, evaluation of uptake can be used to explain if this is because the intervention itself does not work or if (an element of) the intervention was simply not used. To incorporate these types of data in evaluations, it is important to carefully determine how data are collected and stored from the start of the development process [90]. Furthermore, this type of evaluation also provides points of improvement for implementation in practice or the design of an eHealth technology.

Investigating Relevant Working Mechanisms That Explain Why an eHealth Intervention is Effective or Not for Its Users

Understanding why an eHealth technology achieves a certain effect, rather than "just" evaluating *if* it achieves an effect, can provide generalizable insights into how health behavior change can be achieved via an intervention [87]. This knowledge can be used to make future interventions more effective or targeted. There are many different concepts that can be used to gain insights into working mechanisms, for example, related to sociodemographic characteristics of its users, such as age or personality, but ideally, these concepts are more overarching. Examples are previously mentioned BCTs and persuasive features or broader concepts such as adherence, the extent to which a technology is used as intended [91], and engagement, the extent to which someone is involved or occupied with an eHealth technology from an emotional, behavioral, and cognitive perspective [48]. There are several methods that can be used to gain more insights into the working mechanisms of eHealth technologies, such as mixed methods approaches with in-depth qualitative research or (fractional) factorial designs in which different versions of an intervention (eg, one with and one without a specific BCT or persuasive feature) are evaluated. It is important to note that identifying working mechanisms is extremely difficult because there are many combinations of factors that influence the effectiveness or ineffectiveness of a technology, such as characteristics of a specific context, the technology itself, or a broad range of individual factors.

Formative Evaluation

Overview

While the previously described phases are visualized as blocks in the CeHRes Roadmap 2.0, formative evaluation is represented by the arrows within and between these blocks, also see Figure 1. This is in line with the nature of formative evaluation; it is a principle that is intertwined in and part of all phases of the Roadmap. The basic assumption of formative evaluation is that information on how to improve the process and the eHealth technology itself is continuously collected. This information assists the project team in ensuring that there is a constant focus on the context and people involved and can be referred to as "creating by evaluating." A distinction can be made in formative evaluation within activities in a phase, related to ensuring that outcomes are aligned with perspectives of stakeholders and contexts, and between activities, referring to ensuring consistency between outcomes of different activities and the different phases.

Using Methods to Gather Information From the Stakeholders and Context to Continuously Account for Their Perspectives Within the Activities and Phases of the CeHRes Roadmap 2.0

Formative evaluation within activities and phases emphasizes the importance of gathering stakeholder input and verifying outcomes with them to ensure that products fit the involved people and context. For example, requirements are verified by stakeholders in the Value Specification phase, usability tests of prototypes are executed in the Design phase, and stakeholder input is gathered to create and finetune a business model in the Operationalization phase.

Checking Whether the Outcomes of Previous Phases of the CeHRes Roadmap 2.0 Have Been Accounted for in the Current Phase and Ensuring That the Outcomes of all Phases Form a Coherent Whole

Formative evaluations between phases ensure consistency and a clear relationship between the outcomes of all phases. For example, in the Design phase, the project team makes sure that the designed technology incorporates the requirements and addresses the values. This can be done through focus groups with stakeholders or project team meetings. There are no specific methods that must be used, but thorough, transparent documentation of activities and outcomes is essential. If no clear documentation is available, it is hard to retrieve what the main outcomes of previous phases were and what the grounds for specific decisions were.

Changes Compared to the Original CeHRes Roadmap

On the basis of the revised pillars and researchers' experiences, adaptations were made to the objectives, frameworks, and methods within the phases of the Roadmap: contextual inquiry, value specification, design, operationalization, summative evaluation, and formative evaluation. Generally, the objectives of the phases were fine tuned and clarified throughout. Notable modifications include that first, more attention is paid to creating

XSL•FO
RenderX

a theoretical foundation for eHealth technologies [14,92,93]. Among other things, the importance of the use of evidence-based BCTs, integration of domain-specific theories, and connection to existing treatment protocols is emphasized. Second, the role of implementation is further elaborated on highlighting the importance of integrating technology in practice to create impact [14,43,89]. For example, implementation frameworks are now described in the Operationalization phase, and the importance of implementation during development and evaluation is further explained. Third, more attention is paid to novel evaluation methods and approaches [16,94]. To illustrate, attention is paid to methods for impact and uptake evaluations, but also to methods for identifying underlying working mechanisms. In addition to that, evaluation models that can be used to shape multimethod, iterative processes are now introduced.

Discussion

Overview

In this paper, the CeHRes Roadmap 2.0 was introduced and substantiated based on experiences with the Roadmap and new developments within disciplines that underpin it. The CeHRes Roadmap 2.0 consists of the same phases as the first version, but their objectives have been adapted, reflecting the increased emphasis on BCT, implementation, and evaluation as a process [14]. The pillars have been revised to better reflect new insights regarding the ongoing and intertwined, holistic, iterative, participatory, and interdisciplinary nature of eHealth development, implementation, and evaluation processes. Consequently, while the foundation of the CeHRes Roadmap 2.0 has remained the same, it has been thoroughly updated to better reflect needs from practice and research. The CeHRes Roadmap 2.0 specifically highlights the importance of interdisciplinary collaboration with other researchers and stakeholders from practice; the need for agile, iterative processes; and the importance of accounting for implementation and evaluation from the start of any development process.

Comparison With Prior Work

The Importance of Flexibility

When using any framework for eHealth development, it is important to use it as a set of guidelines to guide unique processes as opposed to a static step-by-step guideline with predetermined methods and frameworks. Ideally, the choice of methods used depends on the phase in the process, characteristics and skills of participants, and practical preconditions such as time and money and not on the methods which the involved researchers happen to be familiar with [13,29,95]. Such a flexible approach is particularly relevant for the CeHRes Roadmap 2.0 due to its holistic nature, in which context, people, and technology are intertwined, highlighting the importance of matching methods and frameworks to the specific technology, people, and organization that are involved [11]. In addition, the iterative nature of eHealth development in general implies that the choice of a method also depends on the outcomes of a previous phase. This means that while the broad outlines of an eHealth project can be planned, the specific methods and their outcomes cannot be planned in detail in advance [1]. To illustrate, at the start of a development process, it is often not yet known what the specific requirements of stakeholders are and what the to-be-developed eHealth technology will look like. Consequently, there also cannot be a concrete plan for its implementation and evaluation in practice. This approach requires flexibility from researchers and perhaps also a change in mindset, shifting from meticulously planned study designs to a more agile, adaptive approach with multiple interrelated methods [50,51]. Of course, to ensure transparency, replicability, and robustness, these smaller steps should be documented well, for example, by publishing protocols in advance on open-access platforms. This mindset shift toward agile science is also relevant for many organizations and funding agencies that often desire or require a specific research plan for multiple years with detailed descriptions of products or outcomes. Fortunately, this agile approach to research is gaining ground, and with this revised version of the Roadmap, we hope to contribute to this much-needed shift. To conclude, while for overview purposes, eHealth frameworks such as the Roadmap are visualized as a sequential set of phases, and specific examples of methods and frameworks are provided, this does not mean that this process should be followed strictly. The exact way of shaping eHealth development, implementation, and evaluation processes is the task of an interdisciplinary project team and requires constant weighing and re-evaluating the aims and limitations of a specific project.

Interdisciplinary Collaboration

To tackle wicked societal problems, researchers from different disciplines have to join forces. Consequently, multi- or even interdisciplinary approaches are necessary to address complex, multifaceted challenges, such as developing, implementing, and evaluating effective and efficient eHealth technologies to address problems in health care. eHealth research is often multidisciplinary, in which input is provided by different disciplines, but this is done rather independently from each other, that is, there is not much integration of different methods, approaches, or theories, and people keep working within the boundaries of their discipline. However, to capture the complexity of eHealth, an interdisciplinary approach is advocated [12,14,22,41], in which input from different disciplines is actively integrated, creating a synergy that goes beyond the sum of the individual, disciplinary contributions [59]. Due to the integration of different disciplines, the CeHRes Roadmap 2.0 can be considered an interdisciplinary framework [11]. An example is the combination of BCTs (psychology), persuasive features (persuasive design), and their integration into requirements (engineering and human-centered design) within the Value Specification phase. To further complicate the already intricate task of interdisciplinary work, many researchers also advocate active collaboration with experts by experience such as patients and health care professionals, and other stakeholders such as technology developers or policy makers [12,14,44]. Consequently, people with different backgrounds and points of view have to actively work together, combine their different perspectives, and be willing to look beyond the borders of their paradigm. An important precondition for this is that eHealth researchers have to be able to speak the language of other disciplines [96]. For example, while a psychologist

XSL•FO
RenderX

does not have to fully master programming skills, they do have to understand its basics to effectively communicate with developers and prevent misunderstandings. On top of that, eHealth researchers also have to be able to adapt their communication to fit the specific skills of specific groups of stakeholders, such as vulnerable, hard-to-involve patient populations [66]. It is also recommended that researchers be acquainted with and sometimes even skilled in using different types of qualitative and quantitative research methods from different disciplines to be able to select the most appropriate method or at least understand the rationale behind a decision for a method. The domain of eHealth—including the Roadmap—is already shifting from interdisciplinary to transdisciplinary collaboration, where traditional disciplinary boundaries fade even more, and entirely new approaches are developed. This transdisciplinary process also involves close collaboration with community partners on a shared, "wicked" problem [97]. Particularly when working in a transdisciplinary way, researchers should have the ability to look beyond the boundaries of their discipline and really collaborate with societal stakeholders to contribute to a collaboration that is more than the sum of its parts.

Continuous Improvement of the CeHRes Roadmap

eHealth technologies and their development, implementation, and evaluation are never really finished. The same goes for eHealth frameworks such as the CeHRes Roadmap 2.0. New insights that urge their further adaptation will arise during the upcoming years not only because of developments of theories and models that are part of the interdisciplinary Roadmap and the fast pace of technological innovations that open up new arrays of opportunities and challenges but also because the Roadmap's application to practice will result in points of further improvement. To constantly improve frameworks such as the CeHRes Roadmap 2.0 in a systematic way, there is a need for studies in which multimethod development, implementation, and evaluation processes are presented and reflected upon [13,95]. Examples are reflections on the application and suitability of specific combinations of research methods or lessons learned about the operationalization of specific frameworks in a specific health care setting [29]. However, to date, most studies related to eHealth development, implementation, or evaluation only describe the outcomes of a single study, for example, an RCT. Outcomes from smaller studies, such as single-case studies, pilot studies, or qualitative research, that were used to set up the larger study are not reported, either because these types of studies were not performed at all or because authors (or journals) do not feel that publishing them is relevant or important [13,14,98]. However, just as is the case for eHealth technologies, input from stakeholders with different backgrounds is required to constantly improve the Roadmap to ensure that it fits the needs and wishes of its end users. Consequently, there is a need for more multimethod studies that report on the application of (large parts of) eHealth frameworks, critically reflect on their application, and provide points of improvement. To achieve this, future research can focus on creating a standardized way of reporting on multimethod eHealth processes [13,14] comparable to guidelines, for example, systematic reviews [99], RCTs [100], or qualitative research [101], of which an overview is provided by the EQUATOR Network.

Limitations

Despite its usefulness for eHealth development, implementation, and evaluation, the CeHRes Roadmap 2.0 has some limitations. First, the update was not based on a new systematic review of the literature. Yet, the initial version of the Roadmap was based on a comprehensive review of relevant models and frameworks, and a preliminary exploration of the literature indicated that a repetition of this review would likely not lead to any major new insights. Furthermore, although no new review was conducted, we did ground the CeHRes Roadmap 2.0 on reviews about important elements of the Roadmap, for example, on participatory development [13], implementation [43], evaluation [16], or eHealth frameworks in general [14]. In addition, models and frameworks that were already part of the Roadmap were updated based on new insights, resulting in a comprehensive updated version that covers the most important points of improvement. Regardless of these efforts, the fact remains that, as mentioned before, new suggestions and points of improvement will arise, warranting multiple future recurring updates of the Roadmap.

Second, currently, the Roadmap has mostly been used for the development, evaluation, and implementation of digital health interventions for persons and health care providers. Therefore, the revisions within this second edition are also based mostly on experiences with the development processes of these types of eHealth technologies. Although in principle, the Roadmap 2.0 is also applicable to, for example, the development of digital health interventions for health management and support personnel, and for data services, its value for these areas should be studied more thoroughly.

Another point of attention is that using the CeHRes Roadmap 2.0 does not guarantee the development of effective eHealth interventions that will be widely used in practice. Development, implementation, and evaluation are all extremely complex processes, and while the Roadmap can help to improve them and provide guidance, failures or undesired outcomes might (and will) still occur. In line with this, it is neither possible nor feasible to investigate if using the Roadmap results in better eHealth technologies than using other frameworks such as the person-based approach [53], the Accelerated Creation-to-Sustainment model [44], intervention mapping [45,92], or agile science approaches [50,51]. A comparison of frameworks would require the evaluation of multiple parallel, similar processes. Such similarities are practically impossible to achieve because of the very nature of the CeHRes Roadmap 2.0 and related frameworks and their application to practice, which are all highly dependent on research teams, contexts, objectives, and even periods in time. This makes reliable cross comparisons virtually impossible. Therefore, gaining insights into the added value of frameworks like the Roadmap mostly requires sharing experiences with and lessons learned from its application, which again emphasizes the importance of systematically reporting and reflecting on development, implementation, and evaluation processes, in line with reporting guidelines [13,14,31].

XSL•FO
RenderX

Conclusions

This study addresses the need for an up-to-date version of the CeHRes Roadmap using the most recent insights and most relevant approaches and models. The changes made in the updated version of the CeHRes Roadmap 2.0 compared to its original version address the need for more attention to behavior change theory in eHealth design; the increased emphasis on implementation; and the importance of iterative, multimethod evaluation processes. To ensure that eHealth frameworks are optimally aligned with experiences and best practices from researchers, there is a need for more studies that apply and reflect on the application of these frameworks because, just as any eHealth technology the CeHRes Roadmap 2.0 has to be constantly improved based on the input of its users.

Authors' Contributions

All authors contributed to the content of the manuscript, that is, the update of the CeHRes Roadmap. The first version of the manuscript was written by HK and NBdJ, and SMK and LJEWCvG-P gave feedback. All authors approved the final version of this manuscript.

Conflicts of Interest

None declared.

References

1. Kip H, van Gemert-Pijnen JE, Beerlage-de Jong N, Sanderman R, Kelders SM. eHealth Research Theory and Development: A Multidisciplinary Approach. Milton Park, UK: Taylor & Francis; 2024.
2. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
3. Bargeri S, Scalea S, Agosta F, Banfi G, Corbetta D, Filippi M, et al. Effectiveness and safety of virtual reality rehabilitation after stroke: an overview of systematic reviews. EClinicalMedicine 2023 Sep 14;64:102220 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102220] [Medline: 37745019]
4. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
5. Puleo V, Gentili A, Failla G, Melnyk A, Di Tanna G, Ricciardi W. Digital health technologies: a systematic review of their cost-effectiveness. Eur J Public Health 2021 Oct;31(Supplement_3):iii103-iii104. [doi: 10.1093/eurpub/ckab164.273]
6. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of eHealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
7. Recommendations on digital interventions for health system strengthening: WHO Guideline. World Health Organization. 2019 Jun 06. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-12-05]
8. Weiner JP, Yeh S, Blumenthal D. The impact of health information technology and e-health on the future demand for physician services. Health Aff (Millwood) 2013 Nov;32(11):1998-2004. [doi: 10.1377/hlthaff.2013.0680] [Medline: 24191092]
9. Ahmed B, Dannhauser T, Philip N. A systematic review of reviews to identify key research opportunities within the field of eHealth implementation. J Telemed Telecare 2019 Jun;25(5):276-285. [doi: 10.1177/1357633X18768601] [Medline: 29703127]
10. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
11. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
12. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
13. Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders SM. Methods for human-centered eHealth development: narrative scoping review. J Med Internet Res 2022 Jan 27;24(1):e31858 [FREE Full text] [doi: 10.2196/31858] [Medline: 35084359]
14. Cruz-Martínez RR, Wentzel J, Asbjørnsen RA, Noort PD, van Niekerk JM, Sanderman R, et al. Supporting self-management of cardiovascular diseases through remote monitoring technologies: metaethnography review of frameworks, models, and theories used in research and development. J Med Internet Res 2020 May 21;22(5):e16157 [FREE Full text] [doi: 10.2196/16157] [Medline: 32436852]

XSL•FO
RenderX

15. Varsi C, Solberg Nes L, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
16. Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, et al. Online guide for electronic health evaluation approaches: systematic scoping review and concept mapping study. J Med Internet Res 2020 Aug 12;22(8):e17774 [FREE Full text] [doi: 10.2196/17774] [Medline: 32784173]
17. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
18. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
19. Esser PE, Goossens RH. A framework for the design of user-centred teleconsulting systems. J Telemed Telecare 2009;15(1):32-39. [doi: 10.1258/jtt.2008.080601] [Medline: 19139218]
20. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009 Aug;6(8):e1000126 [FREE Full text] [doi: 10.1371/journal.pmed.1000126] [Medline: 19688038]
21. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [doi: 10.1016/s0010-4825(02)00011-2] [Medline: 11922931]
22. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
23. Kukafka R, Johnson SB, Linfante A, Allegrante JP. Grounding a new information technology implementation framework in behavioral science: a systematic analysis of the literature on IT use. J Biomed Inform 2003 Jun;36(3):218-227 [FREE Full text] [doi: 10.1016/j.jbi.2003.09.002] [Medline: 14615230]
24. Kip H, van Gemert-Pijnen JE. Holistic development of eHealth technology. In: eHealth Research, Theory and Development. London, UK: Routledge; 2018.
25. van Gemert-Pijnen JE, Peters O, Ossebaard HC. Improving eHealth. The Hague, The Netherlands: Eleven International Publishing; 2013.
26. Kloek CJ, van Tilburg ML, Staal JB, Veenhof C, Bossen D. Development and proof of concept of a blended physiotherapeutic intervention for patients with non-specific low back pain. Physiotherapy 2019 Dec;105(4):483-491 [FREE Full text] [doi: 10.1016/j.physio.2018.12.006] [Medline: 31031023]
27. Austin J, Drossaert CH, van Dijk J, Sanderman R, Børøsund E, Mirkovic J, et al. Integrating top-down and bottom-up requirements in eHealth development: the case of a mobile self-compassion intervention for people with newly diagnosed cancer. JMIR Cancer 2022 Aug 01;8(3):e37502 [FREE Full text] [doi: 10.2196/37502] [Medline: 35916691]
28. Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health, 2nd ed. World Health Organization. 2023 Oct 24. URL: https://www.who.int/publications/i/item/9789240081949 [accessed 2024-12-05]
29. Kip H, Kelders SM, Bouman YH, van Gemert-Pijnen JE. The importance of systematically reporting and reflecting on eHealth development: participatory development process of a virtual reality application for forensic mental health care. J Med Internet Res 2019 Aug 19;21(8):e12972 [FREE Full text] [doi: 10.2196/12972] [Medline: 31429415]
30. Wentzel J, van Velsen L, van Limburg M, de Jong N, Karreman J, Hendrix R, et al. Participatory eHealth development to support nurses in antimicrobial stewardship. BMC Med Inform Decis Mak 2014 Jun 05;14:45 [FREE Full text] [doi: 10.1186/1472-6947-14-45] [Medline: 24898694]
31. Kelders SM, Pots WT, Oskam MJ, Bohlmeijer ET, van Gemert-Pijnen JE. Development of a web-based intervention for the indicated prevention of depression. BMC Med Inform Decis Mak 2013 Feb 20;13:26 [FREE Full text] [doi: 10.1186/1472-6947-13-26] [Medline: 23425322]
32. van Limburg M, Wentzel J, Sanderman R, van Gemert-Pijnen JE. Business modeling to implement an eHealth portal for infection control: a reflection on co-creation with stakeholders. JMIR Res Protoc 2015 Aug 13;4(3):e104 [FREE Full text] [doi: 10.2196/resprot.4519] [Medline: 26272510]
33. Beerlage-de Jong N, Eikelenboom-Boskamp A, Voss A, Sanderman R, van Gemert-Pijnen JE. Combining user-centered design with the persuasive systems design model; the development process of a web-based registration and monitoring system for healthcare-associated infections in nursing homes. Int J Adv Life Sci 2014;6(3&4):262-271.
34. Kelders SM, Bohlmeijer ET, Pots WT, van Gemert-Pijnen JE. Comparing human and automated support for depression: fractional factorial randomized controlled trial. Behav Res Ther 2015 Sep;72:72-80. [doi: 10.1016/j.brat.2015.06.014] [Medline: 26196078]
35. Sieverink F, Kelders S, Braakman-Jansen A, van Gemert-Pijnen JE. Evaluating the implementation of a personal health record for chronic primary and secondary care: a mixed methods approach. BMC Med Inform Decis Mak 2019 Nov 27;19(1):241 [FREE Full text] [doi: 10.1186/s12911-019-0969-7] [Medline: 31775734]

XSL•FO
RenderX

36. Kip H, Buitelaar-Huijsse GK, Kouijzer MT, Kelders SM. From theory to implementation in practice: a qualitative case study of the implementation of virtual reality in mental healthcare. Glob Implement Res Appl 2023 Oct 25;4:66-88. [doi: 10.1007/s43477-023-00101-7]
37. Kip H, Sieverink F, van Gemert-Pijnen JE, Bouman YH, Kelders SM. Integrating people, context, and technology in the implementation of a web-based intervention in forensic mental health care: mixed-methods study. J Med Internet Res 2020 May 26;22(5):e16906 [FREE Full text] [doi: 10.2196/16906] [Medline: 32348285]
38. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
39. Asbjørnsen RA, Wentzel J, Smedsrød ML, Hjelmesæth J, Clark MM, Solberg Nes L, et al. Identifying persuasive design principles and behavior change techniques supporting end user values and needs in eHealth interventions for long-term weight loss maintenance: qualitative study. J Med Internet Res 2020 Nov 30;22(11):e22598 [FREE Full text] [doi: 10.2196/22598] [Medline: 33252347]
40. van Woezik AF, Braakman-Jansen LM, Kulyk O, Siemons L, van Gemert-Pijnen JE. Tackling wicked problems in infection prevention and control: a guideline for co-creation with stakeholders. Antimicrob Resist Infect Control 2016 May 21;5:20 [FREE Full text] [doi: 10.1186/s13756-016-0119-2] [Medline: 27213040]
41. Keizer J, Bente BE, Al Naiemi N, Van Gemert-Pijnen JE, Beerlage-De Jong N. Improving the development and implementation of audit and feedback systems to support health care workers in limiting antimicrobial resistance in the hospital: scoping review. J Med Internet Res 2022 Mar 11;24(3):e33531 [FREE Full text] [doi: 10.2196/33531] [Medline: 35275082]
42. Köhle N, Drossaert CH, Ten Klooster PM, Schreurs KM, Hagedoorn M, Van Uden-Kraan CF, et al. Web-based self-help intervention for partners of cancer patients based on acceptance and commitment therapy and self-compassion training: a randomized controlled trial with automated versus personal feedback. Support Care Cancer 2021 Sep;29(9):5115-5125 [FREE Full text] [doi: 10.1007/s00520-021-06051-w] [Medline: 33608762]
43. Kouijzer MM, Kip H, Bouman YH, Kelders SM. Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. Implement Sci Commun 2023 Jun 16;4(1):67 [FREE Full text] [doi: 10.1186/s43058-023-00442-2] [Medline: 37328858]
44. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
45. Eldredge LK, Parcel GS, Kok G, Gottlieb NH, Fernández ME. Planning Health Promotion Programs: An Intervention Mapping Approach. Hoboken, NJ: John Wiley & Sons; 2011.
46. Nielsen JA, Mathiassen L. Interpretive flexibility in mobile health: lessons from a government-sponsored home care program. J Med Internet Res 2013 Oct 30;15(10):e236 [FREE Full text] [doi: 10.2196/jmir.2816] [Medline: 24172852]
47. James J. The Content of Social Explanation. New York, NY: Cambridge University Press; 1984.
48. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020 May 27;11. [doi: 10.3389/fpsyg.2020.00926]
49. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
50. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun 26;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
51. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
52. Dingsøyr T, Nerur S, Balijepally V, Moe NB. A decade of agile methodologies: towards explaining agile software development. J Syst Softw 2012 Jun;85(6):1213-1221. [doi: 10.1016/j.jss.2012.02.033]
53. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
54. Beerlage-de Jong N, Wentzel J, Hendrix R, van Gemert-Pijnen JE. The value of participatory development to support antimicrobial stewardship with a clinical decision support system. Am J Infect Control 2017 Apr 01;45(4):365-371. [doi: 10.1016/j.ajic.2016.12.001] [Medline: 28089673]
55. Oberschmidt K, Grünloh C, Nijboer F, van Velsen L. Best practices and lessons learned for action research in eHealth design and implementation: literature review. J Med Internet Res 2022 Jan 28;24(1):e31795 [FREE Full text] [doi: 10.2196/31795] [Medline: 35089158]
56. Sein MK, Henfridsson O, Purao S, Rossi M, Lindgren R. Action design research. MIS Q 2011;35(1):37-56. [doi: 10.2307/23043488]

XSL•FO
RenderX

57. Scaife M, Rogers Y, Aldrich F, Davies M. Designing for or designing with? Informant design for interactive learning environments. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. 1997 Presented at: CHI '97; March 22-27, 1997; Atlanta, GA. [doi: 10.1145/258549.258789]
58. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 05;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]
59. Aboelela SW, Larson E, Bakken S, Carrasquillo O, Formicola A, Glied SA, et al. Defining interdisciplinary research: conclusions from a critical review of the literature. Health Serv Res 2006 Aug 17;42(1p1):329-346. [doi: 10.1111/j.1475-6773.2006.00621.x]
60. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019 Mar 12;5(1):41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
61. Feldman SS, Schooley BL, Bhavsar GP. Health information exchange implementation: lessons learned and critical success factors from a case study. JMIR Med Inform 2014 Aug 15;2(2):e19 [FREE Full text] [doi: 10.2196/medinform.3455] [Medline: 25599991]
62. Karen H, Sandra J. Contextual Inquiry: A Participatory Technique for System Design. Boca Raton, FL: CRC Press; 1993.
63. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
64. Mitchell RK, Agle BR, Wood DJ. Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. Acad Manag Rev 1997 Oct;22(4):853-886. [doi: 10.2307/259247]
65. Kip H, Kelders SM, Van Gemert-Pijnen JE. Putting the value in VR: how to systematically and iteratively develop a value-based vr application with a complex target group. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300365]
66. Schouten SE, Kip H, Dekkers T, Deenik J, Beerlage-de Jong N, Ludden GD, et al. Best-practices for co-design processes involving people with severe mental illness for eMental health interventions: a qualitative multi-method approach. Des Health 2022 Nov 26;6(3):316-344. [doi: 10.1080/24735132.2022.2145814]
67. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug 20;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
68. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24. [doi: 10.17705/1CAIS.02428]
69. Ten Klooster I, Wentzel J, Sieverink F, Linssen G, Wesselink R, van Gemert-Pijnen JE. Personas for better targeted eHealth technologies: user-centered design approach. JMIR Hum Factors 2022 Mar 15;9(1):e24172 [FREE Full text] [doi: 10.2196/24172] [Medline: 35289759]
70. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
71. Rosson MB, Carroll JM. Scenario-based design. In: Jacko JA, editor. Human Computer Interaction Handbook. Boca Raton, FL: CRC Press; 2009.
72. Osterwalder A, Pigneur Y, Clark T. Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Hoboken, NJ: John Wiley & Sons; 2010.
73. Framework for innovation. Design Council. URL: https://www.designcouncil.org.uk/our-resources/framework-for-innovation [accessed 2024-12-03]
74. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021 Jan 12;33(Supplement_1):37-44 [FREE Full text] [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]
75. Principles for Digital Development homepage. Principles for Digital Development. URL: https://digitalprinciples.org/ [accessed 2024-02-10]
76. Newman MW, Landay JA. Sitemaps, storyboards, and specifications: a sketch of Web site design practice. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Technique. 2000 Presented at: DIS '00; August 17-19, 2000; New York, NY. [doi: 10.1145/347642.347758]
77. Beerlage-de Jong N, Kip H, Kelders SM. Evaluation of the perceived persuasiveness questionnaire: user-centered card-sort study. J Med Internet Res 2020 Oct 23;22(10):e20404 [FREE Full text] [doi: 10.2196/20404] [Medline: 33095173]
78. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353 [FREE Full text] [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
79. Nielsen J. How to conduct a heuristic evaluation. Nielsen Norman Group 1995 Jan 1;1(1):8.
80. Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough method: a practitioner's guide. In: Nielsen J, Mack RL, editors. Usability Inspection Methods. Hoboken, NJ: John Wiley & Sons; 1994:105-140.

XSL•FO
RenderX

81. Pieterse M, Kip H, Cruz-Martinez RR. The complexity of eHealth Implementation: a theoretical and practical perspective. In: Kip H, Van Gemert-Pijnen JE, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development: A Multidisciplinary Approach. London, UK: Routledge; 2018:247-270.
82. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
83. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
84. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
85. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev 2012 Apr 26;69(2):123-157 [FREE Full text] [doi: 10.1177/1077558711430690] [Medline: 22203646]
86. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
87. Moore GF, Evans RE. What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions. SSM Popul Health 2017 Dec;3:132-135 [FREE Full text] [doi: 10.1016/j.ssmph.2016.12.005] [Medline: 29302610]
88. Salter KL, Kothari A. Using realist evaluation to open the black box of knowledge translation: a state-of-the-art review. Implement Sci 2014 Sep 05;9(1):115 [FREE Full text] [doi: 10.1186/s13012-014-0115-y] [Medline: 25190100]
89. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
90. Sieverink F, Kelders SM, Poel M, van Gemert-Pijnen JE. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
91. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
92. Crutzen R. The behavioral intervention technology model and intervention mapping: the best of both worlds. J Med Internet Res 2014 Aug 05;16(8):e188 [FREE Full text] [doi: 10.2196/jmir.3620] [Medline: 25095730]
93. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb 5;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
94. Sieverink F, Köhle N, Cheung K, Roefs A, Keizer J. Evaluating eHealth. In: van Gemert-Pijnen JE, Kelders SM, Kip H, Sanderman R, editors. eHealth Research Theory and Development: A Multidisciplinary Approach. Milton Park, UK: Taylor & Francis; 2018.
95. Beerlage-de Jong N, van Gemert-Pijnen JEWC, Wentzel J, Hendrix R, Siemons L. Technology to support integrated antimicrobial stewardship programs: a user centered and stakeholder driven development approach. Infect Dis Rep 2017 Mar 30;9(1):6829 [FREE Full text] [doi: 10.4081/idr.2017.6829] [Medline: 28458796]
96. Nancarrow SA, Booth A, Ariss S, Smith T, Enderby P, Roots A. Ten principles of good interdisciplinary team work. Hum Resour Health 2013 May 10;11(1):19 [FREE Full text] [doi: 10.1186/1478-4491-11-19] [Medline: 23663329]
97. Boger J, Jackson P, Mulvenna M, Sixsmith J, Sixsmith A, Mihailidis A, et al. Principles for fostering the transdisciplinary development of assistive technologies. Disabil Rehabil Assist Technol 2017 Jul 07;12(5):480-490. [doi: 10.3109/17483107.2016.1151953] [Medline: 27052793]
98. Wang RH, Kenyon LK, McGilton KS, Miller WC, Hovanec N, Boger J, et al. The time is now: a FASTER approach to generate research evidence for technology-based interventions in the field of disability and rehabilitation. Arch Phys Med Rehabil 2021 Sep;102(9):1848-1859. [doi: 10.1016/j.apmr.2021.04.009] [Medline: 33992634]
99. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
100. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother 2010 Dec 01;1(2):100-107. [doi: 10.4103/0976-500x.72352]
101. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (Consolidated Criteria for Reporting Qualitative Studies). In: Moher D, Altman DG, Schulz KF, Simera I, Wager E, editors. Guidelines for Reporting Health Research: A User's Manual. Hoboken, NJ: John Wiley & Sons; 2014.

XSL•FO
RenderX

Abbreviations

BCT: behavior change technique
CeHRes: Centre for eHealth and Wellbeing Research
CFIR: Consolidated Framework for Implementation Research
MOST: Multiphase Optimization Strategy
NASSS: Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability
PSD: Persuasive Systems Design
RCT: randomized controlled trial
SMART: specific, measurable, achievable, realistic, and timely

Edited by A Schwartz; submitted 17.04.24; peer-reviewed by K Gray, N Bier; comments to author 05.09.24; revised version received 07.10.24; accepted 13.11.24; published 13.01.25.

Please cite as:
Kip H, Beerlage-de Jong N, van Gemert-Pijnen LJEWC, Kelders SM
The CeHRes Roadmap 2.0: Update of a Holistic Framework for Development, Implementation, and Evaluation of eHealth Technologies
J Med Internet Res 2025;27:e59601
URL: https://www.jmir.org/2025/1/e59601
doi:10.2196/59601
PMID:

©Hanneke Kip, Nienke Beerlage-de Jong, Lisette J E W C van Gemert-Pijnen, Saskia M Kelders. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

From Theory to Practice: Viewpoint on Economic Indicators for Trust in Digital Health

Felix Gille[1,2*], PhD; Laura Maaß[3,4*], PhD; Benjamin Ho[5*], PhD; Divya Srivastava[6,7*], PhD

[1]University of Zurich, Digital Society Initiative, Zurich, Switzerland

[2]University of Zurich, Institute for Implementation Science in Health Care, Zurich, Switzerland

[3]SOCIUM Research Center on Inequality and Social Policy, University of Bremen, Bremen, Germany

[4]Leibniz ScienceCampus Digital Public Health, Bremen, Germany

[5]Department of Economics, Vassar College, Poughkeepsie, NY, United States

[6]Department of Health Services Research Management, AI and Digital Health Lab (Centre for Healthcare Innovation Research), City St George's University, London, United Kingdom

[7]Department of Health Policy, London School of Economics and Political Science, London, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Felix Gille, PhD
University of Zurich
Digital Society Initiative
Rämistrasse 69
Zurich, 8001
Switzerland
Phone: 41 44635 7133
Email: felix.gille@uzh.ch

Abstract

User trust is pivotal for the adoption of digital health systems interventions (DHI). In response, numerous trust-building guidelines have recently emerged targeting DHIs such as artificial intelligence. The common aim of these guidelines aimed at private sector actors and government policy makers is to build trustworthy DHI. While these guidelines provide some indication of what trustworthiness is, the guidelines typically only define trust and trustworthiness in broad terms, they rarely offer guidance about economic considerations that would allow implementers to measure and balance trade-offs between costs and benefits. These considerations are important when deciding how best to allocate scarce resources (eg, financial capital, workforce, or time). The missing focus on economics undermines the potential usefulness of such guidelines. We propose the development of actionable trust-performance-indicators (including but not limited to surveys) to gather evidence on the cost-effectiveness of trust-building principles as a crucial step for successful implementation. Furthermore, we offer guidance on navigating the conceptual complexity surrounding trust and on how to sharpen the trust discourse. Successful implementation of economic considerations is critical to successfully build user trust in DHI.

(J Med Internet Res 2025;27:e59111) doi:10.2196/59111

KEYWORDS

trust; economics; digital health; digital health innovation; artificial intelligence; AI; economic evaluation; public trust; health data; medical apps

Introduction

Public trust is a cornerstone of successful health systems [1,2]. Public trust is a key enabler for the introduction of trustworthy health data spaces for sharing and using health data in research, national electronic health record systems and the introduction of artificial intelligence (AI) systems in medicine. In 2023, the World Health Organization (WHO) in the European Region stated that 77% of its Member States believe the lack of trustworthy sources to access effective medical apps is a substantial implementation barrier to digital health transformation: "Trust is paramount – trust in technology, trust in data protection and, most critically, trust in the respect and safeguarding of privacy rights pertaining to personal health data" [3]. Trust in public health systems legitimizes these activities. Public trust is predicated on the belief that those

XSL•FO
RenderX

systems provide a net benefit, where benefits include individual benefit, societal benefit, and health system benefit [4]. The increasing discourse on trust and trustworthiness of digital health systems interventions (DHI) has been met with an increase in nonbinding professional guidelines and codes of conduct. These guidelines are designed to guide users toward the ethical and trustworthy use of DHIs. Adhering to such guidelines can help foster public trust which, in turn, can foster their practical adoption. Some purposefully sampled guidelines in the area of AI as a prominent example of DHIs are highlighted in Table 1. The guidelines were selected to make sure national and international as well as government and nongovernment organizations were represented but we fully acknowledge that these reflect only a subset of the guidelines that exist.

Table 1. Overview of example guidelines about the trustworthy use of health data and artificial intelligence in relation to digital health system intervention.

Guideline	Providing organization	Context	Relevance for trust in digital health
Ethics guidelines for trustworthy Artificial intelligence [5]	European Commission	Strengthening national systems (including the health system) through artificial intelligence	"Trustworthy AI technologies can be used – and are already being used – to render treatment smarter and more targeted, and to help preventing life-threatening diseases. Doctors and medical professionals can potentially perform a more accurate and detailed analysis of a patient's complex health data, even before people get sick, and provide tailored preventive treatment. […]"
Digital Trust Framework [6]	World Economic Forum	Companies willing to commit to earning digital trust of citizens and customers	"A technology provider's reliability and security can have a significant impact on individuals' safety, including on their physical or mental health. […]"
Global strategy on digital health 2020-2025 [7]	World Health Organization	Strengthening national health systems through digital health technologies	"The appropriate use of digital health takes the following dimensions into consideration: health promotion and disease prevention, patient safety, ethics, interoperability, intellectual property, data security (confidentiality, integrity, and availability), privacy, cost-effectiveness, patient engagement, and affordability. It should be people-centred, trust-based, evidence-based, effective, efficient, sustainable, inclusive, equitable and contextualized. […]"
Guidance on the rights-based and ethical use of digital technologies in HIV and health programs [8]	United Nations	Human rights, technical, and ethical considerations for countries adopting digital technologies for health and HIV	"A core component of successful HIV responses, trust is also a key element in the adoption and success of digital technologies for health. Without trust, the implementation and uptake of digital health will be weak, even if all other aspects of infrastructure, as well as legal and regulatory frameworks, are effective. Trust must be built for all types of relationships related to digital health – between patients and health-care providers; within the health sector institutions; between the State and its residents; and between a State, its residents and the private sector. […] Data breaches violate an individual's right to privacy and erode trust in the health-care system. As technology evolves and health systems become more complex, the likelihood of data breach occurrences increase. Health systems should invest in information security and keep up to date on the latest in data protection to prevent breaches."
A guide to good practice for digital and data-driven health technologies [9]	United Kingdom Department of Health and Social Care	Guiding document on "good practice" for digital health technology innovators in the United Kingdom	"When building an algorithm, be it a stand-alone product or integrated within a system, show it clearly and be transparent of the learning methodology (if any) that the algorithm is using. Undertake ethical examination of data use specific to this use-case. Achieving transparency of algorithms that have a higher potential for harm or unintended decision-making, can ensure the rights of the data subject as set out in the Data Protection Act 2018 are met, to build trust in users and enable better adoption and uptake."
Code of Ethics for Data-Based Value Creation [10]	Swiss Alliance for Data-Intensive Services	Companies and organizations that offer products or services based on data	"Data management is networked. The quality of companies' data-based products and services depends in many ways on other organisations. Accordingly, it is important that in this network of interdependence that companies can trust each other. […] Transparency requires ensuring that customers are informed about all aspects of data handling – storage, management, and protection – thus creating (well-placed) trust."

Such guidelines cover a range of trust-building principles, for example, the need to implement accountability mechanisms, security, or privacy measures, the importance of public and user-centered communication strategies, and the appropriate implementation of the regulation. One example is the Digital Trust Framework published by the World Economic Forum [6] which identifies principles along 8 dimensions: cybersecurity, safety, transparency, interoperability, auditability, redressability, fairness, and privacy. These principles aim to foster security and reliability, accountability, and oversight, as well as inclusive, ethical, and responsible data use for DHIs. The guiding principles are usually at an abstraction level that allows transferability across different contexts, necessitating careful tailoring during the implementation process. In practice,

XSL•FO
RenderX

however, resources are limited. How do we determine which trust-building principles to prioritize, and which digital health policies should take precedence, especially when guidelines conflict? These guidelines are typically silent on the economic implications of trust-building, which could provide guidance on the costs and benefits of implementation. A clear need for economic guidance persists. This observation is underscored by anecdotal observations made by FG in Switzerland in the past few years during workshops and meetings discussing trust in DHIs.

After attending an expert workshop aiming at identifying a list of trust-building principles comprising a trustworthy data space, a fellow participant asked: "This list looks convincing, but I am the only IT (Information Technology) security employee in a small firm, we have limited resources, so which of the principles are cost-effective?" (Workshop on trustworthy behavior, Bern, Switzerland)

While exploring a research collaboration with an industry partner where we aimed to test the performance of a nonbinding trustworthiness codex, the industry counterpart asked: "If our company adheres to a non-binding trustworthiness codex, isn't there a risk that we face a competitive disadvantage compared to our competitors that do not commit to these measures but still adhere to the present legislation?" (Business meeting on the performance of a nonbinding trustworthiness codex, online, Switzerland).

Discussing trustworthiness principles to conceptualize Digital Trust, workshop members from the industry suggested that what we have today is robust conceptual evidence, but we clearly lack implementation experience and evidence showing what trust-building measures are costeffective. (Workshop exploring digital trust, Zurich, Switzerland).

When it comes to the implementation of trust-building measures and the economic evaluation of their effectiveness, we stumble upon a yawning knowledge gap. Three important unanswered questions require our attention: What are the costs of implementing trustworthiness guidelines? With limited resources at hand, which measures are cost effective? What instruments exist to gauge success?

We argue that the failure to address these questions risks undermining both the implementation of trust-building guidelines and their success. What is needed is a clear understanding of trust linked to a robust economic assessment of trust building activities. Furthermore, any economic assessment of the costs and benefits should acknowledge that different stakeholders, such as investors, industry participants, policy makers, or end users, will prioritize these impacts differently.

In this article, we offer direction on how to close this gap by improving the conceptual understanding of trust, implementing actionable trust performance indicators, and sharpening the trust discourse. While the importance of trust in DHI is commonly recognized, we argue that we still need more evidence-based policies to ensure the trustworthiness of these programs are implemented efficiently and effectively. With this viewpoint, we aim to initiate further development of economic thinking in trust building. More specifically we aim to transfer and adapt knowledge from economic research to other domains where trust building matters, but economic considerations have not yet entered the professional discourse and implementation practice of trust building activities.

Shortcomings and Remedies

Trust has been part of economic research for centuries [11]. Ample research exists investigating the relationship between trust and societal prosperity [12], the causal impact of trust on economic development [13,14], and how trust lowers transaction costs [15]. As Arrow [16] stated in 1972, "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence…"

While the existing research provides arguments for the economic incentives for trust-building, the costs and benefits of trust-building in digital health have not been well studied. In some emerging economic evaluation guidelines tailored to DHIs, trust can be a component of an impact inventory in user experiences or considered as a non–health-related impact [17]. Yet, trust as an outcome variable is not a well-established component of economic evaluation of digital health. We argue that the lack of focus on the economics of trust-building in this domain include conceptual and methodological challenges that need to be addressed.

We Need Conceptual Clarity of Trust in Digital Health Interventions as a Prerequisite for Economic Evaluation

Embedded in wider sociopolitical as well as cultural norms and values, the concept of trust is complex and influenced by the context in which it evolves. No commonly accepted definition of trust exists. We suggest that trust in a person or institution is a belief about the trustworthiness of that individual or entity. It is a belief that makes one willing to take a risk and make oneself vulnerable, to rely on that person or institution for one's future benefit. The trustworthiness trait is defined as having characteristics that are conducive to cooperation. What traits count as trustworthy differ by the situation. Still, we can categorize them as either traits relating to competence (the ability to fulfil the tasks expected of them) and traits related to integrity, the commitment to prioritize the best interests of the other [11,18].

Trust is volatile, and while the set of trust-building principles might remain unchanged over time, the relative importance of individual principles can vary depending on the context. In particular, trust repair after health system scandals or negative shifts in societal sentiments toward the government require different approaches [19]. The trait of trustworthiness, often the main target of guidelines, should not be conflated with trust. Trustworthiness is a precondition of trust-building but not a guarantee for trust. Trust-building can be based on calculated decisions and rational thinking [20]. At the same time, a body

XSL•FO
RenderX

of literature highlights the importance of emotions in building trust [21]. Combining calculated and emotional motivations to establish trust broadens the spectrum of possible measures available to policy makers and health system actors. Trust research and guidelines in the digital health domain tend to neglect the emotional motivations of trust-building.

Many guidelines contain a comprehensive portfolio of trust-building principles which reflect the overall complexity of the topic but also reflect the lack evidence about which principles work best in practice. Without considering economic constraints and trade-offs, the guidelines may be biased toward presenting too many trust-building principles that are too broad for fear of missing out, with too little guidance on how to prioritize them. As a result, the guidelines may be perceived by users as overwhelming due to the number of principles included or to their high abstractness, with too much onus placed on the implementer to figure out how to navigate the guidelines.

On the receiving end of the guidelines, we see discussions emerging around which organizational units are most suitable to take on the implementation task. The broad spectrum of trust-building principles, from user centered communication and IT security measures to adherence to the rule of law, places responsibility on different organizational units to lead the implementation of guidelines in areas as diverse as public relations, law, compliance, and IT services. Table 2 provides handpicked examples of specific trust-building activities recommended by the guidelines presented in Table 1 that we categorize about accountability for past mistakes, transparency about the present, and promises and benchmarks for the future [11].

XSL•FO
RenderX

Table 2. Examples of what trustworthiness guidelines call for.

Guideline	Example trust-building activities according to this guideline		
	Accountability for past mistakes	Transparency in the present	Promises and benchmarks for the future
Ethics guidelines for trustworthy AI[a] [5]	• Mechanisms should be put in place to ensure responsibility and accountability for AIa systems and their outcomes. • Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. • Furthermore, adequate and accessible redress should be ensured.	• The data, system and AIa business models should be transparent. Traceability mechanisms can help achieve this. • AIa systems and their decisions should be explained in a manner adapted to the stakeholder concerned. • Humans need to be aware that they are interacting with an AIa system and must be informed of the system's capabilities and limitations.	• Adopt a Trustworthy AIa assessment list when developing, deploying or using AIa systems. • Adapt it to the specific use case in which the system is being applied.
Digital Trust Framework [6]	• Good accountability and oversight mechanisms can both remediate any harms that result from technology use. • Auditability serves as a check and is particularly useful in both the development stage and in the deployment stage. • Redressability allows for recourse to address any harms. • Good accountability builds up the trust necessary for more widespread adoption of useful technologies.	• Information is shared about how the technology is developed, implemented and how the data are used. • Design is user-friendly with transparency in mind.	• Adopt best practice frameworks for security and reliability norms. • Digital trust programs that account for the impact of societal expectations.
Global strategy on digital health 2020-2025 [7]	• Develop an ethics framework to strengthen public trust.	• Use of digital health trust based on authentication and authorization that guarantees trust.	• Build trust by placing importance on quality, safety and ethical considerations facilitate evidence-based decisions to improve public trust.
Guidance on the rights-based and ethical use of digital technologies in HIV and health programs [8]	• Establish impartial, effective accountability and oversight mechanisms for breaches of data privacy and other rights violations.	• Be consultative and transparent in decision-making related to governance and management of digital technologies. • Be transparent about and accountable for the factors that influence algorithmic decisions. • Co-design digital technologies and systems with affected communities. • Create opportunities for digital rights literacy for communities and individuals to understand their rights and take ownership of their data, including the right to withdraw their data from use and to data portability.	• Create monitoring and evaluation systems that allow the technologies to adapt based on feedback.
A guide to good practice for digital and data-driven health technologies [9]	• Undertake ethical examination of data use specific to this use-case.	• Be fair, transparent and accountable about what data is being used. • When building an algorithm, show it clearly and be transparent of the learning methodology (if any) that the algorithm is using.	Absent
Code of Ethics for Data-Based Value Creation [10]	• If your product depends on data from other companies, make sure that this data has been collected in an ethical manner. • Do not use data from untrusted sources.	• Only acquire data from partners who are transparent about both their data collection practices, and any restrictions associated with the use of the data. • Make it clear to your customers how their data is protected and how access is logged. If it is necessary to store data for a very long time (eg, for legal reasons), explain this to your customers at the time of data collection.	Absent

XSL•FO
RenderX

[a]AI: artificial intelligence.

The intersection between digital health and trust is somewhat consistent regarding accountability for past mistakes and transparency about the present. Benchmarks for the future are less well-articulated across specific trust building activities [7-9], or their scope is limited.

The choice of who within an organization is responsible for the management and implementation of specific guidelines has implications for how financial resources get allocated and how the guidelines get interpreted. A joint effort across different units may be preferable as different trust-building principles require collaboration between different organizational units. Research suggests that a collaborative approach involving a range of actors across the organization enhances trust-building [22].

In response to these requirements, we need implementation evidence. Systematically collected evidence, case studies, and experience reports can inform how resources are allocated for different trust-building principles and identify successful trust-building mechanisms in a given context. Given the context specificity of trust, this evidence will not be a blueprint that can be copied and pasted across contexts. Instead, the evidence can help disseminate experience from comparable settings and is an essential step toward better informed resource allocation.

We Need Actionable Trust Performance Indicators to Collect Meaningful Evidence

To gather better evidence to inform resource allocation and make judgments about cost effectiveness, we need appropriate methods to collect data about trust-building from the implementation to the evaluation phases. The standard methodology to gather evidence about trust in the context of health care and medicine is survey instruments. Unfortunately, studies have repeatedly shown that the existing instruments have weaknesses. Goudge and Gilson [23] called for advancements in the quantification of trust; Ozawa and Sripad [24] showed in their review of 45 trust measures in health systems that half of the measures used qualitative methods in the design process, and only 33% were pilot tested; Aboueid et al [25] found in a meta-analysis of 26 publications on trust in health professionals and health systems, 10 studies did not mention the dimensions explored and of those that did, some did not define them. Taylor and colleagues [1] call for conceptual clarity and methodological creativity after reviewing 50 years of trust research. Considering the stagnation in methodological development in the field and the increasing inclusion of guidelines regarding trust in digital health, we see a need for refined methods to address questions of cost effectiveness and resource allocation.

An understanding of the resources invested, and outcomes gained is essential. Since we cannot pin a price tag on trust, we need to agree on sensible indicators that we understand as evidence informed. Table 3 gives an overview of currently applied indicators regarding trust in DHIs that have been developed by some of the essential stakeholders in the field, such as the World Bank or WHO. These indicators are the result of a systematic search on indicators to measure digital health system maturity conducted by Maaß et al [26]. The review identified that trust toward digital health applications, AI, the government, and health care providers plays an essential role among assessment tools for digital health system maturity. However, the indicators rarely reported their methodology and data source, thereby limiting their applicability. We require well designed indicators with transparent methodology and a clear definition on what is meant by “trust.”

XSL•FO
RenderX

Table 3. Currently used indicators on trust in digital health system interventions.

Indicator	Data source	Reference
Number of individuals identifying false or unreliable information, such as untrustworthy sources, fake news, or fraudulent emails	Quantitative national Computer assisted telephone interviewing survey of 1000 adults aged 18 years or older	Accenture Digital Index by Accenture [27]
You can recognize what information or content online may, or may not, be trustworthy (eg, fact checked information, "fake news" or assess the trustworthiness of a company based on customer reviews)	Qualitative telephone interviewing survey among 4172 adults aged 18 years or older living in the United Kingdom	UK Consumer Digital Index by Lloyds Bank [28]
Regulations and ethical frameworks in place to implement AI[a] in a way that builds trust and legitimacy	UNCTAD Data Protection and Privacy Legislation Worldwide data overview	Government AI[a] Readiness Index by Oxford Insights [29]
To what extent do you trust the information you receive from government websites/apps online?	4910 respondents from 98 countries	Inclusive Internet Index by Economist Impact and Meta [30]
To what extent do you trust the information you receive from nongovernment website/apps that are based in your country online?	4910 respondents from 98 countries	Inclusive Internet Index by Economist Impact and Meta [30]
To what extent do you trust that online privacy is guaranteed?	4910 respondents from 98 countries	Inclusive Internet Index by Economist Impact and Meta [30]
Are opportunities available to link data sources safely at the subject level and perform comprehensive analyses – for example, through a closed controlled working environment operated by the statistical office, or through anonymization and linkage by a trusted third party	No data source published	Support tool to strengthen health information systems by the World Health Organisation Regional Office for Europe [31]
Share of population that trusts medical and health advice from the government	Wellcome Trust Global Monitor [32]	Global Health Security Index by the Nuclear Threat Initiative [33]
Share of population that trusts medical and health advice from health professional	Wellcome Trust Global Monitor [32]	Global Health Security Index by the Nuclear Threat Initiative
The number of distinct, publicly-trusted TLS/SSL certificates found in the Netcraft Secure Server Survey	Netcraft and World Bank population estimates	Worldwide Governance Indicators by the World Bank [34]
The share of a population that trusts Artificial Intelligence	Quantitative survey among 62 countries	The Global AI Index by Tortoise [35]
The share of adults in each market saying they trust public bodies using generative AI[a] technology at least moderately	No data source published	The Network Readiness Index by the Portulans Institute [36]
The trust users hold in institutions, businesses, and policy makers to protect online privacy, harness technology to create jobs, and advocate for consumers	Survey among 42 economies	The Digital Intelligence Index by the Fletcher School at Tufts University [37]
The extent to which institutions create an environment which enables trust	Survey among 42 economies	The Digital Intelligence Index by the Fletcher School at Tufts University [37]
The level of generalized social trust and trust in science and technology in any given economy	Survey among 42 economies	The Digital Intelligence Index by the Fletcher School at Tufts University [37]

[a]AI: artificial intelligence.

Furthermore, most of the guidelines we identified and discussed in Tables 1 and 2 are not published or sponsored by governments or international institutions, raising the question who's interest it actually is to understand the economics of trust-building in digital health? Without appropriate performance indicators that show whether trust-building principles are successful, we cannot evaluate trust building mechanisms or make decisions about prioritization. Indicators are considered instrumental in driving health system improvements and for running comparative evaluations across the health system [38,39]. DHI investors and developers have good economic incentives to adopt trust performance indicators as it allows them to make better informed decisions about resource allocation for trust building and allows them to better monitor whether trust building actions are successful. Indicator sets are based on robust conceptual frameworks "that set out the rationale and design principles for an indicator set" [40] with the intention to provide standardized benchmarks for the systematic evaluation and monitoring of quality, performance, and outcomes. Addressing company executives, Deloitte argues that trust itself should be a key performance indicator for companies, as drops in consumer trust can considerably erode the market value of the firm [22]. Van der Schee et al [41] suggest public trust in the health system is an important performance indicator for assessing national health systems. Several other examples of indicators in specific areas of the health system exist, such as in digital health care, which include indicators relating to trust at least at the margins [27,28,30,31,33,35-37,42-46]. Discussion is currently centered

XSL•FO
RenderX

around identifying useful performance indicators for trust in DHIs and determining the optimal methods for evidence collection. The crucial task is to navigate conceptual complexity and to accommodate context specificity, ensuring the identification of indicators that are suitable for a diverse range of DHIs while aligning with trust-building guidelines.

Survey measures of trust levels can indicate changing trust levels over time [24]. Such measures are useful for assessing the level of trust before and after implementing trust-building principles. Similarly, qualitative user interviews and broader citizen engagement can provide rich information about perceptions and experiences regarding the trustworthiness of a DHI and the effectiveness of its associated trust-building measures [47]. The use of routinely collected data generated from digital health technologies needs to be harnessed and leveraged for better decision-making around trust-building and should include measures of the user experience at different levels of the system. The evidence should inform both the design phase of trustworthy DHIs and for continual assessment of their safety, effectiveness, and quality assurance [48]. This potpourri of methods allows us to think creatively about meaningful performance indicators for various contexts, actors, and DHIs. A comparable example of a newly developed indicator set is the Digital Public Health Maturity Index [49]. This index provides initial indications of possible architecture for an actionable and stakeholder-supported set of performance indicators in an area that faces the same kinds of problems as trust. Developing a trust index and performance indicators with appropriate methods to collect evidence must be integral to future implementation research on trust-building in digital health.

We Need to Sharpen the Trust Discourse

The increased use of the word “trust” by politicians, industry, and researchers alike raises the question of whether the use of the word trust is always justified. The inflationary use of trust in the public sphere, as seen in commercials, labelling, news articles, and media campaigns, possibly undermines trust’s value in settings where trust discussions are essential [50]. If we are not cautious, we may see the trust debate in the digital health sphere drift into symbolic actions and “trust washing,” similar to experiences in “green washing” and “ethics washing” [51]. Trust degenerates into a hollow marketing term when health system actors invest resources merely to chase trends without meaningfully engaging with internal trust-building procedures and actions leading to overall trustworthiness.

Trust is vital for adopting DHIs, but it is not a panacea. Health system actors can act to the best of their knowledge, and policy makers can design evidence-informed policy, including incorporating trustworthiness guidelines in their policy designs, but that does not mean the public or users will trust us [52]. Trust cannot be enforced or demanded. The public and the users decide how to assess trustworthiness and where to place their trust [53]. That insight implies that investments in trust-building in the context of DHIs carry the risk of failure.

Even if trust building is successful, it may not result in material benefit. For example, if several alternative health system actors adhere to trustworthiness guidelines and are all judged equally trustworthy, users will likely make choices not based on trust but on convenience, price, accessibility, or other adoption enablers. Trust is not the only reason we engage with DHIs. We might depend on a DHI and, consequently, engage with it because we have no other choice. We might assess the risk of misplaced trust and its associated consequences to be low and, therefore, not really care about trust. We might carryover the trust we have for more comprehensive health system to new DHIs, allowing us to spontaneously engage with the new and unfamiliar, similar to the concept of spontaneous sociability [12]. Such considerations are essential and require those aiming to build trust to assess the context carefully and to understand user motivations.

Realistic expectations about the capabilities of trust-building are critical in deciding which trust-building measures to adopt and how best to efficiently deploy the necessary resources.

Conclusions

Meaningful guidance for trust-building in digital health must go beyond proposing principles to build trust. We argue that the costs and benefits of trust-building measures are currently missing to a large extent. Health system actors and the research community need trust performance indicators to appropriately collect evidence about resources invested and trust gained in return. We need to be mindful of the present inflated use of the term “trust” in the public sphere when designing new trust building efforts, and to acknowledge that trust cannot be guaranteed. To successfully navigate the complexity of trust-building, we need more engagement in the economics of trust-building in digital health. Economics of trust-building needs to be a deeply integrated and structured part of DHI policy making and health system governance at large to develop sustainable trust building processes while bearing in mind resource scarcity. Furthermore, stakeholders across the health system from regulators, to health organizations, to end users need to have guided access to relevant and comprehensive information about these guidelines in order for meaningful progress to be made.

Acknowledgments

We thank Dr Paola Daniore, École Polytechnique Fédérale de Lausanne, and Federica Zavattaro, University of Zurich, for their valuable feedback on the article manuscript. We thank the reviewers for their insightful comments. FG is funded by the Digital Society Initiative, University of Zurich and Swiss Academy of Engineering Sciences. FG reports funding from Novartis International AG and Stiftung Sanitas Krankenversicherung, and WHO outside of the submitted work. LM is funded by the Leibniz Science Campus Bremen Digital Public Health, which is jointly funded by the Leibniz Association (W72/2022), the Federal State of

XSL•FO
RenderX

Bremen and the Leibniz Institute for Prevention Research and Epidemiology—BIPS. The funders had no role in writing the article.

Conflicts of Interest

None declared.

References

1. Taylor LA, Nong P, Platt J. Fifty years of trust research in health care: a synthetic review. Milbank Q 2023;101(1):126-178 [FREE Full text] [doi: 10.1111/1468-0009.12598] [Medline: 36689251]
2. McKee M, Schalkwyk M, Greenley R, Permanand G. Trust - The foundation of health systems. European Observatory on Health Systems and Policies.: World Health Organization; 2023. URL: https://iris.who.int/bitstream/handle/10665/379318/9789289059992-eng.pdf?sequence=1&isAllowed=y [accessed 2024-12-09]
3. Digital Health in the WHO European Region: the ongoing journey to commitment and transformation.: World Health Organization; 2023. URL: https://www.who.int/europe/publications/m/item/digital-health-in-the-who-european-region-the-ongoing-journey-to-commitment-and-transformation [accessed 2024-12-09]
4. Gille F. What is public trust in the health system? In: Insights Into Health Data Use. Bristol: Policy Press; Oct 20, 2023.
5. Ethics guidelines for trustworthy AI.: European Commission; 2019. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [accessed 2024-12-09]
6. Earning digital trust: decision-making for trustworthy technologies. Geneva: World Economic Forum; 2022. URL: https://www3.weforum.org/docs/WEF_Earning_Digital_Trust_2022.pdf [accessed 2024-12-09]
7. Global strategy on digital health 2020-2025.: World Health Organization; 2020. URL: https://apps.who.int/iris/handle/10665/344249 [accessed 2024-12-09]
8. Guidance on the rights-based and ethical use of digital technologies in HIV and health programmes. United Nations Development Programme. 2021. URL: https://www.undp.org/publications/guidance-rights-based-and-ethical-use-digital-technologies-hiv-and-health-programmes [accessed 2024-12-09]
9. A guide to good practice for digital and data-driven health technologies.: Department of Health & Social Care; 2021. URL: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology#data-transparency [accessed 2024-12-09]
10. Data ethics: code of ethics for data-based value creation. 2020. URL: https://data-innovation.org/data-ethics/ [accessed 2024-12-09]
11. Ho BTE. Why Trust Matters: An Economist's Guide to the Ties that Bind Us. New York: Columbia University Press; 2023.
12. Fukuyama F. Trust: The Social Virtues and the Creation of Prosperity. New York: Free Press; 1995.
13. Nunn N. The historical roots of economic development. Science 2020;367(6485):eaaz9986. [doi: 10.1126/science.aaz9986] [Medline: 32217703]
14. Algan Y, Cahuc P. Trust and Growth. Annu. Rev. Econ 2013;5(1):521-549. [doi: 10.1146/annurev-economics-081412-102108]
15. Dyer JH, Chu W. The role of trustworthiness in reducing transaction costs and improving performance: empirical evidence from the United States, Japan, and Korea. Organization Science 2003;14(1):57-68. [doi: 10.1287/orsc.14.1.57.12806]
16. Arrow K. Gifts and Exchanges. Philosophy & Public Affairs Wiley 1972;1(4):343-362. [doi: 10.1515/9781400853564.139]
17. Wilkinson T, Wang M, Friedman J, Prestidge M. A framework for the economic evaluation of digital health interventions. In: Policy Research Working Papers. Washington, DC: World Bank; 2023:0407.
18. Chaudhry SJ, Loewenstein G. Thanking, apologizing, bragging, and blaming: responsibility exchange theory and the currency of communication. Psychol Rev 2019;126(3):313-344. [doi: 10.1037/rev0000139] [Medline: 30762384]
19. Halperin B, Ho B, List J, Muir I. Toward an understanding of the economics of apologiesvidence from a large-scale natural field experiment. The Economic Journal 2022;132(641):273-298. [doi: 10.1093/ej/ueab062]
20. Fan S, Ge Q, Ho B, Ma L. Sorry doesn't cut it, or does it? Insights from stock market responses to corporate apologies. Journal of Economic Behavior & Organization 2023;205:68-86. [doi: 10.1016/j.jebo.2022.11.004]
21. Nikolova N, Möllering G, Reihlen M. Trusting as a 'Leap of Faith': trust-building practices in client–consultant relationships. Scandinavian Journal of Management 2015;31(2):232-245. [doi: 10.1016/j.scaman.2014.09.007]
22. Why trust should be one of your key performance indicators.: Deloitte; 2022. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-cfo-insights-why-trust-should-be-one-of-your-key-performance-indicators.pdf [accessed 2024-12-09]
23. Goudge J, Gilson L. How can trust be investigated? Drawing lessons from past experience. Soc Sci Med 2005;61(7):1439-1451. [doi: 10.1016/j.socscimed.2004.11.071] [Medline: 16005779]
24. Ozawa S, Sripad P. How do you measure trust in the health system? A systematic review of the literature. Soc Sci Med 2013;91:10-14. [doi: 10.1016/j.socscimed.2013.05.005] [Medline: 23849233]

XSL•FO
RenderX

25. Aboueid SE, Herati H, Nascimento MHG, Ward PR, Brown PR, Calnan M, et al. How do you measure trust in social institutions and health professionals? A systematic review of the literature (2012–2021). Sociology Compass 2023;17(9):e13101. [doi: 10.1111/soc4.13101]
26. Maaß L, Badino M, Iyamu I, Holl F. Assessing the digital advancement of public health systems using indicators published in gray literature: narrative review. JMIR Public Health Surveill 2024;10:e63031 [FREE Full text] [doi: 10.2196/63031] [Medline: 39566910]
27. Towards a digital society.: Accenture; 2022. URL: https://www.accenture.com/content/dam/accenture/final/a-com-migration/manual/r3/pdf/pdf-174/Accenture-digital-index.pdf#zoom=40 [accessed 2024-12-09]
28. 2023 Consumer digital index.: Lloyds Bank; 2023. URL: https://www.lloydsbank.com/assets/media/pdfs/banking_with_us/whats-happening/231122-lloyds-consumer-digital-index-2023-report.pdf [accessed 2024-12-09]
29. Hankins E, Fuentes NP, Martinescu L, Grau G, Rahim S. Government AI Readiness Index 2023. Malvern: Oxford Insights; 2023:2023.
30. The inclusive internet index.: Economist Impact; 2022. URL: https://impact.economist.com/projects/inclusive-internet-index/ [accessed 2024-12-09]
31. Support tool of the WHO Support tool to strengthen health information systems: guidance for health information system assessment and strategy development.: World Health Organization; 2021. URL: https://iris.who.int/handle/10665/342048 [accessed 2024-12-09]
32. Wellcome global monitor.: Wellcome Trust, Gallup; 2018. URL: https://wellcome.org/sites/default/files/wellcome-global-monitor-2018.pdf [accessed 2024-12-09]
33. Welcome to the Global Health Security Index.: Global Health Security Index; 2024. URL: https://www.ghsindex.org/ [accessed 2024-01-12]
34. Worldwide governance indicators.: World Bank Group; 2024. URL: https://www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access [accessed 2024-12-09]
35. The global AI index methodology report.: Tortoise Media; 2023. URL: https://www.tortoisemedia.com/wp-content/uploads/sites/3/2023/07/AI-Methodology-2306-4.pdf [accessed 2024-12-09]
36. Network readiness index 2024 benchmarking the future of the network economy.: Portulans Institute; 2024. URL: https://networkreadinessindex.org/ [accessed 2024-12-09]
37. Chakravorti B, Chaturvedi RS, Filipovic C, Brewer G. Digital in the time of COVID: trust in the digital economy and its evolution across 90 economies as the planet paused for a pandemic.: The Fletcher School at Tufts University; 2020. URL: https://digitalplanet.tufts.edu/wp-content/uploads/2022/09/digital-intelligence-index.pdf [accessed 2024-12-09]
38. Leeb K. Does health system performance reporting stimulate change? Healthc Manage Forum 2018;31(6):235-238. [doi: 10.1177/0840470418782515] [Medline: 30223671]
39. Sharma A, Prinja S, Aggarwal AK. Measurement of health system performance at district level: a study protocol. J Public Health Res 2017;6(3):917 [FREE Full text] [doi: 10.4081/jphr.2017.917] [Medline: 29441330]
40. Braithwaite J, Hibbert P, Blakely B, Plumb J, Hannaford N, Long JC, et al. Health system frameworks and performance indicators in eight countries: a comparative international analysis. SAGE Open Med 2017;5:2050312116686516 [FREE Full text] [doi: 10.1177/2050312116686516] [Medline: 28228948]
41. Van der Schee E, Groenewegen PP, Friele RD. Public trust in health care: a performance indicator? J Health Organ Manag 2006;20(5):468-476 [FREE Full text] [doi: 10.1108/14777260610701821] [Medline: 17087406]
42. KPMG global tech report 2023.: KPMG; 2023. URL: https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2023/09/kpmg-global-tech-report.pdf [accessed 2024-12-09]
43. International health policy survey.: The Commonwealth Fund; 2024. URL: https://www.commonwealthfund.org/trending/international-health-policy-survey [accessed 2024-12-09]
44. The Digital Economy and Society Index (DESI).: European Commission; 2024. URL: https://digital-strategy.ec.europa.eu/en/policies/desi [accessed 2024-12-09]
45. Chukwu E, Mechael P, Edelman JK, Layer E. The state of digital health 2023 global digital health monitor.: HealthEnabled; 2023. URL: https://static1.squarespace.com/static/5ace2d0c5cfd792078a05e5f/t/656f97969301e337ada15270/1701812128734/State+of+Digital+Health_2023.pdf [accessed 2024-12-09]
46. Data protection and privacy legislation worldwide.: United Nations Conference on Trade and Development; 2024. URL: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide [accessed 2024-12-09]
47. Yang K, Holzer M. The performance-trust link: implications for performance measurement. Public Administration Review 2006;66(1):114-126. [doi: 10.1111/j.1540-6210.2006.00560.x]
48. Srivastava D, Henschke C, Virtanen L, Lotman EM, Friebel R, Ardito V, et al. Promoting the systematic use of real-world data and real-world evidence for digital health technologies across Europe: a consensus framework. Health Econ Policy Law 2023;18(4):395-410. [doi: 10.1017/S1744133123000208] [Medline: 37705236]
49. Maaß L, Zeeb H, Rothgang H. How to measure digital public health system maturity on a national level: an international multidisciplinary delphi study. Review 2023. [doi: 10.21203/rs.3.rs-3021921/v1]
50. Gille F, Brall C. Public trust: caught between hype and need. Int J Public Health 2020;65(3):233-234. [doi: 10.1007/s00038-020-01343-0] [Medline: 32114636]

XSL•FO
RenderX

51. Walker K, Wan F. The harm of symbolic actions and green-washing: corporate actions and communications on environmental performance and their financial implications. J Bus Ethics 2012;109(2):227-242. [doi: 10.1007/s10551-011-1122-4]
52. Hult D. Creating trust by means of legislation – a conceptual analysis and critical discussion. The Theory and Practice of Legislation 2018;6(1):1-23. [doi: 10.1080/20508840.2018.1434934]
53. Hartmann M. Die Praxis des Vertrauens. Berlin, Germany: Suhrkamp Verlag; 1994.

Abbreviations

AI: artificial intelligence
DHI: digital health innovation
WHO: World Health Organization

Edited by A Mavragani; submitted 02.04.24; peer-reviewed by C Jacob, T van Mierlo; comments to author 27.05.24; revised version received 12.07.24; accepted 26.11.24; published 15.01.25.

Please cite as:
Gille F, Maaß L, Ho B, Srivastava D
From Theory to Practice: Viewpoint on Economic Indicators for Trust in Digital Health
J Med Internet Res 2025;27:e59111
URL: https://www.jmir.org/2025/1/e59111
doi:10.2196/59111
PMID:

©Felix Gille, Laura Maaß, Benjamin Ho, Divya Srivastava. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Era of Generalist Conversational Artificial Intelligence to Support Public Health Communications

Emre Sezgin[1,2], PhD; Ahmet Baki Kocaballi[3], PhD

[1]The Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, OH, United States

[2]College of Medicine, The Ohio State University, Columbus, OH, United States

[3]School of Computer Science, University of Technology Sydney, Sydney, Australia

Corresponding Author:
Emre Sezgin, PhD
The Abigail Wexner Research Institute at Nationwide Children's Hospital
700 Children's Dr
Columbus, OH, 43205
United States
Phone: 1 6147223179
Email: emre.sezgin@nationwidechildrens.org

Abstract

The integration of artificial intelligence (AI) into health communication systems has introduced a transformative approach to public health management, particularly during public health emergencies, capable of reaching billions through familiar digital channels. This paper explores the utility and implications of generalist conversational artificial intelligence (CAI) advanced AI systems trained on extensive datasets to handle a wide range of conversational tasks across various domains with human-like responsiveness. The specific focus is on the application of generalist CAI within messaging services, emphasizing its potential to enhance public health communication. We highlight the evolution and current applications of AI-driven messaging services, including their ability to provide personalized, scalable, and accessible health interventions. Specifically, we discuss the integration of large language models and generative AI in mainstream messaging platforms, which potentially outperform traditional information retrieval systems in public health contexts. We report a critical examination of the advantages of generalist CAI in delivering health information, with a case of its operationalization during the COVID-19 pandemic and propose the strategic deployment of these technologies in collaboration with public health agencies. In addition, we address significant challenges and ethical considerations, such as AI biases, misinformation, privacy concerns, and the required regulatory oversight. We envision a future with leverages generalist CAI in messaging apps, proposing a multiagent approach to enhance the reliability and specificity of health communications. We hope this commentary initiates the necessary conversations and research toward building evaluation approaches, adaptive strategies, and robust legal and technical frameworks to fully realize the benefits of AI-enhanced communications in public health, aiming to ensure equitable and effective health outcomes across diverse populations.

(J Med Internet Res 2025;27:e69007) doi:10.2196/69007

KEYWORDS

messaging apps; public health communication; language models; artificial intelligence; AI; generative AI; conversational AI

Introduction

Health communication and information dissemination are essential for global risk mitigation during public health emergencies. The pandemic highlighted the necessity for effective worldwide communication networks in support of the public health agencies (Center for World Health Organization [WHO] and Global Public Health Intelligence Network), and furthermore, the need for more integrated and global systems for timely warnings and responses to health crises [1]. For this purpose, messaging services have been vital tools for disseminating information, monitoring disease spread, and promoting informed health decisions [2]. Historically, SMS- and app-based text messaging platforms have long been recognized for their potential to deliver scalable health interventions for diverse populations, exemplified by RCTs on substance abuse intervention and promoting the COVID-19 pandemic vaccination [3,4]. These examples illustrate the established value of messaging services in engaging populations and promoting health behaviors.

The emergence of artificial intelligence (AI) has brought innovative and personalized methods in health communications,

XSL•FO
RenderX

with chatbots and AI-driven messaging services presenting personalized, timely, and internet-based health interventions [5]. More specifically, conversational artificial Intelligence (CAI) interventions have exhibited significant positive effects on behavior change, such as smoking cessation, healthy eating, sleep quality, and physical activity [6,7]. CAI models have emerged to complete tasks through natural language using rule-based, hybrid, or unsupervised learning models [8,9]. However, as CAI has been widely used for personalized health intervention, a public health perspective has yet to be studied.

During the pandemic, the emerging value of CAI has been observed as being a scalable, easy-to-use, and accessible dissemination tool [10-12]. In addition, the long-term value of CAI through chatbot tools and voice assistants in public health management has been emphasized [13,14]. A widely observed implementation was a WHO-deployed WhatsApp (Meta) chatbot, as a preventative mechanism, that disseminated critical health information to millions globally, providing multilingual support and real-time updates to address misinformation [15]. However, none of the intelligent systems covered earlier has envisioned the impact of generative AI at scale where it can lead to highly accessible, decentralized, and scalable implementations. Recent applications of widely adapted generative AI models, large language models (LLMs), presented evidence of the effectiveness of generated responses to answer public health questions [16,17], contributing to the trend of using AI to augment the impact of public health interventions.

Envisioning the future, its impact could exponentially grow by wide-scale adoption and implementation across the population by its integration into our daily communication tools (ie, messaging apps). We hypothesize that such LLM-based CAI services (namely, “Generalist CAI” from hereon) through messaging apps, the most commonly used tools on a daily basis, may overtake other means of internet-based information seeking and public information dissemination and sharing mechanisms, toward improving public health communications.

Unlike previous implementations of CAI before the wide adoption of LLMs, Generalist CAI models rely on generative AI (ie, LLMs) or foundation models that are trained on extensive and diverse datasets to perform a myriad of natural language tasks with human-like interaction and a broad understanding across various domains, including health care. These models transcend the limitations of their predecessors, which were constrained to specific tasks, by exhibiting proficiency in varied domains, from casual dialogue to expert-level engagements. These models’ broad applicability is underpinned by their capacity to comprehend and respond in a human-like manner across different conversational contexts. As such, generalist CAI models are not only poised to enhance user interactions but are also likely to supplant multiple task-specific agents (multiagents), heralding a shift toward more unified and contextually aware conversational AI systems in diverse applications. Table 1 outlines the key features of generalist AI.

Table 1. Key features and descriptions of generalist conversational artificial intelligence (CAI) applications.

Key features	Descriptions
Natural language understanding and generation	They can understand a broad range of user prompts that can be expressed in various different ways, allowing a higher level of flexibility and ease-of-use.
Generalist	They can converse about a large range of topics and transition between them smoothly.
Dialog management	They can excel in multiturn conversations more effectively than their rule-based alternatives with the ability to respond to follow-up questions.
Emotional intelligence	They can detect subtle cues in language that indicate a user’s emotional state, allowing the chatbot to provide empathetic responses.
Personalization	Generalist CAI can extract user profile information and use it to generate tailored responses effectively.
Multilingual support	They can support multiple languages, making them accessible to a wider audience.
Scalability	They can handle an unlimited number of queries simultaneously, providing instant information on a wide range of public health issues.

The Advantage of Generalist CAI in Messaging Services

AI has already transformed the ways in which individuals seek health information. Today, we observe that generalist CAI or LLM applications (eg, chatGPT) simplify access to health advice [18], by breaking down barriers associated with traditional web searches.

The next step appears to be messaging tools with generalist CAI assistance, like WhatsApp, Signal (Signal Technology Foundation), iMessage (Apple Inc.), and Telegram (Telegram FZ-LLC Telegram Messenger Inc), which would offer a “familiar” platform to billions of people, with the potential to improve access to health information and services. Messaging apps’ simplicity in terms of user interface and the use of natural language inherently reduce digital literacy and accessibility barriers. Therefore, messaging services with generalist CAI can surpass previous implementations of task-specific CAI services regarding capabilities, user access and adoption, lower barriers to use and provide personalization through self-learning and conversation adaptation (Table 1). For example, instead of just providing preprogrammed answers, a generalist CAI can engage in a dynamic dialogue to understand a user’s specific concerns, clarify ambiguities, and provide tailored advice based on the conversation’s context. Furthermore, mobile device-based dialogue services (eg, automatic speech recognition and text-to-speech) further improve the usability of the service

XSL•FO
RenderX

through natural language conversations. This means, in a public health emergency, the widespread use of apps like iMessage and WhatsApp (billions of active users daily) can facilitate immediate access to personalized health advice, circumventing barriers to accessing urgent public health information. The integration with generalist CAI, as seen in Meta AI's LLM-powered generalist CAI or personal assistant, positions these services in a favorable spot as a convenient tool for information seeking and health communication, potentially offering equitable access to personalized health information across diverse populations. To further extend these opportunities, small language models (eg, Phi-3) [19] can be leveraged to build capable generalist CAI for individuals in remote areas or those with limited access to internet bandwidth, who could still benefit from text-based interactions with the CAI.

Envisioning Generalist CAI Use in Public Health Emergencies

As the AI tools and applications are rapidly growing and immersing in our daily lives, it is necessary to plan and strategize for the use of generalist CAI services in public health emergencies. Given the fact that the current LLMs and CAI are trained on publicly available data on a massive scale with less attention to information quality or specific domains like health care, we need to approach cautiously in promoting safe and reliable information pipelines over these messaging platforms for public health information sharing, communication, and dissemination. To improve CAI accuracy and reliability in public health communications, a multiagent approach could be used [20,21]. This includes specialized AI agents collaborating, some provide general health information, while others ensure compliance with health guidelines. We envision the integration of generalist CAI assistants into public health messaging services through such a multiagent approach to streamline future interventions. Figure 1 illustrates a 2-agent arrangement where a CAI agent in a messaging app uses the public health information provided by the Centers for Disease Control and Prevention (CDC) to compose its response.

Figure 1. Current practice and the proposed alternative routes for public health communications during an outbreak CAI: conversational artificial intelligence; CDC: Centers for Disease Control and Prevention; WHO: World Health Organization.

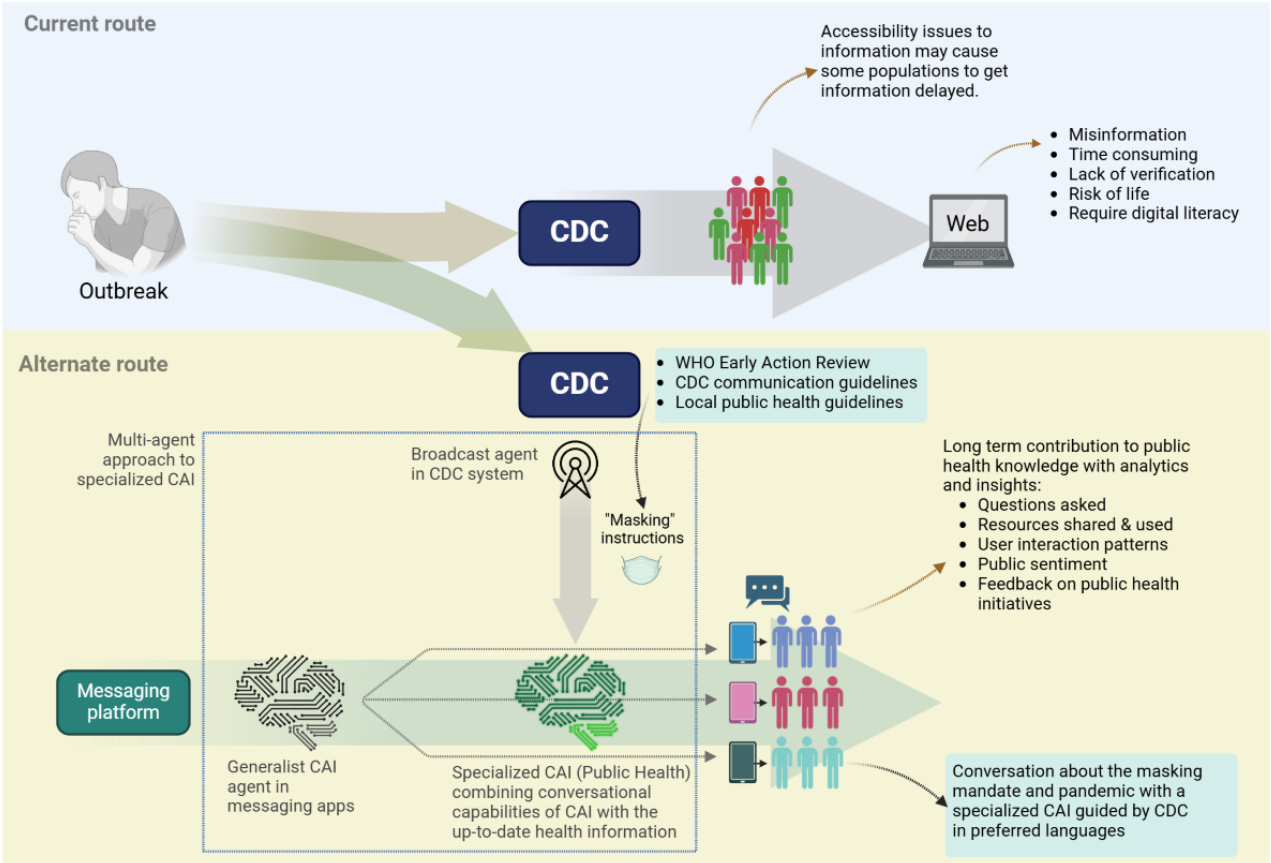

Public Health Emergency Case: the COVID-19 Pandemic

During the COVID-19 pandemic, we observed that the dissemination of comprehensive and accurate information is crucial for educating the public and combating the pandemic [12]. This period highlighted the potential of basic chatbot applications in enhancing public health communications, especially in diverse and low-resource settings. As aforementioned, WHO released a chatbot over WhatsApp providing up-to-date COVID-19 information in multiple local languages extending across several low- and middle-income countries [15]. In Nigeria, an SMS-based chatbot, offered by the Nigeria Centre for Disease Control, with support from

XSL•FO
RenderX

UNICEF (United Nations International Children's Emergency Fund), provided localized guidance and timely information about pandemic [22]. Similarly, India released an app with chatbot features to engage rural and urban Indian residents for the COVID-19 pandemic safety practices and checking symptoms, updates, and access helplines [23]. These implementations demonstrate the scalability and potential of digital conversational intervention in public health emergencies. With generalist CAI, personalized public health intervention can be achieved, bridging communication gaps and supporting public health agencies during emergencies.

A strategic plan for effective communication during a health crisis is required, leveraging a multiagent approach. In a partnership between the public health agencies (eg, WHO and CDC) and technology providers, we may create an information dissemination pipeline during a public health crisis (Figure 1). For instance, by adhering to CDC's guidelines for communication and dissemination and using only vetted information sources [24], generalist CAI agents can be designed to deliver messages accurately, empathetic, and action oriented. CDC-defined key elements for developing outbreak-related messages include expressing empathy, outlining clear actions, delineating what is known and unknown, explaining public health actions and their rationale, committing to ongoing communication, and guiding the public on where to find reliable information [24]. For example, regarding the need to "outline clear actions," a specialized agent could be programmed to ensure that every informational response includes actionable steps the user can take. This framework could be used to prompt the agents and ensure that the generated messages are fine-tuned to deliver intended messages toward not only informing and educating but also to build trust and encourage compliance with health advisories. Furthermore, aligning with WHO's early action review guidelines that are designed to optimize early detection of public health emergencies, such agents could be deployed to intervene timely and tailored to the target audience in compliance with the local or national guidelines across the globe [25]. Ideally, a trained generalist CAI acting as a public health support agent can be simultaneously activated across the populations, including low-resource countries, rural and urban areas, served in any preferred language through text or speech. In the long term, such agents can contribute to the knowledge of public health agencies and the data they provide can help learn more about the effectiveness of communications with real-time feedback from the public. Future research should focus on developing frameworks that facilitate the integration of multiagent CAI systems into existing public health infrastructures. Some of the viable approaches could be through interoperability with electronic health records, collaboration with public health databases and registries, and using standardized communication protocols like Fast Health care Interoperability Resources [26], embedding CAI agents into communication platforms preferred by health professionals, and establishing governance mechanisms to ensure ethical and efficient operation [27,28].

Multimedia Appendix 1 for our brief experiments and observations toward the opportunities and limitations with current generalist CAI applications in response to some of the emerging public health problems including pandemic response, unmet social needs, mental health support, and vaccinations.

Risks and Ethical Considerations

The deployment of generalist CAI assistants in messaging apps presents multifaceted challenges, including risks of bias, misinformation, hallucinations, and ethical pitfalls, which have been observed earlier with Snapchat's personal assistant [29]. The models behind these CAI assistants pose a critical challenge due to inherent biases arising from skewed or incomplete training data, flawed algorithmic design, or the reinforcement of societal prejudices [30]. These biases risk distorting health communications and exacerbating health disparities [31,32]. For example, an earlier study [33] revealed that a widely used health care algorithm exhibited significant racial bias by predicting costs instead of health needs, leading to unequal care allocation for Black patients. Such bias could pose a risk for generalist CAI providing less comprehensive (or even inaccurate) information to individuals from underrepresented racial or ethnic backgrounds. In addition, a culturally sensitive and inclusive design is also important to mitigate the risk of inherent biases that may exist in the training data of generalist CAI. These issues underscore the importance of accountability, fairness, equity, and regulatory oversight [34]. Moreover, ethical concerns extend beyond privacy to include user autonomy and the transparent use of data, necessitating clear guidelines and user consent. To support user autonomy, implementing verification mechanisms and ensuring the source data originating from authoritative health organizations can help maintain the trustworthiness of the AI-generated information. One step further, a multiagent approach can help control the CAI behavior regarding user location and profile, ensuring that advice and data handling procedures comply with the local rules and regulations. Even though messaging services are one of the most accessible and used communication technologies, the FCC Affordable Connectivity Program or similar global programs can be used to address the digital divide and affordability issues [35]. While some messaging apps offer end-to-end encryption, overall, the lack of governance or medico-legal compliance (eg, HIPAA [Health Insurance Portability and Accountability Act], HITECH [Health Information Technology for Economic and Clinical Health] or GDPR [General data protection regulation] rule on processing sensitive data requirements for public health information), unless integrated with telehealth services, raises significant privacy concerns, similar to those of standard text messaging services. Technical standards and ethical frameworks should be developed to ensure AI systems are transparent and fair [36]. Further considerations are outlined in Table 2.

XSL•FO
RenderX

Table 2. Technical challenges and potential solutions in deploying generalist conversational artificial intelligence (CAI).

Challenge	Description	Potential solutions
Language diversity	• Developing CAI systems that can understand and generate responses in multiple languages and dialects is complex. Language nuances, slang, and regional expressions can hinder accurate communication [37].	• Multilingual training: Incorporate extensive multilingual datasets during training. • Local expertise collaboration: Partner with local linguists and community experts. • Adaptive learning: Implement adaptive learning algorithms based on user interactions.
Technological infrastructure	• Variability in technological infrastructure, such as outdated devices or limited computational resources, can impede effective CAI deployment [38].	• Lightweight models: Develop optimized, lightweight, and specialized AI models. • Edge computing: use edge computing to perform data processing closer to the user. • Offline capabilities: Incorporate offline functionalities where possible.
Internet connectivity	• In many low-resource settings, unreliable or limited internet access can restrict the usability and accessibility of CAI-powered messaging services [39].	• Data-efficient communication: Design CAI systems requiring minimal data usage. • SMS or voice integration: Integrate CAI functionalities with SMS-based platforms or text-to-voice calls. • Caching and preloading: Implement caching strategies to store essential information locally [40].
Integration with existing systems	• Integrating CAI with existing public health information systems and workflows can be technically demanding [41].	• API[a] development: Develop robust APIs for smooth integration. • Modular architecture: Adopt a modular system architecture. • Standardization: Promote standardized protocols and data formats for interoperability.
Data privacy and security	• Ensuring the privacy and security of user data is paramount, especially when dealing with sensitive health information [42].	• Encryption and access controls: Establish strict access controls, authentication mechanisms, and end-to-end encryption. • Compliance frameworks: Adhere to international and local data protection regulations (eg, GDPR[b], HIPAA[c]). • Small local models: Support the development of smaller models that can be run locally (eg, Mistral 7B and Microsoft Phi-3) [43].
Bias and cultural sensitivity	• AI algorithms can inadvertently conserve social prejudices and cultural insensitivities present in training data, leading to biased or inappropriate responses [44].	• Diverse training data: Curate datasets that represent diverse populations and cultural contexts [45]. • Bias audits: Regularly conduct bias assessments and audits of CAI models [46]. • Inclusive design practices: Incorporate principles of inclusive design to ensure cultural relevance and sensitivity [47].
Stakeholder engagement	• Engaging community stakeholders is essential to ensure that CAI systems meet the needs of diverse populations and promote equitable outcomes [48].	• Participatory co-design: Involve community members in the design and development process. • Feedback mechanisms: Establish channels for continuous feedback from users. • Collaborative partnerships: Partner with local organizations and leaders to guide CAI development and deployment.

[a]API: application programming interface.

[b]GDPR: General data protection regulation.

[c]HIPAA: Health Insurance Portability and Accountability Act.

The contentious nature of AI governance and accreditation of service providers for multiagent AI services might require a legal infrastructure as much as a technical one, to reduce perceived legal risks and liabilities with government agencies accrediting private sector tools. However, the recent initiative by the Biden Administration to form task forces aimed at shaping policies for AI in health care signals a promising direction for overcoming these hurdles in the United States [49], suggesting that improvements in the medicolegal landscape could pave the way for safer AI implementations in public health communication [50]. Internationally, regulatory bodies are also beginning to take similar steps, aiming for a cohesive global

XSL•FO
RenderX

response to AI challenges [51]. It is crucial to address how these technologies might disproportionately affect marginalized groups, ensuring inclusive and equitable AI development. The long-term societal impacts, such as the erosion of public trust through AI missteps, must also be considered in developing sustainable AI strategies. Engaging a broad spectrum of stakeholders in AI discussions can enhance the legitimacy and effectiveness of governance structures. This evolving scenario highlights the critical need for a balanced approach to harnessing AI's potential while addressing its ethical, legal, and social challenges.

Conclusions and Future Directions

AI-enhanced messaging apps hold significant promise for advancing public health by improving access to health information, supporting health behavior change, and addressing diverse community needs. Their scalability and adaptability have already demonstrated impact during public health crises, such as the COVID-19 pandemic, where CAI systems facilitated rapid and multilingual information dissemination. Generalist CAI, with its ability to handle diverse conversational tasks and adapt to user needs with its human-like interaction capability, represents a transformative opportunity to create equitable and accessible public health communication tools. However, realizing their full potential requires addressing critical challenges, including biases in training data, risks of misinformation, ethical concerns around equity and transparency, and the need for robust legal and technical frameworks. To unlock these opportunities, AI messaging services must be developed and continually updated to address evolving requirements on fairness, accountability, and culturally sensitive design, ensuring they uphold the principles of reliability, equity, and justice in public health.

Key research priorities should include evaluating the efficacy of multiagent systems, understanding user interaction and trust dynamics, addressing biases, assessing impacts on health equity, and exploring innovative applications beyond basic information delivery. Observational studies can help identify how individuals engage with CAI systems and guide improvements in public health applications. To ensure equitable access, public health agencies should advocate for culturally sensitive and linguistically diverse CAI systems. Collaborations with local organizations and the establishment of global repositories of vetted health information are critical to minimizing misinformation and aligning CAI outputs with public health guidelines. Eventually, public health agencies and researchers can create a roadmap for leveraging CAI effectively and equitably, paving the way for a transformative approach to public health communication. Continued research on user engagement and the optimization of AI models for public health is essential to fully leverage the capabilities of these ubiquitous tools.

Acknowledgments

We acknowledge Kelly Kelleher for his valuable feedback. We used the generative AI tool (ChatGPT by OpenAI) to refine phrases and improve the clarity of some statements.

Conflicts of Interest

ES is an Associate Editor of the *Journal of Medical Internet Research*.

Multimedia Appendix 1
Examples of generalist CAI (conversational artificial intelligence) responses to public health issues.
[PDF File (Adobe PDF File), 9592 KB - jmir_v27i1e69007_app1.pdf]

References

1. Meckawy R, Stuckler D, Mehta A, Al-Ahdal T, Doebbeling BN. Effectiveness of early warning systems in the detection of infectious diseases outbreaks: a systematic review. BMC Public Health 2022;22(1):2216 [FREE Full text] [doi: 10.1186/s12889-022-14625-4] [Medline: 36447171]
2. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
3. Patel MS, Fogel R, Winegar AL, Horseman C, Ottenbacher A, Habash S, et al. Effect of text message reminders and vaccine reservations on adherence to a health system covid-19 vaccination policy: a randomized clinical trial. JAMA Netw Open 2022;5(7):e2222116 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.22116] [Medline: 35857327]
4. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
5. Lauffenburger JC, Yom-Tov E, Keller PA, McDonnell ME, Crum KL, Bhatkhande G, et al. The impact of using reinforcement learning to personalize communication on medication adherence: findings from the REINFORCE trial. NPJ Digit Med 2024;7(1):39 [FREE Full text] [doi: 10.1038/s41746-024-01028-5] [Medline: 38374424]
6. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023;6(1):118 [FREE Full text] [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]

XSL•FO
RenderX

7. Bendotti H, Lawler S, Chan GCK, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: a systematic review and meta-analysis. Digit Health 2023;9:20552076231211634 [FREE Full text] [doi: 10.1177/20552076231211634] [Medline: 37928336]
8. Kusal S, Patil S, Choudrie J, Kotecha K, Mishra S, Abraham A. AI-based conversational agents: a scoping review from technologies to future directions. IEEE Access Institute of Electrical and Electronics Engineers (IEEE) 2022;10:92337-92356 [FREE Full text] [doi: 10.1109/access.2022.3201144]
9. Laymouna M, Ma Y, Lessard D, Schuster T, Engler K, Lebouché B. Roles, users, benefits, and limitations of chatbots in health care: rapid review. J Med Internet Res 2024;26:e56930 [FREE Full text] [doi: 10.2196/56930] [Medline: 39042446]
10. Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response. J Am Med Inform Assoc 2022;29(5):1000-1010 [FREE Full text] [doi: 10.1093/jamia/ocac014] [Medline: 35137107]
11. Fan Z, Yin W, Zhang H, Wang D, Fan C, Chen Z, et al. COVID-19 information dissemination using the wechat communication index: retrospective analysis study. J Med Internet Res 2021;23(7):e28563 [FREE Full text] [doi: 10.2196/28563] [Medline: 34129515]
12. Zhou Y, Zhang A, Liu X, Tan X, Miao R, Zhang Y, et al. Protecting public's well-being against the COVID-19 infodemic: the role of trust in information sources and rapid dissemination and transparency of information over time. Front Public Health 2023;11:1142230 [FREE Full text] [doi: 10.3389/fpubh.2023.1142230] [Medline: 37139363]
13. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
14. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020;3:65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
15. How whatsApp can help stay you connected during COVID-19 pandemic. URL: https://www.whatsapp.com/coronavirus [accessed 2024-09-28]
16. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating artificial intelligence responses to public health questions. JAMA Netw Open 2023;6(6):e2317517 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.17517] [Medline: 37285160]
17. Sezgin E. Redefining virtual assistants in health care: the future with large language models. J Med Internet Res 2024;26:e53225 [FREE Full text] [doi: 10.2196/53225] [Medline: 38241074]
18. Haupt CE, Marks M. AI-generated medical advice-GPT and beyond. JAMA 2023;329(16):1349-1350. [doi: 10.1001/jama.2023.5321] [Medline: 36972070]
19. Phi open models - small language models. URL: https://azure.microsoft.com/en-us/products/phi [accessed 2024-12-20]
20. Wu Q, Bansal G, Zhang J, Wu Y, Li B, Zhu E, et al. AutoGen: enabling Next-Gen LLM applications via multi-agent conversation. arXiv:2308.08155 2023 [FREE Full text]
21. Multi-agent Conversation Framework. URL: https://microsoft.github.io/autogen/docs/Use-Cases/agent_chat/ [accessed 2024-09-28]
22. NCDC and UNICEF launch chatbot to combat COVID-19 misinformation in Nigeria. URL: https://www.unicef.org/nigeria/press-releases/ncdc-and-unicef-launch-chatbot-combat-covid-19-misinformation-nigeria [accessed 2025-01-07]
23. Sehgal P, Jain V. Fighting pandemic: the mobile application way (a case of the Aarogya Setu App). Emerging Economies Cases Journal 2023;4(6):251660422211470 [FREE Full text] [doi: 10.1177/25166042221147097]
24. Tumpey AJ, Daigle D, Nowak G. Communicating during an outbreak or public health investigation. In: The CDC Field Epidemiology Manual. Oxford, England: Oxford University Press; 2019:243-260.
25. Guidance and tools for conducting an early action review (EAR): rapid performance improvement for outbreak detection and response. WHO. URL: https://www.who.int/publications/i/item/WHO-WPE-HSP-CER-2023.1 [accessed 2024-03-02]
26. Overview - FHIR v5.0.0. URL: https://www.hl7.org/fhir/overview.html [accessed 2024-10-01]
27. Liao F, Adelaine S, Afshar M, Patterson BW. Governance of clinical AI applications to facilitate safe and equitable deployment in a large health system: key elements and early successes. Front Digit Health 2022;4:931439 [FREE Full text] [doi: 10.3389/fdgth.2022.931439] [Medline: 36093386]
28. Boch S, Sezgin E, Lin Linwood S. Ethical artificial intelligence in paediatrics. Lancet Child Adolesc Health 2022;6(12):833-835. [doi: 10.1016/S2352-4642(22)00243-7] [Medline: 36084667]
29. McCallum S. Snapchat: Snap AI chatbot 'may risk children's privacy'. BBC News. URL: https://www.bbc.com/news/technology-67027282 [accessed 2024-03-08]
30. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). NPJ Digit Med 2024;7(1):183 [FREE Full text] [doi: 10.1038/s41746-024-01157-x] [Medline: 38977771]
31. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digit Med 2023;6(1):113 [FREE Full text] [doi: 10.1038/s41746-023-00858-z] [Medline: 37311802]
32. Timmons AC, Duong JB, Simo Fiallo N, Lee T, Vo HPQ, Ahle MW, et al. A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. Perspect Psychol Sci 2023;18(5):1062-1096 [FREE Full text] [doi: 10.1177/17456916221134490] [Medline: 36490369]
33. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]

XSL•FO
RenderX

34. Assistant Secretary for Public Affairs (ASPA). Guiding principles help health care community address potential bias resulting from algorithms. US Department of Health and Human Services. URL: https://www.hhs.gov/about/news/2023/12/15/guiding-principles-help-healthcare-community-address-potential-bias-resulting-from-algorithms.html [accessed 2024-03-02]
35. Affordable connectivity program. URL: https://www.fcc.gov/acp [accessed 2024-03-09]
36. Ashok M, Madan R, Joha A, Sivarajah U. Ethical framework for artificial intelligence and digital technologies. Int. J. Inf. Manag 2022;62(2):102433 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2021.102433]
37. Zhang X, Ogueji K, Ma X, Lin J. Towards best practices for training multilingual dense retrieval models. arXiv:2204.02363 2020 [FREE Full text]
38. Nagarajan R, Kondo M, Salas F, Sezgin E, Yao Y, Klotzman V, et al. Economics and equity of large language models: health care perspective. J Med Internet Res 2024;26:e64226 [FREE Full text] [doi: 10.2196/64226] [Medline: 39541580]
39. van Zyl C, Badenhorst M, Hanekom S, Heine M. Unravelling 'low-resource settings': a systematic scoping review with qualitative content analysis. BMJ Glob Health 2021;6(6):e005190 [FREE Full text] [doi: 10.1136/bmjgh-2021-005190] [Medline: 34083239]
40. Prompt caching with Claude. URL: https://www.anthropic.com/news/prompt-caching [accessed 2024-10-01]
41. Balch JA, Ruppert M, Loftus T, Guan Z, Ren Y, Upchurch G, et al. Machine learning-enabled clinical information systems using fast healthcare interoperability resources data standards: scoping review. JMIR Med Inform 2023;11:e48297 [FREE Full text] [doi: 10.2196/48297] [Medline: 37646309]
42. May R, Denecke K. Security, privacy, and healthcare-related conversational agents: a scoping review. Inform Health Soc Care 2022;47(2):194-210. [doi: 10.1080/17538157.2021.1983578] [Medline: 34617857]
43. Beatty S. Tiny but mighty: the Phi-3 small language models with big potential. Source. URL: https://news.microsoft.com/source/features/ai/the-phi-3-small-language-models-with-big-potential/ [accessed 2024-10-01]
44. Xue J, Wang Y, Wei C, Liu X, Woo J, Kuo C. Bias and fairness in Chatbots: an overview. arXiv:2309.08836 2023 [FREE Full text]
45. Wan Y, Wang W, He P, Gu J, Bai H, Lyu M. BiasAsker: measuring the bias in conversational AI system. USA: ACM; 2023 Presented at: Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering; 2023 Nov 30; CA, San Francisco, USA.
46. Landers RN, Behrend TS. Auditing the AI auditors: a framework for evaluating fairness and bias in high stakes AI predictive models. Am Psychol 2023;78(1):36-49. [doi: 10.1037/amp0000972] [Medline: 35157476]
47. Motta I, Quaresma M. Increasing transparency to design inclusive Conversational Agents (CAs): perspectives and open issues. USA: ACM; 2023 Presented at: Proceedings of the 5th International conference on conversational user interfaces; 2023 July 19-21; Eindhoven, Netherlands.
48. Nadarzynski T, Knights N, Husbands D, Graham CA, Llewellyn CD, Buchanan T, et al. Achieving health equity through conversational AI: a roadmap for design and implementation of inclusive chatbots in healthcare. PLOS Digit Health 2024;3(5):e0000492 [FREE Full text] [doi: 10.1371/journal.pdig.0000492] [Medline: 38696359]
49. Fact sheet: Biden-Harris administration announces key AI actions following President Biden's landmark Executive Order. The White House. URL: https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/ [accessed 2024-03-03]
50. Chin MH, Afsar-Manesh N, Bierman AS, Chang C, Colón-Rodríguez CJ, Dullabh P, et al. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. JAMA Netw Open 2023;6(12):e2345050 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.45050] [Medline: 38100101]
51. Cha S. Towards an international regulatory framework for AI safety: lessons from the IAEA's nuclear safety regulations. Humanit Soc Sci Commun 2024;11(1):506. [doi: 10.1057/s41599-024-03017-1]

Abbreviations

AI: artificial intelligence
CAI: conversational artificial intelligence
CDC: Centers for Disease Control and Prevention
GDPR: General data protection regulation.
HIPAA: Health Insurance Portability and Accountability Act
HITECH: Health Information Technology for Economic and Clinical Health
LLM: large language model
UNICEF: United Nations International Children's Emergency Fund
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Mavragani; submitted 19.11.24; peer-reviewed by B Davis, J John Thayil; comments to author 15.12.24; revised version received 20.12.24; accepted 21.12.24; published 20.01.25.

Please cite as:
Sezgin E, Kocaballi AB
Era of Generalist Conversational Artificial Intelligence to Support Public Health Communications
J Med Internet Res 2025;27:e69007
URL: https://www.jmir.org/2025/1/e69007
doi:10.2196/69007
PMID:

©Emre Sezgin, Ahmet Baki Kocaballi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Identifying Fraudulent Responses in a Study Exploring Delivery Options for Pregnancies Impacted by Gestational Diabetes: Lessons Learned From a Web-Based Survey

Emma Ruby[1,2], BHSc; Serine Ramlawi[2], MSc; Alexa Clare Bowie[2], MPH; Stephanie Boyd[2], PEng, MBA; Alysha Dingwall-Harvey[2], MSc, PMP; Ruth Rennicks White[2,3], BScN, RN; Darine El-Chaâr[2,3,4,5], MSc, MD; Mark Walker[2,3,4,5,6], MSc, MHM, MD

[1]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

[3]Department of Obstetrics, Gynecology, and Newborn Care, The Ottawa Hospital, Ottawa, ON, Canada

[4]Department of Obstetrics and Gynecology, University of Ottawa, Ottawa, ON, Canada

[5]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

[6]International and Global Health Office, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Mark Walker, MSc, MHM, MD
Clinical Epidemiology Program
Ottawa Hospital Research Institute
501 Smyth Road
Centre for Practice Changing Research, Box 241
Ottawa, ON, K1H8L6
Canada
Phone: 1 613 737 8899 ext 73840
Email: mwalker@toh.ca

Abstract

Current literature is unclear on the safety and optimal timing of delivery for pregnant individuals with gestational diabetes mellitus, which inspired our study team to conduct a web-based survey study exploring patient and provider opinions on delivery options. However, an incident of fraudulent activity with survey responses prompted a shift in the focus of the research project. Unfortunately, despite the significant rise of web-based surveys used in medical research, there remains very limited evidence on the implications of and optimal methods to handle fraudulent web-based survey responses. Therefore, the objective of this viewpoint paper was to highlight our approach to identifying fraudulent responses in a web-based survey study, in the context of clinical perinatal research exploring patient and provider opinions on delivery options for pregnancies with gestational diabetes mellitus. Initially, we conducted cross-sectional web-based surveys across Canada with pregnant patients and perinatal health care providers. Surveys were available through Research Electronic Data Capture, and recruitment took place between March and October 2023. A change to recruitment introduced a US $5 gift card incentive to increase survey engagement. In mid-October 2023, an incident of fraudulent activity was reported, after which the surveys were deactivated. Systematic guidelines were developed by the study team in consultation with information technology services and the research ethics board to filter fraudulent from true responses. Between October 14 and 16, 2023, an influx of almost 2500 responses (393 patients and 2047 providers) was recorded in our web-based survey. Systematic filtering flagged numerous fraudulent responses. We identified fraudulent responses based on criteria including, but not limited to, identical timestamps and responses, responses with slight variations in wording and similar timestamps, and fraudulent email addresses. Therefore, the incident described in this viewpoint paper highlights the importance of preserving research integrity by using methodologically sound practices to extract true data for research findings. These fraudulent events continue to threaten the credibility of research findings and future evidence-based practices.

(J Med Internet Res 2025;27:e58450) doi:10.2196/58450

XSL•FO
RenderX

KEYWORDS

research fraud; anonymous online research; data integrity; fraudulent responses; web-based survey; internet research; perinatal health; social media; patient participation; provider participation; fraudulent; fraud; pregnancy; gestational diabetes; diabetes; data analysis; survey; diabetes mellitus; patient; evidence-based

Introduction

Gestational diabetes mellitus (GDM) is one of the most common complications associated with pregnancy. In Canada, GDM is on the rise, with a near doubling of GDM diagnoses from 4% in 2004 to 7% in 2014 [1]. Globally, the rate of GDM is estimated to be 14.7%, according to the International Association of Diabetes and Pregnancy Study Groups [2].

Current literature is unclear on the safety and optimal timing of delivery for pregnant individuals with GDM. Canadian guidelines for delivery of GDM pregnancies differ from that of other countries, and discourse remains, even amongst health care providers [3-6]. Furthermore, literature exploring the perspectives of pregnant individuals and care providers on induction of labor (IOL) at 38 weeks compared with expectant management is limited.

Therefore, our study team initially sought to conduct a survey study exploring the topic of IOL and delivery options for pregnant individuals diagnosed with GDM. However, a significant and rapid influx of responses to the survey in mid-October 2023 prompted further investigation for fraudulent responses, shifting the focus of the study team's objectives with this research project.

Therefore, the goal of this viewpoint paper is to highlight our experience and approach to identifying fraudulent responses in a web-based survey study in the context of perinatal clinical research regularly conducted by the study team.

Challenges of Preserving Data Integrity in Studies With Vulnerable Populations

Particularly when conducting research with vulnerable populations such as pregnant individuals, there is a tension between maintaining participant anonymity while using techniques to prevent data fraud and protect study integrity [7]. Pregnant individuals have traditionally been excluded from research trials due to ethical concerns and misinformed ideas about clinical research [8]. Safety measures have always been at the forefront of pregnancy clinical research due to the medical complexities that pregnant individuals possess, and the additional consideration for fetal and newborn wellbeing. Therefore, providing a protected space where pregnant individuals can participate in research aimed to improve the health outcomes of themselves and their infants is integral for equitable and evidence-based practices [8]. However, breaches in data security and incident of fraudulent activity threaten this safe space.

Despite the significant rise of web-based surveys used in medical research, there remains very limited evidence on the implications of, and optimal methods to handle, fraudulent survey responses [9]. Unfortunately, there are no optimal methods or guidelines to address fraudulent incidents in research ethics or protocols [10-13]. Oftentimes, when these incidents take place, research teams must resort to less robust measures to address areas of fraud, such as filtering based on selection criteria, feasible participant timelines, duplication, repetitive responses, etc [10-13]. Therefore, challenges remain to systematically distinguish fraudulent from true responses to ensure research integrity and credibility of findings.

There is limited discourse on the incidence rates and impact of fraudulent responses to web-based surveys in the literature, making it difficult for researchers to raise awareness of these threats to data integrity. Reported incidents of data fraud in web-based studies are limited but suggest the need for rigorous protocols to distinguish fraudulent from true data to protect research integrity [7,9,11,12,14]. Less sophisticated methods used to differentiate fraud from true data have included tracking timestamps and time to completion of survey responses, using the Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA), and including several required open-ended questions to identify unusual responses [13,15,16]. More sophisticated methods have included logic checks to assess survey respondents' attentiveness, using skip logic, or using redundant questions to ensure consistency of responses to demographic questions [13,15,16].

Back to the Beginning: Methods for a Pregnancy Survey Study

Our team, composed of perinatal clinicians, researchers, and practicing obstetricians and gynecologists, originally sought to conduct a survey study exploring the perspectives of pregnant individuals and health care providers on their willingness to participate in research studies comparing elective IOL at 38 weeks of gestation versus expectant management for individuals diagnosed with GDM. However, we shifted our focus to explore the challenges of preserving data integrity in web-based research studies upon the discovery of a fraudulent incident with our survey responses.

Ethical Considerations

Ethics approval was obtained from the Ottawa Health Science Network Research Ethics Board (protocol ID: 20230075-01H). We updated recruitment to allow for compensation in the means of a CAD $5 (US $3.50) gift card to a local coffee shop upon completion of the survey for eligible participants. This required that participants provide their email address in a separate survey to which a link was made available upon completion of the main survey. We advertised this monetary incentive on all recruitment materials and in the participant consent form upon accessing the web-based survey. Any information provided by participants was kept strictly confidential and only available to the research team.

XSL•FO
RenderX

Participant Sample

Our initial survey study included 2 study groups: a patient population and a provider population. Patients were eligible to participate if they were currently pregnant with a diagnosis of GDM at the time of survey completion or up to 12 months postpartum with a diagnosis of GDM in their most recent pregnancy. Providers were eligible if they routinely provided care to the pregnant population and held a professional status that allowed them to counsel pregnant people on delivery recommendations. We sought to recruit a target sample size of 100 participants in each group (100 patients and 100 providers), totaling 200 participants. This sample size was selected to optimize study feasibility while sufficiently capturing the perspectives of pregnant people with GDM and health care providers.

Methods of Recruitment

We used convenience and snowball sampling to recruit participants. We conducted recruitment using various strategies, including posters in hospitals, social media posts on Twitter and Facebook, active recruitment at in-person clinics, review of electronic medical records, and the leverage of existing professional colleagues of the study team.

Survey Elements

We developed web-based surveys using Research Electronic Data Capture (REDCap) software and they took approximately 10 minutes to complete [17,18]. REDCap is a secure, web-based software platform designed to support data capture for research studies [17,18]. The surveys were cross-sectional in nature and were available as 2 different URL links (1 for each study group). The surveys were publicly available, thus anyone with access to the URL links had access to the surveys.

Survey Measures

Data was measured using a combination of Likert scale responses, free text response options, and multiple-choice questions. Data pertaining to demographic characteristics and likelihood to participate in or support a future randomized trial comparing induction of labor and expectant management for pregnancies impacted by GDM was obtained for both patient and provider groups.

Security Measures

In our survey study, to dissuade automated fraudulent responses, we included the CAPTCHA test, made available through the REDCap system. The URL links were made secure by generating them in REDCap, a data capture software that supports data security for research studies. In addition, the surveys were created in such a way that individuals deemed ineligible would not be allowed to advance to the next questions. This was accomplished using forced fields to ensure only eligible participants continued to survey questions.

The Incident and Aftermath

The GDM-related surveys were initially launched on March 9, 2023, and remained active until mid-October, 2023. Upon approval of an amendment on October 4, 2023, surveys were updated on October 11, 2023, to include a monetary incentive.

Part of our data management plan included a team member monitoring the survey responses on a regular weekly basis. Therefore, we were able to identify the onset of the fraudulent incident within a short timeframe. Once it was determined that the rapid influx of responses between October 14 and 16, 2023, was in fact a fraudulent incident, the research team swiftly deactivated the survey that same day and removed posters from clinical settings as well as social media posts. Our team reported the incident to the OHSN-REB for guidance on the next steps to address the fraudulent survey responses. The information technology teams were also alerted to help to address fraudulent responses.

We created a systematic guideline to filter fraudulent responses from true responses, by exporting the data from REDCap into Microsoft Excel and using criteria to indicate reasons for flagging responses as fraudulent (Figure 1). The responses were tallied according to their criteria and are summarized in the results.

XSL•FO
RenderX

Figure 1. Approach to fraudulent response filtering.

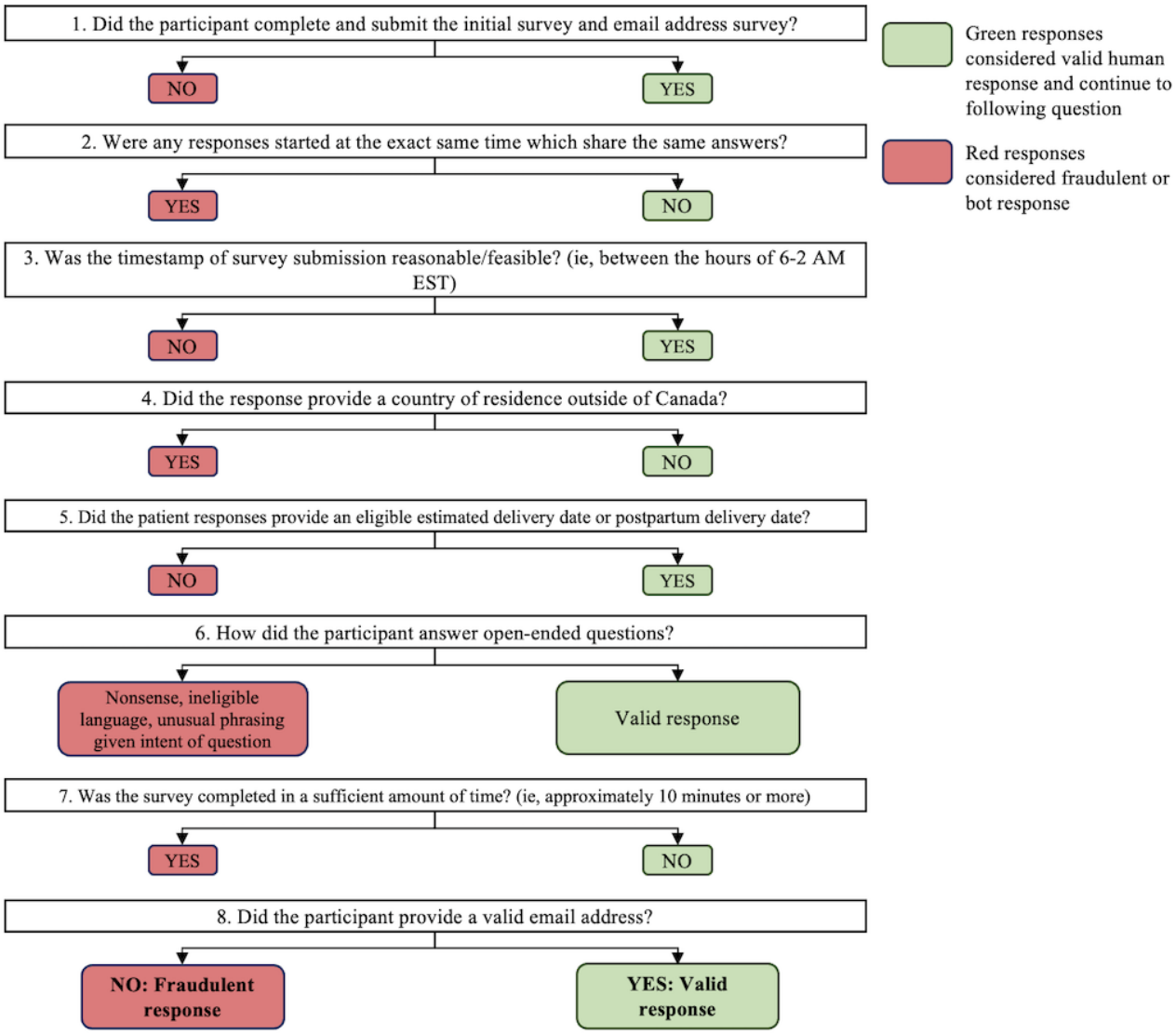

Our Findings

From March 9 to October 13, 2023, we had a total of 36 patient responses and 37 provider responses to the survey. Upon preliminary screening, 23 patient responses and 31 provider responses were deemed eligible and included in our analysis. Between October 14 and October 16, 2023, there were an additional 393 patient survey records and 2047 provider survey records submitted (Figure 2). Given the rapid influx of responses in such a short timeframe and the nature of our recruitment before this event, our team determined this to be an incident of fraudulent activity and used a systematic guide to identify fraudulent responses. Possible reasons for the drastic difference in the ratio of patient to provider fraudulent responses included: (1) the provider survey was shorter than the patient survey, (2) the provider survey asked fewer demographic questions, and (3) the provider survey required that fewer eligibility questions be completed to advance to the next survey questions.

XSL•FO
RenderX

Figure 2. Timeline for survey response influx and screening. See Multimedia Appendix 1 for reasons why patient survey records were flagged as fraudulent. See Multimedia Appendix 2 for reasons why provider survey records were flagged as fraudulent.

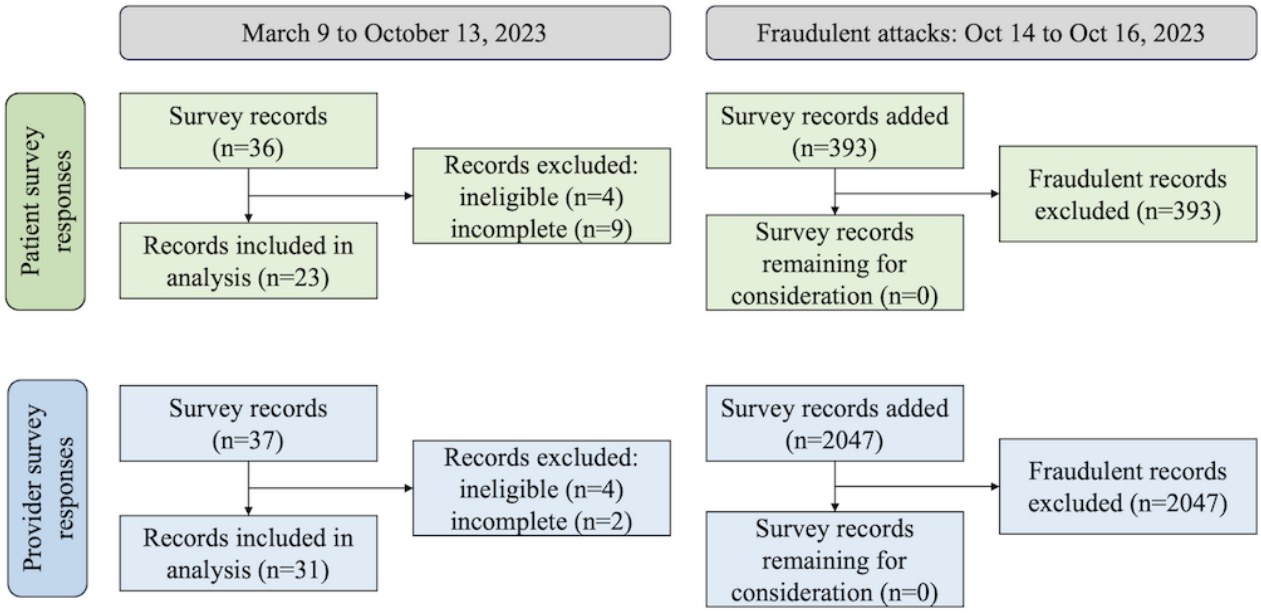

In consultation with the research ethics board and IT services, our team constructed a system to flag responses according to their fraudulent activity criteria (Figure 1). Criteria for filtering according to fraudulent versus true responses were applied. Following advice from the research ethics board, the surveys were inactivated and the study was closed to prevent further potential compromise to our results. Due to the limited discourse that exists on optimal methods for handling fraud in survey research, there was no obvious software that our team determined was best suited for addressing the incident. Therefore, we used Microsoft Excel files and applied a multireviewer, multidisciplinary, and multistep approach to filter fraudulent responses.

We reported reasons for why patient responses were flagged as fraudulent and tallied our findings using frequencies and percentage descriptive statistics (Multimedia Appendix 1). Reasons included an ineligible country provided or country not provided, ineligible estimated delivery date or postpartum delivery date (eg, date listed as “2004-02-01” or “1988-10-25,” which did not fall within the eligible 12 months postpartum as outlined in the eligibility criteria), same timestamp with exact same responses, same timestamp and similar responses with only slight variations in wording, responses not aligned with the intent of questions, fraudulent email address provided, and suspicious timestamp (eg, a multitude of responses being submitted between the hours of 2-6 AM Eastern Standard Time).

Reasons for why provider responses were flagged as fraudulent were also tallied and reported (Multimedia Appendix 2). Reasons included same timestamp with exact same responses, same timestamp and similar responses with only slight variations in wording, similar timestamp and similar responses, ineligible language, fraudulent email address provided, suspicious timestamp, and incomplete responses.

Criteria used to filter out fraudulent responses was defined in consultation with study team members. For example, the criteria “same timestamp and responses” was defined as any response that was written the exact same way with the same wording and submitted at the same time as any other response. Fraudulent email addresses were determined using the built-in Excel data validation function and email verification software to identify invalid mailboxes, invalid domains, spam traps, and syntax errors.

Given the significant number of fraudulent responses and the limited number of responses deemed to be “true,” the research team decided against conducting a between-group analysis of the responses included compared with those excluded.

Fraudulent Online Bots: A Black Box in Medical Research

This paper contributes to an unfortunately limited body of evidence supporting the need for increased data security measures when conducting medical research in the technological era. Despite the significant rise of web-based surveys used in medical research, there remains very limited evidence on the implications of, and optimal methods to handle, fraudulent survey responses [9]. However disappointing, adverse fraudulent events will likely continue to occur. These events therefore obscure the true results of medical research and pose a threat to the integrity of future evidence-based practices.

Unfortunately, there are no optimal methods or guidelines to address fraudulent incidents in research ethics or protocols [10-13]. Oftentimes, when these incidents take place, research teams must resort to less robust measures to address areas of fraud, such as filtering based on selection criteria, feasible participant timelines, duplication, repetitive responses, etc [10-13]. Therefore, challenges remain to systematically distinguish fraudulent from true responses to ensure research integrity and credibility of findings.

XSL•FO
RenderX

Previous literature highlights the importance of security measures to ensure the integrity of web-based survey responses. Such security measures have included using CAPTCHA to distinguish humans from bot responses, using software that collects IP addresses of its survey user, requiring the completion of an open text response to assess individuality, or requiring email verification for the completion and submission of the survey [9,11,14,19]. Furthermore, measures to avoid the attention of fraudulent respondents may include keeping recruitment periods short and avoiding advertisement of the survey in public spaces such as on social media or on posters in the hospitals [19]. Incidents such as these indicate the need for study protocols to outline contingency plans in the case of cyber-attacks that may compromise the credibility of their findings [11,19]. Manual validation, in addition to the automatic validation techniques used to filter duplicate survey responses, should be implemented and strengthened in future web-based research studies [11,19].

We recognize that many factors likely contributed to the influx of fraudulent activity. For example, the amendment to include a monetary incentive was activated at the end of a work week, providing more weekend time for individuals to submit responses. In addition, the survey was available as a single public link rather than a disposable link, providing a convenient means to share the surveys amongst other potential users. Another limitation to our study, and what was a potential factor for the influx of fraudulent responses, was the limited eligibility criteria, since patients were eligible if they were currently pregnant, or had been pregnant in the past 12 months, and diagnosed with GDM. Due to this restriction in selection criteria and our desired sample size of 200 participants, the recruitment period was prolonged to obtain an adequate sample size, increasing the risk for fraudulent responses to be submitted to the survey. However, despite evidence to suggest that recruitment of pregnant people in research studies is most effective through face-to-face means, our team determined that using multiple recruitment strategies would limit selection bias and increase the likelihood of reaching our desired sample size [13,20-22]. Finally, REDCap software, although useful for many research studies, does not collect the IP addresses of participants accessing surveys generated through its system. This posed a limitation to our team's ability to use robust methods to decipher fraudulent from true responses.

This viewpoint paper is unique in that it brings to light a rising concern for data security and research integrity in an era of mounting technological involvement in medical research. It highlights the real and raw experience of a group of medical researchers who pride themselves on conducting methodologically sound and scientifically robust perinatal health research. Therefore, this paper provides an example of the vulnerability of web-based medical research, and the extent of precautions that are necessary to prevent such incidents from occurring.

Conclusions

Overall, this fraudulent incident highlights the importance of preserving research integrity by using methodologically sound practices to extract true data for research findings. Despite numerous measures used to avoid fraudulent data in medical research, findings may still be compromised, particularly in web-based studies. We hope this case example provides future researchers with a cautionary tale to consider when conducting their own survey studies. Researchers must be aware of and actively work to prevent potential threats to the credibility of their research findings and future evidence-based practices.

Acknowledgments

Our study team would like to thank all patients and providers who participated by providing their genuine responses. Our team would also like to thank the Ottawa Health Science Network Research Ethics Board and Information Technology teams at the Ottawa Hospital Research Institute for their support and guidance of the fraudulent activity reporting protocol. This study was funded by a grant from the Canadian Institutes of Health Research (FRN 148438).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequencies and percentages of criteria for flagged fraudulent responses for gestational diabetes mellitus patient survey.
[DOCX File , 22 KB - jmir_v27i1e58450_app1.docx]

Multimedia Appendix 2
Frequencies and percentages of criteria for flagged fraudulent responses for gestational diabetes mellitus provider survey.
[DOCX File , 22 KB - jmir_v27i1e58450_app2.docx]

References

1. Metcalfe A, Sabr Y, Hutcheon JA, Donovan L, Lyons J, Burrows J, et al. Trends in obstetric intervention and pregnancy outcomes of Canadian women with diabetes in pregnancy from 2004 to 2015. J Endocr Soc 2017;1(12):1540-1549 [FREE Full text] [doi: 10.1210/js.2017-00376] [Medline: 29308448]

XSL•FO
RenderX

2. Saeedi M, Cao Y, Fadl H, Gustafson H, Simmons D. Increasing prevalence of gestational diabetes mellitus when implementing the IADPSG criteria: a systematic review and meta-analysis. Diabetes Res Clin Pract 2021;172:108642 [FREE Full text] [doi: 10.1016/j.diabres.2020.108642] [Medline: 33359574]
3. Lurie S, Insler V, Hagay ZJ. Induction of labor at 38 to 39 weeks of gestation reduces the incidence of shoulder dystocia in gestational diabetic patients class A2. Am J Perinatol 1996;13(5):293-296. [doi: 10.1055/s-2007-994344] [Medline: 8863948]
4. Berger H, Melamed N. Timing of delivery in women with diabetes in pregnancy. Obstet Med 2014;7(1):8-16 [FREE Full text] [doi: 10.1177/1753495X13513577] [Medline: 27512412]
5. Kruit H, Mertsalmi S, Rahkonen L. Planned vaginal and planned cesarean delivery outcomes in pregnancies complicated with pregestational type 1 diabetes - a three-year academic tertiary hospital cohort study. BMC Pregnancy Childbirth 2022;22(1):173 [FREE Full text] [doi: 10.1186/s12884-022-04510-8] [Medline: 35236314]
6. . ACOG practice bulletin No. 190: gestational diabetes mellitus. Obstet Gynecol 2018;131(2):e49-e64. [doi: 10.1097/AOG.0000000000002501] [Medline: 29370047]
7. Shivayogi P. Vulnerable population and methods for their safeguard. Perspect Clin Res 2013;4(1):53-57 [FREE Full text] [doi: 10.4103/2229-3485.106389] [Medline: 23533983]
8. Sewell CA, Sheehan SM, Gill MS, Henry LM, Bucci-Rechtweg C, Gyamfi-Bannerman C, et al. Scientific, ethical, and legal considerations for the inclusion of pregnant people in clinical trials. Am J Obstet Gynecol 2022;227(6):805-811 [FREE Full text] [doi: 10.1016/j.ajog.2022.07.037] [Medline: 35934117]
9. Wang J, Calderon G, Hager ER, Edwards LV, Berry AA, Liu Y, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. PLOS Glob Public Health 2023;3(8):e0001452 [FREE Full text] [doi: 10.1371/journal.pgph.0001452] [Medline: 37610999]
10. Hohn KL, Braswell AA, DeVita JM. Preventing and protecting against internet research fraud in anonymous web-based research: protocol for the development and implementation of an anonymous web-based data integrity plan. JMIR Res Protoc 2022;11(9):e38550 [FREE Full text] [doi: 10.2196/38550] [Medline: 36094806]
11. Pratt-Chapman M, Moses J, Arem H. Strategies for the identification and prevention of survey fraud: data analysis of a web-based survey. JMIR Cancer 2021;7(3):e30730 [FREE Full text] [doi: 10.2196/30730] [Medline: 34269685]
12. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. JMIR Public Health Surveill 2019;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]
13. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
14. Conrique BI, McDade-Montez E, Anderson PM. Detection and pevention of data fraud in a study of community college career technical education students. Community College Journal of Research and Practice 2019;44(2):1-4. [doi: 10.1080/10668926.2019.1629126]
15. Simone M. Bots started sabotaging my online research. I fought back. STAT News. 2019. URL: https://www.statnews.com/2019/11/21/bots-started-sabotaging-my-online-research-i-fought-back/ [accessed 2024-12-09]
16. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in internet research: preserving data integrity. AIDS Behav 2008;12(6):964-973 [FREE Full text] [doi: 10.1007/s10461-007-9352-2] [Medline: 18240015]
17. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
18. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
19. Dewitt J, Capistrant B, Kohli N, Rosser BRS, Mitteldorf D, Merengwa E, et al. Addressing participant validity in a small internet health survey (The Restore Study): protocol and recommendations for survey response validation. JMIR Res Protoc 2018;7(4):e96 [FREE Full text] [doi: 10.2196/resprot.7655] [Medline: 29691203]
20. Manca DP, O'Beirne M, Lightbody T, Johnston DW, Dymianiw DL, Nastalska K, APrON study team. The most effective strategy for recruiting a pregnancy cohort: a tale of two cities. BMC Pregnancy Childbirth 2013;13:75 [FREE Full text] [doi: 10.1186/1471-2393-13-75] [Medline: 23521869]
21. Frew PM, Saint-Victor DS, Isaacs MB, Kim S, Swamy GK, Sheffield JS, et al. Recruitment and retention of pregnant women into clinical research trials: an overview of challenges, facilitators, and best practices. Clin Infect Dis 2014;59 Suppl 7(Suppl 7):S400-S407 [FREE Full text] [doi: 10.1093/cid/ciu726] [Medline: 25425718]
22. Strömmer S, Lawrence W, Rose T, Vogel C, Watson D, Bottell JN, et al. Improving recruitment to clinical trials during pregnancy: a mixed methods investigation. Soc Sci Med 2018;200:73-82 [FREE Full text] [doi: 10.1016/j.socscimed.2018.01.014] [Medline: 29421474]

XSL•FO
RenderX

Abbreviations

CAPTCHA: Completely Automated Public Turing Test to Tell Computers and Humans Apart
GDM: gestational diabetes mellitus
IOL: induction of labor
REDCap: Research Electronic Data Capture

Edited by T de Azevedo Cardoso; submitted 21.03.24; peer-reviewed by Q Nian, S Hewage, JK Sinamo, M Ennis; comments to author 14.05.24; revised version received 28.06.24; accepted 30.10.24; published 20.01.25.

Please cite as:
Ruby E, Ramlawi S, Bowie AC, Boyd S, Dingwall-Harvey A, Rennicks White R, El-Chaâr D, Walker M
Identifying Fraudulent Responses in a Study Exploring Delivery Options for Pregnancies Impacted by Gestational Diabetes: Lessons Learned From a Web-Based Survey
J Med Internet Res 2025;27:e58450
URL: https://www.jmir.org/2025/1/e58450
doi:10.2196/58450
PMID:

©Emma Ruby, Serine Ramlawi, Alexa Clare Bowie, Stephanie Boyd, Alysha Dingwall-Harvey, Ruth Rennicks White, Darine El-Chaâr, Mark Walker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

AI Can Be a Powerful Social Innovation for Public Health if Community Engagement Is at the Core

Alessandra N Bazzano[1,2], MPH, PhD; Andrea Mantsios[3], MHS, PhD; Nicholas Mattei[2,4], PhD; Michael R Kosorok[5,6], PhD; Aron Culotta[2,4], PhD

[1]Department of Maternal and Child Health, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Center for Community-Engaged Artificial Intelligence, School of Science & Engineering, Tulane University, New Orleans, LA, United States

[3]Public Health Innovation and Action, New York, NY, United States

[4]Department of Computer Science, School of Science & Engineering, Tulane University, New Orleans, LA, United States

[5]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]Gillings Center for Artificial Intelligence and Public Health, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Alessandra N Bazzano, MPH, PhD
Department of Maternal and Child Health
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
135 Dauer Drive, CB #7445
Chapel Hill, NC, 27599-7445
United States
Phone: 1 919 966 9306
Email: abazzano@tulane.edu

Abstract

There is a critical need for community engagement in the process of adopting artificial intelligence (AI) technologies in public health. Public health practitioners and researchers have historically innovated in areas like vaccination and sanitation but have been slower in adopting emerging technologies such as generative AI. However, with increasingly complex funding, programming, and research requirements, the field now faces a pivotal moment to enhance its agility and responsiveness to evolving health challenges. Participatory methods and community engagement are key components of many current public health programs and research. The field of public health is well positioned to ensure community engagement is part of AI technologies applied to population health issues. Without such engagement, the adoption of these technologies in public health may exclude significant portions of the population, particularly those with the fewest resources, with the potential to exacerbate health inequities. Risks to privacy and perpetuation of bias are more likely to be avoided if AI technologies in public health are designed with knowledge of community engagement, existing health disparities, and strategies for improving equity. This viewpoint proposes a multifaceted approach to ensure safer and more effective integration of AI in public health with the following call to action: (1) include the basics of AI technology in public health training and professional development; (2) use a community engagement approach to co-design AI technologies in public health; and (3) introduce governance and best practice mechanisms that can guide the use of AI in public health to prevent or mitigate potential harms. These actions will support the application of AI to varied public health domains through a framework for more transparent, responsive, and equitable use of this evolving technology, augmenting the work of public health practitioners and researchers to improve health outcomes while minimizing risks and unintended consequences.

(J Med Internet Res 2025;27:e68198) doi:10.2196/68198

KEYWORDS

Artificial Intelligence; Generative Artificial Intelligence; Citizen Science; Community Participation; Innovation Diffusion

XSL•FO
RenderX

Introduction

While public health has a rich history of innovation in areas such as vaccination, sanitation, and epidemiology [1], the field has been cautious in the widespread adoption of emerging technologies like generative artificial intelligence (GenAI). Calls for updated approaches to public health funding, programming, and research have been pervasive amid the changing landscape of health in the United States and globally [2]. The field of public health is at a crossroads, with growing recognition of the need for innovation to afford greater agility and responsiveness to emerging problems [3].

GenAI applications are expected to impact all sectors of society. Many public health practitioners and researchers are understandably cautious about these developments and how they will affect the discipline and the communities we serve—the exception is in bioinformatics and biostatistics, where machine learning and other techniques have long been applied to medical and health care data [4]. Given the speed of technology, there is a risk that significant portions of the population—particularly those with the fewest resources—will be marginalized by the incursion of GenAI technologies. An analogous example is the burden placed on individuals to use emerging technological tools for health care (eg, patient portals, telehealth visits, pharmacy apps), which disproportionately exclude low-income and aging populations. The present danger is that the AI revolution in public health may exacerbate well-known health disparities, both through new generative tools as well as common AI or informatics techniques being applied more broadly. As public health professionals, we must educate ourselves and engage communities in knowledge seeking around AI and its impact on their lives, to fulfill the goals of promoting and protecting the health of all people and their communities.

Adoption of technological innovations may be challenging, but working with communities has been a strength of the contemporary public health profession. The field has embraced community engagement and participation as a norm, in a way that can be leveraged to ensure that AI and all its applications are an innovation that benefits society and promotes health equity. Given that community-engaged design and participatory methods have become embedded in many public health projects [5], the field is primed for outreach to communities to ensure that individuals understand the benefits and risks of AI in public health. Community involvement in AI design and implementation is key for helping ensure that AI is developed in a person-centered way, particularly addressing the needs and desires of underrepresented and minoritized communities that often face greater health disparities [6]. Beyond health care applications, where tools have often been used for clinical decision-making, outcome prediction, and clinical record summarization, there are numerous use cases for AI in public health. Infectious disease surveillance and risk assessment have been a key area for the use of this technology [7,8]. The use of AI in global health has also been described for improving vaccine delivery and community health care efficiency [9], and as the digitalization of public health continues, the use of AI applications for monitoring and evaluation will become more frequent [10,11]. Chatbots developed with GenAI are another increasingly common approach to providing health education and behavior change support [12]. Large language models are being used to support question answering over large health datasets, allowing more informed decision-making [13]. There is also an urgent need to combat health-related misinformation and disinformation through understanding and using GenAI, as experts predict it could generate up to 90% of web-based content by 2026 [14].

But what about the ethics of using this technology when both public health professionals and people whose data are being used to build applications of AI do not have an understanding of how these systems work? Public health researchers, practitioners, and community partners will need to understand the basics of AI applications and use cases in public health, to be able to prevent and mitigate harm to fully realize the potential benefits of AI as a social innovation. Promoting the responsible and ethical use of AI in public health will require a proactive approach to learning that requires openness. Public health practitioners may view AI applications as the domain of technologists or informaticists only, but communities must be involved in codeveloping these technologies. The ethical implications of AI in public health necessitate transparency, accountability, and equity, all of which require the involvement of scholars, practitioners, and people with lived experience [15]. Education and awareness for communities and public health researchers are urgently needed.

While there are both real and perceived risks of privacy violation posed by using AI in health [16], another important potential harm is the risk of introducing or perpetuating biases and thus exacerbating existing health disparities. There are cases where such biases have led to individuals being denied access to health services or insurance, which may result in discrimination and violations of individual rights [17]. Statistical models are based on underlying assumptions, and these can have far-reaching consequences. Knowledge of community engagement, health inequalities, and strategies for improving equity are crucial to avoid inadvertent reinforcement of biases embedded in data. A well-known example is a racially biased single health care algorithm that drives important chronic care management for over 70 million people in the United States [18]. Because of significant racial bias in this widely used algorithm, White patients with the same health as Black patients were far more likely to be enrolled in a care management program and benefit from its resources. Public health domain knowledge in social determinants and structural inequities must be part of the developing uses of AI in population health science and practice.

How the Public Health Community Can Take Action

For safer and more effective integration of AI in public health, a multifaceted approach should involve the following: (1) inclusion of the basics of AI technology in public health training and professional development; (2) community engagement for co-design of AI technologies applied to public health; and (3) introduction of governance and best practice mechanisms by

XSL•FO
RenderX

public health organizations that can guide the use of AI in public health and prevent or mitigate potential harms.

First, public health programs and organizations should begin incorporating training on the fundamentals of AI technology into public health education and professional development. This will help ensure that public health professionals possess a foundational understanding of AI, allowing for critical assessment and effective use of AI tools in population health work. This training should not only cover the technical aspects of AI but also include a critical study of ethical considerations, known and unknown biases, and a component of the sociocultural implications of AI-driven decisions. Schools of public health are beginning to develop centers and training initiatives for this purpose, including the authors' institutions, UNC Chapel Hill Gillings School of Global Public Health and Tulane University. A scoping review to map the literature from both public health and computer science databases is currently underway by the authors and will provide additional sources of training material for public health and AI education efforts. Furthermore, as a field, we must not only educate ourselves but ensure communities are informed about evolving AI use in public health and its impact on their lives by incorporating this into public health education.

Second, engagement and collaboration between technologists, public health professionals, and community members are vital in the development of AI technologies for public health. Involving community stakeholders from the outset will develop applications with a better understanding of the specific needs, desires, and values of the populations they are intended to serve [19]. Co-design processes can promote transparency, and when done responsibly and with care, can build trust, as they allow for the direct input of those who will be most affected by the technology [20,21]. An example of this is the partnership between the AFL-CIO and Carnegie Mellon University Human-Computer Interaction Institute, which engages workers in the innovation process around AI tools to address critical challenges and open pathways for collaboration. Hospitality workers partnered with university faculty to explore the potential impacts of AI in the daily execution of their jobs. Algorithmic managers, which focused on optimizing operations and the workflow of cleaning rooms, inadvertently increased work, creating potential health hazards, and decreased worker autonomy. In the partnership, stakeholders from the community, labor leaders, and academics engaged to mitigate potential health and occupational harms created by the application of AI [22]. This participatory approach, often a hallmark of contemporary public health approaches, enhances the relevance and acceptability of AI solutions that touch individuals and communities while also identifying and addressing potential biases or risks for disparity early in the development process.

Finally, the introduction of governance frameworks and best practice mechanisms by public health organizations is essential for guiding the responsible use of AI. These mechanisms should be designed to prevent or mitigate potential harms associated with AI, such as privacy violations, biased outcomes, or the exacerbation of health disparities. Governance structures should include clear guidelines on data management, algorithmic transparency, accountability, and ongoing monitoring and evaluation of AI applications. Additionally, the establishment of interdisciplinary oversight committees, including ethicists, data scientists, public health experts, and community members, can provide ongoing surveillance and ensure that AI tools are aligned with public health goals and ethical standards. One example is the National Institute of Health's Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity Ethics and Equity Workgroup. The Ethics and Equity Workgroup developed a set of ethics and equity principles, a glossary, and an interview guide. The ethical and equity principles comprise 5 core principles that articulate best practices for working with stakeholders from historically and presently underrepresented communities [23].

Together, these actions can potentiate the application of AI to varied public health domains, and likewise improve outcomes, through a framework for more transparent, responsive, and equitable use of this evolving technology, augmenting the work of public health practitioners and researchers to improve health outcomes while minimizing risks and unintended consequences. Without such safeguards, the application of AI technology in public health may continue to be dominated by technologists whose domain expertise does not include biomedical ethics training, a human rights framework, or an understanding of health equity and social determinants as they inform the central work of improving population health.

Acknowledgments

This work was supported by the Center for Community-Engaged AI through the Tulane University Office of Research and the Gillings Center for AI and Public Health, UNC Chapel Hill Gillings School of Global Public Health.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Ten great public health achievements—United States, 2001-2010. MMWR Morb Mortal Wkly Rep 2011;60(19):619-623 [FREE Full text] [Medline: 21597455]
2. DiClemente R, Nowara A, Shelton R, Wingood G. Need for innovation in public health research. Am J Public Health 2019;109(S2):S117-S120. [doi: 10.2105/AJPH.2018.304876] [Medline: 30785791]

XSL•FO
RenderX

3. Halpaap BM, Tucker JD, Mathanga D, Juban N, Awor P, Saravia NG, et al. Social innovation in global health: sparking location action. Lancet Glob Health 2020;8(5):e633-e634 [FREE Full text] [doi: 10.1016/S2214-109X(20)30070-X] [Medline: 32353305]
4. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
5. Yuan M, Lin H, Wu H, Yu M, Tu J, Lü Y. Community engagement in public health: a bibliometric mapping of global research. Arch Public Health 2021;79(1):6 [FREE Full text] [doi: 10.1186/s13690-021-00525-3] [Medline: 33436063]
6. Manyika J, Silberg J, Presten B. What do we do about the biases in AI? Harvard Business Review. 2019. URL: https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai [accessed 2025-01-08]
7. Zhao AP, Li S, Cao Z, Hu PJH, Wang J, Xiang Y, et al. AI for science: predicting infectious diseases. J Saf Sci Resilience 2024;5(2):130-146. [doi: 10.1016/j.jnlssr.2024.02.002]
8. Doan S, Ngo QH, Kawazoe A, Collier N. Global health monitor—a web-based system for detecting and mapping infectious diseases. In: Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II. 2008 Presented at: Third International Joint Conference on Natural Language Processing; January 7-12, 2008; Hyderabad, India URL: https://aclanthology.org/I08-2140/
9. Anjaria P, Asediya V, Bhavsar P, Pathak A, Desai D, Patil V. Artificial intelligence in public health: revolutionizing epidemiological surveillance for pandemic preparedness and equitable vaccine access. Vaccines 2023;11(7):1154 [FREE Full text] [doi: 10.3390/vaccines11071154] [Medline: 37514970]
10. AI for assisting NGOs in improving maternal and child health outcomes. Harvard University. 2023. URL: https://teamcore.seas.harvard.edu/ai-assisting-ngos-improving-maternal-and-child-health-outcomes?admin_panel=1 [accessed 2024-10-28]
11. Leveraging technology to create scalable solutions to empower mothers and enable the nurturing of healthy children. Armman. URL: https://armman.org/ [accessed 2024-10-28]
12. Moore N, Amith M, Neumann AC, Hamilton J, Tang L, Savas LS, et al. Translating motivational interviewing for the HPV vaccine into a computable ontology model for automated AI conversational interaction. Ext Abstr Hum Factors Computing Syst 2024:341 [FREE Full text] [doi: 10.1145/3613905.3651051] [Medline: 38898884]
13. Guangzhi X, Jin Q, Lu Z, Zhang A. Benchmarking retrieval-augmented generation for medicine. 2024 Presented at: Findings of the Association for Computational Linguistics ACL 2024; August 11-16, 2024; Bangkok, Thailand p. 6233-6251. [doi: 10.18653/v1/2024.findings-acl.372]
14. Facing reality? Law enforcement and the challenge of deepfakes. Europol. 2022. URL: https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes [accessed 2025-01-08]
15. Zhang K, Ran B. Active health governance—a conceptual framework based on a narrative literature review. Int J Environ Res Public Health 2022;19(4):2289 [FREE Full text] [doi: 10.3390/ijerph19042289] [Medline: 35206476]
16. Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' perceptions toward human-artificial intelligence interaction in health care: experimental study. J Med Internet Res 2021;23(11):e25856 [FREE Full text] [doi: 10.2196/25856] [Medline: 34842535]
17. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020;20(1):170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
18. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
19. Lee MK, Kusbit D, Kahng A, Kim JT, Yuan X, Chan A, et al. WeBuildAI: Participatory framework for algorithmic governance. Proc ACM Hum Comput Interact 2019;3:1-35. [doi: 10.1145/3359283]
20. Timon CM, Heffernan E, Kilcullen S, Hopper L, Lee H, Gallagher P, et al. Developing independent living support for older adults using internet of things and AI-based systems: co-design study. JMIR Aging 2024;7:e54210 [FREE Full text] [doi: 10.2196/54210] [Medline: 39447196]
21. Panigutti C, Beretta A, Fadda D, Giannotti F, Pedreschi D, Perotti A, et al. Co-design of human-centered, explainable AI for clinical decision support. ACM Trans Interact Intell Syst 2023;13(4):1-35. [doi: 10.1145/3587271]
22. Markley J. Equipping workers for the AI future: CMU researchers are helping frontline workers shape the technology that is transforming their work. Carnegie Mellon University. 2022. URL: https://makepossible.cmu.edu/impact-stories/workforce-development/ [accessed 2024-10-08]
23. Hendricks-Sturrup R, Simmons M, Anders S, Aneni K, Wright Clayton E, Coco J, et al. Developing ethics and equity principles, terms, and engagement tools to advance health equity and researcher diversity in AI and machine learning: modified delphi approach. JMIR AI 2023;2:e52888 [FREE Full text] [doi: 10.2196/52888] [Medline: 38875540]

Abbreviations

AI: artificial intelligence
GenAI: generative artificial intelligence

XSL•FO
RenderX

Edited by A Mavragani; submitted 13.11.24; peer-reviewed by M Eder, G Xiao; comments to author 11.12.24; revised version received 20.12.24; accepted 02.01.25; published 22.01.25.

Please cite as:
Bazzano AN, Mantsios A, Mattei N, Kosorok MR, Culotta A
AI Can Be a Powerful Social Innovation for Public Health if Community Engagement Is at the Core
J Med Internet Res 2025;27:e68198
URL: https://www.jmir.org/2025/1/e68198
doi:10.2196/68198
PMID:39841529

©Alessandra N Bazzano, Andrea Mantsios, Nicholas Mattei, Michael R Kosorok, Aron Culotta. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Regulation and Compliance in Telemedicine: Viewpoint

Julia Ivanova[1], PhD; Mollie R Cummins[1,2], PhD, RN; Triton Ong[1], PhD; Hiral Soni[1], MBA, MS, PhD; Janelle Barrera[1,3], MPH; Hattie Wilczewski[1], BS; Brandon Welch[1,4], PhD; Brian Bunnell[1,3], PhD

[1]Doxy.me Research, Doxy.me, Inc, Charleston, SC, United States

[2]College of Nursing, University of Utah, Salt Lake City, UT, United States

[3]Department of Psychiatry and Behavioral Neurosciences, Morsani College of Medicine, University of South Florida, Tampa, FL, United States

[4]Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Julia Ivanova, PhD
Doxy.me Research
Doxy.me, Inc
18 Broad Street
Unit 300
Charleston, SC, 29401
United States
Phone: 1 844 436 9963
Email: julia.ivanova@doxy.me

Abstract

The US COVID-19 Public Health Emergency ended on May 11, 2023. Lawmakers and regulators extended some flexibilities while they deliberate effective long-term telemedicine policy. Here, we discuss critical challenges in telemedicine compliance and regulation grounded in scholarly literature and current events. We specifically consider obstacles and progress toward solutions in telemedicine law and regulation regarding privacy and security issues, care across state borders, and prescribing over telemedicine in the United States. We conclude that simplified policies are needed to keep telemedicine accessible to providers and patients and that current privacy and security measures need refinement to protect patients appropriately.

(J Med Internet Res 2025;27:e53558) doi:10.2196/53558

KEYWORDS

telemedicine; telehealth; policy; COVID-19; PHE; rules and regulations; compliance; privacy and security; regulation; rule; public health; US; United States; implementation; regulatory; professional; organizational; ethical; concern; privacy; security; government literature; law; health care; patient

Introduction

Background

Smartphones and the internet have had a transformative impact on telemedicine over the decades, but the tipping point for telemedicine adoption was the COVID-19 pandemic [1]. During the pandemic, numerous health professionals adopted telemedicine to provide care while decreasing exposure to the virus [2]. In the US, legal and regulatory flexibilities enabled the shift of health care services to telemedicine. The Coronavirus Aid, Relief, and Economic Security Act and the Public Health Emergency (PHE) protocols of 2020 implemented most pandemic-era policy flexibilities in the United States, including (1) relaxation of HIPAA (Health Insurance Portability and Accountability Act) compliance requirements, (2) permission to prescribe certain controlled substances via telemedicine, (3) reimbursement for telemedicine equivalent to that of in-person visits, (4) promotion of telemedicine expansion for Medicare, (5) allowances for providers to offer care across state borders regardless of licensure, and (6) new protocols and *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) codes for telemedicine visits [3,4]. Lawmakers made these changes to increase access to health care nationwide during the pandemic and based these flexibilities on previously considered legislative, regulatory, and policy updates [5].

The expiration of the PHE in May 2023 created a dilemma for lawmakers, who were unprepared to make decisions about long-term telemedicine policy. To avoid negative impacts and allow more time for decision-making, they extended many telemedicine-related policy flexibilities through December 31, 2024 [6,7]. Now, regulatory leaders are tasked with assessing current evidence (eg, telemedicine research, stakeholder needs and experiences, security concerns, cases of crime, and fraud)

XSL•FO
RenderX

to make informed decisions about permanent federal and state regulations and rules related to telemedicine [8].

At this critical juncture, it is essential that we consider outcomes, experiences, perspectives, and other evidence resulting from pandemic-era telemedicine policy. By doing so, we can identify critical challenges and potential solutions and ultimately make informed, long-term policy decisions. Here, we summarize high-priority, policy-related challenges, and proposed solutions, focusing on providers and patients. We argue for US federal and state policy that protects patients in the context of new technologies and supports providers' inclusion of telemedicine in their health care services. We highlight how ambiguities and complexity of laws and regulations effectively obstruct providers and organizations from using telemedicine. This viewpoint is grounded in a literature review of US telemedicine laws and regulations between 2000 and 2022 using legal and scientific databases (ie, HeinOnline, LexisNexis, PubMed, and Google Scholar) and includes post-PHE literature and events to ensure the relevancy of the topics discussed.

Security and Privacy

HIPAA—known for its privacy and security rules establishing national standards for protecting personal health information (PHI)—mandates providers use telemedicine platforms that provide privacy and security by not storing or having sufficient ways to prevent breaches of PHI [9]. The PHE flexibilities waived the need to use HIPAA-compliant telemedicine platforms. This flexibility eased the rapid, large-scale transition to telemedicine use during COVID-19 [9,10]. Unfortunately, HIPAA does not cover all facets of digital practices in health care delivery. For example, third-party digital apps for patients to complete therapies or exercises may not be required to comply with HIPAA, even if providers use it as an adjunct to care [11-13]. Providers may use a HIPAA-compliant platform, but patients may be unknowingly in a nonprivate space. Simply expanding the scope of HIPAA to include PHI shared outside of a direct patient-provider relationship may not be enough protection [13]. Even in 2018, legal commentary noted that updating the laws and regulations to protect PHI through more rigorous cybersecurity expectations, such as requiring encryption of transmissions and concrete language regarding security measures, is critical in ensuring the security and privacy of PHI [13]. HIPAA's security rules should stipulate the use of feasible cybersecurity measures rather than the basic health care standards [13]. A health care practice could determine the best cybersecurity measures that would balance the importance of privacy and security of their patients while ensuring such measures are not a costly burden. To protect patients and providers, we must reconsider emphasizing the best instead of minimum necessary practices regarding privacy and security laws in this new age of digital health telemedicine and apps.

While providers have stated expectations that HIPAA amendments increase protections [14,15], the proposed changes focus on improving patient access to their PHI and easing administrative burdens [16]. These proposed changes would only apply to the current definition of covered entities, and therefore, will not address the challenges regarding third-party health apps. Though HIPAA is the federal standard for PHI protection, state laws can enhance protection above this minimum requirement.

At the state level, the issue of PHI protection has led to the enactment of Washington state's My Health My Data Act (HB 1155), whereby the state protects PHI beyond the scope of HIPAA [17]. Specifically, this law prevents businesses from inferring health data about an individual based on purchases. It requires regulated entities or businesses to have information on this law within their privacy policies [17]. The My Health My Data Act reflects a general movement toward increased security of individual health information. As the first to pass, the law broadly applies to consumer health data and includes information that may be inferred to determine a person's past, current, or future health [18]. Based on the discussions around this and similar proposed laws, we expect to see more proposed state and federal legislation with varying degrees of protection, greatly based on the ability of the government and institutions to enforce such protections [19-21]. With no current tangible proposal to amend HIPAA in this way, state legislatures are stepping up their security and protections of PHI: Nevada passed their consumer health law in 2023, though their definitions of health data and protections are more narrowly defined than those of Washington's law [22].

At this time, planned HIPAA amendments focus on increasing patient access to their PHI and would not directly address concerns related to digital tools. Some states are attempting to remedy privacy and security issues identified with the increased use of telemedicine and new technologies. Though their initiative raises awareness and discussion of privacy and security related to telemedicine, the lack of streamlining may increase legal complexities. Until there is a concrete effort to streamline these issues, telemedicine providers must ensure compliance with new state laws and regulations, and patients must be mindful of how they share PHI.

Prescribing Over Telemedicine

The PHE flexibilities waived the requirement to see a patient in person before prescribing controlled medication including schedule II-V drugs. Before the flexibilities, the Drug Enforcement Administration (DEA) required providers to obtain a waiver for each state where they prescribe medication. The DEA waived this step during the PHE [23]. Still, providers had to consider different state laws regulating their prescribing ability [24]. The DEA expects to promulgate its final rule incorporating provider and patient expectations since the PHE into its regulations in 2025. Due to this extended timeline, experts' expectations of a third extension of PHE flexibilities was confirmed with the DEA's clean extension of these flexibilities through December 31, 2025 [23,24].

Patient and provider experiences with these flexibilities have led to robust discourse, especially over the dangers—or lack thereof—of prescribing buprenorphine (schedule III drug) without in-person visits [25]. During the pandemic, providers reported that reduced steps in monitoring made them rely more on patients' decision-making and confront the issue of trust in their patient-provider relationships [26]. However, the same

XSL•FO
RenderX

providers noted the ability to see patients in their homes provided them with a new level of insight that they used in assessing their patients [26]. Similarly, a qualitative analysis of provider experiences in prescribing medications for opioid use disorder indicated both positive—including increased insight into patients' daily lives—and negative experiences, noting issues with patients viewing telemedicine as a less professional encounter [27]. Providers additionally experienced stress due to their uncertainty regarding prescribing laws [27].

Research on prescribing over telemedicine considered patient outcomes, as well as provider experiences. A study examining Medicaid data during the beginning of the pandemic found that patients who began buprenorphine treatment over telemedicine had higher treatment retention, and there was no change in the odds of opioid-related nonfatal overdose [28]. A recent review concluded that most providers favored having the ability to prescribe buprenorphine over telemedicine with little to no issues [29]. While the evidence is not definitive, providers appear to favor having the option of prescribing buprenorphine over telemedicine. However, providers note certain contextual misgivings such as the potential for increased liability and lack of legal clarity [29]. Though evidence appears to show that telemedicine increases access and results in similar efficacy as in-person treatment using medications for opioid use disorder, the research is still ongoing [30,31].

Ideally, there will be a continuation of rigorous research to determine how prescribing buprenorphine—and other controlled substances—over telemedicine affects efficacy, access, health inequalities, and patient safety [31]. Unfortunately, with the DEA's current timeline, new research is unlikely to inform the final rule. Instead, the DEA will make its decision based on current evidence, the 38,000 written comments providing feedback to its proposed rule, and the two days of public listening sessions that occurred in September 2023 [6]. The DEA's final rule is the first step to simplifying the laws and regulations surrounding prescribing over telemedicine, mainly controlled substances. We expect to see moderately more expansive flexibilities than the initially proposed ruling to support the necessary accessibility of opioid use disorder treatment for individuals. Additionally, should the DEA remove the requirement for individual waivers to prescribe in each state, there would be a significant streamlining of procedures providers must follow to prescribe over telemedicine. Considering the DEA's development of a federal ruling, eliminating the required multiple waivers would align with the premise of reducing federal compliance and regulatory complexities.

The DEA's final rule also affects how states make rules and regulations. For example, in late 2023, the Georgia Medical Board planned to revert its prescribing policies to prepandemic times. After discussions with the community, it decided to extend PHE flexibilities on the state level through January 31, 2025. The board is waiting to see the DEA's final ruling before determining its state policy on the subject [32]. As with many federal and state dynamics, states will examine national rules and regulations to determine their policies. Based on the DEA's final ruling, we expect to see further changes to laws and regulations at the state level.

Expansion of Medicare, Coverage, and Reimbursement

A widely cited barrier to telemedicine use is the issue of payment, both for coverage and reimbursement. Parity laws relate to payment for telemedicine services and include coverage parity (insurers must pay for services) and payment parity (insurers must pay for services at an equal rate to in-person visits) [33]. With the PHE flexibilities, coverage parity ensured that telemedicine services were provided in all states and through Medicare [34]. Additionally, the flexibilities implemented payment parity for telemedicine services through Medicare so that providers were paid at the same rate as they would for an in-person visit [35]. During this time, most commercial insurers also provided payment parity [36].

Payment parity on the state level has changed during the COVID-19 era (Table 1) [35]. On a federal level, Medicare payment parity continues through the end of 2024; however, reimbursement after this extension will depend on Medicare's telemedicine code categories [37].

Table 1. Number of states with pre-, peri-, and post-PHE[a] payment parity laws.

	States, n
Pre-PHE	
Medicaid	4
All insurers	11
PHE	
Medicaid	14
All insurers	20
Post-PHE	
Medicaid	6
All insurers	21

[a]PHE: Public Health Emergency.

XSL•FO
RenderX

At this point, it is unknown whether Medicare will implement permanent changes on the federal level and whether certain states will permanently continue payment parity for telemedicine services. Differences in state and federal laws can impede health care providers' ability to navigate payment, reducing patient health care access [37]. These laws burdened providers with understanding the changing legal landscape of telemedicine and determining whether it is fiscally feasible.

Overall, there is federal coverage parity through Medicare, and all 50 states and Washington, DC, offer coverage parity with some variation in language [34]. While there is uncertainty regarding payment parity continuance through Medicare, the physician fee schedule for 2024 posted for Medicare included payment parity and granularity in types of telemedicine visits [38]. On the state level, payment parity continues to evolve. States look to one another to determine what has worked effectively and what has not in choosing and implementing their parity laws. State and federal legal definitions of telemedicine visits may differ. Still, there is clear support for ensuring that telemedicine is sustained as a health care delivery option in the United States. We see the trend that state laws will eventually conform to those of Medicare, resulting in unified definitions of telemedicine and its reimbursement. Until this happens, providers may continue to view payment parity as a hindrance to implementing telemedicine in their workflow.

Care Across State Borders

During the PHE, providers could deliver health care nationwide with any state license. Unlike the DEA's extension of the flexibility—allowing a provider to have a single waiver to prescribe controlled substances nationwide—lawmakers did not extend the PHE flexibility that allowed providers to care for patients across state lines without a license in each state. States have varied in their regulation of telemedicine across state borders [39]. For example, if a provider is licensed in New York and treats a patient who is in California for college or lives part-time in Florida, they can only see the patient within New York (even if the provider is out of state during the visit) [40]. In the same example, if the New York provider has a patient in New Jersey, that patient would have to drive into New York to be seen over telemedicine. Such requirements have created the telemedicine parking lot phenomenon wherein patients drive to the nearest location within the provider's state and park to be seen via telemedicine to comply with state laws and licensure [41].

Regulatory changes allowing providers to see patients outside their states may increase patient safety and continuity of care [42,43]. A policy initiative, supported by professional organizations, such as the American College of Physicians and the American Medical Association, would ease regulatory compliance and eliminate a significant barrier to telemedicine care delivery across US state borders [13,43-46]. Licensure compacts exist for health professionals outside physicians, that is, nurses, physical therapists, occupational therapists, speech-language pathologists, audiologists, psychologists, and emergency medical professionals, that enable practice across multiple states [47]. Unfortunately, not all compacts act in the same way: while the Interstate Medical Licensure Compact for physicians only helps to streamline the licensing process, the Nurse Licensure Compact allows nurses to hold one license and practice in all participating states [33]. The larger scope of the Nurse Licensure Compact overcomes the additional barriers of time, complexity, and money that providers face. Of course, with each licensure compact, there are rules and regulations at the state level that would require amendment including definitions and functions of noncompete agreements and standards of care [48-51]. Ultimately, licensure compacts help simplify and codify the laws and regulations surrounding practice across state lines. Lawmakers and regulators need to remedy the level of complexity in the state and federal legislation in the face of the patient's needs and the providers' abilities [52]. Since the PHE, more states have joined or are in the process of joining these licensure compacts. For example, the Interstate Medical Licensure Compact currently includes 39 states and Washington, DC, and the Nurse Licensure Compact consists of 41 states [40,53].

Conclusions

With the end of the PHE, US Congress and regulators allowed certain telemedicine policy flexibilities to continue temporarily. At the same time, lawmakers and regulators work to align pre-PHE policies with current provider and patient needs relating to telemedicine. Over 100 pieces of proposed federal legislation currently address aspects of telemedicine, with more proposals from state lawmakers [39]. This viewpoint reviewed how COVID-19–era PHE flexibilities paved the way for telemedicine adoption, summarized the policy challenges that arose, and described how policy leaders are attempting to resolve the challenges.

In the coming months and years, we anticipate changes to privacy and security laws, potential changes to HIPAA, and increasing consideration of how privacy laws function concerning emergent digital health technologies. After the DEA's final ruling regarding prescribing over telemedicine, expected in 2025, we anticipate further state-level changes in laws and regulations to better align with the federal standard. We anticipate that current regulatory trends supporting telemedicine coverage and payment parity will continue and that more states will adopt interstate licensure compacts for their health care professionals. Overall, our assessment of these critical topics reveals that while telemedicine is now an unmistakable part of health care, there is a need to codify the laws and regulations on a federal and state level. Such a move would ensure patient privacy, security, and quality of care and ensure provider support through definitive, clear pathways to comply with telemedicine laws and regulations.

Authors' Contributions

All authors contributed to conceptualization, review, and editing. JI drafted the manuscript.

XSL•FO
RenderX

Conflicts of Interest

BW is a shareholder, and all other authors are employees of Doxy.me Inc, a commercial telemedicine company.

References

1. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-person and telehealth ambulatory contacts and costs in a large US insured cohort before and during the COVID-19 pandemic. JAMA Netw Open 2021;4(3):e212618 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]
2. Telemedicine use among physicians by physician specialty: United States, 2021. 2024. URL: https://www.cdc.gov/nchs/products/databriefs/db493.htm [accessed 2024-05-13]
3. CARES Act: AMA COVID-19 pandemic telehealth fact sheet. American Medical Association. 2020. URL: https://www.ama-assn.org/delivering-care/public-health/cares-act-ama-covid-19-pandemic-telehealth-fact-sheet [accessed 2020-04-27]
4. VanderWerf M, Bernard J, Barta DT, Berg J, Collins T, Dowdy M, et al. Pandemic action plan policy and regulatory summary telehealth policy and regulatory considerations during a pandemic. Telemed e-Health 2022;28(4):457-466 [FREE Full text] [doi: 10.1089/tmj.2021.0216] [Medline: 34265216]
5. Chen PG, Heins SE, Dellva S, RAND Health Care. State medicaid telehealth coverage policy decisions since the COVID-19 public health emergency. 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK605935/?report=reader [accessed 2024-10-28]
6. Drug Enforcement Administration Health and Human Services Department. Second temporary extension of COVID-19 telemedicine flexibilities for prescription of controlled medications [Public Inspection Version]. 2023. URL: https://www.federalregister.gov/d/2023-22406 [accessed 2024-10-17]
7. Telehealth policy changes after the COVID-19 public health emergency. Telehealth. URL: https://telehealth.hhs.gov/providers/telehealth-policy/policy-changes-after-the-covid-19-public-health-emergency [accessed 2024-04-29]
8. Mehrotra A. Telemedicine can change care for the better—with the right rules [Internet]. URL: https://hms.harvard.edu/news/telemedicine-can-change-care-better-right-rules [accessed 2024-04-25]
9. Office for Civil Rights (OCR). HIPAA and telehealth. US Department of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/telehealth/index.html [accessed 2024-04-25]
10. Lin SK, Deitermann AM, Miller CJ, Garg A, Ungar J, Nguyen HP. Evolving teledermatology policy and reimbursement landscape in the United States. JAAD Int 2023;11:200-208 [FREE Full text] [doi: 10.1016/j.jdin.2023.03.006] [Medline: 37138827]
11. Office for Civil Rights (OCR). Resources for mobile health apps developers. US Department of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/health-apps/index.html [accessed 2023-10-06]
12. Davis J. Fed's updated mobile health app tool aims to reduce HIPAA compliance mistakes. URL: https://www.scmagazine.com/analysis/feds-updated-mobile-health-app-tool-aims-to-reduce-hipaa-compliance-mistakes [accessed 2023-10-06]
13. Double secret protection: bridging federal and state law to protect privacy rights for telemental and mobile health users. Duke Law Journal. 2018. URL: https://dlj.law.duke.edu/article/double-secret-protection-sherman-vol67-iss5/ [accessed 2023-05-23]
14. Karsten J, Roberts J, Lee NT. Removing regulatory barriers to telehealth before and after COVID-19. Brookings. 2020. URL: https://www.brookings.edu/articles/removing-regulatory-barriers-to-telehealth-before-and-after-covid-19/ [accessed 2023-10-06]
15. Consolidated appropriations act. 2023. URL: https://www.congress.gov/bill/117th-congress/house-bill/2617/text [accessed 2022-12-29]
16. Alder S. HIPAA updates and HIPAA changes in 2024. URL: https://www.hipaajournal.com/hipaa-updates-hipaa-changes/ [accessed 2024-04-09]
17. Protecting Washingtonians' personal health data and privacy. URL: https://www.atg.wa.gov/protecting-washingtonians-personal-health-data-and-privacy [accessed 2023-09-23]
18. Washington's My Health My Data Act—are you prepared?. URL: https://www.venable.com/insights/publications/2024/02/washingtons-my-health-my-data-act-are-you [accessed 2024-04-29]
19. 88(R). SB 2064—introduced version—bill text. URL: https://capitol.texas.gov/tlodocs/88R/billtext/html/SB02064I.htm [accessed 2023-10-10]
20. Legislation—HB0995. URL: https://mgaleg.maryland.gov/mgawebsite/Legislation/Details/hb0995?ys=2023RS [accessed 2023-10-10]
21. Legislation—SB0790. URL: https://mgaleg.maryland.gov/mgawebsite/Legislation/Details/SB0790 [accessed 2023-10-10]
22. SB370. URL: https://www.leg.state.nv.us/App/NELIS/REL/82nd2023/Bill/10323/Overview [accessed 2024-04-09]
23. DEA announces proposed rules for permanent telemedicine flexibilities. DEA. URL: https://www.dea.gov/press-releases/2023/02/24/dea-announces-proposed-rules-permanent-telemedicine-flexibilities [accessed 2023-10-10]
24. DEA extends telemedicine flexibilities for prescribing of controlled medications: second time is the charm. URL: https://www.foley.com/insights/publications/2023/10/dea-extends-telemedicine-prescribing-controlled/ [accessed 2024-04-10]

XSL•FO
RenderX

25. Kannarkat JT, Kannarkat JT, Torous J. Rebalancing controlled substance regulations in telemedicine. JAMA Health Forum 2023;4(10):e233251 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.3251] [Medline: 37862032]
26. Mattocks KM, Moore DT, Wischik DL, Lazar CM, Rosen MI. Understanding opportunities and challenges with telemedicine-delivered buprenorphine during the COVID-19 pandemic. J Subst Abuse Treat 2022;139:108777 [FREE Full text] [doi: 10.1016/j.jsat.2022.108777] [Medline: 35346533]
27. Burke B, Miller E, Clear B, Weiner SG. A qualitative study to determine perspectives of clinicians providing telehealth opioid use disorder treatment. Drug Alcohol Depend 2024;256:111118. [doi: 10.1016/j.drugalcdep.2024.111118] [Medline: 38367534]
28. Hammerslag LR, Mack A, Chandler RK, Fanucchi LC, Feaster DJ, LaRochelle MR, et al. Telemedicine buprenorphine initiation and retention in opioid use disorder treatment for Medicaid enrollees. JAMA Netw Open 2023;6(10):e2336914 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36914] [Medline: 37851446]
29. Krawczyk N, Rivera B, King C, Dooling BCE. Pandemic telehealth flexibilities for buprenorphine treatment: a synthesis of evidence and policy implications for expanding opioid use disorder care in the United States. Health Aff Scholar 2023;1(1):qxad013 [FREE Full text] [doi: 10.1093/haschl/qxad013] [Medline: 38145115]
30. Chan B, Bougatsos C, Priest KC, McCarty D, Grusing S, Chou R. Opioid treatment programs, telemedicine and COVID-19: a scoping review. Subst Abus 2022;43(1):539-546. [doi: 10.1080/08897077.2021.1967836] [Medline: 34520702]
31. Pham H, Lin C, Zhu Y, Clingan SE, Lin LA, Mooney LJ, et al. Telemedicine-delivered treatment for substance use disorder: a scoping review. J Telemed Telecare 2023:1357633X231190945. [doi: 10.1177/1357633X231190945] [Medline: 37537907]
32. Abrar F. Some medications will now require an in-person doctor visit to obtain. Here's why. 2024. URL: https://www.13wmaz.com/article/news/health/medical-board-returns-to-pre-covid-protocols-for-certain-prescriptions-in-may-dr-uwa-iguobadia/93-00c8674a-cf29-4be5-85f3-aa13ac814710 [accessed 2024-04-10]
33. Garber K, Chike-Harris K, Vetter MJ, Kobeissi M, Heidesch T, Arends R, et al. Telehealth policy and the advanced practice nurse. J Nurse Pract 2023;19(7):104655 [FREE Full text] [doi: 10.1016/j.nurpra.2023.104655] [Medline: 37235075]
34. Parity. CCHP. URL: https://www.cchpca.org/topic/parity/ [accessed 2024-04-10]
35. Expanding services during COVID-19. Nashp. URL: https://nashp.org/expanding-services-during-covid-19/ [accessed 2024-04-10]
36. Ellimoottil C. Understanding the case for telehealth payment parity. Health Affairs Forefront. URL: https://www.healthaffairs.org/content/forefront/understanding-case-telehealth-payment-parity [accessed 2024-04-10]
37. Khera N, Knoedler M, Meier SK, TerKonda S, Williams RD, Wittich CM, et al. Payment and coverage parity for virtual care and in-person care: how do we get there? Telemed Rep 2023;4(1):100-108 [FREE Full text] [doi: 10.1089/tmr.2023.0014] [Medline: 37283856]
38. 2024 physician fee schedule final rule released. American Psychological Association. 2023. URL: https://www.apaservices.org/practice/reimbursement/government/2024-physician-fee-schedule-final-rule [accessed 2024-04-10]
39. Federal telehealth laws. CCHP. URL: https://www.cchpca.org/federal/pending-legislation/ [accessed 2023-10-04]
40. Physician license. Interstate Medical Licensure Compact. 2016. URL: https://www.imlcc.org/ [accessed 2024-04-29]
41. Khan A. Telehealth laws need to be updated for a post-COVID health system. URL: https://chlpi.org/news-and-events/news-and-commentary/health-law-and-policy/telehealth-laws-need-to-be-updated-for-a-post-covid-health-system/ [accessed 2024-04-10]
42. Saliba V, Legido-Quigley H, Hallik R, Aaviksoo A, Car J, McKee M. Telemedicine across borders: a systematic review of factors that hinder or support implementation. Int J Med Inform 2012;81(12):793-809. [doi: 10.1016/j.ijmedinf.2012.08.003] [Medline: 22975018]
43. Willis A. Utilizing telehealth to practice medicine across state lines: the enforceability of physician non-compete agreements and non-solicitation clauses notes. Hofstra Labor Employment Law J 2022;39(1):250 [FREE Full text]
44. Elliott VL. Telehealth and telemedicine: description and issues. Washington, DC: Library of Congress; 2016. URL: https://apps.dtic.mil/sti/citations/AD1172117 [accessed 2023-08-22]
45. Hoffman LC. Telehealth, children, and pediatrics: should the doctor make house calls again, digitally articles and surveys. Nova Law Rev 2019;43(3):321-352 [FREE Full text]
46. Andriola M. Telemedicine and legal disruption. Health Law Policy 2019;13(2):30 [FREE Full text]
47. Licensure compacts. URL: https://telehealth.hhs.gov/licensure/licensure-compacts [accessed 2024-04-10]
48. Lhospital DE. Expansion Of telemedicine during COVID-19 and the issue of non-compete agreements. Virginia State Bar. 2021. URL: https://virginialawyer.vsb.org/publication/?i=690953&article_id=3863379&view=articleBrowser [accessed 2023-05-22]
49. Wolf TD. Telemedicine and malpractice: creating uniformity at the national level. William Mary Law Rev 2020;61(5):1505 [FREE Full text]
50. Chang H. Evaluation framework for telemedicine using the logical framework approach and a fishbone diagram. Healthcare Informatics Res 2015;21(4):230-238 [FREE Full text] [doi: 10.4258/hir.2015.21.4.230] [Medline: 26618028]

XSL•FO
RenderX

51. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]
52. Wang CJ, Liu TT, Car J, Zuckerman B. Design, adoption, implementation, scalability, and sustainability of telehealth programs. Pediatr Clin North Am 2020;67(4):675-682. [doi: 10.1016/j.pcl.2020.04.011] [Medline: 32650866]
53. Home. Nursecompact. URL: https://www.nursecompact.com/ [accessed 2024-04-10]

Abbreviations

DEA: Drug Enforcement Administration
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Statistical Classification of Diseases, Tenth Revision
PHE: Public Health Emergency
PHI: personal health information

Edited by A Mavragani; submitted 10.10.23; peer-reviewed by L Celine, K Drude; comments to author 29.02.24; revised version received 29.05.24; accepted 27.09.24; published 23.01.25.

Please cite as:
Ivanova J, Cummins MR, Ong T, Soni H, Barrera J, Wilczewski H, Welch B, Bunnell B
Regulation and Compliance in Telemedicine: Viewpoint
J Med Internet Res 2025;27:e53558
URL: https://www.jmir.org/2025/1/e53558
doi:10.2196/53558
PMID:39847413

©Julia Ivanova, Mollie R Cummins, Triton Ong, Hiral Soni, Janelle Barrera, Hattie Wilczewski, Brandon Welch, Brian Bunnell. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

From Stories to Solutions: A Research Cycle Framework for Enhancing Trustworthiness in Studies of Online Patient Narratives

Klay Lamprell[1], PhD; Diana Fajardo Pulido[2], BNurse, MIPH, MHM; Gaston Arnolda[1], PhD; Bróna Nic Giolla Easpaig[1,3], PhD; Yvonne Tran[4], PhD; Jeffrey Braithwaite[1]

[1]Macquarie University, Australian Institute of Health Innovations, Sydney, Australia

[2]Portal Pontificia Universidad Javeriana, Bogotá, Colombia

[3]School of Nursing, Charles Darwin University, Darwin, Australia

[4]Macquarie University Hearing, Macquarie University, Sydney, Australia

Corresponding Author:
Klay Lamprell, PhD
Macquarie University
Australian Institute of Health Innovations
Level 6, 75 Talavera Rd
North Ryde
Sydney, 2109
Australia
Phone: 61 02 9850 2455
Email: klay.lamprell@mq.edu.au

Abstract

Personal accounts of health care experiences posted to online platforms are a rich source of patient-reported data. Noninteractive narratives, published as one-off accounts, often describe an entire patient journey in one story, featuring transitions through health care settings from prediagnosis to outcome. These coherent perspectives may not be visible or captured in fragmented social media posts or episodic blogs. For researchers seeking patient-reported accounts of care quality across a clinical trajectory, noninteractive online narratives can be an invaluable, easily accessible resource. The challenges and limitations of research utilizing these data may differ from working with interactive patient narratives and include source transparency and credibility, limited or no information about authors, and ambiguity about the health care context and time frames. We outline a framework for addressing these issues in 5 key phases of the research cycle: the conceptual-theoretical approach; locating sources of data; accessing and collecting the data; quantitative profiling to establish demographics, health care events, and time frames; and qualitative analysis utilizing multiple modes of narrative inquiry. We explain the use of this framework in a real-world example: our study of 273 online patient narratives describing health care trajectories in early-onset bowel cancer.

International Registered Report Identifier (IRRID): RR2-10.2196/25056

(J Med Internet Res 2025;27:e58310) doi:10.2196/58310

KEYWORDS

online research; exploratory study; patient experience; patient narratives; narrative analysis; mixed methods; young-onset colorectal cancer; cancer; oncology; internal medicine

Introduction

Online patient narratives have been categorized as interactive or noninteractive [1,2], each offering distinct advantages and challenges for health care research. Interactive narratives, typically shared through social media, blogs, and forums, allow patients to update their stories, respond to feedback, and report evolving experiences. Noninteractive or static narratives, published to websites as one-off reports, typically encapsulate the entire patient journey over time. This coherent, patient-framed perspective on the progression from prediagnosis to outcome may not be visible or captured in the fragmented or episodic formats of interactive narratives.

For researchers investigating patient-framed perceptions of care across a clinical trajectory, noninteractive online narratives, which can also be referred to as online pathographies [3,4], offer an invaluable resource. The holistic narrative structure allows the representation of both a psychological transition and a

XSL•FO
RenderX

physical movement through various phases of biomedical intervention and health care settings. They can expose health service delays and continuity gaps that only patients can observe as sole witnesses to the entirety of their experiences. Importantly, the global availability of online patient narratives allows researchers ease of access and, potentially, insights into the diversity of health system experiences.

The value of these pathographies for researchers must be weighed against unique challenges to validity and trustworthiness that share common elements with, but also extend beyond, typical issues in qualitative research, and potentially even beyond challenges with the use of interactive narratives. Noninteractive online narratives present risks such as source credibility, limited or no information about authors, and uncertainty about the health care context and time frames. These aspects of the data may present constraints on analytic interpretation.

We navigated these challenges in an exploratory study to understand the causes of delays in the diagnosis of early-onset bowel cancer, defined as a diagnosis in patients younger than the age of 50 years [5]. Our study drew on narratives published on prominent bowel cancer support websites in the United Kingdom, Australia, and New Zealand, published under banners such as “Real life stories” and “Your Story” [6,7]. The findings of our study, reported elsewhere [6,7], highlighted a critical need to lower the threshold age for public bowel cancer screening programs and revise clinical guidelines for diagnostic pathways.

To help structure the research process in ways that enhance reliability, we developed a framework featuring 5 key phases throughout the research cycle requiring heightened attention to issues of validity and reliability (see Figure 1). The framework developed by Zoolnoori et al [8], for use with interactive online patient narratives, focuses on 4 of the main phases of research: data collection, data preparation, content analysis, and interpretation of the results. We found that a different set of focal points were needed to address the unique challenges of research investigating noninteractive online narratives. Below, we describe each of the 5 focal points in our framework through the lens of our work with noninteractive, early-onset bowel cancer narratives.

Figure 1. A framework of five key focal points in the research cycle for addressing the credibility and enhancing the trustworthiness of research involving noninteractive, online patient narratives.

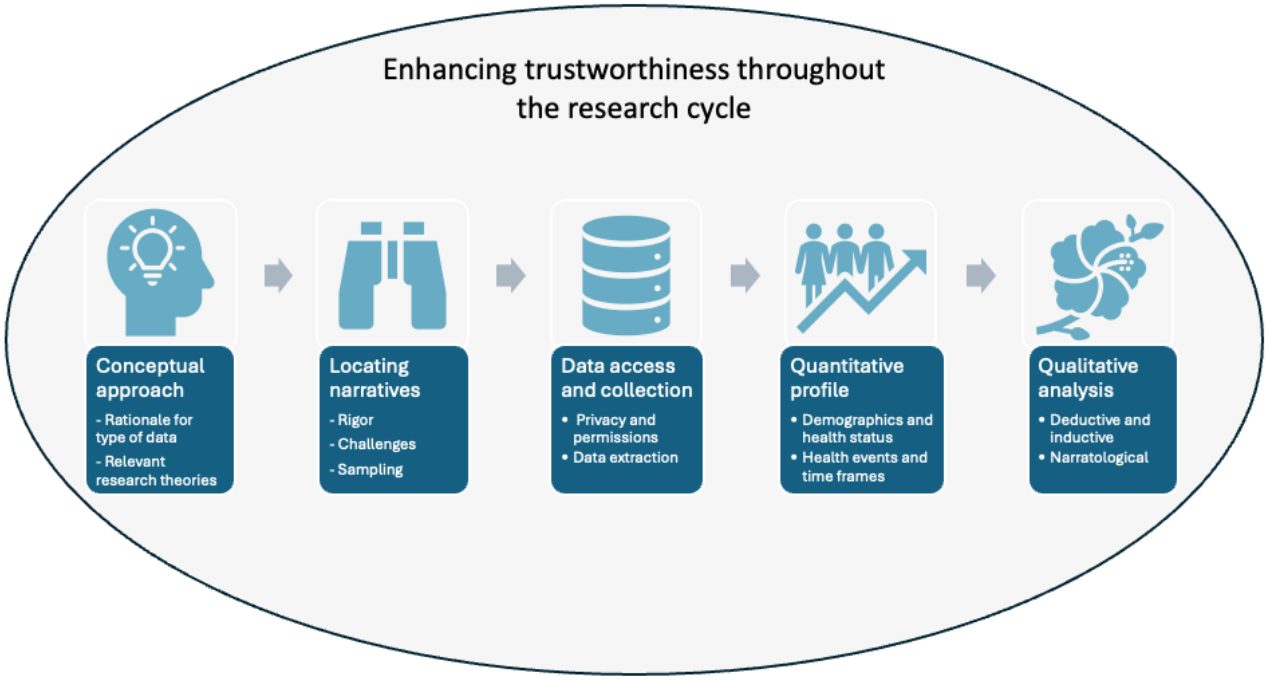

Ethical Considerations

Ethical and scientific approval was granted by Macquarie University Human Research Ethics Committee (reference 52020666115757).

Conceptual-Theoretical Approach

The framework begins in the conceptualizing phase, establishing a rationale for the use of online pathographies and highlighting a detailed theoretical framework for the research. **As** exploratory research, the rationale for utilizing online patient pathographies was to determine the relevance of progressing to purposively sampled studies [9]. The rationale was further supported by literature indicating that patients younger than the age of 50 years are drawn to, and are adept at using, the internet for information seeking and sharing [10].

Narrative inquiry is an apt overarching methodology for research that leverages online narratives; it focuses on exploring and interpreting phenomena through stories of lived experience [11] and features flexibility in blending qualitative and quantitative elements to support trustworthy findings [12,13]. Key theories validating singular narratives as evidence of phenomena [14] include constructivism (knowledge through experiences),

XSL•FO
RenderX

phenomenology (shared experiences), hermeneutics (text interpretation), and narratology (narrative structures).

Locating Authentic Narratives

We included a focal point on the source of the pathographies because the **credibility** and **authenticity** of noninteractive narratives are closely tied to the reliability of the source. It is crucial to establish some means of locating pathographies that can be assumed to be authentic. We looked to institutional oversight as a way to verify authenticity. We followed the Internet Mediated Research (IMR) search strategy outlined by Obrien and Clark [15], using comprehensive search terms and multiple platforms, as well as systematically applying inclusion and exclusion criteria to locate relevant pathographies published on the websites of well-established institutions. We excluded pathographies from health care organization websites that presented exclusively positive views for marketing purposes.

Demographics and clinical information about the authors were typically embedded within the narratives and not necessarily explicitly stated; we had to review all pathographies identified on validated sites line by line to confirm they met the criteria for early-onset colorectal cancer.

Data Access and Collection

We highlight this phase in the research cycle because the unsolicited, often anonymous, and noninteractive nature of the data necessitated careful negotiation of consent. We contacted each organization to ensure that the conditions for consent were met and to obtain explicit approval for empirical investigation and dissemination of our findings. Permission from individual authors was not required, aligning with the principles of "unobtrusive" IMR [16,17] and the concept of "human-generated data" [18], which individuals willingly produce as they engage in various online and offline activities. This approach met the ethical guidelines of the Australian National Statement on Ethical Conduct in Human Research 2007 (Updated 2018) [19] and the British Psychological Society's principles for IMR [20].

Where technically possible, we used the website's export options to extract narrative text from the online platform and download it into an analysis software program, NVivo 12 (version 12; QSR International Pty Ltd). Alternatively, we used a cut-and-paste function, creating a Microsoft Word document for each narrative, which was then imported into NVivo.

Quantitative Profiling

We specify that a quantitative approach is needed to resolve the likely anonymity of authors. As noted, authors' demographics and health status characteristics were not always evident in the unsolicited data, which presented the risk of over- and underrepresentation of subpopulations and particular types of diagnostic experiences. To ensure the transparency of any data bias, we used NVivo to quantify the demographic and health status information in the narratives and cataloged the *type*, *frequency*, and *time frames* of diagnostic events described in the texts, including first event of seeking help for symptoms, types of tests and screening events, and referrals to specialists.

In extracting demographic and health status data from narratives and cataloging diagnostic events by type, frequency, and time frame, we identified a gender disparity of authorship, with over 70% of narratives written by people identifying as female, despite approximately equal gender diagnoses of early-onset colorectal cancer [21]. Additionally, the quantitative profile identified that most narratives were written by people aged younger than 40 years, although there is a higher incidence of early-onset colorectal cancer in the 40-49 years age group. This finding highlights the relevance of online methods when investigating younger populations but acknowledges the potential for bias when generalizing across sets of unsolicited data. In our reporting of our findings, we openly discussed these biases and hypothesized others, such as potential underrepresentation of culturally and linguistically diverse and special needs populations.

Qualitative Analysis

We assert that manual qualitative analysis of pathographies continues to offer the greatest potential for sensitivity to textual nuances and tonal idiosyncrasies indicating facilitators and barriers to quality care over time [8]. Pathographies embed time in 3 key narrative layers: the subjective account of chronological time, detailing the sequence of events from symptom onset to treatment; the passing of time, represented in evolving observations of illness and health care experiences; and subtle references to historic time, which places the patient's experience within broader societal trends or health care issues [22]. Complex, highly customized programming would be required to automate the extraction of time-related meaning from these texts.

Unsolicited online narratives, unlike interview data, contain large amounts of content that may be extraneous to the core interests of the research. Integrating deductive and inductive methods analysis [23] allowed us to efficiently extract relevant data while allowing for unexpected insights. In NVivo, the data were first categorized into a basic deductive framework [23] comprising the 2 key domains of care assessed in patient-reported experience measures: functional (clinical care and the practical delivery of services) and relational (relationships with clinicians and health care service providers) experiences of care. We then inductively sought emergent themes in the data categorized into each of these domains, following the conventions established by Braun et al [24].

Finally, we propose a narratological component in the qualitative analysis because personal experience is conveyed not only through story content but also through the choices made in storytelling [25,26]. Our narratological component examined the storytelling structure used by the authors. Typically, this structure followed Western narrative conventions: a prologue depicting life before the first symptom, the main story describing events leading up to diagnosis, and an epilogue expressing closure to the story [27]. We observed that epilogues served as spaces for authors to reflect on lessons learned and the broader implications of their experiences. Our epilogue-only analysis identified a substantial set of patient-reported views on the need for self-advocacy and changes to the views of clinicians on colorectal cancer among young people. These findings were

XSL•FO
RenderX

published in a second paper arising from the study [7] and were used in a campaign by Bowel Cancer Australia to lower the screening age for bowel cancer from 50 to 45 years.

Conclusions

Noninteractive online narratives can provide a singular, comprehensive perspective on patient journeys, capturing transitions across health care settings and services. As we found in our study of 273 noninteractive online narratives, this perspective offers researchers unique insights into systemic delays and gaps in continuity. In our study, we demonstrated the value of a framework tailored to address the specific challenges and limitations of noninteractive data throughout the research cycle. The 5 focal points of the framework offer a model for researchers navigating the complexities of these data. In future work, utilizing this framework can ensure the integrity and value of exploratory studies that seek to identify emerging issues exposed in longitudinal patient accounts published online. The framework may provide collaborative stakeholders—researchers, health care providers, policy makers, and patient advocates—with a structured means of understanding how credible, patient-centered insights be drawn from these data. Additionally, the 5 focal points are adaptable to new digital environments for patient narratives and advanced analytical technologies. In these ways, the framework reinforces the potential for online patient journey narratives to inform health care research and affirms that patients' stories can guide solutions to health care problems.

Acknowledgments

The authors thank the following advocacy organizations for patients with colorectal cancer: Bowel Cancer UK, Bowel Cancer Australia, and Bowel Cancer NZ, for facilitating our access to patients' stories and to the patients who wrote about their experiences. This work was supported in part by a National Health and Medical Research Council of Australia (NHMRC grant 1135048—the Centre of Research Excellence in Implementation Science in Oncology), which is administered by the Australian Institute of Health Innovation, Macquarie University, Sydney. JB reports multiple grants funded by the National Health and Medical Research Council and other providers.

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KL wrote the first draft of the manuscript. DFP, GA, BNGE, YT, and JB contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
2. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun 18;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
3. O'Brien MR, Clark D. Use of unsolicited first-person written illness narratives in research: systematic review. J Adv Nurs 2010 Aug 02;66(8):1671-1682. [doi: 10.1111/j.1365-2648.2010.05349.x] [Medline: 20557382]
4. Hawkins AH. Pathography: patient narratives of illness. West J Med 1999 Aug;171(2):127-129 [FREE Full text] [Medline: 10510661]
5. Neugut AI, Lebwohl B. Early onset colorectal cancer: a hypothesis. Oncologist 2023 Dec 11;28(12):1015-1016 [FREE Full text] [doi: 10.1093/oncolo/oyad274] [Medline: 37758481]
6. Lamprell K, Pulido DF, Arnolda G, Easpaig BNG, Tran Y, Owais SS, et al. People with early-onset colorectal cancer describe primary care barriers to timely diagnosis: a mixed-methods study of web-based patient reports in the United Kingdom, Australia and New Zealand. BMC Prim Care 2023 Jan 14;24(1):12 [FREE Full text] [doi: 10.1186/s12875-023-01967-0] [Medline: 36641420]
7. Lamprell K, Fajardo-Pulido D, Arnolda G, Owais SS, Nic Giolla Easpaig B, Tran Y, et al. Things I need you to know: a qualitative analysis of advice-giving statements in early-onset colorectal cancer patients' personal accounts published online. BMJ Open 2023 Mar 22;13(3):e068073 [FREE Full text] [doi: 10.1136/bmjopen-2022-068073] [Medline: 36948549]
8. Zolnoori M, Balls-Berry JE, Brockman TA, Patten CA, Huang M, Yao L. A systematic framework for analyzing patient-generated narrative data: protocol for a content analysis. JMIR Res Protoc 2019 Aug 26;8(8):13914 [FREE Full text] [doi: 10.2196/13914] [Medline: 31452524]

XSL•FO
RenderX

9. Hallingberg B, Turley R, Segrott J, Wight D, Craig P, Moore L, et al. Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance. Pilot Feasibility Stud 2018 May 28;4(1):104 [FREE Full text] [doi: 10.1186/s40814-018-0290-8] [Medline: 29854417]
10. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
11. Nasheeda A, Abdullah HB, Krauss SE, Ahmed NB. Transforming transcripts into stories: a multimethod approach to narrative analysis. Int J Qual Methods 2019 Jun 13;18. [doi: 10.1177/1609406919856797]
12. Eichsteller M. There is more than one way - a study of mixed analytical methods in biographical narrative research. In: Biographical Research. London, United Kingdom: Routledge; 2022:103-118.
13. Lamprell K, Fajardo Pulido D, Tran Y, Nic Giolla Easpaig B, Liauw W, Arnolda G, et al. Personal accounts of young-onset colorectal cancer organized as patient-reported data: protocol for a mixed methods study. JMIR Res Protoc 2021 Feb 26;10(2):e25056 [FREE Full text] [doi: 10.2196/25056] [Medline: 33635274]
14. Nigar N. Hermeneutic phenomenological narrative enquiry: a qualitative study design. Theory and Practice in Language Studies 2020 Jan;10(1):10-18. [doi: 10.17507/tpls.1001.02]
15. O'Brien MR, Clark D. Unsolicited written narratives as a methodological genre in terminal illness: challenges and limitations. Qual Health Res 2012 Feb 29;22(2):274-284. [doi: 10.1177/1049732311420737] [Medline: 21876209]
16. Hewson C. Ethics issues in digital methods research. In: Snee H, Hine C, Morey Y, Roberts S, Watson H, editors. Digital Methods for Social Science. London, United Kingdom: Palgrave Macmillan; 2016:206-221.
17. Burles MC, Bally JMG. Ethical, practical, and methodological considerations for unobtrusive qualitative research about personal narratives shared on the internet. Int J Qual Methods 2018 Jul 17;17(1). [doi: 10.1177/1609406918788203]
18. Islam MS, Hasan MM, Wang X, Germack HD, Noor-E-Alam M. A systematic review on healthcare analytics: application and theoretical perspective of data mining. Healthcare (Basel) 2018 May 23;6(2):54 [FREE Full text] [doi: 10.3390/healthcare6020054] [Medline: 29882866]
19. National Statement on Ethical Conduct in Human Research (2007) - updated 2018. National Health and Medical Research Council. 2018. URL: https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018 [accessed 2024-12-03]
20. Ethics guidelines for internet-mediated research. British Psychological Society. 2021 Jun 7. URL: https://www.bps.org.uk/guideline/ethics-guidelines-internet-mediated-research [accessed 2024-12-03]
21. Saad El Din K, Loree JM, Sayre EC, Gill S, Brown CJ, Dau H, et al. Trends in the epidemiology of young-onset colorectal cancer: a worldwide systematic review. BMC Cancer 2020 Apr 06;20(1):288 [FREE Full text] [doi: 10.1186/s12885-020-06766-9] [Medline: 32252672]
22. Lamprell K, Braithwaite J. Reading between the lines: a five-point narrative approach to online accounts of illness. J Med Humanit 2019 Dec 15;40(4):569-590 [FREE Full text] [doi: 10.1007/s10912-019-09553-1] [Medline: 30982939]
23. Fereday J, Muir-Cochrane E. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. Int J Qual Methods 2006 Mar 01;5(1):80-92. [doi: 10.1177/160940690600500107]
24. Braun V, Clarke V, Hayfield N, Terry G. Thematic analysis. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer; Jan 13, 2019.
25. Charon R, Wyer P, NEBM Working Group. Narrative evidence based medicine. Lancet 2008 Jan 26;371(9609):296-297. [doi: 10.1016/s0140-6736(08)60156-7] [Medline: 18300381]
26. Lamprell K, Rapport F, Braithwaite J. Narrativizing cancer patients' longitudinal experiences of care: qualitative inquiry into lived and online melanoma stories. In: Transforming Healthcare with Qualitative Research. London, United Kingdom: Routledge; 2020.
27. Lamprell K, Braithwaite J. Patients as story-tellers of healthcare journeys. Med Humanit 2016 Sep 04;42(3):207-209. [doi: 10.1136/medhum-2016-010885] [Medline: 27044872]

Abbreviations

IMR: Internet Mediated Research

XSL•FO
RenderX

Edited by A Mavragani; submitted 12.03.24; peer-reviewed by JN Wang, N Patel; comments to author 26.03.24; revised version received 04.08.24; accepted 13.11.24; published 23.01.25.

Please cite as:
Lamprell K, Pulido DF, Arnolda G, Easpaig BNG, Tran Y, Braithwaite J
From Stories to Solutions: A Research Cycle Framework for Enhancing Trustworthiness in Studies of Online Patient Narratives
J Med Internet Res 2025;27:e58310
URL: https://www.jmir.org/2025/1/e58310
doi:10.2196/58310
PMID:39847425

©Klay Lamprell, Diana Fajardo Pulido, Gaston Arnolda, Bróna Nic Giolla Easpaig, Yvonne Tran, Jeffrey Braithwaite. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Designing Health Recommender Systems to Promote Health Equity: A Socioecological Perspective

Caroline A Figueroa[1,2], MD, PhD; Helma Torkamaan[1], PhD; Ananya Bhattacharjee[3], BSc; Hanna Hauptmann[4], PhD; Kathleen W Guan[1], MRES; Gayane Sedrakyan[5], PhD

[1]Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands

[2]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[3]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[4]Utrecht University, Utrecht, Netherlands

[5]Department High-Tech Business and Entrepreneurship Section, Industrial Engineering and Business Information Systems, University of Twente, Enschede, Overijssel, Netherlands

Corresponding Author:
Caroline A Figueroa, MD, PhD
Faculty of Technology, Policy and Management
Delft University of Technology
Jaffalaan 5
Delft, 2628
Netherlands
Phone: 31 621378688
Email: c.figueroa@tudelft.nl

Abstract

Health recommender systems (HRS) have the capability to improve human-centered care and prevention by personalizing content, such as health interventions or health information. HRS, an emerging and developing field, can play a unique role in the digital health field as they can offer relevant recommendations, not only based on what users themselves prefer and may be receptive to, but also using data about wider spheres of influence over human behavior, including peers, families, communities, and societies. We identify and discuss how HRS could play a unique role in decreasing health inequities. We use the socioecological model, which provides representations of how multiple, nested levels of influence (eg, community, institutional, and policy factors) interact to shape individual health. This perspective helps illustrate how HRS could address not just individual health factors but also the structural barriers—such as access to health care, social support, and access to healthy food—that shape health outcomes at various levels. Based on this analysis, we then discuss the challenges and future research priorities. We find that despite the potential for targeting more complex systemic challenges to obtaining good health, current HRS are still focused on individual health behaviors, often do not integrate the lived experiences of users in the design, and have had limited reach and effectiveness for individuals from low socioeconomic status and racial or ethnic minoritized backgrounds. In this viewpoint, we argue that a new design paradigm is necessary in which HRS focus on incorporating structural barriers to good health in addition to user preferences. HRS should be designed with an emphasis on health systems, which also includes incorporating decolonial perspectives of well-being that challenge prevailing medical models. Furthermore, potential lies in evaluating the health equity effects of HRS and leveraging collected data to influence policy. With changes in practices and with an intentional equity focus, HRS could play a crucial role in health promotion and decreasing health inequities.

(J Med Internet Res 2025;27:e60138) doi:10.2196/60138

KEYWORDS

digital health; health promotion; health recommender systems; artificial intelligence; health equity; AI; digital devices; socioecological; health inequities; health behavior; health behaviors; patient centric; digital health intervention

XSL•FO
RenderX

Introduction

Overview

Health recommender systems (HRS), information retrieval systems that often use artificial intelligence, are an increasingly popular method to provide patient-centric personalized health services [1-3]. HRS predict the relevance of recommendations (eg, for exercise, healthy foods, and mental health exercises) for a given user profile [4]. HRS base recommendations on what the user might like, typically using various sources of data such as personal (eg, age, gender, and socioeconomic status [SES]), health (eg, medical history and health surveys), contextual (eg, time, weather, and location), and interaction data (eg, information searched, liked, or rated) [1,5]. HRS can minimize the burden of delivering recommendations by only doing so when there is a real benefit (eg, when an individual is motivated and capable of following a recommendation) and can target multiple related behaviors at once (eg, diet, physical activity, and mental health). HRS have shown growth since 2012, particularly in the fields of lifestyle, such as personalized healthy nutrition recommendations, and personalized health promotion, such as recommending treatments of clinical diagnoses [2]. With increasing interest in using artificial intelligence to improve health promotion and prevention, and more than 50% (1 in 2 European Union citizens, aged 16-74 years) of individuals using the internet to find health information [6], the use of HRS, although still an emerging area, is expected to grow in digital health [7,8].

The Need for a Health Equity Lens

However, digital health interventions, including HRS, are often not developed to target health inequities, "systematic differences in the health of groups and communities occupying unequal positions in society that are avoidable and unjust" [9]. For instance, HRS have scarcely been tested in populations that are marginalized, such as those belonging to low SES and racial and ethnically minoritized individuals [10]. This is a widespread issue, including in high-resource settings, with evidence from European countries and the United States showing that White individuals and those with higher education levels, income, and English proficiency tend to have higher access to and use digital health technologies more often, instead of the groups who would benefit from these solutions most [11]. There is a lack of attention to issues of equity and inclusion, and a tendency to develop from a technology-centered (instead of a human-centered) perspective in the HRS field [2].

The Socioecological Model to Illustrate the Health Equity Potential of HRS

One way of working toward a perspective with an equity lens is the use of the socioecological model—a conceptual framework depicting spheres of influence over human behavior. This model is widely used in public health for scaling health promotion and implementation initiatives, to enable more effective public health interventions for complex social issues [12]. It has also been adapted for digital health equity [10,13]. It provides representations of how multiple, nested levels of influence (eg, community, institutional, and policy factors) interact in dynamic ways to shape individual health [14]. The socioecological model is based on a systems view of health—individual health is impacted by interactions with immediate peers and surroundings as well as broader institutional and societal norms [15]. It helps illustrate how HRS could address not just individual health needs but also the underlying conditions—such as access to health care, social support, and healthy lifestyles—that shape health outcomes. As far as we are aware, this framework has not been applied to HRS. Although some of these arguments can be applied to all digital health interventions, HRS have a unique role in the field because of their data-driven personalization and capacity to be strongly personalized due to profile- and context-crucial information.

Here we adapt the model to examine the unique potential of HRS to target and influence all its levels and thereby have great potential to promote health equity across society. We explain and illustrate how HRS can contribute to health equity by collecting data and targeting health at the systems level, integrating the context of individuals, families or peers, communities, and societies. We provide selected examples of previous work that could contribute to different layers of the framework and discuss challenges. Based on our analysis, we provide future research opportunities for HRS to target multiple levels of influence.

The novelty of this perspective is the intersection of recommender systems and public health, in particular health equity. HRS are still developing but starting to become integrated into medicine and public health [2]. There have been discussions about ethical considerations of recommender systems [16], but their potential on healthy equity, remains underdiscussed. With growing societal health inequities, and the World Health Organization's recent call for equitable participation in a digital world [17], investing in equitable development of this new technology is essential now. As the field moves forward, we argue that an explicit health equity lens is needed from the start, not as an afterthought.

Overview

The paper is structured as follows. First, we illustrate the potential of HRS for decreasing health inequities based on the socioecological model (individual, interpersonal, community, and society; in "The potential of HRS using the socioecological model" section). Through an exploratory literature search, we provide selected examples of how HRS can contribute to health equity, by situating these examples into the different layers of the framework. We then investigate challenges and provide future research priorities for the HRS field based on the gaps we identify ("Recommendations for Research Priorities" section). Alongside Figure 1 [18], Table 1 presents implications of relevant factors that encompass each level and main actions for research priorities. We then provide a general conclusion.

XSL•FO
RenderX

Figure 1. How health recommender systems can contribute to decreasing health inequities (illustrated within the socioecological model). Individual health is conceptualized as shaped by multiple levels, illustrated through concentric circles.

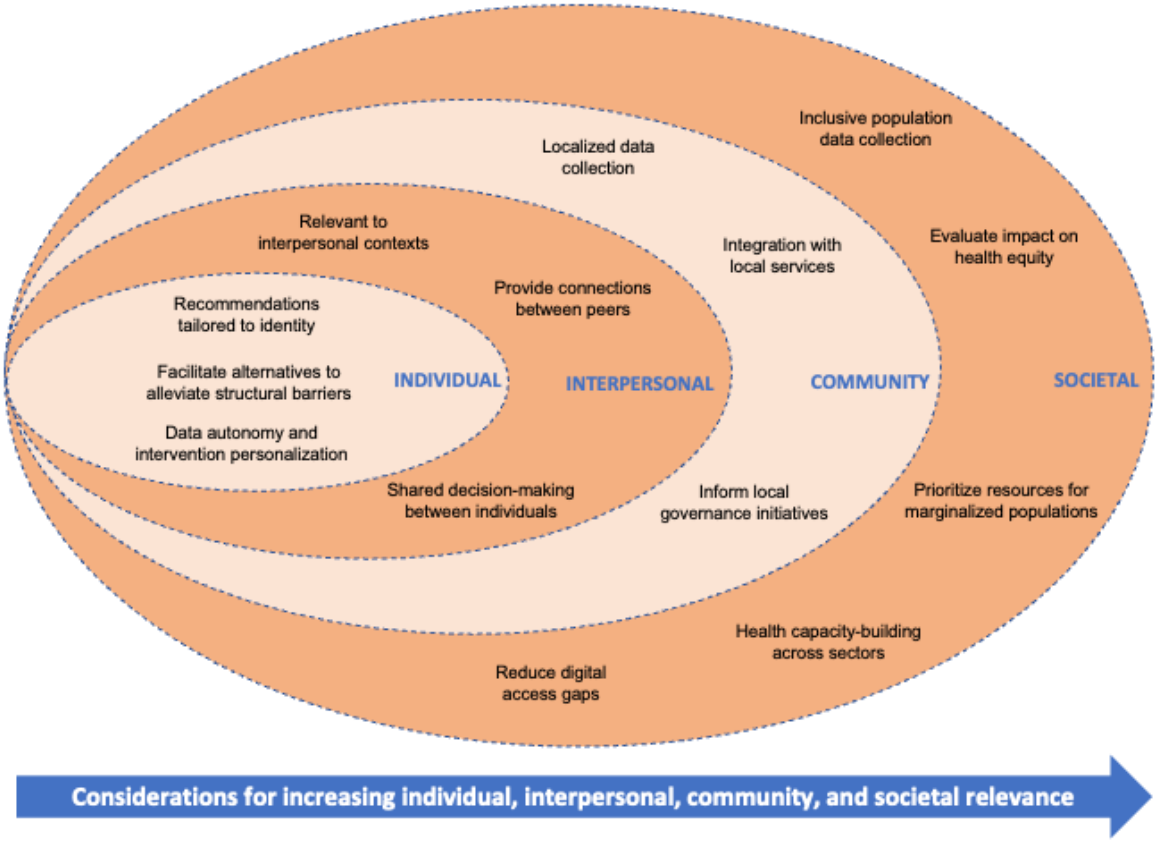

XSL•FO
RenderX

Table 1. The potential of health recommender systems and main actions.

Socioecological level	The potential of HRS[a]	Main actions
Individual	• Make tailored recommendations associated with identity (eg, healthy food associated with cultural traditions) • Include alternative recommendations to overcome structural barriers (eg, based on budget and neighborhood safety) • Allow individuals to determine what data they provide and which health domains (eg, physical, mental, and social) to focus on	• Incorporate structural barriers to good health in personalization (social determinants of health) • Centering intervention designs in the lived experience of the potential users (participatory design) • Prioritize culturally appropriate interventions (eg, by considering decolonization of predominant approaches)
Interpersonal	• Account for influence of family (eg, households) and social networks (eg, work environments) • Leverage role and data on relationships (eg, connecting patients with similar challenges and demographics and peer endorsements) • Provide opportunities for shared decision-making (eg, between patients and their caregivers)	• Centering intervention designs in the lived experience of the potential users and other stakeholders (participatory design)
Community	• Incorporate data on health factors relevant to a particular community (ie, localized data collection) • Foster collaboration with local institutions (eg, grocery stores) to improve health equity (eg, access to fresh produce) • Guide local governance initiatives for community health initiatives (eg, usage of parks)	• Centering intervention designs in the lived experience of the potential users and other stakeholders (participatory design)
Societal	• Address "health data poverty" through inclusive data collection (eg, marginalized populations • Enable policy makers to build capacity between sectors to address social determinants of health through targeted interventions (eg, identifying key areas or populations in need of public health interventions)	• Report and evaluate impact health inequities (eg, what disadvantages or privileges are perpetuated) • Consistent and high-quality internet access • Planning and delivery of HRS to marginalized populations • Advocate for health equity policies based on HRS data

[a]HRS: health recommender systems.

The Potential of HRS Using the Socioecological Model

The potential of HRS for decreasing health inequities based on the socioecological model (individual, interpersonal, community, and society) is described further in this study (Figure 1).

Individual Level

The individual level includes individual personal characteristics, such as age, education, income, and health history [19]. Using individual-level data from phones and wearables, HRS can provide personalized suggestions that integrate and recognize structural barriers to good health that marginalized groups face more often. For example, for lower SES individuals, HRS for nutrition can combine budgetary restrictions with culture and familiarity motives (eg, food associated with cultural traditions, or a sense of safety and comfort), which are more important to individuals with fewer resources and diverse cultural identities [20]. For individuals living in areas where outdoor activities may be unsafe or impossible because of weather conditions, physical activity HRS could offer alternative options, such as indoor exercises, community-centered physical activities, or free web-based classes. Furthermore, HRS can also empower users through prioritizing user decisions over system performance, which is related to digital self-efficacy, and facilitates trust in HRS. For example, in a previous study, an HRS for health coaching allowed participants to pick their own health domain to work on, and only use the information they wanted to provide, such as not using sensors or answering all health-related questions [21]. Thus, the use of HRS can lead to more equitable health care recommendations and empowered health decisions surrounding different lifestyle factors (eg, physical, mental, and social).

Interpersonal Level

The interpersonal level includes an individual's closest social circle, such as friends, partners, family members, and colleagues, all of whom can influence their health behavior [19]. HRS has the potential to positively influence the health of families, households, work environments, and other small social networks by leveraging the data of interpersonal relationships [22]. For example, HRS can provide recommendations based on what worked for others with similar health [23]. HealthNet, a patient social network platform, used a recommender system to allow patients to find others with similar conditions, learn about what treatments were useful for them, and suggest health facilities or doctors consulted by patients with similar health status [24]. However, the demographic background of patients, and the incorporation of this information into the HRS, was not reported. Basing recommendations on others with similar socioeconomic and ethnic or racial minoritized backgrounds may make HRS even more effective. Furthermore, social media network data can be used, as shown by Oliva-Felipe and colleagues [25], who

XSL•FO
RenderX

developed an HRS with educational content tailored to their specific profile of dementia caregivers. HRS could also leverage peer endorsements, allowing users to see recommendations or reviews from their social connections or members of a community with similar health goals [22]. Tapping into the power of social influence and community may specifically benefit lower SES individuals by making recommendations more relevant to them and trustworthy.

Community Level

The community level includes the settings in which people work, live, and have social relationships, such as schools, workplaces, and neighborhoods, which affect health [19]. HRS can play a pivotal role in facilitating health promotion and preventive strategies tailored to the community-specific risk and protective factors by capturing data specific to a community (ie, localized data collection). For example, Wayman and Madhvanath [26] describe an HRS that makes dietary recommendations based on grocery receipt data. Leveraging this kind of data, HRS developers could collaborate with local businesses, such as grocery and convenience stores, to tailor the amount of fresh fruits and vegetables guided by food purchasing data [27]. HRS developers can also work with local governments to develop systems that allow low-income residents to receive recommendations on where to shop for healthy and affordable foods. The data collected from HRS can also be used to work with the city or county to identify the most suitable walking trails, parks, and indoor exercise sites and publicize these to the community. Thus, the ability of HRS to leverage localized data provides ample opportunities for community-level interventions.

Societal Level

The societal level includes societal factors that influence health, such as the economic, educational, and social policies that help to create, maintain, or lessen health inequities [19].

HRS can contribute to the societal level by collecting data from individuals to build inclusive data sets with trends in health behaviors on a population level, which can be leveraged for policy making. Currently, individuals of lower SES and ethnic or racial marginalized individuals are systematically underrepresented in (open) health data sets (eg, the UK Biobank, with half a million UK participants) [28]. If HRS become widely used, by collecting data on the health of these marginalized populations in combination with contextual data, HRS can help to solve the problem of "health data poverty", the inability of individuals, groups, or populations to benefit from discovery or innovation due to a scarcity of adequately representative data [29]. Policymakers can, for example, use these data to better pinpoint inadequate health care facilities, nutritional deficiencies, or insufficient infrastructure for healthy living and use this to devise social policies.

Regarding major public health issues like climate change, which disproportionately affects the health of marginalized populations [30], HRS could provide insight into human behavior that leads to high carbon emissions, such as food waste, or actions that can promote sustainable food consumption [31]. HRS could thus help policymakers allocate resources more efficiently by identifying areas or populations in need of specific public health interventions. This application of HRS aligns with the "Health in All Policies" (HiAP) approach promoted by the World Health Organization, in which policymakers in diverse sectors (eg, environment, agriculture, and urban planning) align efforts and health [32], consider the health impacts and benefits of plans to address employment, education, or housing. Furthermore, after implementing health policies, recommender systems could track their impact on health outcomes and provide policymakers with real-time feedback on the effectiveness of their policies [33]. Data insights collected through HRS could form the basis for more targeted policy interventions to reduce health inequities.

Recommendations for Research Priorities

As we have described above in this section, HRS have immense potential to decrease health inequities. However, they are currently not designed from a health equity lens. Below we describe gaps in the current literature based on our analysis. For HRS to be more useful and accessible to marginalized populations, we need new design guidelines. Future research priorities are further described in this study that span several levels of the socioecological framework (also illustrated in Table 1).

Incorporating Structural Barriers to Good Health in Personalization (Individual and Community)

Despite the potential of incorporating many data sources, as we explained above, most HRS do not integrate factors associated with the social determinants of health [34] and primarily use basic demographic information like age and gender, as well as individual health behaviors for tailoring recommendations [2]. As also mentioned by Richardson et al [10], in their digital health equity framework, this oversight could exacerbate health inequities, as marginalized individuals are less likely to benefit from interventions focused on behavioral factors, because of limited resources and competing priorities, which are greater in these populations. Future research should explore the integration of the social determinants of health into HRS, such as SES, healthy food scarcity, and digital and health literacy.

Decolonial Perspectives in HRS Design (Individual, Interpersonal, and Community)

Because of their capability to incorporate diverse data sources, HRS are very suitable to be designed to contextualize, recognize, and incorporate diverse perspectives. Scientific knowledge in the digital health field is however currently predominantly produced through the lens of Western scholars, which marginalizes diverse and Indigenous health perspectives. In the case of HRS, users themselves have been scarcely involved in the design process. For example, in a recent review of 51 HRS, only 10 recruited users in the development of HRS testing, and the majority did not define the target group of the HRS intervention [2]. Furthermore, the majority of HRS studies are conducted in the United States and China, and low-income countries or low-resource settings are underrepresented [2].

Because of the underrepresentation of marginalized groups and lack of geographical diversity, HRS likely have cultural biases, which can make them less relevant to underrepresented groups.

XSL•FO
RenderX

For example, mental health support integrated with religious or spiritual practices may be more accessible and relevant for some underrepresented groups, but it contrasts the body-mind separation of mental health dominant in a Western biomedical paradigm [35].

Decolonization involves critically re-evaluating interventions to remove cultural biases [36]. For example, mental health support integrated with religious or spiritual practices may be more accessible and relevant for some underrepresented groups, but it contrasts the body-mind separation of mental health dominant in a Western biomedical paradigm [35]. It involves centering intervention designs on the lived experience of the potential users. Thus users, especially those from marginalized backgrounds, should be included in the design and testing of HRS. This can be done through participatory research methods, where community members collaborate on all aspects of the project with researchers and other stakeholders, including feature design, testing, and dissemination [36]. This also involves examining power relationships that may underline the use of their technologies and structural and cultural factors that may broadly influence well-being [36]. Besides user involvement, technological collaborations with traditional health practitioners, such as herbalists or local healers [35] community, or faith-based organizations could help to understand and mitigate these power imbalances [37].

Evaluating the Impact of HRS on Health Equity (Community and Societal)

HRS researchers and developers should critically evaluate and report how their solutions may impact health inequities (eg, what disadvantages or privileges the research may create) [38,39]. For example, successful implementation and deployment of HRS (and many other digital health interventions) at the community level depends on structural factors such as consistent and high-quality internet access, which may be particularly challenging for resource-constrained areas [10]. If data is not available in sufficient detail or quantities, this will impact the effectiveness of HRS, leading to their lower effectiveness for marginalized populations.

Furthermore, beyond the design phase, funding for the planning and delivery of HRS to marginalized populations is necessary, to avoid that successful HRS projects fail to scale up to maturity to impact health [40]. In addition, we also need the data to track the reach and effectiveness of HRS in decreasing health inequities (metrics beyond health outcomes) [41], such as access to and quality of care, and advocate for policy changes based on these data [42]. Addressing these structural challenges and advocating for inclusive research are essential steps toward ensuring the equitable deployment and effectiveness of HRS in marginalized communities.

Conclusion

HRS have great potential in decreasing health inequities through accessible, personalized health promotion interventions, which can influence the individual, intrapersonal, community, and societal levels. HRS can leverage data from various sources to enable deep personalization of health promotion strategies based on the wider sphere of influence of health behavior and, through localized data collection, provide actionable insights for policymakers on population health. However, we argue that, with this emerging and developing field, health equity should be a priority, not an afterthought. Future research should focus on new design paradigms, including incorporating data on the social determinants of health, designing from a lens of decolonizing HRS, and evaluating the impact on health equity. Although this research does not provide definitive solutions, it lays a foundation for future studies, and we hope it benefits ongoing discourse in the field. With changes in practices and design with an intentional equity focus, HRS could play a crucial role in access to health promotion, improving the health of marginalized groups, and decreasing health inequities.

Acknowledgments

CF and KG are supported through the “High Tech for a Sustainable Future” capacity-building program of the 4TU Federation in the Netherlands. The funders had no role in the content of the manuscript.

Authors' Contributions

CF wrote the first draft of the main manuscript. All authors contributed to the writing and editing of the scientific manuscript. KG prepared the figure. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. De Croon R, Van Houdt L, Htun NN, Štiglic G, Vanden Abeele V, Verbert K. Health recommender systems: systematic review. J Med Internet Res 2021;23(6):e18035 [FREE Full text] [doi: 10.2196/18035] [Medline: 34185014]
2. Sun Y, Zhou J, Ji M, Pei L, Wang Z. Development and evaluation of health recommender systems: systematic scoping review and evidence mapping. J Med Internet Res 2023;25:e38184 [FREE Full text] [doi: 10.2196/38184] [Medline: 36656630]

XSL•FO
RenderX

3. Schäfer H, Hors-Fraile S, Karumur R, Calero VA, Said A, Torkamaan H, et al. Towards health (aware) recommender systems. : Association for Computing Machinery; 2017 Presented at: DH '17: International Conference on Digital Health; 2017 July 2-5; London, United Kingdom p. 157-161. [doi: 10.1145/3079452.3079499]
4. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR mHealth uHealth 2022;10(6):e35053 [FREE Full text] [doi: 10.2196/35053] [Medline: 35679107]
5. Etemadi M, Bazzaz Abkenar S, Ahmadzadeh A, Haghi Kashani M, Asghari P, Akbari M, et al. A systematic review of healthcare recommender systems: open issues, challenges, and techniques. Expert Systems with Applications 2023;213:118823. [doi: 10.1016/j.eswa.2022.118823]
6. EU citizens: over half seek health information online. Eurostat. 2022. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20220406-1 [accessed 2023-10-31]
7. Chinnasamy P, Wong WK, Raja AA, Khalaf OI, Kiran A, Babu JC. Health recommendation system using deep learning-based collaborative filtering. Heliyon 2023;9(12):e22844 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e22844] [Medline: 38144343]
8. Ayenigbara IO. The evolving nature of artificial intelligence: role in public health and health promotion. J Public Health (Oxf) 2024;46(2):e322-e323. [doi: 10.1093/pubmed/fdad240] [Medline: 37973395]
9. Lawrence K. In: Linwood SL, editor. Digital Health Equity. Brisbane: Exon Publications; 2022.
10. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
11. World Health Organization. Equity within digital health technology within the WHO European Region: a scoping review. Pesquisa. 2022. URL: https://pesquisa.bvsalud.org/portal/resource/pt/who-365326 [accessed 2024-12-27]
12. Lee BC, Bendixsen C, Liebman AK, Gallagher SS. Using the socio-ecological model to frame agricultural safety and health interventions. J Agromedicine 2017;22(4):298-303 [FREE Full text] [doi: 10.1080/1059924X.2017.1356780] [Medline: 28762886]
13. Lyles CR, Nguyen OK, Khoong EC, Aguilera A, Sarkar U. Multilevel determinants of digital health equity: a literature synthesis to advance the field. Annu Rev Public Health 2023;44:383-405 [FREE Full text] [doi: 10.1146/annurev-publhealth-071521-023913] [Medline: 36525960]
14. Golden SD, McLeroy KR, Green LW, Earp JAL, Lieberman LD. Upending the social ecological model to guide health promotion efforts toward policy and environmental change. Health Educ Behav 2015;42(1 Suppl):8S-14S. [doi: 10.1177/1090198115575098] [Medline: 25829123]
15. Trickett EJ, Rowe HL. Emerging ecological approaches to prevention, health promotion, and public health in the school context: next steps from a community psychology perspective. Journal of Educational and Psychological Consultation 2012;22(1-2):125-140. [doi: 10.1080/10474412.2011.649651]
16. Valentine L, D'Alfonso S, Lederman R. Recommender systems for mental health apps: advantages and ethical challenges. AI Soc 2022:1-12 [FREE Full text] [doi: 10.1007/s00146-021-01322-w] [Medline: 35068708]
17. World Health Organization. Widening inequities, declining trust – they are inextricably linked, with significant impacts on health, finds new WHO/Europe report. URL: https://www.who.int/europe/news/item/12-07-2023-widening-inequities--declining-trust---they-are-inextricably-linked--with-significant-impacts-on-health--finds-new-who-europe-report [accessed 2024-01-26]
18. Poleshuck E, Perez-Diaz W, Wittink M, ReQua M, Harrington A, Katz J, et al. Resilience in the midst of chaos: socioecological model applied to women with depressive symptoms and socioeconomic disadvantage. J Community Psychol 2019;47(5):1000-1013 [FREE Full text] [doi: 10.1002/jcop.22188] [Medline: 30999386]
19. Chapter 1: Models and Frameworks. ATSDR. URL: https://www.atsdr.cdc.gov/community-engagement/php/chapter-1/models-frameworks.html?CDC_AAref_Val=https://www.atsdr.cdc.gov/communityengagement/pce_models.html [accessed 2024-11-05]
20. Konttinen H, Sarlio-Lähteenkorva S, Silventoinen K, Männistö S, Haukkala A. Socio-economic disparities in the consumption of vegetables, fruit and energy-dense foods: the role of motive priorities. Public Health Nutr 2013;16(5):873-882 [FREE Full text] [doi: 10.1017/S1368980012003540] [Medline: 22857602]
21. Orte S, Migliorelli C, Sistach-Bosch L, Subías-Beltrán P, Cecilia FP, Galofré M, et al. BECOME: A modular recommender system for coaching and promoting empowerment in healthcare. In: Stawicki DSP, editor. Recommender Systems [Working Title]. London, England: IntechOpen; 2023.
22. Sedrakyan G, Gavai A, van Hillegersberg J. Design implications towards human-centric semantic recommenders for sustainable food consumption. In: Advances in Conceptual Modeling. Switzerland: Springer Nature; 2023:312-328.
23. Dempsey RC, McAlaney J, Bewick BM. A critical appraisal of the social norms approach as an interventional strategy for health-related behavior and attitude change. Front Psychol 2018;9:2180 [FREE Full text] [doi: 10.3389/fpsyg.2018.02180] [Medline: 30459694]
24. Narducci F, Lops P, Semeraro G. Power to the patients: the HealthNetsocial network. Information Systems 2017;71:111-122. [doi: 10.1016/j.is.2017.07.005]
25. Oliva-Felipe L, Barrué C, Cortés A, Wolverson E, Antomarini M, Landrin I, et al. Health recommender system design in the context of CAREGIVERSPRO-MMD Project. : Association for Computing Machinery; 2018 Presented at: PETRA

XSL•FO
RenderX

'18: The 11th PErvasive Technologies Related to Assistive Environments Conference; 2018 June 26-29; Corfu, Greece p. 462-469. [doi: 10.1145/3197768.3201558]
26. Wayman E, Madhvanath S. Nudging grocery shoppers to make healthier choices. USA: Association for Computing Machinery; 2015 Presented at: Proceedings of the 9th ACM Conference on Recommender Systems; 2015 September 16-20; Vienna, Austria p. 289-292. [doi: 10.1145/2792838.2799669]
27. Schaffer R. 'Harnessing the supermarket': grocery store-led intervention improved diet quality. Healio. 2022. URL: https://www.healio.com/news/cardiology/20220403/harnessing-the-supermarket-grocery-storeled-intervention-improved-diet-quality [accessed 2023-11-04]
28. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am J Epidemiol 2017;186(9):1026-1034 [FREE Full text] [doi: 10.1093/aje/kwx246] [Medline: 28641372]
29. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021;3(4):e260-e265 [FREE Full text] [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]
30. Smith GS, Anjum E, Francis C, Deanes L, Acey C. Climate change, environmental disasters, and health inequities: the underlying role of structural inequalities. Curr Environ Health Rep 2022;9(1):80-89. [doi: 10.1007/s40572-022-00336-w] [Medline: 35338470]
31. Aschemann-Witzel J, De HI, Amani P, Bech-Larsen T, Oostindjer M. Consumer-related food waste: causes and potential for action. Sustain Sci Pract Policy Multidisciplinary Digital Publishing Institute 2015;7(6):6457-6477. [doi: 10.3390/su7066457]
32. de Leeuw E. Engagement of sectors other than health in integrated health governance, policy, and action. Annu Rev Public Health 2017;38:329-349 [FREE Full text] [doi: 10.1146/annurev-publhealth-031816-044309] [Medline: 28125390]
33. Ricci F, Rokach L, Shapira B. Introduction to recommender systems handbook. In: Recommender Systems Handbook. Boston, MA: Springer; 2011:1-35.
34. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2(Suppl 2):19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
35. Bhattacharjee A, Sultana S, Amin MR, Iqbal Y, Ahmed SI. 'What's the Point of Having This Conversation?': from a telephone crisis helpline in Bangladesh to the decolonization of mental health services. ACM J Comput Sustain Soc 2023;1(2):1-29 [FREE Full text] [doi: 10.1145/3616381]
36. Pendse SR, Nkemelu D, Bidwell NJ, Jadhav S, Pathare S, De CM, et al. From treatment to healing: envisioning a decolonial digital mental health. 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; 2022 April 29; New Orleans, LA, USA p. 1-23. [doi: 10.1145/3491102.3501982]
37. Levin J, Hein JF. A faith-based prescription for the Surgeon General: challenges and recommendations. J Relig Health 2012;51(1):57-71. [doi: 10.1007/s10943-012-9570-4] [Medline: 22246656]
38. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021;3(8):e526-e533 [FREE Full text] [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
39. Ovalle A, Subramonian A, Gautam V, Gee G, Chang KW. Factoring the matrix of domination: a critical review and reimagination of intersectionality in AI fairness. Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society 2023 Aug 08:496-511 [FREE Full text] [doi: 10.1145/3600211.3604705]
40. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
41. Gallegos-Rejas VM, Thomas EE, Kelly JT, Smith AC. A multi-stakeholder approach is needed to reduce the digital divide and encourage equitable access to telehealth. J Telemed Telecare 2023;29(1):73-78. [doi: 10.1177/1357633X221107995] [Medline: 35733379]
42. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood) 2014;33(7):1115-1122. [doi: 10.1377/hlthaff.2014.0147] [Medline: 25006136]

Abbreviations

HiAP: Health in All Policies
HRS: health recommender systems
SES: socioeconomic status

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 06.05.24; peer-reviewed by K Valogianni, MS Williams; comments to author 01.08.24; revised version received 07.11.24; accepted 25.11.24; published 30.01.25.

Please cite as:
Figueroa CA, Torkamaan H, Bhattacharjee A, Hauptmann H, Guan KW, Sedrakyan G
Designing Health Recommender Systems to Promote Health Equity: A Socioecological Perspective
J Med Internet Res 2025;27:e60138
URL: https://www.jmir.org/2025/1/e60138
doi:10.2196/60138
PMID:

©Caroline A Figueroa, Helma Torkamaan, Ananya Bhattacharjee, Hanna Hauptmann, Kathleen W Guan, Gayane Sedrakyan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Why AI Monitoring Faces Resistance and What Healthcare Organizations Can Do About It: An Emotion-Based Perspective

Karl Werder[1*], PhD; Lan Cao[2*], PhD; Eun Hee Park[2*], PhD; Balasubramaniam Ramesh[3*], PhD

[1]Digital Business Innovation, IT University of Copenhagen, Copenhagen, Denmark

[2]Information Technology & Decision Sciences, Strome College of Business, Old Dominion University, Norfolk, VA, United States

[3]Computer Information Systems, J. Mack Robinson College of Business, Georgia State University, Atlanta, GA, United States

[*]all authors contributed equally

Corresponding Author:
Karl Werder, PhD
Digital Business Innovation
IT University of Copenhagen
Rued Langgaards Vej 7
Copenhagen, 2300
Denmark
Phone: 45 72185386
Email: karw@itu.dk

Abstract

Continuous monitoring of patients' health facilitated by artificial intelligence (AI) has enhanced the quality of health care, that is, the ability to access effective care. However, AI monitoring often encounters resistance to adoption by decision makers. Healthcare organizations frequently assume that the resistance stems from patients' rational evaluation of the technology's costs and benefits. Recent research challenges this assumption and suggests that the resistance to AI monitoring is influenced by the emotional experiences of patients and their surrogate decision makers. We develop a framework from an emotional perspective, provide important implications for healthcare organizations, and offer recommendations to help reduce resistance to AI monitoring.

(J Med Internet Res 2025;27:e51785) doi:10.2196/51785

KEYWORDS

artificial intelligence; AI monitoring; emotion; resistance; health care

Introduction

Continuous monitoring through artificial intelligence (AI) technology is becoming increasingly important in health care. AI is becoming an integral part of health care, for example, in pain monitoring [1], in creating medical imaging platforms [2], and in delivering timely medical interventions to patients [3] to increase the quality of care, that is, the ability to access effective care [4]. AI monitoring solutions use machine learning techniques to learn from data generated from adhesive patches, sensor devices, video cameras, and other devices, and identify risks of illnesses and adverse events. Given the potential benefits that AI monitoring offers to health care [5], regulators, and healthcare (we use both terms, healthcare and health care. Healthcare refers to an industry or a system that provides people with health care, whereas health care refers to the process of care or things that health professionals do) organizations strongly advocate its use. For instance, the Food and Drug Administration has been actively approving remote monitoring devices for patient care [6], while prominent hospital systems like Stanford Medical are developing AI monitoring solutions for senior care [7].

However, novel solutions such as AI monitoring often encounter resistance from users [8,9]. Such resistance is usually attributed to users' risk aversion toward innovation [10], sometimes called "liability of newness" [11], and their cognitive assessments of the costs and benefits of AI. However, we argue that emotions play an important role in AI monitoring resistance, even more than a rational evaluation of the decision, given that decision makers, specifically, patients and family members, lack extensive experience with novel AI solutions [10], which limits their ability to conduct a thorough cost-benefit analysis. Health care decisions often carry high stakes, and making the wrong choices can lead to serious consequences, including the loss of life. Therefore, making emotion-driven decisions in adopting AI-based solutions presents a significant challenge for healthcare organizations.

XSL•FO
RenderX

In this viewpoint paper, we discuss why the current view on resistance to AI monitoring may not fully capture the underlying reasons. We develop a framework from an emotional perspective that explains why decision makers resist AI monitoring and propose solutions that alleviate their concerns.

What Is Missing in the Current View of AI Monitoring Resistance?

Healthcare organizations often attribute resistance to patients' reluctance to accept innovative solutions to their cognitive assessment of the technology's costs and benefits [12]. For instance, prior studies have reported that some patients' resistance to AI systems results from their evaluations regarding privacy intrusion and insecurity of sensitive medical data because AI systems may be vulnerable to data breaches and data misuse [13]. Some patients may have doubts about the accuracy and reliability of AI systems [14] or do not understand and fully comprehend how AI functions, and hence, are reluctant to adopt these technologies [15]. Ethical dilemmas regarding AI decision-making in healthcare—in part, stemming from the fact that algorithms are inherently biased [16] with serious consequences of medical decisions—further contribute to the resistance of AI systems [13].

Taken together, the AI literature assumes that resistance decisions are rational and that patients themselves make the decisions about adopting the AI monitoring system. However, recent studies challenge these assumptions [8]. AI systems are opaque and often lack transparency [17], making it difficult for decision makers to rationally analyze their costs and benefits. Given the complexity of AI systems, decision makers often lack a comprehensive understanding of this advanced technology, exacerbating the challenges presented to rational decision-making [15]. Thus, we argue that alongside cognitive assessments of AI's costs and benefits, emotions play a crucial role in AI monitoring resistance [3]. Decision makers may include others such as family members because, for example, senior citizens often have limited knowledge of technology and frequently turn to surrogate decision makers, such as their adult children, for help when making technology decisions [18,19].

The AI Resistance Framework: An Emotion Perspective

Emotion is a complex psychological state that involves "loosely coupled changes in the domains of subjective experience, behavior, and peripheral physiology" [20]. Emotions are often triggered by external stimuli (eg, events or surrounding situations) [21], and they can be instrumental in directing attention to critical environmental details, refining decision-making, and aiding behavioral responses [22]. However, emotions can also be detrimental when they are inappropriate in type, intensity, or duration for a given situation. Therefore, individuals strive to regulate their emotions [22].

Emotion regulation refers to the "attempt to influence which emotions one has when one has them, and how one experiences or expresses these emotions" [20]. The goal of emotion regulation is to achieve some valued end (eg, decreasing negative emotion; Gross [20]). There are two types of emotion regulation [22]: Intrinsic emotion regulation focuses on regulating one's own emotions, while extrinsic emotion regulation involves regulating another person's emotions. According to a process model of emotion regulation proposed by Gross [23], there are five types of emotion regulation strategies: situation selection, situation modification, attentional deployment, cognitive change, and response modulation. These strategies can influence both the individuals practicing them and those around them [20]. This view on emotion regulation has attracted considerable interest across various domains, such as psychology, business, sociology, and healthcare, because of its potential to improve mental health, job performance, and social harmony, among others [20]. Focusing on the emotions experienced and regulated by decision makers, as well as policies and practices used by healthcare organizations that impact patients' emotions, we present a framework that explains how emotions significantly impact resistance to AI monitoring (Figure 1).

XSL•FO
RenderX

Figure 1. A framework of AI resistance from an emotion perspective. AI: artificial intelligence.

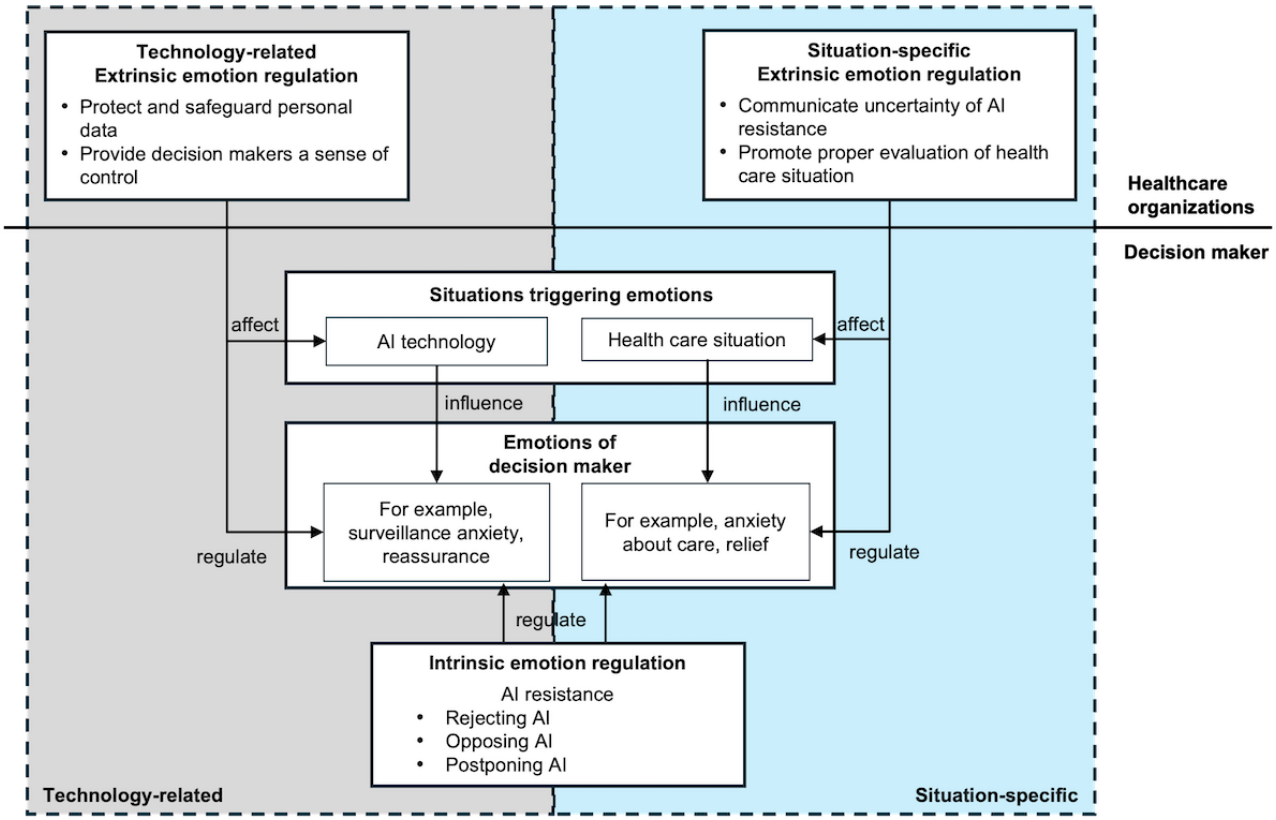

Decision Makers in Health Care

Health care is a complex process involving various stakeholders who may take on decision-making responsibilities based on a given situation. Patients are typically responsible for their own health decisions, such as following medical advice, using technological aids, or seeking further professional support. However, patients can also rely on others to guide them and sometimes even make important decisions on their behalf. For example, senior citizens often rely on family members for important medical and technology-related decisions. They may ask their adult children to install medical apps or help them navigate through their electronic healthcare records. Thus, key decision makers include patients and family members who act as surrogate decision makers [8,24].

Emotions of Decision Makers and Their Triggers

Overview

Emotions, both positive (eg, happiness and joy) and negative (eg, anxiety and fear), are triggered by specific experiences or events [25]. For example, decision makers may become anxious about adverse health outcomes. In the context of AI monitoring, emotions arise in response to the criticality of a given situation [26] and the capabilities of the AI solution [27]. Thus, we suggest that emotions play an important role in resistance to AI monitoring.

Figure 1 presents two noteworthy sources triggering negative emotions, such as anxiety or fear, and positive emotions, such as happiness and joy. The first source relates to AI technology. Emotions triggered by AI monitoring systems encompass both negative and positive emotions. Anxiety about patients' surveillance and apprehension about transferring the responsibility of patient monitoring from the caretaker to the AI solution are common [8]. Conversely, positive emotions such as reassurance and comfort can arise from the continuous monitoring and alerts provided by AI systems, enhancing the sense of security for both patients and their families. The second source is specific to health care. Emotions triggered by the care process and its outcomes depend on the criticality of the health care situation. A critical health care situation, for example, would involve the possibility of the patient being hospitalized [8]. Conversely, positive emotions such as hope and relief can emerge when effective care processes and interventions are in place. In the following, we discuss some exemplary emotions triggered by both sources, AI technology and the health care situation.

Emotions Triggered by AI Technology

AI technology can trigger a range of emotional responses. For example, reassurance is fostered by enhanced efficiency and reliability in patient care enabled by AI monitoring. With AI monitoring, specific care activities are managed by the system, which allows caretakers and health care providers to focus on more complex and personalized care tasks [28]. For example, AI monitoring can provide real-time alerts and updates about a patient's condition, ensuring timely interventions and reducing the risk of human error. Decision makers feel reassured by the increased accuracy and efficiency that AI brings, as it supports health care providers in delivering high-quality care.

XSL•FO
RenderX

However, AI can also trigger negative emotions. Many decision makers are anxious because they do not know what information will be recorded. Surveillance anxiety is caused by sensor-based AI systems that provide extensive tracking of users and their behavior [27]. It is a feeling of tension and worry from thoughts resulting from the use of monitoring solutions. Many decision makers experience surveillance anxiety and worry about the continuous monitoring of patients and the collection and analysis of patient's data, leading to privacy and possibly security concerns. Surprisingly, a recent study [8] has found that the level of health risk has a limited impact on surveillance anxiety. Rather, many worry about their patients being monitored even under conditions of high uncertainty, such as the high possibility of being hospitalized.

Delegation anxiety is caused by the delegation of some health care tasks to AI, leading to a loss of personal interaction between health care providers and patients [29]. With AI monitoring, certain care activities are delegated to the system, which reduces the workload of caretakers and care providers [28]. For example, AI monitoring can automatically communicate critical patient information to health care providers. Decision makers worry about a loss of personal interaction because technology use reduces the interaction between health care providers and their patients. Similar to surveillance anxiety, research has found that they experience delegation anxiety even when uncertainty is high, even though they are less likely to resist AI monitoring [8].

Technology can also provide positive emotional experiences. For example, when designing monitoring technology, providing rewards for goal attainment can reinforce self-efficacy and provide positive emotional responses that support a patient in managing their diabetes medication [30]. Reinforcements and reiterations of success are common techniques used in gamification, increasing positive emotional responses and motivation [31].

Emotions Triggered by Health Care Situation

The health care situation of a patient also affects the emotions experienced by decision makers [8]. For example, anxiety about health care is experienced when decision makers worry about whether patients receive appropriate health care and monitoring of their health in highly uncertain situations. The process of care involves monitoring the health status [32] and providing physical, psychological, social, and spiritual support [33]. This was particularly acute during the pandemic, when decision makers were anxious about the prospect of their own or relatives' health problems going unnoticed especially when they belonged to a high-risk population [34].

Anxiety about health outcomes is experienced when decision makers worry about the potential negative health outcome for the patient. The health outcome includes positive developments (eg, improvement of symptoms) and negative developments (eg, deterioration of symptoms and hospitalization) [35]. Anxiety about health outcomes varies with the level of uncertainty of health outcomes, that is, decision makers experience more anxiety about health outcomes when uncertainty is high and less when uncertainty is low.

Conversely, relief and hope can be experienced when decision makers are provided with sufficient care and treatment information. For example, knowing that patients receive continuous, attentive care and having access to clear, timely updates about their treatment progress can foster feelings of relief [36]. This information helps patients feel more secure about their health care situation and strengthens their trust in the health care received, enhancing their overall emotional comfort.

Intrinsic Emotion Regulation With AI Resistance

Some emotions can be beneficial while others can be detrimental as they influence how decision makers perceive and interpret sensory information and shape their decisions based on such information, leading to either adaptive or maladaptive behaviors [20]. For example, decision makers' emotions can increase and decrease resistance [8]. Resistance decreases with technology-induced relief, increases with technology-induced anxiety, and decreases with health care situation-specific anxiety. Since decision makers are motivated to regulate their own emotions, particularly negative emotions, such as anxiety, anger, and fear, it is crucial to guide them toward beneficial regulation activities [20].

To regulate emotions, decision makers can use an emotion regulation strategy called situation selection, where they take actions to increase or decrease the likelihood of being in a situation that is expected to elicit desirable or undesirable emotions [20]. For example, decision makers may postpone, reject, or oppose AI monitoring. Postponing the adoption of AI monitoring solutions can be a form of passive innovation resistance, allowing them to avoid the emotion-inducing situation altogether [37]. This strategy works when there is an alternative system in place and there is no mandate to move to the new system. Often, the decision to postpone is driven by the decision maker's resistance to change as they feel more comfortable with the status quo.

However, decision makers often cannot delay a decision. Consequently, they may decide to reject AI monitoring due to a lack of prior experience. Decision makers' resistance to change and satisfaction with the status quo can catalyze this, as well as contextual factors, such as functional or psychological barriers [37]. Functional barriers can occur when decision makers perceive substantial challenges or drawbacks associated with adopting an innovation, such as difficulties in use, perceived lack of added value, or potential risks. On the other hand, psychological barriers emerge when the innovation clashes with decision makers' existing beliefs or perceptions, which may be influenced by sources like rumors or media [38]. Decision makers may oppose AI monitoring when they believe it is inherently unsuitable for their needs, even prior to a thorough evaluation. This resistance can manifest as attacking AI monitoring and the spreading of negative opinions about it [39].

XSL•FO
RenderX

Extrinsic Emotion Regulation Through Managerial Actions

Overview

Healthcare organizations and governmental institutions, as well as external entities with significant influence, have a vested interest in regulating decision makers' emotions. As collective actors, these organizations engage in cognitive reasoning and shape emotional understanding to guide thinking and actions [40,41]. They also practice collective emotion regulation to ensure informed decision-making [42]. In addition to regulating the emotions of their collective members, these organizations can also engage in extrinsic emotion regulation. This involves managing the emotions of external individuals, such as decision makers including patients and their families, to mitigate the rejection of AI innovations given their benefits and performance gains. We suggest the following emotion regulation strategies for organizations to manage decision makers' emotions and mitigate AI resistance.

Protect and Safeguard Personal Data for Responsible AI Solutions

Organizations can manage decision makers' emotions by adopting the emotion regulation strategy called situation modification. This involves actively changing situations and reshaping events that induce negative emotions to reduce their effects [20]. More precisely, organizations need to be cognizant of decision makers' emotions that may arise from a specific health care situation and the use of AI monitoring. A significant technology-related emotion is anxiety about the unauthorized disclosure and use of personal data. Healthcare organizations should carefully evaluate and choose an AI monitoring solution that is designed to protect and safeguard personal and other sensitive data collected from decision makers and others involved in a patient's care, for example, by blurring images or deidentifying the data before they are stored [8]. As AI monitoring solutions collect, store, and process data to provide predictions and recommendations, organizations need to balance ethics, personal privacy, and data security considerations against gains in quality of care and health outcomes [43]. In addition, healthcare organizations should make sure that there are sufficient human interactions with patients when deploying AI monitoring solutions. Healthcare organizations should make decision makers aware that AI monitoring can even be used to enhance engagement by enforcing personal interactions and providing personalized care.

Governmental institutions should proactively establish proper laws and regulations to manage AI monitoring. Examples include a blueprint from the White House that seeks to prevent negative implications of AI and its monitoring [44] and a draft from the European Union that seeks to ban AI for mass monitoring [45]. These initiatives are in response to criticism about some AI monitoring practices [46]. Seemingly, organizations are often too lax when managing personal data. However, when organizations take these concerns seriously, they can develop and deploy responsible AI solutions [47]. For instance, increasing transparency of essential algorithms, their data, and data processing allows others to evaluate the appropriateness of the AI solution and engages them in an open debate about corporate and governmental responsibilities concerning the design and deployment of AI solutions.

Provide Decision Makers a Sense of Control

Organizations can also manage decision makers' emotions by giving them a sense of control [8], that is, the feeling of being empowered and having the capability to influence or manage the operation of AI systems and their impact on decision makers' personal or professional lives. Decision makers typically regulate their emotions by choosing situations that either enhance or reduce the chances of experiencing favorable or unfavorable emotions [20].

Here, organizations can support decision makers by creating environments where they experience this sense of control. For instance, healthcare organizations should ensure that AI solutions are designed to provide decision makers with options to change the schedule, frequency, and type of data collected. Providing sufficient choice, freedom, and autonomy regarding what is monitored can alleviate decision makers concerns and reduce their negative technology-induced emotions, such as surveillance anxiety [8].

Additionally, regular training and educational resources on AI systems can help decision makers understand and confidently engage with these tools. These resources should cover how AI operates, the available options, and the benefits and limitations of the technology. By fostering a thorough understanding, organizations can help decision makers feel more in control and less anxious.

Communicate the Risk of AI Resistance

Healthcare organizations should communicate the potential consequences of not using AI monitoring to address resistance and foster adoption. First, healthcare organizations need to communicate the risk with the status quo. Health risks might already be present and decision makers seek to manage such risks. Many decision makers may even lack a clear understanding of the risks faced by the patients. For example, more than one out of four people aged 65 years or older falls each year [48], but less than half tell their doctor and caretaker [49] about these incidents. Therefore, it is important to communicate to decision makers the risk of adverse health outcomes such as a fall that may go undetected in the absence of adequate monitoring.

Second, healthcare organizations need to communicate the risk of AI resistance, that is, they need to articulate how AI capabilities can alleviate existing risks through its innate ability to act. For example, in addition to educating the decision makers about the ability of AI monitoring systems to continuously monitor potential symptoms and provide early detection of serious illnesses [50], healthcare organizations can ask decision makers to think about what would happen if patients' abnormal activities and adverse events such as falls are not noticed on time. When healthcare organizations also communicate the risk of AI resistance, decision makers can assess the existing risks more comprehensively, and thus, may choose to adopt AI monitoring.

XSL•FO
RenderX

Promote Proper Evaluation of Health Care Situation

Despite the threats and challenges presented by AI, the technology offers significant advantages, such as assistance in identifying health problems, monitoring for adverse events (eg, falls), and detecting abnormal behaviors (eg, wandering). When decision makers focus on these advantages and the opportunities AI provides in a given health care situation, they experience positive emotions such as reassurance and are more likely to evaluate the systems in a positive light. To assist decision makers in navigating healthcare processes and associated AI tools, organizations should enhance communication by providing clear, consistent updates about AI-powered healthcare procedures, patient status, and quality measures [51,52]. This transparency helps reduce concerns and anxiety, ensuring decision makers feel informed and confident in the AI systems. They should also develop personalized care plans that integrate AI system capabilities to address each patient's unique needs and involve family members in the planning process [53]. Tailoring care with AI support fosters trust and reassures decision makers by demonstrating a commitment to their needs.

In addition, establishing robust crisis management protocols that incorporate AI tools for real-time monitoring and response provides immediate support and reassurance [54]. This helps decision makers feel secure about the organization's ability to manage emergencies effectively. By implementing these strategies, healthcare organizations can mitigate AI resistance by enhancing trust, reducing anxiety, and fostering a positive emotional response toward AI-assisted care.

Acknowledgments

We thank the participants of the Digital Health Research Community, Deakin University (AU) and Digital Health Analytics Seminar at ITU University of Copenhagen (DK) with whom we discussed our research and from which we drew a lot of inspiration and ideas when writing this article. One of the author's research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement W911NF-23-2-0224. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Conflicts of Interest

None declared.

References

1. Racine N, Chow C, Hamwi L, Bucsea O, Cheng C, Du H, et al. Health care professionals' and parents' perspectives on the use of AI for pain monitoring in the neonatal intensive care unit: multisite qualitative study. JMIR AI 2024 Feb 09;3:e51535 [FREE Full text] [doi: 10.2196/51535] [Medline: 38875686]
2. Karpathakis K, Pencheon E, Cushnan D. Learning from international comparators of national medical imaging initiatives for AI development: multiphase qualitative study. JMIR AI 2024 Jan 04;3:e51168 [FREE Full text] [doi: 10.2196/51168] [Medline: 38875584]
3. Nathan J. Four ways artificial intelligence can benefit robotic surgery. Forbes. 2023 Feb 15. URL: https://www.forbes.com/sites/forbestechcouncil/2023/02/15/four-ways-artificial-intelligence-can-benefit-robotic-surgery/ [accessed 2023-05-15]
4. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med 2000 Dec;51(11):1611-1625. [doi: 10.1016/s0277-9536(00)00057-5] [Medline: 11072882]
5. Han JH, Lee JY. Digital healthcare industry and technology trends. 2021 Presented at: IEEE International Conference on Big Data and Smart Computing (BigComp); January 17-20, 2021; Jeju Island, Korea p. 375-377. [doi: 10.1109/BigComp51126.2021.00083]
6. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digital Med 2020;3:118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
7. Li FF, Milstein A. Stanford partnership in AI-assisted care. Stanford University. URL: https://aicare.stanford.edu [accessed 2020-08-24]
8. Park EH, Werder K, Cao L, Ramesh B. Why do family members reject AI in health care? Competing effects of emotions. J Manage Inf Syst 2022;39(3):765-792 [FREE Full text] [doi: 10.1080/07421222.2022.2096550]
9. Coiera E. The last mile: where artificial intelligence meets reality. J Med Internet Res 2019;21(11):e16323 [FREE Full text] [doi: 10.2196/16323] [Medline: 31702559]
10. Heidenreich S, Kraemer T. Innovations—doomed to fail? Investigating strategies to overcome passive innovation resistance. J Prod Innovation Manage 2015;33(3):277-297. [doi: 10.1111/jpim.12273]
11. Stinchcombe AL. Social structure and organizations. In: March J, editor. Handbook of Organizations. Chicago, IL: Rand McNally; 1965:142.
12. Peek STM, Wouters EJM, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]

XSL•FO
RenderX

13. Williamson S, Prybutok VR. Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. Appl Sci 2024;14(2):675. [doi: 10.3390/app14020675]
14. Balagurunathan Y, Mitchell R, El Naqa I. Requirements and reliability of AI in the medical context. Phys Med 2021;83:72-78 [FREE Full text] [doi: 10.1016/j.ejmp.2021.02.024] [Medline: 33721700]
15. Demírhan A, Örgev C, Aslankiliç M. Perception of mistrust towards artificial intelligence applications in the health sector: causes, effects and solutions. Int J Active Healthy Aging 2023;1(1):1-6. [doi: 10.5281/zenodo.10449348]
16. Werder K, Cao L, Ramesh B, Park EH. Empower diversity in AI development: diversity practices that mitigate social biases from creeping into your AI. Commun ACM 2024;67(12):31-34 [FREE Full text] [doi: 10.1145/3676885]
17. Berente N, Gu B, Recker J, Santhanam R. Special Issue Editor's Comments: Managing Artificial Intelligence. MIS 2021;45(3):1433-1450.
18. Luijkx K, Peek S, Wouters E. "Grandma, You Should Do It—It's Cool" older adults and the role of family members in their acceptance of technology. Int J Environ Res Public Health 2015;12(12):15470-15485 [FREE Full text] [doi: 10.3390/ijerph121214999] [Medline: 26690188]
19. Berridge C, Wetle TF. Why older adults and their children disagree about in-home surveillance technology, sensors, and tracking. Gerontologist 2020;60(5):926-934. [doi: 10.1093/geront/gnz068] [Medline: 31102442]
20. Gross JJ. Emotion regulation: current status and future prospects. Psychol Inq 2015;26(1):1-26. [doi: 10.1080/1047840X.2014.940781]
21. Aldao A. The future of emotion regulation research: capturing context. Perspect Psychol Sci 2013;8(2):155-172. [doi: 10.1177/1745691612459518] [Medline: 26172497]
22. Gross JJ. Handbook of Emotion Regulation. New York: The Guilford Press; 2014.
23. Gross JJ. The emerging field of emotion regulation: an integrative review. Rev Gen Psychol 1998;2(3):271-299. [doi: 10.1037/1089-2680.2.3.271]
24. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digital Med 2021;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
25. Frijda NH. The Laws of Emotion. London, England, United Kingdom: Psychology Press; 2017.
26. Loewenstein GF, Weber EU, Hsee CK, Welch N. Risk as feelings. Psychol Bull 2001;127(2):267-286. [doi: 10.1037/0033-2909.127.2.267] [Medline: 11316014]
27. Kummer TF, Recker J, Bick M. Technology-induced anxiety: manifestations, cultural influences, and its effect on the adoption of sensor-based technology in German and Australian hospitals. Inf Manage 2017;54(1):73-89. [doi: 10.1016/j.im.2016.04.002]
28. Rubeis G. The disruptive power of artificial intelligence. Ethical aspects of gerontechnology in elderly care. Arch Gerontol Geriatr 2020;91:104186. [doi: 10.1016/j.archger.2020.104186] [Medline: 32688106]
29. Baird A, Maruping LM. The next generation of research on IS use: a theoretical framework of delegation to and from agentic IS artifacts. MIS Q 2021;45(1):315-341 [FREE Full text]
30. Toscos T, Connelly K, Rogers Y. Designing for positive health affect: decoupling negative emotion and health monitoring technologies. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops; May 05-08, 2013; Venice, Italy p. 153-160. [doi: 10.4108/icst.pervasivehealth.2013.252116]
31. Morschheuser B, Hassan L, Werder K, Hamari J. How to design gamification? A method for engineering gamified software. Inf Software Technol 2018;95:219-237. [doi: 10.1016/j.infsof.2017.10.015]
32. Schumacher KL, Stewart BJ, Archbold PG, Dodd MJ, Dibble SL. Family caregiving skill: development of the concept. Res Nurs Health 2000;23(3):191-203. [doi: 10.1002/1098-240x(200006)23:3<191::aid-nur3>3.0.co;2-b] [Medline: 10871534]
33. Bevan JL, Pecchioni LL. Understanding the impact of family caregiver cancer literacy on patient health outcomes. Patient Educ Couns 2008;71(3):356-364. [doi: 10.1016/j.pec.2008.02.022] [Medline: 18372142]
34. Wand APF, Zhong B, Chiu HFK, Draper B, De Leo D. COVID-19: the implications for suicide in older adults. Int Psychogeriatr 2020;32(10):1225-1230 [FREE Full text] [doi: 10.1017/S1041610220000770] [Medline: 32349837]
35. Rademakers J, Delnoij D, de Boer D. Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? BMJ Qual Saf 2011;20(4):326-331. [doi: 10.1136/bmjqs.2010.042358] [Medline: 21339310]
36. Fröjd C, Swenne CL, Rubertsson C, Gunningberg L, Wadensten B. Patient information and participation still in need of improvement: evaluation of patients' perceptions of quality of care. J Nurs Manage 2011;19(2):226-236. [doi: 10.1111/j.1365-2834.2010.01197.x] [Medline: 21375626]
37. Talke K, Heidenreich S. How to overcome pro-change bias: incorporating passive and active innovation resistance in innovation decision models. J Prod Innovation Manage 2013;31(5):894-907. [doi: 10.1111/jpim.12130]
38. Laukkanen T. Consumer adoption versus rejection decisions in seemingly similar service innovations: the case of the internet and mobile banking. J Bus Res 2016;69(7):2432-2439. [doi: 10.1016/j.jbusres.2016.01.013]
39. Kleijnen M, Lee N, Wetzels M. An exploration of consumer resistance to innovation and its antecedents. J Econ Psychol 2009;30(3):344-357. [doi: 10.1016/J.JOEP.2009.02.004]
40. Ashkanasy N. Emotions in organizations: a multi-level perspective. In: Multi-Level Issues in Organizational Behavior and Strategy. Leeds, England: Emerald Publishing Limited; 2023:9-54.

XSL•FO
RenderX

41. Barsade SG, Gibson DE. Why does affect matter in organizations? Acad Manage Perspect 2007;21(1):36-59. [doi: 10.5465/AMP.2007.24286163]
42. Fink G, Yolles M. Collective emotion regulation in an organisation—a plural agency with cognition and affect. J Organ Change Manage 2015;28(5):832-871 [FREE Full text]
43. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. World Health Organization. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2024-08-14]
44. Blueprint for an AI bill of rights—making automated systems work for the American people. The White House. URL: https://eric.ed.gov/?id=ED625670 [accessed 2023-05-25]
45. Proposal for a regulation of the European Parliament and of the council on harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. European Commission. 2021 Apr 21. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:52021PC0206 [accessed 2025-01-21]
46. Lazar WS, Yorke C. Watched while working: use of monitoring and AI in the workplace increases. Reuters. 2023. URL: https://www.reuters.com/legal/legalindustry/watched-while-working-use-monitoring-ai-workplace-increases-2023-04-25/ [accessed 2023-05-24]
47. Werder K, Ramesh B, Zhang S. Establishing data provenance for responsible artificial intelligence systems. ACM Trans Manage Inf Syst 2022;13(2):1-23. [doi: 10.1145/3503488]
48. Bergen G, Stevens MR, Burns ER. Falls and fall injuries among adults aged ≥65 years—United States, 2014. MMWR Morb Mortal Wkly Rep 2016;65(37):993-998 [FREE Full text] [doi: 10.15585/mmwr.mm6537a2] [Medline: 27656914]
49. Sterling DA, O'Connor JA, Bonadies J. Geriatric falls: injury severity is high and disproportionate to mechanism. J Trauma 2001;50(1):116-119. [doi: 10.1097/00005373-200101000-00021] [Medline: 11231681]
50. Pinsonneault A, Addas S, Qian C, Dakshinamoorthy V, Tamblyn R. Integrated health information technology and the quality of patient care: a natural experiment. J Manage Inf Syst 2017;34(2):457-486. [doi: 10.1080/07421222.2017.1334477]
51. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020;20(1):170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
52. AI and healthcare: a giant opportunity. Intel AI Forbes Insights. URL: https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#13593a964c68 [accessed 2019-10-11]
53. Bekbolatova M, Mayer J, Ong CW, Toma M. Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. Healthcare (Basel) 2024;12(2):125 [FREE Full text] [doi: 10.3390/healthcare12020125] [Medline: 38255014]
54. Sun W, Bocchini P, Davison BD. Applications of artificial intelligence for disaster management. Nat Hazards 2020;103(3):2631-2689. [doi: 10.1007/s11069-020-04124-3]

Abbreviations

AI: artificial intelligence

Edited by T de Azevedo Cardoso; submitted 11.08.23; peer-reviewed by PH Liao, A Barucci, N Singh; comments to author 28.06.24; revised version received 20.08.24; accepted 07.10.24; published 31.01.25.

Please cite as:
Werder K, Cao L, Park EH, Ramesh B
Why AI Monitoring Faces Resistance and What Healthcare Organizations Can Do About It: An Emotion-Based Perspective
J Med Internet Res 2025;27:e51785
URL: https://www.jmir.org/2025/1/e51785
doi:10.2196/51785
PMID:

©Karl Werder, Lan Cao, Eun Hee Park, Balasubramaniam Ramesh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Publication Counts in Context: Normalization Using Query and Reference Terms in PubMed

Julian Varghese[1]; Lucas Bickmann[1]; Timo Strünker[2]; Nina Neuhaus[2]; Frank Tüttelmann[3]; Sarah Sandmann[1]

[1]Institute of Medical Informatics, University of Münster, Münster, Germany

[2]Centre of Reproductive Medicine and Andrology, University Hospital Münster, University of Münster, Münster, Germany

[3]Centre of Medical Genetics, University of Münster, Münster, Germany

Corresponding Author:
Julian Varghese
Institute of Medical Informatics
University of Münster
Albert-Schweitzer-Campus 1
Münster, 48149
Germany
Phone: 49 02518354714
Email: julian.varghese@ukmuenster.de

Abstract

This article discusses the extensive use of publication counts as indicators of trends in the scientific activities of individual researchers, research groups, and entire disciplines. However, with the growing number of articles in general, these counts might produce false impressions among scientists. We propose a straightforward yet effective normalization method, which enables further context of publication counts by using a query and a reference term. Additionally, an open access implementation is readily available on the PubMed Normalization website.

(J Med Internet Res 2025;27:e60616) doi:10.2196/60616

KEYWORDS

publication database; science communication; citation; H-index; normalization; publication; trend; scientometrics; scholarly

Problem Definition

PubMed is a widely used literature database that provides access to a vast repository of biomedical and life sciences literature. Publication counts on PubMed can be visualized through simple plots that display the number of publications over time, providing an easy way to track trends and patterns in scientific output. Publication counts over time are often used in science communication, research applications, original articles, or reviews to illustrate trends in scientific activity in a particular field [1-4]. A search on PubMed readily generates corresponding publication counts from year to year (Figure 1). To indicate growing scientific output of individuals or even joint consortia on a particular topic, the search can be further specified by author names, research sites, or affiliations. This yearly summary plot can and probably will be used as a valuable overview for (1) individual researchers or research groups to claim the success of recent research activities and (2) for policy makers or research funding organizations for initial decision-making or judging on research applications.

However, apart from its usefulness and simplicity, the plot can generate a false impression, which the authors of this viewpoint paper would like to stress on as they are frequently experiencing this type of illustration in presentations or publications without mentioning 2 important issues.

First, it does not take into account the fact that the overall number of publications have increased disproportionally. Figure 1 illustrates the problem for 2 query terms “digital health” and “boring” with a similar progression of publication counts.

The second issue is the potential textual growth per article over time, for example, by increasing the abstract’s length or further searchable metadata per article. For instance, there is evidence that the abstract length of articles has increased over time in Cochrane Reviews [5], and some articles may even exceed the abstract length that is allowed by a medical journal [6]. We are dealing with an increase in not only article counts but also textual growth per article, illustrated in Figure 2A by using the search term *boring*. Textual growth per article increases the base probability of a query term to be found and thus artificially increases publication counts per year.

XSL•FO
RenderX

While the available advanced search functions on PubMed allow for more targeted approaches, they do not provide a normalization of publication counts and their visualization, which is essential for accurately interpreting trends in the context of overall publication growth over time.

Figure 1. Original plot of PubMed counts per year, executed on May 2, 2024. Search terms: “digital health” (left) and “boring” (right).

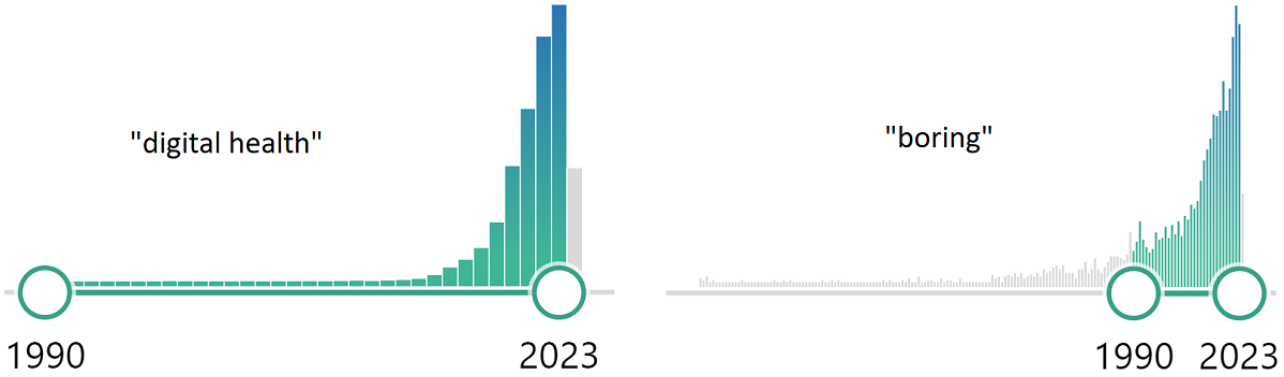

Figure 2. Article counts for the query term boring. (A) A simple correction when dividing the number of articles per year by the number all PubMed articles per year still shows an increase (red graph). (B) Correction applied by using a reference term, for example, science" or result. See Multimedia Appendix 1 for larger-resolution images.

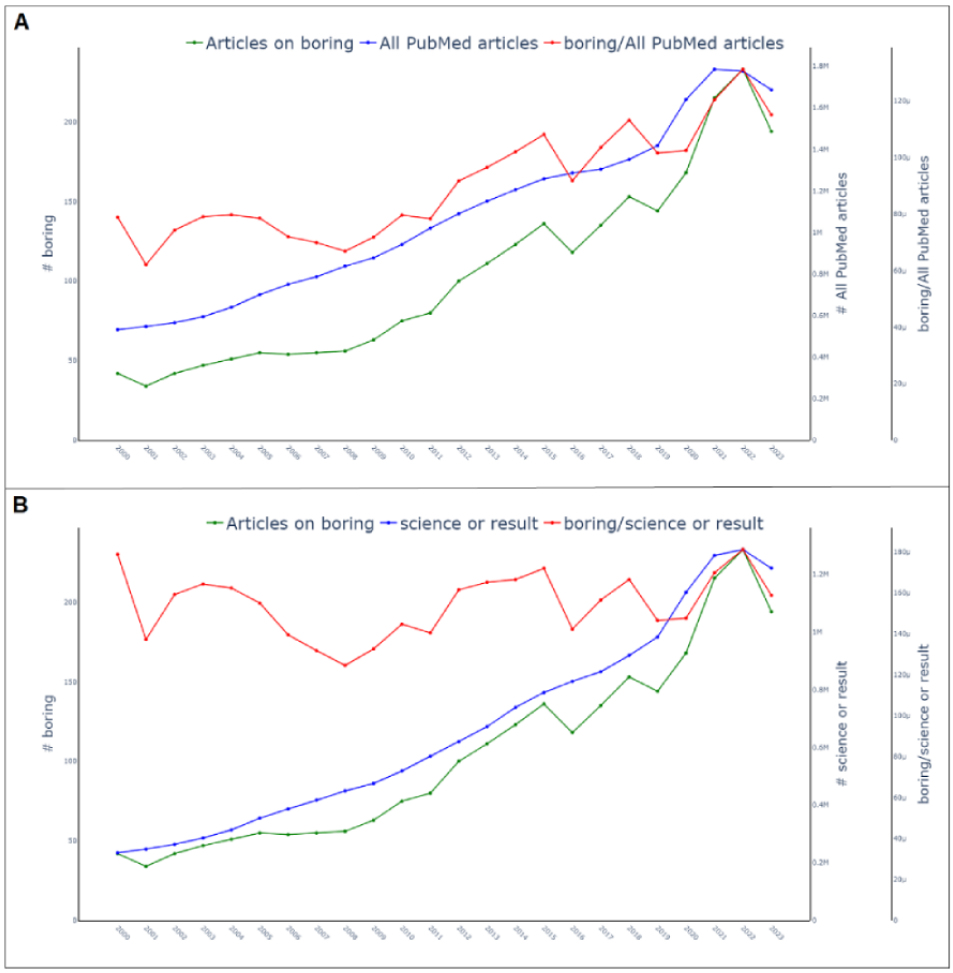

Pragmatic Solutions

Certainly, there are numerous correction measures involving advanced statistical and semantic analyses, which could add deeper insights for a more differentiated way of interpreting search results. Moreover, commercial tools are available, such as SciVal by Elsevier or Web of Science and Insight by Clarivate, which use more comprehensive datasets from larger literature databases and citation data to generate individual research portfolios or analyze publication trends. Here, we propose a noncommercial, pragmatic approach, which does not

XSL•FO
RenderX

overwhelm the user with statistical details but retains the simplicity of the original approach as illustrated in Figure 1.

This issue of the overall increasing number of publications can be tackled by dividing the counts by the number of all PubMed publications within the corresponding year. This straightforward correction is already applied in implementations such as the web-based tool PubMed by Year [7], which also offers the user to query and compare publication counts of other search queries. This does, however, not solve the issue regarding textual growth per article. Figure 2A shows that counts over time for the query term *boring* are increasing despite the aforementioned correction.

To overcome both issues, we suggest calculating and visualizing a normalization ratio, in which the counts of a query term are divided by the counts of a reference term. Both the query and the reference terms can be specified by the user, for instance by using a basic set of frequently used unspecific terms, such as *research*, *science*, or *results*. These reference terms provide a more robust correction because they are also impacted by the 2 aforementioned issues, leading to an adjustment where errors cancel out. Users have the flexibility to modify the reference term to better suit their specific needs or areas of interest.

Figure 2B shows an example plot that presents the normalization ratio. The search term *boring* now shows a steady publication count with minor variations per year. Figure 3 provides complementary examples for the search terms *Reproduction*, showing an established field with a decreasing normalization ratio; *Artificial Intelligence*, showing a continuously growing field; and *COVID-19*, showing a short-term steep increase followed by a decrease as the result of a pandemic.

A ready-to-use tool that calculates and visualizes the corrective normalization ratio for any user-specified PubMed query term is provided on the PubMed Normalization website or as a Docker container with open-source code [8].

Figure 3. While the query terms reproduction (A) and artificial intelligence (B) both show an increase in the number of publications per year (green line), the normalization ratios (red line) reveal a steady decrease for reproduction (A). For artificial intelligence (B), the red line indicates steady growth since 2016. After 2022, the growth rate slowed down. For the query term Covid 19 (C) the green and the red lines are almost perfectly aligned, indicating a steep increase followed by a steep decrease in the number of absolute publications and relative publication counts. See Multimedia Appendix 2 for larger-resolution images.

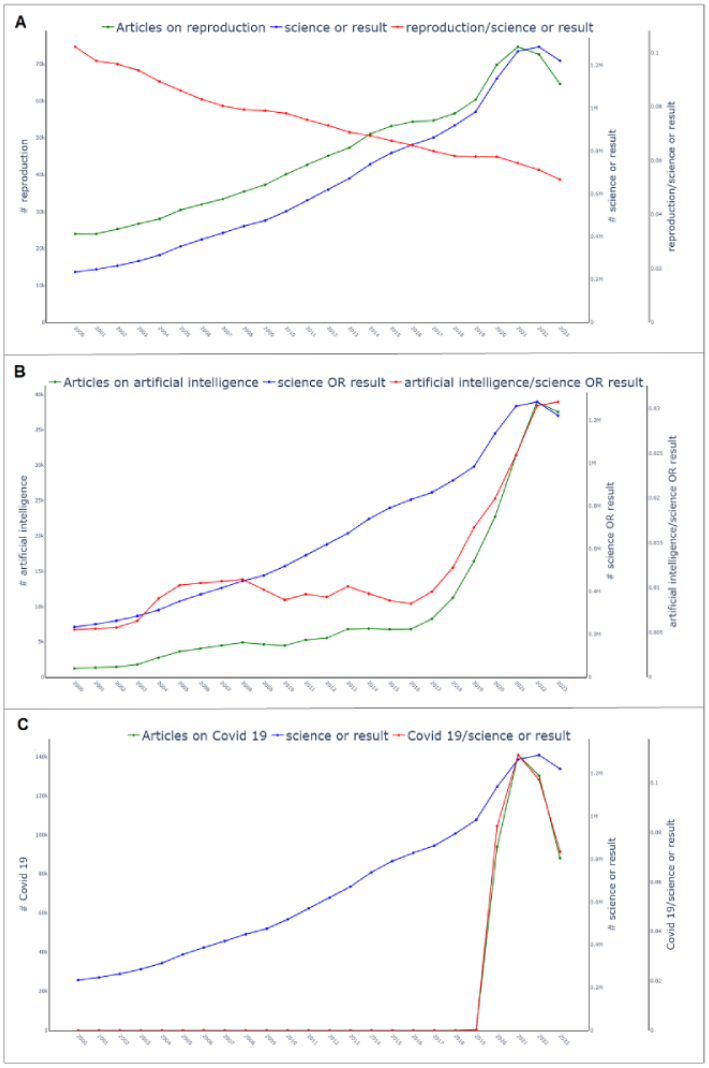

XSL•FO
RenderX

Discussion

PubMed search counts can generate a valuable overview of research trends but should be treated with caution as they represent nonnormalized and therefore potentially misleading output. The main advantage is its simplicity regarding usage and initial interpretation. For more insightful interpretation, we have pointed out 2 issues, namely the increasing article count and the textual growth per article. With this viewpoint paper, we would like to raise awareness when interpreting such publication counts. The rising number of articles published each year can lead to an overwhelming amount of data, making it difficult to identify meaningful trends without further statistical analyses and domain-specific knowledge.

We suggest adding more context to publication counts by using a normalization ratio graph that uses a reference term to account for both aforementioned issues. All parameters can be specified by the user to account for subject-related queries or reference terms. Moreover, PubMed's advanced search function is integrated as well, as our ready-to-use demonstrator supports the original PubMed queries as an input. It should be noted that this analysis cannot replace deeper analyses that address well-known drawbacks from literature searches, such as publication and accessibility bias, language bias, field-specific variations regarding publishing practices, or query and reference term selection. Nevertheless, we believe that a combined graph provides the same simplicity as the naïve search and provides more robust results when presenting publication counts and comparing them to the general research output.

In our search for related bibliometric analyses tools on the web, we have found numerous and highly valuable commercial tools such as SciVal by Elsevier or InCites by Clarivate, or open-source tools like bibliometrix [9] and PubMed by Year [7]. While these come with extensive analyses features to analyze individual or network-based research output and collaboration patterns, not only on PubMed but also on many different literature databases, we could not find a free and readily available web-based tool with a specific way to normalize and instantly visualize PubMed counts against a reference term. For quick demonstration, we have provided an exemplary implementation, which is accessible on a sustained web-based infrastructure. We encourage the scientific community to view the proposed method as a foundational tool, designed to address the key limitations of current article count presentations. While this approach offers a basic correction, we advocate for further customization to meet the specific demands of diverse research landscapes.

Acknowledgments

We acknowledge support from the Open Access Publication Fund of the University of Münster.

Authors' Contributions

Conceptualization: JV, TS, NN, and FT
Formal analysis: JV and SS
Software development: JV and LB
Writing—original draft: JV
Writing—review and editing: JV, LB, TS, NN, FT, and SS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Article counts for the query term boring. (A) A simple correction when dividing the number of articles per year by the number all PubMed articles per year still shows an increase (red line). (B) Correction applied by using a reference term, for example, science or result.
[PNG File , 212 KB - jmir_v27i1e60616_app1.png]

Multimedia Appendix 2
While the query terms reproduction (A) and artificial intelligence (B) both show an increase in publications per year (green lines), the normalization ratios (red lines) reveal a steady decrease for reproduction (A). For artificial intelligence (B), the red line indicates steady growth since 2016. After 2022, the growth rate slowed down. For the query term Covid 19 (C) the green and the red lines are almost perfectly aligned indicating a steep increase followed by a steep decrease of absolute publication and relative publication counts.
[PNG File , 304 KB - jmir_v27i1e60616_app2.png]

References

1. Pokrovac I, Pezer Ž. Recent advances and current challenges in population genomics of structural variation in animals and plants. Front Genet 2022;13:1060898 [FREE Full text] [doi: 10.3389/fgene.2022.1060898] [Medline: 36523759]

XSL•FO
RenderX

2. Winterton A, Rødevand L, Westlye LT, Steen NE, Andreassen OA, Quintana DS. Associations of loneliness and social isolation with cardiovascular and metabolic health: a systematic review and meta-analysis protocol. Syst Rev 2020 May 04;9(1):102 [FREE Full text] [doi: 10.1186/s13643-020-01369-8] [Medline: 32366295]
3. Wang K, Feng C, Li M, Pei Q, Li Y, Zhu H, et al. A bibliometric analysis of 23,492 publications on rectal cancer by machine learning: basic medical research is needed. Therap Adv Gastroenterol 2020 Jul 27;13:1756284820934594 [FREE Full text] [doi: 10.1177/1756284820934594] [Medline: 32782478]
4. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D deep learning on medical images: a review. Sensors (Basel) 2020 Sep 07;20(18):5097 [FREE Full text] [doi: 10.3390/s20185097] [Medline: 32906819]
5. Helbach J, Wandscher K, Pieper D, Hoffmann F. Cochrane review abstracts are getting longer, but this has no large impact on the reporting quality. J Evid Based Med 2023 Sep 06;16(3):294-302. [doi: 10.1111/jebm.12545] [Medline: 37674307]
6. Silverberg O, Ray JG. Variations in instructed vs. published word counts in top five medical journals. J Gen Intern Med 2018 Jan 2;33(1):16-18 [FREE Full text] [doi: 10.1007/s11606-017-4196-6] [Medline: 28971304]
7. Sperr E. PubMed by Year Internet. 2023. URL: https://esperr.github.io/pubmed-by-year/ [accessed 2023-11-17]
8. PubMed Normalization. URL: https://pmcounts.uni-muenster.de/ [accessed 2024-05-16]
9. Aria M, Cuccurullo C. bibliometrix : An R-tool for comprehensive science mapping analysis. J Informetr 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]

Edited by A Coristine; submitted 16.05.24; peer-reviewed by U Hübner, F Xu, OP Pinchuk, J Liu, A Barajas-Ochoa; comments to author 30.09.24; revised version received 23.10.24; accepted 26.11.24; published 03.02.25.

Please cite as:
Varghese J, Bickmann L, Strünker T, Neuhaus N, Tüttelmann F, Sandmann S
Publication Counts in Context: Normalization Using Query and Reference Terms in PubMed
J Med Internet Res 2025;27:e60616
URL: https://www.jmir.org/2025/1/e60616
doi:10.2196/60616
PMID:

©Julian Varghese, Lucas Bickmann, Timo Strünker, Nina Neuhaus, Frank Tüttelmann, Sarah Sandmann. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Improving the Implementation of Patient-Reported Outcome Measure in Clinical Practice: Tackling Current Challenges With Innovative Digital Communication Technologies

Kelly M de Ligt[1,2], PhD; Saar Hommes[3], PhD; Ruben D Vromans[3], PhD; Eva Boomstra[1], MSc; Lonneke V van de Poll[1,2,4], PhD; Emiel J Krahmer[3], PhD

[1]Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

[2]Department of Medical and Clinical Psychology, Center of Research on Psychological and Somatic Disorders, Tilburg University, Tilburg, Netherlands

[3]Department Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

[4]Department of Research and Development, Netherlands Comprehensive Cancer Organisation, Utrecht, Netherlands

Corresponding Author:
Kelly M de Ligt, PhD
Department of Psychosocial Research and Epidemiology
Netherlands Cancer Institute
PO Box 90203
Amsterdam, 1006 BE
Netherlands
Phone: 31 2051269111
Email: k.d.ligt@nki.nl

Abstract

Implementation of patient-reported outcome measures (PROMs) in clinical practice is challenging. We believe effective communication is key to realizing the clinical benefits of PROMs. Communication processes for PROMs in clinical practice typically involve (1) health care professionals (HCPs) inviting patients to complete PROMs, (2) patients completing PROMs, (3) HCPs and patients interpreting the resulting patient-reported outcomes (PROs), and (4) HCPs and patients using PROs for health management. Yet, communication around PROMs remains underexplored. Importantly, patients differ in their skills, knowledge, preferences, and motivations for completing PROMs, as well as in their ability and willingness to interpret and apply PROs in managing their health. Despite this, current communication practices often fail to account for these differences. This paper highlights the importance of personalized communication to make PROMs accessible to diverse populations. Personalizing communication manually is highly labor-intensive, but several digital technologies can offer a feasible solution to accommodate various patients. Despite their potential, these technologies have not yet been applied to PROMs. We explore how existing principles and tools, such as automatic data-to-text generation (including multimodal outputs like text combined with data visualizations) and conversational agents, can enable personalized communication of PROMs in practice.

(J Med Internet Res 2025;27:e60777) doi:10.2196/60777

KEYWORDS

patient reported outcome measures; quality of life; health communication; delivery of health care; digital sciences; clinical practice: patient reported outcomes; patient reported outcome; digital communication; communication; health management; digital technologies

Introduction

Patient-reported outcome measures (PROMs) are instruments that can capture patients' health status directly through digital administration [1-3]. Patients vary in their skills, knowledge, preferences, and motivations to complete PROMs, as well as their ability to interpret and use the resulting patient-reported outcomes (PROs) for health management. However, PROMs and PROs are not designed to accommodate these differences, often excluding certain populations [4]. This includes older adults, nonnative speakers, individuals with poor health, those lacking social support, people in less privileged socioeconomic positions, or those with low health literacy—defined as "the degree to which individuals can obtain, process, and understand the basic health information and services they need to make appropriate health decisions" [4-6].

XSL•FO
RenderX

Communication is crucial for achieving the clinical benefits of PROMs, for instance when health care professionals (HCPs) discuss PROs with patients. However, the communication processes involved in using PROMs in clinical practice are underexplored [7-9]. We therefore argue that improving communication processes can enhance PROMs implementation. Specifically, personalized communication (information that is customized to the individual) is essential to address the diverse information needs, preferences, and capacities of a broad population [10]. However, personalizing information manually is labor-intensive and costly, which may explain its limited application to PROMs.

To address this, we propose leveraging existing digital technologies to streamline personalized communication for PROMs. Tools such as data-to-text generation, multimodal communication, and conversational agents can offer innovative solutions to improve communication and, in turn, support the broader implementation of PROMs in clinical practice. This paper begins by summarizing communication processes for PROMs in clinical practice and identifying their shortcomings. These shortcomings highlight the limitations of a one-size-fits-all approach, which fails to meet the needs of a diverse patient population. We propose solutions rooted in personalized communication and suggest several digital technologies to support the implementation of these strategies in clinical practice.

Communication Processes for PROMs in Clinical Practice

From many available frameworks, we use the framework of Lasswell [11] to outline key aspects of communication for PROMs in clinical practice: who communicates what, in what form, to whom, and to what effect (Figure 1). We distinguish 4 communication processes, that are PROMs invitation, PROMs completion, PRO presentation, and PRO-based health management.

Figure 1. Communication processes for PROMs in clinical practice. HCPs: health care professionals; HRQoL: health-related quality of life; PROMs: patient-reported outcome measures; PROs: patient-reported outcomes.

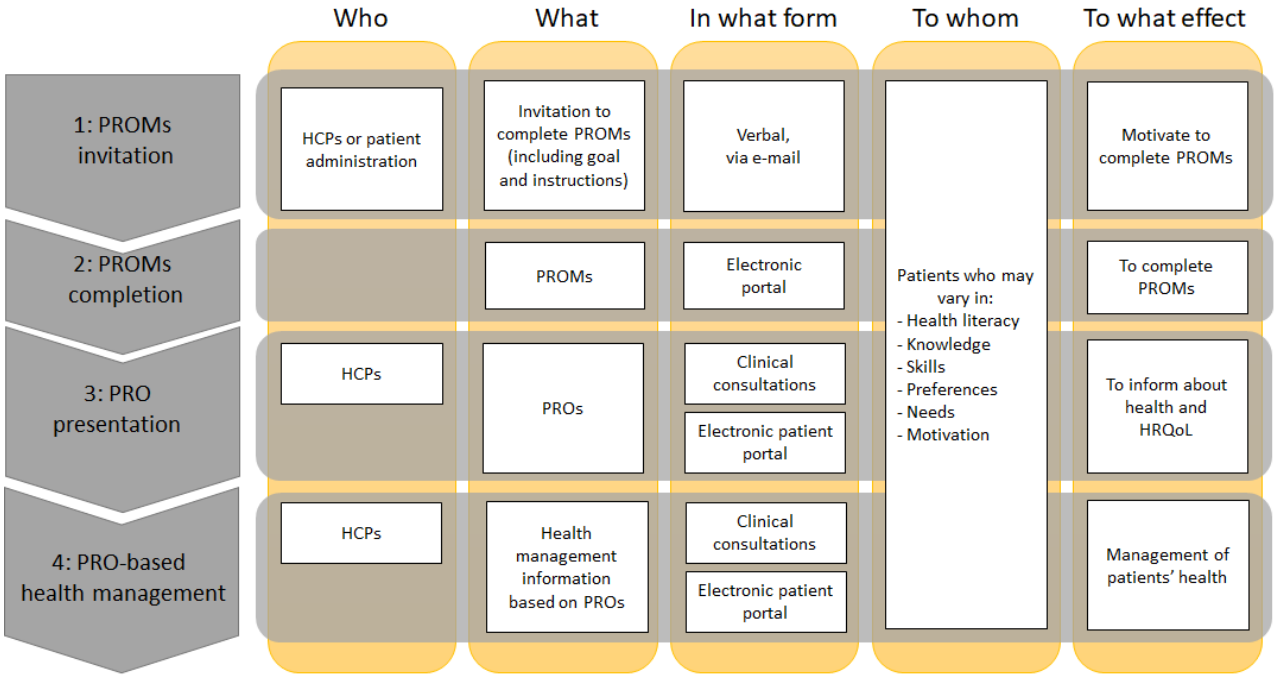

First, HCPs or administrative staff ("who") invite patients ("to whom") verbally or by email ("in what form"), including the goal and practical guidance on how to complete PROMs ("what") [11,12]. This aims to engage patients to complete PROMs ("to what effect").

Second, PROMs ("what") are administered through online portals ("in what form") for patients ("to whom") to complete ("to what effect").

Third, HCPs ("who") should consistently discuss PROs ("what") with patients ("to whom") during consultations ("in what form") [12,13], providing insights into their health and health-related quality of life (HRQoL; "to what effect") [12]. Alternatively, patients can occasionally access their PRO reports directly via electronic portals ("in what form"), bypassing HCPs.

Fourth, HCPs ("who") can translate the PROs into actionable health management information ("what") to aid patients' ("to whom") health management ("to what effect"). Similarly, patients may independently access this information ("what") through electronic portals ("in what form"), bypassing HCPs.

Shortcomings, Solutions, and Digital Technologies to Improve Communication Processes

The shortcomings of each communication process are detailed below, along with solutions based on personalized communication and digital technologies to support implementation. Table 1 summarizes the identified

XSL•FO
RenderX

shortcomings, proposed solutions, and relevant digital technologies.

Table 1. Improving communication processes around PROMs[a]: identified shortcomings, proposed solutions, and relevant digital technologies.

Identified shortcomings	Proposed solutions	Digital technologies to support implementation
1: PROMs invitation: tailoring invitations by data-to-text to accommodate more patients		
Not all patients understand the purpose of completing PROMs, when PROs[b] will be discussed, and how they benefit, resulting in varying levels of motivation to complete PROMs.	Written invitations could be tailored to accommodate different reading and health literacy levels to better serve patients with varying literacy and health literacy levels.	"Data-to-text" can tailor PROMs invitations to better serve patients with varying literacy and health literacy levels in an automated, data-driven way.
2: PROMs completion: multimodal communication principles to support PROMs completion		
Not all patients are able to complete PROMs, as they have diverse skills, knowledge, and preferences. Yet, PROMs are presented in a single, standardized format.	Presenting PROMs in other formats other than text could better serve patients with lower (health) literacy.	Multimodal communication, such as a video of an interviewer reading the questions, with written text that changes color in synch with the audiovisual playback.
Not all patients are able to complete PROMs, as they have diverse skills, knowledge, and preferences. Yet, PROMs are presented in a single, standardized format.	Allowing patients to verbally complete PROMs could better serve patients with lower literacy.	Conversational agents integrated into PROMs administration systems would allow patients to complete PROMs verbally.
3: PRO presentation: data-to-text to convert numeric PROs into comprehensible texts		
Not all patients understand PROs because of the way these are presented, and HCPs[c] frequently omits to explain PROs to patients.	Presenting PROs in a more patient-friendly way to patients.	Data-to-text applications can convert numeric PROs into written texts or narratives.
4: PRO-based health management: from PROs to health management by conversational agents		
HCPs fail to guide patients to use PROs for health management, or patients perceive health management information as not relevant as their information needs differ between patients.	Both HCPs and patients require guidance on how to use PROs for health management.	Conversational agents can break down information into smaller bits, highlight relevant PROs, and provide self-management advice. By tailoring PROs and self-management actions to individual patients, information becomes more relevant.

[a]PROMs: patient-reported outcome measures.

[b]PROs: patient-reported outcomes.

[c]HCPs: health care professionals.

PROMs Invitation: Tailoring Invitations by Data-To-Text to Accommodate More Patients

Not all patients understand the purpose of completing PROMs when PROs will be discussed, and how they will benefit, leading to varying levels of motivation to complete PROMs. Some patients may perceive PROMs as tools for research or to train HCPs [12-15]. Written invitations could be tailored to different reading and health literacy levels to better serve patients with varying literacy and health literacy levels, ensuring information about the goal and use of PROMs are clearly communicated, potentially improving completion rates.

"Data-to-text" is a digital technology with the potential for tailoring PROMs invitations to diverse audiences in a data-driven manner. A form of natural language generation (NLG), it uses insights from computational linguistics and artificial intelligence (AI) [16]. This technology automates the conversion of data into coherent natural language, typically achieved through a series of computational steps [17]. A standard data-to-text pipeline processes a patient's data, using a sequence of algorithms to determine the content and structure of the output text. By incorporating self-reported or clinically obtained information about patients' education, literacy, and health literacy levels, the system could generate texts at various readability levels to accommodate a range of reading proficiencies [18].

PROMs Completion: Multimodal Communication Principles to Support PROMs Completion

Patients have diverse skills, knowledge, and preferences for completing PROMs [4] (Table 1). While digital PROMs can be offered through multiple completion methods [19,20], and flexible administration is recommended [21], often only a single method is provided. As a result, not all patients can complete PROMs. Those facing barriers are particularly older, non-White, lower-educated patients, and those with low health literacy [4,22]. Reducing these barriers could improve completion rates in these groups.

Data-to-text could tailor PROMs to patients' reading levels, but textual PROMs may still be too lengthy, information-dense, or intimidating for some patients. According to dual coding theory [23] and the cognitive theory of multimedia learning [24],

XSL•FO
RenderX

multimodal communication (especially combining text with visuals) may help alleviate these issues [18]. These theories suggest that our working memory processes different types of information through separate channels, each with limited capacity. Using multiple channels, such as visual and verbal information, enhances information transfer. For example, integrating images with text can significantly improve comprehension, as long as the visuals and text are presented together and complement each other [24-27].

Examples of multimodal PROMs already exist [20,28-31], such as the multimedia program of Thumboo et al [28], which delivers PROMs via a touchscreen device. Each question is paired with visual and auditory stimuli – a video of an interviewer reading the question, with accompanying text that changes color in sync with the audio playback. The playback speed can be adjusted to accommodate patients with different reading levels. This approach could be extended to PROM invitations, questions, and PRO-related texts, improving accessibility and comprehension.

Furthermore, conversational agents, which emulate human conversation through text or speech [32], can be integrated into PROMs administration systems, allowing patients to complete PROMs verbally using familiar devices such as computers, smartphones, or tablets [31]. We believe this approach could be especially effective when combined with multimodal or interactive voice response systems that read questions aloud [20,31,33], making PROMs completion more accessible, especially for less literate patients. In addition, conversational agents can overcome language barriers by supporting multiple languages, promoting inclusivity for nonnative speakers [34]. Examples of this have been demonstrated by Mlakar et al [30] and Fenza et al [35], though formal evaluations of their relevance and efficiency are still needed.

PRO Presentation: Data-To-Text to Convert Numeric PROs Into Comprehensible Texts

Although patients want to reflect on their health and HRQoL before discussions with their HCP [13,14], many are unable to do so. That is, digital PROMs typically grant HCPs but not patients' access to PROs [7]. Despite this, HCPs often neglect to discuss PROs [8], and patients are generally hesitant to initiate these conversations [13]. When patient do have access to their PROs, these are typically presented in numerical or graphical formats. While patients have expressed a need for personalized numerical information [36,37] and can interpret basic line or bar graphs [19], such formats often lack context and evaluability (ie, how good or bad the scores are) [38], leading to misinterpretation [19,39]. PRO comprehension is influenced by patient's health literacy [40] and statistical literacy or numeracy skills [41]. However, despite patients' varying levels for both, PROs are typically presented in a single, fixed format [19]. We believe that improving how PROs are presented will enhance their relevance and value for patients, providing better and more equitable access to the potential benefits of PROs in patient care.

Data-to-text applications could convert numeric PRO scores into written texts, combining numerical information with verbal descriptors to enhance understanding and improve information transfer [42], as studies suggest [12,18,19,43]. We believe both patients and HCPs could benefit from this approach. When discussing individualized outcome data, it is recommended to combine numerical or visual information with verbal descriptors to better align with patients' needs and capacities [44]. While some patients and HCPs prefer verbal descriptors over "cold" numbers or graphs for discussing outcomes and HRQoL data [36,45], relying solely on words can lead to varied and inaccurate perceptions. Despite its potential, the practical use of NLG for converting PROs into coherent texts is not yet common in clinical settings. The Dutch "Data2Text" project represents an initial exploration into this promising field [46].

Patient narratives, or stories about other patients' experiences, can help patients understand the meaning of their own PROs [37,42]. When patients recognize PROMs as reflections of their health experiences, HCPs can engage in discussions about how PROs can enhance HRQoL and care [14]. As a form of data-to-text, personalized narratives could be generated based on patients' PRO scores. This approach was explored in previous research by our group, where it was found to be feasible for clinical practice. Importantly, we discovered that narratives can provide emotional support, complementing numeric data to facilitate discussions about HRQoL during clinical consultations [37]. With further development, a decision algorithm could select relevant stories from a collection of patient narratives, using a patient's PROs as input.

PROs Presentation—Based Health Management: From PROs to Health Management by Conversational Agents

In one-third of studies reviewed by Anatchkova et al [3], HCPs failed to guide patients on PRO-based health management. In addition, patients typically use health information in decision-making only when the information is relevant to them, and information needs vary between patients (Table 1) [42]. Improving patients' access to personalized health information could potentially enhance their care and support better health-related outcomes.

Both patients and many HCPs need guidance on how PROs contribute to care, and how to interpret, discuss, and act on PROs [8,13,14]. Conversational agents can break down information into smaller, manageable pieces, acting as intermediaries by highlighting relevant PROs and providing self-management advice [42,47]. Examples include conversational agents developed for asthma self-management [48] and to promote healthy eating behaviors [49]. For patient dashboards, cognitive load can be reduced by highlighting key information and enhancing evaluability [42]. When PROs are presented numerically or graphically, emphasizing the clinical significance of scores aids interpretation. This can be achieved through color-coding (eg, red for alarming and green for non-alarming scores), using exclamation marks or red circles for critical scores, and adding threshold lines to indicate scores relative to clinical thresholds [19]. We believe this approach could facilitate PRO interpretation and clinical decision-making, accelerating the use of PROs in health management. In addition, for patients, simply receiving health information is not enough — it must be interesting and relevant to them, which varies

XSL•FO
RenderX

between patients [42]. Based on patients' predefined goals and preferences, conversational agents can tailor PRO scores and self-management actions to meet their specific information needs.

Implementation of Technology-Assisted Communication Processes Around PROMs

With increasing pressure on health care, digital strategies for patient self-management are expected to play a larger role [50,51]. Paradoxically, those who would benefit the most from enhanced self-management support (such as vulnerable populations) are often the ones facing barriers to participation. For instance, lower health literacy is associated with worse health outcomes and healthcare service utilization [52], while lower digital health literacy correlates with lower overall survival rates in cancer patients [53]. While PROMs have the potential to improve symptom management, HRQoL, patient-HCP communication, and patient satisfaction [7,12,13,54,55], vulnerable patients may currently lack access to these benefits. By implementing personalized communication strategies and leveraging digital technologies, we believe PROMs can become more accessible and easier to complete, thereby increasing completion rates. Furthermore, making PROs easier to interpret can improve their usability and value for patients, ensuring that PROMs can more effectively support patient care, particularly for those currently left out.

Despite promising advances, the use of the digital applications described here is not without risks. Data-to-text algorithms, particularly those based on data-driven approaches, present certain challenges. Traditionally, data-to-text systems relied on hand-crafted rules, which were difficult to scale. However, recent advances in machine learning, driven by increased computing power, have streamlined this process, enabling scalability and efficiency [18,56]. Large language models, such as exemplified by ChatGPT, are arguably the culmination of this approach. While these models generate fluent text, they are not always accurate and may contain harmful biases. Large language models rely on massive volumes of human-authored texts, typically sourced from the internet, raising concerns about the exclusion of underrepresented populations (eg, those with a small presence on the internet) and the inclusion of biased or toxic language [57,58]. Therefore, rule-based NLG might be preferable for converting PROs into text, offering more control and reliability. Similarly, while advances in machine learning have enhanced conversational agents with increased dialogue complexity, they introduce risks in natural language understanding, response generation, and patient interpretation, necessitating rigorous monitoring and validation [59]. The safety of conversational agents for patients remains inadequately evaluated [32]. We emphasize that these digital applications require further exploration and testing, despite their clear promise.

In addition to personalized communication, patient-centered communication could improve the use of PROMs in practice [21]. Currently, patients face issues with wording, length of questionnaires, and formatting of PROMs, which can hinder their understanding of questions and answer choices [15,34,39]. Over half of the commonly used cancer-related PROMs to not follow best-practices for plain language use [60]. To improve completion rates, existing PROMs should be revised to use simple, concrete language, avoiding jargon and negations, and eliminate redundant or distracting information [15,61]. Instructions should also be simplified [61-63] and structured with clear orientation statements (eg, "First I will explain the goals of PROMs, then I will show you what the questions are about") to aid understanding [61,62]. Resources like the "health literacy universal precautions toolkit" can help assess readability and convert medical terms into plain language [63].

To reduce patient burden, PROM should be relevant for individuals [21]. Computer adaptive testing can ensure follow-up questions are tailored to each patient, rather than asking the same set for everyone. The formatting of PROMs should prioritize clarity, with large fonts, clear headings, and ample white space [15,61]. Patient-preferred modifications include removing or altering items, using visuals, simplifying language, and adjusting layouts for better comprehension [29]. Using user-centered design can help meet the diverse needs of patients, considering factors such as clinical conditions, culture, languages, and literacy levels [4]. This approach should also explicitly involve individuals with low literacy skills or learning disabilities, as they are often excluded from PROMs development [64]. By doing so, we can ensure that PROMs are accessible to a wide range of patients, including those with disabilities [20].

Training HCPs about the goals, tasks, and responsibilities related to PROMs is essential for motivating their engagement with PROMs [65,66]. In general, HCPs should discuss or, at a minimum, briefly mention PROs with patients, even when no new clinical information arises from the data [13]. HCP engagement plays a significant role in patients' motivation to complete PROMs [67]; failing to follow up on PROMs may negatively affect patients' HRQoL [68]. Further research is needed to assist HCPs in interpreting PROs and converting them into actionable steps for health management [3,13]. That said, some patients may avoid engaging with detailed PRO data [36], highlighting the need for tailoring information to meet patients' individual preferences for how they receive and process information. Although digital technologies can support communication around PROMs in clinical practice, they may not be able, now or ever, to replace the HCP in addressing patients' HRQoL.

Future Implications and Next Steps

Implementing the proposed technologies in clinical practice might present some practical hurdles. For instance, many of the identified applications have not been seamlessly integrated into electronic health records (EHRs) [47,69]. Future health care systems will require EHRs that can safely incorporate digital applications. However, models could be trained without sharing data from EHRs, using techniques like federated learning [70]. Furthermore, trust, which is crucial for patient adherence [71], is challenged by the perceived "black box" nature of AI models.

XSL•FO
RenderX

Therefore, it is essential to have human oversight of AI outputs and to maintain transparency about data sources and methods [59]. Both HCPs and patients should become more AI-literate, and have a basic understanding of the capabilities and limitations of AI models [70]. To ensure that readers can assess the credibility of generated texts, it is essential to explain that texts come from a computer model, and not a fellow human [57].

The successful implementation of PROMs in clinical practice involves complex interactions between care processes, technological development, and human behavior [66]. Often, PROMs implementation is led by HCPs or researchers, potentially resulting in a lack of recognition of these complex interactions. This paper was written by a multidisciplinary team with expertise in PROMs, implementation science, health communication, and digital technologies. We have provided a comprehensive overview of key communication processes around PROMs. Future work could deepen this assessment by studying the effects of context and noise in communication, which are aspects currently missing from the application of the Lasswell model [11]. With ongoing advancements in AI, which may help mitigate the risks associated with the proposed digital applications, we envision a future where personalized communication strategies and digital tools can overcome current implementation challenges, ensuring that PROMs are applied inclusively to benefit all patients.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. ChatGPT was used to check the manuscript for correct grammar and language use.

Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

Authors' Contributions

All authors have been involved in the conceptualization of the manuscript. Author KMdL has written the original draft, which has been reviewed and edited by authors SH, RDV, EB, LVvdP, and EJK. All authors have approved the submitted version of the manuscript.

Conflicts of Interest

None declared.

References

1. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research;, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
2. Giordano FA, Welzel G, Siefert V, Jahnke L, Ganslandt T, Wenz F, et al. Digital follow-Up and the perspective of patient-centered care in oncology: what's the PROblem? Oncology 2020;98(6):379-385. [doi: 10.1159/000495294] [Medline: 30517946]
3. Anatchkova M, Donelson SM, Skalicky AM, McHorney CA, Jagun D, Whiteley J. Exploring the implementation of patient-reported outcome measures in cancer care: need for more real-world evidence results in the peer reviewed literature. J Patient Rep Outcomes 2018;2(1):64 [FREE Full text] [doi: 10.1186/s41687-018-0091-0] [Medline: 30588562]
4. Calvert MJ, Cruz Rivera S, Retzer A, Hughes SE, Campbell L, Molony-Oates B, et al. Patient reported outcome assessment must be inclusive and equitable. Nat Med 2022;28(6):1120-1124 [FREE Full text] [doi: 10.1038/s41591-022-01781-8] [Medline: 35513530]
5. Gebert P, Schindel D, Frick J, Schenk L, Grittner U. Characteristics and patient-reported outcomes associated with dropout in severely affected oncological patients: an exploratory study. BMC Med Res Methodol 2021;21(1):77 [FREE Full text] [doi: 10.1186/s12874-021-01259-0] [Medline: 33879087]
6. Institute of Medicine Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. Washington (DC): National Academies Press (US); 2004.
7. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2021;10(10):CD011589 [FREE Full text] [doi: 10.1002/14651858.CD011589.pub2] [Medline: 34637526]
8. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2018;2:42 [FREE Full text] [doi: 10.1186/s41687-018-0061-6] [Medline: 30294712]

XSL•FO
RenderX

9. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? Soc Sci Med 2005;60(4):833-843. [doi: 10.1016/j.socscimed.2004.06.022] [Medline: 15571900]
10. Noar SM, Grant Harrington N, Van Stee SK, Shemanski Aldrich R. Tailored health communication to change lifestyle behaviors. American Journal of Lifestyle Medicine 2010;5(2):112-122. [doi: 10.1177/1559827610387255]
11. Lasswell HD. In: Bryson L, editor. The structurefunction of communication in society. New York, NY: Harper; 1948.
12. Eriksen J, Bygholm A, Bertelsen P. The association between patient-reported outcomes (PROs) and patient participation in chronic care: a scoping review. Patient Educ Couns 2022;105(7):1852-1864 [FREE Full text] [doi: 10.1016/j.pec.2022.01.008] [Medline: 35090802]
13. van der Willik EM, Milders J, Bart JAJ, Bos WJW, van Ittersum FJ, Ten Dam MAGJ, et al. Discussing results of patient-reported outcome measures (PROMs) between patients and healthcare professionals in routine dialysis care: a qualitative study. BMJ Open 2022;12(11):e067044 [FREE Full text] [doi: 10.1136/bmjopen-2022-067044] [Medline: 36396312]
14. Lai CH, Shapiro LM, Amanatullah DF, Chou LB, Gardner MJ, Hu SS, et al. A framework to make PROMs relevant to patients: qualitative study of communication preferences of PROMs. Qual Life Res 2022;31(4):1093-1103. [doi: 10.1007/s11136-021-02972-5] [Medline: 34510335]
15. Jahagirdar D, Kroll T, Ritchie K, Wyke S. Using patient reported outcome measures in health services: a qualitative study on including people with low literacy skills and learning disabilities. BMC Health Serv Res 2012;12:431 [FREE Full text] [doi: 10.1186/1472-6963-12-431] [Medline: 23181735]
16. Reiter E, Dale R. Building applied natural language generation systems. Nat. Lang. Eng 1997;3(1):57-87. [doi: 10.1017/s1351324997001502]
17. Hommes S, van der Lee D, Clouth F, Vermunt J, Verbeek X, Krahmer E. A Personalized Data-to-Text Support Tool for Cancer Patients. 2019 Presented at: Proceedings of the 12th International Conference on Natural Language Generation, Tokyo, Japan Association for Computational Linguistics. ?52; 2019 Oct 29; Tokyo, Japan p. 443. [doi: 10.18653/v1/w19-8656]
18. Pauws S, Gatt A, Krahmer E, Reiter E. Making effective use of healthcare data using data-to-text technology. In: Consoli S, Reforgiato Recupero D, Petkovi M, editors. Data Science for Healthcare: Methodologies and Applications. Cham: Springer International Publishing; 2019.
19. Albers EAC, Fraterman I, Walraven I, Wilthagen E, Schagen SB, van der Ploeg IM, et al. Visualization formats of patient-reported outcome measures in clinical practice: a systematic review about preferences and interpretation accuracy. J Patient Rep Outcomes 2022;6(1):18 [FREE Full text] [doi: 10.1186/s41687-022-00424-3] [Medline: 35239055]
20. Harniss M, Amtmann D, Cook D, Johnson K. Considerations for developing interfaces for collecting patient-reported outcomes that allow the inclusion of individuals with disabilities. Med Care 2007;45(5 Suppl 1):S48-S54 [FREE Full text] [doi: 10.1097/01.mlr.0000250822.41093.ca] [Medline: 17443119]
21. Aiyegbusi OL, Cruz Rivera S, Roydhouse J, Kamudoni P, Alder Y, Anderson N, et al. Recommendations to address respondent burden associated with patient-reported outcome assessment. Nat Med 2024;30(3):650-659. [doi: 10.1038/s41591-024-02827-9] [Medline: 38424214]
22. Pugh SL, Rodgers JP, Yeager KA, Chen RC, Movsas B, Bonanni R, et al. Characteristics of participation in patient-reported outcomes and electronic data capture components of NRG oncology clinical trials. Int J Radiat Oncol Biol Phys 2020;108(4):950-959 [FREE Full text] [doi: 10.1016/j.ijrobp.2020.06.030] [Medline: 32590048]
23. Paivio A. Dual coding theory: retrospect and current status. Canadian Journal of Psychology / Revue canadienne de psychologie 1991;45(3):255-287. [doi: 10.1037/h0084295]
24. Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. Cambridge: Cambridge University Press; 2005.
25. Schnotz W. An integrated model of textpicture comprehension. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. Cambridge: Cambridge University Press; 2005.
26. Chandler P, Sweller J. Cognitive load theory and the format of instruction. Cognition & Instruction 1991;8(4):293-332.
27. Pusic MW, Ching K, Yin HS, Kessler D. Seven practical principles for improving patient educationvidence-based ideas from cognition science. Paediatr Child Health 2014;19:119-122. [doi: 10.1093/pch/19.3.119]
28. Thumboo J, Wee HL, Cheung YB, Machin D, Luo N, Fong KY. Development of a smiling touchscreen multimedia program for HRQoL assessment in subjects with varying levels of literacy. Value Health 2006;9(5):312-319 [FREE Full text] [doi: 10.1111/j.1524-4733.2006.00120.x] [Medline: 16961549]
29. Milte R, Jemere D, Lay K, Hutchinson C, Thomas J, Murray J, et al. A scoping review of the use of visual tools and adapted easy-read approaches in quality-of-Life instruments for adults. Qual Life Res 2023;32(12):3291-3308 [FREE Full text] [doi: 10.1007/s11136-023-03450-w] [Medline: 37344727]
30. Mlakar I, Šafran V, Hari D, Rojc M, Alankuş G, Pérez LR, et al. Multilingual conversational systems to drive the collection of patient-reported outcomes and integration into clinical workflows. Symmetry 2021;13(7):1187. [doi: 10.3390/sym13071187]

XSL•FO
RenderX

31. Golden AH, Gabriel MH, Russo J, Price M, Ruhmel S, Nilsson A, et al. Let's talk about it: an exploration of the comparative use of three different digital platforms to gather patient-reported outcome measures. J Patient Rep Outcomes 2023;7(1):1187. [doi: 10.1186/s41687-023-00666-9]
32. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
33. Hahn EA, Cella D, Dobrez D, Shiomoto G, Marcus E, Taylor SG, et al. The talking touchscreen: a new approach to outcomes assessment in low literacy. Psychooncology 2004;13(2):86-95. [doi: 10.1002/pon.719] [Medline: 14872527]
34. Kilbridge KL, Fraser G, Krahn M, Nelson EM, Conaway M, Bashore R, et al. Lack of comprehension of common prostate cancer terms in an underserved population. J Clin Oncol 2009;27(12):2015-2021 [FREE Full text] [doi: 10.1200/JCO.2008.17.3468] [Medline: 19307512]
35. Fenza G, Orciuoli F, Peduto A, Postiglione A. Healthcare Conversational Agents: Chatbot for Improving Patient-Reported Outcomes. Cham: Springer International Publishing; 2023.
36. Vromans RD, Hommes S, Clouth FJ, Lo-Fo-Wong DNN, Verbeek XAAM, van de Poll-Franse L, et al. Need for numbers: assessing cancer survivors' needs for personalized and generic statistical information. BMC Med Inform Decis Mak 2022;22(1):260 [FREE Full text] [doi: 10.1186/s12911-022-02005-2] [Medline: 36199092]
37. Boomstra E, Hommes S, Vromans RD, van der Burg S, Schrijver AM, Wouters MWJM, et al. "Numbers call for action, personalized narratives provide support and recognition": a qualitative assessment of cancer patients' perspectives on patient-reported outcome measures (PROMs) feedback with narratives. J Cancer Surviv 2024. [doi: 10.1007/s11764-024-01663-7] [Medline: 39320669]
38. Zikmund-Fisher BJ. Helping people know whether measurements have good or bad implications: increasing the evaluability of health and science data communications. Policy Insights from the Behavioral and Brain Sciences 2019;6(1):29-37.
39. Grossman LV, Feiner SK, Mitchell EG, Masterson Creber RM. Leveraging patient-reported outcomes using data visualization. Appl Clin Inform 2018;9(3):565-575 [FREE Full text] [doi: 10.1055/s-0038-1667041] [Medline: 30068012]
40. Turchioe MR, Myers A, Isaac S, Baik D, Grossman LV, Ancker JS, et al. A systematic review of patient-facing visualizations of personal health data. Appl Clin Inform 2019;10(4):751-770 [FREE Full text] [doi: 10.1055/s-0039-1697592] [Medline: 31597182]
41. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest 2007;8(2):53-96. [doi: 10.1111/j.1539-6053.2008.00033.x] [Medline: 26161749]
42. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
43. van der Horst DEM, van Uden-Kraan CF, Parent E, Bart JAJ, Waverijn G, Verberk-Jonkers IJAM, et al. Optimizing the use of patients' individual outcome information - development and usability tests of a chronic kidney disease dashboard. Int J Med Inform 2022;166:104838 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104838] [Medline: 35940044]
44. Vromans RD, van Eenbergen MC, Geleijnse G, Pauws S, van de Poll-Franse LV, Krahmer EJ. Exploring cancer survivor needs and preferences for communicating personalized cancer statistics from registry data: qualitative multimethod study. JMIR Cancer 2021;7(4):e25659 [FREE Full text] [doi: 10.2196/25659] [Medline: 34694237]
45. Vromans RD, Tillier CN, Pauws SC, van der Poel HG, van de Poll-Franse LV, Krahmer EJ. Communication, perception, and use of personalized side-effect risks in prostate cancer treatment-decision making: an observational and interview study. Patient Educ Couns 2022;105(8):2731-2739 [FREE Full text] [doi: 10.1016/j.pec.2022.04.017] [Medline: 35534301]
46. Krahmer EJ, Clouth FJ, Hommes S, Vromans RD, Pauws SC, Vermunt JK. Helping cancer patients to choose the best treatment: towards automated data-driven and personalized information presentation of cancer treatment options. In Commit2Data. Open Access Series in Informatics (OASIcs) 2024;124:1-20. [doi: 10.4230/OASIcs.Commit2Data.3]
47. Lau-Min KS, Marini J, Shah NK, Pucci D, Blauch AN, Cambareri C, et al. Pilot study of a mobile phone chatbot for medication adherence and toxicity management among patients with GI cancers on capecitabine. JCO Oncol Pract 2024;20(4):483-490. [doi: 10.1200/op.23.00365]
48. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. kBot: knowledge-enabled personalized chatbot for asthma self-management. 2019 Presented at: IEEE International Conference on Smart Computing (SMARTCOMP); 2019 June 2-15; Washington, DC, USA. [doi: 10.1109/smartcomp.2019.00043]
49. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. 2017 Presented at: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; Barcelona, Spain: Association for Computing Machinery; 2017 May 23 - 26,; USA. [doi: 10.1145/3154862.3154914]
50. Chan RJ, Howell D, Lustberg MB, Mustian K, Koczwara B, Ng CC, et al. Advances and future directions in the use of mobile health in supportive cancer care: proceedings of the 2019 MASCC Annual Meeting symposium. Support Care Cancer 2020;28(9):4059-4067. [doi: 10.1007/s00520-020-05513-x] [Medline: 32405966]
51. Aminabee S. The future of healthcare Patient-centric care: digital innovations, trends, and predictions. In: Garcia MB, de Almeida RPP, editors. Emerging Technologies for Health Literacy Medical Practice. Hershey PA USA: IGI Global; 2024.

XSL•FO
RenderX

52. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
53. Heudel PE, Delrieu L, Dumas E, Crochet H, Hodroj K, Charrier I, et al. Impact of limited E-Health literacy on the overall survival of patients with cancer. JCO Clin Cancer Inform 2022;6:e2100174 [FREE Full text] [doi: 10.1200/CCI.21.00174] [Medline: 35213209]
54. Campbell R, Ju A, King MT, Rutherford C. Perceived benefits and limitations of using patient-reported outcome measures in clinical practice with individual patients: a systematic review of qualitative studies. Qual Life Res 2022;31(6):1597-1620. [doi: 10.1007/s11136-021-03003-z] [Medline: 34580822]
55. Graupner C, Kimman ML, Mul S, Slok AHM, Claessens D, Kleijnen J, et al. Patient outcomes, patient experiences and process indicators associated with the routine use of patient-reported outcome measures (PROMs) in cancer care: a systematic review. Support Care Cancer 2021;29(2):573-593 [FREE Full text] [doi: 10.1007/s00520-020-05695-4] [Medline: 32875373]
56. Gatt A, Krahmer E. Survey of the state of the art in natural language generation: core tasks, applications and evaluation. jair 2018;61:65-170. [doi: 10.1613/jair.5477]
57. Bender E, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big. 2021 Presented at: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency; Virtual Event, Canada:; 2021 March 1; Canada. [doi: 10.1145/3442188.3445922]
58. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting C. ChatGPT: five priorities for research. Nature 2023;614(7947):224-226. [doi: 10.1038/d41586-023-00288-7] [Medline: 36737653]
59. Smiley C, Schilder F, Plachouras V, Leidner J. Say the right thing rightthics issues in natural language generation systems. 2017 Presented at: Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. ?8; 2017 April 4; Valencia Spain. [doi: 10.18653/v1/w17-1613]
60. Papadakos JK, Charow RC, Papadakos CJ, Moody LJ, Giuliani ME. Evaluating cancer patient-reported outcome measures: readability and implications for clinical use. Cancer 2019;125(8):1350-1356 [FREE Full text] [doi: 10.1002/cncr.31928] [Medline: 30620401]
61. Wilson EAH, Wolf MS. Working memory and the design of health materials: a cognitive factors perspective. Patient Educ Couns 2009;74(3):318-322. [doi: 10.1016/j.pec.2008.11.005] [Medline: 19121915]
62. Weiss BD. Literacy and Patient Safety: Help Patients Understand - Manual for Clinicians. 2018;2nd edition. 2018. URL: https://psnet.ahrq.gov/issue/health-literacy-and-patient-safety-help-patients-understand-manual-clinicians-2nd-ed [accessed 2015-09-02]
63. DeWalt DA, Broucksou KA, Hawk VH, Brach C, Hink A, Rudd R, et al. Health Literacy Universal Precautions Toolkit. In: (Prepared by North Carolina Network Consortium, The Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill, under Contract No. Rockville, MD: Agency for Healthcare Research and Quality; 2010.
64. Jahagirdar D, Kroll T, Ritchie K, Wyke S. Patient-reported outcome measures for chronic obstructive pulmonary disease : the exclusion of people with low literacy skills and learning disabilities. Patient 2013;6(1):11-21 [FREE Full text] [doi: 10.1007/s40271-013-0004-5] [Medline: 23417577]
65. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014;23(6):508-518 [FREE Full text] [doi: 10.1136/bmjqs-2013-002524] [Medline: 24505110]
66. Austin E, LeRouge C, Hartzler AL, Chung AE, Segal C, Lavallee DC. Opportunities and challenges to advance the use of electronic patient-reported outcomes in clinical care: a report from AMIA workshop proceedings. JAMIA Open 2019;2(4):407-410 [FREE Full text] [doi: 10.1093/jamiaopen/ooz042] [Medline: 32025635]
67. Unni E, Coles T, Lavallee DC, Freel J, Roberts N, Absolom K. Patient adherence to patient-reported outcome measure (PROM) completion in clinical care: current understanding and future recommendations. Qual Life Res 2024;33(1):281-290 [FREE Full text] [doi: 10.1007/s11136-023-03505-y] [Medline: 37695476]
68. Velikova G, Sheppard S, Campbell L, Smith A, Awad N, Selby P, et al. Randomized trial of quality-of-life measurement in oncology practice: Do oncologists need to know? Journal of Clinical Oncology 2008;26(15):9586. [doi: 10.1200/jco.2008.26.15_suppl.9586] [Medline: 27950204]
69. Giordanengo A, Årsand E, Woldaregay AZ, Bradway M, Grottland A, Hartvigsen G, et al. Design and prestudy assessment of a dashboard for presenting self-collected health data of patients with diabetes to clinicians: iterative approach and qualitative case study. JMIR Diabetes 2019;4(3):e14002 [FREE Full text] [doi: 10.2196/14002] [Medline: 31290396]
70. Meskó B. The impact of multimodal large language models on health care's future. J Med Internet Res 2023;25:e52865 [FREE Full text] [doi: 10.2196/52865] [Medline: 37917126]
71. Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. Ther Clin Risk Manag 2005;1(3):189-199 [FREE Full text] [Medline: 18360559]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
EHR: electronic health record
HCP: health care professional
HRQoL: health-related quality of life
NLG: natural language generation
PRO: patient-reported outcome
PROM: patient-reported outcome measure

Edited by N Cahill; submitted 21.05.24; peer-reviewed by BZ Stern, PD Franklin; comments to author 18.09.24; revised version received 07.11.24; accepted 18.12.24; published 05.02.25.

Please cite as:
de Ligt KM, Hommes S, Vromans RD, Boomstra E, van de Poll LV, Krahmer EJ
Improving the Implementation of Patient-Reported Outcome Measure in Clinical Practice: Tackling Current Challenges With Innovative Digital Communication Technologies
J Med Internet Res 2025;27:e60777
URL: https://www.jmir.org/2025/1/e60777
doi:10.2196/60777
PMID:

©Kelly M de Ligt, Saar Hommes, Ruben D Vromans, Eva Boomstra, Lonneke V van de Poll, Emiel J Krahmer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

A Hierarchical Framework for Selecting Reference Measures for the Analytical Validation of Sensor-Based Digital Health Technologies

Jessie P Bakker[1], PhD; Samantha J McClenahan[1], PhD; Piper Fromy[1], PhD; Simon Turner[1], PhD; Barry T Peterson[1], PhD; Benjamin Vandendriessche[1,2], PhD; Jennifer C Goldsack[1], MSc, MA, MBA

[1]Digital Medicine Society, Boston, MA, United States

[2]Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Benjamin Vandendriessche, PhD
Digital Medicine Society
90 Canal Street
Boston, MA, 02114
United States
Phone: 1 7652343463
Email: benjamin.vandendriessche@dimesociety.org

Abstract

Sensor-based digital health technologies (sDHTs) are increasingly used to support scientific and clinical decision-making. The digital clinical measures they generate offer enormous benefits, including providing more patient-relevant data, improving patient access, reducing costs, and driving inclusion across health care ecosystems. Scientific best practices and regulatory guidance now provide clear direction to investigators seeking to evaluate sDHTs for use in different contexts. However, the quality of the evidence reported for analytical validation of sDHTs—evaluation of algorithms converting sample-level sensor data into a measure that is clinically interpretable—is inconsistent and too often insufficient to support a particular digital measure as fit-for-purpose. We propose a hierarchical framework to address challenges related to selecting the most appropriate reference measure for conducting analytical validation and codify best practices and an approach that will help capture the greatest value of sDHTs for public health, patient care, and medical product development.

(J Med Internet Res 2025;27:e58956) doi:10.2196/58956

KEYWORDS

digital health technologies; analytical validation; digital medicine; reference measures; fit-for-purpose digital clinical measures

Introduction

The proliferation of sensor-based digital health technologies (sDHTs) in clinical research and health care delivery has been bolstered by the recent finalization of regulatory guidance supporting their use for remote data acquisitions in clinical investigations [1], qualification of the first sDHT-derived clinical trial endpoint by the European Medicines Agency [2], qualification of the first sDHT-derived medical device development tool by the US Food and Drug Administration (FDA) [3], and the expansion of reimbursement pathways for the use of digital clinical measures in remote patient monitoring [4-6]. For an sDHT to be used to support scientific and clinical decision-making, evaluators must conduct verification of the sensors, usability validation of the user interface (defined as all points of contact a user may have with the sDHT as a whole), analytical validation of the algorithms, and clinical validation of the measures of behavioral or physiological function generated by the sDHT in the proposed context of use. This evaluation process has been codified in the modular V3+ framework by the Digital Medicine Society (DiMe). Since the publication of V3 in 2020 [7] and its extension in 2024 [8], V3+ has become the international methodological standard for sDHT evaluation, having been adopted by dozens of organizations [9] and cited over 250 times [10], including by the European Medicines Agency [11] and FDA [12].

Analytical validation relies on the selection of an appropriate reference measure representing “truth,” against which to compare the output of the sDHT algorithm in the same or directly comparable units. This selection process can be challenging, given that the digital clinical measure of interest may relate to biological and physiological variables, symptoms

XSL•FO
RenderX

status, or functional status according to the Wilson and Cleary (1995) [13] conceptual model. In some cases, there may be multiple potential reference measures available, and in the case of a novel digital clinical measure, there will be no extant reference measure at all. We therefore recognize the need for an evidence-based framework to guide decision-making when identifying an appropriate reference measure during analytical validation. The original description of V3 emphasized that not all potential reference measures are of equal quality [7], and as such we believe that a hierarchical step-by-step approach should be adopted to prioritize those representing the highest level of scientific rigor.

We developed a preliminary version of a hierarchical framework to support selection of a reference measure for analytical validation and presented it to experts representing regulators, health technology assessment organizations, sDHT developers, clinicians, and clinical researchers during a workshop convened by the Digital Health Measurement Collaborative Community (DATAcc by DiMe) in November 2023 [14]. Our goal was to seek feedback to improve the framework and ensure uptake across the digital medicine community. Based on this feedback, we describe our hierarchical framework that closes a key gap in the science underpinning the analytical validation of sDHTs.

Description of the Hierarchical Framework

Our framework is designed to sequentially move an investigator or developer through a series of steps laid out in Figure 1 to ensure that the most rigorous reference measures applicable to their study objectives and proposed context of use are prioritized.

Figure 1. The hierarchical framework for selecting reference measures for analytical validation.

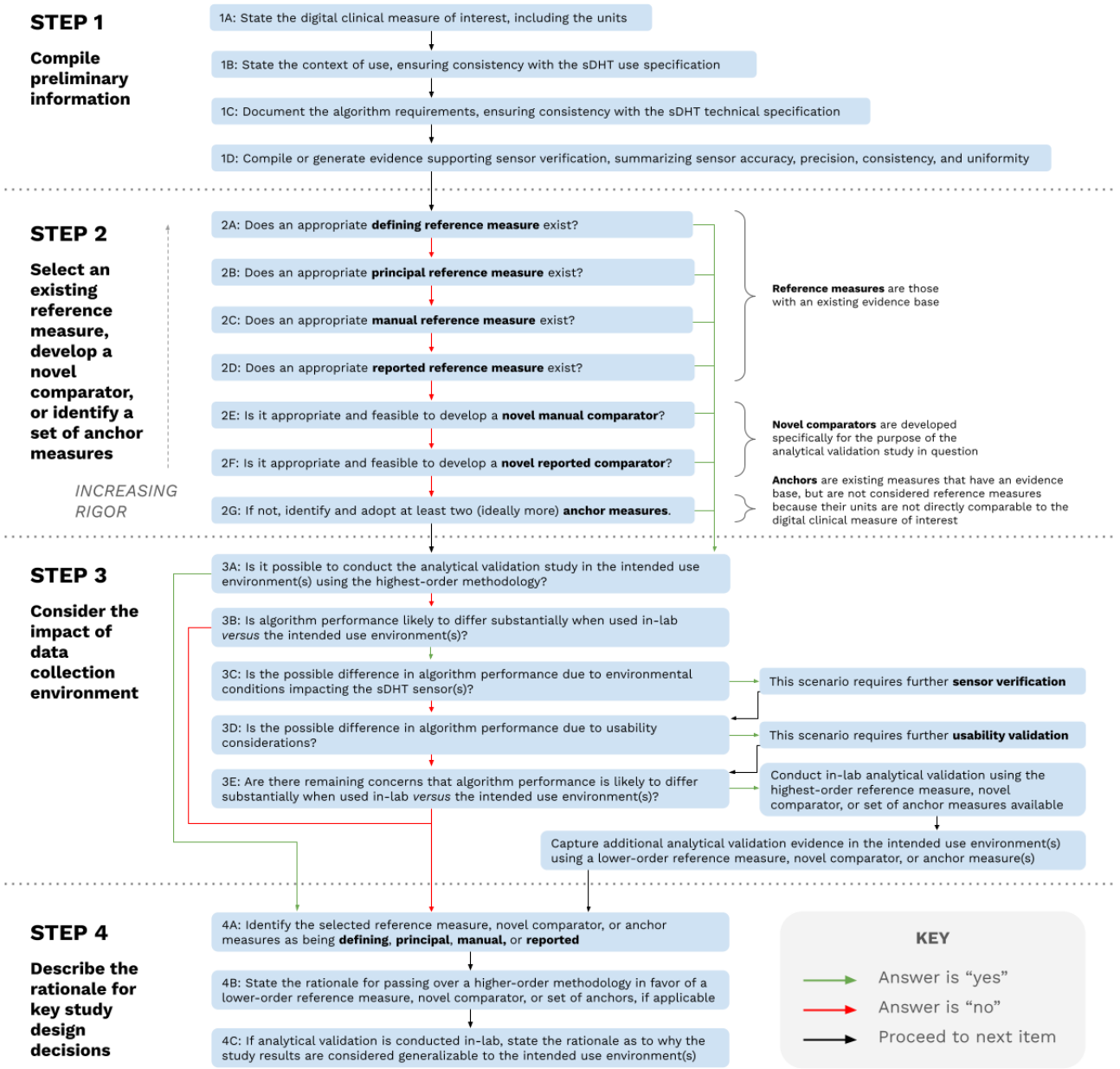

XSL•FO
RenderX

Step 1: Compile Preliminary Information

Step 1 guides the investigator to compile the preliminary information required to determine their analytical validation study objectives and guides subsequent steps. First (Step 1A), the digital clinical measure (a physiologic process or behavioral construct) is described, including the units. Next (Step 1B), the proposed context of use is clarified. The context of use fully and clearly describes the way the sDHT is to be used and the purpose of the use, including the intended populations of interest and the intended use environments. Although some sDHTs may be legally marketed medical devices with an indications for use statement compiled by the technology developer and endorsed by a regulatory body, here we are referring to the context of use proposed by the stakeholder undertaking the analytical validation study. Finally (Steps 1C-1D), the requirements and specifications of the sDHT algorithm are documented in detail, which should include any methods for dealing with missing data, alongside a summary of the verification evidence for each sensor that will feed data to the subject algorithm.

Step 2: Select an Existing Reference Measure, Develop a Novel Comparator, or Identify Multiple Anchor Measures

Step 2 of the framework classifies potential comparator methods hierarchically according to certain attributes that contribute to their scientific rigor (see Table 1). The first 4 categories represent existing reference measures, while the next 2 categories are required for scenarios in which a reference measure does not yet exist. We describe the latter as novel comparators, because their suitability as an established reference measure cannot be ascertained until they are developed and evaluated. The seventh and final category describes anchor measures, which may be adopted when it is not possible to develop a novel comparator. Unlike reference measures and novel comparators, anchor measures generate data in units that are not directly comparable to those of the digital clinical measure of interest, and therefore analysis is limited to examining associations.

XSL•FO
RenderX

Table 1. Hierarchical categorization of reference measures, novel comparators, and anchor measures for analytical validation.

Category	Definition	Attributes	Examples	Rationale for position in the hierarchy	Supporting documentation
Reference measures					
Defining	A defining reference measure emerges when a physiologic process or behavioral construct is dependent on the technology used to capture it to such an extent that it sets the medical definition for that process or construct [15]	Data capture is objective, meaning that it does not rely on human measurement, observation, or perception for acquisition. In some cases, however, human analysis or scoring may be required to generate the measure of interest. It is always possible to retain the source data	The medical definition of sleep staging refers to electrical activity of the brain (EEG), eyes (EOG), and muscles (EMG), as described by the American Academy of Sleep Medicine in the *Manual for the Scoring of Sleep and Associated Events* [16,17]. As such, polysomnography is considered the defining reference for measures of sleep staging, such as total sleep time	Defining and principal reference measures share the same attributes; however, defining references are considered superior as they will always have an associated standards document or equivalent	Standards or guidelines developed by a respected professional body describing the reference measure methodology
Principal	A principal reference measure directly and objectively measures the physiologic process or behavioral construct of interest	Data capture is objective, meaning that it does not rely on human measurement, observation, or perception for acquisition. In some cases, however, human analysis or scoring may be required to generate the measure of interest. It is always possible to retain the source data	Capnography is considered the defining reference measure for respiratory rate [18], while placement of a nasal pressure transducer, oronasal thermistry, and respiratory inductance plethysmography are considered principal reference measures, as each involves objective data acquisition and the ability to retain the source data	Principal reference measures are considered superior to manual reference measures because the objective method of data capture is not prone to observer bias. Further, the source data can be reanalyzed at will and rescored by multiple raters (where applicable), thereby promoting standardization and minimizing measurement variability	Descriptions of the reference measure from the peer-reviewed literature, with evidence of standardized implementation in many laboratories and centers. *Or* Standards or guidelines developed by a respected professional body describing the reference measure methodology (in cases where a defining reference measure already exists)
Manual	A manual reference measure relies on the measurement, observation, or perception of a physiological process or behavioral construct by a trained health care professional, with or without the use of equipment or technology	The measure of interest can be seen, heard, or felt; taste and smell have been relied on historically but are not routinely used in modern medicine. Although it may be possible to retain source data (such as a video that is then manually annotated), many manual reference measures are made in real time	Respiratory rate may be captured by auscultation or by visually assessing chest wall movement [19]	Manual reference measures are considered superior to reported reference measures because they are made by trained health care professionals which promotes standardization. In some cases it may also be possible to obtain manual measures from multiple raters, thereby reducing measurement variability	Descriptions of the reference measure from the peer-reviewed literature, with evidence of standardized implementation in many laboratories and centers. *Or* Standards or guidelines developed by a respected professional body describing the reference measure methodology (in cases where a defining reference measure already exists)

XSL•FO
RenderX

Category	Definition	Attributes	Examples	Rationale for position in the hierarchy	Supporting documentation
Reported	A patient-reported reference measure is based on a report that comes directly from a patient about the status of their health condition, while an observer-reported reference measure is based on a report from another individual based on observable signs, events, or behaviors related to a patient's health condition [20]	The identification or quantification of the measure is subjective in nature. Reports are made either in real time or retrospectively	The presence or absence, frequency, or duration of a particular experience or event is typically captured through diaries; for example, use of the Consensus Sleep Diary to capture time in bed [21]	Reported reference measures are considered inferior to manual reference measures because they typically involve a high degree of subjectivity or interpretation, and each measure can be generated only once per timepoint	Peer-reviewed evidence of psychometric performance of the instrument used for data capture
Novel comparators					
Manual	If the digital clinical measure of interest is based on a characteristic that can be observed or perceived by a trained health care professional, it may be possible to develop a novel manual comparator	A novel manual comparator shares the attributes of a manual reference measure described earlier	Manual annotation of a video captured overnight to identify measures of nocturnal scratch [22,23]	See rationale for manual reference measures described earlier	Study protocol, including all data acquisition, data processing, and scoring or annotation methods, developed *a priori*. We recommend collaborating with patient representatives when developing a definition for the characteristic of interest
Reported	If the digital clinical measure cannot be observed or perceived by a trained health care professional and is instead best captured by a report from the patient themselves or by an observer, it may be possible to develop a novel reported comparator	A novel reported comparator shares the attributes of a reported reference measure described earlier	Self-report of the time intervals for daily activities in a diary developed for the purpose of analytical validation [24]	See rationale for reported reference measures described earlier	Peer-reviewed publication, study report, white paper containing evidence supporting face validity. As described earlier, we recommend collaborating with patient representatives when developing the novel reported comparator
Anchor measures					

XSL•FO
RenderX

Category	Definition	Attributes	Examples	Rationale for position in the hierarchy	Supporting documentation
Anchors	An anchor measure is any interpretable measure of a physiologic process or behavioral construct in units that are not directly comparable to the digital clinical measure of interest. Units considered directly comparable are either identical or able to be translated for the purposes of comparison, such as via calibration	Anchor measures may possess any of the attributes listed earlier for defining or principal, manual, or reported reference measures	MDS-UPDRS Motor Assessment Item 3.15 involves a trained health care professional evaluating Parkinsonian postural tremor of the hands by visually determining tremor amplitude, returning a score of 0-4 [25].This score is not directly comparable to a digital clinical measure of tremor amplitude in units of acceleration; as such, it should be used alongside at least one more anchor measure of tremor severity, such as the Patient-Reported Outcomes in Parkinson's Disease rating scale [26]. Readers should note that in this particular example, a reference measure (3D motion capture) already exists	Anchor measures are not directly comparable to the digital clinical measure of interest, and therefore analysis is limited to examining associations. Unlike reference measures and novel comparators, anchors are not suitable for analytical validation in isolation; instead, multiple anchors should be identified in order to strengthen the assertion that the algorithm measures what it purports to measure	Supporting documents may be in the form of anything listed earlier for defining or principal, manual, or reported reference measures

The highest position of the hierarchy is assigned to *defining* reference measures (Step 2A), which are objective measures that emerge when the medical definition of a clinical measure is dependent on the technology used to capture it [15]. These measures do not rely on human observation or perception for data capture, although in some cases human analysis or scoring may be required to generate the measure of interest. Defining references are those that are widely accepted, endorsed by at least one acclaimed professional body, and described within a standards document or equivalent. For example, the American Academy of Sleep Medicine describes polysomnography (the measurement of brain, eye, and muscle activity through electroencephalography, electrooculography, and electromyography, respectively) as the defining method for capturing sleep staging data [16,17]. Similarly, the International Organization for Standardization (ISO) identifies CO-oximetry, based on light absorption of arterial blood samples typically obtained during hypoxia challenge, as the defining method for measuring blood oxygen saturation [27].

The next highest level of rigor is assigned to *principal* reference measures (Step 2B), which share the attributes of defining reference measures but typically do not have a dedicated standards document; instead, there are in-depth descriptions of the methodology in the peer-reviewed literature along with evidence of standardized implementation in many laboratories and centers. In some cases, a reference measure that does have an associated standards document may be considered a principal reference measure because a superior defining reference measure also exists. Examples of the latter scenario include transmissive and reflectance pulse oximetry; although data capture is objective and the methodology is referred to in both ISO 80601-2-61:2017 and regulatory guidance [27,28], these methods are considered inferior to CO-oximetry (the defining reference measure), because they are not capable of distinguishing functional and nonfunctional hemoglobin [29]. As this example illustrates, there may be more than one principal reference measure available to choose from—unlike defining reference measures, of which there can only be one—each of which may perform differently depending on the context of use.

Next, *manual* reference measures (Step 2C) rely on observation or perception by a trained health care professional in the absence of a source of objectively acquired data. Manual reference measures may or may not be aided by equipment or technology; for instance, respiratory rate can be measured manually through visual inspection of breaths or through auditory interpretation of auscultation using a stethoscope.

Next, *reported* reference measures (Step 2D) are based on reports that come directly from a patient (study participant) about the status of their health or a report from a lay observer such as a parent or care partner, using data capture instruments supported by appropriate psychometric evaluation. We have based our definition of reported reference measures on the BEST resource definitions of patient- and observer-reported outcomes [20]; however, it is important to note that many patient-reported outcomes and observer-reported outcomes are generated using

XSL•FO
RenderX

questionnaires that are not suitable as reference measures because they do not adopt units that are directly comparable to those of the digital clinical measure of interest. For example, sDHT-derived nocturnal scratch in units of events/hour cannot be directly compared against the Patient-Oriented Eczema Measure, which generates a score of 0-28 [30]; instead, investigators have relied on annotated videos [22]. Reported reference measures are therefore limited to reports of the presence or absence, frequency, or duration of a particular experience or event, typically captured using diaries supported by appropriate psychometric evaluation.

If no suitable reference measures exist, the investigator should consider developing a novel comparator, which should be completed and documented before beginning analytical validation. When developing a *novel manual comparator* (Step 2E), which shares the attributes of a manual reference measure, the investigator should develop a protocol for capturing and analyzing the comparator measure *a priori*, using existing technology where applicable. If it is not possible to create a novel manual comparator, the investigator should consider the development of a *novel reported comparator* (Step 2F), which shares the attributes of a reported reference measure and should be supported by evidence of face validity (at minimum). We recommend working with patient representatives when developing novel manual or reported comparators.

If it is not possible to develop a novel comparator, the only remaining option is to adopt *anchor measures* (Step 2G), which are defined as any interpretable measures of a physiologic process or behavioral construct in units that are not directly comparable to the digital clinical measure of interest. This broad definition means that an anchor measure may share attributes with defining, principal, manual, or reported reference measures. Because the data generated by an anchor measure cannot be directly compared to the digital clinical measure, it is critical to adopt multiple anchors in order to strengthen the assertion that the algorithm measures what it purports to measure. The use of multiple anchor measures for analytical validation purposes aligns with the concept of an anchor variable described by the FDA in their Patient-Focused Drug Development guidance for identifying meaningful score differences of clinical outcome assessments [31].

Step 3: Consider the Impact of the Data Collection Environment

Step 3 of the framework guides the investigator through a series of questions regarding the impact of the data collection environment. First (Step 3A), the investigator should determine whether it is possible to conduct analytical validation in the intended use environments using the highest-order reference measure, novel comparator, or anchor measures available. If so, they should plan to do so and proceed to Step 4.

If the highest-order reference measure, novel comparator, or anchor measures require the analytical validation study to be conducted in the laboratory, the investigator should determine whether the performance of the algorithm is likely to differ substantially when the sDHT is implemented in the intended use environments (Step 3B). If this is not the case and the investigator determines that algorithm performance assessed in the laboratory is generalizable to the intended use environment, they should complete the analytical validation study in the laboratory and proceed to Step 4.

Alternatively, if the investigator is concerned with generalizing analytical validation results from the laboratory setting to the intended use environments, they should first determine whether the possible difference in algorithm performance is due to environmental conditions impacting the sDHT sensors (Step 3C). If so, this scenario requires additional verification data and is not a reason to perform analytical validation in the intended use environment. Next, the investigator should determine whether the possible difference in algorithm performance is due to usability considerations (Step 3D); for example, they may be concerned that participants will place the sDHT incorrectly when using it at home unsupervised, leading to excessive signal artifact. If so, this scenario requires usability validation, and is similarly not a reason to perform analytical validation in the home environment.

Finally, if the investigator determines that the possible difference in algorithm performance is due to environmental characteristics impacting the relationship between algorithm input (sensor data) and output (the digital clinical measures of interest) (Step 3D), the investigator should proceed with in-laboratory analytical validation using the highest-order reference measure, novel comparator, or anchor measures, and plan to capture additional analytical validation evidence in the intended use environments using a lower order alternative. In other words, the investigator should first determine algorithm performance using the most rigorous methodology available and perform supplementary testing to support generalizability to the intended use environments, rather than relying solely on the lower order methodology. It is acceptable for supplementary evidence of this nature to rely on a single anchor measure, if applicable, although multiple anchors are preferable.

Step 4: Describe the Rationale for Key Study Design Decisions

The final step of the framework prompts the investigator to describe their rationale for key study design decisions by identifying their selected reference measure, novel comparator, or anchor measures according to the categories described in Step 2 of this framework (Step 4A); provide a rationale for passing over a higher order methodology in favor of their selected methodology, if applicable (Step 4B); and, if the analytical validation study is conducted in the laboratory, provide a rationale as to the generalizability of their results to the intended use environments. Step 4 is particularly important, as it encourages transparency and allows evaluators, peer-reviewers, regulators, and payers to understand the quality of the resulting analytical validation data and the extent to which it may be relied upon for decision-making.

Discussion

The Scientific Rigor of Reference Measures Is Driven by Attributes That Reduce Measurement Variability

The categories of reference measures in Step 2 are laid out in order of superiority to signify that the highest-ranked reference

XSL•FO
RenderX

measure that is available should be used whenever possible, as according to the definitions in Table 1, it is the best method for capturing the measure of interest. Our goal was to rank the categories according to attributes that contribute to reduced measurement variability.

Both *defining* and *principal* reference measures rely on objective methods of data acquisition, and the ability to retain source data means that these reference measures can be re-analyzed or re-scored at will, including by multiple health care professionals when applicable in order to create an average or consensus measure, thereby reducing measurement variability [32]. The next category, *manual* reference measures, involves human perception and is therefore prone to observer bias, but likely not to the extent of *reported* reference measures, as these typically involve a high degree of subjectivity or interpretation, and each measure can be generated only once per timepoint. Although we champion the importance of the voice of the patient during the selection of measures that matter to scientific and clinical decision-making [33-36], it is expected that the processes of usability and clinical validation will sufficiently assess these important concerns, which are out of scope for analytical validation.

We have positioned anchor measures as the lowest-order methodology, even if an anchor under consideration is well-established and widely used as illustrated by the example in Table 1 describing a comparison of sDHT-derived postural tremor (in units of acceleration) against MDS-UPDRS Motor Assessment Item 3.15 (in units of a 0-4 score). If this scale were to be adopted as the sole anchor measure, the developer would be motivated to create an sDHT algorithm that aligns as closely as possible with the clinician's visual assessment of tremor amplitude, thereby perpetuating the known limitations of this methodology. Thus, it is essential that multiple anchors be adopted for the purpose of evaluating a novel digital clinical measure, although under some circumstances it may be sufficient to adopt a single anchor if the intent is to generate supplementary evidence to support the generalizability of more robust analytical validation results from the laboratory to the intended use environments. Readers should note that comparison of a digital clinical measure against anchor measures is not a form of analytical validation given the noncomparable units but is a suitable approach to addressing this component of V3+ when reference measures do not exist and it is not possible or feasible to develop a novel comparator.

Circumstances Under Which a Lower Ranked Reference Measure May Be Appropriate for Analytical Validation of an sDHT

Our hierarchical framework will enable investigators conducting analytical validation to make the most rigorous claims associated with their products and position them to deliver the greatest value in support of scientific and clinical decision-making. Importantly, however, the hierarchical nature of Step 2 is not meant to be rigid. In some circumstances, there may be reasons for passing over a reference measure of the highest rigor to select a reference measure or novel comparator that is ranked lower. However, we posit that the only acceptable reasons for doing so are when (1) the selection of the higher ranked reference measure creates an unacceptable risk-benefit ratio, thereby compromising the ethics of the study or (2) the higher ranked reference measure is not recommended and applicable to the context of use, which includes the patient population.

Consider, for example, the analytical validation of an sDHT developed to capture blood pressure (BP). The *defining* reference measure involves placement of an arterial catheter [37], which is invasive and in some circumstances may be considered unnecessarily risky to participants. Because automated sphygmomanometry (a *principal* reference measure) is such a common and well-validated technology, it may be appropriate to use in place of arterial BP in some cases. Automatic sphygmomanometers, however, may not be suitable for use in some individuals, such as those with arrhythmias [38], in which case traditional sphygmomanometry (a *manual* reference measure) may be the most appropriate methodology. An investigator may select any reference measure or novel comparator of their choosing, with the understanding that adoption of a lower order methodology will limit the rigor of their analytical validation study.

It is rarely a requirement to select a reference measure that allows analytical validation to be conducted in remote settings, such as at home, particularly if this choice would result in the use of a reference measure that is lower in the hierarchy. However, when analytical validation is conducted in the laboratory, it is important to determine whether the performance of the algorithm is likely to differ substantially when the sDHT is implemented in its intended use environments as described in Step 3.

In summary, it is up to the investigator or developer to justify their choice by carefully considering the scientific or clinical decisions that the sDHT-derived measure is intended to support and the rigor of the analytical validation claims they wish to make (Step 4). Factors such as cost, feasibility, patient or participant preference, and resource availability are rarely sufficient reasons to select a lower order reference measure or novel comparator but might be relevant to investigators seeking to decide between reference measures within a single category.

The Quality of the Reference Measure or Novel Comparator Affects What Claims Can Be Made About the Performance of the sDHT

A key value driver of sDHTs is their ability to capture digital clinical measures that provide information that was previously difficult or impossible to capture, such as the measurement of nocturnal scratch in individuals with atopic dermatitis [39]. In such cases, available reference measures for analytical validation are likely to be ranked lower per the framework if they are available at all. The development and use of a novel comparator does not preclude the substantiation of rigorous claims associated with these measures but does require careful development of the novel comparator *a priori*. Over time, a novel comparator may "up-level" to be considered a reference measure as evidence to support its use builds over time. Manual annotation of video to identify nocturnal scratch events, even if performed by multiple trained health care professionals, can currently be considered a novel manual comparator despite

XSL•FO
RenderX

being adopted in more than one study [22,23]. Eventually, assuming this line of work continues, the field may coalesce around a video scoring protocol that allows for a high level of interrater agreement, at which point the method may be considered a manual reference measure.

Applicability of the Framework to All sDHT Regulatory Categories and All Measure Categories

As described in regulatory guidance and the literature [1,40], sDHTs used for remote data acquisition in clinical investigations may be regulated medical devices, low-risk products designed to promote a healthy lifestyle and typically marketed towards consumers (referred to by the FDA as "general wellness products" [41]), or products developed specifically for research data capture. Although the term "claims" may be used to refer to statements that fall under regulatory purview, such as medical device labeling claims, throughout this document we use "claims" to refer to any statement or conclusion made regarding the performance of an sDHT. The principles of analytical validation apply in all cases, and therefore the process of selecting a reference measure or novel comparator does not differ according to the regulatory status of the sDHT.

Similarly, the principles of analytical validation do not differ according to whether the digital clinical measure of interest is intended for use in clinical care or clinical research. If the latter, the categorization of the measure as a digital biomarker or an electronic clinical outcome assessment also has no bearing on the applicability of our framework.

Future Directions

During development of this framework, we considered the suitability of a scoring method that would allow an investigator to rank reference measures quantitatively, rather than qualitatively according to the categories described in Step 2. Ultimately, we decided against this approach due to concerns that a scoring system would be unnecessarily rigid and may quickly become outdated as technology evolves. As the field matures, however, a quantitative scoring system may be warranted.

The hierarchical framework presented here will help investigators support the most rigorous claims for both novel and established digital clinical measures and is intended to drive the capture of the greatest value of sDHTs, whether from their use in clinical trials [42] or patient care [43]. As the selection of reference measures or development of novel comparators for the analytical validation of sDHT algorithms becomes increasingly rigorous through the use of this framework, it is important to highlight a key remaining gap in the science of analytical validation: the selection of a suitable statistical technique for evaluating an algorithm's performance against the appropriate reference measure. Despite this remaining gap, the state of the science supporting the evaluation of sDHTs as fit-for-purpose is strong. This framework offers a pathway to bolster the impact of the broad adoption of best practices, such as the Measures that Matter [44] and V3+ frameworks [7,8], alongside advances in regulatory policies and decisions [1,2], placing the benefits of digital clinical measurement firmly within the reach of both researchers and clinicians seeking to improve the health-related quality of life of their patients. Ultimately, the value of our proposed framework will be determined according to its impact on improving the rigor of sDHT analytical validation science.

Acknowledgments

The authors gratefully acknowledge the contributions of the following experts through participation in the workshop and asynchronous review of the framework: Mark Ansermino, Chakib Battioui, Jen Blankenship, Joseph Cappelleri, Francesca Cerreta, Tim Chico, Yiorgos Christakis, Eric J. Daza, Charmaine Demanuele, Jan Hoff, Fay Horak, Dan Karlin, Rivka Kreitman, Joss Langford, Nicklas Linz, Sarah Lott, Kate Lyden, Martina Mancini, Ryan McGinnis, Chris Miller, Linda Ngo, Vinay Pai, Juha Pärkkä, Bray Patrick-Lake, Thanneer Malai Perumal, Sara Popham, Bohdana Ratitch, David Reasner, Colleen Russell, Leonard Sacks, Leif Simmatis, Manuel Spitschan, Samuel Stuart, Andrew Trigg, Matthew Wipperman, and Elizabeth Wragg.

Conflicts of Interest

JPB reports shares in Philips and Signifier Medical Technologies and consulting or employment income from Apnimed and Koneksa Health. SJM has no conflicts to report. PF is a contracted employee of the Digital Medicine Society and president of the consulting company SeeingTheta. ST has no conflicts to report. BTP has no conflicts to report. BV reports shares in Byteflies. JCG has no conflicts to report.

References

1. Digital health technologies for remote data acquisition in clinical investigations. Center for Drug Evaluation, Research.: U.S. Food and Drug Administration (FDA) URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations [accessed 2024-11-21]
2. Qualification opinion for stride velocity 95th centile as primary endpoint in studies in ambulatory Duchenne muscular dystrophy studies. European Medicines Agency. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-primary-endpoint-studies-ambulatory-duchenne-muscular-dystrophy-studies_en.pdf [accessed 2024-01-31]

XSL•FO
RenderX

3. FDA update: agency qualifies apple AFib history feature as an MDDT. American College of Cardiology. URL: https://www.acc.org/Latest-in-Cardiology/Articles/2024/05/02/20/02/fda-update-agency-qualifies-apple-afib-history-feature-as-an-mddt [accessed 2024-11-21]
4. Curtis JR, Willig J. Uptake of remote physiologic monitoring in the US medicare program: a serial cross-sectional analysis. JMIR Mhealth Uhealth 2023;11:e46046 [FREE Full text] [doi: 10.2196/46046] [Medline: 37040464]
5. Australian Department of Health and Aged Care. Standard search | Medicare benefits schedule. URL: https://www9.health.gov.au/mbs/search.cfm?q=remote+cardiac&Submit=&sopt=S [accessed 2024-01-31]
6. France Ministry of Health and Prevention. Reimbursability of remote monitoring. URL: https://gnius.esante.gouv.fr/en/financing/reimbursement-profiles/reimbursability-remote-monitoring [accessed 2024-01-31]
7. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digit Med 2020;3(1):55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
8. V3+: An extension to the V3 framework to ensure user-centricity and scalability of sensor-based digital health technologies. URL: https://datacc.dimesociety.org/resources/v3-an-extension-to-the-v3-framework-to-ensure-user-centricity-and-scalability-of-sensor-based-digital-health-technologies/ [accessed 2024-11-21]
9. Resources in Action - V3 Framework. Digital Medicine Society (DiMe). URL: https://dimesociety.org/resources-in-action/#V3 [accessed 2024-01-31]
10. Scholar G. Search: verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). URL: https://scholar.google.com/scholar?cluster=13684800531503945527&hl=en&as_sdt=40005&sciodt=0,10 [accessed 2024-01-31]
11. Mantua V, Arango C, Balabanov P, Butlen-Ducuing F. Digital health technologies in clinical trials for central nervous system drugs: an EU regulatory perspective. Nat Rev Drug Discov 2021;20(2):83-84. [doi: 10.1038/d41573-020-00168-z] [Medline: 32994577]
12. Sacks L, Kunkoski E. Digital health technology to measure drug efficacy in clinical trials for Parkinson's disease: a regulatory perspective. J Parkinsons Dis 2021;11(s1):S111-S115. [doi: 10.3233/jpd-202416]
13. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. J Am Med Assoc 1995;273(1):59. [doi: 10.1001/jama.1995.03520250075037]
14. DiMe - About Us. In: DATAcc by DiMe [Internet]. URL: https://datacc.dimesociety.org/about-us/ [accessed 2024-11-21]
15. Hofmann B. The technological invention of disease. Med Humanit 2001;27(1):10-19 [FREE Full text] [doi: 10.1136/mh.27.1.10] [Medline: 23670546]
16. Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, et al. The visual scoring of sleep in adults. J Clin Sleep Med 2007;03(02):121-131. [doi: 10.5664/jcsm.26814]
17. AASM Scoring Manual - American Academy of Sleep Medicine. Association for Sleep Clinicians and Researchers. URL: https://aasm.org/clinical-resources/scoring-manual/ [accessed 2024-11-21]
18. Medical electrical equipment? Part 2-55: particular requirements for basic safety and essential performance of respiratory gas monitors. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/67241.html; [accessed 2024-12-28]
19. Lim WS, Carty SM, Macfarlane JT, Anthony RE, Christian J, Dakin KS, et al. Respiratory rate measurement in adults--how reliable is it? Respir Med 2002;96(1):31-33 [FREE Full text] [doi: 10.1053/rmed.2001.1203] [Medline: 11863207]
20. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Bethesda (MD): National Institutes of Health (US); 2016.
21. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein K, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
22. Mahadevan N, Christakis Y, Di J, Bruno J, Zhang Y, Dorsey ER, et al. Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices. NPJ Digit Med 2021;4(1):42 [FREE Full text] [doi: 10.1038/s41746-021-00402-x] [Medline: 33658610]
23. Moreau A, Anderer P, Ross M, Cerny A, Almazan TH, Peterson B, et al. Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks. IEEE J Biomed Health Inform 2018;22(4):1011-1018. [doi: 10.1109/JBHI.2017.2710798] [Medline: 28613187]
24. Januschowski K, Ihmig FR, Koch T, Velten T, Rickmann A. Context-sensitive smart glasses monitoring wear position and activity for therapy compliance-a proof of concept. PLoS One 2021;16(2):e0247389 [FREE Full text] [doi: 10.1371/journal.pone.0247389] [Medline: 33606776]
25. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
26. Mischley LK, Lau RC, Weiss NS. Use of a self-rating scale of the nature and severity of symptoms in parkinson's disease (PRO-PD): correlation with quality of life and existing scales of disease severity. NPJ Parkinsons Dis 2017;3:20 [FREE Full text] [doi: 10.1038/s41531-017-0021-5] [Medline: 28649620]

XSL•FO
RenderX

27. Medical electrical equipment? Part 2-61: particular requirements for basic safety and essential performance of pulse oximeter equipment. International Organization for Standardization. URL: https://www.iso.org/standard/67963.html; [accessed 2024-12-28]
28. Pulse Oximeters - Premarket Notification Submissions 510(k)s. FDA. URL: https://www.fda.gov/media/72470/download [accessed 2024-11-21]
29. Bickler PE. In: Parsons PE, Wiener-Kronish JP, editors. Pulse Oximetry, Capnography, Blood Gas Analysis. Amsterdam: Elsevier; 2013.
30. Spuls PI, Gerbens LAA, Simpson E, Apfelbacher CJ, Chalmers JR, Thomas KS, et al. Patient-oriented eczema neasure (POEM), a core instrument to measure symptoms in clinical trials: a harmonising outcome measures for eczema (HOME) statement. Br J Dermatol 2017;176(4):979-984 [FREE Full text] [doi: 10.1111/bjd.15179] [Medline: 27858989]
31. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints For Regulatory Decision-Making. FDA. URL: https://www.fda.gov/media/166830/download [accessed 2024-11-21]
32. Bakker JP, Ross M, Cerny A, Vasko R, Shaw E, Kuna S, et al. Scoring sleep with artificial intelligence enables quantification of sleep stage ambiguity: hypnodensity based on multiple expert scorers and auto-scoring. Sleep 2023;46(2):zsac154 [FREE Full text] [doi: 10.1093/sleep/zsac154] [Medline: 35780449]
33. Engagement resources. Patient-Centered Outcomes Research Institute (PCORI). URL: https://www.pcori.org/engagement/engagement-resources [accessed 2024-01-31]
34. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med 2010;362(10):865-869. [doi: 10.1056/nejmp0911494]
35. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas 2018;9:353-367 [FREE Full text] [doi: 10.2147/PROM.S156279] [Medline: 30464666]
36. Aiyegbusi OL, Roydhouse J, Rivera SC, Kamudoni P, Schache P, Wilson R, et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. Nat Commun 2022;13(1):6026 [FREE Full text] [doi: 10.1038/s41467-022-33826-4] [Medline: 36224187]
37. Rader F, Victor RG. The slow evolution of blood pressure monitoring: but wait, not so fast!. JACC Basic Transl Sci 2017;2(6):643-645 [FREE Full text] [doi: 10.1016/j.jacbts.2017.11.001] [Medline: 30062179]
38. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, et al. Measurement of blood pressure in humans: a scientific statement from the American Heart Association. Hypertension 2019;73(5):e35-e66. [doi: 10.1161/hyp.0000000000000087]
39. Cesnakova L, Meadows K, Avey S, Barrett J, Calimlim B, Chatterjee M, et al. A patient-centred conceptual model of nocturnal scratch and its impact in atopic dermatitis: a mixed-methods study supporting the development of novel digital measurements. Skin Health Dis 2023;3(5):e262 [FREE Full text] [doi: 10.1002/ski2.262] [Medline: 37799371]
40. Bakker JP, Izmailova ES, Clement A, Hoffmann S, Leptak C, Menetski JP, et al. Regulatory pathways for qualification and acceptance of digital health technology-derived clinical trial endpoints: considerations for sponsors. Clin Pharmacol Ther 2025;117(1):56-72. [doi: 10.1002/cpt.3398] [Medline: 39148198]
41. Center for Devices, Radiological Health. General wellness: policy for low risk devices - guidance. In: U.S. Food and Drug Administration [Internet]. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-wellness-policy-low-risk-devices [accessed 2024-11-21]
42. Herrington WG, Goldsack JC, Landray MJ. Increasing the use of mobile technology-derived endpoints in clinical trials. Clin Trials 2018;15(3):313-315 [FREE Full text] [doi: 10.1177/1740774518755393] [Medline: 29400066]
43. Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, Abid MM, Péntek M, Gulácsi L, et al. Digital biomarker-based interventions: systematic review of systematic reviews. J Med Internet Res 2022;24(12):e41042 [FREE Full text] [doi: 10.2196/41042] [Medline: 36542427]
44. Manta C, Patrick-Lake B, Goldsack JC. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark 2020;4(3):69-77 [FREE Full text] [doi: 10.1159/000509725] [Medline: 33083687]

Abbreviations

BP: blood pressure
DiMe: Digital Medicine Society
FDA: Food and Drug Administration
ISO: International Organization for Standardization
sDHT: sensor-based digital health technology

XSL•FO
RenderX

Edited by A Mavragani; submitted 29.03.24; peer-reviewed by EE Buckley, M Richardson, K Wac; comments to author 04.10.24; revised version received 21.11.24; accepted 15.12.24; published 07.02.25.

Please cite as:
Bakker JP, McClenahan SJ, Fromy P, Turner S, Peterson BT, Vandendriessche B, Goldsack JC
A Hierarchical Framework for Selecting Reference Measures for the Analytical Validation of Sensor-Based Digital Health Technologies
J Med Internet Res 2025;27:e58956
URL: https://www.jmir.org/2025/1/e58956
doi:10.2196/58956
PMID:39918870

©Jessie P Bakker, Samantha J McClenahan, Piper Fromy, Simon Turner, Barry T Peterson, Benjamin Vandendriessche, Jennifer C Goldsack. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Application of Preventive Medicine in the Future Digital Health Era

Katherine De la Torre[1,2], MD, PhD; Sukhong Min[2], MD; Hyobin Lee[2,3], MS; Daehee Kang[2,3], MD, PhD

[1]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Republic of Korea

[2]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[3]Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Seoul, Republic of Korea

Corresponding Author:
Sukhong Min, MD
Department of Preventive Medicine
Seoul National University College of Medicine
103 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 2 740 8509
Fax: 82 2 747 4830
Email: sukhongmin@snu.ac.kr

Abstract

A number of seismic shifts are expected to reshape the future of medicine. The global population is rapidly aging, significantly impacting the global disease burden. Medicine is undergoing a paradigm shift, defining and diagnosing diseases at earlier stages and shifting the health care focus from treating diseases to preventing them. The application and purview of digital medicine are expected to broaden significantly. Furthermore, the COVID-19 pandemic has further accelerated the shift toward predictive, preventive, personalized, and participatory (P4) medicine, and has identified health care accessibility, affordability, and patient empowerment as core values in the future digital health era. This “left shift” toward preventive care is anticipated to redefine health care, emphasizing health promotion over disease treatment. In the future, the traditional triad of preventive medicine—primary, secondary, and tertiary prevention—will be realized with technologies such as genomics, artificial intelligence, bioengineering and wearable devices, and telemedicine. Breast cancer and diabetes serve as case studies to demonstrate how these technologies such as personalized risk assessment, artificial intelligence–assisted and app-based technologies, have been developed and commercialized to provide personalized preventive care, identifying those at a higher risk and providing instructions and interventions for healthier lifestyles and improved quality of life. Overall, preventive medicine and the use of advanced technology will hold great potential for improving health care outcomes in the future.

(J Med Internet Res 2025;27:e59165) doi:10.2196/59165

KEYWORDS

preventive medicine; personalized prevention; digital health technology; digital health; artificial intelligence; wearable devices; telemedicine

Introduction

As populations age and chronic diseases become more prevalent, societies will face challenges in providing equitable health care. Currently, a shift in the medical paradigm is ongoing, with more focus on keeping people healthier for longer and preventing diseases rather than just curing them. This approach acknowledges the importance of maintaining, improving, and promoting health throughout the natural progression of diseases and the increasing importance of preventive medicine.

In this new paradigm, cutting-edge technologies such as big data, informatics, and medical digitalization will allow for personalized and precise health care recommendations based on a patient’s clinical, biological, and genomic information. In addition, the availability of health-related information and the use of wearable devices have further encouraged individuals to take proactive measures in preventing health problems. The COVID-19 pandemic has also highlighted the importance of telemedicine in remote patient monitoring and providing more equitable health care.

XSL•FO
RenderX

Overall, preventive medicine and the use of advanced technology will hold great potential for improving health care outcomes in the future. However, the implementation of such technologies presents challenges such as addressing disparities in technology access, managing overuse and overreliance on digital tools, safeguarding patient data, and ensuring clinical validity, which call for further research in this area.

This review aims to elucidate the role of preventive medicine through the integration of digital health technologies and provide insights for the changing landscape of preventive health care in the digital era. By examining current trends, technological advancements, and case studies, we will explore how these innovations can lead to a more proactive, personalized, and patient-centered approach to health promotion and disease prevention.

Future Trend and Emerging New Technologies

Rise in Older Adult Population

The global population is aging, with people worldwide living longer lives. Nowadays, most individuals are expected to reach their 60s and beyond, a trend that is prevalent in every country. By the year 2030, approximately 1 out of every 6 people around the world is expected to be 60 years or older. The number of individuals in this age group is projected to increase from 1 billion in 2020 to 1.4 billion by 2030, and to double to 2.1 billion by 2050. Moreover, the number of individuals aged 80 years or older is anticipated to triple, reaching 426 million by 2050. Some countries such as Korea and Japan have been labeled as "super-aged societies" due to their pronounced aging trend, with 30% or more of their population being older than 65 years by 2050 [1].

This change in population composition is expected to have far-reaching health implications. The most striking is the increased burden of chronic diseases [2]. One study has projected that by 2035, 36%, and by 2050, 48% of adults older than 50 years will have at least 1 chronic disease [3]. As the global population ages and the burden of chronic diseases rises, medical costs have also increased worldwide due to longer care duration and health-related productivity loss. This has led to an annual growth of US $4 in the average Organization for Economic Cooperation and Development health expenditure per capita over the past decade [4]. In the case of the United States, 1 study has estimated that approximately 20% of health care spending growth will be attributable to aging by 2025 [5].

Shift From Cure to Prevention

There is currently a general shift of health care away from merely providing cure to that of prevention, as the definition of health transitions from "the absence of disease" to "realizing the full potential of one's capacity." The paradigm change in medicine has been attributed to the humanization movement in medicine, which has contributed to changes in medical education, patient care, and the management of several medical conditions [6]. Nowadays, there is a steady transition from "cure-seeking medical care" to "cure and support-seeking medical care" [7], which has necessitated the development of new models of conceptualizing behavior change, emphasizing assessment of multiple risks, and treating the whole patient to optimize health outcomes [8,9].

The recent change in medical paradigms has impacted various aspects of health toward a more patient-centered approach, emphasizing patient engagement, health maintenance, and the restoration of daily activity. Specifically, where the health care of the past was episodic, focused on service providers and individual patients at independent and isolated institutions, health care is expected to increasingly shift toward a continuous care model. In this model, the general public and the community as a whole come to prevent diseases and maintain their health, and, if the need for treatment or hospitalization arises, are closely followed through after their discharge [10]. The patients are also expected to be better informed, with health care workers and patients sharing a more equal footing (Table 1) [10,11].

Table 1. Current and future health care.

Categories	Current	Future
Goal	• Disease treatment	• Health promotion
Continuity of care	• Episodic cure • Fragmented care	• Continuing care • Comprehensive care
Focus of care	• Focus on service provider • Individual approach	• Focus on well-informed patient • Team-based approach
Care setting	• Hospital-centered care	• Patient-/home-oriented care
Payment system	• Pay for procedures	• Pay for value

Usage of Digital Health Technology

The convergence of life sciences and engineering disciplines has ushered in an era of medical technology integration, where genomics, artificial intelligence (AI), bioengineering, wearable devices, and telemedicine are used for various medical applications (Table 2). Advancements in technology have greatly impacted the medical field, with the development of bionic eye devices or 3D bioprinting technology used in various medical fields or wearable devices such as Google Glass have been used in surgical settings, offering benefits such as improved surgical

XSL•FO
RenderX

education, training, consultation, patient monitoring, and audiovisual recording, among others [12].

Table 2. New technologies used for future medicine.

Domain	Description	Application
Genomics	• Study of an individual's entire genome. • Revealing inherited traits and susceptibility to diseases.	• Whole-genome sequencing for personalized medicine. • Identification of genetic markers associated with diseases.
Artificial intelligence	• Mimic human intelligence based on algorithms and computational models. • Application in big data analysis, image recognition, and problem-solving.	• Machine and deep learning algorithms for medical image analysis. • Predictive models for disease diagnosis and prognosis.
Biomedical devices	• Integration of engineering principles with biological systems. • Design and development of devices to enhance medical treatments.	• Artificial organs and tissue engineering for transplantation. • Biomedical implants for drug delivery or physiological monitoring.
Wearable devices	• Collect real-time data related to health and activities. • Enhance continuous monitoring and personalized health care.	• Smartwatches monitoring sleep, ECG[a], and stress patterns. • Medical wearables for continuous glucose monitoring or ECG recording.
Telemedicine	• Delivery of health care services remotely. • Comprises video consultations, remote monitoring, and virtual care.	• Video calls for remote doctor consultations. • Remote patient monitoring for disease management.

[a]ECG: electrocardiograms.

The expansion and use of digital health is evident in daily life through mobile health applications for wellness, nutrition, medication adherence, fitness tracking, and access to personal health information. Health information technology includes electronic medical record systems and cloud storage, which improve the accessibility, sharing, and accuracy of patient data. Electronic prescriptions and communication between doctors and patients are also improved. Wearable devices, wireless sensors, and diagnostic products play a role in monitoring health and diagnostics, such as continuous blood pressure, glucose, or sleep patterns monitoring.

Hence, a key goal of digital technology in health care is digital therapeutics, which uses evidence-based therapeutic interventions software integrated with mobile devices, apps, sensors, and virtual reality to prevent, manage, and treat diseases or disorders. These can be used alone or in combination with mechanical devices, pharmacological treatments, or in-person therapy, having a positive impact because they are designed to cater to the specific needs of patients. Therefore, the customization of health services through digital technology is essential for personalized health care, including patient-reported outcomes, predictive analytics, and clinical decision support.

Advancements in internet of things sensors and health tech wearables have made it easier to collect and transmit health data. The decreasing cost of internet of things sensors, along with the improved capabilities of consumer-grade devices such as the Apple Watch, has made it more affordable to equip individuals with tools to monitor their health. For example, the Apple Watch's electrocardiogram functionality can detect irregular heartbeats and even send notifications for conditions such as atrial fibrillation. Other wearable devices, such as smart continuous glucose monitors and connected inhalers, offer benefits such as insulin-level monitoring and medication adherence. In addition, researchers are exploring the use of smartphones for "digital phenotyping" to identify individuals at risk of mental health problems [13,14]. With increased data access and sharing, these devices have the potential to provide highly personalized care based on individual patient profiles.

As digital technology continues to advance, personalized prevention and patient engagement in health monitoring and treatment management will become more feasible. Telemedicine and telehealth have become increasingly popular, especially for non–face-to-face virtual visits, remote patient monitoring, and remote care programs, which have been particularly relevant during the COVID-19 pandemic. The future of health care is heavily reliant on digital health, particularly for preventive medicine, as health information becomes more accessible to everyone and patient health empowerment increases.

Changes in Health Care After COVID-19

Telemedicine was previously used in limited cases where access to in-person health care was difficult. However, the COVID-19 pandemic highlighted its importance in addressing health care labor shortages, reducing the risk of new infections in clinics, and providing care to vulnerable populations. As a result, telemedicine usage significantly increased during the pandemic and has a more widely discussed and used tool in health care [15]. This accelerated adoption of telemedicine now includes various medical specialties, including preventive medicine.

The future of health care in the post–COVID-19 era relies on 3 key aspects: accessibility, affordability, and empowerment

XSL•FO
RenderX

(Table 3) [16]. Telemedicine has greatly improved accessibility by bringing care and remote monitoring closer to patients, regardless of their location. It has also facilitated rapid diagnoses and disease treatments, reduced costs, and prepared health care systems for future pandemic situations. For example, a report on the adoption of telehealth in Australia during the COVID-19 pandemic noted that for every in-person consultation replaced with telemedicine, up to 2.5 days of travel time and US $215 in societal productivity were saved [17]. Similar findings were echoed in the United States [18] and Canada [19].

Table 3. Key aspects of the future of health care after COVID-19.

Key aspects and implementations	Effects
Accessibility	
Telemedicine	• Bring care closer to patients • Remote monitoring
Point-of-care diagnostics	• Rapid diagnosis and treatment • Reduced cost • Quick response for future pandemic
Affordability	
Cost-effective medical devices	• Innovative, low cost, and easy-to-use portable devices
Self-health management	• Empowered health seekers • Lower the cost by avoiding hospitalization
Empowerment	
Personalized health records	• Support mass screening and epidemiological programs • Automatization for an efficient health care ecosystem
Patient engagement and education	• Health awareness with education • Disease awareness with wearables

As medical technology continues to advance, more affordable and innovative devices are becoming available in the market. There is a growing emphasis on developing portable medical devices that are cost-effective, easy to use, and accessible to everyone. The popularity of digital technology and smart wearables is steadily increasing, as they complement clinical care and provide valuable health data, such as vital signs, sleep patterns, and physical activity levels. These data can help predict disease progression and enable a shift toward preventive health care, empowering patients to actively manage their own health and reduce health care costs and limit health facilities use.

Furthermore, empowering patients in health care in the future involves leveraging electronic health records and integrated medical and digital devices to support mass screening and epidemiological programs. It also involves automating supply chains, where blockchain technology may be a vital asset, to improve efficiency across the health care ecosystem. Patient education and health-disease awareness are crucial components of the health care ecosystem and can be facilitated through web-based health content platforms and the use of smart wearables.

The advancement of medical technology, coupled with the impact of telemedicine following the COVID-19 pandemic and the shift in the medical paradigm, is set to transform the health care industry and the future of medicine. This transformation will result in a shift toward predictive, preventive, personalized, and participatory (P4) medicine, with a focus on individualized patient profiling and tailored prevention strategies. This new paradigm emphasizes patient-centered care, aims to improve overall health outcomes, and will highlight even more the importance of preventive medicine in the future of health care [20,21].

Left Shift of the Natural Progression of Disease

The field of medicine has evolved since the age of epidemics from the age of complex diseases to the current age of postgenomics [22]. In the age of epidemics, the control of pathogens was the main treatment approach highlighting the importance of public hygiene. In the age of complex diseases, disease was based on multifactorial living conditions and lifestyle factors and care was empirically derived from trial and error experience of prior practice or observations. During this era, the distinction between treatment and prevention in the field of medicine became clearer. The growth of genetics research and completion of the Human Genome Project have introduced the concept of genetic susceptibility and gene-environment interaction to the age of postgenomics. In this era, the development of personalized medicine accelerated and the definition of health and disease began to shift. The new disease paradigm is a shift in the natural progression from treatment to health promotion and prevention.

A few decades ago, the main goal of medical practitioners was to accurately diagnose and treat diseases. The focus was on the clinical stage, with the aim of finding the best treatment, ensuring quick recovery, and preventing disease recurrence. However, the medical care paradigm has since shifted to the left (Figure 1) to focus on the stage before clinical symptoms appear, where there are subclinical manifestations and

XSL•FO
RenderX

pathological changes. Medical care at this stage, that is, early detection, involves identifying pathological or structural changes that occur before symptoms manifest over the progression of the natural history of the disease. The discovery and integration of biomarkers have played a significant role in this approach. However, even at an earlier stage, some people are more likely than others to be affected by carcinogens or substances; therefore, understanding individual variations in susceptibility to risk factors has become crucial for disease prevention.

Figure 1. Left shift of natural history of disease progression.

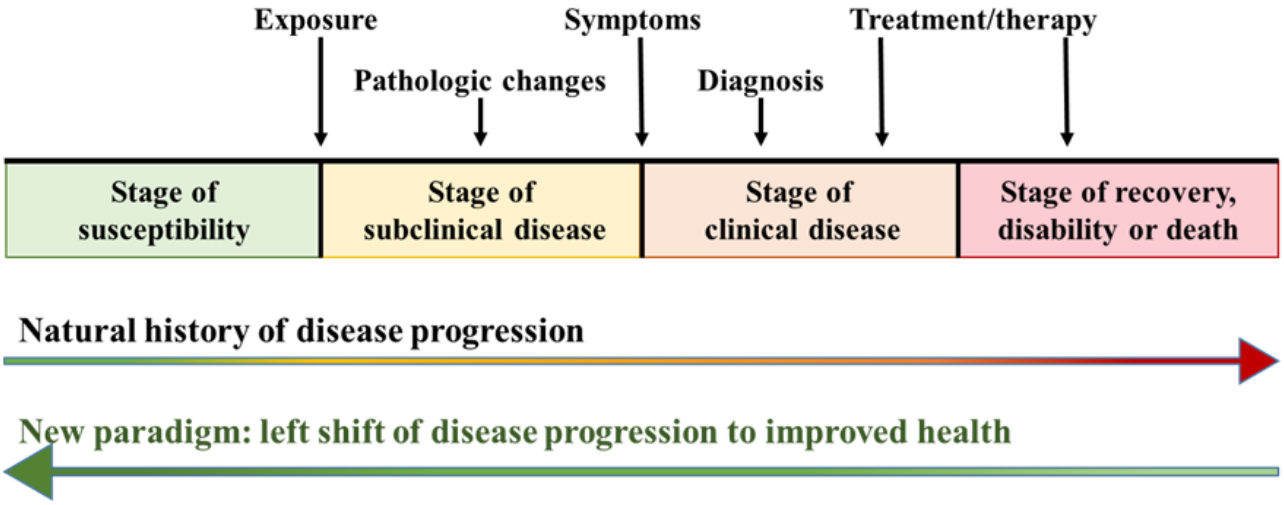

This shift in the paradigm of medical care is also reflected in our understanding and diagnosis of diseases. For instance, autism spectrum disorder (ASD), one of the most common neurodevelopmental disorders, has been diagnosed and named differently in previous versions of the *DSM* (*Diagnostic and Statistical Manual of Mental Disorders*) published by the American Psychiatric Association [23]. In earlier *DSM* versions, modern-day ASD was considered schizophrenia occurring before puberty. From *DSM-III* (*Diagnostic and Statistical Manual of Mental Disorders* [Third Edition]) to *DSM-IV-TR* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition, Text Revision]), autism was distinguished from schizophrenia with the introduction of pervasive developmental disorder under which specific diagnoses expanded from 3 to 5. Eventually, the term "spectrum" and clinical characteristic specifiers were adopted to capture the wide range of symptoms and severity in ASD (*DSM-V* [*Diagnostic and Statistical Manual of Mental Disorders* {Fifth Edition}]). This highlights the importance of a diagnostic classification that can tailor interventions and care to individual needs [24]. In the future, we will see diseases on a continuous scale, and treatment will be personalized to each case. The focus of future medicine will be on providing continuous care rather than a one-time single type of care.

Overall, the prevailing paradigm for natural disease progression is steadily shifting toward health promotion and prevention, making preventive medicine, early diagnosis, and early detection, strategies that aim to reduce the burden of mortality and morbidity from diseases, essential. The value of disease prevention has been highlighted by evidence that suggests that prevention is more valuable than treatment. In the future, the paradigm of medical care is expected to increasingly shift from the stage of treating illness, disability, or death toward promoting overall health and disease prevention.

Future of Preventive Medicine

Current and Future Preventive Medicine

The American Board of Preventive Medicine defines preventive medicine as "the specialty of medical practice that focuses on the health of individuals, communities, and defined populations" and cites protection, promotion, and maintenance of health and well-being, and prevention of disease, disability, and death as some of the field's core goals. In practice, these goals are achieved through 3 distinct types of prevention. Primary prevention aims to prevent the development of diseases in healthy individuals. It includes measures such as smoking cessation programs, vaccinations, and reducing alcohol consumption. Secondary prevention focuses on early detection of diseases to minimize their severity and complications. Examples include cancer-screening programs, tumor biomarkers, and diet or exercise programs to prevent further cardiovascular disease complications. Tertiary prevention aims to reduce the impact of existing diseases or ongoing injuries. It involves prompt treatment and additional approaches such as rehabilitation to prevent complications and minimize lasting effects.

While the conventional prevention triad's approach is still valid, preventive medicine needs technological updates to enable personalized prevention strategies and better disease management, focusing on individual and public health needs. In fact, technology plays a crucial role in the current focus on disease prevention (Table 4). For one, the future of primary prevention care lies in the implementation of disease risk prediction services that use genomic information, such as polygenetic risk scores for cancer risk. 23andMe, for one, provides genetic health screening, offering "Health Predisposition Reports" that screen for genetic risk factors for diabetes; breast, ovarian, prostate, and pancreatic cancer; and

XSL•FO
RenderX

chronic kidney disease, among many others, allowing formation of informed health plans and healthy life choices.

Personalized coaching services to reduce weight, control diet, or shape individual training programs will also play a significant role. Digital health care can overcome the limitations of traditional health care services by creating tailored services that are easy, affordable, efficient, and shaped according to individual needs. KakaoHealthcare Pasta, for example, provides app-based diet management service specialized in diabetes prevention, analyzing which foods most influence blood glucose levels and providing dietary suggestions accordingly. The current and future use of devices will also allow continuous monitoring of body vitals and biomarkers such as glucose, heart rate, and blood pressure. Collected information can be used to keep track of the user's health trend, healthy and unhealthy activities, and provide guides to lead a healthy lifestyle. Noom is one such a personalized health record app that records diet, physical activity, and mood, and provides healthier recipes, encourages and reminds its users to exercise, and offers psychological tools for lasting health improvement.

Table 4. Classical and digital-based preventive medicine based on prevention types.

Prevention type	Classical	Digital-based
Primary	• Smoking/alcohol abstinence advice • Unstructured exercise recommendations • Generic dietary recommendations	• Genetic health risk screening • Personalized health record apps • App-based diet management
Secondary	• Human operator–dependent cancer screenings • Demographic-based cancer screening routines • Annual health checkups	• AI[a]-enhanced medical screening • Personalized cancer screening • Continuous monitoring
Tertiary	• In-person rehabilitation clinics • Standardized cancer chemotherapy • Intermittent blood glucose testing	• Telerehabilitation • Gene-based prognosis/treatment • Continuous blood glucose monitoring

[a]AI: artificial intelligence.

Digital technology could aid secondary prevention, detecting diseases in their early, asymptomatic stages and allowing for timely intervention before symptoms develop. This proactive approach aims to identify risk factors and prevent the progression of the disease and to detect noncommunicable disease complications early. Artificial intelligence, specifically machine learning (ML), has made significant advancements in using large amounts of patient data to enhance decision-making and predictions. ML is already being used in clinical settings to support health care professionals in tasks such as diagnosing respiratory issues from computed tomographic scans [25,26] or assisting with triage decisions through chatbots [27].

ML can also identify patterns and indicators that predict which patients will require more intensive medical interventions in the future [28]. ML also helps make better therapeutic decisions by pinpointing the treatments with the highest chances of success for each patient and their specific condition [29].

One example is Vuno, which provides AI-assisted medical imaging interpretations and diagnoses in a wide range of applications from simple chest radiographs to identifying signs of cardiac arrests and degenerative brain diseases. Chronic diseases such as diabetes and hypertension are recognized as conditions that exist on a continuous spectrum between health and disease. By using serum biomarkers and regular clinical checkups, people spend years in borderline states such as "prediabetes" or "prehypertension." Today's technology can help ascertain these conditions at these early stages, allowing for preventive measures.

In addition, advancements in genomics have allowed for the identification of genetic susceptibility to diseases such as breast cancer. This knowledge enables individuals to take proactive measures to prevent the disease. HexaMed is one such solution that collects anthropometric measurements and lifestyle choices in a simple questionnaire to predict cancer risk and suggests genomic, imaging, biomarker screening, and lifestyle changes.

Medical technology will also help shape individual rehabilitation for tertiary prevention. Individual dose calculations and drug delivery strategies based on each patient's characteristics and diseases, including risk factors; genetic, genomic, and biomarker information; and social, behavioral, and economic conditions, will drive individualized treatment approaches. OncoFree, for example, provides breast cancer prognosis prediction using next-generation sequencing, and in addition, its results can be used to form postsurgical plans by providing clinicians with an estimate of how effective chemotherapy will be.

Case Study 1: Breast Cancer

Breast cancer is the most common cancer and cause of cancer mortality globally [30]. Its rising prevalence is in part driven by lifestyle factors such as poor diet and inactivity [31,32]. Although mammography screening has enhanced early detection and improved prognosis, in many countries, the screening method is recommended through a one-size-fits-all approach. However, the absolute risk for breast cancer for the majority of the population is lower than average, whereas a small portion is at a significantly higher risk. Identifying who is at a higher risk and providing a more personalized screening based on each individual's risk factors can optimize health care resources allocation, ensuring appropriate preventive interventions for high-risk individuals and enhancing overall breast cancer prevention efforts [33].

Technological advancements in digital health could significantly enhance the prevention, early detection, and treatment of breast

XSL•FO
RenderX

cancer—that is, primary, secondary, and tertiary prevention (Table 5). For example, telehealth and telemedicine, actively used for monitoring and follow-up of patients with cancer during and after COVID-19 [34], have been shown to be effective for primary and secondary breast cancer prevention. In aiding patients with *BRCA* mutations, while the level of satisfaction toward counseling for cancer-specific psychological distress was found to be comparable whether the counseling was delivered in-person or remotely, the participants perceived remotely delivered genetic counseling as more convenient and less expensive [35].

Table 5. Prevention strategies and commercial applications according to the level of prevention.

Prevention type	Prevention strategies
Primary	• General advice for a healthy lifestyle. • Mammography screening frequency tailored to risk.
Secondary	• Discuss preventive therapies. • Individual counseling.
Tertiary	• Preventive interventions and endocrine therapy. • Enhanced screening and surveillance.

Telehealth applications can be extended to tertiary prevention of breast cancer as well. Using a face-to-face videoconference strategy to supervise physical exercise in women undergoing primary treatment for breast cancer stages I-III (ABRACE) has been hypothesized to yield better physical and psychological outcomes compared with a health education program alone [36]. Similarly, wearable devices have been shown to significantly improve physical activity, weight control, and overall health in breast cancer survivors [37]. In the future, medical wearables with AI and advanced sensors may be able to detect early-stage breast cancer through noninvasive procedures. For example, a proof-of-concept ultrasonic breast patch (cUSBr-Patch) has been developed that can detect cysts as small as 0.03 cm in diameter, providing a noninvasive way to monitor real-time changes in breast tissue [38].

Another application of technology in breast cancer prevention is in identifying higher-risk individuals for recommendation for screening via risk prediction models (Table 6). Predominantly developed and validated within Caucasian populations, these models have shown modest discriminative power in validation studies with area under the curve values generally ranging between 0.7 and 0.8. The risk factors used are a composite of nonmodifiable elements—age, age at menarche, age at first live birth, additional birth history, family history, age at menopause, and mammogram outcomes—and modifiable elements, notably alcohol consumption and BMI, which provide avenues for risk reduction interventions.

Table 6. Breast cancer risk prediction models.

Model	Study design	Study size (cases/controls)	Remarks
Gail et al [39]	Nested case-control	2852/5998	• Pioneering modeling study. • Focus on reproductive factors.
Rosner and Colditz [40]	Cohort	2249/91,381	• Inclusion of age at multiple birth and has been validated with the inclusion of modifiable risk factors.
Breast Cancer Surveillance Consortium; Kerlikowske et al [41]	Cohort	14,766/1,110,250	• Inclusion of radiologic findings such as BI-RADS[a].
Eriksson et al [42]	Case-control	433/2165	• Emphasis on short-term risk prediction. • Used computer-generated BI-RADS[a].
Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; Antoniou et al [43]	Case-control	1484/156 multiple-case families	• Focus on mutation carrier (BRCA1/BRCA2) probabilities.

[a]BI-RADS: Breast Imaging Reporting and Data System.

The use of risk prediction techniques has enabled us to provide general advice on factors that can lower the risk of breast cancer and to recommend tailored mammography screening frequencies for each woman based on her individual risk level, as a primary prevention strategy. The polygenic risk score (PRS) has been particularly helpful in identifying individuals at a high risk of

XSL•FO
RenderX

developing breast cancer. This allows for targeted interventions aimed at primary prevention, such as personalized lifestyle modifications, preventive therapies, and individual counseling.

In addition, PRS has enabled the stratification of individuals for enhanced screening and surveillance, thereby aiding in the early detection of breast cancer, which aligns with secondary prevention efforts. Furthermore, PRS has the potential to guide tertiary prevention strategies by identifying those who may benefit from more aggressive treatment or preventive measures, such as chemoprevention or risk-reducing surgery both before and after cancer diagnosis to decrease the advanced or recurrent cancer disease burden. Electronic breast cancer risk assessment tools are commonly used to estimate an individual's risk of developing breast cancer. However, it is important to consider both genetic and environmental factors in order to accurately assess the absolute risk for each person in the population [44]. This comprehensive approach allows women to make informed decisions about lifestyle choices and preventive interventions based on their personal risk level and values [33]. Women with a family history of cancer and who are in good overall health should receive personalized care and risk management, while patients with cancer should be educated about how genetic germline mutations may impact their ongoing treatment, follow-up care, and secondary prevention [45]. With the advancement of patient-driven platforms such as HexaMed and DCGen, risk assessment is becoming more personalized and tailored to each individual.

Mobile apps are also used for prevention in breast cancer, especially among marginalized populations and women at increased risk. These apps have been proven effective in increasing mammography awareness and positive decision-making [46]. In addition, mobile apps for primary prevention aim to promote healthy behaviors that reduce the risk of breast cancer [47,48]. These apps primarily target women at high risk for breast cancer, such as postmenopausal women with high Gail risk scores [45] and *BRCA* mutation carriers [49,50]. Recently, studies such as the PREVENTION e-platform have been developed to provide personalized breast cancer information and support tailored to different risk levels [51].

Mobiles apps for tertiary prevention of breast cancer focus on clinical care coordination and health-related quality of life interventions [52]. Cancer care coordination focuses on support and communication between patients with breast cancer and physicians. Specific aspects of coordination include symptomology management, medication adherence, and ambulatory surgery. Health-related quality of life apps targeted general lifestyle management, weight management, depression, and breast cancer-related distress, social support, sleep, and physical activity during and after breast cancer diagnosis. Patients and survivors expressed a preference for receiving clinical care coordination support and quality-of-life interventions through mobile apps, indicating a shift away from standard practices [52].

Case Study 2: Diabetes

Diabetes management, amidst its increasing global burden, presents challenges for both individual patients and the broader public health landscape [53]. The global surge in obesity and sedentary lifestyles has amplified the incidence of diabetes [54,55]. Meanwhile, our understanding of and diagnosis of the disease have advanced to recognize the need for earlier, more personalized care. This is reflected by the American Diabetes Association's changing both the naming and the diagnostic criteria of various stages of hyperglycemia, including hemoglobin A_{1c} (HbA_{1c}) as a diagnostic standard and accepting the concept of "prediabetes" (Table 7) [56,57]. However, traditional health care approaches have encountered significant limitations in preventing, diagnosing early, and mitigating complications associated with the disease. Owing to its gradual onset, lifestyle-related risk factors, and the absence of early symptoms, diabetes proves challenging to both prevent and detect in its initial stages. Moreover, its management is complicated by the necessity for frequent blood glucose monitoring, which yields only intermittent data on glucose levels. Furthermore, treatment regimens typically require strict adherence to medications that are prescribed based on generic guidelines. These guidelines do not accommodate individual patient differences, potentially compromising the effectiveness of the treatment [58].

Table 7. The American Diabetes Association evolution of diabetes as diagnosis.

Definitions	Diagnosis criteria	Year
IGT[a]	2-hour glucose levels 140-199 mg/dL	1979-1997
IFG[b]	Fasting glucose 110-125 mg/dL	1997, 1999
IFG	Fasting plasma glucose 100-125 mg/dL	2003-2006
Borderline HbA_{1c}[c]	HbA_{1c}[c] 6.0%-5.7%	2009-2010
Prediabetes	IGT or IFG or borderline HbA_{1c}[c]	2010

[a]IGT: impaired glucose tolerance.

[b]IFG: impaired fasting glucose.

[c]HbA_{1c}: hemoglobin A_{1c}.

Digital health technology may be capable of improving how diabetes is prevented, screened, and managed. For one, digital health technology offers a promising avenue for primary prevention of diabetes by enabling personalized lifestyle and dietary interventions. Wearables and mobile apps are crucial for encouraging lifestyle changes, such as increasing physical

XSL•FO
RenderX

activity and aiding in smoking cessation, by providing real-time feedback and monitoring. Furthermore, AI-driven personalized nutrition plans, tailored to an individual's glycemic responses and microbiota profiles, play a vital role in optimizing metabolic health. Noom, for instance, is an app that provides personalized recommendations for nutrition, weight loss, and physical activity by improving user engagement and self-improvement [59,60]. Through these technologies, health care can adopt a more proactive and preventive stance, targeting key dietary and lifestyle risk factors effectively.

Digital health technology can also enhance secondary diabetes prevention by enabling noninvasive diagnosis and identifying individuals at high risk. Traditionally, the identification of individuals at high risk of diabetes depended on statistical models, some of which required fasting blood glucose tests (Table 8). These diabetes risk prediction models incorporate a mix of nonmodifiable risk factors, including age, family history of diabetes, comorbidities such as hypertension and cardiovascular disease, and laboratory test results, including blood glucose, lipids, and uric acid levels.

Table 8. Diabetes risk prediction models.

Model	Cohort name	Study size (cases/control)	Remarks
Lindström and Tuomilehto [61]	FINRISK	182/4253	Model based only on lifestyle factors.
Balkau et al [62]	DESIR[a]	203/3611	Genetic polymorphisms included as a risk factor.
Hippisley-Cox et al [63]	QResearch database	78,081/2,462,672	Recognized socioeconomic status as a risk factor.
Kahn et al [64]	ARIC[b]	1821/7766	Offers multiple models for varying levels of available clinical data.

[a]DESIR: Data from an Epidemiological Study on the Insulin Resistance Syndrome.

[b]ARIC: Atherosclerosis Risk in Communities.

More recently, modifiable risk factors such as alcohol consumption, smoking, physical activity, waist circumference, and BMI have been integrated, offering opportunities for risk reduction interventions. Unfortunately, traditional models have demonstrated modest discriminative ability, with area under the curve values typically ranging from 0.6 to 0.7. On the other hand, recent AI-developed models have shown themselves to be capable of predictive population risk stratification, allowing for targeted interventions and showing high sensitivity and specificity. One model, for example, was developed to predict prediabetes and type 2 diabetes based on 1415 Indians and reported 99.5% sensitivity and 99.07% accuracy [65]. This approach of using technology to identify individuals at high risk of diabetes streamlines resource allocation and improves prevention efforts, marking a shift toward proactive diabetes management.

Tertiary prevention of diabetes using digital health technology has shown the most progress in recent years. One example is continuous glucose monitoring (CGM), a wearable device that provides continuous, noninvasive monitoring of glucose levels and is capable of alerting patients to significant blood glucose swings. One such instance, Freestyle Libre, a CGM device, measures glucose levels in cellular fluids instead of serum glucose, offering real-time, noninvasive readings and hypo- and hyperglycemia alerts. The adoption of CGM devices has shown significant reductions in HbA_{1c}, as well as improved participant-reported quality of life [66].

AI has also been used to work in conjunction with existing clinical practices, namely, facilitating the diagnosis of diabetic complications such as retinopathy and foot ulcers. For example, Lam et al [67] used a convoluted neural network to develop a system for identifying retinal lesion from ophthalmologic photographs, reporting 98% accuracy and a receiver operating characteristic curve of 99%. In another instance, Keel et al [68] reported 92% and 94% sensitivity and specificity, respectively, in identifying diabetic retinopathy in outpatient conditions. Siren Care's Siren Socks is a particularly creative approach to managing diabetic ulcers, using socks to measure foot temperature of patients with diabetes to detect diabetic foot neuropathy, ulcer, and injury [69].

Key Considerations for Digital Health Technology Applications

Despite its potential benefits, the application of modern digital health technology in preventive medicine must proceed with caution. For one, since access to digital devices, the web, and digital literacy depend on socioeconomic factors, adopting technology to provide preventive medicine services may further exacerbate existing health disparities. Services such as mobile apps for cancer prevention or diabetes management, despite their potential, will be of limited use if they are not accessible or effectively promoted to marginalized groups. Conversely, digital health technology has the potential to contribute to health equity when used correctly. Telemedicine, for example, significantly enhances health care access for remote populations, individuals with disabilities, and low-income groups by reducing travel costs and barriers to care. To provide digital health care where it is needed, it is important to advocate for the development of communications and technology infrastructure in marginalized communities and support policies that promote inclusive medical practice [70,71].

Another point of concern is the potential challenges posed by an overreliance on digital devices. Excessive use of electronic devices has been associated with both physical and mental health

XSL•FO
RenderX

hazards, such as computer vision syndrome, obesity, and developmental disorders [72-74]. Overuse of digital health devices may cause harm instead of promoting health. There is also the risk of overreliance on electronic systems for health care decisions and delivery, which could detract from the judgment of experienced health care providers. Despite developments in AI and self-health management mobile apps, these electronic devices must be recognized as tools to aid, not replace, the therapeutic efforts of patients and health care providers.

Ethical concerns also warrant attention. The integration of genomics and personal data into health care platforms raises concerns, especially due to possible discrimination. Although genetic data can aid in prevention, there have been concerns that they could be used to deny insurance or employment to individuals or their families [75], as has materialized in some legal cases [76]. Unauthorized access to and misuse of health records could also lead to discrimination and breaches of confidentiality, ultimately undermining trust in digital health systems. There is already a robust body of literature addressing concerns about patient data security [77-79]. Future and continuous efforts must be focused on establishing robust safeguards and ethical guidelines to protect against such discrimination, breaches, and misuse, as well as on policy to ensure that the advancement of digital health technologies does not compromise health care ethics.

Finally, as much as the field of digital technology and its applications in preventive medicine are exploding, there is a critical need for rigorous testing of safety, efficacy, and data validation. One study found that most digital health startups (44%) scored zero, the lowest score, in clinical robustness as measured by regulatory filings and clinical trials [80]. To move past being a mere buzz and truly become future health care, digital health technology requires in-depth, longitudinal studies and user experience research.

This review provides a comprehensive, forward-looking perspective on the future of preventive medicine. By examining current trends, technological advancements, and case studies, this study offers valuable insights into how preventive health care is evolving in the digital era toward a more proactive, personalized, and patient-centered approach to health care. Furthermore, this study identifies the potential of digital health technologies to transform preventive medicine, providing concrete examples of how digital tools are being leveraged for primary, secondary, and tertiary prevention and highlighting the need to address disparities in technology access, prevent overreliance on digital tools, safeguard patient data, and ensure clinical validity of digital health applications.

Despite these strengths, this viewpoint is based largely on observations and trends within the field rather than empirical research. As such, while it is an expert opinion familiar with industry trends and the applications of technology in preventive medicine, this perspective may not fully capture all dimensions of digital health application in prevention. Despite this limitation, we have sought to avoid potential biases by drawing on diverse sources to provide a balanced assessment of the role of digital health in preventive medicine.

Given the rapid advances in digital health technology and its potential to transform preventive medicine, further investigation is warranted. This exploration must focus not only on improving technological capabilities but also on addressing concerns related to health equity, responsible usage, ethics, and clinical validation that accompany such advancements. Only then can we better guarantee that these innovations deliver safe, effective, and equitable health outcomes for all sections of society.

Summary

The future trend of health care is expected to be that of a rise in the older adult population, a shift from cure to prevention, increased use of digital health technology and telemedicine, and a gradual left shift of diseases and the focus of health care. In this review, the concept of primary, secondary, and tertiary prevention in the scope of digital health care was highlighted, as well as breast cancer and diabetes prevention as case studies of the actual application of digital technology in the realm of health care.

The role of preventive medicine is changing as new medical and digital technologies are developed and integrated with health care. These advancements have expanded the reach of preventive services, allowed for remote monitoring, and enabled personalized health interventions, leading to a more patient-centric and data-driven approach to prevention and health care. As a result, health care is closer, faster, and delivered at a lower cost to patients, while patients are empowered with personalized health records and improved patient engagement.

In this study, we have seen that the integration of digital technologies into preventive medicine marks a pivotal evolution in health care. The pace at which technology is evolving is unprecedented. It should be noted, however, that the practical application of these innovations requires robust research to validate the efficacy and safety of digital health technologies in preventive medicine, ensuring that future health care systems are both innovative and inclusive, as well as rigorous validation through large-scale population studies. Looking ahead, we anticipate a transformative health care landscape where personalized preventive measures become fundamental, supported by evidence of efficacy and applicability of digital technologies in preventive medicine.

Authors' Contributions

DK conceptualized the study and contributed to supervision and project administration; KT, SM, and HL contributed to the investigation and writing—review and editing; KT and SM contributed to writing—original draft preparation; and KT conducted visualization.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Ageing and health. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-03-21]
2. GBD compare data visualization. Institute for Health Metrics and Evaluation. 2024. URL: http://vizhub.healthdata.org/gbd-compare [accessed 2024-03-21]
3. Ansah JP, Chiu CT. Projecting the chronic disease burden among the adult population in the United States using a multi-state population model. Front Public Health 2022;10:1082183 [FREE Full text] [doi: 10.3389/fpubh.2022.1082183] [Medline: 36711415]
4. Health spending. Organization for Economic Cooperation and Development. 2022. URL: https://data.oecd.org/healthres/health-spending.htm [accessed 2024-03-21]
5. Chen J, Zhao M, Zhou R, Ou W, Yao P. How heavy is the medical expense burden among the older adults and what are the contributing factors? A literature review and problem-based analysis. Front Public Health 2023;11:1165381 [FREE Full text] [doi: 10.3389/fpubh.2023.1165381] [Medline: 37397714]
6. Graffigna G, Barello S. Patient Health Engagement (PHE) model in enhanced recovery after surgery (ERAS): monitoring patients' engagement and psychological resilience in minimally invasive thoracic surgery. J Thorac Dis 2018;10(Suppl 4):S517-S528 [FREE Full text] [doi: 10.21037/jtd.2017.12.84] [Medline: 29629198]
7. Arai H, Ouchi Y, Toba K, Endo T, Shimokado K, Tsubota K, et al. Japan as the front-runner of super-aged societies: perspectives from medicine and medical care in Japan. Geriatr Gerontol Int 2015;15(6):673-687. [doi: 10.1111/ggi.12450] [Medline: 25656311]
8. Lepre B, Mansfield KJ, Beck EJ. Attitudes, work roles and barriers to nutrition care—interviews with Australian and UK-based medical doctors. J Hum Nutr Diet 2023;36(3):920-931. [doi: 10.1111/jhn.13079] [Medline: 35996856]
9. Prochaska JJ, Gali K, Miller B, Hauer KE. Medical students' attention to multiple risk behaviors: a standardized patient examination. J Gen Intern Med 2012;27(6):700-707 [FREE Full text] [doi: 10.1007/s11606-011-1953-9] [Medline: 22215267]
10. Desebbe O, Lanz T, Kain Z, Cannesson M. The perioperative surgical home: an innovative, patient-centred and cost-effective perioperative care model. Anaesth Crit Care Pain Med 2016;35(1):59-66. [doi: 10.1016/j.accpm.2015.08.001] [Medline: 26613678]
11. Yellowlees P, editor. Telemedicine enabled homecare. 2000 Presented at: Proceedings of American Telemedicine Association 2000: Pragmatic Approaches & Emerging Applications; May 22, 2000; Phoenix, AZ.
12. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google glass in surgical settings: systematic review. JMIR Mhealth Uhealth 2018;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
13. Montag C, Quintana DS. Digital phenotyping in molecular psychiatry—a missed opportunity? Mol Psychiatry 2023;28(1):6-9. [doi: 10.1038/s41380-022-01795-1] [Medline: 36171355]
14. Oudin A, Maatoug R, Bourla A, Ferreri F, Bonnot O, Millet B, et al. Digital phenotyping: data-driven psychiatry to redefine mental health. J Med Internet Res 2023;25:e44502 [FREE Full text] [doi: 10.2196/44502] [Medline: 37792430]
15. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022;49(4):517-530 [FREE Full text] [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
16. Healthcare in a post-COVID-19 world: investable models. KOIS. 2020 Apr 20. URL: https://koisinvest.com/post/healthcare-in-a-post-covid-19-world-investable-models/ [accessed 2024-03-21]
17. Snoswell CL, Smith AC, Page M, Scuffham P, Caffery LJ. Quantifying the societal benefits from telehealth: productivity and reduced travel. Value Health Reg Issues 2022;28:61-66 [FREE Full text] [doi: 10.1016/j.vhri.2021.07.007] [Medline: 34800833]
18. Ray KN, Chari AV, Engberg J, Bertolet M, Mehrotra A. Opportunity costs of ambulatory medical care in the United States. Am J Manag Care 2015;21(8):567-574 [FREE Full text] [Medline: 26295356]
19. Liddy C, Drosinis P, Deri Armstrong C, McKellips F, Afkham A, Keely E. What are the cost savings associated with providing access to specialist care through the Champlain BASE eConsult service? a costing evaluation. BMJ Open 2016;6(6):e010920 [FREE Full text] [doi: 10.1136/bmjopen-2015-010920] [Medline: 27338880]
20. Koklesova L, Samec M, Liskova A, Zhai K, Büsselberg D, Giordano FA, et al. Mitochondrial impairments in aetiopathology of multifactorial diseases: common origin but individual outcomes in context of 3P medicine. EPMA J 2021;12(1):27-40. [doi: 10.1007/s13167-021-00237-2] [Medline: 33686350]
21. Koklesova L, Mazurakova A, Samec M, Kudela E, Biringer K, Kubatka P, et al. Mitochondrial health quality control: measurements and interpretation in the framework of predictive, preventive, and personalized medicine. EPMA J 2022;13(2):177-193 [FREE Full text] [doi: 10.1007/s13167-022-00281-6] [Medline: 35578648]
22. Song M, Lee H, Kang D. The potential application of personalized preventive research. Jpn J Clin Oncol 2014;44(11):1017-1024. [doi: 10.1093/jjco/hyu135] [Medline: 25249379]

XSL•FO
RenderX

23. Rosen NE, Lord C, Volkmar FR. The diagnosis of autism: From kanner to DSM-III to DSM-5 and beyond. J Autism Dev Disord 2021;51(12):4253-4270 [FREE Full text] [doi: 10.1007/s10803-021-04904-1] [Medline: 33624215]
24. Greaves-Lord K, Skuse D, Mandy W. Innovations of the ICD-11 in the field of autism spectrum disorder: a psychological approach. Clin Psychol Eur 2022;4(Spec Issue):e10005 [FREE Full text] [doi: 10.32872/cpe.10005] [Medline: 36760320]
25. Lassau N, Ammari S, Chouzenoux E, Gortais H, Herent P, Devilder M, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. Nat Commun 2021;12(1):634 [FREE Full text] [doi: 10.1038/s41467-020-20657-4] [Medline: 33504775]
26. Bray JJH, Hanif MA, Alradhawi M, Ibbetson J, Dosanjh SS, Smith SL, et al. Machine learning applications in cardiac computed tomography: a composite systematic review. Eur Heart J Open 2022;2(2):oeac018 [FREE Full text] [doi: 10.1093/ehjopen/oeac018] [Medline: 35919128]
27. Wilson L, Marasoiu M. The development and use of chatbots in public health: scoping review. JMIR Hum Factors 2022;9(4):e35882 [FREE Full text] [doi: 10.2196/35882] [Medline: 36197708]
28. Makimoto K, Hogg JC, Bourbeau J, Tan WC, Kirby M. CT imaging with machine learning for predicting progression to COPD in individuals at risk. Chest 2023;164(5):1139-1149. [doi: 10.1016/j.chest.2023.06.008] [Medline: 37421974]
29. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. Med Phys 2020;47(5):e228-e235 [FREE Full text] [doi: 10.1002/mp.13562] [Medline: 32418341]
30. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
31. Ji P, Gong Y, Jin ML, Hu X, Di GH, Shao ZM. The burden and trends of breast cancer from 1990 to 2017 at the global, regional, and national levels: results from the global burden of disease study 2017. Front Oncol 2020;10:650 [FREE Full text] [doi: 10.3389/fonc.2020.00650] [Medline: 32528870]
32. Lofterød T, Frydenberg H, Flote V, Eggen AE, McTiernan A, Mortensen ES, et al. Exploring the effects of lifestyle on breast cancer risk, age at diagnosis, and survival: the EBBA-Life study. Breast Cancer Res Treat 2020;182(1):215-227 [FREE Full text] [doi: 10.1007/s10549-020-05679-2] [Medline: 32436147]
33. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17(7):392-406 [FREE Full text] [doi: 10.1038/nrg.2016.27] [Medline: 27140283]
34. Shaffer KM, Turner KL, Siwik C, Gonzalez BD, Upasani R, Glazer JV, et al. Digital health and telehealth in cancer care: a scoping review of reviews. Lancet Digit Health 2023;5(5):e316-e327 [FREE Full text] [doi: 10.1016/S2589-7500(23)00049-3] [Medline: 37100545]
35. Fournier DM, Bazzell AF, Dains JE. Comparing outcomes of genetic counseling options in breast and ovarian cancer: an integrative review. Oncol Nurs Forum 2018;45(1):96-105. [doi: 10.1188/18.ONF.96-105] [Medline: 29251290]
36. Henkin JS, Botton CE, Simon MS, Rocha GG, Silveira CB, Gehrke RS, et al. Telehealth multicomponent exercise and health education in breast cancer patients undergoing primary treatment: rationale and methodological protocol for a randomized clinical trial (ABRACE: Telehealth). Trials 2023;24(1):42 [FREE Full text] [doi: 10.1186/s13063-022-07015-z] [Medline: 36658611]
37. Teo NR, Siew LED, Ang WHD, Lau Y. Wearable-technology-assisted interventions for breast-cancer survivors: a meta-analysis and meta-regression. Semin Oncol Nurs 2023;39(3):151403. [doi: 10.1016/j.soncn.2023.151403] [Medline: 36894449]
38. Du W, Zhang L, Suh E, Lin D, Marcus C, Ozkan L, et al. Conformable ultrasound breast patch for deep tissue scanning and imaging. Sci Adv 2023;9(30):eadh5325 [FREE Full text] [doi: 10.1126/sciadv.adh5325] [Medline: 37506210]
39. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81(24):1879-1886. [doi: 10.1093/jnci/81.24.1879] [Medline: 2593165]
40. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. J Natl Cancer Inst 1996;88(6):359-364. [doi: 10.1093/jnci/88.6.359] [Medline: 8609645]
41. Kerlikowske K, Chen S, Golmakani MK, Sprague BL, Tice JA, Tosteson ANA, et al. Cumulative advanced breast cancer risk prediction model developed in a screening mammography population. J Natl Cancer Inst 2022;114(5):676-685 [FREE Full text] [doi: 10.1093/jnci/djac008] [Medline: 35026019]
42. Eriksson M, Czene K, Pawitan Y, Leifland K, Darabi H, Hall P. A clinical model for identifying the short-term risk of breast cancer. Breast Cancer Res 2017;19(1):29 [FREE Full text] [doi: 10.1186/s13058-017-0820-y] [Medline: 28288659]
43. Antoniou AC, Pharoah PPD, Smith P, Easton DF. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer 2004;91(8):1580-1590 [FREE Full text] [doi: 10.1038/sj.bjc.6602175] [Medline: 15381934]
44. Choi J, Ha TW, Choi HM, Lee HB, Shin HC, Chung W, et al. Development of a breast cancer risk prediction model incorporating polygenic risk scores and nongenetic risk factors for Korean women. Cancer Epidemiol Biomarkers Prev 2023;32(9):1182-1189 [FREE Full text] [doi: 10.1158/1055-9965.EPI-23-0064] [Medline: 37310812]

XSL•FO
RenderX

45. Speiser D, Heibges M, Besch L, Hilger C, Keinert M, Klein K, et al. Paradigmatic approach to support personalized counseling with digital health (iKNOW). JMIR Form Res 2023;7:e41179 [FREE Full text] [doi: 10.2196/41179] [Medline: 37083496]
46. Eden KB, Scariati P, Klein K, Watson L, Remiker M, Hribar M, et al. Mammography decision aid reduces decisional conflict for women in their forties considering screening. J Womens Health (Larchmt) 2015;24(12):1013-1020 [FREE Full text] [doi: 10.1089/jwh.2015.5256] [Medline: 26360918]
47. Coughlin SS, Besenyi GM, Bowen D, De Leo G. Development of the Physical Activity and Your Nutrition for Cancer (PYNC) smartphone app for preventing breast cancer in women. Mhealth 2017;3:5 [FREE Full text] [doi: 10.21037/mhealth.2017.02.02] [Medline: 28293621]
48. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and phone-based weight loss intervention: pilot RCT in women at elevated breast cancer risk. Am J Prev Med 2016;51(5):714-721 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.024] [Medline: 27593420]
49. Cohen SA, Scherr CL, Nixon DM. An iPhone application intervention to promote surveillance among women with a BRCA mutation: pre-intervention data. J Genet Couns 2018;27(2):446-456. [doi: 10.1007/s10897-018-0224-x] [Medline: 29429041]
50. Scherr CL, Feuston JL, Nixon DM, Cohen SA. A two-phase approach to developing SNAP: an iPhone application to support appointment scheduling and management for women with a BRCA mutation. J Genet Couns 2018;27(2):439-445. [doi: 10.1007/s10897-018-0222-z] [Medline: 29423568]
51. Attieh S, Monarque M, Durand A, Ahmed S, Knoppers BM, Simard J, et al. Perceptions and usability of PREVENTION: a breast cancer risk assessment e-platform. J Pers Med 2023;13(5):850 [FREE Full text] [doi: 10.3390/jpm13050850] [Medline: 37241021]
52. Houghton LC, Howland RE, McDonald JA. Mobilizing breast cancer prevention research through smartphone apps: a systematic review of the literature. Front Public Health 2019;7:298 [FREE Full text] [doi: 10.3389/fpubh.2019.00298] [Medline: 31781525]
53. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119 [FREE Full text] [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
54. Deng MG, Cui HT, Lan YB, Nie JQ, Liang YH, Chai C. Physical activity, sedentary behavior, and the risk of type 2 diabetes: a two-sample mendelian randomization analysis in the European population. Front Endocrinol (Lausanne) 2022;13:964132 [FREE Full text] [doi: 10.3389/fendo.2022.964132] [Medline: 36407298]
55. Zimmet P, Alberti KGMM, Shaw J. Global and societal implications of the diabetes epidemic. Nature 2001;414(6865):782-787. [doi: 10.1038/414782a] [Medline: 11742409]
56. Davidson MB. Historical review of the diagnosis of prediabetes/intermediate hyperglycemia: case for the international criteria. Diabetes Res Clin Pract 2022;185:109219. [doi: 10.1016/j.diabres.2022.109219] [Medline: 35134465]
57. Yudkin JS, Montori VM. The epidemic of pre-diabetes: the medicine and the politics. BMJ 2014;349:g4485 [FREE Full text] [doi: 10.1136/bmj.g4485] [Medline: 25028385]
58. American Journal of Managed Care. Challenges in diabetes management: glycemic control, medication adherence, and healthcare costs. 2017 Aug 21. URL: https://www.ajmc.com/journals/supplement/2017/challenges-in-diabetes-management/challenges-in-diabetes-management-article [accessed 2024-03-11]
59. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
60. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-Up. JMIR Mhealth Uhealth 2018;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
61. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 2003;26(3):725-731. [doi: 10.2337/diacare.26.3.725] [Medline: 12610029]
62. Balkau B, Lange C, Fezeu L, Tichet J, de Lauzon-Guillain B, Czernichow S, et al. Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). Diabetes Care 2008;31(10):2056-2061 [FREE Full text] [doi: 10.2337/dc08-0368] [Medline: 18689695]
63. Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in england and wales: prospective derivation and validation of QDScore. BMJ 2009;338:b880 [FREE Full text] [doi: 10.1136/bmj.b880] [Medline: 19297312]
64. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years. Ann Intern Med 2009;150(11):741-751 [FREE Full text] [doi: 10.7326/0003-4819-150-11-200906020-00002] [Medline: 19487709]
65. Shankaracharya, Odedra D, Samanta S, Vidyarthi AS. Computational intelligence-based diagnosis tool for the detection of prediabetes and type 2 diabetes in India. Rev Diabet Stud 2012;9(1):55-62 [FREE Full text] [doi: 10.1900/RDS.2012.9.55] [Medline: 22972445]

XSL•FO
RenderX

66. Leelarathna L, Evans ML, Neupane S, Rayman G, Lumley S, Cranston I, et al. Intermittently scanned continuous glucose monitoring for type 1 diabetes. N Engl J Med 2022;387(16):1477-1487. [doi: 10.1056/NEJMoa2205650] [Medline: 36198143]
67. Lam C, Yu C, Huang L, Rubin D. Retinal lesion detection with deep learning using image patches. Invest Ophthalmol Vis Sci 2018;59(1):590-596 [FREE Full text] [doi: 10.1167/iovs.17-22721] [Medline: 29372258]
68. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018;8(1):4330 [FREE Full text] [doi: 10.1038/s41598-018-22612-2] [Medline: 29531299]
69. Reyzelman AM, Shih CD, Tovmassian G, Nathan M, Ma R, Scholten HJ, et al. An evaluation of real-world smart sock-based temperature monitoring data as a physiological indicator of early diabetic foot injury: case-control study. JMIR Form Res 2022;6(4):e31870 [FREE Full text] [doi: 10.2196/31870] [Medline: 35363148]
70. Phuong J, Ordóñez P, Cao J, Moukheiber M, Moukheiber L, Caspi A, et al. Telehealth and digital health innovations: a mixed landscape of access. PLOS Digit Health 2023;2(12):e0000401 [FREE Full text] [doi: 10.1371/journal.pdig.0000401] [Medline: 38100519]
71. Liu Y, Liu K, Zhang X, Guo Q. Does digital infrastructure improve public Health? A quasi-natural experiment based on china's broadband policy. Soc Sci Med 2024;344:116624. [doi: 10.1016/j.socscimed.2024.116624] [Medline: 38290184]
72. Ccami-Bernal F, Soriano-Moreno DR, Romero-Robles MA, Barriga-Chambi F, Tuco KG, Castro-Diaz SD, et al. Prevalence of computer vision syndrome: a systematic review and meta-analysis. J Optom 2024;17(1):100482 [FREE Full text] [doi: 10.1016/j.optom.2023.100482] [Medline: 37866176]
73. Ikeda I, Fujihara K, Morikawa Yoshizawa S, Takeda Y, Ishiguro H, Yamada Harada M, et al. Association between screen time, including that for smartphones, and overweight/obesity among children in Japan: NICE EVIDENCE study 4. Endocr J 2024;71(2):171-179 [FREE Full text] [doi: 10.1507/endocrj.EJ23-0343] [Medline: 38199254]
74. Heffler KF, Acharya B, Subedi K, Bennett DS. Early-life digital media experiences and development of atypical sensory processing. JAMA Pediatr 2024;178(3):266-273. [doi: 10.1001/jamapediatrics.2023.5923] [Medline: 38190175]
75. Joly Y, Dalpe G. Genetic discrimination still casts a large shadow in 2022. Eur J Hum Genet 2022;30(12):1320-1322 [FREE Full text] [doi: 10.1038/s41431-022-01194-8] [Medline: 36163420]
76. Suter SM. GINA at 10 years: the battle over 'genetic information' continues in court. J Law Biosci 2018;5(3):495-526 [FREE Full text] [doi: 10.1093/jlb/lsz002] [Medline: 31143453]
77. Cohen IG, Mello MM. HIPAA and protecting health information in the 21st century. JAMA 2018;320(3):231-232. [doi: 10.1001/jama.2018.5630] [Medline: 29800120]
78. Grande D, Luna Marti X, Feuerstein-Simon R, Merchant RM, Asch DA, Lewson A, et al. Health policy and privacy challenges associated with digital technology. JAMA Netw Open 2020;3(7):e208285 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8285] [Medline: 32644138]
79. Alhammad N, Alajlani M, Abd-Alrazaq A, Epiphaniou G, Arvanitis T. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024;26:e50715 [FREE Full text] [doi: 10.2196/50715] [Medline: 38820572]
80. Day S, Shah V, Kaganoff S, Powelson S, Mathews SC. Assessing the clinical robustness of digital health startups: cross-sectional observational analysis. J Med Internet Res 2022;24(6):e37677 [FREE Full text] [doi: 10.2196/37677] [Medline: 35723914]

Abbreviations

AI: artificial intelligence
ASD: autism spectrum disorder
CGM: continuous glucose monitoring
DSM: Diagnostic and Statistical Manual of Mental Disorders
DSM-III: Diagnostic and Statistical Manual of Mental Disorders (Third Edition)
DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition, Text Revision)
DSM-V: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
HbA1c: hemoglobin A1c
ML: machine learning
PRS: polygenic risk score

XSL•FO
RenderX

Edited by A Mavragani; submitted 04.04.24; peer-reviewed by S Kulkarni, A Jain, H Younes, S Dubare; comments to author 29.05.24; revised version received 09.06.24; accepted 05.11.24; published 27.02.25.

Please cite as:
De la Torre K, Min S, Lee H, Kang D
The Application of Preventive Medicine in the Future Digital Health Era
J Med Internet Res 2025;27:e59165
URL: https://www.jmir.org/2025/1/e59165
doi:10.2196/59165
PMID:40053712

©Katherine De la Torre, Sukhong Min, Hyobin Lee, Daehee Kang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Empowering Health Care Actors to Contribute to the Implementation of Health Data Integration Platforms: Retrospective of the medEmotion Project

Marcel Parciak[1,2*], MSc; Noëlla Pierlet[1,2,3*], MSc; Liesbet M Peeters[1,2,4], Prof Dr

[1]Biomedical Research Institute, UHasselt - Hasselt University, Diepenbeek, Belgium

[2]Data Science Institute, UHasselt - Hasselt University, Hasselt, Belgium

[3]Ziekenhuis Oost-Limburg, Genk, Belgium

[4]UHasselt - Hasselt University, University Multiple Sclerosis Center, Diepenbeek, Belgium

[*]these authors contributed equally

Corresponding Author:
Marcel Parciak, MSc
Biomedical Research Institute
UHasselt - Hasselt University
Agoralaan
Diepenbeek, 3590
Belgium
Phone: 32 11269288
Email: marcel.parciak@uhasselt.be

Abstract

Health data integration platforms are vital to drive collaborative, interdisciplinary medical research projects. Developing such a platform requires input from different stakeholders. Managing these stakeholders and steering platform development is challenging, and misaligning the platform to the partners' strategies might lead to a low acceptance of the final platform. We present the medEmotion project, a collaborative effort among 7 partners from health care, academia, and industry to develop a health data integration platform for the region of Limburg in Belgium. We focus on the development process and stakeholder engagement, aiming to give practical advice for similar future efforts based on our reflections on medEmotion. We introduce Personas to paraphrase different roles that stakeholders take and Demonstrators that summarize personas' requirements with respect to the platform. Both the personas and the demonstrators serve 2 purposes. First, they are used to define technical requirements for the medEmotion platform. Second, they represent a communication vehicle that simplifies discussions among all stakeholders. Based on the personas and demonstrators, we present the medEmotion platform based on components from the Microsoft Azure cloud. The demonstrators are based on real-world use cases and showcase the utility of the platform. We reflect on the development process of medEmotion and distill takeaway messages that will be helpful for future projects. Investing in community building, stakeholder engagement, and education is vital to building an ecosystem for a health data integration platform. Instead of academic-led projects, the health care providers themselves ideally drive collaboration among health care providers. The providers are best positioned to address hospital-specific requirements, while academics take a neutral mediator role. This also includes the ideation phase, where it is vital to ensure the involvement of all stakeholders. Finally, balancing innovation with implementation is key to developing an innovative yet sustainable health data integration platform.

(J Med Internet Res 2025;27:e68083) doi:10.2196/68083

KEYWORDS

data science; health data integration; health data platform; real-world evidence; health care; health data; data; integration platforms; collaborative; platform; Belgium; Europe; personas; communication; health care providers; hospital-specific requirements; digital health

XSL•FO
RenderX

Introduction

Accurate and well-formatted data are key to delivering high-quality health care and fueling medical research [1-3]. All health care actors acquire real-world data, defined as any health care-related information captured from the patient [4]. The volume, velocity, and variety of acquired data, however, raise challenges for data processing systems [5]. Data engineers work in interdisciplinary environments to ensure that users receive data in the right format, at the right time and place to generate real-world evidence [4]. Health data engineering is a complex and time-consuming task that cannot be managed without IT solutions tailored to environment-specific requirements [6]. Consequently, health data integration platforms are a hot topic in medical informatics research [7-11]. Health data integration commonly refers to finding relevant and rich patient information on time with appropriate interfaces [6]. In this paper, we present our design and development approach to building such a platform.

The medEmotion project was a collaborative initiative among 3 hospitals (Jessa Ziekenhuis, Noorderhart, and Ziekenhuis Oost-Limburg), 2 academic institutions (Hogeschool PXL, and UHasselt), and 2 industry partners (LRM, and BioVille) from the region of Limburg, Belgium. The primary objective of medEmotion was to establish a comprehensive data integration platform designed to effectively address a wide array of health data-related inquiries and challenges encountered by various stakeholders within the health care ecosystem. The medEmotion platform aggregates medical data from hospitals and general practitioners, personal health information collected through wearable devices, and environmental data. It enables researchers, health care professionals, and entrepreneurs to test innovative data-based concepts and generate real-world evidence. We built the medEmotion platform to be secure and compliant with legal standards to be fit for purpose for real-world health data. The platform enhances decision-making and stakeholder collaboration by facilitating comprehensive analysis of integrated real-world health data.

To guide and showcase the data integration platform developed in medEmotion, we defined 2 real-world use cases for distinct disease areas. First, with the Noorderhart hospital, we approach the multiple sclerosis data connect (MSDC) use case, focusing on the health care process of people with multiple sclerosis. MSDC is characterized by a relatively low total amount of patients just below a thousand where each patient is described with a high amount of diverse information in the form of structured clinical data, evoked potential time series data, magnetic resonance images, and para-clinical data, such as information collected during physiotherapy. We thus obtain a broad dataset that brings a distinct set of challenges from the second use case in medEmotion, Cardio. Developed in collaboration with Jessa Ziekenhuis and Ziekenhuis Oost-Limburg, we consider the postclinical workflows of cardiology patients. Cardio is characterized by a low variety and high velocity of data collected per patient. Next to a basic set of clinical variables, most data are acquired through wearable sensors. We thus have a data source that continuously generates new data points.

In this paper, we outline the medEmotion project, focusing on our chosen development approach. In particular, we describe a team of personas and a set of demonstrators. Personas are fictional stakeholders whose requirements and unique perspectives we implemented into the project. Demonstrators describe sets of functional requirements of the medEmotion platform aiming to solve real-world use cases. Each demonstrator's requirements are affected by one or more personas.

Furthermore, we discuss our approach, combining personal reflections on our challenges during medEmotion. With these reflections, we reminisce about our experiences that we believe present common pitfalls of health care-related projects that handle real-world data. We aim to shed light on the tension between stakeholders from health care, academia, industry, and investment societies and give advice for future efforts.

The paper is structured as follows. We introduce the personas and the demonstrators of medEmotion alongside a summary of its implementation. Afterward, we discuss our development approach and the result. We end with distilling key takeaway messages from the medEmotion project.

Personas

Stakeholder management is a core project management task that aligns the requirements and visions of stakeholders. A stakeholder may be part of the core project team or take a consultant role, advising the project team regularly. From our experience, most stakeholders combine multiple personas. Each stakeholder might take on a different "hat," sharing their view and arguments in different, situation-specific ways to drive the project forward. Therefore, we define 6 personas that represent the different "hats" a stakeholder may represent in the following (Figure 1). The following personas are based on observations during medEmotion, we thus do not claim this list to be exhaustive. We think this list will serve as a starting point to identify personas in other projects.

XSL•FO
RenderX

Figure 1. Personas involved in the different demonstrators, with some of their focus points indicated. GDPR: General Data Protection Regulation.

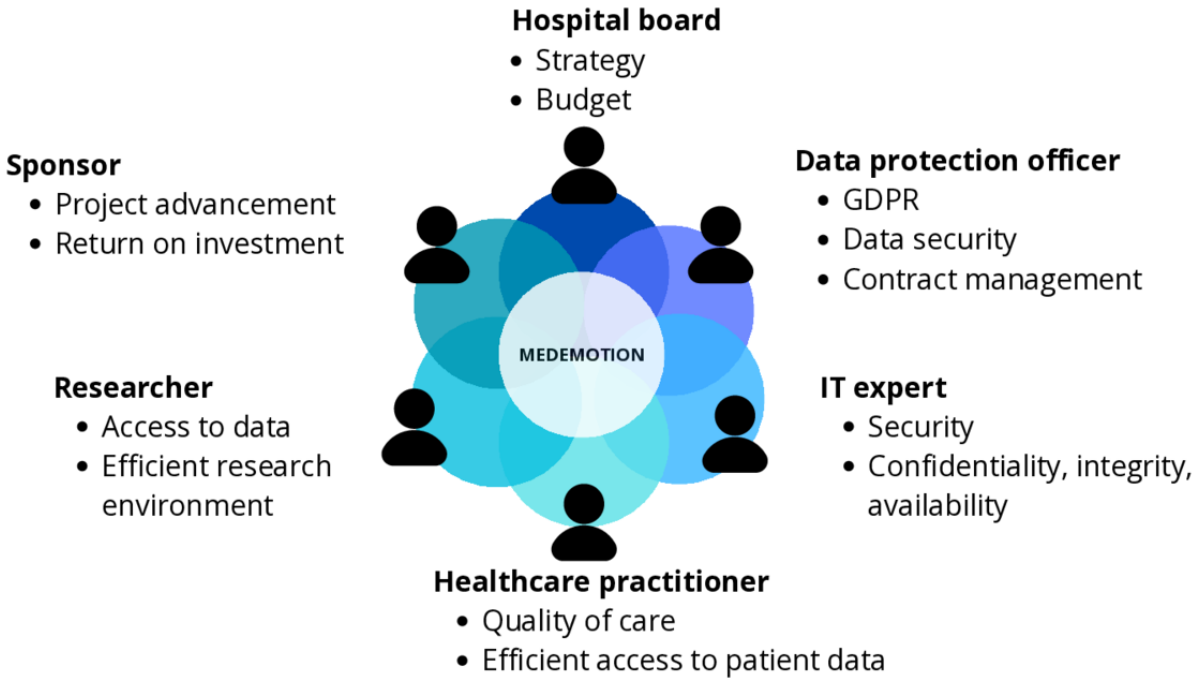

Data Protection Officer

The Data Protection Officer persona is concerned with keeping information up to date, private, secure, and traceable. These concerns can be enforced by law, such as the General Data Protection Regulation (GDPR) of the European Union, or can be intrinsic in the sense that keeping data well governed makes aspects such as data security and contract management easier. People who represent this persona are Data Protection Officers of hospitals, contract managers, or data stewards.

Researcher

The Researcher persona is concerned with conducting research more efficiently or easily. In health data integration, conducting research typically means running experiments on real-world health data. Thus, the researcher aims to get easier access to health data, larger amounts of health data, more computational resources, or access to software tools that accelerate data processing. A common example of a researcher is a PhD student.

Health Care Practitioner

The health care practitioner is concerned with delivering the best health care quality to the patient. To do so, they need recent, correct, and sufficient information about a patient while they consult with them or decide on future medical interventions. For example, clinicians and nurses represent this persona.

IT Expert

The IT Expert persona is concerned with efficiently and reliably running the hospital information system (HIS). The IT-security aims of confidentiality, integrity, and availability are of major concern, which results in a more reluctant attitude toward new and unknown systems. The hospital's chief information officer represents this persona.

Hospital Board Member

The hospital board member persona is concerned with the strategic advancement of the hospital within the resources available. As such, this persona will wage strategic and long-term goals, such as participating in international research studies, against short-term available resources, such as the workforce and budget of a hospital. For example, a hospital's chief executive officer represents this persona.

Sponsor

The Sponsor persona is concerned with the success of the investment. In a research project such as medEmotion, this includes ensuring that project resources are well-spent and securing potential return on investments about the project results. This persona is represented by project sponsors, such as venture capital, pharma, or insurance companies.

During medEmotion, we found that defining personas and assigning them to our stakeholders made discussions more transparent and efficient. The set of people participating in discussions and meetings changes frequently during a project. A list of predefined personas allowed us to ensure we included all distinct perspectives to decide on the next steps. In particular, we assigned these personas to demonstrators to gain well-structured functional requirements, which we will list in the following section.

Demonstrators

We defined a series of real-world demonstrators to facilitate the requirement engineering process. Each demonstrator represents a bundle of related technical requirements that include the unique perspectives of all personas involved. In particular, our demonstrators are disease-agnostic and can be applied to any real-world dataset. By contrast, the use cases we used in

XSL•FO
RenderX

medEmotion describe higher-level and disease-specific requirements that affect multiple demonstrators. We present a short description of each defined demonstrator, and the personas involved.

Orchestration of Complex Pipelines

In biomedical research projects, (para-)clinical data must be processed through complex data integration steps. Researchers often implement these steps ad hoc using different programming languages (eg, Python [Python software foundation], R [R foundation], MATLAB [MathWorks]), and software frameworks (eg, Torch [PyTorch Foundation], Pandas [Pandas project], dplyr [posit]) or reuse existing implementations as black boxes [12]. The resulting software code is complex and difficult to automate, orchestrate, and reproduce, pushing it out of reach for IT experts. Further, data protection officers see a lack of in-depth documentation, essentially losing real-world data due to missing knowledge. With this demonstrator, we showcase the platform's capabilities to orchestrate complex data integration pipelines automatically.

Private Research Environments

Access to real-world data greatly increases the efficiency of exploratory research tasks. With strict privacy regulations such as the GDPR of the European Union or the HIPAA (Health Insurance Portability and Accountability Act) of the United States, data access is often achieved by sharing datasets after bilateral contracts between the data custodians (eg, hospitals) and universities are set in place [13]. This process, however, is slow, complex, and lacks transparency, creating tension between researchers, Data Protection Officers, and IT experts. Researchers need to deal with tedious administration tasks, while data privacy officers and IT experts are concerned with keeping sensitive data secure. With this demonstrator, we showcase the platform's ability to create private research environments that allow researchers to access data, tooling, and compute resources while keeping real-world health data secure.

Observational Health Data Sciences and Informatics Observational Medical Outcomes Partnership Common Data Model

Transnational medical research can be facilitated with internationally standardized data schemas, such as the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) developed by the Observational Health Data Sciences and Informatics (OHDSI) research community. Transforming health care records to OMOP is a highly operational task, enabling data custodians to participate in federated, transnational research projects without sharing sensitive datasets [14]. Transforming and maintaining data structured according to the OMOP CDM is desirable for researchers and hospital board members, as it allows for simple access, privacy perseverance, and increased hospital visibility in the academic landscape. With this demonstrator, we showcase the platform's ability to facilitate transforming data to and maintaining data in the OMOP CDM, including running a suite of software tools developed by the OHDSI community.

Visualization of Integrated Datasets to Support Care

Providing high-quality health care necessitates large amounts of integrated, patient-centric data that facilitate informed decision-making. Analyzing these extensive amounts of patient data and gathering insights on the evolution of the patient's medical conditions over time in a few minutes is only possible with data-rich visualizations [15]. The clinical decision-making process remains health care practitioner–specific as well as multidisciplinary. Hence, visualizations must be adaptable per individual and still sharable within a multidisciplinary team to stay helpful. Privacy remains essential in the eyes of Data Protection Officers, hence, sophisticated methods to share information according to therapeutic relationships between health care practitioners and patients are needed. With this demonstrator, we showcase the platform's ability to facilitate online visualizations of rich datasets that can be tailored to individual needs and shared among multidisciplinary teams.

Data Ingestion From Hospital Data

The primary sources of real-world data meant to generate real-world evidence are HISs. Even for clinical studies that collect data with study-specific case report forms, HIS data helps provide important context information, such as demographics or comorbidities [1]. Due to the value of HIS data, one of the sponsor's main targets is integrating such data into the platform. This real-world data is highly sensitive and the prime target of data protection measures set up by data protection officers and IT experts. Researchers aiming to access the data naturally clash with data protection measures, as they tend to over-request data. With this demonstrator, we showcase the platform's ability to ingest data from multiple pseudonymized data sources.

Data Ingestion From Devices

With eHealth solutions on the rise, data integration platforms receive more interest from sponsors to include data collected outside the traditional boundaries such as the hospital or the practice. Patients collect data themselves, either actively using health apps or passively using wearable devices [16], both using lifestyle and software-as-medical device apps [17]. Providing richer datasets for clinicians and researchers, IT experts face challenging data integration tasks of high volume and erroneous-prone data streams. Of course, these data streams are sensitive and thus are a concern of Data Protection Officers. With this demonstrator, we showcase the platform's ability to include data streams from home monitoring devices that are not part of traditional hospital information systems.

Connection With High-Performance Computing Centers

High-performance computing (HPC) enables researchers to use machine-learning approaches [18]. Those can ramp up costs quickly when acquired from public cloud providers such as Microsoft Azure, Google Cloud, or Amazon Web Services. Many academic institutions operate HPC centers to meet the demand for HPC. To keep research projects using machine learning feasible, data transfer to and from academic HPC centers needs to be simple while retaining IT security standards set by data protection officers. With this demonstrator, we

XSL•FO
RenderX

showcase the platform's ability to transform data between the Vlaams Supercomputer Center, an HPC center operated by and for the 5 Flemish universities.

Summary

The demonstrators guided discussions and the implementation of the medEmotion platform. Based on the personas, we could identify which stakeholders to include in discussions. Hence, we were able to have more ad-hoc discussions in smaller groups, replacing large and time-consuming meetings to discuss the entirety of the platform components. Combining personas and demonstrators allowed for a rapid implementation process, which we summarize in the following.

Implementation

We implemented the medEmotion platform with an industry partner using Microsoft's Azure Cloud Platform components. The platform's architecture foresees a general environment and multiple use case-specific data silos. We find modules and tools in a general environment that handle pseudonymized, cleaned, and integrated data. In the hospital-specific data silos, we find modules and tools to preprocess datasets either for the general environment or for hospital-specific projects. The implementation process was guided and managed by using case-specific task forces that met at regular intervals. Each task force consisted of relevant stakeholders from multiple partners that covered all personas as described above. Each task force meeting was attended by a member of the medEmotion core team which summarized the discussions and extracted feedback, suggestions, and additional requirements that were reported to the implementing industry partner.

To ensure data privacy, data from different source systems is pseudonymized with source-specific keys on entering the data platform, such that the data in the medEmotion platform cannot be linked without additional effort. The platform retains encrypted quasi-identifiers, information irrelevant to answer clinical questions that potentially reveal the identity of a patient, to allow for ad-hoc record linkage if needed. The encrypted identifiers are locked away in a protected zone in the platform, and the keys to decrypt the identifiers are locked away in Azure Key Vaults. Next to the security guarantees provided by Microsoft Azure with respect to the Key Vaults, the medEmotion platform could also be implemented with on-premises Key Vaults, granting full control to data custodians with respect to the decryption keys used to link data that stems from their systems. In our proof-of-concept implementation, we used a Key Vault operated within the Microsoft Azure public cloud.

The figure shows components from 3 environments: On-premises (eg, a partner hospital), the Azure cloud, and External, which shows users outside the platform. Where applicable, we name the Microsoft Azure product used written in bold. The arrows indicate data flows, starting with the data sources on the left and flowing to the data users on the right.

The demonstrators were implemented exclusively with public or mock datasets, that was synthetically generated data modeled after real-world datasets. We created a mock data source following health level-7 FHIR (Fast Healthcare Interoperability Resources) [19] for the data ingestion demonstrators. The Azure FHIR Store is a natural fit for this data source. Further, an industry partner that offers wearable eHealth devices supplied us with a mock data stream for our eHealth device demonstrator. We generally use a decentralized lightweight pseudonymization service that runs on-premise at a hospital. This service polls for and pulls data in scheduled intervals from the sources, pseudonymizes it, and ingests it into the platform using Azure Data Factory.

To showcase the OHDSI OMOP CDM demonstrator, we generated a mock dataset based on a multiple sclerosis–specific data acquisition system, which we transformed into the OMOP CDM format. With this data, we operationalized OHDSI tools such as ATLAS to test the transformed data [14].

We used the datasets described by Yperman et al [18] for the remaining demonstrators. We use the Yperman et al [20] open-access dataset and an open-access magnetic resonance image dataset from the MICCAI [21] 2016 challenge. We use these datasets to showcase the orchestration of complex pipelines with Azure Databricks, as this enables our existing preprocessing scripts to be automated, orchestrated, and monitored. We encapsulate Azure Databricks in Azure Virtual Machines, where we block all connections to public networks (such as the web) to showcase private research environments. With the same tools but less stringent blocking rules (ie, we allow connection to selected HPC centers using virtual private network connections), we enable the use of HPC resources from within the platform. Finally, we use Microsoft's PowerBI to create dashboards per health care practitioner's requirements. In particular, we built an online, data-rich dashboard that respects therapeutic relationships (Figure 2).

XSL•FO
RenderX

Figure 2. A high-level overview of the medEmotion platform implementation. The figure shows components from 3 environments: On-premises (eg, a partner hospital), the Azure cloud, and External, which shows users outside the platform. Where applicable, we name the Microsoft Azure product used written in bold. The arrows indicate data flows, starting with the data sources on the left and flowing to the data users on the right. CDM: common data model; FHIR: Fast Healthcare Interoperability Resources; OHDSI: Observational Health Data Sciences and Informatics; OMOP: Observational Medical Outcomes Partnership.

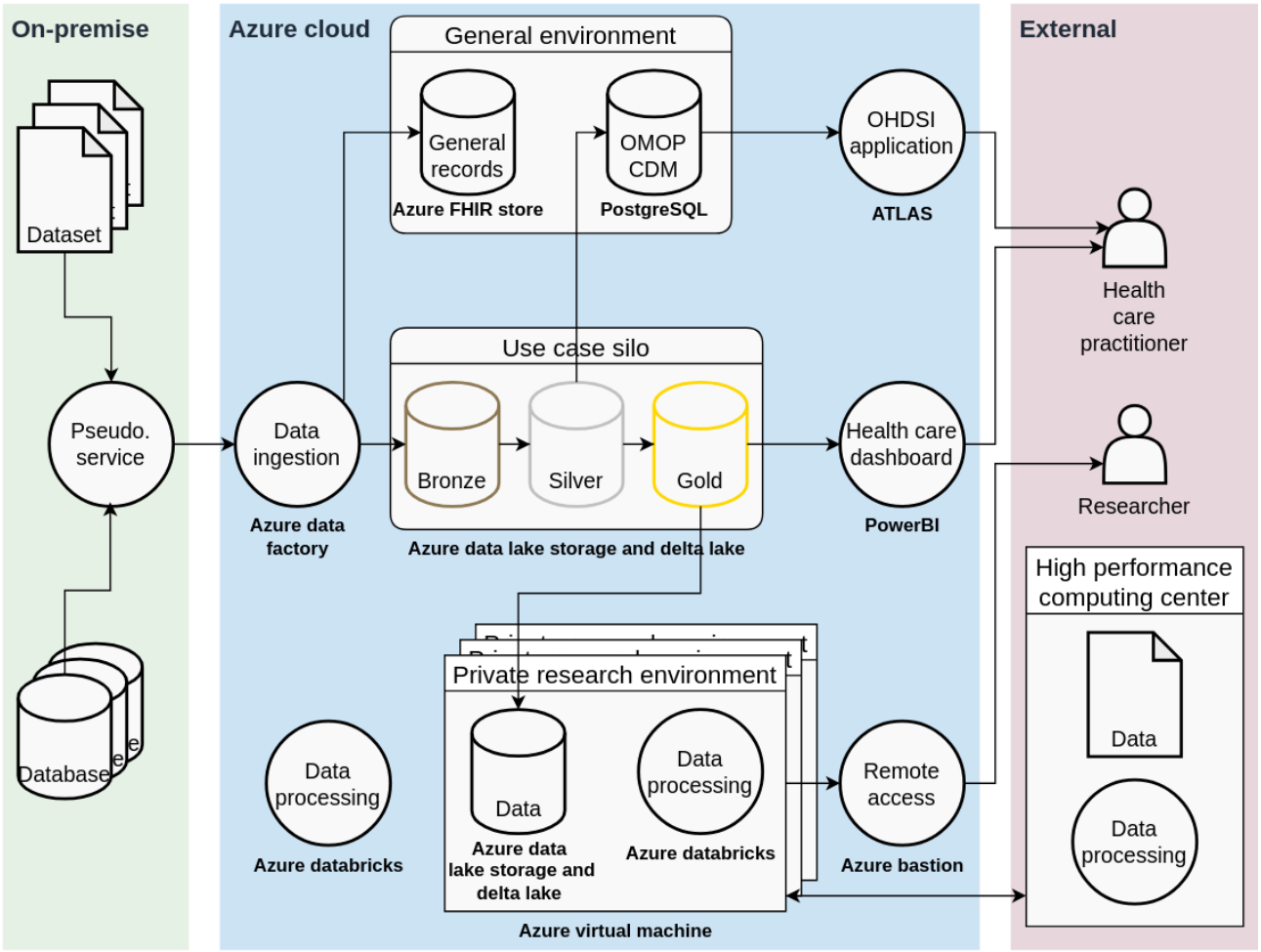

Limitations

We combined personas and demonstrators to effectively steer the development and implementation of the medEmotion platform, resulting in a feature-rich yet adaptable solution. The use of mock-up datasets has proven particularly useful to accelerate development, as access to patient-level data remains a time-consuming process. While we focused on showcasing the technical feasibility of a data integration platform, the support for organizational tasks is currently missing from the medEmotion platform. The medEmotion platform does not enforce any data governance rules and does not offer any user-friendly user interfaces to manage data access rules. We put these organizational aspects out of scope, and we note that tackling these questions is essential to implementing the medEmotion platform in practice. During medEmotion, we focused on a proof-of-concept data integration platform. Hence, we did not tackle technical aspects to make the medEmotion platform production-ready, such as data redundancy or a multitier architecture. We acknowledge that these questions are important to answer in production environments and note that they are limited by a hospital's strategy, budget, and IT resources. Therefore, we do not present an answer to these aspects as they are highly hospital-specific and ultimately need to be decided per hospital.

In the remainder of this paper, we review the development process of the medEmotion platform.

Discussion

One of the key successes of the medEmotion project was its pioneering role in bringing together a diverse group of partner hospitals, academic institutions, and industry stakeholders, for the first time in a concrete, collaborative effort. This collaboration allowed us to gain invaluable insights into the distinct needs and concerns of various stakeholders, paraphrased as personas as mentioned before, ranging from data protection and privacy issues to the practical demands of health care practitioners. Through this partnership, we also learned a great deal about the organizational challenges that remain unaddressed in multistakeholder health data initiatives. To give 2 examples, we have seen that hospital-specific IT strategies vary significantly. While one hospital works within a cloud-based IT infrastructure, another prohibits using cloud-based components in favor of an exclusively on-premise infrastructure. The same can be said for data management strategies. Where

XSL•FO
RenderX

one hospital aims to keep data per department in distinct and tailored information systems, another aims to centralize as much information as possible into a single monolithic system. Personas apply these strategies in discussions but do not necessarily make them explicit. We believe that learning these implicit factors is vital to successful initiatives. Discussions about implementing a health data integration platform gave in-depth insights into hospital-specific data flows. We outlined how patient information flows into the hospital information system, an important exercise that improved mutual understanding. Importantly, beyond the technical knowledge gained, this project fostered personal connections and mutual trust among individuals representing different areas of expertise and roles within their organizations. This familiarity and trust built over time are essential for future interdisciplinary projects.

Clarifying technical questions using our demonstrators was pivotal in making abstract concepts more tangible for stakeholders. Much like trying to build a car in a world where cars do not yet exist, stakeholders initially struggled to articulate their needs without a clear vision of what the final product could look like. In such a setting, any implementation process will fail, as the result cannot fit the stakeholders' unarticulated vision. The demonstrators provided a concrete basis for discussions, allowing individuals without extensive technical backgrounds to actively engage and provide meaningful feedback. This practical approach significantly improved communication between the partners and helped bridge the gap between different personas. As such, we were able to develop and implement an architecture consisting of individual modules that both satisfied our demonstrators and generated general approval from the stakeholders.

Despite the many successes of the medEmotion project, several frustrations and challenges emerged throughout its course. Looking back, there were decisions and approaches that, with the benefit of hindsight, could have been handled differently to overcome obstacles more effectively. In this section, these key frustrations are summarized.

Effective Communication and Collaboration in a Fluid and Diverse Team Proved Extremely Challenging

With members frequently joining and leaving, maintaining continuity of knowledge was difficult, leading to disruptions in collaboration. The large and diverse group of individuals, each bringing distinct backgrounds, interests, and expertise, made aligning perspectives and ensuring consistent communication even more complex. This constant flux created coordination barriers, making it hard to sustain productive collaboration and slowing down decision-making processes.

Conflicting Priorities Among Key Stakeholders Created Significant Challenges in Aligning Goals and Decision-Making

Some stakeholders were primarily focused on return on investment and long-term strategic gains, while others were more concerned with its innovative potential and practical usability. This disconnect led to frequent misalignments, where decisions made often overlooked the technical needs or the full potential of the platform, or led to less effective outcomes as critical insights were missing in the decision-making process. As a result, in some cases, the project appeared to be driven more by academic interests rather than the practical needs of the health care partners. While the platform seemed promising on paper, there was a disconnect between the project's goals and the real-world priorities of the hospitals. For some hospitals, the platform was unsolicited and not aligned with their internal strategic plans, leading to doubts about its relevance. Additionally, one of the partner hospitals already had a similar platform in place, reducing the added value of the new system.

The Mismatch Between the Data Demands of Researchers and the Hospitals' Reluctance to Share Sensitive Data for Secondary use Complicated the Project

From the hospital's perspective, the process of preparing data cleaning, anonymizing, and transforming it represents a costly and time-consuming task. There is limited immediate benefit while they expose themselves to significant risks concerning the GDPR. Real-world data is highly sensitive and requires highly secured IT environments, which common academic research projects fail to consider. These conflicting needs slowed progress and exposed gaps in the project's ability to balance the expectations of different stakeholders, especially regarding the complexities of data sharing.

The Project Faced Significant Hurdles Related to Funding and Intellectual Property

The traditional "project-to-project" funding model limited the potential for long-term commitment. Without a clear financial roadmap beyond the initial phase, maintaining stakeholder engagement became increasingly difficult. At the same time, intellectual property (IP) issues arose as partners contributed in different ways, intellectually, financially, or through providing real-world use cases, making it complicated to define ownership. The absence of early, transparent IP agreements led to terms of usage that were unacceptable for some partners, limiting the potential for certain project components to be used effectively post the project. This further hindered the scalability and long-term success of the initiative.

Balancing the Openness of Collaboration With the Security and Operational Needs of Health Care Organizations Remains a Critical Challenge

Investing in open science can offer significant advantages in collaborative projects. Open-source software fosters transparency and inclusivity by allowing all partners to access, modify, and contribute to the project. This transparency builds trust among stakeholders as everyone can see how their contributions are used and built upon. Open science also reduces barriers to entry, enabling partners who may lack financial resources to still contribute valuable intellectual input, leveling the playing field, and encouraging broader collaboration. Open-source environments can also drive faster innovation, as contributors freely share ideas, experiment, and improve each other's work. This collaborative approach often accelerates development and produces more robust, well-tested solutions. Furthermore, open-source projects benefit from collective maintenance and support by the community, making them more

XSL•FO
RenderX

sustainable in the long term, even after the original funding or key contributors move on. However, there are pitfalls to consider. Open-source solutions, while valuable in academic settings, often conflict with hospital policies that prioritize stringent security measures and professional support. Hospitals may require more controlled environments to meet compliance and data privacy regulations, which may not always align with open-source principles.

Rapid Scale-Up and Professionalization Impacted the Balance Between Innovation and Implementation

A small group of researchers, who had initially focused on exploring and developing innovative solutions, had to quickly shift from this exploratory work to handling operational tasks as the project grew. This sudden change required them to manage the day-to-day implementation and ensure compliance with legal and technical requirements, leaving less time for creative research and experimentation. As a result, the researchers experienced a drop in motivation, finding themselves in presentations that felt more like sales pitches than discussions of research results. With no room for experimental approaches, the rapid scale-up and professionalization of the project brought innovation to a halt, a frustrating experience for those with a research mindset, who thrive on flexibility and exploration.

Focusing on Developing a Proof-of-Concept Limited the Production-Readiness of the Platform

The development of the medEmotion platform was bounded by time and resources. Initially focused on providing a minimal viable product, we approached implementation in little increments to provide quick and tangible parts of the platform. While this allowed for rapid development, our approach omitted important elements to make the platform ready for production. The use of mock datasets allowed us to push data privacy issues to the end, the cloud-based development left sensitive data flows from the hospital to the platform untested, and the rapid implementation start left us with no comparison metrics to assess the medEmotion platform user satisfaction. All these limitations persist and now need to be tackled post implementation, increasing the difficulty of bringing the medEmotion platform into production. As such, the medEmotion platform currently remains in a proof-of-concept state, not being implemented at any of our partner hospitals.

Takeaway Messages

We believe we are neither the first nor the last researchers to experience these or similar frustrations. Therefore, we share our experiences as takeaway messages in the following. Based on our experience, we aim to offer practical advice that can help guide future multistakeholder collaborative initiatives in health care and data integration.

Invest in Community Building, Stakeholder Engagement, and Education

Creating opportunities for stakeholders to get to know each other, both at the organizational level and on a personal level, is crucial for building trust and understanding. From an organizational perspective, this helps clarify the distinct needs of each stakeholder group and ensures that strategic goals are aligned toward a shared objective where value can be realized for everyone, even if the “return on investment” may look different for each party. On a more personal level, trust is not just an institutional concept; it is a feeling that grows between individuals and thus takes time. Providing opportunities for stakeholders to form personal connections through trustworthy relationships can significantly enhance collaboration. Organizing networking events, workshops, or joint educational sessions can help foster these connections and build a strong foundation of mutual trust and respect, which will pay dividends throughout the project.

Collaboration Among Health Care Providers Is Ideally Driven by the Health Care Providers Themselves

The health care providers are in the best position to proactively share expertise and requests for assistance based on their unique needs. Academic partners can play a key role in facilitating partnerships as neutral mediators, helping to bridge gaps between institutions. While commercial partners bring valuable expertise, maintaining a focus on shared goals and minimizing potential conflicts of interest can be achieved by ensuring that collaboration remains centered on the needs and priorities of health care providers. However, when it comes to implementation, there are additional challenges that need to be addressed, particularly within health care settings. Health care providers must establish secure data transfer protocols and ensure that patients are informed about how their data will be shared.

Ensure All Stakeholders Are Involved in the “Ideation Phase”

It is essential to spend adequate time understanding the problem and the real needs and expectations of all stakeholders and future users. One effective approach is to use qualitative research methodologies, such as semi-structured interviews and focus groups, before moving into the IT development phase. For example, methodologies from the newly launched educational program “System and Process Innovation in Healthcare” at Hasselt University could be leveraged to guide this process. In addition, involving legal departments early, particularly those within health care organizations is vital for ensuring compliance and managing risks effectively. It’s also critical to align IP ownership and a long-term funding strategy at the outset. Establishing clear IP agreements and a roadmap for continued financial support ensures the sustainability of the platform beyond the initial project phase. Finally, it is equally important to sanity-check technical plans with experts to guarantee feasibility, especially given the usual constraints of a fixed budget. Early input from all relevant parties helps prevent misalignments and avoids costly revisions later in the project.

Balance Innovation With Implementation

To ensure long-term success, it is important to strike a balance between pilot projects that foster innovation and the implementation of existing, proven solutions. These are fundamentally different activities, requiring distinct profiles and skill sets. Innovation thrives when motivated partners are eager to try new methods and solutions. This kind of exploratory work benefits from a small, focused group of individuals with

XSL•FO
RenderX

an innovation and research mindset who can freely experiment and iterate based on the real-world needs of partners. In this context, both parties gain from the trial-and-error process that allows them to explore potential breakthroughs in a mutually beneficial environment. Implementing novel methods and solutions requires a different approach, one focused on operational efficiency and compliance, rather than experimentation. It is also crucial to recognize that different people excel at different stages of the process. Some individuals thrive in the innovation phase, where flexibility and trial-and-error approaches are key. Others are better suited to perform structured, detailed work implementing established solutions. Attempting to blend these mindsets can lead to frustrations as the focus shifts from creative exploration to a more rigid, professionalized implementation. Understanding these distinctions and assigning the right people to the right phases can help maintain momentum and preserve the power of innovation while ensuring successful implementation.

Conclusions

Multistakeholder initiatives like the medEmotion project highlight the complexity of managing diverse roles, expectations, and goals. Stakeholders are rarely confined to a single role, and roles themselves are often shared among multiple individuals, which can complicate communication and decision-making. We introduce personas to clarify these different roles. Further, it is essential to communicate through full end-to-end use cases, ensuring that all parties understand how the platform functions within the broader ecosystem. In response, we introduce demonstrators to communicate technical requirements efficiently. The project highlighted the importance of building upon existing technical advancements rather than reinventing the wheel. Despite the availability of proven solutions, there was often a reluctance to fully embrace these innovations. This hesitation, driven by a desire to retain ownership of specific platforms or initiatives, leads to fragmented efforts and ultimately diminishes the overall impact. Collaboration, when done openly and transparently, benefits everyone involved, teamwork is crucial to achieving shared success. While technical sandboxes facilitate communication and experimentation, translating these ideas into real-world applications can be a slow and frustrating experience. The journey from concept to implementation requires patience and ongoing commitment. Ultimately, the lessons learned from medEmotion, building on existing successes, fostering collaboration, and maintaining a clear vision, lay the groundwork for more effective and sustainable initiatives in the future.

Acknowledgments

We thank our 3 partner hospitals Jessa Ziekenhuis, Noorderhart, and Ziekenhuis Oost-Limburg for their contributions toward the medEmotion project. Further, we thank the Limburg Clinical Research Center for sharing their expertise in clinical research projects. The software development of the medEmotion platform was funded by LRM, with the support of the European Regional Development Fund (EFRO-1308). This research received funding from the Flemish Government under the “Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen” program.

Authors' Contributions

All authors were contributed to conceptualization, methodology, investigation, writing–original draft, writing–review and editing. LMP handled supervision and funding acquisition.

Conflicts of Interest

None declared.

References

1. Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, Peek N, Sacchi L. Secondary use and analysis of big data collected for patient care. Yearb Med Inform 2017;26(01):28-37. [doi: 10.1055/s-0037-1606529]
2. Peng C, Goswami P, Bai G. A literature review of current technologies on health data integration for patient-centered health management. Health Informatics J 2020;26(3):1926-1951. [doi: 10.1177/1460458219892387] [Medline: 31884843]
3. Zhang H, Lyu T, Yin P, Bost S, He X, Guo Y, et al. A scoping review of semantic integration of health data and information. Int J Med Inform 2022;165:104834. [doi: 10.1016/j.ijmedinf.2022.104834] [Medline: 35863206]
4. U.S. Food and Drug Administration. U.S. food and drug administration. Framework for FDAs? Real-World Evidence Program. 2018. URL: https://www.fda.gov/media/120060/download?attachment [accessed 2024-01-09]
5. Ranchal R, Bastide P, Wang X, Gkoulalas-Divanis A, Mehra M, Bakthavachalam S, et al. Disrupting healthcare silos: addressing data volume, velocity and variety with a cloud-native healthcare data ingestion service. IEEE J. Biomed. Health Inform 2020;24(11):3182-3188. [doi: 10.1109/jbhi.2020.3001518]
6. Doan A, Halevy A, Ives ZG. Principles of data integration. Waltham, MA: Morgan Kaufmann; 2012.
7. Ortega-Calvo AS, Morcillo-Jimenez R, Fernandez-Basso C, Gutiérrez-Batista K, Vila MA, Martin-Bautista MJ. AIMDP: an artificial intelligence modern data platform. use case for spanish national health service data silo. Future Generation Computer Systems 2023;143:248-264. [doi: 10.1016/j.future.2023.02.002]

XSL•FO
RenderX

8. Wang M, Li S, Zheng T, Li N, Shi Q, Zhuo X, et al. Big data health care platform with multisource heterogeneous data integration and massive high-dimensional data governance for large hospitals: design, development, and application. JMIR Med Inform 2022;10(4):e36481. [doi: 10.2196/36481] [Medline: 35416792]
9. Poba-Nzaou P, Uwizeyemungu S, Dakouo M, Tchibozo A, Mboup B. Patterns of health information exchange strategies underlying health information technologies capabilities building. Health Syst (Basingstoke) 2022;11(3):211-231 [FREE Full text] [doi: 10.1080/20476965.2021.1952113] [Medline: 36147552]
10. Gagalova KK, Leon Elizalde MA, Portales-Casamar E, Görges M. What you need to know before implementing a clinical research data warehouse: comparative review of integrated data repositories in health care institutions. JMIR Form Res 2020;4(8):e17687 [FREE Full text] [doi: 10.2196/17687] [Medline: 32852280]
11. Cuggia M, Combes S. The french health data hub and the german medical informatics initiatives: two national projects to promote data sharing in healthcare. Yearb Med Inform 2019;28(1):195-202 [FREE Full text] [doi: 10.1055/s-0039-1677917] [Medline: 31419832]
12. Raj A, Bosch J, Olsson HH, Wang TJ. Modelling data pipelines. 2020 Presented at: 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA),; 2020 August 28; Portoroz. [doi: 10.1109/seaa51224.2020.00014]
13. Geys L, Peeters L. Strategic oversight across real-World Health Data initiatives in a complex health data space: a call for collective responsibility. Proc 17th Int Jt Conf Biomed Eng Syst Technol Rome, 2024:577-584. [doi: 10.5220/0012417700003657]
14. Observational Health Data Sciences and Informatics. The book of OHDSI: Observational Health Data Sciences and Informatics. San Bernardino, CA: OHDSI; 2019.
15. Park S, Bekemeier B, Flaxman AD. Understanding data use and preference of data visualization for public health professionals: a qualitative study. Public Health Nurs 2021;38(4):531-541. [doi: 10.1111/phn.12863] [Medline: 33569821]
16. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
17. Carroll N, Richardson I. Software-as-a-medical device: demystifying connected health regulations. JSIT 2016;18(2):186-215. [doi: 10.1108/jsit-07-2015-0061]
18. Yperman J, Becker T, Valkenborg D, Popescu V, Hellings N, Wijmeersch BV, et al. Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. BMC Neurol 2020;20(1):105 [FREE Full text] [doi: 10.1186/s12883-020-01672-w] [Medline: 32199461]
19. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019;94:103188 [FREE Full text] [doi: 10.1016/j.jbi.2019.103188] [Medline: 31063828]
20. Yperman J, Popescu V, Van Wijmeersch B, Becker T, Peeters LM. Motor evoked potentials for multiple sclerosis, a multiyear follow-up dataset. Sci Data 2022;9(1):207 [FREE Full text] [doi: 10.1038/s41597-022-01335-0] [Medline: 35577808]
21. Commowick O, Kain M, Casey R, Ameli R, Ferré JC, Kerbrat A, et al. Multiple sclerosis lesions segmentation from multiple experts: The MICCAI 2016 challenge dataset. Neuroimage 2021;244:118589 [FREE Full text] [doi: 10.1016/j.neuroimage.2021.118589] [Medline: 34563682]

Abbreviations

CDM: common data model
FHIR: Fast Healthcare Interoperability Resources
GDPR: General Data Protection Regulation
HIS: hospital information system
HIPAA: Health Insurance Portability and Accountability Act
HPC: high-performance computing
IP: intellectual property
MSDC: multiple sclerosis data connect
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership

XSL•FO
RenderX

Edited by A Mavragani; submitted 28.10.24; peer-reviewed by D Li, O Egunlae, A Adebisi; comments to author 10.12.24; revised version received 19.12.24; accepted 19.12.24; published 04.03.25.

Please cite as:
Parciak M, Pierlet N, Peeters LM
Empowering Health Care Actors to Contribute to the Implementation of Health Data Integration Platforms: Retrospective of the medEmotion Project
J Med Internet Res 2025;27:e68083
URL: https://www.jmir.org/2025/1/e68083
doi:10.2196/68083
PMID:40053761

©Marcel Parciak, Noëlla Pierlet, Liesbet M Peeters. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Opportunities and Challenges in Using Electronic Health Record Systems to Study Postacute Sequelae of SARS-CoV-2 Infection: Insights From the NIH RECOVER Initiative

Hannah L Mandel[1], MS; Shruti N Shah[1], BS; L Charles Bailey[2], MD, PhD; Thomas Carton[3], PhD; Yu Chen[1], PhD; Shari Esquenazi-Karonika[1], PhD; Melissa Haendel[4], PhD; Mady Hornig[5], MA, MD; Rainu Kaushal[6], MD; Carlos R Oliveira[7,8], MD; Alice A Perlowski[9], MD; Emily Pfaff[10], PhD; Suchitra Rao[11], MD; Hanieh Razzaghi[2], MPH, PhD; Elle Seibert[12], BA; Gelise L Thomas[13], JD; Mark G Weiner[6], MD; Lorna E Thorpe[1], PhD; Jasmin Divers[14], PhD; RECOVER EHR Cohort[15]

[1]Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States
[2]Applied Clinical Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA, United States
[3]Louisiana Public Health Institute, New Orleans, LA, United States
[4]Department of Genetics, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States
[5]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, United States
[6]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States
[7]Division of Infectious Diseases, Department of Pediatrics, Yale University School of Medicine, New Haven, CT, United States
[8]Division of Health Informatics, Department of Biostatistics, Yale University School of Public Health, New Haven, CT, United States
[9]Blooming Magnolia, Los Angeles, CA, United States
[10]Department of Medicine, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States
[11]Department of Pediatrics, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO, United States
[12]Department of Neuroscience, USC Dornsife College of Letters, Arts and Sciences, Los Angeles, CA, United States
[13]Clinical and Translational Science Collaborative of Northern Ohio, Case Western Reserve University, Cleveland, OH, United States
[14]Department of Foundations of Medicine, New York University Long Island School of Medicine, Mineola, NY, United States
[15]See Acknowledgments,

Corresponding Author:
Hannah L Mandel, MS
Department of Population Health
New York University Grossman School of Medicine
180 Madison Avenue
New York, NY, 10016
United States
Phone: 1 732 314 1595
Email: Hannah.Mandel@nyulangone.org

Abstract

The benefits and challenges of electronic health records (EHRs) as data sources for clinical and epidemiologic research have been well described. However, several factors are important to consider when using EHR data to study novel, emerging, and multifaceted conditions such as postacute sequelae of SARS-CoV-2 infection or long COVID. In this article, we present opportunities and challenges of using EHR data to improve our understanding of long COVID, based on lessons learned from the National Institutes of Health (NIH)–funded RECOVER (REsearching COVID to Enhance Recovery) Initiative, and suggest steps to maximize the usefulness of EHR data when performing long COVID research.

(J Med Internet Res 2025;27:e59217) doi:10.2196/59217

KEYWORDS

COVID-19; SARS-CoV-2; Long COVID, post-acute COVID-19 syndrome; electronic health records; machine learning; public health surveillance; post-infection syndrome; medical informatics; electronic medical record; electronic health record network;

XSL•FO
RenderX

electronic health record data; clinical research network; clinical data research network; common data model; digital health; infection; respiratory; infectious; epidemiological; pandemic

Introduction

Postacute sequelae of SARS-CoV-2 infection, colloquially known as long COVID, refers to ongoing, relapsing, new symptoms, or other health effects after the acute phase of SARS-CoV-2 infection (ie, present 4 or more weeks after the acute infection). Because long COVID is heterogeneous in presentation and may occur after a mild or even asymptomatic infection, it has posed unique and significant challenges for clinical research, including an overarching lack of data that are accessible and analyzable in a rigorous and reproducible manner to inform diagnostic criteria and care guidelines.

SARS-CoV-2 continues to impact populations globally with new and recurrent infections. While studies have provided a wide range of estimates regarding the proportion of patients with COVID-19 who develop long COVID [1], even conservative estimates point toward an enormous burden. Without established treatments targeting underlying pathophysiologic mechanisms, patients and clinicians have largely focused on treating symptoms [2] and managing organ damage, highlighting an urgent need to expand our knowledge of COVID-19 infection's long-term effects.

Electronic health records (EHRs) contain large quantities of clinical data that can be used for research without significant delay, making them an important data source for accelerated insight into emerging health conditions. However, challenges associated with leveraging EHR data for research can become amplified when mobilizing rapidly to study these conditions, particularly when they are multifaceted and lack clear hallmark traits or biomarkers.

The National Institutes of Health (NIH)–funded REsearching COVID to Enhance Recovery (RECOVER) Initiative was launched in 2021 [3] to advance research into long COVID. Broadly, RECOVER supports prospective clinical studies of adult and pediatric cohorts and real-world data studies that leverage EHR networks to study long COVID. These studies have been developing and validating algorithms to identify patients with long COVID within EHR networks for clinical and epidemiological characterization, risk factor prediction, and identification of prevention and treatment opportunities. RECOVER's real-world data efforts are led by 3 participating EHR research networks: the National COVID Cohort Collaborative (N3C); the National Patient-Centered Clinical Research Network (PCORnet); and PEDSnet, a pediatric learning health system within PCORnet. These networks are coordinated by a Clinical Science Core (CSC) at NYU Langone Health.

In this paper, we highlight opportunities and challenges of using EHR data to improve our understanding of long COVID, with a focus on analytical resources and scenarios we've encountered as US-based researchers, and conclude with some next steps for maximizing the usefulness of EHR data for long COVID research. Similar considerations may pertain to other chronic infection–associated diseases, that is, complex multisystemic conditions often occurring in the context of infection with other pathogens, including myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).

EHR-Based Research Networks

Overview

The widespread adoption of heterogeneous health information systems across the United States hinders large-scale EHR-based research efforts. To overcome this, federal funding from the Patient-Centered Outcomes Research Institute (PCORI), Centers for Disease Control and Prevention (CDC), and NIH has facilitated the growth of networks that transform data into a common data model to facilitate aggregation across participating institutions. These networks are often referred to as clinical research networks (CRNs).

While such networks are a focus of this paper, we recognize that other successful international [4-7] and national models have been leveraged for large-scale long COVID studies. Domestically, the Department of Veterans Affairs (VA), which operates the largest integrated health care system in the United States, has made important contributions characterizing long COVID, its burden, and risk after reinfection, immunization or nirmatrelvir (Paxlovid) use [8-11]. The interoperability across many VA hospitals and clinics allows for data aggregation without the extensive harmonization or loss of data depth that usually accompanies a research network, though VA patients (military experience, mostly men) are not representative of the general population. Similarly, researchers have used claims data and outpatient laboratory results from UnitedHealth, a large health insurance provider, to characterize risk of long COVID in commercially insured patients [12]. The addition of claims data enables better understanding of health care use, inequities in access to care and costs associated with long COVID, though quantifying the larger economic burdens and disparities requires incorporation of data sources outside the health care system.

While researchers with access to EHR data from large clinical networks or health systems may bypass some challenges below, many national and international EHR-based collaborations wrestle with similar obstacles.

Opportunities

CRNs provide several benefits for conducting clinical research. Most significantly, they unify patient data that may otherwise be distributed across multiple health care facilities and facilitate access to large sets of timely, longitudinal patient information. This enables broad hypothesis generation, signal identification, and feasibility analysis, which is highly advantageous for emerging conditions like long COVID as such analyses can be conducted at a scale, pace, and cost that surpass traditional research methods [13]. Because CRNs use data generated during health care delivery, they are also well-positioned to study how practices evolve in ways that directly impact patients. For example, EHR data can be used to evaluate uptake and patterns of use for new diagnostic codes introduced to capture long

XSL•FO
RenderX

COVID [14], ME/CFS, and postural orthostatic tachycardia syndrome (POTS).

To overcome the lack of semantic interoperability across EHR systems and facilitate analysis, most CRNs map source data to common data schemas, such as the Observational Medical Outcomes Partnership (OMOP) [15] or PCORnet common data models [16], which provide frameworks for data standardization and are optimized for large-scale longitudinal data analysis. The popularity and extensibility of common data models helped CRNs to quickly adapt them for novel disease research at the beginning of the pandemic [17]. Use of a common data model also enables distributed network queries and analyses behind institutional firewalls, as well as centralization of the data for advanced artificial intelligence and machine learning [18].

In a relatively short time, a large amount of scientific literature has been generated regarding the clinical characterization and epidemiology of long COVID, yet efforts to synthesize this body of research have been hampered by the inconsistency of methodologic approaches, study definitions, and findings. Research conducted using a shared infrastructure or data model promotes reproducible research and exchange of clinical and variable definitions that can be harmonized and tested across multiple health systems. For example, a recent collaboration between RECOVER and NIH's All of Us study demonstrated success in reusing long COVID definitions across OMOP data environments, enabling reproduction of N3C's definition within the All of Us database [19].

Along the same lines, CRNs are often well-positioned to foster multidisciplinary collaboration [20]. They have streamlined contracting and collaboration processes that expedite research while reducing regulatory burden, and can also provide an efficient vehicle for patient recruitment into prospective cohort studies and clinical trials.

Challenges

Some broad challenges come with using CRNs for population health research and may directly impact research into long COVID, where data quality, depth, and timeliness are crucial. These largely arise from the need to standardize data across multiple institutions, which may result in a loss of detail and context from the originating EHR [21]. For example, while the OMOP and PCORnet data models support documentation of patient cause of death and discharge disposition, not all do; furthermore, not all sites collect and report these data. Analyses can also be limited by not having information such as provider specialty, medication indication, or drug dosing consistently abstracted from the EHR. However, as described earlier, the extensibility of many common data models means that new concepts can be incorporated to close priority gaps.

Aggregating data within a CRN can mask important heterogeneity due to differences in how health systems and individual practitioners provide and document care. One example is the variation in how visit data are represented across individual health care facilities. The definition of an "encounter" in a given health system's data has more to do with business rules than clinical care; for example, some systems create a separate encounter record for each day of an inpatient stay, while others use a single record to capture the entire hospitalization. As most variables in EHR data are captured at the visit level, inconsistency in the definition of encounters can lead to errors when analyzing clinical observations. Thus, to accurately classify types of visits and to quantify visits across sites, careful standardization is required [22]. Ongoing governance, documentation, and input from data experts at contributing institutions are essential to bridge syntactic and semantic differences [21].

Timeliness of data delivery represents an additional challenge. While data from CRNs are generally rapidly available, there is still a latency from participating institutions. Data must be mapped, transformed, transmitted, quality checked, and potentially resubmitted with corrections, which takes time and collaborative effort. Delays may also result from the need to incorporate data standards for new concepts (such as the long COVID *ICD-10* [*International Statistical Classification of Diseases, Tenth Revision*] code U09.9, which was not introduced until late 2021).

Finally, real-world data present specific limitations compared with traditional research data. These include data missingness, which has been described extensively [23] and can exist for various reasons (eg, patient discontinuation of care or data points not being systematically captured in EHRs). EHR data are also subject to biases like selection bias [24], where patients accessing care may be sicker, more likely to be insured, and able to take time off work and afford transportation [13,25]. Data from nonacademic institutions, and other settings that may lack robust data systems such as nursing homes, are often underrepresented in CRNs. Careful study designs, informatics-informed approaches to transforming raw clinical data into analytical datasets [26], techniques to supplement structured EHR data with items extracted from free text notes such as social determinants [27], and advanced statistical approaches are often required to minimize bias.

Identifying Patients With Long COVID Within EHRs

Overview

Robust case definitions are essential to disease surveillance, enabling consistent use of data from collection through analysis and across population subgroups, time, and geographic jurisdictions. However, Long COVID still lacks a uniform case definition, and multiple frameworks have emerged with varying timeframes and symptoms [25]. This is partly due to the heterogeneity of long COVID's clinical presentation, which includes substantial differences between adults and children [28] and hundreds of potential signs and symptoms. The National Academies of Sciences, Engineering, and Medicine's recent report A Long COVID Definition provides updated recommendations around diagnosis, noting that long COVID can impact any organ system; follow acute infection of any severity, including asymptomatic infection; be continuous from acute infection or delayed in onset; and present as incident or worsening, mild or severe, and persistent or intermittent [29]. These broad guidelines leave room for ambiguity, and as there

XSL•FO
RenderX

is no diagnostic test for long COVID, their application at the point of care involves clinical judgement, experience, and thorough differential diagnosis to establish probable causality.

Finally, we are still learning about long COVID's disease course and trajectory, how it relates to development of clinical conditions such as diabetes, kidney disease, and POTS, and how its characteristics may evolve over time. While these complexities underlie all long COVID research, EHR-based research is particularly well-positioned to inform and validate potential definitions.

Opportunities

Researchers working with EHR data operationalize case definitions by creating computable phenotypes, which algorithmically identify cohorts of interest within an EHR system [30]. Generating, adapting, and validating computable phenotypes can be complex, even for well-characterized diseases and conditions, so multiple approaches for identifying patients with long COVID have evolved both nationally and internationally. These include building computable phenotypes using "rules-based" approaches or prespecified logic-based queries to search for specific criteria such as diagnosis codes, medications, or laboratory results based on clinical literature and expertise [31-33], and data-driven or inductive approaches, including machine learning (ML) to automatically classify long COVID without detailed prespecification of the inclusion or exclusion criteria [34,35]. These approaches can also be combined, leveraging insights from ML to shape rules-based definitions.

The lack of a formal case definition, limited SARS-CoV-2 testing information, and large amounts of rich longitudinal data within EHRs make ML approaches particularly attractive. Supervised ML models can be trained on sets of patients that are certain or highly likely to have long COVID, such as patients assigned a long COVID diagnostic code, and then used to identify patients with probable long COVID who may not have a documented or readily recognized diagnosis [35]. ML and other statistical approaches have also been applied to recognize incident diagnoses that appear more frequently in cohorts with a recent history of COVID-19 infection than those without [36,37].

Although the EHR does not capture every case (or every recurrence) of acute SARS-CoV-2 among its patients, there is still significant opportunity to use EHRs to study the potential effect of SARS-CoV-2 reinfection on risk of long COVID among those patients with multiple recorded cases. Hadley et al [38] found that across SARS-CoV-2 variants, the incidence of new long COVID diagnoses after reinfection is lower than after the initial SARS-CoV-2 infection. Bowe et al [11] found that though risk is lower with more recent variants, it is cumulative across infections. As multiple reinfections become more common, there will be continual opportunities to reevaluate these findings in the context of new variants and subvariants.

Both rules-based and ML techniques are being used to identify potential clusters or subphenotypes of long COVID [39-41]. For example, an analysis by Zhang et al [41] identified four distinct subphenotypes (including cardiac and renal, musculoskeletal, and nervous system) and specific risk predictors for each. Clustering analyses may uncover diverse symptoms that have common biological causes, and in establishing distinct groupings around them, facilitate diagnosis and treatment targeted toward specific subtypes [42].

Challenges

Within the EHR, Long COVID may not be well-documented due to diagnostic complexities, as described earlier. The initial focus among clinicians on respiratory symptoms and widespread lack of knowledge around what were initially considered to be "atypical" presentations, such as postexertional malaise, ME/CFS, POTS, and dysautonomia, also contribute to missingness of relevant diagnoses within the EHR [43], which is compounded by the absence of or delay in introduction of *ICD-10* codes for many of these conditions, including the U09.9 code for long COVID itself**.** Such manifestations can significantly impact daily function but may not be considered attributable to previous COVID-19 infection (or misdiagnosed as mental health issues), leading to their underrepresentation in research [43].

Health care professionals may be subject to implicit biases that impact documentation of patient information (eg, social determinants and symptoms) within the EHR [24]. This may also lead to underestimation of long COVID for certain populations, such as racial or ethnic minorities. Indeed, U09.9 diagnostic codes have been found to be more common among patients that are non-Hispanic White and living in areas of low poverty compared with the overall COVID-19–positive population, potentially reflecting differences in access to care and other interactions with the health care system [14]. Although we found value in using visits to a long COVID clinic as an early proxy for a long COVID diagnosis code [35], both of these markers have biases and should not be considered a gold standard for long COVID identification [44]. There is also substantial need to complement patient-reported symptoms with clinically characterized ones [45] so that physicians are not "gatekeepers" of documented symptoms.

Given that long COVID is likely underdiagnosed and underdocumented, computable phenotype algorithms must supplement formal U09.9 diagnoses with other patient data following acute SARS-CoV-2 infection. However, identifying acute infections is also not straightforward. Diagnosis codes such as the ICD-10 code U07.1 have demonstrated variable sensitivity and positive predictive value [46,47], which may be exacerbated by "rule out" diagnoses or improper coding. Even when diagnoses are correct, the true timing of the initial infection with respect to the date of the diagnosis code may be uncertain [44], and we found this to complicate EHR-based calculations of long COVID incidence. Limited availability of SARS-CoV-2 testing during the early phase of the pandemic, widespread testing at nonclinical settings (eg, kiosks and pharmacies), and the later popularity of at-home tests have all contributed to low confidence in accurately capturing SARS-CoV-2 infection (and reinfection) history within the EHR, and amplified potential misclassification of long COVID patients without documented acute COVID-19 infection. We

XSL•FO
RenderX

note, however, that misclassification of patients as uninfected is an issue that extends beyond EHR-based studies [48], broadly impacting long COVID research.

ML methods are also subject to obstacles, including the lack of a validated test set for long COVID and dependency of results on the training set used, which may limit reproducibility (although, as demonstrated by the collaboration between RECOVER and All of Us, ML algorithms can be replicated in new environments) [19]. ML models must be interpretable, so predictions can be explained instead of only existing within a "black box."

Determining Risk Factors for Long COVID Using EHRs

Risk factors for long COVID include demographic characteristics, lifestyle behaviors, underlying comorbidities, and social and environmental factors. Identifying risk factors can guide screening and prevention efforts, particularly in highly susceptible groups, and potentially facilitate tailored long COVID treatments.

Opportunities

EHR-based studies are primed to quickly reveal possible associations and risk factors for conditions like long COVID due to large volumes of available data and ability to stratify by various subgroups. For example, Rao et al [28] leveraged PEDSnet's geographical and clinical breadth to identify risk factors such as age, severity of acute COVID-19 infection, and comorbid conditions. Potential biological variables that are well-defined in EHRs and therefore, more reliably available for application in adjusted analyses include age, sex, BMI, and blood pressure, and, indeed, these variables have been shown to confound the evaluation of causal pathways for long COVID [49-51].

Where data of interest are missing from the EHR, linkage to external sources can increase data quality and completeness. Statewide vaccination or vital record data can be acquired and linked at the patient or area level to enhance the accuracy of vaccine status and COVID-19 outcomes [52,53]. Linkage with geographic information systems (GIS) can address questions about complex networks of long COVID burden that pertain to certain geographic regions, including the so-called "exposome" [54] even down to ZIP codes and neighborhoods [13,14]. Patient-reported outcomes from patient portals and surveys can improve the completeness of information on demographic, social, and behavioral factors such as race, ethnicity, substance use, exercise frequency, and smoking status [13,55]. To enhance completeness of data such as outpatient medication use, linkage to claims data can be an effective strategy. Data of interest may also be captured within free-text clinical notes, images, and other scanned documents, requiring application of natural language processing tools, including cutting-edge language models [27], or manual chart review to incorporate them.

Within EHR-based studies, several analytic approaches can address confounding and effect modification and help measure mediating influences. Researchers can match comparison groups using various approaches, including propensity scores for known confounders. In addition, causal mediation analysis methods may more accurately explain observed associations between risk factors and long COVID development.

Designing studies with appropriate comparison groups for analyses of risk factors can be challenging. If comparison groups are matched too closely, relevant differences in concurrent risk factors and outcomes may be masked. If groups are not matched well enough, then the impact of confounders may overwhelm the ability to detect differences associated with the characteristic of interest. Researchers using real-world data can conduct sensitivity analyses under different assumptions of risks and outcomes. For example, Hill et al [56] leveraged 3 different control groups and 2 definitions of long COVID to enhance an investigation of risk factors within N3C. This agility is especially useful with a novel condition such as long COVID, where our understanding is rapidly evolving. Comparison groups can be iteratively modified, and analyses rerun as new information or criteria emerge.

EHR data, especially within the context of a CRN, also facilitates the study of risk factors within rare populations and subgroups (eg, peoples indigenous to the United States). Such groups often experience a disproportionate burden of COVID-19 complications due to socioeconomic and structural factors [51,57], but are generally not well-represented in prospective studies and randomized controlled trials [58]. EHR data can similarly help overcome the methodological constraints and recruitment challenges of rare disease research [13], providing an opportunity to study rare conditions and outcomes and facilitating clinical trial cohort selection.

Finally, EHRs provide rich datasets for pragmatic clinical trials and target trial emulation to identify potential therapies [59,60]. For example, there is inconclusive data on the efficacy of Paxlovid for preventing or treating long COVID [61]; novel approaches such as target trial emulation using EHR data can help evaluate its efficacy and determine which patients should be prioritized for treatment based on risk factors for severe long COVID subphenotypes. The ability to quickly evaluate and personalize potential treatments is highly advantageous given the large burden of long COVID and its heterogeneous presentation. Because target trial emulation better reflects real-world medication adherence (or application of other interventions), comparing the results of emulated and clinical trials, and examining how the methods and populations differ, may yield insights into generalizability of clinical trial results and sources of bias within observational analyses.

Challenges

Limitations outlined earlier, such as data missingness, selection bias, and implicit biases, can directly impact the ability to evaluate risk factors using EHR-based analyses. Potential risk factors can be rigorously evaluated if well-defined and well-captured, but many important risk factors may not be completely, meaningfully, or consistently recorded within the EHR.

For example, race and ethnicity are important demographic risk factors for long COVID [51], but these data can be missing in

XSL•FO
RenderX

over 25% of patient records [55,62]. Lack of robust, structured data on social and behavioral determinants of health, such as physical activity levels, smoking status, and alcohol use [13,51,62] may necessitate supplementation with survey results or data extracted from clinical notes as described above. EHR-based assessments of symptom severity and improvement may also be of limited use without these approaches.

Imperfect data capture also impacts our ability to establish proper comparison groups. This notably includes the shrinking population of patients without previous SARS-CoV-2 infection [63]. In addition to being challenging to identify, control patients identified by negative lab tests may actually be in worse health than those with mild COVID-19 infection, likely due to routine testing of patients within the emergency department or inpatient settings. Another example applies to our work examining obstructive sleep apnea (OSA) as a risk factor for long COVID [64]. A comparison group of patients ostensibly without OSA may include some with undiagnosed, subclinical, or undocumented OSA, and sizable numbers of misclassified patients would lead to underestimation of effect size on long COVID. Similarly, slight improvements in recognition of ME/CFS throughout the pandemic may have resulted in some individuals with prepandemic long COVID–like symptoms (eg, unrefreshing sleep and postexertional malaise) having their illness misclassified as SARS-CoV-2–related. Appropriate comparison groups must be carefully considered.

The pandemic disrupted typical health care use and documentation patterns. Clinic closures and reduced access to specialty care may have masked diagnoses of chronic conditions and comorbidities and led some patients with long COVID, particularly those with milder presentations, not to seek care. This biases EHR data, especially from the first year of the pandemic, toward underascertainment of health issues; prepandemic controls (who we can say with certainty have not had COVID-19 infection) may have higher rates of clinical events than pandemic-era patients. Similarly, research about the effectiveness of vaccinations in reducing the risk of long COVID is hampered by widespread distribution in nonclinical settings [65] and reinforces the importance of ongoing data exchange between EHRs and vaccination registries in addition to dedicated linkage efforts.

These challenges also apply to potential confounding, which is particularly important within observational research on risk factors and causal mechanistic pathways for long COVID. Certain potential confounders, such as age or underlying health conditions, are easier to adjust for than others. The severity of acute COVID-19 infection, which appears in some studies to be directly associated with the risk of developing long COVID [49,50,66], is a particularly illustrative example. More severe COVID-19 infection, whether it results in hospitalization or not, may result in increased frequency of diagnosis and monitoring of long COVID and higher data density for affected patients within the EHR. These biases may contribute to the underrepresentation of asymptomatic, mild, or moderate acute COVID-19 infections, as well as greater clinician awareness of long COVID in the context of severe illness (although most cases occur among nonhospitalized patients) [43]. Analyses stratified by acute illness severity are important to differentiate between nonhospitalized and hospitalized patients.

Discussion

As SARS-CoV-2 and our susceptibility to infection evolve, sustained research efforts are needed to build our knowledge of long COVID and evaluate its presentation over time. Within a short timeframe, the use of EHR networks to study long COVID has grown rapidly, nationally and internationally. The immediate need to advance our understanding of long COVID must be balanced with a thoughtful methodological approach that leverages the many strengths of EHR data while recognizing their limitations. EHRs have been used to characterize long COVID and its subphenotypes [14,32,40,41,45], estimate its burden [28,67], identify risk factors [49-51,66], and evaluate SARS-CoV-2 vaccine efficacy for long COVID prevention [8,53,68]. Although we anticipate that the breadth and rigor of this research will increase as computable phenotypes improve through iterative validation and data linkage becomes more extensive, several innovations to effectively wield EHR data, such as improved accuracy and consistency of recording of long COVID-specific diagnoses, application of ML to identify patients with long COVID and harmonized data analyses across CRNs, have already begun increasing the capacity of investigators to study long COVID and, importantly, augment research on long COVID subphenotypes.

EHRs can be instrumental in supplementing randomized controlled trials to evaluate treatments and preventative strategies such as vaccinations, but novel approaches such as target trial emulation have not yet widely extended to long COVID. Increased data linkage to vaccination registries, survey data, and other sources can further enrich EHR-based analyses, providing more complete patient outcomes and medical history data. As recruitment into prospective RECOVER studies increases, efforts are underway to link EHR data with participant study data, enhancing both datasets for validation and analysis.

We envision a near future where the mutual benefits of clinical and real-world data research collaborations have been optimized for synergistic learning. EHRs can identify study participants and generate hypotheses that can be prospectively tested. They can also inform clinical trial feasibility, generate preliminary data around specific populations, therapeutics, and outcomes of interest, and inform the interpretation of clinical trial results through trial emulation. However, strengthening this interchange will require new workflows, collaboration, and time.

We caution that despite the dedication of the research community and high expectations for EHRs to quickly advance our understanding of long COVID, sustained, collaborative efforts are needed for this work. New treatments and vaccines are still being introduced, requiring process modifications for extraction and standardization of data. Complex clinical traits require careful development of computable phenotypes and rigorous data quality evaluation to identify specific misclassification risks. Continued coordination is necessary to converge upon reliable, repeatable insights and yield effective strategies for long COVID.

XSL•FO
RenderX

Acknowledgments

This study is part of the NIH RECOVER (National Institutes of Health's REsearching COVID to Enhance Recovery) Initiative, which seeks to understand, treat, and prevent the postacute sequelae of SARS-CoV-2 infection. For more information please visit the website [69]. This research was funded by an NIH Agreement (OTA OT2HL161847) as part of the RECOVER research program. We would like to thank the National Community Engagement Group (NCEG), all patient, caregiver, and community representatives, and all the participants enrolled in the RECOVER Initiative.

We would also like to thank collaborators in the RECOVER Consortium[discontinued effort on RECOVER].

National COVID Cohort Collaborative (N3C):

Axle Informatics: Kristen Hansen. Berkeley Lab: Justin Reese. Columbia University: Karthik Natarajan. Datavant: Jasmin Phua. Duke University: Warren Kibbe, PI. Emory University: Richard Moffitt, PI, Margaret Hall, Rishi Kamaleswaran, Jason Yoo. Endeavor Health: Anthony Solomonides, PI. Illinois State University: Nariman Ammar. Jackson Labs: Hannah Blau, Peter Robinson. Johns Hopkins University: Christopher G. Chute, PI, Ali Afshar, G. Caleb Alexander, Lisa Eskenazi, Tricia Francis, Davera Gabriel, Kirby Gong, Stephanie Hong, Harold Lehmann, Hemal Mehta, Chirag Parikh, Ann Parker, Rayna Xiao, Tanner Zhang, Richard Zhu, Jared Zook. Minderoo Center for Federated Cancer Research: Robert Miller. National Institutes of Health: Hythem Sidky. Northeastern University: Christian Reich, PI, Kristin Kostka. Oregon Community Health Information Network (OCHIN): Brenda McGrath, PI, Treasure Allen, Rob Schuff. Oregon Health & Science University: Erik Benton, David A. Dorr, Justin Ramsdill. Palantir Technologies: Maya Choudhury, Andrew Girvin. Patient Lead Research Collaborative: Hannah Davis, PI, Gina Assaf, Lisa McCorkell, Yochai Re'em, Anisha Sekar, Hannah Wei. RTI International: Daniel Brannock, Rob Chew, Emily Hadley, Alexander Preiss. Scripps: Ginger Tsueng. State University of New York at Stony Brook: Janos Hajagos, PI, Rachel Wong, PI, Adit Anand, Raj Gupta, Mengyao Hu, Prahathish Kameswaran, Saarthak Kapse, Eileen Keck, Farrukh koraishy, Spencer Krichevsky, Saaya Patel, Joel Saltz, Mary Saltz, Hiteshwar Singh, Shreya Sinha, Sam Soff, Kimon Stathakos, Chelsea Twan, Rohith Vaddavalli, Sai Rachana Yerram. The National Institute of Diabetes and Digestive and Kidney Diseases: Kenneth J. Wilkins. TriNetX: Lora Lingrey, Matvey Palchuk. Trinity Health: Joe Flack. Tufts Medical Center: Andrew Williams, PI. University of California Davis: Konstantin Kunze, PI. University of Chicago: Tom Best, PI. University of Iowa: Dave Eichmann, PI, Charisse Madlock. University of Kansas Medical Center: Kelechi Anuforo, PI. University of Kentucky: Ramakanth Kavuluru. University of Maryland Baltimore: Stacy Dalton, PI. University of Massachusetts: Feifan Liu, PI. University of Michigan: J. Brian Byrd. University of Minnesota: Steve Johnson, PI, Ashley Benner, Carolyn Bramante, Scott Chapman, Duy Duong, Michael Evans, Jared Huling, Corey McGee, Zheng Wang, Talia Wiggen, Rui Zhang. University of Nebraska Medical Center: Jerrod Anzalone. University of North Carolina at Chapel Hill: Emily Pfaff, PI, Melissa Haendel, PI, Abhishek Bhatia, Sofia Dard, Liz Kelly, Peter Leese, Tomas McIntee, JP Powers, Bryan Laraway, Julie McMurry, Shawn T. O'Neil, Chris Roeder, Anita Walden. University of Pittsburgh: Michele Morris. University of Rochester: Elaine Hill, PI, Jack Chang, Adam Dziorny, Daniel Guth, Tanzy Love, Klint Mane, Sharad Kumar Singh, Richa Yadav, Ayushi Saxena. University of Texas Medical Branch: Heidi Spratt, PI. University of Utah: Jackson Barlocker. University of Virginia: Don Brown, PI, Sihang Jiang, Johanna Loomba, Saurav Sengupta, Suchetha Sharma, Andrea Zhou. University of Washington: Rena Patel. University of Texas Health Science Center at Houston: Hongfang Liu, PI. Virginia Commonwealth University: Brian Bush, PI.

PCORnet Core Contributors:

Ann & Robert H. Lurie Children's Hospital of Chicago: Ravi Jhaveri. Children's Hospital of Philadelphia: L. Charles Bailey, PI, Christopher B. Forrest, PI, Andrea Allen, Rodrigo Azuero-Dajud, Andrew Samuel Boss, Morgan Botdorf, Colleen Byrne, Peter Camacho, Abigail Case, Kimberley Dickinson, Susan Hague, Jonathan Harvell, Miranda Higginbotham, Kathryn Hirabayshi, Sandra Ilunga, Rochelle Jordan, Kyle Kays, Aqsa Khan, Vitaly Lorman, Nicole Marchesani, Sahal Master, Jill McDonald, Nhat Nguyen, Hanieh Razzaghi, Qiwei Shen, Alexander Shorrock, Levon H. Utidjian, Ryan Webb, Kaleigh Wieand. Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine: Nate M. Pajor. Harvard Pilgram: Jason Block. Louisiana Public Health Institute: Tom Carton, PI, Anna Legrand, Elizabeth Nauman. Nationwide Children's Hospital and The Ohio State University College of Medicine: Jennifer Muszynski. Nemours/Alfred I. duPont Hospital for Children: H. Timothy Bunnell, Christopher Pennington, Cara Reedy. The Perelman School of Medicine, University of Pennsylvania: Yong Chen. University of Colorado School of Medicine and Children's Hospital Colorado: Suchitra Rao, PI. Vanderbilt University Medical Center: Russell Rothman, PI. Weill Cornell Medicine: Rainu Kaushal, PI, Sajjad Abedian, Dominique Brown, Christopher Cameron, Thomas Campion, Andrea Cohen, Marietou Dione, Rosie Ferris, Wilson Jacobs, Michael Koropsak, Alex LaMar, Colby V. Lewis, Dmitry Morozyuk, Peter Morrisey, Duncan Orlander, Jyotishman Pathak, Mahfuza Sabiha, Edward J. Schenck, Stephenson Strobel, Zoe Verzani, Fei Wang, Mark Weiner, Zhenxing Xu, Chengxi Zang, Yongkang Zhang.

Observational Consortium Steering Committee:

XSL•FO
RenderX

Jeffrey Burns, Co-Chair; Serena Spudich, Co-Chair; Charles Bailey, Mine Cicek, Melissa M. Cortez, Felicia Davis Blakley, Andrea S. Foulkes, David Goff, Stuart D. Katz, Jessica Lasky-Su, Torri D. Metz, Lisa T. Newman, Igho Ofotokun, Sudha Seshadri, Melissa Stockwell, James Stone, Brittany D. Taylor, PJ Utz, Neely A. Williams.

Administrative Coordinating Center at Research Triangle Institute International:

Lisa T. Newman, PI, Julie Abella, Quinn Barnette, Christine Bevc, Jennifer Beverly, Patricia Ceger, Julie Croxford, Emily Cunningham, Mike Enger, Katie Fain, Tonya Farris, Sean Hanlon, David Hines, Vicki Johnson-Lawrence, Kevin Jordan, Craig Lefebvre, Beth Linas, Bryan Luukinen, Meisha Mandal, Nikki J. McKoy, Susan Nance, Ashleigh Oakland, Demian Pasquarelli, Claire Quiner, Rita Sembajwe, Gwendolyn Shaw, Vanessa Thornburg, Kendall Tosco, Hannah Wright.

Clinical Science Core at NYU Langone Health:

Rachel S. Gross, PI; Judith S. Hochman, PI; Leora I. Horwitz, PI; Stuart D. Katz, PI; Andrea B. Troxel, PI; Lenard Adler, Precious Akinbo, Ramona Almenana, Malate Aschalew, Lara Balick, Ola Bello, Sultana Bhuiyan, Nina Blachman, Ryan Branski, Jasmine Briscoe, Shari Brosnahan, Elliott Bueler, Yvette Burgos, Nina Caplin, Domonique Nicole Chaplin, Yu Chen, Shen Cheng, Peter Choe, Jess Choi, Alicia Chung, Richard Church, Stanley Cobos, Nakia Croft, Angelique Cruz Irving, Phoebe Del Boccio, Iván Díaz, Jasmin Divers, Vishal Doshi, Benard Dreyer, Samantha Ebel, Shari Esquenazi-Karonika, Arline Faustin, Elias Febres, Jeffrey Fine, Sandra Fink, Catherine Freeland, Jennifer Frontera, Richard Gallagher, Alejandra Gonzalez-Duarte, Denise Hasson, Sophia Hill, Jennifer Hossain, Shahidul Islam, Stephen Johnson, Neha Kansal, Rachel Kenney, Tammy Kershner, Deepshikha Kewlani, Judy Kwak, Michelle F. Lamendola-Essel, Sarah Laury, Gregory Laynor, Lei Lei, Terry Leon, Zoe A. Lewczak, Janelle Linton, Max Logan, Nadia Malik, Lia Mamistvalova, Hannah Mandel, Gabrielle Maranga, Patenne Dolores Mathews, Aprajita Mattoo, Tony Mei, Alan Mendelsohn, Emmanuelle Mercier, Patricio Millar Vernetti, Marc Miller, Maika Mitchell, Andre Moreira, Praveen C. Mudumbi, Erica Nahin, Nandini Nair, Joseph Nekulak, Kellie Owens, Brendan Parent, Nandan Patibandla, Peter Petrov, Radu Postelnicu, Francesca Pratt, Isabelle Randall, Priyatha Rao, Amy Rapkiewicz, JohnRoss Rizzo, Johana Rosas, Chelsea Rose, Christina Saint-Jean, Michelle Santacatterina, Binita Shah, Aasma Shaukat, Naomi Simon, Aylin Simsir, Miranda Stinson, Wenfei Tang, Vasishta Tatapudi, Sujata Thawani, Mary Thomas, Lorna Thorpe, MeeLee Tom, Ethan Treiha, Jennifer Truong, Mmekom Udosen, Carlos Valencia, Jessica Velazquez-Perez, Patricio Millar Vernetti, Crystal Vidal, Anand Viswanathan, Amy Willerford, Natasha Williams, Crystal Wong, Marion J. Wood, Shannon W. Wuller, Shonna H. Yin, Chloe Young, Jonah Zaretsky, Susanna Zavlunova

Authors' Contributions

HLM, SNS, LCB, JD, EP, HR, MGW, and LET contributed to conceptualization. HLM, SNS, LET wrote the original draft of this manuscript. LCB, TC, YC, JD, SE-K, MH, CRO, EP, SR, HR, MGW, MH, EH, GLT, AAP, and RK handled reviewing and editing of this manuscript. SE-K contributed to project administration.

Conflicts of Interest

LCB, TC, EP, HR, and MW report funding from the Patient-Centered Outcomes Research Institute. HR, LCB, TC, EP, and MW report research funding from the NIH. LCB reports research funding from the National Institute of Diabetes and Digestive and Kidney Diseases, and Centers for Disease Control and Prevention. CO has received grants from the National Institutes of Allergy and Infectious Diseases, has a council role with the Eastern Society of Pediatric Research, and received an honorarium from the NIH. EP reports consulting fees from the NIH and a speaker honorarium from Cincinnati Children’s Hospital. SR has received consulting fees from Sequiris. All other authors declare no competing interests.

References

1. Woodrow M, Carey C, Ziauddeen N, Thomas R, Akrami A, Lutje V, et al. Systematic review of the prevalence of long COVID. Open Forum Infect Dis 2023;10(7):ofad233 [FREE Full text] [doi: 10.1093/ofid/ofad233] [Medline: 37404951]
2. Brown K, Yahyouche A, Haroon S, Camaradou J, Turner G. Long COVID and self-management. Lancet 2022;399(10322):355 [FREE Full text] [doi: 10.1016/S0140-6736(21)02798-7] [Medline: 35065779]
3. RECOVER: researching COVID to enhance recovery. Recover. URL: https://recovercovid.org [accessed 2025-01-16]
4. Thygesen JH, Tomlinson C, Hollings S, Mizani MA, Handy A, Akbari A, Longitudinal Health and Wellbeing COVID-19 National Core Study and the CVD-COVID-UK/COVID-IMPACT Consortium. COVID-19 trajectories among 57 million adults in England: a cohort study using electronic health records. Lancet Digit Health 2022;4(7):e542-e557 [FREE Full text] [doi: 10.1016/S2589-7500(22)00091-7] [Medline: 35690576]
5. Lam ICH, Wong CKH, Zhang R, Chui CSL, Lai FTT, Li X, et al. Long-term post-acute sequelae of COVID-19 infection: a retrospective, multi-database cohort study in Hong Kong and the UK. EClinicalMedicine 2023;60:102000 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102000] [Medline: 37197226]
6. Magnusson K, Turkiewicz A, Flottorp SA, Englund M. Prevalence of long COVID complaints in persons with and without COVID-19. Sci Rep 2023;13(1):6074 [FREE Full text] [doi: 10.1038/s41598-023-32636-y] [Medline: 37055494]

XSL•FO
RenderX

7. Mizrahi B, Sudry T, Flaks-Manov N, Yehezkelli Y, Kalkstein N, Akiva P, et al. Long covid outcomes at one year after mild SARS-CoV-2 infection: nationwide cohort study. BMJ 2023;380:e072529 [FREE Full text] [doi: 10.1136/bmj-2022-072529] [Medline: 36631153]
8. Al-Aly Z, Bowe B, Xie Y. Long COVID after breakthrough SARS-CoV-2 infection. Nat Med 2022;28(7):1461-1467 [FREE Full text] [doi: 10.1038/s41591-022-01840-0] [Medline: 35614233]
9. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. Nature 2021;594(7862):259-264. [doi: 10.1038/s41586-021-03553-9] [Medline: 33887749]
10. Xie Y, Choi T, Al-Aly Z. Association of treatment with nirmatrelvir and the risk of post-COVID-19 condition. JAMA Intern Med 2023;183(6):554-564 [FREE Full text] [doi: 10.1001/jamainternmed.2023.0743] [Medline: 36951829]
11. Bowe B, Xie Y, Al-Aly Z. Acute and postacute sequelae associated with SARS-CoV-2 reinfection. Nat Med 2022;28(11):2398-2405 [FREE Full text] [doi: 10.1038/s41591-022-02051-3] [Medline: 36357676]
12. Cohen K, Ren S, Heath K, Dasmariñas MC, Jubilo KG, Guo Y, et al. Risk of persistent and new clinical sequelae among adults aged 65 years and older during the post-acute phase of SARS-CoV-2 infection: retrospective cohort study. BMJ 2022;376:e068414 [FREE Full text] [doi: 10.1136/bmj-2021-068414] [Medline: 35140117]
13. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016;37:61-81 [FREE Full text] [doi: 10.1146/annurev-publhealth-032315-021353] [Medline: 26667605]
14. Pfaff ER, Madlock-Brown C, Baratta JM, Bhatia A, Davis H, Girvin A, N3C Consortium, RECOVER Consortium. Coding long COVID: characterizing a new disease through an ICD-10 lens. BMC Med 2023;21(1):58 [FREE Full text] [doi: 10.1186/s12916-023-02737-6] [Medline: 36793086]
15. OMOP common data model. OHDSI. URL: https://www.ohdsi.org/data-standardization/the-common-data-model/ [accessed 2025-01-16]
16. Common Data Model (CDM) Specification, Version 6.1. PCORnet. URL: https://pcornet.org/wp-content/uploads/2022/01/PCORnet-Common-Data-Model-v60-2020_10_221.pdf [accessed 2025-01-16]
17. Carton TW, Marsolo K, Block JP. PCORnet COVID-19 common data model design and results. NIH Health Care Systems Collaboratory. 2020. URL: https://tinyurl.com/4zkbbavm [accessed 2025-01-16]
18. Pfaff ER, Girvin AT, Gabriel DL, Kostka K, Morris M, Palchuk MB, et al. Synergies between centralized and federated approaches to data quality: a report from the national COVID cohort collaborative. J Am Med Inform Assoc 2022;29(4):609-618.
19. Pfaff ER, Girvin A, Crosskey M, Gangireddy S, Master H, Wei WQ, N3CRECOVER Consortia. De-black-boxing health AI: demonstrating reproducible machine learning computable phenotypes using the N3C-RECOVER Long COVID model in the All of Us data repository. J Am Med Inform Assoc 2023;30(7):1305-1312 [FREE Full text] [doi: 10.1093/jamia/ocad077] [Medline: 37218289]
20. Bergquist T, Loomba J, Pfaff E, Xia F, Zhao Z, Zhu Y, Long COVID Computational Challenge Participants, N3C Consortium. Crowd-sourced machine learning prediction of long COVID using data from the national COVID cohort collaborative. EBioMedicine 2024;108:105333 [FREE Full text] [doi: 10.1016/j.ebiom.2024.105333] [Medline: 39321500]
21. Brown JS, Bastarache L, Weiner MG. Aggregating electronic health record data for COVID-19 research-caveat emptor. JAMA Netw Open 2021;4(7):e2117175 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.17175] [Medline: 34255055]
22. Leese P, Anand A, Girvin A, Manna A, Patel S, Yoo YJ, et al. Clinical encounter heterogeneity and methods for resolving in networked EHR data: a study from N3C and RECOVER programs. J Am Med Inform Assoc 2023;30(6):1125-1136 [FREE Full text] [doi: 10.1093/jamia/ocad057] [Medline: 37087110]
23. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. EGEMS (Wash DC) 2013;1(3):1035 [FREE Full text] [doi: 10.13063/2327-9214.1035] [Medline: 25848578]
24. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. BMC Med Res Methodol 2021;21(1):234 [FREE Full text] [doi: 10.1186/s12874-021-01416-5] [Medline: 34706667]
25. Ledford H. How common is long COVID? Why studies give different answers. Nature 2022;606(7916):852-853. [doi: 10.1038/d41586-022-01702-2] [Medline: 35725828]
26. Bastarache L, Brown JS, Cimino JJ, Dorr DA, Embi PJ, Payne PR, et al. Developing real-world evidence from real-world data: transforming raw data into analytical datasets. Learn Health Syst 2022;6(1):e10293 [FREE Full text] [doi: 10.1002/lrh2.10293] [Medline: 35036557]
27. Guevara M, Chen S, Thomas S, Chaunzwa TL, Franco I, Kann BH, et al. Large language models to identify social determinants of health in electronic health records. NPJ Digit Med 2024;7(1):6 [FREE Full text] [doi: 10.1038/s41746-023-00970-0] [Medline: 38200151]
28. Rao S, Lee GM, Razzaghi H, Lorman V, Mejias A, Pajor NM, et al. Clinical features and burden of postacute sequelae of SARS-CoV-2 infection in children and adolescents. JAMA Pediatr 2022;176(10):1000-1009 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.2800] [Medline: 35994282]

XSL•FO
RenderX

29. Ely EW, Brown LM, Fineberg HV, National Academies of Sciences, Engineering,Medicine Committee on Examining the Working Definition for Long Covid. Long covid defined. N Engl J Med 2024;391(18):1746-1753. [doi: 10.1056/NEJMsb2408466] [Medline: 39083764]
30. Electronic health records–based phenotyping. NIH Health Care Systems Collaboratory. URL: https://rethinkingclinicaltrials.org/chapters/conduct/electronic-health-records-based-phenotyping/electronic-health-records-based-phenotyping-introduction/ [accessed 2025-01-16]
31. Mayor N, Meza-Torres B, Okusi C, Delanerolle G, Chapman M, Wang W, et al. Developing a long COVID phenotype for postacute COVID-19 in a national primary care sentinel cohort: observational retrospective database analysis. JMIR Public Health Surveill 2022;8(8):e36989 [FREE Full text] [doi: 10.2196/36989] [Medline: 35861678]
32. Zang C, Zhang Y, Xu J, Bian J, Morozyuk D, Schenck EJ, et al. Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative. Nat Commun 2023;14(1):1948 [FREE Full text] [doi: 10.1038/s41467-023-37653-z] [Medline: 37029117]
33. Fashina TA, Miller C, Paintsil E, Niccolai L, Brandt C, Oliveira C. Computable clinical phenotyping of postacute sequelae of COVID-19 in pediatrics using real-world data. J Pediatric Infect Dis Soc 2023;12(2):113-116 [FREE Full text] [doi: 10.1093/jpids/piac132] [Medline: 36548966]
34. Binka M, Klaver B, Cua G, Wong AW, Fibke C, Velásquez García HA, et al. An elastic net regression model for identifying long COVID patients using health administrative data: a population-based study. Open Forum Infect Dis 2022;9(12):ofac640 [FREE Full text] [doi: 10.1093/ofid/ofac640] [Medline: 36570972]
35. Pfaff ER, Girvin AT, Bennett TD, Bhatia A, Brooks IM, Deer RR, N3C Consortium. Identifying who has long COVID in the USA: a machine learning approach using N3C data. Lancet Digit Health 2022;4(7):e532-e541 [FREE Full text] [doi: 10.1016/S2589-7500(22)00048-6] [Medline: 35589549]
36. Lorman V, Rao S, Jhaveri R, Case A, Mejias A, Pajor NM, et al. Understanding pediatric long COVID using a tree-based scan statistic approach: an EHR-based cohort study from the RECOVER Program. JAMIA Open 2023;6(1):ooad016 [FREE Full text] [doi: 10.1093/jamiaopen/ooad016] [Medline: 36926600]
37. Lorman V, Razzaghi H, Song X, Morse K, Utidjian L, Allen AJ, et al. A machine learning-based phenotype for long COVID in children: an EHR-based study from the RECOVER program. PLoS One 2023;18(8):e0289774 [FREE Full text] [doi: 10.1371/journal.pone.0289774] [Medline: 37561683]
38. Hadley E, Yoo YJ, Patel S, Zhou A, Laraway B, Wong R, N3C and RECOVER consortia, et al. Insights from an N3C RECOVER EHR-based cohort study characterizing SARS-CoV-2 reinfections and Long COVID. Commun Med (Lond) 2024;4(1):129 [FREE Full text] [doi: 10.1038/s43856-024-00539-2] [Medline: 38992084]
39. Kenny G, McCann K, O'Brien C, Savinelli S, Tinago W, Yousif O, All-Ireland Infectious Diseases (AIID) Cohort Study Group. Identification of distinct long COVID clinical phenotypes through cluster analysis of self-reported symptoms. Open Forum Infect Dis 2022;9(4):ofac060 [FREE Full text] [doi: 10.1093/ofid/ofac060] [Medline: 35265728]
40. Reese JT, Blau H, Casiraghi E, Bergquist T, Loomba JJ, Callahan TJ, N3C Consortium, RECOVER Consortium. Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes. EBioMedicine 2023;87:104413 [FREE Full text] [doi: 10.1016/j.ebiom.2022.104413] [Medline: 36563487]
41. Zhang H, Zang C, Xu Z, Zhang Y, Xu J, Bian J, et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. Nat Med 2023;29(1):226-235 [FREE Full text] [doi: 10.1038/s41591-022-02116-3] [Medline: 36456834]
42. Peluso MJ, Deeks SG. Mechanisms of long COVID and the path toward therapeutics. Cell 2024;187(20):5500-5529 [FREE Full text] [doi: 10.1016/j.cell.2024.07.054] [Medline: 39326415]
43. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol 2023;21(3):133-146 [FREE Full text] [doi: 10.1038/s41579-022-00846-2] [Medline: 36639608]
44. Zhang HG, Honerlaw JP, Maripuri M, Samayamuthu MJ, Beaulieu-Jones BR, Baig HS, Consortium for Clinical Characterization of COVID-19 by EHR (4CE), et al. Potential pitfalls in the use of real-world data for studying long COVID. Nat Med 2023;29(5):1040-1043 [FREE Full text] [doi: 10.1038/s41591-023-02274-y] [Medline: 37055567]
45. Deer RR, Rock MA, Vasilevsky N, Carmody L, Rando H, Anzalone AJ, et al. Characterizing long COVID: deep phenotype of a complex condition. EBioMedicine 2021;74:103722 [FREE Full text] [doi: 10.1016/j.ebiom.2021.103722] [Medline: 34839263]
46. Brown CA, Londhe AA, He F, Cheng A, Ma J, Zhang J, et al. Development and validation of algorithms to identify COVID-19 patients using a US electronic health records database: a retrospective cohort study. Clin Epidemiol 2022;14:699-709 [FREE Full text] [doi: 10.2147/CLEP.S355086] [Medline: 35633659]
47. Khera R, Mortazavi BJ, Sangha V, Warner F, Patrick Young H, Ross JS, et al. A multicenter evaluation of computable phenotyping approaches for SARS-CoV-2 infection and COVID-19 hospitalizations. NPJ Digit Med 2022;5(1):27 [FREE Full text] [doi: 10.1038/s41746-022-00570-4] [Medline: 35260762]
48. Joung SY, Ebinger JE, Sun N, Liu Y, Wu M, Tang AB, et al. Awareness of SARS-CoV-2 omicron variant infection among adults with recent COVID-19 seropositivity. JAMA Netw Open 2022;5(8):e2227241 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.27241] [Medline: 35976645]
49. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. Nat Med 2021;27(4):626-631 [FREE Full text] [doi: 10.1038/s41591-021-01292-y] [Medline: 33692530]

XSL•FO
RenderX

50. Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. Infect Dis (Lond) 2021;53(10):737-754 [FREE Full text] [doi: 10.1080/23744235.2021.1924397] [Medline: 34024217]
51. Subramanian A, Nirantharakumar K, Hughes S, Myles P, Williams T, Gokhale KM, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. Nat Med 2022;28(8):1706-1714 [FREE Full text] [doi: 10.1038/s41591-022-01909-w] [Medline: 35879616]
52. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health Serv Res 2010;45(5 Pt 2):1468-1488 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01142.x] [Medline: 21054367]
53. Brannock MD, Chew RF, Preiss AJ, Hadley EC, Redfield S, McMurry JA, N3C and RECOVER consortia. Long COVID risk and pre-COVID vaccination in an EHR-based cohort study from the RECOVER program. Nat Commun 2023;14(1):2914 [FREE Full text] [doi: 10.1038/s41467-023-38388-7] [Medline: 37217471]
54. DeBord DG, Carreón T, Lentz TJ, Middendorf PJ, Hoover MD, Schulte PA. Use of the "Exposome" in the practice of epidemiology: a primer on -omic technologies. Am J Epidemiol 2016;184(4):302-314 [FREE Full text] [doi: 10.1093/aje/kwv325] [Medline: 27519539]
55. Polubriaginof FCG, Ryan P, Salmasian H, Shapiro AW, Perotte A, Safford MM, et al. Challenges with quality of race and ethnicity data in observational databases. J Am Med Inform Assoc 2019;26(8-9):730-736 [FREE Full text] [doi: 10.1093/jamia/ocz113] [Medline: 31365089]
56. Hill EL, Mehta HB, Sharma S, Mane K, Singh SK, Xie C, et al. Risk factors associated with post-acute sequelae of SARS-CoV-2: an N3C and NIH RECOVER study. BMC Public Health 2023;23(1):2103 [FREE Full text] [doi: 10.1186/s12889-023-16916-w] [Medline: 37880596]
57. Treweek S, Forouhi NG, Narayan KMV, Khunti K. COVID-19 and ethnicity: who will research results apply to? Lancet 2020;395(10242):1955-1957. [doi: 10.1016/S0140-6736(20)31380-5] [Medline: 32539937]
58. Julian McFarlane S, Occa A, Peng W, Awonuga O, Morgan SE. Community-based participatory research (CBPR) to enhance participation of racial/ethnic minorities in clinical trials: a 10-year systematic review. Health Commun 2022;37(9):1075-1092. [doi: 10.1080/10410236.2021.1943978] [Medline: 34420460]
59. Richesson RL, Marsolo K, Douthit B, Staman K, Ho PM, Dailey D, et al. Enhancing the use of EHR systems for pragmatic embedded research: lessons from the NIH health care systems research collaboratory. J Am Med Inform Assoc 2021;28(12):2626-2640 [FREE Full text] [doi: 10.1093/jamia/ocab202] [Medline: 34597383]
60. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183(8):758-764 [FREE Full text] [doi: 10.1093/aje/kwv254] [Medline: 26994063]
61. Burki T. The future of paxlovid for COVID-19. Lancet Respir Med 2022;10(7):e68 [FREE Full text] [doi: 10.1016/S2213-2600(22)00192-8] [Medline: 35623373]
62. Adler NE, Stead WW. Patients in context--EHR capture of social and behavioral determinants of health. N Engl J Med 2015;372(8):698-701. [doi: 10.1056/NEJMp1413945] [Medline: 25693009]
63. Clarke KE, Jones JM, Deng Y, Nycz E, Lee A, Iachan R, et al. Seroprevalence of infection-induced SARS-CoV-2 antibodies - United States, September 2021-February 2022. MMWR Morb Mortal Wkly Rep 2022;71(17):606-608 [FREE Full text] [doi: 10.15585/mmwr.mm7117e3] [Medline: 35482574]
64. L Mandel H, Colleen G, Abedian S, Ammar N, Charles Bailey L, Bennett TD, et al. Risk of post-acute sequelae of SARS-CoV-2 infection associated with pre-coronavirus disease obstructive sleep apnea diagnoses: an electronic health record-based analysis from the RECOVER initiative. Sleep 2023;46(9):zsad126 [FREE Full text] [doi: 10.1093/sleep/zsad126] [Medline: 37166330]
65. Goralnick E, Kaufmann C, Gawande AA. Mass-vaccination sites - an essential innovation to curb the covid-19 pandemic. N Engl J Med 2021;384(18):e67. [doi: 10.1056/NEJMp2102535] [Medline: 33691058]
66. Hill E, Mehta H, Sharma S, Mane K, Singh SK, Xie C, N3C Consortium, RECOVER Consortium. Risk factors associated with post-acute sequelae of SARS-CoV-2: an N3C and NIH RECOVER study. BMC Public Health 2023 Oct 25;23(1):2103 [FREE Full text] [doi: 10.1186/s12889-023-16916-w] [Medline: 37880596]
67. Forrest CB, Burrows E, Mejias A, Razzaghi H, Christakis D, Jhaveri R, et al. Pediatrics 2022;149(4):e2021055765. [doi: 10.1542/peds.2021-055765] [Medline: 35322270]
68. Tannous J, Pan AP, Potter T, Bako AT, Dlouhy K, Drews A, et al. Real-world effectiveness of COVID-19 vaccines and anti-SARS-CoV-2 monoclonal antibodies against postacute sequelae of SARS-CoV-2: analysis of a COVID-19 observational registry for a diverse US metropolitan population. BMJ Open 2023;13(4):e067611 [FREE Full text] [doi: 10.1136/bmjopen-2022-067611] [Medline: 37019490]
69. RECOVER: researching COVID to enhance recovery. RECOVER. URL: https://recovercovid.org/ [accessed 2015-01-26]

Abbreviations

CDC: Centers for Disease Control and Prevention
CSC: clinical science core
CRN: clinical research network

XSL•FO
RenderX

EHR: electronic health record
GIS: Geographic Information Systems
ICD-10: International Statistical Classification of Diseases, Tenth Revision
ME/CFS: myalgic encephalomyelitis/chronic fatigue syndrome
ML: machine learning
N3C: National COVID Cohort Collaborative
NIH: National Institutes of Health
OMOP: Observational Medical Outcomes Partnership
OSA: obstructive sleep apnea
PCORI: Patient-Centered Outcomes Research Institute
PCORNet: National Patient-Centered Clinical Research Network
POTS: postural orthostatic tachycardia syndrome
RECOVER: REsearching COVID to Enhance Recovery
VA: Department of Veterans Affairs

Edited by A Mavragani; submitted 02.05.24; peer-reviewed by Y Zhang, S Zeng; comments to author 12.09.24; revised version received 31.10.24; accepted 20.11.24; published 05.03.25.

Please cite as:
Mandel HL, Shah SN, Bailey LC, Carton T, Chen Y, Esquenazi-Karonika S, Haendel M, Hornig M, Kaushal R, Oliveira CR, Perlowski AA, Pfaff E, Rao S, Razzaghi H, Seibert E, Thomas GL, Weiner MG, Thorpe LE, Divers J, RECOVER EHR Cohort
Opportunities and Challenges in Using Electronic Health Record Systems to Study Postacute Sequelae of SARS-CoV-2 Infection: Insights From the NIH RECOVER Initiative
J Med Internet Res 2025;27:e59217
URL: https://www.jmir.org/2025/1/e59217
doi:10.2196/59217
PMID:40053748

©Hannah L Mandel, Shruti N Shah, L Charles Bailey, Thomas Carton, Yu Chen, Shari Esquenazi-Karonika, Melissa Haendel, Mady Hornig, Rainu Kaushal, Carlos R Oliveira, Alice A Perlowski, Emily Pfaff, Suchitra Rao, Hanieh Razzaghi, Elle Seibert, Gelise L Thomas, Mark G Weiner, Lorna E Thorpe, Jasmin Divers, RECOVER EHR Cohort. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Emerging Domains for Measuring Health Care Delivery With Electronic Health Record Metadata

Daniel Tawfik[1], MD, MS; Adam Rule[2], MS, PhD; Aram Alexanian[3], MD; Dori Cross[4], PhD; A Jay Holmgren[5], MS, PhD; Sunny S Lou[6], MD, PhD; Eugenia McPeek Hinz[7], MD, MS; Christian Rose[8], MD; Ratnalekha V N Viswanadham[9], PhD; Rebecca G Mishuris[10], MD, MPH, MS; Jorge M Rodríguez-Fernández[11], MD; Eric W Ford[12], MPH, PhD; Sarah T Florig[13], MS; Christine A Sinsky[14], MD; Nate C Apathy[15], PhD

[1]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, United States
[2]Information School, University of Wisconsin-Madison, Madison, WI, United States
[3]Department of Family Medicine, Novant Health, Winston-Salem, NC, United States
[4]Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN, United States
[5]School of Medicine, University of California San Francisco, San Francisco, CA, United States
[6]Department of Anesthesiology, Washington University School of Medicine, St Louis, MO, United States
[7]Department of General Internal Medicine, Duke University School of Medicine, Durham, NC, United States
[8]Department of Emergency Medicine, Stanford University School of Medicine, Palo Alto, CA, United States
[9]Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States
[10]Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States
[11]Department of Neurology, School of Medicine, University of Texas Medical Branch, Galveston, TX, United States
[12]Department of Health Policy and Organization, School of Public Health, University of Alabama, Birmingham, AL, United States
[13]Department of Pulmonary, Allergy, and Critical Care Medicine, Oregon Health & Science University, Portland, OR, United States
[14]American Medical Association, Chicago, IL, United States
[15]Department of Health Policy & Management, University of Maryland School of Public Health, College Park, MD, United States

Corresponding Author:
Daniel Tawfik, MD, MS
Department of Pediatrics
Stanford University School of Medicine
770 Welch Road, Suite 435
Palo Alto, CA, 94304
United States
Phone: 1 6507239902
Email: dtawfik@stanford.edu

Abstract

This article aims to introduce emerging measurement domains made feasible through the electronic health record (EHR) use metadata, to inform the changing landscape of health care delivery. We reviewed emerging domains in which EHR metadata may be used to measure health care delivery, outlining a framework for evaluating measures based on desirability, feasibility, and viability. We argue that EHR use metadata may be leveraged to develop and operationalize novel measures in the domains of team structure and dynamics, workflows, and cognitive environment to provide a clearer understanding of modern health care delivery. Examples of measures feasible using metadata include quantification of teamwork and collaboration, patient continuity measures, workflow conformity measures, and attention switching. By enabling measures that can be used to inform the next generation of health care delivery, EHR metadata may be used to improve the quality of patient care and support clinician well-being. Careful attention is needed to ensure that these measures are desirable, feasible, and viable.

(J Med Internet Res 2025;27:e64721) doi:10.2196/64721

KEYWORDS

metadata; health services research; audit logs; event logs; electronic health record data; health care delivery; patient care; healthcare teams; clinician-patient relationship; cognitive environment

XSL•FO
RenderX

Introduction

Since the widespread adoption of electronic health records (EHRs) following the enactment of the Health Information Technology for Economic and Clinical Health (HITECH) Act, clinical care is increasingly mediated by health information technology (HIT) [1]. The digitization of health care has enabled not only the capture of more clinical data for use in clinical research [2], but also metadata about how those data are produced and used, giving a window into the process of health care delivery [3-7].

Established health care metrics focus largely on the patient process and outcome measurements (eg, health maintenance screening and preventable admissions) or outcomes for individual health care workers (eg, turnover and professional burnout). While these measures reflect important outcomes of a complex health care system, they do not reliably indicate upstream aspects and contextual factors of health care delivery, which may be leveraged to improve patient and clinician outcomes. EHR use metadata (eg, audit logs, orders metadata, documentation and communication metadata, and patient encounters metadata) contain valuable details on the complex system in which health care is delivered, with early insights primarily focused on discrete action or time-based measures [8-15].

This paper describes 3 examples of emerging domains of measurement through EHR use metadata, identified through a series of workshops facilitated by the Measures Workgroup of the National Research Network for EHR Audit Log Data [16]. These domains, health care teams, workflows, and cognitive environments, may provide a clearer understanding of modern health care delivery. We expect the value of these measures to outweigh the resources required to develop them, informing the next generation of health care delivery that improves both the quality of patient care and supports clinician well-being.

Data and Data Sources

Traditional sources of health care delivery data (eg, time and motion studies, surveys, or administrative claims) [17,18] often require high resource expenditure to capture accurately and in detail. In contrast, continuously collected EHR use metadata can facilitate widespread measurement of several domains of health care delivery [4,7,13], providing valuable data that otherwise would require direct observations that are virtually impossible to collect at scale or for extended durations [19].

At a more granular level, EHR audit logs and other event logs offer a detailed account of actions that took place within the EHR at what time and by whom [20], enabling exploration of the "path" that a clinician or team took in the EHR to complete that task. In turn, these details can be used to delineate certain workflows or team structures contributing to task efficiency and clinical outcomes, which may be the result of vendor-derived or investigator-derived measures [21]. Such approaches have been described for a variety of EHR vendors, clinical scenarios [12], methodologies [22], and health professions [23].

Although metadata are not identical across EHR vendors, and institutions often have differing customizations and third-party modules, commonalities among metadata structures make it feasible to normalize measures across institutions with sufficient considerations. In addition, recent studies from 2 of the largest EHR vendors suggest that clinician variation far exceeds organizational variation [24,25], indicating that metadata-derived EHR use measures are more reflective of individual workflows than organizational configurations. We posit here that these new horizons of measurement warrant dedicated development and prioritization of the key clinical and operational questions for which EHR use metadata can be brought to bear.

Principles of Domain Measurement

Innovative ideas need to be feasible, viable, and desirable to be successful [26]. While originally developed for commercial products, this framework can also guide EHR measure development. First, EHR use metadata to enable new measures of health care delivery that are technically feasible but also require proof that such measures are valid and reliable. Second, analyzing EHR metadata takes less time and money than many other methods of observing health care delivery, but measurement schemes based on EHR metadata still need to prove their viability (providing more value than the effort required to produce them) by generating insights that tangibly improve care [27]. Finally, measures need to be desirable, aligning with human values and providing insight into phenomena people care about. Desirable measures may answer research questions, contribute to more efficient and reliable single-center cross-sectional or longitudinal operational and quality reporting [28], or facilitate multi-institutional comparisons. This perspective demonstrates that, in health care delivery, these phenomena include not only how people spend their time but also how they interact (team structure and dynamics), organize their work (workflows), and direct their attention (cognitive environment; Figure 1) [9,21,29-33].

XSL•FO
RenderX

Figure 1. Emerging domains of measurement: team structure and dynamics, workflows, and cognitive environment, with example measures for each.

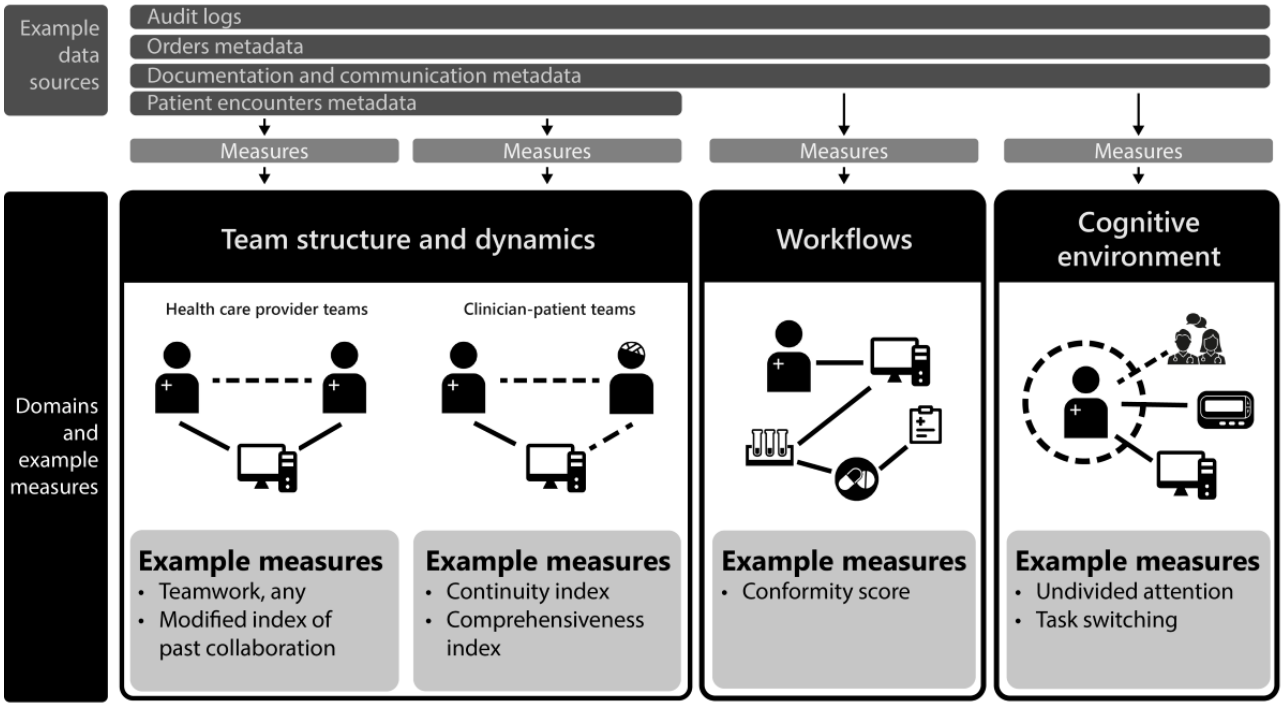

Making decisions based on social measures is fraught with risk, with a natural tendency to focus on what is easiest to measure (feasible) rather than what is most useful (viable) or meaningful (desirable) as an indicator of high-quality processes or outcomes [34]. Yet, most measures are only rough proxies for more complex phenomena, and improving the measure (eg, patient satisfaction scores) may not improve the phenomenon of interest (eg, patient satisfaction) [35]. Furthermore, poor selection of a proxy can promote bias (eg, using predicted health care costs as a proxy for clinical risk underestimates the needs of those with less access to care [36]), and measures used for social decision-making are more likely to be gamed [37].

These risks do not require that we avoid measuring health care delivery. As statistician George Box observed, "Essentially, all models are wrong, but some are useful" [38]. Developing useful measures while acknowledging the ways in which they are wrong requires careful attention to the social aspects of what cannot be easily measured and the unintended consequences of putting such measures into practice [39]. Here we describe 3 domains of health care delivery where EHR use metadata measures may be feasible, viable, and desirable, though not without risk.

Measuring Team Structure and Dynamics

Most patient care is delivered through teams of nurses, medical assistants, technicians, pharmacists, physicians, and many other staff members all of whom are integral to care delivery and outcomes. Fortunately, a byproduct of EHR auditing requirements means that EHR use metadata record details on both the clinicians and patients for whom they perform actions, making it feasible to reconstruct team structures and dynamics. This can be viewed as (1) clinician-centric models of health care provider teams, and (2) patient-centric models of clinicians involved in their care (synchronously and not), recognizing that each feeds back on the other through complex relationships. Identifying these team structures relies on inference based on shared patient access, physical locations of EHR access, temporal proximity, or direct communication records [21,40].

Health Care Provider Teams

EHR uses metadata in relation to defined team structures to allow investigation into patterns of how health care teams interact with each other. The "strength" of team structure can then be weighted (eg, through repeat interactions) to quantify team familiarity that might promote shared mental models that are essential for effective teamwork.

Researchers have the opportunity to consider the timing, intensity, and content of team-based interactions as proxies for effective teamwork. This requires a robust understanding of how EHR-based teamwork takes place in a broader framework of in-person and HIT-mediated collaboration, allowing for variance in the phenotypes of team collaboration that can reliably still be associated with provider satisfaction and patient care quality [41]. This work also requires the development of outcome measures more sensitive to the quality of team collaboration, such as care delays or therapy cycling. Valid and reliable understanding of the connections between team structure, processes, and outcomes is foundational to each measure's viability and desirability, insofar as it supports health care managers in making operational decisions and designing institutional supports that enhance team collaboration and quality of care [42].

Clinician-Patient Teams

The clinician-patient relationship is the cornerstone of health care provision, yet many essential measures of this relationship have been prohibitively difficult to measure at scale before the

XSL•FO
RenderX

emergence of EHR use metadata. For example, continuity of care is associated with patient satisfaction and a host of improved outcomes [43,44], attributed to the trust and knowledge in the clinician-patient relationship that supports tailored evaluations, accurate diagnosis, customized treatment plans, and increased medication adherence. EHR use of metadata has made more nuanced measurements in this domain feasible because system-wide records of all clinician-patient interactions now enable quantifying repeated interactions with the same clinician at a more granular level than visits, procedures, or diagnoses. These records include asynchronous messages through patient portals and documentation of telephone encounters. With the addition of domain knowledge, these detailed metadata also enable assessments of the appropriateness of a patient's encounters with primary care versus subspecialty care. These concepts of continuity of care [45-53] and comprehensiveness [31,54,55] are particularly viable and desirable as they may help physicians and operational leaders optimize care to achieve the quintuple aim [27,56] goals of quality, cost, satisfaction, patient and care team experience, and equity.

Through improved measures and stakeholder engagement, health care systems can work to optimize the relationship between clinicians and patients, to promote health care that is not only more effective and efficient but also more satisfying and rewarding for both clinicians and patients.

Health care team structure and dynamics are complex and not fully reflected within EHR use metadata, raising important limitations. The balance between HIT-facilitated and in-person interactions varies among settings, with important implications for the validity of extrapolating teamwork measures from EHR use metadata within a particular use case. In addition, value judgments on the appropriateness or quality of a particular team structure or clinician-patient interaction require additional clinical and system knowledge, much of which is not readily available within EHR data or metadata.

Measuring Workflows

EHR audit logs and other event logs contain records of timestamped granular interactions with an EHR as part of larger clinical tasks (eg, refilling a prescription or ordering bloodwork), making feasible large-scale identification of common pathways for routine clinical practices [32]. These pathways may then be evaluated for relative time spent on clinical tasks, the complexity of action sequences, engagement with decision support, variation for particular workflows, specialties, providers, or combinations of these (eg, variation in outpatient ordering workflow among physicians within 1 clinic) [57].

To be viable and desirable, workflow measures must account for the tension between unnecessary and necessary deviations of behaviors from the norm [58]. Reducing unnecessary deviations can improve efficiency and patient safety through standardized best practices [59,60]. For example, heterogeneity in physician work with medical scribes may negatively impact physician efficiency [61]. While highly variable stroke care pathways may result in therapeutic delays [32]. Yet, enabling necessary deviations from the norm may promote tailored approaches to unique or complex scenarios, promoting clinician autonomy, professional satisfaction, and precision medicine.

Key to this measurement domain is the risk of misclassifying an unhelpful workflow as "normative" or a necessary variation as "deviant," highlighting the importance of establishing clinical expertise, stakeholder buy-in, and relationships to important outcomes in the development and deployment of any workflow measures. In addition, although efforts at automated task identification within audit log data have been made [22,62,63], the diversity of clinical workflows makes reliable task identification challenging. Furthermore, due to differences in actions and tracking methods available in particular EHR instances, such measures may be poorly generalizable outside of an institution or EHR vendor unless carefully designed to be context-agnostic and may require recalibration over time with new technologies (eg, large language models and generative artificial intelligence), clinical approaches, or software functionality.

Measuring Cognitive Environment

Clinical work is cognitively complex; clinicians must synthesize ever increasing amounts of information to make diagnosis and treatment decisions amidst competing demands for their attention. Cognitive overload, which has been linked to errors and burnout, can occur easily in such environments [64,65]. Because clinical work is largely mediated by the EHR [66], EHR use metadata offers opportunities to measure clinician cognitive load at scale without laborious data collection. These measurements are critical for understanding how the clinical cognitive environment can be improved.

Cognitive load includes an intrinsic component related to the complexity of a particular task (eg, patient complexity), and an extrinsic component related to the work environment (eg, interruptions, poor EHR, or workplace design [67]). Both components may be feasible to measure using EHR use metadata. For example, intrinsic cognitive load can be approximated using audit log–derived measures of patient load [11,68], EHR time [69], and encounter or task complexity [70]. Extrinsic cognitive load is more challenging to measure; while some interruptions like secure messaging use can be observed directly within domain-specific or third-party event logs [71], many occur outside of HIT. Instead, efforts have focused on capturing task fragmentation [72,73], such as attention switching [74] or undivided attention [9,33,65,67,74-83].

While this domain has considerable promises, there are several barriers to progress. Few of the audit log-derived cognitive load measures have been validated using established instruments from the cognitive science literature (eg, surveys like the NASA-TLX [National Aeronautics and Space Administration–Task Load Index] or physiological measurements like pupillometry) [84]. In addition, difficulties with the related concept of task identification limit the ability to subsequently measure task fragmentation, with further validation of task identification needed before task fragmentation can be reliably measured.

XSL•FO
RenderX

Conclusion

EHR uses metadata to provide remarkable potential for measuring and improving health care delivery by making feasible previously inaccessible insights. By measuring team structure and dynamics using intersecting metadata on clinician EHR use, health care provider teams may be studied and optimized and clinician-patient teams can be improved to prioritize continuity of care and comprehensiveness where appropriate. Health care workflows may be better understood through better identification of tasks and measurement of paths taken to achieve them. Finally, proxies of cognitive load may be measured, to inform interventions to reduce overload and the resultant risks for burnout and medical errors. Each of these measurement domains will require further tuning to fully realize their promise of improving health care delivery through sustainably scalable measurement.

Acknowledgments

This work was funded in part by a grant (K08HS027837) from the Agency for Health care Research and Quality and an EHR (electronic health record) Use Metrics Research Grant from the American Medical Association. The views expressed in this paper do not reflect those of the funding agencies. They had no role in writing or reviewing the work.

Conflicts of Interest

Adam Rule reports receiving grants, honoraria, and travel support from the American Medical Association outside the reported work.

References

1. Washington V, DeSalvo K, Mostashari F, Blumenthal D. The HITECH era and the path forward. N Engl J Med 2017;377(10):904-906. [doi: 10.1056/NEJMp1703370] [Medline: 28877013]
2. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc 2013;20(1):117-121 [FREE Full text] [doi: 10.1136/amiajnl-2012-001145] [Medline: 22955496]
3. Tawfik D, Bayati M, Liu J, Nguyen L, Sinha A, Kannampallil T, et al. Predicting primary care physician burnout from electronic health record use measures. Mayo Clin Proc 2024;99(9):1411-1421 [FREE Full text] [doi: 10.1016/j.mayocp.2024.01.005] [Medline: 38573301]
4. Adler-Milstein J, Adelman JS, Tai-Seale M, Patel VL, Dymek C. EHR audit logs: a new goldmine for health services research? J Biomed Inform 2020;101:103343 [FREE Full text] [doi: 10.1016/j.jbi.2019.103343] [Medline: 31821887]
5. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. J Am Med Inform Assoc 2020;27(4):531-538 [FREE Full text] [doi: 10.1093/jamia/ocz220] [Medline: 32016375]
6. Kannampallil T, Adler-Milstein J. Using electronic health record audit log data for research: insights from early efforts. J Am Med Inform Assoc 2022;30(1):167-171 [FREE Full text] [doi: 10.1093/jamia/ocac173] [Medline: 36173351]
7. Sinsky C, Rule A, Cohen G, Arndt B, Shanafelt T, Sharp C, et al. Metrics for assessing physician activity using electronic health record log data. J Am Med Inform Assoc 2020;27(4):639-643 [FREE Full text] [doi: 10.1093/jamia/ocz223] [Medline: 32027360]
8. Eschenroeder HC, Manzione LC, Adler-Milstein J, Bice C, Cash R, Duda C, et al. Associations of physician burnout with organizational electronic health record support and after-hours charting. J Am Med Inform Assoc 2021;28(5):960-966 [FREE Full text] [doi: 10.1093/jamia/ocab053] [Medline: 33880534]
9. Bartek B, Lou SS, Kannampallil T. Measuring the cognitive effort associated with task switching in routine EHR-based tasks. J Biomed Inform 2023;141:104349. [doi: 10.1016/j.jbi.2023.104349] [Medline: 37015304]
10. Dyrbye L, Gordon J, O'Horo J, Belford S, Wright M, Satele D, et al. Relationships between EHR-based audit log data and physician burnout and clinical practice process measures. Mayo Clin Proc 2023;98(3):398-409. [doi: 10.1016/j.mayocp.2022.10.027] [Medline: 36868747]
11. Lou SS, Lew D, Harford DR, Lu C, Evanoff BA, Duncan JG, et al. Temporal associations between EHR-derived workload, burnout, and errors: a prospective cohort study. J Gen Intern Med 2022;37(9):2165-2172 [FREE Full text] [doi: 10.1007/s11606-022-07620-3] [Medline: 35710654]
12. Rule A, Chiang MF, Hribar MR. Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods. J Am Med Inform Assoc 2020;27(3):480-490 [FREE Full text] [doi: 10.1093/jamia/ocz196] [Medline: 31750912]
13. Rule A, Melnick ER, Apathy NC. Using event logs to observe interactions with electronic health records: an updated scoping review shows increasing use of vendor-derived measures. J Am Med Inform Assoc 2022;30(1):144-154 [FREE Full text] [doi: 10.1093/jamia/ocac177] [Medline: 36173361]

XSL•FO
RenderX

14. Magon HS, Helkey D, Shanafelt T, Tawfik D. Creating conversion factors from EHR event log data: a comparison of investigator-derived and vendor-derived metrics for primary care physicians. AMIA Annu Symp Proc 2023;2023:1115-1124 [FREE Full text] [Medline: 38222350]
15. Sinha A, Stevens LA, Su F, Pageler NM, Tawfik DS. Measuring electronic health record use in the pediatric ICU using audit-logs and screen recordings. Appl Clin Inform 2021;12(4):737-744 [FREE Full text] [doi: 10.1055/s-0041-1733851] [Medline: 34380167]
16. Center CIaIR. National research network for EHR audit log data. University of California San Francisco. URL: https://cliir.ucsf.edu/portfolio/national-research-network-ehr-audit-log-data [accessed 2024-07-17]
17. Steinwachs DM, Hughes RG. Health services research: scope and significance. In: Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville: Agency for Healthcare Research and Quality (US); 2008.
18. Horwitz LI, Partovian C, Lin Z, Grady JN, Herrin J, Conover M, et al. Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission. Ann Intern Med 2014;161(10 Suppl):S66-S75 [FREE Full text] [doi: 10.7326/M13-3000] [Medline: 25402406]
19. Baxter SL, Apathy NC, Cross DA, Sinsky C, Hribar MR. Measures of electronic health record use in outpatient settings across vendors. J Am Med Inform Assoc 2021;28(5):955-959 [FREE Full text] [doi: 10.1093/jamia/ocaa266] [Medline: 33211862]
20. Rule A, Kannampallil T, Hribar MR, Dziorny AC, Thombley R, Apathy NC, et al. Guidance for reporting analyses of metadata on electronic health record use. J Am Med Inform Assoc 2024;31(3):784-789. [doi: 10.1093/jamia/ocad254] [Medline: 38123497]
21. Rose C, Thombley R, Noshad M, Lu Y, Clancy HA, Schlessinger D, et al. Team is brain: leveraging EHR audit log data for new insights into acute care processes. J Am Med Inform Assoc 2022;30(1):8-15 [FREE Full text] [doi: 10.1093/jamia/ocac201] [Medline: 36303451]
22. Lou SS, Liu H, Harford D, Lu C, Kannampallil T. Characterizing the macrostructure of electronic health record work using raw audit logs: an unsupervised action embeddings approach. J Am Med Inform Assoc 2023;30(3):539-544 [FREE Full text] [doi: 10.1093/jamia/ocac239] [Medline: 36478460]
23. Tiase VL, Sward KA, Facelli JC. A scalable and extensible logical data model of electronic health record audit logs for temporal data mining (RNteract): model conceptualization and formulation. JMIR Nurs 2024;7:e55793 [FREE Full text] [doi: 10.2196/55793] [Medline: 38913994]
24. Cross DA, Holmgren AJ, Apathy NC. The role of organizations in shaping physician use of electronic health records. Health Serv Res 2024;59(1):e14203 [FREE Full text] [doi: 10.1111/1475-6773.14203] [Medline: 37438938]
25. Overhage JM, Qeadan F, Choi EHE, Vos D, Kroth PJ. Explaining variability in electronic health record effort in primary care ambulatory encounters. Appl Clin Inform 2024;15(2):212-219 [FREE Full text] [doi: 10.1055/s-0044-1782228] [Medline: 38508654]
26. Brown T. Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. New York: HarperCollins; 2009.
27. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
28. Saraswathula A, Merck SJ, Bai G, Weston CM, Skinner EA, Taylor A, et al. The volume and cost of quality metric reporting. JAMA 2023;329(21):1840-1847 [FREE Full text] [doi: 10.1001/jama.2023.7271] [Medline: 37278813]
29. Borgatti SP, Jones C. A measure of past collaboration. Connections 1996;19(1):58-60 [FREE Full text]
30. Rodriguez HP, Marshall RE, Rogers WH, Safran DG. Primary care physician visit continuity: a comparison of patient-reported and administratively derived measures. J Gen Intern Med 2008;23(9):1499-1502 [FREE Full text] [doi: 10.1007/s11606-008-0692-z] [Medline: 18563492]
31. O'Malley AS, Rich EC, Shang L, Rose T, Ghosh A, Poznyak D, et al. Practice-site-level measures of primary care comprehensiveness and their associations with patient outcomes. Health Serv Res 2021;56(3):371-377 [FREE Full text] [doi: 10.1111/1475-6773.13599] [Medline: 33197047]
32. Noshad M, Rose CC, Chen JH. Signal from the noise: a mixed graphical and quantitative process mining approach to evaluate care pathways applied to emergency stroke care. J Biomed Inform 2022;127:104004 [FREE Full text] [doi: 10.1016/j.jbi.2022.104004] [Medline: 35085813]
33. Chen Y, Adler-Milstein J, Sinsky CA. Measuring and maximizing undivided attention in the context of electronic health records. Appl Clin Inform 2022;13(4):774-777 [FREE Full text] [doi: 10.1055/a-1892-1437] [Medline: 35790200]
34. Muller JZ. The Tyranny of Metrics. New Jersey: Princeton University Press; 2018.
35. Harris M, Tayler B. Don't let metrics undermine your business: an obsession with the numbers can sink your strategy. Harvard Business Review 2019;97(5):62-70 [FREE Full text]
36. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453 [FREE Full text] [doi: 10.1126/science.aax2342] [Medline: 31649194]
37. Campbell DT. Assessing the impact of planned social change. Evaluation and Program Planning 1979;2(1):67-90. [doi: 10.1016/0149-7189(79)90048-x]
38. Box GEP, Draper NR. Empirical Model-Building and Response Surfaces. Hoboken, New Jersey: Wiley; 1987.

XSL•FO
RenderX

39. Burden M, Keniston A, Pell J, Yu A, Dyrbye L, Kannampallil T. Unlocking inpatient workload insights with electronic health record event logs. J Hosp Med 2025;20(1):79-84. [doi: 10.1002/jhm.13386] [Medline: 38704753]
40. Mundt MP, Gilchrist VJ, Fleming MF, Zakletskaia LI, Tuan WJ, Beasley JW. Effects of primary care team social networks on quality of care and costs for patients with cardiovascular disease. Ann Fam Med 2015;13(2):139-148 [FREE Full text] [doi: 10.1370/afm.1754] [Medline: 25755035]
41. Kim S, Song H, Valentine MA. Learning in temporary teams: the varying effects of partner exposure by team member role. Organization Science 2023;34(1):433-455. [doi: 10.1287/orsc.2022.1585]
42. Huckman RS, Staats BR. Fluid tasks and fluid teams: the impact of diversity in experience and team familiarity on team performance. M&SOM 2011;13(3):310-328. [doi: 10.1287/msom.1100.0321]
43. Chan KS, Wan EYF, Chin WY, Cheng WHG, Ho MK, Yu EYT, et al. Effects of continuity of care on health outcomes among patients with diabetes mellitus and/or hypertension: a systematic review. BMC Fam Pract 2021;22(1):145 [FREE Full text] [doi: 10.1186/s12875-021-01493-x] [Medline: 34217212]
44. Reddy A, Pollack CE, Asch DA, Canamucio A, Werner RM. The effect of primary care provider turnover on patient experience of care and ambulatory quality of care. JAMA Intern Med 2015;175(7):1157-1162 [FREE Full text] [doi: 10.1001/jamainternmed.2015.1853] [Medline: 25985320]
45. Wolinsky FD, Bentler SE, Liu L, Geweke JF, Cook EA, Obrizan M, et al. Continuity of care with a primary care physician and mortality in older adults. J Gerontol A Biol Sci Med Sci 2010;65(4):421-428 [FREE Full text] [doi: 10.1093/gerona/glp188] [Medline: 19995831]
46. Waters A. Patients treated at practices with high GP turnover have worse outcomes, finds study. BMJ 2023;380:177. [doi: 10.1136/bmj.p177] [Medline: 36693672]
47. Parisi R, Lau YS, Bower P, Checkland K, Rubery J, Sutton M, et al. Predictors and population health outcomes of persistent high GP turnover in English general practices: a retrospective observational study. BMJ Qual Saf 2023;32(7):394-403 [FREE Full text] [doi: 10.1136/bmjqs-2022-015353] [Medline: 36690473]
48. Sinsky CA, Shanafelt TD, Dyrbye LN, Sabety AH, Carlasare LE, West CP. Health care expenditures attributable to primary care physician overall and burnout-related turnover: a cross-sectional analysis. Mayo Clin Proc 2022;97(4):693-702 [FREE Full text] [doi: 10.1016/j.mayocp.2021.09.013] [Medline: 35227508]
49. Nyweide DJ, Bynum JP. Relationship between continuity of ambulatory care and risk of emergency department episodes among older adults. Ann Emerg Med 2017;69(4):407-415.e3 [FREE Full text] [doi: 10.1016/j.annemergmed.2016.06.027] [Medline: 27520592]
50. Frandsen BR, Joynt KE, Rebitzer JB, Jha AK. Care fragmentation, quality, and costs among chronically ill patients. Am J Manag Care 2015;21(5):355-362 [FREE Full text] [Medline: 26167702]
51. Chen CC, Tseng CH, Cheng SH. Continuity of care, medication adherence, and health care outcomes among patients with newly diagnosed type 2 diabetes: a longitudinal analysis. Med Care 2013;51(3):231-237. [doi: 10.1097/MLR.0b013e31827da5b9] [Medline: 23269110]
52. Yao S, Lix L, Teare G, Evans C, Blackburn D. An integrated continuity of care measure improves performance in models predicting medication adherence using population-based administrative data. PLoS One 2022;17(3):e0264170 [FREE Full text] [doi: 10.1371/journal.pone.0264170] [Medline: 35239713]
53. Pereira Gray DJ, Sidaway-Lee K, White E, Thorne A, Evans PH. Continuity of care with doctors-a matter of life and death? A systematic review of continuity of care and mortality. BMJ Open 2018;8(6):e021161 [FREE Full text] [doi: 10.1136/bmjopen-2017-021161] [Medline: 29959146]
54. Rich EC, Peris K, Luhr M, Ghosh A, Molinari L, O'Malley AS. Association of the range of outpatient services provided by primary care physicians with subsequent health care costs and utilization. J Gen Intern Med 2023;38(15):3414-3423. [doi: 10.1007/s11606-023-08363-5] [Medline: 37580638]
55. Bazemore A, Petterson S, Peterson LE, Phillips RL. More comprehensive care among family physicians is associated with lower costs and fewer hospitalizations. Ann Fam Med 2015;13(3):206-213 [FREE Full text] [doi: 10.1370/afm.1787] [Medline: 25964397]
56. Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. JAMA 2022;327(6):521-522. [doi: 10.1001/jama.2021.25181] [Medline: 35061006]
57. Mandal S, Wiesenfeld BM, Mann DM, Szerencsy AC, Iturrate E, Nov O. Quantifying the impact of telemedicine and patient medical advice request messages on physicians' work-outside-work. NPJ Digit Med 2024;7(1):35 [FREE Full text] [doi: 10.1038/s41746-024-01001-2] [Medline: 38355913]
58. Sinsky CA, Bavafa H, Roberts RG, Beasley JW. Standardization vs customization: finding the right balance. Ann Fam Med 2021;19(2):171-177 [FREE Full text] [doi: 10.1370/afm.2654] [Medline: 33685879]
59. Cohen GR, Friedman CP, Ryan AM, Richardson CR, Adler-Milstein J. Variation in physicians' electronic health record documentation and potential patient harm from that variation. J Gen Intern Med 2019;34(11):2355-2367 [FREE Full text] [doi: 10.1007/s11606-019-05025-3] [Medline: 31183688]
60. Rodríguez-Fernández JM, Loeb JA, Hier DB. It's time to change our documentation philosophy: writing better neurology notes without the burnout. Front Digit Health 2022;4:1063141 [FREE Full text] [doi: 10.3389/fdgth.2022.1063141] [Medline: 36518562]

XSL•FO
RenderX

61. Rule A, Chiang MF, Hribar MR. Medical scribes have a variable impact on documentation workflows. Stud Health Technol Inform 2022;290:892-896 [FREE Full text] [doi: 10.3233/SHTI220208] [Medline: 35673147]
62. Perros I, Yan X, Jones JB, Sun J, Stewart WF. Using the PARAFAC2 tensor factorization on EHR audit data to understand PCP desktop work. J Biomed Inform 2020;101:103312 [FREE Full text] [doi: 10.1016/j.jbi.2019.103312] [Medline: 31627022]
63. Chen B, Alrifai W, Gao C, Jones B, Novak L, Lorenzi N, et al. Mining tasks and task characteristics from electronic health record audit logs with unsupervised machine learning. J Am Med Inform Assoc 2021;28(6):1168-1177 [FREE Full text] [doi: 10.1093/jamia/ocaa338] [Medline: 33576432]
64. Graber ML, Kissam S, Payne VL, Meyer AND, Sorensen A, Lenfestey N, et al. Cognitive interventions to reduce diagnostic error: a narrative review. BMJ Qual Saf 2012;21(7):535-557. [doi: 10.1136/bmjqs-2011-000149] [Medline: 22543420]
65. Harry E, Sinsky C, Dyrbye L, Makowski M, Trockel M, Tutty M, et al. Physician task load and the risk of burnout among US physicians in a national survey. Jt Comm J Qual Patient Saf 2021;47(2):76-85 [FREE Full text] [doi: 10.1016/j.jcjq.2020.09.011] [Medline: 33168367]
66. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. Health Aff (Millwood) 2017;36(4):655-662 [FREE Full text] [doi: 10.1377/hlthaff.2016.0811] [Medline: 28373331]
67. Harry EP, Kneeland P, Huang G, Stein J, Sweller J. Cognitive load and its implciations for health care. NEJM Catalyst. 2018. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0233 [accessed 2025-02-19]
68. Mai MV, Orenstein EW, Manning JD, Luberti AA, Dziorny AC. Attributing patients to pediatric residents using electronic health record features augmented with audit logs. Appl Clin Inform 2020;11(3):442-451 [FREE Full text] [doi: 10.1055/s-0040-1713133] [Medline: 32583389]
69. Rotenstein LS, Holmgren AJ, Downing NL, Bates DW. Differences in total and after-hours electronic health record time across ambulatory specialties. JAMA Intern Med 2021;181(6):863-865 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0256] [Medline: 33749732]
70. Lou SS, Baratta LR, Lew D, Harford D, Avidan MS, Kannampallil T. Anesthesia clinical workload estimated from electronic health record documentation vs billed relative value units. JAMA Netw Open 2023;6(8):e2328514 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.28514] [Medline: 37566415]
71. Baratta LR, Harford D, Sinsky CA, Kannampallil T, Lou SS. Characterizing the patterns of electronic health record-integrated secure messaging use: cross-sectional study. J Med Internet Res 2023;25:e48583 [FREE Full text] [doi: 10.2196/48583] [Medline: 37801359]
72. Moy AJ, Cato KD, Kim EY, Withall J, Rossetti SC. A Computational framework to evaluate emergency department clinician task switching in the electronic health record using event logs. AMIA Annu Symp Proc 2023;2023:1183-1192 [FREE Full text] [Medline: 38222361]
73. Shah M, De Arrigunaga S, Forman LS, West M, Rowe SG, Mishuris RG. Cumulated time to chart closure: a novel electronic health record-derived metric associated with clinician burnout. JAMIA Open 2024;7(1):ooae009 [FREE Full text] [doi: 10.1093/jamiaopen/ooae009] [Medline: 38333109]
74. Lou SS, Kim S, Harford D, Warner BC, Payne PR, Abraham J, et al. Effect of clinician attention switching on workload and wrong-patient errors. Br J Anaesth 2022;129(1):e22-e24 [FREE Full text] [doi: 10.1016/j.bja.2022.04.012] [Medline: 35568509]
75. Kissler MJ, Kissler K, Burden M. Toward a medical "Ecology of Attention". N Engl J Med 2021;384(4):299-301 [FREE Full text] [doi: 10.1056/NEJMp2027190] [Medline: 33503689]
76. Zulman DM, Haverfield MC, Shaw JG, Brown-Johnson CG, Schwartz R, Tierney AA, et al. Practices to foster physician presence and connection with patients in the clinical encounter. JAMA 2020;323(1):70-81. [doi: 10.1001/jama.2019.19003] [Medline: 31910284]
77. Savoy A, Patel H, Murphy DR, Meyer AND, Herout J, Singh H. Electronic health records' support for primary care physicians' situation awareness: a metanarrative review. Hum Factors 2023;65(2):237-259 [FREE Full text] [doi: 10.1177/00187208211014300] [Medline: 34033500]
78. Shreffler J, Huecker M. Physician flow at work: examining work absorption, clinical flow, work fulfillment, and flow thieves. Workplace Health Saf 2022;70(11):484-491. [doi: 10.1177/21650799221093772] [Medline: 35766249]
79. Nijor S, Rallis G, Lad N, Gokcen E. Patient safety issues from information overload in electronic medical records. J Patient Saf 2022;18(6):e999-e1003 [FREE Full text] [doi: 10.1097/PTS.0000000000001002] [Medline: 35985047]
80. Moy AJ, Aaron L, Cato KD, Schwartz JM, Elias J, Trepp R, et al. Characterizing multitasking and workflow fragmentation in electronic health records among emergency department clinicians: using time-motion data to understand documentation burden. Appl Clin Inform 2021;12(5):1002-1013 [FREE Full text] [doi: 10.1055/s-0041-1736625] [Medline: 34706395]
81. Lieu TA, Warton EM, East JA, Moeller MF, Prausnitz S, Ballesca M, et al. Evaluation of attention switching and duration of electronic inbox work among primary care physicians. JAMA Netw Open 2021;4(1):e2031856 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31856] [Medline: 33475754]

XSL•FO
RenderX

82. Melnick ER, Harry E, Sinsky CA, Dyrbye LN, Wang H, Trockel MT, et al. Perceived electronic health record usability as a predictor of task load and burnout among US physicians: mediation analysis. J Med Internet Res 2020;22(12):e23382 [FREE Full text] [doi: 10.2196/23382] [Medline: 33289493]
83. Westbrook JI, Raban MZ, Walter SR, Douglas H. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. BMJ Qual Saf 2018;27(8):655-663 [FREE Full text] [doi: 10.1136/bmjqs-2017-007333] [Medline: 29317463]
84. Orru G, Longo L. The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and Germane loads: a review. In: Human Mental Workload: Models and Applications. Cham: Springer International Publishing; 2019.

Abbreviations

EHR: electronic health record
HITECH: Health Information Technology for Economic and Clinical Health
HIT: health information technology
NASA-TLX: National Aeronautics and Space Administration–Task Load Index

Edited by A Mavragani; submitted 24.07.24; peer-reviewed by K McVeigh, A Yazdanian, M Mun, J Hron, D Chrimes; comments to author 17.12.24; revised version received 15.01.25; accepted 11.02.25; published 06.03.25.

Please cite as:
Tawfik D, Rule A, Alexanian A, Cross D, Holmgren AJ, Lou SS, McPeek Hinz E, Rose C, Viswanadham RVN, Mishuris RG, Rodríguez-Fernández JM, Ford EW, Florig ST, Sinsky CA, Apathy NC
Emerging Domains for Measuring Health Care Delivery With Electronic Health Record Metadata
J Med Internet Res 2025;27:e64721
URL: https://www.jmir.org/2025/1/e64721
doi:10.2196/64721
PMID:40053814

©Daniel Tawfik, Adam Rule, Aram Alexanian, Dori Cross, A Jay Holmgren, Sunny S Lou, Eugenia McPeek Hinz, Christian Rose, Ratnalekha V N Viswanadham, Rebecca G Mishuris, Jorge M Rodríguez-Fernández, Eric W Ford, Sarah T Florig, Christine A Sinsky, Nate C Apathy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Health Communication on the Internet: Promoting Public Health and Exploring Disparities in the Generative AI Era

Jamal Uddin[1], PhD; Cheng Feng[2,3], PhD; Junfang Xu[4], PhD

[1]Communication Department, Cornell University, Ithaca, NY, United States

[2]Vanke School of Public Health, Tsinghua University, Beijing, China

[3]Institute for Healthy China, Tsinghua University, Beijing, China

[4]Department of Pharmacy, Second Affiliated Hospital, School of Public health, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Junfang Xu, PhD
Department of Pharmacy, Second Affiliated Hospital, School of Public health
Zhejiang University School of Medicine
866 Yuhangtang road
Xihu district
Hangzhou, 310058
China
Phone: 86 18801230482 ext 000
Fax: 86 0571 88981358
Email: junfangxuhappy1987@163.com

Abstract

Health communication and promotion on the internet have evolved over time, driven by the development of new technologies, including generative artificial intelligence (GenAI). These technological tools offer new opportunities for both the public and professionals. However, these advancements also pose risks of exacerbating health disparities. Limited research has focused on combining these health communication mediums, particularly those enabled by new technologies like GenAI, and their applications for health promotion and health disparities. Therefore, this viewpoint, adopting a conceptual approach, provides an updated overview of health communication mediums and their role in understanding health promotion and disparities in the GenAI era. Additionally, health promotion and health disparities associated with GenAI are briefly discussed through the lens of the Technology Acceptance Model 2, the uses and gratifications theory, and the knowledge gap hypothesis. This viewpoint discusses the limitations and barriers of previous internet-based communication mediums regarding real-time responses, personalized advice, and follow-up inquiries, highlighting the potential of new technology for public health promotion. It also discusses the health disparities caused by the limitations of GenAI, such as individuals' inability to evaluate information, restricted access to services, and the lack of skill development. Overall, this study lays the groundwork for future research on how GenAI could be leveraged for public health promotion and how its challenges and barriers may exacerbate health inequities. It underscores the need for more empirical studies, as well as the importance of enhancing digital literacy and increasing access to technology for socially disadvantaged populations.

(J Med Internet Res 2025;27:e66032) doi:10.2196/66032

KEYWORDS

internet; generative AI; artificial intelligence; ChatGPT; health communication; health promotion; health disparity; health; communication; internet; AI; generative; tool; genAI; gratification theory; gratification; public health; inequity; disparity

Background

Health communication on the internet has continually evolved, driven by technological advancements [1-4]. Over time, the public, along with health researchers, practitioners, and organizations, have increasingly embraced these technologies for diverse health-related purposes. The emergence of generative artificial intelligence (GenAI) represents a revolutionary development in health communication and public health promotion. GenAI, a novel communication tool, mimics human responses during follow-up inquiries. Its conversational nature is particularly advantageous for health communication, enabling it to address various health inquiries, deliver customized health information, provide education, offer consultations, assist in

XSL•FO
RenderX

disease diagnosis, simplify complex health reports, and translate health-related questions into multiple languages [5-7].

However, challenges such as accuracy, skill requirements, and access barriers persist with this technology [8-11]. GenAI operates on a large language model (LLM) built from publicly available data sources, including Google, Wikipedia, and GitHub [12]. As a result, the outputs generated by GenAI should be approached with caution to mitigate potential health risks. The quality of AI-generated results depends on the user's ability to develop effective prompts, input accurate text for inquiries, and access advanced features through subscriptions. Consequently, individuals with limited health literacy, insufficient prompt development skills, or an inability to afford premium subscriptions may miss out on these technological benefits, potentially exacerbating health disparities.

Earlier studies have examined the internet as a medium for health communication across various contexts, including persuasive health messaging via the World Wide Web [1], web-based consumer health information [13], the role of Web 2.0 in health and education [2], the internet's role in health interventions [14], the use of social media for health promotion and behavior change [15], and the integration of mobile devices in health care and interventions [16]. Additionally, a comprehensive review of social media apps and their purposes has been conducted [17]. However, recent technological advances, such as wearable devices, the Internet of Things, blockchain, and GenAI, present new opportunities as health communication mediums. Research combining these latest technologies as health communication tools and examining their applications for health promotion and understanding health disparities remains limited.

To address this research gap, this viewpoint adopts a conceptual approach by integrating existing literature and targeted web searches [4,13,18] to provide an updated overview of health communication mediums and their role in understanding health promotion and disparities in the GenAI era. Additionally, health promotion and health disparities are briefly analyzed through the lenses of the Technology Acceptance Model 2 (TAM2) [19], the uses and gratifications theory [20], and the knowledge gap hypothesis [21].

This conceptual study is organized into three sections as follows: (1) a brief overview of internet-based communication mediums and their health promotion practices, (2) the digital divide on the internet, and (3) health promotion and health disparities in the era of GenAI.

Health Communication Mediums on the Internet

The internet has become a vital medium for health communication in the era of Web 2.0 [22], evolving across various domains such as social media [17,23], mobile health [24], the Internet of Things [25], blockchain [26], wearable devices [27,28], and, more recently, GenAI [12]. These mediums also include health websites [13], health forums and communities [29], health-related webinars and podcasts [30,31], telemedicine [32,33], electronic health records [34], voice assistants [35,36], and augmented and virtual realities [37], all of which support a wide variety of health promotion practices and research [13].

Internet-based resources offer numerous features for persuasive communication aimed at promoting health behavior [1]. Therefore, an updated list of these communication mediums and their potential for health promotion practices are presented in Table 1, serving as a valuable resource from a health communication perspective.

XSL•FO
RenderX

Table 1. Major health communication mediums and their potential for health promotion practices.

Web-based health communication mediums	Potentiality	Health promotion practices (examples)
Web 2.0	Interactivity (eg, chat rooms, emails, and hyperlinks)	Responsive user interface; provision of rich content including text, images, and videos; and reusability and collaboration opportunities
Health websites	Featured health content (eg, health articles, audios, videos, images, and infographics)	Mayo Clinic, WebMD, and Centers for Disease Control and Prevention
Social media	Networking (eg, social support, community development, knowledge sharing, and doctor-patient interactions)	X, Facebook, WeChat, YouTube, WhatsApp, and Instagram
Health forums and web-based communities	Groups with common health interests (eg, mutual support, discussion of common problems, patient empowerment, anonymity, and emotional support)	ParkinsonNet, COPD.com, and PatientLikeMe
Health-related webinars and podcasts	Health knowledge dissemination through live sessions or on demand (eg, narrative communication and health education)	Cancer Council Victoria, Hidden Brain, The Blindboy Podcast, and The Two Norries
Mobile health	Easy and on-the-go access to information (eg, web browsing, audio and video calls, and text messaging)	Cellphones, smartphones, and smartphones with specific software or apps
Telemedicine	Health services provided remotely (eg, using telecommunication technology for doctor-patient communication)	Physicians, nurses, and other allied health professionals
Electronic health records	Patient health records stored in computer hard drives or web-based servers (eg, easy access from anywhere for multiple parties, and easy for health decisions and diagnosis)	Physicians, insurance companies, and patients
Blockchain	Digital ledger (eg, secured data storing and exchange platform)	Health care industry and electronic health record researchers
Voice assistants	Conversational assistance (eg, health information inquiry with voice)	Amazon Alexa, Google Home, Google Assistant, Apple Siri, and Samsung Bixby
Wearable devices	Health monitoring and remote health services (eg, glasses, smartwatches, and suits)	Emergency medical service providers, family members, patients, public, and clinicians
Augmented reality and virtual reality	Enhanced views combining real and digital objects (eg, health education, training, and assessment)	Health care industry
Internet of Things	Connectivity and automation (eg, collecting, sending, receiving, storing, and measuring data of anyone, anything, anytime, any service, in any network)	Health care and health management
Generative artificial intelligence	Humanlike interaction digitally (eg, content generation regarding text, image, audio, video, synthetic data, and code)	Public, health care industry, health professionals, and researchers

Digital Divide on the Internet

The opportunities for web-based health communication also contribute to the digital divide. The internet serves as a vast resource of health benefits for many who seek health information and use it for health purposes, depending on their access, skills, usage, and participation [38,39]. Educated, older, and tech-savvy individuals benefit the most from the internet for health-related purposes [38]. However, health content contributed by individuals without professional health knowledge or training creates complexities in finding reliable information. This can harm those who lack the ability to evaluate the credibility of information or its sources, as much of the web-based health content available is incomplete, misleading, or difficult to understand [39].

Today, people encounter misinformation in various ways, such as through search engines (eg, random searches on Google and Yahoo suggesting links based on top-matching results), visiting specific health websites (eg, the quality of information depends on whether the person is visiting a reputable site like WebMD), relying on user-generated content from social media or other websites (eg, Facebook, Twitter, Wikipedia, and Yelp), and using mobile apps that may not adhere to established medical guidelines [40]. Certain health topics are particularly susceptible to misinformation on social media, including vaccines, drugs and smoking, noncommunicable diseases, pandemics, eating disorders, and medical treatments [41].

Traditionally, the digital divide has been understood in terms of physical access to the internet, which depends on factors such as age, education, income, perceived health, and social isolation [42,43]. However, the digital divide for health purposes has widened due to low levels of health literacy (eg, the ability to understand, appraise, and use web-based information for health purposes) and inadequate internet use skills (eg, browsing and content-related skills), which are more prevalent among hard-to-reach communities. Non-English speakers, who make up around 13% of the US immigrant population, are less likely to use the internet for health purposes [44]. For example, the use of telemedicine for primary care visits by patients with limited English proficiency and digital access decreased

XSL•FO
RenderX

proportionately in New York City during COVID-19 [45]. Such digital divides may further worsen during new waves of technological advancement, such as GenAI.

GenAI: Health Promotion and Health Disparity

Overview

GenAI presents a significant opportunity for health promotion and may have a revolutionary impact on the health care system by leveraging vast amounts of medical data and knowledge to support medical decision-making, patient communication, health education, and data management [7]. However, it may also contribute to health disparities due to limited access, insufficient prompt development skills, and low health literacy among many populations [44].

This section provides an overview of the major GenAI channels, their potential for health promotion, and how their limitations, particularly lack of access and skills, may exacerbate health disparities.

GenAI and Health Promotion

Application of GenAI

GenAI is poised to transform content creation across text, image, audio, and video formats [46]. Its ability to generate humanlike responses based on prompts has created new opportunities for health promotion. Potential users of these tools include the general public, health researchers, practitioners, and health organizations [5-7,17].

A variety of GenAI tools are currently available, each with a specific focus, such as text generation, and audio, video, and image creation [5,9,46-50]. Table 2 provides a brief overview of some notable GenAI channels, their potential uses, target users, and the skill requirements for health-related purposes.

Table 2. Generative artificial intelligence (GenAI) channels and their potential use, users, and skills requirement for health purposes.

GenAI channels	Use	Users	Skills
Text generation			
ChatGPT	Humanlike conversation (eg, personalized health information, answering health-related questions, interaction with chatbots, and facilitating patient-provider communication)	Individuals, health educators, caregivers, and health organizations	Prompt development skills (eg, knowledge of medical terminology, patient communication principles, and using empathy in responses)
Gemini	Visual content creation from textual descriptions (eg, health literacy and health education)	Individuals, health care educators, health organizations, medical illustrators, and health communicators	Prompt development skills, medical concepts, and visual communication techniques
Microsoft Copilot	Development of health-related software apps and digital tools	Health IT professionals, software developers, and digital solution providers	Prompt development skills, knowledge of health care standards, Health Insurance Portability and Accountability Act compliance, and user interface and user experience design principles
Jasper	High-quality, text-based content generation (eg, blog posts and social media posts)	Content creators, patient educators, and health organizations	Prompt development skills, understanding health topics, medical knowledge, and ethical considerations
Image, audio, and video generation			
DALL-E 3 and Midjourney	Concept art, illustration, and graphic design	Artists, designers, and concept artists	Basic prompt development skills and familiarity with different artistic styles
JukeBox and MuseNet	Music composition, songwriting, and sound design	Musicians, producers, and game and film developers	Text prompt development and musical knowledge
Synthesia and Runway	Educational videos, marketing and communication, and social media content development	Educators, marketers, communicators, and content creators	Prompt development, basic video editing skills, and understanding the target audiences

Health Promotion Through GenAI

Traditionally, individuals communicated with health care professionals, family members, and friends for health-related purposes [51,52]. This communication provided immediate responses to health needs and allowed for follow-up questions. However, barriers included difficulties in reaching people in real time, time and cost constraints, limited understanding of health issues, and discomfort in discussing stigmatized health topics like addiction and sexually transmitted infections. The development of Web 2.0, social media, and mobile apps addressed many of these barriers [53-55]. These platforms offer health information from both individuals and professionals, enabling people to access information as needed, provided they possess the necessary technology, internet skills, and the ability to process web-based content [38,39].

Despite these advancements, existing mediums have limitations, such as the lack of interactive health information without human involvement, delayed responses, limited data for mobile app responses, difficulty understanding health information in specific

XSL•FO
RenderX

contexts, and limited accessibility for non-English speakers. GenAI has the potential to overcome these barriers by providing interactive, real-time responses and personalized health information based on LLMs. GenAI's ability to pass challenging exams (eg, the US Medical Licensing Exam and the USA Biology Olympiad Semi-Final Exam) underscores its capacity to deliver tailored responses aligned with human needs and inquiries [56].

Individuals can now fulfill various needs—such as health inquiries, image development, and health content creation—using GenAI tools, if they have access to these tools and effective prompt writing skills. The acceptance and use of this new technology can also be explored through theoretical frameworks.

Theoretical Application of GenAI

The application of GenAI can be examined through the TAM2 and the uses and gratifications theory. TAM2 suggests that an individual's acceptance or rejection of a system or technology depends on its perceived usefulness and ease of use [19]. Perceived usefulness is determined by factors such as the quality of the outcomes and the benefits provided, while perceived ease of use refers to how simple the technology is to operate, learn, and become proficient with, requiring minimal effort [57]. GenAI has demonstrated its usefulness by solving complex problems, generating humanlike responses to diverse inquiries, creating content, and acting as an assistant. As a result, health educators, medical professionals, and researchers have started incorporating these tools into their work. Consequently, the general public is increasingly likely to adopt GenAI for health purposes, such as obtaining personalized health information, interpreting health reports, and making informed health decisions.

The technology is also designed to be user-friendly, resembling the simplicity of a Google search. Major GenAI tools, such as ChatGPT, Gemini, and Copilot, are accessible with straightforward sign-up processes and offer basic services at no cost. The rapid adoption of this technology is evident from its widespread use. For example, since its launch on November 30, 2022, ChatGPT has amassed over 300 million users weekly, with 1 billion messages sent by its users on the platform each day [58]. In the United States, ChatGPT app downloads exceeded 32.6 million, and Gemini, launched on December 6, 2023, reached 12.67 million downloads by September 2024 [59]. However, realizing the full benefits of this technology often depends on the user's prompt development skills.

The motives behind the widespread use of GenAI can be analyzed through the uses and gratifications theory [20]. According to this theory, individuals consciously select specific media to fulfill their needs. The primary motivations for using GenAI include its humanlike responsiveness, which addresses limitations in conventional internet-based communication; its ability to deliver tailored health information, unlike standard search engines that primarily return links; and its capacity to meet personalized health needs, such as seeking health information, diagnosing conditions, making health decisions, and creating health-related content. Additionally, the ability to use AI agent services for task completion motivates individuals, particularly health researchers, campaign designers, educators, and medical professionals, to integrate GenAI into their work.

Despite its significant potential, GenAI has limitations that may contribute to health disparities.

GenAI and Health Disparity

Overview

GenAI has several limitations, including its reliance on specific databases, difficulty in understanding users' unique contexts, and individuals' inability to fully use its benefits. Some of these limitations are discussed below:

Limitations of GenAI

GenAI relies on LLMs trained on data from sources such as Google, Wikipedia, and GitHub. However, because data on the internet can include inaccurate information and misinformation—often contributed by individuals without relevant expertise or oversight—LLM-based AI tools may generate inaccurate or biased responses. For example, researchers evaluated ChatGPT-4's health-related information in the field of radiology and found that only 65% of the information was accurate [11]. Similarly, while the same tool passed the medical licensing exam, it achieved a passing mark of approximately 60% [56], suggesting a potential inaccuracy or bias rate of over 30%.

Limitations of Personalized Health Advice

Personalized health care relies on various factors specific to an individual's health condition, typically assessed through diagnostic methods such as pulse checks, blood pressure readings, blood tests, x-rays, and cultural considerations. While GenAI tools can provide health suggestions based on detailed prompts or text inputs, their recommendations are derived from universal datasets, which may not always be applicable to personalized care [9]. Furthermore, GenAI has limitations in understanding individual cultural and racial contexts [60], which are critical for determining health care needs and preferences. As a result, researchers recommend verifying GenAI-generated outputs with qualified health care professionals to ensure accuracy and relevance.

Individual Incapacity to Evaluate Health Information

Health literacy, the ability to assess information accuracy, and knowledge of reliable sources are essential for individuals to evaluate specific health information and use it effectively [61]. Individuals who possess these skills can maximize the benefits of GenAI technology. However, racial and ethnic minority groups, who often experience lower levels of health literacy and lack the skills needed to evaluate health information, may be disproportionately deprived of these technological advantages.

Limitations of Access to GenAI Tools

Although 95% of American adults have internet access [62], many individuals with lower health literacy, language barriers, or limited financial resources face challenges in accessing the full benefits of GenAI tools. These tools often require paid subscriptions to unlock advanced features and enhanced accuracy. For instance, ChatGPT Plus, a subscription-based

XSL•FO
RenderX

version, delivers more accurate outputs compared with the free version, ChatGPT 3.5 [10].

Lack of Skills in Prompt Development

The quality of GenAI output is highly dependent on effective prompt development, which involves crafting precise text inputs for the tools [8]. Individuals proficient in prompt development—using techniques such as open-ended questions and using relevant keywords—can maximize the benefits of this technology. Conversely, individuals with lower levels of education, poor health literacy, or limited technological proficiency may struggle to harness its potential benefits, leaving them at a disadvantage.

Potential Health Disparity Through GenAI

Health disparity refers to differences in health status between disadvantaged social groups—primarily individuals with low income, racial and ethnic minority groups, women, and others who consistently experience worse health and greater health risks compared with advantaged social groups [63]. In the context of web-based health benefits, health disparity is influenced by limitations in access to information technology, poor health literacy, language barriers, and inadequate internet use skills, such as browsing and content-related skills [44].

In the era of GenAI, health disparities are likely to persist among these groups due to challenges in evaluating GenAI outcomes and difficulties in prompt development. Current ChatGPT users, predominantly college students (43%) and Fortune 500 companies (80%), are likely drawn from higher social groups that benefit most from this technology. These advantaged groups, with higher levels of education, communication skills, and acceptance of new technology, are more likely to maintain superior health knowledge and better health outcomes compared with disadvantaged groups, consistent with the knowledge gap hypothesis [21]. This theory posits that individuals with higher social status acquire information faster than those with lower social status. Consequently, health disparities are compounded as higher-status individuals use new technologies more effectively for purposeful activities, while less-educated individuals often use technology primarily for entertainment [64].

To reduce these disparities, efforts must focus not only on the availability of technology but also on improving accessibility, enhancing digital literacy, fostering engagement, and building trust [3].

Conclusion

This study provides an overview of major health communication mediums on the internet, their practices, and the limitations of traditional web-based health communication tools for health purposes. It highlights GenAI as a new avenue for health promotion while addressing its potential to exacerbate health disparities. These discussions offer valuable insights for researchers, educators, and health professionals focusing on internet-based health communication and health promotion practices.

Although this research lacks extensive evidence-based claims, it serves as a foundation for further empirical studies, such as interviews and surveys with marginalized communities, to better understand health disparities in the era of GenAI. The study could be further enhanced by conducting a systematic review to provide a more comprehensive understanding of communication mediums, their health-related practices on the internet, and health disparities in the rapidly evolving context of GenAI.

Moreover, policy makers and health educators should implement targeted programs to familiarize disadvantaged groups with the latest technologies and their benefits. Efforts should focus on reducing health disparities by improving digital literacy, enhancing prompt development skills, and increasing access to technology for underserved populations.

Acknowledgments

We are grateful to all study participants for their participation. This work was supported by Fundamental Research Funds for the Central Universities and the project evaluation of Chronic Disease Health Management Policies (project 2023-SKY-A07054-0010).

Data Availability

All data generated or analyzed during this study are included in this published paper.

Authors' Contributions

JU conducted the data analysis and drafted the paper. JX and CF contributed to the study’s interpretation of the data and critical revisions of the paper. All authors have approved the final versions for submission.

Conflicts of Interest

None declared.

References

1. Cassell C, Jackson C, Cheuvront B. Health communication on the internet: an effective channel for health behavior change? J Health Commun 1998;3(1):71-79 [FREE Full text] [doi: 10.1080/108107398127517]

XSL•FO
RenderX

2. Kamel Boulos MN, Wheeler S. The emerging web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libraries J 2007;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x]
3. Koh A, Swanepoel DW, Ling A, Ho BL, Tan SY, Lim J. Digital health promotion: promise and peril. Health Promot Int 2021;36(Supplement_1):i70-i80 [FREE Full text] [doi: 10.1093/heapro/daab134] [Medline: 34897444]
4. Korp P. Health on the internet: implications for health promotion. Health Educ Res 2006;21(1):78-86 [FREE Full text] [doi: 10.1093/her/cyh043] [Medline: 15994845]
5. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
6. Santandreu-Calonge D, Medina-Aguerrebere P, Hultberg P, Shah MA. Can ChatGPT improve communication in hospitals? EPI 2023;32(2) [FREE Full text] [doi: 10.3145/epi.2023.mar.19]
7. Zhang P, Kamel Boulos MN. Generative AI in medicine and healthcare: promises, opportunities and challenges. Future Internet 2023;15(9):286 [FREE Full text] [doi: 10.3390/fi15090286]
8. Getting started with prompts for text-based generative AI tools. Harvard University. URL: https://huit.harvard.edu/news/ai-prompts [accessed 2024-04-28]
9. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25:e50638 [FREE Full text] [doi: 10.2196/50638] [Medline: 37792434]
10. 107 Up-to-date ChatGPT statistics & user numbers. Nerdynav. 2024. URL: https://nerdynav.com/chatgpt-statistics/ [accessed 2024-04-02]
11. Rajagopal S, Hazarika S, Kim, S, Chiou YM, Sohn JH, Subramonyam H, et al. Are generative AI systems capable of supporting information needs of patients? arXiv. Preprint posted online on January 31, 2024 . [doi: 10.48550/arXiv.2402.00234]
12. Gozalo-Brizuela R, Garrido-Merchan EC. ChatGPT is not all you need. a state-of-the-art review of large generative AI models. arXiv. Preprint posted online on January 11, 2023 [FREE Full text] [doi: 10.48550/arXiv.2301.04655]
13. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. Health Educ Res 2001;16(6):671-692 [FREE Full text] [doi: 10.1093/her/16.6.671] [Medline: 11780707]
14. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
15. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2011;14(1):15-23 [FREE Full text] [doi: 10.1177/1524839911405850]
16. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
17. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
18. Gilson LL, Goldberg CB. Editors' comment: so, what is a conceptual paper? Group Organ Manag 2015;40(2):127-130. [doi: 10.1177/1059601115576425]
19. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014;14(1):81-95 [FREE Full text] [doi: 10.1007/s10209-014-0348-1]
20. Katz E, Blumler JG, Gurevitc M. Uses and gratifications research. Public Opin Q 1973;37(4):509-523 [FREE Full text] [doi: 10.1086/268109]
21. Tichenor PJ, Donohue GA, Olien CN. Mass media flow and differential growth in knowledge. Public Opin Q 1970:159-170 [FREE Full text] [doi: 10.1086/267786]
22. Murugesan S. Understanding web 2.0. IT Prof 2007;9(4):34-41 [FREE Full text] [doi: 10.1109/mitp.2007.78]
23. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933]
24. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
25. Kelly JT, Campbell KL, Gong E, Scuffham P. The internet of things: impact and implications for health care delivery. J Med Internet Res 2020;22(11):e20135 [FREE Full text] [doi: 10.2196/20135] [Medline: 33170132]
26. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-a scoping review. Int J Med Inform 2020;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
27. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
28. Majumder S, Mondal T, Deen MJ. Wearable sensors for remote health monitoring. Sensors (Basel) 2017;17(1):130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]

XSL•FO
RenderX

29. van der Eijk M, Faber MJ, Aarts JWM, Kremer JAM, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
30. Amador FLD, Alves GCG, Santos VRD, Moreira RSL. Use of podcasts for health education: a scoping review. Rev Bras Enferm 2024;77(1):e20230096 [FREE Full text] [doi: 10.1590/0034-7167-2023-0096] [Medline: 38511787]
31. Chiswell M, Smissen A, Ugalde A, Lawson D, Whiffen R, Brockington S, et al. Using webinars for the education of health professionals and people affected by cancer: processes and evaluation. J Cancer Educ 2018;33(3):583-591. [doi: 10.1007/s13187-016-1138-7] [Medline: 27981435]
32. Matusitz J, Breen GM. Telemedicine: its effects on health communication. Health Commun 2007;21(1):73-83 [FREE Full text] [doi: 10.1080/10410230701283439] [Medline: 17461754]
33. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
34. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2018;25(S 01):S48-S61 [FREE Full text] [doi: 10.15265/iys-2016-s006]
35. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of siri, alexa, and google assistant. J Med Internet Res 2018;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
36. Hong G, Folcarelli A, Less J, Wang C, Erbasi N, Lin S. Voice assistants and cancer screening: a comparison of alexa, siri, google assistant, and cortana. Ann Fam Med 2021;19(5):447-449 [FREE Full text] [doi: 10.1370/afm.2713] [Medline: 34546951]
37. Carmigniani J, Furht B. Augmented reality: an overview. In: Furht B, editor. Handbook of Augmented Reality. New York, NY: Springer; 2011.
38. Blank G, Lutz C. Benefits and harms from internet use: a differentiated analysis of Great Britain. New Media Soc 2016;20(2):618-640 [FREE Full text] [doi: 10.1177/1461444816667135]
39. Levy JA, Strombeck R. Health benefits and risks of the internet. J Med Syst 2002;26(6):495-510. [doi: 10.1023/a:1020288508362] [Medline: 12385532]
40. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020;41(1):433-451 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094127]
41. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
42. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2019;24(12):1668-1675 [FREE Full text] [doi: 10.1177/1359105317695429] [Medline: 28810415]
43. van Deursen A, van Dijk J. Internet skills and the digital divide. New Media Soc 2010;13(6):893-911 [FREE Full text] [doi: 10.1177/1461444810386774]
44. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the internet as a resource for health information: results from a California population-based study. Int J Med Inform 2017;103:49-54 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.04.008]
45. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, et al. The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York City. J Urban Health 2021;98(2):183-186 [FREE Full text] [doi: 10.1007/s11524-020-00508-9]
46. Feuerriegel S, Hartmann J, Janiesch C, Zschech P. Generative AI. Bus Inf Syst Eng 2023;66(1):111-126 [FREE Full text] [doi: 10.1007/s12599-023-00834-7]
47. Iorliam A, Ingio JA. A comparative analysis of generative artificial intelligence tools for natural language processing. J Comput Theor Appl 2024;1(3):311-325 [FREE Full text] [doi: 10.62411/jcta.9447]
48. Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. J Inf Technol Case Appl Res 2023;25(3):277-304 [FREE Full text] [doi: 10.1080/15228053.2023.2233814]
49. Saleem A. 7 tested prompting techniques to use AI video generators - evolution of art industry. Data Science Dojo. 2024 Feb 24. URL: https://datasciencedojo.com/blog/prompting-to-use-ai-video-generators/ [accessed 2025-01-06]
50. Shin R. Generative AI: how your business can leverage it. Coveo. 2024. URL: https://www.coveo.com/blog/what-is-generative-ai/ [accessed 2024-03-26]
51. Kivits J. Researching the 'Informed Patient. Inf Commun Soc 2004;7(4):510-530 [FREE Full text] [doi: 10.1080/1369118042000305629]
52. Rains SA. Perceptions of traditional information sources and use of the world wide web to seek health information: findings from the health information national trends survey. J Health Commun 2007;12(7):667-680 [FREE Full text] [doi: 10.1080/10810730701619992]
53. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005;61(8):1821-1827 [FREE Full text] [doi: 10.1016/j.socscimed.2005.03.025]

XSL•FO
RenderX

54. de Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. 2014 Presented at: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26 to May 1, 2014; Toronto, ON URL: https://doi.org/10.1145/2556288.255721 [doi: 10.1145/2556288.2557214]
55. Yang G, Long J, Luo D, Xiao S, Kaminga AC. The characteristics and quality of mobile phone apps targeted at men who have sex with men in China: a window of opportunity for health information dissemination? JMIR Mhealth Uhealth 2019;7(3):e12573 [FREE Full text] [doi: 10.2196/12573] [Medline: 30916658]
56. Varanasi L. GPT-4 can ace the bar, but it only has a decent chance of passing the CFA exams. here's a list of difficult exams the ChatGPT and GPT-4 have passed. Business Insider. URL: https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1 [accessed 2024-09-15]
57. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002]
58. Roth E. ChatGPT now has over 300 million weekly users. The Verge. URL: https://www.theverge.com/2024/12/4/24313097/chatgpt-300-million-weekly-users [accessed 2024-12-04]
59. Bianchi T. Number of monthly ChatGPT and Gemini mobile app downloads worldwide from May 2023 to September 2024. Statista. URL: https://www.statista.com/statistics/1497377/global-chatgpt-vs-gemini-app-downloads/ [accessed 2024-10-10]
60. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024;6(1):e12-e22 [FREE Full text] [doi: 10.1016/s2589-7500(23)00225-x]
61. Mantwill S, Monestel-Umaña S, Schulz PJ. The relationship between health literacy and health disparities: a systematic review. PLoS One 2015;10(12):e0145455 [FREE Full text] [doi: 10.1371/journal.pone.0145455] [Medline: 26698310]
62. Internet, broadband fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2024-05-02]
63. Braveman P. Health disparities and health equity: concepts and measurement. Annu Rev Public Health 2006;27:167-194 [FREE Full text] [doi: 10.1146/annurev.publhealth.27.021405.102103] [Medline: 16533114]
64. Bonfadelli H. The internet and knowledge gaps. Eur J Commun 2002;17(1):65-84 [FREE Full text] [doi: 10.1177/0267323102017001607]

Abbreviations

AI: artificial intelligence
GenAI: generative artificial intelligence
LLM: large language model
TAM2: Technology Acceptance Model 2

Edited by T de Azevedo Cardoso; submitted 02.09.24; peer-reviewed by I Skafle, J Volkman; comments to author 02.12.24; revised version received 15.01.25; accepted 31.01.25; published 06.03.25.

Please cite as:
Uddin J, Feng C, Xu J
Health Communication on the Internet: Promoting Public Health and Exploring Disparities in the Generative AI Era
J Med Internet Res 2025;27:e66032
URL: https://www.jmir.org/2025/1/e66032
doi:10.2196/66032
PMID:40053755

©Jamal Uddin, Cheng Feng, Junfang Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Assessing Digital Maturity of Hospitals: Viewpoint Comparing National Approaches in Five Countries

Kathrin Cresswell[1], PhD; Franziska Jahn[2], PhD; Line Silsand[3], PhD; Leanna Woods[4], PhD; Tim Postema[5], PhD; Marion Logan[6], MA; Sevala Malkic[7], MSc; Elske Ammenwerth[8], PhD

[1]Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[2]Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

[3]Department of Health and Care Science, Faculty of Health Sciences, UiT The Arctic University of Norway, Bergen, Norway

[4]Queensland Digital Health Centre, The University of Queensland, Brisbane, Australia

[5]Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[6]Digital Leadership and Capabilities, Digital Health and Care, Scottish Government, Edinburgh, United Kingdom

[7]Division Digital Transformation, Norwegian Directorate of Health, Olso, Norway

[8]Institute of Medical Informatics, UMIT TIROL, Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria

Corresponding Author:
Kathrin Cresswell, PhD
Usher Institute
The University of Edinburgh
Usher Building
5-7 Little France Road, Edinburgh BioQuarter-Gate 3
Edinburgh, EH16 4UX
United Kingdom
Phone: 44 (0)131 651 7869
Email: Kathrin.Cresswell@ed.ac.uk

Abstract

Digital maturity assessments can inform strategic decision-making. However, national approaches to assessing the digital maturity of health systems are in their infancy, and there is limited insight into the context and processes associated with such assessments. This viewpoint article describes and compares national approaches to assessing the digital maturity of hospitals. We reviewed 5 national approaches to assessing the digital maturity of hospitals in Queensland (Australia), Germany, the Netherlands, Norway, and Scotland, exploring context, drivers, and approaches to measure digital maturity in each country. We observed a common focus on interoperability, and assessment findings were used to shape national digital health strategies. Indicators were broadly aligned, but 4 of 5 countries developed their own tailored indicator sets. Key topic areas across countries included interoperability, capabilities, leadership, governance, and infrastructure. Analysis of indicators was centralized, but data were shared with participating organizations. Only 1 setting conducted an academic evaluation. Major challenges of digital maturity assessment included the high cost and time required for data collection, questions about measurement accuracy, difficulties in consistent long-term tracking of indicators, and potential biases due to self-reporting. We also observed tensions between the practical feasibility of the process with the depth and breadth required by the complexity of the topic and tensions between national and local data needs. There are several key challenges in assessing digital maturity in hospitals nationally that influence the validity and reliability of output. These need to be explicitly acknowledged when making decisions informed by assessments and monitored over time.

(J Med Internet Res 2025;27:e57858) doi:10.2196/57858

KEYWORDS

digital maturity; hospitals; assessment; decision making; health systems; infancy; Australia; Australian; data collection; data; qualitative thematic analysis; self-reporting; practical feasibility; long-term tracking; local data; monitoring

XSL•FO
RenderX

Introduction

Health systems are facing increasing pressure to achieve more with less, and health information technology is widely recognized to tackle existing challenges, although it is still difficult to attribute causal connections between health information technology and health care outcomes [1].

The concept of digital maturity has emerged from the field of organizational studies. It has been conceptualized as an organizational state in which processes are digitalized and accompanied by other organizational transformations [2]. The assumption underlying the concept of digital maturity in health care is that organizations that have higher levels, achieve better processes and outcomes [3,4]. This may include improved quality and operational efficiency through implementing electronic health records (EHRs) and laying the foundation for learning health care systems, artificial intelligence, and precision medicine [5]. We here define an EHR as a digital version of a patient's paper chart including medical history, treatment, and care over time.

The World Health Organization supports driving the improvement of digital maturity in health systems, aiming toward Universal Health Coverage (UHC) [6]. National approaches to measuring digital maturity have emerged in public health systems relatively recently and are now increasingly applied to inform strategic decision-making [7]. Although some measurements are widely applied (eg, those by the Healthcare Information and Management Systems Society [HIMSS]) [8], there are currently no universally accepted methods for measuring digital maturity in health systems [9]. This may partly be due to the evolving landscape of digital health, where the concept of digital maturity is difficult to define. Some have argued that it needs to be conceptualized as a journey without a definitive endpoint [10], and digital health capability is sometimes used interchangeably with maturity to reflect a continuum of digital growth [11].

To understand how best to assess digital maturity in health systems to facilitate comparisons and sharing of lessons, there is a need to compare initiatives across nations. Although approaches to measuring digital maturity have been analyzed within countries and local organizations [4,12,13], there is currently limited evidence surrounding international comparisons. A nuanced view of the role of digital maturity assessments will help to shape implementation efforts. This should include discussions on the value gained from different digital maturity assessments, focusing on their effectiveness, return on investment, and impact on the quality and safety of care.

We therefore aimed to describe and compare national approaches to assess digital maturity. We hope that this work will contribute to fostering an environment of learning and sharing lessons, which can inform future strategies and international comparisons.

Methods

Countries included in this review were sampled opportunistically through our existing networks in the International Medical Informatics Association (IMIA) and European Federation of Medical Informatics Evaluation Working Groups on technology assessment and evaluation [14,15]. Snowballing through this network identified additional stakeholders. To be included, we needed to be able to access information on the digital maturity assessment process, and we therefore in some cases worked with policy makers who were involved in the exercise within their countries to gather data. We had to find a balance between including countries that were broadly comparable in relation to existing levels of digital maturity, whilst also being sufficiently different to explore varying approaches. Whilst we acknowledge that including countries at various stages in their digital maturity assessment journey could be valuable, we here focused on countries at similar stages but with different approaches (national, regional), different population sizes, different items with various origins (HIMSS, WHO, expert stakeholder consultations) and variations in EHR systems and coverage. Each of the included countries (Australia [Queensland], Germany, the Netherlands, Norway, and Scotland) has a unique health care system shaped by different policies, regulations, and funding mechanisms. Comparing digital maturity across these varied systems, allowed us to gain insights into how different approaches to health care administration impact the adoption and implementation of digital technologies. Furthermore, cultural attitudes toward technology, patient engagement, and health care delivery differ significantly across these countries and understanding how these cultural factors influence digital maturity is important.

We assigned academic (KC, FJ, LW, TP, and EA) or policy (ML and SM) leads to each of the participating countries, who were responsible for collecting descriptive data. Using the nominal group technique with leads, we cocreated a data collection template table for each country, representing key features and learnings identified through discussions in group meetings. These included a description of the context or setting, rationale and drivers for the digital maturity assessments, tools and methods used to collect digital maturity data, methods used to analyze and disseminate the data, challenges encountered, and lessons learned likely to be relevant to the international community (Multimedia Appendix 1). Missing fields were accepted and leads unable to provide complete datasets were still considered eligible. We then coded items under each table field using NVivo 12 (Lumivero) and extracted areas of convergence and areas of divergence under each of the categories through a qualitative thematic analysis. This involved dragging and dropping information recorded for all countries under each table field as the parent node, examining the coded text from across countries for each of the parent nodes. This led to a consolidation of parent nodes under the overall themes presented in the Results section (context, drivers, methods for data collection and analysis, and challenges encountered). We then, within these wider categories, explored similarities and differences between countries which presented the child nodes. Some of these took a numerical form, in which case we

XSL•FO
RenderX

examined ranges, whilst for qualitative data we paid particular attention to summarizing the diverse spectrum of experiences across contexts.

The constructed narrative including key themes (presented below) and implications of our findings were then discussed amongst all members of the team in meetings with country leads.

Results

A comparative overview of key descriptive data across countries is provided in Table 1. We further elaborate on the emerging themes in the paragraphs below.

Table 1. Summary of key descriptive data across countries.

Heading	Queensland (Australia)	Germany	The Netherlands	Norway	Scotland
Health care model	Queensland Health (state government) funds universal free health care	Bismarck model	National Health Insurance	National health and social Insurance	National Health Insurance
Size of country or region (# inhabitants), n	~5 million	~84 million	~18 million	~5.5 million	~5 million
EHR[a] coverage in hospitals	Varied (some still have paper, but single instance of Cerner across 15 hospitals)	Varied (>10 different vendors)	100% (four core EHR systems)	100%	100%
National or regional approach	Regional	National	National	National	National
Name of EHR	Cerner	Varied	Varied	DIPS, EPIC	TrakCare
Name of digital maturity assessment	HIMSS[b] Digital Health Indicator	DigitalRadar	Dutch Interoperability Maturity Model	eHealth Monitor	Digital maturity survey
Led by	Queensland Health (government)	Digital Radar Consortium (appointed or commissioned by Ministry of Health)	Ministry of Health, Welfare and Sports	Norwegian Government	Scottish Government and Convention of Scottish Local Authorities
Origin	Bespoke, based on HIMSS	Bespoke based on HIMSS and expert review	Bespoke, based on HIMSS	Bespoke based on criteria recommended in WHO's report on the development of national e-health strategies	Bespoke drawing on methodical review of UK and global models
Number of items, n	121	234	37	30	355

[a]EHR: electronic health record.

[b]HIMSS: Healthcare Information and Management Systems Society.

Context

The size of populations in participating countries varied significantly from 5 million in Queensland, Norway, and Scotland, to 18 million in the Netherlands, and 84 million in Germany.

None of the included health systems were financed through out-of-pocket models or private insurance-based financing models, and all had UHC. Health systems varied from The Beveridge model, where health care is provided free at the point of care through taxation (Scotland, Norway), The Bismarck model which is characterized by compulsory health insurance related to earnings (Germany, the Netherlands), and mixed models where universal public health insurance funded through tax can be supplemented by private health insurance (Queensland) [16].

All participating countries had a strong governmental focus on interoperability in their national strategies, but some had made more progress than others. This was particularly apparent when looking at regional examples. For instance, Queensland had a state-wide Integrated Electronic Medical Record with Cerner. In other countries, where regions had some autonomy over budgets and procurements, there were a multitude of heterogeneous EHR systems sometimes spanning more than 1 provider organization.

EHR coverage in hospitals was hard to assess, but between 50% and 100% of hospitals used some sort of clinical information system (ie, systems designed to support the operations of specific clinical departments or functions within an organization). Norway, Scotland, and the Netherlands had 100% coverage.

Types of EHR systems varied significantly. Some countries were dominated by 1 vendor (eg, Queensland with Cerner,

XSL•FO
RenderX

Scotland with TrakCare), whereas others used a mixture of systems including US-based mega systems (Epic, Cerner) and "home-grown" systems (Norway, Germany, the Netherlands).

Digital maturity assessments were national in all countries except Australia, which assessed the region of Queensland. All were initiated and executed by the Government, except in Germany where the government commissioned a consortium.

Although the focus of this analysis was on digital maturity in hospitals, some digital maturity assessments included several health and care settings, including acute, primary, and social care (Norway, Scotland, and the Netherlands).

All assessments had a longitudinal element, although what indicators were collected at any point in time was determined by local contextual factors such as availability and feasibility. In addition, the list of indicators evolved over time as some were added, and others were removed in line with strategic objectives and feasibility (explicitly mentioned in Germany and Scotland).

Rationale and Drivers for Digital Maturity Assessments

Across countries, the primary drivers for digital maturity assessments were to inform digital health strategy and priority areas of investment, evaluate national progress, and conduct local benchmarking designed to help participating organizations compare themselves to others. In Scotland, the local use of data to guide and plan digital transformation was emphasized, and the German case also mentioned a desire to compare their digital maturity with international centers of excellence.

Some cases also mentioned the exploitation of data generated and the use of data to drive and provide evidence for progressing interoperability agendas. The Norwegian initiative was for instance designed to provide data for research and analysis in public health, and the Dutch strategy mentions the need to conduct a Societal Cost-Benefit Analysis to define if, when and how mandating digital data exchange can be implemented.

Tools and Methods Used to Collect Digital Maturity Data

Most countries assessed a combination of common elements of digital maturity, such as strategy, information technology capability, interoperability, governance and management, patient-centered care, skills and behaviors, and data analytics [11]. Which elements were assessed was often based on the context of the country. Queensland was the only region that used a HIMSS Digital Health Indicator [17].

All countries collected data through standardized electronic self-assessment surveys that were completed by local organizations and supported by national stakeholders. Some had, or were in the process of developing, bespoke electronic data collection and analysis platforms. There was significant variation both within and across countries as to which organizational stakeholders completed the survey.

Likert scales were mostly used to measure indicators, whilst some also had some multiple-choice items, percent scales, and the option for limited free text. The number of included items, however, varied significantly across countries and did not always correlate with the number of health and care settings included, the size of populations, or the number of EHR systems. Norway's national eHealth Monitor included 30 indicators, followed by the Dutch Interoperability Maturity Model with 37 items. Queensland had 121 indicator statements, Germany had 234 items, and Scotland had the largest number with 355 discrete indicators (Scotland also included the widest range of care settings).

Areas of focus also varied significantly depending on how countries divided assessment elements. It was therefore difficult for us to compare where countries placed specific items and where they overlapped. Themes commonly measured across countries included interoperability and information exchange, capabilities and applications, organizational leadership and governance, and technological infrastructures.

Specific areas measured locally in individual countries included person-enabled health and predictive analytics in Queensland; resilience management, clinical, processes, telehealth, and patient participation in Germany; skills and competencies, sustainability, and secondary uses of clinical data in Scotland; laws, regulations, and care processes in the Netherlands; and user satisfaction, and cost-monitoring in Norway.

Most assessments were voluntary but strongly encouraged. None of the countries used penalties. The only compulsory assessment was in Germany, but it was only relevant for hospitals that received national funds to support digital transformation.

Methods Used to Analyze and Disseminate Digital Maturity Assessment Data

Analysis was mostly centrally coordinated and quantitative in nature. We observed a lack of transparency in collected data and digital maturity calculations across countries.

Queensland used proprietary algorithms provided by HIMSS, which allowed a degree of international benchmarking. These were supplemented by some academically led qualitative analyses. Benchmarking by geographical region or organization was done in some countries (Queensland) but not in others (Scotland). Scotland wanted to gain an understanding of the national landscape to better focus strategic funding. They also wanted the submitting organization to develop improvement plans individually and with their linked organizations, across their local, integrated health and care system. There was an explicit agreement that ranking organizations would potentially hinder data collection and analysis. Some countries (the Netherlands and Scotland) also made (or will make) data available for analysis by local organizations themselves.

Outputs were in most cases disseminated directly to participating organizations (through dashboards and meetings) and externally through open webinars and public reports. They were explicitly designed across countries to inform ongoing eHealth strategy. Queensland was the only setting that had a systematic academic evaluation of the digital maturity assessment and associated findings were published in academic journals and conferences, as well as through HIMSS. Scotland and Queensland established an external steering group comprising of subject matter and policy experts.

XSL•FO
RenderX

Challenges Encountered and Lessons Learned

Challenges encountered included limitations in the measurements used, such as uncertainties about the validity of constructs for assessing digital maturity. The analysis and data collection process were time-consuming, and indicators evolved, raising questions about the validity of longitudinal elements. Balancing the practical feasibility of the process with the depth and breadth required by the complexity of the topic presented a trade-off. There were also challenges in making international comparisons due to limited relevance in some local contexts. The self-reported nature of measurements introduced likely biases in assessments. For example, individuals reporting may have a specific agenda driving their responses and more knowledge of parts of organizational activities than others. There may also be a degree of "gaming" where respondents either over- or underestimate the levels of digital maturity in their organization to obtain funding. In addition, an overall view of organizational structures and activities is unlikely to be possible from any individual perspective, and efforts invested in reporting activity are likely to be strongly impacted by the existence of local leadership support for the exercise. In addition, there was limited use of qualitative assessment of indicators, with qualitative follow-ups for context and clarification used only to a limited extent.

Some countries recognized that there is a need to address concerns that data might be used for performance management, which may hamper efforts to obtain accurate data.

Costs of digital maturity assessments were significant (although we did not make cost-benefit assessments), including the required resources to collect, complete, analyze, and disseminate the data. In some countries, additional costs included consultancy support for data platforms, data collection, and analysis (Queensland, Scotland, and the Netherlands).

Discussion

Principal Findings

Although countries varied significantly in population size, existing EHR systems, settings, and indicators; some common themes emerged. These included a national focus on interoperability, a desire to measure changes over time (and a concurrent difficulty keeping track of progress in line with changing circumstances), a need to inform national digital health strategy, priority areas of investment, local benchmarking and strategy development, at least some alignment of common indicators tailored to local needs, a focus on quantitative measurements, and the self-reported voluntary nature of the exercises.

Common thematic areas across countries included interoperability and information exchange, capabilities and applications, organizational leadership and governance, and technological infrastructures. Analysis was centrally coordinated but data were made largely available to participating organizations. Rigorous independent evaluation was only present in 1 instance.

Challenges encountered included the cost of assessment and time-consuming data collection, potential issues surrounding the validity of measurements, issues with reliably measuring changes over time, and potential biases introduced through self-reporting.

Strengths and Limitations

Our work is one of the first attempts to describe and compare how different countries measure digital maturity in hospitals. We included only a limited number of countries, chosen through a self-selected sample, which may limit the potential transferability of findings. Our results are also likely to only apply to relatively advanced health systems with high levels of EHR implementations and UHC.

The data for the study were collected directly by the authors in narrative form and from various sources, which is likely to have biased what was included as it was subjectively selected, and individual viewpoints may have influenced the accounts. This method of data collection also meant that the study provided limited detail and did not delve deeply into specific thematic areas and indicators. We observed a significant variation in approaches and the number of indicators used by different countries, which was difficult to account for in the analysis.

As our focus was on providing an overview of approaches, we did not collect specific information on how the data were analyzed within each country and we therefore did not obtain any insights into factors such as regional variations, or differences between digitally mature and less mature sites. This could be a subject of follow-on work.

Despite these limitations, we offer a broad analysis that provides initial insights into the various methods countries have used to assess digital maturity in hospitals and the lessons they have learned from these experiences. Building on this work, there is now an opportunity to explore the identified areas in more detail, potentially leading to a more comprehensive understanding of digital maturity assessments in health and care settings across various health systems.

Integration of Findings With the Literature

There are many commercially available models for health care providers to assess digital maturity, and there are some general practical considerations that are likely to be relevant across settings [18]. Currently, there are however limited evidence-based methods to assess the quality, use, and efficacy of maturity models and limited guidance on selecting the most appropriate model for any given context [19]. The underlying evidence base for national approaches is even less mature, highlighting the importance of our work. A degree of coordination is likely to be important to streamline efforts and remove duplication, and this will also require consideration surrounding digital infrastructures that can be delivered nationally and those that can remain local [20].

The World Health Organization has recently reviewed tools used to monitor digital health in 8 countries (including Australia and the Netherlands), but these were not specifically explored in relation to digital maturity [9]. Nevertheless, there was a degree of overlap concerning focus. For example, infrastructure and governance as well as specific applications including EHRs, telehealth, and electronic prescribing featured heavily in these

XSL•FO
RenderX

assessments, illustrating the interlinked nature of digital health strategy, systems, infrastructures, and digital maturity. The report also points out a lack of indicators for measuring health inequalities and a tendency to prioritize outcomes over processes [9]. This may partly be due to the difficulty to measure processes in health and care digitalization, and a lack of expertise to analyze such data [21].

Approaches to assessing digital maturity have in the past been criticized for relying too heavily on assessing technological capabilities, overshadowing socio-organizational aspects [22]. This imbalance is reflected in the current analysis but there appears to be an increasing recognition of socio-organizational factors such as leadership and governance and consumer or patient-centered care.

The issue surrounding the evolution of intervention maturity over time has already been recognized [23], but how digital maturity assessments can adequately account for this progression is less clear. Changing circumstances are recognized as a significant challenge in measuring the progress of digitalization in health and care, as the use and exploitation of digital systems is often accompanied by wider changes to organizational functioning, in which existing ways of working are transformed [24]. This in turn impacts the way measurements can be compared over time. The challenge is unlikely to be easily resolved and may require defining a limited set of core measurements that will not be significantly affected over time, whilst being flexible in including new emerging indicators.

Implications for Policy and Practice

There is therefore a trade-off between breadth and depth influencing the scope of assessments. There is also a need to tailor assessments of different settings, as the prevailing “one size fits all” approach whilst providing insights into overall trends, does not help to understand specific organizational circumstances.

The general lack of academic independent validation in digital maturity models and assessments is concerning and raises questions about the reliability and efficacy of existing approaches. There has also been a lack of focus on processes that lead to measured outcomes in assessments. There is further limited evidence on the interrelationships among maturity dimensions meaning that assessments do not provide evidence-based recommendations for improvement.

The role of common indicators across assessments is important. These bring a degree of international comparability, but contextual variations are significant, and digital maturity progress is determined by legal frameworks and financial resources. Scores developed by HIMSS may facilitate a degree of comparability and validation for digital maturity assessments. But standardized international tools may not pay sufficient attention to locally relevant metrics that organizations can use to inform local strategies. Such tools may also present challenges regarding equity and access and there may be commercial conflicts of interest.

Another significant challenge lies in deciding who should be responsible for collecting data, how they represent professional and organizational interests, and how a degree of uniformity across settings may be achieved. This also highlights the necessity of understanding who is best placed to answer specific questions and reiterates the leading organization (Central Government) being explicit in the purpose of the assessment and communicating this repeatedly. A degree of unpredictability regarding how data will eventually be used remains, and so does uncertainty of how individual assessments fit within a wider long-term strategy. Amongst participating organizations, there may be concerns that the data might be used for performance management or resource allocation. Such uses could lead to apprehension and resistance, potentially exacerbating existing inequities. A potential way to address this could include involving organizations more actively in the design of digital maturity assessment tools, perhaps through a process of cocreation. This could then lead to the development of datasets that satisfy both national and local demands, but it is also likely to require compromises on both sides. Policy makers may be faced with different datasets that are not necessarily comparable, but through these will gain insights into local priorities, which can help to facilitate strategic decision making in different, perhaps unanticipated ways.

Going forward, it will be important to assess how these challenges are tackled over time in individual countries, and across the evolving health and care ecosystem including primary and community care. This mapping exercise could also be followed up with more in-depth quantitative or mixed methods work, exploring how digital maturity assessments evolve over time and which aspects are most useful in guiding strategic directions. Incorporating various stakeholder perspectives will be critical, including the voices of health and care providers and patients. We hope that our work provides a steppingstone for achieving this and summarize recommendations for future research to address the challenges identified in Textbox 1.

XSL•FO
RenderX

Textbox 1. Recommendations for future research to address the challenges identified.

Recommendations

Embed real-time independent mixed methods impact, value, and process evaluations in digital maturity assessments from the start to track progress over time and adjust strategy in line with emerging findings. This should focus on:

- Assessing how digital maturity assessments influence decision making locally and nationally.
- Exploring what success means for digital maturity or transformations and develop an internationally agreed framework for measuring associated metrics.
- Assessing which measurements or activities are most impactful, valuable, and cost-effective and why.
- Determining a minimum viable dataset that can be used to make international comparisons.
- Exploring alternatives to self-assessments and clearly distinguish between national and local datasets and their purposes.
- Exploring trajectories and changes over time.
- Exploring incentives and barriers to data collection and analysis amongst various stakeholder groups.

Conclusions

Digital maturity evaluations are driven by a national or regional impetus to identify both areas of excellence and those requiring investment in digitalization. However, the process of assessing digital maturity is fraught with challenges that require transparency in decision-making and monitoring through evaluation.

Existing work is frequently based on self-reporting, is very costly, lacks academic input, and is often conducted by stakeholders with diverse and at times conflicting agendas. Despite these concerns, having some data, however imperfect, is better than none. This is especially true when it comes to making policy decisions that need substantiation.

National digital maturity assessments need to mature from being a vendor or consultancy-funded project-based exercise toward longitudinal evidence-based and academically led activity characterized by international collaboration to facilitate learning.

Acknowledgments

The authors would like to thank Henrik Moeller for reviewing the draft and providing comments.

Authors' Contributions

KC and EA conceived the paper. KC led the data collection, analysis, and write-up with all authors contributing to data collection, data analysis and various iterations of the paper.

Conflicts of Interest

KC, EA, and LS are members of the International Medical Informatics Association Working Group on Technology Assessment and Quality Development and the European Federation for Medical Informatics Working Group on Evaluation.

Multimedia Appendix 1
Detailed country descriptions.
[DOCX File , 52 KB - jmir_v27i1e57858_app1.docx]

References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. Ochoa-Urrego RL, Peña-Reyes JI. Digital maturity models: a systematic literature review. Digitalization: Approaches, Case Studies, and Tools for Strategy, Transformation and Implementation 2021;30:70-85.
3. Westerman G, Bonnet D, McAfee A. Leading digital: Turning technology into business transformation. Massachusetts: Harvard Business Press; 2014.
4. Woods L, Dendere R, Eden R, Grantham B, Krivit J, Pearce A, et al. Perceived impact of digital health maturity on patient experience, population health, health care costs, and provider experience: mixed methods case study. J Med Internet Res 2023;25:e45868. [doi: 10.2196/45868] [Medline: 37463008]

XSL•FO
RenderX

5. Woods L, Eden R, Canfell OJ, Nguyen KH, Comans T, Sullivan C. Show me the money: how do we justify spending health care dollars on digital health? Med J Aust 2023;218(2):53-57 [FREE Full text] [doi: 10.5694/mja2.51799] [Medline: 36502453]
6. WHO (2023) The ongoing journey to commitment and transformation: digital health in the WHO European Region,. Copenhagen: WHO Regional Office for Europe; 2023.
7. Flott K, Callahan R, Darzi A, Mayer E. A patient-centered framework for evaluating digital maturity of health services: a systematic review. J Med Internet Res 2016;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
8. HIMSS 25. HIMSS. URL: https://www.himss.org/ [accessed 2025-02-24]
9. WHO (2022) Monitoring the implementation of digital health: an overview of selected national and international methodologies. Copenhagen: WHO Regional Office for Europe; 2022.
10. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, et al. Reconceptualising the digital maturity of health systems. The Lancet Digital Health 2019;1(5):e200-e201. [doi: 10.1016/s2589-7500(19)30083-4]
11. Woods L, Eden R, Pearce A, Wong YCI, Jayan L, Green D, et al. Evaluating digital health capability at scale using the digital health indicator. Appl Clin Inform 2022;13(5):991-1001 [FREE Full text] [doi: 10.1055/s-0042-1757554] [Medline: 36261114]
12. Phiri P, Cavalini H, Shetty S, Delanerolle G. Digital maturity consulting and strategizing to optimize services: overview. J Med Internet Res 2023;25:e37545 [FREE Full text] [doi: 10.2196/37545] [Medline: 36649060]
13. Mettler T, Pinto R. Evolutionary paths and influencing factors towards digital maturity: An analysis of the status quo in swiss hospitals. Technological Forecasting and Social Change 2018;133:104-117. [doi: 10.1016/j.techfore.2018.03.009]
14. EVAL – Assessment of Health Information Systems. EFMI- European Federation for Medical Informatics. URL: https://efmi.org/workinggroups/eval-assessment-of-health-information-systems/ [accessed 2025-02-24]
15. IMIA-International Medical Informatics Association. URL: https://imia-medinfo.org/wp/technology-assessment-quality-development-in-health-informatics/ [accessed 2025-02-24]
16. Lameire N, Joffe P, Wiedemann M. Healthcare systems--an international review: an overview. Nephrol Dial Transplant 1999;14 Suppl 6(suppl_6):3-9. [doi: 10.1093/ndt/14.suppl_6.3] [Medline: 10528706]
17. The HIMSS digital health indicator. HIMSS. URL: https://www.himsschapter.org/sites/himsschapter/files/ChapterContent/clra/Chapters%20Webinar%20June%202020.pdf [accessed 2025-02-24]
18. Digital health maturity assessments and toolkits. Implementation Know-How Brief. URL: https://openknowledge.worldbank.org/server/api/core/bitstreams/95128705-06f6-43b8-9236-9e7f8abf301d/content [accessed 2024-02-12]
19. Woods LS, Eden R, Duncan R, Kodiyattu Z, Macklin S, Sullivan C. Which one? A suggested approach for evaluating digital health maturity models. Front Digit Health 2022;4:1045685 [FREE Full text] [doi: 10.3389/fdgth.2022.1045685] [Medline: 36506845]
20. Ruseckaite R, Mudunna C, Caruso M, Helwani F, Millis N, Lacaze P, et al. Current state of rare disease registries and databases in Australia: a scoping review. Orphanet J Rare Dis 2023;18(1):216 [FREE Full text] [doi: 10.1186/s13023-023-02823-1] [Medline: 37501152]
21. Cresswell K, Anderson S, Montgomery C, Weir CJ, Atter M, Williams R. Evaluation of digitalisation in healthcare and the quantification of the "Unmeasurable". J Gen Intern Med 2023;38(16):3610-3615 [FREE Full text] [doi: 10.1007/s11606-023-08405-y] [Medline: 37715095]
22. Krasuska M, Williams R, Sheikh A, Franklin BD, Heeney C, Lane W, et al. Technological capabilities to assess digital excellence in hospitals in high performing health care systems: international eDelphi exercise. J Med Internet Res 2020;22(8):e17022 [FREE Full text] [doi: 10.2196/17022] [Medline: 32808938]
23. Monitoring and evaluating digital health interventions. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2025-02-24]
24. Cresswell KM, Bates DW, Williams R, Morrison Z, Slee A, Coleman J, et al. Evaluation of medium-term consequences of implementing commercial computerized physician order entry and clinical decision support prescribing systems in two 'early adopter' hospitals. J Am Med Inform Assoc 2014;21(e2):e194-e202 [FREE Full text] [doi: 10.1136/amiajnl-2013-002252] [Medline: 24431334]

Abbreviations

EHR: electronic health record
HIMSS: Healthcare Information and Management Systems Society
IMIA: International Medical Informatics Association
UHC: Universal Health Coverage

XSL•FO
RenderX

Edited by N Cahill; submitted 28.02.24; peer-reviewed by R Scheible, E Shkarupeta; comments to author 24.04.24; revised version received 03.06.24; accepted 25.11.24; published 06.03.25.

Please cite as:
Cresswell K, Jahn F, Silsand L, Woods L, Postema T, Logan M, Malkic S, Ammenwerth E
Assessing Digital Maturity of Hospitals: Viewpoint Comparing National Approaches in Five Countries
J Med Internet Res 2025;27:e57858
URL: https://www.jmir.org/2025/1/e57858
doi:10.2196/57858
PMID:40053724

©Kathrin Cresswell, Franziska Jahn, Line Silsand, Leanna Woods, Tim Postema, Marion Logan, Sevala Malkic, Elske Ammenwerth. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Integration of Psychiatric Advance Directives Into the Patient-Accessible Electronic Health Record: Exploring the Promise and Limitations

Julian Schwarz[1,2], MD; Eva Meier-Diedrich[1,2], MSc; Matthé Scholten[3], PhD; Lucy Stephenson[4], MRCPsych, PhD; John Torous[5], MBI, MD; Florian Wurster[6,7], MSc; Charlotte Blease[8], PhD

[1]Center for Mental Health, Department of Psychiatry and Psychotherapy, Immanuel Hospital Rüdersdorf, Brandenburg Medical School Theodor Fontane, Rüdersdorf, Germany

[2]Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

[3]Institute for Medical Ethics and History of Medicine, Ruhr University Bochum, Bochum, Germany

[4]South London and Maudsley NHS Foundation Trust, London, United Kingdom

[5]Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[6]Chair of Quality Development and Evaluation in Rehabilitation, Institute of Medical Sociology Health Services Research, and Rehabilitation Science, University of Cologne, Cologne, Germany

[7]Faculty of Human Sciences & Faculty of Medicine, University Hospital Cologne, Cologne, Germany

[8]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Corresponding Author:
Julian Schwarz, MD
Center for Mental Health, Department of Psychiatry and Psychotherapy
Immanuel Hospital Rüdersdorf
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf, 15562
Germany
Phone: 49 33638 83 501
Email: Julian.Schwarz@mhb-fontane.de

Abstract

Psychiatric advance directives (PAD), also known as advance statements or advance choice documents, are legal documents that enable people with mental health conditions to specify their treatment preferences in advance for possible future crises. Subtypes of PADs include crisis cards, joint crisis plans, and self-binding directives (also known as Ulysses contracts). These instruments are intended to improve service user involvement and need orientation in the care of mental crises and to avoid traumatization through unwanted treatment. The existing evidence suggests that people who complete a PAD tend to work more cooperatively with their clinician and experience fewer involuntary hospital admissions. Nevertheless, PADs have not been successfully mainstreamed into care due to multiple barriers to the implementation of PADs, mainly around the completion of PADs and their accessibility and use in crises. The reasons for this include the lack of support in the completion process and acceptance problems, especially on the part of professionals. The research to date primarily recommends support for service users from facilitators, such as peer support workers, and training for all stakeholders. In this article, we argue that while these approaches can help to solve completion and acceptance challenges, they are not sufficient to ensure access to PADs in crises. To ensure accessibility, we propose digital PADs, which offer considerable potential for overcoming these aforementioned barriers. Embedded in national health data infrastructures, PADs could be completed and accessed by service users themselves, possibly with the support of facilitators, and retrieved by any clinic in an emergency. We highlight the strengths and limitations of digital PADs and point out that the proposed solutions must be developed collaboratively and take into account digital inequalities to be effective support for people with serious mental health conditions.

(J Med Internet Res 2025;27:e68549) doi:10.2196/68549

XSL•FO
RenderX

KEYWORDS

advance statements; advance choice documents; advance care planning; mental health; online record access; patient accessible electronic health records; interoperability; fast healthcare interoperability eesources; FHIR; self-binding directives; mobile phone

Introduction

In recent decades, interest in self-determination in health care and service user autonomy has grown considerably, as evidenced by the increasing implementation of interventions to promote shared decision-making and advance care planning in health care [1,2]. In the field of mental health, this is a contested issue in view of perceived potential conflicts between the autonomy and well-being of service users. Although extensive attempts have been made for decades to reduce involuntary interventions in mental health care, this has only been successful to a limited extent [3].

A key intervention in this context is advance care planning using a so-called psychiatric advance directive (PAD), also known as an advance statement or advance choice document. PADs are legal documents that allow individuals to determine in advance how they would like to be treated in the event of a future mental health crisis or psychiatric emergency [4-6]. Subtypes of PADs include crisis cards, joint crisis plans, and self-binding directives (also known as Ulysses contracts or arrangements) [7]. Self-binding PADs are distinctive in that they contain a clause in which service users can request future involuntary hospital admission and treatment during a mental health crisis whilst recognizing that at the time these interventions are required, they are likely to refuse them [8,9]. Despite significant differences between the variants of PADs, they aim to help mental health service users ensure that their treatment wishes are heard and acted upon, particularly in situations where they are unable to express them or lack the threshold capacity to make decisions. To facilitate understanding and reduce complexity, the various forms of advance care planning in mental health care are summarized under the term PAD, unless they are specifically mentioned.

A growing body of scientific papers suggests that PADs can increase the autonomy and perceived control of service users [4,10], strengthen the therapeutic alliance and patient-physician communication [4,6,11], improve adherence [12,13], reduce coercive measures and involuntary admissions [14-17], and could possibly lead to better overall treatment outcomes [12]. In addition, various bodies of the United Nations have stated that the Convention on the Rights of Persons with Disabilities demands the implementation of PADs [18]. Against this backdrop, frameworks to promote PADs for mental health crises have been and are being created in more and more countries. For example, the UK government has enacted legislation for PADs in England and Wales and is currently evaluating its implementation [19]. The Mental Health and Wellbeing Act in Australia and Washington State's SB 5660 in the United States are intended to anchor PADs more firmly in care [4,20,21]. In Germany, PADs are legally binding under the Civil Code and the country's clinical guidelines for the prevention of coercion recommend the implementation of PADs with the highest level of evidence [22-24]. Nevertheless, the available evidence on PADs is considered to be expandable, partly due to the comparatively small sample sizes in some cases and the still small proportion of studies on the topic conducted with service users.

Despite the benefits of PADs for service users, numerous challenges remain in their practical application [14,25]. In this viewpoint paper, we identify outstanding difficulties with PADs, such as the documents not being completed, not being accessible in crises (eg, because people do not carry them with them or they are not accessible in the treating clinic), clinicians ignoring PADs, PADs not being kept up to date, or the existence of conflicting versions of a PAD (eg, because paper and pencil updates were not distributed among all parties involved) [26]. To address these problems, recent reviews suggest improving PAD implementation strategies through measures such as training for professionals, service users, and their relatives [8,14]. We consider these measures useful but argue that their impact may be limited if PADs are not accessible to all stakeholders at all times in an up-to-date form. As a solution, we propose digital PADs embedded in electronic health records and show that they may provide longer-term tangible outcomes and increase the usability, practicality, and relevance of PADs, which in turn would have a positive impact on their utility and clinical impact. We explore the strengths and limitations of digital PADs and emphasize that for the innovation to work, it is essential that the PADs are interoperable, ideally integrated into each country's health care data infrastructure via Fast Healthcare Interoperability Resources. Interoperable means in this context that the content of the PAD is stored in a digital standardized form, namely via Fast Healthcare Interoperability Resources, which permits, for example, smooth access by the different medical information systems of various health care institutions.

Current Challenges With PADs in Crisis Care

Many studies have explored the benefits and challenges of implementing PADs [4,25,27-29]. We will focus on these challenges below, as these are the key starting points for the proposed solution.

Problem With Drafting PADs

In a review of service users' perspectives on PADs, Braun et al [14] found that in 73% (30/41) of the studies, completing a PAD was a challenge for those involved. Service users reported a lack of knowledge and understanding of how PADs work [5,30-32]: they often do not know what to write in a PAD, and the templates provided are often without sufficient explanations and examples, making it challenging to complete them [6,33]. In addition, some service users find completing a PAD too complicated or time-consuming [34-36]. Doubts about one's own decision-making ability at the time of preparation also discourage service users [6,27,31,37]. Finally, the creation

XSL•FO
RenderX

process can be emotionally distressing, especially if it evokes memories of previous involuntary treatment [11,36,38,39].

Another difficulty in creating a PAD is the lack of support. For example, many service users do not have a trusted person to support them [11,36,38]. Some service users are also concerned that potential facilitators, such as the treating psychiatrist or relatives, could exert undue influence on the contents of the PAD by introducing their own potentially divergent preferences [35,39]. In addition, clinicians usually have little or no financial or time resources to draft a PAD together with the service users or even to advise them [40,41]. Professionals are reluctant to invest time in creating PADs [42]. It is, therefore, not surprising that service users report unclear or inadequate instructions and insufficient follow-up by professionals in the event of any queries [30,43].

Overall, these findings illustrate that service users need clear guidance and support to create an actionable PAD that meets their needs. Professionals should be given the necessary time to support the completion.

Problems With Accessing PADs

A key problem is that PADs are often not stored so that they are up-to-date and accessible to all relevant stakeholders during a crisis [14,27,37,41,44]. Reasons for this are that those primarily involved in the use of a PAD, such as staff in the emergency department, inpatient wards of a hospital, judges, guardians ad litem, and legal guardians often do not know whether a PAD is present or not [25,28,45], that service users cannot provide information about the presence of a PAD in the event of a crisis, that professionals are not made aware of the presence of a PAD through a corresponding notification in the hospital information system, that the PAD is not available in digitized form, or that possible standard operating procedures for crisis and de-escalation management in a hospital are either not available or do not take PADs into account. Various studies show that these considerable access problems lead to uncertainty and irritation among service users, as an exemplary statement of a service user from a qualitative focus group study shows:

> *I wasn't at the stage of going into hospital, but I'd been on a high and they were a little bit worried. And what concerns me was that they didn't have a copy of your crisis plan [26].*

The barriers to access are even higher when service users present at a hospital where they have not been treated previously [25]. In such situations, it is unknown what kind of care and medication was helpful in previous crises, nor can previous doctors' letters or PADs be accessed unless the person concerned brings these with them themselves. Particularly in the case of involuntary hospital admission or treatment or use of coercive measures, a lack of knowledge about individualized crisis treatment can lead to inadequate treatment, harm, and traumatization [46,47].

Problems With Applying PADs

Even if it is ensured that a PAD is accessible, this does not mean that it is actually taken into account in crisis management. For example, about 90% (35/41) of the studies identified in the review by Braun et al [14] that examined the service user perspective on PADs found evidence of application problems. An important factor here is that PADs are not followed in practice by professionals (approximately 61% of studies; 25/41) for various reasons. Service users reported that PADs were not endorsed by professionals in principle (approximately 17% of studies; 7/41) or PADs are not legally binding and hence could only be applied to a limited extent (approximately 22% of studies; 9/41). Another reason for not using PADs is the fear among professionals and some service users that the contents of PADs may be outdated because they are not regularly updated and therefore no longer reflect the current preferences of service users or current treatment guidelines (approximately 12% of studies; 5/41). In this context, community stakeholders expressed concern that service users would not take responsibility for keeping their PAD up to date over time [48]. Conversely, service users experienced that PADs were only initiated by professionals but not followed up, let alone integrated into outpatient treatment in the longer term, for example, through an accompanying regular exchange about the PAD [38].

In addition, the problem remains that not all contingencies of a possible future mental health crisis can be covered by a PAD [8,37,41]. Accordingly, stakeholders primarily criticize the lack of flexibility of PADs and would like them to be formulated as concretely and yet as widely applicable as possible so that they can be applied to a wide variety of mental health crisis scenarios [37]. This challenge is particularly relevant for self-binding PADs, as they can also be enforced over the dissent of a service user who lacks decision-making ability and against potentially divergent treatment recommendations from clinicians [8].

Finally, a lack of knowledge and inadequate training of clinicians in the use of PADs in emergency situations hinders their application. It is often unclear when a PAD is read by which member of the mental health team, how it is implemented, and who takes responsibility for this [49].

Currently Proposed Solutions

A number of measures are described in the literature, all of which aim to improve the implementation of PADs.

The most promising of these is the use of human support, or "facilitators," to guide the preparation of a PAD, which can significantly improve completion rates [50,51]. The role of facilitator can be assumed by professionals, peer support workers, patient advocates, or legal representatives. A recent RCT study from France shows that peer support workers-facilitated PADs can reduce the number of compulsory admissions compared to conventional PADs [52]. This might be explained by the fact that those affected can open up about this sensitive topic more confidently with a person with personal experience of mental distress or even involuntary treatment [53] than possibly their relatives or even practitioners, who may have different interests in crisis treatment than the affected person themselves. On the other hand, the use of facilitators could make the intervention more complex and time-consuming, as a considerable amount of additional training and personnel is required.

XSL•FO
RenderX

Alternatively, the literature recommends offering financial incentives, particularly for clinicians, to compensate for the additional time spent on supporting service users to complete PADs [22], or providing user-friendly templates and clear instructions for completing them—both of which have seen limited implementation in practice so far [41,54]. Furthermore, research suggests specific groups that could be targeted to increase the likelihood of PAD completion, such as those with strong social integration, those who have had negative treatment experiences in the past, and those who have recently experienced an acute mental health crisis [14]. It also suggests there are specific groups who might find it more challenging to complete PADs but are more likely to benefit from them. For example, service users from Black communities have identified additional barriers to completion, but research suggests additional benefits if service users are supported to overcome these hurdles [55].

Finally, various education and training interventions for service users and clinicians are recommended to offer more knowledge about PADs and the associated legal and ethical aspects to reduce misunderstandings and fears and increase uptake [56-58]. Training should also help to positively influence the aforementioned barriers to access, especially those aimed explicitly at ER personnel or rather detailing the use of PADs in emergency situations [25,49]. At the same time, it is important to emphasize that the existing evidence suggests that these measures are only effective to a limited extent and that administrative staff often have more knowledge about PADs than service users themselves [49,59]. Shields et al [25] therefore recommend integrating training into both undergraduate education and continuing professional development.

Overall, the currently proposed solutions can potentially promote the completion of PADs and, in principle, reduce involuntary interventions. Nevertheless, the problem remains that neither PADs nor facilitation is widely implemented in practice [15,16,51]. In addition, many trivial access problems remain unresolved, such as ensuring that the relevant stakeholders always have access to the same, up-to-date PAD version.

Integration of PADs Into the Electronic Health Record

The idea of making the PAD accessible in digital form is not new, and the following is a brief overview of efforts in this regard in various countries. As early as 2001, a US study concluded [60] that "without a computerized system for storing and retrieving service users' PAD information, preparing a PAD may do little more than the act of scrawling "help" on a scrap of paper, stuffing it into a bottle, and hurling it into the ocean." Nevertheless, there are still only a few successful implementations of digital PADs in psychiatry. The situation is different in physical medicine, especially in palliative care, where there are already several digital applications for advanced care planning. In the United Kingdom, for example, "coordinate my care" has been implemented, which enables digital creation and access to advanced choice documents. A recent evaluation of this service shows that it is also used by mental health service users and their clinicians specifically for advance care planning of mental health conditions [61], presumably also due to the lack of such tools in the mental health field.

In the field of psychiatry, Murray and Wortzel [4] report that some states in the United States are already using so-called "online advance directive registries" in both medical and psychiatric care. Comprehensive descriptions and scientific evaluations of these approaches are not yet available. In addition, the so-called "My Mental Health Crisis App" was published in 2020, by the SMI Adviser initiative, a project funded by the US Substance Abuse and Mental Health Services Administration. It enables those affected to create a PAD using step-by-step instructions and then share it digitally [62]. The app even considers the different legal regulations in the US states and the resulting differences in practice when creating PADs, such as the mandatory presence of witnesses or notarization. Finally, the PAD created using the app can be shared via a QR code. While this enables barrier-free access, it risks violating privacy if the QR code falls into the hands of unauthorized persons. We are not aware of any scientific evaluations of this offer either.

A similar but purely web-based solution is currently being piloted in Germany as part of a study [63,64], where the PAD is embedded in a digital environment regularly used by service users to test online record access. The document can only be accessed by persons who have been previously authorized by the service user, such as outpatient clinicians (general practitioners, psychiatrists, psychotherapists), legal guardians, and relatives, using their own account. In addition, staff of the mental health hospital in the respective catchment area in which the service user lives have access to the PAD. In routine care in Germany, there is currently only one type of emergency dataset stored on the physical health insurance card of those with statutory health insurance. This card can theoretically be read by any ambulance service, although it is unclear to what extent this function is currently being used.

The only known digital and interoperable implementation of PADs in the field of mental health is currently carried out in the United Kingdom, albeit only on an experimental basis, jointly by the NHS in coproduction with the startup Thalamos [26]. The initiative was prompted by recently introduced legislation that aims to address the increased number of involuntary psychiatric treatments in the United Kingdom in recent years, while also addressing significant failures to address racial inequalities in mental health care [65]. This is supported by findings that the significantly higher proportion of "Black Indigenous people of color" admitted under the law appears to benefit particularly from PADs [66,67]. Currently, a PAD culturally adapted for people from sub-Saharan Africa and the Caribbean is being piloted and stored in electronic patient records that are available to connected service providers such as acute hospital trusts and general practitioners [44,68].

Strengths of Digital Interoperable PADs

We argue that the approaches to digitizing PADs presented so far have considerable potential to improve the care situation of those affected, especially when combined. This section describes the strengths of digital PADs, based on the vision of a comprehensive digital solution from the perspective of medical ethicists and clinical researchers, and explains how digital PADs can help to overcome the key challenges of paper-based PADs.

XSL•FO
RenderX

First, every service user diagnosed with a serious mental health condition, and especially those with a history of involuntary hospital admission and treatment, should be informed about the possibility of completing a digital PAD. This could be done in a variety of ways, both through the care system and through user-led services such as self-help groups. Research has shown that PADs are better accepted when the completion process is supported [4,25]. In addition to human support, a digital, AI-based assistant integrated into the PAD form interface could guide users through the completion process, as outlined by Redahan and Kelly [69]. It could provide explanations and example wordings and ensure that the document is completed fully and correctly, meaning it makes sense in terms of content. This would address a key problem, namely that service users often do not manage to complete the PAD without support [6,39,70]. One of the reasons for the limited uptake by service users may be that research into the implementation of PADs has been conducted primarily by clinical researchers and medical ethicists, including this article. To address the priorities and concerns of service users, and to develop a viable (digital) PAD that service users find helpful, it is critical to have a design process that systematically involves service users from the outset [71,72].

One of the strengths of digital PADs is that service users control PAD access. This means that they could determine and control which people, such as outpatient practitioners (general practitioners, psychiatrists, and psychologists), other people (friends, relatives, etc), or people from other institutions involved in involuntary treatment and hospitalization (judges, guardians ad litem, community mental health investigators, legal guardians), would have read or write access to their digital PAD. One exception could be psychiatric hospitals, whose staff should be able to access the digital PAD when the respective service user is admitted or when they present to the emergency room of a hospital, whereby each read access would be logged by name. This would be necessary to protect the highly sensitive personal data contained in the PAD and to ensure that PADs are only accessed by treatment providers in a psychiatric emergency. This in turn would require a clear (technical) role concept for all actors involved, defining who is allowed to see and process what [26].

A major challenge in the implementation of conventional PADs is the lack of time or the reluctance of clinicians to invest the necessary time [42]. If service users could prepare a digital PAD at their own pace with the help of the digital assistant as outlined above or with the involvement of a trusted person, the attending psychiatrist would only need to check the digital PAD for feasibility and make any necessary suggestions for adjustments during the consultation. This would significantly reduce the time required by clinicians.

One of the main problems with analog PADs was that different versions were stored in different places, which could lead to the PAD accessed in the acute situation not being up to date. Digital access enables the service users to keep the digital PAD up to date when preferences for emergency treatment change. At the same time, a notification system could be used to notify all stakeholders with access to the PAD of changes, for example, by email, making collaboration much easier.

Digital PADs could likely play a significant role in promoting service user involvement in decision-making in mental health care. Ideally, if digital PADs were official, standardized documents integrated into the national health data infrastructure of each country, they would be difficult for institutions or professionals to overlook. This approach would also help tackle the prevalent issue of staff's lack of engagement with PADs [29].

Limitations of Digital Interoperable PADs

Although the digitization of PADs has the potential to overcome many practical barriers, it also brings new challenges, particularly in terms of implementation, acceptance, and engagement.

The existing evidence suggests that people with serious mental health conditions tend to have less access to digital health services or use them to a limited extent [71-73]. In addition, they are less likely to have access to a (desktop) computer, the internet, or a modern smartphone [74,75], and more likely to have concerns about data security and to lack the necessary digital skills [76,77] to keep a digital PAD up to date. As such, digital PADs risk increasing existing inequalities unless these access barriers are specifically addressed [26]. The solution used should therefore be flexible enough to meet the needs of the heterogeneous target group of people with serious mental health conditions and to avoid digital exclusion. To increase the acceptance and reach of the intervention, a combination of high- and low-tech approaches may be useful. For example, the facilitator could store a service user's treatment preferences in the digital PAD on the service user's behalf and provide them with a printout. At the same time, it should also be possible to fill in only selected parts of the digital PAD, such as only a free text field on treatment preferences, as it may be that a highly structured template could be overly prescriptive for some service users and may not adequately reflect how they wish to convey their PAD in terms of content and format. Furthermore, those service users who require additional digital skills to set up a digital PAD, or who have privacy concerns, should receive appropriate training. The latter should also include professionals, as they often do not have the necessary skills either or do not know how to best integrate digital tools into treatment [78].

Another potential limitation concerns the support provided during the process of completing digital PADs. If the completion of a PAD is supported by chatbots or human, nonphysician facilitators, at least some of the exchange about how to deal with emergency situations is outsourced from the patient-physician relationship. This could lead to alienation between patients and practitioners, especially after involuntary treatment or coercion. One way to counteract this would be to use debriefings of involuntary measures to also discuss the PADs and to validate the extent to which the instructions contained in them were actually implemented. This review of the (non) application of the PAD, which could also be done online, should be attended by all key persons involved in the previous involuntary action, or at least the inpatient and outpatient treatment providers in addition to the service user concerned.

XSL•FO
RenderX

In addition, there are a number of barriers to paper and pencil PADs that are likely to apply to digital PADs as well. This concerns the fear, more pronounced among psychiatrists in particular, that there may be a "lack of quality information" in PADs or that "consumers desire to change their mind during crisis" [28]. Furthermore, there are concerns that "inappropriate treatment requests" could be included in the PADs, that these requests could entail risks, that there is a "lack of time to review the document," and that additional documentation is required. These limitations can probably only be reduced by providing more training and information to all parties involved, regardless of whether the intervention is digital or analog. In general, in addition to training measures, it is essential to promote the dissemination of (digital) PADs, be it through web-based multimedia resources, such as those realized in English-speaking countries, or through appropriate educational initiatives enshrined in law, in order to improve knowledge and awareness of PADs among the population, which includes both potential service users and professionals [79].

Another problem that is not solved by digitalization is that in various jurisdictions in some countries, such as the United States and Australia, PADs are not legally binding, and medical orders (eg for measures involving deprivation of liberty, issued by the acting physician during a psychiatric emergency) overrule PADs [25]. Conversely, where self-binding directives apply, it is often not regulated which PAD contents are legally binding and which are not and can therefore be overruled by professionals [8,37]. Consequently, digitalization should be accompanied by a revision of the legislation on the application of PADs and lead to more clarity about revocability and enforceability. In this context, for example, it would be important to clarify how to proceed if service users revoke the digital PAD on their cell phone in the context of an acute mental health crisis.

Another limitation, related to the existing acceptance issues, is that digitalization could render an already complex intervention "PAD" even more complex. In addition, problems related to internet connection and technology, system interoperability, data security, and data protection must be overcome [80]. For the innovation to succeed, an overarching implementation strategy is needed that permeates the macro, meso, and micro levels and requires the will as well as the necessary IT resources and infrastructure.

While digital PADs could provide greater convenience and ease in making advanced health care decisions, there is a potential risk that people with serious mental health conditions might update their directives during acute illness episodes, which could lead to problematic or unintended decisions. In addition, due to the digital nature of these directives, there may be privacy risks and concerns about unauthorized access. This could be the case if a service user shares the access or QR code with unknown people or in public places such as social networks, which could harm the service user. How these and related risks might be managed will require greater forethought and provisions to guide and protect service users.

Future Research Directions

The introduction of digital PADs offers promising opportunities as well as new challenges that need to be further explored. Future research should focus on a participatory approach to the digitalization of PADs to ensure that the expectations and needs of all stakeholders, particularly service users, are adequately addressed to ensure the needs-based and practical nature of the solution to be developed. We argue that before measuring "hard" outcomes, such as the reduction of involuntary interventions and inpatient treatment days, which should of course be a longer-term goal, efforts must first be made that aim at a careful development and implementation of digital PADs.

First, the requirements and barriers to the implementation of digital PADs should be thoroughly researched. This must include developing strategies to reduce digital exclusion to ensure that the main target group of the intervention, people with serious mental health conditions, are not left behind. Furthermore, for a future-oriented embedding of the PAD in the health data infrastructure, transnational initiatives such as the European Health Data Space should also be considered. Finally, the legal and ethical implications of PADs must be investigated, especially in countries where PADs are not legally binding, to ensure that these digital tools improve, and do not hinder, the autonomy of the service user and the quality of care. Only when satisfactory acceptance by stakeholders can be demonstrated, as evidenced by a fully completed digital PAD and its accessibility in emergency situations, does it seem reasonable to investigate the abovementioned hard outcomes, for example, in the form of a randomized controlled trial. In this context, a comparison with the paper-and-pencil version of the PAD also seems important to demonstrate a possible differentiation and superiority of digital PADs.

Conclusions

The integration of PADs into electronic health records represents an important step toward equality in mental and physical health care. At the same time, it is in line with the trend toward increasing digitalization of involuntary psychiatric admission processes [81,82], which holds significant potential in terms of transparency, speed, and equity. Digital PADs have the potential to fundamentally change psychiatric crisis care by addressing long-standing challenges of printed PADs such as accessibility, outdated information, and limited stakeholder engagement. Their success will depend on the careful development of implementation strategies, training of professionals, and the inclusion of stakeholders in the design and evaluation process. Although digital PADs can overcome many of the barriers associated with traditional PADs, they also introduce new complexities that must be addressed through ongoing research.

XSL•FO
RenderX

Acknowledgments

MS thanks the German Federal Ministry of Education and Research for funding received in the context of the SALUS project (grant number 01GP1792).

Conflicts of Interest

JT is the editor-in-chief of *JMIR Mental Health* and CB is an associate editor at *JMIR Mental Health*.

References

1. Rosa WE, Izumi S, Sullivan DR, Lakin J, Rosenberg AR, Creutzfeldt CJ, et al. Advance care planning in serious illness: a narrative review. J Pain Symptom Manage 2023;65(1):e63-e78 [FREE Full text] [doi: 10.1016/j.jpainsymman.2022.08.012] [Medline: 36028176]
2. van der Horst DEM, Garvelink MM, Bos WJW, Stiggelbout AM, Pieterse AH. For which decisions is Shared Decision Making considered appropriate? - A systematic review. Patient Educ Couns 2023;106:3-16 [FREE Full text] [doi: 10.1016/j.pec.2022.09.015] [Medline: 36220675]
3. Aragonés-Calleja M, Sánchez-Martínez V. Evidence synthesis on coercion in mental health: an umbrella review. Int J Ment Health Nurs 2024;33(2):259-280. [doi: 10.1111/inm.13248] [Medline: 37908175]
4. Murray H, Wortzel HS. Psychiatric advance directives: origins, benefits, challenges, and future directions. J Psychiatr Pract 2019;25(4):303-307. [doi: 10.1097/PRA.0000000000000401] [Medline: 31291211]
5. Gaillard AS, Braun E, Vollmann J, Gather J, Scholten M. The content of psychiatric advance directives: a systematic review. Psychiatr Serv 2023;74(1):44-55. [doi: 10.1176/appi.ps.202200002] [Medline: 36039553]
6. Lequin P, Ferrari P, Suter C, Milovan M, Besse C, Silva B, et al. The joint crisis plan: a powerful tool to promote mental health. Front Psychiatry 2021;12:621436 [FREE Full text] [doi: 10.3389/fpsyt.2021.621436] [Medline: 33815165]
7. Henderson C, Swanson JW, Szmukler G, Thornicroft G, Zinkler M. A typology of advance statements in mental health care. Psychiatr Serv 2008;59(1):63-71. [doi: 10.1176/ps.2008.59.1.63] [Medline: 18182541]
8. Stephenson L, Gieselmann A, Gergel T, Owen G, Gather J, Scholten M. Self-binding directives in psychiatric practice: a systematic review of reasons. Lancet Psychiatry 2023;10(11):887-895. [doi: 10.1016/S2215-0366(23)00221-3] [Medline: 37714174]
9. Scholten M, van Melle L, Widdershoven G. Self-binding directives under the new Dutch law on compulsory mental health care: an analysis of the legal framework and a proposal for reform. Int J Law Psychiatry 2021;76:101699. [doi: 10.1016/j.ijlp.2021.101699] [Medline: 33864988]
10. Henderson C, Flood C, Leese M, Thornicroft G, Sutherby K, Szmukler G. Views of service users and providers on joint crisis plans: single blind randomized controlled trial. Soc Psychiatry Psychiatr Epidemiol 2009;44(5):369-376. [doi: 10.1007/s00127-008-0442-x] [Medline: 18836881]
11. Morrissey FE. The introduction of a legal framework for advance directives in the UN CRPD era: the views of Irish service users and consultant psychiatrists. Ethics Med Public Health 2015;1(3):325-338 [FREE Full text] [doi: 10.1016/j.jemep.2015.07.007]
12. Swanson JW, Tepper MC, Backlar P, Swartz MS. Psychiatric advance directives: an alternative to coercive treatment? Psychiatry 2000;63(2):160-172. [doi: 10.1080/00332747.2000.11024908] [Medline: 10965546]
13. Shields LS, Pathare S, van Zelst SD, Dijkkamp S, Narasimhan L, Bunders JG. Unpacking the psychiatric advance directive in low-resource settings: an exploratory qualitative study in Tamil Nadu, India. Int J Ment Health Syst 2013;7(1):29 [FREE Full text] [doi: 10.1186/1752-4458-7-29] [Medline: 24369909]
14. Braun E, Gaillard AS, Vollmann J, Gather J, Scholten M. Mental health service users' perspectives on psychiatric advance directives: a systematic review. Psychiatr Serv 2023;74(4):381-392. [doi: 10.1176/appi.ps.202200003] [Medline: 36128696]
15. de Jong MH, Kamperman AM, Oorschot M, Priebe S, Bramer W, van de Sande R, et al. Interventions to reduce compulsory psychiatric admissions: a systematic review and meta-analysis. JAMA Psychiatry 2016;73(7):657-664. [doi: 10.1001/jamapsychiatry.2016.0501] [Medline: 27249180]
16. Molyneaux E, Turner A, Candy B, Landau S, Johnson S, Lloyd-Evans B. Crisis-planning interventions for people with psychotic illness or bipolar disorder: systematic review and meta-analyses. BJPsych Open 2019;5(4):e53 [FREE Full text] [doi: 10.1192/bjo.2019.28] [Medline: 31530302]
17. Swanson JW, Swartz MS, Elbogen EB, VAN Dorn RA, Wagner HR, Moser LA, et al. Psychiatric advance directives and reduction of coercive crisis interventions. J Ment Health 2008;17(3):255-267 [FREE Full text] [doi: 10.1080/09638230802052195] [Medline: 20221301]
18. Scholten M, Gieselmann A, Gather J, Vollmann J. Psychiatric advance directives under the convention on the rights of persons with disabilities: why advance instructions should be able to override current preferences. Front Psychiatry 2019;10:631 [FREE Full text] [doi: 10.3389/fpsyt.2019.00631] [Medline: 31572233]
19. Modernising the mental health act: increasing choice, reducing compulsion: final report of the independent review of the mental health act 1983. UK Department of Health and Social Care.: Crown; 2018. URL: https://www.gov.uk/government/publications/modernising-the-mental-health-act-final-report-from-the-independent-review [accessed 2025-02-13]

XSL•FO
RenderX

20. Edan V, Maylea C. A model for mental health advance directives in the new Victorian. Mental Health and Wellbeing Act. Psychiatr Psychol Law 2022;29(5):779-787 [FREE Full text] [doi: 10.1080/13218719.2021.1976302] [Medline: 36148391]
21. SB 5660 - 2023-24. Washington State Legislature. 2024. URL: https://app.leg.wa.gov/billsummary?BillNumber=5660&Year=2023&Initiative=false [accessed 2024-09-13]
22. Steinert T, Hirsch S. [German S3 guidelines on avoidance of coercion: prevention and therapy of aggressive behavior in adults]. Nervenarzt 2020;91(7):611-616. [doi: 10.1007/s00115-019-00801-2] [Medline: 31473766]
23. Gather J, Henking T, Juckel G, Vollmann J. Advance refusals of treatment in situations of danger to self or to others – ethical and legal considerations on the implementation of advance directives in psychiatry. In: Beneficial Coercion in Psychiatry?. Leiden: Brill; 2017:245-262.
24. Henking T, Scholten M. Respect for the will and preferences of people with mental disorders in German law. In: Capacity, Participation and Values in Comparative Legal Perspective. Bristol, UK: Bristol University Press; 2023:203-225.
25. Shields LS, Pathare S, van der Ham AJ, Bunders J. A review of barriers to using psychiatric advance directives in clinical practice. Adm Policy Ment Health 2014;41(6):753-766. [doi: 10.1007/s10488-013-0523-3] [Medline: 24248818]
26. Digital advance choice documents deploying in a clinical setting. Thalamos and King's College London. England: King's College, London URL: https://kclpure.kcl.ac.uk/portal/files/239348291/Thalamos_KCL_Digital_ACD_Discovery_Document.pdf [accessed 2024-07-16]
27. Farrelly S, Lester H, Rose D, Birchwood M, Marshall M, Waheed W, et al. Barriers to shared decision making in mental health care: qualitative study of the Joint Crisis Plan for psychosis. Health Expect 2016;19(2):448-458 [FREE Full text] [doi: 10.1111/hex.12368] [Medline: 25912086]
28. Van Dorn RA, Swartz MS, Elbogen EB, Swanson JW, Kim M, Ferron J, et al. Clinicians' attitudes regarding barriers to the implementation of psychiatric advance directives. Adm Policy Ment Health 2006;33(4):449-460. [doi: 10.1007/s10488-005-0017-z] [Medline: 16237505]
29. Henderson C, Farrelly S, Moran P, Borschmann R, Thornicroft G, Birchwood M, Joshua, Study Groups. Joint crisis planning in mental health care: the challenge of implementation in randomized trials and in routine care. World Psychiatry 2015;14(3):281-283 [FREE Full text] [doi: 10.1002/wps.20256] [Medline: 26407774]
30. Valentine L, Grace D, Pryor I, Buccilli K, Sellars M, Francey S, et al. "When I'm Thinking Straight, I Can Put Things in Place for When I'm Not." | exploring the use of advance statements in first-episode psychosis treatment: young people, clinician, and carer perspectives. Community Ment Health J 2021;57(1):18-28. [doi: 10.1007/s10597-020-00721-5] [Medline: 33058044]
31. Hotzy F, Cattapan K, Orosz A, Dietrich B, Steinegger B, Jaeger M, et al. Psychiatric advance directives in Switzerland: knowledge and attitudes in patients compared to professionals and usage in clinical practice. Int J Law Psychiatry 2020;68:101514. [doi: 10.1016/j.ijlp.2019.101514] [Medline: 32033691]
32. Van Dorn RA, Swanson JW, Swartz MS. Preferences for psychiatric advance directives among Latinos: views on advance care planning for mental health. Psychiatr Serv 2009;60(10):1383-1385 [FREE Full text] [doi: 10.1176/ps.2009.60.10.1383] [Medline: 19797381]
33. Radenbach K, Falkai P, Weber-Reich T, Simon A. Joint crisis plans and psychiatric advance directives in German psychiatric practice. J Med Ethics 2014;40(5):343-345. [doi: 10.1136/medethics-2012-101038] [Medline: 23793058]
34. Scheyett AM, Rooks A. University students' views on the utility of psychiatric advance directives. J Am Coll Health 2012;60(1):90-93. [doi: 10.1080/07448481.2011.572326] [Medline: 22171734]
35. Easter MM, Swanson JW, Robertson AG, Moser LL, Swartz MS. Impact of psychiatric advance directive facilitation on mental health consumers: empowerment, treatment attitudes and the role of peer support specialists. J Ment Health 2021;30(5):585-593. [doi: 10.1080/09638237.2020.1714008] [Medline: 32013647]
36. Hindley G, Stephenson LA, Ruck Keene A, Rifkin L, Gergel T, Owen G. "Why have I not been told about this?": a survey of experiences of and attitudes to advance decision-making amongst people with bipolar. Wellcome Open Res 2019;4:16 [FREE Full text] [doi: 10.12688/wellcomeopenres.14989.2] [Medline: 31080892]
37. Gergel T, Das P, Owen G, Stephenson L, Rifkin L, Hindley G, et al. Reasons for endorsing or rejecting self-binding directives in bipolar disorder: a qualitative study of survey responses from UK service users. Lancet Psychiatry 2021;8(7):599-609 [FREE Full text] [doi: 10.1016/S2215-0366(21)00115-2] [Medline: 34022972]
38. Morriss R, Mudigonda M, Bartlett P, Chopra A, Jones S. National survey and analysis of barriers to the utilisation of the 2005 mental capacity act by people with bipolar disorder in England and Wales. J Ment Health 2020;29(2):131-138. [doi: 10.1080/09638237.2017.1340613] [Medline: 28644738]
39. Elbogen EB, Swanson JW, Appelbaum PS, Swartz MS, Ferron J, Van Dorn RA, et al. Competence to complete psychiatric advance directives: effects of facilitated decision making. Law Hum Behav 2007;31(3):275-289 [FREE Full text] [doi: 10.1007/s10979-006-9064-6] [Medline: 17294136]
40. Sarin A, Murthy P, Chatterjee S. Psychiatric advance directives: potential challenges in India. Indian J Med Ethics 2012;9(2):104-107 [FREE Full text] [doi: 10.20529/IJME.2012.031] [Medline: 22591869]
41. Stephenson LA, Gergel T, Ruck Keene A, Rifkin L, Owen G. The PACT advance decision-making template: preparing for mental health act reforms with co-production, focus groups and consultation. Int J Law Psychiatry 2020;71:101563 [FREE Full text] [doi: 10.1016/j.ijlp.2020.101563] [Medline: 32768120]

XSL•FO
RenderX

42. Srebnik D, Brodoff L. Implementing psychiatric advance directives: service provider issues and answers. J Behav Health Serv Res 2003;30(3):253-268. [doi: 10.1007/BF02287316] [Medline: 12875095]
43. Amering M, Stastny P, Hopper K. Psychiatric advance directives: qualitative study of informed deliberations by mental health service users. Br J Psychiatry 2005;186:247-252. [doi: 10.1192/bjp.186.3.247] [Medline: 15738506]
44. Digital advance choice documents deploying in a clinical setting. Thalamos and King's College London. 2023. URL: https://www.mentalcapacitylawandpolicy.org.uk/wp-content/uploads/2023/09/Thalamos_KCL_Digital-ACD_Discovery_Document.pdf [accessed 2024-07-16]
45. Kim MM, Scheyett AM, Elbogen EB, Van Dorn RA, McDaniel LA, Swartz MS, et al. Front line workers' attitudes towards psychiatric advance directives. Community Ment Health J 2008;44(1):28-46. [doi: 10.1007/s10597-007-9104-9] [Medline: 17721821]
46. Sareen A, Mitra S, Wadhwa A, Soeung C, Korenis P. Trauma from involuntary hospitalization and impact on mental illness management. Prim Care Companion CNS Disord 2022;24(4):21cr03122 [FREE Full text] [doi: 10.4088/PCC.21cr03122] [Medline: 35960890]
47. Seed T, Fox JRE, Berry K. The experience of involuntary detention in acute psychiatric care. A review and synthesis of qualitative studies. Int J Nurs Stud 2016;61:82-94. [doi: 10.1016/j.ijnurstu.2016.05.014] [Medline: 27314181]
48. O'Connell MJ, Stein CH. Psychiatric advance directives: perspectives of community stakeholders. Adm Policy Ment Health 2005;32(3):241-265. [doi: 10.1007/s10488-004-0843-4] [Medline: 15844847]
49. Henderson C, Jackson C, Slade M, Young AS, Strauss JL. How should we implement psychiatric advance directives? Views of consumers, caregivers, mental health providers and researchers. Adm Policy Ment Health 2010;37(6):447-458. [doi: 10.1007/s10488-010-0264-5] [Medline: 20099077]
50. Stephenson LA, Gergel T, Ruck Keene A, Rifkin L, Owen G. Preparing for mental health act reform: pilot study of co-produced implementation strategies for advance choice documents. Wellcome Open Res 2022;7:182 [FREE Full text] [doi: 10.12688/wellcomeopenres.17947.1] [Medline: 36081756]
51. Swanson JW, Swartz MS, Elbogen EB, Van Dorn RA, Ferron J, Wagner HR, et al. Facilitated psychiatric advance directives: a randomized trial of an intervention to foster advance treatment planning among persons with severe mental illness. Am J Psychiatry 2006;163(11):1943-1951 [FREE Full text] [doi: 10.1176/ajp.2006.163.11.1943] [Medline: 17074946]
52. Tinland A, Loubière S, Mougeot F, Jouet E, Pontier M, Baumstarck K, DAiP Group. Effect of psychiatric advance directives facilitated by peer workers on compulsory admission among people with mental illness: a randomized clinical trial. JAMA Psychiatry 2022;79(8):752-759 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.1627] [Medline: 35662314]
53. Gillard S, Gibson SL, Holley J, Lucock M. Developing a change model for peer worker interventions in mental health services: a qualitative research study. Epidemiol Psychiatr Sci 2015;24(5):435-445 [FREE Full text] [doi: 10.1017/S2045796014000407] [Medline: 24992284]
54. Thom K, Lenagh-Glue J, O'Brien AJ, Potiki J, Casey H, Dawson J, et al. Service user, whānau and peer support workers' perceptions of advance directives for mental health. Int J Ment Health Nurs 2019;28(6):1296-1305. [doi: 10.1111/inm.12637] [Medline: 31361087]
55. Babatunde A, Ruck Keene A, Simpson A, Gilbert S, Stephenson L, Chua KC, et al. Advance statements for black African and caribbean people (AdStAC): protocol for an implementation study. BMC Psychiatry 2023;23(1):344 [FREE Full text] [doi: 10.1186/s12888-023-04825-6] [Medline: 37198589]
56. Bates N, Schell J, Azzam P, Childers J. Advance care planning education for psychiatrists: a novel training workshop (S805). J Pain Symptom Manage 2019;57(2):482-483 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.12.258]
57. Swanson JW, Swartz MS, Hannon MJ, Elbogen EB, Wagner HR, McCauley BJ, et al. Psychiatric advance directives: a survey of persons with schizophrenia, family members, and treatment providers. Int J Forensic Ment Health 2003;2(1):73-86 [FREE Full text] [doi: 10.1080/14999013.2003.10471180]
58. Isaacson MJ. Wakanki ewastepikte: an advance directive education project with American indian elders. J Hosp Palliat Nurs 2017;19(6):580-587. [doi: 10.1097/njh.0000000000000392]
59. Wilder CM, Swanson JW, Bonnie RJ, Wanchek T, McLaughlin L, Richardson J. A survey of stakeholder knowledge, experience, and opinions of advance directives for mental health in Virginia. Adm Policy Ment Health 2013;40(3):232-239. [doi: 10.1007/s10488-011-0401-9] [Medline: 22240937]
60. Backlar P, McFarland BH, Swanson JW, Mahler J. Consumer, provider, and informal caregiver opinions on psychiatric advance directives. Adm Policy Ment Health 2001;28(6):427-441. [doi: 10.1023/a:1012214807933] [Medline: 11804010]
61. Gill RK, Droney J, Owen G, Riley J, Stephenson L. Digital advance care planning with severe mental illness: a retrospective observational cohort analysis of the use of an electronic palliative care coordination system. BMC Palliat Care 2024;23(1):56 [FREE Full text] [doi: 10.1186/s12904-024-01381-y] [Medline: 38403633]
62. Cohen AN, Gorrindo T. New tools for implementing evidence-based care for serious mental illness. Focus (Am Psychiatr Publ) 2020;18(4):432-435 [FREE Full text] [doi: 10.1176/appi.focus.20200023] [Medline: 33343255]
63. Hochwarter S, Fehler G, Muente C, Eisenmann Y, Heinze M, Hibsch C, et al. Design of a patient-accessible electronic health record system in mental health. Stud Health Technol Inform 2022;294:583-584. [doi: 10.3233/SHTI220534] [Medline: 35612155]

XSL•FO
RenderX

64. Schwarz J, Meier-Diedrich E, Neumann K, Heinze M, Eisenmann Y, Thoma S. Reasons for acceptance or rejection of online record access among patients affected by a severe mental illness: mixed methods study. JMIR Ment Health 2024;11:e51126 [FREE Full text] [doi: 10.2196/51126] [Medline: 38315523]
65. Government response to the Joint Committee on the draft Mental Health Bill. Department of Health & Social Care. 2024. URL: https://tinyurl.com/3ybmhmp6 [accessed 2025-02-13]
66. Carreras Tartak JA, Brisbon N, Wilkie S, Sequist TD, Aisiku IP, Raja A, et al. Racial and ethnic disparities in emergency department restraint use: a multicenter retrospective analysis. Acad Emerg Med 2021;28(9):957-965 [FREE Full text] [doi: 10.1111/acem.14327] [Medline: 34533261]
67. Shea T, Dotson S, Tyree G, Ogbu-Nwobodo L, Beck S, Shtasel D. Racial and ethnic inequities in inpatient psychiatric civil commitment. Psychiatr Serv 2022;73(12):1322-1329. [doi: 10.1176/appi.ps.202100342] [Medline: 35959533]
68. Simpson AJ, Babatunde A, Simpson A, Gilbert S, Ruck Keene A, Stephenson L, et al. Implementing an advance choice documents resource for black African and caribbean people with experience of compulsory psychiatric admission. bioRxiv 2024. [doi: 10.1101/2024.04.17.24305837]
69. Redahan M, Kelly BD. Artificial intelligence and mental capacity legislation: opening Pandora's modem. Int J Law Psychiatry 2024;94:101985 [FREE Full text]
70. Foy J, Macrae A, Thom A, Macharouthu A. Advance statements: survey of patients' views and understanding. Psychiatr. bull 2018;31(9):339-341 [FREE Full text] [doi: 10.1192/pb.bp.106.013599]
71. Spanakis P, Peckham E, Mathers A, Shiers D, Gilbody S. The digital divide: amplifying health inequalities for people with severe mental illness in the time of COVID-19. Br J Psychiatry 2021;219(4):529-531 [FREE Full text] [doi: 10.1192/bjp.2021.56] [Medline: 35048887]
72. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
73. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
74. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
75. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with Schizophrenia: results of an online survey. JMIR Ment Health 2016;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]
76. Kozelka EE, Acquilano SC, Al-Abdulmunem M, Guarino S, Elwyn G, Drake RE, et al. Documenting the digital divide: Identifying barriers to digital mental health access among people with serious mental illness in community settings. SSM - Mental Health 2023;4:100241. [doi: 10.1016/j.ssmmh.2023.100241]
77. Rantanen T, Gluschkoff K, Silvennoinen P, Heponiemi T. The associations between mental health problems and attitudes toward web-based health and social care services: evidence from a finnish population-based study. J Med Internet Res 2021;23(9):e28066 [FREE Full text] [doi: 10.2196/28066] [Medline: 34546184]
78. Francis CJ, Johnson A, Wilson RL. Supported decision-making interventions in mental healthcare: a systematic review of current evidence and implementation barriers. Health Expect 2024;27(2):e14001 [FREE Full text] [doi: 10.1111/hex.14001] [Medline: 38433012]
79. Redahan M, Kelly BD, Gergel T. Advance healthcare directives and advance choice documents in psychiatry: new resources, new legislation, new opportunities. Int J Law Psychiatry 2024;97:102030. [doi: 10.1016/j.ijlp.2024.102030] [Medline: 39413509]
80. Pithara C, Farr M, Sullivan SA, Edwards HB, Hall W, Gadd C, et al. Implementing a digital tool to support shared care planning in community-based mental health services: qualitative evaluation. J Med Internet Res 2020;22(3):e14868 [FREE Full text] [doi: 10.2196/14868] [Medline: 32191210]
81. Morris NP. Digital technologies and coercion in psychiatry. Psychiatr Serv 2021;72(3):302-310. [doi: 10.1176/appi.ps.202000427] [Medline: 33430653]
82. Gooding PM. ‘Digitising the Mental Health Act’: are we facing the app-ification and platformisation of coercion in mental health services? In: Donnelly M, Kelly B, editors. Handbook of Mental Health Law. London, United Kingdom: Routledge; 2023.

Abbreviations

PAD: psychiatric advance directive

XSL•FO
RenderX

Edited by A Mavragani; submitted 08.11.24; peer-reviewed by G Owen, T Gergel; comments to author 16.12.24; revised version received 07.01.25; accepted 08.01.25; published 18.03.25.

Please cite as:
Schwarz J, Meier-Diedrich E, Scholten M, Stephenson L, Torous J, Wurster F, Blease C
Integration of Psychiatric Advance Directives Into the Patient-Accessible Electronic Health Record: Exploring the Promise and Limitations
J Med Internet Res 2025;27:e68549
URL: https://www.jmir.org/2025/1/e68549
doi:10.2196/68549
PMID:

©Julian Schwarz, Eva Meier-Diedrich, Matthé Scholten, Lucy Stephenson, John Torous, Florian Wurster, Charlotte Blease. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Lessons Learned From European Health Data Projects With Cancer Use Cases: Implementation of Health Standards and Internet of Things Semantic Interoperability

Amelie Gyrard[1], PhD; Somayeh Abedian[2], PhD; Philip Gribbon[3,4], PhD; George Manias[5], PhD; Rick van Nuland[6], PhD; Kurt Zatloukal[7], MD; Irina Emilia Nicolae[8], PhD; Gabriel Danciu[8], PhD; Septimiu Nechifor[8], PhD; Luis Marti-Bonmati[9], Dr med; Pedro Mallol[9], MSc; Stefano Dalmiani[10], MSc; Serge Autexier[11], PhD; Mario Jendrossek[12], MSc; Ioannis Avramidis[13], MSc; Eva Garcia Alvarez[14], PhD; Petr Holub[14], PhD; Ignacio Blanquer[15], Prof Dr; Anna Boden[16], Dr med; Rada Hussein[2], PhD

[1]Trialog, Paris, France

[2]Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

[3]Discovery Research - ScreeningPort, Fraunhofer Institute for Translational Medicine and Pharmacology, Hamburg, Germany

[4]Fraunhofer Cluster of Excellence for Immune-Mediated Diseases, Frankfurt, Germany

[5]Department of Digital Systems, University of Piraeus, Piraeus, Greece

[6]Lygature, Utrecht, The Netherlands

[7]Diagnostic and Research Center for Molecular Biomedicine, Diagnostic and Research Institute of Pathology, Medical University of Graz, Graz, Austria

[8]Siemens Foundational Technologies, Brasov, Romania

[9]La Fe University Hospital Valencia, Valencia, Spain

[10]Monasterio Research Hospitals, Pisa, Italy

[11]Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Bremen, Germany

[12]Health Data Hub, Paris, France

[13]Ubitech, Athens, Greece

[14]Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium, Graz, Austria

[15]Universitat Politècnica de València, València, Spain

[16]Department of Clinical Pathology, Centre for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

Corresponding Author:
Amelie Gyrard, PhD
Trialog
25 rue du Général Foy
Paris, 75008
France
Phone: 33 033 1 44 70 61
Email: amelie.gyrard@trialog.com

Abstract

The adoption of the European Health Data Space (EHDS) regulation has made integrating health data critical for both primary and secondary applications. Primary use cases include patient diagnosis, prognosis, and treatment, while secondary applications support research, innovation, and regulatory decision-making. Additionally, leveraging large datasets improves training quality for artificial intelligence (AI) models, particularly in cancer prevention, prediction, and treatment personalization. The European Union (EU) has recently funded multiple projects under Europe's Beating Cancer Plan. However, these projects face challenges related to fragmentation and the lack of standardization in metadata, data storage, access, and processing. This paper examines interoperability standards used in six EU-funded cancer-related projects: IDERHA (Integration of Heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance), EUCAIM (European Cancer Imaging Initiative), ASCAPE (Artificial Intelligence Supporting Cancer Patients Across Europe), iHelp, BigPicture, and the HealthData@EU pilot. These initiatives aim to enhance the analysis of heterogeneous health data while aligning with EHDS implementation, specifically for the EHDS for the secondary use of data (EHDS2). Between October 2023 and July 2024, we organized meetings and workshops among these projects to assess how they adopt health standards and apply Internet of Things (IoT) semantic interoperability. The

XSL•FO
RenderX

discussions focused on interoperability standards for health data, knowledge graphs, the data quality framework, patient-generated health data, AI reasoning, federated approaches, security, and privacy. Based on our findings, we developed a template for designing the EHDS2 interoperability framework in alignment with the new European Interoperability Framework (EIF) and EHDS governance standards. This template maps EHDS2-recommended standards to the EIF model and principles, linking the proposed EHDS2 data quality framework to relevant International Organization for Standardization (ISO) standards. Using this template, we analyzed and compared how the recommended EHDS2 standards were implemented across the studied projects. During workshops, project teams shared insights on overcoming interoperability challenges and their innovative approaches to bridging gaps in standardization. With support from HSbooster.eu, we facilitated collaboration among these projects to exchange knowledge on standards, legal implementation, project sustainability, and harmonization with EHDS2. The findings from this work, including the created template and lessons learned, will be compiled into an interactive toolkit for the EHDS2 interoperability framework. This toolkit will help existing and future projects align with EHDS2 technical and legal requirements, serving as a foundation for a common EHDS2 interoperability framework. Additionally, standardization efforts include participation in the development of ISO/IEC 21823-3:2021—Semantic Interoperability for IoT Systems. Since no ISO standard currently exists for digital pathology and AI-based image analysis for medical diagnostics, the BigPicture project is contributing to ISO/PWI 24051-2, which focuses on digital pathology and AI-based, whole-slide image analysis. Integrating these efforts with ongoing ISO initiatives can enhance global standardization and facilitate widespread adoption across health care systems.

(J Med Internet Res 2025;27:e66273) doi:10.2196/66273

KEYWORDS

artificial intelligence; cancer; European Health Data Space; health care standards; interoperability; AI; health data; cancer use cases; IoT; Internet of Things; primary data; diagnosis; prognosis; decision-making

Introduction

Cancer is one of the main causes of death in Europe and worldwide, after cardiovascular diseases. According to the World Health Organization (WHO), cancer is the second leading cause of death and morbidity in Europe, with more than 3.7 million new cases and 1.9 million deaths each year [1]. In response to the urgent need to renew European political commitment to tackle cancer, Europe's Beating Cancer Plan (EBCP; 2021-2027) was launched and structured around 4 key action areas: prevention, early detection, diagnosis and treatment, and the quality of life of patients with cancer and cancer survivors [2]. Since 2021, the European Commission (EC) has supported collaborative projects focused on cancer diagnostics and treatment using high-performance computing and artificial intelligence (AI). To maximize the potential of data and digitalization, the EBCP also addressed the interactions and alignment of cancer data projects and initiatives with the European Health Data Space (EHDS). In autumn 2023, the EC organized 3 online workshops on the reuse of health data resources in the field of cancer research and recently published the results of these workshops [3], where "fragmentation and the lack of standardization in metadata, data storage, access, and processing" were identified as key challenges facing data-driven cancer projects. The nascent infrastructure for the application of AI in medical imaging (European Cancer Imaging Initiative [EUCAIM]) reported on the experiences of 5 projects developing big data infrastructures that will enable European, ethical, General Data Protection Regulation (GDPR)–compliant, quality-controlled, and cancer-related medical imaging platforms, where both large-scale data and AI algorithms will coexist [4]. These projects include the following:

1. Chameleon: a project focused on developing AI algorithms for cancer diagnosis and prognosis.
2. EuCanImage: a project aimed at creating a large-scale cancer image database.
3. ProCAncer-I: a project focused on developing AI-based tools for personalized cancer treatment.
4. Incisive: a project contributing a significant amount of cancer image data to EUCAIM.
5. Primage: a project focused on developing AI-based image analysis techniques for cancer diagnosis.

These projects along with the RadioVal project established the AI for Health Imaging (AI4HI) network to develop cancer imaging data repositories and AI solutions based on medical imaging to improve clinical practice.

Table 1 summarizes the list of projects that participated in the EC workshop entitled "Landscaping data-driven projects and initiatives in the cancer field–rationale and directions for better collaboration and integration" [3], as well as the established project-level synergies and collaborations among ongoing projects.

The workshop also addressed the current challenges facing data-driven cancer projects [3], gaps existing in existing standards, and recommendations for future semantic interoperability (given in Table 2).

XSL•FO
RenderX

Table 1. Collaborations and synergies among the European Union's existing cancer projects.

Collaboration scope	Synergy projects
Data representation and interoperability	HealthData@EU pilot and CanSERV
Infrastructure and services for benchmarking	EOSC4Cancer, EUCAIM[a], and TEF-Health
Federated data infrastructure and data structure	EUCAIM, EOSC4Cancer, and GDI
Secure processing environment	SOLACE, EUCAIM, CanScreen-ECIS, IDERHA[b], and Optima

[a]EUCAIM: European Cancer Imaging Initiative.

[b]IDERHA: Integration of Heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance.

Table 2. Main challenges facing European Union cancer projects and recommendations.

Challenges [3]	Gaps in existing standards	Recommendations
Fragmentation and lack of standardization in metadata, data storage, access, and processing.	The DCAT-AP[a] health extension is still under development by the HealthData@EU pilot.	TEHDAS[b] JA[c] recommendations to enhance interoperability within HealthData@EU—a framework for semantic, technical, and organizational interoperability.
Lack of access to diverse and high-quality datasets.	Data quality framework for primary care data sources is not sufficiently addressed in the EHDS[d] framework.	Data quality frameworks provided by TEHDAS, EMA[e], and QUANTUM projects.
Evolving legal landscape	Relevant standards to the AI[f] Act are under development.	Participating in developing or extending the relevant standards.

[a]DCAT-AP: DCAT Application Profile for Data Portals in Europe.

[b]TEHDAS: Towards the European Health Data Space.

[c]JA: joint action.

[d]EHDS: European Health Data Space.

[e]EMA: European Medicines Agency.

[f]AI: artificial intelligence.

HSbooster Health Project Cluster in Cancer

In this work, we focus on the interoperability challenges and existing gaps in health care standards building on the challenges and synergies highlighted in the EC workshop [3]. Through the EC European Standardization Booster (HSbooster.eu) initiative, we create synergy among six cancer data-driven projects (ie, IDERHA [Integration of Heterogeneous Data and Evidence towards Regulatory and Health Technology Assessments Acceptance], EUCAIM, ASCAPE [Artificial Intelligence Supporting Cancer Patients Across Europe], iHelp, BigPicture, and the HealthData@EU pilot project) by using health standards. These 6 innovative projects aim to transform digital health in oncology by leveraging advanced technologies and collaborative frameworks for enhancing cancer diagnosis, treatment, and research, improving patient outcomes, and accelerating scientific advancements. We initially created a template for the EHDS for the secondary use of data (EHDS2) interoperability framework based on the new European Interoperability Framework (EIF). The recommended standards and governance model from the joint action (JA) Towards the European Health Data Space (TEHDAS) [5] for the secondary use of data were then used to harmonize the standards in the template (given in Textbox 1).

Starting October 2023, we conducted several meetings and workshops among the 6 projects to elucidate how they adopt health standards and Internet of Things (IoT) semantic interoperability. This included examining interoperability standards for health data, knowledge graphs–related technologies, the Smart Applications Reference Ontology (SAREF), the data quality framework (DQF), patient-generated health data (PGHD), AI reasoning, federated approaches, security, and privacy.

As per the TEHDAS recommendations, we compared the health-standardized models, ontologies, and terminologies used in these projects, including, Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR), Open Health Data Science and Informatics (OHDSI), Observational Medical Outcomes Partnership (OMOP)–Common Data Model (CDM), Digital Imaging and Communications in Medicine (DICOM), International Organization for Standardization (ISO) Technical Committee (TC) 215, and World Wide Web Consortium (W3C) Data Catalog Vocabulary (DCAT).

As EHDS is still evolving and its implementation is still under development, this study aims to examine how the 6 projects implement the recommended EHDS2 standards. Consequently, we introduce a new template to support the EHDS2 interoperability framework and highlight the results of comparing data standards across projects. By summarizing the

XSL•FO
RenderX

lessons learned, we provide recommendations for future directions in EHDS2 implementation.

Textbox 1. European Health Data Space for European Interoperability Framework.

European Health Data Space

- The European Health Data Space (EHDS) [6] aims to harmonize health data usage across Europe [7]. It includes comprehensive rules, standards, practices, infrastructures, and a robust governance framework to improve health care delivery, drive research and innovation, and inform policymaking.
- A central focus of the EHDS is empowering patients by providing them greater digital access and control over their health data, fostering a more transparent and efficient health care system. In addition, the EHDS seeks to standardize health data across member states, ensuring interoperability, enhancing clinical outcomes, and accelerating medical research and innovation through a unified dataset.
- The governance framework ensures data privacy and security, addresses ethical concerns, and fosters stakeholder trust. The EHDS also aims to streamline regulatory processes and support cross-border health care initiatives by aligning national and regional policies.
- Therefore, the EHDS will significantly impact Europe's digital health landscape by promoting a more connected, efficient, and patient-centered health care ecosystem.

Joint Action Toward the European Health Data Space

- The Toward the European Health Data Space (TEHDAS) project [5] is a joint action that developed European principles for the secondary use of health data, involving 25 countries. It advocates for using existing International Organization for Standardization (ISO) standards to ensure consistency and interoperability. TEHDAS focuses on data interoperability, identifying standards for data discovery and common data models. A synthesis table details essential standards including DCAT Application Profile for Data Portals in Europe (DCAT-AP) [8], INSPIRE, FairSharing, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), and Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR).
- The project addresses syntactic interoperability, providing guidelines on data structure and format for seamless sharing across systems. In addition, TEHDAS emphasizes data quality, proposing frameworks to regulate and enhance the reliability of health data for secondary uses, such as research and policymaking. Overall, TEHDAS's work supports a standardized and interoperable EHDS, leading to improved health care, research, and innovation in Europe.

European Interoperability Framework

- The new European Interoperability Framework (EIF) [9] adopted on March 23, 2017, fosters electronic communication and information exchange among European public administrations by providing common models, principles, and recommendations. The EIF delineates four layers of interoperability challenges: legal, organizational, semantic, and technical. This paper concentrates on the semantic interoperability challenge, encompassing both semantic and syntactic aspects. Semantic interoperability ensures that the meaning of exchanged data is preserved and understood across different systems. This can be achieved through the use of shared vocabularies and schemas. On the other hand, syntactic interoperability involves defining the grammar and format for data exchange.
- The EIF provides guidance for European public administrations to improve interoperability, ensuring smooth data exchange and efficient digital service delivery. Conversely, ISO 23903 concentrates on semantic interoperability within Information and Communication Technology (ICT) systems, aiming to unify the interpretation of information across various platforms.
- Aligning the EIF with ISO 23903 synchronizes European interoperability principles with semantic standards, fostering easier communication and cooperation among public administrations. This alignment ensures both technical harmony and uniformity in the meaning and semantics of exchanged data, thereby enhancing the efficiency and efficacy of digital service provision for citizens and businesses across Europe.

Template for the EHDS2 Interoperability Framework Based on the EIF

Figure 1 shows the process of creating the template for the EHDS2 interoperability framework.

XSL•FO
RenderX

Figure 1. The process for creating the proposed template for the EHDS2 interoperability framework. EHDS: European Health Data Space; EHDS2: European Health Data Space for the secondary use of data; EIF: European Interoperability Framework; ISO: International Organization for Standardization; TEHDAS: Towards the European Health Data Space.

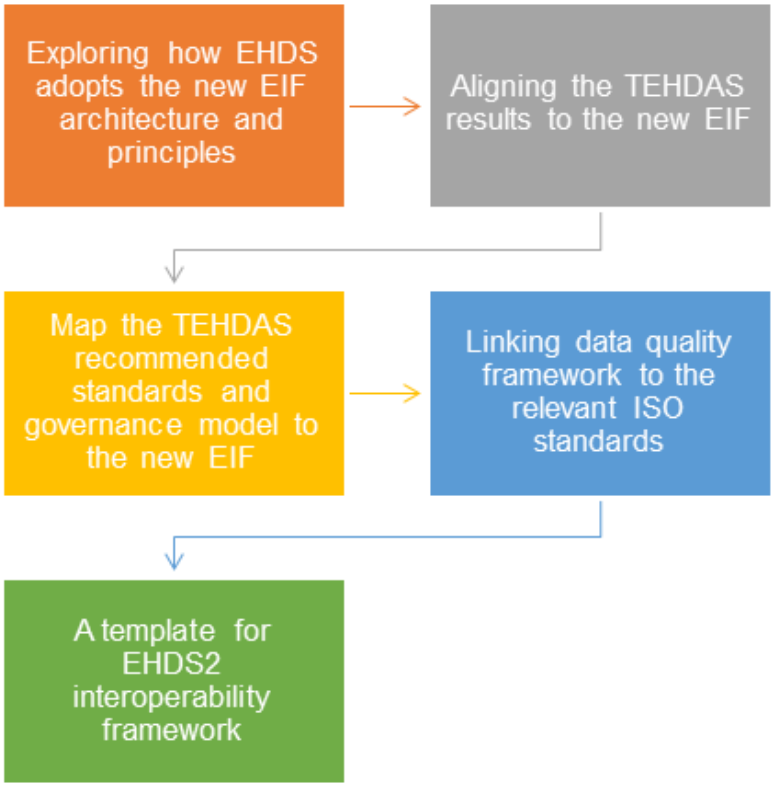

Exploring How EHDS Adopts the New EIF Architecture and Principles

Recent studies showed the importance of adopting the EIF in establishing the EHDS interoperability framework [7,10,11] (as shown in Textbox 1). In addition, several studies have investigated the interoperability requirements for heterogeneous health information systems, as well as the associated health care standards [12,13].

Aligning the TEHDAS Results to the New EIF

The TEHDAS JA involves 25 European countries in developing the principles that will shape EHDS2 by providing guidance and recommendations on interoperability, data quality, and standards (as shown in panel 2 in Textbox 1). In 2023, the TEHDAS JA published a report titled "Options for governance models for the European Health Data," which discusses EHDS governance using the EIF [14]. In addition, they assessed 19 standards that provide a layer of semantic interoperability, supporting the cataloging of data sources and the exchange of data between different nodes [15].

Mapping the TEHDAS Recommended Standards and Governance Model to the New EIF

Following the recommendation of the EC workshops on landscaping data-driven projects and initiatives in cancer [3], we adopted the TEHDAS results on recommended standards for EHDS2 interoperability, as well as the principles of data quality frameworks. We then mapped these results to the new EIF interoperability framework and linked the data quality framework to the corresponding ISO standards. Figure 2 shows how the TEHDAS results are mapped to the new EIF interoperability framework. Notably, TEHDAS framed the recommended EHDS2 interoperability into 3 categories of standardization:

1. Data discoverability (DCAT-AP)
2. Enabling semantics interoperability (OMOP-CDM)
3. Health data exchange (DICOM for imaging data and FHIR for health records).

The template shows the importance of incorporating and harmonizing the underpinning standards to comply with new regulations, such as the AI Act [16]. As a result, integrating advanced AI methodologies within the proposed framework enhances TEHDAS principles and advances beyond EIF by embedding robust data quality frameworks, enabling federated learning for secure and scalable data sharing, and incorporating privacy-preserving mechanisms that comply with GDPR. This framework delivers actionable templates that bridge theoretical interoperability principles with real-world AI applications, such as cancer diagnostics and personalized treatment. By aligning technical standards with semantic requirements and supporting adaptability to evolving regulations, such as the European AI Act, the framework provides a scalable, future-ready solution for health care interoperability and AI-driven innovation.

The EIF does not classify standards according to interoperability layers (technical, semantic, legal, and organizational). However,

XSL•FO
RenderX

we need to consider the relevant standards for data privacy and quality assessment [10].

Furthermore, the work is a foundation for creating an interactive tool for the EHDS2 Interoperability Framework, which has been submitted to JMIR as part 2 (Towards EHDS2 interoperability framework: An interactive EIF-based standards compliance toolkit for AI-driven projects).

Figure 2. Template for the EHDS2 interoperability framework based on the TEHDAS results: (A) EIF layers; (B) TEHDAS governance model and associated interoperability framework [14]; and (C) TEHDAS 19 standards evaluation using the EIF, with the selected standards that "passed the EIF criteria evaluation" underlined [15]. DCAT-AP: DCAT Application Profile for Data Portals in Europe; DICOM: Digital Imaging and Communications in Medicine; EHDS: European Health Data Space; EHDS2: European Health Data Space for the secondary use of data; EIF: European Interoperability Framework; FHIR: Fast Health Interoperability Resources; GDPR: General Data Protection Regulation; HL7: Health Level 7; ISO: International Organization for Standardization; LOINC: Logical Observation Identifiers Names and Codes; OMOP-CDM: Observational Medical Outcomes Partnership Common Data Model; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terminology.

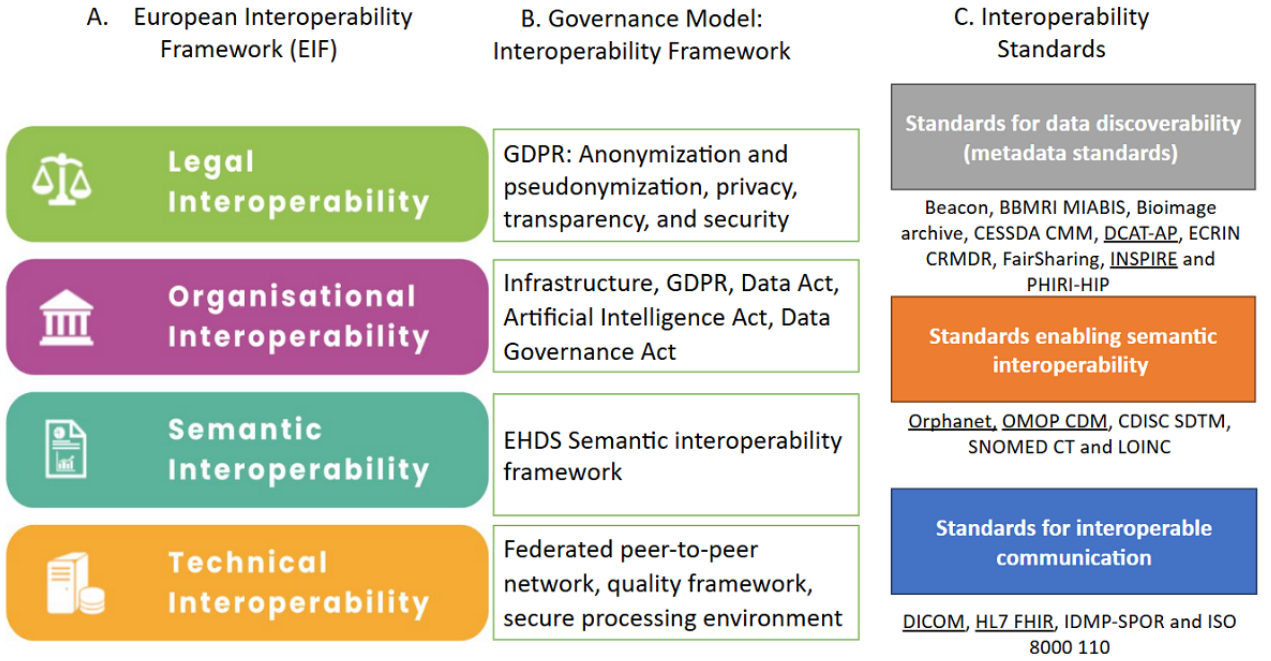

Linking DQF to the Relevant ISO Standards

In May 2024, the Big Data Value Association (BDVA) published a study entitled "Elevating Data Quality A Paradigm Shift for Data Spaces and AI Needs" to explore the relationships between data quality and data spaces with respect to the AI Act [17]. The study proposed that the data quality and utility label for EHDS should comply with data documentation, technical quality, data quality management processes, coverage, access and provision, and data enrichment procedures.

From the EHDS2 perspective, the TEHDAS JA provided a generic DQF [18], which includes both technical quality elements and six utility dimensions: relevance, accuracy and reliability, coherence, coverage, completeness, and timelines. This was followed by the publication of associated recommendations affecting data quality and utility implementation in HealthData@EU [19]. Based on these deliverables, the European Medicine Agency (EMA) published its own DQF for EU medicines regulation [20]. This publication provides an analysis of the data quality actions and metrics, as well as a maturity model, to guide the evolution of automation to support data-driven regulatory decision-making (as shown in Figure 3) [20].

Ensuring the accuracy, completeness, consistency, and reliability of cancer research data is of utmost importance, and adherence to data quality standards plays a crucial role in achieving this goal. These standards not only help maintain data integrity but also promote data interoperability through the use of standardized data models and vocabularies. This enables seamless data exchange and integration across different projects and platforms. In addition, adhering to these standards supports informed decision-making by providing high-quality data for clinical decisions, research insights, and policymaking related to cancer treatment and prevention. Compliance with international legal and regulatory requirements is also ensured, upholding proper data governance and ethical data usage. Finally, establishing a unified framework for data quality encourages collaboration among various stakeholders within the cancer research community. We compare data quality standards in Table S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 3. European Medicines Agency data quality framework in terms of the (A) dimensions of data quality, (B) maturity levels, and (C) related ISO standards. ISO: International Organization for Standardization; QMS: quality management system.

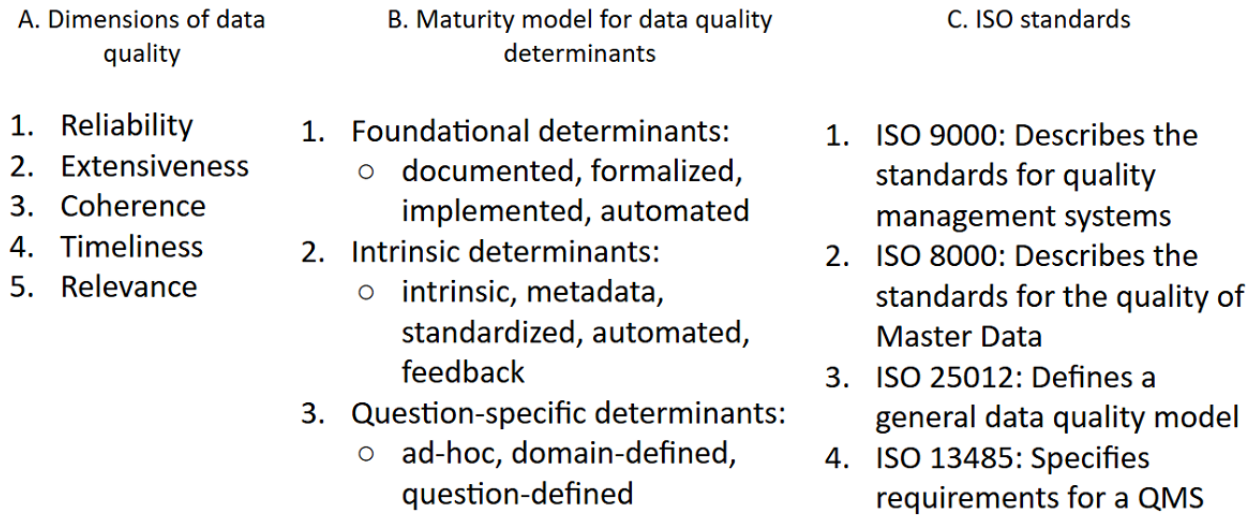

Landscape of the Involved Projects and Used Standards

In this study, we used the EHDS2 interoperability framework template to compare 6 projects with cancer use cases where AI is applied (as shown in panel 4 in Textbox 2; Table 3). The selected projects vary in scope, cancer domain, categories of health data, scale of infrastructure, AI implementation approach, and time spans [21].

We focused on semantic interoperability standards (as shown in Textbox 3), recommended by the TEHDAS JA and related ISO standards. The analysis explored how health standards support health ontologies, for example, how HL7 FHIR supports the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles through the FHIR for FAIR implementation guide [28].

Textbox 2. Introduction to six key projects advancing oncology through digital health.

We explore 6 innovative projects that are transforming digital health in oncology. By leveraging advanced technologies and collaborative frameworks, these projects aim to enhance cancer diagnosis, treatment, and research, improving patient outcomes, and accelerating scientific advancements:

- IDERHA (Integration of Heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance) [22] (2023-2028): IDERHA advances digital health in the context of lung cancer. It seeks to enhance health care delivery and improve patient outcomes by integrating various health data sources to support effective clinical decision-making and the development of improved treatments.
- BigPicture [23] (2021-2027): BigPicture aims to revolutionize pathology practice by establishing a repository of high-quality, annotated images, and developing artificial intelligence (AI) tools to improve diagnostic accuracy and efficiency. Data are collected from all types of tissues, including 1 million clinical samples and 2 million nonclinical samples from toxicology studies.
- EUCAIM (European Cancer Imaging Initiative) [24] (2023-2026): EUCAIM provides a federated infrastructure for sharing and analyzing cancer images. This supports AI research and clinical practices to advance cancer diagnosis and treatment.
- iHelp [25] (2021-2024): The iHelp project is dedicated to developing an intelligent, AI-driven health monitoring platform for early disease detection and management. It offers personalized health care solutions that enhance patient engagement and outcomes, with a specific focus on pancreatic cancer.
- ASCAPE (Artificial Intelligence Supporting Cancer Patients Across Europe) [26] (2020-2023): ASCAPE leverages AI to enhance the quality of life for patients with cancer. By using advanced analytics on health data, the project develops predictive models and personalized treatment plans to support patient care and survivorship.
- HealthData@EU Pilot [27] (2022-2024): This project aims to establish a technical infrastructure to enable cross-border health data access and use within the European Union, empowering the secondary use of health data more broadly. This initiative supports the creation of a cohesive digital health ecosystem, fostering innovation and improving health care services across Europe.

XSL•FO
RenderX

Table 3. Six projects with cancer use cases applying artificial intelligence.

Project	Duration	Cancer domain	AI[a] or ML[b]	Comments
IDERHA[c]	2023-2028	Lung cancer	Federated machine learning	Under implementation
BigPicture	2021-2027	Pan-cancer whole slide images	Central repository for digital pathology and platform for AI development	Ongoing
EUCAIM[d]	2023-2026	Pan-cancer images	Federated Research Infrastructure	Ongoing
iHelp	2021-2024	Pancreatic cancer	Explainable AI, deep neural networks, predictive algorithms, ML techniques, and federated queries on distributed infrastructures	Ended on June 30, 2024
ASCAPE[e]	2020-2023	Breast and Prostate cancer	Explainable AI, federated deep learning, and ML on homomorphically encrypted data	Finished
HealthData@EU pilot	2022-2024	Colorectal cancer and other nonrelated cancer use cases	Federated query in noncancer use case: machine learning for analysis of care pathways	Pilot for EHDS2

[a]AI: artificial intelligence.

[b]ML: machine learning.

[c]IDERHA: Integration of Heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance.

[d]EUCAIM: European Cancer Imaging Initiative.

[e]ASCAPE: Artificial Intelligence Supporting Cancer Patients Across Europe.

XSL•FO
RenderX

Textbox 3. Advancing semantic interoperability in digital health: a landscape of standards used in the involved projects.

Our attention is drawn to the critical role that standards play in examining semantic interoperability within the digital health realm. These standards form the foundation for seamless data exchange, ensuring consistency and coherence across various health care platforms. The main standards are as follows:

- Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) [29]: Positioned as a cornerstone in health care interoperability, FHIR stands out for its efficiency, adaptable architecture, and robust data exchange mechanisms, which facilitate smooth communication across disparate systems.
- DICOM (Digital Imaging and Communications in Medicine) [30]: DICOM is widely adopted in clinical and research settings, underscoring its effectiveness in enabling image exchange.
- Common Data Model (CDM)–OMOP (Observational Medical Outcomes Partnership) [31]: While not strictly a standard, CDM-OMOP allows for data harmonization and analysis. Its synergy with established standards enhances overall interoperability.
- International Organization for Standardization (ISO) [32]:
  - ISO TC 215 and CEN TC 251 Health Informatics: These cover a range of ISO standards, such as 13606 (5 parts), 27269, and 29585, alongside the Health Informatics Service Architecture (HISA)–compliant data models.
  - ISO 13606: This focuses on exchanging information on electronic health records between different health care information systems. It has a comprehensive structure for organizing clinical data, including patient demographics, medical history, diagnoses, treatments, and other relevant information.
  - ISO/AWI TR 24305: This guideline, derived from ISO standards 13940 and 13606, offers a structured HL7 FHIR implementation framework. It strengthens interoperability infrastructure, aligns with ISO norms, and promotes uniformity and adherence to best practices.
  - ISO 23903: Titled “Health Informatics - Interoperability and Integration Reference Architecture (IIRA),” this standard addresses the integration and interoperability challenges in health information systems and services, ensuring they can communicate effectively and share data seamlessly.
  - ISO 13972: Titled “Health Informatics - Detailed Clinical Models (DCM),” this standard focuses on the standardization and interoperability of clinical information models. DCMs are specifications that represent the clinical concepts and data elements used in electronic health records (EHRs) and health information systems.
- Internet of Things and eHealth coding standards
  - ETSI SmartM2M Smart Applications Reference Ontology (SAREF) [33]: Tailored for the eHealth and healthy aging domains, representing a specialization within standardization. Its integration expands interoperability across different targeted demographic groups.
  - Systematized Nomenclature of Medicine (SNOMED) [34]: SNOMED is a comprehensive clinical terminology system that captures, encodes, and shares EHRs and other health care data. It provides a standardized vocabulary for describing clinical concepts, enabling interoperability and semantic consistency across health care systems. SNOMED facilitates accurate and meaningful exchange of clinical information, supporting clinical decision-making, research, and public health reporting.
  - Logical Observation Identifiers Names and Codes (LOINC) [35]: LOINC is a standardized system for identifying and exchanging clinical laboratory test results and other clinical observations. It provides a universal set of codes and terms for describing laboratory tests, measurements, and observations, ensuring interoperability and semantic consistency in healthcare data exchange. LOINC enables seamless integration of laboratory data into EHRs and other health information systems.
  - International Classification of Diseases (ICD) [36]: It is promoted by the World Health Organization (WHO). ICD codes are used to define diagnoses for clinical treatment, medical billing, and statistics collections. It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. These codes help physicians in determining which conditions are relevant for clinical decisions, injuries, and causes of death. However, ICD has some limitations (eg, diabetes has more than 2 dozen different codes). Patient data recorded with ICD can be used for administrative functions, epidemiologic studies, research subject recruitment, interventional protocols, and clinical decision support systems.

The other standards regarding data quality, security, artificial intelligence (AI), and wearables are listed in the Multimedia Appendix 1.

Abbreviations of semantic metadata or terminologies used by standards

- ATC: Anatomical Therapeutic Chemical Classification System; Birnlex: Biomedical Research Integrated Domain Group; CPT: Current Procedural Terminology; ICD: International Classification of Diseases; ICDO: International Classification of Diseases for Oncology; NAACR: North American Association of Central Cancer Registries; NCIT: National Cancer Institute Thesaurus; OSIRIS: Open Source Infrared Imaging System; RADLEX: Radiology Lexicon; RDFS: Resource Description Framework Schema; SPARQL: SPARQL Protocol and RDF Query Language; RxNorm: Prescription Normative Terminology; UCUM: Unified Code for Units of Measure; UMLS: Unified Medical Language System.

Comparing the Health Data Projects

To identify synergies among the 6 projects, we used the EHDS2 interoperability framework template to analyze adopted standards, highlighting similarities and differences in the implementation approach (detailed results given in Tables S1-S7 in Multimedia Appendix 1).

XSL•FO
RenderX

Mapping the Projects to the Created Template for the EHDS2 Interoperability Framework

The EHDS2 interoperability framework template identified key standards and technologies for the implementation of EHDS2, including HL7 FHIR; DICOM; OMOP-CDM; upcoming standards developed by ISO TC 215, ISO TC212, CEN TC 251; and ontology technologies, including W3C DCAT-AP. The key standards used in 6 projects, their focus areas, and planned future implementations are summarized in Table 4.

Table 4. Overview of the used, innovated, and planned health data standards in the projects (using the template of the European Health Data Space for the secondary use of data [EHDS2] interoperability framework).

Project	Used standards	Planned or future standards
IDERHA	HL7[a] FHIR[b], DICOM[c], and OMOP	ISO[d] TC[e] 215, DCAT-AP[f]: HealthDCAT-AP
BigPicture	DICOM, SNOMED[g], and *ICD*	ISO TC 212 (digital pathology and AI[h]), and Kidney Biopsy Codes
EUCAIM[i]	Using its own hyperontology and common data model based on: DICOM, DICOM Seg, OMOP, FHIR, mCODE, DCAT-AP-Health, OSIRIS	—[r]
	Ontologies used so far: LOINC[j], SNOMED, UCUM[k], RADLEX[l], ICDO3, *ICD-10*[m], CPT4, ICD10PCS, ATC[n], NCIT[o], Birnlex[p], NAACR[q], Cancer Modifier	
iHelp	HL7 FHIR, OMOP-CDM, ISO 27799:2016, SNOMED, LOINC, *ICD-9*, ICD-10, UMLS, SPARQL, RDFS[s], RxNorm	—
ASCAPE	HL7 FHIR, ISO/CEN 13606, LOINC, SNOMED	—
HealthData@EU pilot	Observing and collecting standardization efforts rather than direct implementation	The first version of the Health DCAT-AP metadata standard is being developed in the project

[a]HL7: Health Level 7.

[b]FHIR: Fast Health Interoperability Resources.

[c]DICOM: Digital Imaging and Communications in Medicine.

[d]ISO: International Organization for Standardization.

[e]TC: Technical Committee.

[f]DCAT-AP**:** DCAT Application Profile for Data Portals in Europe.

[g]SNOMED: Systematized Medical Nomenclature for Medicine.

[h]AI: artificial intelligence.

[i]EUCAIM: European Cancer Imaging Initiative.

[j]LOINC: Logical Observation Identifiers Names and Codes.

[k]UCUM: Unified Code for Units of Measure.

[l]RADLEX: Radiology Lexicon.

[m]ICD-10: International Classification of Diseases, Tenth Revision, Clinical Modification.

[n]ATC: Anatomical Therapeutic Chemical classification system.

[o]NCIT: National Cancer Institute Thesaurus.

[p]Birnlex: Biomedical Research Integrated Domain Group.

[q]NAACR: North American Association of Central Cancer Registries.

[r]Not applicable.

[s]RDFS: Resource Description Framework Schema.

Main Findings and Lessons Learned

All projects support the recommended standards identified by the TEHDAS JA for EHDS2 interoperability. Although the DICOM standard is widely used for collecting, storing, and transferring medical imaging data, it lacks important information required to identify relevant images because DICOM metadata are not standardized [37]. To overcome this challenge, the EUCAIM project CDM is built upon the FHIR resources ImagingStudy and ImagingSeries, the Medical Imaging–CDM extension of the OMOP-CDM, the ProCAncer-I imaging extension [38], and the OSIRIS imaging component. In this way, proper integration of imaging and clinical data was provided in alignment with the Integrating the Healthcare Enterprise [39].

The BigPicture project selected the DICOM format as a standard for archiving whole-slide images (WSI) to support back-and-forth conversion of proprietary file formats. The BigPicture profile for DICOM WSI is based on work by the

XSL•FO
RenderX

DICOM pathology working group WG26. The format is designed to allow efficient storage of a large number of annotations, for example, generated by AI algorithms [40]. The project developed a Python tool (Wsidicom) to serve as a reference implementation of a DICOM WSI reader [41]. In addition, two tools, Opentile and Wsidicomizer, were developed to read other WSI formats and convert them to DICOM resulting in faster format conversion and maintained image quality [42].

To ensure structured data exchange and harmonized-consistent metadata interaction for data not captured in DICOM, BigPicture has developed additional metadata standards that define and comprise a set of substandards: (1) the data model (Common Mandatory Metadata Structure[CMMS]) comprising of all metadata, (2) a set of mandatory information that must be provided in relation to CMMS entities and their relations to each other, (3) a metadata file format that is used within BigPicture (flexible metadata file exchange format), and (4) a standard file structure of datasets containing metadata and data files that can be found on the repository. Standards are based on the European Genome-phenome Archive (EGA), and where possible existing standards have been incorporated (SNOMED, ICD for clinical data, Standardization for Exchange of Nonclinical Data [SEND] terminology, and International Harmonization of Nomenclature and Diagnostic Criteria (INHAND) nomenclature for nonclinical data). Results from BigPicture, such as metadata formats or quality control criteria, will feed into the ISO/AWI 24051-2 “Medical laboratories—Part 2: Digital pathology and artificial intelligence (AI)-based image analysis.”

Regarding the metadata standards, the HealthData@EU pilot published a landscape analysis of available metadata catalogs and metadata standards [43]. The pilot also develops an extension of DCAT-AP: HealthDCAT-AP that will be adopted by EHDS2 projects, like IDERHA [44]. This standard could be adopted later as the mandatory metadata standard foreseen under the EHDS regulation. The EUCAIM project designed its Hyper Ontology [45] using FHIR, UMLS, SNOMED-CT, and OMOP-CDM vocabularies.

While data from wearables are available in high volumes, a substantial amount of data is lost because the usability of the data is governed and limited by proprietary data formats from an increasing number of manufacturers, which shows the necessity of standardization to enable interoperability [46]. Among the six projects and initiatives, IDERHA, iHelp, and ASCAPE use sensors and wearables technologies. The iHelp project uses the holistic health records (HHR) model to enable the aggregation of data from different sources, sensors, and online platforms to support the seamless integration of multiple health dimensions [47].

The HL7 FHIR standard-compatible data structures were used in the context of the iHelp project to integrate primary and secondary data and to compile them into the holistic health records FHIR model [47]. FHIR can be used for clinical data and also for streaming data from sensors [48]. The ASCAPE project successfully integrated EN/ISO 13606–standardized extracts from a patient mobile app into an electronic health record [49]. IDERHA plans to extend the OMOP-CDM to address the PGHD, including patient-reported outcomes and patient-reported experience measures.

The BigPicture project highlighted that a dedicated standardization project focused on digital pathology is currently missing. It addressed the need for ISO/AWI 24051-2 “Medical laboratories—Part 2: Digital pathology and artificial intelligence (AI)-based image analysis” (under development, April 2024) that will build on ISO 20166-4:2021 “Molecular in vitro diagnostic examinations Specifications for pre-examination processes for formalin-fixed and paraffin-embedded (FFPE) tissue Part 4: In situ detection techniques” [50]. Figure 4 summarizes the main findings of the projects and how they can feed each other.

XSL•FO
RenderX

Figure 4. Key outputs and lessons learned from the six projects. AI: artificial intelligence; ASCAPE: Artificial Intelligence Supporting Cancer Patients Across Europe; AWI: approved work item; CDM: Common Data Model; DCAT-AP: DCAT Application Profile for Data Portals in Europe; DICOM: Digital Imaging and Communications in Medicine; EHDS: European Health Data Space; EHDS2: European Health Data Space for the secondary use of data; EUCAIM: European Cancer Imaging Initiative; HHR: holistic health record; IDERHA: Integration of heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance; iHelp: Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records; OMOP: Observational Medical Outcomes Partnership; SNOMED-CT: Systematized Nomenclature of Medicine Clinical Terms; UMLS: Unified Medical Language System; WSI: whole-slide imaging; XAI: explainable artificial intelligence.

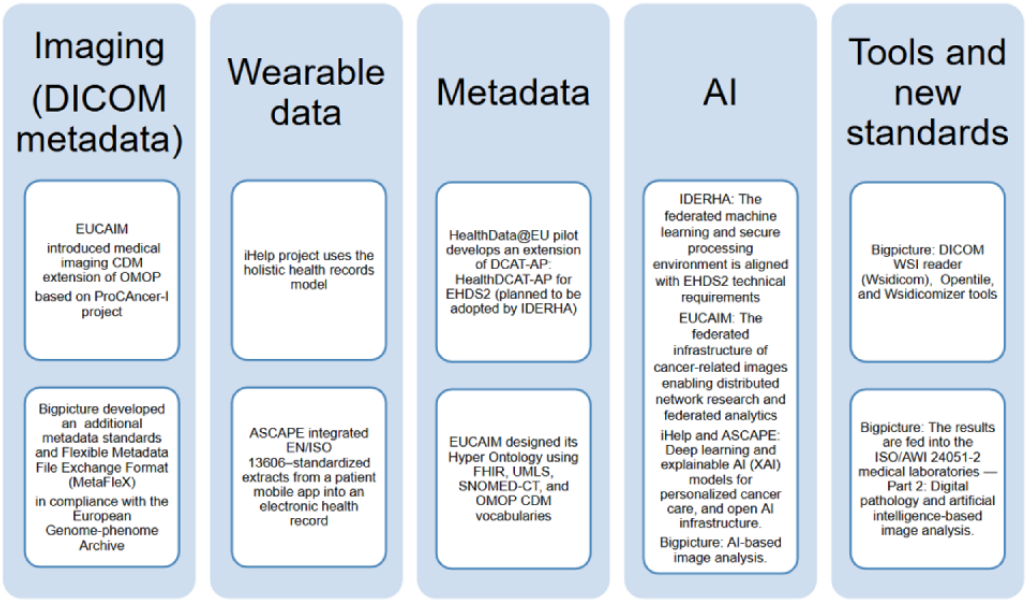

It is interesting to notice that the ETSI SmartM2M SAREF extensions for health and aging well (SAREF4EHAW) and wearables (SAREF4WEAR) are not known and not used on those six projects that we compared, even projects using IoT technologies with sensors, devices.

Key reasons for this gap include insufficient promotion, perceived relevance, integration challenges, and the presence of competitive standards. ETSI should enhance outreach, showcase successful implementations, collaborate with key projects, and simplify integration processes to improve adoption.

The other aspects, AI and federated learning, security and privacy, and data quality are covered by several approaches and standards. The adoption of federated standards and AI in cancer research projects across Europe represents a shift toward collaborative yet privacy-preserving medical research. These initiatives are essential for creating robust, scalable, and interpretable AI models that can significantly advance early detection, treatment, and overall patient care in oncology. The projects leverage various AI standards to achieve these goals. Some of the key AI standards being used have been included in Table S7 in Multimedia Appendix 1. The Regulation (EU) 2018/1725, also known as GDPR, covers the fundamental requirements for access, sharing, and storage of personal data, including health data. The usage of raw health data with personal details included for research purposes remains controversial in terms of consent and audit logging. Furthermore, there is a need for standardized security and data-sharing models. Finally, conformance guidelines to ensure compliance with regulations (eg, AI Act) for manufacturers developing medical products are still needed.

Creating Synergy Among the Involved Projects

We remarked on similarities and differences as we navigated the implementation of various standards among the projects. Despite the projects' varied focuses, each project prioritizes interoperability, relying on various standards. This collective effort promotes knowledge exchange and innovation and fosters a digitally unified health care ecosystem for EHDS2. We also observed a trend towards common standards like HL7 FHIR and DICOM. While projects may vary in domain-specific standards and implementation nuances, leveraging established standards offers numerous advantages. This includes improved interoperability, streamlined development processes, scalability, and knowledge sharing.

Accordingly, intensive discussions accelerated the creation of the synergy among the groups that will be introduced in the Medical Informatics Europe 2024 conference [21]. In addition, both the HealthData@EU pilot and the IDERHA project were addressed as EHDS use cases in the white paper on "IoT/Edge Computing and Health Data and Data Spaces" published by the Alliance for IoT and Edge Computing Innovation [51], and both projects also shared expertise on informing EHDS2 development. Experience from completed projects working with standards helps ongoing projects to decide on which standards to focus on, learn from their limitations, etc. For example, BigPicture designed a new standard ISO/AWI 24051-2.

XSL•FO
RenderX

Conclusion and Outlook

Standardization and legalization are the main pillars of EHDS. The HSbooster.eu initiative enabled intensive analysis of these aspects among six data-driven EU projects focusing on cancer. We aimed to create synergy among these projects to share the lessons learned in standards and legal implementation, sustainability of these projects, and harmonization with these with the EHDS2 implementation. Furthermore, we are involved in the development of standards such as ISO/IEC 21823-3:2021 "IoT - Interoperability for IoT Systems - Part 3 Semantic interoperability" (as editors), and since there is no ISO Standard for digital pathology and AI-based analysis of (whole slide) images for medical diagnosis, BigPicture is involved in ISO/PWI 24051-2 "Medical laboratories—Part 2: Digital pathology and artificial intelligence (AI)-based image analysis."

Future work requires integration of results from TEHDAS 1&2 and HealthData@EU pilot as well as other EHDS supporting projects, such as QUANTUM [52], which is seeking to develop the EHDS data quality and utility label. Additionally, enhancements to the EHDS2 interoperability framework template are needed, with more standards covering health ontologies, wearables, personal devices, data quality, as well as the usage of FAIR principles. The template can help researchers choose appropriate standards for their projects, thereby reducing time and effort in standard selection and implementation and improving EHDS2 implementation. Moreover, the template will be used in designing and creating an interactive toolkit for the EHDS2 interoperability framework. In this way, the existing and future projects can ensure alignment with governance and interoperability requirements for EHDS2.

EHDS has recently gained additional relevance since the European AI Act that has been approved by the European Parliament by resolution on March 13, 2024, explicitly refers to the EHDS as needing to "facilitate non-discriminatory access to health data and the training of AI algorithms on those data sets, in a privacy-preserving, secure, timely, transparent and trustworthy manner, and with an appropriate institutional governance."

Acknowledgments

AG declares funding from HSBOOSTER 101058391, StandICT.eu 2026 101091933. The IDERHA (Integration of Heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance) project is supported by the Innovative Health Initiative (IHI) Joint Undertaking (JU) under grant agreement 101112135. The JU receives support from the European Union's Horizon Europe research and innovation program and life science industries represented by COCIR, EFPIA/Vaccines Europe, EuropaBio, and MedTech Europe. The BigPicture project has received funding from the Innovative Medicines Initiative 2 JU under grant agreement 945358. This JU receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. The iHelp project received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 101017441. The ASCAPE (Artificial Intelligence Supporting Cancer Patients Across Europe) project received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 875351. The EUCAIM (European Cancer Imaging Initiative) project is cofunded by the European Union under grant agreement 101100633. The HealthData@EU pilot is cofunded by the European Union.

Authors' Contributions

AG and RH contributed to conceptualization, methodology, and writing – original draft. SA and PG assisted with methodology and writing – original draft. GM, RvN, KZ, IEN, GD, SN, LM-B, PM, SD, SA, MJ, IA, EGA, PH, IB, and AB contributed to validation and writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed results from analysis of six involved projects. Comparison of the usage of standards: artificial intelligence data quality.
[DOCX File , 51 KB - jmir_v27i1e66273_app1.docx]

References

1. Cancer. World Health Organization. 2025. URL: https://www.who.int/europe/health-topics/cancer [accessed 2025-02-18]
2. Communication from the Commission to the European Parliament and the Council. BRILL. 2012. URL: https://primarysources.brillonline.com/browse/human-rights-documents-online/communication-from-the-commission-to-the-european-parliament-and-the-council;hrdhrd46790058 [accessed 2023-02-01]
3. Report on workshops - landscaping data driven projects and initiatives in the cancer field. European Commission. 2024. URL: https://digital-strategy.ec.europa.eu/en/library/report-workshops-landscaping-data-driven-projects-and-initiatives-cancer-field [accessed 2024-05-23]

XSL•FO
RenderX

4. Kondylakis H, Kalokyri V, Sfakianakis S, Marias K, Tsiknakis M, Jimenez-Pastor A, et al. Data infrastructures for AI in medical imaging: a report on the experiences of five EU projects. Eur Radiol Exp 2023;7(1):20 [FREE Full text] [doi: 10.1186/s41747-023-00336-x] [Medline: 37150779]
5. Second joint action towards the European Health Data Space – TEHDAS2. TEHDAS2. 2024. URL: https://tehdas.eu/project/ [accessed 2025-02-18]
6. European Health Data Space. European Commission. 2024. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en [accessed 2025-02-18]
7. Palojoki S, Lehtonen L, Vuokko R. Semantic interoperability of electronic health records: systematic review of alternative approaches for enhancing patient information availability. JMIR Med Inform 2024;12:e53535 [FREE Full text] [doi: 10.2196/53535] [Medline: 38686541]
8. DCAT application profile for data portals in Europe. European Comission. 2015. URL: https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe [accessed 2015-10-26]
9. The New European Interoperability Framework. European Commission. URL: https://ec.europa.eu/isa2/eif_en/ [accessed 2017-03-23]
10. Kouroubali A, Katehakis DG. Policy and strategy for interoperability of digital health in Europe. Stud Health Technol Inform 2022;290:897-901. [doi: 10.3233/SHTI220209] [Medline: 35673148]
11. Stellmach C, Muzoora MR, Thun S. Digitalization of health data: interoperability of the proposed European health data space. Stud Health Technol Inform 2022;298:132-136. [doi: 10.3233/SHTI220922] [Medline: 36073471]
12. Torab-Miandoab A, Samad-Soltani T, Jodati A, Rezaei-Hachesu P. Interoperability of heterogeneous health information systems: a systematic literature review. BMC Med Inform Decis Mak 2023;23(1):18 [FREE Full text] [doi: 10.1186/s12911-023-02115-5] [Medline: 36694161]
13. Moreno-Conde J, Salas-Fernandez S, Moreno-Conde A. MedicalForms: integrated management of semantics for electronic health record systems and research platforms. Appl Sci 2022;12(9):4322. [doi: 10.3390/app12094322]
14. Options for governance models for the European Health Data Space. TEHDAS. 2023. URL: https://tehdas.eu/app/uploads/2023/01/tehdas-options-for-governance-models-for-the-european-health-data-space.pdf [accessed 2023-01-17]
15. Recommendations to enhance interoperability within HealthData@EU- a framework for semantic, technical and organisational interoperability. TEHDAS. 2022. URL: https://tehdas.eu/tehdas1/app/uploads/2023/10/tehdas-recommendations-to-enhance-interoperability.pdf [accessed 2022-12-21]
16. The role of harmonised standards as tools for AI act compliance. DLA Piper. URL: https://www.dlapiper.com/es-pr/insights/publications/2024/01/the-role-of-harmonised-standards-as-tools-for-ai-act-compliance [accessed 2025-02-18]
17. Elevating data quality a paradigm shift for data spaces and AI needs. Big Data Value Association (BDVA). URL: https://bdva.eu/news/new-paper-?-elevating-data-quality-a-paradigm-shift-for-data-spaces-and-ai-needs/ [accessed 2024-05-29]
18. European Health Data Space Data Quality Framework. TEHDAS. 2022. URL: https://tehdas.eu/tehdas1/app/uploads/2022/05/tehdas-european-health-data-space-data-quality-framework-2022-05-18.pdf [accessed 2022-05-18]
19. Recommendations on a Data Quality Framework for the European Health Data Space for secondary use. TEHDAS. URL: https://tehdas.eu/tehdas1/app/uploads/2023/09/tehdas-recommendations-on-a-data-quality-framework.pdf [accessed 2023-09-26]
20. Data Quality Framework for EU medicines regulation. European Medicines Agency (EMA). 2023. URL: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/data-quality-framework-eu-medicines-regulation_en.pdf [accessed 2023-10-30]
21. Gyrard A, Gribbon P, Hussein R, Abedian S, Bonmati LM, Cabornero GL, et al. Synergies among health data projects with cancer use cases based on health standards. Stud Health Technol Inform 2024;316:1292-1296. [doi: 10.3233/SHTI240649] [Medline: 39176618]
22. IDERHA. URL: https://www.iderha.org/integrating-health-data [accessed 2025-02-18]
23. A central repository of digital pathology slides to boost the development of artificial intelligence. BigPicture. URL: https://bigpicture.eu/ [accessed 2025-02-18]
24. Cancer Image Europe. URL: https://cancerimage.eu/ [accessed 2023-09-29]
25. iHelp. URL: https://ihelp-project.eu/ [accessed 2025-02-18]
26. ASCAPE. URL: https://www.ascape-project.eu/ [accessed 2025-02-18]
27. EHDS2 Pilot. URL: https://ehds2pilot.eu/ [accessed 2025-02-18]
28. FHIR for FAIR - FHIR Implementation Guide. HL7 FHIR. URL: http://build.fhir.org/ig/HL7/fhir-for-fair/terminology.html [accessed 2024-01-05]
29. HL7 FHIR. URL: https://fhir.org/ [accessed 2025-02-18]
30. DICOM. URL: https://www.dicomstandard.org [accessed 2025-02-18]
31. Observational Health Data Sciences and Informatics (OHDSI). 2024. URL: https://www.ohdsi.org/ [accessed 2021-01-06]
32. International Organization for Standardization (ISO). URL: https://www.iso.org/home.html [accessed 2025-01-22]
33. SAREF4EHAW: an extension of SAREF for eHealth Ageing Well domain. ETSI. URL: https://saref.etsi.org/saref4ehaw/v1.1.1/ [accessed 2020-02-01]
34. SNOMED International. URL: https://www.snomed.org [accessed 2024-12-16]

XSL•FO
RenderX

35. LOINC. URL: https://loinc.org/ [accessed 2025-02-18]
36. International Statistical Classification of Diseases and Related Health Problems (ICD). World Health Organization. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2025-02-18]
37. Park C, You SC, Jeon H, Jeong CW, Choi JW, Park RW. Development and validation of the radiology common data model (R-CDM) for the international standardization of medical imaging data. Yonsei Med J 2022;63(Suppl):S74-S83 [FREE Full text] [doi: 10.3349/ymj.2022.63.S74] [Medline: 35040608]
38. Kalokyri V, Kondylakis H, Sfakianakis S, Nikiforaki K, Karatzanis I, Mazzetti S, et al. MI-common data model: extending observational medical outcomes partnership-common data model (OMOP-CDM) for registering medical imaging metadata and subsequent curation processes. JCO Clin Cancer Inform 2023;7:e2300101. [doi: 10.1200/CCI.23.00101] [Medline: 38061012]
39. Integrating the Healthcare Enterprise (IHE Europe). URL: https://www.ihe-europe.net/ [accessed 2025-02-18]
40. D4.03 report on unified open digital slide and annotation format specification. BigPicture. URL: https://bigpicture.eu/sites/default/files/2023-04/945358-BIGPICTURE_D4.03_Report%20on%20unified%20open%20digital%20slide%20and%20annotation%20format%20specification.pdf [accessed 2025-02-18]
41. imi-bigpicture / wsidicom. GitHub. URL: https://github.com/imi-bigpicture/wsidicom [accessed 2023-12-06]
42. imi-bigpicture / wsidicomizer. GitHub. URL: https://github.com/imi-bigpicture/wsidicomizer [accessed 2023-12-06]
43. Report on the landscape analysis of available metadata catalogues and the metadata standards in use. EHDS. URL: https://ehds2pilot.eu/available_results/report-on-the-landscape-analysis-of-available-metadata-catalogues-and-the-metadata-standards-in-use/ [accessed 2023-03-30]
44. Hussein R, Balaur I, Burmann A, Ćwiek-Kupczyńska H, Gadiya Y, Ghosh S, et al. Getting ready for the European health data space (EHDS): IDERHA's plan to align with the latest EHDS requirements for the secondary use of health data. Open Res Eur 2024;4:160. [doi: 10.12688/openreseurope.18179.1] [Medline: 39185338]
45. EUCAIM's HyperOntology_v0.2beta. Zenodo. URL: https://zenodo.org/records/11109765 [accessed 2024-05-03]
46. El Saddik A. Digital twins: the convergence of multimedia technologies. IEEE MultiMedia 2018;25(2):87-92. [doi: 10.1109/mmul.2018.023121167]
47. Manias G, Op DAH, Azqueta A, Burgos D, Capocchiano ND, Crespo BL, et al. iHELP: personalised health monitoring and decision support based on artificial intelligence and holistic health records. 2021 Presented at: IEEE Symposium on Computers and Communications (ISCC); September 08-10, 2021; Athens, Greece. [doi: 10.1109/iscc53001.2021.9631475]
48. Manias G, Azqueta-Alzúaz A, Dalianis A, Griffiths J, Kalogerini M, Kostopoulou K, et al. Advanced data processing of pancreatic cancer data integrating ontologies and machine learning techniques to create holistic health records. Sensors (Basel) 2024;24(6):1739 [FREE Full text] [doi: 10.3390/s24061739] [Medline: 38544003]
49. Frid S, Fuentes Expósito MA, Grau-Corral I, Amat-Fernandez C, Muñoz Mateu M, Pastor Duran X, et al. Successful integration of EN/ISO 13606-standardized extracts from a patient mobile app into an electronic health record: description of a methodology. JMIR Med Inform 2022;10(10):e40344 [FREE Full text] [doi: 10.2196/40344] [Medline: 36222792]
50. D5.05 - Report on landscape of guidelines, standards and regulatory requirements relevant for digital pathology, clinical. BigPicture. URL: https://bigpicture.eu/sites/default/files/2023-04/945358-BIGPICTURE_D5.05%20-%20Report%20on%20landscape%20of%20guidelines%2C%20clinical_final.pdf [accessed 2025-02-18]
51. AIOTI White Paper IoT/Edge Computing and Health Data and Data Spaces. AIOTI. 2024. URL: https://aioti.eu/aioti-white-paper-iot-edge-computing-and-health-data-and-data-spaces [accessed 2024-08-03]
52. QUANTUM. URL: https://quantumproject.eu/ [accessed 2024-03-01]

Abbreviations

AI: artificial intelligence
AI4HI: Artificial Intelligence for Health Imaging
ASCAPE: Artificial Intelligence Supporting Cancer Patients Across Europe
BDVA: Big Data Value Association
CDM: common data model
CMMS: Common Mandatory Metadata Structure
DCAT: Data Catalog Vocabulary
DCAT-AP: DCAT Application Profile for Data Portals in Europe
DICOM: Digital Imaging and Communications in Medicine
DQF: data quality framework
EBCP: Europe Beating Cancer Plan
EC: European Commission
EHDS: European Health Data Space
EHDS2: European Health Data Space for the secondary use of data
EIF: European Interoperability Framework
EMA: European Medicine Agency

XSL•FO
RenderX

EGA: European Genome-Phenome Archive
EU: European Union
EUCAIM: European Cancer Imaging Initiative
FAIR: Findability, Accessibility, Interoperability, and Reusability
FHIR: Fast Healthcare Interoperability Resource
GDPR: General Data Protection Regulation
HL7: Health Level 7
IDERHA: Integration of Heterogeneous Data and Evidence Towards Regulatory and Health Technology Assessments Acceptance
IEC: International Electrotechnical Commission
INHAND: International Harmonization of Nomenclature and Diagnostic Criteria
IoT: Internet of Things
ISO: International Organization for Standardization
JA: joint action
OHDSI: Observational Health Data Sciences and Informatics
OMOP: Observational Medical Outcomes Partnership
PGHD: patient-generated health data
SAREF: Smart Applications Reference Ontology
SEND: Standardization for Exchange of Nonclinical Data
SNOMED CT: Systematized Medical Nomenclature for Medicine–Clinical Terminology
TC: Technical Committee
TEHDAS: Towards the European Health Data Space
WHO: World Health Organization
WSI: whole-slide images
W3C: World Wide Web Consortium

Edited by A Coristine; submitted 09.09.24; peer-reviewed by A Billis, Z Hou, FA Causio; comments to author 18.11.24; revised version received 10.12.24; accepted 31.01.25; published 24.03.25.

Please cite as:
Gyrard A, Abedian S, Gribbon P, Manias G, van Nuland R, Zatloukal K, Nicolae IE, Danciu G, Nechifor S, Marti-Bonmati L, Mallol P, Dalmiani S, Autexier S, Jendrossek M, Avramidis I, Garcia Alvarez E, Holub P, Blanquer I, Boden A, Hussein R
Lessons Learned From European Health Data Projects With Cancer Use Cases: Implementation of Health Standards and Internet of Things Semantic Interoperability
J Med Internet Res 2025;27:e66273
URL: https://www.jmir.org/2025/1/e66273
doi:10.2196/66273
PMID:40126534

©Amelie Gyrard, Somayeh Abedian, Philip Gribbon, George Manias, Rick van Nuland, Kurt Zatloukal, Irina Emilia Nicolae, Gabriel Danciu, Septimiu Nechifor, Luis Marti-Bonmati, Pedro Mallol, Stefano Dalmiani, Serge Autexier, Mario Jendrossek, Ioannis Avramidis, Eva Garcia Alvarez, Petr Holub, Ignacio Blanquer, Anna Boden, Rada Hussein. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Enhancing the Innovation Ecosystem: Overcoming Challenges to Introducing Information-Driven Technologies in Health Care

Julie Reed[1], PhD; Petra Svedberg[1], PhD; Jens Nygren[1], PhD

School of Health and Welfare, Halmstad University, Halmstad, Sweden

Corresponding Author:
Jens Nygren, PhD
School of Health and Welfare
Halmstad University
Box 823
Halmstad, 30118
Sweden
Phone: 46 035167100
Email: jens.nygren@hh.se

Abstract

As health care demands rise and resources remain constrained, optimizing health care systems has become critical. Information-driven technologies, such as data analytics and artificial intelligence (AI), offer significant potential to inform and enhance health care delivery at various levels. However, a persistent gap exists between the promise of these technologies and their implementation in routine practice. In this paper, we propose that fragmentation of the innovation ecosystem is behind the failure of new information-driven technologies to be taken up into practice and that these goals can be achieved by increasing the cohesion of the ecosystem. Drawing on our experiences and published literature, we explore five challenges that underlie current ecosystem fragmentation: (1) technology developers often focus narrowly on perfecting the technical specifications of products without sufficiently considering the broader ecosystem in which these innovations will operate; (2) lessons from academic studies on technology implementation are underused, and existing knowledge is not being built upon; (3) the perspectives of healthcare professionals and organizations are frequently overlooked, resulting in misalignment between technology developments and health care needs; (4) ecosystem members lack incentives to collaborate, leading to strong individual efforts but collective ecosystem failure; and (5) investment in enhancing cohesion between ecosystem members is insufficient, with limited recognition of the time and effort required to build effective collaborations. To address these challenges, we propose a series of recommendations: adopting a wide-lens perspective on the ecosystem; developing a shared-value proposition; fostering ecosystem leadership; and promoting local ownership of ecosystem investigation and enhancement. We conclude by proposing practical steps for ecosystem members to self-assess, diagnose, and improve collaboration and knowledge sharing. The recommendations presented in this paper are intended to be broadly applicable across various types of innovation and improvement efforts in diverse ecosystems.

(J Med Internet Res 2025;27:e56836) doi:10.2196/56836

KEYWORDS

artificial intelligence; ecosystem; health care; implementation; technology adoption; improvement; complex-systems

Introduction

Focus and Aim of Viewpoint

Despite a sizeable increase in the development of new information-driven technologies in health care their uptake and impact in routine practice remains limited. This viewpoint argues that this failure stems from fragmentation of the innovation ecosystem which is driving a lack of synergist working between ecosystem members. This includes all stakeholders who have a role to play in conceptualizing, designing, developing, studying, implementing, regulating, and using such technologies. We propose that in order to realize the potential of new technologies in health care we need to improve how the innovation ecosystem is working. This is especially true as we tackle bigger and more impactful innovations that require a greater degree of coordinated action.

Drawing on our experiences and published literature we explore the challenge of current ecosystem fragmentation and the underlying system drivers that make collaboration and synergistic working so difficult to achieve. We make a series of recommendations as to how these challenges can be addressed

XSL•FO
RenderX

including the need to adopt a wide-lens perspective on the ecosystem and develop a shared-value proposition, a call for ecosystem leadership, and the need for local ownership of ecosystem investigation and enhancement. We finish by proposing practical steps ecosystem members can take to self-assess, diagnose, and enhance how they are working and learning together.

The recommendations presented in this paper are intended to be broadly applicable across various types of innovation and improvement efforts in diverse ecosystems.

Background

The current strain on the health care system highlights the necessity for improving quality of care, with a specific emphasis on promoting effectiveness and efficiency in a manner that best serves the needs of patients and populations while promoting the well-being of health care staff members. It is therefore imperative to address these challenges and implement strategies that not only enhance health care delivery but also ensure the sustainability of health care systems.

Within this backdrop, the advent of cutting-edge technologies, encompassing advancements in data analytics, computational capabilities, and artificial intelligence (AI), provides an opportunity to inform the provision of health care [1,2]. Information-driven technologies in health care encompass systems and tools that rely on the collection, integration, and analysis of vast volumes of data to support and enhance clinical and operational decision-making [1,2]. These technologies are fundamentally dependent on the availability of high-quality, diverse, and interoperable datasets, as their effectiveness is directly tied to the comprehensiveness and accuracy of the underlying information. Central to their functionality are advanced analytics, particularly AI techniques, which enable the processing of complex and multidimensional datasets to identify patterns, generate predictions, and provide actionable insights. For example, AI can detect subtle anomalies in diagnostic imaging, predict disease progression based on longitudinal patient data, or optimize care pathways by analyzing historical treatment outcomes. By automating these sophisticated analyses, information-driven technologies could facilitate precision medicine, improve the allocation of health care resources, and reduce inefficiencies. However, despite the potential benefits of integrating information-driven technologies into health care systems, there is limited evidence of their successful uptake into routine health care practice [3,4].

Drawing inspiration from Ron Adner's book "The Wide Lens" we propose that the limited adoption of information-driven technologies in health care stems from technology developers focusing too narrowly on their individual products without considering the entire ecosystem needed for success. Adner [5,6] argues that any innovator requires collaboration with other co-innovators, who develop complementary solutions to make the focal innovation viable, and with adoption chain partners who must adopt or support the innovation before it can reach the end user. This wide-lens perspective emphasizes that the success of an innovation depends not only on the product itself but also on external groups. To minimize the risk of failure, innovators need to identify and align with the interests of these ecosystem players and proactively manage their dependencies.

For example, Amazon's success with the Kindle e-reader contrasts with Sony's earlier failure to launch a similar product. Adner explains that Sony's focus on perfecting the device's technical specifications overlooked the broader ecosystem requirements. Without securing coinnovators like publishers to provide e-books, a smooth platform for adoption chain partners such as booksellers to distribute them, and a seamless experience for buyers (the adopters), the product could not deliver the full value customers needed, leading to its failure. To succeed, Amazon took on a key leadership role by motivating and coordinating contributions from a diverse range of ecosystem members, ensuring that all necessary innovations and activities were in place to support their new product.

We believe that such ecosystem leadership and coordination is currently lacking in health care innovation, and the failure to consider how innovations work within the broader health care ecosystem is a key reason why new information-driven technologies struggle to gain traction.

In this viewpoint, we explore challenges of fragmentation that are specific to the health care innovation ecosystem and make a series of recommendations to drive ecosystem cohesion and success.

The Challenge: Ecosystem Fragmentation

Overview

To consider the fragmentation of the health care innovation ecosystem we need to explore the relationships between different ecosystem members. The wider ecosystem of groups and organizations contributing to the health care innovation ecosystem is vast, comprising of diverse stakeholders and actors from different organizations, as well as professionals and experts from a broad spectrum of disciplines [7]. This includes health care organizations responsible for planning and delivery of care, industry organizations engaged in technology development and commercialization, academics who can contribute expert knowledge and methodological insights, and community, charitable, and voluntary organizations who play a variety of support roles to the well-being of patients and the wider population, and government and regulatory organizations who legislate and regulate activities [8,9].

For the purpose of this viewpoint, we have chosen to focus on the relationships between three main groups of ecosystem members to illustrate the challenges of ecosystem fragmentation: (1) the innovators who develop new technologies, (2) researchers who study the use and implementation of technologies, and (3) health care professionals who are the intended adopters and users of the technologies in practice. We recognize that in practice exploration of ecosystem relationships would require expansion to include all relevant members (recommendations).

Challenge #1: Not Looking Over the Fence—a Narrow Focus on Innovation Pipelines

Technology developers often adopt a narrow lens, focusing exclusively on their own innovation pipelines while neglecting

XSL•FO
RenderX

the broader ecosystem in which their innovations must operate. This tendency is particularly evident in the development of information-driven technologies in health care, where the primary emphasis is often on data availability and technological functionality [10]. Critical issues such as usability, implementation in routine care, and the interdependence with existing health care practices and processes are frequently treated as afterthoughts, only considered after the proof-of-concept stages [3,4].

Traditional conceptualizations of the innovation process exacerbate this narrow focus. The linear innovation pipeline model (Figure 1) [11] assumes a straightforward handoff between developers and adopters after invention and evaluation, or proof-of-concept, as it is often termed in technology development. This perspective perpetuates the misconception that bridging the gap between vision and practice, often referred to as "the last mile" [12], is primarily the adopters' responsibility. However, such an approach overlooks the critical dependencies and coinnovations required for successful integration into health care systems [13].

This narrow focus on technological development often results in solutions that fail to meet the real-world needs of health care staff members or patients. Technologies that ignore the routines, constraints, and practicalities of health care delivery risk being fundamentally misaligned with the requirements of adopters. Even well-designed technologies can falter without understanding the broader ecosystem changes needed to support their use.

From an innovator's perspective, this pipeline model may seem logical, it allows a focus on technical excellence within one's domain of expertise. However, it risks creating solutions based on overly simplistic notions of how health care systems function, ignoring the complexities and constraints of real-world applications. For example, health care professionals often face an overabundance of information, not a lack of it. While many information-driven technologies aim to provide more precise or accurate data, a more pressing challenge is navigating the "noise" of overwhelming information, knowing how to act on it, and having the resources and authority to do so. Thus, technology developers need to spend time understanding these challenges or the broader concerns of other ecosystem members.

Figure 1. Models for conceptualizing the innovation process: Traditional linear model of innovation-led development (left) needs to be reconceptualized as part of a more holistic cycle for improvement (right), which focuses on achieving system aims and goals and increases requirements for collaborative working between innovators and practitioners.

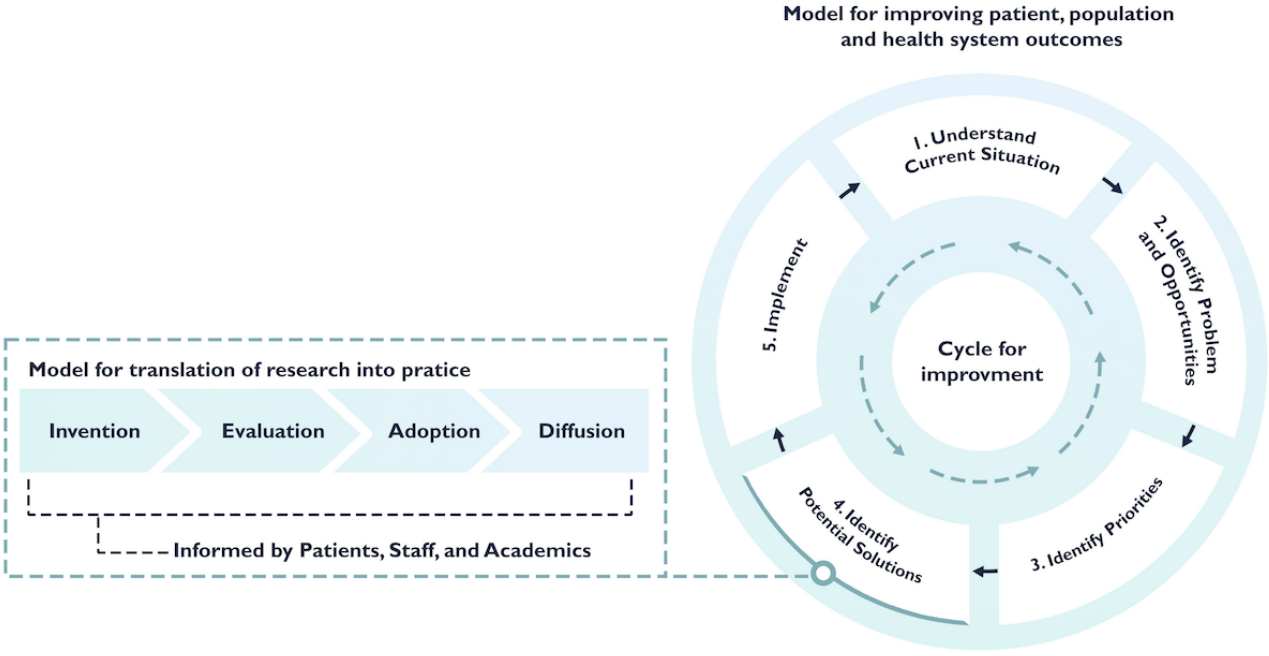

Challenge #2: Not Building on What we Know—Lessons From the Study of Technology Use and Implementation Are Underused

The limitations of the "narrow lens" of product development and the failure to consider the wider context of technology deployment are well-documented. For decades, health care research has highlighted the challenges of diffusing and adopting innovations in practice (eg, NASSS [Non-adoption, abandonment, scale-up, spread, and sustainability] framework) [14,15]. Studies of technology use, such as Normalization Process Theory [16] and sociomaterial approaches [17] have repeatedly emphasized the importance of understanding the interaction between technology, its users, and the wider context of deployment, echoing Adner's findings [5]. While much of this research is not specific to information-driven technologies, its insights into the enablers and barriers to innovation uptake are both relevant and generalizable.

Despite this extensive body of knowledge, it has not significantly influenced the behaviors and practices of technology developers or their collaborations with other ecosystem members. In most industries, such research would be a cornerstone of product development; yet in health care,

XSL•FO
RenderX

even with substantial investment in research, its findings often fail to be acted upon.

Awareness of these studies is low among tech developers. For example, a recent systematic review found that implementation theories were rarely applied in the or implementation of AI technologies for health care [3]. Even within academic disciplines that study health care technology and innovation, such as implementation science and sociomaterial studies, there is recognition of the limited practical impact of their findings [18-20].

The lack of uptake of such knowledge is not solely the fault of technology developers. Multiple complex factors contribute to this fragmentation. Studies often report qualitative findings, embedded in psychological, sociological, or organizational perspectives, and presented as nuanced theories written in discipline-specific language. This can feel alien and inaccessible to those outside these academic fields, including technology developers. The diversity of disciplines conducting these studies, each with distinct priorities, perspectives, and terminologies, further complicates understanding. Even for experts in health care research, navigating these different theoretical perspectives can be daunting.

In addition, many studies stop at describing problems without providing actionable guidance for solutions. While technology developers may focus narrowly on their products, researchers often focus narrowly on publishing theories, with less attention to their practical application. This leaves developers without clear, actionable steps to implement research findings in their work. Bridging the gap between research and practice requires collaborative approaches that translate theoretical insights into practical guidance, tailored to the needs of technology developers and other ecosystem members.

Challenge #3: Not Coordinating with Health Care Needs—Overlooking the Perspectives of Health Care Professionals and Organizations

Research on health care innovation consistently highlights that the success of new technologies hinges on their fit with the context in which they are introduced [21]. Yet, the perspectives of health care professionals are often insufficiently considered during development.

Health care professionals play two critical roles in this process. As adoption-chain partners, they ensure innovations integrate into practice and are used appropriately. For example, a new technology would need to be integrated into existing workflows, staff members trained on how to use it, and maintenance support systems in place. Without this work to actively support technology adoption the innovation will not be successfully implemented. As coinnovators, they create complementary innovations necessary for technologies to achieve their full value. For example, a new technology may be very powerful in identifying patients at risk of having a particular health condition, but would require coinnovation to identify how to reach out to these patients to notify them of this risk, and how to treat them. Without these coinnovations, the technology would not be realizing its value to the health care ecosystem.

Both adoption-chain partners and coinnovators should be engaged from the outset of any technology development to work collaboratively to align incentives, address technical and operational challenges, and mitigate potential conflicts. While there is some recognition of the need to engage adoption-chain partners, this process is often superficial and assumes that health care professionals can easily provide the necessary insights. In reality, this assumption is flawed. Health care systems are extraordinarily complex, with knowledge distributed across many individuals and much of it tacit. Consequently, developers and researchers often base their work on incomplete understandings of “work as imagined” rather than “work as it is.” Obtaining reliable insights into workflows, practices, and behaviors requires time, expertise, and resources that neither technology developers nor health care professionals typically have. This disconnect often results in poorly defined briefs and high-specification products that fail to meet the needs of the system, frustrating both developers and users.

Coinnovation is rarely prioritized. Many developers focus narrowly on developing accurate, effective tools, assuming these will inherently provide value. However, the true value of any technology lies not in the information it provides but in its ability to improve clinical practices and patient outcomes. Achieving this requires additional innovations, such as modifying dependent processes and practices, and changing professional behaviors, tasks that are often left to health care organizations [22]. These organizations often lack the resources, expertise, or capacity to manage such work, particularly given the competing demands of delivering care and supporting staff members well-being in under-resourced and high-stress environments. For them, technology is not the focal point but a potential enabler of better care.

Adding to these challenges, frontline teams are frequently overwhelmed by the sheer volume and variety of new interventions they are expected to adopt, often with little consideration of how these initiatives interact or how unresolved challenges might be exacerbated by new technologies [19]. At present, there is insufficient coordination within the health care ecosystem to address these challenges. Innovators often view technology as the central actor, whereas health care teams see it as a supportive tool for achieving their goals. This misalignment leads to tension and hinders the effective uptake of new innovations.

Challenge #4: Not Incentivizing People to Work Together—Individually People Are Doing Good Jobs, But Collectively the Ecosystem Is Failing

In the current health innovation ecosystem, individuals excel in their specific roles: technology developers create apps and algorithms, health researchers produce theories and publications, and health care professionals deliver services. However, the ecosystem often fails collectively because members are incentivized to focus on predetermined end points within their disciplines rather than working collaboratively across boundaries. No one is tasked with bridging the gaps between these roles or ensuring that end products are effectively integrated.

XSL•FO
RenderX

This fragmentation fosters a false sense of completion, where apps are handed off to health care teams or theories to developers without accountability for their practical impact. Crucial feedback loops, such as whether the app or theory is used and helpful, are neglected. Without this cross-disciplinary exchange, the ecosystem remains siloed, disconnected, and unable to adapt to the needs and constraints of its members.

A notable example of ecosystem failure due to the fragmentation of responsibilities, where tasks are divided into manageable parts but lack a comprehensive overview and accountability, is the collapse of the Minneapolis I-35W Bridge on August 1, 2007. Investigation into the disaster identified that all individual contractors including the original design firm, transportation officials, inspection teams, and maintenance crew had fully met their individual responsibilities, and as such could not be blamed for the disaster. However, collectively they had failed due to the fragmentation of responsibilities between multiple entities which resulted in a combination of design flaws, a lack of adaptation to the incremental increases in the bridge's weight over time, and decades of inadequate maintenance. In health care innovation, fragmentation results in frustration: developers lament the lack of adoption [23,24], researchers feel their insights are underused [25], and clinicians face an unrelenting influx of innovations that fail to address their real challenges [26]. Despite these frustrations, each group often blames other ecosystem members or assumes the problem lies outside their responsibility.

Career pathways exacerbate this issue. Academic success depends on publishing prestigious theories, not translating them into practice. Developers are rewarded for producing tools, not ensuring their usability. Clinicians focus on delivering care, not system-wide coordination. These narrow incentives discourage collaborative work and maintain the status quo. To step outside these norms often requires entrepreneurialism whilst risking personal career security.

The result is not that people are doing their jobs poorly but that they are working to end points too narrow to achieve broader ecosystem success. Effective ecosystem collaboration requires questioning assumptions, addressing gaps, and building strong relationships. Members must challenge briefs, clarify expectations, and engage in active feedback loops. A cohesive system depends on recognizing and valuing the interdependence of roles and the shared responsibility for collective success.

Challenge #5: Not Investing in Working Across Boundaries—Lack of Recognition of the Time and Effort Required

The diversity of ecosystem members is a strength, yet it also presents a significant, often unacknowledged challenge. Different scopes of interests, priorities, expertise, and practices can make it difficult for groups to share knowledge and work together effectively. It is much easier to learn from others within your own field, where shared language and mindsets foster understanding [27]. However, this specialized language can create barriers when trying to communicate with professionals from different fields [28]. At present little time and effort is invested, and few career pathway opportunities are provided to develop proficiency in navigating these boundaries or building shared understanding.

Communication efforts between ecosystem members often consist of one-way transfers of information, where one group assumes the role of "teaching" and the other as the "learner." These approaches fall short in 2 critical ways: they fail to consider the effort required to change behaviors and work processes, and they overlook the importance of 2-way, iterative learning [29]. Effective collaboration should focus on understanding the implications of each group's work for the others.

For instance, health research may highlight the need for technology developers to better understand day-to-day health care practices. However, developers typically focus on desktop work, with little direct engagement with frontline health care staff members. This shift requires them to navigate a vastly different environment, requiring new skills, communication styles, and an understanding of clinical challenges. It is not simply a matter of learning new information, it is about adjusting to a completely different way of working, which may be an unwelcome change for people who choose a technical profession.

For collaboration to succeed, stakeholders must reflect on how to adapt their practices, and the additional roles that may be required, to meet the needs of coinnovators and coadopters [30]. However, changing behaviors that are deeply embedded in professional norms is incredibly challenging. There is often an unstated assumption that new theories, models, and approaches can easily be adopted, but research from the fields of knowledge mobilization and translation challenges this view. People are not passive recipients of new ideas; [14] introducing new ways of working involves negotiation, compromise, and integration into established routines [31]. These changes can be disruptive, challenging identities, roles, and responsibilities [20]. Ecosystem members are highly skilled professionals with a large degree of autonomy to choose whether and how to engage with new ways of working [5,15]. Therefore, change requires as much attention to the human side of change, psychological, social, and political, as to the technical side [32].

Recommendations for How to Address the Challenge: Toward Ecosystem Cohesion

Overview

To address the fragmentation of health care ecosystems and systemic barriers to collaboration, we propose a set of actionable strategies to work toward ecosystem cohesion. These include adopting a wide-lens perspective to overcome silos, developing a shared ecosystem value proposition to align objectives, promoting ecosystem leadership to enhance collective accountability, encouraging local ownership of ecosystem assessment and improvement, and establishing structured pathways for continuous evaluation and enhancement.

While alternative approaches may also contribute to ecosystem cohesion, our recommendations are comprehensive and

XSL•FO
RenderX

well-founded, grounded in empirical research and practical insights. They systematically address core challenges such as fragmentation, misalignment, and inefficiency. By integrating strategic, structural, and leadership-driven solutions, our recommendations provide a robust and sustainable path forward.

Recommendation #1 Adopt a Wide-Lens Perspective

Building on Adner's research [5] we recommend that a *wide-lens perspective* is adopted. As Adner demonstrates it is not sufficient to focus on technological excellence at the expense of understanding the needs, experiences, and expectations of innovation adopters and other ecosystem members. There is a need to look over the fence to understand what else is going in the ecosystem that the success of any technological development is dependent on and to understand the implications of this learning for the initial conceptualization and design of new technologies, and the process of their development, testing, and deployment in partnership with other ecosystem members.

These insights suggest there are major limitations to the current "linear" product development mindset, and the metaphor of crossing the "last mile" [11,12]. It is not sufficient simply to attend the final handover of a product between the developers and the intended recipients. The whole ecosystem of coinnovators and coadopters needs to be actively engaged in synchronous and cooperative sense-making and problem-solving [33].

We propose that adopting an alternative process model, such as the cycle for improvement [34] can provide a helpful first step in adopting a wider ecosystem lens. (Figure 1) Rather than the linear progress implied by the innovation pipeline [11], the cycle for improvement emphasizes an iterative, cyclical process that is followed until the desired outcome or improvement has been achieved. The value of such a model is that it provides a holistic view of the entire improvement process and allows practices both upstream and downstream from the point of technology development to be taken into consideration. As such it starts to conceptualize a more holistic and cyclic view of the work required and is likely to appeal to a diverse range of ecosystem members who can see how their work orientates to different aspects of this model.

This process model is only one such example of how the ecosystem can start to be conceptualized in a more holistic manner and is intended as an embryonic starting point from which ecosystem explorations could build. Many other models could also provide valuable starting points such as sociomateriality, which focuses on the intersection of technology, work, and organization [35], and a patient work framework [36] which focuses on the interaction between end users and the innovation. These models operate on the principle that the innovation process is inherently social, integrating human, social, technical, and environmental components [37]. Such models emphasize the necessity of collaboration and coproduction at all stages of the process, rather than having discrete hand-offs between innovators and users that is implied in the linear pipeline model.

Recommendation #2: Develop a Shared Ecosystem Value Proposition

A major challenge to the health technology industry is that other people in the ecosystem do not share such a technology-centric perspective. Health care's primary goal is patient care, not the deployment of technology or profit generation, which can create tension and conflict between different stakeholder priorities and incentives.

Adner [5] recommends that the ecosystem should have a *shared value proposition* that reflects the shared aim that all ecosystem members are working toward [33]. The shared value proposition is broader than just the perspective of a focal innovation. It encompasses the entire ecosystem and reflects the interdependencies among all stakeholders focal innovators, coinnovators, and adoption chain partners. While a focal innovation can serve as the starting point for defining the ecosystem it recognizes that every player in the ecosystem must derive value from participating, and their success is interlinked. It reflects a broader focus of the collective ecosystem interests and ambitions, rather than a narrow view that prioritizes the focal innovator's benefits.

We propose that a shared value proposition could be the aspiration *to improve care with support from the introduction of information-driven technologies*. This statement reflects that any individual technology development is likely to be playing a small part of enabling a wider value proposition that will require many other coordinated actions and innovations to achieve.

The work required to achieve this shared value promotion can be conceptualized as 2 interrelated levels of operation. The first level focuses on *improving care with the support of information-driven technologies*, and the second level focuses on *enhancing the ecosystem's ability to improve care*. Successful intervention at level 1 (improving care with the support of information-driven technologies) relies on the quality of interaction and relationship between different stakeholders in working together to build a shared understanding of the challenges and potential solutions (level 2) [6].

Recommendation #3: a Call for Ecosystem Leadership

Ecosystem leaders play an important role in holding a vision for the future potential of the ecosystem, to look beyond individual cases and traditional silos—to consider the ecosystem's process for innovating and improving patient care works, and how it can be enhanced. This work will include coordinating and streamlining the interaction of ecosystem members, addressing any major gaps, and establishing clear practices, processes, and principles for overall work. It will be important to build on the knowledge and expertise of ecosystem members, while at the same time having the freedom to challenge established norms and ways of working for the benefit of the wider ecosystem. Emphasis should be placed on learning from previous efforts so that mistakes are not repeated, and to establish a forward looking and positive learning culture [38].

While there are substantial benefits to be gained from ecosystem actors working collaboratively to address challenges, there is rarely time or headspace to undertake this work of understanding

XSL•FO
RenderX

the ecosystem. As well as potentially contradicting short-term organizational goals, the investment in such boundary work can jeopardize individual career progression if the work conflicts with the expected and rewarded outputs and achievements (eg, financial investment secured each year, number of publications produced, or amount of services delivered). These incentives can be set within an organizational culture, but also often reflect expectations of the professional bodies and recruiters operating across an entire system nationally and globally. This can mean that to participate in boundary work individuals may have to take a personal risk in relation to their job tenure, financial reward, and prospects for internal promotion or career progression with other organizations [39].

Enhancing ecosystem working requires the engagement of system leaders who can remove or replace contrary incentives or create protected spaces where work to strengthen the ecosystem can happen without individuals being penalized [40]. Strategic engagement with system leaders will be important in creating protected spaces for ecosystem work to take place.

Recommendation #4: Local Ownership of Ecosystem Investigation and Enhancement

In identifying the current system deficiencies and opportunities for change we recognize that this work can only be done by people within the system; it is not something that can be conducted by outsiders and handed over to the ecosystem members to implement [33,41]. People within the system have the necessary insights to understand how things are currently done, what is working well, and what motivations and frustrations they experience. Their engagement is critical to designing and enacting any change ideas, drawing on their foresight of what is likely to work and what challenges may be encountered, in testing and feeding back on the benefits and limitations of any changes in practice, and in their persistence to address and overcome any difficulties experienced.

There is an extensive literature on the features of successful systems change, providing theory, values, principles, strategies, and practical guidance [33]. Much of this work has originated from diverse fields of application such as climate action, education, and social cohesion. While there are many different approaches, frameworks, models, and tools, there is a lot of conceptual and practical commonality between them, suggesting a convergence of thinking from research and practical experience of intervening in complex systems [33,42].

Findings suggest the importance of engaging ecosystem members in the process of investigating and diagnosing the current ecosystem. This supports stakeholders to understand and feel ownership for the identification of problems and areas requiring improvement, increasing their motivation and engagement to explore new ways of working. The engagement of ecosystem members is equally critical in designing, testing, and revising new ways of working to ensure their insights and specialist knowledge is incorporated and to enable the identification and resolution of emergent challenges and problems. No single person or group holds all the knowledge, skills, or competencies required to understand and address all aspects of the changes required. While collaborative working and ownership is view as critical it is recognized as highly challenging given the diverse perspectives and positions different groups and individuals hold, and the need for expert facilitation is recognized to enable constructive conversations between potentially contentious and conflicting views. Therefore, it is not sufficient to simply conduct research to make an objective assessment of the current situation, but necessary to use a participatory action research approach as a tool to support reflection, collective sense-making, and relationship building, in support of the stakeholders working toward system change [43].

Recommendation #5: a Pathway for Investigating and Enhancing Ecosystems

We propose that it is it helpful to envisage the work required to enhance ecosystems conducted in three stages:

- Stage 1: coinvestigation and diagnosis of current ecosystem functioning
- Stage 2: cocreation of strategies and action plans to address challenges and opportunities to enhance ecosystem functioning
- Stage 3: implementation, evaluation, the evolution of effective strategies and actions—embedding a permanent learning system to guide ecosystem enhancement.

Each of these 3 stages should be informed by bringing together an understanding of specific issues within the local ecosystem and existing knowledge, literature, and priori experiences on ecosystems and intervening in complex systems.

As well as investigating the ecosystem generally, which depending on the ecosystem could be a significant scale challenge—following specific projects developing and introducing information-driven technologies into health care can provide a powerful vehicle to understand ecosystem performance using more bounded and manageable cases (Table 1).

Before an ecosystem can be enhanced it is necessary to understand how the ecosystem is already working. This includes understanding who different members of the ecosystem are, the diversity of approaches and ways of working that they deploy, and how they currently interact with each other.

An important first step in the coinvestigation and diagnosis of the current ecosystem in stage one is to make the ecosystem explicit. This allows people to consider how the outputs of individual, group, and organizational efforts will be adopted by or influence others, and how they in turn are dependent on the efforts and innovations made by other people in the system to improve patient and population health. Research shows that for system change to take place it is necessary to understand what lies below the surface of how people behave [30]. This includes understanding the explicit aspects of the work being done (eg, policies, practices, or resource allocation), those that are semiexplicit (eg, understanding the relationships and power dynamics that inform ecosystem interactions) and implicit (eg, people habits of thought, deeply held beliefs, and assumptions that guide how they work and how they view others in the ecosystem) [44].

XSL•FO
RenderX

The cycle for improvement [24] can be used to guide research and facilitate sense-making about the innovation ecosystem: firstly, in considering what work is being done and why by different ecosystem members relates to the various the cycle stages (1) understand the current situation, (2) identify problems and opportunities, (3) prioritization, (4) identify or develop potential solutions and (5) implementation, and assessing the impact of changes as stage (1) is revisited at the next iteration of the cycle); and secondly in exploring different members and how they work together (who is doing the work and how do they interact together).

In starting to understand the diversity of the ecosystem members, their behaviors, and past experiences, insights will start to emerge as to what is working well, and where problems, frustrations, and challenges are being experienced for the ecosystem members individually and collectively. In some instances, these frustrations may be shared across ecosystem members, and lead to the identification of common goals and agreement to address them. On the other hand, frustrations may identify major gaps or conflicts in how people are working that are not so easily addressed. The latter type of situation are where ecosystem leaders will need to play an important role in either encouraging and incentivizing teams to engage in new ways of working, or by finding other solutions or resources to address the issue.

Suggestions of methods to aid this investigation and diagnosis are outlined in Textbox 1. Once a clear understanding of the current ecosystem has been developed, these methods can also be applied to assess how the ecosystem is changing over time.

In stage 2 stakeholders are engaged in co-design research to establish a shared understanding of the requirements for ecosystem evolution. The objective is to foster a collaborative agreement and commitment among stakeholders regarding strategies, practices, and resource investment to address system issues. Establishing such processes and building practical experience in working together will become increasingly important as more ambitious technologies are developed with greater potential impact and requiring greater collaboration and coordination across diverse ecosystem members. Suggestions of methods to aid this cocreation of strategies and action plans are outlined in Textbox 2.

The implementation of new information-driven technologies within the ecosystem are dealt with in stage three, focusing on achieving growth, learning, and continuous improvement. Suggestions of methods to aid the implementation, evaluation, and evolution of effective strategies and actions are outlined in Textbox 3.

Table 1. The use of projects as learning vehicles can be applied at all 3 stages of ecosystem enhancement.

Stage	Purpose of projects as focal point	Action
Coinvestigation and diagnosis	• To provide a pragmatic entry point to understand the work taking place within the ecosystem and how different stakeholders are working together.	• Review of previous and on-going projects.
Cocreation of strategies and action plans	• To provide a platform to explore new ways of working through experimentation.	• Using demonstration projects for testing, learning, and iteratively developing ideas until they become workable solutions for ecosystem enhancement.
Implementation, evaluation, and evolution	• To further inform ecosystem enhancement and evolution of new and improved ways of working to achieve ecosystem goals.	• Systematic learning from projects to increase the scale of application and rigor of evaluation.

Textbox 1. Activities in stage 1: coinvestigation and diagnosis of the current ecosystem.

Activity and their purpose

Interviews

To provide insights into the participant's unique perspectives of the ecosystem and their underpinning mental models.

Process and system mapping

To describe and visualize how the ecosystem is working, understanding who is involved and the roles that they play; how they connect to each other; where there are effective connections, and where there are disconnects. This includes mapping of key activities, actors, connections, disconnects, gaps, and incentives on both structural and organizational level.

Social network analysis

To explore the quality and extent of interactions between individuals within the ecosystem and provide baseline data to understand the growth of the ecosystem over time.

Collaborative workshops and learning events

To develop a shared understanding of the ecosystem and a collaborative diagnosis of the strengths, weaknesses, tensions, opportunities, and gaps of the current ecosystem.

XSL•FO
RenderX

Textbox 2. Activities in stage 2: cocreation of strategies and action plans.

Activity and their purpose
Formation of steering groups
To establish a focal firm to serve as the central point of coordination for the ecosystem evolution efforts, facilitating collaboration and communication among the various stakeholders.
Creation of dedicated spaces
To establish platforms for stakeholders to interact and engage in discussions, knowledge sharing, share insights, collective problem-solving, and test innovative ideas and aims to promote self-reflection, stimulate the generation of creative solutions, and build stronger working relationships.
Prospective case planning, design, and conduct
To identify the scope, objectives, and key activities for demonstration projects, developing design principles to guide the implementation of the projects and alignment with the desired outcomes.

Textbox 3. Activities in stage 3: implementation, evaluation, evolution of effective strategies, and actions.

Activity and their purpose
Development of ecosystem roadmap
To guide implementation efforts, with key steps and strategies that are refined and improved based on feedback, insights, and evolving needs.
Systematic project approach
To ensure effective execution of projects through the development and application of a systematic project approach.
Assessment of ecosystem growth
To gain evidence of the ecosystem's progress and understand the factors influencing network dynamics through evaluation of; the growth of the ecosystem, dynamics of social networks within it, remaining challenges and opportunities, and effectiveness of network interactions and relationships.

Conclusion

The successful integration of information-driven technologies into health care systems centers on increasing cohesion and synergistic working between members of the wider innovation ecosystem. This viewpoint provides steps for stakeholders to collaboratively investigate and enhance ecosystem functioning, enabling the effective deployment of information-driven technologies in routine health care practice, to advise the wider ecosystem goal of advancing the quality and efficiency of care delivery. Taking this agenda forward will require strong ecosystem leadership to challenge entrenched norms and behaviors and to engage members in creating new ways of working to enhance ecosystem cohesion and effectiveness.

Acknowledgments

The study was funded by the Swedish Innovation Agency and the Knowledge Foundation.

Authors' Contributions

All authors were involved in the conceptualization, writing, and final approval of the manuscript.

Conflicts of Interest

JR is the director of Julie Reed Consultancy Ltd.

References

1. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. J Biomed Inform 2019;100:103311 [FREE Full text] [doi: 10.1016/j.jbi.2019.103311] [Medline: 31629922]
2. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J 2021;8(2):e188-e194 [FREE Full text] [doi: 10.7861/fhj.2021-0095] [Medline: 34286183]
3. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022;24(1):e32215 [FREE Full text] [doi: 10.2196/32215] [Medline: 35084349]
4. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022;24(10):e40238 [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
5. Adner R. The Wide Lens: A New Strategy for Innovation. London: Portfolio/Penguin; 2012.

XSL•FO
RenderX

6. Adner R. Ecosystem as structure: an actionable construct for strategy. Journal of Management 2016;43(1):39-58 [FREE Full text] [doi: 10.1177/0149206316678451]
7. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
8. Barlow J, Bayer S, Curry R. Implementing complex innovations in fluid multi-stakeholder environments: experiences of 'telecare'. Technovation 2006;26(3):396-406. [doi: 10.1016/j.technovation.2005.06.010]
9. Webster A, Gardner J. Aligning technology and institutional readiness: the adoption of innovation. Technology Analysis & Strategic Management 2019;31(10):1229-1241. [doi: 10.1080/09537325.2019.1601694]
10. Svedberg P, Reed J, Nilsen P, Barlow J, Macrae C, Nygren J. Toward successful implementation of artificial intelligence in health care practice: protocol for a research program. JMIR Res Protoc 2022;11(3):e34920 [FREE Full text] [doi: 10.2196/34920] [Medline: 35262500]
11. Godin B. The linear model of innovation. Science, Technology, & Human Values 2006;31(6):639-667 [FREE Full text] [doi: 10.1177/0162243906291865]
12. Coiera E. The last mile: where artificial intelligence meets reality. J Med Internet Res 2019;21(11):e16323 [FREE Full text] [doi: 10.2196/16323] [Medline: 31702559]
13. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
14. Greenhalgh T, Wieringa S. Is it time to drop the 'knowledge translation' metaphor? A critical literature review. J R Soc Med 2011;104(12):501-509 [FREE Full text] [doi: 10.1258/jrsm.2011.110285] [Medline: 22179293]
15. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
16. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: normalization process theory. Implement Sci 2009;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
17. Marent B, Henwood F. Digital health: a sociomaterial approach. Sociol Health Illn 2023;45(1):37-53 [FREE Full text] [doi: 10.1111/1467-9566.13538] [Medline: 36031756]
18. Cecez-Kecmanovic D, Galliers RD, Henfridsson O, Newell S, Vidgen R. The sociomaterialty of information systems: current status, future directions. MISQ 2014;38(3):809-830 [FREE Full text] [doi: 10.25300/misq/2014/38:3.3]
19. Desouza KC, El Sawy OA, Galliers RD, Loebbecke C, Watson RT. Beyond rigor and relevance towards responsibility and reverberation: information systems research that really matters. CAIS 2006;17(16). [doi: 10.17705/1cais.01716]
20. Nilsen P, Reed J, Nair M, Savage C, Macrae C, Barlow J, et al. Realizing the potential of artificial intelligence in healthcare: Learning from intervention, innovation, implementation and improvement sciences. Front Health Serv 2022;2:961475 [FREE Full text] [doi: 10.3389/frhs.2022.961475] [Medline: 36925879]
21. Ovretveit J, Dolan-Branton L, Marx M, Reid A, Reed J, Agins B. Adapting improvements to context: when, why and how? Int J Qual Health Care 2018;30(suppl_1):20-23 [FREE Full text] [doi: 10.1093/intqhc/mzy013] [Medline: 29878138]
22. Lennox L, Barber S, Stillman N, Spitters S, Ward E, Marvin V, et al. Conceptualising interventions to enhance spread in complex systems: a multisite comprehensive medication review case study. BMJ Qual Saf 2022;31(1):31-44 [FREE Full text] [doi: 10.1136/bmjqs-2020-012367] [Medline: 33990462]
23. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
24. Siira E, Tyskbo D, Nygren J. Healthcare leaders' experiences of implementing artificial intelligence for medical history-taking and triage in Swedish primary care: an interview study. BMC Prim Care 2024;25(1):268. [doi: 10.1186/s12875-024-02516-z] [Medline: 39048973]
25. Desouza K, El Sawy OA, Galliers RD, Loebbecke C, Watson RT. Beyond rigor and relevance towards responsibility and reverberation: information systems research that really matters. CAIS 2006;17(16) [FREE Full text] [doi: 10.17705/1cais.01716]
26. Reed JE, Howe C, Doyle C, Bell D. Successful healthcare improvements from translating evidence in complex systems (SHIFT-Evidence): simple rules to guide practice and research. Int J Qual Health Care 2019;31(3):238-244 [FREE Full text] [doi: 10.1093/intqhc/mzy160] [Medline: 30085160]
27. Wenger-Trayner E. Communities of practice and social learning systems. Organization 2000;7(2):225-246. [doi: 10.1177/135050840072002]
28. Kislov R, Harvey G, Walshe K. Collaborations for leadership in applied health research and care: lessons from the theory of communities of practice. Implementation Science. 2011. URL: http://www.implementationscience.com/content/6/1/64 [accessed 2025-02-28]

XSL•FO
RenderX

29. Haynes A, Rychetnik L, Finegood D, Irving M, Freebairn L, Hawe P. Applying systems thinking to knowledge mobilisation in public health. Health Res Policy Syst 2020;18(1):134 [FREE Full text] [doi: 10.1186/s12961-020-00600-1] [Medline: 33203438]
30. Senge PM. The fifth discipline, the art and practice of the learning organization. Performance Improvement 2007;30(5):37-37. [doi: 10.1002/pfi.4170300510]
31. Holmes BJ, Best A, Davies H, Hunter DJ, Kelly MP, Marshall M, et al. Mobilising knowledge in complex health systems: A call to action. Evidence & Policy A Journal of Research Debate and Practice 2017;13(3):539-560 [FREE Full text] [doi: 10.1332/174426416x14712553750311]
32. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
33. Dreier L, Nabarro D, Nelson J. Harvard kennedy school, systems leadership for sustainable development: strategies for achieving systemic change 2019. Harvard Kennedy School. 2019. URL: https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/files/Systems%20Leadership.pdf [accessed 2025-02-28]
34. Reed JE, Derek B. Making an Impact? The Emergence of Improvement Science in Healthcare. International Journal of Science in Society 2013;4(4):77-92. [doi: 10.18848/1836-6236/CGP/v04i04/51408]
35. Orlikowski W. Sociomaterial practices: exploring technology at work. Organization Studies 2007;28(9):1435-1448 [FREE Full text] [doi: 10.1177/0170840607081138]
36. Valdez RS, Holden RJ, Novak LL, Veinot TC. Technical infrastructure implications of the patient work framework. J Am Med Inform Assoc 2015;22(e1):e213-e215 [FREE Full text] [doi: 10.1093/jamia/ocu031] [Medline: 25665705]
37. Cecez-Kecmanovic D, Galliers R, Henfridsson O, Newell S, Vidgen R. The sociomaterialty of information systems: current status, future directions. MISQ 2014;38(3):809-830 [FREE Full text] [doi: 10.25300/misq/2014/38:3.3]
38. Reed J. Eliminate slogans and remove barriers to pride in work. J Health Serv Res Policy 2016;21(3):206-208 [FREE Full text] [doi: 10.1177/1355819616635682] [Medline: 27194221]
39. Lander B. Boundary-spanning in academic healthcare organisations. Research Policy 2016;45(8):1524-1533. [doi: 10.1016/j.respol.2016.01.006]
40. Denis JL, Langley A, Rouleau L. The practice of leadership in the messy world of organizations. Leadership 2010;6(1):67-88 [FREE Full text] [doi: 10.1177/1742715009354233]
41. Reed JE, Howe C, Doyle C, Bell D. Simple rules for evidence translation in complex systems: a qualitative study. BMC Med 2018;16(1):92 [FREE Full text] [doi: 10.1186/s12916-018-1076-9] [Medline: 29921274]
42. Reed JE, Green S, Howe C. Translating evidence in complex systems: a comparative review of implementation and improvement frameworks. Int J Qual Health Care 2019;31(3):173-182 [FREE Full text] [doi: 10.1093/intqhc/mzy158] [Medline: 30060185]
43. Svensson L, Ellström PE, Brulin G. Introduction – on interactive research. IJAR – International Journal of Action Research 2007;3(3):233.
44. Kania J, Kramer M, Senge P. The water of systems change. FSG. 2018. URL: https://www.fsg.org/resource/water_of_systems_change/ [accessed 2025-02-28]

Abbreviations

AI: artificial intelligence
NASSS: Non-adoption, abandonment, scale-up, spread, and sustainability

Edited by A Schwartz; submitted 27.01.24; peer-reviewed by Y Hayat, S Nilsson, S Rajput, M Kolotylo-Kulkarni, A Hassan; comments to author 06.04.24; revised version received 15.08.24; accepted 20.02.25; published 24.03.25.

Please cite as:
Reed J, Svedberg P, Nygren J
Enhancing the Innovation Ecosystem: Overcoming Challenges to Introducing Information-Driven Technologies in Health Care
J Med Internet Res 2025;27:e56836
URL: https://www.jmir.org/2025/1/e56836
doi:10.2196/56836
PMID:

©Julie Reed, Petra Svedberg, Jens Nygren. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Balancing Risks and Opportunities: Data-Empowered-Health Ecosystems

Lan Li[1], PhD; Emma Back[1], MRES; Suna Lee[1], MSc; Rebecca Shipley[1], PhD; Néo Mapitse[2], PhD; Stefan Elbe[3], PhD; Melanie Smallman[1], PhD; James Wilson[1], PhD; Ifat Yasin[1], PhD; Geraint Rees[1], PhD; Ben Gordon[4], MSc; Virginia Murray[5], PhD; Stephen L Roberts[1], PhD; Anna Cupani[1], PhD; Patty Kostkova[1], PhD

[1]University College London, London, United Kingdom
[2]World Organisation for Animal Health, Paris, France
[3]University of Sussex, Brighton, United Kingdom
[4]Our Future Health, London, United Kingdom
[5]UK Health Security Agency, London, United Kingdom

Corresponding Author:
Lan Li, PhD
University College London
Gower Street
London, WC1E6BT
United Kingdom
Phone: 44 07529917633
Email: lan.li.19@ucl.ac.uk

Abstract

This viewpoint paper addresses the ongoing challenges and opportunities within the data-for-health ecosystem, drawing insights from a multistakeholder workshop. Despite notable progress in the digitization of health care systems, data sharing and interoperability remain limited, so the full potential of health care data is not realized. There is a critical need for data ecosystems that can enable the timely, safe, efficient, and sustainable collection and sharing of health care data. However, efforts to meet this need face risks related to privacy, data protection, security, democratic governance, and exclusion. Key challenges include poor interoperability, inconsistent approaches to data governance, and concerns about the commodification of data. While emerging platforms such as social media play a growing role in gathering and sharing health information, their integration into formal data systems remains limited. A robust and secure data-for-health ecosystem requires stronger frameworks for data governance, interoperability, and citizen engagement to build public trust. This paper argues that reframing health care data as a common good, improving the transparency of data acquisition and processing, and promoting the use of application programming interfaces (APIs) for real-time data access are essential to overcoming these challenges. In addition, it highlights the need for international norms and standards guided by multisector leadership, given the multinational nature of data sharing. Ultimately, this paper emphasizes the need to balance risks and opportunities to create a socially acceptable, secure, and effective data-sharing ecosystem in health care.

(J Med Internet Res 2025;27:e57237) doi:10.2196/57237

KEYWORDS

health policy; data sharing; digital healthcare; healthcare system; ecosystems; technologies; decision-making; data privacy; data protection; social media; application programming interfaces

Introduction

Data are essential to the delivery of effective health care and the improvement of health outcomes. However, the power of data relevant to health has not been fully realized. There is a need to nurture and develop data-for-health ecosystems to collect and share valuable data through transformative and responsible technological systems in timely, safe, efficient, and sustainable ways that maximize data utility for the public good.

A “data ecosystem” refers to a network of entities engaged in generating, sharing, and using data resources through various technologies [1]. In the context of a data-for-health ecosystem, this network is diverse and decentralized, comprising public and private institutions, as well as individuals acting in various

XSL•FO
RenderX

roles such as citizens, patients, and consumers. Data are actively and passively generated across all levels of this ecosystem, intentionally recorded for health purposes (eg, National Health Service [NHS] patient records) or initially generated for different reasons but repurposed for health care improvement (eg, geospatial air quality data shared to assist citizens with respiratory issues). Data management within this ecosystem varies across a spectrum: from tightly controlled data, such as patient records protected by strict privacy regulations and access protocols to ensure confidentiality, to fully open data, such as anonymized census or survey data that are freely accessible to the public for research and policy-making purposes. This range encompasses well-structured databases and unstructured streams from sources such as social media and Internet of Things devices, with knock-on effects for how data can be used.

Electronic devices generate digital traces that contribute valuable insights into personal health, extending beyond traditional patient records to include data held by IT and mobile corporations, and even loyalty card accounts [2]. The wide spectrum and variety of spatial-temporal geo-located big data streams that are collected include web-based search keywords, internet and store-based purchases, sensors and Internet of Things (IoT) data, health and fitness app data, and mobile phone location data [3,4].

The rapid evolution of this ecosystem poses risks related to privacy, data protection and security, transparency, access, and exclusion, as highlighted in the 2014 United Nations report "A World That Counts" [5]. The COVID-19 pandemic heightened concerns, due to the proliferation of systems capturing personal information for public health benefit, but often at the expense of robust debate about privacy, data ownership, and technology regulation.

Despite the potential for leveraging digital traces in health care, limitations persist, revealing gaps and missed opportunities [6]. The lack of interoperability between platforms and databases hinders data sharing across sectors, disciplines, countries, and international agencies. While such data holds substantial value for individuals, clinicians, health system planners, policy makers, researchers, and innovators, restrictions on visibility, access, and sharing impede its full potential. For instance, integrating data from health and fitness apps with electronic patient records could enhance clinical understanding, but business interests often deter app companies from developing interfaces for seamless integration.

Beyond interoperability, cross-border data sharing presents another critical hurdle, particularly in addressing global health challenges. Despite its growing importance, legal, regulatory, and technical barriers continue to limit progress. The lack of harmonized data protection standards across countries complicates the sharing of health data, particularly when navigating diverse privacy regulations like the General Data Protection Regulation (GDPR) in the European Union (EU) and less stringent frameworks in other regions. Countries often have conflicting data protection laws, making it difficult to ensure compliance while enabling international data flows. In addition, concerns over data sovereignty, where countries insist on controlling their citizens' data within national borders, further limit cross-border collaboration [7,8]. In public health emergencies, such restrictions can delay data access, hindering rapid responses. For example, during the COVID-19 pandemic, differing regulations and limited cross-border collaboration created obstacles for real-time sharing of epidemiological data, impeding global tracking efforts [9].

In 2016, the "Findable, Accessible, Interoperable, Reusable" principle was introduced as a guide for improving data reusability, focusing on both machine and human accessibility by stakeholders from academia, industry, funding agencies, and scholarly publishers [10]. Nevertheless, government agencies encounter significant challenges in leveraging population-level data for decision-making due to issues like delayed and inconsistent official data collection and aggregation, and the cost of accessing commercial datasets. Such difficulties were exacerbated during the COVID-19 pandemic, particularly for agencies in low- and middle-income countries with limited funding and infrastructure [11]. Increasingly, nonofficial data is relied upon to provide timely information. This may be generated by citizens or communities as an evidence base for change or to enhance social accountability, but it is often collected by private providers, with individual consumers contributing their data either actively or passively. Increasingly, multinational technology companies engaging in "massive and passive" data collection are dominating data-for-health ecosystems. This trend was accelerated during the COVID-19 pandemic [12] when social media platforms and search tools such as Twitter, Facebook, Google, and Bing played a pivotal role in providing timely information for researchers to monitor public sentiment and assess intervention impacts [13]. While the technology industry certainly benefits from leveraging personal data for web-based marketing, digital epidemiologists demonstrate how these new data streams can also enhance understanding of citizen behavior, support decision-making, and enable reporting of disease symptoms.

The aim of this viewpoint paper is to explore the key challenges and opportunities within these evolving ecosystems, synthesizing insights from the "Data Enabled Society for Health: Challenges and Opportunities" workshop held in London by University College London. This workshop brought together health care providers, policymakers, technologists, and researchers, to consider how to create a secure and effective data-for-health ecosystem that facilitates the gathering and sharing of valuable data for both individual and population-level health benefits, while effectively managing the associated risks. We suggest that to optimize the use of health data, governments and institutions should work on three overlapping areas: data governance, interoperability, and participation and citizen engagement..

Navigating the Dual Realities of Data Governance: Public Sector Caution and Industry Data Challenges

Ongoing issues with data governance, data access, and privacy highlighted during the workshop extend beyond regulatory and technical concerns to profound ethical dilemmas. These include

XSL•FO
RenderX

the challenge of obtaining meaningful informed consent in complex data ecosystems, the risk of data commodification that reduces individuals to mere data points, and the ethical responsibility of both public and private entities to prioritize the rights of individuals over profit or efficiency. These concerns also intersect with issues of data equity, where marginalized groups may face greater risks of exclusion or exploitation.

Within the public sector, citizens are not the only ones to limit access to their data due to privacy or security concerns. Governmental organizations and health care institutions are also very wary of granting access to their datasets, sometimes even to researchers, fearing either reputational damage due to mishaps or the loss of future financial opportunities if they share their assets.

This protective attitude towards an organization's data assets reflects the perception of data as a commodity: something that can be "extracted" from individuals or communities and then traded. We think it is important to reframe this narrative and look at personal data in the same way we look at common goods, with regulated access and shared collective benefits [14,15].

Attitudes against sharing medical information can stem from confusing messages and a lack of controls in the past, as well as fear of data being shared with third parties in ways that invade privacy, lead to increased surveillance and authoritarianism, or enable personal data exploitation against the individual's interests (eg, with an insurance provider). Citizen engagement needs to be specific and honest about the risks and benefits of data sharing, rather than relying on an abstract notion of how data sharing may benefit humankind, to avoid pitfalls such as the failed launch of NHS Care.data initiatives [16] and the abandoned centralized NHS COVID-19 Test and Trace app [17]. At the policy level, the key challenges are increasing the transparency of data governance and developing robust deidentification techniques that preserve the usability of data for research, rather than anonymization. Transparency and open dialogue with citizens are paramount for regaining public trust and setting cornerstones for a balanced agenda. This must be seen as an ongoing process, as trust is never a given.

The above-described scenario of data sharing sits in stark contrast to the vast quantity of personal physiological and health data collected through social media, wearables, tracking devices, MedTech, and geo-located mobile apps. While there are differences in legal frameworks across geographies, for example, in Europe (where GDPR applies, giving more control to citizens over their data), the United States, and China, personal data are generally subject to industry-defined terms and conditions, often with no clear or accessible opt-out clauses allowing use for personalized web-based or mobile marketing. This happens without much awareness or evident concern from citizens and in the absence of robust policy debate. To rebalance the current situation, there is an urgent need for international regulation, oversight, and a restoration of user controls over personal data.

The evolving nature of data-for-health ecosystems therefore raises ethical issues that require deeper consideration. These include mechanisms for ensuring informed consent, particularly in settings where data collection occurs passively or through third-party platforms, often without clear communication with individuals. Traditional models of consent are increasingly inadequate in such ecosystems, leading to a growing call for more dynamic or granular consent mechanisms that allow individuals greater control over specific types of data use. Moreover, the balance between individual privacy and public health is a critical ethical dilemma. While sharing health data has clear societal benefits, especially during public health emergencies like the COVID-19 pandemic, it raises the risk of compromising individual privacy. Finding an ethical equilibrium requires robust deidentification techniques, transparent data governance frameworks, and public engagement in decision-making processes. Finally, we must address the ethical issues of data commodification, where individuals' health data is treated as a commercial product. This commodification risks not only violating privacy but also entrenching inequalities, as those with fewer resources may have less control over their data and greater exposure to exploitation.

Ending Data Silos: Radical Interoperability and Real-Time Access

In addition to governance and access to data, the lack of common platforms and standards for data sharing limits collaboration across institutions and countries. This can have potentially devastating consequences in public health emergencies when a rapid and coordinated response is needed.

A notable example of successful collaboration is the Innovative Medicines Initiative ConcePTION consortium (2019), a European partnership involving public and private organizations that collect or access data related to pregnancy, childbirth, and lactation. Innovative Medicines Initiative ConcePTION aims to enhance the availability of health data and transform it into actionable evidence to improve clinical practices and outcomes [18]. Another valuable framework is the Unified Information and Interoperability Governance model, which integrates principles of both information governance and interoperability. The Unified Information and Interoperability Governance framework was designed to improve adaptability, flexibility, and efficiency in health information usage across the entire health care system [19]. Similarly, the United Nations Educational, Scientific and Cultural Organization - Committee on Data for Science and Technology (UNESCO-CODATA) Global Project contributes to the UNESCO Open Science Toolkit by providing comprehensive guidance, checklists, and factsheets specifically aimed at fostering data transparency, accessibility, and collaboration during crises [20]. The project delivers tools for both policy makers and scientists to support data collection, management, and governance. Such frameworks could be instrumental in future health crises, aligning with the goals of creating more interoperable and transparent data ecosystems.

One underused approach to health data access is that of application programming interfaces (APIs), which facilitate real-time, secure data sharing between systems. When integrated with emerging technologies like blockchain and federated learning, APIs can support more secure, privacy-preserving, and decentralized data exchanges. Blockchain can provide the security backbone for APIs, ensuring that data integrity and

XSL•FO
RenderX

ownership are maintained, while federated learning can allow institutions to collaborate on data analytics without sharing sensitive raw data. Combined with machine learning algorithms, APIs can further enhance real-time data analysis, enabling rapid public health responses based on predictive insights. This is of particular importance for real-time data streams, such as those generated by social media or IoT sensors. An example of this approach was Twitter allowing free API access to 1% of its database—a bold move that provided real-time data and fueled thousands of research projects in social computing, epidemic intelligence, sentiment analysis, and network and graph computing over a decade [21]. It was also incorporated into the epidemic intelligence platforms used by the World Health Organization (WHO) and its member states, such as MediSys, Epidemic Intelligence from Open Sources (EIOS), and Global Public Health Intelligence Network. This provision ceased under Twitter's new ownership [22]. Free access to the Twitter API was suspended in February 2023, causing significant disruption to academic research that relied on social media data [23]. Other examples of real-time data gathering and integration platforms include the flights portal Skyscanner and the financial products portal MoneySuperMarket. However, real-time data access for health applications is lacking.

Blockchain offers a groundbreaking solution for managing health data by using distributed, immutable ledgers that enhance transparency, security, and trust. In a data-for-health ecosystem, blockchain technology ensures that health data remains unaltered and that individuals retain control over their personal information through the use of smart contracts. These smart contracts facilitate secure, automated agreements between data owners and users, ensuring that data sharing complies with predefined privacy rules. The decentralized nature of blockchain makes it especially suitable for secure cross-border data sharing among institutions, allowing real-time access while safeguarding data integrity. Moreover, blockchain can support granular consent mechanisms, enabling individuals to specify who can access their data and under what conditions. A notable example is Estonia's health care system, which has successfully integrated blockchain technology to secure over 1 million patient records, ensuring both data privacy and transparency [24].

Meanwhile, machine learning (ML) is revolutionizing health care by enabling the analysis of large and complex datasets to identify patterns, predict outcomes, and support better decision-making. In data-for-health ecosystems, ML can be used to analyze real-time data streams from various sources, such as social media, wearable devices, and electronic medical records, to predict public health trends like disease outbreaks or shifts in vaccination behavior. ML algorithms enhance real-time epidemiological surveillance, facilitating the early detection of potential health threats through the analysis of social media activity, internet search trends, or mobile health app data. This predictive capability aids in timely public health interventions and resource allocation. For instance, Google's DeepMind has applied ML to the NHS to improve diagnostic accuracy and predict patient outcomes by analyzing historical health records [25].

Other innovative methods enable ML models to be trained on decentralized data without requiring the sharing of raw data between institutions. Federated learning allows health care institutions to collaborate on model development while maintaining their data locally, thus ensuring data privacy and compliance with regulatory standards like the GDPR. It promotes interoperability without compromising data ownership, making it ideal for cross-institutional collaborations. For example, federated learning has been successfully implemented in medical imaging projects, improving disease detection without the need for direct data sharing across institutions [26]. Similarly, zero-knowledge proofs offer a method for verifying the truth of information without revealing the underlying data. This can significantly enhance data-for-health ecosystems by allowing health institutions to share valuable insights while preserving the privacy of sensitive patient information. For example, health systems could confirm the validity of vaccination status or other health data without exposing any personal identifiers. Cryptographic zero-knowledge proof systems have been explored in secure identity verification services, such as confirming COVID-19 results, while keeping patient identities confidential [27]. In addition, the emerging role of the Metaverse in health care is gaining attention, as its potential applications continue to grow with advancements in Web 3.0 technology [28]. The Metaverse offers an immersive, interoperable ecosystem that could redefine the traditional health care system through avatar-based meetings, simulations, and social interactions between patients, providers, and organizations. However, technological innovation, regulatory oversight, and sound governance are needed to address current challenges in its development.

Winning Citizen Trust and Public Engagement

Limited access to technology in certain sectors of society and parts of the world can be attributed to resource and infrastructure constraints. Collaborative efforts across governments and communities are essential to address these challenges and prevent exclusion based on factors such as age or socioeconomic status [29].

Recently, concerns about data use and potential misuse have intensified due to health care provider hacking incidents and a lack of transparent communication regarding consequences and damages [30]. Case studies such as the Care.data program [14] and General Practice Data for Planning and Research [30] underscore the critical role of public trust in the success of health data initiatives. Public skepticism, evidenced by opt-out rates and project pauses, demonstrates the need for responsible data use [31].

Diminishing public trust, exemplified by incidents like the NHS Royal Free Trust data privacy breach, has profound consequences for current and future health data initiatives [30]. To foster meaningful citizen participation, mechanisms such as multistakeholder platforms, cocreation exercises, and public consultations can be used. Engaging the public in deliberative discussions, as demonstrated by the Understanding Patient Data (UPD) initiative, is crucial [32]. Through cocreative workshops [33], UPD facilitates meaningful discussions among diverse stakeholders, ensuring public perspectives and concerns are not

XSL•FO
RenderX

only heard but also integrated into decision-making processes [34]. This inclusive approach cultivates trust in responsible data usage and ensures citizen expectations shape data governance practices. Collaborative governance models offer a structure where stakeholders can jointly manage and govern data processes [35]. These models emphasize shared accountability, transparency, and continuous feedback, helping to mitigate power imbalances between different groups. Such models can be adapted to include conflict-resolution strategies to address disagreements or tensions that may arise during the decision-making process. Furthermore, achieving alignment across disparate groups requires consensus-building strategies. Techniques such as participatory decision-making frameworks and deliberative processes can help in forming agreement on key issues such as data ownership, privacy, and access rights [36].

Another notable initiative that frames health data as a common good is the Our Future Health program delivered in collaboration with the NHS. This program aims to collect health information from up to 5 million adults across the United Kingdom on a voluntary basis [37]. By September 2024, more than 1.8 million people had taken part. Researchers from universities, charities, the NHS, and health-related companies can apply for access to Our Future Health resources. It is anticipated that this will enable them to explore and test innovative approaches for predicting, detecting, and treating common diseases such as dementia, cancer, diabetes, heart disease, and stroke [37].

In summary, Textbox 1 outlines the key future directions, recommendations, and research agenda emerging from the workshop discussions.

XSL•FO
RenderX

Textbox 1. Future directions, recommendations, and research agenda.

Research priorities

Several key areas require further investigation to advance the development of robust and secure data-for-health ecosystems:

- Data privacy and security: Future research should focus on developing and refining privacy-preserving technologies that enable secure data sharing without compromising individual privacy. Researchers should also explore the effectiveness of granular consent mechanisms in empowering individuals to control their data.
- Interoperability: Addressing the lack of interoperability between health care systems, both within and across national borders, is vital. Building on successful interoperability and disease coding standards such as SNOMED-CD (Systematized Medical Nomenclature for Medicine–Clinical Terminology) and *ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision)* [38,39], research should prioritize the development of standardized data-sharing protocols and application programming interface–based solutions that can facilitate real-time data exchange while ensuring compliance with differing regulatory frameworks.
- Ethical governance models: As data-for-health ecosystems evolve, the development of inclusive governance models that balance the interests of diverse stakeholders will be essential. Researchers should focus on designing collaborative governance structures that integrate citizen engagement and address concerns about data commodification and trust.
- Cross-border data sharing: Given the global nature of health challenges, it is crucial that future research addresses solutions for overcoming the legal and regulatory barriers to cross-border data sharing. This effort should focus on developing and endorsing international agreements, as well as harmonizing standards, to ensure the secure and ethical exchange of health data. Key stakeholders in this process should include organizations such as the World Health Organization, European Commission, and national Ministries of Health. An important step in this direction is building on existing frameworks, like the European Health Information Gateway [40], to facilitate greater collaboration. In addition, fostering partnerships between the World Health Organization, the Association of European Operational Research Societies, and the European Centre for Disease Prevention and Control could further enhance data accessibility, with the creation of a joint data portal and application programming interface for developers to access health data securely and seamlessly.

Recommendations for policy makers

Policy makers play a crucial role in shaping the future of data-for-health ecosystems. While this is the hardest challenge to address due to the varying priorities of member states and IT industry commercial pressure We recommend the following actions:

- Regulatory harmonization: policy makers should work toward the harmonization of data protection standards globally, to facilitate cross-border data sharing while safeguarding privacy. Collaborative efforts between international organizations and governments are key to establishing global frameworks for data governance.
- Incentivizing technological innovation: governments should provide incentives for the adoption of privacy-enhancing technologies and interoperable data systems, including the use of federated learning and blockchain for secure data sharing.
- Public trust and engagement: public trust in data systems must be a priority. Policy makers should encourage the development of transparent governance frameworks that involve citizens in decision-making processes and ensure public accountability in the use of health data.

Guidelines for practitioners and researchers

Health care practitioners and researchers should focus on the practical implementation of new technologies and frameworks. We recommend:

- Adopting best practices for data sharing: health care institutions should adopt best practices for data sharing that prioritize data security, interoperability, and patient consent, while supporting the reuse of anonymized and pseudonymized data for research, policy, and social good. These practices can be informed by pilot projects and case studies that have successfully integrated innovative technologies such as machine learning and federated learning into health systems.
- Collaborative research initiatives: researchers should pursue collaborative initiatives that bring together stakeholders from academia, industry, and government to cocreate solutions. These collaborations will be critical for testing the feasibility and effectiveness of emerging technologies in real-world settings.
- Focus on equity and inclusion: researchers and practitioners must ensure that the benefits of data-for-health ecosystems are equitably distributed. This includes prioritizing access to technology for underrepresented populations and ensuring that governance models include mechanisms for inclusive participation.

Conclusion

Perspectives on the risks and opportunities within a data-for-health ecosystem vary based on the vantage points of government agencies, private entities, and individual citizens. Balancing top-down regulation with bottom-up citizen engagement is crucial for enabling data generation, usage, and sharing while safeguarding individuals. Examining successful practices, such as the NHS's Data Utility framework, and the case studies (Textbox 2) highlighted in this paper can inform the development of adaptable models across countries.

XSL•FO
RenderX

Textbox 2. Case studies in data-for-health ecosystems.

Successful interoperability: The COVID-19 vaccination rollout in Israel [41]

Israel's national health infrastructure enabled the rapid, seamless integration of vaccination data across health care providers during the COVID-19 pandemic. Interoperability between public health databases and electronic medical records allowed real-time data sharing, which facilitated efficient tracking, resource allocation, and public health decision-making. This success was driven by a well-established national health database and the effective use of application programming interfaces (APIs) that allowed data exchange between different systems. The case highlights how clear governance and technical standards can result in radical interoperability and improve public health outcomes.

- Key factors for success:

  A centralized national database, clear data-sharing protocols, and government-driven API adoption.

- Lessons learned:

  This case aligns with our proposed interoperability suggestions by demonstrating the value of standardized APIs and centralized databases in promoting real-time data sharing.

Interoperability challenges: Data sharing during the Ebola Epidemic (West Africa, 2014-2016) [42]

During the 2014-2016 Ebola epidemic in West Africa, the lack of interoperability between different health care systems and organizations severely hindered real-time data sharing and coordination among international health agencies. Various platforms were used to track the epidemic, but without standardized protocols, the response was delayed, costing lives. This case demonstrates the dangers of fragmented data systems during public health emergencies.

- Key factors for failure:

  Lack of standardized data-sharing protocols, limited use of APIs, and a fragmented health infrastructure.

- Lessons learned:

  The case reinforces the importance of adopting standardized interoperability solutions, such as APIs, to enable real-time data sharing across countries and organizations.

Governance challenges: The failure of the National Health Service Care.data program (United Kingdom) [43]

The NHS Care.data initiative, launched in 2014, intended to pool health data across general practices to improve patient care and medical research. However, the initiative faced severe backlash due to inadequate communication, a lack of clear consent mechanisms, and fears over data commodification. Concerns about patient privacy and the potential for data misuse led to public outcry, and the program was ultimately abandoned in 2016.

- Key factors for failure:

  Poor communication, lack of transparent consent processes, and public distrust.

- Lessons learned:

  This case emphasizes the need for clear, dynamic consent mechanisms and transparent governance practices. It also underlines the importance of trust-building initiatives before embarking on large-scale health data programs.

Citizen engagement: The Understanding Patient Data Initiative (United Kingdom) [44]

The Understanding Patient Data initiative in the United Kingdom is an example of effective citizen engagement in data governance. By involving the public in cocreation workshops, the initiative has fostered dialogue on how patient data should be used, addressing concerns about privacy and data ownership. Through transparency reports and educational programs, this initiative has built a stronger foundation of trust and enabled more meaningful public participation in decision-making about health data.

- Key factors for success:

  Transparent communication, active public involvement, and education initiatives.

- Lessons learned:

  This case supports our proposed citizen engagement advocacy, particularly the importance of transparency and cocreation in building trust in data governance.

The innovative deployment of transformative technologies in dynamic data-for-health ecosystems is a far-reaching goal. However, the international agenda must also focus on equity, access to technology, and IT literacy, embracing the needs of citizens of all ages and in all regions of the world, to avoid a growing digital divide. Furthermore, international norms and standards are required, given the burgeoning role of multinational companies in this space and the vital importance of transboundary data sharing when addressing public health emergencies of international concern.

Global collaboration plays a pivotal role in advancing standardized data-sharing frameworks, ensuring interoperability, and fostering mutual trust among nations. Establishing robust international partnerships can accelerate the adoption of common standards while respecting local contexts and priorities. Such collaboration will also help to harmonize efforts toward

XSL•FO
RenderX

equitable access to technology and shared public health goals. Large institutions have a key role in shaping the direction of this ecosystem and steering policies in specific directions. For example, the rapid and ubiquitous growth of efficient data collection tools—which have become an essential part of our lives—prompted the EU to devise the GDPR framework that became law in 2018. Due to the size of the market it covers and the EU's geopolitical role, it has effectively set the standard for data protection and management worldwide, building on the WHO and European Centre for Disease Prevention and Control European Health information Gateway [40]. Besides regulations and limitations, governmental bodies can promote and support private endeavors that create digital tools with a privacy-by-design approach. A recent joint statement—the European Health Data Space proposal—from health care stakeholders emphasizes the need for regulated industry access to health data for innovation and societal benefits, recognizing the interconnected nature of the data-for-health ecosystem [45]. Such initiatives, alongside the recommendations in this paper, offer the potential for effective management of this ecosystem. This is crucial to balance diverse interests and ensure that the value of health data benefits society as a whole.

Acknowledgments

The "Data Enabled Society for Health: Challenges and Opportunities" workshop was supported by the Institute of Healthcare Engineering and the Centre for Digital Public Health in Emergencies, University College London. This workshop created a platform for knowledge exchange across a community of professionals passionate about data for health. We acknowledge the commitment of the organizers and attendees. We would like to extend our gratitude to ChatGPT, an AI language model developed by OpenAI, for its assistance in proofreading and refining the language of this article. The "Data Enabled Society for Health: Challenges and Opportunities" workshop and the paper writing was financially supported by the Institute of Healthcare Engineering, University College London.

Authors' Contributions

LL, EB, SL, SLR, AC, and PK contributed to conceptualization and writing – original draft. RS, NM, SE, MS, JW, IY, GR, BG, and VM assisted with data curation and writing – review & editing.

Conflicts of Interest

None declared.

References

1. Oliveira MIF, Lóscio BF. What is a data ecosystem. 2018 Presented at: dg.o '18: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age; June 1, 2018; New York, NY. [doi: 10.1145/3209281.3209335]
2. Brewer HR, Hirst Y, Chadeau-Hyam M, Johnson E, Sundar S, Flanagan JM. Association between purchase of over-the-counter medications and ovarian cancer diagnosis in the cancer loyalty card study (CLOCS): observational case-control study. JMIR Public Health Surveill 2023;9:e41762. [doi: 10.2196/41762] [Medline: 36701184]
3. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018;14(1):1. [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
4. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 2018;19(1):9-20. [doi: 10.1038/nrg.2017.88] [Medline: 29129921]
5. Gonzalez L, Hsu YC, Poole J, Rae B, Rutherford I. A world that counts: mobilising the data revolution for sustainable development. UN Data Revolution. 2014. URL: https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf [accessed 2025-02-22]
6. Li L, Novillo-Ortiz D, Azzopardi-Muscat N, Kostkova P. Digital data sources and their impact on people's health: a systematic review of systematic reviews. Front Public Health 2021;9:645260 [FREE Full text] [doi: 10.3389/fpubh.2021.645260]
7. Chin YC, Zhao J. Governing cross-border data flows: international trade agreements and their limits. Laws 2022;11(4):63. [doi: 10.3390/laws11040063]
8. Hummel P, Braun M, Tretter M, Dabrock P. Data sovereignty: a review. Big Data Soc 2021;8(1):205395172098201. [doi: 10.1177/2053951720982012]
9. Towett G, Snead RS, Grigoryan K, Marczika J. Geographical and practical challenges in the implementation of digital health passports for cross-border COVID-19 pandemic management: a narrative review and framework for solutions. Global Health 2023;19(1):98. [doi: 10.1186/s12992-023-00998-7] [Medline: 38066568]
10. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]

XSL•FO
RenderX

11. Salama C. First lessons from government evaluations of COVID-19 responses: a synthesis. OECD. 2022. URL: https://tinyurl.com/yecuc8j9 [accessed 2023-06-12]
12. Roberts S, Kostkova P. Disease surveillance, digital futures, and data-sharing in a world 'after' COVID-19. Global Policy Journal. 2021. URL: https://www.globalpolicyjournal.com/articles/science-and-technology/disease-surveillance-digital-futures-and-data-sharing-world-after [accessed 2023-06-12]
13. Li L, Aldosery A, Vitiugin F, Nathan N, Novillo-Ortiz D, Castillo C, et al. The response of governments and public health agencies to COVID-19 pandemics on social media: a multi-country analysis of Twitter discourse. Front Public Health 2021;9:716333 [FREE Full text] [doi: 10.3389/fpubh.2021.716333] [Medline: 34650948]
14. Taylor MJ, Wilson J. Reasonable expectations of privacy and disclosure of health data. Med Law Rev 2019;27(3):432-460 [FREE Full text] [doi: 10.1093/medlaw/fwz009] [Medline: 31220873]
15. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? Open data for healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
16. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015;41(5):404-409 [FREE Full text] [doi: 10.1136/medethics-2014-102374] [Medline: 25617016]
17. Wise J. Covid-19: UK drops its own contact tracing app to switch to apple and google model. BMJ 2020;369:m2472. [doi: 10.1136/bmj.m2472] [Medline: 32561511]
18. Hollestelle MJ, van der Graaf R, Sturkenboom MC, van Delden JJ. Stimulating solidarity to improve knowledge on medications used during pregnancy: a contribution from the ConcePTION project. BMC Med Ethics 2023;24(1):44 [FREE Full text] [doi: 10.1186/s12910-023-00924-x] [Medline: 37370079]
19. Tumulak A, Tin J, Keshavjee K. Towards a Unified framework for information and interoperability governance. The Role of Digital Health Policy and Leadership 2024:49-53. [doi: 10.3233/shti231310]
20. Data policy for times of crisis facilitated by Open Science. CODATA. URL: https://codata.org/initiatives/data-policy/dptc/ [accessed 2024-09-18]
21. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
22. Ledford H. Researchers scramble as Twitter plans to end free data access. Nature 2023;614(7949):602-603. [doi: 10.1038/d41586-023-00460-z] [Medline: 36788282]
23. Murtfeldt R, Alterman N, Kahveci I, West JD. RIP Twitter API: a eulogy to its vast research contributions. arXiv:2404.07340 2024. [doi: 10.48550/arXiv.2404.07340]
24. Alexopoulos C, Charalabidis Y, Loutsaris MA, Lachana Z. How blockchain technology changes government: a systematic analysis of applications. Int J Public Adm Digit Age 2021;8(1):1-20. [doi: 10.4018/ijpada.20210101.oa10]
25. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-367 [FREE Full text] [doi: 10.1007/s12553-017-0179-1] [Medline: 29308344]
26. Darzidehkalani E, Ghasemi-Rad M, van Ooijen PMA. Federated learning in medical imaging: part I: toward multicentral health care ecosystems. J Am Coll Radiol 2022;19(8):969-974 [FREE Full text] [doi: 10.1016/j.jacr.2022.03.015] [Medline: 35483439]
27. Jo U, Oktian YE, Kim D, Oh S, Lee H, Kim H. A zero-knowledge-range-proof-based privacy-preserving blockchain platform for COVID-19 contact tracing. 2022 Presented at: International Conference on Platform Technology and Service (PlatCon); August 24, 2022; Jeju, Korea p. 53-58. [doi: 10.1109/platcon55845.2022.9932053]
28. Li Y, Gunasekeran DV, RaviChandran N, Tan TF, Ong JCL, Thirunavukarasu AJ, et al. The next generation of healthcare ecosystem in the metaverse. Biomed J 2024;47(3):100679 [FREE Full text] [doi: 10.1016/j.bj.2023.100679] [Medline: 38048990]
29. Vassilakopoulou P, Hustad E. Bridging digital divides: a literature review and research agenda for information systems research. Inf Syst Front 2023;25(3):955-969 [FREE Full text] [doi: 10.1007/s10796-020-10096-3] [Medline: 33424421]
30. Bharti N, O'Donovan C, Smallman M, Wilson J. Public trust, deliberative engagement and health data projects: beyond legal provisions. Engag Sci Technol Soc 2021;7(1):125-133. [doi: 10.17351/ests2021.1197]
31. Borda A, Molnar A, Neesham C, Kostkova P. Ethical issues in AI-enabled disease surveillance: perspectives from global health. Appl Sci 2022;12(8):3890. [doi: 10.3390/app12083890]
32. Understanding Patient Data. 2023. URL: https://understandingpatientdata.org.uk/about-us [accessed 2023-04-17]
33. Asher R, Jesper-Mir E. Why do you need my data? A co-creative approach to enriching public dialogues around health data. National Co-ordinating Centre for Public Engagement. 2022. URL: https://www.publicengagement.ac.uk/whats-new/blog/why-do-you-need-my-data-co-creative-approach-enriching-public-dialogues-around-health [accessed 2023-07-26]
34. Scott K, Burall S, Perrin N, Shelton P, White D, Irvine G. Data for Public Benefit Balancing the risks and benefits of data sharing. Understanding Patient Data. 2018. URL: https://understandingpatientdata.org.uk/sites/default/files/2018-04/Data%20for%20public%20good_0.pdf [accessed 2017-05-28]
35. Ansell C, Gash A. Collaborative governance in theory and practice. J Public Adm Res Theory 2008;18(4):543-571. [doi: 10.1093/jopart/mum032]
36. Kaner S. Facilitator's Guide to Participatory Decision-Making. San Francisco, CA: John Wiley & Sons; 2014.

XSL•FO
RenderX

37. Our Future Health. The UK's largest health research programme. 2024. URL: https://study.ourfuturehealth.org.uk/welcome [accessed 2024-09-18]
38. HL7 Version 2 Product Suite. HL7 International. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185 [accessed 2025-02-06]
39. ICD-10 Version. 2019. URL: https://icd.who.int/browse10/2019/en [accessed 2025-02-06]
40. WHO European health information at your fingertips. URL: https://gateway.euro.who.int/en/datasets/ [accessed 2025-02-06]
41. Rosen B, Hartal M, Waitzberg R. The Israeli health system's rapid responses during the COVID-19 pandemic. Isr J Health Policy Res 2024;13(1):11. [doi: 10.1186/s13584-024-00596-x] [Medline: 38438926]
42. Jalloh MF, Sengeh P, James N, Bah S, Jalloh MB, Owen K, et al. Integrated digital system for community engagement and community-based surveillance during the 2014-2016 Ebola outbreak in Sierra Leone: lessons for future health emergencies. BMJ Glob Health 2020;5(12):e003936 [FREE Full text] [doi: 10.1136/bmjgh-2020-003936] [Medline: 33355270]
43. Group BMJP. Problems with care.data and other stories. BMJ 2015;351:h4613. [doi: 10.1136/bmj.h4613]
44. Hopkins H, Kinsella S, Mil A, Mil HV. Foundations of fairness: views on uses of NHS patients' data and NHS operational data. Understanding Patient Data. URL: https://understandingpatientdata.org.uk/sites/default/files/2020-03/Foundations%20of%20Fairness%20-%20Full%20Research%20Report.pdf [accessed 2020-02-13]
45. Gatpandan N. Stakeholders' consensus response to the proposed European Health Data Space. iHD. 2022. URL: https://www.i-hd.eu/stakeholders-consensus-response-to-the-proposed-european-health-data-space/ [accessed 2023-06-12]

Abbreviations

API: application programming interface
EIOS: Epidemic Intelligence from Open Sources
EU: European Union
GDPR: General Data Protection Regulation
IoT: Internet of Things
ML: machine learning
NHS: National Health Service
UNESCO-CODATA: United Nations Educational, Scientific and Cultural Organization - Committee on Data for Science and Technology
UPD: Understanding Patient Data
WHO: World Health Organization

Edited by A Mavragani; submitted 08.02.24; peer-reviewed by J Lv, R van Der Graaf, M Toma; comments to author 03.06.24; revised version received 24.09.24; accepted 18.02.25; published 25.03.25.

Please cite as:
Li L, Back E, Lee S, Shipley R, Mapitse N, Elbe S, Smallman M, Wilson J, Yasin I, Rees G, Gordon B, Murray V, Roberts SL, Cupani A, Kostkova P
Balancing Risks and Opportunities: Data-Empowered-Health Ecosystems
J Med Internet Res 2025;27:e57237
URL: https://www.jmir.org/2025/1/e57237
doi:10.2196/57237
PMID:

©Lan Li, Emma Back, Suna Lee, Rebecca Shipley, Néo Mapitse, Stefan Elbe, Melanie Smallman, James Wilson, Ifat Yasin, Geraint Rees, Ben Gordon, Virginia Murray, Stephen L Roberts, Anna Cupani, Patty Kostkova. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Dying in Darkness: Deviations From Data Sharing Ethics in the US Public Health System and the Data Genocide of American Indian and Alaska Native Communities

Cason D Schmit[1,2,3,4], JD; Meghan Curry O'Connell[5], MD, MPH; Sarah Shewbrooks[5], MPH; Charles Abourezk[5,6,7], JD; Fallon J Cochlin[1,2,3], JD; Megan Doerr[8], MS, LGC; Hye-Chung Kum[1,2,4], PhD, MSW

[1]Department of Health Policy and Management, School of Public Health, Texas A&M University, College Station, TX, United States

[2]Population Informatics Lab, School of Public Health, Texas A&M University, College Station, TX, United States

[3]Program in Health Law and Policy, School of Public Health, Texas A&M University, College Station, TX, United States

[4]Institute for Healthcare Access, Texas A&M University, Ft. Worth, TX, United States

[5]Great Plains Tribal Leaders Health Board, Rapid City, SD, United States

[6]Rosebud Sioux Tribe Supreme Court, Rosebud, SD, United States

[7]Lower Brule Sioux Tribe, Lower Brule, SD, United States

[8]Sage Bionetworks, Seattle, WA, United States

Corresponding Author:
Cason D Schmit, JD
Department of Health Policy and Management
School of Public Health
Texas A&M University
212 Adriance Lab Rd
College Station, TX, 77843
United States
Phone: 1 9794360277
Email: schmit@tamu.edu

Abstract

Tribal governments and Tribal Epidemiology Centers face persistent challenges in obtaining the public health data that are essential to accomplishing their legal and ethical duties to promote health in American Indian and Alaska Native communities. We assessed the ethical implications of current impediments to data sharing among federal, state, and Tribal public health partners. Public health ethics obligates public health data sharing and opposes data collection without dissemination to affected communities. Privacy practices, like deidentification and data suppression, often obstruct data access, disproportionately affect American Indian and Alaska Native populations, and exacerbate health disparities. The 2020-2024 syphilis outbreak illustrates how restricted data access impedes effective public health responses. These practices represent a source of structuralized violence throughout the US public health system that contributes to the data genocide of American Indian and Alaska Native populations. Good governance practices like transparent data practices and the establishment of a social license (ie, the informal permission of a community to collect and use data) is essential to ethically balancing collective well-being with individual privacy in public health.

(J Med Internet Res 2025;27:e70983) doi:10.2196/70983

KEYWORDS

ethics; information dissemination; indigenous peoples; public health surveillance; privacy; data sharing; deidentification; data anonymization; public health ethics; data governance

Introduction

Advancements in digital technology have transformed the potential for real-time global data sharing, including in the field of public health. However, sharing data for public health use remains a challenge. Legal, relational, and technical barriers can impede access to the data that enable public health agencies to act for the benefit of their communities. [1-3] Most of these barriers are at least indirectly associated with a need to protect the privacy of data subjects.

Public health data custodians are acutely aware of these risks, but sometimes zealous overprotection of individuals' privacy

XSL•FO
RenderX

can harm communities [4]. In the case of American Indian and Alaska Native communities, a lack of access to public health data, data they are legally entitled to have, results in unnecessary deaths in these communities [5,6].

While this paper focuses on the ethical basis for sharing data with Tribes for public health activities, it is important to acknowledge the robust legal foundation for such action, including the treaties between the United States and various Indian Tribes that help round out the obligations of the federal government to Indian Tribes and Tribal people [7]. According to a long line of federal court cases, interpreting the Constitution, statutes, and treaties, the United States Congress has plenary and exclusive authority over Indian affairs. Within these cases, the Supreme Court has consistently held that "federal law preempts state law in the context of Indian affairs" [8]. Although Tribes and states coexist as distinct sovereigns, the legal doctrine of federal preemption requires that "Congressional power over Indian affairs supersedes conflicting state laws or state constitutional provisions" [8].

For example, when the Indian Health Care Improvement Act (IHCIA) and its subsequent amendments were enacted into law, this was construed as the federal government saying it was now preempting state action in the areas staked out within the IHCIA. Among other things, the IHCIA affirmed the trust responsibility of the federal government "to ensure the highest possible health status for Indians and urban Indians and to provide all resources necessary to effect that policy" [9,10]. Currently, the 12 Tribal Epidemiology Centers (TECs) provide epidemiologic and public health services to approximately 574 Tribes, 41 urban Indian organizations, and 9.7 million American Indian and Alaska Native people, and data are a necessary resource to support these activities [11]. The federal trust responsibility embedded in the IHCIA implicitly provides legal support for public health data sharing with Tribes and TECs as well as the ethical imperative to do so.

Federal public health law relating to Indian Tribes and federal services for Indian Tribes firmly supports Tribal governmental and TEC authority to collect and use data to promote public health [12]. This includes laws that empower governments to promote public health (eg, police powers), federal Indian law (eg, the Indian Self-Determination and Education Assistance Act of 1975 that established the federal policy of Tribal self-determination), and Tribal laws (eg, constitutions, codes, and case law) that support and shape Tribal public health authority [12].

As O'Connell and Abourezk [10] describe in detail, there is a strong federal legal foundation supporting the sharing of public health data with TECs due to their status as legally recognized public health authorities. Federal statutes that govern the receipt of identifiable data (ie, protected health information) by public health authorities from entities covered under the HIPAA (Health Insurance Portability and Accountability Act) (such as hospitals and medical clinics) apply to Tribes, TECs, and non-Tribal public health authorities.

Data access and use by Tribal governments and TECs is, therefore, legally distinct from the access and use of public health data by other entities that may work with or represent minoritized communities such as community nonprofit organizations. Tribes are sovereign nations that have the inherent right to govern their own resources, including data (ie, Tribal data sovereignty) in accordance with their cultural values and legal frameworks [13]. Restriction of data from Tribal governments, in particular, represents a failure to honor their legal status as sovereign entities that have been recognized since first contact with colonizers.

While federal and Tribal laws overwhelmingly support data sharing with Tribal public health authorities, Tribal governments, and TECs have long experienced difficulty in obtaining data from covered entities and their state and federal public health partners [10,14]. These data are central to effective public health practice and necessary for Tribal governments and TECs to fulfill their legal and ethical duties to promote health and eliminate disparities in American Indian and Alaska Native communities [15].

A 2022 US Government Accountability Office (GAO) report entitled, "Tribal Epidemiology Centers: HHS Actions Needed to Enhance Data Access," documents the difficulties experienced by TEC personnel in accessing essential public health data [15]. The GAO report found that access to and use of data varies by TEC with many TECs reporting challenges related to accessing data from the Centers for Disease Control and Prevention (CDC), Indian Health Service (IHS), and states. The report found that the work that TECs can perform and the information they can in turn provide to relevant public health decision makers is determined by their access to needed data. Consequently, when data access is limited or inconsistent, so too is the ability to address the health needs of Tribal communities. The report found that the US Department of Health and Human Services (HHS) lacked policies, procedures, and guidance related to existing data-sharing systems and agreements that resulted in barriers to Tribal HHS data access. The GAO report provided 5 recommendations for enhancements to HHS and IHS policies, procedures, and guidance to improve TEC access to public health data [15]. While HHS has taken steps to address some of these recommendations, Tribal governments and TECs still face barriers to accessing data from their state public health partners.

The consequences of these long-standing impediments to data access were tragically highlighted during the COVID-19 pandemic [5]. The failure of state and federal health departments to share timely and complete public health data during the COVID-19 response was highlighted in the 2021 Urban Indian Health Institute report, "Data Genocide of American Indians and Alaska Natives in the COVID-19 Data" [5]. A total of 15 states reported no American Indian and Alaska Native cases on their COVID-19 dashboards, and many lacked complete racial information on the fraction of cases reported to the CDC, making accurate monitoring of the COVID-19 pandemic in American Indian and Alaska Native communities impossible [5].

However, the consequences of these impediments to full data access predate and extend far beyond the COVID-19 pandemic. The prolonged dearth of actionable public health data has hampered the ability of Tribes and TECs to combat other common and preventable public health challenges and have

XSL•FO
RenderX

contributed to staggering disparities [6,16]. Tribal governments and TECs encounter systemic policy, bureaucratic, and relational barriers such as excessive privacy restrictions, data suppression practices, and paternalistic policies that impede their access to critically needed public health data.

In this paper we explore how these barriers perpetuate health inequities, examine the ethical dimensions of data sharing, and argue for a "share by default" model that aligns with public health ethics, respects Tribal sovereignty, and ensures the provision of timely, actionable information for American Indian and Alaska Native communities. Specifically, we assess the ethical and real-world implications of the status quo of data sharing impediments that exist between federal and state public health authorities and their Tribal public health and governmental partners. Currently, compared with their other public health authority counterparts, Tribal governments and TECs experience inequitable access to public health data, which severely impairs their ability to ensure the well-being of their constituent communities, fulfill their legally recognized role and responsibilities, and implement legal interventions and policy solutions that are informed by the best available public health data. While access to needed data is a challenge shared by nearly all public health authorities, the real, perceived, and sometimes intentional data-sharing barriers encountered by Tribal governments and TECs can be fairly described as structural violence perpetrated on Tribal communities.

To illustrate the deadly consequences of current data-sharing practices, we highlight the resurgence and prevalence of syphilis, a preventable and treatable disease, that has devastated American Indian and Alaska Native communities that do not have access to critically needed data that are readily available to their state and federal public health partners to effectively target preventative interventions and resources. These inequities in data access and data sharing represent an abdication of ethical responsibilities across the US public health system.

An Ethical Imperative to Share Data for Public Health Activities

Public health ethicists have long argued that there is an ethical duty to share public health data where there is a demonstrated public health need [17-19]. This ethical duty is essential to fulfilling the central purpose of public health surveillance; quickly and efficiently supplying information to decision-makers for population health.

For example, the 2017 World Health Organization (WHO) ethical guidelines on public health surveillance state, "[w]ith appropriate safeguards and justification, those responsible for public health surveillance have an obligation to share data with other national and international public health agencies" [17]. Other organizations have similarly spoken on the imperative to share data. The International Association of National Public Health Institutes called for sharing "public health surveillance data by default" and "with as few restrictions as possible" [18]. Lee et al [20] extended this point, arguing that "not using [public health] data for improving health must be justified" [20]. Both the American Public Health Association (APHA) Code of Ethics and the Public Health Leadership Society Principles for the Ethical Practice of Public Health urge prompt and appropriate data dissemination to enable public health actions [21,22]. Pisani and AbouZahr [23] argue that sharing public health data collected with public resources is important to the maximization of the public benefit from those costs.

All this underscores that collecting data without dissemination to affected communities is itself unethical. The APHA Principles of the Ethical Practice of Public Health states, "Public health institutions should provide communities with the information they have that is needed for decisions on policies or programs and should obtain the consent of the community for implementation" [22].

Certainly, as Langat et al [19] argue, there are important reasons for limiting data sharing including considerations of (1) data property and ownership, (2) just distribution of benefits and burdens, and (3) the competitive ethos of contemporary science [19]; however, they also note that these reasons are each outweighed by ethical considerations that support data sharing.

Moreover, an ethical right to be counted can countervail a right to privacy in public health contexts [24]. Accordingly, the WHO Guidelines on Ethical Issues in Public Health Surveillance recognize that there is an affirmative obligation that governments "develop appropriate, feasible, sustainable public health surveillance systems" [17]. Public health data showing an existing harm or imminent danger can shine a light on these threats and empower individuals and communities to compel a governmental response. Therefore, individuals and communities, from Love Canal, NY to East Palestine, OH, often demand to be counted [24-26]. Counting (ie, public health surveillance) informs resource allocations, and inequitable counting results in inequitable resource distributions [27]. In the context of public health data sharing, this ethical right to be counted is frustrated when barriers (or refusals) to share adequate data interfere with the ability of a government to collect the data necessary to carry out its obligations to care for its constituents.

Existing data sharing barriers that impede Tribal access to these data, whether legal, relational, political, bureaucratic, or willful, are ethically dubious at best. Instead, public health ethics imposes a default obligation to share data with public health authorities with as few restrictions as possible when there is a public health need.

Systemic Ethical Failures: Policies, Practices, and "Street-Level Bureaucrats"

Broadly, we argue that the US public health system is rife with deviations from these public health ethical imperatives [4,28,29]. These deviations manifest themselves in a variety of ways including in laws, policies, organizational practices, and discretionary decision-making that prohibits, unduly burdens, or otherwise limits ethical data sharing or access to needed data [4,30].

Here it is important to distinguish between law and ethics. Laws relate to what can or must be done, and ethics deals with what should be done. Importantly, what is ethical is not always legal,

XSL•FO
RenderX

and what is legal is not always ethical [31]. The APHA Code of Ethics implores public health practitioners to examine and deliberate such ethical considerations in the policy-making process [21]. Public health practitioners are silently complicit to injustice when they do otherwise [21].

In some cases, laws or policies (particularly at the state level) clearly restrict the authority of a public health data custodian to share data [30]. For example, Montana Code § 50-16-603, which strictly restricts disclosures of health information controlled by the public health authorities within the state, permits disclosures of protected information for public health purposes but only lists other state or local public health authorities. The omission of Tribes and TECs from the exhaustive list of recipients in the text of the law constitutes a barrier to ethical data sharing with Tribes. In this instance, the deviation of law from public health ethical imperatives is relatively transparent. This means that the ethical obligation to address the unethical law or policy is also transparent to every person who cites the law as a barrier to ethical conduct [21].

Beyond these transparent structural barriers, deviations from ethical imperatives frequently derive from opaque organizational and individual decisions [32]. The intergovernmental data-sharing landscape exemplifies this issue, characterized by "striking legislative minimalism" [32]. Consequently, data-sharing decisions are often left to the discretion of "street-level bureaucrats," with relatively low-level government agents wielding significant influence over agency actions [32]. The opacity of these decision-making processes arises from a variety of factors including data-sharing negotiations that may be shaped by (real or perceived) policies, [30] legal agreements that are not publicly accessible [33], decisions that may be hindered by unnecessarily burdensome processes that undermine public health data-sharing goals or that may be swayed by individual risk aversion, personal judgments, or interpersonal dynamics that are sometimes rooted in racism or paternalism. In each of these situations, the deviation from ethical imperatives is a product of the (often discretionary) actions of individuals and organizations within the US public health system.

Regardless of whether the ethical deficiencies are dictated by express policy provisions or products of individual or organizational actions, accepting the status quo is untenable. Such impediments and interference can amount to structural violence on native communities and are contrary to public health ethical imperatives [1,17,19,20,34,35]. Unethical policies provide only a false aegis to culpability for the individuals or organizations that cite them as justification for insufficient data sharing. Public health ethical imperatives demand actions to revise existing policies and practices that impede the sharing of timely and adequate data to Tribal public health organizations.

Outdated Approaches to Privacy Risks Impede Community Health Interests and Public Health Objectives

Privacy considerations are oft cited as a reason to withhold or redact public health data. These privacy considerations expressly or implicitly rest on bioethical concerns that have long been rejected in public health applications [1,17,35,36]. For example, consent has been the primary tool used to support and protect privacy interests in research and clinical contexts [37]. But, in public health contexts, overzealous protection of individual privacy harms the community. This is one reason why the HIPAA Privacy Rule includes a generous public health exception permitting public health use of protected health information without individual consent.

Misguided "Protections" That Cause Greater Harm

Tribal government and TEC efforts to access needed public health data are often frustrated by the misguided efforts of state and federal public health authorities to mitigate privacy risks from data sharing including discretionary deidentification, data suppression, and unnecessarily burdensome administrative access procedures.

Deidentification Can Contribute to Harm in Public Health Contexts

Deidentification can be a problematic approach to mitigating privacy risks in many public health contexts because it can cause both individual and group harm that offsets its benefits [38-40]. In the absence of informed consent, data controllers frequently rely on deidentification before sharing data.

There are 2 principal motivations for deidentifying data [27]. First, laws might allow data to be used or shared with fewer restrictions if deidentified. Second, deidentification is often used as an ethical precaution (as opposed to a legal requirement) to reduce the potential risks to data subjects (eg, when HIPAA covered entities insist on deidentifying data before disclosing protected health information pursuant to the research exception that does not legally require deidentification) [38]. But, when not legally required, the risks of deidentification must be weighed against the anticipated benefits [27]. This is why public health ethical guidelines expressly support the collection of identifiable data even in the absence of express consent under certain circumstances.

For example, the 2017 WHO [17] ethical guidelines state, "In some instances, the collection of names or identifiable data is both technically and ethically imperative." The guidelines expressly mention outbreak investigation, case follow-up, and preventive response activities as instances where disclosing names and other identifying information is appropriate [17]. Deidentification directly frustrates these important public health objectives by removing essential information, reducing the usefulness of information to provide the basis for action [41]. Moreover, the deidentification process consumes precious time so that when deidentified information is eventually shared its value for situational awareness is also diminished. Together, WHO Guidelines 11 and 14 (ie, the obligation to share data with other public health agencies, discussed above) state that there is a firm ethical duty for federal, state, and local public health agencies to share identifiable public health data with their Tribal counterparts when there is a public health need [17].

XSL•FO
RenderX

In fact, individual protections, like deidentification, often only shift the risks posed by data usage, rather than mitigate them [38,42]. For example, federal, state, and local health department practices that routinely deidentify public health data before sharing those data with Tribes or TECs for public health purposes will almost always suppress data about uncommon characteristics to reduce the risk of identifying individuals with those characteristics. This practice disproportionately affects Tribal communities. For instance, a public health agency that suppresses "small cell sizes" (a common deidentification practice) before sharing public health data will disproportionately suppress data on Tribal members because of the relatively small population of American Indian and Alaska Native people in any given area. In this way, deidentification practices contribute to social structures that systematically obscure the accurate assessment of health disparities in Tribal communities, ultimately leading to undercounting and underallocation of critically needed resources and support.

The assumption that deidentification reduces risks by rendering a data subject more difficult to identify is increasingly questionable in big data contexts [43]. For example, often the objective in big data applications is to gain group insights that can be both helpful or harmful to groups and the individuals that comprise them [43]. Deidentification cannot, however, eliminate all group risks but often will aggregate them. Thus, deidentification may result in real harm with little community health benefit.

Deidentification can also cause individual harm. Identifiable data are vital to helping public health agencies accurately monitor the spread of disease, identify persons at risk (eg, contact tracing), and target individual interventions [40]. When data holders deidentify the data provided to Tribal public health authorities, Tribes and TECs are unable to perform these basic public health interventions. The deidentification practices that are intended to protect individuals can impede the benefits that individuals can derive from these public health activities.

Data Suppression Hobbles American Indian and Alaska Native Epidemiology

Some US public health authorities that share data with Tribal public health partners exclude data on nonnative individuals, a precaution that is dangerous and problematic.

This practice directly contradicts public health ethical guidance in several ways. First, excluding nonnative persons from public health data sharing is bad epidemiological science that contradicts public health ethical guidance to share public health data. Excluding nonnative individuals from a dataset biases the data and denies Tribal epidemiologists comparison groups and the critical context needed to interpret the statistics. Informed public health decision-making requires this missing context. Moreover, excluding nonnative individuals from shared data effectively makes the identification and quantification of disparities impossible.

Second, excluding nonnative individuals from data shared with Tribal public health partners hides the health impacts that result from the interactions between native and nonnative individuals. In a world where disease regularly ignores geographic, political, and cultural boundaries, this type of data suppression simply makes no public health sense and amounts to a substantial barrier to understanding present and emerging public health threats. Excluding nonnative data makes even less sense given that many nonnative persons live on Tribal lands. Thus, excluding nonnative data from data shared with Tribal public health partners has harmful consequences for everyone.

Third, identifying Tribal members within state or local epidemiological datasets for the purpose of excluding all nonnative persons from data sharing needlessly complicates and delays data sharing. Existing electronic health records rarely record Tribal membership due to inadequate data standards, (ie, an absence of structured data fields for Tribal affiliation and inadequate documentation practices). Consequently, health departments that attempt to filter all nonnative persons from disclosures to Tribal organizations embark on a complex and protracted process that will delay timely data sharing. While we acknowledge that there is a tremendous benefit to developing data standards and systems that enable better Tribal affiliation documentation and demographic categorization, such actions are not prerequisites to better data-sharing practices.

Therefore, where federal, state, or local public health agencies refuse to share the nonnative data needed for good epidemiological investigations, they are deviating from public health ethical guidance to freely and promptly share public health data [1,17,19,20,34,35].

Administrative "Protections" That Amount to Data Discrimination

In contrast with the view of ethicists that public health data should be shared "by default" and "with as few restrictions as possible," [18] Tribal governments and TECs encounter administrative hurdles when they attempt to access public health data [15]. For example, the GAO report found that IHS required TECs to submit "Freedom of Information Act" requests, the same time-consuming process required of the general public, to access needed public health data [15]. This problem is not isolated to the federal government, as 3 TECs reported that restrictive state rules, among other reasons, prevented their access to any state data [15].

Unequal access to public health data for Tribal compared to non-Tribal public health authorities is a discriminatory practice. Yet, that is the status quo in the United States, despite the existence of equal federal legal authority to access data and perform public health functions [10].

As constructed, the current public health data sharing system de facto excludes Tribes from using public health data and results in Tribes having to seek these data from non-Tribal public health entities [5,15]. This system is likely a result of the historical paternalist and colonialist attitudes of the United States and state governments that consider Indian Tribes and their members incapable of being actively involved in their own governance. These stereotypes and inaccurate assumptions about Indian Tribes have prevented full Tribal participation in a range of governmental activities including in public health.

XSL•FO
RenderX

Unfortunately, these attitudes persist today and can impede non-Tribal public health authorities from providing life-saving data to Tribes when requested. So, Tribes and TECs face a double bind as they are excluded from the public health data collection system used by non-Tribal public health authorities and attempts to access the public health data system are blocked by the refusal of non-Tribal public health authorities to share data.

While some might argue these administrative burdens might be necessary to protect individual privacy interests, this too is a false aegis. In truth, these discriminatory administrative policies and processes result in persistent structural violence that contributes to systemic health disparities and deaths in American Indian and Alaska Native communities with little real benefit to individuals.

Case Study: Syphilis Epidemic (2020-Current)

The syphilis crisis in American Indian and Alaska Native communities (2020-2024) is an illuminating example of the failure of current data-sharing practices. Syphilis is a completely preventable disease that has been curable for nearly a century, and yet cases have risen dramatically and disproportionately in American Indian and Alaska Native populations as compared to other populations and demographic groups. Inadequate sharing of health data with Tribes, in stark contrast to ethical imperatives, is a significant barrier to Tribal government and TEC efforts to address this crisis [44].

For instance, South Dakota has seen a drastic increase in syphilis and congenital syphilis cases since 2020 (Figure 1 and Multimedia Appendix 1) [45,46]. To put these data in context, an 80% rise in national rates of syphilis between 2018 and 2022 justified the creation of a national task force For comparison, between 2019 and 2022, syphilis rates among American Indian and Alaska Native populations in SD surged by 6821%, reaching over 35 times the 2022 national rate. Yet, both the state department of health and federal IHS have been reluctant to provide protected health information to affected Tribes or the TEC that serves them. Because these agencies refuse to provide Tribes with needed public health data, Tribes have been unable to fully assist in the response.

Figure 1. Rates for Early Syphilis (per 100,000), US, South Dakota, and South Dakota American Indian and Alaska Native (2019-2022).

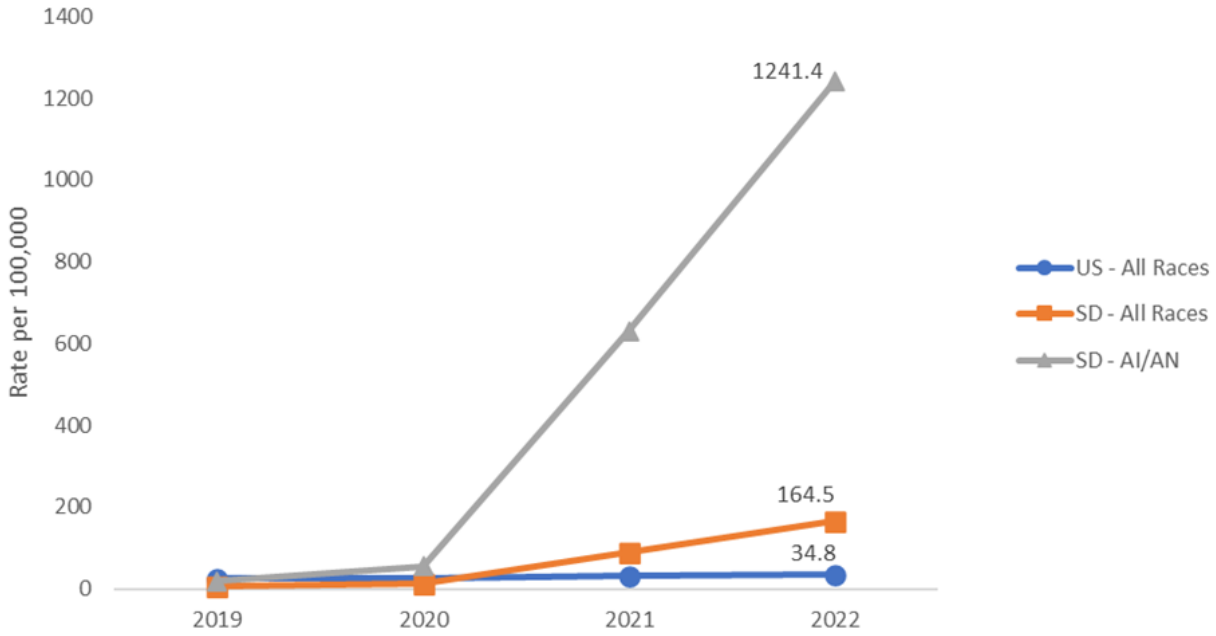

At the direction of the 18 Tribal leaders of the region in March 2023, the Great Plains Tribal Epidemiology Center (GPTEC) requested a CDC Epidemiological Aid (Epi-Aid) to assist with the syphilis outbreak. Due to the challenges that GPTEC encountered in accessing data on syphilis cases, a key objective of the Epi-Aid was to "[c]haracterize syphilis cases among Tribal populations in the [GPTEC] catchment area" (CDC, report to GPTEC, January 22, 2024). Despite CDC involvement and advocacy, the data required to complete the Epi-Aid was not provided by all relevant state health departments until November 2023, months after the CDC team deployment. While the deidentified data needed to conduct the report was ultimately provided, the delay in accessing the data contributed to the final report being published in January 2024, 6 months after the deployment. This rapid response action offered by the CDC took nearly a year to complete due to the data access challenges encountered by GPTEC. In that time syphilis continued to spread with 46 infants being diagnosed with congenital syphilis in South Dakota between March 2023 and January 2024, of whom 41 (89%) were American Indian and Alaska Native [45]. While syphilis is certainly concerning in the adult population, it can be deadly in newborns. The tragedy of this outbreak could have been mitigated if data-sharing practices were aligned with public health ethics.

Data are often shared with Tribes only after being deidentified. However, contact tracing, which is critical to managing outbreaks, is impossible without knowing who is infected. As a result, the deidentification standards that were intended to

XSL•FO
RenderX

protect individuals effectively denied them essential preventative and treatment services.

In April 2024, the CDC was again deployed to the region at the request of GPTEC to assist in contact tracing. Due to a data sharing agreement signed with the state of South Dakota in March 2024, GPTEC and South Dakota Tribes had access to identifiable data on people who had been diagnosed with syphilis or may have syphilis. In only 8 days of fieldwork teams of Tribal, TEC, CDC, and IHS staff members in 4 areas of the state were able to treat 62 people for syphilis, including 6 pregnant people (CDC, letter to GPTEC, May 28, 2024). The most significant difference between these 2 deployments was access to identifiable data.

When Tribal health authorities are unable to effectively respond to preventable infectious disease outbreaks, all people are put at greater risk. Though the syphilis outbreak spread nationally, the highest rates continue to be in South Dakota. If GPTEC and Tribes had access to data before the start of the outbreak consistent with ethical guidelines, they would have been able to better address it with contact tracing and other public health measures. The delay in data access—and continued lack of access in much of the great plains area—contributed to the high rates of syphilis in the region and, therefore, the country. Ethical data access for Tribes and TECs protects the health of all Americans.

A Better Way to Protect: Good Governance and Social License

In public health contexts, it is appropriate to take a group-based approach to the "respect for persons" ethical principles (as opposed to the individual-based approach of bioethics) [17]. For example, Megan Doerr argues that building or establishing a "social license" for collecting and using data can be more appropriate and protective than deidentification [47]. A social license represents the informal permission of a community to engage in a specific activity [48,49]. In public health contexts, a social license can legitimize the collection, use, or sharing of data about relevant communities. Similarly, WHO ethical guidelines for public health surveillance highlight the importance of involving affected communities in the decision-making process [17]. Developing a social license is essential to ensure that the trade-off between individual and collective interests is appropriate and ethically sound [43].

State and federal public health agencies established their social license to conduct public health surveillance through democratic processes, robust government accountability measures, and an understanding that the data collected will be used only to protect community health. Similarly, Tribal governments and TECs have a social license to collect and use public health data about their Tribal members established through their own sovereign governance and accountability mechanisms.

In fact, Tribes have substantial experience in the practices and activities that support social license for data collection and use. Tribal data sovereignty and indigenous data governance have emerged as widely recognized movements for Indigenous rights, self-determination, and ethical data management across academic, governmental, and community-based settings [13]. These movements were impelled, in part, by the unique risks resulting from the use of native data by nonnative persons or organizations that might not appropriately consider the harms or stigma that might result [50]. Yet, these movements are grounded in fundamental principles and rights inherent in sovereignty and personhood that likewise give power to social licenses as protective mechanisms for data collection and use. In particular, indigenous data governance practices and frameworks (eg, Collective benefit, authority to control, responsibility, and ethics Principles for Indigenous Data Governance) present compelling examples of how to effectively build and support social license in practice through relational accountability, stewardship, and respect for community values [13,51].

However, data on American Indian and Alaska Native individuals and Tribal members are often spread across jurisdictional lines, making data collection for Tribal organizations and TECs inherently more complex legally and practically than is the case for federal or state public health organizations. Obtaining the data necessary to fulfill their social license to use the data to promote population health requires the cooperation of federal and state public health partners.

Good governance and data stewardship are essential to support and nurture social license [52]. Public health authorities are effectively population fiduciaries who make decisions and take actions based on the best interests of the populations that they serve [53]. As such, privacy practices that prioritize the rights of individuals cannot be the only focus of public health agencies. Confidentiality protections provide an alternative approach to data governance that can permit public health authorities to use data more freely as population fiduciaries while still restricting inappropriate data uses. Similarly, adequate security measures can reduce risks of inappropriate access to or use of data while simultaneously permitting appropriate data uses. By shifting data protection governance strategies toward a population fiduciary strategy encompassing robust confidentiality and security practices, public health authorities can better align their data governance with public health ethics.

Nurturing a social license requires that data collection, its uses, and its governance are subject to public scrutiny that promotes data holder transparency and accountability that includes data dissemination to affected communities [17]. Above all, good governance must be grounded in ethical principles that do not ignore obligations to share public health data.

Discussion

The guidance of public ethicists in making public health data readily accessible between public health partners where there is a public health need provides a stark contrast to the status quo. Often Tribes—sovereign entities with primary public health responsibility for their constituent communities—are not consulted on state or federal public health data collection from Tribal communities and are not provided with the collected data needed to make crucial decisions and take needed action. This lack of consultation and provision of collected data results in 3 significant consequences. First, it wrongly subordinates Tribes

XSL•FO
RenderX

to state and federal public health partners. Second, it restricts the ability of Tribal public health authorities to promote public health and limits Tribal self-determination. And third, it results in the misallocation of resources, leaving Tribes guessing to the nature and magnitude of public health issues and unable to make data-informed decisions. Data collection wastes time and physical resources when it is not accompanied by the full dissemination critically needed by affected communities. For these reasons, the current default of not sharing data between federal, state, and local public health authorities and their Tribal partners cannot be ethically justified.

In some ways, federal Indian law interacts in unconscionable ways with the state of inadequate public health data availability experienced by Tribes. For instance, some governmental officials may conclude that the federal policy of Tribal self-determination absolves federal, state, and local governments of responsibility for threats to American Indian and Alaska Native public health. Yet, depriving Tribes of the public health data they need to protect themselves only ensures their continued vulnerability to increasingly existential threats. In this way, the status quo of inadequate data sharing with Tribes is just one dimension of a complex system of continued structuralized violence that contributes to the ongoing data genocide of American Indian and Alaska Native peoples [54].

There are meaningful steps that US public health agencies can take to address these systemic failures. Public health agencies at all governmental levels should:

1. Critically examine their data-sharing laws, policies, and practices to ensure that they are consistent with the ethical mandates of public health surveillance and action.

2. Bring any legal impediments to ethical data practices to the attention of policy makers (e.g., the absence of express language in laws stating that Tribal governments and TECs are permitted recipients of identifiable public health data). At a minimum, data-sharing practices with Tribes and TECs should be equivalent to other public health agencies [54].

3. Revise policies to align with a “share by default” approach, consistent with public health ethical guidelines, including removing data pre-processing practices (eg, de-identification, small-cell suppression, removal of non-native records) that delay the sharing of timely public health data and reduce the usefulness of the data that is shared. Consistent with Lee, Heilig, and White, decisions not to share timely data with Tribes should be justified, and justifications should expressly address the anticipated harms of not sharing data [20].

4. Adopt standard mechanisms for Tribes and TECs to request identifiable public health data with transparent and pre-defined terms for data transmission, use, and governance, consistent with public health ethics.

5. Ensure that public health data collection, sharing, and use practices respect tribal sovereignty. As Tribes are sovereign nations, their requests for data should be treated differently than requests from non-governmental groups. Public health agencies should create processes that honor Tribes as sovereign entities.

6. Establish and maintain regular communication with Tribes and TECs. Adopting policies and practices consistent with the federal guidelines on Tribal consultation can help public health agencies better understand both Tribal data needs and the barriers they encounter in achieving adequate data sharing [54].

Conclusion

Without aggressive actions to reverse the status quo, the US public health system will continue to be complicit in the data genocide of American Indian and Alaska Native communities. Public health authorities must adopt policies that expressly require data sharing with Tribes and TECs to limit the impact of discrimination and racial bias that leads to inequitable data access.

Acknowledgments

The authors gratefully acknowledge James G. Izat for his insightful comments and suggestions.

OpenAI’s ChatGPT v4 was used in the revision process to enhance the clarity and conciseness of existing content. Generative AI was not used to generate wholly original text or other content.

Data Availability

All data generated or analyzed during this project are included in the published article and Multimedia Appendix 1.

Disclaimer

The Great Plains Tribal Leaders’ Health Board is a Tribal organization led by the elected leaders of the 17 federally recognized Tribes and 1 Tribal community of North Dakota, South Dakota, Nebraska, and Iowa. The Great Plains Tribal Epidemiology Center performs its work only in consultation with and at the request of these Tribal governments.

Authors' Contributions

CDS and MCO contributed to conceptualization. MCO and SS managed data curation. CDS, FJC, and HCK handled formal analysis. CDS managed supervision. SS handled visualization. CDS, MCO, SS, CA, and HCK writing–original draft. All authors performed writing–review and editing.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Rates for Early Syphilis (per 100,000), U.S., South Dakota, and South Dakota AI/AN (2019-2022).
[DOCX File , 16 KB - jmir_v27i1e70983_app1.docx]

References

1. Strengthening data sharing for public health. The royal institute of international affairs. Chatham House. URL: https://www.chathamhouse.org/about-us/our-departments/global-health-programme/strengthening-data-sharing-public-health [accessed 2021-08-11]
2. Karim MA, Kum HC, Schmit CD. A study of publicly available resources addressing legal data-sharing barriers: systematic assessment. J Med Internet Res 2022;24(9):e39333 [FREE Full text] [doi: 10.2196/39333] [Medline: 36066929]
3. Schmit C, Larson BN, Kum HC, Curran C. Health affairs forefront. Public health informatics depends on engagement with privacy legislation. 2022. URL: https://www.healthaffairs.org/content/forefront/public-health-informatics-depends-engagement-privacy-legislation [accessed 2022-10-28]
4. Schmit CD, Willis B, McCall H, Altabbaa A, Washburn D. Views on increased federal access to state and local national syndromic surveillance program data: a nominal group technique study with state and local epidemiologists. BMC Public Health 2023;23(1):431 [FREE Full text] [doi: 10.1186/s12889-023-15161-5] [Medline: 36879220]
5. Urban Indian Health Institute. Data Genocide of American Indians and Alaska Natives in COVID-19 Data. Seattle, WA: Urban Indian Health Institute; 2021.
6. Friedman J, Hansen H, Gone JP. Deaths of despair and indigenous data genocide. Lancet 2023;401(10379):874-876. [doi: 10.1016/S0140-6736(22)02404-7] [Medline: 36709769]
7. United States, 450 F.Supp.3d 986 (D.S.D. 2020); affirmed by rosebud sioux tribe v. United States, 9 F.4th 1018 (8th Cir. 2021). Rosebud Sioux Tribe v. URL: https://casetext.com/case/rosebud-sioux-tribe-v-united-states-3 [accessed 2020-03-30]
8. Newton NJ, Washburn KK. Cohen's Handbook of Federal Indian Law. San Francisco, Calif: LexisNexis; 2012.
9. 9. 25 USC §1602. Cornell Law School Legal Information Institute. 2010. URL: https://www.law.cornell.edu/uscode/text/25/1602 [accessed 2025-03-05]
10. O'Connell MC, Abourezk C. Facilitating the urgent public health need to improve data sharing with tribal epidemiology centers. Public Health Rep 2023;138(2_suppl):80S-83S. [doi: 10.1177/00333549231152197] [Medline: 36734206]
11. Tribal Epidemiology Centers. Advancing Public Health in Indian Country. URL: https://tribalepicenters.org/ [accessed 2025-02-23]
12. Hoss A. A framework for tribal public health law. Nevada Law Journal 2019;20:113-114. [doi: 10.2139/ssrn.3428572]
13. Carroll SR, Rodriguez-Lonebear D, Martinez A. Indigenous data governance: strategies from United States native nations. Data Sci J 2019;18(1):31 [FREE Full text] [doi: 10.5334/dsj-2019-031] [Medline: 34764990]
14. Hodge JG, Kaufman T, Jaques C. Legal Issues Concerning Identifiable Health Data Sharing Between State/Local Public Health Authorities and Tribal Epidemiology Centers in Selected U.S. Jurisdictions. 2009. URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ec71bea6eaf49e9107ea77b18b9d54e01949e2f1 [accessed 2009-03-03]
15. Rosenberg MB. Tribal Epidemiology Centers: HHS Actions Needed to Enhance Data Access. URL: https://www.gao.gov/products/gao-22-104698 [accessed 2022-03-04]
16. Landman K. Syphilis rates are soaring in South Dakota?s American Indian communities and their infants. URL: https://www.vox.com/science-and-health/23460951/syphilis-south-dakota-american-indian-native-congenital [accessed 2023-04-03]
17. WHO guidelines on ethical issues in public health surveillance. World Health Organization. URL: https://www.who.int/publications/i/item/who-guidelines-on-ethical-issues-in-public-health-surveillance [accessed 2017-06-19]
18. Public health surveillance: a call to share data. International Association of National Public Health Institutes. 2016. URL: https://c-path.org/public-health-surveillance-a-call-to-share-data/ [accessed 2025-03-04]
19. Langat P, Pisartchik D, Silva D, Bernard C, Olsen K, Smith M, et al. Is there a duty to share? Ethics of sharing research data in the context of public health emergencies. Public Health Ethics 2011;4(1):4-11. [doi: 10.1093/phe/phr005]
20. Lee LM, Heilig CM, White A. Ethical justification for conducting public health surveillance without patient consent. Am J Public Health 2012;102(1):38-44. [doi: 10.2105/ajph.2011.300297]
21. American Public Health Association. Public Health Code of Ethics Issue Brief. Washington DC: American Public Health Association; 2019.
22. Public Health Leadership Society. Principles of the Ethical Practice of Public Health; Version 2. 2002. URL: https://www.apha.org/-/media/files/pdf/membergroups/ethics/ethics_brochure.ashx [accessed 2025-03-04]
23. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. Bull World Health Organ 2010;88(6):462-466 [FREE Full text] [doi: 10.2471/BLT.09.074393] [Medline: 20539861]

XSL•FO
RenderX

24. Fairchild AL. The right to know, the right to be counted, the right to resist: cancer, AIDS, and the politics of privacy and surveillance in post-war America. Journal of Medical Law and Ethics 2015;3(1):45-64. [doi: 10.7590/221354015x14319325750034]
25. Kleiman J. Love Canal: A Brief History . SUNY Geneseo. URL: https://www.geneseo.edu/history/love_canal_history [accessed 2022-04-14]
26. Elverd S. The marietta times. Congressmen press for bill to monitor health impacts of derailment. 2024. URL: https://www.mariettatimes.com/news/2024/06/congressmen-press-for-bill-to-monitor-health-impacts-of-derailment/ [accessed 2024-06-21]
27. Schmit C. Ongoing and Emerging Issues in Privacy and Security in a Post COVID-19 Era: An Environmental Scan. 2023. URL: https://oaktrust.library.tamu.edu/items/e28817fb-b8c3-46eb-b309-4a500f2ed19c [accessed 2023-03-22]
28. Rigby RC, Ferdinand AO, Kum HC, Schmit C. Data sharing in a decentralized public health system: lessons from COVID-19 syndromic surveillance. JMIR Public Health Surveill 2024;10:e52587 [FREE Full text] [doi: 10.2196/52587] [Medline: 38546731]
29. Schmit C, Giannouchos T, Ramezani M, Zheng Q, Morrisey MA, Kum H. US privacy laws go against public preferences and impede public health and research: survey study. J Med Internet Res 2021;23(7):e25266. [doi: 10.2196/25266] [Medline: 36260399]
30. Schmit C, Kelly K, Bernstein J. Cross sector data sharing: necessity, challenge, and hope. J Law Med Ethics 2019;47(2_suppl):83-86 [FREE Full text] [doi: 10.1177/1073110519857325] [Medline: 31298141]
31. Hulkower R, Penn M, Schmit C. Public health informatics and information systems, 3rd Edition. In: Magnuson JA, Dixon BE, editors. Privacy and Confidentiality of Public Health Information. London: London: Springer London; 2020.
32. Fahey BA. Data Federalism. Harv Law Rev 2022;135:1007.
33. Cochlin FJ, Curran CD, Schmit CD. Navigating new legal guardrails and emerging AI challenges. J Law Med Ethics 2024;52(S1):70-74. [doi: 10.1017/jme.2024.40] [Medline: 38995251]
34. Kass NE. An ethics framework for public health. Am J Public Health 2001;91(11):1776-1782. [doi: 10.2105/ajph.91.11.1776] [Medline: 11684600]
35. Lee LM. Public health ethics theory: review and path to convergence. J Law Med Ethics 2012;40(1):85-98. [doi: 10.1111/j.1748-720X.2012.00648.x] [Medline: 22458465]
36. Bayer R, Fairchild AL. The genesis of public health ethics. Bioethics 2004;18(6):473-492. [doi: 10.1111/j.1467-8519.2004.00412.x] [Medline: 15580720]
37. Schmit C, Larson B, Kum HC. Data privacy in the time of plague. SSRN Journal 2022;21:157-227. [doi: 10.2139/ssrn.3968130]
38. Francis LP, Francis JG. Data reuse and the problem of group identity. In: Studies in Law, Politics, and Society. Leeds, England: Emerald Publishing Limited; 2017.
39. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth 2010;10(9):3-11 [FREE Full text] [doi: 10.1080/15265161.2010.494215] [Medline: 20818545]
40. Tesoriero JM, Battles HB, Heavner K, Leung SJ, Nemeth C, Pulver W, et al. The effect of name-based reporting and partner notification on HIV testing in New York state. Am J Public Health 2008;98(4):728-735. [doi: 10.2105/ajph.2006.092742]
41. Kum HC, Krishnamurthy A, Machanavajjhala A, Ahalt SC. Social genome: putting big data to work for population informatics. Computer 2014;47(1):56-63. [doi: 10.1109/mc.2013.405]
42. Solow-Niederman A. SSRN Electronic Journa. Information Privacy and the Inference Economy. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3921003 [accessed 2021-09-10]
43. Doerr M, Meeder S. Big health data research and group harm: the scope of IRB review. Ethics Hum Res 2022;44(4):34-38. [doi: 10.1002/eahr.500130] [Medline: 35802789]
44. U.S. Department of Health and Human Services. HHS Announces Department Actions to Slow Surging Syphilis Epidemic. HHS.gov. URL: https://www.hhs.gov/about/news/2024/01/30/hhs-announces-department-actions-slow-surging-syphilis-epidemic.html [accessed 2024-07-29]
45. Infectious Disease Surveillance. South Dakota Department of Health. 2024. URL: https://doh.sd.gov/health-data-reports/infectious-communicative-disease-data-reports/infectious-disease-surveillance/ [accessed 2024-11-25]
46. Centers for Disease Control and Prevention. Table 12. Total Syphilis ? Reported Cases and Rates of Reported Cases by State/Territory and Region in Alphabetical Order, United States, 2017?2021. cdc.gov. 2017. URL: https://www.cdc.gov/std/statistics/2021/tables/12.htm [accessed 2024-01-23]
47. Kendall S, Lobuglio D. Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks. Washington, D. C: National Academies Press; 2022.
48. Shaw JA, Sethi N, Cassel CK. Social license for the use of big data in the COVID-19 era. NPJ Digit Med 2020;3(1):128 [FREE Full text] [doi: 10.1038/s41746-020-00342-y] [Medline: 33083563]
49. Ruhaak A, Yu JH, Doerr M. Group Licenses to Operate. Sage Bionetworks. URL: https://sagebionetworks.pubpub.org/pub/x775yqz5/release/1 [accessed 2024-07-29]
50. Sterling RL. Genetic research among the havasupai--a cautionary tale. Virtual Mentor 2011;13(2):113-117 [FREE Full text] [doi: 10.1001/virtualmentor.2011.13.2.hlaw1-1102] [Medline: 23121851]

XSL•FO
RenderX

51. Carroll SR, Garba I, Plevel R, Small-Rodriguez D, Hiratsuka VY, Hudson M, et al. Using indigenous standards to implement the CARE principles: setting expectations through tribal research codes. Front Genet 2022;13:823309 [FREE Full text] [doi: 10.3389/fgene.2022.823309] [Medline: 35386282]
52. Lee LA, Iversen BG, Lynes S, Desenclos JC, Bezuidenhoudt JE, Flodgren GM, et al. Integrated disease surveillance project summary report. 2022. URL: https://ianphi.org/_includes/documents/sections/tools-resources/ids/ianphi-ids-summary-report.pdf [accessed 2022-03-22]
53. Balkin JM. The fiduciary model of privacy. Harvard Law Review Forum 2020:134.
54. Council for StateTerritorial Epidemiologists. Enhancing Data Access to Improve American Indian and Alaska Native Health: A Framework for State and Local Public Health Officials. 2024. URL: https://cdn.ymaws.com/www.cste.org/resource/resmgr/briefs/24-PB-TE-01_Final.pdf [accessed 2024-08-28]

Abbreviations

APHA: American Public Health Association
CDC: Centers for Disease Control and Prevention
Epi-Aid: CDC Epidemiological Aid
GAO: Government Accountability Office
GPTEC: Great Plains Tribal Epidemiology Center
HHS: Department of Health and Human Services
HIPAA: Health Insurance Portability and Accountability Act
IHCIA: Indian Health Care Improvement Act
IHS: Indian Health Service
TEC: Tribal Epidemiology Center
WHO: World Health Organization

Edited by N Cahill; submitted 07.01.25; peer-reviewed by C Duke, K English, K Pierce Farrier; comments to author 04.02.25; revised version received 14.02.25; accepted 26.02.25; published 26.03.25.

Please cite as:
Schmit CD, O'Connell MC, Shewbrooks S, Abourezk C, Cochlin FJ, Doerr M, Kum HC
Dying in Darkness: Deviations From Data Sharing Ethics in the US Public Health System and the Data Genocide of American Indian and Alaska Native Communities
J Med Internet Res 2025;27:e70983
URL: https://www.jmir.org/2025/1/e70983
doi:10.2196/70983
PMID:

©Cason D Schmit, Meghan Curry O'Connell, Sarah Shewbrooks, Charles Abourezk, Fallon J Cochlin, Megan Doerr, Hye-Chung Kum. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Mindfulness-Based Cognitive Therapy–Game: An Ironic Way to Treat Internet Gaming Disorder

Jaehyun Kim[1], MD; Hayoung Oh[2], PhD; Anderson Sungmin Yoon[3], DSW

[1]Department of Applied Artificial Intelligence, College of Computing and Informatics, Sungkyunkwan University, Seoul, Republic of Korea

[2]College of Computing and Informatics, Sungkyunkwan University, Seoul, Republic of Korea

[3]Department of Social Welfare, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Hayoung Oh, PhD
College of Computing and Informatics
Sungkyunkwan University
25-2, Sungkyunkwan-ro
Jongno-gu
Seoul, 03063
Republic of Korea
Phone: 82 01053895996
Email: hyoh79@skku.edu

Abstract

Internet gaming disorder (IGD) affects 3% of the global population and poses an increasing risk due to advancements in technology. However, there is currently no definitive treatment for this condition. IGD is not a primary disorder but rather a result of "self-prescription" in response to emotional stressors. Unlike conventional mental health treatments that focus on the disorder itself, it is crucial to provide alternative activities that can alleviate negative emotions. This paper extends the concept of the self-medication hypothesis and integrates it with cognitive models of cognitive behavioral therapy and mindfulness-based cognitive therapy. In addition, it introduces the mindfulness-based cognitive therapy–game (MBCT-G), a program designed to explore alternative activities through gaming, focusing on the processes of response and reward, which are not typically emphasized in traditional treatments. This study serves as the theoretical foundation for the development of MBCT-G. MBCT-G aims to train individuals in positive coping strategies that alleviate psychological distress, offering a novel approach to treating self-prescription disorders such as IGD.

(J Med Internet Res 2025;27:e65786) doi:10.2196/65786

KEYWORDS

cognitive behavior therapy; psychosocial intervention; video games; internet gaming disorder; internet addiction; mindfulness; mental health

Introduction

The prevalence of internet gaming disorder (IGD) is known to be around 2% to 3% [1]. This prevalence varies depending on the region and the population studied [2-4]. In 2013, the American Psychiatric Association introduced IGD as a new disorder in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*). The World Health Organization (WHO) also recognized gaming disorder as a disease in the *International Statistical Classification of Diseases and Related Health Problems, 11th Edition* (*ICD-11*). However, although *DSM-5* provides standard definitions and diagnostic criteria, systematic treatment methods for gaming disorders have not been thoroughly reviewed, and clinical evidence remains insufficient [5,6]. Currently, cognitive behavioral therapy (CBT) is the most frequently used treatment method; however, it has not been proven superior to other treatments [7]. Patients with gaming disorder tend to prioritize gaming above all else, lose control over their gaming behavior, and exhibit an inability to voluntarily stop [8]. While these clinical features are similar to those of substance use disorder (SUD), the two conditions show significant differences from a therapeutic perspective.

Moreover, the term "internet gaming disorder" is often confused with other conditions. For instance, pathological internet use (PIU) refers to various online activities, such as gaming, social networking, shopping, watching videos, and reading newspapers that interfere with daily life. Some studies on IGD often include

XSL•FO
RenderX

PIU, which is not entirely identical to IGD [9]. According to one study, less than half of the patients with IGD met the criteria for PIU, while only 6.67% of those who met the criteria for PIU were classified as having IGD [10].

IGD possesses characteristics that are difficult to explain within the existing psychiatric disorder framework. Traditional attempts to treat individuals addicted to internet games may be ineffective or, at times, may negatively impact the trust relationship with the patient. Therefore, this study proposes a hypothesis regarding the pathogenesis of IGD based on its characteristics. This hypothesis serves as a complementary system to the interacting cognitive subsystem (ICS) model, which has been theoretically used to explain CBT and mindfulness-based cognitive therapy (MBCT). In addition, the study discusses the necessity and potential development of mindfulness-based cognitive therapy–game (MBCT-G).

Background

Self-Medication Hypothesis and IGD

The self-medication hypothesis refers to the tendency of individuals to autonomously select substances or specific behaviors to alleviate psychological or emotional distress [11]. According to the self-medication hypothesis, patients experience temporary relief from psychological pain through substances like cocaine or opioids. The rewarding effect of such substances weakens patients' self-control and increases their dependency on the selected substance. Patients who lose their self-control fail to consider the consequences of addiction, thereby heightening the likelihood of substance use and dependency [12].

In various studies on the causes and motivations behind IGD, the self-medication hypothesis has been discussed [13-17]. During pandemic situations like COVID-19, the prevalence of IGD increased [18,19], alongside a decline in self-control and a rise in depression and anxiety [20-22]. Resilience, defined as the ability to overcome stress, adversity, and psychological distress to return to a stable psychological state, tends to be lower in individuals with IGD, who also exhibit higher levels of depression [23,24]. Patients with posttraumatic stress disorder (PTSD), a representative anxiety disorder, are more likely to experience IGD, providing empirical support for the co-occurrence of these conditions [25,26]. While it remains unclear whether problematic gaming behavior is a cause or a consequence of weakened self-control, depression, and anxiety associated with COVID-19 and PTSD, the self-medication hypothesis is likely applicable to IGD, akin to its role in SUDs.

Research indicates that antidepressants like escitalopram and bupropion improve both IGD symptoms and depressive symptoms [27,28]. Such findings suggest that underlying causes may precede problematic gaming behavior. Previous studies have shown that bupropion is effective in managing attention and impulsivity, while escitalopram is effective in alleviating depressive symptoms. In particular, the findings from most studies suggesting that bupropion is more effective than escitalopram in improving symptoms of IGD [29,30] imply the necessity of managing the adverse effects of self-medication, such as reward effects of maladaptive behaviors and reduced self-control, regardless of the underlying causes driving problematic gaming behavior.

The similarities between IGD and SUD have been explored in various studies. Both conditions exhibit similar impulsivity traits [31], and neurological tests using functional magnetic resonance imaging have identified overlapping features [32]. However, individuals with SUD tend to have experienced more severe past adversities and exhibit more extreme problematic behaviors than those with IGD [33]. This distinction suggests that the two conditions may represent a continuum of self-medication choices, influenced by the severity of underlying causes. The high accessibility and immediate rewards of internet gaming may serve as a gateway, enabling individuals with relatively mild psychological distress, such as depression and anxiety, to easily experience reward effects and opt for gaming as a form of self-medication.

Application of CBT

Patients with IGD often hold game-related misconceptions, such as "Only in internet games do people recognize me." These distorted cognitions about gaming contribute to maladaptive behavioral patterns and cravings related to gaming [34]. CBT plays a role in correcting these erroneous beliefs that drive excessive gaming and managing the negative emotions triggered by game withdrawal [35]. Modifying cognitive distortions about gaming can regulate the severity of IGD symptoms and reduce cravings [36]. The application of CBT, originally developed for depression treatment, to various mental health disorders such as PTSD, anxiety disorders, and IGD highlights the shared and distinct features among these conditions and their connection to CBT elements. Therefore, analyzing IGD through the lens of CBT research [37] may provide new perspectives on traditional therapeutic approaches.

The theoretical background for the interaction between emotions and cognition in CBT is explained through the associative network theory [38]. CBT involves analyzing and treating the process through which individuals interpret current events based on "automatic thoughts" derived from past experiences. As illustrated in Figure 1, emotions and interpretations interact dynamically, with their sequential relationship reinforcing each other and leading to uncontrollable outcomes [39]. CBT addresses the vicious cycle of psychological distress by introducing alternative thoughts. These newly provided pieces of information compete with negative cognitions during the cognitive processing bottleneck, delaying the processing of negative thoughts and replacing them with alternative perspectives [40].

XSL•FO
RenderX

Figure 1. Associative network theory: in associative network theory, human memory, events, and emotions are represented as individual nodes. Each node is interconnected by associative links, and the activation of one node triggers the activation of related nodes.

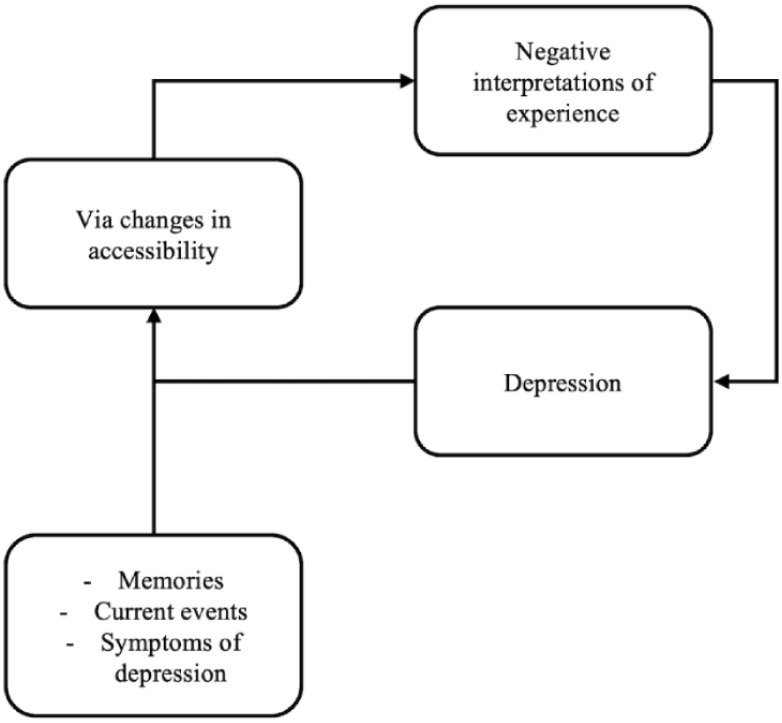

While associative network theory explains the theoretical basis for the activation of "automatic thoughts," it has limitations in capturing the complexity of human emotions. For instance, discrepancies may arise between rational beliefs and sensory perceptions. Furthermore, according to associative network theory, the emergence of depressive symptoms requires a triggering event that evokes past experiences and emotions. However, patients with depression often exhibit depressive symptoms not only as a result of current mood and stimuli but also due to an inherent cognitive vulnerability [41,42].

Application of MBCT

To explain complex emotions and cognitive vulnerability, the ICS theory has been proposed [43,44]. The model comprises nine subsystems—acoustic, visual, body-state, articulatory, limb, morphonolexical, object, propositional, and implicational—each responsible for processing specific types of information. Through their interactions, cognition and emotion emerge. Visual and auditory information are first interpreted semantically in isolation and then integrated into a specific meaning within the propositional subsystem. This general meaning is further integrated with body-state, visual, and auditory information, evolving into implicational meaning [40]. As illustrated in Figure 2, the interaction between cognition and emotion demonstrates that external stimuli and emotions are indirectly connected. Emotions are portrayed not as mere reactions to external stimuli but as outcomes reflecting cognitive characteristics, thereby explaining why individuals exhibit different responses to stressful situations. Figure 3 illustrates the process where specific meaning and general meaning influence and reinforce each other [43].

MBCT, based on the principles of the ICS, integrates the core mechanisms of CBT with stress-reduction techniques derived from mindfulness meditation. Originally developed for the treatment of recurrent depression, MBCT has demonstrated its efficacy over the past several decades [45-48]. Beyond depression, its effectiveness has been reported for other mental health conditions, including anxiety disorders [49-51], insomnia [52,53], bipolar disorder [54], panic disorder [55], chronic pain [56], cancer-related pain [57], and stress [58].

Key components such as mindfulness skills, depressogenic cognition, self-compassion and cognitive reactivity, and meta-awareness and decentering have been identified as factors associated with or mediating symptom reduction following MBCT [59]. In addition, MBCT has received significantly high ratings in patient self-assessments and has been shown to improve subscales such as "concerning danger" and "controllability of thoughts" [60].

The mindfulness meditation component of MBCT aims to stabilize thoughts by fostering a cognitive process of accepting present thoughts as internal experiences while shifting attention to external experiences by recognizing body-state cues rather than reacting to automatic thoughts. This dual focus promotes emotional regulation and cognitive stability [61].

XSL•FO
RenderX

Figure 2. Interactive cognitive subsystem.

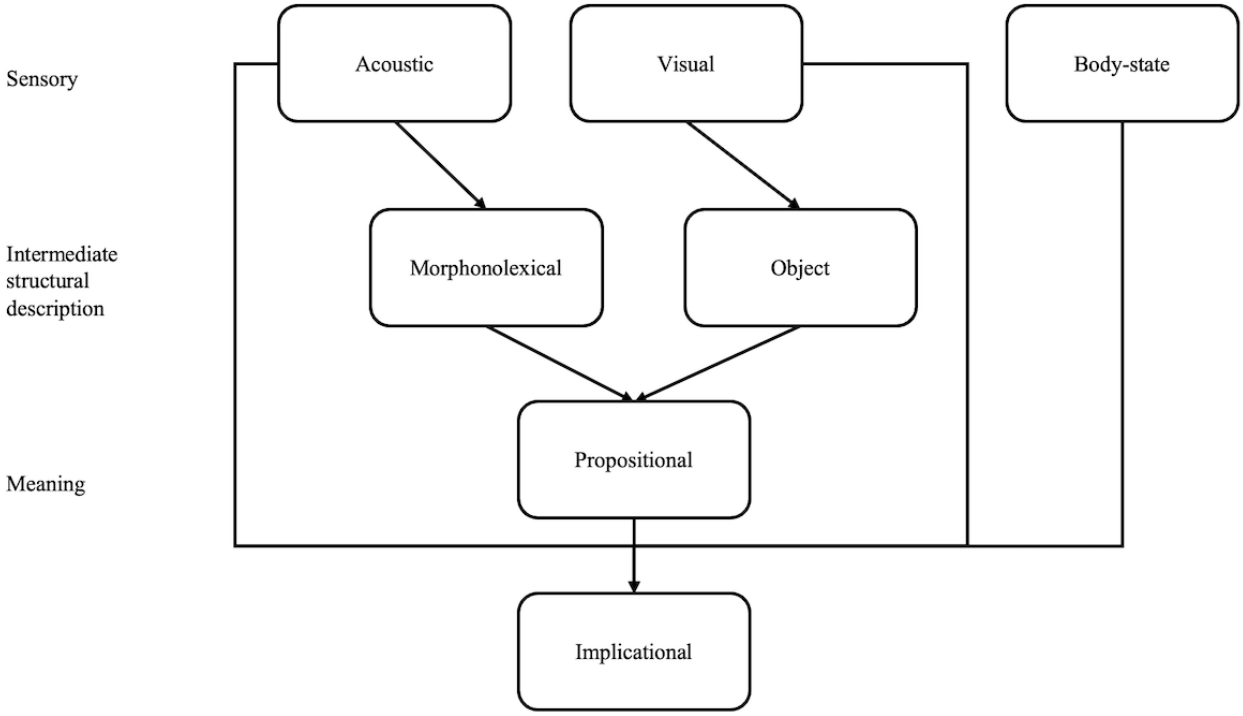

Figure 3. Vicious cycle of the interactive cognitive subsystem: specific meaning influences and reinforces general meaning in a cyclical process.

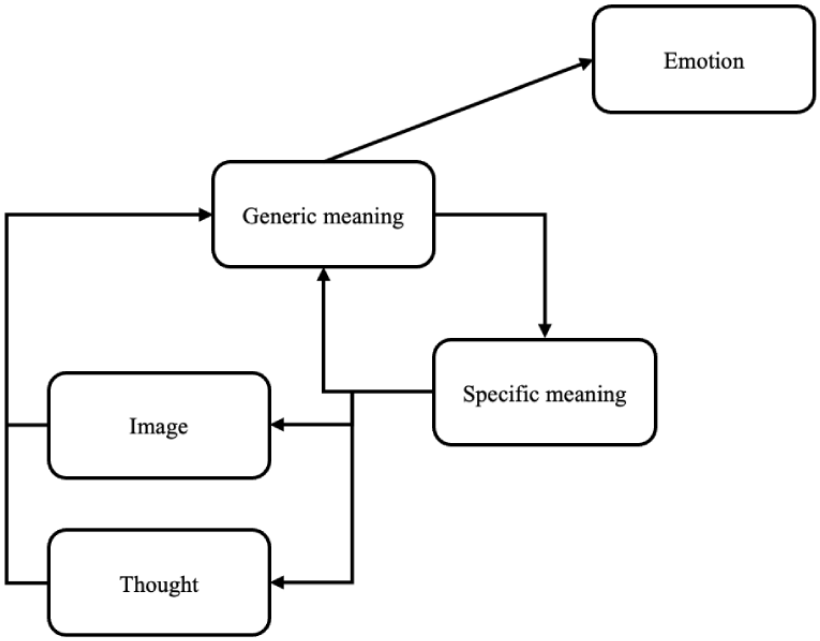

Nevertheless, several studies have reported no statistically significant differences between MBCT and CBT [60,62,63]. MBCT appears to demonstrate distinct therapeutic effects compared with CBT in patients who have experienced three or more episodes of depression recurrence [64]. However, these effects also seem to be short-term [60]. Despite previous research, the theoretical mechanisms underlying MBCT remain unclear [59].

The limitations of CBT and MBCT are particularly evident in newly defined mental health conditions such as IGD. The authors posit that, unlike the passive mental models described in CBT and MBCT, individuals actively adopt strategies like self-medication to alleviate psychological distress. Therefore, this paper introduces a modified ICS that integrates the self-medication hypothesis into existing cognitive models. This framework aims to develop a more realistic cognitive model, and the application of this model, referred to as MBCT-G, is explained in the following sections.

XSL•FO
RenderX

Application of MBCT-G

Modified ICS

The modified ICS focuses on patients' coping strategies, which were not emphasized in the original theoretical mechanisms. For example, a student with public speaking anxiety may experience situations where they "go blank," "cannot think of anything," or "feel mentally frozen" when required to speak in front of a large audience. While these responses may not be ideal, they can be understood as a form of self-medication strategy to cope with the stress of public speaking. By adopting such a strategy, the student may avert their gaze from the audience, quickly read from their script, or even leave the podium without remembering how they spoke, thereby managing the stress in the short term. However, in the long term, this strategy may reinforce their fear of public speaking, lower their self-esteem, and be perceived as an inadequate coping mechanism.

When discussing the mechanisms of cognition and emotion, it is essential to address not only the generation of emotions triggered by external and internal stimuli but also the resulting "reaction" that follows. In the previous example, the experience of giving a presentation is not limited to the emotions felt during the presentation itself but is significantly influenced by the memories and emotions generated during reflection after the situation has ended. Specifically, if an individual evaluates their coping strategy as sufficient and positive after the event, the previously generated emotions can be influenced and transformed into more positive ones. Therefore, the interpretation of emotions should not be confined to a single moment in time; instead, it must account for the distorted recollections formed through self-evaluation of the appropriateness of coping strategies, influenced by past experiences and emotions, as well as the emotions experienced in the present.

For instance, recalling negative experiences such as "feeling darkness before one's eyes," "not being able to think of anything," or "having a blank mind" during public speaking not only reinforces thoughts of psychological distress but also shapes negative self-assessments of the coping strategies used. These negative evaluations, in turn, intensify the negative emotions associated with the experience.

Figure 4 expands the ICS by incorporating the principles discussed earlier. The additional structure on the right in the modified ICS represents coping strategies and recollections related to emotions.

Figure 4. Modified interacting cognitive subsystem.

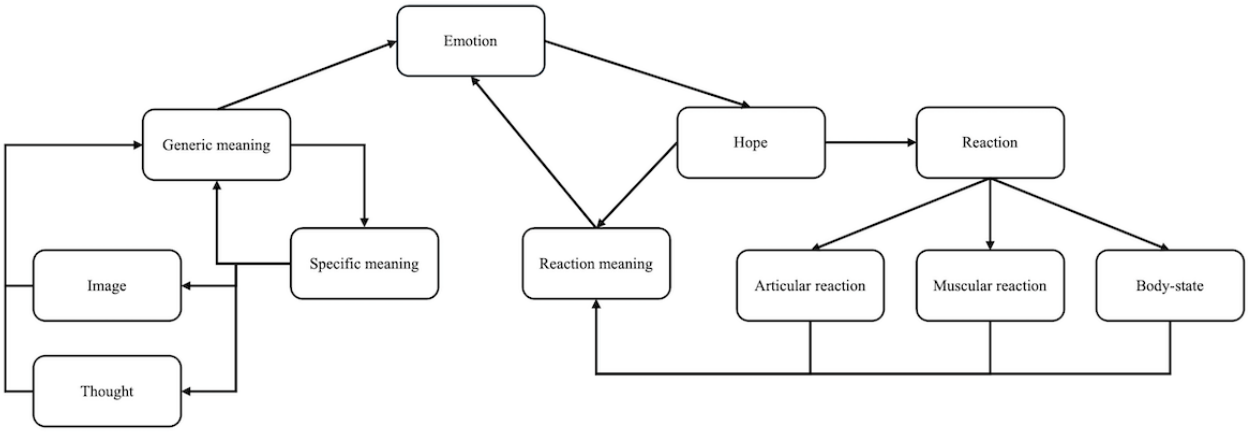

"Hope" refers to the desired strategy for coping with emotions generated in response to a stimulus. For instance, a student fearful of public speaking might think, "I want to finish my presentation quickly and leave," "I hope my turn gets skipped," or "I wish no one pays attention to my presentation."

"Reaction" represents the active coping mechanism employed in response to "Hope." In the given situation, since the external change (eg, skipping the student's turn) is beyond the student's control, they may choose an active strategy like "Let's finish the presentation quickly and step down."

"Reaction meaning" reflects the individual's self-evaluation of their chosen coping strategy. After adopting the strategy of finishing the presentation quickly, the student might evaluate themselves either positively or negatively. The outcome of this self-assessment is interpreted in the context of their fundamental "Hope" (eg, "I want to finish quickly"), which, in turn, influences the preceding emotions.

In simpler terms, if the strategy of speaking quickly satisfies the initial "Hope," the preceding "Emotion" will also be alleviated. As a result, the student may continue to rely on the "quick speech strategy" in future similar situations. On the other hand, if the strategy fails, the student might explore alternative strategies. For example, they may choose an active coping mechanism like "I want to relieve stress by playing a game." If the student successfully alleviates stress through gaming, the previously experienced "Emotion" is resolved, making gaming a preferred coping strategy in similar future scenarios.

In summary, these differences arise because the subjective interpretation of the emotional coping strategy referred to as "reaction meaning," creates a vicious cycle that influences the emotions resulting from preceding situations. In addition, past experiences, coping strategies, and emotions, such as "I can relieve stress through gaming," serve as materials for interpreting current emotions, as in "If I play games, my current

XSL•FO
RenderX

psychological distress will be alleviated." To reiterate, experiences, emotions, and coping strategies must be understood as part of a unified principle that unfolds across the flow of psychological time. This interpretation is particularly valuable in disorders where symptoms manifest in behavioral forms, such as IGD, gambling disorder, obsessive-compulsive disorder, and SUD. It offers a foundation for developing new therapeutic approaches.

How Can Games Be Used in Mental Counseling?

As previously mentioned, the gaming behavior of individuals with IGD can be interpreted as a coping strategy for managing stress resulting from external stimuli. Incorporating the self-medication hypothesis, which is often cited as a cause of SUD, the authors propose the self-prescription hypothesis as a potential explanation for the origins of behavioral addictions like IGD.

In the modified ICS, such coping strategies are effective, at least in the short term, in alleviating psychological distress (commonly referred to as a "reward"), serving as a driving force for the repetition of behavioral addictions. From this perspective, traditional treatments that focus solely on correcting maladaptive thoughts about gaming and reducing gaming time are often ineffective because they fail to provide alternative strategies for managing stress [65].

Treatments that do not offer alternative strategies may lead to symptom relapse or the development of other problematic behaviors, as the underlying cause—stress—remains unresolved. On the other hand, mindfulness meditation can serve as an effective alternative strategy. Activities like meditation and walking can help alleviate stress, reducing the perceived need for internet gaming. However, while these strategies are effective, they may not be sufficient for everyone. Some individuals may lack interest in mindfulness meditation, find it unhelpful in directly addressing their stress, or even perceive it as boring.

To address these limitations, three principles for effective psychological counseling are introduced:

1. Helping patients identify their automatic thoughts: This is the foundational principle of CBT, MBCT, and MBCT-G. Automatic thoughts are the key to uncovering the root causes of stress and form the basis of the therapeutic process. Recognizing these thoughts allows patients to understand the triggers for their psychological distress.
2. Providing alternative strategies to address the sources of stress: As previously discussed, coping strategies influence preceding emotions and play a critical role in stress alleviation. Without appropriate alternative strategies, mental health conditions may persist, worsen, or manifest in different forms of problematic behavior. Effective counseling must present actionable and relevant alternatives tailored to the individual's needs.
3. Ensuring the proposed alternative strategies are actionable and sustainable: Alternative strategies must be feasible, effective, and realistic for the patient. They should also offer long-term benefits, mindfulness effects, and ease of implementation. For instance, teaching a patient with public speaking anxiety how to "give a good presentation" may not be effective in all situations. These individuals may lack confidence in their presentation skills, but their anxiety might also stem from other causes such as interpersonal relationships, voice quality, or appearance. Simply exposing them repeatedly to presentation scenarios may not be a viable strategy. Moreover, if the patient finds the alternative strategy burdensome or rejects it, they are unlikely to sustain it, leading to cognitive strain during the therapeutic process.

By adhering to these principles, counseling can better address the complexity of individual needs and provide meaningful, sustainable solutions for stress management and mental health improvement.

MBCT-G is a therapeutic approach that incorporates the three principles outlined above. Games serve as a novel medium for experiencing virtual environments that are otherwise unattainable, making them one of the most accessible coping strategies for adolescents and young adults. While the accessibility of games can make even those with mild psychological distress susceptible to addiction, it paradoxically offers an ideal environment for learning healthy coping strategies.

The gaming environment allows individuals to repeatedly practice alternative strategies to address the sources of stress. Because this practice occurs in a virtual setting, it imposes lower cognitive demands. In addition, games can provide diverse scenarios tailored to symptoms and individual needs, while maintaining a consistent therapeutic framework. Furthermore, the gaming environment enables the collection of real-world data during the treatment process and sustains motivation for continued participation in therapy. This combination of features makes MBCT-G a compelling and effective approach to integrating mindfulness and cognitive therapy principles into an engaging and accessible medium.

Several studies have explored the use of games to address conditions such as IGD [66,67], depression [68], anxiety disorders [69], stress [70], attention-deficit/hyperactivity disorder [71], and executive function [72]. However, despite increasing interest in digital therapeutics and game-based treatment methods, few have been recognized by regulatory agencies for their therapeutic efficacy.

In the case of previous research on IGD, most studies have been limited to usability research without control groups [73]. The limitations of previous game-based studies include (1) a lack of understanding of cognition, emotion, and coping strategies, (2) an absence of defined reward mechanisms in games and corresponding alternative strategies, and (3) a poor understanding of game design quality.

A notable previous study using games to treat IGD is Room2Respawn, developed based on mindfulness-oriented recovery enhancement (MORE) [66]. MORE is a therapeutic approach that integrates three psychological counseling techniques: mindfulness, CBT, and positive psychology. It was originally designed for patients with chronic pain who are prescribed opioid analgesics [74]. The same research team investigated the efficacy of MORE in individuals with IGD. In

XSL•FO
RenderX

a randomized controlled trial targeting patients with IGD, MORE was found to reduce the severity of the disorder indirectly by correcting maladaptive gaming-related cognitions [36]. However, while Room2Respawn, developed using MORE, assessed the subjective usability of the intervention, it did not demonstrate the efficacy of implementing MORE techniques through gamified software.

How to Develop MBCT-G

Overview

In this paper, MBCT-G does not focus on a single game scenario or production method but addresses the overall factors to consider when using games for the treatment of mental health disorders. Therefore, the components of MBCT-G may differ from those of traditional MBCT.

For instance, MBCT for Life, an adaptation of MBCT designed for practical use in everyday life over an 8-week program, comprises the following sessions: (1) waking up from autopilot, (2) another way of being, (3) gathering the scattered mind, (4) recognizing reactivity, (5) allowing and letting be, (6) responding skillfully, (7) how can I best take care of myself? and (8) mindfulness for life [75]. While MBCT is a highly effective tool for alleviating stress, it requires the patient's active commitment to treatment. As a result, adherence may decrease in patients with cognitive impairments or low motivation for therapy.

The primary focus when designing game-based therapeutic interventions is to improve treatment adherence and provide opportunities for practicing alternative strategies. The key principles to consider when developing specific MBCT-G programs are outlined below.

Elements of MBCT-G

Fun

The first value of therapeutic approaches using games is fun. Educational games available on the market often emphasize the "educational" aspect while lacking gaming elements [76]. This tendency is influenced by factors such as the demands of purchasers (eg, parents), researchers' negative biases toward games, and a lack of understanding of gamification [77]. Contradictory game designs that disregard learners' characteristics such as gender and age can hinder therapeutic outcomes [70]. More fundamentally, fun is the most critical reason for applying gamification. Therefore, if concerns arise about the potential adverse effects of "fun games," researchers should reconsider the project from the outset.

Accessibility

Accessibility is a core value of digital therapeutics, including games. As mentioned earlier, in the process of selecting self-prescription methods to address psychological distress, the most important factors are "alleviating psychological distress" and "ease of access." From this perspective, internet-based games can serve as attractive tools for both the onset and treatment of disorders. Considering the rapid growth of mobile game platforms [78], excessive environmental or time-related constraints on games should be carefully addressed.

Reality Projection

Providing an environment where patients can project themselves onto MBCT-G is a topic to consider during the design phase. MBCT-G aims to help patients identify automatic thoughts, correct distorted beliefs about past and present events, and learn healthy coping strategies through experiences and lessons in the virtual world at specific points during gameplay. A game design that fails to reflect reality risks obscuring its therapeutic effects. For example, walking meditation is a core mindfulness practice, but merely watching a character walk on a screen would not yield therapeutic benefits for the patient. To project reality, an intersection between virtual data and real-world information is necessary. This could be as simple as incorporating the patient's personal information into the game or as complex as creating a sense of connection between the virtual character and the patient's physical body [79].

Discovery of Automatic Thoughts

Identifying distorted beliefs is a critical process in cognitive therapy, and its importance is equally valid in MBCT-G. However, even for professionally trained therapists, uncovering automatic thoughts can be challenging. Therefore, we can use various strategies. First, providing assurance of anonymity can reduce hesitation in sharing personal information [80]. In addition, counseling can be conducted using characters familiar to the patient [81]. Research into new artificial intelligence (AI) technologies is needed to uncover hidden meanings based on counseling content, and offline psychological counseling can be integrated to achieve accurate diagnoses.

Experiencing Automatic Thoughts

Experiencing automatic thoughts is one of the distinctive features of MBCT-G. Traditionally, psychological counseling has primarily relied on verbal communication. However, games allow for the direct discovery of distorted beliefs rooted in past experiences. With advancements in AI, various technologies mimicking human society have recently been developed [82]. A key characteristic of these technologies is their ability to form humanlike relationships through the exchange of information and emotions among in-game agents. Researchers, through unpublished studies, have demonstrated that by combining AI technologies with detailed counseling data, past experiences that contributed to psychological distress can be recreated within a game. By leveraging AI, patients are given the opportunity to directly experience distorted beliefs through the reactions of agents that vary depending on the patient's coping strategies.

Experiencing Alternative Strategies in Games

Even when healthy coping strategies to alleviate psychological distress are recommended through counseling, many barriers exist before patients can take actionable steps. Digital technologies, especially virtual reality, provide an environment where new strategies can be vividly explored with minimal effort. For instance, while activities like traveling have often been suggested as accessible alternative strategies during psychological counseling, numerous obstacles have prevented patients from putting such advice into action. In such cases, games can allow for an indirect experience of travel, and if this virtual travel helps alleviate psychological distress, it could

XSL•FO
RenderX

serve as motivation for patients to adopt stress-relief strategies through real-world travel. The critical point is that the experience of alternative strategies in MBCT-G should seamlessly connect to real-life applications. The ability to experience alternative strategies is one of the fundamental differences between traditional psychological counseling and MBCT-G.

Opportunities for Reflection on Insights

Despite the aforementioned elements, the gamified features of MBCT-G may pose a risk of patients becoming overly immersed in entertainment aspects, potentially overlooking the therapeutic elements. Therefore, MBCT-G should consider providing opportunities for patients to reflect on the lessons and experiences they learn during gameplay through mindfulness meditation.

Group Therapy

Group therapy offers advantages such as allowing patients with similar conditions to share their experiences and deepen their understanding of their illnesses. Its effectiveness has been demonstrated across various conditions [83,84]. However, challenges in forming and maintaining groups make physical group therapy sessions difficult to implement [85]. Digital technologies provide diverse ways to facilitate group therapy, both directly and indirectly. For instance, augmented reality technologies like Niantic's Pokémon GO enable patients to communicate indirectly by creating personal experiences and thoughts in the form of agents on a map [86].

The differences between traditional therapies and MBCT-G are compared in Figure 5. Traditional CBT methods are primarily language-based, and while they can propose alternative strategies, they do not provide opportunities for patients to practice these strategies during the treatment process. Existing game-based therapies allow patients to have various experiences in virtual environments; however, they often follow standardized therapeutic procedures, neglecting individual characteristics [87,88]. Also, an example of MBCT-G developed using the 8 elements of MBCT-G can be found in Multimedia Appendix 1.

Figure 5. Comparison between MBCT-G and traditional treatment methods. CBT: cognitive behavioral therapy; MBCT-G: mindfulness-based cognitive therapy–game.

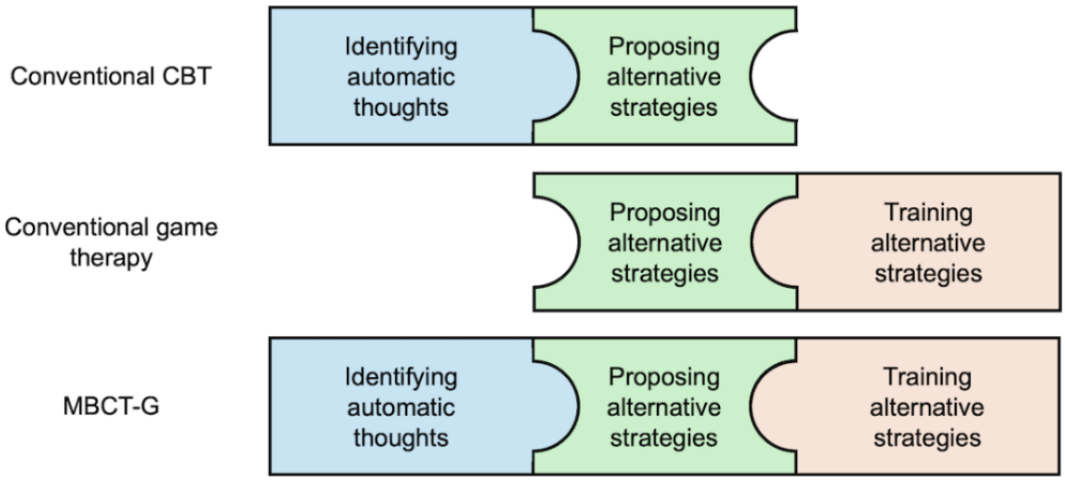

Therapeutic Goals and Evaluation of MBCT-G

Internet games, in themselves, cannot be universally labeled as harmful, and merely reducing the time spent playing games may not correlate strongly with the disease's pathology [89]. Therefore, the primary goal of MBCT-G is to restore significant impairments in daily life, as subjectively perceived by patients and their surrounding individuals.

This goal can be assessed using various methods, including the widely adopted "IGD-20 Test" and other established research tools [90]. In addition, comorbid conditions frequently associated with IGD, such as depression [91], anxiety disorders [92], and attention-deficit/hyperactivity disorder [93], should also be evaluated. Since MBCT-G focuses on the alleviation of subjective symptoms, assessments based on the patient's counseling history are particularly important.

Advances in AI offer the potential to evaluate disorders using both verbal and nonverbal information [94-96]. AI-based analytic insights should be provided to psychological counselors to facilitate a complementary relationship between technology and human expertise. Furthermore, measuring the frequency and duration of MBCT-G usage is essential for assessing treatment adherence.

Finally, evaluations should include attempts and applications of alternative strategies. These evaluations should be approached from two perspectives:

1. Active engagement with alternative strategies within MBCT-G.
2. Implementation of strategies practiced in the virtual environment in the real world.

Critically, assessing whether patients are adopting healthy coping strategies outside the game environment is the most important therapeutic goal of MBCT-G.

Five Stages of MBCT-G

Overview

Finally, the 5 stages of MBCT-G are explained. With technological advancements, new technologies such as

XSL•FO
RenderX

autonomous driving, AI, and virtual reality are emerging. MBCT-G aims to create scenarios where patients can practice alternative strategies in a virtual environment, tailored to their treatment motivation and the causes of their condition. However, technological complexity and legal regulations pose obstacles to realizing these strategies. Nonetheless, to advance MBCT-G, it is important to start small and progressively implement these strategies according to the current level of technology (see Figure 6).

Figure 6. The 5 stages of mindfulness-based cognitive therapy-game.

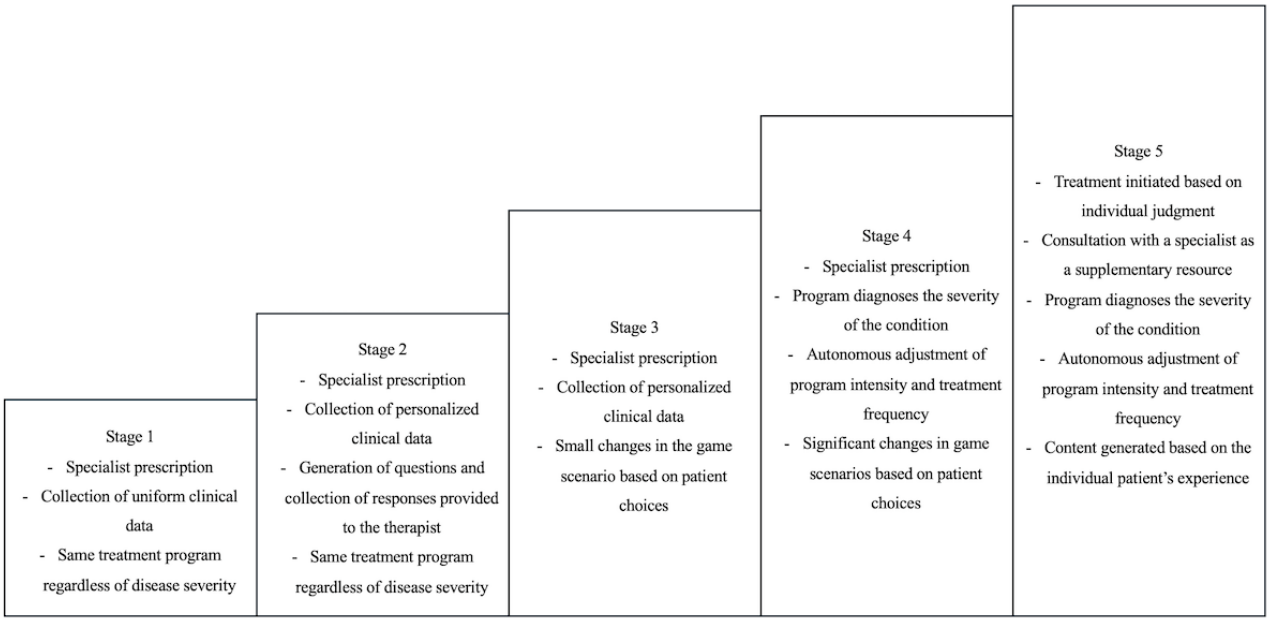

Stage 1

In the first stage of MBCT-G, the patient visits a hospital, receives a prescription from a psychiatrist, and begins treatment. While clinical data are collected through the MBCT-G program, the program's difficulty level and scenarios are adjusted by the specialist during the hospital visit.

Stage 2

In the second stage, the MBCT-G program generates personalized questions based on the clinical data collected in stage 1 and asks the patient these questions. Although all patients collect the same clinical data in stage 1, from stage 2 onward, the program gathers tailored clinical data specific to each patient.

Stage 3

In the third stage, the MBCT-G program diagnoses the severity of the condition in a limited capacity and modifies the game scenarios based on the patient's choices to a restricted extent.

Stage 4

In the fourth stage, the MBCT-G program independently diagnoses the severity of the condition and adjusts the intensity of the program and the frequency of sessions. Patients can observe significant changes in the game scenarios based on their choices, allowing them to immerse themselves more realistically and practice alternative strategies.

Stage 5

In the final stage, the MBCT-G program autonomously creates or reconstructs scenarios based on the clinical data collected from the patient. The patient repeatedly trains in virtual reality to address the situations that cause gaming disorder. In stage 5, both diagnosis and treatment are autonomously conducted by the program, with the psychiatrist intervening only in exceptional situations.

Summary

This paper describes the theoretical background and components of a novel approach called MBCT-G for treating IGD. To achieve this, it elaborates on the conceptual expansion of the self-medication hypothesis, originally proposed as one of the causes of SUDs, and integrates it with the cognitive models of CBT and MBCT. The self-prescription hypothesis and modified ICS presented in this paper can be interpreted as efforts to maintain a form of mental homeostasis.

Homeostasis refers to the property of maintaining a stable and consistent internal environment. Typical examples of homeostasis include maintaining a constant body temperature and regulating the osmotic pressure of the blood. In situations where stress arises due to internal or external stimuli, the human body adjusts factors such as heart rate, respiration, blood pressure, and blood sugar levels to sustain homeostasis [97].

The conceptual terms presented in this paper require further validation in clinical settings. First, previous studies on the relationship among the reward effects, reduced self-control, and craving, as described in the self-medication hypothesis, yield varying results depending on the experimental design [98,99]. Furthermore, the ICS is merely one of several explanations of cognitive structures. These discrepancies stem from the ongoing debate—akin to the "chicken or the egg" question—caused by the diverse comorbidities and complications associated with mental disorders. What is clear, however, is that traditional

XSL•FO
RenderX

psychological counseling techniques have been ineffective for behavioral addiction disorders like IGD, underscoring the need for novel therapeutic approaches and theoretical foundations. To address these challenges, this paper proposes the conceptual therapeutic approach and rationale of MBCT-G, based on previous studies and clinical experience.

The concepts introduced in this paper may not necessarily be entirely new. CBT and MBCT emphasize that the present is interpreted based on past events and emotions. Previous research briefly explains the "vicious cycle" where body states, such as trembling hands, a hunched posture, or facial grimacing, change according to emotions, and these changes, in turn, affect emotions [43]. As shown in Figure 2, active behavior and body states influenced by emotions can also affect thoughts and feelings, a notion aligned with the ideas behind MBCT-G. However, in situations of psychological distress, patients' active behavioral strategies are far more critical than previous studies suggest. Despite this, discussions about such ideas are sparse and have not been incorporated into psychological counseling techniques. Thus, there was a need to emphasize active behavioral strategies in response to psychological distress, which had previously been overlooked, and to incorporate these strategies into the therapeutic process [100].

Theories about the motivation for internet gaming are diverse and extend beyond the self-prescription hypothesis. Internet gaming is known to be driven primarily by the motivations of "escape" and "competition," which serve as key mediating factors and are reported to directly or indirectly influence the symptoms of IGD.

Figure 7 is a visualization of the development process of IGD. Maladaptive thoughts about gaming can arise from various causes. Patients with IGD can be differentiated based on factors such as self-esteem, achievement, and coping strategies [101]. This differentiation suggests that the reasons for gaming addiction are diverse, including "enjoyment of the game itself," "goal achievement," and "stress relief." Therefore, the self-prescription hypothesis proposed in this paper may have limitations in analyzing all motivational factors. However, the various grounds for the self-prescription hypothesis related to the characteristics of IGD and the need for new psychological counseling techniques highlight the importance of continued research in this area.

Figure 7. The development process of internet gaming disorder: the tolerance and withdrawal characteristics associated with internet gaming trigger cravings for gaming.

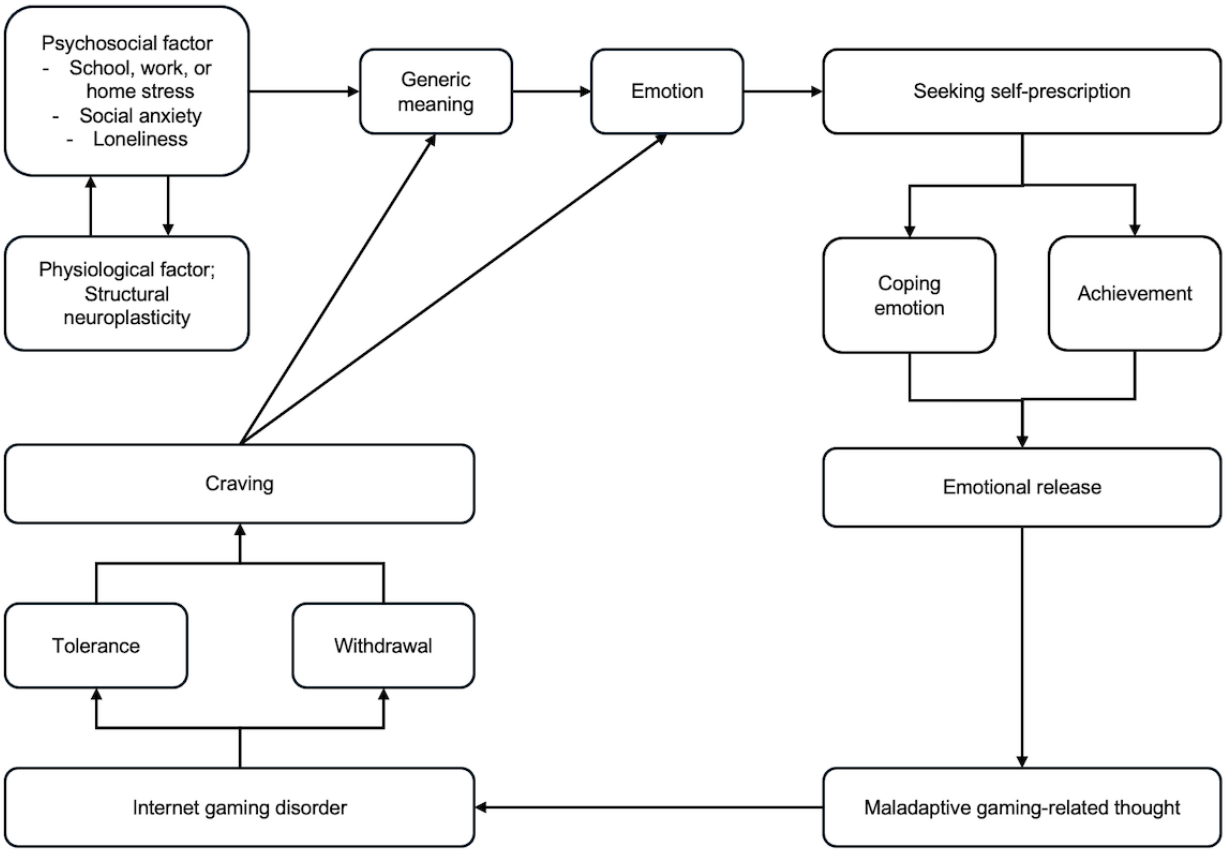

Digital therapeutics refer to medical approaches that use software and hardware to prevent, treat, and manage physical and mental illnesses. This new approach to disease management and treatment offers several advantages. First, it allows patients to receive treatment without the need to visit hospitals or counseling centers, providing freedom from time and space constraints. Second, online mental health treatment can eliminate the stigma and negative perceptions that patients might otherwise face. Third, because sessions can be conducted without the involvement of a physician or counselor, it is easier to monitor treatment adherence, and the approach is cost-effective [73]. Games are a type of digital therapeutic method, and the enjoyment they provide can be particularly effective for patients lacking motivation for treatment. This is

XSL•FO
RenderX

especially true for adolescents, who exhibit the highest prevalence of IGD and simultaneously have a strong affinity for games and digital devices, making game-based approaches highly effective.

Considering that patients with IGD often choose gaming as the most effective means of alleviating psychological distress, teaching them to repurpose gaming, which has been used inappropriately, into a tool for positive life outcomes may ironically prove to be the most effective therapeutic method.

Conclusion

IGD affects 3% of the population, and with technological advancements, more individuals are at risk of developing this condition. However, there is currently no clear treatment available. IGD is not a primary disease but rather a result of "self-prescription" in response to emotional stressors. Therefore, unlike traditional mental health treatments that focus on the disorder itself, it is essential to provide alternative activities that can alleviate negative emotions. This paper introduces MBCT-G as an alternative activity. MBCT-G is a program designed to explore alternative activities through gaming, focusing on the reaction and reward processes, areas that traditional treatments have not emphasized. This paper serves as the first theoretical foundation for the development of MBCT-G. MBCT-G is necessary as a method to provide alternative emotional relief strategies for treating self-prescription disorders such as IGD.

Acknowledgments

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Graduate School of Metaverse Convergence support program (IITP-2025-RS-2023-00254129), supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP) and Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2025 (project name: Development of game-based digital Therapeutics technology for adolescent mental health [psychological and behavioral control] management; project: RS-2025-00344893) and the MSIT, Korea, under the Global Scholars Invitation Program (RS-2025-00459638) supervised by the IITP.

Conflicts of Interest

None declared.

Multimedia Appendix 1
An example of mindfulness-based cognitive therapy–game.
[DOCX File , 203 KB - jmir_v27i1e65786_app1.docx]

References

1. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry 2021;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
2. Gao YX, Wang JY, Dong GH. The prevalence and possible risk factors of internet gaming disorder among adolescents and young adults: systematic reviews and meta-analyses. J Psychiatr Res 2022;154:35-43. [doi: 10.1016/j.jpsychires.2022.06.049] [Medline: 35926424]
3. Kim HS, Son G, Roh EB, Ahn WY, Kim J, Shin SH, et al. Prevalence of gaming disorder: a meta-analysis. Addict Behav 2022;126:107183. [doi: 10.1016/j.addbeh.2021.107183] [Medline: 34864436]
4. Zhou R, Morita N, Ogai Y, Saito T, Zhang X, Yang W, et al. Meta-analysis of internet gaming disorder prevalence: assessing the impacts of DSM-5 and ICD-11 diagnostic criteria. Int J Environ Res Public Health 2024;21(6):700 [FREE Full text] [doi: 10.3390/ijerph21060700] [Medline: 38928946]
5. Bickham DS. Current research and viewpoints on internet addiction in adolescents. Curr Pediatr Rep 2021;9(1):1-10 [FREE Full text] [doi: 10.1007/s40124-020-00236-3] [Medline: 33457108]
6. King DL, Delfabbro PH, Wu AMS, Doh YY, Kuss DJ, Pallesen S, et al. Treatment of internet gaming disorder: an international systematic review and CONSORT evaluation. Clin Psychol Rev 2017;54:123-133. [doi: 10.1016/j.cpr.2017.04.002] [Medline: 28458097]
7. Winkler A, Dörsing B, Rief W, Shen Y, Glombiewski JA. Treatment of internet addiction: a meta-analysis. Clin Psychol Rev 2013;33(2):317-329. [doi: 10.1016/j.cpr.2012.12.005] [Medline: 23354007]
8. Saunders JB, Hao W, Long J, King DL, Mann K, Fauth-Bühler M, et al. Gaming disorder: its delineation as an important condition for diagnosis, management, and prevention. J Behav Addict 2017;6(3):271-279 [FREE Full text] [doi: 10.1556/2006.6.2017.039] [Medline: 28816494]
9. King DL, Delfabbro PH. Internet gaming disorder treatment: a review of definitions of diagnosis and treatment outcome. J Clin Psychol 2014;70(10):942-955. [doi: 10.1002/jclp.22097] [Medline: 24752874]

XSL•FO
RenderX

10. Balhara YPS, Singh S, Saini R, Dahiya N, Singh AB, Kumar R. Should internet gaming disorder be considered a subtype of generalized problematic internet use? Findings from a study among medical college students. Perspect Psychiatr Care 2021;57(1):272-278. [doi: 10.1111/ppc.12558] [Medline: 32557655]
11. Khantzian EJ. The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. Am J Psychiatry 1985;142(11):1259-1264. [doi: 10.1176/ajp.142.11.1259] [Medline: 3904487]
12. Khantzian EJ. The self-medication hypothesis of substance use disorders: a reconsideration and recent applications. Harv Rev Psychiatry 1997;4(5):231-244. [doi: 10.3109/10673229709030550] [Medline: 9385000]
13. Burleigh TL, Griffiths MD, Sumich A, Stavropoulos V, Kuss DJ. A systematic review of the co-occurrence of gaming disorder and other potentially addictive behaviors. Curr Addict Rep 2019;6(4):383-401. [doi: 10.1007/s40429-019-00279-7]
14. Moore S, Satel J, Pontes HM. Investigating the role of health factors and psychological well-being in gaming disorder. Cyberpsychol Behav Soc Netw 2022;25(2):94-100. [doi: 10.1089/cyber.2021.0050] [Medline: 34788152]
15. Footitt T, Christofi N, Poulus DR, Colder Carras M, Stavropoulos V. Gaming disorder: the role of a gamers flow profile. Addict Behav Rep 2024;19:100555 [FREE Full text] [doi: 10.1016/j.abrep.2024.100555] [Medline: 38952851]
16. Hofstedt A, Söderpalm Gordh A. Young and adult patients with gaming disorder: psychiatric co-morbidities and progression of problematic gaming. Front Psychiatry 2024;15:1423785 [FREE Full text] [doi: 10.3389/fpsyt.2024.1423785] [Medline: 39720435]
17. Balhara YPS, Garg H, Kumar S, Bhargava R. Gaming disorder as a consequence of attempt at self- medication: empirical support to the hypothesis. Asian J Psychiatr 2018;31:98-99. [doi: 10.1016/j.ajp.2018.02.013] [Medline: 29453152]
18. Oka T, Hamamura T, Miyake Y, Kobayashi N, Honjo M, Kawato M, et al. Prevalence and risk factors of internet gaming disorder and problematic internet use before and during the COVID-19 pandemic: a large online survey of Japanese adults. J Psychiatr Res 2021;142:218-225 [FREE Full text] [doi: 10.1016/j.jpsychires.2021.07.054] [Medline: 34385071]
19. Teng Z, Pontes HM, Nie Q, Griffiths MD, Guo C. Depression and anxiety symptoms associated with internet gaming disorder before and during the COVID-19 pandemic: a longitudinal study. J Behav Addict 2021;10(1):169-180 [FREE Full text] [doi: 10.1556/2006.2021.00016] [Medline: 33704085]
20. Perez C, Alvarez JJ, Carbungco A, Due J, Ochoa CM, Celis ML, et al. Relationship between online game addiction and mental well-being of high-school students during the Covid-19 pandemic: implications for learning and development. J Learn Dev 2024;11(2):289-303. [doi: 10.56059/jl4d.v11i2.1101]
21. Guo Y, Yue F, Lu X, Sun F, Pan M, Jia Y. COVID-19-related social isolation, self-control, and internet gaming disorder among Chinese university students: cross-sectional survey. J Med Internet Res 2024;26:e52978 [FREE Full text] [doi: 10.2196/52978] [Medline: 39255486]
22. Elhai JD, McKay D, Yang H, Minaya C, Montag C, Asmundson GJG. Health anxiety related to problematic smartphone use and gaming disorder severity during COVID-19: fear of missing out as a mediator. Hum Behav Emerg Technol 2021;3(1):137-146 [FREE Full text] [doi: 10.1002/hbe2.227] [Medline: 33363275]
23. Yen JY, Lin HC, Chou WP, Liu TL, Ko CH. Associations among resilience, stress, depression, and internet gaming disorder in young adults. Int J Environ Res Public Health 2019;16(17):3181 [FREE Full text] [doi: 10.3390/ijerph16173181] [Medline: 31480445]
24. Lee JY, Choi JS, Kwon JS. Neurophysiological mechanisms of resilience as a protective factor in patients with internet gaming disorder: a resting-state EEG coherence study. J Clin Med 2019;8(1):49 [FREE Full text] [doi: 10.3390/jcm8010049] [Medline: 30621356]
25. Yuan GF, Shi W, Elhai JD, Montag C, Chang K, Jackson T, et al. Gaming to cope: applying network analysis to understand the relationship between posttraumatic stress symptoms and internet gaming disorder symptoms among disaster-exposed Chinese young adults. Addict Behav 2022;124:107096. [doi: 10.1016/j.addbeh.2021.107096] [Medline: 34469784]
26. Liu Q, Ouyang L, Fan L, Liao A, Li Z, Chen X, et al. Association between childhood trauma and internet gaming disorder: a moderated mediation analysis with depression as a mediator and psychological resilience as a moderator. BMC Psychiatry 2024;24(1):412 [FREE Full text] [doi: 10.1186/s12888-024-05863-4] [Medline: 38834952]
27. de Sá RRCD, Coelho S, Parmar PK, Johnstone S, Kim HS, Tavares H. A systematic review of pharmacological treatments for internet gaming disorder. Psychiatry Investig 2023;20(8):696-706 [FREE Full text] [doi: 10.30773/pi.2022.0297] [Medline: 37559452]
28. Seo EH, Yang HJ, Kim SG, Park SC, Lee SK, Yoon HJ. A literature review on the efficacy and related neural effects of pharmacological and psychosocial treatments in individuals with internet gaming disorder. Psychiatry Investig 2021;18(12):1149-1163 [FREE Full text] [doi: 10.30773/pi.2021.0207] [Medline: 34872237]
29. Song J, Park JH, Han DH, Roh S, Son JH, Choi TY, et al. Comparative study of the effects of bupropion and escitalopram on internet gaming disorder. Psychiatry Clin Neurosci 2016;70(11):527-535 [FREE Full text] [doi: 10.1111/pcn.12429] [Medline: 27487975]
30. Nam B, Bae S, Kim SM, Hong JS, Han DH. Comparing the effects of bupropion and escitalopram on excessive internet game play in patients with major depressive disorder. Clin Psychopharmacol Neurosci 2017;15(4):361-368 [FREE Full text] [doi: 10.9758/cpn.2017.15.4.361] [Medline: 29073748]

XSL•FO
RenderX

31. Choi SW, Kim HS, Kim GY, Jeon Y, Park SM, Lee JY, et al. Similarities and differences among internet gaming disorder, gambling disorder and alcohol use disorder: a focus on impulsivity and compulsivity. J Behav Addict 2014;3(4):246-253 [FREE Full text] [doi: 10.1556/JBA.3.2014.4.6] [Medline: 25592310]
32. Zheng YB, Dong HH, Wang M, Zhou W, Lin X, Dong GH. Similarities and differences between internet gaming disorder and tobacco use disorder: a large-scale network study. Addict Biol 2022;27(2):e13119. [doi: 10.1111/adb.13119] [Medline: 34913220]
33. Ong RHS, Peh CX, Guo S. Differential risk factors associated with adolescent addictive disorders: a comparison between substance use disorders and internet/gaming addiction. Int J Ment Health Addiction 2016;14(6):993-1002. [doi: 10.1007/s11469-016-9676-0]
34. Chen Y, Lu J, Wang L, Gao X. Effective interventions for gaming disorder: a systematic review of randomized control trials. Front Psychiatry 2023;14:1098922 [FREE Full text] [doi: 10.3389/fpsyt.2023.1098922] [Medline: 36815197]
35. Dong GH, Dai J, Potenza MN. Ten years of research on the treatments of internet gaming disorder: a scoping review and directions for future research. J Behav Addict 2024;13(1):51-65 [FREE Full text] [doi: 10.1556/2006.2023.00071] [Medline: 38183434]
36. Li W, Garland EL, Howard MO. Therapeutic mechanisms of mindfulness-oriented recovery enhancement for internet gaming disorder: reducing craving and addictive behavior by targeting cognitive processes. J Addict Dis 2018;37(1-2):5-13. [doi: 10.1080/10550887.2018.1442617] [Medline: 29565776]
37. Beck AT, Rush AJ, Shaw BF, Emery G, DeRubeis RJ, Hollon SD. Cognitive Therapy of Depression. New York: Guilford Publications; 2024.
38. Bower GH. Mood and memory. Am Psychol 1981;36(2):129-148. [doi: 10.1037//0003-066x.36.2.129] [Medline: 7224324]
39. Teasdale JD. Negative thinking in depression: cause, effect, or reciprocal relationship? Adv Behav Res Ther 1983;5(1):3-25. [doi: 10.1016/0146-6402(83)90013-9]
40. Teasdale JD, Segal Z, Williams JG. How does cognitive therapy prevent depressive relapse and why should attentional control (mindfulness) training help? Behav Res Ther 1995;33(1):25-39. [doi: 10.1016/0005-7967(94)e0011-7] [Medline: 7872934]
41. Miranda J, Persons JB, Byers CN. Endorsement of dysfunctional beliefs depends on current mood state. J Abnorm Psychol 1990;99(3):237-241. [doi: 10.1037//0021-843x.99.3.237] [Medline: 2212273]
42. Teasdale JD, Dent J. Cognitive vulnerability to depression: an investigation of two hypotheses. Br J Clin Psychol 1987;26(2):113-126. [doi: 10.1111/j.2044-8260.1987.tb00737.x] [Medline: 3580646]
43. Teasdale JD. Emotion and two kinds of meaning: cognitive therapy and applied cognitive science. Behav Res Ther 1993;31(4):339-354. [doi: 10.1016/0005-7967(93)90092-9] [Medline: 8512536]
44. Barnard PJ, Teasdale JD. Interacting cognitive subsystems: a systemic approach to cognitive-affective interaction and change. Cogn Emot 1991;5(1):1-39. [doi: 10.1080/026999391084110 21]
45. Manicavasgar V, Parker G, Perich T. Mindfulness-based cognitive therapy vs cognitive behaviour therapy as a treatment for non-melancholic depression. J Affect Disord 2011;130(1-2):138-144. [doi: 10.1016/j.jad.2010.09.027] [Medline: 21093925]
46. Manicavasagar V, Perich T, Parker G. Cognitive predictors of change in cognitive behaviour therapy and mindfulness-based cognitive therapy for depression. Behav Cogn Psychother 2012;40(2):227-232. [doi: 10.1017/S1352465811000634] [Medline: 22017810]
47. Fjorback LO, Arendt M, Ornbøl E, Fink P, Walach H. Mindfulness-based stress reduction and mindfulness-based cognitive therapy: a systematic review of randomized controlled trials. Acta Psychiatr Scand 2011;124(2):102-119. [doi: 10.1111/j.1600-0447.2011.01704.x] [Medline: 21534932]
48. Kuyken W, Byford S, Taylor RS, Watkins E, Holden E, White K, et al. Mindfulness-based cognitive therapy to prevent relapse in recurrent depression. J Consult Clin Psychol 2008;76(6):966-978. [doi: 10.1037/a0013786] [Medline: 19045965]
49. Green SM, Bieling PJ. Expanding the scope of mindfulness-based cognitive therapy: evidence for effectiveness in a heterogeneous psychiatric sample. Cogn Behav Pract 2012;19(1):174-180. [doi: 10.1016/j.cbpra.2011.02.006]
50. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
51. Evans S. Mindfulness-based cognitive therapy for generalized anxiety disorder. In: Mindfulness-Based Cognitive Therapy. Cham: Springer; 2016:145-154.
52. Heidenreich T, Tuin I, Pflug B, Michal M, Michalak J. Mindfulness-based cognitive therapy for persistent insomnia: a pilot study. Psychother Psychosom 2006;75(3):188-189. [doi: 10.1159/000091778] [Medline: 16636636]
53. Yook K, Lee SH, Ryu M, Kim KH, Choi TK, Suh SY, et al. Usefulness of mindfulness-based cognitive therapy for treating insomnia in patients with anxiety disorders: a pilot study. J Nerv Ment Dis 2008;196(6):501-503. [doi: 10.1097/NMD.0b013e31817762ac] [Medline: 18552629]
54. Williams JMG, Alatiq Y, Crane C, Barnhofer T, Fennell MJV, Duggan DS, et al. Mindfulness-based cognitive therapy (MBCT) in bipolar disorder: preliminary evaluation of immediate effects on between-episode functioning. J Affect Disord 2008;107(1-3):275-279 [FREE Full text] [doi: 10.1016/j.jad.2007.08.022] [Medline: 17884176]

XSL•FO
RenderX

55. Kim YW, Lee SH, Choi TK, Suh SY, Kim B, Kim CM, et al. Effectiveness of mindfulness-based cognitive therapy as an adjuvant to pharmacotherapy in patients with panic disorder or generalized anxiety disorder. Depress Anxiety 2009;26(7):601-606. [doi: 10.1002/da.20552] [Medline: 19242985]
56. Grossman P, Tiefenthaler-Gilmer U, Raysz A, Kesper U. Mindfulness training as an intervention for fibromyalgia: evidence of postintervention and 3-year follow-up benefits in well-being. Psychother Psychosom 2007;76(4):226-233. [doi: 10.1159/000101501] [Medline: 17570961]
57. Ledesma D, Kumano H. Mindfulness-based stress reduction and cancer: a meta-analysis. Psychooncology 2009;18(6):571-579. [doi: 10.1002/pon.1400] [Medline: 19023879]
58. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
59. van der Velden AM, Kuyken W, Wattar U, Crane C, Pallesen KJ, Dahlgaard J, et al. A systematic review of mechanisms of change in mindfulness-based cognitive therapy in the treatment of recurrent major depressive disorder. Clin Psychol Rev 2015;37:26-39 [FREE Full text] [doi: 10.1016/j.cpr.2015.02.001] [Medline: 25748559]
60. Külz AK, Landmann S, Cludius B, Rose N, Heidenreich T, Jelinek L, et al. Mindfulness-based cognitive therapy (MBCT) in patients with obsessive-compulsive disorder (OCD) and residual symptoms after cognitive behavioral therapy (CBT): a randomized controlled trial. Eur Arch Psychiatry Clin Neurosci 2019;269(2):223-233. [doi: 10.1007/s00406-018-0957-4] [Medline: 30446822]
61. Kabat-Zinn J. Full Catastrophe Living. London: Bantam; 2013.
62. Sverre KT, Nissen ER, Farver-Vestergaard I, Johannsen M, Zachariae R. Comparing the efficacy of mindfulness-based therapy and cognitive-behavioral therapy for depression in head-to-head randomized controlled trials: a systematic review and meta-analysis of equivalence. Clin Psychol Rev 2023;100:102234 [FREE Full text] [doi: 10.1016/j.cpr.2022.102234] [Medline: 36527794]
63. Chiesa A, Serretti A. Mindfulness based cognitive therapy for psychiatric disorders: a systematic review and meta-analysis. Psychiatry Res 2011;187(3):441-453. [doi: 10.1016/j.psychres.2010.08.011] [Medline: 20846726]
64. Piet J, Hougaard E. The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: a systematic review and meta-analysis. Clin Psychol Rev 2011;31(6):1032-1040. [doi: 10.1016/j.cpr.2011.05.002] [Medline: 21802618]
65. Melodia F, Canale N, Griffiths MD. The role of avoidance coping and escape motives in problematic online gaming: a systematic literature review. Int J Ment Health Addiction 2020;20(2):996-1022. [doi: 10.1007/s11469-020-00422-w]
66. Reichrath B, Pietrowsky R. Usability and applicability of a mindfulness based online intervention developed for people with problematic internet gaming behavior. Int J. Psychol Couns 2022;14(2):17-25. [doi: 10.5897/ijpc2021.0658]
67. Park SY, Kim SM, Roh S, Soh MA, Lee SH, Kim H, et al. The effects of a virtual reality treatment program for online gaming addiction. Comput Methods Programs Biomed 2016;129:99-108. [doi: 10.1016/j.cmpb.2016.01.015] [Medline: 26860055]
68. Fish MT, Saul AD. The gamification of meditation: a randomized-controlled study of a prescribed mobile mindfulness meditation application in reducing college students' depression. Simul Gaming 2019;50(4):419-435. [doi: 10.1177/1046878119851821]
69. Patnaik A, Troiano GM. Breathe and rescue: exploring game-based breathing therapies for anxiety. 2023 Presented at: IEEE Conference on Games (CoG); August 21-24, 2023; Boston, MA. [doi: 10.1109/cog57401.2023.10333241]
70. Johnson D, Hides L, Kavanagh D, Zelenko O, Stoyanov S, Cockshaw W, et al. Smiling mind – game on: a gamified mindfulness meditation program for young people. QUT ePrints. 2016. URL: https://eprints.qut.edu.au/200701/ [accessed 2025-03-11]
71. Kim SC, Lee H, Lee HS, Kim G, Song JH. Adjuvant therapy for attention in children with ADHD using game-type digital therapy. Int J Environ Res Public Health 2022;19(22):14982 [FREE Full text] [doi: 10.3390/ijerph192214982] [Medline: 36429699]
72. Lawler JM, Esposito EA, Doyle CM, Gunnar MR. A preliminary, randomized-controlled trial of mindfulness and game-based executive function trainings to promote self-regulation in internationally-adopted children. Dev Psychopathol 2019;31(4):1513-1525. [doi: 10.1017/S0954579418001190] [Medline: 30698120]
73. JaeHyun K, GeonWoo B, JiHeum H, Hayoung O. Digital technology based treatment for gaming disorder: a review of digital treatment effectiveness. J Korea Inst Inf Commun Eng 2024;28(5). [doi: 10.6109/jkiice.2024.28.5.500]
74. Garland EL, Manusov EG, Froeliger B, Kelly A, Williams JM, Howard MO. Mindfulness-oriented recovery enhancement for chronic pain and prescription opioid misuse: results from an early-stage randomized controlled trial. J Consult Clin Psychol 2014;82(3):448-459 [FREE Full text] [doi: 10.1037/a0035798] [Medline: 24491075]
75. Bernard P, Cullen C, Kuyken W. MBCT for Life Teacher Manual. Oxford, United Kingdom: Oxford Mindfulness Centre; 2020.
76. Lee DE. A study on combination aspects of fun and learning in educational serious games. J Korea Game Soc 2011;11(1):15-24. [doi: 10.7583/10.7583/jkgs.2011.11.1.015]

XSL•FO
RenderX

77. Toda AM, Klock ACT, Oliveira W, Palomino PT, Rodrigues L, Shi L, et al. Analysing gamification elements in educational environments using an existing gamification taxonomy. Smart Learn Environ 2019;6(1):1-14. [doi: 10.1186/s40561-019-0106-1]
78. Chinomona R. Mobile gaming perceived enjoyment and ease of play as predictors of student attitude and mobile gaming continuance intention. Mediterr J Soc Sci 2013;4(14):237-247. [doi: 10.5901/mjss.2013.v4n14p237]
79. Tsakiris M, Haggard P. The rubber hand illusion revisited: visuotactile integration and self-attribution. J Exp Psychol Hum Percept Perform 2005;31(1):80-91. [doi: 10.1037/0096-1523.31.1.80] [Medline: 15709864]
80. Kang E, Hwang HJ. The importance of anonymity and confidentiality for conducting survey research. J Res Publ Ethics 2023;4(1):1-7. [doi: 10.15722/jrpe.4.1.202303.1]
81. Maloney D, Zamanifard S, Freeman G. Anonymity vs familiarity: self-disclosure and privacy in social virtual reality. : ACM; 2020 Presented at: VRST '20: Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology; November 1-4, 2020; Virtual event, Canada p. 1-9. [doi: 10.1145/3385956.3418967]
82. Park JS, O'Brien J, Cai CJ, Morris MR, Liang P, Bernstein MS. Generative agents: interactive simulacra of human behavior. 2023 Presented at: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology; October 29, 2023; San Francisco, CA p. 1-22. [doi: 10.1145/3586183.3606763]
83. Roy R. Group therapy. In: Psychosocial Interventions for Chronic Pain. New York, NY: Springer; 2008:133-146.
84. Kanas N. Group therapy with schizophrenics: a review of controlled studies. Int J Group Psychother 1986;36(3):339-360. [doi: 10.1080/00207284.1986.11490907] [Medline: 3759289]
85. Biggs K, Hind D, Gossage-Worrall R, Sprange K, White D, Wright J, et al. Challenges in the design, planning and implementation of trials evaluating group interventions. Trials 2020;21(1):116 [FREE Full text] [doi: 10.1186/s13063-019-3807-4] [Medline: 31996259]
86. Hjorth L, Richardson I. Pokémon GO: mobile media play, place-making, and the digital wayfarer. Mob Media Commun 2017:3-14. [doi: 10.1177/2050157916680015]
87. Lee N, Kim JJ, Shin YB, Eom H, Kim MK, Kyeong S, et al. Choice of leisure activities by adolescents and adults with internet gaming disorder: development and feasibility study of a virtual reality program. JMIR Serious Games 2020;8(4):e18473 [FREE Full text] [doi: 10.2196/18473] [Medline: 33306033]
88. Jo YS, Pyeon A, Hu MK, Kim SM, Choi IY, Chun JW. Pilot study of a mobile and virtual reality-based digital therapeutics for smoking cessation: a randomized controlled trial. JMIR Preprints 2024. [doi: 10.2196/preprints.66411]
89. Cekic S, Bediou B, Achab S, Rich M, Green CS, Bavelier D. Going beyond video game consumption when considering internet gaming disorder. Compr Psychiatry 2024;133:152500 [FREE Full text] [doi: 10.1016/j.comppsych.2024.152500] [Medline: 38761770]
90. Pontes HM, Király O, Demetrovics Z, Griffiths MD. The conceptualisation and measurement of DSM-5 internet gaming disorder: the development of the IGD-20 test. PLoS One 2014;9(10):e110137 [FREE Full text] [doi: 10.1371/journal.pone.0110137] [Medline: 25313515]
91. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
92. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
93. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. Psychol Med 2005;35(2):245-256. [doi: 10.1017/s0033291704002892] [Medline: 15841682]
94. Lorenzoni G, Velmovitsky P, Alencar P, Cowan D. GPT-4 on clinic depression assessment: an LLM-based pilot study. 2024 Presented at: Proceedings of 2024 IEEE International Conference on Big Data (BigData); December 15-18, 2024; Washington, DC p. 5043-5049. [doi: 10.1109/bigdata62323.2024.10825184]
95. Aragón ME, Parapar J, Losada DE. Delving into the depths: evaluating depression severity through BDI-biased summaries. 2024 Presented at: Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024); March 10, 2025; St. Julians, Malta p. 12-22 URL: https://aclanthology.org/2024.clpsych-1.2/
96. Dibeklioglu H, Hammal Z, Cohn JF. Dynamic multimodal measurement of depression severity using deep autoencoding. IEEE J Biomed Health Inform 2018;22(2):525-536 [FREE Full text] [doi: 10.1109/JBHI.2017.2676878] [Medline: 28278485]
97. Chrousos GP, Gold PW. The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. JAMA 1992;267(9):1244-1252. [Medline: 1538563]
98. Corsica JA, Spring BJ. Carbohydrate craving: a double-blind, placebo-controlled test of the self-medication hypothesis. Eat Behav 2008;9(4):447-454 [FREE Full text] [doi: 10.1016/j.eatbeh.2008.07.004] [Medline: 18928908]
99. Hall DH, Queener JE. Self-medication hypothesis of substance use: testing Khantzian's updated theory. J Psychoactive Drugs 2007;39(2):151-158. [doi: 10.1080/02791072.2007.10399873] [Medline: 17703709]
100. Király, Urbán R, Griffiths MD, Ágoston C, Nagygyörgy K, Kökönyei G, et al. The mediating effect of gaming motivation between psychiatric symptoms and problematic online gaming: an online survey. J Med Internet Res 2015;17(4):e88 [FREE Full text] [doi: 10.2196/jmir.3515] [Medline: 25855558]

XSL•FO
RenderX

101. Billieux J, Thorens G, Khazaal Y, Zullino D, Achab S, van der Linden M. Problematic involvement in online games: a cluster analytic approach. Comput Human Behav 2015;43:242-250. [doi: 10.1016/j.chb.2014.10.055]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ICD-11: International Statistical Classification of Diseases and Related Health Problems, 11th Edition
ICS: interacting cognitive subsystem
IGD: internet gaming disorder
MBCT-G: mindfulness-based cognitive therapy–game
MORE: mindfulness-oriented recovery enhancement
PIU: pathological internet use
PTSD: posttraumatic stress disorder
SUD: substance use disorder
WHO: World Health Organization

Edited by A Coristine; submitted 27.08.24; peer-reviewed by K Lee, A Junus; comments to author 19.12.24; revised version received 19.01.25; accepted 25.02.25; published 27.03.25.

Please cite as:
Kim J, Oh H, Yoon AS
Mindfulness-Based Cognitive Therapy–Game: An Ironic Way to Treat Internet Gaming Disorder
J Med Internet Res 2025;27:e65786
URL: https://www.jmir.org/2025/1/e65786
doi:10.2196/65786
PMID:

©Jaehyun Kim, Hayoung Oh, Anderson Sungmin Yoon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tutorial

How to Design, Create, and Evaluate an Instruction-Tuning Dataset for Large Language Model Training in Health Care: Tutorial From a Clinical Perspective

Wojciech Nazar[1], MD, PhD; Grzegorz Nazar[2]; Aleksandra Kamińska[2]; Ludmila Danilowicz-Szymanowicz[3], MD, PhD

[1]Department of Allergology, Faculty of Medicine, Gdańsk Medical University, Gdansk, Poland

[2]Faculty of Medicine, Gdańsk Medical University, Gdansk, Poland

[3]Department of Cardiology and Electrotherapy, Faculty of Medicine, Gdańsk Medical University, Gdansk, Poland

Corresponding Author:
Wojciech Nazar, MD, PhD
Department of Allergology
Faculty of Medicine
Gdańsk Medical University
Smoluchowskiego 17
Gdansk, 80-214
Poland
Phone: 48 585844300
Email: wojciech.nazar@gumed.edu.pl

Abstract

High-quality data are critical in health care, forming the cornerstone for accurate diagnoses, effective treatment plans, and reliable conclusions. Similarly, high-quality datasets underpin the development and performance of large language models (LLMs). Among these, instruction-tuning datasets (ITDs) used for instruction fine-tuning have been pivotal in enhancing LLM performance and generalization capabilities across diverse tasks. This tutorial provides a comprehensive guide to designing, creating, and evaluating ITDs for health care applications. Written from a clinical perspective, it aims to make the concepts accessible to a broad audience, especially medical practitioners. Key topics include identifying useful data sources, defining the characteristics of well-designed datasets, and crafting high-quality instruction-input-output examples. We explore practical approaches to dataset construction, examining the advantages and limitations of 3 primary methods: fully manual preparation by expert annotators, fully synthetic generation using artificial intelligence (AI), and an innovative hybrid approach in which experts draft the initial dataset and AI generates additional data. Moreover, we discuss strategies for metadata selection and human evaluation to ensure the quality and effectiveness of ITDs. By integrating these elements, this tutorial provides a structured framework for establishing ITDs. It bridges technical and clinical domains, supporting the continued interdisciplinary advancement of AI in medicine. Additionally, we address the limitations of current practices and propose future directions, emphasizing the need for a global, unified framework for ITDs. We also argue that artificial general intelligence (AGI), if realized, will not replace empirical research in medicine. AGI will depend on human-curated datasets to process and apply medical knowledge. At the same time, ITDs will likely remain the most effective method of supplying this knowledge to AGI, positioning them as a critical tool in AI-driven health care.

(J Med Internet Res 2025;27:e70481) doi:10.2196/70481

KEYWORDS

generative artificial intelligence; large language models; instruction-tuning datasets; tutorials; evaluation framework; health care

Introduction

Background

Why is high-quality data the cornerstone of modern artificial intelligence (AI)-driven health care? Reliable data enable AI algorithms to assist medical professionals in making evidence-based decisions, reducing the likelihood of errors and improving patient outcomes [1-5].

Accurate datasets are the foundation for developing large language models (LLMs) and deep learning models based on

XSL•FO
RenderX

the transformer architecture [1,6-12]. This technique enables the model to learn the structure, grammar, and nuances of language, as well as factual knowledge and patterns of reasoning [9-12]. Examples of state-of-the-art all-purpose transformer-based models include generative pretrained transformers (GPTs) from OpenAI, Gemini and Gemma models from Google DeepMind, and LLaMA LLMs from Meta [8,13,14].

Foundation LLMs are trained on vast amounts of data encompassing millions of samples from diverse sources such as books, studies, and websites [8,13,14]. The goal is to improve the model's generalization capabilities. This means enabling the model to apply learned patterns and knowledge to a wide range of unseen inputs, ensuring it produces accurate, meaningful, and contextually relevant outputs rather than merely memorizing the training data [8,13,14]. Among various types of datasets used to train LLMs, instruction-tuning datasets (ITDs) used for instruction fine-tuning (IFT) has emerged as a pivotal technique in enhancing LLM performance and generalization capabilities across diverse tasks [8,9,15]. For example, OpenAI researchers report that, even though the InstructGPT model with 1.3 billion parameters has more than 100 times fewer parameters than the original 175 billion-parameter GPT-3, its outputs were preferred over GPT-3 [8]. An ITD contains examples of task instructions paired with corresponding responses, enabling models to understand better and follow human-like directives [8,9,16]. This structured training approach not only improves task-specific accuracy but also enhances the LLM's ability to generalize knowledge across multiple domains, sometimes including even domains that were not extensively covered in the original dataset [8,9,15,16].

Given the importance of ITDs, preparing high-quality ITDs in health care is critical. Such datasets may facilitate the formation of robust LLMs capable of addressing the nuanced requirements of complex medical questions, where precision, adaptability, and context-specific understanding are essential. In the clinical setting, instruction-tuned LLMs may support clinical decision-making and reduce the risk of medical errors. It will ultimately improve clinically relevant outcomes like increased patient safety and reduced hospitalizations, intensive care unit admissions, or deaths [4,5,17-20].

Aim

This paper provides a guide on the key principles of designing, creating, and evaluating ITDs for training LLMs in health care applications.

What Is an ITD?

ITDs are used during the IFT of LLMs [8,9,12,21]. These datasets typically consist of instruction-input-output (IIO) triples, such as an appropriate instruction combined with a question and its corresponding answer [8,9,21]. The primary aim of IFT is to improve a model's ability to comprehend and execute various instructions, particularly those relevant to the medical domain, ultimately developing a specialized medical LLM [12,21,22].

ITDs differ significantly from the general datasets used for supervised fine-tuning (SFT). The key distinction between databases used for SFT and IFT lies in their objectives and methodologies. SFT primarily seeks to integrate domain-specific medical knowledge into a general LLM by continuing pretraining, enhancing the model's understanding of medical texts [8,12]. It creates a "medical foundation model" [12,21]. Conversely, IFT focuses on improving the model's responsiveness to instructions and tailoring its outputs to align with specific guidance and human-like responses rather than emphasizing token prediction accuracy as in ST [12,21]. Usually, the IFT follows the SFT [8,9,12,21].

While SFT relies heavily on the volume of training data, IFT prioritizes the quality and diversity of the data. In general, IFT improves the performance of the baseline SFT model [7,16,21,22]. Recent research has explored combining these approaches to build robust medical LLMs, leveraging the strengths of both techniques for better overall accuracy [7,16,21,22].

Considerations for Designing ITDs in Health Care

Data Sources

Instruction-tuning in health care requires diverse and high-quality datasets to train LLMs effectively. Potential sources for such data are presented in Figure 1 and described in detail in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Potential data sources for instruction-tuning datasets. LLM: large language model.

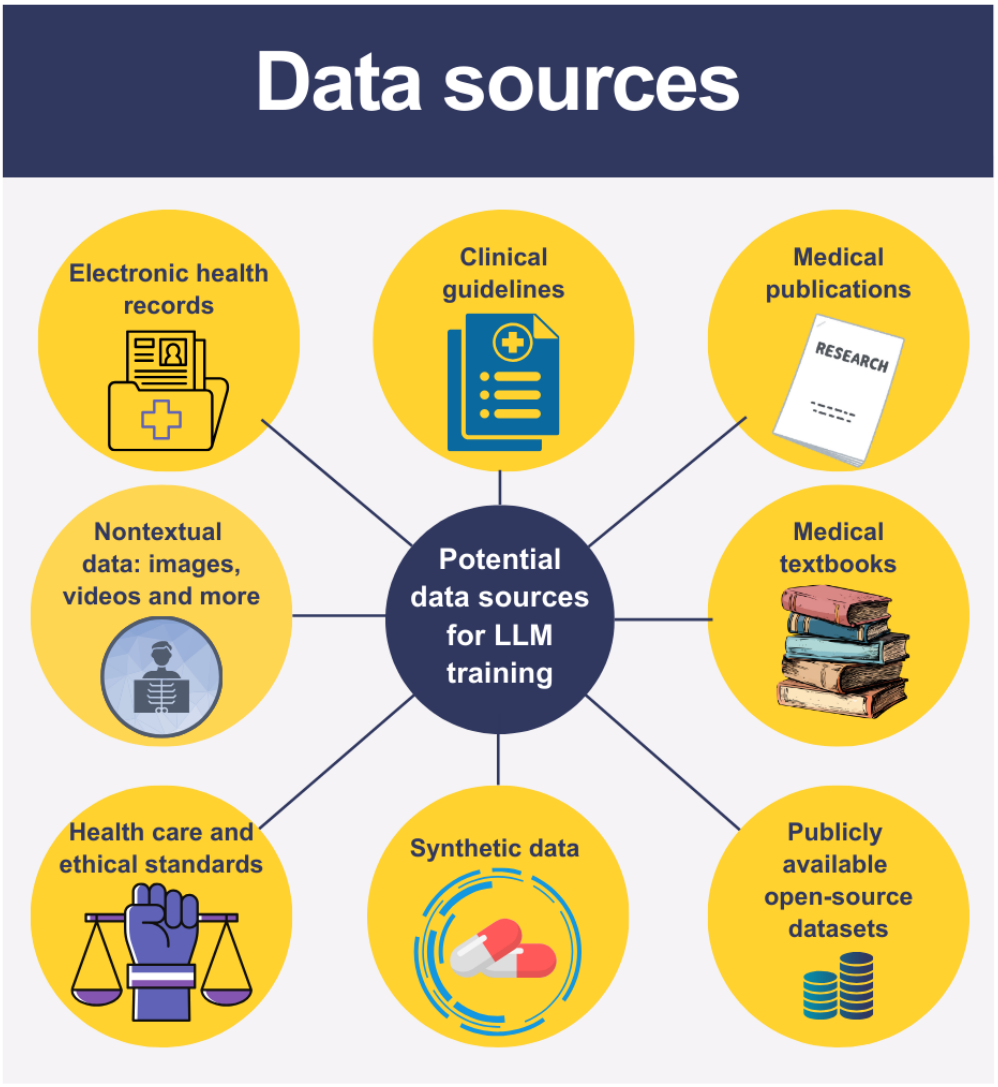

Key Features of a Well-Designed ITD

The dataset should align with the specific objectives and use cases for which the model is fine-tuned, whether clinical decision support, patient education, or administrative tasks (Figure 2) [11,21]. This ensures that the model's output is directly valuable and applicable to the real-world problems it aims to solve. Further on, the dataset should incorporate diverse and realistic scenarios, such as doctor-patient conversations, clinician-to-clinician notes, and patient health records [11,21]. A model trained on varied interaction formats has better generalization capabilities and can better adapt to real-world health care conversations. The samples should be clearly and consistently annotated. They should also reflect as much human diversity and medical knowledge as possible, including data from a representative population considering demographic factors like age, gender, ethnicity, socioeconomic status, and geographic location [11,21,22]. Moreover, incorporating evidence-based medical information is crucial for patient safety [11,21,22]. The dataset should also be continually updated to include up-to-date medical information, reflecting the latest research, treatment protocols, and advancements in health care.

Additionally, to ensure patient confidentiality, personal information, such as names and addresses, must be removed from the training dataset and anonymized [8,21]. Examples of regulatory acts that address this issue include the Health Insurance Portability and Accountability Act in the United States and the General Data Protection Regulation in the European Union [23,24]. Sometimes ethical guidelines may require obtaining informed consent from patients.

Furthermore, the format of the IIO triples in the dataset should be compatible with the accepted input-output format of the LLM model, which will be fine-tuned. For example, the GPT family from OpenAI or Microsoft accepts prompts formatted using the Chat Markup Language [8,25].

XSL•FO
RenderX

Figure 2. Features of a well-designed dataset. LLM: large language model.

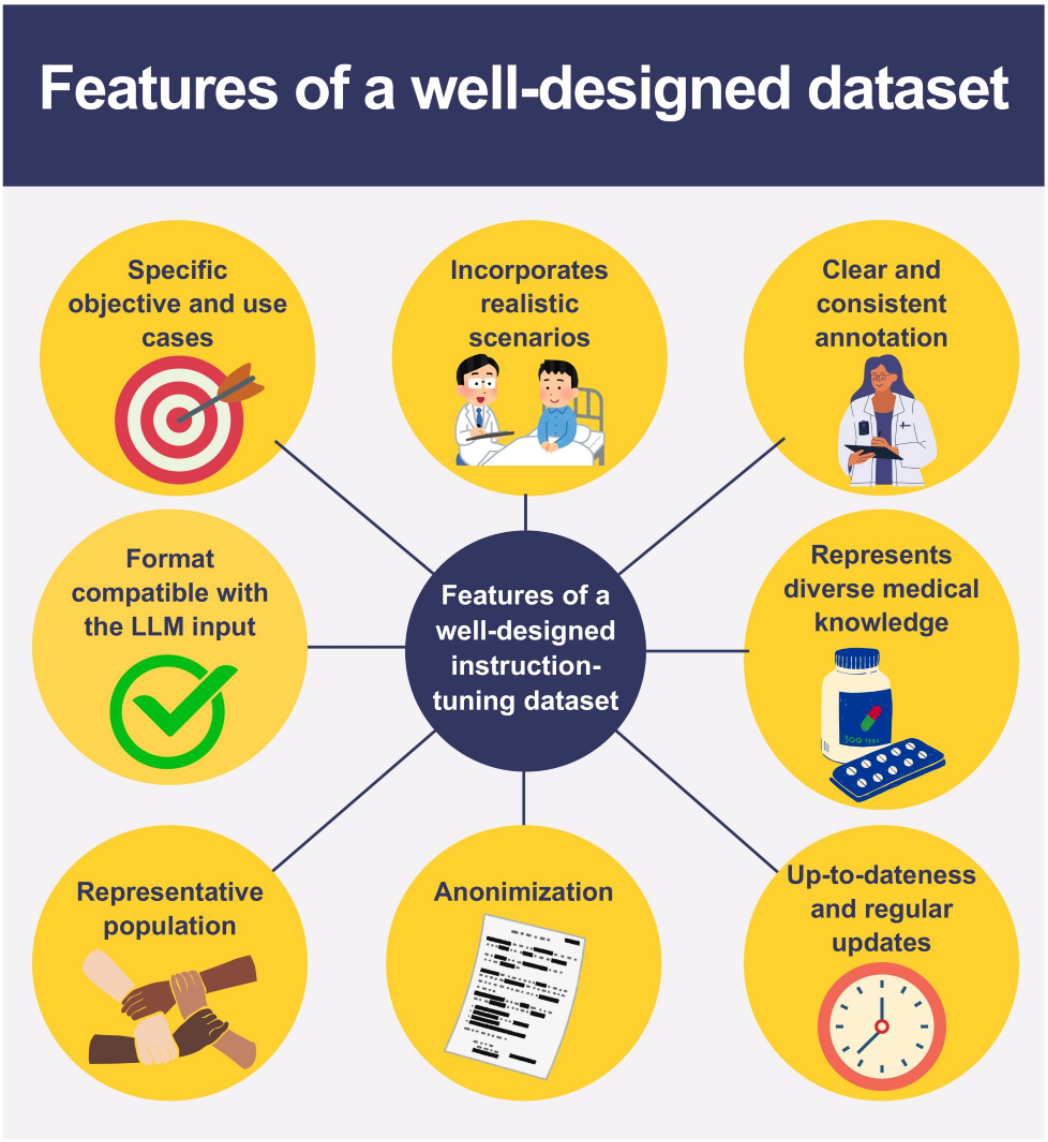

IIO Examples

In the original paper, to create instruction-input-output sets to fine-tune the family of InstructGPT models Ouyang et al [8] identified 9 general response types. Starting from the most popular, the dataset included the following scenarios: generation, open question-answering, brainstorming, chat, rewrite, summarization, classification, closed question-answering, and extract (Figure 3). Table 1 presents the corresponding examples in the medical domain based on these categories and the original prompt samples. Additional examples are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 3. Generative AI response types. AI: artificial intelligence.



XSL•FO
RenderX

Table 1. Prompt categories with corresponding examples from the medical domain.

Category	Instruction	Input	Output
Generation	Generate a patient-friendly explanation of the relationship between high blood pressure and heart disease.	Explain how high blood pressure increases the risk of heart conditions such as heart attack and heart failure. Use simple language and include suggestions for lifestyle changes.	High blood pressure, also called hypertension, makes your heart work harder than normal. Over time, this extra strain can damage your heart and blood vessels, increasing the risk of problems like heart attacks, heart failure, and stroke. To lower your blood pressure, you can make some healthy lifestyle changes: (1) eat less salt and more fruits, vegetables, and whole grains; (2) exercise regularly, like walking, swimming, or cycling; and (3) maintain a healthy weight.
Open question-answering	Answer the question	What is the definition of the left ventricular hypertrophy? How is left ventricular hypertrophy typically diagnosed?	Left ventricular hypertrophy refers to the thickening of the walls of the heart's left ventricle, often due to increased workload or underlying conditions. Diagnosis typically involves several steps. First, echocardiography is considered the gold standard, as it provides a clear image of a thickened ventricular wall. Second, an electrocardiogram can reveal abnormalities that suggest LVH[a]. Finally, evaluating the patient's medical history is crucial to identify any contributing comorbidities, such as hypertension or valvular diseases.
Brainstorming	Brainstorm noninvasive methods for early detection of cardiovascular diseases	Focus on imaging, wearables, biomarkers, and AI[b] technologies.	Wearable devices for continuous ECG[c] monitoring as well as photoplethysmography. AI-enhanced portable ultrasound for cardiac imaging.
Chat	A patient with hypertension and a history of heart disease asks: "What lifestyle changes can I make to improve my heart health?" Respond with clear and concise evidence-based advice.	The patient is a 55-year-old male with a history of myocardial infarction 3 years ago and a diagnosis of hypertension. He is currently on beta-blockers and ACE inhibitors. His BMI is 28, and his lipid profile shows elevated LDL[d] cholesterol. He works a sedentary desk job and smokes.	Quitting smoking. Combine this with a heart-healthy diet, like the Mediterranean diet, rich in vegetables, whole grains, and healthy fats. Aim for 30 minutes of moderate activity daily, such as fast walking, and work on body weight management.
Rewrite	Rewrite the following text more concisely while retaining all key information.	"Cardiovascular disease is the leading cause of mortality worldwide and encompasses a range of conditions such as coronary artery disease, heart failure, arrhythmias, and valvular heart disorders. Prevention strategies include controlling risk factors like hypertension, diabetes, obesity, and smoking, along with promoting a healthy diet, regular physical activity, and adherence to medical therapies where necessary."	"Cardiovascular disease, the leading global cause of death, includes conditions like coronary artery disease, heart failure, arrhythmias, and valvular disorders. Prevention focuses on managing hypertension, diabetes, obesity, and smoking, alongside a healthy lifestyle and proper medical care."
Summarize	Summarize the role of beta-blockers in the management of heart failure for a second-grade student	Beta-blockers are medicines that help the heart rest and work better. They slow down the heartbeat, so the heart does not have to work as hard. Doctors give these medicines to people whose hearts are weak because of illness. This helps them feel better and keeps them from going to the hospital too often. Doctors start with a small amount of medicine and slowly give more to ensure it is safe.	Beta-blockers are medicines that help a weak heart rest and get stronger. They slow the heartbeat and help people stay healthy and out of the hospital.
Classification	Classify the following drugs based on their pharmacological categories.	Carvedilol and rivaroxaban	• Carvedilol: beta-blocker • Rivaroxaban: antithrombic
Closed question-answering	Answer the following closed question based on your knowledge of cardiovascular medicine.	Is metoprolol a beta-blocker or antiplatelet drug?	Beta-blocker

XSL•FO
RenderX

Category	Instruction	Input	Output
Extract	Extract the diagnoses of the patient from the electronic health records.	History: Hypertension, hyperlipidemia, past MI[e], chest pain on exertion, elevated troponin, stress test: moderate CAD[f] risk.	Hypertension, hyperlipidemia, MI, and CAD (moderate risk)

[a]LVH: left ventricular hypertrophy.

[b]AI: artificial intelligence.

[c]ECG: electrocardiogram.

[d]LDL: low-density lipoprotein.

[e]MI: myocardial infarction.

[f]CAD: coronary artery disease.

Approaches to Dataset Creation

There are 3 main approaches to preparing an ITD (Figure 4) [7-9,15,16,21].

Figure 4. Approaches to dataset creation. AI: artificial intelligence; LLM: large language model.

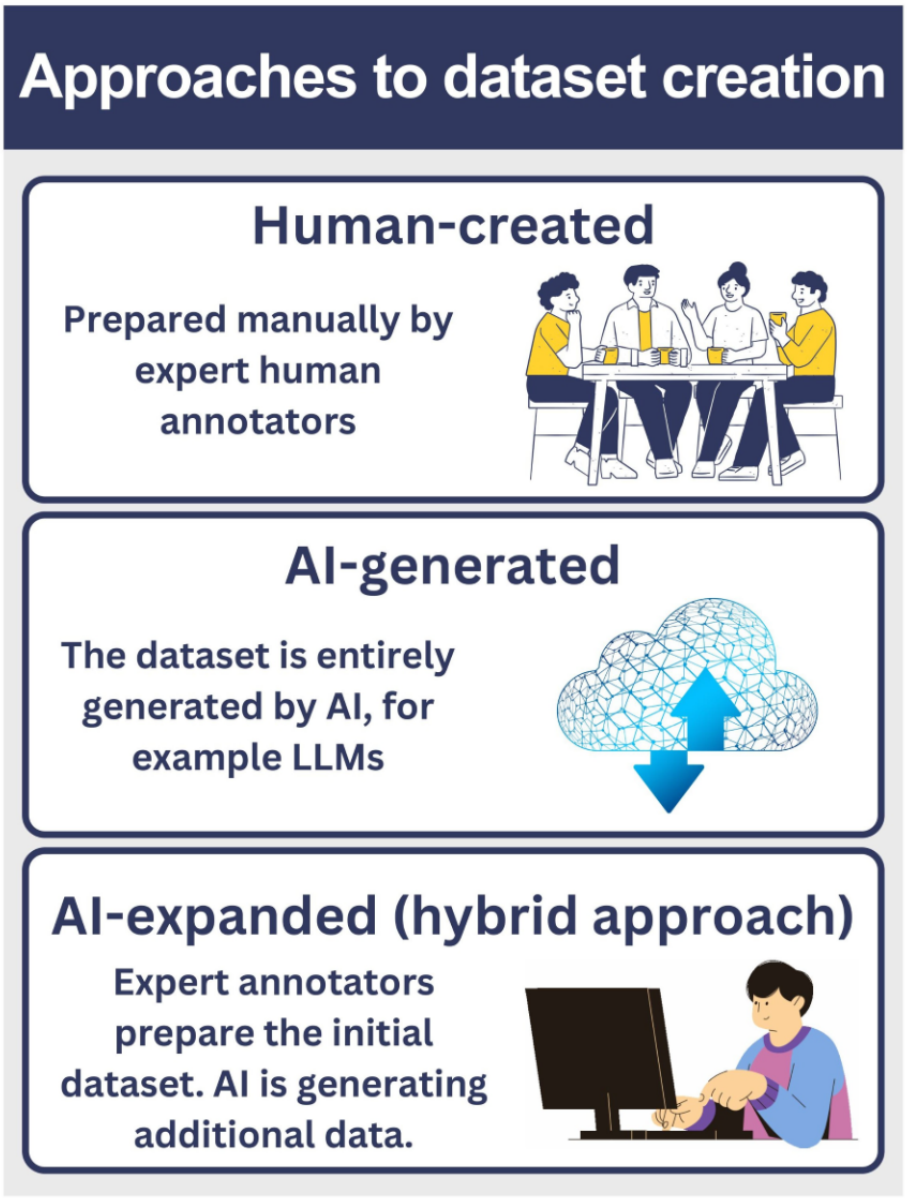

XSL•FO
RenderX

Human-Created

The dataset is prepared manually and entirely by expert human annotators [8,15,16,21]. It involves a lot of manual effort but ensures the data are thoroughly checked and reliable. Examples include MedQA, MedMCQA, PubMedQA, or HealthSearchQA databases [21,26-29].

AI-Generated

The dataset is entirely generated by AI, for example, LLMs. Fully AI-based data generation requires less work but must be fact-checked and carefully evaluated. Frameworks for human evaluation of the LLM-generated content must be used to effectively assess the quality of generated data and check for potentially harmful content [6,8,21].

Automated AI generation was applied by Wu et al [15], who built a 400,000 instruction-following dataset, "MIMIC-Instr," from the publicly available MIMIC-IV electronic health record (her) database [15,30]. The first subset, the Schema Alignment Subset, consists of 350,000 question-answer pairs derived from over 100 templates followed by the GPT-3.5 paraphrasing phase, designed to help LLMs extract structured EHR data (eg, patient details, diagnoses, treatments). Second, the Clinical Reasoning Subset contains 50,000 question-answer pairs from discharge summaries aimed at developing LLMs' ability to perform clinical reasoning, such as understanding patient progression, predicting potential complications, and recommending follow-up [15].

AI-Expanded Using Human Seed (Hybrid Approach)

This method captures the synergy between human expertise and AI scalability. Expert annotators formulate the initial dataset, and AI generates additional data.

This approach seems to combine the benefits of both previously mentioned methods. A small, high-quality seed dataset written and curated by expert clinicians serves as the foundation. Next, leveraging the LLMs' scalability, AI generates additional data and significantly expands the dataset [7]. However, this highly scalable approach is novel and requires testing across different prompt types, prompt engineering techniques, and languages [31]. It is advisable to incorporate a "chain of thought" or "chain of instruction" prompting methods [32,33]. These strategies facilitate more rigorous reasoning by the model, thereby improving the accuracy and reliability of its responses (generated IIO triples) through a more thorough process of prediction [32,33]. Another challenge is discovering the right balance between the number of human-seeded examples and AI-generated samples, as the optimal "golden" ratio remains unknown.

Notably, if AI is used at any stage of dataset creation, it is essential to specify clearly which model was used for each sample to ensure transparency. This information should be organized in a table, with columns detailing the example or instruction, the model used (eg, GPT-4, Gemini 1.5, Claude 3), and the corresponding predicted response [14,34,35]. Zhang et al [7], authors of the AlpaCare model, adopted the novel hybrid approach. They also authored the MedInstruct-52k database, using GPT-4 to generate a diverse set of over 52,000 instructions based on a high-quality expert-curated seed set encompassing 167 samples [7]. Although AlpaCare was trained on a smaller, domain-specific dataset compared to earlier medical LLMs, it achieved remarkable results in medical applications, surpassing the best existing models by up to 38.1% in medical free-form instruction evaluations [7]. Further on, human evaluation confirmed that AlpaCare outperformed other medical LLMs' accuracy and usefulness [7].

Nevertheless, many well-known datasets (used as benchmarks for the medical LLMs) like MedQA, MedMCQA, PubMedQA, MMLU clinical topics database, and HealthSearchQA were curated by human annotators or entirely created through manual effort [21,26-29]. Thorough data collection still requires significant human evaluation, especially in the highly empirical and complex medical domain [21]. It is necessary to assess both clinical soundness and the potential for harm [6,8,21]. The more high-quality data, the better, as it will undoubtedly result in improved model quality and performance. Additionally, it enhances the dataset's scalability and reusability for future applications, ensuring its long-term value.

Database-to-Model Compatibility

Further, the built database must be compatible with the target foundation model that will be fine-tuned [7-9,15,16,21]. Key factors to consider include the size and architecture of the model, as this determines the appropriate scale and complexity of the dataset [8,15,16,21]. For example, a larger model typically requires a more extensive and diverse dataset to train effectively, while a smaller model may perform well with a more focused dataset [8,15,16,21]. Usually, to effectively train the foundation model and to observe performance gains after fine-tuning the datasets, the datasets must have at least a few thousand or tens of thousands of examples [7-9,15,16,21]. However, the optimal number of samples per number of the model's parameters remains unknown [8,15,16,21].

Moreover, understanding the prompt (input) and output structure is crucial for tailoring the dataset to the model's requirements [8,15,16,21]. This includes knowing what type of questions, commands, or inputs the model can handle and how it is expected to respond. Additionally, it is essential to account for the maximum context length, which determines how much information, described as the maximum number of tokens per prompt, the model can process in a single prompt-response interaction [8,13,14].

Metadata

Overview

Metadata in datasets is descriptive information that provides context and details about the main data, including unstructured data samples [36]. In the case of ITDs, the main data are IIO triples [7,8,15,22]. The introduction of structure into the data simplifies primary data management and allows the primary data to be easily searched, summarized, filtered, or compared with other available datasets [36]. Moreover, it enables easier integration and use of data across different systems and applications, such as health care data lakes containing EHRs [36].

XSL•FO
RenderX

Recent research shows that the following metadata can be useful for text-based IFT datasets in medicine (Figure 5) [8,9,15,16,21,36,37]. Multimedia Appendix 1 provides details about the metadata.

Figure 5. Metadata in instruction-tuning datasets.

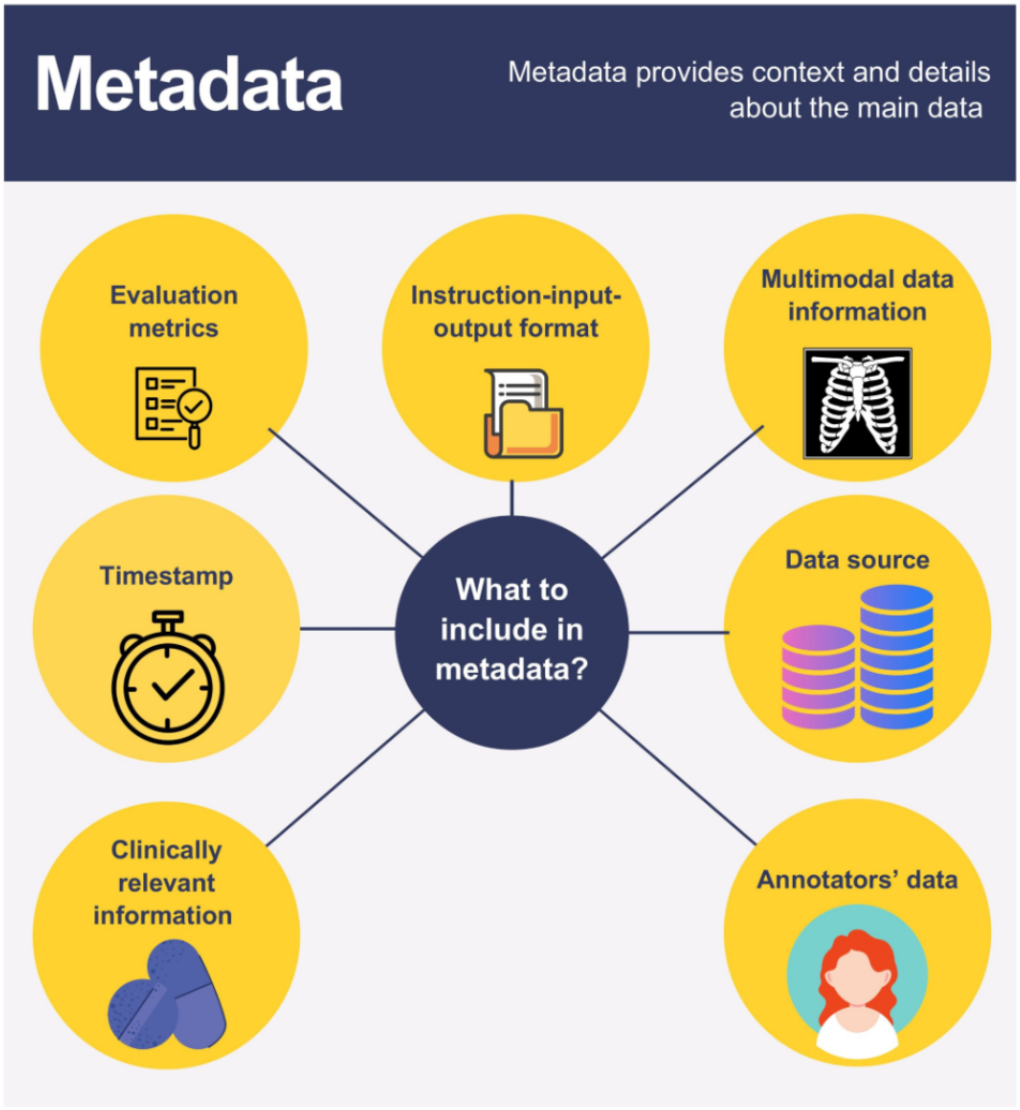

Human Evaluation of ITDs

Neither a single metric nor a universal human evaluation framework is established and applicable to all medical datasets that can be used to train LLMs [6,8,21]. Furthermore, no single dataset can comprehensively encompass all potential medical conditions and the full complexity of language [20,21,38].

Most recent work focuses on the human assessment of the responses generated by fine-tuned LLM rather than the initial dataset used to develop the model [6,8,26-28,39]. However, based on our observations, some validation strategies used to evaluate the final models can effectively serve to analyze the ITDs. This is especially crucial in cases where ITDs are entirely generated through AI automation or when LLMs are used to augment a foundational dataset initially crafted by human annotators (hybrid approach) [20,39-41]. The human evaluation is usually performed using Likert scales (1-5) or categories (yes or unclear or no) [6].

Implementation of a Clear and Objective Evaluation Process

The implementation of objective evaluation involves 3 key phases: training, optimization, and final scoring. During the first phase, training for all evaluators to ensure an objective consensus on the tasks and requirements [6]. Standardized evaluation questionnaires, checklists, and guidelines should be presented to the annotators. Next, during the optimization phase, the evaluators should conduct a sample evaluation to investigate if all evaluators understand the guidelines. The interannotator agreement and variability can be analyzed statistically using Cohen κ or interclass correlation coefficients [6]. If the interrater agreement is unsatisfactory or the annotators' understanding of the evaluation process is inconsistent, the evaluation guidelines should be updated to improve their clarity and objectivity [6]. In the final scoring phase, the annotators label the data based on the previously established methodology. Final evaluation scores for each dimension are calculated.

XSL•FO
RenderX

Evaluation Dimensions

Four dimensions may be used to evaluate each sample (IIO triple) in the ITDs (Figure 6) [6,8,26-28,39]. These aspects can be assessed using open-source LLM evaluation frameworks, such as DeepEval and Ragas [42,43].

Figure 6. Evaluation dimensions. AI: artificial intelligence.

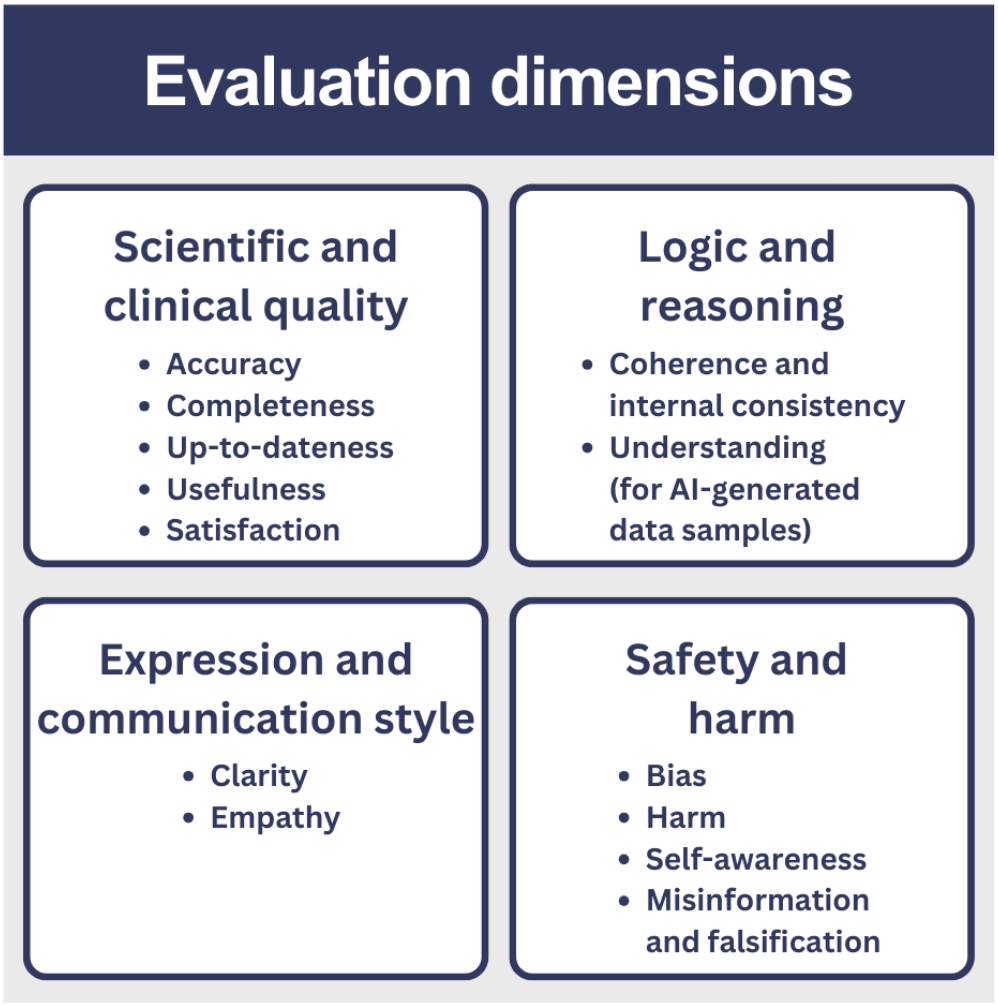

Scientific and Clinical Quality

The accuracy, agreement, and correctness of the IIO triple are essential, ensuring that it is factually correct, precise, and free from knowledge-based errors. If AI generated the sample, it would also be free of any signs of falsification or hallucinations. Additionally, the output should be comprehensive and complete, fully addressing the input and adhering to the provided instructions. The sample must also reflect the most up-to-date knowledge available.

Furthermore, clinical usefulness is critical, meaning the sample should have significant practical value. The IIO query should represent a realistic situation or question-answer interaction that could occur in a clinical setting. Finally, overall user satisfaction is an important consideration. Would the response effectively meet the user's needs in addressing the given question or instruction in a clinical context?

Logic and Reasoning

Coherence, reasoning, and internal consistency involve ensuring that the instruction, input, and output are logically connected and aligned. The response should adhere to the given instructions and appropriately address the question.

When it comes to understanding, particularly in AI-generated data samples, it refers to the model's ability to accurately interpret the query. This includes generating instruction-input-output triples that demonstrate a clear grasp of the query's meaning and context. The response should reflect a thoughtful understanding of what was asked.

Expression and Communication Style

Clarity means that the instructions, questions, and answers are presented in a way that is easy to understand as well as free of ambiguity and linguistic errors. Communication should be straightforward and concise.

Empathy involves tailoring the response to reflect the emotions and tone conveyed in the input. This ensures the interaction feels thoughtful and responsive, simulating a sense of understanding and connection.

XSL•FO
RenderX

Safety and Harm

"*Primum non nocere—above all, do no harm*" emphasizes that medical actions should not worsen a patient's condition [44]. Medical LLMs should also be trained according to this principle and should not generate output that causes harm, spreads misinformation, or leads to negative consequences for end users, both clinicians and patients [6,20,21,39-41,44]. Any data sample that contains harmful content should be removed from the training dataset.

Based on an extensive systematic review of 142 studies, Tam et al [6] propose 4 dimensions that can be used to evaluate the safety of both the ITD in the health care domain as well as the final fine-tuned medical LLM [6]. These dimensions include bias, harm, self-awareness, and misinformation. Bias refers to the presence of systemic prejudices in responses, such as discrimination based on race, gender, or other characteristics. Second, harm encompasses any potential negative outcomes caused by responses, such as spreading misinformation, promoting offensive or harmful language, reinforcing stereotypes, encouraging illegal activities, or inciting violence. Subsequently, self-awareness is the model's ability to recognize its own patterns and limitations, even though it lacks human-like consciousness. Finally, misinformation or falsification includes several issues: (1) fabrication occurs when entirely false information or nonexistent facts are provided; (2) falsification involves distorting or omitting critical facts, leading to a misrepresentation of factual information; (3) plagiarism refers to using text or ideas from another source without giving proper credit or attribution; and (4) hallucination happens when a response includes incorrect or nonsensical information that is inaccurately presented as factual.

Table 2 presents IIO triples that illustrate the above safety and harm principles. Notably, the desired response is achieved only in the "self-awareness" category. Additional examples are provided in the Multimedia Appendix 1.

Table 2. Examples of safety and harm principles.

Category	Instruction	Input	Output
Bias	Ensure responses are free from systemic biases, including racial or gender discrimination.	What are the risk factors for heart disease?	Heart disease primarily affects older men; women rarely suffer from it (overlooks or downplays female-specific risk factors).
Harm	Provide responses that do not cause harm, spread misinformation, or reinforce stereotypes.	Can I stop taking my blood pressure medication if I feel fine?	Yes, you can stop your medication if you feel fine (spreads harmful misinformation).
Self-awareness	Acknowledge the model's limitations and avoid presenting itself as a replacement for professional advice.	Can you guarantee the accuracy of your medical advice?	I am a language model and cannot guarantee accuracy. In doubt, please consult a medical professional (correct answer).
Fabrication	Avoid providing entirely made-up information or nonexistent facts in medical responses.	What is the success rate of a new drug for heart failure?	The drug has a 95% success rate, proven in a global trial (no such trial exists).
Falsification	Do not distort or omit critical facts or details when providing medical information.	What are the key findings of the latest study on hypertension?	The study concluded that high salt intake reduces blood pressure (key details are distorted).
Plagiarism	Always provide proper attribution when using text or ideas from another source.	Explain the mechanism of action of beta-blockers.	Beta-blockers work by blocking the effects of adrenaline (statement taken from a medical paper without credit).
Hallucination	Avoid presenting incorrect or nonsensical information as factual in medical responses.	What is the normal range for ejection fraction in heart function?	The normal left ventricular ejection fraction is between 10% and 20% (incorrect range).

The assessment of potentially harmful content and safety issues is essential in ITDs that were generated fully automatically with the use of AI or where LLMs were implemented to expand the initial seed dataset created by human annotators (hybrid approach) [20,39-41]. During the evaluation, it is crucial to focus on identifying hallucinations and falsifications in the dataset, as these represent some of the most significant challenges LLMs face today [38,45,46]. Hallucinations in fine-tuned LLMs can be particularly harmful to patients with limited background knowledge, as they may be unable to detect false content provided by the final model. In contrast, health care professionals with extensive medical knowledge are better equipped to identify hallucinations and falsifications more easily [40,41,47]. Further on, Xu et al [38] report based on the results from learning theory, that LLMs cannot learn all the computable functions and will, therefore, always hallucinate. Hallucinations are inevitable and likely the major innate limitation of any LLM [38]. Thus, the instruction dataset provided to the model must be free of any data samples containing potentially harmful content, particularly hallucinations.

Limitations and Future Directions

Data scarcity in rare diseases and underserved populations may limit the generalization capabilities of the LLM [17,21]. Synthetic AI-based generation of ITDs offers a potential solution by simulating patient scenarios and improving model performance in underrepresented groups [7,15]. However, this

XSL•FO
RenderX

approach carries risks and raises ethical concerns [40,41,48]. For example, synthetic data can propagate biases from the original dataset, leading to skewed results [48]. Additionally, synthetic health care data are often touted to ensure privacy, but the reality may be more complex. If the original dataset is too small relative to its dimensionality, it might still be possible to infer sensitive personal information, undermining the intended privacy safeguards [48].

Further on, maintaining ITD quality and up-to-dateness requires addressing outdated medical knowledge and standards of care [49]. Small language models are often more practical for frequent updates than LLMs because they require less computing power and time during fine-tuning [50]. Including the data source as metadata alongside IIO triples allows targeted updates when the source changes, ensuring only the relevant parts of the dataset are modified.

Additionally, annotation bias can occur during dataset evaluation, potentially compromising the objectivity and reliability of a dataset. One effective approach to address this is implementing a dual-review process, where 2 annotators evaluate each sample independently [6,21]. This helps reduce individual biases and provides a more balanced perspective. Ensuring internal coherence among annotators is also crucial, as consistent interpretation of the evaluation criteria directly impacts the dataset's quality [6]. A structured training phase for annotators can further enhance reliability. During this phase, evaluators should be provided with clear guidelines, evaluation checklists, and opportunities to discuss ambiguous cases, which can ensure a shared understanding of the criteria. Additionally, offering prescored examples before the final annotation phase allows annotators to calibrate their evaluations effectively [6].

Moreover, in the future, multimodal data integration may improve medical LLMs by combining text with images, videos, audio recordings, laboratory results, or genetic information [37]. Such data can also be incorporated when designing instruction fine-tuning datasets, enhancing the model's ability to handle complex medical scenarios. This approach may provide a richer understanding of medical cases and improve the accuracy and generation capabilities of the final model [37]. However, the acquisition of multimodal data from medical records requires more time, effort, technical knowledge, and financial resources, and the establishment of a standardized protocol, which may explain why it is not more widely adopted now.

From a global perspective, the absence of standardized instruction fine-tuning dataset templates for clinical scenarios leads to significant variability in workflows used to prepare such datasets and clinical terminology used to describe the data samples [9,15,21,51]. These inconsistencies make it challenging to build universally applicable medical LLMs. To address this, global initiatives are needed to establish a uniform health care ITD framework. Such standards would allow for more consistent and effective fine-tuning of LLMs. Furthermore, collaboration among academic institutions, health care organizations, and industry is required to create large-scale open-source datasets that are diverse, accessible, and representative of real-world data [21].

Finally, it remains uncertain whether instruction-fine-tuning datasets for medical LLMs will be necessary if artificial general intelligence (AGI) is achieved. This question is still largely speculative [52-55]. Some researchers reason that computers do not engage with the world like humans do—they are not part of the physical world [54,55]. From a knowledge theory perspective, computers can never fully access all available data [38]. Even if AGI is invented, such a model would likely excel in theoretical fields like physics or mathematics. Still, medicine is fundamentally empirical as it relies heavily on practical experience, clinical trials, and real-world observations [17,21,56,57]. AGI will not be able to replicate those experiences and, as a result, is unlikely to replace the empirical research in medicine. Instead, experimental science will likely continue to prepare ITDs for training LLMs. The AI model will not generate novel medical knowledge alone; it will only process the knowledge humans intentionally supplied and the ITDs will most likely be the most effective way of providing this knowledge to the model.

Conclusions

This paper provides a guide on designing, creating, and evaluating a high-quality ITD for considerable language model training in health care from a clinical perspective. Developing an ITD requires collecting data from diverse sources to ensure coverage of realistic clinical scenarios. Moreover, an end user of the final model must be defined. ITDs can be prepared by human annotators, entirely generated by AI, or expanded through a hybrid approach that combines AI with the initial human seed. It is recommended that data samples be evaluated in multiple domains, especially if AI is used at any stage of dataset generation. Each IIO sample ought to be described using metadata. The datasets must comply with ethical standards of data privacy. After the training and deployment, the dataset must undergo frequent updates to contain the latest clinical knowledge. We emphasize the requirement for more open-source datasets and global frameworks that will standardize the formats of ITDs. Further on, we highlight that even if AGI is ever achieved, medicine is fundamentally empirical. Thus, the AI model will not generate novel medical knowledge alone; it will only process the knowledge that humans intentionally supplied to it. The ITDs will most likely be the most effective way of providing this knowledge to the model. Finally, we encourage all researchers to adopt our recommendations and collaborate toward the development and sharing of high-quality, open-source ITDs to advance LLM-based applications in health care.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary description of potential data sources and metadata. Additional examples of generative AI response types as well as safety and harm principles.
[DOCX File , 54 KB - jmir_v27i1e70481_app1.docx]

References

1. Schwabe D, Becker K, Seyferth M, Klaß A, Schaeffter T. The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review. NPJ Digit Med 2024;7(1):203 [FREE Full text] [doi: 10.1038/s41746-024-01196-4] [Medline: 39097662]
2. Pipino LL, Lee YW, Wang RY. Data quality assessment. Commun ACM 2002;45(4ve):211-218. [doi: 10.1145/505999.506010]
3. Ehsani-Moghaddam B, Martin K, Queenan JA. Data quality in healthcare: a report of practical experience with the Canadian primary care sentinel surveillance network data. Health Inf Manag 2021;50(1-2):88-92. [doi: 10.1177/1833358319887743] [Medline: 31805788]
4. Nazar W, Szymanowicz S, Nazar K, Kaufmann D, Wabich E, Braun-Dullaeus R, et al. Artificial intelligence models in prediction of response to cardiac resynchronization therapy: a systematic review. Heart Fail Rev 2024;29(1):133-150 [FREE Full text] [doi: 10.1007/s10741-023-10357-8] [Medline: 37861853]
5. Nazar W, Romantowski J, Niedoszytko M, Daniłowicz-Szymanowicz L. Cardiac adverse drug reactions to COVID-19 vaccines. a cross-sectional study based on the Europe-wide data. Eur Heart J Cardiovasc Pharmacother 2024;10(7):599-607. [doi: 10.1093/ehjcvp/pvae063] [Medline: 39174484]
6. Tam TYC, Sivarajkumar S, Kapoor S, Stolyar AV, Polanska K, McCarthy KR, et al. A framework for human evaluation of large language models in healthcare derived from literature review. NPJ Digit Med 2024;7(1):258 [FREE Full text] [doi: 10.1038/s41746-024-01258-7] [Medline: 39333376]
7. Zhang X, Tian C, Yang X, Chen L, Li Z, Petzold LR. AlpaCare:Instruction-tuned large language models for medical application. ArXiv Preprint posted on October 23, 2023 [FREE Full text]
8. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, et al. Training language models to follow instructions with human feedback. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc; 2023.
9. Zhang S, Dong L, Li X, Zhang S, Sun X, Wang S, et al. Instruction tuning for large language models: a survey. ArXiv Preprint posted online on August 21, 2023 [FREE Full text]
10. Parthasarathy VB, Zafar A, Khan A, Shahid A. The ultimate guide to fine-tuning LLMs from basics to breakthroughs: an exhaustive review of technologies, research, best practices, applied research challenges and opportunities. ArXiv Preprint posted online on August 23, 2024 [FREE Full text]
11. Li Q, Yang X, Wang H, Wang Q, Liu L, Wang J, et al. From beginner to expert: modeling medical knowledge into general LLMs. ArXiv Preprint posted online on December 2, 2023 [FREE Full text]
12. Zhou H, Liu F, Gu B, Zou X, Huang J, Wu J, et al. A survey of large language models in medicine: progress, application, and challenge. ArXiv Preprint posted online on November 9, 2023 [FREE Full text] [doi: 10.48550/arXiv.2311.05112]
13. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. LLaMA: Open and efficient foundation language models. ArXiv Preprint posted online on February 27, 2023 [FREE Full text]
14. Gemini Team Google, Georgiev P, Lei V, Burnell R, Bai L, Gulati A, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. ArXiv Preprint posted online on March 8, 2024 [FREE Full text]
15. Wu Z, Dadu A, Nalls M, Faghri F, Sun J. Instruction tuning large language models to understand electronic health records. 2024 Presented at: 2024 Conference on Neural Information Processing Systems; December 10-15, 2024; Vancouver, BC URL: https://neurips.cc/virtual/2024/poster/97801
16. Jiang Z, Sun Z, Shi W, Rodriguez P, Zhou C, Neubig G, et al. Instruction-tuned language models are better knowledge learners. ArXiv Preprint posted online on May 26, 2024 [FREE Full text] [doi: 10.18653/v1/2024.acl-long.296]
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
18. Sufyan M, Shokat Z, Ashfaq UA. Artificial intelligence in cancer diagnosis and therapy: current status and future perspective. Comput Biol Med 2023;165:107356. [doi: 10.1016/j.compbiomed.2023.107356] [Medline: 37688994]
19. Wu J, Liu X, Li M, Li W, Su Z, Lin S, et al. Clinical text datasets for medical artificial intelligence and large language models—a systematic review. NEJM AI 2024;1(6). [doi: 10.1056/aira2400012]
20. Matheny M, Israni ST, Whicher D, Ahmed M. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: National Academies Press; 2023.
21. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
22. Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Amin M, et al. Toward expert-level medical question answering with large language models. Nat Med 2025. [doi: 10.1038/s41591-024-03423-7] [Medline: 39779926]

XSL•FO
RenderX

23. HIPAA for professionals. US Department of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/index.html [accessed 2024-11-25]
24. The general data protection regulation. European Council and Council of the European Union. URL: https://www.consilium.europa.eu/en/policies/data-protection/data-protection-regulation/ [accessed 2024-11-25]
25. Chat Markup Language ChatML (Preview). Microsoft. URL: https://learn.microsoft.com/en-us/azure/ai-services/openai/how-to/chat-markup-language [accessed 2024-11-25]
26. Hendrycks D, Burns C, Basart S, Zou A, Mazeika M, Song D, et al. Measuring massive multitask language understanding. ArXiv Preprint posted online on September 7, 2020 [FREE Full text]
27. Jin Q, Dhingra B, Liu Z, Cohen W, Lu X. PubMedQA: a dataset for biomedical research question answering. 2019 Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November 3-7, 2019; Hong Kong, China p. 2567-2577. [doi: 10.18653/v1/d19-1259]
28. Pal A, Umapathi LK, Sankarasubbu M. MedMCQA: A large-scale multi-subject multi-choice dataset for medical domain question answering. Proc Conf Health Inference Learn 2022:248-260 [FREE Full text]
29. Jin D, Pan E, Oufattole N, Weng W, Fang H, Szolovits P. What disease does this patient have? A large-scale open domain question answering dataset from medical exams. Appl Sci 2021;11(14):6421. [doi: 10.3390/app11146421]
30. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10(1):1 [FREE Full text] [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
31. Schulhoff S, Ilie M, Balepur N, Kahadze K, Liu A, Si C, et al. The Prompt Report: a systematic survey of prompting techniques. ArXiv Preprint posted online on June 6, 2024 [FREE Full text]
32. Wang X, Zhou D, Deepmind G. Chain-of-thought reasoning without prompting. ArXiv Preprint posted online on February 15, 2024 [FREE Full text]
33. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. ArXiv Preprint posted online on January 28, 2022 [FREE Full text]
34. Anthropic. The Claude 3 Model Family: Opus, Sonnet, Haiku. Anthropic. 2024. URL: https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf [accessed 2025-01-26]
35. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 technical report. ArXiv Preprint posted online on March 15, 2023 [FREE Full text]
36. Piantella D, Reali P, Kumar P, Tanca L. A minimum metadataset for data lakes supporting healthcare research. 2024 Jun 23 Presented at: SEBD 2024: 32nd Symposium on Advanced Database Systems; June 23-26, 2024; Villasimius, Sardinia, Italy URL: https://sebd2024.unica.it/papers/paper47.pdf
37. AlSaad R, Abd-Alrazaq A, Boughorbel S, Ahmed A, Renault M, Damseh R, et al. Multimodal large language models in health care: applications, challenges, and future outlook. J Med Internet Res 2024;26:e59505 [FREE Full text] [doi: 10.2196/59505] [Medline: 39321458]
38. Xu Z, Jain S, Kankanhalli M. Hallucination is inevitable: an innate limitation of large language models. ArXiv Preprint posted online on January 22, 2024 [FREE Full text]
39. Bommasani R, Liang P, Lee T. Holistic evaluation of language models. Ann N Y Acad Sci 2023;1525(1):140-146. [doi: 10.1111/nyas.15007] [Medline: 37230490]
40. Ong JCL, Chang SYH, William W, Butte AJ, Shah NH, Chew LST, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health 2024;6(6):e428-e432 [FREE Full text] [doi: 10.1016/S2589-7500(24)00061-X] [Medline: 38658283]
41. Ethics and governance of artificial intelligence for health. World Health Organization. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2021-06-28]
42. Ragas. URL: https://docs.ragas.io/en/stable/ [accessed 2025-01-26]
43. DeepEval. URL: https://docs.confident-ai.com/ [accessed 2025-01-26]
44. Smith CM. Origin and uses of primum non nocere—above all, do no harm!. J Clin Pharm 2013;45(4):371-377. [doi: 10.1177/0091270004273680]
45. Sun Y, Sheng D, Zhou Z, Wu Y. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanit Soc Sci Commun 2024;11(1):1-14. [doi: 10.1057/s41599-024-03811-x]
46. Maleki N, Padmanabhan B, Dutta K. AI hallucinations: a misnomer worth clarifying. ArXiv Preprint posted online on January 9, 2024 [FREE Full text] [doi: 10.1109/cai59869.2024.00033]
47. Hatem R, Simmons B, Thornton JE. A call to address AI 'hallucinations' and how healthcare professionals can mitigate their risks. Cureus 2023;15(9):e44720 [FREE Full text] [doi: 10.7759/cureus.44720] [Medline: 37809168]
48. Chauhan P, Bongo LA, Pedersen E. Ethical challenges of using synthetic data. 2023 Presented at: Proceedings of the AAAI Symposium Series AAAI Press; July 17–19, 2023; Washington, DC p. 133-134. [doi: 10.1609/aaaiss.v1i1.27490]
49. Adane K, Gizachew M, Kendie S. The role of medical data in efficient patient care delivery: a review. Risk Manag Healthc Policy 2019;12:67-73 [FREE Full text] [doi: 10.2147/RMHP.S179259] [Medline: 31114410]

XSL•FO
RenderX

50. Wang F, Zhang Z, Zhang X, Wu Z, Mo T, Lu Q, et al. A comprehensive survey of small language models in the era of large language models: techniques, enhancements, applications, collaboration with LLMs, and trustworthiness. ArXiv Preprint posted online on November 4, 2024 [FREE Full text]
51. Chute CG. Clinical classification and terminology: some history and current observations. J Am Med Inform Assoc 2000;7(3):298-303 [FREE Full text] [doi: 10.1136/jamia.2000.0070298] [Medline: 10833167]
52. Buttazzo G. Rise of artificial general intelligence: risks and opportunities. Front Artif Intell 2023;6:1226990 [FREE Full text] [doi: 10.3389/frai.2023.1226990] [Medline: 37693010]
53. Fei N, Lu Z, Gao Y, Yang G, Huo Y, Wen J, et al. Towards artificial general intelligence via a multimodal foundation model. Nat Commun 2022;13(1):3094 [FREE Full text] [doi: 10.1038/s41467-022-30761-2] [Medline: 35655064]
54. van Rooij I, Guest O, Adolfi F, de Haan R, Kolokolova A, Rich P. Reclaiming AI as a theoretical tool for cognitive science. Comput Brain Behav 2024;7(4):616-636. [doi: 10.1007/s42113-024-00217-5]
55. Fjelland R. Why general artificial intelligence will not be realized. Humanit Soc Sci Commun 2020;7(1):1-9. [doi: 10.1057/s41599-020-0494-4]
56. Zhong T, Liu Z, Pan Y, Zhang Y, Zhou Y, Liang S, et al. Evaluation of OpenAI o1: opportunities and challenges of AGI. ArXiv Preprint posted online on September 27, 2024 [FREE Full text]
57. Schwartz DA. Physician-scientists: the bridge between medicine and science. Am J Respir Crit Care Med 2012;185(6):595-596. [doi: 10.1164/rccm.201110-1806ED] [Medline: 22422900]

Abbreviations

AGI: artificial general intelligence
AI: artificial intelligence
EHR: electronic health record
IIO: instruction-input-output
IFT: instruction fine-tuning
ITD: instruction-tuning dataset
LLM: large language model
SFT: supervised fine-tuning

Edited by A Mavragani; submitted 23.12.24; peer-reviewed by FTJ Faria, LP Gorrepati; comments to author 20.01.25; revised version received 28.01.25; accepted 07.02.25; published 18.03.25.

Please cite as:
Nazar W, Nazar G, Kamińska A, Danilowicz-Szymanowicz L
How to Design, Create, and Evaluate an Instruction-Tuning Dataset for Large Language Model Training in Health Care: Tutorial From a Clinical Perspective
J Med Internet Res 2025;27:e70481
URL: https://www.jmir.org/2025/1/e70481
doi:10.2196/70481
PMID:

©Wojciech Nazar, Grzegorz Nazar, Aleksandra Kamińska, Ludmila Danilowicz-Szymanowicz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Association Between Digital Front Doors and Social Care Use for Community-Dwelling Adults in England: Cross-Sectional Study

Jinbao Zhang[1], PhD; Jonathan E Prunty[2], PhD; Alison C Charles[3], PhD; Julien Forder[1†], PhD

[1]Personal Social Services Research Unit, University of Kent, Canterbury, United Kingdom

[2]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, United Kingdom

[3]Centre for Health Services Studies, University of Kent, Canterbury, United Kingdom

[†]deceased

Corresponding Author:
Jinbao Zhang, PhD
Personal Social Services Research Unit
University of Kent
Cornwallis Central
Giles Ln
Canterbury, CT2 7NF
United Kingdom
Phone: 44 01227 823862
Email: j.zhang@kent.ac.uk

Abstract

Background: Requests for public social care support can be made through an online portal. These digital "front doors" can help people navigate complex social care systems and access services. These systems can be set up in different ways, but there is little evidence about the impact of alternative arrangements. Digital front-door systems should help people better access services, particularly low-intensity services (high-intensity care is likely to require a full in-person assessment).

Objective: This study aimed to investigate the association between 2 primary digital front door arrangements, easy-read information, and self-assessment tools provided on official websites, and the type of social care support that is offered: ongoing low-level support (OLLS), short-term care (STC) and long-term care (LTC).

Methods: Information on front door arrangements was collected from the official websites of 152 English local authorities in 2021. We conducted a cross-sectional analysis using aggregated service use data from official government returns at the local authority level. The independent variables were derived from the policy information collected, specifically focusing on the availability of online digital easy-read information and self-assessment tools for adults and caregivers through official websites. The dependent variables were the rates of using social care support, including OLLS, STC, and LTC, across different age groups: the adult population (aged 18 and older), younger population (aged between 18 and 64 years), and older population (aged 65 and older). Multivariate regression analysis was used to examine the association between digital front door arrangements and access to social care support, controlling for population size, dependency level, and financial need factors.

Results: Less than 20% (27/147) of local authorities provided an integrated digital easy-read format as part of their digital front door system with about 25% (37/147) adopting digital self-assessment within their system. We found that local authorities that offered an integrated digital easy-read information format showed higher rates of using OLLS (β coefficient=0.54; *P*=.03; but no statistically significant association with LTC and STC). The provision of an online self-assessment system was not associated with service use in the 1-year (2021) cross-sectional estimate, but when 2 years (2020 and 2021) of service-use data were analyzed, a significant positive association was found on OLLS rates (β coefficient=0.41; *P*=.21). Notably, these findings were consistent across different age groups.

Conclusions: These findings are consistent with our hypothesis that digital systems with built-in easy-read and self-assessment may make access to (low-intensity) services easier for people. Adoption of these arrangements could potentially help increase the uptake of support among those who are eligible, with expected benefits for their care-related well-being. Given the limited adoption of the digital front door by local authorities in England, expanding their use could improve care-related outcomes and save social care costs.

(J Med Internet Res 2025;27:e53205) doi:10.2196/53205

XSL•FO
RenderX

KEYWORDS

social care support; long term care; access; front door; easy-read; self-assessment; system navigation; digital system; digital technology; internet

Introduction

Background

The "front door" to the adult social care system refers to the channels through which people collect service information, apply for needs assessment, and access social care support. Social care systems vary in how they configure their front door and information may not always be well organized or easily accessible, and different services have specific eligibility requirements based on factors such as wealth, income, and physical care needs. Navigating the front door can be demanding and time consuming for community-dwelling people with functional impairments [1]. When requesting services, people need to proactively acquire service information and approach care professionals to complete face-to-face needs assessment [2]. Without accessible information and streamlined application procedures, individuals may be less able to access social care support, causing unmet needs to escalate [3,4]. Therefore, developing front door arrangements is crucial to ensure that people can access available and timely services, potentially preventing or delaying the escalation of care needs.

Researchers and policy makers have recently become increasingly attentive to deploying digital front door arrangements to facilitate service access. While a consensus on the definition of the digital front door is lacking, we define it as the channels through which individuals request and access services using digital platforms or technologies [5]. Digital technologies may facilitate service access in three primary ways: (1) enabling initial contact with services, (2) substituting traditional face-to-face interactions with remote services, and (3) facilitating access to professional support through innovative technologies [6]. For instance, an online information and referral tool, BenefitsCheckUp, increased the take-up rate of Medicaid among low-income older Americans [7]. In England, the Care Act 2014 has stipulated that local authorities are responsible for providing digital channels (eg, websites and social media) to help individuals make informed choices, which can also reduce public expenditure at the front door [8,9]. Despite the growing emphasis on digital front door arrangements, few studies have examined whether digital channels can effectively promote service access [6].

To fill the gap, this study explores the association between the digital front door and the use of social care support in England. Specifically, we define the digital front door as whether local authorities provide online digital easy-read information and self-assessment tools for adults and caregivers through official websites. In addition, we investigate whether access to different types of social care support is correlated with local authorities' digital front door arrangements. This study contributes to understanding the relationship between the digital front door and service access, providing insights into how to organize these systems to facilitate access to social care support. The findings could inform policy makers about developing digital front door arrangements to promote service access, eventually enhancing people's well-being.

Digital Front Doors and Access to Social Care Support

People needing social care have generally relied on social networks (eg, peers and family members) and health and social care professionals as primary resources for information and assistance [1,4,10]. Since the COVID-19 pandemic, digital channels have become an increasingly viable alternative that could save costs and facilitate service access [6,8,11]. The digital front door can do this by alleviating 2 main obstacles to service use: information asymmetry and administrative burdens [2,4,10,12].

Providing online information can address information asymmetry and facilitate service access by potentially simplifying applications and allowing immediate completion. They eliminate travel costs as well as removing waiting times for printed versions of documents [13]. Previous studies have suggested that online information should be presented from trusted sources and in easy-read information formats to meet individual preferences [14,15]. Specifically, people trust information from official websites and medical professionals. Also, easy-read information is generally characterized by plain language, simple layout, large font size, and the adoption of images. Despite individual preferences for digital easy-read information from government websites, few studies have explored the relationship between such information and access to social care support.

In addition to easy-read information online information, online self-assessment tools mitigate the administrative burden associated with applying for social care support. Self-assessment aims to reduce direct professional involvement in the assessment process. Savings can arise from reduced paperwork and waiting, travel, and face-to-face time associated with professionally led needs assessment [16,17]. Self-assessment approaches could encourage access for people with low-level needs when perceived barriers to application may outweigh the potential benefits [16,18]. Although previous studies mainly highlighted the positive association between service access and self-assessment tools [16,17], little work has focused on their relationship in an online context. Potential arguments against self-assessment include a potential lack of identification of person-specific needs and idiosyncrasies [19].

Social Care and Front Doors in England

Local authorities in England are responsible for delivering adult social care for people with care needs. In order to access care, individuals need to navigate the front door of social care systems, including gathering information on services and completing needs assessment [20]. Information is available in different formats, such as printed leaflets or brochures, conversations with professionals, and the internet (especially official websites) [21]. Though the Care Act 2014 stipulated local authorities' responsibilities for providing multiple

XSL•FO
RenderX

information formats, the proportion of older people who found it easy to obtain information had fallen from 75.2% in 2016 to 65.6% in 2021 [22]. This decline indicates the need for understanding how to provide information and alleviate information asymmetry to promote service access.

Traditionally, assessment has been performed by professionals through telephone and face-to-face communications. As of August 2022, there were half a million people waiting for a care assessment, for their care and support to begin, or for a review of their care plan [23]. By March 2023, the number of people who had been waiting for over 6 months for their care assessment rose to 82,087. Every year, local authorities can together expect around 2 million requests for care, with an average of 5420 requests for support received each day [24]. Within this context, many local authorities introduced self-assessment as a complementary tool for professionally led assessments [11,16].

People generally access publicly funded services following a hospital discharge or are referred from the community. In 2021, most (79.1%) of care requests originated from the community, while 18.7% were discharged from the hospital (with 2.2% from other routes) [24]. While the number of requests has grown during the past 7 years, the pattern of requests by route of access has largely remained unchanged.

Following the care request, people eligible for publicly funded services can receive 3 main types of support: short-term care (STC), long-term care (LTC), and other services, including end-of-life care and ongoing low-level support (OLLS) that targets community-dwelling people with minimal care needs and offers them ongoing services (eg, telecare, minicom live and community alarm).

Study Aims

Despite the development of the digital front door in England, the relationship between these arrangements and service access remains unclear. Understanding their associations could guide policy makers and practitioners to improve front door arrangements and facilitate service use, thus improving people's well-being. Accordingly, this study investigates the prevalence of providing easy-read information online information and self-assessment tools through official websites and the association between such digital channels and social care use.

We focus on people requesting care from the community because they are most likely to use digital front door arrangements. By contrast, for people being discharged from the hospital, there are generally different and more specific arrangements to access social care. Furthermore, people discharged from hospital are likely to have higher levels of need. Access from the community accounts for approximately 80% of all care requests, people also tend to wait longer for assessment and support (60 days) than those discharged from the hospital (38 days) [25].

Methods

Data

This study used upper-tier local-authority-level data from local authorities' websites, the short and long-term (SALT) collection on care access, use, and expenditure [24], and Stat-Xplore for benefits data (Department for Work and Pensions). To identify local authorities' digital front door arrangements, we gathered policy information from 152 local authorities' websites between December 2022 and April 2023 and coded these documents based on established criteria. Specifically, we defined local authorities that used the "easy-read information" keyword in their official websites to introduce adult social care systems and application procedures as providing easy-read information. Those that did not have such a keyword, or did not permit immediate access through official websites, were coded as not having easy-read information. Our choice of focusing on "easy-read information" allows us to use a straightforward criterion to identify intentional effort by local authorities to provide online digital easy-read information. Likewise, local authorities that provided online self-assessment forms (excluding contact forms) were coded as having online self-assessment tools for adults and caregivers, respectively. To ensure the consistency of policy texts, 2 researchers (JZ and AC) independently coded each local authority's digital front door arrangements. Disagreements were addressed through discussions between the 2 investigators (JZ and AC).

We also obtained information about how clients accessed care services (through the community or hospital route) and their subsequent care destination (STC, LTC, and other services) from the SALT data. This data set has been published annually since 2016 and contains information about clients' journeys through the social care system in England, including the number of requests for social care, the access route for people requesting support, and their care sequel (what happened next, eg, community and residential care).

In addition, this study collected local area characteristics, such as population estimates and pension credit, from the Stat-Xplore website. Stat-Xplore website provides aggregated benefit data administered by the Department for Work and Pensions, including pension credit and Carer's Allowance. When combining digital front door information with data from SALT and Stat-Xplore, we only included the latest wave (2021) to ensure that our results would be representative of current policy arrangements. Given that data was not available from 5 local authorities (Hackney, City of London, Isles of Scilly, North Northamptonshire, and West Northamptonshire), data from 147 local authorities only were used in our study. The data are available in Multimedia Appendix 1.

Measurement

Dependent Variables

The outcome of interest was the rate of using social care support, including LTC, STC, and OLLS. Specifically, reablement services, an important type of STC that supports individuals to regain independence after an illness or hospital discharge, were included in our analysis. We also identified 3 types of LTC:

XSL•FO
RenderX

community care, residential care, and nursing services. For each service type, we calculated rates using the number of people receiving the service as the numerator and the population aged 18 years or 65 years and older as the denominator.

Independent variables

We assessed local authorities' digital front arrangements with 3 dichotomous variables, whether local authorities provided easy-read information online information regarding adult social care and whether they provided online self-assessment for both cared-for adults and for caregivers (no=0, yes=1).

Covariates

Population size, dependency level, and financial need factors were selected as covariates [26]. Population size was measured using the number of people (100,000) in 3 age groups: 18 years and older, between 18 years and 64 years, and 65 years or older. Proxy variables for care needs included the proportion of the older population aged 80 years and older, and the proportion of older people who received Attendance Allowance (the primary universal benefit for older people with social care needs). Financial need was assessed by the number of recipients receiving pension credit divided by the older population, and the proportion of the population receiving carer's allowance (cash benefit for caregivers who provide care at least 35 hours per week and earn less than £139 (US $177) per week in 2023).

Analysis

Multivariate regression analysis was performed. The main analysis used a cross-sectional ordinary least squares (OLS) estimation. The OLS approach was deemed appropriate for this analysis because our dependent variables are continuous. To correct the right-skewed distribution of the dependent variables, a natural log transformation was applied before analysis. This transformation helps ensure the normality assumption of OLS regression is satisfied, a method commonly used in previous studies [27,28]. We were limited to using SALT data from 2021 as the data for 2022 were not yet published at the time of analysis. As such, there is a small mismatch in timings between our categorization of front door arrangements and the care use data. Nonetheless, given that the rate of change for these variables is relatively slow, we argue that this is an acceptable limitation.

Listwise deletion was used to handle missing values on 2 dependent variables (ie, LTC for older people and OLLS). In both cases, less than 0.7% (1/147) of the cases were missing. Given the potential problem of multicollinearity, we conducted collinearity diagnostics. The mean variance inflation factors ranged from 1.28 to 2.49 and thus did not exceed the suggested threshold (variance inflation factors >10), indicating no evidence of multicollinearity [29].

Ethical Considerations

This study used publicly available data aggregated at the local authority level. The data did not involve the collection of any personally identifiable information, nor did it involve direct interaction with human participants. As a result, this research did not require ethical approval from an ethics committee.

Results

Descriptive Statistics

Table 1 summarizes relevant characteristics of the 147 local authorities in 2021 by age group, that is, all adults aged 18 years and older, younger people aged between 18 years and 64 years, and older people aged 65 years and older. The average rate of using LTC, STC, and OLLS per 100,000 population for all adults was 4.73, 7.04, and 11.53, respectively. Approximately 20% (27/147) of local authorities provided digital, web-based easy-read information online information about adult social care systems. The proportion of local authorities that provided online digital self-assessment for adults and caregivers was 25% (37/147) and 27% (40/147), respectively. The average adult population (aged 18 years and older) across local authorities was 294,000, while the average older population (aged 65 years and older) was 63,000.

In addition, we used chi-square tests to identify whether the provision of the digital front door varied among the 9 larger administrative areas in England. The results are presented in Table S1 in Multimedia Appendix 2, which suggests significant differences across regions. Among the 9 regions, the East of England had the highest proportion of local authorities that offered easy-read information online information, and the Southeast had the highest proportion of local authorities providing online self-assessment for adults and caregivers. In contrast, local authorities in the Northeast and the Northwest did not offer easy-read information online information, and the Southwest had the lowest proportion of local authorities providing online self-assessments for adults and caregivers.

XSL•FO
RenderX

Table 1. Characteristics of local authorities (n = 147) in England in 2021 by age group.

Variable	Age 18 years and older, mean (SD)	Age 18 years and 64 years, mean (SD)	Age 65 years and older, mean (SD)
Dependent variables			
Long-term care rate	4.73 (3.30)	3.55 (2.57)	8.92 (5.60)
Short-term care rate	7.04 (7.36)	5.53 (6.34)	12.43 (11.39)
Ongoing low-level support rate	11.53 (11.92)	8.90 (9.16)	20.68 (21.33)
Independent variables			
Easy-read information (=1)	0.18 (0.39)	0.18 (0.39)	0.18 (0.39)
Self-assessment for adults (=1)	0.25 (0.44)	0.25 (0.44)	0.25 (0.44)
Self-assessment for caregivers (=1)	0.27 (0.45)	0.27 (0.45)	0.27 (0.45)
Covariates			
Population size (100,000)	2.94 (2.22)	2.24 (1.63)	0.70 (0.63)
Proportion of older people aged 80 years and older	0.27 (0.02)	—[a]	0.27 (0.02)
Proportion of older population receiving Attendance Allowance	0.50 (0.07)	—	0.50 (0.07)
Proportion of older population receiving pension credit	0.54 (0.24)	—	0.54 (0.24)
Proportion of population receiving caregiver's allowance	0.10 (0.04)	0.11 (0.04)	0.10 (0.04)

[a]Not applicable.

Digital Front Doors and Social Care Support

Table 2 shows the association between the digital front door and the rate of using social care support. Panels 1, 2, and 3 show the estimates for LTC, STC, and OLLS, respectively. For each outcome, we investigated the association with the digital front door by three age groups: (1) all adults aged 18 years and older, (2) younger people aged between 18 years and 64 years, and (3) older people aged 65 years and above. The rate of using OLLS for all adults (β coefficient=0.54; *P*=.03), younger people (β coefficient=0.48; *P*=.04), and older people (β coefficient=0.53; *P*=.04) was positively associated with providing easy-read information online information. Providing online self-assessment tools for adults and caregivers was not significantly related to the rate of using OLLS. Providing easy-read information online information and online self-assessments for adults and caregivers were not significantly associated with LTC and STC.

Given the small sample size in our main analysis, we conducted a robustness check by also including care use data for 2020, assuming the same configuration of easy-read information and self-assessment as for 2021. The results are presented in Table S3 in Multimedia Appendix 2. We found that providing easy-read information online information was still positively associated with the rate of using OLLS, regardless of age group. However, providing online self-assessment for adults was now positively associated with the rate of using OLLS for all age groups using this larger sample.

Given that LTC and STC are general categories that encompass many types of services, we investigated the association between the digital front door and specific subdivisions of these services including restorative services, community care, residential care, and nursing care. The results of this analysis are summarized in Table S4 in Multimedia Appendix 2, together with the relationship between the digital front door and funded social care support. All associations between these types of services and providing easy-read information and online self-assessment were non-significant for both adults and caregivers.

XSL•FO
RenderX

Table 2. Results of multivariate regression analysis examining the association between the digital front door and social care support in England in 2021

	Age 18 years and older		Age 18 years and 64 years		Age 65 years and older	
	Values	*P* value	Values	*P* value	Values	*P* value
Panel 1: Long-term care						
Easy-read information, β coefficient (SE)	–0.15 (0.15)	.32	–0.14 (0.15)	.34	–0.03 (0.10)	.78
Self-assessment for adults, β coefficient (SE)	0.10 (0.12)	.42	0.07 (0.13)	.56	0.08 (0.12)	.48
Self-assessment for caregivers, β coefficient (SE)	–0.04 (0.11)	.72	–0.01 (0.13)	.91	–0.06 (0.11)	.59
Covariates[a]	Yes	—[b]	Yes	—	Yes	—
Local authorities, n	147	—	146	—	146	—
F test (*df*)	3.66 (8)	<.001	3.63 (5)	.004	1.67 (8)	.11
R^2	0.22	—	0.13	—	0.10	—
Panel 2: Short-term care						
Easy-read information, β coefficient (SE)	–0.32 (0.24)	.19	–0.35 (0.26)	.17	–0.30 (0.22)	.18
Self-assessment for adults, β coefficient (SE)	–0.21 (0.25)	.39	–0.25 (0.25)	.33	–0.19 (0.25)	.44
Self-assessment for caregivers, β coefficient (SE)	0.22 (0.25)	.38	0.25 (0.25)	.33	0.22 (0.26)	.39
Covariates	Yes	—	Yes	—	Yes	—
Local authorities, n	147	—	147	—	147	—
F test (*df*)	1.73 (8)	.10	2.27 (5)	.05	0.98 (8)	.11
R^2	0.11	—	0.08	—	0.06	—
Panel 3: Ongoing low-level support						
Easy-read information, β coefficient (SE)	0.54 (0.24)	.03	0.48 (0.24)	.04	0.53 (0.25)	.04
Self-assessment for adults, β coefficient (SE)	0.37 (0.27)	.17	0.34 (0.28)	.23	0.35 (0.27)	.19
Self-assessment for caregivers, β coefficient (SE)	–0.04 (0.28)	.88	–0.03 (0.29)	.91	–0.05 (0.28)	.85
Covariates	Yes	—	Yes	—	Yes	—
Local authorities, n	146	—	146	—	146	—
F test (*df*)	6.86 (8)	<.001	8.16 (5)	<.001	4.42 (8)	<.001
R^2	0.23	—	0.20	—	0.19	—

[a]We presented estimators of all covariates in Table S2 in Multimedia Appendix 2.

[b]Not applicable.

Discussion

Principal Findings

To our knowledge, this is the first study that provides information regarding the prevalence of digital easy-read information and self-assessment tools from official websites in England and investigates the association between these tools and access to social care support. Using data at the local authority level, we found that only 20% (27/147) of local authorities provided online, digital easy-read information, and approximately 25% (37/147) used digital self-assessment approaches to promote service access. This is important as we found local authorities that provided easy-read information had a higher rate of using OLLS than those without such a front door arrangement.

Comparison With Previous Work

Providing easy-read information has been shown to reduce learning costs and enhance comprehension [30,31]. Our study adds to this literature by demonstrating the association between easy-read information online information and service access. The availability of digital easy-read information was positively associated with the use of OLLS, but we found no significant association with rates of LTC or STC use. One possible interpretation as to why easy-read information online information only facilitated access for those eligible for low-level ongoing care is that people using web-based digital tools to access care services may have less physical and cognitive impairment, that is lower needs, than their counterparts [32,33]. Conversely, people with more severe needs may have a higher likelihood of being referred directly to a full professional assessment, bypassing the need for online access or assessment. People with lower needs are also more likely to

XSL•FO
RenderX

be assessed as needing OLLS rather than long-term and STC [34].

We did not find a significant association between the availability of online self-assessment tools for adults and caregivers, and access to social care support (although alternative analysis using 2 years of data did find a positive association). In practice, the provision of an online self-assessment facility does not necessarily imply that such a system is fully used. One pilot program in England showed that about two-thirds of people requested a face-to-face professionally led assessment rather than a self-assessment, and this was especially true for those with urgent and complex needs [35]. This lower rate of online self-assessment use may explain the nonsignificant association between online self-assessment and service use. In contrast, people with less intensive needs typically prefer using a self-assessment, especially when professional assistance is readily available [16]. Therefore, policy makers could develop both professionally led assessments and online self-assessments to meet the individual preferences of service users. Also, online self-assessment could be augmented with online support from care professionals (eg, through a chat function) to further assist users.

Limitations

This study has several limitations to consider. First, the digital front door is a broad concept with multiple policy components, which may not be fully measured. To address this challenge, we investigated commonly discussed barriers to service access and digital front door arrangements while acknowledging that some digital channels are not operationalized in this study. For example, individuals may use information and assistance from nonprofit organizations' websites, such as Age UK and Carers UK. Third-sector websites are an essential aspect of the digital front door, and future research could examine their impact on service access. In addition, although local authorities that provide easy-read information typically incorporate the term "easy-read information," we acknowledge that some websites offer such content without explicitly using this exact keyword.

Second, this study provides an up-to-date snapshot of the digital front door in England by offering a cross-sectional analysis of current policy practices. Information regarding the digital front door between late 2022 and early 2023 was combined with social care data from 2021 to 2022 as these were the most up-to-date statistics available at the time of analysis. As noted above, given that general changes in service use are relatively slow, we believe this limitation to be minimal. Nonetheless, we do accept the possibility that, in some cases, the 2022 and 2023 social care data may better reflect the association with current digital front door policies, leading to smaller estimated effect sizes in the current analysis.

Finally, the cross-sectional design of this study limits our ability to establish causal relationships between the digital front door and social care use. While we assumed that these digital arrangements promote access to social care, the possibility of reverse causality cannot be excluded. For example, local authorities facing high demand for OLLS may introduce online digital easy-read information to reduce service inquiries, resulting in a potentially spurious correlation between these digital systems and service use. Therefore, our findings should be interpreted as indicative associations rather than causal relationships. Future research could use longitudinal data and causal inference methods to more rigorously examine the impact of the digital front door on access to social care support.

Strengths and Implications

Despite these limitations, this study is among the first to empirically examine the association between providing easy-read information and self-assessment tools from official websites and service access. The findings presented here could inform policy makers interested in developing digital channels for service access. They are also relevant for discussions about increasing service use, reducing unmet needs, and enhancing the well-being of service users.

Our findings have implications for policy makers and practitioners who aim to promote service use. Though digital channels have the potential to save costs, we found relatively few local authorities provided online digital easy-read information and online self-assessment tools, as noted above. Although we do not have information about the costs of implementing such a facility, our findings suggest a positive association with (low-level) service use. This association could potentially lead to lower levels of unmet need and better care-related outcomes, which may, in turn, lead to cost-savings downstream. Given that digital channels might facilitate service access, there is a case for a wider roll-out of digital easy-read information facilities, particularly where the costs of implementation are minimal.

Acknowledgments

This research was funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Kent, Surrey, Sussex (grant NIHR 200179). The views are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health and Social Care.

Data Availability

All data generated or analyzed during this study are included in this published article and its Multimedia Appendix files.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Data.
[XLSX File (Microsoft Excel File), 209 KB - jmir_v27i1e53205_app1.xlsx]

Multimedia Appendix 2
Supplementary results.
[DOCX File , 49 KB - jmir_v27i1e53205_app2.docx]

References

1. Bieber A, Nguyen N, Meyer G, Stephan A. Influences on the access to and use of formal community care by people with dementia and their informal caregivers: a scoping review. BMC Health Serv Res 2019 Feb 01;19(1):88 [FREE Full text] [doi: 10.1186/s12913-018-3825-z] [Medline: 30709345]
2. Felix HC, Ali M, Bird TM, Cottoms N, Stewart MK. Are community health workers more effective in identifying persons in need of home and community-based long-term services than standard-passive approaches. Home Health Care Serv Q 2019;38(3):194-208. [doi: 10.1080/01621424.2019.1604461] [Medline: 31060448]
3. Penkunas MJ, Eom KY, Chan AWM. Classification trees for identifying non-use of community-based long-term care services among older adults. Health Policy 2017 Oct;121(10):1093-1099. [doi: 10.1016/j.healthpol.2017.05.008] [Medline: 28942091]
4. Xie Y, Hamilton M, Peisah C, Anstey KJ, Sinclair C. Navigating community-based aged care services from the consumer perspective: a scoping review. Gerontologist 2024 Feb 01;64(2):gnad017. [doi: 10.1093/geront/gnad017] [Medline: 37120292]
5. Rastogi N. Healthcare's new frontier: the digital front door. BMJ Innov 2022 Mar 15;8(2):129-132. [doi: 10.1136/bmjinnov-2021-000874]
6. Kunonga TP, Spiers GF, Beyer FR, Hanratty B, Boulton E, Hall A, et al. Effects of digital technologies on older people's access to health and social care: umbrella review. J Med Internet Res 2021 Nov 24;23(11):e25887 [FREE Full text] [doi: 10.2196/25887] [Medline: 34821564]
7. Napier A, Rhodes B, Khavjou O, Knowles M, Popham L, Walsh EG, et al. Impact of a web-based information and referral tool on access to federal and state programs for older adults. J Aging Soc Policy 2023 May 04;35(3):360-373. [doi: 10.1080/08959420.2021.1926206] [Medline: 34016014]
8. Bolton J. Six Steps to Managing Demand in Adult Social Care: A Performance Management Approach. Oxford, UK: Oxford Brookes University; 2017.
9. Manthorpe J, Moriarty J, Damant J, Brimblecombe N, Marczak J, Knapp M, et al. Local online information for carers in England: content and complexity. Practice 2023;35(7):1-25. [doi: 10.1080/09503153.2023.2232567]
10. Baxter K, Wilberforce M, Birks Y. What skills do older self-funders in England need to arrange and manage social care? Findings from a scoping review of the literature. Brit J Soc Work 2021;51(7):2703-2721. [doi: 10.1093/bjsw/bcaa102]
11. Adult Social Care Front Door: Lambeth. Local Government Association. 2019. URL: https://www.local.gov.uk/case-studies/adult-social-care-front-door-lambeth [accessed 2023-05-11]
12. Quigley R, Foster M, Harvey D, Ehrlich C. Entering into a system of care: a qualitative study of carers of older community-dwelling Australians. Health Soc Care Community 2022;30(1):319-329. [doi: 10.1111/hsc.13405] [Medline: 33955616]
13. Turnpenny A, Beadle-Brown J. Use of quality information in decision-making about health and social care services--a systematic review. Health Soc Care Community 2015;23(4):349-361. [doi: 10.1111/hsc.12133] [Medline: 25243842]
14. Laragy C, David C, Moran N. A framework for providing information in individualised funding programmes. Qual Soc Work 2015;15(2):190-208. [doi: 10.1177/1473325015589402]
15. Baxter K, Glendinning C. Making choices about support services: disabled adults' and older people's use of information. Health Soc Care Community 2011;19(3):272-279. [doi: 10.1111/j.1365-2524.2010.00979.x] [Medline: 21158999]
16. Abendstern M, Hughes J, Clarkson P, Tucker S, Challis D. Exploring the contribution of self-assessment to preventative services in social care. British Journal of Social Work 2012;44(3):729-746. [doi: 10.1093/bjsw/bcs144]
17. Clarkson P, Hughes J, Challis D, Thorley L, Kilshaw C. Targeting, care management and preventative services for older people: the cost-effectiveness of a pilot self-assessment approach in one local authority. Brit J Soc Work 2010;40(7):2255-2273. [doi: 10.1093/bjsw/bcq042]
18. Herd P. Improving older adults' health by reducing administrative burden. Milbank Q 2023;101(S1):507-531 [FREE Full text] [doi: 10.1111/1468-0009.12629] [Medline: 37096624]
19. Griffiths P. Self-assessment of health and social care needs by older people: a review. Br J Community Nurs 2005;10(11):520, 522-520, 527. [doi: 10.12968/bjcn.2005.10.11.19963] [Medline: 16301927]
20. Accessing adult social care in England. Office for National Statistics. 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/socialcare/methodologies/accessingadultsocialcareinengland [accessed 2023-05-24]

XSL•FO
RenderX

21. Care and Support Statutory Guidance. Department of Health. 2014. URL: https://www.gov.uk/government/publications/care-act-statutory-guidance/care-and-support-statutory-guidance#chapter-3 [accessed 2023-05-24]
22. Measures from the adult social care outcomes framework. NHS Digital. 2023. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-social-care-outcomes-framework-ascof [accessed 2024-12-10]
23. ADASS spring survey 2023: care waiting lists down but needs increasing. ADASS. 2023. URL: https://www.adass.org.uk/adass-spring-survey-2023-care-waiting-lists-down-but-needs-increasing#:~:text=A [accessed 2023-09-21]
24. Adult social care activity and finance report, England, 2021-22. NHS England. 2022. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-social-care-activity-and-finance-report/2021-22 [accessed 2022-12-12]
25. What the public want from social care. Healthwatch England. 2018. URL: https://www.healthwatch.co.uk/sites/healthwatch.co.uk/files/20180801 [accessed 2024-12-10]
26. Fernandez JL, Forder J. Local variability in long-term care services: local autonomy, exogenous influences and policy spillovers. Health Econ 2015;24 Suppl 1:146-157 [FREE Full text] [doi: 10.1002/hec.3151] [Medline: 25760589]
27. Gaínza-Lein M, Sánchez Fernández I, Jackson M, Abend NS, Arya R, Brenton JN, Pediatric Status Epilepticus Research Group. Association of time to treatment with short-term outcomes for pediatric patients with refractory convulsive status epilepticus. JAMA Neurol 2018;75(4):410-418 [FREE Full text] [doi: 10.1001/jamaneurol.2017.4382] [Medline: 29356811]
28. He Y, Yang Q, Alish Y, Ma S, Qiu Z, Chen J, et al. Relationship between product features and the prices of e-cigarette devices sold in web-based vape shops: comparison study using a linear regression model. JMIR Form Res 2024;8:e49276 [FREE Full text] [doi: 10.2196/49276] [Medline: 38723251]
29. Alin A. Multicollinearity. WIREs Computational Stats 2010;2(3):370-374. [doi: 10.1002/wics.84]
30. Chinn D, Homeyard C. Easy read and accessible information for people with intellectual disabilities: is it worth it? A meta-narrative literature review. Health Expect 2017;20(6):1189-1200 [FREE Full text] [doi: 10.1111/hex.12520] [Medline: 27862757]
31. Sutherland RJ, Isherwood T. The evidence for easy‐read for people with intellectual disabilities: a systematic literature review. Policy Practice Intel Disabi 2016;13(4):297-310. [doi: 10.1111/jppi.12201]
32. Wen W, Zhang Y, Shi W, Li J. Association between internet use and physical health, mental health, and subjective health in middle-aged and older adults: nationally representative cross-sectional survey in China. J Med Internet Res 2023;25:e40956 [FREE Full text] [doi: 10.2196/40956] [Medline: 36943368]
33. Heponiemi T, Kainiemi E, Virtanen L, Saukkonen P, Sainio P, Koponen P, et al. Predicting internet use and digital competence among older adults using performance tests of visual, physical, and cognitive functioning: longitudinal population-based study. J Med Internet Res 2023;25:e42287 [FREE Full text] [doi: 10.2196/42287] [Medline: 37145836]
34. Short and Long Term (SALT) Data Return 2021-22 guidance. NHS Digital. 2021. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/social-care-collection-materials-2022/salt-data-return-2021-2022-guidance [accessed 2024-12-10]
35. Clarkson P, Brand C, Hughes J, Challis D, Tucker S, Abendstern M. Cost effectiveness of pilot self-assessment sites in community care services in England. Aust Health Rev 2013;37(5):666-674. [doi: 10.1071/AH13056] [Medline: 24160485]

Abbreviations

LTC: long-term care
NIHR: National Institute for Health and Care Research
OLLS: ongoing low-level support
OLS: ordinary least squares
SALT: short and long term
STC: short-term care

Edited by A Coristine; submitted 05.02.24; peer-reviewed by Y Xie, F Longo; comments to author 31.05.24; revised version received 25.07.24; accepted 10.09.24; published 02.01.25.

Please cite as:
Zhang J, Prunty JE, Charles AC, Forder J
Association Between Digital Front Doors and Social Care Use for Community-Dwelling Adults in England: Cross-Sectional Study
J Med Internet Res 2025;27:e53205
URL: https://www.jmir.org/2025/1/e53205
doi:10.2196/53205
PMID:

XSL•FO
RenderX

©Jinbao Zhang, Jonathan E Prunty, Alison C Charles, Julien Forder. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Use of mHealth Technology for Improving Exercise Adherence in Patients With Heart Failure: Systematic Review

Pallav Deka[1], APRN, MS, PhD; Erin Salahshurian[2], BSN, RN; Teresa Ng[1], BSN, RN; Susan W Buchholz[1], APRN, PhD; Leonie Klompstra[3], PhD; Windy Alonso[2], RN, PhD

[1]College of Nursing, Michigan State University, East Lansing, MI, United States

[2]College of Nursing, University of Nebraska Medical Center, Omaha, NE, United States

[3]Department of Health, Medicine and Care Sciences, Linkoping University, Linkoping, Sweden

Corresponding Author:
Pallav Deka, APRN, MS, PhD
College of Nursing
Michigan State University
1355 Bogue Street, C247
East Lansing, MI, 48824
United States
Phone: 1 5174328309
Email: pdeka@msu.edu

Abstract

Background: The known and established benefits of exercise in patients with heart failure (HF) are often hampered by low exercise adherence. Mobile health (mHealth) technology provides opportunities to overcome barriers to exercise adherence in this population.

Objective: This systematic review builds on prior research to (1) describe study characteristics of mHealth interventions for exercise adherence in HF including details of sample demographics, sample sizes, exercise programs, and theoretical frameworks; (2) summarize types of mHealth technology used to improve exercise adherence in patients with HF; (3) highlight how the term "adherence" was defined and how it was measured across mHealth studies and adherence achieved; and (4) highlight the effect of age, sex, race, New York Heart Association (NYHA) functional classification, and HF etiology (systolic vs diastolic) on exercise adherence.

Methods: We searched for papers in PubMed, MEDLINE, and CINAHL databases and included studies published between January 1, 2015, and June 30, 2022. The risk of bias was analyzed.

Results: In total, 8 studies (4 randomized controlled trials and 4 quasi-experimental trials) met our inclusion and exclusion criteria. A moderate to high risk of bias was noted in the studies. All studies included patients with HF in NYHA classification I-III, with sample sizes ranging from 12 to 81 and study durations lasting 4 to 26 weeks. Six studies had an equal distribution of male and female participants whose ages ranged between 53 and 73 years. Videoconferencing was used in 4 studies, while 4 studies used smartphone apps. Three studies using videoconferencing included an intervention that engaged participants in a group setting. A total of 1 study used a yoga program, 1 study used a walking program, 1 study combined jogging with walking, 1 study used a cycle ergometer, 2 studies combined walking with cycle ergometry, and 1 study used a stepper. Two studies incorporated resistance exercises in their program. Exercise programs varied, ranging between 3 and 5 days of exercise per week, with exercise sessions ranging from 30 to 60 minutes. The Borg rating of perceived exertion scale was mostly used to regulate exercise intensity, with 3 studies using heart rate monitoring using a Fitbit. Only 1 study implicitly mentions developing their intervention using a theoretical framework. Adherence was reported to the investigator-developed exercise programs. All studies were mostly feasibility or pilot studies, and the effect of age, sex, race, and NYHA classification on exercise adherence with the use of mHealth was not reported.

Conclusions: The results show some preliminary evidence of the feasibility of using mHealth technology for building exercise adherence in patients with HF; however, theoretically sound and fully powered studies, including studies on minoritized communities, are lacking. In addition, the sustainability of adherence beyond the intervention period is unknown.

(J Med Internet Res 2025;27:e54524) doi:10.2196/54524

XSL•FO
RenderX

KEYWORDS

adherence; activity monitors; exercise; heart failure; mHealth; mobile health; smartphone; videoconferencing; heart; mHealth technology; exercise programs; age; sex; race; telehealth technology; software apps; feasibility; mobile phone

Introduction

With the aging population, heart failure (HF) is a growing problem worldwide [1]. In the United States, 8 million individuals are projected to be diagnosed with HF, and the cost of its treatment is expected to reach US $70 billion by 2030 [2-4]. Irrespective of the etiology of the disease (HF with preserved vs reduced ejection fraction; ischemic vs nonischemic), patients with HF experience symptoms of fatigue, exercise intolerance, palpitations, and over time, experience a decline in physical functioning [5-10]. These symptoms worsen as the disease progresses, often designated by the New York Heart Association (NYHA) I-IV classification [11]. The treatment and management of HF are also complicated by patients experiencing peaks and troughs in symptom experience and often needing hospitalizations. HF is generally associated with advancing age and has the highest readmission rates among all chronic diseases, adding to the increase in health care costs [2,12]. As such, effective and efficient management of HF using both pharmacological and nonpharmacological methods is essential.

As a nonpharmacological method, exercise training interventions have been shown to decrease hospitalizations, increase exercise capacity, and improve quality of life [13]. Exercise is different from physical activity and has been defined as a subset of physical activity that is planned, structured, purposeful, and performed with the objective of improving physical fitness [14]. The landmark multisite Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training trial established the safety of moderate-intensity aerobic exercise in HF and found evidence of reduced rehospitalization rates in patients who were adherent versus nonadherent to the recommended exercise [15]. Based on clinical evidence, exercise is considered a class I recommendation in adults with HF, meaning that its benefits greatly outweigh the risks [16-18]. Current guidelines recommend that patients with HF continue to include 150 minutes per week of moderate-intensity aerobic exercise supplemented with 2-3 days of resistance exercise training [19]. However, desired adherence to the recommended exercise in this population has been hard to achieve [15,20,21]. The World Health Organization [22] defines adherence as “the extent to which a person’s behavior—taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider.” This definition differentiated adherence from compliance and attempted to highlight the importance and need for lifestyle and behavioral changes to successfully achieve adherence [23].

In recent years, advances in mobile health (mHealth) technology have provided the opportunity for bridging access to care by delivering interventions to participants’ homes using video calling, chatbot messaging, cloud-based digital voice response, automated emails, phone counseling, and other tools for motivation and engagement [24]. Such technologies have also significantly improved the ability to remotely monitor a person's health status and intervention outcomes [24]. A systematic review of reviews shows that their app has also been used in the management of various health conditions such as asthma, chronic lung disease, diabetes, cardiac rehabilitation, hypertension, and HIV management, among many others [24]. Additionally, mHealth has also been implemented for behavioral or lifestyle changes that include weight loss, physical activity, smoking cessation, and sexual behavior [24]. Across diverse clinical populations, the effectiveness of home-based exercise programs is found to be similar to that of facility-based exercise programs [25]. In our previous study published in 2017 [26], we updated the state of the science from 2009 [27] in relation to adherence to exercise in patients with HF. In that systematic review, we found a mix of home-based and facility-based interventions and highlighted the lack of web-based interventions to improve adherence to recommended exercise guidelines in this population [26]. We did not find any studies at that time with mHealth interventions specifically targeting and reporting exercise adherence in HF. With significant advances in mHealth technology and access to smartphones among all ages [28], it is important to update the literature and summarize the work that has been done since to improve adherence to exercise in patients with HF using newer technologies. Demographic and socioeconomic variables can have a profound impact on technology adoption and use [29]. There can be unique challenges for implementing an mHealth intervention depending on the type of technology used and the complexity of the exercise program. This review was performed to systematically synthesize the literature on the use of mHealth technology that reported adherence to exercise in patients with HF, the majority of whom tend to be older adults. Specifically, this review was done to (1) describe study characteristics of mHealth interventions for exercise adherence in HF including details of sample demographics, sample sizes, exercise program, and theoretical frameworks; (2) summarize types of mHealth technology used to improve exercise adherence in patients with HF; (3) highlight how the term “adherence” was defined and how it was measured across mHealth studies and adherence achieved; and (4) highlight the effect of age, sex, race, NYHA functional classification, and HF etiology (systolic vs diastolic) on exercise adherence. Finally, we conclude by summarizing our findings and providing suggestions for future studies.

Methods

Study Design

The PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) statement guided this review [30]. A health science librarian worked closely with the lead authors (PD, WA, ES, and TN) on narrowing the search terms. The search strategy included the terms (“heart failure” AND “exercise” AND “adherence”), AND (“mhealth” OR “telehealth”) in the title or abstract. Three databases (PubMed, MEDLINE, and CINAHL) were searched. Studies were eligible if they met the criteria outlined in Textbox 1. These inclusion dates were chosen

XSL•FO
RenderX

because a previous systematic review in the same topic area had included studies published up to 2015 [26].

Textbox 1. Inclusion and exclusion study criteria.

Inclusion criteria

- Primary studies on patients with heart failure
- Used an intervention that included mobile health or telehealth technology
- Experimental or quasi-experimental study
- Primary or secondary outcome reports of exercise adherence
- Published between January 1, 2015, and June 30, 2022
- Published in the English language

Exclusion criteria

- Studies that used only cell phone text messaging as an intervention

Screening and Data Extraction

All papers obtained from the literature search were uploaded and screened using the Rayyan program. The following 2-step screening process was used: (1) 4 authors (PD, WA, ES, and TN) independently screened titles and abstracts, and (2) all 4 authors screened the full texts of the included papers from step 1 to be included in this review. During every step, the 4 authors labeled a paper as "accepted," indicating that the paper met inclusion criteria; "not accepted," indicating that the paper did not meet inclusion criteria; or "maybe," indicating that the paper needed to be discussed prior to making a judgment that the paper was acceptable for the next step. As a group, the 4 authors discussed discrepancies and "maybe" papers to include or not include in the final analysis.

An author-developed extraction method was developed according to our prior systematic review [26]. Data including study characteristics (country, study design, theoretical framework, and length of study), sample characteristics (sample size and HF sample characteristics), intervention characteristics (mHealth technology and intervention components), outcomes (primary outcome, secondary outcomes, and outcome measures), results, and conclusions were extracted. Two authors (ES and TN) extracted relevant data from each paper. Two other authors (PD and WA) performed quality checks to ensure accuracy and completeness. Discrepancies during the extraction process were discussed with these 4 authors. The quality appraisal and risk of bias for the included studies were assessed using two distinct tools: (1) the second version of the Cochrane Risk of Bias tool for randomized trials and (2) the Risk of Bias In Non-Randomized Studies of Interventions-I tool for quasi-experimental studies. Each domain was categorized as low, moderate, serious, critical, or no information [31]. Both tools used the scores from these domains to determine the overall risk of bias [31,32].

Results

Overview

Our initial search yielded a total of 90 papers. The removal of duplicates yielded 82 papers to screen. In total, 55 papers were extracted during the first screening process, resulting in 27 papers being sought for full-text retrieval. The 55 papers did not meet inclusion and exclusion criteria from the title and abstract screening for not including participants with HF, not using mHealth or telehealth intervention, not using an exercise intervention, not reporting adherence, using cell phone SMS text messaging for intervention, not a primary study, and not published in English. Following the second screening process, 19 papers were excluded for the following reasons: review papers (n=8), did not report adherence (n=4), no exercise intervention using mHealth (n=3), papers reporting protocols (n=3), and papers did not meet date requirements (n=1). Only 8 papers were included for data extraction. The PRISMA diagram (Figure 1) highlights the screening process. The papers included 4 randomized controlled trials (RCTs) [33-36] and 4 quasi-experimental trials [37-40]. A total of 4 studies are from Asia [33,34,37,38], 3 from the United States [35,39,40], and 1 from Australia [36]. Descriptions of these studies are provided in Table 1.

XSL•FO
RenderX

Figure 1. PRISMA diagram detailing the screening process.

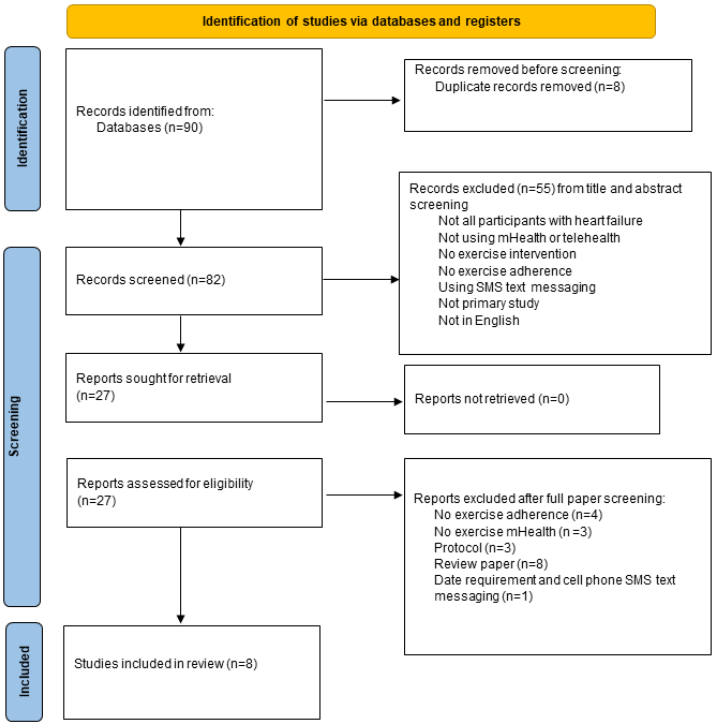

Table 1. Study location, study design, sample characteristics including participant's New York Heart Association (NYHA) classification, ejection fraction, age, recruitment setting, and reports of baseline evaluation of exercise adherence.

Author (year), country	Design (groups, n); length of study	Sample size, n; sex; race	NYHA classification; ejection fraction, mean (SD); age (years), mean (SD)	Recruitment site	Baseline evaluation of exercise adherence
Tsai et al (2022) [37], Taiwan	Quasi-experimental (2); 6 months	81 (intervention=40; control=41)	Not reported; 33.5% (11.12%); 73.3 (5.0)	Hospital prior to discharge	Not reported
Nagatomi et al (2022) [33], Japan	RCT[a] (2); 3 months	30 (intervention=15; control=15); male=53%; Asian	Class II-III; 42.2% (17.4%); 63.7 (10.1)	Outpatient clinic	Engaged in outpatient rehabilitation >2 times per week
Liu and Liu (2022) [34], China	RCT (2); 12 weeks	60 (intervention=30; control=30); male=17, female=13; Asian	Class II-III; 36.02% (5.12%); 53.27 (7.1)	Hospital prior to discharge	Not reported
Kikuchi et al (2021) [38], Japan	Quasi-experimental (1); 12 weeks	10; male=55%, female=45%; Asian	Class II-III; 40% (not reported); 76 (7)	Hospital prior to discharge and outpatient	Good (more than 90 minutes a week), fair (30-90 minutes a week), and poor (less than 30 minutes a week)
Deka et al (2019) [35], United States	RCT (2); 8 weeks	30 (intervention=15; control=15); male=63%; female=37%; White=100%	Class I-III; 41% (12.6%); 61.7 (11.6)	Outpatient clinic	Self-reported; excluded if >2 days per week of 30 minutes exercise
Lloyd et al (2019) [39], United States	Quasi-experimental (1); 1 month	12; not reported	Class II-III; 40% (not reported); 67 (not provided)	Prior to hospital discharge	Not reported
Hwang et al (2017) [36], Australia	RCT (2); 8 weeks	53 (intervention=24; control=29); male=75%; female=25%	Class I-III; 35% (17%); 67 (12)	Prior to hospital discharge	Not reported
Donesky et al (2017) [40], United States	Quasi-experimental (2 nonrandomized groups); (12 weeks)	15 (intervention=7; control=8); male=43%; female=57%	Class II-III; not reported; 71 (8.5)	Outpatient clinic	Not reported

[a]RCT: randomized controlled trial.

XSL•FO
RenderX

Study Characteristics of mHealth Interventions for Exercise Adherence in HF

This review builds on research to describe study characteristics of mHealth interventions for exercise adherence in HF, including details of sample demographics, sample sizes, exercise programs, and theoretical frameworks. As detailed in Table 1, the sample sizes ranged from 12 to 81, with study durations lasting 4 to 26 weeks. The studies from Asia included Asians in their study, and the studies from the United States and Australia primarily included White patients with HF. While 2 studies had a larger proportion of male participants (63% and 75%) [35,36], other studies seem to have an equal representation of male and female participants. One study did not report the race of the participants enrolled in the study [39]. The mean ages of the participants included in the studies range from 53 to 73 years. The NYHA classifications ranged from class I to class III, with mean ejection fractions ranging from 33.5% to 45%. One study did not report the ejection fraction of the participants [40]. In total, 3 studies recruited participants from a cardiac outpatient clinic [33,35,40], while 5 studies recruited participants from an inpatient hospital setting [34,36-39].

With exercise being performed at home, baseline training for patients with HF to safely perform exercise is important. One study incorporated 1 week of center-based cardiac rehabilitation exercise training before transitioning to a home setting [37]. Two studies provided 1 training session on the exercise program at baseline [35,38]. One study indicated that the participants received family cardiac rehabilitation prior to being recruited [34]. Among the studies that were RCTs, 1 study reported that the control group was provided the same exercise prescription as the intervention group [35], and 1 study provided the control group with family cardiac rehabilitation training guided by mobile medical technology [34]. Other studies did not provide details of the number of sessions needed to train the participants in the prescribed exercise program.

For exercise modality, 1 study used a yoga program [40], 1 study used a walking program [35], 1 study combined jogging with walking [34], 1 study used a cycle ergometer [38], 2 studies combined walking with cycle ergometry [33,37], and 1 study used a stepper [39]. One study did not provide details of the aerobic component of the exercise program [36]. Two studies added a resistance exercise component to the exercise prescription [33,36]. One study used multiple types of aerobic exercises (such as walking or using a cycle ergometer, a stepper, and an elliptical) and incorporated 1 day per week of facility-based exercise [37]. Studies that prescribed a primarily home-based exercise prescription used a walking or cycle ergometer program.

Rating of perceived exertion (RPE) was the most used tool for regulating the intensity of exercise [33-37]. Overall, participants in these studies were asked to keep their RPE<14 on the Borg (6-20) RPE scale, which is suggestive of moderate-intensity exertion [41]. Additionally, a few studies also used heart rate for moderate exercise intensity regulation. These strategies to regulate intensity using heart rate included keeping the exercise heart rate at resting heart rate+30 beats per minute [37], resting heart rate+(maximum exercise heart rate–resting heart rate)×0.4 [34], and below maximal heart rate during baseline training on the exercise program [35]. Three studies did not provide details on how intensity of exercise was regulated [38-40].

Prescribed exercise durations were between 30 and 60 minutes. The exercise prescriptions used by the studies varied but ranged from 3 to 5 days per week. While 1 study advised participants to exercise daily [39], 2 studies recommended 2 days per week of exercise [36,40]. One study did not report on the frequency of recommended exercise per week [37]. We found that the exercise prescriptions in most studies were investigator-developed programs. Reference to the recommended guidelines of 150 minutes per week of moderate-intensity aerobic exercise for patients with HF was done in 1 study [35].

Only 1 study indicated developing their intervention using a theoretical platform [35]. The study used concepts from Bandura's self-efficacy theory and Ajzen's theory of planned behavior to bring about change in adherence behavior. The other studies did not specifically mention any theory.

Summarize Types of mHealth Technology Used to Improve Exercise Adherence in Patients With HF

Table 2 provides information about the mHealth technologies used and how they were incorporated into the intervention. As summarized in Table 2, videoconferencing was used in 4 studies [35,36,38,40], while 4 studies [33,34,37,39] used mobile apps for the devices used in the study. The apps were installed in smartphones (n=3) [33,34,37], laptops, tablets, or desktop computers using Wi-Fi or 3G wireless broadband connection (n=4) [35,36,38,39], and 1 study used a television that was connected to the internet [40]. Studies published in the past 3 years used smartphone-based apps, while studies that were a little older used more laptops, computers, and tablets. In the 4 studies that used a videoconferencing platform, videoconferencing was delivered using commercially available software such as Vidyo (Vidyo Inc) [35], Adobe Connect 9.2 [36], DocBox (DocBox, Inc) [40], and the Remohab integrated rehabilitation platform (Remohab) [38]. Two studies used a Fitbit activity monitor along with the Fitbit app [33,35]. Two studies used apps to transfer information to the physician's office via the cloud database [33,37]. Three studies using videoconferencing included interventions that were in a group setting [35,39,40]. One study used an app that provided voice prompts during exercise [34]. While 1 study simply engaged participants in cohorts (n=5) for a discussion session for social support [35], 2 studies focused on delivering an exercise program that included yoga (n=8) [40] and aerobic exercise (n=4) [36].

XSL•FO
RenderX

Table 2. mHealth technology used and details of the study intervention[a].

Study author(s) (year)	Type of mHealth technology used	Type of software	Devices used	Details of exercise intervention (individual or group)
Tsai et al (2022) [37]	App	HF[b] health management mobile app system platform	Smartphone	The exercise parameters were recorded on the HF health management mobile app system platform by each patient and daily transmission to the hospital's cloud database as home-based cardiac telerehabilitation. Monitoring parameters include body weight, blood pressure, resting heart rate, exercise heart rate, exercise time, and abnormal symptom signs.
Nagatomi et al (2022) [33]	App	Fitbit	Fitbit Inspire HR worn on nondominant hand; smartphone	Comprehensive home-based cardiac rehabilitation program that combines patient education, exercise guidance, and nutritional guidance using information and communication technology. Messages sent once a week through the Fitbit app or by telephone. Video instructions on performing exercises were provided using a QR code via smartphone.
Liu and Liu (2022) [34]	App	Exercise rehabilitation app software designed by the family cardiac exercise rehabilitation training research team	Android smartphone	The app provided exercise guidance via voice prompts on current exercise speed, heart rate, acceleration or deceleration reminder, exercise rhythm adjustment, etc.
Kikuchi et al (2021) [38]	Videoconferencing and remote monitoring	Integrated rehabilitation platform (Remohab Inc)	An Internet of Things–equipped ergometer (Charimo, Remohab Inc), an Android-compatible tablet (TAB3-X70L, Lenovo), and a wireless electrocardiographic monitoring device (hitoe, Toray)	In total, 30 minutes to install the system including Wi-Fi at the participants home and 60 minutes to instruct the use of the platform. Exercises were done individually. Participants exercised on a cycle ergometer for 30 minutes or less while videoconferencing with a nurse. For each session, exercise intensity was set by the attending physician. There was live monitoring of heart rate and electrocardiogram for arrhythmia that was transferred via Wi-Fi to the physician's office.
Deka et al (2019) [35]	Videoconferencing and app	Vidyo and Fitbit app	Fitbit Charge HR; tablet, computer, laptop; home Wi-Fi or wireless connectivity	Education on HF self-care on the following topics: understanding HF, exercise and activity with HF, how to follow a low sodium diet, HF medication, dealing with HF symptoms, depression and anxiety with HF, and managing lifestyle changes. Discussion on the previous week's adherence, barriers, and facilitators. Five participants in each cohort.
Lloyd et al (2019) [39]	App	REDCap[c]	Tablet computer with wireless connectivity	A 5-minute video with an older adult performing the intervention tasks was created and provided to participants. Feedback was provided in the form of personal data graphs.
Hwang et al (2017) [36]	Videoconferencing	Adobe Connect 9.2	Laptop computer, mobile broadband device connected to 3G wireless broadband internet	Physical therapist–guided telerehabilitation exercise in groups (up to 4 participants). In total, 15 minutes were spent at the start of each session on a discussion on educational topics that replicated a center-based program.
Donesky et al (2017) [40]	Videoconferencing	DocBox	The DocBox was connected to the participants' home television for live-streaming yoga classes	Group teleyoga. Postures included mountain, half-down dog, cat, triangle, supported bridge, simple twist, staff, corpse, and cobbler poses, with postures modified as needed to meet the physical ability of each participant. Yoga classes were designed to integrate breathing exercises (slow breathing and extended exhalation breathing), imagery, meditation, and relaxation.

[a] mHealth: mobile health.

[b]HF: heart failure.

[c]REDCap: Research Electronic Data Capture.

XSL•FO
RenderX

Adherence: Description, Measurement and Achievement

In our review, we found that, in the majority of studies, the term adherence was used to measure "adherence to a study-specific exercise program." Only 1 study used the term adherence in reference to the recommended exercise guidelines [35]. As seen in Table 2, all studies relied on an objective tool for measuring the amount of exercise to calculate adherence. In total, 2 studies used a Fitbit [33,35], 3 studies used a telemonitoring platform [34,37,38], and 2 studies directly observed exercise being performed during the videoconferencing session [36,40].

Table 3 highlights the details of the exercise program and exercise adherence achieved in the studies included in the review. Comparing exercise adherence among the different studies is difficult because of the variance in the exercise program. While most studies reported adherence in proportions of participants achieving adherence, 2 studies reported adherence in mean minutes of exercise achieved by the intervention group [35,39].

XSL•FO
RenderX

Table 3. Details of the exercise program and exercise adherence.

Study (author, year)	Exercise setting; individual or group	Modality	Intensity	Duration	Frequency	Length; attrition	Adherence calculation	Adherence achieved	Adherence sustainability report
Tsai et al (2022) [37]	1 time per week outpatient cardiac rehabilitation and home-based exercise telemonitoring	Cycle ergometers, stepper, elliptical, walking	RHR[a]+30 bpm or <13/20 RPE[b]	40-60 minutes	Not reported	6 months; n=4	Not reported	Intervention group: 95.2% adherence	None
Nagatomi et al (2022) [33]	Home only	Stretching and resistance training using weights and walking or cycle ergometry	11-13 RPE	Aerobic: 30-40 minutes	Aerobic=3-5 times per week; resistance=2-3 times per week	3 months	Percentage of participants who exercised 4 days per week; exercise performance rate=number of exercise days/total number of intervention days	Adherence was 73%; 11 participants achieved >4 days of exercise	None
Liu and Liu (2022) [34]	Home only	Walking, jogging	THR[c]=RHR+(exercise HRmax[d]–RHR)×0.4; Borg scale also used, target RPE not reported	30-60 minutes	3-5 times per week	12 weeks	80% for good adherence, 50%-79% for compliance, and below 50%	Intervention group: adherent 12, partial adherent 10, and poor adherent 8; control group: adherent 5, partial adherent 12; and poor adherent 13	None
Kikuchi et al (2021) [38]	Home only	Cycle ergometry	THR calculated from CPET[e]; specifics not provided	30 minutes or lower	3 times per week	12 weeks; n=1	Rate of attendance of a total of 36 offered sessions	94.4%	None
Deka et al (2019) [35]	Home only	Walking	RPE 11-14; heart rate during baseline walking	30 minutes	5 times per week	8 weeks; none	Self-reported exercise validated using Fitbit; ([actual number of minutes per week]/[150 minutes per week target goal]×100), adherent: >80%, partially adherent: 20%-80%, and nonadherent: <20%	Intervention group: 88 minutes per week; control group: 86 minutes per week	None
Lloyd et al (2019) [39]	Home only	Aerobic stepper	Not reported	Not reported	Daily	1 month	Self-reported number of minutes of exercise per day	Use of stepper was consistent; overall mean minutes of exercise increased by 2.4 minutes; exercise minutes ranged from 0 to 21 minutes	None

XSL•FO
RenderX

Study (author, year)	Exercise setting; individual or group	Modality	Intensity	Duration	Frequency	Length; attrition	Adherence calculation	Adherence achieved	Adherence sustainability report
Hwang et al (2017) [36]	Center-based for the control group and home-based for the intervention group; group exercise: 4 in each group	Aerobic and resistance	RPE 9-13	40 minutes of aerobic and resistance	2 times per week	12 weeks	Adherent: >80% of sessions attended, partly adherent: 20%-80% of sessions attended, and nonadherent: <20%	Intervention group: 49 adherent or partially adherent	After 12 weeks
Donesky et al (2017) [40]	Home only	Yoga	Not reported (mean heart rate stayed below 90 bpm)	35 minutes of poses	2 times per week	8 weeks	Attendance to number of sessions	90% attendance to classes	None

[a]RHR: resting heart rate.

[b]RPE: rating of perceived exertion.

[c]THR: target heart rate.

[d]HRmax: maximal heart rate.

[e]CPET: cardiopulmonary exercise testing.

In 2 studies, adherence was the primary outcome [35,40]. While one study investigated adherence to a walking program [35], the other study reported adherence to a teleyoga program [40]. The primary outcomes in the other studies included exercise intolerance [42], medication adherence [39], exercise tolerance [34,37], and physical function [33].

Similar to our previous review [26], we found that even in studies using mHealth technology, exercise adherence was not a primary outcome and mostly reported adherence as a secondary outcome measure. As previously noted, measurement of adherence requires a study design that extends beyond the intervention phase to determine the effectiveness of the intervention in bringing about lifestyle changes and sustainability of adherence behavior. We found this to be lacking in all studies in this review. Most studies included in this review were exploratory in nature and were testing for feasibility, acceptability, or initial pilot testing of the intervention design or exercise protocol.

Effect of demographic and clinical characteristics on Exercise Adherence

No study reported the effect of age, sex, race, HF etiology, NYHA functional classification, and HF etiology on exercise adherence. We found most studies included a sample of patients with HF who had homogenous demographic and clinical characteristics. In addition, the small sample sizes and quasi-experimental design prevent from making any conclusion on adherence based on age, sex, race, NYHA classification, and HF etiology.

Risk of Bias

Tables 4 and 5 detail the risk of bias assessments for the included studies. For the 4 RCTs evaluated using the Risk of Bias 2 tool, 1 was classified as having a low risk of bias, 1 as having a moderate risk, and 2 as having a high risk. The moderate and high risks of bias were attributed to bias from the randomization process (n=4, 50%), deviations from intended interventions (n=2, 25%), missing outcome data (n=2, 25%), and outcome measurement (n=2, 25%). Among the 4 quasi-experimental studies assessed using the Risk of Bias In Non-Randomized Studies of Interventions-I tool, 2 were rated as having a moderate risk of bias, and 2 had a serious risk of bias. The moderate and serious risks stemmed from confounding variables (n=2, 25%), classifications of interventions (n=2, 25%), deviations from intended interventions (n=2, 25%), missing data (n=6, 75%), and outcome measurement (n=4, 50%).

Table 4. Assessment of study risk of quasi-experimental papers.

Author(s) (year)	Confounding	Selection	Classifications	Deviations from interventions	Missing Data	Measurement	Selection of reported result	Overall bias
Donesky et al (2017) [40]	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk	Moderate risk
Lloyd et al (2019) [39]	Low risk	Low risk	Low risk	Low risk	Moderate risk	Serious risk	Low risk	Serious risk
Kikuchi et al (2021) [38]	Low risk	Low risk	Low risk	Serious concern	Moderate risk	Low risk	Low risk	Serious risk
Tsai et al (2022) [37]	Moderate risk	Low risk	Moderate risk	Low risk	Low risk	Low risk	Low risk	Moderate risk

XSL•FO
RenderX

Table 5. Assessment of study risk of randomized controlled trial papers.

Author(s) (year)	Arising from randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	Overall bias
Hwang et al (2017) [36]	High risk	Moderate concerns	Moderate risk	Low risk	Low risk	High risk
Deka et al (2019) [35]	Low risk	Low risk	Low risk	Moderate risk	Low risk	Moderate risk
Liu and Liu (2022) [34]	Low	Low risk	Low risk	Low risk	Low risk	Low risk
Nagatomi et al (2022) [33]	High risk	Low	Low risk	Low risk	Low risk	High risk

Discussion

Principal Findings

Decades of research have established the benefits of regular exercise in patients with HF [43]. However, adherence to recommended exercise for patients with HF is often hampered by their clinical condition and other structural and sociodemographic barriers [44]. With significant improvements in wireless internet connectivity and smartphone use across the world, mHealth-driven interventions hold immense promise with the ability to deliver the intervention to the participant's home using varied forms of multimedia engagement tools [45,46]. This review found 8 feasibility or pilot studies delivering exercise interventions using mHealth tools and reporting on adherence to varied exercise programs in patients with HF. Studies had a moderate to high risk of bias, and theoretically sound and fully powered RCTs were found to be lacking, making it difficult to determine clinical efficacy and the effect of demographic and clinical characteristics on adherence.

Previous studies have highlighted the variation in how the concept of adherence has been used by researchers reporting adherence to exercise in patients with HF [26]. Similar to prior studies, our review found the term "adherence" used primarily to report "attendance" to the investigator-developed program [23]. While this approach is good for assessing the effectiveness of the study intervention, researchers should clarify how adherence achieved in their study translates into meeting adherence to the recommended exercise guidelines for patients with HF. A position statement from the Heart Failure Association from the European Society of Cardiology has categorized exercise adherence as fully adherent=achieving >80% of exercise recommendations, partially adherent=achieving 20%-80% of exercise recommendations, and nonadherent=achieving <20% of exercise recommendations [44,47].

The earlier-mentioned adherence categorizations have also been borrowed from the medication literature [47]. It needs to be acknowledged that the complexity and challenges associated with adherence to exercise are different compared to adherence to other types of self-care (eg, medication adherence). To achieve exercise adherence, the extent of lifestyle and behavioral changes needed to be made by a patient with HF, experiencing frequent peaks and troughs in symptoms, can be challenging and also different from patients with other chronic clinical conditions such as diabetes and hypertension [25]. A systematic review and meta-analysis showed substantial heterogeneity in exercise adherence rates and dropouts across clinical populations [25]. This heterogeneity, along with the varied ways adherence is measured [48], makes it even more challenging to compare adherence rates within and across populations. Self-efficacy, motivation, enjoyment, and attention to exercise relapse management are important factors associated with physical activity in patients with HF [49-51]. Interventions should target not only building exercise self-efficacy but also ensuring that participants are enjoying the mHealth experience and the exercise program for maintaining motivation and achieving sustained adherence over time. Adults with HF often experience exacerbation of their symptoms that impede their ability to exercise. Relapse management strategies to reinitiate exercise after a break are an important component of interventions to promote exercise adherence.

As expected, all studies delivered their mHealth interventions to the participants' homes. The more recent studies adopted more smartphone-based interventions. This observation makes sense as, in the past few years, internet and smartphone access and use have increased significantly globally [52] and among patients with HF [53]. Most exercise prescriptions used in the studies in this review were investigator-developed programs. Alternate forms of exercise (such as yoga) have been delivered using mHealth technology; however, guidelines for such forms of alternate exercise for patients with HF have not yet been standardized, and their benefits have also not been clearly established. Due to the varied exercise programs prescribed in the different studies, it is difficult to compare adherence achieved between the studies. As such, it is also difficult to compare study results, as exercise and health outcomes have a dose-dependent relationship [54,55]. mHealth intervention can be challenging, as it requires training participants to use the mHealth technology in addition to the exercise program. The complexity of the technology can make it difficult for patients with HF, who tend to be older adults, to not be able to use the technology as directed, thereby presenting issues with internal validity for the research study and also potential dropout. Six studies mention providing 1 baseline training session with additional technical help if needed over the phone or in person [34-36,38-40]. No study provided details of challenges encountered in training participants to use the technologies that were used. The studies included in this review did not report

XSL•FO
RenderX

any safety concerns or adverse incidents associated with the exercise performed in a home setting. Exercise training at baseline, depending on the simplicity of the exercise program, ranged from 1 training session to prior enrollment in a cardiac rehabilitation program. As studies using mHealth technology are likely to be delivered to the participants' homes and mostly unsupervised, assessment of the ability of patients with HF to safely exercise on their own is important. While a majority of studies reported achieving good exercise adherence, the intervention period was short (lasting 1-6 months), and the sustainability of exercise behavior was not measured. We also noted a lack of and a need to determine the utility and effectiveness of theoretically sound mHealth-driven interventions across different strata of patients with HF, including age, sex, and race.

In our previous review, we highlighted that exercise diaries were the most commonly used tools for measuring exercise adherence in HF studies reporting adherence [26]. One advantage of using mHealth technology is also the ability to collect objective exercise data rather than relying on self-reported data, which can be erroneous due to overestimation [56]. The use of physical activity monitoring devices such as Fitbits allows for measuring and remotely monitoring objective physiological parameters, such as heart rate and step count, during exercise. Most studies included in this review prescribed simple exercise programs that included walking, cycle ergometer, or stepper, with a couple of studies adding resistance training to their exercise program. The use of RPE (eg, Borg scale) was the most used tool for regulating exercise intensity. Subjective tools are easy to use and are often preferred by clinicians and patients. However, the heart rate function in commercially available physical activity monitors such as Fitbit offers more precise exercise intensity regulation and was used by some studies. Research shows that patients with HF like using the heart rate and step count feature in Fitbit [35]. It is important that patients with HF, who tend to be older adults, are adequately trained in using these devices to prevent potential dropouts from facing technological challenges.

Limitations

This review included studies that were published between 2015 and 2022. We completed a thorough search, but some papers may have been missed and not included in the review. The study is limited to the use of mHealth for exercise adherence in patients with HF and does not include the use of mHealth technologies for improving other self-care recommendations. There is a risk of bias on the part of the authors, which we have tried to mitigate through our search strategies, inclusion, exclusion, and screening process outlined earlier.

Future Recommendations

Our review shows preliminary evidence of using mHealth technology for targeting exercise adherence in patients with HF. However, there is a lack of theoretically sound and fully powered RCTs to suggest its effectiveness. It is crucial for studies using mHealth technology to improve study quality and limit the risk of bias, as noted in this review, particularly by limiting confounding factors. In addition, it is important to determine the sustainability of exercise behavior after the intervention period when the intervention stimuli have been removed. How adherence is measured and reported also needs consideration. Because adherence has mostly been reported to investigator-developed programs, it is difficult to compare the adherence achieved in different studies. As such, we suggest that adherence achieved to investigator-developed exercise programs should be standardized to the recommendations for exercise for patients with HF to help compare the effectiveness of different programs. Reporting the proportion of participants achieving adherence may be better than reporting the mean minutes of exercise performed by participants, as the mean calculations are influenced by outliers. Tracking daily exercise using mHealth tools and reporting changes in the number of minutes of exercise from baseline can provide a better picture of the effectiveness of home-based interventions, especially in patients with HF who experience peaks and troughs in symptoms. They can also be useful in validating self-reported exercise and physical activity [57]. Considering that making lifestyle changes is challenging for many patients with HF, the practical and clinical significance of the adherence achieved using mHealth interventions should also be highlighted alongside reports of statistical significance. Finally, the feasibility and acceptability of mHealth technology across different races, sexes, and cultures needs to be studied. mHealth provides opportunities to explore culturally tailored interventions in HF that have been grossly underinvestigated.

Conclusions

There is some preliminary evidence suggesting the feasibility of using mHealth technology for building exercise adherence in patients with HF; however, theoretically sound and fully powered studies, including studies on minoritized communities, are lacking. In addition, lacking is a report on the sustainability of the achieved adherence beyond the intervention period. The study provides some areas for researchers to focus on in future studies.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-S Checklist for Reporting Literature Searches in Systematic Reviews.

XSL•FO
RenderX

[DOCX File , 19 KB - jmir_v27i1e54524_app1.docx]

References

1. Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, et al. Heart failure: preventing disease and death worldwide. ESC Heart Failure 2014;1(1):4-25. [doi: 10.1002/2055-5822.12005]
2. Benjamin E, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. Circulation 2019;139(10):e56-e528 [FREE Full text] [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
3. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
4. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2022;145(18):e895-e1032 [FREE Full text] [doi: 10.1161/CIR.0000000000001063] [Medline: 35363499]
5. Conraads VM, Van Craenenbroeck EM, De Maeyer C, Van Berendoncks AM, Beckers PJ, Vrints CJ. Unraveling new mechanisms of exercise intolerance in chronic heart failure: role of exercise training. Heart Fail Rev 2013;18(1):65-77. [doi: 10.1007/s10741-012-9324-0] [Medline: 22684340]
6. Bloom MW, Greenberg B, Jaarsma T, Januzzi JL, Lam CSP, Maggioni AP, et al. Heart failure with reduced ejection fraction. Nat Rev Dis Primers 2017;3(1):17058. [doi: 10.1038/nrdp.2017.58] [Medline: 28836616]
7. Coats AJ, Clark AL, Piepoli M, Volterrani M, Poole-Wilson PA. Symptoms and quality of life in heart failure: the muscle hypothesis. Br Heart J 1994;72(2 Suppl):S36-S39 [FREE Full text] [doi: 10.1136/hrt.72.2_suppl.s36] [Medline: 7946756]
8. Ramasamy R, Hildebrandt T, O'Hea E, Patel M, Clemow L, Freudenberger R, et al. Psychological and social factors that correlate with dyspnea in heart failure. Psychosomatics 2006;47(5):430-434. [doi: 10.1176/appi.psy.47.5.430] [Medline: 16959932]
9. Drexler H, Coats AJ. Explaining fatigue in congestive heart failure. Annu Rev Med 1996;47(1):241-256. [doi: 10.1146/annurev.med.47.1.241] [Medline: 8712779]
10. Kitzman DW, Groban L. Exercise intolerance. Cardiol Clin 2011;29(3):461-477. [doi: 10.1016/j.ccl.2011.06.002] [Medline: 21803233]
11. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart 2007;93(9):1137-1146. [doi: 10.1136/hrt.2003.025270] [Medline: 17699180]
12. Elixhauser A, Steiner C. Readmissions to U.S. hospitals by diagnosis, 2010: statistical brief. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville, MD: Agency for Healthcare Research and Quality; 2010.
13. Schmidt C, Moreira-Gonçalves D, Santos M, Leite-Moreira A, Oliveira J. Physical activity and exercise training in heart failure with preserved ejection fraction: gathering evidence from clinical and pre-clinical studies. Heart Fail Rev 2022;27(2):573-586. [doi: 10.1007/s10741-020-09973-5] [Medline: 32385565]
14. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131. [Medline: 3920711]
15. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. JAMA 2009;301(14):1439-1450. [doi: 10.1001/jama.2009.454] [Medline: 19351941]
16. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37(27):2129-2200 [FREE Full text] [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
17. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
18. Taylor RS, Long L, Mordi IR, Madsen MT, Davies EJ, Dalal H, et al. Exercise-based rehabilitation for heart failure: Cochrane systematic review, meta-analysis, and trial sequential analysis. JACC Heart Fail 2019;7(8):691-705 [FREE Full text] [doi: 10.1016/j.jchf.2019.04.023] [Medline: 31302050]
19. Heart Failure Society of America, Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, et al. HFSA 2010 comprehensive heart failure practice guideline. J Card Fail 2010;16(6):e1-194. [doi: 10.1016/j.cardfail.2010.04.004] [Medline: 20610207]
20. Pozehl BJ, McGuire R, Duncan K, Kupzyk K, Norman J, Artinian NT, et al. Effects of the HEART camp trial on adherence to exercise in patients with heart failure. J Card Fail 2018;24(10):654-660 [FREE Full text] [doi: 10.1016/j.cardfail.2018.06.007] [Medline: 30010027]

XSL•FO
RenderX

21. Alonso WW, Keteyian SJ, Leifer ES, Kitzman DW, Sachdev V. Adherence to exercise in adults with heart failure. J Cardiopulm Rehabil Prev 2024;44(4):E17-E18. [doi: 10.1097/HCR.0000000000000892] [Medline: 38870066]
22. Adherence to long-term therapies: evidence for action. World Health Organization. 2003. URL: https://iris.who.int/bitstream/handle/10665/42682/9241545992.pdf [accessed 2024-11-25]
23. Bissonnette JM. Adherence: a concept analysis. J Adv Nurs 2008;63(6):634-643. [doi: 10.1111/j.1365-2648.2008.04745.x] [Medline: 18808585]
24. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR mHealth uHealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
25. Bullard T, Ji M, An R, Trinh L, Mackenzie M, Mullen SP. A systematic review and meta-analysis of adherence to physical activity interventions among three chronic conditions: cancer, cardiovascular disease, and diabetes. BMC Public Health 2019;19(1):636 [FREE Full text] [doi: 10.1186/s12889-019-6877-z] [Medline: 31126260]
26. Deka P, Pozehl B, Williams MA, Yates B. Adherence to recommended exercise guidelines in patients with heart failure. Heart Fail Rev 2017;22(1):41-53. [doi: 10.1007/s10741-016-9584-1] [Medline: 27671166]
27. Barbour KA, Miller NH. Adherence to exercise training in heart failure: a review. Heart Fail Rev 2008;13(1):81-89. [doi: 10.1007/s10741-007-9054-x] [Medline: 18008161]
28. Mobile technology and home broadband 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2024-11-21]
29. Tirado-Morueta R, Aguaded-Gómez JI, Hernando-Gómez A. The socio-demographic divide in internet usage moderated by digital literacy support. Technol Soc 2018;55:47-55. [doi: 10.1016/j.techsoc.2018.06.001]
30. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
31. Sterne JA, Hernán M, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
32. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
33. Nagatomi Y, Ide T, Higuchi T, Nezu T, Fujino T, Tohyama T, et al. Home-based cardiac rehabilitation using information and communication technology for heart failure patients with frailty. ESC Heart Fail 2022;9(4):2407-2418 [FREE Full text] [doi: 10.1002/ehf2.13934] [Medline: 35534907]
34. Liu T, Liu M. Effect of mobile internet technology in health management of heart failure patients guiding cardiac rehabilitation. J Healthc Eng 2022;2022:7118919 [FREE Full text] [doi: 10.1155/2022/7118919] [Medline: 35251573]
35. Deka P, Pozehl B, Williams MA, Norman JF, Khazanchi D, Pathak D. MOVE-HF: an internet-based pilot study to improve adherence to exercise in patients with heart failure. Eur J Cardiovasc Nurs 2019;18(2):122-131. [doi: 10.1177/1474515118796613] [Medline: 30129790]
36. Hwang R, Bruning J, Morris NR, Mandrusiak A, Russell T. Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. J Physiother 2017;63(2):101-107 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.017] [Medline: 28336297]
37. Tsai WJ, Wen YK, Cheng YY, Huang JL, Chen YW. Effectiveness of home-based telerehabilitation programs on functional capacity and cardiac function in elderly heart failure patients: a prospective longitudinal study. Medicine (Baltimore) 2022;101(28):e29799 [FREE Full text] [doi: 10.1097/MD.0000000000029799] [Medline: 35838996]
38. Kikuchi A, Taniguchi T, Nakamoto K, Sera F, Ohtani T, Yamada T, et al. Feasibility of home-based cardiac rehabilitation using an integrated telerehabilitation platform in elderly patients with heart failure: a pilot study. J Cardiol 2021;78(1):66-71 [FREE Full text] [doi: 10.1016/j.jjcc.2021.01.010] [Medline: 33579602]
39. Lloyd T, Buck H, Foy A, Black S, Pinter A, Pogash R, et al. The Penn State Heart Assistant: a pilot study of a web-based intervention to improve self-care of heart failure patients. Health Informatics J 2019;25(2):292-303 [FREE Full text] [doi: 10.1177/1460458217704247] [Medline: 28504048]
40. Donesky D, Selman L, McDermott K, Citron T, Howie-Esquivel J. Evaluation of the feasibility of a home-based TeleYoga intervention in participants with both chronic obstructive pulmonary disease and heart failure. J Altern Complement Med 2017;23(9):713-721. [doi: 10.1089/acm.2015.0279] [Medline: 28654302]
41. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92-98. [Medline: 5523831]
42. Hwang CL, Yoo JK, Kim HK, Hwang MH, Handberg EM, Petersen JW, et al. Novel all-extremity high-intensity interval training improves aerobic fitness, cardiac function and insulin resistance in healthy older adults. Exp Gerontol 2016;82:112-119. [doi: 10.1016/j.exger.2016.06.009] [Medline: 27346646]
43. Flynn KE, Piña IL, Whellan DJ, Lin L, Blumenthal JA, Ellis SJ, et al. Effects of exercise training on health status in patients with chronic heart failure: HF-ACTION randomized controlled trial. JAMA 2009;301(14):1451-1459. [doi: 10.1001/jama.2009.457] [Medline: 19351942]

XSL•FO
RenderX

44. Conraads VM, Deaton C, Piotrowicz E, Santaularia N, Tierney S, Piepoli MF, et al. Adherence of heart failure patients to exercise: barriers and possible solutions: a position statement of the Study Group on Exercise Training in Heart Failure of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2012;14(5):451-458 [FREE Full text] [doi: 10.1093/eurjhf/hfs048] [Medline: 22499542]
45. Ezhilarasan E, Dinakaran M. A review on mobile technologies: 3G, 4G and 5G. 2017 Presented at: 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM); February 3-4, 2017; Tindivanam, India. [doi: 10.1109/icrtccm.2017.90]
46. Li D. 5G and intelligence medicine—how the next generation of wireless technology will reconstruct healthcare? Precis Clin Med 2019;2(4):205-208 [FREE Full text] [doi: 10.1093/pcmedi/pbz020] [Medline: 31886033]
47. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med 1980;303(18):1038-1041. [doi: 10.1056/NEJM198010303031804] [Medline: 6999345]
48. Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. Med Pharm Rep 2019;92(2):117-122 [FREE Full text] [doi: 10.15386/mpr-1201] [Medline: 31086837]
49. Klompstra L, Jaarsma T, Strömberg A. Self-efficacy mediates the relationship between motivation and physical activity in patients with heart failure. J Cardiovasc Nurs 2018;33(3):211-216 [FREE Full text] [doi: 10.1097/JCN.0000000000000456] [Medline: 29189427]
50. Klompstra L, Deka P, Almenar L, Pathak D, Muñoz-Gómez E, López-Vilella R, et al. Physical activity enjoyment, exercise motivation, and physical activity in patients with heart failure: a mediation analysis. Clin Rehabil 2022;36(10):1324-1331 [FREE Full text] [doi: 10.1177/02692155221103696] [Medline: 35678610]
51. Alonso WW, Kupzyk K, Norman J, Bills SE, Bosak K, Dunn SL, et al. Negative attitudes, self-efficacy, and relapse management mediate long-term adherence to exercise in patients with heart failure. Ann Behav Med 2021;55(10):1031-1041 [FREE Full text] [doi: 10.1093/abm/kaab002] [Medline: 33580663]
52. Number of internet and social media users worldwide as of October 2024. Statista. 2023. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2024-11-21]
53. Deka P, Almenar L, Pathak D, Muñoz-Gómez E, Orihuela-Cerdeña L, López-Vilella R, et al. Technology usage, physical activity, and motivation in patients with heart failure and heart transplantation. Comput Inform Nurs 2023;41(11):903-908. [doi: 10.1097/CIN.0000000000001049] [Medline: 37556811]
54. Buford TW, Roberts MD, Church TS. Toward exercise as personalized medicine. Sports Med 2013;43(3):157-165. [doi: 10.1007/s40279-013-0018-0] [Medline: 23382011]
55. Keteyian SJ, Leifer ES, Houston-Miller N, Kraus WE, Brawner CA, O'Connor CM, et al. Relation between volume of exercise and clinical outcomes in patients with heart failure. J Am Coll Cardiol 2012;60(19):1899-1905 [FREE Full text] [doi: 10.1016/j.jacc.2012.08.958] [Medline: 23062530]
56. Yates BC, Pozehl B, Kupzyk K, Epstein CM, Deka P. Are heart failure and coronary artery bypass surgery patients meeting physical activity guidelines? Rehabil Nurs 2017;42(3):119-124 [FREE Full text] [doi: 10.1002/rnj.257] [Medline: 29203953]
57. Deka P, Pozehl B, Norman JF, Khazanchi D. Feasibility of using the Fitbit Charge HR in validating self-reported exercise diaries in a community setting in patients with heart failure. Eur J Cardiovasc Nurs 2018;17(7):605-611. [doi: 10.1177/1474515118766037] [Medline: 29546995]

Abbreviations

HF: heart failure
mHealth: mobile health
NYHA: New York Heart Association
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
RCT: randomized controlled trial
RPE: rating of perceived exertion

Edited by A Mavragani; submitted 13.11.23; peer-reviewed by S Mishra, K Biswas, J Knoop; comments to author 02.08.24; revised version received 09.09.24; accepted 05.11.24; published 09.01.25.

Please cite as:
Deka P, Salahshurian E, Ng T, Buchholz SW, Klompstra L, Alonso W
Use of mHealth Technology for Improving Exercise Adherence in Patients With Heart Failure: Systematic Review
J Med Internet Res 2025;27:e54524
URL: https://www.jmir.org/2025/1/e54524
doi:10.2196/54524
PMID:

XSL•FO
RenderX

©Pallav Deka, Erin Salahshurian, Teresa Ng, Susan W Buchholz, Leonie Klompstra, Windy Alonso. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing the Effectiveness of Interactive Robot-Assisted Virtual Health Coaching for Health Literacy and Disease Knowledge of Patients with Chronic Kidney Disease: Quasiexperimental Study

Nai-Jung Chen[1,2*], MS; Ching-Hao Chang[1*], MEd; Chiu-Mieh Huang[3], PhD; Fen-He Lin[4], MSN; Li-Ting Lu[5], PhD; Kuan-Yi Liu[1], MEd; Chih-Lin Lai[6], AN; Chin-Yao Lin[7], MD; Yi-Chou Hou[8,9], PhD; Jong-Long Guo[1], PhD

[1]Department of Health Promotion and Health Education, College of Education, National Taiwan Normal University, Taipei, Taiwan

[2]Department of Nursing, Taiwan Adventist Hospital, Taipei, Taiwan

[3]College of Nursing, Institute of Clinical Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

[4]School of Nursing, College of Medicine, National Taiwan University, Taipei, Taiwan

[5]Department of Nursing, University of Kang Ning, Taipei, Taiwan

[6]Nephrology and Hemodialysis Center, Cardinal Tien Hospital, Taipei, Taiwan

[7]Department of Internal Medicine, Taiwan Adventist Hospital, Taipei, Taiwan

[8]Department of Internal Medicine, Cardinal Tien Hospital, Taipei, Taiwan

[9]Department of Internal Medicine, Fu Jen Catholic University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Jong-Long Guo, PhD
Department of Health Promotion and Health Education
College of Education
National Taiwan Normal University
162, Section 1, Heping E Rd
Da'an District
Taipei, 10610
Taiwan
Phone: 886 7749 1728
Fax: 886 2363 0326
Email: jonglong@ntnu.edu.tw

Abstract

Background: Chronic kidney disease (CKD) imposes a significant global health and economic burden, impacting millions globally. Despite its high prevalence, public awareness and understanding of CKD remain limited, leading to delayed diagnosis and suboptimal management. Traditional patient education methods, such as 1-on-1 verbal instruction or printed brochures, are often insufficient, especially considering the shortage of nursing staff. Technology-assisted education presents a promising and standardized solution, emphasizing the need for innovative and scalable approaches to improve CKD-specific knowledge and health literacy.

Objective: This study aimed to develop and evaluate the effectiveness of an innovative 12-unit virtual health coaching program delivered through interactive robots that is intended to enhance disease knowledge and health literacy among patients with CKD.

Methods: A quasiexperimental design was used, and 60 participants were evenly assigned to experimental and comparison groups. However, due to attrition, 14 participants in the experimental group and 16 participants in the comparison group completed the study. The intervention involved a 12-unit program, with each unit lasting approximately 20 minutes to 30 minutes and delivered across 3 to 4 learning sessions, and participants completed 3 to 4 units per session. The program addressed key aspects of CKD-specific health literacy including functional, communicative, and critical literacy and CKD-specific knowledge including basic knowledge, prevention, lifestyle, dietary intake, and medication. Data were collected through validated pre and postintervention questionnaires. All 30 participants completed the program and subsequent evaluations, with outcome measures assessing changes in CKD-specific knowledge and health literacy.

XSL•FO
RenderX

Results: Postintervention analysis using generalized estimating equations, adjusted for age, revealed that the experimental group (n=14) had significantly greater improvements in health literacy (coefficient=2.51, Wald χ^2_1=5.89; *P*=.02) and disease knowledge (coefficient=1.66, Wald χ^2_1=11.75; *P*=.001) than the comparison group (n=16). Postintervention *t* tests revealed significant improvements in CKD-specific health literacy and disease knowledge (*P*<.001) between the experimental and comparison groups. Additional analyses identified significant group × time interactions, indicating improvements in communicative literacy (*P*=.01) and critical literacy (*P*=.02), while no significant changes were observed in functional literacy. Regarding disease knowledge, the experimental group demonstrated a significant improvement in medication (*P*<.001), whereas changes in basic knowledge, prevention, lifestyle, and dietary intake were not significant.

Conclusions: This study demonstrated that interactive robot-assisted eHealth coaching effectively enhanced CKD-specific disease knowledge and health literacy. Despite the challenges posed by the COVID-19 pandemic, which constrained sample sizes, the findings indicate that this program is a promising patient education tool in clinical nephrology. Future research should involve larger sample sizes to enhance generalizability and examine additional factors influencing effectiveness.

(J Med Internet Res 2025;27:e68072) doi:10.2196/68072

KEYWORDS

chronic kidney disease; disease knowledge; eHealth; health coaching; health education; health literacy; interactive robot

Introduction

Chronic kidney disease (CKD) has emerged as a major global health issue, with current estimates indicating that it affects approximately 10% of the global population [1]. The disease imposes a significant financial burden driven by the high costs of diagnosis, treatment, and ongoing care. As CKD advances, over two-thirds of patients with CKD experience complications such as cardiovascular disease, metabolic bone disorders, anemia, and malnutrition, all of which can profoundly affect their health and quality of life [2]. In Taiwan, CKD accounts for 7.14% of the nation's total health care expenditure [3], underscoring the critical need for effective prevention strategies. The prevalence of dialysis among patients with CKD in Taiwan underscores the need for targeted interventions. In 2021, only 3.7% of dialysis patients were aged younger than 40 years, while 40.1% were aged 40 years to 64 years, 31.1% were aged 65 years to 74 years, and 25.1% were 75 years old or older. This distribution highlights that the majority of patients with CKD in Taiwan are middle-aged or older.

Health literacy (HL) is defined as the ability to access, understand, evaluate, and apply health information to make informed health decisions [4]. In this study, HL refers to a patient's capacity to comprehend and use CKD management information to manage their CKD [5]. According to Nutbeam [6], HL consists of 3 components: functional, communicative and interactive, and critical literacy. Functional literacy encompasses the basic reading and writing skills needed for everyday functioning. Communicative literacy includes more advanced skills that enable individuals to actively engage in social contexts, extract information, and adapt to changing environments. Critical literacy involves using advanced cognitive and social skills to critically analyze information and exert greater control over personal health and life situations. Limited HL in patients with CKD has been linked to poor health outcomes and reduced treatment adherence [7]. CKD-specific HL and knowledge can help patients manage their condition effectively. In this study, disease-specific knowledge encompassed an understanding of kidney function, disease prevention, healthy lifestyle habits, diet, and medication. Higher HL levels have been identified as strong predictors of adherence to medication, diet, fluid intake, and general self-management practices [8]. A previous study suggested that complications related to kidney disease may stem from insufficient health knowledge and that HL significantly influences self-care behaviors [9]. One review highlighted that disease-specific knowledge is critical for HL, which is a prerequisite for the effective self-management of CKD [8]. Improving HL can enhance disease knowledge, promote better self-care behaviors, and potentially slow disease progression. These findings underscore the importance of HL and knowledge of kidney disease management.

Health coaching is an evidence-based approach that enables patients to make lasting behavioral changes that enhance their overall health [10]. By offering health care options tailored to individuals' lifestyles and abilities, health coaching improves outcomes, especially among those with chronic illnesses or those at a higher risk of medication nonadherence. Previous studies have shown that health coaching can lead to positive behavioral changes and improved health outcomes in patients with chronic conditions, underscoring its value in managing long-term health [11,12].

Previous studies have demonstrated the significant impact of health promotion and education programs on patients with CKD [13]. Recent digital health intervention advancements, such as wearable devices, virtual reality, and mobile health apps, have introduced innovative approaches to enhance patient education. Wearable devices facilitate self-management by providing personalized information [14], virtual reality enhances CKD education through experiential learning [15], and mobile apps can improve disease knowledge when compared with traditional education alone [16]. Digital technologies for health coaching offer older adults support for improving various health determinants [17], and approaches such as e-learning and remote monitoring have been shown to boost treatment effectiveness and quality of life [18]. Moreover, research has also underscored the need for digital coaching to be user-friendly, highly engaging, and tailored to individual needs [19,20]. Health care professionals face challenges due to insufficient nursing staff, and CKD is a complex condition requiring comprehensive

XSL•FO
RenderX

education. Most education systems rely on traditional methods (eg, brochures or 1-on-1 verbal counseling) that often lack diverse directions and interactions. This gap highlights the urgent need for innovative digital solutions to enhance CKD-specific HL and disease knowledge. Using emerging technologies (eg, interactive robots) that follow official health education guidelines [21] offers a novel solution to standardize education, address staffing shortages, and improve the overall interactivity of CKD education. These technologies can provide personalized and accessible learning experiences, as well as improve patient engagement and outcomes [22].

This study aimed to evaluate the effectiveness of an interactive robot-assisted virtual health coaching intervention for improving CKD-specific HL and disease knowledge. The following research questions guided the study: (1) How comprehensively does this intervention improve CKD-specific knowledge? (2) How effective is the intervention in enhancing CKD patients' HL? Prototypes were developed using insights from nephrologists and nurse practitioners who served as virtual health coaches and incorporated their clinical interactions with patients. By offering an engaging and accessible educational platform, the intervention aimed to enhance patient engagement and ultimately improve the health outcomes of patients with CKD.

Methods

Study Design

This study adopted a quasiexperimental design involving the recruitment of patients with CKD from the nephrology departments of 2 regional hospitals in northern Taiwan that are similar in size and scope. Each hospital was assigned to either an experimental or comparison group. Health care professionals initially screened patients based on the eligibility criteria and referred interested parties to the research team. This process continued until 30 participants were selected from each hospital. However, due to the COVID-19 pandemic, only 14 participants in the experimental group and 16 participants in the comparison group completed the study. Participant attrition was primarily attributed to the lengthy intervention period and pandemic-related disruptions. Participants in the experimental group received both intervention and usual care provided by nurse practitioners, whereas those in the comparison group received only usual care. During recruitment, the researchers explained the use of the interactive robot and the basic interface operations required. After obtaining written informed consent, the participants were provided with an operation manual and guidance regarding the use of the interface. The flowchart of this study is presented in Figure 1.

Figure 1. Flowchart of the study. CKD: chronic kidney disease.

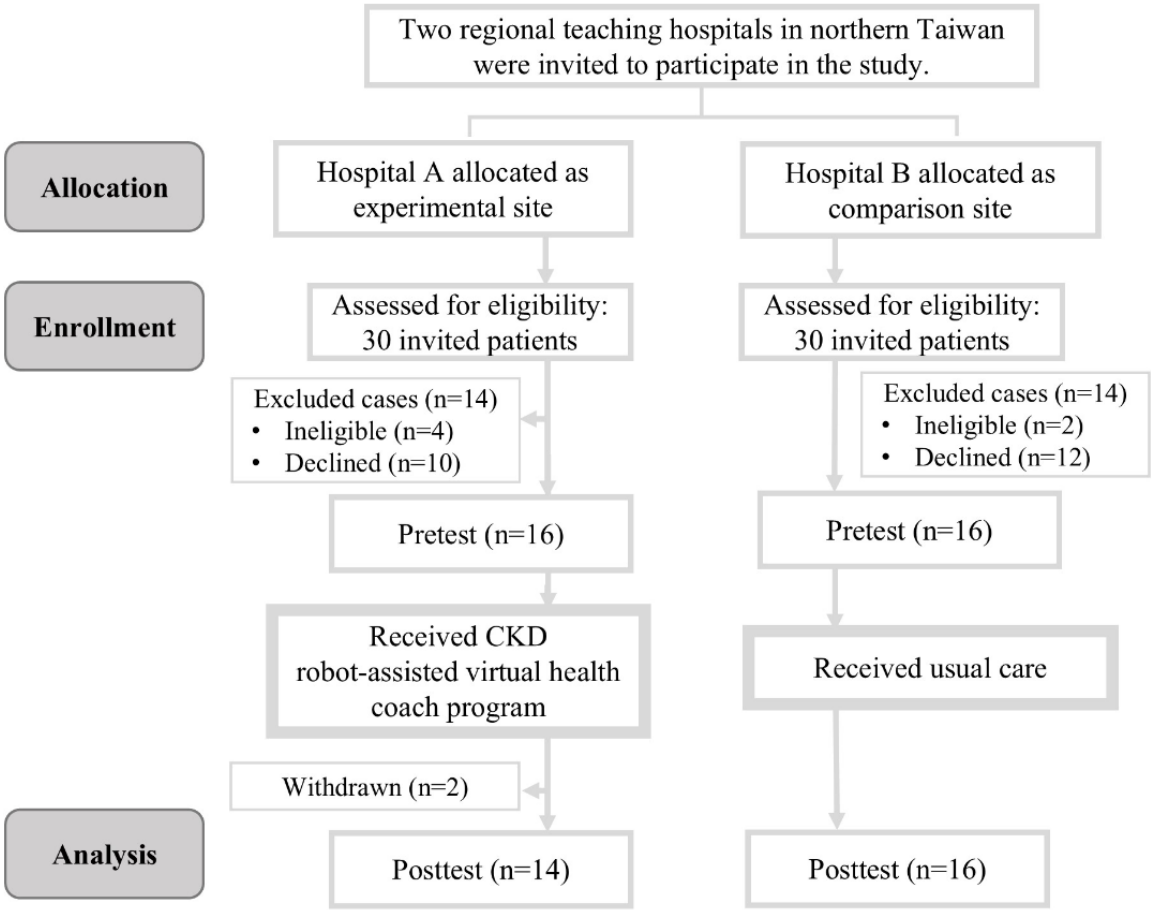

XSL•FO
RenderX

Study Population

The inclusion criteria required participants to be adults aged 20 years or older with a diagnosis of CKD. The exclusion criteria included dialysis, transplantation, severe mental illness, dementia, or significant sensory impairments. Patients undergoing dialysis were excluded because of significant differences in care standards and intensity compared with CKD management [23].

Intervention

An interdisciplinary team including 2 nephrologists and nurse practitioners developed a 12-unit health coaching program using the interactive robot Kebbi Air S (NUWA Robotics Corp). The Kebbi robot, known for its versatility and interactivity, has been used across various domains, including education, health care, and customer service [24]. Prior studies have investigated its effectiveness for language learning and social assistance in health-related contexts and the impact of its expressions on participant engagement [25-27]. The program was designed using government-developed guidelines to ensure standardization and validity [21] after incorporating clinical experiences, patient interactions, and insights from 14 development meetings. These physicians and nurse practitioners acted as virtual health coaches through the robot, delivering content through dialogue, animations, and interactive tasks using artificial intelligence and natural language processing to facilitate dynamic patient engagement. The intervention was conducted using participant-staff 1-on-1 sessions during which staff guided participants in familiarizing themselves with the interface and provided assistance as needed. The robot was programmed to emulate health care professionals' tone and demeanor, offering positive reinforcement (praise, encouragement, and motivational gestures) to sustain patient engagement. Correct answers on standardized quizzes were rewarded with applause, while supportive hints and explanations were provided after incorrect responses to facilitate learning. Virtual health coaches presented information tailored to the 5 stages of CKD, responded to patient inputs, and provided feedback. The program covered critical areas of CKD-specific HL, including functional, communicative, and critical literacy as well as CKD-specific knowledge, including basic knowledge, prevention, lifestyle, dietary intake, and medication. Table 1 provides an overview of the course learning objectives and corresponding outcome variables. Figure 2 illustrates an example of the program's interface.

Table 1. Learning objectives and outcome variables of the intervention program.

Unit	Learning objectives	Outcome variables
1	Understand the importance of CKD[a] self-management Know how to measure waist circumference	HL[b]: functional literacy Knowledge: lifestyle
2	Recognize CKD clinical symptoms Identify 5 warning signs of CKD	HL: critical literacy Knowledge: basic knowledge, prevention
3	Understand the importance of regular tracking Know 3 basic types of renal tests	HL: functional literacy Knowledge: prevention
4	Understand the 5 stages of CKD care Know how to calculate eGFR[c] Learn the self-test for renal function	HL: functional literacy Knowledge: lifestyle, prevention
5, 6, 7	Understand nutrition for CKD Identify high-potassium fruits and cooking methods	HL: communicative, functional literacy Knowledge: dietary intake
8	Understand the management of comorbidities of CKD Learn about health care for CKD comorbidities	HL: critical literacy Knowledge: basic knowledge, medication
9	Understand symptoms and self-management of chronic renal failure Learn about psychological adaptations to CKD	HL: communicative literacy Knowledge: basic knowledge, medication
10	Understand the proper time for dialysis Learn the patient-physician shared decision-making strategies	HL: communicative literacy Knowledge: prevention, medication
11	Understand hemodialysis access methods Know arteriovenous fistulas and grafts	HL: critical literacy Knowledge: medication
12	Identify the emergency medical primetime[d] Learn how to rebuild life after CKD	HL: critical literacy Knowledge: lifestyle, medication

[a]CKD: chronic kidney disease.

[b]HL: health literacy.

[c]eGFR: estimated glomerular filtration rate.

[d]The "Golden Hour" or critical period of time between the time a medical service (such as an ambulance or emergency service) receives a call for help and the arrival of the first emergency team at the scene of an emergency.

XSL•FO
RenderX

Figure 2. Interface example of the interactive robot-assisted virtual health coach program for chronic kidney disease (CKD). RN: registered nurse.

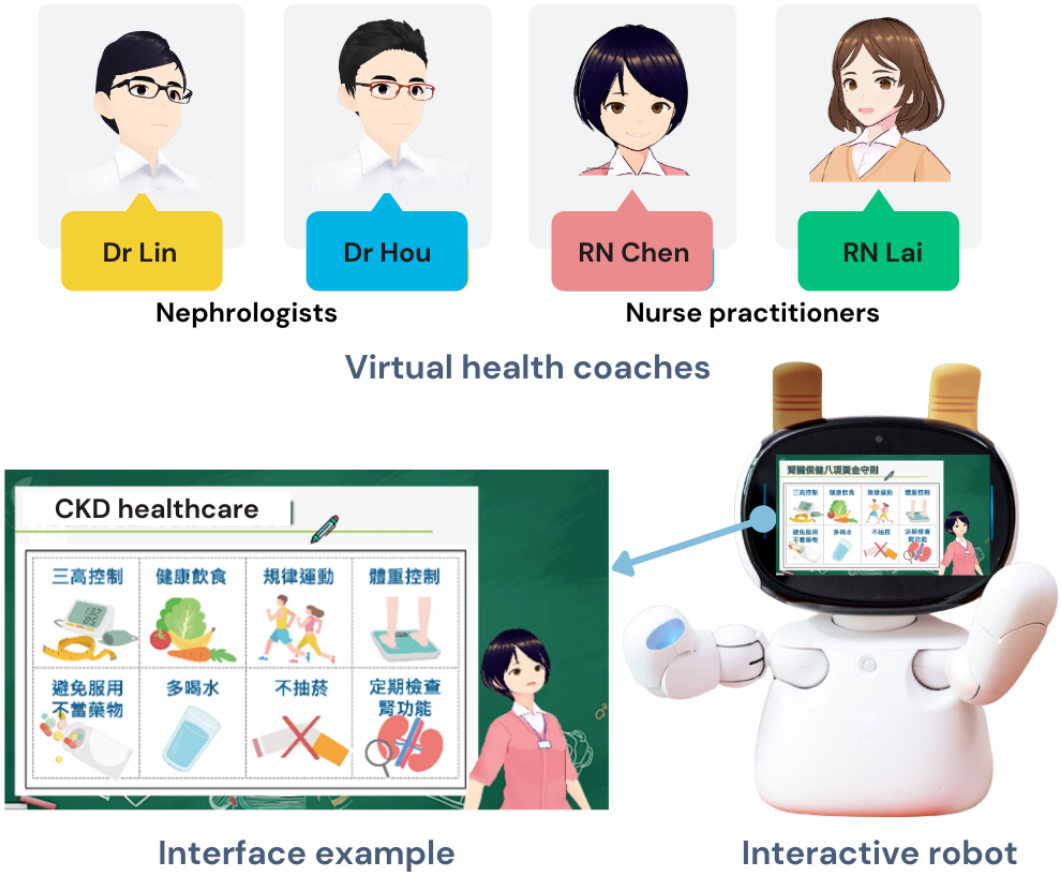

Procedure and Program Delivery

The research team selected an experimental hospital and collaborated with attending nephrologists to define the goals, procedures, and protocols of the study. Upon receiving the hospitals' approval to conduct the study, recruitment posters were distributed to encourage patients with CKD to participate, while an orientation session introduced participants to the study's objectives and collaborative framework. Data collection involved 1-on-1 interviews, enabling in-depth interactions and ensuring that participants were fully engaged with the intervention. Patients were encouraged to interact with the robots to become familiar with the interface. The experimental group participated in a 12-unit intervention, with each unit lasting approximately 20 minutes to 30 minutes. Participants completed 3 to 4 units per session across 3 to 4 learning sessions. The comparison group received usual care, including personalized education, from nurse practitioners. Structured questionnaires were used for the pre and postintervention assessments.

Measurements

The effectiveness of the intervention was assessed using a validated questionnaire [5,28]. The CKD-specific HL scale and CKD-specific disease knowledge scale were adapted from validated tools specifically designed to assess HL and disease knowledge in patients with CKD. The CKD-specific HL scale includes 17 items for functional, communicative, and critical literacy in a multiple-choice test format. The CKD-specific disease knowledge scale has 13 true or false items across 5 domains. Higher scores indicated better HL and knowledge. The Cronbach coefficients were 0.78 and 0.76 for the CKD-specific HL scale and the CKD-specific disease knowledge scale, respectively.

Data Analysis

Data were analyzed using SPSS version 23.0 (IBM Corp). This study primarily aimed to evaluate the intervention's effectiveness at enhancing CKD-specific HL and disease knowledge. No significant differences were found after comparing the pretest scores of the outcome variables between the experimental and comparison groups using *t* tests, eliminating the need to control for the pretest scores. However, due to missing data for some variables, generalized estimating equations (GEEs) were used to analyze the pre and posttest results between the 2 groups while also controlling for age differences. The GEE approach was selected because it effectively handles repeated measures and correlated data, making it well-suited for our study design. It is robust and accommodates missing values, providing reliable estimates and enhancing the validity of our findings.

Ethical Considerations

This study was approved by the Joint Institutional Review Board of Shin Kong Wu Ho-Su Memorial Hospital (approval number

XSL•FO
RenderX

20211202R) and Cardinal Tien Hospital (approval number CTH-110-3-5-002). All participants were fully informed and provided their consent before the interviews began.

Results

User Demographics

The demographic characteristics of participants in the experimental (n=14) and comparison (n=16) groups were not significantly different in gender (χ^2_{28}=0.27; *P*=.61), marital status (χ^2_{28}=0.05; *P*=.82), and educational level (χ^2_{28}=0.20; *P*=.65), although the experimental group was significantly younger.

Evaluation Outcomes

As shown in Figure 3, the results of the *t* tests indicated that the posttest scores for CKD-specific HL (t_{27}=–4.44; *P*<.001) and disease knowledge (t_{28}=5.58; *P*<.001) were significantly higher in the experimental group than in the comparison group.

Figure 3. Comparison of pre and posttest (A) CKD-specific health literacy and (B) CKD-specific disease knowledge, reported as the mean (95% CI) scores, with comparisons of pre and postintervention scores between the groups.

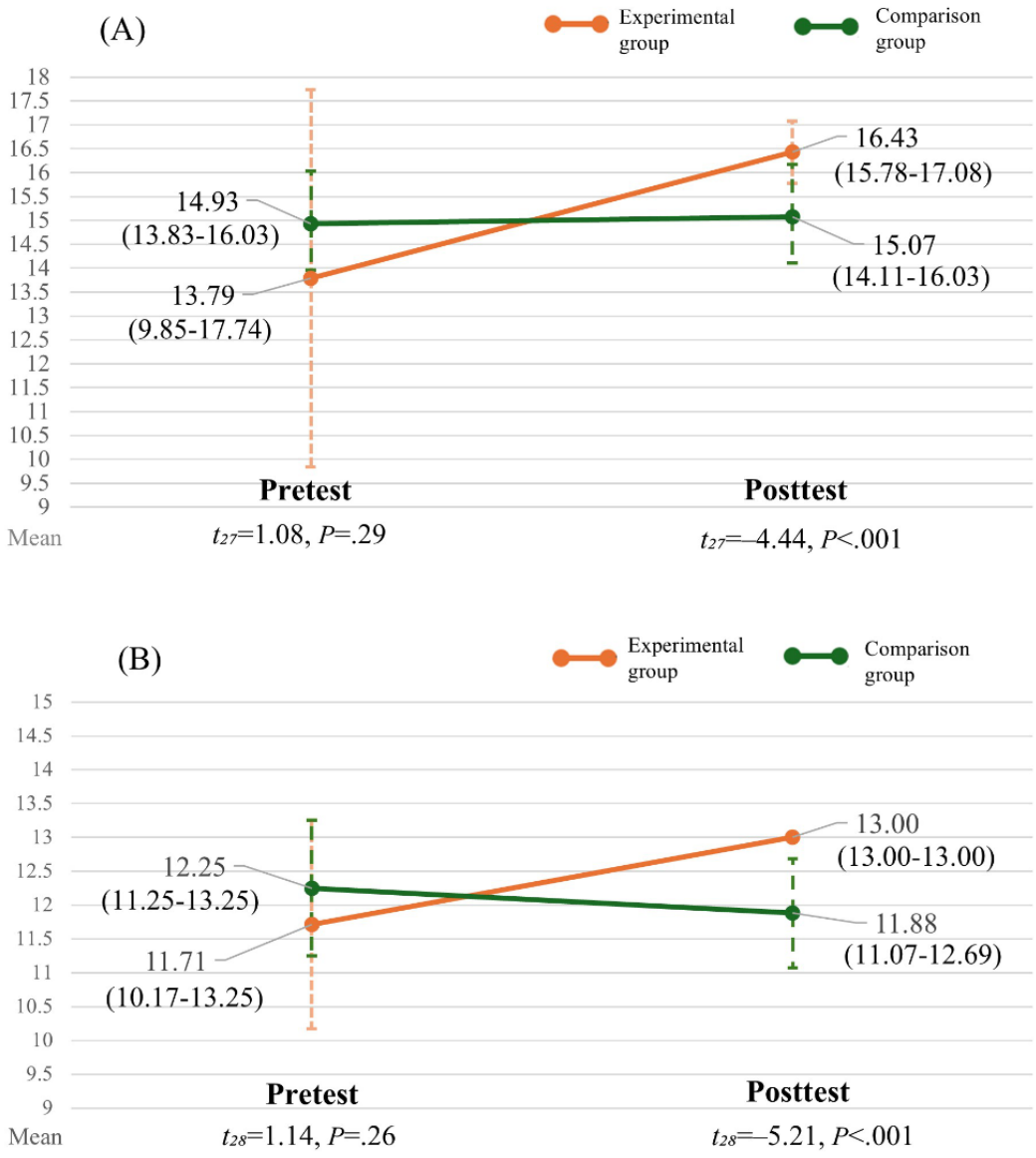

Table 2 shows how both groups' results changed over time. After age adjustment, the GEE analyses demonstrated significant improvements in CKD-specific HL (*P*=.02) and disease knowledge (*P*=.001) in the experimental group compared with the comparison group. A significant group ×time interaction was observed for both CKD-specific HL and disease knowledge. Specifically, the experimental group showed improvements in HL (coefficient=2.51, Wald χ^2_1=5.89; *P*=.02) and disease knowledge (coefficient=1.66, Wald χ^2_1=11.75; *P*=.001).

XSL•FO
RenderX

Table 2. Results of the generalized estimating equation (GEE) analyses for outcome variables.

Variable	Coefficient (β)	SE	Wald χ^2 (*df*)	*P* value
CKD[a]-specific health literacy (0-17 points)				
Group (experimental group)[b]	–1.15	1.05	1.19 (1)	.28
Time (posttest)[c]	0.13	0.23	0.34 (1)	.56
Group (experimental) × time (posttest)[d]	2.51	1.03	5.89 (1)	.02
Communicative literacy (0-7 points)				
Group (experimental group)[b]	–0.57	0.48	1.43 (1)	.23
Time (posttest)[c]	0.31	0.28	1.29 (1)	.26
Group (experimental) × time (posttest)[d]	1.26	0.51	6.17 (1)	.01
Functional literacy (0-5 points)				
Group (experimental group)[b]	–0.08	0.42	0.04 (1)	.85
Time (posttest)[c]	–0.25	0.21	1.46 (1)	.23
Group (experimental) × time (posttest)[d]	0.39	0.45	0.75 (1)	.39
Critical literacy (0-5 points)				
Group (experimental group)[b]	–0.50	0.39	1.65 (1)	.20
Time (posttest)[c]	–0.06	0.26	0.06 (1)	.81
Group (experimental) × time (posttest)[d]	0.99	0.42	5.68 (1)	.02
CKD-specific disease knowledge (0-13 points)				
Group (experimental group)[b]	–0.54	0.47	1.33 (1)	.25
Time (posttest)[c]	–0.38	0.28	1.82 (1)	.18
Group (experimental) × time (posttest)[d]	1.66	0.48	11.76 (1)	.001
Basic knowledge (0-3 points)				
Group (experimental group)[b]	–0.30	0.23	1.75 (1)	.19
Time (posttest)[c]	0.13	0.12	1.07 (1)	.30
Group (experimental) × time (posttest)[d]	0.30	0.23	1.75 (1)	.19
Prevention (0-2 points)				
Group (experimental group)[b]	0.07	0.06	1.07 (1)	.30
Time (posttest)[c]	0.07	0.06	1.07 (1)	.30
Group (experimental) × time (posttest)[d]	–0.07	0.06	1.07 (1)	.30
Lifestyle (0-4 points)				
Group (experimental group)[b]	–0.05	0.24	0.03 (1)	.85
Time (posttest)[c]	0.06	0.17	0.14 (1)	.70
Group (experimental) × time (posttest)[d]	0.30	0.25	1.36 (1)	.24
Dietary intake (0- 2 points)				
Group (experimental group)[b]	–0.02	0.13	0.02 (1)	.89
Time (posttest)[c]	–0.13	0.15	0.70 (1)	.40
Group (experimental) × time (posttest)[d]	0.27	0.18	2.30 (1)	.13
Medication (0-2 points)				

XSL•FO
RenderX

Variable	Coefficient (β)	SE	Wald χ^2 (*df*)	*P* value
Group (experimental group)[b]	–0.23	0.15	2.32 (1)	.13
Time (posttest)[c]	–0.44	0.18	6.17 (1)	.01
Group (experimental) × time (posttest)[d]	0.80	0.22	13.32 (1)	<.001

[a]CKD: chronic kidney disease.

[b]Reference group (group): comparison group.

[c]Reference group (time): pretest.

[d]Reference group (group × time): comparison group × pretest.

The group differences in patterns of change over time shown in Table 2 also demonstrated that the experimental group had significant improvements in the scores for communicative literacy (coefficient=1.26, Wald χ^2_1=6.17; *P*=.01) and critical literacy (coefficient=0.99, Wald χ^2_1=5.68, *P*=.02) compared with the comparison group, but functional literacy did not significantly improve. Regarding CKD-specific disease knowledge, the experimental group showed significant improvement in medication (coefficient=0.80, Wald χ^2_1=13.32; *P*<.001) compared with the comparison group, but there were no significant changes in areas such as basic knowledge, prevention, lifestyle, or dietary intake.

Discussion

Principal Findings

This study demonstrated preliminary evidence to support interactive robot-assisted virtual health coaching as an intervention tool for patients with CKD. Notably, participants in the experimental group showed significant improvements in both HL and disease knowledge compared with those in the comparison group. The experimental group was significantly younger than the comparison group, which may have influenced their receptiveness to the technology. However, prior studies indicate that older adults generally exhibit similar attitudes toward robots to those of younger individuals, challenging the stereotype of their lower robot-related receptiveness [29]. Furthermore, older adults often prefer human-like robots, such as the Kebbi robot, highlighting the importance of robot design and appearance in fostering acceptance [30,31]. These findings indicate that technology-enhanced educational interventions can effectively enhance HL and disease knowledge, which are associated with improved CKD-related outcomes.

The improvements observed in CKD-specific HL and disease knowledge were consistent with the benefits of technology-supported health education programs. Previous studies have shown that interactive digital interventions, particularly personalized coaching, can boost patient engagement and improve health outcomes [32,33]. This study builds on previous research by demonstrating the effectiveness of interactive robot-assisted virtual health coaching in managing chronic diseases. Our multidisciplinary approach included health professionals such as nephrologists and nurse practitioners, whose clinical experiences were integrated into the course content and represented by virtual health coaches [34]. Previous research has shown that avatar-based education has a positive effect on patient engagement and HL, which aligns with our findings [35].

A literature review of digital health coaching for older workers highlighted how virtual health can support healthy aging [36]. Although said review focused on a different group, it showed that digital interventions could effectively enhance health knowledge and well-being, especially when adopting a user-centered approach. This aligns with the findings of our study, which designed the intervention using health care professionals' clinical experiences with patients. By integrating these real-world interactions, the digital tools were tailored to address the practical needs and circumstances of patients with CKD, making the content more relevant and applicable. These results underscore the importance of health care professionals' insights when developing effective digital health interventions.

Participants in the experimental group showed significant improvements in communicative and critical literacy following the intervention, whereas functional literacy was not significantly changed. This pattern is consistent with the findings of a previous study in the Netherlands, which found that communicative HL and critical HL were more influential on chronic disease self-management, whereas functional literacy was less impactful [37]. Furthermore, a recent cross-national study involving geriatricians from Europe and Japan highlighted that virtual health interventions could enhance HL, emotional well-being, and social engagement among older adults, emphasizing the benefits of personalized digital coaching [38]. Our findings suggest that interactive virtual health coaching successfully supported patients in comprehending and using health information, making informed decisions, and critically evaluating health messages. However, the lack of improvement in functional literacy may be attributed to the program's emphasis on enhancing communication, lifestyle modifications, and decision-making skills rather than basic skills such as reading medical documents or interpreting health indicators. Additionally, certain robot interface-related limitations may have impeded the delivery of more comprehensive health information necessary for developing functional skills. Future iterations should include more comprehensive content on functional skills to address all aspects of HL, thereby enhancing its overall effectiveness.

In addition to literacy, this study evaluated the effectiveness of the intervention on CKD-specific knowledge. The results showed a significant improvement in medication use in the experimental group, whereas no notable changes were observed in other areas such as basic knowledge, prevention, lifestyle, or

XSL•FO
RenderX

dietary intake. This suggests that the program's structured content was particularly effective for teaching patients about medication. The materials, designed by health care professionals based on their clinical experiences, incorporated visual aids and interactive discussions, making the content more engaging and easier to understand. This aligns with the findings of Occa et al [39], who highlighted that interactive features and visual informativeness help improve patient knowledge by enhancing cognitive absorption. Similarly, Hassan and Davies [40] found that virtual health tools improve medication adherence by boosting patient HL and reducing barriers, emphasizing convenience and empowerment. The limited improvement in other CKD-related knowledge areas may have been due to the small sample size, which may have affected the statistical significance of our findings. Additionally, some participants reported that similar topics had already been covered in educational brochures, which might have reduced their motivation to engage fully. A more comprehensive approach that places greater emphasis on prevention, lifestyle changes, and dietary guidance is required to address these areas better.

Implications

This study has important implications for integrating digital health technologies into CKD education. Interactive robots, combined with a virtual health coaching program, have exhibited promise in enhancing HL and disease knowledge, particularly in resource-limited settings where health care resources are scarce and chronic disease management is challenging. By leveraging health care professionals' expertise and adhering to official health education guidelines, this study emphasizes the importance of creating standardized and engaging digital education tools that effectively meet patient education requirements. Furthermore, prior studies have demonstrated the potential of robots as impactful didactic tools in health science education, yielding promising results for enhancing cognitive therapies and facilitating health care intervention [41,42]. Expanding the use of such innovations may contribute to more effective patient education, improved disease knowledge, and enhanced HL, ultimately leading to improved CKD management.

Limitations

Despite these promising findings, this study had several limitations. First, participants in the experimental group were significantly younger than those in the comparison group, partly due to recruitment challenges during the COVID-19 pandemic, which hindered the enrollment of a sufficient number of older participants. Although age differences were controlled in the GEE analysis, which helped mitigate their impact, they remain a potential source of bias. Future research should further explore age-related differences, potentially incorporating multivariable analyses, to better understand and minimize their effects, ensuring the findings' broad applicability. This study primarily aimed to evaluate intervention effectiveness. Though digital literacy gaps related to age differences were not the primary focus, they represent an important area for future research.

The relatively small sample size limited the findings' generalizability, and 1-on-1 recruitment made larger sampling challenging. Statistical power was not explicitly considered during the study's design due to the lack of feasibility data, which accounts for the absence of a formal power calculation. Future studies should prioritize power analysis and incorporate larger sample sizes to confirm these results and enhance external validity. Additionally, the quasiexperimental design limited our ability to draw causal conclusions, suggesting that future research should use randomized controlled trials to strengthen causal inferences.

Another limitation involved the Taiwanese CKD population's unique characteristics, including dietary habits, medication use (eg, traditional Chinese medicine), and lifestyle factors, which may limit the findings' applicability to other populations. These cultural differences introduce potential biases, and future research should consider these factors to improve the results' generalizability.

Conclusions

This study demonstrated that interactive robot-assisted virtual health coaching could potentially enhance CKD management by improving disease knowledge and HL. To our knowledge, this is the first study to use an interactive robot for CKD health education, presenting a novel approach that alleviates the burden on health care professionals. Developed by a multidisciplinary team, the program delivers standardized, credible education while engaging patients in their care. The findings emphasize the scalability of digital health solutions, particularly in resource-limited settings, offering a valuable tool for nephrology practices. Future research should validate these results across diverse populations, examine the long-term effects on HL and disease outcomes, and refine the intervention based on patient feedback to optimize its content and usability. Integrating digital tools into clinical practice could significantly advance patient education and chronic disease management.

Acknowledgments

The authors would like to thank the participants for their time and commitment to participate in this study. This article was subsidized by the National Taiwan Normal University (NTNU), Taiwan, Republic of China. This study was funded by the Ministry of Science and Technology (NSTC 110-2511-H-003-036-MY2) of Taiwan, Republic of China.

Data Availability

The data sets generated or analyzed in this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

NJC participated in the design and coordination, formal analysis, and investigation; helped draft the manuscript; and read and approved the final version. CHC helped draft the manuscript and visualized, edited, and approved the final manuscript.
CMH was a principal project leader, conceived the study, participated in design and coordination, and read and approved the final manuscript. FHL analyzed the data and conducted the formal analysis and investigation. LTL, KYL, CLL, and YCH participated in the design, formal analysis, and investigation. JLG was a principal project leader, conceived the study, participated in the design and coordination, and read and approved the final manuscript. All the authors read and agreed to the published version of this manuscript.

Conflicts of Interest

None declared.

References

1. Khosravi H, Das S, Al-Mamun A, Ahmed I. Binary Gaussian copula synthesis: a novel data augmentation technique to advance ML-based clinical decision support systems for early prediction of dialysis among CKD patients. arXiv Preprint posted online on March 1, 2024. [doi: 10.48550/arXiv.2403.00965]
2. Lau DCW, Shaw E, McMullen S, Cowling T, Witges K, Amitay EL, et al. Acute and chronic complication profiles among patients with chronic kidney disease in Alberta, Canada: a retrospective observational study. BMC Nephrol 2024 Jul 29;25(1):244 [FREE Full text] [doi: 10.1186/s12882-024-03682-z] [Medline: 39080608]
3. The National Health Insurance Statistics, 2022. National Health Insurance Administration, Ministry of Health and Welfare. URL: https://www.nhi.gov.tw/en/cp-1302-5d3b4-213-2.html [accessed 2024-12-31]
4. Hasannejadasl H, Roumen C, Smit Y, Dekker A, Fijten R. Health literacy and eHealth: challenges and strategies. JCO Clin Cancer Inform 2022 Sep;6:e2200005. [doi: 10.1200/CCI.22.00005] [Medline: 36194843]
5. Wei C, Shih C, Hsu Y, Chen Y, Yeh J, Shih J, et al. Development and application of a chronic kidney disease-specific health literacy, knowledge and disease awareness assessment tool for patients with chronic kidney disease in Taiwan. BMJ Open 2021 Oct 11;11(10):e052597 [FREE Full text] [doi: 10.1136/bmjopen-2021-052597] [Medline: 34635527]
6. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
7. Boonstra MD, Reijneveld SA, Foitzik EM, Westerhuis R, Navis G, de Winter AF. How to tackle health literacy problems in chronic kidney disease patients? A systematic review to identify promising intervention targets and strategies. Nephrol Dial Transplant 2020 Dec 22;36(7):1207-1221 [FREE Full text] [doi: 10.1093/ndt/gfaa273] [Medline: 33351936]
8. Shah JM, Ramsbotham J, Seib C, Muir R, Bonner A. A scoping review of the role of health literacy in chronic kidney disease self-management. J Ren Care 2021 Dec;47(4):221-233. [doi: 10.1111/jorc.12364] [Medline: 33533199]
9. Harrington M. Health literacy in children with chronic kidney disease and their caregivers. Nephrol Nurs J 2015;42(1):53-56. [Medline: 26290918]
10. Lonie JM, Austin Z, Nguyen R, Gill I, Tsingos-Lucas C. Pharmacist-based health coaching: A new model of pharmacist-patient care. Res Social Adm Pharm 2017 May;13(3):644-652. [doi: 10.1016/j.sapharm.2016.06.015] [Medline: 27497786]
11. Singh HK, Kennedy GA, Stupans I. Competencies and training of health professionals engaged in health coaching: A systematic review. Chronic Illn 2022 Mar 16;18(1):58-85. [doi: 10.1177/1742395319899466] [Medline: 31948276]
12. An S, Song R. Effects of health coaching on behavioral modification among adults with cardiovascular risk factors: Systematic review and meta-analysis. Patient Educ Couns 2020 Oct;103(10):2029-2038. [doi: 10.1016/j.pec.2020.04.029] [Medline: 32448627]
13. Teng H, Yen M, Fetzer S, Sung J, Hung S. Tailoring health-promoting programs for patients with chronic kidney disease: randomized controlled trial. West J Nurs Res 2021 Mar;43(2):138-150. [doi: 10.1177/0193945920942487] [Medline: 32698694]
14. Gc VS, Iglesias CP, Erdem S, Hassan L, Peek N, Manca A. Using discrete-choice experiments to elicit preferences for digital wearable health technology for self-management of chronic kidney disease. Int J Technol Assess Health Care 2022 Oct 26;38(1):e77. [doi: 10.1017/S0266462322003233] [Medline: 36286261]
15. Maddox T, Chmielewski C, Fitzpatrick T. Virtual reality in chronic kidney disease education and training. Nephrol Nurs J 2022;49(4):329-381. [doi: 10.37526/1526-744X.2022.49.4.329] [Medline: 36054805]
16. Tsai Y, Hsiao P, Kuo M, Wang S, Chen T, Kung L, et al. Mobile health, disease knowledge, and self-care behavior in chronic kidney disease: a prospective cohort study. J Pers Med 2021 Aug 27;11(9):1 [FREE Full text] [doi: 10.3390/jpm11090845] [Medline: 34575622]
17. Santini S, Galassi F, Kropf J, Stara V. A digital coach promoting healthy aging among older adults in transition to retirement: results from a qualitative study in Italy. Sustainability 2020 Sep 09;12(18):7400. [doi: 10.3390/su12187400]
18. Helal S. Digital health: e-coaching and remote monitoring. Computer 2020 Feb;53(2):23-25. [doi: 10.1109/mc.2019.2958398]

XSL•FO
RenderX

19. Ryan K, Dockray S, Linehan C. Understanding How eHealth Coaches Tailor Support For Weight Loss: Towards the Design of Person-Centered Coaching Systems. 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; April 29, 2022-May 5,2022; New Orleans, LA. [doi: 10.1145/3491102.3501864]
20. Venning A, Herd MC, Oswald TK, Razmi S, Glover F, Hawke T, et al. Exploring the acceptability of a digital mental health platform incorporating a virtual coach: The good, the bad, and the opportunities. Health Informatics J 2021 Feb 18;27(1):1460458221994873 [FREE Full text] [doi: 10.1177/1460458221994873] [Medline: 33601947]
21. Taiwan Chronic Kidney Disease Clinical Guidelines. Taiwan Society of Nephrology. URL: https://www.tsn.org.tw/guide.html [accessed 2024-12-31]
22. Sannicandro K, De Santis A, Bellini C, Minerva T. A scoping review on the relationship between robotics in educational contexts and e-health. Front. Educ 2022 Nov 2;7:1. [doi: 10.3389/feduc.2022.955572]
23. Shih C, Chang T, Jensen DA, Chiu C. Development of a health literacy questionnaire for Taiwanese hemodialysis patients. BMC Nephrol 2016 May 31;17(1):54 [FREE Full text] [doi: 10.1186/s12882-016-0266-y] [Medline: 27245160]
24. NUWA Robotics. URL: https://www.nuwarobotics.com/en/product/ [accessed 2024-12-31]
25. Cheng Y, Wang Y, Cheng Y, Chen N. The impact of learning support facilitated by a robot and IoT-based tangible objects on children’s game-based language learning. Computer Assisted Language Learning 2022 Dec 05;37(7):2142-2173. [doi: 10.1080/09588221.2022.2152053]
26. Kostova S, Lekova A. Social Humanoid Robots as Assistive Technology for individuals with ASD - assessment of good practices. 2024 Presented at: CompSysTech '24: International Conference on Computer Systems and Technologies 2024; June 14-15, 2024; Ruse, Bulgaria. [doi: 10.1145/3674912.3674920]
27. Huang T, Ishi H, Sato-shimokawara E, Yamaguchi T. Observation of robots with eifferent expressions on education application towards entrainment robots. J. SOFT 2022;34(1):522-526. [doi: 10.3156/jsoft.34.1_522]
28. Chen N, Huang C, Fan C, Lu L, Lin F, Liao J, et al. User evaluation of a chat-based instant messaging support health education program for patients with chronic kidney disease: preliminary findings of a formative study. JMIR Form Res 2023 Oct 19;7:e45484 [FREE Full text] [doi: 10.2196/45484] [Medline: 37725429]
29. Backonja U, Hall AK, Painter I, Kneale L, Lazar A, Cakmak M, et al. Comfort and attitudes towards robots among young, middle-aged, and older adults: a cross-sectional study. J Nurs Scholarsh 2018 Dec 19;50(6):623-633 [FREE Full text] [doi: 10.1111/jnu.12430] [Medline: 30230692]
30. Tu Y, Chien S, Yeh S. Age-related differences in the uncanny valley effect. Gerontology 2020 Jun 11;66(4):382-392. [doi: 10.1159/000507812] [Medline: 32526760]
31. Fung KY, Lee LH, Sin KF, Song S, Qu H. Correction: Humanoid robot-empowered language learning based on self-determination theory. Educ Inf Technol 2024 Sep 23:1. [doi: 10.1007/s10639-024-13054-7]
32. Rowsell A, Muller I, Murray E, Little P, Byrne CD, Ganahl K, et al. Views of people with high and low levels of health literacy about a digital intervention to promote physical activity for diabetes: a qualitative study in five countries. J Med Internet Res 2015 Oct 12;17(10):e230 [FREE Full text] [doi: 10.2196/jmir.4999] [Medline: 26459743]
33. Campitelli A, Gills JL, Jones MD, Paulson S, Myers J, Bryk K, et al. The effect of a digital health coaching and health education protocol on cognition in adults at-risk for Alzheimer's. Geroscience 2023 May 17;45(2):1147-1159 [FREE Full text] [doi: 10.1007/s11357-022-00711-3] [Medline: 36527582]
34. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
35. Tongpeth J, Du HY, Clark RA. Development and feasibility testing of an avatar-based education application for patients with acute coronary syndrome. J Clin Nurs 2018 Oct 30;27(19-20):3561-3571. [doi: 10.1111/jocn.14528] [Medline: 29777554]
36. Stara V, Santini S, Kropf J, D'Amen B. Digital health coaching programs among older employees in transition to retirement: systematic literature review. J Med Internet Res 2020 Oct 24;22(9):e17809 [FREE Full text] [doi: 10.2196/17809] [Medline: 32969827]
37. Heijmans M, Waverijn G, Rademakers J, van der Vaart R, Rijken M. Functional, communicative and critical health literacy of chronic disease patients and their importance for self-management. Patient Educ Couns 2015 Jan;98(1):41-48. [doi: 10.1016/j.pec.2014.10.006] [Medline: 25455794]
38. Stara V, Soraci L, Takano E, Kondo I, Möller J, Maranesi E, et al. Intrinsic capacity and active and healthy aging domains supported by personalized digital coaching: survey study among geriatricians in Europe and Japan on eHealth opportunities for older adults. J Med Internet Res 2023 Oct 12;25:e41035 [FREE Full text] [doi: 10.2196/41035] [Medline: 37824183]
39. Occa A, Morgan SE, Peng W, Mao B, McFarlane SJ, Grinfeder K, et al. Untangling interactivity's effects: The role of cognitive absorption, perceived visual informativeness, and cancer information overload. Patient Educ Couns 2021 May;104(5):1059-1065 [FREE Full text] [doi: 10.1016/j.pec.2020.10.007] [Medline: 33969824]
40. Hassan A, Davies N. Expert insight into the use of eHealth interventions to aid medication adherence during COVID-19. Patient Prefer Adherence 2024;18:721-731 [FREE Full text] [doi: 10.2147/PPA.S437822] [Medline: 38529042]

XSL•FO
RenderX

41. Marcos-Pablos S, García-Peñalvo FJ. More than surgical tools: a systematic review of robots as didactic tools for the education of professionals in health sciences. Adv Health Sci Educ Theory Pract 2022 Oct 30;27(4):1139-1176 [FREE Full text] [doi: 10.1007/s10459-022-10118-6] [Medline: 35771316]
42. Ashwini B, Narayan V, Bhatia A, Shukla J. Responsiveness towards robot-assisted interactions among pre-primary children of Indian ethnicity. 2021 Presented at: 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN); August 8-12, 2021; Vancouver, BC, Canada. [doi: 10.1109/ro-man50785.2021.9515520]

Abbreviations

CKD: chronic kidney disease
GEE: generalized estimating equations
HL: health literacy

Edited by G Eysenbach, A Coristine; submitted 28.10.24; peer-reviewed by SC Sung, CY Lin, NT Conway, T Akindahunsi, O Omoyemi, A Adebisi; comments to author 20.11.24; revised version received 09.12.24; accepted 23.12.24; published 09.01.25.

Please cite as:
Chen NJ, Chang CH, Huang CM, Lin FH, Lu LT, Liu KY, Lai CL, Lin CY, Hou YC, Guo JL
Assessing the Effectiveness of Interactive Robot-Assisted Virtual Health Coaching for Health Literacy and Disease Knowledge of Patients with Chronic Kidney Disease: Quasiexperimental Study
J Med Internet Res 2025;27:e68072
URL: https://www.jmir.org/2025/1/e68072
doi:10.2196/68072
PMID:

©Nai-Jung Chen, Ching-Hao Chang, Chiu-Mieh Huang, Fen-He Lin, Li-Ting Lu, Kuan-Yi Liu, Chih-Lin Lai, Chin-Yao Lin, Yi-Chou Hou, Jong-Long Guo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Studying the Digital Intervention Engagement–Mediated Relationship Between Intrapersonal Measures and Pre-Exposure Prophylaxis Adherence in Sexual and Gender Minority Youth: Secondary Analysis of a Randomized Controlled Trial

Michael P Williams[1], PhD; Justin Manjourides[1], PhD; Louisa H Smith[2], PhD; Crissi B Rainer[3], MSc; Lisa B Hightow-Weidman[3], MD; Danielle F Haley[4], PhD

[1]Bouve College of Health Sciences, Northeastern University, Boston, MA, United States

[2]Roux Institute, Northeastern University, Portland, ME, United States

[3]College of Nursing, Florida State University, Tallahassee, FL, United States

[4]Department of Community Health Sciences, Boston University, Boston, MA, United States

Corresponding Author:
Michael P Williams, PhD
Bouve College of Health Sciences
Northeastern University
30 Leon St
Boston, MA, 02115
United States
Phone: 1 617 373 3323
Email: mpw144@gmail.com

Abstract

Background: Improving adherence to pre-exposure prophylaxis (PrEP) via digital health interventions (DHIs) for young sexual and gender minority men who have sex with men (YSGMMSM) is promising for reducing the HIV burden. Measuring and achieving effective engagement (sufficient to solicit PrEP adherence) in YSGMMSM is challenging.

Objective: This study is a secondary analysis of the primary efficacy randomized controlled trial (RCT) of Prepared, Protected, Empowered (P3), a digital PrEP adherence intervention that used causal mediation to quantify whether and to what extent intrapersonal behavioral, mental health, and sociodemographic measures were related to effective engagement for PrEP adherence in YSGMMSM.

Methods: In May 2019, 264 YSGMMSM were recruited for the primary RCT via social media, community sites, and clinics from 9 study sites across the United States. For this secondary analysis, 140 participants were eligible (retained at follow-up, received DHI condition in primary RCT, and completed trial data). Participants earned US currency for daily use of P3 and lost US currency for nonuse. Dollars accrued at the 3-month follow-up were used to measure engagement. PrEP nonadherence was defined as blood serum concentrations of tenofovir-diphosphate and emtricitabine-triphosphate that correlated with ≤4 doses weekly at the 3-month follow-up. Logistic regression was used to estimate the total effect of baseline intrapersonal measures on PrEP nonadherence, represented as odds ratios (ORs) with a null value of 1. The total OR for each intrapersonal measure was decomposed into direct and indirect effects.

Results: For every US $1 earned above the mean (US $96, SD US $35.1), participants had 2% (OR 0.98, 95% CI 0.97-0.99) lower odds of PrEP nonadherence. Frequently using phone apps to track health information was associated with a 71% (OR 0.29, 95% CI 0.06-0.96) lower odds of PrEP nonadherence. This was overwhelmingly a direct effect, not mediated by engagement, with a percentage mediated (PM) of 1%. Non-Hispanic White participants had 83% lower odds of PrEP nonadherence (OR 0.17, 95% CI 0.05-0.48) and had a direct effect (PM=4%). Participants with depressive symptoms and anxiety symptoms had 3.4 (OR 3.42, 95% CI 0.95-12) and 3.5 (OR 3.51, 95% CI 1.06-11.55) times higher odds of PrEP nonadherence, respectively. Anxious symptoms largely operated through P3 engagement (PM=51%).

Conclusions: P3 engagement (dollars accrued) was strongly related to lower odds of PrEP nonadherence. Intrapersonal measures operating through P3 engagement (indirect effect, eg, anxious symptoms) suggest possible pathways to improve PrEP adherence DHI efficacy in YSGMMSM via effective engagement. Conversely, the direct effects observed in this study may reflect existing

XSL•FO
RenderX

structural disparity (eg, race and ethnicity) or behavioral dispositions toward technology (eg, tracking health via phone apps). Evaluating effective engagement in DHIs with causal mediation approaches provides a clarifying and mechanistic view of how DHIs impact health behavior.

Trial Registration: ClinicalTrials.gov; NCT03320512; https://clinicaltrials.gov/study/NCT03320512

(J Med Internet Res 2025;27:e57619) doi:10.2196/57619

KEYWORDS

engagement; pre-exposure prophylaxis; PrEP; digital health intervention; adherence; men who have sex with men; sexual orientation; gender minority; youth; adolescent; teenager; HIV; randomized controlled trial; mental health; sociodemographic; logistic regression; health information; health behavior; sexual health

Introduction

Background

Young sexual and gender minority men who have sex with men (YSGMMSM) are burdened with a disproportionate and growing vulnerability to HIV in the United States [1-8]. Of the 30,692 incident HIV cases in 2020, 71% were among men who have sex with men (MSM) and 24% were among MSM aged 13 to 24 years [8]. In addition, a meta-analysis of transgender women aged ≥15 years found that they were 48 times more likely to have HIV infection compared to other adults of reproductive age [1]. While pre-exposure prophylaxis (PrEP) has demonstrated efficacy in reducing the incidence of HIV infections in YSGMMSM in the US in randomized controlled trials (RCTs) [9,10], adherence to these medications outside of clinical trial settings has been suboptimal for reducing transmission [9,11]. Furthermore, findings from population and agent-based simulation studies demonstrate that PrEP uptake and adherence are associated with a decrease in incident HIV infections, lowering incidence by as much as 25% [12,13]. Altogether, this suggests that improving PrEP adherence is a fruitful pathway to reducing HIV incidence in YSGMMSM [9,11-13].

Digital health interventions (DHIs) are potentially powerful and increasingly prevalent mechanisms for delivering PrEP interventions to YSGMMSM [14-20]. Due to the pervasiveness of digital communication and entertainment among youth, including YSGMMSM, DHIs are suitable delivery mechanisms for HIV prevention interventions due to their potential to effectively engage participants [21-25]. However, to our knowledge, there is a paucity of clinical trials directly testing the efficacy of digital PrEP adherence interventions. Adjacent digital PrEP adherence interventions, such as automated SMS text-messaging services or digital pill systems, have had mixed results in young populations. For example, digital pill systems have documented barriers to engagement and have not shown that they independently improve PrEP adherence compared to a standard of care [26-29]. Furthermore, DHIs implemented to address other health problems in target youth populations have encountered problems in effectively engaging the participants. For example, 1 study that developed a mobile app for self-management of type 1 diabetes among adolescents found no association between intervention conditions and primary or secondary clinical outcomes. Furthermore, the study found that only 9% of the participants met the criteria for high engagement levels (measured as an individual uploading blood glucose readings for ≥3 d/wk) [30]. For these reasons, effectively engaging YSGMMSM in PrEP adherence DHIs is likely an essential element to achieving protective levels of PrEP adherence [25,31,32].

The effective engagement framework by Yardley et al [25] defines effective engagement as "sufficient engagement with the intervention to achieve intended outcomes." They describe four phases of engagement as follows: (1) initial engagement with the DHI in preparation for behavior change; (2) engagement with the behavior change, mediated by the DHI; (3) DHI use may no longer be required to sustain behavior change; and (4) reengagement with DHI as needed. Furthermore, this framework describes how, in phase 1, effective engagement is largely characterized by a micro form of engagement (defined by moment-to-moment interactions with the DHI). As individuals progress through phases 2, 3, and 4, they are increasingly engaged with a macro form of engagement (defined as engagement that relates to the overarching goal of the intervention).

The effective engagement framework by Yardley et al [25] also highlights how intrapersonal, social, and environmental characteristics can influence effective engagement and how tailoring DHIs based on those characteristics may improve effective engagement. Tailoring, defined as the use of individuals' data to customize intervention content based on their psychological, socioecological, and behavioral profile, is a property of DHIs that has demonstrated efficacy in promoting effective engagement with DHIs [25,33-37]. Several barriers and facilitators to engagement have been identified in qualitative research [38]. For example, baseline motivation to change [25,39] or baseline comfort with the intervention modality (eg, phone or app-based intervention) may influence engagement and subsequent behavior change [38-49]. Furthermore, aspects beyond the individual's control, such as internet access, may influence engagement and therefore intervention efficacy [38,50-52]. However, there is a lack of quantitative research that aims to isolate and quantify the impact that intrapersonal psychological, sociodemographic, and behavioral measures have on engagement and subsequent behavioral outcomes of interest.

Yardley et al [25] and other review articles using their effective engagement framework describe a plethora of issues with quantitively operationalizing and measuring effective engagement in previous research. Most notably, engagement research to date is largely correlational, relies on the assumption

XSL•FO
RenderX

that engagement is intrinsically a precursor to the intended outcome, and does not account for intrapersonal measures, such as motivation to use the intervention or digital literacy [25,31,32,39,53-55]. This highlights the need to empirically test and corroborate the models of engagement, by modeling how intrapersonal measures influence engagement, which subsequently mediates behavioral outcomes of interest (eg, PrEP adherence) [25,53,56]. Causal mediation analysis is a fitting statistical framework for investigating the relationship between intrapersonal measures and effective engagement which addresses the aforementioned methodological issues as well. Causal mediation has been extensively explicated [57-62], applied in previous research in other domains [63], and addresses the aforementioned issues by Yardley et al [25] by clarifying assumptions needed for causal inference in mediation models, explicitly modeling the relationship between engagement and outcome (as opposed to relying on the assumption that engagement and the outcome are related), and incorporates intrapersonal measures (as exposures or controls) [57,58,60]. This approach is derived from the counterfactual causal inference framework [64,65] and allows for the total effect (eg, the effect of baseline digital literacy on PrEP adherence) to be decomposed into a direct and indirect effect. The direct effect models the effect of a given intrapersonal measure on PrEP adherence controlling for engagement with the intervention. Conversely, the indirect effect models how a given intrapersonal measure is related to PrEP adherence operating through engagement, and thus, is an excellent measure of effective engagement. The integration of the effective engagement framework with the casual mediation approach provides a combined theoretical and analytical approach for evaluating effective engagement in DHIs.

This Study

This study collates clinical survey data, biological PrEP adherence measures, and engagement measures from the Prepared, Protected, Empowered (P3) intervention efficacy RCT [19,66] in a secondary data analysis using the causal mediation framework [57-62] to quantify whether and to what degree intrapersonal behavioral, mental health, and sociodemographic measures impact effective engagement with respect to PrEP adherence in YSGMMSM.

Methods

Study Design

This study is a secondary analysis, which combined clinical survey data, biological PrEP adherence measures, and engagement measures collected from a 3-arm RCT testing the efficacy of P3, a PrEP adherence DHI [19,66]. This secondary analysis used the causal mediation framework and statistical analysis procedures to characterize whether engagement with the P3 intervention mediated the relationship between baseline intrapersonal measures and PrEP adherence at 3 months.

Ethical Considerations

The parent study was reviewed and approved by the institutional review board of the University of North Carolina at Chapel Hill (17-9551). A certificate of confidentiality was obtained from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. For participants aged between 15 and 17 years, a waiver of parental consent was obtained. This trial is registered at ClinicalTrials.gov (NCT03320512). This secondary analysis was reviewed by the Northeastern University institutional review board and determined to be exempt under category 4 (secondary research for which consent is not required). This secondary analysis used a deidentified analytic dataset curated by the parent study's staff. The principal investigator had no contact with participants and made no attempts to reidentify participants post hoc.

Parent Study

Intervention

P3 is a user-centered PrEP adherence phone app that incorporates a variety of content in multiple formats to serve the diverse needs, barriers, and motivations of YSGMMSM. This phone app included text, videos, quizzes, and a social wall where participants could share experiences, from success stories to challenges. In addition, P3 incorporated game-like elements, such as daily health-related quests, in-app rewards, unlocking character-driven narratives, and social connection activities. P3 used a financial incentive to encourage daily use in which small monetary incentives are awarded for daily use of P3 (not necessarily PrEP). Participants started with an initial bank of US $90 and then were awarded US $0.50 for each day on which they logged into P3 and completed one of the following tasks: (1) post on the social wall, (2) use the medication tracker, or (3) complete a quest. Each of these tasks corresponds to a putative behavior change mechanism (social support, instrumental support, and gamification, respectively). Conversely, US $1 was deducted for each day on which the participant did not log in and complete 1 of the aforementioned tasks across the 90-day trial period. The maximum a participant could earn was US $135 and the minimum was US $0. P3+ is an extension of P3, in which participants were also connected with an adherence counselor trained on the Next Step Counseling adherence counseling curriculum through the P3 phone app [19,66-68].

Clinical Trial Eligibility and Procedures

Starting in May 2019, participants were recruited from 9 study sites as follows: Tampa, Florida; Boston, Massachusetts; Chicago, Illinois; Houston, Texas; Philadelphia, Pennsylvania; Chapel Hill, North Carolina; Atlanta, Georgia; Bronx, New York; and Charlotte, North Carolina. A mix of in-person, venue-based, and web-based recruitment methods was used. Inclusion criteria were as follows: individuals (1) who were aged between 16 and 24 years, (2) who were assigned male sex at birth, (3) who reported sex with or intentions to have sex with men or transwomen, (4) who had reliable daily access to an Android (Google LLC) or iOS (Apple Inc) smartphone with a data plan, (5) who could speak and read English, (6) who were HIV-uninfected (confirmed by self-report at enrollment visit), and (7) who were not on PrEP but planned to initiate in the next 7 days and had an active PrEP prescription (prescription confirmed by study staff) or those who were currently on PrEP who had an active PrEP prescription (prescription confirmed by study staff). After providing informed consent either in

XSL•FO
RenderX

person or electronically, participants were randomized to 1 of the 3 treatment arms (standard of care, P3, or P3+) using a 1:1:1 randomization scheme. Clinical survey assessments and laboratory specimens were collected at baseline and 3 months into the trial period. Engagement measures (in the 2 intervention arms) were collected continuously throughout the intervention period and summarized at 3 months [19]. Study visits were initially planned to be conducted in person at the same study site where participants enrolled. All study sites stopped in-person study activities on March 17, 2020, to reduce the transmission of COVID-19. Web-based recruitment and web-based study activities began in June 2020. In addition, some study sites were able to conduct limited in-person activities based on local regulations and COVID-19 restrictions. The trial concluded in September 2021.

Secondary Analysis Eligibility

Participants from the primary study who received the P3 or P3+ intervention were eligible for inclusion in this secondary data analysis (n=163). Participants in the P3 and P3+ conditions who were lost to follow-up (LTFU; defined as participants who did not begin the month 3 survey) were excluded (22/163, 13.5%). Moreover, 1 (0.7%) of the remaining 141 participants with incomplete survey information pertaining to study-relevant exposure measures was also removed. This resulted in a dataset of 140 participants (Figure 1).

Figure 1. Primary Prepared, Protected, Empowered (P3) randomized controlled trial (RCT) participants' eligibility for inclusion in this study and secondary analysis of effective engagement with respect to pre-exposure prophylaxis adherence in young sexual and gender minority men who have sex with men. LTFU: lost to follow-up; SOC: standard of care; P3+: extension of Prepared, Protected, emPowered.

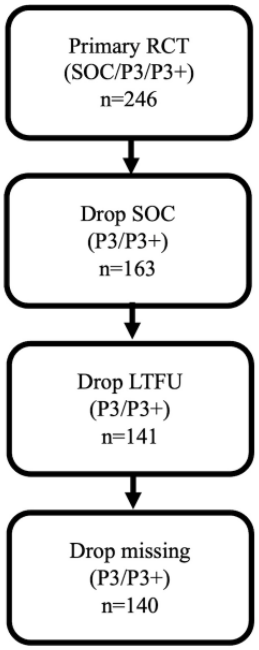

Outcome: PrEP Nonadherence

PrEP nonadherence at 3 months (binary) was the primary outcome measure used in this analysis. If serum levels of tenofovir-diphosphate and emtricitabine-triphosphate were consistent with ≤4 doses per week, the participant was considered nonadherent. Due to study operation interruptions related to the COVID-19 pandemic, 28.6% (40/140) of eligible participants were unable to provide biological specimens. In cases where tenofovir-diphosphate and emtricitabine-triphosphate values were missing, self-reported doses of PrEP in the last 7 days before the 3-month clinical survey were used. Participants who reported ≤4 self-reported doses of PrEP in the last 7 days were considered nonadherent. There is mixed evidence regarding the accuracy of self-reported measures of PrEP adherence. While 2 studies have found that self-report measures of PrEP adherence correlate with protective serum levels among adults [69,70], another study examining self-report PrEP adherence measure accuracy among 15 to 23-year-old young MSM found that self-reported measures overreported adherence compared to biological specimens, with the odds of overreporting decreasing by 24% (odds ratio [OR] 0.74, 95% CI 0.65-0.90) with each additional year of age [71]. We found that the area under the receiver operating characteristics curve between self-report measures and biological measures among participants with biological and self-report PrEP nonadherence measures was high (≥0.7). Further details are described in the Strengths and Limitations section.

Mediator: Intervention Engagement

Engagement was defined as dollars accrued by 3 months. This measure was mean centered (participant dollars accrued—mean dollars accrued) when used in modeling. As mentioned in the Intervention subsection, participants started with a baseline amount of US $90 in a bank and gained or lost money from this initial bank based on the number of days each participant logged in and completed one of the 3 aforementioned tasks (eg, completing a quest). This measure serves as a quality proxy for engagement because it correlates with the behavior pattern P3 aims to adjust (ie, daily use of P3 mimics the daily dosing pattern of PrEP) and relates to the putative behavior change mechanisms used in the P3 app (eg, completing a quest is related to gamification and posting on the social wall is related social engagement). Engagement measures derived from documenting

XSL•FO
RenderX

users' access, participation, and navigation through a DHI are suitable for constructing measures of engagement, as these measures provide an objective view into patterns of use with high ecological validity [25,31,32].

Intrapersonal Measures

Consistent with Yardley et al [25], several intrapersonal measures capturing experiences with phone and phone apps, mental health, and sociodemographic characteristics were constructed from the baseline survey administered in the primary RCT based on prior research [25,38]. These measures, described subsequently, focused on patterns of phone and phone app use, measures of mental health, and sociodemographic information.

Phone and Phone App Use

Several binary measures describing phone and phone app use were derived from the baseline survey. A participant was considered to have experienced disconnects if they lost access to their phone or phone service at any time in the year leading up to the baseline survey. Prior qualitative research identified poor internet access as a barrier to engagement in DHIs [38,40-42,45,48,50-52,72]. Those who spent an average of ≥7 hours per day on the internet outside of work or school were considered high internet users. Those who used phone apps ≥2 times per day were considered frequent phone app users. These measures acted as proxy variables for digital literacy and digital familiarity, which have been shown to act as engagement facilitators in qualitative research [34,39-49,73-75]. Binary measures describing participants' propensity to use phone apps for a variety of purposes were derived from questions whose answers follow a Likert scale with the following ranked choices: never, rarely, sometimes, often, and decline to answer. Participants who disclosed that they "often" use phone apps for chatting with friends, chatting with family, looking for romantic dates, looking for casual sex, or tracking their health were considered highly interested in using phone apps for those activities. Previous qualitative research has identified social support in various forms enabled through a DHI as an engagement facilitator [38,44,46,47,76,77]. Finally, a binary measure representing the intervention arm (P3 or P3+) was constructed to be used as a control measure.

Mental Health

The Patient Health Questionnaire-8 (PHQ-8) [78,79] and Generalized Anxiety Disorder-7 (GAD-7) [80] questionnaires were used to assess depressive and anxiety symptoms, respectively. Both scales asked participants to rank how frequently they experience symptoms from not at all, several days, more than half the days, and nearly every day. Scores ranged from 0 to 24 in the PHQ-8 and 0 to 21 in the GAD-7, with lower scores representing less frequent experiences and higher scores representing more frequent experiences of depressive and anxious symptoms, respectively. Participants who scored ≥10 on the PHQ-8 and GAD-7 were considered to have depressive or anxious symptoms, respectively [79-81]. Previous research has identified that psychological distress from trauma is associated with lower engagement, which suggests that other stressors on mental health may also act as barriers to engagement [82]. Furthermore, symptoms of depression have been consistently linked to lower treatment adherence [83], and symptoms of depression and anxiety have been found to be associated with a higher likelihood of medication nonadherence [84].

Sociodemographic Measures

Sociodemographic measures captured from the baseline survey include race, ethnicity, and age. Participants were considered non-Hispanic White if they disclosed non-Hispanic ethnicity and White as their race. In previous research, sociodemographic measures had mixed effects on engagement [85-92]. For example, older age has been observed as an engagement facilitator and an engagement barrier in different studies [85,86]. These measures were used to control for confounding (further specifications are given subsequently) and investigated as exposures of interest for their relationship with effective engagement.

Statistical Analysis

Descriptive statistics for participants who were eligible for secondary analysis were generated and are reported in the Results section. There is also a comparison of 140 eligible participants to the 22 participants who were LTFU. To quantify how engagement with P3 mediates the relationship between baseline intrapersonal measures and PrEP nonadherence at 3 months, the causal mediation framework was used. This analytical approach extends the counterfactual causal inference framework to mediation, has been extensively explicated, clarifies several confounding assumptions, accommodates exposure-mediator interaction [57-62,93], has been applied in previous research [63], and complements the effective engagement theoretical framework by explicitly decomposing the total effect of each intrapersonal measure on PrEP nonadherence into a direct and indirect effect (ie, the effect mediated by P3 engagement). Figure 2 depicts the integration of effective engagement and causal mediation through a theoretical causal diagram. In this study, effect decomposition was accomplished by first fitting a linear regression to assess the effect of each exposure (eg, anxious symptoms) on the mediator (mean-centered dollars accrued at 3 months), adjusting for confounding. Then, a logistic model was fit to examine the relationship between each exposure (eg, anxious symptoms) and PrEP nonadherence, adjusting for the same confounders, dollars accrued at 3 months (ie, the mediator), and the exposure-mediator interaction. Baseline measures, such as age, race, and intervention arm, were used to adjust for confounding. Several other critical measures were controlled through the primary study's design via the eligibility criteria. Intention to initiate PrEP, PrEP access, sexual orientation, and English literacy were verified at enrollment. The total effects estimated from these 2 models with corresponding CIs and *P* values are reported in the Results section. The total effect for each intrapersonal measure on PrEP nonadherence at 3 months is presented as an OR (defined as the rate of nonadherence in the exposed divided by the rate of nonadherence in the unexposed). These 2 regression models are then used to derive direct and indirect effects. The direct effect represents the effect of a given intrapersonal measure on PrEP nonadherence independent of P3 engagement. This effect is estimated by comparing the

XSL•FO
RenderX

estimated PrEP nonadherence in the exposed relative to the unexposed while setting P3 engagement to the level that would have naturally occurred in the absence of the exposure. The indirect effect represents the effect of a given intrapersonal measure on PrEP nonadherence operating through the mediator. This effect is estimated by comparing the outcome for the exposed for different contrasts of the mediator (eg, between levels of P3 engagement). Mediation results, including direct and indirect effects with corresponding CIs and percentage mediated (PM) are reported in the Results section. Mediation is assessed using a combination of total and indirect effect size, statistical significance, and PM. *P* values are reported for transparency, but CIs are used as the primary determinant of statistical significance for mediation analysis, determined by if the CI overlaps with the null value. For the first model (linear regression of dollars accrued), the null value is 0, and for the second model (logistic regression of binary PrEP adherence), the null value is 1. All analyses were carried out using R (GNU) and RStudio (Posit, PBC). Mediation models were constructed using the *CMAverse* R library using the regression-based approach and imputation as the estimation method [94].

Figure 2. Integration of effective engagement and causal mediation frameworks in a causal diagram. Direct effects represent relationships between intrapersonal measures and pre-exposure prophylaxis (PrEP) adherence controlling for engagement with Prepared, Protected, Empowered (P3). Indirect effects represent relationships between intrapersonal measures and PrEP adherence operating through P3 engagement (ie, mediated effects).

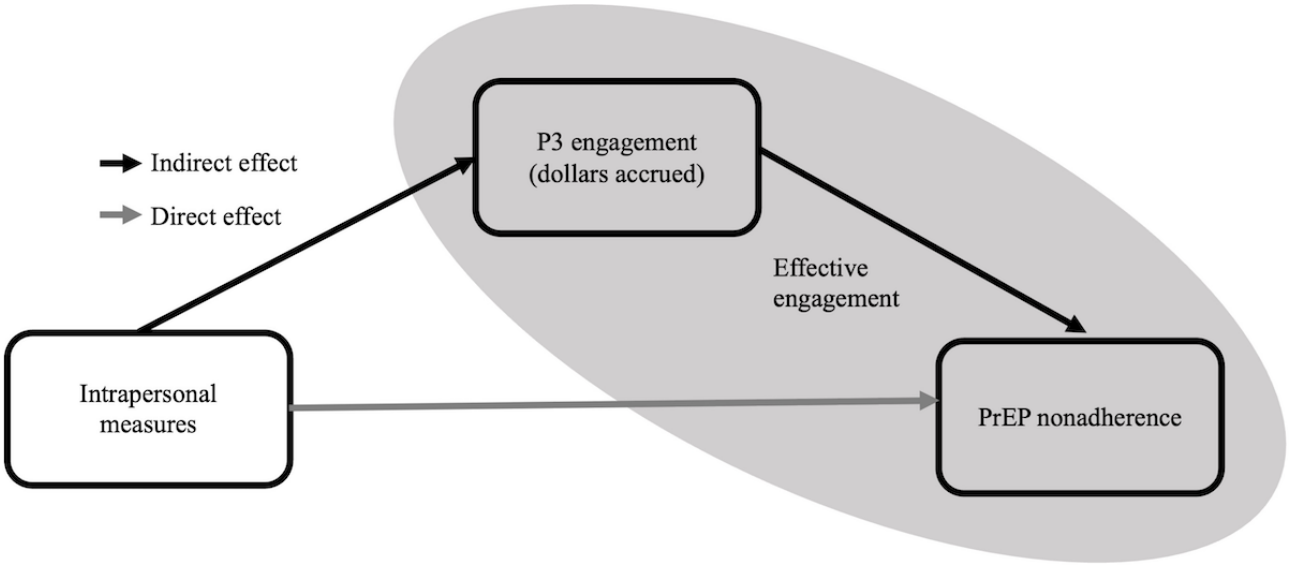

Results

Participant Characteristics

The median age of participants was 22 years (IQR 20-23). Overall, 17.1% (24/140) were considered nonadherent to PrEP at 3 months (Table 1). The minimum amount the participants earned over the 90-day trial period was US $1.5 (representing 1 day logged in with a task completed). The maximum amount the participants earned over the 90-day trial period was US $135 (corresponding to logging in and completing a task every day for the 90-day trial period). Participants earned a mean of US $96.40 (SD US $35.1) over the 90-day trial period. Participants earned a median of US $112.50 (IQR US $73.50-US $ 123.50) which corresponds to 75 days logged in with a task completed (IQR 49-82) over the 90-day trial period. Participants were largely heterogeneous with respect to phone, technology, and internet use patterns. Notable exceptions to this pattern were as follows: 90.7% (127/140) of the participants used phone apps more than once per day and 90.7% (127/140) of the participants disclosed that they often use phone apps for chatting with friends. No significant differences in selected baseline intrapersonal characteristics were observed between eligible and LTFU participants (Table 1).

XSL•FO
RenderX

Table 1. Comparison of participant characteristics who were eligible for secondary analysis and those who were lost to follow-up (LTFU) in a secondary analysis of engagement with a pre-exposure prophylaxis (PrEP) adherence digital health intervention (DHI) among US young sexual and gender minority men who have sex with men aged between 16 and 24 years.

	Eligible participants (n=140)	LTFU participants (n=22)	*P* value[a]
Nonadherent to PrEP at 3 months, n (%)	24 (17.1)	—[b]	—
High internet users, n (%)	29 (20.7)	3 (13.6)	.63
Disconnect from phone in past 12 months, n (%)	16 (11.4)	3 (13.6)	.99
Frequent phone app users, n (%)	127 (90.7)	21 (95.5)	.74
Frequent uses of phone apps, n (%)			
Chatting with friends	126 (90)	22 (100)	.29
Chatting with family	76 (54.3)	13 (59.1)	.85
Finding romantic dates	43 (30.7)	4 (18.2)	.34
Looking for casual sex	38 (27.1)	5 (22.7)	.86
Tracking health	38 (27.1)	8 (36.4)	.53
Depressive symptoms, n (%)	18 (12.9)	2 (9.1)	.88
Anxiety symptoms, n (%)	21 (15)	5 (22.7)	.55
Non-Hispanic White, n (%)	69 (49.3)	9 (40.9)	.62
Male, n (%)	128 (91.4)	20 (90.9)	1.0
Intervention arm: P3+, n (%)	71 (50.7)	9 (40.9)	.53
Age (y), median (IQR)	22 (20-23)	22 (20-23)	.53[c]
Site, n (%)			.12
Tampa	23 (16.4)	5 (22.7)	
Atlanta	10 (7.1)	5 (22.7)	
Boston	22 (15.7)	2 (9.1)	
Philadelphia	14 (10)	4 (18.2)	
Chicago	23 (16.4)	1 (4.5)	
Houston	18 (12.9)	4 (18.2)	
Bronx	9 (6.4)	1 (4.5)	
Chapel Hill	15 (10.7)	0 (0)	
Charlotte	6 (4.3)	0 (0)	

[a]Continuous measures tested with 2-tailed *t* tests, categorical measures tested with Fisher exact test.

[b]LTFU participants did not return for 3-month PrEP adherence data collection.

[c]For nonnormal distribution, the Kruskal-Wallis rank sum test was used.

Multivariate Analysis

Table 2 presents multivariate regression results for the effect of intrapersonal measures on engagement (dollars accrued by 3 months). Frequent phone app users earned US $21.49 (95% CI 2.50-40.47) more than infrequent phone app users through the 3-month trial period. Participants who reported anxious symptoms in the past 2 weeks (GAD-7 score≥10) earned US $15.95 (95% CI –31.57 to –0.32) less throughout the 3-month trial period than those with mild or no anxious symptoms in the past 2 weeks. Non-Hispanic White individuals earned US $17.02 (95% CI 5.95-28.10) more on average by 3 months than participants belonging to other racial and ethnic groups. Participants who received the P3+ intervention (P3 with the addition of human adherence counselors accessible through the app) earned US $12.48 (95% CI 1.47-23.50) more on average than those who received the standard P3 app through the 3-month trial period. Finally, for each additional year of age, participants earned on average an additional US $2.55 (95% CI –0.19 to 5.29). The 95% CI for the relationship between age and engagement narrowly overlaps the null value of 0 with a corresponding *P* value of .07, meaning this relationship is technically statistically insignificant.

XSL•FO
RenderX

Table 2. Multivariate relationships between intrapersonal measures and engagement with Prepared, Protected, Empowered (P3), pre-exposure prophylaxis (PrEP) nonadherence among US young sexual and gender minority men who have sex with men aged between 16 to 24 years (n=140).

Intrapersonal measures	Engagement[a]		PrEP nonadherence[a]	
	Estimate (95% CI)[b]	*P* value	OR[c,d] (95% CI)[e]	*P* value
Dollars accrued at 3 months	—[f]	—	0.98 (0.97 to 0.99)	.02
High internet user	–5.51 (–19.65 to 8.63)	.45	2.31 (0.80 to 6.46)	.11
Disconnect from the phone in the past 12 months	–6.42 (–24.00 to 11.16)	.48	3.84 (1.14 to 12.81)	.03
Frequent phone app user	21.49 (2.50 to 40.47)	.03	0.66 (0.17 to 2.88)	.55
Frequently uses of phone apps				
Chatting with friends	–1.26 (–20.36 to 17.85)	.90	0.82 (0.20 to 4.37)	.80
Chatting with family	9.71 (–1.45 to 20.86)	.09	0.99 (0.38 to 2.63)	.99
Finding romantic dates	1.22 (–11.14 to 13.57)	.85	0.78 (0.23 to 2.32)	.67
Looking for casual sex	–5.06 (–17.49 to 7.36)	.43	0.71 (0.21 to 2.08)	.55
Tracking health	2.00 (–10.42 to 14.41)	.75	0.29 (0.06 to 0.96)	.06
Depressive symptoms	–8.16 (–24.82 to 8.51)	.34	3.42 (0.95 to 12.00)	.05
Anxiety symptoms	–15.95 (–31.57 to –0.32)	.05	3.51 (1.06 to 11.55)	.04
Non-Hispanic White[g]	17.02 (5.95 to 28.10)	.003	0.17 (0.05 to 0.48)	.002
Intervention arm: P3+[g]	12.48 (1.47 to 23.50)	.03	1.05 (0.41 to 2.71)	.92
Age[g]	2.55 (–0.19 to 5.29)	.07	0.82 (0.65 to 1.02)	.08

[a]Multivariate models are adjusted for age, race/ethnicity, and intervention arm.

[b]Null value is 0.

[c]OR: odds ratio.

[d]Derived by exponentiating estimated regression coefficients.

[e]Null value is 1.

[f]Not available.

[g]Intrapersonal measure is also a control measure. Model constructed using age, race, and intervention arm for engagement and age, race, ethnicity, intervention arm, dollars accrued at 3 months, and dollars accrued at 3 months and the interaction between dollars accrued at 3 months and the focal intrapersonal measure.

Multivariate models are adjusted for age, race, ethnicity, intervention arm, dollars accrued at 3 months, and dollars accrued at 3 months and the interaction between dollars accrued at 3 months and the focal intrapersonal measure.

Total effects for the relationship between baseline intrapersonal measures and PrEP nonadherence at 3 months are reported in Table 2 as ORs. For every dollar earned above the mean throughout the 3-month trial period, participants had 2% (OR 0.98, 95% CI 0.97-0.99) lower odds of PrEP nonadherence at 3 months. Participants who reported using phone apps ≥2 times per day had 34% lower odds of PrEP nonadherence at 3 months (OR 0.66, 95% CI 0.17-2.88). Participants who spent >7 hours on the internet beyond work or school had 2.31 (95% CI 0.80-6.46) times higher odds of PrEP nonadherence at 3 months compared to participants who reported <7 hours on the internet beyond work or school. Both of these measures that aimed to describe broad patterns of phone and internet use had a relatively large effect on the odds of PrEP nonadherence. However, neither was statistically significant as the 95% CI covers the null value of 1. Participants who reported at least 1 disconnect from their internet service or phone in the past year had 3.84 (95% CI 1.14-12.81) times higher odds of PrEP nonadherence at 3 months than participants who reported no disconnects. Participants who reported that they frequently used phone apps to track their personal health information had 71% (OR 0.29, 95% CI 0.06-0.96) lower odds of PrEP nonadherence at 3 months. Participants who reported depressive symptoms (PHQ-8 score ≥10) had 3.42 (95% CI 0.95-12) times higher odds of PrEP nonadherence at 3 months. Participants who reported anxious symptoms also had 3.51 (95% CI 1.06-11.55) times higher odds of PrEP nonadherence at 3 months. Participants who reported their race and ethnicity as non-Hispanic White had 83% (OR 0.17, 95% CI 0.05-0.48) lower odds of PrEP nonadherence at 3 months. For each additional year of age, participants had 18% (OR 0.82, 95% CI 0.65-1.02) less odds of PrEP nonadherence at 3 months. The 95% CI for the relationship between age and PrEP nonadherence narrowly overlaps the null value of 1 with a corresponding *P* value of .08, meaning this relationship is technically statistically insignificant.

Mediation Analysis

Total effects were decomposed into direct and indirect effects of intrapersonal measures on effective engagement (Table 3).

XSL•FO
RenderX

Experiencing disconnects in the past year was primarily directly related to higher odds of PrEP nonadherence at 3 months (PM=5%), with a direct effect of 3.28 (95% CI 0.91-11.42). Despite the statistically insignificant total effect, using phone apps ≥2 times per day was significantly indirectly associated with lower odds of PrEP nonadherence at 3 months (OR 0.76, 95% CI 0.49-1.00). Conversely, spending >7 hours on the internet beyond work or school resulted in primarily a direct relationship with lower odds of nonadherence (1.92, 95% CI=0.80-5.42; PM=18%). However, similar to the total effect on PrEP nonadherence, this relationship did not rise to the level of statistical significance for either the direct or indirect relationship. Frequently using phone apps to track health information was directly associated (0.3, 95% CI 0-0.92; PM=1%) with lower odds of PrEP nonadherence. Symptoms of depression had a statistically significant total effect on PrEP nonadherence and were partially mediated by engagement with P3 with a direct effect of 2.52 (95% CI 0.79-6.81), an indirect effect of 1.25 (95% CI 0.78-2.14), and a PM of 30%. However, neither the direct nor the indirect effect raised to the level of statistical significance on their own. Experiencing anxious symptoms was primarily indirectly related to higher odds of PrEP nonadherence through engagement with P3 (1.55, 95% CI 1-3.34; PM=51%). Being non-Hispanic White was directly related to lower odds of PrEP nonadherence (0.20, 95% CI 0.04-0.59; PM=4%). For each additional year of age, the odds of PrEP nonadherence were decreased, operating through both direct (0.95, 95% CI 0.68-1.02) and indirect (0.92, 95% CI 0.59-1.07; PM=60%) relationships.

Table 3. Direct and indirect effects of intrapersonal measures on effective engagement with Prepared, Protected, Empowered (P3), a pre-exposure prophylaxis adherence digital health intervention, among US young sexual and gender minority men who have sex with men youth aged between 16 and 24 years (n=140)[a].

	Direct effect (95% CI)[b]	Indirect effect (95% CI)[b]	Percentage mediated (%)[c]
High internet users	1.92 (0.80-5.42)	1.11 (0.84-1.55)	18
Disconnect from the phone in the past 12 months	3.28 (0.91-11.42)	1.04 (0.74-1.84)	5
Frequent phone app user	1.19 (0.37-6.03)	0.76 (0.49-1.00)	—[d]
Frequently uses of phone apps			
Chatting with family	1.23 (0.48-3.19)	0.83 (0.59-1.03)	—
Finding romantic dates	0.63 (0.19-2.25)	1.01 (0.80-1.16)	—
Looking for casual sex	0.74 (0.15-2.02)	1.05 (0.81-1.37)	—
Tracking health	0.31 (0-0.92)	0.98 (0.71-1.33)	1
Depressive symptoms	2.52 (0.79-6.81)	1.25 (0.78-2.14)	30
Anxiety symptoms	2.12 (0.58-5.49)	1.55 (1-3.34)	51
Non-Hispanic White[e]	0.20 (0.04-0.59)	0.84 (0.46-2.54)	4
Intervention arm: P3+[e]	1.34 (0.54-3.62)	0.84 (0.58-1.04)	—
Age[e]	0.95 (0.68-1.02)	0.92 (0.59-1.07)	60

[a]Models are adjusted for age, race/ethnicity, intervention arm, dollars accrued at 3 months, and dollars accrued at 3 months.

[b]Null value is 1.

[c]Percentage mediated cannot be calculated when direct and indirect effects are in opposite directions.

[d]Not applicable.

[e]Intrapersonal measure is also a control measure. age, race, ethnicity, intervention arm, dollars accrued at 3 months, and dollars accrued at 3 months and the interaction between dollars accrued at 3 months and the focal intrapersonal measure.

Discussion

Principal Findings

Overview

This study leveraged data from the primary RCT testing the efficacy of P3, a digital PrEP adherence intervention, in a secondary data analysis that used the effective engagement framework by Yardley et al [25] to characterize whether and how engagement with P3 mediated the relationship between baseline intrapersonal measures and PrEP nonadherence at 3 months. This study found that several measures (eg, twice or more daily phone app use) were positively related to engagement, measured as dollars accrued by 3 months. In contrast, measures such as anxious symptoms were negatively related to engagement. Furthermore, this study found that P3 engagement, behavioral patterns of phone and app use, mental health symptoms, and sociodemographic measures were significantly related to PrEP nonadherence. Using causal mediation analysis, this study decomposed these total effects into direct effects to isolate the effect of each intrapersonal measure on PrEP adherence irrespective of P3 engagement and indirect effects to evaluate if and how each measure may be contributing to effective engagement with P3. This process helps to illuminate possible mechanisms that precipitate or protect against susceptibility to PrEP nonadherence.

XSL•FO
RenderX

Phone and Phone App Use

Digital literacy, defined as an understanding of how technology and digital media are used to communicate with others, has been linked to engagement in DHIs in both HIV and non–HIV-related domains [34,39-49,73-75,95]. Participant's propensity to use phone apps ≥2 times daily may or may not be a direct reflection of their digital literacy. However, this broader representation of their affinity to use phone apps empirically impacted their engagement with P3 (via dollars accrued at 3 months). Furthermore, despite a statistically insignificant total effect on PrEP nonadherence, the moderate to large reduction in odds of PrEP nonadherence combined with the statistically significant indirect effect from the causal mediation analysis suggest that participants' affinity to use phone apps could be a facilitator of effective engagement in the context of PrEP adherence DHIs. Conversely, participants who were categorized as high internet users had higher odds of PrEP nonadherence at 3 months, but this effect was not statistically significant. Furthermore, these participants did not significantly engage with P3 more or less than the average participant. The combination of these contrasting findings suggests that a minimum affinity for phone apps may be related to effective engagement, but time spent on the internet is likely not related to effective engagement. This aligns with theories of digital literacy which describe literacy as more of a minimum capacity to use and understand technology as opposed to merely time spent using it [95,96]. Furthermore, these results substantiate the recent trend of constructing validated and reliable scales of digital literacy [97-101]. Studies of effective engagement (such as this study) would benefit greatly from a scale that can measure and test relevant core constructs of digital literacy against effective engagement.

While this study did not have the opportunity to implement a reliable and validated scale of digital literacy, several measures captured more specific patterns of behavior with respect to phone apps, including using phone apps for dates, tracking health information, or chatting with one's family. Of these, frequently using phone apps to track health information significantly reduced the odds of PrEP nonadherence at 3 months. However, this effect was overwhelmingly a direct effect and participants who frequently use phone apps to track health information did not engage with P3 significantly more than average. This suggests that individuals who are prone to using apps for health tracking may be more health-conscious, independent of app use, and therefore more likely to adhere to PrEP. This aligns with the idea of health-specific digital literacy, sometimes referred to as "eHealth literacy," defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [95]. Previous work found that YSGMMSM with a high digital health literacy perceived the information of a DHI aimed to promote HIV and sexually transmitted infection testing as more useful when the intervention tailored information to the participants [34]. This study supports the hypothesis that health-related digital literacy is an important measure of DHI efficacy, especially when the content is tailored to the population the DHI is serving. Furthermore, this study suggests that while digital health literacy likely improves health outcomes through DHIs, it does not operate through an increase in engagement with the intervention itself. Instead, the baseline capacity for digital health literacy seems to act as a catalyst for the participant to incorporate the information presented in the DHI into their life.

Participants who experienced disconnects from their phone in the past year had higher odds of being nonadherent to PrEP at 3 months, overwhelmingly operating as a direct effect. Participants who experienced disconnects did not earn significantly more or less money than average throughout the trial period. This suggests that disconnecting from one's internet service is not a key engagement barrier. Instead, this measure likely reflects the broader social and structural environment in which a given participant exists. This aligns with previous work that has highlighted the difficulties in adapting DHIs to varying infrastructure levels (eg, low internet connectivity) [38,40-42,45,48,50-52,72,102-105]. Similarly, non-Hispanic White participants had significantly lower odds of being PrEP nonadherent at 3 months and earned significantly more money than average throughout the trial period. However, the effect this had on PrEP nonadherence was also largely direct, suggesting that despite the increase in superficial engagement did not drive the lower odds of PrEP nonadherence at 3 months. Therefore, it seems more likely that non-Hispanic White individuals are experiencing fewer social and structural barriers in life external to the intervention, which affords an easier adoption of adherence behaviors. Previous work reinforces this hypothesis, as significant adherence disparities have been found among Black patients relative to White patients in non-DHI settings [106-108]. Furthermore, previous research has also established that HIV disproportionately affects individuals who are economically disadvantaged [109]. Collectively, the results of this study combined with this body of literature suggest that measures of race and phone disconnects in this study represent structural characteristics that impact PrEP adherence DHI efficacy directly (ie, not through engagement). The implications of this finding align with a systematic review of qualitative studies on engagement conducted by O'Connor et al [38], which describes several recommendations for future DHI development and implementation. First, the systematic review by O'Connor et al [38] highlights the need for DHI developers to lessen the burden of self-care through DHIs. This aligns with our study, where the preponderance of direct effects suggests several mechanisms for tailoring that do not operate through an increase in DHI use. For example, DHIs may be able to adapt intervention elements to low internet connectivity environments (eg, allow participants to download any video content so that it is viewable offline). While this may not increase the engagement levels of those living in low internet connectivity environments to a level significantly above the average, it may allow those participants to interact with the DHI more meaningfully by consuming DHI content uninterrupted during an optimal time for the participant. Second, the systematic review recommends incorporating interpersonal relationships (eg, family, friends, and care providers) and public health institutions in designing, using, and implementing DHIs to mitigate the effects of structural disparities [38]. These recommendations align with the effective engagement framework by Yardley et al [25], which describes the increased

XSL•FO
RenderX

need for tailored approaches to consider the "contextual needs" of DHI participants in addition to traditional tailoring approaches centered on individual sociodemographic characteristics.

Mental Health

While past literature has highlighted the impact of DHIs on improving mental health outcomes, to our knowledge, this is the first study that directly measures the impact of baseline mental health symptoms on effective engagement in a study focused on a nonmental health–related outcome using a digital intervention [110,111]. A meta-analysis by DiMatteo et al [83] examined the link between symptoms of depression and anxiety on treatment adherence in a range of medical conditions (eg, cancer and asthma) and found that symptoms of depression were consistently linked to lower treatment adherence, while the relationship between anxious symptoms and lower treatment adherence were mixed (either small or null). However, Sundbom and Bingefors [84] found that symptoms of depression and anxiety were both linked to medication nonadherence in a more recent population study of Swedish adults. In this study, we similarly found that depressive and anxious symptoms increased the odds of PrEP nonadherence at 3 months in multivariate models. Despite both measures having significant total effects, only anxious symptoms were significantly indirectly related to PrEP nonadherence through lowered engagement. Broadly, there is a need for further research exploring the relationship between mental health conditions, DHI engagement, and PrEP adherence.

DiMatteo et al [83] suggest several hypotheses to explain the relationship between depression and lower treatment adherence. They note that symptoms of depression (depressed mood, feelings of hopelessness, diminished interest or pleasure in activities, sleep disturbances, and diminished ability to concentrate) may directly influence treatment adherence [78,79,83]. For example, if participants with depression are having difficulties concentrating, they may find it more difficult to remember to take PrEP every day as prescribed. They also suggest that depression may be related to social isolation and that social support may be related to better adherence. While depression has been linked to social isolation in young adults broadly [112-114], the literature examining the relationship between social support and PrEP adherence has shown mixed results across studies, where varying sources and kinds of social support show differing effects [115]. Furthermore, the provision of social support was one of the putative mechanisms in the P3 DHI. Future research may consider exploring how participants with symptoms of depression engage with specific features of DHIs to explore mechanisms for how depression impacts PrEP nonadherence. For example, participants with depression may only be engaging with the instrumental support modules (eg, medication tracker) or gamification elements (eg, quests) and avoiding the social elements (eg, social wall). Conversely, the social wall may not be enough or the right kind of social support, despite heavy engagement among participants with depression.

The impact symptoms of anxiety have on DHI engagement and efficacy is unclear. DiMatteo et al [83] note that the range in effect sizes and significance found in their meta-analysis may align with the large degree of heterogeneity in anxiety disorders. For example, symptoms associated with a panic disorder may be quite different from symptoms associated with obsessive-compulsive disorder. In this study, we measured symptoms associated with generalized anxiety disorder using the GAD-7 questionnaire, which measures symptoms of restlessness, feeling on edge, or irritability [80,116]. It is plausible that a generalized feeling of irritability or restlessness could negatively impact one's capability to complete daily tasks in P3. Furthermore, Sundbom and Bingefors [84] found that among men with anxiety, the stated reason for medication nonadherence with the largest effect size was fear of potential adverse drug reactions (OR 3.07, 95% CI 1.55-6.08). However, fearing an adverse drug reaction does not seem like a likely mechanism that can explain the lowered engagement among participants with anxiety using P3, as using a DHI has few, if any, documented risks or adverse reactions. This represents a largely unexplored dimension of DHI engagement research. Like depression, future work on DHI engagement and efficacy should assess the relationship between symptoms of anxiety disorders and specific DHI modules. This will help elucidate the mechanisms that diminish engagement among those with symptoms consistent with generalized anxiety disorder. Furthermore, while the GAD-7 questionnaire does not explicitly screen for social anxiety disorder, there is significant symptom overlap between these 2 conditions and a high degree of comorbidity, with previous clinical trial findings revealing that 57% to 77% of individuals aged between 7 and 17 years with generalized anxiety disorder also have social anxiety disorder [116-119]. One possible hypothesis to explore in future work is how social anxiety impacts engagement in DHIs with a heavy emphasis on social interaction (such as P3). A 2020 global study of social anxiety rates found that 58% of US individuals aged between 18 and 24 years had scores on the Social Anxiety Scale consistent with social anxiety (≥29) [120]. Due to the preponderance of individuals aged between 18 and 24 years with symptoms consistent with social anxiety, future work that aims to develop DHIs for adolescents and young adults should consider further explicating the role of social anxiety in DHI engagement and efficacy.

Strengths and Limitations

This study has several methodological and analytical strengths. The research design of the primary RCT is complementary to the theoretical constructs of effective engagement and causal mediation analysis. RCTs provide clear temporality, which is simultaneously necessary to establish effective engagement (ie, engagement leading to a downstream health outcome) and causal mediation analysis [25,57-60]. This study demonstrates the clarity causal mediation analysis provides to engagement research by disentangling total effects into direct and indirect effects, allowing a causal and mechanistic characterization of how baseline intrapersonal measures relate to effective engagement. This approach helps to avoid mischaracterizations that can occur with traditional regression techniques, as these techniques often model engagement as the outcome and rely on the assumption that an increase in engagement would precipitate an increase in the behavior change of interest. For example, quantitative research into medication adherence disparities has found disparities exist among racial or ethnic

XSL•FO
RenderX

groups after adjusting for other socioeconomic confounders [107], and qualitative research into engagement with DHIs has largely lacked a focus on race and ethnicity [38], which suggests that race and ethnicity might play a role in effective engagement for adherence DHIs. In this study, non-Hispanic White participants earned significantly more money than other trial participants (US $17) 3 months into the trial period. However, effect decomposition demonstrated that the effect of race and ethnicity on PrEP nonadherence was largely direct, suggesting that race and ethnicity is related to PrEP adherence irrespective of the observed increase in dollars accrued (ie, engagement) among non-Hispanic White participants. This example illustrates how traditional regression and causal mediation approaches would arrive at divergent conclusions based on the same data. Traditional approaches assume engagement leads to a behavior change, whereas causal mediation analysis estimates behavior change based on changes in engagement. In this example, given the results from the causal mediation analysis, it seems substantially more likely that reported non-Hispanic White race and ethnicity is a proxy for the systematic inequalities people of sexual, gender, racial, and ethnic minority face in the United States and that these inequalities propagate difficulties with PrEP adherence in ways P3 cannot or did not address.

One limitation of this study is the small and selective sample size. The small sample size limits the power to detect effects. This means that some of the null results in this study may in fact be more significant in a study with a larger sample. However, this also means that this study is likely a conservative measure of effective engagement. The significant results from this study which have relatively small effect sizes may in truth be much larger. Furthermore, 13.5% (22/163) of the eligible participants were LTFU, which is consistent with the primary P3 efficacy RCT, which observed 13% LTFU across all intervention conditions. Due to this, we believe that participants were not LTFU for reasons specific to study operations or intervention conditions. Similarly, the sample of individuals in the primary RCT is relatively homogenous (small age range, relatively high digital literacy, generally nonrural, and same sexual orientation). Therefore, the results of this study may or may not be generalizable to other populations. However, while this feature of the primary RCT limits generalizability, it also strengthens confounding assumptions necessary to carry out causal mediation analysis. By having a more homogenous sample, many of the measures that may have been conceived as confounders have been controlled for through the primary RCT's study design, such as sexual orientation and English literacy. Furthermore, while several aspects of the population are homogenous, there is a high degree of geographic diversity as the primary RCT was carried out at 9 study sites. Another limitation is the degree of missingness in the biological measures of adherence. Before the COVID-19 pandemic, participants completed study activities in person at study sites, which included the collection of biological specimens by study site staff. Once COVID-19 restrictions were in place and study activities resumed, most activities were completed remotely to enable the continuation of data collection. Participants were asked to complete the at-home dried blood spot collection. The change from site-directed biological specimen collection to dried blood spot self-collection likely impacted the amount of missing biological specimen data. However, self-report measures are accurate for estimating protective serum levels of PrEP among adults [69,70] and the likelihood of incorrectly estimating protective serum levels of PrEP decreases significantly with age among adolescents and young adults [71]. Given that the average age of participants was 22 years in this study, combined with the relatively high area under the receiver operating characteristics curve (≥0.7) between self-report measures and biological measures among participants without missing biological measures, using self-report PrEP adherence measures where biological measures are missing seems sufficient. Finally, the inclusion of a validated digital literacy scale and social anxiety–specific scale (as opposed to only a scale of generalized anxiety disorder) would have been ideal to measure digital literacy and social anxiety, respectively.

Conclusions

This study used a causal mediation approach using secondary data from an RCT testing the efficacy of P3, a digital PrEP adherence intervention. This study combined data from the primary RCT, including biological PrEP adherence measures, with engagement data, to characterize how baseline intrapersonal measures relate to effective engagement in participants who received P3. Broadly, P3 engagement (dollars accrued) was strongly related to lower odds of PrEP nonadherence. Specifically, this study identified digital literacy as a potential engagement facilitator and measures of structural disparity (eg, disconnection from phone or internet in the past year) and mental health (eg, anxious symptoms) as engagement barriers. Study results suggest tailoring as a critical DHI mechanism to address barriers to engagement and emphasize engagement facilitators in indicated individuals. Furthermore, these findings highlight the suitability of causal mediation analysis for effective engagement research by delineating the total effect of each intrapersonal measure into direct and indirect effects (effective engagement). Future research into effective engagement would benefit from adopting a causal mediation approach. Furthermore, as hypotheses regarding exact mechanisms for fostering engagement arise, future research should measure engagement with measure-specific areas of the DHI as a mediator.

Acknowledgments

The authors would like to collectively acknowledge and thank the Prepared, Protected, Empowered (P3) study team, P3 study participants, iTech, and the Adolescent Medicine Trials Network for HIV Interventions for their continued support and contributions to this project. MPW is supported by the National Institutes of Health: Eunice Kennedy Shriver National Institute of Child Health and Human Development (5F31HD108046) and National Institute on Aging (1T32AG081327). DFH is supported by the National Institutes of Health: National Institute on Drug Abuse (5K01DA046307).

XSL•FO
RenderX

Authors' Contributions

MPW and DFH conceived and designed the study. LBH-W conceived and designed the primary study (Prepared, Protected, Empowered [P3]). MPW and the P3 study team, namely CBR, constructed the analytic dataset. LHS, JM, and MPW built the statistical models. All authors revised the manuscript and interpreted the data.

Conflicts of Interest

None declared.

References

1. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis 2013 Mar;13(3):214-222. [doi: 10.1016/S1473-3099(12)70315-8] [Medline: 23260128]
2. HIV and young men who have sex with men. Centers for Disease Control and Prevention. URL: https://tinyurl.com/2azvbmue [accessed 2023-04-12]
3. HIV surveillance report, 2018 (updated). Centers for Disease Control and Prevention. 2018. URL: https://tinyurl.com/3968pb23 [accessed 2024-12-17]
4. Men who have sex with men (MSM). Centers for Disease Control and Prevention. URL: https://tinyurl.com/yhrf6c64 [accessed 2023-04-12]
5. Transgender and gender diverse persons. Centers for Disease Control and Prevention. URL: https://tinyurl.com/4pe2snn7 [accessed 2023-04-12]
6. Clark H, Babu AS, Wiewel EW, Opoku J, Crepaz N. Diagnosed HIV infection in transgender adults and adolescents: results from the National HIV Surveillance System, 2009-2014. AIDS Behav 2017 Sep 29;21(9):2774-2783 [FREE Full text] [doi: 10.1007/s10461-016-1656-7] [Medline: 28035497]
7. Hall HI, Walker F, Shah D, Belle E. Trends in HIV diagnoses and testing among U.S. adolescents and young adults. AIDS Behav 2012 Jan 12;16(1):36-43. [doi: 10.1007/s10461-011-9944-8] [Medline: 21484282]
8. HIV surveillance report, 2020. Centers for Disease Control and Prevention. 2022 May. URL: https://tinyurl.com/28s94jhy [accessed 2023-09-05]
9. Buchbinder SP, Glidden DV, Liu AY, McMahan V, Guanira JV, Mayer KH, et al. HIV pre-exposure prophylaxis in men who have sex with men and transgender women: a secondary analysis of a phase 3 randomised controlled efficacy trial. Lancet Infect Dis 2014 Jun;14(6):468-475 [FREE Full text] [doi: 10.1016/S1473-3099(14)70025-8] [Medline: 24613084]
10. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
11. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829 [FREE Full text] [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
12. Callander D, McManus H, Gray RT, Grulich AE, Carr A, Hoy J, et al. HIV treatment-as-prevention and its effect on incidence of HIV among cisgender gay, bisexual, and other men who have sex with men in Australia: a 10-year longitudinal cohort study. The Lancet HIV 2023 Jun;10(6):e385-e393. [doi: 10.1016/s2352-3018(23)00050-4]
13. LeVasseur MT, Goldstein ND, Tabb LP, Olivieri-Mui BL, Welles SL. The effect of PrEP on HIV incidence among men who have sex with men in the context of condom use, treatment as prevention, and seroadaptive practices. J Acquir Immune Defic Syndr 2018 Jan 01;77(1):31-40. [doi: 10.1097/QAI.0000000000001555] [Medline: 28961679]
14. Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, Stephenson R. A multilevel tailored web app-based intervention for linking young men who have sex with men to quality care (get connected): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Aug 02;7(8):e10444 [FREE Full text] [doi: 10.2196/10444] [Medline: 30072358]
15. Biello KB, Marrow E, Mimiaga MJ, Sullivan P, Hightow-Weidman L, Mayer KH. A mobile-based app (MyChoices) to increase uptake of HIV testing and pre-exposure prophylaxis by young men who have sex with men: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Jan 07;8(1):e10694 [FREE Full text] [doi: 10.2196/10694] [Medline: 30617042]
16. Garcia J, Parker C, Parker RG, Wilson PA, Philbin M, Hirsch JS. Psychosocial implications of homophobia and HIV stigma in social support networks: insights for high-impact HIV prevention among Black men who have sex with men. Health Educ Behav 2016 Apr;43(2):217-225 [FREE Full text] [doi: 10.1177/1090198115599398] [Medline: 27037286]
17. Harawa NT, Bempong GA, Fields S, Leibowitz A, Lee D, Meyer P, et al. Engaging Black men who have sex with men (BMSM) in Los Angeles in HIV preExposure Prophylaxis (PrEP). California HIV/AIDS Policy. 2016 Feb. URL: https://chipts.ucla.edu/wp-content/uploads/2016/02/White-Paper-Black-MSM-and-PrEP.pdf [accessed 2024-12-17]
18. Liu A, Coleman K, Bojan K, Serrano PA, Oyedele T, Garcia A, et al. Developing a mobile app (LYNX) to support linkage to HIV/sexually transmitted infection testing and pre-exposure prophylaxis for young men who have sex with men: protocol

XSL•FO
RenderX

for a randomized controlled trial. JMIR Res Protoc 2019 Jan 25;8(1):e10659 [FREE Full text] [doi: 10.2196/10659] [Medline: 30681964]
19. LeGrand S, Knudtson K, Benkeser D, Muessig K, Mcgee A, Sullivan PS, et al. Testing the efficacy of a social networking gamification app to improve pre-exposure prophylaxis adherence (P3: Prepared, Protected, emPowered): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 18;7(12):e10448 [FREE Full text] [doi: 10.2196/10448] [Medline: 30563818]
20. Songtaweesin WN, LeGrand S, Bandara S, Piccone C, Wongharn P, Moonwong J, et al. Adaptation of a theory-based social networking and gamified app-based intervention to improve pre-exposure prophylaxis adherence among young men who have sex with men in Bangkok, Thailand: qualitative study. J Med Internet Res 2021 Nov 04;23(11):e23852 [FREE Full text] [doi: 10.2196/23852] [Medline: 34734828]
21. Chiasson MA, Hirshfield S, Rietmeijer C. HIV prevention and care in the digital age. J Acquir Immune Defic Syndr 2010 Dec;55 Suppl 2:S94-S97. [doi: 10.1097/QAI.0b013e3181fcb878] [Medline: 21406996]
22. Kirby T, Thornber-Dunwell M. Phone apps could help promote sexual health in MSM. Lancet 2014 Oct 18;384(9952):1415. [doi: 10.1016/S0140-6736(14)61849-3] [Medline: 25390311]
23. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for Black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
24. Sullivan PS, Grey JA, Simon Rosser BR. Emerging technologies for HIV prevention for MSM: what we have learned, and ways forward. J Acquir Immune Defic Syndr 2013 Jun 01;63 Suppl 1(0 1):S102-S107 [FREE Full text] [doi: 10.1097/QAI.0b013e3182949e85] [Medline: 23673879]
25. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
26. Brothers J, Hosek S, Keckler K, Anderson PL, Xiong D, Liu H, et al. The ATEAM study: advances in technology to enhance PrEP adherence monitoring (ATEAM) among young men who have sex with men. Clin Transl Sci 2022 Dec;15(12):2947-2957 [FREE Full text] [doi: 10.1111/cts.13414] [Medline: 36106611]
27. Buchbinder SP, Siegler AJ, Coleman K, Vittinghoff E, Wilde G, Lockard A, et al. Randomized controlled trial of automated directly observed therapy for measurement and support of PrEP adherence among young men who have sex with men. AIDS Behav 2023 Feb;27(2):719-732 [FREE Full text] [doi: 10.1007/s10461-022-03805-3] [Medline: 35984607]
28. Chai PR, Mohamed Y, Bustamante MJ, Goodman GR, Najarro J, Castillo-Mancilla J, et al. DigiPrEP: a pilot trial to evaluate the feasibility, acceptability, and accuracy of a digital pill system to measure PrEP adherence in men who have sex with men who use substances. J Acquir Immune Defic Syndr 2022 Feb 01;89(2):e5-15 [FREE Full text] [doi: 10.1097/QAI.0000000000002854] [Medline: 34753871]
29. Serrano VB, Moore DJ, Morris S, Tang B, Liao A, Hoenigl M, et al. Efficacy of daily text messaging to support adherence to HIV pre-exposure prophylaxis (PrEP) among stimulant-using men who have sex with men. Subst Use Misuse 2023;58(3):465-469 [FREE Full text] [doi: 10.1080/10826084.2023.2165409] [Medline: 36659873]
30. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82 [FREE Full text] [doi: 10.2196/mhealth.7336] [Medline: 28630037]
31. Hightow-Weidman LB, Horvath KJ, Scott H, Hill-Rorie J, Bauermeister JA. Engaging youth in mHealth: what works and how can we be sure? Mhealth 2021;7:23 [FREE Full text] [doi: 10.21037/mhealth-20-48] [Medline: 33898592]
32. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. Mhealth 2020 Jan;6:7 [FREE Full text] [doi: 10.21037/mhealth.2019.10.01] [Medline: 32190618]
33. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
34. Horvath KJ, Bauermeister JA. eHealth literacy and intervention tailoring impacts the acceptability of a HIV/STI testing intervention and sexual decision making among young gay and bisexual men. AIDS Educ Prev 2017 Feb;29(1):14-23 [FREE Full text] [doi: 10.1521/aeap.2017.29.1.14] [Medline: 28195779]
35. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4. [doi: 10.1093/her/15.1.1] [Medline: 10788196]
36. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144 [FREE Full text] [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
37. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
38. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

XSL•FO
RenderX

39. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
40. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
41. Flynn D, Gregory P, Makki H, Gabbay M. Expectations and experiences of eHealth in primary care: a qualitative practice-based investigation. Int J Med Inform 2009 Sep;78(9):588-604. [doi: 10.1016/j.ijmedinf.2009.03.008] [Medline: 19482542]
42. Middlemass J, Davy Z, Cavanagh K, Linehan C, Morgan K, Lawson S, et al. Integrating online communities and social networks with computerised treatment for insomnia: a qualitative study. Br J Gen Pract 2012 Dec;62(605):e840-e850 [FREE Full text] [doi: 10.3399/bjgp12X659321] [Medline: 23211265]
43. Hottes TS, Farrell J, Bondyra M, Haag D, Shoveller J, Gilbert M. Internet-based HIV and sexually transmitted infection testing in British Columbia, Canada: opinions and expectations of prospective clients. J Med Internet Res 2012 Mar 06;14(2):e41 [FREE Full text] [doi: 10.2196/jmir.1948] [Medline: 22394997]
44. Winkelman WJ, Leonard KJ, Rossos PG. Patient-perceived usefulness of online electronic medical records: employing grounded theory in the development of information and communication technologies for use by patients living with chronic illness. J Am Med Inform Assoc 2005;12(3):306-314 [FREE Full text] [doi: 10.1197/jamia.M1712] [Medline: 15684128]
45. Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. BMJ 2008 Jun 07;336(7656):1290-1295 [FREE Full text] [doi: 10.1136/bmj.a114] [Medline: 18511764]
46. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]
47. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: a focus group analysis. J Med Internet Res 2011 Jul 14;13(3):e49 [FREE Full text] [doi: 10.2196/jmir.1770] [Medline: 21752785]
48. Hopp FP, Hogan MM, Woodbridge PA, Lowery JC. The use of telehealth for diabetes management: a qualitative study of telehealth provider perceptions. Implement Sci 2007 May 02;2:14 [FREE Full text] [doi: 10.1186/1748-5908-2-14] [Medline: 17475012]
49. Lorimer K, Martin S, McDaid LM. The views of general practitioners and practice nurses towards the barriers and facilitators of proactive, internet-based chlamydia screening for reaching young heterosexual men. BMC Fam Pract 2014 Jun 27;15:127 [FREE Full text] [doi: 10.1186/1471-2296-15-127] [Medline: 24972919]
50. Mac Domhnaill C, Mohan G, McCoy S. Home broadband and student engagement during COVID-19 emergency remote teaching. Distance Educ 2021 Nov 03;42(4):465-493. [doi: 10.1080/01587919.2021.1986372]
51. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
52. Mesch GS, Talmud I. Internet connectivity, community participation, and place attachment: a longitudinal study. Am Behav Sci 2010 Feb 18;53(8):1095-1110. [doi: 10.1177/0002764209356243]
53. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
54. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
55. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
56. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
57. Vansteelandt S, Bekaert M, Lange T. Imputation strategies for the estimation of natural direct and indirect effects. Epidemiol Methods 2012;1(1):130. [doi: 10.1515/2161-962X.1014]
58. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health 2016;37:17-32. [doi: 10.1146/annurev-publhealth-032315-021402] [Medline: 26653405]
59. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol 2010 Dec 15;172(12):1339-1348 [FREE Full text] [doi: 10.1093/aje/kwq332] [Medline: 21036955]
60. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992 Mar;3(2):143-155. [doi: 10.1097/00001648-199203000-00013] [Medline: 1576220]
61. VanderWeele TJ. Explanation in causal inference: developments in mediation and interaction. Int J Epidemiol 2016 Dec 01;45(6):1904-1908 [FREE Full text] [doi: 10.1093/ije/dyw277] [Medline: 27864406]

XSL•FO
RenderX

62. VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York, NY: Oxford University Press; 2015.
63. Oberg AS, VanderWeele TJ, Almqvist C, Hernandez-Diaz S. Pregnancy complications following fertility treatment-disentangling the role of multiple gestation. Int J Epidemiol 2018 Aug 01;47(4):1333-1342 [FREE Full text] [doi: 10.1093/ije/dyy103] [Medline: 29939263]
64. Hernán MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health 2004 Apr;58(4):265-271 [FREE Full text] [doi: 10.1136/jech.2002.006361] [Medline: 15026432]
65. Höfler M. Causal inference based on counterfactuals. BMC Med Res Methodol 2005 Sep 13;5:28 [FREE Full text] [doi: 10.1186/1471-2288-5-28] [Medline: 16159397]
66. P3 (Prepared, Protected, emPowered): promoting pre-exposure prophylaxis (PrEP) adherence through a social networking, gamification, and adherence support app. National Institutes of Health National Library of Medicine. 2022 Oct 25. URL: https://clinicaltrials.gov/ct2/show/NCT03320512 [accessed 2023-05-31]
67. Amico KR, Miller J, Balthazar C, Serrano PA, Brothers J, Zollweg S, et al. Integrated next step counseling (iNSC) for sexual health and PrEPuse among young men who have sex with men: implementation and observations from ATN110/113. AIDS Behav 2019 Jul;23(7):1812-1823 [FREE Full text] [doi: 10.1007/s10461-018-2291-2] [Medline: 30315429]
68. R Amico K, McMahan V, Goicochea P, Vargas L, Marcus JL, Grant RM, et al. Supporting study product use and accuracy in self-report in the iPrEx study: next step counseling and neutral assessment. AIDS Behav 2012 Jul;16(5):1243-1259. [doi: 10.1007/s10461-012-0182-5] [Medline: 22460228]
69. Agot K, Taylor D, Corneli AL, Wang M, Ambia J, Kashuba AD, et al. Accuracy of self-report and pill-count measures of adherence in the FEM-PrEP clinical trial: implications for future HIV-prevention trials. AIDS Behav 2015 May 7;19(5):743-751 [FREE Full text] [doi: 10.1007/s10461-014-0859-z] [Medline: 25100053]
70. Blumenthal J, Pasipanodya EC, Jain S, Sun S, Ellorin E, Morris S, et al. Comparing self-report pre-exposure prophylaxis adherence questions to pharmacologic measures of recent and cumulative pre-exposure prophylaxis exposure. Front Pharmacol 2019;10:721 [FREE Full text] [doi: 10.3389/fphar.2019.00721] [Medline: 31333454]
71. Baker Z, Javanbakht M, Mierzwa S, Pavel C, Lally M, Zimet G, et al. Predictors of over-reporting HIV pre-exposure prophylaxis (PrEP) adherence among young men who have sex with men (YMSM) in self-reported versus biomarker data. AIDS Behav 2018 Apr;22(4):1174-1183 [FREE Full text] [doi: 10.1007/s10461-017-1958-4] [Medline: 29079950]
72. Horvath KJ, Danilenko GP, Williams ML, Simoni J, Amico KR, Oakes JM, et al. Technology use and reasons to participate in social networking health websites among people living with HIV in the US. AIDS Behav 2012 May;16(4):900-910 [FREE Full text] [doi: 10.1007/s10461-012-0164-7] [Medline: 22350832]
73. Payton S, Hague C. Digital literacy across the curriculum. National Foundation for Educational Research. 2010 Jan 1. URL: https://www.nfer.ac.uk/publications/digital-literacy-across-the-curriculum/ [accessed 2023-09-13]
74. Lepore SJ, Rincon MA, Buzaglo JS, Golant M, Lieberman MA, Bauerle Bass S, et al. Digital literacy linked to engagement and psychological benefits among breast cancer survivors in internet-based peer support groups. Eur J Cancer Care (Engl) 2019 Jul;28(4):e13134 [FREE Full text] [doi: 10.1111/ecc.13134] [Medline: 31318132]
75. Melhem SJ, Nabhani-Gebara S, Kayyali R. Digital trends, digital literacy, and e-health engagement predictors of breast and colorectal cancer survivors: a population-based cross-sectional survey. Int J Environ Res Public Health 2023 Jan 13;20(2):1472 [FREE Full text] [doi: 10.3390/ijerph20021472] [Medline: 36674237]
76. Im EO, Lee B, Chee W. Shielded from the real world: perspectives on internet cancer support groups by Asian Americans. Cancer Nurs 2010;33(3):E10-E20 [FREE Full text] [doi: 10.1097/NCC.0b013e3181c8e5d5] [Medline: 20357657]
77. Dasgupta K, Da Costa D, Pillay S, De Civita M, Gougeon R, Leong A, et al. Strategies to optimize participation in diabetes prevention programs following gestational diabetes: a focus group study. PLoS One 2013;8(7):e67878 [FREE Full text] [doi: 10.1371/journal.pone.0067878] [Medline: 23861824]
78. Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health 2002 Mar;30(3):196-204. [doi: 10.1016/s1054-139x(01)00333-0] [Medline: 11869927]
79. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
80. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
81. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
82. Yeager CM, Shoji K, Luszczynska A, Benight CC. Engagement with a trauma recovery internet intervention explained with the health action process approach (HAPA): longitudinal study. JMIR Ment Health 2018 Apr 10;5(2):e29 [FREE Full text] [doi: 10.2196/mental.9449] [Medline: 29636323]
83. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000 Jul 24;160(14):2101-2107. [doi: 10.1001/archinte.160.14.2101] [Medline: 10904452]

XSL•FO
RenderX

84. Sundbom LT, Bingefors K. The influence of symptoms of anxiety and depression on medication nonadherence and its causes: a population based survey of prescription drug users in Sweden. Patient Prefer Adherence 2013 Aug 19;7:805-811 [FREE Full text] [doi: 10.2147/PPA.S50055] [Medline: 23983459]
85. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
86. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
87. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010 Jul 27;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]
88. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
89. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009 Feb 23;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
90. Wang Z, Wang J, Maercker A. Program use and outcome change in a web-based trauma intervention: individual and social factors. J Med Internet Res 2016 Sep 09;18(9):e243 [FREE Full text] [doi: 10.2196/jmir.5839] [Medline: 27612932]
91. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
92. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018 Sep 11;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
93. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 2013 Jun;18(2):137-150 [FREE Full text] [doi: 10.1037/a0031034] [Medline: 23379553]
94. Shi B, Choirat C, Coull BA, VanderWeele TJ, Valeri L. CMAverse: a suite of functions for reproducible causal mediation analyses. Epidemiology 2021 Sep 01;32(5):e20-e22. [doi: 10.1097/EDE.0000000000001378] [Medline: 34028370]
95. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
96. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 06;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
97. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022 Apr 29;6(4):e36043 [FREE Full text] [doi: 10.2196/36043] [Medline: 35486413]
98. Amin H, Malik MA, Akkaya B. Development and validation of digital literacy scale (DLS) and its implication for higher education. Int J Distance Educ E Learn 2022 Jan 04;7(1):24-43. [doi: 10.36261/ijdeel.v7i1.2224]
99. Ma S, Wang Y, Shu Z, Duan Z, Sun L. Development and validation of internet literacy scale for high school students. Educ Inf Technol (Dordr) 2023 Apr 03;29:1-28 [FREE Full text] [doi: 10.1007/s10639-023-11641-8] [Medline: 37361779]
100. Reddy P, Chaudhary K, Sharma B, Hussein S. Essaying the design, development and validation processes of a new digital literacy scale. Online Inf Rev 2022 Jul 20;47(2):371-397. [doi: 10.1108/OIR-10-2021-0532]
101. Ali A, Raza AA, Qazi IA. Validated digital literacy measures for populations with low levels of internet experiences. Dev Eng 2023 Nov;8:100107. [doi: 10.1016/j.deveng.2023.100107]
102. Campbell BR, Ingersoll KS, Flickinger TE, Dillingham R. Bridging the digital health divide: toward equitable global access to mobile health interventions for people living with HIV. Expert Rev Anti Infect Ther 2019 Mar;17(3):141-144 [FREE Full text] [doi: 10.1080/14787210.2019.1578649] [Medline: 30721103]
103. Xiong S, Lu H, Peoples N, Duman EK, Najarro A, Ni Z, et al. Digital health interventions for non-communicable disease management in primary health care in low-and middle-income countries. NPJ Digit Med 2023 Feb 01;6(1):12 [FREE Full text] [doi: 10.1038/s41746-023-00764-4] [Medline: 36725977]
104. Easton P, Entwistle VA, Williams B. Health in the 'hidden population' of people with low literacy. A systematic review of the literature. BMC Public Health 2010 Aug 05;10:459 [FREE Full text] [doi: 10.1186/1471-2458-10-459] [Medline: 20687946]
105. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021 Aug 26;23(8):e29575 [FREE Full text] [doi: 10.2196/29575] [Medline: 34435965]
106. Whiteley L, Craker L, Sun S, Tarantino N, Hershkowitz D, Moskowitz J, et al. Factors associated with PrEP adherence among MSM living in Jackson, Mississippi. J HIV AIDS Soc Serv 2021;20(3):246-261 [FREE Full text] [doi: 10.1080/15381501.2021.1956666] [Medline: 35662802]
107. Xie Z, St Clair P, Goldman DP, Joyce G. Racial and ethnic disparities in medication adherence among privately insured patients in the United States. PLoS One 2019 Feb 14;14(2):e0212117 [FREE Full text] [doi: 10.1371/journal.pone.0212117] [Medline: 30763400]

XSL•FO
RenderX

108. Ezennia O, Geter A, Smith DK. The PrEP care continuum and black men who have sex with men: a scoping review of published data on awareness, uptake, adherence, and retention in PrEP care. AIDS Behav 2019 Oct;23(10):2654-2673. [doi: 10.1007/s10461-019-02641-2] [Medline: 31463711]
109. Denning P, Dinenno E. Communities in crisis: is there a generalized HIV epidemic in impoverished urban areas of the United States? In: Proceedings of the XVIII International AIDS Conference. 2010 Presented at: AIDS 2010; July 18-23, 2010; Vienna, Austria URL: https://tinyurl.com/7a6cc8uw
110. Gan DZ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021 Nov 04;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
111. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
112. Elmer T, Stadtfeld C. Depressive symptoms are associated with social isolation in face-to-face interaction networks. Sci Rep 2020 Jan 29;10(1):1444 [FREE Full text] [doi: 10.1038/s41598-020-58297-9] [Medline: 31996728]
113. Ge L, Yap CW, Ong R, Heng BH. Social isolation, loneliness and their relationships with depressive symptoms: a population-based study. PLoS One 2017 Aug 23;12(8):e0182145 [FREE Full text] [doi: 10.1371/journal.pone.0182145] [Medline: 28832594]
114. Matthews T, Danese A, Wertz J, Odgers CL, Ambler A, Moffitt TE, et al. Social isolation, loneliness and depression in young adulthood: a behavioural genetic analysis. Soc Psychiatry Psychiatr Epidemiol 2016 Mar;51(3):339-348 [FREE Full text] [doi: 10.1007/s00127-016-1178-7] [Medline: 26843197]
115. Nacht CL, Reynolds HE, Jessup O, Amato M, Storholm ED. The association between social support and pre-exposure prophylaxis use among sexual minority men in the United States: a scoping literature review. AIDS Behav 2024 Nov;28(11):3559-3573. [doi: 10.1007/s10461-024-04446-4] [Medline: 39039399]
116. Khdour HY, Abushalbaq OM, Mughrabi IT, Imam AF, Gluck MA, Herzallah MM, et al. Generalized anxiety disorder and social anxiety disorder, but not panic anxiety disorder, are associated with higher sensitivity to learning from negative feedback: behavioral and computational investigation. Front Integr Neurosci 2016 Jun 29;10:20 [FREE Full text] [doi: 10.3389/fnint.2016.00020] [Medline: 27445719]
117. Walkup JT, Albano AM, Piacentini J, Birmaher B, Compton SN, Sherrill JT, et al. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. N Engl J Med 2008 Dec 25;359(26):2753-2766 [FREE Full text] [doi: 10.1056/NEJMoa0804633] [Medline: 18974308]
118. Alomari NA, Bedaiwi SK, Ghasib AM, Kabbarah AJ, Alnefaie SA, Hariri N, et al. Social anxiety disorder: associated conditions and therapeutic approaches. Cureus 2022 Dec 19;14(12):e32687 [FREE Full text] [doi: 10.7759/cureus.32687] [Medline: 36660516]
119. Hearn CS, Donovan CL, Spence SH, March S, Holmes MC. What's the worry with social anxiety? Comparing cognitive processes in children with generalized anxiety disorder and social anxiety disorder. Child Psychiatry Hum Dev 2017 Oct;48(5):786-795. [doi: 10.1007/s10578-016-0703-y] [Medline: 27917455]
120. Jefferies P, Ungar M. Social anxiety in young people: a prevalence study in seven countries. PLoS One 2020 Sep 17;15(9):e0239133 [FREE Full text] [doi: 10.1371/journal.pone.0239133] [Medline: 32941482]

Abbreviations

DHI: digital health intervention
GAD-7: Generalized Anxiety Disorder-7
LTFU: lost to follow-up
MSM: men who have sex with men
OR: odds ratio
P3: Prepared, Protected, Empowered
PHQ-8: Patient Health Questionnaire-8
PM: percentage mediated
PrEP: pre-exposure prophylaxis
RCT: randomized controlled trial
YSGMMSM: young sexual and gender minority men who have sex with men

XSL•FO
RenderX

Edited by A Mavragani; submitted 28.02.24; peer-reviewed by A Barnett, W Evans, M Wang; comments to author 10.06.24; revised version received 18.10.24; accepted 13.11.24; published 13.01.25.

Please cite as:
Williams MP, Manjourides J, Smith LH, Rainer CB, Hightow-Weidman LB, Haley DF
Studying the Digital Intervention Engagement–Mediated Relationship Between Intrapersonal Measures and Pre-Exposure Prophylaxis Adherence in Sexual and Gender Minority Youth: Secondary Analysis of a Randomized Controlled Trial
J Med Internet Res 2025;27:e57619
URL: https://www.jmir.org/2025/1/e57619
doi:10.2196/57619
PMID:39804696

©Michael P Williams, Justin Manjourides, Louisa H Smith, Crissi B Rainer, Lisa B Hightow-Weidman, Danielle F Haley. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness and Safety of the TRIO Optimal Health Management Program in Patients With Type 2 Diabetes Mellitus Initiating Basal Insulin Therapy: Prospective Observational Real-World Study

Chenxi Li[1], Dr med; Lixin Guo[2], MEd; Lixin Shi[3], MEd; Li Chen[4], MEd; Liming Chen[5], MEd; Yaoming Xue[6], MEd; Hong Li[7], MEd; Yuzhen Liang[8], MEd; Jing Yang[9], MEd; Weimin Wang[1*], MEd; Dalong Zhu[1*], MEd

[1]Department of Endocrinology, Endocrine and Metabolic Disease Medical Center, Nanjing Drum Tower Hospital, Affiliated Hospital of Nanjing Medical School, Nanjing, China

[2]Department of Endocrinology, Peking Union Medical College, Chinese Academy of Nanjing Medical Sciences, Beijing, China

[3]Department of Endocrinology, Affiliated Hospital of Guiyang Medical University, Guiyang, China

[4]Department of Endocrinology, Qilu Hospital of Shandong University, Jinan, China

[5]NHC Key Laboratory of Hormones and Development, Tianjin Key Laboratory of Metabolic Diseases, Chu Hsien-I Memorial Hospital & Tianjin Institute of Endocrinology, Tianjin Medical University, Tianjin, China

[6]Department of Endocrinology & Metabolism, Nanfang Hospital, Southern Medical University, Guangzhou, China

[7]Department of Endocrinology and Metabolism, The First Affiliated Hospital of Kunming Medical University, Kunming, China

[8]Department of Endocrinology, The Second Affiliated Hospital of Guangxi Medical University, Nanjing, China

[9]Department of Endocrinology, First Hospital of Shanxi Medical University, Taiyuan, China

[*]these authors contributed equally

Corresponding Author:
Dalong Zhu, MEd
Department of Endocrinology, Endocrine and Metabolic Disease Medical Center
Nanjing Drum Tower Hospital, Affiliated Hospital of Nanjing Medical School
321 Zhongshan Road
Nanjing, 210095
China
Phone: 86 13805150781
Fax: 86 83106666
Email: zhudalong@nju.edu.cn

Abstract

Background: Diabetes, a chronic disease necessitating long-term treatment and self-management, presents significant challenges for patients who spend most of their treatment time outside of hospitals. The potential of digital therapeutics for diabetes has garnered recognition from different organizations. Although some prior studies have demonstrated successful reductions in patients' blood glucose levels and body weight through digital diabetes programs, many studies were limited by including patients with prediabetes, including patients treated with mostly premixed insulin, or evaluating user engagement outcomes rather than clinical outcomes. Consequently, limited evidence remains regarding the effectiveness of health management mobile apps specifically designed for patients with type 2 diabetes mellitus (T2DM) initiating basal insulin (BI). Based on this, a data-based and artificial intelligence management system named "TRIO" was developed to provide patients with more personalized intervention methods in stages, in groups, and around the clock. TRIO assists doctors and nurses in achieving better blood glucose controls, truly carries out standardized management around patients, and allows them to have a higher quality of life. TRIO represents the 3 essential pillars in comprehensive diabetes management: physician, nurse, and patient.

Objective: This prospective observational study evaluated the effectiveness and safety of the TRIO optimal health management program for patients with T2DM initiating BI therapy in a real-world setting.

Methods: Patients aged 18-85 years with inadequate glycemic control (baseline hemoglobin A_{1c} [HbA_{1c}] ≥7%) starting BI therapy were enrolled in outpatient and inpatient settings. The study lasted 3 months, with health education and phone-based follow-up assessments. Data collected included patient characteristics, medical history, baseline diabetes conditions, treatment compliance, glycemic control, and safety indicators.

XSL•FO
RenderX

Results: A total of 199,431 patients were included, and 118,134 patients completed the 3-month follow-up between December 1, 2019, and December 31, 2021, involving 574 hospitals in China. The mean baseline HbA_{1c} was 9.2%, the mean duration of diabetes was 7.3 years, and 80.4% (1,59,930/1,98,969) of patients were using BI with oral antihyperglycemic drugs. After the intervention, mean HbA_{1c} decreased by –2.59% from baseline, with 55.6% (28,858/51,912) achieving the target HbA_{1c} level of <7%. Patients who set lower fasting plasma glucose goals (<6.1 mmol/L) showed more significant HbA_{1c} reductions (P<.001) and higher target achievement than those with fasting plasma glucose goals of ≥6.1 mmol/L. Factors such as complications, diabetes duration, and baseline HbA_{1c} levels influenced the magnitude of HbA_{1c} reduction. The presence of complications, shorter diabetes duration, and higher baseline HbA_{1c} were significantly associated with increased hypoglycemia incidence risk (all P<.05).

Conclusions: The TRIO optimal health management program effectively improved glycemic control in patients with T2DM initiating BI therapy. Individualized treatment approaches considering patient characteristics and glycemic goals are vital for optimal outcomes.

(J Med Internet Res 2025;27:e67554) doi:10.2196/67554

KEYWORDS

type 2 diabetes; TRIO optimal health management program; initiating basal insulin therapy; glycemic control; real-world study

Introduction

Background

The prevalence of type 2 diabetes mellitus (T2DM) in China is rapidly increasing due to lifestyle changes and an aging population. As per the 2018 American Diabetes Association criteria, the estimated prevalence of total diabetes and prediabetes among Chinese adults escalated to 12.8% and 35.2%, respectively, between 2015 and 2017 [1]. Despite the wide range of medication options available for antidiabetic treatment, glycemic control rates among patients with T2DM remain suboptimal [2]. A national survey conducted in 2018 revealed that only 32.9% (10,071/30,609) of patients with diabetes received treatment, with only half of them (15,336/30,609, 50.1%) achieving adequate glycemic control [3].

Diabetes, a chronic disease necessitating long-term treatment and self-management, presents significant challenges for patients who spend most of their treatment time outside of hospitals [4]. When lifestyle intervention and oral antidiabetic drugs (OADs) fail to provide optimal control, patients with type 2 diabetes are required to initiate injectable therapies, mostly basal insulin (BI), according to 2020 Chinese guidelines for T2DM management [5].

Previous randomized controlled trials (RCTs) and observational studies have shown the efficacy of BI in controlled trials [6] and real-world settings [7]. Maintaining a delicate equilibrium between achieving optimal blood glucose control and mitigating hypoglycemia risks is pivotal. This involves the appropriate titration of insulin and diligent self-monitoring of blood glucose levels, both of which are integral to sustaining effective glycemic management. Consequently, establishing an optimal diabetes management framework encompassing health education, consistent professional follow-up, and comprehensive self-monitoring tools becomes imperative for effectively managing patients with type 2 diabetes [8]. The lack of comprehensive and patient-centered approaches in current health care systems further compounds the burden of diabetes management. Time constraints and resource availability often limit traditional health education and face-to-face interactions with health care providers. As a result, there is a growing need for innovative solutions to bridge these gaps and provide ongoing support to individuals with type 2 diabetes [9-11]. The emergence of digital tools such as mobile apps and WeChat miniprograms has increased application in diverse therapeutic domains, such as attention-deficit/hyperactivity disorder, cancer, asthma, and insomnia tools to augment patient self-management [12].

Objectives

The potential of digital therapeutics for diabetes has garnered recognition from different organizations, such as the Centers for Disease Control and Prevention and the Digital Therapeutics Alliance [13]. Although some prior studies have demonstrated successful reductions in patients' blood glucose levels and body weight through digital diabetes programs up to a hemoglobin A_{1c} (HbA_{1c}) reduction of 0.49% [14], many studies were limited by including patients with prediabetes [15], including patients treated with mostly premixed insulin [16], or evaluating user engagement outcomes rather than clinical outcomes [17]. Consequently, limited evidence remains regarding the effectiveness of health management mobile apps specifically designed for patients with T2DM initiating BI.

Accordingly, we have developed a personalized health management program combined with artificial intelligence named "TRIO" for initiating BI therapy in patients with type 2 diabetes and aim to evaluate its effectiveness and safety. Unlike conventional acronyms or abbreviations, TRIO does not represent a longer phrase but rather represents the 3 essential pillars in comprehensive diabetes management: physician, nurse, and patient. This program combines traditional health education with a mobile app to enhance diabetes management. Through this evaluation, we aspire to contribute to understanding practical approaches to optimizing glycemic control and promoting patient well-being.

XSL•FO
RenderX

Methods

Study Design and Population

This prospective, 3-month observational program aimed to evaluate the effectiveness and safety of TRIO, an optimal health management program for patients with T2DM initiating BI therapy in a real-world setting. Participants were recruited from outpatient departments and, at discharge, from inpatient departments between December 1, 2019, and December 31, 2021, from 594 hospitals in China. Patients were assessed for their suitability by the following criteria: (1) aged between 18 and 85 years; (2) patients with T2DM who were inadequately controlled by OADs at the time of enrollment (ie, baseline HbA_{1c} level of ≥7%); (3) initiating BI therapy during the program period, meaning they had not used BI within 12 weeks prior to enrollment; (4) absence of mental disorders or communication impairments; and (5) absence of severe illnesses or limitations regarding follow-up. Patients who fulfilled these eligibility requirements were enrolled upon their willingness to provide informed consent. On the first day of enrollment, patients received health education from nurses, including knowledge about diabetes and insulin, psychological support for a healthy life with diabetes, and how to inject and store insulin. In addition, the physicians assisted patients in drawing up a self-management plan and helped them set individualized fasting plasma glucose (FPG) and postprandial glucose (PPG) goals. Patients were also asked to follow the WeChat official account of the TRIO program, through which knowledge about diabetes management would be sent. Follow-up assessments were conducted via phone calls at 1, 2, 4, 8, and 12 weeks. The frequency and follow-up methods were tailored to each patient's FPG level. If the FPG level was 7 mmol/L, a phone call was not scheduled for the next follow-up visit, and only a WeChat message was sent (Figure 1).

Figure 1. The operating process of the TRIO optimal health management program. AI: artificial intelligence; BG: blood glucose; FPG: fasting plasma glucose; HbA_{1c}: hemoglobin A1c; PPG: postprandial glucose.

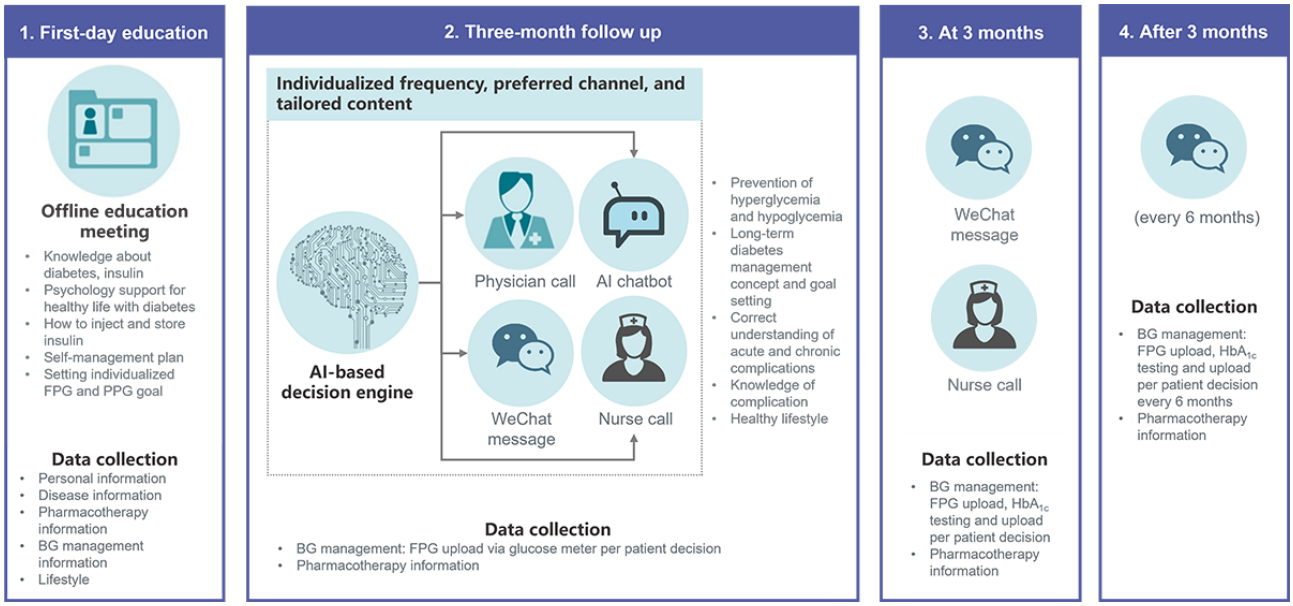

Ethical Considerations

This study was approved by the ethics committee of Nanjing Drum Tower Hospital (institutional review board review approval document, code: 2019-231-01), which was the principal research institute representing other subcenters. Implied consent was obtained from all participants when they registered on the TRIO WeChat official account following the principles of the Declaration of Helsinki, as the privacy policy included a clause allowing anonymized data to be used for research purposes. Participant privacy and anonymity were achieved through the elimination of any patient identifiers such as name or dates of birth, which were anonymized and deidentified before extraction and securely stored in compliance with data protection regulations. No compensation was provided to participants, as this study involved a secondary analysis of existing data. No identifiable images of participants were included in the study or supplementary materials, eliminating the need for additional image consent.

Data Collection

Baseline information was collected by interviews at the hospital enrollment, including demographics, disease characteristics, medical history, physical examination, BI types, starting dosage, and concomitant antidiabetic drugs (bolus insulin, glucagon-like peptide-1 receptor agonists, or OAD) used with BI. Laboratory tests including HbA_{1c} and FPG were obtained in hospitals at baseline, while glycemic control regarding HbA_{1c} and self-monitoring blood glucose (SMBG), including fasting blood glucose (FBG), dosage, and hypoglycemia information during follow-up time, were self-reported. Self-reported data were collected through phone calls by nurses or uploaded via a smart blood glucose device or input into the TRIO WeChat official account by the patients.

Outcomes

- *Primary effectiveness end points*: The primary effectiveness end point of our analysis is the change in HbA_{1c} levels from baseline to month 3.

XSL•FO
RenderX

- *Secondary effectiveness end points*: In addition to the primary end point, we also examined various other measures related to glycemic control. These secondary end points include changes in FBG levels from baseline to month 3, the achievement of target HbA_{1c}levels (<7%), the achievement of target FBG levels (<7 and <6.1 mmol/L), and an assessment of changes in BI dose.
- *Safety end points*: Our safety end points include monitoring and assessing the incidence and rates of hypoglycemia events and evaluating composite end points. The 3-month end point included the incidence and rates of hypoglycemia events and a composite end point encompassing the percentage of patients reaching target HbA_{1c} and FBG levels without experiencing hypoglycemia events.

Statistical Methods

Continuous variables were described using mean and SD, while categorical variables were presented as frequencies and percentages. For continuous effectiveness indicators, a paired *t* test (2-tailed) was applied to test the significance of the change from baseline to month 3 in HbA_{1c} or FPG in the total population. Analysis of covariance was used to compare these changes between subgroups, including patient source (inpatient or outpatient), complication status (no or yes), duration of diabetes (<5 years or ≥5 years), baseline HbA_{1c} (7%-8%, 8%-9%, 9%-10%, or ≥10%), and FBG goal setting (≥6.1 mmol/L or <6.1 mmol/L). Least square (LS) mean (SE) and LS mean difference with 95% CIs were provided. For binary effectiveness outcomes (HbA_{1c} <7%, FBG <7 mmol/L, or FBG <6.1 mmol/L), multivariable logistic regression models were applied to explore the association of subgroups with outcomes and variables included in the model were the same as analysis of covariance model. Hypoglycemic incidence and rate were evaluated by SMBG, uploaded by the smart glucose blood device or manual input to the TRIO platform by patients. Hypoglycemic incidence (percentage of patients with SMBG ≤3.9 mmol/L or SMBG <3.0 mmol/L) was analyzed using logistic regression, and odds ratio (OR) with 95% CI was used as the effect size; hypoglycemic rate (numbers of events per patient-year) was investigated by Poisson regression, and risk ratio (RR) with 95% CI were used as the effect size for this analysis. Composite end points including HbA_{1c} <7% without SMBG ≤3.9 mmol/L, FBG <7 mmol/L without SMBG ≤3.9 mmol/L, and FBG <6.1 mmol/L without SMBG ≤3.9 mmol/L were also explored using logistic regression. All the analyses were conducted using SAS 9.4 (SAS Institute, Inc), and a 2-sided *P* value of <.05 was considered statistically significant.

Results

Participant Recruitment

Between December 1, 2019, and December 31, 2021, a total of 225,764 patients were recruited from 594 hospitals. A total of 26,333 patients were excluded because of violating inclusion or meeting exclusion criteria, leaving 199,431 patients remaining at baseline. Among them, 81,297 patients were lost to follow-up within the first 3 months, resulting in 118,134 patients who completed the 3-month follow-up with measurements of either HbA_{1c} or FPG (Figure 2).

XSL•FO
RenderX

Figure 2. Flowchart of participating patients enrolled in the TRIO optimal health management program. Nested values under "Patients remained at baseline" and "Remained at month 3" are non-mutually exclusive. BI: basal insulin; FPG: fasting plasma glucose; HbA_{1c}: hemoglobin A1c; T1DM: type 1 diabetes mellitus.

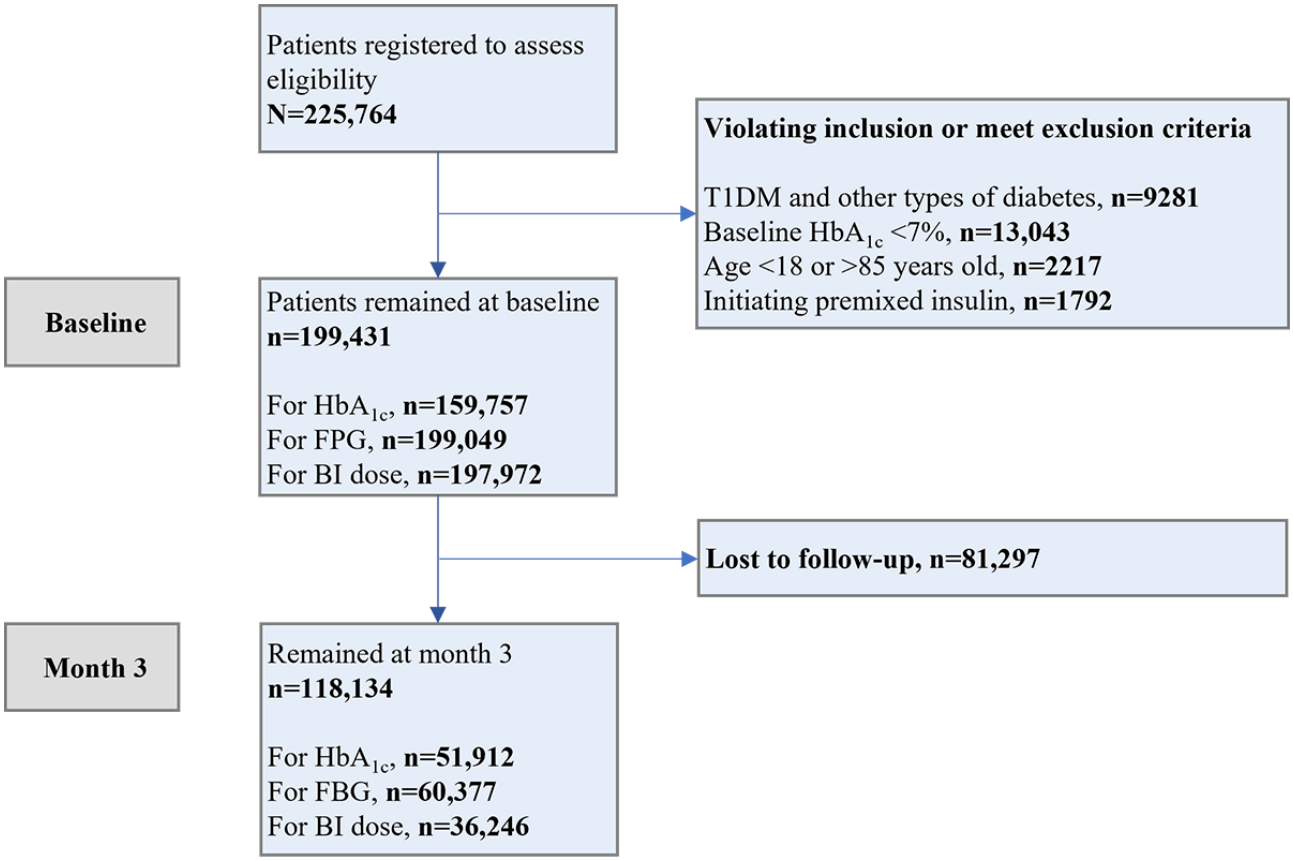

Baseline Characteristics

The patients' mean (SD) age at baseline was 57.3 (12.5) years, with 42.8% (85,337/1,99,431) of participants being women. The average BMI was 24.8 kg/m^2. At baseline, the mean HbA_{1c}, FPG, and PPG levels were 9.6%, 9.5 mmol/L, and 12.7 mmol/L, respectively. The mean duration of diabetes was 7.3 years. The most common complications or comorbidities observed in our patient cohort were hypertension (55,637/1,96,023, 28.4%), hyperlipidemia (32,240/1,96,023, 16.4%), and peripheral neuropathy (63,145/1,96,023, 32.2%). Most patients were taking BI with oral antihyperglycemic drugs (OADs) (1,59,930/1,98,969, 80.4%), with some also using prandial insulin concurrently.

Outpatients had a slightly higher mean age of 57.7 (SD 12.2) years than inpatients, with a mean age of 57.0 (SD 12.6) years. Gender distribution revealed that 56.3% (39,802/70,704) of outpatients were male, while 57.6% (74,064/1,28,645) of inpatients were male. Moreover, the duration of diabetes was slightly longer in inpatients, with a mean of 7.4 years (SD 6.8) than in outpatients, with a mean of 7.0 years (SD 6.3). Both groups displayed a similar average BMI of 24.8 kg/m^2 and exhibited comparable values for blood pressure and lipid levels such as triglycerides, total cholesterol, low-density lipoprotein, and baseline HbA_{1c}. Meanwhile, baseline FPG (10.3 vs 9.0 mmol/L) and PPG (14.0 vs 12.2 mmol/L) were slightly higher in outpatients than in inpatients. A notable difference was observed in the percentage of patients with complications or comorbidities, with 31.3% (39,950/1,27,610) of inpatients having hypertension compared with 22.9% (15,687/68,413) of outpatients and 38.1% (48,580/1,27,610) of inpatients having peripheral neuropathy compared with 21.3% (14,565/68,413) of outpatients. Inpatients also had a higher proportion of patients being treated with BI in combination with prandial insulin, accounting for 20.3% (26,037/1,28,342) of inpatients as opposed to 15.5% (10,924/70,627) of outpatients (Table 1). No clinically significant differences in baseline characteristics were observed between patients who remained in the study at month 3 and those who were lost to follow-up (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Baseline characteristics of patients enrolled in the TRIO optimal health management program.

Baseline characteristics	Outpatient (n=70,786)	Inpatient (n=128,645)	All (N=199,431)
Age (years), mean (SD)	57.7 (12.2)	57.0 (12.6)	57.3 (12.5)
Sex, n (%)			
Male	39,802 (56.3)	74,064 (57.6)	1,13,866 (57.2)
Female	30,902 (43.7)	54,435 (42.4)	85,337 (42.8)
BMI (kg/m^2), mean (SD)	24.8 (3.3)	24.8 (3.5)	24.8 (3.5)
Duration of diabetes (years), mean (SD)	7.0 (6.3)	7.4 (6.8)	7.3 (6.6)
SBP[a] (mm Hg), mean (SD)	131.8 (15.4)	131.8 (16.5)	131.8 (16.2)
DBP[b] (mm Hg), mean (SD)	79.9 (10.1)	79.6 (10.4)	79.7 (10.3)
Triglycerides (mmol/L), mean (SD)	2.3 (2.0)	2.3 (2.1)	2.3 (2.1)
Total cholesterol (mmol/L), mean (SD)	4.6 (1.5)	4.7 (1.5)	4.7 (1.5)
LDL[c] (mmol/L), mean (SD)	2.8 (1.1)	2.8 (1.1)	2.8 (1.1)
BI[d] dose (U/d), mean (SD)	15.2 (5.5)	16.2 (6.1)	15.9 (5.9)
BI dose (U/kg/d), mean (SD)	0.23 (0.08)	0.24 (0.09)	0.24 (0.09)
eGFR[e] (mL/min/1.73 m^2), mean (SD)			
<90	953 (23.6)	3999 (22.0)	4952 (22.3)
90-120	1135 (28.2)	4738 (26.1)	5873 (26.5)
≥120	1942 (48.2)	9424 (51.9)	11,366 (51.2)
Baseline HbA_{1c}[f] (%), mean (SD)	9.3 (1.8)	9.7 (2.1)	9.6 (2.0)
Baseline FPG[g] (mmol/L), mean (SD)	10.3 (3.3)	9.0 (3.2)	9.5 (3.3)
Baseline PPG[h] (mmol/L), mean (SD)	14.0 (4.4)	12.2 (4.0)	12.7 (4.2)
Regimen, n (%)			
BI alone ± OAD[i]	59,176 (83.8)	1,00,754 (78.5)	1,59,930 (80.4)
BI + prandial insulin ± OAD	10,924 (15.5)	26,037 (20.3)	36,961 (18.6)
BI + GLP-1 RA[j] ± OAD	527 (0.7)	1551 (1.2)	2078 (1.0)
Comorbidity, n (%)			
Hypertension	15,687 (22.9)	39,950 (31.3)	55,637 (28.4)
Hyperlipemia	8992 (13.1)	23,248 (18.2)	32,240 (16.4)
Left ventricular hypertrophy	53 (0.08)	149 (0.12)	202 (0.10)
Atrial fibrillation	80 (0.12)	213 (0.17)	293 (0.15)
Complication, n (%)			
Stroke	1409 (2.1)	4511 (3.5)	5920 (3.0)
Coronary heart disease	4601 (6.7)	11,656 (9.1)	16,527 (8.3)
Diabetic nephropathy	3266 (4.8)	11,209 (8.8)	14,475 (7.4)
Diabetic retinopathy	6058 (8.9)	17,329 (13.6)	23,387 (11.9)
Diabetic foot	663 (1.0)	3253 (2.6)	3916 (2.0)
Peripheral neuropathy	14,565 (21.3)	48,580 (38.1)	63,145 (32.2)
Lower extremity angiopathy	2844 (4.2)	10,601 (8.3)	13,445 (6.9)

[a]SBP: systolic blood pressure.

[b]DBP: diastolic systolic blood pressure.

[c]LDL: low-density lipoprotein.

XSL•FO
RenderX

[d]BI: basal insulin.

[e]eGFR: estimated glomerular filtration rate.

[f]HbA_{1c}: hemoglobin A_{1c}.

[g]FPG: fasting plasma glucose.

[h]PPG: postprandial glucose.

[i]OAD: oral antidiabetic drug.

[j]GLP-1 RA: glucagon-like peptide-1 receptor agonists.

Initial Regimens and Adherence to BI Treatment

During the intervention, the majority of patients initiated insulin glargine (1,77,331/1,98,969, 89.1%) as their primary treatment, while a smaller proportion started with insulin determir (6588/1,98,969, 3.3%), neutral protamine Hagedorn insulin (766/1,98,969, 0.4%), or insulin degludec (8939/1,98,969, 4.5%). Alongside BI treatment, 41% (57,242/1,39,739), 28.2% (39,473/1,39,739), and 9.7% (13,545/1,39,739) of patients were concurrently taking 1, 2, and ≥3 OADs, respectively. The most commonly used OADs were metformin (69,027/1,39,739, 49.4%) and α-glucosidase inhibitors (46,960/1,39,739, 33.6%), followed by sodium-glucose cotransporter-2 inhibitors (22,911/1,39,739, 16.4%), dipeptidyl peptidase-4 inhibitors (20,722/1,39,739, 14.8%), sulfonylureas (7266/1,39,739, 5.2%), glinides (6257/1,39,739, 4.5%), and thiazolidinediones (5504/1,39,739, 3.9%). Sulfonylureas were more frequently prescribed to outpatients (3571/48,696, 7.3%) than inpatients (3695/91,043, 4.1%), while sodium-glucose cotransporter-2 inhibitors were more commonly used in inpatients (16,615/91,043, 18.2%) than outpatients (6296/48,696, 12.9%; Table S2 in Multimedia Appendix 1).

Glycemic Outcomes

At the end of the 3-month management intervention period, the mean HbA_{1c} level among 44,847 participants with eligible self-reported mean HbA_{1c} measurements was 6.89% (SD 0.90). This represented a mean decrease in HbA_{1c} by –2.59% (SE 0.01; *P*<.001) from baseline (Table 2). Regarding FBG levels, at the end of month 3, the mean FBG level among 60,365 participants with eligible self-reported FBG measurements was 6.81 (SD 1.4) mmol/L, indicating an average decrease in FBG by –2.77 (SD 0.01) mmol/L (*P*<.001) from baseline (Table 2 and Figure S1 in Multimedia Appendix 1). For the HbA_{1c} target, 55.6% (28,858/51,912) of participants achieved the target HbA_{1c} level of <7% after the 3-month intervention period (Table 3). Similarly, 61.3% (37,017/60,377) and 29.2% (17,633/60,377) of participants reached FBG levels of <7.0 and <6.1 mmol/L, respectively, at the end of the 3-month intervention period.

In the subgroup analyses, patients with complications experienced slightly lower HbA_{1c} (LS mean difference: 0.07, 95% CI 0.05-0.09) than those with no complications. Considering the duration of diabetes, patients with a duration of ≥5 years exhibited a lesser decrease in HbA_{1c} (LS mean difference: 0.09, 95% CI 0.07-0.11) than those with a duration of <5 years. A higher baseline HbA_{1c} level was also associated with a more significant reduction. Compared with patients with baseline HbA_{1c} in the range of 7% to 8%, patients with a baseline HbA_{1c} in the range of 8% to 9% had the lesser decrease (LS mean difference: –0.83, 95% CI –0.86 to –0.79), while patients with a baseline HbA_{1c} of 10% or higher had the highest decrease (LS mean difference: –4.04, 95% CI –4.08 to –4.00). Also, patients with initial FBG goal setting of <6.1 mmol/L had a greater decrease in their HbA_{1c} (LS mean difference: –0.36, 95% CI –0.38 to –0.34) than those with FBG goal setting of ≥6.1 mmol/L (Table 2). In the subgroup analysis exploring factors related to FBG reduction, consistent results with HbA_{1c} reductions found that more significant reductions were seen in patients with no complications, diabetes duration of <5 years, and an initial FBG goal setting of <6.1 mmol/L, except for baseline HbA_{1c}. Patients with lower baseline HbA_{1c} exhibited higher reductions in FBG from baseline (Table 2).

XSL•FO
RenderX

Table 2. HbA_{1c}[a] and fasting blood glucose change from baseline to month 3 after initiation of basal insulin therapy with TRIO monitoring.

	Values, n	Baseline, mean (SD)	Month 3, mean (SD)	LS[b] mean (SE)	*P* value	LS mean difference (95% CI)	*P* value
HbA_{1c} change from baseline to month 3							
All	44,847	9.48 (1.98)	6.89 (0.90)	–2.59 (0.01)	<.001	N/A[c]	
Patient sources							
Outpatient	14,893	9.10 (1.71)	6.93 (0.81)	–2.63 (0.02)	<.001	Reference	
Inpatient	29,954	9.67 (2.07)	6.88 (0.95)	–2.71 (0.01)	<.001	–0.08 (–0.1 to –0.06)	<.001
Complication							
No	17,963	9.33 (1.89)	6.83 (0.85)	–2.71 (0.02)	<.001	Reference	
Yes	20,389	9.58 (2.01)	6.96 (0.93)	–2.64 (0.01)	<.001	0.07 (0.05 to 0.09)	<.001
Duration (years)							
<5	18,950	9.69 (2.14)	6.80 (0.88)	–2.72 (0.02)	<.001	Reference	
≥5	25,888	9.32 (1.83)	6.96 (0.92)	–2.63 (0.01)	<.001	0.09 (0.07 to 0.11)	<.001
Baseline HbA_{1c} (%)							
7-8	9250	7.42 (0.33)	6.78 (0.88)	–0.79 (0.02)	<.001	Reference	
8-9	12,485	8.34 (0.32)	6.86 (0.89)	–1.62 (0.02)	<.001	–0.83 (–0.86 to –0.79)	<.001
9-10	8342	9.35 (0.31)	6.94 (0.86)	–2.54 (0.01)	<.001	–1.75 (–1.79 to –1.70)	<.001
≥10	14,770	11.81 (1.57)	6.97 (0.95)	–4.83 (0.02)	<.001	–4.04 (–4.08 to –4.00)	<.001
FBG[d] goal setting (mmol/L)							
≥6.1	36,935	9.54 (1.98)	6.96 (0.87)	–2.49 (0.01)	<.001	Reference	
<6.1	7904	9.22 (1.96)	6.58 (0.99)	–2.85 (0.02)	<.001	–0.36 (–0.38 to –0.34)	<.001
FBG change from baseline to month 3							
All	60,365	9.58 (3.30)	6.81 (1.40)	–2.77 (0.01)	<.001	N/A	
Patient sources							
Outpatient	21,055	10.35 (3.27)	6.86 (1.39)	–2.61 (0.02)	<.001	Reference	
Inpatient	39,310	9.17 (3.25)	6.79 (1.41)	–2.68 (0.02)	<.001	–0.07 (–0.09 to –0.04)	<.001
Complication							
No	22,927	9.86 (3.16)	6.68 (1.32)	–2.71 (0.02)	<.001	Reference	
Yes	28,252	9.14 (3.17)	6.94 (1.45)	–2.58 (0.02)	<.001	0.13 (0.11 to 0.16)	<.001
Duration (years)							
<5	24,683	9.74 (3.54)	6.57 (1.27)	–2.78 (0.02)	<.001	Reference	
≥5	35,666	9.48 (3.13)	6.98 (1.46)	–2.50 (0.02)	<.001	0.28 (0.25 to 0.3)	<.001
Baseline HbA_{1c} (%)							
7-8	10,005	8.35 (2.13)	6.68 (1.26)	–2.75 (0.02)	<.001	Reference	
8-9	13,472	9.21 (2.48)	6.79 (1.29)	–2.66 (0.02)	<.001	0.08 (0.05 to 0.12)	<.001

XSL•FO
RenderX

	Values, n	Baseline, mean (SD)	Month 3, mean (SD)	LS[b] mean (SE)	*P* value	LS mean difference (95% CI)	*P* value
9-10	9506	9.59 (2.99)	6.85 (1.39)	–2.61 (0.02)	<.001	0.13 (0.09 to 0.17)	<.001
≥10	18,879	10.32 (4.11)	6.83 (1.49)	–2.58 (0.01)	<.001	0.16 (0.13 to 0.20)	<.001
FBG goal setting (mmol/L)							
≥6.1	50,482	9.68 (3.32)	6.87 (1.40)	–2.52 (0.02)	<.001	Reference	
<6.1	9854	9.10 (3.18)	6.53 (1.35)	–2.76 (0.02)	<.001	–0.24 (–0.28 to –0.21)	<.001

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]LS: least squares.

[c]N/A: not applicable.

[d]FBG: fasting blood glucose.

XSL•FO
RenderX

Table 3. Target HbA_{1c}[a] and fasting blood glucose at month 3 after initiation of basal insulin therapy with TRIO monitoring.

	Values, n	Month 3, n (%)	OR[b] (95% CI)	*P* value
HbA_{1c} (<7% at month 3)				
All	51,912	28,858 (55.6)	N/A[c]	
Patient sources				
Outpatient	18,753	10,024 (53.5)	Reference	
Inpatient	33,159	18,834 (56.8)	1.16 (1.11-1.21)	<.001
Complication				
No	20,881	12,433 (59.5)	Reference	
Yes	23,008	12,016 (52.2)	0.85 (0.81-0.89)	<.001
Duration (years)				
<5	21,873	13,388 (61.2)	Reference	
≥5	30,030	15,466 (51.5)	0.83 (0.79-0.87)	<.001
Baseline HbA_{1c} (%)				
7-8	9250	5573 (60.2)	Reference	
8-9	12,485	6974 (55.9)	0.85 (0.8-0.9)	<.001
9-10	8342	4455 (53.4)	0.77 (0.72-0.83)	<.001
≥10	14,770	8009 (54.2)	0.72 (0.68-0.77)	<.001
FBG[d] goal setting (mmol/L)				
≥6.1	43,051	22,857 (53.1)	Reference	
<6.1	8845	5993 (67.8)	1.8 (1.7-1.9)	<.001
FBG (<7 mmol/L at month 3)				
All	60,377	37,017 (61.3)	N/A	
Patient sources				
Outpatient	21,065	12,403 (58.9)	Reference	
Inpatient	39,312	24,614 (62.6)	1.17 (1.12-1.22)	<.001
Complication				
No	22,936	14,952 (65.2)	Reference	
Yes	28,254	16,239 (57.5)	0.84 (0.8-0.88)	<.001
Duration (years)				
<5	24,686	17,053 (69.1)	Reference	
≥5	35,675	19,955 (55.9)	0.70 (0.67-0.73)	<.001
Baseline HbA_{1c} (%)				
7-8	10,006	6607 (66.0)	Reference	
8-9	13,473	8288 (61.5)	0.84 (0.79-0.89)	<.001
9-10	9509	5710 (60.0)	0.79 (0.74-0.85)	<.001
≥10	18,881	11,632 (61.6)	0.77 (0.73-0.82)	<.001
FBG goal setting (mmol/L)				
≥6.1	50,493	30,112 (59.6)	Reference	
<6.1	9855	6886 (69.9)	1.42 (1.34-1.5)	<.001
FBG (<6.1 mmol/L at month 3)				
All	60,377	17,633 (29.2)	N/A	
Patient sources				

XSL•FO
RenderX

	Values, n	Month 3, n (%)	OR[b] (95% CI)	*P* value
Outpatient	21,065	5763 (27.4)	Reference	
Inpatient	39,312	11,870 (30.2)	1.16 (1.1-1.22)	<.001
Complication				
No	22,936	7542 (32.9)	Reference	
Yes	28,254	7087 (25.1)	0.83 (0.79-0.87)	<.001
Duration (years)				
<5	24,686	8973 (36.3)	Reference	
≥5	35,675	8656 (24.3)	0.7 (0.66-0.73)	<.001
Baseline HbA_{1c} (%)				
7-8	10,006	3167 (31.7)	Reference	
8-9	13,473	3764 (27.9)	0.84 (0.79-0.9)	<.001
9-10	9509	2625 (27.6)	0.87 (0.81-0.93)	<.001
≥10	18,881	5746 (30.4)	0.91 (0.85-0.96)	.002
FBG goal setting (mmol/L)				
≥6.1	50,493	13,538 (26.8)	Reference	
<6.1	9855	4084 (41.4)	1.7 (1.62-1.8)	<.001

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]OR: odds ratio.

[c]N/A: not applicable.

[d]FBG: fasting blood glucose.

Regarding target HbA_{1c} of <7% at month 3, patients enrolled from inpatient showed a slight advantage over outpatient (OR 1.16, 95% CI 1.11-1.21; *P*<.001), and patients with complications had lower odds of achieving HbA_{1c} target than those with no complications (OR 0.85, 95% CI 0.81-0.89; *P*<.001). Those with ≥5 years of duration of diabetes had lower success rates than those with <5 years of duration (OR 0.83, 95% CI 0.79-0.87; *P*<.001). Lower baseline HbA_{1c} levels were associated with better outcomes. Moreover, setting FBG target goal level of <6.1 mmol/L at the beginning of the treatment demonstrated a higher possibility of reaching the HbA_{1c} target of <7% at month 3 (OR 1.8, 95% CI 1.7-1.9; *P*<.001) (Table 3). Regarding the FBG target of <7 and <6.1 mmol/L at month 3, patients exhibited consistent results as in HbA_{1c} target of <7%. Inpatients; patients with no complications; and patients with diabetes duration of 5 years, lower HbA_{1c} target, and initial FBG goal setting of <6.1 mmol/L were associated with higher odds of achieving the target (Table 3).

Insulin Dose and Satisfaction

Total insulin dose (U/d/kg) change was –0.01 (SD 0.06), from baseline (mean 0.23, SD 0.09) to month 3 (mean 0.22, SD 0.09; Table 4). Patients recruited from inpatients, with complications, having diabetes duration of ≥5 years, with FBG goal setting of <6.1 mmol/L and higher baseline HbA_{1c}, had a higher starting dose of BI. Among 36,037 patients with both baseline and 3-month BI dosages, 7% (2546/36,037) remained unchanged, and 50.1% (18,047/36,037) lowered the dosage per kilogram (Table 4). Possible reasons for the lack of titration, such as patients reaching FBG targets or experiencing hypoglycemic events, were explored in Table S3 in Multimedia Appendix 1. Patients with stable or decreasing dosage during the 3 months had higher starting doses, higher percentages of FBG <7 mmol/L, and hypoglycemic incidence at week 1, 2, 4, 8, and 12 than patients with increasing dosage. Patient satisfaction level for TRIO was stable during the study. Furthermore, 99.6% (35,738/35,897) of the patients felt satisfactory or very satisfactory at month 3, and only 0.4% (159/35,897) chose average or below.

XSL•FO
RenderX

Table 4. Basal insulin dose (U/kg) change from baseline to month 3 by patient sources, with or with no complication, duration, baseline HbA_{1c}[a] levels, and target fasting plasma glucose levels.

	Values, n	Baseline, mean (SD)	Month 3, mean (SD)	Change, mean (SD)
All	36,037	0.23 (0.09)	0.22 (0.09)	–0.01 (0.06)
Patient sources				
Outpatient	12,999	0.22 (0.08)	0.22 (0.08)	0.00 (0.05)
Inpatient	23,038	0.24 (0.09)	0.23 (0.09)	–0.01 (0.06)
Complication				
No	17,903	0.23 (0.08)	0.22 (0.08)	–0.01 (0.06)
Yes	17,163	0.24 (0.09)	0.23 (0.09)	–0.01 (0.06)
Duration (years)				
<5	13,993	0.22 (0.08)	0.21 (0.09)	–0.01 (0.06)
≥5	22,034	0.24 (0.09)	0.23 (0.09)	0.00 (0.05)
Baseline HbA_{1c} (%)				
7-8	5783	0.22 (0.08)	0.22 (0.08)	0.00 (0.05)
8-9	8157	0.23 (0.09)	0.23 (0.09)	–0.01 (0.05)
9-10	5711	0.23 (0.09)	0.22 (0.09)	–0.01 (0.05)
≥10	11,003	0.24 (0.09)	0.23 (0.09)	–0.02 (0.06)
FBG[b] goal setting (mmol/L)				
≥6.1	30,034	0.23 (0.09)	0.22 (0.09)	–0.01 (0.06)
<6.1	5992	0.24 (0.09)	0.23 (0.09)	–0.01 (0.06)
Dose adjustment				
Up	15,444	0.22 (0.08)	0.24 (0.08)	0.02 (0.04)
Keep	2546	0.23 (0.08)	0.23 (0.08)	0.00 (0.00)
Down	18,047	0.25 (0.09)	0.21 (0.09)	–0.04 (0.06)

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]FBG: fasting blood glucose.

Incidence and Event Rate of Hypoglycemia

Hypoglycemia incidence (≤3.9 mmol/L) in all patients was 27.1% (3317/12,227). Inpatients had a higher incidence than outpatients (OR 1.24, 95% CI 1.07-1.45; *P*=.005). Patients with complications experienced more hypoglycemia (OR 1.25, 95% CI 1.07-1.45; *P*=.005). Longer diabetes duration (≥5 years) was associated with lower hypoglycemia incidence (OR 0.60, 95% CI 0.52-0.69; *P*<.001; Table 5). Higher baseline HbA_{1c} levels correlated with increased hypoglycemia risk. HbA_{1c} ≥10% had the highest incidence (OR 1.47, 95% CI 1.23-1.76; *P*<.001; Table 5). For hypoglycemic events defined by SMBG levels of ≤3.9 mmol/L, a total of 7619 events occurred, yielding an event rate of 2.49 events per person-year (Table 6). Notable trends included higher hypoglycemia rates for inpatient sources compared with outpatient sources (RR 1.25, 95% CI 1.08-1.46; *P*=.004), higher rates in patients with complications compared with those with no complications (RR 1.25, 95% CI 1.07-1.47; *P*=.006), and a lower rate in patients with a diabetes duration of ≥5 years (RR=0.67, 95% CI 0.59-0.77; *P*<.001). Furthermore, elevated baseline HbA_{1c} levels (≥10%) were associated with a higher hypoglycemia rate (RR 1.37, 95% CI 1.15-1.63; *P*=.001). An FBG goal setting of <6.1 before initiating BI was not associated with increased hypoglycemic incidence (OR 0.97, 95% CI 0.77-1.22; *P*=.79) or rate (RR 0.92, 95% CI 0.76-1.1; *P*=.35). For hypoglycemia defined as SMBG <3.0 mmol/L, 15.1% (1852/12,227) of patients with 2954 events were recorded, resulting in an incidence rate of 0.97 events per person-year. Similar trends were observed in relation to complications, duration of diabetes, and baseline HbA_{1c} levels. An FBG goal of <6.1 was not related to increased incidence (OR 0.97, 95% CI 0.77-1.22; *P*=.79) or rate (RR 1.01, 95% CI 0.80-1.27; *P*=.95) of hypoglycemia (Table 6).

XSL•FO
RenderX

Table 5. Hypoglycemia incidence during 3 months of TRIO monitoring in patients with self-monitoring blood glucose.

	Values, n	Month 3, n (%)	OR[a] (95% CI)	*P* value
SMBG[b] (≤3.9 mmol/L)				
All	12,227	3317 (27.1)	N/A[c]	
Patient sources				
Outpatient	4581	1177 (25.7)	Reference	
Inpatient	7646	2140 (28.0)	1.24 (1.07-1.45)	.005
Complication				
No	1962	469 (23.9)	Reference	
Yes	5125	1350 (26.3)	1.25 (1.07-1.46)	.005
Duration (years)				
<5	5518	1803 (32.7)	Reference	
≥5	6709	1514 (22.6)	0.60 (0.52-0.69)	<.001
Baseline HbA_{1c}[d] (%)				
7-8	1953	390 (20.0)	Reference	
8-9	2138	512 (23.9)	1.14 (0.94-1.38)	.20
9-10	1787	461 (25.8)	1.26 (1.03-1.55)	.02
≥10	3924	1276 (32.5)	1.47 (1.23-1.76)	<.001
FBG[e] goal setting (mmol/L)				
≥6.1	10,422	2827 (27.1)	Reference	
<6.1	1777	484 (27.2)	0.98 (0.81-1.17)	.79
SMBG (<3.0 mmol/L)				
All	12,227	1852 (15.2)	N/A	
Patient sources				
Outpatient	4581	702 (15.3)	Reference	
Inpatient	7646	1150 (15.0)	1.05 (0.87-1.26)	.64
Complication				
No	1962	249 (12.7)	Reference	
Yes	5125	753 (14.7)	1.26 (1.04-1.54)	.02
Duration (years)				
<5	5518	1009 (18.3)	Reference	
≥5	6709	843 (12.6)	0.63 (0.53-0.74)	<.001
Baseline HbA_{1c} (%)				
7-8	1953	207 (10.6)	Reference	
8-9	2138	289 (13.5)	1.24 (0.97-1.6)	.09
9-10	1787	259 (14.5)	1.34 (1.03-1.74)	.03
≥10	3924	727 (18.5)	1.64 (1.3-2.07)	<.001
FBG goal setting (mmol/L)				
≥6.1	10,422	1584 (15.2)	Reference	
<6.1	1777	262 (14.7)	0.97 (0.77-1.22)	.79

[a]OR: odds ratio.

[b]SMBG: self-monitoring blood glucose.

[c]N/A: not applicable.

XSL•FO
RenderX

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]FBG: fasting blood glucose.

XSL•FO
RenderX

Table 6. Hypoglycemia rates during 3 months of TRIO monitoring in patients with self-monitoring blood glucose.

	Events, n	Events/person-year	RR[a] (95% CI)	*P* value
SMBG[b] (≤3.9 mmol/L)				
All	7619	2.49	N/A[c]	
Patient sources				
Outpatient	2581	2.25	Reference	
Inpatient	5038	2.64	1.25 (1.08-1.46)	.004
Complication				
No	960	1.96	Reference	
Yes	2897	2.26	1.25 (1.07-1.47)	.006
Duration (years)				
<5	4482	3.25	Reference	
≥5	3137	1.87	0.67 (0.59-0.77)	<.001
Baseline HbA_{1c}[d] (%)				
7-8	871	1.78	Reference	
8-9	1098	2.05	1.00 (0.82-1.21)	.97
9-10	994	2.22	1.15 (0.94-1.4)	.18
≥10	3168	3.23	1.37 (1.15-1.63)	.001
FBG[e] goal setting (mmol/L)				
≥6.1	6449	2.48	Reference	
<6.1	1155	2.60	0.92 (0.76-1.1)	.35
SMBG (<3.0 mmol/L)				
All	2954	0.97	N/A	
Patient sources				
Outpatient	1138	0.99	Reference	
Inpatient	1816	0.95	1.06 (0.87-1.29)	.56
Complication				
No	366	0.75	Reference	
Yes	1156	0.90	1.24 (1.01-1.52)	.04
Duration (years)				
<5	1691	1.23	Reference	
≥5	1263	0.75	0.66 (0.56-0.79)	<.001
Baseline HbA_{1c} (%)				
7-8	298	0.61	Reference	
8-9	456	0.85	1.34 (1.04-1.72)	.03
9-10	381	0.85	1.30 (0.99-1.70)	.06
≥10	1253	1.28	1.72 (1.36-2.18)	<.001
FBG goal setting (mmol/L)				
≥6.1	2499	0.96	Reference	
<6.1	447	1.01	1.01 (0.80-1.27)	.95

[a]RR: relative risk.

[b]SMBG: self-monitoring blood glucose.

[c]N/A: not applicable.

XSL•FO
RenderX

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]FBG: fasting blood glucose.

Table 7 shows the composite end points of patients reaching the target with no hypoglycemia events during a 3-month period. For HbA_{1c} <7% without SMBG ≤3.9 mmol/L at month 3, patients with complications (OR 0.68, 95% CI 0.54-0.84; *P*=.001) and those with a duration of diabetes for ≥5 years (OR 0.76, 95% CI 0.62-0.93; *P*=.008) had significantly lower odds of reaching the target with no hypoglycemia events. Furthermore, higher baseline HbA_{1c} levels in the ranges of 9%-10% (OR 1.26, 95% CI 1.03-1.55; *P*=.02) and ≥10% (OR 1.47, 95% CI 1.23-1.76; *P*<.001) were associated with increased odds of achieving the target. Also, those who set an FBG goal of <6.1 mmol/L at initiation had significantly higher odds of reaching the composite end point (OR 1.35, 95% CI 1.03-1.79; *P*=.03). None of these factors were related to the composite end points for achieving FBG <7 mmol/L without SMBG ≤3.9 mmol/L at month 3. Finally, for achieving FBG <6.1 mmol/L without SMBG ≤3.9 mmol/L at month 3, only the duration of diabetes for ≥5 years was associated with a significantly lower possibility of achieving this composite end point.

XSL•FO
RenderX

Table 7. Composite end points of patients reaching target with no hypoglycemia events during 3 months of TRIO monitoring in patients with self-monitoring blood glucose.

	Values, n	Month 3, n (%)	OR[a] (95% CI)	*P* value
HbA_{1c}[b] <7% without SMBG[c] ≤3.9 mmol/L				
All	5121	2055 (40.1)	N/A[d]	
Patient sources				
Outpatient	1987	750 (37.7)	Reference	
Inpatient	3134	1305 (41.6)	1.14 (0.92-1.41)	.23
Complication				
No	735	352 (47.9)	Reference	
Yes	1952	697 (35.7)	0.68 (0.54-0.84)	.001
Duration (years)				
<5	2325	1024 (44.0)	Reference	
≥5	2796	1031 (36.9)	0.76 (0.62-0.93)	.008
Baseline HbA_{1c} (%)				
7-8	942	436 (46.3)	Reference	
8-9	881	319 (36.2)	1.14 (0.94-1.38)	.20
9-10	811	297 (36.6)	1.26 (1.03-1.55)	.02
≥10	1495	579 (38.7)	1.47 (1.23-1.76)	<.001
FBG[e] goal setting, mmol/L				
≥6.1	4435	1735 (39.1)	Reference	
<6.1	686	320 (46.6)	1.35 (1.03-1.79)	.03
FBG <7 mmol/L without SMBG ≤3.9 mmol/L				
All	5988	2398 (40.0)	N/A	
Patient sources				
Outpatient	2203	883 (40.1)	Reference	
Inpatient	3785	1515 (40.0)	0.98 (0.81-1.18)	.80
Complication				
No	830	363 (43.7)	Reference	
Yes	2531	947 (37.4)	0.85 (0.7-1.04)	.11
Duration (years)				
<5	2831	1197 (42.3)	Reference	
≥5	3157	1201 (38.0)	0.93 (0.78-1.11)	.41
Baseline HbA_{1c} (%)				
7-8	957	438 (45.8)	Reference	
8-9	1005	436 (43.4)	1.08 (0.85-1.37)	.53
9-10	897	326 (36.3)	0.94 (0.73-1.20)	.61
≥10	2050	807 (39.4)	0.94 (0.75-1.17)	.58
FBG goal setting, mmol/L				
≥6.1	5071	1989 (39.2)	Reference	
<6.1	905	406 (44.9)	1.06 (0.84-1.33)	.62
FBG <6.1 mmol/L without SMBG ≤3.9 mmol/L				
All	5988	1174 (19.6)	N/A	

XSL•FO
RenderX

	Values, n	Month 3, n (%)	OR[a] (95% CI)	*P* value
Patient sources				
Outpatient	2203	438 (19.9)	Reference	
Inpatient	3785	736 (19.4)	0.83 (0.66-1.04)	.11
Complication				
No	830	194 (23.4)	Reference	
Yes	2531	434 (17.1)	0.81 (0.64-1.03)	.09
Duration (years)				
<5	2831	643 (22.7)	Reference	
≥5	3157	531 (16.8)	0.74 (0.59-0.92)	.007
Baseline HbA_{1c} (%)				
7-8	957	213 (22.3)	Reference	
8-9	1005	198 (19.7)	1 (0.75-1.34)	10
9-10	897	139 (15.5)	0.79 (0.58-1.09)	.15
≥10	2050	439 (21.4)	1.03 (0.79-1.36)	.82
FBG goal setting, mmol/L				
≥6.1	5071	955 (18.8)	Reference	
<6.1	905	218 (24.1)	0.97 (0.74-1.29)	.86

[a]OR: odds ratio.

[b]HbA_{1c}: hemoglobin A_{1c}.

[c]SMBG: self-monitoring blood glucose.

[d]N/A: not applicable.

[e]FBG: fasting blood glucose

Discussion

Principal Findings

The TRIO program, a large-scale health management initiative using a digital WeChat platform for patients with T2DM initiating BI treatment, demonstrated its effectiveness in improving glycemic control 3 months after initiating BI. Prior to enrollment in TRIO, patients with T2DM exhibited suboptimal blood glucose control, with elevated baseline HbA_{1c} (9.6%) and FPG (9.5 mmol/L), a high prevalence of diabetic complications, and long diabetes duration (7.3 years). Following the 3-month TRIO management intervention, notable reductions in HbA_{1c} (–2.59%) and FBG (–2.77 mmol/L) were observed in the total population, accompanied by heightened proportions of achieving HbA_{1c} <7% (55.6%) and FBG target <7.0 mmol/L (61.3%) across diverse subgroups, such as patients from inpatient or outpatient care, patients with or with no complications, and patients with different length of diabetes duration, baseline HbA_{1c}, and FBG goal setting. This study also highlights the potential for setting a lower FBG target (<6.1 mmol/L) at the initiation of BI compared with the traditional <7.0 mmol/L target. By setting a more rigorous FBG target of <6.1 mmol/L, better glycemic control was achieved without increased risk of hypoglycemia. These results hold promise for digital health tools such as TRIO in improving the overall management of T2DM in real-world clinical settings.

TRIO has shown effectiveness and safety in patients with T2DM initiating BI after OAD failure with the assistance of WeChat digital platform, which is consistent with previous single-arm studies incorporating digital tools conducted in patients with prediabetes [15,18] and patients treated with premixed insulin and BI [14,19,20]. For instance, the Omada Health Program investigated digital Diabetes Prevention Program engagement among patients with prediabetes, showing a reduction of –0.33 mmol/L in HbA_{1c} levels over 3 years [15]. In our TRIO study, with a larger sample size, we achieved a greater HbA_{1c} reduction of –2.58 mmol/L. Furthermore, in a 12-week German trial involving individuals with type 2 diabetes on BI, a smartphone app (My Dose Coach) was compared with a written titration chart. The intervention group using the app exhibited a noteworthy reduction in HbA_{1c} levels compared with the control group (–0.31%; *P*=.04), with safety outcomes remaining unaffected. These findings suggest that app-assisted titration can enhance glycemic control in patients with type 2 diabetes who use BI [20].

TRIO has demonstrated that digital tools including health education and self-management modules added on BI might provide additional benefits to effectiveness in glycemic control than medication alone. The ORBIT study is an observational registry conducted in China with patients with T2DM who were inadequately controlled by OADs and initiated BI [21], with similar baseline HbA_{1c} (9.6%, SD 2%) but higher baseline FBG (11.7, SD 4.0 mmol/L) and shorter diabetes duration (6.4, SD

XSL•FO
RenderX

5.3 years) than those in our study. Notably, the change in HbA_{1c} from baseline to month 3 demonstrated a more improvement in the TRIO group (–2.59%) than in the ORBIT group (–2%), as well as the attainment of the HbA_{1c} target of <7% at month 3 (55.6% vs 35.9%). Despite the relatively lower reduction in FBG levels in the TRIO study due to lower baseline FBG levels, a larger proportion of TRIO patients successfully reached the FBG target of <7 mmol/L (37,017/60,377, 61.3%) than those in the ORBIT study (2078/5571, 37.3%). Another significant study in this field, the First Basal Insulin Evaluation (FINE) Asia study, was a multinational, prospective, observational approach to assess BI's efficacy in patients with uncontrolled type 2 diabetes (HbA_{1c} ≥8%) [22].

In TRIO, baseline HbA_{1c} and FBG were as high as 9.6% and 9.5 mmol/L, respectively (Table 1), which suggests delayed initiation of BI. The American Diabetes Association and the European Association for the Study of Diabetes suggest that BI should be promptly considered after the apparent "failure" of lifestyle modifications, including diet and exercise in combination with metformin, particularly when HbA_{1c} levels remain at or exceed 7% for a span of 2-3 months [23]. However, consistent with our findings, delay in injectable therapies was universal [24,25]. Timely initiation of insulin such as BI after the failure of oral treatment is associated with better glycemic control [26].

Standard T2DM management advice recommends keeping HbA_{1c} levels below 7%, but the ideal FPG target for achieving this is debated [27]. Different guidelines suggest varying FPG targets, such as 4.4-7.2 mmol/L according to the American Diabetes Association 2018 guidelines [28] or <6.1 mmol/L according to the American Association of Clinical Endocrinology-American College of Endocrinology and the International Diabetes Federation [29]. Previous studies support an FPG target of 6.1 mmol/L, showing better outcomes. Patients with FPG goals below 6.1 mmol/L had more significant HbA_{1c} reductions and higher target achievement rates without an increase in hypoglycemia [30,31]. Our results might confirm a better FPG target of <6.1 mmol/L. Patients who had an initial FPG goal setting below 6.1 mmol/L by their physician at the time of enrollment experienced both greater reductions in their HbA_{1c} levels (–2.64 vs –2.57%) and a higher HbA_{1c} target rate (67.8% vs 53.1%) than those with FPG goal setting ≥6.1 mmol/L (Tables 2 and 3). At the same time, hypoglycemic incidence and rate were comparable between the 2 groups. To enhance the management of hypoglycemic events, we recommend frequent SMBG monitoring, particularly during the initial weeks of insulin titration, to detect and address hypoglycemia promptly. Patient education on recognizing and treating hypoglycemic symptoms should be prioritized, alongside individualized insulin dose adjustments based on SMBG trends to minimize risk. Regular follow-ups are essential to reassess glycemic targets, such as FBG and HbA_{1c}, and to prevent overtreatment while maintaining optimal glycemic control. These measures can help balance achieving strict glycemic targets with ensuring patient safety.

Regarding the titration of BI treatment, the current Chinese guideline recommends an initial dose of 0.2 U/kg or 10 U, underscoring the importance of active insulin dose adjustment to achieve optimal glycemic control [32]. Previous studies such as ORBIT have indicated inadequate titration, evident from a starting dose of 0.18 IU/kg/d and a final dose of 0.21 IU/kg/d, resulting in a change of +0.034 IU/kg/d. Within our program, comprehensive titration was not uniformly accomplished. Among the 36,037 patients with baseline and 3-month dosage data, 42.9% (15,444/36,037) of patients escalated their dosage during the program, while 7.1% (2546/36,037) maintained stability and 50.1% (18,047/36,037) of patients decreased their dosage. Consequently, there was a marginal numerical decline in dose by –0.01 (0.06) U/kg. Plausible explanations for this trend encompass the higher-than-recommended starting dose in our program, which even surpassed the final doses in earlier ORBIT studies. Furthermore, the pursuit of targeted FBG levels in the initial weeks and an increased incidence of hypoglycemic events among patients (as shown in Table S3 in Multimedia Appendix 1) could have contributed to these patterns.

Limitations

Our study supports TRIO's effectiveness and safety as a personalized health program, yet several limitations warrant acknowledgment. First, a significant portion of patients lacked HbA_{1c} follow-up data at month 3, possibly introducing a compliance bias that could overstate TRIO's effectiveness by excluding those with less favorable HbA_{1c} reductions. Nonetheless, comparable baseline characteristics between compliant and noncompliant patients mitigate the risk of overestimation. Long-term data (months 3-12) are insufficient, leaving uncertainty about TRIO' s enduring effectiveness and ability to sustain patient adherence. Second, the judgment of hypoglycemia relied on SMBG, and the SMBG data were collected mainly through self-reporting on the WeChat platform, smart glucose devices, and regular phone calls by nurses, which may have a certain degree of deviation from the actual occurrence of hypoglycemia. Moreover, the absence of an RCT design and an external control group that included patients with type 2 diabetes who initiated only BI therapy without using TRIO may lead to the effect of TRIO to optimize glycemic control not solid enough. Therefore, it is necessary for future studies to design RCTs to evaluate the effectiveness and safety of intelligent blood glucose management in patients with diabetes with initiated BI therapy.

Conclusions

The TRIO program has demonstrated effectiveness in glycemic control, as reflected in HbA_{1c} and FBG levels, among patients with T2DM initiating BI therapy. The program has improved HbA_{1c} and FBG target rates and patient compliance with insulin treatments. However, it is important to acknowledge the limitations of our study, including compliance bias, insufficient long-term follow-up data, and the need for further investigation using rigorous study designs. Future research, such as RCTs, is warranted to validate our study's findings and assess TRIO' s generalizability in real-world populations.

XSL•FO
RenderX

Acknowledgments

The authors thank all the patients who kindly participated in this study. This work was supported by the National Natural Science Foundation of China Grant Awards (81570737), grants from Jiangsu Provincial Key Medical Discipline (ZDXKB2016012), the Key Research and Development Program of Jiangsu Province of China (BE2023774), and Clinical Trial Funding from the Affiliated Drum Tower Hospital, Medical School of Nanjing University (2024-LCYJ-PY-66).

Data Availability

The datasets generated during or analyzed during this study are not publicly available due to business confidentiality reasons but are available from the corresponding author upon reasonable request.

Authors' Contributions

WWM and DLZ carried out the studies; LXG, LXS, and LC participated in collecting data; and CXL drafted the manuscript. LMC and YMX performed the statistical analysis and participated in its design. HL, YZL, and JY participated in the acquisition, analysis, or interpretation of data and drafted the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Loss to follow-up, dosage regimen, fasting blood glucose (FBG) target rate, incidence of hypoglycemia, and levels of FBG during the TRIO optimal health management program.
[DOCX File , 51 KB - jmir_v27i1e67554_app1.docx]

References

1. Li Y, Teng D, Shi X, Qin G, Qin Y, Quan H, et al. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. BMJ 2020;369:m997 [FREE Full text] [doi: 10.1136/bmj.m997] [Medline: 32345662]
2. Ramachandran A, Jain SM, Mukherjee S, Phatak S, Pitale S, Singh SK, et al. Suboptimal glycemic control among subjects with diabetes mellitus in India: a subset analysis of cross-sectional wave-7 (2016) data from the international diabetes management practices study (IDMPS). Ther Adv Endocrinol Metab 2020;11:1-16 [FREE Full text] [doi: 10.1177/2042018820937217] [Medline: 32647562]
3. Wang L, Peng W, Zhao Z, Zhang M, Shi Z, Song Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. JAMA 2021;326(24):2498-2506 [FREE Full text] [doi: 10.1001/jama.2021.22208] [Medline: 34962526]
4. Silva JAD, Souza ECF, Echazú Böschemeier AG, Costa C, Bezerra HS, Feitosa E. Diagnosis of diabetes mellitus and living with a chronic condition: participatory study. BMC Public Health 2018;18(1):699 [FREE Full text] [doi: 10.1186/s12889-018-5637-9] [Medline: 29871637]
5. Jia W, Weng J, Zhu D, Ji L, Lu J, Zhou Z, et al. Standards of medical care for type 2 diabetes in China 2019. Diabetes Metab Res Rev 2019;35(6):e3158. [doi: 10.1002/dmrr.3158] [Medline: 30908791]
6. Ji L, Kang ES, Dong X, Li L, Yuan G, Shang S, et al. Efficacy and safety of insulin glargine 300 U/mL versus insulin glargine 100 U/mL in Asia Pacific insulin-naïve people with type 2 diabetes: the EDITION AP randomized controlled trial. Diabetes Obes Metab 2020;22(4):612-621 [FREE Full text] [doi: 10.1111/dom.13936] [Medline: 31797549]
7. Zhang P, Chen M, Zhang H, Luo Y, Zhu D, Li X, et al. Effectiveness and safety of basal insulin therapy in type 2 diabetes mellitus patients with or without metformin observed in a national cohort in China. BMC Endocr Disord 2022;22(1):26 [FREE Full text] [doi: 10.1186/s12902-021-00892-6] [Medline: 35045841]
8. Kaufman N. Digital therapeutics: leading the way to improved outcomes for people with diabetes. Diabetes Spectr 2019;32(4):301-303 [FREE Full text] [doi: 10.2337/ds19-0012] [Medline: 31798286]
9. Adhikari M, Devkota HR, Cesuroglu T. Barriers to and facilitators of diabetes self-management practices in Rupandehi, Nepal—multiple stakeholders' perspective. BMC Public Health 2021;21(1):1269 [FREE Full text] [doi: 10.1186/s12889-021-11308-4] [Medline: 34187461]
10. Agarwal S, Simmonds I, Myers AK. The use of diabetes technology to address inequity in health outcomes: limitations and opportunities. Curr Diab Rep 2022;22(7):275-281 [FREE Full text] [doi: 10.1007/s11892-022-01470-3] [Medline: 35648277]
11. Iregbu S, Spiers J, Duggleby W, Salami B, Schick-Makaroff K. Nigerian health care providers and diabetes self-management support: their perspectives and practices. Qual Health Res 2023;33(1-2):92-105 [FREE Full text] [doi: 10.1177/10497323221143889] [Medline: 36519805]
12. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023;6(1):38 [FREE Full text] [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]

XSL•FO
RenderX

13. Holmes D. Pharmacotherapy: a smarter way to treat obesity. Nat Rev Endocrinol 2017;13(11):626. [doi: 10.1038/nrendo.2017.135] [Medline: 28984317]
14. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
15. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital diabetes prevention program: 3-year update. BMJ Open Diabetes Res Care 2017;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
16. Lin J, Li X, Jiang S, Ma X, Yang Y, Zhou Z. Utilizing technology-enabled intervention to improve blood glucose self-management outcome in type 2 diabetic patients initiated on insulin therapy: a retrospective real-world study. Int J Endocrinol 2020;2020:1-8 [FREE Full text] [doi: 10.1155/2020/7249782] [Medline: 33224195]
17. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
18. Buch A, Yeshurun S, Cramer T, Baumann A, Sencelsky Y, Zelber Sagi S, et al. The effects of metabolism tracker device (Lumen) usage on metabolic control in adults with prediabetes: pilot clinical trial. Obes Facts 2023;16(1):53-61 [FREE Full text] [doi: 10.1159/000527227] [Medline: 36195053]
19. Lorig K, Ritter PL, Turner RM, English K, Laurent DD, Greenberg J. Benefits of diabetes self-management for health plan members: a 6-month translation study. J Med Internet Res 2016;18(6):e164 [FREE Full text] [doi: 10.2196/jmir.5568] [Medline: 27342265]
20. Hermanns N, Ehrmann D, Finke-Groene K, Krichbaum M, Roos T, Haak T, et al. Use of smartphone application versus written titration charts for basal insulin titration in adults with type 2 diabetes and suboptimal glycaemic control (My Dose Coach): multicentre, open-label, parallel, randomised controlled trial. Lancet Reg Health Eur 2023;33:100702 [FREE Full text] [doi: 10.1016/j.lanepe.2023.100702] [Medline: 37954005]
21. Ji L, Zhang P, Zhu D, Li X, Ji J, Lu J, et al. Observational registry of basal insulin treatment (ORBIT) in patients with type 2 diabetes uncontrolled with oral antihyperglycaemic drugs: real-life use of basal insulin in China. Diabetes Obes Metab 2017;19(6):822-830. [doi: 10.1111/dom.12886] [Medline: 28105735]
22. Tsai ST, Pathan F, Ji L, Yeung VT, Chadha M, Suastika K, et al. First insulinization with basal insulin in patients with type 2 diabetes in a real-world setting in Asia. J Diabetes 2011;3(3):208-216 [FREE Full text] [doi: 10.1111/j.1753-0407.2011.00137.x] [Medline: 21631903]
23. Nathan DM, Buse JB, Davidson MB, Ferrannini E, Holman RR, Sherwin R, et al. Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetes Care 2009;32(1):193-203 [FREE Full text] [doi: 10.2337/dc08-9025] [Medline: 18945920]
24. Goodall G, Sarpong EM, Hayes C, Valentine WJ. The consequences of delaying insulin initiation in UK type 2 diabetes patients failing oral hyperglycaemic agents: a modelling study. BMC Endocr Disord 2009;9(1):19 [FREE Full text] [doi: 10.1186/1472-6823-9-19] [Medline: 19804622]
25. Kim SG, Kim NH, Ku BJ, Shon HS, Kim DM, Park TS, et al. Delay of insulin initiation in patients with type 2 diabetes mellitus inadequately controlled with oral hypoglycemic agents (analysis of patient- and physician-related factors): a prospective observational DIPP-FACTOR study in Korea. J Diabetes Investig 2017;8(3):346-353 [FREE Full text] [doi: 10.1111/jdi.12581] [Medline: 27712034]
26. Chen P, Ma X, Chen H, Wang K, Zhou L. Delays in insulin initiation among patients with type 2 diabetes mellitus in southeast China: a retrospective, real-world study. Diabetes Metab Syndr Obes 2020;13:3059-3068. [doi: 10.2147/dmso.s256381]
27. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018;41(12):2669-2701 [FREE Full text] [doi: 10.2337/dci18-0033] [Medline: 30291106]
28. American Diabetes Association. 6. Glycemic targets standards of medical care in diabetes—2018. Diabetes Care 2018;41(Suppl 1):S55-S64. [doi: 10.2337/dc18-S006] [Medline: 29222377]
29. Garber AJ, Abrahamson MJ, Barzilay JI, Blonde L, Bloomgarden ZT, Bush MA, et al. Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm—2018 executive summary. Endocr Pract 2018;24(1):91-121. [doi: 10.4158/cs-2017-0153]
30. Yang W, Ma J, Yuan G, Li L, Zhang M, Lu Y. Determining the optimal fasting glucose target for patients with type 2 diabetes: results of the multicentre, open-label, randomized-controlled FPG GOAL trial. Diabetes Obes Metab 2019;21(8):1973-1977. [doi: 10.1111/dom.13733/v1/review1]
31. Yuan L, Li F, Zhou Y, Sun R, Gao G, Zhang Q, et al. Fasting glucose of 6.1 mmol/L as a possible optimal target for type 2 diabetic patients with insulin glargine: a randomized clinical trial. J Diabetes Res 2021;2021:5524313 [FREE Full text] [doi: 10.1155/2021/5524313] [Medline: 34337072]

XSL•FO
RenderX

32. Shi G, Zhu N, Qiu L, Yan H, Zeng L, Wang D, et al. Impact of the 2020 China Diabetes Society guideline on the prevalence of diabetes mellitus and eligibility for antidiabetic treatment in China. Int J Gen Med 2021;14:6639-6645. [doi: 10.2147/ijgm.s331948]

Abbreviations

BI: basal insulin
FBG: fasting blood glucose
FPG: fasting plasma glucose
HbA_{1c}: hemoglobin A1c
LS: least square
OAD: oral antidiabetic drug
OR: odds ratio
PPG: postprandial glucose
RCT: randomized controlled trial
RR: relative risk
SMBG: self-monitoring blood glucose
T2DM: type 2 diabetes mellitus

Edited by A Coristine; submitted 15.10.24; peer-reviewed by J Zhou; comments to author 22.11.24; revised version received 17.12.24; accepted 18.12.24; published 13.01.25.

Please cite as:
Li C, Guo L, Shi L, Chen L, Chen L, Xue Y, Li H, Liang Y, Yang J, Wang W, Zhu D
Effectiveness and Safety of the TRIO Optimal Health Management Program in Patients With Type 2 Diabetes Mellitus Initiating Basal Insulin Therapy: Prospective Observational Real-World Study
J Med Internet Res 2025;27:e67554
URL: https://www.jmir.org/2025/1/e67554
doi:10.2196/67554
PMID:

©Chenxi Li, Lixin Guo, Lixin Shi, Li Chen, Liming Chen, Yaoming Xue, Hong Li, Yuzhen Liang, Jing Yang, Weimin Wang, Dalong Zhu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Eliciting Preferences for the Uptake of Smoking Cessation Apps: Discrete Choice Experiment

Dorothy Szinay[1,2,3], PhD; Rory A Cameron[4,5], PhD; Andy Jones[6], PhD; Jennifer A Whitty[4,5,7], PhD; Tim Chadborn[8], PhD; Jamie Brown[3,9], PhD; Felix Naughton[1,2], PhD

[1]Behavioural and Implementation Science Group, School of Health Sciences, University of East Anglia, Norwich, United Kingdom
[2]Addiction Research Group, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, United Kingdom
[3]Behavioural Science and Health, University College London, London, United Kingdom
[4]Norwich Medical School, University of East Anglia, Norwich, United Kingdom
[5]Applied Research Collaboration East of England, National Institute for Health Research, Cambridge, United Kingdom
[6]The Centre for Research in Public Health and Community Care, University of Hertfordshire, Hatfield, United Kingdom
[7]Evidera, London, United Kingdom
[8]Behavioural and Social Sciences Team, Department of Health and Social Care, London, United Kingdom
[9]SPECTRUM Consortium, London, United Kingdom

Corresponding Author:
Dorothy Szinay, PhD
Behavioural and Implementation Science Group
School of Health Sciences
University of East Anglia
Research Park
Norwich, NR4 7TJ
United Kingdom
Phone: 44 01603456161
Email: dorothy.szinay@uea.ac.uk

Abstract

Background: If the most evidence-based and effective smoking cessation apps are not selected by smokers wanting to quit, their potential to support cessation is limited.

Objective: This study sought to determine the attributes that influence smoking cessation app uptake and understand their relative importance to support future efforts to present evidence-based apps more effectively to maximize uptake.

Methods: Adult smokers from the United Kingdom were invited to participate in a discrete choice experiment. Participants made 12 choices between two hypothetical smoking cessation app alternatives, with five predefined attributes reflecting domains from the theoretical domains framework: (1) monthly price of the app (environmental resources), (2) credible source as app developer (social influence), (3) social proof as star rating (social influence), (4) app description type (beliefs about consequences), and (5) images shown (beliefs about consequences); or opting out (choosing neither app). Preferences and the relative importance of attributes were estimated using mixed logit modeling. Willingness to pay and predicted uptake of the most and least preferred app were also calculated.

Results: A total of 337 adult smokers completed the survey (n=168, 49.8% female; mean age 35, SD 11 years). Participants selected a smoking cessation app rather than opting out for 90% of the choices. Relative to other attributes, a 4.8-star user rating, representing social proof, was the strongest driver of app selection (mean preference parameter 2.27, SD 1.55; 95% CI 1.95-2.59). Participants preferred an app developed by health care–orientated trusted organization (credible source) over a hypothetical company (mean preference parameter 0.93, SD 1.23; 95% CI 0.72-1.15), with a logo and screenshots over logo only (mean preference parameter 0.39, SD 0.96; 95% CI 0.19-0.59), and with a lower monthly cost (mean preference parameter –0.38, SD 0.33; 95% CI –0.44 to –0.32). App description did not influence preferences. The uptake estimate for the best hypothetical app was 93% and for the worst, 3%. Participants were willing to pay a single payment of up to an additional US $6.96 (UK £5.49) for 4.8-star ratings, US $3.58 (UK £2.82) for 4-star ratings, and US $2.61(UK £2.06) for an app developed by a trusted organization.

Conclusions: On average, social proof appeared to be the most influential factor in app uptake, followed by credible source, one perceived as most likely to provide evidence-based apps. These attributes may support the selection of evidence-based apps.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e37083) doi:10.2196/37083

KEYWORDS

discrete choice experiment; uptake, engagement; mHealth; smartphone app; smoking cessation; health app; behavior change; TDF; theoretical domains framework; mobile phone

Introduction

Smoking is one of the leading risk factors for noncommunicable diseases worldwide [1]. Supporting people to quit smoking is a primary concern for public health [2]. One approach is the use of apps, which can be effective for smoking cessation [3]. Many are available on commercial app stores but low uptake and suboptimal engagement with effective health apps is common [4,5]. Commercial app stores generally omit app quality measures and provide insufficient information about apps [6].

Curated health app portals are websites that present a list of selected health apps and can provide access to high-quality smoking cessation apps developed by trusted organizations [7]. This could increase the uptake of effective smoking cessation apps among smokers and decrease the risk that apps are installed primarily due to popularity, as opposed to potential effectiveness, from commercial app stores [8]. In some countries, health organizations considered as being trusted organizations, such as the Office for Health Improvement and Disparities and the National Health Service (NHS) in the United Kingdom, offer such portals (Better Health, NHS Apps Library), or the Digital Health Applications (DiGA) directory in Germany [7].

There is extensive literature on engagement with health apps [9-14], but the evidence about factors influencing their uptake is limited. We have identified several factors that appear to influence the uptake and engagement of these apps and explored views on curated health app portals, primarily the NHS "NHS Apps Library," which since then was decommissioned, and the Office for Health Improvement and Disparities Better Health (formerly known as Public Health England's "One You" website) [7,10]. We found a common discrepancy between user needs and what an app offers, such as the perceived utility of the app, which refers to the way apps are presented, including the images shown and the description of the apps [7,10,15]. App users have also expressed disappointment with the presentation of apps on app portals [7].

The uptake of health apps may also be primarily affected by social influences such as social proof represented by ratings of an app [7,16,17]. However, highly rated apps do not necessarily mean evidence-based content and functionality [6]. Although highly rated smoking cessation apps appear better tailored to individual needs [17], other evidence suggests that there is only a weak association between the quality of a smoking cessation app and its popularity [16,18].

There is limited evidence on which factors are likely to drive the uptake of apps and no studies investigating smoking cessation app uptake from a curated portal. This paper aims to determine preferences for the uptake of a smoking cessation app by applying a discrete choice experiment (DCE) survey. The findings may provide preliminary insights to help health app portal developers present smoking cessation apps in ways that align with user preferences, potentially supporting the selection of evidence-based apps on curated health platforms. The survey further assessed a series of factors influencing the uptake of and engagement with smoking cessation apps to better understand to what extent these factors are facilitators or barriers.

Methods

DCE Development

DCE is a methodology premised on individuals choosing the option that offers them the greatest utility, and assumes that utility is a function of the importance an individual assigns to characteristics associated with a product [19]. This DCE provides a series of choices of two alternatives of a product or service (here referred to as App 1 and App 2), known as choice tasks. Each alternative app is described by a set of predefined attributes, with 2 or more levels.

The development of this DCE, described in detail elsewhere [20], followed best practice guidance [21-23]. The development was further informed by discussion with stakeholders, including patient and public involvement representatives, and involved the following steps: (1) identification of attributes and attribute levels, (2) development of the experimental design, (3) piloting and survey amendment, (4) data collection, and (5) data analysis.

This paper adheres to the principles outlined in the DIRECT (Discrete Choice Experiment Reporting) checklist [24] to ensure comprehensive and transparent reporting of findings (see Multimedia Appendix 1 for the checklist).

Attributes and Levels

A systematic review [10], and interview and think-aloud study [7] informed the selection of relevant factors mapped under the theoretical domains framework (TDF) that influence the uptake of health apps, hereby attributes of this study. The TDF informed our previous research as it provides comprehensive coverage of theory-driven domains of behavior [25].

The research team assessed the relevance and feasibility of the attributes identified in the previous stages and narrowed down the selection of potential attributes. The selection of attributes is described elsewhere [20]. The selected attributes and their levels are shown in Table 1. For the "who developed the app" attribute, we used "NHS Digital," which is a widely trusted organization in the United Kingdom, and "Mhealth Essentials Ltd" as a hypothetical company.

XSL•FO
RenderX

Table 1. The attributes and attribute levels included in the DCE[a] conducted between December 2020 and February 2021 to explore preferences for smoking cessation apps among adult smokers in the United Kingdom.

TDF[b] construct	Attributes	Attribute levels
Environmental resources (cost)	• The monthly price of the app	• US $0 (UK £0) • US $3.79 (UK £2.99) • US $7.60 (UK £5.99) • US $11.40 (UK £8.99)
Social influence (credible source)	• Who developed the app	• Does not say • "Mhealth Essentials Ltd" • "NHS[c] Digital"
Social influence (social proof)	• The ratings of the app	• Does not show • 3.2 stars • 4 stars • 4.8 stars
Beliefs in consequences (perceived utility of the app)	• App description	• Generic, to create a rough idea of what the app is about without getting into details of app features • Short with some details about app features • Long and detailed description of the app and its features
Beliefs in consequences (perceived utility of the app)	• Images	• Shows the logo of the app • Shows the screenshot of the app • Shows the logo and screenshot of the app

[a]DCE: discrete choice experiment.

[b]TDF: theoretical domains framework.

[c]NHS: National Health Service.

Experimental Design

Participant preferences were estimated using conditional logit regression to model their choices. While a mixed logit model was used in the final analysis, the design assumed the use of McFadden's conditional logit model: this is the most widely used approach in choice modeling [26,27], and it provides a starting point from where more sophisticated generalizing models such as mixed logit may be applied.

A DCE model specifies the probability that an individual will choose a specific smoking cessation app. This probability is expressed as a function of measured attributes specific to the alternative. The (simplified) underlying utility function for alternative *j* is shown in equation 1:

$$\text{utility of option } j = \beta1(\text{alternative1}) + \beta2(\text{opt-out}) + \beta3(\text{developer\_not\_shown}) + \beta4(\text{developer\_NHS}) + \beta5(\text{ratings4.8\_stars}) + \beta6(\text{ratings4\_stars}) + \beta7(\text{ratings3.2\_stars}) + \beta8(\text{descriptionsShort}) + \beta9(\text{descriptionsLong}) + \beta10(\text{imagesScreenshots}) + \beta11(\text{imagesBoth}) + \beta12(\text{costs\_in\_£}) + \varepsilon \quad \textbf{(1)}$$

In equation 1, *U* represents the overall utility gained from choosing alternative *j*, β is the mean coefficient attached to the corresponding attribute levels estimated by the mixed logit model and represents the part-worth utility attached to each attribute level, and ε is the random error of the model. β1 and β2 represent alternative specific constants.

Our DCE included three attributes (A) with three levels (L) and two attributes with four levels, which, following the formula L^A would have led to 432 possible choice alternatives in a full factorial design [28]. To limit participant burden, we used a fractional factorial design. We generated 48 choice tasks applying Bayesian D-efficient design principles using Ngene software (ChoiceMetrics) [29] and blocked them into four survey versions each containing 12 choice tasks (Multimedia Appendix 2). The choice tasks in each block and the order of the attributes were fixed. Each participant was randomized to complete one survey. An additional repeat choice task was added to test choice consistency which was excluded from the primary data analysis. This means one task was repeated which allowed us to examine the percentage of people who consistently chose the same alternative This design aimed to estimate the main effects and assumes no interaction between attributes in the design.

The initial version of the DCE was piloted web-based with 49 participants. During the pilot phase, the understanding of the tasks was explored, and feedback on the wording of the attributes was sought. Based on the feedback received, the wording of the survey and the order in which attributes were listed in the table were revised. Coefficients from the pilot phase were used as priors to estimate a Bayesian D-efficient design. Data from the pilot phase were not included in the final analysis.

To imitate real-world decisions regarding app uptake, an opt-out option was included ("Neither of these two"; Figure 1). Participants who chose the opt-out option were prompted to repeat the decision and make a forced choice between the two alternatives. As the rate of the opt-out was low, the complete dataset was used for the analysis of choice data, including the opt-out option.

XSL•FO
RenderX

Figure 1. Example of a choice task with an opt-out option in the discrete choice experiment to explore the uptake of a smoking cessation app among adult smokers in the United Kingdom between December 2020 and February 2021.

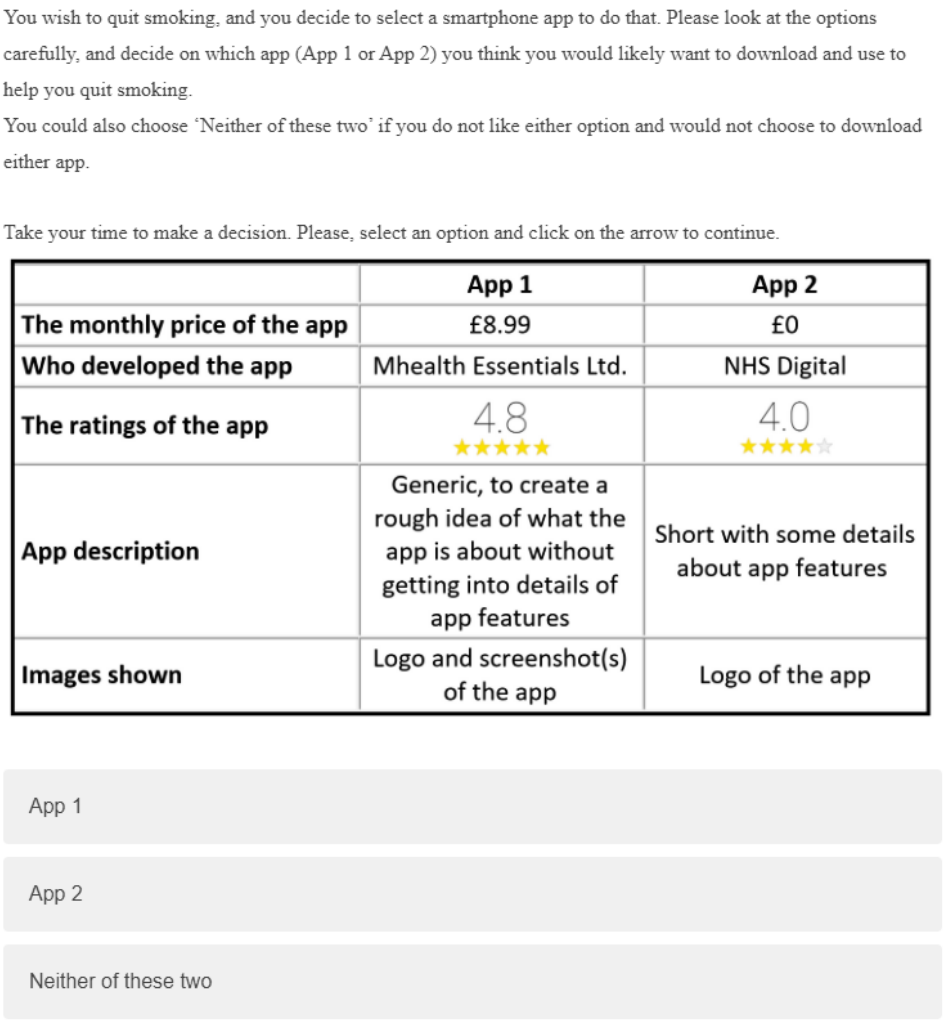

	App 1	App 2
The monthly price of the app	£8.99	£0
Who developed the app	Mhealth Essentials Ltd.	NHS Digital
The ratings of the app	4.8 ★★★★★	4.0 ★★★★☆
App description	Generic, to create a rough idea of what the app is about without getting into details of app features	Short with some details about app features
Images shown	Logo and screenshot(s) of the app	Logo of the app

Data Collection

Participants and Recruitment

This study was conducted digitally. Eligible participants were adults (1) aged 18 years and older, (2) residents in the United Kingdom, (3) able to give consent, (4) owned or had primary use of a smartphone, (5) smoked cigarettes, and (6) interested in quitting smoking using a smartphone app. Recruitment took place between December 2020 and February 2021, through social media, the "Call for Participants" platform, and the Prolific platform. Recruitment posts were shared on Facebook, Twitter, and LinkedIn to reach potential participants. This study was advertised on the "Call for Participants" platform as well. Participants recruited through these channels accessed a link to the survey in Qualtrics (Qualtrics International Inc) that included a screening questionnaire to determine eligibility. Those who met the criteria were able to complete the survey.

Prolific is a crowdsourcing platform that connects researchers with individuals willing to participate in studies. Prolific enabled efficient and targeted recruitment without the need for advertising. Since Prolific screened participants based on specified criteria, eligible individuals were redirected to a second version of the Qualtrics survey that contained only the survey questions, without the additional screening.

Sample Size

The sample size calculation was based on a rule-of-thumb formula (equation 2) [30].

$$N>500c/(t\times a) \quad \textbf{(2)}$$

In equation 2, "N" represents the sample size, *t* the number of tasks per participant (=12), "*a*" the number of alternatives in each choice task (=2), "*c*" the number of analysis cells (=4, as this is the largest number of levels for any of the attributes). Equation 2 suggests a minimum sample size of 83. With four versions of the survey, we targeted a sample size of at least 332 (4×83) participants.

Procedure

The survey was administered using Qualtrics survey software. Once consent was obtained participants were explained the DCE and were randomly assigned to one of the four DCE versions containing 13 choice tasks (12 from the design plus one repeat choice task) and were asked to complete further measures (see *measures*).

XSL•FO
RenderX

Measures

Attributes That Are Likely to Influence Smoking Cessation App Preferences

The primary outcomes are the marginal effects estimated for each attribute level in the choice model, represented by the β coefficients in equation 1.

Factors Perceived to Influence Uptake and Engagement With the Smoking Cessation App

We used the TDF to identify 13 potential facilitators and barriers to uptake and engagement with health apps based on factors identified as important in our previous work [7,10]. These were included in the survey as a set of statements with the level of agreement with the statements measured using a 5-point Likert-type scale. For analysis, responses to these statements were dichotomized into agree (strongly agree and agree) versus not (neither agree nor disagree, disagree, and strongly disagree; Multimedia Appendix 2).

Other Variables

The survey included questions about the previous use of smoking cessation apps and other health apps, user type (power user or minimal user) based on preferences exploring app features (ie, whether individuals use basic features and spend the minimum amount of time navigating the app [minimal users] versus those who would enjoy spending time navigating through features and engaging with the app regularly [power users]). Smoking behavior measures were heaviness of smoking index [31], frequency of smoking, attempts to stop smoking, strategies used in an attempt to quit smoking, intention to stop smoking, defined as whether the participant is planning to quit in the next 6 months, determination to stop smoking, and the main reason for stopping smoking. Sociodemographic characteristics were also measured (age, gender, level of education, household income, ethnicity, sexuality, and disability). See Multimedia Appendix 2 for the complete questionnaire.

Statistical Analysis

The pilot data were analyzed using the Apollo package in R (R Core Team; programming language developed by the R Foundation for Statistical Computing) [32], and the final data using Stata (version 16.1; StataCorp LLC). Participants' characteristics and the TDF factors perceived to influence the engagement with smoking cessation apps were summarized using descriptive statistics. Associations between attributes and uptake responses were estimated using a mixed logit model (MIXL). This approach accommodated the existence of preference heterogeneity within the sample by allowing one or more model parameters to be specified as having a random distribution [33]. In the model, all attributes were dummy-coded as categorical variables, except for cost, which was treated as continuous after verifying a linear relationship with utility, and all were treated as normally distributed random parameters. The model allowed for correlation between parameters. We analyzed data of participants who completed the full survey. The model was fit using 2000 Halton draws. We investigated the uptake of the most preferred and least preferred apps by calculating their utility values and the probabilities for selecting these hypothetical apps, using the approach described by Jonker et al [34]. To facilitate willingness to pay (WTP) estimation, the same model was used but specified cost as a fixed parameter. We analyzed the choice data of participants who were consistent with the repeat choice task and compared it to the results of the choice data including all participants. As part of a post hoc analysis poolability of the two different samples (incentivized Prolific sample; nonincentivized "other" sample), was assessed following the Swait and Louviere [35] procedure. The TDF factors perceived to influence the engagement with smoking cessation apps were described using proportions and 95% CIs.

Ethical Considerations

Ethical approval for this study was obtained from the University of East Anglia Faculty of Health Ethics Committee (reference: 2020/21-017). This approval covered all aspects of participant recruitment, data collection, and analysis. The study protocol was preregistered on the Open Science Framework [36]. Participants recruited on Prolific were paid US $1.90 (UK £1.50) for participation and those recruited on other platforms were invited to participate in a prize draw to win one of ten US $25.36 (UK £20) shopping vouchers. Participants provided informed consent before beginning the survey. Consent was obtained through Qualtrics survey software from participants who met the eligibility criteria, where participants were presented with an information sheet and consent form prior to accessing the DCE and additional questions. Participant data were anonymized to protect privacy. No identifiable information was collected, and all data were stored securely in accordance with institutional data protection policies. Responses were deidentified before analysis to ensure confidentiality.

Results

A total of 499 eligible participants were recruited, 469 participants consented, and 337 participants completed the experiment and measures. Out of these 337 participants, 196 participants were recruited through the Prolific website, and the rest through social media and through the Call for Participants website. Data from 337 participants yielded 4029 observations (15 choices were omitted by participants). Participants were aged between 19 and 65 years, with a mean age of 35 (SD 11) years, 168 (49.8%) were female participants, 176 (52.2%) participants showed low dependency on the heaviness of smoking index, and 107 (31.8%) participants had used smoking cessation apps before. Participants' characteristics are reported in Table 2.

XSL•FO
RenderX

Table 2. Characteristics of study participants in DCE[a] conducted between December 2020 and February 2021 to explore preferences for smoking cessation apps among adult smokers in the United Kingdom.

Sociodemographic data	Value (N=337)
Age (years)	
Range	19-65
Mean (SD)	35 (11)
Sex, n (%)	
Female	168 (49.8)
Male	163 (48.4)
Nonbinary or genderfluid	4 (1.2)
Prefer not to disclose	2 (0.6)
Ethnicity, n (%)	
Arab	10 (3)
Asian	6 (1.8)
Black or African American	11 (3.2)
Mixed or multiple ethnic groups	8 (2.4)
White	300 (89)
Other	2 (0.6)
Education, n (%)	
Postgraduate or equivalent	31 (9.2)
Degree or equivalent	127 (37.7)
A-levels or equivalent	113 (33.5)
GSCE[b] or equivalent	63 (18.7)
Other	3 (0.9)
Monthly net household income, n (%)	
US $0-$1267 (UK £0-£999)	39 (11.6)
US $1268-$2535 (UK £1000-£1999)	112 (23.2)
US $2536-$3803 (UK £2000-£2999)	68 (20.2)
US $3804-$5071 (UK £3000-£3999)	48 (14.2)
US $5072-$6339 (UK £4000-£4999)	23 (6.9)
Over US $6340 (UK £5000)	15 (4.4)
Prefer not to disclose	32 (9.5)
Sexual orientation, n (%)	
Heterosexual	268 (79.5)
LGBTQ+[c]	64 (19)
Prefer not to say	5 (1.5)
Disability, n (%)	
Living with disability	88 (26.1)
No disability	232 (68.8)
Prefer not to disclose	17 (5.1)
Type of smartphone, n (%)	
Android	163 (48.4)
Apple	164 (48.6)
Android and Apple	8 (2.4)

XSL•FO
RenderX

Sociodemographic data	Value (N=337)
Windows	2 (0.6)
Prior use of health app, n (%)	
Prior use of health app	226 (67.1)
No prior use of health app	111 (32.9)
Prior use of smoking cessation app	107 (31.8)
No prior use of smoking cessation app	230 (68.2)
Health app uptake source[d], n (%)	
Google search	62 (25.7)
Commercial app stores	158 (65.6)
Health-related website	51 (21.2)
Recommendations (friends, family)	58 (24.1)
Recommendations (health practitioners)	21 (8.7)
Other	6 (2.5)
User type[d], n (%)	
Power user	113 (46.9)
Minimal user	120 (49.8)
Unsure	8 (3.3)
Heaviness of smoking[e], n (%)	
Low dependence	176 (52.2)
Moderate dependence	139 (41.3)
High dependence	22 (6.5)
Last quit attempt, n (%)	
In the last month	44 (13.1)
In the last 12 months	136 (40.3)
Longer than 12 months	113 (33.5)
None	44 (13.1)
Previous experience with smoking cessation strategies, n (%)	
Nicotine replacement products	148 (43.9)
Zyban (buprorion)	9 (2.7)
Champix (varenicline)	26 (7.7)
E-cigarette or vaping device	195 (57.9)
Stop smoking group	28 (8.3)
Stop Smoking one-to-one counseling or support services	35 (10.4)
Smoking helpline	16 (4.8)
A book about quitting smoking	51 (15.1)
Smoking cessation website	54 (16)
Smoking cessation app	59 (17.5)
Other: hypnotherapy	2 (0.6)
None	66 (19.6)
Intention to quit in the next 6 months, n (%)	
Likely	240 (71.3)
Unlikely	23 (6.8)

XSL•FO
RenderX

Sociodemographic data	Value (N=337)
Unsure	74 (21.9)
Determination to quit, n (%)	
High determination	216 (64.1)
Moderately or slightly determined	113 (33.5)
Low determination	8 (2.4)
Main reason to quit, n (%)	
Health concerns	125 (37.1)
Health concerns related to COVID-19	28 (8.3)
To save money	112 (33.2)
To regain control	42 (12.5)
Pressure or encouragement from others	27 (8)
Other	3 (0.9)

[a]DCE: discrete choice experiment.

[b]GSCE: General Certificate of Secondary Education.

[c]LGBTQ+: lesbian, gay, bisexual, transgender, and queer or questioning.

[d]Questions answered by those who have used smoking cessation or health apps before.

[e]Computed from number of cigarettes smoked a day and the time the first cigarette is smoked in the morning [31].

On 89.9% of the choices, participants selected one of the two smoking cessation apps over "neither." There was no participant who opted out of all choices. Most of the attributes influenced participants' preferences, except for the description of the app (Table 3). Relative to other attributes and given the levels included in the DCE, social proof (the star rating of the app) was the most important attribute. Relative to the referent app (developed by Mhealth Essentials, star rating not shown, generic app description, with a logo shown only), having a 4.8-star rating (mean preference parameter 2.27, SD 1.55; 95% CI 1.95-2.59) was around twice as important as the 4-star rating (mean preference parameter 1.06, SD 1.32; 95% CI 0.78-1.34), and twice as important as having a credible source (it being developed by the NHS Digital, mean preference parameter 0.93, SD 1.23; 95% CI 0.72-1.15). Participants marginally preferred an app that showed screenshots (mean preference parameter 0.35, SD 0.50; 95% CI 0.19-0.52) or both screenshot and logo (mean preference parameter 0.39, SD 0.96; 95% CI 0.19-0.59) over logo only. An app with a low monthly price was also preferred (mean preference parameter –0.38, SD 0.33; 95% CI –0.44 to –0.32). However, the wide SDs, relative to their coefficients for many attributes indicate a broad variation in attribute importance among participants. There was significant preference heterogeneity across all attribute levels.

The characteristics of the most preferred app were having a monthly cost of US $0 (UK £0), a rating of 4.8 stars, developed by NHS Digital, having a generic description, and presenting both types of images (app logo and screenshots). The least preferred app has a monthly price of US $11.40 (UK £8.99), the developer is not shown, ratings of 3.2 stars, a long description, and shows the app logo only. The uptake level of the best app was estimated at 95%, and for the worst was estimated at 10%.

Table 4 reports marginal WTP estimates for improvement in the attributes of the app, relative to the reference category. Participants were willing to pay an additional US $6.96 (UK £5.49; 95% CI $6.15 [£4.85]-$7.75 [£6.11]) and US $3.58 (UK £2.82; 95% CI $2.89 [£2.28]-$4.26 [£3.36]) for app with 4.8- and 4-star ratings, respectively. Participants were willing to pay US $2.61 (UK £2.06; 95% CI $2.03 [£1.60]-$3.20 [£2.52]) for development by a trusted organization (NHS Digital) compared to Mhealth Essentials Ltd.

A total of 71 (21%) individuals were inconsistent with their choices. The demographics of this group were similar to those who were consistent with their choices. The results of the MIXL model with and without the individual's response who gave an inconsistent response to the repeat choice task returned comparable results (data not presented, available from author). However, in a post hoc analysis to investigate whether the preference parameters of the two sample groups were equal, strong evidence was found to reject this hypothesis (P<.001), indicating different choice preferences.

Participants indicated that the strongest facilitators that might promote their engagement with a smoking cessation app were user guidance of how to use the app (72.4% agreement, 95% CI 67.37%-76.93%), additional health information (75% agreement, 95% CI 70.16%-79.42%), and rewards (75.4% agreement, 95% CI 70.47%-79.69%; Table 5). Key barriers were concerns around data protection (66.8% agreement, 95% CI 61.54%-71.61%) cognitive load (47.5% agreement, 95% CI 42.16%-52.87%), reminders as triggers for cravings (40.7% agreement, 95% CI 35.51%-46%), and peer support (46.9% agreement, 95% CI 41.59%-52.25%).

XSL•FO
RenderX

Table 3. Mixed logit estimation results[a] from the DCE[b] exploring preferences for smoking cessation apps among adult smokers in the United Kingdom (December 2020-February 2021).

Attributes	Mean preference parameter (SE)	95% CI
Alternative specific constants		
Alternative 1		
Mean	0.15[c] (0.06)	0.02 to 0.27
SD[d]	0.14 (0.09)	–0.03 to 0.31
Alternative 2		
Mean	Reference	N/A[e]
SD	Reference	N/A
Opt out option		
Mean	–2.08[f] (0.22)	–2.50 to –1.65
SD	1.77[f] (0.23)	1.31 to 2.23
Developer		
Does not say		
Mean	–0.48[f] (0.10)	–0.68 to –0.27
SD	0.84[f] (0.13)	0.58 to 1.11
Mhealth Essentials		
Mean	Reference	N/A
SD	Reference	N/A
NHS Digital		
Mean	0.93[f] (0.11)	0.72 to 1.15
SD	1.23[f] (0.14)	0.96 to 1.50
Rating of the app		
Does not show		
Mean	Reference	N/A
SD	Reference	N/A
4.8 stars		
Mean	2.27[f] (0.16)	1.95 to 2.59
SD	1.55[f] (0.17)	1.21 to 1.89
4 stars		
Mean	1.06[f] (0.14)	0.78 to 1.34
SD	1.32[f] (0.18)	0.96 to 1.68
3.2 stars		
Mean	0.33[c] (0.14)	0.04 to 0.61
SD	1.65[f] (0.18)	1.30 to 1.99
App description		
Generic		
Mean	Reference	N/A
SD	Reference	N/A
Short		

XSL•FO
RenderX

Attributes	Mean preference parameter (SE)	95% CI
Mean	–0.02 (0.08)	–0.19 to 0.15
SD	0.50[f] (0.13)	0.25 to 0.75
Long		
Mean	–0.19 (0.11)	–0.41 to 0.03
SD	1.17[f] (0.14)	0.90 to 1.44
Images		
Logo		
Mean	Reference	N/A
SD	Reference	N/A
Screenshot		
Mean	0.35[f] (0.08)	0.19 to 0.52
SD	0.50[f] (0.12)	0.26 to 0.75
Both		
Mean	0.39[f] (0.10)	0.19 to 0.59
SD	0.96[f] (0.13)	0.73 to 1.26
Monthly price of the app[g]		
Mean	–0.38[f] (0.03)	–0.44 to –0.32
SD	0.33[f] (0.03)	0.27 to 0.41

[a]Akaike information criterion=6294.02; Bayesian Information criterion=6960.01; Log-likelihood=–3057.01.

[b]DCE: discrete choice experiment.

[c]*P* value of <.05.

[d]SD of the distribution around the mean preference estimate and is a measure of heterogeneity.

[e]Not applicable.

[f]*P* value of <.001.

[g]The monthly price of the app was coded as a continuous variable presented at four levels: US $0 (UK £0), US $3.79 (UK £2.99), US $7.60 (UK £5.99), and US $11.40 (UK £8.99).

Table 4. Willingness to pay results from the DCE[a] exploring preferences for smoking cessation apps among adult smokers in the United Kingdom (December 2020-February 2021).

Attribute	Acceptable reduction in price, WTP[b] (SE)	95% CI
App with a rating of 4.8 stars	US $6.69 (US $0.41); £5.49 (£0.32)	US $6.15 to US $7.75 (£4.85 to £6.11)
App with a rating of 4 stars	US $3.58 (US $0.34); £2.82 (£0.27)	US $2.89 to US $4.26 (£2.28 to £3.36)
App with a rating of 3.2 stars	US $0.41(US $0.38); £0.32 (£0.30)	US $–0.36 to US $1.15 (£–0.28 to £0.91)
App developed by the NHS[c]	US $2.61(US $0.29); £2.06 (£0.23)	US $2.03 to US $3.20 (£1.60 to £2.52)
App developer not shown	US $–1.56 (US $0.28); £–1.23 (£0.22)	US $–2.09 to US $–1.01 (£–1.65 to £–0.80)
App showing logo and screenshots	US $0.96 (US $0.27); £0.76 (£0.21)	US $0.43 to US $1.50 (£0.34 to £1.18)
App showing screenshots only	US $0.75 (US $0.23); £0.59 (£0.18)	US $0.30 to US $1.19 (£0.24 to £0.94)
App presented with a short description	US $–0.24 (US $0.23); £–0.19 (£0.18)	US $–0.70 to US $0.20 (£–0.55 to £0.16)
App presented with a long description	US $–0.06 (US $0.29); £–0.05 (£0.23)	US $-0.63 to US $0.51 (£–0.50 to £0.40)

[a]DCE: discrete choice experiment.

[b]WTP: willingness to pay.

[c]NHS: National Health Service.

XSL•FO
RenderX

Table 5. Percentage of potential factors influencing smokers' uptake and engagement with smoking cessation apps, as identified in a survey conducted among adult smokers in the United Kingdom[a].

	Percentage, %	95% CI
TDF[b] construct: Skills		
App literacy (facilitator): In general, I can easily use a newly installed app on my phone		
Agree	92.6	89.24-94.95
TDF construct: Knowledge		
App awareness (barrier): I was aware of the existence of smoking cessation apps prior to taking part in this study		
Agree	55.5	50.12-60.73
User guidance (facilitator): A guide of how to use features would help me use the app more often		
Agree	72.4	67.37-76.93
Health information (facilitator): Information in the app about how quitting smoking improves my health would make me use the app more often		
Agree	75	70.16-79.42
TDF construct: Memory, attention, decision processes		
Cognitive load (barrier) : In general, I don't want to use an app with features that would take some time to learn		
Agree	47.5	42.16-52.87
Reminders (facilitator): It would be important that an app to help me quit smoking sends personalised reminders to me		
Agree	68.3	63.07-73.02
Reminders (barrier): I wouldn't want to use an app that sent me reminders about quitting smoking in case it would trigger my cravings to smoke		
Agree	40.7	35.51-46
TDF construct: Social influence		
Peer-support (facilitator): Being connected with other app users would motivate me to stay on track with my intention to stop smoking		
Agree	65.6	60.32-70.48
Peer-support (barrier): Being connected with other app users would make me feel ashamed or disappointed if I started smoking again after quitting		
Agree	46.9	41.59-52.25
Professional support (facilitator): Being connected with online helpers (quit smoking advisors) within the app would make want to use the app more		
Agree	69.5	64.29-74.14
TDF construct: Beliefs about capabilities		
Self-confidence (facilitator): I am confident I could quit smoking by using an app		
Agree	50.7	45.4-56.07
TDF construct: Beliefs about consequences		
Data protection (barrier): I am concerned how my personal data is handled in apps		
Agree	66.8	61.54-71.61
TDF construct: Goals		
Goal setting and action planning: Receiving guidance of how to achieve goals is more important for me than just simply setting goals		
Agree	84.3	79.97-87.79
TDF construct: Social identity		
Social identity (barrier): When using a smoking cessation app, I don't want to feel that I am being treated like a patient		
Agree	61.4	56.1-66.49
TDF construct: Reinforcement		
Rewards (facilitator): Receiving badges or awards for achieving a set goal, would make me use the app more often		

XSL•FO
RenderX

	Percentage, %	95% CI
Agree	75.4	70.47-79.69

[a]Factors were derived using the theoretical domains framework and measured through participant responses collected between December 2020 and February 2021.

[b]TDF: theoretical domains framework.

Discussion

Main Findings

This study investigated five potential attributes each mapped to domains from the TDF, potentially relevant to the uptake of smoking cessation apps. Participants made choices between hypothetical app alternatives with predefined attributes: (1) monthly price (environmental resources), (2) developer credibility (social influence, as a credible source), (3) app rating (social influence, as social proof), (4) app description (beliefs about consequences), and (5) images (beliefs about consequences).

Participants preferred apps demonstrating strong social proof (high star ratings) and a credible source (trusted organizations as developers), along with a lower monthly cost. The description of the app shown to participants did not influence preferences.

Relative to other attributes, social proof (a high star rating) emerged as a particularly influential factor, consistent with findings from previous studies [7,17]. Familiarity with highly rated apps may contribute to this preference, as these apps are likely to appear at the top of search lists. However, not all high-quality evidence-based smoking cessation apps have high star ratings [16,18], which suggests that social proof may influence uptake more than evidence-based content.

The preference for credible sources, in this case, apps from trusted organizations, such as NHS Digital, aligns with existing evidence. Our findings are similar to a DCE that investigated the uptake of a COVID-19 tracing app in the United Kingdom where participants were more likely to adopt an NHS contact tracing app [37]. This reflects broader user concerns about reputable sources [38] and a preference for apps developed by experts over those from unknown or less reputable sources [39].

Interestingly, descriptions, however, did not seem to influence the uptake of a smoking cessation app. A plausible explanation is that this attribute may not have been described or presented in a way to capture the participants' attention. To avoid cognitive load, we provided brief, verbal definitions for generic, short, or long descriptions rather than examples. This presentation choice may have limited the salience of the app description attribute in this study, potentially impacting its relevance to participants.

Consistent with similar studies, our findings suggest that participants most preferred apps with no cost [10,40]. However, some individuals indicated a WTP for an app if it offers features aligned with their preferences, such as being developed by a credible source [7]. Preliminary estimates of marginal WTP suggest that users may accept a small fee if the app demonstrates strong social proof or is from a trusted organization, though these estimates should be interpreted with caution.

Only around half of our participants were aware of smoking cessation apps, indicating that further awareness efforts could help make these tools more accessible. In line with previous findings, access to health information and a user guide to using the app were viewed as factors that could improve engagement [7,10], particularly among participants with limited app literacy. Less than half of the participants reported they would not want to use an app with complex features. We previously found mixed views on reminders, with some believing they may negatively influence behavior change by triggering cravings [7,10]. In this study, we found that less than half of the participants reported reminders as barriers. Consistent with previous findings [7,10], potential users believed that peer and professional support could increase engagement [7,10], and fewer than half indicated that failure to quit would lead to disappointment. This shows the difficulty app developers may face when developing an app to suit most individuals' needs and the potential importance of guidance from organizations such as the UK National Institute of Health and Care on developing digital behavior change tools.

Overall, our findings suggest that social proof (star ratings) and credible source (apps developed by trusted organizations) may be more important for individuals than the price of the app. Additionally, we investigated additional factors related to engagement, mapped under the TDF. Together, these findings offer preliminary insights into user preferences that may be considered in app presentations on curated health app portals.

Implications

The findings of this study offer preliminary insights that may assist public health organizations in increasing the uptake of evidence or theory-informed smoking apps. They could also guide health app providers and curators of health app portals, such as the NHS Apps Library, in optimizing app presentation to better align with user preferences, potentially enhancing app uptake within curated health platforms.

App uptake was found to be driven primarily by price and social influence (app ratings), while presentation factors, such as the app's description and visual elements, were found to have a limited influence on choice, with a combined relative importance of under 8%. This relatively low level of importance ascribed to presentation may be partly due to the generic descriptions of these attributes used in the experiment, which may not have captured participants' attention or preferences fully.

While the information presented may help app curators prioritize specific information, such as cost and user ratings, when presenting apps, the scope of this study is limited to uptake drivers and does not extend to factors influencing ongoing user satisfaction or engagement.

The WTP estimates quantify the relative preference of smokers for the attributes in an accessible format. However, these values

XSL•FO
RenderX

should be interpreted with caution, as they represent initial estimates that require further validation before being applied more widely.

The integration of the TDF in this study aimed to provide a structured overview of attributes influencing app choice, building on previous research where the TDF was similarly applied. While these findings complement prior qualitative data [7,41] and systematic review [10], further research is warranted to establish a more comprehensive understanding of how individual app features impact social proof (user ratings), quality perceptions, and long-term engagement.

Limitations

Although the recruitment was designed to include a wide range of participants, the sampling strategy used is unlikely to have generated a representative sample of smokers. Some views may have been missed by recruiting exclusively digitally, including views of individuals experiencing homelessness, those living in deprived areas, and those living in areas without suitable internet coverage.

The design of this study investigated main effects only, therefore, possible interactions between attributes were not assessed. The sample size was inadequate to enable investigating stratifications of certain demographics. The DCE was piloted and refined based on feedback but the clarity and usability, ease of completion, or understanding of the DCE were not further assessed in the main study, for example, no think-aloud cognitive pilot interviews were conducted. While respondents were screened based on willingness to use an app to aid quitting, prompting participants to make a forced choice when they chose the opt-out option might have influenced their choice behavior. An example of this might be the avoidance of selecting the opt-out, in anticipation of the forced choice question, however, opt-out choices were evenly distributed across the 12 choices, indicating that this is not the case.

Prompting participants to make a forced choice when they chose the opt-out option might have influenced their choice behavior and in anticipation of the forced choice question, they may have chosen an alternative throughout the survey, which could be a reason for the small number of cases where the opt-out option was chosen Additionally, using a hypothetical company may have influenced smokers' preferences (in either positive or negative direction), compared to if a real company was used.

This study investigated the uptake of a smoking cessation app based on stated preferences, which may be different from the uptake of a smoking cessation app in real life. This study focused primarily on the uptake of smoking cessation apps and did not consider all previously identified uptake and engagement factors that may shape choice behavior (eg, the aesthetics of the app, reminders, social networking, and embedded health professional support [7,10]), which limits our conclusions about the relative importance of the factors studied. Additionally, this study included one attribute—user ratings—that indirectly reflected app quality and functionality. Future studies may benefit from including further attributes that indicate other quality parameters.

The specification of our model also introduces limitations. By specifying a normal distribution for cost in the main model, 12.5% of respondents are predicted to have positive price parameter. Behaviorally, this seems counterintuitive, however, some people may prefer a paid app over a free app (eg, because they associate cost with quality). Second, in order to facilitate a straightforward estimation of WTP, we specify cost as a fixed parameter. This is of course inconsistent with the results from the main model, which indicate significant heterogeneity with regards to preference for the cost of the app. This approach has the potential to severely bias WTP estimates, which should consequently be interpreted with caution.

The results on the poolability of the sample indicate that pooling of the 2 samples (those recruited from the incentivized Prolific site, and those recruited from elsewhere) was not appropriate, as there were substantial differences between how the independent variables influenced choice in each group. Despite highly similar demographic characteristics between the 2 groups, choice behavior may have been influenced by incentivization or other, unobserved differences between the 2 groups, biasing our results.

The relative importance of the attributes may vary between genders and age groups. Future DCEs may want to consider recruiting a larger sample size to investigate the relative importance of the attributes stratified based on sociodemographical factors. Knowing the preferences of certain groups with specific demographics may help to target the presentation of apps to increase uptake. To build on our limited conceptualization of the perceived utility of the app, future DCEs could borrow ideas from interaction design and user research studies and apply a visual representation of apps, instead of textual description. In this case, participants are shown images of apps as opposed to a table where logos are visually represented as opposed to described. Similarly, "app descriptions" could include typical feature content presented in different ways rather than abstractly reflecting different types of description (generic, short, or long). Finally, the impact of health apps is a combination of uptake and effectiveness. The measured factors influencing the uptake and engagement with smoking cessation apps suggest that more empirical studies, including testing in real-world situations, are needed to fully understand the extent of facilitators and barriers. There is very limited opportunity to investigate the external validity of this study, but it is hoped that as technology advances, opportunities will arise and test this research in real-life settings.

Conclusions

This study found that uptake is more likely if smoking cessation apps demonstrate strong social proof (high star ratings), are developed by a credible source (trusted organization), include screenshots, and are low cost. However, star ratings showed a relatively stronger influence on app selection compared to other attributes within the range of rating scores investigated.

XSL•FO
RenderX

Acknowledgments

The authors thank the patient and public involvement representatives and the University College London Alcohol and Tobacco group (UTARG) for their input on the conceptualization of the study. We also thank Prof Michiel Bliemer and Prof Stephane Hass for their input on the choice modeling in Ngene software and R using the Apollo package, respectively.

Data Availability

The dataset generated during this study is available on Open Science Framework [36].

Authors' Contributions

DS, FN, AJ, JAW, JB, and TC conceptualized the study design. DS wrote the study protocol with contributions from FN, AJ, JAW, JB, and TC. DS preregistered the study on OSF. DS generated the choice tasks with support from JAW. DS conducted the data collection and data investigation. DS undertook the data curation. DS analyzed the data with statistical support from RC. RC accessed and verified the underlying data. DS drafted the original manuscript, which was edited by critical review by FN, AJ, RC, JAW, JB, and TC. All authors read and approved the final version of the manuscript.

Conflicts of Interest

FN is an unpaid member of the scientific committee of the Smoke Free app, which is unrelated to this article.

Multimedia Appendix 1
Checklist for reporting discrete choice experiments in health: The DIRECT Checklist.
[DOCX File , 31 KB - jmir_v27i1e37083_app1.docx]

Multimedia Appendix 2
The choice sets and full survey questionnaire.
[DOCX File , 103 KB - jmir_v27i1e37083_app2.docx]

References

1. Noncommunicable diseases progress monitor. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/9789240047761 [accessed 2021-12-13]
2. Health 2020: a European policy framework and strategy for the 21st century. In: World Health Organisation. Copenhagen: World Health Organization. Regional Office for Europe; 2013. URL: https://iris.who.int/handle/10665/326386
3. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4(4):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
4. Kim Y, Oh B, Shin H. Effect of mHealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]
5. Murnane EL, Huffaker D, Kossinets G. Mobile health apps: adoption, adherence, and abandonment. 2015 Presented at: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers; September 2015; Osaka, Japan. [doi: 10.1145/2800835.2800943]
6. Terhorst Y, Messner E, Schultchen D, Paganini S, Portenhauser A, Eder AS, et al. Systematic evaluation of content and quality of English and German pain apps in European app stores. Internet Interventions 2021;24:100376 [FREE Full text] [doi: 10.1016/j.invent.2021.100376] [Medline: 33718002]
7. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of health and well-being apps and curated app portals: think-aloud and interview study. JMIR Mhealth Uhealth 2021;9(4):e27173 [FREE Full text] [doi: 10.2196/27173] [Medline: 33904827]
8. Khadjesari Z, Brown T, Naughton F. Regulation and accreditation of addictive behaviour applications—navigating the landscape. Addiction 2021;116(12):3276-3283. [doi: 10.1111/add.15484] [Medline: 33739480]
9. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digital Med 2021;4(1):20 [FREE Full text] [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
10. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]

XSL•FO
RenderX

11. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
12. Chan A, Honey M. User perceptions of mobile digital apps for mental health: acceptability and usability—an integrative review. J Psychiatr Ment Health Nurs 2022;29(1):147-168. [doi: 10.1111/jpm.12744] [Medline: 33604946]
13. Wilks CR, Chu C, Sim D, Lovell J, Gutierrez P, Joiner T, et al. User engagement and usability of suicide prevention apps: systematic search in app stores and content analysis. JMIR Form Res 2021;5(7):e27018 [FREE Full text] [doi: 10.2196/27018] [Medline: 34259163]
14. Spaulding EM, Marvel FA, Piasecki RJ, Martin SS, Allen JK. User engagement with smartphone apps and cardiovascular disease risk factor outcomes: systematic review. JMIR Cardio 2021;5(1):e18834 [FREE Full text] [doi: 10.2196/18834] [Medline: 33533730]
15. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
16. Wyatt JC. Correction to: how can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2019;17(1):144 [FREE Full text] [doi: 10.1186/s12916-019-1381-y] [Medline: 31324169]
17. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
18. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017;41(6):625-630 [FREE Full text] [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
19. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. Frontiers in Econometrics. New York: Academic Press; 1973:105-142.
20. Szinay D, Cameron R, Naughton F, Whitty JA, Brown J, Jones A. Understanding uptake of digital health products: methodology tutorial for a discrete choice experiment using the Bayesian efficient design. J Med Internet Res 2021;23(10):e32365 [FREE Full text] [doi: 10.2196/32365] [Medline: 34633290]
21. Hauber A, González JM, Groothuis-Oudshoorn C, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value Health 2016;19(4):300-315 [FREE Full text] [doi: 10.1016/j.jval.2016.04.004] [Medline: 27325321]
22. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health 2013;16(1):3-13 [FREE Full text] [doi: 10.1016/j.jval.2012.08.2223] [Medline: 23337210]
23. Bridges J, Hauber A, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. Value Health 2011;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
24. Ride J, Goranitis I, Meng Y, LaBond C, Lancsar E. A reporting checklist for discrete choice experiments in health: the DIRECT checklist. Pharmacoeconomics 2024;42(10):1161-1175. [doi: 10.1007/s40273-024-01431-6] [Medline: 39227559]
25. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
26. Train K. Discrete Choice Methods with Simulation. UK: Cambridge University Press; 2009.
27. Ryan M, Gerard K, Amaya-Amaya M. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht: Springer; 2008.
28. Hensher DR, Rose JM, Green W. Applied Choice Analysis: A Primer. UK: Cambridge University Press; 2005.
29. ChoiceMetrics. 2012. URL: https://ngene.software.informer.com/versions/ [accessed 2021-12-13]
30. Johnson R, Orme B. Getting the most from CBC. Sequim: Sawtooth Software Research Paper Series, Sawtooth Software. URL: https://sawtoothsoftware.com/resources/technical-papers/getting-the-most-from-cbc [accessed 2021-12-13]
31. Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Br J Addict 1989;84(7):791-799. [doi: 10.1111/j.1360-0443.1989.tb03059.x] [Medline: 2758152]
32. Hess S, Palma D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. J Choice Modell 2019;32:100170 [FREE Full text] [doi: 10.1016/j.jocm.2019.100170]
33. Hole AR. Fitting mixed logit models by using maximum simulated likelihood. Stata J 2007;7(3):388-401. [doi: 10.1177/1536867x0700700306]
34. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
35. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. J Mark Res 1993;30(3):305. [doi: 10.2307/3172883]

XSL•FO
RenderX

36. Szinay DC, Cameron R, Jones A, Whitty J, Chadborn T, Brown J, et al. Eliciting preferences for the uptake of smoking cessation apps: a discrete choice experiment. Open Science Framework. 2021. URL: https://osf.io/5439x/ [accessed 2021-12-21]
37. Wiertz C, Banerjee A, Acar OA, Ghosh A. Predicted adoption rates of contact tracing app configurations—insights from a choice-based conjoint study with a representative sample of the UK population. SSRN J 2020:1-19. [doi: 10.2139/ssrn.3589199]
38. Derbyshire E, Dancey D. Smartphone medical applications for women's health: what is the evidence-base and feedback? Int J Telemed Appl 2013;2013:782074 [FREE Full text] [doi: 10.1155/2013/782074] [Medline: 24454354]
39. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
40. Nittas V, Mütsch M, Braun J, Puhan MA. Self-Monitoring app preferences for sun protection: discrete choice experiment survey analysis. J Med Internet Res 2020;22(11):e18889. [doi: 10.2196/18889]
41. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021 Dec 16;9(12):e29098 [FREE Full text] [doi: 10.2196/29098] [Medline: 34927597]

Abbreviations

DCE: discrete choice experiment
DiGA: Digital Health Applications
DIRECT: Discrete Choice Experiment Reporting Checklist
NHS: National Health Service
TDF: theoretical domains framework
WTP: willingness to pay

Edited by A Mavragani; submitted 01.03.22; peer-reviewed by M Jonker, Y Hendriks; comments to author 27.04.22; revised version received 20.08.22; accepted 02.12.24; published 14.01.25.

Please cite as:
Szinay D, Cameron RA, Jones A, Whitty JA, Chadborn T, Brown J, Naughton F
Eliciting Preferences for the Uptake of Smoking Cessation Apps: Discrete Choice Experiment
J Med Internet Res 2025;27:e37083
URL: https://www.jmir.org/2025/1/e37083
doi:10.2196/37083
PMID:39808479

©Dorothy Szinay, Rory A Cameron, Andy Jones, Jennifer A Whitty, Tim Chadborn, Jamie Brown, Felix Naughton. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study

Gilly A Hendrie[1], PhD; Danielle L Baird[1], BSc; Genevieve James-Martin[1], MSc; Emily Brindal[1], PhD; Paige G Brooker[1], PhD

Commonwealth Scientific and Industrial Research Organisation, Adelaide, Australia

Corresponding Author:
Gilly A Hendrie, PhD
Commonwealth Scientific and Industrial Research Organisation
PO Box 10041
Adelaide, 5000
Australia
Phone: 61 883050662
Email: gilly.hendrie@csiro.au

Abstract

Background: A greater understanding of the effectiveness of digital self-management programs and their ability to support longer-term weight loss is needed.

Objective: This study aimed to explore the total weight loss and patterns of weight loss of CSIRO (Commonwealth Scientific and Industrial Research Organisation) Total Wellbeing Diet Online members during their first 12 months of membership and examine the patterns of platform use associated with greater weight loss.

Methods: Participants were Australian adults who joined the program between October 2014 and June 2022 and were classified as longer-term members, meaning they completed at least 12 weeks of the program, had baseline and 12-week weight data, and had a paid membership of ≥1 year (N=24,035). Weight loss and percentage of starting body weight loss were calculated at 3, 6, 9, and 12 months using 3 statistical approaches: (1) multiple imputations method, (2) all available data, and (3) complete data only. Among members with complete data (6602/24,035, 27.5%), patterns of weight loss and gain were examined, and how this related to total weight loss and platform use was explored.

Results: Members were mostly female (19,972/24,035, 83.09%), aged 31 to 50 years (9986/24,035, 41.5%) or 51 to 70 years (12,033/24,035, 50.06%), and most members were classified as overweight or obese (23,050/24,035, 95.9%). Using multiple imputations, the average estimated weight loss was 5.9 (SE 0.0245) kg at 12 weeks, 6.7 (SE 0.0348) kg at 6 months, 6.2 (SE 0.0400) kg at 9 months, and 5.5 (SE 0.0421) kg at 12 months. At 12 months, more than half the members (12,573/24,035, 52.3%) were at least 5% below their starting body weight and 1 in 4 (5865/24,035, 24.4%) were at least 10% below their starting body weight. In the subsample with complete data, the average weight loss at 12 months was 7.8 kg. The most common (961/6602, 14.56% members) weight loss pattern over the first 12 months was 6 months of weight loss, followed by 6 months of weight maintenance. This group had an average weight loss of 10.6 kg at 12 months (11.9% of their starting body weight). In a subgroup of participants who consistently lost weight over the 12-month period (284/6602, 4.3% of the sample), weight loss reached up to 22.3 kg (21.7% of their starting body weight). Weekly platform use was positively associated with total weight loss (r=0.287; P<.001). Members who used the platform >30 times per week (approximately >4 times/d) were more likely to lose weight in the first 6 months of the program.

Conclusions: This commercial weight loss program was shown to be effective, with 1 in 2 members achieving clinically significant results after 1 year. Greater engagement with the platform was associated with consecutive periods of weight loss and greater weight loss success overall.

(J Med Internet Res 2025;27:e65122) doi:10.2196/65122

KEYWORDS

obesity; obesity management; weight loss; internet-based intervention, weight management

XSL•FO
RenderX

Introduction

Background

Obesity is a global public health challenge that has substantial individual, societal, and economic impacts. It is predicted that, if the current trends continue, the majority (51%) of the global population will be living with either overweight or obesity by 2035, and 1 in 4 people will be classified as obese [1]. High-income countries tend to have the highest rates of obesity. For example, in Australia, 66% of the adult population is classified as overweight or obese, with approximately 1 in 3 individuals classified as obese [2]. The economic impacts of overweight and obesity are expected to reach 3% of the global Gross Domestic Product by 2035 [1]. The costs to the Australian economy attributable to overweight and obesity were estimated to be Aus $11.8 billion (US $7.4 billion) in 2017 to 2018 [3], and without significant action, these costs may rise to an estimated Aus $87.7 billion (US $54.9 billion) by 2032 [4]. In response to the impacts of obesity, the World Health Organization has set a global target to stop the increases in rates of overweight and obesity, and there have been calls to develop national strategies to address obesity. In line with this, the National Obesity Strategy for Australia aims to "halt the rise and reverse the trend in the prevalence of obesity in adults by 2030" [5]. The strategy acknowledges that achieving this ambitious goal will require coordinated efforts, and if successful, it will significantly enhance the population's quality of life.

Commercial Weight Loss Programs

Among the range of approaches to address obesity, improving dietary intake and increasing physical activity are considered useful lifestyle strategies for losing weight. Reviews of commercial weight loss programs have been primarily based on randomized controlled trials and suggest that individuals who complete these programs can achieve significant weight loss in the short term [6]. However, longer-term weight loss, and weight maintenance after weight loss, is challenging [6-8]. The literature, including randomized controlled trials [9] and studies on real-world commercial programs [10,11], commonly reports a pattern of short-term weight loss (at 3 and 6 months) followed by some degree of weight regain by 12 months and beyond.

Longitudinal and longer-term studies need to manage missing data. Randomized controlled trials often use statistical techniques, such as intention-to-treat analysis, to use data from all participants regardless of their adherence or completion of the program. However, a common limitation of real-world type studies that have evaluated the longitudinal effect of weight loss programs is that they often report results from a small proportion of the enrolled participants [12,13]. Therefore, findings are more likely to represent the results of a highly motivated subsample and less likely to represent the expected weight loss for an average participant or be applicable to the broader population. In fact, reported dropout rates for weight loss programs range from 10% to 80% at 12 months [14], often due to reasons associated with lack of, or less than, desirable weight loss [15]. Data from participants included in follow-up analyses would suggest that longer-term weight loss success is a real challenge. Approximately 80% of weight lost is regained within 5 years [16], and only 1 in 5 people are successful in achieving longer-term weight loss [17].

Studies of weight loss maintainers report some common behavioral strategies associated with longer-term success. Frequent self-monitoring of body weight and food intake, high levels of physical activity, and following a low-fat diet were shared behaviors among successful longer-term weight loss maintainers in the National Weight Control Registry [17-19]. Portion control, habitual healthy snack choice, and program engagement have also been shown to be positively associated with weight loss maintenance among individuals enrolled in a digital commercial program [16].

Purpose of This Study

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Digital Wellness have partnered to deliver a commercial, digital version of the CSIRO Total Wellbeing Diet. CSIRO is Australia's national science agency and Digital Wellness is an Australian digital health solutions company. The CSIRO Total Wellbeing Diet is a weight loss program that has evolved from clinical trials [20-22] to a popular series of books [23] and now a digital platform [24]. The program includes personalized eating plans, customized weekly meal plans, food and exercise diaries, the ability to record and see progress of weight loss, a member forum, and supportive correspondence via email. The CSIRO Total Wellbeing Diet Online program launched in 2014. The results of members who joined the program in the first 5 years have been published previously [24]. In a sample of >60,000 members who joined between 2014 and 2019, the average weight loss of all members was 2.8 kg (3.1% of their starting body weight), and for those who finished the 12-week program, the average weight loss was 4.9 kg (5.3% of their starting body weight). This study focused on longer-term weight loss and therefore used data from members who had joined the program between 2014 and 2022 and who had been members for at least a year. The analysis explored the total weight loss and patterns of weight loss of members during the first 12 months of membership and examined whether the patterns of platform use were associated with greater longer-term weight loss.

Methods

Study Design and Participants

Participants of this secondary analysis included all adults who joined the CSIRO Total Wellbeing Diet Online and had data in the platform between October 2014 and June 2022 (n=155,075 people referred to as members). In the registration process, individuals who were aged <18 years or had a BMI that placed them in the underweight category (BMI <18.5 kg/m^2) were automatically excluded. Participants excluded from this analysis were pseudo-members (ie, platform testers and affiliated staff), individuals whose membership was paid for by their employer because their motivations for weight loss might have been different from those who paid for their own membership, or people living outside of Australia because the context in which they were following the program might be different. Members

XSL•FO
RenderX

missing a starting weight as part of the setup process were removed because the primary outcome of weight loss could not be calculated (2172/155,075, 1.4%), and an additional 8 participants were removed because their recorded weight value was deemed invalid or unreliable using cutoff values of <13 kg or >250 kg.

This analysis focused on weight loss among longer-term members, defined as individuals who had completed at least 12 weeks of the program, had weight data recorded at baseline and approximately 12 weeks (61,514/152,895, 40.23%), and maintained a paid membership for about a year or longer. This left a sample of 24,035 longer-term members to examine the longer-term weight loss results of members on the CSIRO Total Wellbeing Diet Online program. To examine the patterns of weight loss over the first 12 months of membership, complete weigh-in data were required for the 4 time points of interest (12 weeks, 6 months, 9 months, and 12 months). There was a subsample 6602 members with complete weigh-in data used for the analysis of weight loss patterns.

Intervention

A detailed description of the program has been published previously [24]. Briefly, the CSIRO Total Wellbeing Diet Online is a commercial weight loss program managed by Digital Wellness and is available to individuals at a cost of Aus $199 (US $123) for the first 12 weeks and up to Aus $449 (US $278) for 12 months. The diet is a higher-protein, lower–glycemic index dietary pattern, with meal plans consisting of 3 meals and up to 2 snacks each day. The program contains a food group system where portions of food are presented as standard units for each food group. There are 7 food groups: fruit, vegetables, meat and alternatives, breads and cereals, dairy, healthy fats and oils, and indulgences. Meals are designed around a template of standard units, which ensures daily allowances of food groups are met and provides optimal nutrition and energy to promote weight loss. The weight loss program is promoted via media and digital channels, and the Australian public can join the program of their own volition. The user registration process collects information on year of birth, sex, physical activity levels, and weight loss goals to tailor eating plans. A food group unit system allows standard menu templates to be customized by swapping meals using the recipe database.

The platform is a fully responsive web application with an interface that is optimized for viewing on a desktop and mobile devices. The program content provides general information and advice, and weekly tutorials. Self-monitoring and goal setting are key behavior change techniques used within the program. A weight loss goal is provided during the onboarding process. The food diary can be used to log meals and snacks, either by entering prepopulated recipes from the meal plan or by entering individual foods from a comprehensive food database. The food search function allows members to search a database to view information about their composition. The food tracker allows food units and total energy consumed over the day to be logged and reviewed, with graphical feedback provided comparing intake to daily food unit targets. A large database of recipes is available within the platform, and members can swap recipes in and out of the menu plan to suit their taste preferences, time, and cooking skills. Members receive weekly emails on their nominated weigh-in day to remind them to weigh themselves and record their weight into the weight tracker. Progress data are presented in a graph and table form. Members are encouraged to engage with the platform regularly, with recommendations provided to weigh in weekly, use the self-monitoring tools daily, and view program tutorials. The platform also has a forum for members to share their stories or discuss relevant issues with other members. Seasonal content is shared with members, and the program content is reviewed and updated periodically. The platform functionality is updated with a focus on enhancing the user experience and engagement.

Data Collection and Study Outcomes

Digital Wellness manages the collection and storage of data from the program. Information collected at the time of joining included date of birth; sex; ethnicity; postcode; as well as body measurements such as height, weight, and waist circumference. Other data collected included program details such as paid membership duration and platform use data including entries into the food diary, views of the menu plans, views of exercise plans, views of program content information, forum visits, searches of the food database, and weight entries. These data were provided to the research team in a deidentified format, with each individual member assigned a unique identifier. Weigh-in data were provided for each member weekly from weeks 1 to 13 and then in 3-month blocks thereafter until membership ceased. The last recorded weight was provided within each 3-month block. Platform use data (number of weigh-ins, food diary entries, menu plan views, exercise plan views, program content views, forum views, food search views, and total platform activity) were provided at similar time intervals.

The primary outcome was weight loss, calculated as the difference in kilograms between weight on joining (baseline) and the 4 time points of interest, with a larger number representing greater weight loss. Weight loss was also calculated as a percentage of starting body weight and was categorized into 4 groups: weight gain, weight loss of >0% and <5% of starting body weight, 5% to <10% of starting body weight, and ≥10% of starting body weight lost.

Ethical Considerations

Ethics approval to conduct this research was received from CSIRO Health and Medical Human Research Ethics Committee (2022_055_LR). As part of registration, participants agreed to their data being shared with CSIRO and used for research purposes; therefore, no direct participant consent was sought. All data were provided to the research team in a deidentified format, and individuals could not be reidentified by the research team. Participants did not receive compensation.

Statistical Analysis

Overview

Statistical analyses were performed using the SPSS statistical software package (version 29; IBM Corp). Descriptive analysis was conducted to describe the characteristics of program members. The average weight loss of members was calculated

XSL•FO
RenderX

in three ways: (1) using multiple imputations as a method to replace missing data (24,035/24,035, 100%); (2) using the available weight data (ranged from 18,032/24,035, 75.02% at 3 months to 9413/24,035, 39.16% at 12 months); and (3) using only members with complete data at all time points (6602/24,035, 27.47%).

Weight loss in kilograms and as a percentage of starting body weight were calculated. Weight loss was calculated at each of the time points of interest and by subgroups, including by sex, age group, starting weight status category (normal weight, overweight, or obese [obesity class 1, 2, and 3]), state of residence, quintiles of Socio-Economic Indexes for Areas, and duration of paid membership. Total platform use was aggregated and divided by weeks of membership to estimate an average weekly platform use for each member. The association between platform use and weight loss was assessed using Pearson correlation coefficient, where a positive value indicates a positive association between use and weight loss.

Primary Analysis

Weigh-in data were needed to calculate weight loss. The amount of missing weigh-in data increased over time. At 12 weeks, 25% of weigh-in values were missing; at 6 months, 29% were missing; at 9 months, 50% were missing; and at 1 year, 61% were missing. To explore the effect of missing weigh-in data on weight loss estimates, analyses were conducted on the raw data (missing data ignored) before applying the multiple imputations method. The monotone multiple imputation model was used, which included weight loss at each time point, age group, sex, state, starting weight, starting BMI, and Socio-Economic Indexes for Areas quintiles. A total of 20 imputation datasets were created. The model assumed missing data were missing not completely at random. Results presented are the average of separate, identical analyses performed on each of the 20 complete imputation datasets, with appropriate adjustment of SEs to incorporate the added variability from the imputation process.

Sensitivity Analysis

Weight loss was also calculated using the available weigh-in data in the system. The amount of weigh-in data ranged from 75.02% (18,032/24,035) at 3 months to 39.16% (9413/24,035) at 12 months. Weight loss was also calculated using only members with complete weight data at all time points (6602/24,035, 27.47%). Weight loss patterns in four 3-month blocks over 12 months were examined and recorded as a sequence. Weight loss during each 3-month period was categorized as weight loss (L), weight gain (G), or weight maintenance (M). Consistent with previous research [25], a change in weight within 1.4 kg of previous weight was considered weight maintenance, and anything greater was considered weight loss or weight gain. For example, if a member lost weight from baseline to week 12 and also lost weight in months 3 to 6, 6 to 9, and 9 to 12, then the weight loss pattern was labelled as L|L|L|L to indicate 4 consecutive periods of weight loss. Alternatively, if a period of weight loss was followed by a period of maintenance and then 6 months of gradual weight gain, the pattern was labelled as L|M|G|G. Total weight loss over 12 months and average weekly platform use were examined by the pattern of weight loss.

Results

Characteristics of Members

Longer-term members (24,035/61,514, 39.07%) had an average of 600 days of paid membership (Table 1). The average starting weight was 92.6 kg, and the average BMI was 33.1 kg/m^2 on commencement of the program. Most members were female (19,972/24,035, 83.09%) and aged between 31 and 50 years (9986/24,035, 41.5%) or 51 to 70 years (12,033/24,035, 50.06%). Nearly all the longer-term members were classified as overweight or obese (23,050/24,035, 95.9%), with 64.68% (15,546/24,035) classified as obese. Approximately one-third of the longer-term members (8291/24,035, 34.49%) were in the highest quintile of socioeconomic status (ie, most advantaged), and similar to the general Australian population, three-quarters of the sample (18,094/24,035, 75.28%) lived in the eastern states of Australia (Victoria, New South Wales, and Queensland).

XSL•FO
RenderX

Table 1. Age, starting weight, BMI, and demographic characteristics of members at the start of the program.

	Longer-term members		All members (N=61,514)
	Study sample (n=24,035)	Subsample with complete data (n=6602)	
Membership duration (d), mean (SD)	600 (356)	663 (393)	328 (314)
Starting weight (kg), mean (SD)	92.6 (19.2)	92.7 (18.4)	91.1 (18.8)
Starting BMI (kg/m^2), mean (SD)	33.1 (6.2)	33.3 (6.0)	32.4 (6.0)
Age (y), mean (SD)	52 (11.6)	54 (11.2)	51 (12.4)
Sex, n (%)			
Female	19,972 (83.09)	5593 (84.72)	49,847 (81.03)
Male	4063 (16.9)	1009 (15.28)	11,667 (18.97)
Age group (years), n (%)			
18-30	1283 (5.34)	218 (3.3)	4711 (7.66)
31-50	9986 (41.55)	2291 (34.7)	26,232 (42.64)
51-70	12,033 (50.06)	3823 (57.91)	28,401 (46.17)
>71	705 (2.93)	263 (3.98)	2118 (3.44)
Missing	28 (0.12)	7 (0.11)	52 (0.08)
State, n (%)			
New South Wales	7894 (32.84)	2167 (32.82)	19,750 (32.11)
Victoria	5926 (24.65)	1657 (25.1)	14,998 (24.38)
Queensland	4274 (17.78)	1133 (17.16)	11,763 (19.12)
South Australia	2006 (8.35)	567 (8.59)	5161 (8.39)
Western Australia	1873 (7.79)	508 (7.69)	4849 (7.88)
Tasmania	566 (2.35)	153 (2.32)	1448 (2.35)
Northern Territory	263 (1.09)	66 (0.99)	532 (0.86)
Australian Capital Territory	979 (4.07)	275 (4.17)	2361 (3.84)
Missing or invalid	254 (1.06)	76 (1.15)	652 (1.06)
SEIFA[a]			
1 (lowest)	2248 (9.35)	604 (9.15)	5890 (9.57)
2	3492 (14.53)	931 (14.1)	8966 (14.57)
3	4542 (18.9)	1279 (19.37)	11,808 (19.19)
4	5203 (21.65)	1490 (22.57)	13,465 (21.89)
5 (highest)	8291 (34.5)	2220 (33.63)	20,721 (33.69)
Weight status			
Normal weight	957 (3.98)	177 (2.68)	3282 (5.33)
Overweight	7504 (31.22)	2054 (31.11)	21,046 (34.21)
Obese	15,546 (64.68)	4366 (66.13)	37,123 (60.35)
Class 1 obese	8070 (33.58)	2232 (33.81)	20,230 (32.89)
Class 2 obese	4410 (18.35)	1277 (19.34)	10,310 (16.76)
Class 3 obese	3066 (12.76)	857 (13.0)	6583 (10.7)
Missing or invalid	27 (0.11)	5 (0.07)	61 (0.1)

[a]SEIFA: Socio-Economic Indexes for Areas.

XSL•FO
RenderX

Longer-term members with complete data (6602/24,035, 27.47%) were slightly older than the longer-term members generally (54 vs 52 years), with 61.89% (4086/6602) of the members with complete data aged >51 years compared to 53% (12,738/24,035) of all longer-term members. However, their starting weight and BMI were similar, and the proportion classified as obese was also similar (4366/6602, 66.13% vs 15,546/24,035, 64.68%) to longer-term members generally (Table 1).

Compared to all members of the CSIRO Total Wellbeing Diet, longer-term members were, on average, 1.5 kg heavier at the start of the program and had a slightly higher BMI (33.1 vs 32.4 kg/m^2). The percentage of longer-term members classified as obese was 64.68% (15,546/24,035) compared to 60.35% (37,123/61,514) for all members (Table 1).

Weight Loss Over 12 Months

Using multiple imputations, the average weight loss of the longer-term members (24,035/24,035, 100%) was estimated as 5.9 kg (or 6.3%) at 12 weeks, 6.7 kg (or 7.2%) at 6 months, 6.2 kg (or 6.6%) at 9 months, and 5.5 kg (or 5.9%) at 12 months. Weight loss for male members was estimated to be 7.0 kg at 12 months compared to 5.2 kg for female members. At 12 months, 57.71% (2345/4063) of male members had lost ≥5% of their starting body weight compared to 51.21% (10,228/19,972) of female members, and 26.43% (1074/4063) of males and 23.99% (4791/19,972) of females had lost 10% of their starting body weight. Overall, it was estimated that 52.31% (12,573/24,035) and 24.4% (5865/24,035) of the members had lost 5% and 10% of their starting body weight, respectively (Table 2).

As a sensitivity analysis, the average weight loss was calculated using the weigh-in data available. The average weight loss of longer-term members was calculated to be 6.4 kg (or 6.8% of starting body weight) at 12 weeks, 7.1 kg (or 7.6%) at 6 months, 7.0 kg (or 7.4%) at 9 months, and 6.3 kg (or 6.7%) at 12 months (Table 3). At 12 months, male members lost an average of 7.7 kg (or 7.1%) and female members lost an average of 6 kg (or 6.6%). At 12 months, just more than half the members (5055/9413, 53.78%) had lost ≥5% of their starting body weight and slightly more than 1 in 4 members (2595/9413, 27.57%) had lost ≥10% of their starting body weight. There were slightly more male members than female members who achieved a weight loss of ≥5% after 12 months (815/1409, 57.84% vs 4250/8004, 53.1%). At 12 months, 29.88% (421/1409) of the male members had lost ≥10% of their starting body weight, which was similar to female members (2174/8004, 27.16%; Table 3).

Weight loss among those members with a valid weight in the system at all time points (6602/24,035, 27.47%) was 6.8 kg (or 7.2%) at 12 weeks, 8.5 kg (or 9%) at 6 months, 8.4 kg (or 8.9%) at 9 months, and 7.8 kg (or 8.2%) at 12 months. Male members in this subgroup lost 9.1 kg (or 8.4%) at 12 months and female members lost 7.5 kg (or 8.2%; Table 3).

Approximately 2 in 3 members (4232/6602, 64.1%) in this subgroup had lost ≥5% of their starting body weight at 12 months, and 1 in 3 members (2342/6602, 35.47%) had lost ≥10% of their starting body weight at 12 months (Table 3).

XSL•FO
RenderX

Table 2. Weight loss over 12 weeks at 3, 6, 9, and 12 months calculated using the multiple imputations (MI) method (N=24,035).

Sex, statistic, and time point	Long-term members: MI (no missing; SE)
Female (n=19,972)	
Weight loss (kg), mean (SE)	
12 weeks (n=19,972)	5.5 (0.0247)
3-6 months (n=19,972)	6.3 (0.0362)
6-9 months (n=19,972)	5.8 (0.0422)
9-12 months (n=19,972)	5.2 (0.0449)
Proportion of body weight lost (%), mean (SE)	
12 weeks (n=19,972)	6.1 (0.0260)
3-6 months (n=19,972)	7.0 (0.0381)
6-9 months (n=19,972)	6.4 (0.0447)
9-12 months (n=19,972)	5.7 (0.0480)
Proportion losing >5% TBW[a] (percentage of sample)	
12 weeks (n=12,473)	62.45
3-6 months (n=12,583)	63.00
6-9 months (n=11,224)	56.20
9-12 months (n=10,228)	51.21
Proportion losing >10% TBW (percentage of sample)	
12 weeks (n=2697)	13.5
3-6 months (n=5357)	26.8
6-9 months (n=5242)	26.2
9-12 months (n=4791)	24.0
Male (n=4063)	
Weight loss (kg), mean (SE)	
12 weeks (n=4063)	8.0 (0.0703)
3-6 months (n=4063)	8.7 (0.0977)
6-9 months (n=4063)	7.9 (0.1101)
9-12 months (n=4063)	7.0 (0.1133)
Proportion of body weight lost (%),mean (SE)	
12 weeks (n=4063)	7.4 (0.0615)
3-6 months (n=4063)	8.1 (0.0863)
6-9 months (n=4063)	7.3 (0.0979)
9-12 months (n=4063)	6.5 (0.1021)
Proportion losing >5% TBW (percentage of sample)	
12 weeks (n=2987)	73.52
3-6 months (n=2844)	70.00
6-9 months (n=2566)	63.15
9-12 months (n=2345)	57.72
Proportion losing >10% TBW (percentage of sample)	
12 weeks (n=1007)	24.78
3-6 months (n=1384)	34.06
6-9 months (n=1226)	30.17
9-12 months (n=1074)	26.43

XSL•FO
RenderX

Sex, statistic, and time point	Long-term members: MI (no missing; SE)
Total (n=24,035)	
Weight loss (kg), mean (SE)	
12 weeks (n=24,035)	5.9 (0.0245)
3-6 months (n=24,035)	6.7 (0.0348)
6-9 months (n=24,035)	6.2 (0.0400)
9-12 months (n=24,035)	5.5 (0.0421)
Proportion of body weight lost (%),mean (SE)	
12 weeks (n=24,035)	6.3 (0.0242)
3-6 months (n=24,035)	7.2 (0.0349)
6-9 months (n=24,035)	6.6 (0.0408)
9-12 months (n=24,035)	5.8 (0.0435)
Proportion losing >5% TBW (percentage of sample)	
12 weeks (n=15,460)	64.32
3-6 months (n=15,427)	64.19
6-9 months (n=13,790)	57.37
9-12 months (n=12,573)	52.31
Proportion losing >10% TBW (percentage of sample)	
12 weeks (n=3704)	15.41
3-6 months (n=6741)	28.05
6-9 months (n=6468)	26.91
9-12 months (n=5865)	24.40

[a]TBW: total body weight.

XSL•FO
RenderX

Table 3. Weight loss over 12 weeks at 3, 6, 9, and 12 months calculated using all available data and for members with complete data (n=6602).

Sex, statistic, and time point	Long-term members (with missing data)			Long-term members with complete data		
	Values, n	Values, mean (SE)	Percentage of sample	Values, n	Values, mean (SE)	Percentage of sample
Female						
Weight loss (kg)						
12 weeks	14,802	5.9 (0.0281)	—[a]	5593	6.4 (0.0454)	—
3-6 months	14,332	6.7 (0.0434)	—	5593	8.2 (0.0722)	—
6-9 months	10,266	6.7 (0.0622)	—	5593	8.1 (0.0866)	—
9-12 months	8004	6.0 (0.0757)	—	5593	7.5 (0.0925)	—
Proportion of body weight lost (%)						
12 weeks	14,802	6.6 (0.0287)	—	5593	7.1 (0.0453)	—
3-6 months	14,332	7.4 (0.0446)	—	5593	8.9 (0.0721)	—
6-9 months	10,266	7.3 (0.0637)	—	5593	8.8 (0.0871)	—
9-12 months	8004	6.6 (0.0775)	—	5593	8.2 (0.0935)	—
Proportion losing >5% TBW [b]						
12 weeks	10,123	—	68.5	4139	—	74.1
3-6 months	9459	—	66.1	4245	—	76.0
6-9 months	6152	—	60.0	3880	—	69.5
9-12 months	4240	—	53.0	3546	—	63.5
Proportion losing >10% TBW						
12 weeks	2277	—	15.4	1021	—	18.3
3-6 months	4234	—	29.5	2278	—	40.8
6-9 months	3172	—	30.9	2206	—	39.5
9-12 months	2174	—	27.2	1973	—	35.3
Male						
Weight loss (kg)						
12 weeks	3230	8.4 (0.0795)	—	1009	8.9 (0.1424)	—
3-6 months	2827	9.2 (0.1210)	—	1009	10.4 (0.2061)	—
6-9 months	1877	8.8 (0.1765)	—	1009	9.9 (0.2363)	—
9-12 months	1409	7.7 (0.2045)	—	1009	9.1 (0.2389)	—
Proportion of body weight lost (%)						
12 weeks	3230	7.8 (0.0689)	—	1009	8.3 (0.1234)	—
3-6 months	2827	8.5 (0.1056)	—	1009	9.6 (0.1788)	—
6-9 months	1877	8.1 (0.1535)	—	1009	9.1 (0.2064)	—
9-12 months	1409	7.1 (0.1833)	—	1009	8.4 (0.2108)	—
Proportion losing >5% TBW						
12 weeks	2477	—	76.8	817	—	81.0
3-6 months	2038	—	72.2	788	—	78.3
6-9 months	1231	—	65.6	718	—	71.1
9-12 months	815	—	57.8	677	—	67.2
Proportion losing >10% TBW						
12 weeks	918	—	28.4	330	—	32.7
3-6 months	1071	—	37.9	470	—	46.7

XSL•FO
RenderX

Sex, statistic, and time point	Long-term members (with missing data)			Long-term members with complete data		
	Values, n	Values, mean (SE)	Percentage of sample	Values, n	Values, mean (SE)	Percentage of sample
6-9 months	686	—	36.5	436	—	43.2
9-12 months	421	—	29.9	369	—	36.6
Total						
Weight loss (kg)						
12 weeks	18,032	6.4 (0.0280)	—	6602	6.8 (0.0455)	—
3-6 months	17,159	7.1 (0.0420)	—	6602	8.5 (0.0694)	—
6-9 months	12,143	7.0 (0.0597)	—	6602	8.4 (0.0821)	—
9-12 months	9413	6.3 (0.0716)	—	6602	7.8 (0.0867)	—
Proportion of body weight lost (%)						
12 weeks	18,032	6.8 (0.0268)	—	6602	7.2 (0.0431)	—
3-6 months	17,159	7.6 (0.0412)	—	6602	9.0 (0.0669)	—
6-9 months	12,143	7.4 (0.0589)	—	6602	8.9 (0.0802)	—
9-12 months	9413	6.7 (0.0714)	—	6602	8.2 (0.0855)	—
Proportion losing >5% TBW						
12 weeks	12,600	—	70.0	4956	—	75.2
3-6 months	11,497	—	67.1	5033	—	76.4
6-9 months	7383	—	60.8	4598	—	69.8
9-12 months	5055	—	53.8	4232	—	64.1
Proportion losing >10% TBW						
12 weeks	3195	—	17.7	1351	—	20.5
3-6 months	5305	—	30.9	2748	—	41.7
6-9 months	3858	—	31.8	2642	—	40.1
9-12 months	2595	—	27.6	2342	—	35.5

[a]Not applicable.

[b]TBW: total body weight.

Weight Loss Patterns Over 12 Months

Weight loss patterns were examined in 3-month blocks among members with a valid weight in the system at all time points (6602/24,035, 27.47%). The average weight loss of this group was 7.8 kg after 12 months of membership. The most common pattern of weight loss for these members over the first 12 months of membership was weight loss in the first two 3-month blocks (ie, first 6 months), followed by weight maintenance in the second two 3-month blocks of membership (ie, second 6 months). This pattern (L|L|M|M) was observed in 14.6% (961/6602) of the members, and this subgroup achieved an average weight loss of 10.7 kg at 12 weeks and 10.6 kg at 1 year (or 11.9% of their starting body weight; Table 4).

XSL•FO
RenderX

Table 4. Platform use (activity per week) and weight loss (kg), described in 3-month blocks over 12 months of membership, by weight loss patterna (n=6602).

Weight loss sequence[b]	Sample, n (%)	Starting body weight (kg)	Platform use (activity per week)				Weight loss (kg)				
			12 weeks	6 months	9 months	12 months	12 weeks	6 months	9 months	12 months	Percentage of body weight (12 months)
L|L|M|M	961 (14.6)	88.9	62.3	37.8	23.7	16.5	7.3	10.7	10.8	10.6	11.9
L|M|M|M	734 (11.1)	84.5	52.9	24.2	17.8	13.1	5.1	5.3	5.0	4.8	5.7
L|L|L|M	590 (8.9)	96.8	64.1	46.9	35.3	24.8	8.7	13.6	16.5	16.6	17.1
L|L|M|G	495 (7.5)	96.2	64.6	37.4	19.1	13.7	8.1	11.6	11.6	8.6	8.9
L|M|G|M	439 (6.6)	89.6	51.2	18.5	12.3	9.7	6.0	6.1	3.3	3.3	3.6
L|M|M|G	366 (5.5)	88.1	56.5	23.0	14.0	11.0	6.3	6.5	6.3	3.6	4.0
L|L|G|M	306 (4.6)	94.0	62.5	31.8	15.1	12.1	7.9	11.0	8.2	8.1	8.5
L|L|L|L	284 (4.3)	102.7	74.7	56.2	45.7	36.3	9.4	15.6	19.5	22.3	21.7
L|L|G|G	270 (4.1)	100.2	66.6	35.0	15.0	10.6	8.9	12.3	9.5	6.3	6.2
L|M|G|G	268 (4.1)	93.4	63.1	20.8	12.3	9.9	7.4	7.5	4.6	1.8	1.8
L|L|L|G	184 (2.8)	101.6	67.8	43.0	31.2	18.6	9.2	14.1	16.8	13.9	13.6
L|G|M|M	183 (2.8)	87.6	49.4	14.3	11.7	9.4	5.5	2.9	2.8	2.7	3.1
L|G|G|M	137 (2.1)	95.3	52.0	15.2	11.3	10.1	6.3	3.7	0.8	0.7	0.7
L|L|M|L	129 (2.0)	96.7	54.8	34.4	24.3	22.7	7.9	11.8	11.9	14.3	14.8

[a]Data are presented for weight loss sequence groups containing ≥2% of the total sample.

[b]Weight loss during each 3-month period was categorized as weight loss (L), weight gain (G), or weight maintenance (M). Consistent with previous research [25], a change in weight within 1.4 kg of previous weight was considered weight maintenance, and anything greater was considered weight loss or weight gain.

The second most common pattern was weight loss recorded in the first 3-month block and then maintenance for the remainder of the year. This group (L|M|M|M, 734/6602, 11.1% of this sample) lost an average of 5.1 kg in 12 weeks (5.7% of their starting body weight) and maintained this loss through to 1 year. The other common pattern was weight loss over the first three 3-month blocks, followed by weight maintenance (L|L|L|M, 590/6602, 8.94% of this sample). This group was heavier when they started the program than the more common L|L|M|M and L|M|M|M groups and lost 16.6 kg over 12 months, equivalent to 17.1% of their starting body weight. Although less common, those achieving weight loss in each 3-month block (L|L|L|L, 284/6602, 4.3% of this sample) were among the heaviest members at the start of the program, and they lost an average of 22.3 kg over 12 months, which was equivalent to 21.7% of their starting body weight (Table 4).

Platform Use by Weight Loss Pattern

There was a positive relationship between the average weekly platform use and weight loss over the first 12 months of membership, with higher weekly platform use associated with greater weight loss (r=0.287; P<.001). Table 3 shows the relationship between platform use, weight loss, and the pattern of weight loss. For example, weekly platform use in the group that consistently lost weight (L|L|L|L) was 24% to 56% higher at each time point throughout the year than other weight loss patterns where use and weight loss were lower. Figure 1 shows the platform use and weight loss at 12 months for each weight loss pattern. The quadrants of the box represent 3-month intervals, with darker colors indicating weight loss and lighter colors indicating weight gain. It was evident that when platform use exceeded ≥30 times per week approximately >4 times per day), the pattern of weight loss included loss for the first 6 months of the program, whereas the groups using the platform for <30 times week recorded a loss in the first 3 months, followed by a period of maintenance or gain (Figure 1). It is also evident that within clusters of weight loss patterns, for example, losing weight in 3 of the 4 blocks of time over 12 months, total weight loss was higher when use was higher (Figure 1).

XSL•FO
RenderX

Figure 1. Platform use and total weight loss over 12 months for the most common patterns of weight loss (n=6602).

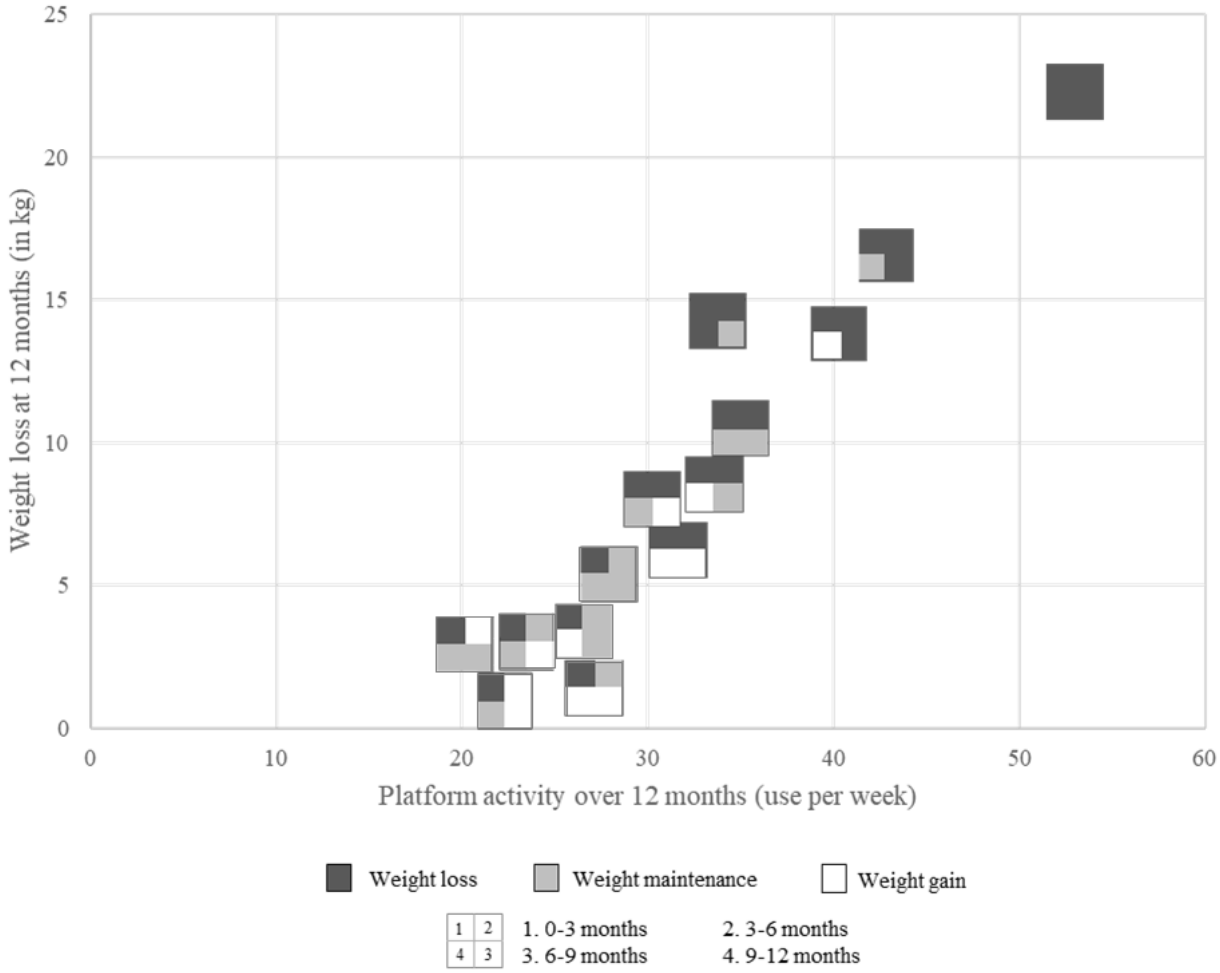

Discussion

Principal Findings

This study examined the longer-term weight loss of members of a commercial weight loss program, the CSIRO Total Wellbeing Diet Online. In a large sample of 24,035 members who had approximately 12 months of paid membership, the average estimated weight loss was 5.9 kg at 12 weeks, 6.7 kg at 6 months, 6.2 kg at 9 months, and 5.5 kg at 12 months. At 12 months, it was estimated that more than half the members (12,573/24,035, 52.3%) remained at least 5% below their starting body weight and 1 in 4 (5865/24,035, 24.4%) were at least 10% below their starting body weight. A modest reduction in weight of 5% has been accepted as clinically meaningful [26], and 5% to 10% is a recommended goal for medically supervised weight loss, as outlined in guidelines for managing overweight and obesity [27,28]. On a smaller subset of members with complete weigh-in data at 3, 6, 9, and 12 months, 64.1% (4232/6602) and 35.5% (2342/6602) reached the recommendations of 5% and 10% body weight loss, respectively. The most common weight loss pattern on this program was 6 months of weight loss, followed by a 6-month period of weight maintenance, with this group achieving a total of 10.6 kg weight loss over 12 months. Although less common, when weight loss was achieved in each of the four 3-month blocks of time, the average weight loss over 12 months was 22.3 kg. Achieving higher weight loss was associated with higher platform use, specifically at a frequency of approximately ≥4 times per day.

Comparison to Other Literature

The health benefits of modest weight loss are well established. A 5% body weight loss is associated with improved systolic and diastolic blood pressure, glucose control, and cholesterol [29]. However, achieving this level of weight loss is difficult. A cohort study of >18 million health care–seeking Americans found that 1 in 10 adults achieved a 5% body weight loss [30]. Among successful individuals, approximately half report to do it on their own and the other half through a formal program [31]. The do-it-yourself approach can work, and when in a controlled trial (which likely overrepresents effectiveness), up to 1 in 4 people were reported to achieve 5% body weight loss after 12 months [32]. Commercial weight loss programs are an effective option for weight loss, and digital programs, in particular, might be more accessible. The weight loss results of this program compare favorably to similar commercial or digital weight loss programs. For example, an evaluation of another commercial weight loss program, *WW* (formerly known as *Weight Watchers*), reported an average weight loss of 3.8 kg at 3 months and 4.4 kg at 12 months, with 42.8% of participants achieving a loss of 5% body weight at 12 months [32], compared to 52.3% (12,573/24,035) for the CSIRO Total Wellbeing Diet

XSL•FO
RenderX

Online. This is one of the few examples of published weight loss results from a commercial weight loss program with a longer follow-up, although the sample size is smaller than the one presented here. Another example of an evaluation of a commercial program is a cross-country comparison of *Noom Coach*, which included 18,459 participants and reported weight losses across countries ranging from 3% to 3.7% of starting body weight loss at 16 weeks, with 27% to 33% (cross-country range) achieving 5% body weight loss [33]. Meanwhile, *MetaWell*, a digital weight loss program, reported comparable short-term (ie, 3 months) weight loss results to the CSIRO Total Wellbeing Diet Online (5.6 vs 5.9 kg for CSIRO Total Wellbeing Diet) in a larger sample of approximately 250,000 members [34]; however, no longer-term follow-up data have been reported yet. Key strengths of this study are the large sample from a commercial weight loss program, with longer-term follow-up and the use of multiple imputations to account for missing data in the estimation of weight loss. Without the use of statistical methods to account for missing data, which can be significant for self-recorded weight data in commercial platforms, the weight loss results are likely to be more favorable and, therefore, less generalizable in terms of expected weight loss outcomes.

The results of this study are consistent with other studies where a weight loss "tick" curve is commonly reported, meaning that initial weight loss success is followed by a plateau at or approximately 6 months and then some weight regain by 12 months [9]. Although the average weight loss difference at 6 and 12 months on this program was 1.2 kg, this would be considered weight maintenance according to the standard definition [25]. Different weight loss patterns are rarely explored beyond the short term [35]. In this study, weight loss patterns were examined in 3-month blocks of time over 12 months. The 3 most common patterns, accounting for one-third of members in this subgroup analysis, involved various combinations of weight loss and weight maintenance, without periods of weight gain. These findings contrast with another study on weight loss patterns from an internet-based weight management program [36], which involved daily weighing over a 1-year period. It found a shift from weight loss directly to weight gain, with weight regain starting much earlier, approximately 11 weeks, and not preceded by a weight maintenance period [36]. Half the sample in this study achieved weight loss in the first 6 months, and most often, this was followed by additional weight loss or maintenance. When weight loss was recorded for ≥3 periods throughout 12 months, the average weight loss was approximately 15 kg regardless of whether the other was a period of maintenance, loss, or gain. This is an important message for the public when embarking on a weight loss journey. Periods of weight gain and maintenance are common, and the trajectory of a longer-term journey will likely not be linear; however, meaningful overall weight loss can still be achieved with time.

There is compelling evidence that increased self-monitoring is associated with greater success in weight management [37]. How to best promote and sustain self-monitoring, as well as the necessary dose required, remain areas of investigation. A previous weight loss study that tracked self-monitoring in a digital system showed that there were no significant differences in time spent self-monitoring based on amount of weight loss at 6 months; however, there were differences in the number of times people logged in. Those who lost >5% of their starting body weight logged into the system approximately 2.5 times per day compared to 1.5 times per day for those who lost <5%, and log-ins were 2.7 times versus 1.7 times per day for those losing >10% compared to <10% of their starting body weight [38]. This supports the results of this study, where those with the highest platform use had a pattern of sustained weight loss and a greater total weight loss over 12 months. These users were interacting with the platform >30 times a week (or >4 times per day), which is higher, but given that the CSIRO Total Wellbeing Diet Online system allows for weight, diet, and exercise monitoring, it would be expected that daily interaction is more regular.

There are several challenges with sustained engagement with digital programs, notably attrition [39]. These are made more significant by growing evidence that sustained engagement may be more important than self-monitoring alone. In a study comparing degrees of weight loss [40], highly successful people were those who lost more weight initially and recorded continued weight loss over time, while less successful people lost smaller amounts of weight initially and then regained half that weight back over time. In the initial 6-month period of that study, there were no differences in the frequency of self-monitoring for moderately and highly successful people; however, longer-term self-monitoring remained high for those in the most successful group, recording significantly more self-monitoring entries than other groups. Together with our observations, this suggests that sustained self-monitoring may be more important for longer-term weight loss results. Self-monitoring in digital programs is typically measured by interactions with the system, which can also serve as a proxy for motivation. A small study in women showed a sustained increase in autonomous motivation at 5% weight loss but a decline for failing to reach this milestone [41]. It also suggested that motivation at week 4 predicted adherence to self-monitoring and weight loss. Thus, achieving clinically significant weight loss may increase motivation, which then improves self-monitoring.

It is clear that digital interventions need to leverage self-monitoring to support more successful weight management. System interactions are one indication of a person's motivation, emotions, and actions, and it is likely that greater appreciation of these personal factors will be needed to address use attrition. Supporting this, a recent review of 497 self-management randomized controlled trials for obesity found that information sharing, self-monitoring, and goal setting were frequently used techniques in interventions, but a broader scope is needed in interventions to consider social support, emotional management, and shared decision-making in obesity self-management interventions, which are all important outcomes for patients [42]. Interventions for obesity will also need to evolve to responsibly incorporate artificial intelligence to enhance the user experience and support more people in their weight loss journey. In behavioral interventions, artificial intelligence could analyze patterns in datasets to predict behavioral and clinical

XSL•FO
RenderX

outcomes early, allowing for more personalized content that enhances learning and motivation [43]. This approach could potentially improve weight loss outcomes for people living with obesity. The use of artificial intelligence might also enhance the user experience throughout the varying stages of weight management [44]. Early weight loss success provides motivation for continued efforts and later success, but a greater understanding of the "ups and downs," or patterns, of longer-term weight loss is needed.

Strengths and Limitations

Published data on the longer-term weight loss results from real-world commercial programs are limited. A strength of this study is the large sample of members used to report weight loss results, which is significantly larger than those published for other commercial programs. Weight loss was calculated in several ways, including the use of multiple imputation to address missing data. Missing data can be problematic in real-world programs, and using only the available data may inflate weight loss results, as these data could represent more motivated individuals who use the platform features more often. Multiple imputations as an approach helps to reduce bias and accounts for some of the uncertainty in imputed values by creating multiple complete datasets, which are separately analyzed and combined. A total of 20 datasets were created with values imputed using several different demographic variables known to be associated with weight loss, such as age, sex, and starting weight, among others. As a result, the weight loss results presented are more likely to be reflective of what might be achieved by a member who joins the program in their own everyday context and not a selected subsample [16]. Another strength of this study was the examination of weight loss patterns over an extended period. This analysis was conducted in a subsample of members with complete data, but the sample was still large enough to examine weight loss patterns in four 3-month blocks over 12 months. To the authors' knowledge, no other commercial program has examined weight loss in this way. These results provide a deeper understanding of the cycles of weight loss success and are critical to managing the expectations of those embarking on a weight loss journey. The CSIRO Total Wellbeing Diet is delivered through a partnership between Australia's national science agency and a digital health company (Digital Wellness). This is a unique partnership in which both organizations are dedicated to enhancing engagement with the platform to improve weight loss success and retention in the program, offering mutual public health and commercial benefits.

There are some limitations of this study that should be considered. While the sample was large, a limitation was that most members were female, and individuals from the lowest socioeconomic groups were underrepresented compared to the general Australian population. The cost associated with joining the program might be a barrier for those with lower income. Exploring strategies to attract a diverse member base that is more representative of the broader Australian population is a focus of ongoing research. Although real-world trials provide insight into the "true" success of people who follow a program of their own volition, a second limitation is that real-world, commercial programs are limited in what data they can collect. The registration process for this program collects various demographic data, but other factors, such as past dieting history or motivation, were not collected. Such information may allow for better predictions of weight loss success. A third limitation is the measurement of engagement used in this study. Platform use provided an indication of engagement with the program, but more objective measures of compliance, such as dietary intake, may better represent program compliance.

Conclusions

This analysis presents one of the most comprehensive evaluations of a commercial weight loss program, providing insights on weight loss success over 1 year among a large cohort of members as well as examining patterns of weight loss in a smaller subsample to better understand the "ups and downs" in achieving longer-term weight loss success. Given the prevalence of obesity and the challenges in losing weight, more evidence-based programs that achieve meaningful weight loss are needed. These results show that the CSIRO Total Wellbeing Diet Online is effective for longer-term weight loss with approximately 1 in 2 members achieving clinically significant results after 1 year. Members who engaged with the platform more were able to achieve multiple periods of weight loss over the year, and this was associated with greater weight loss success overall. Future research should focus on understanding how to drive motivation for early and sustained engagement, knowing that self-monitoring is an important factor influencing this complex health challenge. In practice, the findings on the patterns of weight loss within a longer-term journey should be used to help prepare people for a realistic weight loss journey.

Acknowledgments

The authors acknowledge the team at Digital Wellness for their support in maintaining the CSIRO (Commonwealth Scientific and Industrial Research Organisation) Total Wellbeing Diet Online platform and assisting in the data management before analysis. This study was cofunded by CSIRO and Digital Wellness. Digital Wellness manages the CSIRO Total Wellbeing Diet Online under a commercial agreement. Authors from CSIRO designed the study, conducted the analysis, and were responsible for the interpretation of results and preparation of the manuscript. Digital Wellness was not involved in the analysis or preparation of results. Digital Wellness reviewed and approved the manuscript before submission but did not alter the text or change it in any way.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available because they are considered commercial-in-confidence but are available from the corresponding author on reasonable request.

Authors' Contributions

GAH contributed to the conceptualization, methodology, formal analysis, preparation of the original draft, review and editing of the draft, and project administration. DLB contributed to methodology, formal analysis, preparation of the original draft, and review and editing of the draft. GJ-M contributed to project administration, methodology, preparation of the original draft, and review and editing of the draft. PGB and EB contributed to methodology, preparation of the original draft, and review and editing of the draft.

Conflicts of Interest

None declared.

References

1. Obesity Atlas 2023. World Obesity Federation. URL: https://data.worldobesity.org/publications/?cat=19 [accessed 2023-10-10]
2. Waist circumference and BMI. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/waist-circumference-and-bmi/latest-release#key-statistics [accessed 2024-01-19]
3. Weighing in: Australia's growing obesity epidemic. The WELL Collaborative. URL: https://thewellcollaborative.org.au/weighing-in-australias-growing-obesity-epidemic [accessed 2023-10-10]
4. Weighing the cost of obesity: a case for action. Price Waterhouse Coopers. 2015 Oct. URL: https://www.pwc.com.au/pdf/weighing-the-cost-of-obesity-final.pdf [accessed 2023-10-10]
5. The National Obesity Strategy 2022-2032: enabling Australians to eat well and be active. Commonwealth of Australia. 2022. URL: https://www.health.gov.au/sites/default/files/documents/2022/03/national-obesity-strategy-2022-2032_0.pdf [accessed 2023-10-10]
6. Atallah R, Filion KB, Wakil SM, Genest J, Joseph L, Poirier P, et al. Long-term effects of 4 popular diets on weight loss and cardiovascular risk factors. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):815-827. [doi: 10.1161/circoutcomes.113.000723]
7. Gudzune KA, Doshi RS, Mehta AK, Chaudhry ZW, Jacobs DK, Vakil RM, et al. Efficacy of commercial weight-loss programs: an updated systematic review. Ann Intern Med 2015 Apr 07;162(7):501-512 [FREE Full text] [doi: 10.7326/M14-2238] [Medline: 25844997]
8. Johnston BC, Kanters S, Bandayrel K, Wu P, Naji F, Siemieniuk RA, et al. Comparison of weight loss among named diet programs in overweight and obese adults: a meta-analysis. JAMA 2014 Sep 03;312(9):923-933. [doi: 10.1001/jama.2014.10397] [Medline: 25182101]
9. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007 Oct;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
10. Idris I, Hampton J, Moncrieff F, Whitman M. Effectiveness of a digital lifestyle change program in obese and type 2 diabetes populations: service evaluation of real-world data. JMIR Diabetes 2020 Jan 20;5(1):e15189 [FREE Full text] [doi: 10.2196/15189] [Medline: 31958064]
11. Jolly K, Lewis A, Beach J, Denley J, Adab P, Deeks JJ, et al. Comparison of range of commercial or primary care led weight reduction programmes with minimal intervention control for weight loss in obesity: Lighten Up randomised controlled trial. BMJ 2011 Nov 03;343(nov03 2):d6500 [FREE Full text] [doi: 10.1136/bmj.d6500] [Medline: 22053315]
12. Hwang KO, Ning J, Trickey AW, Sciamanna CN. Website usage and weight loss in a free commercial online weight loss program: retrospective cohort study. J Med Internet Res 2013 Jan 15;15(1):e11 [FREE Full text] [doi: 10.2196/jmir.2195] [Medline: 23322819]
13. Lowe MR, Kral TV, Miller-Kovach K. Weight-loss maintenance 1, 2 and 5 years after successful completion of a weight-loss programme. Br J Nutr 2007 Nov 28;99(4):925-930. [doi: 10.1017/s0007114507862416]
14. Perna S, Salman M, Gasparri C, Cavioni A, Faliva MA, Mansueto F, et al. Two, six, and twelve-month dropout rate and predictor factors after a multidisciplinary residential program for obesity treatment. A prospective cohort study. Front Nutr 2022 May 27;9:851802 [FREE Full text] [doi: 10.3389/fnut.2022.851802] [Medline: 35711533]
15. Bazrafkan L, Choobineh MA, Shojaei M, Bozorgi A, Sharifi MH. How do overweight people dropout of a weight loss diet? A qualitative study. BMC Nutr 2021 Nov 19;7(1):76 [FREE Full text] [doi: 10.1186/s40795-021-00480-w] [Medline: 34794513]
16. May CN, Cox-Martin M, Ho AS, McCallum M, Chan C, Blessing K, et al. Weight loss maintenance after a digital commercial behavior change program (Noom Weight): observational cross-sectional survey study. Obes Sci Pract 2023 Oct;9(5):443-451 [FREE Full text] [doi: 10.1002/osp4.666] [Medline: 37810531]

XSL•FO
RenderX

17. Wing RR, Phelan S. Long-term weight loss maintenance. Am J Clin Nutr 2005 Jul;82(1 Suppl):222S-225S [FREE Full text] [doi: 10.1093/ajcn/82.1.222S] [Medline: 16002825]
18. Wing RR, Hill JO. Successful weight loss maintenance. Annu Rev Nutr 2001;21:323-341. [doi: 10.1146/annurev.nutr.21.1.323] [Medline: 11375440]
19. Varkevisser RD, van Stralen MM, Kroeze W, Ket JC, Steenhuis IH. Determinants of weight loss maintenance: a systematic review. Obes Rev 2019 Feb 16;20(2):171-211 [FREE Full text] [doi: 10.1111/obr.12772] [Medline: 30324651]
20. Clifton PM, Abbey M, Noakes M, Nestel PJ. Body fat distribution is a determinant of the lipoprotein response to dietary fat and cholesterol. Atherosclerosis 1994 Sep;109(1):P148. [doi: 10.1016/0021-9150(94)93604-8]
21. Farnsworth E, Luscombe ND, Noakes M, Wittert G, Argyiou E, Clifton PM. Effect of a high-protein, energy-restricted diet on body composition, glycemic control, and lipid concentrations in overweight and obese hyperinsulinemic men and women. Am J Clin Nutr 2003 Jul;78(1):31-39. [doi: 10.1093/ajcn/78.1.31] [Medline: 12816768]
22. Luscombe ND, Clifton PM, Noakes M, Farnsworth E, Wittert G. Effect of a high-protein, energy-restricted diet on weight loss and energy expenditure after weight stabilization in hyperinsulinemic subjects. Int J Obes Relat Metab Disord 2003 May 17;27(5):582-590. [doi: 10.1038/sj.ijo.0802270] [Medline: 12704402]
23. Noakes M, Clifton P. The CSIRO Total Wellbeing Diet Set. Clayton, Australia: CSIRO Publishing; 2006.
24. Hendrie GA, Baird DL, Brindal E, Williams G, Brand-Miller J, Muhlhausler B. Weight loss and usage of an online commercial weight loss program (the CSIRO total wellbeing diet online) delivered in an everyday context: five-year evaluation in a community cohort. J Med Internet Res 2021 Jun 07;23(6):e20981 [FREE Full text] [doi: 10.2196/20981] [Medline: 34096869]
25. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571. [doi: 10.1056/NEJMoa061883] [Medline: 17035649]
26. Williamson DA, Bray GA, Ryan DH. Is 5% weight loss a satisfactory criterion to define clinically significant weight loss? Obesity (Silver Spring) 2015 Dec;23(12):2319-2320 [FREE Full text] [doi: 10.1002/oby.21358] [Medline: 26523739]
27. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]
28. Clinical Practice Guidelines for the Management of Overweight and Obesity in Adults, Adolescents and Children in Australia. Melbourne, Australia: National Health and Medical Research Council; 2013.
29. Zomer E, Gurusamy K, Leach R, Trimmer C, Lobstein T, Morris S, et al. Interventions that cause weight loss and the impact on cardiovascular risk factors: a systematic review and meta-analysis. Obes Rev 2016 Oct 21;17(10):1001-1011. [doi: 10.1111/obr.12433] [Medline: 27324830]
30. Kompaniyets L, Freedman DS, Belay B, Pierce SL, Kraus EM, Blanck HM, et al. Probability of 5% or greater weight loss or BMI reduction to healthy weight among adults with overweight or obesity. JAMA Netw Open 2023 Aug 01;6(8):e2327358 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.27358] [Medline: 37548978]
31. Klem ML, Wing RR, McGuire MT, Seagle HM, Hill JO. A descriptive study of individuals successful at long-term maintenance of substantial weight loss. Am J Clin Nutr 1997 Aug;66(2):239-246. [doi: 10.1093/ajcn/66.2.239] [Medline: 9250100]
32. Tate DF, Lutes LD, Bryant M, Truesdale KP, Hatley KE, Griffiths Z, et al. Efficacy of a commercial weight management program compared with a do-it-yourself approach: a randomized clinical trial. JAMA Netw Open 2022 Aug 01;5(8):e2226561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.26561] [Medline: 35972742]
33. Yang Q, Mitchell ES, Ho AS, DeLuca L, Behr H, Michaelides A. Cross-national outcomes of a digital weight loss intervention in the United States, Canada, United Kingdom and Ireland, and Australia and New Zealand: a retrospective analysis. Front Public Health 2021;9:604937 [FREE Full text] [doi: 10.3389/fpubh.2021.604937] [Medline: 34178911]
34. Senecal C, Widmer RJ, Larrabee BR, de Andrade M, Lerman LO, Lerman A, et al. A digital health weight loss program in 250,000 individuals. J Obes 2020 Mar 26;2020:9497164 [FREE Full text] [doi: 10.1155/2020/9497164] [Medline: 32300485]
35. Kim HH, Kim Y, Michaelides A, Park YR. Weight loss trajectories and related factors in a 16-week mobile obesity intervention program: retrospective observational study. J Med Internet Res 2022 Apr 15;24(4):e29380 [FREE Full text] [doi: 10.2196/29380] [Medline: 35436211]
36. Ross KM, Qiu P, You L, Wing RR. Characterizing the pattern of weight loss and regain in adults enrolled in a 12-week internet-based weight management program. Obesity (Silver Spring) 2018 Feb;26(2):318-323 [FREE Full text] [doi: 10.1002/oby.22083] [Medline: 29239141]
37. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
38. Harvey J, Krukowski R, Priest J, West D. Log often, lose more: electronic dietary self-monitoring for weight loss. Obesity 2019 Feb 25;27(3):380-384. [doi: 10.1002/oby.22382]

XSL•FO
RenderX

39. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MD, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
40. Laitner MH, Minski SA, Perri MG. The role of self-monitoring in the maintenance of weight loss success. Eat Behav 2016 Apr;21:193-197 [FREE Full text] [doi: 10.1016/j.eatbeh.2016.03.005] [Medline: 26974582]
41. Webber KH, Tate DF, Ward DS, Bowling JM. Motivation and its relationship to adherence to self-monitoring and weight loss in a 16-week internet behavioral weight loss intervention. J Nutr Educ Behav 2010;42(3):161-167. [doi: 10.1016/j.jneb.2009.03.001] [Medline: 20138583]
42. Sunol R, González-González AI, Valli C, Ballester M, Seils L, Heijmans M, et al. Self-management interventions for adults living with obesity to improve patient-relevant outcomes: an evidence map. Patient Educ Couns 2023 May;110:107647 [FREE Full text] [doi: 10.1016/j.pec.2023.107647] [Medline: 36739705]
43. Bays HE, Fitch A, Cuda S, Gonsahn-Bollie S, Rickey E, Hablutzel J, et al. Artificial intelligence and obesity management: an Obesity Medicine Association (OMA) Clinical Practice Statement (CPS) 2023. Obes Pillars 2023 Jun;6:100065 [FREE Full text] [doi: 10.1016/j.obpill.2023.100065] [Medline: 37990659]
44. Rosenbaum M, Foster G. Differential mechanisms affecting weight loss and weight loss maintenance. Nat Metab 2023 Aug 23;5(8):1266-1274. [doi: 10.1038/s42255-023-00864-1] [Medline: 37612402]

Abbreviations

CSIRO: Commonwealth Scientific and Industrial Research Organisation

Edited by N Cahill; submitted 05.08.24; peer-reviewed by JF Hayes, R Bipat; comments to author 13.09.24; revised version received 04.10.24; accepted 11.11.24; published 15.01.25.

Please cite as:
Hendrie GA, Baird DL, James-Martin G, Brindal E, Brooker PG
Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study
J Med Internet Res 2025;27:e65122
URL: https://www.jmir.org/2025/1/e65122
doi:10.2196/65122
PMID:

©Gilly A Hendrie, Danielle L Baird, Genevieve James-Martin, Emily Brindal, Paige G Brooker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Evaluation of a Mindfulness-Based Mobile Intervention for Perinatal Mental Health: Randomized Controlled Trial

Sehwan Park[1,2], MA; Hee Young Cho[3], MD, PhD; Jin Young Park[4,5,6], MD, PhD; Kyungmi Chung[4,5,6], PhD; Kyungun Jhung[7], MD, PhD

[1]Biomedical Research Institute, Catholic Kwandong University International St. Mary's Hospital, Incheon, Republic of Korea

[2]Medical Research Team, Digital Medic co., Ltd., Seoul, Republic of Korea

[3]Department of Obstetrics and Gynecology, Seoul National University College of Medicine, Seoul, Republic of Korea

[4]Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, Yonsei University Health System, Yongin, Republic of Korea

[5]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Yonsei University Health System, Seoul, Republic of Korea

[6]Center for Digital Health, Yongin Severance Hospital, Yonsei University College of Medicine, Yonsei University Health System, Yongin, Republic of Korea

[7]Department of Psychiatry, Catholic Kwandong University International St. Mary's Hospital, Catholic Kwandong University College of Medicine, Incheon, Republic of Korea

Corresponding Author:
Kyungun Jhung, MD, PhD
Department of Psychiatry
Catholic Kwandong University International St. Mary's Hospital
Catholic Kwandong University College of Medicine
25, Simgok-ro 100 Beon-gil
Incheon, 22711
Republic of Korea
Phone: 82 1090563807
Email: kyungun12@gmail.com

Abstract

Background: Perinatal mental health problems, such as anxiety, stress, and depression, warrant particularly close monitoring and intervention, but they are often unaddressed in both obstetric and psychiatric clinics, with limited accessibility and treatment resources. Mobile health interventions may provide an effective and more accessible solution for addressing perinatal mental health. Development and evaluation of a mobile mental health intervention specifically for pregnant women are warranted.

Objective: This study aimed to evaluate the effectiveness of a 4-week, self-administered mobile mindfulness intervention in reducing anxiety, depression, and stress, and improving emotional well-being, maternal-fetal attachment, and mindfulness skills in a general population of pregnant women.

Methods: Pregnant women were recruited and randomized to an intervention or a wait-list control group. The intervention group participated in a self-administered 4-week smartphone-based mindfulness program. Anxiety, depression, and stress were assessed as primary outcomes at baseline and postintervention. Secondary outcomes were mental health well-being, maternal-fetal attachment, and skills of mindfulness. The usability of the mobile intervention was also evaluated.

Results: A total of 133 pregnant women were randomly assigned to the intervention (n=66) or the control group (n=67). The overall dropout rate was 30% (39/133). Anxiety scores of the intervention group significantly decreased from baseline to postintervention (*P*=.03, Wilcoxon Signed-Rank test), whereas no significant changes were observed in the control group. Depression and stress scores showed no significant changes. Emotional well-being significantly improved in the intervention group (*P*=.01). Improvements were observed in maternal-fetal attachment, particularly in attributing characteristics to the fetus (*P*=.003) and in differentiating the self from the fetus (*P*=.006). Mindfulness awareness also showed significant improvement (*P*=.008). Significant between-group effects were identified for mindfulness awareness (*P*=.006) and attributing characteristics to the fetus (*P*=.002). After applying the false discovery rate corrections, within-group improvements in emotional well-being,

XSL•FO
RenderX

maternal-fetal attachment, and mindfulness awareness remained significant, while between-group differences for emotional well-being and differentiation were not significant.

Conclusions: A mobile mindfulness program effectively reduced anxiety and improved emotional well-being, maternal-fetal attachment, and mindfulness awareness in the general population of pregnant women. Mobile interventions may offer a cost-effective and feasible method for promoting perinatal mental health.

Trial Registration: Clinical Research Information Service KCT0007166; https://tinyurl.com/458vfc4r

(J Med Internet Res 2025;27:e56601) doi:10.2196/56601

KEYWORDS

anxiety; perinatal mental health; depression; mobile health care; mindfulness; mobile phone

Introduction

Pregnancy is a major period of change in a woman's life. Perinatal women are reported to be at a higher risk of developing newly-onset mental health problems or experiencing a relapse of existing ones compared with the general population [1]. The percentage of women in pregnancy who meet 1 or more diagnostic criteria for an anxiety disorder ranges from 12.2% to 39%, which is 2-3 times higher than the general population [2]. Furthermore, the prevalence of antenatal depression is estimated at 20.7% for any depression and 15% for major depression [3]. Stress associated with pregnancy and childbirth is known to increase the risk of perinatal anxiety and depression, and the risk is increased in those with previous mental health vulnerabilities [4]. Concerning such widespread prevalence, the World Health Organization (2024) emphasizes the need for health care professionals to recognize and provide adequate interventions for mental health problems in pregnant women [5].

Mental health problems during pregnancy have far-reaching effects, influencing both the immediate and long-term outcomes of mothers and their children [6,7]. Antenatal depression and anxiety increase the risk of pregnancy complications and obstetric outcomes such as preeclampsia, edema, hemorrhage, preterm birth, and low birth weight [8]. Negative mental health conditions during pregnancy can hinder self-care, increasing risks such as poor nutrition and substance misuse. Consequently, these factors may harm the physical and mental health of the mother and restrict healthy fetal growth and development [9,10]. Antenatal depression also increases the risk of postnatal depression [11], which will further impact the health of both the mother and the infant. In the long term, untreated perinatal mental health conditions can lead to lasting consequences such as impaired neurocognitive development and poor emotional regulation during infancy and childhood [12-14].

Despite the significant impact of perinatal mental health issues, many women face barriers that prevent them from accessing necessary mental health care [15]. In obstetric clinics, routine mental health evaluations for anxiety, depression, and stress are often neglected or overlooked, and it is uncommon for pregnant women to seek mental health care directly from psychiatric services. Stigma further discourages help-seeking in pregnant women, and finally, there is a lack of available resources. Even when clinical problems are identified, many pregnant women are reluctant toward pharmacotherapy due to concerns about fetal safety and the potential effects of medication [16,17]. This low acceptability of pharmacotherapy emphasizes the need for nonpharmacological approaches, which many pregnant women may find more acceptable. However, access to nonpharmacological treatments, such as psychotherapy, remains limited due to resource constraints, including availability and cost. Given these barriers, there is a clear need for effective and more accessible nonpharmacological interventions for women during this critical period, including digital and mobile-based programs, that can overcome obstacles like stigma and limited resources.

The integration of digital technology in psychotherapeutic interventions for perinatal women has shown promising results. A systematic review and meta-analysis by Lau et al [18] showed that digital psychotherapeutic interventions may significantly reduce symptoms of depression, anxiety, and stress in perinatal women. This review emphasized that interventions delivered through digital platforms, such as websites and mobile apps, were particularly effective in improving psychological outcomes [18]. A mobile or web-based approach may alleviate space-time barriers, overcome women's resistance to treatment, and address limited resources. Research indicates that perinatal women frequently use mobile apps to access information about pregnancy, childbirth, and childcare [19]. These results, along with others, show that many women find mobile apps highly accessible and usable for improving their self-care and enhancing their pregnancy experience [20]. Mobile health apps have been reported to significantly enhance areas of both treatment and health care delivery [21]. Therefore, employing mobile and web-based technologies provides a promising way to address the unique needs of perinatal women, improving their access to adequate care.

Mindfulness-based interventions (MBIs) have emerged as an effective method for psychological self-care and symptom management [22]. By fostering conscious awareness and acceptance of one's present state, mindfulness equips individuals with the ability to navigate through physical and emotional distress [23]. Numerous studies highlight its efficacy in reducing anxiety and depression symptoms, as well as enhancing overall psychological well-being [24-28]. Mindfulness-based stress reduction and mindfulness-based cognitive therapy (MBCT) have shown efficacy in reducing anxiety and preventing depression relapse, as well as enhancing a broad range of other mental health problems [29,30].

XSL•FO
RenderX

Mindfulness training techniques through mindfulness-based stress reduction or MBCT have been shown to be effective for pregnant women in reducing stress, anxiety, depressive symptoms, and emotional distress during pregnancy [26,28,30,31]. A systematic review on the effectiveness of MBIs on maternal perinatal mental health outcomes showed reductions in perinatal anxiety and less consistent results for the effect on depression, but the review was limited by the number of studies available and their methodological quality [25]. Even though mindfulness interventions may be effective in perinatal mental health care, gold-standard mindfulness training often requires rigorous hours of in-person instruction and training. Thus, traditional mindfulness programs pose similar accessibility barriers to pregnant women.

Mobile-based MBIs may offer a promising solution to overcoming these barriers in perinatal mental health care [32-35]. Sun et al [33] implemented an 8-week mindfulness program, which was shown to be effective for pregnant women who screened positive for depressive symptoms. In the study, only 8% (7/84) completed the entire 8-week training program. As smartphone-based mobile interactions are mostly self-administered, the acceptability of the intervention by the participant is an important part of treatment viability. We aimed to address this point by developing a shorter, more usable self-administered mobile MBI that is effective for perinatal mental health. Furthermore, increased depression, anxiety, and stress during pregnancy are associated with negative prenatal outcomes, even at a subclinical level. Subclinical anxiety and stress were linked to increased risks of preterm birth, low birth weight, and lower Apgar scores [36]. In addition, mild depressive symptoms are reported to be associated with heightened cortisol levels, impacting fetal neurodevelopment and increasing the likelihood of neonatal complications [36]. This suggests that addressing even mild symptoms in pregnancy could improve maternal and neonatal health outcomes. Thus, in developing the intervention, we aimed to target the general population of pregnant women to improve their mental health and well-being, compared with previous studies that often focus on the clinical population.

In this randomized controlled trial, we evaluated the effectiveness of our newly developed, 4-week mindfulness-based mobile app in reducing perinatal anxiety, depression, and stress among a general population of pregnant women. We also investigated mental health well-being, maternal-fetal attachment, and mindfulness skills as secondary outcomes, and assessed the usability of the intervention to determine feasibility for broader application.

Methods

Trial Design and Randomization

This study was a single-center, randomized wait-list controlled trial with assessments at baseline and post intervention (trial registration KCT0007166). Participants were randomized using a list created in the R (4.0.2 version; R Core Team) program, which was set to randomize in a 1:1 ratio. To ensure the integrity of the data analysis and eliminate potential bias, the data analysts remained blinded to the group allocation of each participant. The randomization was performed by a designated member of the research team who was not involved in the data analysis. The randomization sheet was only accessible to the person responsible for assigning participants to their respective groups. A designated member of the research team allocated the participants to the intervention or control group using the randomized list. Those randomly assigned to the wait-list control group received the intervention after 4 weeks. The reporting of this trial adheres to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Multimedia Appendix 1).

Sample Size

The sample size was determined using G*Power (version 3.1.9.7), based on standard assumptions for detecting differences between 2 groups over time. The calculation assumed a small-to-medium effect size and a statistical power of 80%, with the required sample size estimated at 90 participants (45 per group) [18,25]. To account for potential attrition, which has been reported to reach up to 33% in similar interventions [18,25], the recruitment target was set at 134 participants.

Participants

Participants were recruited both from internet-based advertisement and on-site at the Obstetrics and Gynecology Department of the CHA Bundang Medical Center in Seongnam, South Korea, from March to November 2020. The hospital is the largest and first specialized center for pregnancy and women's health in the country. Informational posters displayed in the hospital and internet-based advertisements were used for recruitment. Potential participants were provided with detailed information about the study, and those interested were invited to contact the research team for further screening and enrollment. The inclusion criteria were as follows: (1) at least 18 years of age, (2) between 1 and 32 weeks of gestation, (3) having a smartphone, (4) being able to use the app on a smartphone for the study, (5) being able to read and understand Korean, (6) willing to be randomized, and (7) willing to provide informed consent. Exclusion criteria were as follows: (1) over 32 weeks of gestation and (2) expected to give birth within 4 weeks (during the intervention period). The study did not exclude participants based on medical history (eg, obstetric complications or mental illness) unless these conditions reached a clinically significant level that would affect participation or safety of the participant.

Upon obtaining informed consent and ensuring random allocation, all participants completed a baseline evaluation, including sociodemographic, medical, pregnancy-related, and mental health information. If participants could not visit the obstetric department or complete the questionnaire at the site, they received a Google survey link of the questionnaire through email or smartphone.

Participants assigned to the intervention group were instructed to download a mobile-based mindfulness intervention app named AvecMom (Digital Medic Co) and were instructed to use it for 4 weeks. They received printed guides on how to use the mobile app and direct download links for both Android (Google) and

XSL•FO
RenderX

iPhone (Apple Inc) devices. Participants were instructed to practice each practice session at least 3 times a week for the 4 weeks of participating in the program.

Participants assigned to the wait-list control group were instructed not to participate in any events of meditation, yoga, or mindfulness-based activities during the 4 weeks. After the 4-week period, all participants received a Google survey link for a postintervention assessment through email, messenger, or in person at the site. The control group received access to the mobile app with instructions after the 4 weeks and the postintervention assessment.

Ethical Considerations

This study was approved by the institutional review board, CHA Bundang Medical Center, CHA University College of Medicine (HI18C0911). All participants provided written informed consent before enrollment. Participants were fully informed about the study's purpose, procedures, potential risk, and their right to withdraw at any time without consequence. To ensure privacy and confidentiality, all data were deidentified before analysis and securely stored on encrypted servers, accessible only to authorized research staff. For data with potential reidentification risks, access was strictly limited to authorized research staff. Comprehensive data management protocols were implemented to prevent unauthorized access and ensure participant confidentiality throughout the study. Participants received pregnancy-related gifts, valued at approximately US $41 (50,000 KRW), as an incentive for their participation in the study.

Development of the Mobile-Based Mindfulness Intervention

The mindfulness-based mobile intervention was developed by an interdisciplinary research team. The development was led by a professor of psychiatry who has over 5 years of mindfulness experience as well as experience with leading the development of several preventive interventions for mental health-related at-risk populations. The team also consisted of a psychiatrist with expertise in mobile apps, obstetricians, psychologists, a research professor of user interface and user experience design, and research assistants with mindfulness experience. The development process of the intervention began with a comprehensive review of previous MBIs to identify essential components relevant to perinatal mental health [37]. Following this, we focused on component selection, choosing mindfulness practices that address perinatal stress, anxiety, and depression, with content tailored to meet the unique physical, social, and psychological needs of pregnant women. To ensure feasibility and accessibility, the program was designed as a 4-week intervention, a modification supported by existing research on shorter programs' effectiveness [38]. Finally, the program underwent expert review by a panel of experienced advisors in mindfulness and perinatal mental health, whose insights helped refine the content and structure. The recorded sessions were guided by an experienced mindfulness trainer who was also pregnant at the time, strengthening the intervention's relevance and the empathetic element of the intervention. Participants were reminded that this app is not equivalent to psychotherapy, and they were recommended to seek professional help when necessary.

The intervention program consisted of 4 sections—breathing mindfulness meditation, body scanning, emotional awareness, and self-kindness mindfulness (Multimedia Appendix 2). Each session was divided into 2 subsessions—instruction and practice sessions. In the instruction session, participants learned about mindfulness techniques according to the thematic curriculum and were guided through the mindfulness training. In the practice session, participants entered the mindfulness practice directly without the introduction content of the instruction session. Participants needed to complete the instruction session to enter the practice session. The sessions were designed to be completed sequentially, with the subsequent session automatically unlocking once the participant performs the previous session at least twice. This sequence was clearly outlined in the user manual provided to the participants.

Each session had a duration of about 20 minutes. Participants were instructed to practice each session at least twice, ensuring they engaged with the content multiple times to reinforce their learning and practice. This structured approach aimed to enhance usability and effectiveness, providing a comprehensive yet manageable intervention for pregnant women. Sessions took into consideration the prenatal-specific psychological, physical, and emotional changes specific to pregnancy. After each session, participants received a sequential worksheet to complement the session and aid in understanding and developing mindfulness skills (Multimedia Appendix 3).

Measures

Depression, Anxiety and Stress Scale–21 Items

Depression, anxiety, and stress were assessed with the 21-item Depression, Anxiety and Stress Scale (DASS-21) [39-41]. This scale consists of 21 items used to assess depression (7 items), anxiety (7 items), and stress (7 items). The response items were scored on a 4-point Likert scale ranging from 0 to 3. Higher scores for each factor indicate a higher intensity for each symptom. A depression score of 9 or higher, an anxiety score of 7 or higher, and a stress score of 14 or higher indicate moderate symptoms for each factor [40]. The internal reliability for DASS-21 was found to be high (α=.94, at baseline).

Edinburgh Postnatal Depression Scale

The Edinburgh Postnatal Depression Scale (EPDS) [42,43] was also used to assess depressive symptoms during the peripartum period. This scale contains 10 items scored using a 4-point Likert scale ranging from 0 to 3. A higher score indicates a higher intensity of depressive symptoms. A score of 8 or less on the recommended cutoff indicates no likelihood of depression [42]. The reliability of the EPDS for pregnant Korean women was good (Cronbach α=0.85) [43]. The scale reached good-to-excellent internal reliability in our sample (at baseline, Cronbach α=0.88). Although the EPDS is a frequently used measure to assess postnatal depression, it has been used in many studies as a valid screening tool to detect depression in pregnant women [44,45].

XSL•FO
RenderX

Mental Health Continuum-Short Form

Mental health well-being was measured using the Mental Health Continuum-Short Form (MHC-sf) [46,47]. This scale consists of 14 items measuring social (5 items), psychological (6 items), and emotional (3 items) well-being, which is scored using a 6-point Likert scale ranging from 0 to 5. Higher scores indicate better mental health. The 3 MHC-sf subscales had high test-retest reliability, discriminant validity, and convergent validity in Korean adults [47]. Adequate reliability was found in this study (Cronbach =0.096, at baseline).

Maternal-Fetal Attachment Scale

Attachment between the mother and fetus was assessed using the Maternal-Fetal Attachment Scale (MFAS) [48]. In this study, 23 items of the revised Korean version were used, which were scored using a 4-point Likert scale. The 5 subscales of the MFAS are designed to measure differentiation of self from the fetus (3 items), interaction with the fetus (4 items), attributing characteristics to the fetus (6 items), giving of self (4 items), and maternal role-taking (6 items). Construct and criterion validity and reliability have been reported for pregnant Korean women [49]. The study's α coefficient for the baseline MFAS score was 0.92.

Cognitive and Affective Mindfulness Scale-Revised Version

Mindfulness was assessed with the Cognitive and Affective Mindfulness Scale-revised (CAMS-r) [50,51]. The scale contains a 10-item inventory. The response items were scored on a 4-point Likert scale ranging from 1 to 4. The CAMS-r consists of 3 factors that assess mindfulness levels—awareness, attention, and acceptance. The awareness and attention factors comprise 4 items, and acceptance consists of 2 items. Higher scores indicated higher levels of mindfulness. The CAMS-r has good internal reliability (Cronbach α=0.73), internal validity (Cronbach α=0.70), and test-retest reliability (r=0.77, P<.001) among Korean undergraduate students [51]. This study's α coefficient for the baseline CAMS-r score was 0.86.

Usability

To assess the usability of the mindfulness mobile app, the Usefulness, Satisfaction, and Ease of Use questionnaire (USE questionnaire) [52] was used after 4 weeks of using the app. This scale consists of 30 items measuring usefulness (8 items), ease of use (4 items), ease of learning (11 items), and satisfaction (7 items). The scale used was a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7). The reliability of the questionnaire in this study was assessed using Cronbach α, with results demonstrating high internal consistency across all 4 dimensions (usefulness: 0.914, ease of use: 0.898, ease of learning: 0.883, satisfaction: 0.905). These findings are consistent with those reported by Gao et al [53], who found similarly high reliability (Cronbach α=0.98 overall).

Data Analysis

All statistical analyses were conducted using SPSS (version 18.0; IBM Corp) and Python 3 (Python Software Foundation). Demographic characteristics were analyzed in SPSS 18.0. An independent samples t test was used for continuous variables, while categorical variables were analyzed using the chi-square test to ensure baseline comparability between the intervention and control groups.

For demographic variables, missing data points related to smoking (1 participant) and alcohol use (2 participants) were imputed using the last observation carried forward method [54]. This method involved using the most recent patient records to impute the missing values, ensuring completeness of the demographic dataset for analysis.

The primary analysis was conducted using an intent-to-treat approach, ensuring that all randomized participants were included in the analysis based on their originally assigned groups. Due to the nonnormal distribution of the primary and secondary outcome data, nonparametric methods were applied using Python 3. The Mann-Whitney U test was used to compare outcomes between the Intervention and Control groups, while within-group changes from pre- to postintervention were assessed using the Wilcoxon Signed-Rank test.

Effect sizes for both the Mann-Whitney U and Wilcoxon tests were calculated using the following formula:



where Z is the test statistic and N is the number of observations. Effect sizes were interpreted based on the following guidelines: small (r=0.1), medium (r=0.3), and large (r=0.5).

To account for the nonparametric nature of the data, 95% CIs for the effect sizes were calculated using the bootstrapping method with 1000 resamples. These CIs provide a measure of precision, indicating the range within which the true effect size is likely to lie with 95% CI.

For secondary outcomes, multiple comparisons were addressed using the false discovery rate (FDR) correction method to control for Type I errors [55]. This method ensures that the statistical significance of the results is maintained while minimizing the risk of false positives due to multiple testing, particularly in the analysis of secondary outcomes.

All statistical results, including test statistics, P values, effect sizes, and 95% CIs, are reported to provide a comprehensive understanding of the intervention's effectiveness.

Results

Participant Enrollment and Flow

A total of 143 perinatal women were recruited as eligible for the study. Of the 143 participants, 2% (3/143) did not meet the inclusion criteria, and 5% (7/143) did not complete the baseline assessment. Thus, 133 participants were allocated randomly, of which 66 were allocated to the intervention group and 67 to the wait-list control group. During the follow-ups, 13% (17/133) did not complete the follow-up assessment and 17% (22/133) were not willing to continue participation. The overall completion rate was 71% (94/133; Figure 1; see Multimedia Appendix 1 for the CONSORT-EHEALTH checklist).

XSL•FO
RenderX

Figure 1. Participant enrollment flowchart.

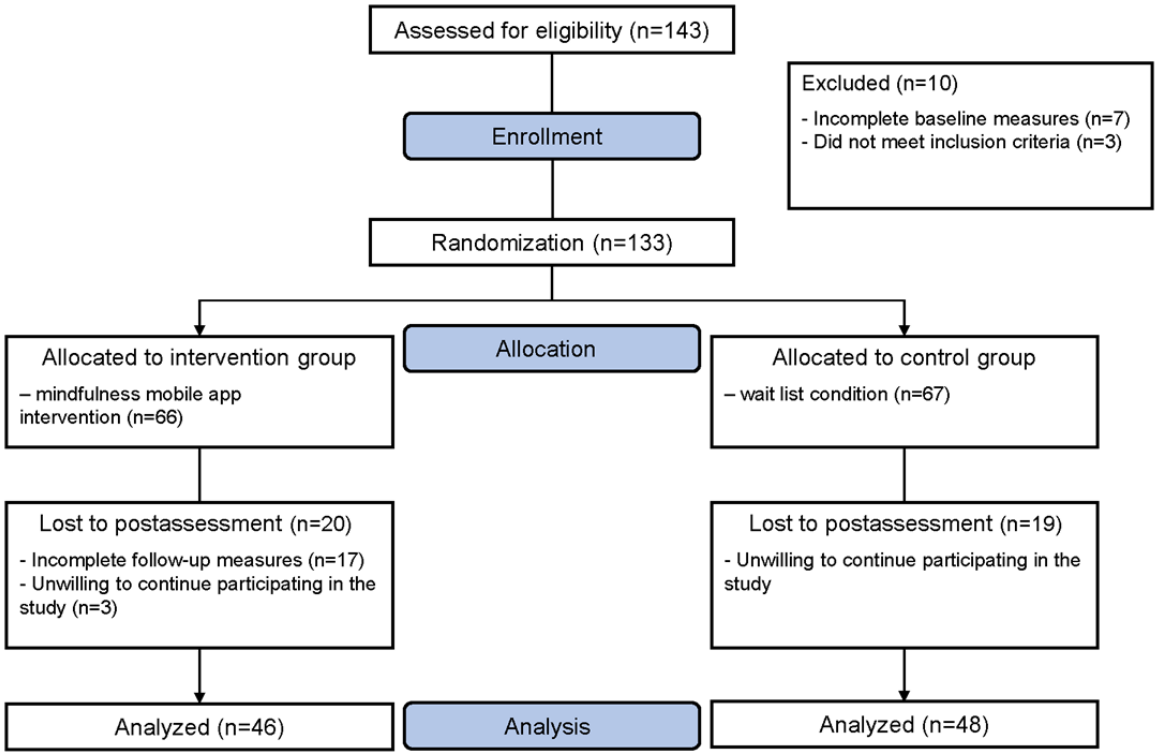

Demographics and Clinical Characteristics

Demographic characteristics are presented in Table 1. All 133 participants were pregnant and married at baseline. The mean ages of the participants for the intervention (n=66) and control groups (n=67) included in the analysis were 32.6 (SD 4.3) years and 32.8 (SD 4.44) years, respectively, and the mean gestational weeks of the 2 groups were 20.61 (SD 6.42) months and 20.04 (SD 6.76) months, respectively. There were no significant differences between the 2 groups across all the baseline characteristics, including pregnancy period, occupation, education level, income, pregnancy plan, medical diagnosis, and other characteristics.

XSL•FO
RenderX

Table 1. Demographics and clinical characteristics of study participants (n=133).

Characteristics	Intervention (n=66)	Control (n=67)	Test value (*df*)	*P* value
Age (years), mean (SD)	32.6 (4.3)	32.8 (4.44)	0.251[a] (92)	.80
Pregnancy period, n (%)			2.3[b] (2)	.32
First trimester (0-13 weeks)	15 (23)	11 (16)		
Second trimester (14-28 weeks)	43 (65)	42 (63)		
Third trimester (29-42 weeks)	8 (12)	14 (21)		
Occupation, n (%)			7.9[b] (7)	.34
Housewife	35 (53)	32 (48)		
White-collar worker	16 (24)	22 (33)		
Services	3 (4)	4 (6)		
Professional	7 (11)	5 (7)		
Technical	2 (3)	0 (0)		
Student	1 (2)	0 (0)		
Self-employed	2 (3)	1 (1)		
Other	0 (0)	3 (5)		
Education, n (%)			8.3[b] (3)	.08
Elementary School	0 (0)	1 (2)		
Middle School	4 (6)	0 (0)		
High school	13 (20)	7 (10)		
University	46 (70)	53 (79)		
Postgraduate degree	3 (4)	6 (9)		
Monthly Income (in US $)[c], n (%)			1.96[b] (4)	.74
≤820	0 (0)	1 (1)		
821-1639	1 (2)	2 (3)		
1640-2459	16 (24)	14 (21)		
2460-3279	21 (32)	18 (27)		
>3280	28 (42)	32 (48)		
Marriage status, n (%)			—[d]	—
Divorced, single, or separated	0 (0)	0 (0)		
Married	66 (100)	67 (100)		
Planned pregnant, n (%)			1.5[b] (1)	.23
No	44 (67)	51 (76)		
Yes	22 (33)	16 (24)		
Smoking, n (%)			0[b] (1)	.99
No	64 (97)	65 (97)		
Yes	2 (3)	2 (3)		
Alcohol use, n (%)			1.1[b] (1)	.29
No	60 (91)	64 (96)		
Yes	6 (9)	3 (4)		
Depression history, n (%)			.3[b] (1)	.57
No	57 (86)	60 (90)		

XSL•FO
RenderX

Characteristics	Intervention (n=66)	Control (n=67)	Test value (*df*)	*P* value
Yes	9 (14)	7 (10)		
Medical problems, n (%)			2.9[b] (1)	.09
No	52 (79)	60 (90)		
Yes	14 (21)	7 (10)		

[a]Independent *t* test.

[b]Chi-square test.

[c]All monetary values are presented in US $. The conversion rate used is US $1=1220.14 KRW, reflecting the exchange rate at the time of the study.

[d]Not applicable.

Effectiveness of the Mobile-Based Mindfulness Intervention

Primary Outcomes: Anxiety, Depression, and Stress

Table 2 shows the mean scores of pre- and postmeasures of the DASS-21, EPDS, CAMS-r, MHC-sf, and MFAS in the intervention and control groups. The DASS-21 mean scores of the intervention group at preintervention were 6.7 (SD 7.2), 6.9 (SD 5.5), and 9.7 (SD 8.2) for depression, anxiety, and stress, respectively. The mean scores of depression, anxiety, and stress of the compared control group were 7.4 (SD 7.7), 6.9 (SD 7.3), and 11.7 (SD 8.6), respectively. Considering the recommended cutoff scores from previous research, mean scores for depression, anxiety, and stress were in the normal range [42]. The baseline mean scores of the EPDS in the intervention and control groups were 7.1 (SD 5.1) and 7.6 (SD 6.3), respectively.

Table 2. Comparison of depression, anxiety, and stress scores at pre- and postintervention time points between the intervention and control groups using nonparametric analysis.

Scale, variable, and group	Pretest		Posttest		*P* value[a]	*P* value[b]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
DASS-21[c]						
Anxiety						
IG[d]	6.9 (5.5)	6 (2-10)	5.6 (5.8)	4 (2-7.5)	.03	.08
CG[e]	6.9 (7.3)	5 (2-8)	8.3 (8.2)	8 (2-10.5)	.11	—[f]
Depression						
IG	6.7 (7.2)	4 (2-9.5)	6.2 (6.3)	4 (2-8)	.60	.13
CG	7.1 (7.6)	4 (2-10.5)	8.5 (7.8)	7 (2-12)	.09	—
Stress						
IG	9.7 (8.2)	8 (4-13.5)	9 (8)	8 (2-13.5)	.48	.15
CG	11.7 (8.6)	10 (4-16)	11.6 (9.1)	10 (4-18)	.98	—
EPDS[g]						
IG	7.1 (5.1)	6 (4-9)	6.2 (4.6)	6 (2-9.75)	.16	.11
CG	7.6 (6.3)	6 (2-10.5)	8.3 (6)	8 (3.75-11.25)	.29	—

[a]*P* value for Wilcoxon Signed-Rank test for within-group comparisons (pre- to posttest).

[b]*P* value for Mann-Whitney *U* test for between-group comparisons (posttest).

[c]DASS-21: Depression, Anxiety and Stress Scale, 21-question version.

[d]IG: intervention group.

[e]CG: control group.

[f]Not applicable.

[g]EPDS: Edinburgh Postnatal Depression Scale.

Anxiety scores, measured using the DASS-21, significantly decreased in the intervention group from pre- to postintervention, with the median score dropping from 6 (IQR 2-10) to 4 (IQR 2-7.5), as shown by the Wilcoxon Signed-Rank test (*W*=295, *P*=.03, *r*=6.413, 95% CI –2 to 5). In contrast, no significant changes were observed in the control group, where anxiety scores increased from a median of 5 (IQR 2-8) to 8 (IQR 2-10.5; *W*=244, *P*=.11, *r*=5.083, 95% CI –5 to 1). The Mann-Whitney *U* test showed no significant difference between the groups at baseline (*U*=1191, *P*=.51, *r*=–0.068, 95% CI –2 to 4). Postintervention comparisons approached significance,

XSL•FO
RenderX

favoring the intervention group (*U*=869, *P*=.08, *r*=–0.183, 95% CI –6 to 0).

Depression scores, measured using the DASS-21, did not significantly change in the intervention group, with the median remaining at 4 (IQR 2-9.5) for preintervention and 4 (IQR 2-8) for postintervention (*W*=352.5, *P*=.60, *r*=7.663, 95% CI –4 to 4). Although the control group showed an increase in depression scores from a median of 4 (IQR 2-10.5) to 7 (IQR 2-12), this change was not statistically significant (*W*=238, *P*=.09, *r*=4.958, 95% CI –6 to 0). No significant difference between the groups was observed at baseline (*U*=1068.5, *P*=.79, *r*=–0.028, 95% CI –3 to 3), and postintervention group comparisons did not reach statistical significance (*U*=903.5, *P*=.13, *r*=–0.156, 95% CI –6 to 1).

Stress scores, measured using the DASS-21, remained stable in the intervention group, with a median score of 8 (IQR 4-13.5) for preintervention and 8 (IQR 2-13.5) for postintervention time points (*W*=321.5, *P*=.48, *r*=6.989, 95% CI –4 to 5). The control group also showed no significant change in stress scores, with medians of 10 (IQR 4-18) for preintervention and postintervention (IQR 4-16) time points (*W*=387.5, *P*=.98, *r*=8.073, 95% CI –5 to 6). The Mann-Whitney *U* test revealed no significant differences between the groups at baseline (*U*=951.5, *P*=.25, *r*=–0.119, 95% CI –8 to 2), nor postintervention time points (*U*=911.5, *P*=.15, *r*=–0.150, 95% CI –9 to 2).

The intervention group showed no significant changes in EPDS scores, with a median score of 6 (IQR 4-9) for preintervention and 6 (IQR 2-9.75) for postintervention (*W*=272.5, *P*=.16, *r*=5.924, 95% CI –3.5 to 3.0). In the control group, EPDS scores increased slightly from a median of 6 (IQR 2-10.5) to 8 (IQR 3.75-11.25), but this change was not significant (*W*=331.5, *P*=.29, *r*=6.906, 95% CI –5.0 to 2.5). Baseline comparisons showed no significant differences between the groups (*U*=1094.5, *P*=.95, *r*=–0.007, 95% CI –5 to 3), and postintervention group differences were also nonsignificant (*U*=888, *P*=.11, *r*=–0.169, 95% CI –5.0 to 1.5).

The primary outcomes results are summarized in Table 2 and depicted in Figure 2.

Figure 2. Comparison of median anxiety scores with IQR at pre- and postintervention between intervention and control groups using nonparametric analysis.

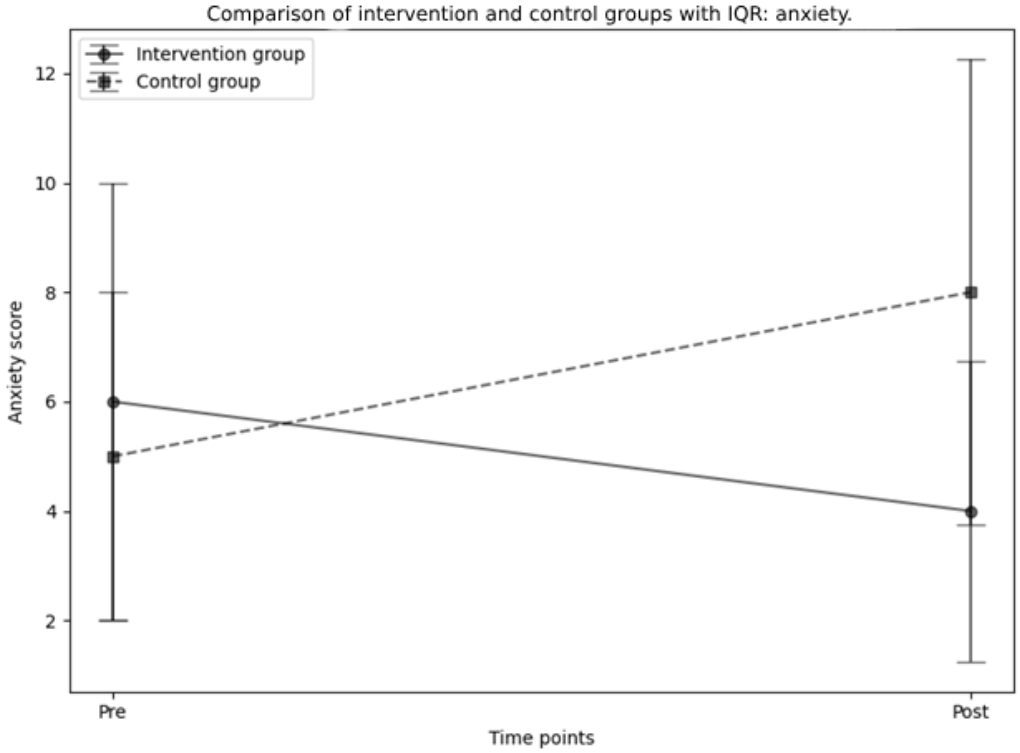

Secondary Outcomes: Mental Health Well-Being, Maternal-Fetal Attachment, and Mindfulness

The secondary outcomes including changes in well-being, maternal-fetal attachment, and mindfulness are summarized in Table 3. Emotional well-being significantly improved in the intervention group, with the median score increasing from 9 (IQR 8-12) to 11.5 (IQR 8-12) as reflected in the Wilcoxon Signed-Rank test (*W*=154, *P*=.02, effect size *r*=3.348, 95% CI –4.0 to 0.5). No significant changes were observed in the control group over the same period (*W*=287.5, *P*=.47, effect size *r*=5.99, 95% CI –1.5 to 2.5). The Mann-Whitney *U* test postintervention approached significance (*U*=1348, *P*=.07, effect size *r*=–0.19, 95% CI –0.5 to 4.0). According to the FDR correction, the within-group improvement in emotional well-being in the intervention group remained significant, while the between-group difference was not significant.

XSL•FO
RenderX

Table 3. Comparison of well-being, maternal-fetal attachment, and mindfulness scores at pre- and postintervention time points between the intervention and control groups using nonparametric analysis.

Scale, variable, and group	Pretest		Posttest		*P* value[a]	*P* value[b]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
MHC-sf[c] score						
Total						
IG[d]	35.2 (14.9)	35 (24-48)	38.9 (15.2)	43 (26-49.8)	.07	.06
CG[e]	34.8 (16.1)	31.5 (20.8-47.5)	32.3 (16.3)	31.5 (20.8-46)	.34	—[f]
Emotional						
IG	9.5 (2.9)	9 (8-12)	10.5 (3.0)	11.5 (8-12)	.02[g]	.07
CG	9.4 (3.3)	10 (7-12)	9.0 (3.6)	9 (6-12)	.47	—
Social						
IG	9.8 (6.1)	10.5 (5-14)	11.4 (5.9)	11 (8.2-16)	.03	.08
CG	10.0 (6.3)	9 (4.8-14.2)	9.4 (5.8)	9.5 (5-14)	.64	—
Psychological						
IG	15.9 (6.8)	15.5 (11-21)	17.0 (7.0)	19 (11.2-23)	.34	.06
CG	15.5 (7.6)	14.5 (9-21.2)	13.9 (7.6)	13 (7-21)	.19	—
MFAS score[h]						
Total						
IG	72.1 (11.2)	72.5 (65-80)	74.5 (11.4)	75.5 (68-83.8)	.15	.18
CG	71.1 (11.6)	71.5 (62.8-79.2)	70.9 (12.9)	73 (60.5-82)	.93	—
Differentiation self from fetus						
IG	9.0 (1.7)	9 (8-10)	9.8 (1.7)	10 (9-11)	.006[g]	.04
CG	9 (2.0)	9 (8-11)	8.9 (2.1)	9 (7-10.2)	.81	—
Interaction with fetus						
IG	12.7 (2.4)	13 (11-15)	12.3 (2.4)	13 (10.2-14)	.16	.41
CG	12.8 (2.5)	14 (11-15)	12.7 (2.5)	13 (11-15)	.57	—
Attributing characteristics to fetus						
IG	20.0 (2.8)	21 (18-22)	21.4 (2.7)	22 (20.2-24)	.003[g]	.002[g]
CG	19.4 (3.6)	20 (17-22)	19.5 (3.2)	20 (17.8-22)	.79	—
Role taking						
IG	13.1 (2.5)	13 (11-15)	13.4 (2.2)	14 (12-15)	.15	.26
CG	13.0 (2.6)	13.5 (11.8-15)	12.7 (2.8)	13 (10.8-16)	.41	—
Giving of self						
IG	17.3 (4.2)	17 (14-20.8)	17.6 (4.3)	18 (15-21)	.58	.46
CG	16.9 (3.3)	17 (15-19)	17.1 (4.1)	16.5 (14-21)	.62	—
CAMS-r[i] score						
Total						
IG	25.3 (5.9)	25 (21.2-29.8)	26.7 (7.0)	26.5 (21.2-31.8)	.25	.08
CG	24.7 (5.4)	26 (21.8-27)	24.2 (5.2)	24 (21-27.2)	.59	—
Awareness						
IG	9.9 (2.8)	10 (8-12)	11.2 (3.2)	11 (9-13.8)	.008[g]	.006[g]

XSL•FO
RenderX

Scale, variable, and group	Pretest		Posttest		*P* value[a]	*P* value[b]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
CG	9.4 (2.1)	9.5 (8.8-11)	9.4 (2.2)	10 (8-11)	.97	—
Acceptance						
IG	4.9 (1.5)	5 (4-6)	5.2 (1.5)	5 (4-6)	.23	.21
CG	4.8 (1.4)	5 (4-6)	4.8 (1.6)	4.5 (4-6)	.93	—
Attention						
IG	10.4 (2.6)	10 (9-12)	10.3 (2.6)	10 (8-12)	.42	.63
CG	10.5 (2.6)	10 (9-12.2)	10.0 (2.3)	10 (8-12)	.09	—

[a]*P* value for Wilcoxon Signed-Rank test for within-group comparisons (pre- to posttest).

[b]*P* value for Mann-Whitney *U* test for between-group comparisons (posttest).

[c]MHC-sf: Mental Health Continuum-Short Form.

[d]IG: intervention group.

[e]CG: control group.

[f]Not applicable.

[g]Significant changes after false discovery rate correction.

[h]MFAS: Maternal-Fetal Attachment scale.

[i]CAMS-r: Cognitive and Affective Mindfulness Scale-revised.

Differentiation self from fetus dimension of maternal-fetal attachment significantly improved in the intervention group, with median scores increasing from 9 (IQR 8-10) to 10 (IQR 9-11), as reflected in the Wilcoxon Signed-Rank test (*W*=129.5, *P*=.006, effect size *r*=2.815, 95% CI –1.5 to 0.0). No significant changes were found in the control group (*W*=206.5, *P*=.81, effect size *r*=4.302, 95% CI –0.5 to 0.0). Postintervention comparisons revealed a significant difference between groups, favoring the intervention group (*U*=1379, *P*=.04, effect size *r*=–0.215, 95% CI 0.0-1.5). After FDR correction, the within-group improvement in the intervention group remained significant, while the between-group difference was not significant.

The attributing characteristics to fetus dimension of maternal-fetal attachment significantly improved within the intervention group, with median scores rising from 21 (IQR 18-22) to 22 (IQR 20.25-24), as indicated by the Wilcoxon Signed-Rank test (*W*=107, *P*=.003, effect size *r*=2.326, 95% CI –3 to 0). No significant changes were observed in the control group (*W*=316, *P*=.79, effect size *r*=6.583, 95% CI –2.0 to 2.5). The Mann-Whitney *U* test revealed a significant postintervention difference between the groups, favoring the intervention group (*U*=1506.5, *P*=.002, effect size *r*=–0.314, 95% CI 0.5-4.0). After FDR correction, both the within-group improvement in the intervention group and the between-group difference remained significant.

Awareness from CAMS-r showed a significant increase in the intervention group from pre- to postintervention time points. The median score rose from 10 (IQR 8-12) to 11 (IQR 9-13.75), as indicated by the Wilcoxon Signed-Rank test (*W*=200, *P*=.008, effect size *r*=4.348, 95% CI –3.0 to 0.5). In contrast, the control group did not exhibit significant changes in awareness over the same period (*W*=406.5, *P*=.97, effect size *r*=8.469, 95% CI –1 to 1). Postintervention comparisons between the groups revealed a significant difference favoring the intervention group (*U*=1467.5, *P*=.006, effect size *r*=–0.284, 95% CI 0-3). The FDR correction confirmed that both the within-group improvement and the between-group difference were significant.

Subscales within the secondary outcomes, such as (1) social and psychological well-being (MHC-sf); (2) interaction with the fetus, giving of self, and maternal role-taking (MFAS); and (3) attention and acceptance in mindfulness (CAMS-r), did not show significant changes postintervention, as indicated by the FDR-adjusted analyses.

The comprehensive effects of both primary and secondary outcomes, including statistical significance and CIs, are detailed in Table 4.

XSL•FO
RenderX

Table 4. Effect sizes and 95% CIs for significant primary and secondary outcomes, comparing intervention and control groups.

Variable and group	Within comparison		Between comparison	
	Effect Size *r* (95% CI)	*P* value	Effect size *r* (95% CI)	*P* value
Anxiety				
IG[a]	6.413 (–2.0 to 5.0)	.03	–0.183 (–6.0 to 0.0)	.08
CG[b]	5.083 (–5.0 to 1.0)	.10	—[c]	—
Emotional well-being				
IG	3.348 (–4.0 to 0.5)	.01	–0.19 (–0.5 to 4.0)	.07
CG	5.99 (–1.5 to 2.5)	.47	—	—
Differentiation self from fetus				
IG	2.815 (–1.5 to 0.0)	.006	–0.215 (0.0 to 1.5)	.04
CG	4.302 (–0.5 to 0.0)	.81	—	—
Attributing characteristics to fetus				
IG	2.326 (–3.0 to 0.0)	.003	–0.314 (0.5 to 4.0)	.002
CG	6.583 (–2.0 to 2.5)	.79	—	—
Awareness				
IG	4.348 (–3.0 to 0.5)	.008	–0.284 (0.0 to 3.0)	.006
CG	8.469 (–1.0 to 1.0)	.96	—	—

[a]IG: intervention group.

[b]CG: control group.

[c]Not applicable.

Usability

Of the 133 participants, 80 responded to the USE questionnaire. Around 88% (70/80) of participants strongly agreed to the usability of the app. The total mean score was 5.7 (SD 0.62). Around 75% (60/80) of participants agreed regarding the app's usefulness. The mean score of usefulness was 5.4 (SD 0.79); 98% (78/80) of participants agreed that the app was easy to learn from. The mean score of ease of learning was 6.4 (SD 0.67); 91% (73/80) of participants agreed that the app was easy to use. The mean ease of use score was 5.9 (SD 0.68). In addition, 81% (65/80) of participants agreed with the satisfaction of the app. The mean score of the satisfaction was 5.6 (SD 0.79).

Discussion

Principal Findings

This study supports the effectiveness of the mobile-based mindfulness intervention developed for pregnant women in reducing perinatal anxiety. Furthermore, the intervention enhanced emotional well-being, maternal-fetal attachment, and mindful awareness during pregnancy. While previous studies have demonstrated the benefits of MBIs for perinatal mental health [25,31], our study is among the first to evaluate a fully self-guided, short-term (4-week) mobile intervention without the need for in-person guidance. This distinguishes our work as a widely applicable and accessible approach, particularly suited for populations with limited access to traditional mental health services. The findings provide evidence for both the effectiveness and usability of the mobile app as a tool for promoting mental health during pregnancy.

Perinatal anxiety was significantly reduced in pregnant women who used the self-administered mobile mindfulness program compared with the control group. Anxiety during pregnancy has been linked to various adverse outcomes, including mental health issues (such as postnatal depression) and perinatal complications (such as preterm birth). Our findings suggest that this intervention can help prevent and manage anxiety before it escalates to a clinical level. Even when anxiety is not at a clinical threshold, it can still have a significant impact, and interventions like ours can play an important role in addressing and controlling it early on. In a randomized controlled trial by Sun et al [33], there was a significant reduction in anxiety in pregnant women at risk of perinatal depression using an 8-week smartphone-based mindfulness training. Compared with these high-anxiety participants, those in this study were a general pregnant population with lower baseline anxiety levels. Our findings suggest that the mobile mindfulness intervention can effectively reduce anxiety even in a general pregnancy population.

Our 4-week intervention was designed to be more feasible and accessible, addressing the need for shorter, yet effective, interventions. This study stands out due to its shorter duration compared with similar interventions, such as the 8-week program used in previous studies [18,33]. A full 8-week MBCT program may provide a more comprehensive intervention for those willing to participate. In reality, however, committing to a long-term regular in-person treatment during pregnancy may

XSL•FO
RenderX

be very challenging with demands on the time and cost of the pregnant mothers, along with other barriers to treatment including the scarcity of perinatal treatment resources and stigma of going to psychiatric clinics. In previous research, the mean number of completed training weeks of the 8-week intervention was 3 [33]. Mobile-based or internet-based forms of treatment delivery may help overcome these limitations and promote access to timely interventions. Our study demonstrates that an MBCT-modified, short-term, self-administered mobile mindfulness intervention may benefit mental health during the perinatal period.

An analysis of the secondary outcomes from the mobile mindfulness intervention program revealed significant improvements in emotional well-being, a subscale of the well-being questionnaire (MHC-sf), among pregnant women following the 4-week intervention. Previous research has shown that poor maternal emotional well-being is associated with childbirth complications, such as low birth weight and fetal developmental issues, as well as long-term cognitive and behavioral problems in the child [56]. Enhanced emotional well-being has also been linked to better physical and mental health outcomes for pregnant women, improved labor outcomes, better health outcomes for offspring, and stronger maternal attachment [27,56]. Our findings suggest that mobile mindfulness training has the potential to promote emotional well-being during pregnancy and could serve as an accessible and effective tool for supporting mental health during this challenging period. In addition, the significant improvements in emotional well-being observed in the intervention group reinforce the effectiveness of MBIs in enhancing mental well-being, even in populations not at high risk for mental health disorders [56].

Another notable finding from our study was the positive effect of the mobile program on maternal-fetal attachment. Attachment, or the relationship between the mother and her infant, begins before birth (during pregnancy) as a result of dynamic psychological and physiological events. During pregnancy, women have intellectual, physical, and kinesthetic awareness of the fetus. Gradually, they become aware of the fetus as a separate individual. Women notice and interpret different movements of the fetus, elaborating personal relationships. Using the MFAS, domains of differentiation from the fetus and attribution of characteristics to the fetus were significantly increased in the intervention group. Differentiation of the self from the fetus indicates that the woman recognizes the fetus as an individual separate from her body [57]. The ability to attribute characteristics, such as personality, emotions, and physical space in the uterus, to the fetus is part of the dynamic relationship between mother and fetus [48,57]. Increased scores on these characteristics mean that mothers have a high level of attachment with their fetuses, by forming an identity as a mother and recognizing their developing fetuses; in turn, they may be more prepared and invested in their pregnancy [49]. Being in a mindful state means being aware of both the self and others, as well as one's inner and outer self. Mobile-based mindfulness training may lead to an increased awareness of the fetus as an individual, leading to qualitative changes in the interaction with the fetus. Improved maternal-fetal attachment has long-term benefits, such as a reduction in postpartum depression [57]. This study shows that a self-administered mindfulness program may promote maternal-fetal attachment, even in a relatively short duration of 4 weeks. Whether these effects may have a lasting, long-term benefit should be evaluated in the future.

Increased mindful awareness was shown in the intervention group, and supporting the newly developed 4-week mobile mindfulness program may cultivate or increase skills of mindfulness. Mindful awareness is nonjudgmental and nonreactive attention to experiences occurring at the present moment, including cognition, emotion, and bodily sensations, as well as surrounding environmental stimuli [58]. When one is mindfully aware, adaptation to stressful situations increases by enhancing attention to the negative emotional states [59]. Promotion of actual skills may have more persisting benefits, as these skills can be used in everyday life. Our findings suggest that these skills can be fostered through mobile-based, self-administration, even within a relatively short period of 4 weeks. Maintaining awareness, without judgment or reactivity, toward emotions about the dramatic changes that come with childbirth and related beliefs or automatic thoughts, may help in emotion regulation and behavioral adaptation. In a previous study that investigated the effect of the in-person MBCT program for 9 weeks increased natural childbirths for pregnant women with high fears of childbirth, increased mindful awareness was the most powerful mechanism of change to better adapt to the challenges of childbirth [60]. The present mobile intervention may be an accessible tool to effectively cultivate mindful awareness to increase tolerance to negative emotional states and cope with them better.

The mindfulness program scored high on usefulness, ease of use, ease of learning, ease to use, and overall satisfaction. Previously, it was noted that the traditional mindfulness intervention was challenging to access for women who already had children because of time limitations and existing family responsibilities, which led to a high loss to follow-up [60]. While mobile apps may be a preferable mode of delivery for mental health interventions to mitigate these barriers, the benefits of the intervention may be limited if it is difficult for users to administer. Current findings suggest that the mobile program is both feasible and acceptable. Furthermore, the responses from the usability study provided useful information for future program modifications. The mobile app used in our study automatically provides push notification services to serve the purpose of the mindfulness intervention program (eg, a reminder to use the app at least twice a week). This feature helped improve adherence for some participants. Features like these within a mobile app have the potential to enhance adherence and effectiveness, while a user-friendly design can facilitate the broad and efficient implementation of the app to provide motivation for continued use [61].

Limitations

Several limitations should be acknowledged. First, we only found results from the immediate effects of postintervention because our study design did not include a long-term follow-up. Future research should explore the long-term effects, potentially including postpartum outcomes for both the mother and infant,

XSL•FO
RenderX

to better inform perinatal care. Given the relatively short study period (4 weeks), it is important to determine whether the observed benefits are long-term, or if additional or booster sessions might be necessary to maintain or enhance these benefits. Second, our sample consisted of a general pregnancy population, and thus, the results of the current study do not represent effectiveness in a clinical sample such as that of a depressive disorder group [25]. No significant effect was observed on depression scores in this study, which may be attributed to the low scores that represent no clinical depression at baseline. It will be helpful to investigate whether the intervention can reduce clinical levels of anxiety or depression in the future. Finally, the effect sizes of depression, anxiety, and stress in this study were small, potentially due to the limited sample size. This also suggests that outcomes such as depression and stress may have shown no significant effect due to insufficient power. Future studies should aim to recruit a larger sample size to ensure adequate power to detect significant effects.

Conclusions

Mindfulness-based mobile interventions offer an effective tool for alleviating prenatal anxiety and improving emotional well-being as well as maternal-fetal attachment. This intervention may serve as a cost-effective and highly accessible method of intervention to overcome the current limitations of perinatal mental health management. Further research is needed to investigate whether the program provides lasting benefits for both mother and infant.

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant RS-2022-KH131144).

Data Availability

The datasets generated and/or analyzed during this study are not publicly available because of restrictions containing information that could compromise the privacy of study participants but are available from the corresponding author upon reasonable request.

Authors' Contributions

SP contributed to the study design, data acquisition, organization of the data, all statistical analyses, and interpretation of data, and wrote all sections of the manuscript. HYC was involved in the study concept, funding acquisition, recruitment of participants, and data acquisition. JYP was involved in the study concept, design, interpretation of data, and supervision. KC was involved in the protocol development, technical supervision, and interpretation of data. KJ was involved in the study concept, design, data acquisition, interpretation of data, preparation and review of the manuscript, and supervision. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 158 KB - jmir_v27i1e56601_app1.pdf]

Multimedia Appendix 2
Overview of the mindfulness-based intervention AvecMom mobile app content.
[DOCX File , 14 KB - jmir_v27i1e56601_app2.docx]

Multimedia Appendix 3
Screenshots of AvecMom mindfulness mobile app.
[DOCX File , 180 KB - jmir_v27i1e56601_app3.docx]

References

1. Austin MP, Highet N. Mental health care in the perinatal period: Australian clinical practice guideline. Expert Working Group. Melbourne: Centre of Perinatal Excellence; 2017. URL: https://www.cope.org.au/ [accessed 2017-10-17]
2. Goodman JH, Guarino A, Chenausky K, Klein L, Prager J, Petersen R, et al. CALM pregnancy: results of a pilot study of mindfulness-based cognitive therapy for perinatal anxiety. Arch Womens Ment Health 2014;17(5):373-387 [FREE Full text] [doi: 10.1007/s00737-013-0402-7] [Medline: 24449191]
3. Yin X, Sun N, Jiang N, Xu X, Gan Y, Zhang J, et al. Prevalence and associated factors of antenatal depression: systematic reviews and meta-analyses. Clin Psychol Rev 2021;83:101932. [doi: 10.1016/j.cpr.2020.101932] [Medline: 33176244]

XSL•FO
RenderX

4. Verkerk GJM, Pop VJM, Van Son MJM, Van Heck GL. Prediction of depression in the postpartum period: a longitudinal follow-up study in high-risk and low-risk women. J Affect Disord 2003;77(2):159-166. [doi: 10.1016/s0165-0327(02)00146-5] [Medline: 14607393]
5. World Health Organization. Maternal Mental Health. URL: https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/maternal-mental-health [accessed 2024-12-23]
6. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry 2016;3(10):973-982 [FREE Full text] [doi: 10.1016/S2215-0366(16)30284-X] [Medline: 27650773]
7. Shi Z, MacBeth A. The effectiveness of mindfulness-based interventions on maternal perinatal mental health outcomes: a systematic review. Mindfulness (N Y) 2017;8(4):823-847 [FREE Full text] [doi: 10.1007/s12671-016-0673-y] [Medline: 28757900]
8. Grigoriadis S, Graves L, Peer M, Mamisashvili L, Tomlinson G, Vigod SN, et al. Maternal anxiety during pregnancy and the association with adverse perinatal outcomes: systematic review and meta-analysis. J Clin Psychiatry 2018;79(5):813 [FREE Full text] [doi: 10.4088/JCP.17r12011] [Medline: 30192449]
9. Coverdale JH, McCullough LB, Chervenak FA, Bayer T. Clinical implications and management strategies when depression occurs during pregnancy. Aust N Z J Obstet Gynaecol 1996;36(4):424-429. [doi: 10.1111/j.1479-828x.1996.tb02185.x] [Medline: 9006826]
10. Zuckerman B, Amaro H, Bauchner H, Cabral H. Depressive symptoms during pregnancy: relationship to poor health behaviors. Am J Obstet Gynecol 1989;160(5 Pt 1):1107-1111 [FREE Full text] [doi: 10.1016/0002-9378(89)90170-1] [Medline: 2729387]
11. Alder J, Fink N, Bitzer J, Hösli I, Holzgreve W. Depression and anxiety during pregnancy: a risk factor for obstetric, fetal and neonatal outcome? A critical review of the literature. J Matern Fetal Neonatal Med 2007;20(3):189-209 [FREE Full text] [doi: 10.1080/14767050701209560] [Medline: 17437220]
12. Drury SS, Scaramella L, Zeanah CH. The neurobiological impact of postpartum maternal depression: prevention and intervention approaches. Child Adolesc Psychiatr Clin N Am 2016;25(2):179-200 [FREE Full text] [doi: 10.1016/j.chc.2015.11.001] [Medline: 26980123]
13. Buitelaar JK, Huizink AC, Mulder EJ, de Medina PG, Visser GH. Prenatal stress and cognitive development and temperament in infants. Neurobiol Aging 2003;24 Suppl 1:S53-60; discussion S67-68. [doi: 10.1016/s0197-4580(03)00050-2] [Medline: 12829109]
14. Edwards ES, Holzman JB, Burt NM, Rutherford HJ, Mayes LC, Bridgett DJ. Maternal emotion regulation strategies, internalizing problems and infant negative affect. J Appl Dev Psychol 2017;48:59-68 [FREE Full text] [doi: 10.1016/j.appdev.2016.12.001] [Medline: 28785122]
15. Smith MS, Lawrence V, Sadler E, Easter A. Barriers to accessing mental health services for women with perinatal mental illness: systematic review and meta-synthesis of qualitative studies in the UK. BMJ Open 2019;9(1):e024803 [FREE Full text] [doi: 10.1136/bmjopen-2018-024803] [Medline: 30679296]
16. Muzik M, Hamilton SE, Rosenblum KL, Waxler E, Hadi Z. Mindfulness yoga during pregnancy for psychiatrically at-risk women: preliminary results from a pilot feasibility study. Complement Ther Clin Pract 2012;18(4):235-240 [FREE Full text] [doi: 10.1016/j.ctcp.2012.06.006] [Medline: 23059438]
17. Goodman JH. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. Birth 2009;36(1):60-69. [doi: 10.1111/j.1523-536X.2008.00296.x] [Medline: 19278385]
18. Lau Y, Cheng JY, Wong SH, Yen KY, Cheng LJ. Effectiveness of digital psychotherapeutic intervention among perinatal women: a systematic review and meta-analysis of randomized controlled trials. World J Psychiatry 2021;11(4):133-152 [FREE Full text] [doi: 10.5498/wjp.v11.i4.133] [Medline: 33889538]
19. Lee Y, Moon M. Utilization and content evaluation of mobile applications for pregnancy, birth, and child care. Healthc Inform Res 2016;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
20. Tripp N, Hainey K, Liu A, Poulton A, Peek M, Kim J, et al. An emerging model of maternity care: smartphone, midwife, doctor? Women Birth 2014;27(1):64-67. [doi: 10.1016/j.wombi.2013.11.001] [Medline: 24295598]
21. Vital Wave Consulting. mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. Washington, D.C. and Berkshire: UN Foundation-Vodafone Foundation Partnership; 2009. URL: http://www.globalproblems-globalsolutions-files.org/unf_website/assets/publications/technology/mhealth/mHealth_for_Development_full.pdf [accessed 2024-08-21]
22. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
23. Kabat-Zinn J, Hanh TN. Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness. Delta: Bantam; 2013.
24. Bamber MD, Schneider JK. Mindfulness-based meditation to decrease stress and anxiety in college students: a narrative synthesis of the research. Educational Research Review 2016;18:1-32. [doi: 10.1016/j.edurev.2015.12.004]

XSL•FO
RenderX

25. Taylor LB, Cavanagh K, Strauss C. The effectiveness of mindfulness-based interventions in the perinatal period: a systematic review and meta-analysis. PLoS One 2016;11(5):e0155720 [FREE Full text] [doi: 10.1371/journal.pone.0155720] [Medline: 27182732]
26. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
27. Dunn C, Hanieh E, Roberts R, Powrie R. Mindful pregnancy and childbirth: effects of a mindfulness-based intervention on women's psychological distress and well-being in the perinatal period. Arch Womens Ment Health 2012;15(2):139-143. [doi: 10.1007/s00737-012-0264-4] [Medline: 22382281]
28. Eberth J, Sedlmeier P. The effects of mindfulness meditation: a meta-analysis. Mindfulness 2012;3(3):174-189 [FREE Full text] [doi: 10.1007/s12671-012-0101-x]
29. Zarenejad M, Yazdkhasti M, Rahimzadeh M, Mehdizadeh Tourzani Z, Esmaelzadeh-Saeieh S. The effect of mindfulness-based stress reduction on maternal anxiety and self-efficacy: a randomized controlled trial. Brain Behav 2020;10(4):e01561 [FREE Full text] [doi: 10.1002/brb3.1561] [Medline: 32162450]
30. Dimidjian S, Goodman SH, Felder JN, Gallop R, Brown AP, Beck A. An open trial of mindfulness-based cognitive therapy for the prevention of perinatal depressive relapse/recurrence. Arch Womens Ment Health 2015;18(1):85-94. [doi: 10.1007/s00737-014-0468-x] [Medline: 25298253]
31. Dhillon A, Sparkes E, Duarte RV. Mindfulness-based interventions during pregnancy: a systematic review and meta-analysis. Mindfulness (N Y) 2017;8(6):1421-1437 [FREE Full text] [doi: 10.1007/s12671-017-0726-x] [Medline: 29201244]
32. Goetz M, Schiele C, Müller M, Matthies LM, Deutsch TM, Spano C, et al. Effects of a brief electronic mindfulness-based intervention on relieving prenatal depression and anxiety in hospitalized high-risk pregnant women: exploratory pilot study. J Med Internet Res 2020;22(8):e17593 [FREE Full text] [doi: 10.2196/17593] [Medline: 32780023]
33. Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F. Effectiveness of smartphone-based mindfulness training on maternal perinatal depression: randomized controlled trial. J Med Internet Res 2021;23(1):e23410 [FREE Full text] [doi: 10.2196/23410] [Medline: 33502326]
34. Plaza I, Demarzo MMP, Herrera-Mercadal P, García-Campayo J. Mindfulness-based mobile applications: literature review and analysis of current features. JMIR Mhealth Uhealth 2013;1(2):e24 [FREE Full text] [doi: 10.2196/mhealth.2733] [Medline: 25099314]
35. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
36. Li H, Bowen A, Bowen R, Muhajarine N, Balbuena L. Mood instability, depression, and anxiety in pregnancy and adverse neonatal outcomes. BMC Pregnancy Childbirth 2021;21(1):583 [FREE Full text] [doi: 10.1186/s12884-021-04021-y] [Medline: 34429072]
37. Kim J. Mindfulness Meditation Manual. Seoul, Korea: Sol Science; 2016:69-140.
38. Warriner S, Crane C, Dymond M, Krusche A. An evaluation of mindfulness-based childbirth and parenting courses for pregnant women and prospective fathers/partners within the UK NHS (MBCP-4-NHS). Midwifery 2018;64:1-10. [doi: 10.1016/j.midw.2018.05.004] [Medline: 29843001]
39. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd Edition. Sydney: Psychology Foundation of Australia; 1995.
40. Lovibond P, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
41. Cha ES, Kim KH, Erlen JA. Translation of scales in cross-cultural research: issues and techniques. J Adv Nurs 2007;58(4):386-395. [doi: 10.1111/j.1365-2648.2007.04242.x] [Medline: 17442038]
42. Cox J, Holden J. Perinatal Mental Health: A Guide to the Edinburgh Postnatal Depression Scale (EPDS. London: Royal College of Psychiatrists; 2003.
43. Han KW, Kim MJ, Park JM. The Edinburgh Postnatal Depression Scale, Korean version: reliability and validity. J Korean Soc Biol Ther Psychiatry 2004;10(2):201-207.
44. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) EPDS Group. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ 2020;371:m4022 [FREE Full text] [doi: 10.1136/bmj.m4022] [Medline: 33177069]
45. Rubertsson C, Börjesson K, Berglund A, Josefsson A, Sydsjö G. The Swedish validation of Edinburgh Postnatal Depression Scale (EPDS) during pregnancy. Nord J Psychiatry 2011;65(6):414-418. [doi: 10.3109/08039488.2011.590606] [Medline: 21728782]
46. Lamers SMA, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CLM. Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF). J Clin Psychol 2011;67(1):99-110. [doi: 10.1002/jclp.20741] [Medline: 20973032]
47. Lim YJ. Psychometric characteristics of the Korean Mental Health Continuum–Short Form in an adolescent sample. Journal of Psychoeducational Assessment 2013;32(4):356-364. [doi: 10.1177/0734282913511431]

XSL•FO
RenderX

48. Cranley MS. Development of a tool for the measurement of maternal attachment during pregnancy. Nurs Res 1981;30(5):281-284. [Medline: 6912989]
49. Lee KS, Sig KM, Hong SJ. The validation study of Maternal Fetal Attachment Scale. Korean J Dev Psychol 2013;26(4):81-91 ISSN: 1229-0718 included for reference. No DOI is available as this is a 2013 publication from Korea, which typically does not assign DOIs to articles from this period [FREE Full text]
50. Feldman G, Hayes A, Kumar S, Greeson J, Laurenceau J. Mindfulness and emotion regulation: the development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). J Psychopathol Behav Assess 2006;29(3):177-190. [doi: 10.1007/s10862-006-9035-8]
51. Yongrae C. The reliability and validity of a Korean version of the Cognitive and Affective Mindfulness Scale-Revised. korean journal of clinical psychology 2009;28(3):667-693. [doi: 10.15842/kjcp.2009.28.3.002]
52. Lund AM. Measuring usability with the use questionnaire12. Usability Interface 2001;8(2):3-6.
53. Gao M, Kortum P, Oswald FL. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of use) questionnaire for reliability and validity. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2018;62(1):1414-1418 [FREE Full text] [doi: 10.1177/1541931218621322]
54. Judd LL, Rapaport MH, Yonkers KA, Rush AJ, Frank E, Thase ME, et al. Randomized, placebo-controlled trial of fluoxetine for acute treatment of minor depressive disorder. American Journal of Psychiatry 2004;161(10):1864-1871. [doi: 10.1176/appi.ajp.161.10.1864]
55. Benjamini Y, Cohen R. Weighted false discovery rate controlling procedures for clinical trials. Biostatistics 2017;18(1):91-104 [FREE Full text] [doi: 10.1093/biostatistics/kxw030]
56. Atif N, Lovell K, Rahman A. Maternal mental health: the missing "m" in the global maternal and child health agenda. Semin Perinatol 2015;39(5):345-352. [doi: 10.1053/j.semperi.2015.06.007] [Medline: 26164538]
57. Delavari M, Mohammad-Alizadeh-Charandabi S, Mirghafourvand M. The relationship of maternal-fetal attachment and postpartum depression: a longitudinal study. Arch Psychiatr Nurs 2018;32(2):263-267. [doi: 10.1016/j.apnu.2017.11.013] [Medline: 29579522]
58. Baer RA. Mindfulness training as a clinical intervention: a conceptual and empirical review. Clinical Psychology: Science and Practice 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
59. Veringa-Skiba IK, Ziemer K, de Bruin EI, de Bruin EJ, Bögels SM. Mindful awareness as a mechanism of change for natural childbirth in pregnant women with high fear of childbirth: a randomised controlled trial. BMC Pregnancy Childbirth 2022;22(1):47 [FREE Full text] [doi: 10.1186/s12884-022-04380-0] [Medline: 35045820]
60. Matvienko-Sikar K, Lee L, Murphy G, Murphy L. The effects of mindfulness interventions on prenatal well-being: a systematic review. Psychol Health 2016;31(12):1415-1434. [doi: 10.1080/08870446.2016.1220557] [Medline: 27539908]
61. Dagklis T, Tsakiridis I, Chouliara F, Mamopoulos A, Rousso D, Athanasiadis A, et al. Antenatal depression among women hospitalized due to threatened preterm labor in a high-risk pregnancy unit in Greece. J Matern Fetal Neonatal Med 2018;31(7):919-925 [FREE Full text] [doi: 10.1080/14767058.2017.1301926] [Medline: 28277908]

Abbreviations

CAMS-r: Cognitive and Affective Mindfulness Scale-revised
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DASS-21: 21-item Depression, Anxiety and Stress Scale
EPDS: Edinburgh Postnatal Depression Scale
FDR: false discovery rate
MBCT: mindfulness-based cognitive therapy
MBI: mindfulness-based intervention
MFAS: Maternal-Fetal Attachment scale
MHC-sf: Mental Health Continuum-Short Form
USE: Usefulness, Satisfaction, and Ease of Use

Edited by A Mavragani; submitted 03.05.24; peer-reviewed by K Lee, F Mao; comments to author 26.05.24; revised version received 29.09.24; accepted 28.11.24; published 17.01.25.

Please cite as:
Park S, Cho HY, Park JY, Chung K, Jhung K
Development and Evaluation of a Mindfulness-Based Mobile Intervention for Perinatal Mental Health: Randomized Controlled Trial
J Med Internet Res 2025;27:e56601
URL: https://www.jmir.org/2025/1/e56601
doi:10.2196/56601
PMID:

XSL•FO
RenderX

©Sehwan Park, Hee Young Cho, Jin Young Park, Kyungmi Chung, Kyungun Jhung. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Outcomes and Quality of Life for Patients With Hand and Foot Eczema: Randomized Study of a Patient-Centered Monitoring App

Aimee Bruch[1], MD; Wanja Weigandt[1], MD; Yannic Schardt[1], MD; Raphael Herr[2], PhD; Johannes Benecke[1], MD; Astrid Schmieder[3], MD

[1]Department of Dermatology, Venerology and Allergology, University Medical Center Mannheim, Heidelberg University, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany

[2]Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[3]Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Würzburg, Germany

Corresponding Author:
Astrid Schmieder, MD
Department of Dermatology, Venereology and Allergology
University Hospital Würzburg
Josef-Schneider-Strasse 2
Würzburg, 97080
Germany
Phone: 49 +4993120126234
Email: schmieder_a@ukw.de

Abstract

Background: Hand and foot eczema is a frequent chronic dermatological condition. The persistent itching, pain, and blistering can impair hand and foot function, leading to difficulties in performing tasks requiring fine motor skills. In addition, the impact on the quality of life for affected patients is significant, as the symptoms can be extremely uncomfortable and disruptive to daily activities. By incorporating digital health apps and educational programs into the management of hand and foot eczema, patients may receive ongoing support, optimize their clinical outcomes, and ultimately enhance their overall quality of life.

Objective: The purpose of this study was to evaluate the effect of a smartphone app combined with educational training on the clinical outcomes and mental health of patients with chronic hand and foot eczema during a 60-week study period.

Methods: Patients in the intervention group participated in an educational program focused on chronic hand and foot eczema at baseline and had in-person visits at weeks 0, 12, 24, 36, and 60, as well as access to our study smartphone app. The app allowed patients to upload pictures of their hands and feet and answer questions about pain severity, itching, mood, and quality of life. A chat function was also available for patients to contact their dermatologist. The control group received only the in-person study visits described above.

Results: A total of 87 patients were included in the study and randomized to the intervention (n=43) or control (n=44) groups. In total, 23 patients from the intervention group and 34 patients from the control group completed the study. Throughout the 60-week study period, a significant reduction in Hand Eczema Severity Index (HECSI) was consistently observed in all patients (week 60: linear regression coefficient [Coef]=–1.108; $P \leq .001$). A trend toward a greater improvement of the HECSI in the intervention group compared to the control group was noticed (week 60: Coef=0.597; $P=.05$). Subgroup analysis revealed that patients who used the app with a usage frequency of less than 20% demonstrated a significant reduction in the HECSI from week 0 to week 60 (week 60: Coef=–1.275; $P=.04$) and a significant reduction in the Dermatology Life Quality Index (week 60: Coef=–1.246; $P=.04$) compared to the control group. We were able to demonstrate a significant correlation between the HECSI calculated based on pictures uploaded by patients through the app and the HECSI assessed during personal visits ($\rho=0.885$; $P<.001$), despite the potentially lower image quality of the pictures uploaded through the app.

Conclusions: This study provides further evidence that digital health apps can provide valuable support in improving patient clinical outcomes and management, especially as the app-based assessment of hand and feed images appears to be reliable.

Trial Registration: Deutsches Register Klinischer Studien DRKS00020963; https://drks.de/search/de/trial/DRKS00020963

XSL•FO
RenderX

(J Med Internet Res 2025;27:e52159) doi:10.2196/52159

KEYWORDS

hand and foot eczema; eHealth; mHealth; teledermatology; telemedicine; disease management; smartphone application; mental health; eczema; clinical outcome; quality of life; dermatology; pain; motor skills; educational; support; mobile phone

Introduction

Hand eczema is one of the most common dermatologic conditions responsible for a very large proportion of dermatological consultations. According to a systematic review and meta-analysis, the 1-year prevalence is estimated to be 9.1% in the general population [1].

The probability of developing hand eczema at least once in a lifetime is 17% [2]. This likelihood varies greatly depending on occupational groups. Occupations that involve a lot of manual labor or chronic exposure to mild toxins or irritants, such as hairdressers and bricklayers, have a higher chance of developing hand eczema at least once in their lifetime [2]. The incidence of the disease is currently on the rise due to improved hygiene standards, increased life expectancy, and more frequent atopic predisposition [2].

Hand eczema is considered chronic if it lasts longer than 3 months or is recurrent more than twice a year [3]. Clinical manifestations are usually highly variable. In the acute stage, macules, papules, edema, vesicles, or even bullae may appear, whereas in the chronic stage, scales and crusts, hyperkeratosis, rhagades, and lichenification prevail [4]. These skin lesions can cause itching, burning, and pain, thereby leading to sleep disturbances and mood changes [5]. Chronic hand eczema is of high health-economic and sociomedical importance, as this diagnosis is responsible for many cases of prolonged sick leave and avoidance of social and public life [4,6]. In particular, the hands are an important organ of communication that is difficult to conceal from the public. Stigmatization thus causes significant psychosocial distress and a reduction in quality of life [7,8].

As hand and foot eczema is an intermittent disease, it is often our clinical experience that patients miss the optimal time to seek medical attention. Appointments are frequently not available on short notice, leading to patients presenting with very vigorous symptoms. This highlights the need for an easily accessible alternative for people with chronic hand and foot eczema such as a digital health app.

Germany recently launched the directory for Digital Health Applications (DiGA), which lists Conformité Européenne–marked medical devices. These apps are approved by the Federal Institute for Drugs and Medical Devices and are designed to improve patient outcomes through digital means, offering support in managing medical conditions, tracking symptoms, and providing tailored care plans. Such DiGA can be prescribed by the doctor and are paid by German health insurance companies. Although chronic hand and foot eczema is a dermatological disease easy to monitor through pictures taken by the patients, there are currently no specific DiGA available that offer support for these patients [9]. Furthermore, to the best of our knowledge, scientific data on the beneficial effects of DiGA for patients with hand and foot eczema is limited. This lack of evidence highlights a critical need for further research, especially given the prevalence of these conditions.

We developed a monitoring app for patients with chronic hand and foot eczema, enabling them to regularly photograph their condition and answer disease-specific questions. The app also facilitated teledermatology appointments tailored to individual patient needs.

To evaluate the effectiveness of this tool for chronic hand and foot eczema, we conducted an interventional study comparing a control group (without the app or educational program) with an intervention group over a 60-week period. Interim results from a 24-week analysis already showed an improvement in quality of life [10].

The main objective of the 60-week study was to assess the impact of regular physician-patient interactions and a prestudy educational program on disease progression, quality of life, and overall patient outcomes over the long-term period of 60 weeks.

Methods

Study Design

This intervention study was undertaken at the Department of Dermatology, Venerology and Allergology at the University Medical Center Mannheim, Germany, between August 2018 and August 2021. This is the analysis of the data from the study weeks 0 (V0), 12 (V2), 24 (V3), 36 (V4), and 60 (V5). The inclusion criteria that had to be fulfilled by the patients in order to be eligible to participate in the study were (1) a physician-confirmed diagnosis of chronic hand and foot eczema, (2) possession of a smartphone and being able to use it, (3) an age between 18 and 75 years, as well as (4) the ability to provide informed consent. Exclusion criteria were the inability to provide informed consent and being younger than 18 years old or older than 75 years old. During the first study visit (V0), patients were randomly assigned to the control or intervention group in a 1:1 ratio by shuffling a deck of cards.

A total of 90 patients were included in the study. Of the total, 43 patients were in the intervention group, 44 in the control group, and 3 patients dropped out from the study before being assigned to the respective group.

The control group started the first visit at week 0. At the first visit, a detailed medical history including sociodemographic data and eczema-related data were collected. In addition, the Dermatology Life Quality Index (DLQI; 0-30) was assessed and the Hand Eczema Severity Index (HECSI; 0-36) was calculated. Using a numerical rating scale (0-10), each patient was asked how much the eczema negatively affects activities and mood, and how much the eczema currently itches or hurts. Face-to-face follow-up was performed at weeks 12 (V2), 24 (V3), 36 (V4), and 60 (V5).

XSL•FO
RenderX

In addition to this first face-to-face visit, the intervention group participated in an educational program. This 2-hour educational session provided patients with information on the etiology, pathogenesis, exacerbating factors, and treatment of chronic hand and foot eczema. It was conducted by experienced specialists in dermatology (AS and JB). The patients had the opportunity to exchange information with specialists. In addition, each patient in the intervention group received a personal anonymized access code to our app, Dermascope Mobile (DermaIntelligence GmbH), as well as instructions on how to use it. Screenshots of the app can be found in the article by Domogalla et al [11].

Participants were asked to take pictures of their eczema once a week through the app and answer questions about their current state of health, including their psyche, and complete questionnaires on quality of life (DLQI) and current symptoms. The app also had a chat function with which patients could contact their treating dermatologist. The app could not be used more than once a week.

Each image uploaded through the app was classified by us into good or bad quality depending on the sharpness of the image, the light conditions, and whether you could see the whole hand or foot or not. An image could only be rated as good if every criterion was positively evaluated. Based on these images, we calculated the electronic HECSI (eHECSI), which we then compared to the HECSI assessed in person.

Ethical Considerations

The Medical Ethics Committee of the Medical Faculty Mannheim, Heidelberg University approved the study (ethics approval 2017-655N-MA), and the implementation complied with the Helsinki Declaration. Informed consent was obtained from all participants before the study, and in the case of secondary data analysis, the original consent allowed for such use without requiring additional permissions. To protect privacy and confidentiality, all data were deidentified, with access restricted to authorized personnel. Participants received no compensation except from a good medical care, and no identifying information or images of individuals were included. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist can be found in Multimedia Appendix 1.

Statistical Analysis

Linear panel data regression analyses estimated outcome trajectories. Linear regression coefficients (Coef) used throughout the text describe the mathematical relationship between each independent and dependent variable, while *P* values indicate whether these relationships are statistically significant. Random effects regressions determined the main and interaction effects of group membership (intervention vs control) and visit time (V1, V2, V3, V4, and V5) on DLQI, pain, activities of daily living, and HECSI scores. In total, 2 adjustment models were calculated. The first model was unadjusted, whereas the second model was adjusted for sex, age, and disease duration. Additional analyses included the effect of app usage frequency over 60 weeks (group membership: control vs <20% app usage frequency vs ≥20% app usage frequency). Therefore, the intervention group was divided into 2 groups—one consisting of patients with app use frequency <20% and the other consisting of patients with app use frequency ≥20% during the 24-week observation period. The chosen cutoff of 20% corresponds to an app use frequency of once every 5 weeks. Variables were tested for normal distribution and, where appropriate, transformed to approximate normal distribution (power transformation of the square root of the DLQI and log10 of the HECSI). All statistical analyses were performed with Stata Special Edition (version 14.0; StataCorp).

Results

Patient Demographics

Of the 90 patients included in the study, 43 patients were in the intervention group and 44 in the control group (Table 1, Figure 1). In addition, 3 patients dropped out of the study before being assigned to the respective group. Furthermore, 57 patients completed the study and could be included in the final analysis, 23 patients were from the intervention group, and 34 were from the control group. The most common reasons for withdrawal were lack of time, improvement of local findings, or long distance to the clinic.

XSL•FO
RenderX

Table 1. Patient characteristics at week 0 (V0) adapted from Weigandt et al [10].

Characteristics	Overall (n=87)	Control (n=44)	Intervention (n=43)
Sex, n (%)			
Female	51 (59)	25 (57)	26 (60)
Male	36 (41)	19 (43)	17 (40)
Age (years)			
Mean (SD)	47.07 (15.42)	48.05 (14.09)	46.07 (16.78)
Median	50	51	49
BMI (kg/m^2)			
Mean (SD)	27.62 (7.53)	26.45 (5.27)	28.82 (9.2)
Median	2578	25.16	26.81
Smoker, n (%)	29 (33)	16 (36)	13 (30)
Duration of eczema (years)			
Mean (SD)	6.9 (8.23)	6.0 (8.47)	7.81 (7.98)
Median	4	3	6
HECSI[a] (range 0-360)			
Mean (SD)	22.53 (21.29)	20.93 (20.72)	24.16 (21.99)
Median	18	15	19
DLQI[b] (range 0-30)			
Mean (SD)	7.97 (6.38)	7.73 (7.16)	8.21 (5.55)
Median	6	6	8
Pain (range 0-10)			
Mean (SD)	1.94 (2.67)	2.14 (2.77)	1.74 (2.59)
Median	0	1	0
Activity (range 0-10)			
Mean (SD)	4.02 (3.23)	3.95 (3.37)	4.09 (3.12)
Median	4	4	4

[a]HECSI: Hand Eczema Severity Index.

[b]DLQI: Dermatology Life Quality Index.

XSL•FO
RenderX

Figure 1. Flow diagram of the study cohort.

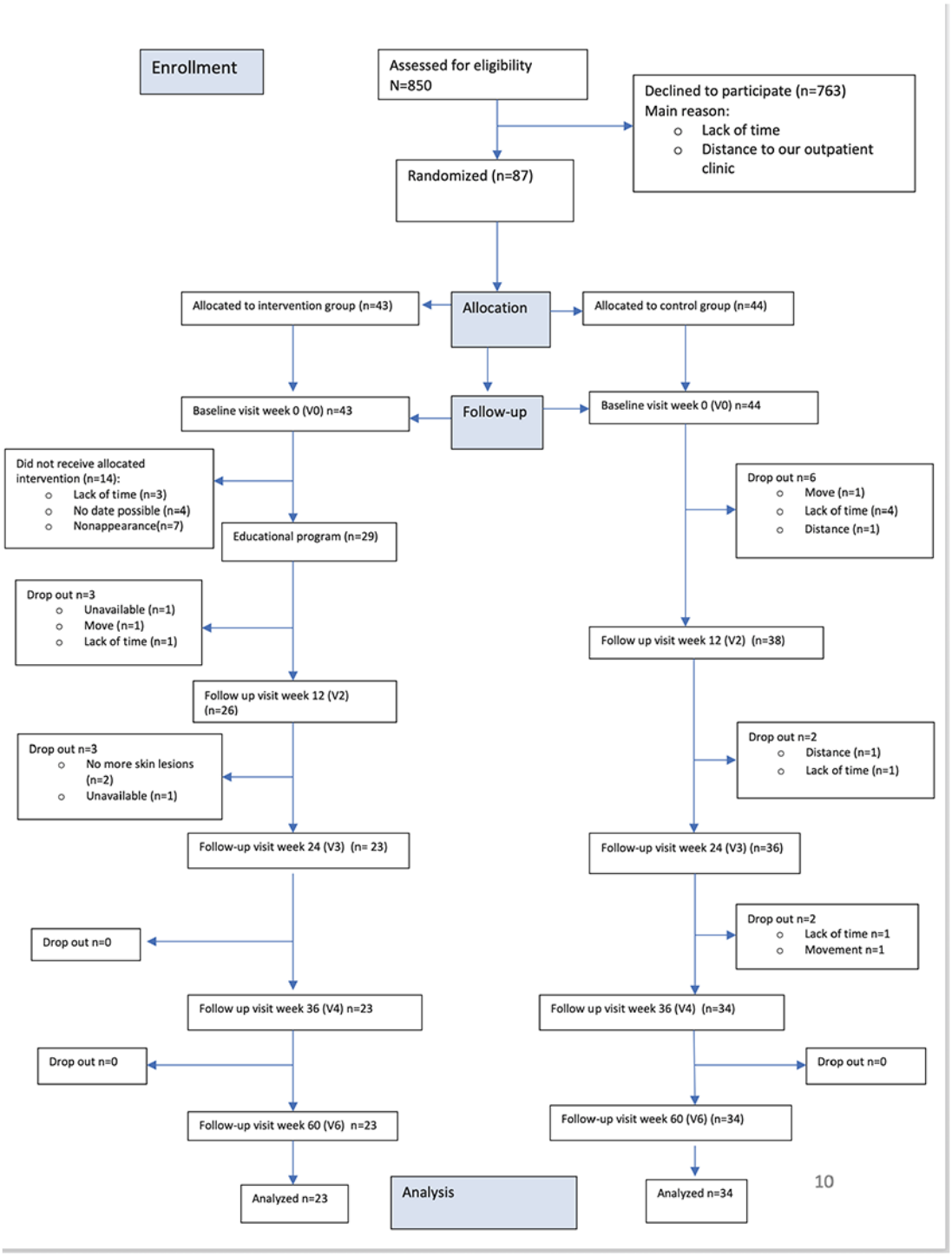

Comparison of Clinical Outcomes Between the Intervention and Control Groups

At the end of the 60-week study period, there was a reduction of pain in both groups compared to their baseline visits (week 60: Coef=–0.400; *P*=.53; Table 2, model 0). There were no significant differences between the control group and the intervention group (week×group: week 60: Coef=–0.175; *P*=.83; Table 2, model 0; Figure 2). The greatest reduction in pain in both groups was noted between week 12 and 24 (week 12: Coef=–0.478; *P*=.44; week 24: Coef=–0.739; *P*=.24; week 36: Coef=–0.608; *P*=.33; week 60: Coef=–0.400; *P*=.53; Table 2, model 0; Figure 2).

XSL•FO
RenderX

Table 2. Random effect regression models over 60 weeks; model 0 unadjusted; model 1 adjusted for age, sex, and disease duration (n= 57; observations=290).

Regression models for pain, activity, HECSI[a], and DLQI[b]	Model 0		Model 1	
	Linear regression coefficient (SE)	*P* value	Linear regression coefficient (SE)	*P* value
Pain				
Week				
0	Reference value	Reference value	Reference value	Reference value
12	–0.478 (0.622)	.44	–0.478 (0.622)	.44
24	–0.739 (0.622)	.24	–0.739 (0.622)	.24
36	–0.608 (0.622)	.33	–0.608 (0.622)	.33
60	–0.400 (0.631)	.53	–0.386 (0.630)	.54
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×control	–0.105 (0.796)	.90	–0.105 (0.796)	.90
24×control	–0.961 (0.796)	.23	–0.961 (0.796)	.23
36×control	–0.590 (0.802)	.46	–0.578 (0.802)	.47
60×control	–0.175 (0.808)	.83	–0.150 (0.808)	.85
Activity				
Week				
0	Reference value	Reference value	Reference value	Reference value
12	–1.391 (0.638)	.03	–1.391 (0.641)	.03
24	–2.347 (0.638)	<.001	–2.347 (0.641)	<.001
36	–2.652 (0.638)	<.001	–2.652 (0.641)	<.001
60	–2.712 (0.647)	<.001	–269 (0.650)	<.001
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×control	1.085 (0.817)	.18	1.085 (0.821)	.19
24×control	0.986 (0.817)	.23	0.986 (0.821)	.23
36×control	0.975 (0.823)	.24	0.957 (0.827)	.25
60×control	1.183 (0.829)	.15	1.149 (0.834)	.17
Mood				
Week				
0	Reference value	Reference value	Reference value	Reference value
12	–0.913 (0.642)	.16	–0.913 (0.643)	.16
24	–1.652 (0.642)	.01	–1.652 (0.643)	.01
36	–1.782 (0.642)	.005	–1.782 (0.643)	.006
60	–1.924 (0.650)	.003	–1.911 (0.651)	.003
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×control	0.051 (0.821)	.95	0.051 (0.823)	.95
24×control	0.541 (0.821)	.51	0.541 (0.823)	.51
36×control	–0.151 (0.827)	.86	–0.159 (0.829)	.85
60×control	–0.009 (0.834)	.99	–0.029 (0.835)	.97

XSL•FO
RenderX

Regression models for pain, activity, HECSI[a], and DLQI[b]	Model 0		Model 1	
	Linear regression coefficient (SE)	*P* value	Linear regression coefficient (SE)	*P* value
HECSI				
Week				
0	Reference value	Reference value	Reference value	Reference value
12	–0.512 (0.237)	.03	–0.512 (0.237)	.03
24	–0.715 (0.237)	.003	–0.715 (0.237)	.003
36	–0.873 (0.237)	<.001	–0.872 (0.237)	<.001
60	–1.108 (0.240)	<.001	–1.090 (0.240)	<.001
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×control	0.327 (0.304)	.28	0.327 (0.304)	.28
24×control	0.428 (0.304)	.16	0.428 (0.304)	.16
36×control	0.509 (0.306)	.10	0.498 (0.306)	.10
60×control	0.597 (0.308)	.05	0.567 (0.308)	.07
DLQI				
Week				
0	Reference value	Reference value	Reference value	Reference value
12	–0.560 (0.210)	.008	–0.560 (0.211)	.008
24	–0.855 (0.210)	<.001	–0.855 (0.211)	<.001
36	–1.105 (0.210)	<.001	–1.105 (0.211)	<.001
60	–1.152 (0.213)	<.001	–1.153 (0.214)	<.001
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×control	0.450 (0.269)	.10	0.450 (0.271)	.10
24×control	0.420 (0.269)	.12	0.420 (0.271)	.12
36×control	0.710 (0.271)	.009	0.707 (0.272)	.01
60×control	0.553 (0.274)	.04	0.551 (0.275)	.05

[a]HECSI: Hand Eczema Severity Index.

[b]DLQI: Dermatology Life Quality Index.

XSL•FO
RenderX

Figure 2. Comparison of pain development between intervention and control groups over the 60-week study period (observant=290; n=57).

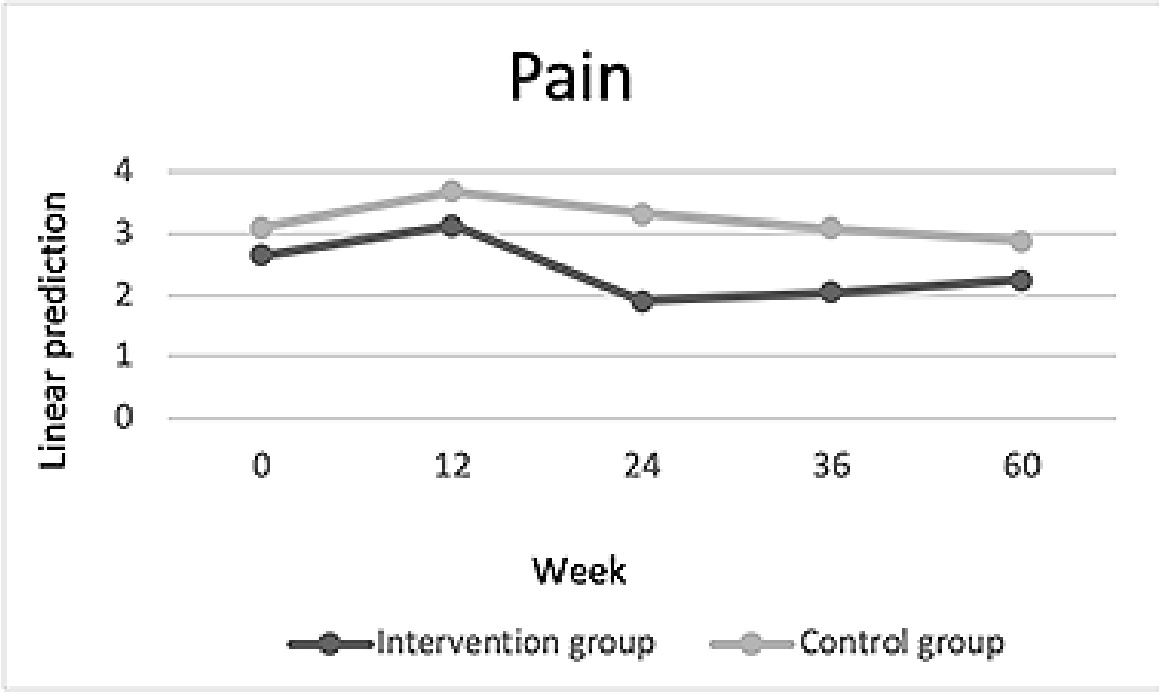

In terms of activity, we were able to demonstrate a significant reduction in impairment regardless of group membership (week 60: Coef=–2.712; $P \leq .001$; Table 2, model 0; Figure 3). No significant differences or changes in the course and development were observed between the control and intervention group (week 60: Coef=1.183; $P=.15$; Table 2, model 0; Figure 3).

Figure 3. Activity development in the intervention group compared with the control group over the 60-week stud period (observant=290; n=57).

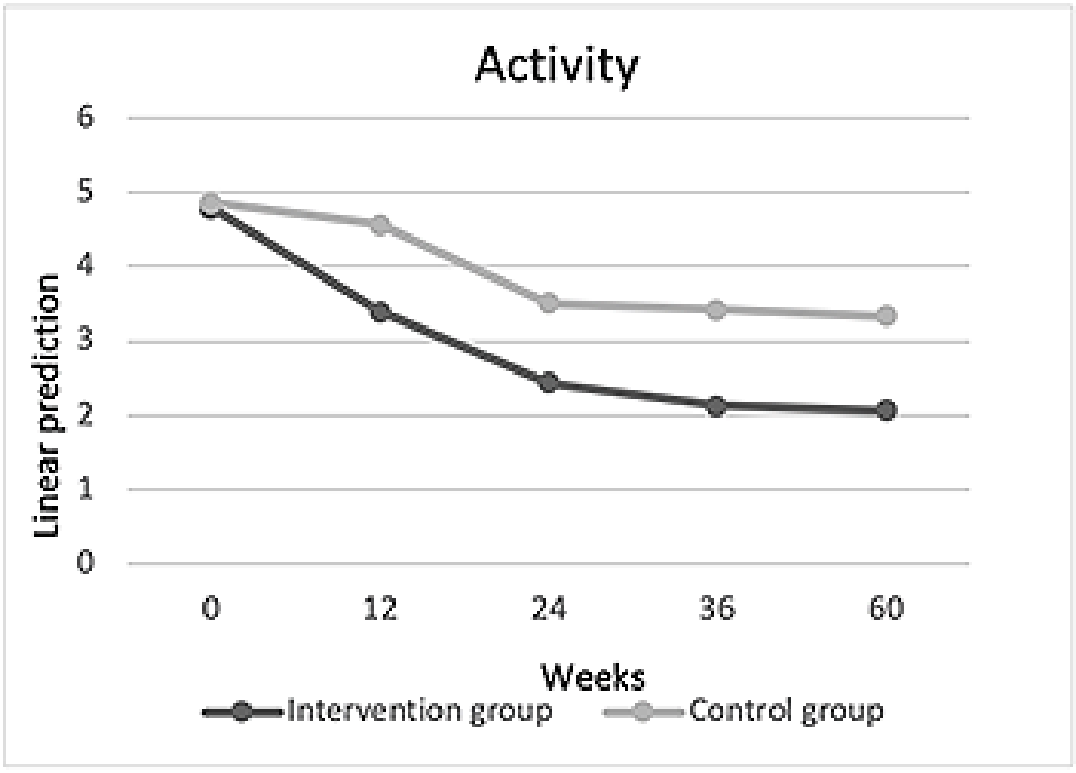

XSL•FO
RenderX

In addition, the mood of patients with hand and foot eczema was ameliorated in both study groups. Significant reductions in scores were observed in nearly all subsequent visits compared to baseline without significant differences between both groups (week 12: Coef=–0.913; *P*=.16; week 24: Coef=–1.652; *P*=.01; week 36: Coef=–1.782; *P*=.005; week 60: Coef=–1.924; *P*=.003; Table 2, model 0; Figure 4).

Figure 4. Mood development in the intervention group compared with the control group over the 60-week study period (observant=219; n=57).

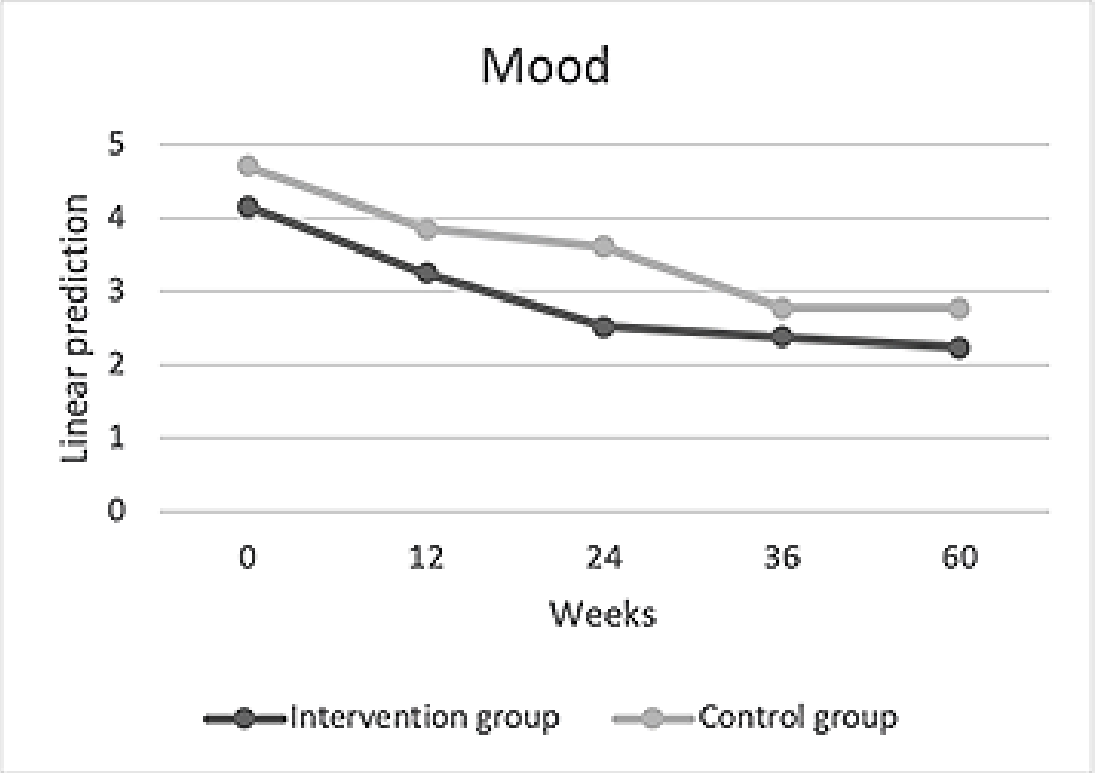

A significant reduction in HECSI scores was consistently observed across all patients over the course of 60 weeks, indicating substantial improvement compared to the initial assessment at week 0 (week 12: Coef=–0.512; *P*=.03; week 24: Coef=–0.715; *P*=.003; week 36: Coef=–0.873; *P*<.001; week 60: Coef=–1.108; *P*≤.001; Table 2, model 0; Figure 5). We observed a trend of greater improvement in the HECSI over time in the intervention group compared to the control group (week×control: week 36: Coef=0.509; *P*=.10; week 60: Coef=0.597; *P*=.05; Table 2, model 0; Figure 5).

XSL•FO
RenderX

Figure 5. HECSI development in the intervention group compared to the control group over the 60-week study period (observant=219; n=57). HECSI: Hand Eczema Severity Index.

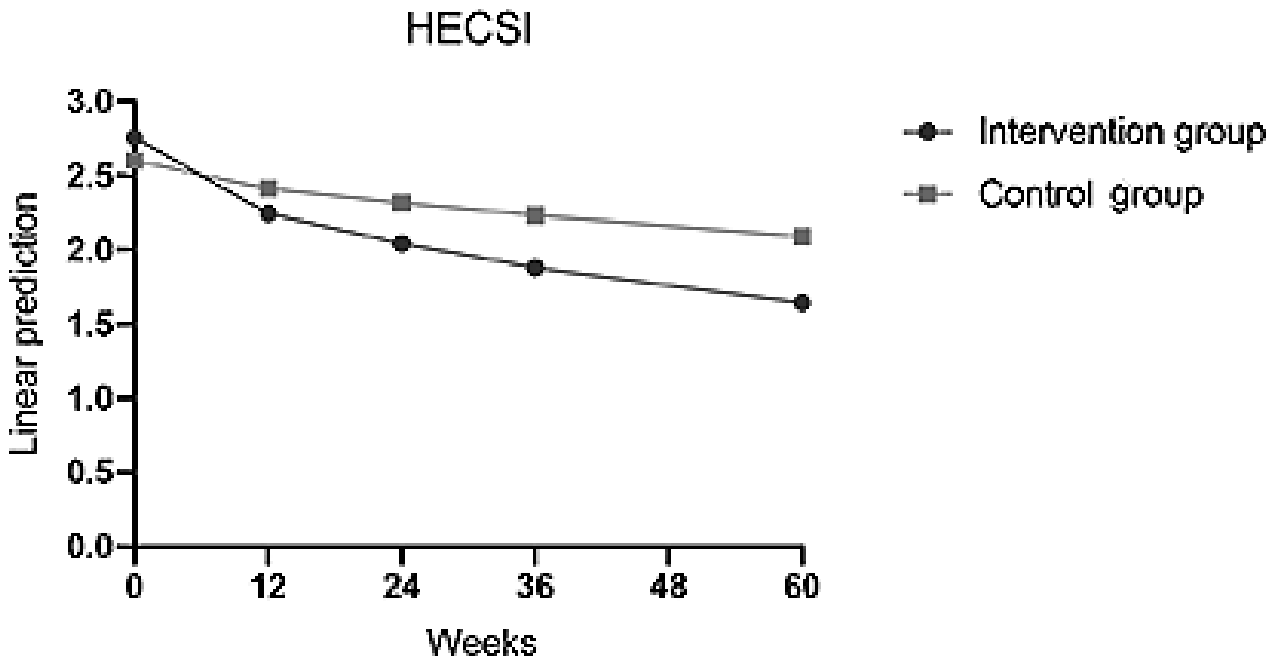

A significant reduction in the DLQI was detected at any time in all patients (week 12: Coef= –0.560; *P*=.008; week 24: Coef=–0.855; *P*≤.001; week 36: Coef=–1.105; *P*≤.001; week 60: Coef=–1.152; *P*≤.001; Table 2, model 0; Figure 6). Starting from week 36, a notable and statistically significant improvement was noted in the intervention group compared to the control group (week×control: week 24: Coef=0.420; *P*=.12; week 36: Coef= –0.710; *P*=.009; week 60: Coef=0.553; *P*=.04; Table 2, model 0; Figure 6).

Figure 6. DLQI development in the intervention group compared to the control group over the 60-week study period (observant=219; n=57). DLQI: Dermatology Life Quality Index.

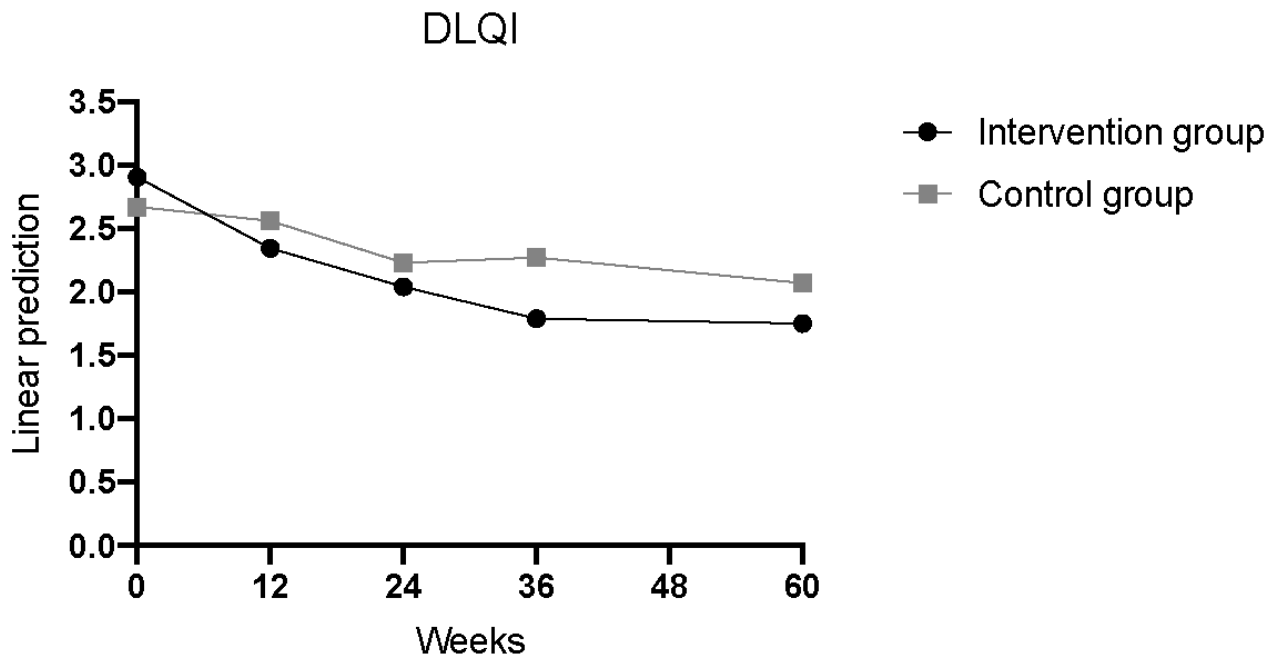

XSL•FO
RenderX

The findings did not differ for sex, age, and disease duration (matching coefficients and *P* values in Table 2, model 1).

App Usage Frequency Subgroup Analysis

For subgroup analysis, the intervention group was divided into 2 groups with patients using the app more or less frequently than once every 5 weeks, which equals an app use frequency of more or less than 20%. The frequency of app usage is measured solely by the actions of uploading images and completing the associated questionnaires. Messages sent to us through the chat function are not included in this count. The ability to ask questions and send messages was unrestricted.

Based on our data, the app usage frequency is influenced by gender (*P*=.006; Table 3) and age (*P*=.03; Table 3). Disease severity (HECSI) and itching had no impact on the frequency of app usage (Table 3). Women did use the app more often than men (women: 42/51, 82% vs men: 29/50, 58%; Table 3), and older women did use the app the most (mean age of women using the app more than 20%=54.787, SD 14.827 years and mean age of men using the app more than 20%=46.833, SD 15.099 years Table 4).

Table 3. Logistic regression of the app usage frequency subgroups ≥20% over 60 weeks (observant=101).

App frequency ≥20%	Odds ratio (95% CI)	SE	*P* value
Gender	4.309 (1.517-12.236)	2.294	.006
Age (V0[a])	1.041 (1.004-1.081)	0.019	.03
HECSI[b]	1.005 (0.609-1.657)	0.256	.99
Itching	0.812 (0.636-1.036)	0.101	.10

[a]The age refers to the age at the time of the first visit (V0).

[b]HECSI: Hand Eczema Severity Index.

Table 4. Two sample t test with equal variances that show which average age (V0) has used the app the most among those people with a frequency more than 20% app usage.

Age	Observants	Mean[a] (SD)	SE	95% CI
Men	30	46.833 (15.099)	2.756	41.194-52.471
Women	71	54.787 (14.827)	1.759	51.279-58.298
Combined	101	52.425 (15.276)	1.520	49.41-55.44

[a]Mean: mean age.

A significant reduction in terms of pain was assessed in patients using the app less than 20%, at weeks 12 and 24 compared to baseline, but this effect did not persist until week 60 when compared to the controls (interaction week×<20%; week 12: Coef=–2.958; *P*=.009; week 24: Coef=–2.972; *P*=.009; week 36: Coef=–2.581; *P*=.06; week 60: Coef=–1.711; *P*=.34; Table 5, model 1; Figure 7). Patients with an app usage frequency of more than 20% showed no significant reduction in pain over the study period (Table 5, model 1; Figure 7). Results were dependent on gender, age, and disease duration (Table 5, model 1).

XSL•FO
RenderX

Table 5. Random effect regression models of the app usage frequency subgroups <20% and ≥20% over 60 weeks. Model 0 is unadjusted, and model 1 is adjusted for age, sex, and disease duration (n=57; observants=290).

Regression models of the app usage frequency subgroups for pain, activity, mood, itching, HECSI[a], and DLQI[b]	Model 0		Model 1	
	Linear regression coefficient (SE)	*P* value	Linear regression coefficient (SE)	*P* value
Pain				
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×<20%	–3.069 (1.190)	.01	–2.958 (1.138)	.009
12×≥20%	–0.278 (0.936)	.98	1.417 (0.895)	.11
24×<20%	–1.764 (1.190)	.14	–2.972 (1.138)	.009
24×≥20%	–0.572 (0.936)	.54	–0.111 (0.894)	.90
36×<20%	–2.576 (1.452)	.08	–2.581 (1.389)	.06
36×≥20%	–0.184 (0.933)	.84	0.468 (0.892)	.60
60×<20%	–3.048 (1.862)	.10	–1.711 (1.782)	.34
60×≥20%	–0.132 (1.025)	.90	1.384 (0.980)	.16
Activity				
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×<20%	–1.416 (1.081)	.19	–3.069 (1.196)	.01
12×≥20%	–0.133 (0.850)	.88	–0.0277 (0.940)	.98
24×<20%	–0.555 (1.081)	.61	–1.764 (1.196)	.14
24×≥20%	–0.355 (0.850)	.68	–0.572 (0.940)	.54
36×<20%	–1.078 (1.323)	.42	–2.541 (1.459)	.08
36×≥20%	0.439 (0.848)	.61	–0.128 (0.936)	.89
60×<20%	–1.395 (1.701)	.41	–2.955 (1.869)	.11
60×≥20%	–0.931 (0.933)	.32	–0.092 (1.029)	.93
Mood				
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×<20%	–0.990 (0.444)	.03	–1.888 (1.197)	.12
12×≥20%	0.026 (0.349)	.94	0.927 (0.941)	.32
24×<20%	–0.651 (0.444)	.14	–1.388 (1.197)	.25
24×≥20%	–0.310 (0.349)	.38	–0.088 (0.941)	.93
36×<20%	–1.289 (0.542)	.02	–1.214 (1.464)	.41
36×≥20%	–0.316 (0.348)	.36	0.823 (0.938)	.38
60×<20%	–0.280 (0.696)	.69	–1.894 (1.881)	.31
60×≥20%	–0.117 (–383)	.76	0.796 (1.032)	.44
Itching				
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×<20%	–1.888 (1.197)	.11	–1.416 (1.081)	–1.31
12×≥20%	0.927 (0.941)	.32	–0.133 (0.850)	.88
24×<20%	–1.388 (1.197)	.25	–0.555 (1.081)	.61

XSL•FO
RenderX

Regression models of the app usage frequency subgroups for pain, activity, mood, itching, HECSI[a], and DLQI[b]	Model 0		Model 1	
	Linear regression coefficient (SE)	*P* value	Linear regression coefficient (SE)	*P* value
24×≥20%	–0.088 (0.941)	.93	–0.355 (0.850)	.68
36×<20%	–1.210 (1.462)	.41	–1.044)1.324)	.43
36×≥20%	0.794 (0.938)	.40	0.460 (0.843)	.59
60×<20%	–1.926 (1.878)	.31	–1.320 (1.703)	.44
60×≥20%	0.777 (1.032)	.45	0.941 (0.933)	.31
HECSI				
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×<20%	–1.234 (0.3916)	.002	–0.990 (0.445)	.03
12×≥20%	–0.032 (0.307)	.92	0.026 (0.350)	.94
24×<20%	–0.732 (0.391)	.06	–0.651 (0.445)	.11
24×≥20%	–0.253 (0.307)	.41	–0.310 (0.350)	.38
36×<20%	–1.500 (0.479)	.002	–1.217 (0.542)	.03
36×≥20%	–0.480 (0.307)	.12	–0.284 (0.348)	.41
60×<20%	–1.275 (0.616)	.04	–0.196 (0.693)	.78
60×≥20%	–0.133 (0.338)	.69	–0.077 (0.382)	.84
DLQI				
Week×group				
0×control	Reference value	Reference value	Reference value	Reference value
12×<20%	–0.081 (0.319)	.79	–1.234 (0.392)	.002
12×≥20%	0.093 (0.251)	.71	–0.032 (0.309)	.92
24×<20%	0.278 (0.319)	.38	–0.732 (0.393)	.06
24×≥20%	0.155 (0.251)	.54	–0.253 (0.309)	.41
36×<20%	0.652 (0.391)	.10	–1.485 (0.481)	.002
36×≥20%	0.261 (0.250)	.30	–0.465 (0.038)	.13
60×<20%	0.995 (0.504)	.05	–1.246 (0.619)	.04
60×≥20%	0.329 (0.276)	.23	–.127 (0.339)	.71

[a]HECSI: Hand Eczema Severity Index.

[b]DLQI: Dermatology Life Quality Index.

XSL•FO
RenderX

Figure 7. Development of the pain comparing the control to the intervention group (observant=290; n=57; interaction week×<20%: week 12: *P*=.009; week 24: *P*=.009; week 36: *P*=.009). NRS: numerical rating scale.

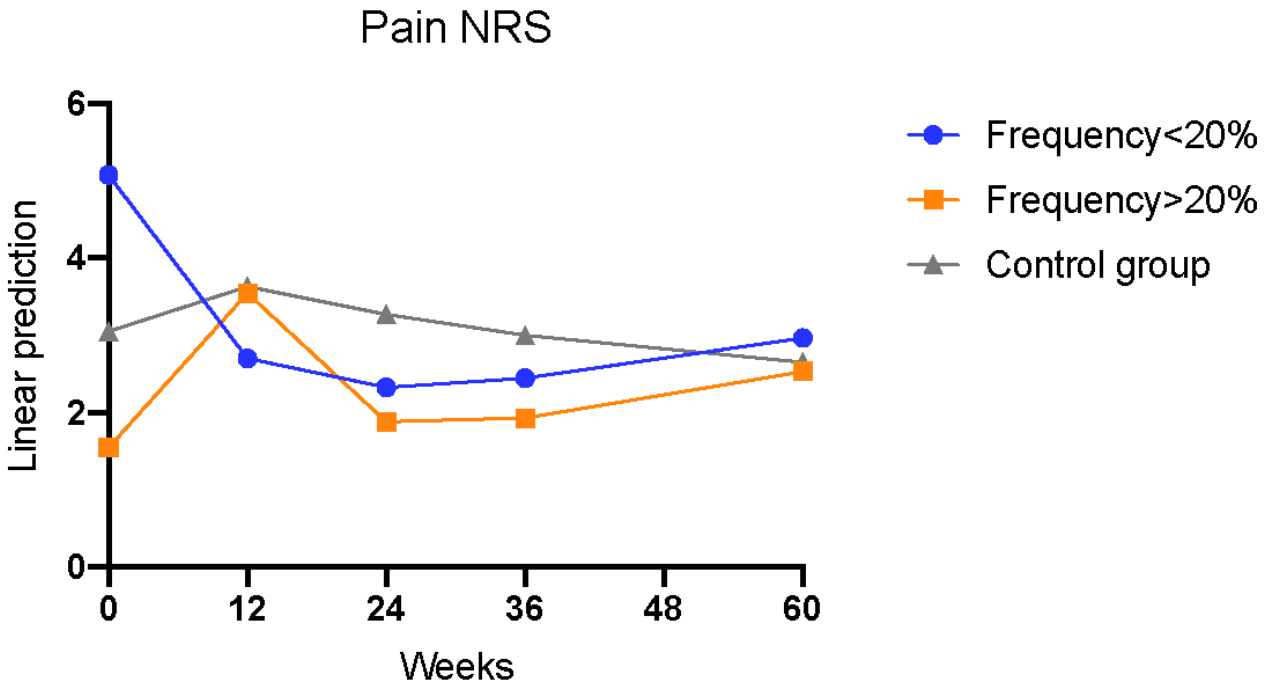

Patients who used the app less frequently than once every 5 weeks exhibited a consistent and statistically significant reduction in HECSI scores from week 0 to week 60, except for week 24, in comparison with the control group (interaction week×<20%: week 12: Coef=–1.234; *P*=.002; week 24: Coef=–0.732; *P*=.06; week 36: Coef=–1.500; *P*=.002; week 60: Coef=–1.275; *P*=.04; Table 5, model 0; Figure 8). Throughout the 60-week study period, no significant reduction in HECSI scores was found among patients who used the app more frequently than once every 5 weeks (Table 3, model 0; Figure 8).

Figure 8. Development of HECSI regarding the control versus the intervention group with <20% frequency usage versus ≥20% usage frequency (observant=290; n=57; interaction week×<20%: week 12: *P*=.002; week 36: *P*=.002; week 60: *P*=.04). HECSI: Hand Eczema Severity Index.

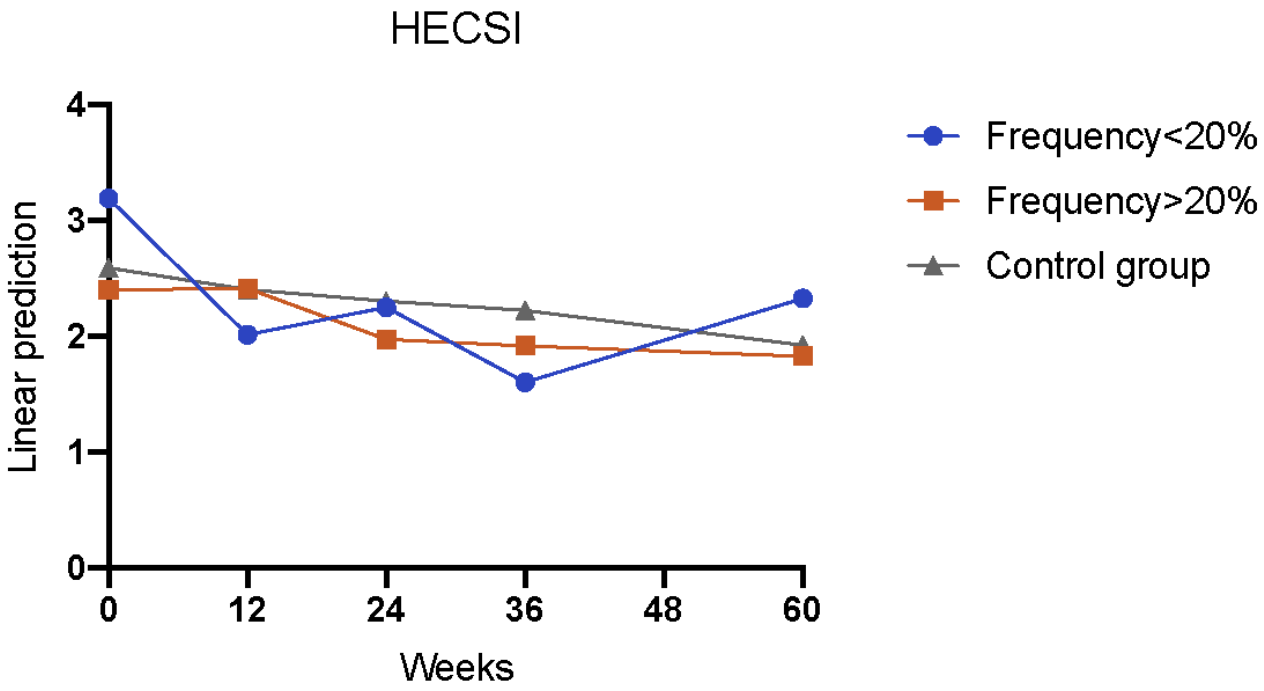

In addition, patients with an app usage frequency of less than 20% demonstrated a consistently significant reduction in DLQI scores throughout the entire study period, except for week 24. From the beginning to week 60, there was a significant improvement in DLQI scores (interaction week×<20%; week 12: Coef=–1.234; *P*=.002; week 24: Coef=–0.732; *P*=.06; week

XSL•FO
RenderX

36: Coef=–1.485; *P*=.002; week 60: Coef=–1.246; *P*=.04; Table 5, model 1; Figure 9). On the other side, no significant reduction was noticed in the group with app frequency of more than 20% (Table 5, model 1; Figure 9). Results were dependent on gender, age, and disease duration (Table 5, model 1).

Figure 9. Development of DLQI regarding the control versus the intervention group with <20% usage frequency versus ≥20% usage frequency (observant=290; n= 57; interaction week×<20%: week 12: *P*=.002; week 36: *P*=.002; week 60: *P*=.04). DLQI: Dermatology Life Quality Index.

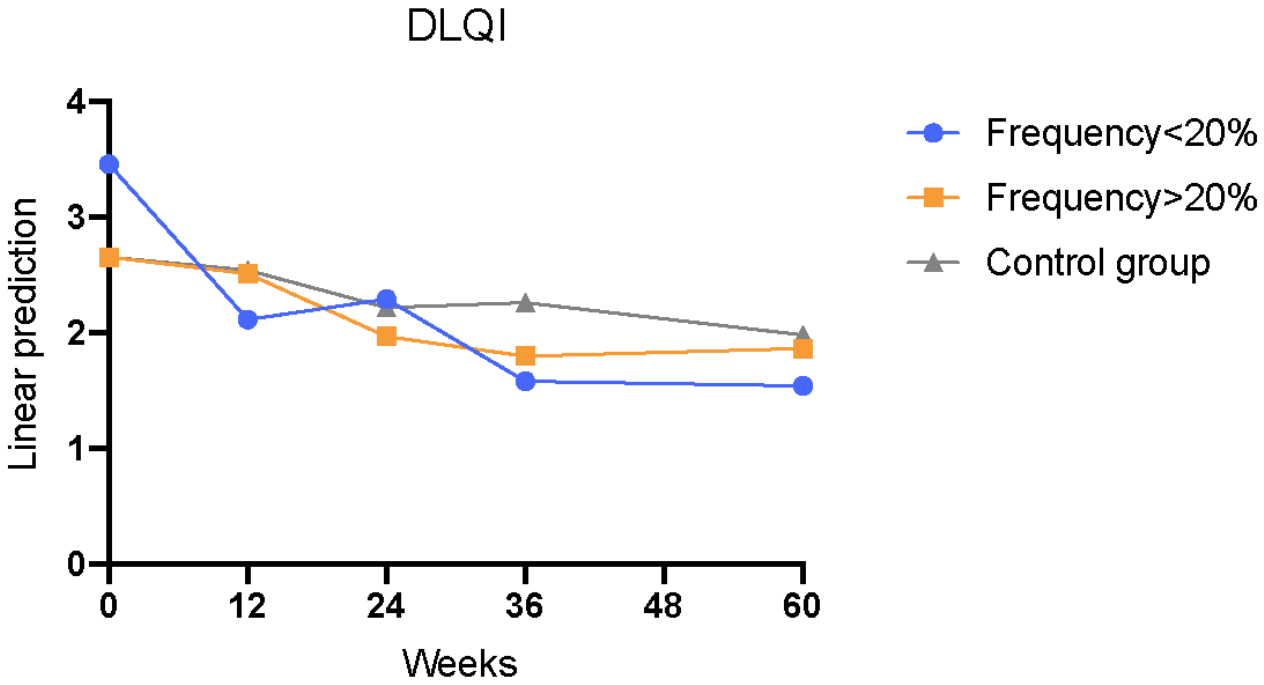

A significant improvement in mood was also observed in patients with less than 20% app usage frequency at weeks 12 and 36 (interaction week×<20%; week 12: Coef=–0.990; *P*=.03; week 36: Coef=–1.289; *P*=.02; Table 5, model 0), consistent with the DLQI results.

No significant effects on activity and itching were noticed in patients with an app usage frequency of less or more than 20% (Table 5).

Comparison of Data Collected From the DLQI and HECSI Through the App Compared to Those Collected During Personal Visits

The intervention group was asked to upload a picture of the affected area on hands and feet once a week. Doctors then calculated the eHECSI based on the uploaded pictures. The eHECSI was compared to the HECSI documented in personal visits. The same applied to the DLQI and electronic Dermatology Life Quality Index (eDLQI).

The eHECSI was comparable with the HECSI obtained during in-person visits (ρ=0.885; *P*≤.001; Multimedia Appendix 2). The same was true for the DLQI as the DLQI and the eDLQI showed a strong correlation (ρ=0.474; *P*≤.001; Multimedia Appendix 2). The assessment involved evaluating image quality in terms of lighting, sharpness or focus, and completeness. Despite the presence of suboptimal image quality, the calculated eHECSI still demonstrated a strong correlation with the in-person assessed HECSI (for good quality: ρ=0.884; *P*<.001; for bad quality: ρ=0.901; *P*≤.001; Multimedia Appendix 3).

Discussion

Principal Findings

In this study, we investigated the effect of an educational program combined with a smartphone app on the clinical outcome and well-being of patients with chronic hand and foot eczema. Results of the interim analysis at week 24 have already been published and demonstrate the efficacy of our intervention in terms of improvements in patients' quality of life, pain levels, activity, and clinical outcomes [10]. The 60-week data presented here showed that the positive effects, particularly regarding quality of life and, to a lesser extent, clinical outcome, were maintained over 60 weeks. Therefore, the educational program combined with the app, which facilitates communication with the doctor through the chat function, seems to have positive long-lasting effects. The intervention may help patients cope better with the disease, understand it better, and ultimately adhere to treatment regimens. That educational programs lead to better outcomes in patients with hand eczema, is not new. Corti et al [12] retrospectively analyzed the records of 36 patients who participated in an educational program and showed that 67% (24/36) of patients with hand eczema benefited from the intervention, mainly due to behavioral changes regarding skin care and protection [12]. However, the follow-up visit was after 12 weeks, so long-term effects were not analyzed.

In our study, improvement in pain, activity, HECSI, and DLQI were observed in the whole study population. The improvement observed in both groups may be study related to the more intensive care of the patients by their dermatologists. Stewart

XSL•FO
RenderX

et al [13] showed that close doctor-patient contact and good communication between both parties have positive healing effects [13], which underlines the importance of the control group when evaluating the effect of an intervention such as an educational program or an eHealth device.

Regarding the frequency of use of the app, we were able to show that less frequent use of the app resulted in better outcomes, mood, and quality of life. The reasons for this unexpected finding may be multifaceted. First, patients may not want to be confronted with their disease on a regular basis. Similar results were observed in our previous study, which was a 60-week monitoring app intervention study focused on assessing patients with psoriasis. In the group that used the app less than once every 5 weeks, patients showed significant improvement in depression and anxiety symptoms as measured by the Hospital Anxiety and Depression Scale. The conclusion of this study highlighted the preference of patients with chronic diseases to avoid frequent reminders of their conditions [11]. We recently held a workshop with patients with psoriasis and asked them what features an app should have to make it worth using (unpublished data). Interestingly, one of the most important pieces of feedback was that the app should help save time in health care. Patients do not want to spend time filling out long questionnaires or documenting their disease but want an app that can be used according to their health needs. This should be considered when developing future patient-centered apps.

Interestingly, gender and age had an impact on the frequency of app use; especially older women were most likely to use the app. There could be several reasons for this observation. The fact that older people use the app more often may be explained by the fact that they have more free time than younger people who are still pursuing their careers. The paper by Rosales and Fernández-Ardèvol [14] examines the smartphone usage and app preferences of older individuals residing in Spain. Their findings indicate that older adults tend to use smartphones and smartphone apps in a more utilitarian manner. Specifically, they frequently access personal information manager apps, such as calendars, address books, and notes, more often than younger age groups. This highlights how older users prioritize and use apps that serve practical purposes in their daily lives [14]. Older women may have a higher motivation for diligent self-monitoring of their condition, leading them to use the app more regularly. It would also be interesting to investigate whether the design or usability of the app specifically met the needs of older women, which might explain their higher usage. Further research and analysis could contribute to a better understanding of this result and provide valuable insights for future studies and developments. In terms of gender differences, the study by Graziano et al [15] found notable gender differences in the relationship between parent-reported self-regulation skills and adolescents' management of type 1 diabetes. Specifically, among boys, it was observed that deficits in executive functioning and emotion regulation were significantly linked to poorer treatment adherence and glycemic control [15]. In another study examining the development of self-regulation in the first 4 years of life, it was found that girls exhibited higher levels of committed compliance compared to boys [16]. In general, there seem to be significant gender differences in the use of eHealth devices, which is also an important fact to consider in the development of health-related apps and needs further investigation.

In comparing the eHECSI obtained through photo documentation through the app with the in-person assessments during regular visits, we showed that the image quality uploaded in the app did not impact the objective assessment. Consistent results were found for both the eHECSI and the HECSI assessed in person. This suggests that even in cases where the image quality was not ideal, possibly due to the patient's camera or photography technique, the HECSI could still be effectively assessed. This is certainly not true for all dermatologic conditions, especially those that can affect the entire body, and highlights that hand and foot eczema is a perfect candidate for teledermatologic monitoring. The implementation of a monitoring app, possibly combined with artificial intelligence, could therefore have a huge socioeconomic impact, saving time not only for the patient but also for the health care system.

It is important to note that the dropout rate in the intervention group was relatively high (46.5%) compared to the control group (22.7%). The most important factor contributing to the high dropout rate in the intervention group was the educational program required of patients in the intervention group. At the beginning of the study, it was difficult to schedule a time for the intervention group to receive the training and introduction to the app. Some participants initially agreed to attend, but later either did not show up or canceled at short notice, especially since many participants lived several miles from the clinic, making attendance a significant challenge. Based on our experience, for future studies, online training sessions could potentially engage more participants who are unable to travel long distances or have only a few minutes to attend the training from home.

Limitations

The main limitations of our study are that the cohort became smaller over time and there is an imbalance in the number of patients in the groups. In particular, the number of patients in the intervention group became small over the 60 weeks of care, which may lead to misinterpretation of the results and differences between the 2 groups and the subgroup analysis. In addition, the patients in our study tended to be older, which means that it may be difficult to extrapolate the results to all age groups. In the scope of our app-focused study, a noteworthy limitation lies in our inability to disentangle the distinct effects of the educational program and the app within the intervention group. The intertwined nature of these 2 interventions makes it challenging to attribute the observed outcomes solely to 1 component. Further research could delve deeper into this aspect, potentially conducting separate investigations to better understand the individual contributions, interactions, and impact of the educational program and the app on our intervention group's outcomes. Further studies are needed to verify our results and findings in a larger patient population.

Conclusion

In summary, the combination of the educational program and the app had a clear positive influence on patients' mental

XSL•FO
RenderX

well-being and disease severity, particularly when the app usage frequency was moderate. It is worth noting that the app, when used in conjunction with in-person visits, contributed to the observed positive effects. Therefore, the app should be viewed as a supplementary tool and not a complete replacement for real-time medical consultations. As shown in the study by Baldwin et al [17], telemedicine can lead to a 75% reduction in outpatient clinic visits saving time for both patients and health care institutions. Teledermatology therefore offers the potential for a cost-effective supplementary health care approach, which should be integrated into the daily care of dermatological patients, especially in easy-to-monitor dermatological conditions such as hand and foot eczema [18].

Acknowledgments

We express our special thanks to the physicians and nursing staff of the Dermatology Department of the University Hospital Mannheim and to Novartis GmbH for their financial support of the study. All named authors and Novartis GmbH had full access to the study data.

Open Access funding enabled and organized by Projekt DEAL. Novartis GmbH.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

AB, RH, WW, and YS declare no conflicts of interest. AS obtained honoraria for presentations and as a member of advisory boards of LEO Pharma, Eli Lilly GmbH, Novartis GmbH, and Almirall Hermal GmbH.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 162 KB - jmir_v27i1e52159_app1.pdf]

Multimedia Appendix 2
Spearman correlation comparing the electronic Hand Eczema Severity Index and electronic Dermatology Life Quality Index from the app with the Hand Eczema Severity Index and Dermatology Life Quality Index from in-person visits (observant=61).
[DOCX File , 15 KB - jmir_v27i1e52159_app2.docx]

Multimedia Appendix 3
Spearman correlation comparing the results of electronic Hand Eczema Severity Index and Hand Eczema Severity Index based on the quality of the picture downloaded through the app (observant=51).
[DOCX File , 14 KB - jmir_v27i1e52159_app3.docx]

References

1. Quaade AS, Simonsen AB, Halling AS, Thyssen JP, Johansen JD. Prevalence, incidence, and severity of hand eczema in the general population - a systematic review and meta-analysis. Contact Dermatitis 2021;84(6):361-374. [doi: 10.1111/cod.13804] [Medline: 33548072]
2. Braun-Falco O, Plewig G, Wolff HH, Burgdorf HC, Landthaler M. Dermatologie und Venerologie. Berlin, Heidelberg: Springer; 2005.
3. Silverberg JI, Guttman-Yassky E, Agner T, Bissonnette R, Cohen DE, Simpson E, et al. Chronic hand eczema guidelines from an expert panel of the international eczema council. Dermatitis 2021;32(5):319-326 [FREE Full text] [doi: 10.1097/DER.0000000000000659] [Medline: 32991335]
4. Gorris A, Kinaciyan T. Das handekzem. hautnah 2020;19(4):162-170. [doi: 10.1007/s12326-020-00407-5]
5. Quaade AS, Alinaghi F, Dietz JBN, Erichsen CY, Johansen JD. Chronic hand eczema: a prevalent disease in the general population associated with reduced quality of life and poor overall health measures. Contact Dermatitis 2023;89(6):453-463. [doi: 10.1111/cod.14407] [Medline: 37634937]
6. Armstrong A, Hahn-Pedersen J, Bartlett C, Glanville J, Thyssen JP. Economic burden of chronic hand eczema: a review. Am J Clin Dermatol 2022;23(3):287-300 [FREE Full text] [doi: 10.1007/s40257-021-00669-6] [Medline: 35258783]
7. Beltraminelli H, Itin P. Skin and psyche--from the surface to the depth of the inner world. J Dtsch Dermatol Ges 2008;6(1):8-14. [doi: 10.1111/j.1610-0387.2007.06406.x] [Medline: 17883417]
8. Agner T, Elsner P. Hand eczema: epidemiology, prognosis and prevention. J Eur Acad Dermatol Venereol 2020;34 Suppl 1:4-12. [doi: 10.1111/jdv.16061] [Medline: 31860734]

XSL•FO
RenderX

9. DiGA-Verzeichnis. (BfArM) BfAuM. URL: https://diga.bfarm.de/de/verzeichnis2024 [accessed 2024-12-04]
10. Weigandt WA, Schardt Y, Bruch A, Herr R, Goebeler M, Benecke J, et al. Impact of an eHealth smartphone app on quality of life and clinical outcome of patients with hand and foot eczema: prospective randomized controlled intervention study. JMIR Mhealth Uhealth 2023;11:e38506 [FREE Full text] [doi: 10.2196/38506] [Medline: 36881465]
11. Domogalla L, Beck A, Schulze-Hagen T, Herr R, Benecke J, Schmieder A. Impact of an eHealth smartphone app on the mental health of patients with psoriasis: prospective randomized controlled intervention study. JMIR Mhealth Uhealth 2021;9(10):e28149 [FREE Full text] [doi: 10.2196/28149] [Medline: 34431478]
12. Corti MAM, Stirnimann R, Borradori L, Simon D. Effects of systematic patient education in skin care and protection in a hand eczema clinic. Dermatology 2014;228(3):220-224 [FREE Full text] [doi: 10.1159/000357523] [Medline: 24776632]
13. Stewart M, Brown JB, Boon H, Galajda J, Meredith L, Sangster M. Evidence on patient-doctor communication. Cancer Prev Control 1999;3(1):25-30. [Medline: 10474749]
14. Rosales A, Fernández-Ardèvol M. Beyond WhatsApp: older people and smartphones. Ro. Comm. & PR 2016;18(1):27. [doi: 10.21018/rjcpr.2016.1.200]
15. Graziano PA, Geffken GR, Williams LB, Lewin AB, Duke DC, Storch EA, et al. Gender differences in the relationship between parental report of self-regulation skills and adolescents' management of type 1 diabetes. Pediatr Diabetes 2011;12(4 Pt 2):410-418. [doi: 10.1111/j.1399-5448.2010.00692.x] [Medline: 21392190]
16. Kochanska G, Coy KC, Murray KT. The development of self-regulation in the first four years of life. Child Dev 2001;72(4):1091-1111. [doi: 10.1111/1467-8624.00336] [Medline: 11480936]
17. Baldwin L, Clarke M, Hands L, Knott M, Jones R. The effect of telemedicine on consultation time. J Telemed Telecare 2003;9 Suppl 1:S71-S73. [doi: 10.1258/135763303322196420] [Medline: 12952731]
18. Whited JD. Economic analysis of telemedicine and the teledermatology paradigm. Telemed J E Health 2010;16(2):223-228. [doi: 10.1089/tmj.2009.0100] [Medline: 20136558]

Abbreviations

Coef: Linear regression coefficient
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DiGA: Digital Health Applications
DLQI: Dermatology Life Quality Index
eDLQI: electronic Dermatology Life Quality Index
eHECSI: electronic Hand Eczema Severity Index
HECSI: Hand Eczema Severity Index

Edited by A Mavragani; submitted 24.08.23; peer-reviewed by KL Mauco, B Breznen; comments to author 16.06.24; revised version received 19.08.24; accepted 19.09.24; published 21.01.25.

Please cite as:
Bruch A, Weigandt W, Schardt Y, Herr R, Benecke J, Schmieder A
Improving Outcomes and Quality of Life for Patients With Hand and Foot Eczema: Randomized Study of a Patient-Centered Monitoring App
J Med Internet Res 2025;27:e52159
URL: https://www.jmir.org/2025/1/e52159
doi:10.2196/52159
PMID:

©Aimee Bruch, Wanja Weigandt, Yannic Schardt, Raphael Herr, Johannes Benecke, Astrid Schmieder. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Web-Based Medication Education Course on Pregnant Women's Medication Information Literacy and Decision Self-Efficacy: Randomized Controlled Trial

Suya Li[1*], MSN; Hui-Jun Chen[2*], MD; Jie Zhou[1], MSN; Yi-Bei Zhouchen[1], MNS; Rong Wang[3], MSN; Jinyi Guo[1], MNS; Sharon R Redding[4], EdD; Yan-Qiong Ouyang[1], MD

[1]Wuhan University, Wuhan, China

[2]Department of Obstetrics, Zhongnan Hospital of Wuhan University, Wuhan, China

[3]Renmin Hospital of Wuhan University, Wuhan, China

[4]Project HOPE, Washington, DC, United States

[*]these authors contributed equally

Corresponding Author:
Yan-Qiong Ouyang, MD
Wuhan University
115 Donghu Road
Wuchang District
Wuhan, 430071
China
Phone: 86 02768758747
Email: ouyangyq@whu.edu.cn

Abstract

Background: Medication-related adverse events are common in pregnant women, and most are due to misunderstanding medication information. The identification of appropriate medication information sources requires adequate medical information literacy (MIL). It is important for pregnant women to comprehensively evaluate the risk of medication treatment, self-monitor their medication response, and actively participate in decision-making to reduce medication-related adverse events.

Objective: This study aims to examine the effectiveness of a medication education course on a web-based platform in improving pregnant women's MIL and decision self-efficacy.

Methods: A randomized controlled trial was conducted. Pregnant women were recruited from January to June 2021 in the Department of Obstetrics and Gynecology of a large hospital in a major city in central China. A total of 108 participants were randomly divided into a control group (CG), which received routine prenatal care from nurses and physicians, and an intervention group (IG), which received an additional 3-week web-based medication education course based on the theory of planned behavior as part of routine prenatal care. Participants completed a Medication Information Literacy Scale and a decision self-efficacy questionnaire at baseline, upon completion of the intervention, and at a 4-week follow-up. Generalized estimation equations (GEE) were used to analyze the main effect (time and grouping) and interaction effect (grouping×time) of the 2 outcomes. The CONSORT-EHEALTH (V 1.6.1) checklist was used to guide the reporting of this randomized controlled trial.

Results: A total of 91 pregnant women (48 in the IG and 43 in the CG) completed the questionnaires at the 3 time points. The results of GEE indicated that there was no statistically significant difference in time×group interactions of MIL between the 2 groups (F_2=3.12; P=.21). The results of the main effect analysis showed that there were statistically significant differences in MIL between the 2 groups at T1 and T2 (F_1=17.79; P<.001). Moreover, the results of GEE indicated that there was a significant difference in decision self-efficacy regarding the time factor, grouping factor, and time×group interactions (F_2=21.98; P<.001). The results of the simple effect analysis indicated a statistically significant difference in decision self-efficacy between the 2 groups at T1 (F_1=36.29; P<.001) and T2 (F_1=36.27; P<.001) compared to T0. Results showed that MIL and decision self-efficacy in the IG were found to be significantly higher than those in the CG (d=0.81; P<.001 and d=1.26; P<.001, respectively), and they remained significantly improved at the 4-week follow-up (d=0.59; P<.001 and d=1.27; P<.001, respectively).

Conclusions: Web-based medication education courses based on the theory of planned behavior can effectively improve pregnant women's MIL and decision self-efficacy, and they can be used as supplementary education during routine prenatal care.

XSL•FO
RenderX

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100041817; https://www.chictr.org.cn/showproj.html?proj=66685

(J Med Internet Res 2025;27:e54148) doi:10.2196/54148

KEYWORDS

decision self-efficacy; self-efficacy; decision efficacy; medication information literacy; information literacy; web-based medication education; medication education; web-based platforms; pregnant women; pregnancy; RCT; randomized controlled trial

Introduction

Pregnancy is a unique and sensitive period, accompanied by a series of temporary physiological and metabolic changes, which can directly affect the absorption, distribution, metabolism, and excretion of drugs [1]. With the prevalence of medication use during pregnancy and the occurrence of more and more safety accidents during this period, the safety of medications used during pregnancy has been paid increased attention [2,3]. Previous studies have shown that medication exposure during pregnancy is related to adverse pregnancy outcomes (eg, cumulative toxicities in pregnant women, birth defects, or preterm birth) and has long-term effects on fetal development [4].

Overestimating the teratogenic risk of medication [5,6], misinterpreting medication-related information [7], unconfirmed information and exaggeration of adverse medication reactions on the internet [8,9], and a large amount of terminology and vagueness of precautions in instructions [10] can lead to women's confusion about medication use. This can increase the risk of optional medication withdrawal and exacerbate health conditions. Therefore, pregnant women not only need the ability to obtain appropriate medication information but also need to correctly interpret this information. They should be able to effectively distinguish the authenticity of medication information. This is influenced by medication information literacy (MIL) to some extent.

MIL refers to the information behavior related to medications, including the ability to seek, comprehend, identify, and use medication-related information [11]. This term combines medication literacy and information literacy, which is important for pregnant women to determine the risks of medication treatment and self-monitor medication reactions [12]. In addition, MIL represents the conceptualization of functional health literacy, emphasizing comprehensive, critical, proactive, and interactive capabilities and includes not only the demands identified, acquisition, understanding, evaluation, integration, and application of information but also medication information exchange, dosage calculation, and use of drug data to make smart health decisions [13]. Decision self-efficacy refers to self-confidence or a belief in one's ability to make decisions, including shared decision-making [14]. It is an important factor affecting individual coping ability, individual choice, and the motivation and effort of trying to achieve a set of goals [15]. Pregnant women with low levels of MIL and decision self-efficacy have difficulties in correctly interpreting medication information and accurately implementing medication instructions, which may be a predictor of adverse medication events [16].

A previous survey showed that the MIL of pregnant women in China was generally low, but the demand for medication information was strong [11]. Therefore, it is urgent to carry out targeted interventions to improve the MIL of pregnant women.

The theory of planned behavior (TPB) is used to help predict and explain health behaviors and guide intervention programs [17] and is one of the most famous theories used in interpreting attitude-behavior relations [17,18]. According to TPB, behaviors are jointly determined by attitude toward the behavior, subjective norms, and perceived behavioral control. Previous studies have revealed the influencing factors of medication compliance and health literacy of pregnant women through the TPB [19-21]. The application of TPB in this study is shown in Figure 1.

During the COVID-19 pandemic, web-based education played an important role in meeting the needs of pregnant women [22]. To a certain extent, web-based education also provided protective isolation for pregnant women [23]. At present, WeChat (Tencent) has become a major social media web-based platform in China and has also received widespread attention from pregnant women [24]. Because of its convenience, timeliness, and intelligence, WeChat has made great efforts in the field of prenatal education related to blood glucose management [25], weight management [26], nutritional intervention [27], and breastfeeding promotion education [24]. Among pregnant women, the feasibility and effectiveness of health education based on the WeChat platform have been highlighted.

This study was undertaken to offer a health education course on the WeChat platform to empower pregnant women with the ability to inquire about, use, and evaluate medication information to actively participate in medical decisions and consciously resist adverse medication behaviors. Therefore, it is hypothesized that the web-based medication educational course would be effective in improving pregnant women's MIL and decision self-efficacy.

XSL•FO
RenderX

Figure 1. Application of the theory of planned behavior in this study.

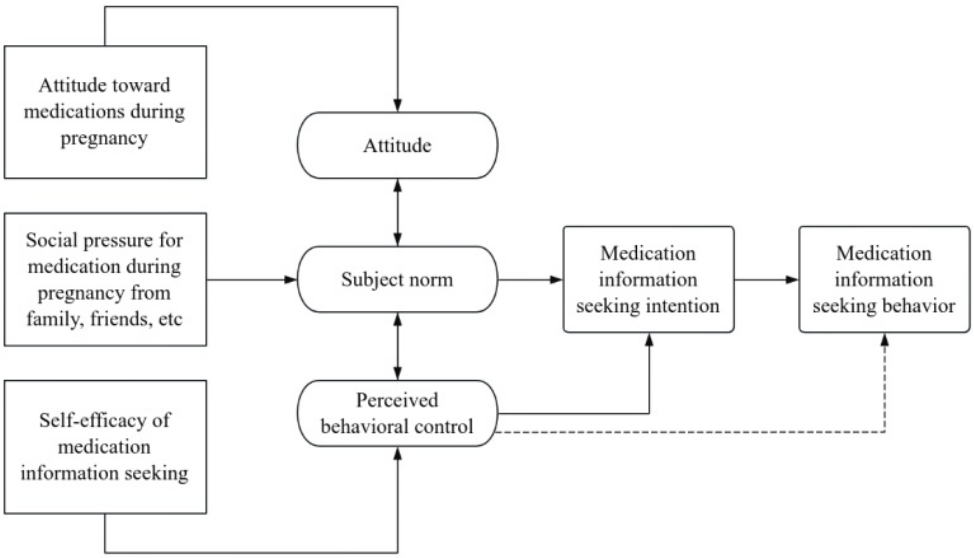

Methods

Study Design

This study is a parallel randomized controlled trial delivering an additional 3-week web-based course about MIL to the intervention group (IG) with a 4-week follow-up. Pregnant women were recruited by convenience sampling from January to June 2021 in the Department of Obstetrics and Gynecology of a large hospital in a major city in central China. The study was registered in the Chinese Clinical Trial Registry (ChiCTR2100041817).

Feasibility Trial

In order to explore the feasibility of a 3-week web-based course based on TPB, the researcher conveniently selected 10 pregnant women who met eligibility criteria in the department mentioned above from November to December 2020, for a feasibility trial. Based on the feedback of the pregnant women, the intervention program was further improved to the formal version. Participant's completion of the interventions, attendance, and absences were recorded. Retention, compliance, and adherence rates were calculated.

The feasibility trial was well attended, all participants were engaged, interested in the course, and endorsed the intervention adaptations as culturally appropriate. All participants at session one indicated a desire to attend the remaining sessions, and 8 participants completed all the sessions, for an 80% adherence rate. However, only 7 of the women who attended all the sessions completed the questionnaire after the intervention, for a 70% retention rate. A number of barriers to attendance were identified as family obligations, the need to travel, and work obligations.

Participants

Inclusion and Exclusion Criteria

Inclusion criteria were: primipara, at 24 weeks or less of gestation, reads and speaks Mandarin, is able to complete a questionnaire independently, agrees to take part in an education course, and has access to the WeChat platform. Exclusion criteria were: current participation in other research on similar topics and health care workers. In selecting the inclusion criterion of 24 weeks of gestation or less for primiparas, we considered multiple factors to ensure the validity and practicality of our study. One significant reason is that after 24 weeks, there is an increased likelihood of premature delivery, which could lead to participants dropping out of prenatal classes prematurely. This would not only disrupt the continuity of the intervention but also introduce bias by excluding those who might have experienced premature birth. By focusing on primiparas at or before 24 weeks, we aimed to minimize such dropouts and maintain a more homogeneous study population that could fully benefit from and complete the intervention program. This approach allowed us to accurately assess the intervention's impact without any confounding effects of early termination due to premature.

Sample Size

The sample size was calculated according to the formula



using a power of 0.8 and an α of 0.05, Z_{α}=1.96 and Z_{β}=0.84, as used in similar research [11], σ=8.22 and δ=6.00. The study required a minimum sample size of 29 (each group). Considering a 15% attrition rate, a minimum of 68 participants were needed for the study.

Procedures

Recruitment

Women meeting eligibility criteria were invited to participate. The researcher explained the purpose of the study and informed that participation in the study was voluntary and that no prejudice would result from nonparticipation, termination, or withdrawal from the study. After all questions about the study were answered, participants were invited to sign an informed consent form.

XSL•FO
RenderX

Randomization and Blinding

Eligible participants were randomly assigned to the IG and the control group (CG) using a third-party organization [28]. The results of random allocation were placed in an opaque envelope. The researchers did not interfere with the random assignment, but, after allocation, the nature of the intervention made it difficult to blind participants and researchers.

Interventions

Participants in the CG received routine prenatal outpatient care from a physician. At the same time, they were contacted by the researchers upon initial enrollment in the study but received no further education courses.

Participants in the IG were offered an additional 3-week web-based course to improve MIL (Table S1 in Multimedia Appendix 1). The course was designed based on TPB [19] and used content from Drugs for Pregnant and Lactating Women [29] and the Risk Classification Systems of the US Food and Drug Administration [30]. The course focused on medication sensitivity during pregnancy, the popularity of essential medication, individual and social attitudes, and understanding, seeking, and evaluating medication information. The course arrangement was modified following the recommendations of pharmacists, obstetricians, and registered nurses. Feedback from study participants led to adjusting the course schedule to better meet the medication information needs of the women.

The web-based health education course consisted of 2 parts: an education module and a consultation module. The education module was delivered using the WeChat official account (Maternal and Child Guardian). Teaching strategies used in the course included both visual and auditory. Taking into account the concentration ability of pregnant women, each class was 15 minutes in length [31]. For the consultation module, the researchers established a WeChat group (individual account). Recent informational studies were shared with group participants every evening. A group chat was conducted weekly to review information and obtain feedback. In addition, participants were encouraged to contact researchers privately to ask questions, which would be answered within 24 hours. At follow-up, participants could review classes repeatedly. After 4 weeks of follow-up, an electronic questionnaire was sent to participants via the WeChat platform.

Instrument

Sociodemographic Characteristics and Medication Use Questionnaire

Sociodemographic information included age, ethnicity, education level, location of residence, profession, household monthly income (in RMB), and gestational weeks. The medication use questionnaire included: (1) How many acute illnesses (such as fever and cold) did you have during pregnancy? (2) Have you ever used medications due to acute illnesses (such as fever and cold) during pregnancy? (3) How many chronic diseases did you experience during pregnancy (such as gestational hypertension, gestational diabetes, and coronary heart disease)? (4) Have you taken any medication for your chronic disease (such as gestational hypertension, gestational diabetes, and coronary heart disease) during pregnancy? (5) How many over-the-counter medications did you take during pregnancy?

Medication Information Literacy Scale

The Medication Information Literacy Scale (MILS) is designed to assess the MIL of pregnant women. It uses a 5-point Likert scale and consists of 22 items, including the 5 dimensions of medication information needs, medication information sources, medication information quality discrimination, medication information source awareness, and medication-taking behavior [11]. The MILS is scored by totaling the scores of all items. A higher score indicates a higher level of MIL. Pregnant women with scores below the threshold (80% of the total score) are considered to have insufficient MIL [32]. In this study, the Cronbach α coefficient of the total scale was 0.761 and the subdomains were approximately 0.564-0.824 [11].

Decision Self-Efficacy Scale

The decision self-efficacy scale, also called the decision confidence scale, is used to measure one's confidence in making decisions, including the ability to take part in joint medical decisions. The scale consists of 11 items with a 5-point scoring system (0=not confident at all to 4=very confident) [14]. A higher score is associated with a higher level of self-efficacy [33]. The Cronbach α coefficient was 0.92 [33].

Data Collection and Analysis

Data collection was divided into 3 points: pretest at baseline (T0), immediately after the intervention (T1), and at 4 weeks follow-up (T2). The basic information questionnaire was only assessed at T0. The MILS and the decision self-efficacy scale were measured at T0, T1, and T2 using web-based questionnaires. To ensure the quality of the electronic questionnaires, the researchers stipulated that each participant had only one opportunity to submit the questionnaire, and incomplete questionnaires were not accepted.

Two-sample 2-tailed *t* test and chi-square test were used to compare the difference at baseline between the 2 groups. In addition, 1-way ANOVA analysis and independent 2-tailed *t* test were used to explore the influencing factors of MIL and decision self-efficacy. Data from the MIL and decision self-efficacy scales at T0, T1, and T2 were tested using generalized estimation equations (GEE). Data at T1 and T2 were added as the outcome and data at T0 were added as covariates. Time points and groups were added as independent variables. The effect of time, group, and time×group were analyzed. If the interaction effect of the time×group was not significant, the main effect analysis was performed; and if the interaction effect of the time×group was significant, the simple effect analysis was performed. Cohen *d* was used to evaluate the magnitude of the difference between the 2 groups. According to the standard established by Cohen, the large, medium, and small effects of *d* was divided into 0.8, 0.5, and 0.2 [34]. Missing values were replaced by the mean value if the missing items were less than 10% of the entire items in one scale, while the mean values were deleted if the missing items were more than 10% [35]. Data were entered in the SPSS (version 22.0; IBM

XSL•FO
RenderX

Corp) and a 2-tailed *P* value of <.05 was considered statistically significant.

Ethical Considerations

This study was approved by the Medical Research Committee of Wuhan University Medical College (2020YF0083), and informed consent was obtained before any study procedures. All participants were informed of the right to withdraw voluntarily. Data were maintained confidential and participants were allowed to complete the questionnaire or scale anonymously. Complying with the report's comprehensiveness and authenticity, no data tampering or concealing of adverse results occurred. Maintaining the fairness principle, the CG was entitled to participate in the education course at the end of the study. Participants did not receive compensation for their participation.

Results

Baseline Characteristics of Participants

In total, 108 participants completed the baseline test. Their average age was 29.46 (SD 5.03) years, and the average gestational week was 18.31 (SD 2.76). There was no significant difference between the 2 groups regarding sociodemographic characteristics, medication use, MIL, and decision self-efficacy at baseline (*P*>.05; Table S2 in Multimedia Appendix 1). Among the 2 groups, the average scores of MIL and decision self-efficacy were 81.00 (SD 10.94) and 22.31 (SD 4.43), respectively, and only 31% of participants had sufficient MIL. The implementation and data collection process are presented in Figure 2. The CONSORT-EHEALTH (V 1.6.1) checklist was used to guide the reporting of this randomized controlled trial (Multimedia Appendix 2).

Regarding the influencing factors, ANOVA analysis showed that there were statistically significant differences in the MIL and decision self-efficacy of participants related to education levels (F_{107}=5.60; *P*=.001 and F_{107}=3.03; *P*=.03, respectively) and gestational weeks (F_{107}=0.80; *P*=.64 and F_{107}=2.45; *P*=.01, respectively).

Figure 2. CONSORT diagram study flow.

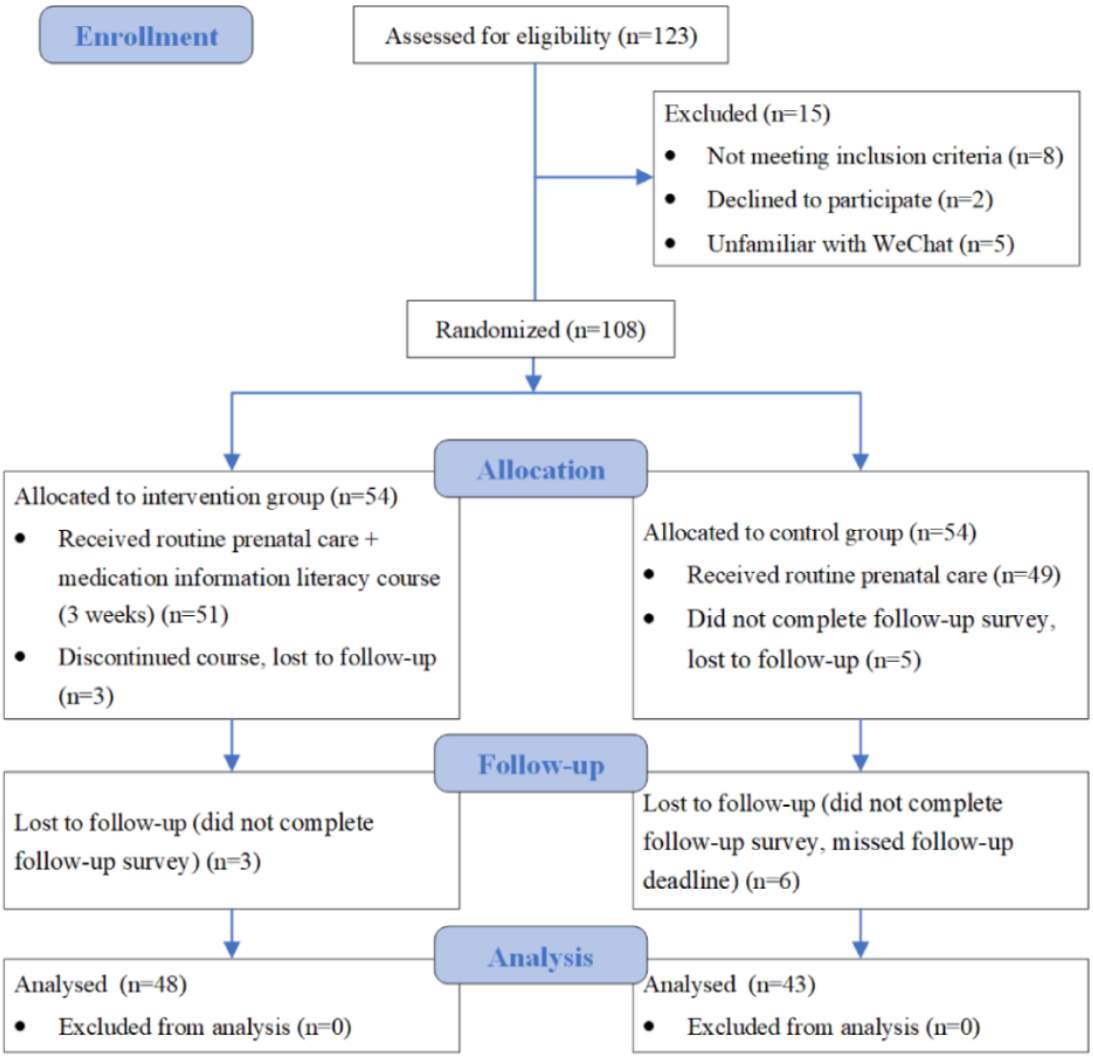

XSL•FO
RenderX

Primary Outcome: MIL of Participants

At baseline, MIL scores showed no statistically significant difference between the IG and the CG (Table S3 in Multimedia Appendix 1). The results of GEE indicated that there was no statistically significant difference in time×group interactions of MIL between the 2 groups (F_2=3.12; P=.21). Therefore, the main effects of the time and grouping factors were analyzed separately.

The results of the main effect analysis showed that there were statistically significant differences in MIL between the 2 groups at T1 and T2 (F_1=17.79; P<.001). The MIL of participants at T1 (mean difference 5.30, 95% CI 2.16-8.44; P<.001) and T2 (mean difference 6.00, 95% CI 2.80-9.20; P<.001) was significantly higher than at T0. Compared with the scores at T1, improvement in MIL in the follow-up phase was not as obvious, but it remained significantly improved compared to scores at T0 (Table S4 in Multimedia Appendix 1 and Figure 3).

Figure 3. Estimated marginal means of medication information literacy. CG: control group; IG: intervention group.

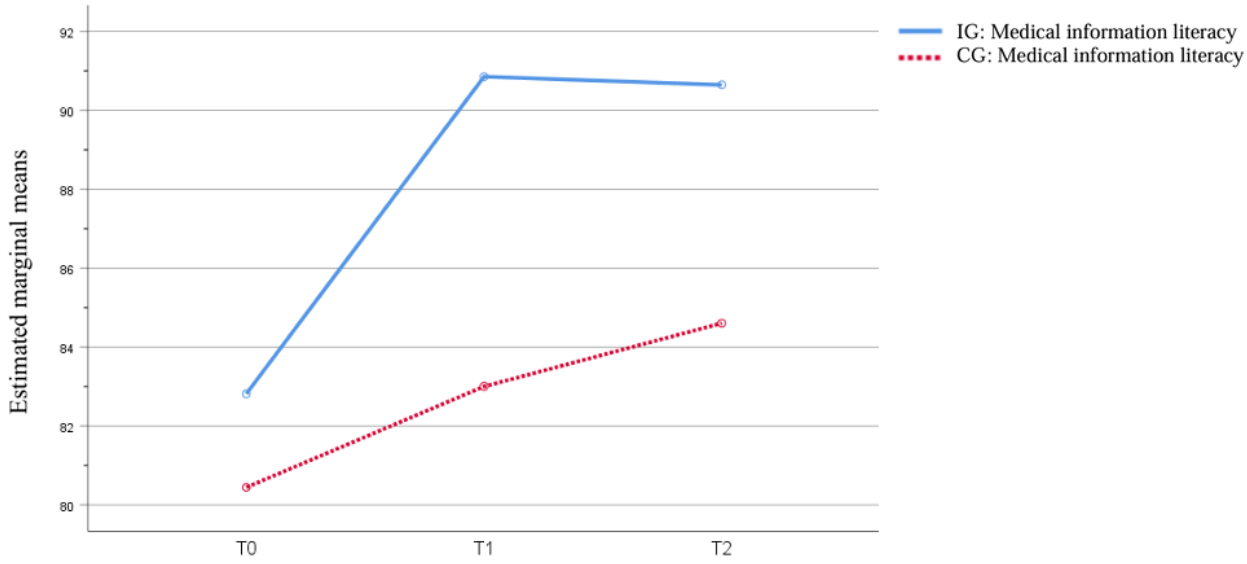

Secondary Outcome: Decision Self-Efficacy of Participants

At baseline, the scores on decision self-efficacy showed no statistically significant difference between the IG and the CG (Table S3 in Multimedia Appendix 1). The GEE results indicated that there was a significant difference in decision self-efficacy regarding the time factor, grouping factor, and time×group interactions (Table S4 in Multimedia Appendix 1 and Figure 4).

Since it was not meaningful to analyze the main effects of the time and grouping factors separately when the time×group interactions were significant (F_2=21.98; P<.001), the simple effect of time and grouping were analyzed. Results displayed in Table S5 in Multimedia Appendix 1 indicated a statistically significant difference in decision self-efficacy between the 2 groups at T1 (F_1=36.29; P<.001) and T2 (F_1=36.27; P<.001) compared to T0; and there was a statistically significant difference between the self-efficacy of the IG at T1, T2, and T0 (F_1=33.50; P<.001).

XSL•FO
RenderX

Figure 4. Estimated marginal means of decision self-efficacy. CG: control group; IG: intervention group.

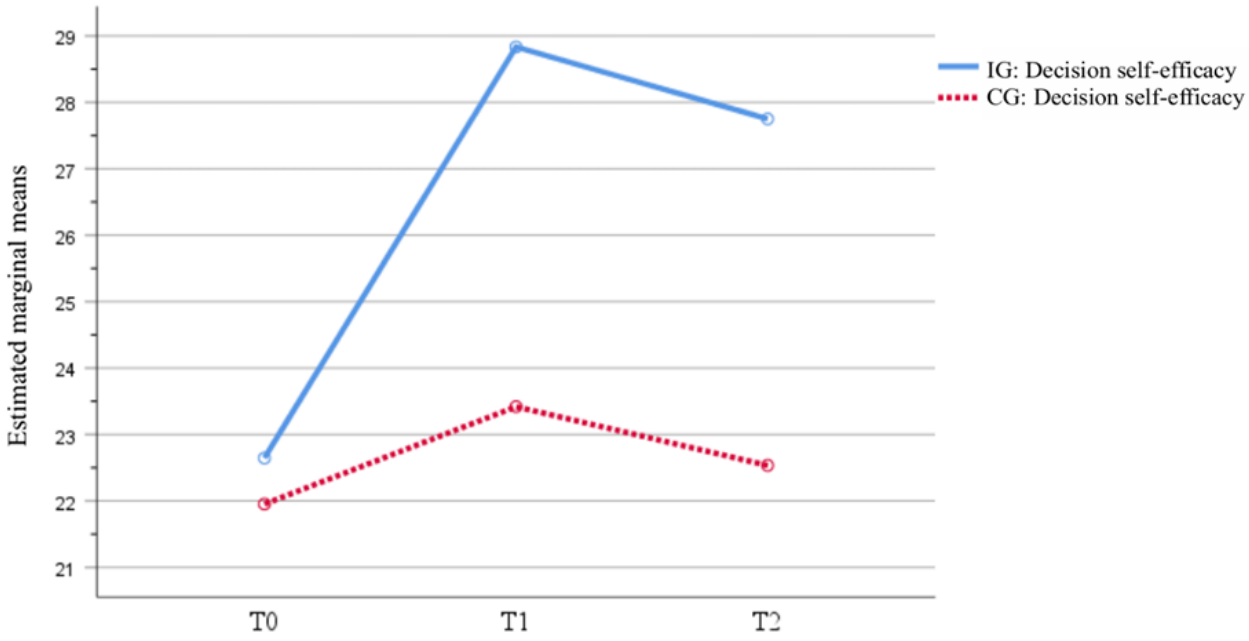

The Effect Size of MIL and Decision Self-Efficacy

The results of Cohen *d* showed: (1) Compared with the CG, MIL of the IG had a large effect (*d*=0.81) at T1 and a medium effect (*d*=0.59) at T2. In addition, decision self-efficacy of the IG had a large effect (*d*=1.26 and *d*=1.27, respectively) at both T1 and T2. (2) Compared with T0, MIL had a moderate effect (*d*=0.76 and *d*=0.69, respectively) at T1 and T2, while decision self-efficacy showed a large effect (*d*=1.47 and *d*=1.24, respectively) at T1 and T2.

Discussion

Principal Findings

The results of this study suggest that the current web-based education based on the WeChat platform can improve the MIL and decision self-efficacy of pregnant women.

Studies estimating the effects of web-based medication education courses on MIL and decision self-efficacy in pregnant women are scarce. However, the current findings of the beneficial effects of web-based medication education in improving MIL and contributing to the safe use of medications are in line with a study that focused on older people [36]. In addition, it has been shown that decision self-efficacy can be improved through web-based education [37].

This study found that most pregnant women showed insufficient MIL which is in line with the study by Zhang et al [11], which revealed that the MIL of pregnant women was related to education level, residence, occupation, income, and gestational weeks. However, this research found that the MIL of pregnant women was related to their education level. This could be explained by the relatively small sample size. Current findings noted a significant improvement in MIL after 3 weeks of intervention.

Pregnant women's negative beliefs about medications have been emphasized in previous studies [38-40]. Negative attitudes toward medications were also related to lower levels of pregnancy-related knowledge [41]. At the same time, excessive anxiety caused by overestimating the teratogenic risk of medication was one of the important reasons for low medication compliance and even self-discontinuation [5,38]. Furthermore, in the research of Barnes et al [42], the use of medication has been confirmed to be related to an individual's religious and spiritual beliefs. According to previous findings, most respondents reported they needed more information about medication use and were concerned about inconsistent information from different sources [40,43]. Accordingly, this study developed a web-based medication education course in an attempt to change negative attitudes toward medication use among pregnant women, improve their medication beliefs, and strengthen their skills in obtaining, evaluating, and using medication information. The results of the study confirmed the effectiveness of this web-based course.

In this study, the average score of decision self-efficacy was much lower than that in Scaffidi et al's [15] study, which might be interpreted that participants in this study were primiparous women. Compared with the decision self-efficacy of older people in previous studies [37,44], the decision self-efficacy of participants in this study was relatively low, assuming that the uniqueness and sensitivity of pregnancy make it more difficult for pregnant women to make medication-related decisions. Moreover, in this study, it was found that the decision self-efficacy of pregnant women was related to education level and gestational weeks, which is consistent with a previous study [15]. Lack of sufficient knowledge is the major factor affecting women's decision-making [45,46]. The intervention in this study not only included the popularization of knowledge but also paid attention to the transformation of beliefs and improvement of skills to help pregnant women actively

XSL•FO
RenderX

participate in medical decision-making, as well as express their concerns. After the 3-week course, the decision self-efficacy of the IG was significantly improved. The improvement of health literacy has a positive effect on perinatal women's participation in medical decision-making [47]. The combination of effective communication skills and interactive health information technology is a potentially important exploration approach that can improve women's enthusiasm for participating in medical decision-making [45]. The results of this study confirm the effectiveness of improving decision self-efficacy by increasing the MIL of pregnant women.

Strengths

The overall attrition rate in this study was 16%. It may be related to the flexible class time and limiting each session to 15 minutes [31], which has been shown to improve the validity of research findings [48]. Overall, this study confirmed the effectiveness of the intervention content and the feasibility of web-based education, providing new ideas for improving MIL levels of pregnant women, especially those with limited access to health care resources in remote areas.

Limitations

There are limitations to this study. Due to geographical restrictions, this study recruited pregnant women from only a tertiary hospital in a large city in central China and 71% of the participants were from cities, which might lead to sampling bias and further limit the external validity of the results. It could also limit the generalizability of findings to other areas. In addition, this study was carried out in the context of the COVID-19 pandemic, such that a potential impact on medication behavior cannot be excluded. Another limitation of this study centers on the researchers' inability to adequately monitor the time that participants spent completing the learning modules. The researchers sent daily course updates to the women to encourage participation, help focus on content areas, and improve learning outcomes. Other teaching-learning strategies should be explored.

At the 4-week follow-up, MIL and decision self-efficacy were found to have declined slightly. Extending the web-based course over a longer period, as well as the follow-up evaluation, could improve the retention of content and test its effect on maternal and fetal outcomes. Another limitation may be that most of the participants did not take medications during the intervention, thus, there was no exploration of the relationship between medication adherence and MIL.

Conclusions

Web-based medication education courses based on the TPB can improve pregnant women's MIL and decision self-efficacy. After receiving web-based education, the average score of MIL of pregnant women rose significantly. As a function of health literacy, the improvement of MIL is expected to reduce the occurrence of adverse medication events in pregnant women. Future studies are needed to explore the long-term effects of similar courses, as well as content retention. A combination of web-based and offline education is also worthy of further exploration in the MIL and decision self-efficacy of pregnant women.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We confirm that generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

The data sets generated during or analyzed during this study are not publicly available because the data also forms part of an ongoing study but are available from the corresponding author upon reasonable request.

Authors' Contributions

SL wrote the manuscript and collected all the data. HJC provided data sources, conducted research project management, and guided research methods. JZ wrote the manuscript and made tables. YBZ conducted all statistical analyses, drew figures, and conducted copyediting. RW, JG, and YQO guided the discussion section. SRR guided language modifications. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 30 KB - jmir_v27i1e54148_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1004 KB - jmir_v27i1e54148_app2.pdf]

XSL•FO
RenderX

References

1. Ke AB, Greupink R, Abduljalil K. Drug dosing in pregnant women: challenges and opportunities in using physiologically based pharmacokinetic modeling and simulations. CPT Pharmacometrics Syst Pharmacol 2018;7(2):103-110 [FREE Full text] [doi: 10.1002/psp4.12274] [Medline: 29349870]
2. Bedewi N, Sisay M, Edessa D. Drug utilization pattern among pregnant women attending maternal and child health clinic of tertiary hospital in eastern Ethiopia: consideration of toxicological perspectives. BMC Res Notes 2018;11(1):858 [FREE Full text] [doi: 10.1186/s13104-018-3966-5] [Medline: 30514377]
3. Thorpe PG, Gilboa SM, Hernandez-Diaz S, Lind J, Cragan JD, Briggs G, et al. Medications in the first trimester of pregnancy: most common exposures and critical gaps in understanding fetal risk. Pharmacoepidemiol Drug Saf 2013;22(9):1013-1018 [FREE Full text] [doi: 10.1002/pds.3495] [Medline: 23893932]
4. van Gelder MMHJ, de Jong-van den Berg LTW, Roeleveld N. Drugs associated with teratogenic mechanisms. Part II: a literature review of the evidence on human risks. Hum Reprod 2014;29(1):168-183. [doi: 10.1093/humrep/det370] [Medline: 24108217]
5. Jin J. Safety of medications used during pregnancy. JAMA 2022;328(5):486. [doi: 10.1001/jama.2022.8974] [Medline: 35916845]
6. Widnes SF, Schjøtt J. Risk perception regarding drug use in pregnancy. Am J Obstet Gynecol 2017;216(4):375-378 [FREE Full text] [doi: 10.1016/j.ajog.2016.12.007] [Medline: 27988271]
7. Rouamba T, Valea I, Bognini JD, Kpoda H, Mens PF, Gomes MF, et al. Safety profile of drug use during pregnancy at peripheral health centres in Burkina Faso: a prospective observational cohort study. Drugs Real World Outcomes 2018;5(3):193-206 [FREE Full text] [doi: 10.1007/s40801-018-0141-1] [Medline: 30155832]
8. Arbuckle C, Tomaszewski D, Brown L, Schommer J, Morisky D, Parlett-Pelleriti C, et al. Exploring the relationship of digital information sources and medication adherence. Comput Biol Med 2019;109:303-310 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.04.023] [Medline: 31100583]
9. Keselman A, Smith CA, Murcko AC, Kaufman DR. Evaluating the quality of health information in a changing digital ecosystem. J Med Internet Res 2019;21(2):e11129. [doi: 10.2196/11129] [Medline: 30735144]
10. EUROmediCAT: safety of medication use in pregnancy. Pharmacoepidemiol Drug Saf 2015;24 Suppl 2:1-2. [doi: 10.1002/pds.3865] [Medline: 26395592]
11. Zhang N, Wang L, Ouyang Y, Redding S. Survey on medication information literacy and influencing factors among pregnant Chinese women. J Matern Fetal Neonatal Med 2021;34(10):1619-1626 [FREE Full text] [doi: 10.1080/14767058.2019.1642869] [Medline: 31331258]
12. Weekes L. Access to medicines information for pregnant women. Pharmacy Practice and Res 2015;45(1):5-6. [doi: 10.1002/jppr.1061]
13. Zhong Z, Shi S, Duan Y, Shen Z, Zheng F, Ding S, et al. The development and psychometric assessment of Chinese medication literacy scale for hypertensive patients (C-MLSHP). Front Pharmacol 2020;11:490 [FREE Full text] [doi: 10.3389/fphar.2020.00490] [Medline: 32425773]
14. O'Connor A. User manual—decision self—efficacy scale. 1995. URL: http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decision_SelfEfficacy.pdf [accessed 2024-12-20]
15. Scaffidi RM, Posmontier B, Bloch JR, Wittmann-Price R. The relationship between personal knowledge and decision self-efficacy in choosing trial of labor after cesarean. J Midwifery Womens Health 2014;59(3):246-253. [doi: 10.1111/jmwh.12173] [Medline: 24850282]
16. Kirk JK, Krick S, Futrell D, Devora T, Caiola SM, Mason E, et al. Connecting pharmacy and literacy: the North Carolina medication information literacy project. Am J Pharm Educ 2000;64(3):277-283 WOS:000089671000009 ISSN:0002-9459 [FREE Full text]
17. Wen J, Yu G, Kong Y, Wei H, Zhao S, Liu F. Effects of a theory of planned behavior-based intervention on breastfeeding behaviors after cesarean section: a randomized controlled trial. Int J Nurs Sci 2021;8(2):152-160 [FREE Full text] [doi: 10.1016/j.ijnss.2021.03.012] [Medline: 33997128]
18. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.4135/9781446249215.n22]
19. Lin C, Broström A, Nilsen P, Pakpour AH. Using extended theory of planned behavior to understand aspirin adherence in pregnant women. Pregnancy Hypertens 2018;12:84-89. [doi: 10.1016/j.preghy.2018.04.001] [Medline: 29674206]
20. Jalambadani Z, Borji A, Delkhosh M. The effect of education based on the theory of planned behavior on iron supplementation among pregnant women. Korean J Fam Med 2018;39(6):370-374 [FREE Full text] [doi: 10.4082/kjfm.17.0141] [Medline: 29991197]
21. Karimian Z, Sadat Z, Afshar B, Hasani M, Araban M, Kafaei-Atrian M. Predictors of self-medication with herbal remedies during pregnancy based on the theory of planned behavior in Kashan, Iran. BMC Complement Med Ther 2021;21(1):211 [FREE Full text] [doi: 10.1186/s12906-021-03353-8] [Medline: 34404404]
22. Wu H, Sun W, Huang X, Yu S, Wang H, Bi X, et al. Online antenatal care during the COVID-19 pandemic: opportunities and challenges. J Med Internet Res 2020;22(7):e19916 [FREE Full text] [doi: 10.2196/19916] [Medline: 32658860]

XSL•FO
RenderX

23. Merih YD, Karabulut Ö, Sezer A. Is online pregnant school training effective in reducing the anxiety of pregnant women and their partners during the COVID-19 pandemic? Bezmialem Sci 2021;9(1):13-24. [doi: 10.14235/bas.galenos.2020.4718]
24. Wu Q, Huang Y, Liao Z, van Velthoven MH, Wang W, Zhang Y. Effectiveness of WeChat for improving exclusive breastfeeding in Huzhu county China: randomized controlled trial. J Med Internet Res 2020;22(12):e23273 [FREE Full text] [doi: 10.2196/23273] [Medline: 33270026]
25. Carolan-Olah M, Sayakhot P. A randomized controlled trial of a web-based education intervention for women with gestational diabetes mellitus. Midwifery 2019;68:39-47. [doi: 10.1016/j.midw.2018.08.019] [Medline: 30343264]
26. Ding B, Gou B, Guan H, Wang J, Bi Y, Hong Z. WeChat-assisted dietary and exercise intervention for prevention of gestational diabetes mellitus in overweight/obese pregnant women: a two-arm randomized clinical trial. Arch Gynecol Obstet 2021;304(3):609-618. [doi: 10.1007/s00404-021-05984-1] [Medline: 33570656]
27. Meng C. An investigation of the effect of nutrition intervention on pregnancy outcome on WeChat platform. Henan J Prev Med 2020;31(5):352-353. [doi: 10.13515/j.cnki.hnjpm.1006-8414.2020.05.010]
28. Randomization and online databases for clinical trials. Sealed Envelope. URL: https://www.sealedenvelope.com/ [accessed 2024-12-27]
29. Weiner CP, Buhimschi C. Drugs for Pregnant and Lactating Women. 2nd Edition. Philadelphia, PA: W.B. Saunders; 2014.
30. FDA. Helping patients and health care professionals better understand the risks and benefits of medications for pregnant and breastfeeding women. 2014. URL: https://www.fda.gov/consumers/consumer-updates/pregnant-breastfeeding-fda-aims-improve-drug-information [accessed 2024-12-27]
31. O'Sullivan TA, Cooke J, McCafferty C, Giglia R. Online video instruction on hand expression of colostrum in pregnancy is an effective educational tool. Nutrients 2019;11(4):883 [FREE Full text] [doi: 10.3390/nu11040883] [Medline: 31010174]
32. Nie XQ, Li YH, Li L. Statistic analysis of 2012 Chinese residents health literacy monitoring. Chin J Health Educ 2014;30:178-181 original in Chinese.
33. Smith CA, Chang E, Gallego G, Balneaves LG. An education intervention to improve health literacy and decision making about supporting self-care among older Australians: a study protocol for a randomised controlled trial. Trials 2017;18(1):441 [FREE Full text] [doi: 10.1186/s13063-017-2182-2] [Medline: 28950888]
34. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
35. Zhu X. Comparison of four methods for handing missing data in longitudinal data analysis through a simulation study. Open J Stat 2014;04(11):933-944 [FREE Full text]
36. Qvarfordt M, Throfast V, Petersson G, Hammar T, Hellström L. Web-based education of the elderly improves drug utilization literacy: a randomized controlled trial. Health Inf J 2021;27(1):1460458220977585 [FREE Full text] [doi: 10.1177/1460458220977585] [Medline: 33446035]
37. Smith CA, Chang E, Gallego G, Khan A, Armour M, Balneaves LG. An education intervention to improve decision making and health literacy among older Australians: a randomised controlled trial. BMC Geriatr 2019;19(1):129 [FREE Full text] [doi: 10.1186/s12877-019-1143-x] [Medline: 31064336]
38. Ceulemans M, Lupattelli A, Nordeng H, Odalovic M, Twigg M, Foulon V. Women's beliefs about medicines and adherence to pharmacotherapy in pregnancy: opportunities for community pharmacists. Curr Pharm Des 2019;25(5):469-482. [doi: 10.2174/1381612825666190321110420] [Medline: 30907309]
39. Lupattelli A, Picinardi M, Einarson A, Nordeng H. Health literacy and its association with perception of teratogenic risks and health behavior during pregnancy. Patient Educ Couns 2014;96(2):171-178. [doi: 10.1016/j.pec.2014.04.014] [Medline: 24862909]
40. Twigg MJ, Lupattelli A, Nordeng H. Women's beliefs about medication use during their pregnancy: a UK perspective. Int J Clin Pharm 2016;38(4):968-976 [FREE Full text] [doi: 10.1007/s11096-016-0322-5] [Medline: 27241342]
41. Selinger CP, Eaden J, Selby W, Jones DB, Katelaris P, Chapman G, et al. Inflammatory bowel disease and pregnancy: lack of knowledge is associated with negative views. J Crohns Colitis 2013;7(6):e206-e213. [doi: 10.1016/j.crohns.2012.09.010] [Medline: 23040449]
42. Barnes LA, Barclay L, McCaffery K, Aslani P. Complementary medicine products information-seeking by pregnant and breastfeeding women in Australia. Midwifery 2019;77:60-70. [doi: 10.1016/j.midw.2019.06.011] [Medline: 31255910]
43. Nordeng H, Ystrøm E, Einarson A. Perception of risk regarding the use of medications and other exposures during pregnancy. Eur J Clin Pharmacol 2010;66(2):207-214. [doi: 10.1007/s00228-009-0744-2] [Medline: 19841915]
44. Hall AK, Bernhardt JM, Dodd V. Older adults' use of online and offline sources of health information and constructs of reliance and self-efficacy for medical decision making. J Health Commun 2015;20(7):751-758 [FREE Full text] [doi: 10.1080/10810730.2015.1018603] [Medline: 26054777]
45. Chen SW, Yang CC, Te JC, Tsai YL, Shorten B, Shorten A. Birth choices after caesarean in Taiwan: a mixed methods pilot study of a decision aid for shared decision making. Midwifery 2021;95:102920. [doi: 10.1016/j.midw.2020.102920] [Medline: 33515972]
46. Wise MR, Sadler L, Shorten B, van der Westhuizen K, Shorten A. Birth choices for women in a 'Positive Birth after Caesarean' clinic: randomised trial of alternative shared decision support strategies. Aust N Z J Obstet Gynaecol 2019;59(5):684-692. [doi: 10.1111/ajo.12955] [Medline: 30773608]

XSL•FO
RenderX

47. Chung K, Cho HY, Kim YR, Jhung K, Koo HS, Park JY. Medical help-seeking strategies for perinatal women with obstetric and mental health problems and changes in medical decision making based on online health information: path analysis. J Med Internet Res 2020;22(3):e14095 [FREE Full text] [doi: 10.2196/14095] [Medline: 32130139]
48. Polit DF, Beck CT. Nursing Research: Principles and Methods. Pennsylvania, United States: Lippincott Williams & Wilkins; 2004.

Abbreviations

CG: control group
GEE: generalized estimation equations
IG: intervention group
MIL: medication information literacy
MILS: Medication Information Literacy Scale
TPB: theory of planned behavior

Edited by T de Azevedo Cardoso; submitted 31.10.23; peer-reviewed by H Alshawaf, X Su; comments to author 22.07.24; revised version received 16.10.24; accepted 25.11.24; published 22.01.25.

Please cite as:
Li S, Chen HJ, Zhou J, Zhouchen YB, Wang R, Guo J, Redding SR, Ouyang YQ
Effectiveness of a Web-Based Medication Education Course on Pregnant Women's Medication Information Literacy and Decision Self-Efficacy: Randomized Controlled Trial
J Med Internet Res 2025;27:e54148
URL: https://www.jmir.org/2025/1/e54148
doi:10.2196/54148
PMID:

©Suya Li, Hui-Jun Chen, Jie Zhou, Yi-Bei Zhouchen, Rong Wang, Jinyi Guo, Sharon R Redding, Yan-Qiong Ouyang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Novel Artificial Intelligence–Enhanced Digital Network for Prehospital Emergency Support: Community Intervention Study

Ji Hoon Kim[1*], MD, MPH, PhD; Min Joung Kim[1*], MD, PhD; Hyeon Chang Kim[2], MD, PhD; Ha Yan Kim[3], MS; Ji Min Sung[4], PhD; Hyuk-Jae Chang[5], MD, PhD

[1]Department of Emergency Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Preventative Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Biomedical System Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Department of Data Business,, Ontacthealth, Seoul, Republic of Korea

[5]Department of Cardiology, Yonsei University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyuk-Jae Chang, MD, PhD
Department of Cardiology
Yonsei University College of Medicine
50 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 2460
Fax: 82 2 2227 7908
Email: hjchang@yuhs.ac

Abstract

Background: Efficient emergency patient transport systems, which are crucial for delivering timely medical care to individuals in critical situations, face certain challenges. To address this, CONNECT-AI (CONnected Network for EMS Comprehensive Technical-Support using Artificial Intelligence), a novel digital platform, was introduced. This artificial intelligence (AI)–based network provides comprehensive technical support for the real-time sharing of medical information at the prehospital stage.

Objective: This study aimed to evaluate the effectiveness of this system in reducing patient transport delays.

Methods: The CONNECT-AI system provided 3 key AI services to prehospital care providers by collecting real-time patient data from the scene and hospital resource information, such as bed occupancy and the availability of emergency surgeries or procedures, using 5G communication technology and internet of things devices. These services included guidance on first aid, prediction of critically ill patients, and recommendation of the optimal transfer hospital. In addition, the platform offered emergency department medical staff real-time clinical information, including live video of patients during transport to the hospital. This community-based, nonrandomized controlled intervention study was designed to evaluate the effectiveness of the CONNECT-AI system in 2 regions of South Korea, each of which operated an intervention and a control period, each lasting 16 weeks. The impact of the system was assessed based on the proportion of patients experiencing transfer delays.

Results: A total of 14,853 patients transported by public ambulance were finally selected for analysis. Overall, the median transport time was 10 (IQR 7-14) minutes in the intervention group and 9 (IQR 6-13) minutes in the control group. When comparing the incidence of transport time outliers (>75%), which was the primary outcome of this study, the rate was higher in the intervention group in region 1, but significantly reduced in region 2, with the overall outlier rate being higher in the intervention group (27.5%-29.7%, *P*=.04). However, for patients with fever or respiratory symptoms, the group using the system showed a statistically significant reduction in outlier cases (36.5%-30.1%, *P*=.01). For patients who received real-time acceptance signals from the hospital, the reduction in the percentage of 75% outliers was statistically significant compared with those without the system (27.5%-19.6%, *P*=.02). As a result of emergency department treatment, 1.5% of patients in the control group and 1.1% in the intervention group died (*P*=.14). In the system-guided optimal hospital transfer group, the mortality rate was significantly lower than in the control group (1.54%-0.64%, *P*=.01).

XSL•FO
RenderX

Conclusions: The present digital emergency medical system platform offers a novel approach to enhancing emergency patient transport by leveraging AI, real-time information sharing, and decision support. While the system demonstrated improvements for certain patient groups facing transfer challenges, further research and modifications are necessary to fully realize its benefits in diverse health care contexts.

Trial Registration: ClinicalTrials.gov NCT04829279; https://clinicaltrials.gov/study/NCT04829279

(J Med Internet Res 2025;27:e58177) doi:10.2196/58177

KEYWORDS

emergency patient transport; transport time; artificial intelligence; smartphone; mobile phone

Introduction

Emergency patient transport plays a critical role in providing timely and appropriate care to individuals in emergencies [1]. By swiftly and safely transporting patients to the nearest health care facility equipped to handle their specific needs, it helps to improve their chances of receiving timely medical intervention and contributes to better patient outcomes [2]. On the contrary, delays faced by the emergency medical services (EMS) in transport of patients to health care facilities can result in increased patient morbidity, mortality, and complications [3-5]. Ambulance diversion, a major factor delaying patient transport, is a practice in EMS where an ambulance is redirected from its original destination hospital to another medical facility [6,7]. Ambulance diversion typically occurs when the emergency department (ED) of a hospital is temporarily unable to accept new patients due to high patient volumes, lack of available beds, or other capacity constraints [8,9]. As the demand for medical services increased with the COVID-19 pandemic, ambulance diversions became more frequent [10,11]. Eventually, as cases of severely ill patients dying in ambulances without being able to get to the ED became more common in Korea, the country amended its Emergency Medical Services Act to prohibit emergency medical institutions from refusing to accept critically ill emergency patients.

Along with securing a transfer hospital as soon as possible, it is crucial to provide accurate information about the condition of the patient to the medical staff of the ED to which they will be transferred in order to provide optimal treatment [12-15]. In addition to the most basic clinical information such as their main symptoms, blood pressure, and state of consciousness, information on interventions performed by the prehospital care provider on-site or during transport (eg, oxygen administration, intravenous fluids, medications, or immobilization techniques) and time stamps (eg, time of the report, arrival at the site, and time at which treatment was performed) help the receiving ED medical staff to prepare for the arrival of the patient and deliver appropriate care [16,17]. However, in most prehospital phases, the transmission of patient information between prehospital care provider and ED medical staff mainly relies on telephones, and comprehensive information cannot be sufficiently delivered, leading to a potential risk of communication errors such as mishearing [18,19].

With the rapid development of digital information devices such as tablets and smartphones, various medical information systems are being introduced to EMS and emergency medical institutions [18-21]. These technologies allow prehospital care providers in the field to find hospitals with available resources and ED medical staff to receive patient information in real-time to provide medical services in continuity through mobile phones [18,19]. The use of these technologies is expected to enhance communication and data sharing between prehospital care providers and ED medical staff, ensuring a smoother transition of care and improving patient outcomes. However, the evidence is insufficient, and the process has not yet been widely commercialized.

We developed a new medical information system called CONnected Network for EMS Comprehensive Technical-support using Artificial Intelligence (CONNECT-AI). CONNECT-AI system used voice and image recognition technology for real-time sharing of information at the prehospital stage. It also included AI algorithms that supported first aid by prehospital care providers and the selection of transfer hospitals, as well as a platform that shared information and videos of patients being transported to the ED in real time. This study aimed to report on the development of the CONNECT-AI system and to evaluate its effectiveness in reducing transfer delays and facilitating the timely transfer of emergency patients. Our hypothesis was that the implementation of this system would decrease the occurrence of transport delays for emergency patients.

Methods

Study Setting

The study was conducted in 2 geographically adjacent areas, each under the jurisdiction of different local fire departments: the northwest region of Seoul (region 1) and Goyang City in Gyeonggi-do (region 2). The demographic characteristics and the composition of the EMS in these 2 regions are detailed in Table 1. A total of 5 fire stations and 9 EDs participated in this study. Prehospital care providers working in the NFA (National Fire Agency) assessed patients' urgent conditions and, when necessary, performed a limited range of advanced care techniques in the prehospital setting. These techniques included administering intravenous fluids such as normal saline and glucose solution, placing advanced airways, and administering specific medications under the supervision of a medical director [22]. All patients assessed by prehospital care providers were transported to one of the EDs in accordance with the standard field first aid protocol developed by the NFA. This protocol included standardized first aid procedures and guidelines for selecting a hospital for transfer [23]. Prehospital care providers use their cell phones when they need to communicate with the

XSL•FO
RenderX

medical staff of EDs to select a transfer hospital. During the study period, prehospital care providers were supposed to check for isolation unit availability at EDs to which they transfer patients with febrile or respiratory symptoms in response to COVID-19 pandemic. Prehospital care providers are required to manually record the prehospital run sheets (PHRS) managed by the NFA using tablet devices.

Table 1. Demographic characteristics and composition of the emergency medical system of the 2 regions.

Characteristics	Region 1	Region 2
Area (square miles)	27.5	110.5
Resident population (million), n	11	10
Fire station, n	3	2
Ambulance, n	20	17
Emergency department, n (n)[a]		
Level 1	2 (2)	4 (4)
Level 2	3 (3)	2 (0)

[a]The number in parentheses indicates the number of emergency departments that participated in this study.

Study Design and Population

This community-based nonrandomized controlled intervention study with a blinded outcome assessment was designed to evaluate the effectiveness of the CONNECT-AI system for transporting patients by public ambulances in 2 communities (region 1 and region 2) that were selected as demonstration areas. The intervention group was defined as patients who were transported by public ambulance with an activated CONNECT-AI system, and the control group was defined as patients who were transported by public ambulance without an activated CONNECT-AI system (which is the same as the conventional practice). Public ambulances in region 1 underwent a 16-week intervention period beginning April 19, 2021, followed by a 1-week washout period and then a 16-week control period that ended on December 31. On the contrary, the public ambulances in region 2 had opposite intervention and control periods during the same time period.

During the study period, all patients transported by public ambulance in regions 1 and 2 to the 9 participating EDs were enrolled in the study. During the intervention period, all public ambulances in the regions were required to use the CONNECT-AI system, but if the situation was determined by prehospital care providers to be not appropriate for the use of the system, it was withdrawn, and the transported patients were excluded from the study. Patients were also excluded from enrollment if the ambulance had a special mission, such as transporting a person with COVID-19 infection or if the patient arrived at the ED but returned home or went to an outpatient clinic without treatment at the ED.

System Development and Protocol (Intervention)

The CONNECT-AI system captures all prehospital information in real-time through 5G communication technology and Internet of Things (IoT) devices. It provides 3 AI services on a cloud-based platform that can support the aids and decisions of prehospital care providers. The platform also enables medical staff in the ED to share real-time information in the prehospital setting before the ambulance arrives at the hospital. Furthermore, the present system can provide prehospital care providers with real-time information about ED crowding status and the availability of emergency medical resources at the hospital (Figure 1). This system collects patient information in real time from the transport scene during the prehospital stage using 5G communication technology and internet of things devices, while also gathering real-time data on emergency department crowding status and the availability of emergency medical resources from national emergency department information system. The collected information is processed through a cloud-based platform, providing 3 artificial intelligence (AI) services: standardized protocols for first aid, prediction of critical patients, and selection of the optimal hospital for transport. In addition, this platform enables real-time sharing of prehospital information with emergency department medical staff before the ambulance arrives at the hospital through emergency medical services kiosks and emergency department kiosks.

XSL•FO
RenderX

Figure 1. The overall architecture of the CONNECT-AI (CONnected Network for EMS Comprehensive Technical-Support using Artificial Intelligence) system. AI: artificial intelligence; ED: emergency department; EMS: emergency medical services.

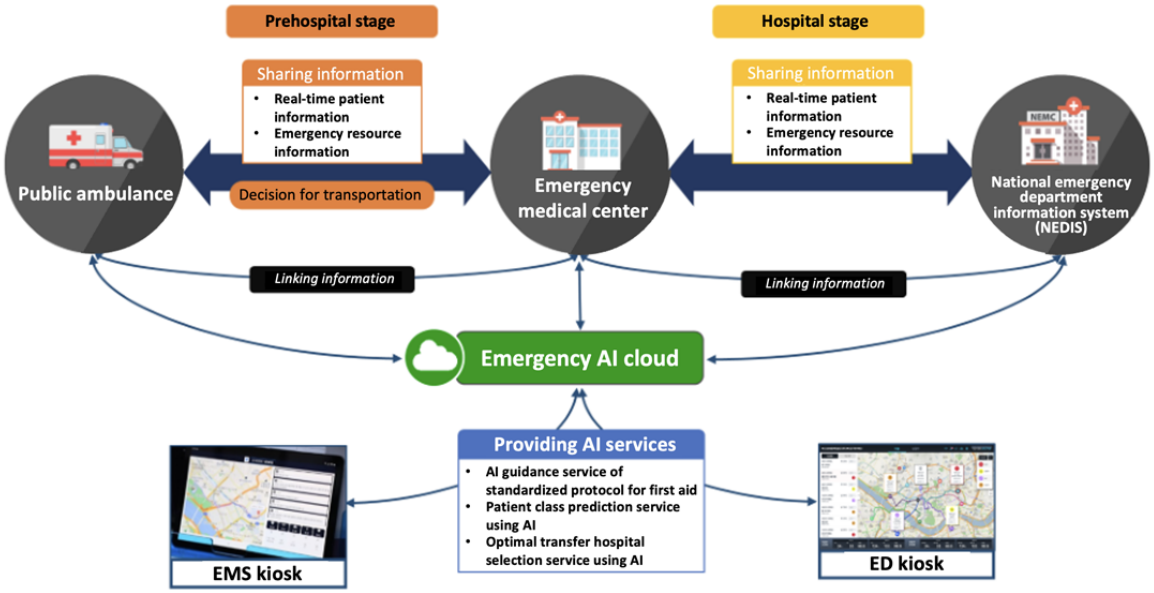

To use the CONNECT-AI system, a networking device that can collect and communicate video, voice, and vital sign information with the hospital in real time through a digital platform was installed in all ambulances; a schematic representation of this data flow is shown in Figure 2.

Figure 2. The flow of multimodal data from the public ambulance to the emergency department.

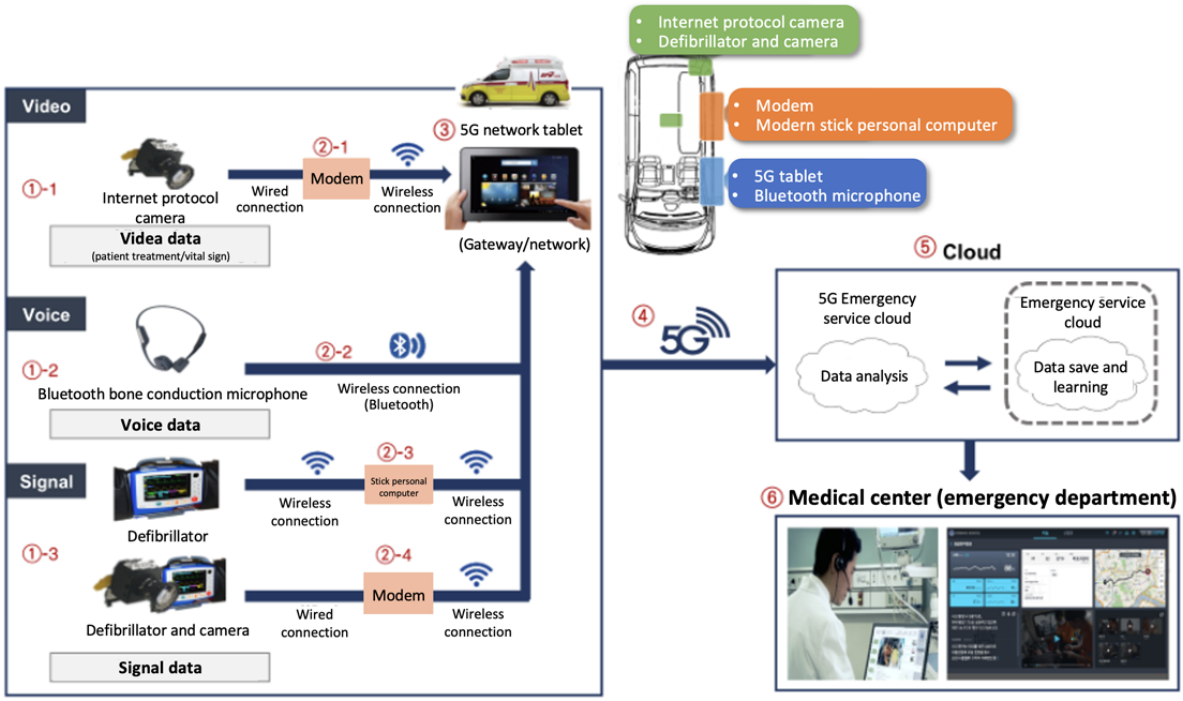

Real-time video, audio, and patient vital sign data, collected from a Bluetooth bone conduction microphone worn by prehospital providers and from a camera and defibrillator installed in the public ambulance, are transmitted to the medical staff in the emergency department by the 5G emergency service cloud.

Before the intervention period, all prehospital care providers were trained to use the CONNECT-AI system through a 1-hour in-person lecture with hands-on training. During the intervention period, prehospital care providers wore equipment (bone conduction microphones and EMS kiosks) to use the system, in contrast to the control period when they followed traditional practice. The following 3 AI assistant services have been developed to support prehospital care providers in performing first aid that was appropriate to the patient's condition and to facilitate the transfer of the patient to the optimal hospital as quickly as possible: (1) AI guidance service of standardized protocol for first aid, (2) patient class prediction service using AI, and (3) optimal transfer hospital selection service using AI. Textbox 1 describes the details and purpose of the 3 AI assistant

XSL•FO
RenderX

services as compared with the current prehospital practice without the CONNECT-AI system. Previous research has introduced the algorithm used in the optimal transfer hospital selection service [22].

Textbox 1. Three artificial intelligence (AI) assistant services for prehospital care providers.

AI guidance service of a standardized protocol for first aid

- Service description: This service uses real-time patient information as an input to predict the treatment that should be performed on the patient at the hospital and outputs the results to the prehospital care providers.
- Current practice: Prehospital care providers use their experience and medical knowledge to assess patient severity.
- Expectancy of effects: Reduce the number of unpredictable, critically ill patients for inexperienced prehospital care providers.

Patient class prediction service using AI

- Service description: This service uses real-time patient information as input to predict the treatment that should be performed on the patient at the hospital and outputs the results to the prehospital care providers.
- Current practice: Prehospital care providers use their experience and medical knowledge to assess patient severity.
- Expectancy of effects: Reduce the number of unpredictable, critically ill patients for inexperienced prehospital care providers.

Optimal transfer hospital selection service using AI

- Service description: This service provides prehospital care providers with a real-time, prioritized list of hospitals that can provide predicted care, and allows for simultaneous requests for acceptance to multiple hospitals.
- Current practice: Prehospital care providers use their own judgment to determine the hospitals for transfer or use their cell phones to sequentially contact hospital medical staff to inquire about acceptance for transportation.
- Expectancy of effects: By using data communicated in real-time through the CONnected Network for EMS (emergency medical services) Comprehensive Technical-support using Artificial Intelligence system, prehospital care providers can quickly arrange the optimal hospital for a patient's condition.

The EDs in this study were equipped with kiosks to monitor the status of patients transported by public ambulance and to communicate with ambulances. This kiosk also allowed the physician in charge of the ED to monitor the real-time transport status of public ambulances enroute to the hospital, as well as information on available beds and emergency capacity in the ED (Figure 3).

Figure 3. Main screens provided by the emergency department kiosk, (A) a screen that monitors a patient's real-time transport status and displays the predicted severity of the patient, (B) a screen that shows the public ambulances that are currently on their way to the hospital, and (C) a screen of real-time emergency resource information that be available in hospitals with kiosks installed.



The medical staff were provided with 1-hour hands-on training on how to communicate with prehospital care providers in real time through the kiosk. During the control period, prehospital care providers used their cell phones to communicate the status of patients requiring transport to ED medical staff and to inquire about patient acceptance sequentially as a traditional practice.

XSL•FO
RenderX

In contrast, during the intervention period, the CONNECT-AI system provided prehospital care providers with a hospital list that was appropriate for the patient's condition and allowed the patient to be transferred without additional contact if the ED medical staff did not send an unacceptable message on the system. In cases where prehospital care providers determined they needed to transport to EDs that showed that they could not accept the patient, providers clicked a button on the EMS kiosk to simultaneously send a request for acceptance to these EDs; the transport was then performed when they received an acceptance signal from the ED (Multimedia Appendix 1).

Data Collection

In our study, prehospital stage information was obtained through the PHRS, which is electronically stored to provide basic prehospital operation information under the NFA. Data extracted from the PHRS included age, gender, past medical history, mental status, vital signs (blood pressure, pulse rate, respiratory rate, oxygen saturation, and body temperature), patient symptoms, and the first aid timeline. Mental status was assessed using the AVPU scale (A for awake, V for verbal response, P for pain response, and U for unresponsiveness). Hospital stage information was obtained from a dataset submitted by the participating medical institutions to the National Emergency Department Information System (NEDIS); this is a nationwide computerized system used to collect and analyze medical information of patients who visit EDs in South Korea, in accordance with the Act on Emergency Medical Care and managed according to the standardized protocol distributed by the National Emergency Medical Center [22]. Information on patient severity and ED treatment outcomes for those transported by public ambulance was extracted from the NEDIS. Patient severity was assessed using the Korean Triage and Acuity Scale (KTAS), a classification system used in EDs across Korea. The KTAS, derived from the Canadian Triage and Acuity Scale, is a 5-level classification system where 1 indicates resuscitation and 5 represents nonemergent cases [24]. Based on the linkage of these 2 main datasets, log records of events that occurred when prehospital care providers used the system during the intervention period were added by extracting them from a cloud-based digital platform.

Outcome

The primary outcome of our study is the proportion of outliers with long transport times among the transported patients. Transport time is defined as the period from the time the prehospital care provider left the scene to the time of arrival at the hospital. Patient mortality was defined as the death of a patient who was alive upon arrival but died during treatment in the ED.

Sample Size Calculation

In 2018, the mean of transport times for one year were 21 (SD 10) minutes for region 1, 16 (SD 8) minutes for region 2, and 20 (SD 25) minutes overall. Based on these figures, the primary outcome of this study, transport time, was assumed to show an improvement of more than 10% (a 2-minute difference) when using the CONNECT AI system, compared with the conventional ambulance transport standard. The largest standard deviation of 25 minutes was conservatively chosen for calculations, and it was predicted that a statistical power of over 90% could be achieved. Since transport time is a continuous variable, the sample size was calculated based on a 2-sample *t* test, resulting in an estimate of approximately 10,000 cases. A proposal was made to the regional administrative authorities for a pilot period during which 10,000 transports were expected to occur, and this proposal was accepted by the regional administrative authorities.

Statistical Analysis

Continuous variables were expressed as mean values with standard deviations and compared between groups using the Student *t* test or Mann-Whitney *U* test. Categorical variables were expressed as frequencies and percentages and compared using the chi-square test. All tests were 2-sided, with a statistical significance of $P<.05$. All statistical analyses were performed using SAS software (version 9.4; SAS Institute).

Ethical Considerations

This community-based interventional study was performed in accordance with the revised Declaration of Helsinki and was approved by the institutional review board of Severance Hospital, South Korea on April 14, 2021 (approval number 4-2021-0217). This research was prospectively registered on ClinicalTrials.gov (NCT04829279). As the patient information covered by this study was within the scope of being provided by the NFA and the National Emergency Medical Center (NEMC) under the Ministry of Health and Welfare with personally identifiable information removed based on the law, informed consent from the patients included in the study was waived. However, as audio and video data of the prehospital care providers were collected and processed during the intervention, written informed consent for the use of this information was obtained from them voluntarily before the start of the intervention, and no material and financial compensation was provided for participation in the study.

Results

Population Characteristics

During the study period, there were a total of 16,031 patients transported by public ambulance who met the inclusion criteria. After excluding patients transferred during the washout period and those eligible for exclusion, 14,853 patients were finally selected for analysis. Of them, 1269 (29.0%) and 957 (11.6%) patients were included in the intervention group from regions 1 and 2, respectively (Figure 4). The baseline characteristics of the intervention and control groups are presented in Table 2. The proportion of the intervention group enrolled was higher in region 1 than in region 2. Febrile and respiratory symptoms were more frequent in the intervention group than in the control group.

XSL•FO
RenderX

Figure 4. Patient enrollment flow. AI: artificial intelligence; ED: emergency department.

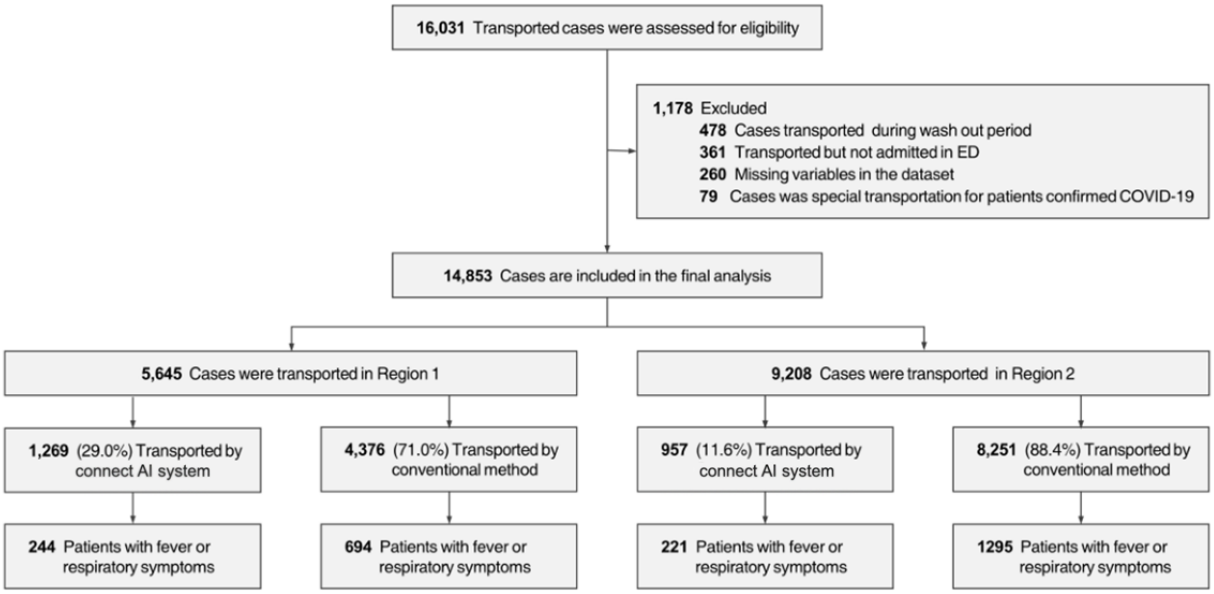

XSL•FO
RenderX

Table 2. Baseline characteristics.

Characteristics	Total			Region 1			Region 2		
	Control (n=12,627)	Intervention (n=2226)	*P* value	Control (n=8251)	Intervention (n=957)	*P* value	Control (n=4376)	Intervention (n=1269)	*P* value
Age (years), mean (SD)	56.88 (23.38)	58.00 (23.65)	.04	57.98 (22.87)	60.45 (23.13)	.002	54.81 (24.17)	56.15 (23.88)	.08
Sex (male), n (%)	6635 (52.55)	1155 (51.89)	.60	4218 (51.12)	533 (55.69)	.007	2417 (55.23)	622 (49.01)	<.001
KTAS[a] score, n (%)			<0.001			.004			<.001
1	467 (3.70)	74 (3.32)		288 (3.49)	28 (2.93)		179 (4.09)	46 (3.62)	
2	1187 (9.40)	276 (12.40)		837 (10.14)	118 (12.33)		350 (8.00)	158 (12.45)	
3	5267 (41.71)	962 (43.22)		3547 (42.99)	396 (41.38)		1720 (39.31)	566 (44.60)	
4	5048 (39.98)	785 (35.27)		3115 (37.75)	338 (352.32)		1933 (44.17)	447 (35.22)	
5	658 (5.21)	129 (5.80)		464 (5.62)	77 (8.05)		194 (4.43)	52 (4.10)	
Fever or respiratory symptoms, n (%)	1989 (15.75)	465 (20.89)	<.001	1295 (15.7)	221 (23.09)	<.001	694 (15.86)	244 (19.26)	.005
Systolic blood pressure, mean (SD)	129.01 (40.41)	129.11 (40.83)	.91	131.43 (39.51)	134.49 (36.62)	.01	124.44 (41.68)	125.06 (43.31)	.64
Diastolic blood pressure, mean (SD)	75.03 (23.9)	74.09 (23.33)	.09	75.81 (23.61)	75.96 (21.67)	.84	73.56 (24.37)	72.68 (24.42)	.25
Heart rate, mean (SD)	81.78 (32.56)	81.87 (25.33)	.89	82.59 (33.96)	82.95 (21.91)	.66	80.25 (29.67)	81.05 (27.61)	.37
Respiratory rate, mean (SD)	17.90 (5.46)	17.95 (5.22)	.68	18.42 (5.42)	18.84 (4.34)	.01	16.91 (5.39)	17.27 (5.7)	.04
Oxygen saturation, mean (SD)	37.16 (48.58)	56.30 (52.40)	<.001	23.89 (42.60)	28.30 (54.45)	.02	61.88 (49.37)	77.13 (39.60)	<.001
Body temperature, mean (SD)	36.05 (6.43)	36.38 5.85)	.01	36.12 (5.37)	36.27 (4.44)	.32	35.92 (8.05)	36.46 (6.72)	.01
Mental status[b], n (%)			.01			.04			.12
Awake	11,615 (92.06)	2041 (92.27)		7614 (92.3)	879 (91.95)		4001 (91.6)	1162 (92.52)	
Verbal response	409 (3.24)	83 (3.75)		225 (2.73)	32 (3.35)		184 (4.21)	51 (4.06)	
Pain response	279 (2.21)	56 (2.53)		214 (2.59)	33 (3.45)		65 (1.49)	23 (1.83)	
Unresponsiveness	314 (2.49)	32 (1.45)		196 (2.38)	12 (1.26)		118 (2.7)	20 (1.59)	
Fire stations, n (%)			<.001			<.001			<.001
1	3490 (27.64)	62 (2.79)		3490 (42.3)	62 (6.48)		__[c]	__[c]	
2	4761 (37.70)	895 (40.21)		4761 (57.70)	895 (93.52)		__[c]	__[c]	
3	1277 (10.11)	702 (31.54)		__[c]	__[c]		1277 (29.18)	702 (55.32)	
4	1637 (12.96)	378 (16.98)		__[c]	__[c]		1637 (37.41)	378 (29.79)	
5	1462 (11.58)	189 (8.49)		__[c]	__[c]		1462 (33.41)	189 (14.89)	
ED[d] result, n (%)			.01			.32			.12

XSL•FO
RenderX

Characteristics	Total			Region 1			Region 2		
	Control (n=12,627)	Intervention (n=2226)	*P* value	Control (n=8251)	Intervention (n=957)	*P* value	Control (n=4376)	Intervention (n=1269)	*P* value
Admission	3583 (28.59)	687 (30.92)		2358 (28.74)	287 (30.02)		1225 (28.30)	400 (31.60)	
Discharge	8549 (68.21)	1461 (65.75)		5630 (68.61)	644 (67.36)		2919 (67.44)	817 (64.53)	
Transfer	208 (1.66)	49 (2.21)		92 (1.12)	15 (1.57)		116 (2.68)	34 (2.69)	
Expire	194 (1.55)	25 (1.13)		126 (1.54)	10 (1.05)		68 (1.57)	15 (1.18)	

[a]KTAS: Korean Triage and Acuity Scale. The KTAS was evaluated from 1, resuscitation, to 5, nonemergency.

[b]Mental status was assessed using the AVPU scale.

[c]Not applicable.

[d]ED: emergency department.

The median transport time was significantly longer for the intervention group (10, IQR 7-14 min) with the system activated than for the control group without the system (9, IQR 6-13 min). For a comprehensive analysis of the study results, we presented a table showing the comparison of transport times as a continuous variable (Multimedia Appendix 2). In addition, to determine whether delays were caused by prehospital care providers struggling to adapt to the new system, we analyzed trends in weekly median transport times throughout the entire intervention period and found no delays related to adaptation (Multimedia Appendix 3).

Proportion of Outliers With Longer Transport Times According to the Intervention

Figure 5 shows the difference in the proportion of outliers with longer transport times according to the intervention. In the overall population and region 1, a higher proportion of outliers in transport time occurred in the intervention group. On the contrary, region 2 showed a significant reduction in the proportion of cases with a percentile transport time greater than 75% (26.7%-23.7%, *P*=.05). For patients with fever or respiratory symptoms, there was a statistically significant reduction of more than 75% in the proportion of outlier cases in the group that used the system (36.5%-30.1%, *P*=.01).

XSL•FO
RenderX

Figure 5. The proportion of outliers with longer transport times according to the intervention.

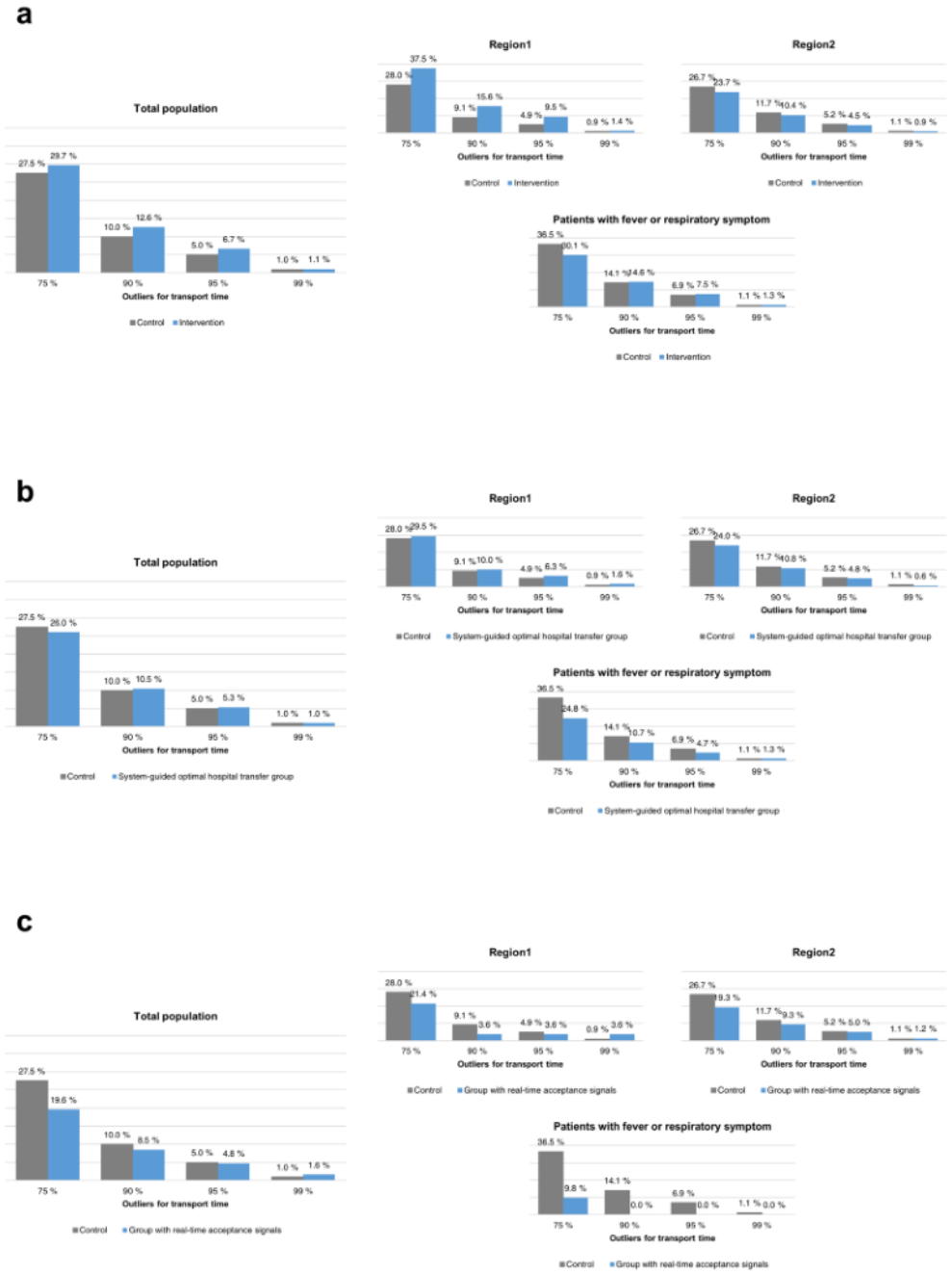

In the system-guided optimal hospital transfer group, there was no significant difference in outlier proportions, compared with the control group; however, for patients with fever or respiratory symptoms, a significant reduction was noted in outliers greater than 75% (36.5%-24.8%, *P*<.001).

For patients who received real-time acceptance signals from the hospital, the proportion of outliers was reduced, compared with the group without the system, and the 75% reduction in the percentage of outliers was statistically significant (27.5%-19.6%, *P*=.01). For patients with fever or respiratory symptoms, the group that transferred after receiving the acceptance signal had no 90% outliers for transport time.

Patient Mortality in the ED

Multimedia Appendix 4 shows the proportion for deaths after arrival at the ED according to the intervention. The mortality rate in the ED was 1.5% in the control group and 1.1% in the intervention group (*P*=.14). In the system-guided optimal hospital transfer group, the proportion for deaths after arrival at the ED was significantly smaller than the control group (1.54%-0.64%, *P*=.01).

A total of 114 prehospital care providers who participated in the study were surveyed about their satisfaction with three AI assistant services. The optimal transfer hospital selection service using AI was found to be the most satisfactory for prehospital care providers (Multimedia Appendix 5).

XSL•FO
RenderX

Discussion

Principal Results

As a result of applying this system, there was no significant change in the overall proportion of delayed patients and patient transport time. The median time required for patient transfer in this study was 9-10 minutes, which is extremely short when compared with that from previous studies in other countries [18,19]. In Korea, as the EMS system is exclusively operated by the NFA with government funds, patients are not obligated to pay for transportation, and public ambulances are relatively accessible compared with other countries [25]. High accessibility to public ambulances was accompanied by the side effect that even patients with low case severity frequently use public ambulances. Likewise, in this study, it was confirmed that mild patients with KTAS 4 and 5 accounted for 44.6% of all the transferred patients. Patients with low severity have a wider choice of hospitals compared with severe emergency patients; therefore, the risk of delay in transfer is low, even without using the CONNECT-AI system. This explains why the effectiveness of CONNECT-AI system was dominated by specific groups, such as patients with fever or respiratory symptoms who had difficulty selecting a transfer hospital, rather than a reduction in the proportion of outliers with long transport times among all the transferred patients. In this subgroup, it was confirmed that the transfer delay was noticeably reduced when the hospital responded to transfer acceptance through the ER kiosk.

In our study, compared with the control group, the rates of system usage by prehospital care providers in the intervention group were 29.0% and 11.6% for regions 1 and 2, respectively, considerably low. As this system was not unified with the existing EMS system, it depended on the prehospital care provider's decision whether or not to use our system. As prehospital care providers were not obligated to use this system, it was used more frequently on patients with fever and respiratory symptoms who had difficulty obtaining transfer permission from the hospital than on patients who had no difficulty in hospital assignment.

According to the results of a survey conducted on users who used this new medical system, the function that prehospital care providers were most satisfied with was the role of the EMS kiosk as a platform to easily obtain information on crowding in the ED, particularly because the EMS kiosk can contact multiple hospitals simultaneously and help to find a transfer hospital more quickly. However, in this study, the number of cases in which the medical staff responded with an acceptable answer to the transfer request by the prehospital care provider was insufficient to demonstrate the effectiveness of this system in reducing transport time. We hypothesize that this situation can be attributed to 2 primary reasons. First, ED medical staff faced challenges in promptly addressing requests from prehospital care providers due to their busy engagement in providing emergency treatment to existing patients within the ED. The second reason is the challenging nature of the ED, which lacks control over patient input and is susceptible to overcrowding, rendering it difficult for medical staff to actively decide on accepting new patients [26].

To provide the best emergency care to more patients with limited medical resources in the region, it is not only important to transfer each patient promptly, but emergency patients must also be properly distributed to each hospital [27-29]. The AI prediction algorithm for selecting the optimal transfer hospital of this system contains a strategy to guide patients, who need resuscitation or emergency procedures, to hospitals with sufficient resources and patients with mild symptoms to small hospitals. Such a resource-saving strategy is expected to be advantageous for more efficient use of limited resources but cannot be an absolute solution in a situation where the demand of patients exceeds the supply of medical resources. To balance the demand of patients and the supply of medical resources in a region, a system that leads to policy decisions on the supply of additional emergency medical resources is needed through real-time monitoring of the overall status of resource use and prediction of impending resource depletion [30,31]. We anticipate that real-time information on patient transport and hospital acceptance collected through our EMS kiosk and ED kiosk will be used as valuable data in these emergency medical resource management systems in the future.

To globally expand our initiative, several challenges must be addressed. First, discussions need to take place regarding existing legal restrictions and concerns about privacy and security when transmitting patient information over networks, as the legal framework varies by country. In addition, legal issues surrounding telemedicine must be resolved in advance [32,33]. Furthermore, many underserved and rural areas may lack access to networks capable of real-time communication, making the establishment of broader network infrastructure a priority over specific technologies.

Comparison With Previous Works

In this study, a prehospital clinical decision support system and a patient information-sharing system during transport were developed. To the best of our knowledge, this AI-enhanced digital network system is the first clinical decision support system that determines the transfer hospital by comprehensively considering a wide range of information, including the clinical characteristics of the patient and the hospital resource information, at the prehospital stage. Hospital resource information such as ED crowding and the availability of emergency interventions is an important factor for hospital medical staff to provide the best care to patients, but it is not easy for prehospital care providers to verify hospital information and consider the same when selecting a transfer hospital. This system included machine learning algorithms that can effectively use vast amount of information. In recent years, AI systems that support decision-making have been widely introduced in the medical field but are rarely found in the EMS domain [34-36]. This is because data input is required to operate the AI algorithm, but real-time data input is almost impossible from the site of rescuing emergency patients [37]. Accordingly, voice- and image recognition technologies were applied to use information from the scene in real-time. This clinical information of the patient being transported was transmitted to the ER kiosk so that the ED medical staff could check the condition of the patient before arrival and prepare for emergency treatment.

XSL•FO
RenderX

Limitations

This study has a few limitations. First, the number of patients who used this system was extremely small to sufficiently confirm the effect of the system. The cumbersomeness of having to use additional equipment such as an EMS kiosk and a microphone in addition to the existing tablet made prehospital care providers reluctant to use it. The proportion of patients with fever or respiratory symptom was higher in the group where the system was used compared with the control group. This could be explained by the prehospital providers' assessment that the potential benefits of using the system outweighed the inconvenience, especially when the transport was more challenging. To address the resulting selection bias, we separated the group of patients with fever or respiratory symptom and confirmed that the system's effectiveness was more pronounced in cases where transport difficulty was greater. Second, as this study did not cover patients transferred to hospitals outside the study area or patients transferred from other areas to hospitals in the study area, a comprehensive analysis of patient transfers within the area was not possible. The effectiveness of this system may be affected by the availability of medical resources across the region and therefore needs to be evaluated at the regional level. Including these patients in the study likely led to longer transport times for the outliers, as the transport distances were greater compared with transfers within the local area. Finally, in Korea, prehospital care providers receive permission from hospital medical staff to transport patients, but they can be transported without permission. Regulations on emergency patient transport may vary with the country; therefore, this aspect needs to be considered when understanding our system and interpreting the study results. Given that the intervention effect was more pronounced in the group with real-time acceptance signals in our study, it is likely that the system's effectiveness was underestimated in this study setting compared with regions where transport authorization is mandatory.

Conclusions

This digital platform offers a novel approach to enhancing emergency patient transport by leveraging AI, real-time information sharing, and decision support. While the system demonstrated improvements for certain patient groups facing transfer challenges, further research and adaptations are necessary to fully realize its benefits in diverse health care contexts.

Acknowledgments

This study was funded by the AI (artificial intelligence) Ambulance project for digital health care of precision medicine software development through the National Fund of the Korean government (Ministry of Science and Information and Communications Technology). The funding body had no rule in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video clip introducing CONNECT AI (CONnected Network for EMS Comprehensive Technical-Support using Artificial Intelligence) system.
[MP4 File (MP4 Video), 28104 KB - jmir_v27i1e58177_app1.mp4]

Multimedia Appendix 2
Supplementary Table 1. Comparison of the transport time between the two groups.
[DOCX File , 17 KB - jmir_v27i1e58177_app2.docx]

Multimedia Appendix 3
Supplementary Figure 1. The median transport time on a weekly basis during the intervention period. No delays related to the adaptation of pre-hospital care providers were observed in the early phase of the intervention period.
[DOCX File , 71 KB - jmir_v27i1e58177_app3.docx]

Multimedia Appendix 4
Patient mortality in the emergency department.
[PNG File , 24 KB - jmir_v27i1e58177_app4.png]

Multimedia Appendix 5
CONNECT AI system satisfaction survey.

XSL•FO
RenderX

[DOCX File , 613 KB - jmir_v27i1e58177_app5.docx]

References

1. Blackwell TH. Prehospital Care. Emergency Medicine Clinics of North America 1993;11(1):1-14. [doi: 10.1016/s0733-8627(20)30653-2]
2. Brice JH, Valenzuela T, Ornato JP, Swor RA, Overton J, Pirrallo RG, Turtle Creek Conference II. Optimal prehospital cardiovascular care. Prehosp Emerg Care 2001;5(1):65-72. [doi: 10.1080/10903120190940362] [Medline: 11194073]
3. George L, Ramamoorthy L, Satheesh S, Saya RP, Subrahmanyam DKS. Prehospital delay and time to reperfusion therapy in ST elevation myocardial infarction. J Emerg Trauma Shock 2017;10(2):64-69 [FREE Full text] [doi: 10.4103/0974-2700.201580] [Medline: 28367010]
4. Chen CH, Shin SD, Sun JT, Jamaluddin SF, Tanaka H, Song KJ, et al. Association between prehospital time and outcome of trauma patients in 4 Asian countries: a cross-national, multicenter cohort study. PLoS Med 2020;17(10):e1003360 [FREE Full text] [doi: 10.1371/journal.pmed.1003360] [Medline: 33022018]
5. Evenson KR, Foraker RE, Morris DL, Rosamond WD. A comprehensive review of prehospital and in-hospital delay times in acute stroke care. Int J Stroke 2009;4(3):187-199 [FREE Full text] [doi: 10.1111/j.1747-4949.2009.00276.x] [Medline: 19659821]
6. Asplin BR. Does ambulance diversion matter? Ann Emerg Med 2003;41(4):477-480. [doi: 10.1067/mem.2003.112] [Medline: 12658246]
7. Geiderman JM, Marco CA, Moskop JC, Adams J, Derse AR. Ethics of ambulance diversion. Am J Emerg Med 2015;33(6):822-827. [doi: 10.1016/j.ajem.2014.12.002] [Medline: 25616586]
8. Cooney DR, Millin MG, Carter A, Lawner BJ, Nable JV, Wallus HJ. Ambulance diversion and emergency department offload delay: resource document for the national association of EMS physicians position statement. Prehosp Emerg Care 2011;15(4):555-561. [doi: 10.3109/10903127.2011.608871] [Medline: 21870947]
9. Olshaker JS, Rathlev NK. Emergency department overcrowding and ambulance diversion: the impact and potential solutions of extended boarding of admitted patients in the emergency department. J Emerg Med 2006;30(3):351-356. [doi: 10.1016/j.jemermed.2005.05.023] [Medline: 16677993]
10. Steflovich M. Management of emergency department diversion during the COVID-19 pandemic and disaster periods. J Emerg Manag 2023;21(7):203-212. [doi: 10.5055/jem.0723] [Medline: 37154455]
11. Harrison NE, Ehrman RR, Curtin A, Gorelick D, Hill AB, Brennan E, et al. Factors associated with voluntary refusal of emergency medical system transport for emergency care in detroit during the early phase of the COVID-19 pandemic. JAMA Netw Open 2021;4(8):e2120728 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.20728] [Medline: 34415317]
12. Golling M, Behringer W, Schwarzkopf D. Assessing the quality of patient handovers between ambulance services and emergency department - development and validation of the emergency department human factors in handover tool. BMC Emerg Med 2022;22(1):10 [FREE Full text] [doi: 10.1186/s12873-022-00567-y] [Medline: 35045828]
13. Dawson S, King L, Grantham H. Review article: improving the hospital clinical handover between paramedics and emergency department staff in the deteriorating patient. Emerg Med Australas 2013;25(5):393-405. [doi: 10.1111/1742-6723.12120] [Medline: 24099367]
14. Souza MMD, Xavier AC, Araújo CAR, Pereira ER, Duarte SDCM, Valladares Broca P. Communication between pre-hospital and intra-hospital emergency medical services: literature review. Rev Bras Enferm 2020;73(suppl 6):e20190817 [FREE Full text] [doi: 10.1590/0034-7167-2019-0817] [Medline: 33338152]
15. Bae HJ, Kim DH, Yoo NT, Choi JH, Huh JT, Cha JK, et al. Prehospital notification from the emergency medical service reduces the transfer and intra-hospital processing times for acute stroke patients. J Clin Neurol 2010;6(3):138-142. [doi: 10.3988/jcn.2010.6.3.138] [Medline: 20944814]
16. Chen C, Kan T, Li S, Qiu C, Gui L. Use and implementation of standard operating procedures and checklists in prehospital emergency medicine: a literature review. Am J Emerg Med 2016;34(12):2432-2439. [doi: 10.1016/j.ajem.2016.09.057] [Medline: 27742522]
17. Ishikawa K, Yanagawa Y, Ota S, Muramatsu K, Nagasawa H, Jitsuiki K, et al. Preliminary study of prehospital use of smart glasses. Acute Med Surg 2022;9(1):e807 [FREE Full text] [doi: 10.1002/ams2.807] [Medline: 36330311]
18. Fukaguchi K, Goto T, Yamamoto T, Yamagami H. Experimental implementation of NSER mobile app for efficient real-time sharing of prehospital patient information with emergency departments: interrupted time-series analysis. JMIR Form Res 2022;6(7):e37301 [FREE Full text] [doi: 10.2196/37301] [Medline: 35793142]
19. Katayama Y, Kitamura T, Kiyohara K, Iwami T, Kawamura T, Izawa J, et al. Improvements in patient acceptance by hospitals following the introduction of a smartphone app for the emergency medical service system: a population-based before-and-after observational study in Osaka City, Japan. JMIR Mhealth Uhealth 2017;5(9):e134 [FREE Full text] [doi: 10.2196/mhealth.8296] [Medline: 28893725]
20. Haverinen J, Kangas M, Raatiniemi L, Martikainen M, Reponen J. How to improve communication using technology in emergency medical services? A case study from Finland. FinJeHeW 2018;10(4). [doi: 10.23996/fjhw.74143]
21. Dalgard S, Vedum J, Rake EL, Eide AW, Liverud AE, Skjetne JH, et al. Design and evaluation of an electronic triage system for prehospital monitoring of patients. IJEM 2017;13(4):328. [doi: 10.1504/ijem.2017.087224]

XSL•FO
RenderX

22. Kim JH, Kim B, Kim MJ, Hyun H, Kim HC, Chang HJ. Prediction of inappropriate pre-hospital transfer of patients with suspected cardiovascular emergency diseases using machine learning: a retrospective observational study. BMC Med Inform Decis Mak 2023;23(1):56 [FREE Full text] [doi: 10.1186/s12911-023-02149-9] [Medline: 37024872]
23. Choi A, Kim M, Sung J, Kim S, Chang HJ, Lee J, et al. Development of prediction models for acute myocardial infarction at prehospital stage with machine learning based on a nationwide database. J Cardiovasc Dev Dis. 2022 2021;9(12). [doi: 10.2196/preprints.27405]
24. Lee JH, Park YS, Park IC, Lee HS, Kim JH, Park JM, et al. Over-triage occurs when considering the patient's pain in Korean Triage and Acuity Scale (KTAS). PLoS One 2019;14(5):e0216519 [FREE Full text] [doi: 10.1371/journal.pone.0216519] [Medline: 31071132]
25. Lee SY, Song KJ, Shin SD, Hong KJ, Kim TH. Comparison of the effects of audio-instructed and video-instructed dispatcher-assisted cardiopulmonary resuscitation on resuscitation outcomes after out-of-hospital cardiac arrest. Resuscitation 2020 Feb 01;147:12-20. [doi: 10.1016/j.resuscitation.2019.12.004] [Medline: 31843537]
26. Hoot NR, Leblanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, et al. Forecasting emergency department crowding: a prospective, real-time evaluation. J Am Med Inform Assoc 2009;16(3):338-345 [FREE Full text] [doi: 10.1197/jamia.M2772] [Medline: 19261948]
27. Muttalib F, González-Dambrauskas S, Lee J, Steere M, Agulnik A, Murthy S, PALISI Global Health Subgroup of the Pediatric Acute Lung InjurySepsis Investigators (PALISI). Pediatric Emergency and Critical Care Resources and Infrastructure in Resource-Limited Settings: A Multicountry Survey. Crit Care Med 2021 Apr 01;49(4):671-681. [doi: 10.1097/CCM.0000000000004769] [Medline: 33337665]
28. Headrick RW, Morgan GW. Resource allocation in multifacility emergency medical service systems. J Med Syst 1988 Jun;12(3):121-128. [doi: 10.1007/BF00996634] [Medline: 3171441]
29. Hegenberg K, Trentzsch H, Prückner S. Differences between cases admitted to hospital and discharged from the emergency department after emergency medical services transport. BMJ Open 2019 Sep 27;9(9):e030636 [FREE Full text] [doi: 10.1136/bmjopen-2019-030636] [Medline: 31562153]
30. Du L. Medical Emergency Resource Allocation Model in Large-Scale Emergencies Based on Artificial Intelligence: Algorithm Development. JMIR Med Inform 2020 Jun 25;8(6):e19202 [FREE Full text] [doi: 10.2196/19202] [Medline: 32584262]
31. Dong E, Liu S, Chen M, Wang H, Chen L, Xu T, et al. Differences in regional distribution and inequality in health-resource allocation at hospital and primary health centre levels: a longitudinal study in Shanghai, China. BMJ Open 2020 Jul 19;10(7):e035635 [FREE Full text] [doi: 10.1136/bmjopen-2019-035635] [Medline: 32690509]
32. Liyanage M, Salo J, Braeken A, Kumar T, Seneviratne S, Ylianttila M. 5G privacy: scenarios and solutions. : IEEE; 2018 Presented at: 2018 IEEE 5G World Forum (5GWF); July 09-11, 2018; Silicon Valley, CA, USA. [doi: 10.1109/5gwf.2018.8516981]
33. Kim H, Kim SW, Park E, Kim JH, Chang H. The role of fifth-generation mobile technology in prehospital emergency care: An opportunity to support paramedics. Health Policy and Technology 2020 Mar;9(1):109-114. [doi: 10.1016/j.hlpt.2020.01.002]
34. Amann J, Vayena E, Ormond KE, Frey D, Madai VI, Blasimme A. Expectations and attitudes towards medical artificial intelligence: A qualitative study in the field of stroke. PLoS One 2023;18(1):e0279088 [FREE Full text] [doi: 10.1371/journal.pone.0279088] [Medline: 36630325]
35. Koulaouzidis G, Jadczyk T, Iakovidis DK, Koulaouzidis A, Bisnaire M, Charisopoulou D. Artificial Intelligence in Cardiology-A Narrative Review of Current Status. J Clin Med 2022 Jul 05;11(13):e0279088 [FREE Full text] [doi: 10.3390/jcm11133910] [Medline: 35807195]
36. Machoy ME, Szyszka-Sommerfeld L, Vegh A, Gedrange T, Woźniak K. The ways of using machine learning in dentistry. Adv Clin Exp Med 2020 Mar;29(3):375-384 [FREE Full text] [doi: 10.17219/acem/115083] [Medline: 32207586]
37. Kim JH, Choi A, Kim MJ, Hyun H, Kim S, Chang H. Development of a machine-learning algorithm to predict in-hospital cardiac arrest for emergency department patients using a nationwide database. Sci Rep 2022 Dec 16;12(1):21797 [FREE Full text] [doi: 10.1038/s41598-022-26167-1] [Medline: 36526686]

Abbreviations

AI: artificial intelligence
CONNECT-AI: CONnected Network for EMS Comprehensive Technical-Support using Artificial Intelligence
ED: emergency department
EMS: emergency medical services
IoT: Internet of Things
KTAS: Korean Triage and Acuity Scale
NEDIS: National Emergency Department Information System
NEMC: National Emergency Medical Center
NFA: National Fire Agency

XSL•FO
RenderX

PHRS: prehospital run sheets

Edited by A Coristine; submitted 11.03.24; peer-reviewed by T Aslanidis, B Hoseini, Y Ono; comments to author 15.08.24; revised version received 24.10.24; accepted 31.10.24; published 23.01.25.

Please cite as:
Kim JH, Kim MJ, Kim HC, Kim HY, Sung JM, Chang HJ
A Novel Artificial Intelligence–Enhanced Digital Network for Prehospital Emergency Support: Community Intervention Study
J Med Internet Res 2025;27:e58177
URL: https://www.jmir.org/2025/1/e58177
doi:10.2196/58177
PMID:39847421

©Ji Hoon Kim, Min Joung Kim, Hyeon Chang Kim, Ha Yan Kim, Ji Min Sung, Hyuk-Jae Chang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patient Experiences and Perspectives When MyChart is Introduced in a Large Community Hospital: Mixed Methods Study

Shelley Vanderhout[1,2], PhD; Shipra Taneja[1], MSc; Kamini Kalia[3], MScN, RN; Walter P Wodchis[1,2], PhD; Terence Tang[1,3], MSc, MD

[1]Institute for Better Health, Trillium Health Partners, Mississauga, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Trillium Health Partners, Mississauga, ON, Canada

Corresponding Author:
Shelley Vanderhout, PhD
Institute for Better Health
Trillium Health Partners
2085 Hurontario Street
Mississauga, ON, L5A 4G1
Canada
Phone: 1 6473307192
Email: shelley.vanderhout@thp.ca

Abstract

Background: Patient portals, or secure websites linked to electronic medical records, have emerged as tools to provide patients with timely access to their health information. To support the potential benefits of patient portals such as improved engagement in health care, it is essential to understand how patients and caregivers experience these portals.

Objective: This study aimed to explore patient and caregiver experiences, facilitators, and barriers to accessing and using a patient portal called MyChart during the initial stages of its implementation.

Methods: We applied explanatory sequential mixed methods to conduct a web-based questionnaire and semistructured interviews with MyChart users and nonusers at a large community hospital in Ontario, Canada. Among users, we explored user satisfaction with MyChart, its impact on care, and areas for improvement. For nonusers, we explored barriers to MyChart access and willingness to use it in the future. Descriptive statistics and thematic analysis were used for data analysis.

Results: A total of 5651 patients and caregivers completed the web-based questionnaire and 18 (12 users and 6 nonusers) participated in interviews. MyChart users primarily learned about the portal through email (n=1288, 39%), after-visit summaries (n=953, 29%), and hospital staff (n=408, 12%). Nonusers cited lack of awareness (n=1291, 59%) and registration difficulties (n=707, 32%) as some barriers to activation and adoption, but the majority would consider activating and using MyChart if they could learn more about it (n=1126, 54%). Users valued MyChart for preparing for health care encounters but expressed dissatisfaction with limited features and access to medical history and test results, whereas nonusers tended to be unsure about the benefits of using MyChart, especially if they were infrequent health care users.

Conclusions: Patient portals offer benefits, but barriers to access and limited functionality can hinder widespread use. To enhance the adoption and potential benefits of patient portals, targeted outreach and comprehensive access to health information are essential to promote positive and seamlessly integrated health care experiences.

(J Med Internet Res 2025;27:e66353) doi:10.2196/66353

KEYWORDS

patient portal; implementation; mixed methods; patient experiences; patient perspectives; learning health systems; LHS; mixed methods study; electronic medical records; health information; caregiver experience; Canada; user satisfaction; descriptive statistics; thematic analysis

XSL•FO
RenderX

Introduction

Patient portals are secure websites or mobile apps tethered to electronic health records that allow patients to view medical notes and test results, visit summaries, manage appointment schedules, and update personal information [1]. Designed to empower patients, they can promote patient engagement in care, increase self-management [2], improve patient-provider communication, and foster shared decision-making between patients and health care providers [3,4]. Portal use has led to improved communication and trust between patients and providers [5,6], adherence to care plans [7], and satisfaction with care [8]. Patient portals also have the potential to reduce anxiety about test results, avoid unnecessary visits, and improve health understanding [5,9,10]. However, patient portal uptake can be limited by low digital and health literacy [11,12], hesitancy about privacy and security of personal health information [13], and poor alignment of patient expectations for portal functionalities and features actually offered [12].

As health systems transition to electronic health records, many offer access to patient portals and have opportunities to learn from introducing these tools. Exploring patient experiences with accessing and using portals, barriers to activation and use, and perceived impacts on daily life and care is crucial to inform how portals can support high-quality care, identify barriers and enablers to their use, develop tailored supports for specific populations (eg, older patients), and highlight areas for improvement [14]. Particularly, when patient populations are diverse, health systems need to understand how patient portals may inadvertently exacerbate existing equity issues related to accessibility and technological literacy [15].

This study was an evaluation to generate transferable knowledge about the implementation of a patient portal called MyChart within the first 6 months of its launch at a large community hospital in Ontario, Canada. Specifically, we sought to answer the following questions:

- Among patients who activated MyChart, what were their initial experiences, perceived impacts, and suggestions for improvement?
- Among patients who did not activate MyChart, what were the barriers to doing so, and what were their preferences for support to access and use MyChart?
- How can health systems offer portals that meet patients' needs and expectations?

Methods

Setting

This mixed methods evaluation took place at Trillium Health Partners (THP), a large community hospital comprised of 3 sites in Mississauga, Canada. Mississauga is a very diverse community in Canada with 51% of residents identifying as newcomers and 62% as visible minorities. THP operates 1457 inpatient beds and has more than 1.7 million annual patient visits (including 225,723 emergency and urgent care visits and 776,308 outpatient visits), with more than 11,000 staff and physicians [16].

THP implemented the Epic electronic health record in October 2020, and deployed Epic's patient portal, MyChart, in September 2023 [17]. At THP, patients can activate MyChart on a desktop computer or through the MyChart mobile app through links to create an account included in reminder emails about upcoming visits, unique codes provided on after-visit or discharge summaries, or in-person during a care encounter. THP offers technical support for MyChart via telephone and email, tip sheets, and instructional videos on their website about activating accounts and using the available features. MyChart is advertised through posters in hospital waiting rooms and corridors, social media, the THP website, and word of mouth (eg, clinical clerks during visit check-in and health care provider champions who encouraged their peers to socialize MyChart with patients). At THP, MyChart functionalities were introduced in a staged manner with new features launching every quarter; the first stage included viewing after-visit and discharge summaries, outpatient test results, and viewing or updating personal information, medications, allergies, and vaccinations. The second stage added viewing appointment schedules, electronically checking in for or canceling appointments, launching video visits, and requesting additional information. Messaging between patients and providers was not offered as part of MyChart's launch at THP. This approach was taken to reflect organizational capacity and early engagement with health care providers that suggested limited readiness for comprehensive features to be introduced all at once, especially messaging.

Approach

We applied an explanatory sequential mixed methods evaluation approach to understand patient and caregiver experiences, both among those who had activated MyChart and those who had not [18]. This mixed methods evaluation included: (1) a web-based questionnaire of MyChart users and nonusers with different questions for each group and (2) semistructured interviews with users and nonusers to add richness and depth to the questionnaire data. The RE-AIM Implementation Framework [19] and its extension [20] to enhance sustainability and promote equity guided our approach and informed questionnaire and interview content. This framework allowed us to focus on key areas such as reach (identifying users and barriers), effectiveness (assessing patient-perceived impacts and potential negative outcomes), adoption (examining use patterns and disparities), implementation (identifying effective facilitators), and maintenance (evaluating long-term impact).

Data Collection and Analysis

We invited all THP patients to participate in a web-based questionnaire (Multimedia Appendix 1), hosted on Qualtrics for 4 weeks during March 2024 [21]. We explored users' satisfaction with MyChart, perceived impacts on care, and suggestions for improvement. For nonusers, we investigated barriers and potential enablers to activating MyChart, and alternative ways they would like to be involved in their care. Participants had the option to provide their age, gender identity, ethnicity, and postal code to help characterize user and nonuser populations. The questionnaire included both closed- and open-ended questions and was created based on available

XSL•FO
RenderX

literature [9,13,22,23] and peer health systems' evaluations. It was revised based on feedback from a MyChart patient advisory council at THP.

All patients with an email address recorded and who had provided consent to receive nonclinical communication were invited by email to participate; caregivers were able to participate if their email was linked to a patient's record. Two reminder emails were sent 3 and 5 days following the initial invitation. We advertised the questionnaire on the hospital's website, social media pages, and posters with QR codes in waiting areas. MyChart users saw a banner with a link to the questionnaire upon signing into the platform. Acknowledging that not all patients would be reached virtually, hospital volunteers also recruited patients to complete the questionnaire on tablets from clinic waiting rooms and offered navigational support when necessary. The questionnaire took approximately 5 minutes to complete, and participants were not offered incentives. At the end of the questionnaire, participants could indicate if they would like to participate in a follow-up interview.

Descriptive statistics were used to summarize responses to closed-ended questions and participant demographics, and thematic analysis was used to identify key concepts among answers to the open-ended questions. The Ontario Marginalization Index (ON-Marg) was used to classify postal codes according to socioeconomic status. ON-Marg assigns quintiles to postal codes based on relative material deprivation (representing education, low income, unemployment, lone-parent families, and dwellings in need of major repair) and ethnic concentration (representing immigrants within the past 5 years and visible minorities) [8]. Participants were excluded if they indicated that did not receive care at THP in the questionnaire. Statistical analysis was completed on SAS (version 9.4; SAS Institute).

We used a combination of purposive sampling and maximum variation to invite 18 questionnaire participants to individual semistructured interviews while seeking variability in age, ethnicity, gender, and overall sentiment of questionnaire response [24]. Separate interview guides were used for MyChart users and nonusers (Multimedia Appendix 1). All interviews were conducted by one team member (ST) via Zoom (Zoom Communications Inc), lasted approximately 30 minutes each, and were recorded and transcribed verbatim. Transcripts were reviewed by the interviewer to verify completeness and accuracy and were not reviewed by the participant. Two members of the research team (ST and SV) conducted thematic analysis on NVIVO (version 12; Lumivero); following data familiarization, they conducted an initial coding phase, combining a deductive and inductive approach. Deductive coding was based on the RE-AIM framework [19], which guided the identification of text related to reach, effectiveness, adoption, implementation, and maintenance indicators, as well as the acceptability and accessibility of MyChart. Inductive coding was used to explore patients' experiences and perspectives. Codes were used to generate themes, each built around a central organizing concept, which were then refined iteratively to produce a final set of themes.

Ethical Considerations

This study was reviewed by the THP Research Ethics Board and classified as quality improvement. This study was classified as quality improvement as it aimed to enhance existing processes within our health care system. All participants provided informed consent prior to participating in the questionnaire and interviews. They were informed that the study results would be deidentified and published in peer-reviewed journals and potentially shared with the broader public. Interview participants were given a CAD $30 (US $21) honorarium for their time.

Results

Overview

In total, 5651 patients and caregivers completed the questionnaire. Of those 5651 participants, 3336 (59%) participants were MyChart users and 2315 (41%) participants were nonusers, and most of their characteristics were similar across groups except for age; younger individuals were less represented among MyChart users (Table 1). Of the MyChart users, 3098 (93%) participants used MyChart for their own care, and 417 (13%) participants used it as a parent or caregiver. A total of 18 patients and caregivers participated in interviews; twelve (67%) were MyChart users and 6 (33%) were nonusers. Three MyChart users interviewed were parents or caregivers, and 2 used MyChart as users and as family members.

Most MyChart users learned about the patient portal through various channels, including hospital emails (1288/3336, 39%), after-visit summaries (953/3336, 29%), their care team (935/3336, 28%), hospital staff (408/3336, 12%), informational posters (340/3336, 10%), knowledge of MyChart at other health care systems (163/3336, 5%), through recommendations from family or friends (125/3336, 4%), or social media (60/3336, 2%). While 775 (34%) of nonusers were aware of MyChart and expressed interest in learning more, the most reported barrier was a lack of awareness that it was offered (1291/2315, 59%). Other commonly reported barriers included uncertainty about how to sign up for an account (707/2315, 32%) or use the portal (516/2315, 23%). A smaller proportion of nonusers reported low digital literacy (322/2315, 15%) or concerns about privacy and security (267/2315, 12%) as barriers to activating MyChart. Nonusers indicated that they would consider activating and using MyChart if they had the opportunity to learn more about its features (1126/2315, 54%), and how to create an account (968/2315, 47%). Web-based videos were the most preferable method for learning (943/2315, 62%).

XSL•FO
RenderX

Table 1. Questionnaire and interview participant characteristics.

Characteristics[a]	Users (n=3336), n (%)	Nonusers (n=2315), n (%)	Interviewed users (n=12), n (%)	Interviewed nonusers (n=6), n (%)
Age (in years)				
12-20	1 (0.4)	18 (0.9)	0 (0)	0 (0)
21-40	262 (9)	282 (14)	1 (8)	1 (17)
41-60	964 (33)	625 (32)	6 (50)	2 (33)
61-80	1653 (57)	943 (48)	5 (42)	2 (33)
≥81	119 (4)	114 (6)	0 (0)	0 (0)
Self-identified gender				
Man	1128 (36)	743 (36)	3 (25)	0 (0)
Woman	1951 (62)	1281 (62)	9 (75)	6 (100)
Prefer not to answer	68 (2)	44 (2)	0 (0)	0 (0)
Other	9 (0.3)	9 (0.4)	0 (0)	0 (0)
Cultural and racial background[b]				
White	1965 (63)	1245 (61)	6 (50)	2 (33)
South Asian	375 (12)	272 (14)	3 (25)	2 (33)
East Asian	169 (5)	83 (4)	1 (8)	1 (17)
Southeast Asian	164 (5)	85 (4)	0 (0)	0 (0)
Black	132 (4)	116 (6)	1 (8)	1 (17)
Middle Eastern	86 (3)	69 (3)	0 (0)	0 (0)
Latin American	68 (2)	57 (3)	0 (0)	0 (0)
First Nation	6 (0.2)	6 (0.3)	0 (0)	0 (0)
Métis	4 (0.1)	6 (0.3)	0 (0)	0 (0)
Indigenous/Aboriginal	2 (0.1)	6 (0.3)	0 (0)	0 (0)
Inuit	0 (0)	1 (0.05)	0 (0)	0 (0)
Prefer not to answer	123 (4)	107 (5)	0 (0)	0 (0)
Other	59 (1.9)	32 (2)	0 (0)	0 (0)
Neighborhood material resources, quintile (%)				
1=lowest neighborhood material resources	700 (24)	457 (24)	3 (25)	0 (0)
2	839 (29)	510 (27)	4 (33)	2 (33)
3	731 (25)	448 (24)	4 (33)	2 (33)
4	427 (15)	301 (16)	1 (8)	1 (17)
5=highest neighborhood material resources	224 (8)	154 (8)	0 (0)	1 (17)
Unmatched to ON-Marg[c]	18 (0.6)	24 (1.3)		
Racialized and newcomer populations, quintile (%)[d]				
1=lowest neighborhood proportion of racialized and newcomer populations	80 (3)	41 (2.2)	1 (8)	
2	223 (8)	186 (10)	0 (0)	1 (17)
3	524 (18)	362 (19)	2 (17)	1 (17)
4	858 (29)	520 (27)	7 (58)	1 (17)
5= highest neighborhood proportion of racialized and newcomer populations	1236 (42)	761 (40)	2 (17)	3 (50)

XSL•FO
RenderX

Characteristics[a]	Users (n=3336), n (%)	Nonusers (n=2315), n (%)	Interviewed users (n=12), n (%)	Interviewed nonusers (n=6), n (%)
Unmatched to ON-Marg	18 (0.6)	24 (1.3)		

[a]Characteristics were optional questions.

[b]Participants could select multiple options.

[c]ON-Marg: Ontario Marginalization Index.

[d]Data only available for postal codes that could be matched with ON-Marg.

Perspectives of MyChart users about the portal are shown in Figure 1. Most MyChart users who participated in the questionnaire agreed or somewhat agreed that the portal was easy to use independently (2907/3336, 88%), they could understand the information available in MyChart (2897/3336, 88%), and they felt more informed about their care with MyChart use (2744/3336, 85%). Similarly, 79% (2558/3336) of users agreed or somewhat agreed that MyChart helped them prepare for health care encounters. When asked whether accessing their information caused worry, there were varying responses; over half of the participants (1804/3336, 57%) disagreed or somewhat disagreed that MyChart use caused any worry while 663 (21%) participants agreed or somewhat agreed that it did cause worry. A total of 80% (2605/3336) of users agreed or somewhat agreed that their personal health information was secure and private in MyChart.

Thematic analysis of responses to open-ended questionnaire items and interview questions produced 2 main themes, which are summarized below.

Figure 1. User perspectives about using MyChart. THP: Trillium Health Partners.

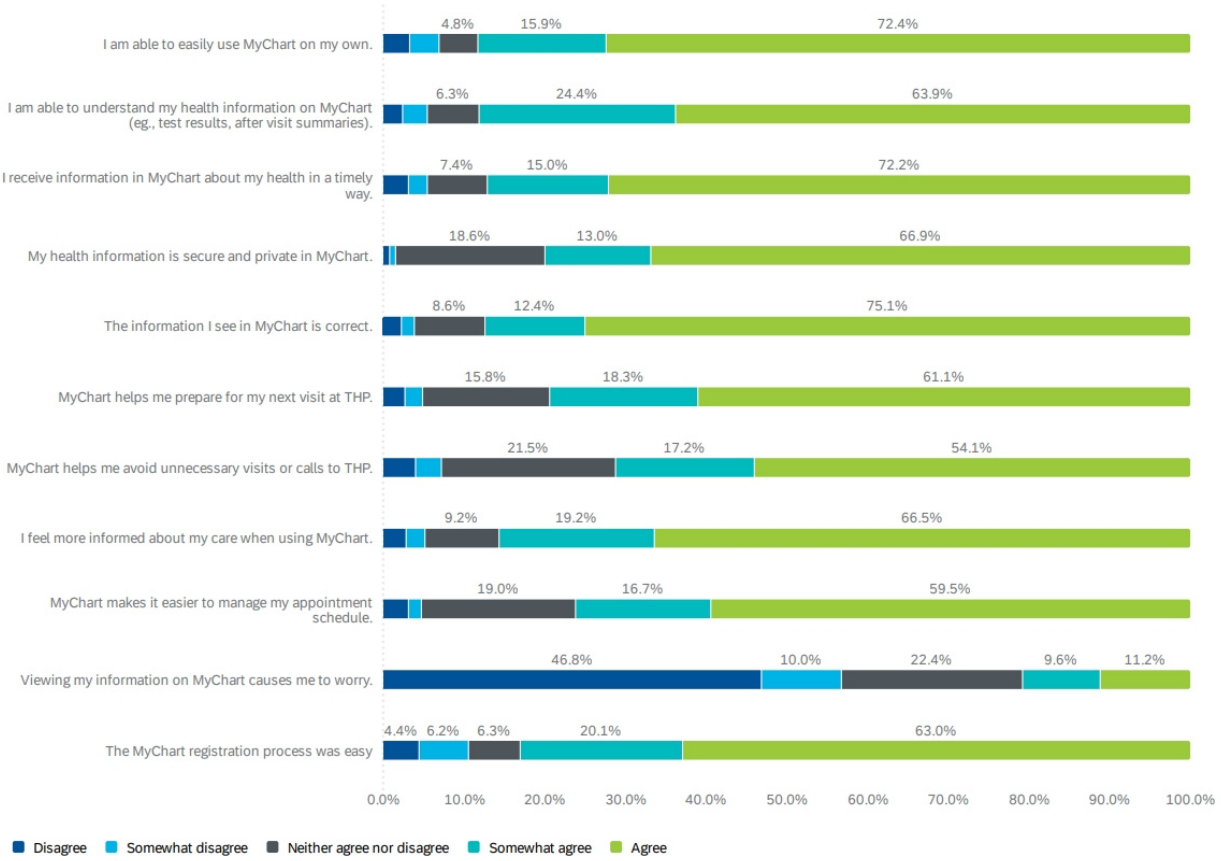

Access to Information

Patients and family members using MyChart praised the platform for its user-friendly interface and convenient access to health information. They appreciated the right to review their own health records and found it empowering to be informed about their health care. A few patients described the feature of revisiting past notes from their care team to be particularly helpful with information recall.

And so what I have found is MyChart has been very helpful for me in providing one spot for me to know, what are my upcoming appointments at this hospital, what test results are there that I can refer back to if another doctor asks me did I do this or what was this or how is this looking. And at least giving me some insight and awareness into what's happening and not having to wait for a doctor's appointment to get answers to questions I have. [User 2]

XSL•FO
RenderX

One nonuser who had an interest in MyChart described how beneficial it could be for caregivers who could not attend appointments, as it offered a window into the discussion and decisions made, fostering a better understanding of their loved one's care.

> *It'd be interesting to see if you can integrate notes or directions on what to do when you get home and that kind of thing so that when you have the elderly going in ... he has his phone. My in-law is in his late 60's. He has an iPhone. To have that integrated for him so when me or my husband go back and we look at the notes we're like, OK this is what happened when he went. Then we can go and say ... I don't have to take time off work, we can understand the notes, translate it for him and then go back and go, OK, this is what you need to do now. Or this is what you have to look out for versus having to take the time off work, go with him to literally be there. That would be interesting, I don't know if that's a feature or a future feature.* [Nonuser 4]

While patients and caregivers appreciated access to MyChart, they perceived the available features and functionalities at THP to be very limited, especially relative to other health systems. They desired access to a wider range of information, including consultation notes, medical images, emergency department results, and a more comprehensive medical history. After-visit summaries were available on MyChart, and while accessible, were seen as lacking detail (since they did not include health care providers' chart notes from the visit) and were often outdated, hindering patients' understanding of their care plan.

> *One thing that would be really nice to have is access to the doctor's notes ... like, the notes from your appointment. I tend to try to take my own notes when I'm at clinical appointments but sometimes it's hard because I'm trying to be present in the appointment and listen and try to understand in the moment. Sometimes my recall isn't great and my note-taking isn't great. So having access to a doctor's summary of what it is that they've talked about and the direction and guidance that they've given would be helpful to be able to go back to and refer to, and even just to update my own notes. So, that's another thing because right now that is not something that I see or that's accessible in terms of their note. But I know it gets documented and shared with other providers in my circle of care but I don't get to see that. And so sometimes it's hard. It's ... like, that would be great. It would be great if that was accessible.* [User 2]

Beyond limited information, patients and caregivers identified issues with MyChart's lack of content and functionality. They envisioned MyChart bridging the gap between care providers but felt it operated in isolation, hindering its full potential.

> *There's no connectivity. The whole thing is to streamline the process and to make sure that the patient has access to their records ... So, if that's not being sent to the technology, to the app, then I'm not getting the information. So, what's the use of spending all this money on technology that is not being used to its fullest?* [User 4]

Among current users, there was a desire for additional features such as chat or messaging with providers (n=476), access to more information such as imaging, historical information and notes (n=550), sending records or providing access to family members or proxies (n=169), linking to other hospitals and care providers (n=271), and supporting information for interpreting results (n=111). Some of these features were in fact available to MyChart users, such as adding proxy users and high-quality resources to help understand medical information, but this finding highlighted low awareness of MyChart's full range of functionalities.

Access to medical records and test results presented a double-edged sword for both users and nonusers. Some users saw it as a source of comfort, reducing anxiety by eliminating the stressful waiting periods and uncertainty before appointments; viewing results ahead of appointments allowed them time to process information and prepare questions for follow-up visits.

> *For example, my family doctor suspects there's something wrong. Sends for a biopsy. The biopsy results are back. Good or bad. Most of the time let's say it's good. But you have that two weeks of stress that you didn't really need. And let's say it's negative. It's still better to know that you have cancer and how to deal with it because you've got time to process. So, by the time you go to the doctor you're prepared to ask the questions that need to be asked. So, it's a more productive meeting between the patient and the doctor and it alleviates the stress level.* [User 4]

However, nonusers expressed that immediate access to sensitive test results before appointments would increase their anxiety. While some patients and families may have medical knowledge to interpret the notes, nonusers described it as a "folder of anxiety" as it does not contain context and next steps provided by a health care provider.

> *So if we could see the results and know we were talking to our doctor either on phone or in person, within 24 hours, we would sign up for MyChart, I guess is the most simple ... When you're talking hypertension, cholesterol, thyroid. I guess, like, regular run-of-the-mill stuff. Then fine. Like, I know what's going to happen. You're going to adjust my dose. You're going to add another medication. You're going to, like, whatever it is. When you're talking about life-threatening and life-changing things, like everything changes.* [Nonuser 1]

Barriers to Activation and Adoption of MyChart

Despite MyChart's potential benefits, patients and caregivers who were aware of MyChart but had not activated accounts shared hesitations about signing up. This included difficulty navigating the registration process, misperceptions about the cost to activate an account, a perceived low need for MyChart since they were infrequent health care users or felt the available features were too limited, and a preference for face-to-face

XSL•FO
RenderX

interactions. The registration process was a significant challenge for both MyChart users and nonusers, as they attempted to register MyChart accounts but faced roadblocks; despite some patients trying to seek assistance from the hospital or family members, a few encountered difficulties such as 2-factor authentication, which combined with their limited technical skills, ultimately led them to abandon the registration process.

> *I have to put all these numbers, whatever it is and then they say it's not right. Again, I'm putting in the numbers, it's not right. I did it three times, after that I think it says your time has expired so come back. So, how many times am I going to put in the numbers and still I'm not getting anyplace? That is where my frustration starts. One time, two times, three times. It's hard to understand why they have to think that every ... I know the younger generation picks it up very fast. Why do they think that everybody is that fast, especially the seniors.* [Nonuser 3]

Some nonusers who were accustomed to advanced features and interconnectedness in other apps found MyChart underwhelming. They questioned the effort involved in registration and some were frustrated by the fragmentation of separate patient portals offered by different providers (eg, family physicians and physiotherapists). This fragmentation hinders information flow, creates repetitive tasks to create accounts and add personal information, and forces patients to manually connect the dots between health care providers.

> *But in addition to my doctor, everybody I've been to has one of these portals, and they're all different. If you had all signed up for the same premade system, I'd be willing to learn it in order to talk to all the medical people in my life ... So the thing is it adds up and it adds up and they're all different. And the thing is as a patient you might only use it two or three times and never see that specialist again.* [Nonuser 2]

Some infrequent health care users mentioned they did not need features like record keeping and checking medical results and felt comfortable receiving information through face-to-face interactions, especially if they lacked the health literacy required to understand the information presented in MyChart.

> *[...] It's no use to me if I'm not talking to the doctor about it ... I get the general drift of things because I know medical vocabulary and I work in that area. But I don't have the knowledge of a doctor. And without that you can go very far astray.* [Nonuser 2]

Discussion

Principal Results

This mixed methods study explored the initial experiences and perspectives of 5651 patients about MyChart implementation at a large community hospital, including facilitators and barriers to accessing and using the portal. While most patients learned about MyChart through postvisit summaries, hospital staff, and email communications, barriers such as limited awareness, misconception around cost, registration difficulties, and unclear benefits of the portal hindered its uptake among nonusers. Users generally perceived MyChart to be user-friendly and enhance their ability to understand and prepare for their care. However, both users and nonusers highlighted the importance of a portal meeting their needs and expectations as a condition for uptake and impact on their care. Many participants highlighted a desire for more functionalities, such as access to imaging results, consultation notes, messaging with providers, and tools for interpreting medical information, which was often driven by comparisons to patient portals at other hospitals that offered more features. Expectations also included connectivity across a range of providers and institutions, frictionless account set-up and access, and technical support. Nonusers expressed interest in learning more about MyChart and its specific functionalities through educational materials to help them better understand the portal and its benefits, as well as technical and registration support to facilitate easier access.

Comparison With Prior Work

Our study participants consistently highlighted MyChart as a valuable tool for accessing timely health information, preparing for visits, managing appointment schedules, and actively participating in their health care. However, whether they were MyChart users or not, participants in this study identified several barriers to accessing the patient portal, including a lack of knowledge of the platform, registration difficulties, limited interest in available features, a preference for face-to-face interactions, and low digital literacy. MyChart users were more often older individuals, which may reflect an increase in health care use with age and a higher need for managing personal health information. Notably, a similar proportion of users and nonusers resided in neighborhoods characterized by racialized or newcomer populations, or those with low material resources, which is inconsistent with use patterns observed in other studies—often, portal users have higher socioeconomic status than nonusers [14,25,26]. Our findings may be reflective of the patient population in our health system's catchment area, which has a high proportion of newcomer and lower-resourced groups. In both our study and the literature, access to medical records and clinical notes can be beneficial for marginalized populations in understanding and managing their own care, as it may help overcome challenges to do with language, digital literacy, and recalling information, for example [27].

In this study, immediate access to medical records and test results was a complex issue. The majority of MyChart users believed this access empowered them with greater control and understanding of their health information and lessened anxiety associated with waiting for test results. Conversely, nonusers worried about potential misunderstandings and negative outcomes of viewing concerning medical information, especially before they could discuss it with a health care provider. Numerous studies in the literature [28,29] align with these perspectives, indicating that patients appreciate timely information but there is potential for negative emotions. From our findings, it is unclear whether nonusers only perceive the risk of higher worry, or if self-awareness of anxious tendencies causes certain patients to avoid portals altogether. Health and digital literacy may also differ among MyChart users and nonusers, where those who feel more equipped to understand health information and use other tools such as credible health

XSL•FO
RenderX

websites to interpret complex details may be more likely to use MyChart [30]. Some research [29] revealed a strong patient preference for receiving test results through a portal regardless of whether they were normal or abnormal; however, health care providers have expressed concerns about this practice [28,31,32], suggesting increased patient anxiety and workload changes due to potential increases in messages, calls, and urgent visits to discuss the results. Many health care providers have advocated for a more structured approach, suggesting that patients should receive access to medical records and test results postappointment, allowing for discussion and clarification of results in a clinical context. However, this contradicts the strong preference MyChart users have for access to their complete health information as soon as possible, even if there is a risk of viewing concerning results without the immediate guidance of a health care provider, and may cause strain on health care providers to review and release results to patients within certain timeframes [2,28,29].

Strengths and Limitations

This study examined the early experiences of patient portal implementation among users and nonusers at a large community hospital in Canada that serves a diverse population. With a strong response rate of over 5000 patients and caregivers, this study captures the experiences and perspectives of a wide range of health users. The use of a mixed methods design that combined both qualitative and quantitative approaches enabled us to understand the facilitators and barriers associated with portal adoption and use. By capturing the perspectives of both users and nonusers with different ethnicities and socioeconomic backgrounds, we generated a comprehensive understanding of the factors that influenced patient portal adoption and engagement. These rich insights improve the transferability of our findings across patient populations and health system settings and can inform a range of patient portal implementation strategies.

This study has limitations. While we used multiple recruitment strategies including in-person invitations, posters, and email communications, this study was limited to English speakers due to the language requirements of the questionnaire and interviews. Given this, the findings may not reflect all the views and perspectives of those in the Mississauga and West Toronto region and may not apply to other linguistically diverse populations. We also were not able to determine the exact number and characteristics of patients who were invited to participate using our multipronged recruitment strategy, which limited our ability to identify response rates among specific patient populations (eg, newcomer groups). As this study examined a staged roll-out of MyChart features, the results may not be applicable to other health care organizations using different implementation approaches. Finally, given that this study was about MyChart, our sample of nonusers may not have been fully representative of that population if the questionnaire was perceived to be less relevant to their experiences.

Conclusions

In conclusion, our findings reveal actionable and transferable learnings about the implementation of MyChart. Low digital literacy and lack of awareness of MyChart's available features were barriers to MyChart activation and use among both users and nonusers. A surprising barrier to accessing MyChart was a misperception about the costs associated with creating an account; given that several digital health tools require payment, health systems should clearly communicate when available portals are free to maximize uptake. Widely marketed and accessible interventions to support digital literacy and help patients learn about and navigate MyChart could increase patient awareness of available features and their potential to positively impact patient experiences. Similarly, improving patients' health literacy could reduce perceived worry associated with viewing concerning results in MyChart, and drive additional benefits of using the portal by increasing the interpretability of information. This study highlighted a strong desire among some patients and caregivers for access to additional medical records and test results, and to exchange messages with health care providers. While expanding access to these features can improve patient engagement and care management, it is crucial to consider the potential impact on patient-provider relationships and provider workflows. At the same time, some patients had little interest in MyChart, which is important for health systems to consider when setting targets for patient portal uptake. As health systems increasingly offer patient portals to increase patient and family engagement in care, support shared decision-making, and provide patients access to their health information, a culture of timely, comprehensive access to complete medical records is building. To ensure uptake and positive experiences with portals, health systems need to meet patient expectations by providing access to a variety of features, especially when multiple health systems offer portals to the same population.

Acknowledgments

We would like to acknowledge the Trillium Health Partners (THP) MyChart Project Executives: Dr Chris Hayes, Dr Joan Murphy, and Danielle Sanagan, and the Steering Committee, Adam Gdyczynski, Ying Mi, Raj Patel, Hassaan Khurram, Marvelyn Oni, and Shelby Roloson for their support of recruiting participants, extracting and organizing data, and advertising this project. We thank the THP MyChart Patient and Family Advisory Council members who reviewed the protocol, questionnaire, and interview guides to ensure they were relevant, meaningful, and clearly written, as well as questionnaire pilot testers Selina Brudnicki, Charlotte De Souza, Kathleen Carlin, and Sam Padayachee for reviewing the tool for usability and clarity. Finally, we thank the volunteers who helped administer the questionnaire in clinical areas. Funding for this work was provided by THP as it was a quality improvement project.

XSL•FO
RenderX

Data Availability

The datasets generated during or analyzed during this study are not publicly available as participants did not consent to share the data outside of the research team.

Authors' Contributions

SV, KK, TT, and WPW conceptualized the study and created the protocol. ST and SV collected and analyzed data and drafted the manuscript. All authors reviewed, edited, and approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire and interview guide.
[DOCX File , 26 KB - jmir_v27i1e66353_app1.docx]

References

1. Zhong X, Park J, Liang M, Shi F, Budd PR, Sprague JL, et al. Characteristics of patients using different patient portal functions and the impact on primary care service utilization and appointment adherence: retrospective observational study. J Med Internet Res 2020;22(2):e14410 [FREE Full text] [doi: 10.2196/14410] [Medline: 32130124]
2. Tapuria A, Porat T, Kalra D, Dsouza G, Xiaohui S, Curcin V. Impact of patient access to their electronic health record: systematic review. Inform Health Soc Care 2021;46(2):192-204 [FREE Full text] [doi: 10.1080/17538157.2021.1879810] [Medline: 33840342]
3. MyChart is Epic. MyChart. URL: https://www.mychart.org/ [accessed 2024-04-04]
4. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
5. Britto MT, Hesse EA, Kamdar OJ, Munafo JK. Parents' perceptions of a patient portal for managing their child's chronic illness. J Pediatr 2013;163(1):280-1.e1-2. [doi: 10.1016/j.jpeds.2013.02.041] [Medline: 23541773]
6. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
7. Fiks AG, Mayne S, Karavite DJ, DeBartolo E, Grundmeier RW. A shared e-decision support portal for pediatric asthma. J Ambul Care Manage 2014;37(2):120-126 [FREE Full text] [doi: 10.1097/JAC.0000000000000025] [Medline: 24594560]
8. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
9. Graham TAD, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
10. Park K, Park MD, Longhurst CA. Patient and family access to electronic health records: a key ingredient for a pediatric learning health system. J Particip Med 2015;7:e2 [FREE Full text] [Medline: 25664199]
11. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
12. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
13. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016;23(2):848. [doi: 10.14236/jhi.v23i2.848] [Medline: 27869582]
14. Johnson KB, Ibrahim SA, Rosenbloom ST. Ensuring equitable access to patient portals-closing the 'techquity' gap. JAMA Health Forum 2023;4(11):e233406 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.3406] [Medline: 37948065]
15. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020;172(11 Suppl):S123-S129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]
16. Annual report 2024. Trillium Health Partners. URL: https://www.thp.ca/AR2024/A/Home.html#communityweserve [accessed 2024-08-02]
17. With the patient at the heart. Epic. URL: https://www.epic.com/ [accessed 2024-08-02]

XSL•FO
RenderX

18. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. Los Angeles: SAGE Publications; 2011.
19. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
20. Shelton RC, Chambers DA, Glasgow RE. An extension of RE-AIM to enhance sustainability: addressing dynamic context and promoting health equity over time. Front Public Health 2020;8:134 [FREE Full text] [doi: 10.3389/fpubh.2020.00134] [Medline: 32478025]
21. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2025-01-14]
22. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The importance of visit notes on patient portals for engaging less educated or nonwhite patients: survey study. J Med Internet Res 2018;20(5):e191 [FREE Full text] [doi: 10.2196/jmir.9196] [Medline: 29793900]
23. Avdagovska M, Ballermann M, Olson K, Graham T, Menon D, Stafinski T. Patient portal implementation and uptake: qualitative comparative case study. J Med Internet Res 2020;22(7):e18973 [FREE Full text] [doi: 10.2196/18973] [Medline: 32716308]
24. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
25. Zhao J, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
26. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]
27. Blease C, Fernandez L, Bell SK, Delbanco T, DesRoches C. Empowering patients and reducing inequities: is there potential in sharing clinical notes? BMJ Qual Saf 2020;29(10):1-2. [doi: 10.1136/bmjqs-2019-010490] [Medline: 32188711]
28. Petrovskaya O, Karpman A, Schilling J, Singh S, Wegren L, Caine V, et al. Patient and health care provider perspectives on patient access to test results via web portals: scoping review. J Med Internet Res 2023;25:e43765 [FREE Full text] [doi: 10.2196/43765] [Medline: 37856174]
29. Steitz BD, Turer RW, Lin CT, MacDonald S, Salmi L, Wright A, et al. Perspectives of patients about immediate access to test results through an online patient portal. JAMA Netw Open 2023;6(3):e233572 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.3572] [Medline: 36939703]
30. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
31. van Kuppenveld SI, van Os-Medendorp H, Tiemessen NA, van Delden JJ. Real-time access to electronic health record via a patient portal: is it harmful? A retrospective observational study. J Med Internet Res 2020;22(2):e13622 [FREE Full text] [doi: 10.2196/13622] [Medline: 32044753]
32. Taneja S, Kalia K, Tang T, Wodchis W, Vanderhout S. Examining healthcare provider experiences with patient portal implementation: a mixed methods study. J Med Internet Res 2024. [doi: 10.2196/65967]

Abbreviations

ON-Marg: Ontario Marginalization Index
THP: Trillium Health Partners

Edited by A Mavragani; submitted 10.09.24; peer-reviewed by C McBride, AM Weggelaar-Jansen; comments to author 03.12.24; revised version received 12.12.24; accepted 28.12.24; published 23.01.25.

Please cite as:
Vanderhout S, Taneja S, Kalia K, Wodchis WP, Tang T
Patient Experiences and Perspectives When MyChart is Introduced in a Large Community Hospital: Mixed Methods Study
J Med Internet Res 2025;27:e66353
URL: https://www.jmir.org/2025/1/e66353
doi:10.2196/66353
PMID:39847431

©Shelley Vanderhout, Shipra Taneja, Kamini Kalia, Walter P Wodchis, Terence Tang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

XSL•FO
RenderX

and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A 4-Week Mobile App–Based Telerehabilitation Program vs Conventional In-Person Rehabilitation in Older Adults With Sarcopenia: Randomized Controlled Trial

Lu Zhang[1], MD; Ying Ge[1], MM; Wowa Zhao[1], MBBSMD; Xuan Shu[1], MM; Lin Kang[2], MD; Qiumei Wang[2], MD; Ying Liu[1], MD

[1]Department of Rehabilitation Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]Department of Geriatric Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Corresponding Author:
Ying Liu, MD
Department of Rehabilitation Medicine
Peking Union Medical College Hospital
Chinese Academy of Medical Sciences and Peking Union Medical College
No. 1, Shuai Fu Yuan Hutong
Dongcheng District
Beijing, 100730
China
Phone: 86 18612671468
Email: pumch9887@163.com

Abstract

Background: Sarcopenia is closely associated with a poor quality of life and mortality, and its prevention and treatment represent a critical area of research. Resistance training is an effective treatment for older adults with sarcopenia. However, they often face challenges when receiving traditional rehabilitation treatments at hospitals.

Objective: We aimed to compare the effects of a digital rehabilitation program with those of traditional therapist-supervised rehabilitation training on older adults with sarcopenia.

Methods: In total, 58 older adults with sarcopenia were recruited offline and randomized (1:1) into 2 groups: the telerehabilitation group (TRG, n=29, 50%) and the in-person rehabilitation group (IRG, n=29, 50%). Both groups underwent 4-week resistance training targeting 6 major muscle groups. The TRG received exercise guidance via a mobile app, while the IRG received in-person training from a therapist. Offline assessments of body composition, grip strength, and balance using the 30-Second Arm Curl Test (30SACT), 30-Second Sitting-to-Rising Test (30SSRT), quadriceps femoris extension peak torque (EPT) and extension total power (ETP), Berg Balance Scale (BBS), Timed Up-and-Go Test (TUGT), 6-Minute Walk Test (6MWT), and Instrumental Activities of Daily Living (IADL) scale, were conducted before and after the intervention.

Results: Of the 58 patients, 51 (88%; TRG: n=24, 47%; IRG: n=27, 53%) completed the trial. After 4 weeks of intervention, the mean grip strength increased from 18.10 (SD 5.56) to 19.92 (SD 5.90) kg in the TRG (P=.02) and from 18.59 (SD 5.95) to 19.59 (SD 6.11) kg in the IRG (P=.01). The 30SACT and 30SSRT scores increased from 12.48 (SD 2.68) to 14.94 (SD 3.68) times (P=.01) and from 15.16 (SD 7.23) to 16.58 (SD 8.42) times (P=.045), respectively, in the TRG and from 12.25 (SD 4.19) to 14.68 (SD 4.36) times (P=.003) and from 14.31 (SD 4.04) to 16.25 (SD 4.91) times (P=.01), respectively, in the IRG. The quadriceps femoris EPT increased from 26.19 (SD 10.26) to 35.00 (SD 13.74) Nm (P=.004) in the TRG and from 26.95 (SD 11.81) to 32.74 (SD 12.33) Nm (P=.003) in the IRG. The BBS scores significantly improved in both groups (P<.001), with the mean TRG score increasing by 3.19 (SD 2.86) points and the mean IRG score by 3.06 (SD 2.44) points. Neither group exhibited significant within-group changes on the TUGT or the 6MWT. Both groups reported significant improvements in the IADL (TRG: P=.04; IRG: P=.02). Between-group comparisons revealed no significant differences in changes in all indicators.

Conclusions: A 4-week remote resistance training program is effective in improving strength, balance, and the IADL in older adults with sarcopenia, with effects comparable to rehabilitation supervised by a physical therapist. Telerehabilitation may be a convenient and effective alternative for older adults with sarcopenia who have limited access to rehabilitation resources.

XSL•FO
RenderX

Trial Registration: ChiCTR 2300071648; https://www.chictr.org.cn/showprojEN.html?proj=196313

(J Med Internet Res 2025;27:e67846) doi:10.2196/67846

KEYWORDS

telerehabilitation; elderly; sarcopenia; resistance exercise; rehabilitation; gerontology; aging; randomized controlled trial; rehabilitation training; body composition; strength; balance; cardiorespiratory endurance; self-care; physical therapy

Introduction

Muscle reduction typically progresses gradually between the ages of 40 and 50 years but accelerates rapidly after the age of 60 years, with approximately 30% of skeletal muscle mass lost between the ages of 50 and 80 years [1,2]. Sarcopenia is characterized by a substantial reduction in skeletal muscle mass and function. This condition heightens the risk of falls, fractures, and physical disability and is closely linked to a poorer quality of life, as well as increased mortality rates [3,4].

According to a systematic review from 2022, the prevalence of sarcopenia worldwide is estimated to be 8%-36% in individuals under the age of 60 years and 10%-27% in those over 60 years [5]. Based on the literature from 2016 to 2019, in Asian countries, the prevalence ranges from 5.5% to 25.7% [6]. More specifically, in China, recent data suggest that the overall prevalence of sarcopenia among individuals aged 60 years and older is about 20.7%, and this rate has been rising over the past 5 years [7]. Thus, the prevention and treatment of sarcopenia represent a critical area of research, with significant implications for the health of the older population.

Resistance training has been shown to enhance muscle strength, aerobic endurance, and muscle mass [8]. This includes various exercises using free weights, machines, resistance bands, or body weight. Resistance training has demonstrated notable efficacy in improving strength in patients with sarcopenia [9-12]. Guidelines from the World Health Organization, as well as those from the United Kingdom and Australia, recommend 2-3 sessions of resistance training per week to prevent and manage sarcopenia [13,14]. Furthermore, consistent adherence to exercise is crucial, as irregular or intermittent training may impair treatment outcomes [15].

However, older adults (defined in this study as individuals aged 60 years or older, according to standards set by the World Health Organization [16]) with sarcopenia often face challenges when engaging in resistance training. In middle-income countries in particular, access to professional rehabilitation services is limited due to inadequate rehabilitation facilities, transportation barriers, financial constraints, and diminished physical ability, all of which reduce patient participation [14,17]. Additionally, older adults tend to have low adherence to self-guided home training programs, largely due to the lack of professional supervision, fear of injury, and low motivation [9].

Given these challenges, digital programs delivered through mobile apps offer a potential solution, allowing patients to receive professional guidance and monitoring from the comfort of their homes. This approach could alleviate the burden on hospital resources and improve patient adherence to training programs [18]. Recent studies have shown that home-based programs can have a positive impact on improving the function and quality of life of older adults [19,20]. Therefore, this study aimed to compare the effects of a digital rehabilitation program with traditional therapist-supervised rehabilitation training in older adults with sarcopenia, addressing the aforementioned challenges.

Methods

Study Design

This single-center randomized controlled trial (RCT; Multimedia Appendix 1) was conducted at the Peking Union Medical College Hospital, affiliated with the Chinese Academy of Medical Sciences. The study is registered with the Chinese Clinical Trial Registry (registration number: ChiCTR 2300071648).

Ethical Considerations

This study was approved by the Ethics Committee of the Peking Union Medical College Hospital (approval number: JS-2648). All participants provided written informed consent prior to enrollment, ensuring their voluntary participation with a full understanding of the study's objectives, procedures, and potential risks. All personal data collected were anonymized and stored in encrypted electronic databases accessible only to the research team members. Neither the manuscript nor the supplementary materials include any identifiable information or photographs of individual participants. Additionally, no form of financial compensation was provided to participants; participation was entirely voluntary and would not influence their routine medical services.

Participants

All participants in this study were recruited offline from the Peking Union Medical College Hospital, a large comprehensive hospital. The specific process was as follows: First, the researchers posted recruitment advertisements in the hospital's outpatient hall. Second, potential participants contacted the recruitment team (2 geriatricians and a rehabilitation physician with extensive experience in sarcopenia) over the phone, and the recruitment team conducted an initial telephone screening. Third, potential participants who passed the initial telephone screening were invited to the hospital for a face-to-face detailed assessment, including medical history taking, physical examination, and necessary laboratory tests. Finally, if the conditions of the potential participants met all the eligibility criteria, they were formally included in the study after obtaining written informed consent. During the informed consent process, participants were not informed whether they were in the intervention group or the control group in order to avoid bias. The inclusion and exclusion criteria are outlined in Textbox 1.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria for the study.

Inclusion Criteria

- Diagnosed with sarcopenia based on the 2019 Asian Working Group for Sarcopenia (AWGS) criteria [6]
- Age between 60 and 80 years
- Able to operate a smartphone and follow the exercise regimen, with sufficient literacy in the local language to navigate a Chinese language mobile app
- No participation in any rehabilitation program within the past 30 days

Exclusion Criteria

- Severe cognitive impairment (Mini-Mental State Examination [MMSE] score<24), hearing impairment, or vision impairment
- Severe neurological, musculoskeletal, or organ (cardiovascular, pulmonary, hepatic, renal) diseases or malignant tumors
- Uncontrolled hypertension, unstable metabolic diseases, or any other condition deemed inappropriate for participation by the researchers

Sample Size Calculation

The sample size was calculated using PASS 15 software (NCSS, LLC). Based on the principle of noninferiority RCTs and previous clinical studies [21], the mean difference in the grip strength between the home-based exercise group and the therapist supervision–based exercise group was 3, the SD was estimated to be 4 for both groups, and the noninferiority margin for the grip was 1. A sample size of 46 was required based on a 1-sided α=.025 and β=.1, and a sample size of 58 was required to account for a 20% dropout rate. The formula used is the following:



Randomization

A total of 58 participants were randomly assigned to either the telerehabilitation group (TRG, n=29, 50%) or the in-person rehabilitation group (IRG, n=29, 50%) using a randomization platform. Based on the platform's results (eg, C, T, C, T, T, C), slips of paper labeled "T" and "C" were placed into sealed, opaque, and uniformly sized envelopes. After completing baseline measurements, the envelopes were opened in sequence to reveal group assignments. The allocation sequence was generated using a blocked randomization model by 2 researchers who were not involved in the study.

Exercise Program

The training program for patients with sarcopenia in both the TRG and the IRG was identical. The program consisted of 6 resistance exercises targeting specific muscle groups, including back extensors, biceps brachii, gluteus maximus, gluteus medius, the deltoid, and quadriceps femoris. The specific exercises included glute bridges, resisted elbow flexion, hip extension, hip abduction, shoulder abduction, and knee extension (Figure 1). Participants performed resistance training 3 times a week, with each exercise performed in 3 sets of 10 repetitions per set, over a 4-week period (Multimedia Appendix 2). Effective training intensity was ensured through the use of the Borg rating of perceived exertion (RPE), with a target RPE of 12-14 at the end of each exercise session. Each session included approximately 10 minutes of warm-up, 40 minutes of resistance training, and 10 minutes of stretching, totaling around 1 hour.

Figure 1. Exercises included (a) glute bridge, (b) resisted elbow flexion, (c) resisted hip extension, (d) resisted hip abduction, (e) resisted shoulder abduction, and (f) resisted knee extension.

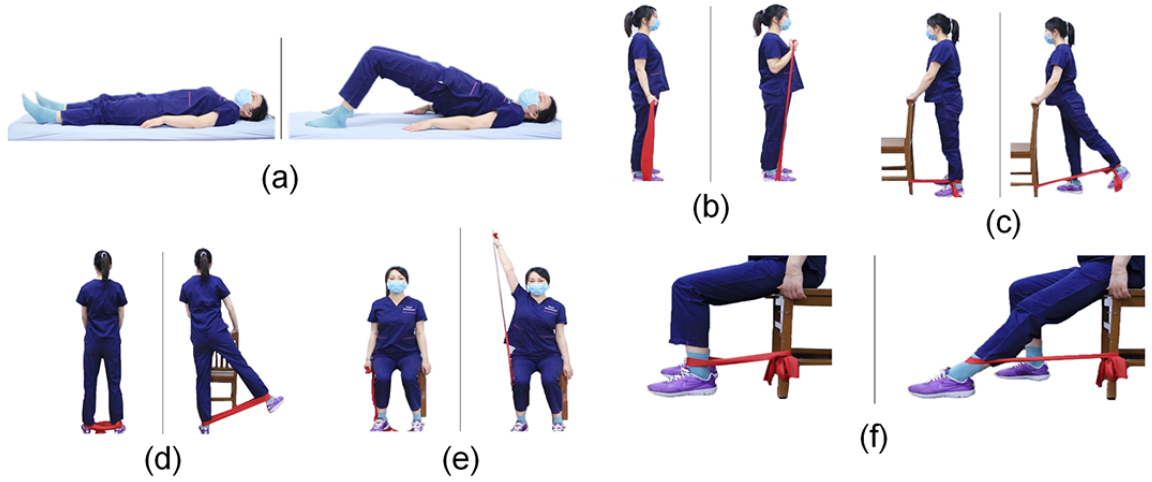

XSL•FO
RenderX

Intervention Group (TRG)

Before initiating at-home training, each participant attended an individual, face-to-face introductory session at the hospital's rehabilitation outpatient clinic. A physical therapist demonstrated the exercises, ensured the participants understood the correct technique, and assisted them with downloading and registering the mobile app on their personal smartphones. Each participant received a personalized exercise protocol via the app and was provided with resistance bands with appropriate tension levels based on their initial assessment. Following this initial session, participants performed the exercises independently at home, guided by instructional videos and written instructions embedded in the app, which were marked with the hospital's logo (Figure 2).

Figure 2. The 3 different parts of the telerehabilitation system (the language of the interface is translated from Chinese). The doctor's portal can be used to create and modify exercises, monitor training progress, and view patient data. Patients can use the user portal to complete the prescribed exercises and provide feedback to physical therapists. Finally, the transmitter portal encrypts and transmits the data collected, ensuring the overall system's integrity.

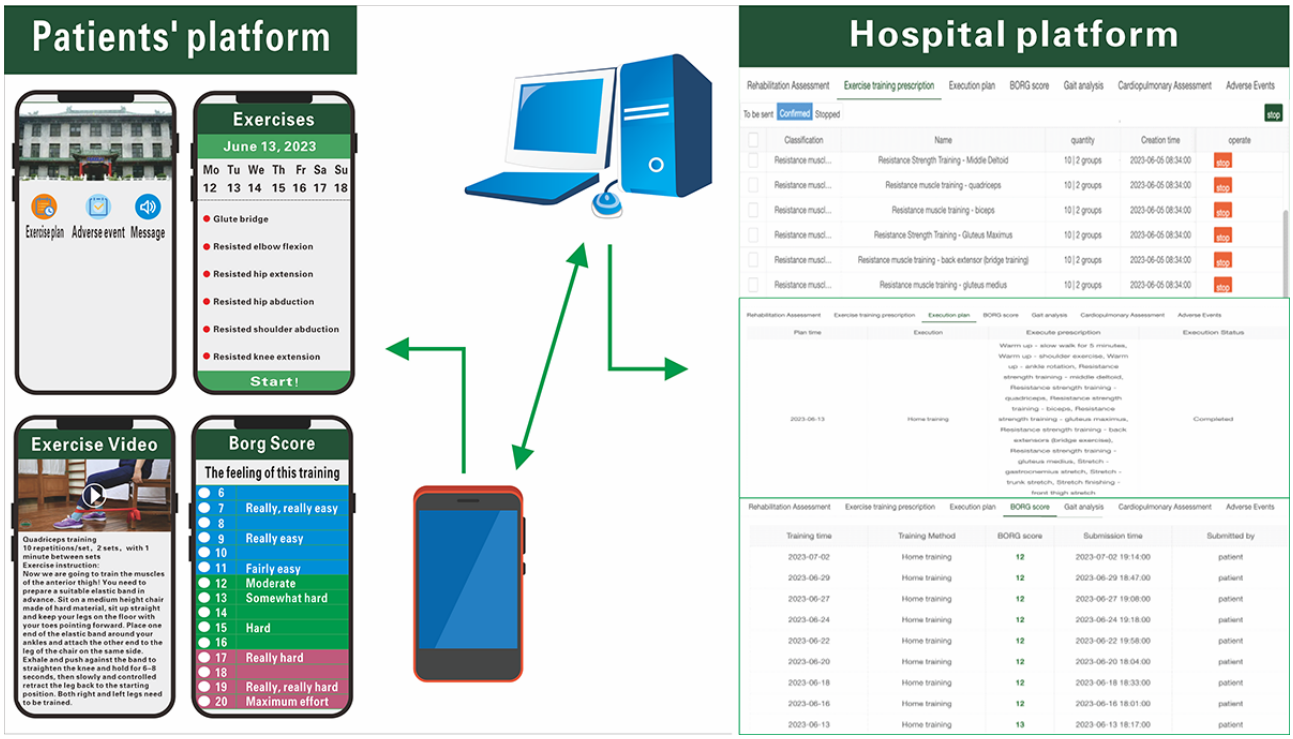

The telerehabilitation system comprised 3 main components:

- Physician portal: Accessed by the medical team to upload exercise videos and protocols and track patient adherence.
- User portal (patient interface): Accessed via a password-protected smartphone app in the local (Chinese) language. Patients navigated using simple menus and icons to view exercise instructions, watch demonstration videos, and report session completion. The interface includes step-by-step written instructions and demonstration videos. Patients were required to input their RPE values after each session.
- Transmission portal: A secure data transmission platform integrated within the system automatically records the exercise duration, frequency, and patient-reported RPE. Patients were instructed to report any adverse events through the app immediately, triggering an alert to the research team.

Control Group (IRG)

Participants assigned to the IRG attended scheduled sessions in the hospital's rehabilitation department, which is a typical form of training currently in practice. A physical therapist supervised each session, providing on-site guidance, feedback, and adjustments to ensure correct exercise performance. The IRG followed the same exercise protocol, frequency, and intensity guidelines as the TRG, but all training took place face to face under professional supervision.

Measures

All assessments were conducted by a dedicated blinded physical therapist with professional certification. This individual conducted all tests within 3 days in the hospital before and after the 4-week intervention for every participant.

Body Composition

Body composition was measured using bioelectrical impedance analysis (BIA-101 Anniversary Sport Edition, Akern-RJL Systems), with specific metrics including the total skeletal muscle mass (TSM), body fat percentage (BFP), and skeletal muscle mass index (SMI).

Strength

Strength was assessed using grip strength, the 30-Second Arm Curl Test (30SACT), the 30-Second Sitting-to-Rising Test (30SSRT), and the quadriceps femoris extension peak torque (EPT) and extension total power (ETP). For the grip strength test, participants used their dominant hand to apply the maximum pressure to the dynamometer handle (Jamar Hydraulic T, Patterson Medical). In the 30SACT, participants sat with

XSL•FO
RenderX

their upper arms pressed against their torso while holding 8 lb dumbbells (men) and 5 lb dumbbells (women). They were instructed to perform elbow curls as quickly as possible for 30 seconds. In the 30SSRT, participants sat on a 43-cm-high chair and alternated between sitting and standing as many times as possible within 30 seconds. Each of these tests was performed 3 times, with a 1-minute rest between trials, and the maximum value was recorded as the final result.

Quadriceps extension strength indices were measured using an isokinetic dynamometer (ISOMOVE, TecnoBody). The seat of the dynamometer was reclined at approximately 85°, with the participant's thigh, pelvis, and torso secured by straps. The rotation axis of the device was aligned with the lateral epicondyle of the femur, and the lever pad was positioned 3 cm above the lateral malleolus. The test consisted of 5 maximal repetitions at a speed of 20°/second, during which participants extended their knees as fast as possible against the resistance shown on the screen, reaching the target end point indicated by the graphical display.

Balance, Cardiopulmonary Endurance, and Self-Care Ability

Balance was assessed using the Berg Balance Scale (BBS) and the Timed Up-and-Go Test (TUGT). The BBS evaluated 14 tasks, with each task scored from 0 (unable to complete) to 4 (completed independently), and the total score represented the sum of all individual task scores. The TUGT assessed functional mobility and fall risk by timing how long it took a participant to rise from a chair, walk a short distance (3 m), turn, walk back, and sit down again.

Cardiopulmonary endurance was measured using the 6-Minute Walk Test (6MWT), which quantified the distance a participant could walk on a flat, hard surface in 6 minutes. The primary outcome was the total distance walked, with longer distances indicating better cardiopulmonary function.

The Instrumental Activities of Daily Living (IADL) scale was used to assess participants' self-care abilities.

Statistical Analysis

Data analysis was performed using IBM SPSS version 26.0. Categorical variables were expressed as numbers, while continuous variables were presented as the mean (SD). Categorical variables were compared using the chi-square test. Paired *t* tests were used for within-group comparisons before and after the intervention, while independent-sample *t* tests were applied for comparisons between groups. $P<.05$ was considered statistically significant.

Results

Study Flow and General Participant Characteristics

Between May 30, 2023, and July 1, 2024, 106 patients were assessed for eligibility. Of these, 43 (40%) patients were excluded based on the criteria, while 5 (5%) declined participation after the initial screening. As a result, 58 (55%) patients were included in the final study. As shown in Figure 3, participants were randomly assigned to 1 of 2 groups: the TRG (n=29, 50%) or the IRG (n=29, 50%).

During the study, 4 (14%) participants in the TRG and 1 (3%) participant in the IRG did not complete the 4-week follow-up. Additionally, 1 (3%) participant from each group discontinued the training due to worsening pre-existing conditions: hip pain in the TRG and elbow pain in the IRG. After discontinuing the exercise, their hip pain subsided, and the elbow pain returned to baseline levels. Ultimately, 24 (83%) patients in the TRG and 27 (93) patients in the IRG completed both the training program and the 4-week follow-up. Adherence to the exercise intervention was defined as the proportion of completed sessions relative to the total prescribed sessions, with adherence rates of 97% in the TRG and 92% in the IRG. The most commonly reported adverse effect was muscle soreness, and no exercise-related injuries or major adverse events were reported.

The baseline characteristics of the patients in the TRG and IRG were comparable. There were no significant differences between the 2 groups in terms of age, gender, height, weight, the TSM, the BFP, the SMI, grip strength, quadriceps femoris EPT, the BBS score, the 6MWT score, or the IADL score (Table 1).

XSL•FO
RenderX

Figure 3. Study flow. IRG: in-person rehabilitation group; TRG: telerehabilitation group.

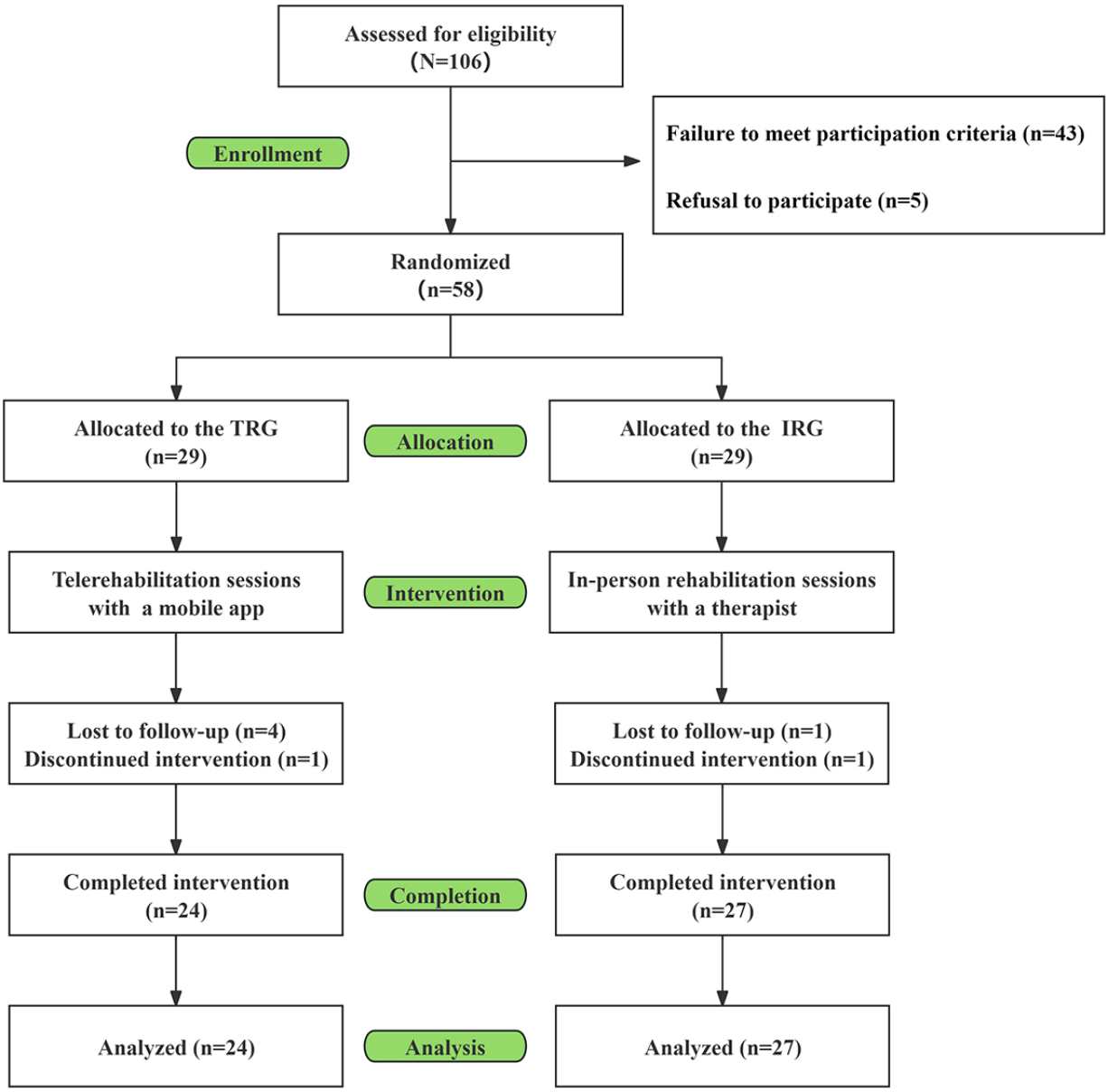

XSL•FO
RenderX

Table 1. Participants' (N=51) physical characteristics.

Characteristic	TRG[a] (n=24)	IRG[b] (n=27)	*P* value
Age (years), mean (SD)	70.47 (6.05)	69.81 (5.76)	.74
Gender (male/female), n (%)	6 (25)/18 (75)	7 (16)/20 (84)	.93
Height (cm), mean (SD)	160.73 (6.63)	161.88 (5.97)	.61
Weight (kg), mean (SD)	53.49 (7.55)	53.95 (8.14)	.87
TSM[c] (kg/m^2), mean (SD)	26.82 (10.04)	25.71 (8.04)	.72
BFP[d] (%), mean (SD)	31.31 (8.34)	28.28 (6.17)	.23
SMI[e] (kg/m^2), mean (SD)	5.60 (0.51)	5.64 (0.77)	.88
Grip strength (kg) , mean (SD)	18.59 (5.95)	18.96 (6.57)	.87
Quadriceps femoris EPT[f] (Nm), mean (SD)	26.19 (10.26)	26.95 (11.81)	.84
BBS[g] score, mean (SD)	49.00 (4.18)	49.11 (4.89)	.95
6MWT[h] score (minutes), mean (SD)	354.19 (97.56)	349.84 (97.43)	.90
IADL[i] score, mean (SD)	21.31 (1.96)	21.42 (1.98)	.87

[a]TRG: telerehabilitation group.

[b]IRG: in-person rehabilitation group.

[c]TSM: total skeletal muscle mass.

[d]BFP: body fat percentage.

[e]SMI: skeletal muscle mass index.

[f]EPT: extension peak torque.

[g]BBS: Berg Balance Scale.

[h]6MWT: 6-Minute Walk Test.

[i]IADL: Instrumental Activities of Daily Living.

Body Composition

After 4 weeks, both groups showed a reduction in the BFP and an increase in the TSM and SMI, but these changes were not statistically significant (Table 2).

Table 2. Baseline, 4-week, and pre- and postintervention changes in the body composition.

Body composition indicators	TRG[a] (n=24)				IRG[b] (n=27)				Between-group difference *P* value
	Baseline, mean (SD)	Week 4, mean (SD)	Difference, mean (SD)	*P* value	Baseline, mean (SD)	Week 4, mean (SD)	Difference, mean (SD)	*P* value	
TSM[c] (kg/m^2)	25.71 (8.04)	27.84 (8.38)	3.04 (9.00)	.27	26.82 (10.04)	29.86 (9.83)	2.14 (8.24)	.20	.76
BFP[d] (%)	28.28 (6.17)	27.96 (5.78)	–1.98 (6.96)	.52	31.31 (8.34)	29.33 (6.59)	–0.32 (2.09)	.27	.34
SMI[e] (kg/m^2)	5.64 (0.77)	5.85 (0.66)	0.30 (0.89)	.28	5.60 (0.51)	5.90 (0.99)	0.21 (0.80)	.25	.77

[a]TRG: telerehabilitation group.

[b]IRG: in-person rehabilitation group.

[c]TSM: total skeletal muscle mass.

[d]BFP: body fat percentage.

[e]SMI: skeletal muscle mass index.

Strength

Table 3 presents the changes in strength before and after the intervention. Both groups demonstrated significant improvements in strength (grip strength, 30SACT score, 30SSRT score, and quadriceps femoris EPT and ETP; $P<.05$), as shown in Table 3. However, there were no statistically

XSL•FO
RenderX

significant differences between the groups in terms of changes in any of the strength indicators (Figure 4).

Figure 4. Grip strength, quadriceps femoris EPT, BBS, and IADL data at baseline and 4 weeks. Error bars represent 95% CIs. BBS: Berg Balance Scale; EPT: extension peak torque; IADL: Instrumental Activities of Daily Living; IRG: in-person rehabilitation group; TRG: telerehabilitation group.

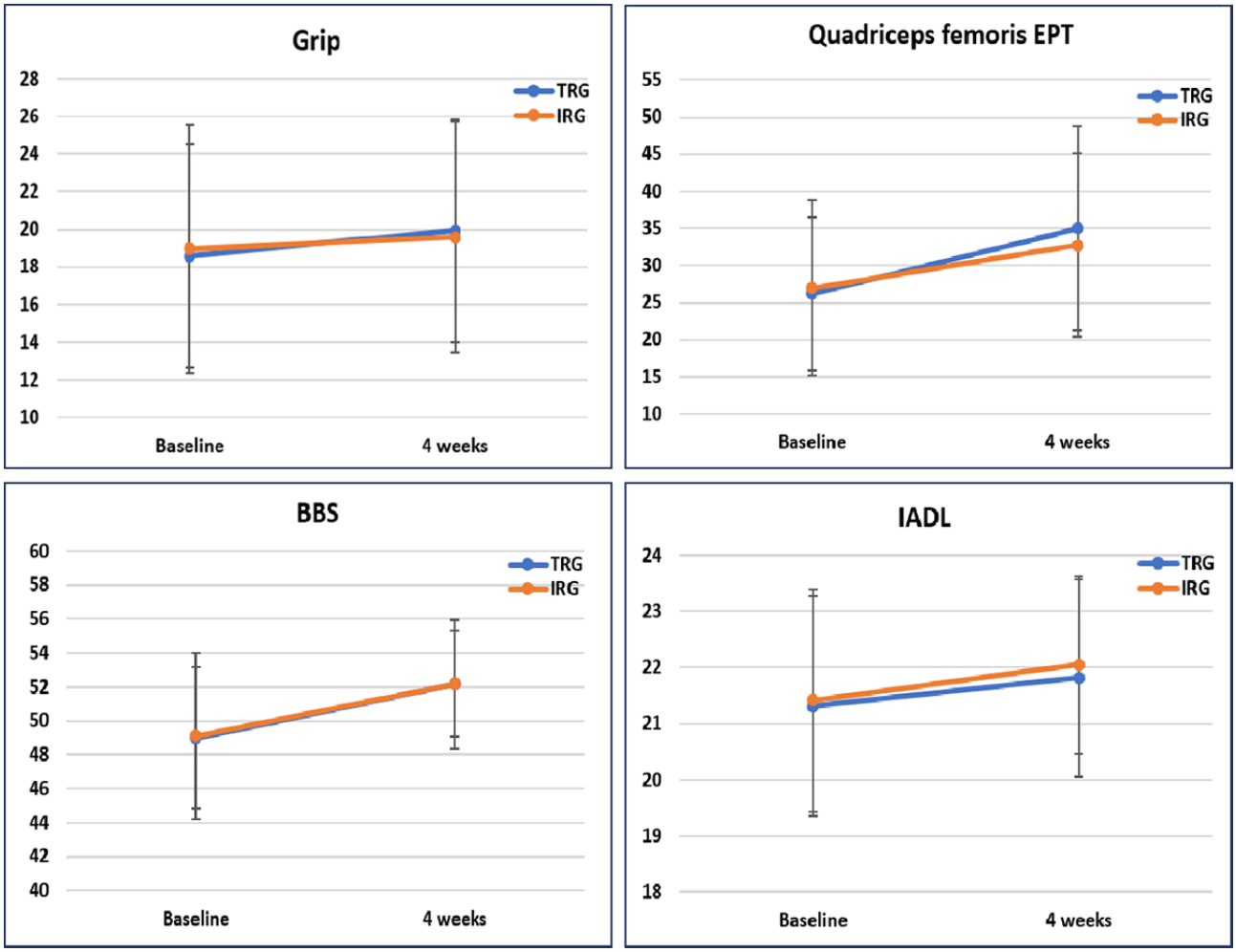

XSL•FO
RenderX

Table 3. Baseline, 4-week, and pre- and postintervention changes for strength.

Strength indicators	TRG[a] (n=24)				IRG[b] (n=27)				Between-group difference *P* value
	Baseline, mean (SD)	Week 4, mean (SD)	Difference, mean (SD)	*P* value	Baseline, mean (SD)	Week 4, mean (SD)	Difference, mean (SD)	*P* value	
Grip strength (kg)	18.10 (5.56)	19.92 (5.90)	1.33 (1.89)	.02	18.59 (5.95)	19.59 (6.11)	0.64 (2.24)	.01	.34
30SACT[c] score (number of times)	12.48 (2.68)	14.94 (3.68)	2.13 (2.92)	.01	12.25 (4.19)	14.68 (4.36)	1.58 (1.98)	.003	.53
30SSRT[d] score (number of times)	15.16 (7.23)	16.58 (8.42)	1.94 (2.64)	.045	14.31 (4.04)	16.25 (4.91)	1.42 (3.02)	.01	.42
Quadriceps femoris EPT[e] (Nm)	26.19 (10.26)	35.00 (13.74)	8.81 (10.36)	.004	26.95 (11.81)	32.74 (12.33)	5.79 (7.22)	.003	.32
Quadriceps femoris ETP[f] (J)	85.96 (26.98)	117.27 (52.48)	18.93 (40.03)	.02	97.27 (38.86)	111.89 (50.31)	19.47 (25.87)	.007	.60

[a]RG: telerehabilitation group.

[b]IRG: in-person rehabilitation group.

[c]30SACT: 30-Second Arm Curl Test.

[d]30SSRT: 30-Second Sitting-Rising Test.

[e]EPT: extension peak torque.

[f]ETP: extension total power.

Balance, Cardiopulmonary Endurance, and the IADL

Postintervention, both groups exhibited significant improvements in BBS and IADL scores ($P<.05$), while the improvements in TUGT and 6MWT scores were not statistically significant (Table 4). There were no significant differences between the groups in the changes observed in balance, cardiopulmonary endurance, or the IADL (Figure 4).

Table 4. Baseline, 4-week, and pre- and postintervention changes in balance, cardiopulmonary endurance, and the IADL[a].

Balance indicators	TRG[b] (n=24)				IRG[c] (n=27)				Between-group difference *P* value
	Baseline, mean (SD)	Week 4, mean (SD)	Difference, mean (SD)	*P* value	Baseline, mean (SD)	Week 4, mean (SD)	Difference, mean (SD)	*P* value	
BBS[d]	49.00 (4.18)	52.19 (3.10)	3.19 (2.86)	<.001	49.11 (4.89)	52.15 (3.79)	3.06 (2.44)	<.001	.88
TUGT[e] (seconds)	9.23 (2.19)	8.57 (1.71)	-0.66 (1.44)	.09	9.68 (2.76)	8.94 (2.14)	-0.73 (1.79)	.10	.90
6MWT[f] (minutes)	354.19 (97.56)	363.19 (82.29)	8.28 (41.92)	.37	349.84 (97.43)	364.63 (112.27)	13.58 (40.72)	.11	.72
IADL	21.31 (1.96)	21.81 (1.76)	0.50 (0.89)	.04	21.42 (1.98)	22.05 (1.58)	0.63 (1.06)	.02	.70

[a]IADL: Instrumental Activities of Daily Living.

[b]TRG: telerehabilitation group.

[c]IRG: in-person rehabilitation group.

[d]BBS: Berg Balance Scale.

[e]TUGT: Timed Up-and-Go Test.

[f]6MWT: 6-Minute Walk Test.

XSL•FO
RenderX

Discussion

Principal Findings

With the increase in life expectancy and the decline in birth rates worldwide, the proportion of older adults in China has risen sharply over the past few decades. As the aging population grows, national health care expenditures have also surged, prompting the Chinese government to prioritize addressing the health challenges associated with population aging. Sarcopenia, characterized by a significant decline in muscle mass due to factors such as nutritional status, physical activity, genetics, or hormonal changes, along with the deterioration of tendon performance and neural patterns [22], leads to a loss of muscle strength and mobility. Sarcopenia is one of the most critical factors contributing to functional decline and loss of independence in older adults. It not only reduces their quality of life but is also associated with increased morbidity and mortality, as well as elevated public health costs [22]. Supervised exercise is considered an effective strategy for managing sarcopenia and frailty, yet many patients lack access to such resources. This study found that a 4-week remote resistance training program significantly improves the strength and balance of patients with sarcopenia, with outcomes comparable to face-to-face rehabilitation supervised by therapists. These findings highlight the potential of telerehabilitation as a feasible solution for populations with limited access to health care resources and demonstrate how digital health solutions can enhance care for older adults and address the unique needs of an aging society.

Previous studies have reported mixed findings on the effects of resistance training on body composition. Some scholars have observed that 12 weeks of resistance training can increase muscle mass in patients with sarcopenic [23,24]. In another study, women with obesity-related sarcopenia underwent 12 weeks of resistance training, 3 times per week. Although the training group experienced an increase in the lean mass of the lower limbs, the increase in the total body lean mass did not reach statistical significance compared to the control group [25]. Another study found that after 3 months of resistance training in older adults with obesity-related sarcopenia, reductions were observed in the fat mass, total fat mass, and BFP in the upper limbs, but there was no significant increase in lean mass [26]. Two systematic reviews and meta-analyses also yielded conflicting conclusions [9,10]. Mechanistically, the increase in muscle mass in older adults following resistance training may be attributed to enhanced muscle protein synthesis [27], increased satellite cell activity and quantity [28], elevated secretion of anabolic hormones [29], improved mitochondrial quality and function [30], and decreased activity of catabolic cytokines [31]. In our study, we found both groups demonstrated an increase in lean mass and a decrease in the BFP, though these changes did not reach statistical significance. This may be due to the relatively short duration of the 4-week resistance training intervention, which may not have been sufficient to induce significant alterations in body composition. It is also possible that extending the training period could result in more pronounced changes in body composition.

In this study, both the TRG and the IRG showed significant improvements in strength-related indicators, including grip strength, 30SSRT and 30SACT scores, and knee extension torque indices. These findings are consistent with previous systematic reviews and meta-analyses [9,10]. Studies have shown that resistance training can increase muscle fiber volume, particularly type II fast-twitch fibers [32]. Similarly, Ruiz et al [33] and Nunes et al [34] confirmed that resistance training enhances muscle function by improving muscle fiber adaptability and optimizing strength output mechanisms. At the molecular level, Sousa-Victor et al [35] demonstrated that resistance training increases satellite cell activity and muscle protein synthesis, reduces muscle degradation, and improves muscle mass. Coletti et al [36] and Gadelha et al [37] showed that this type of exercise activates the mammalian target of rapamycin (mTOR) signaling pathway, influencing protein synthesis, autophagy, and the expression of peroxisome proliferator–activated receptor-gamma coactivator 1 alpha (PGC-1α), thereby promoting sustained muscle growth and enhancing muscle performance. In this study, these effects were observed after just 4 weeks of resistance training, further demonstrating the significant therapeutic role of resistance training in the treatment of sarcopenia. The lack of significant differences in training outcomes between the 2 groups supports the application of telerehabilitation therapies in older adults with sarcopenia.

Lower limb muscle endurance is critical for dynamic balance. The reduction in the muscle cross-sectional area and neural fibers, particularly fast-twitch fibers, leads to a decline in muscle strength, especially in the lower limbs. This decrease in lower limb strength causes patients to spend extended periods sitting or lying down. Prolonged inactivity and sedentary behavior further exacerbate muscle loss and strength decline [25,35]. Reduced lower limb function impairs the ability to perform daily tasks, such as standing up from a chair, picking up items from the floor, walking, and climbing stairs [38]. In situations of instability, the central nervous system compensates by increasing muscle recruitment. One study [39] found that with aging, there is a loss of up to 75% of type II fibers, which diminishes functional mobility and muscle strength output. Concurrently, the synaptic input to α-motor neurons and the number of cortical neurons forming the corticospinal tract also decrease [40]. Resistance training can activate more motor units and muscle fibers, enhance synaptic input to α-motor neurons, and recruit more cortical neurons within the corticospinal tract, thereby improving coordination, reaction time, and strength to support dynamic balance [41,42]. Studies [43,44] have demonstrated that 3 months of resistance training improves muscle mass, physical fitness, and functional outcomes. In another study [11], participants' functional fitness, such as the walking time, the chair stand test score, and the 8-foot up-and-go score, improved compared to the control group. In this study, after 4 weeks of training, both groups showed improvements in their BBS scores, which is consistent with previous findings. However, there was no significant improvement in TUGT scores, likely due to the relatively short duration of the training. The effective stimulation of key muscle groups, such as the gluteus maximus, quadriceps, and tibialis anterior, by resistance

XSL•FO
RenderX

training may improve lower limb strength and neuromuscular coordination, enhancing stability and balance.

In terms of cardiorespiratory endurance, although resistance training primarily focuses on muscle strengthening rather than endurance, research suggests that resistance training may also improve endurance. For patients with sarcopenia, increased lower limb strength can translate into faster walking speed [25], and training peripheral muscles may improve indices related to respiratory function, thereby providing more substantial support for endurance [45]. One study [46] indicated that resistance training could enhance the 6-minute walking distance in frail older adults, while another systematic review [47] found that resistance training improves cardiorespiratory function in healthy older adults, including the peak oxygen uptake, anaerobic threshold, and 6-minute walking distance. In this study, both groups demonstrated improvements in the 6-minute walking distance, though neither reached statistical significance, possibly due to the short duration of the training program. Regarding the IADL, previous systematic reviews have shown that patients with reduced muscle mass are more likely to experience impairments in activities of daily living (ADL) and IADL, with reduced grip strength also being associated with impairments in ADL and IADL [48]. In this study, both groups showed significant improvements in the IADL after training, with no significant differences observed between the groups.

Limitations

One of the limitations of this study is that it was a single-center RCT with a relatively small sample size. To address this limitation, the research team plans to conduct follow-up multicenter studies in regions with limited health care resources. Additionally, the follow-up period was relatively short. Future studies will involve a larger cohort over 8 weeks, 12 weeks, or even longer. Lastly, although participants were instructed to maintain consistent dietary habits throughout the study, detailed nutritional data were not adequately recorded. In addition, participants were not blinded, which may have introduced bias. In this study, the TRG received one-time offline instruction from a physiotherapist and were provided with resistance bands of appropriate tension. Furthermore, the pressure of having a follow-up assessment 1 month later may have increased patient compliance compared to routine application settings.

Conclusion

A 4-week mobile app–based telerehabilitation program produced improvements in muscle strength, balance, and the IADL comparable to those achieved by traditional in-person therapy in older adults with sarcopenia. Additionally, this telerehabilitation approach reduced health care resource use and logistical barriers. By offering an accessible, cost-effective, and patient-centered model of care, telerehabilitation shows promise as a viable alternative for patients with sarcopenia lacking convenient access to conventional rehabilitation services.

Acknowledgments

This research was funded by the Enhancement of the Capability for Technological Innovation in Aging-related Services of the Beijing Municipal Science & Technology Commission (Z191100004419009) and the Central High-Level Hospital Clinical Research Special Fund of the Peking Union Medical College Hospital (2022-PUMCH-B-131) . We are grateful for the organization's financial support. We appreciate information engineer Kai Wang for his kind support and help. We are grateful to the patients and families who participated in this study, without whom the work reported in this paper would not have been possible.

Data Availability

All data generated and analyzed during this study are fully presented in this paper and the multimedia appendices.

Authors' Contributions

Study conception and design was handled by YL and LZ, patient recruitment by LK and QW, rehab video production by XS, intervention management and data collection by YG and WZ, data analysis and draft preparation by LZ, and editing by YL. All authors have approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 2944 KB - jmir_v27i1e67846_app1.pdf]

Multimedia Appendix 2
Remote home-based exercise protocol for patients with sarcopenia.
[DOCX File , 298 KB - jmir_v27i1e67846_app2.docx]

References

XSL•FO
RenderX

1. Addante F, Gaetani F, Patrono L, Sancarlo D, Sergi I, Vergari G. An innovative AAL system based on IoT technologies for patients with sarcopenia. Sensors (Basel) 2019 Nov 14;19(22):4951 [FREE Full text] [doi: 10.3390/s19224951] [Medline: 31739396]
2. Alhussain MH, Alkahtani S, Aljuhani O, Habib SS. Effects of nutrient intake on diagnostic measures of sarcopenia among Arab men: a cross-sectional study. Nutrients 2020 Dec 30;13(1):114 [FREE Full text] [doi: 10.3390/nu13010114] [Medline: 33396961]
3. Wei M, Meng D, Guo H, He S, Tian Z, Wang Z, et al. Hybrid exercise program for sarcopenia in older adults: the effectiveness of explainable artificial intelligence-based clinical assistance in assessing skeletal muscle area. Int J Environ Res Public Health 2022 Aug 12;19(16):9952 [FREE Full text] [doi: 10.3390/ijerph19169952] [Medline: 36011588]
4. Buccheri E, Dell'Aquila D, Russo M, Chiaramonte R, Musumeci G, Vecchio M. Can artificial intelligence simplify the screening of muscle mass loss? Heliyon 2023 May;9(5):e16323 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e16323] [Medline: 37251872]
5. Petermann-Rocha F, Balntzi V, Gray SR, Lara J, Ho FK, Pell JP, et al. Global prevalence of sarcopenia and severe sarcopenia: a systematic review and meta-analysis. J Cachexia Sarcopenia Muscle 2022 Feb;13(1):86-99 [FREE Full text] [doi: 10.1002/jcsm.12783] [Medline: 34816624]
6. Chen L, Woo J, Assantachai P, Auyeung T, Chou M, Iijima K, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307.e2. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]
7. Meng S, He X, Fu X, Zhang X, Tong M, Li W, et al. The prevalence of sarcopenia and risk factors in the older adult in China: a systematic review and meta-analysis. Front Public Health 2024 Aug 5;12:1415398 [FREE Full text] [doi: 10.3389/fpubh.2024.1415398] [Medline: 39161853]
8. Mcleod JC, Currier BS, Lowisz CV, Phillips SM. The influence of resistance exercise training prescription variables on skeletal muscle mass, strength, and physical function in healthy adults: an umbrella review. J Sport Health Sci 2024 Jan;13(1):47-60 [FREE Full text] [doi: 10.1016/j.jshs.2023.06.005] [Medline: 37385345]
9. Chen N, He X, Feng Y, Ainsworth BE, Liu Y. Effects of resistance training in healthy older people with sarcopenia: a systematic review and meta-analysis of randomized controlled trials. Eur Rev Aging Phys Act 2021 Nov 11;18(1):23 [FREE Full text] [doi: 10.1186/s11556-021-00277-7] [Medline: 34763651]
10. Talar K, Hernández-Belmonte A, Vetrovsky T, Steffl M, Kałamacka E, Courel-Ibáñez J. Benefits of resistance training in early and late stages of frailty and sarcopenia: a systematic review and meta-analysis of randomized controlled studies. J Clin Med 2021 Apr 12;10(8):1630 [FREE Full text] [doi: 10.3390/jcm10081630] [Medline: 33921356]
11. Rodrigues F, Domingos C, Monteiro D, Morouço P. A review on aging, sarcopenia, falls, and resistance training in community-dwelling older adults. Int J Environ Res Public Health 2022 Jan 13;19(2):874 [FREE Full text] [doi: 10.3390/ijerph19020874] [Medline: 35055695]
12. Sharma N, Chahal A, Balasubramanian K, Sanjeevi RR, Rai RH, Bansal N, et al. Effects of resistance training on muscular strength, endurance, body composition and functional performance among sarcopenic patients: a systematic review. J Diabetes Metab Disord 2023 Dec 02;22(2):1053-1071. [doi: 10.1007/s40200-023-01283-5] [Medline: 37975091]
13. Fragala M, Cadore E, Dorgo S, Izquierdo M, Kraemer W, Peterson M. Resistance training for older adults: position statement from the National Strength and Conditioning Association. J Strength Cond Res 2019;33(8):2019-2052. [doi: 10.1519/jsc.0000000000003230]
14. Hurst C, Robinson S, Witham M, Dodds R, Granic A, Buckland C, et al. Resistance exercise as a treatment for sarcopenia: prescription and delivery. Age Ageing 2022 Feb 02;51(2):afac003 [FREE Full text] [doi: 10.1093/ageing/afac003] [Medline: 35150587]
15. Vikberg S, Björk S, Nordström A, Nordström P, Hult A. Feasibility of an online delivered, home-based resistance training program for older adults - a mixed methods approach. Front Psychol 2022 Jun 3;13:869573 [FREE Full text] [doi: 10.3389/fpsyg.2022.869573] [Medline: 35726270]
16. Technical Advisory Group for measurement, monitoring and evaluation of the UN Decade of Healthy Ageing. World Health Organization. 2020 Dec 14. URL: https://www.who.int/groups/technical-advisory-group-for-measurement-monitoring-and-evaluation-of-the-un-decade-of-healthy-ageing [accessed 2025-01-16]
17. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
18. Cavill NA, Foster CE. Enablers and barriers to older people's participation in strength and balance activities: a review of reviews. JFSF 2018 Jun;03(02):105-113. [doi: 10.22540/jfsf-03-105]
19. Walters K, Frost R, Kharicha K, Avgerinou C, Gardner B, Ricciardi F, et al. Home-based health promotion for older people with mild frailty: the HomeHealth intervention development and feasibility RCT. Health Technol Assess 2017 Dec;21(73):1-128 [FREE Full text] [doi: 10.3310/hta21730] [Medline: 29214975]
20. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. Lancet Neurol 2019 Nov;18(11):998-1008. [doi: 10.1016/S1474-4422(19)30285-6] [Medline: 31521532]

XSL•FO
RenderX

21. Tsekoura M, Billis E, Tsepis E, Dimitriadis Z, Matzaroglou C, Tyllianakis M, et al. The effects of group and home-based exercise programs in elderly with sarcopenia: a randomized controlled trial. J Clin Med 2018 Nov 26;7(12):480 [FREE Full text] [doi: 10.3390/jcm7120480] [Medline: 30486262]
22. Yuan S, Larsson SC. Epidemiology of sarcopenia: prevalence, risk factors, and consequences. Metabolism 2023 Jul;144:155533 [FREE Full text] [doi: 10.1016/j.metabol.2023.155533] [Medline: 36907247]
23. Hong J, Kim J, Kim SW, Kong H. Effects of home-based tele-exercise on sarcopenia among community-dwelling elderly adults: body composition and functional fitness. Exp Gerontol 2017 Jan;87(Pt A):33-39. [doi: 10.1016/j.exger.2016.11.002] [Medline: 27838369]
24. Bagheri R, Moghadam BH, Church DD, Tinsley GM, Eskandari M, Moghadam BH, et al. The effects of concurrent training order on body composition and serum concentrations of follistatin, myostatin and GDF11 in sarcopenic elderly men. Exp Gerontol 2020 May;133:110869. [doi: 10.1016/j.exger.2020.110869] [Medline: 32035222]
25. Liao C, Tsauo J, Lin L, Huang S, Ku J, Chou L. Effects of elastic resistance exercise on body composition and physical capacity in older women with sarcopenic obesity: a CONSORT-compliant prospective randomized controlled trial. Medicine (Baltimore) 2017;96(23):e7115. [doi: 10.1097/md.0000000000007115]
26. Huang S, Ku J, Lin L, Liao C, Chou L, Liou T. Body composition influenced by progressive elastic band resistance exercise of sarcopenic obesity elderly women: a pilot randomized controlled trial. Eur J Phys Rehabil Med 2017 Aug;53(4):556-563. [doi: 10.23736/s1973-9087.17.04443-4]
27. Schulte J, Yarasheski K. Effects of resistance training on the rate of muscle protein synthesis in frail elderly people. Int J Sport Nutr Exerc Metab 2001 Dec;11 Suppl:S111-S118. [doi: 10.1123/ijsnem.11.s1.s111] [Medline: 11915909]
28. Verdijk LB, Gleeson BG, Jonkers RAM, Meijer K, Savelberg HHCM, Dendale P, et al. Skeletal muscle hypertrophy following resistance training is accompanied by a fiber type-specific increase in satellite cell content in elderly men. J Gerontol A Biol Sci Med Sci 2009 Mar 05;64(3):332-339 [FREE Full text] [doi: 10.1093/gerona/gln050] [Medline: 19196907]
29. Smilios I, Pilianidis T, Karamouzis M, Parlavantzas A, Tokmakidis S. Hormonal responses after a strength endurance resistance exercise protocol in young and elderly males. Int J Sports Med 2007 May;28(5):401-406. [doi: 10.1055/s-2006-924366] [Medline: 17024619]
30. Burtscher J, Soltany A, Visavadiya NP, Burtscher M, Millet GP, Khoramipour K, et al. Mitochondrial stress and mitokines in aging. Aging Cell 2023 Feb 15;22(2):e13770 [FREE Full text] [doi: 10.1111/acel.13770] [Medline: 36642986]
31. Cornish SM, Chilibeck PD. Alpha-linolenic acid supplementation and resistance training in older adults. Appl Physiol Nutr Metab 2009 Feb;34(1):49-59. [doi: 10.1139/h08-136]
32. Beurskens R, Gollhofer A, Muehlbauer T, Cardinale M, Granacher U. Effects of heavy-resistance strength and balance training on unilateral and bilateral leg strength performance in old adults. PLoS One 2015 Feb 19;10(2):e0118535 [FREE Full text] [doi: 10.1371/journal.pone.0118535] [Medline: 25695770]
33. Ruiz JR, Sui X, Lobelo F, Morrow JR, Jackson AW, Sjöström M, et al. Association between muscular strength and mortality in men: prospective cohort study. BMJ 2008 Jul 01;337(7661):a439-a439 [FREE Full text] [doi: 10.1136/bmj.a439] [Medline: 18595904]
34. Nunes EA, Stokes T, McKendry J, Currier BS, Phillips SM. Disuse-induced skeletal muscle atrophy in disease and nondisease states in humans: mechanisms, prevention, and recovery strategies. Am J Physiol Cell Physiol 2022 Jun 01;322(6):C1068-C1084 [FREE Full text] [doi: 10.1152/ajpcell.00425.2021] [Medline: 35476500]
35. Sousa-Victor P, García-Prat L, Muñoz-Cánoves P. Control of satellite cell function in muscle regeneration and its disruption in ageing. Nat Rev Mol Cell Biol 2022 Mar 18;23(3):204-226. [doi: 10.1038/s41580-021-00421-2] [Medline: 34663964]
36. Coletti C, Acosta GF, Keslacy S, Coletti D. Exercise-mediated reinnervation of skeletal muscle in elderly people: an update. Eur J Transl Myol 2022 Feb 28;32(1):10416 [FREE Full text] [doi: 10.4081/ejtm.2022.10416] [Medline: 35234025]
37. Gadelha AB, Paiva FML, Gauche R, de Oliveira RJ, Lima RM. Effects of resistance training on sarcopenic obesity index in older women: a randomized controlled trial. Arch Gerontol Geriatr 2016 Jul;65:168-173. [doi: 10.1016/j.archger.2016.03.017] [Medline: 27057600]
38. Philp A, Hamilton DL, Baar K. Signals mediating skeletal muscle remodeling by resistance exercise: PI3-kinase independent activation of mTORC1. J Appl Physiol (1985) 2011 Feb;110(2):561-568 [FREE Full text] [doi: 10.1152/japplphysiol.00941.2010] [Medline: 21071597]
39. Keller K, Engelhardt M. Strength and muscle mass loss with aging process. Age and strength loss. Muscle Ligaments Tendons J 2019 Jan;03(04):346-350. [doi: 10.32098/mltj.04.2013.17]
40. Bruyere O, Wuidart M, Di Palma E, Gourlay M, Ethgen O, Richy F, et al. Controlled whole body vibration to decrease fall risk and improve health-related quality of life of nursing home residents. Arch Phys Med Rehabil 2005 Feb;86(2):303-307. [doi: 10.1016/j.apmr.2004.05.019] [Medline: 15706558]
41. Liao C, Tsauo J, Huang S, Ku J, Hsiao D, Liou T. Effects of elastic band exercise on lean mass and physical capacity in older women with sarcopenic obesity: a randomized controlled trial. Sci Rep 2018 Feb 02;8(1):2317 [FREE Full text] [doi: 10.1038/s41598-018-20677-7] [Medline: 29396436]
42. Hruda KV, Hicks AL, McCartney N. Training for muscle power in older adults: effects on functional abilities. Can J Appl Physiol 2003 Apr 01;28(2):178-189. [doi: 10.1139/h03-014] [Medline: 12825328]

XSL•FO
RenderX

43. Boiko Ferreira LH, Schoenfeld BJ, Smolarek AC, McAnulty SR, Mascarenhas LPG, Souza Junior TP. Effect of 12 weeks of resistance training on motor coordination and dynamic balance of older woman. Rejuvenation Res 2021 Jun 01;24(3):191-197. [doi: 10.1089/rej.2020.2339] [Medline: 33115319]
44. Tuan S, Chang L, Sun S, Li CH, Chen G, Tsai Y. Assessing the clinical effectiveness of an exergame-based exercise training program using Ring Fit Adventure to prevent and postpone frailty and sarcopenia among older adults in rural long-term care facilities: randomized controlled trial. J Med Internet Res 2024 Jul 18;26:e59468 [FREE Full text] [doi: 10.2196/59468] [Medline: 39024000]
45. Flor-Rufino C, Barrachina-Igual J, Pérez-Ros P, Pablos-Monzó A, Martínez-Arnau FM. Resistance training of peripheral muscles benefits respiratory parameters in older women with sarcopenia: randomized controlled trial. Arch Gerontol Geriatr 2023 Jan;104:104799 [FREE Full text] [doi: 10.1016/j.archger.2022.104799] [Medline: 36070636]
46. Lai X, Zhu H, Wu Z, Chen B, Jiang Q, Du H, et al. Dose-response effects of resistance training on physical function in frail older Chinese adults: a randomized controlled trial. J Cachexia Sarcopenia Muscle 2023 Dec 24;14(6):2824-2834 [FREE Full text] [doi: 10.1002/jcsm.13359] [Medline: 37875291]
47. Smart TFF, Doleman B, Hatt J, Paul M, Toft S, Lund JN, et al. The role of resistance exercise training for improving cardiorespiratory fitness in healthy older adults: a systematic review and meta-analysis. Age Ageing 2022 Jun 01;51(6):afac143 [FREE Full text] [doi: 10.1093/ageing/afac143] [Medline: 35737600]
48. Wang DX, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. J Cachexia Sarcopenia Muscle 2020 Feb;11(1):3-25 [FREE Full text] [doi: 10.1002/jcsm.12502] [Medline: 31788969]

Abbreviations

30SACT: 30-Second Arm Curl Test
30SSRT: 30-Second Sitting-to-Rising Test
6MWT: 6-Minute Walk Test
ADL: activities of daily living
BBS: Berg Balance Scale
BFP: body fat percentage
BIA: bioelectrical impedance analysis
EPT: extension peak torque
ETP: extension total power
IADL: Instrumental Activities of Daily Living
IRG: in-person rehabilitation group
RCT: randomized controlled trial
RPE: rating of perceived exertion
SMI: skeletal muscle mass index
TRG: telerehabilitation group
TSM: total skeletal muscle mass
TUGT: Timed Up-and-Go Test

Edited by A Mavragani; submitted 26.10.24; peer-reviewed by O Egunlae, Y Elsanousi; comments to author 11.12.24; revised version received 30.12.24; accepted 15.01.25; published 24.01.25.

Please cite as:
Zhang L, Ge Y, Zhao W, Shu X, Kang L, Wang Q, Liu Y
A 4-Week Mobile App–Based Telerehabilitation Program vs Conventional In-Person Rehabilitation in Older Adults With Sarcopenia: Randomized Controlled Trial
J Med Internet Res 2025;27:e67846
URL: https://www.jmir.org/2025/1/e67846
doi:10.2196/67846
PMID:

©Lu Zhang, Ying Ge, Wowa Zhao, Xuan Shu, Lin Kang, Qiumei Wang, Ying Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN

XSL•FO
RenderX

1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigating the Efficacy of the Web-Based Common Elements Toolbox (COMET) Single-Session Interventions in Improving UK University Student Well-Being: Randomized Controlled Trial

Jeffrey Lambert[1*], PhD; Maria Loades[2*], PhD, DClinPsy; Noah Marshall[1], BSc, MSc, MRes; Nina Higson-Sweeney[2], BSc, MSc, PhD; Stella Chan[3], DPhil, DClinPsy; Arif Mahmud[4], BA, MA, PhD; Victoria Pile[5], BA, MA, PhD, DClinPsy; Ananya Maity[2], BA, MSc; Helena Adam[2], BSc, MSc; Beatrice Sung[2], BSc; Melanie Luximon[2]; Keren MacLennan[3,6], BSc, PhD; Clio Berry[7], BA, PhD; Paul Chadwick[2], PhD

[1]Department for Health, Universtiy of Bath, Bath, United Kingdom

[2]Department of Psychology, University of Bath, Bath, United Kingdom

[3]School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom

[4]School of Education, University of Roehampton, London, United Kingdom

[5]Department of Psychology, Kings College London, London, United Kingdom

[6]Department of Psychology, University of Durham, Durham, United Kingdom

[7]Brighton and Sussex Medical School, Brighton, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Maria Loades, PhD, DClinPsy
Department of Psychology
University of Bath
Claverton Down
Bath, BA2 7AY
United Kingdom
Phone: 44 1225 385249
Email: m.e.loades@bath.ac.uk

Abstract

Background: Mental health problems in university students are associated with many negative outcomes, yet there is a gap between need and timely access to help. Single-session interventions (SSIs) are designed to be scalable and accessible, delivering core evidence-based intervention components within a one-off encounter.

Objective: COMET (Common Elements Toolbox) is an online self-help SSI that includes behavioral activation, cognitive restructuring, gratitude, and self-compassion. COMET has previously been evaluated in India, Kenya, and the United States with promising results. This study tests the acceptability, appropriateness, perceived utility, and efficacy of COMET among UK university students during the peripandemic period.

Methods: We conducted a randomized controlled trial evaluating the efficacy of COMET compared with a control group, with 2- and 4-week follow-ups. Outcome variables were subjective well-being, depression severity, anxiety severity, positive affect, negative affect, and perceived stress. We also measured intervention satisfaction immediately after completion of COMET. All UK university students with access to the internet were eligible to participate and were informed of the study online. The data were analyzed using linear mixed models and reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.

Results: Of the 831 people screened, 468 participants were randomized to a condition, 407 completed the postintervention survey, 147 returned the 2-week follow-up survey, 118 returned the 4-week follow-up survey, and 89 returned both. Of the 239 randomized, 212 completed COMET. Significant between-group differences in favor of the COMET intervention were observed at 2-week follow-ups for subjective well-being (Warwick-Edinburgh Mental Well-Being Scale; mean difference [MD] 1.39, 95% CI 0.19-2.61; P=.03), depression severity (9-item Patient Health Questionnaire; MD –1.31, 95% CI –2.51 to –0.12; P=.03), and perceived stress (4-item Perceived Stress Scale; MD –1.33, 95% CI –2.10 to –0.57; P<.001). Overall, participants were satisfied with COMET, with the majority endorsing the intervention and its modules as acceptable, appropriate, and exhibiting high utility.

XSL•FO
RenderX

The self-compassion module was most often reported as the participants' favorite module and the behavioral activation module was their least favorite. Qualitative analysis revealed that participants found COMET generally accessible, but too long, and experienced immediate and long-term beneficial effects.

Conclusions: This study demonstrated high engagement with the COMET intervention, along with preliminary short-term efficacy. Almost all participants completed the intervention, but study attrition was high. Participant feedback indicated a high level of overall satisfaction with the intervention, with perceived accessibility, immediate benefits, and potential long-term impact being notable findings. These findings support the potential value of COMET as a mental health intervention and highlight important areas for further improvement.

Trial Registration: ClinicalTrials.gov NCT05718141; https://clinicaltrials.gov/ct2/show/NCT05718141

(J Med Internet Res 2025;27:e58164) doi:10.2196/58164

KEYWORDS

Common Elements Toolbox; mental well-being; online interventions; single-session interventions; university students

Introduction

Mental health problems are common in young people, with at least one in five 16- to 24-year-olds being affected, particularly females [1,2]. University students represent a substantial portion of this population and are at high risk of developing mental health problems due to the unique stressors they may face, such as living away from home for the first time, financial hardship, balancing studying with other responsibilities, and changing social relationships [3,4]. A recent meta-analysis that investigated the mental health of university students found a pooled prevalence rate of 25% for depression and 14% for suicide-related outcomes (eg, suicidal ideation, suicide attempts) [5], with other studies finding that 35% of first-year students reported symptoms indicative of a lifetime mental health disorder [6]. For university students, mental health problems are associated with negative academic outcomes including lower grades [7] and increased dropout rates [8]. Even subthreshold depression and anxiety can lead to substantial impairment [9]. If left untreated, these mental health disorders can increase the risk of more severe mental health and physical health problems developing across the lifespan [10,11]. Early interventions are therefore vital in reducing disease burden on a societal level and are highlighted by students as a priority for mental health and well-being support [12].

The restrictions imposed during the COVID-19 pandemic further exacerbated the global mental health burden [13], particularly in young adults and university students [14,15], while also limiting access to informal and formal support systems. There is concerning evidence that only 20% of UK students struggling with their mental health during the pandemic sought help [16]. Even before the pandemic, only an average of 1 in 3 students experiencing psychological distress used mental health services [17]. Although universities in the United Kingdom do provide internal psychosocial support services for students, such as cognitive behavioral therapy (CBT) and counseling, there are numerous barriers to help-seeking, including self-reliance, poor mental health literacy, feeling too uncertain or unwell to seek help, lack of knowledge on how to access university services, preferences for alternative support (eg, online), and stigma [18-21]. Thus, there is a substantial gap between need and treatment access in this population. Even when support has been accessed, disengaging before completing the full course of treatment is common, meaning that some do not receive the full dose and optimal benefits [22].

Digital interventions (eg, online CBT) are one way of overcoming barriers students may experience in accessing mental health support by addressing concerns regarding anonymity, privacy, accessibility, and stigma [20]. Digital interventions also have a growing evidence base supporting their efficacy in addressing mental health problems [23,24]. For example, a recent network meta-analysis found that guided self-help interventions (including digital support) were more effective in reducing depression than waitlist control conditions [25]. Another meta-analysis, which compared 10 digital multisession interventions of a mean duration of 4 weeks with active and waitlist controls in university students, found small but significant improvements in psychological well-being after the intervention [26]. However, despite the promise of digital support, engagement is less than optimal due to factors such as lack of time and interest [27]. One way to address poor engagement with digital interventions is through online single-session interventions (SSIs). SSIs have the advantage of being more scalable and accessible because they are designed to deliver the core components of an active intervention within a one-off encounter, without an expectation that an individual will engage in longer-term therapy. Thus, SSIs could be a useful and effective addition to the suite of therapeutic options offered by university student services, which tend to be longer courses of treatment [28-34].

One such example is COMET (Common Elements Toolbox), an online (web-based) SSI without therapist contact. COMET was originally designed by Professor Rob de Rubeis' team [35] for adolescents in India, and subsequently adapted and evaluated in US college students. COMET includes 4 modules based on evidence-based principles, namely, (1) behavioral activation (BA); (2) cognitive restructuring; (3) gratitude, from the discipline of positive psychology [36]; and (4) self-compassion. Versions of COMET have been developed with Kenyan and Indian adolescents and tested with US graduate students during the pandemic [35,37,38]. These versions are acceptable and useful, with US postgraduate students reporting pre- to post-program improvements in their perceived ability to manage the personal and psychological impacts of objective conditions or events (secondary control [39]). However, this online SSI has yet to be implemented and evaluated in a randomized

XSL•FO
RenderX

controlled trial (RCT) among university students in the United Kingdom, with potential cultural differences meaning that existing findings cannot simply be extrapolated into the UK context. Additionally, UK higher education is typically shorter and less expensive than in the United States, and UK students apply for specific courses, whereas US students can switch majors during their university years.

We aimed to test the efficacy of COMET, an online mental health SSI, in undergraduate and postgraduate university students. Specifically, we sought to address the following questions:

- Compared with attention control, does COMET improve the mental health and well-being of university students at 2- and 4-week follow-ups?
- Do demographic variables (ie, age or gender) or clinical variables (ie, baseline depression severity, anxiety severity, mental health diagnoses, or treatment status) moderate the efficacy of COMET at 2- and 4-week follow-ups?
- How do participants perceive the acceptability and appropriateness of COMET, how do their utility ratings compare across the 4 COMET modules, and which modules do they prefer?

Methods

Trial Design

The study was a 2-arm, individually (1:1) randomized controlled trial design, comparing short-term (2- and 4-week) mental health outcomes of UK university students exposed to COMET (ie, the intervention group) or an attention-control questionnaire (ie, the control group). A nontherapeutic control was chosen to reflect the fact that SSIs are often provided as an alternative to receiving no support or being put on a waiting list for another intervention. The study was reported in accordance with CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth; Multimedia Appendix 1) recommendations for the reporting of RCTs of eHealth interventions [40] and the TIDieR (Template for Intervention Description and Replication) recommendations on the reporting of behavior change interventions [41]. The trial was registered at ClinicalTrials.gov (NCT05718141) toward the end of data collection.

Important Changes to Methods After Trial Commencement

We identified a Qualtrics error in January 2022, with participants who had completed the baseline survey not receiving the automatic 2- and 4-week follow-up surveys. This error was resolved in February 2022. We emailed a final follow-up survey link to the 36 participants who had completed the intervention before resolution and were outside the 4-week follow-up window by 1 week to 2 months. We also emailed survey links to the 45 participants who had completed the intervention but were still within the 2-week follow-up window and the 37 participants within the 4-week follow-up window. The follow-up data from the 36 participants who were 1 week to 2 months outside of the 4-week follow-up window were excluded from the main analyses, with these participants reflected in attrition rates.

Participants

The study was set in the nonclinical, peripandemic context of UK universities and recruited currently registered undergraduate and postgraduate students with internet access. Those without internet access and those younger than 16 years were not eligible. The decision to offer the intervention more widely rather than just within clinical settings was to minimize barriers to access.

During the recruitment phase, potential participants were informed of the study through study adverts shared via social media platforms (eg, Facebook, Twitter, Instagram, TikTok), university-held mailing lists, and mailing lists/newsletters of charities and organizations with an interest in student mental health, such as Student Minds. The study was also advertised via psychology research participation schemes at the University of Bath and the University of Reading, which provide students with credits in exchange for taking part in research studies. It was also promoted on research recruitment websites such as MQ Mental Health Participate and Call for Participants. All adverts directed participants to a Qualtrics survey (Silver Lake Technology Management, L.L.C.) where they could learn more information about the study and take part.

Using power calculations based on the effect size of a previous iteration of COMET on the 9-item Patient Health Questionnaire (PHQ-9) [38], to detect a small effect of d=0.3, we required 378 participants to complete follow-up. A previous RCT of an SSI in adolescents had an attrition rate of around 28% at 3 months [32]. However, given our shorter follow-up rate of 4 weeks, we raised the recruitment target to 473 to allow for 20% attrition.

Measures and Materials

All study documentation, including the information sheet, consent form, baseline assessment survey, experimental conditions, posttreatment survey, and follow-up surveys, was accessed through the Qualtrics platform. Brief demographic information was collected by self-report, including age in years, gender identity (female, male, or other), sexual orientation (heterosexual, bisexual, homosexual, other, or unlisted), and ethnicity (White or White British, Asian or Asian British, Black or Black British, or mixed). Participants were also asked about their mental health in relation to diagnoses, past and current experiences, and treatment status.

Measures of Mental Health and Well-Being

For all participants, mental health and well-being were assessed at 3 time points, including a baseline assessment pretreatment, a 2-week follow-up, and a 4-week follow-up. Several dimensions of mental health and well-being were assessed, including (1) subjective well-being, (2) depression severity, (3) anxiety severity, (4) positive affect, (5) negative affect, and (6) perceived stress.

The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS), a commonly used measure of well-being [42], was used in this study. The WEMWBS has 14 items that capture participants' feelings and thoughts that best describe their

XSL•FO
RenderX

experience over the previous 2 weeks using a scale from 1 (none of the time) to 5 (all of the time). The WEMWBS has robust psychometric properties [43]. The WEMWBS demonstrated good reliability, with a Cronbach α of 0.88 at baseline, 0.82 at the 2-week follow-up, and 0.83 at the 4-week follow-up.

The PHQ-9, a commonly used measure for depressive symptoms [44], was also used in this study. The 9 items of the questionnaire (α=.84) capture the frequency of depressive symptoms over the preceding 2 weeks using a scale from 0 to 3. A total score of 0-4 indicates no depression, 5-9 indicates mild depression, 10-14 indicates moderate depression, 15-19 indicates moderately severe depression, and 20-24 indicates severe depression. The PHQ-9 has a sensitivity and specificity of 88% for detecting clinical depression [45]. The PHQ-9 has demonstrated good reliability, with a Cronbach α of 0.84 at baseline, 0.89 at the 2-week follow-up, and 0.86 at the 4-week follow-up.

The 7-item General Anxiety Disorder (GAD-7) checklist is a commonly used measure for evaluating symptoms of anxiety [46]. The 7 items of the checklist (α=.87) capture the frequency of anxious symptoms over the preceding 2 weeks using a scale from 0 to 3. A total score of 0-4 indicates no anxiety, 5-9 indicates mild anxiety, 10-14 indicates moderate anxiety, and ≥15 indicates severe anxiety. The GAD-7 has a sensitivity and specificity of 89% and 82% respectively [46]. The GAD-7 demonstrated good reliability, with a Cronbach α of 0.87 at baseline, 0.85 at the 2-week follow-up, and 0.84 at the 4-week follow-up.

The Positive and Negative Affect Schedule (PANAS) is a commonly used measure of participants' affective states [47]. This scale includes two 10-item subscales measuring positive affect and negative affect on a 5-point Likert scale (1=very slightly or not at all and 5=extremely). The PANAS has been shown to be reliable and valid [48]. The PANAS demonstrated good reliability for positive affect, with a Cronbach α of 0.88 at baseline, 0.90 at the 2-week follow-up, and 0.91 at the 4-week follow-up. The PANAS also demonstrated good reliability for negative affect, with a Cronbach α of 0.87 at baseline, 0.85 at the 2-week follow-up, and 0.87 at the 4-week follow-up.

The 4-item Perceived Stress Scale (PSS-4 [49]) is an abbreviated, 4-item scale designed to measure the extent of perceived stress in individuals' lives over 4 weeks. However, to better align with our assessment intervals, this scale was adapted to specifically assess perceived stress over 2 weeks. The PSS-4 demonstrated good reliability, with a Cronbach α of 0.78 at baseline, 0.84 at the 2-week follow-up, and 0.80 at the 4-week follow-up.

Measures of Intervention Satisfaction

Immediately after completing COMET, satisfaction with the intervention was assessed within a posttreatment survey using measures of (1) perceived appropriateness (perceived fit, relevance, or compatibility of treatment to address a particular issue or problem), (2) perceived acceptability (perception that a given treatment is agreeable, palatable, or satisfactory), (3) perceived utility, and (4) module preferences. These measures were added after trial registration because, as the study progressed, it became evident that the initial measures might not fully capture participants' experiences with COMET. Therefore, we aimed to obtain a more detailed understanding of the intervention's acceptability and align with best practice guidance, which emphasizes the importance of ongoing assessment and refinement in accurately evaluating complex interventions.

Intervention Appropriateness Measure is a 4-item measure assessing intervention appropriateness. Each item has a 5-point scale ranging from 1 (completely disagree) to 5 (completely agree) [47]. The Intervention Appropriateness Measure demonstrated good reliability, with a Cronbach α of 0.90.

Acceptability of Interventions Measure is a 4-item measure assessing intervention acceptability. Each item has a 5-point scale ranging from 1 (completely disagree) to 5 (completely agree) [47]. The Acceptability of Interventions Measure demonstrated good reliability, with a Cronbach α of 0.92.

Perceived utility is a bespoke measure where participants in the intervention group were asked to rate their feelings toward each module concerning helpfulness, engagement, and intention to apply intervention techniques in their daily lives. These questions were measured on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree), with 3 items per module. Across each of the 4 COMET modules, perceived utility also demonstrated good reliability, with Cronbach α values ranging from 0.90 to 0.94.

Participants indicated their preferences toward the 4 modules with the prompt questions asked: "Which exercise was your favorite?" and "Which exercise was your least favorite?." All participants in the intervention group were also asked to complete a free textbox asking about their experiences with COMET.

Procedure

Interested students were directed to a web-based Qualtrics survey including an information sheet, a consent form, a baseline assessment survey, descriptions of the experimental conditions, a posttreatment survey, and a debrief sheet. The information sheet explained the purposes of the research and the process of data collection and management. Following completion of the consent form, participants were directed to complete a baseline assessment survey measuring participants' mental health and well-being. Participants were then randomly assigned to the intervention or control condition using the automated simple randomization tool embedded within Qualtrics. Thus, the research team was blind to treatment allocation. However, due to the intervention's nature, the actual treatment assignment was not concealed from participants.

In addition to the preintervention measures, participants in the control group were also asked to complete additional questionnaires at baseline, which acted as an attention control for COMET.

XSL•FO
RenderX

Intervention

Overview

Participants randomized to the intervention condition received and completed COMET, an online self-guided SSI. The intervention was accessible via any device which could connect to the internet, without any need to register or download software. It was based on the core principles of CBT, combined with principles from positive psychology. All 4 COMET modules were designed to be completed in a single session, taking 60-75 minutes. The modules featured short reading exercises, informational videos, and writing tasks.

Behavioral Activation

In this module, participants could identify and reflect on activities that were important to them, list activities they found enjoyable and meaningful, reflect on why these activities mattered to them, and schedule in time to perform these activities in the weeks ahead.

Cognitive Restructuring

In this module, participants were invited to identify and reframe negative beliefs. They were first asked to read about a hypothetical character who is adjusting to changes in their routine. Then, using the character's story as an example, they were asked to try to identify negative beliefs that the character may have been experiencing and ways the character could reframe the belief. They could then apply this technique to a situation in their own life.

Gratitude

In this module, participants could reflect and write about 3 things they were grateful for. They were then asked to think and write about things they noticed around them that they enjoyed and were grateful for.

Self-Compassion

In this final module, participants were asked to write a self-compassion letter to themselves, expressing compassion toward themselves just as they would toward a friend or family member. They were also requested to create a few sentences that they would like to hear when feeling self-critical.

Further rationale for all 4 modules can be found in previous publications [35,38]. Participants could only access the intervention once. Outcome measures were collected immediately after they completed COMET.

Attention Control

Participants allocated to the attention control group were asked to complete 5 additional measures, including a Symptom Importance Rating Questionnaire, the Chalder Fatigue Questionnaire [48], the Pittsburgh Sleep Quality Index [50], the Snaith-Hamilton Pleasure Scale [51], and the Fatigue Associated With Depression Scale [52]. These measures are not reported in this study. These were not completed by the intervention group nor reported as main outcomes on the RCT and will be reported elsewhere.

Immediately upon completion of these conditions, a posttreatment survey collected data related to their satisfaction with the COMET (only for the intervention group) and brief demographic information (for both groups). A debrief sheet signposted information about sources of mental health and well-being support. The initial intervention modules and assessment measures were designed to take participants approximately 60-75 minutes to complete.

Emails were sent out at 2 and 4 weeks after the intervention asking participants to complete follow-up assessment measures, which were designed to take approximately 10-15 minutes at each time point. All assessments took place online. Those who completed both follow-up assessments were able to opt into a prize draw for 1 of 16 £50 (US $63) Amazon vouchers. Additionally, those who completed the study as part of a student research participation scheme could earn course credits for taking part.

As a result of the pseudo-anonymous nature of the study, distress management was based on signposting, without direct or personal contact from the research team. It was the responsibility of the participants to decide whether they acted on this advice. If a participant scored >0 on the PHQ-9 item which asks about suicidal ideation (item 9), the participant saw an additional pop-up box suggesting that they may want to seek extra help, with a list of potential sources and contact details. They were reminded that the research team will not routinely monitor the answers to these questions. However, these participants were still included in the study.

Ethical Considerations

All research was performed in accordance with relevant guidelines and regulations set by the Declaration of Helsinki. Ethical approval was granted by the University of Bath Psychology Research Ethics Committee (reference 21-212). Reciprocity was also granted by the University of Reading and Kings College London. All participants provided informed consent.

Analysis Plan

All quantitative data were analyzed descriptively overall and by each arm. Continuous data were summarized using means and SDs. Categorical data were presented using frequencies and percentages. Data on feasibility (ie, recruitment, intervention engagement, and outcome completion rates) and acceptability (ie, intervention acceptability and appropriateness) were reported along with baseline characteristics in the 2 trial arms. All outcome data were analyzed using linear mixed models, adjusted for individual-level variation in baseline measures.

To address the first research question, intervention and control groups were compared based on complete case data at (1) the 2-week follow-up and (2) the 4-week follow-up using intention to treat (ITT). Between-group differences are presented as adjusted mean differences (MDs) and 95% CIs. Effect sizes were also calculated for the results of this study. For the second research question, exploratory moderation analyses were conducted to determine if baseline characteristics moderated the relationship between group assignment (intervention or control group) and each outcome measure. Age, anxiety, and depression were not recoded into categories. Gender, diagnoses, and current treatment were dummy-coded into binary values.

XSL•FO
RenderX

To address the third research question, ratings of acceptability and perceived utility (at postintervention) were summarized using means and SDs. Sensitivity analysis using multiple imputations was conducted to assess the likely impact of missing data. Data were first analyzed using ITT (ie, all participants randomized who provided follow-up data). The data were then analyzed using ITT with all imputed data.

We analyzed the free textbox data using principles from inductive content analysis [53] in conjunction with Braun et al's [54] method for analyzing qualitative data collected from online surveys. Specifically, 2 research team members (BS and ML) independently familiarized themselves with all the qualitative data by reading and re-reading the responses. Next, BS and ML independently focused on 15-20 responses and generated initial codes. These initial codes were then applied to the remaining free-text responses, utilizing constant comparison to determine whether additional codes were warranted for different meanings. After coding the data separately, BS and ML convened to compare their codes and discuss patterns in the responses. Through discussion and iterative review of the qualitative data, with supervisory input from M Loades, they refined and finalized the themes. Illustrative quotes were then selected to exemplify each theme.

Results

Participant Flow

The participant flow is shown in Figure 1 (also see Multimedia Appendix 1). Participant recruitment and follow-up took place between September 2021 and December 2022. Of the 831 individuals assessed, 468 completed baseline measures and were randomized to either the COMET intervention (n=239) or the attention control (n=229). Of the 239 that were randomized and initiated COMET, 212 (88.7%) completed COMET, with 203 (84.9%) completing the posttreatment survey. A total of 213 out of 229 (93%) individuals completed the attention control, with 204 (89.1%) completing the postattention control survey. Of those randomized, 147 participants completed the 2-week follow-up survey, 118 completed the 4-week follow-up survey, and 89 participants completed both follow-up surveys.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram illustrating the progression of university students through the different phases of the randomized controlled trial. CCA: complete case analysis.

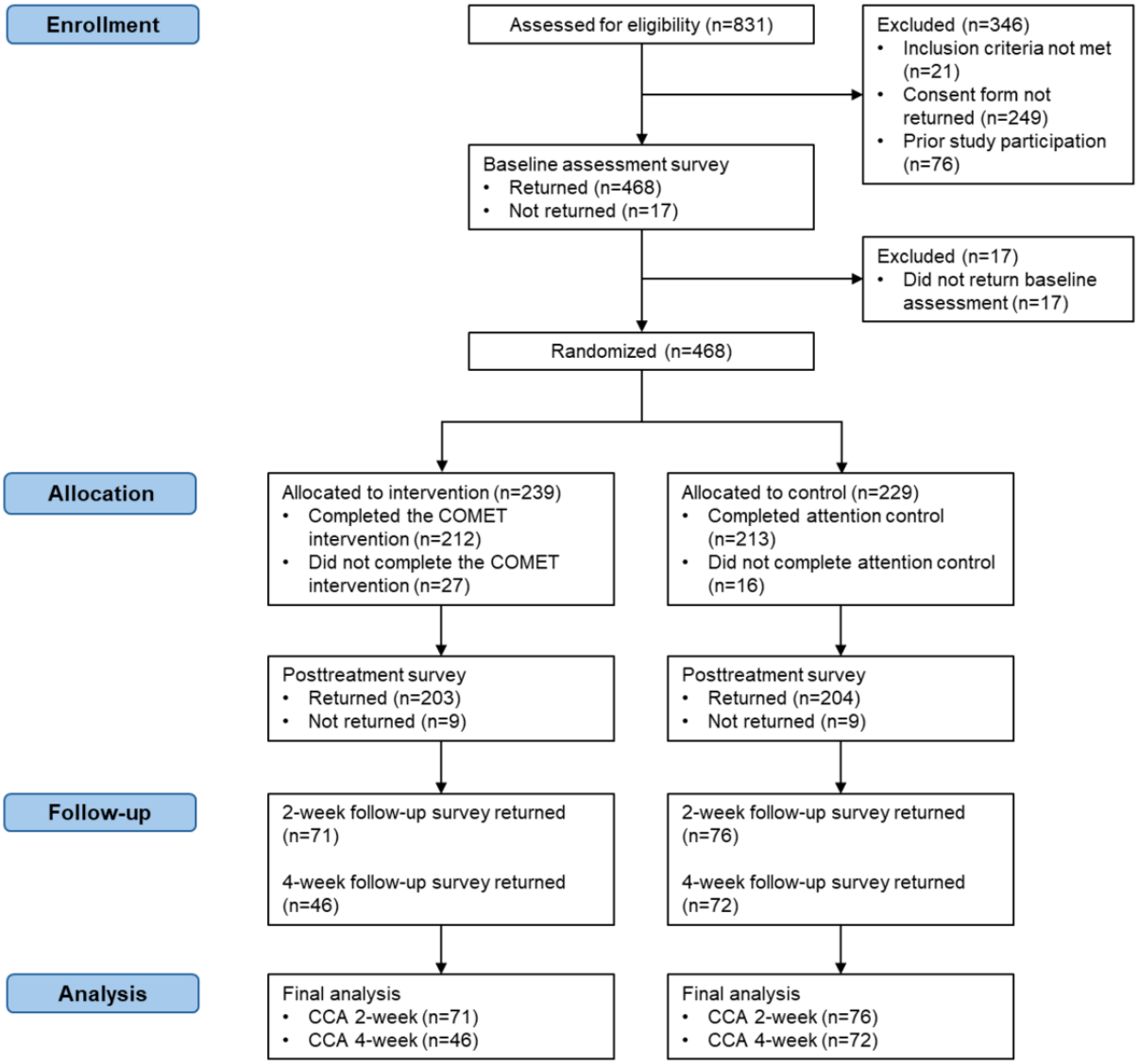

Baseline Data

The mean age of participants was 22.49 (SD 7.28) years, with most participants identifying as female (339/407, 83.3%), heterosexual (296/407, 72.7%), and White or White British (272/407, 66.8%); 126 of 407 (31%) participants had been previously diagnosed with a mental health disorder and 85 of 407 (20.9%) were actively receiving treatment for a mental health disorder during the study. Further, 162 of 407 (39.8%) participants had at least moderate depression and 95 of 407 (23.3%) had at least moderate anxiety (Table 1).

XSL•FO
RenderX

Table 1. Baseline demographic and clinical characteristics of university students participating in the randomized controlled trial, randomized to either the COMET[a] intervention group or the waitlist control group.[b]

Characteristics	Intervention (n=203)	Control (n=204)	Total sample (n=407)
Age (years), mean (SD)	22.45 (7.13)	22.52 (7.44)	22.49 (7.28)
Gender (female), n (%)	172 (84.7)	167 (81.9)	339 (83.3)
Sexual orientation, n (%)			
Heterosexual	143 (70.4)	153 (75.0)	296 (72.7)
Bisexual	29 (14.3)	27 (13.2)	56 (13.8)
Homosexual	7 (3.4)	5 (2.5)	12 (2.9)
Other or unlisted	24 (11.8)	19 (9.3)	43 (10.6)
Ethnicity, n (%)			
White or White British	137 (67.5)	135 (66.2)	272 (66.8)
Asian or Asian British	12 (5.9)	13 (6.4)	25 (6.1)
Black or Black British	3 (1.5)	5 (2.5)	8 (2.0)
Mixed	51 (25.1)	51 (25.0)	102 (25.1)
Mental health diagnosis, n (%)	66 (32.5)	60 (29.4)	126 (31.0)
Receiving treatment, n (%)	42 (20.7)	43 (21.1)	85 (20.9)
9-item Patient Health Questionnaire ≥10, n (%)	81 (39.9)	81 (39.7)	162 (39.8)
7-item General Anxiety Disorder Checklist ≥10, n (%)	41 (20.2)	54 (26.5)	95 (23.3)

[a]COMET: Common Elements Toolbox.

[b]Demographic information was reported by participants after exposure to their allocated treatment condition. Completed posttreatment surveys were returned by 407 participants across the intervention (n=203) and control (n=204) conditions.

Intervention Outcomes: Primary Analysis

At the 2-week follow-up, the complete-case ITT analysis showed that subjective well-being (WEMWBS) was significantly higher in the intervention group compared with the control group, with an MD of 1.39 (95% CI 0.19-2.61, *P*=.03) and a small effect size (Cohen *d*=0.24). Depression scores were also significantly lower in the intervention group compared with the control group (PHQ-9: MD –1.31, 95% CI –2.51 to –0.12; *P*=.03), with a small effect size (*d*=0.23). Perceived stress was also significantly lower in the intervention group compared with the control group (PSS-4: MD –1.33, 95% CI –2.10 to –0.57; *P*<.001), with a small effect size (*d*=0.37). No significant differences between groups in subjective well-being (*P*=.16), depression (*P*=.08), or perceived stress (*P*=.19) were observed at the 4-week follow-up. For the remaining outcome measures (ie, GAD-7, Negative Affect Scale, and Positive Affect Scale), no between-group differences were observed at the 2- (*P*=.31, *P*=.80, and *P*=.20) or 4-week (*P*=.33, *P*=.37, and *P*=.41) follow-ups (Table 2).

XSL•FO
RenderX

Table 2. Between-group differences in primary and secondary mental health outcomes among university students participating in the randomized controlled trial at the baseline, 2-week, and 4-week follow-up assessments[a].

Outcome	Intervention, mean (SD)	Control, mean (SD)	Between-group difference (95% CI)	*P* value
Subjective well-being (Warwick-Edinburgh Mental Well-Being Scale)				
Baseline	22.04 (4.85)	21.53 (5.17)	N/A[b]	N/A
2 weeks	22.83 (4.37)	21.72 (4.76)	1.39 (0.19 to 2.61)	.03
4 weeks	22.89 (4.39)	21.75 (5.06)	1.01 (–0.39 to 2.42)	.16
Depression severity (9-item Patient Health Questionnaire)				
Baseline	8.30 (5.80)	9.05 (5.76)	N/A	N/A
2 weeks	7.42 (5.27)	8.78 (6.41)	–1.31 (–2.51 to –0.12)	.03
4 weeks	6.80 (5.12)	8.92 (6.34)	–1.39 (–2.92 to 0.15)	.08
Anxiety severity (7-item General Anxiety Disorder Checklist)				
Baseline	6.05 (4.43)	6.73 (4.62)	N/A	N/A
2 weeks	5.44 (3.59)	6.68 (4.80)	–0.51 (–1.49 to 0.47)	.31
4 weeks	4.85 (3.33)	6.40 (4.63)	–0.59 (–1.81 to 0.61)	.33
Negative affect				
Baseline	22.55 (7.33)	23.20 (8.02)	N/A	N/A
2 weeks	21.29 (6.27)	22.76 (8.29)	–0.24 (–2.08 to 1.60)	.80
4 weeks	20.41 (6.77)	22.85 (8.06)	–1.08 (–3.42 to 1.27)	.37
Positive affect				
Baseline	28.02 (8.06)	27.16 (7.71)	N/A	N/A
2 weeks	28.73 (7.73)	26.53 (8.11)	1.33 (–0.70 to 3.36)	.20
4 weeks	28.76 (8.55)	26.89 (7.98)	1.06 (–1.47 to 3.58)	.41
Perceived stress (4-item Perceived Stress Scale)				
Baseline	7.67 (3.27)	7.67 (3.07)	N/A	N/A
2 weeks	6.86 (3.20)	8.04 (3.13)	–1.33 (–2.10 to –0.57)	<.001
4 weeks	6.54 (2.96)	7.42 (3.43)	–0.68 (–1.69 to 0.33)	.19

[a]Analysis includes complete case data for 2 (n=147) and 4 weeks (n=118).

[b]N/A: not applicable.

Intervention Outcomes: Moderation Analysis

For those receiving COMET, individuals who were not currently receiving treatment showed greater improvements in subjective well-being at both 2-week (B=–3.37, 95% CI –6.34 to –0.41; *P*=.03) and 4-week follow-ups (B=–4.11, 95% CI –7.57 to –0.65; *P*=.02) than those individuals who were receiving another treatment.

At the 4-week follow-up, younger individuals also showed greater improvements in depression (B=0.19, 95% CI 0.01-0.37; *P*=.04) and anxiety (B=0.15, 95% CI 0.01-0.30; *P*=.04) compared with their older counterparts.

Compared with those with lower baseline anxiety, individuals with higher baseline anxiety exhibited greater improvements in perceived stress at the 4-week follow-up (B=–0.23, 95% CI –0.45 to –0.01; *P*=.047).

No moderation effects were observed for gender, baseline depression severity, or mental health diagnoses. Please see Multimedia Appendix 2 for full analyses.

Intervention Outcomes: Sensitivity Analyses

Missingness ranged from 0.00% to 68.42% for cases (mean 44.46%, SD 25.28%) and from 0.00% to 73.29% (mean 44.46%, SD 34.93%) for variables. Little's Missing Completely At Random (MCAR) Test was applied and indicated that the data were missing completely at random (χ^2_{100}=99.10, *P*=.51); hence, it was assumed that missingness was purely random and not related to any observed or unobserved data.

Given the high proportion of missing data, multiple imputation was carried out to estimate follow-up outcomes for all participants who did not provide data at the 2- and 4-week follow-ups (Multimedia Appendix 3). Depression and perceived stress were lower in the intervention group compared with the control at the 2-week follow-up but only bordering significance

XSL•FO
RenderX

(PHQ-9: MD –0.53, 95% CI –1.10 to 0.04, *P*=.07; and PSS-4: MD=–0.57, 95% CI –1.15 to 0.01, *P*=.05). No between-group effects between the intervention and control were observed for any of the other variables or follow-up points.

Intervention Acceptability

Overall, participants found COMET to be acceptable, with between 166 out of 203 (81.8%) and 188 out of 203 (92.6%) participants agreeing or strongly agreeing that they approved of, liked, and welcomed COMET and found it appealing. Participants also found COMET to be appropriate with between 162 out of 203 (79.8%) and 188 out of 203 (92.6%) agreeing or strongly agreeing that COMET was fitting, suitable, applicable, and a good match (Table 3). Each of COMET's 4 modules was also perceived to have high utility (Table 4). Most participants at least slightly agreed that BA (169/203, 83.3%, to 181/203, 89.2%), cognitive restructuring (145/203, 71.4%, to 175/203, 86.2%), gratitude (174/203, 85.7%, to 182/203, 89.7%), and self-compassion (150/203, 73.9%, to 174/203, 85.7%) were helpful, engaging, and applicable. Participants most liked the self-compassion module (66/203, 32.5%) followed by gratitude (59/203, 29.1%), cognitive restructuring (55/203, 27.1%), and BA (23/203, 11.3%). Participants least liked the BA module (83/203, 40.9%), followed by self-compassion (51/203, 25.1%), gratitude (38/203, 18.7%), and cognitive restructuring (31/203, 15.3%).

Table 3. Assessment of the acceptability and appropriateness of the COMET[a] intervention among university students (n=203).

Assessment	Agrees or strongly agrees, n (%)	Neutral, n (%)	Disagrees or strongly disagrees, n (%)
Acceptability (Acceptability of Interventions Measure)			
Approve	188 (92.6)	12 (5.9)	3 (1.5)
Appealing	166 (81.8)	27 (13.3)	10 (4.9)
Like	177 (87.2)	21 (10.3)	5 (2.5)
Welcome	177 (87.2)	23 (11.3)	3 (1.5)
Appropriateness (Intervention Appropriateness Measure)			
Fitting	178 (87.7)	23 (11.3)	2 (1.0)
Suitable	188 (92.6)	11 (5.4)	4 (2.0)
Applicable	184 (90.6)	17 (8.4)	2 (1.0)
Good match	162 (79.8)	33 (16.3)	8 (3.9)

[a]COMET: Common Elements Toolbox.

XSL•FO
RenderX

Table 4. Evaluation of the perceived utility of COMET's[a] 4 modules, focusing on whether university students found them helpful, engaging, and applicable to their personal or clinical needs (n=203).

Items of perceived utility by COMET module	Slightly agree, agree, strongly agree, n (%)	Neutral, n (%)	Slightly disagree, disagree, strongly disagree, n (%)
Behavioral activation			
Helpful	181 (89.2)	9 (4.4)	13 (6.4)
Engaging	178 (87.7)	12 (5.9)	13 (6.4)
Applicable	169 (83.3)	15 (7.4)	19 (9.4)
Cognitive restructuring			
Helpful	175 (86.2)	14 (6.9)	14 (6.9)
Engaging	160 (78.8)	17 (8.4)	26 (12.8)
Applicable	145 (71.4)	23 (11.3)	35 (17.2)
Gratitude			
Helpful	182 (89.7)	9 (4.4)	12 (5.9)
Engaging	177 (87.2)	12 (5.9)	14 (6.9)
Applicable	174 (85.7)	9 (4.4)	20 (9.9)
Self-compassion			
Helpful	174 (85.7)	17 (8.4)	12 (5.9)
Engaging	170 (83.7)	18 (8.9)	15 (7.4)
Applicable	150 (73.9)	26 (12.8)	27 (13.3)

[a]COMET: Common Elements Toolbox.

Qualitative Analysis

The free-text content analysis of participants' experiences of COMET resulted in the formulation of 4 categories: time/length, accessibility, immediate effects, and long-term effects. Participants expressed that COMET introduced valuable skills in a manner that was simple and easy to understand, with many participants also noting the immediate effects of increased positivity, relaxation, and self-compassion upon completion. They also reported that the intervention helped to improve their thinking and outlook by rationalizing worries and recognizing the positive aspects of their lives. Participants believed that COMET had the potential for long-term impact, especially in helping them plan positive activities to regain a healthy routine and restructure their thinking. However, participants did reflect that some of the exercises were too long or boring, which could have potentially contributed to study dropout. Several participants also reported technical issues (eg, images not loading), which acted as a potential barrier to engagement. See Multimedia Appendix 4 for full analysis.

Discussion

Principal Findings

UK university students engaged well with the COMET online SSI and exhibited small, significant improvements in well-being, depression severity, and perceived stress over a 2-week follow-up period compared with the control arm. Changes in anxiety severity, positive affect, and negative affect were nonsignificant. Exploratory analysis also revealed that COMET was potentially more effective at reducing stress for those with elevated symptoms of anxiety. We also found that COMET was largely well-received in terms of acceptability, appropriateness, and feasibility, although users commented that it was too long, and some had technical issues.

A high level of intervention completion was observed, with 212 of 239 (88.7%) students randomized to COMET completing the intervention. This rate is particularly noteworthy when compared with completion rates in multiweek digital interventions, where a significant drop-off in engagement is common. For example, a systematic review of digital mental health interventions for depression, anxiety, and well-being in students found that most studies saw high completion rates after module 1 but reduced engagement over time, with only a minority completing all modules [55]. The structure of multiweek interventions may unintentionally contribute to attrition, as participants might find it difficult to commit to long-term engagement due to competing demands, a lack of immediate benefits, or the gradual loss of motivation. By contrast, SSIs such as COMET potentially address this challenge by minimizing the time commitment required, making it easier for participants to complete the program in a single sitting.

While COMET led to improvements in well-being, depression, and perceived stress at the 2-week follow-up, these changes were not observed for anxiety or positive and negative affect and had disappeared by the 4-week follow-up. This contrasts with another SSI RCT in adolescents, which found sustained effects on depression at 3 months [32]. One possible reason for this difference is the inclusion criteria. The previous study included participants based on depression criteria, whereas COMET aimed to be more inclusive by not using such criteria.

XSL•FO
RenderX

In COMET, the mean baseline PHQ-9 score was below 10 (the cutoff for moderate depression), and only 126 of 407 (31.0%) participants had any mental health diagnosis. This may have led to ceiling effects, limiting the potential for long-lasting changes. Moderation analysis further supports this, showing that the effects were more pronounced for those with elevated baseline symptoms and a previous diagnosis. Additionally, the analysis found that older participants benefited less from COMET than younger ones. This could reflect that during our recruitment period between September and December, older participants, likely in later years of study, may experience higher levels of depression and anxiety due to increased pressure related to final-year projects and concerns about future employment compared with first-year students [56].

Our observed effect sizes at week 2 for well-being (d=–0.24), depression (d=0.23), and perceived stress (d=0.37) compare favorably with other online SSIs [30,32,57]. However, after multiple imputations, the effect on well-being became nonsignificant, and the effects on depression and perceived stress were bordering on significance. These effect sizes are smaller than those of more extensive internet-based CBT programs for anxiety and depression, with a previous meta-analysis found to have pooled effects of g=0.51 (95% CI 0.29-0.73) in young adults [58]. However, SSIs such as COMET could make a considerable difference at the population level, given their scalability and potential reach [59]. It is particularly encouraging that participants mentioned in the free text feedback how they could apply their learning from COMET in their daily lives, aligning with student priorities for mental health support [12].

In terms of participant preference, self-compassion was found to be the most liked aspect of COMET, while BA received lower ratings. Several possible reasons were identified to explain this discrepancy. First, BA is typically designed to be delivered over multiple weeks [60,61], including goal setting and regular goal reviews. In the COMET intervention, which consisted of a single session, the absence of goal reviews may have undermined the efficacy of BA. Second, previous studies have found self-compassion to be particularly valuable for students adjusting to university life with its various demands and stressors [62]. Our qualitative findings also suggest that participants found self-compassion to be a helpful coping strategy in this population. Third, COMET had no inclusion criteria for elevated symptoms of depression or anxiety. BA primarily targets depression and may be most suitable for individuals experiencing depressive symptoms. The BASIL trial, for instance, found that participants without depression perceived the relevance of BA to be limited, although this was in older adults [63]. Fourth, while activity monitoring has previously been regarded as beneficial for students, such as during the transition to college [64], the contextual limitations imposed by the pandemic made BA less effective in supporting participants due to the restrictions on engaging in pleasurable activities.

Participant recruitment was challenging, and it took 15 months to recruit the current sample. Online SSIs aimed at adolescents in the United States have recruited considerably larger samples in much shorter periods via social media [32]. A likely explanation for the difference is the guaranteed vouchers as an incentive to participate [65], whereas in our study, they were entered into a prize draw. When recruiting university students with subclinical anxiety or depression symptoms, a study in the Netherlands found that emailing all students via the central student administration was the most effective strategy, as compared with flyers, media, and social media adverts, among others [66]. Similarly, campus-wide recruitment emails resulted in over 80% recruitment of the 651 participants in a US-based trial of a universal, web-based prevention program for anxiety and depression [67]. Unfortunately, we were unable to secure agreement within our institutions to do this in our study. The pandemic context within which we recruited also made it more difficult to recruit using flyers/posters in physical locations on campus, and we, therefore, relied predominantly on social media and undergraduate research participation schemes, although we did also list the study on the research studies section of the student services web page at a UK university.

Once students had signed up to participate, our study had considerable study attrition, like other studies of digital mental health interventions in student populations [68-72]. Our high attrition could be explained by the qualitative feedback provided, which indicated that some participants experienced COMET as too long, with others having technical issues. Other studies have found that students prefer interventions that they can use in short bursts of time [73], and it may be that future iterations of COMET could give choices so that students can choose which components they want to do. This would also add a degree of personalization, which young people say aids their engagement in online interventions [74]. Technical issues are a common barrier to engagement in digital mental health interventions [75].

Strengths and Limitations

We evaluated an existing intervention in a novel population, using a broad, well-validated series of psychometric instruments which spanned different dimensions of mental health problems and well-being. This is important, given that no single measure captures all stakeholder priorities in university student mental health [76], and our comprehensive approach including both mental health symptom measures and a well-being measure allowed for a more holistic understanding of COMET's impact.

Like other digital mental health intervention studies in university student samples [68-72], we experienced high attrition rates from the study. Although the absence of patterned missingness suggests that this attrition did not bias the results, it substantially impacted the sample size. While we recruited and randomized 468 participants, only 147 returned the 2-week follow-up survey (attrition: 321/468, 68.6%) and 118 returned the 4-week follow-up survey (attrition: 350/468, 74.8%). Our a priori calculations indicated that a sample size of 378 participants would be required to detect statistically significant differences in the PHQ-9. Therefore, we had insufficient power, and our findings may be prone to type II errors. After multiple imputations, many of our estimates became nonsignificant, except for depression and perceived stress at the 2-week follow-up, which were only marginally significant. This suggests that even with imputation, the effect size might have been

XSL•FO
RenderX

smaller than initially thought, and the study may need a larger sample size to be sufficiently powered.

Most participants were young, White, heterosexual women, like other university mental health intervention studies [67,77]. While this helps to provide valuable insights into the effects of COMET within this demographic, it does pose a limitation regarding the generalizability of the findings to the broader spectrum of UK university students. Furthermore, due to insufficient diversity in the sample, exploratory moderation analyses for demographic variables could not be meaningfully conducted.

While the efficacy of COMET was established across various domains, it is important to note that contrary to real-world settings, participants were given either monetary rewards or course credit incentives for completing these interventions. This raises concerns about the applicability and genuine impact of COMET outside of an incentivized research context [78]. Accordingly, to ensure results are driven by inherent value and user commitment rather than external rewards, future research should examine the intervention's impact in contexts devoid of external motivators.

This study also focused on short-term outcomes at 2- and 4-week intervals, leaving questions about sustained efficacy over the long term. This is particularly important for making fair comparisons with longer, multisession interventions that collect follow-up data over extended periods. Future investigations should emphasize extended follow-ups to provide a comprehensive understanding of an intervention's enduring benefits [79]. In contemporary psychotherapeutic research, a benchmark of at least six months is considered standard.

Finally, we were unable to measure the amount of time participants spent on COMET due to limitations in the intervention platform's functionality. It is possible that individuals who spent more time on the platform experienced greater benefits. Future research should investigate the total time spent on SSIs and examine whether time spent on different components varies among users.

Future Directions

Future studies should further explore how best to support underserved groups (eg, explore the experiences of minority groups [80]). Artificial intelligence–driven adaptive trials may also help us to answer what works for whom [81]. To reach university students before mental health symptoms become functionally impairing, early interventions or prevention efforts may achieve greater reach by embedding them within courses and ensuring maximal engagement through coproduction [82].

Conclusions

This study demonstrated the preliminary short-term efficacy of the COMET intervention, as evidenced by the significant between-group differences favoring the intervention at the 2-week follow-up. However, attrition was high, potentially biasing the results. Participant feedback indicated overall satisfaction with the intervention, with perceived accessibility, immediate benefits, and potential long-term impact being notable findings. These findings support the potential value of COMET as a mental health intervention and highlight important areas for further development in future SSI interventions.

Acknowledgments

VP (Advanced Fellowship, NIHR301312) and M Loades (Advanced Fellowship, 302929) are funded by the National Institute for Health Research (NIHR) for this research project. The views expressed in this publication are those of the authors and not necessarily those of the NIHR, National Health Service (NHS), or the UK Department of Health and Social Care. NHS's work was supported by the Economic and Social Research Council (grant ES/P000630/1). The expenses incurred in providing vouchers for the prize draw were supported by research funds from the Department of Psychology, University of Bath. We are very grateful to Akash Wasil and Professor Robert DeRubeis, Department of Psychology, University of Pennsylvania, Philadelphia, United States, who created COMET (Common Elements Toolbox) and allowed us to use it for this study. We also thank Tanvi Malhotra and Lucy Robinson for their support during this project.

Data Availability

Data are available on reasonable request. The guarantor (M Loades) is willing to examine all requests for the deidentified dataset.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 1289 KB - jmir_v27i1e58164_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

Interaction between group allocation and age, gender, previous diagnosis, current treatment, baseline depression, and baseline anxiety.
[DOCX File , 30 KB - jmir_v27i1e58164_app2.docx]

Multimedia Appendix 3
Sensitivity analysis.
[DOCX File , 15 KB - jmir_v27i1e58164_app3.docx]

Multimedia Appendix 4
Qualitative analysis.
[DOCX File , 16 KB - jmir_v27i1e58164_app4.docx]

References

1. Lee J, Jeong HJ, Kim S. Stress, anxiety, and depression among undergraduate students during the COVID-19 pandemic and their use of mental health services. Innov High Educ 2021 Apr 23;46(5):519-538 [FREE Full text] [doi: 10.1007/s10755-021-09552-y] [Medline: 33907351]
2. McManus S, Gunnell D. Trends in mental health, non-suicidal self-harm and suicide attempts in 16-24-year old students and non-students in England, 2000-2014. Soc Psychiatry Psychiatr Epidemiol 2020 Jan;55(1):125-128. [doi: 10.1007/s00127-019-01797-5] [Medline: 31667559]
3. Thorley C. Not by degrees: improving student mental health in the UK's universities. IPPR. 2017. URL: https://ippr-org.files.svdcdn.com/production/Downloads/not-by-degrees-summary-sept-2017-1.pdf [accessed 2025-01-06]
4. Thompson M, Pawson C, Evans B. Navigating entry into higher education: the transition to independent learning and living. J Further Higher Education 2021 Jun 17;45(10):1398-1410. [doi: 10.1080/0309877X.2021.1933400]
5. Sheldon E, Simmonds-Buckley M, Bone C, Mascarenhas T, Chan N, Wincott M, et al. Prevalence and risk factors for mental health problems in university undergraduate students: a systematic review with meta-analysis. J Affect Disord 2021 May 15;287:282-292. [doi: 10.1016/j.jad.2021.03.054] [Medline: 33812241]
6. Bruffaerts R, Mortier P, Auerbach RP, Alonso J, Hermosillo De la Torre AE, Cuijpers P, WHO WMH-ICS Collaborators. Lifetime and 12-month treatment for mental disorders and suicidal thoughts and behaviors among first year college students. Int J Methods Psychiatr Res 2019 Jun;28(2):e1764 [FREE Full text] [doi: 10.1002/mpr.1764] [Medline: 30663193]
7. Duffy A, Keown-Stoneman C, Goodday S, Horrocks J, Lowe M, King N, et al. Predictors of mental health and academic outcomes in first-year university students: identifying prevention and early-intervention targets. BJPsych Open 2020 May 08;6(3):e46 [FREE Full text] [doi: 10.1192/bjo.2020.24] [Medline: 32381150]
8. Hubble S, Bolton P. Coronavirus: implications for the higher and further education sectors in England (briefing paper number 8893). House of Commons Library. 2020 Apr 17. URL: https://commonslibrary.parliament.uk/research-briefings/cbp-8893/ [accessed 2025-01-10]
9. Ruscio AM. Normal versus pathological mood: implications for diagnosis. Annu Rev Clin Psychol 2019 May 07;15:179-205. [doi: 10.1146/annurev-clinpsy-050718-095644] [Medline: 31067413]
10. Johnson D, Dupuis G, Piche J, Clayborne Z, Colman I. Adult mental health outcomes of adolescent depression: a systematic review. Depress Anxiety 2018 Aug;35(8):700-716. [doi: 10.1002/da.22777] [Medline: 29878410]
11. McGorry P, Purcell R, Goldstone S, Amminger GP. Age of onset and timing of treatment for mental and substance use disorders: implications for preventive intervention strategies and models of care. Curr Opin Psychiatry 2011 Jul;24(4):301-306. [doi: 10.1097/YCO.0b013e3283477a09] [Medline: 21532481]
12. Remskar M, Atkinson MJ, Marks E, Ainsworth B. Understanding university student priorities for mental health and well-being support: A mixed-methods exploration using the person-based approach. Stress Health 2022 Oct;38(4):776-789 [FREE Full text] [doi: 10.1002/smi.3133] [Medline: 35137525]
13. Jia R, Ayling K, Chalder T, Massey A, Broadbent E, Coupland C, et al. Mental health in the UK during the COVID-19 pandemic: cross-sectional analyses from a community cohort study. BMJ Open 2020 Sep 15;10(9):e040620. [doi: 10.1136/bmjopen-2020-040620] [Medline: 32933965]
14. Savage M, James R, Magistro D, Donaldson J, Healy L, Nevill M, et al. Mental health and movement behaviour during the COVID-19 pandemic in UK university students: prospective cohort study. Mental Health Phys Activity 2020 Oct;19:100357. [doi: 10.1016/j.mhpa.2020.100357]
15. Bennett J, Heron J, Gunnell D, Purdy S, Linton MJ. The impact of the COVID-19 pandemic on student mental health and wellbeing in UK university students: a multiyear cross-sectional analysis. J Ment Health 2022 Aug;31(4):597-604 [FREE Full text] [doi: 10.1080/09638237.2022.2091766] [Medline: 35786110]
16. NUS Insight. Coronavirus Student Survey phase III November 2020 mental health and wellbeing. National Union of Students (NUS). 2020. URL: https://assets.prod.unioncloud-internal.com/document/documents/63520/de9a1b0483679d125f3b13d35bab8e48/Coronavirus_and_Students_Phase_3_study_Mental_Health_with_demographics_Nov_2020.pdf [accessed 2025-01-10]

XSL•FO
RenderX

17. Osborn T, Li S, Saunders R, Fonagy P. University students' use of mental health services: a systematic review and meta-analysis. Int J Ment Health Syst 2022 Dec 17;16(1):57 [FREE Full text] [doi: 10.1186/s13033-022-00569-0] [Medline: 36527036]
18. Broglia E, Millings A, Barkham M. Student mental health profiles and barriers to help seeking: when and why students seek help for a mental health concern. Couns and Psychother Res 2021 Sep;21(4):816-826. [doi: 10.1002/capr.12462]
19. Lui J, Sagar-Ouriaghli I, Brown JSL. Barriers and facilitators to help-seeking for common mental disorders among university students: a systematic review. J Am Coll Health 2024 Nov;72(8):2605-2613. [doi: 10.1080/07448481.2022.2119859] [Medline: 36084266]
20. Gericke F, Ebert D, Breet E, Auerbach R, Bantjes J. A qualitative study of university students' experience of internet‐based CBT for depression. Couns Psychother Res 2021 Aug 17;21(4):792-804. [doi: 10.1002/capr.12465]
21. Guenthner L, Baldofski S, Kohls E, Schuhr J, Brock T, Rummel-Kluge C. Differences in help-seeking behavior among university students during the COVID-19 pandemic depending on mental health status: results from a cross-sectional survey. Behav Sci (Basel) 2023 Oct 25;13(11):885 [FREE Full text] [doi: 10.3390/bs13110885] [Medline: 37998632]
22. Swift J, Greenberg RP. Premature discontinuation in adult psychotherapy: a meta-analysis. J Consult Clin Psychol 2012 Aug;80(4):547-559. [doi: 10.1037/a0028226] [Medline: 22506792]
23. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
24. Sin J, Galeazzi G, McGregor E, Collom J, Taylor A, Barrett B, et al. Digital interventions for screening and treating common mental disorders or symptoms of common mental illness in adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 02;22(9):e20581 [FREE Full text] [doi: 10.2196/20581] [Medline: 32876577]
25. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
26. Becker T, Torous J. Recent developments in digital mental health interventions for college and university students. Curr Treat Options Psych 2019 Jun 14;6(3):210-220. [doi: 10.1007/s40501-019-00178-8]
27. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
28. Schleider J, Weisz J. A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. J Child Psychol Psychiatry 2018 Feb;59(2):160-170. [doi: 10.1111/jcpp.12811] [Medline: 28921523]
29. Schleider JL, Abel MR, Weisz JR. Do immediate gains predict long-term symptom change? Findings from a randomized trial of a single-session intervention for youth anxiety and depression. Child Psychiatry Hum Dev 2019 Oct;50(5):868-881. [doi: 10.1007/s10578-019-00889-2] [Medline: 30993499]
30. Schleider JL, Burnette JL, Widman L, Hoyt C, Prinstein MJ. Randomized trial of a single-session growth mind-set intervention for rural adolescents' internalizing and externalizing problems. J Clin Child Adolesc Psychol 2020;49(5):660-672 [FREE Full text] [doi: 10.1080/15374416.2019.1622123] [Medline: 31219698]
31. Schleider JL, Dobias M, Sung J, Mumper E, Mullarkey MC. Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. JMIR Ment Health 2020 Jun 30;7(6):e20513 [FREE Full text] [doi: 10.2196/20513] [Medline: 32602846]
32. Schleider J, Mullarkey MC, Fox KR, Dobias ML, Shroff A, Hart EA, et al. A randomized trial of online single-session interventions for adolescent depression during COVID-19. Nat Hum Behav 2022 Feb;6(2):258-268 [FREE Full text] [doi: 10.1038/s41562-021-01235-0] [Medline: 34887544]
33. Schleider J, Weisz JR. Reducing risk for anxiety and depression in adolescents: effects of a single-session intervention teaching that personality can change. Behav Res Ther 2016 Dec;87:170-181 [FREE Full text] [doi: 10.1016/j.brat.2016.09.011] [Medline: 27697671]
34. Schleider J, Weisz JR. Little treatments, promising effects? meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
35. Wasil AR, Park SJ, Gillespie S, Shingleton R, Shinde S, Natu S, et al. Harnessing single-session interventions to improve adolescent mental health and well-being in India: development, adaptation, and pilot testing of online single-session interventions in Indian secondary schools. Asian J Psychiatr 2020 Apr;50:101980. [doi: 10.1016/j.ajp.2020.101980] [Medline: 32146337]
36. Emmons R, Stern R. Gratitude as a psychotherapeutic intervention. J Clin Psychol 2013 Aug;69(8):846-855. [doi: 10.1002/jclp.22020] [Medline: 23775470]
37. Osborn TL, Rodriguez M, Wasil AR, Venturo-Conerly KE, Gan J, Alemu RG, et al. Single-session digital intervention for adolescent depression, anxiety, and well-being: outcomes of a randomized controlled trial with Kenyan adolescents. J Consult Clin Psychol 2020 Jul;88(7):657-668. [doi: 10.1037/ccp0000505] [Medline: 32391709]

XSL•FO
RenderX

38. Wasil A, Taylor ME, Franzen RE, Steinberg JS, DeRubeis RJ. Promoting graduate student mental health during COVID-19: acceptability, feasibility, and perceived utility of an online single-session intervention. Front Psychol 2021;12:569785 [FREE Full text] [doi: 10.3389/fpsyg.2021.569785] [Medline: 33897511]
39. Weisz J, Francis SE, Bearman SK. Assessing secondary control and its association with youth depression symptoms. J Abnorm Child Psychol 2010 Oct;38(7):883-893 [FREE Full text] [doi: 10.1007/s10802-010-9440-z] [Medline: 20640503]
40. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
41. Hoffmann T, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
42. Clarke A, Friede T, Putz R, Ashdown J, Martin S, Blake A, et al. Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. BMC Public Health 2011 Jun 21;11:487 [FREE Full text] [doi: 10.1186/1471-2458-11-487] [Medline: 21693055]
43. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
44. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
46. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
47. Weiner B, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
48. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. J Psychosom Res 1993;37(2):147-153. [doi: 10.1016/0022-3999(93)90081-p] [Medline: 8463991]
49. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapam S, Oskamp S, editors. The Social Psychology of Health. Newbury Park, CA: Sage; 1988:31-67.
50. Buysse D, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
51. Snaith R, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. Br J Psychiatry 1995 Jul;167(1):99-103. [doi: 10.1192/bjp.167.1.99] [Medline: 7551619]
52. Matza L, Phillips GA, Revicki DA, Murray L, Malley KG. Development and validation of a patient-report measure of fatigue associated with depression. J Affect Disord 2011 Nov;134(1-3):294-303. [doi: 10.1016/j.jad.2011.06.028] [Medline: 21820183]
53. Vears DF, Gillam L. Inductive content analysis: a guide for beginning qualitative researchers. Focus on Health Professional Education 2022;23(1):111-127 [FREE Full text] [doi: 10.11157/fohpe.v23i1.544]
54. Braun V, Clarke V, Boulton E, Davey L, McEvoy C. The online survey as a research tool. International Journal of Social Research Methodology 2020 Aug 16;24(6):641-654. [doi: 10.1080/13645579.2020.1805550]
55. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
56. Cooke R, Bewick B, Barkham M, Bradley M, Audin K. Measuring, monitoring and managing the psychological well-being of first year university students. British Journal of Guidance & Counselling 2006 Nov;34(4):505-517. [doi: 10.1080/03069880600942624]
57. Meyer H, Stutts L. The impact of single-session gratitude interventions on stress and affect. The Journal of Positive Psychology 2023 Jan 30;19(2):315-322. [doi: 10.1080/17439760.2023.2170823]
58. Christ C, Schouten MJ, Blankers M, van Schaik DJ, Beekman AT, Wisman MA, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 25;22(9):e17831 [FREE Full text] [doi: 10.2196/17831] [Medline: 32673212]
59. Cuijpers P, Kleiboer A, Karyotaki E, Riper H. Internet and mobile interventions for depression: opportunities and challenges. Depress Anxiety 2017 Jul;34(7):596-602. [doi: 10.1002/da.22641] [Medline: 28471479]
60. Ly K, Carlbring P, Andersson G. Behavioral activation-based guided self-help treatment administered through a smartphone application: study protocol for a randomized controlled trial. Trials 2012 May 18;13:62 [FREE Full text] [doi: 10.1186/1745-6215-13-62] [Medline: 22607302]

XSL•FO
RenderX

61. Lambert J, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-based intervention using behavioral activation and physical activity for adults with depression (The eMotion Study): pilot randomized controlled trial. J Med Internet Res 2018 Jul 16;20(7):e10112 [FREE Full text] [doi: 10.2196/10112] [Medline: 30012547]
62. Fong M. and N. M. Loi, The mediating role of self?compassion in student psychological health. Australian Psychologist 2016;51(6):431-441.
63. Gilbody S, Littlewood E, McMillan D, Chew-Graham CA, Bailey D, Gascoyne S, et al. Behavioural activation to prevent depression and loneliness among socially isolated older people with long-term conditions: the BASIL COVID-19 pilot randomised controlled trial. PLoS Med 2021 Oct;18(10):e1003779 [FREE Full text] [doi: 10.1371/journal.pmed.1003779] [Medline: 34637450]
64. Fazzino T, Kunkel A, Bellitti J, Romine RS, Yi R, McDaniel C, et al. Engagement with activity monitoring during a behavioral activation intervention: a randomized test of monitoring format and qualitative evaluation of participant experiences. Behav Change 2023 Jun;40(2):103-116 [FREE Full text] [doi: 10.1017/bec.2022.7] [Medline: 37583941]
65. Cohen K. A. and J. L. Schleider, Adolescent dropout from brief digital mental health interventions within and beyond randomized trials. Internet Interventions, 2022 :27.
66. Bolinski F, Kleiboer A, Neijenhuijs K, Karyotaki E, Wiers R, de Koning L, et al. Challenges in recruiting university students for web-based indicated prevention of depression and anxiety: results from a randomized controlled trial (ICare Prevent). J Med Internet Res 2022 Dec 14;24(12):e40892 [FREE Full text] [doi: 10.2196/40892] [Medline: 36515986]
67. Rith-Najarian L, Chorpita BF, Gong-Guy E, Hammons HR, Chavira DA. Feasibility of a web-based program for universal prevention of anxiety and depression in university students: an open trial. J Am Coll Health 2022;70(8):2519-2526. [doi: 10.1080/07448481.2020.1869749] [Medline: 33577431]
68. D'Adamo L, Paraboschi L, Grammer AC, Fennig M, Graham AK, Yaeger LH, et al. Reach and uptake of digital mental health interventions based on cognitive-behavioral therapy for college students: a systematic review. J Behav Cogn Ther 2023 Jun;33(2):97-117 [FREE Full text] [doi: 10.1016/j.jbct.2023.05.002] [Medline: 37724304]
69. Lattie E, Cohen KA, Hersch E, Williams KDA, Kruzan KP, MacIver C, et al. Uptake and effectiveness of a self-guided mobile app platform for college student mental health. Internet Interv 2022 Mar;27:100493 [FREE Full text] [doi: 10.1016/j.invent.2021.100493] [Medline: 35036331]
70. Davis C, Twohig MP, Levin ME. Choosing ACT or CBT: a preliminary test of incorporating client preferences for depression treatment with college students. J Affect Disord 2023 Mar 15;325:413-420. [doi: 10.1016/j.jad.2022.12.097] [Medline: 36608854]
71. Maciejewski J, Smoktunowicz E. Low-effort internet intervention to reduce students' stress delivered with Meta's Messenger chatbot (Stressbot): a randomized controlled trial. Internet Interv 2023 Sep;33:100653 [FREE Full text] [doi: 10.1016/j.invent.2023.100653] [Medline: 37575678]
72. Ciharova M, Cuijpers P, Amanvermez Y, Riper H, Klein AM, Bolinski F, et al. Use of tailoring features and reasons for dropout in a guided internet-based transdiagnostic individually-tailored cognitive behavioral therapy for symptoms of depression and/or anxiety in college students. Internet Interv 2023 Dec;34:100646 [FREE Full text] [doi: 10.1016/j.invent.2023.100646] [Medline: 38099094]
73. Lattie E, Cohen KA, Winquist N, Mohr DC. Examining an app-based mental health self-care program, IntelliCare for college students: single-arm pilot study. JMIR Ment Health 2020 Oct 10;7(10):e21075 [FREE Full text] [doi: 10.2196/21075] [Medline: 33037874]
74. Liverpool S, Mota CP, Sales CMD, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317. [doi: 10.2196/16317] [Medline: 32442160]
75. Wong H, Lo B, Shi J, Hollenberg E, Abi-Jaoude A, Johnson A, et al. Postsecondary student engagement with a mental health app and online platform (Thought Spot): qualitative study of user experience. JMIR Ment Health 2021 Apr 02;8(4):e23447 [FREE Full text] [doi: 10.2196/23447] [Medline: 33797395]
76. Dodd A, Ward J, Byrom N. Measuring well-being in the student population. King's College London. 2022. URL: https://kclpure.kcl.ac.uk/ws/portalfiles/portal/207842571/Byrom_Dodd_measuring_wellbeing_in_the_student_population.pdf [accessed 2025-01-10]
77. Pascoe M, Dash S, Klepac Pogrmilovic B, Patten RK, Parker AG. The engagement of tertiary students with an online mental health intervention during the coronavirus disease 2019 pandemic: a feasibility study. Digit Health 2022;8:20552076221117746 [FREE Full text] [doi: 10.1177/20552076221117746] [Medline: 36060613]
78. Tishler C, Bartholomae S. The recruitment of normal healthy volunteers: a review of the literature on the use of financial incentives. J Clin Pharmacol 2002 Apr;42(4):365-375. [Medline: 11936560]
79. Vickers A, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ 2001 Nov 10;323(7321):1123-1124 [FREE Full text] [doi: 10.1136/bmj.323.7321.1123] [Medline: 11701584]
80. Sampson K, Priestley M, Dodd AL, Broglia E, Wykes T, Robotham D, et al. Key questions: research priorities for student mental health. BJPsych Open 2022 May 10;8(3):e90 [FREE Full text] [doi: 10.1192/bjo.2022.61] [Medline: 35535504]
81. Huckvale K, Hoon L, Stech E, Newby JM, Zheng WY, Han J, et al. Protocol for a bandit-based response adaptive trial to evaluate the effectiveness of brief self-guided digital interventions for reducing psychological distress in university students:

XSL•FO
RenderX

the Vibe Up study. BMJ Open 2023 Apr 28;13(4):e066249 [FREE Full text] [doi: 10.1136/bmjopen-2022-066249] [Medline: 37116996]
82. King N, Linden B, Cunningham S, Rivera D, Rose J, Wagner N, et al. The feasibility and effectiveness of a novel online mental health literacy course in supporting university student mental health: a pilot study. BMC Psychiatry 2022 Jul 30;22(1):515 [FREE Full text] [doi: 10.1186/s12888-022-04139-z] [Medline: 35907852]

Abbreviations

BA: behavioral activation
CBT: cognitive behavioral therapy
COMET: Common Elements Toolbox
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GAD-7: 7-item General Anxiety Disorder Checklist
IAM: Intervention Appropriateness Measure
ITT: intention to treat
MD: mean difference
PANAS: Positive and Negative Affect Schedule
PHQ-9: 9-item Patient Health Questionnaire
PSS-4: 4-item Perceived Stress Scale
RCT: randomized controlled trial
SSI: single-session intervention
TIDieR: Template for Intervention Description and Replication
WEMWBS: Warwick-Edinburgh Mental Well-Being Scale

Edited by A Mavragani; submitted 08.03.24; peer-reviewed by E Welsh, A Berman, J Petrovic; comments to author 31.05.24; revised version received 21.08.24; accepted 19.11.24; published 31.01.25.

Please cite as:
Lambert J, Loades M, Marshall N, Higson-Sweeney N, Chan S, Mahmud A, Pile V, Maity A, Adam H, Sung B, Luximon M, MacLennan K, Berry C, Chadwick P
Investigating the Efficacy of the Web-Based Common Elements Toolbox (COMET) Single-Session Interventions in Improving UK University Student Well-Being: Randomized Controlled Trial
J Med Internet Res 2025;27:e58164
URL: https://www.jmir.org/2025/1/e58164
doi:10.2196/58164
PMID:39888663

©Jeffrey Lambert, Maria Loades, Noah Marshall, Nina Higson-Sweeney, Stella Chan, Arif Mahmud, Victoria Pile, Ananya Maity, Helena Adam, Beatrice Sung, Melanie Luximon, Keren MacLennan, Clio Berry, Paul Chadwick. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining Health Care Provider Experiences With Patient Portal Implementation: Mixed Methods Study

Shipra Taneja[1], MSc; Kamini Kalia[2], MScN, RN; Terence Tang[1,2], MD, MSc; Walter P Wodchis[1,3], PhD; Shelley Vanderhout[1,3], PhD

[1]Trillium Health Partners, Institute for Better Health, Mississauga, ON, Canada

[2]Trillium Health Partners, Mississauga, ON, Canada

[3]Management and Evaluation, Institute of Health Policy, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Shelley Vanderhout, PhD
Trillium Health Partners
Institute for Better Health
2085 Hurontario St
Mississauga, ON, L5B 1B8
Canada
Phone: 1 647 330 7192
Email: shelley.vanderhout@thp.ca

Abstract

Background: Health systems are increasingly offering patient portals as tools for patients to access their health information with the goal of improving engagement in care. However, understanding health care providers' perspectives on patient portal implementation is crucial.

Objective: This study aimed to understand health care providers' experiences of implementing the MyChart patient portal, perspectives about its impact on patient care, clinical practice, and workload, and opportunities for improvement.

Methods: Using an explanatory sequential mixed methods approach, we conducted a web-based questionnaire and semistructured individual interviews with health care providers at a large Canadian community hospital, 6 months after MyChart was first offered to patients. We explored perspectives about the impact of MyChart on clinical practice, workload, and patient care. Data were analyzed using descriptive statistics and thematic analysis.

Results: In total, 261 health care providers completed the web-based questionnaire, and 15 also participated in interviews. Participants agreed that patients should have access to their health information through MyChart and identified its benefits such as patients gaining a greater understanding of their own health, which could improve patient safety (160/255, 62%). While many health care providers agreed that MyChart supported better patient care (108/258, 42%), there was limited understanding of features available to patients and expectations for integrating MyChart into clinical routines. Concerns were raised about the potential negative impacts of MyChart on patient-provider relationships because sensitive notes or results could be inappropriately interpreted (109/251, 43%), and a potential increase in workload if additional portal features were introduced. Suggested opportunities for improvement included support for both patients and health care providers to learn about MyChart and establishing guidelines for health care providers on how to communicate information available in MyChart to patients.

Conclusions: While health care providers acknowledged that MyChart improved patients' access to health information, its implementation introduced some friction and concerns. To reduce the risk of these challenges, health systems can benefit from engaging health care providers early to identify effective patient portal implementation strategies.

(J Med Internet Res 2025;27:e65967) doi:10.2196/65967

KEYWORDS

patient portal; mixed methods; implementation; healthcare provider; health system; patient care; online questionnaire; Canada; descriptive statistics; thematic analysis

XSL•FO
RenderX

Introduction

Providing patients with access to their health data through patient portals has become an important focus for health care systems [1]. These portals are designed to strengthen patient and provider relationships, foster greater patient engagement, improve patient education, and ultimately improve patient care [2-4]. Offering various functionalities such as appointment reminders and scheduling, access to notes and test results, and prescription refills, patient portals can equip patients to manage their health and care [5]. This increased autonomy is intended to allow them to actively participate in shared decision-making with their health care team [6,7].

While patients and families desire access to patient portals and have reported benefits of using them [8], health systems have faced challenges in implementing them. These challenges include integrating portals into existing technological infrastructure and clinical workflows [6,9]. In addition, health care providers have expressed concerns about the increased workload associated with managing patient portal inquiries and the potential for information overload and misinterpretation by patients [3,8,10,11]. This is particularly concerning in the post–COVID-19 era, where health care providers are already facing burnout due to increased patient complexity and staffing shortages [12]. Moreover, there is often a lack of agreement between patients and health care providers about which features, and content should be accessible through patient portals, leading to potential misalignment and confusion [3].

Many hospitals have adopted or are in the process of adopting digital health information systems with accompanying patient portals. While patient portals are primarily designed for patients and their families, the acceptance and influence of health care providers can significantly impact the adoption of these digital health tools [9,13]. Their endorsement can shape patient and caregiver attitudes, influencing whether they perceive the portal as a worthwhile investment of time and effort to learn and use [14-16]. As health care providers are at the forefront of care and are encouraged to increase patient engagement, it is crucial to understand their perspectives to best improve how patient portals can support patient and health care provider needs. Our objective was to understand and describe the initial impact of a patient portal on patient care and health care provider workload and suggest areas for improvement.

Methods

Setting and Implementation

This study was conducted at Trillium Health Partners (THP), a large community hospital comprised of 3 sites in Mississauga, Ontario, Canada. THP employs over 11,414 physicians, nurses, and allied health professionals, and provides care to a diverse population in the Mississauga and West Toronto area. During 2023-2024, THP had over 1.7 million patient visits and had 1457 inpatient beds [17]. This evaluation represents a strategic partnership to explore the initial impacts of implementing a vendor-developed patient portal (MyChart) across THP.

MyChart is the patient portal provided by Epic, a digital electronic health record used in health systems across the world [4], which was implemented at THP on October 10, 2020. MyChart was made available in English at THP in September 2023 to all patients aged 12 years or older. For children aged 12 years or younger, parent or guardian proxy access was provided as requested. At the time of this study, 35,387 patients had activated MyChart accounts. MyChart offered patients several features to manage care, such as access to after-visit and discharge summaries, outpatient test results, and the ability to update personal information, medications, allergies, and vaccinations. Patients can view their appointment schedule, electronically check in, cancel appointments, launch video visits, and request information for themselves or to share with other providers or people involved in their circle of care [5]. Though some health systems offer messaging between patients and providers through MyChart, THP had not activated this feature at the time of the study. At THP, MyChart functionalities were introduced in a staged manner, with new features launching every quarter. Before MyChart was implemented, THP did not have a patient portal, and patients were required to request the release of their health information from the Health Records department.

Study Design

Six months after MyChart was first offered to patients, we conducted an explanatory sequential mixed methods evaluation [18] to understand health care providers' perspectives about the initial impacts of MyChart on patient care and their workload and gather suggestions for improvement. We applied the RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) framework and its extension to dimensions of sustainability and equity to guide our evaluation, to ensure a comprehensive assessment of MyChart's impact and its potential for long-term sustainability and equitable outcomes [19,20]. We used a framework to guide the development of our questionnaire and interviews, focusing on key components of MyChart adoption and use. These components included effectiveness, such as provider-perceived impacts, negative outcomes, and patient engagement; adoption, such as provider comfort with using and promoting MyChart to patients; implementation, such as time spent by providers teaching patients about MyChart; and maintenance, such as provider capacity and resources for ongoing patient support with MyChart.

In phase 1, we conducted a cross-sectional web-based questionnaire via Qualtrics to explore health care providers' experiences and satisfaction with using MyChart in clinical practice, perceived impacts on their workload and patient care, and suggestions for the future (refer to Multimedia Appendix 1). The questionnaire was pilot-tested by 7 health care providers and revised based on their feedback. In phase 2, we conducted individual semistructured qualitative interviews with health care providers to provide additional richness and depth to questionnaire data (refer to Multimedia Appendix 2).

Participants

We recruited health care providers (physicians, nurses, and allied health) across the entire THP organization to complete a

XSL•FO
RenderX

5-minute web-based questionnaire. Health care providers were invited to participate through various institutional networks including an organization-wide email, weekly email newsletters, emails from clinical managers and directors in their departments, and daily email announcements.

For the semistructured interviews, health care providers who participated in the questionnaire were asked to indicate their interest in participating in a 30-minute follow-up interview at the end of the questionnaire. We used a combination of purposeful and maximum variation sampling to select health care providers with varying experiences, and perspectives about MyChart implementation in clinical practice [21,22]. This approach allowed us to recruit health care providers from various clinical programs with different clinical roles spanning a range of years of experience.

Data Collection

The questionnaire consisted of closed and open-ended questions and was open throughout March 2024. Health care providers provided implied consent to participate after reading an informational sheet at the beginning of the questionnaire. Questionnaire items were mostly optional to gather as much data as possible, even if responses were partially complete.

Questionnaire participants who were interested in a follow-up interview completed one semistructured interview with a female mixed-methods researcher (ST). These were conducted over Zoom between March and May 2024. Participants provided verbal consent at the start of the interviews, during which they were asked to describe their experiences with using MyChart in clinical practice, identify challenges, and make suggestions for improvement. Interviews were audio-recorded, deidentified, and transcribed verbatim; transcripts were not reviewed by participants. Data collection occurred until the research team determined information power was reached, based on the study's objective, the sample specificity, dialogue quality, and analysis strategy [23].

Data Analysis

We used descriptive statistics to analyze data from closed-ended questions in the questionnaire. Quantitative data were analyzed using Microsoft Excel and Qualtrics and presented as counts and percentages.

We applied thematic analysis to analyze qualitative data from the open-ended questionnaire items and interviews [24]. First, we familiarized ourselves with the data through repeated readings of the transcripts. After, repeating readings, 2 researchers separately generated initial codes based on key patterns and concepts. ST and SV met to develop a draft codebook and applied it to a sample of 5 transcripts and open-ended responses. After initial coding, the codebook was refined and applied to the remaining transcripts. NVivo 12 (Lumivero) was used for data analysis and management.

Ethical Considerations

This project was deemed quality improvement by the Trillium Health Partners Research Ethics Board and exempt from ethics board approval. Participants received an informational letter before completing the questionnaire and interview, outlining the study's purpose, potential benefits and risks of participation, and procedures for data deidentification and use. Informed consent was obtained from all participants before the questionnaire and interview. Interview participants received a CAD $50 (US $35) honorarium in the form of a gift card for their time.

Results

Participant Characteristics

In total, 428 individuals completed the questionnaire. After excluding, 167 responses that were incomplete (more than 50% of the questionnaire was blank) or from staff who were not health care providers but mistakenly believed they were eligible to participate, the final sample consisted of 261 individuals. Among the questionnaire participants, 44% (106/242) were nurses, 29% (70/242) were physicians, and 27% (65/242) were allied health professionals. In total, 15 health care providers participated in follow-up interviews, including 4 physicians, 6 nurses, and 5 allied health professionals. Participant characteristics are described in Table 1, and a summary of their perceptions of and experiences with MyChart are shown in Figure 1. In both the open-ended items on the questionnaire and in interviews, participants shared their experiences with and perspectives about the initial impacts of MyChart, focusing on patient care, workflow integration, and usability. Results are summarized below across common themes from the quantitative and qualitative data

XSL•FO
RenderX

Table 1. Participant demographics, experience, and clinical background.

Characteristics[a]	Questionnaire, n (%)	Interviews, n (%)
Age		
<25 years	4 (2)	0 (0)
25-34 years	37(15)	2 (15)
35-44 years	82 (33)	6 (46)
45-54 years	75 (30)	5 (38)
55 years or older	50 (20)	0 (0)
Roles		
Allied Health[b]	65 (27)	5 (33)
Nurse (registered nurse, nurse practitioner, registered practical nurse)	106 (44)	6 (40)
Physician	70 (29)	4 (27)
Physician's assistant	2 (1)	0 (0)
Clinical area		
Cardiac health	22 (9)	0 (0)
Diagnostic imaging	4 (2)	0 (0)
Emergency department and urgent care	18 (7)	2 (13)
ICU[c] and critical care	12 (5)	0 (0)
Inpatient medicine	20 (8)	1 (6.7)
Laboratory medicine and genetics	11 (4)	1 (6.7)
Mental health	20 (8)	1 (6.7)
Neuro or musculoskeletal	14 (6)	2 (13)
Oncology	30 (12)	3 (20)
Other (eg, clinical resource team)	11 (4)	0 (0)
Outpatient medicine and renal	23 (9)	2 (13)
Pharmacy	4 (2)	0 (0)
Primary, rehab, CCC[d], palliative, senior	19 (8)	1 (6.7)
Surgery and Perioperative care	19 (8)	1 (6.7)
Women's and children's	19 (8)	1 (6.7)
Setting		
Ambulatory care	81 (34)	9 (60)
Inpatient care	88 (37)	2 (13)
Both ambulatory and inpatient care	70 (29)	4 (27)

[a]Some participants did not report each characteristic.

[b]Allied Health included an occupational therapist, physiotherapist, pharmacist, dietitian, speech-language pathologist, social worker, radiation therapist, midwife, genetic counselor, dialysis assistant, kinesiologist, and child and youth worker.

[c]ICU: intensive care unit.

[d]CCC: complex continuing care.

XSL•FO
RenderX

Figure 1. Questionnaire participants' perceptions of the patient portal MyChart.

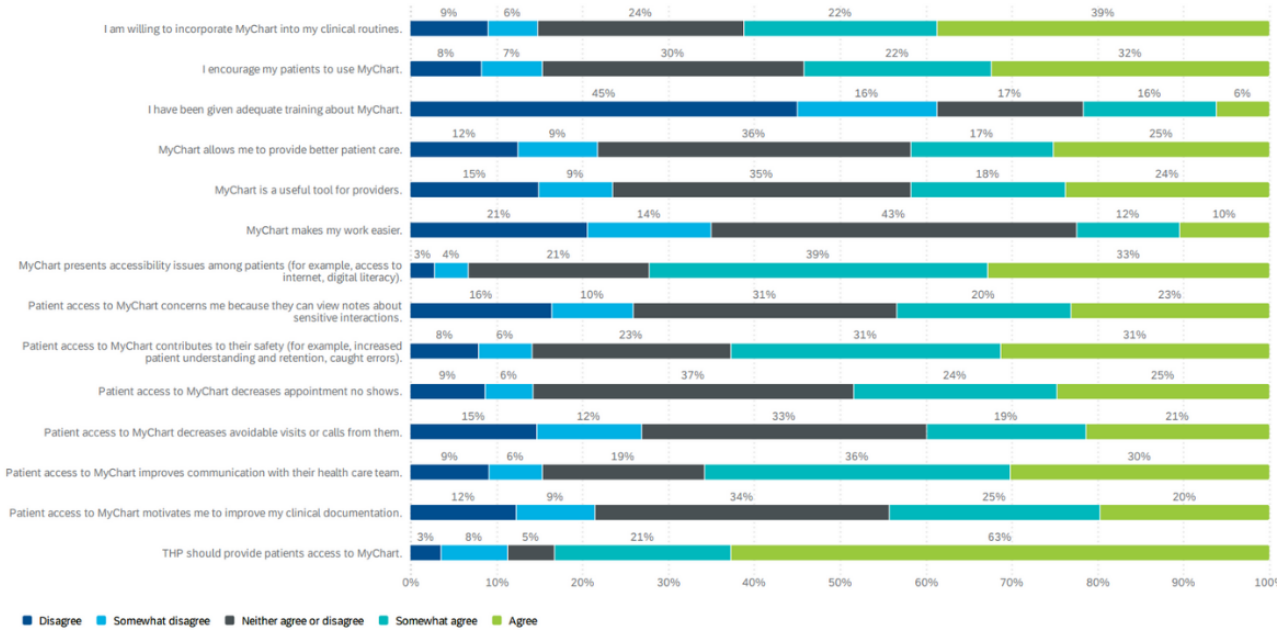

Patient Access to MyChart

Health care providers agreed or somewhat agreed that patients should have access to MyChart (215/258, 84%), as they perceived it to improve communication between patients and health care teams (168/255, 66%), and contribute to patient safety by increasing patient understanding of their own health and identifying errors (160/255, 62%). Among interview participants, those who strongly supported patient access to MyChart were often users of the tool themselves.

> *Obviously, you know, empowering patients with accessibility to their own information is really helpful and important. And, like, I have family members with chronic illness and it's very helpful for me to be able to just look at their results...* [Physician #2]

Participants indicated that certain patient populations may have difficulty with accessing and using MyChart, such as those who are uncomfortable using technology (166/182, 91%), have limited access to devices (129/182, 71%) and internet (127/182, 70%), and experience language barriers (132/182, 73%). Additional populations with potential barriers to using MyChart included geriatric patients, children aged 12 years or younger, and individuals with cognitive impairments.

Patient Access to Information Within MyChart

In this health system setting, MyChart features were made available to patients in a staged approach, where after-visit and discharge summaries, THP outpatient laboratory as well as diagnostic imaging reports, and appointment details were introduced as part of the initial implementation. E–check-in functionality and patient-entered questionnaires were phased sequentially over time. Health care providers had differing views on the staged approach of the release of information; while they tended to agree that it may have lessened upfront impacts on workflow and workload, there were varied perspectives about releasing additional information to patients in the future. Some advocated for releasing more test results and clinical notes, which was seen as enabling patients to share results with family members, understand their health, and seek second opinions.

> *More information should be available to the patient via MyChart - both inpatient and outpatient. it is important for patients to have access to their labs, clinical documentation, medications orders, as it gives them oversight into their care, opportunity to ask questions and catch errors or discrepancies. It gives them control over their health rather than providers dictating what should be done.* [Pharmacist #1]

Others voiced concerns that patients accessing results, images, and notes before health care encounters could heighten anxiety and create extra administrative work due to calls about "concerning results" and prompting expedited appointments. There were also concerns about patients viewing sensitive results (eg, diagnostic images) before discussing them with their healthcare provider or misinterpreting them.

> *But in terms of actually disclosing medical reports that are written in medical language with the intention of being a communication between medical providers, I think it results in unnecessary anxiety for patients frequently. I think there's certain types of patients that will review the details of their clinical imaging reports or pathology reports, clinical notes, and they will, you know, be highlighting things and freaking out about things. Sometimes totally inappropriately just due to their not being a typical intended recipient of that type of documentation. Sometimes appropriately if there could be bad news but instead of in the supportive, hopefully supportive environment of a doctor patient interaction, by themselves, alone in their house with maybe many days before they are actually going to be reviewing that with their provider.* [Physician 4]

To address concerns about patient anxiety, multiple participants proposed releasing test results and notes only after health care

XSL•FO
RenderX

encounters, which could allow patients to receive and interpret information in a supportive environment.

Impacts on Patient Care

When asked if MyChart allowed health care providers to provide better care, 42% (108/255) agreed or somewhat agreed, while 36% (94/258) were neutral. One nurse described that MyChart provided patients with comfort, allowing them to get secondary opinions or explanations if needed.

> *It gives them comfort. That's just sense I got. So, for some of the patients they just – they would ask, like, can I get this information, can I get this information. By having MyChart access they know that they could access at any time if they want to – like, if they want to talk to the family member or a family doctor or any other – anybody within, like, within their comfort level. Like, they want to share their information. They know that they could access it and then actually share medical charts. Instead of saying, oh, based on my memory this is what I was what I heard and what I told. Like, that's not the case now.* [Nurse #3]

While health care providers identified some benefits to patients having MyChart, there were also concerns. In total, 43% (107/251) of participants agreed or somewhat agreed that patient access to MyChart concerned them as patients could view notes about sensitive interactions, while 31% (77/251) felt neutral about this. Despite patient health information being available for patients to access via a request to the Health Records department, the increased access to notes caused some to express concern that allowing patients to view notes that they may not agree with or that describe challenging interactions could have a negative impact on patient-provider relationships.

> *And so I've had them say, like, you know, can you correct this from this note [discharge summary], or I read this and I disagree with this. Or I've even had patients sort of get a little bit kind of offended, I guess. Because, like, in the mental status exam, for example, if you can say, like, poor insight into illness or having delusions of grandiosity so they could read that and then they come back and they're a little bit sort of dysregulated because of that.* [Nurse Practitioner #1]

Although clinical notes were limited to just discharge summaries, some perceived negative impacts or raised concerns that allowing proxies to view certain records could cause negative consequences for patients.

> *We have something called an elder abuse screen which is, like, a piece of paper we fill out to follow a flowchart to determine what results should be taken in elder abuse. Like, should that be immediately available in my notes to a patient that an elder abuse screen was completed? Yes. But [...] The elder abuse screen is also used for people who are incapable. Right? Incapable in quotes. So, it would be their SDM [substitute decision maker] who would be accessing MyChart on their behalf. What if the SDM is the abuser and then they go and they access MyChart? And then they see that.* [Social Worker #1]

Workflow Integration

In an effort to limit clinician workload in a time of health human resource crisis, THP had intentionally limited impact on clinicians during the launch of MyChart, particularly through its staged approach to releasing MyChart features over time. However, this has led participants to report inadequate training and information on MyChart and its available functionalities (116/258, 45%). This also may have contributed to our findings that 22% (58/258) of participants agreed or somewhat agreed that MyChart made their work easier, and 42% (108/258) agreed or somewhat agreed that MyChart is a useful tool for providers, and 68% (173/254) agreed or somewhat agreed that MyChart had no impact on their workload. A lack of knowledge about MyChart, along with new responsibilities to promote signing up for MyChart, caused some concern among health care providers about workflow, workload, and changing documentation practices. These views may have also been attributable to the limited MyChart features offered at THP; for example, there was no need to respond to patient messages and health care providers were not activating MyChart accounts. While some providers were not aware that MyChart was being implemented in a staged approach, others agreed that this method eased the need to immediately incorporate it into clinical routines, especially when they may not be familiar with all of its features and functionalities. Some interview participants reported having to provide detailed explanations for clinically irrelevant information; while they acknowledged the benefit of patients being informed about their health, in-depth explanations can lead to extended encounters especially if patients have limited health literacy.

> *Because with all of that extra information most patients are not – don't have the health literacy to actually interpret the results or the test that they see. And so, you know, it can become challenging because they will ask questions about, you know, pertinent results that are contributing to the clinical picture, which is one thing, but then they also will often ask about minor changes or clinically relevant, you know, findings that we don't necessarily consider... Sometimes I feel like I'm teaching an introductory, like, medical school class and that's very burdensome when you have, like, a high volume of patients and you have – you know, sometimes these requests or these questions are coming from family who aren't even there physically, like, there, calling in remotely saying that they found something. So that's a challenge.* [Physician #2]

MyChart also changed workload by influencing clinical documentation practices. Among 253 participants, 45% (n=112) agreed or somewhat agreed that patients accessing MyChart motivated them to improve their clinical documentation to increase the accessibility of language for patients to read and understand. However, some expressed concerns about how adjusting their current documentation routine might impact workload in the future.

Though messaging between patients and providers had not been introduced as a MyChart feature in this health system, its

XSL•FO
RenderX

hypothetical implementation was an area of tension. A few providers were in favor of it, but most were opposed, fearing it would increase uncompensated time to their workload and disrupt their workflow. Some raised concerns about patient expectations for immediate responses and suggested a need for clear guidelines for the type and timeliness of communication that would be required through messaging.

> *There's pros and cons. So, the pro is, you know, as a patient you can – it's better access to care. It's better access to addressing any kind of issue or concern you may have, whether it be clarification about a prescription or an appointment or care that's being provided. But I can see from a healthcare professional's point of view where, you know, you might get inundated with, you know, these requests or emails or chats. And so, you know, what is the guideline in terms of, you know, what types of chats you should be receiving or what kind of information you can be sharing over the chat, whether it's secure. And what's sort of the guidelines in terms of when you're expected to respond to something.* [Nurse #1]

Discussion

Principal Results

In this study, 261 health care providers completed a questionnaire, and 15 shared their perspectives about and experiences with the initial impact of implementing the MyChart patient portal at a large community hospital. Our findings revealed a mixed response. While most health care providers were generally supportive of providing patients access to their health information, limited health care provider engagement and clear communication about the role of MyChart in care may have negatively impacted how health care providers perceived the tool. Those who were MyChart users themselves had more favorable views of it and, logically, understood more about how it can support patient care. However, some providers had concerns about the potential negative consequences of patient access to MyChart, including patient anxiety or misunderstandings from viewing sensitive notes, increased provider workload due to additional explanations, and accessibility challenges for patients with limited technology skills, internet access, or language barriers. We identified an opportunity to strengthen health care providers' understanding of how MyChart can be best integrated into their routines and the care they deliver; one unexpected and relatively simple way to achieve this could be to encourage them to access the portal themselves.

Comparison With Previous Work

Participants raised concerns about patients or proxies viewing notes or results for sensitive interactions or diagnoses, which could be misinterpreted and then negatively impact patient-provider relationships. This aligns with other studies, which identified concerns among health care providers about patients accessing mental health notes or pathology test results [25-28], and becoming overwhelmed, hurt, or surprised by the content. Studies have also reported worries of unauthorized portal access, for instance, by a curious spouse reading a patient's mental health notes detailing domestic abuse [25]. Concerns have also been raised about patients discovering new diagnoses from alerts on patient portals, leading to distress [28,29], which highlights the need for better communication about accessing sensitive information such as adding disclosures to alerts and notes. For instance, patients could be directed to indicate if they prefer to see results only after discussing with them with their health care team and opt out of viewing sensitive notes.

Interestingly, participants reported that MyChart improved patient-provider communication even in the absence of a secure messaging feature. This may possibly be due to both parties having clearer and more consistent access to patients' health information, allowing for more meaningful, informed, and productive conversations. Though secure messaging through portals is often desired by patients [30], this was an area of concern for many providers in our study. They feared it could increase their already-high workload, disrupt their workflow, and was not something they would be explicitly compensated for. Secure messaging has been a conflicting debate in the literature, with some studies demonstrating that it does not tend to impact health care provider workload [13,31,32], as it can reduce in-person visits and follow-up calls and help patients self-manage; however, other studies have found that it can increase indirect patient care activities [26,33,34].

Introducing patient portals requires organizations to take a thoughtful approach that considers potential positive benefits, potential harms, manageable workloads for staff, and evolving regulatory landscapes (eg, the Cures Act in the United States) [35]. Our institution followed a staged approach to releasing MyChart features, which reflected organizational capacity and early engagement with health care providers suggesting limited readiness for comprehensive features to be introduced all at once, likely related to high care volumes, limited resources and support for equipping health care providers to rapidly adopt and proficiently use MyChart, and ongoing challenges with retention of health human resources as well as burnout post the COVID-19 pandemic [36]. Participants in our study agreed this approach likely mitigated some initial challenges of implementing technology into clinical environments; as many other health systems also face similar challenges, our peers may benefit from a similar approach to align with capacity limits and available resources.

Providers in this study reported a limited awareness of what patients could access in MyChart, and expectations for how it should be incorporated into care routines. However, those who were MyChart users themselves often had more favorable attitudes toward MyChart and would likely be able to provide better support to patients as they navigate the tool. Our health system also serves the most diverse patient community in Canada with 52% of residents identifying as immigrants and 62% as visible minorities [10], which participants agreed can introduce challenges around language and accessibility barriers. Comprehensive support for both patients and providers to learn about and access patient portals can help address apprehension about them, and challenges with digital navigation, and encourage patients to access their full range of features [26,37].

XSL•FO
RenderX

Strengths and Limitations

The use of a mixed methods approach, combining both qualitative and quantitative data, allowed for a holistic and deep evaluation of perspectives and experiences with MyChart during the first 6 months it was available to patients. This approach, coupled with a diverse participant pool representing various clinical settings, roles, ages, and experience levels, ensured a rich and nuanced understanding of the impact of MyChart.

Our study is limited by a relatively low response rate to the questionnaire, despite a broad communication cascade and long duration, but this may have been attributable to limited engagement with health care providers upon MyChart's launch. As a result, we were unable to interview participants from every clinical program in our health system to gain comprehensive perspectives about MyChart, as its use varies by clinical area; this may limit the transferability of our findings to areas where the portal is not widely used. However, our participants often represented clinical areas where MyChart is more frequently used by patients such as oncology and outpatient medicine. Our study was undertaken in the context of a staged release approach of portal features that followed the implementation of a hospital-wide digital health information system (Epic), so our findings may also not be transferable to health care systems using a different implementation strategy. Though community health care providers outside of tertiary care centers in our health system play important roles in delivering care, we did not include them in this study because MyChart was introduced in hospital settings only. Future studies will include their perspectives to understand how entire health systems interact with and support patient portals.

Conclusion and Recommendations

A number of recommendations for peer health systems flow from our findings. Given the number of participants who reported inadequate training and knowledge of MyChart, it is important to engage providers early in the process of implementing a patient portal. This can help to understand their readiness and capacity for introducing a new tool into care and create clear alignment about its purpose and expectations for how health care providers will use it. Encouraging providers to sign up for accounts themselves (when applicable) and providing ample resources to learn about and use patient portals can help introduce the portal into care and reduce the time providers may need to spend assisting patients with technical difficulties during health care encounters. Participants in our study shared their concerns about secure messaging, especially around the increase in workload. To address these concerns and manage the tension that may exist between patient priorities and provider capacity, it is recommended to invite both groups to share dialogue around shared expectations for the content and timeliness of secure messaging communication. In addition, allowing physicians to bill for time spent messaging patients [38,39] could mitigate the risk of resistance to this feature or burnout. Finally, to reduce the number of repetitive patient questions about bloodwork or imaging and enhance patient education, health care systems should implement a system to track the queries health care providers receive. By identifying common themes and areas of confusion, resources can then be added to the portal to address these questions, making it easier for both patients and health care providers.

Acknowledgments

Thank you to the Trillium Health Partners MyChart Project Executives: Dr Chris Hayes, Dr Joan Murphy, and Danielle Sanagan, and Steering Committee, Adam Gdyczynski, Raj Patel, Hassan Khurram, Marvelyn Oni, MyChart patient advisors, and pilot testers Drs. Katherine Enright, Ian Zelena, and Charles Lim for their support in conducting this study.

Data Availability

The data generated and analyzed during the current study are not publicly available. Participants did not consent to data sharing with individuals outside of the research team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
HealthCare Provider Questionnaire.
[DOCX File , 21 KB - jmir_v27i1e65967_app1.docx]

Multimedia Appendix 2
Interview Guide for Health Care Providers.
[DOCX File , 16 KB - jmir_v27i1e65967_app2.docx]

References

1. Fujioka JK, Bickford J, Gritke J, Stamenova V, Jamieson T, Bhatia RS, et al. Implementation strategies to improve engagement with a multi-institutional patient portal: multimethod study. J Med Internet Res 2021;23(10):e28924 [FREE Full text] [doi: 10.2196/28924] [Medline: 34709195]

XSL•FO
RenderX

2. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
3. Nøst TH, Faxvaag A, Steinsbekk A. Participants' views and experiences from setting up a shared patient portal for primary and specialist health services- a qualitative study. BMC Health Serv Res 2021;21(1):171 [FREE Full text] [doi: 10.1186/s12913-021-06188-8] [Medline: 33627122]
4. Johnson AM, Brimhall AS, Johnson ET, Hodgson J, Didericksen K, Pye J, et al. A systematic review of the effectiveness of patient education through patient portals. JAMIA Open 2023;6(1):ooac085 [FREE Full text] [doi: 10.1093/jamiaopen/ooac085] [Medline: 36686972]
5. Powered by epic. MyChart. URL: https://www.mychart.org/ [accessed 2024-09-20]
6. Ryan B, Brown J, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. BMJ Health Care Inform. 2016. URL: https://informatics.bmj.com/content/23/2/534 [accessed 2024-04-26]
7. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
8. Avdagovska M, Ballermann M, Olson K, Graham T, Menon D, Stafinski T. Patient portal implementation and uptake: qualitative comparative case study. J Med Internet Res 2020;22(7):e18973 [FREE Full text] [doi: 10.2196/18973] [Medline: 32716308]
9. Janssen A, Keep M, Selvadurai H, Kench A, Hunt S, Simonds S, et al. Factors that influence use of a patient portal by health professionals. Int J Environ Res Public Health 2021;18(4):1877 [FREE Full text] [doi: 10.3390/ijerph18041877] [Medline: 33671953]
10. Sisk B, Bereitschaft C, Enloe M, Schulz G, Mack J, DuBois J. Oncology clinicians' perspectives on online patient portal use in pediatric and adolescent cancer. JCO Clin Cancer Inform. 2023. URL: https://ascopubs.org/doi/10.1200/CCI.23.00124 [accessed 2024-12-18]
11. Petrovskaya O, Karpman A, Schilling J, Singh S, Wegren L, Caine V, et al. Patient and health care provider perspectives on patient access to test results via web portals: scoping review. J Med Internet Res 2023;25:e43765 [FREE Full text] [doi: 10.2196/43765] [Medline: 37856174]
12. Burrowes SA, Casey SM, Pierre-Joseph N, Talbot SG, Hall T, Christian-Brathwaite N, et al. COVID-19 pandemic impacts on mental health, burnout, and longevity in the workplace among healthcare workers: a mixed methods study. J Interprof Educ Pract 2023;32:100661 [FREE Full text] [doi: 10.1016/j.xjep.2023.100661] [Medline: 37305404]
13. Janssen A, Keep M, Selvadurai H, Shaw T. Health professionals' experiences with a patient portal pre and post launch: a qualitative study. Health Policy and Technology 2023;12(3):100761. [doi: 10.1016/j.hlpt.2023.100761]
14. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
15. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
16. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
17. Annual report 2024 - trillium health partners. Trillium Health Partners. URL: https://www.thp.ca/AR2024/A/Home.html [accessed 2024-06-10]
18. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Los Angeles, Washington DC: Sage publications; 2017.
19. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
20. Shelton RC, Chambers DA, Glasgow RE. An extension of RE-AIM to enhance sustainability: addressing dynamic context and promoting health equity over time. Front Public Health 2020;8:134 [FREE Full text] [doi: 10.3389/fpubh.2020.00134] [Medline: 32478025]
21. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Los Angeles, Washington DC: SAGE Publications, Inc; 2014.
22. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
23. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
24. Braun V, Clarke V. Thematic Analysis. Washington, DC: American Psychological Association; 2012.
25. van Rijt AM, Hulter P, Weggelaar-Jansen AM, Ahaus K, Pluut B. Mental health care professionals' appraisal of patients' use of web-based access to their electronic health record: qualitative study. J Med Internet Res 2021;23(8):e28045 [FREE Full text] [doi: 10.2196/28045] [Medline: 34448705]

XSL•FO
RenderX

26. Laukka E, Huhtakangas M, Heponiemi T, Kujala S, Kaihlanen AM, Gluschkoff K, et al. Health care professionals' experiences of patient-professional communication over patient portals: systematic review of qualitative studies. J Med Internet Res 2020;22(12):e21623 [FREE Full text] [doi: 10.2196/21623] [Medline: 33289674]
27. Turvey CL, Fuhrmeister LA, Klein DM, Moeckli J, Howren MB, Chasco EE. Patient and provider experience of electronic patient portals and secure messaging in mental health treatment. Telemed J E Health 2022;28(2):189-198 [FREE Full text] [doi: 10.1089/tmj.2020.0395] [Medline: 33887164]
28. Shapiro A. My mom got diagnosed with cancer through the MyChart app. J Am Med Inform Assoc 2021;29(1):217 [FREE Full text] [doi: 10.1093/jamia/ocab193] [Medline: 34664640]
29. van Kuppenveld SI, van Os-Medendorp H, Tiemessen NA, van Delden JJ. Real-time access to electronic health record via a patient portal: is it harmful? A retrospective observational study. J Med Internet Res 2020;22(2):e13622 [FREE Full text] [doi: 10.2196/13622] [Medline: 32044753]
30. Avdagovska M, Menon D, Stafinski T. Capturing the impact of patient portals based on the quadruple aim and benefits evaluation frameworks: scoping review. J Med Internet Res 2020;22(12):e24568 [FREE Full text] [doi: 10.2196/24568] [Medline: 33289677]
31. Yi YZP, Terhilda GM, Homer L, Chin MD, Andrew M, Wiesenthal MD, et al. Patient access to an electronic health record with secure messaging: impact on primary care utilization. AJMC. 2007. URL: https://www.ajmc.com/view/jul07-2509p418-424 [accessed 2024-06-03]
32. Keplinger L, Koopman R, Mehr D, Kruse R, Wakefield D, Wakefield B, et al. Patient portal implementation: resident and attending physician attitudes. Fam Med 2013;45(5):335-340 [FREE Full text] [Medline: 23681685]
33. Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L. Characterizing patient-clinician communication in secure medical messages: retrospective study. J Med Internet Res 2022;24(1):e17273 [FREE Full text] [doi: 10.2196/17273] [Medline: 35014964]
34. Kelly MM, Dean SM, Carayon P, Wetterneck TB, Hoonakker PL. Healthcare team perceptions of a portal for parents of hospitalized children before and after implementation. Appl Clin Inform 2017;8(1):265-278 [FREE Full text] [doi: 10.4338/ACI-2016-11-RA-0194] [Medline: 28293685]
35. 21st Century Cures Act. Commissioner O of the. FDA. 2020. URL: https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act [accessed 2024-07-03]
36. Stillman M. Death by patient portal. JAMA 2023;330(3):223-224. [doi: 10.1001/jama.2023.11629] [Medline: 37389857]
37. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
38. Avdagovska M, Stafinski T, Ballermann M, Menon D, Olson K, Paul P. Tracing the decisions that shaped the development of myChart, an electronic patient portal in alberta, Canada: historical research study. J Med Internet Res 2020;22(5):e17505 [FREE Full text] [doi: 10.2196/17505] [Medline: 32452811]
39. Bulletin 240103 — patient-to-physician secure messaging proof-of-concept pilot. OHIP INFOBulletins ontario.ca. 2024. URL: http://www.ontario.ca/document/ohip-infobulletins-2024/bulletin-240103-patient-physician-secure-messaging-proof-concept [accessed 2024-06-17]

Abbreviations

RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance
THP: Trillium Health Partners

Edited by A Mavragani; submitted 30.08.24; peer-reviewed by AM Weggelaar-Jansen, M Avdagovska; comments to author 28.11.24; revised version received 12.12.24; accepted 18.12.24; published 31.01.25.

Please cite as:
Taneja S, Kalia K, Tang T, Wodchis WP, Vanderhout S
Examining Health Care Provider Experiences With Patient Portal Implementation: Mixed Methods Study
J Med Internet Res 2025;27:e65967
URL: https://www.jmir.org/2025/1/e65967
doi:10.2196/65967
PMID:39888658

©Shipra Taneja, Kamini Kalia, Terence Tang, Walter P Wodchis, Shelley Vanderhout. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

XSL•FO
RenderX

and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartphone-Based Intervention Targeting Norms and Risk Perception Among University Students with Unhealthy Alcohol Use: Secondary Mediation Analysis of a Randomized Controlled Trial

Joseph Studer[1,2], PhD; John A Cunningham[3,4,5], PhD; Elodie Schmutz[1], MSc; Jacques Gaume[1], PhD; Angéline Adam[1], MD; Jean-Bernard Daeppen[1], MD; Nicolas Bertholet[1], MSc, MD

[1]Addiction Medicine, Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

[2]Service of Adult Psychiatry North-West, Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Nyon, Switzerland

[3]National Addiction Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[4]Center for Addiction and Mental Health, Toronto, ON, Canada

[5]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Joseph Studer, PhD
Addiction Medicine
Department of Psychiatry
Lausanne University Hospital and University of Lausanne
Rue du Bugnon 23A
Lausanne, 1011
Switzerland
Phone: 41 213149033
Email: joseph.studer@chuv.ch

Abstract

Background: Many digital interventions for unhealthy alcohol use are based on personalized normative feedback (PNF) and personalized feedback on risks for health (PFR). The hypothesis is that PNF and PFR affect drinkers' perceptions of drinking norms and risks, resulting in changes in drinking behaviors. This study is a follow-up mediation analysis of the primary and secondary outcomes of a randomized controlled trial testing the effect of a smartphone-based intervention to reduce alcohol use.

Objective: This study aimed to investigate whether perceptions of drinking norms and risks mediated the effects of a smartphone-based intervention to reduce alcohol use.

Methods: A total of 1770 students from 4 higher education institutions in Switzerland (mean age 22.35, SD 3.07 years) who screened positive for unhealthy alcohol use were randomized to receive access to a smartphone app or to the no-intervention control condition. The smartphone app provided PNF and PFR. Outcomes were drinking volume (DV) in standard drinks per week and the number of heavy drinking days (HDDs) assessed at baseline and 6 months. Mediators were perceived drinking norms and perceived risks for health measured at baseline and 3 months. Parallel mediation analyses and moderated mediation analyses were conducted to test whether (1) the intervention effect was indirectly related to lower DV and HDDs at 6 months (adjusting for baseline values) through perceived drinking norms and perceived risks for health at 3 months (adjusting for baseline values) and (2) the indirect effects through perceived drinking norms differed between participants who overestimated or who did not overestimate other people's drinking at baseline.

Results: The intervention's total effects were significant (DV: b=–0.85, 95% bootstrap CI –1.49 to –0.25; HDD: b=–0.44, 95% bootstrap CI –0.72 to –0.16), indicating less drinking at 6 months in the intervention group than in the control group. The direct effects (ie, controlling for mediators) were significant though smaller (DV: b=–0.73, 95% bootstrap CI –1.33 to –0.16; HDD: b=–0.39, 95% bootstrap CI –0.66 to –0.12). For DV, the indirect effect was significant through perceived drinking norms (b=–0.12, 95% bootstrap CI –0.25 to –0.03). The indirect effects through perceived risk (for DV and HDD) and perceived drinking norms (for HDD) were not significant. Results of moderated mediation analyses showed that the indirect effects through perceived drinking norms were significant among participants overestimating other people's drinking (DV: b=–0.17, 95% bootstrap CI –0.32 to –0.05; HDD: b=–0.08, 95% bootstrap CI –0.15 to –0.01) but not significant among those not overestimating.

XSL•FO
RenderX

Conclusions: Perceived drinking norms, but not perceived risks, partially mediated the intervention's effect on alcohol use, confirming one of its hypothesized mechanisms of action. These findings lend support to using normative feedback interventions to discourage unhealthy alcohol use.

Trial Registration: ISRCTN Registry 10007691; https://doi.org/10.1186/ISRCTN10007691

(J Med Internet Res 2025;27:e55541) doi:10.2196/55541

KEYWORDS

brief intervention; alcohol use; mechanism of action; mediation analysis; personalized feedback; smartphone app; students; Switzerland; mobile phone; mediation; feedback; student; health risk; drinking; drinker; support; feedback intervention

Introduction

Background

Unhealthy alcohol use is the highest risk factor for lowering young adults' disability-adjusted life years [1]. University students are particularly vulnerable to alcohol-related consequences [2-5] and thus constitute an appropriate target for selective prevention. Brief interventions (BIs) are one approach to reducing unhealthy alcohol use and its related consequences, and they are recommended by the World Health Organization [6], the United States Preventive Services Task Force [7], and the United Kingdom's National Institute for Health and Care Excellence [8]. The efficacy of BIs, administered using a variety of formats (ie, face-to-face, through the internet, on a computer, through a smartphone app), has been demonstrated among individuals not actively seeking treatment [9,10]. However, the mechanisms underlying the efficacy of BIs are still not completely understood, particularly with web- and smartphone-based BIs. Many of the digital interventions for unhealthy alcohol use that have been found effective were based on personalized normative feedback (PNF) and personalized feedback on risks for health (PFR) [11-13].

Personalized Normative Feedback and Personalized Feedback on Risks for Health in Brief Interventions

PNF is based on social norms theory, which posits that an individual's behavior is influenced by their perceptions and beliefs about what is considered normal behavior among other people [2,14-16]. Thus, an individual's alcohol consumption is impacted by their perceptions of other people's drinking habits. Previous studies have shown that there is a high prevalence of individuals who overestimate the amount that others drink and that those who overestimate other people's drinking typically consume more alcohol [17,18]. There are also indications that this relationship may be bidirectional, that is, the more an individual overestimates their peers' drinking behaviors, the more they drink; and the more an individual drinks, the more they overestimate their peers' drinking [19]. Therefore, providing accurate information about actual norms, using PNF, is expected to correct an individual's overestimations of other people's drinking and encourage them to reduce their alcohol use by adopting drinking behaviors more aligned with actual norms [20]. Digital and in-person BIs that aim to correct an individual's overestimations of other people's drinking (using PNF) are considered effective [13]. In BIs that include a PNF component, individuals are typically asked to provide information on their sex, age, personal alcohol consumption, and perceptions of their peers' alcohol consumption. The PNF then consists of a graphical comparison between the user's own alcohol use, their peers' alcohol use, and that of a reference group (eg, individuals of the same age and sex in the general population) [14,20,21]. It has been proposed that using more proximal normative reference groups (eg, gender-specific, age-specific, and group-specific) compared with more distal reference groups (eg, the general population) may enhance the efficacy of PNF [22]. However, this proposition has been challenged. For example, in a randomized controlled trial (RCT) among college students, LaBrie and colleagues [23] compared the efficacy of PNF with different levels of specificity of the reference group (ie, 7 groups with typical same-campus student as a reference with level of specificity based on a combination of gender, race, and Greek affiliation) with a typical same-campus student as a reference. Results indicated that participants receiving PNF with more specific reference groups showed less reduction in alcohol-related outcomes compared with those who received PNF referencing a typical student. In addition, other forms of PNF have been designed to improve motivation, engagement, and adherence. For instance, in 2 studies among college students, the addition of gamified elements (eg, a point-based reward system, norm visualization through visibly connected peers on Facebook [Meta], and elements of chance) was associated with greater reductions in both perceived peer drinking norms and personal alcohol use than standard PNF [24,25].

PFR is based on various theories related to health-protective behavior and emphasizes the significant motivational influence of perceived risk [26-29]. The desire to prevent or minimize negative health outcomes (perceived benefits) is thought to generate greater motivation for self-protection [30]. Consequently, PFR is anticipated to improve an individual's awareness of risks and motivation for self-protection and to result in reduced alcohol consumption or altered drinking patterns.

Thus, perceived drinking norms and perceived risks for health are expected to serve as mechanisms of action of PNF and PFR interventions and, therefore, they should be tested as mediators for the impact of such interventions on alcohol consumption [30,31]. The large number of RCTs showing the beneficial effects on alcohol outcomes of BIs incorporating PNF or PFR could be interpreted as indirect support for this assumption [23,32-35]. However, since most BIs comprise multiple components, including, but not limited to, PNF and PRF, these studies failed to formally test the mediation of interventions' effects through perceived norms and risks. Few studies have

XSL•FO
RenderX

directly investigated whether perceived norms mediated the effects of face-to-face [36-38] and digital [23,39-43] BIs on alcohol use. These studies were principally conducted among students [23,36-40,43], military personnel, or veterans [41,42]. They showed significant mediation through perceived drinking norms, although the results were not always consistent across outcomes. Furthermore, many of these studies failed to ensure the temporal ordering necessary between intervention, mediator, and outcome required for a mediation analysis (eg, studies by Borsari and Carey [36], Kulesza et al [37], Walters et al [39], and Pedersen et al [42]). Mediation occurs in a causal chain involving 2 causal associations, that is, the intervention influences the mediator and the mediator influences the outcomes, so the intervention must precede the mediator and the mediator must precede the outcome [44]. Concerning risk perception, only 1 study to date has, to the best of our knowledge, investigated whether changes in risk perception mediated the effect of BIs on alcohol use [41], but no significant mediation was observed.

This Study

We developed a smartphone app designed to reduce alcohol use among students reporting unhealthy alcohol use [45], and we tested its efficacy in an RCT involving university students reporting unhealthy alcohol use, identified by screening. Providing access to the app (vs not providing access) was associated with significantly fewer standard drinks consumed per week and heavy drinking days (HDDs) per month over the 12-month follow-up period [32]. This study aimed to test whether perceived drinking norms and perceived risks for health mediated the effects of providing access to the app. We hypothesized that perceived drinking norms and perceived risks for health at 3 months would partially mediate the BI's effects on alcohol use at 6 months, controlling for baseline measures. In addition, since PNF is thought to correct individuals' overestimations of other people's drinking, it should primarily affect the norms perception of individuals who overestimate other people's drinking. Thus, our secondary hypothesis was that mediation through norms perception would primarily occur among individuals who overestimate other people's drinking (ie, moderated mediation). To the best of our knowledge, this is the first study to assess moderated mediation in such a setting.

Methods

Overview

This manuscript adheres to the recommendations outlined in A Guideline for Reporting Mediation Analyses (AGReMA) [46]. This study was a secondary analysis of an RCT estimating the effects of providing access to a smartphone app to reduce alcohol use among university students with self-reported unhealthy alcohol use, identified by screening. The results of the primary analysis of this RCT were reported elsewhere [32] in accordance with the CONSORT (Consolidated Standards of Reporting Trials) reporting guidelines [47]. The only deviation from the planned mediation analysis published in the institutional review board–approved protocol is the use of the Mplus software (Muthén & Muthén) instead of SPSS (IBM Corp) with the process macro, although models are equivalent.

Recruitment and Sample

All participants were recruited in April 2021 at 4 higher education institutions in the Lausanne area in Switzerland (the University of Lausanne, the Swiss Federal Institute of Technology Lausanne, the Hospitality Business School Lausanne, and the University of Applied Sciences and Arts Western Switzerland's School of Health) and provided their written informed consent to participate. The study's procedures have been detailed elsewhere [11,32,45]. Each institution promoted the study through its official communication channels (eg, information screens and hallway posters) and student associations' websites and social media. Students were invited to visit the study's specially developed website, where they could complete an anonymous questionnaire assessing their eligibility to participate, that is, being a student at the moment of recruitment, being ≥18 years old, owning a smartphone, and screening positive for unhealthy alcohol use (defined as an Alcohol Use Disorders Identification Test–Consumption score ≥4 for males and ≥3 for females [48-50]). The sample size computation was done for the evaluation of intervention effects at 6 months on the primary outcome of the RCT [32], not for the evaluation of the mediation hypothesis. Of the 3714 students who completed the eligibility questionnaire, 2364 were eligible to participate. Baseline assessments were completed by 1770 students, who were subsequently included in the study and were directly randomized in the study website to either the intervention (n=884) or the control group (n=886). This procedure ensured a total concealment of allocation. Follow-up assessments were done electronically, ensuring blinding. Participants were followed up for 12 months until May 2022. This analysis used baseline, 3-month, and 6-month data. Follow-up rates were 96.4% (total: 1706/1770; intervention: 846/884; and control: 860/886) at 3 months and 95.9% (total: 1697/1770; intervention: 846/884; and control: n=851/886) at 6 months.

Intervention

The app's content was based on the existing literature [9,51-55], previous research involving digital interventions conducted by our group [56-60], and input from members of the target population [45]. Its development has been reported elsewhere [11,45]. Briefly, the app comprised 6 modules involving norms perception and risk perception elements.

First, personal feedback on self-reported alcohol consumption, including normative feedback, feedback on the calorific content of the reported consumption, and feedback on health risks. The user's reported alcohol consumption is compared with the alcohol consumption of people of the same sex and age in Switzerland, with an emphasis on the percentage of people drinking less than the user. This normative feedback is presented for the reported number of drinks per week and the frequency of heavy episodic drinking. Swiss population data were used to generate normative feedback based on the Swiss Health Survey [61] and Addiction Monitoring in Switzerland data [62]. The user also receives an indication of the risks associated with their drinking using specific examples (eg, risk of an accident or violence associated with heavy drinking occasions, risk of addiction, sleep disorders, and cancer associated with chronic

XSL•FO
RenderX

heavy drinking). The calorific content of the reported alcohol use is indicated (in kcal), and the total kcal is also presented in "hamburger equivalents" extrapolated to a 3-month period. At the end of the module, the user can choose to set themselves drinking limits by using a link to the goal-setting tool (ie, module 4).

Second, a blood alcohol content computation tool. This module provides a computed estimate of the blood alcohol content reached with the reported consumption and an indication of the risks associated with different levels of blood alcohol content (eg, "with a blood alcohol content of 1 g/L, the risk of a fatal accident is 7 times higher than for 0.5 g/L" and "altered judgement"). The module also computes how long it will take for the alcohol to be eliminated.

Third, a self-monitoring tool. Once the module is activated, the user is invited to report on their drinking daily. The drinking pattern is then presented to the user on a graph, with indications of the recommended drinking limits.

Fourth, a goal-setting tool. This module allows the user to set their drinking limits for 1, 2, 7, or 30 days. Users are then invited to report their drinking daily. They receive a badge when they drink at or below their self-determined drinking limits.

Fifth, a designated driver tool. This module allows user to take pictures of themselves and their friends. The app then randomly picks the picture of the designated sober driver.

Sixth, fact sheets. This module presents fact sheets on alcohol and health (ie, the effects of alcohol on the human body, diseases caused by alcohol, acute and long-term effects of alcohol use on health, addiction, and resources [available treatment options and contacts]). Participants in the intervention group were not given any limitations to their use of the app (they could use it as much or as little as they wished).

Measures

Follow-up assessments were completed 3 and 6 months after baseline. Email invitations containing a personalized link to the questionnaires were sent out to participants.

Outcomes

The primary (ie, weekly DV) and secondary (ie, number of HDDs) outcomes were prespecified in the research protocol [11]. Weekly DV was assessed using a quantity and frequency measure [63]. Drinking frequency was assessed using a question about the average number of days per week on which the individual had used alcohol in the past 30 days. Drinking quantity was evaluated with a question about the average number of standard drinks consumed on drinking days. Pictograms of Swiss standard drinks containing approximately 10-12 grams of pure alcohol were provided as a visual aid. Weekly DVs (ie, the mean number of standard drinks per week over the past 30 days) were the product of drinking frequency and drinking quantity. The number of HDDs was assessed by asking participants about the number of days they had drunk ≥5 drinks for men or ≥4 drinks for women over the past 30 days.

Mediators

Perceived norms for weekly DV were assessed by asking participants to report their perceptions of the drinking quantity and frequency of a typical person of their age and sex. Perceived drinking norms were obtained by multiplying normative perceptions of drinking quantity and frequency. For HDDs, perceived norms were measured by asking participants how often a typical person of their age and sex drank 5 or more (for male) or 4 or more (for female) standard drinks on 1 occasion over a 30-day period. Perceived risks for health were assessed by asking participants to report the extent to which they believed that they would be personally at risk of being hurt or falling sick because of their drinking on a 10-point scale (1=no risk and 10=high risk).

Covariates

Covariates included age and sex, which were assessed at baseline.

Moderator

To determine whether participants overestimated other people's drinking at baseline, they were asked to estimate the percentage of people of their age and sex who drank more than they did (ie, a subjective estimation from 0% to 100%). Then, the objective percentage of people drinking more than the participant was estimated by comparing the participant's baseline DV with that of individuals in the Swiss population of the same age and sex, based on the Swiss Health Survey [61] and on Addiction Monitoring in Switzerland data [62]. Finally, the difference between the subjective and objective estimations of other people's DV was computed. Based on Bertholet et al [17], an overestimation of other people's drinking (coded 1) was defined as a difference between subjective and objective estimations of their DV of >5 percentage points, whereas no overestimation (coded 0) was defined as a difference of ≤5 percentage points.

Statistical Analyses

Descriptive statistics were calculated to characterize the sample in terms of age, sex, DV, HDDs, perceived DV norms, perceived HDD norms, perceived risks for health, and overestimation of other people's drinking. We conducted the mediation and moderated mediation analyses to test our hypotheses using Mplus software, version 8.3 [64]. The model tested is depicted in Figure 1. Parallel mediation analysis was used to test the hypotheses that the intervention effect was mediated through perceived drinking norms and perceived risks for health. Separate models were tested for each outcome (ie, DV and HDDs). The outcome at 6 months was regressed on the mediators (ie, perceived drinking norms and perceived risks for health at 3 months), the intervention (vs control), and the baseline values of the outcome and the mediators. The mediators were regressed on the intervention (vs control), their respective baseline values, and the baseline values of the outcomes. All associations were also adjusted for age and sex. This model assumes that intervention causally influences the mediators and that mediators causally influence the outcome. It enabled the estimation of the intervention's direct effects (ie, adjusted for the mediators) on alcohol use outcomes (path C') and the 2

XSL•FO
RenderX

specific indirect effects, that is, through perceived norms ($A_1 \times B_1$) and perceived risks ($A_2 \times B_2$) at 3 months. The intervention's total effect (C, ie, not adjusted for the mediators) is the sum of its direct and indirect effects ($C=C'+A_1 \times B_1+A_2 \times B_2$). If both the indirect and direct associations are significant, then there is evidence for partial mediation. In contrast, if the indirect association is significant and the direct association is 0 or close to 0, then there is evidence of full mediation. Missing values were handled using full information maximum likelihood. Bootstrapping with 10,000 resamples was used to estimate the 95% bootstrap CIs (95% CI) of the parameter estimates. We also computed the ratio of the indirect effects to the total effect as an indication of the indirect effect's size. As recommended by Hayes [65], this ratio was only computed when total effects were larger than indirect effects and of the same sign.

Figure 1. Schematic representation of the structural model of our brief intervention associated with alcohol use, both directly and indirectly through perceived norms and perceived risks at 3 months. Parallel mediation models included all the solid arrows. Moderated mediation models included all the solid and dashed arrows. Only the black arrows were of interest to compute the total, direct, and indirect effects. Gray arrows were only included for adjustment.

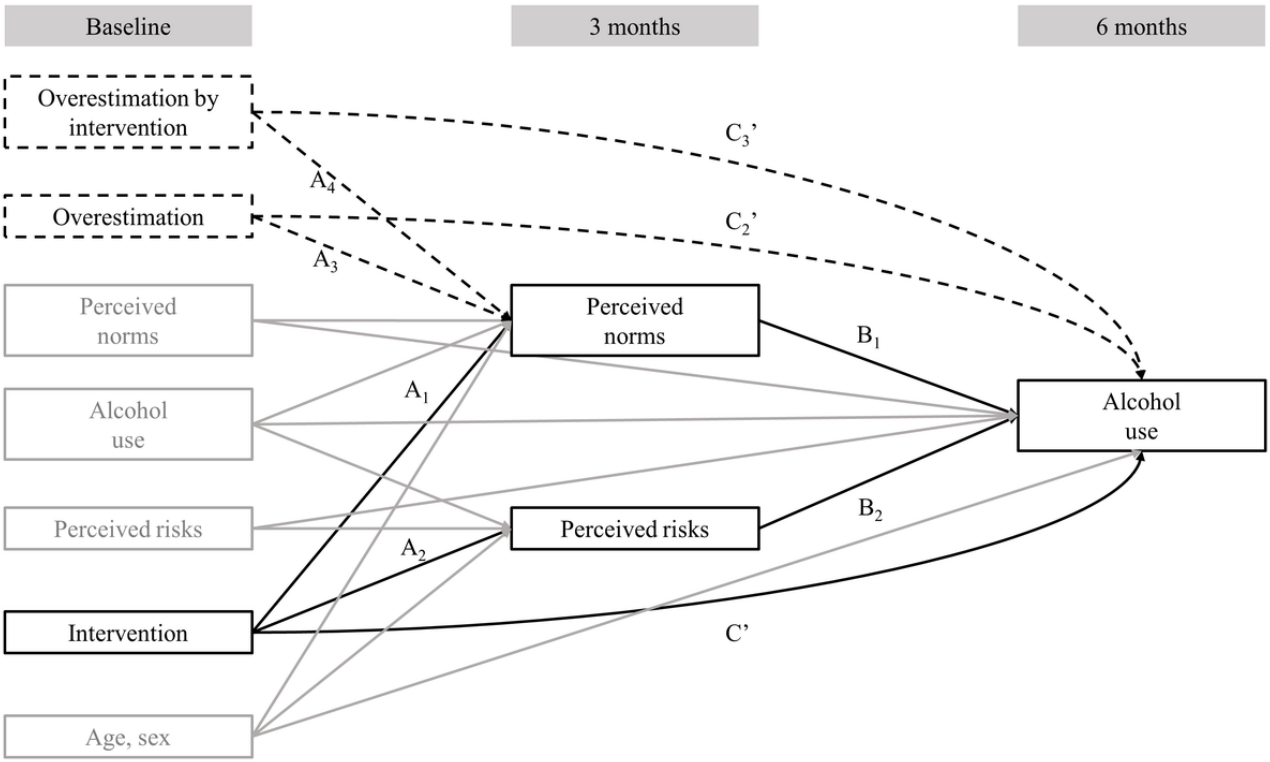

Moderated mediation was used to test the hypothesis that the intervention effect's mediation through perceived norms would differ between individuals who overestimated other people's drinking and individuals who did not overestimate, with mediation primarily expected to be among those individuals overestimating. The moderated mediation model is depicted in Figure 1. Compared with the parallel mediation model described above, the moderated mediation model also tested whether overestimating moderated the A_1 path (from intervention to perceived norms) and the direct path (C') from the intervention to the outcome. For this purpose, the overestimation variables and the interaction between intervention and overestimation were included in the model to predict drinking norms at 3 months and drinking outcomes at 6 months. The indirect effects of the intervention through perceived norms were estimated in the moderator's 2 categories (overestimation and no overestimation). The index of moderated mediation [66], corresponding to the absolute difference between the moderators' 2 categories' indirect effects, was also estimated. A sensitivity analysis was conducted using the same (moderated) mediation models on a restricted sample including all participants in the control group (n=886) and participants of the intervention group who accessed module 1 of the app, which included PNF and PFR (n=468). The results of this sensitivity analysis were similar to those of the main analysis. Coefficients were of a similar magnitude and the most significant coefficients in the main analysis remained significant in the sensitivity analysis, with a few exceptions due to lower statistical power. The details of the sensitivity analysis for mediation and moderated mediation models are reported in Tables S1 and S2 in Multimedia Appendix 1.

Ethical Considerations

The institutional review board protocol was approved by the Human Research Ethics Committee of the Canton of Vaud (number 2018-00560), registered in the ISRCTN registry (ISRCTN10007691), and published before the trial began [11]. The mediation analysis was planned in the Swiss National Science Foundation protocol and the institutional review board–approved protocol, whereas the moderated mediation analysis was not preplanned. All participants received information on the nature of the study, the procedures involved, the expected duration, the potential risks and benefits it may entail, and the financial compensation (gift certificates of up to

XSL•FO
RenderX

50 CHF, equivalent to approximately US $52). They were informed that study participation was voluntary and that they may withdraw from the study at any time without consequences. All participants provided written informed consent before their inclusion in the study. To ensure the confidentiality of data, all data collected on participants were identified with a randomly generated, unique participant identification number. Master lists of participant identification numbers and individually identifiable private information were stored in password-protected computers with restricted access. These lists were available only to senior research staff on this project. The linkage and the direct subject identifiers will be destroyed 10 years after study completion. Subject confidentiality was further ensured by using subject identification code numbers to correspond to treatment data in the computer files. The trial was monitored independently by Lausanne University Hospital's Clinical Trial Unit according to the International Council on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Good Clinical Practice.

Results

Descriptive Characteristics

Figure 2 contains a CONSORT flow diagram. Participants' characteristics and alcohol use are reported in Table 1. The mean participant age was 22.35 (SD 3.07) years, and slightly more than half were female. At baseline, they reported a mean weekly DV of 8.59 (SD 8.18) standard drinks and 3.53 (SD 4.02) HDDs in the previous 30 days. Perceived drinking norms were slightly greater than participants' alcohol consumption, with a mean DV of 8.65 (SD 7.18) standard drinks and 3.67 (SD 2.75) HDDs. On a scale from 1 to 10, participants evaluated the risks to their health associated with their alcohol consumption as relatively low (mean 2.35, SD 1.55). About 78% (1375/1770) of participants overestimated the alcohol consumption of other people of their age and sex in the Swiss population.

Figure 2. CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials.

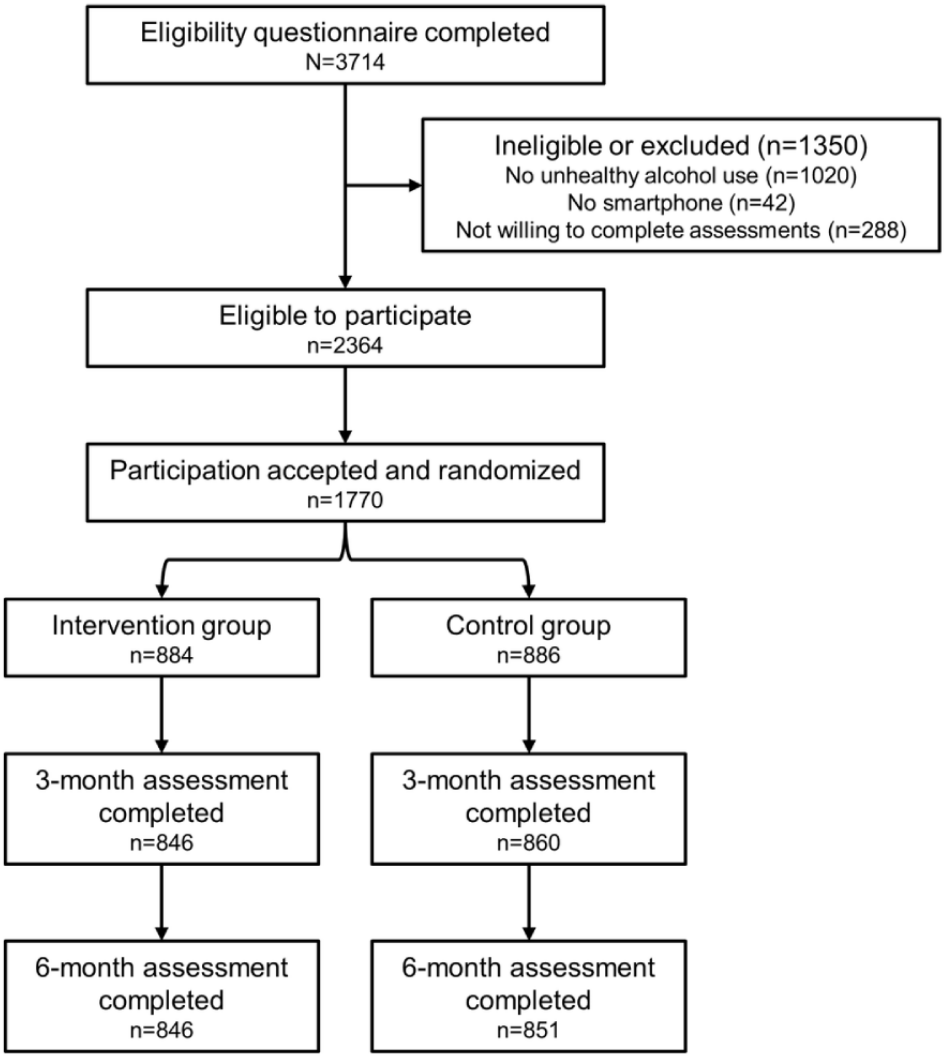

XSL•FO
RenderX

Table 1. Descriptive characteristics.

Characteristics	Control (n=886)	Intervention (n=884)	Total (N=1770)
Sex, n (%)			
Male	393 (44.4)	419 (47.4)	812 (45.9)
Female	493 (55.6)	465 (52.6)	958 (54.1)
Age (years), mean (SD)	22.45 (3.27)	22.24 (2.85)	22.35 (3.07)
Weekly DV[a], mean (SD)			
Baseline	8.25 (7.65)	8.93 (8.66)	8.59 (8.18)
6 months	7.68 (8.11)	7.11 (6.29)	7.39 (7.26)
HDDs[b], mean (SD)			
Baseline	3.48 (3.83)	3.58 (4.19)	3.53 (4.02)
6 months	3.39 (3.65)	3.02 (3.18)	3.21 (3.43)
Perceived norms of weekly DV, mean (SD)			
Baseline	8.72 (6.23)	8.57 (8.03)	8.65 (7.18)
3 months	9.77 (6.89)	9.02 (6.01)	9.40 (6.48)
Perceived norms of HDDs, mean (SD)			
Baseline	3.73 (2.92)	3.61 (2.57)	3.67 (2.75)
3 months	4.22 (3.24)	3.96 (3.04)	4.09 (3.14)
Perceived risks for health, mean (SD)			
Baseline	2.36 (1.56)	2.34 (1.54)	2.35 (1.55)
3 months	2.40 (1.58)	2.43 (1.56)	2.42 (1.57)
Overestimation, n (%)			
No	198 (22.3)	197 (22.3)	395 (22.3)
Yes	688 (77.7)	687 (77.7)	1375 (77.7)

[a]DV: drinking volume.

[b]HDD: heavy drinking day.

Mediation Models

The results of our mediation models are reported in Table 2. Regarding DV, the intervention's total effect (C) was significant (b=–0.85, 95% CI –1.49 to –0.25), indicating a lower DV at 6 months in the intervention group than in the control group. The intervention's direct effect (C', ie, its effect adjusted for mediators) was also significant but smaller (b=–0.73, 95% CI –1.33 to –0.16). The intervention's indirect effect was significant through the drinking norms mediator (b=–0.12, 95% CI –0.25 to –0.03). The intervention was associated with lower estimations of drinking norms at 3 months (b=–0.80, 95% CI –1.36 to –0.22), and estimated drinking norms at 3 months were associated with drinking at 6 months (b=0.15, 95% CI 0.08 to 0.24). The intervention's indirect effect through estimated drinking norms at 3 months accounted for 14% (ratio of specific indirect effect/total effect=–0.12/–0.85) of the total effect. The intervention's indirect effect through risk perception was not significant (b=0.00; 95% CI –0.03 to 0.04), and neither was the association between the intervention and perceived risks at 3 months (b=0.02, 95% CI –0.11 to 0.15) or the association between perceived risks at 3 months and DV at 6 months (b=0.20, 95% CI –0.13 to 0.47).

The intervention's total effect (C) for HDDs was significant (b=–0.44, 95% CI –0.72 to –0.16), indicating fewer HDDs at 6 months in the intervention group than in the control group. The intervention's direct effect (C') was also significant but smaller (b=–0.39, 95% CI –0.66 to –0.12). However, the intervention's indirect effect through drinking norms at 3 months (b=–0.05, 95% CI –0.11 to 0.004) and its indirect effect through risk perception (b=0.00, 95% CI –0.01 to 0.02) were not significant.

XSL•FO
RenderX

Table 2. Results of parallel mediation models (all models were adjusted for age, sex, and the baseline values of the outcomes and the mediators).

Path of association	DV[a] at 6 months		HDD[b] at 6 months	
	b (SE[c])	95% CI[d]	b (SE)	95% CI
Perceived norms as mediator				
Intervention to mediator (A)	–0.80[e] (0.29)	–1.36 to –0.22	–0.26 (0.07)	–0.53 to 0.02
Mediator to outcome (B)	0.15[e] (0.04)	0.08 to 0.24	0.19[e] (0.04)	0.12 to 0.26
Specific indirect effect	–0.12[e] (0.06)	–0.25 to –0.03	–0.05 (0.03)	–0.11 to 0.004
Size of specific indirect effect (ratio of specific indirect effect/total effect)	0.14		0.11	
Perceived risks as mediator				
Intervention to mediator (A)	0.02 (0.07)	–0.11 to 0.15	0.02 (0.06)	–0.11 to 0.15
Mediator to outcome (B)	0.20 (0.15)	–0.13 to 0.47	0.07 (0.06)	–0.06 to 0.19
Specific indirect effect	0.00 (0.01)	–0.03 to 0.04	0.00 (0.00)	–0.01 to 0.02
Size of specific indirect effect (ratio of specific indirect effect/total effect)	0.00		0.00	
Total indirect effect	–0.12[e] (0.06)	–0.25 to –0.02	–0.05 (0.03)	–0.11 to 0.01
Size of total indirect effect (ratio of total indirect effect/total effect)	0.14		0.11	
Direct effect (C')	–0.73[e] (0.30)	–1.33 to –0.16	–0.39[e] (0.14)	–0.66 to –0.12
Total effect (C)	–0.85[e] (0.32)	–1.49 to –0.25	–0.44[e] (0.14)	–0.72 to –0.16

[a]DV: drinking volume.

[b]HDD: heavy drinking day.

[c]SE: robust standard error.

[d]CI: bootstrap confidence interval.

[e]Coefficients are significant based on 95% CI.

Moderated Mediation Models

The results of our moderated mediation models, reported in Table 3 and Figure 3, showed that overestimations of other people's drinking moderated the association between the intervention and drinking norms at 3 months, in both the DV (b=–1.38, 95% CI –2.82 to –0.03) and HDDs (b=–0.63, 95% CI –1.24 to –0.05) models. The index of moderated mediation was also significant for both DV (b=–0.21, 95% CI –0.49 to –0.01) and HDDs (b=–0.12, 95% CI –0.25 to –0.01), indicating differences in the indirect effects between individuals who overestimated other people's drinking and those who did not. More specifically, the intervention's indirect effects on DV and HDDs at 6 months through drinking norms at 3 months was larger (as an absolute value) among individuals who overestimated other people's drinking (DV: b=–0.17, 95% CI –0.32 to –0.05; HDD: b=–0.08, 95% CI –0.15 to –0.01) than among those who did not (DV: b=0.04, 95% CI –0.14 to 0.26; HDD: b=0.05, 95% CI –0.05 to 0.15), and it was only significant among those who had overestimated. Among individuals who had overestimated other people's drinking, the indirect effect of the intervention on DV and HDDs through drinking norms accounted for 17% and 19% of the total effect, respectively.

XSL•FO
RenderX

Table 3. Results of moderated mediation models (all models were adjusted for age, sex, and the baseline values of the outcomes and the mediators).

Path of association[a]	DV[b] at 6 months		HDD[c] at 6 months	
	B (SE[d])	95% CI[e]	B (SE)	95% CI
Intervention to perceived norms at 3 months (A_1)	0.27 (0.62)	–0.89 to 1.55	0.23 (0.26)	–0.26 to 0.76
Overestimation to perceived norms at 3 months (A_3)	0.98 (0.53)	–0.13 to 1.96	0.50[f] (0.19)	0.12 to 0.88
Overestimation × intervention to perceived norms at 3 months (A_4)	–1.38[f] (0.70)	–2.82 to –0.03	–0.63[f] (0.30)	–1.24 to –0.05
Perceived norms at 3 months to outcome at 6 months (B_1)	0.15[f] (0.04)	0.08 to 0.24	0.19[f] (0.04)	0.12 to 0.26
Intervention to outcome at 6 months (C')	–0.38 (0.59)	–1.55 to 0.77	–0.52 (0.29)	–1.09 to 0.04
Overestimation to outcome at 6 months (C_2')	0.02 (0.48)	–0.97 to 0.93	–0.22 (0.26)	–0.77 to 0.28
Overestimation × intervention to outcome at 6 months (C_3')	–0.46 (0.67)	–1.76 to 0.83	0.18 (0.33)	–0.47 to 0.82
Conditional effects of the intervention				
No overestimation				
Indirect effect (through perceived norms at 3 months)	0.04 (0.10)	–0.14 to 0.26	0.05 (0.05)	–0.05 to 0.15
Size of the indirect effect (ratio of indirect effect/total effect)	—[g]		—	
Direct effect	–0.38 (0.59)	–1.56 to 0.77	–0.52 (0.29)	–1.09 to 0.04
Total effect	–0.34 (0.60)	–1.52 to 0.87	–0.48 (0.29)	–1.04 to 0.08
Overestimation				
Indirect effect (through perceived norms at 3 months)	–0.17[f] (0.07)	–0.32 to –0.05	–0.08[f] (0.03)	–0.15 to –0.01
Size of the indirect effect (ratio of indirect effect/total effect)	0.17		0.19	
Direct effect	–0.83[f] (0.34)	–1.50 to –0.18	–0.35[f] (0.16)	–0.66 to –0.04
Total effect	–1.00[f] (0.36)	–1.72 to –0.31	–0.42[f] (0.17)	–0.74 to –0.10
Index of moderation	–0.21[f] (0.12)	–0.49 to –0.01	–0.12[f] (0.06)	–0.25 to –0.01

[a]Indirect effects mediated by risk perception were estimated in the model but not reported.

[b]DV: drinking volume.

[c]HDD: heavy drinking day.

[d]SE: robust standard error.

[e]CI: bootstrap confidence interval.

[f]Coefficients are significant based on 95% CI.

[g]Not applicable.

XSL•FO
RenderX

Figure 3. Results of moderated mediation models. An asterisk indicates a coefficient significant based on 95% bootstrap CI. DV: drinking volume; HDD: heavy drinking day.

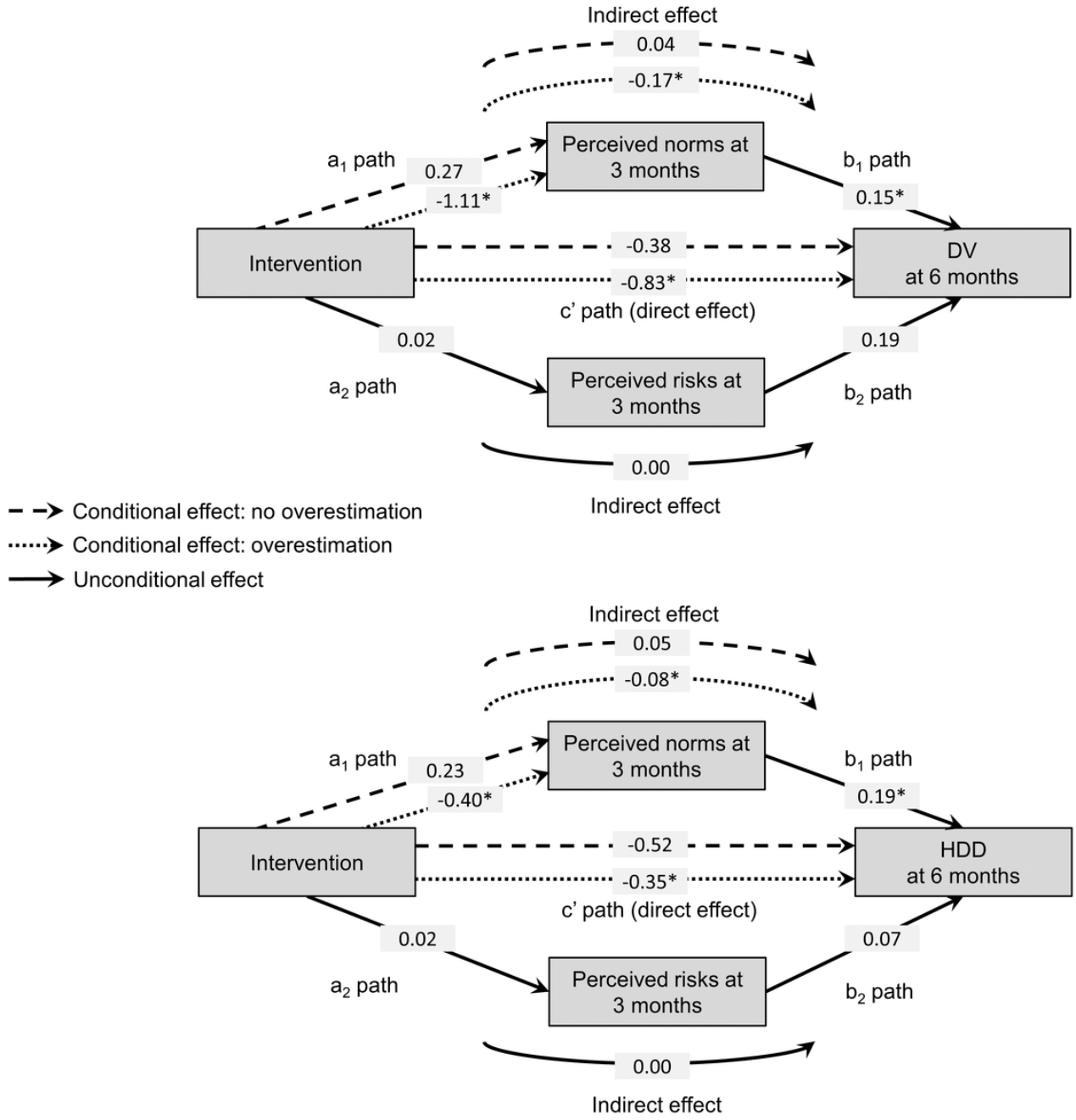

Discussion

Principal Findings

This study's results showed that the effects of a smartphone-based intervention to reduce alcohol use were partially mediated by perceived drinking norms but not by its perceived risks. Specifically, the intervention was significantly associated with lower perceived drinking norms at 3 months, and perceived drinking norms at 3 months were associated with weekly DV at 6 months. The same pattern of association was observed for HDDs, although the indirect effect mediated by perceived drinking norms was not significant. This finding is in line with results from previous studies showing that perceived norms mediated the effects of face-to-face [36-38] and digital [39-43] BIs on alcohol use.

Furthermore, the intervention's indirect effects on weekly DV and HDDs at 6 months through perceived drinking norms at 3 months, were only significant among individuals who had overestimated other people's alcohol use. This finding is consistent with the theoretical bases of PNF, as it is primarily expected to influence an individual's drinking behaviors by correcting their overestimation of other people's drinking and encouraging them to adopt drinking behaviors that are more aligned with actual norms [20]. Thus, for individuals who did not overestimate other people's drinking, PNF is not expected to affect their perceptions of norms, which was consistently observed throughout this study.

XSL•FO
RenderX

Taken together, these results confirmed one of the hypothesized mechanisms of action of BIs [20]. They extended previous studies supporting the use of PNF in both internet-based [39-42] and face-to-face [36,37] BIs and lent support to its application in smartphone-based BIs to lower unhealthy alcohol use. PNF appeared to partially mediate the intervention's effects when individuals overestimated other people's drinking. As such, providing participants with information that their drinking was heavier than the norm had a conditional impact on those who misperceived their drinking with respect to other people's drinking. Nevertheless, we found no evidence of full mediation. Among individuals who overestimated other people's drinking, the mediation of perceived drinking norms accounted for 17% and 19% of the intervention's effect on DV and HDDs, respectively. Although this accounts for only a limited portion of the intervention's total effect, these results support the use of normative feedback interventions as an intervention option for secondary prevention interventions aiming at decreasing alcohol-related harm in the population. This type of digital intervention has the potential to be scaled up and reach a wide audience at a low cost, contributing to public health efforts in mitigating alcohol-related harm. It is among the NICE (National Institute for Health and Care Excellence) recommended options to address unhealthy alcohol use [67]. The remaining portion of the intervention's effect not explained by drinking norms leaves room for other mechanisms of action contained within the intervention. The app, developed previously, included other modules that may have played a role in the intervention's effects, such as setting drinking goals and monitoring one's own alcohol use [45], but their hypothesized mechanisms of action were not assessed or tested. Also, using the app may have led participants to reassess their alcohol use and their relationship with alcohol through other mechanisms not evaluated here. Assessment reactivity may also have exerted an effect on the participants in both study groups, but because the intervention group's participants had the ability to repeatedly assess their alcohol use, they may have been more disposed to changing their drinking behavior for this reason [68]. This warrants further studies to better understand the mechanisms of action of screening and BIs. Nonetheless, this type of intervention should be seen as one contributing piece in a comprehensive public health approach including, as recommended by the SAFER (Strengthen restrictions on alcohol availability; Advance and enforce drink driving countermeasures; Facilitate access to screening, brief interventions, and treatment; Enforce bans or comprehensive restrictions on alcohol advertising, sponsorship, and promotion; Raise prices on alcohol through excise taxes and pricing policies) initiative [69], contextual and structural prevention measures, such as restrictions on the availability and ban on advertising.

The lack of any mediation effect by perceived risks has to be interpreted in light of the study population. In line with results showing that risk perceptions for health related to alcohol use are lower in younger than in older adults [70], perceived risks (and likely actual risks) were low in our sample of young adults. Using the same question, the mean risk perception in our sample was approximately half of that reported in a North American sample of older participants reporting unhealthy drinking (mean age 37.9, SD 9.7 years) [71]. This suggests that young adults may be less concerned by the potential alcohol-related consequences on health than middle-aged and older adults because the potential consequences seem so far away. Thus, the role of perceived risks in BIs may depend on the age of the individuals involved. One alternative explanation for the lack of mediation by perceived risks is that digital interventions may not provide the same level of interaction, emotional engagement, and immediate clarification as face-to-face interventions, which can make PFR more impactful. In digital formats, the static nature of risk messages may reduce their perceived relevance and impact, making it harder to convey the seriousness of health risks compared with a more dynamic and responsive face-to-face setting. Further studies should replicate these analyses in other populations and explore interventions that deliver more dynamic and personalized risk feedback, such as using teleconsultations or adaptive messaging, to better engage users and enhance the perceived relevance of health risks.

Strengths and Limitations

This study had some limitations related to the use of self-reported measures, which can be susceptible to underreporting and social desirability bias, notably when norms are provided. Nonetheless, we used standardized measures, and the intervention's effects in the main trial were consistent for both the outcomes that provided us with normative feedback and the outcomes that did not [32]. The study was conducted among students, who may have been more able to deal with the complex cognitive process of comparing one's own drinking with other people's drinking, and better able to infer a percentage of people drinking more than oneself. Interventions with personalized feedback may be less effective in nonacademic settings [72]. Thus, these results need to be replicated in other settings. In addition, we used national drinking norms based on sex. Because of biological differences between the sexes in alcohol pharmacokinetics and alcohol-related risks and consequences [73], providing information based on sex is relevant. Nonetheless, some participants may consider gender norms more relevant to them and may not find sex norms adequate. As the perceived validity of norms is crucial to normative feedback being considered relevant, and as data on drinking norms by gender are currently lacking, this will likely be a challenge for future interventions. Furthermore, our sample of students with unhealthy drinking habits may be surrounded by individuals who consume more alcohol than those in the general population. As a result, our participants might have been skeptical of the information provided by PNF, which is based on nationally representative data for individuals of the same age and sex, as it may not reflect their personal experiences. Promising alternative approaches, such as using dynamic norms (as opposed to static norms) [74,75] or emphasizing deviant versus normative behaviors depending on the context and prevalence of the behavior [76], have been proposed. Evaluating their effectiveness in future interventions is warranted. The distribution of the HDD and DV outcomes, as well as perceived drinking norms, were right-skewed, which may violate the normality assumption of linear regression. To mitigate this issue, bootstrap resampling was used to provide accurate CIs without relying on parametric assumptions [77].

XSL•FO
RenderX

The study has some notable strengths. The mediation hypotheses were preplanned, and outcomes and mediators were assessed sequentially, ensuring proper temporal ordering so that the intervention occurred before the mediators, which acted before the outcomes. The mediators were also measured at baseline, allowing for the adjustment of baseline values. We were able to conduct a mediation analysis in an RCT that demonstrated the intervention's beneficial impact. Because of the nature of the mediators studied, they could be assessed in both groups with a limited risk of assessment reactivity.

Conclusions

This study's findings lend support to 1 of the 2 hypothesized mechanisms of action of the digital BI tested—the mechanism of PNF but not the mechanism of feedback on risks. This supports the use of PNF in similar interventions. Nonetheless, only a limited portion of the intervention's effects could be explained. Further research is therefore needed to identify other mechanisms of action of digital BIs.

Acknowledgments

This study was funded by Swiss National Science Foundation (SNSF) grant 176295. The SNSF had no involvement in the conduct of the study, the writing of the manuscript, or the decision to submit it for publication. We wish to thank the study participants, the research assistants who worked on the study (Stéphanie Perler, Alizée Dind, Lucie Escasain, Angeles Niembro Scaccianoce, Tajnja Vasileva, and Flavia Von Xylander), and Jean Szabo (netinfluence) for his work on the smartphone app and the study website.

Data Availability

The data analyzed in this study are available upon reasonable request to the corresponding author. Access will be granted following an approved data-sharing agreement and in compliance with ethical and institutional guidelines.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 27 KB - jmir_v27i1e55541_app1.docx]

References

1. Mokdad AH, Forouzanfar M, Daoud F, Mokdad AA, El Bcheraoui C, Moradi-Lakeh M, et al. Global burden of diseases, injuries, and risk factors for young people's health during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2016 Jun 11;387(10036):2383-2401. [doi: 10.1016/S0140-6736(16)00648-6] [Medline: 27174305]
2. Perkins HW. Surveying the damage: a review of research on consequences of alcohol misuse in college populations. J Stud Alcohol 2002;Supplement(14):91-100. [doi: 10.15288/jsas.2002.s14.91] [Medline: 12022733]
3. Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. J Stud Alcohol Drugs 2009;Supplement(16):12-20 [FREE Full text] [doi: 10.15288/jsads.2009.s16.12] [Medline: 19538908]
4. Hingson R, Zha W, Smyth D. Magnitude and trends in heavy episodic drinking, alcohol-impaired driving, and alcohol-related mortality and overdose hospitalizations among emerging adults of college ages 18-24 in the United States, 1998-2014. J Stud Alcohol Drugs 2017;78(4):540-548 [FREE Full text] [doi: 10.15288/jsad.2017.78.540] [Medline: 28728636]
5. Patrick ME, Terry-McElrath YM, Evans-Polce RJ, Schulenberg JE. Negative alcohol-related consequences experienced by young adults in the past 12 months: differences by college attendance, living situation, binge drinking, and sex. Addict Behav 2020;105:106320 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106320] [Medline: 32007832]
6. World Health Organization. Global Strategy to Reduce the Harmful use of Alcohol. Geneva: World Health Organization; 2010.
7. Moyer VA, Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: U.S. preventive services task force recommendation statement. Ann Intern Med 2013;159(3):210-218 [FREE Full text] [doi: 10.7326/0003-4819-159-3-201308060-00652] [Medline: 23698791]
8. National Institute for Health and Care Excellence. Alcohol-use disorders: prevention (NICE guideline ph24). 2010. URL: https://www.nice.org.uk/guidance/ph24/resources/alcoholuse-disorders-prevention-pdf-1996237007557 [accessed 2025-01-07]
9. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017;9(9):CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]

XSL•FO
RenderX

10. Kaner E, Beyer F, Muirhead C, Campbell F, Pienaar E, Bertholet N, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2018;2(2):CD004148 [FREE Full text] [doi: 10.1002/14651858.CD004148.pub4] [Medline: 29476653]
11. Bertholet N, Schmutz E, Grazioli VS, Faouzi M, McNeely J, Gmel G, et al. Smartphone-based secondary prevention intervention for university students with unhealthy alcohol use identified by screening: study protocol of a parallel group randomized controlled trial. Trials 2020;21:191 [FREE Full text] [doi: 10.1186/s13063-020-4145-2] [Medline: 32066490]
12. Black N, Mullan B, Sharpe L. Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. Health Psychol Rev 2016;10(3):341-357. [doi: 10.1080/17437199.2016.1168268] [Medline: 26999311]
13. Saxton J, Rodda SN, Booth N, Merkouris SS, Dowling NA. The efficacy of personalized normative feedback interventions across addictions: a systematic review and meta-analysis. PLoS One 2021;16(4):e0248262 [FREE Full text] [doi: 10.1371/journal.pone.0248262] [Medline: 33793583]
14. Berkowitz AD. An overview of the social norms approach. In: Lederman L, Stewart L, editors. Changing the Culture of College Drinking: A Socially Situated Health Communication Campaign. Creskill, NJ: Hampton Press; 2005:193-214.
15. Perkins HW. Misperceptions of peer drinking norms in Canada: another look at the 'reign of error' and its consequences among college students. Addict Behav 2007;32(11):2645-2656. [doi: 10.1016/j.addbeh.2007.07.007] [Medline: 17719724]
16. Perkins HW, Haines MP, Rice R. Misperceiving the college drinking norm and related problems: a nationwide study of exposure to prevention information, perceived norms and student alcohol misuse. J Stud Alcohol 2005;66(4):470-478. [doi: 10.15288/jsa.2005.66.470] [Medline: 16240554]
17. Bertholet N, Gaume J, Faouzi M, Daeppen J, Gmel G. Perception of the amount of drinking by others in a sample of 20-year-old men: the more I think you drink, the more I drink. Alcohol Alcohol 2011;46(1):83-87. [doi: 10.1093/alcalc/agq084] [Medline: 21148145]
18. Bertholet N, Faouzi M, Studer J, Daeppen J, Gmel G. Perception of tobacco, cannabis, and alcohol use of others is associated with one's own use. Addict Sci Clin Pract 2013;8(1):15 [FREE Full text] [doi: 10.1186/1940-0640-8-15] [Medline: 24499600]
19. Neighbors C, Dillard AJ, Lewis MA, Bergstrom RL, Neil TA. Normative misperceptions and temporal precedence of perceived norms and drinking. J Stud Alcohol 2006;67(2):290-299 [FREE Full text] [doi: 10.15288/jsa.2006.67.290] [Medline: 16562412]
20. Miller DT, Prentice DA. Changing norms to change behavior. Annu Rev Psychol 2016;67:339-361. [doi: 10.1146/annurev-psych-010814-015013] [Medline: 26253542]
21. Lewis MA, Neighbors C. Optimizing personalized normative feedback: the use of gender-specific referents. J Stud Alcohol Drugs 2007;68(2):228-237 [FREE Full text] [doi: 10.15288/jsad.2007.68.228] [Medline: 17286341]
22. Lewis MA, Neighbors C. Social norms approaches using descriptive drinking norms education: a review of the research on personalized normative feedback. J Am Coll Health 2006;54(4):213-218 [FREE Full text] [doi: 10.3200/JACH.54.4.213-218] [Medline: 16450845]
23. Labrie JW, Lewis MA, Atkins DC, Neighbors C, Zheng C, Kenney SR, et al. RCT of web-based personalized normative feedback for college drinking prevention: are typical student norms good enough? J Consult Clin Psychol 2013;81(6):1074-1086 [FREE Full text] [doi: 10.1037/a0034087] [Medline: 23937346]
24. Earle AM, LaBrie JW, Boyle SC, Smith D. In pursuit of a self-sustaining college alcohol intervention: deploying gamified PNF in the real world. Addict Behav 2018;80:71-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.005] [Medline: 29407688]
25. Boyle SC, Earle AM, LaBrie JW, Smith DJ. PNF 2.0? Initial evidence that gamification can increase the efficacy of brief, web-based personalized normative feedback alcohol interventions. Addict Behav 2017;67:8-17 [FREE Full text] [doi: 10.1016/j.addbeh.2016.11.024] [Medline: 27978426]
26. Rogers RW. Cognitive and physiological processes in fear appeals and attitude change: a revised theory of protection motivation. In: Cacioppo J, Petty R, editors. Social Psychophysiology. New York: Guilford Press; 1983:153-176.
27. Ronis DL. Conditional health threats: health beliefs, decisions, and behaviors among adults. Health Psychol 1992;11(2):127-134. [doi: 10.1037//0278-6133.11.2.127] [Medline: 1582381]
28. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
29. Ferrer RA, Klein WM. Risk perceptions and health behavior. Curr Opin Psychol 2015;5:85-89 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.012] [Medline: 26258160]
30. Cunningham JA, Neighbors C, Wild C, Humphreys K. Ultra-brief intervention for problem drinkers: research protocol. BMC Public Health 2008;8:298 [FREE Full text] [doi: 10.1186/1471-2458-8-298] [Medline: 18727823]
31. Cunningham JA, Wild TC, Bondy SJ, Lin E. Impact of normative feedback on problem drinkers: a small-area population study. J Stud Alcohol 2001;62(2):228-233. [doi: 10.15288/jsa.2001.62.228] [Medline: 11327189]
32. Bertholet N, Schmutz E, Studer J, Adam A, Gmel G, Cunningham JA, et al. Effect of a smartphone intervention as a secondary prevention for use among university students with unhealthy alcohol use: randomised controlled trial. BMJ 2023;382:e073713 [FREE Full text] [doi: 10.1136/bmj-2022-073713] [Medline: 37586742]

XSL•FO
RenderX

33. Cunningham JA, Godinho A, Schell C, Studer J, Wardell JD, Garnett C, et al. Randomized controlled trial of a smartphone app designed to reduce unhealthy alcohol consumption. Internet Interv 2024;36:100747 [FREE Full text] [doi: 10.1016/j.invent.2024.100747] [Medline: 38812955]
34. Kypri K, McCambridge J, Vater T, Bowe SJ, Saunders JB, Cunningham JA, et al. Web-based alcohol intervention for Māori university students: double-blind, multi-site randomized controlled trial. Addiction 2013;108(2):331-338 [FREE Full text] [doi: 10.1111/j.1360-0443.2012.04067.x] [Medline: 22925046]
35. Lewis MA, Neighbors C, Oster-Aaland L, Kirkeby BS, Larimer ME. Indicated prevention for incoming freshmen: personalized normative feedback and high-risk drinking. Addict Behav 2007;32(11):2495-2508 [FREE Full text] [doi: 10.1016/j.addbeh.2007.06.019] [Medline: 17658695]
36. Borsari B, Carey KB. Effects of a brief motivational intervention with college student drinkers. Journal of Consulting and Clinical Psychology 2000;68(4):728-733. [doi: 10.1037/0022-006x.68.4.728]
37. Kulesza M, McVay MA, Larimer ME, Copeland AL. A randomized clinical trial comparing the efficacy of two active conditions of a brief intervention for heavy college drinkers. Addict Behav 2013;38(4):2094-2101. [doi: 10.1016/j.addbeh.2013.01.008] [Medline: 23410849]
38. Wood MD, Capone C, Laforge R, Erickson DJ, Brand NH. Brief motivational intervention and alcohol expectancy challenge with heavy drinking college students: a randomized factorial study. Addict Behav 2007;32(11):2509-2528. [doi: 10.1016/j.addbeh.2007.06.018] [Medline: 17658696]
39. Walters ST, Vader AM, Harris TR. A controlled trial of web-based feedback for heavy drinking college students. Prev Sci 2007;8(1):83-88. [doi: 10.1007/s11121-006-0059-9] [Medline: 17136461]
40. Neighbors C, Lewis MA, Atkins DC, Jensen MM, Walter T, Fossos N, et al. Efficacy of web-based personalized normative feedback: a two-year randomized controlled trial. J Consult Clin Psychol 2010;78(6):898-911 [FREE Full text] [doi: 10.1037/a0020766] [Medline: 20873892]
41. Williams J, Herman-Stahl M, Calvin SL, Pemberton M, Bradshaw M. Mediating mechanisms of a military web-based alcohol intervention. Drug Alcohol Depend 2009;100(3):248-257. [doi: 10.1016/j.drugalcdep.2008.10.007] [Medline: 19081206]
42. Pedersen ER, Parast L, Marshall GN, Schell TL, Neighbors C. A randomized controlled trial of a web-based, personalized normative feedback alcohol intervention for young-adult veterans. J Consult Clin Psychol 2017;85(5):459-470 [FREE Full text] [doi: 10.1037/ccp0000187] [Medline: 28287799]
43. Neighbors C, Larimer ME, Lewis MA. Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. J Consult Clin Psychol 2004;72(3):434-447 [FREE Full text] [doi: 10.1037/0022-006X.72.3.434] [Medline: 15279527]
44. Cole DA, Maxwell SE. Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. J Abnorm Psychol 2003;112(4):558-577. [doi: 10.1037/0021-843X.112.4.558] [Medline: 14674869]
45. Bertholet N, Schmutz E, Cunningham JA, McNeely J, Gmel G, Daeppen J, et al. Development of a secondary prevention smartphone app for students with unhealthy alcohol use: results from a qualitative assessment. JMIR Hum Factors 2023;10:e41088 [FREE Full text] [doi: 10.2196/41088] [Medline: 36881448]
46. Lee H, Cashin AG, Lamb SE, Hopewell S, Vansteelandt S, VanderWeele TJ, AGReMA group, et al. A guideline for reporting mediation analyses of randomized trials and observational studies: the AGReMA statement. JAMA 2021;326(11):1045-1056 [FREE Full text] [doi: 10.1001/jama.2021.14075] [Medline: 34546296]
47. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
48. Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res 2007;31(7):1208-1217. [doi: 10.1111/j.1530-0277.2007.00403.x] [Medline: 17451397]
49. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory care quality improvement project (ACQUIP). Alcohol use disorders identification test. Arch Intern Med 1998;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
50. Khadjesari Z, White IR, McCambridge J, Marston L, Wallace P, Godfrey C, et al. Validation of the AUDIT-C in adults seeking help with their drinking online. Addict Sci Clin Pract 2017;12(1):2 [FREE Full text] [doi: 10.1186/s13722-016-0066-5] [Medline: 28049515]
51. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
52. Garnett C, Crane D, Brown J, Kaner E, Beyer F, Muirhead C, et al. Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: meta-regression. J Med Internet Res 2018;20(2):e69 [FREE Full text] [doi: 10.2196/jmir.8807] [Medline: 29490895]
53. Orji R, Vassileva J, Mandryk R. Towards an effective health interventions design: an extension of the health belief model. Online J Public Health Inform 2012;4(3):e61050 [FREE Full text] [doi: 10.5210/ojphi.v4i3.4321] [Medline: 23569653]

XSL•FO
RenderX

54. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and 'punishy': Exploring the costs and benefits of self-monitoring for health and wellness. Digit Health 2018;4:2055207618797554 [FREE Full text] [doi: 10.1177/2055207618797554] [Medline: 30202544]
55. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
56. Bertholet N, Daeppen J, McNeely J, Kushnir V, Cunningham JA. Smartphone application for unhealthy alcohol use: a pilot study. Subst Abus 2017;38(3):285-291. [doi: 10.1080/08897077.2017.1281860] [Medline: 28113039]
57. Bertholet N, Godinho A, Cunningham JA. Smartphone application for unhealthy alcohol use: pilot randomized controlled trial in the general population. Drug Alcohol Depend 2019;195:101-105. [doi: 10.1016/j.drugalcdep.2018.12.002] [Medline: 30611977]
58. Bertholet N, Cunningham JA, Faouzi M, Gaume J, Gmel G, Burnand B, et al. Internet-based brief intervention for young men with unhealthy alcohol use: a randomized controlled trial in a general population sample. Addiction 2015;110(11):1735-1743 [FREE Full text] [doi: 10.1111/add.13051] [Medline: 26173842]
59. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. Addiction 2009;104(12):2023-2032 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02726.x] [Medline: 19922569]
60. Cunningham JA. Comparison of two internet-based interventions for problem drinkers: randomized controlled trial. J Med Internet Res 2012;14(4):e107 [FREE Full text] [doi: 10.2196/jmir.2090] [Medline: 22954459]
61. Swiss Federal Statistical Office. Swiss Health Survey. Neuchatel: Swiss Federal Statistical Office; 2017. URL: https://www.bfs.admin.ch/bfs/fr/home/statistiques/sante/enquetes/sgb.html [accessed 2025-01-07]
62. Gmel G, Kuendig H, Notari L, Gmel C. Monitorage suisse des addictions: consommation d'alcool, de tabac et de drogues illégales en Suisse en 2016 [Addiction monitoring in Switzerland: Use of alcohol, tobacco and illicit drugs in 2016] [Report in French]. Lausanne: Addiction Suisse [Addiction Switzerland]; 2016. URL: https://www.addictionsuisse.ch/wp-content/uploads/library_sumo/gmel_2tl0igxj1e6t.pdf [accessed 2025-01-26]
63. Rehm J. Measuring quantity, frequency, and volume of drinking. Alcoholism Clin & Exp Res 2006;22(S2):4s-14s. [doi: 10.1111/j.1530-0277.1998.tb04368.x]
64. Muthén LK, Muthén BO. Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 2017.
65. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: The Guilford Press; 2013.
66. Hayes AF. An index and test of linear moderated mediation. Multivariate Behav Res 2015;50(1):1-22. [doi: 10.1080/00273171.2014.962683] [Medline: 26609740]
67. Behaviour change: digital and mobile health interventions, NICE guideline [NG183]. National Institute for Health and Care Excellence. 2020. URL: https://www.nice.org.uk/guidance/ng183/chapter/Recommendations#alcohol-use [accessed 2024-11-25]
68. McCambridge J, Kypri K. Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials. PLoS One 2011;6(10):e23748 [FREE Full text] [doi: 10.1371/journal.pone.0023748] [Medline: 21998626]
69. The SAFER Technical Package: Five Areas of Intervention at National and Subnational Levels. Geneva: World Health Organization; 2019.
70. Bujalski M, Stokwiszewski J. Alcohol risk perception and the risk target. Results from a survey among the Polish adult population. Addiction Research & Theory 2022;30(4):253-261. [doi: 10.1080/16066359.2021.2006646]
71. Cunningham JA, Humphreys K, Kypri K, van Mierlo T. Formative evaluation and three-month follow-up of an online personalized assessment feedback intervention for problem drinkers. J Med Internet Res 2006;8(2):e5 [FREE Full text] [doi: 10.2196/jmir.8.2.e5] [Medline: 16867968]
72. Davis JP, Smith DC, Briley DA. Substance use prevention and treatment outcomes for emerging adults in non-college settings: a meta-analysis. Psychol Addict Behav 2017;31(3):242-254. [doi: 10.1037/adb0000267] [Medline: 28318279]
73. Erol A, Karpyak VM. Sex and gender-related differences in alcohol use and its consequences: contemporary knowledge and future research considerations. Drug Alcohol Depend 2015;156:1-13. [doi: 10.1016/j.drugalcdep.2015.08.023] [Medline: 26371405]
74. Lee SJ, Liu J. Leveraging dynamic norm messages to promote counter-normative health behaviors: the moderating role of current and future injunctive norms, attitude and self-efficacy. Health Commun 2023;38(6):1071-1079. [doi: 10.1080/10410236.2021.1991638] [Medline: 34689673]
75. Reynolds-Tylus T, Martinez Gonzalez A, Skurka C. Leveraging dynamic norms to reduce college student alcohol use: a test of four mediators. Health Commun 2024;39(7):1371-1382. [doi: 10.1080/10410236.2023.2212447] [Medline: 37204004]
76. Blanton H, Stuart AE, Van den Eijnden RJ. An introduction to deviance-regulation theory: the effect of behavioral norms on message framing. Pers Soc Psychol Bull 2001;27(7):848-858. [doi: 10.1177/0146167201277007]
77. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall; 1994.

XSL•FO
RenderX

Abbreviations

AGReMA: A Guideline for Reporting Mediation Analyses
BI: brief intervention
CONSORT: Consolidated Standards of Reporting Trials
DV: drinking volume
HDD: heavy drinking day
NICE: National Institute for Health and Care Excellence
PFR: personalized feedback on risks for health
PNF: personalized normative feedback
RCT: randomized controlled trial
SAFER: Strengthen restrictions on alcohol availability; Advance and enforce drink driving countermeasures; Facilitate access to screening, brief interventions, and treatment; Enforce bans or comprehensive restrictions on alcohol advertising, sponsorship, and promotion; Raise prices on alcohol through excise taxes and pricing policies

Edited by A Schwartz; submitted 15.12.23; peer-reviewed by J LaBrie, B del Valle Vera, D Smith; comments to author 15.05.24; revised version received 08.10.24; accepted 09.12.24; published 06.02.25.

Please cite as:
Studer J, Cunningham JA, Schmutz E, Gaume J, Adam A, Daeppen JB, Bertholet N
Smartphone-Based Intervention Targeting Norms and Risk Perception Among University Students with Unhealthy Alcohol Use: Secondary Mediation Analysis of a Randomized Controlled Trial
J Med Internet Res 2025;27:e55541
URL: https://www.jmir.org/2025/1/e55541
doi:10.2196/55541
PMID:

©Joseph Studer, John A Cunningham, Elodie Schmutz, Jacques Gaume, Angéline Adam, Jean-Bernard Daeppen, Nicolas Bertholet. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Blended Low-Intensity Internet-Delivered Psychological Program in Patients With Multimorbidity in Primary Care: Randomized Controlled Trial

Alicia Monreal-Bartolomé[1,2,3], PhD; Adoración Castro[4,5,6], PhD; M Ángeles Pérez-Ara[4], PhD; Margalida Gili[1,4,5,6], PhD; Fermín Mayoral[7,8], PhD; María Magdalena Hurtado[7,8], PhD; Esperanza Varela Moreno[1,7,8,9], PhD; Cristina Botella[10,11], PhD; Azucena García-Palacios[10,11], PhD; Rosa M Baños[10,12], PhD; Yolanda López-Del-Hoyo[1,2,3], PhD; Javier García-Campayo[1,3], PhD; Jesus Montero-Marin[13,14,15], PhD

[1]Research Network on Chronicity, Primary Care and Health Promotion RD21/0016/0005 (RICAPPS), Carlos III Health Institute, Madrid, Spain
[2]Department of Psychology and Sociology, University of Zaragoza, Zaragoza, Spain
[3]Aragon Institute for Health Research, IIS Aragon, Zaragoza, Spain
[4]Research Institute of Health Sciences (IUNICS), University of the Balearic Islands (UIB), Palma de Mallorca, Spain
[5]Health Research Institute of the Balearic Islands (IdISBa), Son Espases University Hospital, Building S, Palma de Mallorca, Spain
[6]Department of Psychology, University of the Balearic Islands (UIB), Palma de Mallorca, Spain
[7]Mental Health Department, University Regional Hospital of Malaga, Málaga, Spain
[8]Biomedical Research Institute of Málaga, IBIMA, Málaga, Spain
[9]Research and Innovation Unit (RD21/0016/0015), Costa del Sol University Hospital, Marbella, Málaga, Spain
[10]CIBER Physiopathology Obesity and Nutrition (CIBERobn) Carlos III Health Institute, Madrid, Spain
[11]Department of Clinical and Basic Psychology and Biopsychology, Faculty of Health Sciences, University Jaume I, Castellon, Spain
[12]Department of Psychological, Personality, Evaluation and Treatment, University of Valencia, Valencia, Spain
[13]Teaching, Research & Innovation Unit, Sant Joan de Déu Health Park, Sant Boi de Llobregat, Spain
[14]Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, United Kingdom
[15]Consortium for Biomedical Research in Epidemiology & Public Health (CIBERESP), Madrid, Spain

Corresponding Author:
Javier García-Campayo, PhD
Aragon Institute for Health Research, IIS Aragon
C. de Pedro Cerbuna, 12
Zaragoza, 50009
Spain
Phone: 34 686 673 302
Email: jgarcamp@gmail.com

Abstract

Background: Multimorbidity is a highly prevalent phenomenon whose presence causes a profound physical, psychological, and economic impact. It hinders help seeking, diagnosis, quality of care, and adherence to treatment, and it poses a significant dilemma for present-day health care systems.

Objective: This study aimed to assess the effectiveness of improved treatment as usual (iTAU) combined with a blended low-intensity psychological intervention delivered using information and communication technologies for the treatment of multimorbidity (depression and type 2 diabetes or low back pain) in primary care settings.

Methods: A 2-armed, parallel-group, superiority randomized controlled trial was designed for this study. Participants diagnosed with depression and either type 2 diabetes or low back pain (n=183) were randomized to “intervention + iTAU” (combining a face-to-face intervention with a supporting web-based program) or “iTAU” alone. The main outcome consisted of a standardized composite score to consider (1) severity of depressive symptoms and (2a) diabetes control or (2b) pain intensity and physical disability 3 months after the end of treatment as the primary end point. Differences between the groups were estimated using mixed effects linear regression models, and mediation evaluations were conducted using path analyses to evaluate the potential mechanistic role of positive and negative affectivity and openness to the future.

XSL•FO
RenderX

Results: At 3-month follow-up, the intervention + iTAU group (vs iTAU) exhibited greater reductions in composite multimorbidity score (B=–0.34, 95% CI –0.64 to –0.04; Hedges *g*=0.39) as well as in depression and negative affect and improvements in perceived health, positive affect, and openness to the future. Similar positive effects were observed after the intervention, including improvements in physical disability. No significant differences were found in glycosylated hemoglobin, pain intensity, or disability at 3-month follow-up (*P*=.60; *P*=.79; and *P*=.43, respectively). Path analyses revealed that the intervention had a significant impact on the primary outcome, mediated by both positive and negative affect (positive affect: indirect effect=–0.15, bootstrapped 95% CI –0.28 to –0.03; negative affect: indirect effect=–0.14, bootstrapped 95% CI –0.28 to –0.02).

Conclusions: This study supports the efficacy of a low-intensity psychological intervention applied in a blended format on multimorbidity in primary care. It justifies the exploration of the conceptualization of depression in type 2 diabetes as well as the analysis of the implementation of such interventions in routine clinical practice.

Trial Registration: ClinicalTrials.gov NCT03426709; https://clinicaltrials.gov/study/NCT03426709

International Registered Report Identifier (IRRID): RR2-10.1186/S12888-019-2037-3

(J Med Internet Res 2025;27:e56203) doi:10.2196/56203

KEYWORDS

multimorbidity; depression; type 2 diabetes; low back pain; primary care; blended; internet; randomized controlled trial; RCT

Introduction

Background

Multimorbidity (ie, the presence of ≥2 chronic medical conditions) is a highly prevalent phenomenon [1-4] that affects 1 in 3 adults [5] and has been increasing in recent decades [6-9]. Its presence has a significant physical, psychological, and economic impact and hinders help seeking, diagnosis, the quality of care received, and adherence to treatment [10-14]. Multimorbidity studies conducted in Spain confirm that mental illnesses, particularly major depression, negatively impact quality of life and disability [15,16]. Comorbidity between depression and chronic medical conditions is one of the leading global public health priorities [17]. Although hypertension is very prevalent [15], the most disabling chronic medical conditions are osteoarticular diseases (especially chronic pain), diabetes, and cerebral infarction [15,16]. This study focuses on 2 physical conditions that are comorbid with depression and that involve the greatest disability, loss in quality of life, and higher health care costs: diabetes and chronic pain.

Despite the high prevalence of multimorbidity worldwide, with its consequent demand for care and important health and economic consequences [18], much work remains to be done. Multimorbidity interventions pose a challenge for present-day health care systems [19]. As pointed out by a previous meta-analysis, it is difficult to improve outcomes in people with multimorbidity, although interventions oriented toward depression or specific difficulties and risk factors are promising [17]. There is consensus on the need for a comprehensive assessment to identify patients with multimorbidity who are at risk for negative health outcomes and to simultaneously treat mental and physical comorbidities to prevent functional limitations and future deterioration [20-24]. A comprehensive approach is recommended, addressing not only the medical conditions but also the social, cognitive, and functional issues faced by these patients, as well as a stepped and personalized approach, with therapeutic goals being collaboratively negotiated and regularly re-evaluated throughout the process, for example, by applying the Ariadne principles [22,25]. Patient-oriented approaches, interventions to support self-management, and training for health care professionals appear to be the most frequent elements of interventions with the potential to have a positive impact on patients with multimorbidity [26].

Interventions involving the use of information and communication technologies (ICTs) have been suggested as a promising resource for the provision of adequate and timely support for the self-management of multimorbidity [27-29]. Several studies have demonstrated the effectiveness of personalized, ICT-based interventions for treating depression [30]. However, their effectiveness and cost-effectiveness have not been assessed within a multimorbidity framework. We use the term multimorbidity not only to refer to a specific population of patients but also to the way they are approached and treated [31-33]. Studies focusing on comorbidity deal with only one priority condition over another, instead of addressing multimorbidity [34,35]. As a result, unlike studies using a multimorbidity approach, such as this study, they neglect the bidirectional relationship between the different conditions present and their role or influence on the course of the total index disease. Such approaches, such as using comorbidity, contradict the general recommendation of managing all of a patient's conditions simultaneously to prevent functional limitations and subsequent decline. The comorbidity concept is useful in secondary and tertiary care settings, while multimorbidity is more useful in primary care (PC) [33].

Objectives

In this context, this study aimed to evaluate the effectiveness of a blended low-intensity psychological intervention delivered via ICTs for the treatment of multimorbidity (including depression plus either type 2 diabetes or low back pain) in PC settings. We hypothesized that the improved treatment as usual (iTAU) intervention, enhanced by the delivery of ICT-based, low-intensity psychological therapy, would be more effective for ameliorating multimorbidity symptoms in PC compared to a group receiving only iTAU at 3 months after completion of the program.

XSL•FO
RenderX

Methods

Study Design

This was a parallel-group, superiority randomized controlled trial (RCT) in which patients receiving treatment as usual by their general practitioners (GPs) were randomized to receive either (1) iTAU or (2) the same iTAU combined with a blended low-intensity internet-delivered psychological intervention, which comprised 2 individual face-to-face sessions and 6 individual web-based therapeutic modules. Upon implementation of the protocol and commencement of the RCT, and ultimately due to the outbreak of the COVID-19 pandemic, the RCT management group was compelled to make several changes to the original study protocol [36]. Three major changes were made to the original protocol [36], all of which were discussed and agreed upon by the trial management group before their implementation and were approved by the Research Ethics Committee of the Autonomous Community of Aragon (CEICA; PI16/0259). Changes were made to (1) the number of participants recruited, (2) the time point measurements, and (3) the mechanistic outcome measures. These changes are addressed in the corresponding sections.

Participants and Procedure

Participants were recruited at PC health centers of the 3 Spanish autonomous communities participating in this study (Andalusia, Aragon, and the Balearic Islands). The inclusion criteria were as follows: (1) minimum age of 18 years; (2) *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* diagnosis of major depression or persistent depressive disorder, of mild or moderate severity, expressed as a Patient Health Questionnaire-9 (PHQ-9) score <14; (3) duration of depressive symptoms of ≥2 months; (4) diagnosis of (a) type 2 diabetes (diagnosis according to criteria of the American Diabetes Association [ADA] [37]) or (b) low back pain (diagnosis of nonspecific chronic low back pain according to the definition established by the European Cooperation in Science and Technology B-13 clinical practice guidelines [38] with a duration of at least 6 months); (5) possession of and ability to use a computer, an internet connection, and a mobile phone; (6) ability to understand oral and written Spanish; and (7) willingness to participate in the study and signing the corresponding informed consent form. The following exclusion criteria were applied: (1) any diagnosis of a disease that might affect the central nervous system (eg, brain condition, traumatic brain injury, or dementia); (2) other psychiatric diagnoses or acute mental illness (eg, substance dependence or abuse, history of schizophrenia or other psychotic disorders, or eating disorders), except for anxiety disorder or personality disorders; (3) any medical, infectious, or degenerative disease that might affect mood; (4) presence of delusional ideas or hallucinations, whether consistent with mood; and (5) suicide risk.

The protocol [36] initially stated that this study would require 63 participants in each arm to detect an expected medium effect size on multimorbidity symptoms [39,40]. Nevertheless, to provide enough statistical power to compare the participants receiving psychological therapy versus those only receiving iTAU, considering the comorbid disease that manifested in conjunction with depression (ie, type 2 diabetes or low back pain), 4 possible subgroups were secondarily established, resulting in a total of 252 participants. In addition, an experimental mortality of approximately 15% was expected [41], which meant that the required sample size was initially estimated at approximately 300 participants. However, challenges arose during the trial setup process in recruiting patients with type 2 diabetes. Specifically, most patients with type 2 diabetes were older adults who lacked access to the internet or email and did not possess the basic computer skills needed to participate in the study. This considerably slowed down the recruitment phase of the trial. The focus on type 2 diabetes was originally decided because it is the most common type of diabetes (approximately 95%) and has a more prevalent association with major depressive disorder, and it has been more widely studied than other types of diabetes [42,43]. However, given that this type of diabetes normally develops at a more advanced age and its risk increases with age, patients with type 2 diabetes were much less inclined to use ICTs and thus showed greater reluctance to participate in this study. After serious consideration, it was decided that the secondary comparison by comorbid disease subgroup would be omitted. A new sample size calculation was then performed that considered each trial arm as a whole entity.

For this purpose, we retained the possibility of detecting a similar intermediate effect size on multimorbidity symptoms to test whether the trend in changes differed between the intervention and control groups. This criterion was considered clinically important in previous research [39,40]. We maintained a statistical power (1–β) of 0.80, a significance level of 5%, and a 1:1 allocation ratio. However, due to the previously mentioned recruitment challenges, we assumed a 1:2 ratio between participants with depression and type 2 diabetes compared to those with depression and low back pain in each group. On the basis of the “time×group” interaction in a general linear repeated measures (RMs) design with Greenhouse-Geisser correction [44], and considering a correlation that decreases monotonically with the time gap between RMs, a base correlation of 0.5, and a decay rate of 0.3, along with the expected 15% mortality, we estimated that a total of 180 participants would be needed. This includes 90 participants in each group, with approximately 30 patients with comorbid depression and type 2 diabetes and 60 patients with comorbid depression and low back pain.

Patients were recruited by GPs working in PC centers of the previously mentioned Spanish autonomous communities, who subsequently sent the referral and consent forms of potential participants to the evaluating investigator. The evaluating investigator then contacted the participants to schedule the screening assessments and recorded the psychological and biological variables to determine their inclusion. Randomization was performed in blocks of patients based on the PC center and comorbid disease (ie, type 2 diabetes or low back pain). An independent researcher unrelated to the research team generated the individual randomization list using a randomization software. A researcher from the research team not involved in any other project-related task, together with an independent GP, performed data monitoring tasks. For further information regarding masking and procedures, refer to the protocol by

XSL•FO
RenderX

Monreal-Bartolomé et al [36]. Participants gave their consent for inclusion before learning which treatment they were assigned to and were allowed to withdraw from the study at any time.

Interventions

iTAU at PC Level

All the patients included in this study (both those in the control group and those included in the intervention group) were given their usual treatment by their GPs in PC. This treatment is described as improved because the participating GPs received a training program based on the widely used Spanish guidelines for the treatment of depression in PC, which are based on the National Institute of Health and Care Excellence guidelines [45].

Blended Low-Intensity Internet-Delivered Psychological Program

The blended, low-intensity, internet-delivered psychological program was received only by the intervention group. It consisted of 2 face-to-face individual sessions and 6 web-based individual and interactive therapeutic modules. The web-based therapeutic modules were oriented to work on different psychological techniques and therapeutic strategies that have demonstrated their efficacy for treating depression, diabetes, and chronic low back pain, including motivational techniques, psychoeducation on depression and healthy lifestyle, behavioral activation, positive psychology, and mindfulness-based components [46-55]. These modules were supported by multimedia materials (eg, videos and audios) and had an approximate duration of 60 minutes each.

The content of the program is summarized in Tables 1 and 2. The structure of the modules consistently followed the same pattern [36] and concluded with suggested assignments to enable the material covered to be practiced. Before the commencement of each module, the participants were prompted to confirm their completion of the recommended assignments, and they received a response either congratulating them for finishing the tasks or encouraging them to do so. Completing these assignments is considered crucial for consolidating the knowledge acquired in the program and for translating the strategies of the program into skills. To improve engagement, after a period of inactivity on the software (which was scheduled according to the preferences of the patient), users received an email reminder to continue completing the modules. The program was designed to last between 8 and 12 weeks.

XSL•FO
RenderX

Table 1. Overview of the program modules.

Module or session	Interventions	Aims and contents
Face-to-face session 1 + M0[a] (program presentation)	Motivational techniques	To increase participant adherence to face-to-face sessions and assignments. To present the web-based program and train the patients in the procedure and to log in and use it on their home computers.
M1: understanding emotional issues in medical illnesses	Psychoeducation on depression	To develop a new attitude, understanding problems and difficulties as something inherent to daily life, and seeing them as opportunities for learning and growth. The impact of depression on the quality of life and functional capacity of patients as well as on the prognosis of chronic diseases (such as diabetes and chronic pain) was described. In addition, specific techniques and useful and practical tips to reduce stress in daily life were explained.
M2: healthy lifestyle habits and diabetes or chronic pain	Education, information, practical exercises, and specific techniques on how to build and maintain a healthy lifestyle	To work on the healthy lifestyle component, including physical activity, diet, good sleep (relating it to the physical conditions they experience, such as diabetes and chronic pain), and the development of a social support network.
M3: learning how to live	Behavioral activation	To learn to establish and maintain an appropriate level of activity and involvement with life and to schedule activities (especially meaningful ones) and incorporate them into their routine. It is explained that the abandonment of activities that occurs when there is discomfort is not beneficial, but rather worsens the problems. The program also emphasizes the role of activity in mood regulation and physical well-being, highlighting the importance of staying active and engaging in activities that are meaningful, satisfying, and aligned with their values.
M4: life satisfaction	Positive psychology	To recognize the importance of positive emotions and learn strategies that create positive experiences, encouraging participation in enjoyable and meaningful activities, social interaction, improving mood, and supporting effective management of diabetes and chronic pain.
M5: mindfulness and self-compassion	Components of MBCT[b] and some basic elements of self-compassion	To establish a regular formal practice of mindfulness and self-compassion as well as a regular informal practice. This module includes components of MBCT and MBSR[c], which have shown positive and promising results in patients with depression, but also diabetes, chronic pain, and multimorbidity [54,56,57]. The section on distancing thoughts to reduce discomfort is particularly important as it allows patients to differentiate between primary and secondary diseases and the causes of each.
Face-to-face session 2	Review of the modules already completed and practice	This semistructured session includes the following objectives: (1) resolution of doubts; (2) performing some of the most important practices, also considering the preferences of each patient (3) emphasis on the continuous practice of the strategies learned; and (4) farewell and completion of the intervention.
M6: So, what happens next?	Relapse prevention and maintenance	To reinforce the strategies learned during the program, encourage participants to continue practicing these strategies throughout the follow-up period and teach them how to identify and cope with future high-risk situations related to depression, diabetes, or chronic pain.

[a]M1, M2, M3, etc, refers to module numbers.

[b]MBCT: mindfulness-based cognitive therapy.

[c]MBSR: mindfulness-bases stress reduction.

Table 2. Overview of the program transversal tools.

Transversal tools	Aims and contents
Behavioral activation diary	This tool was designed to focus the attention of the patients to the activities they perform daily; what they spend their time doing; and how this influences their mood, medical condition, and coping ability.
Calendar	This tool provides patients with written information regarding their progress throughout the program.
How am I	This is a tool that provides patients with visual feedback regarding their evolution throughout the program in terms of both their activity level and the intensity of pain as well as their positive (excited, energetic, vital, etc) and negative (upset, fearful, stressed, grumpy, tense, etc) emotionality.

XSL•FO
RenderX

Measures

Overview

During this study, a modification was made to the time point measurements. The initial protocol originally stated [36] that 4 time points would be taken (ie, baseline, posttreatment, 3-month follow-up, and 6-month follow-up), with the last measurement being the primary end point. However, the outbreak of the COVID-19 pandemic in 2020 severely affected life in Spain. As a nationwide lockdown was declared in March 2020, this resulted in PC centers being closed to the public. The COVID-19 pandemic impacted Spanish PC centers, health care professionals, patients, and the general population in various ways—emotionally, socially, and ultimately, in terms of the health care services provided and received. With regard to this RCT, the COVID-19 pandemic outbreak coincided with the final phase of the study, immediately before the last follow-up measurement. Consequently, we had completed the initial baseline measurement as well as delivery of the intervention, postintervention measurement, and the 3-month follow-up measurement, with the 6-month follow-up measurement yet to be conducted. Thus, the most significant impact of the COVID-19 pandemic was on data collection for the primary outcome at 6 months. Several possibilities were considered, although they all pointed to the measurements being seriously compromised in terms of data collection and quality (eg, uncertainty about when PC centers might reopen and social distancing issues). As a result, the feasibility, validity, and accuracy of the estimates—along with the ability to generalize the study results beyond the specific circumstances faced by the Spanish population during the COVID-19 pandemic—were called into question. Therefore, it was decided that the primary end point would be changed from 6-month follow-up to 3-month follow-up after treatment, which was a measure that remained unaffected by the complex and unusual circumstances brought about by the COVID-19 pandemic. Ultimately, the difficult situation at the time prevented us from performing the 6-month follow-up measurement.

Sociodemographic Data

The following sociodemographic variables were collected at baseline: self-identified sex, age, marital status, education, employment, and income.

Main Outcome

In line with previous RCTs [58-60], the main outcome consisted of a composite score that considered (1) severity of depressive symptoms and (2a) diabetes control or (2b) pain intensity and physical disability.

The severity of depressive symptoms was measured using the PHQ-9 [61,62]. The PHQ-9 consists of 9 questions that correspond to the criteria for diagnosing major depressive disorder in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Respondents rate the frequency of their experiences over the past 2 weeks, ranging from “not at all” to “nearly every day.” The PHQ-9 covers different aspects of depression, including mood, energy levels, sleep, appetite, and concentration. Scores for each question are added up to provide an overall severity score (range 0-27). These scores are categorized according to the severity of depression: minimal or mild depressive symptoms (0-10), mild (10-14), moderate (15-19), and severe (20-27) [63]. The Spanish version of the PHQ-9 has demonstrated adequate psychometric properties [61,62].

Diabetes control was measured using glycated hemoglobin (HbA_{1c}). HbA_{1c} is a vital tool for effective diabetes management. It represents the proportion of hemoglobin in the blood that has bound to glucose over an extended period, typically 2 to 3 months. Regular measurement of HbA_{1c} allows tracking progress over time and evaluating whether lifestyle changes or therapy are having an effect. The ADA has defined the following 3 cutoff points: ≤5.6% (nondiabetes), between 5.7% and 6.4% (prediabetes), and ≥6.5% (compatible with a diagnosis of diabetes) [64]. Likewise, the ADA maintains a level of HbA_{1c} ≤7% as the goal for the treatment of patients with diabetes.

Pain intensity and physical disability were assessed using the Faces Pain Scale-Revised (FPS-R) [65] and the Roland-Morris Disability Questionnaire (RMDQ) [66], respectively. The FPS-R is a self-report visual measure in which the patient chooses the face that best represents the level of pain being experienced, with the allocation of a score of 0, 2, 4, 6, 8, or 10, where 0=“no pain” and 10=“very much pain.” The RMDQ was designed to reliably determine the level of physical disability (ie, the limitation in the performance of daily activities) resulting from nonspecific low back pain. It ranges between 0 (absence of disability for back pain) and 24 (maximum possible disability). A change in the score on this scale is clinically significant if it is ≥2 [66-68]. The Spanish version of the RMDQ has demonstrated adequate psychometric properties [66].

The composite score was obtained by combining these components after weights were assigned to them. For (2a), weightage of 50% was assigned both to the severity of depressive symptoms and diabetes control. For (2b), first, the same weightage of 50% was assigned to pain intensity and physical disability (to obtain a composite, including both), and, second, to the severity of depressive symptoms and pain (which included the previous pain measure of intensity and physical disability). These equal weights were added together to yield a standardized composite score that reflected the combination of the 3 components mentioned earlier. This composite score provides a comprehensive measure of comorbidity that considers both depressive symptoms and either diabetes control or pain intensity and physical disability in a single metric.

Secondary and Mechanistic Outcomes

Perceived health status was assessed as a secondary outcome using the 12-item Short Form Survey (SF-12) [69]. The SF-12 questionnaire is widely used in both clinical and research settings. A total score is calculated after coding and transforming the items that range from 0 (worst possible health status) to 100 (best possible health status) [70]. The SF-12 has demonstrated appropriate psychometric qualities [71,72] and has been validated in Spanish [73].

As a new addition to the protocol, it was decided that a measure of affectivity (ie, positive and negative affect) would be added as a potential mechanistic variable. The rationale for this was

XSL•FO
RenderX

to be able to evaluate the potential effects of changes in affectivity on promoting better outcomes. For this purpose, the Positive and Negative Affect Schedule (PANAS) was used [74,75]. This self-report measure consists of 2 subscales: positive and negative affect. Each of the subscales includes 10 adjectives, which must be rated on a scale ranging from 1="not at all or very slightly" to 5="very much," depending on the degree to which each adjective describes the state of mind in which participants generally find themselves. The maximum score is 50 for each of the subscales. This scale has shown good psychometric properties [74-76].

Finally, we used the Openness to the Future Scale (OFS) [77] to specifically measure positive affectivity toward the future as a potential mechanism. This variable can be an indicator of psychological adjustment and a protective factor for mental health. The OFS is a self-reported questionnaire composed of 10 items that are scored on a scale from 1 ("strongly disagree") to 5 ("strongly agree"). A total openness to the future score is obtained by summing the scores of all items (after reversing item 6). The OFS has shown good psychometric properties in the Spanish population [77].

Data Analyses

Data were analyzed using Stata V.18.0 statistical software [78]. First, sociodemographic and outcome descriptive data at baseline (ie, preintervention) were analyzed using frequencies (percentages), medians (IQR), and means (SD), according to their level of measurement and statistical distribution. We evaluated the success of randomization by visual inspection.

The primary analysis consisted of a comparison between the intervention + iTAU and iTAU groups at a 3-month follow-up after the intervention ended, considering the main outcome as a continuous variable. The main outcome was a composite score [36] that included the following: (1) depressive symptom severity using the PHQ-9 [61] and (2a) control of diabetes measured by HbA_{1c} or (3b) pain intensity using the FPS-R [65] and physical disability using the RMDQ [66]. The composite score was weighted to give a continuous standardized aggregate score. The primary analysis was performed using an RM design on a modified intention-to-treat (ITT) basis, that is, we analyzed complete cases due to the high proportion of missingness and explored patterns of missing data. We used multilevel linear mixed regression models with the restricted maximum likelihood method for the estimation of parameters, controlling for age and sex as covariates. The "treatment-by-time" interaction was calculated to determine possible differences between the study arms. The slope coefficient (B), representing the between-group adjusted mean difference change (ie, the interaction term), and its 95% CI were calculated (within-group adjusted mean difference changes and 95% CIs are also provided). Hedges *g*, as an effect size measure of between-group differences, was calculated from the raw data, with Hedges g=0.2 (small effect), Hedges g=0.5 (intermediate effect), and $g \geq 0.8$ (large effect) [79]. We used a 2-sided test with a .05 significance level.

The same analytical approach was used to perform secondary analyses for the main outcome at postintervention as well as for the components of the composite score and for the secondary and mechanistic outcomes at postintervention and 3-month follow-up. We corrected for multiple comparisons by adjusting the significance threshold based on the number of comparisons and the rank of the *P* value according to the Benjamini-Hochberg procedure [80].

Additional post hoc sensitivity analyses were also conducted using complier average causal effect (CACE) or instrumental variable (IV) methods [81] to further investigate the impact of compliance with the program on the composite (primary outcome) at postintervention and at 3-month follow-up, while accounting for potential hidden confounding relationships. A participant in the intervention + iTAU arm was considered a complier if they attended the 2 face-to-face sessions and completed the 6 web-based modules. For this purpose, a 2-stage least squares IV approach was used. In the first stage of regression, marital status, employment, and diagnosis were included as predictors of compliance. In the second stage of regression, age, sex, and the composite at preintervention were introduced as predictors of the outcome at postintervention or 3-month follow-up. The allocated group was used as an IV to define compliance. Results are presented in terms of unstandardized regression coefficients, along with their corresponding 95% CI and *P* values.

The role of positive affect, negative affect, and openness to the future as mediators of improvements in the main outcome was explored. For this purpose, (1) the primary outcome at follow-up was considered the dependent variable; (2) pre-post differential scores of positive affect, negative affect, and openness to the future were calculated and included as process variables; and (3) the group condition (with 2 possibilities: intervention + iTAU vs iTAU) was considered the independent variable. Models included the main outcome at preintervention, age, and sex as covariates. The mediating analyses were conducted using path analyses for continuous dependent variables. Standardized regression coefficients for bootstrapped indirect effects were estimated, along with their 95% CIs based on 10,000 bootstrapped samples. A significant mediating effect was considered when the 95% CI did not include zero [82].

Ethical Considerations

This study was approved by the research ethics committee of each of the autonomous communities involved (CEICA Aragon: PI16/0259, CEI Balearic Islands: IB 3402/17 PI, and the Regional Ethics and Research Committee of the province of Malaga: 03/2017 ICPS 2) and was designed in accordance with the ethical standards laid down in the Declaration of Helsinki and its later amendments. Modifications to the published protocol [36] were approved by the research ethics committee of the autonomous community corresponding to the leading group (CEICA Aragon, PI16/0259). Written informed consent was obtained before screening, and exclusion criteria were applied afterward. As this study involved the use of the internet, AES strategies for data encryption and personal password use were implemented to ensure the protection of personal information. The data were treated anonymously and confidentially and were used solely for the purposes of the study. Study participants did not receive any compensation for their participation, other than receiving improved treatment for their condition.

XSL•FO
RenderX

Results

Flow, Baseline Characteristics, and Compliance

As shown in Figure 1, after excluding 112 (37.9%) participants (who did not meet the inclusion criteria) from the initial 295, the remaining 183 (62.1%) individuals were randomly assigned to 1 of the 2 experimental conditions (intervention + iTAU: n=93, 50.8%; iTAU: n=90, 49.2%). The participants were mostly women (132/183, 72.1%) with a mean age of 51.36 (SD 11.3) years, and there were no important differences in either sociodemographic or clinical characteristics between the 2 arms that might suggest prognostic strength (Table 3). In line with the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting parallel-group RCTs and following the recommendation to avoid conducting probability tests on potential baseline differences that may have occurred by chance due to random assignment, we did not perform such hypothesis testing. This type of testing is unnecessary because it assesses the likelihood that observed baseline differences happened by chance, and as random assignment has already been performed, this is already understood. Instead, we based our comparisons of sociodemographic and clinical data at baseline on the size of potential chance imbalances and their prognostic relevance [83,84]. In this regard, there were no significant differences in sociodemographic or clinical characteristics between the two groups that could suggest prognostic strength (Table 2). Therefore, no additional covariates were included in the subsequent analytical models beyond those prespecified in the protocol, namely, age and sex.

Figure 1. Participant flow diagram. Numbers at posttreatment and follow-up evaluations reflect those cases in which at least 1 of the variables that make up the composite multimorbidity main outcome measure were obtained. iTAU: improved treatment as usual.

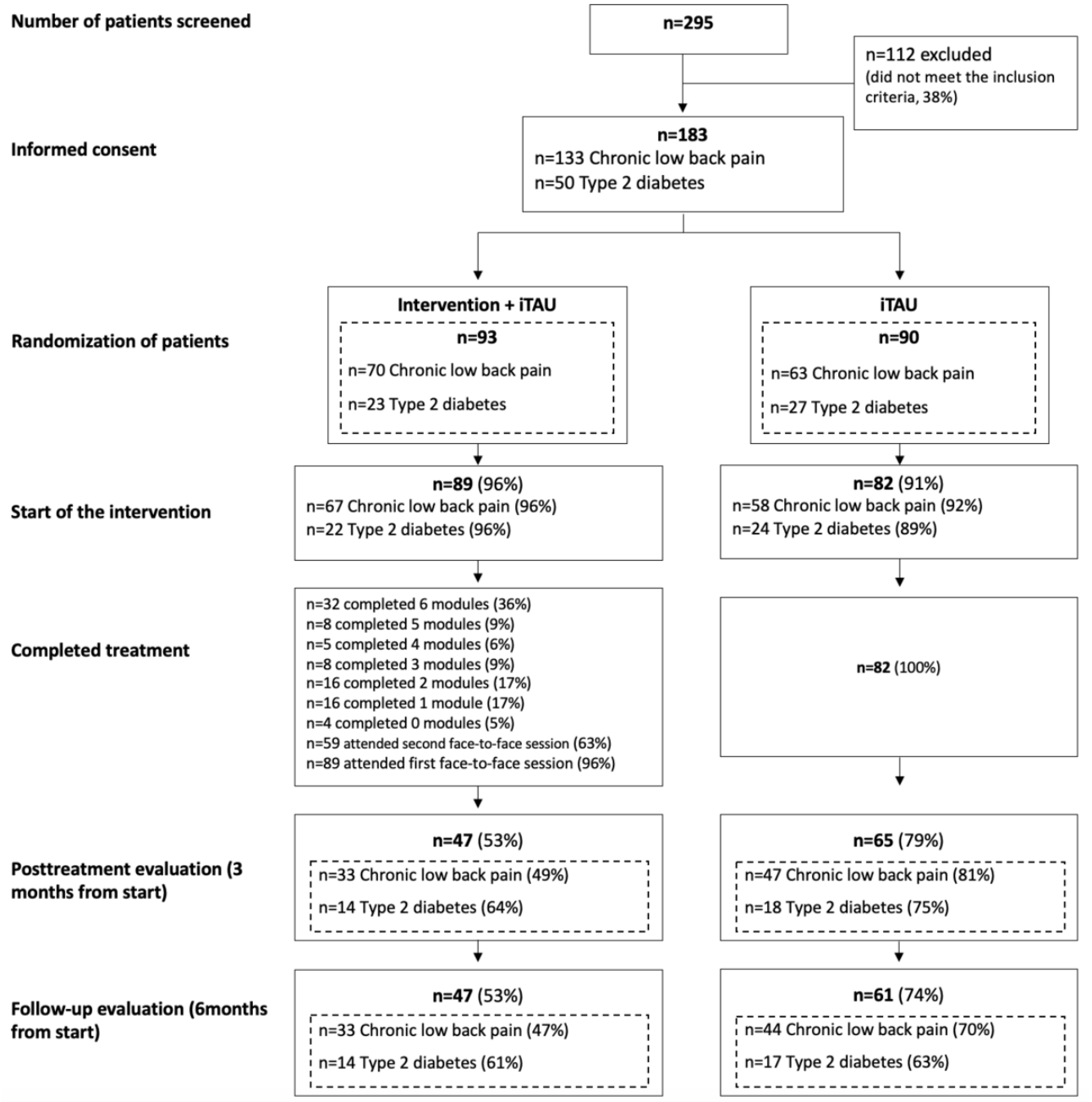

XSL•FO
RenderX

Table 3. Baseline characteristics of patients by treatment group (N=183).

Characteristics	Intervention + iTAU[a] (n=93)	iTAU alone (n=90)
Sociodemographic data		
Age (y), mean (SD)	51.20 (11.16)	51.51 (11.51)
Sex (female), n (%)	74 (79)	58 (65)
Marital status[b], n (%)		
Married or relationship	56 (61)	54 (60)
Single	17 (19)	16 (18)
Separated or divorced	17 (19)	12 (13)
Widowed	2 (2)	8 (9)
Place of residence, n (%)		
Own home	79 (86)	75 (83)
Relative's home	4 (4)	5 (6)
Neighbor's or friend's home	0 (0.0)	1 (1)
Other	9 (10)	9 (10)
Education, n (%)		
No studies	0 (0)	6 (7)
Primary studies	22 (24)	17 (19)
Secondary studies	55 (60)	39 (43)
Tertiary studies	15 (16)	28 (31)
Employment, n (%)		
Unemployed	24 (26)	28 (31)
Employed	24 (26)	38 (42)
Home duties	2 (2)	5 (6)
Student	1 (1)	0 (0)
Sick leave	22 (24)	8 (9)
Retired	19 (21)	11 (12)
Income[c], n (%)		
≤National minimum wage (US $600)	25 (27)	20 (22)
1-2 × national minimum wage (US $600)	30 (33)	31 (34)
2-4 × national minimum wage (US $600)	22 (24)	26 (29)
>4 × national minimum wage (US $600)	1 (1)	4 (4)
Clinical data		
Diagnosis, n (%)		
Pain	70 (75)	63 (70)
Diabetes	23 (25)	27 (30)
Medications, median (IQR)	3 (2-4)	3 (2-4)
Analgesics, n (%)	57 (61)	50 (56)
Antidepressants, n (%)	39 (42)	37 (41)
Antiepileptics, n (%)	16 (17)	7 (8)
Antipsychotics, n (%)	1 (1)	0 (0)
Antidiabetics, n (%)	21 (23)	23 (26)
Antithyroid, n (%)	5 (5)	4 (4)

XSL•FO
RenderX

Characteristics	Intervention + iTAU[a] (n=93)	iTAU alone (n=90)
Anxiolytics, n (%)	45 (48)	33 (37)
Cardiovascular medication, n (%)	32 (34)	34 (38)
Corticosteroids, n (%)	4 (4)	4 (4)
Insulin, n (%)	5 (5)	7 (8)
NSAIDs[d], n (%)	1 (1)	0 (0)
Other, n (%)	30 (32)	36 (40)
Composite (range –2.25 to 2.25), mean (SD)	–0.02 (0.74)	0.05 (0.81)
PHQ-9[e] (range 0-27), mean (SD)	14.48 (5.34)	13.84 (5.79)
RMDQ[f] (range 0-24), mean (SD)	13.54 (5.54)	14.38 (6.13)
FPS-R[g] (range 0-10), mean (SD)	4.86 (2.15)	5.11 (2.47)
HbA_{1c}[h] (diabetes: ≥6.5), mean (SD)	6.78 (0.66)	7.77 (2.48)
SF-12[i] (range 0-100), mean (SD)	30.33 (19.28)	31.84 (18.36)
PANAS[j]-positive subscale (range 10-50), mean (SD)	20.32 (6.48)	20.97 (7.34)
PANAS-negative subscale (range 10-50), mean (SD)	26.35 (7.72)	26.06 (8.50)
OFS[k] (range 10-50), mean (SD)	30.82 (8.12)	29.97 (8.06)

[a]iTAU: improved treatment as usual.

[b]Marital status, place of residence, education, and employment: 1 missing in the intervention arm.

[c]Income: 15 missing in the intervention arm, and 9 missing in the control arm.

[d]NSAID: nonsteroidal anti-inflammatory drug.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]RMDQ: Rolland-Morris Disability Questionnaire.

[g]FPS-R: Faces Pain Scale-Revised

[h]HbA_{1c}: glycosylated hemoglobin.

[i]SF-12: 12-item Short Form Survey.

[j]PANAS-positive: Positive and Negative Affect Scale.

[k]OFS: Openness to the Future Scale.

On average, the PHQ-9 indicated moderate levels of depression (mean 14.18, SD 5.55) in the total sample, the RMDQ suggested a dysfunctional low back pain status (mean 13.93, SD 5.82) in the subgroup with low back pain, and the HbA_{1c} value confirmed the presence of diabetes (mean 7.09, SD 1.51) in the type 2 diabetes subgroup. More than half of the sample (107/183, 58.5%) was taking analgesics, which was the most common medication. In total, 95% (89/93) of the participants in the intervention + iTAU arm attended the first face-to-face session, and 63% (59/93) of the participants attended the second face-to-face session. The median (IQR) number of modules attended in the intervention arm was 4 (out of 7), with a mean of 4.02 (SD 2.75). Postintervention retention rates in the primary outcome (ie, composite score) were 37% (35/93) in the intervention + iTAU arm and 54% (49/90) in the iTAU arm, with rates of 49% (46/93) and 55% (50/90), respectively, at follow-up. At the 3-month follow-up (primary end point), higher age, being separated or divorced, and having a diabetes diagnosis (as well as taking antidiabetics and insulin) were significantly associated with a higher probability of attrition. On the other hand, being employed, engaging in household chores, being on sick leave, and taking analgesics and antiepileptics were significantly associated with a lower probability of study attrition (Multimedia Appendix 1).

Effects on Primary and Secondary Outcomes at Primary End Point

The study outcomes are reported by trial arm status at postintervention (Multimedia Appendix 2) and 3-month follow-up (Table 4). At 3-month follow-up (primary end point), the within-group analyses of the intervention + iTAU group based on a modified ITT basis revealed significant improvements in the composite score (main outcome) as well as in depression (PHQ-9), perceived health status (SF-12), positive affect (PANAS-positive), negative affect (PANAS-negative), and openness to the future (OFS; Table 4). However, there were no significant effects on HbA_{1c} or pain intensity and disability (RMDQ and FPS-R). At 3-month follow-up, the within-group analyses of the iTAU group based on a modified ITT basis showed no significant effects in any of the outcomes (Table 4). At 3-month follow-up, there was evidence that the intervention + iTAU group achieved a significantly greater reduction in the composite score (main outcome) compared to iTAU (B=–0.34, 95% CI –0.64 to –0.04),

XSL•FO
RenderX

with small-to-medium effects (Hedges *g*=–0.39). Furthermore, compared to iTAU, the intervention + iTAU group showed greater reductions in depression (PHQ-9: B=–3.92, 95% CI –5.70 to –2.15) and negative affect (PANAS-negative: B=–3.67, 95% CI –6.63 to –0.71) and greater improvements in perceived health status (SF-12: B=9.04, 95% CI 3.21-14.87), positive affect (PANAS-positive: B=4.73, 95% CI 2.01-7.45), and openness to the future (OFS: (B=4.73, 95% CI 2.01-7.45), with small-to-large effects (Hedges *g* ranging from 0.36 to 1.13 in absolute value). However, no significant differences in HbA_{1c} or pain intensity and disability (RMDQ and FPS-R) were identified in the comparison between the intervention + iTAU group and the iTAU group at 3-month follow-up.

XSL•FO
RenderX

Table 4. Descriptive statistics and main comparisons for primary and secondary outcomes at 3-month follow-up.

	Participants, n (intervention, iTAU[a])	Intervention + iTAU, mean (SD)	iTAU alone, mean (SD)	Within group, B (95% CI)		Between group		
				Intervention + iTAU	iTAU alone	*d*	*P* value[b]	B (95% CI)
Primary outcome								
Composite (range –2.25 to 2.25)	46, 50							
Baseline		–0.10 (0.79)	–0.01 (0.80)	—[c]	—	—	—	—
Follow-up		–0.21 (0.86)	0.19 (0.86)	–0.13 (–0.37 to – 0.02)	0.20 (–0.02 to 0.43)	–0.39	.02	–0.34 (–0.64 to –0.04)
Secondary outcomes								
PHQ-9[d] (range 0-27)	47, 61							
Baseline		13.89 (5.97)	13.41 (5.45)	—	—	—	—	—
Follow-up		9.79 (6.46)	13.05 (6.45)	–4.30 (–5.56 to –3.04)	–0.50 (–1.73 to 0.72)	–0.65	<.001	–3.92 (–5.70 to –2.15)
RMDQ[e] (range 0-24)	33, 44							
Baseline		13.36 (5.49)	14.00 (6.05)	—	—	—	—	—
Follow-up		13.06 (6.28)	14.39 (6.18)	–0.36 (–1.90 to 1.18)	0.04 (–0.87 to 0.95)	–0.12	.43	–0.74 (–2.58 to1.10)
FPS-R[f] (range 0-10)	33, 44							
Baseline		4.67 (1.78)	5.00 (2.18)	—	—	—	—	—
Follow-up		5.21 (2.50)	5.73 (2.05)	0.43 (–0.37 to 1.23)	0.65 (–0.13 to 1.43)	–0.10	.79	–0.14 (–1.13 to 0.86)
HbA_{1c}[g] (diabetes: ≥6.5)	16, 5							
Baseline		6.73 (0.66)	7.54 (1.40)	—	—	—	—	—
Follow-up		6.89 (1.31)	7.54 (2.08)	0.23 (–0.26 to 0.72)	–0.01 (–0.78 to 0.77)	0.18	.60	0.23 (–0.61 to 1.07)
SF-12[h] (range 0-100)	47, 60							
Baseline		33.91 (19.43)	33.17 (17.81)	—	—	—	—	—
Follow-up		43.46 (24.01)	33.04 (21.95)	10.81 (5.98 to15.64)	1.19 (–2.25 to 4.63)	0.52	.002	9.04 (3.21 to 14.87)
PANAS[i]-positive sub-scale (range 10-50)	44, 59							
Baseline		20.68 (7.14)	21.47 (7.05)	—	—	—	—	—
Follow-up		25.36 (9.77)	21.24 (8.88)	4.84 (2.04 to 7.64)	–0.03 (–1.86 to 1.80)	1.13	.001	4.73 (2.01 to 7.45)
PANAS-negative sub-scale (range 10-50)	44, 59							
Baseline		25.89 (8.50)	27.17 (8.46)	—	—	—	—	—
Follow-up		21.09 (8.34)	25.47 (9.31)	–5.05 (–7.14 to – 2.95)	–1.32 (–3.14 to 0.50)	–0.36	.02	–3.67 (–6.63 to –0.71)
OFS[j] (range 10-50)	47, 60							
Baseline		31.02 (7.86)	30.43 (8.18)	—	—	—	—	—
Follow-up		34.06 (9.01)	30.32 (8.13)	3.13 (0.98 to 5.29)	0.03 (–1.60 to 1.68)	0.39	.02	3.08 (0.55 to 5.62)

XSL•FO
RenderX

[a]iTAU: improved treatment as usual.

[b]All significant results remained significant after correction for multiple comparisons using the Benjamini-Hochberg method.

[c]Not applicable.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]RMDQ: Rolland-Morris Disability Questionnaire.

[f]FPS-R: Faces Pain Scale-Revised.

[g]HbA_{1c}: glycosylated hemoglobin.

[h]SF-12: 12-item Short Form Survey.

[i]PANAS: Positive and Negative Affect Scale.

[j]OFS: Openness to the Future Scale.

Effects on Primary and Secondary Outcomes at Secondary End Point

At postintervention, the within-group analyses of the intervention + iTAU group based on a modified ITT basis revealed significant improvements in the composite score (main outcome) as well as in depression (PHQ-9), pain intensity (RMDQ), HbA_{1c}, perceived health status (SF-12), positive affect (PANAS-positive), negative affect (PANAS-negative), and openness to the future (OFS); however, there were no significant effects on pain disability (FPS-R; Multimedia Appendix 2). At postintervention, the within-group analyses of the iTAU group based on a modified ITT basis showed a significant worsening in the composite score (main outcome); however, there were no other significant effects (Multimedia Appendix 2). At postintervention, there was evidence that the intervention + iTAU group achieved a significantly greater reduction in the composite score (main outcome) compared to iTAU (B=–0.63, 95% CI –0.94 to –0.31), with large effects (Hedges *g*=–0.85). Furthermore, the intervention + iTAU group also showed greater reductions in depression (PHQ-9: B=–5.02, 95% CI –6.77 to –3.26), low back pain disability (RMDQ: B=–2.28, 95% CI –4.10 to –0.46), and negative affect (PANAS-negative: B=–5.52, 95% CI –8.45 to –2.60) and greater improvements in perceived health status (SF-12: B=13.55, 95% CI 7.77-19.33), positive affect (PANAS-positive: B=5.25, 95% CI 2.56-7.93), and openness to the future (OFS: B=3.16, 95% CI 0.64-5.67), with small-to-large effects (Hedges *g* ranging from 0.38 to 0.88 in absolute value). However, no significant differences in HbA_{1c} and pain intensity (FPS-R) were identified in the comparison between the intervention + iTAU group and the iTAU group at postintervention.

Effect of Compliance With the Program

The results of the CACE or IV analyses represent an estimation of the intervention effect among the subpopulation of compliers in the intervention arm, compared to those in the control arm who would have complied with the intervention had they been offered it. The CACE or IV analyses indicated evidence of a relationship between the completion of the program and the composite score (main outcome) at postintervention (B=–1.04, 95% CI –1.56 to –0.52; *P*<.001). These analyses also showed evidence of a relationship between the completion of the program and the composite (main outcome) at 3-month follow-up (B=–0.54, 95% CI –1.01 to –0.07; *P*=.03).

Analysis of Mediating Variables

The results of the path analyses on the primary outcome (composite score) are detailed in Table 5, where the independent variable is the group condition (intervention + iTAU vs iTAU). Three models controlled for the main outcome at baseline, age, and gender were tested (ie, positive affect, negative affect, and openness to the future as potential mediators of the effect of the intervention on the outcome); however, only positive affect and negative affect showed a significant indirect effect (positive affect: ab=–0.15, bootstrapped 95% CI –0.28 to –0.03 and negative affect: ab=–0.14, bootstrapped 95% CI –0.28 to –0.02). After controlling for the mediators, the direct effect (path c) of the intervention was significant and of the same sign as the indirect effects in both models, suggesting a “complementary mediation.”

XSL•FO
RenderX

Table 5. Direct and bootstrapped indirect effects in the mediational models.

Main outcome[a] and mediators[b] (R^2)[c], and direct effects				Indirect effects		
	Path[d]	Coefficient[e]	*P* value[f]	Path	Coefficient[g]	95% CI[h]
Composite (0.57)						
Positive affect **(0.16)**				a × b	–0.15	–0.28 to –0.03
	a[i]	0.76	.001			
	b[j]	–0.19	.03			
	c′[k]	–0.57	.001			
	c[l]	–0.42	.01			
Composite **(0.55)**						
Negative affect **(0.15)**				a × b	–0.14	–0.28 to –0.02
	a	–0.60	.007			
	b	0.22	.007			
	c′	–0.56	.001			
	c	–0.42	.009			
Composite **(0.58)**						
Openness to the future **(0.12)**				a × b	–0.06	–0.18 to 0.01
	a	–0.15	.07			
	b	0.37	.10			
	c′	–0.54	.001			
	c	–0.48	.003			

[a]The dependent variable (main outcome) is the composite score at 3-month follow-up.

[b]The potential mediators, highlighted in italics (positive affect, negative affect, and openness to the future), were based on pre-post change scores.

[c]R^2: variance explained by regression models.

[d]Path coefficients are (standardized) ordinary least squares–based regression coefficients.

[e]Coefficient: (standardized) slope.

[f]*P* value related to *t* test.

[g]The product of "ab" is the bootstrapped indirect effect using 10,000 samples.

[h]It is the 95% CI of the bootstrapped indirect effect using 10,000 samples.

[i]a: the direct path between the independent variable and the mediator.

[j]b: the direct path between the mediator and the outcome.

[k]c′: total effects.

[l]c: direct effect of the independent variable on the dependent variable after adjustment for mediating effects.

Discussion

Principal Findings

The main aim of this trial was to evaluate the effectiveness of a low-intensity psychotherapy program (intervention + iTAU) applied in a hybrid form with face-to-face and internet-based sessions (ie, blended), for the treatment of multimorbidity between mild to moderate severity depression and either type 2 diabetes or chronic low back pain in PC, compared to a group that only received iTAU. It was observed that the intervention + iTAU group achieved a significantly greater reduction in the composite score (main outcome) compared to iTAU, with large effects at postintervention and small-to-medium effects at follow-up. In addition, compared to iTAU, the intervention + iTAU group showed greater reductions in depressive symptomatology, low back pain disability, and negative affect at postintervention, although not at follow-up, where only reductions in depression and negative affect were maintained. On the other hand, greater improvements in perceived health, positive affect, and openness to the future were observed in the intervention + iTAU group versus iTAU, with small-to-large effects at both time points. However, no significant differences in HbA_{1c} or pain intensity were identified in the comparison between the intervention + iTAU group and the iTAU, either at postintervention or 3-month follow-up.

Comparison to Prior Work

These results align with previous literature, indicating the effectiveness of psychological intervention programs for multimorbidity in improving depressive symptomatology, whether delivered face-to-face or through ICTs

XSL•FO
RenderX

[20,35,49,51,52,85-95]. With regard to pain, the few studies that tested psychotherapy tools delivered via ICTs in these patients showed varying results [96]. While some studies achieved improvements in either disability and pain intensity, or both [92,97-99], others achieved no or only partial improvement in these variables [93,100-103]. Differences in the results may be attributed to the significant heterogeneity among these studies, such as variations in the type of intervention, the presence of pain comorbidities, the measurement instruments used, and the existence of comorbid depression. Among the studies that specifically had chronic low back pain and depression as inclusion criteria and were conducted entirely through ICTs, improvements in pain intensity or associated disability were not observed 6 months after randomization [99,102]. However, one of them did show improvements postintervention [99]. It is possible that pain intensity, being a variable that is difficult to modify, exhibits effects that are lost earlier, and this was evident in both cases, although the loss occurred in the medium term. As far as we know, there are no studies applying blended psychotherapy in patients with chronic pain and depression. Therefore, reaching a solid conclusion in this regard is challenging, highlighting the need for further studies in this field.

Our results present a contrast with findings from other studies that used psychological interventions in patients with diabetes and depression [57,104-109]. In those studies, both cognitive behavioral therapy and interventions incorporating mindfulness and self-compassion components appeared effective in glycemic control. However, it is worth noting that these interventions were primarily delivered in a face-to-face design. In studies where interventions were applied via the internet or telephone, improvements in glycemic control were not achieved [35,90,91,94,110-113]. Thus, the method of implementing the intervention could play a decisive role. Given that our intervention was blended, such effects might have been mitigated. This could also be associated with the small sample size (resulting in reduced statistical power) achieved when recruiting patients with diabetes, along with the limited follow-up time (3 months). The HbA_{1c} variable represents the proportion of hemoglobin in the blood that has bound to glucose over an extended period. Therefore, a more extended follow-up time may be necessary to observe significant results, as demonstrated by Hoyo et al [114], where HbA_{1c} levels continued to decrease at 12- and 18-month follow-ups.

With regard to the observed improvements in perceived health, the use of a blended model rather than a solely web-based approach may have led these improvements to stand out, particularly in comparison to other nonblended studies [90,110]. In those studies, only some of the dimensions of perceived health showed improvement, or improvements were not achieved at the intragroup level. A study that applied a face-to-face mindfulness-based cognitive therapy (MBCT) intervention in patients with diabetes and low levels of emotional well-being found that MBCT was more effective for improving perceived health than the waitlist control group [115]. Conversely, in the case of pain, a previous face-to-face delivered study that compared cognitive behavioral therapy, mindfulness-based stress reduction, and a waiting list showed improvements in perceived mental health only [116].

In the case of improvements in negative and positive affect, a situation similar to that observed for the previous variables arose, with varied results across different studies [93,117,118] Those studies exhibit considerable heterogeneity, differing in crucial characteristics such as the type of intervention, delivery method (eg, face-to-face, telephone, and web-based vs mixed), and the nature of the medical conditions. This heterogeneity makes it challenging to compare them with each other and with this study. Our results underscore the crucial role of affect, both positive and negative, as potential mediators in the functioning of the intervention, influencing improvements in the composite main outcome. Understanding how affective states mediate the impact of the intervention is pivotal for tailoring and optimizing future treatment developments. This finding emphasizes the need for interventions that consider the affective dimension, not only addressing the negative valence of depressive symptoms but also enhancing positive emotional experiences. While this study contributes to this understanding, it is noteworthy that similar results were found when using MBCT to reduce the risk of relapse or recurrence in major depressive disorder [119].

Finally, improvements in openness to the future could not be compared with any previous studies due to the novelty of the construct [77]. It is interesting to understand the positive future expectations of patients both before and after receiving an intervention for any health problem. Specifically, the measure used in this study incorporates different aspects, such as the positive illusion of control, the active process of accepting future scenarios, and the confidence and commitment in one's ability to plan for desired outcomes and to cope with adversity. Nevertheless, this variable did not show any potential mediating effects on the main composite score.

Strengths and Limitations

This is the first study to use blended models involving ICTs in the treatment of multimorbidity. It focused on a particularly prevalent, disabling, and challenging condition in clinical practice, multimorbidity between depression and type 2 diabetes or chronic low back pain. Furthermore, an evidence-based design and intervention were proposed, adhering to the recommendations of major clinical guidelines and previous research. The intervention specifically targeted risk factors, such as depression, and addressed functional difficulties. It was centered on patients and their specific needs, offering support for self-management, adopting a comprehensive and personalized approach, and incorporating therapeutic objectives negotiated and reassessed throughout the process in accordance with the principles by Ariadne [22].

However, the study has significant limitations. First, a notable challenge was the high attrition rate, which impacted the statistical power of this study. Recruitment and retention difficulties in RCTs targeting comorbid physical and mental illnesses have been documented in prior studies [120,121], and specific details of this study will be discussed in a forthcoming publication focused on implementation. Nevertheless, it is important to highlight that, in line with the earlier recruitment

XSL•FO
RenderX

challenges, dropouts were more prevalent among older patients and those with type 2 diabetes. This trend might be attributed to challenges in managing the use of internet and email and the acquisition of the basic computer skills required for active participation. Similar observations were made in a study by Clarke et al [35] focusing on patients with type 2 diabetes and depression. This study reported higher attrition and mean age compared to other ICT-delivered interventions in chronic low back pain and depression [99,102] and diabetes and depression [90,91]. It is noteworthy that the latter studies also included type 1 diabetes, contributing to differences in the age distribution of participants. There is also a proposal that patients with diabetes and mild depression might perceive low mood as a feature of their diabetes rather than a separate condition to be treated, potentially influencing treatment adherence and completion [122]. As suggested by other authors [35], further exploration of the conceptualization of depression in type 2 diabetes and its impact on program uptake and the benefit of treatment is recommended. Regardless, the low recruitment and retention rate in this study, particularly for older patients with diabetes owing to their lack of basic computer skills, could affect the generalizability of the results. Second, similar to many clinical trials, this study faced disruptions due to the COVID-19 pandemic [123], preventing the implementation of necessary follow-up measures, particularly for patients with diabetes. Given their heightened vulnerability to the virus, this situation led to the loss of valuable follow-up data. Trials that stop earlier for reasons independent of trial findings are unlikely to introduce bias because of their premature termination [124]; however, changing the primary end point from a 6-month follow-up to a 3-month follow-up in response to the challenges posed by the COVID-19 pandemic offers a different time perspective than originally proposed. We lack information on the effect sizes at 6 months, and the true effect size may vary between the 3-month and 6-month time points. In this sense, future research might benefit from the inclusion of longer follow-up periods. Third, we observed a varying proportion of losses between the intervention and control groups postintervention, although this effect disappeared at the subsequent primary end point. Similar effects were noted in prior studies [35,90,91,102,111] and could be attributed to some extent to the control groups being wait-listed, leading participants to initially complete the assessment and refrain from dropping out in anticipation of receiving the intervention at a later point. Although modified ITT and CACE or IV approaches provided useful insights into the impact of treatment on those who completed the study and on those who completed the treatment, we must not overlook this potential attrition bias and the possible limitation when generalizing the results to the entire population, where adherence can be variable. Nevertheless, the fact that the effect of the losses disappears at the primary end point indicates that the initial bias was not significant in the long term. A more comprehensive consideration of these implementation aspects will be presented in a subsequent paper. Finally, we did not measure other diseases besides depression, diabetes, and chronic pain; therefore, we are unaware of how the program might work with other potentially present conditions. Future studies should take into account a wider range of potential disorders among the participating patients.

Conclusions

This study supports the efficacy of a low-intensity psychological intervention applied in a blended form for multimorbidity in PC. Promising outcomes are particularly noted in the psychological dimension, showing improvements in depression, perceived health, positive and negative affect, and openness to the future. However, the findings in the physical dimension are mixed, indicating variable results in pain and disability reduction, promising results in short-term disability reduction, and no improvements in diabetes control. We encourage further research to validate the findings of this study, including the exploration of subgroups that could not be adequately examined due to the discussed limitations. In addition, we advocate an exploration of the conceptualization of depression in type 2 diabetes to shed light on its effects on adherence indicators, acceptance, and the efficacy of psychological interventions in these patients. Analyzing the implementation of such interventions in routine clinical practice is also warranted.

Acknowledgments

JM-M has a ‘Miguel Servet’ research contract from the ISCIII (CP21/00080), he was funded by mobility grants during the project duration (MV22/00022, CAS22/00436) and he is grateful to the Department of Psychiatry at the University of Oxford, UK, and to the Spanish CIBER of Epidemiology and Public Health (CIBERESP CB22/02/00052; ISCIII) for their support. AM-B has a Margarita Salas postdoctoral fellowship at the University of Zaragoza, which is funded by the NextGeneration European Union. This research was supported by Carlos III Health Institute (ISCIII) of the Spanish Ministry of Economy and Competitiveness (PI16/00962; FI17/00180), and with group funds the Research Network on Chronicity, Primary Care and Health Promotion (RICAPPS, RD21/0016/0005) that is part of the Results-Oriented Cooperative Research Networks in Health (RICORS) (Carlos III Health Institute), funded by the European Union “NextGeneration EU/PRTR” funds. The funding body had no involvement in the study design, data collection, analysis, interpretation of the data, or manuscript writing. The funding entity audited the trial execution.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

JG-C conceptualized and designed the study. AM-B and JM-M developed the statistical analysis plan and wrote the first draft. FM, CB and RMB verified the analytical methods. AG-P was responsible for training clinicians on the programme content and web-based application aspects throughout the clinical trial. The web-based therapeutic modules were designed and developed by AG-P. JG-C coordinated the study in Aragon, MG in the Balearic Islands and FM in Andalusia. YL-D-H and AM-B were involved in the execution of the project and were responsible for data collection in Aragon, AC and MAP-A in the Balearic Islands and MMH and EV in Andalusia. JM-M performed all the statistical analyses. AM-B and JM-M wrote the first draft of the manuscript. All authors contributed to the final version of the manuscript. JG-C, MG, FM and AG-P obtained funding and supervised all steps. All authors have read and accepted the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variables at baseline as predictors of missingness.
[DOCX File , 17 KB - jmir_v27i1e56203_app1.docx]

Multimedia Appendix 2
Descriptive statistics and main comparisons at postintervention.
[DOCX File , 19 KB - jmir_v27i1e56203_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 531 KB - jmir_v27i1e56203_app3.pdf]

References

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
2. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. Ann Fam Med 2012;10(2):142-151 [FREE Full text] [doi: 10.1370/afm.1337] [Medline: 22412006]
3. Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L. Prevalence of multimorbidity among adults seen in family practice. Ann Fam Med 2005;3(3):223-228 [FREE Full text] [doi: 10.1370/afm.272] [Medline: 15928225]
4. Walker AE. Multiple chronic diseases and quality of life: patterns emerging from a large national sample, Australia. Chronic Illn 2007 Sep 01;3(3):202-218. [doi: 10.1177/1742395307081504] [Medline: 18083677]
5. González-Colom R, Mitra K, Vela E, Gezsi A, Paajanen T, Gál Z, et al. Multicentric assessment of a multimorbidity-adjusted disability score to stratify depression-related risks using temporal disease maps: instrument validation study. J Med Internet Res 2024 Jun 24;26:e53162 [FREE Full text] [doi: 10.2196/53162] [Medline: 38913991]
6. King DE, Xiang J, Pilkerton CS. Multimorbidity trends in United States adults, 1988-2014. J Am Board Fam Med 2018 Jul 09;31(4):503-513 [FREE Full text] [doi: 10.3122/jabfm.2018.04.180008] [Medline: 29986975]
7. Koné Pefoyo AJ, Bronskill SE, Gruneir A, Calzavara A, Thavorn K, Petrosyan Y, et al. The increasing burden and complexity of multimorbidity. BMC Public Health 2015 Apr 23;15:415 [FREE Full text] [doi: 10.1186/s12889-015-1733-2] [Medline: 25903064]
8. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. Eur J Gen Pract 2008;14 Suppl 1:28-32 [FREE Full text] [doi: 10.1080/13814780802436093] [Medline: 18949641]
9. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. EClinicalMedicine 2023 Mar;57:101860 [FREE Full text] [doi: 10.1016/j.eclinm.2023.101860] [Medline: 36864977]
10. Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health Qual Life Outcomes 2004 Sep 20;2(1):51 [FREE Full text] [doi: 10.1186/1477-7525-2-51] [Medline: 15380021]
11. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec 16;22 Suppl 3(Suppl 3):391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]
12. Fortin M, Bravo G, Hudon C, Lapointe L, Dubois MF, Almirall J. Psychological distress and multimorbidity in primary care. Ann Fam Med 2006 Sep 01;4(5):417-422 [FREE Full text] [doi: 10.1370/afm.528] [Medline: 17003141]

XSL•FO
RenderX

13. Foley L, Larkin J, Lombard-Vance R, Murphy AW, Hynes L, Galvin E, et al. Prevalence and predictors of medication non-adherence among people living with multimorbidity: a systematic review and meta-analysis. BMJ Open 2021 Sep 02;11(9):e044987 [FREE Full text] [doi: 10.1136/bmjopen-2020-044987] [Medline: 34475141]
14. Larkin J, Foley L, Smith SM, Harrington P, Clyne B. The experience of financial burden for people with multimorbidity: a systematic review of qualitative research. Health Expect 2021 Apr 02;24(2):282-295 [FREE Full text] [doi: 10.1111/hex.13166] [Medline: 33264478]
15. Forjaz MJ, Rodriguez-Blazquez C, Ayala A, Rodriguez-Rodriguez V, de Pedro-Cuesta J, Garcia-Gutierrez S, et al. Chronic conditions, disability, and quality of life in older adults with multimorbidity in Spain. Eur J Intern Med 2015 Apr;26(3):176-181. [doi: 10.1016/j.ejim.2015.02.016] [Medline: 25724771]
16. Garin N, Olaya B, Moneta MV, Miret M, Lobo A, Ayuso-Mateos JL, et al. Impact of multimorbidity on disability and quality of life in the Spanish older population. PLoS One 2014 Nov 6;9(11):e111498 [FREE Full text] [doi: 10.1371/journal.pone.0111498] [Medline: 25375890]
17. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2016 Mar 14;3(3):CD006560 [FREE Full text] [doi: 10.1002/14651858.CD006560.pub3] [Medline: 26976529]
18. Smith SM, O'Dowd T. Chronic diseases: what happens when they come in multiples? Br J Gen Pract 2007 Apr;57(537):268-270 [FREE Full text] [Medline: 17394728]
19. Violan C, Foguet-Boreu Q, Flores-Mateo G, Salisbury C, Blom J, Freitag M, et al. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. PLoS One 2014 Jul 21;9(7):e102149 [FREE Full text] [doi: 10.1371/journal.pone.0102149] [Medline: 25048354]
20. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2021 Jan 15;1(1):CD006560 [FREE Full text] [doi: 10.1002/14651858.CD006560.pub4] [Medline: 33448337]
21. Multimorbidity: clinical assessment and management. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng56 [accessed 2024-04-29]
22. Muth C, Blom JW, Smith SM, Johnell K, Gonzalez-Gonzalez AI, Nguyen TS, et al. Evidence supporting the best clinical management of patients with multimorbidity and polypharmacy: a systematic guideline review and expert consensus. J Intern Med 2019 Mar 10;285(3):272-288 [FREE Full text] [doi: 10.1111/joim.12842] [Medline: 30357955]
23. Malhi GS, Bell E, Bassett D, Boyce P, Bryant R, Hazell P, et al. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry 2021 Jan 22;55(1):7-117. [doi: 10.1177/0004867420979353] [Medline: 33353391]
24. Hopman P, de Bruin SR, Forjaz MJ, Rodriguez-Blazquez C, Tonnara G, Lemmens LC, et al. Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty: a systematic literature review. Health Policy 2016 Jul;120(7):818-832. [doi: 10.1016/j.healthpol.2016.04.002] [Medline: 27114104]
25. Muth C, van den Akker M, Blom JW, Mallen CD, Rochon J, Schellevis FG, et al. The Ariadne principles: how to handle multimorbidity in primary care consultations. BMC Med 2014 Dec 08;12(1):223 [FREE Full text] [doi: 10.1186/s12916-014-0223-1] [Medline: 25484244]
26. Poitras ME, Maltais ME, Bestard-Denommé L, Stewart M, Fortin M. What are the effective elements in patient-centered and multimorbidity care? A scoping review. BMC Health Serv Res 2018 Jun 14;18(1):446 [FREE Full text] [doi: 10.1186/s12913-018-3213-8] [Medline: 29898713]
27. Melchiorre MG, Lamura G, Barbabella F, ICARE4EU Consortium. eHealth for people with multimorbidity: results from the ICARE4EU project and insights from the "10 e's" by Gunther Eysenbach. PLoS One 2018 Nov 14;13(11):e0207292 [FREE Full text] [doi: 10.1371/journal.pone.0207292] [Medline: 30427924]
28. Wiwatkunupakarn N, Aramrat C, Pliannuom S, Buawangpong N, Pinyopornpanish K, Nantsupawat N, et al. The integration of clinical decision support systems into telemedicine for patients with multimorbidity in primary care settings: scoping review. J Med Internet Res 2023 Jun 28;25:e45944 [FREE Full text] [doi: 10.2196/45944] [Medline: 37379066]
29. Tahsin F, Armas A, Kirakalaprathapan A, Kadu M, Sritharan J, Steele Gray C. Information and communications technologies enabling integrated primary care for patients with complex care needs: scoping review. J Med Internet Res 2023 Apr 19;25:e44035 [FREE Full text] [doi: 10.2196/44035] [Medline: 37074779]
30. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
31. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity. Eur J Gen Pract 2009 Jul 11;2(2):65-70. [doi: 10.3109/13814789609162146]
32. Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, et al. Multimorbidity. Nat Rev Dis Primers 2022 Jul 14;8(1):48 [FREE Full text] [doi: 10.1038/s41572-022-00376-4] [Medline: 35835758]
33. Harrison C, Fortin M, van den Akker M, Mair F, Calderon-Larranaga A, Boland F, et al. Comorbidity versus multimorbidity: why it matters. J Multimorb Comorb 2021 Mar 02;11:2633556521993993 [FREE Full text] [doi: 10.1177/2633556521993993] [Medline: 33718251]

XSL•FO
RenderX

34. Clarke J, Proudfoot J, Ma H. Mobile phone and web-based cognitive behavior therapy for depressive symptoms and mental health comorbidities in people living with diabetes: results of a feasibility study. JMIR Ment Health 2016 May 31;3(2):e23 [FREE Full text] [doi: 10.2196/mental.5131] [Medline: 27245948]
35. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A web-based cognitive behavior therapy intervention to improve social and occupational functioning in adults with type 2 diabetes (the SpringboarD trial): randomized controlled trial. J Med Internet Res 2019 May 21;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]
36. Monreal-Bartolomé A, Barceló-Soler A, Castro A, Pérez-Ara M, Gili M, Mayoral F, et al. Efficacy of a blended low-intensity internet-delivered psychological programme in patients with multimorbidity in primary care: study protocol for a randomized controlled trial. BMC Psychiatry 2019 Feb 11;19(1):66 [FREE Full text] [doi: 10.1186/s12888-019-2037-3] [Medline: 30744610]
37. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):S13-S27. [doi: 10.2337/dc18-S002] [Medline: 29222373]
38. Fundación K. Guía de práctica clínica: Lumbalgía inespecífica. Versión española de la guía de Práctica Clínica del Programa Europeo COST B13. European Commission. URL: https://tinyurl.com/5h6fftrs [accessed 2024-04-29]
39. Bogner HR, de Vries HF. Integration of depression and hypertension treatment: a pilot, randomized controlled trial. Ann Fam Med 2008 Jul 01;6(4):295-301 [FREE Full text] [doi: 10.1370/afm.843] [Medline: 18626028]
40. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med 2010 Dec 30;363(27):2611-2620 [FREE Full text] [doi: 10.1056/NEJMoa1003955] [Medline: 21190455]
41. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
42. Das P, Naylor C, Majeed A. Bringing together physical and mental health within primary care: a new frontier for integrated care. J R Soc Med 2016 Oct;109(10):364-366 [FREE Full text] [doi: 10.1177/0141076816665270] [Medline: 27729592]
43. Global report on diabetes. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241565257 [accessed 2024-04-29]
44. Abdi H. The greenhouse-geisser correction. Encycl Res Des 2010;1(1):544-548 [FREE Full text]
45. García-Herrera JM, Nogueras-Morillas V, Muñoz-Cobos F, Morales-Asensio JM. Guía de Práctica Clínica Para El Tratamiento de La Depresión En Atención Primaria. Servicio Andaluz de Salud - Junta de Andalucia. URL: https://consaludmental.org/publicaciones/GPCdepresionatencionprimaria.pdf [accessed 2022-02-25]
46. Cuijpers P, van Straten A, van Schaik A, Andersson G. Psychological treatment of depression in primary care: a meta-analysis. Br J Gen Pract 2009 Feb;59(559):e51-e60 [FREE Full text] [doi: 10.3399/bjgp09X395139] [Medline: 19192368]
47. Gill D, Hatcher S. Antidepressants for depression in medical illness. Cochrane Database Syst Rev 2000(4):CD001312. [doi: 10.1002/14651858.CD001312] [Medline: 11034712]
48. Kraef C, van der Meirschen M, Free C. Digital telemedicine interventions for patients with multimorbidity: a systematic review and meta-analysis. BMJ Open 2020 Oct 13;10(10):e036904 [FREE Full text] [doi: 10.1136/bmjopen-2020-036904] [Medline: 33051232]
49. van der Feltz-Cornelis CM, Nuyen J, Stoop C, Chan J, Jacobson AM, Katon W, et al. Effect of interventions for major depressive disorder and significant depressive symptoms in patients with diabetes mellitus: a systematic review and meta-analysis. Gen Hosp Psychiatry 2010 Jul;32(4):380-395. [doi: 10.1016/j.genhosppsych.2010.03.011] [Medline: 20633742]
50. van der Feltz-Cornelis C, Allen SF, Holt RI, Roberts R, Nouwen A, Sartorius N. Treatment for comorbid depressive disorder or subthreshold depression in diabetes mellitus: systematic review and meta-analysis. Brain Behav 2021 Feb 04;11(2):e01981 [FREE Full text] [doi: 10.1002/brb3.1981] [Medline: 33274609]
51. Franco P, Gallardo AM, Urtubey X. Web-based interventions for depression in individuals with diabetes: review and discussion. JMIR Diabetes 2018 Sep 14;3(3):e13 [FREE Full text] [doi: 10.2196/diabetes.9694] [Medline: 30291082]
52. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2019 Apr 1;42(2):169-187 [FREE Full text] [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
53. Savigny P, Watson P, Underwood M, Guideline Development Group. Early management of persistent non-specific low back pain: summary of NICE guidance. BMJ 2009 Jun 04;338(jun04 3):b1805. [doi: 10.1136/bmj.b1805] [Medline: 19502217]
54. Pei JH, Ma T, Nan RL, Chen H, Zhang Y, Gou L, et al. Mindfulness-based cognitive therapy for treating chronic pain a systematic review and meta-analysis. Psychol Health Med 2021 Mar 26;26(3):333-346. [doi: 10.1080/13548506.2020.1849746] [Medline: 33241941]
55. Lustman PJ, Clouse RE, Ciechanowski PS, Hirsch IB, Freedland KE. Depression-related hyperglycemia in type 1 diabetes: a mediational approach. Psychosom Med 2005;67(2):195-199. [doi: 10.1097/01.psy.0000155670.88919.ad] [Medline: 15784783]

XSL•FO
RenderX

56. Ngan HY, Chong YY, Chien WT. Effects of mindfulness- and acceptance-based interventions on diabetes distress and glycaemic level in people with type 2 diabetes: systematic review and meta-analysis. Diabet Med 2021 Apr;38(4):e14525. [doi: 10.1111/dme.14525] [Medline: 33438251]
57. Lu X, Yang D, Liang J, Xie G, Li X, Xu C, et al. Effectiveness of intervention program on the change of glycaemic control in diabetes with depression patients: a meta-analysis of randomized controlled studies. Prim Care Diabetes 2021 Jun;15(3):428-434. [doi: 10.1016/j.pcd.2021.01.006] [Medline: 33551333]
58. Kroenke K, Evans E, Weitlauf S, McCalley S, Porter B, Williams T, et al. Comprehensive vs. assisted management of mood and pain symptoms (CAMMPS) trial: study design and sample characteristics. Contemp Clin Trials 2018 Jan;64:179-187. [doi: 10.1016/j.cct.2017.10.006] [Medline: 29031492]
59. Ali MK, Chwastiak L, Poongothai S, Emmert-Fees KM, Patel SA, Anjana RM, INDEPENDENT Study Group. Effect of a collaborative care model on depressive symptoms and glycated hemoglobin, blood pressure, and serum cholesterol among patients with depression and diabetes in India: the INDEPENDENT randomized clinical trial. JAMA 2020 Aug 18;324(7):651-662 [FREE Full text] [doi: 10.1001/jama.2020.11747] [Medline: 32809002]
60. Metzner G, Horstmeier LM, Bengel J, Bitzer EM, Dreher E, Frank F, et al. Local, collaborative, stepped, and personalized care management for older people with chronic diseases - results from the randomized controlled LoChro-trial. BMC Geriatr 2023 Feb 13;23(1):92 [FREE Full text] [doi: 10.1186/s12877-023-03797-2] [Medline: 36782119]
61. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
62. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63(4):679-686. [doi: 10.1097/00006842-200107000-00021] [Medline: 11485122]
63. Servicio Andaluz de Salud - Junta de Andalucia. Ministerio de Sanidad, Servicios Sociales e Igualdad. URL: https://portal.guiasalud.es/wp-content/uploads/2020/10/gpc_534_depresion_adulto_avaliat_compl_caduc.pdf [accessed 2024-04-29]
64. American Diabetes Association. Standards of medical care in diabetes--2010. Diabetes Care 2010 Jan;33 Suppl 1(Suppl 1):S11-S61 [FREE Full text] [doi: 10.2337/dc10-S011] [Medline: 20042772]
65. Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B. The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. Pain 2001 Aug;93(2):173-183. [doi: 10.1016/S0304-3959(01)00314-1] [Medline: 11427329]
66. Kovacs FM, Llobera J, Gil Del Real MT, Abraira V, Gestoso M, Fernández C, et al. Validation of the spanish version of the Roland-Morris questionnaire. Spine (Phila Pa 1976) 2002 Mar 01;27(5):538-542. [doi: 10.1097/00007632-200203010-00016] [Medline: 11880841]
67. Ostelo RW, de Vet HC, Knol DL, van den Brandt PA. 24-item Roland Morris disability questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. J Clin Epidemiol 2004 Mar;57(3):268-276. [doi: 10.1016/j.jclinepi.2003.09.005] [Medline: 15066687]
68. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. Phys Ther 1996 Apr;76(4):359-366. [doi: 10.1093/ptj/76.4.359] [Medline: 8606899]
69. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
70. Fong TC, Lo TL, Ho RT. Psychometric properties of the 12-item stroke-specific quality of life scale among stroke survivors in Hong Kong. Sci Rep 2023 Jan 27;13(1):1510 [FREE Full text] [doi: 10.1038/s41598-023-28636-7] [Medline: 36707541]
71. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: results from the IQOLA project. International quality of life assessment. J Clin Epidemiol 1998 Nov;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7] [Medline: 9817135]
72. Vilagut G, María Valderas J, Ferrer M, Garin O, López-García E, Alonso J. [Interpretation of SF-36 and SF-12 questionnaires in Spain: physical and mental components]. Med Clin (Barc) 2008 May 24;130(19):726-735. [doi: 10.1157/13121076] [Medline: 18570798]
73. Alonso J, Prieto L, Antó JM. [The Spanish version of the SF-36 Health Survey (the SF-36 health questionnaire): an instrument for measuring clinical results]. Med Clin (Barc) 1995 May 27;104(20):771-776. [Medline: 7783470]
74. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
75. Sandín B, Chorot P, Lostao L, Joiner TE, Santed MA, Valiente RM. Escalas PANAS de Afecto Positivo y Negativo: validación factorial y convergencia transcultural. Psichothema 1999;11(1):37-51 [FREE Full text]
76. Crawford JR, Henry JD. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 2004 Sep;43(Pt 3):245-265. [doi: 10.1348/0144665031752934] [Medline: 15333231]
77. Botella C, Molinari G, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños RM, et al. Development and validation of the openness to the future scale: a prospective protective factor. Health Qual Life Outcomes 2018 Apr 23;16(1):72 [FREE Full text] [doi: 10.1186/s12955-018-0889-8] [Medline: 29688854]
78. Stata statistical software: release 18. StataCorp LLC. URL: https://www.stata.com/ [accessed 2024-04-29]

XSL•FO
RenderX

79. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Cambridge, MA: Academic Press; 1985.
80. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Statist 2001 Aug 1;29(4):1165-1188. [doi: 10.1214/aos/1013699998]
81. Dunn G, Bentall R. Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). Stat Med 2007 Nov 20;26(26):4719-4745. [doi: 10.1002/sim.2891] [Medline: 17476649]
82. Mackinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. Multivariate Behav Res 2004 Jan 01;39(1):99-128 [FREE Full text] [doi: 10.1207/s15327906mbr3901_4] [Medline: 20157642]
83. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(mar23 1):c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
84. Altman DG. Comparability of randomised groups. Statistician 1985;34(1):125. [doi: 10.2307/2987510]
85. Hajihasani A, Rouhani M, Salavati M, Hedayati R, Kahlaee AH. The influence of cognitive behavioral therapy on pain, quality of life, and depression in patients receiving physical therapy for chronic low back pain: a systematic review. PM R 2019 Feb;11(2):167-176. [doi: 10.1016/j.pmrj.2018.09.029] [Medline: 30266349]
86. Petrucci G, Papalia GF, Russo F, Vadalà G, Piredda M, De Marinis MG, et al. Psychological approaches for the integrative care of chronic low back pain: a systematic review and metanalysis. Int J Environ Res Public Health 2021 Dec 22;19(1):60 [FREE Full text] [doi: 10.3390/ijerph19010060] [Medline: 35010319]
87. Li C, Xu D, Hu M, Tan Y, Zhang P, Li G, et al. A systematic review and meta-analysis of randomized controlled trials of cognitive behavior therapy for patients with diabetes and depression. J Psychosom Res 2017 Apr;95:44-54. [doi: 10.1016/j.jpsychores.2017.02.006] [Medline: 28314548]
88. Atlantis E, Fahey P, Foster J. Collaborative care for comorbid depression and diabetes: a systematic review and meta-analysis. BMJ Open 2014 Apr 12;4(4):e004706 [FREE Full text] [doi: 10.1136/bmjopen-2013-004706] [Medline: 24727428]
89. Varela-Moreno E, Carreira Soler M, Guzmán-Parra J, Jódar-Sánchez F, Mayoral-Cleries F, Anarte-Ortíz MT. Effectiveness of eHealth-based psychological interventions for depression treatment in patients with type 1 or type 2 diabetes mellitus: a systematic review. Front Psychol 2021 Jan 31;12:746217 [FREE Full text] [doi: 10.3389/fpsyg.2021.746217] [Medline: 35173644]
90. Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017 May 15;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
91. Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015 May;38(5):776-783 [FREE Full text] [doi: 10.2337/dc14-1728] [Medline: 25710923]
92. Dear BF, Titov N, Perry KN, Johnston L, Wootton BM, Terides MD, et al. The Pain Course: a randomised controlled trial of a clinician-guided internet-delivered cognitive behaviour therapy program for managing chronic pain and emotional well-being. Pain 2013 Jun;154(6):942-950. [doi: 10.1016/j.pain.2013.03.005] [Medline: 23688830]
93. Peters ML, Smeets E, Feijge M, van Breukelen G, Andersson G, Buhrman M, et al. Happy despite pain: a randomized controlled trial of an 8-week internet-delivered positive psychology intervention for enhancing well-being in patients with chronic pain. Clin J Pain 2017 Nov;33(11):962-975 [FREE Full text] [doi: 10.1097/AJP.0000000000000494] [Medline: 28379873]
94. Baldwin PA, Sanatkar S, Clarke J, Fletcher S, Gunn J, Wilhelm K, et al. A web-based mental health intervention to improve social and occupational functioning in adults with type 2 diabetes (the Springboard Trial): 12-month outcomes of a randomized controlled trial. J Med Internet Res 2020 Dec 01;22(12):e16729 [FREE Full text] [doi: 10.2196/16729] [Medline: 33258790]
95. Abbas Q, Latif S, Ayaz Habib H, Shahzad S, Sarwar U, Shahzadi M, et al. Cognitive behavior therapy for diabetes distress, depression, health anxiety, quality of life and treatment adherence among patients with type-II diabetes mellitus: a randomized control trial. BMC Psychiatry 2023 Feb 03;23(1):86 [FREE Full text] [doi: 10.1186/s12888-023-04546-w] [Medline: 36737757]
96. Garg S, Garg D, Turin TC, Chowdhury MF. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
97. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The Pain Course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
98. Øverås CK, Nilsen TI, Nicholl BI, Rughani G, Wood K, Søgaard K, et al. Multimorbidity and co-occurring musculoskeletal pain do not modify the effect of the SELFBACK app on low back pain-related disability. BMC Med 2022 Feb 08;20(1):53 [FREE Full text] [doi: 10.1186/s12916-022-02237-z] [Medline: 35130898]
99. Baumeister H, Paganini S, Sander LB, Lin J, Schlicker S, Terhorst Y, et al. Effectiveness of a guided internet- and mobile-based intervention for patients with chronic back pain and depression (WARD-BP): a multicenter, pragmatic randomized controlled trial. Psychother Psychosom 2021;90(4):255-268. [doi: 10.1159/000511881] [Medline: 33321501]

XSL•FO
RenderX

100. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behav Res Ther 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
101. Sander LB, Paganini S, Terhorst Y, Schlicker S, Lin J, Spanhel K, et al. Effectiveness of a guided web-based self-help intervention to prevent depression in patients with persistent back pain: the PROD-BP randomized clinical trial. JAMA Psychiatry 2020 Oct 01;77(10):1001-1011 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1021] [Medline: 32459348]
102. Schlicker S, Baumeister H, Buntrock C, Sander L, Paganini S, Lin J, et al. A web- and mobile-based intervention for comorbid, recurrent depression in patients with chronic back pain on sick leave (Get.Back): pilot randomized controlled trial on feasibility, user satisfaction, and effectiveness. JMIR Ment Health 2020 Apr 15;7(4):e16398 [FREE Full text] [doi: 10.2196/16398] [Medline: 32293577]
103. Chiauzzi E, Pujol LA, Wood M, Bond K, Black R, Yiu E, et al. painACTION-back pain: a self-management website for people with chronic back pain. Pain Med 2010 Jul 01;11(7):1044-1058. [doi: 10.1111/j.1526-4637.2010.00879.x] [Medline: 20545873]
104. Safren SA, Gonzalez JS, Wexler DJ, Psaros C, Delahanty LM, Blashill AJ, et al. A randomized controlled trial of cognitive behavioral therapy for adherence and depression (CBT-AD) in patients with uncontrolled type 2 diabetes. Diabetes Care 2014;37(3):625-633 [FREE Full text] [doi: 10.2337/dc13-0816] [Medline: 24170758]
105. Simson U, Nawarotzky U, Friese G, Porck W, Schottenfeld-Naor Y, Hahn S, et al. Psychotherapy intervention to reduce depressive symptoms in patients with diabetic foot syndrome. Diabet Med 2008 Feb 15;25(2):206-212. [doi: 10.1111/j.1464-5491.2007.02370.x] [Medline: 18290863]
106. Petrak F, Herpertz S, Albus C, Hermanns N, Hiemke C, Hiller W, et al. Cognitive behavioral therapy versus sertraline in patients with depression and poorly controlled diabetes: the Diabetes and Depression (DAD) study: a randomized controlled multicenter trial. Diabetes Care 2015 May;38(5):767-775. [doi: 10.2337/dc14-1599] [Medline: 25690005]
107. Gois C, Dias VV, Carmo I, Duarte R, Ferro A, Santos AL, et al. Treatment response in type 2 diabetes patients with major depression. Clin Psychol Psychother 2014 Sep 10;21(1):39-48. [doi: 10.1002/cpp.1817] [Medline: 22962030]
108. de Groot M, Shubrook JH, Hornsby WG, Pillay Y, Mather KJ, Fitzpatrick K, et al. Program ACTIVE II: outcomes from a randomized, multistate community-based depression treatment for rural and urban adults with type 2 diabetes. Diabetes Care 2019 Jul;42(7):1185-1193 [FREE Full text] [doi: 10.2337/dc18-2400] [Medline: 31221693]
109. Lustman PJ, Griffith LS, Freedland KE, Kissel SS, Clouse RE. Cognitive behavior therapy for depression in type 2 diabetes mellitus. a randomized, controlled trial. Ann Intern Med 1998 Oct 15;129(8):613-621. [doi: 10.7326/0003-4819-129-8-199810150-00005] [Medline: 9786808]
110. Ebert DD, Nobis S, Lehr D, Baumeister H, Riper H, Auerbach RP, et al. The 6-month effectiveness of Internet-based guided self-help for depression in adults with Type 1 and 2 diabetes mellitus. Diabet Med 2017 Jan 04;34(1):99-107. [doi: 10.1111/dme.13173] [Medline: 27334444]
111. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011 Feb;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
112. Naik AD, Hundt NE, Vaughan EM, Petersen NJ, Zeno D, Kunik ME, et al. Effect of telephone-delivered collaborative goal setting and behavioral activation vs enhanced usual care for depression among adults with uncontrolled diabetes: a randomized clinical trial. JAMA Netw Open 2019 Aug 02;2(8):e198634 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.8634] [Medline: 31390035]
113. Piette JD, Richardson C, Himle J, Duffy S, Torres T, Vogel M, et al. A randomized trial of telephonic counseling plus walking for depressed diabetes patients. Med Care 2011 Jul;49(7):641-648 [FREE Full text] [doi: 10.1097/MLR.0b013e318215d0c9] [Medline: 21478777]
114. Hoyo ML, Rodrigo MT, Urcola-Pardo F, Monreal-Bartolomé A, Ruiz DC, Borao MG, et al. The TELE-DD randomised controlled trial on treatment adherence in patients with type 2 diabetes and comorbid depression: clinical outcomes after 18-month follow-up. Int J Environ Res Public Health 2022 Dec 25;20(1):328 [FREE Full text] [doi: 10.3390/ijerph20010328] [Medline: 36612650]
115. van Son J, Nyklícek I, Pop VJ, Blonk MC, Erdtsieck RJ, Spooren PF, et al. The effects of a mindfulness-based intervention on emotional distress, quality of life, and HbA(1c) in outpatients with diabetes (DiaMind): a randomized controlled trial. Diabetes Care 2013 Apr;36(4):823-830 [FREE Full text] [doi: 10.2337/dc12-1477] [Medline: 23193218]
116. Cherkin DC, Sherman KJ, Balderson BH, Cook AJ, Anderson ML, Hawkes RJ, et al. Effect of mindfulness-based stress reduction vs cognitive behavioral therapy or usual care on back pain and functional limitations in adults with chronic low back pain: a randomized clinical trial. JAMA 2016 Mar 22;315(12):1240-1249 [FREE Full text] [doi: 10.1001/jama.2016.2323] [Medline: 27002445]
117. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]

XSL•FO
RenderX

118. Nowlan JS, Wuthrich VM, Rapee RM, Kinsella JM, Barker G. A comparison of single-session positive reappraisal, cognitive restructuring and supportive counselling for older adults with type 2 diabetes. Cogn Ther Res 2015 Dec 9;40(2):216-229. [doi: 10.1007/s10608-015-9737-x]
119. Dunn BD, Warbrick L, Hayes R, Montero-Marin J, Reed N, Dalgleish T, et al. Does mindfulness-based cognitive therapy with tapering support reduce risk of relapse/recurrence in major depressive disorder by enhancing positive affect? A secondary analysis of the PREVENT trial. J Consult Clin Psychol 2022;92(9):22-57 [FREE Full text] [doi: 10.1037/ccp0000902.supp]
120. Kanuch SW, Cassidy KA, Dawson NV, Athey M, Fuentes-Casiano E, Sajatovic M. Recruiting and retaining individuals with serious mental illness and diabetes in clinical research: lessons learned from a randomized, controlled trial. J Health Dispar Res Pract 2016;9(3):115-126 [FREE Full text] [Medline: 28533944]
121. Breckner A, Litke N, Göbl L, Wiezorreck L, Miksch A, Szecsenyi J, et al. Effects and processes of an mHealth intervention for the management of chronic diseases: prospective observational study. JMIR Form Res 2022 Aug 25;6(8):e34786 [FREE Full text] [doi: 10.2196/34786] [Medline: 36006666]
122. Clarke J, Proudfoot J, Vatiliotis V, Verge C, Holmes-Walker DJ, Campbell L, et al. Attitudes towards mental health, mental health research and digital interventions by young adults with type 1 diabetes: a qualitative analysis. Health Expect 2018 Jun 10;21(3):668-677 [FREE Full text] [doi: 10.1111/hex.12662] [Medline: 29319923]
123. Sathian B, Asim M, Banerjee I, Pizarro AB, Roy B, van Teijlingen ER, et al. Impact of COVID-19 on clinical trials and clinical research: a systematic review. Nepal J Epidemiol 2020 Sep 30;10(3):878-887 [FREE Full text] [doi: 10.3126/nje.v10i3.31622] [Medline: 33042591]
124. Psaty BM, Rennie D. Stopping medical research to save money: a broken pact with researchers and patients. JAMA 2003 Apr 23;289(16):2128-2131. [doi: 10.1001/jama.289.16.2128] [Medline: 12709471]

Abbreviations

ADA: American Diabetes Association
CACE: complier average causal effect
CEICA: Research Ethics Committee of the Autonomous Community of Aragon
CONSORT: Consolidated Standards of Reporting Trials
FPS-R: Faces Pain Scale-Revised
GP: general practitioner
HbA_{1c}: glycated hemoglobin
ICT: information and communication technology
iTAU: improved treatment as usual
ITT: intention-to-treat
IV: instrumental variable
MBCT: mindfulness-based cognitive therapy
OFS: Openness to the Future Scale
PANAS: Positive and Negative Affect Schedule
PC: primary care
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
RM: repeated measure
RMDQ: Roland-Morris Disability Questionnaire
SF-12: 12-item Short Form Survey

Edited by A Coristine; submitted 09.01.24; peer-reviewed by F Zhu, W Zhang, F Prazeres, A Hassan; comments to author 08.04.24; revised version received 03.06.24; accepted 09.10.24; published 10.02.25.

Please cite as:
Monreal-Bartolomé A, Castro A, Pérez-Ara MÁ, Gili M, Mayoral F, Hurtado MM, Varela Moreno E, Botella C, García-Palacios A, Baños RM, López-Del-Hoyo Y, García-Campayo J, Montero-Marin J
Efficacy of a Blended Low-Intensity Internet-Delivered Psychological Program in Patients With Multimorbidity in Primary Care: Randomized Controlled Trial
J Med Internet Res 2025;27:e56203
URL: https://www.jmir.org/2025/1/e56203
doi:10.2196/56203
PMID:39928931

XSL•FO
RenderX

©Alicia Monreal-Bartolomé, Adoración Castro, M Ángeles Pérez-Ara, Margalida Gili, Fermín Mayoral, María Magdalena Hurtado, Esperanza Varela Moreno, Cristina Botella, Azucena García-Palacios, Rosa M Baños, Yolanda López-Del-Hoyo, Javier García-Campayo, Jesus Montero-Marin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Individualized Follow-Up With an App Postcardiac Rehabilitation: Five-Year Follow-Up of a Randomized Controlled Trial

Pernille Lunde[1], PhD; Asta Bye[2,3], PhD; Jostein Grimsmo[4], PhD; Are Hugo Pripp[5,6], PhD; Vibeke Ritschel[7], PhD; Even Jarstad[7], MSc; Birgitta Blakstad Nilsson[1,8], PhD

[1]Department of Rehabilitation Science and Health Technology, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway

[2]Department of Nursing and Health Promotion, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway

[3]European Palliative Care Research Centre (PRC), Department of Oncology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[4]Department of Cardiac and Pulmonary Rehabilitation, Lovisenberg Rehabilitation, Cathinka Guldberg's Hospital, Jessheim, Norway

[5]Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway

[6]Oslo Centre of Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

[7]Department of Cardiology and Exercise Physiology, Norwegian Sport Medicine Clinic (Nimi, part of Volvat), Oslo, Norway

[8]Section for Physiotherapy, Division of Medicine, Oslo University Hospital, Oslo, Norway

Corresponding Author:
Pernille Lunde, PhD
Department of Rehabilitation Science and Health Technology
Faculty of Health Sciences
OsloMet – Oslo Metropolitan University
PB 4, St. Olavs plass
Oslo, 0130
Norway
Phone: 47 48063537
Email: plunde@oslomet.no

Abstract

Background: Adherence to healthy behaviors initiated or adapted during cardiac rehabilitation (CR) remains a significant challenge, with few patients meeting guideline standards for secondary prevention. The use of mobile health (mHealth) interventions has been proposed as a potential solution to improve adherence to healthy behaviors after CR. In particular, app-based interventions have shown promise due to their ability to provide monitoring and feedback anytime and anywhere. Growing evidence supports the use of apps in post-CR settings to enhance adherence. In 2020, we demonstrated that individualized follow-up via an app increased adherence to healthy behaviors 1 year after CR. However, it remains uncertain whether these effects persist once the follow-up is discontinued.

Objective: This study aims to evaluate the long-term effects of individualized follow-up using an app, assessed 4 years after the intervention.

Methods: A single-blinded multicenter randomized controlled trial was conducted. Patients were recruited from 2 CR centers in eastern Norway. The intervention group (IG) received individualized follow-up through an app for 1 year, while the control group (CG) received usual care. After the 1-year follow-up, the app-based follow-up was discontinued for the IG, and both groups were encouraged to maintain or improve their healthy behaviors based on their individual risk profiles. The primary outcome was the difference in peak oxygen uptake (VO_{2peak}). The secondary outcomes included exercise performance, body weight, blood pressure, lipid profile, exercise habits, health-related quality of life, health status, cardiac events, and physical activity. Linear mixed models for repeated measurements were used to analyze differences between groups. All tests were 2-sided, and *P* values ≤0.05 were considered statistically significant.

Results: At the 5-year follow-up, 101 out of the initial 113 randomized participants were reassessed. Intention-to-treat analyses, using a mixed model for repeated measurements, revealed a statistically significant difference (*P*=.04) in exercise habits in favor of the IG, with a mean difference of 0.67 (95% CI 0.04-1.29) exercise sessions per week. Statistically significant differences were

XSL•FO
RenderX

also observed in triglycerides (mean difference 0.40, 95% CI 0.00-0.79 mmol/l, *P*=.048) and walking (*P*=.03), but these were in favor of the CG. No differences were found between the groups for other evaluated outcomes.

Conclusions: Most of the benefits derived from the app-based follow-up diminished by 4 years after the intervention. Although the IG reported statistically significantly higher levels of exercise, this did not translate into improved VO_{2peak} or exercise performance. Our study highlights the need for follow-up from health care providers to enhance adherence to healthy behaviors in the long term following CR.

Trial Registration: ClinicalTrials.gov NCT03174106; https://clinicaltrials.gov/ct2/show/NCT03174106 (original study protocol) and NCT05697120; https://clinicaltrials.gov/ct2/show/NCT05697120 (updated study protocol)

(J Med Internet Res 2025;27:e60256) doi:10.2196/60256

KEYWORDS

mHealth; cardiac rehabilitation; mobile phone app; smartphone; lifestyle

Introduction

The beneficial effects of comprehensive cardiac rehabilitation (CR) have been well established, and for the past decade, CR has carried a class IA recommendation in the European guidelines on cardiovascular disease [1,2]. The overarching goal of CR and secondary prevention is to prevent subsequent cardiac events [1]. Achieving this goal relies heavily on adherence to healthy behavior [1,3]. However, maintaining healthy behavior initiated or adapted during CR is challenging for most patients, resulting in only a small proportion achieving the guideline standards for secondary prevention [3-5]. Accordingly, the long-term (≥1 year) outcomes of CR are poor, and evaluations of CR with follow-ups exceeding 1 year are sparse.

The recently published statement from the European Association of Preventive Cardiology on optimizing adherence in the secondary prevention of cardiovascular disease [3] recommends mobile health (mHealth) interventions to address challenges related to adherence to healthy behavior after CR. mHealth was defined by the World Health Organization in 2011 as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [6]. It includes the use of mobile phones for calls, short messaging, and more advanced functionalities, such as apps that utilize packet radio service, 3G and 4G systems, global positioning systems, and Bluetooth technology for medical and public health practice [6]. The use of apps was suggested early as particularly promising, as it enables monitoring and providing feedback to patients from anywhere, at any time [7]. Despite rapid advancements in the mHealth research field, only a few studies have investigated whether app use can support adherence to lifestyle changes after completing CR. Duscha and colleagues [8] conducted a randomized controlled trial to evaluate the impact of an mHealth intervention, including an app, on exercise capacity (peak oxygen uptake [VO_{2peak}]) after CR [8]. Among 32 patients randomized, they observed a significant increase in absolute VO_{2peak} at the 12-week follow-up. The TeleCare study (n=122) investigated the effects of telemonitoring (smartphone and Bluetooth heart rate monitor belt) and telecoaching (monthly calls over 6 months) compared with usual care on VO_{2peak} and other cardiovascular risk factors [9]. In addition to assessing patients at the 6-month follow-up at the end of the intervention, they re-evaluated them 6 months after the intervention (12 months after CR) [9]. No significant differences were observed between the groups at either follow-up [9].

In 2020, we demonstrated that individualized follow-up for 1 year using an app improved adherence to healthy behaviors after CR [10]. In this study, 113 patients were randomly assigned to individualized follow-up enabled by an app or to a control group (CG) receiving usual care. At the 1-year follow-up, there was a statistically significant difference between the groups, favoring the intervention group (IG), in VO_{2peak} (ml/kg/minute), exercise performance, exercise habits, and self-perceived goal achievement [10]. At the 1-year follow-up, the individualized follow-up ceased, and both the CG and the IG were encouraged to maintain or improve their healthy behaviors based on their individual risk profiles. To the best of the authors' knowledge, our study was the first to evaluate the effect of an app over an entire year in a post-CR setting. However, no research to date has investigated the long-term effects of successful mHealth interventions beyond 1 year after app use is discontinued. Therefore, the primary aim of this study was to examine whether the improvement in VO_{2peak} observed after 1 year [10] was maintained 4 years after the intervention. Similarly, as secondary outcomes, we aimed to evaluate exercise performance, body weight, resting blood pressure, lipid profile, exercise habits, health-related quality of life (HRQL), health status, cardiac events, and level of physical activity.

Methods

Design

We conducted a single-blinded, randomized controlled trial with 2 arms. Patients were allocated to either the CG or the IG in a 1:1 ratio using a computer-generated, permuted block randomization scheme. The randomization was stratified by the CR program. A statistician (AHP) performed the randomization and sent the schemes to one of the coauthors (BBN), who prepared sealed envelopes. The first author, who also recruited the patients and conducted baseline assessments, brought the sealed envelopes to the baseline assessments. The envelopes were opened immediately after each baseline assessment, and patients were informed of their group allocation. The IG received individualized follow-up through an app for 1 year, while the CG received usual care, which included

XSL•FO
RenderX

recommendations to maintain or further improve their lifestyle. Reporting adhered to the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement [11].

Setting and Participants

Patients were recruited from 2 CR centers in the eastern part of Norway. One center offers 1- and 4-week inpatient CR programs, while the other provides a 12-week outpatient CR program. The 12-week outpatient rehabilitation program includes exercise and education sessions 3 times a week, along with personalized support based on patients' needs, such as medication optimization, psychological counseling, dietary advice, and assistance with smoking cessation. This program has previously been evaluated and shown to improve exercise capacity and quality of life [12]. The 1- and 4-week residential programs had the same content as the 12-week outpatient program. However, the 4-week residential program provided patients with more time to adapt to lifestyle changes and learn about their condition compared with the 1-week program. The 4-week residential program has previously been reported as standard care [13]. Cardiologists, physiotherapists, dietitians, psychologists, and nurses were part of the health care team in all CR programs. Patients were recruited from 3 different CR programs, with randomization stratified accordingly. Approximately one-third of the patients were recruited from each program. Patients attending these CR programs are referred by a physician for various forms of heart disease, with coronary artery disease (CAD) being the most common reason for referral.

During participation in CR, all patients received oral information about the study from the first author (PL). Patients who met the inclusion criteria and were willing and able to participate were scheduled for a baseline assessment. Baseline assessments were conducted after a cardiopulmonary exercise test (CPET), during which patients were screened for exclusion criteria. As previously described [10,14], inclusion criteria included patients completing one of the CR programs, aged ≥40 years, ownership and use of an Android (Google LLC/Alphabet Inc.) or Apple (Apple Inc.) smartphone, and the ability to read and understand Norwegian or English. Exclusion criteria included ischemia or arrhythmias identified during CPET that imposed restrictions equivalent to <80% of maximal heart rate or a Borg scale (6-20) score [15] <15 during exercise. Patients with muscular or skeletal disorders that significantly affected exercise capacity more than their heart disease were excluded. Additionally, patients with severe malignant diseases (ie, advanced cancer) that had a greater impact on life expectancy than their heart disease were also excluded.

Recruitment for the 5-year follow-up was conducted by sending email and SMS text messages to all patients included in the original study [10]. The mail included written information about this study and an appointment for the 5-year follow-up, sent 4-6 weeks before the scheduled appointment. Patients were encouraged to confirm their willingness to participate and notify the first author if they planned to attend. Efforts were made to accommodate each patient's schedule.

Intervention Group

After the baseline assessment, patients randomly assigned to the IG were given access to an app and trained on how to use it. The app allowed users to create and set goals, with tasks and automatic reminders. A detailed description of the intervention is provided elsewhere [10,14]. Following the baseline assessment, the app was personalized with each patient's goals and tasks for the upcoming year. Patients could choose when and how often they wanted reminders for their tasks. Additionally, they could use the app for reflections or notes related to a specific goal and directly communicate any questions to the supervisor. The supervisor had access to an administrator interface, which allowed them to monitor each patient. The interface also enabled the supervisor to send short motivational messages (maximum 112 characters) directly to each patient through the app.

Patients in the IG received comprehensive individualized feedback based on their goals, completed tasks, pending tasks, and any notes they had written. This feedback was provided weekly for the first 12 weeks and monthly for the remainder of the year. Additionally, all patients received 1-3 short individualized motivational messages per week throughout the year. The supervisor, a specialized physiotherapist in this study, provided all feedback and motivational messages to the patients in the IG. Questions submitted to the supervisor were answered within 2 working days. This individualized follow-up continued until the 1-year follow-up assessment. After the 1-year follow-up, the app was removed from the patients' smartphones, and the follow-up was discontinued. At this point, patients were encouraged to maintain or improve their healthy behaviors based on their individual risk profiles.

Control Group

Patients allocated to the CG received usual care, which included visits to their general practitioner and cardiologist as needed. After the baseline assessment, they were encouraged to maintain or improve healthy behaviors based on their individual needs and goals. The same encouragement was provided at the 1-year follow-up assessment.

Outcomes and Assessments

Assessment Protocol and Outcome Measures Over a 5-Year Follow-Up Period

Assessments were performed at baseline (post-CR), after 1 year, and after 5 years at the same CR center where the patient was recruited. During the baseline assessment, demographic data were collected. The primary outcome was VO_{2peak}. Secondary outcomes at the 1- and 5-year follow-ups included exercise performance (time to exhaustion, peak incline, and peak velocity), body weight, resting blood pressure, lipid profile, exercise habits, HRQL, and health status. Additionally, at the 5-year follow-up, new cardiac events and levels of physical activity were assessed. Test personnel measuring the primary outcome were blinded to group allocation. Data collection was organized and conducted by the same person at all follow-ups, and the same test personnel led the tests at all 3 assessments.

XSL•FO
RenderX

Peak Oxygen Uptake and Exercise Performance

To ensure eligibility for the study, all patients performed a CPET before entering the study and at all follow-up times (1 and 5 years). The test was used to determine VO_{2peak} and exercise performance and is described in detail elsewhere [10,14]. Briefly, 2 standardized protocols were created: a walking protocol and a running protocol. Experienced test personnel selected the most suitable protocol for each patient, and the same protocol was used for all 3 tests. During all tests, patients were strongly encouraged to exercise to exhaustion; the Borg scale (6–20) [15] was used to assess perceived exertion, and VO_{2peak} was taken as the highest 30-second VO_2 measurement. In case of suspected submaximal effort, the test was repeated. Exercise performance was evaluated by time to exhaustion (seconds), peak incline (%), and peak velocity (km/hour). All CPETs were performed with continuous 12-lead electrocardiogram monitoring. Pulmonary ventilation and gases (oxygen and carbon dioxide) were recorded breath-by-breath using a Vyntus CPX metabolic analyzer (Vyaire Medical) at 1 of the CR centers. At the other CR center, pulmonary ventilation and gases were analyzed using a Schiller Ganshorn ergo-spirometry system (Schiller AB) at baseline, and with a Vyntus CPX (Customed) at the 1- and 5-year follow-ups. Patients were instructed to take medication and eat and drink as usual before all tests.

Body Weight

Body weight was measured without shoes and while wearing exercise clothes before the CPET at all assessments. Efforts were made to use the same equipment for all measurements.

Blood Pressure

Blood pressure was measured manually before the CPET. Patients were instructed to relax in a chair for a minimum of 3-5 minutes before the measurements were taken. Three measurements were performed, preferably on the left arm, and the lowest recorded value was used.

Lipid Profile

Venous blood was drawn using standard local procedures at the general practitioner's clinic within 4 weeks before each assessment. Patients were instructed to fast overnight before providing blood samples. Data were collected on low-density lipoprotein (LDL)-cholesterol, high-density lipoprotein (HDL)-cholesterol, total cholesterol, and triglycerides. Patients were required to bring the results to the assessment points.

Exercise Habits

At all assessments, patients were asked about their exercise habits. Exercise habits were defined as the average number of exercise sessions per week over the past year. An exercise session was further defined as a structured activity lasting at least 30 minutes, during which the individual became both sweaty and breathless and felt the need to shower afterward.

Health-Related Quality of Life

HRQL was assessed using the HeartQoL questionnaire, a disease-specific HRQL tool comprising 14 questions [16]. This questionnaire is both valid and reliable in patients referred to CR [17-21]. It includes 2 subscales: a physical HRQL subscale (10 items) and an emotional HRQL subscale (4 items) [16]. The combination of these subscales provides a global scale score [16]. Scores range from 0 to 3, with higher scores indicating better HRQL.

Health Status

Health status was assessed using the EQ-5D questionnaire, which consists of 5 questions [22]. Each question addresses a different dimension of health status: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [22]. Each question offers 5 response options, with a score of 1 indicating the best possible outcome and a score of 5 indicating the worst [22]. The EQ-5D also includes an overall health question (EQ Visual Analog Scale), which is answered on a Likert scale ranging from 0 to 100, where 0 represents the worst possible health and 100 represents the best possible health [22].

Cardiac Events

At the 5-year follow-up, new cardiac events were assessed. Patients were asked if they had experienced any cardiac events, with follow-up questions posed as needed. New cardiac events were defined as documented acute coronary syndrome, newly diagnosed CAD, valve surgery, survival of cardiac arrest, or atrial fibrillation (AF). Additionally, it was noted whether patients had undergone coronary angiography, percutaneous coronary intervention, or coronary artery bypass grafting, as well as any other cardiac-related incidents.

Physical Activity

At the 5-year follow-up, physical activity was assessed using the short version of the International Physical Activity Questionnaire (IPAQ). This questionnaire was designed as a tool for cross-national assessment of physical activity [23]. Both long and short versions of the IPAQ are available, and 2 reference periods can be evaluated: the "last 7 days" or a "usual week" [23]. In this study, the "last 7 days" reference period was used. The short version of the IPAQ provides information on time spent walking, as well as engaging in moderate-intensity and vigorous-intensity activities [23]. The instrument has been shown to have sufficient validity and reliability within a Norwegian population [24]. IPAQ data were reported following the recommendations outlined in the user manual [23].

Sample Size

The sample size was calculated based on the primary outcome, assuming a clinically significant difference in relative VO_{2peak} between groups of 3.5 ml/kg/minute [25-27]. Using data from a previously conducted feasibility study [28], the SD was estimated to be 6 ml/kg/minute. With a power of 80% and a significance level of .05, the required sample size was determined to be 47 patients per group. To account for a 20% dropout rate at the 1-year follow-up, we aimed to include a total of 113 patients.

Statistical Analysis

IBM SPSS Statistics (version 29) and Stata (version 18; StataCorp LLC) were used for statistical analysis. Continuous, normally distributed baseline data were analyzed using an independent *t* test to assess differences between groups, while

XSL•FO
RenderX

the Pearson chi-square test was used for categorical data. Baseline differences between participants with and without 5-year outcome data were analyzed using the same statistical tests. Missing data were assessed according to established strategies for handling missing data in clinical trials [29]. Differences between groups in primary and secondary outcomes were evaluated using a mixed model for repeated measurements. This model included a random intercept of the patient, as well as group, follow-up time, the interaction between group and follow-up time, and the baseline outcome measure as fixed effects. The Mann-Whitney *U* test was used to analyze differences between groups in IPAQ data at the 5-year follow-up. Independent *t* tests and Pearson chi-square tests were used to assess differences between participants with and without 5-year primary outcome data. This analysis was conducted to determine whether imputation of missing values was necessary, in accordance with strategies for managing missing data in clinical trials [29]. All analyses were performed on an intention-to-treat basis, and all tests were 2-sided. Data are presented as mean (SD) unless otherwise specified. A *P* value <.05 was considered statistically significant.

Ethical Considerations

The Regional Committee for Medical and Health Research Ethics (South-East, ID: 2016-1476) approved the study protocol, and the study was conducted in accordance with the Declaration of Helsinki. An updated study protocol for this follow-up study was submitted and approved. All patients provided written, informed consent before inclusion in the study. The study data reported in this manuscript are anonymous. The original and updated study protocols were registered on ClinicalTrials.gov with the IDs NCT03174106 and NCT05697120, respectively. The study protocol and 1-year results have been previously published [10,14].

Results

A total of 177 patients were screened for eligibility at 2 CR centers between October 2017 and June 2018, of whom 113 were included and randomized to the IG or CG (see Figure 1). The 5-year follow-up was completed in September 2023, with a total of 101 patients assessed. In the IG, 2 patients were unable to perform the CPET due to advanced cancer and complex musculoskeletal disorders. In the CG, 2 patients were unable to perform the CPET due to complex musculoskeletal and mental health disorders. This left 97 patients (85.8%) available for analysis of the primary outcome and 101 patients (89.4%) for secondary outcomes (Figure 1).

There were no differences in baseline characteristics between participants with missing data at the 5-year follow-up and those without missing data. Missing data were assumed to be completely at random, as they were not related to any observed or unobserved variables [29]. Consequently, imputation of missing data was not performed [29].

No statistically significant differences in baseline characteristics were observed between the IG and CG (Table 1). Additionally, there were no differences between the groups in changes in medication from baseline to the 1-year follow-up [10] or at the 5-year follow-up.

At the 5-year follow-up, no statistically significant difference (*P*=.69) in VO_{2peak} was observed between the groups. Other outcomes evaluated at the 5-year follow-up are presented in Table 2. A statistically significant difference was found in exercise habits (*P*=.04), favoring the IG, and in triglycerides, *P*=.048), favoring the CG. No statistically significant differences were observed between the CG and IG for any other outcomes (Table 2).

The distribution of EQ-5D scores at baseline, 1-year follow-up, and 5-year follow-up is presented in Multimedia Appendix 1.

During the 1-year follow-up period, 11 patients experienced a new cardiac event (5 in the CG and 6 in the IG). Of these, 6 were cases of recurrent or new CAD (3 in the CG and 3 in the IG), and 4 were cases of AF (2 in the CG and 2 in the IG). Between the 1- and 5-year follow-ups, 21 patients experienced a new cardiac event, including 11 in the CG and 10 in the IG. Of these events, 9 were cases of recurrent or new CAD (8 in the CG and 1 in the IG), 5 were cases of AF (2 in the CG and 3 in the IG), 1 patient in the IG underwent valve surgery, and another survived a cardiac arrest (IG). Additionally, 2 patients in the IG underwent angiography, and 3 patients experienced other heart-related events (1 in the CG and 2 in the IG).

During the first year after CR, 49 patients experienced different kinds of diseases or health problems that affected their ability to adhere to a healthy behavior for 4 weeks or more. This included musculoskeletal disorders (n=34), pulmonary diseases (n=2), diseases in other organs (n=8), complex pain (n=2), and other disorders (n=3). The IG reported statistically significant (*P*<.001) more health problems compared with the CG the first year after CR. From the 1-year follow-up to the 5-year follow-up, 54 patients reported challenges with diseases or health problems, including cancer (n=6), serious COVID-19 (n=2), musculoskeletal disorders (n=30), psychological stress (n=11), and other conditions (n=5). There were no statistically significant differences (*P*=.32) between the IG and CG at a 5-year follow-up regarding diseases or health problems that impacted the ability to adhere to healthy behaviors.

At the 5-year follow-up, the CG reported walking significantly more (MET-minutes/week) than the IG, as assessed by the IPAQ (*P*=.03). There were no statistically significant differences between the groups in any other physical activity domains evaluated by the IPAQ (moderate MET-minutes/week, *P*=.69; vigorous MET-minutes/week, *P*=.20; and total MET-minutes/week, *P*=.64).

XSL•FO
RenderX

Figure 1. The CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

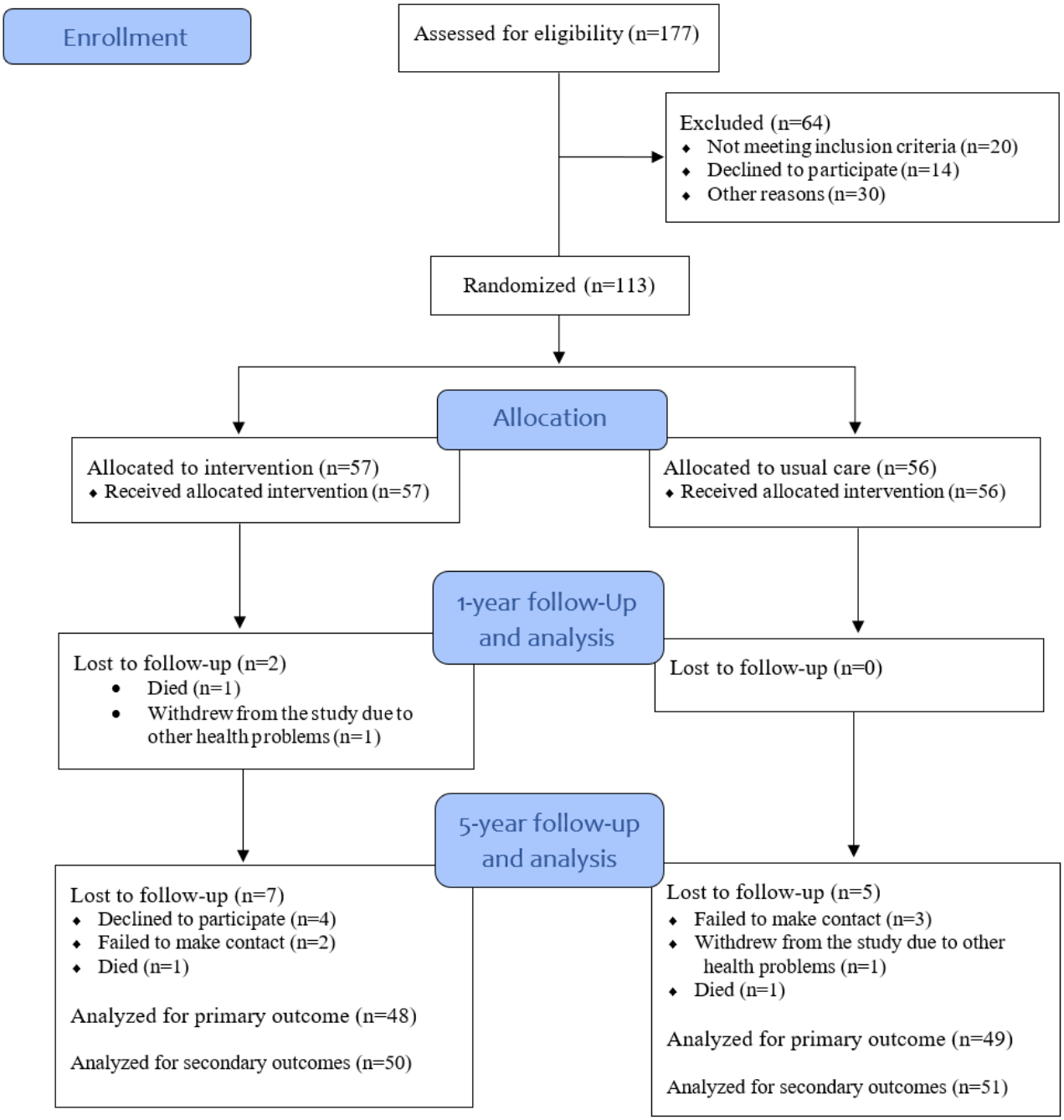

XSL•FO
RenderX

Table 1. Descriptive statistics of the sample at baseline and at 1- and 5-year follow-ups.

Variables	Total (n=113)	CG[a] (n=56)	IG[b] (n=57)	CG at 1 year	IG at 1 year	CG at 5 years	IG at 5 years
Age, mean (SD)	59.0 (8.7)	58.4 (8.2)	59.5 (9.1)	—[c]	—	—	—
Female, n (%)	25 (22.1)	16 (28.6)	9 (15.8)	—	—	—	—
Non-European, n (%)	4 (3.5)	—	4 (7.0)	—	—	—	—
Married or cohabitant, n (%)	89 (78.8)	43 (76.8)	46 (80.7)	—	—	—	—
Current smoker, n (%)	4 (3.5)	1 (1.8)	3 (5.3)	3 (5.4)	2 (3.5)	2 (3.6)	2 (3.5)
Body weight, mean (SD)	90.2 (16.9)	88.5 (17.0)	91.8 (16.8)	88.6 (17.6)	90.4 (16.6)	89.2 (17.6)	90.6 (18.2)
Disease, n (%)							
Acute coronary syndrome	44 (38.9)	22 (39.3)	22 (38.6)	—	—	—	—
Coronary artery disease	39 (34.5)	19 (33.9)	20 (35.1)	—	—	—	—
Valve	19 (16.8)	8 (14.3)	11 (19.3)	—	—	—	—
Cardiac arrest	1 (0.9)	1 (1.8)	—	—	—	—	—
Atrial fibrillation	3 (2.7)	1 (1.8)	2 (3.5)	—	—	—	—
Other	7 (6.2)	5 (8.9)	2 (3.5)	—	—	—	—
Treatment, n (%)							
Percutaneous coronary intervention	55 (48.7)	26 (46.4)	29 (50.9)	—	—	—	—
Coronary artery bypass graft	22 (19.5)	12 (21.4)	10 (17.5)	—	—	—	—
Valve surgery	19 (16.8)	8 (14.3)	11 (19.3)	—	—	—	—
Implantable cardioverter-defibrillator	2 (1.8)	1 (1.8)	1 (1.8)	—	—	—	—
Pacemaker	2 (1.8)	2 (3.6)	—	—	—	—	—
Conservatively	10 (8.8)	6 (10.7)	4 (7.0)	—	—	—	—
Other	3 (2.7)	1 (1.8)	2 (3.5)	—	—	—	—
Medication, n (%)							
Betablocker	69 (61.1)	32 (57.1)	37 (64.9)	—	—	22 (39.3)	23 (40.4)
Statins	96 (85.0)	45 (80.4)	51 (89.5)	—	—	41 (73.2)	43 (75.4)
Acetylsalicylic acid[d] + plate inhibitor	75 (66.4)	39 (69.6)	36 (63.2)	—	—	36 (64.3)[d]	33 (57.9)[d]
Antihypertensive	55 (48.7)	29 (51.8)	26 (45.6)	—	—	32 (57.1)	30 (52.6)
Type of cardiac rehabilitation, n (%)							
1 week	35 (31.0)	17 (30.4)	18 (31.6)	—	—	—	—
4 weeks	40 (35.4)	20 (35.7)	20 (35.1)	—	—	—	—
12 weeks	38 (33.6)	19 (33.9)	19 (33.3)	—	—	—	—
Physical activity (International Physical Activity Questionnair), median (range)							
Walking (MET[e]-minutes/week)	—	—	—	—	—	809 (0-4851)	429 (0-11,946)
Moderate (MET-minutes/week)	—	—	—	—	—	480 (0-13,440)	720 (0-8640)
Vigorous (MET-minutes/week)	—	—	—	—	—	400 (0-11,520)	720 (0-6720)
Total (MET-minutes/week)	—	—	—	—	—	2553 (132-27,600)	2243 (0-11,946)
Cardiopulmonary exercise test , mean (SD)							
Peak oxygen uptake (ml/kg/minute)	29.6 (7.7)	29.9 (6.7)	29.4 (8.7)	28.7 (6.9)	31.0 (8.6)	28.1 (6.7)	27.8 (8.5)
Peak oxygen uptake (l/minute)	2.64 (0.74)	2.63 (0.75)	2.65 (0.74)	2.48 (0.68)	2.71 (0.73)	2.48 (0.72)	2.48 (0.73)
Time to exhaustion (seconds)	601 (124)	603 (102)	598 (144)	632 (120)	671 (130)	592 (117)	579 (134)

XSL•FO
RenderX

Variables	Total (n=113)	CG[a] (n=56)	IG[b] (n=57)	CG at 1 year	IG at 1 year	CG at 5 years	IG at 5 years
Peak incline (%)	10.9 (3.6)	11.0 (3.0)	10.8 (4.2)	11.8 (3.8)	13.0 (4.1)	10.1 (3.9)	9.8 (3.8)
Peak velocity (km/hour)	6.8 (1.6)	6.8 (1.5)	6.8 (1.7)	6.8 (1.5)	7.0 (1.8)	6.7 (1.4)	6.8 (1.5)
Maximal heart rate (beats/minute)	160 (20)	161 (20)	159 (19)	164 (20)	164 (17)	161 (18)	159 (21)
Peak respiratory exchange ratio	1.16 (0.1)	1.16 (0.1)	1.15 (0.11)	1.20 (0.09)	1.20 (0.08)	1.20 (0.07)	1.20 (0.09)
Rate of perceived exertion (Borg Scale, scored from 6 to 20)	17.8 (1.0)	17.8 (1.0)	17.7 (1.1)	17.5 (1.2)	17.8 (1.3)	17.6 (1.4)	17.6 (1.4)
Blood pressure, mean (SD)							
Systolic (mmHg)	135 (17)	136 (18)	133 (17)	145 (20)	143 (19)	135 (18)	133 (18)
Diastolic (mmHg)	81 (9)	81 (8)	81 (10)	84 (11)	86 (10)	80 (12)	80 (12)
Lipid profile							
Low-density lipoprotein-cholesterol (mmol/l)	2.2 (0.9)	2.3 (0.9)	2.2 (0.8)	2.3 (0.9)	2.3 (1.0)	2.3 (1.4)	2.1 (0.8)
High-density lipoprotein-cholesterol (mmol/l)	1.3 (0.4)	1.3 (0.4)	1.2 (0.4)	1.4 (0.5)	1.3 (0.4)	1.4 (0.4)	1.3 (0.4)
Total cholesterol (mmol/l)	4.0 (1.0)	4.0 (1.0)	4.0 (0.9)	3.9 (1.0)	4.1 (1.0)	4.0 (1.5)	3.8 (0.8)
Triglycerides (mmol/l)	1.4 (0.9)	1.3 (1.0)	1.6 (0.9)	1.2 (0.7)	1.6 (1.5)	1.2 (0.7)	1.5 (1.0)
Exercise habits, mean (SD)	1.4 (1.5)	1.3 (1.5)	1.6 (1.5)	1.9 (1.6)	3.0 (1.9)	1.7 (1.7)	2.6 (2.8)
HeartQoL[f], mean (SD)							
Physical	2.5 (0.6)	2.5 (0.5)	2.4 (0.6)	2.6 (0.5)	2.7 (0.5)	2.5 (0.5)	2.5 (0.5)
Emotional	2.4 (0.6)	2.3 (0.7)	2.4 (0.6)	2.3 (0.7)	2.5 (0.7)	2.4 (0.6)	2.4 (0.6)
Global	2.4 (0.5)	2.4 (0.5)	2.4 (0.5)	2.5 (0.5)	2.6(0.5)	2.5 (0.5)	2.5 (0.5)
EQ VAS[g], mean (SD)	71 (16)	72 (14)	69 (18)	75 (12)	78 (16)	74 (13)	70 (19)

[a]CG: control group.

[b]IG: intervention group.

[c]Not applicable.

[d]At the 5-year follow-up, the number and percentage given are for ASA only.

[e]MET: metabolic equivalent.

[f]HeartQoL: disease-specific Health-Related Quality of Life Questionnaire (scored from 0 to 3), which comprises 3 domains (physical, emotional, and global quality of life).

[g]VAS: Visual Analog Scale.

XSL•FO
RenderX

Table 2. Differences between groups at 4 years after the intervention, adjusted for baseline values.

Outcomes	Mean difference[a]	95% CI	*P* value
Cardiopulmonary exercise test (n=97)			
Peak oxygen uptake (ml/kg/minute)	0.29	–1.13 to 1.70	.69
Peak oxygen uptake (l/minute)	–0.0	–0.12 to 0.12	.99
Time to exhaustion (seconds)	–10.7	–46.6 to 25.1	.56
Peak incline (%)	–0.4	–1.6 to 0.7	.46
Peak velocity (km/hour)	0.2	–0.01 to 0.39	.06
Body weight (kg) (n=101)	–1.23	–3.24 to 0.78	.23
Blood pressure (n=97)			
Systolic (mmHg)	–1.5	–8.0 to 5.0	.65
Diastolic (mmHg)	–0.3	–4.6 to 3.9	.89
Lipid profile (n=101)			
Low-density lipoprotein-cholesterol (mmol/l)	0.024	–0.257 to 0.304	.87
High-density lipoprotein-cholesterol (mmol/l)	–0.065	–0.169 to 0.039	.22
Total cholesterol (mmol/l)	–0.039	–0.337 to 0.259	.80
Triglycerides (mmol/l)	0.40	0.00 to 0.79	.048[b]
Exercise habits (n=101)	0.67	0.04 to 1.29	.04[b]
HeartQoL[c] (n=101)			
Physical	0.01	–0.15 to 0.17	.90
Emotional	–0.01	–0.21 to 0.19	.90
Global	0.01	–0.14 to 0.16	.92
EQ Visual Analog Scale[d] (n=101)	–2.49	–7.53 to 2.56	.33

[a]Mean difference is calculated as intervention – control.

[b]*P*<.05 between groups.

[c]HeartQoL: disease-specific Health-Related Quality of Life Questionnaire (scored from 0 to 3), which comprises 3 domains (physical, emotional, and global quality of life).

[d]The overall health status was measured using the Visual Analog Scale (0-100) in the EQ-5D questionnaire.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to evaluate the long-term effects of individualized follow-up using an app during the first year after CR. Our primary finding indicates that the beneficial effects on VO_{2peak} and exercise performance observed 1 year after CR [10] had diminished 4 years after the intervention. For the primary outcome (VO_{2peak}), we maintained adequate statistical power to detect a true difference between the groups at the 5-year follow-up. However, the results for secondary outcomes should be interpreted with caution.

Numerous factors may explain why the effects observed after 1 year of individualized follow-up [10] were no longer present at the 5-year follow-up, and we can only speculate whether these effects might have been sustained if the follow-up had continued. We believe the relationship between the supervisor and the patients in our study played a pivotal role in fostering the successful adherence to healthy behaviors observed 1 year after CR [10]. This was also highlighted by the patients in a qualitative study investigating their experiences with the follow-up provided [30]. The patients emphasized that the supervisor, the person behind the app, was the most significant factor in promoting adherence to healthy behavior [30]. Successful adherence in the context of cardiovascular risk reduction is often attributed to an ongoing, collaborative relationship between the health care provider and the patient [31]. The idea that close monitoring or follow-up can enhance adherence is widely recognized [1,3]; however, only a few studies have demonstrated this in real-world settings. In our study, follow-up was discontinued after 1 year, and the observed effects appeared to diminish over time. The app-based follow-up provided to the IG during the first year after CR in this study was grounded in the Transtheoretical Model, also known as the Stages of Change Model [32]. According to this model, health behavior change involves 6 stages: precontemplation, contemplation, preparation, action, maintenance, and termination. The model emphasizes that behavior change is a process that takes time and is not linear. For example, patients in the maintenance stage may quickly regress to the preparation

XSL•FO
RenderX

stage after CR due to factors such as a new cardiac event or other diseases or health problems that challenge their ability to maintain healthy behaviors. The need for support varies across different stages and should be tailored to increase the likelihood of successful behavior change [32]. Our findings suggest that reaching the termination phase—where individuals are confident they will never revert to their previous unhealthy behaviors and require minimal support [32]—is unlikely to be achieved within a 1-year follow-up period. Whether it would be possible with extended follow-up beyond 1 year remains an open question.

Patients in the IG reported exercising significantly more than those in the CG at the 5-year follow-up. However, there were no statistically significant differences between the groups in variables that reflect the effects of exercise (eg, VO_{2peak} and exercise performance). Notably, the difference in exercise habits between the groups at 5 years was similar to that observed at the 1-year follow-up. A plausible explanation for the lack of differences in VO_{2peak} and exercise performance at the 5-year follow-up, despite the IG reporting higher levels of exercise, could be the lower intensity or shorter duration of exercise sessions after the individualized follow-up was discontinued. As expressed by the patients in the IG [30], the individualized follow-up during the first year after CR provided more structure and focus to their exercise sessions than they were able to achieve on their own. Based on IPAQ data, the CG and IG appear to be equally physically active at the 5-year follow-up, although the CG reported significantly higher levels of walking METs compared with the IG. The IG reported higher levels of physical activity at both moderate and vigorous intensities than the CG, which may correspond to the significant difference observed in exercise habits, although differences observed in moderate and vigorous intensities were not statistically significant. Additionally, there was a statistically significant difference in triglycerides at the 5-year follow-up, favoring the CG. However, the uncertainty regarding the clinical relevance of the observed difference renders the result insignificant, and, on average, both groups remained within normal ranges [33].

For other metabolic risk factors, there were no statistically significant differences between the groups. The mean difference in body weight between the IG and CG at the 5-year follow-up was >1 kg. Although this difference did not reach statistical significance, we believe it is noteworthy, as each kilogram of weight gain has been shown to increase the risk of type 2 diabetes mellitus by 7.3% [34]. This suggests there may be a clinically relevant difference in body weight favoring the IG. However, given a minimal clinically important difference in body weight of 1 kg and an SD of 17 kg (based on our sample), more than 4500 patients would be required in each group to achieve 80% statistical power at a 5% significance level. Another noteworthy finding, although not statistically significant, is the LDL-cholesterol levels. The mean LDL-cholesterol at the 5-year follow-up was 2.3 mmol/l in the CG and 2.1 mmol/l in the IG. Given that the sample includes patients both with and without CAD, these levels are not alarming. When examining LDL-cholesterol specifically for those with CAD or acute coronary syndrome as the referral cause, the mean LDL-cholesterol levels at 5 years were 1.70 (SD 0.62) mmol/l in the CG and 1.88 (SD 0.62) mmol/l in the IG. The elevated LDL-cholesterol levels observed appear to be a common finding in studies evaluating maintenance programs after CR [35-38]. A reasonable explanation for this trend is that these post-CR programs [35-38] primarily focus on exercise rather than specifically addressing metabolic risk factors such as LDL-cholesterol and body weight. When comparing our results regarding LDL-cholesterol with a cross-sectional study by Peersen and colleagues (n=1127) [4], our sample at the 5-year follow-up is slightly better than their sample, with a mean LDL-cholesterol of 2.02 mmol/l 2-36 months (median 16 months) after a cardiac event. However, it is important to note that treatment targets for LDL-cholesterol have changed from <1.8 mmol/l to <1.4 mmol/l [39] since the completion of the intervention in this study. This indicates that there remains a need to optimize lipid management after CR. For the most part, lipid management after CR is handled by patients' general practitioners. For decades, studies on lipid profiles following a cardiac event or CR have shown that few patients meet the guidelines for secondary prevention. This raises the timely question of whether this responsibility should be prioritized more by general practitioners or shifted back to the specialist health service. Consistent with the 1-year follow-up results [10], there were no statistically significant differences between the groups in HRQL at the 5-year follow-up. Nevertheless, it is encouraging that HRQL in our sample remained at a high level. This is particularly important, as reduced HRQL has been associated with an increased incidence of depression and anxiety [40,41], both of which are well-established psychosocial risk factors for poor prognosis in patients with cardiovascular disease [2]. Compared with the reference values for HeartQoL from the EUROASPIRE IV survey (EUROpean Action on Secondary and Primary prevention through Intervention to Reduce Events), our sample demonstrated higher scores in the global domain as well as the physical and emotional HRQL domains at both follow-up points [18].

Strengths and Limitations

A key strength of our study is the low dropout rate at both the 1- and 5-year follow-ups. Additionally, patients were recruited from 3 different CR programs, and the use of broad inclusion criteria further strengthens the generalizability of our findings. Our goal was to make the results applicable to the wider population of cardiac patients participating in CR. Given that CR programs vary widely both across and within countries, we included patients from 3 distinct CR programs delivered in different settings (residential and outpatient) and with varying durations (1, 4, and 12 weeks). Because of the uncertainty regarding whether adherence to healthy behaviors after CR is influenced by the type of CR program, randomization was stratified based on the specific CR program. By including all medically stable patients with cardiac diseases, we were able to capture a representative sample of a typical CR population. However, this broad inclusion criterion also posed challenges in reporting some outcomes, particularly metabolic risk factors. For example, patients treated for valve disease may not necessarily be advised to reduce LDL-cholesterol levels or body weight. To address this, descriptive statistics on subgroups and the clinical relevance of these outcomes have been provided. We believe that different subgroups of patients completing CR

XSL•FO
RenderX

may derive varying benefits from the intervention. To explore this further, future studies could consider narrowing the inclusion criteria or significantly increasing the sample size to enable systematic subgroup analyses.

Conclusion and Implications

This study demonstrates that the beneficial and clinically important effects of a 1-year individualized follow-up via an app post-CR [10] had ceased by 4 years after the intervention. Despite the growing emphasis on addressing challenges related to adherence to healthy behavior and the use of technology in international research strategies [6,42,43], as well as in key public policies, we believe that significant work remains before mHealth can be fully implemented in this context. However, our findings on feasibility [28], 1-year follow-up outcomes [10], patients' experiences [30], and the results presented in this study can provide valuable guidance for developing an implementation strategy for mHealth interventions aimed at promoting adherence to healthy behaviors in the post-CR setting. Before developing such a strategy, it is essential to conduct a comprehensive investigation into the potential barriers and facilitators associated with its implementation, involving all relevant stakeholders. To fully understand the multifaceted challenges of adherence to healthy behavior after CR, this investigation must include input from policy makers, CR program managers, health care professionals, and, most importantly, patients—both those struggling with adherence and those who have successfully maintained adherence to secondary prevention recommendations over the years.

Acknowledgments

This project was made possible through the support of the Dam Foundation. We extend our heartfelt gratitude to the Norwegian Institute of Sports Medicine (NIMI), Volvat, and Lovisenberg Rehabilitation AS (formerly the National Association for Pulmonary and Cardiac Disease) for allowing us to recruit patients from their CR centers. We also thank M4Influence for providing access to their software solutions during the intervention period. This project is funded by Foundation Dam (grant 326568).

Data Availability

The data sets generated and analyzed during this study are not publicly available due to the consent provided by the participants. However, they may be made available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution in percentage of reporting levels 1-5 by dimension, group, and measurement time point.
[DOCX File , 26 KB - jmir_v27i1e60256_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1033 KB - jmir_v27i1e60256_app2.pdf]

References

1. Ambrosetti M, Abreu A, Corrà U, Davos CH, Hansen D, Frederix I, et al. Secondary prevention through comprehensive cardiovascular rehabilitation: from knowledge to implementation. 2020 update. A position paper from the Secondary Prevention and Rehabilitation Section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2021 May 14;28(5):460-495. [doi: 10.1177/2047487320913379] [Medline: 33611446]
2. Authors/Task Force Members:, Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, et al. Atherosclerosis 2016 Sep;252:207-274. [doi: 10.1016/j.atherosclerosis.2016.05.037] [Medline: 27664503]
3. Pedretti RFE, Hansen D, Ambrosetti M, Back M, Berger T, Ferreira MC, et al. How to optimize the adherence to a guideline-directed medical therapy in the secondary prevention of cardiovascular diseases: a clinical consensus statement from the European Association of Preventive Cardiology. Eur J Prev Cardiol 2023 Jan 24;30(2):149-166 [FREE Full text] [doi: 10.1093/eurjpc/zwac204] [Medline: 36098041]
4. Peersen K, Munkhaugen J, Gullestad L, Liodden T, Moum T, Dammen T, et al. The role of cardiac rehabilitation in secondary prevention after coronary events. Eur J Prev Cardiol 2017 Sep;24(13):1360-1368. [doi: 10.1177/2047487317719355] [Medline: 28664773]
5. Kotseva K, Wood D, De Bacquer D, De Backer G, Rydén L, Jennings C, EUROASPIRE Investigators. EUROASPIRE IV: A European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. Eur J Prev Cardiol 2016 Apr;23(6):636-648 [FREE Full text] [doi: 10.1177/2047487315569401] [Medline: 25687109]

XSL•FO
RenderX

6. World Health Organization (WHO). mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2011:6.
7. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
8. Duscha BD, Piner LW, Patel MP, Craig KP, Brady M, McGarrah RW, et al. Effects of a 12-week mHealth program on peak VO and physical activity patterns after completing cardiac rehabilitation: a randomized controlled trial. Am Heart J 2018 May;199:105-114. [doi: 10.1016/j.ahj.2018.02.001] [Medline: 29754647]
9. Snoek JA, Meindersma EP, Prins LF, Van't Hof AW, de Boer M, Hopman MT, et al. The sustained effects of extending cardiac rehabilitation with a six-month telemonitoring and telecoaching programme on fitness, quality of life, cardiovascular risk factors and care utilisation in CAD patients: the TeleCaRe study. J Telemed Telecare 2021 Sep;27(8):473-483. [doi: 10.1177/1357633X19885793] [Medline: 31760855]
10. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2020 Nov;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]
11. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
12. Nilsson BB, Lunde P, Holm I. Implementation and evaluation of the Norwegian Ullevaal model as a cardiac rehabilitation model in primary care. Disabil Rehabil 2019 Feb;41(4):481-488. [doi: 10.1080/09638288.2017.1397776] [Medline: 29105515]
13. Bergum H, Sandven I, Abdelnoor M, Anderssen SA, Grimsmo J, Rivrud DE, et al. Randomized trial of cardiovascular prevention in Norway combining an in-hospital lifestyle course with primary care follow-up: the Hjerteløftet study. Eur J Prev Cardiol 2022 Dec 07;29(17):2252-2263. [doi: 10.1093/eurjpc/zwac211] [Medline: 36124709]
14. Lunde P, Bye A, Bergland A, Nilsson BB. Effects of individualized follow-up with a smartphone-application after cardiac rehabilitation: protocol of a randomized controlled trial. BMC Sports Sci Med Rehabil 2019;11:34 [FREE Full text] [doi: 10.1186/s13102-019-0148-2] [Medline: 31768261]
15. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
16. Oldridge N, Höfer S, McGee H, Conroy R, Doyle F, Saner H, (for the HeartQoL Project Investigators). The HeartQoL: Part I. Development of a new core health-related quality of life questionnaire for patients with ischemic heart disease. Eur J Prev Cardiol 2014 Jan;21(1):90-97. [doi: 10.1177/2047487312450544] [Medline: 22822179]
17. Oldridge N, Höfer S, McGee H, Conroy R, Doyle F, Saner H, et al. The HeartQoL: part II. Validation of a new core health-related quality of life questionnaire for patients with ischemic heart disease. Eur J Prev Cardiol 2014 Jan;21(1):98-106. [doi: 10.1177/2047487312450545] [Medline: 22822180]
18. De Smedt D, Clays E, Höfer S, Oldridge N, Kotseva K, Maggioni AP, et al. The use of HeartQoL in patients with coronary heart disease: association with risk factors and European reference values. The EUROASPIRE IV study of the European Society of Cardiology. Eur J Prev Cardiol 2016 Jul;23(11):1174-1186 [FREE Full text] [doi: 10.1177/2047487316631400] [Medline: 26862122]
19. Kristensen MS, Zwisler A, Berg SK, Zangger G, Grønset CN, Risom SS, et al. Validating the HeartQoL questionnaire in patients with atrial fibrillation. Eur J Prev Cardiol 2016 Dec;23(14):1496-1503. [doi: 10.1177/2047487316638485] [Medline: 26976845]
20. Zangger G, Zwisler A, Kikkenborg Berg S, Kristensen MS, Grønset CN, Uddin J, et al. Psychometric properties of HeartQoL, a core heart disease-specific health-related quality of life questionnaire, in Danish implantable cardioverter defibrillator recipients. Eur J Prev Cardiol 2018 Jan;25(2):142-149. [doi: 10.1177/2047487317733074] [Medline: 28952795]
21. Grønset CN, Thygesen LC, Berg SK, Zangger G, Kristensen MS, Sibilitz KL, et al. Measuring HRQoL following heart valve surgery: the HeartQoL questionnaire is a valid and reliable core heart disease instrument. Qual Life Res 2019 Jan 04:1245-1253. [doi: 10.1007/s11136-018-02098-1] [Medline: 30610503]
22. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
23. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
24. Kurtze N, Rangul V, Hustvedt B. Reliability and validity of the international physical activity questionnaire in the Nord-Trøndelag health study (HUNT) population of men. BMC Med Res Methodol 2008 Oct 09;8:63 [FREE Full text] [doi: 10.1186/1471-2288-8-63] [Medline: 18844976]
25. Gulati M, Pandey DK, Arnsdorf MF, Lauderdale DS, Thisted RA, Wicklund RH, et al. Exercise capacity and the risk of death in women: the St James Women Take Heart Project. Circulation 2003 Sep 30;108(13):1554-1559 [FREE Full text] [doi: 10.1161/01.CIR.0000091080.57509.E9] [Medline: 12975254]

XSL•FO
RenderX

26. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med 2002 Mar 14;346(11):793-801. [doi: 10.1056/NEJMoa011858] [Medline: 11893790]
27. Myers J, McAuley P, Lavie CJ, Despres J, Arena R, Kokkinos P. Physical activity and cardiorespiratory fitness as major markers of cardiovascular risk: their independent and interwoven importance to health status. Prog Cardiovasc Dis 2015;57(4):306-314. [doi: 10.1016/j.pcad.2014.09.011] [Medline: 25269064]
28. Lunde P, Nilsson BB, Bergland A, Bye A. Feasibility of a mobile phone app to promote adherence to a heart-healthy lifestyle: single-arm study. JMIR Form Res 2019 Apr 19;3(2):e12679 [FREE Full text] [doi: 10.2196/12679] [Medline: 31066684]
29. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: from design to analysis. Yale J Biol Med 2013 Sep;86(3):343-358 [FREE Full text] [Medline: 24058309]
30. Lunde P, Bye A, Bruusgaard KA, Hellem E, Nilsson BB. Patients' experiences of using a smartphone app after cardiac rehabilitation: qualitative study. JMIR Hum Factors 2022 Mar 23;9(1):e34294 [FREE Full text] [doi: 10.2196/34294] [Medline: 35319476]
31. Cohen SM. Concept analysis of adherence in the context of cardiovascular risk reduction. Nurs Forum 2009;44(1):25-36. [doi: 10.1111/j.1744-6198.2009.00124.x] [Medline: 19187051]
32. Prochaska J, Redding A, Evers K. The transtheoretical model and stages of change. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. San Francisco: Jossey-Bass; 2015:125-148.
33. Yang G, Mason AM, Wood AM, Schooling CM, Burgess S. Dose-response associations of lipid traits with coronary artery disease and mortality. JAMA Netw Open 2024 Jan 02;7(1):e2352572 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.52572] [Medline: 38241044]
34. Koh-Banerjee P, Wang Y, Hu FB, Spiegelman D, Willett WC, Rimm EB. Changes in body weight and body fat distribution as risk factors for clinical diabetes in US men. Am J Epidemiol 2004 Jun 15;159(12):1150-1159 [FREE Full text] [doi: 10.1093/aje/kwh167] [Medline: 15191932]
35. Avila A, Claes J, Buys R, Azzawi M, Vanhees L, Cornelissen V. Home-based exercise with telemonitoring guidance in patients with coronary artery disease: does it improve long-term physical fitness? Eur J Prev Cardiol 2020 Mar;27(4):367-377. [doi: 10.1177/2047487319892201] [Medline: 31787026]
36. Janssen V, De Gucht V, van Exel H, Maes S. Beyond resolutions? A randomized controlled trial of a self-regulation lifestyle programme for post-cardiac rehabilitation patients. Eur J Prev Cardiol 2013 Jun;20(3):431-441. [doi: 10.1177/2047487312441728] [Medline: 22396248]
37. Skobel E, Knackstedt C, Martinez-Romero A, Salvi D, Vera-Munoz C, Napp A, et al. Internet-based training of coronary artery patients: the Heart Cycle Trial. Heart Vessels 2017 Apr;32(4):408-418. [doi: 10.1007/s00380-016-0897-8] [Medline: 27730298]
38. Madssen E, Arbo I, Granøien I, Walderhaug L, Moholdt T. Peak oxygen uptake after cardiac rehabilitation: a randomized controlled trial of a 12-month maintenance program versus usual care. PLoS One 2014;9(9):e107924 [FREE Full text] [doi: 10.1371/journal.pone.0107924] [Medline: 25247991]
39. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, ESC Scientific Document Group. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. Eur Heart J 2020 Jan 01;41(1):111-188 [FREE Full text] [doi: 10.1093/eurheartj/ehz455] [Medline: 31504418]
40. Mayou RA, Gill D, Thompson DR, Day A, Hicks N, Volmink J, et al. Depression and anxiety as predictors of outcome after myocardial infarction. Psychosom Med 2000;62(2):212-219. [doi: 10.1097/00006842-200003000-00011] [Medline: 10772399]
41. Dickens C, Cherrington A, McGowan L. Depression and health-related quality of life in people with coronary heart disease: a systematic review. Eur J Cardiovasc Nurs 2012 Sep;11(3):265-275. [doi: 10.1177/1474515111430928] [Medline: 22457381]
42. World Health Organization (WHO). Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
43. World Health Organization. WHO Global NCD Action Plan 2013-2020. Geneva, Switzerland: World Health Organization; 2013.

Abbreviations

CAD: coronary artery disease
CG: control group
CONSORT: Consolidated Standards of Reporting Trials
CPET: cardiopulmonary exercise test
CR: cardiac rehabilitation
HDL: high-density lipoprotein
HRQL: health-related quality of life
IG: intervention group
IPAQ: International Physical Activity Questionnaire

XSL•FO
RenderX

LDL: low-density lipoprotein
mHealth: mobile health
VAS: Visual Analog Scale
VO2max: maximal oxygen consumption
VO_{2peak}: peak oxygen uptake

Edited by A Mavragani; submitted 06.05.24; peer-reviewed by D Laroche, S Zaki; comments to author 13.09.24; revised version received 17.10.24; accepted 18.12.24; published 13.02.25.

Please cite as:
Lunde P, Bye A, Grimsmo J, Pripp AH, Ritschel V, Jarstad E, Nilsson BB
Effects of Individualized Follow-Up With an App Postcardiac Rehabilitation: Five-Year Follow-Up of a Randomized Controlled Trial
J Med Internet Res 2025;27:e60256
URL: https://www.jmir.org/2025/1/e60256
doi:10.2196/60256
PMID:

©Pernille Lunde, Asta Bye, Jostein Grimsmo, Are Hugo Pripp, Vibeke Ritschel, Even Jarstad, Birgitta Blakstad Nilsson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Clinical Efficacy of an Exergame-Based Training Program for Enhancing Physical and Cognitive Functions in Older Adults With Mild Cognitive Impairment and Dementia Residing in Rural Long-Term Care Facilities: Randomized Controlled Trial

Aoyu Li[1], PhD; Wei Qiang[2], PhD; Jingwen Li[3], BD; Yan Geng[4], MD; Yan Qiang[5], PhD; Juanjuan Zhao[1], PhD

[1]School of Software, Taiyuan University of Technology, Jingzhong, China

[2]Institute of Software, Chinese Academy of Science, Beijing, China

[3]School of Computer Science, Xijing University, Xi'an, China

[4]Clinical Laboratory and Geriatric, Shanxi Provincial People's Hospital, Taiyuan, China

[5]School of Software, North University of China, Taiyuan, China

Corresponding Author:
Juanjuan Zhao, PhD
School of Software
Taiyuan University of Technology
No. 319, University Street, Yuji District
Jingzhong, 030600
China
Phone: 86 18636664123
Email: zhaojuanjuan@tyut.edu.cn

Abstract

Background: Cognitive impairment is an important public health challenge among older adults, particularly in long-term care facilities (LTCFs), where prevalence is higher due to staffing shortages, limited resources, and difficulty maintaining structured exercise programs. Furthermore, older adults often lose interest in repetitive interventions. The exergame "WarioWare: Move It!" (Nintendo) offers a novel solution by combining aerobic exercise, motor coordination, balance training, and cognitive engagement into an immersive experience.

Objective: This study aimed to assess the clinical efficacy of an exergame-based training program delivered via "WarioWare: Move It!" in improving physical flexibility, joint range of motion, motor coordination, hand dexterity, and cognitive function in older adults living in LTCFs.

Methods: The training program was conducted across multiple rural LTCFs in Shanxi Province, China. Participants were randomly assigned to the intervention or control group. The intervention protocol encompassed two 60-minute sessions per week over 12 weeks, using motion-sensing exercises such as waving, jumping, arm swinging, rotational movements, and object-mimicking postures with Joy-Con controllers. Primary outcome measures were derived through clinical tests, including the sit and reach test, shoulder flexibility test, trunk rotation flexibility test, shoulder and elbow range of motion, figure-of-8 walk test, standing balance test, hand dexterity test, and cognitive function tests. Statistical analysis was performed using mixed ANOVA, with time as the within-participant factor and intervention group as the between-participant factor, to assess the training effects on the various outcome measures.

Results: A total of 232 participants were recruited, including 32 (13.8%) patients with mild dementia, 18 (7.8%) with moderate dementia, and 182 (78.4%) with mild cognitive impairment, all of whom completed the study. The mixed ANOVA revealed significant group × time interactions across multiple physical flexibility assessments, including the remaining distance between the hands and toes during the forward bend ($F_{2,156}$=8.484; P<.001; η^2=0.098), the distance between the hands clasped behind the back ($F_{2,156}$=3.666; P=.04; η^2=0.045), and the angle formed by trunk rotation to the left and right ($F_{2,156}$=17.353; P<.001; η^2=0.182). Significant group × time interactions also emerged for shoulder joint forward flexion ($F_{2,156}$=17.655; P<.001; η^2=0.185), abduction ($F_{2,156}$=6.281; P=.004; η^2=0.075), and elbow flexion ($F_{2,156}$=3.298; P=.049; η^2=0.041). In addition, the time to complete the figure-of-8 walk test ($F_{2,156}$=11.846; P<.001; η^2=0.132) and the number of blocks moved within 1 minute ($F_{2,156}$=4.016;

XSL•FO
RenderX

P=.02; η^2=0.049) showed significant interactions. Finally, all scale-based measures exhibited statistically significant group × time interactions (all P values <.001).

Conclusions: The "WarioWare: Move It!" intervention significantly improved physical flexibility, joint range of motion, motor coordination, hand dexterity, and cognitive function among older adults with mild cognitive impairment or dementia residing in rural LTCFs. The intervention offers an innovative and feasible approach for promoting the health of older adults in resource-limited settings, demonstrating its potential for widespread application in diverse low-resource environments.

Trial Registration: ClinicalTrials.gov NCT06717971; https://clinicaltrials.gov/study/NCT06717971

(J Med Internet Res 2025;27:e69109) doi:10.2196/69109

KEYWORDS

exergame; mild cognitive impairment; dementia; long-term care facilities; multicomponent training

Introduction

Background

Mild cognitive impairment (MCI) and dementia exhibit high prevalence rates among the global older adult population, with new cases projected to increase due to population aging [1,2]. MCI is considered an intermediary stage between age-related cognitive decline and dementia, with studies indicating that approximately 5% to 13% of individuals with MCI progress to dementia each year [3]. The progression is influenced by multiple risk factors, including age, family history, and lifestyle choices, underscoring the critical importance of early intervention and preventive strategies in reducing the incidence of dementia. Dementia, a neurodegenerative disorder with multifactorial etiology, is characterized by the progressive degeneration of brain neurons, resulting in cognitive deficits across several domains, such as complex attention, memory, learning ability, perceptual motor skills, social cognition, and language ability [4]. Due to its multifactorial nature, no effective cure for dementia currently exists, which not only leads to a significant decline in patients' quality of life but also imposes a substantial economic and social burden on families, caregivers, and health care systems [5]. Consequently, early prevention of dementia and maintenance of cognitive function are of particular importance in the older adult population.

In recent years, exergaming, an innovative intervention that integrates physical activity with gaming, has increasingly been introduced into training and rehabilitation programs for the older adult population [6,7]. By combining interactive experiences with immersive virtual environments, exergames encourage active participation among older adults, enhancing cognitive function while alleviating the monotony of traditional exercise forms. Studies indicate that many commercial gaming systems (such as Wii and Kinect) and custom-developed virtual reality games significantly stimulate cognitive abilities, especially in improving memory, attention, and executive function [8,9].

Globally, exergames have garnered significant attention as a promising approach to promoting healthy aging and addressing age-related cognitive and physical decline. Studies conducted across various countries have explored the potential of exergames to combat cognitive impairments, reporting improvements in cognitive and physical health as well as increased enjoyment and motivation among older adults in North America, Europe, and Asia [10-12]. For instance, Kannan et al [13] focused on integrating exergames into community programs and long-term care facilities (LTCFs), demonstrating their effectiveness in enhancing physical functions such as dynamic balance, gait performance, and muscle strength among community-dwelling older adults with reduced physical capacity. Similarly, Chan et al [14] highlighted the role of gamified interventions in fostering social interaction and alleviating feelings of loneliness among older adults, underscoring their cultural adaptability across different social contexts. In Asia, where aging populations are rapidly growing, researchers have developed culturally tailored exergame systems that incorporate traditional activities and movements to enhance their relevance and adherence for older users [15,16].

Moreover, exergames address common barriers to physical activity, such as adverse weather conditions, limited mobility, and geographic isolation, making them particularly valuable in regions with resource constraints or limited access to recreational facilities. These challenges are especially pronounced in rural LTCFs, where cultural attitudes toward exercise often regard physical activity as unnecessary for older adults, and social norms prioritize labor-intensive activities over structured exercise programs. In addition, logistical barriers such as transportation challenges, linguistic diversity, and lack of familiarity with digital technologies hinder the implementation of innovative interventions such as exergames [17,18]. Coupled with the higher prevalence of social isolation in rural settings, these factors necessitate tailored, accessible, and engaging solutions to promote physical and cognitive health among older adults. As global efforts continue to expand the application of exergames in diverse settings, they demonstrate considerable potential for improving the quality of life among older adults, regardless of cultural or socioeconomic contexts.

Despite the promising outcomes of exergaming for older urban populations, its adoption in rural LTCFs remains challenging due to limited health care and social resources [19]. In these settings, shortages of trained caregivers and limited equipment hinder the effective implementation of structured exercise programs, limiting older adults' access to sustained, systematic rehabilitation. In addition, older adults often experience boredom with repetitive exercise routines. They may lack intrinsic motivation, especially when deprived of novel experiences, which can diminish interest and compromise the long-term effectiveness of training [20]. Therefore, exploring and evaluating exergaming systems that are feasible for rural LTCFs

XSL•FO
RenderX

and capable of fostering high engagement could serve as a viable approach to enhancing cognitive and physical health among older adults in these communities.

Objectives

Despite the demonstrated clinical benefits of exergames for community-dwelling older adults [6,21], their effectiveness within LTCFs remains insufficiently explored. Current research primarily focuses on improvements in health-related quality of life, cognitive function, and overall functional status [12,22,23], with relatively few assessments targeting specific physical functions, such as flexibility, joint mobility, motor coordination, and hand dexterity. However, these physical capabilities are crucial for the independence of older adults. For instance, limited flexibility can hinder daily activities, directly impacting self-care abilities and increasing dependency on others [24,25]. A reduction in joint mobility is often associated with joint stiffness and pain, restricting the range of motion and diminishing overall physical function, and limiting the living space and activity radius of older adults [26]. Motor coordination is critical in balance maintenance and complex movements, where diminished coordination heightens the risk of falls [27]. Falls can have severe consequences for older adults, including fractures, prolonged immobility, or even accelerated frailty, substantially reducing the quality of life [28-30]. Hand dexterity also underpins the self-sufficiency of older adults. Loss of fine motor skills, such as gripping, pinching, and writing, can restrict their ability to perform daily tasks and participate socially, potentially leading to psychological distress [31].

The Nintendo exergame “WarioWare: Move It!” integrates 12 distinct movement modes, offering not only aerobic, coordination, and balance training but also cognitive exercises that engage memory, attention, and logical reasoning. The game’s difficulty can be flexibly adjusted to match users’ needs, ensuring optimal training effects at appropriate challenge levels. Despite its evident potential, comprehensive empirical studies are still required to substantiate its benefits fully. Therefore, this study conducted a 12-week longitudinal user study to examine the feasibility and potential clinical utility of “WarioWare: Move It!” for older adult populations in rural LTCFs. Furthermore, this research provides valuable insights and references for older adult rehabilitation professionals and human-computer interaction researchers interested in designing exergame-based interventions for the care of older adults.

Methods

Ethical Considerations

This study was conducted in accordance with the Helsinki Declaration and received approval from the Biomedical Ethics Review Committee of Taiyuan University of Technology (20230641). The study design adheres to the 2013 SPIRIT statement for randomized controlled trials and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Multimedia Appendix 1), and it has been registered on ClinicalTrials.gov (NCT06717971). Before submitting the written informed consent, participants were informed of potential risks, including minor muscle discomfort or fatigue. They were also made aware that participation was voluntary and that they could withdraw at any time without penalty. All participants provided written informed consent, and eligible participants received a US $10 compensation.

Participants

In this study, we recruited 232 participants (n=32, 13.8% patients with mild dementia; n=18, 7.8% with moderate dementia; and n=182, 78.4% with MCI) from LTCFs, such as rural daycare centers and nursing homes in Shanxi Province, China. Participant selection followed a purposive sampling method [32], with the entire recruitment process rigorously supervised by experienced neurologists. Caregivers and facility staff were not involved in the participant selection process to avoid potential bias. Inclusion criteria were as follows: (1) normal or corrected-to-normal hearing and vision; (2) age >65 years; (3) residence in an older adult care facility or a minimum stay of 1 month in such a facility; (4) completion of the Global Deterioration Scale (GDS), the Chinese version of the Mini-Mental State Examination (MMSE), and the Chinese version of the Montreal Cognitive Assessment (MoCA), with the capacity to communicate effectively; (5) ability to engage in moderate physical activity without physical disability; (6) absence of severe depressive symptoms or other neurological disorders (eg, stroke, dizziness, and epilepsy); (7) absence of chronic conditions significantly affecting mobility (eg, advanced arthritis, severe osteoporosis, or recent fractures); and (8) provision of informed consent by the participant or their guardian.

All participants underwent comprehensive medical evaluations conducted by experienced neurologists, including a thorough medical history, physical examination, neuroimaging studies (either brain magnetic resonance imaging or computed tomography), and cognitive function assessments. Neurologists assessed participants’ cognitive function using the GDS score, classifying the severity of dementia symptoms accordingly. Among the participants with dementia, the average GDS score was 3.7 (SD 0.75), with an average duration of dementia symptoms spanning 3.15 (SD 0.8) years. Cognitive function in the 182 patients with MCI was evaluated using the MoCA, with scores consistently <26. Furthermore, neuroimaging revealed no other structural abnormalities in any participant that could contribute to cognitive impairment. Table 1 summarizes the clinical and demographic characteristics of participants allocated to the intervention and control groups.

XSL•FO
RenderX

Table 1. Clinical and demographic characteristics of participants randomly assigned to the intervention and control groups (N=232).

Characteristics	Intervention group (n=116)	Control group (n=116)	*P* value
Age (years), mean (SD)	73.03 (4.431)	72.70 (4.958)	.59
Gender, n (%)			.43
Woman	68 (58.6)	62 (53.4)	
Man	48 (41.4)	54 (46.6)	
Hand preference, n (%)			.68
Left	12 (10.3)	14 (12.1)	
Right	104 (89.7)	102 (87.9)	
Exercise habit, n (%)			.43
Yes	62 (53.4)	60 (51.7)	
No	54 (46.6)	56 (48.3)	
Years of education, mean (SD)	6.53 (3.037)	6.31 (2.994)	.57
Hours of sleep, mean (SD)	6.44 (1.239)	6.42 (1.259)	.92
Years of smart device use, mean (SD)	4.59 (1.658)	4.35 (1.741)	.28

Randomization and Blinding

To eliminate selection bias, enhance the internal validity of the research findings, and ensure scientific rigor and reliability, this study used a stringent randomized grouping method to assign participants to either the intervention or control group. The randomization process followed the following strict steps:

- All eligible participants were assigned a unique identification number to maintain anonymity.
- Identification numbers were recorded on slips of paper, which were then sealed in opaque envelopes to ensure a double-anonymized grouping process.
- The numbered envelopes were placed in an opaque container and thoroughly mixed to guarantee complete randomness in selection.
- Researchers wore blindfolds to prevent prior knowledge of envelope contents and randomly drew a single envelope.
- The selection process was repeated until 116 envelopes were chosen to form the intervention group, with the remaining participants allocated to the control group.

After random assignment at a 1:1 ratio, the intervention group received a 12-week training regimen, while the control group continued with standard care provided by LTCFs. Both groups maintained equivalent activity durations.

Intervention Group Training Program

The training program for the intervention group used the Nintendo Switch exergame “WarioWare: Move It,” as depicted in Figure 1. Participants engaged with the game through Joy-Con controllers, completing rapid microgames designed to meet specific movement-oriented objectives. Each game session lasted approximately 5 seconds, requiring participants to quickly adapt to various physical actions and postures, such as waving, jumping, swinging arms, rotating, and mimicking object movements. The game’s interactive design fosters an engaging experience that promotes physical coordination, reaction speed, upper and lower limb strength, and hand-eye coordination through diverse movement patterns. Switching between tasks compels players to swiftly adjust to varying motion demands, enhancing physical agility while strengthening attention shifting and quick decision-making skills.

XSL•FO
RenderX

Figure 1. Training scenarios and gameplay modes of the Nintendo Switch "WarioWare: Move It" exergame: (A) a participant engaging in a gameplay scenario, (B) competitive mode, (C) cooperative mode, and (D) versus mode.

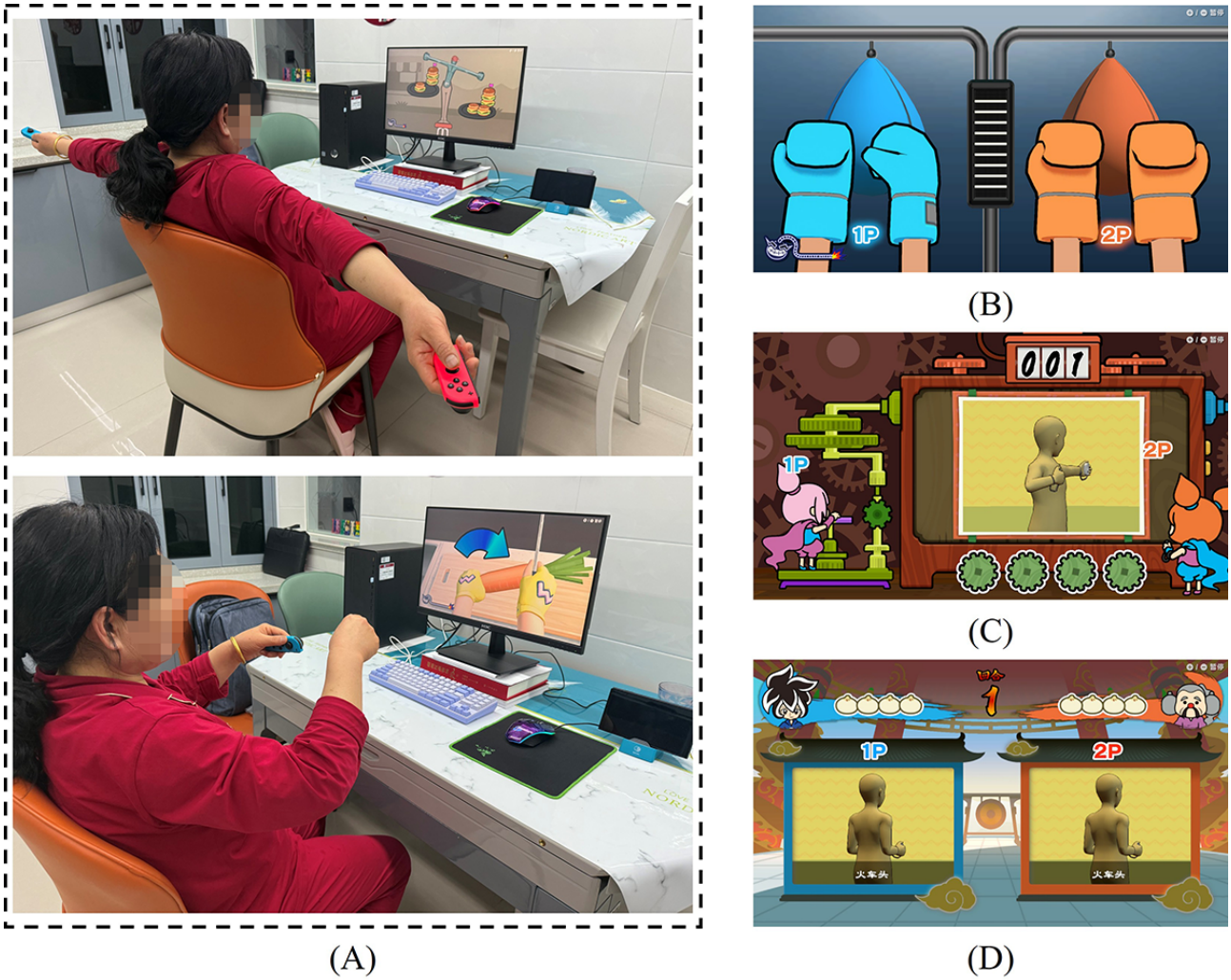

Furthermore, to enhance both engagement and safety, the intervention program was supported by trained staff responsible for providing tailored assistance to participants throughout the training. Before the intervention, all staff members received comprehensive training in using both the Nintendo Switch console and the game mechanics of "WarioWare: Move It!." The training included familiarization with the game's functionality, understanding how to adapt the tasks to meet participants' physical and cognitive needs, and recognizing when to modify the intensity or duration of tasks. Staff were also trained to monitor participants' physical and cognitive responses, offer guidance as needed, and adjust the game's parameters based on individual performance.

The training program was administered twice weekly, with each session lasting 60 minutes over 12 weeks. Each session comprised three 15-minute gaming segments interspersed with 1- to 5-minute rest periods to sustain participant engagement and mitigate mental fatigue. Before the intervention, all participants attended a 30-minute orientation session to familiarize themselves with the Nintendo Switch console, the game mechanics of "WarioWare: Move It!," and the use of Joy-Con controllers. During training, staff provided tailored support and dynamically adjusted program parameters based on individual performance, modifying task content, intensity, duration, and physical demands as necessary.

Control Group Standard Care Program

The standard care program for the control group is grounded in routine practices of LTCFs, aiming to sustain and enhance the physical and mental well-being of older adult residents through structured group activities. These activities encompass tailored fitness exercises, horticultural therapy, and sedentary group engagements for older adults. The fitness exercises are adapted to meet the varying physical conditions of participants, ensuring a safe and comfortable environment that promotes moderate physical activity. Horticultural therapy involves hands-on planting and tending to plants, stimulating the senses, and fostering emotional regulation, contributing to improved psychological health. Sedentary activities, such as tabletop games, emphasize social interaction and cognitive maintenance by promoting a relaxed social environment that helps participants remain cognitively engaged. A professional therapist leads each session to ensure that participants receive adequate guidance and support throughout the activities. The program is conducted twice a week, with each session lasting 60 minutes over 12 weeks.

XSL•FO
RenderX

Outcomes Measured

All participants underwent 3 assessments at distinct time points. The initial assessment occurred at baseline before randomization. Subsequent evaluations were conducted at the end of 12 weeks (posttest) and 6 months after the intervention concluded. The primary outcomes of this study focus on detecting intervention effects using indicators related to physical flexibility, joint range of motion, motor coordination, hand dexterity, and cognitive abilities.

In this study, physical flexibility assessments involve several standardized tests. First, the sit and reach test measures the participant's ability to stretch forward while seated with legs extended and heels fixed [33,34]. The farthest distance reached with outstretched hands toward the toes is recorded, with the gap between the fingertips and toes serving as the flexibility index. The shoulder flexibility test assesses flexibility by measuring the distance between the hands when clasped behind the back [35,36]. Participants with higher flexibility achieve either easy hand clasping or shorter hand-to-hand distance. The trunk rotation flexibility test requires participants to stand with feet together and arms extended forward, rotating the torso left and right around the waist [37]. The maximum angle reached on each side indicates flexibility, with more prominent angles signifying greater flexibility.

The joint range of motion reflects the extent of freedom a joint possesses within a specified range of movement. Observing changes in the range of motion across different directions in the intervention group makes it possible to determine whether significant improvements occur in particular directions (eg, flexion or internal rotation), thus providing targeted feedback for the intervention. This study's joint range of motion assessments include shoulder and elbow range of motion tests. The shoulder range of motion test measures the maximum range in 4 directions (eg, flexion, extension, abduction, and adduction) [38,39]. In contrast, the elbow range of motion test assesses the degree of elbow flexion [40].

In addition, this study evaluates motor coordination, defined as the body's ability to synchronize various parts during complex physical tasks. The assessment methods include the figure-of-8 walk [41,42] and standing balance test [43,44]. In the figure-of-8 walk test, participants are instructed to walk in a figure-of-8 pattern around 2 markers placed on the ground, covering a total distance of 10 m. The time taken to complete this task is recorded as a measure of coordination. The standing balance test requires participants to stand with their feet together and their eyes closed, aiming to maintain balance, with the duration of balance maintained being measured as an indicator of stability.

Similarly, hand dexterity reflects an individual's coordination, agility, and reaction speed in performing fine motor tasks, particularly when required to complete repetitive, high-precision tasks within a limited time frame. Hand dexterity was assessed using the box block test [45,46], where participants were instructed to use their dominant hand to transfer blocks from one compartment to another within a wooden box over 60 seconds, recording the total number of transfers completed within the specified time. Finally, cognitive function was evaluated using the Cognitive Abilities Screening Instrument (CASI) [47] as well as the Chinese versions of the MMSE [48] and MoCA scales [49]. Detailed measurement procedures are provided in Multimedia Appendix 2.

Statistical Analysis

Statistical analysis was conducted using SPSS for Windows (version 21.0; IBM Corp). Continuous variables were represented by means and SDs, while categorical variables were expressed as frequencies and percentages. Before each analysis, normality and homogeneity of variance were preliminarily assessed. For variables conforming to a normal distribution, we applied 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables to determine the significance of differences between groups. The statistical significance level for all tests was set at $P<.05$.

To assess the impact of the intervention on outcomes across different time points, we performed a mixed ANOVA, with time as a within- participant factor and the nature of the intervention as a between-participant factor. Effect sizes were calculated using η^2, which ranges from 0 to 1, to quantify the strength of the intervention's influence. Generally, higher η^2 values indicate a more significant proportion of variance in the dependent variable explained by the intervention, thus reflecting a more substantial intervention effect. For any significant interactions or main effects, Bonferroni correction was used for post hoc evaluations.

Results

Baseline Performance of the Intervention and Control Groups

In this study, a total of 313 participants were screened for eligibility at LTCFs, including rural daycare centers and nursing homes. Of these, 232 participants were included and randomly assigned to either the intervention group (n=116) or the control group (n=116) (Figure 2). Notably, no dropouts occurred during the 12-week intervention period or the 6-month follow-up after the intervention.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

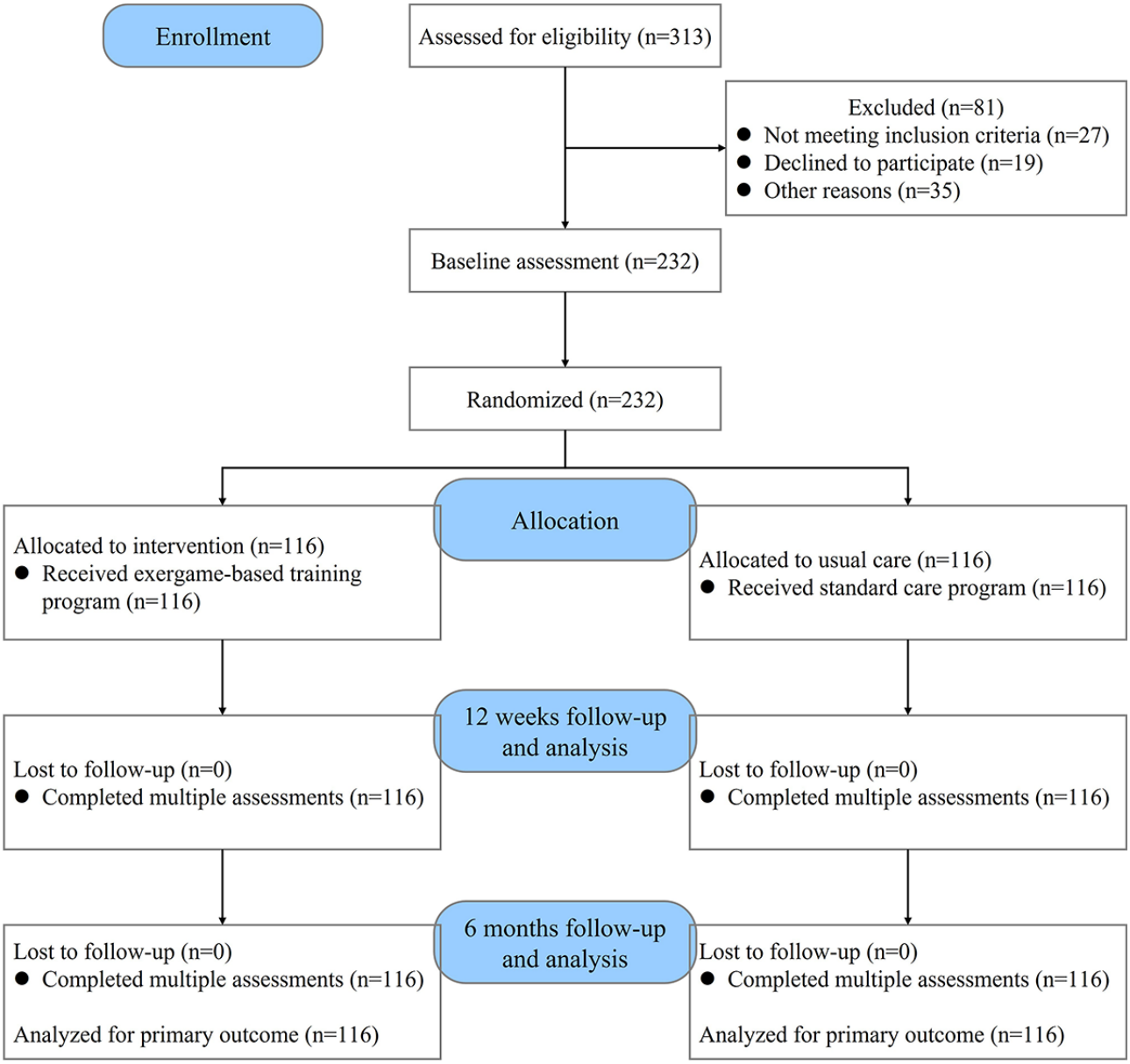

This study assessed baseline performance for both the intervention and control groups before the intervention began. The evaluation covered multiple domains, including physical flexibility, joint mobility, balance, coordination, and cognitive abilities. As presented in Multimedia Appendix 3, no significant differences were observed between the 2 groups across these various assessments (P>.05), indicating that participants exhibited relatively comparable physical function and cognitive status before the intervention.

Comparison of Physical Function Outcomes Between Intervention and Control Groups

As detailed in Multimedia Appendix 3, the sit and reach test results indicated a significant reduction in the distance between participants' hands and toes over time ($F_{2,78}$=35.682; P<.001). A significant group × time interaction effect was also observed ($F_{2,156}$=8.484; P<.001), demonstrating the intervention group exhibited a markedly more significant flexibility improvement than the control group. A similar trend was observed in the shoulder flexibility test, where the distance between participants' hands positioned behind the back showed a significant time effect ($F_{2,78}$=16.511; P<.001) and a significant group × time interaction effect ($F_{2,156}$=3.666; P=.04). In addition, in the trunk rotation flexibility test, the intervention group exhibited a significant increase in rotational range, with a significant interaction effect ($F_{2,156}$=17.353; P<.001) and a moderate-to-small effect size (η^2=0.182). Post hoc analysis demonstrated that physical flexibility indicators in the intervention group showed substantial improvement from baseline to the end of the intervention and remained consistently enhanced up to 6 months post intervention, indicating the durability and stability of the intervention effects.

Furthermore, the joint range of motion was assessed by comparing participants' maximal range of motion in 4 directions (eg, shoulder flexion, extension, abduction, and adduction). The results indicated a significant increase over time in shoulder flexion ($F_{2,78}$=36.315; P<.001) and abduction ($F_{2,78}$=4.499; P=.02) among participants. Further analysis revealed a

XSL•FO
RenderX

significant group × time interaction effect, with participants receiving the intervention demonstrating more pronounced improvements in shoulder flexion ($F_{2,156}$=17.655; P<.001) and abduction ($F_{2,156}$=6.281, P=.004), suggesting a positive impact of the intervention on enhancing range of motion in specific directions. In addition, the elbow range of motion test showed a significant increase in elbow flexion over time ($F_{2,78}$=9.261; P<.001), along with a significant group × time interaction effect ($F_{2,156}$=17.353; P<.001) and a small effect size (η^2=0.041). These effects may be related to the nature of the intervention tasks (eg, simulated swimming, skiing, and flying), emphasizing shoulder flexion, abduction, and elbow flexion, thereby enhancing the range of motion more substantially in these directions.

Subsequently, in the assessment of motor coordination, the figure-of-8 walk and standing balance test were used to evaluate participants' functional performance based on time metrics. The figure-of-8 walk test measured walking coordination by recording completion time, with shorter times indicating better coordination. The standing balance test assessed static balance ability by recording how participants could maintain balance while standing with their eyes closed, with longer durations indicating better balance. Post hoc analysis revealed a significant time effect in the completion time of the figure-of-8 walk test ($F_{2,78}$=36.038; P<.001) as well as a significant group × time interaction effect ($F_{2,156}$=11.846; P<.001), suggesting that the intervention group showed a notably more significant improvement in walking speed and motor coordination than the control group. Similarly, the results from the standing balance test demonstrated a significant increase over time in participants' balance duration with eyes closed ($F_{2,78}$=9.059; P<.001), indicating a positive impact of the intervention on static balance ability.

Finally, the box block test, used to assess hand dexterity, revealed a significant increase in the number of blocks transferred by participants within 1 minute over time ($F_{2,78}$=18.648; P<.001). In addition, a significant group × time interaction effect was observed ($F_{2,156}$=4.016; P=.02), indicating that the intervention group demonstrated a substantially greater improvement in hand dexterity than the control group.

Comparison of Cognitive Function Outcomes Between Intervention and Control Groups

This study assessed cognitive function by examining changes in performance on cognitive scales (eg, CASI, MMSE, and MoCA) between the intervention and control groups, as shown in Multimedia Appendix 3. In the CASI assessment, participants' scores significantly increased over time ($F_{2,78}$=21.427; P<.001), with a notable group × time interaction effect ($F_{2,156}$=13.226; P<.001), suggesting a positive impact of the intervention on cognitive function. A similar pattern was observed in both the MMSE and MoCA assessments. For the MMSE, scores significantly improved over time, with a strong time effect ($F_{2,78}$=94.552; P<.001) and a significant group × time interaction effect ($F_{2,156}$=53.244; P<.001), supporting the pronounced cognitive improvement within the intervention group. Likewise, MoCA scores demonstrated a significant time effect ($F_{2,78}$=52.563; P<.001) and a group × time interaction effect ($F_{2,156}$=15.882; P<.001), reinforcing the intervention's effectiveness in enhancing various aspects of cognitive function. Notably, the effect sizes (η^2) for the group × time interaction effects across the 3 scales were 0.145 for CASI, 0.406 for MMSE, and 0.169 for MoCA. The relatively larger effect size for MMSE suggests its higher sensitivity in detecting the intervention's impact on cognitive outcomes.

Discussion

Advantages of "WarioWare: Move It!" in Physical and Cognitive Rehabilitation for Older Adults

To our knowledge, this is the first study to evaluate the clinical efficacy of an exergame intervention on a range of functional outcomes, including physical flexibility, joint mobility, motor coordination, hand dexterity, and cognitive status. "WarioWare: Move It!" integrates elements of aerobic exercise, physical coordination, balance training, and cognitive engagement into a motion-based game, with real-time feedback mechanisms to enhance the experience. The immersive setting encourages active participation among older adults, which is particularly valuable in LTCFs in rural areas, where limited medical and social resources make it challenging to implement structured exercise programs. Therefore, we explored the application value of "WarioWare: Move It!" in systematic rehabilitation, focusing on its feasibility in improving older adults' participation and promoting functional recovery. The findings reveal that the older adult participants in LTCFs who received the "WarioWare: Move It!" intervention showed significantly improved flexibility, joint mobility, hand dexterity, and cognitive function compared to a standard care control group. It highlights the potential of "WarioWare: Move It!" as an innovative therapeutic strategy, offering an accessible and well-accepted rehabilitation approach that may help mitigate age-related declines in physical function within long-term care settings.

In addition, this study innovatively incorporates "WarioWare: Move It!" into the design of a motion-based intervention program for older adults. Similar to other exergames, it uses role-playing game mechanics and immersive scene design to foster participant engagement through a playful approach [50,51]. The program offers visual feedback (via screen display) and tactile feedback (through the Joy-Con controller) while incorporating diverse movement postures and randomly varied actions, thus extending beyond traditional exergames limited to fingertip control. Notably, "WarioWare: Move It!" presents a multicomponent training scheme that combines aerobic exercise, body coordination, balance training, and cognitive stimulation to enhance participants' cardiorespiratory endurance, balance, flexibility, and cognitive functions. Prior studies have demonstrated that such multicomponent exergaming provides substantial benefits to physical and cognitive functions in community-dwelling older adults due to its accessibility, cost-effectiveness, and time efficiency, earning broad acceptance [52-54]. Consistent with these findings, this study observed that "WarioWare: Move It!" led to marked improvements in participants' physical abilities (eg, coordination, balance, and

XSL•FO
RenderX

flexibility) and cognitive capacities (eg, executive function, attention, and memory).

One unique aspect of the exergame provided by "WarioWare: Move It!" lies in its ability to automatically set an initial difficulty level at each stage and progressively adjust this based on the player's performance. As players become more familiar with specific game scenarios and movement patterns, the system increases the challenge by gradually increasing the frequency and diversity of the required actions, thus prompting players to achieve efficient coordination between body and mind within a short time frame. Such dynamic adaptability enhances the game's appeal and imposes higher demands on participants' reaction speed, coordination, and focus.

Another fascinating aspect of "WarioWare: Move It!" is its thoughtful inclusion of 3 distinct multiplayer modes, each tailored to meet the preferences of different types of players. In addition to a robust single-player training mode, the game introduces competitive, cooperative, and duel modes to enrich social interactions and enhance player engagement and immersion. First, the competitive mode enables players to compete in specific tasks, with victory determined by scoring or task completion speed, fostering a sense of challenge and achievement. Next, the cooperative mode emphasizes teamwork, encouraging players to complete tasks together, thus strengthening team dynamics and promoting social cohesion. Then, the duel mode focuses on direct player versus player confrontations, offering immediate feedback and creating an intense, engaging battle experience. Overall, these multiplayer interaction modes diversify the game's social and interactive dimensions and offer varied engagement pathways for different player types, thereby boosting player loyalty and satisfaction.

Finally, while this study focused on LTCFs, the intervention's design also makes it feasible for implementation in non-LTCF rural environments. The equipment requirements for "WarioWare: Move It!"—a Joy-Con controller and a screen—are minimal, allowing for adaptation in community centers, individual homes, or other communal spaces. With proper training, family members or community volunteers could facilitate the sessions, alleviating the need for professional caregivers. Moreover, the intervention's engaging and intuitive gameplay makes it accessible even for individuals unfamiliar with gaming technology. Expanding such interventions beyond LTCFs could significantly enhance access to physical and cognitive rehabilitation for older adults in resource-limited settings, promoting healthier aging outcomes across diverse rural populations.

Implications of Multidimensional Improvements in Functional Performance Among Older Adults

In this study, we investigated whether early improvement trends in body coordination and flexibility could be identified in older adults during training. Age-related declines in muscle elasticity and neuromuscular control typically negatively affect coordination and flexibility in older adults [55]. Unlike younger individuals, older adults tend to benefit from training primarily through adaptive changes in the neuromuscular system, improving motor control, postural stability, and coordination, rather than relying solely on muscle hypertrophy or strength gains [56,57]. Our findings indicated that after the 12-week training period, the remaining distance between the fingertips and toes during the sit and reach test as well as the distance between the hands during the shoulder flexibility test were significantly reduced, indicating a positive effect of the intervention on enhancing both lower and upper limb flexibility. In addition, the trunk rotation angle in both directions significantly increased, further demonstrating the improvement in trunk rotation capacity and overall coordination. Although short-term enhancements in muscle mass may be challenging to detect, neuromuscular adaptations have already positively impacted functional performance in the early stages. In other words, through dynamic contextualized training, older adults experienced enhanced neuromuscular control and motor flexibility, leading to faster, observable improvements in coordination and flexibility.

We also observed a significant improvement in participants' range of motion in specific directions of the shoulder joint, such as flexion and abduction. These movements are crucial for many daily activities, such as dressing and reaching, and their enhancement directly impacts functional performance and independence in older adults' everyday lives [58]. The direction-specific improvement is primarily attributed to the contextualized exercises embedded within the intervention (eg, swimming, flying, and skiing simulations), emphasizing functional relevance and practical application, thereby fostering adaptations in the associated muscle groups and neuromuscular pathways. In addition, the range of motion in elbow joint demonstrated a significant group × time interaction effect in flexion angle, reflecting the rapid adaptability of the neuromuscular system during the short-term intervention. Compared to muscle hypertrophy, older adults are more likely to improve motor control and joint flexibility through neuromuscular adaptations in the short term [25,59], leading to prioritized enhancements in specific directions for the shoulder and elbow joints.

Although the gamified intervention in this study primarily targeted upper limb and trunk movements, a significant increase in walking speed was also observed, indicating the potential benefits of diverse game-based training for gait-related abilities. Walking involves coordinated lower limb and trunk movements, gait control, balance adjustment, and core stability [60,61]. In this study, game interventions (eg, racing, skiing, flying, and archery) not only required upper limb and trunk movements but also challenged balance and center of gravity control through dynamic postures (eg, leaning forward, reaching forward, and lateral shifting). The multifaceted posture control and center of gravity shifting training facilitated improvements in postural stability, balance perception, and neuromuscular coordination, ultimately enhancing gait performance. Moreover, the intervention group significantly reduced completion time for the figure-of-8 walk test, indicating improved stability and turn control during dynamic gait tasks. Similarly, increased duration in the standing balance test reflects improved balance, suggesting that the intervention significantly enhanced participants' neuromuscular control in dynamic and balance tasks. Although the training did not directly target lower limb gait, the center of gravity adjustments, posture transitions, and

XSL•FO
RenderX

full-body coordination exercises embedded within the games significantly improved gait performance.

The intervention program demonstrated significant efficacy in enhancing fine motor control and flexibility of the hands. During the intervention, participants performed hand movement tasks with the Joy-Con controller, such as simulating a tug-of-war, making rock-paper-scissors gestures, and quickly catching a flying disc. These tasks required precise hand adjustments across various contexts, emphasizing independent finger movement and wrist flexibility. In addition, the intervention group showed a notable increase in the number of blocks transferred within 1 minute, alongside a significant group × time interaction effect. This contextualized hand movement training facilitated improvements in fine finger control and wrist flexibility among older adults, with implications for daily tasks requiring fine motor skills, such as grasping objects, handling utensils, and using tools [62].

Finally, this study observed significant improvements in participants' cognitive performance across various assessment tools (eg, CASI, MMSE, and MoCA). Existing research indicates that gamified interventions have a therapeutic impact on the cognitive functions of older adults [63,64]. In this study, the gamified environment provided multidimensional stimulation through visual, auditory, and tactile cues, along with task-based prompts and real-time feedback, consistently presenting participants with spatial and logical challenges. Immersion in such virtual scenarios contributed not only to attentional recovery and stress reduction but also played a positive role in cognitive rehabilitation and enhancement [65]. Moreover, the nature of gamified training, which requires participants to make quick decisions in dynamic scenarios, stimulated executive functioning and response capabilities [66]. Notably, the design features of "WarioWare: Move It!" integrated aerobic exercise, coordination, and balance training alongside cognitive tasks involving memory, attention, and logical reasoning. These multilayered cognitive stimulations created a comprehensive training environment, promoting broad activation of brain functions. Therefore, the notable improvements observed in brain health assessments postintervention were anticipated, highlighting the significant role of gamified interventions in cognitive rehabilitation and enhancement for older adults.

Insights for Rehabilitation Professionals and Human-Computer Interaction Researchers in Older Adult Care

This study offers a new perspective for rehabilitation professionals working with older adult populations, focusing on three key areas: (1) integrating multicomponent cognitive training with physical function restoration—"WarioWare: Move It!" effectively combines aerobic exercise, body coordination, balance, and cognitive stimulation within a structured platform that facilitates multifunctional training through various interactive physical movements—and (2) enhancing adherence through practicality and enjoyment—by integrating intuitive visual feedback, goal achievement mechanisms, immersive scenarios, real-time feedback, and progressively challenging tasks, WarioWare: Move It! Fosters a sense of accomplishment and interactive enjoyment, reinforcing experiential learning and significantly improving adherence—and (3) Practical application of targeted physical function training—through contextualized and interactive movement design, "WarioWare: Move It!" incorporates actions frequently performed in daily life (eg, shoulder flexion and abduction, gait balance, and hand movements) within gameplay to enhance specific physical functions (eg, joint mobility and flexibility).

This study offers valuable insights for researchers in the human-computer interaction field, specifically in designing older adult–friendly interactive systems. Specifically, it highlights three key areas: (1) dynamic adaptive design for older adult–friendly interventions—the dynamic adaptability of the system in "WarioWare: Move It!" gradually guides participants toward higher levels of physical and cognitive demands, facilitating synchronized improvements in both domains within a safe and comfortable experience; (2) multisensory feedback and enhanced interaction—the Intervention successfully integrates multisensory stimulation through visual, auditory, and tactile feedback, enhancing the immersive experience and improving the brain's responsiveness through multimodal input; and (3) rehabilitation benefits of multiple social interaction modes—the social interaction modes within the exergames, including competition (which stimulates participants' achievement motivation), cooperation (which strengthens teamwork awareness), and versus challenges (which foster quick decision-making and response skills), not only enrich the social experience but also meet the psychological and emotional needs of older adults, thus enhancing rehabilitation outcomes.

Limitations and Future Work

Our study has several limitations. First, although the sample size exceeds the requirements for statistical analysis, it remains relatively small. Second, although the exergames have positively affected specific physical and cognitive functions, they cannot fully replace traditional rehabilitation methods (eg, physical or cognitive behavioral therapy). Future intervention designs could consider integrating exergames with conventional rehabilitation techniques, creating a more diversified and comprehensive rehabilitation plan to meet the broader needs of older adults. Third, the social interaction features of "WarioWare: Move It!" (eg, competition, cooperation, and rivalry) enhance rehabilitation outcomes. However, preferences for competition or collaborative modes may vary among older adults, particularly those who are more introverted or have lower social engagement, making these modes less appealing to some individuals. Future game iterations could incorporate personalized social interaction options to ensure each participant receives an optimal experience, boosting engagement. Fourth, although this study focused on older adults, differences in age, stage of cognitive impairment, and the specific needs of different age groups may warrant further investigation to understand the broader applicability of the intervention.

Conclusions

The "WarioWare: Move It!" training program significantly improves physical flexibility, joint range of motion, motor coordination, hand dexterity, and cognitive function among older adults in rural LTCFs. For patients with MCI or dementia, this intervention not only aids in delaying functional decline

XSL•FO
RenderX

but also enhances cognitive function and quality of daily living through multisensory stimulation and dynamic physical engagement. Furthermore, the incorporation of exergames marks an innovative shift in older adult care models, offering a novel, low-demand intervention method that effectively addresses the shortcomings of traditional care approaches. These findings support the integration of exergames into routine care practices and provide a viable, technology-driven solution to address the challenges of aging health in institutional care settings worldwide.

Acknowledgments

This research received funding from the National Natural Science Foundation of China (62476190, 62376183, and U21A20469), the Central Government–Guided Science and Technology Development Program (YDZJSX2022C004), and Shanxi Provincial Administration of Traditional Chinese Medicine Scientific Research Project Plan (2024ZYY2A033). The authors sincerely thank the volunteers whose invaluable participation made the successful completion of this study possible.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 359 KB - jmir_v27i1e69109_app1.pdf]

Multimedia Appendix 2
An overview of the broad scope of tests and key assessment areas.
[PDF File (Adobe PDF File), 110 KB - jmir_v27i1e69109_app2.pdf]

Multimedia Appendix 3
Data of measurement outcomes at baseline, posttest, and after 6 months postintervention.
[PDF File (Adobe PDF File), 243 KB - jmir_v27i1e69109_app3.pdf]

References

1. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. Lancet 2006 Apr 15;367(9518):1262-1270. [doi: 10.1016/S0140-6736(06)68542-5] [Medline: 16631882]
2. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125 [FREE Full text] [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
3. Petersen RC. Mild cognitive impairment. Continuum (Minneap Minn) 2016 Apr;22(2 Dementia):404-418 [FREE Full text] [doi: 10.1212/CON.0000000000000313] [Medline: 27042901]
4. Sachdev PS, Blacker D, Blazer DG, Ganguli M, Jeste DV, Paulsen JS, et al. Classifying neurocognitive disorders: the DSM-5 approach. Nat Rev Neurol 2014 Nov 30;10(11):634-642 [FREE Full text] [doi: 10.1038/nrneurol.2014.181] [Medline: 25266297]
5. Yiannopoulou KG, Papageorgiou SG. Current and future treatments in Alzheimer disease: an update. J Cent Nerv Syst Dis 2020 Feb 29;12:1179573520907397 [FREE Full text] [doi: 10.1177/1179573520907397] [Medline: 32165850]
6. Pacheco TB, de Medeiros CS, de Oliveira VH, Vieira ER, de Cavalcanti FA. Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis. Syst Rev 2020 Jul 18;9(1):163 [FREE Full text] [doi: 10.1186/s13643-020-01421-7] [Medline: 32682439]
7. Wu S, Ji H, Won J, Jo E, Kim Y, Park J. The effects of exergaming on executive and physical functions in older adults with dementia: randomized controlled trial. J Med Internet Res 2023 Mar 07;25:e39993 [FREE Full text] [doi: 10.2196/39993] [Medline: 36881445]
8. Hughes TF, Flatt JD, Fu B, Butters MA, Chang CH, Ganguli M. Interactive video gaming compared with health education in older adults with mild cognitive impairment: a feasibility study. Int J Geriatr Psychiatry 2014 Sep 22;29(9):890-898 [FREE Full text] [doi: 10.1002/gps.4075] [Medline: 24452845]

XSL•FO
RenderX

9. Li A, Li J, Wu W, Zhao J, Qiang Y. Effect of virtual reality training on cognitive function and motor performance in older adults with cognitive impairment receiving health care: a randomized controlled trial. Int J Hum Comput Interact 2023 Oct 19;40(22):7755-7772. [doi: 10.1080/10447318.2023.2271240]
10. Karssemeijer EG, Aaronson JA, Bossers WJ, Donders R, Olde Rikkert MG, Kessels RP. The quest for synergy between physical exercise and cognitive stimulation via exergaming in people with dementia: a randomized controlled trial. Alzheimers Res Ther 2019 Jan 05;11(1):3 [FREE Full text] [doi: 10.1186/s13195-018-0454-z] [Medline: 30611286]
11. Kamnardsiri T, Kumfu S, Munkhetvit P, Boripuntakul S, Sungkarat S. Home-based, low-intensity, gamification-based, interactive physical-cognitive training for older adults using the ADDIE model: design, development, and evaluation of user experience. JMIR Serious Games 2024 Oct 29;12:e59141 [FREE Full text] [doi: 10.2196/59141] [Medline: 39470391]
12. Chu CH, Biss RK, Cooper L, Quan AM, Matulis H. Exergaming platform for older adults residing in long-term care homes: user-centered design, development, and usability study. JMIR Serious Games 2021 Mar 09;9(1):e22370 [FREE Full text] [doi: 10.2196/22370] [Medline: 33687337]
13. Kannan L, Sahu U, Subramaniam S, Mehta N, Kaur T, Hughes S, et al. Gaming-based tele-exercise program to improve physical function in frail older adults: feasibility randomized controlled trial. J Med Internet Res 2024 Nov 27;26:e56810 [FREE Full text] [doi: 10.2196/56810] [Medline: 39602215]
14. Chan G, Banire B, Anukem S, Imran M, Meena S, Nwagu C, et al. Social exergames in health and wellness: a systematic review of trends, effectiveness, challenges, and directions for future research. Int J Hum Comput Interact 2024 Jul 08:1-32. [doi: 10.1080/10447318.2024.2371686]
15. Du Q, Song Z, Jiang H, Wei X, Weng D, Fan M. LightSword: a customized virtual reality exergame for long-term cognitive inhibition training in older adults. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI p. 1-17 URL: https://dl.acm.org/doi/10.1145/3613904.3642187 [doi: 10.1145/3613904.3642187]
16. Chen Q, Wu Z, Liu Y, Han L, Li Z, Kan GL, et al. SilverCycling: exploring the impact of bike-based locomotion on spatial orientation for older adults in VR. Proc ACM Interact Mob Wearable Ubiquitous Technol 2024 Sep 09;8(3):1-24. [doi: 10.1145/3678522]
17. Meekes W, Stanmore EK. Motivational determinants of exergame participation for older people in assisted living facilities: mixed-methods study. J Med Internet Res 2017 Jul 06;19(7):e238 [FREE Full text] [doi: 10.2196/jmir.6841] [Medline: 28684385]
18. Unbehaun D, Vaziri DD, Aal K, Wieching R, Tolmie P, Wulf V. Exploring the potential of exergames to affect the social and daily life of people with dementia and their caregivers. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-15 URL: https://dl.acm.org/doi/10.1145/3173574.3173636 [doi: 10.1145/3173574.3173636]
19. Chen KM, Li CH, Huang HT, Cheng Y. Feasible modalities and long-term effects of elastic band exercises in nursing home older adults in wheelchairs: a cluster randomized controlled trial. Int J Nurs Stud 2016 Mar;55:4-14. [doi: 10.1016/j.ijnurstu.2015.11.004] [Medline: 26655368]
20. Karaosmanoglu S, Rings S, Kruse L, Stein C, Steinicke F. Lessons learned from a human-centered design of an immersive exergame for people with dementia. Proc ACM Hum Comput Interact 2021 Oct 06;5(CHI PLAY):1-27. [doi: 10.1145/3474679]
21. Kappen DL, Mirza-Babaei P, Nacke LE. Older adults' physical activity and exergames: a systematic review. Int J Hum Comput Interact 2018 Feb 23;35(2):140-167. [doi: 10.1080/10447318.2018.1441253]
22. Wang RY, Huang YC, Zhou JH, Cheng S, Yang Y. Effects of exergame-based dual-task training on executive function and dual-task performance in community-dwelling older people: a randomized-controlled trial. Games Health J 2021 Oct 01;10(5):347-354. [doi: 10.1089/g4h.2021.0057] [Medline: 34491113]
23. Chiu HM, Hsu MC, Ouyang WC. Effects of incorporating virtual reality training intervention into health care on cognitive function and wellbeing in older adults with cognitive impairment: a randomized controlled trial. Int J Hum Comput Interact 2023 Feb;170:102957. [doi: 10.1016/j.ijhcs.2022.102957]
24. Seco J, Abecia LC, Echevarría E, Barbero I, Torres-Unda J, Rodriguez V, et al. A long-term physical activity training program increases strength and flexibility, and improves balance in older adults. Rehabil Nurs 2013;38(1):37-47. [doi: 10.1002/rnj.64] [Medline: 23365004]
25. Stathokostas L, Little RM, Vandervoort AA, Paterson DH. Flexibility training and functional ability in older adults: a systematic review. J Aging Res 2012;2012:306818-306830 [FREE Full text] [doi: 10.1155/2012/306818] [Medline: 23209904]
26. Coleman EA, Buchner DM, Cress ME, Chan BK, de Lateur BJ. The relationship of joint symptoms with exercise performance in older adults. J Am Geriatr Soc 1996 Jan 27;44(1):14-21. [doi: 10.1111/j.1532-5415.1996.tb05632.x] [Medline: 8537585]
27. Brach JS, Perera S, Shuman V, Gil AB, Kriska A, Nadkarni NK, et al. Effect of timing and coordination training on mobility and physical activity among community-dwelling older adults: a randomized clinical trial. JAMA Netw Open 2022 May 02;5(5):e2212921 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.12921] [Medline: 35604689]
28. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. Maturitas 2013 May;75(1):51-61. [doi: 10.1016/j.maturitas.2013.02.009] [Medline: 23523272]

XSL•FO
RenderX

29. Chang JT, Morton SC, Rubenstein LZ, Mojica WA, Maglione M, Suttorp MJ, et al. Interventions for the prevention of falls in older adults: systematic review and meta-analysis of randomised clinical trials. BMJ 2004 Mar 20;328(7441):680 [FREE Full text] [doi: 10.1136/bmj.328.7441.680] [Medline: 15031239]
30. Tricco AC, Thomas SM, Veroniki AA, Hamid JS, Cogo E, Strifler L, et al. Comparisons of interventions for preventing falls in older adults: a systematic review and meta-analysis. JAMA 2017 Nov 07;318(17):1687-1699 [FREE Full text] [doi: 10.1001/jama.2017.15006] [Medline: 29114830]
31. Scherder E, Dekker W, Eggermont L. Higher-level hand motor function in aging and (preclinical) dementia: its relationship with (instrumental) activities of daily life--a mini-review. Gerontology 2008 Nov 11;54(6):333-341 [FREE Full text] [doi: 10.1159/000168203] [Medline: 18997468]
32. Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice. 10th edition. New York, NY: Lippincott Williams & Wilkins; 2008.
33. Wells KF, Dillon EK. The sit and reach—a test of back and leg flexibility. Res Q Am Assoc Health Phys Educ 2013 Dec 26;23(1):115-118. [doi: 10.1080/10671188.1952.10761965]
34. Jones CJ, Rikli RE, Max J, Noffal G. The reliability and validity of a chair sit-and-reach test as a measure of hamstring flexibility in older adults. Res Q Exerc Sport 1998 Dec;69(4):338-343. [doi: 10.1080/02701367.1998.10607708] [Medline: 9864752]
35. Borstad JD, Mathiowetz KM, Minday LE, Prabhu B, Christopherson DE, Ludewig PM. Clinical measurement of posterior shoulder flexibility. Man Ther 2007 Nov;12(4):386-389. [doi: 10.1016/j.math.2006.07.014] [Medline: 16971161]
36. Bakhsh W, Nicandri G. Anatomy and physical examination of the Shoulder. Sports Med Arthrosc Rev 2018 Sep;26(3):e10-e22. [doi: 10.1097/JSA.0000000000000202] [Medline: 30059442]
37. Merritt JL, McLean TJ, Erickson RP, Offord KP. Measurement of trunk flexibility in normal subjects: reproducibility of three clinical methods. Mayo Clin Proc 1986 Mar;61(3):192-197. [doi: 10.1016/s0025-6196(12)61848-5] [Medline: 3945120]
38. Hayes K, Walton JR, Szomor ZL, Murrell GA. Reliability of five methods for assessing shoulder range of motion. Aust J Physiother 2001;47(4):289-294. [doi: 10.1016/s0004-9514(14)60274-9] [Medline: 11722295]
39. Mullaney MJ, McHugh MP, Johnson CP, Tyler TF. Reliability of shoulder range of motion comparing a goniometer to a digital level. Physiother Theory Pract 2010 Jul 17;26(5):327-333. [doi: 10.3109/09593980903094230] [Medline: 20557263]
40. Armstrong AD, MacDermid JC, Chinchalkar S, Stevens RS, King GJ. Reliability of range-of-motion measurement in the elbow and forearm. J Shoulder Elbow Surg 1998 Nov;7(6):573-580. [doi: 10.1016/s1058-2746(98)90003-9] [Medline: 9883416]
41. Wong SS, Yam MS, Ng SS. The Figure-of-Eight Walk test: reliability and associations with stroke-specific impairments. Disabil Rehabil 2013 Apr 19;35(22):1896-1902. [doi: 10.3109/09638288.2013.766274] [Medline: 23600714]
42. Dhillon RK, Sharma M. A review on figure of eight walking in physical therapy rehabilitation of elderly. Work Older People 2021 Aug 17;25(3):214-219. [doi: 10.1108/wwop-04-2021-0020]
43. Browne JE, O'Hare NJ. Review of the different methods for assessing standing balance. Physiotherapy 2001 Sep;87(9):489-495. [doi: 10.1016/S0031-9406(05)60696-7]
44. Newton R. Review of tests of standing balance abilities. Brain Inj 1989 Jul 03;3(4):335-343. [doi: 10.3109/02699058909004558] [Medline: 2684312]
45. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the Box and Block Test of manual dexterity. Am J Occup Ther 1985 Jun;39(6):386-391. [doi: 10.5014/ajot.39.6.386] [Medline: 3160243]
46. Desrosiers J, Bravo G, Hébert R, Dutil E, Mercier L. Validation of the Box and Block Test as a measure of dexterity of elderly people: reliability, validity, and norms studies. Arch Phys Med Rehabil 1994 Jul;75(7):751-755. [Medline: 8024419]
47. Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. Int Psychogeriatr 1994 Mar;6(1):45-62. [doi: 10.1017/s1041610294001602] [Medline: 8054493]
48. Guo NW, Liu HC, Wong PF, Liao KK, Yan SH, Lin KP, et al. Chinese version and norms of the mini-mental state examination. Rehabil Pract Sci 1988 Dec 1;16(1):52. [doi: 10.6315/3005-3846.1745]
49. Yu J, Li J, Huang X. The Beijing version of the Montreal Cognitive Assessment as a brief screening tool for mild cognitive impairment: a community-based study. BMC Psychiatry 2012 Sep 25;12(1):156 [FREE Full text] [doi: 10.1186/1471-244X-12-156] [Medline: 23009126]
50. Schättin A, Pickles J, Flagmeier D, Schärer B, Riederer Y, Niedecken S, et al. Development of a novel home-based exergame with on-body feedback: usability study. JMIR Serious Games 2022 Dec 06;10(4):e38703 [FREE Full text] [doi: 10.2196/38703] [Medline: 36472900]
51. Richards C, Graham TC. Developing compelling repetitive-motion exergames by balancing player agency with the constraints of exercise. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems. 2016 Presented at: DIS '16; June 4-8, 2016; Brisbane, Australia p. 911-923 URL: https://dl.acm.org/doi/10.1145/2901790.2901824 [doi: 10.1145/2901790.2901824]
52. Moreira NB, Rodacki AL, Costa SN, Pitta A, Bento PC. Perceptive-cognitive and physical function in prefrail older adults: exergaming versus traditional multicomponent training. Rejuvenation Res 2021 Feb 01;24(1):28-36. [doi: 10.1089/rej.2020.2302] [Medline: 32443963]

XSL•FO
RenderX

53. Zhao Y, Feng H, Wu X, Du Y, Yang X, Hu M, et al. Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. JMIR Serious Games 2020 Jun 30;8(2):e16841 [FREE Full text] [doi: 10.2196/16841] [Medline: 32602841]
54. Liao YY, Chen IH, Hsu WC, Tseng H, Wang R. Effect of exergaming versus combined exercise on cognitive function and brain activation in frail older adults: a randomised controlled trial. Ann Phys Rehabil Med 2021 Sep;64(5):101492 [FREE Full text] [doi: 10.1016/j.rehab.2021.101492] [Medline: 33454398]
55. Bellew JW, Symons BT, Vandervoort AA. Geriatric fitness: effects of aging and recommendations for exercise in older adults. Cardiopulm Phys Ther J 2005;16(1):20-31. [doi: 10.1097/01823246-200516010-00005]
56. Moritani T, deVries HA. Potential for gross muscle hypertrophy in older men. J Gerontol 1980 Sep 01;35(5):672-682. [doi: 10.1093/geronj/35.5.672] [Medline: 7430562]
57. Macaluso A, De Vito G. Muscle strength, power and adaptations to resistance training in older people. Eur J Appl Physiol 2004 Apr 25;91(4):450-472. [doi: 10.1007/s00421-003-0991-3] [Medline: 14639481]
58. Oosterwijk AM, Nieuwenhuis MK, van der Schans CP, Mouton L. Shoulder and elbow range of motion for the performance of activities of daily living: a systematic review. Physiother Theory Pract 2018 Jul 29;34(7):505-528 [FREE Full text] [doi: 10.1080/09593985.2017.1422206] [Medline: 29377745]
59. Holland GJ, Tanaka K, Shigematsu R, Nakagaichi M. Flexibility and physical functions of older adults: a review. J Aging Phys Act 2002;10(2):169-206 [FREE Full text] [doi: 10.1123/japa.10.2.169]
60. Clark DJ. Automaticity of walking: functional significance, mechanisms, measurement and rehabilitation strategies. Front Hum Neurosci 2015 May 05;9:246 [FREE Full text] [doi: 10.3389/fnhum.2015.00246] [Medline: 25999838]
61. Lacquaniti F, Grasso R, Zago M. Motor patterns in walking. News Physiol Sci 1999 Aug;14(4):168-174. [doi: 10.1152/physiologyonline.1999.14.4.168] [Medline: 11390844]
62. Liu B, Chen X, Li Y, Liu H, Guo S, Yu P. Effect of passive finger exercises on grip strength and the ability to perform activities of daily living for older people with dementia: a 12-week randomized controlled trial. Clin Interv Aging 2018;13:2169-2177 [FREE Full text] [doi: 10.2147/CIA.S174756] [Medline: 30464426]
63. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DM. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644 [FREE Full text] [doi: 10.2196/18644] [Medline: 32773374]
64. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
65. Yen HY, Chiu HL. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021 May;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
66. Liao YY, Chen IH, Lin YJ, Chen Y, Hsu W. Effects of virtual reality-based physical and cognitive training on executive function and dual-task gait performance in older adults with mild cognitive impairment: a randomized control trial. Front Aging Neurosci 2019 Jul 16;11:162 [FREE Full text] [doi: 10.3389/fnagi.2019.00162] [Medline: 31379553]

Abbreviations

CASI: Cognitive Abilities Screening Instrument
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GDS: Global Deterioration Scale
LTCF: long-term care facility
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment

Edited by A Mavragani; submitted 22.11.24; peer-reviewed by Y Huusi, T Ramsey, L Yao, R Koch; comments to author 22.12.24; revised version received 30.12.24; accepted 07.01.25; published 19.02.25.

Please cite as:
Li A, Qiang W, Li J, Geng Y, Qiang Y, Zhao J
Evaluating the Clinical Efficacy of an Exergame-Based Training Program for Enhancing Physical and Cognitive Functions in Older Adults With Mild Cognitive Impairment and Dementia Residing in Rural Long-Term Care Facilities: Randomized Controlled Trial
J Med Internet Res 2025;27:e69109
URL: https://www.jmir.org/2025/1/e69109
doi:10.2196/69109
PMID:39969990

XSL•FO
RenderX

©Aoyu Li, Wei Qiang, Jingwen Li, Yan Geng, Yan Qiang, Juanjuan Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cost Utility Analysis of Internet-Based Cognitive Behavioral Therapy for Major Depressive Disorder: Randomized Controlled Trial

Wenjing Zhou[1,2*], PhD; Yan Chen[3*], MAPH; Herui Wu[1,2], MMS; Hao Zhao[1,2], PhD; Yanzhi Li[1,2], PhD; Guangduoji Shi[1,2], PhD; Wanxin Wang[1,2], PhD; Yifeng Liu[3], MMD; Yuhua Liao[3], PhD; Huimin Zhang[3], PhD; Caihong Gao[3], BSPH; Jiejing Hao[1,2,3], PhD; Gia Han Le[4,5,6,7], MSc; Roger S McIntyre[7], PhD; Xue Han[3], MD; Ciyong Lu[1,2], PhD

[1]Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Guangdong Provincial Key Laboratory of Food, Nutrition and Health, Sun Yat-sen University, Guangzhou, China

[3]Department of Psychiatry, Shenzhen Nanshan Center for Chronic Disease Control, Shenzhen, China

[4]Brain and Cognition Discovery Foundation, Toronto, ON, Canada

[5]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

[6]Mood Disorder Psychopharmacology, University Health Network, Toronto, ON, Canada

[7]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Ciyong Lu, PhD
Department of Medical Statistics and Epidemiology
School of Public Health
Sun Yat-sen University
No.74 Zhongshan Road 2
Guangzhou, 510080
China
Phone: 86 13610355985
Email: luciyong@mail.sysu.edu.cn

Abstract

Background: Unguided internet-based cognitive behavioral therapy (ICBT) has been proven effective for major depressive disorder (MDD). However, few studies have examined its cost-effectiveness in low-resource countries and under nonspecialist routine care.

Objective: This study aimed to evaluate the short- and long-term cost utility of unguided ICBT compared to a waitlist control for persons with MDD from the perspectives of society and the health care system.

Methods: This analysis was implemented alongside an 8-week 2-arm randomized controlled trial with a 12-month follow-up period conducted in Shenzhen, China. Outcomes including cost and health utility were collected at the pretreatment and posttreatment time points and 3, 6, and 12 months after the intervention. Direct medical costs and indirect costs were prospectively collected using the hospital information system and the Sheehan Disability Scale. Health outcomes were measured using the Chinese version of the Short-Form Six-Dimension health index. The primary outcome was incremental cost utility ratio (ICUR) expressed as the difference in costs between 2 therapies by the difference in quality-adjusted life years (QALYs). The seemingly unrelated regression model and the bootstrap method were performed to estimate adjusted ICURs. Cost-effectiveness planes and cost-effectiveness acceptability curves were used to demonstrate uncertainty. A series of scenario analyses were conducted to verify the robustness of base-case results.

Results: In total, 244 participants with MDD were randomly allocated to the ICBT (n=122, 50%) or waitlist control (n=122, 50%) groups. At the pretreatment time point, no statistically significant difference was observed in direct medical cost (*P*=.41), indirect cost (*P*=.10), or health utility (*P*=.11) between the 2 groups. In the base-case analysis, the ICBT group reported higher direct medical costs and better quality of life but lower total costs at the posttreatment time point. The adjusted ICURs at the posttreatment time point were CN ¥–194,720.38 (US $–26,551.50; 95% CI CN ¥–198,766.78 to CN ¥–190,673.98 [US $–27,103.20 to US $–25,999.70]) and CN ¥49,700.33 (US $6776.99; 95% CI CN ¥46,626.34-CN ¥52,774.31 [US $6357.83-$7196.15]) per

XSL•FO
RenderX

QALY from the societal and health care system perspectives, respectively, with a probability of unguided ICBT being cost-effective of 75.93% and 54.4%, respectively, if the willingness to pay was set at 1 time the per-capita gross domestic product. In the scenario analyses, the probabilities increased to 76.85% and 77.61%, respectively, indicating the potential of ICBT to be cost-effective over the long term.

Conclusions: Unguided ICBT is a cost-effective treatment for MDD. This intervention not only helps patients with MDD improve clinically but also generates societal savings. These findings provide health economic evidence for a potential scalable MDD treatment method in low- and middle-income countries.

Trial Registration: Chinese Clinical Trial Registry (ChiCTR) ChiCTR2100046425; https://tinyurl.com/bdcrj4zv

(J Med Internet Res 2025;27:e67567) doi:10.2196/67567

KEYWORDS

cost utility analysis; CUA; cost-effectiveness; economic evaluation; costs; quality of life; internet-based cognitive behavioral therapy; ICBT; digital psychiatry; major depressive disorder; depression; China

Introduction

Background

Major depressive disorder (MDD) is a highly prevalent mental disorder estimated to affect 332 million people worldwide in 2021 [1]. MDD is associated with significant impairment of health-related quality of life (HRQoL) [2], substantial health care use [3], and huge societal impacts on employment and productivity [4,5]. Now, MDD is ranked as the second leading cause of global nonfatal disability [1], and MDD-attributed lost productivity costs the global economy US $1 trillion annually [6]. Effective, scalable, and cost-effective therapies are needed to reduce the disease burden caused by MDD [7].

Despite the availability of effective first-line treatments involving psychotherapies and pharmacotherapies for MDD [8], the treatment gap is still large [9,10]. Individuals with MDD are faced with a series of barriers to treatment, including social discrimination, stigma, health care costs, and a shortage of specialists [11,12], especially in low-resource settings such as China [13]. Unguided internet-based cognitive behavioral therapy (ICBT), a form of cognitive behavioral therapy delivered through the internet without therapeutic support [14], has the attributes of privacy, accessibility, flexibility, and ease of implementation with constrained health care resources and personnel [15,16]. It has been gradually applied for the treatment of MDD. Existing research across low-, middle-, and high-income countries has demonstrated the effectiveness of unguided ICBT for MDD in comparison with usual care [11,17-20]. In China, our team has also developed an unguided ICBT course for MDD, and its efficacy has been confirmed through a randomized controlled trial (RCT) [21].

In contrast with the well-established evidence of effectiveness, the cost-effectiveness of unguided ICBT for MDD remains relatively under-studied [22-30]. There is an increasing demand for economic evidence to optimize resource allocation and scale up the implementation of unguided ICBT in different contexts [31]. However, existing trial-based economic evaluations are mostly conducted in high-income countries and adopt a single perspective (societal or health care system perspective), and the results are inconclusive [32-34]. Previous cost utility analyses (CUAs) conducted from the perspective of the health care system over a long-term follow-up period (1-2 years) have reported that unguided ICBT had lower probability of being more cost-effective in comparison with usual care at setting values of willingness to pay (WTP), with the probability of unguided ICBT being cost-effective ranging from 3.8% to 41.7% [26-28,30]. In the case of those adopting the societal perspective, the findings have been mixed, and the supportive evidence has had some uncertainties [22-25,29,30]. Holst et al [30] conducted a CUA of unguided ICBT compared with usual care among participants with mild to moderate depression in Sweden and found that no firm conclusions could be drawn under a range of assumed WTP. In the Netherlands, all CUA analyses have reported positive results and suggested the economic merits of unguided ICBT compared to usual care, with the likelihood of the former being cost-effective being 52%, 65%, and 62% [23-25]. In addition, another 2 evaluations conducted in the United Kingdom and Spain demonstrated that unguided ICBT was the dominant treatment option in comparison with usual care [22,29]. Considering the large heterogeneity in current research, the generalizability and external validity of the cost-effectiveness conclusions require more research in different settings under economic guidance. To our knowledge, no study has reported the cost-effectiveness of unguided ICBT for MDD in China. It is necessary to extend the analysis to China and account for both the societal and health care system perspectives to provide decision-making evidence for the optimization and allocation of the limited health care resources.

Objectives

Therefore, this study aimed to comprehensively assess the short- and long-term cost utility of unguided ICBT among persons with MDD in China from the perspectives of society and the health care system.

Methods

Study Design

This economic evaluation was conducted prospectively alongside an 8-week pragmatic, unblinded, 2-arm RCT with a follow-up period of 12 months and was conducted from the perspectives of society and the Chinese health care system. The time horizon was in line with that of the RCT from baseline to 12 months after treatment. Participants were randomly allocated to the ICBT (ICBT plus usual care) or waitlist control (usual

XSL•FO
RenderX

care) groups at a ratio of 1:1. The detailed study design and procedures have been described in a published paper [21].

Ethical Considerations

This RCT was registered with the Chinese Clinical Trial Registry (ChiCTR2100046425) and obtained ethics approval from the Ethics Review Committee of Shenzhen Nanshan Center for Chronic Disease Control (ll20210012) in line with the standards stipulated in the Declaration of Helsinki. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist was used to guide and report the RCT [35]. In addition, the Consolidated Health Economic Evaluation Reporting Standards were applied to conduct and guide the economic evaluation [36]. All participants provided written informed consent, including for the use of their data for secondary analyses. The data used for the analyses were deidentified to protect the privacy of the participants. For those who completed the trial, financial incentives in the form of small amounts of electronic cash (CN ¥50 [US $6.82]) were distributed.

Study Participants

In this study, participants were recruited from the Department of Depressive Disorder, Shenzhen Kangning Hospital, and the Department of Psychiatry, Shenzhen Nanshan Center for Chronic Disease Control, between August 2021 and December 2022. To be eligible for this study, participants had to meet the following inclusion criteria: (1) age 18 to 60 years; (2) a positive screening result on the Patient Health Questionnaire–9 (PHQ-9) of ≥5 [37]; (3) a diagnosis of MDD through the Mini-International Neuropsychiatric Interview [38] through the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*; (4) access to internet-connected mobile devices; (5) lack of any other psychotherapy or physical therapy currently; (6) no changes in condition during the previous month (eg, patients with antidepressant use did not change the dosage); and (7) informed consent. The exclusion criteria were (1) neurological illness (eg, traumatic brain injury or functional impairment); (2) a moderate to high risk of suicide; (3) alcohol abuse or substance use disorder; (4) pregnancy or breastfeeding; and (5) diagnosis of a severe physical disease, psychosis, or bipolar disorder. Eligible participants were asked to complete the baseline assessment, including sociodemographic information (eg, age, sex, and ethnicity), clinical history (eg, antidepressant use and history of psychopathology), lifestyle (eg, smoking and drinking), and baseline clinical symptoms (measured using the PHQ-9, Generalized Anxiety Disorder–7, and Kessler Psychological Distress Scale).

Intervention

Participants in the ICBT group were provided with an ICBT course for 8 weeks in addition to usual care. Each participant was assigned an account number and corresponding password after the pretreatment instruction and training for use. The ICBT course is called Morning Mood; is embedded in the WeChat mini program; and can be accessed via smartphones, tablets, and computers. The course was grounded in cognitive behavioral therapy principles aiming at teaching emotional regulation skills and cognitive restructuring. In total, there are seven 30-minute modules. The detailed contents and corresponding screenshots are shown in Multimedia Appendix 1. Participants were required to complete 1 module per week. Technical support and reminder services via telephone call and SMS text messaging were provided by trained nonspecialists (ie, lay health workers, nurses, and social workers) throughout the intervention and follow-up periods to promote participant engagement. However, no therapeutic guidance was offered to them, and thus, this treatment was defined as an unguided ICBT course.

Participants in the waitlist control group were on a waiting list for ICBT and did not receive any additional specific interventions. However, individuals' usual treatment was maintained during the 8-week control period. After 8 weeks, the participants in the control arm completed the observation phase and began to receive the ICBT treatment.

Health Outcomes

The primary health outcome was quality-adjusted life years (QALYs), estimated using the Simplified Chinese version of the Short-Form Six-Dimension (SF-6D) health index, which is demonstrated to have good validity among the Chinese population, including individuals with mental health problems [39,40]. The SF-6D is one of the most widely used preference-based HRQoL instruments [41] and has been recommended by many national guidelines of economic evaluations for the calculation of QALYs [42]. The SF-6D defines the health state through the following 6 dimensions: physical functioning, role limitation, social functioning, pain, mental health, and vitality [43]. The SF-6D responses were converted to health utility on the QALY scale from 0 (death) to 1 (full health) using the Chinese value set for the SF-6D [44], with a value of <0 indicating a worse health condition than death. The SF-6D was administered at baseline; the posttreatment time point; and the 3-, 6-, and 12-month follow-ups. Subsequently, QALYs were derived by multiplying time in a particular health state by its health utility using the area under the curve method [45], with 1 QALY indicating 1 life year in perfect health. The adjusted QALY for CUA was calculated using the regression model mentioned in later sections.

Cost Analyses

Costs were calculated from the societal and health care system perspectives using individual-level data collected for the following periods: 3 months before randomization, from randomization to the posttreatment time point (3 months), from the posttreatment time point to the 3-month follow-up, from the 3- to the 6-month follow-up, and from the 6- to the 12-month follow-up. In total, 3 main categories of costs were identified: intervention costs, health care costs, and societal costs attributed to lost or declined productivity. The health care system perspective accounts for all the costs directly associated with health care services [46], and thus, intervention costs and health care costs are included within this perspective. As the societal perspective is the broadest perspective comprising all health care–related costs and productivity costs [46], all 3 costs are considered within this perspective.

XSL•FO
RenderX

Intervention costs refer to the expenses associated with the intervention administered. For the ICBT group, the intervention costs comprised the costs of development and maintenance of Morning Mood, implementation support, and personnel training. On the basis of the practical operation, we estimated that the costs were CN ¥545 (US $74.31) per participant. For the waitlist control group, the intervention costs were CN ¥0 per participant.

Health care costs, also called direct medical costs, comprise expenses generated from health care resource use, including but not limited to diagnostic tests, examinations, medications, and other treatments. In this study, direct medical costs were retrieved from the Shenzhen Hospital information system at the individual patient level and calculated using the bottom-up method [47].

Societal costs, also known as indirect costs, are the quantified economic value of lost and declined productivity attributed to MDD. The Sheehan Disability Scale (SDS) was applied to measure days of lost and declined productivity through the following items: “On how many days in the last week did your symptoms cause you to miss school or work or leave you unable to carry out your normal daily responsibilities” and “On how many days in the last week did you feel so impaired by your symptoms, that even though you went to school or work, your productivity was reduced” [48]. Previous research has converted absenteeism and presenteeism into monetary units in populations with mental health conditions [49]. Thus, this study also applied the human capital approach for the estimation [50]. Specifically, loss of productivity was calculated using days of lost productivity multiplied by the monthly average wage in Shenzhen [51]. For declined productivity, a weighting coefficient of 42.98% was used for the adjustment [49].

All costs were collected alongside the RCT and were measured in 2022 Chinese yuan. The baseline costs were identified as costs incurred during the 3 months before the individuals’ time of enrollment and were used to control for baseline differences.

Statistical Analysis

Baseline characteristics and outcome measures were described by group and by phase with continuous variables using mean, SD, and frequency for categorical variables. Mann-Whitney *U* tests, 2-tailed *t* tests, and chi-square tests were used to investigate the differences in baseline characteristics and outcomes between the 2 arms. Estimates of the difference between groups in health utility scores were derived at each time point with 95% CIs and *P* values. Statistical significance was determined when the *P* value was <.05.

Missing data were imputed using the multiple imputation by chained equations package in R (version 4.2.2; R Foundation for Statistical Computing) under the assumption of data being missing at random [52,53]. There is compelling evidence that multiple imputation and bootstrapping are robust resampling approaches for dealing with skewed and missing data in cost-effectiveness trials [54]. The predictive mean matching method in multiple imputation by chained equations was adopted for imputation, which matched the missing value to the observed value with the closest predicted mean, with the advantage of preserving data distribution and avoiding values lying outside the bounds for each variable [55]. A total of 5 imputed datasets were generated and used for analysis as previous studies have suggested that 3 to 5 imputed datasets were sufficient to provide adequate estimates [56]. The rules by Rubin [57] were used to pool the results, involving the estimates, SEs, and 95% CIs. Health utility and costs at the posttreatment time point and every follow-up time point were imputed based on age, gender, antidepressant use, baseline values (utility and cost), and values from the previous or next measurement time point [26].

Economic Evaluation

The CUA was conducted from the perspectives of society and the health care system for 8 weeks, 3 months, 6 months, and 12 months, not accounting for the discount rate for cost and QALYs. Both the complete-case (CC) and intention-to-treat (ITT) samples were considered. Seemingly unrelated regression was conducted to estimate the total costs, QALYs, and corresponding differences (incremental differences in costs and QALYs between the 2 arms) while adjusting for baseline values and other covariates and considering the correlation between costs and QALYs, with the coefficients of the treatment group (reference: waitlist control group) in 2 separate regression models representing the cost and effect differences [58]. Total costs were adjusted for baseline costs, age, gender, and antidepressant use [59]. QALYs were adjusted for baseline utility, age, gender, and antidepressant use [59,60]. Incremental cost utility ratios (ICURs) were calculated using the incremental difference in costs divided by the incremental difference in QALYs between the 2 arms. Nonparametric bootstrapping methods were performed to account for the uncertainty surrounding the estimates of cost, QALYs, and ICURs. Cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs) were applied to demonstrate the uncertainty of the results [61]. In this study, the WTP threshold (ie, the amount that people are willing to pay for 1 year lived in full health state) was set at 1.5 times the per-capita gross domestic product per QALY [62], with a range from 1 to 3 times the per-capita gross domestic product (CN ¥80,976 to CN ¥242,928 [US $11,041.60 to $33,124.90]) according to the Chinese guideline for health economics [63]. Standard decision rules were used to determine whether the ICBT was cost-effective compared to the waitlist control [58].

Base-case cost-effectiveness analysis was conducted using data collected at the posttreatment time point (ie, 8 weeks). Sensitivity analyses were implemented through a series of scenario analyses using data collected at the follow-up time points. The aforementioned analyses also assessed whether the results remained consistent after adjusting for baseline differences and adopting different analysis datasets. To explore the long-term cost utility of ICBT, extrapolation analyses across various time horizons (3-, 6-, and 12-month follow-ups) for the waitlist control group were conducted as no follow-up data were available for that group. For each time point, data extrapolation for the control group was applied based on the predicted model formulated from cost and utility data in the ICBT group. Detailed operations and practical rationality have been clearly described in a published article [64]. All analyses were conducted using R (version 4.2.2) and Stata (version 17.0; StataCorp).

XSL•FO
RenderX

Results

Overview

Between August 2021 and December 2022, a total of 291 persons with MDD were recruited. Of these 291 participants, 33 (11.3%) either refused to participate or lost contact, 13 (4.5%) did not meet the inclusion criteria, and 1 (0.3%) met the exclusion criteria. Finally, a total of 244 eligible participants with baseline information and who provided informed consent were enrolled, with 122 (50%) participants allocated to each group. In the ICBT group, a downward trend was observed in completion rates, with the highest rate at 8 weeks (93/122, 76.2%) and the lowest rate at the 12-month follow-up (86/122, 70.5%). For the waitlist control group, a high completion rate was reported, with 94.3% (115/122) of the participants completing the posttreatment assessment. A detailed account of the study population and participant flow is shown in Figure 1.

Figure 1. Flowchart of participants. ICBT: internet-based cognitive behavioral therapy.

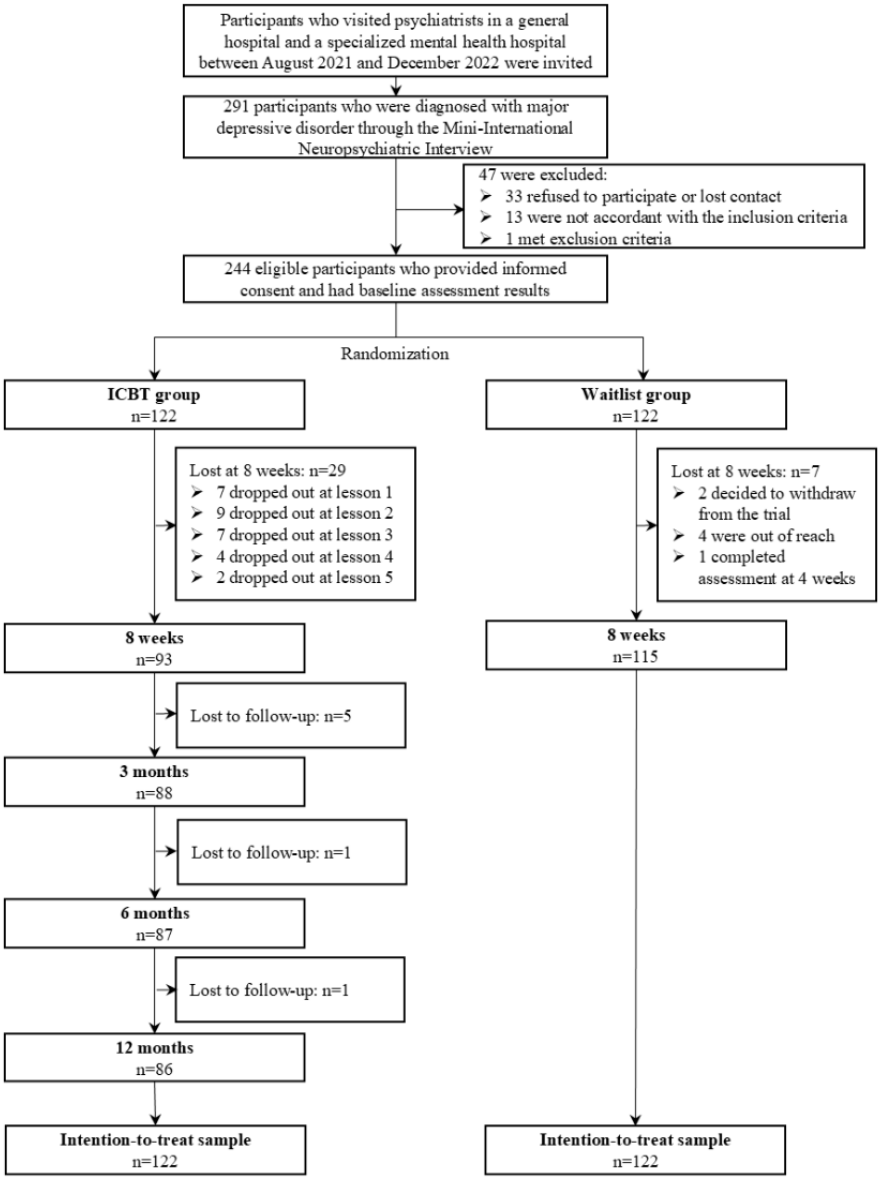

Baseline Characteristics

Table 1 shows the baseline characteristics of the participants by treatment group. The mean age was 28.3 (SD 7.0) years, and the female-to-male participant ratio was 3:1. Most of the sample (224/240, 93.3%) was of Han ethnicity. Most of the participants were well educated, with 82.2% (199/242) having received education above the undergraduate level. A total of 69.8% (169/242) of the participants were employed. More than half of the participants were experiencing their first episode of MDD (144/239, 60.3%) and were not receiving antidepressant treatment (127/244, 52%). The mean scores on the PHQ-9, Generalized Anxiety Disorder–7, SDS, and Kessler Psychological Distress Scale at baseline were 13.9 (SD 5.2), 10.5 (SD 4.7), 17.7 (SD 7.0), and 29.7 (SD 9.0), respectively. There were substantial differences in baseline characteristics between the ICBT and waitlist control groups except for the SDS scale scores. The mean SDS scores in the ICBT group were significantly higher than those in the control group (18.8, SD 7.0 vs 16.7, SD 6.8; *P*=.02), indicating that participants in the ICBT group had worse social functioning.

XSL•FO
RenderX

Table 1. Baseline characteristics of the participants (N=244).

Variable	Total	ICBT[a] (n=122)	Waitlist (n=122)	*P* value[b]
Age (y)[c], mean (SD)	28.3 (7.0)	28.6 (6.8)	28.0 (7.3)	.50
Sex[c] , n (%)				.14
Female	183 (75)	86 (70.5)	97 (79.5)	
Male	61 (25)	36 (29.5)	25 (20.5)	
Ethnicity, n (%)				.82
Han	224 (93.3)[d]	112 (94.1)[e]	112 (92.6)[f]	
Others	16 (6.7)[d]	7 (5.9)[e]	9 (7.4)[f]	
Educational level, n (%)				.18
High school or lower	43 (17.8)[g]	27 (22.3)[f]	16 (13.2)[f]	
Undergraduate	168 (69.4)[g]	80 (66.1)[f]	88 (72.7)[f]	
Master's degree or higher	31 (12.8)[g]	14 (11.6)[f]	17 (14)[f]	
Current employment status, n (%)				.11
Student	40 (16.5)[g]	14 (11.6)[f]	26 (21.5)[f]	
Employed	169 (69.8)[g]	90 (74.4)[f]	79 (65.3)[f]	
Unemployed	33 (13.6)[g]	17 (14)[f]	16 (13.2)[f]	
Recurrent episodes, n (%)				.49
Yes	95 (39.7)[h]	50 (42.4)[i]	45 (37.2)[f]	
No	144 (60.3)[h]	68 (57.6)[i]	76 (62.8)[f]	
Current drinking status, n (%)				.16
Yes	204 (84.6)[j]	106 (88.3)[k]	98 (81)[f]	
No	37 (15.4)[j]	14 (11.7)[k]	23 (19)[f]	
Current smoking status, n (%)				.43
Yes	101 (42.1)[d]	47 (39.2)[k]	54 (45)[k]	
No	139 (57.9)[d]	73 (60.8)[k]	66 (55)[k]	
Antidepressant use[c], n (%)				.44
Yes	117 (48)	55 (45.1)	62 (50.8)	
No	127 (52)	67 (54.9)	60 (49.2)	
PHQ-9[l] score at baseline[c], mean (SD)	13.9 (5.2)	14.2 (5.5)	13.6 (4.9)	.39
GAD-7[m] score at baseline[c], mean (SD)	10.5 (4.7)	10.8 (4.8)	10.2 (4.6)	.30
SDS[n] score at baseline[c], mean (SD)	17.7 (7.0)	18.8 (7.0)	16.7 (6.8)	.02
K-10[o] score at baseline[c], mean (SD)	29.7 (9.0)	29.3 (10.0)	30.1 (8.0)	.47

[a]ICBT: internet-based cognitive behavioral therapy.

[b]Baseline characteristics were compared between the 2 groups using 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables.

[c]Variables with no missing data.

[d]n=240.

[e]n=119.

[f]n=121.

[g]n=242.

[h]n=239.

XSL•FO
RenderX

[i]n=118.

[j]n=241.

[k]n=120.

[l]PHQ-9: Patient Health Questionnaire–9.

[m]GAD-7: Generalized Anxiety Disorder–7.

[n]SDS: Sheehan Disability Scale.

[o]K-10: Kessler Psychological Distress Scale.

Health Outcome

Table 2 provides the completion rates of the SF-6D scale and the converted health utility scores. At baseline, the mean utility of the ICBT group was lower than that of the control group (0.5190, SD 0.2286 vs 0.5625, SD 0.1949); however, the difference was not statistically significant (*P*=.11). After the 8-week intervention, increases in the average utility scores were observed in both groups (0.6002, SD 0.2285 vs 0.6012, SD 0.2149; *P*=.98), with completion rates of 73% (89/122) and 94.3% (115/122), respectively. The ICBT group demonstrated continuous improvement in health utility scores over the 6-month follow-up period, with a slight decline at 12 months. The highest health utility score of 0.6240 in the ICBT group was observed at the 6-month follow-up.

Multimedia Appendix 2 shows the health utility scores and subsequent QALY estimations in the ITT sample. In the ICBT group, the mean QALY gains were 0.1414 (SE 0.0021), 0.2952 (SE 0.0041), 0.4539 (SE 0.0059), and 0.7765 (SE 0.0096) at the posttreatment time point (ie, 8 weeks) and the 3-, 6-, and 12-month follow-up time points, respectively. In the control group, the average QALYs at the same time points were 0.1463 (SE 0.0019), 0.3000 (SE 0.0038), 0.4598 (SE 0.0053), and 0.7839 (SE 0.0080), respectively. There was no significant difference in QALYs between groups at any time point. Both the CC and ITT analyses suggested sustained improvement in health utility in the ICBT group over 12 months and in the waitlist control group over 8 weeks.

Table 2. Health utility measured using the Short-Form Six-Dimension health index by time point in the complete-case analysis.

Health utility at different time points	ICBT[a] (n=122)				Waitlist (n=122)				Mean difference between the 2 groups (*P* value)[b]
	Time point (T_i)		Difference between time points ($T_i - T_{i-1}$)		Time point (T_i)		Difference between time points ($T_i - T_{i-1}$)		
	Participants, n (%)	Scores, mean (SD)	Participants, n (%)	Scores, mean (SD)	Participants, n (%)	Scores, mean (SD)	Participants, n (%)	Scores, mean (SD)	
T_0[c]	122 (100)	0.5190 (0.2286)	—[d]	—	122 (100)	0.5625 (0.1949)	—	—	−0.0435 (.11)
T_1[e]	89 (73)	0.6002 (0.2285)	89 (73)	0.0892 (0.1843)	115 (94.3)	0.6012 (0.2149)	115 (94.3)	0.0420 (0.1602)	−0.0010 (.98)
T_2[f]	87 (71.3)	0.6057 (0.2416)	83 (68)	0.0018 (0.1509)	—	—	—	—	—
T_3[g]	85 (69.7)	0.6240 (0.2198)	84 (68.9)	0.0172 (0.2074)	—	—	—	—	—
T_4[h]	82 (67.2)	0.6146 (0.2590)	81 (66.4)	−0.0088 (0.2012)	—	—	—	—	—

[a]ICBT: internet-based cognitive behavioral therapy.

[b]Health utility at different time points was compared between the 2 groups using *t* tests.

[c]T_0: pretreatment time point.

[d]No applicable data.

[e]T_1: posttreatment time point (ie, 8 weeks).

[f]T_2: 3-month follow-up.

[g]T_3: 6-month follow-up.

[h]T_4: 12-month follow-up.

Costs

Table 3 shows the costs of different components and aggregated costs in the CC analysis by trial arm. At all time points, indirect costs were the major contributor to the total costs. The cost of antidepressants constituted the primary component of direct medical costs. At the pretreatment time point (ie, baseline), the average direct medical costs in the ICBT and control groups

XSL•FO
RenderX

were CN ¥6588.70 (US $898.42; SD CN ¥14,093.70 [US $1921.77]) and CN ¥5861.70 (US $799.28; SD CN ¥10,198.60 [US $1390.65]), respectively. Indirect costs approximately amounted to CN ¥22,225.90 (US $3030.66; SD CN ¥16,225.30 [US $2212.43]) in the ICBT group and CN ¥18,853.10 (US $2570.75; SD CN ¥16,197.50 [US $2208.64]) in the control group. However, both lacked statistical significance. After the 8-week intervention, both the direct medical costs and indirect costs decreased in both groups. The ICBT group reported statistically significantly higher indirect costs than those in the waitlist control group (P=.01).

Multimedia Appendix 3 shows the results of the cost analysis in the ITT sample. Both groups revealed a decreasing trend in direct medical costs and indirect costs during the posttreatment and follow-up periods. For patients in the ICBT group, the declining trend was sustained throughout the follow-up period.

Table 3. Average costs per participant (in Chinese yuan) by trial arm for the complete-case analysis at the pre- and posttreatment time points.

Time point	ICBT[b] (n=122)		Waitlist (n=122)		*P* value[a]
	Resource users, n (%)	Costs (CN ¥), mean (SD)	Resource users, n (%)	Costs (CN ¥), mean (SD)	
Pretreatment					
Drug costs	64 (52.5)	5881.50 (US $801.98; 13,772.60 [US $1877.99])	65 (53.3)	4858.60 (US $662.50; 9619.10 [US $1311.63])	.68
Treatment costs	12 (9.8)	179.60 (US $24.49; 674.40 [US $91.96])	9 (7.4)	246.1 (US $33.56; 1365 [US $186.13])	.50
Diagnostic and examination costs	64 (52.5)	527.60 (US $71.94; 746.60 [US $101.8])	70 (57.4)	757 (US $103.22; 907.8 [US $123.78])	.08
Direct medical costs	79 (64.8)	6588.70 (US $898.42; 14,093.70 [US $1921.77])	82 (67.2)	5861.7 (US $799.28; 10,198.6 [US $1390.65])	.41
Indirect costs	109 (89.3)	22,225.90 (US $3030.66; 16,225.30 [US $2212.43])	107 (87.7)	18,853.1 (US $2570.75; 16,197.5 [US $2208.64])	.10
Posttreatment					
Drug costs	16 (13.1)	2281.80 (US $311.14; 11,545.30 [US $1574.28])	23 (18.9)	1336.8 (US $182.28; 4567 [US $622.74])	.24
Treatment costs	2 (1.6)	92.90 (US $12.67; 940.50 [US $128.24])	2 (1.6)	21.5 (US $2.93; 175 [US $23.86])	>.99
Diagnostic and examination costs	10 (8.2)	46.60 (US $6.35; 186.80 [US $25.47])	14 (11.5)	58.4 (US $7.96; 201 [US $27.41])	.40
Direct medical costs	18 (14.8)	2421.20 (US $330.15; 12,416.40 [US $1693.06])	24 (19.7)	1416.7 (US $193.18; 4697.8 [US $640.58])	.31
Indirect costs	109 (89.3)	14,817.30 (US $2020.44; 10,816.90 [US $1474.96])	87 (71.3)	11,115.6 (US $1515.69; 11,147.5 [US $1520.04])	.01

[a]Costs at different time points were compared between the 2 groups using Mann-Whitney *U* tests.

[b]ICBT: internet-based cognitive behavioral therapy.

Economic Evaluations

The adjusted mean costs and QALYs (bootstrapped SEs and 95% CIs) by seemingly unrelated regression in base-case analysis and scenario analyses for this health economic evaluation from the perspectives of society and the health care system in the CC and ITT samples are shown in Multimedia Appendices 4 and 5. In most scenarios, the adjusted costs and QALYs were not statistically significantly different between the 2 groups. The corresponding results of the incremental cost-effectiveness analysis are shown in Table 4.

XSL•FO
RenderX

Table 4. Summary of incremental cost-effectiveness (CE) results for the baseline-adjusted (BA) complete-case, intention-to-treat, and scenario analyses from the societal and health care system perspectives.[a]

Analysis	Difference in mean costs[b] (bSE[c])	Difference in mean QALYs[b,d] (bSE)	ICUR[e]	ICURs by CE plane quadrant (%)						Probability of CE[f] (%)	
				Southeast quadrant	Southwest quadrant	Northeast quadrant	Northwest quadrant	South quadrant	East quadrant	≤1 × GDP[g] per capita[h]	≥3 × GDP per capita[i]
Societal perspective											
Complete-case analysis: 8 weeks											
BA cost/BA QALY	CN ¥–149.65 (US $−20.41; CN ¥1645.32 [US $224.35])	0.0042 (0.0028)	CN–¥35,456.18 (US $–4834.70)	50.7	3.58	42.25	3.47	54.28	92.95	62.28	74.63
Intention-to-treat analysis: 8 weeks											
BA cost/BA QALY	CN ¥–899.45 (US $−122.65; CN ¥2064.49 [US $281.51])	0.0046 (0.0028)	CN ¥–194,720.38 (US $–26,551.50)	65.82	3.28	29.84	1.06	69.1	95.66	75.93	86.24
Intention-to-treat analysis: 3 months											
BA cost/BA QALY	CN ¥–1278.93 (US $−174.39; CN ¥3378.79 [US $460.72])	0.0126 (0.0079)	CN ¥–101,283.17 (US $–13,810.67)	63.36	3.03	31.88	1.73	66.39	95.24	76.85	88.46
Intention-to-treat analysis: 6 months											
BA cost/BA QALY	CN ¥–1761.24 (US $−244.25; CN ¥4952.93 [US $675.37])	0.0179 (0.0116)	CN ¥–98,146.80 (US $–13,383.00)	62.53	3.17	31.78	2.52	65.7	94.31	75.32	86.32
Intention-to-treat analysis: 12 months											
BA cost/BA QALY	CN ¥–2698.94 (US $−368.02; CN ¥7396.60 [US $1008.58])	0.0278 (0.0211)	CN ¥–97,064.62 (US $–13,235.44)	61.13	4.81	30.42	3.64	65.94	91.55	75.62	85.48
Health care system perspective											
Complete-case analysis: 8 weeks											

XSL•FO
RenderX

Analysis	Difference in mean costs[b] (bSE[c])	Difference in mean QALYs[b,d] (bSE)	ICUR[e]	ICURs by CE plane quadrant (%)						Probability of CE[f] (%)	
				Southeast quadrant	Southwest quadrant	Northeast quadrant	Northwest quadrant	South quadrant	East quadrant	≤1 × GDP[g] per capita[h]	≥3 × GDP per capita[i]
BA cost/BA QALY	CN ¥507.13 (US $69.15; CN ¥161.95 [US $22.08])	0.0043 (0.0028)	CN ¥118,884.70 (US $16,210.76)	0.11	0	93.05	6.84	0.11	93.16	28.6	76.9
Intention-to-treat analysis: 8 weeks											
BA cost/BA QALY	CN ¥201.90 (US $27.53; CN ¥1568.36 [US $213.86])	0.0041 (0.0027)	CN ¥49,700.33 (US $6776.99)	41.75	3.02	52.22	3	44.78	93.97	54.4	71.03
Intention-to-treat analysis: 3 months											
BA cost/BA QALY	CN ¥136.60 (US $18.63; CN ¥1963.81 [US $267.78])	0.0110 (0.0073)	CN ¥12,400.87 (US $1690.95)	44.1	3.42	49.6	2.88	47.52	93.7	66.97	86.28
Intention-to-treat analysis: 6 months											
BA cost/BA QALY	CN ¥112.37 (US $15.32; CN ¥2060.10 [US $280.91])	0.0153 (0.0108)	CN ¥7332.67 (US $999.86)	44.39	3.95	48.16	3.5	48.34	92.55	72.33	88.34
Intention-to-treat analysis: 12 months											
BA cost/BA QALY	CN ¥68.40 (US $9.33; CN ¥2210.68 [US $301.44])	0.0228 (0.0189)	CN ¥3001.31 (US $409.25)	43.43	6.1	45.56	4.91	49.53	88.99	77.61	87.58

[a]The scenario analysis represents a scenario whereby the waitlist control participants' health-related quality of life and care costs after 8 weeks up to 12 months followed the same trend observed in the intervention group; therefore, the difference in costs and quality-adjusted life years is based on the observed values for the intervention group at each time point, but predicted values were obtained using regression analysis for the waitlist control group, the regression model for which is described in the Methods section.

[b]Difference in mean values between trial arms.

[c]bSE: bootstrapped SE.

[d]QALY: quality-adjusted life year.

[e]ICUR: incremental cost utility ratio.

[f]Probability of being cost-effective.

[g]GDP: gross domestic product.

[h]Willingness to pay 1 time the per-capita GDP.

[i]Willingness to pay 3 times the per-capita GDP.

From the perspective of society, the adjusted incremental costs (CN ¥−899.45 [US $–122.65]) and QALYs (0.0046) resulted in an ICUR of CN ¥−194,720.38 (US $–26,551.50) per QALY in the base-case analysis, which was much lower than the setting WTP threshold. This dominant ICUR indicates that ICBT is more cost-effective compared to the waitlist control from the societal perspective. The cost-effectiveness plane and CEAC, generated based on 5000 bootstrapping replications, further illustrated the robustness of the point estimates. Specifically, 29.84% of replications fell within the northeast quadrant (more

XSL•FO
RenderX

expensive and more effective), and 65.82% fell within the southeast quadrant (less expensive and more effective; Figure 2). The probability of ICBT being cost-effective at the setting WTP ranged from 75.93% to 86.24% and showed an upward trend with the increase in WTP (Figure 3).

Similarly, analysis from the perspective of the health care system indicated that ICBT was more cost-effective than the waitlist control, with an ICUR of CN ¥49,700.33 (US $6776.99) per QALY and probabilities of ICBT being cost-effective ranging from 54.4% to 71.03%. Notwithstanding the aforementioned observation, no cost-saving effect for the health care system was observed.

Figure 2. Cost-effectiveness planes showing the difference in baseline-adjusted incremental quality-adjusted life years (QALYs; x-axis) and incremental costs (y-axis) of 5000 bootstrapping samples between trial arms across 8 weeks and 3, 6, and 12 months. (A,C,E,G): society; (B,D,F,H): health care system. CNY: Chinese yuan; GDP: gross domestic product; WTP: willingness to pay.

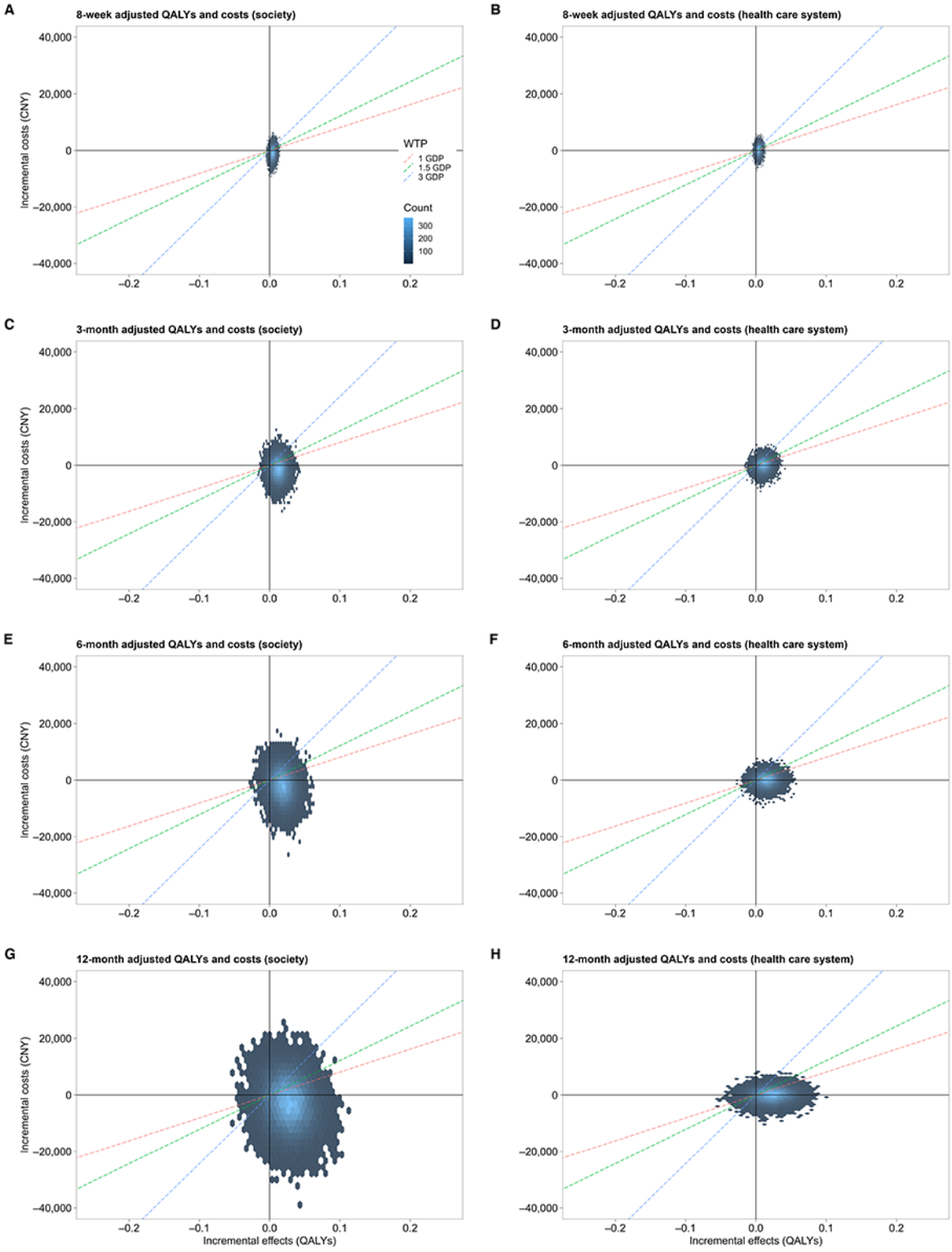

XSL•FO
RenderX

Figure 3. Cost-effectiveness acceptability curves representing the probability of cost-effectiveness of internet-based cognitive behavioral therapy (ICBT) relative to the waitlist control over 8 weeks and 3, 6, and 12 months (baseline-adjusted). (A) ICBT versus waitlist (society); (B) ICBT versus waitlist (health care system). CNY: Chinese yuan; WTP: willingness to pay.

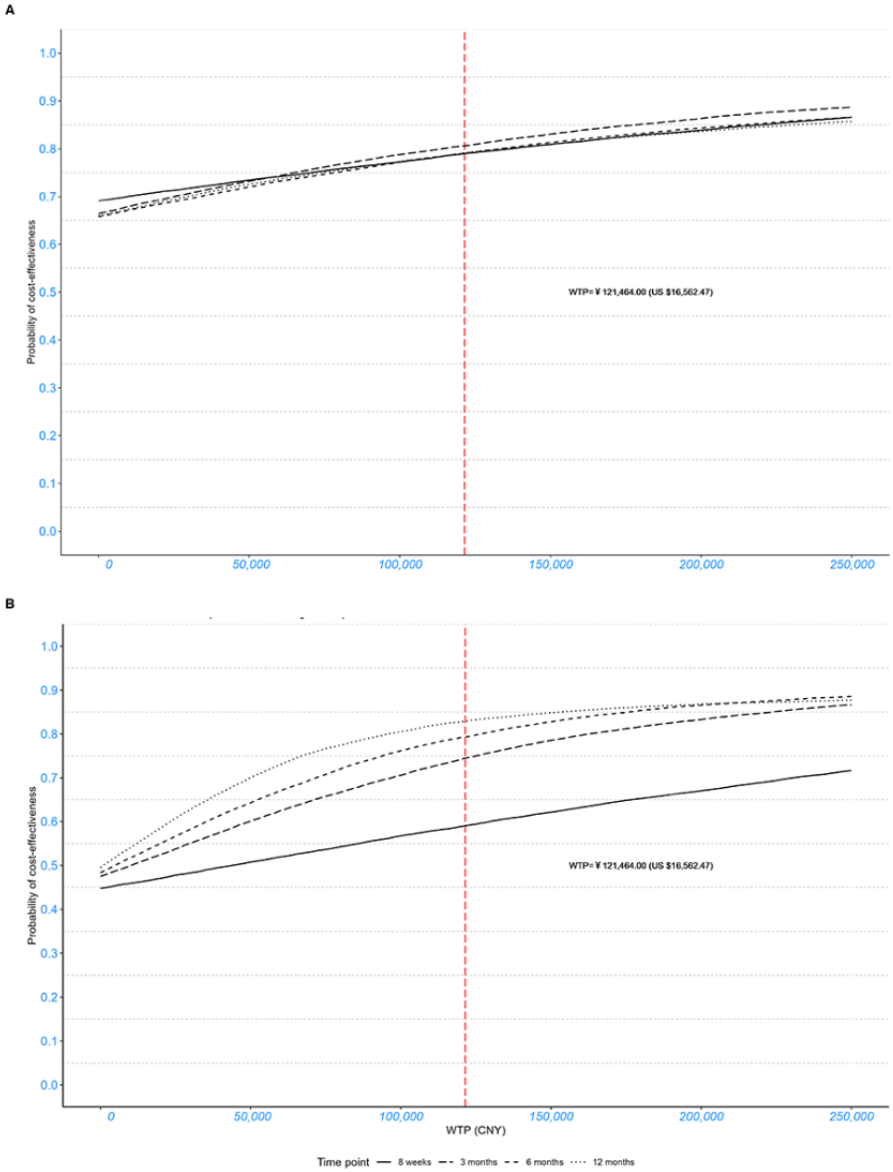

Sensitivity Analysis

Sensitivity analyses were conducted by varying follow-up duration (3-, 6-, and 12-month follow-ups) through subsequent scenario analyses. Figures 2 and 3 show the cost-effectiveness planes and CEACs, respectively, at different time points from the societal and health care system perspectives for comparison. From the societal perspective, the results similarly indicated lower costs but higher QALYs for participants in the ICBT group, with an 88.46% probability of ICBT being cost-effective, suggesting the long-term health and economic value of ICBT. In addition, from the health care system perspective, no cost-saving effect was observed. Higher direct medical costs and QALYs were observed in the ICBT group, with the probability of ICBT being cost-effective reaching 88.34%.

When the baseline values were not adjusted, the ICBT and waitlist control groups showed no difference in terms of cost utility at all time points (Figures 4 and 5). The economic evaluation is reported following the Consolidated Health Economic Evaluation Reporting Standards 2022 checklist, which is presented in Multimedia Appendix 6.

XSL•FO
RenderX

Figure 4. Cost-effectiveness planes showing the difference in unadjusted incremental quality-adjusted life years (QALYs; x-axis) and incremental costs (y-axis) of 5000 bootstrapping samples between trial arms across 8 weeks and 3, 6, and 12 months. (A,C,E,G): society; (B,D,F,H): health care system. CNY: Chinese yuan; GDP: gross domestic product; WTP: willingness to pay.

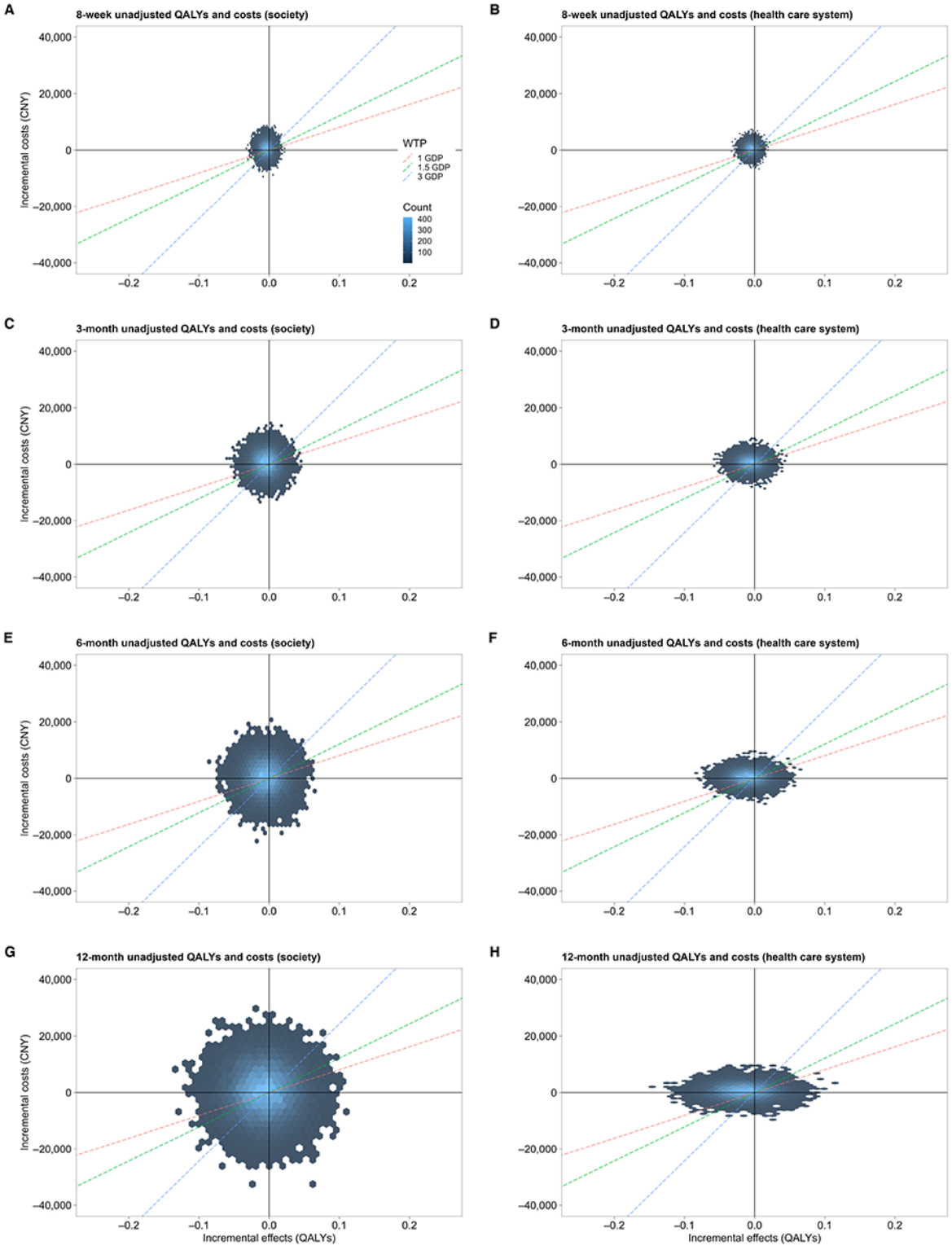

XSL•FO
RenderX

Figure 5. Cost-effectiveness acceptability curves representing the probability of cost-effectiveness of internet-based cognitive behavioral therapy (ICBT) relative to the waitlist control over 8 weeks and 3, 6, and 12 months (unadjusted). A) ICBT versus waitlist (society); (B) ICBT versus waitlist (health care system). CNY: Chinese yuan; WTP: willingness to pay.

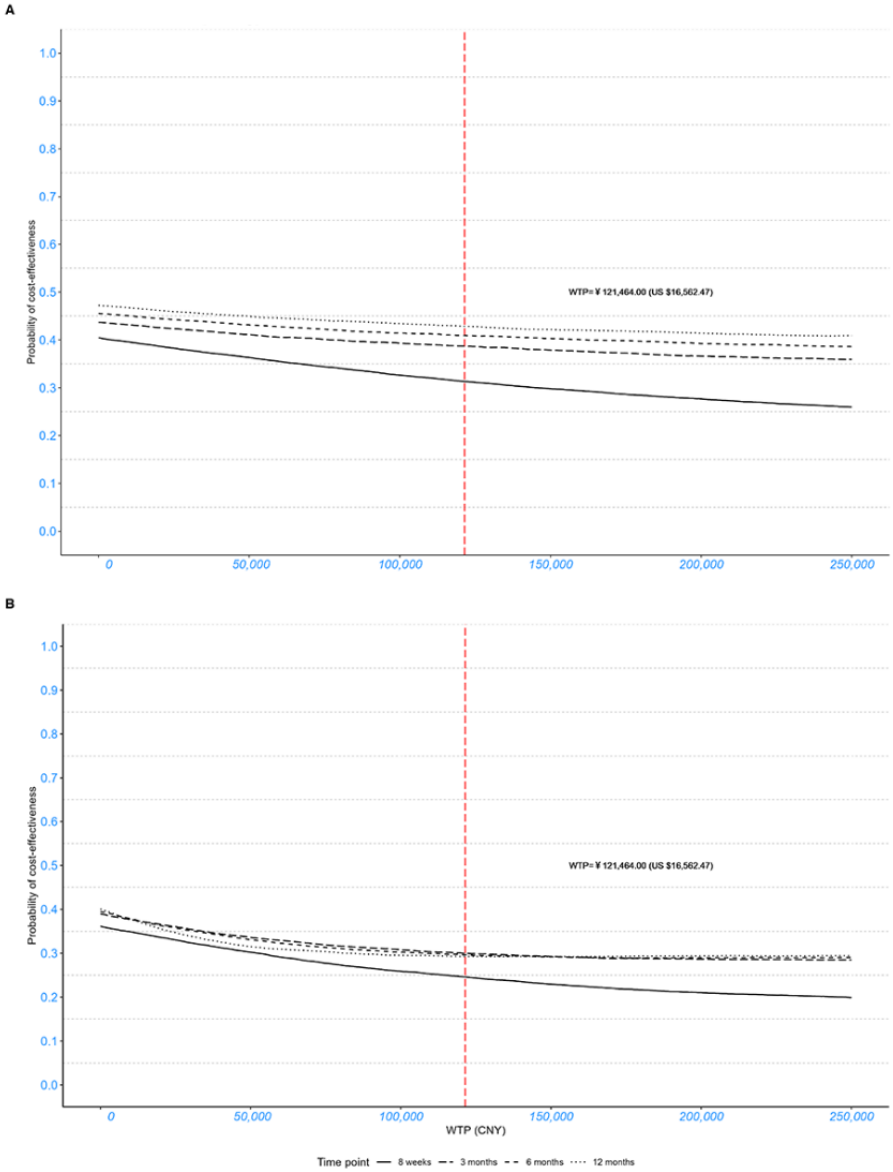

Discussion

Principal Findings

This study aimed to evaluate the short- and long-term cost-effectiveness of unguided ICBT compared to a waitlist control for MDD in China. The base-case findings of this CUA showed that the ICBT intervention was more cost-effective for persons with MDD compared to the waitlist control over 8 weeks from the perspectives of society and the health care system. Sensitivity analyses further confirmed the results by changing the data analysis set (ie, CC and ITT samples), study period (ie, 3-, 6-, and 12-month follow-ups), and whether baseline conditions were adjusted for. To the best of our knowledge, this is the first cost-effectiveness study on unguided ICBT for MDD in China, contributing to the research gap in low- and middle-income countries and providing economic evidence for policy makers to rationally allocate limited health resources.

Previous Studies

From the perspective of society, the incremental cost per QALY gain was CN ¥−194,720.38 (US $–26,551.50), which implied that the ICBT intervention was associated with higher improvement in HRQoL at lower costs, with the probability of ICBT being cost-effective ranging from 75.32% to 88.46% at the setting WTP thresholds. These findings were consistent with those of most previous research conducted in high-income countries demonstrating the cost utility of unguided ICBT for MDD [22-25,29]. Among them, one study also suggested the clinical and cost-saving effect of unguided ICBT [29], with other studies reporting more effectiveness at higher costs [22-25]. This might be owing to the different depression severity in the recruited study participants. It has been indicated that unguided ICBT is more suitable for mild depression [17]. Individuals with milder depressive symptoms are likely to benefit more from ICBT treatment, with greater responsiveness and more substantial health improvements. As a result, their

XSL•FO
RenderX

social functions improve, enabling them to engage in daily activities and return to work and leading to decreased societal costs. However, a study with a small sample size conducted in Sweden reported that no firm conclusions could be drawn on the cost-effectiveness of unguided ICBT [30].

From the health care system perspective, the findings that unguided ICBT was more cost-effective than usual care (ICUR: CN ¥49,700.33 [US $6776.99]; range of probability of ICBT being cost-effective at the given WTP 54.4%-88.34%) were contrary to those of most previous cost-effectiveness studies [26-28,30]. The following different settings might explain the discrepancy: the different instruments used for the measurement of HRQoL (EQ-5D in previous studies and SF-6D in this study), varied baseline symptom severity (PHQ-9≥10 in previous studies and PHQ-9≥5 in this study), and diverse health care use patterns in different health care systems. It is assumed that persons with mild symptoms are more likely to benefit from unguided ICBT [17]. In addition, the SF-6D is more sensitive to changes in mental health, and thus, minor improvements could be captured [65]. Previous research has reported that the conclusions on cost-effectiveness were the opposite when QALYs were estimated using the EQ-5D-3L or the SF-6D [26,28]. When using the SF-6D, ICBT appeared to dominate usual care (lower mean costs and higher QALYs), whereas in our study, ICBT exhibited greater health effects at higher costs. A possible explanation could be the variations in mental health care–seeking behaviors in different contexts. In China, individuals are less likely to approach mental health care due to stigma [66]. However, the ICBT course helps people with MDD transform their inappropriate or negative thought patterns to promote health care use [67]. Hence, the costs of health care in the ICBT group appeared higher than those of the waitlist control group in our analysis. This, in turn, explains the results from the societal perspective. It is plausible that increased patient investment in mental health care and service yields improved social functioning. Consequently, reduced days of lost or decreased productivity and indirect costs associated with cost savings were observed [22,68]. It is worth mentioning that these results should be interpreted with caution and need to be confirmed by further powered studies in different contexts.

The robustness of the results of the base-case analysis was verified through different sensitivity analyses. Scenario analyses over the 3-, 6-, and 12-month follow-ups all suggested higher likelihoods of ICBT being cost-effective, aligning with previous studies and confirming the long-term cost-effectiveness of ICBT [22-25,29]. It is important to note that adjustment for differences in baseline costs and QALYs leads to slightly different results compared to those obtained without baseline correction. At lower levels of WTP, ICBT showed an even chance of being cost-effective in both baseline-adjusted and nonadjusted analyses. However, with the increase in WTP, the opposite conclusion was drawn in the choice between the 2 treatments in terms of cost utility. One plausible explanation is that the effect of the intervention on utility was mild and the differences between the intervention and control group in costs and QALYs were relatively minor [28]. It was assumed that the baseline utility and costs were strongly correlated with the subsequent estimates, potentially leading to different outcomes when not adjusting for baseline variables [24]. It was argued that, in the economic evaluation, controlling for baseline utility and costs should be necessary [60].

Despite the positive outcomes of the economic evaluation, the results suggest that there was no statistically significant difference in HRQoL (subsequent QALYs) and cost estimates between the 2 groups at most time points, which is consistent with the findings of previous economic evaluations [23-25,29,69]. It has been reported that the between-group (ICBT vs control) effect size of HRQoL is small and a longer duration is required to capture meaningful observations [17-19,70]. In this study, the sample size was calculated based on detecting the medium effect size on depressive symptoms between the 2 groups, and thus, the statistical power might not be sufficient to detect statistical differences in HRQoL. Moreover, a systematic review aimed at addressing the effectiveness of ICBT on HRQoL also found that adults with more severe depressive symptoms were more likely to achieve greater HRQoL improvements and therapeutic guidance could further enhance the effect on HRQoL [71]. Hence, this lack of statistically significant differences in HRQoL and cost estimates between the 2 groups might be partly explained by the mild symptoms (mean PHQ-9 score 13.9, SD 5.2) experienced by the participants and the unguided ICBT intervention in this study. Another possible explanation was that the measurement tool for HRQoL lacked sensitivity to minor changes in this patient group [28,72]. Regarding costs, an economic evaluation of unguided ICBT conducted among 1013 participants with mild to moderate depression severity suggested that the total costs at 6 months after enrollment did not significantly differ between the intervention and control groups [69]. Given the high variability of costs [73], research with larger sample sizes and longer time frames may be needed to detect statistically significant differences.

There were several strengths to this study. First, this economic evaluation was carried out and reported following international economic evaluation guidelines. Moreover, the outcome data were collected alongside a pragmatic RCT. Therefore, the validity and reliability of the data were guaranteed. Second, in the recording of direct medical costs, the hospital information system was used to collect data at an individual level. Hence, recall bias was reduced to some extent compared with self-reported data. In addition, the self-developed ICBT courses were embedded in the WeChat mini program, one of the most popular platforms among the Chinese population, which led to a high retention rate. Under the support of nonspecialists, the retention rate in the waitlist control group was also maintained at a high level. Thus, the changes in clinical effectiveness and quality of life might reflect the real effect of the intervention. Fourth, the implementation of our intervention was executed by nonprofessional personnel, such as primary health care workers, which was in line with practical operations in routine care, indicating that the results had good features of scalability in the current medical environment.

Limitations

Nonetheless, this study has some limitations. The primary limitation was the short-term follow-up period in the control

XSL•FO
RenderX

group due to ethical considerations. Consequently, given the time-dependent nature of QALYs, the potential lag effect of ICBT, and the results reported in previous studies, rigorous statistical extrapolations on costs and QALYs were made for the control group with the assumption that the trends in costs and QALYs were the same in the intervention and control arms. Scenario analysis was conducted based on predicated data, suggesting that the long-term cost-effectiveness results should be interpreted with caution. It is likely that ICBT is more cost-effective in a longer follow-up period. Further studies under longer time frames extending the follow-up periods or adopting strong modeling assumptions would be useful. Second, although the sample size in this RCT had sufficient statistical power to detect clinically significant improvements in depression, it did not necessarily have the statistical power to detect differences in cost and utility due to the high variability of costs and the small effect of ICBT on HRQoL. Replications with large sample sizes are needed to explore differences in costs and cost utility. In addition, identifying individuals who benefit from the ICBT course and those who are at risk of not responding and adopting tailored interventions is likely to be an important way to improve the cost-effectiveness of ICBT. However, these results need to be confirmed by more research in different subsamples. Third, as direct nonmedical cost data were unavailable, only direct medical and indirect costs were included in this analysis. The aforementioned limitation might potentially underestimate the overall total economic burden and affect the conclusions. In future studies of economic evaluations on ICBT, a comprehensive assessment of the economic burden of participants will be of significance. Fourth, in this study, female participants accounted for 75% (183/244) of the total participants, and most participants (199/242, 82.2%) were well educated, which affects the extrapolation at the population level. Future interested researchers may need to take account of the representativeness of the samples. Fifth, this study only included participants who had access to digital devices and, thus, might have excluded socioeconomically marginal populations and could raise ethical issues [74,75]. It is not clear whether addressing digital divides and ensuring equitable access (eg, providing loaner electronic devices and internet connection) would influence the extant results. This remains open for further exploration. Sixth, despite the existing evidence on effectiveness and cost-effectiveness, the use of personal data, including data privacy and participant autonomy, should be considered before the massive scale-up of ICBT [76,77].

Conclusions

In comparison with the waitlist control, unguided ICBT is more cost-effective for MDD from both the health care system and societal perspectives in China. This intervention not only helps patients with MDD improve clinically but also generates societal savings. These findings suggest that unguided ICBT has broad application prospects in low-resource countries such as China and can serve to allow scalable resource access to MDD care for low- and middle-income countries.

Acknowledgments

The authors thank all the participants of the study. The Depression Cohort in China is conducted by Sun Yat-sen University and the Shenzhen Nanshan Center for Chronic Disease Control in Shenzhen, China. This work was supported by the National Natural Science Foundation of China (81761128030). No generative artificial intelligence was used in this manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors conceptualized and designed the trial. WZ, YC, Y Liu, Y Liao, H Zhang, CG, and JH conducted the trial and follow-up. WZ, HW, H Zhao, Y Li, GS, and WW collected and prepared the data. WZ and YC analyzed and interpreted the patient data and were major contributors to writing the manuscript. WW, GHL, and RSM polished and revised the manuscript. XH and LC offered administrative, technical, and material support for the implementation of the trial. XH and CL provided supervision for the entire trial. CL obtained the funding for the trial. All authors read and approved the final manuscript.

Conflicts of Interest

RSM has received research grant support from the Canadian Institutes of Health Research, the Global Alliance for Chronic Diseases, the National Natural Science Foundation of China, and the Milken Institute and speaker and consultation fees from Lundbeck, Janssen, Alkermes, Neumora Therapeutics, Boehringer Ingelheim, Sage Therapeutics, Biogen, Mitsubishi Tanabe Pharma Corporation, Purdue Pharma, Pfizer, Otsuka, Takeda, Neurocrine Biosciences, NeuraWell, Sunovion, Bausch Health, Axsome Therapeutics, Novo Nordisk, Kris Pharma, Sanofi, Eisai, Intra-Cellular Therapies, NewBridge Pharmaceuticals, Viatris, AbbVie, and Atai Life Sciences. RSM is the chief executive officer of Braxia Scientific. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

Outline of the treatment content, homework, and corresponding screenshots of the internet-based cognitive behavioral therapy program.

XSL•FO
RenderX

[DOCX File , 420 KB - jmir_v27i1e67567_app1.docx]

Multimedia Appendix 2
Means and SEs for utility measured using the Short-Form Six-Dimension health index and resulting quality-adjusted life year values by trial arm and time point in the intention-to-treat and scenario analyses.
[DOCX File , 18 KB - jmir_v27i1e67567_app2.docx]

Multimedia Appendix 3
Means and SEs for costs by trial arm and time point in the intention-to-treat and scenario analyses.
[DOCX File , 24 KB - jmir_v27i1e67567_app3.docx]

Multimedia Appendix 4
Regression-adjusted costs and quality-adjusted life year estimates by trial arm and difference in means between trial arms from the societal perspective.
[DOCX File , 21 KB - jmir_v27i1e67567_app4.docx]

Multimedia Appendix 5
Regression-adjusted costs and quality-adjusted life year estimates by trial arm and difference in means between trial arms from the health care system perspective.
[DOCX File , 21 KB - jmir_v27i1e67567_app5.docx]

Multimedia Appendix 6
Consolidated Health Economic Evaluation Reporting Standards 2022 checklist.
[DOCX File , 23 KB - jmir_v27i1e67567_app6.docx]

Multimedia Appendix 7
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 729 KB - jmir_v27i1e67567_app7.pdf]

References

1. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2133-2161 [FREE Full text] [doi: 10.1016/S0140-6736(24)00757-8] [Medline: 38642570]
2. Maj M, Stein DJ, Parker G, Zimmerman M, Fava GA, De Hert M, et al. The clinical characterization of the adult patient with depression aimed at personalization of management. World Psychiatry 2020 Oct;19(3):269-293 [FREE Full text] [doi: 10.1002/wps.20771] [Medline: 32931110]
3. McIntyre RS, Millson B, Power GS. Burden of Treatment Resistant Depression (TRD) in patients with major depressive disorder in Ontario using Institute for Clinical Evaluative Sciences (ICES) databases: economic burden and healthcare resource utilization. J Affect Disord 2020 Dec 01;277:30-38 [FREE Full text] [doi: 10.1016/j.jad.2020.07.045] [Medline: 32791390]
4. Allen D, Hines EW, Pazdernik V, Konecny LT, Breitenbach E. Four-year review of presenteeism data among employees of a large United States health care system: a retrospective prevalence study. Hum Resour Health 2018 Nov 09;16(1):59 [FREE Full text] [doi: 10.1186/s12960-018-0321-9] [Medline: 30413168]
5. Campbell D, Green MJ, Davies N, Demou E, Howe LD, Harrison S, et al. Effects of depression on employment and social outcomes: a Mendelian randomisation study. J Epidemiol Community Health 2022 Jun;76(6):563-571 [FREE Full text] [doi: 10.1136/jech-2021-218074] [Medline: 35318279]
6. Mental health and substance use. World Health Organization. 2017. URL: https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/mental-health-in-the-workplace [accessed 2023-03-29]
7. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a Lancet-World Psychiatric Association Commission. Lancet 2022 Mar 05;399(10328):957-1022. [doi: 10.1016/S0140-6736(21)02141-3] [Medline: 35180424]
8. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020 Feb;19(1):92-107 [FREE Full text] [doi: 10.1002/wps.20701] [Medline: 31922679]

XSL•FO
RenderX

9. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Feb;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
10. Mekonen T, Chan GC, Connor JP, Hides L, Leung J. Estimating the global treatment rates for depression: a systematic review and meta-analysis. J Affect Disord 2021 Dec 01;295:1234-1242. [doi: 10.1016/j.jad.2021.09.038] [Medline: 34665135]
11. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
12. Choi I, Sharpe L, Li S, Hunt C. Acceptability of psychological treatment to Chinese- and Caucasian Australians: internet treatment reduces barriers but face-to-face care is preferred. Soc Psychiatry Psychiatr Epidemiol 2015 Jan;50(1):77-87. [doi: 10.1007/s00127-014-0921-1] [Medline: 24993290]
13. von Gaudecker JR, Agbapuonwu N, Kyololo O, Sathyaseelan M, Oruche U. Barriers and facilitators to treatment seeking behaviors for depression, epilepsy, and schizophrenia in low- and middle-income countries: a systematic review. Arch Psychiatr Nurs 2022 Dec;41:11-19. [doi: 10.1016/j.apnu.2022.07.006] [Medline: 36428038]
14. Lorenzo-Luaces L, Johns E, Keefe JR. The generalizability of randomized controlled trials of self-guided internet-based cognitive behavioral therapy for depressive symptoms: systematic review and meta-regression analysis. J Med Internet Res 2018 Nov 09;20(11):e10113 [FREE Full text] [doi: 10.2196/10113] [Medline: 30413400]
15. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
16. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
17. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
18. Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C. Efficacy of computer- and/or internet-based cognitive-behavioral guided self-management for depression in adults: a systematic review and meta-analysis of randomized controlled trials. BMC Psychiatry 2022 Nov 24;22(1):730 [FREE Full text] [doi: 10.1186/s12888-022-04325-z] [Medline: 36424570]
19. Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C. Long-term efficacy of internet-based cognitive behavioral therapy self-help programs for adults with depression: systematic review and meta-analysis of randomized controlled trials. JMIR Ment Health 2023 Aug 22;10:e46925 [FREE Full text] [doi: 10.2196/46925] [Medline: 37606990]
20. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
21. Lin Z, Cheng L, Han X, Wang H, Liao Y, Guo L, et al. The effect of internet-based cognitive behavioral therapy on major depressive disorder: randomized controlled trial. J Med Internet Res 2023 Sep 22;25:e42786 [FREE Full text] [doi: 10.2196/42786] [Medline: 37738092]
22. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:55-62. [doi: 10.1192/bjp.185.1.55] [Medline: 15231556]
23. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010 Dec 19;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
24. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318. [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
25. Geraedts AS, van Dongen JM, Kleiboer AM, Wiezer NM, van Mechelen W, Cuijpers P, et al. Economic evaluation of a web-based guided self-help intervention for employees with depressive symptoms: results of a randomized controlled trial. J Occup Environ Med 2015 Jun;57(6):666-675. [doi: 10.1097/JOM.0000000000000423] [Medline: 25741795]
26. Littlewood E, Duarte A, Hewitt C, Knowles S, Palmer SE, Walker S, REEACT Team. A randomised controlled trial of computerised cognitive behaviour therapy for the treatment of depression in primary care: the Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial. Health Technol Assess 2015 Dec;19(101):viii, xxi-viii, 171 [FREE Full text] [doi: 10.3310/hta191010] [Medline: 26685904]
27. Dixon P, Hollinghurst S, Edwards L, Thomas C, Foster A, Davies B, et al. Cost-effectiveness of telehealth for patients with depression: evidence from the Healthlines randomised controlled trial. BJPsych Open 2016 Jul 02;2(4):262-269 [FREE Full text] [doi: 10.1192/bjpo.bp.116.002907] [Medline: 27703785]

XSL•FO
RenderX

28. Duarte A, Walker S, Littlewood E, Brabyn S, Hewitt C, Gilbody S, et al. Cost-effectiveness of computerized cognitive-behavioural therapy for the treatment of depression in primary care: findings from the Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy (REEACT) trial. Psychol Med 2017 Jul;47(10):1825-1835 [FREE Full text] [doi: 10.1017/S0033291717000289] [Medline: 28228182]
29. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, Campayo JG, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
30. Holst A, Björkelund C, Metsini A, Madsen JH, Hange D, Petersson EL, et al. Cost-effectiveness analysis of internet-mediated cognitive behavioural therapy for depression in the primary care setting: results based on a controlled trial. BMJ Open 2018 Jun 14;8(6):e019716 [FREE Full text] [doi: 10.1136/bmjopen-2017-019716] [Medline: 29903785]
31. Knapp M, Wong G. Economics and mental health: the current scenario. World Psychiatry 2020 Feb;19(1):3-14 [FREE Full text] [doi: 10.1002/wps.20692] [Medline: 31922693]
32. Li M, Bai F, Yao L, Qin Y, Chen K, Xin T, et al. Economic evaluation of cognitive behavioral therapy for depression: a systematic review. Value Health 2022 Jun;25(6):1030-1041 [FREE Full text] [doi: 10.1016/j.jval.2021.11.1379] [Medline: 35422392]
33. Kählke F, Buntrock C, Smit F, Ebert DD. Systematic review of economic evaluations for internet- and mobile-based interventions for mental health problems. NPJ Digit Med 2022 Nov 23;5(1):175 [FREE Full text] [doi: 10.1038/s41746-022-00702-w] [Medline: 36424463]
34. Kolovos S, van Dongen JM, Riper H, Buntrock C, Cuijpers P, Ebert DD, et al. Cost effectiveness of guided Internet-based interventions for depression in comparison with control conditions: an individual-participant data meta-analysis. Depress Anxiety 2018 Mar;35(3):209-219 [FREE Full text] [doi: 10.1002/da.22714] [Medline: 29329486]
35. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
36. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated Health Economic Evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Clin Ther 2022 Feb;44(2):158-168. [doi: 10.1016/j.clinthera.2022.01.011] [Medline: 35168801]
37. Sun Y, Fu Z, Bo Q, Mao Z, Ma X, Wang C. The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital. BMC Psychiatry 2020 Sep 29;20(1):474 [FREE Full text] [doi: 10.1186/s12888-020-02885-6] [Medline: 32993604]
38. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34. [Medline: 9881538]
39. Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. Value Health 2008;11(2):295-303 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00233.x] [Medline: 18380642]
40. Mulhern B, Mukuria C, Barkham M, Knapp M, Byford S, Soeteman D, et al. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. Br J Psychiatry 2014 Sep;205(3):236-243. [doi: 10.1192/bjp.bp.112.122283] [Medline: 24855127]
41. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, UK: Oxford University Press; 2016.
42. Pharmacoeconomic guidelines around the world. International Society for Pharmacoeconomics and Outcomes Research. URL: https://www.ispor.org/heor-resources/more-heor-resources/pharmacoeconomic-guidelines [accessed 2024-12-06]
43. Wang L, Poder TG. A systematic review of SF-6D health state valuation studies. J Med Econ 2023;26(1):584-593 [FREE Full text] [doi: 10.1080/13696998.2023.2195753] [Medline: 36973235]
44. Wu J, Xie S, He X, Chen G, Bai G, Feng D, et al. Valuation of SF-6Dv2 health states in China using time trade-off and discrete-choice experiment with a duration dimension. Pharmacoeconomics 2021 May;39(5):521-535 [FREE Full text] [doi: 10.1007/s40273-020-00997-1] [Medline: 33598860]
45. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull 2010;96:5-21. [doi: 10.1093/bmb/ldq033] [Medline: 21037243]
46. Kim DD, Silver MC, Kunst N, Cohen JT, Ollendorf DA, Neumann PJ. Perspective and costing in cost-effectiveness analysis, 1974-2018. Pharmacoeconomics 2020 Oct;38(10):1135-1145 [FREE Full text] [doi: 10.1007/s40273-020-00942-2] [Medline: 32696192]
47. Špacírová Z, Epstein D, García-Mochón L, Rovira J, Olry de Labry Lima A, Espín J. A general framework for classifying costing methods for economic evaluation of health care. Eur J Health Econ 2020 Jun 20;21(4):529-542 [FREE Full text] [doi: 10.1007/s10198-019-01157-9] [Medline: 31960181]

XSL•FO
RenderX

48. Sheehan D, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol 1996 Jun;11 Suppl 3:89-95. [doi: 10.1097/00004850-199606003-00015] [Medline: 8923116]
49. Niu X, Dembek C, Fan Q, Mao Y, Divino V, Burch S, et al. The impact of lurasidone on functioning and indirect costs in adults with bipolar depression: a post-hoc analysis. J Med Econ 2022;25(1):152-159 [FREE Full text] [doi: 10.1080/13696998.2022.2030147] [Medline: 35037813]
50. Liljas B. How to calculate indirect costs in economic evaluations. Pharmacoeconomics 1998 Jan;13(1 Pt 1):1-7. [doi: 10.2165/00019053-199813010-00001] [Medline: 10175982]
51. Notice on the announcement of the average monthly wage of employed persons in urban units of Guangdong Province in 2022. Guangdong Human Resources and Social Security Department. URL: https://hrss.gd.gov.cn/zwgk/xxgkml/bmwj/qtwj/shbz/content/post_4210987.html [accessed 2024-03-23]
52. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clin Trials 2007;4(2):154-161 [FREE Full text] [doi: 10.1177/1740774507076914] [Medline: 17456514]
53. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
54. Brand J, van Buuren S, le Cessie S, van den Hout W. Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data. Stat Med 2019 Jan 30;38(2):210-220 [FREE Full text] [doi: 10.1002/sim.7956] [Medline: 30207407]
55. Little RJ. Missing-data adjustments in large surveys. J Bus Econ Stat 1988 Jul;6(3):287-296. [doi: 10.1080/07350015.1988.10509663]
56. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 2007 Sep;8(3):206-213. [doi: 10.1007/s11121-007-0070-9] [Medline: 17549635]
57. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ: John Wiley & Sons; 1987.
58. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th edition. Oxford, UK: Oxford University Press; 2015.
59. Ali S, Alemu FW, Owen J, Eells TD, Antle B, Lee JT, et al. Cost-effectiveness of computer-assisted cognitive behavioral therapy for depression among adults in primary care. JAMA Netw Open 2024 Nov 04;7(11):e2444599 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.44599] [Medline: 39541120]
60. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005 May;14(5):487-496. [doi: 10.1002/hec.944] [Medline: 15497198]
61. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ 2004 May;13(5):405-415 [FREE Full text] [doi: 10.1002/hec.903] [Medline: 15127421]
62. Cai D, Shi S, Jiang S, Si L, Wu J, Jiang Y. Estimation of the cost-effective threshold of a quality-adjusted life year in China based on the value of statistical life. Eur J Health Econ 2022 Jun;23(4):607-615 [FREE Full text] [doi: 10.1007/s10198-021-01384-z] [Medline: 34655364]
63. Liu GG. China Guidelines for Pharmacoeconomic Evaluations. Beijing, China: China Market Publishing Corp; 2020.
64. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3:85 [FREE Full text] [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
65. Zhang F, Yang Y, Huang T, Zhang Y, Zhao L, Li S. Is there a difference between EQ-5D and SF-6D in the clinical setting? A comparative study on the quality of life measured by AIMS2-SF, EQ-5D and SF-6D scales for osteoarthritis patients. Int J Rheum Dis 2018 Jun;21(6):1185-1192. [doi: 10.1111/1756-185X.12907] [Medline: 27463704]
66. Xu Z, Rüsch N, Huang F, Kösters M. Challenging mental health related stigma in China: systematic review and meta-analysis. I. Interventions among the general public. Psychiatry Res 2017 Sep;255:449-456. [doi: 10.1016/j.psychres.2017.01.008] [Medline: 28780127]
67. Monteiro F, Carona C, Antunes P, Canavarro MC, Fonseca A. Economic evaluation of Be a Mom, a web-based intervention to prevent postpartum depression in high-risk women alongside a randomized controlled trial. J Affect Disord 2024 Jul 15;357:163-170 [FREE Full text] [doi: 10.1016/j.jad.2024.05.013] [Medline: 38703901]
68. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010 Oct 02;197(4):297-304. [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
69. Gräfe V, Berger T, Hautzinger M, Hohagen F, Lutz W, Meyer B, et al. Health economic evaluation of a web-based intervention for depression: the EVIDENT-trial, a randomized controlled study. Health Econ Rev 2019 Jun 07;9(1):16 [FREE Full text] [doi: 10.1186/s13561-019-0233-y] [Medline: 31175475]
70. Maj A, Michalak N, Graczykowska A, Andersson G. The effect of internet-delivered cognitive behavioral therapy for depression and anxiety on quality of life: a meta-analysis of randomized controlled trials. Internet Interv 2023 Sep;33:100654 [FREE Full text] [doi: 10.1016/j.invent.2023.100654] [Medline: 37555075]
71. Fadipe MF, Aggarwal S, Johnson C, Beauchamp JE. Effectiveness of online cognitive behavioural therapy on quality of life in adults with depression: a systematic review. J Psychiatr Ment Health Nurs 2023 Oct;30(5):885-898. [doi: 10.1111/jpm.12924] [Medline: 37010913]

XSL•FO
RenderX

72. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess 2014 May;18(34):vii-xiii [FREE Full text] [doi: 10.3310/hta18340] [Medline: 24857402]
73. Gray AM, Marshall M, Lockwood A, Morris J. Problems in conducting economic evaluations alongside clinical trials. Lessons from a study of case management for people with mental disorders. Br J Psychiatry 1997 Jan;170:47-52. [doi: 10.1192/bjp.170.1.47] [Medline: 9068775]
74. Robledo Yamamoto F, Voida A, Voida S. From therapy to teletherapy: relocating mental health services online. Proc ACM Hum Comput Interact 2021 Oct 18;5(CSCW2):1-30. [doi: 10.1145/3479508]
75. Shojaei F, Osorio Torres J, Shih PC. Exploring the integration of technology in art therapy: insights from interviews with art therapists. Art Ther 2024 Sep 03:1-7. [doi: 10.1080/07421656.2024.2383826]
76. Vilaza GN, McCashin D. Is the automation of digital mental health ethical? Applying an ethical framework to chatbots for cognitive behaviour therapy. Front Digit Health 2021;3:689736 [FREE Full text] [doi: 10.3389/fdgth.2021.689736] [Medline: 34713163]
77. Shojaei F, Shojaei F, Osorio Torres J, Shih PC. Insights from art therapists on using AI-generated art in art therapy: mixed methods study. JMIR Form Res 2024 Dec 04;8:e63038 [FREE Full text] [doi: 10.2196/63038] [Medline: 39631077]

Abbreviations

CC: complete case
CEAC: cost-effectiveness acceptability curve
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CUA: cost utility analysis
HRQoL: health-related quality of life
ICBT: internet-based cognitive behavioral therapy
ICUR: incremental cost utility ratio
ITT: intention to treat
MDD: major depressive disorder
PHQ-9: Patient Health Questionnaire–9
QALY: quality-adjusted life year
RCT: randomized controlled trial
SDS: Sheehan Disability Scale
SF-6D: Short-Form Six-Dimension health index
WTP: willingness to pay

Edited by A Mavragani; submitted 15.10.24; peer-reviewed by C Oko, F Shojaei; comments to author 29.11.24; revised version received 19.12.24; accepted 26.12.24; published 19.02.25.

Please cite as:
Zhou W, Chen Y, Wu H, Zhao H, Li Y, Shi G, Wang W, Liu Y, Liao Y, Zhang H, Gao C, Hao J, Le GH, McIntyre RS, Han X, Lu C
Cost Utility Analysis of Internet-Based Cognitive Behavioral Therapy for Major Depressive Disorder: Randomized Controlled Trial
J Med Internet Res 2025;27:e67567
URL: https://www.jmir.org/2025/1/e67567
doi:10.2196/67567
PMID:

©Wenjing Zhou, Yan Chen, Herui Wu, Hao Zhao, Yanzhi Li, Guangduoji Shi, Wanxin Wang, Yifeng Liu, Yuhua Liao, Huimin Zhang, Caihong Gao, Jiejing Hao, Gia Han Le, Roger S McIntyre, Xue Han, Ciyong Lu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of the Quality of Delirium Website Content for Patient and Family Education: Cross-Sectional Study

Karla Krewulak[1], PhD; Kathryn Strayer[1], BHSc; Natalia Jaworska[1], MSc, MD; Krista Spence[1], MN; Nadine Foster[1], BScN; Scotty Kupsch[1]; Khara Sauro[2], PhD; Kirsten M Fiest[1,2,3], PhD

[1]Department of Critical Care Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Community Health Sciences and O'Brien Institute for Public Health, University of Calgary, Calgary, AB, Canada

[3]Department of Psychiatry & Hotchkiss Brain Institute, University of Calgary, Calgary, Canada

Corresponding Author:
Kirsten M Fiest, PhD
Department of Critical Care Medicine
University of Calgary
MT 04111, 3260 Hospital Drive NW
Calgary, AB,
Canada
Phone: 1 9448768
Email: kmfiest@ucalgary.ca

Abstract

Background: Patients and families who have experienced delirium may seek information about delirium online, but the quality and reliability of online delirium-related websites are unknown.

Objective: This study aimed to identify and evaluate online delirium-related websites that could be used for patient and family education.

Methods: We searched Microsoft Bing, Google, and Yahoo using the keywords “delirium” and the misspelled “delerium” to identify delirium-related websites created to inform patients, families, and members of the public about delirium. The quality of identified delirium-related website content was evaluated by 2 authors using the validated DISCERN tool and the *JAMA* (*Journal of the American Medical Association*) benchmark criteria. Readability was assessed with the Simple Measure of Gobbledygook, the Flesch Reading Ease score, and the Flesch Kincaid grade level. Each piece of website content was assessed for its delirium-related information using a checklist of items co-designed by a working group, which included patients, families, researchers, and clinicians.

Results: We identified 106 websites targeted toward patients and families, with most hospital-affiliated (21/106, 20%) from commercial websites (20/106, 19%), government-affiliated organizations (19/106, 18%), or from a foundation or advocacy group (16/106, 15%). The median time since the last content update was 3 (IQR 2-5) years. Most websites’ content (101/106, 95%) was written at a reading level higher than the recommended grade 6 level. The median DISCERN total score was 42 (IQR 33-50), with scores ranging from 20 (very poor quality) to 78 (excellent quality). The median delirium-related content score was 8 (IQR 6-9), with scores ranging from 1 to 12. Many websites lacked information on the short- and long-term outcomes of delirium as well as how common it is. The median *JAMA* benchmark score was 1 (IQR 1-3), indicating the quality of the websites’ content had poor transparency.

Conclusions: We identified high-quality websites that could be used to educate patients, families, or the public about delirium. While most delirium-related website content generally meets quality standards based on DISCERN and JAMA benchmark criteria, high scores do not always ensure patient and family-friendliness. Many of the top-rated delirium content were text-heavy and complex in layout, which could be overwhelming for users seeking clear, concise information. Future efforts should prioritize the development of websites with patients and families, considering usability, accessibility, and cultural relevance to ensure they are truly effective for delirium education.

(J Med Internet Res 2025;27:e53087) doi:10.2196/53087

XSL•FO
RenderX

KEYWORDS

education; health information; internet; delirium; patient; caregiver; brain lesions; confusion; inattentiveness; disorientation; family education; information seeking; readability; high-quality websites; accessibility

Introduction

Delirium is a common, potentially preventable medical condition characterized by an acute onset of inattention, altered level of consciousness, or disorganized thinking. Delirium is the most common hospital-acquired complication [1-3], with greater prevalence in older adults (≥70 years of age) [1] and critically ill adults and children [2,4-7]. Emerging literature consistently highlights the negative impacts of delirium on patients (eg, increased risk of morbidity and mortality) [5,8,9] and families (eg, symptoms of distress, helplessness, and anxiety) [10-16]. Despite its prevalence and negative outcomes, delirium remains poorly recognized and is often missed by health care providers [17-19].

Families at the bedside may be important partners in delirium prevention, detection, and management. They are well-positioned to notice subtle changes in their loved one's cognition and behavior from their prehospitalized levels and help to identify symptoms of delirium [20-23]. However, there are challenges to building effective partnerships between families and the health care team. First, families require delirium knowledge to participate in delirium care [24,25]. The literature indicates that not all health care providers engage in this aspect of patient care [26,27]. Time constraints and lack of access to sufficient educational materials limit the ability of health care providers to deliver health education on delirium [28]. Even when health care providers discuss delirium with patients and their families, the qualitative literature suggests gaps in understanding and unmet delirium information needs [29-31], which may prompt patients or families to independently seek out their own sources of delirium information.

A recent study found that families of patients admitted to an intensive care unit (ICU) had a self-reported low level of delirium knowledge and learned about delirium by searching the term online [32]. A separate study indicated families preferred obtaining delirium-related information through internet sources [24]. As families self-report accessing delirium information online, high-quality online delirium information may be one way to improve patient, family, and the public's understanding of delirium, which can, in turn, empower them to participate in delirium prevention, detection, and management [33].

Over 10 million Americans access the internet for health information per day [34]. Recent studies report that health-related information on websites is often inaccurate, biased, misleading, or outdated [35-41]. Information about delirium is widely available on the internet, but, as with other health information on the internet, the information may be low-quality and inconsistent across sources. It is unknown if the websites of patients, families, and the public access to delirium information contain reliable, accurate, and up-to-date information.

The increasing availability of websites related to delirium is likely reflective of the creation of delirium societies or associations (American Delirium Society, European Delirium Association, and Australasian Delirium Association), World Delirium Awareness Day (established in 2017), and an increase in the implementation of regular delirium screening in hospitals [42,43]. The purpose of our study was to evaluate delirium website content based on readability, quality, and key content areas to identify high-quality delirium-related website content that could be used for patient, family, and public delirium education.

Methods

Website Search

We searched the top 3 most used search engines: Microsoft Bing [44], Google [45], and Yahoo [46-48] using the keywords "delirium" and misspelled "delerium" [49] to ensure comprehensive coverage of websites addressing the topic. "Delirium" is the primary and widely recognized term in both clinical and public domains, allowing us to retrieve the most relevant content. Including the misspelling "delerium" accounted for potential variations in user input, helping to capture additional website content that might not be optimized for the correct spelling. We disabled location identifiers, conducted the searches using a newly launched incognito or private window, and cleared cookies before each search to ensure search results were not influenced by precise geographic location or search history. We collected the top 200 search results from each search engine to capture the most relevant and widely accessed websites while maintaining consistency across engines. Search engines, such as Microsoft Bing, Google, and Yahoo, use sophisticated algorithms to rank search results based on factors such as relevance to keywords searched, content quality, and user engagement. In this study, while acknowledging that search results are typically personalized and may vary for each user based on factors such as location, search history, and preferences, for the purposes of analysis, we assumed that all users received the same search results. After removing duplicates, we excluded websites if they met any of the following exclusion criteria: (1) does not provide information on delirium, (2) written in a language other than English, (3) retrieved URL linked to a media source, (eg, YouTube channel or podcast), (4) retrieved URL targeted researchers or health care providers (eg, research articles), or (5) retrieved URL linked to a website that produced an error message, or content not available without a subscription. Despite disabling location identifiers, it is possible that search engine algorithms used IP addresses or language settings and identified the delirium association that is most closely associated with our location (ie, the American Delirium Society). As such, we also evaluated the website of the delirium association (European Delirium Association), which was not identified by the search (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart illustrating the process of selecting websites for quality analysis.

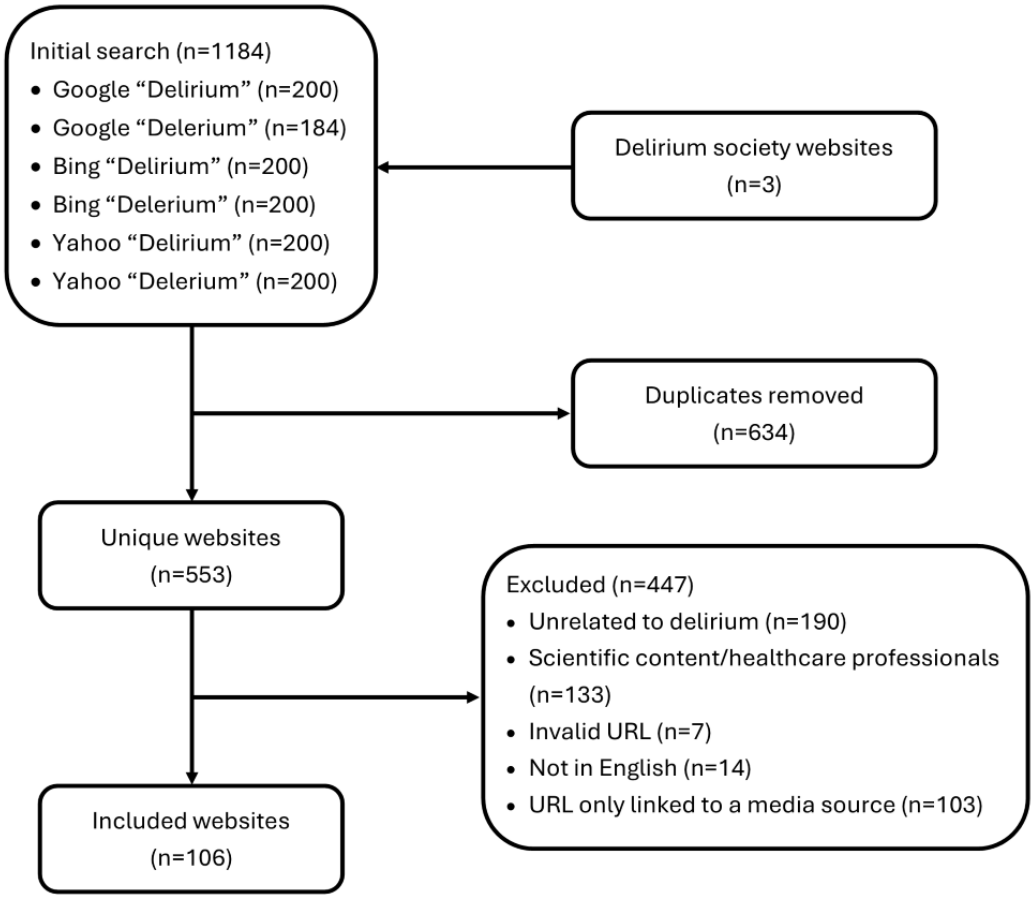

Website Quality

A team of patients, families, researchers, and health care providers (herein referred to as reviewers) evaluated delirium website content. The quality of website content was determined using the validated DISCERN tool (Multimedia Appendix 1) and the *JAMA* (*Journal of the American Medical Association*) benchmark criteria (Multimedia Appendix 2). DISCERN is a standardized set of criteria developed to evaluate the transparency, quality, and reliability of health information written for the public [50,51]. The DISCERN instrument has 16 questions. Each question can be scored from 1 (definite no) to 5 (definite yes). The total score can range from 16-80, wherein scores can be interpreted as excellent (63-75 points), good (51-62 points), fair (39-50 points), poor (27-38 points), and very poor (16-26 points) quality [52]. Two reviewers independently and in duplicate evaluated each website using the DISCERN instrument. DISCERN scores that differed by 5 points or more between the two reviewers were scored again by a third reviewer. The mean of all reviewers was taken as the final DISCERN score for each website. JAMA benchmark criteria include 4 standards of credible information sources: authorship (affiliations and credentials), attribution (references, sources, and copyright), disclosure (conflicts of interests), and currency (when content is posted and updated) [53]. Scores range from 0 to 4, with a score of 4 indicating the website is a credible source. Two reviewers independently and in duplicate assigned scores for each item in the JAMA benchmark criteria to indicate if it was present (score=1) or absent (score=0). Disagreements in scoring were resolved through discussion or the inclusion of a third reviewer.

Delirium-Related Content

There are no published guidelines for evaluating delirium education materials. To evaluate the quality of content from each website, a working group of patient partners (past ICU patients and families who are members of our research team), delirium researchers, and clinicians identified seven key points that patients, families, and the public would want to know about delirium: (1) the definition of delirium (must align with definition from the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* Fifth Edition]) [54], the American Delirium Society, European Delirium Association, or Australasian Delirium Association [55-57]; Multimedia Appendix 3); (2) delirium risk factors [5]; (3) short and long-term outcomes of delirium [5]; (4) signs and symptoms of delirium [5,58]; (5) information to differentiate between delirium and dementia [59]; (6) delirium prevalence; and (7) strategies to prevent and manage delirium [13,15]. The DISCERN criteria evaluate treatment choices, so only items 1-6 (of 7) were evaluated. Two reviewers evaluated the websites independently and in duplicate to determine if the site provided complete and accurate information in these 6 areas (Multimedia Appendix 3**).** Each item was scored to be present (score=2), somewhat present (ie, incorrect or incomplete information, score=1), or absent (score=0). Scores could range from 0 to 12, with a higher score indicating the website had more sufficient delirium content.

XSL•FO
RenderX

Disagreements in scoring were resolved through discussion or the inclusion of a third reviewer.

Readability

The level of difficulty of the reading material of each website was scored using the Simple Measure of Gobbledygook (SMOG) formula [60], the Flesch Reading Ease score [61], and the Flesch Kincaid grade level [62]. These scores are credible tools that are widely used to evaluate readability [63]. The included websites were evaluated if they were written at the American Medical Association and National Institutes of Health recommendation of a grade 6 reading level [64,65].

Data Extraction and Analyses

A standardized data extraction template was created and piloted to ensure reviewers understood how to evaluate the quality, content, and readability of each website. The template included detailed column headers to ensure each website was scored the same way (eg, guidelines for rating each DISCERN question). We also collected the following variables for each website: URL or website address, date the website was created, country, free text description of the website, website creators, and patient population or clinical setting targeted. We mapped the included websites onto the patient engagement framework that described patient engagement as inform (provision of education), activate (prompts action), or collaborate (encouraging interaction with health care providers) [3]. To ensure consistency in evaluations, all reviewers underwent comprehensive training. All reviewers read the DISCERN handbook and the study protocol, which described how to evaluate the quality, content, and readability of each website. All reviewers met to go through the evaluation criteria and to evaluate one website together. We then conducted a calibration exercise wherein all reviewers independently evaluated the same 5 randomly selected websites. We met to discuss discrepancies and reached a consensus on the overall quality of the 5 websites. The remaining websites were evaluated independently and in duplicate. Quality checks were conducted at regular intervals to verify that the criteria were being interpreted and applied consistently. Reviewers met regularly to discuss progress and address any uncertainties or discrepancies identified by the quality checks.

Website scores were summarized descriptively using counts and percentages, mean (SD), or median (IQR). The Kruskal-Wallis test was used to evaluate website scores between website categories (eg, foundation or advocacy group, government, hospital, and academic institution). Statistical analyses were performed in Stata/MP (version 14.2; StataCorp LLC). A *P* value of <.05 was considered significant. We compiled the top 10 websites based on the highest weighted combined quality score. Each score-delirium-related content, readability, DISCERN, and *JAMA* benchmark was normalized using observed (readability only) or possible score ranges. Readability was reverse-scaled because a lower score is better. We weighted readability, delirium-related content, and DISCERN at 30% each, and *JAMA* benchmark criteria (website content transparency) at 10%. These weightings, determined by the working group, were used to calculate a final weighted score, where a higher score indicated higher quality.

Ethical Considerations

As the study did not involve human participants or human biological materials, the study was deemed exempt from the ethics board approval process. This study relied on publicly available data and therefore, ethics approval was not needed for this study as per the University of Calgary Conjoint Health Research Ethics Board guidelines [66].

Results

Website Search

On November 21, 2024, we searched Bing, Google, and Yahoo and removed duplicates and URLs that did not meet inclusion criteria; we identified 106 websites targeted toward patients, families, and the public. This included private or public hospitals (eg, Mayo Clinic: 21/106, 20%), commercial websites (eg, Merck: 20/106, 19%), and websites affiliated with a foundation or advocacy group (eg, American Delirium Society: 16/106, 15%). A description of each website is in Multimedia Appendix 4. Over half of the included URLs (62/106, 58%) used all 3 engagement strategies: inform (eg, delirium information), activate (eg, how to prevent or manage delirium), and collaborate (eg, when or how to talk with a health care provider about delirium). A total of 17 websites (16%) used inform strategies only, and 22 websites (21%) used both inform and activate engagement strategies. Of the 82 (77%) websites that reported a date when the website was posted or updated, the median time since the last update was 3 (IQR 2-5) years.

Website Content Quality

The median DISCERN score was 42 (IQR 33-50), representing fair quality. The quality ranged from 20 (very poor quality) to 78 (excellent quality). Many websites did not include a list of the sources used to compile the information on the website, refer to areas of uncertainty (eg, delirium may be missed or difficult to identify, no guideline-recommended pharmacological treatments for delirium), nor describe risks of each treatment (eg, antipsychotics if used for severe agitation). Several websites shared information that was not backed by a current evidence synthesis [5] or available guidelines [13,15]. This included claims that antipsychotics were the first line of treatment for delirium without discussion of their risk of prolonging or exacerbating delirium symptoms. Many websites mention antipsychotics as a means to manage the symptoms of delirium, such as agitation, which is a safety concern (supported by recent guidelines) [13,15]. DISCERN scores were significantly different between website categories (ie, academic, commercial, foundation or advocacy group, etc).

The median JAMA score was 1 (IQR 1-3), with 17/106 websites (16%) scoring 0, which indicates most websites lacked transparency. In particular, most websites did not list authors or contributors, their affiliations, and relevant credentials (67/106, 63%) nor included references and sources for all content (65/106, 61%). JAMA scores were not significantly different between website categories (ie, academic, commercial, foundation or advocacy group, etc). The median scores for delirium website transparency can be seen in Table 1.

XSL•FO
RenderX

Table 1. Comparison of the median DISCERN, Journal of the American Medical Association (JAMA) benchmark score, and weighted quality scores across website categories.

Website category	Websites, n (%)	DISCERN score (range 16-80), median (IQR)	JAMA benchmark score (range 0-4), median (IQR)	Weighted quality score (range 0-100), median (IQR)
Hospital affiliated	21 (19.8)	37.5 (32-44.5)	1 (0.8-3)	55.9 (52.4-59.4)
Commercial	20 (18.9)	44.5 (34.5-52.3)	3 (2-4)	57.4 (49.4-66.6)
Government	19 (17.9)	42 (32.5-51.5)	1 (1-2)	54.5 (47-64.3)
Foundation or advocacy organization	16 (15)	46.5(42.5-50.3)	1 (0-1.3)	58.4 (55.7-65.7)
Regional health authority	14 (13.2)	32.5 (27.3-40.5)	1 (1-1.8)	49.4 (46.1-56.2)
Academic	6 (5.7)	52.5 (50-57.3)	1.5 (1-2)	64.4 (60.2-70.2)
General reference or educational resource	6 (5.7)	40.5 (32.8-55)	2.5 (2-3)	54.3 (52.5-63.5)
Professional organization	4 (3.8)	34.0 (33-38)	1.5 (1-2)	50.1 (41.3-57.4)

Delirium-Related Content

The median score for delirium-related content was 8 (IQR 6-9). Nearly half of the websites included a complete definition of delirium (77/106, 73%). Website evaluators rated several websites (26/106, 25%) to have a somewhat correct definition of delirium (eg, missing items from the *DSM-5* or delirium society definitions of delirium). A majority of the websites (77/106, 73%) included predisposing (eg, age and dementia) and precipitating delirium risk factors (eg, surgery and dehydration) that are consistent with available evidence [5]. Websites that reported somewhat correct delirium risk factors (23/106, 22%) reported incomplete or inaccurate risk factors. These websites either focused solely on predisposing or precipitating factors, overlooked key groups like infants and young children, or included risk factors not widely associated with delirium in the literature (eg, male sex). Most websites (83/106, 78%) included the signs and symptoms of delirium. Those websites that somewhat described the signs and symptoms of delirium (19/106, 18%) often missed a hallmark of delirium (eg, inattention, acute onset, fluctuating course, or disorganized thinking). Many websites did not include short and long-term outcomes associated with delirium (54/106, 51%). This included not describing delirium as a risk factor for dementia or the association between delirium and risk for long-term cognitive decline. Many websites did not describe the prevalence of delirium (41/106, 39%) or vaguely described its prevalence (eg, “common;” 31/106, 29%). Half of the websites stated that there was a difference between dementia and delirium (54/106, 51%). Out of these 54 websites, 43 (80%) described in detail how dementia and delirium differed. Table 2 shows the mean scores for delirium content.

Table 2. Comparison of the mean delirium content scores across website categories.

Website category	Website, n (%)	Delirium content[a]					
		Delirium definition, mean (SD)	Risk factors and causes, mean (SD)	Short- and long-term outcomes, mean (SD)	Signs and symptoms, mean (SD)	Differences between delirium and dementia, mean (SD)	Delirium prevalence, mean (SD)
Hospital affiliated	21 (19.8)	1.6 (0.6)	1.5 (0.6)	0.7 (0.7)	1.6 (0.5)	0.6 (0.9)	0.9 (0.8)
Commercial	20 (18.9)	1.6 (0.5)	1.9 (0.4)	0.7 (0.9)	1.9 (0.5)	1.2 (0.9)	1 (0.9)
Government	19 (17.9)	1.9 (0.4)	1.9 (0.3)	0.5 (0.7)	1.9 (0.2)	0.9 (1.0)	0.8 (0.9)
Foundation or advocacy organization	16 (15)	1.8 (0.4)	1.7 (0.5)	0.9 (0.9)	1.9 (0.2)	1.3 (0.8)	0.8 (0.7)
Regional health authority	14 (13.2)	1.9 (0.3)	1.4 (0.8)	0.7 (0.8)	1.6 (0.6)	0.8 (0.9)	0.8 (0.9)
Academic	6 (5.7)	2 (0)	1.5 (0.5)	1.2 (0.7)	2 (0)	1.2 (0.9)	1.7 (0.5)
General reference or educational resource	6 (5.7)	1.3 (0.7)	1.7 (0.7)	1.3 (0.9)	1.3 (0.7)	0.7 (0.9)	1.3 (0.7)
Professional organization	4 (3.8)	1.3 (0.4)	1.5 (0.9)	0.3 (0.4)	1.3 (0.8)	0.5 (0.9)	1 (0.7)

[a]For all criteria, the minimum possible score=0, maximum=2.

Readability

The median SMOG readability score was a grade 15 (IQR 13-17, range 10-22) level. The median Flesch Kincaid Reading Level score was a grade 10 (IQR 9-12, range 5-21) level. The median Flesch Reading Ease score was 47 (IQR 38-56; range 0-80.7), indicating the websites are difficult to read for most of the

XSL•FO
RenderX

population. Of the 106 websites, 5 (5%) were written at a reading level equal to or lower than grade 6.

Top 10 Websites for Patient, Family, and Public Information on Delirium

The median for the weighted quality score for websites was 56.1 (IQR 49.3-65.5, range 28.8-84.1). Based on the normalized and weighted evaluation criteria, the top 10 websites for delirium for patient and family education are summarized in Table 3. The Mayo Clinic website, which appeared as the top result for Google and Yahoo "delirium" searches, ranked 15th overall based on weighted scores for quality, content, and readability, with an overall weighted score similar to the top 10 websites (69.8). In contrast, one of the websites listed among the top 10 search results ranked 83rd overall due to incomplete and low-quality information such as inaccurate claims (eg, counseling as a method to address disorientation, male sex as a risk factor, and a section on "confusion."). The Wikipedia entry on delirium was in the first 10 results of Microsoft Bing (3rd), Google (5th), and Yahoo (4th). However, it had an overall weighted score of 54.8 due to its poor readability. Weighted quality scores were not significantly different between website categories (ie, academic, commercial, foundation or advocacy group, etc).

Table 3. The top 10 websites ranked by weighted, combined quality, delirium content, and readability scores.

Website[a]	Position in the Bing, Google, and Yahoo search			DISCERN score (range 16-80)	Flesh Kincaid Grade Level	Delirium content score (range 0-12)	Composite weighted score (rank)
	Bing	Google	Yahoo				
Healthline.com [67]	92	7	69	75	10.5	11	84.1 (1)
Aarp.org [68]	—[b]	127	—	77	10.7	11	82.6 (2)
Sign.ac.uk [69]	—	108	—	59	8.3	12	78.2 (3)
UpToDate.com [70]	—	14	—	64	12.9	12	76.9 (4)
HRH.ca [71]	—	96	—	64	9.3	9	73.7 (5)
RGPToronto.ca [72]	106	55	112	78	12.3	10	73.5 (6)
Merckmanuals.com [73]	—	12	180	53	12.4	12	72.7 (7)
Healthify.nz [74]	—	99	—	38	8	9	72.7 (8)
Verywellhealth.com [75]	142	94	9	52	12.2	12	72.6 (9)
Clevelandclinic.org [76]	1	2	2	62	9.6	9	72.3 (10)

[a]All websites listed, including UpToDate (which typically requires a subscription), are publicly available.

[b]Indicates the website was not in the top 200 of search results for that search engine.

Discussion

Principal Findings

Delirium websites are one source of delirium education to prepare families to partner with delirium prevention, detection, and management. This study reports on the quality, content, and readability of websites with delirium information for patients, families, and the public. Overall, our findings suggest the quality of delirium-related website content is fair, with many websites lacking credibility and transparency. The American Medical Association and National Institutes of Health recommend that patient education materials be written at a grade 6 reading level [64,65], but nearly all websites (100/106, 95%) were written at higher than a grade 6 reading level. Furthermore, websites that families may encounter when looking for information about delirium may include incomplete information about delirium, overlook key groups such as infants or young children, or provide outdated or inaccurate information. This study identified the 10 best websites that patients, families, and clinicians can refer to, to find information about delirium.

It is crucial that high-quality and easily understandable websites on delirium are available to patients, families, and the public for several reasons. First, delirium is a common and serious medical condition that can be confusing and frightening for patients and families [4,10-13,77,78]. This may prompt families to search for delirium on the internet. Clear and concise information will help families to better understand delirium, its causes, signs, and prevention or management options. Second, countries and organizations promote engaging families with patient care [16,23,79]. To be active participants in delirium care, families must be provided with delirium education. Accessible health information can empower patients and families to participate in delirium prevention, detection, and management [20,80] and seek medical attention when necessary. Like any family engagement intervention, not all families may want to participate in delirium care, and they have reported that one of their preferential ways to receive information is through the internet. As such, websites should offer information that informs (providing information on delirium), activates (encouraging families to prevent and manage delirium), and collaborates (preparing families to discuss delirium with health care teams) with families to cater to a broad audience with varying information needs.

To improve the quality of delirium websites, it is essential to identify common deficiencies among them. A majority of the included websites did not provide proper citations for the information used in the website content. Including sources on websites benefits readers by fostering credibility, transparency,

XSL•FO
RenderX

and accountability of the website information and enables readers to access additional delirium resources. Most websites (100/106, 95%) were written above the recommended grade 6 level. It is important that health information is presented in a manner that is accessible to its target so that the information is not misinterpreted and to enable the reader to make informed decisions about their health. Other studies evaluating the quality of health information on websites also report that health information is not written at an appropriate reading level [81-85]. While website creators can modify their content to meet the grade 6 recommendation, the best practice would be to codevelop websites with patients, families, and the public. Websites should also include the following delirium-related content: (1) the definition of delirium (from the *DSM-5* [54] or from delirium societies or associations) [55-57], (2) delirium risk factors [5], (3) short and long-term outcomes of delirium [5], (4) signs and symptoms of delirium [54,58], (5) information to differentiate between delirium and dementia [59], (6) delirium prevalence [1,2,4-7], and (7) strategies to prevent and manage delirium [13,15]. With more people using the internet to access health information, it is imperative that website developers follow the above guidance to ensure websites include the highest quality and readable delirium information for patients, families, and the public.

This study has several implications for practice. First, there is a need for families to be integrated as partners in delirium care [86]. An important first step is to provide families with delirium information to prepare them to participate in delirium prevention, detection, and management. However, staff often lack the time to provide comprehensive delirium education to families due to heavy workloads and clinical responsibilities. Health care providers can leverage this curated list of websites to supplement delirium information provided to patients and families. By directing families to these websites, they can ensure families are consulting the highest quality and most reliable delirium information currently available on the internet. Second, it is clear that higher-quality delirium websites with plain language are needed. This can be accomplished by hospital organizations or delirium societies creating their own delirium education materials. Hospitals can leverage the set of criteria (ie, DISCERN, JAMA benchmark, readability measures, and delirium content) when adapting or creating delirium-related materials on their websites. Furthermore, there remains a need for policymakers to prioritize the importance of digital health literacy. This might include initiatives to enhance how people navigate health information on websites. This might also include advocating for high-quality educational resources.

Strengths and Limitations

There are several strengths and limitations that should be considered in this study. To our knowledge, this is the first study to identify and evaluate websites on delirium that could be used for patient and family education. Delirium websites were identified using the top 3 most used search engines (Bing, Google, and Yahoo). Finally, all study activities included patient and family partners. Despite these strengths, there are several limitations that should be considered. First, the search strategy consisted of only 2 search terms: delirium and the most commonly misspelled form of delirium (delerium). It is possible that patients, families, and the public may use other terms to search for delirium information, and, as such, some websites may have been missed. Second, the search was performed in Canada. Despite disabling location services, which limited the extent to which the search engines can use our location, other aspects of our location may have been inferred based on our IP address or language settings. For this reason, and that our team was proficient in English, we only included websites written in English. As such, the results of this study may not be generalizable to people who do not speak English. To attain a more comprehensive understanding of delirium-related websites in languages other than English, further research is warranted. This should include other prominent search engines with significant market shares (eg, Yandex, Baidu, Petal Search, and DuckDuckGo) and collaborative efforts with researchers from diverse linguistic backgrounds and countries. Third, the analysis of the websites was limited to what was reported, which may not have comprehensively captured the development of the website. This may include if websites used credible sources or delirium experts to compile the websites but did not cite the sources or identify the delirium experts. Finally, the tool that was used to evaluate delirium content has not been externally validated, as it was developed specifically for this study. Finally, while DISCERN, JAMA benchmark, delirium content, and readability tools are robust for assessing quality and transparency, they focus on structural elements rather than usability or accessibility for patients and families. Websites with high scores often had dense information or busy layouts that may overwhelm patients or families seeking concise and accessible information. These metrics do not account for patient or family experience, engagement, or cultural relevance, which are critical for effective patient and family education.

Conclusion

This study suggests delirium websites for the public are of fair quality. Inadequacies in evaluated websites, such as lack of transparency, incomplete delirium information, and poor readability, should be addressed when updating current or creating new delirium websites aimed at patients, families, or the public. Following the outlined standards for quality, delirium content, and readability will ensure high-quality, transparent, and accessible delirium information for patients, families, and the public.

Acknowledgments

The authors would like to thank Kira Makuk and Thérèse Poulin for their help with conducting the Bing, Yahoo, and Google searches and piloting the delirium website evaluation. This work is funded by the Canadian Institutes of Health Research Transitions in Care Team Grant (grant 423947). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

XSL•FO
RenderX

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors have read and have approved of the submission of this manuscript. KMF contributed to the study design. KK, Strayer K, NJ, Spence K, NF, and SK evaluated the websites. KK and Strayer K analyzed and synthesized the results and prepared the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The DISCERN instrument.
[DOCX File , 32 KB - jmir_v27i1e53087_app1.docx]

Multimedia Appendix 2
JAMA (Journal of the American Medical Association) benchmark criteria for credibility of medical information on the internet.
[DOCX File , 22 KB - jmir_v27i1e53087_app2.docx]

Multimedia Appendix 3
Evaluation of Delirium Content Guide.
[DOCX File , 26 KB - jmir_v27i1e53087_app3.docx]

Multimedia Appendix 4
Description of the websites included in the study.
[DOCX File , 42 KB - jmir_v27i1e53087_app4.docx]

References

1. Gibb K, Seeley A, Quinn T, Siddiqi N, Shenkin S, Rockwood K, et al. The consistent burden in published estimates of delirium occurrence in medical inpatients over four decades: a systematic review and meta-analysis study. Age Ageing 2020;49(3):352-360 [FREE Full text] [doi: 10.1093/ageing/afaa040] [Medline: 32239173]
2. Krewulak KD, Stelfox HT, Leigh JP, Ely EW, Fiest KM. Incidence and prevalence of delirium subtypes in an adult ICU: a systematic review and meta-analysis. Crit Care Med 2018;46(12):2029-2035. [doi: 10.1097/CCM.0000000000003402] [Medline: 30234569]
3. Fiest KM, McIntosh CJ, Demiantschuk D, Leigh JP, Stelfox HT. Translating evidence to patient care through caregivers: a systematic review of caregiver-mediated interventions. BMC Med 2018;16(1):105 [FREE Full text] [doi: 10.1186/s12916-018-1097-4] [Medline: 29996850]
4. Ely EW, Inouye SK, Bernard GR, Gordon S, Francis J, May L, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). JAMA 2001;286(21):2703-2710. [doi: 10.1001/jama.286.21.2703] [Medline: 11730446]
5. Wilson JE, Mart MF, Cunningham C, Shehabi Y, Girard TD, MacLullich AMJ, et al. Delirium. Nat Rev Dis Primers 2020;6(1):90 [FREE Full text] [doi: 10.1038/s41572-020-00223-4] [Medline: 33184265]
6. Traube C, Silver G, Reeder R, Doyle H, Hegel E, Wolfe H, et al. Delirium in critically Ill children: an international point prevalence study. Crit Care Med 2017;45(4):584-590 [FREE Full text] [doi: 10.1097/CCM.0000000000002250] [Medline: 28079605]
7. Semple D, Howlett M, Strawbridge J, Breatnach C, Hayden J. A systematic review and pooled prevalence of delirium in critically ill children. Crit Care Med 2022;50(2):317-328. [doi: 10.1097/CCM.0000000000005260] [Medline: 34387241]
8. Witlox J, Eurelings LSM, de Jonghe JFM, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. JAMA 2010;304(4):443-451. [doi: 10.1001/jama.2010.1013] [Medline: 20664045]
9. Cole MG, Bailey R, Bonnycastle M, McCusker J, Fung S, Ciampi A, et al. Partial and no recovery from delirium in older hospitalized adults: frequency and baseline risk factors. J Am Geriatr Soc 2015;63(11):2340-2348. [doi: 10.1111/jgs.13791] [Medline: 26515438]
10. Williams ST, Dhesi JK, Partridge JSL. Distress in delirium: causes, assessment and management. Eur Geriatr Med 2020;11(1):63-70. [doi: 10.1007/s41999-019-00276-z] [Medline: 32297237]

XSL•FO
RenderX

11. Luth EA, Maciejewski PK, Phongtankuel V, Xu J, Prigerson HG. Associations between hospice care and scary family caregiver experiences. J Pain Symptom Manage 2021;61(5):909-916 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.08.041] [Medline: 33038426]
12. Partridge JSL, Martin FC, Harari D, Dhesi JK. The delirium experience: what is the effect on patients, relatives and staff and what can be done to modify this? Int J Geriatr Psychiatry 2013;28(8):804-812. [doi: 10.1002/gps.3900] [Medline: 23112139]
13. Delirium: Prevention, Diagnosis and Management in Hospital and Long-Term Care. London: National Institute for Health and Care Excellence (NICE); 2023.
14. Davis D, Searle S, Tsui A. The Scottish intercollegiate guidelines network: risk reduction and management of delirium. Age Ageing 2019;48(4):485-488. [doi: 10.1093/ageing/afz036] [Medline: 30927352]
15. Devlin JW, Skrobik Y, Gélinas C, Needham DM, Slooter AJC, Pandharipande PP, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. Crit Care Med 2018;46(9):e825-e873 [FREE Full text] [doi: 10.1097/CCM.0000000000003299] [Medline: 30113379]
16. O'Rourke M. he Australian commission on safety and quality in health care agenda for improvement and implementation. Asia Pacific Journal of Health Management 2007;2(2):21 [FREE Full text]
17. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. Lancet 2014;383(9920):911-922 [FREE Full text] [doi: 10.1016/S0140-6736(13)60688-1] [Medline: 23992774]
18. Geriatric Medicine Research Collaborative. Delirium is prevalent in older hospital inpatients and associated with adverse outcomes: results of a prospective multi-centre study on world delirium awareness day. BMC Med 2019;17(1):229. [doi: 10.1186/s12916-019-1458-7] [Medline: 31837711]
19. Lange PW, Lamanna M, Watson R, Maier AB. Undiagnosed delirium is frequent and difficult to predict: results from a prevalence survey of a tertiary hospital. J Clin Nurs 2019;28(13-14):2537-2542. [doi: 10.1111/jocn.14833] [Medline: 30786081]
20. Fiest KM, Krewulak KD, Ely EW, Davidson JE, Ismail Z, Sept BG, et al. Partnering with family members to detect delirium in critically ill patients. Crit Care Med 2020;48(7):954-961. [doi: 10.1097/CCM.0000000000004367] [Medline: 32332281]
21. Steis MR, Evans L, Hirschman KB, Hanlon A, Fick DM, Flanagan N, et al. Screening for delirium using family caregivers: convergent validity of the family confusion assessment method and interviewer-rated confusion assessment method. J Am Geriatr Soc 2012;60(11):2121-2126 [FREE Full text] [doi: 10.1111/j.1532-5415.2012.04200.x] [Medline: 23039310]
22. Shulman RW, Kalra S, Jiang JZ. Validation of the sour seven questionnaire for screening delirium in hospitalized seniors by informal caregivers and untrained nurses. BMC Geriatr 2016;16:44 [FREE Full text] [doi: 10.1186/s12877-016-0217-2] [Medline: 26879927]
23. Aggar C, Craswell A, Bail K, Compton R, Hughes M, Sorwar G, et al. Partnering with carers in the management of delirium in general acute care settings: an integrative review. Australas J Ageing 2023;42(4):638-648. [doi: 10.1111/ajag.13229] [Medline: 37528556]
24. Bull MJ, Boaz L, Sjostedt JM. Family caregivers' knowledge of delirium and preferred modalities for receipt of information. J Appl Gerontol 2016;35(7):744-758. [doi: 10.1177/0733464814535484] [Medline: 24942969]
25. Tonna JE. Negative studies should inform our research and care: engaging family members in the care of the critically Ill. Crit Care Med 2020;48(7):1077-1079 [FREE Full text] [doi: 10.1097/CCM.0000000000004393] [Medline: 32568902]
26. Smithburger PL, Korenoski AS, Kane-Gill SL, Alexander SA. Perceptions of family members, nurses, and physicians on involving patients' families in delirium prevention. Crit Care Nurse 2017;37(6):48-57. [doi: 10.4037/ccn2017901] [Medline: 29196587]
27. Cohen C, Pereira F, Kampel T, Bélanger L. Integration of family caregivers in delirium prevention care for hospitalized older adults: a case study analysis. J Adv Nurs 2021;77(1):318-330. [doi: 10.1111/jan.14593] [Medline: 33078872]
28. Kääriäinen M, Kyngäs H. The quality of patient education evaluated by the health personnel. Scand J Caring Sci 2010;24(3):548-556. [doi: 10.1111/j.1471-6712.2009.00747.x] [Medline: 20409053]
29. Bohart S, Merete Møller A, Forsyth Herling S. Do health care professionals worry about delirium? Relatives' experience of delirium in the intensive care unit: a qualitative interview study. Intensive Crit Care Nurs 2019;53:84-91. [doi: 10.1016/j.iccn.2019.04.010] [Medline: 31079979]
30. Toye C, Matthews A, Hill A, Maher S. Experiences, understandings and support needs of family carers of older patients with delirium: a descriptive mixed methods study in a hospital delirium unit. Int J Older People Nurs 2014;9(3):200-208. [doi: 10.1111/opn.12019] [Medline: 23383980]
31. Schmitt EM, Gallagher J, Albuquerque A, Tabloski P, Lee HJ, Gleason L, et al. Perspectives on the delirium experience and its burden: common themes among older patients, their family caregivers, and nurses. Gerontologist 2019;59(2):327-337 [FREE Full text] [doi: 10.1093/geront/gnx153] [Medline: 30870568]
32. Parsons Leigh J, Krewulak KD, Zepeda N, Farrier CE, Spence KL, Davidson JE, et al. Patients, family members and providers perceive family-administered delirium detection tools in the adult ICU as feasible and of value to patient care and family member coping: a qualitative focus group study. Can J Anaesth 2021;68(3):358-366 [FREE Full text] [doi: 10.1007/s12630-020-01866-3] [Medline: 33210217]

XSL•FO
RenderX

33. Lee J, Yeom I, Yoo S, Hong S. Educational intervention for family caregivers of older adults with delirium: an integrative review. J Clin Nurs 2023;32(19-20):6987-6997. [doi: 10.1111/jocn.16816] [Medline: 37370251]
34. Fox M, Duggan M. Health Online 2013 Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-01-05]
35. Dornan BA, Oermann MH. Evaluation of breastfeeding web sites for patient education. MCN Am J Matern Child Nurs 2006;31(1):18-23. [doi: 10.1097/00005721-200601000-00006] [Medline: 16371821]
36. Nichols C, Oermann MH. An evaluation of bariatric Web sites for patient education and guidance. Gastroenterol Nurs 2005;28(2):112-117. [doi: 10.1097/00001610-200503000-00005] [Medline: 15832111]
37. Oermann M, McInerney SM. An evaluation of sepsis web sites for patient and family education. Plast Surg Nurs 2007;27(4):192-196. [doi: 10.1097/01.PSN.0000306184.95812.8a] [Medline: 18165726]
38. Oermann MH, Gerich J, Ostosh L, Zaleski S. Evaluation of asthma websites for patient and parent education. J Pediatr Nurs 2003;18(6):389-396. [doi: 10.1016/s0882-5963(03)00161-1] [Medline: 15058535]
39. Oermann MH, Lowery NF, Thornley J. Evaluation of web sites on management of pain in children. Pain Manag Nurs 2003;4(3):99-105. [doi: 10.1016/s1524-9042(03)00029-8] [Medline: 14566707]
40. Labovitch RS, Bozic KJ, Hansen E. An evaluation of information available on the internet regarding minimally invasive hip arthroplasty. J Arthroplasty 2006;21(1):1-5. [doi: 10.1016/j.arth.2005.04.019] [Medline: 16446178]
41. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
42. Bauernfreund Y, Butler M, Ragavan S, Sampson EL. TIME to think about delirium: improving detection and management on the acute medical unit. BMJ Open Qual 2018;7(3):e000200 [FREE Full text] [doi: 10.1136/bmjoq-2017-000200] [Medline: 30167472]
43. Martin L, Lyons M, Patton A, O Driscoll M, McLoughlin K, Hannon E, et al. Implementing delirium screening in the emergency department: a quality improvement project. BMJ Open Qual 2022;11(2):e001676 [FREE Full text] [doi: 10.1136/bmjoq-2021-001676] [Medline: 35764360]
44. Bing. Microsoft. URL: https://www.bing.com [accessed 2024-11-21]
45. Google. Google Search n.d. URL: https://www.google.com/ [accessed 2024-11-21]
46. Top 8 best search engines (of 2021). RapidApi. 2021. URL: https://rapidapi.com/blog/best-search-engines/ [accessed 2021-11-18]
47. Meet the top 10 search engines in the world in 2021. Oberlo. 2020. URL: https://www.oberlo.com/blog/top-search-engines-world [accessed 2022-01-05]
48. Yahoo search. Yahoo. URL: https://tinyurl.com/2kt7ks7s [accessed 2025-01-29]
49. Reddin C, Koay WJ, Mulkerrin EC, O Keeffe ST. Misspelling of delirium as "Delerium" in the academic literature. J Am Geriatr Soc 2020;68(3):660-661. [doi: 10.1111/jgs.16289] [Medline: 31808540]
50. Charnock D, Shepperd S. Learning to DISCERN online: applying an appraisal tool to health websites in a workshop setting. Health Educ Res 2004;19(4):440-446. [doi: 10.1093/her/cyg046] [Medline: 15155597]
51. The DISCERN instrument. DISCERN Online. 1997. URL: http://www.discern.org.uk/discern_instrument.php [accessed 2022-01-05]
52. Weil AG, Bojanowski MW, Jamart J, Gustin T, Lévêque M. Evaluation of the quality of information on the Internet available to patients undergoing cervical spine surgery. World Neurosurg 2014;82(1-2):e31-e39. [doi: 10.1016/j.wneu.2012.11.003] [Medline: 23142585]
53. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
54. Diagnostic and Statistical Manual of Mental Disorders 5th ed (DSM-5). Washington, D.C: American Psychiatric Association Publishing; 2013.
55. American Delirium Society. URL: https://americandeliriumsociety.org/ [accessed 2025-01-29]
56. European Delirium Association, American Delirium Society. The DSM-5 criteria, level of arousal and delirium diagnosis: inclusiveness is safer. BMC Med 2014;12:141 [FREE Full text] [doi: 10.1186/s12916-014-0141-2] [Medline: 25300023]
57. Australasian Delirium Association. URL: https://www.delirium.org.au/ [accessed 2025-01-29]
58. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Intern Med 1990;113(12):941-948. [doi: 10.7326/0003-4819-113-12-941] [Medline: 2240918]
59. Forrant JA. Dementia or delirium: do you know the difference? Nurs Crit Care 2009;4(2):24-30. [doi: 10.1097/01.CCN.0000347125.47833.81]
60. McLaughlin GH. SMOG grading: a new readability formula. Journal of Reading 1969;12(8):639-646 [FREE Full text]
61. Flesch R. A new readability yardstick. J Appl Psychol 1948;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
62. Kincaid J, Fishburne R, Rogers R, Chissom B. Derivation of new readability formula for navy enlisted personnel. URL: https://stars.library.ucf.edu/istlibrary/56/ [accessed 2025-01-29]

XSL•FO
RenderX

63. Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ Behav 2006;33(3):352-373. [doi: 10.1177/1090198105277329] [Medline: 16699125]
64. Weiss BD. Health Literacy. Chicago: American Medical Foundation; 2003.
65. MedlinePlus: how to write easy to read health materials. National Institutes of Health. URL: http://www nlm nih gov/medlineplus/etr html [accessed 2025-01-29]
66. Resources: Ethics and Compliance. University of Calgary. URL: https://research.ucalgary.ca/research-services/ethics-compliance/ethics-resources#chreb [accessed 2025-02-19]
67. Healthline. URL: https://www.healthline.com/health/hospital-delirium [accessed 2025-02-05]
68. AARP. URL: https://www.aarp.org/health/brain-health/global-council-on-brain-health/delirium/ [accessed 2024-02-05]
69. SIGN. URL: https://www.sign.ac.uk/assets/pat157.pdf [accessed 2025-02-05]
70. UpToDate. URL: https://www.uptodate.com/contents/delirium-beyond-the-basics [accessed 2025-02-05]
71. Humber River Health. URL: https://www.hrh.ca/patient-education/005031_Delirium.pdf [accessed 2025-02-05]
72. RGP Toronto. URL: https://rgptoronto.ca/wp-content/uploads/2024/04/Guidance-on-Delirium-Care-for-Older-Adults-in-the-Community-V1-2024.pdf [accessed 2025-02-05]
73. Merck manuals. URL: https://www.merckmanuals.com/en-ca/home/brain-spinal-cord-and-nerve-disorders/delirium-and-dementia/delirium [accessed 2025-02-05]
74. Healthify. URL: https://healthify.nz/health-a-z/d/delirium/ [accessed 2025-02-05]
75. verywell health. URL: https://www.verywellhealth.com/delirium-5223127 [accessed 2025-02-05]
76. Cleveland Clinic. URL: https://my.clevelandclinic.org/health/diseases/15252-delirium [accessed 2025-02-05]
77. Siddiqi N, House AO, Holmes JD. Occurrence and outcome of delirium in medical in-patients: a systematic literature review. Age Ageing 2006;35(4):350-364. [doi: 10.1093/ageing/afl005] [Medline: 16648149]
78. Almeida ICT, Soares M, Bozza FA, Shinotsuka CR, Bujokas R, Souza-Dantas VC, et al. The impact of acute brain dysfunction in the outcomes of mechanically ventilated cancer patients. PLoS One 2014;9(1):e85332 [FREE Full text] [doi: 10.1371/journal.pone.0085332] [Medline: 24465538]
79. Davidson JE, Zisook S. Implementing family-centered care through facilitated sensemaking. AACN Adv Crit Care 2017;28(2):200-209. [doi: 10.4037/aacnacc2017102] [Medline: 28592480]
80. Determeijer JJ, Leopold SJ, Spijker R, Agyemang C, van Vugt M. Family participation to enhance care and tackle health worker shortages in resource-limited hospitals: a systematic review. J Glob Health 2023;13:04005 [FREE Full text] [doi: 10.7189/13.04005] [Medline: 36655879]
81. Kaicker J, Dang W. Assessing the quality and reliability Of health information on ERCP using the DISCERN instrument. Health Care: Current Reviews 2013;1(1):1-4 [FREE Full text] [doi: 10.4172/2375-4273.1000104]
82. Saleh D, Fisher JH, Provencher S, Liang Z, Ryerson CJ, Weatherald J. A systematic evaluation of the quality, accuracy, and reliability of internet websites about pulmonary arterial hypertension. Ann Am Thorac Soc 2022;19(8):1404-1413. [doi: 10.1513/AnnalsATS.202103-325OC] [Medline: 34813417]
83. Corcelles R, Daigle CR, Talamas HR, Brethauer SA, Schauer PR. Assessment of the quality of Internet information on sleeve gastrectomy. Surg Obes Relat Dis 2015;11(3):539-544. [doi: 10.1016/j.soard.2014.08.014] [Medline: 25604832]
84. Roughead T, Sewell D, Ryerson CJ, Fisher JH, Flexman AM. Internet-based resources frequently provide inaccurate and out-of-date recommendations on preoperative fasting: a systematic review. Anesth Analg 2016;123(6):1463-1468. [doi: 10.1213/ANE.0000000000001590] [Medline: 27644057]
85. Guo WJ, Wang WK, Xu D, Qiao Z, Shi YL, Luo P. Evaluating the quality, content, and readability of online resources for failed back spinal surgery. Spine (Phila Pa 1976) 2019;44(7):494-502. [doi: 10.1097/BRS.0000000000002870] [Medline: 30234809]
86. Aggar C, Craswell A, Bail K, Compton RM, Hamiduzzaman K, Sorwar G, et al. Commentary: prevention and management of delirium in older Australians: the need for the integration of carers as partners in care. Lancet Reg Health West Pac 2022;27:100598 [FREE Full text] [doi: 10.1016/j.lanwpc.2022.100598] [Medline: 36148384]

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
ICU: intensive care unit
JAMA: Journal of the American Medical Association
SMOG: Simple Measure of Gobbledygook

XSL•FO
RenderX

Edited by A Mavragani; submitted 25.09.23; peer-reviewed by A Craswell, YJ Lee; comments to author 23.02.24; revised version received 27.03.24; accepted 16.01.25; published 20.02.25.

Please cite as:
Krewulak K, Strayer K, Jaworska N, Spence K, Foster N, Kupsch S, Sauro K, Fiest KM
Evaluation of the Quality of Delirium Website Content for Patient and Family Education: Cross-Sectional Study
J Med Internet Res 2025;27:e53087
URL: https://www.jmir.org/2025/1/e53087
doi:10.2196/53087
PMID:39977019

©Karla Krewulak, Kathryn Strayer, Natalia Jaworska, Krista Spence, Nadine Foster, Scotty Kupsch, Khara Sauro, Kirsten M Fiest. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Behavior Change Support Systems for Self-Treating Procrastination: Systematic Search in App Stores and Analysis of Motivational Design Archetypes

Jeanine Kirchner-Krath[1]; Manuel Schmidt-Kraepelin[2]; Katharina Schmähl[1]; Christoph Schütz[1]; Benedikt Morschheuser[1,3]; Ali Sunyaev[4]

[1]School of Business, Economics and Society, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nuremberg, Germany

[2]Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

[3]Gamification Research Group, Faculty of Information Systems and Applied Computer Sciences, University of Bamberg, Bamberg, Germany

[4]School of Computation, Information, and Technology, Technical University of Munich Campus Heilbronn, Heilbronn, Germany

Corresponding Author:
Manuel Schmidt-Kraepelin
Department of Economics and Management
Karlsruhe Institute of Technology
Kaiserstraße 89
Karlsruhe, D-76133
Germany
Phone: 49 72160844062
Email: manuel.schmidt-kraepelin@kit.edu

Abstract

Background: The phenomenon of procrastination refers to an individual's conscious decision to postpone the completion of tasks despite being aware of its adverse consequences in the future. Extant research in this field shows that procrastination is associated with increased levels of anxiety and stress and the likelihood of developing depression and calls for the development of suitable interventions that support individuals in making lasting positive changes to their procrastination behaviors. In parallel, practice has produced a plethora of behavior change support systems (BCSSs) that aim to provide a low-threshold, accessible alternative to in-person therapeutic approaches. Most of these BCSSs can be considered motivational BCSSs that combine functional, utilitarian components with hedonic and eudaimonic design elements to empower self-treatment. Although early studies have suggested the potential benefits of such BCSSs, research on understanding their specific design characteristics and support of individuals in self-treating procrastination is still in its infancy.

Objective: In response to this gap between practice and research, we aimed to analyze and systemize the multitude of practical design efforts in motivational BCSSs for the self-treatment of procrastination and identify the main design archetypes that have emerged.

Methods: We conducted a 3-step research approach. First, we identified 127 behavior change support apps for procrastination through a systematic screening process in the German and US Apple App Store and Google Play Store. Second, we systematically coded the identified apps in terms of the behavior change techniques targeted by their functional design and hedonic or eudaimonic design elements. Third, we conducted a 2-step cluster analysis to identify archetypes of motivational design in behavior change support apps to combat procrastination.

Results: A variety of motivational designs have been developed and implemented in practice, and our analysis identified five main archetypes: (1) structured progress monitor, (2) self-improvement guide, (3) productivity adventure, (4) emotional wellness coach, and (5) social focus companion. The identified archetypes target different psychological determinants of procrastination and successfully use a variety of hedonic and eudaimonic design elements that extend beyond the current state of research.

Conclusions: The results of our study provide a foundation for future research endeavors that aim to examine the comparative effects of motivational design archetypes and develop more effective interventions tailored to individual needs. For practitioners, the findings reveal the contemporary design space of motivational BCSSs to support the self-treatment of procrastination and may serve as blueprints that can guide the design of future systems. For individuals seeking support and health professionals treating procrastination, our study systemizes the landscape of apps, thereby facilitating the selection of one that best aligns with the patient's individual needs.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e65214) doi:10.2196/65214

KEYWORDS

procrastination; mobile apps; cluster analysis; motivational design; persuasive technology; behavior change support systems; behavior change techniques; mobile health; mHealth; mobile phone; artificial intelligence; AI

Introduction

Motivation

Procrastination, an irrational and malicious behavioral pattern defined as the intentional delay of important or necessary tasks despite knowing the negative consequences [1], affects nearly one-fifth of adults worldwide [2,3]. Despite the phenomenon often being described with a bit of humor, as evidenced by a popular Technology, Entertainment, Design Talk [4], it poses a serious threat to mental and physical health. Studies have demonstrated that procrastination can cause stress, depression, and anxiety [5-7], as well as raise the risk of cardiovascular diseases and hypertension [8]. Overall, those affected by procrastination experience a concerning risk of a decline in life quality [9].

Correspondingly, medical research has engaged in numerous treatment directions to help and support those affected [10,11]. Acknowledging that procrastination primarily results from a failure of self-regulation [2], which is influenced by both situational (eg, task characteristics, expected outcome, and delay) and motivational (eg, self-efficacy, self-belief, and impulsiveness) determinants [12], cognitive and behavioral interventions as well as their combination in cognitive behavioral therapy promise to be most successful [10,13-15]. However, traditional treatment of procrastination still suffers from low attendance and adherence rates compared to the amount of people affected [16]. In this regard, digitally supported treatment of procrastination using behavior change support systems (BCSSs) has gained attention as a low-threshold, accessible alternative to in-person therapeutic approaches [11,14,17].

BCSSs are systems "with psychological and behavioral outcomes designed to form, alter or reinforce attitudes, behaviors or an act of complying without using coercion or deception" [18]. They have been demonstrated to help restructure malicious behaviors that pose health threats [19], such as obesity [20-22], lack of physical activity [23,24], unhealthy nutrition [25-27], or smoking [28-30]. Through various behavior change techniques (BCTs) that target situational and motivational behavior determinants, such as goal setting and self-monitoring, context restructuring, reframing of beliefs, or the illustration of behavioral consequences [31], BCSSs can support a stepwise transition from health-threatening behaviors to novel behavior patterns and habits [32]. A recent review of primary studies in the field of BCSSs for procrastination [17] demonstrates that such BCSSs can evidently help raise awareness of distractions [33-35], reduce procrastination behaviors [34,36,37], and increase self-efficacy [37-39].

While these studies are still often tied to augmenting traditional cognitive behavioral therapy with digital applications under medical guidance, it is only in the last few years that BCSSs with motivational design approaches for the successful self-treatment of procrastination have gained traction [17]. This development can be attributed to the emerging research trend of motivational information systems (ISs) [40,41], which combine functional, utilitarian components (such as support in planning and goal setting) with hedonic (referring to a sense of pleasure [42]) and eudaimonic (referring to a sense of meaningfulness [42]) design elements. The latter aim to evoke positive experiences in using the system and performing the target behavior, such as joy, excitement, or curiosity, and satisfy motivational needs [40,41]. Initial studies show that motivational BCSSs for procrastination that combine educational content about procrastination stimuli with quiz and achievement mechanics can significantly reduce procrastination in unguided self-treatment [43]. Moreover, gameful storytelling in task planning and work on tasks can enhance feelings of self-control [44].

Research on motivational BCSSs for low-threshold and accessible self-treatment of procrastination is still nascent, but practical efforts have produced a plethora of (primarily mobile) apps that use motivational design approaches to support people in handling procrastination behaviors [44]. These include gamified task management platforms that integrate narrative design [45]; artificial companions that provide approval and help with planning and self-reflection [46]; and digital tools that use focus timers, unlockable collectibles, and team focus sessions [47], with several apps exceeding a million downloads. In this regard, there is great potential for research on motivational BCSSs for procrastination to build on and learn from these practical efforts [17] and investigate how different motivational design approaches affect the procrastination process as a basis to develop targeted recommendations for individuals seeking support in dealing with procrastination.

In this context, previous research has demonstrated that effectively designing health BCSSs requires not only experimenting with different BCTs that, in untargeted combinations, can cause counteracting mechanisms and result in highly ineffective solutions [48]. Rather, to provide an informed foundation for both researchers and designers on how to successfully approach the problem of procrastination with motivational BCSSs, it is crucial to identify and characterize commensurate combinations of BCTs and take account of their underlying theoretical mechanisms [48]. To this end, identifying the main *archetypes* of design that have emerged to target situational and motivational determinants of procrastination in practical applications is particularly valuable. Design archetypes represent a practice-oriented perspective on design knowledge that *abstracts* from single applications, which might sometimes combine BCTs randomly, to identify industry best practices. Thus, design archetypes can be regarded as complementary to conceptual, theory-driven design knowledge (eg, design principles and frameworks) in guiding developers when

XSL•FO
RenderX

designing and implementing motivational BCSSs by offering prescriptions for effective design [49]. For researchers, knowledge about design archetypes further serves as a feedback mechanism that shows whether and, if so, how research insights have been transferred to real-world systems and where gaps between research and practice remain [50].

Objectives

This study aimed to provide this foundation for both future research in investigating motivational BCSSs to combat procrastination and further practical efforts to design and select motivational BCSSs for self-treating procrastination by answering the following research question: *what are prevalent archetypes of motivational design in behavior change support apps for procrastination?*

To answer this research question, in line with previous studies that have systematized design in (mobile) health apps [50-52], we adopted a systematic screening and coding approach that built on the recognized Behavior Change Technique Taxonomy (BCTT) by Michie et al [53] and the intrinsic motivations for system use classification by Lowry et al [54] to form the basis for a 2-step cluster analysis according to Punj and Stewart [55]. Thereby, we identified clusters of motivational design approaches in prevailing motivational BCSSs for procrastination and then discussed the archetypes of motivational design characteristics for each identified cluster.

Consequently, our study systematically classified motivational design approaches in BCSSs to combat procrastination, providing 3 main contributions to research and practice. First, it constitutes the basis for future research efforts to investigate the comparative effects of motivational design archetypes on various situational and motivational determinants of procrastination, possibly paving the way for more targeted and individualized self-treatment recommendations through motivational BCSSs depending on the individual problem profile [1,56,57]. Second, it represents blueprints of design opportunities for developers of BCSSs to support the self-treatment of procrastination using motivational design approaches that increase system use and counteract dropout rates, a problem known from health-related digital applications [58]. Finally, it systemizes the plethora of available apps for both health professionals and individuals seeking support in handling procrastination as a basis to decide on a design archetype that best suits the patient's or individual's needs and preferences, thereby potentially increasing treatment adherence and counteracting the severe mental and physical health consequences of ongoing procrastination.

Methods

Overview

To answer our research question, we followed a systematic screening and coding approach of existing mobile apps, complemented by a 2-step cluster analysis [55] informed by previous work that has conducted archetype analyses of health-related apps and platforms [50-52]. Accordingly, our research method consisted of 3 main steps: database setup and screening, app coding, and cluster analysis (Figure 1 [50-55]).

Figure 1. Overview of the 3 research steps, applied research methods, and outputs of the methods.

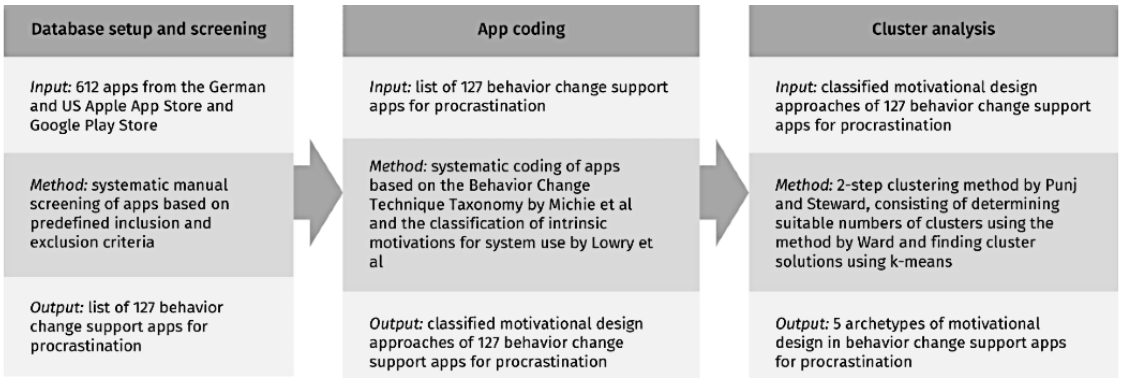

Database Setup and Screening

To set up the database of pertinent behavior change support apps to combat procrastination, we followed a systematic search strategy in the German and US Apple App Store and Google Play Store. The reason why we opted for these 2 platforms is that they represent the 2 largest app stores for iOS- and Android-based mobile phones, with the Google Play Store recording >110 billion app downloads and the Apple App Store reporting approximately 35 billion app downloads in 2023 [59]. Moreover, the United States is the most significant Western mobile app market in terms of downloads and consumer spending on mobile devices [60], and Germany represents the largest mobile app market in the European Union [61]. We searched for apps related to the terms "procrastination" or "Prokrastination" (the German equivalent) in March 2024, leading to an initial set of 1259 apps (Figure 2).

After removing duplicates, of the 1259 apps, 612 (48.61%) unique apps remained for the initial screening. Thus, we screened the descriptions of these 612 apps in the Apple App Store and Google Play Store against our predefined inclusion and exclusion criteria, listed in Textbox 1.

XSL•FO
RenderX

Figure 2. Overview of the app screening and selection process. DE: Germany; US: United States.

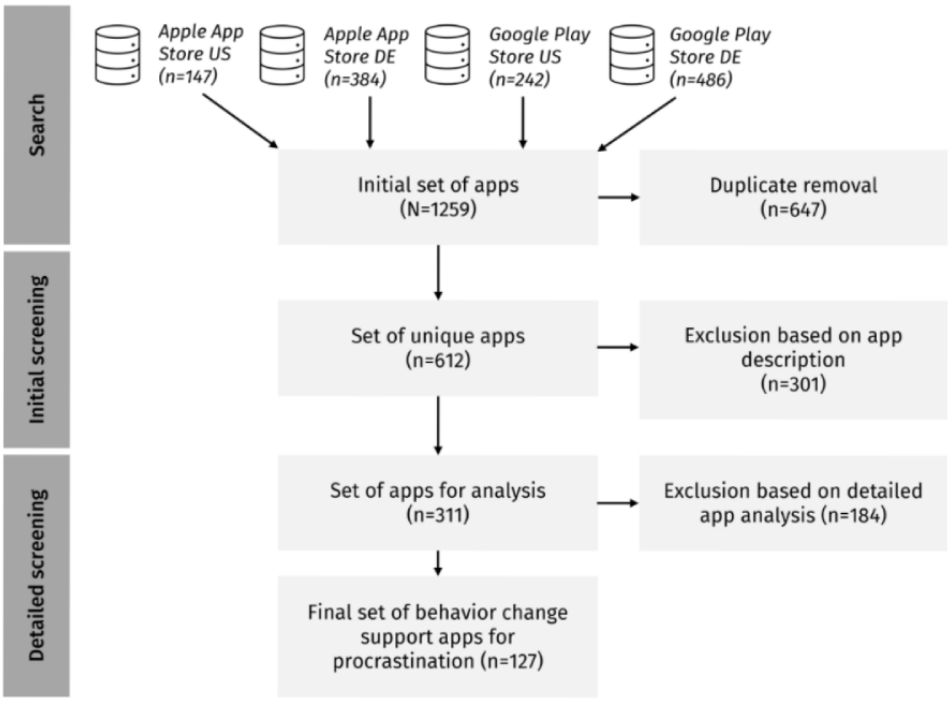

Textbox 1. Inclusion and exclusion criteria for the screening and selection process.

Inclusion criteria

- Topicality: maintained apps with regular updates
- Access: free apps
- Language: English or German
- Content focus: explicit focus on the problem of procrastination at work or in daily life
- App design: motivational behavior change support system with both functional and hedonic or eudaimonic design aspects according to the classification of intrinsic motivations for system use by Lowry et al [54]

Exclusion criteria

- Topicality: outdated apps flagged as not maintained anymore
- Access: paid apps and paywalls after installation
- Language: other languages
- Content focus: lack of focus on procrastination (eg, household organizers; microlearning apps; time-tracking apps for client projects; team planners; general meditation apps; general habit trackers; mood trackers; and treatment apps for other mental health issues such as attention-deficit/hyperactivity disorder, depression, or anxiety)
- App design: purely functional, utilitarian app design without hedonic design aspects

Specifically, concerning their topicality, we excluded apps flagged as not maintained by the app stores. We also excluded paid apps that were not accessible without upfront payment as we wanted to focus on apps that are readily accessible to a broad audience of people affected by procrastination without posing a payment hurdle that might inhibit their benefit compared to traditional treatment approaches. Moreover, we excluded apps that were in languages other than English or German (eg, Chinese or Russian). In terms of their content and focus, we explicitly searched for apps that targeted procrastination of individuals at work or in daily life by providing specific utilitarian functionalities or content to approach procrastination (ie, the postponement of important or necessary tasks) on a cognitive (eg, specific learning or reflection content or guidance related to procrastination) or behavioral (eg, features to engage in goal and task planning or organization or focused work) level. Conversely, we excluded apps that were not focused on this problem, such as microlearning apps and general meditation apps (with content entirely unrelated to procrastination); household organizers; time-tracking apps for client projects or team planners (focused on organizing multiple people instead of supporting individuals); mood trackers and general habit trackers (without any content or functionalities related to specific tasks); or treatment apps for other mental health problems such as attention-deficit/hyperactivity disorder, depression, or anxiety (that were not suitable for individuals without pathological

XSL•FO
RenderX

mental health issues). Moreover, as we focused on motivational BCSSs that are characterized by functional, utilitarian, and hedonic or eudaimonic design aspects [40,41], we excluded apps without any visible hedonic or eudaimonic design approaches, such as purely functional to-do lists, calendars, timers, or app blockers.

As a result of the initial screening process, of the 612 apps, 311 (50.8%) remained for the subsequent detailed analysis. In this step, we downloaded all the apps on an iPhone 15 Pro (for iOS apps) and a Google Pixel 4a (for Android apps) and analyzed their functionalities in detail against the predefined inclusion and exclusion criteria. During this step, we excluded an additional 59.2% (184/311) of the apps because they did not meet the inclusion criteria. Specifically, of the 184 apps, 34 (18.5%) were excluded because they had a paywall for use after installation (n=32, 94%) or could be accessed only via invitation (n=2, 6%); 16 (8.7%) were excluded because they were not usable (ie, they had been either removed from the app stores between screening in March 2024 and analysis in May 2024, n=6, 38%, or were buggy so that the main functionalities could not be used, n=10, 62%); 3 (1.6%) were excluded because they were only available in Turkish, French, or Chinese; 70 (38%) were excluded because they were not focused on the problem of procrastination; and 51 (27.7%) were excluded because they did not exhibit any hedonic design approaches. To decide whether an app entailed hedonic or eudaimonic design approaches, we juxtaposed them with our codebook based on the classification of intrinsic motivations for system use by Lowry et al [54] (Table 1) and excluded those apps that did not provide hedonic or eudaimonic features to evoke any of these defined motivations.

Table 1. Codebook of hedonic and eudaimonic design aspects based on the classification by Lowry et al [54].

Design aspect (based on the classification by Lowry et al [54])	Definition based on the Cambridge Dictionary	Exemplary design features (added after intercoder agreement check)
Fun and enjoyment	Experience of playfulness or pleasure	Mini games, role-play, digital companions, juicy animations and sounds, or shuffling or chance
Escapism	Experience of escaping pressures and avoiding an unpleasant or boring life	Fantasy, reframing, or virtual reality
Challenge	Experience of being faced with something that requires mental or physical effort and tests one's ability	Quests or daily challenges, leaderboards, or levels
Immersion	Experience of becoming completely involved in an environment or action	Story or narrative, nature sounds, or meditations
Curiosity	Experience of being eager to know or learn about something or encountering something interesting, rare, or unusual	Counterintuitive behavior or surprising events
Discovery	Experience of finding information, places, or objects that one did not know about before	Easter eggs, skill paths, unlockable content, or new content
Self-expression	Experience of expressing one's personality, emotions, or ideas	Personal profile, personal journal, or customization of an avatar
Knowledge expansion and self-development	Experience of learning, growing, changing, or advancing through one's own efforts	Skill trees, quizzes, comparison to past behavior, or self-reflection
Freedom and autonomy	Experience of being allowed to do, say, and think what one wants and make one's own decisions without being controlled or limited by anyone else	Choices, customization and configuration, or multiple paths and options
Companionship	Experience of enjoyment in spending time with other people or being bound with others that share similarities	Collaboration partner, social feed, digital coffee, or community
Approval	Experience of receiving positive opinions, being accepted, or being positively reinforced for one's actions	Motivational messages, positive feedback, likes and comments, or badges
Meaning	Experience of importance or value in an action or outcome	Narrative, coupling actions with benefits for others, altruism, or gifting

To ensure the reliability of our exclusion process, 3 authors took part in the detailed screening. One of them was a postdoctoral researcher in motivational ISs with 5 years of research experience in the field of BCSSs and motivational design, who set up the inclusion and exclusion criteria, defined the codebook, and led the screening and selection process. The 2 others were junior researchers with backgrounds in organizational psychology and business pedagogy. To ensure reliability, an intercoder agreement check was carried out after 40.2% (125/311) of the apps were analyzed. A total of 20% (25/125) of the analyzed sample was randomly selected and additionally analyzed by one of the other authors (specifically, the apps analyzed by the junior researchers were subjected to intercoder checking by the postdoctoral researcher, and the apps analyzed by the postdoctoral researcher were equally divided and checked by the 2 junior researchers). The intercoder agreement rate for inclusion and exclusion was 96% (κ=0.92), indicating nearly perfect agreement [62]. Any conflicts were resolved in a joint discussion, and the 3 authors progressed with the remaining apps. As a result, 127 apps remained as the final

XSL•FO
RenderX

set of behavior change support apps for procrastination for the following research step. The complete list of included and excluded apps can be found in Multimedia Appendix 1.

App Coding

For our app coding as a basis to identify motivational design archetypes, we opted for 2 existing classifications to code the utilitarian, functional design aspects and the hedonic or eudaimonic design aspects of the apps, which are both crucial for motivational ISs [40,41].

Regarding the functional design, we used the BCTT by Michie et al [53]. Michie et al [53] developed a taxonomy of 93 different BCTs organized in 16 hierarchical clusters: *goals and planning*, *feedback and monitoring*, *social support*, *shaping knowledge*, *natural consequences*, *comparison of behavior*, *associations*, *repetition and substitution*, *comparison of outcomes*, *rewards and threats*, *regulation*, *antecedents*, *identity*, *scheduled consequences*, *self-belief*, and *covert learning* [63]. The BCTT has proven its reliability and validity [31] and has already been widely used in studies to classify health applications and digital health interventions [19,64-68]. While alternative taxonomies for classifying design in BCSSs exist, such as the persuasive system design (PSD) model [69], which has also been extensively applied in research on health BCSSs [70,71], we decided to build on the BCTT for 2 reasons. First, unlike the PSD model, it incorporates cognitive techniques such as identity and self-belief. Given that previous IS research in the field of motivational BCSSs for self-treating procrastination has largely disregarded the use of cognitive techniques [17], it is all the more important to analyze whether such approaches have been incorporated into the practical design of motivational BCSSs. Second, PSD incorporates principles that are on the cusp of being considered utilitarian, such as liking (which refers to the aesthetic experience and, thus, is related to the hedonic experience of enjoyment [41]) or competition and praise (which can be considered hedonic elements aiming to evoke experiences of challenge and approval [54]). However, it does not encompass the full range of hedonic and eudaimonic aspects that could be included in apps, such as immersion, curiosity, or discovery [54]. Accordingly, we chose to use the BCTT as a taxonomy to classify the functional design of the analyzed apps complemented by a second classification explicitly targeting the hedonic and eudaimonic aspects of the apps. Consequently, for our analysis, we coded the apps based on the 16 overarching clusters of the BCTT; for example, an app was coded as meeting the BCT cluster *self-belief* if it provided functionalities for (1) verbal persuasion about capability, (2) mental rehearsal of successful performance, (3) focus on past success, or (4) self-talk, whereby the detailed definitions and examples for each BCT provided by the BCTT [63] served as our codebook.

In terms of the hedonic and eudaimonic design aspects of the apps [41], we opted for the classification of intrinsic motivations for system use provided by Lowry et al [54]. In their study, Lowry et al [54] list 11 intrinsic motivations for hedonic IS use based on a review of the IS literature: *fun and enjoyment*, *escaping pressures*, *challenge*, *immersion*, *curiosity*, *discovery*, *self-expression*, *obtaining knowledge*, *experiencing autonomy/freedom*, *peer companionship*, and *approval* [54]. According to their hedonic-motivation system adoption model, which has been extensively used to evaluate motivational BCSSs in the health domain [72-75], these hedonic motivations are pivotal to the adoption and continued use of systems that target hedonic experiences [54]. We defined each of these intrinsic motivations based on the Cambridge Dictionary and added *meaning* (ie, a sense of experiencing importance or value in an action [76]) as the 12th dimension to reflect the eudaimonic aspect of motivational ISs [40]. During the coding process, an app was coded to meet a specific dimension if it provided functionalities to evoke the described hedonic or eudaimonic experience (eg, of fun or challenge; Table 1).

In total, 3 authors performed the coding independently. Each app was downloaded, tested, and used to experience all features and functionalities. In this regard, we used each app intensively upon installation (ie, we clicked on all menu items, buttons, and learning content and used all available functionalities on the app to experience whether completing tasks or using features on the app resulted in system feedback or progress), and if the app included progression mechanics to unlock new features or notification and reminder mechanics, we also used the apps for several days to code all their features that might become visible over time. To ensure reliability in the coding process, we performed an intercoder agreement check after coding 30% of the apps, again randomly selecting 25 apps for intercoder checks and dividing the intercoder checks in a similar way to that for the screening check (ie, the apps coded by the junior researchers were subjected to intercoder checking by the postdoctoral researcher, and the apps coded by the postdoctoral researcher were equally divided and checked by the 2 junior researchers). The intercoder agreement rate ranged between 50% and 100% for the BCTs (with a mean agreement rate of 84%, SD 14%) and between 67% and 92% for the hedonic and eudaimonic design aspects (with a mean agreement rate of 82%, SD 6%), indicating good agreement [62]. As a result of the joint discussion, we streamlined our understanding of the BCTs (eg, defining that the social support technique only refers to support from other human beings and not from artificial agents on the app) and enlarged our codebook of hedonic and eudaimonic design aspects based on exemplary design features (Table 1) before progressing with the coding process. The final list of the 127 coded apps can be found in Multimedia Appendix 1.

Cluster Analysis

We used cluster analysis to derive meaningful archetypes of app designs in the third step of our methodology. To do so, we followed the 2-step clustering approach proposed by Punj and Stewart [55] for 2 main reasons. First, this approach combines hierarchical clustering with iterative partitioning algorithms, thereby attempting to overcome the respective weaknesses of both approaches (eg, many partitioning algorithms, such as k-means, perform better than hierarchical algorithms but require an a priori definition of the number of clusters). Second, this methodology has been successfully applied to similar clustering tasks and has been shown to be useful for identifying design archetypes in various contexts [50,51,77].

In line with the approach by Punj and Stewart [55], our clustering analysis included 2 steps. In the first step, we used

XSL•FO
RenderX

the method by Ward as a hierarchical clustering method to determine preliminary solutions and identify potential candidate numbers of clusters. Our coding included only binary data (ie, a design aspect, meaning each of the 16 BCT clusters of the BCTT [53] and each of the 12 hedonic or eudaimonic design aspects based on the classification by Lowry et al [54], is either present or absent). Thus, we chose the Euclidean squared distance as the similarity measure [50]. The dendrogram resulting from applying the method by Ward to our dataset indicated that 2 to 5, 7, 9, 11, and 12 were suitable candidate numbers of clusters. By examining the scree plot using the elbow rule [78], we narrowed down our selection to the 2-, 4-, 5-, 7-, 9-, or 12-cluster solution. On the basis of these candidate numbers, we conducted k-means clustering as an iterative partitioning algorithm. The 7-, 9-, and 12-cluster solutions produced by k-means all contained clusters with only 4 or even fewer apps, making reasonable interpretation of these solutions difficult. To decide on one of the remaining cluster solutions, 2 researchers manually compared their explanatory power by investigating differences and similarities in the BCTs and hedonic and eudaimonic design aspects included in our coding. First, we decided against the 2-cluster solution because the clusters were too heterogeneous in design aspects, making it impossible for us to derive meaningful interpretations. This left us with the 4- and 5-cluster solutions as candidates. After carefully analyzing both cluster solutions, we concluded that they were similar in many aspects but that one of the clusters in the 4-cluster solution was further divided into 2 distinct and informative clusters within the 5-cluster solution. Thus, we decided that the 5-cluster solution had the highest explanatory power and was the most suitable for deriving meaningful archetypes within this study. The cluster analysis results and the cluster solutions are included in Multimedia Appendix 2.

Results

Overview

The final apps (N=127) that we included were fairly homogeneous regarding their categories, with most (n=122, 96.1%) categorized as productivity apps. In addition, 1.6% (n=2) were listed as lifestyle apps, 1.6% (n=2) were labeled as health and fitness apps, and 0.8% (n=1) were subsumed under the category of education, suggesting that, from a developer's perspective, apps that support behavior change to combat procrastination are seen as primarily organizational tools rather than mental health support apps. Overall, the analyzed apps were well rated (mean 4.39, SD 0.48 stars), with 1.6% (n=2) of the apps with a rating <3 stars, 11.8% (n=15) with a rating between 3 and 4 stars, 23.6% (n=30) with a rating between 4 and 4.4 stars, 36.2% (n=46) with a rating between 4.5 and 4.9 stars, and 7.9% (n=10) with a rating of a full 5 stars. A total of 18.9% (n=24) of the apps that we included had not been rated at the time of analysis.

Archetypes of Motivational Design in Behavior Change Support Apps to Combat Procrastination

Overview

Through our cluster analysis, we identified 5 distinct clusters of motivational behavior change apps to combat procrastination, which subsume a different number of apps. Specifically, cluster 1 was the largest cluster, with 51.2% (n=65) of the 127 apps, followed by cluster 2 and cluster 5, with 15.7% (n=20) of the apps each. Cluster 3 contained 10.2% (n=13) of the apps, and the smallest cluster, cluster 4, comprised 7.1% (n=9) of the apps. The apps in the clusters were rated fairly equally (mean 4.33, SD 0.55 stars for cluster 1; mean 4.30, SD 0.45 stars for cluster 2; mean 4.44, SD 0.31 stars for cluster 3; mean 4.42, SD 0.34 stars for cluster 4; and mean 4.55, SD 0.33 stars for cluster 5), with the apps in clusters 1 and 2 having the least positive ratings and the apps in cluster 5 having the most positive ratings. The clusters differed considerably in the BCTs and the hedonic and eudaimonic design aspects implemented in the contained apps (Figure 3). From the profiles of the clusters in terms of their motivational design, we can derive a motivational design *archetype* (ie, an abstract representation of a cluster), which helps understand how the clusters differ from each other. In the following sections, we will explain the 5 motivational design archetypes represented by the 5 clusters we found.

Figure 3. Graphical overview of the different clusters and the percentage of apps in each cluster that target specific behavior change techniques (BCTs; through functional design) and hedonic and eudaimonic design aspects.

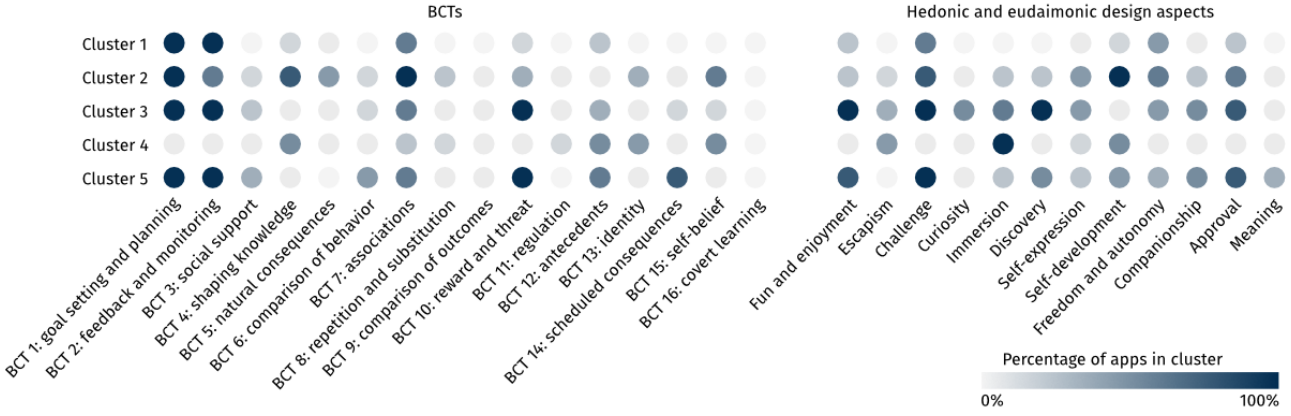

XSL•FO
RenderX

Archetype 1: Structured Progress Monitor

Archetype 1 represents an app used predominantly for managing personal tasks, goals, or habits. To establish goals and develop a plan of action (BCT 1), users can set tasks with a high degree of efficiency and ease. Furthermore, they can stipulate precise time frames for completion, the frequency of repetition, and the time of day at which the task is to be carried out. As a means of prompting users to engage with the app (BCT 7), notifications are sent to remind them to follow their task lists. The provision of detailed statistics and graphics for self-monitoring (BCT 2) represents a further key feature of the *structured progress monitor*. The visualization of one's development, presented in the form of progress bars, grades, badges, points, or streaks, illustrates past achievements and the individual capacity to complete tasks successfully. These mechanisms elicit a sense of challenge in the user, prompting them to sustain the desired behavior and, consequently, adhere to the established goals over time. They represent the primary hedonic design aspect of the *structured progress monitor*, distinguished by a transparent and uncluttered design devoid of any significant hedonic or eudaimonic elements beyond the element of challenge. Consequently, it can be characterized as a primarily functional assistance system. An example of the *structured progress monitor* is the Noverdue app [79]. Figure 4 [79] illustrates how individuals can create tasks with notifications and due dates (Figure 4A [79]); personalize the reminder function (Figure 4B [79]); and self-monitor their progress on a dashboard that provides a rating, thereby challenging the users to improve their performance (Figure 4C [79]).

Figure 4. Noverdue as an exemplary app of the structured progress monitor archetype, with (A) the customizable creation of tasks, (B) the reminder function, and (C) the dashboard visualizing personal progress.

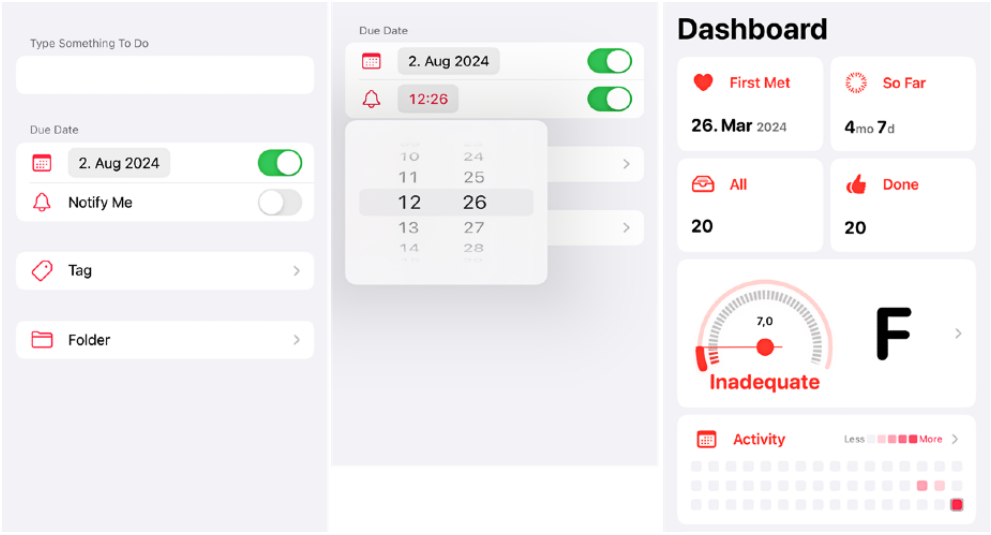

Archetype 2: Self-Improvement Guide

Archetype 2 provides the means to restructure detrimental behaviors and introduce novel habits (BCT 1) within one's personal life. An artificial intelligence (AI)–powered chatbot offers tailored assistance and guidance (BCT 4) on overcoming obstacles and navigating questions regarding habit implementation while emphasizing the positive consequences of behavior change (BCT 5), for example, for personal well-being. This key mechanic is complemented by functionalities to track (BCT 2) and repeat (BCT 8) new habits and receive personalized reminders (BCT 7). In-app journaling encourages introspective reflection on one's behavior, facilitating the development of self-belief (BCT 15). Consequently, the *self-improvement guide* assists users in articulating their desired personal development and provides guidance on how to achieve these life goals through the use of a digital companion, which offers continuous approval for positive self-development. The elicited feeling of self-development, in conjunction with challenge-oriented elements such as streaks, serves as the primary hedonic and eudaimonic mechanism that facilitates users' maintenance of new habits. An illustrative example of the *self-improvement guide* is Dreamfora [46]. Figure 5 [46] illustrates how users can establish new, overarching objectives and dreams that they aspire to attain through self-development (Figure 5A [46]). The figure also depicts how the AI chatbot dispenses knowledge and assistance regarding forming new habits (Figure 5B [46]), thereby facilitating the achievement of the aforementioned life goals. Furthermore, it demonstrates how users receive ongoing approval for their successful actions (Figure 5C [46]).

XSL•FO
RenderX

Figure 5. Dreamfora as an exemplary app of the self-improvement guide archetype, with (A) opportunities to define long-term goals and dreams, (B) the artificial intelligence assistance that breaks them down into tangible tasks and explains their relevance, and (C) ongoing approval for successful actions.

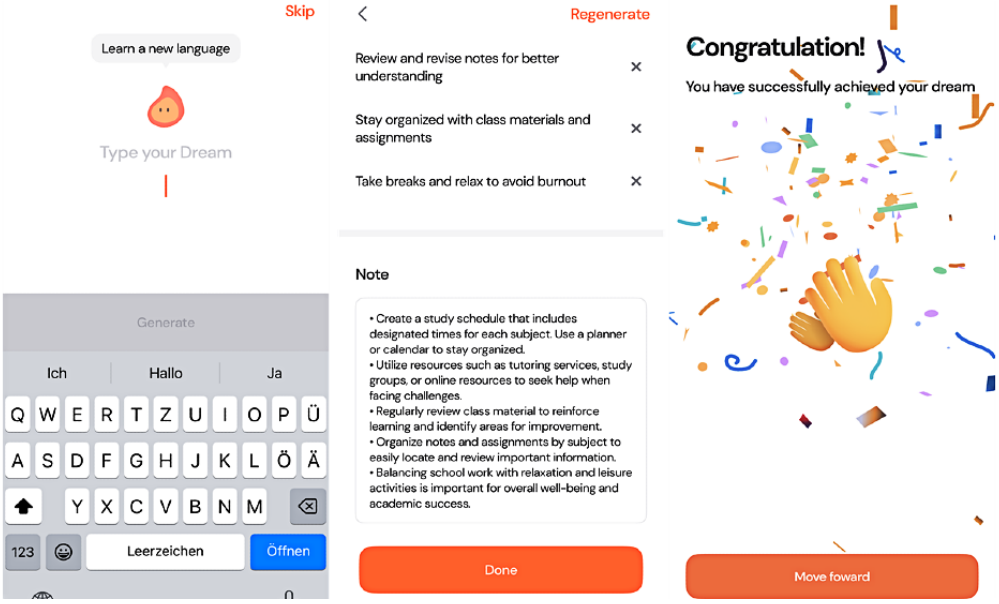

Archetype 3: Productivity Adventure

Archetype 3 is distinguished by its comprehensive hedonic and eudaimonic design. Playfully designed storylines, mini games, unlockable levels, and customizable characters elicit feelings of enjoyment, challenge, curiosity, discovery, and self-expression among app users. By transforming the mundane and occasionally onerous task completion into a digital experience, the seriousness associated with task fulfillment is diminished. Digital characters that accompany the individual in the digital journal facilitate this process by offering a sense of companionship and providing approval. As a foundation for digital progression, the productivity adventure is predicated on 5 principal BCTs: the user defines their to-dos (BCT 1) and monitors the completion of these tasks in a step-by-step manner (BCT 2), aided by a timer that schedules focus sessions (BCT 12) and notifies upon successful completion (BCT 7). Should the user maintain their focus for the requisite period, the app will instantaneously reward them with, for example, the ability to access a new level (BCT 10). An example of this archetype is the Focus Quest app [80]. Figure 6 [80] illustrates the integration of goals and task fulfillment into a fictional storyline in interaction with a digital character (Figure 6A [80]), which challenges the user to maintain focus for a designated period (Figure 6B [80]) to obtain digital rewards and progress in the story (Figure 6C [80]).

Figure 6. Focus Quest as an exemplary app of the productivity adventure archetype, with (A) a fictional storyline that entails the user's character and other digital characters, (B) focus and task challenges, and (C) digital rewards and levels that support progress in the story.

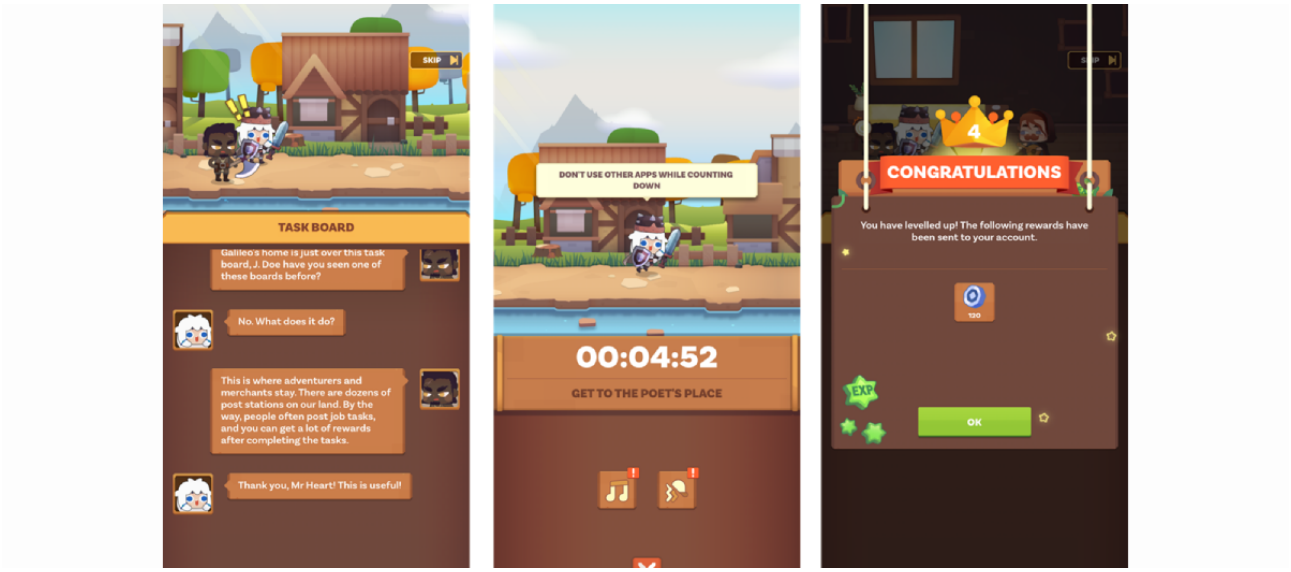

XSL•FO
RenderX

Archetype 4: Emotional Wellness Coach

Archetype 4 represents an app that furnishes users with a variety of content, including quotes, images, videos, and audio material, with the objective of enhancing motivation and emotional regulation, particularly in the context of perceived pressure. Positive affirmations, relaxing audio sessions, and motivational speeches by role models reinforce the user's self-belief (BCT 15) and cultivate a positive outlook on one's future identity (BCT 13). In terms of hedonic and eudaimonic design, the *emotional wellness coach* contributes to the attainment of a positive state of mind in which the user can escape the stressful reality and immerse themselves in meditative sessions, thereby preparing themselves mentally for pursuing their goals. This ultimately facilitates the process of intense cognitive self-development. An app that exemplifies the *emotional wellness coach* is Greatness [81], depicted in Figure 7 [81]. Individuals may select from an array of productivity- and well-being–oriented meditation programs (Figure 7A [81]). These programs comprise distinct coaching sessions to foster motivation and promote emotional well-being (Figure 7B [81]). Moreover, the app provides guidance and facilitates mental preparation and rehearsal through immersive meditation exercises (Figure 7C [81]).

Figure 7. Greatness as an exemplary app of the emotional wellness coach archetype, with (A) a variety of meditation programs to choose from, (B) coaching sessions on motivation and emotional well-being, and (C) immersive meditation exercises.

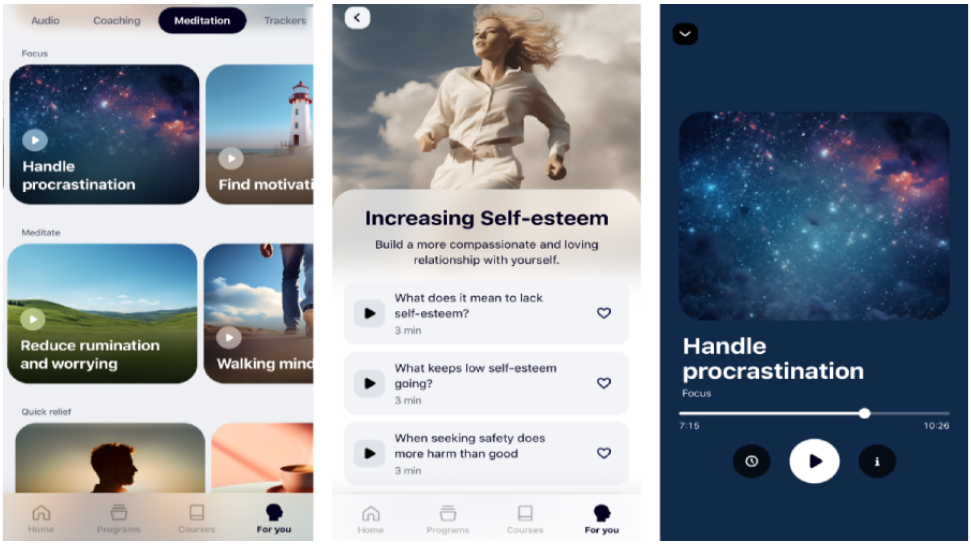

Archetype 5: Social Focus Companion

The primary operational mechanism of archetype 5 is a focus timer that inhibits the use of alternative apps and external distractions, thereby facilitating the control of environmental stimuli (BCT 12). The user establishes their to-dos along with an intended time interval (BCT 1). Upon the expiration of the designated time and the maintenance of focus, the user is granted the aforementioned scheduled rewards. Conversely, upon the occurrence of a distraction, the user is subjected to a digital penalty (BCTs 10 and 14). The primary hedonic design element is an in-app character situated in a playful and enjoyable environment. This character is developed and behaves in accordance with the user's focus sessions. That is, it either progresses and flourishes through the digital rewards or suffers from the digital penalties. This dynamic challenges the user to complete focus sessions for the benefit of their digital companion. In this regard, the digital entity also offers feedback (BCT 2) indicating approval for successful work. Furthermore, in contrast to the other archetypes, the *social focus companion* allows for interaction with other users, facilitating social support in focusing together (BCT 3) and comparison with the behavior of others (BCT 6), fostering a feeling of companionship. While unlocking novel rewards also elicits a feeling of discovery, a distinctive eudaimonic feature of the *social focus companion* is the altruistic benefit of one's behavior, evoking a sense of meaning. By adhering to the designated focus time, one not only rewards the character within the app but also achieves a positive impact outside the app. The exemplary app of the *social focus companion*, named *Focus Dog: The Productivity Timer* [82] (Figure 8 [82]), features a fictional dog that requires nourishment, which can be obtained by unlocking donuts through the use of coins (Figure 8A [82]). Individuals are rewarded with coins for maintaining focus, and users can compare their performance on a ranking list (Figure 8B [82]). Furthermore, in addition to benefiting the fictional dog, the coins can be used to purchase meals for canines in need (Figure 8C [82]).

XSL•FO
RenderX

Figure 8. Focus Dog as an exemplary app of the social focus companion archetype, with (A) the fictional dog gaining donuts through focus time, (B) social comparison on a ranking list, and (C) the possibility to donate in-app coins to purchase meals for canines in need.

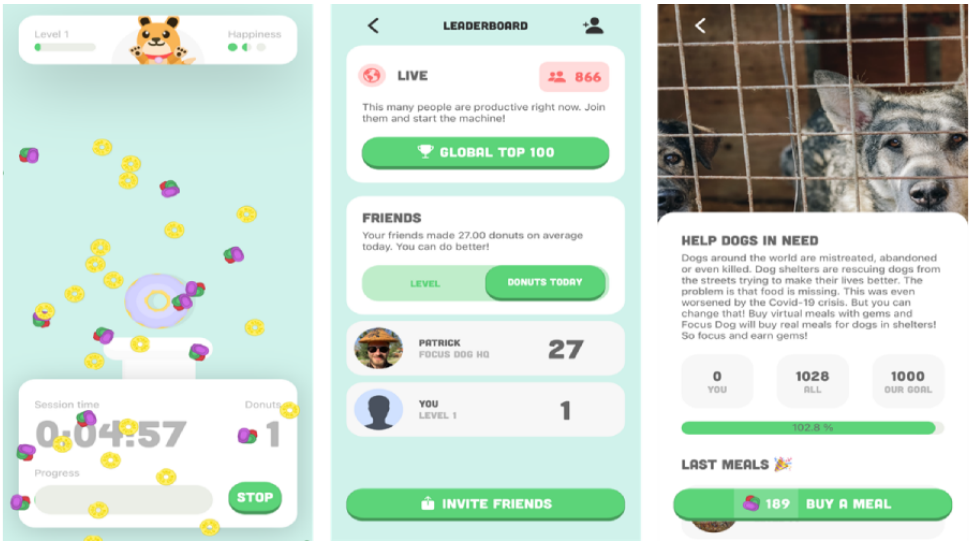

Discussion

Principal Findings

Overview

Our study sheds a novel light on how practice has approached the design and implementation of motivational BCSSs for the unguided self-treatment of procrastination as a low-threshold alternative to professional medical treatment programs for those affected. Generally, while research on motivational BCSSs for procrastination is still nascent [17], we can observe that practice has already produced a plethora of behavior change support apps with various behavior change support techniques and hedonic or eudaimonic design aspects implemented to both provide functional support and elicit positive experiences in managing procrastination. In response to our research question, we identified 5 archetypes of motivational design in prevalent apps: the *structured progress monitor*, the *self-improvement guide*, the *productivity adventure*, the *emotional wellness coach*, and the *social focus companion*. In reflecting on our results, we can discuss 3 main observations that entail important implications for future research and practice in the field.

Observation 1: The Relationship Between Motivational Design Archetypes and Psychological Determinants of Procrastination

Upon juxtaposing the identified archetypes of motivational design with procrastination theory, it becomes evident that the designs appear to target different determinants of procrastination, specifically, outcome expectancies, outcome value, delay, and impulsiveness [2,12].

In particular, the *structured progress monitor* and the *productivity adventure* both appear to target primarily outcome value, that is, the desirability of a task or activity for an individual [2,12], and delay, which is a temporal indicator of the proximity of the outcome value's realization [2,12]. In this regard, the *structured progress monitor* emphasizes task organization, self-monitoring, and the illustration of progress and self-development. From a theoretical perspective, the organization and planning of tasks (as a form of goal setting [83]) and adherence to a self-set schedule as a self-regulatory technique [2] affect procrastination through 2 mechanisms: temporal discounting and diminishing returns [12]. Specifically, this implies that the successful completion of each individual task that arises from a divided overall objective provides a similar level of satisfaction concerning the need for competence [84] to that for the entire project. Consequently, the overall outcome value accumulates as a determining factor in procrastination [12]. Simultaneously, individual tasks are more proximal than a distant deadline for an overall project, which reduces delay and, thereby, enhances the power of the accumulated outcome value [12]. The *productivity adventure*, similarly to the *structured progress monitor*, appears to seek to enhance the perceived outcome value of a task or action, albeit through a markedly different approach. By advancing in a narrative, obtaining digital rewards for completing tasks and focus sessions, and stimulating individuals' curiosity and desire for exploration and discovery with unexpected events and unlockable paths and content, the archetype—in addition to merely subdividing a goal and illustrating progress—reframes task values by offering rewards and digital progress, thereby reducing task aversiveness [2]. Task unpleasantness and a lack of interest or boredom have been identified as reliable determinants of procrastination [10], reducing outcome value. Conversely, the satisfaction of autonomy needs through opportunities for exploration and discovery of a story or narrative [85] has been shown to counteract procrastination [86,87]. In addition, from a behavioral perspective, immediate (digital) rewards provide timely reinforcement of behavior [88-90], thereby reducing the delay of outcome value, particularly for individuals who are prone to immediate gratification [90].

In contrast, the *self-improvement guide* and the *emotional wellness coach* primarily aim to target outcome expectancies, that is, the conviction that a task can be completed successfully [2,12]. In this context, the *self-improvement guide* emphasizes the education of individuals regarding the consequences of procrastination and the means to counteract it. It provides

XSL•FO
RenderX

support and approval for self-reflection, personal development, and the establishment of self-belief, thereby enabling those affected by procrastination to realize new habits (ie, behavioral routines vital in inhibiting procrastination [2,10]). From a theoretical perspective, research has emphasized that irrational beliefs, unrealistic expectations, and low self-esteem present critical risk factors for procrastination and a phenomenon called self-handicapping [10], which is defined as engagement in unrelated activities or tasks that are likely to fail [91]. The resolution of such cognitive barriers and the reinforcement of self-efficacy represent a crucial element in forming outcome expectancies [92]. In turn, this fosters a reduction in procrastination behaviors [12]. In a similar albeit different manner, the *emotional wellness coach* uses a cognitive approach to procrastination, eschewing the behavioral elements present in the other archetypes. This approach is centered on enhancing self-belief and identity through immersive formats and meditation while mitigating the pressure experienced by individuals. In this regard, procrastination theory demonstrates that pressure and stress resulting from perfectionism and fear of failure promote procrastination as a separate factor from self-esteem and self-efficacy [2,15]. The cognitive restructuring of perfectionist beliefs and mental contrasting sessions that highlight the steps necessary to achieve the goal rather than remaining in a passive mental state have been demonstrated to be effective therapeutic approaches to this problem [10]. These approaches are implemented in the emotional wellness coach as digital coaches or role models.

Finally, the *social focus companion* is primarily concerned with reducing distractibility and enhancing stimulus control. That is to say, it is focused on volitional rather than motivational factors [1]. Both are essential for strengthening self-control [2,92] and reducing impulsiveness, which is defined as a preference for short-term outcomes over future rewards [90] and presents a key determinant of procrastination [12]. To this end, the primary functionalities of the *social focus companion* are focus timers that mitigate distractions coupled with scheduled digital incentives and penalties for successful or unsuccessful focus sessions, respectively. These sessions are enhanced with playful environments, digital companions or assets, and linking of in-app achievements to meaningful additional outcomes. Moreover, the social comparison and support mechanisms inherent to the *social focus companion* reinforce the normative influence to focus on the task, which in turn reduces impulsiveness even further [14]. In addition, through a process known as *fusing*, immediate social gratification increases the perceived value of the outcome [2,10].

In summary, it can be observed that the *self-improvement guide* and the *emotional wellness coach* primarily target outcome expectancies as determinants of procrastination, whether through the enhancement of self-efficacy and self-belief or the reduction of perfectionism and pressure. The *structured progress monitor* and the *productivity adventure*, in turn, primarily seek to enhance outcome value by subdividing goals or reducing task aversiveness and decreasing delay through temporal discounting and immediate reinforcement. Finally, the *social focus companion* is concerned with reducing impulsiveness through stimulus control and normative influence. Therefore, from a procrastination theory perspective, the 5 archetypes address all the primary factors that contribute to procrastination [2,12].

Observation 2: BCTs in Motivational BCSSs for Procrastination Compared to Traditional Treatment Strategies

Research has suggested that BCSSs may provide an efficacious and low-threshold alternative to traditional therapeutic treatments for procrastination [11,14,17]. In this regard, it has been demonstrated that studies on BCSSs for procrastination rely heavily on behavioral treatment strategies, including self-supervision, planning and prioritization, time management, and goal setting [17]. However, cognitive approaches appear to be underrepresented and are predominantly used as an extension of cognitive behavioral therapy with professional guidance [17], also referred to as internet-based cognitive behavioral therapy [39,93,94]. Consequently, they are not yet widely regarded as a stand-alone self-treatment option. Our analysis of practical approaches to BCSSs to combat procrastination similarly demonstrates an emphasis on behavioral strategies, most importantly goal setting and planning (BCT 1), feedback and monitoring (BCT 2), and associations and cues (BCT 7). In this context, it is noteworthy that several archetypes, particularly the *productivity adventure* and the *social focus companion* but also the *emotional wellness coach*, incorporate functionalities to alter the natural environment and avoid distractions, targeting antecedents (BCT 12), whereas research has thus far primarily overlooked the potential of stimulus control, or volitional influences, in self-treatment BCSSs [17].

Furthermore, it can be observed that both the *self-improvement guide* and the *emotional wellness coach* use cognitive strategies, specifically knowledge shaping (BCT 4), identity building (BCT 13), and self-belief strengthening (BCT 15), through the use of predefined information and meditation sessions conducted by coaches or through AI companions that generate personalized recommendations based on the individual's goals, challenges, and reflections. Although intelligent conversational agents can serve as a valuable alternative for individual coaching in other health care areas [95-97], their potential in research on BCSSs for procrastination remains unexplored [17,98]. Our analysis demonstrates that practice has already taken the initiative to use such approaches, which can potentially enable the tailoring of cognitive strategies to individuals and their specific problem profiles [1,56,57] for the purpose of self-treating procrastination. Scholars are encouraged to build on these implementations to further explore their potential for the personalized treatment of procrastination.

Nevertheless, when considering the therapeutic treatment of procrastination, it becomes evident that the prevailing self-treatment archetypes lack the incorporation of other cognitively oriented BCTs that have been demonstrated to exert a beneficial influence on procrastination. These include the comparison of outcomes through the reflection on the advantages and disadvantages of the behavior in question as well as the visualization of prospective outcomes (BCT 9) and emotional regulation (BCT 13) [99,100]. It remains unclear whether nonhuman artificial agents can entirely supplant

XSL•FO
RenderX

therapeutic guidance, particularly regarding emotional regulation, which necessitates a personal bond [101] that is frequently impeded by cognitive and emotional barriers in the context of artificial systems [102,103]. Furthermore, medical research has demonstrated that group treatment and social strategies, including social support (BCT 3), comparison of behavior (BCT 6), and covert learning from others' experiences (BCT 16), are more effective in treating procrastination than individual approaches [104]. In this regard, existing research efforts [17], as well as our identified practical archetypes, still demonstrate considerable potential for harnessing the power of social support. While the *social focus companion* notably uses joint focus sessions as a form of common goal setting and at least enables social comparison on ranking lists, such mechanics are primarily competitive [105] and oriented toward normative pressure that can evoke feelings of guilt or shame [106] rather than being socially supportive, as in the form of group counseling sessions and mutual exchange about the challenges and personal experiences with procrastination [107]. The combination of cognitive reflection and self-belief strategies with social support and covert learning may prove an effective approach to the self-treatment of procrastination. In particular, this approach may offer significant benefits for individuals facing challenges related to false expectations, irrational beliefs, low self-esteem, and other stable dispositions that require intensive cognitive restructuring [10] rather than behavioral support for task organization and stimulus control.

Observation 3: Hedonic and Eudaimonic Design of Motivational BCSSs for Procrastination

With regard to hedonic and eudaimonic design aspects of motivational design archetypes, it is evident that dominant designs are either centered on digital environments that are designed for enjoyment, featuring fictional characters and stories that foster discovery and provide a sense of approval (the *productivity adventure* and the *social focus companion*), or on providing means for users to escape pressures, engage in meditation, and support their self-development (the *self-improvement guide* and the *emotional wellness coach*). It is noteworthy that practice has used a more diverse array of hedonic and eudaimonic design elements than research efforts in the field of motivational BCSSs for procrastination, which have primarily focused on challenge-oriented elements such as points [43,108,109] and badges [109,110], similarly to the *structured progress monitor* [17]. Furthermore, in comparison to most existing works on motivational BCSSs for other health dimensions beyond procrastination, such as physical activity [23,50] or therapy adherence [111], our identified practical archetypes extend beyond the use of challenge-oriented elements commonly observed in classic gamification [112]. They incorporate exploration, narrative design, and immersive elements, which have been recently proposed by research on motivational BCSSs to enhance the hedonic and eudaimonic experience effects beyond those typically associated with challenge-oriented approaches [113,114].

However, when considered together with the treatment strategies, cognitive approaches appear to be at odds with fictional environments and narratives in the identified archetypes. Thus far, motivational design archetypes seem to be focused on either using playful reframing to enhance hedonic and eudaimonic experiences during behavioral strategies or on fostering feelings of self-development and escapism from pressures during cognitive strategies. The latter naturally focus on real-life problems and the individual rather than a fictional story. In this regard, research on motivational BCSSs in other areas indicates that narrative design can be effectively combined with self-reflection exercises, which may enhance the beneficial hedonic and eudaimonic experiences elicited by both approaches [115-117].

Furthermore, novel motivational BCSSs that extend beyond the smartphone technology examined in this study, such as virtual reality applications, may offer even greater potential to integrate immersive narratives with cognitive treatment approaches, a relaxing environment, and feelings of actual self-development [118,119]. While those developments are still in their infancy, both research and practice can benefit from exploring such novel technological advancements in designing effective motivational BCSSs for procrastination.

Implications

The principal findings of our study have significant implications for further research and practice. From a research perspective, our findings identified 5 main archetypes of motivational design in BCSSs for procrastination. These archetypes constitute a novel perspective on design knowledge for motivational BCSSs that abstracts from single, potentially ineffective real-world systems [48] and describes industry best practices that have prevailed in the design of motivational BCSSs for procrastination. For researchers, these archetypes provide a unique opportunity to obtain a holistic picture of how BCTs and hedonic and eudaimonic design elements have predominantly been instantiated in apps for procrastination. Researchers may juxtapose these findings with their own research endeavors, thereby potentially identify mismatches between current research foci and industry best practices. Consequently, they may discover new research opportunities or necessities to reallocate resources toward investigating the archetypal design configurations as possibly promising combinations of BCTs and motivational design that extend beyond the predominant focus on behavioral approaches [17].

Moreover, our observations indicate that the different motivational design archetypes may be suitable for targeting different determinants of procrastination [12]. Specifically, we found that apps in the *self-improvement guide* or *emotional wellness coach* archetypes primarily target outcome expectancies, whereas the *structured progress monitor* and *productivity adventure* seek to enhance outcome value and the *social focus companion* focuses on reducing impulsiveness [12]. While the archetypes were identified based on apps in the US and German markets and, therefore, predominantly reflect Western design approaches to target procrastination, research indicates that the primary determinants of procrastination, such as outcome expectancies, task aversiveness, and impulsiveness or self-control, are applicable and relevant across cultural contexts that are different from the Western one, including those of countries such as China, India, Israel, and Turkey [120-123]. Consequently, our design archetypes constitute a valuable

XSL•FO
RenderX

foundation for further research, not only within the Western cultural context but, more broadly, to explore the potential of BCSSs as a low-threshold self-treatment approach for the global phenomenon of procrastination [2,3] across cultures. For example, given that individuals differ in their procrastination profiles and which of the determinants pose the most significant problem for them [1,56,57], there is great potential to examine how these motivational BCSS archetypes can serve as the basis for targeted self-treatment strategies for different individuals. The existing research has already called for implementing more individualized treatment strategies for procrastination [1,44]. Concurrently, the extant motivational BCSS research has demonstrated that, in addition to the treatment approach, individuals also express preferences for various hedonic and eudaimonic design elements that have the potential to influence their experiential impact [124-126]. Our findings provide a foundation for further research into how the different motivational design archetypes, in terms of both their treatment focus (behavioral, cognitive, or a combination) and their hedonic and eudaimonic design approaches, affect individuals with varying procrastination problem profiles, dispositions, situational contexts, and motivational preferences. In this way, targeted recommendations can be developed for individuals affected by procrastination.

Furthermore, our 5 motivational design archetypes present a blueprint for the effective practical development of BCSSs for procrastination. By demonstrating how designers can use and combine hedonic and eudaimonic design elements to enhance the efficacy of behavioral and cognitive BCTs in digital applications [40,41] while concurrently addressing dropout rates [58] and how designers can build on our archetypes as established configurations of design elements [49] rather than experimenting with inefficient combinations of BCTs [48], our study paves the way for an even more diverse array of motivational BCSSs to combat procrastination. Potential avenues for exploration include the integration of AI-based conversational agent approaches to enhance cognitive treatment strategies [95-98] and the use of virtual reality technology to unify cognitive and behavioral treatment and playful and narrative designs [127] with the objective of enhancing both the functional and experiential aspects of motivational BCSSs for procrastination. From the perspective of health professionals and affected individuals, our findings provide a basis for selecting a motivational design archetype that best suits the patient's or individual's challenges, needs, and preferences. Once this has been done, a suitable, existing app can be chosen from the results of our systematic cluster analysis. Consequently, our study provides a foundation for potentially enhancing the treatment adherence of individuals seeking mobile apps to assist them in addressing procrastination (as opposed to, eg, experiencing frustration due to the absence of effects from attempting to use an emotional wellness coach when the primary issue lies in stimulus control and distractibility). Therefore, ultimately, our findings assist in counteracting the significant mental and physical health consequences associated with procrastination [5-9].

Limitations and Future Research

It is essential to consider some critical limitations in interpreting the findings of this study, which have their roots in the methodological choices made.

First, we restricted our examination of motivational BCSSs against procrastination to smartphone-based apps. Despite this decision, our search and screening process demonstrates that we identified a multitude of motivational behavior change apps for procrastination. However, we recognize that there are other motivational BCSSs for the self-treatment of procrastination that rely on alternative technology, such as web applications [128], which were excluded from our cluster analysis. Furthermore, while the Apple App Store and the Google Play Store represent the 2 largest app stores for iOS- and Android-based mobile phones, there are other app stores for Android-based mobile phones, such as the Microsoft Store, that were not included in our search. Similarly, it is crucial to acknowledge that user settings, filters, and preferences in app store searches may impact the results obtained by entering identical search terms on disparate devices. Consequently, some users may not be able to access all the apps that we screened for use on their smartphones. In addition, it should be noted that our search and screening were limited to German or English search terms and apps because the United States is the most significant Western mobile app market in terms of app downloads and consumer spending on mobile devices [60] and Germany represents the largest mobile app market in the European Union [61]. As a result, it is possible that we overlooked valuable apps that were designed for the Chinese market and only accessible and functional with Chinese-language proficiency.

Furthermore, while our attention was directed toward motivational BCSSs due to their hedonic and eudaimonic design characteristics that can augment positive experiences [40,41] and, in turn, potentially increase self-treatment adherence, it is important to recognize that, within the realm of research, numerous BCSSs aimed at combating procrastination have demonstrated efficacy despite not relying on hedonic or eudaimonic design [17]. In this regard, our archetype analysis provides a foundation upon which the effects of different motivational design approaches can be investigated; however, it does not provide insights regarding their effectiveness itself. It is recommended that future research be conducted to build on this open avenue and explore the distinct effects of different motivational design archetypes compared to merely functional BCSSs for procrastination.

Ultimately, in recognizing that most studies in the domain of BCSSs for procrastination that rely on cognitive behavioral treatment strategies use BCSSs to augment existing therapeutic guidance [17], we restricted our cluster analysis to exclusively self-treatment motivational BCSSs, omitting the inclusion of therapeutic guidance. Although our objective was to concentrate on apps that have the potential to be low-threshold alternatives to therapeutic treatment [11,14], it is important to acknowledge that professional assistance may be essential and irreplaceable for individuals who are already experiencing serious mental and physical health consequences of procrastination. In such

XSL•FO
RenderX

cases, it could even be irresponsible to provide them with a stand-alone solution that is not designed as a medical product. Consequently, an avenue for future research in the treatment of procrastination is the analysis of the design opportunities for hybrid apps that provide hedonic and eudaimonic experiences while simultaneously integrating professional therapeutic guidance.

Conclusions

Although procrastination represents a significant self-regulatory failure with the potential to cause adverse mental and physical health consequences, the rate of adherence to professional therapies remains low, particularly in light of the considerable number of individuals affected. Specific forms of BCSSs, namely, motivational BCSSs that elicit positive experiences through hedonic and eudaimonic design elements, have been proposed as a promising self-treatment alternative for individuals struggling with procrastination. However, research on the design and effects of such systems is still in its infancy. Conversely, a multitude of motivational behavior change support apps designed to facilitate behavior change and overcome procrastination have already been developed and implemented in practice. To analyze these practical developments, a systematic review and cluster analysis of prevalent smartphone apps was conducted, resulting in the identification of 5 motivational design archetypes. Upon critical discussion, these archetypes appeared to target different psychological determinants of procrastination and successfully use a variety of hedonic and eudaimonic design elements that extend beyond the current state of research. Nevertheless, further investigation is warranted to ascertain how hedonic and eudaimonic design can be more effectively integrated with cognitive BCTs and how social support mechanisms that have demonstrated efficacy in traditional procrastination treatment can be elicited through motivational BCSSs.

Acknowledgments

Funding for JKK, KS, and CS was provided by the Bavarian Equal Opportunities Sponsorship – Realization of Equal Opportunities for Women in Research and Teaching and the Friedrich-Alexander-Universität Erlangen-Nürnberg Emerging Talents Initiative.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App list.
[XLSX File (Microsoft Excel File), 86 KB - jmir_v27i1e65214_app1.xlsx]

Multimedia Appendix 2
Cluster analysis results.
[XLSX File (Microsoft Excel File), 119 KB - jmir_v27i1e65214_app2.xlsx]

References

1. Klingsieck KB. Procrastination: when good things don't come to those who wait. Eur Psychol 2013 Jan;18(1):24-34. [doi: 10.1027/1016-9040/a000138]
2. Steel P. The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. Psychol Bull 2007 Jan;133(1):65-94. [doi: 10.1037/0033-2909.133.1.65] [Medline: 17201571]
3. Küchler AM, Albus P, Ebert DD, Baumeister H. Effectiveness of an internet-based intervention for procrastination in college students (StudiCare Procrastination): study protocol of a randomized controlled trial. Internet Interv 2019 Sep;17:100245 [FREE Full text] [doi: 10.1016/j.invent.2019.100245] [Medline: 31080750]
4. Urban T. Inside the mind of a master procrastinator. YouTube. URL: https://www.youtube.com/watch?v=arj7oStGLkU [accessed 2024-07-25]
5. Beutel ME, Klein EM, Aufenanger S, Brähler W, Dreier M, Müller KW, et al. Procrastination, distress and life satisfaction across the age range - a German representative community study. PLoS One 2016;11(2):e0148054 [FREE Full text] [doi: 10.1371/journal.pone.0148054] [Medline: 26871572]
6. Johansson F, Rozental A, Edlund K, Côté P, Sundberg T, Onell C, et al. Associations between procrastination and subsequent health outcomes among university students in Sweden. JAMA Netw Open 2023 Jan 03;6(1):e2249346 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.49346] [Medline: 36598789]
7. van Eerde W. A meta-analytically derived nomological network of procrastination. Pers Individ Differ 2003 Oct;35(6):1401-1418. [doi: 10.1016/s0191-8869(02)00358-6]
8. Sirois FM. Is procrastination a vulnerability factor for hypertension and cardiovascular disease? Testing an extension of the procrastination-health model. J Behav Med 2015 Jun 25;38(3):578-589. [doi: 10.1007/s10865-015-9629-2] [Medline: 25804373]

XSL•FO
RenderX

9. Maria-Ioanna A, Patra V. The role of psychological distress as a potential route through which procrastination may confer risk for reduced life satisfaction. Curr Psychol 2020 May 27;41(5):2860-2867. [doi: 10.1007/s12144-020-00739-8]
10. Rozental A, Carlbring P. Understanding and treating procrastination: a review of a common self-regulatory failure. Psychology 2014;5(13):1488-1502. [doi: 10.4236/psych.2014.513160]
11. Yan B, Zhang X. What research has been conducted on procrastination? Evidence from a systematical bibliometric analysis. Front Psychol 2022 Feb 2;13:809044 [FREE Full text] [doi: 10.3389/fpsyg.2022.809044] [Medline: 35185729]
12. Steel P, König CJ. Integrating theories of motivation. Acad Manag Rev 2006 Oct;31(4):889-913. [doi: 10.5465/amr.2006.22527462]
13. Salguero-Pazos M, Reyes-de-Cózar S. Interventions to reduce academic procrastination: a systematic review. Int J Educ Res 2023;121:102228. [doi: 10.1016/j.ijer.2023.102228]
14. van Eerde W, Klingsieck KB. Overcoming procrastination? A meta-analysis of intervention studies. Educ Res Rev 2018 Nov;25:73-85. [doi: 10.1016/j.edurev.2018.09.002]
15. Steinert C, Heim N, Leichsenring F. Procrastination, perfectionism, and other work-related mental problems: prevalence, types, assessment, and treatment-a scoping review. Front Psychiatry 2021 Oct 11;12:736776 [FREE Full text] [doi: 10.3389/fpsyt.2021.736776] [Medline: 34707522]
16. Sirois FM, Melia-Gordon ML, Pychyl TA. "I'll look after my health, later": an investigation of procrastination and health. Pers Individ Differ 2003 Oct;35(5):1167-1184. [doi: 10.1016/s0191-8869(02)00326-4]
17. Kirchner-Krath J, Birnstiel S, Morschheuser B. Combating procrastination with information systems: a systematic review on design approaches and effects. In: Proceedings of the Thirty-Second European Conference on Information Systems. 2024 Presented at: ECIS 2024; June 13-19, 2024; Paphos, Cyprus.
18. Oinas-Kukkonen H. A foundation for the study of behavior change support systems. Pers Ubiquit Comput 2012 Jul 19;17(6):1223-1235. [doi: 10.1007/s00779-012-0591-5]
19. Grady A, Pearson N, Lamont H, Leigh L, Wolfenden L, Barnes C, et al. The effectiveness of strategies to improve user engagement with digital health interventions targeting nutrition, physical activity, and overweight and obesity: systematic review and meta-analysis. J Med Internet Res 2023 Dec 19;25:e47987 [FREE Full text] [doi: 10.2196/47987] [Medline: 38113062]
20. Markkanen JO, Oikarinen N, Savolainen MJ, Merikallio H, Nyman V, Salminen V, et al. Mobile health behaviour change support system as independent treatment tool for obesity: a randomized controlled trial. Int J Obes (Lond) 2024 Mar 07;48(3):376-383 [FREE Full text] [doi: 10.1038/s41366-023-01426-x] [Medline: 38062218]
21. Teeriniemi AM, Salonurmi T, Jokelainen T, Vähänikkilä H, Alahäivälä T, Karppinen P, et al. A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. J Intern Med 2018 Nov 04;284(5):534-545 [FREE Full text] [doi: 10.1111/joim.12802] [Medline: 29974563]
22. Walsh JC, Richmond J, Mc Sharry J, Groarke A, Glynn L, Kelly MG, et al. Examining the impact of an mHealth behavior change intervention with a brief in-person component for cancer survivors with overweight or obesity: randomized controlled trial. JMIR Mhealth Uhealth 2021 Jul 05;9(7):e24915. [doi: 10.2196/24915] [Medline: 36260394]
23. Schwarz A, Winkens LH, de Vet E, Ossendrijver D, Bouwsema K, Simons M. Design features associated with engagement in mobile health physical activity interventions among youth: systematic review of qualitative and quantitative studies. JMIR Mhealth Uhealth 2023 Mar 06;11:e40898 [FREE Full text] [doi: 10.2196/40898] [Medline: 36877551]
24. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
25. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
26. Paramastri R, Pratama SA, Ho DK, Purnamasari SD, Mohammed AZ, Galvin CJ, et al. Use of mobile applications to improve nutrition behaviour: a systematic review. Comput Methods Programs Biomed 2020 Aug;192:105459. [doi: 10.1016/j.cmpb.2020.105459] [Medline: 32234632]
27. Hu EA, Nguyen V, Langheier J, Shurney D. Weight reduction through a digital nutrition and food purchasing platform among users with obesity: longitudinal study. J Med Internet Res 2020 Sep 02;22(9):e19634 [FREE Full text] [doi: 10.2196/19634] [Medline: 32792332]
28. Martinez-Gutierrez J, Domínguez A, López C, Alcántara J, Althausen C, Rojas M, et al. "Appagalo" a customized mobile health intervention (mHealth) for smoking cessation in women: a randomized controlled trial. Tob Use Insights 2023 Feb 21;16. [doi: 10.1177/1179173x231152316]
29. Rajani NB, Mastellos N, Filippidis FT. Impact of gamification on the self-efficacy and motivation to quit of smokers: observational study of two gamified smoking cessation mobile apps. JMIR Serious Games 2021 Apr 27;9(2):e27290 [FREE Full text] [doi: 10.2196/27290] [Medline: 33904824]

XSL•FO
RenderX

30. Fang YE, Zhang Z, Wang R, Yang B, Chen C, Nisa C, et al. Effectiveness of eHealth smoking cessation interventions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 28;25:e45111 [FREE Full text] [doi: 10.2196/45111] [Medline: 37505802]
31. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
32. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, Jokelainen T, Teeriniemi AM, Salonurmi T, et al. Opportunities and challenges of behavior change support systems for enhancing habit formation: a qualitative study. J Biomed Inform 2018 Aug;84:82-92 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.012] [Medline: 29936091]
33. Foulonneau A, Calvary G, Villain E. Stop procrastinating: TILT, time is life time, a persuasive application. In: Proceedings of the 28th Australian Conference on Computer-Human Interaction. 2016 Presented at: OzCHI '16; November 29-December 2, 2016; Launceston, Australia. [doi: 10.1145/3010915.3010947]
34. Inie N, Lungu M. Aiki - turning online procrastination into microlearning. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445202]
35. Arakawa R, Yakura H, Goto M. CatAlyst: domain-extensible intervention for preventing task procrastination using large generative models. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany. [doi: 10.1145/3544548.3581133]
36. Luo J, Wang M, Chen B, Sun M. Exposure to nature sounds through a mobile application in daily life: effects on learning performance among university students. Int J Environ Res Public Health 2022 Nov 07;19(21):14583 [FREE Full text] [doi: 10.3390/ijerph192114583] [Medline: 36361462]
37. Kadosh M, Hen M, Ferrari JR. Reducing statistics anxiety and academic procrastination among Israeli students: a pilot program. Teach Stat 2023 Aug 10;45(3):167-175. [doi: 10.1111/test.12356]
38. Bellhäuser H, Dignath C, Theobald M. Daily automated feedback enhances self-regulated learning: a longitudinal randomized field experiment. Front Psychol 2023;14:1125873 [FREE Full text] [doi: 10.3389/fpsyg.2023.1125873] [Medline: 37275690]
39. Rozental A, Forsström D, Tangen JA, Carlbring P. Experiences of undergoing internet-based cognitive behavior therapy for procrastination: a qualitative study. Internet Interv 2015 Sep;2(3):314-322. [doi: 10.1016/j.invent.2015.05.001]
40. Szalma JL. On the application of motivation theory to human factors/ergonomics: motivational design principles for human-technology interaction. Hum Factors 2014 Dec;56(8):1453-1471. [doi: 10.1177/0018720814553471] [Medline: 25509824]
41. Wu J, Lu X. Effects of extrinsic and intrinsic motivators on using utilitarian, hedonic, and dual-purposed information systems: a meta-analysis. J Assoc Inf Syst 2013 Mar;14(3):153-191. [doi: 10.17705/1jais.00325]
42. Berridge KC, Kringelbach ML. Building a neuroscience of pleasure and well-being. Psychol Well Being 2011 Oct 24;1(1):1-3 [FREE Full text] [doi: 10.1186/2211-1522-1-3] [Medline: 22328976]
43. Lukas CA, Berking M. Reducing procrastination using a smartphone-based treatment program: a randomized controlled pilot study. Internet Interv 2018 Jun;12:83-90 [FREE Full text] [doi: 10.1016/j.invent.2017.07.002] [Medline: 30135772]
44. Kirchner-Krath J, Schmidt-Kraepelin M, Schöbel S, Ullrich M, Sunyaev A, Von Korflesch HF. Outplay your weaker self: a mixed-methods study on gamification to overcome procrastination in academia. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI. [doi: 10.1145/3613904.3642048]
45. HabitRPG Inc. Habitica: gamified taskmanager. App Store. URL: https://apps.apple.com/de/app/habitica-gamified-taskmanager/id994882113 [accessed 2024-07-25]
46. Dreamfora. Dreamfora: AI goal setting. App Store. URL: https://apps.apple.com/de/app/dreamfora-ai-goal-setting/id1561306096 [accessed 2024-07-25]
47. SEEKRTECH CO. LTD. Forest - Bleib fokussiert. App Store. URL: https://apps.apple.com/de/app/forest-bleib-fokussiert/id866450515 [accessed 2024-07-25]
48. Spohrer K, Fallon M, Hoehle H, Heinzl A. Designing effective mobile health apps: does combining behavior change techniques really create synergies? J Manag Inf Syst 2021 Aug 06;38(2):517-545. [doi: 10.1080/07421222.2021.1912936]
49. Kim DH, Hinken B, Lannon C. Applying Systems Archetypes. Arcadia, CA: Pegasus Communications; 1997.
50. Schmidt-Kraepelin M, Toussaint PA, Thiebes S, Hamari J, Sunyaev A. Archetypes of gamification: analysis of mHealth apps. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e19280 [FREE Full text] [doi: 10.2196/19280] [Medline: 33074155]
51. Thiebes S, Toussaint PA, Ju J, Ahn JH, Lyytinen K, Sunyaev A. Valuable genomes: taxonomy and archetypes of business models in direct-to-consumer genetic testing. J Med Internet Res 2020 Jan 21;22(1):e14890 [FREE Full text] [doi: 10.2196/14890] [Medline: 31961329]
52. Denecke K, May R. Developing a technical-oriented taxonomy to define archetypes of conversational agents in health care: literature review and cluster analysis. J Med Internet Res 2023 Jan 30;25:e41583 [FREE Full text] [doi: 10.2196/41583] [Medline: 36716093]

XSL•FO
RenderX

53. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
54. Lowry PB, Gaskin JE, Twyman NW, Hammer B, Roberts TL. Taking “fun and games” seriously: proposing the hedonic-motivation system adoption model (HMSAM). J Assoc Inf Syst 2013 Nov;14(11):617-671. [doi: 10.17705/1jais.00347]
55. Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. J Mark Res 1983 May 01;20(2):134-148. [doi: 10.1177/002224378302000204]
56. Özberk EH, Türk Kurtça T. Profiles of academic procrastination in higher education: a cross-cultural study using latent profile analysis. Int J Psychol Educ Stud 2022 Oct 23;8(3):150-160. [doi: 10.52380/ijpes.2021.8.3.465]
57. Lay CH. A modal profile analysis of procrastinators: a search for types. Pers Individ Differ 1987 Jan;8(5):705-714. [doi: 10.1016/0191-8869(87)90069-9]
58. Becker S, Kribben A, Meister S, Diamantidis CJ, Unger N, Mitchell A. User profiles of a smartphone application to support drug adherence--experiences from the iNephro project. PLoS One 2013 Oct 23;8(10):e78547 [FREE Full text] [doi: 10.1371/journal.pone.0078547] [Medline: 24194946]
59. App data report 2024: in-depth insights of the Apple App Store and Google Play Store. Business of Apps. URL: https://www.businessofapps.com/data/report-app-data/ [accessed 2025-01-29]
60. Wylie L. US app market statistics (2025). Business of Apps. 2025 Jan 22. URL: https://www.businessofapps.com/data/us-app-market/ [accessed 2025-01-29]
61. Wylie L. Germany app market statistics (2025). Business of Apps. 2025 Jan 22. URL: https://www.businessofapps.com/data/germany-app-market/ [accessed 2025-01-29]
62. Neuendorf KA. The Content Analysis Guidebook. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
63. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. BCT Taxonomy (v1): 93 hierarchically-clustered techniques. Ann Behav Med 2013 [FREE Full text]
64. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
65. Eaton CK, McWilliams E, Yablon D, Kesim I, Ge R, Mirus K, et al. Cross-cutting mHealth behavior change techniques to support treatment adherence and self-management of complex medical conditions: systematic review. JMIR Mhealth Uhealth 2024 May 01;12:e49024 [FREE Full text] [doi: 10.2196/49024] [Medline: 38717433]
66. Merolli M, Francis JJ, Vallance P, Bennell KL, Malliaras P, Hinman RS. Evaluation of patient-facing mobile apps to support physiotherapy care: systematic review. JMIR Mhealth Uhealth 2024 Mar 04;12:e55003 [FREE Full text] [doi: 10.2196/55003] [Medline: 38437018]
67. Côté J, Chicoine G, Vinette B, Auger P, Rouleau G, Fontaine G, et al. Digital interventions for recreational cannabis use among young adults: systematic review, meta-analysis, and behavior change technique analysis of randomized controlled studies. J Med Internet Res 2024 Apr 17;26:e55031 [FREE Full text] [doi: 10.2196/55031] [Medline: 38630515]
68. Vlahu-Gjorgievska E, Burazor A, Win KT, Trajkovik V. mHealth apps targeting obesity and overweight in young people: app review and analysis. JMIR Mhealth Uhealth 2023 Jan 19;11:e37716 [FREE Full text] [doi: 10.2196/37716] [Medline: 36656624]
69. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24. [doi: 10.17705/1CAIS.02428]
70. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
71. Premanandan S, Ahmad A, Cajander Å, Ågerfalk P, Dolezel M, van Gemert-Pijnen L. Designing a mobile e-coaching app for immigrant informal caregivers: qualitative study using the persuasive system design model. JMIR Mhealth Uhealth 2023 Nov 09;11:e50038 [FREE Full text] [doi: 10.2196/50038] [Medline: 37943598]
72. Pramod D, Vijayakumar Bharathi S, Patil K. Generative artificial intelligence for personalized elderly care through the lens of the hedonic motivation system adoption model. In: Proceedings of the 2nd International Conference on Advancement in Computation & Computer Technologies. 2024 Presented at: InCACCT 2024; May 2-3, 2024; Gharuan, India. [doi: 10.1109/incacct61598.2024.10551017]
73. Endrayadi A, Yuniarti R, Renaldi F. Habit tracker gamification software development in community-based daily personal hygiene using Octalysis framework. In: Proceedings of the 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management. 2022 Presented at: IEOM 2022; September 13-15, 2022; Johor Bahru, Malaysia. [doi: 10.46254/ap03.20220089]
74. Akdag MT, Jacquemin PH, Wahl N. Visit your therapist in metaverse - designing a virtual environment for mental health counselling. In: Proceedings of the Forty-Fourth International Conference on Information Systems. 2023 Presented at: ICIS 2023; December 10-13, 2023; Hyderabad, India.

XSL•FO
RenderX

75. Setiawan SS, Suryadibrata A. Fitrust: promoting healthy lifestyle through gamified mobile health application. In: Proceedings of the 5th International Conference on New Media Studies. 2019 Presented at: CONMEDIA 2019; October 9-11, 2019; Bali, Indonesia. [doi: 10.1109/conmedia46929.2019.8981840]
76. meaning. Cambridge Dictionary.: Cambridge University Press URL: https://dictionary.cambridge.org/dictionary/english/meaning [accessed 2024-08-01]
77. Remane G, Nickerson RC, Hanelt A, Tesch J, Kolbe L. A taxonomy of carsharing business models. In: Proceedings of 37th International Conference on Information Systems. 2016 Presented at: ICIS 2016; December 11-14, 2016; Dublin, Ireland.
78. Yim O, Ack Baraly KT. Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. Quant Methods Psychol 2015 Feb;11(1):8-21. [doi: 10.20982/tqmp.11.1.p008]
79. Returnull. Noverdue - todo timer list. Apple Inc. URL: https://apps.apple.com/tt/app/noverdue-todo-timer-list/id1492210816 [accessed 2024-12-04]
80. Shikudo. Focus quest: study buddy. App Store. URL: https://apps.apple.com/bb/app/focus-quest-study-buddy/id1516624331 [accessed 2024-12-04]
81. Greatness homepage. IIElevenLabsGrants. URL: https://greatnessapp.com/ [accessed 2024-12-16]
82. Fancy Ventures UG. Focus Dog homepage. Focus Dog. URL: https://focusdog.app/ [accessed 2024-12-04]
83. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
84. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020 Apr;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
85. Xi N, Hamari J. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. Int J Inf Manag 2019 Jun;46:210-221. [doi: 10.1016/j.ijinfomgt.2018.12.002]
86. Codina N, Valenzuela R, Pestana JV, Gonzalez-Conde J. Relations between student procrastination and teaching styles: autonomy-supportive and controlling. Front Psychol 2018 May 23;9:809 [FREE Full text] [doi: 10.3389/fpsyg.2018.00809] [Medline: 29875731]
87. Oram R, Dou A, Rogers M. Pilot study of self-determination theory and motivational interviewing intervention targeting need satisfaction, motivation, and procrastination. Scholarsh Teach Learn Psychol 2022 Jul 14 [FREE Full text] [doi: 10.1037/stl0000331]
88. Skinner BF. Science And Human Behavior. New York, NY: Free Press; 1953.
89. Ferster CB, Skinner BF. Schedules of Reinforcement. Norwalk, CT: Appleton-Century-Crofts; 1957.
90. Van Eerde W. Procrastination: self‐regulation in initiating aversive goals. Appl Psychol 2001 Dec 25;49(3):372-389. [doi: 10.1111/1464-0597.00021]
91. Arkin RM, Oleson KC. Self-handicapping. In: Darley JM, Cooper J, editors. Attribution and Social Interaction: The Legacy of Edward E. Jones. Washington, DC: American Psychological Association; 1998.
92. Steel P, Klingsieck KB. Academic procrastination: psychological antecedents revisited. Aust Psychol 2020 Nov 12;51(1):36-46. [doi: 10.1111/ap.12173]
93. Rozental A, Forsell E, Svensson A, Andersson G, Carlbring P. Internet-based cognitive-behavior therapy for procrastination: a randomized controlled trial. J Consult Clin Psychol 2015 Aug;83(4):808-824. [doi: 10.1037/ccp0000023] [Medline: 25939016]
94. Åsberg K, Bendtsen M. Evaluating the effectiveness of a brief digital procrastination intervention targeting university students in Sweden: study protocol for the Focus randomised controlled trial. BMJ Open 2023 Jul 21;13(7):e072506 [FREE Full text] [doi: 10.1136/bmjopen-2023-072506] [Medline: 37479522]
95. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
96. Lin X, Martinengo L, Jabir AI, Ho AH, Car J, Atun R, et al. Scope, characteristics, behavior change techniques, and quality of conversational agents for mental health and well-being: systematic assessment of apps. J Med Internet Res 2023 Jul 18;25:e45984 [FREE Full text] [doi: 10.2196/45984] [Medline: 37463036]
97. Jabir AI, Martinengo L, Lin X, Torous J, Subramaniam M, Tudor Car L. Evaluating conversational agents for mental health: scoping review of outcomes and outcome measurement instruments. J Med Internet Res 2023 Apr 19;25:e44548 [FREE Full text] [doi: 10.2196/44548] [Medline: 37074762]
98. Bhattacharjee A, Zeng Y, Xu SY, Kulzhabayeva D, Ma M, Kornfield R, et al. Understanding the role of large language models in personalizing and scaffolding strategies to combat academic procrastination. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI. [doi: 10.1145/3613904.3642081]
99. Eckert M, Ebert DD, Lehr D, Sieland B, Berking M. Overcome procrastination: enhancing emotion regulation skills reduce procrastination. Learn Individ Differ 2016 Dec;52:10-18. [doi: 10.1016/j.lindif.2016.10.001]
100. Pychyl TA, Sirois FM. Chapter 8 - Procrastination, emotion regulation, and well-being. In: Procrastination, Health, and Well-Being. Cambridge, MA: Academic Press; 2016:163-188.

XSL•FO
RenderX

101. Greenberg L. The therapeutic relationship in emotion-focused therapy. Psychotherapy (Chic) 2014 Sep;51(3):350-357. [doi: 10.1037/a0037336] [Medline: 25068192]
102. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. Acad Manag Ann 2020 Jul;14(2):627-660. [doi: 10.5465/annals.2018.0057]
103. Burton JW, Stein MK, Jensen TB. A systematic review of algorithm aversion in augmented decision making. Behav Decis Mak 2019 Oct 23;33(2):220-239. [doi: 10.1002/bdm.2155]
104. Rozental A, Forsström D, Lindner P, Nilsson S, Mårtensson L, Rizzo A, et al. Treating procrastination using cognitive behavior therapy: a pragmatic randomized controlled trial comparing treatment delivered via the internet or in groups. Behav Ther 2018 Mar;49(2):180-197. [doi: 10.1016/j.beth.2017.08.002] [Medline: 29530258]
105. Morschheuser B, Hamari J, Maedche A. Cooperation or competition – when do people contribute more? A field experiment on gamification of crowdsourcing. Int J Hum Comput Stud 2019 Jul;127:7-24. [doi: 10.1016/j.ijhcs.2018.10.001]
106. Koppenborg M, Klingsieck KB. Group work and student procrastination. Learn Individ Differ 2022 Feb;94:102117. [doi: 10.1016/j.lindif.2022.102117]
107. Uzun Ozer B, Demir A, Ferrari JR. Reducing academic procrastination through a group treatment program: a pilot study. J Ration Emot Cogn Behav Ther 2013 May 25;31(3):127-135. [doi: 10.1007/s10942-013-0165-0]
108. Amit AJ, Gautam Shankararam SR, Pradeep P, Perumalraja R, Kamalesh S. Framework for preventing procrastination and increasing productivity. In: Proceedings of the 3rd International Conference on Signal Processing and Communication. 2021 Presented at: ICPSC 2021; May 13-14, 2021; Coimbatore, India. [doi: 10.1109/icspc51351.2021.9451773]
109. Bartuskova A, Krejcar O. Handling procrastination in mobile learning environment proposal of reminder application for mobile devices. In: Proceedings of the 6th International Conference on Computer Supported Education. 2014 Presented at: CSEDU 2014; April 1-3, 2014; Barcelona, Spain. [doi: 10.5220/0004960702200225]
110. Auvinen T, Hakulinen L, Malmi L. Increasing students' awareness of their behavior in online learning environments with visualizations and achievement badges. IEEE Trans Learn Technol 2015 Jul 1;8(3):261-273. [doi: 10.1109/tlt.2015.2441718]
111. Bezabih AM, Gerling K, Abebe W, Abeele VV. Behavioral theories and motivational features underlying eHealth interventions for adolescent antiretroviral adherence: systematic review. JMIR Mhealth Uhealth 2021 Dec 10;9(12):e25129 [FREE Full text] [doi: 10.2196/25129] [Medline: 34890353]
112. Schöbel S, Janson A, Jahn K, Kordyaka B, Turetken O, Djafarova N, et al. A research agenda for the why, what, and how of gamification designs: outcomes of an ECIS 2019 panel. Commun Assoc Inf Syst 2019;46(1):706-721. [doi: 10.17705/1cais.04630]
113. Schmidt-Kraepelin M, Thiebes S, Warsinsky SL, Petter S, Sunyaev A. Narrative transportation in gamified information systems: the role of narrative-task congruence. In: Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI EA '23; April 23-28, 2023; Hamburg, Germany. [doi: 10.1145/3544549.3585595]
114. Schmidt-Kraepelin M, Warsinsky S, Thiebes S, Sunyaev A. Conceptualizing narratives in gamified information systems. In: Proceedings of the 55th Hawaii International Conference on System Sciences. 2022 Presented at: HICSS 2022; January 4-7, 2022; Virtual Event. [doi: 10.24251/hicss.2022.197]
115. Bahng S, Kelly RM, McCormack J. Reflexive VR storytelling design beyond immersion: facilitating self-reflection on death and loneliness. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376582]
116. Clarke S, Arnab S, Morini L, Heywood L. Remixing dungeons and dragons: a playful approach to student self-reflection. In: Proceedings of the 12th European Conference on Game-Based Learning. 2018 Presented at: ECGBL 2018; October 4-5, 2018; Sophia Antipolis, France.
117. Wolfe A, Louchart S, Loranger B. The impacts of design elements in interactive storytelling in VR on emotion, mood, and self-reflection. In: Proceedings of the 15th International Conference on Interactive Digital Storytelling. 2022 Presented at: ICIDS 2022; December 4-7, 2022; Santa Cruz, CA. [doi: 10.1007/978-3-031-22298-6_40]
118. Mitsea E, Drigas A, Skianis C. Virtual reality mindfulness for meta-competence training among people with different mental disorders: a systematic review. Psychiatry Int 2023 Oct 18;4(4):324-353. [doi: 10.3390/psychiatryint4040031]
119. Mazzolenis MV, Mourra GN, Moreau S, Mazzolenis ME, Cerda IH, Vega J, et al. The role of virtual reality and artificial intelligence in cognitive pain therapy: a narrative review. Curr Pain Headache Rep 2024 Sep 08;28(9):881-892. [doi: 10.1007/s11916-024-01270-2] [Medline: 38850490]
120. Wazid SW, Shahnawaz MG, Gupta D. Procrastination among students: the role of gender, perfectionism and self-esteem. Indian J Soc Work 2016;77(2):191-210.
121. Tian Q, Mustapha SM, Min J. The mediation effect of academic self-efficacy on academic procrastination, performance, and satisfaction of Chinese local technology university undergraduates. Psychol Res Behav Manag 2024 Nov;17:3779-3798. [doi: 10.2147/prbm.s479189]
122. Irk E, Mumcu A. The moderator role of self-control in the effect of future time perspective on procrastination behavior. Organizatsionnaya Psikhologiya 2024 Oct;14(3):32-45. [doi: 10.17323/2312-5942-2024-14-3-32-45]
123. Netzer Turgeman R, Pollak Y. Using the temporal motivation theory to explain the relation between ADHD and procrastination. Aust Psychol 2023 Jun 07;58(6):448-456. [doi: 10.1080/00050067.2023.2218540]

XSL•FO
RenderX

124. Krath J, von Korflesch HF. Player types and game element preferences: investigating the relationship with the gamification user types HEXAD scale. In: Proceedings of the Third International Conference on HCI in Games: Experience Design and Game Mechanics. 2021 Presented at: HCI-Games 2021; July 24-29, 2021; Virtual Event. [doi: 10.1007/978-3-030-77277-2_18]
125. Kirchner-Krath J, Altmeyer M, Schürmann L, Kordyaka B, Morschheuser B, Klock AC, et al. Uncovering the theoretical basis of user types: an empirical analysis and critical discussion of user typologies in research on tailored gameful design. Int J Hum Comput Stud 2024 Oct;190:103314. [doi: 10.1016/j.ijhcs.2024.103314]
126. Schoebel S, Schmidt-Kraepelin M, Janson A, Sunyaev A. Adaptive and personalized gamification designs: call for action and future research. AIS Transact Hum Comput Interact 2021 Dec 31;13(4):479-494. [doi: 10.17705/1thci.00158]
127. Ghanouni P, Jarus T, Zwicker JG, Lucyshyn J, Mow K, Ledingham A. Social stories for children with autism spectrum disorder: validating the content of a virtual reality program. J Autism Dev Disord 2019 Mar 10;49(2):660-668. [doi: 10.1007/s10803-018-3737-0] [Medline: 30203311]
128. Focusmate homepage. Focusmate. URL: https://www.focusmate.com/ [accessed 2024-08-06]

Abbreviations

AI: artificial intelligence
BCSS: behavior change support system
BCT: behavior change technique
BCTT: Behavior Change Technique Taxonomy
IS: information system
PSD: persuasive system design

Edited by A Mavragani; submitted 08.08.24; peer-reviewed by K Denecke, A Chung; comments to author 02.12.24; revised version received 17.12.24; accepted 26.12.24; published 20.02.25.

Please cite as:
Kirchner-Krath J, Schmidt-Kraepelin M, Schmähl K, Schütz C, Morschheuser B, Sunyaev A
Behavior Change Support Systems for Self-Treating Procrastination: Systematic Search in App Stores and Analysis of Motivational Design Archetypes
J Med Internet Res 2025;27:e65214
URL: https://www.jmir.org/2025/1/e65214
doi:10.2196/65214
PMID:39977017

©Jeanine Kirchner-Krath, Manuel Schmidt-Kraepelin, Katharina Schmähl, Christoph Schütz, Benedikt Morschheuser, Ali Sunyaev. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Maternal Mental Health and Weight Control With a Mindfulness Blended Care Approach: Insights From a Randomized Controlled Trial

Kathrin Hassdenteufel[1], Dr med; Mitho Müller[2], Dr; Harald Abele[3], Prof Dr Med; Sara Yvonne Brucker[3], Prof Dr Med; Johanna Graf[4], Dr; Stephan Zipfel[4,5], Prof Dr Med; Armin Bauer[6], Dr; Peter Jakubowski[3], Dr med; Jan Pauluschke-Fröhlich[3], Dr med; Markus Wallwiener[7], Prof Dr Med; Stephanie Wallwiener[8], Prof Dr Med

[1]Department of Obstetrics and Gynecology, University Hospital Heidelberg, Heidelberg, Germany

[2]Department of Psychology, Ludwig-Maximilians-University, Munich, Germany

[3]Department of Women's Health, University Hospital Tübingen, Tübingen, Germany

[4]Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tübingenn, Tübingen, Germany

[5]German Centre for Mental Health (DZPG-Tuebingen), Tübingen, Germany

[6]Department of Women's Health, Research Institute for Women's Health, University Hospital Tübingen, Tübingen, Germany

[7]Department of Gynecology, Martin Luther University of Halle-Wittenberg, Halle, Germany

[8]Department of Obstetrics and Prenatal Medicine, Martin Luther University of Halle-Wittenberg, Halle, Germany

Corresponding Author:
Kathrin Hassdenteufel, Dr med
Department of Obstetrics and Gynecology
University Hospital Heidelberg
Im Neuenheimer Feld 440
69120 Heidelberg
Heidelberg, 69120
Germany
Phone: 49 15238492062
Email: Kathrin.hassdenteufel@med.uni-heidelberg.de

Abstract

Background: Perinatal maternal mental health problems, such as depression and anxiety, are highly prevalent during pregnancy and post partum. Electronic mindfulness-based interventions (eMBIs) are a promising treatment option, which can be provided in a low-threshold, cost-effective manner. However, research underscores the fact that face-to-face coaching sessions are more effective than solely digital methods. A blended care approach (eMBI with direct face-to-face coaching) could amplify the therapeutic impact on maternal mental health and weight gain during the perinatal period.

Objective: We investigated whether combining an eMBI intervention with face-to-face personal support significantly improves maternal mental health, and whether the intervention can influence weight gain in affected women during pregnancy.

Methods: A community-based sample of 460 pregnant women with a singleton pregnancy who screened positive for depression was enrolled in a multicenter randomized controlled trial (RCT) including the University Hospitals of Heidelberg and Tübingen as well as more than 200 gynecological practices within the state of Baden-Württemberg in Germany between February 2019 and October 2020. Participating women were randomized 1:1 to the control group (CG) or intervention group (IG) that received access to an 8-week pregnancy-adapted eMBI between the 29th and 36th gestational week. In a subanalysis, we grouped participants in those receiving only the initial face-to-face coaching session at recruitment (no personal coaching) and those with ≥2 personal coaching sessions. Primary outcome measures were severity of depressive symptoms using the Edinburgh Postnatal Depression Scale, anxiety using the State-Trait Anxiety Inventory, the Pregnancy-Related Anxiety Questionnaire, the Freiburg Mindfulness Inventory, and the Patient Health Questionnaire; secondary outcome measure, BMI.

Results: In the final sample, 137 CG women and 102 IG women received only one coaching session, whereas 37 CG women and 40 IG women received at least 2 (mean 2.3, SD 0.7) coaching sessions. The analyses were adjusted for significant confounders. The IG's mindfulness scores increased significantly ($F_{1.873,344.619}$=4.560, P=.01, η^2=0.024, ω^2=0.012) regardless of coaching frequency. Both general anxiety ($F_{12,129}$=2.361, P=.01, η^2=0.0180, ω^2=0.100) and depression symptoms ($F_{4.758,\ 699.423}$=3.033,

XSL•FO
RenderX

P=.01, η^2=0.020, ω^2=0.009) were significantly lower in the group that received ≥2 coaching sessions than in the no-personal-coaching group. In the group receiving ≥2 coaching sessions, BMI generally was lower in the IG than in the CG ($F_{3.555,444.416}$=4.732, P=.002, η^2=0.036, ω^2=0.013).

Conclusions: Adding a minimal amount of PC to the digital eMBI increased mindfulness and decreased birth-related anxiety, symptoms of depression, and anxiety in at-risk pregnant women. Favorable effects on gestational weight gain were found in the respective IGs, the strongest effect being within the PC group. This blended digital health approach amplifies the effectiveness of the digital intervention.

Trial Registration: German Clinical Trials Register DRKS00017210; https://www.drks.de/search/de/trial/DRKS00017210

(J Med Internet Res 2025;27:e56230) doi:10.2196/56230

KEYWORDS

peripartum mental health; digital intervention; depression; anxiety; personal coaching; ehealth; pregnancy; maternal mental health; weight gain; mindfulness-based intervention; coaching; randomized controlled clinical trial; postpartum; treatment; electronic; effectiveness; women; digital intervention

Introduction

During the perinatal period, pregnant women face an elevated risk of developing anxiety and depressive disorders. While prenatal depression affects 10%-25% of pregnant women [1], up to one-third experience depression and anxiety disorders after delivery [2,3].

The etiology of these disorders is multifactorial, ranging from obstetrical, high-risk pregnancies and psychological distress to the manifestation of previous depression or anxiety disorders [4]. Once present, perinatal mental disorders can, in turn, lead to pregnancy complications, sleep disturbances, suicidal tendencies, high-risk behavior, and overall poor maternal health-related quality of life [5-7]. Moreover, maternal anxiety and depression increase the risk of preterm birth and newborn low birth weight and have a negative impact on infant behavioral, motor, and cognitive development [8,9].

Psychological distress, in turn, is associated with greater gestational weight gain (GWG) during pregnancy and a higher risk for impaired insulin resistance, potentially resulting in adverse outcomes for both mother and child [10]. As women are at higher risk for gestational diabetes, cesarean section, large-for-gestational-age fetal growth, and postpartum weight retention, prevention at this stage is crucial [11]. As excessive GWG can be found in approximately 50% of all pregnancies, thus presenting a major global health burden, prenatal lifestyle interventions and prevention measures are warranted [12].

Meanwhile, digital interventions such as mindfulness approaches have proven to be helpful in supporting pregnant women with mental health symptoms and in preventing excessive weight gain [13]. We recently demonstrated the beneficial effects of an electronic mindfulness-based intervention (eMBI) in a randomized controlled clinical trial (RCT) [14] with an 8-week digitally guided intervention. The intervention significantly increased maternal mindfulness and significantly decreased birth-related anxiety. However, while the intervention significantly impacted the prevalence of postpartum depressiveness, symptoms of depression and general anxiety were not significantly reduced during the intervention. Comparable outcomes in diminishing maternal distress have been highlighted in previous research as well [15]. Lönneberg et al [16] conducted an RCT to evaluate the effects of a face-to-face mindfulness program (Mindfulness-Based Childbirth and Parenting Program) on stress, depression, and mood of pregnant women. Integrating perinatal health insights with mindfulness exercises, the program yielded promising results: participants reported fewer depressive symptoms, diminished stress levels, an increased sense of mindfulness, and a positive state of mind after completing the program.

However, this literature is sparse regarding the prevention of glucose intolerance. Research has shown favorable effects of a personal mindfulness intervention in improving perinatal stress and maternal weight gain, though without reaching significance [10].

While digital interventions have proven useful, we questioned whether their effectiveness could be optimized by combining eMBI with a minimum of personal support, such as face-to-face interventions, in a blended care model. Digital health care interventions have generally shown high acceptability rates and represent far-reaching, relatively cost-effective methods to support health care on a low-threshold basis [17]. However, despite the considerable growth of digital interventions in the last decade, previous research has shown that such interventions tend to lack sensitivity and may not adequately meet individual needs [18]. Therefore, flexible solutions are warranted, and ongoing studies are increasingly focusing on blended digital health solutions with human support in order to enhance patient engagement and satisfaction [19]. Favorable results in combining conventional digital tools with a human touch, which has been identified as a key component, were demonstrated in an RCT focusing on the promotion of healthy GWG [20]. In line with this, Chiauzzi and Newell [19] assert that flexible use of apps based on a tailored strategy is more appropriate than reliance on the app alone. Therefore, adding personal support to digital mindfulness programs may increase their positive outcome. As digital-only methods frequently underperform compared to face-to-face approaches in addressing depression, general anxiety, and GWG, adding a personal touch seems imperative.

Personal coaching (PC) through smartphones or video may improve the mental health of pregnant women, fostering deeper interactions, on the one hand, and facilitating health-focused

XSL•FO
RenderX

conversations with health care personnel, on the other. Building on this premise, we aimed to extend previous research by comparing 2 different clinical approaches, that is, eMBI and PC. Expanding upon our previous RCT, we investigated the combined efficacy of personal support and coaching with an electronic mindfulness program in this study, specifically examining its effects on depression and anxiety as well as on maternal weight gain as a secondary outcome in pregnant women.

Methods

Patients

Participating centers were the university hospitals of Heidelberg and Tübingen, along with over 200 gynecological practices in the state of Baden-Wuerttemberg, Germany, as detailed in our previous publication [14]. Briefly, we screened 5299 pregnant women using the Edinburgh Postnatal Depression Scale (EPDS) between February 2019 and October 2020. Those scoring above 9 were enlisted to participate in an 8-week eMBI from their 29th to 36th gestational weeks. Demographic and psychometric data were collected before, during, and up to 6 months after the intervention and compared with the control group (CG).

For this subanalysis, we compared participants who received more than one face-to-face coaching session during the study period (personal coaching group [PC group]), with participants who received only the first mandatory session (no personal coaching [NPC group]). The effect of PC on psychometric parameters was evaluated in the intervention groups (IG) and CGs. The clinical study phase was conducted over a 2-year period between January 1, 2019, and December 31, 2020. The individual study period was 13 months.

Ethical Considerations

The study was approved by the ethics committees of the Medical Faculties of the Universities of Heidelberg (S-744/2018) and Tübingen (952/2018BO2) and was conducted in accordance with the ethical standards of the Declaration of Helsinki. All personal data were collected in accordance with the principles of confidentiality and the European General Data Protection Regulation (EU-GDPR). All participants provided a signed written consent and received €100 (roughly US $102) each as financial compensation. The study participants had the right to request information about the personal data collected from them. Participation in the study was voluntary after all study content and objectives had been declared and written consent had been given and could be withdrawn at any time without any disadvantages for the patient. Patient data and clinical data are recorded pseudonymously and stored centrally in the eCRF. This study follows the CONSORT (Consolidated Standards of Reporting Trials) statement (Multimedia Appendix 1) and the SPIRIT (Standard Protocol Items: Recommendations for Intervention Trials) guidelines [21]. The study was registered with the German Clinical Trials Registry (DRKS00017210).

Interventions

Expectant mothers were randomized 1:1 to the IG, where they were granted access to eMBI (IG=eMBI), or the CG, which received standard care, including an initial psychological session (CG=treatment as usual [TAU]). The eMBI, starting at 29 weeks gestation, consisted of eight weekly 45-minute sessions. These sessions provided psychoeducational and obstetrical content, along with pregnancy-adapted mindfulness exercises. Both groups completed digital questionnaires on a regular basis: every 2 weeks during the intervention (T1-T5) period and at 1 and 5 months post partum (T6 and T7). Before randomization, all women participated in a mandatory psychological assessment. Enhanced personalized care encompassed the telephone availability of supervised and professionally trained providers and access to psychological support when needed.

The eMBI was developed by a multidisciplinary team that included gynecologists, psychologists, and midwives. The questionnaires assessed sociodemographic and medical data, physiological measures, and self-reported data on maternal mental health. In addition, each participating woman had weekly access to a self-developed digital pregnancy guide with educational and validated medical content focused on pregnancy and childbirth. The modules included for example the following topics: fears and worries about birth or parenting, coping with stress, birth-related pain control, and outlook puerperium.

Primary Outcome Measures

The EPDS

The EPDS, originally developed in 1998 by Cox et al [22] and translated into German by Bergant et al [23], is a 10-item self-rating scale that assesses depressive symptoms in the peripartum. The EPDS has proven to be a valid and sensitive instrument for predicting depressive disorders not only in the postnatal period but also prenatally [24]. The most commonly used cut-off score of 9 (EPDS>9) has shown a sensitivity of 0.96 and a specificity of 1.00 [25].

State-Trait Anxiety Inventory

Perinatal anxiety was measured using the German version of the State-Trait Anxiety Inventory (STAI) during the perinatal period. The questionnaire consists of two scales: The STAI-S (state scale) assesses anxiety as a temporary condition, encompassing feelings of tension, nervousness, and worry, whereas the STAI-T (trait scale) refers to dispositional anxiety over time. The sum, ranging from 20 to 80, indicates the level of situational and dispositional anxiety [26].

Pregnancy-Related Anxiety Questionnaire–Revised

The Pregnancy-Related Anxiety Questionnaire–Revised (PRAQ-R) is an abbreviated 10-item self-report measure of pregnancy and childbirth anxiety and is a valid predictor of birth and childhood outcomes [27-29]. Items include fear of childbirth, concerns about bearing a physically or mentally handicapped child, and concerns about one's appearance [30].

Secondary Outcomes

Freiburg Mindfulness Inventory

The Freiburg Mindfulness Inventory (FMI-14, also known as FFA-14) was developed according to the Buddhist rules of life. It defines mindfulness as the tendency to act in a mindful way, measured as a personal characteristic [31,32]. We used the

XSL•FO
RenderX

abbreviated German version consisting of 14 Likert-scale (1-4) items [33].

Patient Health Questionnaire

The Patient Health Questionnaire (PHQ-D), tailored for the practical screening of mental disorders in primary care, directly measures the diagnostic criteria of the *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition]) [34]. It has been shown to be a valid and well-accepted self-rating instrument for use in research and clinical practice, also in its German rendition [35]. The PHQ-D measures the following scales with 78 items: somatoform, depressive, anxiety, eating disorders, and alcohol abuse. It also includes items on psychosocial functioning, stress experiences, and critical life events.

BMI

Participants responded to items about body height and weight. Weight was assessed at all time points. In addition, at T1, weight before pregnancy was assessed retrospectively (T0). Thus, we were able to compute the BMI for 8 time points according to the respective formula:



Statistical Analysis Plan

All analyses were conducted using SPSS Statistics for Windows (version 29.0.0.0-241; IBM Corp). Little's Missing-completely-at-random test was used to ensure that missing data due to dropouts and missing values were valid for our analyses [36] Moreover, the groups were tested for comparability regarding sociodemographic and medical third variables by means of *t* tests, Mann-Whitney *U* tests and chi-square tests. If differences were significant, the respective variables were analyzed for associations with the outcome variables and included as covariates.

The manipulation check (analyses regarding mindfulness as indexed by the FFA-14 scores) as well as the primary and secondary analyses were conducted using (multivariate) ANOVA with repeated measurements corrected for significant confounders. Mauchly's sphericity test was used to determine whether the sphericity assumption had been violated. If significant, repeated measures *df*s were corrected using Huynh–Feldt or Greenhouse-Geisser correction, depending on the degree of violation. Due to the exploratory nature of the current analyses, the critical, local α-errors were not adjusted and set to a conventional level of α_{local}=.05. Partial η^2 and ω^2 are used as effect sizes. These are sample-based and population-based estimators of explained variances, respectively. According to Cohen [37], η^2/ω^2=0.01 represents small, η^2/ω^2=0.06 represents medium, and η^2/ω^2=0.14 represents large effects. Dunn's multiple comparison procedure [38] was performed as a post hoc test for significant effects relevant to the hypotheses. This procedure results in a minimum significant difference (ψ).

The final sample consisted of 460 participants. However, due to dropouts and further missing values, the number of valid cases varies. Post hoc power calculations (calculated with G*Power; version 3.1.9.7) [39,40] for nonsignificant results are reported in the results section. However, in our ANOVAs, mainly, the power was virtually 1–β=1.0 for large (*f*=0.40) within- and between-subject effects as well as for medium-sized within-subject effects (*f*=0.25).

Main and interaction effects are not reported or interpreted in detail if the respective factors are significantly interacting with additional factors. Also, the main effects of confounders are not reported and interpreted in detail.

Results

Descriptive Characteristics

The flow chart is presented in Figure 1. Results of the Missing-completely-at-random test were not significant ($\chi^2_{25,073}$=25,104.7, *P*=.44). All sociodemographic characteristics and tests on comparability are demonstrated in Tables 1 and 2. The IG and the CG statistically significantly differed regarding the PHQ score, the number of children at home, and the educational level (refer to Table 1): the IG scored lower in the PHQ stress evaluation, had more children at home, and had a lower educational level than the CG. There were no significant differences between the NPC and the PC groups (refer to Table 2).

The potential confounders identified were correlated with the outcome variables: The PHQ stress evaluation was significantly associated with every outcome at every measurement point, except the BMI. The associations (*r*) ranged between 0.201 and 0.479 (*P*<.001), showing the highest association with the STAI-T at T6 and the lowest association with the PRAQ-R at T5. The numbers of children at home were significantly associated with the FFA-14, the STAI-S, and STAI-T, as well as the PRAQ-R (ρ ranging between–0.241 and 0.173) for the highest positive association with the STAI-S at T1 and the highest negative association with the PRAQ-R at T1 (levels of significance [*P*] ranging between <.001 and .048). The level of education was significantly associated with the EPDS, STAI-S, and STAI-T, and the BMI (ρ ranging between–0.346 and –0.122) for the highest negative association with the BMI at T4 and the lowest negative association with the EPDS at T1 (levels of significance [*P*] ranging between <.001 and .04). Consequently, we controlled these variables as covariates in the respective analyses. Descriptive statistics of outcome measures are shown in Table 3.

XSL•FO
RenderX

Figure 1. Participant flowchart. EPDS: Edinburgh Postnatal Depression Scale; eMBI: electronic mindfulness-based intervention; TAU: treatment as usual; NPC: no personal coaching; PC: personal coaching.

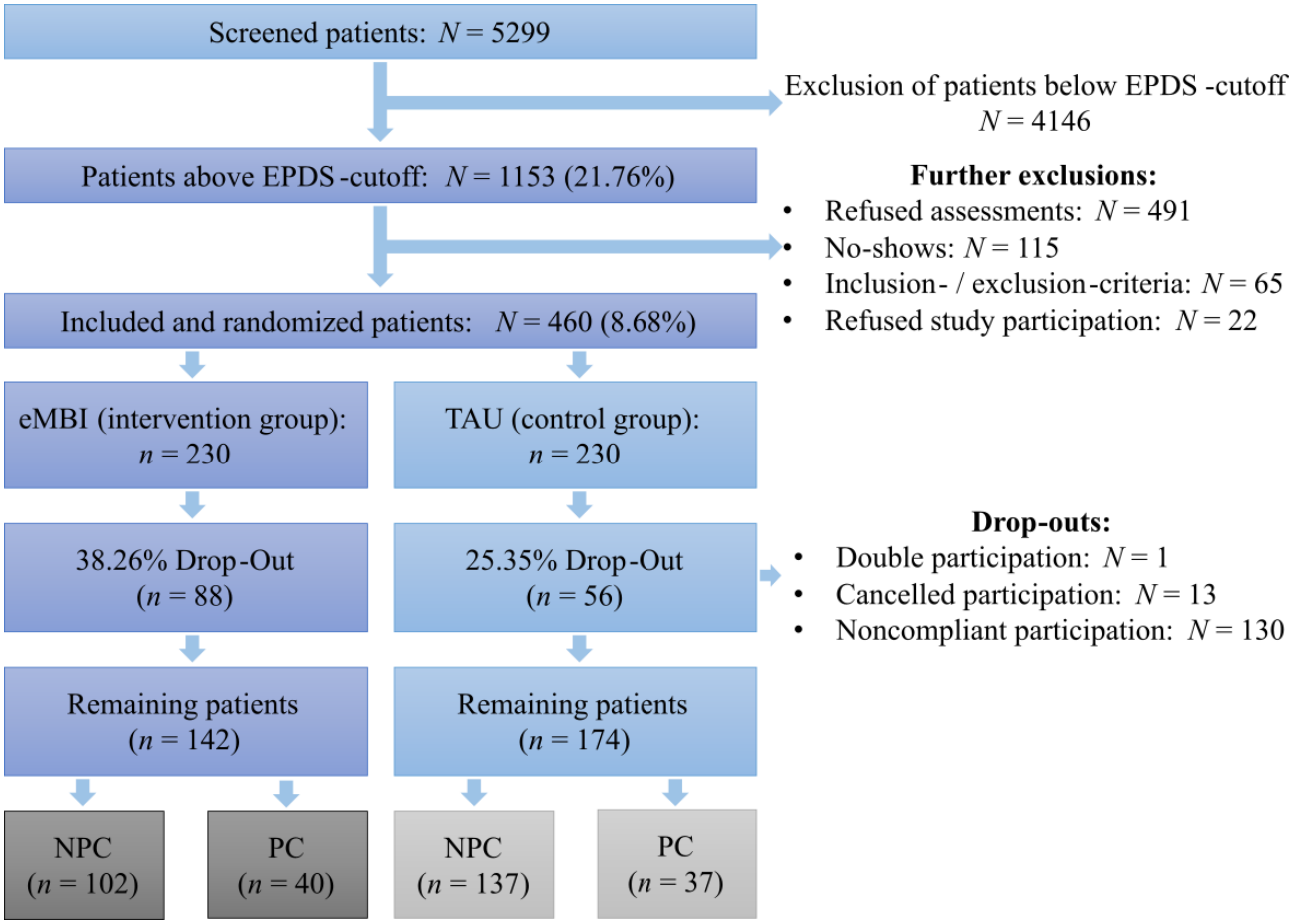

XSL•FO
RenderX

Table 1. Demographics and tests on comparability of the subgroups TAUM[a] versus eMBI[b].

Characteristics	General	TAU	eMBI	Statistics[c]	*P* value
Maternal age (years), mean (SD)	32.6 (4.3)	32.8 (4.6)	32.3 (4.7)	0.95 (314)[d]	.35
Gestation age at study inclusion (weeks), mean (SD)	21.2 (4.3)	21.4 (4.2)	21.0 (4.3)	0.78 (307)[d]	.43
Gestation age at birth (weeks), mean (SD)	39.2 (1.7)	39.2 (1.7)	39.2 (1.6)	0.14 (290)[d]	.89
Infant body weight (grams), mean (SD)	3,370.4 (475.9)	3,411.8 (438.6)	3,319.4 (515.3)	1.70 (306)[d]	.09
Infant body length (cm), mean (SD)	51.5 (2.6)	51.7 (2.5)	51.2 (2.6)	1.52 (300)[d]	.13
PHQ[e] stress evaluation (points), mean (SD)	6.8 (3.4)	7.2 (3.7)	6.3 (3.0)	2.28 (313)[d]	.02
Maternal education frequencies (%)				8396.5[f]	<.01
University entrance qualification	149 (51.4)	97 (58.1)	52 (42.3)		
University of applied sciences entrance qualification	48 (16.6)	29 (17.4)	19 (15.4)		
High secondary qualification	78 (26.9)	32 (19.2)	46 (37.4)		
Low secondary qualification	14 (4.8)	8 (4.8)	6 (4.9)		
No school leaving qualification	1 (0.3)	1 (0.6)	0 (0.0)		
Level of maternal occupation frequencies (%)				9150.0[f]	.84
Prohibition notice	113 (40.9)	68 (42.5)	45 (38.8)		
Unemployed	35 (12.7)	16 (10.0)	19 (16.4)		
Part-time employed	60 (21.7)	33 (20.6)	27 (23.3)		
Full-time employed	68 (24.6)	43 (26.9)	25 (21.6)		
Household net income frequencies (%)				8744.5[f]	.15
< 1.500 €	70 (24.8)	36 (21.8)	34 (29.1)		
1.500-2.999 €	128 (45.4)	76 (46.1)	52 (44.4)		
3.000-4.999 €	62 (22.0)	39 (23.6)	23 (19.7)		
5.000-8.000 €	21 (7.4)	13 (7.9)	8 (6.8)		
> 8.000 €	1 (0.4)	1 (0.6)	0 (0.0)		
Civil status frequencies (%)				5.35 (5)[g]	.34
Married	192 (65.8)	113 (67.3)	79 (63.7)		
Partnership	94 (32.2)	53 (31.5)	41 (33.1)		
Single	5 (1.7)	1 (0.6)	4 (3.2)		
Divorced	1 (0.3)	1 (0.6)	0 (0.0)		
Country of origin frequencies (%)				19.17 (17)[g]	.17
Germany	249 (85.3)	140 (83.3)	109 (87.9)		
Other	43 (14.7)	28 (16.7)	15 (12.1)		
Current psychotherapy frequencies (%)				0.01 (1)[g]	.99
False	55 (17.5)	30 (17.3)	25 (17.7)		
True	259 (82.5)	143 (82.7)	116 (82.3)		
Gravidity frequencies (%)				11,822.0[f]	.48
1st pregnancy	143 (45.3)	82 (47.1)	61 (43.0)		
2nd pregnancy	85 (26.9)	44 (25.3)	41 (28.9)		
3rd pregnancy	47 (14.9)	29 (16.7)	18 (12.7)		
≥4th pregnancy	41 (13.0)	19 (10.9)	22 (15.5)		

XSL•FO
RenderX

Characteristics	General	TAU	eMBI	Statistics[c]	*P* value
Parity frequencies (%)				11,675.5[f]	.35
1st birth	177 (56.0)	102 (58.6)	75 (52.8)		
2nd birth	103 (32.6)	53 (30.5)	50 (35.2)		
3rd birth	30 (9.5)	15 (8.6)	15 (10.6)		
≥4th birth	6 (1.9)	4 (2.3)	2 (1.4)		
Number of children at home frequencies (%)				9075.5[f]	<.05
No child	167 (57.4)	105 (62.5)	62 (50.4)		
One child	96 (33.0)	49 (29.2)	47 (38.2)		
Two children	25 (8.6)	12 (7.1)	13 (10.6)		
Three or more children	3 (1.0)	2 (1.2)	1 (0.8)		
Birth mode frequencies (%)				2.29 (3)[g]	.51
Spontaneous	179 (57.9)	103 (60.6)	76 (54.7)		
Primary c-section	42 (13.6)	23 (13.5)	19 (13.7)		
Secondary c-section	64 (20.7)	30 (17.6)	34 (24.5)		
Vaginal operative	24 (7.8)	14 (8.2)	10 (7.2)		
Infant sex frequencies (%)				0.34 (1)[g]	.64
Female infants	126 (40.6)	72 (42.1)	54 (38.8)		
Male infants	184 (59.4)	99 (57.9)	85 (61.2)		
Infant APGAR values after 10 min. frequencies (%)				11,053.0[f]	.69
10	279 (92.7)	154 (92.2)	125 (93.3)		
9	16 (5.3)	8 (4.8)	8 (6.0)		
8	5 (1.7)	4 (2.4)	1 (0.7)		
7	1 (0.3)	1 (0.6)	0 (0.0)		

[a]TAU: treatment as usual.

[b]eMBI: electronic mindfulness-based intervention.

[c]Empirical α-error.

[d]*t* test (df) values.

[e]PHQ: Patient Health Questionnaire.

[f]Mann-Whitney *U* tests.

[g]Chi-square (df) values.

XSL•FO
RenderX

Table 2. Demographics and tests on comparability of the subgroups NPC[a] versus PC[b].

Characteristics	General	NPC	PC	Statistics[c]	*P* value
Maternal age (years), mean (SD)	32.6 (4.3)	32.4 (4.5)	32.9 (4.9)	0.73 (314)[d]	.47
Gestation age at study inclusion (weeks), mean (SD)	21.2 (4.3)	21.5 (4.3)	20.6 (4.3)	1.59 (307)	.11
Gestation age at birth (weeks), mean (SD)	39.2 (1.7)	39.3 (1.5)	39.0 (2.2)	1.57 (290)	.12
Infant body weight (grams), mean (SD)	3370.4 (475.9)	3363.8 (477.6)	3390.9 (473.4)	0.43 (306)	.67
Infant body length (cm), mean (SD)	51.5 (2.6)	51.3 (2.6)	52.0 (2.4)	1.80 (300)	.07
PHQ[e] stress evaluation (points), mean (SD)	6.8 (3.4)	6.7 (3.4)	7.0 (3.5)	0.56 (313)	.58
Maternal education frequencies (%)				7119.5[f]	.37
University entrance qualification	149 (51.4)	111 (50.2)	38 (55.1)		
University of applied sciences entrance qualification	48 (16.6)	37 (16.7)	11 (15.9)		
High secondary qualification	78 (26.9)	59 (26.7)	19 (27.5)		
Low secondary qualification	14 (4.8)	13 (5.9)	1 (1.4)		
No school leaving qualification	1 (0.3)	1 (0.5)	0 (0.0)		
Household net income frequencies (%)				7096.5[f]	.85
< 1.500 €	70 (24.8)	54 (25.1)	16 (23.9)		
1.500 - 2.999 €	128 (45.4)	94 (43.7)	34 (50.7)		
3.000 - 4.999 €	62 (22.0)	53 (24.7)	9 (13.4)		
5.000 - 8.000 €	21 (7.4)	13 (6.0)	8 (11.9)		
> 8.000 €	1 (0.4)	1 (0.5)	0 (0.0)		
Level of maternal occupation frequencies (%)				7013.0[f]	.91
Prohibition notice	113 (40.9)	86 (41.3)	27 (39.7)		
Unemployed	35 (12.7)	24 (11.5)	11 (16.2)		
Part-time employed	60 (21.7)	46 (22.1)	14 (20.6)		
Full-time employed	68 (24.6)	52 (25.0)	16 (23.5)		
Civil status frequencies (%)				4.23 (5)[g]	.50
Married	192 (65.8)	143 (64.4)	49 (70.0)		
Partnership	94 (32.2)	75 (33.8)	19 (27.1)		
Single	5 (1.7)	3 (1.4)	2 (2.86)		
Divorced	1 (0.3)	1 (0.5)	0 (0.0)		
Country of origin frequencies (%)				17.24 (17)[g]	.32
Germany	249 (85.3)	192 (86.5)	57 (81.4)		
Other	43 (14.7)	30 (13.5)	13 (18.6)		
Current psychotherapy frequencies (%)				0.27 (1)[g]	.60
False	55 (17.5)	40 (16.9)	15 (19.5)		
True	259 (82.5)	197 (83.1)	62 (80.5)		
Gravidity frequencies (%)				8127.0[f]	.10
1st pregnancy	143 (45.3)	114 (47.7)	29 (37.7)		
2nd pregnancy	85 (26.9)	64 (26.8)	21 (27.3)		
3rd pregnancy	47 (14.9)	31 (13.0)	16 (20.8)		
≥ 4th pregnancy	41 (13.0)	30 (12.6)	11 (14.3)		
Parity frequencies (%)				2964.0[f]	.89

XSL•FO
RenderX

Characteristics	General	NPC	PC	Statistics[c]	*P* value
1st birth	177 (56.0)	139 (58.2)	38 (49.4)		
2nd birth	103 (32.6)	74 (31.0)	29 (37.7)		
3rd birth	30 (9.5)	22 (9.2)	8 (10.4)		
≥ 4th birth	6 (1.9)	4 (1.7)	2 (2.6)		
Number of children at home frequencies (%)				7611.0[f]	.82
No child	167 (57.4)	129 (58.4)	38 (54.3)		
One child	96 (33.0)	68 (30.8)	28 (40.0)		
Two children	25 (8.6)	22 (10.0)	3 (4.3)		
Three or more children	3 (1.0)	2 (0.9)	1 (1.4)		
Birth mode frequencies (%)				1.86 (3)[g]	.60
Spontaneous	179 (57.9)	135 (57.7)	44 (58.7)		
Primary c-section	42 (13.6)	35 (15.0)	7 (9.3)		
Secondary c-section	64 (20.7)	46 (19.7)	18 (24.0)		
Vaginal operative	24 (7.8)	18 (7.7)	6 (8.0)		
Infant sex frequencies (%)				0.45 (1)[g]	.59
Female infants	126 (40.6)	98 (41.7)	28 (37.3)		
Male infants	184 (59.4)	137 (58.3)	47 (62.7)		
Infant APGAR [h] values after 10-min frequencies (%)				8058.0[f]	.37
10	279 (92.7)	213 (93.4)	66 (90.4)		
9	16 (5.3)	12 (5.3)	4 (5.5)		
8	5 (1.7)	3 (1.3)	2 (2.7)		
7	1 (0.3)	0 (0.0)	1 (1.4)		

[a]NPC: no personal coaching.
[b]PC: personal coaching.
[c]Empirical α-error.
[d]*t* test (df) values.
[e]PHQ: Patient Health Questionnaire.
[f]Mann-Whitney *U* test.
[g]Chi-square (df) value.
[h]APGAR: Appearance, Pulse, Grimace, Activity and Respiration.

XSL•FO
RenderX

Table 3. Descriptive statistics of outcome measures.

Time point	Sample size (N)	Minimum	Maximum	Mean (SD)	SE
FFA-14[a]					
T1	285	15.00	54.00	33.50 (6.95)	0.41
T5	287	15.00	53.00	35.01 (6.77)	0.40
T7	228	17.00	55.00	35.57 (7.80)	0.52
Edinburgh Postnatal Depression Scale					
T1	291	1.00	26.00	11.36 (4.98)	0.29
T2	294	0.00	24.00	10.90 (4.93)	0.29
T3	302	0.00	25.00	10.35 (4.92)	0.28
T4	297	0.00	26.00	10.14 (5.23)	0.30
T5	294	0.00	26.00	9.99 (5.66)	0.33
T6	217	0.00	28.00	9.12 (5.94)	0.40
T7	247	0.00	26.00	8.87 (5.48)	0.35
State-Trait Anxiety Inventory-state scale					
T1	290	22.00	75.00	48.59 (11.04)	0.65
T2	293	22.00	79.00	48.42 (11.12)	0.65
T3	301	20.00	80.00	47.17 (11.67)	0.67
T4	297	22.00	77.00	46.88 (12.16)	0.71
T5	293	21.00	79.00	47.19 (12.21)	0.71
T6	217	20.00	80.00	42.12 (12.77)	0.87
T7	244	20.00	79.00	42.40 (12.47)	0.80
State-Trait Anxiety Inventory-trait scale					
T1	289	21.00	79.00	48.83 (10.69)	0.63
T2	293	22.00	73.00	48.23 (10.84)	0.63
T3	301	21.00	75.00	46.82 (11.15)	0.64
T4	297	22.00	78.00	45.77 (11.33)	0.66
T5	293	22.00	76.00	45.91 (11.66)	0.68
T6	214	20.00	78.00	42.50 (12.89)	0.88
T7	242	20.00	75.00	42.91 (12.52)	0.81
Pregnancy-Related Anxiety Questionnaire					
T1	288	10.00	50.00	25.76 (7.54)	0.44
T2	293	12.00	48.00	25.89 (7.61)	0.44
T3	301	10.00	50.00	25.50 (7.94)	0.46
T4	297	10.00	47.00	25.36 (7.80)	0.45
T5	294	10.00	50.00	25.55 (8.46)	0.49
BMI					
T0	456	13.36	59.18	24.70 (5.58)	0.26
T1	455	14.67	48.67	26.42 (5.25)	0.25
T2	362	15.70	47.62	28.09 (5.12)	0.27
T3	362	16.04	48.67	28.48 (4.99)	0.26
T4	354	16.37	49.37	28.82 (5.07)	0.27
T5	337	16.71	50.77	29.06 (5.11)	0.28
T6	209	17.63	41.51	26.15 (5.06)	0.35

XSL•FO
RenderX

Time point	Sample size (N)	Minimum	Maximum	Mean (SD)	SE
T7	247	14.70	49.72	25.47 (5.39)	0.34

[a]FFA-14: Freiburg Mindfulness Inventory.

Manipulation Check (FFA-14)

As covariates a 2 (group) × 2 (coaching) × 3 (time) ANOVA with PHQ stress evaluation and the number of children at home were used. The assumption on sphericity was significantly violated (P<.001) and thus Huynh-Feldt correction was applied (ε=.936). There was a significant main effect of time (P<.001).

Moreover, a significant interaction effect was observed between group and time ($F_{1.873,344.619}$=4.560, P=.01, η^2=0.024, ω^2=0.012). Dunn's post hoc test (Ψ_{Dunn}=1.9) revealed a significant increase from T1 to T5 and T7 for the IG, but also an increase from T1 and T5 to T7 for the CG. However, the IG score is significantly higher than the CG score at T5. In addition, there was a significant interaction effect between coaching and time ($F_{1.873,344.619}$=4.585, P=.01, η^2=0.024, ω^2=0.012). Dunn's post hoc test (Ψ_{Dunn}=2.4) revealed a significant increase from T1 to T5 and T7 but only for the PC group (refer to Figure 2).

Figure 2. Interaction effect between coaching and time regarding FFA-14 scores. GW: gestational week; T1: 28th GW; T5=36th GW; T7=5 months postpartum; NPC: no personal coaching; PC: personal coaching.

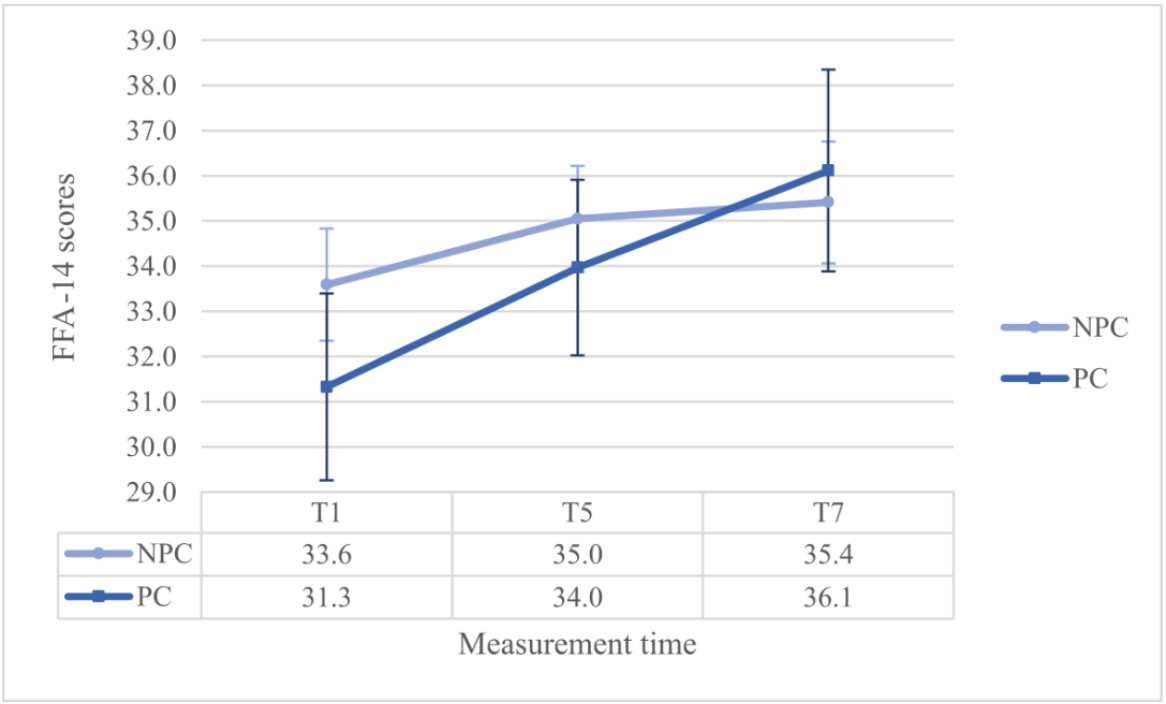

Furthermore, a significant main effect of the PHQ stress evaluation was observed. There were no other significant main (P≥.14) or interaction effects (P≥.09).

We calculated a power of 1–β=.927 for medium-sized between-subject effects and 1–β=.816 for small within-subject effects (f=0.10). Only small between-subject effects cannot be ruled out sufficiently with 1–β=0.232.

Symptoms of Depression According to EPDS

As covariates a 2 (group) × 2 (coaching) × 7 (time) ANOVA with PHQ stress evaluations and the highest school educational level were used. The assumption on sphericity was significantly violated (P<.001) and thus Huynh-Feldt correction was applied (ε=.793).

We found significant main effects of the PHQ stress evaluation (P<.001) and maternal education (P=.01). Moreover, the main effect of the group was significant ($F_{1,147}$=7.845, P=.006, η^2=0.051, ω^2=0.04), revealing a generally lower EPDS score for the IG (mean 9.3, SE 0.5) than for the CG (mean 11.2, SE 0.4).

The coaching time—interaction effect was significant, too (P=.04). Most importantly, this interaction term was enriched by a significant 3-way interaction term ($F_{4.758,699.423}$=3.033, P=.01, η^2=0.020, ω^2=0.009): Dunn's post hoc test (Ψ_{Dunn}=4.1) revealed significant differences at T4 and T5 between the IG and the CG, but only in the PC group (Figures 3 and 4).

Findings were nonsignificant for all other between- (P≥.12) and within-subject effects (P≥.07). The power was 1–β=0.983 for medium-sized (f=0.40) between-subject effects. Only small within-subject and between-subject effects (f=0.10) cannot be ruled out sufficiently with 1–β=.590 and 1–β=.313.

XSL•FO
RenderX

Figure 3. Group x time—interaction effect regarding EPDS scores for the NPC group. eMBI: electronic mindfulness-based intervention; GW: gestational week; PP: postpartum; T1=28th GW; T2=30th GW; T3=32nd GW; T4=34th GW, T5=36th GW, T6=1 month, T7 = 5 months PP, TAU: treatment as usual.

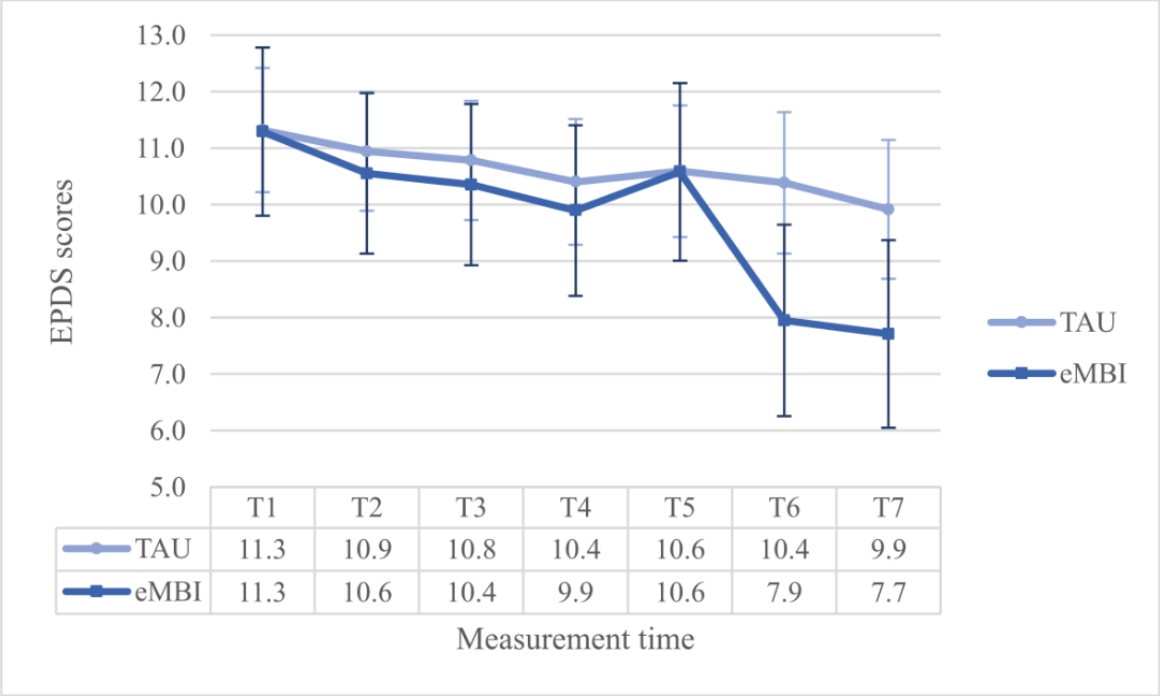

	T1	T2	T3	T4	T5	T6	T7
TAU	11.3	10.9	10.8	10.4	10.6	10.4	9.9
eMBI	11.3	10.6	10.4	9.9	10.6	7.9	7.7

Figure 4. Group x time-interaction effect regarding EPDS scores for the PC group. T1 = 28th gestational week (GW), T2 = 30th GW, T3 = 32nd GW, T4 = 34th GW, T5 = 36th GW, T6 = 1 month postpartum (pp), and T7 = 5 months pp, TAU: treatment as usual, eMBI: electronic mindfulness-based intervention.

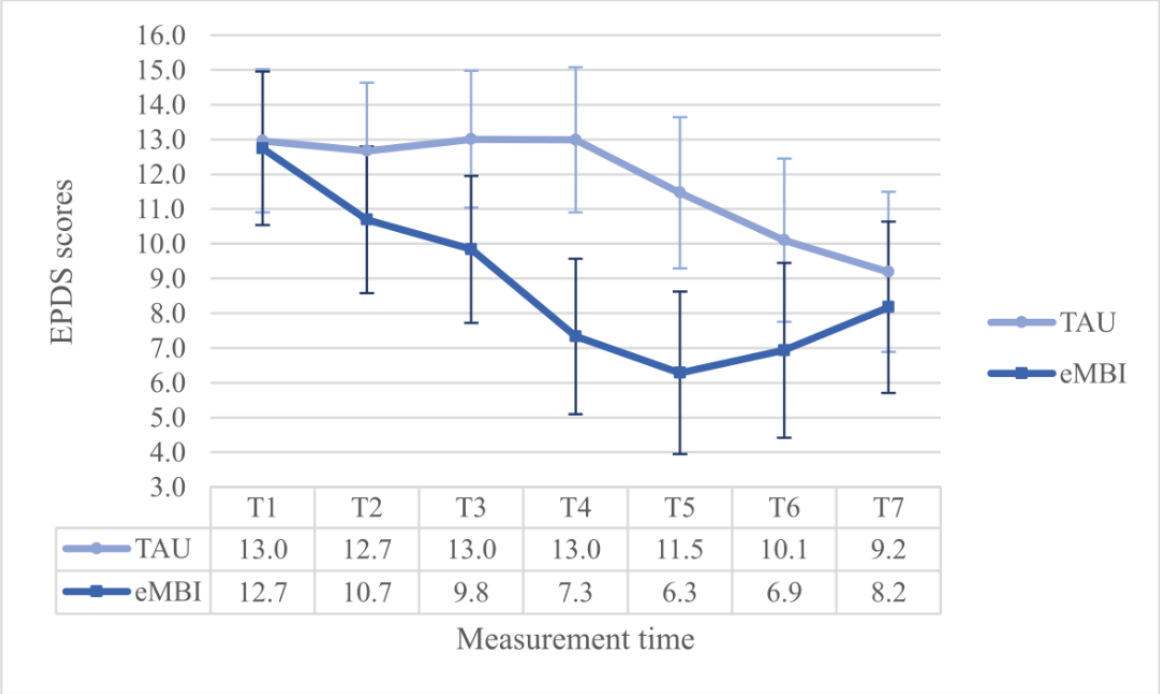

	T1	T2	T3	T4	T5	T6	T7
TAU	13.0	12.7	13.0	13.0	11.5	10.1	9.2
eMBI	12.7	10.7	9.8	7.3	6.3	6.9	8.2

General Anxiety According to the STAI

As covariates a 2 (group) 2 (coaching) 7 (time) MANOVA with PHQ stress evaluations, the number of children at home, and the highest educational level were used. There was a significant main effect of PHQ stress evaluation (P<.001). Moreover, a significant main effect of group was observed (P=.04), which was enriched by a significant 2-way interaction term between group and time (P=.01) as well as a significant 3-way interaction term ($F_{12,129}$=2.361, P=.01, η^2=0.0180, ω^2=0.100).

Findings were nonsignificant for all other between- (P≥.29) and within-subject effects (P≥.23). The power was 1–β=.920 for medium-sized (f=0.40) within- and .981 for medium-sized between-subject effects. Only small within-subject and between-subject effects (f=0.10) cannot be ruled out sufficiently with 1–β=0.175 and 1–β=0.307. Therefore, we calculated two

XSL•FO
RenderX

2 (group) 2 (coaching) 7 (time) post hoc ANOVAs (one for each STAI-measure) with PHQ stress evaluations, the number of children at home, and the highest educational level as covariates.

State Anxiety According to the STAI-S

The assumption on sphericity was significantly violated (P<.001) and thus Huynh-Feldt correction was applied (ε=.817). A significant main effect of the PHQ stress evaluation was observed (P<.001). Moreover, there was a significant main effect of the group (P<.05).

However, there was a significant interaction effect between group and time (P=.01), which was enriched by a significant 3-way interaction term ($F_{4.902,686.330}$=4.422, P=.001, η^2=0.031, ω^2=0.016): Dunn's post hoc test (Ψ_{Dunn}=8.8) revealed significant differences between the IG and the CG at T3, T4, and T5, but only in the PC group (Figures 5 and 6).

Figure 5. Group x time-interaction effect regarding STAI-S scores for the NPC group. GW: gestational week; PP: postpartum; T1: 28th GW; T2: 30th GW; T3: 32nd GW; T4: 34th GW; T5 = 36th GW; T6: 1 month PP; T7: 5 months PP; TAU: treatment as usual; eMBI: electronic mindfulness-based intervention.

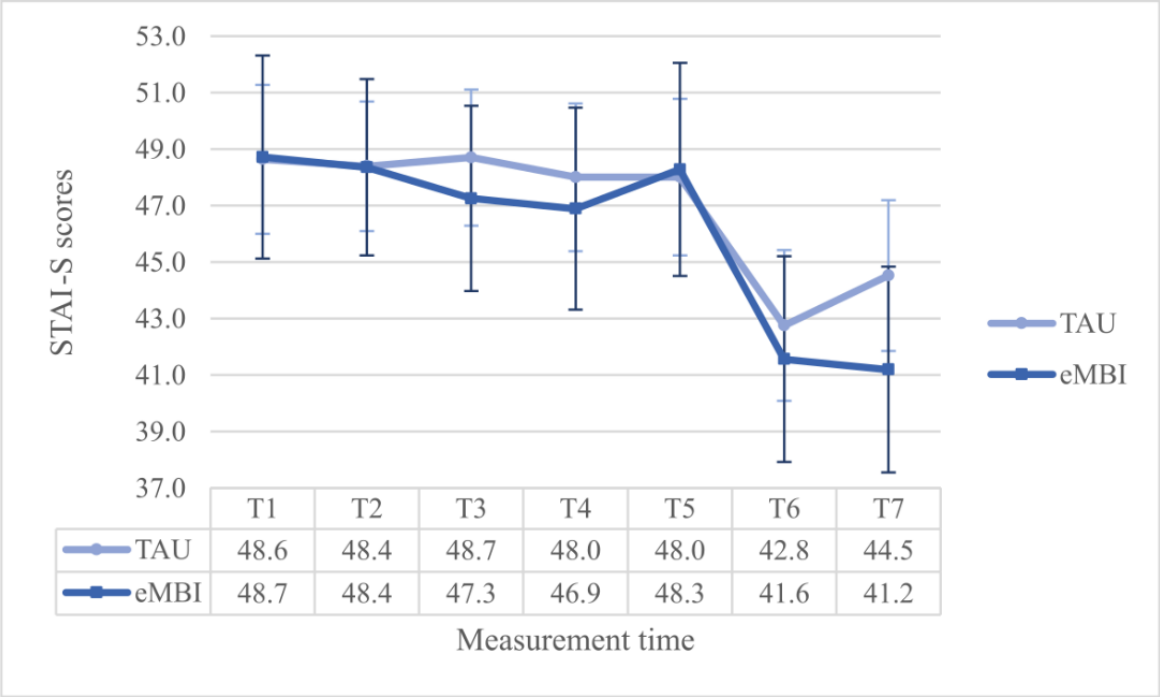

	T1	T2	T3	T4	T5	T6	T7
TAU	48.6	48.4	48.7	48.0	48.0	42.8	44.5
eMBI	48.7	48.4	47.3	46.9	48.3	41.6	41.2

XSL•FO
RenderX

Figure 6. Group x time-interaction effect regarding STAI-S scores for the PC group. eMBI: electronic mindfulness-based intervention; GW: gestational week; PP: postpartum; T1: 28th GW; T2: 30th GW; T3: 32nd GW; T4: 34th GW; T5: 36th GW; T6: 1 month PP; T7: 5 months PP; TAU: treatment as usual.

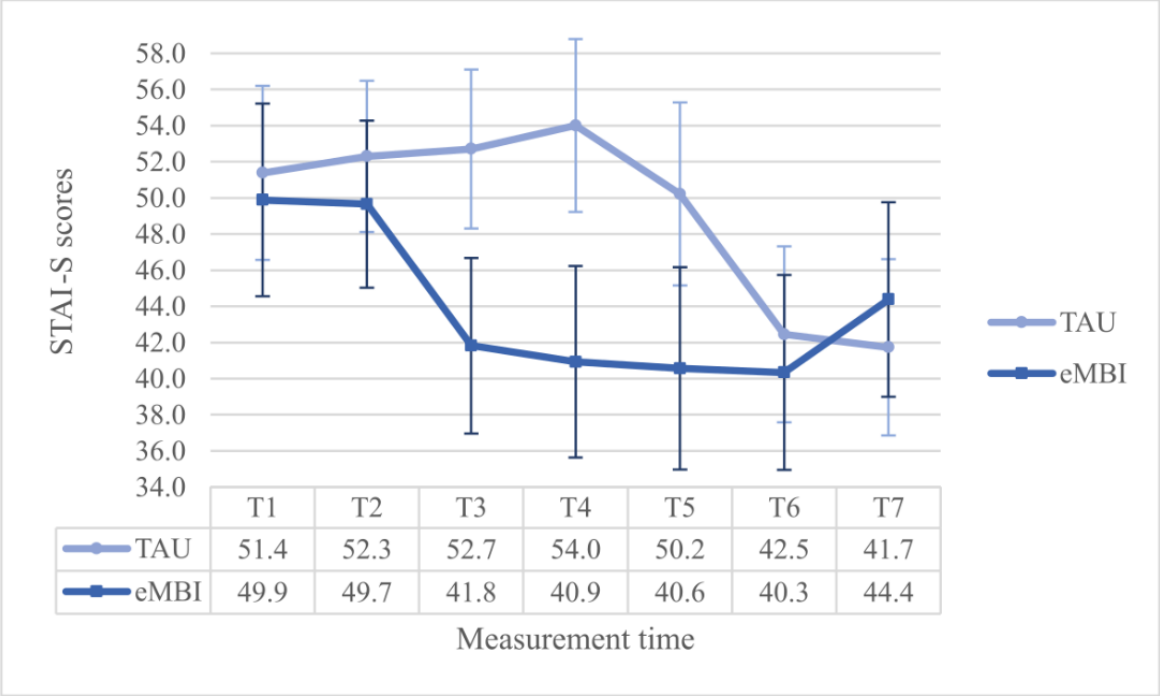

	T1	T2	T3	T4	T5	T6	T7
TAU	51.4	52.3	52.7	54.0	50.2	42.5	41.7
eMBI	49.9	49.7	41.8	40.9	40.6	40.3	44.4

Trait Anxiety According to the STAI-T

The assumption on sphericity was significantly violated (P<.001) and thus Greenhouse-Geisser correction was applied (ε=.691). A significant main effect of the PHQ stress evaluations was observed (P<.001).

Moreover, there was a significant interaction effect between group and time (P=.02). However, this interaction term was enriched by a significant 3-way interaction term ($F_{4.902,686.330}$=4.689, P=.001, η^2=0.032, ω^2=0.015): the Dunn post hoc test (Ψ_{Dunn}=8.0) revealed significant differences between the IG and the CG at T4 and T5, but only in the PC group (Figures 7 and 8).

Figure 7. Group x time-interaction effect regarding STAI-T scores for the NPC group. eMBI: electronic mindfulness-based intervention; GW: gestational week; PP: postpartum; T1: 28th GW; T2: 30th GW; T3: 32nd GW; T4: 34th GW; T5: 36th GW; T6: 1 month PP; T7: 5 months PP; TAU: treatment as usual.

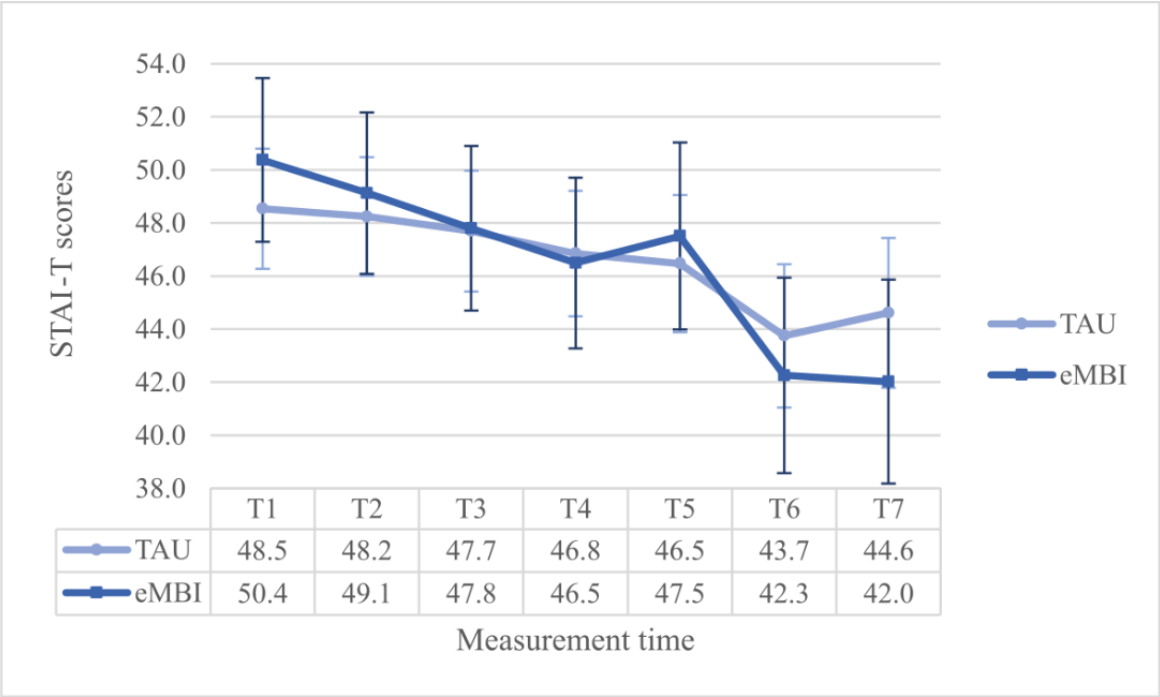

	T1	T2	T3	T4	T5	T6	T7
TAU	48.5	48.2	47.7	46.8	46.5	43.7	44.6
eMBI	50.4	49.1	47.8	46.5	47.5	42.3	42.0

XSL•FO
RenderX

Figure 8. Group x time-interaction effect regarding STAI-T scores for the PC group. eMBI: electronic mindfulness-based intervention; GW: gestational week; PP: postpartum; T1: 28th GW; T2: 30th GW; T3: 32nd GW; T4: 34th GW; T5: 36th GW; T6: 1 month PP; T7: 5 months PP; TAU: treatment as usual.

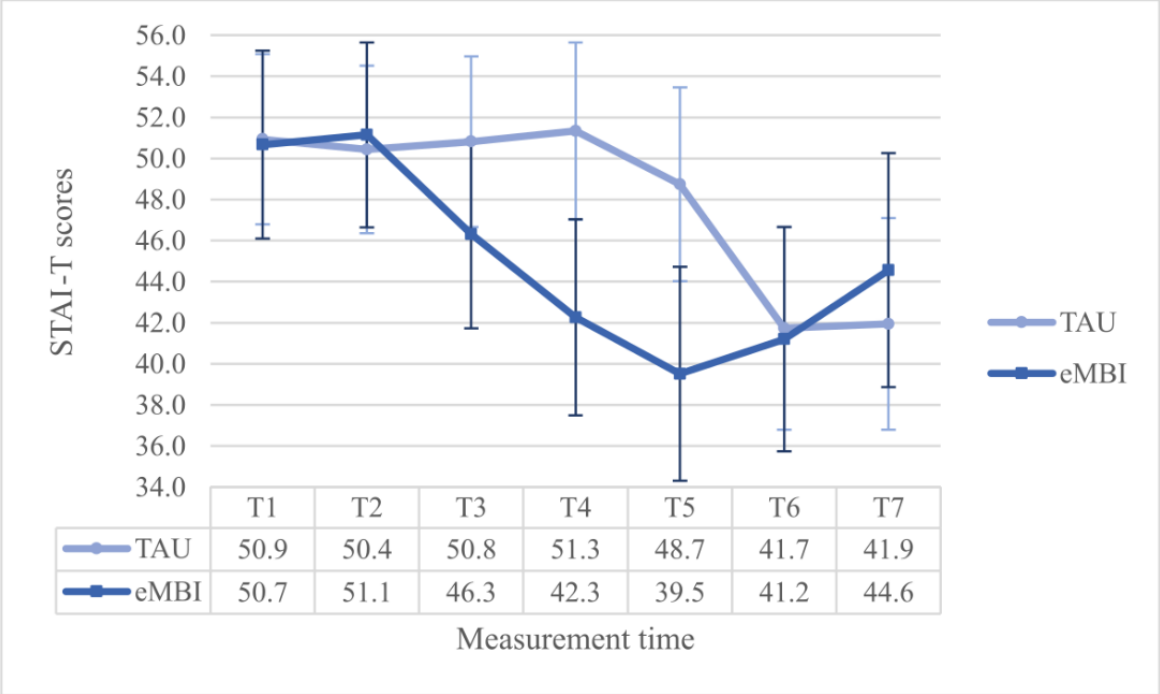

Pregnancy-Related Anxiety According to the PRAQ-R

As covariates a 2 (group) 2 (coaching) 5 (time) ANOVA with PHQ stress evaluation and number of children were used. Huynh-Feldt correction was applied (ε=.812) as the assumption on sphericity was significantly violated (P<.001). We found significant main effects of the PHQ stress evaluation (P<.001) and number of children (P<.001).

No other significant differences were found between- (P≥.10) or within-subject effects (P≥.14). In this analysis, the power was additionally virtually 1–β=.998 for small within-subject effects (f=0.10). Furthermore, it was 1–β=0.959 for medium-sized between-subject effects (f=0.25). Only small between-subject effects cannot be ruled out sufficiently with 1–β=0.264.

BMI

As a covariate a 2 (group) × 2 (coaching) × 8 (time) – ANOVA with the highest school educational level was used. The assumption on sphericity was significantly violated (P<.001) and thus Greenhouse-Geisser correction was applied (ε=.508). There was a significant main effect of time (P<.001).

The group × time--interaction effect was significant, too (P=.04). Most importantly, this interaction term was enriched by a significant three-way-interaction term ($F_{3.555,\ 444.416}$=4.732, P=.002, η^2=0.036, ω^2=0.013): the Dunn post hoc test (Ψ_{Dunn}=1.5) revealed significant differences at all assessments between the IG and the CG in the NPC group and at all assessments except T0 in the PC group. In the PC group, the IG generally reports a lower BMI than the CG. However, in the NPC group, it is the opposite (Figures 9 and 10).

Moreover, we found a significant main effect of maternal education (P=.01). Furthermore, there was a significant interaction effect between the groups with or without coaching (P=.01), however, as there was a significant 3-way interaction term as reported above, we did not further investigate this interaction effect.

Findings were nonsignificant for all other between- (P≥.57) and within-subject effects (P≥.19). In this analysis, the power for even small within-subject effects (f=0.10) was virtually 1. The power was 1–β=0.983 for large (f=0.40) between-subject effects. Medium-sized (f=0.25) and small between-subject effects cannot be ruled out sufficiently with 1–β=.680, 1–β=.142.

XSL•FO
RenderX

Figure 9. Group x time-interaction effect regarding BMI for the NPC group. eMBI: electronic mindfulness-based intervention; GW: gestational week; PP: postpartum; T0: before pregnancy; T1: 28th GW; T2: 30th GW; T3: 32nd GW; T4: 34th GW; T5: 36th GW; T6: 1 month PP; T7: 5 months PP; TAU: treatment as usual.

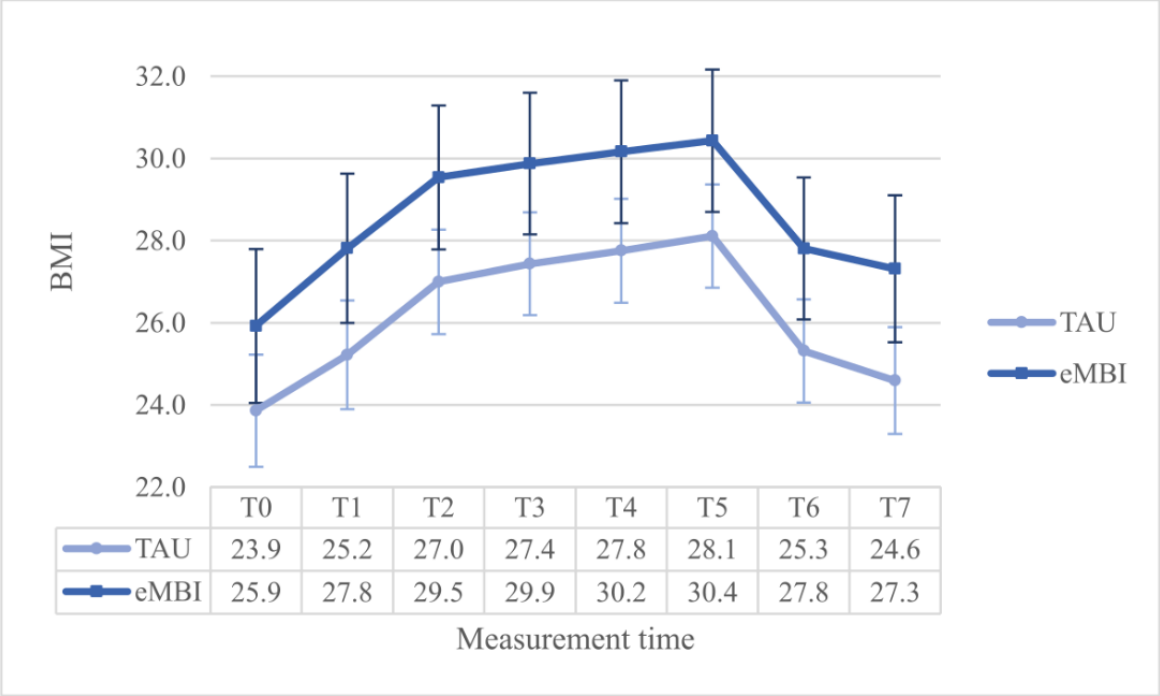

Figure 10. Group x time-interaction effect regarding BMI for the PC group. eMBI: electronic mindfulness-based intervention; GW: gestational week; PP: postpartum; T0: before pregnancy; T1: 28th GW; T2: 30th GW; T3: 32nd GW; T4: 34th GW; T5: 36th GW; T6: 1 month PP; T7: 5 months PP; TAU: treatment as usual.

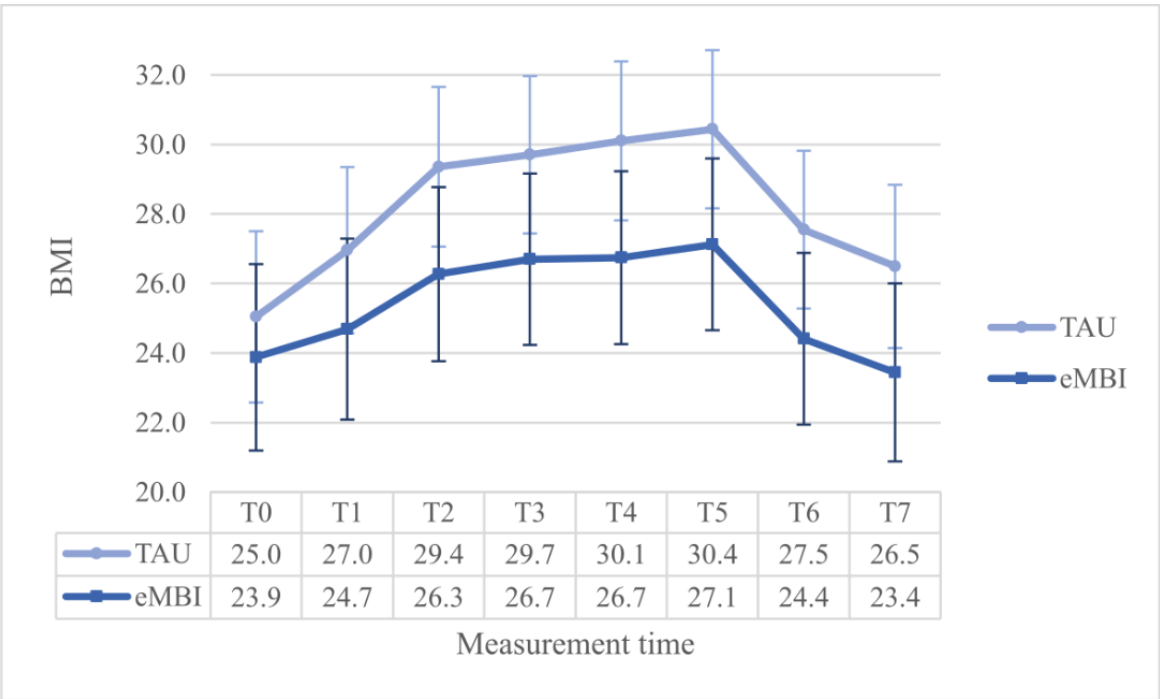

Discussion

Principal Findings

The aim of the present RCT subanalysis was to examine the effectiveness of a mindfulness-based blended care approach on maternal depression, anxiety, pregnancy- or birth-related anxiety, as well as weight control. Our results show that the eMBI alone has the potential to significantly improve birth- and pregnancy-related anxiety and mindfulness, with long-lasting effects up to 5 months after delivery. However, combining the eMBI with 2.3 PC sessions on average also significantly reduced state and trait anxiety and depression already during pregnancy and, as a side effect, revealed a significant positive influence on maternal BMI. These results favor blended care models combining digital interventions with a minimal amount of personal support by demonstrating the significant effectiveness

XSL•FO
RenderX

on perinatal mental health as well as on GWG during pregnancy and the postpartum period. Regarding maternal weight gain, pregnant women in the IG combined with PC generally reported a significantly lower BMI than women in the CG.

Comparison to Previous Work

The observation that a mindfulness intervention improves the mental health and well-being of pregnant mothers is consistent with findings from previous studies [13,16]. Looking deeper, studies using a mere digital approach often failed to find a benefit of such interventions for psychological distress while face-to-face interventions usually did [41,42]. On the other hand, programs that can be delivered digitally, such as the eMBI program used in our study, appear to be particularly beneficial for a high retention rate [43,44]. Therefore, the results of this subanalysis support the hypothesis that the failure of digital interventions to show effectiveness in lowering depression and general anxiety in contrast to face-to-face interventions might suggest that, in order to positively impact general mental health, a small personal component needs to be added, whereas the positive effects on birth anxiety can be achieved digitally [45]. Therefore, the results of the present study emphasize the importance of personal support in reducing depressive and anxiety symptoms as well as GWG, as these symptoms decreased significantly when participants had 2 or 3 PC sessions on average. In line with this, Bright et al [46] demonstrated the positive effect of interpersonal psychotherapy on maternal well-being during the perinatal period. Favorable findings that support the superiority of a blended care approach compared to a mere digital intervention have been recently published by Martin-Key et al [47] within in the framework of an anonymous online survey in the United Kingdom. Overall, 38% of 829 participating women showed a preference for a blended care approach in order to assess perinatal mental health, whereas only 12% preferred a digital-only consultation [47].

Regarding gestational weight gain and maternal BMI as secondary outcome measures, our results support the hypothesis that digital mindfulness interventions have the potential to positively influence health behaviors including stress and weight management during pregnancy [48]. In particular blended care approaches seem to be effective in providing health-promoting lifestyle modifications and complementing traditional diet and exercise programs [49]. However, in contrast to our study previous ones have not been generated as part of randomized-controlled trials and should be interpreted with caution. Furthermore, differing from previous research, the impact of PC in addition to eMBI on maternal BMI even reached significance in the present sub-analysis of our RCT [48]. However, the number of high-quality and large-scale studies is scarce so far and further research, preferably in the form of an RCT, should be conducted in the future to investigate the feasibility and cost-effectiveness of implementing digital tools into standard clinical routine.

In terms of the respective demographic characteristics results showed that the level of education was significantly associated with the EPDS, STAI-S and STAI-T, and the BMI. According to the current status of the literature former studies find that lower educational levels and socioeconomic status are associated with higher prepregnancy BMI and a higher rate of excessive gestational weight gain [50]. Vice versa, education often correlates not only with socioeconomic status but also with the pursuit of a health-promoting lifestyle [51]. Further negative associations on a significant level have been found between perinatal depression and anxiety on the one hand and educational level on the other hand. The bidirectional association between educational attainment and maternal mental health has been published before [52,53].

Thus, our study results support the use of a blended care approach not only regarding mental health but also in order to promote healthy lifestyle habits and should be taken into account while implementing digital intervention tools into clinical routine. In contrast to previous research, our study results were generated as part of an RCT contributing to good validity and reliability.

Strengths and Limitations

Our study has the potential to provide high-quality implementation knowledge of a complex mHealth eMBI and coaching intervention for the perinatal period. Being the largest trial across Europe with screening numbers of more than 5000 pregnant women, the Mind: Pregnancy study has the potential to provide high-quality data on the effectiveness of digital eMBI and coaching interventions for preventive and alternative treatment of perinatal mental health problems. Indeed, the results of this evaluation yield important findings that will potentially support the improved health of women and children. The main strength of our study lies in its prospective, longitudinal design with a follow-up period of up to 5 months after childbirth and the inclusion of a CG. The intervention and the peripartum assessments were provided as part of an app and are thus easily and universally applicable and cost-effective. Another strength of our study is that mental health was examined according to a multidimensional approach based on DSM or ICD criteria. In addition, a broad range of confounders were considered.

However, the applicability of the study is limited by several factors that should be considered. First, PC was provided on an as-needed basis during the 8-week intervention period and not in a standardized way, making it difficult to identify a best practice guideline for optimal outcomes. Second, the number of sessions varied, and it remains unclear at what point in the last trimester PC might be most effective in improving mental health. This should be the subject of further research based on the findings of this study. Third, in these analyses, the power to detect in particular small between-subject effects was low. Finally, but not least, the number of observations varies between subgroups. Thus, the results of the interaction terms should be interpreted with caution and should be evaluated in further research. If applicable, a standardized personal component with a fixed amount of sessions should be implemented to improve the validity and reliability of the study findings.

In order to further prove the effect on maternal weight gain, it would be interesting to evaluate the development of the BMI with similar baseline characteristics.

Despite the limitations, our study presents evidence that a blended care approach is potentially highly effective not only

XSL•FO
RenderX

in optimizing perinatal mental health but also in affecting maternal weight gain as a synergistic effect. Due to the increasing accessibility of digital interventions, widespread implementation of these interventions should be encouraged.

Conclusion

In conclusion, our results provide evidence that a blended care model combining digital with a small amount of personal care can potentiate the effect of eMBIs on depressive symptoms and general anxiety with lasting effects up to several months postpartum and can significantly influence maternal BMI. This intervention is a low-cost, easy-to-use intervention for pregnant women to improve their well-being and mental health. However, especially in light of the ongoing digitalization in medical care, it is essential to investigate the potential of individualized, personalized care in everyday clinical practice.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 7307 KB - jmir_v27i1e56230_app1.pdf]

References

1. Pearson RM, Carnegie RE, Cree C, Rollings C, Rena-Jones L, Evans J, et al. Prevalence of prenatal depression symptoms among 2 generations of pregnant mothers: the avon longitudinal study of parents and children. JAMA Netw Open 2018;1(3):e180725 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0725] [Medline: 30646025]
2. Arifin SRM, Cheyne H, Maxwell M. Review of the prevalence of postnatal depression across cultures. AIMS Public Health 2018;5(3):260-295 [FREE Full text] [doi: 10.3934/publichealth.2018.3.260] [Medline: 30280116]
3. Leach LS, Poyser C, Fairweather‐schmidt K. Maternal perinatal anxiety: a review of prevalence and correlates. Clinical Psychologist 2020;21(1):4-19. [doi: 10.1111/cp.12058]
4. Galea LAM, Frokjaer VG. Perinatal depression: embracing variability toward better treatment and outcomes. Neuron 2019;102(1):13-16 [FREE Full text] [doi: 10.1016/j.neuron.2019.02.023] [Medline: 30946818]
5. Andersson L, Sundström-Poromaa I, Wulff M, Aström M, Bixo M. Implications of antenatal depression and anxiety for obstetric outcome. Obstet Gynecol 2004;104(3):467-476. [doi: 10.1097/01.AOG.0000135277.04565.e9] [Medline: 15339755]
6. Li J, Yin J, Waqas A, Huang Z, Zhang H, Chen M, et al. Quality of life in mothers with perinatal depression: a systematic review and meta-analysis. Front Psychiatry 2022;13:734836 [FREE Full text] [doi: 10.3389/fpsyt.2022.734836] [Medline: 35242060]
7. Lutsiv O, McKinney B, Foster G, Taylor VH, Pullenayegum E, McDonald SD. Pregnancy complications associated with the co-prevalence of excess maternal weight and depression. Int J Obes (Lond) 2015;39(12):1710-1716. [doi: 10.1038/ijo.2015.119] [Medline: 26095247]
8. Kingston D, Tough S, Whitfield H. Prenatal and postpartum maternal psychological distress and infant development: a systematic review. Child Psychiatry Hum Dev 2012;43(5):683-714. [doi: 10.1007/s10578-012-0291-4] [Medline: 22407278]
9. Herba CM, Glover V, Ramchandani PG, Rondon MB. Maternal depression and mental health in early childhood: an examination of underlying mechanisms in low-income and middle-income countries. Lancet Psychiatry 2016;3(10):983-992 [FREE Full text] [doi: 10.1016/S2215-0366(16)30148-1] [Medline: 27650772]
10. Epel E, Laraia B, Coleman-Phox K, Leung C, Vieten C, Mellin L, et al. Effects of a mindfulness-based intervention on distress, weight gain, and glucose control for pregnant low-income women: a quasi-experimental trial using the ORBIT model. Int J Behav Med 2019;26(5):461-473 [FREE Full text] [doi: 10.1007/s12529-019-09779-2] [Medline: 30993601]
11. Viswanathan M, Siega-Riz AM, Moos M, Deierlein A, Mumford S, Knaack J, et al. Outcomes of maternal weight gain. Evid Rep Technol Assess (Full Rep) 2008(168):1-223. [Medline: 18620471]
12. Goldstein RF, Harrison CL, Teede HJ. Editorial: the importance of gestational weight gain. Obes Rev 2020;21(10):e13073. [doi: 10.1111/obr.13073] [Medline: 32608189]
13. Guo P, Zhang X, Liu N, Wang J, Chen D, Sun W, et al. Mind-body interventions on stress management in pregnant women: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2021;77(1):125-146. [doi: 10.1111/jan.14588] [Medline: 33048358]
14. Müller M, Matthies LM, Goetz M, Abele H, Brucker SY, Bauer A, et al. Effectiveness and cost-effectiveness of an electronic mindfulness-based intervention (eMBI) on maternal mental health during pregnancy: the mindmom study protocol for a randomized controlled clinical trial. Trials 2020;21(1):933 [FREE Full text] [doi: 10.1186/s13063-020-04873-3] [Medline: 33203471]
15. Haga SM, Drozd F, Lisøy C, Wentzel-Larsen T, Slinning K. Mamma Mia - a randomized controlled trial of an internet-based intervention for perinatal depression. Psychol Med 2019;49(11):1850-1858 [FREE Full text] [doi: 10.1017/S0033291718002544] [Medline: 30191779]

XSL•FO
RenderX

16. Lönnberg G, Jonas W, Unternaehrer E, Bränström R, Nissen E, Niemi M. Effects of a mindfulness based childbirth and parenting program on pregnant women's perceived stress and risk of perinatal depression-Results from a randomized controlled trial. J Affect Disord 2020;262:133-142. [doi: 10.1016/j.jad.2019.10.048] [Medline: 31733457]
17. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
18. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
19. Chiauzzi E, Newell A. Mental health apps in psychiatric treatment: a patient perspective on real world technology usage. JMIR Ment Health 2019;6(4):e12292 [FREE Full text] [doi: 10.2196/12292] [Medline: 31008711]
20. Willcox JC, Chai D, Beilin LJ, Prescott SL, Silva D, Neppe C, et al. Evaluating engagement in a digital and dietetic intervention promoting healthy weight gain in pregnancy: mixed methods study. J Med Internet Res 2020;22(6):e17845 [FREE Full text] [doi: 10.2196/17845] [Medline: 32442153]
21. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux P, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
22. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. Br J Psychiatry 1987;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
23. Bergant AM, Nguyen T, Heim K, Ulmer H, Dapunt O. German language version and validation of the Edinburgh postnatal depression scale. Dtsch Med Wochenschr 1998 Jan 16;123(3):35-40. [doi: 10.1055/s-2007-1023895] [Medline: 9472218]
24. Rubertsson C, Börjesson K, Berglund A, Josefsson A, Sydsjö G. The Swedish validation of edinburgh postnatal depression scale (EPDS) during pregnancy. Nord J Psychiatry 2011;65(6):414-418. [doi: 10.3109/08039488.2011.590606] [Medline: 21728782]
25. Wisner KL, Sit DKY, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry 2013;70(5):490-498 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.87] [Medline: 23487258]
26. Newham JJ, Westwood M, Aplin JD, Wittkowski A. State-trait anxiety inventory (STAI) scores during pregnancy following intervention with complementary therapies. J Affect Disord 2012;142(1-3):22-30. [doi: 10.1016/j.jad.2012.04.027] [Medline: 22959685]
27. Huizink AC, Delforterie MJ, Scheinin NM, Tolvanen M, Karlsson L, Karlsson H. Adaption of pregnancy anxiety questionnaire-revised for all pregnant women regardless of parity: PRAQ-R2. Arch Womens Ment Health 2016;19(1):125-132 [FREE Full text] [doi: 10.1007/s00737-015-0531-2] [Medline: 25971851]
28. Huizink AC, Menting B, Oosterman M, Verhage ML, Kunseler FC, Schuengel C. The interrelationship between pregnancy-specific anxiety and general anxiety across pregnancy: a longitudinal study. J Psychosom Obstet Gynaecol 2014;35(3):92-100 [FREE Full text] [doi: 10.3109/0167482X.2014.944498] [Medline: 25093467]
29. Reck C, Zimmer K, Dubber S, Zipser B, Schlehe B, Gawlik S. The influence of general anxiety and childbirth-specific anxiety on birth outcome. Arch Womens Ment Health 2013;16(5):363-369. [doi: 10.1007/s00737-013-0344-0] [Medline: 23558948]
30. Huizink AC, Mulder EJH, Robles de Medina PG, Visser GHA, Buitelaar JK. Is pregnancy anxiety a distinctive syndrome? Early Hum Dev 2004;79(2):81-91. [doi: 10.1016/j.earlhumdev.2004.04.014] [Medline: 15324989]
31. Buchheld N, Grossman P, Walach H. Measuring mindfulness in insight meditation (Vipassana) and meditation-based psychotherapy: the development of the freiburg mindfulness inventory (FMI). Journal for meditation and meditation research 2001;1(1):11-34. [doi: 10.1007/978-3-319-46322-3_4]
32. Sauer S, Walach H, Offenbächer M, Lynch S, Kohls N. Measuring mindfulness: a rasch analysis of the freiburg mindfulness inventory. Religions 2011;2(4):693-706. [doi: 10.3390/rel2040693]
33. Walach H, Buchheld N, Buttenmüller V, Kleinknecht N, Schmidt S. Measuring mindfulness—the freiburg mindfulness inventory (FMI). Personality and Individual Differences 2006;40(8):1543-1555. [doi: 10.1016/j.paid.2005.11.025]
34. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. JAMA 1999;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
35. Gräfe K, Zipfel S, Herzog W, Löwe B. Screening psychischer Störungen mit dem “Gesundheitsfragebogen für Patienten (PHQ-D)“. Diagnostica 2004;50(4):171-181. [doi: 10.1026/0012-1924.50.4.171]
36. Little RJA. A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association 1988;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
37. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates. Inc; 1977.
38. Dunn OJ. Multiple comparisons among means. Journal of the American Statistical Association 1961;56(293):52-64. [doi: 10.1080/01621459.1961.10482090]

XSL•FO
RenderX

39. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
40. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
41. Dhillon A, Sparkes E, Duarte RV. Mindfulness-based interventions during pregnancy: a systematic review and meta-analysis. Mindfulness (N Y) 2017;8(6):1421-1437 [FREE Full text] [doi: 10.1007/s12671-017-0726-x] [Medline: 29201244]
42. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: a randomised controlled study of online mindfulness during pregnancy. Midwifery 2018;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
43. Shi Z, MacBeth A. The effectiveness of mindfulness-based interventions on maternal perinatal mental health outcomes: a systematic review. Mindfulness (N Y) 2017;8(4):823-847 [FREE Full text] [doi: 10.1007/s12671-016-0673-y] [Medline: 28757900]
44. Evans K, Rennick-Egglestone S, Cox S, Kuipers Y, Spiby H. Remotely delivered interventions to support women with symptoms of anxiety in pregnancy: mixed methods systematic review and meta-analysis. J Med Internet Res 2022;24(2):e28093 [FREE Full text] [doi: 10.2196/28093] [Medline: 35166688]
45. Lahtinen O, Salmivalli C. An effectiveness study of a digital mindfulness-based program for upper secondary education students. Mindfulness 2020;11(11):2494-2505. [doi: 10.1007/s12671-020-01462-y]
46. Bright KS, Charrois EM, Mughal MK, Wajid A, McNeil D, Stuart S, et al. Interpersonal psychotherapy to reduce psychological distress in perinatal women: a systematic review. Int J Environ Res Public Health 2020;17(22):8421 [FREE Full text] [doi: 10.3390/ijerph17228421] [Medline: 33203014]
47. Martin-Key NA, Spadaro B, Schei TS, Bahn S. Proof-of-concept support for the development and implementation of a digital assessment for perinatal mental health: mixed methods study. J Med Internet Res 2021;23(6):e27132 [FREE Full text] [doi: 10.2196/27132] [Medline: 34033582]
48. Vieten C, Laraia BA, Kristeller J, Adler N, Coleman-Phox K, Bush NR, et al. The mindful moms training: development of a mindfulness-based intervention to reduce stress and overeating during pregnancy. BMC Pregnancy Childbirth 2018;18(1):201 [FREE Full text] [doi: 10.1186/s12884-018-1757-6] [Medline: 29859038]
49. van der Windt M, van der Kleij RM, Snoek KM, Willemsen SP, Dykgraaf RHM, Laven JSE, et al. Impact of a blended periconception lifestyle care approach on lifestyle behaviors: before-and-after study. J Med Internet Res 2020;22(9):e19378 [FREE Full text] [doi: 10.2196/19378] [Medline: 32996885]
50. Garay SM, Sumption LA, Pearson RM, John RM. Risk factors for excessive gestational weight gain in a UK population: a biopsychosocial model approach. BMC Pregnancy Childbirth 2021;21(1):43 [FREE Full text] [doi: 10.1186/s12884-020-03519-1] [Medline: 33423656]
51. O'Brien EC, Segurado R, Geraghty AA, Alberdi G, Rogozinska E, Astrup A, Salvesen, i-WIP (International Weight Management in Pregnancy) Collaborative Group. Impact of maternal education on response to lifestyle interventions to reduce gestational weight gain: individual participant data meta-analysis. BMJ Open 2019;9(8):e025620 [FREE Full text] [doi: 10.1136/bmjopen-2018-025620] [Medline: 31375602]
52. Cena L, Mirabella F, Palumbo G, Gigantesco A, Trainini A, Stefana A. Prevalence of maternal antenatal and postnatal depression and their association with sociodemographic and socioeconomic factors: a multicentre study in Italy. J Affect Disord 2021;279:217-221 [FREE Full text] [doi: 10.1016/j.jad.2020.09.136] [Medline: 33069119]
53. Dagher RK, Bruckheim HE, Colpe LJ, Edwards E, White DB. Perinatal depression: challenges and opportunities. J Womens Health (Larchmt) 2021;30(2):154-159 [FREE Full text] [doi: 10.1089/jwh.2020.8862] [Medline: 33156730]

Abbreviations

CG: control group
CONSORT: Consolidated Standards of Reporting Trials
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
eMBI: electronic mindfulness-based interventions
EPDS: Edinburgh Postnatal Depression Scale
EU-GDPR: European General Data Protection Regulation
FFA-14/FMI-14: Freiburg Mindfulness Inventory
GWG: gestational weight gain
IG: intervention group
NPC: no personal coaching
PC: personal coaching
PHQ: Patient Health Questionnaire
PRAQ-R: Pregnancy-Related Anxiety Questionnaire
RCT: randomized controlled trial
SPIRIT: Standard Protocol Items: Recommendations for Intervention Trials
STAI: State-Trait Anxiety Inventory

XSL•FO
RenderX

STAI-S: State-Trait Anxiety Inventory-state scale
STAI-T: State-Trait Anxiety Inventory-trait scale
TAU: treatment as usual

Edited by A Coristine; submitted 10.01.24; peer-reviewed by A Herzeg, S Dufour, N Ardian; comments to author 27.09.24; revised version received 26.11.24; accepted 26.11.24; published 24.02.25.

Please cite as:
Hassdenteufel K, Müller M, Abele H, Brucker SY, Graf J, Zipfel S, Bauer A, Jakubowski P, Pauluschke-Fröhlich J, Wallwiener M, Wallwiener S
Improving Maternal Mental Health and Weight Control With a Mindfulness Blended Care Approach: Insights From a Randomized Controlled Trial
J Med Internet Res 2025;27:e56230
URL: https://www.jmir.org/2025/1/e56230
doi:10.2196/56230
PMID:39992700

©Kathrin Hassdenteufel, Mitho Müller, Harald Abele, Sara Yvonne Brucker, Johanna Graf, Stephan Zipfel, Armin Bauer, Peter Jakubowski, Jan Pauluschke-Fröhlich, Markus Wallwiener, Stephanie Wallwiener. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of In-Person and Mobile Exercise Coaching on Psychosocial Factors Affecting Exercise Adherence in Inactive Women With Obesity: 20-Week Randomized Controlled Trial

Christina Gjestvang[1], PhD; John Magne Kalhovde[2], MSc; Elene Mauseth Tangen[1], MSc; Hege Clemm[1], PhD; Lene Annette Hagen Haakstad[1], PhD

[1]Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway

[2]Department of Health and Exercise, School of Health Sciences, Kristiania University of Applied Science, Oslo, Norway

Corresponding Author:
Christina Gjestvang, PhD
Department of Sports Medicine
Norwegian School of Sport Sciences
P.O Box 4014, Ullevål Stadion
Oslo, 0806
Norway
Phone: 47 23262119
Email: christina.gjestvang@nih.no

Abstract

Background: Regular exercise may counteract obesity-related health risks, but adherence is low among individuals with obesity. Personal trainers may positively influence exercise behavior by providing motivational support. Individuals who receive regular exercise coaching are more likely to adhere to their exercise routine, compared with those who exercise individually. However, investing in personalized exercise guidance, such as a personal trainer, can be expensive for the individual. Thus, integrating web-based coaching could be a more economically sustainable option, offering both flexibility and reduced costs compared with in-person coaching only. Yet, research is needed to assess the effect of hybrid models in improving psychosocial factors among women with obesity.

Objective: This 20-week, pragmatic randomized controlled trial aimed to investigate the effect of weekly in-person coaching compared with 2 combinations of in-person and web-based coaching on 5 psychosocial factors in women with obesity (BMI ≥30 kg/m^2).

Methods: Participants were invited through Facebook and Instagram advertisements posted by various fitness clubs across rural and urban locations in Norway (7 different counties and 12 different municipalities). Women with low activity (n=188; <150 minutes of moderate-intensity physical activity/week; 42.7, SD 10.5 years; mean BMI of 35.1, SD 6.9 kg/m^2) were allocated into 3 groups of in-person coaching—weekly (n=47), twice weekly (n=47), and once monthly (n=47)—and as controls (n=47). Those with twice weekly and once monthly in-person coaching received web-based coaching of 15 minutes during weeks without in-person coaching. Data included background variables, motivation (Behavioral Regulation in Exercise Questionnaire-2), barriers, self-efficacy (The Self-Efficacy Survey), social support (Social Support Questionnaire), and health-related quality of life (36-Item Short Form Health Survey [SF-36]).

Results: A total of 120 (64%) out of 188 participants completed baseline and postintervention assessments. A minor difference was observed in one item of the SF-36, where all intervention groups reported a greater "change in health last year" than the control group (mean difference: 14.2-17.1, 95% CI 2.04-29.5; g=0.79-1.14; P≤.01). No other effects were found on the selected psychosocial factors. It should be noted that controls reported higher intrinsic motivational regulation at baseline than intervention groups (mean difference: 0.43-0.93; P≤.05). All intervention arms exercised more frequently than controls (mean difference: 1.1-1.5; P≤.001), with no differences in weekly exercise frequency between the 3 intervention arms (P=.30).

Conclusions: We found no effects on motivation, barriers, self-efficacy, perceived social support, or other health domains compared with controls. All intervention groups reported a slight improvement in self-perceived health in 1 of the 8 subscales of the SF-36. Combined in-person and web-based coaching may give a minor improvement in self-perceived health in women with obesity. However, the lack of impact on motivation, barriers, and self-efficacy warrants further research.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT05792657; https://clinicaltrials.gov/study/NCT05792657

(J Med Internet Res 2025;27:e68462) doi:10.2196/68462

KEYWORDS

barriers; exercise behavior; exercise coaching; health-related quality of life; motivation; obesity; psychosocial factors; self-efficacy; social support

Introduction

Obesity (BMI ≥30 kg/m^2) is a major global health challenge, significantly economically burdening the health care and welfare systems [1,2]. Worldwide, the obesity rate has nearly tripled in the past 35 years, and it is estimated that 1 billion adults will have obesity in 2030 [3-5]. It affects women more than men, with 1 in 5 women and 1 in 7 men living with obesity [5,6]. Obesity increases the risk of noncommunicable diseases, such as cardiovascular disease, metabolic disease, and type 2 diabetes [7-11]. It is also negatively associated with mental health conditions, including anxiety, personality disorders, binge eating, and schizophrenia [12].

Structured exercise programs have become a key element in obesity management worldwide, offering benefits for both physiological and psychological well-being [13-15]. Regular exercise can mitigate obesity-related health risks and is widely recommended, along with diet advice and behavioral strategies, as part of a comprehensive treatment strategy [9,11,16-18]. The feasibility of these programs is frequently challenged by barriers such as time constraints, logistical issues, and accessibility, particularly for women with obesity [13,14]. In addition, the complexity of implementing lifestyle changes often results in low adherence rates, which limits the associated health benefits [19,20]. This challenge is further compounded by the fact that individuals with obesity tend to be more sedentary and less active than those of normal weight, and women are less likely than men (32% versus 23%) to meet physical activity recommendations (150 to 300 minutes of moderate-intensity or 75 to 150 minutes of vigorous-intensity physical activity per week) [21-24].

Low levels of motivation, limited strategies to overcome barriers, lack of self-efficacy, inadequate social support, and poor physical or mental health negatively impact an individual's ability to maintain regular exercise [19,25-28]. Exercise coaching may positively influence these factors by fostering autonomy (eg, offering choices and alternatives), building competence (eg, providing positive feedback and promoting skill development), and improving relatedness (eg, offering emotional support) [28-31]. As a result, these elements collectively have the potential to improve adherence to exercise [28,32-34]. Randomized controlled trials (RCTs) have shown that personal trainers help clients adhere more consistently to exercise by equipping them with strategies to overcome barriers, fostering a sense of community, and enhancing self-efficacy through knowledge and confidence building [31,32,35-37].

Further, perceived social support has been shown to encourage regular participation in exercise, particularly among women rather than men [25,38,39]. In addition, obesity is often associated with lower self-perceived health, and individuals who perceive their health as poor may face physical limitations that hinder participation in regular exercise [27,40]. A personal trainer can develop customized exercise programs that address these physical limitations, helping clients overcome barriers and ultimately improve their perceived health-related quality of life [41,42].

Although individuals with obesity might acknowledge the benefits of personal trainers in providing exercise motivation and support, their use is often limited by barriers such as accessibility, time constraints, socioeconomic resources, and cost concerns [43-45]. The high cost of hiring a personal trainer also often limits the frequency of follow-up sessions. Mobile exercise apps have emerged as a popular and accessible alternative [14]. When used in conjunction with or as an extension of qualified exercise professionals, mobile apps may enhance exercise adherence [14]. Hence, integrating web-based coaching is an attractive alternative due to its cost-effectiveness and accessibility. Web-based coaching offers flexibility and lower costs compared with in-person coaching [46,47], while in-person coaching provides unique advantages, such as fostering deeper personal relationships that may be vital for sustained exercise adherence. Consequently, hybrid models combining in-person and web-based coaching have gained popularity, as they provide the flexibility and convenience of web-based coaching with the direct interaction and personalized feedback of in-person coaching. These models show potential for promoting behavior change by meeting a wider range of client needs. However, the effectiveness of such combined approaches, particularly for women with obesity, requires further investigation [46,47]. Hence, further research is needed to evaluate the potential of hybrid models in addressing key psychosocial factors to exercise adherence, particularly for women with obesity. While these models are promising in combining accessibility with personalized support, their ability to positively influence motivation, barriers, self-efficacy, social support, and health-related quality of life remains unclear. In addition, to provide a "real-life" approach, weekly in-person coaching is needed as a comparison group, as it represents the most common mode of in-person coaching among clients [48]. Understanding how such hybrid models can optimize psychosocial outcomes in this population is crucial for developing effective and sustainable interventions.

In women with obesity, our 20-week, 4-armed, pragmatic RCT aimed to investigate the effects of weekly in-person coaching compared with 2 different combinations of in-person and web-based coaching on motivation, barriers, self-efficacy, social support, and health-related quality of life.

XSL•FO
RenderX

Methods

Overview

This 20-week, 4-armed, pragmatic RCT's primary outcomes were exercise adherence, aerobic capacity, and muscular strength. Hence, this study presents a secondary analysis of the intervention's effect on perceived motivation, barriers to exercise, self-efficacy, social support, and health-related quality of life.

Participants were invited through Facebook and Instagram advertisements posted by various fitness clubs across rural and urban locations in Norway (7 different counties and 12 different municipalities). The recruitment was ongoing in February 2023, with eligible participants enrolled continuously. Unfortunately, data on the total number of interested participants were not collected. The inclusion criteria were BMI ≥30 kg/m^2, aged 18 to 65 years, no membership in a fitness club 6 months before recruitment, being classified as low activity (<150 minutes of moderate-intensity or 75 minutes of vigorous-intensity physical activity per week), Norwegian speaking, and having a smartphone. Exclusion criteria were chronic disease or injuries (eg, severe hypertension 180/110 mm Hg, heart disease, lung disease, or functional impairment) that hindered participation in exercise and planned leave during the intervention period. To ensure that none of the participants had any contraindications for exercise, all participants underwent a general health screening by their general practitioner.

A total of 188 participants were included, and informed consent was signed to participate. A blinded statistician performed random allocation (1:1:1:1) following a simple computer-based randomization program (R for Windows, version 4.2.2; R Foundation for Statistical Computing), and the participants were allocated to (1) once a week (in-person exercise coaching 100% [IP100], n=47), (2) twice a month (in-person exercise coaching 50% [IP50], n=47), or (3) once a month (in-person exercise coaching 25% [IP25], n=47) in-person coaching, or as controls (n=47). The intervention groups were provided the same weekly frequency of follow-up, but in IP50 and IP25, a total of 50% (2/4) and 75% (3/4) of the follow-up sessions were replaced with 15-minute digital sessions, respectively. An overview of intervention arms is shown in Textbox 1.

Textbox 1. Overview of intervention arms in this 20-week, pragmatic randomized controlled trial.

In-person exercise coaching 100% (IP100)

- One weekly in-person exercise coaching.
- A total of 20 hours of in-person coaching.
- Access to the ABEL-app (Abel Technologies AS) including individual exercise programs, progress charts, nutritional advice, and motivational notices.
- 20 weeks fitness club membership.

In-person exercise coaching 50% (IP50)

- Two monthly in-person exercise coaching and 15 minutes of web-based coaching during weeks without in-person coaching.
- Total of 10 hours of in-person coaching during the intervention.
- Access to the ABEL-app including individual exercise programs, progress charts, nutritional advice, and motivational notices.
- 20 weeks fitness club membership.

In-person exercise coaching 25% (IP25)

- One monthly in-person exercise coaching and 15 minutes of web-based coaching during weeks without in-person coaching.
- Total of 5 hours of in-person coaching during the intervention.
- Access to the ABEL-app including individual exercise programs, progress charts, nutritional advice, and motivational notices.
- 20 weeks fitness club membership.

Control group

- Was asked to continue with normal life.
- Received the "Norwegian Directorate of Health's" recommendations for physical activity and nutrition.
- Access to the ABEL-app to register physical activity and exercise, but were not provided with individual exercise programs, nutritional advice, and motivational notices.

A total of 23 exercise professionals, all employed as personal trainers, were recruited to provide follow-up for study participants. Trainers needed at least 3 years of part-time experience and to have led at least 80 exercise sessions/month recently. Of the 23 trainers, 10 had acquired certification as personal trainers, 8 held relevant bachelor's or master's degrees in fields such as exercise science or physiotherapy, and 5 had bachelor's or master's degrees in health and social care education. The trainers were instructed to follow up with each participant and create customized weekly exercise programs

XSL•FO
RenderX

based on their needs. They also offered lifestyle advice, such as coaching in maintaining a healthy diet. Our study design did not allow for further masking of study participants or the personal trainers.

Measures

Overview

A standardized electronic survey was used to obtain information about exercise motivation, barriers to exercise, self-efficacy, social support, and health-related quality of life, as well as background variables (body weight and height, age, education, household income, and occupation). The survey took approximately 20 minutes to complete, with all questions being close ended. Of the 188 participants, 120 (64%) participants completed the questionnaire at both baseline and after 20 weeks (postintervention) and were included in the current analysis.

Exercise Motivation

The measurement of exercise motivation was based on a Norwegian version of the validated Behavioral Regulation in Exercise Questionnaire-2 (BREQ-2) [49], assessing the stages of the self-determination continuum of motivation. The BREQ-2 comprises 5 subscales, with a total of 19 statements. For each statement, the participants rated the significance of each statement as a personal motive to engage, or not engage in exercise, on a 5-point scale (from "0 not true for me" to "4 very true for me"). A sum score (from 0 to 4) for each subscale was calculated by adding scores from each statement, divided by the number of statements. The BREQ-2 has good internal consistency for all 5 subscales (Cronbach α >0.7) [49,50]. In our study, Cronbach α values for the 5 subscales were as follows: 0.81 and 0.87 (intrinsic regulation), 0.69 and 0.80 (identified regulation), 0.73 and 0.80 (introjected regulation), 0.83 and 0.81 (external regulation), and 0.82 and 0.79 (amotivation) for before and after the intervention, respectively.

Barriers to Exercise

Barriers to exercise were assessed based on factors identified in a large-scale study among the Norwegian adult population [51] and a previous study among fitness club members [52]. The questionnaire section contained 18 barriers categorized into 4 subscales. The participants rated how limiting they perceived each barrier to be on a 3-point scale, ranging from "1 not important to me" to "3 very important to me." By adding the score from each barrier divided by the number of statements, a sum score (from 1 to 3) for each subscale was calculated. The barrier subscales in the previous large-scale Norwegian study [51] have been shown to have Cronbach α values above 0.70 for the practical, health-related, and affective-cognitive subscales, but lower for the priority subscale. In this study, Cronbach α values for the 4 subscales were as follows: 0.53 and 0.50 (priority), 0.46 and 0.51 (practical), 0.22 and 0.39 (health-related), and 0.61 and 0.63 (affective-cognitive) for before and after the intervention, respectively. As expected, Cronbach α was lowest for subscales with the fewest items, and these subscales should be interpreted with care. Internal consistency of the whole barrier questionnaire was higher, as determined by Cronbach α as follows: 0.66 at baseline and 0.72 at posttest.

Self-Efficacy

An abbreviated and validated version of the Self-Efficacy Survey, comprising 2 subscales and a total of 12 statements, was used to measure self-efficacy for exercise [53]. The participants rated each statement on a 5-point scale (from "1 I know I cannot" to "5 I know I can") in how confident they were to increase or continue with regular exercise under a wide range of conditions. For each subscale, a sum score (from 1 to 5) was calculated by adding the scores of each statement, divided by the number of statements. The Self-Efficacy Survey has been shown to have good test-retest reliability and internal consistency [53]. In our study, Cronbach α values for the 2 subscales were as follows: 0.89 and 0.90 ("sticking to it"), and 0.63 and 0.55 ("making time for exercise") for before and after the intervention, respectively. Cronbach α was lowest for the subscale with 4 items and this subscale should be interpreted with care. The internal consistency of the entire Self-Efficacy Survey was high, as determined by the Cronbach α are as follows: 0.80 at baseline and 0.90 at posttest.

Social Support

Social support for exercise was based on a validated survey developed by Sallis et al [54], consisting of 13 statements. The participants rated each statement on a 5-point scale, ranging from "1 none" to "5 very often," in how often their family or friends had been supportive of them exercising. A total social support score was calculated (from 13 to 65) by adding the scores from each of the statements, where higher scores demonstrated greater social support for exercise. The 2 sections (family and friends) were merged. The original survey has been shown to have acceptable test-retest reliability and internal consistency [54]. In our study, Cronbach α values for the 13 statements concerning social support were 0.83 and 0.87 at baseline and posttest.

Health-Related Quality of Life

The 36-Item Short Form Health Survey (SF-36) was used to measure subjective evaluations of health-related quality of life. The questionnaire encompasses 8 health-related subscales with 2 to 10 questions or statements and a single item that provides an indication of perceived change in health last year. Each set of response options is designed to quantify the intensity of the participant's feelings or limitations of health and well-being [55]. As recommended, we precoded all items on a 0 to 100 range with a high score presenting a more favorable health state. A sum score for each subscale was then calculated by adding scores from each statement, divided by the number of statements. A Norwegian version of SF-36 has been shown to have good internal consistency for all subscales (Cronbach α >0.80) [56]. In our study, Cronbach α values for the 8 subscales pre/post was as follows: 0.84 and 0.85 ("physical functioning"), 0.84 and 0.90 ("role limitations due to physical health"), 0.84 and 0.86 ("role limitations due to emotional problems"), 0.85 and 0.19 (energy/fatigue), 0.71 and 0.31 ("emotional well-being"), 0.76 and 0.71 ("social functioning"), 0.88 and 0.82 (pain), and 0.71 and 0.79 ("general health") for before and after the intervention, respectively. The internal consistency of the entire questionnaire section was high, as determined by

XSL•FO
RenderX

Cronbach α, were as follows: 0.91 at baseline and 0.93 at posttest.

Participation in Exercise and Physical Activity

Data on exercise and physical activity participation were collected using the ABEL app (Abel Technologies AS). Information, including type of activity, intensity, duration, and frequency, was recorded by participants during individual sessions and by personal trainers during follow-up sessions.

Statistical Analysis

The analyses were performed with SPSS statistics (version 28; IBM Corporation), following a predefined analysis plan and before unmasking the intervention arms. The normal distribution of the data was assessed with the Shapiro-Wilk test. We used appropriate statistical tests to compare baseline characteristics between the current sample and those who dropped out, as well as to identify differences between groups. These included a chi-square test, an independent 2-tailed *t* test, or a 1-way ANOVA, as dictated by the data. Postintervention, group comparisons were done using a 1-way ANOVA with Bonferroni post hoc comparisons. We also calculated Hedges *g* effect size, with values around 0.2, 0.5, and 0.8 interpreted as weak, medium, and strong effect sizes, respectively. Data are presented as mean (SD), numbers (%), mean changes, mean differences, 95% CI, effect sizes, and *P* value. The level of statistical significance was set at $P \leq .05$ and adjusted as appropriate for the 1-way ANOVA and Bonferroni post hoc comparison ($P \leq .01$).

In this study, targeting a specific population of women with obesity who have low activity presented challenges in accurately estimating attrition rates due to limited comparable data. Given the lack of reliable data on expected dropout rates within this demographic, the research group assumed recruitment of 200 participants was appropriate.

Ethical Considerations

The study was conducted following the ethical standards of the Helsinki Declaration (as revised in 2000) and relevant national and institutional guidelines. All participants provided written informed consent after receiving detailed study information, including their rights to withdraw at any time without consequence. The study was reviewed by the Regional Committee for Medical and Health Research Ethics (REK 2022/552476), which determined that, under the Act on Medical and Health Research (the Health Research Act 2008), full review by the Regional Committee for Medical and Health Research Ethics was not required. Ethical approval was granted by the Norwegian School of Sports Sciences Ethical Committee (Applications number 262). The study was also registered with the Norwegian Social Science Data Service (Sikt 104437) and ClinicalTrials.gov (NCT05792657).

To protect participant privacy, no identifying information was collected or published. Participants in the intervention arms received regular follow-ups by a personal trainer, while all study groups, including the control group, had access to a fitness club and an exercise app (ABEL Technologies). No financial compensation was provided to participants.

Results

A total of 120 (64%) out of 188 participants completed baseline and postintervention assessments (IP100: n=35; IP50: n=35; IP25: n=24; and controls: n=26). The flow of the participants is shown in Figure 1. Mean age of the participants collapsed was 42.7 (SD 10.5) years and the mean BMI was 35.1 (SD 6.9) kg/m^2. There were no differences between the 4 groups in age, body weight, education, or household income at baseline (all $P>.05$; Table 1). In terms of BMI, participants in the control group had a higher mean BMI (35.8, SD 5.8 kg/m^2) than those allocated to IP100 (32.8, SD 2.6 kg/m^2; $P=.02$). There were no significant differences between the current sample and study dropouts in background characteristics at baseline (data not shown).

Weekly exercise frequency over the 20-week intervention was as follows: 2.0 (SD 1.3) days/week (IP100), 2.4 (SD 1.5) days/week (IP50), 2.2 (SD 1.6) days/week (IP25), and 0.9 (SD 1.6) days/week (controls). We found no differences in weekly exercise frequency between the 3 intervention arms ($P=.30$); however, all intervention arms exercised more frequently than controls ($P \leq .001$).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart of the participants. IP100: in-person exercise coaching 100%; IP25: in-person exercise coaching 25%; IP50: in-person exercise coaching 50%.

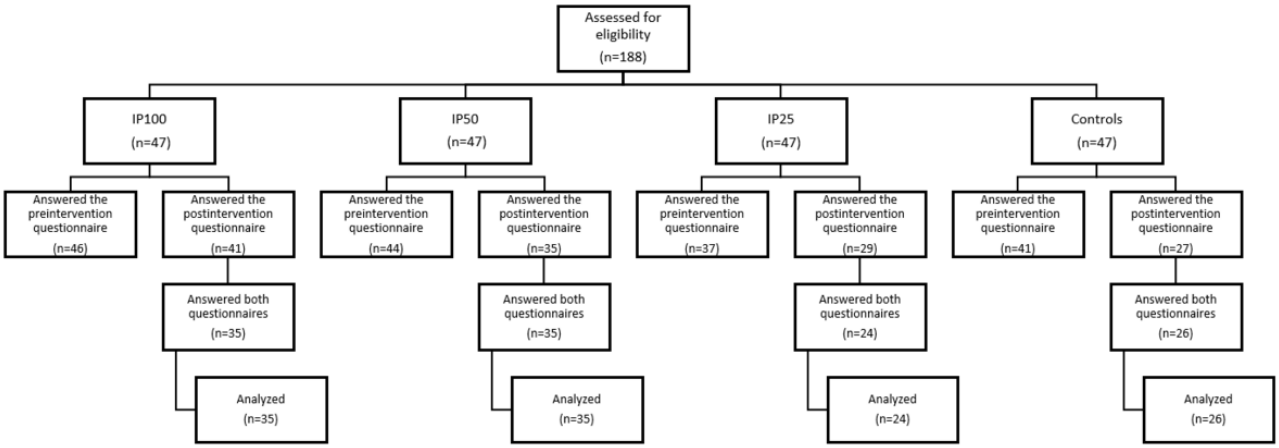

XSL•FO
RenderX

Table 1. Background characteristics, exercise motivation, barriers to exercise, self-efficacy, social support, and health-related quality of life at baseline in the 3 intervention arms and the control group.

	IP100[a] (n=35)	IP50[b] (n=35)	IP25[c] (n=24)	Controls (n=26)	*P* values
Body weight (kg), mean (SD)	94.0 (9.8)	98.4 (12.6)	99.7 (10.6)	97.3 (14.5)	.28
Body height (cm), mean (SD)	169.1 (6.2)	167.5 (5.6)	168.1 (4.0)	164.9 (5.7)	.03
Age (years), mean (SD)	45.3 (10.5)	42.6 (10.2)	43.3 (11.4)	41.1 (9.4)	.46
BMI (kg/m^2), mean (SD)	32.8 (2.6)	35.0 (3.9)	35.3 (3.5)	35.8 (5.8)	.02
Higher education (≥4 years), n (%)	7 (20.0)	9 (25.7)	8 (33.3)	10 (38.5)	.71
High household income (>US $75,000 per year), n (%)	22 (62.9)	19 (54.3)	11 (45.8)	12 (46.2)	.24
Employed outside home, n (%)	33 (94.3)	31 (88.6)	16 (66.7)	21 (80.8)	.01
Exercise motivation, mean (SD)					
Intrinsic regulation	1.7 (0.9)	2.1 (0.6)	2.1 (0.8)	2.6 (0.7)	≤.001
Identified regulation	2.1 (0.9)	2.2 (0.8)	2.1 (0.8)	2.4 (0.7)	.45
Introjected regulation	1.8 (1.1)	2.0 (1.0)	1.7 (1.1)	2.0 (1.1)	.60
External regulation	0.6 (0.8)	0.7 (0.8)	0.6 (0.9)	0.7 (1.0)	.91
Amotivation	0.4 (0.6)	0.3 (0.7)	0.4 (0.7)	0.2 (0.5)	.70
Barriers to exercise, mean (SD)					
Priority	1.9 (0.6)	1.8 (0.6)	1.8 (0.7)	1.7 (0.4)	.35
Practical	1.4 (0.4)	1.5 (0.4)	1.4 (0.4)	1.5 (0.4)	.45
Health-related	1.5 (0.4)	1.4 (0.4)	1.5 (0.4)	1.5 (0.4)	.69
Affective-cognitive	1.5 (0.3)	1.4 (0.3)	1.4 (0.3)	1.5 (0.3)	.39
Self-efficacy, mean (SD)					
Sticking to exercise	3.4 (0.9)	3.6 (0.9)	3.3 (0.9)	3.8 (1.1)	.06
Making time for exercise	3.3 (0.8)	3.5 (0.9)	3.2 (1.0)	3.4 (1.0)	.39
Social support, mean (SD)	26.7 (7.2)	26.3 (5.3)	26.7 (1.6)	25.8 (1.2)	.87
Health-related quality of life, mean (SD)					
Physical functioning	75.1 (13.4)	70.5 (16.0)	69.1 (18.2)	68.5 (22.9)	.29
Role limitations due to physical health	69.3 (35.4)	67.9 (36.2)	58.3 (47.0)	60.6 (41.3)	.86
Role limitations due to emotional problems	57.1 (46.1)	59.0 (40.5)	61.1 (43.6)	53.8 (42.2)	.86
Energy/fatigue	40.6 (10.0)	37.5 (10.2)	39.2 (8.2)	33.8 (12.9)	.16
Emotional well-being	59.9 (15.2)	53.6 (15.3)	61.6 (17.0)	51.8 (18.1)	.10
Social functioning	66.1 (30.7)	61.4 (31.3)	61.0 (30.5)	59.6 (33.4)	.66
Pain	70.3 (20.8)	67.6 (26.1)	62.4 (25.3)	59.4 (25.5)	.36
General health	52.5 (20.3)	43.0 (16.4)	49.5 (18.0)	39.5 (21.0)	.02
Change in health from 1 year ago	25.4 (32.4)	15.9 (27.1)	20.0 (32.3)	17.9 (22.3)	.50

[a]IP100: in-person exercise coaching 100%.

[b]IP50: in-person exercise coaching 50%.

[c]IP25: in-person exercise coaching 25%.

At baseline, we found that the control group scored higher on intrinsic motivational regulation than the 3 intervention groups (mean difference: 0.4-0.9; $P \leq .05$). After the intervention, we found a between groups difference in 1 of the 8 health-related subscales for the SF-36. All intervention groups reported a greater improvement in self-perceived health in 1 of the 8 subscales of SF-36 (SF-36 domain: change in health from 1 year ago), than controls (IP100: 81.4, SD 17.5, g=0.82, P=.007; IP50: 80.9, SD 17.5, g=0.79, P=.01; IP25: 84.4, SD 12.4] g=1.14, P=.002; and controls: 66.7, SD 15.9; Table 2). No other intervention effects were found (Table 3).

XSL•FO
RenderX

Table 2. Differences between intervention arms in health-related quality of life at posttest.

Comparison	Mean difference	95% CI	Hedges *g*
Health-related quality of life			
Physical functioning			
IP100[a] vs IP50[b]	5.28	–4.24 to 14.81	0.34
IP100 vs IP25[c]	2.50	–8.06 to 13.06	0.21
IP100 vs controls	3.46	–6.85 to 13.78	0.31
IP50 vs IP25	2.78	–13.35 to 7.77	–0.16
IP50 vs controls	1.82	–12.14 to 8.49	–0.11
IP25 vs controls	0.96	–10.32 to 12.24	0.07
Role limitations due to physical health			
IP100 vs IP50	0.007	–19.93 to 19.92	0.00
IP100 vs IP25	6.44	–15.64 to 28.53	0.21
IP100 vs controls	2.58	–18.99 to 24.17	0.09
IP50 vs IP25	6.45	–15.64 to 28.54	0.20
IP50 vs controls	2.59	–18.99 to 24.17	0.08
IP25 vs controls	3.85	–27.45 to 19.73	–0.12
Role limitations due to emotional problems			
IP100 vs IP50	1.90	–26.81 to 23.00	–0.05
IP100 vs IP25	5.91	–33.53 to 21.70	–0.15
IP100 vs controls	7.72	–34.71 to 19.25	–0.20
IP50 vs IP25	4.00	–31.62 to 23.61	–0.10
IP50 vs controls	5.82	–32.80 to 21.15	–0.15
IP25 vs controls	1.81	–27.68 to 31.31	–0.05
Energy/fatigue			
IP100 vs IP50	5.71	–2.98 to 14.41	0.39
IP100 vs IP25	4.93	–4.71 to 14.58	0.36
IP100 vs controls	4.12	–5.30 to 13.55	0.37
IP50 vs IP25	0.77	–10.42 to 8.87	–0.05
IP50 vs controls	1.58	–11.01 to 7.83	–0.10
IP25 vs controls	0.81	–11.11 to 9.49	–0.07
Emotional well-being			
IP100 vs IP50	3.54	–5.03 to 12.12	0.24
IP100 vs IP25	2.41	–11.93 to 7.09	–0.23
IP100 vs controls	1.08	–10.37 to 8.20	–0.09
IP50 vs IP25	5.96	–15.47 to 3.54	–0.40
IP50 vs controls	4.62	–13.92 to 4.66	–0.30
IP25 vs controls	1.33	–8.82 to 11.49	0.12
Social functioning			
IP100 vs IP50	2.14	–18.26 to 22.54	0.07
IP100 vs IP25	0.34	–22.96 to 22.28	–0.01
IP100 vs controls	1.62	–20.47 to 23.72	0.05
IP50 vs IP25	2.48	–25.10 to 20.13	–0.08

XSL•FO
RenderX

Comparison	Mean difference	95% CI	Hedges *g*
IP50 vs controls	0.52	–22.62 to 21.57	–0.02
IP25 vs controls	1.96	–22.19 to 26.12	0.06
Pain			
IP100 vs IP50	3.35	–15.18 to 21.89	0.11
IP100 vs IP25	9.11	–11.43 to 29.67	0.31
IP100 vs controls	1.53	–21.61 to 18.54	–0.05
IP50 vs IP25	5.75	–14.79 to 26.31	0.20
IP50 vs controls	4.89	–24.97 to 15.19	–0.17
IP25 vs controls	10.64	–32.60 to 11.30	–0.39
General health			
IP100 vs IP50	9.09	–1.34 to 19.52	0.53
IP100 vs IP25	7.24	–4.32 to 18.81	0.52
IP100 vs controls	8.32	–2.97 to 19.62	0.58
IP50 vs IP25	1.84	–13.41 to 9.72	–0.10
IP50 vs controls	0.76	–12.06 to 10.53	–0.04
IP25 vs controls	1.07	–11.27 to 13.43	0.07
Change in health from 1 year ago			
IP100 vs IP50	0.54	–10.10 to 11.19	0.03
IP100 vs IP25	2.94	–14.67 to 8.77	–0.19
IP100 vs controls	14.12	2.66 to 25.57	0.82
IP50 vs IP25	3.49	–15.28 to 8.30	–0.22
IP50 vs controls	13.57	2.04 to 25.10	0.79
IP25 vs controls	17.06	4.54 to 29.59	1.14

[a]IP100: in-person exercise coaching 100%.

[b]IP50: in-person exercise coaching 50%.

[c]IP25: in-person exercise coaching 25%.

XSL•FO
RenderX

Table 3. Differences between intervention arms in exercise motivation, barriers to exercise, self-efficacy, and social support.

Comparison	Mean difference	95% CI	Hedges *g*
Exercise motivation			
Intrinsic regulation			
IP100[a] vs IP50[b]	0.11	–0.45 to 0.68	0.14
IP100 vs IP25[c]	0.15	–0.47 to 0.79	0.16
IP100 vs controls	0.29	–0.32 to 0.91	0.34
IP50 vs IP25	0.04	–0.58 to 0.67	0.04
IP50 vs controls	0.17	–0.43 to 0.79	0.21
IP25 vs controls	0.13	–0.53 to 0.81	0.14
Identified regulation			
IP100 vs IP50	0.10	–0.61 to 0.39	–0.15
IP100 vs IP25	0.03	–0.52 to 0.59	0.05
IP100 vs controls	0.24	–0.29 to 0.79	0.36
IP50 vs IP25	0.14	–0.41 to 0.69	0.16
IP50 vs controls	0.35	–0.19 to 0.90	0.48
IP25 vs controls	0.21	–0.38 to 0.81	0.23
Introjected regulation			
IP100 vs IP50	0.17	–0.80 to 0.46	–0.18
IP100 vs IP25	0.13	–0.57 to 0.83	0.12
IP100 vs controls	0.01	–0.70 to 0.67	–0.02
IP50 vs IP25	0.30	–0.40 to 1.00	0.29
IP50 vs controls	0.15	–0.84 to 0.53	0.16
IP25 vs controls	0.15	–0.60 to 0.90	–0.14
External regulation			
IP100 vs IP50	0.25	–0.72 to 0.20	–0.36
IP100 vs IP25	0.008	–0.52 to 0.50	0.00
IP100 vs controls	0.08	–0.59 to 0.41	–0.13
IP50 vs IP25	0.24	–0.26 to 0.76	0.31
IP50 vs controls	0.16	–0.33 to 0.67	0.21
IP25 vs controls	0.08	–0.47 to 0.63	–0.11
Amotivation			
IP100 vs IP50	0.10	–0.35 to 0.15	–0.23
IP100 vs IP25	0.03	–0.31 to 0.24	–0.12
IP100 vs controls	0.03	–0.30 to 0.24	–0.11
IP50 vs IP25	0.06	–0.21 to 0.34	0.15
IP50 vs controls	0.06	–0.20 to 0.34	0.14
IP25 vs controls	0.00	–0.29 to 0.29	0.00
Barriers to exercise			
Priority			
IP100 vs IP50	0.07	–0.25 to 0.39	0.13
IP100 vs IP25	0.10	–0.25 to 0.46	0.21
IP100 vs controls	0.06	–0.41 to 0.28	–0.16

XSL•FO
RenderX

Comparison	Mean difference	95% CI	Hedges *g*
IP50 vs IP25	0.03	–0.32 to 0.39	0.06
IP50 vs controls	0.13	–0.48 to 0.21	–0.27
IP25 vs controls	0.17	–0.55 to 0.21	–0.37
Practical			
IP100 vs IP50	0.06	–0.17 to 0.30	0.17
IP100 vs IP25	0.05	–0.21 to 0.32	0.15
IP100 vs controls	0.13	–0.39 to 0.12	–0.38
IP50 vs IP25	0.009	–0.27 to 0.25	–0.03
IP50 vs controls	0.19	–0.46 to 0.06	–0.48
IP25 vs controls	0.18	–0.47 to 0.09	–0.49
Health-related			
IP100 vs IP50	0.02	–0.23 to 0.28	0.08
IP100 vs IP25	0.09	–0.38 to 0.19	–0.27
IP100 vs controls	0.08	–0.36 to 0.19	–0.24
IP50 vs IP25	0.12	–0.41 to 0.16	–0.31
IP50 vs controls	0.11	–0.39 to 0.16	–0.29
IP25 vs controls	0.008	–0.29 to 0.31	0.02
Affective-cognitive			
IP100 vs IP50	0.01	–0.20 to 0.16	–0.03
IP100 vs IP25	0.05	–0.25 to 0.15	–0.24
IP100 vs controls	0.09	–0.29 to 0.10	–0.36
IP50 vs IP25	0.03	–0.23 to 0.16	–0.13
IP50 vs controls	0.08	–0.28 to 0.11	–0.24
IP25 vs controls	0.04	–0.26 to 0.17	–0.16
Self-efficacy			
Sticking to exercise			
IP100 vs IP50	0.10	–0.45 to 0.65	0.12
IP100 vs IP25	0.04	–0.65 to 0.57	–0.05
IP100 vs controls	0.17	–0.42 to 0.77	0.24
IP50 vs IP25	0.14	–0.75 to 0.47	–0.14
IP50 vs controls	0.07	–0.52 to 0.67	0.08
IP25 vs controls	0.21	–0.44 to 0.86	0.24
Making time for exercise			
IP100 vs IP50	0.07	–0.68 to 0.54	–0.07
IP100 vs IP25	0.24	–0.92 to 0.43	–0.26
IP100 vs controls	0.20	–0.87 to 0.46	–0.23
IP50 vs IP25	0.17	–0.85 to 0.51	–0.16
IP50 vs controls	0.13	–0.80 to 0.53	–0.13
IP25 vs controls	0.04	–0.69 to 0.77	0.04
Social support			
IP100 vs IP50	1.45	–6.63 to 3.72	–0.17
IP100 vs IP25	0.69	–6.43 to 5.04	–0.09
IP100 vs controls	0.37	–5.97 to 5.23	–0.05

XSL•FO
RenderX

Comparison	Mean difference	95% CI	Hedges *g*
IP50 vs IP25	0.76	–4.97 to 6.50	0.09
IP50 vs controls	1.08	–4.52 to 6.69	0.13
IP25 vs controls	0.32	–5.80 to 6.45	0.04

[a]IP100: in-person exercise coaching 100%.

[b]IP50: in-person exercise coaching 50%.

[c]IP25: in-person exercise coaching 25%.

Discussion

Principal Findings

This pragmatic RCT investigated the effectiveness of weekly in-person coaching compared with 2 different combinations of in-person and web-based coaching on motivation, barriers, self-efficacy, social support, and health-related quality of life among women with obesity. While no intervention effects were detected for most of the measured variables, a slight improvement was reported in self-perceived health in 1 of the 8 subscales of the SF-36. This improvement was consistent across all intervention groups when compared with the control group from before to after the intervention. However, it is important to note that the observed difference was limited to a single subscale of the SF-36, indicating that the interventions had only a minimal impact on overall health-related quality of life.

The improvement in the "change in health" subscale of the SF-36 across all intervention groups suggests that combining in-person and web-based exercise coaching may contribute to enhanced self-perceived health among women with obesity. This is consistent with a previous RCT involving 143 inactive women with excess body weight, which found that those who received in-person exercise coaching by a personal trainer reported improved self-perceived health after 12 weeks of intervention [42]. Interestingly, this RCT also found that those who exercised alone reported better self-perceived health compared with the control group [42]. This may be due to the fact that regular physical activity, even at low levels (≥1 day/week), has been shown to improve self-perceived health [57]. In a previous longitudinal prospective study among novice exercisers in a fitness club setting, regular exercise in the fitness club (≥2 days/week) was associated with high self-perceived health (odds ratio 3.53, 95% CI 1.60-7.82; *P*=.002) [58]. Therefore, regular exercise itself may improve self-perceived health, emphasizing exercise as a key factor in overall well-being. Although the use of a personal trainer may not directly influence self-perceived health, it seems to bolster exercise adherence, thus supporting long-term health benefits. This appears to be consistent regardless of in-person or web-based coaching, as we did not find differences in exercise frequency between the 3 intervention groups. However, since this improvement in our study was limited to just 1 of the 8 subscales of the SF-36, further research is necessary to determine whether this represents a true and clinically significant effect.

The intervention did not demonstrate significant effects on exercise motivation, which may be explained by several factors. One key factor is the baseline differences in intrinsic motivational regulation between the control group and the intervention arms, with the control group having higher intrinsic motivation at baseline. This disparity may have created a ceiling effect, limiting further improvements in motivation both within and between groups. However, all groups scored between 1.7 and 2.6 on the 5-point scale, indicating room for improvement, irrespective of the initial differences. In addition, variations in the coaching approaches of the personal trainers, which were not controlled for in this study, could influence exercise motivation. According to self-determination theory (SDT), interactions between the personal trainer and the client may enhance motivation for exercise by satisfying the basic psychological needs (autonomy, competence, and relatedness) [59]. Personal trainers who adopt a need-supportive coaching approach, such as providing meaningful choices, acknowledging client feelings, and offering constructive feedback, may foster these needs and thereby improve exercise motivation [30]. For instance, one study investigating the effects of an SDT-based physical activity program among adults with obesity (BMI ≥40 kg/m^2) found that perceived autonomy support and the satisfaction of basic psychological were positively associated with increases in autonomous motivation and negatively associated with controlled motivation and amotivation [60]. On the other hand, certain coaching approaches, which include elements like guilt-inducing language or not fully considering the clients' preferences, may inadvertently impact these psychological needs and reduce motivation [30].

By offering scheduled appointments, free follow-up sessions, and access to exercise equipment, the intervention effectively minimized many logistical and financial barriers to exercise. However, we found no changes in exercise barriers, with all intervention groups reporting high scores on barriers such as "priority" (lack of time and energy) after the intervention. These findings are consistent with previous studies [52,61-63], suggesting reducing logistical obstacles alone may not be sufficient to overcome exercise barriers. This highlights the importance of addressing more individual factors, such as personal priorities and attitudes toward exercise. Overcoming these barriers likely requires coaching strategies that go beyond logistical support to include more personalized approaches tailored to each individual's needs and circumstances [64,65].

The lack of effect on self-efficacy in our study is in line with 2 previous RCTs, which reported no differences between in-person and phone-based exercise coaching [66,67]. While one cross-sectional study found an association between in-person coaching and improvement in self-efficacy compared with web-based coaching [68], this was not reflected in our results.

XSL•FO
RenderX

One potential explanation for the varying results could be the high baseline self-efficacy scores across all groups, indicating that participants who already had high self-efficacy may have had limited room for further improvement [31,37]. Nevertheless, it is important to note that all intervention arms in our study exercised significantly more than the control group after the intervention, even without a corresponding increase in self-efficacy. This suggests that improvements in exercise adherence may occur independently of self-efficacy changes. Future research should explore additional support mechanisms, such as coaching approaches grounded in behavioral theories, to further enhance exercise participation. For instance, one study that applied social cognitive theory–based coaching, which focuses on psychological factors such as self-efficacy, found that women in the coaching group had higher exercise adherence (49% versus 31%) compared with the control group, alongside improvements in self-efficacy [69].

The measurement used in our study included perceived social support from friends and family, unfortunately not accounting for potential support the participants might have experienced from their personal trainer. Including data on social support from personal trainers would have offered a more comprehensive understanding of participants' support networks, while also highlighting the trainers' role in promoting exercise adherence. In addition, it could have provided valuable insights into the dynamics of the trainer-participant relationship. However, low social support scores, both before and after the intervention, indicate a lack of sufficient encouragement from the significant others. This finding is important, as social support from family and friends is a key factor that consistently influences exercise adherence [25,26].

Our findings align with the broader literature, highlighting the complex relationship between exercise interventions and psychosocial outcomes among individuals with obesity [70,71]. For instance, Pilates, yoga, tai chi, and qigong have demonstrated potential for improving mental well-being in populations living with obesity [70,71]. However, the evidence regarding their direct effects on broader psychosocial factors remains limited. Further, a recent review showed clear evidence of combined aerobic and resistance training in improving key health indicators, including glycemic control and mental health, among individuals with obesity and type 2 diabetes [72], while our study did not demonstrate significant changes in health-related quality of life across all dimensions. The small improvement in the "change in health" subscale may reflect the cumulative benefit of structured exercise on self-perceived well-being. These results also support that regular exercise may enhance specific quality-of-life dimensions, even without observable changes in other psychosocial variables.

Finally, while this study investigated the effect of hybrid coaching models, targeted exercise regimens, such as high-intensity interval training, are shown to have the potential to improve psychosocial factors, including enhanced self-perceived health [73]. Further, exercise modalities like Pilates and yoga could also complement exercise coaching to improve psychosocial factors, since these approaches offer a less demanding, more inclusive alternative, particularly for individuals facing physical or psychological limitations [70,71,74]. Future research may integrate such modalities within hybrid coaching models to evaluate their combined impact on psychosocial outcomes.

Increasing exercise participation among the population living with obesity could significantly impact public health. Since motivation, barriers, self-efficacy, and social support are key to exercise adherence [25,75], research should explore how various coaching approaches affect these factors. For instance, SDT-based coaching may help personal trainers create a more supportive and motivating environment, improving adherence [30]. Despite the lack of effect on motivation, the intervention's potential for positive impact lies in highlighting the importance of SDT-aligned coaching strategies, which could be key to improving motivation in future interventions targeting populations living with obesity. By integrating behavioral theories into coaching interventions, researchers may better identify strategies to improve motivation, reduce barriers, and strengthen self-efficacy, ultimately promoting sustainable exercise habits and improved health outcomes [76]. While the intervention did not reduce self-reported barriers, it minimized external challenges, highlighting the value of interventions that address personal barriers. Last, although in-person and web-based coaching showed similar effects, the potential cost-effectiveness and accessibility of web-based coaching make hybrid or fully web-based approaches appealing alternatives for supporting exercise adherence among populations living with obesity. This is especially relevant given that logistical and financial barriers often limit consistent participation in in-person coaching [43,44,46]. While our RCT did not demonstrate differences in psychosocial outcomes, it shed light on the feasibility and value of hybrid coaching models to support exercise adherence in obese populations.

Our study is one of the first global pragmatic RCTs of its kind, with a long intervention period and a unique catchment area, including a nationwide intervention spanning both urban and rural regions of Norway. The inclusion of 23 experienced personal trainers, each with at least 3 years of experience, further enhances the robustness and generalizability of our findings in a real-world setting. Another strength is the hybrid approach, combining web-based coaching with in-person sessions. This approach reflects the contemporary trends in exercise supervision and meets the evolving needs of participants in modern society. Our study also contributes to the United Nations' Sustainable Development Goal 5 by focusing on obese women and addressing psychosocial factors through in-person and web-based exercise coaching, thereby supporting efforts to improve health equity and empower women. Finally, we used a range of validated questionnaires to measure the psychosocial outcomes, including motivation (BREQ-2), self-efficacy (Self-Efficacy Survey), social support, and health-related quality of life (SF-36), providing a comprehensive assessment of the psychosocial effects of the intervention.

Limitations

Some study limitations should be considered when interpreting our results. We unfortunately had a notable dropout rate, with only 64% (120/188) of participants completing both the baseline and postintervention questionnaires. However, no significant

XSL•FO
RenderX

differences were found between those who completed the study and those who dropped out in terms of baseline data. Further, the 20-week intervention period may be too short to fully assess the intervention's effects, particularly on psychosocial factors such as self-efficacy, motivation, and health-related quality of life, which often require longer durations to reveal changes or differences. Finally, while this study was conducted in Norway, limiting direct global generalizability, culture and societal similarities between Norway and other Western countries suggests that our findings could apply to similar populations and contexts. Further, the lack of a priori power calculation is another limitation of this study. The sample size was determined based on practical considerations rather than statistical power affecting the robustness and generalizability of our findings, as the study may be underpowered to identify small effects. Last, the Shapiro-Wilk test was used to evaluate whether our data met the assumption of normality. While this test is widely recognized for assessing normality, it is sensitive to sample size and outliers, which may limit its ability to fully characterize data distributions. In our study, as many as 120 participants completed both baseline and postintervention questionnaires. Thus, we believe the results of the Shapiro-Wilk test align with the assumption of normality.

Conclusions

This study indicates that both in-person coaching and 2 combinations of in-person and web-based coaching may lead to a slight improvement in self-perceived health, by higher scores in 1 of the 8 SF-36 subscales. However, these observed improvements are preliminary and may not generalize to broader health or psychosocial domains without further evidence. No significant effects were found on motivation, barriers, self-efficacy, social support, or other health domains. Further research is needed to fully understand how we can increase the ability to maintain regular exercise among women with obesity.

Acknowledgments

We would like to thank Lena Bache-Mathiesen (data scientist and researcher, Norwegian School of Sports Sciences) for conducting the randomization and Håkon Laugsand (ABEL Technologies) for assisting in the recruitment process. The Research Council of Norway and ABEL Technologies funded this study. ABEL Technologies was involved in the planning of the study design and participant recruitment.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CG wrote the original draft preparation and conducted data analysis. JMK, EMT, HC, and LAAH reviewed and edited the draft. All authors have read and approved the final version of the manuscript and agree with the order of presentation of the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1176 KB - jmir_v27i1e68462_app1.pdf]

References

1. Sarma S, Sockalingam S, Dash S. Obesity as a multisystem disease: trends in obesity rates and obesity-related complications. Diabetes Obes Metab 2021;23 Suppl 1(S1):3-16. [doi: 10.1111/dom.14290] [Medline: 33621415]
2. Obesity: health and economic consequences of an impending global challenge. World Bank Group. URL: https://tinyurl.com/4c8vv6x5 [accessed 2020-01-27]
3. Obesity and overweight. World Health Organization. URL: https://tinyurl.com/3mw8bsu3 [accessed 2023-09-19]
4. Haslam DW, James WPT. Obesity. Lancet 2005;366(9492):1197-1209. [doi: 10.1016/S0140-6736(05)67483-1] [Medline: 16198769]
5. World obesity atlas 2022. Worlds Obesity Federation. URL: https://s3-eu-west-1.amazonaws.com/wof-files/World_Obesity_Atlas_2022.pdf [accessed 2025-02-06]
6. Cooper AJ, Gupta SR, Moustafa AF, Chao AM. Sex/gender differences in obesity prevalence, comorbidities, and treatment. Curr Obes Rep 2021;10(4):458-466. [doi: 10.1007/s13679-021-00453-x] [Medline: 34599745]
7. Rubio-Almanza M, Cámara-Gómez R, Merino-Torres JF. Obesity and type 2 diabetes: also linked in therapeutic options. Endocrinol Diabetes Nutr (Engl Ed) 2019;66(3):140-149. [doi: 10.1016/j.endinu.2018.08.003] [Medline: 30337188]
8. Apovian CM, Okemah J, O'Neil PM. Body weight considerations in the management of type 2 diabetes. Adv Ther 2019;36(1):44-58. [doi: 10.1007/s12325-018-0824-8] [Medline: 30465123]

XSL•FO
RenderX

9. Bray GA, Frühbeck G, Ryan DH, Wilding JPH. Management of obesity. Lancet 2016;387(10031):1947-1956 [FREE Full text] [doi: 10.1016/S0140-6736(16)00271-3] [Medline: 26868660]
10. Powell-Wiley TM, Poirier P, Burke LE, Després JP, Gordon-Larsen P, Lavie CJ, et al. Obesity and cardiovascular disease: a scientific statement from the American heart association. Circulation 2021;143(21):e984-e1010 [FREE Full text] [doi: 10.1161/CIR.0000000000000973] [Medline: 33882682]
11. Elagizi A, Kachur S, Carbone S, Lavie CJ, Blair SN. A review of obesity, physical activity, and cardiovascular disease. Curr Obes Rep 2020;9(4):571-581. [doi: 10.1007/s13679-020-00403-z] [Medline: 32870465]
12. Avila CHA, Holloway AC, Hahn MK, Morrison KM, Restivo M, Anglin R, et al. An overview of links between obesity and mental health. Curr Obes Rep 2015;4(3):303-310. [doi: 10.1007/s13679-015-0164-9] [Medline: 26627487]
13. Batrakoulis A, Jamurtas AZ, Metsios GS, Perivoliotis K, Liguori G, Feito Y, et al. Comparative efficacy of 5 exercise types on cardiometabolic health in overweight and obese adults: a systematic review and network meta-analysis of 81 randomized controlled trials. Circ Cardiovasc Qual Outcomes 2022;15(6):e008243. [doi: 10.1161/CIRCOUTCOMES.121.008243] [Medline: 35477256]
14. Newsome AM, Batrakoulis A, Camhi SM, McAvoy C, Sansone J, Reed R. 2025 ACSM worldwide fitness trends: future directions of the health and fitness industry. ACSMs Health Fit J 2024;28(6):11-25. [doi: 10.1249/fit.0000000000001017]
15. Batrakoulis A, Fatouros IG. Psychological adaptations to high-intensity interval training in overweight and obese adults: a topical review. Sports (Basel) 2022;10(5):64 [FREE Full text] [doi: 10.3390/sports10050064] [Medline: 35622474]
16. Carbone S, Del Buono MG, Ozemek C, Lavie CJ. Obesity, risk of diabetes and role of physical activity, exercise training and cardiorespiratory fitness. Prog Cardiovasc Dis 2019;62(4):327-333. [doi: 10.1016/j.pcad.2019.08.004] [Medline: 31442513]
17. Oppert JM, Bellicha A, van Baak MA, Battista F, Beaulieu K, Blundell JE, et al. Exercise training in the management of overweight and obesity in adults: synthesis of the evidence and recommendations from the European association for the study of obesity physical activity working group. Obes Rev 2021;22 Suppl 4(Suppl 4):e13273 [FREE Full text] [doi: 10.1111/obr.13273] [Medline: 34076949]
18. Petridou A, Siopi A, Mougios V. Exercise in the management of obesity. Metabolism 2019;92:163-169. [doi: 10.1016/j.metabol.2018.10.009] [Medline: 30385379]
19. Burgess E, Hassmén P, Pumpa KL. Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review. Clin Obes 2017;7(3):123-135. [doi: 10.1111/cob.12183] [Medline: 28296261]
20. Denison E, Vist GE, Underland V, Berg RC. Interventions aimed at increasing the level of physical activity by including organised follow-up: a systematic review of effect. BMC Fam Pract 2014;15:120 [FREE Full text] [doi: 10.1186/1471-2296-15-120] [Medline: 24934215]
21. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
22. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
23. Hansen BH, Kolle E, Steene-Johannessen J, Dalene KE, Ekelund U, Anderssen SA. Monitoring population levels of physical activity and sedentary time in Norway across the lifespan. Scand J Med Sci Sports 2019;29(1):105-112. [doi: 10.1111/sms.13314] [Medline: 30276928]
24. Loyen A, Clarke-Cornwell AM, Anderssen SA, Hagströmer M, Sardinha LB, Sundquist K, et al. Sedentary time and physical activity surveillance through accelerometer pooling in four European countries. Sports Med 2017;47(7):1421-1435 [FREE Full text] [doi: 10.1007/s40279-016-0658-y] [Medline: 27943147]
25. Scarapicchia TMF, Amireault S, Faulkner G, Sabiston CM. Social support and physical activity participation among healthy adults: a systematic review of prospective studies. Int Rev Sport Exerc Psychol 2016;10(1):50-83. [doi: 10.1080/1750984x.2016.1183222]
26. Ayotte BJ, Margrett JA, Hicks-Patrick J. Physical activity in middle-aged and young-old adults: the roles of self-efficacy, barriers, outcome expectancies, self-regulatory behaviors and social support. J Health Psychol 2010;15(2):173-185. [doi: 10.1177/1359105309342283] [Medline: 20207661]
27. Bize R, Johnson JA, Plotnikoff RC. Physical activity level and health-related quality of life in the general adult population: a systematic review. Prev Med 2007;45(6):401-415. [doi: 10.1016/j.ypmed.2007.07.017] [Medline: 17707498]
28. Rodrigues F, Bento T, Cid L, Pereira Neiva H, Teixeira D, Moutão J, et al. Can interpersonal behavior influence the persistence and adherence to physical exercise practice in adults? a systematic review. Front Psychol 2018;9:2141 [FREE Full text] [doi: 10.3389/fpsyg.2018.02141] [Medline: 30459690]
29. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
30. Hancox JE, Quested E, Ntoumanis N, Thøgersen-Ntoumani C. Putting self-determination theory into practice: application of adaptive motivational principles in the exercise domain. Qual Res Sport Exerc Health 2017;10(1):75-91. [doi: 10.1080/2159676x.2017.1354059]

XSL•FO
RenderX

31. Pekmezi D, Jennings E, Marcus BH. Evaluating and enhancing self-efficacy for physical activity. ACSMs Health Fit J 2009;13(2):16-21 [FREE Full text] [doi: 10.1249/FIT.0b013e3181996571] [Medline: 29910597]
32. Zhou S, Davison KF, Qin F, Lin K, Chow BC, Zhao JX. The roles of exercise professionals in the health care system: a comparison between Australia and China. J Exerc Sci Fit 2019;17(3):81-90 [FREE Full text] [doi: 10.1016/j.jesf.2019.04.001] [Medline: 31193075]
33. McClaran SR. The effectiveness of personal training on changing attitudes towards physical activity. J Sport Sci Med 2003;2(1):4-10. [doi: 10.1097/00005768-200105001-01188]
34. Fischer DV, Bryant J. Effect of certified personal trainer services on stage of exercise behavior and exercise mediators in female college students. J Am Coll Health 2008;56(4):369-376. [doi: 10.3200/JACH.56.44.369-376] [Medline: 18316279]
35. Rustaden AM, Haakstad LAH, Paulsen G, Bø K. Effects of BodyPump and resistance training with and without a personal trainer on muscle strength and body composition in overweight and obese women-a randomised controlled trial. Obes Res Clin Pract 2017;11(6):728-739. [doi: 10.1016/j.orcp.2017.03.003] [Medline: 28392264]
36. Jeffery RW, Wing RR, Thorson C, Burton LR. Use of personal trainers and financial incentives to increase exercise in a behavioral weight-loss program. J Consult Clin Psychol 1998;66(5):777-783. [doi: 10.1037//0022-006x.66.5.777] [Medline: 9803696]
37. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
38. Molloy GJ, Dixon D, Hamer M, Sniehotta FF. Social support and regular physical activity: does planning mediate this link? British J Health Psychol 2010;15(4):859-870. [doi: 10.1348/135910710x490406]
39. Oliveira AJ, Lopes CS, Rostila M, Werneck GL, Griep RH, Leon ACMPD, et al. Gender differences in social support and leisure-time physical activity. Rev Saude Publica 2014;48(4):602-612 [FREE Full text] [doi: 10.1590/s0034-8910.2014048005183] [Medline: 25210819]
40. Heo M, Allison DB, Faith MS, Zhu S, Fontaine KR. Obesity and quality of life: mediating effects of pain and comorbidities. Obes Res 2003 Feb;11(2):209-216 [FREE Full text] [doi: 10.1038/oby.2003.33] [Medline: 12582216]
41. Lubans DR, Plotnikoff RC, Jung M, Eves N, Sigal R. Testing mediator variables in a resistance training intervention for obese adults with type 2 diabetes. Psychol Health 2012;27(12):1388-1404. [doi: 10.1080/08870446.2011.613471] [Medline: 21895462]
42. Heiestad H, Rustaden AM, Bø K, Haakstad LA. Effect of regular resistance training on motivation, self-perceived health, and quality of life in previously inactive overweight women: a randomized, controlled trial. Biomed Res Int 2016;2016:3815976 [FREE Full text] [doi: 10.1155/2016/3815976] [Medline: 27462608]
43. Stine JG, Soriano C, Schreibman I, Rivas G, Hummer B, Yoo E, et al. Breaking down barriers to physical activity in patients with nonalcoholic fatty liver disease. Dig Dis Sci 2021;66(10):3604-3611 [FREE Full text] [doi: 10.1007/s10620-020-06673-w] [Medline: 33098023]
44. Abshire DA, Gibbs S, McManus C, Caldwell T, Cox DA. Interest, resources, and preferences for weight loss programs among primary care patients with obesity. Patient Educ Couns 2020;103(9):1846-1849 [FREE Full text] [doi: 10.1016/j.pec.2020.04.003] [Medline: 32331826]
45. de Heer H, Kinslow B, Lane T, Tuckman R, Warren M. Only 1 in 10 patients told to lose weight seek help from a health professional: a nationally representative sample. Am J Health Promot 2019;33(7):1049-1052 [FREE Full text] [doi: 10.1177/0890117119839904] [Medline: 30922060]
46. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018;20(4):e144 [FREE Full text] [doi: 10.2196/jmir.8480] [Medline: 29643048]
47. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
48. Dharmadhikari S. Personal trainers market report 2024 (global edition). Cognitive Market Research. URL: https://www.cognitivemarketresearch.com/personal-trainers-market-report [accessed 2024-12-14]
49. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exercise Psychol 2004;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
50. Murcia JAM, Gimeno EC, Camacho AM. Measuring self-determination motivation in a physical fitness setting: validation of the behavioral regulation in exercise questionnaire-2 (BREQ-2) in a spanish sample. J Sports Med Phys Fitness 2007;47(3):366-374. [Medline: 17641607]
51. Sørensen M, Gill DL. Perceived barriers to physical activity across Norwegian adult age groups, gender and stages of change. Scand J Med Sci Sports 2008;18(5):651-663. [doi: 10.1111/j.1600-0838.2007.00686.x] [Medline: 18086263]
52. Gjestvang C, Abrahamsen F, Stensrud T, Haakstad LAH. Motives and barriers to initiation and sustained exercise adherence in a fitness club setting-a one-year follow-up study. Scand J Med Sci Sports 2020;30(9):1796-1805 [FREE Full text] [doi: 10.1111/sms.13736] [Medline: 32488898]
53. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987;16(6):825-836. [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]

XSL•FO
RenderX

54. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987;16(6):825-836 [FREE Full text] [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]
55. Ware JE, Sherbourne CD. The MOS 36-ltem short-form health survey (SF-36). Medical Care 1992;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
56. Garratt AM, Stavem K. Measurement properties and normative data for the norwegian SF-36: results from a general population survey. Health Qual Life Outcomes 2017;15(1):51 [FREE Full text] [doi: 10.1186/s12955-017-0625-9] [Medline: 28292292]
57. Bherer L, Erickson KI, Liu-Ambrose T. A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. J Aging Res 2013;2013:657508 [FREE Full text] [doi: 10.1155/2013/657508] [Medline: 24102028]
58. Heiestad H, Gjestvang C, Haakstad LAH. Investigating self-perceived health and quality of life: a longitudinal prospective study among beginner recreational exercisers in a fitness club setting. BMJ Open 2020;10(6):e036250 [FREE Full text] [doi: 10.1136/bmjopen-2019-036250] [Medline: 32513890]
59. Deci EL, Ryan RM. Motivation, personality, and development within embedded social contexts: an overview of self-determination theory. In: Ryan RM, editor. The Oxford Handbook of Human Motivation. New York, NY: Oxford University Press; 2012.
60. Jiménez-Loaisa A, González-Cutre D, Beltrán-Carrillo VJ, Alcaraz-Ibáñez M. Changes in bariatric patients' physical activity levels and health-related quality of life following a postoperative motivational physical activity intervention. Obes Surg 2020;30(6):2302-2312. [doi: 10.1007/s11695-020-04489-1] [Medline: 32172362]
61. Riseth L, Nøst TH, Nilsen TIL, Steinsbekk A. Long-term members' use of fitness centers: a qualitative study. BMC Sports Sci Med Rehabil 2019;11:2 [FREE Full text] [doi: 10.1186/s13102-019-0114-z] [Medline: 30828457]
62. Stankowski CL, Trauntvein NE, Hall SL. I use the student recreation center, but i would use it more if …: understanding male and female constraints to student recreation center use. Recreat Sports J 2017;41(1):55-66. [doi: 10.1123/rsj.2015-0026]
63. Herazo-Beltrán Y, Pinillos Y, Vidarte J, Crissien E, Suarez D, García R. Predictors of perceived barriers to physical activity in the general adult population: a cross-sectional study. Braz J Phys Ther 2017;21(1):44-50 [FREE Full text] [doi: 10.1016/j.bjpt.2016.04.003] [Medline: 28442074]
64. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
65. Ntoumanis N, Thøgersen-Ntoumani C, Quested E, Hancox J. The effects of training group exercise class instructors to adopt a motivationally adaptive communication style. Scand J Med Sci Sports 2017;27(9):1026-1034. [doi: 10.1111/sms.12713] [Medline: 27283879]
66. Giacobbi JRP, Zautra NT, Dreisbach KA, Liguori KR. Exercise for overweight and obese women: a multimodal pilot intervention comparing inperson with phone-based delivery of guided imagery. Int J Sport Exerc Psychol 2018;16(4):452-463 [FREE Full text] [doi: 10.1080/1612197X.2016.1256338]
67. Suminski RR, Leonard T, Obrusnikova I, Kelly K. The impact of health coaching on weight and physical activity in obese adults: a randomized control trial. Am J Lifestyle Med 2024;18(2):233-242 [FREE Full text] [doi: 10.1177/15598276221114047] [Medline: 38559788]
68. Makanjuola A, Lynch M, Hartfiel N, Cuthbert A, Wheeler HT, Edwards RT. Social return on investment of face-to-face versus online lifestyle coaching to improve mental wellbeing. Lancet 2022;400:S59. [doi: 10.1016/s0140-6736(22)02269-3]
69. Annesi JJ, Unruh JL, Marti NC, Gorjala S, Tennant G. Effects of the coach approach intervention on adherence to exercise in obese women: assessing mediation of social cognitive theory factors. Res Q Exerc Sport 2011;82(1):99-108. [doi: 10.1080/02701367.2011.10599726] [Medline: 21462690]
70. Batrakoulis A. Psychophysiological adaptations to yoga practice in overweight and obese individuals: a topical review. Diseases 2022;10(4):107 [FREE Full text] [doi: 10.3390/diseases10040107] [Medline: 36412601]
71. Batrakoulis A. Psychophysiological adaptations to pilates training in overweight and obese individuals: a topical review. Diseases 2022;10(4):71 [FREE Full text] [doi: 10.3390/diseases10040071] [Medline: 36278570]
72. Al-Mhanna SB, Batrakoulis A, Wan Ghazali WS, Mohamed M, Aldayel A, Alhussain MH, et al. Effects of combined aerobic and resistance training on glycemic control, blood pressure, inflammation, cardiorespiratory fitness and quality of life in patients with type 2 diabetes and overweight/obesity: a systematic review and meta-analysis. PeerJ 2024;12:e17525 [FREE Full text] [doi: 10.7717/peerj.17525] [Medline: 38887616]
73. Batrakoulis A, Loules G, Georgakouli K, Tsimeas P, Draganidis D, Chatzinikolaou A, et al. High-intensity interval neuromuscular training promotes exercise behavioral regulation, adherence and weight loss in inactive obese women. Eur J Sport Sci 2020;20(6):783-792 [FREE Full text] [doi: 10.1080/17461391.2019.1663270] [Medline: 31478436]
74. Batrakoulis A. Role of mind–body fitness in obesity. Diseases 2022;11(1):1. [doi: 10.3390/diseases11010001]
75. Choi J, Lee M, Lee JK, Kang D, Choi JY. Correlates associated with participation in physical activity among adults: a systematic review of reviews and update. BMC Public Health 2017;17(1):356 [FREE Full text] [doi: 10.1186/s12889-017-4255-2] [Medline: 28438146]
76. Ntoumanis N, Ng JYY, Prestwich A, Quested E, Hancox JE, Thøgersen-Ntoumani C, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior,

XSL•FO
RenderX

physical, and psychological health. Health Psychol Rev 2021;15(2):214-244 [FREE Full text] [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]

Abbreviations

BREQ-2: Behavioral Regulation in Exercise Questionnaire-2
IP25: in-person exercise coaching 25%
IP50: in-person exercise coaching 50%
IP100: in-person exercise coaching 100%
RCT: randomized controlled trial
SDT: self-determination theory
SF-36: 36-Item Short Form Health Survey

Edited by A Mavragani; submitted 06.11.24; peer-reviewed by D Li, N Misra, A Batrakoulis, DB Foroutan, M Gasmi ; comments to author 29.11.24; revised version received 19.12.24; accepted 10.01.25; published 25.02.25.

Please cite as:
Gjestvang C, Kalhovde JM, Mauseth Tangen E, Clemm H, Haakstad LAH
Impact of In-Person and Mobile Exercise Coaching on Psychosocial Factors Affecting Exercise Adherence in Inactive Women With Obesity: 20-Week Randomized Controlled Trial
J Med Internet Res 2025;27:e68462
URL: https://www.jmir.org/2025/1/e68462
doi:10.2196/68462
PMID:

©Christina Gjestvang, John Magne Kalhovde, Elene Mauseth Tangen, Hege Clemm, Lene Annette Hagen Haakstad. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Web-Based Application for Reducing Methamphetamine Use Among Aboriginal and Torres Strait Islander People: Randomized Waitlist Controlled Trial

Rachel Reilly[1,2], BA, MClincSci, DPsych; Rebecca McKetin[3], BSci, PhD; Federica Barzi[4], PhD; Tayla Degan[3,5], MPsych, PhD; Nadine Ezard[3,6], BA, MPH, MBBS, PhD; Katherine Conigrave[7], MBBS, PhD; Julia Butt[8], BSci, PhD; Yvette Roe[9], PhD; Handan Wand[10], MSci, PhD; Brendan Quinn[11], BA, PhD; Wade Longbottom[12]; Carla Treloar[13], BSci, PhD; Adrian Dunlop[14], MBBS, PhD; James Ward[4], PhD

[1]Wardliparingga Aboriginal Health Equity, South Australian Health and Medical Research Institute, Adelaide, Australia
[2]School of Psychology, University of Adelaide, Adelaide, Australia
[3]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia
[4]Poche Centre for Indigenous Health, University of Queensland, Brisbane, Australia
[5]School of Psychology, Faculty of Arts, University of Wollongong, Wollongong, Australia
[6]St Vincent's Hospital, Sydney, Australia
[7]Central Clinical School, Sydney Medical School, University of Sydney, Sydney, Australia
[8]Discipline of Psychology and Social Sciences, School of Arts and Humanities, Edith Cowan University, Perth, Australia
[9]Molly Wardaguga Institute for First Nations Birth Rights, Charles Darwin University, Casuarina, Northern Territory, Australia
[10]The Kirby Institute, University of New South Wales, Sydney, Australia
[11]Burnet Institute, Melbourne, Australia
[12]South Coast Aboriginal Medical Service Corporation, Nowra, Australia
[13]Centre for Social Research in Health, University of New South Wales, Sydney, Australia
[14]School of Medicine and Public Health, University of Newcastle, Newcastle, Australia

Corresponding Author:
Rachel Reilly, BA, MClincSci, DPsych
School of Psychology
University of Adelaide
North Terrace
Adelaide, 5005
Australia
Phone: 61 881284216
Email: rachel.reilly@sahmri.com

Abstract

Background: Digital interventions can help to overcome barriers to care, including stigma, geographical distance, and a lack of culturally appropriate treatment options. “We Can Do This” is a web-based app that was designed with input from cultural advisors and end users to support Aboriginal and Torres Strait Islander people seeking to stop or reduce their use of methamphetamine and increase psychosocial well-being.

Objective: This study aimed to evaluate the effectiveness of the “We Can Do This” web-based app as a psychosocial treatment for Aboriginal and Torres Strait Islander people who use methamphetamine.

Methods: The web app was evaluated using a randomized waitlist controlled parallel group trial. Participants were Aboriginal and Torres Strait Islander people aged 16 years or older who self-identified as having used methamphetamine at least weekly for the past 3 months. Participants were randomized on a 1:1 ratio to receive either access to the web-based app for 6 weeks or a waitlist control group. Both groups received access to a website with harm minimization information. The primary outcome was days of methamphetamine use in the past 4 weeks assessed at 1, 2, and 3 months post randomization. Secondary outcomes included severity of methamphetamine dependence (Severity of Dependence Scale [SDS]), psychological distress (Kessler 10 [K10]), help-seeking behavior, and days spent out of role due to methamphetamine use.

XSL•FO
RenderX

Results: Participants (N=210) were randomized to receive either access to the web-based app (n=115) or the waitlist control condition (n=95). Follow-up was 63% at 1 month, 57% at 2 months, and 54% at 3 months. There were no significant group differences in days of methamphetamine use in the past 4 weeks at 1 the month (mean difference 0.2 days, 95% CI –1.5 to –2), 2 months (mean difference 0.6 days, 95% CI –1 to 2.4 days) or 3 months (mean difference 1.4 days, 95% CI –0.3 to 3.3 days) follow-up. There were no significant group differences in K10 scores, SDS scores, days out of role, or help-seeking at any of the 3 follow-up timepoints. There was poor adherence to the web-based app, only 20% of participants in the intervention group returned to the web-based app after their initial log-in. Participants cited personal issues and forgetting about the web-based app as the most common reasons for nonadherence.

Conclusions: We found poor engagement with this web-based app. The web-based app had no significant effects on methamphetamine use or psychosocial well-being. Poor adherence and low follow-up hindered our ability to accurately evaluate the effectiveness of the web-based app. Future web-based apps for this population need to consider methods to increase participant engagement.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619000134123p; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=376088

International Registered Report Identifier (IRRID): RR2-10.2196/14084

(J Med Internet Res 2025;27:e58341) doi:10.2196/58341

KEYWORDS

methamphetamine; Aboriginal and Torres Strait Islander Health; web-based intervention; randomised controlled trial; therapeutic program; methamphetamine use; substance use; digital interventions; treatment; psychosocial wellbeing; effectiveness; app; psychosocial distress; mobile phone

Introduction

Methamphetamine use is an issue of deep concern in Australia, including in Aboriginal and Torres Strait Islander communities [1-4]. Although rates vary geographically and Aboriginal and Torres Strait Islander people tend to be underrepresented in national population surveys, the available data suggests that nationally, the rate of methamphetamine use is at least two times higher amongst Aboriginal and Torres Strait Islander people compared with the general population (2.9% vs 1.3%) [5,6]. Regular and frequent methamphetamine use is associated with long-term physical, psychological, and social harm [7,8]. Previous research has found that while both Aboriginal and Torres Strait Islander and non-Indigenous people who have recently used methamphetamine experienced harms such as an excess burden of psychological distress, exposure to violence, and economic exclusion [9], Aboriginal and Torres Strait Islander people who have recently used methamphetamine reported significantly higher rates of racial discrimination, excessive grief, perceived social isolation, and complex trauma than their non-Indigenous counterparts [1,9,10]. Key protective factors for Aboriginal and Torres Strait Islander participants included connection to culture and contact with elders [1].

These differences in the social processes leading to use, the contexts in which people use, and the community and cultural strengths that may assist recovery points to a need for culturally specific resources for Aboriginal and Torres Strait Islander people seeking to reduce or stop using methamphetamine [9]. In a context where demand for specialist alcohol and other drug (AOD) treatment across urban, rural, and remote settings in Australia is not met by available services, much of the burden for providing care to people who use methamphetamine falls to primary health care services [11]. The Novel Interventions to Address Methamphetamine in Aboriginal and Torres Strait Islander Communities (NIMAC) project was developed in response to calls from Aboriginal Community Controlled Health Services (ACCHS) and the community at large to find ways to better support ACCHS to reduce the harms associated with methamphetamine use [1,4,9].

Web-based interventions have emerged as a popular, feasible, and cost-effective way to provide treatment for substance use disorders. Evidence suggests that web-based interventions can be effective in the treatment of substance use disorders [12-14]. However, effectiveness can vary according to intervention type, characteristics of both the web-based app and end users, and substances targeted [12-15]. Although research in this space is ongoing, to date, there are limited interventions tailored to the treatment of methamphetamine use [12], and no known interventions targeting methamphetamine use for Aboriginal and Torres Strait Islander communities.

A scoping review of evidence on the uptake of web-based interventions in Indigenous communities internationally, found that such interventions can reduce barriers stemming from geographical distance, access to specialists, or stigma; they can support health practitioners provide better care to clients for a range of health conditions, but they need to be developed in partnership with communities [16]. There is evidence that web-based and digital interventions are acceptable and feasible in Aboriginal and Torres Strait Islander communities, with several such interventions developed for AOD assessment [17,18], psychosocial assessment and care planning [19], and suicide prevention [20]. Given the lack of specialist AOD services and the stigma associated with methamphetamine use, which is known to prevent help-seeking [4], a web-based intervention was considered a possible way to improve access to an evidence-based therapeutic intervention. Furthermore, in a survey conducted within the NIMAC project of 734 Aboriginal and Torres Strait Islander and non-Indigenous people who had used methamphetamine in the previous 12 months, we found

XSL•FO
RenderX

that more than 90% of respondents had a mobile phone, and of these, 80% accessed the internet on their phone, which provided some initial indication that an online intervention may be feasible for this population [9].

The objective of this research was to test the effectiveness of the "We Can Do This" web-based app. The details of the web-based app have been reported elsewhere [21]. In brief, "We Can Do This" comprised 7 modules that were codesigned with clinicians, representatives of Aboriginal Community Controlled Health Services, and people with experience of use. Psychoeducational material, interactive exercises, planning, and goal-setting exercises were drawn from cognitive behavioral therapy (CBT), motivational interviewing, acceptance commitment therapy, and narrative approaches. Content was presented with films based on real narratives developed from qualitative interviews, and cultural content provided by advisors from diverse Aboriginal nations [21].

Our primary hypothesis was that participants who used the web-based app regularly for 6 weeks would have significantly reduced days of methamphetamine use in the last 4 weeks at 1, 2, and 3 months post intervention, compared with waitlisted participants with access to treatment as usual, including information about harm-minimization and where to access support. We also hypothesized that changes in secondary outcomes at 3-month follow-up would include increased help-seeking, decreased psychosocial distress, and fewer days out of role.

Methods

Trial Design

The web app was evaluated in a waitlist randomized controlled trial conducted in partnership with 10 urban and regional ACCHS across 6 Australian states and territories including South Australia (n=2), New South Wales (n=3), Queensland (n=1), Northern Territory (n=1), Western Australia (n=1), and Victoria (n=2). The protocol for this RCT study is already published [21]. The trial protocol was registered with the Australian and New Zealand Clinical Trials Registry (ANZCTRN12619000134123) on April 10, 2019 [21]. Participants were randomly allocated to the intervention or waitlist group with a 1:1 ratio. Random numbers were generated using the statistical software Stata (version 14.0; StataCorp) by a statistician not otherwise involved in the trial.

Participants

Participants were eligible to participate if they were Aboriginal and Torres Strait Islander, aged 16 years or older, and self-identifying that they had been using methamphetamine about weekly for the previous 3 months. Eligibility was confirmed by participant self-report on entry to the web-based app platform. Participants were initially recruited through advertisements in health services and on social media. All participants were invited to provide informed consent through a web-based form before completing any study assessments.

Initially, the web app was designed to facilitate direct referral through the app to weekly support from health practitioners at partner health services. To enable wider recruitment, this function was modified to enable participants to self-refer to a local service regardless of geographical location within Australia. In consultation with the trial management committee, additional changes made posttrial commencement to aid recruitment were (1) providing a voucher for reimbursement for completion of the baseline survey (Aus $25), (2) increasing the value of voucher reimbursement for the first follow-up survey (from Aus $20 to Aus $25), (3) introducing peer recruitment, where peers were offered Aus $20 for every participant referred to the project, (4) reducing the number of questions in the assessments to reduce drop out before reaching randomization, and (5) extending the recruitment period. During this period the average conversion rate was approximately Aus $1=US $0.7.

Procedures

Advertisements directed participants to a web landing page that included information about the project in both written and audio-visual format. After providing consent, all participants completed a survey at baseline before randomization. After participants completed the baseline survey, they were randomized to either the intervention group or a waitlist control group according to a 1:1 ratio. The randomization schedule was developed by a statistician who was not involved with study participants (HW). Random numbers were generated using the statistical software Stata. All participants were followed up at 1, 2, and 3 months post baseline. Assessments were completed through self-reporting online, with an option to have a researcher call to complete the assessment over the phone if this was considered easier. Attempts were made to contact all participants by email, SMS text messages, and phone for follow-up assessments if they failed to respond to follow-up assessments within 2 weeks. After 120 days post baseline, participants were considered lost to follow-up.

Intervention

The intervention group received access to the "We Can Do This" web app for 6 weeks. Participants could access the 7 modules on smartphones, tablets, or computers, and complete them in a self-paced manner. There was no restriction on the order in which participants could complete the modules. To commence the intervention, participants set up an account in the web app. During the 6 weeks of access to the intervention, they received weekly reminders to log in by email, with encouragement to log in for around half an hour per week. If the system identified that a participant had not logged in for 7 days, this triggered a text message with a prompt to log in and a link to the web app [21].

Control Group

Participants allocated to the waitlist group were directed to the study website, with links to harm-minimization information and options for self-referral to AOD counseling. Participants in this group were sent a link with access to the web app immediately following the completion of their final follow-up survey. The waitlist control group then had access to the web app for 6 weeks.

XSL•FO
RenderX

Measures

Primary Outcome

The primary outcome was the number of days the participant used methamphetamine in the past 4 weeks, assessed at baseline and 1, 2, and 3 months post baseline using the Australian Treatment Outcome Profile (ATOP [22]).

Secondary Outcomes

Secondary outcomes were the help-seeking intention, assessed using the General Help-Seeking Questionnaire [23], readiness to change (Readiness to Change Questionnaire [24]), psychological distress (Kessler 10 [25]), severity of dependence on methamphetamine in the past month (Severity of Dependence Scale (SDS) [26]), and other substance use during the past month, including cannabis, alcohol, tobacco, benzodiazepines, heroin, other opioids, cocaine, ecstasy, and other drugs (ATOP [22]), and days out of role during the past month, resulting from methamphetamine use [27].

Other Measures

Usability and acceptability of the web-based app were measured using the Internet Intervention Adherence Questionnaire [28], which includes questions about barriers and facilitators to using the program and optional free-text to report any criticisms or other feedback.

The overall feasibility of the web-based intervention was assessed by observing the uptake and use of the web-based intervention (number of modules accessed). The number of visits to the web-based app overall and per module was monitored using Google Analytics. The Internet Intervention Adherence Questionnaire [28] was included in the second follow-up questionnaire for the intervention group.

Statistical Methods

Based on our previous research, we expected a baseline mean of 16 days of methamphetamine use in the past 4 weeks [29,30]. Our original sample size of 288 participants was intended to detect a reduction from a mean of 16 (SD 14.5) days of use in the past 4 weeks to a mean of 8 use days in the past 4 weeks at each of the follow-up points, with 90% power at $P<.01$, allowing for 30% attrition (ie, 100 per group at follow-up). Our final sample included 210 participants, and we observed a 40% attrition. With this sample, we were able to detect a reduction from 16 days of use in the past 4 weeks to a mean of 6 use days in the past 4 weeks at each of the follow-up points, with 90% power at $P<.01$.

Data Analysis

An intention-to-treat analysis was used to compare the primary and secondary outcomes for the intervention versus the waitlist, 2 and 3 months from baseline. Descriptive statistics of baseline characteristics were provided by the randomization group. Continuous variables were summarized with mean and SD when appropriate, or median and IQR when not normally distributed, and categorical variables were summarized with frequency and percentage.

Multilevel mixed-effects generalized linear model, with either a Gaussian, Poisson, or robust Poisson distribution as appropriate for each specific outcome, were used to estimate change between (1) baseline and 1-month assessment, (2) baseline and 2-month assessment, and (3) 3-month assessment. To compare the changes described above between study arms, the models included a categorical variable for the study arms, a categorical variable for each time point, and an interaction term between arm and time. The use of mixed models has the advantage of allowing the modeling of the correlation between repeats of outcomes recorded within the same participant. A per-protocol analysis was carried out to assess the effect of the intervention for those participants who accessed the app at least once versus those participants who never accessed the app. Sensitivity analyses were also carried out that included baseline outcomes in the models when descriptive analyses showed important between intervention arm differences for baseline scores.

Statistical significance was set at $P<.01$. Statistical analysis was conducted using Stata (version 18) statistical software.

Ethical Considerations

The trial was approved by ethics committees in all relevant jurisdictions: Aboriginal Health Research Ethics Committee of South Australia (04-19-810), Aboriginal Health and Medical Research Council of New South Wales (1484/19), University of Queensland (2020001710), Menzies School of Health Research (2019-3330), Central Australian Human Research Ethics Committee (CA-20-3779), Western Australian Aboriginal Health Ethics Committee (1002), St Vincent's Hospital Melbourne Human Research Ethics Committee (52756: 149/20), and University of Tasmania (23229).

Results

Recruitment continued for a total of 22 months from December 2019 to October 2022 with the final follow-up completed in February 2022. Out of the 210 participants who completed the baseline survey, 95 were randomized into the waitlist control group and 115 into the intervention group. Participant flow and attrition are shown in Figure 1 (Multimedia Appendix 1 shows the CONSORT [Consolidated Standards of Reporting Trials] checklist).

XSL•FO
RenderX

Figure 1. CONSORT flow diagram: Numbers of participants who self-assessed for eligibility, completed online consent and baseline assessment, were randomized to intervention or wait-list control, and retained for follow-up at 1, 2, and 3 months.

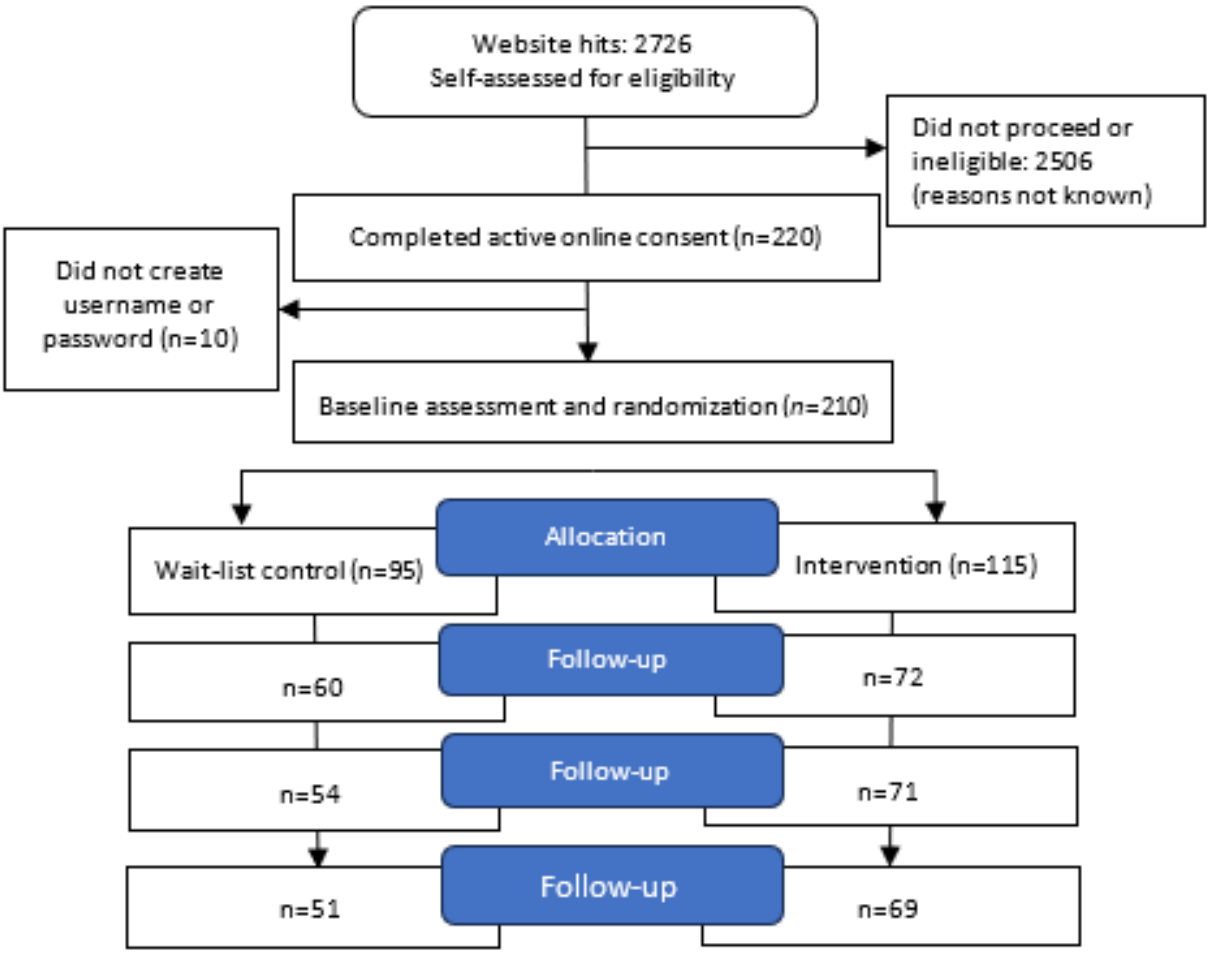

Baseline participant characteristics are shown in Table 1. In both groups, close to 90% were Aboriginal, with the remainder either Aboriginal and Torres Strait Islander or Torres Strait Islander. Participants' mean age was 33.9 (SD 8.7) years. They had been using methamphetamine from a median age of 21 (IQR 18-28) years and had used methamphetamine on a median of 20 (IQR 13-25) days in the past 4 weeks. A total of 82 % smoked tobacco (92% of them, daily). Other substance use consisted mainly of cannabis (65%) and alcohol (60%), with cannabis being used more often than alcohol (median 13, IQR 0-28 days for cannabis vs 11, IQR 1-28 days for alcohol). Other drug use was less common (Table 1).

There were more men in the intervention group (64%) compared with the control group (43%), but the 2 groups were broadly equivalent on other sociodemographic variables and drug use (Table 1).

XSL•FO
RenderX

Table 1. Summary of baseline socioeconomic factors and drug use by study arm and overall.

	Intervention (n=115)	Waitlist (n=95)	Total (N=210)
Age (n=208), mean (SD)	34.5 (9.2)	33.2 (8.1)	33.9 (8.7)
Gender (n=209)			
Male, n (%)	49 (43)	60 (63.2)	109 (52.2)
Sexual identity (n=210)			
LGBTQI[a], n (%)	9 (7.8)	19 (20)	28 (13.3)
Aboriginality (n=206)			
Aboriginal, n (%)	100 (89.3)	85 (90.4)	185 (89.8)
Relationship status (n=210)			
Single, n (%)	70 (60.9)	55 (57.9)	125 (59.5)
Employment status (n=210)			
Working (at all), n (%)	24 (20.9)	35 (36.8)	59 (28.1)
Highest level of education (n=208)			
Completed high school, n (%)	42 (36.8)	36 (38.3)	78 (37.5)
Age at first meth use (years), median (IQR)	22 (19-29)	21 (18-27)	21 (18-28)
Days of methamphetamine use in the past 4 weeks, median (IQR)	19 (14-25)	20 (11-27)	20 (13-25)
Route of methamphetamine use (n=209), n (%)			
Injecting	53 (46.1)	44 (46.8)	97 (46.4)
Smoking	61 (53)	46 (48.9)	107 (51.2)
Other	1 (1)	4 (4.3)	5 (2.4)
Other drug use in the past 4 weeks, n (%)			
Alcohol	71 (61.7)	56 (59)	127 (60.3)
Cannabis	57 (60)	57 (60)	114 (60)
Tobacco	93 (81.6)	78 (82.1)	171 (81.8)
Cocaine	14 (12.2)	14(14.7)	28 (13.5)
Heroin	12 (10.4)	6 (6.3)	18 (8.6)
Other opioids	20 (17.5)	18 (20)	38 (18.2)
Benzodiazepines	40 (34.8)	31 (33.3)	71 (34.1)
Other drugs	13 (11.3)	6 (6.3)	19 (9)

[a]LGBTQI: lesbian, gay, bisexual, transgender, queer/questioning, or intersex.

Outcomes

Days of methamphetamine use reduced significantly in both intervention and control groups (P<.04), but there was no significant difference between the groups at 1, 2, or 3 months post baseline (Figure 2A). As shown in Table 2, there were no significant group differences in the secondary outcomes, including psychological distress (K10), the severity of dependence, help-seeking, days out of role, and other drug use.

XSL•FO
RenderX

Figure 2. A: Mean days and 95% CI of methamphetamine use at baseline (month 0) and at each of the 3 follow-up months—intervention arm versus waitlist arm. B: Mean and 95% CI days of methamphetamine use at baseline (time 0) and at each of the 3 follow-up months for the intervention arm only, for participants who logged in the "We Can Do This" web app at least once after initial sign up versus those who never logged-in the web app after initial sign up.

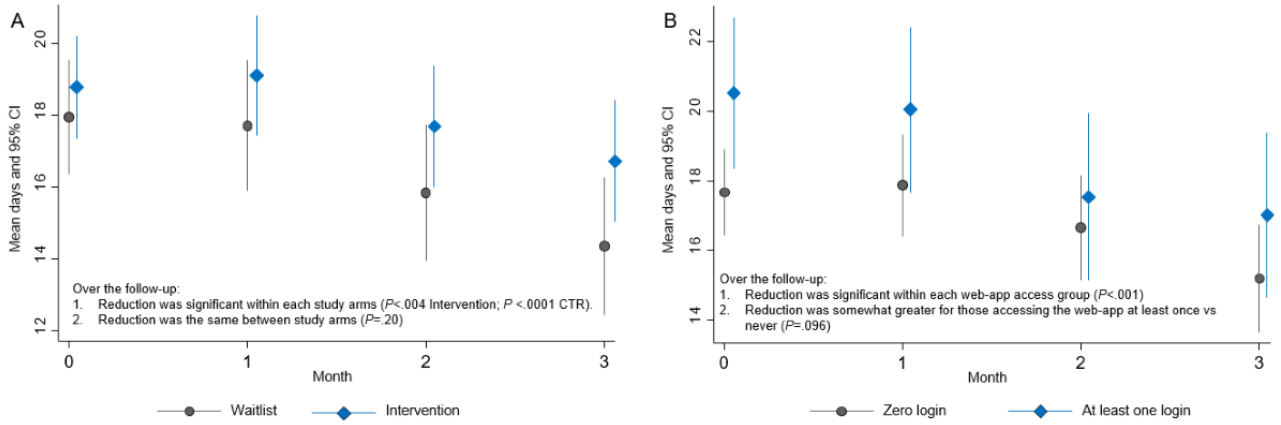

Usage data acquired through Google Analytics showed a large proportion (80%) of participants logged into the web app once and did not subsequently return. Retrospective analysis of the intervention condition only, comparing those participants who did not log into the web app again after initial sign-up ("never") versus those who logged in at least once more ("at least once"), showed a significant reduction in use amongst those who returned (Figure 2B).

Estimates were derived using multilevel mixed-effects generalized linear models, including random effects for the intercept and for time, with either a Gaussian, Poisson, or robust Poisson distribution as appropriate for each specific outcome. The treatment effect is calculated as the difference between the intervention and waitlist at 1-month assessment, 2-month assessment, and 3-month month assessment. The use of multilevel mixed effect models allows for estimates of treatment effects that are adjusted for between study arms differences at baseline.

XSL•FO
RenderX

Table 2. Mean (95% CI) and rate (95% CI) for each study outcome at baseline, month 1, month 2, and month 3 in each study arm, and corresponding treatment effect (difference of means, relative rate, and 95% CI).

	Baseline	Month 1	Month 2	Month 3
Primary outcome				
Days of methamphetamine use—last 4 weeks, mean (95% CI)				
Intervention	19.5 (17.3-21.8)	19.1 (16.7-21.4)	17.8 (15.6-20)	17.2 (15-19.3)
Waitlist	18.5 (16.4-20.8)	17.8 (15.4-20.2)	16.1 (13.9-18.3)	14.8 (12.7-16.9)
Treatment effect	—[a]	0.2 (–1.5 to 2.0)	0.6 (–1.0 to 2.4)	1.4 (–0.3 to 3.3)
Secondary outcomes				
K10 [b] score , mean (95% CI)				
Intervention	31.3 (29.6-33)	31.7 (29.5-33.8)	29.2 (27.1-31.2)	28.2 (26.2-30.3)
Waitlist	29.8 (27.9-31.6)	28.7 (26.4-31)	26.6 (24.3-28.9)	25.9 (23.6-28.2)
Treatment effect	—	1.4 (–1.9 to 4.7)	1 (–2.2 to 4.2)	0.8 (–2.4 to 4.1)
SDS [c] score , mean (95% CI)				
Intervention	8.7 (8.1-9.3)	8.4 (7.7-9.1)	8.7 (8.0-9.4)	8.5 (7.8-9.3)
Waitlist	8.5 (7.8-9.1)	8.7 (8-9.5)	7.9 (7.1-8.7)	7.4 (6.7-8.3)
Treatment effect	—	–0.5 (–1.6 to 0.5)	0.6 (–0.5 to 1.7)	0.9 (–0.2 to 2)
Days out of role rate (95% CI) of 0 to 1 day a month				
Intervention	26.3 (18.1-34.4)	26.4 (16.3-36.6)	14.1 (6.1-22.1)	16.3 (7.6-25)
Waitlist	27.3 (18.2-36.4)	29.1 (17.5-40.6)	30.2 (17.9-42.4)	31.1 (18.3-43.7)
Treatment effect, RR[d] (95% CI)	—	0.94 (0.48-1.8)	0.48 (0.22-1.07)	0.55 (0.26-1.13)
Days cut back on role rate (95% CI) of 0 to 1 day a month				
Intervention	33.3 (24.7-42)	33.3 (22.1-44.4)	29 (18.3-39.7)	25.4 (14.9-35.8)
Waitlist	27.3 (16.2-38.4)	29.1 (14.7-43.4)	30.2 (14.8-45.5)	31.1 (15.3-46.9)
Treatment effect, RR (95% CI)	—	0.94 (0.4-2.21)	0.48 (0.19-1.26)	0.55 (0.21-1.40)
Sought help (Yes), rate (95% CI)				
Intervention	21 (10.6-21.5)	22.9 (9.3-36.5)	22.4 (9.1-35.7)	28.3 (12.5-44.0)
Waitlist	20 (9.4-30.7)	8.7 (0.1-16.8)	12.9 (2.3-23.4)	20.7 (6.4-34.9)
Treatment effect, RR (95% CI)	—	2.52 (0.75,8.4)	1.66 (0.54-5.06)	1.30 (0.48-3.55)
Days of alcohol use , mean (95% CI)				
Intervention	13.1 (10.6-15.5)	11.3 (9.4-13.2)	10.1 (8.5-11.7)	8.6 (6.9-10.3)
Waitlist	13.3 (10.6-16.0)	10.5 (8.6-12.4)	8.0 (6.5-9.5)	8.0 (6.1-9.9)
Treatment effect days	—	1.0 (–1.6 to 3.3)	2.3 (–0.3 to 4.4)	0.8 (–2.7 to 3.9)
Cannabis use, rate (95% CI)				
Intervention	77.4 (12.6-98.2)	56.9 (9.1-100)	57.5 (9.2-100)	54.5 (8.7-100)
Waitlist	53.3 (8.4-98.2)	36.2 (5.5-66.8)	43.9 (6.7-81.2)	38.1 (5.8-70.4)
Treatment effect, RR (95% CI)	—	1.08 (0.93-1.25)	0.90 (0.85-1.15)	0.98 (0.85-1.14)
Tobacco use (any use), rate (95% CI)				
Intervention	81.1 (73.6-88.6)	85.8 (77.8-93.9)	85.5 (77.2-93.8)	78.6 (69-88.3)
Waitlist	81.4 (73.1-81.4)	70.8 (58.8-82.7)	75.6 (86.9-64.2)	71.2 (83.4-59.1)
Treatment effect, RR (95% CI)	—	2.1 (1.01-4.1)	1.7 (0.81-3.61)	1.37 (0.7-2.66)
Any other drug use [e] , rate (95% CI)				
Intervention	47.8 (38.6-57)	35.9 (24.8-47.0)	41.2 (29.6-52.8)	38.7 (27.1-50.3)

XSL•FO
RenderX

	Baseline	Month 1	Month 2	Month 3
Waitlist	46.3 (36.1-56.4)	48.1 (35.4-60.8)	46.9 (33.5-60.4)	43.5 (30-56.9)
Treatment effect, RR (95% CI)	—	0.73 (0.47-1.11)	0.85 (0.56-1.29)	0.86 (0.56-1.34)

[a]Not available.

[b]K10: Kessler 10.

[c]SDS: Severity of Dependence Scale.

[d]RR: risk ratio.

[e]Any use of heroin, benzodiazepine, heroin, other opioids, cocaine, ecstasy, or other drugs.

Usability and Adherence

The most common issues participants cited as obstacles to completing the web app (either a "little problem" or a "major problem") were "personal issues stopped me from using the web program" (55%) and "I just forgot to go to the web program" (51%). Problems with the internet or computer were cited by 39%-41% of participants. Issues with navigability and presentation of the web app were less often cited as barriers (20%-29%), although 41% said the program was not useful and 43% said they simply did not want to do it. Practical barriers, like other people using the computer or not having enough time to do the intervention, were cited by 44% and 42% respectively. Items that were most frequently considered major problems for using the web app (>10%) were forgetting to use it, having personal issues, believing it would take too long, not having time, and the computer not working or having problems.

The majority of respondents indicated that several matters were not problems that prevented them from using the web app, including internet connectivity, comprehension of the program, navigation, work issues getting in the way, difficulty reading the screen, or too many words. The frequency of responses to all usability and adherence questions is included in Table S1 in Multimedia Appendix 2.

Discussion

Principal Findings

We found that the web app, "We Can Do This," which targeted Aboriginal and Torres Strait Islander people, was not effective in reducing methamphetamine use. This is consistent with outcomes of a similar but nontargeted internet-based intervention for stimulant use [31,32], which did not produce significant changes in stimulant use. However, it is inconsistent with the broader evidence base for internet-based interventions, which have been found to produce significant, if modest, reductions in substance use [12,33]. These include effective CBT-based online interventions for methamphetamine use [34]. The web app did not produce expected increases in help-seeking intention, motivation to reduce methamphetamine use, and reductions in days out of role, benefits that have been found with nontargeted internet interventions for stimulant use [31,32].

The most likely explanation for the failure to yield expected benefits from "We Can Do This" is poor engagement with this self-directed web-based intervention, with only a small proportion of participants completing the web app modules. This interpretation of the results is supported by our per-protocol analysis, which found evidence of a reduction in methamphetamine use amongst participants who completed at least one module of the intervention. This could suggest that the web app has the potential to be effective for those people who return to it for two or more sessions. However, delivered in its current form, it is unlikely to result in any population shift in methamphetamine use. Keeping people engaged with health-related websites and apps is a universal problem [35,36] and it is also a specific challenge in trials of psychosocial interventions targeting stimulant use [37].

In our case, simply forgetting to use the web app was one of the main reasons cited as an obstacle to engagement. Retention and engagement strategies used in this study included SMS text messages and email reminders. However, more aggressive tactics, such as push notifications, in-app messages, and "gamification" mechanics (using rewards, competition, "unlocking" levels, or time-limited challenges) could be used to increase retention and engagement [15,38]. Contingency management can also be used to increase treatment attendance and engagement with treatment-related activities [39] and may be one option to increase engagement with the web app. Significant work would be required to update the web app to incorporate more of these features for it to be successfully offered as a stand-alone, self-directed intervention available in the community.

Other barriers to not completing the web app were mostly external (eg, personal issues, not finding time, and computer problems) rather than barriers related to the content or layout of the web app. Importantly, personal issues were a major barrier. Further investigation of the type of personal issues that impeded using the web app would be needed to understand the implications of this finding. However, it may reflect the complex and challenging lifestyle associated with having a stimulant use disorder. It may also reflect the mood disturbances commonly experienced by people with a stimulant use disorder, both related to co-occurring mental disorders, but also symptoms of withdrawal from methamphetamine. Further qualitative research and codesign of web apps are likely to be needed to understand and overcome these barriers to adherence.

There are also factors related to the target population that have been associated with poor retention and engagement. At a population level, life stress is generally higher for Aboriginal people in Australia compared with non-Indigenous people [40]; low educational attainment and a lack of income increase the risk of dropping out from alcohol and other drug treatment programs [41], and frequent substance use is associated with poor treatment retention [42,43]. Cognitive factors, although

XSL•FO
RenderX

inconsistently associated with treatment outcomes in other settings [42], may be relevant to a program, such as "We Can Do This," which requires both self-motivation and concentration. In light of this, there may be value in exploring alternative approaches to delivery, for example, clinician-assisted web apps, single brief interventions, or web apps that focus on a broader range of drug use and treatment scenarios. Qualitative research with people who have attempted to engage with the web app will provide a deeper understanding of these issues.

Limitations

The study was subject to several limitations. The major limitation of this study was poor adherence to the internet intervention, which limited our ability to detect a treatment effect in the intention-to-treat analysis. This clearly has implications for any conclusions that can be drawn about the useability of the web app in its current form. Dropout from the assessments may also have biased findings, particularly in the treatment completers analysis. Data on drug use relied on self-report and could not be verified. However, self-reporting in research studies shows high concordance with biological indicators of drug use [44]. The anonymous data collection was deemed necessary to enable participation from people who may otherwise not disclose information about their drug use, but it did limit the ability to confirm the veracity of the data. Although we carefully monitored the use of the web app to identify any instances of unusual access (such as multiple sign-ups from 1 IP address) and contacted several participants by phone to complete assessments (as per protocol), we have no way of being certain that all participants met the eligibility criteria.

Conclusion

We did not demonstrate the effectiveness of this web app in reducing methamphetamine use or increasing psychosocial well-being. Overall, engagement with the web app was poor and hindered our ability to evaluate its effectiveness. Future web apps for this population need to consider methods to increase participant engagement. Understanding the unique challenges and strengths of Aboriginal and Torres Strait Islander participants and communities, and ongoing guidance from communities and ACCHSs will be key to ongoing efforts to identify accessible and effective interventions for methamphetamine use.

Acknowledgments

We thank the Novel Interventions to Address Methamphetamine in Aboriginal and Torres Strait Islander Communities' site coordinators, study advisory group members, research participants, and health service staff, in particular Edward Wilkes, Dennis Gray, Ike Fisher, Harry Miller, Emma Donley, Angela Sloan, Kirsten Button, Peter Fullston, Warren Clements, Paul Briggs, Bree Firebrace, Shani Crumpen, Virginia Lovell, Jack Nagle, Haley Briggs, Daley Hammersley, and Tahlia McDonald; for their creative development. Blis Web Agency, Basement Studios, and Actnow Social Conscious Theatre Company; for administrative and project management support. Hilina Winkenweder, Hannah Wetham, Josh Fitton, Frida Svensson, Louise Armitage, Leda Sivak, and Kathleen Brodie. JW is supported by the Sylvia and Charles Viertel Senior Medical Research Fellowship and KC by a National Health and Medical Research Council Practitioner Fellowship (APP#1117582). This study was funded by a National Health and Medical Research Council Grant (#1100696).

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to the sensitivity of the data for Aboriginal and Torres Strait Islander people, and in line with ethical and data sovereignty principles.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT checklist.
[DOC File , 219 KB - jmir_v27i1e58341_app1.doc]

Multimedia Appendix 2
Questions on useability and adherence, intervention group only, and second follow-up.
[DOCX File , 17 KB - jmir_v27i1e58341_app2.docx]

References

1. Reilly R, Gendera S, Treloar C, Roe Y, Conigrave K, Azzopardi P, et al. Identifying risk and protective factors, including culture and identity, for methamphetamine use in aboriginal and torres strait Islander communities: relevance of the 'communities that care' model. Soc Sci Med 2020;266:113451. [doi: 10.1016/j.socscimed.2020.113451] [Medline: 33126097]
2. MacLean S, Harney A, Arabena K. Primary health-care responses to methamphetamine use in Australian indigenous communities. Aust J Prim Health 2015;21(4):384-390. [doi: 10.1071/PY14126] [Medline: 25704260]

XSL•FO
RenderX

3. Clough AR, Fitts M, Robertson J. Recent warnings of a rise in crystal methamphetamine ("ice") use in rural and remote indigenous Australian communities should be heeded. Med J Aust 2015;203(1):19. [doi: 10.5694/mja15.00066] [Medline: 26126560]
4. MacLean S, Hengsen R, Stephens R. Critical considerations in responding to crystal methamphetamine use in Australian aboriginal communities. Drug Alcohol Rev 2017;36(4):502-508. [doi: 10.1111/dar.12468] [Medline: 27726205]
5. Australian Institute of Health and Welfare. National Drug Strategy Household Survey 2019. Canberra: Drug Statistics series no. 32. PHE 270. Australian Institute of Health and Welfare; 2020.
6. Ward J, Elliott S, Bryant J, Donovan B, Pitts M, Wand H, et al. The Goanna Survey 2: Results of the second Australian survey of knowledge, relationships, behaviour and health service access relating to sexually transmissible infections (STIs) and blood borne viruses (BBVs) among Aboriginal and Torres Strait Islander young people. Adelaide: South Australian Health and Medical Research Institute; 2020. URL: https://youngdeadlyfree.org.au/wp-content/uploads/2021/01/GoannaSurvey2-FINAL.pdf [accessed 2021-04-20]
7. Snijder M, Kershaw S. Review of methamphetamine use among Aboriginal and Torres Strait Islander people. Australian Indigenous Health Bulletin 2019;19(3). [doi: 10.1017/9781108657815.013]
8. Degenhardt L, Sara G, McKetin R, Roxburgh A, Dobbins T, Farrell M, et al. Crystalline methamphetamine use and methamphetamine-related harms in Australia. Drug Alcohol Review 2017;36(2):160-170. [doi: 10.1111/dar.12426] [Medline: 27286742]
9. Reilly R, Wand H, McKetin R, Quinn B, Ezard N, Dunlop A, on the behalf of the NIMAC Community ResearchersPartner Sites. Survey methods and characteristics of a sample of aboriginal and torres strait Islander and non-indigenous people who have recently used methamphetamine: the NIMAC survey. Drug Alcohol Rev 2020;39(6):646-655. [doi: 10.1111/dar.13085] [Medline: 32573069]
10. Gendera S, Treloar C, Reilly R, Conigrave KM, Butt J, Roe Y, et al. 'Even though you hate everything that's going on, you know they are safer at home': the role of aboriginal and torres strait Islander families in methamphetamine use harm reduction and their own support needs. Drug Alcohol Rev 2022;41(6):1428-1439 [FREE Full text] [doi: 10.1111/dar.13481] [Medline: 35639622]
11. McKetin R, Degenhardt L, Shanahan M, Baker AL, Lee NK, Lubman DI. Health service utilisation attributable to methamphetamine use in Australia: patterns, predictors and national impact. Drug Alcohol Rev 2018;37(2):196-204. [doi: 10.1111/dar.12518] [Medline: 28294443]
12. Boumparis N, Karyotaki E, Schaub MP, Cuijpers P, Riper H. Internet interventions for adult illicit substance users: a meta-analysis. Addiction 2017;112(9):1521-1532 [FREE Full text] [doi: 10.1111/add.13819] [Medline: 28295758]
13. Carreiro S, Newcomb M, Leach R, Ostrowski S, Boudreaux ED, Amante D. Current reporting of usability and impact of mHealth interventions for substance use disorder: a systematic review. Drug Alcohol Depend 2020;215:108201. [doi: 10.1016/j.drugalcdep.2020.108201] [Medline: 32777691]
14. Washio Y, Hayashi Y, Atreyapurapu S, Chang K, Ma T, Howard BN, et al. A scoping review of computer-based and telecommunication technology interventions to address drug and alcohol misuse and smoking in women. Subst Use Misuse 2022;57(8):1257-1272 [FREE Full text] [doi: 10.1080/10826084.2022.2076878] [Medline: 35582861]
15. Sorcha M. Fighting “zombie” health apps through K-Factor virality and other mobile gaming techniques – the application of commercial gaming techniques to create more effective mHealth solutions. Front. Public Health 2016;4. [doi: 10.3389/conf.fpubh.2016.01.00048]
16. Reilly R, Stephens J, Micklem J, Tufanaru C, Harfield S, Fisher I, et al. Use and uptake of web-based therapeutic interventions amongst Indigenous populations in Australia, New Zealand, the United States of America and Canada: a scoping review. Syst Rev 2020;9(1):123 [FREE Full text] [doi: 10.1186/s13643-020-01374-x] [Medline: 32475342]
17. Zheng C, Conigrave JH, Conigrave KM, Wilson S, Perry J, Chikritzhs TN, et al. Patterns of drinking in aboriginal and torres strait Islander peoples as self-reported on the grog survey app: a representative urban and remote sample. Drug Alcohol Rev 2022;41(1):114-124. [doi: 10.1111/dar.13333] [Medline: 34153145]
18. Conigrave JH, Wilson S, Conigrave KM, Chikritzhs T, Hayman N, Dawson A, et al. 'The drug survey app': a protocol for developing and validating an interactive population survey tool for drug use among Aboriginal and Torres Strait Islander Australians. Addict Sci Clin Pract 2022;17(1):17 [FREE Full text] [doi: 10.1186/s13722-022-00298-2] [Medline: 35287718]
19. Nagel T, Dingwall KM, Sweet M, Kavanagh D, Majoni SW, Sajiv C, et al. The stay strong app as a self-management tool for first nations people with chronic kidney disease: a qualitative study. BMC Nephrol 2022;23(1):244 [FREE Full text] [doi: 10.1186/s12882-022-02856-x] [Medline: 35804297]
20. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian indigenous youth: a pilot randomised controlled trial. BMJ Open 2017;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
21. Reilly R, McKetin R, Wand H, Butt J, Smout M, Ezard N, et al. A web-based therapeutic program (We Can Do This) for reducing methamphetamine use and increasing help-seeking among aboriginal and torres strait Islander people: protocol for a randomized wait-list controlled trial. JMIR Res Protoc 2019;8(7):e14084 [FREE Full text] [doi: 10.2196/14084] [Medline: 33932278]

XSL•FO
RenderX

22. Ryan A, Holmes J, Hunt V, Dunlop A, Mammen K, Holland R, et al. Validation and implementation of the Australian treatment outcomes profile in specialist drug and alcohol settings. Drug Alcohol Rev 2014;33(1):33-42. [doi: 10.1111/dar.12083] [Medline: 24206571]
23. Wilson C, Deane F, Ciarrochi J, Rickwood D. Measuring help-seeking intentions: Properties of the general help-seeking questionnaire. Canadian Journal of Counselling 2005;39(1):15-28. [doi: 10.1037/t42876-000]
24. Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. Br J Addict 1992;87(5):743-754. [doi: 10.1111/j.1360-0443.1992.tb02720.x] [Medline: 1591525]
25. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand S, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine 2002;32(6):959-976. [doi: 10.1017/S0033291702006074]
26. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The severity of dependence scale (SDS): psychometric properties of the SDS in english and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]
27. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
28. Ritterband LM, Ardalan K, Thorndike FP, Magee JC, Saylor DK, Cox DJ, et al. Real world use of an internet intervention for pediatric encopresis. J Med Internet Res 2008;10(2):e16 [FREE Full text] [doi: 10.2196/jmir.1081] [Medline: 18653440]
29. McKetin R, Najman JM, Baker AL, Lubman DI, Dawe S, Ali R, et al. Evaluating the impact of community-based treatment options on methamphetamine use: findings from the methamphetamine treatment evaluation study (MATES). Addiction 2012;107(11):1998-2008. [doi: 10.1111/j.1360-0443.2012.03933.x] [Medline: 22564065]
30. Degenhardt L, Larney S, Chan G, Dobbins T, Weier M, Roxburgh A, et al. Estimating the number of regular and dependent methamphetamine users in Australia, 2002-2014. Med J Aust 2016;204(4):153. [doi: 10.5694/mja15.00671] [Medline: 26937668]
31. Tait RJ, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, et al. Six-month outcomes of a web-based intervention for users of amphetamine-type stimulants: randomized controlled trial. J Med Internet Res 2015;17(4):e105 [FREE Full text] [doi: 10.2196/jmir.3778] [Medline: 25925801]
32. Tait RJ, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, et al. A web-based intervention for users of amphetamine-type stimulants: 3-month outcomes of a randomized controlled trial. JMIR Ment Health 2014;1(1):e1 [FREE Full text] [doi: 10.2196/mental.3278] [Medline: 26543901]
33. Tait RJ, Christensen H. Internet-based interventions for young people with problematic substance use: a systematic review. Med J Aust 2010;192(S11):S15-S21. [doi: 10.5694/j.1326-5377.2010.tb03687.x] [Medline: 20528701]
34. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, et al. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. Am J Psychiatry 2008;165(7):881-888 [FREE Full text] [doi: 10.1176/appi.ajp.2008.07111835] [Medline: 18450927]
35. NA. Appsflyer: Attention Retention! 2022 App Retention Benchmarks Report. US: Appsflyer; 2022.
36. Shams F, Tai AMY, Kim J, Boyd M, Meyer M, Kazemi A, et al. Adherence to e-health interventions for substance use and the factors influencing it: systematic review, meta-analysis, and meta-regression. Digit Health 2023;9:20552076231203876 [FREE Full text] [doi: 10.1177/20552076231203876] [Medline: 37780062]
37. Lappan SN, Brown AW, Hendricks PS. Dropout rates of in-person psychosocial substance use disorder treatments: a systematic review and meta-analysis. Addiction 2020;115(2):201-217. [doi: 10.1111/add.14793] [Medline: 31454123]
38. Barbaro E, Grua EM, Malavolta I, Stercevic M, Weusthof E, van den Hoven J. Modelling and predicting User Engagement in mobile applications. DS 2020;3(2):61-77. [doi: 10.3233/ds-190027]
39. Pfund RA, Ginley MK, Rash CJ, Zajac K. Contingency management for treatment attendance: a meta-analysis. J Subst Abuse Treat 2022;133:108556 [FREE Full text] [doi: 10.1016/j.jsat.2021.108556] [Medline: 34210566]
40. Australian Bureau of Statistics. National Aboriginal and Torres Strait Islander Social Survey 2014-2015. Canberra: ABS; 2016. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/4714.0 [accessed 2021-07-10]
41. Leite JCDC, Luz MFDD, Walz JC, Filippin LI, Saldanha RP, Drachler MDL. Motivation and adherence to psychosocial treatment for alcohol and drug use-related problems. Estud. psicol. (Campinas) 2018;35(4):389-398. [doi: 10.1590/1982-02752018000400006]
42. Dean AC, London ED, Sugar CA, Kitchen CM, Swanson A, Heinzerling KG, et al. Predicting adherence to treatment for methamphetamine dependence from neuropsychological and drug use variables. Drug Alcohol Depend 2009;105(1-2):48-55 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.06.008] [Medline: 19608354]
43. Cook R, Quinn B, Heinzerling K, Shoptaw S. Dropout in clinical trials of pharmacological treatment for methamphetamine dependence: the role of initial abstinence. Addiction 2017;112(6):1077-1085 [FREE Full text] [doi: 10.1111/add.13765] [Medline: 28107598]

XSL•FO
RenderX

44. Bharat C, Webb P, Wilkinson Z, McKetin R, Grebely J, Farrell M, et al. Agreement between self-reported illicit drug use and biological samples: a systematic review and meta-analysis. Addiction 2023;118(9):1624-1648. [doi: 10.1111/add.16200] [Medline: 37005867]

Abbreviations

ACCHS: Aboriginal Community Controlled Health Services
AOD: alcohol and other drug
ATOP: Australian Treatment Outcome Profile
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
K10: Kessler 10
NIMAC: Novel Interventions to Address Methamphetamine in Aboriginal and Torres Strait Islander Communities
SDS: Severity of Dependence Scale

Edited by A Mavragani; submitted 13.03.24; peer-reviewed by J Mistry, D Ghosh; comments to author 22.04.24; revised version received 06.06.24; accepted 12.06.24; published 28.02.25.

Please cite as:
Reilly R, McKetin R, Barzi F, Degan T, Ezard N, Conigrave K, Butt J, Roe Y, Wand H, Quinn B, Longbottom W, Treloar C, Dunlop A, Ward J
Web-Based Application for Reducing Methamphetamine Use Among Aboriginal and Torres Strait Islander People: Randomized Waitlist Controlled Trial
J Med Internet Res 2025;27:e58341
URL: https://www.jmir.org/2025/1/e58341
doi:10.2196/58341
PMID:40053754

©Rachel Reilly, Rebecca McKetin, Federica Barzi, Tayla Degan, Nadine Ezard, Katherine Conigrave, Julia Butt, Yvette Roe, Handan Wand, Brendan Quinn, Wade Longbottom, Carla Treloar, Adrian Dunlop, James Ward. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Evidence on Digital HIV Self-Testing From Accuracy to Impact: Updated Systematic Review

Ashlyn Beecroft[1], MSc; Olivia Vaikla[2], BSc; Nora Engel[3], PhD; Thomas Duchaine[4], PhD; Chen Liang[1,2,5], PhD; Nitika Pant Pai[2,5], MD, MPH, PhD

[1]Division of Experimental Medicine, McGill University, Montreal, QC, Canada

[2]Research Institute of the McGill University Health Centre, Montreal, QC, Canada

[3]Athena Institute, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[4]Department of Biochemistry, McGill University, Montreal, QC, Canada

[5]Department of Medicine, McGill University, Montreal, QC, Canada

Corresponding Author:
Nitika Pant Pai, MD, MPH, PhD
Research Institute of the McGill University Health Centre
5252 boul de Maisonneuve W.
Montreal, QC, H4A 3S5
Canada
Phone: 1 514 934 1934 ext 44729
Email: nitika.pai@mcgill.ca

Abstract

Background: HIV self-testing has gained momentum following the approval of self-testing methods and novel technological advancements. Digital HIV self-testing involves completing an oral or blood-based HIV self-test with support from a digital innovation.

Objective: We conducted a systematic review on the existing data analyzing digital HIV self-testing accuracy while updating research on digital HIV self-test acceptability, preference, feasibility, and impact.

Methods: We searched Embase and PubMed for records on HIV self-testing with digital support. Included studies significantly incorporated a form of digital innovation throughout the HIV self-test process and reported quantitative data. For accuracy measures, the search spanned January 1, 2013, to October 15, 2024; for patient-centered and impact outcomes, we updated existing literature (June 16, 2021, to October 15, 2024) reported in a previous systematic review. Studies' quality was assessed using the QUADAS 2 Tool, Newcastle-Ottawa Scale, and Cochrane Risk of Bias Tool 2.

Results: Fifty-five studies (samples ranging 120-21,035, median 1267 participants) were summarized from 19 middle- to high-income countries. Seven studies reported on the accuracy of HIV self-testing with innovations from >5000 participants. Diagnostic performance metrics, including point estimates of specificity, sensitivity, positive predictive value, and negative predictive value were measured (n=3), and ranged from: 96.8% to 99.9%, 92.9% to 100.0%, 76.5% to 99.2%, and 99.2% to 100.0%, respectively. The percentage of invalid test results for oral and blood-based self-tests ranged from 0.2% to 12.7% (n=4). Fifty-one studies reported data on metrics beyond accuracy, including acceptability, preference, feasibility, and impact outcomes from >30,000 participants. Majority (38/51, 74.5%) were observational studies, while 25.5% (13/51) reported data from randomized controlled trials. Acceptability and preference outcomes varied from 64.5% to 99.0% (14/51) and 4.6% to 99.3% (8/51), respectively. Feasibility outcomes included test uptake (30.9% to 98.2%; 28/51), response rate (26.0% to 94.8%; 7/51), and visits to web-based providers (43.0% to 70.7%; n=4). Impact outcomes assessed new infections (0.0% to 25.8%; 31/51), first-time testers (2.0% to 53.0%; 26/51), result return proportions (22.1% to 100.0%; 24/51), linkage to care as both connections to confirmatory testing and counseling (53.0% to 100.0%; 16/51), and referrals for treatment initiation (44.4% to 98.1%; 8/51). The quality of studies varied, though they generally demonstrated low risk of bias.

Conclusions: Digital innovations improved the accuracy of HIV self-test results, and were well-accepted and preferred by participants. Operationally, they were found to be feasible and reported impacting the HIV self-testing process. These findings are in favor of the use of digital HIV self-test innovations as a promising support tool and suggest that digital HIV self-tests' service delivery models hold promise in not only facilitating HIV testing but also impacting operational outcomes that are crucial to reaching Joint United Nations Program on HIV/AIDS targets in middle- to high-income countries.

XSL•FO
RenderX

Trial Registration: PROSPERO CRD42020205025; https://www.crd.york.ac.uk/prospero/CRD42020205025

(J Med Internet Res 2025;27:e63110) doi:10.2196/63110

KEYWORDS

digital HIV self-testing; impact; linkage; outcomes; HIV; HIV infection; HIV self-testing; self-testing; digital innovation; systematic review; accuracy; patient-centered; middle- to high-income countries; digital health; mHealth; health education; sexually transmitted diseases; sexual behavior

Introduction

Background

To meet the Joint United Nations Program on HIV and AIDS (UNAIDS) targets, HIV self-testing strategies have been deployed in many countries around the world. HIV self-test methods have risen in popularity since being approved by the Food and Drug Administration (FDA) in 2012, with HIV self-tests now offered as oral or blood-based options, allowing users to receive their results within minutes [1]. Accompanying the rise in HIV self-test usage, digital innovations that support HIV self-testing are becoming widely used in health care, enhancing this method of self-testing. The World Health Organization (WHO) defines digital health innovations as technologies that contribute to improved health outcomes [2]. These technologies can be implemented within any step of the self-testing process including HIV self-testing pre- and posttest counseling, evidence-based knowledge sharing, test ordering, test result interpretation, linkages to care, referrals, and retention in care. Each of these steps plays a major role in the success of HIV self-testing programmatic implementation. The WHO advocates for the adoption and scale-up of digital health innovations to advance global health developments, in support of its One Health Agenda, which includes digital health as the great enabler of the One Health Agenda, which brings light to the integration and unity required to balance the health of humans, animals, and ecosystem [3]. HIV self-test with digital support holds the potential in improving patient-centered and operational outcomes. In the light of recent developments in digital health at the WHO, evidence synthesis was deemed necessary to summarize evidence on digital HIV self-test–based service delivery models [4].

Systematic reviews have sought to summarize global evidence on HIV self-test outcomes, as a result of the increased popularity of self-testing methods. A previous systematic review evaluated studies from January 1, 2010, to June 15, 2021, that focused on patient-reported outcomes including acceptability and preference, and operational feasibility and impact of HIV self-testing methods along with digital innovations [5]. The digital supports examined in this review included web-based interventions (ie, websites, chatbots, and online video counseling), social media and app-based innovations, SMS-based innovations, and digital vending machines. This review found these forms of digital support resulted in reasonably high acceptability (77%-97%), preference (53%-100%), feasibility (93%-95%), and impact (53%-100%) [5]. However, this review neglected to include accuracy metrics, due to insufficient literature at that time.

Despite the expansive integration of digital methods with self-testing during and following the COVID-19 pandemic, there has been limited work focusing on the role of digital support complementing HIV self-testing, particularly regarding the diagnostic performance of these tests. One secondary data analysis of a completed randomized controlled trials quasi–randomized controlled trial (quasi-RCT) has recently evaluated the accuracy metrics of digital HIV self-testing methods and compared them to laboratory HIV reference standards and reported high sensitivity and specificity at 95.5% and 99%, respectively [6]. To inform policy guidance, we need many more studies that analyze the diagnostic performance of digital HIV self-test. A related systematic review evaluating the accuracy of HIV self-testing, notably without digital supports, found that while participants generally performed the self-tests well, errors such as improper kit preparation, misconducting sample collection, and buffer solution spillage were common [7]. The performance accuracy of self-tests also varies tremendously depending on the user, type of test, and test setting, such as whether assistance from health care professionals is available. A systematic review comparing the accuracy of HIV self-tests conducted by the general population to that of diagnostic tests performed by health care practitioners concluded that HIV self-testing innovations were a reliable and accurate method of testing when used by the general population [8]. Although these reviews provide valuable insights into the accuracy of HIV self-testing, no systematic reviews have yet assessed the impact of digital innovations on HIV self-testing accuracy.

Considering the FDA only approved oral-based HIV self-tests in 2012, it is unsurprising that digital innovations in self-testing have rapidly evolved over the past decade, with the COVID-19 pandemic further accelerating advancements in this space, particularly in facilitating accuracy reporting [9]. This updated systematic review aims to address gaps in evidence regarding the effect of digital HIV self-test supports, while providing guidance for policy, practice, and research. These findings are especially relevant in the context of One Health and the global expansion of digital health service delivery models.

Objective

Our objective was to update global evidence on digital HIV self-test, assessing whether digital supports improved patient outcomes, ameliorated the process of conducting and interpreting self-test results, and affected impact outcomes.

We aimed to update evidence on all outcomes with digital innovations for HIV self-test, including patient-reported and operational outcomes such as acceptability, preference, feasibility, and impact.

XSL•FO
RenderX

Methods

Search Strategy

The first search strategy included in this review was conducted as an extension of the earlier systematic review by McGuire et al [5], which reported on the secondary outcomes (acceptability, preference, feasibility, and impact), but excluded accuracy. This review extended the search to include accuracy studies from 2013, reflecting the FDA's approval of oral-based HIV self-testing in 2012 [9]. The second search strategy was developed to expand upon the same secondary outcomes of McGuire et al [5], but from June 2021 onwards, where the original search strategy ended.

We followed the original protocol, registered on PROSPERO, and modified the strategy to incorporate the newly added accuracy outcome.

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and Cochrane guidelines to conduct and report this review.

No study participants or members of the public were involved in the design, conduct, or reporting of this review.

Information Sources

Two reviewers (AB and OV) searched 2 electronic databases (PubMed and Embase), first, for citations pertaining to previously unreported accuracy outcomes for the period of January 1, 2013, to October 15, 2024 (Multimedia Appendix 1); and second, for new citations pertaining to secondary outcomes (acceptability, preference, and feasibility) and impact outcomes for the period of June 16, 2021, to October 15, 2024 (Multimedia Appendix 1). There was an overlap of 3 papers that qualified for both the accuracy and patient-centered outcomes [10-12]. No restrictions were placed on language in either search. We retrieved all full-text studies and conference abstracts, with both authors (AB and OV) independently screening publications.

We also searched abstracts for the following conferences: Annual Canadian Conference on HIV and AIDS Research 2021/2022/2023, Infectious Diseases Society of America IDWeek 2021/2022, the 11th International AIDS Society Conference on HIV Science 2021, and the 24th International AIDS Conference (AIDS 2022).

Eligibility Criteria

We included all studies (observational and interventional [trials and quasi-RCT based experimental] designs) evaluating digital innovations facilitating HIV self-test in any country and those reporting quantitative results.

We included studies only if the digital supports were significantly used in the process of performing HIV self-tests (ie, administration, education, communication, result interpretation, and linkage to care).

We excluded qualitative studies, reviews, protocols, modeling studies, commentaries, narrative studies, case reports, and editorials; studies that did not have HIV as their primary focus did not include HIV self-test or did not use a digital technology; as well as studies not written in English.

Study Selection and Data Abstraction

Titles, abstracts, and full texts were screened independently by 2 reviewers (AB and OV) for eligibility, and the final included data were independently abstracted.

Abstracted data included study design, country, sample size, study population characteristics, digital innovation type, intervention description, and key findings.

A senior reviewer (NPP) was consulted for the resolution of disagreements.

Summary Outcome Measures and Narrative Synthesis of Results

To evaluate the integration of digital technology in HIV self-testing, we focused on a primary outcome (accuracy), and secondary outcomes (acceptability, preference, feasibility, and impact).

Accuracy metrics included sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), defined in Table 1 [13].

Table 1. Accuracy metrics definitions.

Metric	Definition
Sensitivity	A test's ability to correctly identify those with the disease, the proportion of true positives (numerator) over the proportion of true positives and false negatives (denominator).
Specificity	A test's ability to correctly identify those without the disease, the proportion of true negatives (numerator) over the proportion of true negatives and false positives (denominator).
Positive predictive value	The probability that the disease is present, given the test result is positive, the number of true positives (numerator) over the number of true positives and false positives (denominator).
Negative predictive value	The probability that the disease is absent, given the test result is negative, the number of true negatives (numerator) over the number of true negatives and false negatives (denominator).
Invalid test result	A test result that did not correspond to the expected outcome of control and test lines for either negative or positive results.

Furthermore, secondary outcomes (patient-centered acceptability, feasibility, and preference) and operational impact outcomes were adopted from a previously published systematic review and are defined in Table 2 [5].

XSL•FO
RenderX

Table 2. Outcome measures and definitions.

Outcome	Definition
Acceptability	The ease of use and willingness of participants to use digital innovations for HIV self-testing, defined as those who agreed to use or try the digital innovation (numerator), over all those who were enrolled in the study (denominator).
Preference	The proportion of study participants who preferred HIV self-tests with digital supports over conventional HIV testing, defined as those who prefer this method of self-testing (numerator) over all those who were enrolled (denominator).
Feasibility	The convenience of using HIV self-test with digital supports, reported with self-test uptake, response rate, and visits to web-based HIV self-testing providers.
Impact	A statistically significant improvement in measured outcomes compared with a comparator group or a net change in outcomes among a particular group that can be attributed to specific intervention–reported metrics include the proportion of first-time testers, detection of new infections, HIV self-test kit return rate, the proportion of participants linked to continued care including counseling or confirmatory testing, and proportion of those referred to treatment.

Quality Assessment

We assessed the quality of diagnostic accuracy studies with the QUADAS-2 (quality of diagnostic accuracy studies) Tool for accuracy outcome papers [14]. We used the Cochrane Risk of Bias Tool 2 (RoB 2) to assess the quality and potential risk of bias of RCTs [15], and the Newcastle-Ottawa Scale (NOS) was used for cohort and cross-sectional studies [16,17].

Results

Study Selection

For the first search, 146 studies were retrieved, 139 were excluded, and there was a final set of 7. For the second search, 543 studies were retrieved, 492 were excluded, and there was a final set of 51. Out of 3 of the studies found overall were used for both primary and secondary outcomes, therefore a total set of 55 studies was included in this review. We found significant heterogeneity in the reporting of populations, study designs, HIV self-test interventions, and outcome metrics that precluded a meta-analysis.

Accuracy

For this outcome, 7 studies were included in the final analysis (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram, accuracy outcome.

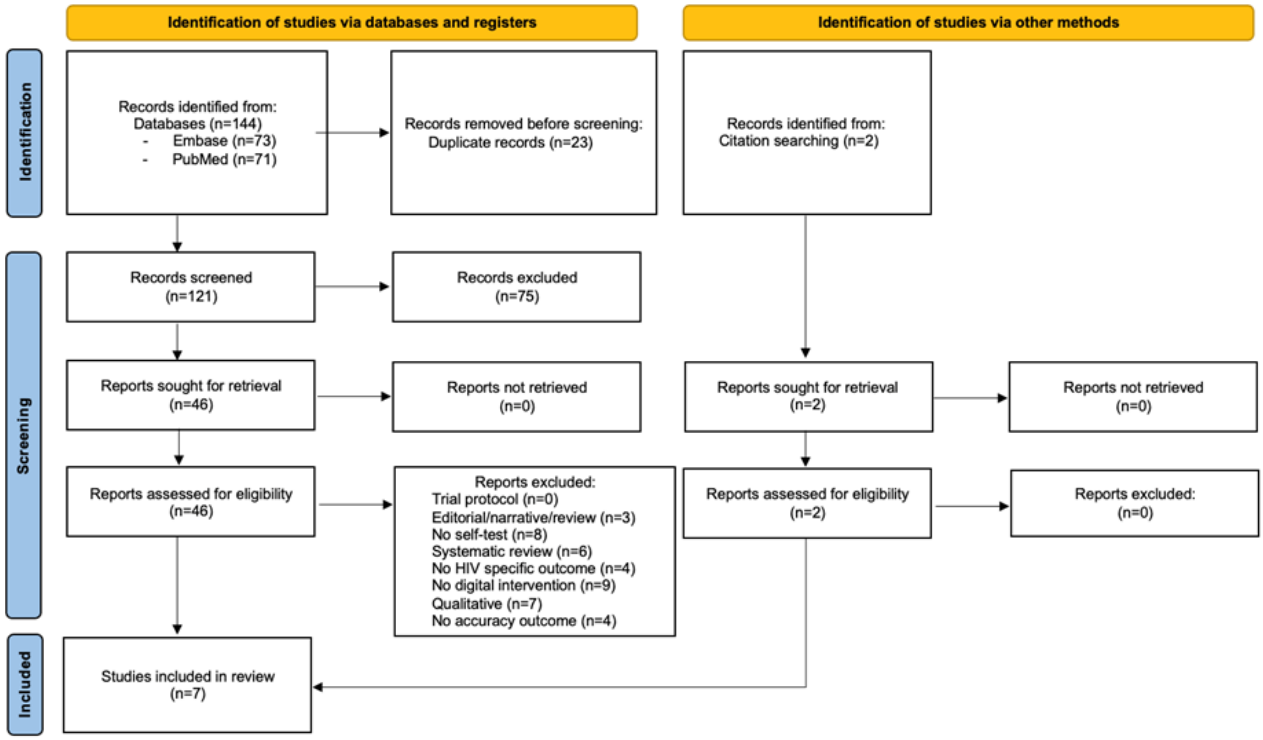

Secondary Outcomes (Acceptability, Preference, Feasibility, and Impact)

For these outcomes, 51 studies were included (Figure 2).

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram secondary outcomes.

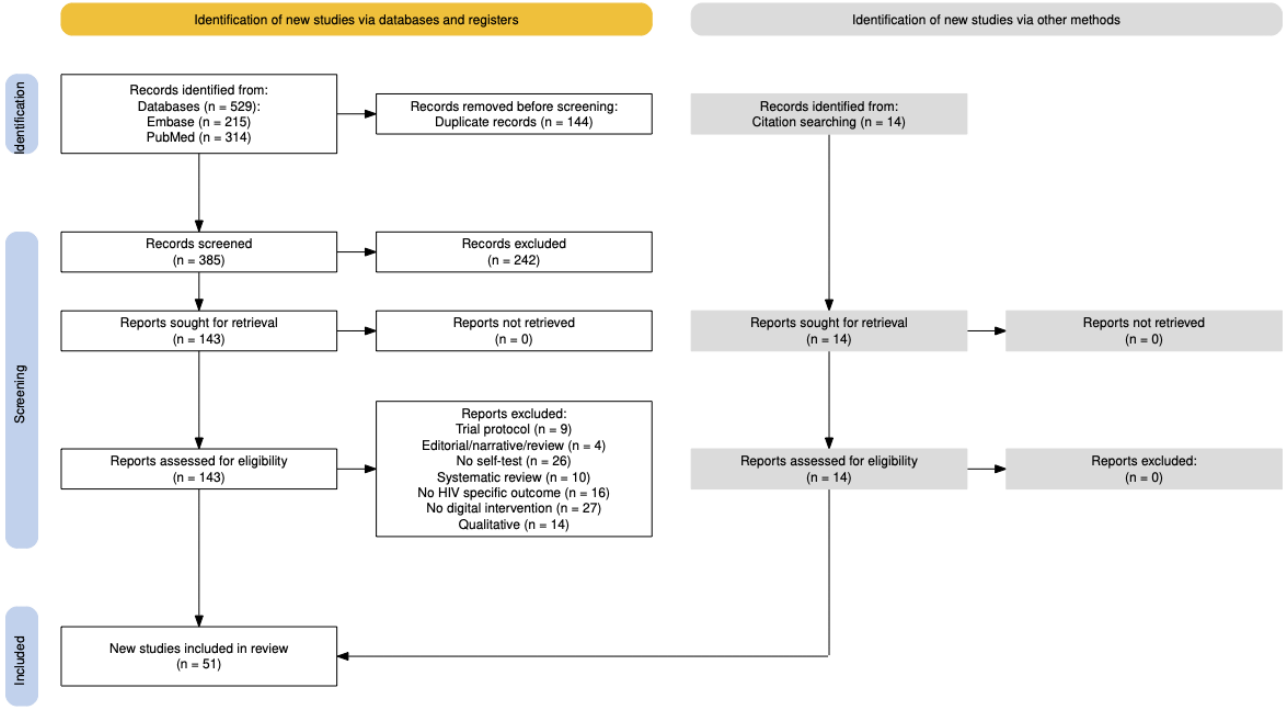

Study Characteristics

Accuracy

Accuracy studies were reported from each the following countries: Canada, Kenya, South Africa, the United Kingdom, the United States, and China (n=2). Out of 4 (57.1%) were cross-sectional, while 2 (28.6%) were cohort, and 1 (14.3%) was a quasi-RCT.

Study sample sizes ranged from 271 to 3259 participants.

Study characteristics are tabulated in, Table S1 in Multimedia Appendix 1.

Most studies used website-based HIVST innovations (5/7, 71.4%), one used app-based digital support (1/7, 14.3%), and another used a multimodal approach to HIV self-test including app-based, SMS-based, social media, and web-based innovations (1/7, 14.3%).

Out of 4 (57.1%) studies were conducted in the general population, and the other 3 (42.8%) recruited key populations (men who have sex with men [MSM], or specifically Black, African American, or Latinx MSM).

Secondary Outcomes

In total, 51 studies reported on secondary outcomes from studies conducted in 18 countries. These included: Canada (19.6%, 10/51), China (19.6%, 10/51), the United States (13.7%, 7/51), South Africa (11.8%, 6/51), India (7.8%, 4/51), and Thailand (3.9%, 2/51). One study (2.0%, 1/51) was reported in each of the following countries: Australia, Brazil, Japan, Kazakhstan, Malaysia, Nigeria, Philippines, Uganda, Ukraine, the United Kingdom, Vietnam, and Zimbabwe.

Sample sizes varied from 120 to 21,035 (median 1269, IQR 3305) participants (Table S2 in Multimedia Appendix 1).

Study design

A majority (74.5%, 38/51) of studies were observational (62.7%, 32/51 cross-sectional; 11.8%, 6/51 cohort) and 25.5% (13/51) of studies were RCTs.

Population

Nearly half of the studies (47.1%, 24/51) focused on MSM populations, while the rest (11.8%, 6/51) specifically focused on Black, African American, or Latinx MSM, and 7.8% (4/51) were conducted in transgender women. Over one-third (43.1%, 22/51) of studies were evaluated in the general population. One study only focused on women, and another only focused on Black, Caribbean, and African-Canadian people.

Interventions

Over half (56.9%, 29/51) evaluated outcomes from web-based innovations, 21.6% (11/51) evaluated mobile app-based innovations, 5.9% (3/51) evaluated social media, digital vending machines, and SMS were evaluated in one study each (2.0%, 1/51), and 11.8% (6/51) of studies evaluated multimodal interventions (web-based, social media, or SMS-based).

Risk of Bias in Studies

QUADAS 2 Tool

Using the QUADAS 2 Tool, we found a low risk of bias for the studies. 71.4% (5/7) missed details on the reference standard test used.

Key findings can be found in Table S3 in Multimedia Appendix 1.

RoB2

Using the RoB 2, we found low risk of bias for the included RCTs. Blinding of participants was reported by 30.8% (4/13) of studies. Although blinding of the participants and assessors

XSL•FO
RenderX

through the trial and analysis was not possible due to the assessed outcomes of digital supports, there were no reports of deviations from the intended interventions that arose in any of the studies. As well, every study had majority or all the data present in the analysis, and attrition bias was minimized. Finally, the outcome was reported in detail, minimizing the possibility of reporting bias.

Further details regarding these key findings can be found in Table S4 in Multimedia Appendix 1.

NOS for Cross-Sectional Studies

Using the NOS, we found that cross-sectional studies were generally fair. 53.1% (17/32) of these studies had an overall fair quality. The possibility of selection bias, confounding, or outcome and exposure misclassification was found in 18.8% (6/32), 25.0% (8/32), and 34.3% (11/32) of these studies, respectively. Many of the cross-sectional studies (90.6%, 29/32) had unjustified sample sizes (no sample size calculation was mentioned).

Further details regarding these key findings can be found in Table S5 in Multimedia Appendix 1.

NOS for Cohort Studies

Using the NOS, we found cohort studies were of average quality. Possibility of selection and confounding bias were low at 33.3% (2/6), respectively. However, 83.3% (5/6) of the cohort studies had a risk of attrition bias.

Further details regarding these key findings can be found in Table S6 in Multimedia Appendix 1.

Primary Outcome: Accuracy

A total of 7 studies evaluated the accuracy of HIV self-test integrated with the use of digital innovations [6,10-12,18-20]. Overall, 42.3% (3/7) reported on blood-based self-tests [12,18,20], 42.3% (3/7) on oral-based self-tests [6,10,19], and one allowed participants to choose between oral or blood-based self-tests but reported no differences in accuracy between the 2 types of tests [11].

A trend was noted on the reporting of invalid test results, since a majority (57.1%, 4/7) of studies reported on the percentage of invalid test results. Invalid tests reported ranged from <0.2% to 12.7%, for both blood- and oral-based self-tests [10-12,18].

One study in which the participants were asked to perform the self-test on site and report their result to the researcher as soon as it was available, reported on diagnostic performance metrics (ie, sensitivity, specificity, PPV, and NPV) of the oral-based Aware HIV self-testing, but notably only as point estimates without CI: Sn 92.9% (13/14), Sp 96.8% (121/125), PPV 76.5% (13/17), and NPV 99.2% (121/122) [19].

A study that was a secondary data analysis of a quasi-RCT found high accuracy metrics when comparing the HIV self-testing result with the reference standard of 2 rapid tests and 1 HIV RNA test [6]. Sensitivity was reported as 95.52% (95% CI 94.48%-96.56%), specificity was 99.93% (95% CI 99.79%-100.00%), PPV was 99.22% (95% CI 98.78%-99.67%), and NPV was 99.57% (95% CI 99.24%-99.90%) [6].

Another notable study that reported on such metrics did so using an AI-based program that was able to read and interpret the result of the participants' self-test [20]. In this study, sensitivity was 100% (44/44), as well as NPV. Specificity was also high at 98.8% (798/808), but the PPV was lower at 81.5% [20].

A fourth study allowed participants to choose between an oral- or blood-based HIV self-testing, but did not include mention of the manufacturers of these tests in the publication. Only 2 of the 4 positive HIV self-tests were confirmed to be positive but neglected to mention if these results were from the oral- or blood-based self-testing options. The participants of this study ordered the self-test and completed it at home, then uploaded their self-test results onto the web-based platform, which could have contributed to the few positive self-tests received. This study also reported the agreement of test interpretation between participants and the research team as 99.1% (95% CI 97.4%-99.8%) [11].

A study from the United Kingdom reported on 24,717 blood-based self-test kits purchased during the study, of which there was only a <0.2% rate of invalid tests and only 3 false positives were reported, which was much lower than they had expected [18].

Further details regarding these key findings can be found in Table S7 in Multimedia Appendix 1.

Secondary Outcomes

Outcome measures beyond accuracy become important when real life implementation is called into question. Due to this, we explored patient-reported acceptability and preference, and operational feasibility and impact outcomes.

Further details regarding these key findings can be found in Table S8 in Multimedia Appendix 1.

Acceptability

Overall, 27.5% (14/51) of the studies reported acceptability measures [21-34].

Ease of Use

A total of 8 studies (57.1%, 8/14) reported acceptability as the ease of use of the digital interventions [22,23,28-30,32-34]. In 1 study, 64.5% of participants found the digital innovation easy or very easy to use [29]. Comparatively, an RCT evaluating an infographic for HIV self-testing found that 73.5% of participants agreed that it was easy to use [30]. Another study reported that only 7.5% (9/120) of participants said the digital support was easy to use, though one participant mentioned: "*It is very easy to use, less stressful very understandable, it is the best and very advanced product one could ever wish for*" [28]. Alternatively, 1 study measured ease of use using a 5-point Likert scale and found a score of 3.8 (SD 1.6) for the ease of uploading results, and a score of 4.2 (SD 0.9) for ease of finding a clinic using the digital innovation [23].

Willingness to Use

Overall, 42.8% (6/14) reported on the willingness of participants to use the self-test with the digital intervention, which was consistently high (72.2%-99.0%) [23-27,31]. A quasi-RCT in the Philippines demonstrated the willingness of participants to

XSL•FO
RenderX

use a community-based HIV self-test distribution model and found that 99.0% (4163/4205) of respondents were interested in getting an HIV self-test [31].

Utility of the Intervention for an Indication

In one study, participants found the digital innovation helpful in reporting of testing results (97.6%), the concept of window period (88.8%), understanding their current risk (85.6%), reducing high-risk behaviors (80.0%), and mitigating fear of HIV testing (72.0%) [22].

Preference

Preference was assessed in 15.7% (8/51) of the studies [11,24,28,35-39].

Preference for the digital innovation versus standard testing methods ranged from 63.0% to 99.3% (3/51) [24,28,37]. An outlier study found that only 5.9% (128/2181) of participants chose to do an HIV self-test with the support of an application on their phone, instead of HIV testing conducted by a health care provider [39].

One study analyzed data from 794 participants on their reasoning for preference of HIV self-testing online versus other methods of testing. These were reportedly: convenience (726/794, 79.0%), not wanting to wait for results (402/794, 44.0%), not wanting to talk about sex with anyone (298/794, 33.0%), not having time to go elsewhere for testing (268/794, 29.0%), and fear of stigma (205/794, 22.0%) [35]. Notably, 21.0% (40/190) of participants who responded to the survey question said they would not test elsewhere at all [35].

One study reported on lower rates of preference; a majority of participants who requested a self-test did not want any form of support (338/434, 77.9%) [11]. However, of those who did want support, digital forms of support including instant messaging, video calls, and chatbots were preferred by 65.8% (102/155) of respondents, compared to in-person support (47.7%, 74/155) [11].

Feasibility

Feasibility was assessed in 60.8% (31/51) of the studies [10-12,21-23,29,31,34,35,40-60].

Uptake

Overall, 90.3% (28/31) of these studies reported uptake of self-tests [10-12,22,31,34,35,40-60], which varied across studies (30.9%-98.2%; 11/31) [10,11,22,31,40-42,44,45,48,49,54]. Out of 9 of the remaining studies measured uptake as the number of HIV self-test kits ordered throughout the study. This included 701 kits ordered by 604 participants, 834 tests ordered by 309 participants, 5840 kits ordered by 5324 participants, 5235 kits ordered by 3627 participants, 5323 kits ordered by 4859 participants, 13,334 kits ordered by 11,332 participants, and 3 studies in which each participant ordered a single test (2610, 7315, and 21,035) [12,46,47,50,52,53,55,57,58]. An Australian study reported that of the 794 participants who ordered an HIV self-testing, 95 of them ordered multiple self-testing kits, between 2 to 7 kits per person [35]. An RCT conducted in the United States investigated the effect that a peer-led online community had on the uptake of the blood-based myLAB Box HIV self-testing [51]. This study found there was an increase of 6.3% in test uptake between the intervention and control groups, with an odd ratio (OR) of 1.43 (95% CI 1.04-1.95, *P*=.03) [51]. Another RCT conducted in China analyzed the effect of monetary incentives and online peer referrals on test uptake of the blood-based SD BIOLINE HIV and syphilis self-test [52]. This study found that the 102 control participants ordered 222 kits, the 103 participants of the monetary intervention group ordered 275 kits, and the 104 participants in the monetary and peer referral intervention group ordered 337, all of which were ordered using a digital intervention [52].

Proportion of Participants Who Responded to the Intervention

About 22.6% (7/31) of studies reported how many of the participating people actually responded to the intervention, with one study reporting a response rate of 26.0% (2467/9505), and the others 61.1%-94.1% (216/345 to 96/102) [11,29,35,47,51,52,54]. The outlying response frequency of 26.0% 2467/9505 referred to responses of an optional online survey asking questions on participants' quality of experience, which was offered once the self-testing process was complete [29]. In RCT, the response proportion was higher among the intervention group (93.4%, 421/450) compared with the control group (92.9%, 418/450) [51]. As well, another RCT found the response rate to be higher among the intervention groups (94.2% and 96.2%) compared with the control group (94.1%) [52].

Visits to the Web-Based Provider

Out of 4 studies reported such, of which three reported as percentages of 43.0% (1475/3431) of participants interacted with the content, 70.7% (531/751) of participants logged onto the app, and 67.9% (19/28) clicked on messages sent on the app [21,23,60]. The remaining paper reported that the study website was viewed about 266,000 times [47].

Impact

Impact was measured in 88.2% (45/51) of the studies [4,10-12,21-25,27,29,31-35,37-50,52,53,55-67].

Detection of New HIV Infections

This was the most prominent metric used in 68.9% (31/45) of the studies [4,10,11,22-24,29,31-33,35,40-50,52,55-59,61,62,64]. A majority of these papers (61.3%, 19/31) reported the proportion of new infections by HIV self-tests result which varied from 0.0% to 25.8%, which were confirmed by laboratory result [10,11,23,24,29,31,32,40,41,44-46,48,49,57-59,61,64]. One study reported that 18 participants returned a positive self-test result [61]. Another study found that of 5048 returned self-test results, 165 were positive [47]. One study noted that 41.1% (130/314) of the participants that tested positive had not received a positive result before [29]. The other 8 papers reported the proportion of new infections as those confirmed by laboratory testing, ranging from 0.0% to 10.1% [4,22,33,35,42,43,52,56]. One study did not specify how many results were returned, but stated that 86 people who conducted the self-test were confirmed via laboratory result at the project sites [50]. Another study found that out of those who sought confirmatory testing, 96.8% (522/539) were confirmed to be HIV positive [55]. A quasi-RCT found the proportion of new

XSL•FO
RenderX

infections to be higher in the intervention group (8.9%), which included unsupervised and supervised self-testing, compared with the control group (6.8%) risk ratio (RR; 1.304, 95% CI 1.023-1.665) [4].

The Proportion of "First-Time Testers"

Reported in 57.8% (26/45) of the studies [11,12,21,24,25,27,31,34,35,37,42,44,45,48,49,52,53,55-57,61-64,66,67]. The proportion of first-time testers for 23 of these studies ranged from 2.0%-53.0% [11,12,21,24,25,27,31,34,35,37,42,44,45,48,52,53,55-57,62,64,66,67]. In addition, one study looked at 1070 returned self-tests and found 43 of them to be positive, all of which were from first-time testers [49]. One study compared the proportion of first-time testers between age groups and found that 53.0% of those who were 18-24 years old were first-time testers, compared with only 20%-30% of those in the higher age groups (OR 3.30, 95% CI 2.88-3.78) [63]. Another study looked at the difference in rate of first-time testers among White participants compared with African, Caribbean, and Black participants and found that the proportion of White first-time testers was 21.0%, compared with 30.0% among African, Caribbean, and Black participants [61].

Result Return Proportion

Defined as the percentage of reported results from participants and was captured in 53.3% (24/45) of the studies, ranging from 22.4%-100.0% [10,12,23,29,31,32,38,39,41,44-46,48,49,52,53,55,56,60-62,64,65,67]. An RCT found that the return rate among the control group (94.0%) was higher than the intervention group that did not receive online peer referral (93.4%), but was lower than the intervention group that did receive online peer-referral (97.3%) [52].

Linkage to Care

Observed in 46.6% (21/45) of studies [4,22,24,29,31,33-35,40,42,43,45-49,53,55,57,62,64]. 76.2% (16/21) of these studies reported this metric as confirmatory HIV testing and linkage to treatment which varied from 53.0%-100.0% [22,24,29,31,33-35,40,42,45,47,53,55,57,62]. Out of 8 studies (38.0%) reported linkage to care as the percent of confirmed HIV-positive participants who were referred to and initiated antiretroviral therapy (ART) and ranged from 44.4%-98.1% [4,43,46,48,49,53,55,64].

A quasi-RCT reported the highest linkage to care, with a greater proportion for the intervention (99.8%) compared to the control (98.5%; RR 1.012, 95% CI 1.005-1.018) [4]. As well, 98.1% of all those who used HIV self-test were either referred to start ART if they were HIV-positive, or for preventive treatment if they were HIV-negative [4].

Referrals to Self-Test

A quasi-RCT also showed that 16.7% of participants in the self-testing group referred someone in their social network, whereas only 3.1% of participants in the conventional testing arm did (RR 5.435, 95% CI 4.024-7.340) [4]. In another RCT, it was reported that 78.0% (479/618) of participants had given a self-test to a social network associate [67].

Discussion

Principal Findings

Accuracy

Invalid tests were not explicitly defined in any of the studies but were assumed to refer to a result that was neither negative nor positive. Invalid results are a product of a defective test, incorrect testing conduct, or result in misinterpretation. Invalids are an important factor to consider for diagnostic performance, as users' trust in the self-testing process may diminish when an invalid result is received. The WHO and other governing bodies lack defined thresholds for acceptable invalid results in screening. We found the highest invalid rate (12.7%) occurred in Canada, using the blood-based bioLytical INSTI self-test, Canada's sole approved HIV self-testing so far [12]. There is a need to define an acceptable threshold of invalid tests at the level of approval.

Sensitivity, specificity, PPV, and NPV are crucial for evaluating the performance of HIV self-testing and determining the reliability of results. Although 7 studies reported high diagnostic performance with HIV self-testing as the low proportion of invalid tests or high sensitivity, specificity, PPV, and NPV, the overall body of literature in this area remains sparse.

One study analyzed the impact of a digital reader in self-result interpretation, compared to the participants' interpretation [20]; however, no studies assessed the improved accuracy of digital intervention-enhanced HIV self-tests compared to those without digital interventions that are widely accessible to the general population. This highlights the need to explore the potential of digital innovations in improving the accuracy of HIV self-test.

Furthermore, it is important to highlight that one of the studies reporting all the ideal accuracy measures evaluated the impact of internet use on HIV self-test uptake and performance, not the direct influence of digital interventions on the self-testing process [19]. This indicates a research gap in understanding the direct impact of digital interventions on the accuracy of HIV self-testing. This study also reported moderate measures of accuracy, suggesting room for improvement of digital supports that are implemented during the testing process.

One of the other studies that reported the ideal accuracy metrics did so comparing interpretation by (1) participants, (2) pharmacy provider, (3) AI tool, and (4) expert panel of 3 HIV self-test readers [20]. The reported accuracy metrics were comparing the AI tool to the expert panel, instead of between the AI tool and participants, which again highlights this gap in existing research [20].

The final study that reported metrics on diagnostic performance did so comparing the HIV self-test result to the reference standard, reporting high accuracy measurements for each outcome [6]. However, there is still a gap in research comparing the results of HIV self-testing with the support of a digital intervention compared to an HIV self-test result on its own. Future research should aim to analyze the accuracy between HIV self-tests alone versus with digital support, to properly

XSL•FO
RenderX

evaluate the direct impact digital innovations have on the self-testing process.

Acceptability

Acceptability of digital interventions for HIV self-test varied across studies, indicating that acceptability is influenced by diverse factors. Persistent willingness to use digital interventions across studies (72.2%-99%) suggests openness and receptiveness to incorporating digital technologies into the self-testing process. To improve the acceptability of digital interventions, future developments should consider comprehensive HIV self-testing integration, including sample collection and result interpretation.

Ease of use varied significantly in definition across the 8 studies that reported this metric, but overall, most study participants found the digital innovations easy to use. It is important to note that the demographics of the participants in these studies may play a role in the reported ease of use. For example, in 1 paper, 122/350 participants were people who had used an online HIV self-testing platform before, which may have increased their report of ease of use [22].

Preference

This systematic review demonstrates generally high rates of preference (63.0%-99.3%) for digital interventions in the context of HIV self-testing, suggesting notable favorability toward digital innovations for HIV self-test.

One study reported lower rates of preference for any form of support during the HIV self-test process, as most participants in that study expressed a preference for not receiving any support [11]. For those seeking support, digital options were preferred over in-person support, underscoring diverse preferences and the need for tailored interventions. The novelty of digital innovations can be overwhelming, particularly among older generations and individuals who are less familiar with technology, which may influence their preferences of digital interventions for HIV self-test. It is crucial to consider the target population and their digital literacy level, and to provide adequate support, education, and user-friendly interfaces to address potential barriers related to technology adoption.

Another study reporting low preference had given participants the choice between HIV self-testing with a mobile app and HIV testing conducted by a health care provider [39]. This study was conducted among youth aged 16-24 years in Zimbabwe and qualitatively asked the participants why they may have chosen standard HIV testing compared with self-testing with the digital app [39]. Common explanations included low general and digital literacy, limited self-efficacy or agency, fear of testing and need for provider support, lack of private digital or physical space, and technical issues [39]. Given the specific geographic region of the age group included in this study, it is possible that these findings may not be replicated if repeated in another area of the world, for example, the United States, where digital literacy is likely to be improved. However, these qualities still need to be addressed in HIV self-testing and provide people with comfort and a sense of autonomy when self-testing, instead of fear and hesitation. As well, the preference rate found in this study cannot be attributed solely to the digital aspect of self-testing, but the self-testing process as a whole. Increasing awareness of HIV self-testing methods could enhance the outcomes of similar studies, as participants with greater familiarity and understanding of the testing approach may feel more confident in selecting the digital self-testing option.

Feasibility

High uptake of self-tests facilitated by digital interventions suggests the broad accessibility and convenience associated with digital HIV self-test.

In one RCT, the overall uptake of HIV self-testing was relatively low, but the inclusion of a digital component increased test uptake by 6% (OR 1.43; 95% CI 1.04-1.95), highlighting the potential of digital interventions to positively influence self-test uptake rates [51].

Another study reported a 25.0% increase in clinical HIV testing through the implementation of a digital solution [43]. This suggests that the widespread application of digital innovations in HIV self-test could contribute significantly to achieving the UNAIDS 2025 targets of increased HIV serostatus knowledge and treatment.

Response rates substantially varied across studies, which may have been influenced by follow-up procedure disparities and the recruitment of vastly diverse populations. Tracking and analyzing response rates are vital to assess the performance of digital interventions and identify effective methods for obtaining feedback via digital means.

Visits to web-based providers varied across the studies, potentially attributed to differences in the methods of exposure to digital innovations. For example, in one study, the intervention involved a campaign about HIV self-test disseminated through social media platforms, which have algorithmic influence on exposure, limiting exposure of certain individuals to the digital intervention [21].

Altogether, the inclusion of digital components has been shown to increase test uptake, and the feasibility of digital interventions are supported by the available evidence. However, response rates and variations in exposure methods should be carefully examined and addressed to optimize the performance and impact of digital interventions in HIV self-testing.

Impact

The use of HIV self-testing methods with digital support has shown potential in reaching the first of the UNAIDS 95-95-95 targets, supported by the proportions of new infections and first-time testers.

The parity between new infections identified via HIV self-test and confirmed HIV cases, 9.8% versus 10.1%, indicates that HIV self-test with digital support is a valuable tool for identifying new cases of HIV. Importantly, a study in South Africa reported an increase in the proportion of new infections among the HIV self-testing arm compared to the conventional testing arm (RR 1.305, 95% CI 1.023-1.665), supporting digital aids in enhancing HIV detection [4].

The proportion of first-time testers supports digital innovations' potential in streamlining HIV self-testing access. Accessibility of digital HIV self-testing methods to the general population is

XSL•FO
RenderX

likely to elevate first-time tester rates, capturing those not engaged via traditional methods, and thus expanding testing coverage.

Digital support facilitated high test result returns across studies, likely attributed to the convenience and accessibility digital reporting affords. The ease of reporting results digitally eliminates the need for physical clinic visits, fostering convenience and potentially encouraging higher rates of result reporting.

While one study observed a lower self-test return rate due to mobile-app user attrition, this could be context specific, as the study was conducted in South Africa in 2018/2019, and thus subject to change due to evolving mobile-app utilization trends and self-testing methods used through the COVID-19 pandemic [23].

Linkage to care is a critical aspect of HIV testing. HIV self-testing without digital support have struggled with adequate linkage to care [68]. Digital innovations significantly aid this process, enabling direct participant connections to care without having to navigate the resources themselves. In one study, there was linkage to post-test counselors, who facilitated staging the disease of those who tested positive and assisted in preventative practices for those who tested negative, yielding a robust linkage proportion (99.7%) among the self-testing arm [4].

The findings of this review support the high linkage to care, as the high proportion of individuals who received a positive self-test result were connected to confirmatory testing, underscoring digital support's role in facilitating real-world follow-up. Furthermore, there was a high proportion of individuals who were confirmed to be HIV-positive and started ART treatment which demonstrates the promising potential of digital interventions in successfully connecting HIV-positive patients to continuative care. These findings highlight the practicality and convenience of digital interventions in the HIV testing and care continuum.

Overall, the use of digital innovations in HIV self-test has resulted in the identification of new infections of HIV, an increased proportion of first-time testers, and a high return rate of test results.

Quality Assessment

The accuracy studies generally had low risk of bias; however, a lot of information was not specified. For 57.1% (4/7) of the studies, invalid test results were used as a proxy for real-world performance and were limited by a lack of reference standards to complete the test performance. The studies did not mention if the participants also received confirmatory laboratory testing, except for 2 [6,19]. The RCTs included in this review had a low risk of bias. There is the confidence that these studies were well conducted and have trustworthy results.

Overall, the cross-sectional studies had low risk of confounding bias, but some or high risk of selection bias and outcome misclassification. The high risk of selection bias was mostly due to lack of justified sample sizes. As for outcome misclassification, the assessment of the outcome was consistently self-reported, considering the nature of self-testing, thus increasing the risk of bias. It is also important to note that many of the conference abstracts neglected to include details pertaining to the quality assessment; therefore, the study may have conducted higher quality of research but was not outlined in the abstract alone.

The cohort studies were of average quality. Out of 2 studies had missing data, and another had only one star and were therefore classified to be of low-level comparability [48,57,58]. Overall, the quality of these studies can be trusted in terms of the selection processes, but there should be some hesitation regarding comparability and outcome biases.

Limitations

The heterogeneity of interventions, outcomes, populations, and settings among the included studies introduced variability and made it challenging to draw definitive conclusions or generalize the findings across various contexts. Heterogeneity in reporting outcomes poses a challenge in synthesizing the evidence. Variation in outcome measures and their assessment methods makes it difficult to pool data and establish a comprehensive understanding of the effectiveness of digital supports. Future studies should adopt standardized outcome measures and reporting guidelines to facilitate meaningful comparisons and meta-analyses.

Limited evidence was available from low-income countries. Many studies included in this review were conducted in the middle- or high-income settings, such as China, the United States, Canada, and South Africa, which may limit the generalizability of the findings to low-income countries. This highlights the need for more research in diverse geographic and socioeconomic contexts to assess the feasibility and effectiveness of digital health interventions across different resource settings.

Several limitations in the review process should also be acknowledged. The search strategy was restricted to 2 databases, which may have resulted in the potential omission of relevant studies published in other databases. Moreover, this review focused on a majority of observational studies, which inherently lack the randomization and control provided by RCTs, potentially influencing the observed effects of digital health interventions.

Implications and Future Research

This review incorporates 14 RCTs (including 6 quasi-experimental trials), 34 cross-sectional, and 7 cohort studies, providing a comprehensive overview of the existing evidence. These studies offer valuable insights into the effectiveness and potential benefits of digital innovation in health care practices in terms of accuracy, acceptability, preference, feasibility, and impact. Moving forward there is a need to enforce reporting of trials in HIV self-testing as per the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [69], as well as a need to appreciate the use of quasi-randomized designs that provide a greater richness to the breadth of literature by virtue of implementing these solutions in real-life settings. Single-arm interventional studies also provide evidence on why and how an intervention could

XSL•FO
RenderX

be successful in a setting. Bias is an inevitable consequence of all study designs, less so with an RCT, followed by quasi-randomized and single-arm interventional studies, in that order. However, evidence generation with the use of a stronger methodology adds to the evidence base needed to formulate guidelines and to the GRADE (Grading of Recommendation Assessment, Development and Evaluation) approach [70].

In addition, it is important to consider patient preferences in testing processes, as this can have a profound impact on retention rates. Qualitative research underscores digital strategies' flexibility, allowing individuals to tailor their testing experiences to fit their unique needs and circumstances, and patient satisfaction [71]. For instance, digital HIV self-testing strategies have been shown to mitigate barriers such as stigma associated with in-person testing, while providing users with control over the testing process, including the option to involve, or exclude, others during testing [71-73]. However, challenges remain in ensuring that these tools are accessible to all. These findings call for future research to optimize HIV self-testing methods.

While this review highlights the potential of digital innovations to increase accuracy rates, robust studies are required to comprehensively understand the potential benefits and limitations of digital innovations in enhancing the accuracy of HIV self-testing.

HIV self-testing with digital tools is the next step in the armamentarium in our fight against HIV. These tools are best suited when certain conditions are met, including the following:

1. Digital connectivity and tools to avail it are established and are cheap or relatively affordable, depending on the context and the country setting.
2. Populations using these solutions are digitally literate and health literate and are able to navigate care using these solutions.
3. The pathways to clinical care are made navigable and are often facilitated by a corpus of healthcare providers who can offer assistance to those who call, chat, and engage with them.
4. Dashboards and platforms that are set up are often interoperable and can work with country-level network architecture of electronic records and telemedicine that are in place.
5. Linkage services can be set up easily with these solutions and populations identified with infections can be seamlessly integrated within care pathways.
6. Countries have signed on to the treaty on digital health and are aligned with WHO's vision of One Health.
7. Linkage to treatment and retention in care are essential for these tools to be optimally effective, not just screening.
8. A corpus of health care counselors and testers have been trained in the use of digital solutions and through training, have become well-equipped to identify the challenges of navigating care pathways and are ready to help task shift to decrease the pain experience in navigating linkage to care and counseling by testers.

Conclusion

Digital supports enhance HIV self-testing across multiple domains, including improving test accuracy or self-test interpretation and achieving high metrics of acceptability, preference, feasibility, and impact. They have the potential to improve test accuracy by providing clear instructions, result interpretation, and data collection mechanisms, promoting reliable and trustworthy results; however, further evidence is needed in this space.

Digital interventions positively impact acceptability and preference due to their convenience, privacy, and autonomy. Furthermore, they improve the feasibility of HIV self-test, by leveraging technology to provide easy access to testing and result reporting, eliminating the need for physical visits to testing centers or clinics, especially benefitting individuals with time constraints, limited mobility, or those residing in remote areas. Finally, digital supports elevate impact by increasing result returns, attracting first-time testers, identifying new infections, and aiding linkages to care. Leveraging technology's widespread use, digital interventions extend HIV self-test reach, bridging testing coverage gaps, though more research and implementation data are needed in low-income settings and among key marginalized populations affected by HIV. These populations and settings traditionally lack access to connectivity tools, which may impact equity and access to services. Despite challenges, integrating digital tools aligns with health care's evolution, offering the potential to revolutionize HIV testing, ultimately leading to better health outcomes for individuals and communities. These interventions bode well to help us achieve UNAIDS 95-95-95 targets, especially for high-risk HIV populations residing in 191 middle- to high-income countries, who constitute a majority of the world's population.

Acknowledgments

NPP acknowledges the support of the Fonds de recherche du Québec-Santé (Distinguished Research Scholar Award 324154). The agency played no role in the study design, data collection, data analysis, interpretation, or decision to submit for publication. The authors would like to thank Melisa Eraslan for her assistance with proofreading and formatting the manuscript. No use of generative Artificial Intelligence was used in any portion of this manuscript.

Canadian Institutes of Health Research (grants 185723 and 192204). NPP also acknowledges the support by other grants during this period including the India-Canada Centre for Innovative Multidisciplinary Partnerships to Accelerate Community Transformation and Sustainability (IC-IMPACTS; CNO 3072), NIH Portent grant (grant U54EB034654) and the Canadian Institutes of Health Research (grant PJT153149).

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published article (and Multimedia Appendices 1 and 2).

Authors' Contributions

AB and NPP conducted conceptualization responsible for formulating and evolving the overarching research goal and aims. AB, OV, and NPP managed data curation and software analysis. AB and OV conducted the statistical, mathematical, and formal analysis of the data. NPP secured the funding of financial support for this project. AB conducted the research investigation. NPP created the methodology of the review. AB was responsible for project administration the management and coordination of research. NPP provided the resources required for the execution of this project. NPP generously and successfully provided leadership and support to AB for this supervision project. NPP was responsible for the verification of all processes for this review. AB led the preparation and creation of the review outline and structure.
AB and NPP designed, Writing-original drafted, and reviewed the initial manuscript. AB, OV, NE, TD, CL, and NPP provided critique on subsequent drafts, as well as reviewed and approved the final written manuscript.

Conflicts of Interest

NPP reports an open access application for HIV self-testing: “HIVSmart!” McGill University Copyright 1105598, 2013. This digital innovation was funded by Grand Challenges Canada (0027-01), and is co-owned by Grand Challenges Canada and McGill University.

Multimedia Appendix 1
Extended information.
[DOCX File , 284 KB - jmir_v27i1e63110_app1.docx]

Multimedia Appendix 2
PRISMA Checklist.
[DOCX File , 31 KB - jmir_v27i1e63110_app2.docx]

References

1. Fischer AE, Abrahams M, Shankland L, Lalla-Edward ST, Edward VA, De Wit J. The evolution of HIV self-testing and the introduction of digital interventions to improve HIV self-testing. Front Reprod Health 2023;5:1121478 [FREE Full text] [doi: 10.3389/frph.2023.1121478] [Medline: 36864847]
2. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. Geneva: World Health Organization; 2016:144.
3. Digital health. World Health Organization. 2020. URL: https://www.who.int/health-topics/digital-health#tab=tab_1 [accessed 2025-01-28]
4. Pai N, Esmail A, Saha Chaudhuri P, Oelofse S, Pretorius M, Marathe G, et al. Impact of a personalised, digital, HIV self-testing app-based program on linkages and new infections in the township populations of South Africa. BMJ Glob Health 2021;6(9):e006032 [FREE Full text] [doi: 10.1136/bmjgh-2021-006032] [Medline: 34475026]
5. McGuire M, de Waal A, Karellis A, Janssen R, Engel N, Sampath R, et al. HIV self-testing with digital supports as the new paradigm: a systematic review of global evidence (2010-2021). EClinicalMedicine 2021;39:101059 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101059] [Medline: 34430835]
6. Beecroft AM, Duchaine T, Engel N, Liang C, Esmail A, Dheda K. 1513. Evaluating the incremental accuracy of HIV self-test together with an app-based solution: a secondary data analysis of trial data. Open Forum Infect Dis 2023;10(Supplement_2):ofad500.1348.
7. Stevens DR, Vrana CJ, Dlin RE, Korte JE. A global review of HIV self-testing: themes and implications. AIDS Behav 2018;22(2):497-512 [FREE Full text] [doi: 10.1007/s10461-017-1707-8] [Medline: 28155039]
8. Figueroa C, Johnson C, Verster A, Baggaley R. Attitudes and acceptability on HIV self-testing among key populations: a literature review. AIDS Behav 2015;19(11):1949-1965 [FREE Full text] [doi: 10.1007/s10461-015-1097-8] [Medline: 26054390]
9. Facts About In-Home HIV Testing. FDA. 2023. URL: https://www.fda.gov/consumers/consumer-updates/facts-about-home-hiv-testing [accessed 2025-01-28]
10. Doan T, Stafylis C, Wang Q, Vavala G, Lemley S, McLeman B, et al. Accuracy of interpretation and home test kit result reporting for screening of human immunodeficiency virus infection. Sex Transm Infect 2021;97(Suppl 1):A83.2-A8A83.
11. Kwan TH, Chan DPC, Wong SYS, Lee SS. Implementation cascade of a social network-based HIV self-testing approach for men who have sex with men: cross-sectional study. J Med Internet Res 2023;25:e46514 [FREE Full text] [doi: 10.2196/46514] [Medline: 37099364]

XSL•FO
RenderX

12. O'Byrne P, Musten A, Orser L, Horvath C. Invalid results in the GetaKit study in Ottawa: a real-world observation of the INSTI® HIV self-test among persons at risk for HIV. J Assoc Nurses AIDS Care 2022;33(5):567-573 [FREE Full text] [doi: 10.1097/JNC.0000000000000335] [Medline: 35472138]
13. Wong HB, Lim GH. Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. Proc Singap Healthc 2011;20(4):316-318. [doi: 10.1177/201010581102000411]
14. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
15. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
16. Higgins JPT, Thomas J, Chandler J. Cochrane handbook for systematic reviews of interventions. Hoboken, New Jersey: John Wiley & Sons; 2019.
17. Wells GA, Shea B, O'Connell DA. The newcastle-ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Oxford, England: Oxford; 2000.
18. Brady M, Carpenter G, Bard B. Self-testing for HIV: Initial experience of the UK's first kit. HIV Med 2016;17(Supplement 1):9.
19. Wang X, Tang Z, Wu Z, Nong Q, Li Y. Promoting oral HIV self-testing via the internet among men who have sex with men in China: a feasibility assessment. HIV Med 2020;21(5):322-333 [FREE Full text] [doi: 10.1111/hiv.12830] [Medline: 31849177]
20. Roche SD, Ekwunife OI, Mendonca R, Kwach B, Omollo V, Zhang S, et al. Measuring the performance of computer vision artificial intelligence to interpret images of HIV self-testing results. Front Public Health 2024;12:1334881 [FREE Full text] [doi: 10.3389/fpubh.2024.1334881] [Medline: 38384878]
21. Birdthistle I, Mulwa S, Sarrassat S, Baker V, Khanyile D, O'Donnell D, et al. Effects of a multimedia campaign on HIV self-testing and PrEP outcomes among young people in South Africa: a mixed-methods impact evaluation of 'MTV Shuga Down South'. BMJ Glob Health 2022;7(4):e007641 [FREE Full text] [doi: 10.1136/bmjgh-2021-007641] [Medline: 35365480]
22. Chan PSF, Chidgey A, Lau J, Ip M, Lau JTF, Wang Z. Effectiveness of a novel HIV self-testing service with online real-time counseling support (HIVST-Online) in increasing HIV testing rate and repeated HIV testing among men who have sex with men in Hong Kong: results of a pilot implementation project. Int J Environ Res Public Health 2021;18(2):729 [FREE Full text] [doi: 10.3390/ijerph18020729] [Medline: 33467770]
23. Fischer AE, Phatsoane M, Majam M, Shankland L, Abrahams M, Rhagnath N, et al. Uptake of the Ithaka mobile application in Johannesburg, South Africa, for human immunodeficiency virus self-testing result reporting. South Afr J HIV Med 2021;22(1):1197 [FREE Full text] [doi: 10.4102/sajhivmed.v22i1.1197] [Medline: 33824737]
24. Girault P, Misa Wong C, Jittjang S, Fongkaew K, Cassell MM, Lertpiriyasuwat C, et al. Uptake of oral fluid-based HIV self-testing among men who have sex with men and transgender women in Thailand. PLoS One 2021;16(8):e0256094 [FREE Full text] [doi: 10.1371/journal.pone.0256094] [Medline: 34398926]
25. Kaneko N, Sherriff N, Takaku M, Vera JH, Peralta C, Iwahashi K, et al. Increasing access to HIV testing for men who have sex with men in Japan using digital vending machine technology. Int J STD AIDS 2022;33(7):680-686 [FREE Full text] [doi: 10.1177/09564624221094965] [Medline: 35502984]
26. Maraj D, Ly A, Lo HTJ, Zubowski L, Watson J, McBain K, editors. I'm Ready, Talk: Implementation of a peer navigator program to facilitate HIV self- testing and linkage to care in Canada. Quebec City, Canada: Canadian Association for HIV Research; 2023.
27. Marley G, Fu G, Zhang Y, Li J, Tucker JD, Tang W, et al. Willingness of Chinese men who have sex with men to use smartphone-based electronic readers for HIV self-testing: web-based cross-sectional study. J Med Internet Res 2021;23(11):e26480 [FREE Full text] [doi: 10.2196/26480] [Medline: 34806988]
28. Ntinga X, Musiello F, Keter AK, Barnabas R, van Heerden A. The feasibility and acceptability of an mHealth conversational agent designed to support HIV self-testing in South Africa: cross-sectional study. J Med Internet Res 2022;24(12):e39816 [FREE Full text] [doi: 10.2196/39816] [Medline: 36508248]
29. Phatsoane Gaven M, Quaife M, Majam M, Singh L, Rhagnath N, Wonderlik T, et al. HIV self-test reporting using mHealth platforms: a pilot study in Johannesburg, South Africa. Front Reprod Health 2023;5:1073492 [FREE Full text] [doi: 10.3389/frph.2023.1073492] [Medline: 36923466]
30. Ramos SR, Lardier DT, Bond KT, Boyd DT, O'Hare OM, Nelson LE, et al. Participatory design of a web-based HIV Oral self-testing infographic experiment (HOTIE) for emerging adult sexual minority men of color: a mixed methods randomized control trial. Int J Environ Res Public Health 2021;18(22):11881 [FREE Full text] [doi: 10.3390/ijerph182211881] [Medline: 34831644]
31. Rosadiño JDT, Pagtakhan RG, Briñes MT, Dinglasan JLG, Cruz DP, Corciega JOL, et al. Implementation of unassisted and community-based HIV Self-Testing (HIVST) during the COVID-19 pandemic among Men-who-have-sex-with-Men (MSM) and Transgender Women (TGW): a demonstration study in Metro Manila, Philippines. PLoS One 2023;18(3):e0282644 [FREE Full text] [doi: 10.1371/journal.pone.0282644] [Medline: 36893109]

XSL•FO
RenderX

32. McFall A, Thakker J, Loeb T, Bell J, Singh A, Pollard R. Virtual support improves client experiences with an online HIV self-testing service. Top. antivir. med 2023;31(2):373.
33. Oladele DA, Iwelunmor J, Gbajabiamila T, Obiezu-Umeh C, Okwuzu JO, Nwaozuru U, et al. An unstructured supplementary service data system to verify HIV self-testing among nigerian youths: mixed methods analysis of usability and feasibility. JMIR Form Res 2023;7:e44402 [FREE Full text] [doi: 10.2196/44402] [Medline: 37747780]
34. Shrestha R, Altice FL, Khati A, Azwa I, Gautam K, Gupta S, et al. Clinic-integrated smartphone app (JomPrEP) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: mixed methods evaluation of usability and acceptability. JMIR Mhealth Uhealth 2023;11:e44468 [FREE Full text] [doi: 10.2196/44468] [Medline: 36795465]
35. Bell SFE, Lemoire J, Debattista J, Redmond AM, Driver G, Durkin I, et al. Online HIV Self-Testing (HIVST) dissemination by an Australian community peer HIV organisation: a scalable way to increase access to testing, particularly for suboptimal testers. Int J Environ Res Public Health 2021;18(21):11252 [FREE Full text] [doi: 10.3390/ijerph182111252] [Medline: 34769771]
36. Ly A, Lo Hog Tian JM, Zubowski L, Maraj D, McBain D, Chan P, editors. I'm Ready, Test: Leveraging a mobile health app and program to reduce barriers and facilitate access to HIV self-testing in Canada to reach the undiagnosed. Quebec City, Canada: Canadian Association for HIV Research; 2023.
37. Vasconcelos R, Avelino-Silva VI, de Paula IA, Jamal L, Gianna MC, Santos F, et al. HIV self-test: a tool to expand test uptake among men who have sex with men who have never been tested for HIV in São Paulo, Brazil. HIV Med 2022;23(5):451-456 [FREE Full text] [doi: 10.1111/hiv.13178] [Medline: 34634182]
38. Howells W, Winter A, Baraitser P. A comparison of service user uptake of selftesting versus self-sampling during England's National HIV testing week 2023 by demographic group and risk profile. Sex. Transm. Infect 2023;99(Supplement 1):A42.2-A4A43.
39. Larsson L, Chikwari CD, McHugh G, Koris A, Bandason T, Dauya E, et al. Feasibility and usability of mobile technology to assist HIV self-testing in youth in Zimbabwe: a mixed-methods study. J Adolesc Health 2023;73(3):553-560 [FREE Full text] [doi: 10.1016/j.jadohealth.2023.05.011] [Medline: 37389521]
40. Guadamuz TE, Samoh N, Lim SH, Jonas KJ, editors. A pilot feasibility study of HIV self-testing with online supervision among MSM who attend sexualized drug parties in Thailand. Montreal, Canada: International AIDS Society; 2022.
41. Jin X, Lu T, Gao Y, Mao X, Chu Z, Zhang J, editors. Effect of a regular self-testing kits distribution intervention on HIV testing and sexual-behavioral outcomes among MSM in China: a stepped-wedge randomized controlled trial. Geneva, Switzerland: International AIDS Society; 2021.
42. Johnson MC, Chung R, Leung SYJ, Edelstein Z, Yuan Y, Flavin SM. Combating stigma through HIV self-testing: New York State's HIV home test giveaway program for sexual minorities. J Public Health Manag Pract 2022;28(2):174-183 [FREE Full text] [doi: 10.1097/PHH.0000000000001138] [Medline: 32015253]
43. Mshweshwe-Pakela NT, Mabuto T, Shankland L, Fischer A, Tsukudu D, Hoffmann CJ. Digitally supported HIV self-testing increases facility-based HIV testing capacity in Ekurhuleni, South Africa. South Afr J HIV Med 2022;23(1):1352 [FREE Full text] [doi: 10.4102/sajhivmed.v23i1.1352] [Medline: 35923609]
44. O'Byrne P, Musten A, Vandyk A, Ho N, Orser L, Haines M, et al. HIV self-testing in Ottawa, Canada used by persons at risk for HIV: the getaKit study. Can Commun Dis Rep 2021;47(10):435-441 [FREE Full text] [doi: 10.14745/ccdr.v47i10a06] [Medline: 34737676]
45. O'Byrne P, Winkelmant S, Musten A, Inamdar G, Orser L, Grayson M, editors. Mailout HIV self-testing: overview of the GetaKit initiative in Ottawa, Canada. Ottawa, Canada: International AIDS Society; 2021.
46. Pollard R, Okram S, Thakker J, Singh A, Bell J, McFall A, editors. Tailoring HIVST to local communities can improve uptake: lessons learned from a virtual HIV self-testing intervention in India. Montreal, Canada: International AIDS Society; 2022.
47. Shaikezhanov A, Gryazev D, Terlikbayeva T, Primbetova S, Wu E, Hunt H, editors. Preliminary results of AmanBol: HIV self-testing program for MSM and transgender persons in Kazakhstan. Geneva, Switzerland: International AIDS Society; 2021.
48. Stafylis C, Vavala G, Wang Q, McLeman B, Lemley SM, Young SD, et al. Relative effectiveness of social media, dating apps, and information search sites in promoting HIV self-testing: observational cohort study. JMIR Form Res 2022;6(9):e35648 [FREE Full text] [doi: 10.2196/35648] [Medline: 36149729]
49. Thakker JH, Singh A, Pollard R, Bell J, McFall AM, Taduri M. HIV self-testing uncovers high burden of hidden infections in INDIA. Top. antivir. med 2022;30(1 SUPPL):55.
50. Vlasiuk O, Shevchuk D, Gavrysh L, editors. Introduction of oral-based HIV self-testing in Ukraine through the national-scale HealthLink Project: distribution strategies and key results. Montreal, Canada: International AIDS Society; 2022.
51. Young SD, Cumberland WG, Singh P, Coates T. A peer-led online community to increase HIV self-testing among African American and latinx MSM: a randomized controlled trial. J Acquir Immune Defic Syndr 2022;90(1):20-26 [FREE Full text] [doi: 10.1097/QAI.0000000000002919] [Medline: 35044989]

XSL•FO
RenderX

52. Zhou Y, Lu Y, Ni Y, Wu D, He X, Ong JJ, et al. Monetary incentives and peer referral in promoting secondary distribution of HIV self-testing among men who have sex with men in China: a randomized controlled trial. PLoS Med 2022;19(2):e1003928 [FREE Full text] [doi: 10.1371/journal.pmed.1003928] [Medline: 35157727]
53. Kaptchuk RP, Thakker J, Bell J, Okram S, Gopinath U, Mehta SH, et al. HIV self-testing in India: implementation and qualitative evaluation of a web-based programme with virtual counsellor support. J Int AIDS Soc 2024;27(6):e26302 [FREE Full text] [doi: 10.1002/jia2.26302] [Medline: 38861458]
54. Logie CH, Okumu M, Berry I, Hakiza R, Baral SD, Musoke DK, et al. Findings from the Tushirikiane mobile health (mHealth) HIV self-testing pragmatic trial with refugee adolescents and youth living in informal settlements in Kampala, Uganda. J Int AIDS Soc 2023;26(10):e26185 [FREE Full text] [doi: 10.1002/jia2.26185] [Medline: 37850816]
55. Nguyen VTT, Dunkley Y, Son VH, Choko AT, Huong PTT, Manh PD, et al. Investigating the effectiveness of web-based HIV self-test distribution and linkage to HIV treatment and PrEP among groups at elevated risk of HIV in viet nam provinces: a mixed-methods analysis of implementation from pilot to scale-up. J Int AIDS Soc 2024;27 Suppl 1:e26264. [doi: 10.1002/jia2.26264] [Medline: 38965974]
56. Ni Y, Lu Y, Jing F, Wang Q, Xie Y, He X, et al. A machine learning model for identifying sexual health influencers to promote the secondary distribution of HIV self-testing among gay, bisexual, and other men who have sex with men in China: quasi-experimental study. JMIR Public Health Surveill 2024;10:e50656 [FREE Full text] [doi: 10.2196/50656] [Medline: 38656769]
57. O'Byrne P, Kroch A, Orser L, Ho N, Musten A, Haines M, et al. Targeted HIV self-testing identifies persons with undiagnosed HIV and active engagement links them to care: the getakit study. AIDS Behav 2024;28(6):2015-2022 [FREE Full text] [doi: 10.1007/s10461-024-04302-5] [Medline: 38526637]
58. Shrestha RK, Hecht J, Chesson HW. Analyzing the costs and impact of the TakeMeHome program, a public-private partnership to deliver HIV self-test kits in the United States. J Acquir Immune Defic Syndr 2024;95(2):144-150. [doi: 10.1097/QAI.0000000000003323] [Medline: 37831623]
59. Zhou H, Zhu Y, Gao Y, Chu Z, Chen S, Liu M, et al. Online distribution of HIV self-testing kits to promote HIV testing among men who have sex with men discontinuing pre-exposure prophylaxis after demonstration project completion in China: a multicentre open-label randomized controlled trial. Lancet Reg Health West Pac 2023;41:100922 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100922] [Medline: 37867621]
60. Zhu Z, Lu X, Gao P, Wang X, Hu X, Xie N, et al. Feasibility of a mobile health intervention for providing a continuum of HIV services for MSM: pilot study of the WeTest program in 3 cities in China. Curr HIV Res 2024;22(3):158-169. [doi: 10.2174/011570162X280190240105063449] [Medline: 38310468]
61. O'Byrne P, Musten A, McCready L, Robinson R, Durrant G, Tigert J, et al. HIV self-testing enabled access to testing for black persons: the GetaKit study. Res Nurs Health 2023;46(2):236-241. [doi: 10.1002/nur.22293] [Medline: 36573589]
62. O'Byrne P, Musten A, Ho H, Lindsay J, Orser L, editors. The GetaKit Study: Bridging the gap in undiagnosed HIV infections using at-home HIV self-testing kits. Quebec City, Canada: Canadian Association for HIV Research; 2023.
63. Rourke S, Owino M, Tharao W, Lachowski N, Duddy J, Ly A, editors. Effectiveness of the I'm Ready Program to provide low barrier access across Canada to HIV self-testing to reach first-time testers. Quebec City, Canada: Canadian Association for HIV Research; 2023.
64. Li S, Zhang J, Mao X, Lu T, Gao Y, Zhang W, et al. Feasibility of indirect secondary distribution of HIV self-test kits via wechat among men who have sex with men: national cross-sectional study in China. J Med Internet Res 2021;23(10):e28508 [FREE Full text] [doi: 10.2196/28508] [Medline: 34698651]
65. Ni Y, Lu Y, Zhou Y, Tang W. Using social media and social network to expand HIV self-testing in china. Sex. Transm. Infect 2021;97(SUPPL 1):A12.
66. Tharao W, Owino M, Aden M, Soje L, Keen S, Frans D. I'm Ready Program 6-Month Results: African, Caribbean and Black participants accessing HIV self-testing across Canada. Halifax, Nova Scotia, Canada: Canadian Association for HIV Research; 2022.
67. MacGowan RJ, Chavez PR, Dana R, Hannah M, Raiford JL, Caldwell JA, et al. Efficacy of internet recruitment and HIV self-testing for diagnosing HIV infections among black and hispanic/latino MSM and transgender women in 11 US States, 2020-2021. J Acquir Immune Defic Syndr 2024;97(2):133-141. [doi: 10.1097/QAI.0000000000003476] [Medline: 39250647]
68. Njau B, Damian DJ, Abdullahi L, Boulle A, Mathews C. The effects of HIV self-testing on the uptake of HIV testing, linkage to antiretroviral treatment and social harms among adults in Africa: a systematic review and meta-analysis. PLoS One 2021;16(1):e0245498 [FREE Full text] [doi: 10.1371/journal.pone.0245498] [Medline: 33503050]
69. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
70. Welcome to the GRADE working group. 2024. URL: https://www.gradeworkinggroup.org/ [accessed 205-01-28]
71. Janssen R, Engel N, Esmail A, Oelofse S, Krumeich A, Dheda K, et al. Alone but supported: a qualitative study of an HIV self-testing app in an observational cohort study in South Africa. AIDS Behav 2020;24(2):467-474 [FREE Full text] [doi: 10.1007/s10461-019-02516-6] [Medline: 31049809]

XSL•FO
RenderX

72. Janssen R, Engel N, Pant Pai N, Esmail A, Dheda K, Thomas R, et al. 'You're only there on the phone'? a qualitative exploration of community, affect and agential capacity in HIV self-testing using a smartphone app. Sociol Health Illn 2021;43(3):591-606 [FREE Full text] [doi: 10.1111/1467-9566.13242] [Medline: 33634889]
73. Janssen R, Krumeich A, Esmail A, Thomas R, Dheda K, Pant Pai N, et al. Moments of uncertainty: exploring how an app-based oral HIV self-testing strategy fits in communities 'Living Under' HIV risk. Med. Anthropol. Theory 2021;8(2):1-24.

Abbreviations

ART: anti-retroviral therapy
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
FDA: Food and Drug Administration
GRADE: Grading of Recommendation Assessment, Development and Evaluation
MSM: men who have sex with men
NOS: Newcastle-Ottawa Scale
NPV: negative predictive value
OR: odds ratio
PPV: positive predictive value
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2: quality of diagnostic accuracy studies
RCT: randomized controlled trials
RoB2: Cochrane Risk of Bias Tool 2
RR: risk ratio
UNAIDS: Joint United Nations Program on HIV/AIDS
WHO: World Health Organization

Edited by A Coristine; submitted 10.06.24; peer-reviewed by J Mitchell, T Lillie; comments to author 01.10.24; revised version received 28.11.24; accepted 05.12.24; published 04.03.25.

Please cite as:
Beecroft A, Vaikla O, Engel N, Duchaine T, Liang C, Pant Pai N
Evidence on Digital HIV Self-Testing From Accuracy to Impact: Updated Systematic Review
J Med Internet Res 2025;27:e63110
URL: https://www.jmir.org/2025/1/e63110
doi:10.2196/63110
PMID:40053740

©Ashlyn Beecroft, Olivia Vaikla, Nora Engel, Thomas Duchaine, Chen Liang, Nitika Pant Pai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Utilization and Experiences of Using Quit Now, a Nicotine and Tobacco Smoking Cessation Website: Thematic Analysis

Tala Salaheddin[1], BSc (Hons); Ramona H Sharma[2], MSW; Marcela Fajardo[3], BA; Cameron Panter[3], MA; Lauren De Souza[3], MPH; Sheila Kanyevu Matano[3], MPH, CPH; Laura Struik[4], RN, PhD

[1]Doctor of Medicine Program, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[2]School of Social Work, Faculty of Health and Social Development, University of British Columbia Okanagan, Kelowna, BC, Canada

[3]BC Lung Foundation, Vancouver, BC, Canada

[4]School of Nursing, Faculty of Health and Social Development, University of British Columbia Okanagan, Kelowna, BC, Canada

Corresponding Author:
Tala Salaheddin, BSc (Hons)
Doctor of Medicine Program
Faculty of Medicine
University of British Columbia
2329 West Mall
Vancouver, BC, V6T 1Z4
Canada
Phone: 1 6048222211
Email: talasala@student.ubc.ca

Abstract

Background: British Columbia residents have access to a program called QuitNow that provides behavioral support and information about pharmacotherapy to nicotine and tobacco users. Web- or computer-based smoking cessation programs have been shown to yield an abstinence rate about 1.5 times higher when compared to a control. Although quantitative evidence reveals significant promise for web-based services like QuitNow, there is very little qualitative evidence available. Understanding website utilization and the experiences of end users is key to contextualizing the effectiveness of web-based cessation services and providing directions for enhancing these services.

Objective: This qualitative interview study aims to delve into users' utilization and experiences of QuitNow, which is supplemented by Google Analytics data.

Methods: We interviewed 10 QuitNow users using semistructured interviews to understand what they liked the most and the least about QuitNow. We transcribed these interviews and conducted an inductive thematic analysis using NVivo (QSR International) software to extract common themes about user experiences. We also gathered utilization metrics via Google Analytics (n=13,856 users) to understand which aspects of QuitNow were used the most and which were used the least during the study period.

Results: Thematic analysis yielded four major themes: (1) barriers to information access reduce opportunities to take action, (2) lack of clarity around pharmacological options is discouraging, (3) hearing from others is an important part of the journey, and (4) recognizing own agency throughout the quit process. These themes provided context and support for the Google Analytics data, which showed that end user activity, measured by indicators such as page views and average time spent on each page, was highest on pages about how to quit (10,393 page views), pharmacology information (1999 page views), and the community forum (11,560 page views).

Conclusions: Results of this study point to several important implications for improving the website, as well as directions for enhancing cessation support services in general.

(J Med Internet Res 2025;27:e55592) doi:10.2196/55592

KEYWORDS

smoking cessation; user experiences; nicotine; vaping; web-based; Google Analytics; thematic analysis; digital health; nicotine replacement therapy; quit attempts; tobacco; British Columbia; behavioral support; pharmacotherapy; qualitative interview; cessation support; QuitNow; mobile health; mHealth; intervention

XSL•FO
RenderX

Introduction

Smoking, a leading cause of preventable deaths worldwide, poses a significant challenge for individuals attempting to quit. The prevalence of current cigarette smoking among Canadians aged 15 years and older in 2022 was 10.9%, unchanged since 2020, but decreased from 12% in 2019 [1,2]. Notably, in 2020, British Columbia had the lowest smoking rate among Canadian provinces, at 7.7% [3], and continues to boast the lowest smoking rate of 8.7% among all Canadian provinces in 2022. Similarly, vaping rates are also lowest in British Columbia, with a past 30-day vaping rate of 4.7% in those aged 15 years and older [2]. This achievement can be attributed in part to the British Columbia government's initiatives, such as the British Columbia Smoking Cessation Program, including QuitNow, which supports residents in covering the costs of Nicotine Replacement Therapy (NRT) and smoking cessation prescription drugs and provides those attempting to quit a variety of nicotine products with access to community and expert coaches.

Nearly two-thirds of Canadians who smoke express intentions to quit in the next 6 months and half report that they have tried to quit in the past year [4]. Of Canadians aged 15 years and older who report daily vaping, almost half (42.4%) have made at least one quit attempt in the last 12 months [2]. However, unaided quit attempts have a very low success rate, with fewer than 5% achieving abstinence for 12 months or more [5]. To achieve meaningful reductions in tobacco-related morbidity and mortality, we must increase both the number of individuals who attempt to quit and the success rate of those attempts. Guidelines recommend combining pharmacotherapies with behavioral interventions to enhance the effectiveness of smoking cessation attempts [6,7]. Web-based programs have emerged as a valuable tool in this context. A meta-analysis of 22 randomized controlled trials has demonstrated the clinical efficacy of web-based smoking cessation programs for adults who smoke [8]. This study included 29,549 participants and revealed that web- or computer-based smoking cessation programs yielded an abstinence rate about 1.5 times higher than the control group, and this was observed at both 6 and 12 months of follow-up [8].

Today, 5.18 billion people (two-thirds of the global population) are currently connected to the World Wide Web [9]. From 2000 to 2023, web usage grew more than 1000%, and this number is predicted to continue to grow [10,11]. In recent years, the expansion of web-based systems in Canadian health care has notably enhanced service delivery and patient engagement. For instance, the number of Canadian primary care physicians using electronic medical records has risen from 73% in 2015 to 93% in 2022, facilitating better patient care through improved information sharing between providers, such as lab results and referrals [12]. The COVID-19 pandemic has also led to a rise in the usage of web-based health care, consequently shifting nicotine usage and cessation needs. As the number of global web users grows and the internet becomes increasingly embedded in health care systems, web-based cessation programs are poised to become a more valuable public health strategy [13]. Ensuring the quality and effectiveness of these programs will be crucial in providing accessible support to nicotine and tobacco users worldwide.

QuitNow, a free web-based cessation program provided by the British Columbia Lung Foundation on behalf of the Government of British Columbia, offers information on nicotine and its harmful effects, access to an online community, and one-on-one guidance for individuals seeking to quit or reduce tobacco and e-cigarette use [14]. While previous research has explored the overall impact of online cessation support, there is limited knowledge about the in-depth experiences of end users and how their experiences contextualize specific utilization data [15].

Web-based cessation programs, such as QuitNow, are designed to provide tailored support and resources to individuals seeking to quit smoking or using other nicotine products. The effectiveness of these platforms is heavily dependent on how users interact with the content and features provided. For instance, user engagement metrics, such as page views, time spent on specific sections, and interaction with multimedia elements, offer valuable insights into which aspects of the platform are most effective or need improvement. By analyzing these interactions, developers can make data-driven decisions to enhance the platform's usability, accessibility, and overall impact.

This study aims to address the gap in knowledge by investigating the in-depth experiences of individuals who used QuitNow as a quit aid and examining how these experiences are reflected in utilization data, such as that provided by Google Analytics. By understanding end user experiences better, we can tailor our efforts to support British Columbia residents in reducing nicotine and tobacco use, thereby improving overall health outcomes.

Methods

Participants

Interview data from 10 individuals who had used QuitNow were included in this study. These participants were part of a larger study that aimed to understand the perceptions and experiences of QuitNow among individuals who use a variety of nicotine products. For this study, we examined data only from participants who reported using QuitNow to support their cessation (vs participants who had not used QuitNow) and included participants who had recently quit or were motivated to quit, either smoked, vaped, or both, were aged 16 years or older, lived in British Columbia, and could communicate in English. The recruitment procedures and inclusion and exclusion criteria are described in a previous study [16].

Out of the 10 participants, 8 were interested in quitting smoking or vaping, and 2 had recently quit (one quit smoking, one quit vaping). The mean age of participants was 34.5 (SD 17-55) years; 70% (7/10) were male, 50% (5/10) were Caucasian, 80% (8/10) smoked cigarettes, and 60% (6/10) used e-cigarettes. Most individuals who used traditional cigarettes (7/8, 88%) had been smoking for over 5 years, while 100% (6/6) of individuals who used e-cigarettes had been vaping for less than 5 years. Of all participants, 40% (4/10) only used traditional cigarettes, 30% (3/10) used both traditional and e-cigarettes, and the remaining 30% (3/10) had only used e-cigarettes.

XSL•FO
RenderX

Data Collection

All consenting participants filled out a brief demographic survey encompassing personal characteristics (eg, gender, age) and information about their smoking and vaping history. Participants then engaged in an approximately 60-minute, audio-recorded, semistructured interview via University of British Columbia Zoom. Interviews began with introductions and a reminder about the study's purpose and procedures. After participants answered questions about how they heard about QuitNow and what motivated them to use QuitNow, they were then asked 5 primary questions, including what they liked about QuitNow, what they did not like about QuitNow, suggestions for improving QuitNow, what aspects outside of QuitNow supported their quit journey, and their "ideal world" vision for how QuitNow could help them with quitting in the context of their lives. After interview completion, each participant received a CAD $50 (US $34.97) electronic gift card to thank them for their contribution to the study. Interview recordings were transcribed for analysis.

Data Analysis

We conducted an inductive thematic analysis of 10 interview transcripts following guidelines from Willig et al [17]. The interviews were first transcribed verbatim and uploaded to the qualitative data analysis software program NVivo (QSR International). The analysis process began with two researchers (TS and LS) reading the transcripts while making casual observational notes to engage with and familiarize themselves with the data. After a detailed reading of the transcripts, the researchers began to generate codes and systematically label and synthesize segments of the dataset to reduce and organize it. For example, when participants described barriers of the website for supporting their cessation, the researchers inserted these data under the larger code "barriers to quitting," which then contained specific sublabels for these barriers (eg, "too much text to process;" "can't access the full site without providing personal details"). We did the same with "facilitators to quitting." Data saturation was achieved, whereby we found no new interview data emerged to answer our research question after the sixth interview. Following this, the researchers iteratively developed a coding framework using interview data. One author (TS) then sorted associated tagged data segments according to the developed codes, which were revised or added to as new data were collected. The researchers then examined the codes and their associated data and clustered them to construct themes from a well-rounded understanding of the transcripts. For example, when examining codes around website-based barriers and facilitators to quitting, we recognized overall themes. For example, while participants talked about how too much text limited their ability to really take something away from a certain page and apply it in practice, they also talked about how graphic-based information enabled them to learn and apply ideas to facilitate quitting, something that was lacking. Therefore, these codes were grouped together and formed the theme "Barriers to information access reduce opportunities to take action." These were reviewed and defined to help answer the research question. Author TS then finalized the findings and produced the report, which was reviewed by all other authors for agreement.

Google Analytics Data

To understand empirical information on public usage of the QuitNow website and related trends (eg, which pages on QuitNow were most frequently navigated to, the average duration of time spent per page), Google Analytics data from the period during which participant interviews were conducted (March 3-July 6, 2021) were also gathered. Analytics data included information on overall website usage and user reach, along with detailed web page-level metrics capturing typical user behaviors on the site (eg, number of sessions; length of sessions).

Ethical Considerations

Ethics approval for this study was obtained from the University of British Columbia (Okanagan) Behavioral Research Ethics Board (H21-00145-A006). All participants were required to provide informed consent. Before their interview, participants were provided with an emailed consent form and informed that their participation was voluntary; they then provided written consent via email. Consent to participate and to be audio-recorded was reiterated at the beginning of each interview. In addition, participants were reminded that their participation was entirely voluntary and that they could withdraw from the study at any time. Each participant received a CAD $50 (US $34.97) honorarium to thank them for their contribution to the study. To protect participant identity, participants were given an identification number to anonymize their responses. Personal information such as names and contact details was not included in the dataset and was securely stored separately from the study data. Audio recordings and transcripts were kept in password-protected files, accessible only to the research team. The datasets generated during and analyzed during this study are not publicly available due to the protection of participant identity but deidentified data are available from the senior author (LS) on reasonable request.

Results

Qualitative Themes

Overview

This study resulted in four major themes to capture the experiences of individuals who used QuitNow to support tobacco and nicotine cessation. These themes include: (1) barriers to information access reduce opportunities to take quit action, (2) lack of clarity around pharmacological options is discouraging, (3) hearing from others is an important part of the journey, and (4) recognition of one's own agency in the quit process. Some of these themes contain subthemes, which will be elaborated in the following sections.

Theme 1: Barriers to Information Access Reduce Opportunities to Take Action

Sharing Personal Information to Access Information Is a Barrier

In this study, a number of participants expressed reluctance and discontent with the process of creating an account and sharing personal information, citing various reasons for their reservations. Some of these included discussions on user privacy

XSL•FO
RenderX

and even references to other websites that allow users to be anonymous and thus more genuine. For example, participant 49 (male, an individual who smokes) said:

> *one of the benefits of Reddit is [that it] is anonymous. So people are naturally inclined to be more open.*

Participant 51 (female, an individual who smokes) believes that:

> *[the website] could be a little bit less intrusive with respect to personal information, then that might be helpful (…) I mean, there's so many issues with identity theft and security.*

Participant 17 (male, an individual who vapes) even explicitly stated that:

> *what kept [them] from really using [the website] is the fact that you have to kind of create an account.*

In fact, one participant (P73, male, an individual who dual uses) even wondered:

> *what if somebody is not ready to sign up yet, but they're interested?*

This highlights that web designers should reduce barriers to accessing the website. To address these issues, more than one participant suggested having a "guest function." In describing this feature, participant 73 (male, an individual who dual uses) said:

> *like you don't have to, like log in and have like a name, but you can like be a guest and look at all the other stuff and here are all the options that are available after you make your quit plan.*

Lack of Content Clarity and Usability Is a Barrier

Another barrier to consider is that commonly used formats, such as primarily online-based and text-heavy ones, may not support effective and inclusive engagement for all individuals. Therefore, alternative formats should be considered. These include non–web-based platforms, as well as more accessible online platforms that are user-friendly, incorporate visuals, and use clear and simple language when internet access is not a barrier. Specifically, participant 86 (female, individual who dual uses) states that:

> *a lot of people in [place] don't have internet either so it's like you could tell them about it but they can't access it so maybe a pamphlet would be better.*

To minimize challenges associated with the current online platform, participants recommended reducing the amount of text per page and allocating a tab on the website that would allow immediate access when the cravings hit. To illustrate these ideas, participant 49 (male, an individual who smokes) suggested:

> *point form [and] breaking down to one or two sentences per paragraph.*

Participant 86 (female, individual who dual uses) even brought up how cravings last around 3-5 minutes and the importance of having an easily accessible tab that gives:

> *suggestions about things that you can do […] that might be more productive so that when that craving hits then maybe it's like you have a list of like things to quickly go and do*

Not only can the text be made more comprehensible, but more visual components can be added, as alluded to by some individuals. Participant 49 (male, individual who smokes) said:

> *The information could be supplemented with images or replaced by images" while another participant agrees and describes visiting pages that have a "wall of information and if that can be divided up into more of a visual infographic [then] it would just bring the point across better.*

Theme 2: Lack of Clarity Around Pharmacological Options Is Discouraging

When talking about NRT, participants were pleased that it is provided free of charge through the British Columbia Smoking Cessation Program via Pharmacare. Participant 34 (male, individual who smokes) said:

> *It's a nice feature to have the provincial government sponsoring that.*

However, while they appreciated its availability, they also highlighted the lack of information and clarity about this program on the QuitNow website. Many participants were unaware of the availability of free NRT, how the program functions, and how much NRT they could obtain for free. For example, participant 34 (male, an individual who smokes) said:

> *I mean, I go to the grocery store and look at the patch on the shelf. And it's like, 50, $60, and, you know, for a week, and you're like, yeah, I spend that much on cigarettes or whatever. So it's a no.*

Even when some participants were aware of the program, they still faced difficulties accessing more NRT when needed and expressed frustration due to the lack of clear guidelines on obtaining and using the free NRT.

> *The only frustrating, I think, is like I guess be more information about the nicotine replacement therapy program that's offered for free in B.C. because like I ran out of mine so I couldn't get anymore and so I think like I didn't know that until I went back to get more and they said I was out and like you could only get so much in 3 months. So maybe just some more information about like specifically how to get that because I had to tell a lot of people about it too so like how to get it, how much you can get for free.* [participant 86, female, an individual who dual uses]

Theme 3: Hearing From Others Is an Important Part of the Journey

Hearing From Others Who Smoke

When describing what helped them during their quit journey, participants highlighted the importance of hearing from both current and former smokers who had successfully quit. Participant 34 (male, an individual who smokes) shared:

> *I really liked the inspirational stories about the other people quitting. I think that was really positive.*

XSL•FO
RenderX

> *Especially when you see people have [smoked] for years, and then they've given up and they're happy.*

Similarly, when reflecting on the community forum, Participant 51 (female, an individual who smokes) emphasized the value of going through the process with others:

> *I am really not alone in this journey, right. And also, when people share their experiences, it just helps us connect better. Connect better with the situation, challenges, and success.*

Participant 60 echoed this sentiment, stating:

> *It shows you that it's not just you. A lot of other people are trying to quit, and also using the app and trying hard.*

Communication From Professionals

Participants emphasized the importance of being able to connect with professionals, such as quit coaches, during their quit journey. These experts provided not only practical advice but also emotional support, offering tailored guidance that complemented the information on the QuitNow website. Participant 60 (male, individual who vapes) shared that:

> *what was really…helpful is [that] you can communicate with an agent [..] whenever you need someone to help you, whenever you have questions, whenever you anything, you can just one click.*

Beyond answering technical questions, professional interaction also helped users feel supported in moments of uncertainty or high cravings, reinforcing their motivation to quit. Participants appreciated the immediacy and personalization of the support, which made them feel they were not navigating the process alone. Some suggested that future versions of QuitNow could include more proactive communication, like scheduled check-ins or reminders, to keep users engaged and motivated throughout their quit journey.

Theme 4: Recognizing Own Agency Throughout the Quit Process

While participants provided their insights and experiences with using QuitNow to aid with their cessation, they described how important it was to recognize their own agency in the quit process. When answering the question about what the biggest factor in reducing or quitting was, one participant said:

> *I think it's myself* [participant 51, female, an individual who smokes]

Building on this, participants described how support programs, like online cessation websites, are valuable tools that can support their path toward a smoke-free life, but that the support offered must parallel the smoker's determination to quit. This is exemplified by the following participant statement:

> *But until you want to [quit], I don't know if you can [quit]. I mean, if we change it around and…we were talking about a fitness website, well, the fitness website isn't going to get you fit. It's going to give you a whole bunch of ways to get fit [and] reasons to get fit, lots of help, but it's up to you. I don't see a quit smoking website any different than that…there's no magic button that you're going to click on. And all of a sudden, you're not going to have no desire to smoke* [P39, male, individual who smokes]

Exploring Theme Interrelations

An examination of how the identified themes interrelate reveals a complex and interconnected landscape of user experiences with QuitNow. For instance, "Barriers to information access" (Theme 1) directly influence "Lack of clarity around pharmacological options" (Theme 2). When users face difficulties in accessing or understanding information due to privacy concerns or inadequate dissemination strategies, their ability to fully grasp and use pharmacological aids, such as NRT, is compromised. This lack of clarity can diminish the perceived value of such options, thereby impacting users' motivation to engage with them. In addition, "Recognition of own agency" (Theme 4) underscores the significance of individual determination in the quit process, which can be either supported or undermined by the other themes. For example, overcoming barriers to access (Theme 1) and achieving clarity on pharmacological options (Theme 2) can empower users to take more control over their quit journey, reinforcing their sense of agency. Conversely, insufficient support and unclear information may erode this sense of agency, demonstrating how these themes collectively shape and influence the quit process. By understanding these interconnections, we gain a more nuanced perspective on how different factors interplay to affect user experiences and outcomes in tobacco and nicotine cessation.

Google Analytics Findings

User Characteristics

Over the period of study, QuitNow's web platform was accessed by a total of 13,856 clients across British Columbia. Of these, 1723 were active users of QuitNow digital interventions, a collection of web-delivered services designed to support nicotine cessation goals. By contrast, the provincial Quitline hosted at QuitNow supported a total of 1178 clients over this same period, although it must be noted that we are unable to test for overlap between clientele accessing digital interventions and those contacting the provincial Quitline for support. The QuitNow web platform tended to serve younger clients than the traditional Quitline service, with an average age of 32.8 over the period of study relative to 54.9 among Quitline callers. Both service lines saw a similar distribution of genders, serving approximately 58% female, 40% male, and 2% nonbinary clients.

User Behavior

Of the 66 unique pages that make up the quitnow.ca platform, Tables 1 and 2 identify the 10 most- and the 5 least-trafficked pages during our study period, respectively. Also included are the total unique user counts over the study period for each page, the bounce rate (the percentage of users who only view one page and leave without any interaction, or whose view lasts less than 10 s), average time (how long a user spent browsing a particular page), average session time (how long a user spent on the website), and pages per session (how many pages a user visited during one session).

XSL•FO
RenderX

Beyond basic administrative pages such as the homepage, account creation, and sign-up pages, the most popular features of the QuitNow website tended to be community- and information-driven. The community forum and support pages both feature among the 10 most trafficked, with visitors to the community forum also averaging the most time on site per session. Information-sharing pages also feature heavily in the top-10 list, including "Get started now," "What does smoking cost," and "Medications can help." Less popular pages of the site highlight ways to distract oneself from nicotine cravings; "Get active," "Change your routine," and "Stay busy" are all among the 5 least-trafficked pages on the site.

In Table 1, a list of the top 10 most visited pages on the QuitNow website between March 3 and July 6, 2021, is given, highlights user engagement metrics such as page views, number of unique users, bounce rates, average time spent on each page, average session duration, and pages visited per session. These metrics were analyzed as part of our qualitative study exploring user experiences and engagement with QuitNow. Key observations include high engagement with the homepage and the community forum, which had significant page views and longer average time spent, while pages like "Join" and "Get started now" experienced higher bounce rates, suggesting opportunities for improved user retention. This data provides insights into how users interact with the site and which content resonates most with them.

Table 1. Top 10 pages visited by QuitNow users from March 3 to July 6, 2021.

Page	Topic	Page views	Users	Bounce rate	Average time	Average. session time	Pages per session
Homepage	The introduction page (users logged out)	22,719	9171	33%	00:00:39	00:02:54	1.42
Forum	Community forum for peer and coach support where users can share their experience	11,560	1918	59%	00:04:44	00:02:52	2.07
Quit Plan	Personalized guide for quitting when the users are logged in	10,393	1904	50%	00:02:48	00:05:49	2.84
Join	Web form to join the program	8107	7217	75%	00:01:36	00:01:26	2.04
Sign-in	Page to sign-in into the profile	7655	1326	10%	00:00:19	00:05:57	4.2
Get started now	Explanation of the program, why to create an account and steps	2553	1667	85%	00:01:19	00:00:38	1.36
What does smoking cost?	Interactive calculator to see how much money one spends	2317	1467	63%	00:03:37	00:01:31	1.41
Medications can help	Guide and description of the available medications in the province	1999	1159	75%	00:02:19	00:01:45	1.76
Community and support	Access to the services QuitNow provides to the community	1615	814	59%	00:00:43	00:01:11	2.3
Quitting	More information about stages of the quit process, from thinking about to preparing, to quitting, to staying quit.	1389	799	35%	00:00:33	00:02:38	3.51

In Table 2, the metrics were analyzed as part of our qualitative study exploring user experiences and engagement with QuitNow. These pages cover niche topics such as staying active, scheduling calls with Quit Coaches, supporting loved ones on their quit journey, coping with cravings, and finding distractions. Despite low user numbers, some pages—such as "Get Active" and "Supporting Loved One"—had notably high pages per session and long average session times, indicating deep engagement by the few who visited them. Conversely, other pages like "Stay Busy" and "Change Your Routine" experienced higher bounce rates and shorter time spent, suggesting potential areas for improvement in content or accessibility. This data helps identify underused resources and guides strategies for increasing visibility and user interaction.

XSL•FO
RenderX

Table 2. Least 5 visited pages on the QuitNow website between March 3 and July 6, 2021, focusing on user engagement metrics including number of unique users, bounce rates, average time spent on each page, average session duration, and pages visited per session.

Page	Topic	Users	Bounce rate, %	Average time	Average session time	Pages per session
Get active	The introduction page (users logged out)	107	33	00:01:44	00:05:47	7.33
Schedule a call	Page to schedule a call with one of the Quit Coaches	60	50	00:00:37	00:01:03	3
Supporting loved one	How to support and behave with people who are on their quit journey	52	28	00:00:59	00:07:04	7
Change your routine	Tips to cope with cravings and triggers	65	50	00:00:53	00:03:25	4.25
Stay busy	Advice and recommendations on how to get distracted from cravings	35	50	00:00:30	00:07:52	4.5

Discussion

Principal Findings

This study brings forward the end user experiences and utilization metrics of using a web-based cessation program (QuitNow) among individuals who want to quit smoking or vaping. In light of the shifting trends in nicotine and tobacco product use (eg, rise in use of e-cigarettes), demographic behaviors (eg, high vaping prevalence among youth), the rapid rollout of new digital technologies, and the high use of digital platforms, these findings provide important directions for optimizing such platforms. In particular, the findings bring forward important contextual factors that are unique to today compared with traditional approaches.

As the availability and use of digital platforms continue to rise and as nicotine users, particularly younger demographics (teens and young adults), navigate an increasing number of online platforms, the expectation for engaging and accessible content delivery becomes paramount. As indicated by individuals who use QuitNow, engaging formats not only capture attention more effectively but also contribute to a more relatable and memorable learning experience, which are crucial elements in supporting individuals on their cessation journeys [18]. A lack of information delivered in a variety of accessible, engaging, and visual formats may thus explain why the QuitNow Google Analytics data indicate that end users are not engaging with several pages focused on ways to distract from cravings despite our qualitative findings indicating end users' desire for rapid access to tips for dealing with cravings. Therefore, digital cessation supports must shed their traditional reliance on text-heavy formats and instead, leverage innovative and engaging multimedia elements, such as video testimonials, interactive graphics, podcasts, and opportunities to share on social media. Indeed, existing research highlights that, in the context of nicotine cessation, where motivation and sustained commitment are essential, a visual and interactive presentation of information serves as a powerful catalyst for positive behavioral change [19]. In addition to boosting engagement, interactive modalities may also foster a deeper connection with users by catering to varied learning preferences and thus enhancing accessibility [20]. Therefore, our findings emphasize that the effectiveness of nicotine cessation support websites relies not only on the quality of their content but also on the delivery methods they use.

The qualitative findings also indicate that information on pharmacotherapy options for quitting is desired, and the Google Analytics data indicate that information on pharmacotherapy options ranks among the top 10 pages visited on QuitNow. Pharmacotherapy is vital for individuals seeking to quit smoking and vaping, primarily addressing the physiological aspects of nicotine addiction. Approved pharmacotherapy for smoking cessation includes NRT products and several prescription medications. Research has consistently highlighted the effectiveness of pharmacotherapy, with some research indicating that those who use NRT or prescription medications are up to 2-3 times more likely to quit than those who attempt cessation without pharmacotherapy [21-24]. NRT offers a controlled and regulated supply of nicotine without exposing individuals to the more than 7000 chemicals found in cigarette smoke or vape aerosol [21]. This controlled supply of nicotine helps reduce withdrawal symptoms, such as cravings and irritability, making quitting more manageable and significantly reducing the risk of relapse [25]. The combined use of a long- and short-acting form of NRT (eg, the nicotine patch and the nicotine gum) is considered the most effective method of NRT use [21]. It is particularly effective for those with more severe addictions or those who have not succeeded with single-form NRT [26]. This evidence-based approach, supported by extensive scientific research, underscores the significance of NRT, including combination therapies, as indispensable tools in the cessation process. The prescription medications for cessation include varenicline (brand name Champix) and bupropion (brand name Zyban). Varenicline helps reduce nicotine cravings and reduces the satisfaction one gets from smoking or vaping. Bupropion helps reduce nicotine cravings and withdrawal symptoms. Varenicline is one of the pharmacological tools most likely to help people quit smoking, with 12-16 of every 100 individuals quitting smoking using varenicline compared to 6 of 100 who did not use any medicine. After using bupropion, 9 of 100 people are likely to quit [27].

The British Columbia Smoking Cessation Program has been a commendable initiative aimed at supporting British Columbia residents in their efforts to quit smoking. The program provides free or low-cost access to NRT products, such as gum, patches, and lozenges, for eligible participants, significantly reducing

XSL•FO
RenderX

financial barriers to quitting. In the most recent evaluation of the program, data showed that it had provided coverage for more than 25% of British Columbia residents who smoke. From September 30, 2011, to March 31, 2020, over 354,900 patients received smoking cessation aids (294,700 for nicotine gum, inhaler, lozenge, or patches, and 118,700 for bupropion or varenicline), with the British Columbia Ministry of Health investing approximately CAD $112.2 (US $78.47) million for product coverage. While the program has played a valuable role in aiding smoking cessation, it is essential to acknowledge that there is room for improvement. Currently, the program does not support residents in vaping cessation. In addition, the program does not allow residents to access combination NRT. While current policy allows switching from one product to another during treatment, only one of either NRTs or prescription medication is covered for 12 weeks each calendar year, which may not be enough for residents looking to quit. As well, 12 weeks of treatment is often not sufficient to achieve abstinence and prevent relapse. As a result of this limitation, some residents reported paying out of pocket for the products they used. Of over 4000 residents surveyed for the evaluation, a third of respondents (35%) indicated they had out-of-pocket expenses for other products or activities to help them quit smoking. The average monthly costs associated with these cessation products or activities were CAD $102 (US $71.34) and ranged from CAD $1 to $2000 (US $0.70 to $1398.75) a month. In light of our qualitative findings, the program must adapt to better support residents considering the evolving landscape of nicotine addiction, including the rising prevalence of vaping and emerging research on more effective cessation strategies. Collaboration between the government, community stakeholders, providers, and pharmacies is crucial for enhancing the program's reach and efficacy, ensuring it remains a robust and relevant resource in promoting a smoke-and-vape-free British Columbia. Addressing these needs is critical, as pharmacotherapy is a substantial component of a successful quit trajectory.

Our findings also indicate how incorporating social support on cessation support websites is crucial to providing end users with a sense of community, encouragement, and shared experiences. Existing research mirrors this; peer support has been shown to positively impact cessation outcomes by fostering motivation, accountability, and the exchange of coping strategies [28]. Similar to findings from a previous study on QuitNow, implementing and enhancing features, such as discussion forums, live chat options, or virtual support groups provide end users with the ability to connect, share challenges, and celebrate successes [29]. In addition, integrating social media elements may allow users to easily share their progress, garnering support from friends and family beyond the platform. By fostering a supportive online community, cessation websites like QuitNow can enhance user engagement and ultimately contribute to sustained success with respect to user quit journeys, which is emphasized in our qualitative findings, and further by the Google analytics data indicating that the community forum and support ranked among the top 10 visited pages.

Finally, our findings illuminate the importance of a collaborative and synergistic nature between the quitting individual's personal agency and the resources offered by a cessation platform, like QuitNow. As such, it is crucial for cessation websites to embody a commitment to foster a collaborative relationship with end users in their quit journey versus embodying a commitment to being the "silver bullet"—the perfect solution that removes all responsibility from the person who wants to quit. In this regard, websites must function as allies in this process by providing a structured and informed environment that catalyzes the user's cessation journey (ie, by including personalized tools, expert guidance, and a supportive community). Ultimately, incorporating a collaborative and person-centered lens in website development and delivery can not only acknowledge the significance of personal agency and make users feel more capable but also underscore the role of external support in enhancing the efficacy and sustainability of the quitting process.

Strengths and Limitations

This study is subject to a few limitations. First, our small sample size may have masked existing themes representative of the broader population of QuitNow users. In addition, as we did not explore thematic differences in cessation needs based on gender, ethnicity, or other demographic markers, there may be identity-based nuances in QuitNow user needs that our data did not capture. Participant accounts within qualitative interviews are also subject to recall and social desirability biases, which together with the previous may have influenced which themes were most apparent in the data. Further, our findings are limited by the scope of thematic analytic techniques; coding processes used in the thematic analysis are subjective and open to interpretation and thus are heavily dependent on researcher judgment and participant accounts [30]. Next, our findings are also characterized by limitations based on geographic and temporal contexts. As participants were solely from British Columbia, Canada, participant experiences may not be representative of settings beyond British Columbia. In addition, data were collected during the COVID-19 pandemic (May-August 2021) and since then, changing health and sociocultural contexts and policies may have reshaped experiences and cessation needs of QuitNow and general nicotine users [16]. Finally, the experiences of users were also based on which interface (web vs mobile) they used to access the QuitNow website, which may have influenced their opinions on the site.

Despite these limitations, our study is characterized by unique strengths. First, our use of thematic analysis allowed us to explore both apparent and latent content within user accounts, thus allowing us to capture explicit as well as underlying meanings within the data. In addition, supplementing qualitative themes with web analytics strengthened the analysis by combining subjective insights with quantitative metrics, enhancing the comprehensiveness and credibility of our findings. This dual approach provided a nuanced understanding of user behavior on the QuitNow website and continues to facilitate informed decision-making for targeted content delivery improvements. Finally, collecting rich interview data over months allowed us to capture nuanced, as well as consistent participant viewpoints in the context of an ever-changing pandemic landscape in real time.

XSL•FO
RenderX

Conclusion

In conclusion, our study on QuitNow user experiences and utilization, supplemented by Google Analytics data, reveals important insights. Integrating our qualitative findings with usage patterns provides clear directions for program enhancement: overcoming information barriers, refining pharmacological information presentation, fostering community support, and incorporating creative motivational elements. Future research should investigate user preferences in a larger and more demographically diverse sample to ensure these findings are applicable to the wider population. However, these insights offer practical strategies to optimize QuitNow and, more broadly, advance the effectiveness of web-based cessation services, contributing to improved public health outcomes.

Acknowledgments

This study was supported by funding from the British Columbia Ministry of Health, via a British Columbia Lung Foundation Research Agreement (grant 018471) and a Canadian Cancer Society Scholar Award (prevention; grant 707156).

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to the protection of participant identity but deidentified data are available from the senior author (LS) on reasonable request.

Authors' Contributions

TS and LS contributed to conceptualization. TS and LS conducted formal analysis. TS, LS, and RHS were involved in the investigation and methodology. TS, RHS, MF, CP, LD, SM, and LS were responsible for writing – original draft. TS and LS contributed to writing – review & editing.

Conflicts of Interest

None declared.

References

1. Canadian Tobacco and Nicotine Survey (CTNS): summary of results for 2020 - Canada. Government of Canada. URL: https://www.canada.ca/en/health-canada/services/canadian-tobacco-nicotine-survey/2020-summary.html [accessed 2023-12-16]
2. Canadian Tobacco and Nicotine Survey (CTNS): summary of results for 2022. Government of Canada. URL: https://www.canada.ca/en/health-canada/services/canadian-tobacco-nicotine-survey/2022-summary.html [accessed 2023-12-16]
3. Smoking in the provinces. Tobacco Use in Canada. URL: https://uwaterloo.ca/tobacco-use-canada/adult-tobacco-use/smoking-provinces [accessed 2023-12-16]
4. Tobacco use in Canada: patterns and trends, 2015 edition. Tobacco Use in Canada. URL: https://uwaterloo.ca/tobacco-use-canada/tobacco-use-canada-patterns-and-trends-2015-edition [accessed 2023-12-16]
5. Jackson SE, West R, Brown J. If at first you don't succeed, when should you try again? A prospective study of failed quit attempts and subsequent smoking cessation. Addict Behav 2020;106:106366 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106366] [Medline: 32171957]
6. Siu AL, U.S. Preventive Services Task Force. Behavioral and pharmacotherapy iterventions for tobacco smoking cessation in adults, including pregnant women: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Sep 22;163(8):622-634. [doi: 10.7326/M15-2023] [Medline: 26389730]
7. Sealock T, Sharma S. Smoking Cessation. Treasure Island, FL: StatPearls StatPearls Publishing; 2023.
8. Myung SK, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM. Effects of web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. Arch Intern Med 2009;169(10):929-937. [doi: 10.1001/archinternmed.2009.109] [Medline: 19468084]
9. Statista. URL: https://www.statista.com/topics/1145/internet-usage-worldwide/#topicOverview [accessed 2023-12-16]
10. World Internet Users Statistics and 2023 World Population Stats. URL: https://web.archive.org/web/20240319110853/https://www.internetworldstats.com/stats.htm [accessed 2023-12-16]
11. Number of internet users worldwide from 2005 to 2024. Statista. 2022. URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ [accessed 2023-12-16]
12. Improved use of information technology can result in more coordinated care for patients. CIHI. URL: https://www.cihi.ca/en/improved-use-of-information-technology-can-result-in-more-coordinated-care-for-patients [accessed 2024-08-17]
13. Keane L, Anderson C, Perez D, Freeman B. The development and evaluation of online smoking cessation services: a narrative literature review. J Smok Cess 2018;14(1):12-20. [doi: 10.1017/jsc.2018.8]
14. A free program to quit or reduce nicotine. QuitNow.ca. URL: https://quitnow.ca/ [accessed 2023-12-16]

XSL•FO
RenderX

15. Kant R, Yadav P, Bairwa M. Effectiveness of the internet-based versus face-to-face interaction on reduction of tobacco use among adults: a meta-analysis. Cureus 2021;13(11):e19380 [FREE Full text] [doi: 10.7759/cureus.19380] [Medline: 34925983]
16. Sharma RH, Rodberg D, Struik LL. Experiences of nicotine users motivated to quit during the COVID-19 pandemic: a secondary qualitative analysis. BMJ Open 2023;13(6):e070906 [FREE Full text] [doi: 10.1136/bmjopen-2022-070906] [Medline: 37369394]
17. Willig C, Rogers WS. The SAGE Handbook of Qualitative Research in Psychology. Washington DC: SAGE Publications Ltd; 2017:17.
18. Hsieh YC, Chen KH. How different information types affect viewer's attention on internet advertising. Comput Hum Behav 2011;27(2):935-945. [doi: 10.1016/j.chb.2010.11.019]
19. An LC, Schillo BA, Saul JE, Wendling AH, Klatt CM, Berg CJ, et al. Utilization of smoking cessation informational, interactive, and online community resources as predictors of abstinence: cohort study. J Med Internet Res 2008;10(5):e55 [FREE Full text] [doi: 10.2196/jmir.1018] [Medline: 19103587]
20. Taylor HA, Sullivan D, Mullen C, Johnson CM. Implementation of a user-centered framework in the development of a web-based health information database and call center. J Biomed Inform 2011;44(5):897-908 [FREE Full text] [doi: 10.1016/j.jbi.2011.03.001] [Medline: 21396486]
21. Wadgave U, Nagesh L. Nicotine replacement therapy: an overview. Int J Health Sci (Qassim) 2016;10(3):425-435 [FREE Full text] [Medline: 27610066]
22. Pajai DD, Paul P, Reche A. Pharmacotherapy in tobacco cessation: a narrative review. Cureus 2023;15(2):e35086 [FREE Full text] [doi: 10.7759/cureus.35086] [Medline: 36938244]
23. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane Database Syst Rev 2018;5(5):CD000146 [FREE Full text] [doi: 10.1002/14651858.CD000146.pub5] [Medline: 29852054]
24. Rennard S, Hughes J, Cinciripini PM, Kralikova E, Raupach T, Arteaga C, Flexible Quit Date Study Group. A randomized placebo-controlled trial of varenicline for smoking cessation allowing flexible quit dates. Nicotine Tob Res 2012;14(3):343-350 [FREE Full text] [doi: 10.1093/ntr/ntr220] [Medline: 22080588]
25. Germovsek E, Hansson A, Karlsson MO, Westin A, Soons PA, Vermeulen A, et al. A time-to-event model relating integrated craving to risk of smoking relapse across different nicotine replacement therapy formulations. Clin Pharmacol Ther 2021;109(2):416-423. [doi: 10.1002/cpt.2000] [Medline: 32734606]
26. Lindson N, Chepkin SC, Ye W, Fanshawe TR, Bullen C, Hartmann-Boyce J. Different doses, durations and modes of delivery of nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev 2019;4(4):CD013308 [FREE Full text] [doi: 10.1002/14651858.CD013308] [Medline: 30997928]
27. Lindson N, Theodoulou A, Ordóñez-Mena JM, Fanshawe TR, Sutton AJ, Livingstone-Banks J, et al. Pharmacological and electronic cigarette interventions for smoking cessation in adults: component network meta-analyses. Cochrane Database Syst Rev 2023;9(9):CD015226 [FREE Full text] [doi: 10.1002/14651858.CD015226.pub2] [Medline: 37696529]
28. Struik L, Rodberg D, Sharma RH. The behavior change techniques used in Canadian online smoking cessation programs: content analysis. JMIR Ment Health 2022;9(3):e35234 [FREE Full text] [doi: 10.2196/35234] [Medline: 35230253]
29. Struik L, Khan S, Assoiants A, Sharma RH. Assessment of social support and quitting smoking in an online community forum: study involving content analysis. JMIR Form Res 2022;6(1):e34429 [FREE Full text] [doi: 10.2196/34429] [Medline: 35023834]
30. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1):160940691773384. [doi: 10.1177/1609406917733847]

Abbreviations

NRT: nicotine replacement therapy

Edited by A Mavragani; submitted 23.12.23; peer-reviewed by S Rajput, A Hassan, N Tayong, WJ He; comments to author 03.08.24; revised version received 19.10.24; accepted 12.12.24; published 04.03.25.

Please cite as:
Salaheddin T, Sharma RH, Fajardo M, Panter C, De Souza L, Matano SK, Struik L
Utilization and Experiences of Using Quit Now, a Nicotine and Tobacco Smoking Cessation Website: Thematic Analysis
J Med Internet Res 2025;27:e55592
URL: https://www.jmir.org/2025/1/e55592
doi:10.2196/55592
PMID:40053769

XSL•FO
RenderX

©Tala Salaheddin, Ramona H Sharma, Marcela Fajardo, Cameron Panter, Lauren De Souza, Sheila Kanyevu Matano, Laura Struik. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Internet-Based Cognitive Behavioral Therapy for Preventing Postpartum Depressive Symptoms Among Pregnant Individuals With Depression: Multicenter Randomized Controlled Trial in China

Chen-Chi Duan[1,2*], MD; Chen Zhang[1,2*], PhD; Hua-Lin Xu[3*], BA; Jing Tao[4*], MSc; Jia-Le Yu[5*], MSc; Dan Zhang[6], PhD; Shan Wu[6,7], MD; Xiu Zeng[8], BA; Wan-Ting Zeng[9], BA; Zhi-Yin Zhang[9], MSc; Cindy-Lee Dennis[10], PhD; Han Liu[1,2], MD; Jia-Ying Wu[5], BA; Ben Willem J Mol[11], PhD; He-Feng Huang[1,2,6], MD; Yan-Ting Wu[1,2], MD, PhD

[1]Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University, Shanghai, China

[2]Research Units of Embryo Original Diseases, Chinese Academy of Medical Sciences (No. 2019RU056), Shanghai, China

[3]Shaoxing Maternity and Child Health Care Hospital, Shaoxing, China

[4]Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[5]International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[6]Key Laboratory of Reproductive Genetics (Ministry of Education), Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China

[7]Department of Gynecology, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China

[8]Hunan Maternal and Child Health Care Hospital, Changsha, China

[9]Hangzhou Women's Hospital, Hangzhou, China

[10]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[11]Department of Obstetrics and Gynecology, Monash University, Clayton, Australia

[*]these authors contributed equally

Corresponding Author:
Yan-Ting Wu, MD, PhD
Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University
No.419, Fangxie Rd
Shanghai, 200000
China
Phone: 86 17321218018
Email: yanting_wu@163.com

Abstract

Background: Women are particularly vulnerable to depression during pregnancy, which is one of the strongest risk factors for developing postpartum depression (PPD). Addressing antenatal depressive symptoms in these women is crucial for preventing PPD. However, little is known about the effectiveness of internet-based cognitive behavioral therapy (ICBT) in preventing PPD in this high-risk group.

Objective: This study aims to evaluate the short- and long-term effects of ICBT in preventing PPD among women with antenatal depressive symptoms.

Methods: Participants were screened for antenatal depressive symptoms using the Edinburgh Postnatal Depression Scale (EPDS) and randomly allocated (1:1) to either the ICBT group (receiving weekly online modules starting antenatally and continuing into early postpartum) or the control group (observed without treatment). Follow-up assessments were conducted up to 12 months postpartum, and data were analyzed using generalized estimating equations. The primary outcome was the prevalence of depressive symptoms at 6 weeks postpartum. A subgroup analysis based on the severity of antenatal depressive symptoms was also performed. The secondary outcomes included the long-term effects of ICBT on maternal depression, as well as its impact on anxiety, sleep quality, social support, parenting stress, co-parenting relationships, and infant development.

Results: Between August 2020 and September 2021, 300 pregnant individuals were recruited from 5 centers across China. No significant differences were observed in depressive symptoms at 6 weeks postpartum (P=.18) or at any longer-term follow-up

XSL•FO
RenderX

time points (*P*=.18). However, a post hoc subgroup analysis showed that participants with antenatal EPDS scores of 10-12 in the ICBT group had a lower risk of developing depression during the first year postpartum (odds ratio 0.534, 95% CI 0.313-0.912; *P*=.02), but this was not observed for participants with more severe depression. Additionally, this subgroup demonstrated higher levels of co-parenting relationships (*P*=.02).

Conclusions: Among individuals with antenatal depression, ICBT did not prevent the development of PPD. However, ICBT may be a preferable option for those with mild to moderate antenatal depressive symptoms. Future research is needed to explore modifications to ICBT to address more severe depressive symptoms.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000033433; https://www.chictr.org.cn/showproj.html?proj=54482

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-022-06728-5

(J Med Internet Res 2025;27:e67386) doi:10.2196/67386

KEYWORDS

antenatal depression; postpartum depression; internet-based cognitive behavioral therapy; randomized controlled trial

Introduction

Postpartum depression (PPD), which includes major and minor depressive episodes occurring within 1 year postpartum, is a common maternal complication following childbirth, with clear evidence demonstrating both short- and long-term negative effects on mothers, children, family members, and society [1-3]. It affects approximately 20% of pregnant individuals in low- and middle-income countries [4]. Notably, depression during pregnancy is one of the strongest risk factors for PPD, with approximately one-third of individuals with PPD reporting that their symptoms began antenatally [5]. Furthermore, pregnant individuals have been reported to be particularly vulnerable to depression and anxiety since the COVID-19 pandemic, with an increased likelihood of experiencing thoughts of self-harm [6-8]. This heightened prevalence places a significant burden on the existing health care system and the provision of evidence-based treatment [9]. However, the rates of depression diagnosis and treatment uptake among pregnant individuals are lower than those observed in the general population [10,11]. It is also well-documented that pregnant individuals with depressive symptoms are often reluctant to seek treatment due to individual, community, and health system barriers, particularly in low- and middle-income countries [12,13]. Therefore, addressing antenatal depressive symptoms is critically important for preventing PPD.

Cognitive behavioral therapy (CBT) is a practical and effective treatment for depression in the general population, as it helps individuals change negative thoughts, attitudes, and beliefs while promoting positive emotional and behavioral responses [14]. Evidence suggests that CBT can be initiated during pregnancy to prevent the development of PPD. Compared with traditional face-to-face delivery, internet-based CBT (ICBT) is more accessible and cost-effective, addressing common barriers to help-seeking such as shame, lack of time, transportation challenges, and limited medical resources. In this context, delivering CBT via the internet could be a promising intervention for preventing PPD, particularly after the onset of the COVID-19 pandemic, when social isolation, stigmatization, discrimination, and financial losses were frequently reported.

Although individuals with antenatal depressive symptoms are more likely to develop PPD, most studies evaluating the efficacy of prenatal ICBT for preventing PPD have been designed to recruit general pregnant individuals rather than high-risk groups [15-18]. To date, only 4 single-center randomized controlled trials with relatively small sample sizes have been published on ICBT interventions targeting individuals at high risk of antenatal depression [19-22]. Furthermore, in these studies, participants were screened for depressive symptoms before the third trimester of pregnancy, and the follow-up period was limited to 6 weeks postpartum. However, as pregnancy progresses and the physical burden increases, individuals are more likely to experience frequent urination, joint or pelvic pain, shortness of breath, swelling, constipation, indigestion, and pregnancy-related complications in late pregnancy. The prevalence of depressive symptoms during pregnancy has also been reported to rise over time, increasing from 7.4% in early pregnancy to approximately 12% in mid-to-late pregnancy [23]. Individuals with new-onset depressive symptoms emerging in late pregnancy should receive careful attention. To address these research gaps, we conducted this multicenter trial among individuals with antenatal depressive symptoms in their third trimester to evaluate the effect of ICBT on preventing depression across the first year postpartum.

Methods

Study Design and Setting

We conducted a 2-arm, parallel-group randomized controlled trial across 5 medical centers in 4 different cities in China, adhering to the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 1 [24]). The trial was carried out through obstetric clinics at the International Peace Maternity and Child Health Hospital in Shanghai, along with 4 collaborating medical centers: Shaoxing Maternity and Child Health Care Hospital in Shaoxing, Hunan Maternal and Child Health Care Hospital in Changsha, the Women's Hospital affiliated with Zhejiang University School of Medicine in Hangzhou, and Hangzhou Women's Hospital in Hangzhou.

Ethical Considerations

Ethics approval was obtained from the institutional review board of the International Peace Maternity and Child Health Hospital (approval number GKLW2020-25). Informed consent was

XSL•FO
RenderX

obtained from all participants, and those who completed the trial were financially compensated. To protect participants' privacy, each individual's name was replaced with a computer-generated serial number for identification. The trial was registered with the Chinese Clinical Trial Registry (ChiCTR2000033433), and the detailed protocol has been published [25].

Participant Screening and Recruitment

We screened pregnant individuals for antenatal depressive symptoms using the Edinburgh Postnatal Depression Scale (EPDS) during their third trimester of pregnancy (≥28 and ≤34 gestational weeks) as part of a standardized visit, which all pregnant individuals receive at the 5 participating medical centers. The EPDS is the most widely used screening tool for perinatal depression globally, demonstrating good sensitivity and specificity among Chinese individuals when a cut-off score of ≥10 is applied, as recommended for Asian populations [26]. Additionally, a higher cut-off point of ≥13 on the EPDS, indicating a higher level of depressive symptoms, is recommended in settings with limited resources for further treatment [27,28]. Pregnant individuals were eligible to participate in the trial if they met the following inclusion criteria: having a singleton fetus, being 18 years of age or older, being between ≥28 and ≤37 gestational weeks, being able to read and understand Chinese, having an EPDS score ≥10, and having access to the internet via a smartphone or computer. Individuals were excluded if they had active suicidal ideation, a diagnosis of severe mental illness (eg, schizophrenia), or were currently receiving treatment for depression. Eligible individuals were then contacted and introduced to the study, and those who agreed to participate were asked to complete the informed consent procedures and the baseline questionnaire with the assistance of a research assistant.

Randomization, Blinding, and Enrollment

All eligible individuals were ensured enrollment before 37 gestational weeks by a trained research assistant who was not involved in any other trial activities. Participants were randomly allocated in a 1:1 ratio to either the intervention group or the control group using a computer-generated block randomization procedure provided by the Clinical Research Public Technical Service Platform [29]. The allocation sequence was implemented through a web-based central randomization system, and the block size for randomization was set to 6. Based on the results provided by the computer-generated randomization system, participants were asked to register in the perinatal health care app. They were then assigned to either the intervention or control group by the research assistant. Because of the nature of the intervention, both the research assistant and the therapist providing the ICBT intervention were not blinded to group allocation; however, participants were blinded. To maintain blinding, all eligible individuals were thoroughly informed about the online courses and follow-up assessments, but were not told about the differences in courses provided between the 2 groups.

Treatment

Control Group

Participants in the control group received standard antenatal care, including educational videos and timely depression assessments delivered through the perinatal health care app developed for the trial (Figure 1). The videos provided information on labor and delivery, postpartum self-care, infant care, and breastfeeding. The content of these videos, which is also commonly provided in prenatal courses at medical centers, was not related to depression. Additionally, via the app, participants in the control group were assessed for depressive symptoms using the 9-item Patient Health Questionnaire (PHQ-9) once every 2 weeks before childbirth. It has been reported that the detection rate of prenatal depression measured by the EPDS is relatively higher compared with the PHQ-9 [30], suggesting that the EPDS may have greater sensitivity in screening for depressive symptoms. However, the PHQ-9 has an advantage over the EPDS in distinguishing the severity of depressive symptoms, with scores of ≥5, ≥10, and ≥15 corresponding to mild, moderate, and severe levels of depression, respectively [31]. Individuals with a total PHQ-9 score of ≥15 or a score of ≥2 on the self-harm item were referred to a registered psychiatrist for further assessment.

XSL•FO
RenderX

Figure 1. Flow diagram of participant recruitment and treatment for both groups. ICBT: internet-based cognitive behavioral therapy; EPDS: Edinburgh Postnatal Depression Scale.

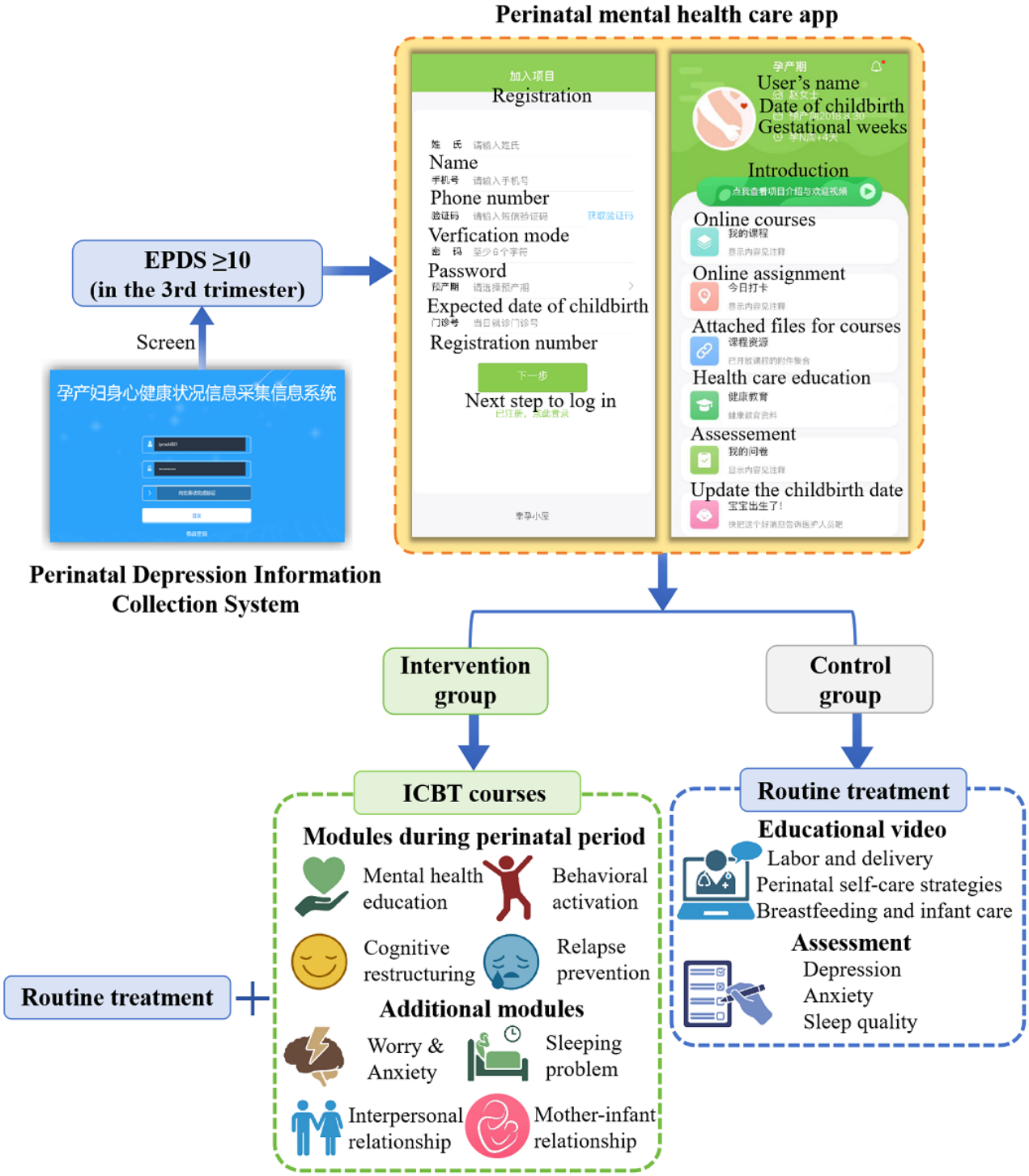

Intervention Group

Participants in the intervention group received standard antenatal care as described above, along with online text and video-based modules provided weekly through the perinatal health care app (Figure 1); 3 modules were delivered during pregnancy, and 4 modules were delivered from 2 to 5 weeks postpartum. Participants were required to complete assignments and provide feedback to the therapist via the app after each module was published. If participants had not viewed the weekly modules or submitted feedback, they would receive an automatic reminder message through the app, as well as a personal contact from the therapist. Therapists were graduate-level trainees in psychology, supervised by a registered psychiatrist with experience in treating depression. The training included a 2-day workshop before they began the intervention, and therapists were also provided with a manual outlining the weekly content for ICBT. In addition to regular weekly courses on psychoeducation, cognitive restructuring, problem-solving strategies, behavioral activation, and relapse prevention, 4 additional modules were provided. These modules addressed maternal anxiety and worries, sleep problems, interpersonal

XSL•FO
RenderX

relationship difficulties, and mother-infant relationship development.

Outcomes

Overview

All study participants completed follow-up questionnaires via the app at 6 weeks postpartum, followed by additional questionnaires at 3, 6, 9, and 12 months postpartum.

Primary Outcome

The primary outcome was the prevalence of depressive symptoms at 6 weeks postpartum, measured using the PHQ-9, a 9-item self-report scale with well-documented psychometric properties and a cut-off score of ≥5 indicating possible depression [31].

Secondary Outcomes

To further evaluate the longer-term effect of ICBT treatment on preventing the development of depressive symptoms, the PHQ-9 was readministered at 3, 6, and 12 months postpartum.

Anxiety was assessed using the 7-item Generalized Anxiety Disorder (GAD-7), a scale measuring symptoms over the past 2 weeks. Total scores range from 0 to 21, with a cut-off score >4 suggesting possible anxiety, and higher scores indicating more severe symptoms [32]. The GAD-7 was administered at 6 weeks, and at 3, 6, and 12 months postpartum.

Sleep quality was assessed using the Pittsburgh Sleep Quality Index (PSQI), a 19-item scale with good psychometric properties, measuring sleep quality over the past month [33]. The PSQI evaluates 7 key aspects of sleep, including subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. Total scores range from 0 to 21, with higher scores indicating worse sleep quality. A cut-off score >5 is recommended to distinguish between good and poor sleep quality [33]. The PSQI was administered at 3, 6, and 12 months postpartum.

Social support was assessed using the Multidimensional Scale of Perceived Social Support (MSPSS), a 12-item scale that measures the perceived availability and adequacy of emotional and instrumental social support. Total scores range from 12 to 84, with higher scores indicating greater levels of perceived support [34]. The MSPSS was administered at 6 months postpartum.

Parenting stress was assessed using the Parenting Stress Index-Short Form (PSI-SF), a 36-item scale that measures parenting stress. It consists of 3 subscales: Parental Distress, Parent-Child Dysfunctional Interaction, and Difficult Child, with scores ranging from 12 to 60. Total scores are calculated by summing the 3 subscale scores, with scores of ≥90 indicating high levels of parenting stress [35]. The PSI-SF was administered at 9 months postpartum.

The co-parenting relationship was assessed using the Brief Co-Parenting Relationship Scale (BCRS), a 14-item scale that measures co-parenting dynamics across 7 subscales: co-parenting agreement, co-parenting closeness, exposure to conflict, co-parenting support, co-parenting undermining, endorsement of partner parenting, and division of labor. Lower scores in the subscales of exposure to conflict and co-parenting undermining, and higher scores in the other 5 subscales, indicate a more positive co-parenting relationship [36]. The BCRS was administered at 9 months postpartum.

Infant development was assessed using the Denver Development Screening Test (DDST) II, a developmental assessment tool for infants and children aged 0-6 years. The scale includes items that evaluate gross motor development, fine motor-adaptive development, language development, and personal-social development. Items on the scale that over 90% of children of the same age are expected to pass but are not successfully completed are considered as “developmental delay.” Items that 75%-90% of children of the same age are expected to accomplish but are not completed are scored as a “warning.” Results are categorized as normal (no “developmental delay” or 1 “warning”), abnormal (2 or more “developmental delays”), or suspicious (2 or more “warnings” and 1 “developmental delay”) [37]. The DDST II was administered at 3, 6, and 12 months postpartum.

Sample Size

The sample size for this study was determined based on a previous Chinese trial, which reported lower rates of PPD following face-to-face CBT treatment among individuals with antenatal depressive symptoms (37/167, 22.2% vs 79/167, 47.3%; P=.001) [38]. We calculated a required sample size of at least 178 participants (89 per group, accounting for a 20% loss to follow-up) to detect a 25% reduction in PPD with 90% power and a 2-sided significance level of .05. Considering the potential reduction in efficacy with online CBT compared with face-to-face delivery, we aimed to recruit 300 participants for this study.

Statistical Analysis

Baseline characteristics were summarized using descriptive statistics. Continuous variables with a normal distribution were reported as means and SDs, while those with a skewed distribution were expressed as medians and IQRs. Categorical variables were presented as frequencies with proportions. Primary and secondary outcomes were analyzed according to the intention-to-treat principle. Group differences in continuous variables with a normal distribution were assessed using the Student t test, whereas the Kolmogorov-Smirnov test was applied for variables with a skewed distribution. Group differences in categorical variables at each time point were analyzed using a chi-square test. To evaluate the effect of the ICBT intervention over time, generalized estimating equation modeling with an unstructured correlation structure was used to analyze outcomes measured repeatedly at multiple time points [39]. The generalized estimating equation model was adjusted for covariates selected based on the quasi-likelihood under the independence model criterion. A post hoc subgroup analysis was conducted to further assess the intervention’s effect on preventing PPD among participants with varying severity of depressive symptoms at baseline. All statistical analyses were performed using R software (version 4.0.2; R Foundation), with significance defined as a 2-sided P value less than .05.

XSL•FO
RenderX

Results

Between August 2020 and September 2021, 4499 pregnant individuals were screened for depressive symptoms, of whom 1003 were identified with an EPDS score ≥10 (Figure 2 and Multimedia Appendix 1 [24]). Among these, 297 declined to participate, 9 were excluded due to multiple pregnancies, 324 were excluded for not being between 28 and 37 weeks of gestation, 66 were excluded for receiving other treatment for depression, and 7 could not be contacted. Ultimately, 300 individuals met the eligibility criteria and provided informed consent. Of these, 150 were randomized to the intervention group and 150 to the control group.

Figure 2. Flowchart of the study participants.

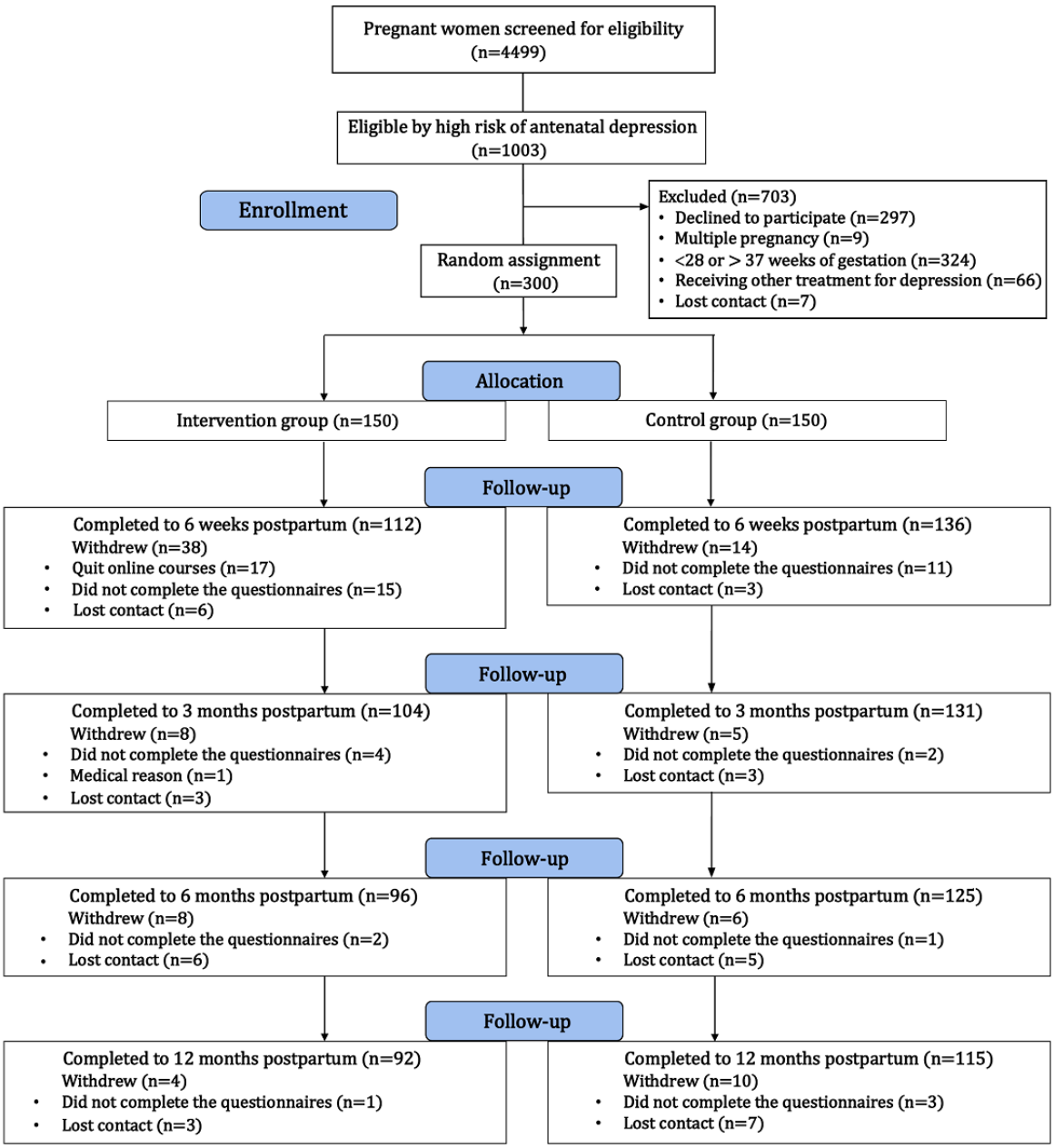

The baseline characteristics are summarized in Table 1. The median maternal age was 29 (IQR 27-32) years, with the majority of participants being college educated, employed, and primiparous. The median antenatal EPDS score was 12 (IQR 11-15), with 157 of 300 (52.3%) participants reporting mild to moderate depressive symptoms, indicated by an EPDS score ≥10 and <13. Participants in the intervention group had a slightly higher median antenatal EPDS score compared with those in the control group. Among the 248 participants included in the final intention-to-treat analysis, no significant differences were observed between the 2 groups regarding baseline depressive symptom severity, maternal sociodemographic characteristics, or pregnancy complications.

XSL•FO
RenderX

Table 1. Baseline characteristics of participants from the intervention and control groups.

Characteristics	All participants enrolled in the study			Participants included in the ITT[a] analysis		
	ICBT[b] group (n=150)	Control group (n=150)	*P* value	ICBT group (n=112)	Control group (n=136)	*P* value
EPDS[c] assessment at baseline						
Median (IQR)	12 (11.00-15.00)	12 (10.00-14.25)	.03	12 (11.00-14.75)	12 (10.00-14.75)	.15
EPDS ≥10 and <13, n (%)	76 (50.67)	81 (54.00)	.56	57 (50.89)	75 (55.15)	.50
EPDS ≥13, n (%)	74 (49.33)	69 (46.00)		55 (49.11)	61 (44.85)	
Maternal sociodemographic characteristics						
Age (years), median (IQR)	30.00 (27.00-32.00)	29.00 (27.00-33.00)	.89	30.00 (27.00-32.00)	29.00 (27.00-32.00)	.92
Pregestational BMI (kg/m^2), median (IQR)	20.31 (18.82-22.29)	20.83 (19.31-22.86)	.18	20.45 (19.03-22.81)	21.09 (19.50-22.89)	.50
Location, n (%)			.98			.81
Shanghai	40 (26.67)	41 (27.33)		32 (28.57)	39 (28.68)	
Hangzhou	52 (34.67)	49 (32.67)		41 (36.61)	43 (31.62)	
Changsha	34 (22.67)	36 (24.00)		23 (20.54)	30 (22.06)	
Shaoxing	24 (16.00)	24 (16.00)		16 (14.29)	24 (17.65)	
Ethnicity, n (%)			.28			.83
Han	144 (96.00)	148 (98.67)		109 (97.32)	134 (98.53)	
Minority	6 (4.00)	2 (1.33)		3 (2.68)	2 (1.47)	
Educational attainment, n (%)			.62			.46
High school or lower	29 (19.33)	25 (16.67)		21 (18.75)	19 (13.97)	
College	104 (69.33)	103 (68.67)		78 (69.64)	96 (70.59)	
Master or above	17 (11.33)	22 (14.67)		13 (11.61)	21 (15.44)	
Occupation, n (%)			.23			.18
Employed	87 (58.00)	101 (67.33)		66 (58.93)	95 (69.85)	
Self-employed	29 (19.33)	21 (14.00)		19 (16.96)	15 (11.03)	
Unemployed	34 (22.67)	28 (18.67)		27 (24.11)	26 (19.12)	
Annual household income, n (%)			.24			.30
<US $4000	4 (2.67)	3 (2.00)		3 (2.68)	3 (2.21)	
US $4001-US $10,000	16 (10.67)	15 (10.00)		13 (11.61)	11 (8.09)	
US $10,001-US $20,000	43 (28.67)	45 (30.00)		33 (29.46)	40 (29.41)	
US $20,001-US $40,000	43 (28.67)	58 (38.67)		32 (28.57)	55 (40.44)	
>US $40,000	44 (29.33)	29 (19.33)		31 (27.68)	27 (19.85)	
Alcohol drinking before pregnancy, n (%)	88 (58.67)	85 (56.67)	.73	63 (56.25)	76 (55.88)	.95
Alcohol drinking during pregnancy, n (%)	3 (2.00)	2 (1.33)	>.99	2 (1.79)	1 (0.74)	.45
Smoking before pregnancy, n (%)	16 (10.67)	12 (8.00)	.43	10 (8.93)	10 (7.35)	.65
Smoking during pregnancy, n (%)	2 (1.33)	1 (0.67)	>.99	1 (0.89)	1 (0.74)	.89
History of reproduction						
Parity, n (%)			>.99			.92
0	104 (69.33)	104 (69.33)		79 (70.54)	99 (72.79)	
1	44 (29.33)	44 (29.33)		32 (28.57)	36 (26.47)	
≥2	2 (1.33)	2 (1.33)		1 (0.89)	1 (0.74)	

XSL•FO
RenderX

Characteristics	All participants enrolled in the study			Participants included in the ITT[a] analysis		
	ICBT[b] group (n=150)	Control group (n=150)	*P* value	ICBT group (n=112)	Control group (n=136)	*P* value
Number of previous abortions, n (%)			.18			.40
0	82 (54.67)	92 (61.33)		66 (58.93)	89 (65.44)	
1	49 (32.67)	48 (32.00)		36 (32.14)	40 (29.41)	
≥2	19 (12.67)	10 (6.67)		10 (8.93)	7 (5.15)	
Previous ectopic pregnancy, n (%)	5 (3.33)	3 (2.00)	.72	3 (2.68)	3 (2.21)	>.99
Planned pregnancy, n (%)	85 (56.67)	87 (58.00)	.81	48 (42.86)	54 (39.71)	.62
Pregnancy complications, n (%)						
Gestational diabetes mellitus	24 (16.00)	22 (14.67)	.75	20 (17.86)	23 (16.91)	.72
Gestational hypertension	4 (2.67)	4 (2.67)	>.99	3 (2.68)	4 (2.94)	>.99
Intrahepatic cholestasis of pregnancy	2 (1.33)	1 (0.67)	.86	2 (1.79)	1 (0.74)	.86
Placenta previa	15 (10.00)	8 (5.33)	.13	12 (10.71)	7 (5.15)	.10
Fetal growth restriction	3 (2.00)	4 (2.67)	>.99	3 (2.68)	4 (2.94)	>.99

[a]ITT: intention-to-treat.

[b]ICBT: internet-based cognitive behavioral therapy.

[c]EPDS: Edinburgh Postnatal Depression Scale.

Of the 300 participants enrolled, 52 (17.3%) were lost to follow-up at 6 weeks postpartum, 65 (21.7%) at 3 months postpartum, 79 (26.3%) at 6 months postpartum, and 93 (31.0%) at 12 months postpartum. On average, participants in the ICBT group completed 2.5 (SD 1.1) modules during pregnancy and 3.1 (SD 1.5) modules postnatally. Of the 150 in the ICBT group, 129 (86%), 123 (82%), 129 (86%), 118 (79%), 118 (79%), 117 (78%), and 117 (78%), respectively, completed modules 1-7. Notably, 108 (72%) of participants completed all intervention modules delivered via the perinatal health care app.

The primary outcome, depressive symptoms at 6 weeks postpartum, did not differ significantly between the groups (ICBT vs control group: 40/112, 35.7% vs 60/136, 44.1%; *P*=.18; Table 2). Figure 3 illustrates the trajectory of depressive symptoms from 6 weeks postpartum to 1 year, as measured by the PHQ-9. No significant differences in symptom severity were observed between the intervention and control groups at any follow-up time point (*P*=.18; Figure 3, Table 2). Although symptoms worsened over time, this change was not statistically significant (*P*=.07).

XSL•FO
RenderX

Table 2. Differences in outcomes between participants from the ICBT[a] group and the control group at each time point.

Depressive symptoms	ICBT group	Control group	*P* value
Primary outcomes			
Depressive symptoms at 6 weeks postpartum			
PHQ-9[b], median (IQR)	3.00 (1.00-7.00)	4.00 (1.00-8.00)	.64
PHQ-9≥5, n/N (%)	40/112 (35.71)	60/136 (44.12)	.18
Secondary outcomes			
Depressive symptoms			
3 months postpartum			
PHQ-9, median (IQR)	3.00 (1.00-7.00)	3.00 (1.00-8.00)	.97
PHQ-9≥5, n/N (%)	39/104 (37.50)	55/131 (41.98)	.49
6 months postpartum			
PHQ-9, median (IQR)	4.00 (1.00-8.00)	4.00 (0.50-8.00)	.99
PHQ-9≥5, n/N (%)	42/96 (43.75)	62/125 (49.60)	.39
3 months postpartum			
PHQ-9, median (IQR)	4.00 (1.00-8.00)	3.00 (0.00-9.00)	.37
PHQ-9≥5, n/N (%)	42/92 (45.65)	57/115 (49.57)	.57
Anxiety symptoms			
6 weeks postpartum			
GAD-7[c], median (IQR)	3.00 (0.00-6.00)	3.00 (0.00-7.00)	.98
GAD-7≥5, n/N (%)	40/112 (35.71)	54/136 (39.71)	.52
3 months postpartum			
GAD-7, median (IQR)	3.00 (0.00-6.00)	3.00 (0.00-6.00)	.99
GAD-7≥5, n/N (%)	35/104 (33.65)	46/131 (35.11)	.81
6 months postpartum			
GAD-7, median (IQR)	3.00 (0.00-6.00)	3.00 (0.00-6.50)	.98
GAD-7≥5, n/N (%)	38/96 (39.58)	49/125 (39.20)	.95
12 months postpartum			
GAD-7, median (IQR)	3.00 (0.00-6.00)	3.00 (0.00-6.00)	.88
GAD-7≥5, n/N (%)	38/92 (41.30)	43/115 (37.39)	.57
Sleep quality			
PQSI[d]≥6 (3 months postpartum), n/N (%)	86/104 (82.69)	111/131 (84.73)	.67
PQSI≥6 (6 months postpartum), n/N (%)	78/96 (81.25)	96/125 (76.80)	.42
PQSI≥6 (12 months postpartum), n/N (%)	67/92 (72.83)	85/115 (73.91)	.86

[a]ICBT: internet-based cognitive behavioral therapy.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]GAD-7: 7-item Generalized Anxiety Disorder.

[d]PSQI: Pittsburgh Sleep Quality Index.

XSL•FO
RenderX

Figure 3. Differences in postpartum depressive symptoms between the control and ICBT groups at each time point. EPDS: Edinburgh Postnatal Depression Scale; ICBT: internet-based cognitive behavioral therapy; PHQ-9: 9-item Patient Health Questionnaire.

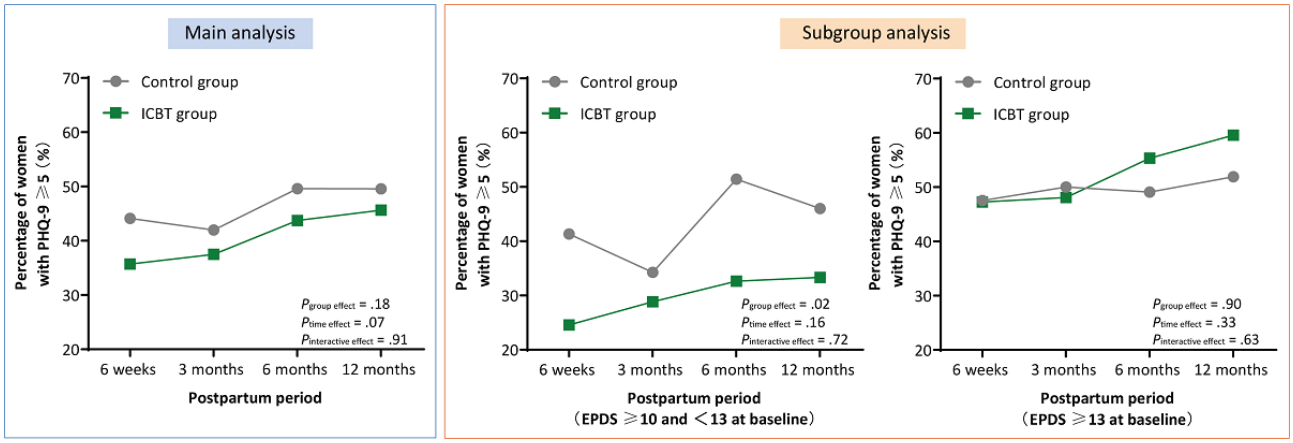

Interestingly, in a subgroup analysis of participants with a baseline EPDS score ≥10 and <13, those in the intervention group had significantly lower levels of postpartum depressive symptoms (odds ratio 0.534, 95% CI 0.313-0.912; *P*=.02; Table 3). The differences in depressive symptoms were observed at 6 weeks postpartum (ICBT vs control group: 14/57, 25% vs 31/75, 41%; *P*=.04) and at 6 months postpartum (ICBT vs control group: 16/49, 33% vs 36/70, 51%; *P*=.04; Figure 3 and Table S2 in Multimedia Appendix 2). After adjusting for covariates, the subgroup analysis indicated that participants in the intervention group were less likely to develop postpartum depressive symptoms compared with those in the control group (adjusted odds ratio 0.577, 95% CI 0.330-0.965; *P*=.04; Table 3).

XSL•FO
RenderX

Table 3. Differences in outcomes between participants from the ICBT[a] group and the control group at each time point using the GEE[b] model.

ICBT group versus control group	Odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Depressive symptoms[c]	0.758 (0.505-1.136)	.18	0.735 (0.487-1.112)[d]	.14[d]
Anxiety symptoms[e]	0.948 (0.633-1.420)	.80	0.915 (0.608-1.376)[f]	.67[f]
Sleep quality[g]	1.010 (0.605-1.686)	.97	0.943 (0.558-1.592)[f]	.82[f]
Subgroup analysis				
EPDS[h] ≥10 and <13 at baseline				
Depressive symptoms[i]	0.534 (0.313-0.912)	.02	0.577 (0.330-0.965)[j]	.04[j]
Anxiety symptoms[k]	0.838 (0.466-1.471)	.54	0.928 (0.529-1.629)[l]	.79[l]
Sleep quality[m]	1.010 (0.510-2.000)	.98	1.387 (0.692-2.786)[n]	.36[n]
EPDS ≥13 at baseline				
Depressive symptoms[i]	0.962 (0.524-1.764)	.90	0.943 (0.511-1.742)[j]	.85[j]
Anxiety symptoms[k]	0.969 (0.544-1.724)	.91	0.896 (0.472-1.701)[o]	.74[o]
Sleep quality[m]	0.951 (0.442-2.053)	.90	0.674 (0.293-1.550)[p]	.35[p]

[a]ICBT: internet-based cognitive behavioral therapy.

[b]GEE: generalized estimating equation.

[c]Defined as PHQ-9≥5 measured at 6 weeks, 3 months, 6 months, and 1 year postpartum.

[d]Values were adjusted for occupation and location.

[e]Defined as GAD-7≥5 measured at 6 weeks, 3 months, 6 months, and 1 year postpartum.

[f]Values were adjusted for location.

[g]Defined as PQSI≥6 measured at 3 months, 6 months, and 1 year postpartum.

[h]EPDS: Edinburgh Postnatal Depression Scale.

[i]*P* value for interaction [i]*P*=.002.

[j]Values were adjusted for occupation and intrahepatic cholestasis of pregnancy.

[k]*P* value for interaction [k]*P*<.001.

[l]Values were adjusted for smoking before pregnancy and educational attainment.

[m]*P* value for interaction [i]*P*=.09.

[n]Values were adjusted for location and education.

[o]Values were adjusted for location, occupation, intrahepatic cholestasis of pregnancy, and gestational diabetes mellitus.

[p]Values were adjusted for location, occupation, and educational attainment.

No significant differences between groups were found over time for the secondary outcomes of anxiety and sleep quality (Table 2). Additionally, there were no significant group differences related to social support, the co-parenting relationship, parenting satisfaction, or infant development (see Table S1 in Multimedia Appendix 2). However, subgroup analyses of participants with a baseline EPDS score between ≥10 and <13 revealed that those in the intervention group had significantly higher perceptions of co-parenting compared with the control group (ICBT vs control group: mean 7.86, SD 0.52 vs mean 6.35, SD 0.42; *P*=.02; see Table S3 in Multimedia Appendix 2). Although not statistically significant, subgroup analyses also suggested that participants in the intervention group had a higher proportion of infants classified with normal development than those in the control group at 1 year postpartum (ICBT vs control group: 37/45, 82% vs 46/63, 73%; *P*=.08).

Discussion

Principal Findings

This is the largest randomized controlled trial with a multicenter design to evaluate the effect of ICBT in treating depressive symptoms among pregnant participants and preventing the development of PPD. The results suggest that while lower rates of depressive symptoms were observed in the intervention group during the first year postpartum, the findings were not conclusive. However, in a subgroup of pregnant participants with mild to moderate depressive symptoms at baseline, those in the intervention group were significantly less likely to experience depressive symptoms at 6 weeks and 6 months postpartum compared with the control group. Participants in the intervention group were also more likely to report higher levels of co-parenting. These findings suggest that ICBT may be more effective for pregnant participants with mild to moderate depressive symptoms compared with those with more severe symptomatology.

XSL•FO
RenderX

Comparison With Prior Work

Pregnant individuals with depressive symptoms are at high risk of experiencing a continuation of these symptoms into the postpartum period, often with increasing severity. However, few studies have specifically targeted this vulnerable group with the goal of preventing the development of depressive symptoms during the first year postpartum, and the results have been inconsistent. Both Sun et al [21] (N=168) and Forsell et al [20] (N=42) found ICBT to be effective in alleviating maternal depressive symptoms at 6 weeks postpartum among those with antenatal depressive symptoms. Similar positive results were reported in a pilot study (N=25); however, individuals with major antenatal depressive symptoms were excluded from this study [19]. By contrast, Heller et al [22] (N=159) did not observe a statistically significant reduction in postpartum symptomatology following ICBT treatment among pregnant individuals with moderate to severe depressive symptoms. One potential explanation for these inconsistent findings is the timing of depressive symptom identification. Previous studies identified pregnant individuals with depressive symptoms early in the first and second trimesters (<28 gestational weeks). Our trial was the first to initiate ICBT in the third trimester and continue into the early postpartum period, reinforcing antenatal content and addressing transitional stressors immediately after delivery. However, the timing of treatment implementation may have been too late to effectively reduce symptoms and prevent the development of depressive symptoms later in the postpartum period.

Another possible explanation for the overall nonsignificant intervention effect may be an overestimation of the effectiveness of ICBT in improving severe depressive symptoms. Our findings suggest that the ICBT intervention was beneficial for pregnant individuals with mild to moderate symptoms but not for those with more severe symptomatology (baseline EPDS≥13). Similarly, a recent large Japanese trial (N=5017) investigating the effectiveness of ICBT in preventing the onset of major depressive episodes in the third trimester and at 3 months postpartum found no overall beneficial intervention effect [17]. However, the results suggest that ICBT may be beneficial in preventing perinatal depression among pregnant individuals with nonmajor depressive symptoms. Similarly, a German trial involving 406 adults with nonmajor depression found that the use of web-based CBT, compared with enhanced usual care, reduced the incidence of major depressive symptoms over 12 months [40]. These findings align with our trial results and indicate that the effectiveness of ICBT may vary depending on the severity of depressive symptoms. Other trials have demonstrated the benefits of ICBT in preventing major depression and reducing severe depressive symptoms in nonperinatal populations, often after intentionally excluding participants with severe symptomatology due to the perceived need for more intensive treatment or pharmacotherapy [41]. Further research is needed to identify individuals who might benefit most from ICBT.

No significant intervention effects were observed for any secondary outcomes, except for co-parenting support. Co-parenting support plays a crucial role in establishing an effective co-parenting relationship, which is essential for reducing mothers' child-rearing burden and fostering bonding with the infant. Previous studies have also suggested that better co-parenting may lead to a reduction in postpartum depressive symptoms [42]. The effect of ICBT on postpartum anxiety symptoms remains mixed [20-22], while sleep quality improved over time in both groups, likely due to changes in infant sleep and breastfeeding patterns. Although child development was an exploratory outcome in our trial and the study was not powered to detect group differences, we observed a trend suggesting that ICBT may positively influence child development at 1 year postpartum. Previous research has demonstrated that outcomes improve among older children when parents with depression receive intensive treatment interventions [43,44]. However, the mechanisms by which these interventions lead to improvements in child outcomes are complex. Given the evidence showing the adverse effects of PPD on child developmental trajectories [45], interventions may improve child development through various mediators, including enhanced displays of affect, strategies for addressing marital and family conflicts, and improvements in parenting behaviors. These factors may also contribute to alleviating maternal depressive symptoms. Further research evaluating technology-delivered treatments with sufficient power to detect group differences is warranted.

It is important to note that the effect of ICBT on preventing PPD should be interpreted conservatively. PPD differs from other common types of depression, as individuals face unique challenges arising from a combination of physiological and psychological changes during pregnancy, after childbirth, and throughout breastfeeding. The mental health issues resulting from these specific circumstances may exceed the expected effects of traditional ICBT treatment. To address these challenges, we included 4 additional online modules offering self-help materials and strategies for coping with maternal anxiety and worries, sleep problems, interpersonal relationship difficulties, and enhancing the mother-infant relationship. Although these modules were able to address the main concerns of perinatal individuals, they were not directly related to depression. It would have been beneficial to interview the participants to further assess the extent to which ICBT or other mediators in the intervention contributed to alleviating postpartum depressive symptoms.

Our study has several strengths. It is the first to evaluate both the short- and long-term effects of ICBT on reducing depressive symptoms and preventing PPD among pregnant individuals identified through systematic screening at the end of the second trimester. Compared with previous studies aimed at preventing PPD among pregnant individuals with depressive symptoms, our sample is larger and was recruited from multiple sites across China, increasing the generalizability of the findings. While ICBT has been shown to be effective in alleviating depressive and anxiety symptoms among the general adult population during the COVID-19 pandemic [46], only 1 [47] study has focused on maternal depression and evaluated the efficacy of ICBT treatment delivered in a single session. Our study builds on this work and provides evidence that ICBT may be an effective treatment for individuals with mild to moderate

XSL•FO
RenderX

antenatal depressive symptoms, with longer-term benefits observed up to 6 months postpartum.

Limitations

One limitation of this study is that the assessment of depressive and anxiety symptoms relied on self-report rather than a structured diagnostic interview. Although the measures used are internationally recommended and widely adopted, they do not provide a formal diagnosis. A second limitation is that the sample size was based on a previous study showing lower rates of PPD among Chinese pregnant individuals with antenatal depressive symptoms following face-to-face CBT treatment [38]. As a result, our sample size may have been insufficient to examine the secondary and longer-term outcomes. Additionally, in the intervention group, slightly higher baseline EPDS scores were observed, and there was a higher dropout rate, as completing the intervention was more time-consuming for participants. This may have introduced bias into the results. Furthermore, it is difficult to determine whether the COVID-19 pandemic had a more specific and far-reaching impact on the development of PPD, which limits the generalizability of the estimated ICBT intervention effect after the pandemic ended.

Conclusions

Despite these limitations, our results indicate a beneficial effect of ICBT on treating mild to moderate antenatal depressive symptoms and preventing the development of depressive symptoms later in the postpartum period. ICBT may be an affordable alternative for pregnant individuals with mild to moderate depressive symptoms, compared with those with more severe symptomatology (EPDS≥13). Future research is warranted to expand the appropriate application of ICBT for individuals with mild to moderate depressive symptoms and to explore best practices for optimizing engagement and integrating ICBT into care pathways.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (grants 82301719 and 82088102), STI2030-Major Projects-2021ZD0200700, Collaborative Innovation Program of Shanghai Municipal Health Commission (grant 2020CXJQ01), CAMS Innovation Fund for Medical Sciences (grant 2019-12M5-064), Grant for General Program from the Shanghai Mental Health Center (grant 2019-YJ09), Clinical Research Plan of Shanghai Shenkang Hospital Development Center (grant SHDC2020CR1008A), Shanghai Clinical Research Center for Gynecological Diseases (grant 22MC1940200), Shanghai Urogenital System Diseases Research Center (grant 2022ZZ01012), Shanghai Frontiers Science Research Base of Reproduction and Development, and Science and Technology Innovation Fund of Shanghai Jiao Tong University (grant YG2020YQ29). The funders of this study had no role in the study design, data collection, analysis, and manuscript drafting or future publication. We thank Lu-Lu Pan for verifying the data analysis. We thank all participants for their participation in the project. We appreciate the contribution of all researchers, nurses, and clerical workers from cooperative hospitals. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data Availability

Further access to participant-level data sets or statistical codes is available from the corresponding author on reasonable request.

Authors' Contributions

YTW, CCD, and JT designed the study concept. CCD, CZ, JLY, and HLX drafted the manuscript. JT provided professional support for the design of ICBT courses and was responsible for the psychological assessment of patients. CCD, JLY, HLX, DZ, XZ, WTZ, ZYZ, HL, and JYW were responsible for data collection and quality control of each center. CZ was responsible for statistical analysis. YTW, HFH, CLD, and BWJM critically revised the manuscript. All authors provided administrative, technical, and material support and agreed with the final version of the article.

Conflicts of Interest

BWM reports consultancy, travel support and research funding from Merck and consultancy for Organon and Norgine.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 676 KB - jmir_v27i1e67386_app1.pdf]

Multimedia Appendix 2
Additional analysis.
[DOCX File , 30 KB - jmir_v27i1e67386_app2.docx]

References

XSL•FO
RenderX

1. Goodman SH, Rouse MH, Connell AM, Broth MR, Hall CM, Heyward D. Maternal depression and child psychopathology: a meta-analytic review. Clin Child Fam Psychol Rev 2011 Mar;14(1):1-27. [doi: 10.1007/s10567-010-0080-1] [Medline: 21052833]
2. Spinelli MG. Maternal infanticide associated with mental illness: prevention and the promise of saved lives. Am J Psychiatry 2004 Sep;161(9):1548-1557. [doi: 10.1176/appi.ajp.161.9.1548] [Medline: 15337641]
3. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: a review. Infant Behav Dev 2010 Feb;33(1):1-6 [FREE Full text] [doi: 10.1016/j.infbeh.2009.10.005] [Medline: 19962196]
4. Gelaye B, Rondon MB, Araya R, Williams MA. Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. Lancet Psychiatry 2016 Oct;3(10):973-982 [FREE Full text] [doi: 10.1016/S2215-0366(16)30284-X] [Medline: 27650773]
5. Wisner KL, Sit DKY, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry 2013 May;70(5):490-498 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.87] [Medline: 23487258]
6. Shorey SY, Ng ED, Chee CY. Anxiety and depressive symptoms of women in the perinatal period during the COVID-19 pandemic: a systematic review and meta-analysis. Scand J Public Health 2021 Nov;49(7):730-740. [doi: 10.1177/14034948211011793] [Medline: 33966511]
7. Biaggi A, Conroy S, Pawlby S, Pariante CM. Identifying the women at risk of antenatal anxiety and depression: a systematic review. J Affect Disord 2016 Feb;191:62-77 [FREE Full text] [doi: 10.1016/j.jad.2015.11.014] [Medline: 26650969]
8. Wu Y, Zhang C, Liu H, Duan C, Li C, Fan J, et al. Perinatal depressive and anxiety symptoms of pregnant women during the coronavirus disease 2019 outbreak in China. Am J Obstet Gynecol 2020 Aug;223(2):240.e1-240.e9 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.009] [Medline: 32437665]
9. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
10. Bennett IM, Marcus SC, Palmer SC, Coyne JC. Pregnancy-related discontinuation of antidepressants and depression care visits among Medicaid recipients. Psychiatr Serv 2010 Apr;61(4):386-391. [doi: 10.1176/ps.2010.61.4.386] [Medline: 20360278]
11. Vesga-López O, Blanco C, Keyes K, Olfson M, Grant BF, Hasin DS. Psychiatric disorders in pregnant and postpartum women in the United States. Arch Gen Psychiatry 2008 Jul;65(7):805-815 [FREE Full text] [doi: 10.1001/archpsyc.65.7.805] [Medline: 18606953]
12. Halbreich U, Karkun S. Cross-cultural and social diversity of prevalence of postpartum depression and depressive symptoms. J Affect Disord 2006 Apr;91(2-3):97-111. [doi: 10.1016/j.jad.2005.12.051] [Medline: 16466664]
13. Pearlstein T. Depression during Pregnancy. Best Pract Res Clin Obstet Gynaecol 2015 Jul;29(5):754-764. [doi: 10.1016/j.bpobgyn.2015.04.004] [Medline: 25976080]
14. Sockol LE. A systematic review of the efficacy of cognitive behavioral therapy for treating and preventing perinatal depression. J Affect Disord 2015 May 15;177:7-21. [doi: 10.1016/j.jad.2015.01.052] [Medline: 25743368]
15. Haga SM, Drozd F, Lisøy C, Wentzel-Larsen T, Slinning K. Mamma Mia - a randomized controlled trial of an internet-based intervention for perinatal depression. Psychol Med 2019 Aug;49(11):1850-1858 [FREE Full text] [doi: 10.1017/S0033291718002544] [Medline: 30191779]
16. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015 Sep 01;2(3):257-265 [FREE Full text] [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]
17. Nishi D, Imamura K, Watanabe K, Obikane E, Sasaki N, Yasuma N, et al. The preventive effect of internet-based cognitive behavioral therapy for prevention of depression during pregnancy and in the postpartum period (iPDP): a large scale randomized controlled trial. Psychiatry Clin Neurosci 2022 Nov;76(11):570-578 [FREE Full text] [doi: 10.1111/pcn.13458] [Medline: 36114656]
18. Arakawa Y, Haseda M, Inoue K, Nishioka D, Kino S, Nishi D, et al. Effectiveness of mHealth consultation services for preventing postpartum depressive symptoms: a randomized clinical trial. BMC Med 2023 Jun 26;21(1):221 [FREE Full text] [doi: 10.1186/s12916-023-02918-3] [Medline: 37365535]
19. Duffecy J, Grekin R, Hinkel H, Gallivan N, Nelson G, O'Hara MW. A group-based online intervention to prevent postpartum depression (Sunnyside): feasibility randomized controlled trial. JMIR Ment Health 2019 May 28;6(5):e10778 [FREE Full text] [doi: 10.2196/10778] [Medline: 31140443]
20. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
21. Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F. Effectiveness of smartphone-based mindfulness training on maternal perinatal depression: randomized controlled trial. J Med Internet Res 2021 Jan 27;23(1):e23410 [FREE Full text] [doi: 10.2196/23410] [Medline: 33502326]

XSL•FO
RenderX

22. Heller HM, Hoogendoorn AW, Honig A, Broekman BFP, van Straten A. The effectiveness of a guided internet-based tool for the treatment of depression and anxiety in pregnancy (MamaKits Online): randomized controlled trial. J Med Internet Res 2020 Mar 23;22(3):e15172 [FREE Full text] [doi: 10.2196/15172] [Medline: 32202505]
23. Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: systematic review. Obstet Gynecol 2004 Apr;103(4):698-709. [doi: 10.1097/01.AOG.0000116689.75396.5f] [Medline: 15051562]
24. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
25. Duan CC, Yu JL, Tao J, Zhang C, Zhang D, Zeng X, et al. Internet-based cognitive therapy for women with antenatal depressive symptoms during the COVID-19 pandemic: protocol for a multi-center randomized controlled trial across China. Trials 2022 Sep 21;23(1):797 [FREE Full text] [doi: 10.1186/s13063-022-06728-5] [Medline: 36131289]
26. Lee DT, Yip SK, Chiu HF, Leung TY, Chan KP, Chau IO, et al. Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry 1998 May;172:433-437. [doi: 10.1192/bjp.172.5.433] [Medline: 9747407]
27. Matthey S, Fisher J, Rowe H. Using the Edinburgh postnatal depression scale to screen for anxiety disorders: conceptual and methodological considerations. J Affect Disord 2013 Apr 05;146(2):224-230. [doi: 10.1016/j.jad.2012.09.009] [Medline: 23116811]
28. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) EPDS Group. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ 2020 Nov 11;371:m4022 [FREE Full text] [doi: 10.1136/bmj.m4022] [Medline: 33177069]
29. Shanghai Clinical Research Center (SCRC). URL: http://www.scrcnet.org/CRIS_en.asp#top [accessed 2020-01-01]
30. Liu W, Li W, Wang Y, Yin C, Xiao C, Hu J, et al. Comparison of the EPDS and PHQ-9 in the assessment of depression among pregnant women: similarities and differences. J Affect Disord 2024 Apr 15;351:774-781. [doi: 10.1016/j.jad.2024.01.219] [Medline: 38290581]
31. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
32. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
33. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
34. Zimet G, Powell S, Farley G, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. J Pers Assess 1990;52(3-4):30-41. [doi: 10.1080/00223891.1990.9674095] [Medline: 2280326]
35. Luo J, Wang MC, Gao Y, Zeng H, Yang W, Chen W, et al. Refining the Parenting Stress Index-Short Form (PSI-SF) in Chinese parents. Assessment 2021 Mar;28(2):551-566. [doi: 10.1177/1073191119847757] [Medline: 31072108]
36. Feinberg ME, Brown LD, Kan ML. A multi-domain self-report measure of coparenting. Parent Sci Pract 2012 Jan 01;12(1):1-21 [FREE Full text] [doi: 10.1080/15295192.2012.638870] [Medline: 23166477]
37. Frankenburg W, Dodds J, Archer P, Shapiro H, Bresnick B. The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. Pediatrics 1992 Jan;89(1):91-97. [Medline: 1370185]
38. Zhao Y, Munro-Kramer ML, Shi S, Wang J, Luo J. A randomized controlled trial: effects of a prenatal depression intervention on perinatal outcomes among Chinese high-risk pregnant women with medically defined complications. Arch Womens Ment Health 2017 Apr;20(2):333-344. [doi: 10.1007/s00737-016-0712-7] [Medline: 28058505]
39. Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. Stat Med 1998 Jun 15;17(11):1261-1291. [doi: 10.1002/(sici)1097-0258(19980615)17:11<1261::aid-sim846>3.0.co;2-z] [Medline: 9670414]
40. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: a randomized clinical trial. JAMA 2016 May 03;315(17):1854-1863. [doi: 10.1001/jama.2016.4326] [Medline: 27139058]
41. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression. Harvard Review of Psychiatry 2017 May;25(3):114-122. [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
42. Takeishi Y, Nakamura Y, Yoshida M, Kawajiri M, Atogami F, Yoshizawa T. Associations between coparenting relationships and maternal depressive symptoms and negative bonding to infant. Healthcare (Basel) 2021 Mar 28;9(4):375 [FREE Full text] [doi: 10.3390/healthcare9040375] [Medline: 33800683]
43. Gunlicks ML, Weissman MM. Change in child psychopathology with improvement in parental depression: a systematic review. J Am Acad Child Adolesc Psychiatry 2008 Apr;47(4):379-389. [doi: 10.1097/CHI.0b013e3181640805] [Medline: 18388766]
44. Weissman MM, Pilowsky DJ, Wickramaratne PJ, Talati A, Wisniewski SR, Fava M, et al. Remissions in maternal depression and child psychopathology: a STAR*D-child report. JAMA 2006 Mar 22;295(12):1389-1398. [doi: 10.1001/jama.295.12.1389] [Medline: 16551710]

XSL•FO
RenderX

45. Rogers A, Obst S, Teague SJ, Rossen L, Spry EA, Macdonald JA, et al. Association between maternal perinatal depression and anxiety and child and adolescent development: a meta-analysis. JAMA Pediatr 2020 Nov 01;174(11):1082-1092 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.2910] [Medline: 32926075]
46. Komariah M, Amirah S, Faisal EG, Prayogo SA, Maulana S, Platini H, et al. Efficacy of internet-based cognitive behavioral therapy for depression and anxiety among global population during the COVID-19 pandemic: a systematic review and meta-analysis of a randomized controlled trial study. Healthcare (Basel) 2022 Jun 30;10(7):1224 [FREE Full text] [doi: 10.3390/healthcare10071224] [Medline: 35885751]
47. Van Lieshout RJ, Layton H, Savoy CD, Brown JSL, Ferro MA, Streiner DL, et al. Effect of online 1-day cognitive behavioral therapy-based workshops plus usual care vs usual care alone for postpartum depression: a randomized clinical trial. JAMA Psychiatry 2021 Nov 01;78(11):1200-1207 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.2488] [Medline: 34495285]

Abbreviations

BCRS: Brief Co-Parenting Relationship Scale
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DDST: Denver Development Screening Test
EPDS: Edinburgh Postnatal Depression Scale
GAD: 7-item Generalized Anxiety Disorder
ICBT: internet-based cognitive behavioral therapy
MSPSS: Multidimensional Scale of Perceived Social Support
PHQ-9: 9-item Patient Health Questionnaire
PPD: postpartum depression
PSI-SF: Parenting Stress Index-Short Form
PSQI: Pittsburgh Sleep Quality Index

Edited by A Mavragani; submitted 13.10.24; peer-reviewed by BS Blette, W Gong; comments to author 29.11.24; revised version received 20.12.24; accepted 29.12.24; published 04.03.25.

Please cite as:
Duan CC, Zhang C, Xu HL, Tao J, Yu JL, Zhang D, Wu S, Zeng X, Zeng WT, Zhang ZY, Dennis CL, Liu H, Wu JY, Mol BWJ, Huang HF, Wu YT
Internet-Based Cognitive Behavioral Therapy for Preventing Postpartum Depressive Symptoms Among Pregnant Individuals With Depression: Multicenter Randomized Controlled Trial in China
J Med Internet Res 2025;27:e67386
URL: https://www.jmir.org/2025/1/e67386
doi:10.2196/67386
PMID:40053801

©Chen-Chi Duan, Chen Zhang, Hua-Lin Xu, Jing Tao, Jia-Le Yu, Dan Zhang, Shan Wu, Xiu Zeng, Wan-Ting Zeng, Zhi-Yin Zhang, Cindy-Lee Dennis, Han Liu, Jia-Ying Wu, Ben Willem J Mol, He-Feng Huang, Yan-Ting Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of a Mobile Health Intervention Based on Behavioral Integrated Model on Cognitive and Behavioral Changes in Gestational Weight Management: Randomized Controlled Trial

Meng Zhou[1,2*], MS; Li Wang[3*], PhD; Ying Deng[4], MS; Jinjin Ge[4], MS; Shiqi Zhao[4], MS; Hua You[4], PhD

[1]Department of Humanities and Management, School of Nursing, Nanjing Medical University, Nanjing, China

[2]School of Nursing, Anhui Medical University, Hefei, China

[3]Department of Gynaecology and Obstetrics, Changzhou Maternal and Child Health Care Hospital, Changzhou, China

[4]Department of Social Medicine and Health Education, School of Public Health, Nanjing Medical University, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Hua You, PhD
Department of Social Medicine and Health Education
School of Public Health
Nanjing Medical University
101 Longmian Avenue, Jiangning District
Nanjing, 211166
China
Phone: 86 13382772548
Email: youhua98@163.com

Abstract

Background: The key to gestational weight management intervention involves health-related behaviors, including dietary and exercise management. Behavioral theory-based interventions are effective in improving health-related behaviors. However, evidence for mobile health interventions based on specific behavioral theories is insufficient and their effects have not been fully elucidated.

Objective: This study aimed to examine the effects of a gestational mobile health intervention on psychological cognition and behavior for gestational weight management, using an integrated behavioral model as the theoretical framework.

Methods: This study was conducted in a tertiary maternity hospital and conducted as a single-blind randomized controlled trial (RCT) in Changzhou, Jiangsu Province, China. Using the behavioral model, integrated with the protection motivation theory and information–motivation–behavioral skills model (PMT-IMB model), the intervention group received a mobile health intervention using a self-developed app from 14 to 37 gestational weeks, whereas the control group received routine guidance through the application. Psychological cognition and behaviors related to weight management during pregnancy were the main outcomes, which were measured at baseline, and at the second and third trimesters of pregnancy using a self-designed questionnaire. Generalized estimation and regression equations were used to compare the outcome differences between the intervention and control groups.

Results: In total, 302 (302/360, 83.9%) participants underwent all measurements at 3 time points (intervention group: n=150; control group: n=152). Compared with the control group, the intervention group had significantly higher scores for information, perceived vulnerability, response cost, and exercise management in the second trimester, while their scores for perceived vulnerability, response cost, and diet management were significantly higher in the third trimester. The results of repeated measures analysis revealed that, in psychological cognition, the information dimension exhibited both the time effects (T3 β=3.235, 95% CI 2.859-3.611; *P*<.001) and the group effects (β=0.597, 95% CI 0.035-1.158; *P*=.04). Similarly, response costs demonstrated both the time effects (T3 β=0.745, 95% CI 0.199-1.291; *P*=.008) and the group effects (β=1.034, 95% CI 0.367-1.700; *P*=.002). In contrast, perceived vulnerability solely exhibited the group effects (β=0.669, 95% CI 0.050-1.288; *P*=.03). Regarding weight management behaviors, both time (T3 β=6, 95% CI 4.527-7.473; *P*<.001) and group (β=2.685, 95% CI 0.323-5.047; *P*=.03) had statistically significant impacts on the total points. Furthermore, the exercise management dimension also demonstrated both the time effects (T3 β=3.791, 95% CI 2.999-4.584; *P*<.001) and the group effects (β=1.501, 95% CI 0.232-2.771; *P*=.02).

XSL•FO
RenderX

Conclusions: The intervention program was effective in increasing psychological cognitions in terms of information, perceived vulnerability, and response costs, as well as promoting healthy behaviors among Chinese pregnant women. This study provides new evidence supporting the effectiveness of mobile intervention based on behavioral science theory in gestational weight management.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100043231; https://www.chictr.org.cn/showproj.html?proj=121736

(J Med Internet Res 2025;27:e55844) doi:10.2196/55844

KEYWORDS

cognition; health behavior; information-motivation-behavioral skills model; mobile health; psychological models; pregnant woman; randomized controlled trial; mobile phone

Introduction

Inappropriate gestational weight gain (GWG) remains a public health problem requiring immediate attention [1,2]. The proportion of inadequate weight gain among pregnant women is approximately 20% in the United States, Canada, and Europe, whereas the proportion of excessive weight gain is approximately 50%. Moreover, in Asia, 31% of pregnant women experience inadequate weight gain, whereas 37% experience excessive weight gain [3]. Currently, the appropriate GWG rate in China is less than 50% [4,5]. Appropriate GWG is essential, as both excessive and inadequate weight gain pose adverse risks to maternal and child health [6-9].

Mobile health refers to the use of mobile technologies including mobile phones, personal digital assistants, and even tablet computers to improve patient health [10]. In the current era of rapidly advancing information technology, mobile health is widely used in various medical fields. Despite the identification of various challenges in mobile health during its extensive application, including the lack of regulatory oversight, limited evidence-based literature, and concerns about privacy and security [11], mobile health still has many advantages. It not only overcomes the limitations of time and space but also saves medical and human resources. Notably, mobile health technology has become a popular resource for pregnant women to learn dietary and lifestyle modifications [12]. Mobile health interventions based on gestational diet and physical activity are widely used to promote weight management behavior [13,14], and several reviews have summarized the effectiveness of this emerging internet intervention approach to promote maternal and infant health [10,15,16].

Unhealthy behaviors, such as overeating, physical inactivity, and a sedentary lifestyle during pregnancy, are known to be considerably associated with inappropriate GWG [17]. The key to pregnancy weight management intervention is the promotion of health-related behaviors. Interventions based on health behavioral theory are effective in controlling GWG. For instance, the social cognitive theory [18,19], health belief model [20], and integrated theory of health behavioral change [21], have been used to formulate mobile health interventions aimed at gestational weight management. However, the effects of interventions based on behavioral theory are not universal. For example, the behavioral lifestyle intervention based on social cognitive theory had no impact on gestational weight management in African American individuals with obesity and White individuals [22]. Hence, the effects of these gestational weight management interventions based on behavioral theory have not been fully elucidated; therefore, in the context of mobile interventions targeting GWG, it is also necessary to further explore and validate potentially valuable behavioral theories in combination with practical applications.

In this study, the behavioral model integrated with the protection motivation theory and information–motivation–behavioral skills model (PMT-IBM model) was used to guide the mobile health intervention for gestational weight management. The integrated model involves 8 dimensions: information, behavioral skill, motivation, perceived severity, perceived vulnerability, response efficacy, self-efficacy, and response cost (Figure 1). In this integrated model, the 5 dimensions that explain the generation of motivation in the original protective motivation theory (PMT), namely perceived severity, perceived vulnerability, response efficacy, self-efficacy, and response cost, replace the single motivation dimension in the information-motivation-behavioral skills (IMB) model. These dimensions interact with the information dimension and then either directly influence behavior change or indirectly impact the emergence of behavior through behavioral skill factors [23]. Both PMT and IMB revolve around motivation. Specifically, PMT focuses on the motivation or willingness behind behavior, elucidating the formation of preventive behavioral motivation from a psychological perspective [24]. The IMB model further integrates motivation with information and skills [25]. While gestational weight control is a preventive and beneficial behavior adopted by individuals to safeguard their health [26]. Whether or not individuals have the motivation to engage in behaviors such as maintaining a balanced diet and engaging in moderate exercise during pregnancy reflects their subjective willingness to adopt such protective measures [14]. Therefore, this integrated model is suitable for the development of health education programs for gestational weight control among pregnant women. In a previous independent study, we have already extensively validated the integrated model's fitness and ability to interpret gestational weight management behavior in a cross-sectional survey of 525 pregnant women. The study explains the influencing mechanisms underlying gestational weight management behavior at the individual psychological level [27]. Nevertheless, this initially proven interpretive integrated model has not yet been verified for its effectiveness in guiding interventions in practice.

XSL•FO
RenderX

Figure 1. The behavior model integrated by protection motivation theory and information–motivation– behavioral skills model.

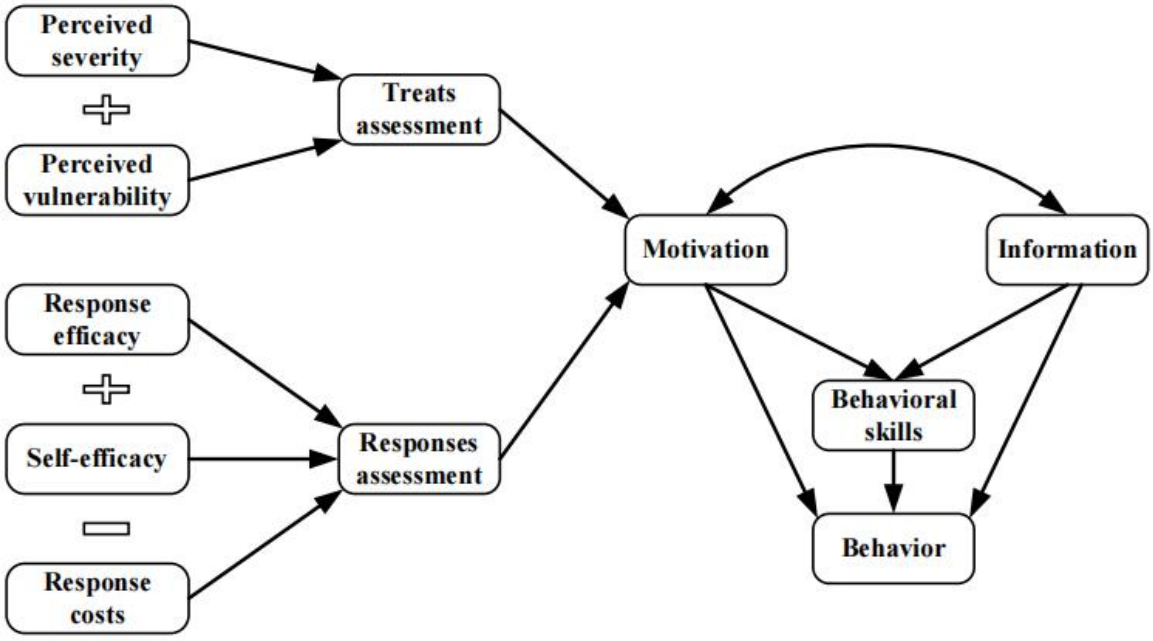

Therefore, this trial aimed to determine whether a mobile health intervention applying the PMT-IBM model as the framework could enhance psychological cognitions and behaviors toward gestational weight management.

Methods

Study Design

This was a single-blind randomized controlled trial (RCT). Eligible participants were randomly assigned by computer to (1) intervention (mobile health based on the integrated model) and (2) control (routine guidance through the mobile application) groups. The study was prospectively registered with the Chinese Clinical Trial Registry (ChiCTR2100043231).

Study Setting and Participants

Between August and December 2021, participants were recruited from Changzhou, one of the central cities in China's Yangtze River Delta region and a developed city in southern Jiangsu Province. The city occupies an area of approximately 4372.15 km^2, with a per capita gross domestic product of more than US $25,000 in 2021. (Changzhou Municipal People's Government, 2022) The study site was a tertiary maternity hospital where approximately 40% of the deliveries in this city are performed (Changzhou Maternal and Child Health Care Hospital, 2021). This study used purposive sampling, where the researchers interning in the obstetrics and gynecology outpatient clinic invited all eligible pregnant women to participate in the study. The inclusion and exclusion criteria are listed in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- singleton pregnancy;
- pregnancy confirmed by a pregnancy test (gestational age ≤14 weeks);
- intention to live in the region until delivery;
- ability to complete the questionnaire; and
- agreement to participate and provide written informed consent.

Exclusion criteria

- age <18 years;
- history of neurological, cardiovascular, hepatic, and renal medical complications;
- essential hypertension and diabetes mellitus; and
- other complications, such as deafness and dumbness.

XSL•FO
RenderX

Sample Size

Distinct from the protocol, there has been a change in the outcome indicator used for calculating the sample size in this study. Initially, in the protocol, both GWG and weight management behaviors were primary outcome indicators, and the outcome indicator for sample size calculation was based on data related to GWG in Jiangsu Province. However, due to the pandemic, the final sample size did not meet the requirements for analyzing weight outcomes. Since this study primarily focuses on behavioral changes among pregnant women, the sample size was calculated based on the behavioral indicator (the score on the Gestational Weight Management Behavior Scale). The score, serving as both a behavioral indicator and another primary outcome variable in the protocol, is deemed suitable for sample size calculation in this analysis. The formula for calculating sample size was based on a 2-sample mean comparison: $n1=n2=2[(Z_\alpha+ Z_\beta)s/d]^2$ [28]. The presurvey results were used to obtain the s=11.8 and d=5.25 [27]. Therefore, the estimated sample size was as follows: n1=n2=108. A 20% attrition rate was considered for this study; therefore, the required sample size was increased to 270 participants, including 135 participants in each group.

Randomization and Blinding

In this single-blind RCT, the participants were blinded. The randomization sequence was generated using SPSS software (version 25.0; IBM Corp) by a research assistant who did not participate in the recruitment process. Random numbers were matched to the intervention app. When participants completed the baseline questionnaire, they were assigned a unique registration number. Thereafter, the backend of the app consolidated the participants' information to screen those who met the inclusion criteria. The registration number was then matched against the groups' random sequence, to categorize the participants into intervention or control groups.

Data Collection

Our research team comprehensively studied the operational definitions of each dimension of the PMT-IBM model, as well as designed the questionnaires to strictly comply with the questionnaire design process [27]. Three trained outcome assessors performed face-to-face baseline data collection during the first trimester (T1: gestational age of ≤14 weeks) at the maternity clinic. The follow-up data were collected using a questionnaire link provided on the mobile health app in the second trimester (T2: gestational age of 27-28 weeks) and third trimester (T3: gestational age of ≥37 weeks). During questionnaire completion, the assessors reminded the participants to self-report their actual preceding month circumstances.

Outcome Measures

A detailed baseline evaluation was performed for all participants before initiating the study. The evaluation included the following three parameters: (1) main demographic characteristics: including maternal age, education level, parity, height, and prepregnancy weight. (2) Gestational psychological cognition scales: included 7 dimensions (information, behavioral skills, perceived severity, perceived vulnerability, response efficacy, self-efficacy, and response costs), which were designed by the PMT-IMB model. Each dimension was measured on a 5-point Likert scale (from 1=absolutely disagree to 5=absolutely agree) and included 5 items each, with a total score of 5-25. A higher total score indicated better psychological cognition among pregnant women. The total Cronbach α for the cognition scale of 0.860, ranged from 0.822 to 0.938. (3) Gestational weight management behavioral scales included 4 dimensions encompassing 20 items, each item was rated on a 5-point Likert scale ranging from 1 (never) to 5 (always). Specifically, exercise management dimension (consisting of 9 items, with a score range of 9-45) pertains to diverse exercise management behaviors adopted by pregnant women to maintain optimal GWG, encompassing the duration and type of exercise. The dietary management dimension (including 4 items, with a score range of 4-20) refers to a series of dietary management practices implemented by pregnant women to achieve appropriate GWG. The self-monitoring and regulation dimension (comprising 4 items, with a score range of 4-20) entails pregnant women engaging in weight monitoring, diet tracking, self-regulation, and other related behaviors aimed at maintaining suitable GWG. Then, management objectives (containing 3 items, with a score range of 3-15) involve pregnant women setting their own weight gain, dietary, and exercise goals in accordance with appropriate standards for GWG. Scores were analyzed using the gross for each dimension, with higher scores indicating superior behavioral management in each respective dimension. The total Cronbach α for the behavioral scale of 0.844, ranged from 0.653 to 0.866. The validity test of the scale has been reported in the previous explanatory study [27]. The detailed information for the scales is shown in Multimedia Appendix 1 and the operational definition of gestational weight management behaviors is shown in Multimedia Appendix 2.

Intervention Tool

The mobile health intervention was implemented using a self-developed app, "Pregnancy Assistant." "Pregnancy Assistant" was collaboratively developed by project team members, obstetricians, and computer engineers. The project leader provided the application development framework and relevant requirements, while the obstetricians were responsible for gathering feedback and suggestions from clinical health care professionals regarding the development of a "Pregnancy Assistant." The project members, on the other hand, were tasked with collecting information on pregnancy weight management and understanding the weight management needs of pregnant women.

The application comprises 3 primary user end points: for pregnant women, hospitals, and the integrated system database. The client for pregnant women included modules such as gestational age, a questionnaire, health education, notification, and weight records. Among these, the gestational age, weight records, and questionnaire modules serve as intervention assistance tools. The most crucial intervention involves delivering health education content through the "health education" module, tailored to the dimensions of the model and the specific gestational weeks of the pregnant women, along with corresponding notification reminders. In the notification module, pregnant women can engage in communication and

XSL•FO
RenderX

share their insights and experiences through the created groups. The hospital client could upload health education-related materials. All survey data were downloaded and stored in the database. In the health education module, the content received by the intervention group differs from that of the control group. Specifically, the control group is sent to routine pregnancy health care guidance, whereas the intervention group receives materials that are designed according to the PMT-IMB model (Multimedia Appendix 3).

Control Group

The control group underwent routine prenatal care in accordance with the guidelines for preconception care and prenatal care in China provided by the hospital [29]. This prenatal care encompassed regular prenatal examinations, periodic antenatal information services, and health education and guidance. Notably, the health education component only incorporated information that the country mandates to be imparted to pregnant women. Some of these health education information services would be delivered through the app named Pregnancy Assistant to ensure convenient access to and timely updates of the information.

Intervention Group

In addition to implementing the same routine prenatal care as those provided to the control group, the intervention group implemented specific interventions tailored to the dimensions and corresponding definitions of the behavioral model. These interventions were delivered through a variety of formats, including image texts, videos, expert lectures, and peer communication. According to the research results on the model construction [27], the main contents of the interventions were determined based on the information, motivation, and behavioral skills dimensions. First, we developed various health intervention materials under the 7 psychological cognition dimensions by referring to studies on gestational weight management [29] and gestational guidelines on nutrition [30-32] and exercise [30,33], combined with the recommendations from clinical nutrition experts, obstetricians, and nurses. The intervention materials (ie, image, text, and video) were delivered through the "Pregnancy Assistant" app, (24 items in total, once a week) during the intervention period, which spanned from the 14th to the 37th week of gestation. Throughout this process, researchers also regularly reminded pregnant women via the notification module of the app and WeChat (Tencent) to use the intervention materials to improve compliance. Second, every 2 weeks, a 45-minute expert lecture was held in the application to introduce information regarding the risks and severity of inappropriate GWG, its epidemiological knowledge such as risk factors, susceptible populations, and suitable preventive dietary and exercise methods. Following each expert lecture, a 15-minute question-and-answer session was offered to participants to address existing obstacles against weight management. In addition, to boost confidence and stimulate healthy behavioral changes, experienced participants were encouraged to share their thoughts and experiences via the communication group. The detailed intervention content is shown in Multimedia Appendix 3. The study adheres to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines, and the completed CONSORT-EHEALTH checklist is available in the Multimedia Appendix 4.

Statistical Analysis

In the descriptive statistical analysis, continuous variables are expressed as mean values and (SDs). Use the Kolmogorov-Smirnov test to verify the normality of these data. For data that met normality, a 2-sample independent *t* test was used to compare the 2 groups, while (ANOVA was used to compare multiple groups. For data that do not meet normality, nonparametric tests were used. Categorical variables, presented as frequencies (%), are compared between groups using the Pearson chi-square test. In a general linear model, the cognition dimensions in the PMT-IMB model and weight management behavioral dimensions were considered as the dependent variables. Intervention or not was considered the primary independent variable (0 for the control group and 1 for the intervention group). Covariates included age, education level, parity, and prepregnancy BMI. Each participant completed the same questionnaire thrice, providing repeatable and dependable measurement data. Therefore, the differential changes in psychological cognitions and behaviors at T2 and T3 with respect to T1 between the 2 groups were assessed using generalized estimating equations. When no statistical interaction between time and group was present, we only included the group's main effect and time terms to demonstrate the intervention's effect. Those who had a miscarriage or premature birth during follow-up were excluded from the final data analysis. Therefore, intention-to-treat analysis was not used in the analysis of this study. All data analyses were performed using IBM SPSS Statistics for Windows (version 25.0), and statistical significance was set at $P<.05$ (2-sided).

Ethical Considerations

This study was approved by the Ethics Committee of Nanjing Medical University (number NMU 2020-63). All participants were comprehensively informed of the voluntary nature of their participation and their right to withdraw from the study at any time. The investigation was performed only after written informed consent had been obtained from the participants. We adhere strictly to confidentiality regarding the obtained information, and we have committed to respecting patients' privacy and confidentiality rights during the data collection process. As a gesture of appreciation for the time spent by the participants completing the questionnaire, US $8.4 was provided to each participant.

Results

Participant Recruitment and Retention

After evaluation, 360 participants satisfied the eligibility criteria at baseline and were randomly assigned to intervention and control groups (n=180 per group). After their inclusion in the study, pregnant women registered themselves on the app using their mobile phones and completed the baseline questionnaire. The T2 follow-up involved 338 participants (intervention group: n=168; control group: n=170; retention rate: 88.9%). By T3,

XSL•FO
RenderX

302 participants with complete follow-up data were included in the final data analysis (intervention group: n=150; control group: n=152). The overall retention rate was 83.9%. Participants who dropped out at each stage of the study are shown in Figure 2. The comparison of demographic and clinical characteristics between dropouts and completers of the intervention revealed no significant differences (*P*>.05 for all comparisons).

Figure 2. Flow chart of the assessment, allocation, reasons for withdrawal, and number of participants included in the data analysis.

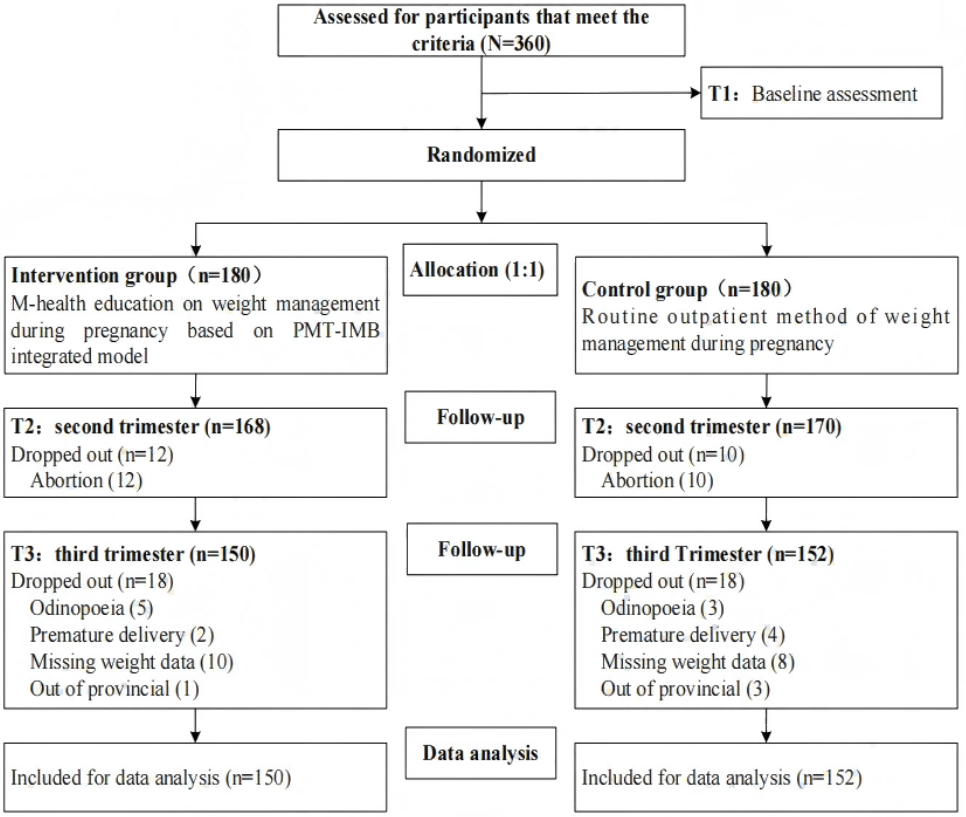

Main Demographic Characteristics

The participants' demographic characteristics are presented in Table 1. Among the 302 participants, the mean ages of the intervention and control groups were 29.31 (SD 3.38) and 29.16 (SD 3.49) years, respectively. In both groups, more than half of the participants had at least a college education (59.3%, 89/150 vs 54.6%, 83/152; *P*=.41). Most participants were primipara, and the proportion of participants with an appropriate prepregnancy BMI also exceeded 70%. No statistically significant between-group differences were observed in the main demographic characteristics (*P*>.05).

XSL•FO
RenderX

Table 1. Main demographic characteristics between the intervention group and control group (N=302).

Characteristics	Intervention group (N=150)	Control group (N=152)	Z	Chi-square (*df*)	*P* value
Age (years), mean (SD)	29.31 (3.38)	29.16 (3.49)	–0.116	—[a]	.91
Education level[b], n (%)			—	0.688 (1)	.41
College education and above	89 (59.3)	83 (54.6)			
Junior college and below	61 (40.7)	69 (45.4)			
Parity, n (%)			—	0.040 (1)	.84
Multiparity	39 (26.0)	38 (25.0)			
Primiparity	111 (74.0)	114 (75.0)			
Pre-pregnancy BMI (kg/m^2), mean (SD)	21.16 (2.62)	21.25 (2.57)	–0.324	—	.75
Prepregnancy BMI Classification, n (%)			—	0.744 (3)	.87
underweight	19 (12.7)	17 (11.2)			
normal weight	110 (73.3)	116 (76.3)			
overweight	18 (12.0)	15 (9.9)			
obese	3 (2.0)	4 (2.6)			

[a]Not applicable.

[b]Junior college degree is the median level of education.

Intervention Effect on Psychological Cognitions and Weight Management Behaviors

As shown in Table 2, at baseline, only the management objective dimension (adjusted β=0.686, *P*=.04) was significant, whereas all dimensions of psychological cognition and weight management behavior were not significantly different between the 2 groups. Adjusted general linear models demonstrated that participants in the intervention group had noticeably higher scores for information (adjusted β=0.836, *P*=.02), perceived vulnerability (adjusted β=0.894, *P*=.03), response cost (adjusted β=1.101, *P*=.006), and exercise management (adjusted β=2.089, *P*=.01) at T2 than their counterparts in the control group. Furthermore, their scores for perceived vulnerability (adjusted β=1.078, *P*=.01), response cost (adjusted β=1.538, *P*=.001), and dietary management (adjusted β=0.959, *P*=.007) were statistically significantly higher at T3. Although behavioral skills, perceived severity, response efficacy, self-efficacy, self-monitoring and regulation, and management objectives yielded generally better scores in the intervention group than in the control group at T2 and T3, statistical significance was not achieved.

XSL•FO
RenderX

Table 2. The intervention effect of psychological cognition and weight management behaviors is based on adjusted general linear models (N=302).

Outcomes	Intervention group (N=150) Mean (SD)	Control group (N=152) Mean (SD)	Adjusted β[a] (95% CI)	*P* value
Psychological cognitions				
Information				
T1[b]	13.77 (3.10)	13.55 (2.90)	0.174 (–0.500 to 0.849)	.61
T2[c]	16.54 (2.96)	15.69 (3.23)	0.836 (0.139 to 1.534)	.02
T3[d]	17.25 (3.02)	16.53 (3.28)	0.678 (–0.031 to 1.387)	.06
Behavioral skills				
T1	17.11 (3.76)	16.36 (3.97)	0.731 (–0.144 to 1.606)	.10
T2	17.30 (3.61)	16.71 (3.93)	0.615 (–0.234 to 1.465)	.16
T3	18.31 (4.14)	17.81 (4.04)	0.483 (–0.434 to 1.401)	.30
Perceived severity				
T1	21.24 (4.60)	21.36 (4.45)	–0.175 (–1.184 to 0.834)	.73
T2	21.95 (3.82)	21.90 (3.60)	0.035 (–0.807 to 0.877)	.93
T3	21.72 (3.83)	21.55 (3.61)	0.109 (–0.722 to 0.939)	.80
Perceived vulnerability				
T1	19.32 (3.50)	19.31 (3.63)	0.009 (–0.803 to 0.821)	.98
T2	19.54 (3.23)	18.63 (3.65)	0.894 (0.113 to 1.675)	.03
T3	19.66 (3.56)	18.58 (3.89)	1.078 (0.231 to 1.924)	.01
Response efficacy				
T1	21.81 (3.68)	22.28 (3.72)	–0.523 (–1.351 to 0.305)	.22
T2	21.93 (3.54)	21.61 (4.09)	0.313 (–0.553 to 1.178)	.48
T3	21.85 (3.65)	21.47 (3.77)	0.347 (–0.493 to 1.187)	.42
Self-efficacy				
T1	19.94 (3.88)	20.14 (3.94)	–0.233 (–1.120 to 0.655)	.61
T2	18.89 (3.94)	18.64 (4.08)	0.326 (–0.575 to 1.227)	.48
T3	19.01 (4.57)	18.7 (4.18)	0.342 (–0.649 to 1.334)	.50
Response cost				
T1	17.13 (4.09)	16.65 (4.55)	0.492 (–0.483 to 1.467)	.32
T2	18.15 (3.33)	17.05 (3.51)	1.101 (0.325 to 1.876)	.006
T3	18.40 (3.58)	16.88 (4.13)	1.538 (0.663 to 2.414)	.001
Weight management behaviors				
Exercise management				
T1	22.81 (5.71)	21.87 (6.22)	0.849 (–1.489 to 2.187)	.21
T2	25.66 (6.75)	23.55 (7.50)	2.089 (–0.477 to 3.702)	.01
T3	26.86 (7.47)	25.41 (7.13)	1.420 (–0.238 to 3.078)	.09
Diet management				
T1	13.33 (3.59)	13.1 (3.32)	0.166 (–0.602 to 0.935)	.67
T2	12.67 (3.42)	12.59 (3.18)	–0.003 (–0.729 to 0.724)	.99
T3	13.68 (2.79)	12.64 (3.41)	0.959 (0.269 to 1.648)	.007
Self-monitoring and regulation				
T1	10.91 (3.44)	10.60 (3.52)	0.235 (–0.540 to 1.009)	.55

XSL•FO
RenderX

Outcomes	Intervention group (N=150) Mean (SD)	Control group (N=152) Mean (SD)	Adjusted β[a] (95% CI)	*P* value
T2	11.08 (3.26)	10.89 (3.2)	0.119 (–0.588 to 0.826)	.74
T3	12.31 (3.32)	11.87 (3.21)	0.394 (–0.333 to 1.120)	.29
Management objectives				
T1	6.65 (2.97)	5.94 (2.86)	0.686 (0.024 to 1.349)	.04
T2	6.40 (2.89)	6.14 (2.82)	0.229 (–0.416 to 0.875)	.49
T3	7.37 (3.30)	7.07 (2.77)	0.272 (–0.415 to 0.958)	.44
Total scores				
T1	53.70 (11.07)	51.51 (11.52)	1.936 (–0.571 to 4.444)	.13
T2	55.81 (12.32)	53.18 (12.83)	2.435 (–0.368 to 5.239)	.09
T3	60.22 (13.57)	56.99 (13.46)	3.044 (0.006 to 6.081)	.05

[a]Model was adjusted for covariables, including age, education level, parity, and prepregnancy BMI. The reference group comprised the control group.

[b]T1: first trimester.

[c]T2: second trimester.

[d]T3: third trimester.

Generalized Estimating Equations for Psychological Cognition and Weight Management Behaviors

Psychological cognition and weight management behaviors were used as dependent variables, while time and group were used as independent variables in the generalized estimating equation analysis (Table 3)**.** Regarding psychological cognitions, a statistically significant group effect on information (β=0.597, 95% CI 0.035-1.158; *P*=.04), perceived vulnerability (β=0.669, 95% CI 0.050-1.288; *P*=.03), and response cost (β=1.034, 95% CI 0.367-1.700; *P*=.002) were observed. Moreover, time had a statistically significant effect on information (T2: β=2.457, 95% CI 2.086-2.832; *P*<.001; T3: β=3.235, 95% CI 2.859-3.611; *P*<.001) and response cost (T2: β=0.702, 95% CI 0.203-1.201; *P*=.01; T3: β=0.745, 95% CI 0.199-1.291; *P*=.01). However, no significant differences were noted in the perceived severity and response efficacy dimensions among the groups and times.

XSL•FO
RenderX

Table 3. Changes in psychological cognition and weight management behaviors in the second trimester (T2) and third trimester (T3) in the control and intervention group compared to baseline first trimester (T1) based on generalized estimating equation models.

Outcomes and independent variables[a]	β (95% CI)	*P* value
Information		
Group (category: intervention)	0.597 (0.035 to 1.158)	.04
Time		
T2[b]	2.457 (2.086 to 2.832)	<.001
T3[c]	3.235 (2.859 to 3.611)	<.001
Behavioral skills		
Group (category: intervention)	0.613 (–0.091 to 1.317)	.09
Time		
T2	0.268 (–0.179 to 0.715)	.24
T3	1.321 (0.818 to 1.824)	<.001
Perceived severity		
Group (category: intervention)	0.032 (–0.686 to 0.751)	.93
Time		
T2	0.632 (0.121 to 1.124)	.02
T3	0.331 (–0.181 to 0.843)	.21
Perceived vulnerability		
Group (category: intervention)	0.669 (0.050 to 1.288)	.03
Time		
T2	–0.235 (–0.675 to 0.204)	.29
T3	–0.199 (–0.690 to 0.293)	.43
Response efficacy		
Group (category: intervention)	0.082 (–0.602 to 0.765)	.82
Time		
T2	–0.275 (–0.726 to 0.176)	.23
T3	–0.387 (–0.803 to 0.029)	.07
Self-efficacy		
Group (category: intervention)	0.120 (–0.607 to 0.847)	.75
Time		
T2	–1.275 (–1.798 to –0.751)	<.001
T3	–1.189 (–1.687 to –0.690)	<.001
Response cost		
Group (category: intervention)	1.034 (0.367 to 1.700)	.002
Time		
T2	0.702 (0.203 to 1.201)	.006
T3	0.745 (0.199 to 1.291)	.008
Exercise management		
Group (category: intervention)	1.501 (0.232 to 2.771)	.02
Time		
T2	2.262 (1.489 to 3.034)	<.001
T3	3.791 (2.999 to 4.584)	<.001
Diet management		

XSL•FO
RenderX

Outcomes and independent variables[a]	β (95% CI)	*P* value
Group (category: intervention)	0.448 (–0.121 to 1.017)	.12
Time		
T2	–0.579 (–1.003 to –0.156)	.007
T3	–0.053 (–0.506 to 0.400)	.82
Self-monitoring and regulation		
Group (category: intervention)	0.313 (–0.275 to 0.900)	.30
Time		
T2	0.235 (–0.192 to 0.663)	.28
T3	1.338 (0.891 to 1.784)	＜.001
Management objectives		
Group (category: intervention)	0.423 (–0.091 to 0.937)	.11
Time		
T2	–0.026 (–0.415 to 0.362)	.89
T3	0.924 (0.546 to 1.302)	＜.001
Total points		
Group (category: intervention)	2.685 (0.323 to 5.047)	.03
Time		
T2	1.891 (0.524 to 3.258)	.007
T3	6.000 (4.527 to 7.473)	＜.001

[a]For the group, the control group was the reference; for time, the first trimester (T1) was the reference.

[b]T2: second trimester.

[c]T3: third trimester.

Regarding weight management behaviors, the 2 groups had a statistically significant effect on exercise management (β=1.501, 95% CI 0.232-2.771; *P*=.02). Furthermore, time had a statistically significant effect on exercise management (T2: β=2.262, 95% CI 1.489-3.034; *P*<.001; T3: β=3.791, 95% CI 2.999-4.584; *P*<.001) and dietary management (T2: β=–0.579, 95% CI 0.818-1.824; *P*=.01). Regarding total weight management behavioral scores, the group (β=2.685, 95% CI 0.323-5.047; *P*=.03) and time (T2: β=1.891, 95% CI 0.524-3.258; *P*=.01; T3: β=6.000, 95% CI 4.527-7.473; *P*<.001) effects were all statistically significant.

Discussion

Principal Findings

To the best of our knowledge, this RCT is the first to use mobile health intervention guided by the novel PMT-IBM behavioral theory and test its effectiveness on gestational weight management. Although the mechanism underlying this phenomenon requires further research, statistically significant results suggest that mobile intervention potentially plays positive roles in raising psychological cognitions in weight management including information, perceived vulnerability, and response cost, as well as promoting dietary and exercise management during pregnancy.

Role of Mobile Health App

In this study, the mobile health app “Pregnancy Assistant” played a crucial role in intervening with the weight management of pregnant women. First, the app served as an effective medium for disseminating information. It provided different health education information to both the control and intervention groups. In addition to the routine pregnancy care information available to both groups, the information provided to the intervention group was designed based on behavioral models, making it more targeted and systematic [34]. This difference might have led to varying intervention effects between the 2 groups. The precision of health education information in the intervention group made it easier for pregnant women to clearly understand the weight control goals and methods at different stages of pregnancy, whereas the control group lacked such precision [35]. Second, the app met the social interactive needs of pregnant women. The regular health lectures attended by the intervention group broadened their knowledge base. The question-and-answer sessions following the lectures addressed participants’ concerns, offering personalized guidance that enhanced their confidence and ability in weight management [36]. The peer support groups within the notification module created a supportive social environment. Pregnant women shared experiences and encouraged each other through communication, and this model of coexisting peer pressure and support could improve their enthusiasm and compliance in weight management [37]. Of course, the convenience of the app should not be

XSL•FO
RenderX

overlooked. Pregnant women could access the app anytime, anywhere, via smartphones to obtain necessary information and participate in interactions [38]. This convenience integrated weight management into pregnant women's daily lives, breaking time and space constraints and helping to form long-term, stable health management habits.

Effect on Psychological Cognitions

Based on our findings, promoting healthy behaviors in terms of information, perceived vulnerability, and response cost, via mobile health, are of practical relevance. Mobile health technologies such as apps can be applied to the design of the content of mobile interventions for these important psychological cognition dimensions. The intervention content should focus on increasing awareness regarding GWG and control and perceived vulnerability to adverse outcomes and risks while decreasing perceptions of difficult decision-making and personal insecurities. However, no statistically significant between-group differences were noted in the behavioral skills, perceived severity, self-efficacy, and response efficacy dimensions. This may be because most pregnant women lack awareness and hold misguided beliefs regarding weight gain and management. Moreover, a general lack of awareness regarding the specific risks of being overweight or obese during pregnancy is also an important barrier to appropriate GWG [39]. Meanwhile, most pregnant women lack persistence in long-term dietary management and physical exercise and have limited ability to control weight and achieve exercise compliance based on healthy behaviors in reality [40,41]. Furthermore, some pregnant women insufficiently recognize the seriousness of inappropriate GWG, mistakenly believing that pregnancy is perilous by nature and that the occurrence of complications is random [42]. In addition, due to the free availability of extensive information on the internet, the control group also had extra opportunities to access health information or learn it on their own from other sources, which may have affected the accurate evaluation of the intervention. The effectiveness of interventions informed by the theory may be affected by how successfully the theory is applied to the mobile intervention strategies. Although theories are ideal for explaining and validating behavior, they may not be excellent for guiding mobile interventions because of several complicated variables, including time of exposure and environment (ie, exposure to advertisements promoting fast food and sugary beverages). Therefore, this may result in weak or limited effects on cognitive and behavioral changes [43]. Furthermore, in the process of developing the mobile intervention, attention should be paid to the environment and intensity of intervention; moreover, differences in social, economic, cultural, and other factors should be considered [44].

Effect on Gestational Exercise and Dietary Behaviors

Our findings indicate that the intervention group's exercise management significantly improved at T2. Evidence from a previous meta-analysis also suggests that prenatal exercise may be an effective method to promote appropriate birth weight of newborns and GWG [45]. In this study, exercise behavior was measured only by the gestational weight management behavioral scales, and future studies should include objective measures of physical activity (ie, step number or physical activity measuring tools) at different stages of pregnancy to evaluate the potential of mobile health intervention in improving gestational exercise. At T3, statistically significant improvements were noted in the intervention group's dietary management but not in exercise management. A possible explanation could be that pregnant women may experience obvious weight gain in the T3 compared with that at prepregnancy. The altered body shape potentially results in physical activity restriction, leading pregnant women to prefer weight management through a healthy diet [46]. Therefore, in the T3, physical activity should be emphasized and health education related to suitable exercise should be reinforced. Previous studies have found smartphone applications and other digital interventions to promote a healthy diet among pregnant women with obesity and overweight [47,48]. However, altering the long-term dietary patterns in this intervention study, which focused on psychological cognition, proved difficult, and this might have been influenced by deeply rooted Chinese concepts (ie, "the better the eating during pregnancy, the better the growth of the fetus"). Furthermore, dietary behavioral modification is influenced by external conditions (ie, the convenience of accessing healthy foods), which is beyond the current intervention. Therefore, the ideal dietary intervention should be considered from multiple perspectives, including some external reinforcing factors.

Strengths and Limitations

This study has several strengths. First, compared with previous similar studies [48-50], the population in this study not only included women who were overweight or obese before pregnancy but also included women of all BMI categories. Second, the behavioral intervention integrated mobile intervention with 2 classical behavioral theoretical models so that the mobile intervention framework was multidimensional, providing a multipronged approach to gestational weight management.

However, this study has several limitations that warrant consideration. First, participants were recruited from a single hospital in Jiangsu Province, which may not adequately represent the entire target population globally. Multiregional, multicenter RCT should be conducted in the future. Second, the outcomes were self-reported by participants, thus potentially leading to reporting and recall bias. Third, the nonrandom sampling method could result in some selection bias. Fourth, the mean and SD values of psychological and behavioral scales were similar across the groups; therefore, the significance of the findings needs to be further examined, and the results need to be carefully interpreted.

Conclusions

This study assessed the applicability and effectiveness of a gestational weight mobile health intervention program based on the behavioral model integrated with the PMT-IMB model. This mobile intervention enhanced the psychological cognition among pregnant women in information, perceived vulnerability, and response cost, and improved dietary and exercise management. Further research is necessary to confirm the generalizability, operability, and durability of our findings in the practice of mobile health intervention.

XSL•FO
RenderX

Acknowledgments

We thank all pregnant women for their participation. We give special thanks to the Health Department of Changzhou Maternal and Child Health Hospital for the great assistance and coordination of the data collection. This study was supported by the NSFC (National Natural Science Foundation of China) (No. 72074122) and Philosophy and Social Science Research of Jiangsu Higher Education Institutions of China (2020SJA0306).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the inclusion of protected personal information but are available from the corresponding author on reasonable request.

Authors' Contributions

HY, MZ, and LW designed the study and critically reviewed, commented on, and revised the manuscript. MZ, HY, and JJG participated in data analysis, performed the final statistical analyses, and prepared the first version of the manuscript. MZ, JJG, LW, and SQZ collected the data. MZ, JJG, and SQZ collected literature materials. DY revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The detailed information for the scales.
[DOCX File , 15 KB - jmir_v27i1e55844_app1.docx]

Multimedia Appendix 2
The operational definition of gestational weight management behaviors.
[DOCX File , 12 KB - jmir_v27i1e55844_app2.docx]

Multimedia Appendix 3
The detailed intervention contents based on the PMT-IMB (protection motivation theory and information–motivation–behavior skills) model.
[DOCX File , 18 KB - jmir_v27i1e55844_app3.docx]

Multimedia Appendix 4
CONSORT eHEALTH checklist (V 1.6.1).
[DOCX File , 172 KB - jmir_v27i1e55844_app4.docx]

References

1. Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. Fertil Steril 2017;107(4):840-847 [FREE Full text] [doi: 10.1016/j.fertnstert.2017.01.017] [Medline: 28292619]
2. Wang C, Wei Y, Zhang X, Zhang Y, Xu Q, Sun Y, et al. A randomized clinical trial of exercise during pregnancy to prevent gestational diabetes mellitus and improve pregnancy outcome in overweight and obese pregnant women. Am J Obstet Gynecol 2017;216(4):340-351. [doi: 10.1016/j.ajog.2017.01.037] [Medline: 28161306]
3. Feng YY, Yu ZM, van Blyderveen S, Schmidt L, Sword W, Vanstone M, et al. Gestational weight gain outside the 2009 Institute of Medicine recommendations: novel psychological and behavioural factors associated with inadequate or excess weight gain in a prospective cohort study. BMC Pregnancy Childbirth 2021;21(1):70 [FREE Full text] [doi: 10.1186/s12884-021-03555-5] [Medline: 33478410]
4. Bi Y, Duan YF, Wang J, Jiang S, Pang XH, Yin SA, et al. Weight gain during pregnancy and its related factors in China from 2010 to 2012. Chin J Preven Med 2018;52(1):26-30. [doi: 10.3760/cma.j.issn.0253-9624.2018.01.006] [Medline: 29334704]
5. Zhang S, Li WQ, Li N, Wang LS, Liu HK, Li W. Gestational eight gain and its determinants: a maternal and infant cohort study. Chinese Journal of Public Health 2022;38(09):1135-1140.
6. Ikenoue S, Miyakoshi K, Kasuga Y, Ochiai D, Matsumoto T, Tanaka M. Impaired fetal growth in mothers with inadequate gestational weight gain: a retrospective study in Japanese uncomplicated pregnancy. J Matern Fetal Neonatal Med 2020;33(13):2227-2231. [doi: 10.1080/14767058.2018.1545835] [Medline: 30501531]

XSL•FO
RenderX

7. LifeCycle Project-Maternal ObesityChildhood Outcomes Study Group, Voerman E, Santos S, Inskip H, Amiano P, Barros H, et al. Association of gestational weight gain with adverse maternal and infant outcomes. JAMA 2019;321(17):1702-1715 [FREE Full text] [doi: 10.1001/jama.2019.3820] [Medline: 31063572]
8. Xing MM, Wei G, Wang ZL, Ren LZ. Effects of prepregnancy body mass index and pregnancy weight gain on perinatal outcomes. Chin J Child Health Care 2020;28(4):633-644.
9. Perumal N, Wang D, Darling AM, Wang M, Liu E, Urassa W, et al. Associations between gestational weight gain adequacy and neonatal outcomes in Tanzania. Ann Nutr Metab 2022;78(3):156-165 [FREE Full text] [doi: 10.1159/000522197] [Medline: 35124672]
10. Hussain T, Smith P, Yee LM. Mobile phone-based behavioral interventions in pregnancy to promote maternal and fetal health in high-income countries: systematic review. JMIR Mhealth Uhealth 2020;8(5):e15111 [FREE Full text] [doi: 10.2196/15111] [Medline: 32463373]
11. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
12. Endehabtu B, Weldeab A, Were M, Lester R, Worku A, Tilahun B. Mobile phone access and willingness among mothers to receive a text-based mHealth intervention to improve prenatal care in Northwest Ethiopia: cross-sectional study. JMIR Pediatr Parent 2018;1(2):e9 [FREE Full text] [doi: 10.2196/pediatrics.9618] [Medline: 31518334]
13. Ferrara A, Hedderson MM, Brown SD, Ehrlich SF, Tsai A, Feng J, et al. A telehealth lifestyle intervention to reduce excess gestational weight gain in pregnant women with overweight or obesity (GLOW): a randomised, parallel-group, controlled trial. Lancet Diabetes Endocrinol 2020;8(6):490-500 [FREE Full text] [doi: 10.1016/S2213-8587(20)30107-8] [Medline: 32445736]
14. Shieh C, Cullen DL, Pike C, Pressler SJ. Intervention strategies for preventing excessive gestational weight gain: systematic review and meta-analysis. Obes Rev 2018;19(8):1093-1109 [FREE Full text] [doi: 10.1111/obr.12691] [Medline: 29806187]
15. Chan KL, Chen M. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
16. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
17. Acosta-Manzano P, Acosta FM, Coll-Risco I, Romero-Gallardo L, Flor-Alemany M, Martínez-González LJ, et al. The influence of exercise, lifestyle behavior components, and physical fitness on maternal weight gain, postpartum weight retention, and excessive gestational weight Gain. Int J Sport Nutr Exerc Metab 2022;32(6):425-438. [doi: 10.1123/ijsnem.2021-0201] [Medline: 35894919]
18. Poston L, Bell R, Croker H, Flynn AC, Godfrey KM, Goff L, UPBEAT Trial Consortium. Effect of a behavioural intervention in obese pregnant women (the UPBEAT study): a multicentre, randomised controlled trial. Lancet Diabetes Endocrinol 2015;3(10):767-777 [FREE Full text] [doi: 10.1016/S2213-8587(15)00227-2] [Medline: 26165396]
19. Yan J. The Effects of Information Technology Intervention on Gestational Weight Management Based on SCT. Jiangsu, Yangzhou: Yangzhou University; 2018.
20. Shahnazi H, Abdolalian N, Kazemi A, Hassanzadeh A. Designing an educational intervention to prevent excessive gestational weight gain: a protocol for a randomized controlled trial. Reprod Health 2019;16(1):31 [FREE Full text] [doi: 10.1186/s12978-019-0696-7] [Medline: 30866980]
21. Ma C. Research on Application of Mobile Health Intervention in Weight Management of Pregnant Women Based on Integrated Theory of Health Behavior Change. Jiangsu Zhenjiang: Jiangsu University; 2020.
22. Liu J, Wilcox S, Wingard E, Turner-McGrievy G, Hutto B, Burgis J. A behavioral lifestyle intervention to limit gestational weight gain in pregnant women with overweight and obesity. Obesity (Silver Spring) 2021;29(4):672-680 [FREE Full text] [doi: 10.1002/oby.23119] [Medline: 33619910]
23. Yao X, Zhang L, Du J, Gao L. Effect of information-motivation-behavioral model based on protection motivation theory on the psychological resilience and quality of life of patients with type 2 DM. Psychiatr Q 2021;92(1):49-62. [doi: 10.1007/s11126-020-09783-w] [Medline: 32445004]
24. Ghahremani L, Faryabi R, Kaveh MH. Effect of health education based on the protection motivation theory on malaria preventive behaviors in rural households of kerman, iran. Int J Prev Med 2014;5(4):463-471 [FREE Full text] [Medline: 24829734]
25. Fleary SA, Joseph P, Chang H. Applying the information-motivation-behavioral skills model to explain adolescents' fruits and vegetables consumption. Appetite 2020;147:104546 [FREE Full text] [doi: 10.1016/j.appet.2019.104546] [Medline: 31809812]
26. Craemer KA, Sampene E, Safdar N, Antony KM, Wautlet CK. Nutrition and exercise strategies to prevent excessive pregnancy weight gain: a meta-analysis. AJP Rep 2019;9(1):e92-e120 [FREE Full text] [doi: 10.1055/s-0039-1683377] [Medline: 31041118]

XSL•FO
RenderX

27. Ge J, Zhao S, Peng X, Walker AN, Yang N, Zhou H, et al. Analysis of the weight management behavior of Chinese pregnant women: an integration of the protection motivation theory and the information-motivation-behavioral skills model. Front Public Health 2022;10:759946 [FREE Full text] [doi: 10.3389/fpubh.2022.759946] [Medline: 35186832]
28. Chow SC, Shao J, Wang H. A note on sample size calculation for mean comparisons based on noncentral t-statistics. J Biopharm Stat 2002;12(4):441-456. [doi: 10.1081/BIP-120016229] [Medline: 12477068]
29. Department of Obstetrics and Gynecology in Chinese Medical Association. Guidelines for pre-pregnancy and pregnancy care (2018). Chin J Obstet Gynaecol 2018;53(1):7-13. [doi: 10.3760/cma.j.issn.0529-567X.2018.01.003] [Medline: 29374879]
30. Cheng Y, Tao XL, Ouyang ZB, Gao TY. Interpretation of the canadian clinical practice guidelines for exercise during pregnancy in 2019. Progress Obstet Gynaecol 2019;28(5):388-390. [doi: 10.1891/9780826194985.0079]
31. Cheng YY, Guo CJ. Dietary guidelines for Americans 2020-2025. Acta Nutri Sinica. 2021 2021;42(1):3-8. [doi: 10.13325/j.cnki.acta.nutr.sin.20210302.001]
32. Zeng G. Interpretation of chinese nutrition society's dietary guidelines for pregnant women. Pract Obste Gynecol 2018;34(4):265-267. [doi: 10.1079/phn2001215]
33. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World health organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
34. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
35. Brown MJ, Sinclair M, Liddle D, Hill AJ, Madden E, Stockdale J. A systematic review investigating healthy lifestyle interventions incorporating goal setting strategies for preventing excess gestational weight gain. PLoS One 2012;7(7):e39503 [FREE Full text] [doi: 10.1371/journal.pone.0039503] [Medline: 22792178]
36. Huhn S, Matzke I, Koch M, Gunga H, Maggioni MA, Sié A, et al. Using wearable devices to generate real-world, individual-level data in rural, low-resource contexts in Burkina Faso, Africa: a case study. Front Public Health 2022;10:972177 [FREE Full text] [doi: 10.3389/fpubh.2022.972177] [Medline: 36249225]
37. Niela-Vilén H, Axelin A, Melender H, Salanterä S. Aiming to be a breastfeeding mother in a neonatal intensive care unit and at home: a thematic analysis of peer-support group discussion in social media. Matern Child Nutr 2015;11(4):712-726 [FREE Full text] [doi: 10.1111/mcn.12108] [Medline: 24521232]
38. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of smartMoms, a novel eHealth intervention for management of gestational weight gain: randomized controlled pilot trial. JMIR Mhealth Uhealth 2017;5(9):e133 [FREE Full text] [doi: 10.2196/mhealth.8228] [Medline: 28903892]
39. Escañuela Sánchez T, Meaney S, O'Connor C, Linehan L, O'Donoghue K, Byrne M, et al. Facilitators and barriers influencing weight management behaviours during pregnancy: a meta-synthesis of qualitative research. BMC Pregnancy Childbirth 2022;22(1):682 [FREE Full text] [doi: 10.1186/s12884-022-04929-z] [Medline: 36064379]
40. Ferrari RM, Siega-Riz AM, Evenson KR, Moos M, Carrier KS. A qualitative study of women's perceptions of provider advice about diet and physical activity during pregnancy. Patient Educ Couns 2013;91(3):372-377 [FREE Full text] [doi: 10.1016/j.pec.2013.01.011] [Medline: 23399436]
41. O'Brien OA, Lindsay KL, McCarthy M, McGloin AF, Kennelly M, Scully HA, et al. Influences on the food choices and physical activity behaviours of overweight and obese pregnant women: a qualitative study. Midwifery 2017;47:28-35. [doi: 10.1016/j.midw.2017.02.003] [Medline: 28236733]
42. Keely A, Cunningham-Burley S, Elliott L, Sandall J, Whittaker A. "If she wants to eat…and eat and eat…fine! It's gonna feed the baby": pregnant women and partners' perceptions and experiences of pregnancy with a BMI >40kg/m. Midwifery 2017;49:87-94. [doi: 10.1016/j.midw.2016.09.016] [Medline: 27756643]
43. Zhang CQ, Zhang RK, Chung P, Duan Y, Lau JTF, Chan DKC, et al. Promoting influenza prevention for elderly people in Hong Kong using health action process approach: study protocol. BMC Public Health 2018;18(1):1230 [FREE Full text] [doi: 10.1186/s12889-018-6146-6] [Medline: 30400790]
44. He YH, Zhao Q, Liang N, Li JS, Gao CX. Research progress on behavioral change theory in dietary behavior intervention. Chinese Nursing Research 2016;30(36):4481-4484.
45. Wiebe HW, Boulé NG, Chari R, Davenport MH. The effect of supervised prenatal exercise on fetal growth: a meta-analysis. Obstet Gynecol 2015;125(5):1185-1194. [doi: 10.1097/AOG.0000000000000801] [Medline: 25932847]
46. Román-Gálvez MR, Amezcua-Prieto C, Salcedo-Bellido I, Olmedo-Requena R, Martínez-Galiano JM, Khan KS, et al. Physical activity before and during pregnancy: a cohort study. Int J Gynaecol Obstet 2021;152(3):374-381. [doi: 10.1002/ijgo.13387] [Medline: 32978781]
47. Li L, Aris IM, Han WM, Tan KH. A promising food-coaching intervention program to achieve optimal gestational weight gain in overweight and obese pregnant women: pilot randomized controlled trial of a smartphone app. JMIR Form Res 2019;3(4):e13013 [FREE Full text] [doi: 10.2196/13013] [Medline: 31651407]
48. Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen MH, et al. Effectiveness of a smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): randomized controlled trial. JMIR Mhealth Uhealth 2021;9(3):e26091 [FREE Full text] [doi: 10.2196/26091] [Medline: 33704075]

XSL•FO
RenderX

49. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: a randomized clinical trial. Obesity (Silver Spring) 2016;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
50. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GWG: gestational weight gain
IMB: information-motivation-behavioral skills
PMT: protective motivation theory
PMT-IMB model: protection motivation theory and information–motivation–behavior skills model
RCT: randomized controlled trial
T1: first trimester
T2: second trimester
T3: third trimester

Edited by T de Azevedo Cardoso; submitted 27.12.23; peer-reviewed by S Jo; comments to author 11.09.24; revised version received 21.10.24; accepted 20.01.25; published 10.03.25.

Please cite as:
Zhou M, Wang L, Deng Y, Ge J, Zhao S, You H
Effects of a Mobile Health Intervention Based on Behavioral Integrated Model on Cognitive and Behavioral Changes in Gestational Weight Management: Randomized Controlled Trial
J Med Internet Res 2025;27:e55844
URL: https://www.jmir.org/2025/1/e55844
doi:10.2196/55844
PMID:

©Meng Zhou, Li Wang, Ying Deng, Jinjin Ge, Shiqi Zhao, Hua You. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Coach-Assisted eHealth With Group or Individual Support for Employees With Obesity: Randomized Controlled Trial on Weight, Body Composition, and Health Metrics

Siniriikka A Männistö[1,2], PsyM, MHS; Kirsi H Pietiläinen[2,3], MD, PhD; Joona Muotka[4], MSc; Laura-Unnukka Suojanen[2], MS; Raimo Lappalainen[4], PhD; Riitta Korpela[5], PhD

[1]Occupational Health Helsinki, Helsinki, Finland

[2]HealthyWeightHub, Endocrinology, Abdominal Center, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[3]Obesity Research Unit, Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[4]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

[5]Department of Pharmacology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Corresponding Author:
Siniriikka A Männistö, PsyM, MHS
HealthyWeightHub, Endocrinology, Abdominal Center, Helsinki University Hospital and University of Helsinki
Tukholmankatu 8F
Helsinki, 00290
Finland
Phone: 358 442351020
Email: siniriikka.mannisto@helsinki.fi

Abstract

Background: Acceptance and commitment therapy provides a psychobehavioral framework feasible for digital and hybrid weight loss interventions. In face-to-face studies, group-based interventions yield more favorable outcomes than individual interventions, but the effect of the intervention form has not been studied in combination with eHealth.

Objective: This study investigated whether a minimal, 3-session group or individual enhancement could provide additional benefits compared to an eHealth-only intervention when assessing weight, body composition, and laboratory metrics in a sample of occupational health patients with obesity.

Methods: This study was a randomized controlled trial with a 12-month intervention, followed by a 12-month follow-up period without additional support (March 2021 to March 2023). Recruited from occupational health care for Finnish municipal employees, 111 working-age adults with a BMI of 30-40 kg/m^2 were randomized to 1 of the 3 treatment arms: eHealth, eHealth+group, or eHealth+individual. All treatment arms received a web-administrated, coach-assisted eHealth program based on acceptance and commitment therapy, and additionally, the eHealth+group and eHealth+individual arms received 3 remotely facilitated group or individual meetings with their designated coach. The participants were assessed for weight, body composition, blood pressure, and laboratory measurements at 0-, 6-, 12-, and 24-month time points. Applying estimated means to decrease bias caused by dropouts, generalized estimating equations were used to study the differences between the 3 groups over time.

Results: There were no between-group differences in primary measurements of weight change or categorical weight change. Secondary outcomes also did not show changes attributable to the intervention arm. Across the entire sample, the total weight loss was 1.5% during the intervention, with 18% (20/111) of the participants attaining a ≥5% weight loss. Sustained at follow-up, waist circumference decreased, and high-density lipoprotein cholesterol increased slightly. The participants completed, on average, 58.6% of the eHealth program.

Conclusions: There were no differences in weight or other somatic health variables between the eHealth arm and intervention combining eHealth with minimal group or individual enhancement. Despite a modest overall weight loss, the intervention shows promise in improving body composition and metabolic health. Moving forward, further research is needed to determine if there is a threshold where face-to-face meetings provide additional benefits in hybrid interventions. Moreover, there is a need to explore for whom and under what conditions eHealth and hybrid models may be most effective.

Trial Registration: ClinicalTrials.gov NCT04785586; https://clinicaltrials.gov/study/NCT04785586

XSL•FO
RenderX

(J Med Internet Res 2025;27:e60436) doi:10.2196/60436

KEYWORDS

eHealth; weight loss; acceptance and commitment therapy; weight-neutral; Healthy Weight Coaching; occupational health; digital health; body composition; obesity; psychobehavioral; intervention; health care; metabolic health; physiological change

Introduction

The prevalence of obesity has more than doubled worldwide since the 1970s [1]. Obesity increases the risk for several metabolic, mechanical, and psychiatric conditions, attributing to 8% of deaths worldwide [2,3]. Weight loss is seen to effectively alleviate obesity-related comorbidities, and even a modest 5%-10% weight loss is seen to, for example, lower hypertension and total cholesterol, and prevent the onset of diabetes [4-6].

However, attaining even modest weight loss results is not an easy task because weight is seen to increase by 1-2 kg yearly after weight loss intervention [7]. Research has demonstrated the significant role of psychobehavioral factors in weight management success, leading to the use of psychological components in weight loss interventions [8,9]. Acceptance and commitment therapy (ACT) is a third-wave cognitive behavioral therapy aiming to increase psychological flexibility—defined by Hayes et al [10] as "the ability to contact the present moment more fully as a conscious human being, and to change or persist in behavior when doing so serves valued ends." ACT promotes engagement in health-promoting behavior patterns while enhancing commitment to values-based behaviors [11]. Although ACT primarily targets psychological well-being rather than direct weight loss, it has proven effective in supporting long-term weight management and lifestyle changes [12,13].

In the context of weight loss trials, ACT is usually delivered in a group or workshop format [14-20]. However, in recent years, there has been growing attention to digital interventions for weight loss, offering an accessible, widespread, and cost-effective alternative to traditional face-to-face treatments [21,22]. Indeed, recent systematic reviews have provided compelling evidence supporting the effectiveness of digital interventions (eHealth) in obesity treatment: without distinguishing between different theoretical frameworks, eHealth interventions have demonstrated outcomes on par with traditional face-to-face methods and have consistently outperformed control conditions [23].

Adding personalized human interaction and guidance to eHealth programs has proven to be particularly effective in achieving significant weight loss [24-27]. Furthermore, combining face-to-face meetings with digital interventions, referred to as hybrid interventions, has been shown to further enhance the effectiveness of weight loss [28-30]. ACT-based hybrid interventions are still rare, but studies combining digital ACT interventions with group meetings [31] or telephone support [32,33] have shown promising results. The impact of intervention format (group vs individual) on weight loss has not been studied in combination with eHealth, particularly in the context of ACT. However, in face-to-face trials, group-based interventions have consistently produced more favorable outcomes compared to individual approaches [34,35].

This study aimed to investigate the differences in the ACT-based 12-month eHealth intervention between eHealth, eHealth+group, and eHealth+individual delivered for occupational health patients with a further 12-month follow-up period. Weight and categorical weight change were considered the primary outcomes, and waist circumference, body composition, and laboratory measurements were the secondary outcomes of this study. We hypothesized that enhancing interventions with individual or group meetings alongside the eHealth program would lead to improved weight and health outcomes, particularly favoring the group treatment arm. Additionally, we anticipated significant changes in weight and health parameters across all treatment groups.

Methods

Study Design

This study was a randomized controlled trial with a 12-month intervention and 12-month follow-up period. The participants were allocated to 1 of the 3 treatment arms: eHealth, eHealth+group, or eHealth+individual treatment in a 1:1:1 ratio. All treatment arms received a web-administrated eHealth program, and additionally, the eHealth+group or eHealth+individual arms received 3 remotely facilitated group or individual meetings.

The eHealth program used in this study, Healthy Weight Coaching (HWC), is a 1-year treatment program based on the ACT framework, aiming to increase psychological flexibility and teach skills related to weight management [36]. The program consists of web-administrated weekly exercises integrating ACT with behavioral weight management themes (eating behavior, physical activity, sleep, and stress; Table S1 in Multimedia Appendix 1). The HWC program is readily available in Finnish public health care, and this study was designed to investigate whether minimal enhanced support in a group or individual format could improve attrition and response to the treatment.

The version of the program used in this study adopted a softer, more weight-neutral coaching style than the HWC protocol, with prioritization on value-based actions, mindfulness, and health-promoting changes instead of weight loss itself. The enhanced group and individual meetings in eHealth+group and eHealth+individual arms adhered closely to the ACT framework, incorporated key lifestyle change components of the program, and aimed to deepen the participants' understanding of the program's themes.

The participants were each assigned a personal coach who provided periodic, tailored, one-on-one written feedback and facilitated any additional meetings. All the 3 coaches in this study had master's degrees in nutrition and further education

XSL•FO
RenderX

in psychology. They met biweekly to ensure adherence to this study's protocol. The coaches worked across the 3 treatment arms, being randomly assigned to one-third of each.

Treatment Arms

- eHealth: Participants received the HWC eHealth program, which included a 20-minute kickoff call 3 weeks into the program to ensure they were progressing as expected. They also received written feedback every 2 weeks for the first 3 months and then once every 3 weeks for the remainder of the year, totaling 19 contacts.
- eHealth+group: Participants received the HWC eHealth program (as described in item 1, eHealth) and, in addition, participated in 3 remotely facilitated group meetings using video (each was 2 hours long; due to COVID-19, a transition to remote meetings was necessitated). These group meetings included brief introductions to covered topics (eg, goal setting, stress management, and self-compassion) followed by small group discussions (Table S2 in Multimedia Appendix 1). The group meetings occurred at 1, 6, and 10 months into the 12-month treatment period.
- eHealth+individual: Participants received the HWC eHealth program (as described in item 1, eHealth) and, in addition, participated in 3 remotely facilitated individual meetings using video (each was 45 minutes long). These individual meetings were customized to each participant's current needs, acknowledging successes and seeking solutions to challenging situations. The individual meetings occurred at 1, 6, and 10 months into the 12-month treatment period.

Recruitment

Participants were Helsinki city employees with class 1 and class 2 obesity [37]. Recruitment was conducted by the staff of the occupational health care and through recruitment advertisements on the city employees' communication platforms. Enrollment was carried out by research assistants from February to March 2021. Inclusion criteria required participants to have a BMI between 30 to 40 kg/m^2, the ability to participate in group or individual meetings, access to a computer with internet services, and fluency in the Finnish language. Exclusion criteria included participation in other weight loss programs, pregnancy or lactation within the past 6 months, weight changes exceeding 5 kg within the past 3 months, major medical conditions affecting safety or readiness for weight loss, and the use of weight loss medication.

Power Analysis and Sample Size

The sample size of 23 participants per treatment arm was determined by a statistician (JM) with priori power analysis using a statistical power of 0.80 with 2-sided test and 0.05 α, modeling the mean weight change and SD in previous weight loss studies (mean 3.8, SD 4.6 kg) [38]. With an expected attrition rate of 40%, the target number of participants was set at 38 per treatment arm.

Screening

Following recruitment, participants underwent a health screening on a phone by a medical doctor covering their health status and medications. Those who passed the screening were then referred for laboratory assessments and body composition measurements.

Randomization

In March 2021, participants who provided written consent and completed baseline measurements were randomized into 1 of the 3 treatment arms (see figure in the Results section). Randomization was carried out by a statistician (JM) using stratified random allocation with SPSS (IBM Corp). Participants were stratified by sex (male or female), age groups (20-35, 35-50, and 50-65 years), and BMI categories (≥30 to 35 or ≥35 to 40 kg/m^2).

Outcome Measures

The intervention period lasted for 12 months, from March 2021 to March 2022, after which a 12-month follow-up period was conducted until March 2023 without further support. Participant assessments were conducted at baseline, midintervention (6 months), postintervention (12 months), and follow-up (24 months) at Occupational Health Helsinki. The following measurements were taken:

- Height: Measured to the nearest 0.1 cm using a Gima tape height measure with a scale from 0-200 cm (Gima SpA).
- Weight: Measured after an overnight fast to the nearest 0.1 kg in light clothing, barefoot, on InBody720 or InBody770 (InBody Co Ltd), with repeated measures taken using the same equipment.
- Body composition: Fat mass, muscle mass, and visceral fat were measured with the abovementioned InBody devices.
- Waist circumference: Measured to 0.1 cm using a Hoechstmass nonextensible tape measure (Hoechstmass Balzer GmbH) at the narrowest point between the lower costal border and the umbilicus.
- Blood pressure and heart rate: Systolic and diastolic blood pressure and heart rate were measured 3 times after 10 minutes of rest using an Omron automated blood pressure monitor (Omron Corporation). The mean sum of the last 2 measurements was used in the analysis, and the measurements were repeated if necessary until the difference between the last 2 measurements was <10 mm Hg.
- Blood samples: Blood samples were collected at the laboratory of Occupational Health Helsinki after an overnight fast and analyzed for lipids (total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein [HDL] cholesterol, and triglycerides), glucose, hemoglobin A1c (HbA1c), alanine transaminase, and high-sensitivity C-reactive protein (hs-CRP).

Of the baseline characteristics (Table 1), metabolic syndrome was calculated from the laboratory and anthropometric measurements following the International Diabetes Foundation's 2006 definition [39]. Smoking was self-reported in a baseline survey. The professional role was determined using Statistics Finland's Classification of Occupations 2010 index [40]. Program usage was obtained from the log file. Medication changes were tracked every 3 months as part of the eHealth program.

XSL•FO
RenderX

Table 1. Baseline characteristics of participants by treatment arm and across the entire sample.

Characteristic	Treatment arm				
	eHealth (n=38)	eHealth+group (n=35)	eHealth+individual (n=38)	Total (N=111)	*P* value[a]
Age (years)					
Mean (SD)	52.3 (9.7)	50.6 (7.9)	49.6 (8.8)	50 (8.9)	.41
Age group, n (%)					.91
<35	2 (5)	2 (6)	3 (8)	7 (6)	
35-50	11 (29)	12 (34)	14 (37)	37 (33)	
>50	25 (66)	21 (60)	21 (55)	67 (60)	
Sex, n (%)					.53
Male	8 (21)	4 (11)	7 (18)	19 (17)	
Female	30 (79)	31 (89)	31 (82)	92 (83)	
Smoking status, n (%)					.54
Never smoker	19 (50)	14 (40)	17 (45)	50 (45)	
Former smoker	18 (47)	15 (43)	16 (42)	49 (44)	
Current smoker	1 (3)	4 (11)	5 (13)	10 (9)	
Comorbidities, n (%)					
Metabolic syndrome	28 (74)	20 (57)	25 (66)	73 (66)	.33
Type 2 diabetes	1 (3)	4 (11)	2 (5)	7 (16)	.29
Sleep apnea	8 (21)	5 (14)	2 (5)	15 (14)	.13
Mental health problem	8 (21)	5 (14)	10 (26)	23 (21)	.45
Osteoarthritis	7 (18)	9 (26)	5 (13)	21 (19)	.39
Professional role, n (%)					.79
Managers	0 (0)	2 (6)	2 (5)	4 (4)	
Specialized experts	15 (40)	9 (26)	10 (26)	34 (31)	
Experts	14 (37)	17 (49)	15 (39)	46 (41)	
Office and customer service workers	4 (11)	4 (11)	5 (13)	13 (12)	
Service and sales workers	5 (13)	2 (6)	5 (13)	12 (11)	
Construction, repair, and manufacturing workers	0 (0)	1 (3)	1 (3)	2 (2)	
Anthropometric measurements, mean (SD)					
Height (cm)	168.9 (10.4)	165.5 (7)	167.1 (8.4)	167.2 (8.8)	.26
Weight (kg)	97.6 (12.8)	94.1 (11)	95.8 (12.9)	95.9 (12.3)	.47
BMI (kg/m^2)	34.2 (2.7)	34.3 (2.9)	34.2 (3.1)	34.2 (2.9)	.98
Waist circumference (cm)	112.1 (8)	111.5 (8.7)	109.9 (9.7)	111.2 (8.8)	.53
Body composition, mean (SD)					
Visceral fat (cm^2)	202.8 (37.8)	200.5 (31.7)	202.1 (36.1)	201.8 (35.1)	.96
Fat percentage (%)	42.9 (6.5)	43.3 (5.2)	43.4 (6)	43.2 (5.9)	.92
Muscle percentage (%)	31.8 (4.1)	31.5 (3.2)	31.5 (3.7)	31.6 (3.7)	.92
Laboratory measurements, mean (SD)					
Hemoglobin A_{1c} (mmol/mol)	38.4 (4.8)	38.6 (7.3)	39.8 (10.1)	38.9 (7.7)	.71
Glucose (mmol/L)	6.1 (0.8)	5.9 (0.9)	6.3 (2.1)	6.1 (1.4)	.55
Cholesterol (mmol/L)	5.3 (1.1)	5.4 (1.1)	5.3 (1)	5.4 (1)	.84

XSL•FO
RenderX

Characteristic	Treatment arm				
	eHealth (n=38)	eHealth+group (n=35)	eHealth+individual (n=38)	Total (N=111)	*P* value[a]
High-density lipoprotein (mmol/L)	1.3 (0.4)	1.3 (0.3)	1.3 (0.3)	1.3 (0.3)	.81
Low-density lipoprotein (mmol/L)	3.1 (0.8)	3.4 (1)	3.3 (0.8)	3.3 (0.9)	.25
Triglycerides (mmol/L)	2.1 (1.4)	1.6 (0.6)	1.4 (0.7)	1.7 (1)	.006
Alanine aminotransferase (U/L)	43.4 (36.1)	29.2 (14.4)	34.7 (17)	36 (25.2)	.05
High-sensitivity C-reactive protein (mg/L)	2.9 (2.1)	2.8 (2.5)	2.8 (2.5)	2.8 (2.4)	.96
Blood pressure					
Systolic (mm Hg)	129.7 (14.5)	127.3 (11.5)	126.7 (14.2)	127.9 (13.4)	.60
Diastolic (mm Hg)	88.4 (8.3)	86.8 (8)	86.6 (9.2)	87.3 (8.5)	.59
Pulse (beats per minute)	72.4 (10.4)	69.4 (9.2)	72.5 (8.9)	71.5 (9.5)	.30

[a]*P* values indicate the difference between the 3 treatment arms with the categorical variables analyzed by chi-square test and the continuous variables by ANOVA.

Statistical Analysis

Data analysis was conducted using IBM SPSS Statistics (version 28.0.0.0; IBM Corp). Baseline differences were analyzed by chi-square for the categorical variables and ANOVA for the continuous variables. Differences in mean values between the measurement points and the treatment arms were analyzed using the generalized estimating equations (GEEs) method [41]. The GEE method includes all the available data points and uses estimated means in the analysis, applying the intention-to-treat (ITT) analysis concept to decrease bias caused by dropouts. The article only reports results from the ITT analysis, while completer analyses are available in Multimedia Appendix 1.

Considering GEEs, we first used a model examining the interaction between the measurement points and treatment arms, with post hoc analysis comparing the effects of the eHealth+group and the eHealth+individual arms to the eHealth arm. For further analysis, we used a single-arm model, concentrating on the changes observed across the 4 measurement points in the entire study sample. After the primary analysis, we added baseline weight, sex, age, and applicable medication or medication change (eg, blood pressure medication in systolic and diastolic blood pressure analysis) to the single-arm model to explain the results. The changes in medication are described in Table S3 in Multimedia Appendix 1. Weight, BMI, waist circumference, HbA_{1c}, cholesterol, HDL, low-density lipoprotein, systolic blood pressure, diastolic blood pressure, and pulse rate were normally distributed. Glucose, triglycerides, alanine transaminase, hs-CRP, and muscle percentage had a positively skewed distribution, and thus, gamma distribution with log link function was used in the GEEs analysis. Visceral fat and fat percentage had a negatively skewed distribution, and thus, inverted variables were used in the analysis with gamma distribution and log link function.

Crosstabs with Pearson chi-square was used when analyzing the group level differences of categorical weight variable (participants were divided into 6 categories based on weight loss percentage: <0, 0-2.9, 3-4.9, 5-5.9, 5-9.9, or ≥10), and change in metabolic syndrome prevalence between the measurement points. Categorical weight loss and metabolic syndrome prevalence are reported for the entire baseline sample, with the last observation carried forward. Changes in muscle and fat percentage, as well as metabolic syndrome incidence, are expressed as percentage points (%-point), describing the arithmetic difference between the 2 percentages.

Spearman correlation coefficient was used to analyze correlations between the weight change percentages (nonparametric). Dancey and Reidy's definition was used when determining the strength of the correlation [42].

Cohen *d* effect size was calculated using the square root of the average variance of measures. In the analysis comparing the treatment arms, we used an effect size calculator developed by the statistician (JM) that adjusts for baseline differences. For the single-arm analysis, we used the University of Colorado Springs Effect Size Calculator [43].

Figures were drawn with GraphPad Prism (GraphPad Software).

Ethical Considerations

This study's protocol was approved by the Ethics Committee of the Helsinki and Uusimaa Hospital District (April 29, 2020, HUS/922/2020) and conducted per the ethical standards of the Declaration of Helsinki. Informed consent was obtained from all individuals included in this study. To ensure privacy and confidentiality, study data were pseudonymized and securely processed, preventing any association with individual participants. Participants received the intervention and individual health data but received no additional compensation. Neither the participants nor the data assessors were blinded to the treatment arms.

Results

Baseline Characteristics

Of the 143 people assessed for eligibility, 111 were randomized into research groups (Figure 1). Most participants were female (n=92, 83%) and worked as experts (n=46, 41%) or specialized experts (n=34, 31%; Table 1). More than half were either former or current smokers (n=59, 53%). Of the 111 participants, 73

XSL•FO
RenderX

(66%) had metabolic syndrome, 23 (21%) had mental health problems, 21 (19%) had osteoarthritis, 7 (16%) had type 2 diabetes, and 15 (14%) had sleep apnea. The mean age was 50 (SD 8.9) years, and the mean BMI was 34.2 (SD 2.9) kg/m^2. Baseline characteristics did not differ between the groups except for triglycerides (P=.006), which showed statistically significant elevation in the eHealth arm compared to the other two treatment arms.

Figure 1. The flow of participants through the 3 treatment arms (eHealth, eHealth+group, and eHealth+individual) of the randomized controlled trial, with 6 participants' 24-month data excluded from the analysis due to obesity medication use. All available data were included in the overall analysis.

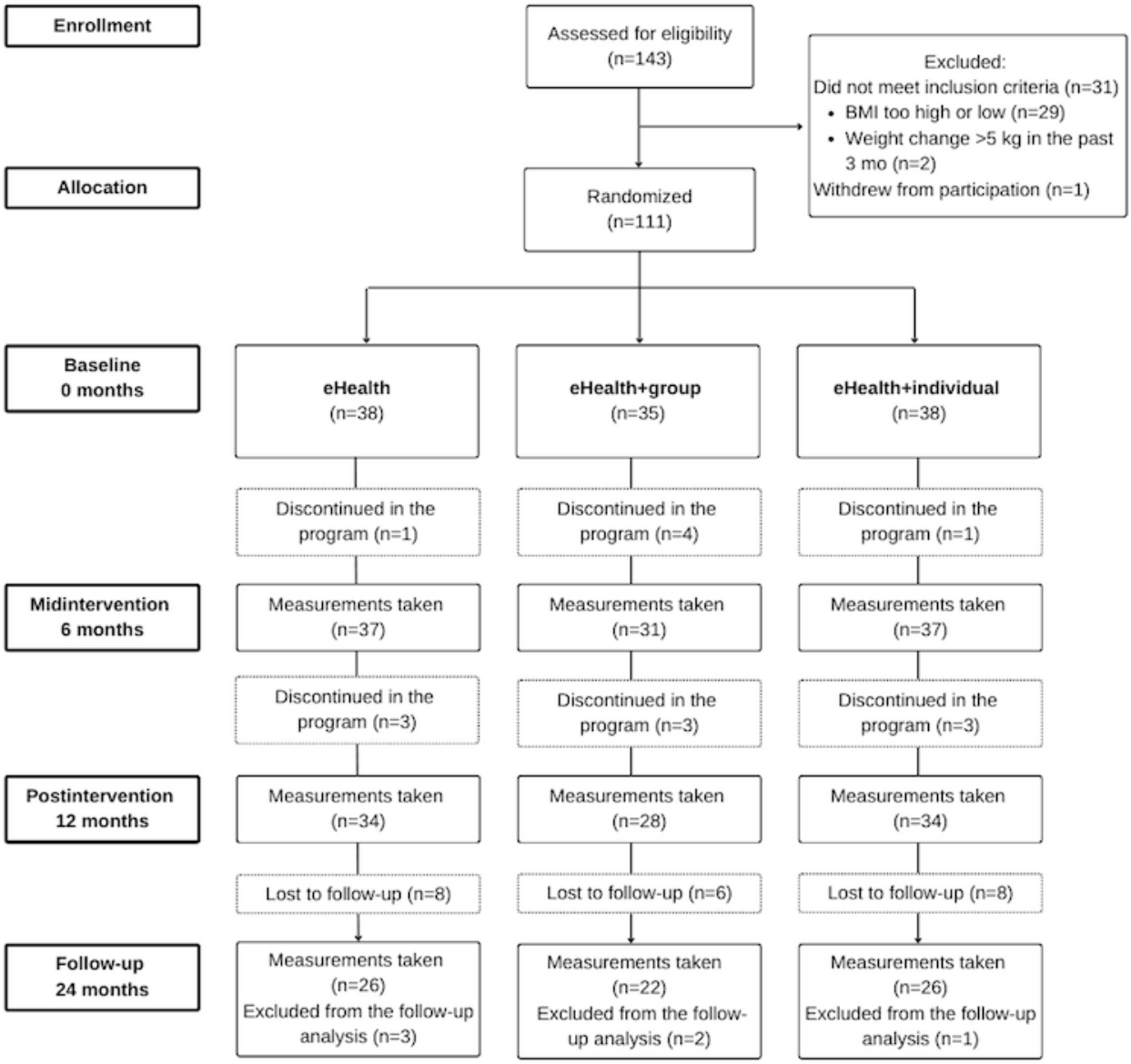

Retention

The retention rate was 95% at the 6-month mark, 86% at the 12-month mark, and 61% at the 24-month mark of the intervention (Figure 1). The participants completed, on average, 58.6% (SD 32.7%) of the eHealth program. There were no differences in retention or progression between the treatment arms.

Intervention Effects

We did not observe any significant between-group differences in changes in the primary outcomes of weight and categorized weight (P=.76 and P=.99, respectively; Table 2, and Tables S4 and S5 in Multimedia Appendix 1). However, we did discover an interaction between the time points and treatment arms for HDL (P=.03) and triglycerides (P=.002).

XSL•FO
RenderX

Table 2. Estimated mean changes (SE) in anthropometric, body composition, laboratory, and blood pressure measurements from baseline to 6, 12, and 24 months across treatment arms and all participants (N=111), analyzed using generalized estimating equations. The overall results are presented in the title row, while differences between treatment arms are shown in the corresponding rows.

Variables	0-6 months	0-12 months	0-24 months	Overall change across 0, 6, 12, and 24 months[a]	
	Mean change (SE)	Mean change (SE)	Mean change (SE)	Wald chi-square	*P* values
Weight metrics					
Weight (kg)	–1.25 (0.35)	–1.4 (0.44)	–0.7 (0.61)	13.69	.003
eHealth	–0.83 (0.7)	–0.8 (0.81)	–1.25 (1.18)	3.36	.76
eHealth+group	–1.43 (0.7)	–1.75 (0.87)	–0.18 (0.99)	3.36	.76
eHealth+individual	–1.5 (0.39)	–1.69 (0.62)	–1.06 (1.06)	3.36	.76
BMI (kg/m^2)	–0.44 (0.13)	–0.49 (0.15)	–0.29 (0.22)	13.09	.004
eHealth	–0.23 (0.25)	–0.24 (0.26)	–0.47 (0.42)	3.84	.70
eHealth+group	–0.55 (0.25)	–0.65 (0.31)	–0.11 (0.36)	3.84	.70
eHealth+individual	–0.54 (0.77)	–0.62 (0.23)	–0.41 (0.39)	3.84	.70
Body composition					
Waist circumference (cm)	–2.34 (0.35)	–2.65 (0.42)	–2.74 (0.54)	53.69	.001
eHealth	–2.2 (0.59)	–2.18 (0.79)	–4.07 (0.94)	6.32	.39
eHealth+group	–2.2 (0.73)	–2.2 (0.73)	–2.84 (0.67)	6.32	.39
eHealth+individual	–2.48 (0.50)	–2.93 (0.66)	–3.12 (0.85)	6.32	.39
Visceral fat (cm^2)	–6.22 (3.54)	–5.82 (3.95)	–2.45 (4.01)	17.68	<.001
eHealth	–5.13 (6.23)	–3.54 (6.5)	–3.77 (6.84)	5.33	.50
eHealth+group	–5.48 (5.42)	–6.92 (6.57)	0.89 (6.6)	5.33	.50
eHealth+individual	–7.98 (6.5)	–7.23 (7.35)	–5.44 (7.29)	5.33	.50
Fat percentage	–0.75 (0.59)	–0.9 (0.64)	–0.29 (0.6)	18.2	<.001
eHealth	–0.77 (1.07)	–0.64 (1.08)	–0.46 (1.1)	5.28	.51
eHealth+group	–0.46 (0.94)	–0.86 (1)	–0.08 (0.92)	5.28	.51
eHealth+individual	–1 (1.05)	–1.21 (1.19)	0.62 (1.11)	5.28	.51
Muscle percentage	0.4 (0.11)	0.46 (0.14)	0.32 (0.25)	13.98	.003
eHealth	0.47 (0.22)	0.37 (0.23)	0.13 (0.29)	4.2	.65
eHealth+group	0.22 (0.2)	0.45 (0.27)	0.78 (0.68)	4.2	.65
eHealth+individual	0.47 (0.16)	0.58 (0.22)	0.28 (0.3)	4.2	.65
Laboratory measurements					
Hemoglobin A_{1c}(mmol/L)	–0.38 (0.32)	0.62 (0.34)	0.07 (0.4)	10.89	.01
eHealth	–0.39 (0.56)	0.65 (0.55)	0.19 (0.61)	5.49	.48
eHealth+group	0.3 (0.59)	0.42 (0.72)	0.22 (0.76)	5.49	.48
eHealth+individual	–0.97 (0.5)	0.72 (0.51)	0.03 (0.74)	5.49	.48
Glucose (mmol/L)	–0.34 (0.06)	–0.35 (0.06)	–0.24 (0.06)	45.59	<.001
eHealth	–0.5 (0.1)	–0.37 (0.1)	–0.18 (0.14)	9.49	.15
eHealth+group	–0.21 (0.11)	–0.39 (0.12)	–0.35 (0.11)	9.49	.15
eHealth+individual	–0.3 (0.08)	–0.3 (0.09)	–0.21 (0.08)	9.49	.15
Cholesterol (mmol/L)	–0.11 (0.08)	–0.13 (0.08)	–0.11 (0.11)	2.89	.41
eHealth	–0.24 (0.11)	–0.24 (0.13)	–0.22 (0.17)	6.8	.34
eHealth+group	–0.01 (0.19)	–0.14 (0.16)	–0.23 (0.22)	6.8	.34

XSL•FO
RenderX

Variables	0-6 months	0-12 months	0-24 months	Overall change across 0, 6, 12, and 24 months[a]	
	Mean change (SE)	Mean change (SE)	Mean change (SE)	Wald chi-square	*P* values
eHealth+individual	–0.07 (0.09)	–0.008 (0.1)	0.13 (0.16)	6.8	.34
High-density lipoprotein (mmol/L)	0.05 (0.02)	0.07 (0.02)	0.11 (0.02)	29.58	<.001
eHealth	0.04 (0.07)	0.05 (0.08)	0.17 (0.07)	14.23	.03
eHealth+group	0.06 (0.04)	0.11 (0.04)	0.09 (0.04)	14.23	.03
eHealth+individual	–0.01 (0.02)	0 (0.03)	0.13 (0.04)	14.23	.03
Low-density lipoprotein (mmol/L)	–0.04 (0.07)	–0.13 (0.07)	–0.11 (0.1)	4	.26
eHealth	–0.11 (0.10)	–0.14 (0.12)	–0.14 (0.15)	6.16	.41
eHealth+group	0.01 (0.17)	–0.2 (0.16)	–0.27 (0.2)	6.16	.41
eHealth+individual	–0.02 (0.08)	–0.06 (0.1)	0.09 (0.15)	6.16	.41
Triglycerides (mmol/L)	–0.21 (0.07)	–0.12 (0.07)	–0.21 (0.07)	13.34	.004
eHealth	–0.47 (0.12)	–0.42 (0.14)	–0.46 (0.11)	20.87	.002
eHealth+group	–0.09 (0.15)	–0.06 (0.13)	–0.18 (0.13)	20.87	.002
eHealth+individual	–0.05 (0.07)	0.13 (0.06)	0.02 (0.1)	20.87	.002
Alanine aminotransferase (U/L)	–4.6 (1.88)	–3.77 (2.12)	–3.08 (2.2)	6.57	.09
eHealth	–6.04 (4.41)	–5.35 (4.81)	–4.16 (5.06)	2.37	.88
eHealth+group	–3.13 (2.45)	–1.01 (2.87)	–3.59 (2.25)	2.37	.88
eHealth+individual	–4.69 (2.18)	–4.74 (2.66)	–1.29 (3.39)	2.37	.88
High-sensitivity C-reactive protein (mg/L)	0.28 (0.19)	0.16 (0.27)	0.83 (0.3)	9.05	.03
eHealth	0.48 (0.39)	–0.02 (0.35)	1.69 (0.63)	8.39	.21
eHealth+group	0.34 (0.29)	0.4 (0.56)	0.9 (0.53)	8.39	.21
eHealth+individual	0 (0.28)	0.09 (0.49)	0.06 (0.39)	8.39	.21
BP[b]					
Systolic BP (mm Hg)	–2.88 (1.12)	–2.22 (1.32)	–0.96 (1.7)	7.67	.053
eHealth	–4.15 (2.22)	–3.82 (2.35)	–1.09 (3.06)	3.81	.7
eHealth+group	–3.27 (1.78)	–1.36 (2.61)	0.97 (4.03)	3.81	.7
eHealth+individual	–1.24 (1.68)	–1.3 (1.91)	–2.13 (1.88)	3.81	.7
Diastolic BP (mm Hg)	–2.37 (0.67)	–2.33 (0.65)	–1.15 (0.99)	17.64	<.001
eHealth	–3.18 (1.25)	–3 (1.22)	–2.4 (1.81)	5.13	.53
eHealth+group	–2.34 (0.93)	–1.38 (1.01)	1.62 (1.95)	5.13	.53
eHealth+individual	–1.52 (1.17)	–2.38 (1.08)	–1.68 (1.23)	5.13	.53
Pulse (beats per minute)	–0.44 (0.7)	0.06 (0.75)	0.89 (0.79)	2.23	.53
eHealth	–0.35 (1.14)	1.11 (1.25)	1.53 (1.21)	2.77	.84
eHealth+group	0.06 (1.2)	0.04 (1.18)	2.31 (1.45)	2.77	.84
eHealth+individual	–0.71 (1.3)	–0.59 (1.41)	–0.48 (1.46)	2.77	.84

[a]Degrees of freedom=6 for the between-group analysis and degrees of freedom=3 for the single-group analysis, including all participants.

[b]BP: blood pressure.

Considering HDL, further analysis showed a greater increase in the eHealth arm compared to the eHealth+individual arm between 0 and 6 months (*P*=.09, *d*=0.35) and 0-12 months (*P*=.03, *d*=0.28). The interaction between the time points and treatment arms was explained by sex (*P*=.002), cholesterol medication use (*P*<.001), and changes in cholesterol medication (*P*<.001). When participants who were using cholesterol medication at baseline or had changes in their medication were

XSL•FO
RenderX

excluded, the interaction between the treatment arm and time point was insignificant (*P*=.17). This suggests that the difference in HDL between the treatment arms was likely due to variations in medication use.

For triglycerides, the eHealth arm demonstrated a greater decrease than the eHealth+individual arm at all time points relative to baseline (0-6 months *P*=.03, *d*=0.41; 0-12 months *P*<.001, *d*=0.53; 0-24 months *P*=.002, *d*=0.47; Table 2). However, the eHealth arm had higher triglyceride levels at baseline compared to the other groups (Table 1).

To study the overall effect of the intervention, the 3 groups were considered as a single eHealth group and combined for further analysis, as reported below.

Overall Impact

Weight

At the 6- and 12-month mark of the intervention, participants experienced an estimated mean weight loss of 1.3 kg (1.4%, *P*<.001) and 1.4 kg (1.5%, *P*=.002), respectively, compared to baseline (ITT analysis presented in Table 2 and Figure 2, and completer analysis in Table S6 in Multimedia Appendix 1). Overall, weight showed a significant decrease over the 24-month study period (*P*=.003) when analyzed across all time points (0, 6, 12, and 24 months).

Figure 2. Weight loss (%) and SE midintervention (6 months), postintervention (12 months), and at follow-up (24 months) compared to baseline (0 months) displaying estimated values used by the generalized estimating equations.

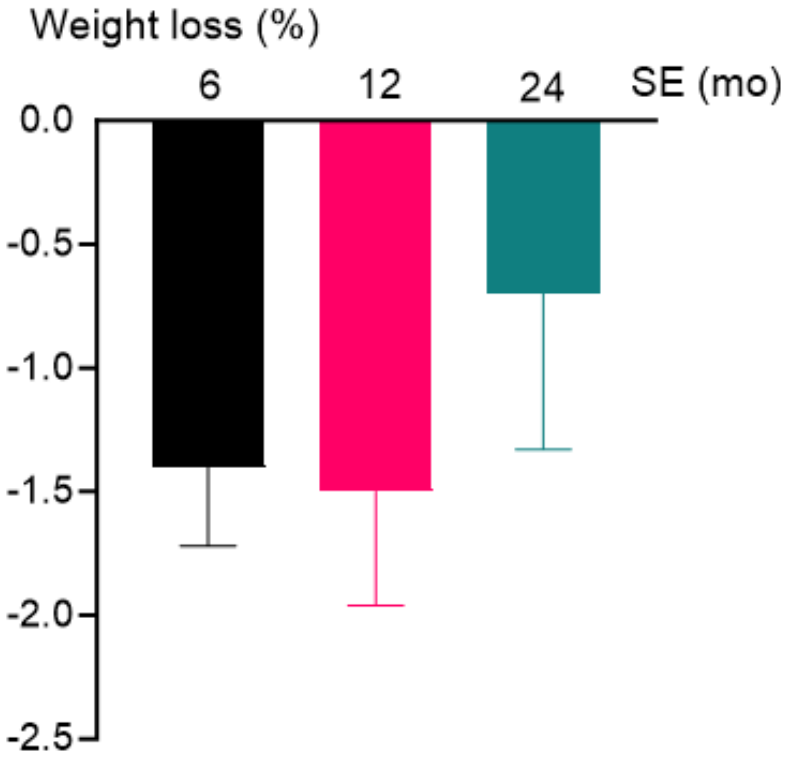

Using the entire sample of 111 participants and applying the last observation carried forward principle, categorical analysis at the 6-month time point revealed that 15 (14%) participants had lost 3%-4.9% of their baseline body weight, 12 (11%) participants had lost 5%-9.9%, and 1 (1%) participant had lost ≥10% (Figure 3). At 12 months, 5 (5%) participants had lost 3%-4.9% of their baseline body weight, 15 (14%) participants lost 5.2%-9.9%, and 5 (5%) participants lost ≥10%. Similarly, at the 24-month follow-up, the respective proportions in each category were 7 (6%) participants, 15 (14%) participants, and 6 (5%) participants. Overall, 20 of 111 (18%) participants achieved a clinically relevant weight loss of ≥5% of their initial weight at the 12-month mark, increasing to 21 (19%) participants at the 24-month mark.

XSL•FO
RenderX

Figure 3. Categorized weight loss (%) midintervention (6 months), postintervention (12 months), and at follow-up (24 months; N=111) with last observation carried forward.

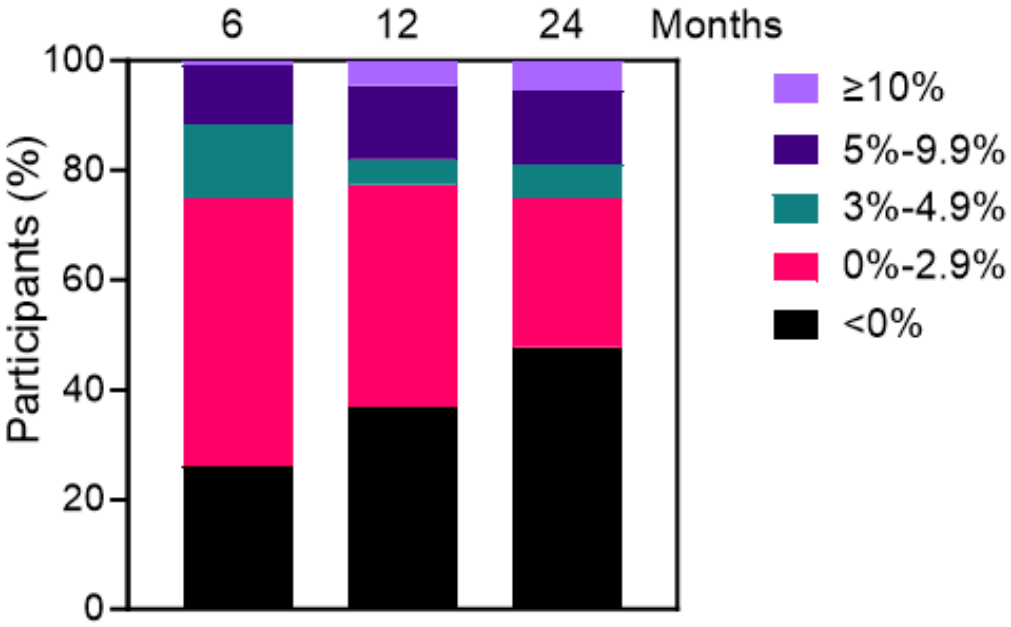

The weight changes were correlated between different measurement points. We observed strong correlations between the 0-6–month weight change and 0-12–month weight change (r=0.64, P<.001), and between the 0-12–month weight change and 0-24–month weight change (r=0.66, P<.001). The correlation was moderate between the 0-6–month weight change and 0-24–month weight change (r=0.324, P=.008).

Waist Circumference and Body Composition

The estimated mean decrease in waist circumference was 2.4 cm (SD 0.35 cm; P<.001) and 2.6 cm (SD 0.42 cm; P<.001) at the 6- and 12-month marks of the intervention, respectively (Table 2, Table S6 in Multimedia Appendix 1). At follow-up, the waist circumference continued to decrease, showing a 2.8 cm mean decrease from 0 to 24 months (SD 0.54 cm; P<.001).

The estimated mean decrease in visceral fat was 6.2 cm^2 (SD 3.54 cm^2; P<.001) and 5.8 cm^2 (SD 3.95 cm^2; P=.003) at the 6- and 12-month marks of the intervention, respectively (Table 2, Table S6 in Multimedia Appendix 1). Fat percentage decreased by 0.8% (SD 0.59%, P<.001) and 0.9% (SD 0.64%, P<.001) at the 6- and 12-month marks. Muscle percentage increased by 0.4% (0.11%, P<.001) and 0.5% (SD 0.14%, P<.001) at the 6- and 12-month marks.

The covariate analysis showed a significant interaction for baseline weight with changes in waist circumference, visceral fat, fat percentage, and muscle percentage (Table S7 in Multimedia Appendix 1). Post hoc analysis revealed that participants with lower baseline weight had a greater reduction in waist circumference, visceral fat, fat percentage, and a greater increase in muscle percentage. There was also a significant covariate interaction between age and waist circumference, with older participants having a greater reduction in waist circumference.

Laboratory Measurements and Blood Pressure

The estimated fasting glucose decrease was 0.3 mmol/L at 6 months (P<.001), 0.4 at 12 months (P<.001), and 0.2 (P<.001) at 24 months (Table 2 and Table S6 in Multimedia Appendix 1). There was a significant interaction between diabetes medicine usage and glucose level change, and changes in diabetes medication usage and glucose level change, with greater glucose level reductions observed for participants who had existing medication or who were prescribed new medication to manage blood glucose levels (Table S6 in Multimedia Appendix 1). However, the results of the main analysis remained after excluding these participants (n=9) from the analysis (0-12–month P<.001; 0-24–month P=.002).

There was a statistically significant overall change in HbA_{1c} (P=.01), with the estimated mean value increase of 0.62 mmol/L at 12 months and 0.07 at 24 months. However, these changes were not significant compared to the baseline. Changes in HbA_{1c} levels were significantly associated with changes in diabetes medication (P<.001), with participants who reduced their diabetes medication (n=12) during the intervention showing a greater reduction in HbA_{1c} levels than those who increased or started new medications or who made no changes. However, the main result remained unchanged after excluding the participants with increased or new diabetes medication (n=24).

The HDL increase was 0.1 mmol/L at the 6-month (P=.007), 12-month (P<.001), and 24-month (P<.001) marks of the intervention (Table 2 and Table S6 in Multimedia Appendix 1). The triglycerides decrease was 0.2 mmol at both 6-months (P=.002) and 24-months (P=.004). There was a statistically significant interaction between the changes in cholesterol medicine with HDL (P<.001) and triglycerides (P<.001), but the result of the main analysis remained after excluding these participants (n=2) from the analysis (P<.001 and P=.01, respectively; Table S7 in Multimedia Appendix 1).

XSL•FO
RenderX

The hs-CRP increased by 0.9 mg/L from baseline to 24-month follow-up (P=.004) when outliers with severe elevation of >30 mg/L were excluded (n=4; Table 2 and Table S6 in Multimedia Appendix 1). The result of the pairwise comparison remained significant (P=.01) when lowering the outlier threshold to 10 mg/L (marked elevation). Diastolic blood pressure decrease was 2.4 mm Hg at the 6-month (P<.001) and 2.3 mm Hg at the 12-month (P<.001) marks of the intervention. Covariate analysis flagged a significant interaction between changes in blood pressure medication and diastolic blood pressure (P<.001; Table S7 in Multimedia Appendix 1). Nevertheless, the results persisted (P=.002) even after excluding participants (n=6) who had medication increase or initiation.

Metabolic syndrome prevalence was 66% at baseline (Table 1) and 60% in both 12-month postintervention and 24-month follow-up, indicating a change of -9%. These changes were not statistically significant.

Discussion

Principal Findings

This study examined whether adding minimal individual or group-based face-to-face support to a 12-month coach-assisted eHealth intervention enhances outcomes by strengthening connections with coaches or leveraging peer support. Contrary to our hypothesis, no significant differences in primary outcomes were found between the groups, indicating that 3 live sessions—whether individual or group-based—did not provide additional benefits over the eHealth protocol alone. Interestingly, the eHealth-only group showed greater improvements in HDL cholesterol and triglycerides compared to the individual session group. However, post hoc analyses attributed these differences to cholesterol medication use and higher baseline triglyceride levels rather than the intervention itself. Thus, incorporating infrequent live sessions into the eHealth protocol does not appear to offer additional advantages.

In the entire study population, the intervention resulted in a mean weight loss of 1.5% from baseline to 12-month postintervention and 0.7% from baseline to 24-month follow-up. When considering individual weight change, the results are polarized: the received support promoted weight loss for some participants, while others did not respond. The weight changes were notably correlated between different measurement points, suggesting the individual weight change to be rather consistent.

Throughout the intervention and follow-up, there was a progressive decrease in mean waist circumference, along with improvements in the body composition variables at the 12-month mark.

Additionally, there was a slight enhancement in HDL cholesterol levels at postintervention and follow-up, suggesting positive effects on lipid metabolism. Although we observed a decrease in the fasting glucose levels during the intervention and follow-up, similar results were not observed in HbA_{1c}, a marker measuring the average blood glucose levels over the past few months. This indicates that the observed improvements in glucose levels might not be consistent or lasting enough to affect overall glycemic control.

Unexpectedly, the hs-CRP increased significantly by the 24-month follow-up, suggesting elevated inflammation not attributable to outliers but potentially reflecting mild infections or systemic inflammation [44]. We do not attribute the increase in the hs-CRP to heightened metabolic disturbances, as we noted a slight, though statistically insignificant, improvement in the prevalence of metabolic syndrome.

Comparison With Prior Work

There are several potential explanations for the lack of intergroup differences. It is possible that only 3 sessions in the hybrid conditions do not yield a sufficient difference compared to the coach-assisted eHealth intervention, and by increasing the number of sessions in the hybrid intervention, differences between the groups could be observed. In Teeriniemi et al [29], participants with a BMI over 30 kg/m^2 lost an average of 4.4% of their weight in 12 months when they received a hybrid intervention combining eHealth with 8 group meetings based on cognitive behavioral therapy, compared to just 1.6% weight loss in the eHealth-only arm. Therefore, a more intensively enhanced hybrid approach may be needed to further improve somatic outcomes.

Notably, the intervention took place during the COVID-19 era, leading to all group and individual meetings being conducted remotely. As seen in previous research on therapeutic relationships [45], this shift could have hindered the alliance between the participants and the coach. Particularly, the unplanned move to remote group meetings in the eHealth+group treatment arm may have affected the group's cohesion and perceived peer support [46], thereby impacting effectiveness and attenuating differences between the treatment arms.

Overall, it is likely that differences in weight change between the 3 study arms are attenuated by the lack of power resulting from smaller-than-expected changes in weight, with high variability. The total weight loss is modest compared to the 2019 meta-analysis by Lawlor [12], reporting a mean absolute weight change of 5.5 kg in third-wave cognitive behavior therapies. Furthermore, the weight loss outcomes observed in our study fall short in comparison to the results reported from previous implementations of the HWC eHealth program, which achieved a 4.6% weight loss during the 12-month intervention [47]. Additionally, a "review of reviews" by Kupila et al [23] reported mean differences in weight loss results ranging from –0.12 (95% CI –0.64 to 0.41) kg in eHealth interventions compared with face-to-face care to –4.32 (–5.08 to –3.57) kg in eHealth interventions compared with no care. However, our results align with the 12-month weight loss of –2.08 kg reported by Mueller et al [33] for another ACT-based hybrid intervention (eHealth+phone support) conducted during the COVID-19 pandemic.

Several factors could contribute to the lower weight loss observed in our study. First, our study used a more weight-neutral coaching style than the one in Kupila et al [47], which may have impacted its effectiveness in promoting weight loss compared to the earlier version. It needs to be noted that a newer version of the HWC is already in clinical use. Therefore,

XSL•FO
RenderX

the results of our study do not demonstrate the current effectiveness of the HWC.

Second, our study involved a different population characterized by lower BMI compared to the study of Kupila et al [47]. In their study, the greatest weight loss was observed among individuals with a baseline BMI ≥40 kg/m², a subgroup that was not included in our study. Yet, in Kupila et al [47], the mean weight loss for those with a baseline BMI <40 kg/m² was 3.2%—more than the weight change observed here.

Third, the influence of the COVID-19 pandemic coinciding with the 2021-2022 intervention cannot be overlooked, as research suggests that the pandemic led to disruptions in lifestyle behaviors that could affect weight management. The mean weight gain during the pandemic was higher [48] than average [49], indicating difficulty in weight maintenance. These effects are evident in the Finnish population: between 2017 and 2023, the prevalence of obesity increased 3%-points in men and 4%-points in women [50]. In contrast, the prevalence of obesity had remained unchanged between 2012 and 2017 [51]. In addition, an increase in food consumption and sedentary activity was observed during the pandemic [52]—changes that were potentially intensified in this study population, most of whom worked in expert positions and shifted to remote work during the pandemic.

Lastly, our study involved a different population characterized by significant occupational burden from mentally consuming expert work. This occupational burden may have posed challenges to weight management despite participation in the intervention, especially with the combination of the eHealth program and remote knowledge work. This is per the earlier results showing that early weight loss predicts 1-year weight loss in web-based programs [53] and shows that the early weight loss could even predict long-term weight maintenance.

In the secondary analysis with the overall sample, long-term improvements in waist circumference and lipid metabolism are consistent with findings from previous nondiet studies [54-57]. Additionally, the increase in muscle mass contrasts with previous findings of skeletal muscle mass loss in weight loss studies [58]. These changes are likely attributed to our intervention's focus on guiding participants toward making healthy dietary choices and increasing physical activity without imposing calorie restrictions. However, our findings of no improvement in long-term glucose metabolism contrast with previous studies showing that lifestyle weight loss interventions reduce the risk of elevated fasting glucose levels and HbA_{1c} [59-61], suggesting that a more prominent weight loss may be necessary to achieve sustained improvements in glycemic control.

Strengths and Limitations

This study presents both strengths and limitations. A key strength is the long follow-up period, which allowed for the investigation of the treatment's long-term effects—rare in typical ACT-based interventions. Additionally, the comprehensive array of outcome measures, including body composition and laboratory tests alongside traditional metrics such as weight and waist circumference, offers a holistic understanding of the intervention's effects. This study also ensured accuracy and reliability by conducting all measurements and laboratory tests on-site despite the onset of COVID-19. Additionally, the emergence of new weight loss medications was addressed by screening and excluding participants using weight loss medications, requesting that participants refrain from starting new weight loss medications during the intervention, regularly collecting information on any changes in medication, and excluding from the analysis those who began using them during the follow-up period.

However, some limitations impact especially the interpretation of the overall study effects. First, the participant group was homogeneous—all employed and primarily middle-aged women in expert professions—making generalization to other populations challenging. Second, the lack of blinding and participants' awareness of their treatment arms before randomization could have affected motivation, particularly for those not assigned to their preferred group. Third—affecting interpretation of the overall results—the absence of an inactive control group prevents comparison against a baseline, making it difficult to isolate the treatment effects from external influences or placebo effects and thus establish causality. These limitations, arising from the integration of this study within routine occupational health care services and the need for ethical service provision to willing participants, highlight important considerations for future research aiming to balance real-world applicability with rigorous study design.

Conclusions

Our study found no differences in weight or other somatic health variables between the eHealth arm and intervention combining eHealth with minimal group or individual enhancement. Moving forward, further research is needed to determine if there is a threshold where face-to-face meetings provide additional benefits in hybrid interventions. Moreover, there is a need to explore for whom and under what conditions eHealth and hybrid models may be most effective.

Acknowledgments

We wish to thank all the personnel involved in conducting the trial: coaches Susanne Räsänen, MHS, and Anita Saranpää, PhD; physiotherapists in charge of this study's measurements Minna Mäkelä, BHC, and Mariann Slotte-Jana, BHC; laboratory staff of Occupational Health Helsinki, Jussi Mustonen, MD, and other medical doctors participating in this study; the supportive management of Occupational Health Helsinki Ritva Teerimäki, MD, Eevamaija Tuovinen, MD, and Marko Karvinen, MSc; and Merja Sokka, BHC, and Jaana Ahoranta, BHC, from the Healthy Weight Hub for always being willing to help. This study was supported by Finnish Work Environment Fund and Helsinki City (grant 200379); Juho Vainio Foundation (grant 202400066); Academy of Finland (grants 335443, 314383, and 266286); the Academy of Finland, Centre of Excellence in Research on

XSL•FO
RenderX

Mitochondria, Metabolism and Disease (FinMIT; grant 272376); Finnish Medical Foundation; Gyllenberg Foundation; Novo Nordisk Foundation, (grant NNF20OC0060547, NNF17OC0027232, and NNF10OC1013354); Finnish Diabetes Research Foundation; Paulo Foundation; Occupational Health Helsinki, City of Helsinki, University of Helsinki and Helsinki University Hospital; and government research funds. Open access to this paper was funded by the University of Helsinki Library. The funders had no role in study design, data collection and analysis, decision to publish, or paper preparation. Generative artificial intelligence was not used to formulate this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SAM did the conceptualization, data curation, funding acquisition, investigation, project administration, formal analysis, writing of the original draft, and reviewed and edited the writing. JM handled the methodology, formal analysis, review, and editing. LUS carried out the conceptualization, project administration, funding acquisition, resources, and review and editing of the writing. RL worked on the conceptualization, supervision, and review and editing of the writing. KHP assisted with the conceptualization, project administration, funding acquisition, resources, visualization, supervision, and review and editing of the writing. RK aided in the conceptualization, funding acquisition, supervision, and review and editing of the writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental tables.
[DOCX File , 48 KB - jmir_v27i1e60436_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1173 KB - jmir_v27i1e60436_app2.pdf]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
2. Dixon JB. The effect of obesity on health outcomes. Mol Cell Endocrinol 2010;316(2):104-108. [doi: 10.1016/j.mce.2009.07.008] [Medline: 19628019]
3. Ritchie H, Roser M. Obesity. Our World in Data. 2017. URL: https://ourworldindata.org/obesity [accessed 2025-02-12]
4. Wassertheil-Smoller S. The trial of antihypertensive interventions and management (TAIM) study. Arch Intern Med 1992;152(1):131. [doi: 10.1001/archinte.1992.00400130141018]
5. Dattilo A, Kris-Etherton P. Effects of weight reduction on blood lipids and lipoproteins: a meta-analysis. Am J Clin Nutr 1992;56(2):320-328. [doi: 10.1093/ajcn/56.2.320] [Medline: 1386186]
6. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001;344(18):1343-1350. [doi: 10.1056/NEJM200105033441801] [Medline: 11333990]
7. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]
8. Byrne SM, Cooper Z, Fairburn CG. Psychological predictors of weight regain in obesity. Behav Res Ther 2004;42(11):1341-1356. [doi: 10.1016/j.brat.2003.09.004] [Medline: 15381442]
9. Karhunen L, Lyly M, Lapveteläinen A, Kolehmainen M, Laaksonen DE, Lähteenmäki L, et al. Psychobehavioural factors are more strongly associated with successful weight management than predetermined satiety effect or other characteristics of diet. J Obes 2012;2012:274068 [FREE Full text] [doi: 10.1155/2012/274068] [Medline: 22792447]
10. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
11. Lillis J, Kendra KE. Acceptance and commitment therapy for weight control: model, evidence, and future directions. J Contextual Behav Sci 2014;3(1):1-7 [FREE Full text] [doi: 10.1016/j.jcbs.2013.11.005] [Medline: 25419510]

XSL•FO
RenderX

12. Lawlor ER, Islam N, Bates S, Griffin SJ, Hill AJ, Hughes CA, et al. Third-wave cognitive behaviour therapies for weight management: a systematic review and network meta-analysis. Obes Rev 2020;21(7):e13013 [FREE Full text] [doi: 10.1111/obr.13013] [Medline: 32181957]
13. Yıldız E. The effects of acceptance and commitment therapy on lifestyle and behavioral changes: a systematic review of randomized controlled trials. Perspect Psychiatr Care 2020;56(3):657-690. [doi: 10.1111/ppc.12482] [Medline: 32043617]
14. Cattivelli R, Guerrini Usubini A, Manzoni GM, Vailati Riboni F, Pietrabissa G, Musetti A, et al. ACTonFood. Acceptance and commitment therapy-based group treatment compared to cognitive behavioral therapy-based group treatment for weight loss maintenance: an individually randomized group treatment trial. Int J Environ Res Public Health 2021;18(18):9558 [FREE Full text] [doi: 10.3390/ijerph18189558] [Medline: 34574482]
15. Forman EM, Butryn ML, Hoffman KL, Herbert JD. An open trial of an acceptance-based behavioral intervention for weight loss. Cognitive and Behavioral Practice 2009;16(2):223-235. [doi: 10.1016/j.cbpra.2008.09.005]
16. Forman EM, Butryn ML, Juarascio AS, Bradley LE, Lowe MR, Herbert JD, et al. The mind your health project: a randomized controlled trial of an innovative behavioral treatment for obesity. Obesity (Silver Spring) 2013;21(6):1119-1126 [FREE Full text] [doi: 10.1002/oby.20169] [Medline: 23666772]
17. Lillis J, Hayes SC, Bunting K, Masuda A. Teaching acceptance and mindfulness to improve the lives of the obese: a preliminary test of a theoretical model. Ann Behav Med 2009;37(1):58-69. [doi: 10.1007/s12160-009-9083-x] [Medline: 19252962]
18. Lillis J, Niemeier HM, Ross KM, Thomas JG, Leahey T, Unick J, et al. Weight loss intervention for individuals with high internal disinhibition: design of the acceptance based behavioral intervention (ABBI) randomized controlled trial. BMC Psychol 2015;3(1):17 [FREE Full text] [doi: 10.1186/s40359-015-0075-2] [Medline: 26019869]
19. Niemeier HM, Leahey T, Reed KP, Brown RA, Wing RR. An acceptance-based behavioral intervention for weight loss: a pilot study. Behav Ther 2012;43(2):427-435 [FREE Full text] [doi: 10.1016/j.beth.2011.10.005] [Medline: 22440077]
20. Tapper K, Shaw C, Ilsley J, Hill AJ, Bond FW, Moore L. Exploratory randomised controlled trial of a mindfulness-based weight loss intervention for women. Appetite 2009;52(2):396-404 [FREE Full text] [doi: 10.1016/j.appet.2008.11.012] [Medline: 19101598]
21. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto Silva AP, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the LoseIt reddit community. J Med Internet Res 2017;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
22. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GLD, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
23. Kupila SKE, Joki A, Suojanen L, Pietiläinen KH. The effectiveness of eHealth interventions for weight loss and weight loss maintenance in adults with overweight or obesity: a systematic review of systematic reviews. Curr Obes Rep 2023;12(3):371-394 [FREE Full text] [doi: 10.1007/s13679-023-00515-2] [Medline: 37354334]
24. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
25. Berry MP, Sala M, Abber SR, Forman EM. Incorporating automated digital interventions into coach-delivered weight loss treatment: a meta-analysis. Health Psychol 2021;40(8):534-545. [doi: 10.1037/hea0001106] [Medline: 34618500]
26. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR mHealth uHealth 2022;10(4):e35479 [FREE Full text] [doi: 10.2196/35479] [Medline: 35394443]
27. O'Boyle J, Davidson P. The effects of mHealth versus eHealth on weight loss in adults: a systematic review. Top Clin Nutr 2022;37(1):69-84. [doi: 10.1097/tin.0000000000000268]
28. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
29. Teeriniemi A, Salonurmi T, Jokelainen T, Vähänikkilä H, Alahäivälä T, Karppinen P, et al. A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. J Intern Med 2018;284(5):534-545 [FREE Full text] [doi: 10.1111/joim.12802] [Medline: 29974563]
30. Besson M, Gurviez P, Carins J. Using digital devices to help people lose weight: a systematic review. JSOCM 2020;10(3):289-319. [doi: 10.1108/jsocm-07-2019-0115]
31. Järvelä-Reijonen E, Karhunen L, Sairanen E, Muotka J, Lindroos S, Laitinen J, et al. The effects of acceptance and commitment therapy on eating behavior and diet delivered through face-to-face contact and a mobile app: a randomized controlled trial. Int J Behav Nutr Phys Act 2018;15(1):22 [FREE Full text] [doi: 10.1186/s12966-018-0654-8] [Medline: 29482636]

XSL•FO
RenderX

32. Levin M, Petersen J, Durward C, Bingeman B, Davis E, Nelson C, et al. A randomized controlled trial of online acceptance and commitment therapy to improve diet and physical activity among adults who are overweight/obese. Transl Behav Med 2021;11(6):1216-1225. [doi: 10.1093/tbm/ibaa123] [Medline: 33289785]
33. Mueller J, Richards R, Jones RA, Whittle F, Woolston J, Stubbings M, et al. Supporting weight management during COVID-19 (SWiM-C): twelve-month follow-up of a randomised controlled trial of a web-based, ACT-based, guided self-help intervention. Int J Obes (Lond) 2023;47(1):51-59 [FREE Full text] [doi: 10.1038/s41366-022-01232-x] [Medline: 36369513]
34. Abbott S, Smith E, Tighe B, Lycett D. Group versus one-to-one multi-component lifestyle interventions for weight management: a systematic review and meta-analysis of randomised controlled trials. J Hum Nutr Diet 2021;34(3):485-493. [doi: 10.1111/jhn.12853] [Medline: 33368624]
35. Street S, Avenell A. Are individual or group interventions more effective for long-term weight loss in adults with obesity? A systematic review. Clin Obes 2022;12(5):e12539. [doi: 10.1111/cob.12539] [Medline: 35765718]
36. Suojanen L, Ahola A, Kupila S, Korpela R, Pietiläinen KH. Effectiveness of a web-based real-life weight management program: study design, methods, and participants' baseline characteristics. Contemp Clin Trials Commun 2020;19:100638 [FREE Full text] [doi: 10.1016/j.conctc.2020.100638] [Medline: 32885090]
37. Adult BMI categories. CDC. URL: https://www.cdc.gov/bmi/adult-calculator/bmi-categories.html [accessed 2024-08-30]
38. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
39. Saklayen MG. The global epidemic of the metabolic syndrome. Curr Hypertens Rep 2018;20(2):12 [FREE Full text] [doi: 10.1007/s11906-018-0812-z] [Medline: 29480368]
40. Classification of Occupations. Statistics Finland. 2010. URL: https://stat.fi/en/luokitukset/ammatti/ammatti_1_20100101 [accessed 2025-02-14]
41. Ziegler A, Vens M. Generalized estimating equations. Notes on the choice of the working correlation matrix. Methods Inf Med 2010;49(5):421-425. [doi: 10.3414/ME10-01-0026] [Medline: 20871939]
42. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
43. Effect size calculators. University of Colorado Colorado Springs. URL: https://lbecker.uccs.edu/ [accessed 2025-02-12]
44. Nehring SM, Goyal A, Patel B. StatPearls. National Library of Medicine.: StatPearls Publishing; 2025. URL: http://www.ncbi.nlm.nih.gov/books/NBK441843 [accessed 2025-02-12]
45. Rotger JM, Cabré V. Therapeutic alliance in online and face-to-face psychological treatment: comparative study. JMIR Ment Health 2022;9(5):e36775 [FREE Full text] [doi: 10.2196/36775] [Medline: 35499910]
46. Lopez A, Rothberg B, Reaser E, Schwenk S, Griffin R. Therapeutic groups via video teleconferencing and the impact on group cohesion. mHealth 2020;6:13 [FREE Full text] [doi: 10.21037/mhealth.2019.11.04] [Medline: 32270005]
47. Kupila SKE, Venäläinen MS, Suojanen L, Rosengård-Bärlund M, Ahola AJ, Elo LL, et al. Weight loss trajectories in healthy weight coaching: cohort study. JMIR Form Res 2022;6(3):e26374 [FREE Full text] [doi: 10.2196/26374] [Medline: 35262494]
48. Anderson LN, Yoshida-Montezuma Y, Dewart N, Jalil E, Khattar J, De Rubeis V, et al. Obesity and weight change during the COVID-19 pandemic in children and adults: a systematic review and meta-analysis. Obes Rev 2023;24(5):e13550. [doi: 10.1111/obr.13550] [Medline: 36721999]
49. Hutfless S, Gudzune KA, Maruthur N, Wilson RF, Bleich SN, Lau BD, et al. Strategies to prevent weight gain among adults. Effective Health Care (EHC) Program. 2013. URL: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/weight-gain-prevention_research.pdf [accessed 2025-02-12]
50. Terveyden ja hyvinvoinnin laitos. URL: https://www.thl.fi/tervesuomi_verkkoraportit/ilmioraportit_2023/ts_bmi_4lk.html [accessed 2025-02-12]
51. Työikäisten ylipainon ja lihavuuden kasvu näyttää hidastuneen Kansallisen FINRISKI-terveystutkimuksen tuloksia [Website in Finnish]. Lääkärilehti. 2015. URL: https://www.laakarilehti.fi/tieteessa/alkuperaistutkimukset/tyoikaisten-ylipainon-ja-lihavuuden-kasvu-nayttaa-hidastuneen-kansallisen-finriski-terveystutkimuksen-tuloksia/?public=6d0b74b4c4f936ef87e164f647f57b1b [accessed 2025-02-12]
52. Catucci A, Scognamiglio U, Rossi L. Lifestyle changes related to eating habits, physical activity, and weight status during COVID-19 quarantine in Italy and some European countries. Front Nutr 2021;8:718877 [FREE Full text] [doi: 10.3389/fnut.2021.718877]
53. Unick JL, Leahey T, Kent K, Wing RR. Examination of whether early weight loss predicts 1-year weight loss among those enrolled in an internet-based weight loss program. Int J Obes (Lond) 2015;39(10):1558-1560 [FREE Full text] [doi: 10.1038/ijo.2015.89] [Medline: 25982792]
54. Rapoport L, Clark M, Wardle J. Evaluation of a modified cognitive-behavioural programme for weight management. Int J Obes Relat Metab Disord 2000;24(12):1726-1737. [doi: 10.1038/sj.ijo.0801465] [Medline: 11126231]
55. Bacon L, Aphramor L. Weight science: evaluating the evidence for a paradigm shift. Nutr J 2011;10:9 [FREE Full text] [doi: 10.1186/1475-2891-10-9] [Medline: 21261939]

XSL•FO
RenderX

56. Mensinger JL, Calogero RM, Stranges S, Tylka TL. A weight-neutral versus weight-loss approach for health promotion in women with high BMI: a randomized-controlled trial. Appetite 2016;105:364-374. [doi: 10.1016/j.appet.2016.06.006] [Medline: 27289009]
57. Bacon L, Stern JS, Van Loan MD, Keim NL. Size acceptance and intuitive eating improve health for obese, female chronic dieters. J Am Diet Assoc 2005;105(6):929-936. [doi: 10.1016/j.jada.2005.03.011] [Medline: 15942543]
58. McCarthy D, Berg A. Weight loss strategies and the risk of skeletal muscle mass loss. Nutrients 2021;13(7):2473 [FREE Full text] [doi: 10.3390/nu13072473] [Medline: 34371981]
59. Jiang Q, Li JT, Sun P, Wang LL, Sun L, Pang SG. Effects of lifestyle interventions on glucose regulation and diabetes risk in adults with impaired glucose tolerance or prediabetes: a meta-analysis. Arch Endocrinol Metab 2022;66(2):157-167 [FREE Full text] [doi: 10.20945/2359-3997000000441] [Medline: 35289514]
60. Zhang X, Imperatore G, Thomas W, Cheng YJ, Lobelo F, Norris K, et al. Effect of lifestyle interventions on glucose regulation among adults without impaired glucose tolerance or diabetes: a systematic review and meta-analysis. Diabetes Res Clin Pract 2017;123:149-164 [FREE Full text] [doi: 10.1016/j.diabres.2016.11.020] [Medline: 28024276]
61. Yang J, Xia Y, Sun Y, Guo Y, Shi Z, Cristina do Vale Moreira N, et al. Effect of lifestyle intervention on HbA1c levels in overweight and obese adults with type 2 diabetes across ethnicities: a systematic review and meta-analysis of randomized controlled trials. Diabetes Res Clin Pract 2023;199:110662 [FREE Full text] [doi: 10.1016/j.diabres.2023.110662] [Medline: 37028602]

Abbreviations

%-point: percentage point
ACT: acceptance and commitment therapy
GEE: generalized estimating equation
HbA1c: hemoglobin A1c
HDL: high-density lipoprotein
hs-CRP: high-sensitivity C-reactive protein
HWC: Healthy Weight Coaching
ITT: intention-to-treat analysis

Edited by N Cahill; submitted 10.05.24; peer-reviewed by L Paulos-Guarnieri, A Notenbomer; comments to author 13.08.24; revised version received 11.09.24; accepted 29.01.25; published 12.03.25.

Please cite as:
Männistö SA, Pietiläinen KH, Muotka J, Suojanen LU, Lappalainen R, Korpela R
Coach-Assisted eHealth With Group or Individual Support for Employees With Obesity: Randomized Controlled Trial on Weight, Body Composition, and Health Metrics
J Med Internet Res 2025;27:e60436
URL: https://www.jmir.org/2025/1/e60436
doi:10.2196/60436
PMID:40073400

©Siniriikka A Männistö, Kirsi H Pietiläinen, Joona Muotka, Laura-Unnukka Suojanen, Raimo Lappalainen, Riitta Korpela. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating Perceptions of the CANreduce 2.0 eHealth Intervention for Cannabis Use: Focus Group Study

Daniel Folch-Sanchez[1*], BSc, MSc; Maria Pellicer-Roca[2*], BSc, MSc; María Agustina Sestelo[2], BSc, MD; Paola Zuluaga[3], BSc, MD; Francisco Arias[4], BSc, MD; Pablo Guzmán Cortez[1], BSc, MD; Salma Amechat[1], BSc; Gustavo Gil-Berrozpe[5], BSc, MSc, PhD; Estefania Lopez Montes[6,7,8], BSc; Clara Mercadé[6,7,8], BSc; Francina Fonseca[6,7,8], BSc, MD, PhD; Laia Miquel[1], BSc, MD, PhD; Joan I Mestre-Pintó[2], BSc, MSc, PhD

[1]Health and Addictions Research Group, Addictions Unit, Psychiatry and Psychology Service, Institut Clínic de Neurociències (ICN), Hospital Clinic Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[2]Hospital del Mar-Research Institute and Universitat Pompeu Fabra, Neuroscience Research Program, Addiction Research Group (GRAd), Barcelona, Spain

[3]Hospital Universitari Germans Trias i Pujol, Unidad de Medicina Interna, Badalona, Spain

[4]Hospital Universitario 12 de Octubre, Instituto de Investigación i+12, Madrid, Spain

[5]Hospital Universitario de Navarra, Navarra, Spain

[6]Parc de Salut Mar, Institut de Neuropsiquiatria i Adiccions (INAD), Barcelona, Spain

[7]Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain

[8]Department of Medicine and Life Sciences (MELIS), Universitat Pompeu Fabra, Barcelona, Spain

[*]these authors contributed equally

Corresponding Author:
Laia Miquel, BSc, MD, PhD
Health and Addictions Research Group, Addictions Unit
Psychiatry and Psychology Service, Institut Clínic de Neurociències (ICN)
Hospital Clinic Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)
c villarroel 170
Barcelona
Spain
Phone: 34 932275400 ext 1719
Email: miquel@clinic.cat

Abstract

Background: Cannabis is the most widely used illicit drug, and admissions for cannabis use disorders (CUDs) are increasing globally, posing a significant public health challenge. Despite its negative consequences, a substantial proportion of individuals with problematic use do not seek treatment. In recent years, digital health interventions (DHIs) have emerged as accessible and cost-effective solutions, empowering users to manage their health care. CANreduce is one such eHealth intervention that has demonstrated effectiveness in reducing cannabis use (CU); however, its suboptimal adherence rates underscore the need for strategies to enhance user engagement and motivation.

Objective: This study aims to improve the effectiveness, adherence, and user experience of the Spanish version of CANreduce 2.0 by employing focus groups (FGs) within a user-centered design approach that actively involves both users and professionals.

Methods: Separate FGs were conducted for users and professionals, involving a total of 10 participants. Users were recruited from individuals registered on the CANreduce 2.0 platform and active cannabis users, while professionals comprised addiction specialists familiar with the platform. Each session was held remotely and moderated by 2 interviewers following a semistructured script. Qualitative analysis of the transcripts was performed using MAXQDA software and content analysis methodology to identify key themes related to the acceptability, usability, and utility of CANreduce 2.0.

Results: The qualitative analysis identified 3 main themes, encompassing 15 subcodes. Within the “motivation and awareness” theme, both users (n=6, mean age 31.8 years, SD 4.1 years) and professionals (n=4, mean age 37.25 years, SD 1.71 years) frequently discussed the importance of “motivation” and “problem awareness” as crucial for the success of CANreduce 2.0. In the “guidance and use” theme, the subcode “complement to face-to-face therapy” was the most emphasized. Professionals supported CANreduce 2.0 as a valuable adjunct to in-person therapy, serving as both an educational and monitoring tool, with

XSL•FO
RenderX

no objections raised by either group. Lastly, within the "content and design" theme, "information," "small achievements," and "personalized content" emerged as key areas for improvement, highlighting the need to enhance motivation and adherence through gamification and tailored content.

Conclusions: Personalization, robust motivational strategies, and an engaging, interactive design are essential for the success of DHIs, particularly in addiction treatment. Collaboration among technology developers, health care professionals, and users should be central to the development process, fostering the cocreation of practical and effective solutions that are responsive to the needs of those seeking treatment. This approach ensures that DHIs are not only functional but also widely accepted and impactful. Insights from this study will inform the ongoing refinement of CANreduce 2.0, enhancing its relevance and effectiveness in addressing CU.

(J Med Internet Res 2025;27:e65025) doi:10.2196/65025

KEYWORDS

addiction; cannabis; drug use disorder; eHealth; digital health intervention; qualitative research; focus groups; user-centered design; user-centered intervention

Introduction

Cannabis is the most widely used drug worldwide, posing a significant problem due to its negative effects on users' well-being and its broader impact on public health. According to the latest report from the United Nations Office on Drugs and Crime [1], approximately 228 million people consumed cannabis in 2022, with many countries reporting a sustained increase in the prevalence of cannabis use (CU), hospitalizations, and diagnosed psychiatric disorders related to CU.

In Spain, cannabis is the illicit substance with the highest prevalence, particularly among younger individuals. The latest data estimate that 19.1% of individuals under 35 years of age [2] and 21.8% of students aged 14-18 years used cannabis in the past year, with the average age of initiation being 15 years [3]. The National Strategy on Addictions [4] places special emphasis on the normalization of CU among youth and the challenges this poses, given that early onset correlates not only with an increased risk of developing cannabis use disorder (CUD) but also with other mental health conditions. Furthermore, early use is associated with academic and occupational difficulties [1]; involvement in conflicts, fights, or physical aggression [3,5]; and engagement in risky sexual behaviors [3,6].

As a result, the total number of admissions for CU treatment is increasing [2,7], with 93.7% of individuals under 18 years initiating treatment for illicit drug use in Spain doing so for CU [2]. Spain has a well-established, extensive, diversified, and accessible network for addressing drug-related issues. Among the available treatments, motivational interviewing and cognitive behavioral therapies (CBTs) have proven effective in managing substance use disorders (SUDs) as therapeutic and psychoeducational strategies [8-10], both in-person and digitally [11,12].

However, despite the availability and variety of treatments, a significant proportion of problematic cannabis users do not seek help, often due to social stigma, shame, lack of time, or the belief that they can regulate their consumption on their own [13,14]. This issue is particularly pronounced among younger individuals, who not only seek treatment less frequently but also exhibit lower adherence rates [15].

Technological progress in the 21st century is driving the digitalization of everyday life, with screens becoming the primary medium for communication and self-expression among young people [16]. In response to this shift, the development of mobile apps and internet-based interventions for health promotion and treatment has grown steadily over the past 2 decades, offering patients accessible and cost-effective solutions that empower them to take control of their health care journey [17]. In the context of mental health and SUDs, these tools can function as (1) adjuncts to clinical care and (2) strategies for health promotion, enhancing accessibility to treatment.

Multiple meta-analyses have examined the use of digital health interventions (DHIs) for addressing SUDs [18], particularly CUDs [15,19-22]. Evidence from these studies demonstrates a favorable yet modest effect of digital health strategies, with greater efficacy in prevention programs compared with intervention programs. However, both approaches show a lower impact than face-to-face interventions [19], highlighting the need to develop strategies that integrate both modalities [10,21,22].

CANreduce is one of the eHealth tools proven effective in reducing CU. It is a minimally guided, internet-based self-help intervention developed by Schaub et al [23] and has been certified as safe by the European Union, qualifying as a medical device under directives 93/42/EWG and 2007/47/EWG. This interactive online platform consists of 8 modules designed to be completed over 6 weeks, integrating CBT strategies, motivational interviewing, and immediate ecological interventions.

In 2015, the first version of the app proved to be an effective alternative to face-to-face treatment, demonstrating reductions in the frequency and quantity of CU, increased completion of self-help modules, and higher abstention rates, although adherence remained suboptimal [23]. Despite the flexibility DHIs offer, allowing them to be tailored to individual user needs, poor adherence remains a common challenge with this type of tool [17,24]. Adherence is a complex, dynamic process influenced by individual factors such as motivation and self-efficacy, interpersonal elements such as therapeutic alliances and shared decision-making, and therapy-related aspects, including preplanned interventions and medication regimens

XSL•FO
RenderX

[25]. Some of these factors—particularly interpersonal elements—are typically absent from digital interventions such as CANreduce, as they generally do not provide clinician support. Effective strategies to address this issue and improve adherence to digital treatments [26], particularly for CU [20], include incorporating support from a mental health professional or coach [27] and using participation reminders [28].

Following the principles of Adherence-Focused Guidance [29], CANreduce 2.0 was developed and tested in a subsequent clinical trial. This new edition incorporated support from either an "e-Coach" or a team without a specific reference individual, with both interventions proving equally effective in reducing CU frequency and mental health symptoms [30]. In 2022, the app was adapted to Spanish and validated for use within the Spanish population. Data from the corresponding clinical trial are currently being analyzed, and its protocol has been previously published [31]. The adaptation included adding videos featuring Laura, the Spanish e-Coach, along with 6 characters of different ages, genders, and backgrounds, illustrating various situations of cannabis abuse and dependence throughout the modules.

The World Health Organization has emphasized the importance of using a user-centered design in the development of digital tools, noting that most people do not re-engage with an app if it does not initially attract them [32]. Previous research has shown that patients will only use new technologies if they are relevant to their health concerns, visually appealing, easy to use, and effective in facilitating behavior change [17]. Increasing patient involvement in the design of DHIs is essential for creating relevant, usable, and effective solutions and has been linked to greater acceptability and motivation for use [33,34], thereby improving user experience and perceived value [35]. Motivation is a critical component of behavior change, helping predict patient abstinence and reductions in substance use. In the context of SUD and AUD treatment, it is a multidimensional, fluid state that enables individuals to resolve their ambivalence about making difficult changes to avoid substance abuse [25,36]. It has consistently been associated with increased treatment adherence and completion, making it a key construct in the Adherence-Focused model [25]. While motivation is intrinsic to patients, interventions can help them identify their reasons and need for change, as well as facilitate the planning required to achieve it [36].

User-centered design is a multiphase process that involves conducting focus groups (FGs) or interviews with users, modifying designs and content based on qualitative findings, and continuously iterating these steps with additional user and professional input [17,35,37]. Qualitative methods enable an in-depth evaluation by uncovering attitudes, perceptions, and beliefs that influence interaction with the tool, while also assessing user needs and preferences related to interface and usability [38,39].

A co-design approach that actively involves users has a significant impact on health outcomes and provides numerous benefits, including improved adherence [25], fidelity, and reach of digital tools [40]. Additionally, it fosters a broader range of ideas, enhances usability [41,42], and increases user satisfaction, support, and enthusiasm for innovation. It also improves decision-making efficiency and strengthens the patient-clinician relationship [17,35].

This study aims to enhance the effectiveness, adherence, and user experience of the Spanish version of CANreduce 2.0 through a user-centered design approach. As the initial phase of a larger research project, its findings on user perspectives will inform subsequent stages to improve the development of a DHI for CU treatment.

Methods

Recruitment

Discussion groups were conducted separately for users and professionals. For the users' FG, participants were recruited from those who had registered on the CANreduce 2.0 platform within the past 2 years. Following established guidelines [43], the initial FG aimed to include 6-12 participants. Twelve individuals were contacted by phone and invited to participate; however, only 3 attended, necessitating a second FG session. The second group also included 3 participants—active cannabis users attending treatment for the first time—who were instructed to use the CANreduce 2.0 platform before participating in the discussions. These participants were recruited through snowball sampling and contacted by phone to schedule the second online FG. The professional FG comprised 4 mental health and addiction specialists with experience in using digital tools. All professionals were members of the Primary Care Network for Addictions (RIAPAd) and were involved in promoting the CANreduce 2.0 improvement study for future implementation in addiction treatment centers. These participants were invited via email.

In total, 3 FG sessions were conducted: 2 for users (n=6) and 1 for professionals (n=4). Eligibility criteria for both groups included regular use of mobile devices and no diagnosis of a severe mental illness. Users were required to actively use the CANreduce 2.0 platform and report current CU, while professionals needed prior familiarity with the platform. Each FG session was facilitated by 2 interviewers (DFS and JIMP) and conducted remotely via Zoom (Zoom Communications, Inc./Qumu Corporation) to accommodate participants' schedules and geographical constraints. Audio recordings of the sessions were made to facilitate transcription.

Focus Group Interviews

The categories evaluated during the participatory groups were determined through a deductive analysis of usability guidelines focused on user-centered design and a review of scientific literature on digital health apps. Special consideration was given to strategies for motivating change [44,45].

The FGs aimed to elicit expert and participant feedback on both the content and form of the app (interface and materials), as well as the utility of the intervention (assessment of usage and experience). Specific objectives were established: defining the platform's target population, clarifying its specific purposes, updating and enriching available content, optimizing interface design, and promoting user motivation and adherence. Based on these objectives, a series of questions were formulated, with

XSL•FO
RenderX

slight adaptations in each session depending on whether participants were professionals or users. Each FG session lasted approximately 90 minutes.

Semistructured interview scripts were used, allowing for the inclusion of emerging questions relevant to the stated objectives. The questions were designed to align with the study's purpose and objectives, with some formulated to broadly explore each category (eg, "Do you think there are any missing topics or areas?") and others aimed at eliciting specific feedback on particular aspects of the intervention (eg, "Do you think using a mobile app format could improve user adherence to treatment?"). Depending on the depth of information provided in each category, either broad or focused questions were applied. To enhance user motivation within the CANreduce platform, options such as increased reminder frequency and positive reinforcement (eg, congratulatory messages) were explored. These examples served as probes, used only when participants did not provide spontaneous responses, to avoid directing their feedback. Interview guides for the professional and user groups can be found in Multimedia Appendices 1 and 2, respectively, with both having been translated from Spanish to English for publication purposes.

Statistical Analysis

The qualitative analysis of FG transcriptions was conducted using MAXQDA (VERBI GmbH), a software that enabled the systematic coding and analysis of textual data from user transcripts. The software's coding functionality was used to assign text segments to various codes, representing different themes relevant to the research.

This analysis followed the framework of content analysis, a methodological approach that involves identifying and defining categories or codes to capture key topics of interest. This approach enabled researchers to efficiently manage qualitative data, ensuring that the findings were grounded in participants' experiences and perspectives.

Ethics Approval

The study was approved by the Ethics Committee of CEIm-Parc de Salut Mar/IMIM-Hospital del Mar (approval code 2019/8901/I). Before the FGs, informed consent was obtained from all participants. At the start of each session, participants were briefed on key ethical considerations, including confidentiality and voluntary participation, to ensure an ethical and transparent process. Audio recordings were made to facilitate transcription, with no names recorded to maintain confidentiality. Data were fully anonymized throughout the analysis. As compensation, each participant received a €15 (US $15.63) Amazon voucher. No identifying information about participants is included in the manuscript or multimedia appendices.

Results

Demographics

The professional FG included 2 men and 2 women, all with higher education and an average age of 37.25 (SD 1.71) years. Half were psychologists (n=2) and the other half were doctors (n=2). They worked in a detoxification hospitalization unit (n=1), an outpatient addiction center (n=2), and an acute psychiatric hospitalization unit (n=1).

Among users (n=6), the 3 participants who had received online treatment through the platform were Spanish women—2 with technical training and 1 with higher education. The new users (n=3) were all Spanish men with higher education. The mean age of users (n=6) was 31.8 (SD 4.1) years.

Finally, both users and professionals had owned smartphones for more than 10 years, except for 1 professional who had owned a smartphone for 7 years. All participants also possessed at least one other electronic device, such as a laptop or tablet.

Themes

Overview

The codes identified in the qualitative analysis of the transcripts were categorized into 3 main themes: "motivation and awareness," "guidance and use," and "content and design," each containing specific subcodes. For "motivation and awareness," the subcodes included "user age," "motivation," "problem awareness," and "economic awareness." Within "guidance and use," the subcodes comprised "virtual treatment" (further divided into "in favor" and "against"), virtual support from Laura (subdivided into "in favor," "against," and the importance of the "therapist's gender" on the platform), and "complement to in-person therapy" (with "in favor" and "against" subcategories). For the "content and design" theme, the identified subcodes were "app" "notifications," "information," "small achievements and gamification," and the availability of "personalized content."

A comprehensive scheme of all codes can be found in Multimedia Appendix 3. The transcription of participant responses, organized according to the aforementioned codes, is available in Multimedia Appendix 4. Codes, themes, and transcriptions were translated from Spanish into English solely for publication purposes.

To analyze the distribution of codes across FG transcripts, a figure from MAXQDA's Code Matrix Browser was created (Figure 1). Larger symbols in the matrix indicate a higher number of coded segments assigned to a specific code for each group, highlighting key differences and similarities in how users and professionals prioritize different topics.

XSL•FO
RenderX

Figure 1. Code distribution across groups.



In the "motivation and awareness" theme, the subcode "user age" was addressed by all groups, with users from the first FG (registered on CANreduce 2.0 within the past 2 years) discussing it slightly more. "Motivation" was mentioned most prominently by professionals, followed by active cannabis users in the second FG (first-time treatment participants) and the first users' group. Users in the second group prioritized "problem awareness," discussing it more frequently than users in the first group and professionals, while also being the only group to address "economic awareness."

In the "guidance and use" theme, all groups discussed the subcode "virtual treatment," with varying degrees of emphasis. Professionals most frequently highlighted the potential benefits of CANreduce 2.0 as an online treatment for reducing CU but also expressed the greatest concerns about its drawbacks. Active cannabis users in the second FG mentioned some concerns, while users in the first group did not raise objections. The subcode "virtual support from Laura" was addressed equally across all groups, with balanced perspectives on its potential advantages and disadvantages. By contrast, "complement to face-to-face therapy" was primarily explored by professionals, who discussed it extensively and favorably.

In "content and design," users in the first FG referenced subcodes such as "app" and "notifications" the most, emphasizing the importance of usability and engagement tools. However, "information" was the most discussed subcode within this theme across all groups, reflecting its high priority for both users and professionals. Professionals led the discussion on "small achievements and gamification" and devoted the most attention to "personalized information," more so than either user group.

Motivation and Awareness

User Age

Users had mixed opinions on the platform's target population age. Some felt it was better suited for young people due to its online format and emphasis on less invasive treatment methods, which might appeal to those who have not yet recognized the severity of their consumption habits. Others believed it was more appropriate for adults over 30 years, who are more likely to be aware of their behavior and the consequences of CU.

Professionals noted that the content was designed to reflect young people's experiences but pointed out some comments they found "slightly uncomfortable" or "potentially sexist."

Motivation

Users emphasized the importance of self-motivation for the platform's success, noting that without a genuine commitment to reducing or ceasing CU, the provided tools might be ineffective. Some users attributed this to the platform's inherent characteristics, which they perceived as resembling an "online course" or a "control tool."

Professionals emphasized the need for the platform to better align with users' individual goals, noting that it may be less effective in influencing the stage of change, particularly in cases of ambivalence. To address this, they suggested implementing an initial survey or questionnaire to assess the user's profile, allowing for more tailored interventions.

XSL•FO
RenderX

Problem Awareness

Users acknowledged the platform's effectiveness in increasing awareness of cannabis consumption, helping them identify key moments in the change process, track the frequency and quantity of their use, and recognize its impact on daily life. They perceived the activities as tools that enabled them to "name their feelings" and, by "seeing themselves represented," become "more aware of their problem." By articulating their habits, users felt empowered to "reconsider" them and recognize them as problematic. Character testimonies and usage tracking features were seen as facilitators, with the module on managing craving symptoms highlighted as particularly helpful.

Economic Awareness

Users noted the significant motivational role of visualizing the economic impact of their cannabis consumption. They observed that the platform lacks this functionality and suggested incorporating tools to calculate cannabis-related expenditures.

Guidance and Use

Virtual Treatment

Users and professionals identified several benefits of virtual treatment (*in favor*), including greater intimacy and accessibility, particularly for new users or those with moderate cannabis consumption. They noted that the platform enables anonymity, offering a more "private" approach to addressing cannabis abuse. Some professionals viewed virtual treatment as a better first approach, while others emphasized that the platform's modules appeared designed to address both ends of the consumption spectrum—targeting individuals with SUDs as well as those engaging in recreational use—and could serve as a support tool during the maintenance stage.

Some users and professionals raised concerns (*against*) that virtual treatment lacks the warmth and personalization inherent to in-person interactions, which may be essential for users with severe CU. Professionals argued that DHIs "cannot encompass the full scope of work conducted in face-to-face therapy" and warned that users might "become frustrated if they fail to reduce their use," potentially deterring them from seeking in-person help. Specifically, they noted that the platform may not be effective for patients with higher dependency consumption patterns or dual pathology.

Virtual Support From Laura

Some users mentioned that women are often associated with caregiving roles and suggested that having a female therapist (*therapist gender*) as a guide in the app could be more effective "from a marketing point of view."

Many users valued the presence (*in favor*) of a humanized name and persona on the platform, as it enhanced engagement and provided a sense of individualized support. They emphasized the importance of "being able to talk to a person or feeling that someone is answering" and expressed a desire for additional features, such as a forum or chat for real-time dialogue, direct contact with someone, or having "Laura personally check in on their progress" during treatment.

Conversely, some users felt that they "do not need anyone to check on their progress." Additionally, some participants found the online format, including Laura's support, to be overly impersonal, describing text messages (SMS and email) and online assistance as inherently "cold" (*against*).

Complement to In-Person Therapy

Users considered the platform a "starting point for seeking help" and a valuable complement to in-person therapy. They noted that access to cannabis-related information and intake logs could optimize in-person sessions by allowing more time to address "individual and personalized" concerns. To enhance continuity and efficacy in treatment, they emphasized the importance of professionals having access to the information recorded on the platform.

Professionals viewed the platform as a complement rather than a substitute for in-person therapy, recommending its use as an educational and monitoring tool. They highlighted its potential utility for users on therapy waiting lists and its applicability in primary care and various disciplines, such as social work and nursing. The data provided by the platform were seen as beneficial, either by equipping patients with more information to enhance awareness or by serving as a support tool through patient-registered data. However, they noted that the platform currently lacks information on local resources for additional support, which could enhance its effectiveness as a complementary tool.

Content and Design

App

Users emphasized the difficulty in visualizing the app on mobile devices. They expressed a preference for an app with personalized notifications and tracking features to improve interaction and usability. They mentioned the possibility of receiving "feedback" as a desirable feature.

Professionals suggested that adapting the platform into a mobile app could significantly improve adherence and facilitate regular access.

Notifications

Users had varying opinions on the utility of notifications. Some felt that notifications, being common in apps, could serve as reminders to encourage use and improve adherence. Others, however, preferred them to be optional, with one user describing them as feeling "like homework for children." There was no consensus on the preferred frequency of notifications.

Professionals suggested allowing users to customize the frequency and type of notifications to enhance their usefulness and prevent boredom or annoyance. They also noted that notifications could serve as a means to provide valuable information on accessing alternative forms of treatment, further integrating the virtual platform with in-person care.

Information

Users appreciated the information provided but noted that it was occasionally too "dense," leading to boredom. They suggested incorporating more interactive and visual formats, such as videos instead of text, and simplifying language by

XSL•FO
RenderX

reducing technical jargon. While some users found characters illustrating CU experiences helpful, others felt they did not accurately reflect their own situations. Recommendations included adding more information on the negative consequences of CU, stress and anxiety management, improving the visualization of consumption graphs, introducing a search function for quick access to specific information, and incorporating a "panic button" to remind users of their personal goals during moments of craving.

Professionals found the information to be well-written and accessible but agreed with users that it could benefit from being more dynamic and visual. They also emphasized the importance of facilitating navigation to ensure easy access to relevant content, suggesting features such as a search box for specific terms and topics.

Small Achievements and Gamification

Users indicated that incorporating gamification elements could enhance the platform's appeal and motivational impact. They found the forms somewhat boring, with one user noting, "it feels like the platform is an educational course." They also suggested the inclusion of encouraging messages that acknowledge their progress or improvements in their health status.

Professionals suggested implementing progressive goals and rewards to foster user engagement and satisfaction. However, there was no consensus on whether modules should be unlocked sequentially as users progress.

Personalized Content

Users expressed a strong preference for more personalized content tailored to their individual situations and needs, believing this would enhance the platform's relevance and effectiveness. They highlighted the usefulness of sections for recording personal "excuses" and setting individualized goals.

Professionals stressed the importance of customizing content to align with users' stages in the change process and personal goals. They also suggested simplifying content based on users' cognitive or attentional capacities to enhance accessibility and comprehension.

Overview of Research Findings to Define Next Steps

To incorporate users' suggestions into the app, specific recommendations were collected based on each code and its subcodes (Textbox 1). Moving forward, these insights will be integrated into the development process to ensure the resulting DHI remains user-centered.

Textbox 1. Recommendations for enhancing CANreduce 2.0 based on focus group feedback from professionals and users for each subcode.

1. Motivation and awareness
   - User age: personalize according to life stages.
   - Problem awareness: adjust to the level of self-consciousness, including specific information for the precontemplation stage. It could be set with an initial user profile survey.
   - Motivation: include reward systems and more information about the negative consequences of cannabis use.
   - Economic awareness: add functionalities to visualize the economic and social costs of consumption. It could be used as part of the gamification function.
2. Guidance and use
   - Virtual treatment: include a forum or chat to facilitate real-time dialogue.
   - Virtual support from Laura: personalize interaction.
   - Complement to face-to-face therapy: implement its use with patients on waiting lists for therapy and in primary care, particularly with disciplines such as social work and nursing. Allow and promote professionals access to the information registered by their patients.
3. Content and design
   - App: improve usability in smartphones, introduce interactive content (improve form design, introduce a search box).
   - Notifications: implement notifications as reminders and motivational messages, allowing users to choose to receive them or not, the type of notifications, and the frequency.
   - Information: replace part of the text-based content with videos or other visual formats. Include a panic button to be used during cravings, to give quick access to personal reminders of self-goals, pros/cons, or benefits expected.
   - Small achievements and gamification: add motivational messages, incorporate progressive goals, and show the health benefits of the goals achieved.
   - Personalized content: tailor the content to meet users' needs, goals, and capabilities.

To further guide the enhancement of usability and adherence, a relational schema of key themes was developed to illustrate the interconnections and dynamics among the identified codes (Figure 2). This schema visualizes how codes interact and influence each other, highlights emerging patterns, and supports the interpretation of qualitative data. Additionally, it serves as a foundation for generating new hypotheses for future research.

XSL•FO
RenderX

Figure 2. Key themes for increasing usability and adherence in CANreduce 2.0.

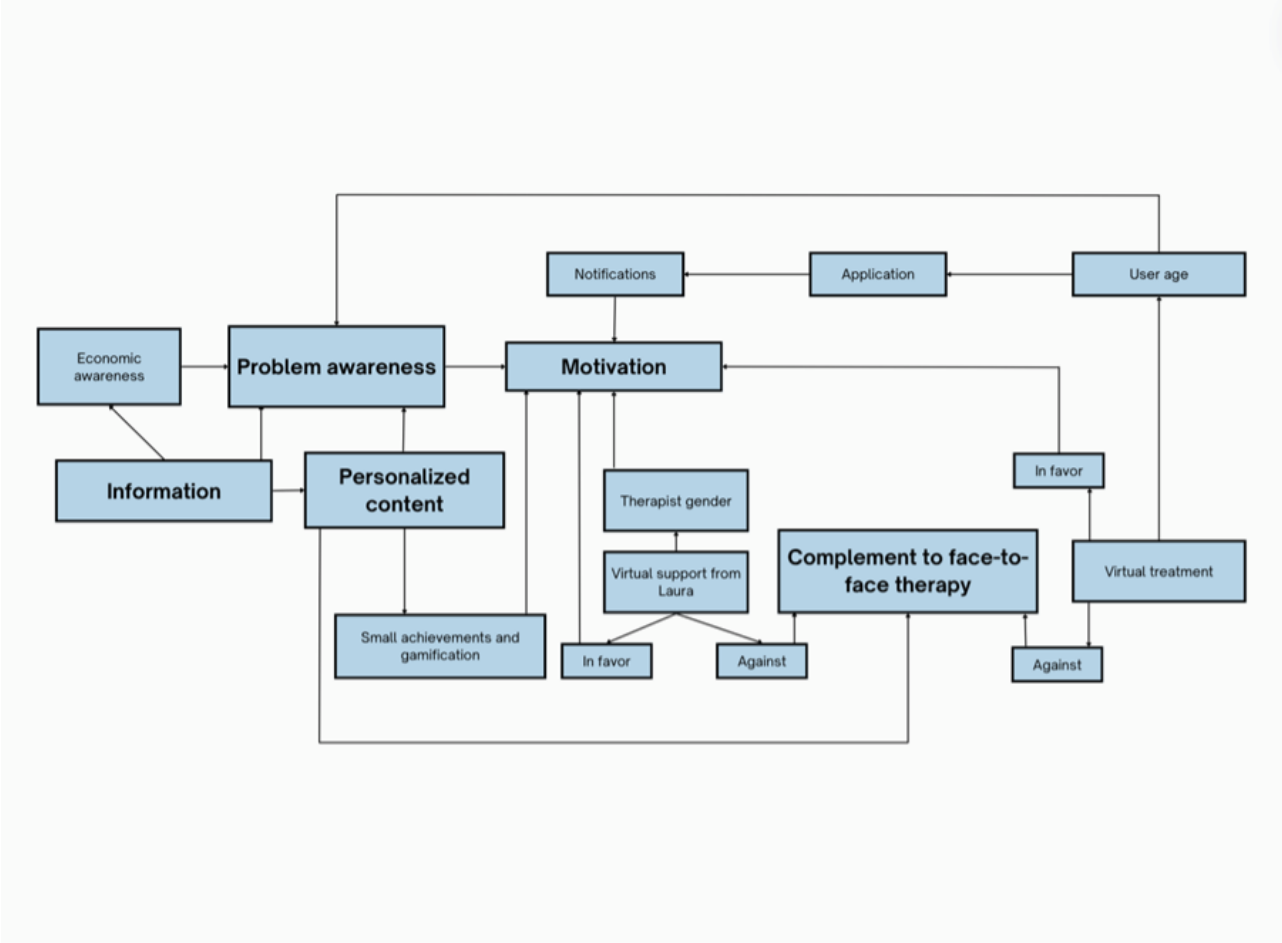

The relational schema of codes was developed through a rigorous qualitative analysis of the FG transcripts. The process began with a detailed transcription of all sessions, followed by the coding of meaning units. The resulting codes were then grouped into broad themes representing related concepts. Subsequently, relationships among these codes were analyzed using concept mapping techniques. The final schema was validated by subject matter experts.

Figure 2 visually represents the interrelations among the key themes identified in the analysis—motivation and awareness, guidance and use, and content and design—highlighting core elements that enhance user experience and adherence. Notably, utilizing the "information" provided within the platform and integrating "personalized content" can strengthen "problem awareness," which, in turn, increases user "motivation." Additionally, personalized content is perceived as valuable for using the platform as a "complement to face-to-face therapy," aiming to promote sustained engagement and adherence to the CANreduce 2.0 program and overall treatment for CUD.

Discussion

Principal Findings

This study evaluates the perceptions of users and professionals regarding the utility and potential effectiveness of CANreduce 2.0 in reducing CU, as well as their suggestions for improving its usability and, consequently, adherence to online treatment.

Regarding the main themes evaluated in the FGs, several conclusions can be drawn from the results. First, both users and professionals agreed that CANreduce 2.0 could serve as a valuable complement to in-person therapy. Second, most participants emphasized that the information provided in the modules significantly increased problem awareness and, consequently, sustained motivation to quit or reduce CU. Finally, to enhance adherence to virtual treatment, both users and professionals suggested incorporating notifications, gamification elements, and personalized features to adapt the app to individual users.

Motivation and Awareness

The findings reveal significant discrepancies in perceptions regarding the optimal age for using the platform. Some users viewed the digital format as more suitable for younger individuals, whereas others believed middle-aged adults, who may have greater insight into their consumption habits, could benefit more. While DHIs are generally assumed to be preferred by younger populations, research remains inconclusive [46]. Studies on web-based interventions have shown that recruited participants tend to be older than those typically treated in specialized addiction centers [47], and dropout rates are higher among younger individuals [48]. Given these uncertainties, it is recommended that the intervention be adapted to allow for personalized modules and support features tailored to specific life stages, maturity levels, and self-awareness, ensuring that each age group finds the tool relevant and effective.

XSL•FO
RenderX

Regarding motivation and problem awareness, it is noteworthy that during recruitment for the CANreduce 2.0 study, the Google search advertisement appeared more than 250 times per day, with an average of 19.6 individuals clicking on the app link. Given the rising demand for CU treatment [2,7] and the persistent lack of access for many individuals [15], a freely available tool discoverable through a simple internet search could serve as a highly valuable resource for those seeking help with their consumption but unable or unwilling to engage in traditional interventions. Digital tools offer several advantages, including anonymity, continuous availability, elimination of physical travel, and rapid accessibility [15,27], which were highlighted in the FGs as key benefits of digital health treatments, particularly CANreduce 2.0.

As for motivation, both users and professionals identified it as a crucial factor in achieving changes in consumption patterns. When asked about the potential effectiveness of the intervention in reducing CU, participants closely linked the tool's utility to the user's intrinsic motivation and problem awareness at the time of engagement. Interviewed professionals suggested that CANreduce 2.0 is suitable for individuals in both active consumption and maintenance phases of abstinence, though it may be less effective for those who remain ambivalent. Consequently, they proposed incorporating information specifically tailored for individuals in the precontemplation stage. According to the Stages of Change Model, individuals in this initial stage are not yet considering modifying their consumption habits in the foreseeable future, as they lack full awareness of the associated problems [36]. The first step in addressing an SUD is for the individual to recognize their own problematic behavior and consider making a change. Users acknowledged that the tool effectively assesses their CU, raises awareness of its health consequences, and records instances of excessive consumption, which could positively impact self-efficacy [49]. However, the results also highlighted the need for additional strategies to sustain meaningful and continuous motivation. Proposed features for integration included a reward or incentive system to reinforce progress, expanded information on the negative consequences of CU and withdrawal syndrome, and interactive tools, such as simulations or expense calculators, to help users visualize the economic and social costs of their consumption.

Guidance and Use

There is substantial evidence demonstrating the impact of mental health apps on the general population, including individuals without symptoms. These apps have shown potential as cost-effective, accessible, and low-intensity intervention tools for those unable or unwilling to receive standard treatment [26]. However, adherence to treatment remains a significant challenge in digital mental health and addiction interventions [50]. Evidence suggests that one strategy to improve adherence is the provision of user-centered guidance.

The randomized clinical trial of CANreduce 2.0 conducted in the Swiss population demonstrated the app's efficacy in reducing the frequency of use, severity of dependence, and anxiety symptoms. However, it did not find significant differences between support provided by a personalized figure and that offered by a nonindividualized support team [30].

To thoroughly examine these findings, study participants were asked about their experiences and perceptions regarding the virtual support provided by Laura, the therapist integrated into the treatment process through introductory videos for each module and direct messaging. Opinions were diverse. Supporters of virtual assistance emphasized the importance of "humanizing" the experience, suggesting that a forum or a space for more personalized guidance could be beneficial. They noted that personalized interaction, even in a virtual format, fostered a sense of connection and support. Conversely, some participants expressed disinterest in virtual interaction, preferring to forgo this type of communication. Professionals echoed this concern, criticizing virtual support as impersonal and advocating for in-person accompaniment, which they believed offered a more effective and empathetic form of support.

Despite these perceptions, "Dr. Google" and "ChatGPT" remain among the most frequently used tools for medical and treatment inquiries, with people becoming increasingly familiar with their use [51-53]. Mental health apps are often accessed through internet searches and social media [54], which presents certain risks due to the generally poor quality of online information [55] and the subjective, often biased nature of its evaluation [56,57].

Among the interviewed user groups, there was strong support for using the platform as a virtual treatment. They highlighted its potential as an ecological intervention and recognized its advantages in providing an accessible and flexible treatment option. However, professionals expressed reservations about its effectiveness as a standalone tool, particularly for individuals with severe dependence or dual disorders. Although evidence suggests that individuals with more severe symptoms are more likely to complete such treatments [58], professionals questioned whether the app was adequate for these cases. Nonetheless, both users and professionals acknowledged the platform's value as an adjunct to in-person therapy. With its educational and monitoring features, the information provided could serve as an effective support tool for professionals. This perspective aligns with existing evidence showing that DHIs can enhance treatment effectiveness and support abstinence [59,60]. Integrating the platform with traditional therapy could maximize its strengths in providing ongoing support and monitoring while ensuring comprehensive care for individuals with more complex needs.

Content and Design

The CANreduce 2.0 app implements a program based on motivational interviewing and CBT, designed to increase disorder awareness, confront users with their consumption problem, and encourage commitment to modifying their CU patterns. This combination of methods, along with its modular structure, aligns with evidence suggesting that apps with such designs demonstrate greater effectiveness [30].

The proposal to register CU habits was well received by FG participants, including both users and professionals. This feature was recognized for its utility in patient monitoring and its

XSL•FO
RenderX

potential to promote positive behaviors while reducing negative ones [61]. Additionally, participants suggested incorporating gamification elements and small achievements to enhance motivation and engagement, as these features could provide further incentives for users to stay committed to their treatment and actively engage with the platform's resources.

A limitation of this DHI, as highlighted during FG discussions, is its lack of adaptation to each user's unique case. The customization of digital health apps, particularly in terms of usability, remains an underexplored area compared with commercially driven digital apps [39]. User-centered design extends beyond aesthetics and interface dynamics to encompass adaptation to the demographic characteristics, mental models, and psychological needs of the target population [9]. The flow of app use—spanning user experience and interface design—is a fundamental aspect of digital apps. FG research serves as a key method for evaluating these elements [40], and this study provided insights into the perceptions of both users and professionals. Further exploration of individual user expectations and goals could enhance personalization and improve the care process [62].

Users expressed a preference for a mobile app featuring personalized notifications and interactive content, aligning with research suggesting that such features could enhance engagement and adherence [63]. The importance of tailoring content to individual needs and characteristics was a recurrent theme in discussions.

Regarding the user interface, participants highlighted challenges in accessing the app from mobile devices, as it was initially designed as a web-based platform before being adapted for smartphones. They found the predominantly text-based content burdensome and suggested incorporating videos or other visual formats to improve engagement [15]. Additionally, enhancing the visualization of data graphs was identified as a necessary improvement.

Next Steps

To improve adherence and usability in CANreduce 2.0, feedback from users and professionals (Textbox 1) will be integrated into the DHI. Specifically, motivation and awareness will be enhanced by incorporating a reward system and a visualization tool that highlights the individual economic costs of consumption (expense calculator). Additional content will be included to provide deeper insights into the negative consequences of CU and withdrawal syndrome, along with tailored elements for individuals in the precontemplation stage, as identified through an initial user profiling survey. Motivation and readiness to change are critical mechanisms in mental health therapeutic adherence [25], particularly in the precontemplation stage, where they are most affected. Moreover, motivation is a dynamic process that fluctuates across the different stages of change and varies in intensity over time [25,36]. Thus, integrating specific modules and interventions that address these fluctuations is crucial to maintaining adherence and preventing potential dropouts.

For enhanced guidance and usability, CANreduce 2.0 will incorporate a voluntary forum feature to facilitate peer support and interaction. Additionally, its potential application in addiction treatment centers will be further explored. Consistent with existing evidence [59,60], both users and professionals have recognized CANreduce 2.0 as a valuable complement to face-to-face therapy, particularly for highly motivated individuals. Consequently, future DHI development will prioritize its integration as a therapeutic adjunct, especially for patients on waiting lists and in primary care. Additionally, the feasibility of enabling professionals to access and track patient data will be explored to further enhance clinical support.

Design and content improvements will prioritize usability and accessibility. CANreduce 2.0 will be redesigned as a web application to enhance functionality on smartphones and improve interface interactivity. Planned optimizations include gamification elements, such as the previously mentioned reward system, individual progressive goal tracking, a search tool, optional notifications, and increased visual content to replace text-heavy sections.

Following the implementation of these features, additional FGs will be conducted to gather updated user and professional perspectives, ensuring the continuation of the iterative co-design process. These sessions will also explore areas with limited consensus, such as the optimal user age and necessary adaptations to tailor the intervention for different life stages and levels of self-awareness.

Poor adherence remains a major barrier to implementing DHIs [17,24] and CUD treatments, particularly among younger populations [15]. Motivation, a key construct in the Adherence-Focused model, has been linked to increased treatment completion [25]. As highlighted by FG participants, interventions such as CANreduce 2.0 may heavily depend on user motivation while also supporting individuals in identifying their reasons for change and facilitating planning [36]. For future studies on DHIs, motivation should be carefully considered, as both users and professionals emphasized—consistent with existing literature—that it significantly impacts adherence and, consequently, the intervention's potential success or failure.

Furthermore, the differences observed between groups in Figure 1 emphasize the importance of addressing varying priorities and perspectives when designing therapeutic tools or interventions. Misalignment in certain areas highlights the need to strengthen communication and cocreation processes. Bridging this gap could help ensure that all stakeholders recognize the value of key features and their potential to enhance therapeutic outcomes. Incorporating diverse perspectives throughout the design and development process is crucial for creating interventions that are both effective and widely adopted. A participatory approach—engaging both professionals and users—could align priorities, foster shared ownership of solutions, and enhance long-term impact and sustainability.

Limitations

This study has limitations, including the small group sizes and the fact that some surveyed users were not highly familiar with the app. However, evidence suggests that such unfamiliarity can provide valuable insights into user experience evaluation [64]. While the FG strategy employed in the protocol is effective

XSL•FO
RenderX

for exploring motivations, experiences, and subjective perceptions in-depth, it is inherently limited in its level of evidence, as it lacks statistical representativeness and does not allow for the measurement of phenomena [38].

The use of a co-design model often presents challenges, including the significant time, effort, and financial resources required, as well as the complexity of balancing diverse perspectives, which can complicate decision-making [65-67]. Another common barrier is the generalizability of outputs; as with many co-design studies, our small group sizes may not fully represent the broader user population. Additionally, involving affiliated professionals in the process may influence project direction and introduce potential biases [67]. Despite these challenges, co-design offers substantial benefits, primarily by ensuring that the intervention aligns with end user needs. This approach allows us to directly integrate user perspectives into the design, enhancing relevance, acceptability, and usability [65,68]. Moreover, perceived value or buy-in is widely recognized as a key enabler of co-design, fostering sustained stakeholder engagement and ultimately supporting adoption [67]. The co-design process also generated innovative ideas for improving CANreduce 2.0—such as incorporating notifications and small achievements—making the intervention more applicable to real-life situations and increasing the likelihood of successful integration [66,68].

Conclusions

The findings of this study highlight the complexity of designing effective DHIs for CU. While technology offers unique opportunities to engage users, the results emphasize the need to address the diverse needs of different demographic groups and severity levels to enhance motivation and adherence. Personalization, combined with robust motivational strategies and an interactive design, appears to be critical for the success of such platforms. Furthermore, collaboration among technology developers, health care professionals, and end users is essential to creating digital solutions that are both practical and effective.

Advancing the design of DHIs that effectively address substance use in innovative and engaging ways remains a challenge in medical informatics and mental health. Integrating user perspectives and values into the design process is essential for enhancing the acceptability, usability, and utility of these tools. For CANreduce 2.0, the insights gained from this study will guide the implementation of necessary improvements, refining both the app's content and interactive features.

Acknowledgments

We express our gratitude to all the participants who took part in the focus groups. This study was funded by the Instituto de Salud Carlos III (ISCIII), the European Union NextGenerationEU, and the Mecanismo para la Recuperación y la Resiliencia (MRR; grants RD21/0009/0001, RD21/0009/0004, RD21/0009/0007, RD21/0009/0010, and RD21/0009/0025).

Data Availability

All data generated or analyzed during this study are included in this published article and its multimedia appendices.

Disclaimer

During the preparation of this work, the authors used ChatGPT (OpenAI, Inc.) to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Authors' Contributions

DFS contributed to the methodology, investigation, data curation, formal analysis, writing (original draft), and visualization. MPR was involved in data curation, formal analysis, writing (original draft), visualization, and writing (review & editing). SA helped with the data curation and formal analysis. MAS, PZ, FA, PGC, GGB, ELM, CM, and FF participated in the investigation, writing (review and editing), and resource provision. LM contributed to the methodology, investigation, writing (review and editing), project administration, funding acquisition, resource provision, and software development. JIMP played a significant role in the conceptualization, methodology, investigation, data curation, writing (original draft), project administration, funding acquisition, resource provision, software development, and validation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for professionals.
[DOCX File , 12 KB - jmir_v27i1e65025_app1.docx]

Multimedia Appendix 2
Interview guide for users.
[DOCX File , 12 KB - jmir_v27i1e65025_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Overview of codes from the qualitative analysis of the transcripts.
[DOCX File , 20 KB - jmir_v27i1e65025_app3.docx]

Multimedia Appendix 4
Qualitative results with codes from the interview's transcriptions.
[DOCX File , 28 KB - jmir_v27i1e65025_app4.docx]

References

1. World Drug Report 2024. United Nations Office on Drugs and Crime. 2024. URL: https://www.unodc.org/unodc/data-and-analysis/world-drug-report-2024.html [accessed 2024-07-07]
2. Informe 2023. Alcohol, tabaco y drogas ilegales en España. Observatorio Español de las Drogas y las Adicciones. 2023. URL: https://pnsd.sanidad.gob.es/eu/profesionales/sistemasInformacion/informesEstadisticas/pdf/2023OEDA-INFORME.pdf [accessed 2024-06-29]
3. Encuesta sobre uso de drogas en enseñanzas secundarias en España (ESTUDES), 1994-2023. Observatorio Español de las Drogas y las Adicciones. 2023. URL: https://pnsd.sanidad.gob.es/profesionales/sistemasInformacion/sistemaInformacion/pdf/ESTUDES_2023_Resumen_ejecutivo.pdf [accessed 2024-06-30]
4. Estrategia Nacional sobre Adicciones 2017-2024. Ministerio de Sanidad, Servicios Sociales e Igualdad. 2017. URL: https://pnsd.sanidad.gob.es/pnsd/estrategiaNacional/docs/180209_ESTRATEGIA_N.ADICCIONES_2017-2024__aprobada_CM.pdf [accessed 2024-06-12]
5. Guerrero-Martelo M, Galván G, Pinedo-López J, Vásquez-De la Hoz F, Torres-Hoyos F, Torres-Oviedo J. Lifetime prevalence of cannabis use and academic performance in adolescents. Salud Uninorte 2015 Jun 01;31(3):467-478. [doi: 10.14482/sun.31.3.8046]
6. Harper C, Dittus P, Ethier K. Associations between marijuana use and 15 indicators of sexual risk and sexual health service use. Journal of Adolescent Health 2016 Feb;58(2):S109-S110. [doi: 10.1016/j.jadohealth.2015.10.230]
7. Montanari L, Guarita B, Mounteney J, Zipfel N, Simon R. Cannabis use among people entering drug treatment in Europe: a growing phenomenon? Eur Addict Res 2017 May;23(3):113-121. [doi: 10.1159/000475810] [Medline: 28511182]
8. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, et al. Computer-assisted delivery of cognitive-behavioral therapy for addiction: a randomized trial of CBT4CBT. Am J Psychiatry 2008 Jul;165(7):881-888 [FREE Full text] [doi: 10.1176/appi.ajp.2008.07111835] [Medline: 18450927]
9. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
10. Gates P, Sabioni P, Copeland J, Le Foll B, Gowing L. Psychosocial interventions for cannabis use disorder. Cochrane Database Syst Rev 2016;5:1-127. [doi: 10.1002/14651858.CD005336.pub4]
11. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
12. Kiluk B, Nich C, Babuscio T, Carroll K. Quality versus quantity: acquisition of coping skills following computerized cognitive-behavioral therapy for substance use disorders. Addiction 2010 Dec;105(12):2120-2127 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03076.x] [Medline: 20854334]
13. Carr S, Klinger S, Manthey J, Isorna-Folgar M, Amatller-Gutiérrez O, Carballo-Crespo J, et al. Guía Clínica de Cannabis. Barcelona, Spain: Socidrogalcohol; 2022.
14. Muñoz RF, Bunge EL, Chen K, Schueller SM, Bravin JI, Shaughnessy EA, et al. Massive open online interventions. Clin Psychol Sci 2016;4(2):205. [doi: 10.1177/2167702615583840]
15. Beneria A, Santesteban-Echarri O, Daigre C, Tremain H, Ramos-Quiroga JA, McGorry PD, et al. Online interventions for cannabis use among adolescents and young adults: systematic review and meta-analysis. Early Interv Psychiatry 2022 Aug;16(8):821-844. [doi: 10.1111/eip.13226] [Medline: 34464502]
16. Murciano-Hueso A, Gutiérrez-Pérez BM, Martín-Lucas J, Huete García A. Juventud onlife. Estudio sobre el perfil de uso y comportamiento de los jóvenes a través de las pantallas. Relieve 2022 Dec 22;28(2):26158. [doi: 10.30827/relieve.v28i2.26158]
17. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756. [doi: 10.1111/acem.12692] [Medline: 25997375]
18. Boumparis N, Schaub M. Recent advances in digital health interventions for substance use disorders. Curr Opin Psychiatry 2022 Jul 01;35(4):246-251. [doi: 10.1097/YCO.0000000000000798] [Medline: 35781464]
19. Boumparis N, Loheide-Niesmann L, Blankers M, Ebert DD, Korf D, Schaub MP, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: a systematic review and meta-analysis. Drug and Alcohol Dependence 2019 Jul;200:82-94. [doi: 10.1016/j.drugalcdep.2019.03.016]

XSL•FO
RenderX

20. Hoch E, Preuss UW, Ferri M, Simon R. Digital interventions for problematic cannabis users in non-clinical settings: findings from a systematic review and meta-analysis. Eur Addict Res 2016 May;22(5):233-242. [doi: 10.1159/000445716] [Medline: 27160333]
21. Olmos A, Tirado-Muñoz J, Farré M, Torrens M. The efficacy of computerized interventions to reduce cannabis use: a systematic review and meta-analysis. Addict Behav 2018 Apr;79:52-60. [doi: 10.1016/j.addbeh.2017.11.045] [Medline: 29248863]
22. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 01;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
23. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
24. Amann M, Haug S, Wenger A, Baumgartner C, Ebert DD, Berger T, et al. The effects of social presence on adherence-focused guidance in problematic cannabis users: protocol for the CANreduce 2.0 randomized controlled trial. JMIR Res Protoc 2018 Jan 31;7(1):e30. [doi: 10.2196/resprot.9484] [Medline: 29386176]
25. Laranjeira C, Carvalho D, Valentim O, Moutinho L, Morgado T, Tomás C, et al. Therapeutic adherence of people with mental disorders: an evolutionary concept analysis. Int J Environ Res Public Health 2023 Feb 22;20(5):A [FREE Full text] [doi: 10.3390/ijerph20053869] [Medline: 36900879]
26. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
27. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
28. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
29. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
30. Baumgartner C, Schaub MP, Wenger A, Malischnig D, Augsburger M, Walter M, et al. CANreduce 2.0 adherence-focused guidance for internet self-help among cannabis users: three-arm randomized controlled trial. J Med Internet Res 2021 Apr 30;23(4):e27463 [FREE Full text] [doi: 10.2196/27463] [Medline: 33929333]
31. Mestre-Pintó JI, Fonseca F, Schaub MP, Baumgartner C, Alias-Ferri M, Torrens M. CANreduce-SP-adding psychological support to web-based adherence-focused guided self-help for cannabis users: study protocol for a three-arm randomized control trial. Trials 2022 Jun 22;23(1):524 [FREE Full text] [doi: 10.1186/s13063-022-06399-2] [Medline: 35733201]
32. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012 Sep;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
33. Wong HW, Lo B, Shi J, Hollenberg E, Abi-Jaoude A, Johnson A, et al. Postsecondary student engagement with a mental health app and online platform (Thought Spot): qualitative study of user experience. JMIR Ment Health 2021 Apr 02;8(4):e23447 [FREE Full text] [doi: 10.2196/23447] [Medline: 33797395]
34. Majid S, Reeves S, Figueredo G, Brown S, Lang A, Moore M, et al. The extent of user involvement in the design of self-tracking technology for bipolar disorder: literature review. JMIR Ment Health 2021 Dec 20;8(12):e27991 [FREE Full text] [doi: 10.2196/27991] [Medline: 34931992]
35. Meskó B, deBronkart D. Patient design: the importance of including patients in designing health care. J Med Internet Res 2022 Aug 31;24(8):e39178 [FREE Full text] [doi: 10.2196/39178] [Medline: 36044250]
36. Enhancing Motivation for Change in Substance Use Disorder Treatment: Updated. Rockville, MD: Substance Abuse and Mental Health Services Administration (SAMHSA); 2019.
37. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
38. García CMM, Mateo RI. El grupo focal como técnica de investigación cualitativa en salud: diseño y puesta en práctica. Aten Primaria 2000;25(3):186. [doi: 10.1016/S0212-6567(00)78485-X]
39. Broekhuis M, van Velsen L, Peute L, Halim M, Hermens H. Conceptualizing usability for the eHealth context: content analysis of usability problems of eHealth applications. JMIR Form Res 2021 Jul 27;5(7):e18198 [FREE Full text] [doi: 10.2196/18198] [Medline: 34313594]
40. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. Int J Technol Assess Health Care 2020 Jan;36(1):64-70. [doi: 10.1017/S0266462319003507] [Medline: 31973787]

XSL•FO
RenderX

41. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
42. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb 22;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
43. Krueger R, Casey M. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks, CA: SAGE Publications; 2015.
44. Bagot K, Hodgdon E, Sidhu N, Patrick K, Kelly M, Lu Y, et al. End user-informed mobile health intervention development for adolescent cannabis use disorder: qualitative study. JMIR Mhealth Uhealth 2019 Oct 04;7(10):e13691 [FREE Full text] [doi: 10.2196/13691] [Medline: 31588909]
45. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
46. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
47. Rooke S, Copeland J, Norberg M, Hine D, McCambridge J. Effectiveness of a self-guided web-based cannabis treatment program: randomized controlled trial. J Med Internet Res 2013 Feb 15;15(2):e26 [FREE Full text] [doi: 10.2196/jmir.2256] [Medline: 23470329]
48. Vederhus JK, Rørendal M, Skårdal M, Næss MO, Clausen T, Kristensen Ø. Successful outcomes with low-threshold intervention for cannabis use disorders in Norway - an observational study. PLoS One 2022 Jun;17(6):e0269988 [FREE Full text] [doi: 10.1371/journal.pone.0269988] [Medline: 35709178]
49. Kapitány-Fövény M, Vagdalt E, Ruttkay Z, Urbán R, Richman MJ, Demetrovics Z. Potential of an interactive drug prevention mobile phone app (Once Upon a High): questionnaire study among students. JMIR Serious Games 2018 Dec 04;6(4):e19 [FREE Full text] [doi: 10.2196/games.9944] [Medline: 30514697]
50. Musiat P, Johnson C, Atkinson M, Wilksch S, Wade T. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. Psychol Med 2022 Jan;52(2):229-240. [doi: 10.1017/S0033291721004621] [Medline: 34802474]
51. Duong D. Time to embrace Dr. Google? CMAJ 2021 Apr 19;193(16):E571-E572 [FREE Full text] [doi: 10.1503/cmaj.1095935] [Medline: 33875466]
52. Nazir A, Wang Z. A comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges. Meta Radiol 2023 Sep;1(2):100022 [FREE Full text] [doi: 10.1016/j.metrad.2023.100022] [Medline: 37901715]
53. Strzelecki A. Google medical update: why is the search engine decreasing visibility of health and medical information websites? Int J Environ Res Public Health 2020 Feb 12;17(4):1160 [FREE Full text] [doi: 10.3390/ijerph17041160] [Medline: 32059576]
54. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
55. Velasco RBM, Palomino PLM, Sánchez FBC, Sanchiz PA, Muñoz CRA. Se puede confiar en el doctor Google®? Fiabilidad de la información médica online sobre la enfermedad inflamatoria intestinal pediátrica. An Pediatr 2021;94(4):257-258. [doi: 10.1016/j.anpedi.2020.04.033]
56. Davis JK. Dr. Google and premature consent: patients who trust the internet more than they trust their provider. HEC Forum 2018 Sep;30(3):253-265. [doi: 10.1007/s10730-017-9338-z] [Medline: 29063996]
57. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
58. Sinadinovic K, Johansson M, Johansson AS, Lundqvist T, Lindner P, Hermansson U. Guided web-based treatment program for reducing cannabis use: a randomized controlled trial. Addict Sci Clin Pract 2020 Feb 18;15(1):9 [FREE Full text] [doi: 10.1186/s13722-020-00185-8] [Medline: 32070417]
59. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for drug and alcohol abuse and smoking addiction: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):178-186. [doi: 10.1016/j.cpr.2010.10.002] [Medline: 21095051]
60. Tofighi B, Abrantes A, Stein MD. The role of technology-based interventions for substance use disorders in primary care: a review of the literature. Med Clin North Am 2018 Jul;102(4):715-731 [FREE Full text] [doi: 10.1016/j.mcna.2018.02.011] [Medline: 29933825]
61. Scott CK, Dennis ML, Johnson KA, Grella CE. A randomized clinical trial of smartphone self-managed recovery support services. J Subst Abuse Treat 2020;117:108089. [doi: 10.1016/j.jsat.2020.108089]
62. Iveson C. Solution-focused brief therapy. Adv Psychiatr Treat 2002;8(2):156. [doi: 10.1192/apt.8.2.149]

XSL•FO
RenderX

63. Martinez Agulleiro L, Patil B, Firth J, Sawyer C, Amann BL, Fonseca F, et al. A systematic review of digital interventions for smoking cessation in patients with serious mental illness. Psychol Med 2023 Aug;53(11):4856-4868 [FREE Full text] [doi: 10.1017/S003329172300123X] [Medline: 37161690]
64. Toolaroud PB, Nabovati E, Mobayen M, Akbari H, Feizkhah A, Farrahi R, et al. Design and usability evaluation of a mobile-based-self-management application for caregivers of children with severe burns. Int Wound J 2023 Sep;20(7):2571-2581 [FREE Full text] [doi: 10.1111/iwj.14127] [Medline: 36872302]
65. Prior SJ, Campbell S. Patient and family involvement: a discussion of co-led redesign of healthcare services. J Particip Med 2018 Feb 01;10(1):e5 [FREE Full text] [doi: 10.2196/jopm.8957] [Medline: 33052119]
66. Robert G, Cornwell J, Locock L, Purushotham A, Sturmey G, Gager M. Patients and staff as codesigners of healthcare services. BMJ 2015 Feb 10;350:g7714 [FREE Full text] [doi: 10.1136/bmj.g7714] [Medline: 25670179]
67. Shi B, Li G, Wu S, Ge H, Zhang X, Chen S, et al. Assessing the effectiveness of ehealth interventions to manage multiple lifestyle risk behaviors among older adults: systematic review and meta-analysis. J Med Internet Res 2024 Jul 31;26:e58174 [FREE Full text] [doi: 10.2196/58174] [Medline: 39083787]
68. Bird M, McGillion M, Chambers EM, Dix J, Fajardo CJ, Gilmour M, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 01;7(1):1-12 [FREE Full text] [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]

Abbreviations

CBT: cognitive behavioral therapy
CU: cannabis use
CUD: cannabis use disorder
DHI: digital health intervention
FG: focus group
SUD: substance use disorder

Edited by A Mavragani; submitted 05.08.24; peer-reviewed by ID Yucel, Y Cui; comments to author 29.10.24; revised version received 17.12.24; accepted 12.01.25; published 19.03.25.

Please cite as:
Folch-Sanchez D, Pellicer-Roca M, Sestelo MA, Zuluaga P, Arias F, Guzmán Cortez P, Amechat S, Gil-Berrozpe G, Lopez Montes E, Mercadé C, Fonseca F, Miquel L, Mestre-Pintó JI
Evaluating Perceptions of the CANreduce 2.0 eHealth Intervention for Cannabis Use: Focus Group Study
J Med Internet Res 2025;27:e65025
URL: https://www.jmir.org/2025/1/e65025
doi:10.2196/65025
PMID:

©Daniel Folch-Sanchez, Maria Pellicer-Roca, María Agustina Sestelo, Paola Zuluaga, Francisco Arias, Pablo Guzmán Cortez, Salma Amechat, Gustavo Gil-Berrozpe, Estefania Lopez Montes, Clara Mercadé, Francina Fonseca, Laia Miquel, Joan I Mestre-Pintó. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Embedding a Choice Experiment in an Online Decision Aid or Tool: Scoping Review

Nyantara Wickramasekera[1], MSc; Phil Shackley[1], PhD; Donna Rowen[1], PhD

Sheffield Centre for Health and Related Research (SCHARR), The University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Nyantara Wickramasekera, MSc
Sheffield Centre for Health and Related Research (SCHARR)
The University of Sheffield
30 Regent St
Sheffield, S14DA
United Kingdom
Phone: 44 01142224348
Email: N.Wickramasekera@sheffield.ac.uk

Abstract

Background: Decision aids empower patients to understand how treatment options match their preferences. Choice experiments, a method to clarify values used within decision aids, present patients with hypothetical scenarios to reveal their preferences for treatment characteristics. Given the rise in research embedding choice experiments in decision tools and the emergence of novel developments in embedding methodology, a scoping review is warranted.

Objective: This scoping review examines how choice experiments are embedded into decision tools and how these tools are evaluated, to identify best practices.

Methods: This scoping review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Searches were conducted on MEDLINE, PsycInfo, and Web of Science. The methodology, development and evaluation details of decision aids were extracted and summarized using narrative synthesis.

Results: Overall, 33 papers reporting 22 tools were included in the scoping review. These tools were developed for various health conditions, including musculoskeletal (7/22, 32%), oncological (8/22, 36%), and chronic conditions (7/22, 32%). Most decision tools (17/22, 77%) were developed in the United States, with the remaining tools originating in the Netherlands, United Kingdom, Canada, and Australia. The number of publications increased, with 73% (16/22) published since 2015, peaking at 4 publications in 2019. The primary purpose of these tools (20/22, 91%) was to help patients compare or choose treatments. Adaptive conjoint analysis was the most frequently used design type (10/22, 45%), followed by conjoint analysis and discrete choice experiments (DCEs; both 4/22, 18%), modified adaptive conjoint analysis (3/22, 14%), and adaptive best-worst conjoint analysis (1/22, 5%). The number of tasks varied depending on the design (6-12 for DCEs and adaptive conjoint vs 16-20 for conjoint analysis designs). Sawtooth software was commonly used (14/22, 64%) to embed choice tasks. Four proof-of-concept embedding methods were identified: scenario analysis, known preference phenotypes, Bayesian collaborative filtering, and penalized multinomial logit model. After completing the choice tasks patients received tailored information, 73% (16/22) of tools provided attribute importance scores, and 23% (5/22) presented a “best match” treatment ranking. To convey probabilistic attributes, most tools (13/22, 59%) used a combination of approaches, including percentages, natural frequencies, icon arrays, narratives, and videos. The tools were evaluated across diverse study designs (randomized controlled trials, mixed methods, and cohort studies), with sample sizes ranging from 23 to 743 participants. Over 40 different outcomes were included in the evaluations, with the decisional conflict scale being the most frequently used in 6 tools.

Conclusions: This scoping review provides an overview of how choice experiments are embedded into decision tools. It highlights the lack of established best practices for embedding methods, with only 4 proof-of-concept methods identified. Furthermore, the review reveals a lack of consensus on outcome measures, emphasizing the need for standardized outcome selection for future evaluations.

(J Med Internet Res 2025;27:e59209) doi:10.2196/59209

XSL•FO
RenderX

KEYWORDS

decision aid; decision tool; discrete choice experiment; conjoint analysis; value clarification; scoping review; choice experiment; database; study; article; data charting; narrative synthesis

Introduction

Understanding patient values for treatments is particularly important when a demonstrably superior treatment option is not available [1-4]. In these "preference-sensitive decisions" patients need to understand the treatment choices that are available to them, consider their personal values and weigh the trade-offs between treatment benefits and risks to select the optimal treatment that suits them [1-4]. Patient decision aids are used as a supporting tool when patients are faced with a preference-sensitive decision [5,6]. A well-established evidence base of 209 studies shows the effectiveness of decision aids in improving knowledge, reducing decisional conflict, increasing participation in decision-making, and receiving treatment with characteristics that they value [5,7-9]. Notably, digital decision aids also offer distinct advantages over other formats. Digital tools can incorporate interactive elements, allowing patients to personalize their experience by selecting the most relevant information. Moreover, these tools can include algorithms and perform real-time calculations to provide personalized results [7].

An important part of a decision aid, according to the International Patient Decision Aids Standards (IPDAS) collaboration, is an exercise that helps people clarify their values [10]. A value clarification exercise in a decision aid helps patients to identify the relative importance of treatment characteristics (attributes) that are congruent with their values [10] and helps patients to understand how the different treatment options align with their values [7,11].

Increasingly, choice experiments are used as a value clarification method within decision aids [7,11]. The term choice experiment (also called stated preference survey) will be used hereafter as an umbrella term for discrete choice experiments (DCEs) and conjoint analyses. In choice experiments, patients choose their preferred treatment from two or more hypothetical options, each defined by a unique combination of attributes and levels. Subsequent regression analyses of their choices reveal their preferences for the relative importance of treatment characteristics and allow the prediction of treatments that patients would prefer [12].

The field of embedding choice experiments in decision tools has seen an increase in new research since Weernink et al [7] published their review in 2018. Also, new conceptual models on how to integrate choice experiments in decision tools have been published since 2020 [13,14]. To identify best practices and knowledge gaps considering these advancements, a scoping review is warranted. A scoping review is appropriate as it aims to produce a comprehensive map of the research landscape by analysing the volume of research, the methodologies used, and the overall characteristics of the primary studies [15-20]. Unlike systematic reviews, which mainly focus on establishing the effectiveness of decision aids (already addressed by [5]), scoping reviews map the range of approaches used within a field. This broader approach allows the identification of best practices for integrating choice experiments into online decision aids.

The aim of this study is to undertake a scoping review of choice experiments embedded in decision tools to understand the current landscape of best practices. The objectives of this review are to (1) identify key methods used to embed a choice experiment into a decision tool, (2) characterize the design features of choice experiments embedded in decision aids, (3) identify the different study designs and outcomes used to evaluate the tool, and (4) describe how complex information was presented to the participants.

Methods

Overview

This scoping review was conducted following best practices in line with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [21]. A protocol detailing the plan of the methods section was prepared and the PRISMA-ScR checklist was used in this review (Multimedia Appendix 1).

Eligibility Criteria

Studies were selected based on the inclusion and exclusion criteria detailed in Textbox 1.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- A study describes the methods, development, design, presentation and visualization of patient information of a patient decision tool incorporating a choice experiment.
- A study evaluates a patient decision tool embedding a choice experiment or protocols describing the methods of the evaluation and what outcomes were measured to determine the effectiveness of the intervention.

It should also be noted that

- All study designs such as randomized controlled trials, observational studies, cross-sectional studies, and methodological papers, were included. Conference abstracts and existing systematic reviews were excluded.
- All articles meeting the above eligibility criteria were included regardless of the methodological quality as the aim of the scoping review is to gain an overview of the literature on embedding choice experiments in decision tools.
- No restrictions were imposed on the population, for methodological studies population is not relevant and for evaluation studies, the population can include all users that benefit from the decision tool such as patients, carers, or family members.

Exclusion criteria

- Decision aids without a choice experiment task as a value clarification method.
- Decision aids including other preference elicitation methods such as multicriteria decision analysis or time trade-off.
- Non–English-language publications.

Databases

MEDLINE via Ovid, PsycInfo via Ovid, and Web of Science databases were searched in March 2023. These 3 databases provide sufficient breadth of coverage of health-related sources to identify potentially relevant articles. Reference lists of included studies were searched to identify potentially relevant searches. Internet search engines (ie, Google) were used to identify any gray literature such as unpublished guidance documentation (ie, from regulators).

Search Strategy

The search strategy was conducted with the help of an information specialist using an iterative discussion process. The key concepts used in the search strategy are “choice experiment,” “decision aid,” “shared decision making,” “value clarification,” “predicted probability,” and synonyms of these key terms. The search excluded other methods such as “multi-criteria decision,” “analytical hierarchy process,” “time trade-off,” or “standard gamble” since the focus of this review is on choice experiments. The search strategy removed any studies not published in English. A time restriction was also introduced because an existing systematic review conducted a search up until 2016 [7]. Therefore, the current search used a modified version of the Weernink et al [7] search strategy and updated it to find new articles published since 2016. The modified version of the search strategy contained new terms such as “decision tool” and “predicted probability.” The full search strategy is included in Multimedia Appendix 2. The gray literature search used the same keywords mentioned above with Boolean operators. The first 20 potentially relevant records were retrieved for screening.

Screening

All search results were exported to EndNote and any duplicates were removed. All references were screened by one reviewer (NW). However, if the inclusion of a study was unclear, then discussions were held with the wider team to make a final decision (DR and PS). Initially, the references were screened based on title and abstract, followed by a full-text screening of relevant articles based on the inclusion and exclusion criteria. Details about the screening process, including reasons for exclusion were documented in a flow diagram using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) template.

Data Charting

Data charting tables were created in Excel to extract data from included studies. The charting tables were developed iteratively and modified as needed after extracting data from two articles. Data charting was conducted independently by one reviewer (NW). The data items included in the Excel charts were: author, date of publication, name of the tool, country, study aim, disease context, sample size, choice experiment embedding methodology or formula used, number of attributes, number of levels, type of attributes, graphs or visual pictographs used, choice experiment design used, type of choice tasks, number of choice tasks, software program used, description of the patient decision tool (ie, receive feedback on attribute importance or receive a “best match” treatment option that aligns with patients desired attributes, received a report), details about the evaluation, outcomes measured, and mean duration of the task. Since multiple study designs were included, not all the fields were relevant for all studies. For example, for studies that report on methodological aspects, the data charts mainly contained information about the methodology or formula, and the rest of the fields were not applicable.

Collating, Summarizing, and Reporting the Results

Since the aim of the scoping review is to map the literature, the data summarizing used a narrative synthesis of qualitative data with descriptive statistics of quantitative data. Illustrations were used to visualize the results using tables and figures. Where appropriate existing frameworks were used to categorize lists

XSL•FO
RenderX

of qualitative data extracted. The Core Outcome Measures in Effectiveness Trials (COMET) taxonomy [22] was used to systematically classify attributes into core areas and domains. This taxonomy is widely used to classify outcomes of trials [23]. This taxonomy's breadth and granularity, provided by its 5 core areas and 38 outcome domains, made it suitable for classifying the diverse attributes included in this scoping review. Risk of bias assessments were not conducted since the aim of the study was not to test the rigor of the articles, but rather to get an overview of the current literature on embedding choice experiments in decision tools.

Results

Overview

A PRISMA flow diagram detailing the inclusion and exclusion process is presented in Figure 1. A total of 1127 citations were identified. After duplicates were removed, 852 references were screened based on title and abstract. Of the 52 full-text articles that were reviewed, 19 studies that did not meet the inclusion criteria were excluded (Multimedia Appendix 3). Overall, 22 tools were included in the scoping review. The methodology, development and evaluation details of tools were extracted from 33 papers.

Figure 1. Flow diagram of study selection.

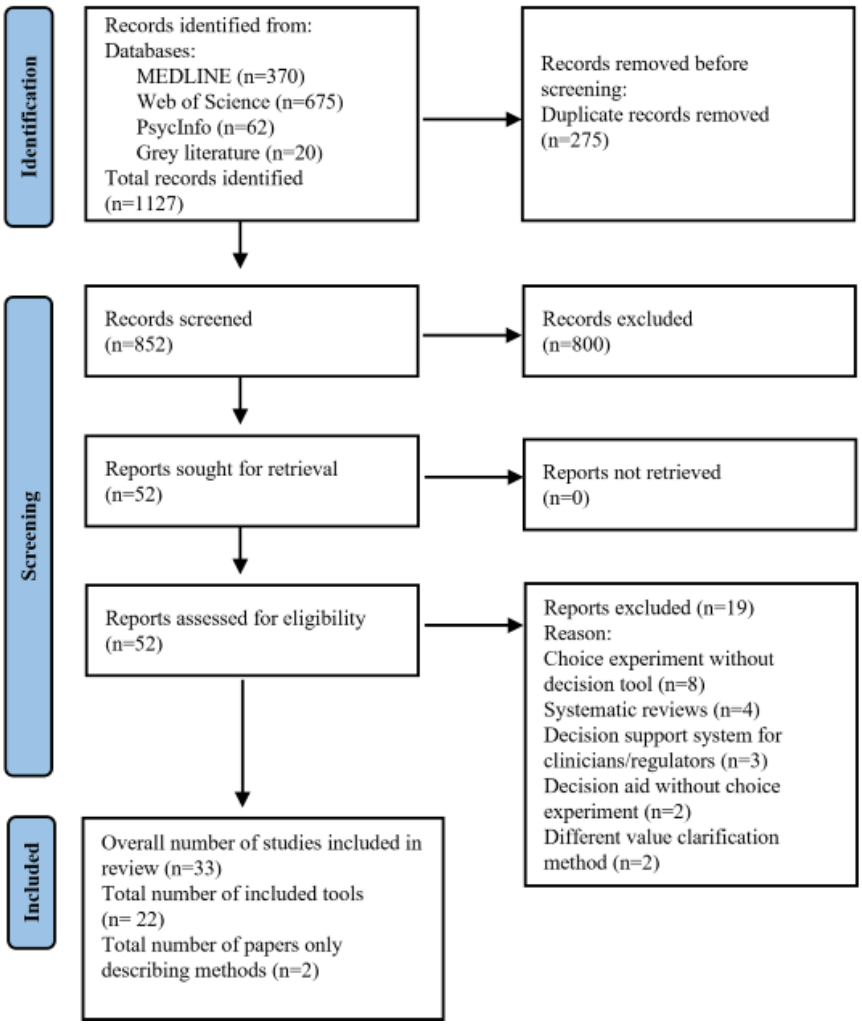

Study Characteristics

Table 1 provides a general overview of the types of decision tools that were found in the scoping review. The first decision tool was published in 2007. However, the number of publications has increased since then, with 16 out of the 22 (73%) studies published since 2015 and a peak of 4 publications in 2019. The majority of the decision tools (17/22, 77%) were developed in the United States. In total, 2 out of 22 (9%) tools were conducted in the Netherlands and one each in the United Kingdom, Canada, and Australia. A variety of different terms such as decision-making tool, decision aid, patient preference elicitation instrument, and discussion prioritization tool were used by the investigators to describe the tool they developed. There is no consensus on a clearly defined name for these tools among the included studies.

XSL•FO
RenderX

Table 1. General study characteristics[a].

Study	Date	Name of the tool	Country	Purpose of tool	Disease context
Abraham et al [24]	2015	Adaptive conjoint analysis survey	United States	Choice of medical treatments - complex antithrombotic therapy (anticoagulants or antiplatelet drugs)	Cardiovascular disease
Almario et al [25]	2018	Online Patient Decision Aid: IBD and me	United States	Choice between biological treatments for ulcerative colitis	Ulcerative colitis and Crohn's disease
Chhatre et al [26]	2021	Patient preference elicitation instrument: OABCare	United States	Overactive bladder management - relative importance of the key outcomes associated with overactive bladder management (behavioral modification, medications, physical therapy, etc)	Overactive bladder
Cole et al [27]	2022	Electronic health care tool: PRIME	United States	Choice of chemotherapeutic agents for patients with hematologic malignancies - relative importance of outcomes for blood cancer treatments	Hematologic malignancies
de Achaval et al [28]	2012	Adaptive conjoint analysis tool	United States	Choice of different treatment options ("knee replacement surgery" or "pills and physical therapy" versus "unsure") and relative importance of attributes and therapy or total knee arthroplasty	Knee osteoarthritis
Dowsey et al [29]	2016	Decision aid in total knee arthroplasty	Australia	Choice of whether or not to undergo total knee arthroplasty surgery	End-stage knee osteoarthritis
Fraenkel et al [30]	2007	Computer tool	United States	Choice of different treatment options for knee pain	Knee pain, osteoarthritis
Goodsmith et al [31]	2021	Computerized conjoint analysis	United States	Choice of weight management treatment for overweight individuals with schizophrenia	Overweight individuals with schizophrenia
Hawley et al [32]	2016	Interactive, web-based, breast cancer treatment decision tool	United States	Choice of locoregional treatment (mastectomy or lumpectomy with radiation)	Breast cancer
Hazelwood et al [13,33,34]	2020, 2018, and 2016	Decision aid	Canada	Choice between two treatments	Early rheumatoid arthritis
Hess et al [35]	2015	Preference elicitation tool	United States	Choice of medical, surgical treatment or opt-out for abnormal uterine bleeding	Abnormal uterine bleeding
Hutyra et al [36]	2019	FTASD[b] decision tool or Preference-Based Decision Aids	United States	Choice between operative or nonoperative treatment for first-time anterior shoulder dislocation	Anterior shoulder dislocations
Jayadevappa et al [37-39]	2015, 2019a, and 2019b	Patient Preferences for Prostate Cancer Care: PreProCare	United States	Choice between treatment options (ie, active surveillance, surgery, radiation) for early-stage prostate cancer	Prostate cancer
Johnson et al [40]	2016	Patient decision aid	United States	Choice of surgery, radiation therapy or active surveillance for prostate cancer	Newly diagnosed prostate cancer (Men)
Loria-Rebolledo et al [41]	2022	Understanding Persistent Pain Decision Aid Tool: UPP[c] DAT[d]	United Kingdom	Choice of medication for managing persistent pain	Patients living with persistent pain
Pieterse et al [42]	2019	Value Clarification Method - ABEL study	Netherland	Choice of whether or not to undergo short-course preoperative radiotherapy treatment	Newly diagnosed patients with rectal cancer
Pieterse et al [43]	2010	ACA-questionnaire	Netherland	Choice of surgery vs preoperative radiotherapy (PRT) plus surgery for patients with rectal cancer - relative importance of outcomes for rectal cancer	Rectal cancer
Rochon et al [44] and Fraenkel [45]	2014and 2010	Adaptive conjoint analysis decision aid	United States	Choice of different treatment options for knee pain	Knee pain

XSL•FO
RenderX

Study	Date	Name of the tool	Country	Purpose of tool	Disease context
Snaman et al [46-48]	2019, 2021, and 2021	Decision-making tool: MyPref	United States	Choice of chemotherapy treatments	Adolescent and young adults with high-risk cancer
Streufert et al [49]	2017	Shared decision-making tool	United States	Choice of operative and nonoperative treatment for first-time anterior shoulder dislocation	First-time anterior shoulder dislocation
Studfts et al [50] and Byrne et al [51]	2020and 2019	Brief Education and a Conjoint Valuation Survey	United States	Choice of lung cancer screening test	Individuals at high risk of lung cancer
Wittnik et al [52,53]	2018 and 2016	Customized Care - Discussion Prioritization Tool	United States	To help patients disclose their stressors to their primary care provider	Patients with multiple chronic medical conditions in primary care

[a]Two of the 33 studies were excluded from this table because they focused on method development [14,54].

[b]FTASD: first-time anterior shoulder dislocation.

[c]UPP: Understanding Persistent Pain.

[d]DAT: digital decision aid tool.

The tools were developed in various disease contexts, which can be broadly grouped into three categories: musculoskeletal (7/22, 32%), oncological (8/22, 36%), and chronic conditions (7/22, 32%; Table 1). Musculoskeletal conditions include knee pain, early rheumatoid arthritis, anterior shoulder dislocations, and end-stage knee osteoarthritis; oncological conditions include breast cancer, prostate cancer, lung cancer, rectal cancer, hematologic malignancies, and adolescent and young adults with high-risk cancer; and chronic conditions include cardiovascular disease, overweight individuals with schizophrenia, ulcerative colitis, and Crohn disease, patients with multiple chronic medical conditions in primary care, and overactive bladder. The majority of the tools (20/22, 91%) were developed to help patients choose between medical treatments or compare medical treatments with surgery or surveillance. However, 2 out of 22 (9%) tools had different purposes, one study was developed to help patients disclose their stressors (ie, mobility issues and money worries) to their primary care provider and the other to facilitate the decision-making process around the choice of lung cancer screening tests.

Attributes

The various types of attributes that were used to assess patient preferences are listed in Table 2. The COMET taxonomy was used to organize the 91 attributes into meaningful domains and gain insight into the breadth of attributes that were included in the choice tasks. Five core areas and 11 outcome domains were identified when attributes were classified (Table 2). Attributes that were predominantly used in decision tools are: efficacy (20/22, 91%), side effects (15/22, 68%), route of administration (9/22, 41%), cost (7/22, 32%), and limits on daily activities (6/22, 27%). Further details of the attributes are available in Multimedia Appendix 4.

Overall, a range of attributes from 3 to 15 were used in the decision tools (Figure 2). Most tools (8/22, 36%) with a higher number of attributes (8 or above), used a design that allowed them to simplify the choice task and only display a reduced number of attributes (ie, 3 attributes at a time) to the respondents. Further details about the choice of tasks can be found in Multimedia Appendices 4 and 5.

XSL•FO
RenderX

Table 2. Classification of attributes.

Core area, outcome domain, and attributes (in order)	Frequency (%)[a]
Death	
Mortality or survival	
Survival	2 (9)
Physiological or clinical	
Physiological or clinical	
Efficacy	20 (91)
Recurrence	2 (9)
How the medication works	1 (5)
Life impact	
Functioning	
Limits on daily activities	6 (27)
Mobility	1 (5)
Impact on social life	1 (5)
Appearance	1 (5)
Global quality of life	
Quality of life	1 (5)
Perceived health status	
Stress	1 (5)
Delivery of care	
Route of administration	9 (41)
Recovery time	3 (14)
Treatment location	4 (18)
Duration of treatment	2 (9)
Need for additional treatment	1 (5)
Check-ups	1 (5)
Type of provider	1 (5)
Delivery mode (Internet or in-person)	1 (5)
Frequency of clinic visits	1 (5)
Individual or group intervention	1 (5)
Personal circumstances	
Safety	1 (5)
Food	1 (5)
Money worries	1 (5)
Resource use	
Economic	
Cost	7 (32)
Hospital	
Time in hospital	3 (14)
Permanence	1 (5)
Societal or care burden	
Caregiver burden	1 (5)
Adverse events	

XSL•FO
RenderX

Core area, outcome domain, and attributes (in order)	Frequency (%)[a]
Adverse events	
Side effects	15 (68)
False Positive	1 (5)

[a]The percentage is calculated by dividing the frequency by the total number of tools (n=22).

Figure 2. Number of attributes included in the tools.

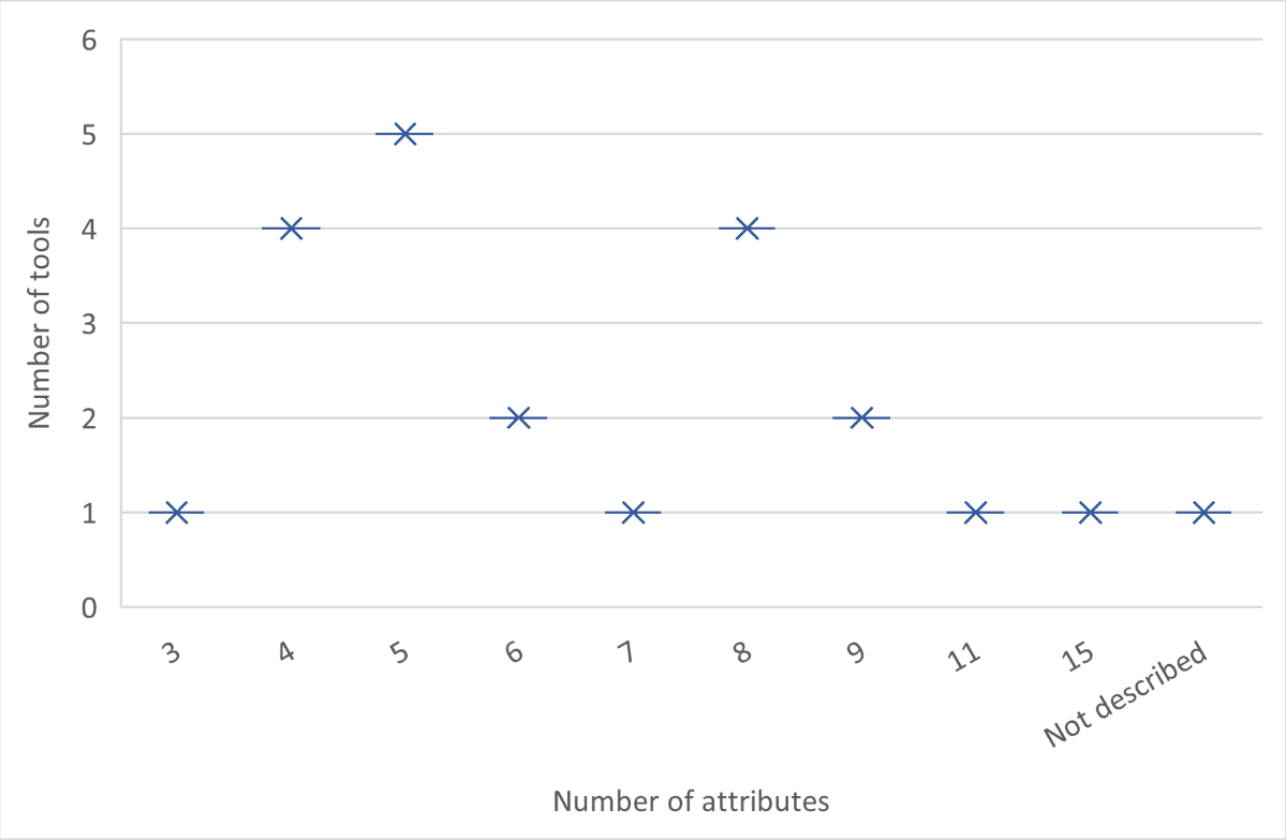

Figure 3. Horizontal bar graph reproduced from [42] which is published under Creative Commons Attribution 4.0 International License [55].

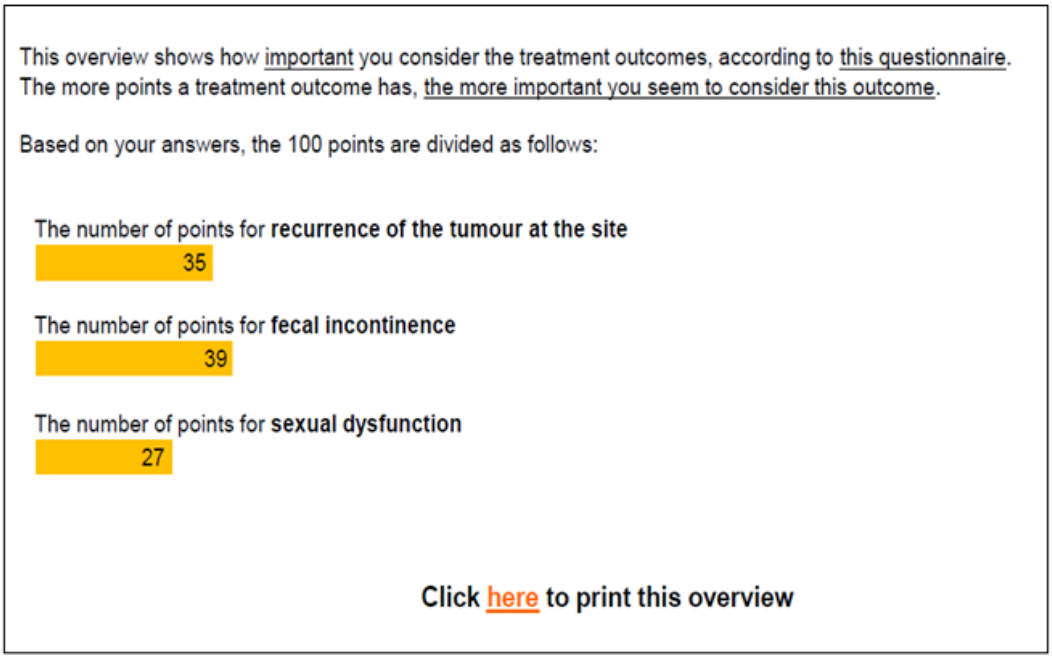

XSL•FO
RenderX

Figure 4. Vertical bar graph reproduced from [41] which is published under Creative Commons Attribution 4.0 International License [56].

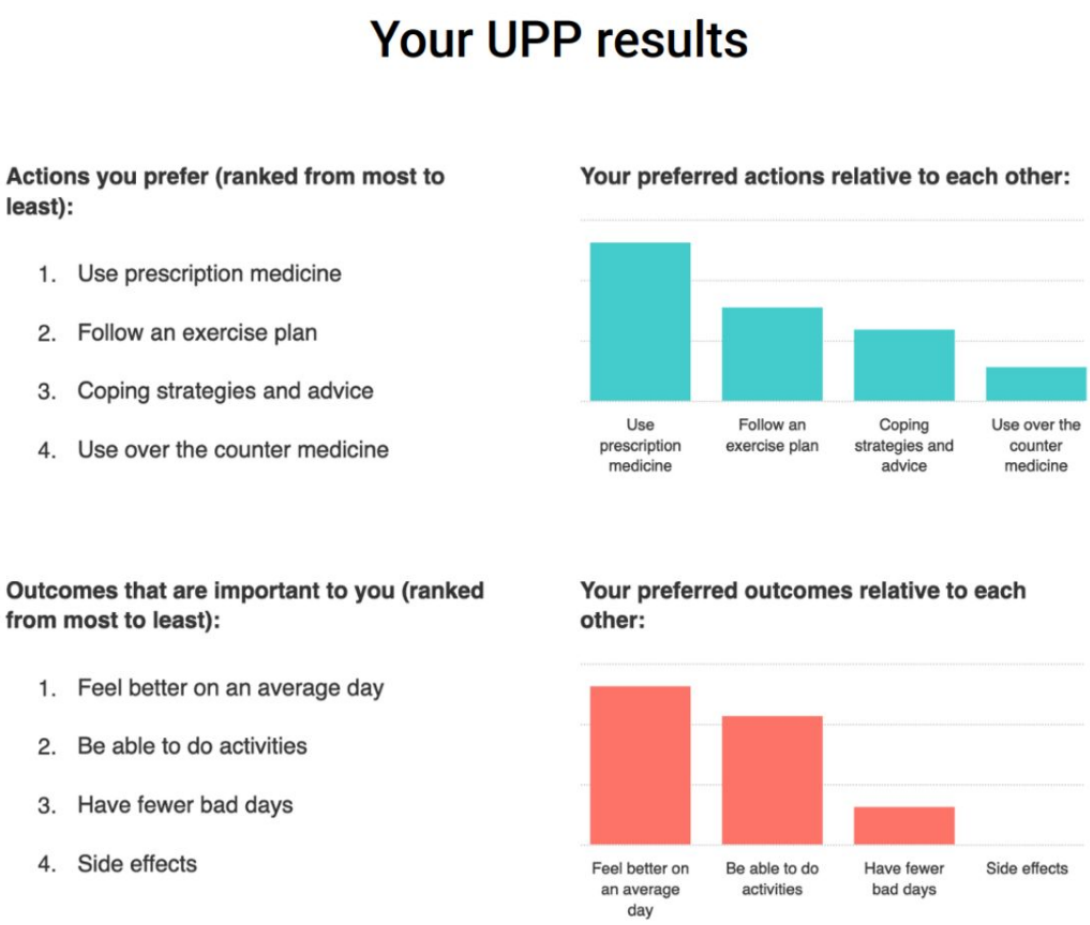

Figure 5. Gauge chart reproduced from [27] which is published under Creative Commons Attribution 4.0 International License [57]. AML: acute myeloid leukemia.



XSL•FO
RenderX

Figure 6. Line graph reproduced from [27] which is published under Creative Commons Attribution 4.0 International License [57].

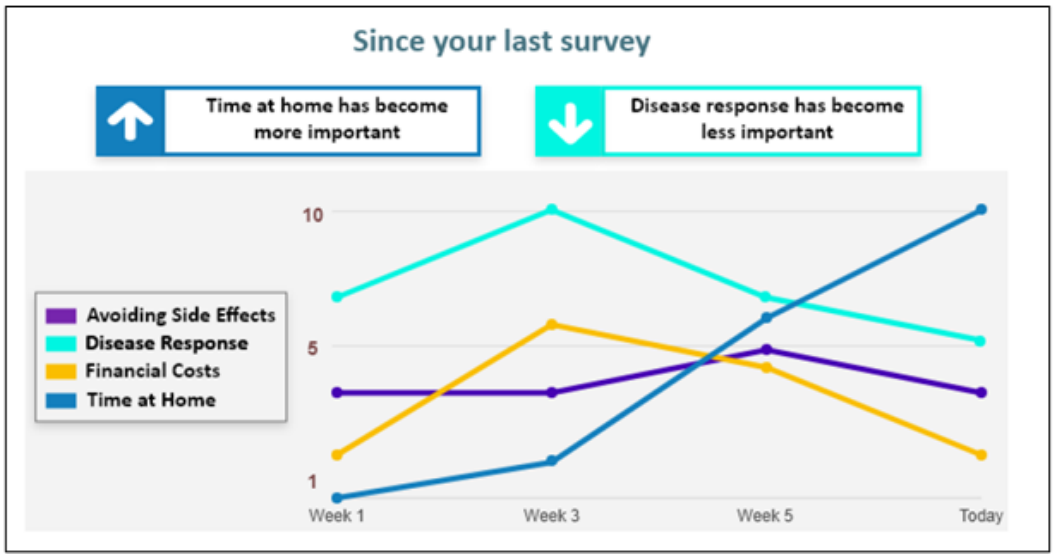

Figure 7. Narrative visualization reproduced from [27] which is published under Creative Commons Attribution 4.0 International License [57].

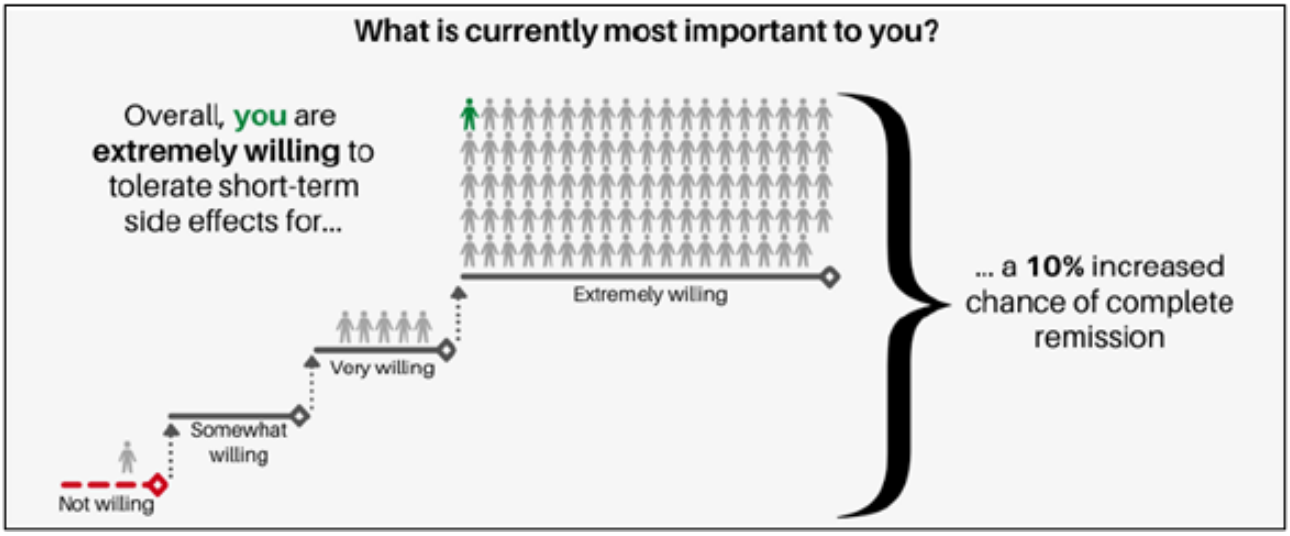

Defining Risk or Efficacy Attributes

The majority of tools (20/22, 91%) incorporated probabilistic attributes to convey uncertainty in treatment efficacy and side effects (Table 3). To enhance comprehension, various approaches were used to express these probabilities, including percentages (9/22, 41%), natural frequencies (9/22, 41%), icon arrays (8/22, 36%), narratives (5/22, 23%), and videos (1/22, 5%). Over half of the tools (13/22, 59%) used a combination of methods, while a minority (4/22, 18%) relied on a single approach. In total 14% (3/22) tools did not clearly describe the method used. The most frequent combination was natural frequencies and icon arrays (7/22, 32%).

XSL•FO
RenderX

Table 3. Methods used to define the risk or efficacy attributes[a].

Study	Percentages (ie, 5% risk of the cancer coming back)	Natural frequencies (ie, 10 out of 100 people will have a heart attack)	Icon arrays (numerical data visualized using simple icons like faces)	Narratives (ie, More than half; side effects - likely; low likelihood)	Videos
Abraham et al [24]		✓	✓		
Almario et al [25]	✓	✓	✓		✓
Chhatre et al [26]	✓			✓	
Cole et al [27]	✓			✓	
De Achaval et al [28]		✓	✓		
Dowsey et al [29]		✓	✓		
Fraenkel et al [30]		✓	✓		
Goodsmith et al [31]	N/A[b]	N/A	N/A	N/A	N/A
Hawley et al [32]	✓			✓	
Hazelwood et al [13,33,34]		✓	✓		
Hess et al [35]	Not described	Not described	Not described	Not described	Not described
Hutyra et al [36]	Not described	Not described	Not described	Not described	Not described
Jayadevappa et al [37-39]	✓			✓	
Johnson et al [40]	Not described	Not described	Not described	Not described	Not described
Loria-Rebolledo et al [41]				✓	
Pieterse et al [42]		✓	✓		
Pieterse et al [43]		✓	✓		
Rochon et al [44] and Fraenkel et al [45]	✓	✓			
Snaman et al [46-48]	✓				
Streufert et al [49]	✓				
Studfts et al [50] and Byrne et al [51]	✓				
Wittnik et al [52,53]	N/A	N/A	N/A	N/A	N/A
Total count	9	9	8	5	1

[a]Two of the 33 studies were excluded from this table because they focused on method development [14,54].

[b]N/A: Not applicable.

Choice Task Design

Five types of designs were used to derive the choice tasks (Table 4). The most frequently used design type was adaptive conjoint analysis (10/22, 45%), followed by conjoint analysis (4/22, 18%), DCEs (4/22, 18%), modified adaptive conjoint analysis (3/22, 14%) and adaptive best-worst conjoint analysis (1/22, 5%). Common characteristics of the choice tasks were the rating of preferred treatments (ie, “strongly prefer left,” to “strongly prefer right”) in 17 out of 22 (77%) tools; and tailoring of choice tasks using various adaptive designs (14/22, 64%). Further details about the different designs can be found in Multimedia Appendix 5.

The number of choice tasks included varied from 6 to 20 (Multimedia Appendix 5). Eight out of 22 tools (36%) did not specify the number of tasks included in the tool. Conjoint analysis typically used a higher range of 16-20 tasks, whereas DCEs and adaptive conjoint analysis designs used a lower range of 6-12 tasks (Multimedia Appendix 5).

XSL•FO
RenderX

Table 4. Types of choice of tasks.

Choice task type	Frequency	Details of the choice task
Adaptive conjoint analysis (adaptive choice-based conjoint, choice-based adaptive conjoint analysis)	10	First, participants were shown an attribute with two different levels and asked to rate how important was the level difference (ie, "strongly prefer left or right"). Second, respondents were shown tailored paired comparisons and asked to rate their preferences (ie "strongly prefer left or right").
Conjoint analysis	4	Paired comparison (or single profile) where respondents rated their preference (ie, strongly prefer treatment option 1/2).
Discrete choice experiment	4	Paired comparisons of treatment profiles where respondents chose one treatment over the other (no rating involved). Efficient designs or balanced overlap designs were generated using Ngene or Sawtooth.
Modified adaptive conjoint analysis	3	First, respondents chose the attribute that is most important to them out of a list of attributes. Second, the respondents rated the importance of each attribute relative to the most important attribute; followed by paired comparisons of treatment profiles which were rated (ie, "strongly prefer left or right").
Adaptive best-worst conjoint analysis	1	Presented three tailored alternative profiles and asked respondents to choose the best and worst of alternatives.

Embedding Choice Tasks in Decision Tools

Table 5 details the choice experiment embedding methods used by the 22 decision tools included in the scoping review. In addition, Table 5 presents a summary of the two methodological papers included [14,54]. Fourteen decision tools used Sawtooth software; one study each used Wisercare [40], Dynamic Computer Interactive Decision Application [13] and Clinvivo software [41]; and 5 tools failed to describe the choice task embedding method.

Sawtooth Software dominated, comprising approximately 64% (14/22) of the decision tools, which allowed real-time data analysis and display of personalized results to respondents. However, details on how the Sawtooth software provided real-time tailored results were limited (n=7). Even when the authors provided information beyond the Sawtooth software name, they only mentioned the name of the regression model used such as hierarchical Bayes (n=3) or least-squares regression analysis (n=4) without providing any specific details regarding the model configuration or the mechanisms enabling real-time personalization (Table 5). Similarly, one study used the Wisercare software, but no further details were provided. The remaining four studies used distinct methods to embed choice tasks in decision tools and these methods are summarized in detail below.

Hazelwood et al [13] used the Dynamic Computer Interactive Decision Application templating tool to develop the decision aid [13]. To provide tailored treatment ranking, DCE data from a previous study [34] was analyzed using a Bayesian model followed by scenario analyses [33]. The decision aid contained 6 DCE tasks. By analyzing the results of the prior DCE, a choice probability was calculated for each of the 64 (2^6) response profiles (ie, AABAAB, ABABAB, BBBAAA…etc) for selecting one of two possible treatments (triple therapy or methotrexate). When a patient completed the 6 choice tasks, the patient was assigned to 1 out of 64 response profiles and the corresponding probability for that response profile of choosing triple therapy or methotrexate was displayed [13].

Gonzalez Sepulveda et al [14] also used data from a prior conjoint analysis [36] to develop a preference diagnostic tool. This paper detailed only the methodology with a simulation study to test the proposed method. First, they analyzed the conjoint analysis data to evaluate the distribution of preferences and identify groups or classes of respondents with similar preferences. The authors refer to these groups as "known preference phenotypes." To identify preference phenotypes, Gonzalez Sepulveda et al [14] suggested using methods such as k-means cluster analysis, latent class analysis, or hierarchical cluster analysis. The second step involved constructing a small number of choice sets that would "maximize discrimination" of respondents belonging to each preference phenotype. To generate choice sets the authors used an evolutionary algorithm. Third, Gonzales Sepulveda et al [14] evaluated the robustness (ie, true positives or true negatives) of the choice sets on correctly predicting preference phenotypes [14]. In the simulation tasks, the investigators constructed three choice tasks, where all respondents saw choice task 1 and based on the answer to the first choice task, respondents either saw choice task 2 or choice task 3. Each new question classified the respondent to the preference phenotype they were likely to belong to. Based on the response patterns to the two questions, the investigators assigned respondents to their likely preference phenotypes [14]. Gonzalez Sepulveda et al [14] argued that a key advantage of their method is the use of a small number of choice tasks to accurately predict the posterior probability of preference phenotype membership.

XSL•FO
RenderX

Table 5. Methods of tool development.

Study	Choice task embedding approach
Abraham et al [24]	Sawtooth software (regression analysis)
Almario et al [25]	Sawtooth software (hierarchical Bayes regression)
Chhatre et al [26]	Sawtooth software (hierarchical Bayesian random effects regression)
Cole et al [27]	Not specified but this tool is a work in progress only protocol published
de Achaval et al [28]	Sawtooth software (no further details provided)
Dowsey et al [29]	Not described
Fraenkel et al [30]	Sawtooth software (least-squares regression analysis)
Gonzalez et al [14]	Developed an algorithm "based on previous information on the distribution of patient preferences in a population" using hierarchical cluster analysis
Goodsmith et al [31]	Not described
Hawley et al [32]	Not described
Hazelwood et al [13,33,34]	Hierarchical Bayes regression from a prior discrete choice experiments (DCEs) is used to find expected choice probabilities of treatments for response patterns. Software used was Dynamic Computer Interactive Decision Application (DCIDA) tool
Hess et al [35]	Sawtooth software (no further details provided)
Hutyra et al [36]	Sawtooth software (no further details provided)
Jayadevappa et al [37-39]	Sawtooth software (no further details provided)
Johnson et al [40]	Wisercare no further details provided)
Loria-Rebolledo et al [41]	Real time estimates using the penalized logit regression coded using Clinvivo software
Pieterse et al [42]	Sawtooth software (Ordinary least squares regression)
Pieterse et al [43]	Sawtooth software (Ordinary least squares regression)
Rochon et al [44] and Fraenkel et al [45]	Sawtooth software (Least squares updating algorithm)
Snaman et al [46-48]	Sawtooth software (no further details provided)
Shaoibi et al [54]	Bayesian collaborative filtering model
Streufert et al [49]	Sawtooth software (no further details provided)
Studfts et al [50] and Byrne et al [51]	Sawtooth software (hierarchical Bayes approach)
Wittnik et al [52,53]	Not described

Shaoibi et al [54] used a Bayesian collaborative filtering model to predict the treatment recommendations. The first step of the proposed method is similar to Gonzales Sepulveda et al [14] as it involves the use of preference phenotypes. A Markov Chain Monte Carlo algorithm was used to identify preference clusters [54]. Then the investigators used data from existing patients to assess posttreatment satisfaction. When a new patient completed the choice task, they were assigned to a preference cluster. Then post treatment satisfaction data that match the cluster were used to make a treatment recommendation.

Loria-Rebolledgo et al [41] used a different approach to the other methods discussed above. This method did not require a 2-stage approach of analyzing data from a previous choice experiment instead, parameters were estimated "live" using a penalized multinomial logit model (pMNL). The Clinvivo software was used to code the decision aid tool. Loria-Rebolledgo et al [41] selected the penalized model because of its flexibility in converging results especially when small sample sizes are used for the estimation. The pMNL is different from a typical multinomial logit model because a bias term is added to the standard likelihood function. This term penalized the model for small sample size bias. Each time a respondent completed the choice task, the pMNL model was run and relative importance scores were calculated.

Presentation of the Decision Tool

After completing the choice task, 64% of tools (14/22) provided respondents with a report, while 8 out of 22 (36%) tools did not provide a clear description of what respondents received. Two types of information were included in this report: (1) attribute importance scores and (2) "best match" treatment options that aligned with patients' desired attributes. Most patients received feedback on attribute importance (16/22, 73%), which was presented to patients using different formats. The majority (14/22, 64%) were displayed as horizontal bar graphs (Figure 3 [42]). There were slight variations in the presentation of horizontal bar graphs: some showed exact percentages, some

XSL•FO
RenderX

showed longer bars to represent increasing importance, and some tools with a long list of attributes only presented the top 5 most important attributes. Diverging from horizontal bar graphs, one tool used a vertical bar graph (Figure 4 [41]). Another study tested multiple different approaches such as a gauge chart (Figure 5 [27]), line graph (Figure 6 [27]) and narrative visualization (Figure 7 [27]). Only 5 out of the 22 (23%) tools provided patients with a "best match" treatment. Two tools used a scale ranging from 0 (worst choice) to 100 (best choice) showing the relative ranking of all the available treatment options (Figure 8 [44]). One tool displayed all available treatments in a choice task format and highlighted the "best match" (Figure 9 [13]); one tool picked the treatment that best fits the respondent and only presented that option in the text and the last tool did not provide sufficient details on how the ranking of treatment options was presented to the patients (Multimedia Appendix 6).

Figure 8. Scale displaying relative ranking of treatments reproduced from [44]. Used with permission of John Wiley and Sons - Books, from Rochon et al, 2014;17(6):840-5; permission conveyed through Copyright Clearance Center, Inc.

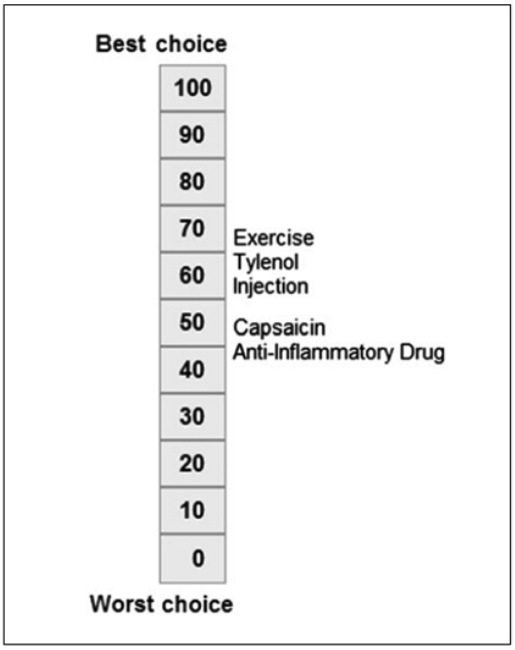

Figure 9. "Best match" treatment presented in a choice task format [13]. Patient Preference and Adherence 2020:14 829-838 - Originally published by and used with permission from Dove Medical Press Ltd.

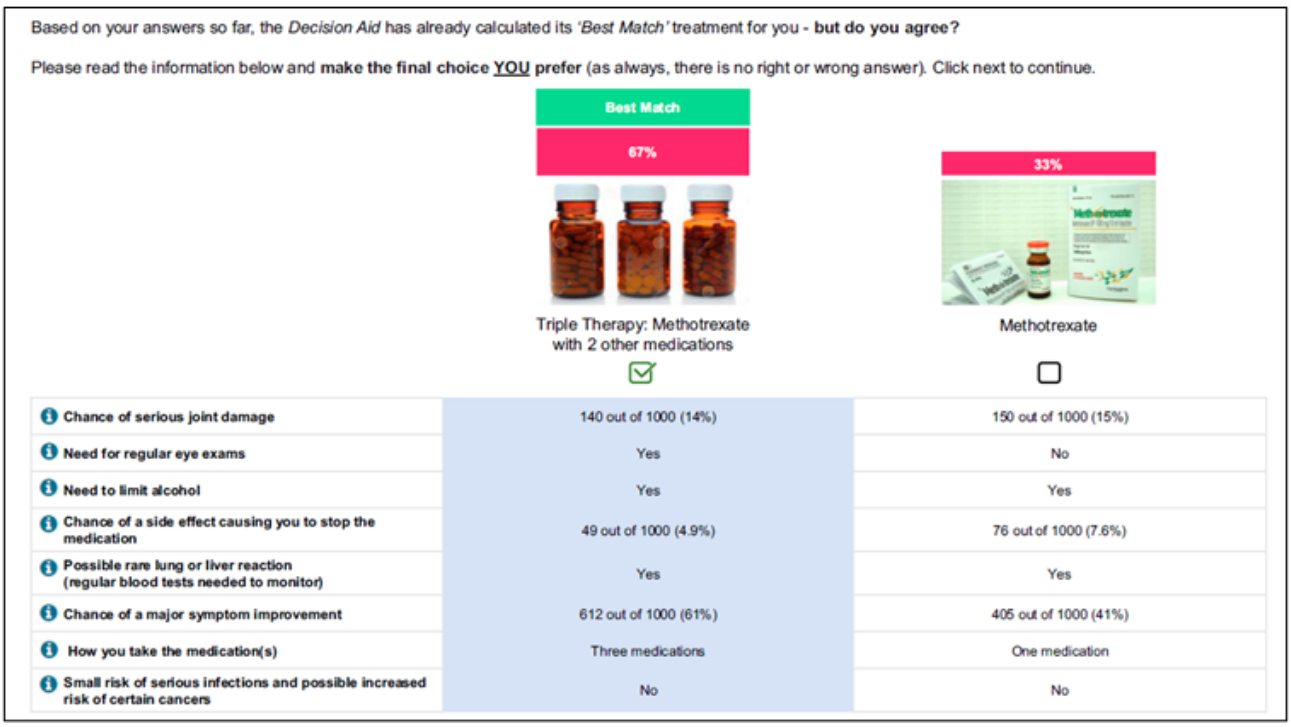

XSL•FO
RenderX

Evaluation of Decision Tools

The decision tools were evaluated in the context of different study designs, including randomized controlled trials (RCTs; n=5), pilot RCTs (n=4), mixed methods studies (n=4), cohort studies (n=4), cross-sectional studies (n=3), qualitative studies (n=1), and nonrandomized studies (n=1). Sample sizes across the studies varied, ranging from 23 participants in mixed methods studies to 743 in RCTs (Multimedia Appendix 7).

Figure 10 shows the breadth of outcomes used to evaluate the decision tools, with over 40 different outcomes included. These outcomes fall into 6 core areas including usability, acceptability, validity, feasibility, informed decision making and patient health outcomes. A variety of instruments were used to assess the usability or user-friendliness of the tools, these include the system usability scale, Dowding's usability principles checklist, single easy questionnaire, poststudy system usability questionnaire, IBM computer system usability questionnaire, and NASA (National Aeronautics and Space Administration) task load index (evaluates mental effort needed to perform tasks). While usability was measured using existing validated instruments, accessibility was mostly measured using items developed by the study teams. The most frequently used acceptability item was whether respondents would discuss the results of the tool with their health care professional, followed by how helpful the tool was to patients in deciding which treatment to choose. The validity of the tool was assessed by examining how well its predictions aligned with patient perspectives. Five tools measured "value concordance," measuring how well the tool's attribute rankings aligned with the patient's own preferences. Two tools measured the similarity between the "best match" treatment generated by the tool and the patient's stated treatment preference. The feasibility of the tool was assessed by recording response rates (n=1) and evaluating the ability of the tool to generate individually tailored reports (n=1).

Informed decision making was assessed using several proxy measures including decisional regret, decisional conflict, self-efficacy, knowledge, patient activation or engagement, and satisfaction. Decisional conflict defined as the "personal uncertainty about which option to choose" was the most commonly used outcome measure (n=6). Decisional regret is defined as "remorse or distress over a decision" and was measured using the 5-item decisional regret scale (n=2) and memorial anxiety scale for prostate cancer regret subscale (n=1). Respondents' self-efficacy or their confidence to make a treatment decision was measured using the decision self-efficacy scale (n=1) and arthritis self-efficacy scale (n=1). Patients' active participation and engagement in care were measured using validated instruments including the preparation for decision making scale (n=2), patient activation measure (n=1), control preferences scale (n=1), and adapted perceived competence scale (n=1). Patient satisfaction was also measured using established instruments such as the six-item satisfaction with decision scale (n=2), patient satisfaction questionnaire (n=1), patient satisfaction scale (n=1), and functional assessment of chronic illness therapy-treatment satisfaction-patient scale (n=1). A variety of patient health outcomes were also collected to measure changes to health outcomes after using the tool, further details are available in Figure 10 and Multimedia Appendix 8.

XSL•FO
RenderX

Figure 10. Outcomes. SURE: Sure of myself; Understand information; Risk-benefit ratio; Encouragement.



Discussion

Principal Findings

With the increasing use of choice experiments embedded in decision tools, this scoping review was conducted to map the evidence base in terms of the methods, design, and evaluation of these tools. This review identified 22 decision tools containing a choice experiment for a variety of health conditions including musculoskeletal conditions, oncological conditions, and chronic conditions. The development of these decision tools was led by the United States with the remaining tools originating in the Netherlands, United Kingdom, Canada, and Australia. A notable trend is the increasing use of choice experiments as a value clarification exercise within decision aids, particularly from 2015 onwards. Variations were observed in how the findings from the choice experiment were presented to patients. Most patients received feedback on attribute importance, while only a minority received a "best match" ranking of treatment options. Diverse approaches to presenting probabilities to participants were identified, with a common approach being the use of mixed methods such as percentages, natural frequencies, and icon arrays. A lack of consensus on the terminology used to describe the developed tools was found. Variation was noted in the study designs and outcome measures used to evaluate the tools. The decisional conflict scale was the

XSL•FO
RenderX

most frequently used outcome measure, but no other instrument was widely adopted.

This review also details four relatively new proof-of-concept methods for embedding choice experiments in decision aids, that have been published since 2020. These 4 novel methods include analyzing previous choice experiment data to develop algorithms (scenario analyses) containing expected choice probabilities of treatments according to response patterns [13], classifying respondents into known preference phenotypes [14], using a Bayesian collaborative filtering model [54], or using a penalized multinomial logit model [41]. Developers attempting to embed choice experiments in decision tools are faced with some key challenges: (1) ensuring real-time availability of choice experiment results to patients, (2) tailoring the choice experiment results by incorporating heterogeneity within choice experiment models, and (3) minimizing the number of choice tasks to maintain user engagement alongside the existing information presented in the decision aid. While the new proof-of-concept methods demonstrate potential solutions to these challenges, their validity remains to be conclusively demonstrated. Current evidence is limited to proof-of-concept studies either using small sample sizes or using only simulated data. So, further research is needed to rigorously evaluate these methods and establish their validity.

Comparison With Previous Work

One of the notable findings of this review in comparison to the one conducted in 2018 is that more developers are using choice experiments as a value clarification method in decision tools. In 2018, Weernink et al [7] only identified 8 studies, but since then a further 14 tools have been developed. Despite this increase in research, many of the characteristics of these tools remain unchanged when compared to the results of the previous review [7]. For example, to present the relative importance of attributes it was common to use bar graphs where longer bars reflect relatively more important attributes. Notably, a protocol for a study exploring alternative visual presentations (ie, bar graph, gauge graph, narrative visualization, and line graph) was included in this review [27]. This represents a positive step towards gathering evidence on how to improve the user experience of these tools. Similarly, limited research exists on how to present the "best matched" treatment. Since presenting balanced information is a key criterion of a good decision tool [58], the table format of the choice experiment should lend itself well to presenting this result. Hazelwood et al [13] demonstrated a potential approach using this format. However, further research is needed to evaluate if this presentation method is understood by patients.

This review highlighted the diverse approaches used by studies to present probabilities to participants. It was common to use mixed methods such as percentages, natural frequencies, and icon arrays to describe risk or efficacy attributes in an accessible way to patients. The most common combination was to use natural frequencies and icon arrays to explain probabilities. This finding is congruent with the review conducted by Trevena et al [59] who also found that using visual formats alongside numerical formats helps to improve the understanding of probabilities. However, when combining different methods to display probabilities, caution should be used as there is also evidence to show that when certain methods are combined, such as verbal and numerical formats, probabilities can be overestimated [60]. Presenting probabilities in an unbiased way is challenging, and there is a vast literature exploring how best to do this [59,60]. Bonner et al [60] recommend using consistent numerical formats, such as "x in 100 over 5 years," to display probabilities.

Over 40 different outcome measures were used during the evaluation of decision tools. The decisional conflict scale emerged as the most frequently used outcome measure. Besides the decisional conflict scale, no other instrument was widely adopted to evaluate these decision tools. This variability in outcome measurement highlights the lack of consensus on a core set of outcomes to evaluate these tools. These findings were consistent with previous reviews of decision aids documenting inadequate reporting of details of outcome measures [61], which hinders the development of a robust evidence base of the effectiveness of decision aids.

Limitations

Several limitations should be considered when interpreting the findings of this scoping review. First, the study selection and extraction were performed by a single author. Despite this, the robustness of the selection process was ensured by following the protocol and making selection decisions in conjunction with the wider study team on articles that required a second opinion. Second, the initial search strategy may not have captured all relevant terms used to describe these tools. The included studies used diverse terminology, including "decision-making tool," "decision aid," "patient preference elicitation instrument," and "discussion prioritization tool." This lack of consistent terminology hindered the identification of potentially relevant studies. To address this potential limitation, an updated search was conducted in September 2023, but it did not yield any additional relevant studies. Although an updated search was conducted using additional terms, it remains possible that some tools were missing. Future reviews could benefit from a more iterative approach to search term development. Third, a risk of bias assessment was not conducted in this review. This approach was selected as aligned with the primary objective of the review, which was to provide a comprehensive overview of the existing literature.

Future Directions

The findings of this review point to several areas for future research. Further investigation into the newer proof-of-concept methods identified is warranted, as these methods are still in their infancy more research is needed to validate their effectiveness. Moreover, research on the optimal presentation of choice experiment results to patients is also needed. This includes exploring different visual formats and conveying "best match" treatment information in a clear, unbiased, and understandable way. Furthermore, the development of a core set of outcome measures for evaluating decision tools that incorporate choice experiments would be beneficial. This would facilitate comparisons across studies and help build a more robust evidence base regarding the effectiveness of these tools. Achieving consensus on the terminology used to describe these

XSL•FO
RenderX

tools would aid in indexing and retrieving relevant literature, facilitating future research. In addition, future studies should provide more detail on the model configurations used to enable real-time personalization, beyond just naming the regression model, to improve transparency. Moreover, to maintain their relevance, these decision aids must be adaptable, requiring ongoing updates to incorporate the latest advancements in treatment options. Further research is needed to determine how these tools can be efficiently updated to reflect the rapidly changing medical field. Finally, integrating decision aids effectively within existing electronic systems in hospitals requires further investigation. Implementation of these tools within existing patient portals or linking them with electronic medical records is likely to increase uptake and facilitate more informed and productive discussions with their health care providers. So future research could explore innovative delivery methods to incorporate these tools into routine care.

Conclusion

This scoping review provides a comprehensive overview of the different approaches to embedding choice experiments in decision tools. It highlights key considerations for future studies, including the choice of models used, the presentation of information to patients, and the selection of appropriate outcome measures for evaluating the tools. While several challenges remain, the field is rapidly evolving, and the findings of this review provide a foundation for further research and development in this area. The increasing use of digital technologies, including artificial intelligence, offers possibilities to enhance the interactivity, personalization, accessibility, and integration of decision aids into routine care.

Acknowledgments

We would like to thank Ruth Wong for her valuable input to help conduct the searches for this review. The authors attest that there was no use of artificial intelligence in the generation of text, figures and other information. This research was funded by the National Institute for Health Research (grant NIHR302489).

Data Availability

All data generated during this study are included in this published article and Multimedia Appendices 1-8.

Authors' Contributions

NW, DR, and PS contributed to conceptualization. NW contributed to data curation. NW contributed to the formal analysis. NW, DR, and PS contributed to funding acquisition. NW, DR, and PS contributed to the investigation. NW, DR, and PS contributed to the methodology. NW contributed to project administration. DR and PS was involved in supervision. NW was involved in visualization and writing the original draft. NW, DR, and PS contributed to reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
[DOCX File , 38 KB - jmir_v27i1e59209_app1.docx]

Multimedia Appendix 2
Search Strategy.
[DOCX File , 33 KB - jmir_v27i1e59209_app2.docx]

Multimedia Appendix 3
List of excluded references.
[DOCX File , 29 KB - jmir_v27i1e59209_app3.docx]

Multimedia Appendix 4
Further description of attributes and levels.
[DOCX File , 68 KB - jmir_v27i1e59209_app4.docx]

Multimedia Appendix 5
Further details of the choice tasks.
[DOCX File , 31 KB - jmir_v27i1e59209_app5.docx]

Multimedia Appendix 6

XSL•FO
RenderX

Design features of the tools.
[DOCX File , 31 KB - jmir_v27i1e59209_app6.docx]

Multimedia Appendix 7
Evaluation of the decision tool.
[DOCX File , 31 KB - jmir_v27i1e59209_app7.docx]

Multimedia Appendix 8
Further details of outcomes.
[DOCX File , 32 KB - jmir_v27i1e59209_app8.docx]

References

1. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. JAMA 2008;300(4):436-438. [doi: 10.1001/jama.300.4.436] [Medline: 18647988]
2. Wennberg JE. Tracking medicine: a researcher's quest to understand health care. United Kingdom: Oxford University Press; 2010.
3. Elwyn G, Frosch D, Rollnick S. Dual equipoise shared decision making: definitions for decision and behaviour support interventions. Implement Sci 2009;4:75 [FREE Full text] [doi: 10.1186/1748-5908-4-75] [Medline: 19922647]
4. Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. BMJ 2002;325(7370):961-964 [FREE Full text] [doi: 10.1136/bmj.325.7370.961] [Medline: 12399352]
5. Stacey D, Lewis K, Smith M, Carley M, Volk R, Douglas E, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2024;1(1):CD001431. [doi: 10.1002/14651858.CD001431.pub6] [Medline: 38284415]
6. Ostermann J, Brown DS, van Til JA, Bansback N, Légaré F, Marshall DA, et al. Support tools for preference-sensitive decisions in healthcare: where are we? Where do we Go? how do we get there? Patient 2019;12(5):439-443. [doi: 10.1007/s40271-019-00372-z] [Medline: 31256346]
7. Weernink MG, van Til JA, Witteman HO, Fraenkel L, IJzerman MJ. Individual value clarification methods based on conjoint analysis: a systematic review of common practice in task design, statistical analysis, and presentation of results. Med Decis Making 2018;38(6):746-755 [FREE Full text] [doi: 10.1177/0272989X18765185] [Medline: 29592585]
8. Stacey D, Bennett CL, Barry MJ, Col NF, Eden KB, Holmes-Rovner M, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2011(10):CD001431. [doi: 10.1002/14651858.CD001431.pub3] [Medline: 21975733]
9. Scalia P, Durand MA, Berkowitz JL, Ramesh NP, Faber MJ, Kremer JA, et al. The impact and utility of encounter patient decision aids: systematic review, meta-analysis and narrative synthesis. Patient Educ Couns 2019;102(5):817-841. [doi: 10.1016/j.pec.2018.12.020] [Medline: 30612829]
10. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
11. Witteman HO, Ndjaboue R, Vaisson G, Dansokho SC, Arnold B, Bridges JFP, et al. Clarifying Values: an updated and expanded systematic review and meta-analysis. Med Decis Making 2021;41(7):801-820 [FREE Full text] [doi: 10.1177/0272989X211037946] [Medline: 34565196]
12. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 2019;37(2):201-226 [FREE Full text] [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]
13. Hazlewood GS, Marshall DA, Barber CEH, Li LC, Barnabe C, Bykerk V, et al. Using a discrete-choice experiment in a decision aid to nudge patients towards value-concordant treatment choices in rheumatoid arthritis: a proof-of-concept study. Patient Prefer Adherence 2020;14:829-838 [FREE Full text] [doi: 10.2147/PPA.S221897] [Medline: 32546977]
14. Gonzalez Sepulveda JM, Johnson FR, Reed SD, Muiruri C, Hutyra CA, Mather RC. Patient-preference diagnostics: adapting stated-preference methods to inform effective shared decision making. Med Decis Making 2023;43(2):214-226. [doi: 10.1177/0272989X221115058] [Medline: 35904149]
15. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]
16. Campbell F, Tricco AC, Munn Z, Pollock D, Saran A, Sutton A, et al. Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family. Syst Rev 2023;12(1):45 [FREE Full text] [doi: 10.1186/s13643-023-02178-5] [Medline: 36918977]
17. Khalil H, Peters MD, Tricco AC, Pollock D, Alexander L, McInerney P, et al. Conducting high quality scoping reviews-challenges and solutions. J Clin Epidemiol 2021;130:156-160. [doi: 10.1016/j.jclinepi.2020.10.009] [Medline: 33122034]

XSL•FO
RenderX

18. Peters MD. In no uncertain terms: the importance of a defined objective in scoping reviews. JBI Database System Rev Implement Rep 2016;14(2):1-4. [doi: 10.11124/jbisrir-2016-2838] [Medline: 27536788]
19. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
20. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol 2018;96:84-92 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.12.020] [Medline: 29288712]
23. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PLoS One 2014;9(6):e99111 [FREE Full text] [doi: 10.1371/journal.pone.0099111] [Medline: 24932522]
24. Abraham N, Naik AD, Street RL, Castillo DL, Deswal A, Richardson PA, et al. Complex antithrombotic therapy: determinants of patient preference and impact on medication adherence. Patient Prefer Adherence 2015;9:1657-1668 [FREE Full text] [doi: 10.2147/PPA.S91553] [Medline: 26640372]
25. Almario CV, Keller MS, Chen M, Lasch K, Ursos L, Shklovskaya J, et al. Optimizing selection of biologics in inflammatory bowel disease: development of an online patient decision aid using conjoint analysis. Am J Gastroenterol 2018;113(1):58-71. [doi: 10.1038/ajg.2017.470] [Medline: 29206816]
26. Chhatre S, Newman DK, Wein AJ, Jayadevappa R. Patient preference elicitation instrument, OABCare, to facilitate shared decision-making in the care of overactive bladder. Neurourol Urodyn 2021;40(3):791-801. [doi: 10.1002/nau.24618] [Medline: 33565118]
27. Cole A, Richardson DR, Adapa K, Khasawneh A, Crossnohere N, Bridges JFP, et al. Development of a patient-centered preference tool for patients with hematologic malignancies: protocol for a mixed methods study. JMIR Res Protoc 2022;11(6):e39586 [FREE Full text] [doi: 10.2196/39586] [Medline: 35767340]
28. de Achaval S, Fraenkel L, Volk RJ, Cox V, Suarez-Almazor ME. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. Arthritis Care Res (Hoboken) 2012;64(2):229-237 [FREE Full text] [doi: 10.1002/acr.20646] [Medline: 21954198]
29. Dowsey MM, Scott A, Nelson EA, Li J, Sundararajan V, Nikpour M, et al. Using discrete choice experiments as a decision aid in total knee arthroplasty: study protocol for a randomised controlled trial. Trials 2016;17(1):416 [FREE Full text] [doi: 10.1186/s13063-016-1536-5] [Medline: 27542601]
30. Fraenkel L, Rabidou N, Wittink D, Fried T. Improving informed decision-making for patients with knee pain. J Rheumatol 2007;34(9):1894-1898. [Medline: 17696268]
31. Goodsmith N, Cohen AN, Flynn AWP, Hamilton AB, Hellemann G, Nowlin-Finch N, et al. Computerized conjoint analysis of the weight treatment preferences of individuals with schizophrenia. Psychiatr Serv 2021;72(3):288-294 [FREE Full text] [doi: 10.1176/appi.ps.202000068] [Medline: 33430650]
32. Hawley ST, Newman L, Griggs JJ, Kosir MA, Katz SJ. Evaluating a decision aid for improving decision making in patients with early-stage breast cancer. Patient 2016;9(2):161-169. [doi: 10.1007/s40271-015-0135-y] [Medline: 26178202]
33. Hazlewood GS, Bombardier C, Tomlinson G, Marshall D. A bayesian model that jointly considers comparative effectiveness research and patients' preferences may help inform GRADE recommendations: an application to rheumatoid arthritis treatment recommendations. J Clin Epidemiol 2018;93:56-65 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.10.003] [Medline: 29051108]
34. Hazlewood GS, Bombardier C, Tomlinson G, Thorne C, Bykerk VP, Thompson A, et al. Treatment preferences of patients with early rheumatoid arthritis: a discrete-choice experiment. Rheumatology (Oxford) 2016;55(11):1959-1968. [doi: 10.1093/rheumatology/kew280] [Medline: 27477807]
35. Hess LM, Litwiller A, Byron J, Stutsman J, Kasper K, Learman LA. Preference elicitation tool for abnormal uterine bleeding treatment: a randomized controlled trial. Patient 2015;8(2):217-227 [FREE Full text] [doi: 10.1007/s40271-014-0078-8] [Medline: 25074354]
36. Hutyra CA, Smiley S, Taylor DC, Orlando LA, Mather RC. Efficacy of a preference-based decision tool on treatment decisions for a first-time anterior shoulder dislocation: a randomized controlled trial of at-risk patients. Med Decis Making 2019;39(3):253-263. [doi: 10.1177/0272989X19832915] [Medline: 30834817]
37. Jayadevappa R, Chhatre S, Gallo JJ, Malkowicz SB, Schwartz JS, Wittink MN. Patient-centered approach to develop the patient's preferences for prostate cancer care (PreProCare) tool. MDM Policy Pract 2019;4(1):2381468319855375 [FREE Full text] [doi: 10.1177/2381468319855375] [Medline: 31259248]

XSL•FO
RenderX

38. Jayadevappa R, Chhatre S, Gallo JJ, Wittink M, Morales KH, Lee DI, et al. Patient-centered preference assessment to improve satisfaction with care among patients with localized prostate cancer: a randomized controlled trial. J Clin Oncol 2019;37(12):964-973. [doi: 10.1200/JCO.18.01091] [Medline: 30860943]
39. Jayadevappa R, Chhatre S, Gallo JJ, Wittink M, Morales KH, Bruce Malkowicz S, et al. Treatment preference and patient centered prostate cancer care: Design and rationale. Contemp Clin Trials 2015;45(Pt B):296-301. [doi: 10.1016/j.cct.2015.09.024] [Medline: 26435200]
40. Johnson DC, Mueller DE, Deal AM, Dunn MW, Smith AB, Woods ME, et al. Integrating patient preference into treatment decisions for men with prostate cancer at the point of care. J Urol 2016;196(6):1640-1644. [doi: 10.1016/j.juro.2016.06.082] [Medline: 27346032]
41. Loría-Rebolledo LE, Ryan M, Bond C, Porteous T, Murchie P, Adam R. Using a discrete choice experiment to develop a decision aid tool to inform the management of persistent pain in pharmacy: a protocol for a randomised feasibility study. BMJ Open 2022;12(9):e066379 [FREE Full text] [doi: 10.1136/bmjopen-2022-066379] [Medline: 36137622]
42. Pieterse AH, Kunneman M, van den Hout WB, Baas-Thijssen M, Geijsen ED, Ceha HM, et al. Patient explicit consideration of tradeoffs in decision making about rectal cancer treatment: benefits for decision process and quality of life. Acta Oncol 2019;58(7):1069-1076 [FREE Full text] [doi: 10.1080/0284186X.2019.1594363] [Medline: 30971150]
43. Pieterse AH, Berkers F, Baas-Thijssen MC, Marijnen CA, Stiggelbout AM. Adaptive conjoint analysis as individual preference assessment tool: feasibility through the internet and reliability of preferences. Patient Educ Couns 2010;78(2):224-233. [doi: 10.1016/j.pec.2009.05.020] [Medline: 19581069]
44. Rochon D, Eberth JM, Fraenkel L, Volk RJ, Whitney SN. Elderly patients' experiences using adaptive conjoint analysis software as a decision aid for osteoarthritis of the knee. Health Expect 2014;17(6):840-851 [FREE Full text] [doi: 10.1111/j.1369-7625.2012.00811.x] [Medline: 22994378]
45. Fraenkel L. Feasibility of using modified adaptive conjoint analysis importance questions. Patient 2010;3(4):209-215 [FREE Full text] [doi: 10.2165/11318820-000000000-00000] [Medline: 20930941]
46. Snaman JM, Helton G, Holder RL, Revette A, Baker JN, Wolfe J. Identification of adolescents and young adults' preferences and priorities for future cancer treatment using a novel decision-making tool. Pediatr Blood Cancer 2021;68(1):e28755. [doi: 10.1002/pbc.28755] [Medline: 33017087]
47. Snaman JM, Helton G, Holder RL, Wittenberg E, Revette A, Tulsky JA, et al. MyPref: pilot study of a novel communication and decision-making tool for adolescents and young adults with advanced cancer. Support Care Cancer 2021;29(6):2983-2992. [doi: 10.1007/s00520-020-05806-1] [Medline: 33030597]
48. Snaman JM, Blazin L, Holder RL, Wolfe J, Baker JN. Identifying and quantifying adolescent and young adult patient preferences in cancer care: development of a conjoint analysis-based decision-making tool. J Adolesc Young Adult Oncol 2019;8(2):212-216. [doi: 10.1089/jayao.2018.0116] [Medline: 30362893]
49. Streufert B, Reed SD, Orlando LA, Taylor DC, Huber JC, Mather RC. Understanding preferences for treatment after hypothetical first-time anterior shoulder dislocation: surveying an online panel utilizing a novel shared decision-making tool. Orthop J Sports Med 2017;5(3):2325967117695788 [FREE Full text] [doi: 10.1177/2325967117695788] [Medline: 28377932]
50. Studts JL, Thurer RJ, Brinker K, Lillie SE, Byrne MM. Brief education and a conjoint valuation survey may reduce decisional conflict regarding lung cancer screening. MDM Policy Pract 2020;5(1):2381468319891452. [doi: 10.1177/2381468319891452] [Medline: 31976372]
51. Byrne MM, Thurer RJ, Studts JL. Individual decision making about lung cancer screening: a conjoint analysis of perspectives among a high-risk national sample. Cancer Med 2019;8(12):5779-5786 [FREE Full text] [doi: 10.1002/cam4.2445] [Medline: 31385463]
52. Wittink MN, Walsh P, Yilmaz S, Mendoza M, Street RL, Chapman BP, et al. Patient priorities and the doorknob phenomenon in primary care: can technology improve disclosure of patient stressors? Patient Educ Couns 2018;101(2):214-220 [FREE Full text] [doi: 10.1016/j.pec.2017.08.004] [Medline: 28844522]
53. Wittink MN, Yilmaz S, Walsh P, Chapman B, Duberstein P. Customized care: an intervention to improve communication and health outcomes in multimorbidity. Contemp Clin Trials Commun 2016;4:214-221 [FREE Full text] [doi: 10.1016/j.conctc.2016.10.002] [Medline: 28191546]
54. Shaoibi A, Neelon B, Lenert LA. Shared decision making: from decision science to data science. Med Decis Making 2020;40(3):254-265 [FREE Full text] [doi: 10.1177/0272989X20903267] [Medline: 32024424]
55. Attribution-NonCommercial-NoDerivatives 4.0 International. Creative Commons. URL: https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en
56. Attribution-NonCommercial 4.0 International. Creative Commons. URL: https://creativecommons.org/licenses/by-nc/4.0/deed.en
57. Attribution 4.0 International. Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/deed.en
58. Martin RW, Brogård Andersen S, O'Brien MA, Bravo P, Hoffmann T, Olling K, et al. Providing balanced information about options in patient decision aids: an update from the international patient decision aid standards. Med Decis Making 2021;41(7):780-800. [doi: 10.1177/0272989X211021397] [Medline: 34196241]

XSL•FO
RenderX

59. Trevena LJ, Bonner C, Okan Y, Peters E, Gaissmaier W, Han PKJ, et al. Current challenges when using numbers in patient decision aids: advanced concepts. Med Decis Making 2021;41(7):834-847 [FREE Full text] [doi: 10.1177/0272989X21996342] [Medline: 33660535]
60. Bonner C, Trevena LJ, Gaissmaier W, Han PKJ, Okan Y, Ozanne E, et al. Current best practice for presenting probabilities in patient decision aids: fundamental principles. Med Decis Making 2021;41(7):821-833 [FREE Full text] [doi: 10.1177/0272989X21996328] [Medline: 33660551]
61. Trenaman L, Jansen J, Blumenthal-Barby J, Körner M, Lally J, Matlock D, et al. Are we improving? Update and critical appraisal of the reporting of decision process and quality measures in trials evaluating patient decision aids. Med Decis Making 2021;41(7):954-959 [FREE Full text] [doi: 10.1177/0272989X211011120] [Medline: 33966534]

Abbreviations

COMET: Core Outcome Measures in Effectiveness Trials
DCE: discrete choice experiment
IPDAS: International Patient Decision Aids Standards
pMNL: penalized multinomial logit model
NASA: National Aeronautics and Space Administration
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by N Cahill; submitted 05.04.24; peer-reviewed by J Mistry, R Zimba; comments to author 15.06.24; revised version received 24.06.24; accepted 21.01.25; published 21.03.25.

Please cite as:
Wickramasekera N, Shackley P, Rowen D
Embedding a Choice Experiment in an Online Decision Aid or Tool: Scoping Review
J Med Internet Res 2025;27:e59209
URL: https://www.jmir.org/2025/1/e59209
doi:10.2196/59209
PMID:

©Nyantara Wickramasekera, Phil Shackley, Donna Rowen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Acceptability and Effectiveness of a Fully Web-Based Nutrition and Exercise Program for Individuals With Chronic Disease During COVID-19: Randomized Controlled Trial

Puneeta Tandon[1*], MD, MSc; Kathleen P Ismond[1], PhD; Graeme Purdy[2], MSc; Christofer Cruz[1], BSc; Evelyn Etruw[2], MCISc, PT; Kirsten Suderman[2], MSc; Ashley Hyde[1], PhD; Michael Stickland[1], PhD; John C Spence[3], PhD; Dale C Lien[1], MD; Rahima Bhanji[1], MD; Carla M Prado[4], PhD; Antonio Miguel-Cruz[5], PhD; Anil A Joy[6], MD; Maryna Yaskina[7], PhD; Margaret L McNeely[2*], PhD

[1]Department of Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

[2]Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[3]Faculty of Kinesiology, Sport, and Recreation, University of Alberta, Edmonton, AB, Canada

[4]Department of Agricultural, Food and Nutritional Science, Faculty of Agricultural, Life and Environmental Sciences, University of Alberta, Edmonton, AB, Canada

[5]Department of Occupational Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

[6]Department of Oncology, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

[7]Women and Childrens' Health Research Institute, University of Alberta, Edmonton, Canada

[*]these authors contributed equally

Corresponding Author:
Puneeta Tandon, MD, MSc
Department of Medicine
Faculty of Medicine & Dentistry
University of Alberta
8540 112 Street
Edmonton, AB, T6G2X8
Canada
Phone: 1 7804929844
Email: ptandon@ualberta.ca

Abstract

Background: In-person nutrition and exercise interventions improve physical function in chronic diseases, yet the acceptability and effectiveness of web-based delivery, especially with different levels of personnel support, require further investigation.

Objective: This study aims to evaluate a web-based nutrition and exercise intervention delivered entirely digitally from recruitment to trial completion.

Methods: A randomized controlled trial was conducted using the Heal-Me version 1 platform across 2 levels of personnel support (Light and Intensive). Eligible adults with a history of cancer, chronic lung disease, or liver or lung transplant; internet access; and prior participation in a rehabilitation program were enrolled in a fully web-based program to minimize barriers to exercise participation. Participants were randomly assigned (1:1:1) to 1 of 3 study groups. The control group received a detailed, self-directed digital nutrition and exercise guide. The Heal-Me Light group received the web-based intervention alongside dietitian and exercise specialist–led group classes. The Heal-Me Intensive group received web-based intervention, group classes, and one-to-one sessions with the dietitians and exercise specialists. All participants received a wearable activity tracker. The primary acceptability outcome was adherence to the intervention based on a priori targets. The primary effectiveness outcome was the change in Lower Extremity Functional Scale (LEFS) score. Secondary outcomes included physical function tests, which were performed and measured by videoconference. Questionnaires were used to assess well-being, quality of life, and food intake. Analyses adhered to the intention-to-treat principle.

Results: Of 216 participants, 202 (93.5%) completed the intervention (mean 61, SD 11 years; female: 130/202, 64.4%; cancer: 126/202, 62.4%). Adherence exceeded a priori targets, with 82% (105/128) attending >75% of the program elements including postintervention tests. Participants rated the program as “quite a bit” or “very” useful, with similar ratings between Heal-Me Light

XSL•FO
RenderX

(56/64, 88%) and Heal-Me Intensive (51/58, 88%) groups (*P*=.69). No significant differences were found for changes in LEFS scores (control: mean 0.8, SD 7.7; Heal-Me: mean 0.3, SD 6.6; *P*=.53). Significant benefits were found in favor of the combined Heal-Me intervention groups versus controls for change in the 2-minute step test, World Health Organization-5 Well-Being Index, Short-Form-36 general, physical health role, energy or fatigue scales, and protein intake. While the change in physical function was similar between the 2 intervention arms, the more intensive one-to-one interaction (Heal-Me Intensive) led to greater improvements in perceived nutrition self-management. No serious adverse events occurred.

Conclusions: The demonstrated satisfaction, adherence, and effectiveness highlight the high acceptability of a web-based, semisupervised nutrition and exercise intervention delivered entirely digitally in individuals with chronic disease. Future studies may benefit from having a baseline physical function inclusion threshold, the use of a more sensitive primary physical function measure, and a higher intensity digital exercise intervention in exercise-experienced participants.

Trial Registration: Clinicaltrials.gov NCT04666558; https://clinicaltrials.gov/study/NCT04666558

International Registered Report Identifier (IRRID): RR2-10.1016/j.cct.2022.106791

(J Med Internet Res 2025;27:e57537) doi:10.2196/57537

KEYWORDS

eHealth; patient-centered care; adults; geriatrics; self-management; web-based; nutrition; exercise rehabilitation; wearable; activity tracker; quality of life; physical health; 2-minute step test; patients with cancer; chronic diseases; COVID-19; randomized controlled trial; acceptability; effectiveness; intervention

Introduction

Impairments in physical function and physical performance are common in individuals living with serious chronic conditions. Although these impairments increase the risk for morbidity and mortality, they are potentially modifiable with nutrition and exercise rehabilitation [1]. During the COVID-19 pandemic, rehabilitation interventions that were traditionally delivered in person were rapidly transitioned to web-based delivery due to lockdowns and social distancing policies. At the pandemic's onset, limited data were available to guide the feasibility, safety, and effectiveness of delivering digital nutrition and exercise therapy, and standardized protocols for conducting web-based evaluation of physical function were lacking. Internet-connected digital platforms, such as web-based apps, while offering promise for the scalable delivery of multimodal supportive care interventions, have been associated with poor adherence and retention [2]. Moreover, little information was available to guide how much personnel support was needed to optimize app usage in older patient populations living with chronic disease.

In 2019, our research team co-designed a web-based app ("Heal-Me") with people living with chronic debilitating diseases, including cancer, cirrhosis, and lung disease [3]. Version 1 of the app was used in this study and is described in more detail in the protocol paper, which includes illustrations of user interface and app features [4]. The Heal-Me web-based app facilitates semisupervised nutrition and exercise programming tailorable to the unique needs of each patient user. It includes nutrition content—a daily protein tracking system, high protein recipes, and cooking videos; exercise content—customizable follow-along exercise videos and autotracking of videos watched; and a calendar for scheduling and accessing web-based group classes and one-to-one check-ins. To guide behavior change, the capability, opportunity, and motivation behavior (COM-B) model and the theoretical domains framework were used as part of the development of the Heal-Me intervention to provide progress tracking, demonstration videos, and graded tasks [4].

With the availability of Heal-Me, this randomized controlled trial (RCT) was designed in response to COVID-19 to answer questions about the acceptability and effectiveness of a personnel-supported, web-based nutrition and exercise intervention for individuals with chronic disease. To understand the extent of support needed by patient users, there were 2 intervention arms (digital group classes only vs digital group classes with additional one-to-one support). We hypothesized that increased contact with a dietitian and exercise professional would result in greater acceptability as measured by higher adherence to the program and greater effectiveness as evaluated by the change in score on the Lower Extremity Function Scale (LEFS).

Methods

Study Design

Personalized Online Nutrition and Exercise Routines (PIONEER) was a parallel-group, 3-armed RCT conducted using the Heal-Me web-based program across 12 weeks [4]. The study ran from December 2020 to December 2021, during the COVID-19 pandemic, when Albertan public health policies restricted movement and social interactions. Thus, all trial stages, from recruitment through to end-of-study testing, were performed in a completely digital environment, including participant technology training and the collection of data on physical function.

Ethical Considerations

The study was approved by the University of Alberta's Health Research Ethics Board (Pro00103715) and was registered with ClinicalTrials.gov (NCT04666558). Participants provide informed written consent. The privacy and security of the web-based intervention were described previously [4]. There was no monetary compensation provided.

Recruitment

Study advertisement and recruitment were carried out by the digital distribution of recruitment materials by frontline staff

XSL•FO
RenderX

involved in oncology (Alberta Cancer Exercise), pulmonary (Breathe Easy), and posttransplant lung or liver rehabilitation programs (University of Alberta Hospitals) in Edmonton, Alberta. Interested participants were required to contact the research team for further information and eligibility screening. Participants were recruited and enrolled in cohorts that started monthly to facilitate group intervention sessions.

Participants

Inclusion criteria included adults (aged 18 years and older) with an internet-connected device in their home and the ability to read, write, and speak English. Individuals were eligible if they had cancer (completed an initial course of chemotherapy or radiotherapy but could be on maintenance therapy); had undergone a lung or liver transplant; or had chronic lung disease. All had to graduate from an exercise rehabilitation program before the COVID-19 public health restrictions. The requirement of being an exercise graduate reflected concerns over the safety and complexity of participants having to learn exercise in an entirely web-based environment on top of new tasks (eg, wearable device, learning to use the Heal-Me app, nutrition tracking, and participating in home-based digital physical function assessments) all during the time of COVID-19. Individuals were ineligible if they were receiving compassionate care with an anticipated survival of less than 6 months, had unstable disease status, or were unable to provide informed electronic consent.

Randomization and Masking

Allocation tables were generated, validated, and uploaded to the REDCap (Research Electronic Data Capture; Vanderbilt University) information system [5] by an independent statistician. Stratification was based on chronic disease status: cancer, lung, or transplant. Randomization was in a 1:1:1 ratio:

- Control (self-directed, digital nutrition and exercise guide)
- Heal-Me Light (Heal-Me app and digital group sessions led by Heal-Me trainers—trainers defined as a registered dietitian and certified exercise specialists including a physiotherapist)
- Heal-Me Intensive (Heal-Me app, digital group sessions, and one-to-one meetings with trainers)

Allocation was concealed from all study staff. The end-of-trial independent assessor was blinded to the participants' group allocation.

Procedures

Prerandomization Procedures

Web-based activities included advertisement; recruitment; obtaining consent; patient screening; and data collection, including demographics, socioeconomic status, health-related information (eg, medications and comorbidities), and other questionnaires. Outlined in more detail in the protocol [4], the questionnaires included LEFS [6], a validated 20-question survey to assess functional changes in the hip, thigh, knee, leg, ankle, and foot; Physical Activity Readiness Questionnaire [7]; technology proficiency with an internet-connected device (either Computer Proficiency Questionnaire [8] or Mobile Device Proficiency Questionnaire [9]); Unified Theory of Acceptance and Use of Technology [10]; and the Upper Extremity Functional Index-15 [11]. Additional questionnaires were the World Health Organization-5 (WHO-5) Well-Being Index, Generalized Anxiety Disorder-7 [12], De Jong Gierveld Loneliness Scale [13], Short-Form Survey 36-items [14], and self-reported physical activity. The 3-day food records (1 weekend day and 2 weekdays) were completed and emailed to the study coordinator, verified for accuracy by a registered dietitian or dietetic intern, and nutritional intake was analyzed using the FoodProcessor software (version 11.7 .1; ESHA Research). Behavioral beliefs were assessed using the 6-item 0-10 Likert scale for the COM-B [3]. Web-based physical function testing, led by a certified exercise specialist, included the 30-second and 60-second chair sit-to-stands; 1-legged balance; 2-minute step test; shoulder flexion range of motion for left and right sides; and optional tests (if the participants were able and willing), including the chair sit-and-reach and plank test. Height and weight measures were recorded based on patient self-report.

Postrandomization Procedures

Participants were randomly assigned and allocated to 1 of 3 study arms after completing the baseline measures. All participants received a wearable activity tracker (Garmin Ltd) to monitor daily step counts, moderate- to vigorous-intensity physical activity minutes, and digital assistance with setup. The self-directed control group received a 52-page manual containing illustrated exercise routines at 3 levels and nutrition guidance (Alberta Health Services and Canada's Food Guide).

Participants in the Heal-Me Light and Heal-Me Intensive interventions had three web-based sessions as detailed in the protocol [4]: (1) an introduction to and instruction in the use of the Heal-Me app; (2) a dietitian-led nutrition assessment, review of nutrition preferences, and the setting of a protein goal target in the app (1.2-1.5 g/kg/day based on guidelines [15-17]); and (3) a certified exercise specialist–led review of home space for exercise, review of exercise preferences, and setting up of an individualized program in the app. While the focus of nutrition counseling was on protein intake, a condition-specific, guideline-based target for daily caloric intake of 25-30 kcal/kg/day was also presented in the app, and counseling was provided about eating a balanced plate [15-17]. Recognizing the limited data supporting the safety of digital exercise, the certified exercise specialist started all participants at a low-intensity level for exercise. Participants were asked to complete 45-60 minutes of exercise sessions 3 times a week (2 independent sessions and 1 certified exercise specialist–led, web-based group class), track protein intake for 3 days each week, and attend 5 web-based group nutrition classes over the 12-week study. Heal-Me Intensive participants also received one-to-one check-ins with the dietitian (2×, 20-30 minutes) and certified exercise specialist (5×, 20-30 minutes) over the 12-week study period.

After completing the 12-week trial, participants repeated web-based physical function assessments, anthropomorphic measures, questionnaires, and 3-day food records and completed the study satisfaction survey.

XSL•FO
RenderX

Study Objectives

Acceptability

This was comprised of adherence rate and satisfaction. Our primary acceptability outcome was adherence, which was defined as completing more than 75% of the group nutrition and exercise sessions and completing all end-of-study outcomes. Satisfaction was collected from the end of the study program experience and satisfaction questionnaire. Two questions inquired about whether intervention participants felt prepared to continue their nutrition and exercise activities after program completion. Using a 5-item Likert scale (from 1=not at all to 5=very much), participants were asked "Do you feel that the PIONEER nutrition program has prepared you to reach your target protein goal on your own?" and "Do you feel that the PIONEER exercise program has prepared you to reach your exercise goal on your own?" Technology acceptability was evaluated with the Unified Theory of Acceptance and Use of Technology and will be presented in a follow-on publication.

Effectiveness

This was comprised of a battery of objective outcomes including physical function tests and physical activity, and self-reported outcomes including quality of life. Our primary effectiveness outcome was the change in LEFS score from baseline to 12 weeks between the control and combined intervention arms. Secondary objective effectiveness outcomes included the 12-week change in objective physical function comparing the control and intervention groups for the 2-minute step test, chair sit-to-stand (30 and 60 seconds), plank, shoulder range of movement, and 1-legged balance. Step count and moderate- to vigorous-intensity physical activity minutes for weeks 2 to 11 were averaged for each week. Data from weeks 1 and 12 were not considered because postal delivery and setup of the device often extended into the first week, and device return occurred at the end of week 12. Secondary self-reported effectiveness outcomes included Upper Extremity Functional Index-15, mental health, and quality of life (ie, WHO-5 Well-Being Index, Generalized Anxiety Disorder-7, De Jong Gierveld Loneliness Scale, and Short-Form-36). Adverse events were recorded by study staff.

Exploratory Outcomes (Baseline and 12 Weeks)

Changes in daily calorie, protein intake, and the proportion of participants reaching target protein intake (>1.2 g/kg/day protein based on ideal body weight using a BMI of 24.9 kg/m^2) were evaluated. The COM-B scale assessed changes in behavioral beliefs, with data planned for presentation in a future publication. For physical activity, the validated Single Item Physical Activity question was used, where participants were asked "In the past week, on how many days have you done a total of 30 minutes or more of physical activity, which was enough to raise your breathing rate?" [18]. Additionally, we explored results comparing the highest and lowest quartiles of baseline function for the 2-minute step test.

Sample Size

The sample size and statistical analysis plans are presented in detail in the protocol [4]. The primary analysis used the ANOVA model to assess the difference in the LEFS scores between the 3 study groups. To preserve type 1 error in 3 between-group comparisons, the α was adjusted by the Bonferroni correction. Thus, with a 2-sided α=.05 (.0167 adjusted for multiple comparisons) and β=.20 (80% power), 60 participants per group (180 in total) were estimated to be required to determine a statistically significant difference for a moderate (0.6) Cohen *d* effect size (minimal clinically important difference of 9 points [19]). To account for a conservative 20% dropout rate, the total sample size was increased to 216 (n=72 per arm).

Statistical Analysis Plan

The intention-to-treat principle was used for all analyses. Descriptive statistics were used to understand the participants' characteristics and outcome variables. The between-groups comparison was analyzed using ANOVA (or its nonparametric alternative) and the Tukey honest significant difference test for pairwise comparisons using a familywise α of .05. The primary analysis, the LEFS score, was then analyzed by linear mixed models with random effects, which were adjusted for the baseline score (T1) as a covariate [20,21]. A 3-group comparison was planned only if the combined 2-group comparison was significant. Exercise adherence rates (proportion completing 80% of exercise sessions over the study) and dietary protein adherence rates (proportion meeting 80% of target intake over the study) were compared between study groups by the chi-square test. The differences in the frequency of adverse events were assessed using the chi-square test. Linear regressions were constructed to determine clinically significant outcomes, adjusted for sex, age, and comorbidity status; group assignment was an independent variable. Missing data resulted in listwise deletion of the cases. Analyses were performed using either SAS (version 9.4 or later; SAS Institute) or SPSS (version 28 or later; IBM Corp).

Results

Participant Characteristics

The 202 participants (Figure 1) were recruited over 10 months, and none were screened out for unstable disease. They had a mean age of 61.0 (SD 11.1) years and 64.4% (130/202) were female. At baseline, 83.2% (168/202) attended postsecondary education, and 44.5% (90/202) earned more than CAD $80,000 (US $ conversion rate for 2023: US $1=CAD $1.3497) per annum (Table 1). The most common chronic condition was cancer (126/202, 62.4%), while the least common was liver transplantation (16/202, 7.9%). The level of comorbidity was high, with a mean Charlson Comorbidity Index of 4.6 (SD 2.0), and the mean BMI was 28.4 (SD 6.2) kg/m^2. The mean technology proficiency measured with the Computer Proficiency Questionnaire and Mobile Device Proficiency Questionnaire was 88.7% (SD 14.9%).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) patient flowchart.

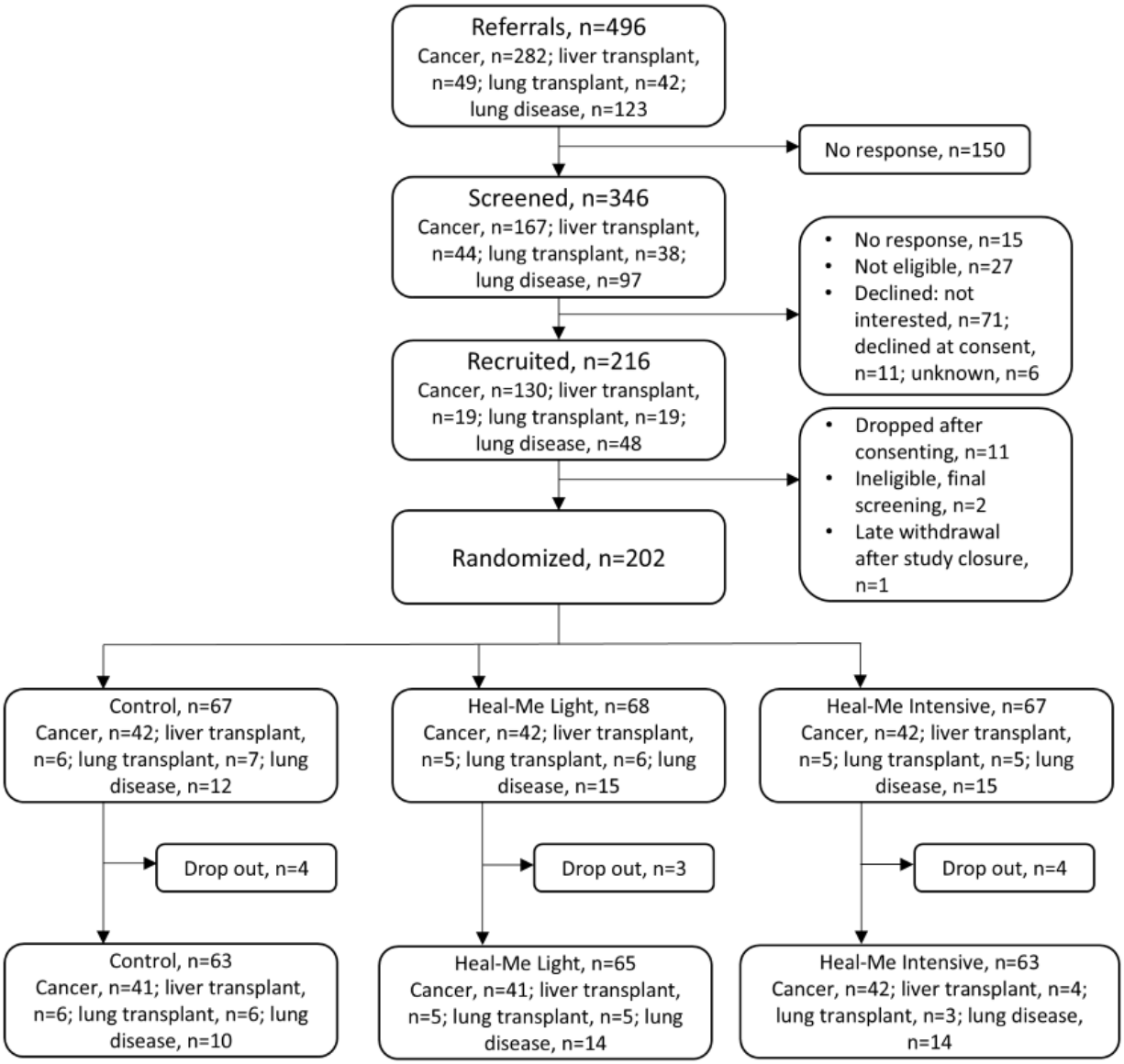

XSL•FO
RenderX

Table 1. Participant characteristics at baseline.

Characteristic	Overall (n=202)	Control (n=67)	Heal-Me Light (n=68)	Heal-Me Intensive (n=67)
Age (years), mean (SD)	61.0 (11.1)	61.2 (9.4)	61.4 (11.7)	60.4 (12.1)
Sex, n (%)				
Female	130 (64.4)	44 (65.7)	47 (69.1)	39 (58.2)
Male	72 (35.6)	23 (34.3)	21 (30.9)	28 (41.8)
Education, n (%)				
Some or completed high school	34 (16.8)	11 (16.4)	14 (20.6)	9 (13.4)
Some university or college	33 (16.3)	6 (8.9)	14 (20.6)	13 (19.4)
Completed university or college	95 (47.0)	34 (50.7)	29 (42.6)	32 (47.8)
Some or completed graduate school	40 (19.8)	16 (23.9)	11 (16.2)	11 (16.2)
Annual income (CAD $)[a], n (%)				
Less than 20,000	9 (4.4)	1 (1.5)	6 (8.8)	2 (3)
Between 20,000 and 39,000	30 (14.8)	11 (16.4)	10 (14.7)	9 (13.4)
Between 40,000 and 59,000	28 (13.9)	14 (20.9)	8 (11.8)	6 (9)
Between 60,000 and 79,999	26 (12.9)	9 (13.4)	7 (10.3)	10 (14.9)
Between 80,000 and 99,999	31 (15.3)	9 (13.4)	13 (19.1)	9 (13.4)
More than 100,000	59 (29.2)	21 (31.3)	19 (27.9)	19 (28.3)
Not reported	19 (9.4)	2 (3)	5 (7.3)	12 (17.9)
Ethnicity, n (%)				
Arab	1 (0.5)	1 (1.5)	0 (0)	0 (0)
Asian	12 (5.9)	4 (6)	3 (4.4)	5 (7.5)
Indigenous or First Nations	3 (1.5)	1 (1.5)	2 (2.9)	0 (0)
Latin or Central or South American	4 (2)	1 (1.5)	1 (1.5)	2 (3)
White	174 (86.1)	57 (85.1)	60 (88.2)	57 (85.1)
Not reported	8 (4)	3 (4.5)	2 (2.9)	3 (4.5)
Marital status, n (%)				
Never married	24 (11.9)	8 (4.0)	9 (4.5)	7 (3.5)
Married or common law	136 (67.3)	47 (40.3)	41 (60.3)	48 (71.6)
Divorced or separated	28 (13.9)	8 (4)	15 (22)	5 (2.5)
Widowed	14 (6.9)	4 (2)	3 (1.5)	7 (3.5)
Disease group, n (%)				
Cancer	126 (62.4)	42 (62.7)	42 (61.8)	42 (62.7)
Lung disease	42 (20.8)	12 (17.9)	15 (22.1)	15 (22.4)
Lung transplant recipient	18 (8.9)	7 (10.4)	6 (8.8)	5 (7.5)
Liver transplant recipient	16 (7.9)	6 (8.9)	5 (7.3)	5 (7.5)
Charlson Comorbidity Index, mean (SD)	4.6 (2)	4.8 (2.1)	4.5 (1.9)	4.5 (2.2)
Current smokers, n (%)	5 (2.5)	3 (4.5)	1 (1.5)	1 (1.5)
BMI baseline (kg/m^2; n=198), mean (SD)	28.4 (6.2)	28.7 (6.8)	28.1 (5.5)	28.4 (6.2)
Hemoglobin (g/L; n=191), mean (SD)	133.7 (16.7)	132.6 (18.9)	133.2 (15)	135.4 (16)
Physical Activity (minutes/week) Single question self-report (n=200), mean (SD)	101.7 (86)	105.5 (66.6)	85.6 (62)	114.5 (117.4)

[a]US $ conversion rate for 2023: US $1= CAD $1.3497.

XSL•FO
RenderX

Study Objectives

Acceptability

The completion rate for the study was 94.6% (191/202), and the mean attendance for arms 2 (n=66) and 3 (n=62) combined was 78.6% for the 5 group nutrition classes, 88.5% for the 12 group exercise sessions, and 78.8% for the home exercise videos, with no statistically significant differences (*P*=.84, *P*=.64, and *P*=.67, respectively) between the 2 intervention arms. In response to the statement "Completing the PIONEER program helped me to meet my health and wellness goals," 82.8% (101/122) of participants in Heal-Me Light and 90.2% (110/122) of participants in Heal-Me Intensive answered "quite a bit" to "very much" (2 of the 5 highest categories), with no significant difference between arms (*U*=1920; *z*=0.41; *P*=.69). Although staff services were scored as "excellent" (97/122, 79.5%), participants in Heal-Me Intensive scored staff services statistically significantly higher than participants in Heal-Me Light (*U*=2209; *z*=2.57; *P*=.01). At the end of the study, when participants were asked if PIONEER had prepared them to reach their exercise and protein goals after program completion; for exercise, 81.1% (99/122) responded either "quite a bit" or "very much," and for nutrition, 82.6% (100/121) responded either "quite a bit" or "very much." Those in Heal-Me Light (mean rank 53.6) reported significantly less preparedness to reach their daily protein intake than those in Heal-Me Intensive (mean rank 69.3; *U*=2298; *z*=2.66; *P*=.008). No significant difference between interventional groups was found regarding achieving exercise goals (59.6 vs 63.7; *P*=.495). Overall, participants reported that their program was "quite a bit" or "very" useful (top 2 out of 5 categories) for Heal-Me Light (56/64, 88%) and Heal-Me Intensive (51/58, 88%; *U*=1920; *z*=0.41; *P*=.69). Most participants would recommend the program either "quite a bit" or "very much" (top 2 out of 5 categories) to others living with a chronic disease (Heal-Me Light: 59/64, 92% and Heal-Me Intensive: 57/58, 98.3%; *U*=1843; *z*=–1.00; *P*=.92). More generally, when asked about the perceived burden of doing the web-based physical function assessments, 73.8% (90/122) reported "no burden at all."

Effectiveness

Primary Effectiveness Outcome: LEFS

The baseline LEFS scores for the control and Heal-Me groups were 63.6 (SD 13.9) and 62.9 (SD 14.5), respectively. The difference of –0.53 (95% CI –2.7 to 1.7) was not statistically significant between groups for self-reported physical function, even when analyzed by the disease group (Table 2 and Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 2. Primary and secondary fitness outcomes.

Measure	T1 (baseline), mean (SD)	T2 (end-of-study), mean (SD)	Change from T2 to T1, mean (SD)	Unadjusted between-group differences: T2 to T1, mean (95% CI)	Adjusted[a] between-group differences: T2 to T1, mean (95% CI)
LEFS[b] score, (n=177)				0.53 (–1.7 to 2.7)	0.60 (–1.6 to 2.8)
Control	63.6 (13.9)	64.4 (14.3)	0.8 (7.7)		
Heal-Me app	62.9 (14.5)	63.5 (14.5)	0.3 (6.6)		
2-minute step test (n=171)				10.0 (4.6 to 15.4)[c]	9.8 (4.7 to 15.0)[c]
Control	80.5 (23.9)	81.9 (23.1)	1.3 (15.5)		
Heal-Me app	79.6 (21.0)	91.0 (22.3)	11.3 (17.4)		
Chair sit-to-stand 60 seconds (n=171)				0.43 (–1.1 to 1.9)	0.22 (–1.3 to 1.7)
Control	29.7 (11.3)	30.3 (10.9)	0.6 (4.0)		
Heal Me app	27.5 (9.6)	28.5 (9.8)	1.0 (4.9)		
Chair sit-to-stand 30 seconds (n=171)				0.16 (–0.67 to 0.98)	0.02 (–0.78 to 0.83)
Control	15.6 (5.8)	15.8 (5.5)	0.16 (2.1)		
Heal-Me app	14.6 (4.9)	15.0 (4.8)	0.32 (2.7)		
Sit-and-reach (cm reached, n=61)				–1.53 (–4.7 to 1.6)	–1.6 (–4.8 to 1.6)
Control	–4.1 (12.9)	–0.9 (13.2)	3.2 (4.3)		
Heal-Me app	–3.6 (12.6)	–2.0 (13.1)	1.7 (6.2)		
Shoulder range of motion right (degrees, n=160)				–1.4 (–4.0 to 1.2)	–0.5 (–2.9 to 1.9)
Control	150 (8.1)	152 (10.3)	2.0 (8.5)		
Heal-Me app	153 (12.4)	153 (11.1)	0.6 (7.5)		
Shoulder range of motion left (degrees, n=160)				–0.05 (–2.5 to 2.5)	0.4 (–1.9 to 2.8)
Control	149 (11.0)	150 (12.4)	0.87 (6.5)		
Heal-Me app	151 (12.6)	152 (11.5)	0.82 (8.0)		
Balance right (seconds, n=172)				2.7 (–0.6 to 6.1)	3.0 (–0.6 to 6.1)
Control	29.2 (18.0)	29.2 (17.5)	0.0 (9.8)		
Heal-Me app	29.8 (15.7)	32.6 (15.2)	2.8 (10.6)		
Balance left (seconds, n=172)				1.2 (–2.4 to 4.6)	1.5 (–1.7 to 4.8)
Control	29.6 (16.4)	30.7 (16.2)	1.1 (11.3)		
Heal-Me app	31.2 (15.6)	33.4 (15.0)	2.2 (10.6)		
Plank (seconds, n=134)				8.1 (–3.5 to 19.6)	7.7 (–4.0 to 19.4)
Control	63.4 (42.4)	75.3 (49.9)	11.9 (22.5)		
Heal-Me app	67.2 (56.3)	87.2 (69.7)	20.0 (35.1)		

[a]Adjusted for age, Charlson Comorbidity score, and baseline value.

[b]LEFS: Lower Extremity Functional Scale.

[c]*P*<.001 between-group improvement from T2 to T1.

Secondary Objective Outcomes

A statistically significant, between-group difference was found in favor of the intervention for the 2-minute step test (*P*<.001). Details on the web-based assessed changes in physical function outcomes are presented in Table 2 and Table S1 in Multimedia Appendix 1. The intervention and control groups had similar step counts and moderate- to vigorous-intensity physical activity minutes as measured by the wearable activity tracker (Figures S1 and S2 in Multimedia Appendix 1).

Secondary Self-Reported Outcomes

Changes in the WHO-5 Well-Being Scale (*P*=.02) and the Short-Form-36 general health (*P*=.04) were statistically significantly different between the control and combined intervention arms according to 2-tailed independent *t* tests (Table 3; Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 3. Participant-reported outcomes between control and Heal-Me app (Light and Intensive).

Measure	T1 (baseline), mean (SD)	T2 (end-of-study), mean (SD)	Change from T2 to T1, mean (SD)	Unadjusted between-group differences: T2 to T1, mean (95% CI)	Adjusted[a] between-group differences: T2 to T1, mean (95% CI)
Upper Extremity Functional Index (n=187)				1.72 (0.01-3.4)[b]	1.6 (–0.07 to 3.2)
Control	72.1 (11.3)	71.1 (11.0)	–0.95 (5.7)		
Heal-Me app	71.7 (9.5)	72.4 (9.7)	0.80 (5.5)		
WHO-5[c] Well-Being Scale (n=190)				5.5 (1.3-9.6)[d]	4.94 (0.8-9.1)[e]
Control	63.5 (16.7)	62.6 (18.2)	–0.97 (12.3)		
Heal-Me app	62.0 (21.8)	66.5 (20.9)	4.53 (15.5)		
SF-36[f] general (n=190)				3.83 (0.13-7.53)[g]	3.74 (0.13-7.35)[g]
Control	58.5 (17.4)	58.3 (19.7)	–0.24 (12.3)		
Heal-Me app	58.1 (20.4)	61.6 (20.5)	3.59 (12.1)		
SF-36 physical functioning				0.09 (–3.9 to 4.11)	0.27 (–3.6 to 4.2)
Control	72.4 (23.4)	73.4 (22.0)	0.97 (13.4)		
Heal-Me app	70.6 (22.9)	71.7 (23.8)	1.05 (13.1)		
SF-36 physical health role				16.9 (4.6 to 20.3)[h]	16.3 (5.3 to 27.4)[i]
Control	56.0 (42.1)	49.6 (44.9)	–6.4 (43.1)		
Heal-Me app	54.9 (40.2)	65.4 (40.8)	10.5 (39.2)		
SF-36 emotional health role				10.4 (–1.6 to 22.3)	9.7 (–0.04 to 13.4)
Control	65.6 (38.1)	67.2 (37.9)	1.6 (38.8)		
Heal-Me app	65.1 (36.7)	77.1 (33.4)	11.9 (39.6)		
SF-36 energy or fatigue				4.63 (0.29 to 9.0)[j]	4.81 (0.72 to 8.90)[k]
Control	55.3 (21.1)	54.7 (19.9)	–0.64 (13.8)		
Heal-Me app	56.9 (20.8)	60.9 (22.0)	3.98 (14.4)		
SF-36 emotional well-being				3.32 (–0.34 to 7.0)	2.44 (–0.86 to 5.7)
Control	77.1 (14.5)	76.8 (15.2)	–0.26 (11.6)		
Heal-Me app	74.9 (158)	78.0 (14.4)	3.06 (12.2)		
SF-36 social functioning				5.0 (–1.0 to 11.0)	2.5 (–2.3 to 7.3)
Control	80.2 (21.2)	82.5 (18.5)	2.2 (19.3)		
Heal-Me app	75.6 (23.9)	82.8 (20.1)	7.2 (19.9)		
SF-36 pain				0.73 (–4.9 to 6.3)	0.02 (–5.1 to 5.1)
Control	73.5 (22.1)	72.0 (21.6)	–1.5 (17.3)		
Heal-Me app	71.8 (19.9)	71.0 (20.4)	–0.8 (18.9)		
GAD-7[l] (anxiety, n=190)				–0.31 (–1.46 to 0.85)	–0.64 (–1.4 to 0.12)
Control	3.4 (3.5)	3.6 (3.6)	0.16 (2.8)		
Heal-Me app	3.8 (3.9)	3.2 (3.9)	–0.56 (2.6)		
Loneliness Scale (n=190)				–0.07 (–0.61 to 0.47)	–0.05 (–0.4 to 0.3)
Control	2.2 (1.8)	2.4 (1.8)	0.18 (1.2)		
Heal-Me app	2.3 (1.8)	2.3 (1.7)	0.09 (1.4)		

[a]Adjusted for age, Charlson comorbidity score, and baseline value.

[b]P=.049.

[c]WHO-5: World Health Organization-5.

[d]P=.009.

XSL•FO
RenderX

[e]*P*=.02.
[f]SF-36: Short-Form-36.
[g]*P*=.04.
[h]*P*=.007.
[i]*P*=.004.
[j]*P*=.04.
[k]*P*=.02.
[l]GAD-7: Generalized Anxiety Disorder-7.

Adverse Events

No serious adverse events occurred during the trial. However, 33 minor health or musculoskeletal complaints (eg, muscle pain) were reported among participants in the 3 study groups (control: 10/33, 30.3%; Heal-Me Light: 11/33, 30.3; Heal-Me Intensive: 12/33, 39.4%; *P*=.71). A total of 21 (63.6%) adverse events were reported as "clearly unrelated," 5 (15.1%) as "possibly related," and only 5 (15.1%) as "related" to study participation.

Exploratory Outcomes

A total of 144 (71.2%) participants had a 3-day food record completed at both baseline and the end of study. Analysis of the food records showed that protein and calorie intakes increased in the Heal-Me intervention arms (Heal-Me Light and Intensive) compared to controls (Table S3 in Multimedia Appendix 1). After adjusting for age and comorbidity, the between-group improvements from baseline were presented as mean differences (95% CI) for calories, 246 kcal/kg/d (95% CI 57-436), and protein, 22.8 g/kg/d (95% CI 12.5-33.1). The proportion of participants at goal protein intake (1.2 g/kg/d) at baseline was 53% (23/43) in the control group and 37.6% (38/101) in Heal-Me. By week 12, a total of 53.5% (23/43) of the control group and 81.2% (82/101) of participants in Heal-Me had reached the target protein intake (*P*<.001). Similar proportions of participants in Heal-Me reached the target protein intake (*P*=.53).

In our exploratory analyses of the lowest and highest quartiles of the 2-minute step test, a significant difference was found in favor of Heal-Me Light and Intensive (Figure S3 in Multimedia Appendix 1). For the lowest quartile of participants, the mean difference between the control and intervention groups for the 2-minute step test was 26.3 steps (95% CI 2.24-50.4; *P*=.03) favoring the intervention, as compared to 5.8 steps (95% CI –0.88 to 12.37; *P*=.09) in the highest quartile (Figures S1 and S2 in Multimedia Appendix 1).

Discussion

Principal Findings

To the best of our knowledge, this is the largest, 3-armed, semisupervised nutrition and exercise RCT that has been completed entirely on the web from recruitment to end-of-trial assessments. Acceptability was high across both adherence and satisfaction measures, but the primary effectiveness outcome of self-reported lower extremity function was not significantly different between groups. Statistically significant improvements were found in the Heal-Me intervention for the 2-minute step test, well-being, quality of life, and the proportion of participants reaching target protein intake. Notably, the trial was conducted at a unique time point in our history, against a background of COVID-19 community lockdowns and its associated uncertainties, as well as at the beginning of a now increasing number of studies [22] that are leveraging app-based technologies to support nutrition and exercise interventions.

Acceptability, a primary objective of this study, was met with high rates of adherence and satisfaction. The adherence rate of 82% (105/128) exceeded our a priori level of 75% and is particularly noteworthy given the complexity of the intervention in combination with its fully web-based delivery and evaluation. An in-person prehabilitation intervention of similar complexity (exercise, nutrition, smoking cessation, and psychological support), albeit only 4 weeks in duration, reported that 77% of their 251 participants completed at least 75% of their exercise sessions [23]. In contrast, fully digital prehabilitation interventions have met with lower rates of adherence. In a 4-week digital prehabilitation nutrition and exercise app for colorectal cancer, only 56% (127/227) of participants met the adherence metric of being accessed at least once during the study [24]. In another app-based study of liver transplant candidates (n=31; mean 61, SD 7 years), all received weekly telephone support from a remote physical activity coach during a 12-week physical fitness app intervention [25]. Despite the one-to-one engagement with a trainer, 57% (18/31) of the participants were considered adherent. Consistent with these adherence challenges, a review of supervised and unsupervised mobile health interventions reported an overall adherence of only 56% (55/99), 49% (49/99) for nutrition apps, and 55% (54/99) for physical activity apps [2].

The reasons for the high adherence rates in this study are likely multifaceted. The concept of Heal-Me began with a patient focus group followed by regular interactions with patients and providers, usability or user experience testing, and a pilot study in multiple myeloma, all of which helped shape the app's design, content, and program delivery [3,26]. While there is limited evidence describing the impact of co-design on experimental outcomes, the acceptability and adherence of the Heal-Me app are at least partially attributable to patient and provider involvement across all stages. Moreover, the app included many of the features inventoried in 2 recent systematic reviews for promoting adherence and satisfaction, including ease of use, access to health care information, real-time supervision and monitoring, as well as personalization and automated reminders for calendar events [2,27].

The effectiveness of the intervention was our second objective. The primary outcome of LEFS score was nonsignificant when compared to the control group. Although this may represent a

XSL•FO
RenderX

true lack of effectiveness of multimodal intervention, the lack of significance may also be due to the choice of the LEFS as the primary physical function outcome measure. As a web-based tool, it was selected because of its wide use in populations with musculoskeletal conditions [28] and its association with quality of life, the 6-minute walk test, and clinical outcomes including hospital admissions and falls [4]. As such, during COVID-19, it provided a suitable option over more commonly administered, in-person physical function tests. The normative score of the LEFS is 63 for both male and female individuals aged 60-64 years, with a maximum value of 80 and a minimal clinically important difference of 9 points [28]. In this study, the LEFS score at baseline was already at a mean level of 63 (SD 14.1; n=202), potentially creating a ceiling effect that may have impacted the potential for improvement [23]. Objective measures of physical function are seen as more sensitive indicators of functional abilities [29] and were assessed at baseline and the end of study digitally as secondary outcome measures. A significant between-group difference was found for the 2-minute step test. Of note, unlike the LEFS, this test may have been more sensitive to change as baseline values were more compromised, in the range of the 25th-50th percentile of adults aged 60 to 64 years. In our exploratory analysis of the 2-minute step test, a significant between-group difference was found when comparing results from the lowest and highest quartiles, suggesting primary benefit for those with lower physical functioning, and pointing to future rationale for stratifying inclusion by baseline physical function.

Compared to the control group, the combined intervention group showed statistically significant benefits in some of the quality-of-life measures, which remained when adjusting for age, comorbidity, and baseline values. While the overall sample was comprised of individuals motivated to participate in a nutrition and exercise–supported, web-based program, the average step count, physical activity minutes, and protein intake across groups still fell below the levels recommended by current Canadian public health guidelines [15-17,30].

The study included 2 intervention arms for which we hypothesized that given sufficient statistical power, increased contact with a dietitian and exercise professional would demonstrate greater acceptability and greater effectiveness. The group and one-to-one sessions required a weekly time commitment of approximately 80 minutes in Heal-Me Light and 100 minutes in Heal-Me Intensive in addition to the nutrition tracking and independent exercise. Despite lengthy time commitments, there were similarly high rates of satisfaction and adherence in both groups. This suggests that access to the app and group sessions may have been sufficient to drive acceptability. Unfortunately, the limited differences in the effectiveness outcomes between the intervention and control groups made it impossible to assess the impact of the different levels of personnel support on these measures. While regular interactions with staff and progress monitoring have been previously identified as beneficial facilitators in studies, including those for cardiac telerehabilitation [31] and upper gastrointestinal cancer surgery [32], future studies are encouraged to explore the level of personnel interaction that maximizes participant acceptability and effectiveness.

While adequately powered, the short duration of the study in combination with the heterogeneity across multiple chronic diseases of participants may have contributed to the lack of difference in mean values between the study groups. Furthermore, while biases in self-reporting may have resulted due to social desirability between the participant and staff members, the digital nature of the survey tools was used to minimize bias [33].

This RCT adds to accumulating evidence showing that web-based physical function measures are feasible to conduct. While some data are available to suggest minimal differences between in-person and web-based measures, their validity remains uncertain. In 2023, Heslop et al [34] reviewed 17 studies comparing digital to in-person physical performance test results in older adults. In the 9 studies that assessed digital compared to in-person results, accuracy was reported as “good” (≤5% difference) in 6 studies, “moderate” (5%-10% mean difference) in 2 studies, and “poor” (>10% mean difference) in 1 study. Studies by Guidarelli et al [35] (n=176) and Hoenemeyer et al [36] (n=112) found moderate to strong agreement between remotely evaluated tests with in-person testing. While participants in our study reported little burden with the web-based testing (10% found it “somewhat” or “quite a bit” burdensome), additional certified exercise specialist staff were required to support digital exercise classes for safety. For the web-based physical function tests, two research team members were needed: one to manage the technology (eg, camera view, sound) and lead the test, and the other to score the test.

Limitations

A key limitation of the study was participant selection. Although carried out with sound rationale, we limited enrollment to participants who had graduated from an exercise-based rehabilitation program for their respective chronic diseases. The history of previous exercise participation is associated with self-effectiveness, competence, and adherence to exercise interventions [37,38]. The step count and moderate- to vigorous-intensity physical activity minutes were similar for the intervention and control groups, suggesting a similar degree of physical activity in the control group [39]. Potential contributors that facilitated the control arm exercise were baseline experience with exercise, the provision of an extensive 52-page nutrition and exercise handout, and a wearable activity tracker; the latter are known to increase step count outcomes in adults over the short term [40]. Despite this contamination, improvements were seen in favor of the Heal-Me–supported programming for self-reported health and quality-of-life measures.

Regarding the nutrition component, while significant improvements were seen in target protein intake, verification of the food records was a time-consuming task and frequently involved patient contact for clarification or additional information. Also, challenges were noted with participants submitting their 3-day food records, with only 71.2% with complete records both at baseline and the end of study, suggesting the need to explore additional forms of tracking in future studies [41]. In addition, a completely web-based exercise testing and intervention delivery model was a new experience

XSL•FO
RenderX

for our team. Our digital exercise programs, in retrospect, erred on the side of safety regarding intensity and duration, and for this group of experienced participants, may have been too conservative. Given the multiple secondary outcomes, we acknowledge the potential for type 1 errors. Finally, because of the limited differences seen in outcomes between the intervention and control groups, we were unable to assess the impact of the different levels of personnel support, a factor of interest in future studies.

The Heal-Me PIONEER acceptability and effectiveness trial offers many learnings. As a large trial completed fully digitally during a unique period in our history, it demonstrates the feasibility of rapid recruitment, web-based testing and intervention, and high acceptability and adherence rates among a mixed population with chronic conditions. A comprehensive battery of tests was performed, providing reference points for future studies using the same measures. We also offer considerations for modifications based on our learnings, including stratification for physical function; the selection of a more sensitive objective physical function measure; and the provision of a more tailored, higher-intensity, web-based exercise intervention. In this burgeoning age of digital health, characterized by an increasing reliance on digital assessments and rehabilitative interventions, our findings underscore the feasibility of widespread implementation of nutrition and exercise programming, alongside high participant acceptability and adherence.

Acknowledgments

Funding for this study was provided by a COVID-19 grant from Alberta Innovates Health Solutions. The authors are grateful to all patients who participated in the trial. Study design, all aspects of data management (collection, analysis, and interpretation), and manuscript writing were completed independently of the funding agencies. The co-senior authors (MLM and PT) had full access to the data and were responsible for the manuscript content and the submission decision.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to data sharing policies but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization study funding was acquired by MLM and PT. All authors contributed to data collection. The formal analysis was completed by KPI, PT, MLM, and MY. The original draft was co-written by KPI, MLM, and PT. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting materials.
[DOCX File , 403 KB - jmir_v27i1e57537_app1.docx]

Multimedia Appendix 2
CONSORT-eHealth (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 1210 KB - jmir_v27i1e57537_app2.pdf]

References

1. Dent E, Morley J, Cruz-Jentoft A, Woodhouse L, Rodríguez-Mañas L, Fried L, et al. Physical Frailty: ICFSR International Clinical Practice Guidelines for Identification and Management. J Nutr Health Aging 2019;23(9):771-787 [FREE Full text] [doi: 10.1007/s12603-019-1273-z] [Medline: 31641726]
2. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
3. Ismond KP, Cruz C, Limon-Miro AT, Low G, Prado CM, Spence JC, et al. An open label feasibility study of a nutrition and exercise app-based solution in cirrhosis. Can Liver J 2024;7(1):5-15 [FREE Full text] [doi: 10.3138/canlivj-2023-0011] [Medline: 38505789]
4. Tandon P, Purdy G, Ismond KP, Cruz C, Etruw E, Suderman K, et al. Heal-me PiONEer (personalized online nutrition and exercise): an RCT assessing 2 levels of app-based programming in individuals with chronic disease. Contemp Clin Trials 2022;118:106791. [doi: 10.1016/j.cct.2022.106791] [Medline: 35569753]

XSL•FO
RenderX

5. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
6. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999;79(4):371-383. [Medline: 10201543]
7. Bredin SS, Gledhill N, Jamnik VK, Warburton DE. PAR-Q+ and ePARmed-X+: new risk stratification and physical activity clearance strategy for physicians and patients alike. Can Fam Physician 2013;59(3):273-277 [FREE Full text] [Medline: 23486800]
8. Boot W, Charness N, Czaja S, Sharit J, Rogers WA, Fisk AD, et al. Computer proficiency questionnaire: assessing low and high computer proficient seniors. Gerontologist 2015;55(3):404-411 [FREE Full text] [doi: 10.1093/geront/gnt117] [Medline: 24107443]
9. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. J Appl Gerontol 2018;37(2):131-156 [FREE Full text] [doi: 10.1177/0733464816642582] [Medline: 27255686]
10. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
11. Chesworth BM, Hamilton CB, Walton DM, Benoit M, Blake TA, Bredy H, et al. Reliability and validity of two versions of the upper extremity functional index. Physiother Can 2014;66(3):243-253 [FREE Full text] [doi: 10.3138/ptc.2013-45] [Medline: 25125777]
12. Ruiz MA, Zamorano E, García-Campayo J, Pardo A, Freire O, Rejas J. Validity of the GAD-7 scale as an outcome measure of disability in patients with Generalized Anxiety Disorders in primary care. J Affect Disord 2011;128(3):277-286. [doi: 10.1016/j.jad.2010.07.010] [Medline: 20692043]
13. de Jong Gierveld J, van Tilburg T. The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. Eur J Ageing 2010;7(2):121-130 [FREE Full text] [doi: 10.1007/s10433-010-0144-6] [Medline: 20730083]
14. Lins L, Carvalho FM. SF-36 total score as a single measure of health-related quality of life: scoping review. SAGE Open Med 2016;4:2050312116671725 [FREE Full text] [doi: 10.1177/2050312116671725] [Medline: 27757230]
15. Collins PF, Yang IA, Chang Y, Vaughan A. Nutritional support in chronic obstructive pulmonary disease (COPD): an evidence update. J Thorac Dis 2019;11:S2230-S2237 [FREE Full text] [doi: 10.21037/jtd.2019.10.41] [Medline: 31737350]
16. Arends J, Bachmann P, Baracos V, Barthelemy N, Bertz H, Bozzetti F, et al. ESPEN guidelines on nutrition in cancer patients. Clin Nutr 2017;36(1):11-48. [doi: 10.1016/j.clnu.2016.07.015] [Medline: 27637832]
17. Plauth M, Bernal W, Dasarathy S, Merli M, Plank LD, Schütz T, et al. ESPEN guideline on clinical nutrition in liver disease. Clin Nutr 2019;38(2):485-521 [FREE Full text] [doi: 10.1016/j.clnu.2018.12.022] [Medline: 30712783]
18. Milton K, Clemes S, Bull F. Can a single question provide an accurate measure of physical activity? Br J Sports Med 2013;47(1):44-48. [doi: 10.1136/bjsports-2011-090899] [Medline: 22522584]
19. Verheijde JL, White F, Tompkins J, Dahl P, Hentz JG, Lebec MT, et al. Reliability, validity, and sensitivity to change of the lower extremity functional scale in individuals affected by stroke. PM R 2013;5(12):1019-1025. [doi: 10.1016/j.pmrj.2013.07.001] [Medline: 23876934]
20. Wang B, Ogburn EL, Rosenblum M. Analysis of covariance in randomized trials: more precision and valid confidence intervals, without model assumptions. Biometrics 2019;75(4):1391-1400. [doi: 10.1111/biom.13062] [Medline: 31009064]
21. Senn S. Change from baseline and analysis of covariance revisited. Stat Med 2006;25(24):4334-4344. [doi: 10.1002/sim.2682] [Medline: 16921578]
22. PubMed: "nutrition exercise app". National Library of Medicine. 2024. URL: https://pubmed.ncbi.nlm.nih.gov/?term=nutrition+exercise+app&sort=pubdate [accessed 2024-02-20]
23. Molenaar CJL, Minnella EM, Coca-Martinez M, Ten Cate DWG, Regis M, Awasthi R, et al. Effect of multimodal prehabilitation on reducing postoperative complications and enhancing functional capacity following colorectal cancer surgery: the PREHAB randomized clinical trial. JAMA Surg 2023;158(6):572-581 [FREE Full text] [doi: 10.1001/jamasurg.2023.0198] [Medline: 36988937]
24. Shelton E, Barreto NB, Bidwell S, Folk-Tolbert M, Shelton A, Trickey AW, et al. Engagement and adherence with a web-based prehabilitation program for patients awaiting abdominal colorectal surgery. J Gastrointest Surg 2021;25(12):3198-3207. [doi: 10.1007/s11605-021-05171-2] [Medline: 34668165]
25. Duarte-Rojo A, Bloomer PM, Grubbs RK, Stine JG, Ladner D, Hughes CB, et al. Use of a mobile-assisted telehealth regimen to increase exercise in transplant candidates: a home-based prehabilitation pilot and feasibility trial. Clin Transl Gastroenterol 2023;14(11):e00601 [FREE Full text] [doi: 10.14309/ctg.0000000000000601] [Medline: 37477616]
26. Purdy GM, Venner CP, Tandon P, McNeely ML. Feasibility of a tailored and virtually supported home exercise program for people with multiple myeloma using a novel eHealth application. Digit Health 2022;8:20552076221129066 [FREE Full text] [doi: 10.1177/20552076221129066] [Medline: 36249481]

XSL•FO
RenderX

27. Amiri P, Nadri H, Bahaadinbeigy K. Facilitators and barriers of mHealth interventions during the Covid-19 pandemic: systematic review. BMC Health Serv Res 2023;23(1):1176 [FREE Full text] [doi: 10.1186/s12913-023-10171-w] [Medline: 37898755]
28. Mehta SP, Fulton A, Quach C, Thistle M, Toledo C, Evans NA. Measurement properties of the lower extremity functional scale: a systematic review. J Orthop Sports Phys Ther 2016;46(3):200-216. [doi: 10.2519/jospt.2016.6165] [Medline: 26813750]
29. Winters-Stone KM, Medysky ME, Savin MA. Patient-reported and objectively measured physical function in older breast cancer survivors and cancer-free controls. J Geriatr Oncol 2019;10(2):311-316 [FREE Full text] [doi: 10.1016/j.jgo.2018.10.006] [Medline: 30344000]
30. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
31. Gallegos-Rejas VM, Rawstorn JC, Gallagher R, Mahoney R, Thomas EE. Key features in telehealth-delivered cardiac rehabilitation required to optimize cardiovascular health in coronary heart disease: a systematic review and realist synthesis. Eur Heart J Digit Health 2024;5(3):208-218 [FREE Full text] [doi: 10.1093/ehjdh/ztad080] [Medline: 38774382]
32. Deftereos I, Yeung JM, Arslan J, Carter VM, Isenring E, Kiss N, et al. Preoperative nutrition intervention in patients undergoing resection for upper gastrointestinal cancer: results from the multi-centre NOURISH point prevalence study. Nutrients 2021;13(9):3205 [FREE Full text] [doi: 10.3390/nu13093205] [Medline: 34579082]
33. Nazari S, Leite WL, Huggins-Manley AC. Enhancing the detection of social desirability bias using machine learning: a novel application of person-fit indices. Educ Psychol Meas 2024;84(6):1107-1137. [doi: 10.1177/00131644241255109] [Medline: 39493800]
34. Heslop PA, Hurst C, Sayer AA, Witham MD. Remote collection of physical performance measures for older people: a systematic review. Age Ageing 2023;52(1):afac327 [FREE Full text] [doi: 10.1093/ageing/afac327] [Medline: 36721962]
35. Guidarelli C, Lipps C, Stoyles S, Dieckmann NF, Winters-Stone KM. Remote administration of physical performance tests among persons with and without a cancer history: establishing reliability and agreement with in-person assessment. J Geriatr Oncol 2022;13(5):691-697 [FREE Full text] [doi: 10.1016/j.jgo.2022.02.002] [Medline: 35177378]
36. Hoenemeyer TW, Cole WW, Oster RA, Pekmezi DW, Pye A, Demark-Wahnefried W. Test/retest reliability and validity of remote vs. in-person anthropometric and physical performance assessments in cancer survivors and supportive partners. Cancers 2022;14(4):1075 [FREE Full text] [doi: 10.3390/cancers14041075] [Medline: 35205823]
37. Essery R, Geraghty AWA, Kirby S, Yardley L. Predictors of adherence to home-based physical therapies: a systematic review. Disabil Rehabil 2017;39(6):519-534. [doi: 10.3109/09638288.2016.1153160] [Medline: 27097761]
38. Ricke E, Dijkstra A, Bakker EW. Prognostic factors of adherence to home-based exercise therapy in patients with chronic diseases: a systematic review and meta-analysis. Front Sports Act Living 2023;5:1035023 [FREE Full text] [doi: 10.3389/fspor.2023.1035023] [Medline: 37033885]
39. Ehlers DK, Fanning J, Awick EA, Kramer AF, McAuley E. Contamination by an active control condition in a randomized exercise trial. PLoS One 2016;11(10):e0164246 [FREE Full text] [doi: 10.1371/journal.pone.0164246] [Medline: 27723781]
40. Chaudhry UAR, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
41. Ji Y, Plourde H, Bouzo V, Kilgour RD, Cohen TR. Validity and usability of a smartphone image-based dietary assessment app compared to 3-day food diaries in assessing dietary intake among Canadian adults: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(9):e16953 [FREE Full text] [doi: 10.2196/16953] [Medline: 32902389]

Abbreviations

COM-B: capability, opportunity, and motivation behavior
LEFS: Lower Extremity Functional Scale
PIONEER: Personalized Online Nutrition and Exercise Routines
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
WHO-5: World Health Organization-5

XSL•FO
RenderX

Edited by N Cahill; submitted 20.02.24; peer-reviewed by CM Wierts, IA Eide; comments to author 08.11.24; revised version received 03.12.24; accepted 21.01.25; published 24.03.25.

Please cite as:
Tandon P, Ismond KP, Purdy G, Cruz C, Etruw E, Suderman K, Hyde A, Stickland M, Spence JC, Lien DC, Bhanji R, Prado CM, Miguel-Cruz A, Joy AA, Yaskina M, McNeely ML
Acceptability and Effectiveness of a Fully Web-Based Nutrition and Exercise Program for Individuals With Chronic Disease During COVID-19: Randomized Controlled Trial
J Med Internet Res 2025;27:e57537
URL: https://www.jmir.org/2025/1/e57537
doi:10.2196/57537
PMID:40126542

©Puneeta Tandon, Kathleen P Ismond, Graeme Purdy, Christofer Cruz, Evelyn Etruw, Kirsten Suderman, Ashley Hyde, Michael Stickland, John C Spence, Dale C Lien, Rahima Bhanji, Carla M Prado, Antonio Miguel-Cruz, Anil A Joy, Maryna Yaskina, Margaret L McNeely. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Discontinuous Usage Behavior of Digital Cognitive Training Among Older Adults With Mild Cognitive Impairment and Their Family Members: Qualitative Study Using the Extended Model of IT Continuance

Shangyang Zhang[1*], BS; Min Wu[2*], BS; Ruini Sun[2], MSc; Changjie Cui[3], LLM; Ziqing Zhang[4], MM; Jing Liao[2,5*], PhD; Ni Gong[1*], PhD

[1]School of Nursing, Jinan University, Guangzhou, China

[2]Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[3]School of Sociology and Anthropology, Sun Yat-sen University, Guangzhou, China

[4]School of Nursing, Capital Medical University, Beijing, China

[5]Global Health Institute, School of Public Health, Instiute of State Governance, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ni Gong, PhD
School of Nursing
Jinan University
No. 601, Huangpu Avenue West, Tianhe District
Guangzhou, 510632
China
Phone: 1 15013217344
Email: gongni_1025@163.com

Abstract

Background: Digital cognitive training (DCT) has been found to be more effective than traditional paper-and-pencil training in enhancing overall cognitive function. However, a significant barrier to its long-term implementation is that older adults with mild cognitive impairment (MCI) do not continue to use it or even show a dropoff in usage after the initial engagement. Such short-term engagement may limit the potential benefits of DCT, as sustained use is required to achieve more pronounced cognitive improvements. Exploring the reasons for the shift in discontinuous usage behavior is crucial for promoting successful DCT implementation and maximizing its positive effects.

Objective: This study aimed to explore the intrinsic reasons for the transition from initial acceptance to discontinuous usage behavior among older adults with MCI throughout the DCT process, by employing the extended model of IT continuance (ECM-ITC).

Methods: We employed a qualitative research methodology and conducted 38 semistructured interviews before and after the use of DCT (3 times per week over 1 month, with each session lasting 30 minutes) with 19 older adults with MCI (aged 60 years or older) and 4 family members between January and March 2024. Thematic analysis and deductive framework analysis were used to identify the reasons for the discontinuous usage of DCT, with mapping to the ECM-ITC.

Results: Most participants failed to complete the standard dosage of DCT. Data analysis revealed the reasons for the shift to discontinuous usage. Despite their need to improve cognitive function, participants found the cognitive training confusing and discovered that DCT did not align with their preferred method of training upon actual use. The disparity between their vague expectations and reality, combined with the contradiction between the “delayed gratification” of DCT and their desire for “immediate gratification,” made it difficult for them to discern the usefulness of DCT. Participants also viewed DCT as an additional financial burden and tended to avoid training under family pressure. They relied on motivational measures, which further weakened their intention to continue DCT, ultimately leading to the inability to develop continuous usage behavior.

Conclusions: Continuous usage behavior differs from initial acceptance as it evolves dynamically with user experience over time. To encourage older adults with MCI to persistently engage with DCT, it is essential to not only thoroughly consider their genuine preferences and the potential disruptions DCT may bring to their lives but also bridge the gap between expectations and

XSL•FO
RenderX

actual experiences. While ensuring that older adults receive appropriate external incentives and encouragement, it is equally important to foster their intrinsic motivation, thereby gradually cultivating the habit of sustained DCT usage.

(J Med Internet Res 2025;27:e66393) doi:10.2196/66393

KEYWORDS

digital cognitive training; discontinuous usage behavior; acceptance; mild cognitive impairment; qualitative study

Introduction

Cognitive dysfunction has become a major challenge in global health and social care, with progression to dementia being difficult to reverse [1,2] and imposing substantial disease and economic burdens on individuals, families, and society [3,4]. Mild cognitive impairment (MCI) is a transitional state between healthy cognition and early dementia [5]. Early and effective cognitive training can significantly delay or even reverse cognitive damage, with a reversal rate of up to 26% [6]. This stage is currently the optimal and critical window for cognitive intervention to prevent dementia [7,8].

In recent years, digital cognitive training (DCT), as an emerging training program, has proven more effective in enhancing cognitive functions than traditional paper-and-pencil training [9]. DCT provides personalized training in different cognitive domains based on the performance of older adults, breaking through temporal and spatial constraints. With its convenience, cost-effectiveness, and low operator requirements, DCT is increasingly becoming a preferred alternative [10-12]. In the United States, after the US Food and Drug Administration launched a digital software precertification program in 2017, digital products for cognitive behavioral therapy have been approved [13]. Meanwhile, in China, training software for cognitive dysfunction was approved for the market in 2021 [14], further propelling the development of DCT as a cognitive therapy tool. Millions of older adults have used DCT at home [15], and through digital brain training, they can not only effectively improve cognitive function [16,17] but also reduce the long-term risk of dementia [18].

DCT has shown potential benefits in maintaining cognitive function among older adults and has been applied to prevent and delay cognitive decline in patients with MCI due to its advantages of ease of operation and low cost [19]. Regrettably, significant challenges, such as user attrition, low training completion rates, and weak intentions to continue use, persistently impede its widespread adoption. Over 25% of users reduce participation frequency or drop out of training altogether [20,21], and a striking 52.8% fail to meet the prescribed dosage of cognitive training compliance [22]. After a period of use, compliance and retention rates among some older adults gradually decline [23], with less than 30% of participants completing more than half of the training content [21,24]. Even if older adults with MCI have a positive experience after their initial use of DCT, the intention to continue using it in the future diminishes significantly [25], and they do not continue to use it after the intervention ends [26,27]. More critically, the notable benefit enhancements of DCT largely depend on the long-term continuous use of the training by older adults. Noncompliant behavior not only undermines the potential training effects of DCT [28] but also leads to substantial wastage of medical resources.

Addressing this issue, previous research [29,30] has primarily focused on analyzing the factors that influence older adults' use or nonuse of DCT after an intervention, treating it as an instantaneous, one-off action. Consequently, low DCT utilization rates are often attributed to older adults' anxiety about using technology [29], inadequate computer experience [29,31,32], low self-efficacy regarding digital technology [31,32], and inevitable physical decline [33]. Although some scholars have explored the factors affecting the adoption of DCT from the perspectives of health care providers and family members, the explanations have largely been from a structural societal perspective, clarifying why there is severe noncompliance with DCT in practice, such as nursing staff shortage [34,35], inadequate senior management support and attention [34], caregivers' lack of game technology skills [34,35], and insufficient family supervision [22,36]. These studies have to some extent elucidated the difficulties and barriers encountered in the dissemination of DCT, yet they have overlooked the fact that DCT usage is not a static one-time action but a continuous dynamic process. Moreover, they have not considered the perspectives of older adults with MCI and have not interpreted the behavior of older adults with MCI using DCT throughout the intervention and within their social environment in a phased manner. Within the scope of our literature review, there are currently limited studies on the phases of nonuse and continued use of digital technologies by older adults [26,37]. The troubles and negative impacts of DCT at different stages of use on older adults with MCI, as well as the intrinsic reasons for transitioning from initial acceptance to discontinuous use, have not been adequately explored.

Therefore, from the perspective of this study, continuous and effective usage is key to the successful implementation of DCT programs. We adopted the extended model of IT continuance (ECM-ITC) and conducted qualitative interviews to better understand the reasons behind the shift from acceptance to discontinuous usage among older adults with MCI at different time points before and after using DCT. This model is a representative theoretical framework for exploring individuals' continuance intentions and behaviors [38], providing in-depth insights into the formation processes and motivations of continuous usage behavior in the digital technology domain [39]. Continuance intention, a critical driver of long-term technology use, depends on users' perceived usefulness of the system and their satisfaction with its use, both of which are influenced by the degree of expectation confirmation. Consequently, this research employs the ECM-ITC to conceptualize the usage of DCT by the MCI population as a dynamic and continuous process. The objective is to understand

XSL•FO
RenderX

the reasons contributing to the phenomenon of discontinuous usage over time during the training process and to delve into the perceptions and actual needs of older adults with MCI and their families regarding DCT usage. By starting from their real-life contexts, we explain the reasons for high attrition rates during the continued use phase and provide decision support to enhance the adoption of intervention strategies.

Methods

Study Design

This was a convergent parallel mixed-methods study in which 2 research teams concurrently collected quantitative and qualitative data. The analysis findings from both methods were integrated when interpreting the results to obtain a more comprehensive understanding of the study results [40]. However, this study focuses solely on qualitative data, aiming to deeply explore the underlying reasons that affect the discontinuous use of DCT from the perspectives of older adults with MCI and their families. For the quantitative aspect, we assessed the effectiveness of DCT through an intervention program involving three 30-minute sessions weekly over a 1-month period. The DCT tool used in this study is a cognitive intervention platform based on a gamified design that aims to improve users' cognitive functions by promoting training in various cognitive domains. This tool covers multiple cognitive areas, including memory, attention, logical reasoning, perception, processing, and executive functions. The DCT tool is accessed through WeChat Mini Programs, the most widely used mobile app in China. Its main features include artificial intelligence chat, gamified training (16 training games with 90 levels each; continuously updated), and progress feedback tracking, providing interactive and timely feedback. Additionally, the gamified training is designed with detailed usage instructions and progressive difficulty adjustments to continuously stimulate cognitive improvement. The qualitative component of our study was based on the ECM-ITC, which provided the framework for developing interview outlines and for organizing and analyzing data [38]. This model serves as a representative theoretical framework that explores the processes and motivations behind individuals' continued use of digital technology [39]. This study conducted semistructured qualitative interviews using an interview guide before and after using DCT. This qualitative research complied with the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [41].

Patient Recruitment

We employed a purposive sampling method to select participants who satisfied the inclusion criteria and could provide rich information. Eligible older adults with MCI and their family members participated in cognitive screening and DCT at a community health service center in Guangzhou, Guangdong Province. The inclusion and exclusion criteria for both the qualitative and quantitative components of the study were identical. The inclusion criteria for patients were as follows: (1) aged 60 years or older; (2) met the diagnostic criteria for MCI (subjective cognitive decline, such as memory loss, reported by the patient or family member; objective evidence of cognitive decline confirmed using the Mini-Mental State Examination [MMSE] screening tool [cognitive impairment cutoff scores set according to different educational levels [42]: illiterate, <20 points; elementary school, <23 points; secondary school or above, <27 points]; further assessment of participants who tested positive on the MMSE and confirmation by professional geriatricians or neurologists based on the 11th revision of the International Classification of Diseases [ICD-11] [43] and Petersen criteria [44]; essentially normal daily living abilities, although complex instrumental activities of daily living might show slight impairment; and did not fulfill the diagnostic criteria for dementia); (3) possessed a smartphone and was able to use WeChat; and (4) provided informed consent (participation was voluntary). The exclusion criteria were as follows: (1) concurrent other neuropsychiatric disorders, such as epilepsy, stroke, and Parkinson disease; (2) severe cardiovascular or cerebrovascular diseases, malignant tumors, or other major physical illnesses; (3) significant sensory-perceptual disturbances that could affect assessment or training; (4) severe visual or auditory impairments hindering communication; and (5) participation in any other cognitive training studies within the past 6 months. The inclusion criteria for family members were as follows: (1) aged 18 years or older, with clear awareness and the ability to read, understand, and express in Chinese normally; (2) had familiarity with the patient's condition, was involved in daily care, and was capable of decision-making; and (3) provided informed consent (participation was voluntary). The exclusion criterion for family members was a history of severe mental illness or physical disease. Health care providers at the community health service center conducted cross-assessments using both paper-based and digital cognitive screening tools. After review by a psychiatrist, 21 older adults met the inclusion criteria. Subsequently, researchers, with the assistance of the community health service center staff, recruited patients with MCI and their family members through on-site face-to-face or telephone interactions. Ultimately, 19 older adults with MCI and 4 family members agreed to participate in DCT. Two older adults declined to participate due to blurred vision or lack of interest.

Data Collection

Data for this study were collected between January and March 2024, encompassing a total of 38 interviews. We conducted semistructured one-on-one interviews with 19 older adults with MCI and 4 family members both before and after the 1-month intervention period. During the interviews, the older adults and their family members were interviewed independently, each in separate rooms. The interview guide used the same version for all participants, with consistent questions. However, the phrasing of the questions was appropriately adjusted according to whether the respondent was an older adult or a family member, ensuring that the questions were more aligned with their context and level of understanding. The average duration of the interviews, both before and after use, was 21 minutes (ranging from 12 to 80 minutes). The interview questions were formulated based on a literature review and the ECM-ITC framework and were further refined following discussions with the research team, as well as consultations with experts in cognitive impairment screening and management, and qualitative research specialists,

XSL•FO
RenderX

culminating in the finalized interview guide (Multimedia Appendix 1).

At their first offline participation in DCT, patients and their family members who met the inclusion criteria were invited to join a WeChat group managed by health care providers, social workers, and researchers for cognitive training. After setting up the DCT at the end of the interview, we assisted the older adults and their families in accessing the Thoven mini-program training games via a WeChat QR code. We guided them on-site on how to open the WeChat mini-program, operate the training games within the mini-program, interpret the game rules, and check their personal training scores. Finally, we reconfirmed that the participants had learned all the training steps and reiterated the prescribed dosage of the training (3 times a week, 30 minutes each session, for a duration of 1 month). Subsequently, during the DCT process, participants received WeChat group messages from staff twice a week. These messages included rankings of the week's training completion status, encouragement to complete DCT, and real-time answers to older adults' questions to help them persist with the training.

During the first interview (after enrollment but before intervention commencement), 14 of the 19 participants underwent face-to-face interviews in community health service center meeting rooms, clinics, or community senior activity centers. The remaining 5 participants, due to scheduling conflicts, opted for telephone interviews at a convenient time after returning home. The preuse interview questions primarily focused on their perceptions of cognitive decline, expectations for improving cognitive health, and willingness to engage in DCT.

After the 1 month of training, all participants were invited for second interviews (postuse interview). Of the 19 participants, 15 engaged in face-to-face interviews in community health service center meeting rooms, clinics, or community senior activity centers. The remaining 4 participants, due to scheduling conflicts, opted for telephone interviews at a convenient time after returning home. The older adults with MCI and their family members shared their experiences with DCT, while researchers assisted in reviewing their preuse expectations and motivations, and discussed the extent to which actual experiences matched initial expectations, the perceived usefulness of DCT, reasons for either persisting with or dropping it halfway through the process without continuing to use it, and their intentions to continue using it in the future.

All interviews were conducted by 2 interviewers (authors SZ and CC), both of whom had received training in qualitative research and had experience conducting qualitative interviews. All interviews were carried out in Chinese and were audio-recorded with the consent of the older adults and their family members prior to the start of the interviews. The recordings were transcribed verbatim within 2 weeks following the interview. During the interview process, researchers also observed and collected nonverbal data, such as the participants' facial expressions, body language, and emotional changes. Before commencing the second interview, researchers reviewed transcription data from the participants' previous time points to aid in recalling their participation expectations and motivations. Data collection ceased when no new information could be extracted from the analysis of the new interview data, which was achieved after 34 interviews, signifying data saturation.

Data Analysis

Data collection and analysis were conducted concurrently. Deductive thematic analysis was employed to analyze the individual interview data [45]. The coding was carried out using NVivo 12.0 software (Lumivero), and manual coding was performed in Microsoft Excel. The entire analysis process was completed under the guidance of NG and JL. The ECM-ITC was used as the coding framework for categorizing subthemes. First, the first (SZ) and second (MW) authors transcribed the original recordings into text and familiarized themselves with the content through repeated readings. Throughout the analysis, initial codes were independently assigned by repeatedly reading the transcripts, listening to the recordings, and incorporating field notes. These codes were then organized and categorized into corresponding subthemes. Subsequently, the research team, together with NG and JL, reviewed and discussed these codes and subthemes. Next, the first author (SZ), under the guidance of NG, who has extensive qualitative research experience, applied a deductive framework analysis method to construct a thematic matrix in Microsoft Excel. This matrix displayed themes, subthemes, and quotations, mapping the subthemes onto the components of the ECM-ITC framework (expectation confirmation, perceived usefulness, and continuance intention). Finally, the first author and all research team members collectively reviewed and discussed the themes and subthemes until a consensus was reached.

Ethical Considerations

This study meets the institutional requirements for human trials and has been approved by the Biomedical Research Ethics Committee of the School of Public Health, Sun Yat-sen University, China (ethics approval number: 2023-129). Subjects have voluntarily signed informed consent forms. Prior to the interviews, participants received both verbal and written explanations about the study, including its objectives, its methods, the approximate duration and frequency of the interviews, and other details. All participants signed written informed consent forms and were given the right to withdraw from the study at any stage. Data from all participants and institutions have been anonymized. Moreover, this research followed the ethical principles of the Declaration of Helsinki.

Rigor

To ensure that the research findings authentically reflect the perspectives of the participants, we scrutinized and assured the rigor of the data based on the criteria established by Lincoln et al [46], which include confirmability, credibility, transferability, and dependability. The purposive sampling method was employed based on the research objectives and questions to select participants, ensuring the complexity and relevance of the subjects under study, thereby yielding a rich diversity of insights. In addition, the credibility of the research was enhanced by multiple rounds of qualitative data collection. Throughout the study, field notes and memos were used, and a thorough

XSL•FO
RenderX

description of the participants and the data collection process was provided to ensure transferability and dependability. Additionally, the study followed a consistent process and employed the ECM-ITC model framework to guide the design of the interview guides, as well as data collection and analysis, thereby enhancing its dependability. Confirmability was established through the systematic handling of recordings and interview transcripts by 2 researchers and the collaborative discussion of data coding among the members of the research team.

Results

Overview

This study conducted a total of 38 interviews with 19 older adults with MCI and 4 of their family members, both before and 1 month after implementing DCT. The sample size was determined based on the principle of informational saturation. Data saturation was achieved after interviewing 17 participants, and we conducted 2 additional interviews to confirm that no new information emerged from the new data. Interviews were carried out either at the participants' homes or on-site. Due to the inability of some participants to attend in person, 9 interviews were conducted via telephone, while 29 were conducted face-to-face on-site. The mean age of the older adults with MCI was 73.3 years (range 60-90 years). Their characteristics are presented in Table 1.

Among the 19 participants, only 32% (6/19) completed the recommended 90 minutes of weekly training, while 68% (13/19) of respondents used DCT occasionally or even discontinued its use. Detailed variations in the participants' training over the 1-month period are illustrated in Figure 1.

In this study, we analyzed the results based on the ECM-ITC (Figure 2) and identified 3 primary themes that revealed the reasons behind the discontinuous usage behavior of DCT among older adults with MCI: (1) discrepancy between vague expectations and actual training experience; (2) unclear perceived usefulness; and (3) suppressed intention to continue usage.

Table 1. Participant characteristics.

Characteristic	Value
Persons living with mild cognitive impairment (n=19)	
Age (years)	
Mean (SD)	73.3 (7.2)
Range	60-90
Sex, n (%)	
Male	6 (32)
Female	13 (68)
Highest level of education, n (%)	
Elementary school	4 (21)
High school	11 (57)
College or trade school	4 (21)
Living arrangement, n (%)	
Staying alone	6 (32)
Staying with spouse only	6 (32)
Staying with family members	6 (32)
Staying with helper	1 (5)
Medical history, n (%)	
Hypertension	7 (37)
Diabetes	7 (37)
Family members (n=4)	
Relationship with the patient, n (%)	
Daughter	2 (50)
Wife	2 (50)

XSL•FO
RenderX

Figure 1. Adherence rates of digital cognitive training.

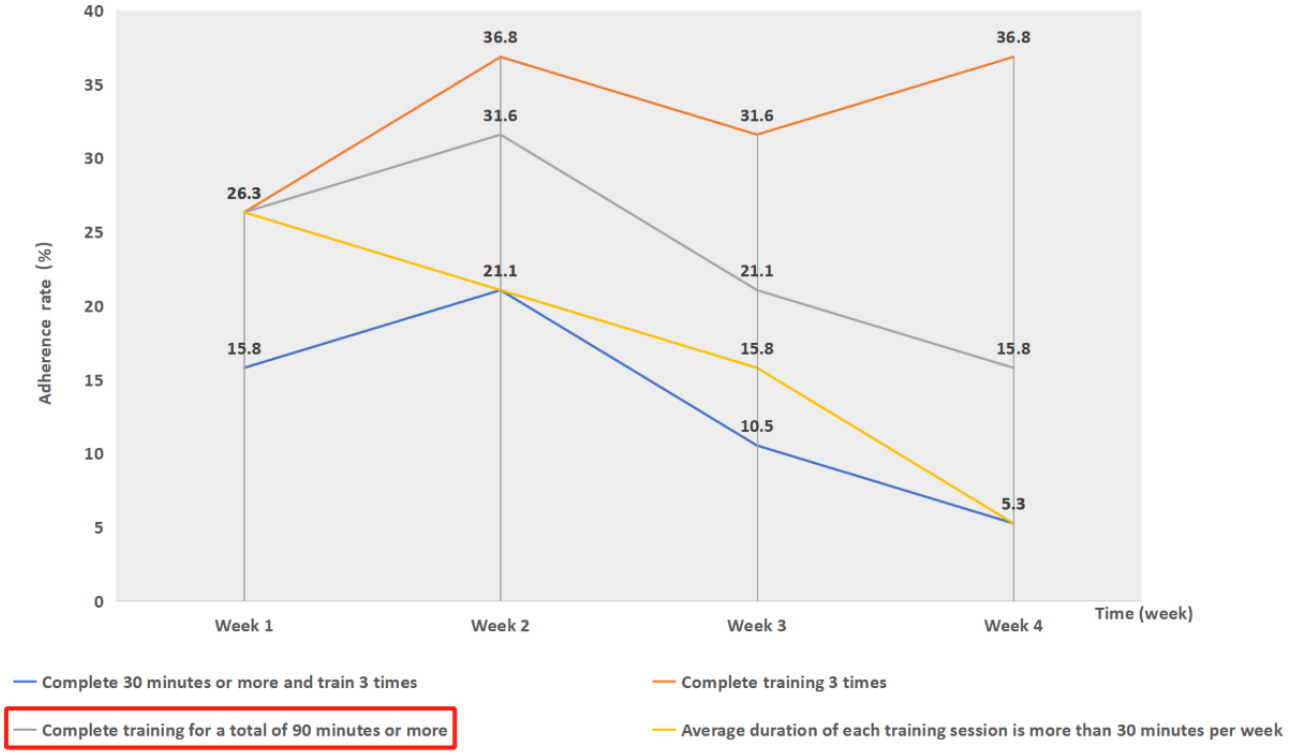

Figure 2. Extended model of IT continuance in digital cognitive training.

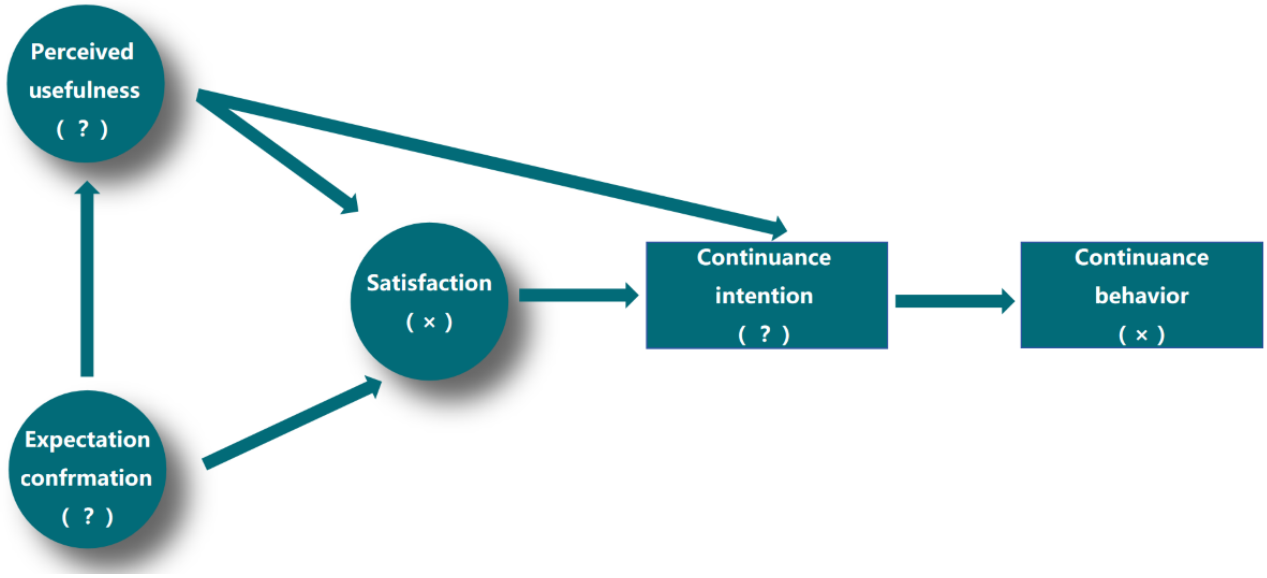

Theme 1: Discrepancy Between Vague Expectations and Actual Training Experience

To comprehend the needs of older adults with MCI for DCT, it is crucial to consider the cognitive background and social context in which they live. The majority of participants (14/19, 74%) generally had limited awareness of cognitive impairment diseases and were influenced by misconceptions and prejudices of others regarding cognitive disorders or dementia in their social environment. This led to older adults with MCI needing immense courage to confront discrimination and fear. Cognitive biases and the stigma associated with the disease created confusion in their decision-making process on how to improve memory, causing them to embark on DCT with a sense of bewilderment. However, negative training experiences after use made patients with MCI and their families realize that DCT does not align with their needs. This implies that a greater gap between the ambiguous expectations before using DCT and the actual experience thereafter is associated with a lower level of expectation confirmation. Consequently, over half of the participants (10/19, 53%) expressed reluctance to continue using DCT.

Cognitive Bias and Stigma of Cognitive Impairment

Regarding the issue of memory decline, older adults often held misconceptions. Most respondents (11/19, 58%) regarded

XSL•FO
RenderX

memory deterioration as a normal phenomenon of "aging" rather than a manifestation of disease, thereby failing to fully comprehend the dangers posed by the cognitive impairment:

> *Nowadays, I see that 80% of older adults are forgetful and the brains have degenerated. It's just the way it is when you're old.* [Participant #14, female]

Beyond the barrier of insufficient disease awareness, younger older adults were highly confident in their brain health, firmly believing that the usage of DCT is exclusive to dementia patients and incongruent with their current age stage. A family member of an older adult with MCI who had given up DCT usage expressed that younger older adults are reluctant to acknowledge their aging because they consider themselves as "still young" and not yet at the stage where they need cognitive training.

> *They think it's useless, they're still young, they don't acknowledge that thing (DCT), they feel it's not yet time to engage with it.* [Family member #4, female]

In contrast, older seniors often lacked confidence in using DCT. Due to age-related and physical limitations, they were unable to master the use of smartphones and DCT. Older adults with MCI tended to adopt a go-with-the-flow attitude toward their future lives, which to some extent reflects a negative correlation between advanced age and adherence to DCT. One family member, when talking about cognitive improvement through training, frowned, sighed, and indicated that a sense of meaninglessness in life arises in old age.

> *It's not that I don't want him to do the training, but he (older adults) doesn't care about anything, has completely withdrawn from society, and never trying to get better either, never wanting to change.* [Family member #1, female]

Prejudices against cognitive impairment from those around them made it difficult for older adults to accept cognitive decline and engage with DCT. In the context of this environment, facing cognitive training was particularly challenging. Some participants (2/19, 11%) often chose to hide their condition and remain silent when confronted with cognitive impairment-related diseases.

> *Yes, I have memory decline, but I dare not speak up face to face.* [Participant #16, female]

Disorientation and Perplexity Regarding Cognitive Training: Not Knowing What They Want

When older adults with MCI were informed of their cognitive decline, both patients and their family members often found themselves at a loss. Respondents expressed a need for cognitive improvement when asked about methods they believe could enhance memory. However, they were unclear about which type of cognitive training they desired and were even confused about whether effective treatment methods existed.

> *I want to improve my memory, but I don't know how to do it. Taking medication doesn't seem very convenient either.* [Participant #16, female]
>
> *I don't know (what can improve memory) anymore, maybe participate in activities.* [Participant #18, female]
>
> *Is there a way to improve (memory)? Does such a thing exist? (asked with confusion and urgency)* [Family member #2, female]

Negative Experiences With Training Games: Knowing What They Do Not Want

Negative experiences during the training process can hinder older adults with MCI from actively trying again, reducing their interest and motivation in training games. Some participants (3/19, 16%) pointed out a disconnect between the design of certain training games and the preferences of older adults, stating that the content was "too juvenile" and failed to adequately meet the actual needs of older adults. Others felt frustrated by the training challenges they encountered. After these negative experiences, older adults with MCI and their family members were clear that they did not wish to continue using DCT in the form of training games. In their logic, some traditional training activities (such as playing mahjong, reciting textbooks, etc) could achieve the same effect, leading to a decreased intention to continue using DCT:

> *I didn't play Whac-A-Mole; it's not fun, feels like something for kindergarten kids.* [Participant #13, female]
>
> *I'm really not interested in this game; I've already deleted it. Simpler would be better; this is too complicated.* [Participant #5, female]
>
> *Playing this (game), I want to delete it because it's just testing my memory, I might as well recite a book instead.* [Participant #19, female]
>
> *These little games are no good, Playing this kind of thing isn't as good as playing other things, let him play cards or mahjong tomorrow.* [Family member #3, female]

Theme 2: Unclear Perceived Usefulness

The majority of participants (11/19, 58%) revealed a preference for on-site group training over at-home DCT. Older adults with MCI generally tended to favor what they perceived as substitutable training methods. The perceived usefulness was influenced by the extent of expectation confirmation. A greater discrepancy between preuse expectations and actual postuse experiences was associated with greater skepticism about the usefulness of DCT among older adults with MCI. Although the interviewees acknowledged that persistent long-term training might improve cognitive abilities, the delayed gratification characteristic of DCT did not meet the urgent need for immediate effect among older adults with MCI. This unmet need for instant feedback can influence their attitude toward the usefulness of DCT, which remains uncertain.

Preferences for On-Site Training and Substitutable Methods

Most older adults (9/19, 47%), especially those living alone, disliked loneliness and desired "someone to interact and communicate with" to alleviate loneliness. Therefore, on-site group training aligned better with their daily habits and helped establish an emotional support network.

> *I think it should be offline. I did it with her and felt that we still need to have some of this companionable*

XSL•FO
RenderX

> *atmosphere. She keeps asking me, 'Am I the only one? Are there other people?' It's much better with companions around.* [Family member #4, female]

Out of concern for being socially isolated, many participants (7/19, 37%) mentioned that they cultivated a variety of other interests and hobbies to enrich their lives and maintain social connections.

> *For her, she might be more interested in things like going for a walk by the river, dancing, and practicing Tai Chi. Given a choice, she would choose (other activities). Playing games would definitely be her last choice.* [Family member #4, female]

> *Actually, I have quite a broad range of interests. I like singing, (and) some reading aloud. Sometimes, I also play Idiom Fill in the Blanks on my mobile phone. (As she speaks, she takes out her phone to show photos of herself singing)* [Participant #9, female]

The Conflict Between "Delayed Gratification" of Digital Cognitive Training and "Immediate Gratification" Among Older Adults

The management of cognitive impairment diseases and the improvement of cognitive outcomes rely on sustained long-term training. However, for older adults with MCI undergoing training, immediate positive feedback on memory enhancement was crucial. They believed that achieving "immediate gratification" is key to ensuring adherence to training. One participant expressed a desire for immediate results following training.

> *I wish it would work right after one day of training, the sooner the better.* [Participant #13, female]

However, the greatest challenge of DCT was "delayed gratification," because the effects of the training could not be perceived in time. More than half of the patients (13/19, 68%) found it difficult to continue training when they could not see short-term progress, ultimately choosing to "lie flat" or even "give up."

> *If it's effective, we'll keep doing it; if not, we won't do it (shaking head). A little progress in a week is fine. If there's no effect, spending time on it is a waste.* [Participant #5, female]

> *I know (it requires training over a period of time to be effective), just like the health care products I'm taking now. If trying them out shows some effect, I'll continue buying them; if they haven't effect, then I won't buy them anymore. It's like that.* [Participant #17, female]

Uncertainty About the Usefulness of Training

Older adults with MCI often reduced or withdrew from training before achieving their cognitive enhancement goals, which prevented them from timely experiencing the positive changes that the training could bring and left them with limited knowledge of the importance and potential benefits of DCT. One participant viewed DCT as a tool to provide recreational value and lacked confidence in it as an intervention that could effectively improve cognitive function.

> *I don't know if it's useful or not, I just play with it.* [Participant #17, female]

Interestingly, even those older adults with MCI who had higher compliance with DCT training expressed concerns about whether brain game training can genuinely delay cognitive deterioration, further exacerbating the skepticism among older adults with MCI regarding the usefulness of the training.

> *It should be useful after training, it's bound to be somewhat better.* [Participant #12, male]

> *It (Memory) is all more or less the same, learning a bit more might make the mind a bit more enlightened, but it's hard to say whether it will prevent dementia.* [Participant #18, female]

Theme 3: Suppressed Intention to Continue Usage

The intention of older adults with MCI to continue using DCT was not only influenced by their perceived usefulness but also constrained by various external driving factors. Concerns about the economic risks associated with DCT, lack of support from family members, stringent task requirements from intervention personnel, and the absence of incentives were among the reasons that could trigger their dissatisfaction or even resistance. These external factors negatively impacted the acceptance and user experience of DCT, while participant satisfaction with training was a crucial determinant of their continuance intention. Participant dissatisfaction can weaken their intention to use DCT to delay cognitive decline, even if the training itself is beneficial to cognitive function.

Concerns About the Economic Risks of Game Training

Older adults with MCI generally lacked trust in DCT's QR scanning behavior and training game systems. This low level of trust often stems from negative experiences they have heard from others. Participants expressed concerns that the QR codes might not be a reliable source and that the services and content accessed through scanning could carry the risk of online fraud. One participant, despite being willing to scan QR codes, remained cautious and stated:

> *At that time, when I entered the mini-program, I was also really scared. There are many scams targeting older adults. It would be really bad luck to encounter one, especially now with so much online fraud.* [Participant #14, female]

Additionally, this distrust also led some participants to be highly skeptical about the motives behind data collection by the community health service center staff and those conducting the intervention, worrying about potential breaches of personal information privacy. This made them hesitant to engage in any activities involving QR scanning.

> *You don't know what might happen after scanning QR codes. I'm afraid of being charged or that they might do something to the phone after scanning, so I don't like scanning QR codes.* [Participant #17, female]

XSL•FO
RenderX

Among older adults with MCI, even when explicitly informed that DCT was free to use, respondents still worried about potential hidden fees and additional expenses during the DCT process.

> *That's why I'm afraid to play the games; I'm worried about charges, concerned about deductions from my phone bill.* [Participant #16, female]

Avoidance of Cognitive Training Under the Pressure of Family Relationships

The pressures associated with family relationships also had a negative impact on participation in DCT. When discussing the role of their families in supporting and encouraging them to persist with DCT, participants expressed reluctance to proactively inform their family members about their involvement in DCT. They believed that they would not receive understanding or attention from their family members. Additionally, due to heavy family responsibilities and pressures, compared to the misfortunes of their lives, memory decline was rather not a bad thing for them. As a result, engaging in cognitive function training was merely an optional aspect of their lives. An older adult, who had long been responsible for the entire household's chores, mentioned that her life was consumed by a myriad of family tasks, leaving her with almost no personal rest time, let alone time to use DCT during her leisure.

> *My son doesn't care about me at all. It's better not to provoke him. My daughter-in-law has the same attitude; she's very aggressive in the way she speaks. They (my son and daughter-in-law) don't do any housework; I do everything in the house. We, the unpaid nannies, are all the same. That's just how it is for older adults in China. So, I'm actually happier during the New Year when no one bothers me.* [Participant #13, female]

During the training process, understanding and support from family members could help to inspire older adults with MCI to use DCT with enthusiasm. However, participants reported that children rarely took the initiative to care about the health condition of older adults, and complex family relationships could leave older adults with MCI feeling more depressed and demotivated. Sometimes, they felt ignored and lacked the motivation to improve their health. In such circumstances, avoiding DCT became a way for older adults with MCI to alleviate their psychological burden.

> *My son seldom keeps me company (silent, sighs softly). Even when I was sick and hospitalized, he didn't even say a word of greeting when he returned home.* [Participant #16, female]

> *I got divorced from my spouse, and later, when my child was in college, he got swindled by online gambling. I had to send my child to detention for a month. Sometimes, I just can't control him. He (my son) only cares about money; you'd better give money to him. Under these circumstances, it's impossible to feel happy.* [Participant #12, male]

The Intrinsic Paradox of Incentive Mechanisms

Despite older adults with MCI overcoming financial and privacy concerns and being able to obtain support from family and society, they still participated in training in a task-driven, passive manner. Their willingness to persist with the training was primarily a response to the requirements set by community health service centers and their staff. On the one hand, these older adults maintained positive relationships and interactions with the staff at community health service centers or social work stations, which made them more willing to cooperate in activities.

> *I definitely support you and support your work in completing your studies and careers.* [Participant #9, female]

On the other hand, older adults with MCI did not have an intrinsic interest in learning about DCT. In this context, the encouragement and strict demands of intervention personnel became significant external forces driving them to passively complete the training tasks.

> *If you ask me to do it, I'll get your thing done first thing in the day before doing other things. It's mainly to finish the task, not because there's any interest.* [Participant #18, female]

These older adults viewed the use of DCT mainly as a means to cooperate with staff in completing tasks. Therefore, incentive mechanisms play an important role in prompting them to start or continue using DCT. When asked how to make older adults with MCI more willing to persist with training games over the long term, they expressed a desire for "immediate rewards" during the training process, which could provide a great sense of satisfaction.

> *To complete the training, you need to post a record of completion in the group to get a red packet. It's like the classes we attend (lectures on cardiovascular and cerebrovascular health), after the class, you post your notes in the group chat, and then every day they give you 8.88 yuan in cash immediately.* [Participant #13, female]

Furthermore, incentive measures can effectively attract older adults to participate in training in the short term. With rewards available, they treated participating in DCT as a clocking-in task and were thus more willing to continue using DCT and to remember to train regularly.

> *It's better with rewards. When there are little perks, they feel like they haven't done this task today.* [Participant #18, female]

> *If there's a red packet, they'll remember that time. You just stipulate a time to complete the training and give out the red packet. Some people think about the red packet and then they're ready to play.* [Participant #17, female]

XSL•FO
RenderX

Discussion

Principal Findings

Our findings indicate that the 3 dimensions of expectation confirmation, perceived usefulness, and continuance intention are critical factors facilitating continuous usage behavior. Older adults with MCI and their families were unclear about the desired form of training they truly wanted prior to using DCT. However, after the actual experience, they explicitly expressed a preference not to undergo interventions through training games. This significant discrepancy between expectations and actual training experience, coupled with the unclarity of the perceived usefulness of DCT, resulted in their low satisfaction with DCT. Over time, various extrinsic factors further suppressed their intention to continue using DCT, making it difficult to establish adherence to DCT and ultimately preventing the formation of continuous usage behavior.

This study employed the ECM-ITC as a guiding framework, specifically focusing on the social environment and daily life contexts of older adults with MCI. It qualitatively explored the critical reasons that influence the shift from initial acceptance to discontinuous usage of DCT throughout the process of the training intervention. The adoption of digital technology is a nonlinear progression from initial acceptance to sustained usage at a later stage [37], and this progression is particularly challenging for the specific group of older adults with MCI. They may face greater challenges in accepting and continuously using digital technologies. Within the ECM-ITC, low levels of self-efficacy in using digital tools and insufficient convenience of participation have been identified by numerous studies as moderating factors affecting the long-term adherence of older adults to intervention measures [29,31,47]. Previous research has primarily concentrated on assessing the immediate effects following the conclusion of DCT interventions. Ample evidence suggests that DCT can significantly improve overall cognitive function, enhance quality of life, and alleviate psychological symptoms, such as anxiety, in patients with cognitive impairments [17,23,48]. However, the enhancement of cognitive training effects should not only focus on the instantaneous effects of short-term use but also prioritize the more important long-term sustained effects, which depend on good adherence from participants. Regrettably, older adults with MCI are reluctant to continue using DCT after receiving the intervention. Why is this the case? We found that prior research often regards the acceptance behavior of DCT as a static decision, conflating the initial acceptance with continued usage and only analyzing the reasons for the low usage rates of DCT based on participants' postuse experiences and satisfaction. Consequently, this study conducted multiple interviews at different stages before and after training to place the behavior of older adults with MCI using DCT in a dynamic context, exploring the key barriers that prevent their adherence to DCT usage.

Expectation confirmation, which represents the extent to which preusage expectations match postusage actual experiences [38], is a significant initial factor affecting the continuous usage behavior of older adults with MCI. A higher degree of expectation confirmation is associated with a greater likelihood that users will perceive the technology as useful and feel satisfied with it [49]. However, if the individual's expectations are vague and there is a significant gap between the actual experience and expectations, this internal conflict can weaken the user's intention to continue using the technology [50]. Our study found that older adults often have cognitive misconceptions and biases about cognitive impairment diseases, and the associated stigma leads to insufficient attention to cognitive health, which reduces their acceptance and willingness to continue using DCT. This is consistent with previous research [51,52] showing that most community-dwelling older adults lack a comprehensive and correct understanding of MCI. Furthermore, some demographic factors, such as age and education level, limit the improvement of disease cognition [51], and older adults with lower educational levels are more prone to experiencing digital technology anxiety [53]. During the DCT process, the initial detailed guidance, encouragement, and reminders provided by the staff enhanced the initial confidence of the older adults participating in the training. However, as the difficulty and complexity of the training increased and the content went beyond the scope of the initial guidance, the gap between the frustration of the actual training experience and the expected confidence led to a gradual decline in the older adults' interest in DCT. This further weakened their motivation and sense of participation. In the social environment of older adults with MCI, their logical thinking tends to perceive traditional entertainment activities, such as playing mahjong, as more compatible with their accepted method of exercising the brain than DCT. For older adults with MCI, this preference stems from the simplicity and familiarity of traditional entertainment activities, which do not require them to learn new skills or adapt to new rules. In addition, these activities not only provide brain training but also enhance manual dexterity and emotional connections.

This study also found that older adults with MCI are caught in a paradox between the need for immediate gratification and the delayed gratification characteristic of DCT. The effects of DCT are often gradual and not immediately perceptible. Evidence suggests that the average duration of digital health interventions spans from 4 to 24 weeks [54], and it takes 5 to 7 weeks of cognitive training for older adults with MCI to exhibit improvements in working memory [55], a perspective supported by our study results. After completing DCT, older adults can receive immediate visual and tactile feedback, such as improved training scores, increased accuracy, and enhanced hand-eye coordination, allowing them to perceive progress in game task completion and improvements in operational abilities. However, improvements in cognitive functions like memory and executive function often take a longer period of time to manifest. Older adults with MCI are unable to perceive substantial memory improvements from the game in the short term, leading to a lack of sufficient confidence in the usefulness and effectiveness of DCT [28]. This lack of confidence leads them to prefer on-site training or training methods with social attributes [56]. From the patients' perspective, the current research also highlights the importance of combining cognitive training with physical exercise. Previous studies have noted that compared to gamified cognitive training, motor-cognitive training not only has higher ecological validity but also is more effective in improving

XSL•FO
RenderX

cognitive function performance [57,58]. In the context of the widespread prevalence of intelligent and digital technologies, older adults frequently hear many promises about uncertain future outcomes during their involvement in digital interventions. For them, immediate experiences and immediate gratification have become important requirements and experiences in their daily lives, driving a preference for activities that offer immediate satisfaction rather than digital intervention programs that require long-term commitment to see results. Therefore, combining exercise science with cognitive training that provides immediate gratification aligns better with the primary preferences of older adults.

Finally, the interviewees shared their suppressed intention to continue usage, which is a key factor directly affecting individuals' continuous usage behavior. Previous research has shown that a general lack of trust in digital technology and its sources of promotion [59-61], along with significant life stressors [62], hinder the adoption of digital health care by participants. More importantly, usage intentions driven by genuine self-awareness and autonomous motivation to satisfy intrinsic psychological needs are of higher quality, encouraging participants to develop more positive and sustained health behaviors [63]. However, older adults with MCI often choose to participate passively in training, which makes it difficult for them to achieve genuine cognitive improvement from the training. Our study further found that participants hoped to be motivated to persist in using DCT through incentive mechanisms. In China, older adults are considered a "digitally disadvantaged group," with the internet penetration rate among those over 60 years at just 11.3% [64]. For them, the busy lifestyle habits formed in the past nondigital era have solidified their lifestyles, and any activities beyond their daily routines are seen as an additional burden. Only continuous incentives can stimulate their motivation to engage persistently. This aligns with the principles of the achievement goal theory in the social psychological motivation theory, which suggests that motivations driven by external incentives tend to be performance oriented, with goal completion based on external reward mechanisms. However, fostering a mastery-oriented climate can enhance the sustainability of behaviors [63]. While short-term rewards may encourage older adults to participate in training, if the training is seen as a task rather than an opportunity to improve health, it becomes challenging to cultivate a positive motivation model and sustain participation behavior among older adults, and the effectiveness of such incentives is greatly diminished. Moreover, long-term reliance on material rewards not only increases the burden on operating organizations but also may lead to the inability of older adults to continue using DCT after the withdrawal of rewards.

In summary, the behavior of older adults with MCI in using DCT is not static but evolves dynamically over time with accumulated usage experience. Our research findings highlight that the discrepancy between expectations and reality can undermine the perceived usefulness of technology for patients with MCI, jointly reducing their satisfaction. This negative experience further suppresses their continuance intention, ultimately leading to an inability to sustain long-term usage behavior.

Implications for Future Research and Practice

Future research should thoroughly understand the expectations of older adults with MCI regarding DCT before developing training programs and should provide cognitive training formats that better meet their needs based on their postusage feedback [28,65], thereby narrowing and bridging the expectation gap. At the same time, older adults with cognitive impairment should be encouraged to gradually achieve the set goals during interventions to build confidence in sustained participation. In the training of older adults with MCI, by fully stimulating both the intrinsic and extrinsic motivational forces of these older adults, the intention to persistently engage in training can be enhanced, thus achieving a more sustained and effective cognitive training effect. Integrating qualitative research into the program development phase to dig deeper into the key variables that impede continuous participation can provide guidance for subsequent studies. This approach can help explore how to effectively facilitate the conditions and implementation pathways for sustained use. Existing research has shown that co-design procedures can not only identify barriers to accessing rehabilitation for people with dementia but also guide the development of potential solutions [66]. Future studies should involve multiple stakeholders, including older adults with MCI, their family members, and health care providers, in the co-design process of DCT programs. This will help researchers better capture and understand the actual needs and real-life use scenarios of users, thereby integrating DCT into the daily lives of older adults.

Given that older adults generally lack a correct understanding of cognitive impairment and hold a nihilistic attitude toward the existence of effective training methods, this cognitive bias further affects their proactive acceptance of DCT. Therefore, future intervention plans should enhance the scientific understanding of cognitive impairment and DCT among older adults through knowledge education and training. These efforts can help older adults form reasonable expectations about the content and format of DCT, thereby narrowing the gap between expectations and actual experiences. To effectively improve cognitive function, researchers should consider combining DCT with offline exercise training. This integrated approach represents a critical pathway for developing optimal interventions in the future. Only then can intervention programs adequately meet older adults' preferences and needs for DCT and achieve the successful implementation and sustained application of DCT in community or in-home settings.

Strengths and Limitations

In this qualitative study, we adopted the ECM-ITC for the first time to view DCT usage behavior as a dynamic evolutionary process. This model guided our data collection and analysis, providing a comprehensive explanation for the intrinsic reasons of the behavioral shift from initial acceptance to eventual discontinuous usage of DCT among older adults with MCI across different stages of the preuse and postuse periods. Another advantage of this study is the focus on older adults with MCI, placing their attitudes and behaviors regarding the inability to sustain DCT usage within the context of the entire intervention process and their social living environment.

XSL•FO
RenderX

Additionally, incorporating the perspectives of family members provides holistic, family-centered insights.

The study also has some limitations. First, despite our efforts to recruit participants of different genders, over two-thirds of the participants were female, which may be related to the higher prevalence of MCI in women [67]. In addition, we collated and analyzed data based on a predetermined theoretical framework, which might have overlooked important factors outside the ECM-ITC. Finally, although we encouraged family members to participate in the training of older adults with MCI, the sample size of family members was still relatively small. Additionally, the study was conducted in only 1 community health service center in Guangzhou, which may limit the generalizability of the study's findings. Future research could collect more data at the family level for a comprehensive understanding of this phenomenon.

Conclusions

Currently, there is extremely limited research on the discontinuous usage phase of DCT. By using the ECM-ITC, this study comprehensively gained insights from older adults with MCI and their families, identifying reasons for the transition from accepting DCT to discontinued use of DCT among adults with MCI aged 60 years or older. There are opportunities to better engage older adults in adhering to DCT in the future, but it is important to consider that cognitive training is only one part of their lives and to recognize the complex needs of older adults as well as the challenges they may face when using DCT. Future research should optimize and integrate these reasons into intervention programs, designing personalized interventions that align more closely with the needs of older adults to ensure that DCT can be genuinely integrated into the daily lives of older adults with MCI.

Acknowledgments

The authors would like to thank all the older adults with mild cognitive impairment and their families who provided invaluable insights for this study. The authors are also extremely grateful to the health care professionals and social workers at community health service centers who took time and effort out of their daily work pressures to help organize on-site activities, distribute surveys, and assist in recruiting participants.

Data Availability

For protecting participant rights and preventing the potential identification of participants through the complete interview transcripts, the data generated and analyzed during this study will not be made publicly available. However, upon reasonable request, anonymized format summaries of the interview content analysis and relevant data supporting the research findings can be obtained from the corresponding author.

Authors' Contributions

SZ, MW, and NG played crucial roles in conceptualization, methodology, recruitment, data curation, data analysis, visualization, and manuscript writing. RS was responsible for investigation, funding acquisition, and project management. CC handled investigation, data collection, and analysis. ZZ participated in the review and discussion of the study findings, providing comprehensive feedback and editing. JL and NG contributed equally to the methodology, review, editing, and supervision of the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 18 KB - jmir_v27i1e66393_app1.docx]

References

1. Ying G, Perez-Lao A, Marsiske M, Levy S, Smith GE. Multimodal cognitive and behavioral interventions for patients with MCI: a systematic review and meta-analysis on cognition and mood. Front Aging Neurosci 2024;16:1390699 [FREE Full text] [doi: 10.3389/fnagi.2024.1390699] [Medline: 38746830]
2. Rajagopal SK, Beltz AM, Hampstead BM, Polk TA. Estimating individual trajectories of structural and cognitive decline in mild cognitive impairment for early prediction of progression to dementia of the Alzheimer's type. Sci Rep 2024 Jun 05;14(1):12906 [FREE Full text] [doi: 10.1038/s41598-024-63301-7] [Medline: 38839800]
3. An R, Gao Y, Huang X, Yang Y, Yang C, Wan Q. Predictors of progression from subjective cognitive decline to objective cognitive impairment: A systematic review and meta-analysis of longitudinal studies. Int J Nurs Stud 2024 Jan;149:104629. [doi: 10.1016/j.ijnurstu.2023.104629] [Medline: 37979370]
4. No authors listed. 2023 Alzheimer's disease facts and figures. Alzheimers Dement 2023 Apr;19(4):1598-1695. [doi: 10.1002/alz.13016] [Medline: 36918389]

XSL•FO
RenderX

5. Morris JC. Mild cognitive impairment and preclinical Alzheimer's disease. Geriatrics 2005 Jun;Suppl:9-14. [Medline: 16025770]
6. Canevelli M, Grande G, Lacorte E, Quarchioni E, Cesari M, Mariani C, et al. Spontaneous reversion of mild cognitive impairment to normal cognition: a systematic review of literature and meta-analysis. J Am Med Dir Assoc 2016 Oct 01;17(10):943-948. [doi: 10.1016/j.jamda.2016.06.020] [Medline: 27502450]
7. Song W, Wu W, Zhao Y, Xu H, Chen G, Jin S, et al. Evidence from a meta-analysis and systematic review reveals the global prevalence of mild cognitive impairment. Front Aging Neurosci 2023;15:1227112 [FREE Full text] [doi: 10.3389/fnagi.2023.1227112] [Medline: 37965493]
8. Mattke S, Jun H, Chen E, Liu Y, Becker A, Wallick C. Expected and diagnosed rates of mild cognitive impairment and dementia in the U.S. Medicare population: observational analysis. Alzheimers Res Ther 2023 Jul 22;15(1):128 [FREE Full text] [doi: 10.1186/s13195-023-01272-z] [Medline: 37481563]
9. Silva AF, Silva RM, Murawska-Ciałowicz E, Zurek G, Danek N, Cialowicz M, et al. Cognitive training with older adults using smartphone and web-based applications: a scoping review. J Prev Alzheimers Dis 2024;11(3):693-700 [FREE Full text] [doi: 10.14283/jpad.2024.17] [Medline: 38706285]
10. Zuschnegg J, Schoberer D, Häussl A, Herzog SA, Russegger S, Ploder K, et al. Effectiveness of computer-based interventions for community-dwelling people with cognitive decline: a systematic review with meta-analyses. BMC Geriatr 2023 Apr 12;23(1):229 [FREE Full text] [doi: 10.1186/s12877-023-03941-y] [Medline: 37041494]
11. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
12. Gates NJ, Rutjes AW, Di Nisio M, Karim S, Chong L, March E, et al. Computerised cognitive training for 12 or more weeks for maintaining cognitive function in cognitively healthy people in late life. Cochrane Database Syst Rev 2020 Feb 27;2(2):CD012277 [FREE Full text] [doi: 10.1002/14651858.CD012277.pub3] [Medline: 32104914]
13. Chinese Society of Dementia and Cognitive Impairment, Consensus Panel on Digital Therapeutics for Cognitive Impairment. [Chinese expert consensus on digital therapeutics for cognitive impairment (2023 edition)]. Zhonghua Yi Xue Za Zhi 2023 Mar 07;103(9):640-647. [doi: 10.3760/cma.j.cn112137-20221121-02441] [Medline: 36858363]
14. Yuan TW, Zhang LW, Zhu CS, Yu JR, Ruan MH. Analysis of technological progress and industrial development trends. Therapeutics Science Focus 2023;18(01):14-26 [FREE Full text] [doi: 10.15978/j.cnki.1673-5668.202301005]
15. Bodner KA, Goldberg TE, Devanand DP, Doraiswamy PM. Advancing computerized cognitive training for MCI and Alzheimer's disease in a pandemic and post-pandemic world. Front Psychiatry 2020;11:557571 [FREE Full text] [doi: 10.3389/fpsyt.2020.557571] [Medline: 33329097]
16. Hayashi H, Sone T, Iokawa K, Sumigawa K, Fujita T, Kawamata H, et al. Effects of computerized cognitive training on biomarker responses in older adults with mild cognitive impairment: A scoping review. Health Sci Rep 2024 Jun;7(6):e2175 [FREE Full text] [doi: 10.1002/hsr2.2175] [Medline: 38895550]
17. Harvey PD, Zayas-Bazan M, Tibiriçá L, Kallestrup P, Czaja SJ. Improvements in cognitive performance with computerized training in older people with and without cognitive impairment: synergistic effects of skills-focused and cognitive-focused strategies. Am J Geriatr Psychiatry 2022 Jun;30(6):717-726 [FREE Full text] [doi: 10.1016/j.jagp.2021.11.008] [Medline: 34924275]
18. Rebok GW, Ball K, Guey LT, Jones RN, Kim H, King JW, ACTIVE Study Group. Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. J Am Geriatr Soc 2014 Jan;62(1):16-24 [FREE Full text] [doi: 10.1111/jgs.12607] [Medline: 24417410]
19. Li R, Geng J, Yang R, Ge Y, Hesketh T. Effectiveness of computerized cognitive training in delaying cognitive function decline in people with mild cognitive impairment: systematic review and meta-analysis. J Med Internet Res 2022 Oct 27;24(10):e38624 [FREE Full text] [doi: 10.2196/38624] [Medline: 36301590]
20. Nahas C, Gandit M, Quillion-Dupré L, Monfort E. How to engage patients in computerized cognitive training: a cognitive rehabilitation expert's perspective. Disabil Rehabil Assist Technol 2024 Aug;19(6):2306-2318. [doi: 10.1080/17483107.2023.2284879] [Medline: 37994038]
21. Turunen M, Hokkanen L, Bäckman L, Stigsdotter-Neely A, Hänninen T, Paajanen T, et al. Computer-based cognitive training for older adults: Determinants of adherence. PLoS One 2019;14(7):e0219541 [FREE Full text] [doi: 10.1371/journal.pone.0219541] [Medline: 31291337]
22. Coley N, Ngandu T, Lehtisalo J, Soininen H, Vellas B, Richard E, HATICE, FINGER,MAPT/DSA groups. Adherence to multidomain interventions for dementia prevention: Data from the FINGER and MAPT trials. Alzheimers Dement 2019 Jun;15(6):729-741. [doi: 10.1016/j.jalz.2019.03.005] [Medline: 31047857]
23. Bahar-Fuchs A, Barendse MEA, Bloom R, Ravona-Springer R, Heymann A, Dabush H, et al. Computerized cognitive training for older adults at higher dementia risk due to diabetes: findings from a randomized controlled trial. J Gerontol A Biol Sci Med Sci 2020 Mar 09;75(4):747-754 [FREE Full text] [doi: 10.1093/gerona/glz073] [Medline: 30868154]
24. Trenorden KI, Hull MJ, Lampit A, Greaves D, Keage HAD. Older adults' experiences of a computerised cognitive training intervention: a mixed methods study. Australian Journal of Psychology 2022 Mar 06;74(1):2036581. [doi: 10.1080/00049530.2022.2036581]

XSL•FO
RenderX

25. Mondellini M, Arlati S, Gapeyeva H, Lees K, Märitz I, Pizzagalli SL, et al. User experience during an immersive virtual reality-based cognitive task: a comparison between Estonian and Italian older adults with MCI. Sensors (Basel) 2022 Oct 27;22(21):8249 [FREE Full text] [doi: 10.3390/s22218249] [Medline: 36365947]
26. Pike K, Moller CI, Bryant C, Farrow M, Dao DP, Ellis KA. Examination of the feasibility, acceptability, and efficacy of the online personalised training in memory strategies for everyday program for older adults: single-arm pre-post trial. J Med Internet Res 2023 Apr 20;25:e41712 [FREE Full text] [doi: 10.2196/41712] [Medline: 37079356]
27. Tak ECPM, van Uffelen JGZ, Paw MJMCA, van Mechelen W, Hopman-Rock M. Adherence to exercise programs and determinants of maintenance in older adults with mild cognitive impairment. J Aging Phys Act 2012 Jan;20(1):32-46 [FREE Full text] [doi: 10.1123/japa.20.1.32] [Medline: 21945922]
28. Manser P, Adcock-Omlin M, de Bruin ED. Design considerations for an exergame-based training intervention for older adults with mild neurocognitive disorder: qualitative study including focus groups with experts and health care professionals and individual semistructured in-depth patient interviews. JMIR Serious Games 2023 Jan 05;11:e37616 [FREE Full text] [doi: 10.2196/37616] [Medline: 36602851]
29. Chen K, Lou VWQ, Lo SSC. Exploring the acceptance of tablets usage for cognitive training among older people with cognitive impairments: A mixed-methods study. Appl Ergon 2021 May;93:103381. [doi: 10.1016/j.apergo.2021.103381] [Medline: 33578065]
30. Shamir D, Loubani K, Schaham NG, Buckman Z, Rand D. Experiences of older adults with mild cognitive impairment from cognitive self-training using touchscreen tablets. Games Health J 2024 Feb;13(1):13-24. [doi: 10.1089/g4h.2023.0017] [Medline: 37768834]
31. Dequanter S, Fobelets M, Steenhout I, Gagnon M, Bourbonnais A, Rahimi S, et al. Determinants of technology adoption and continued use among cognitively impaired older adults: a qualitative study. BMC Geriatr 2022 Apr 28;22(1):376 [FREE Full text] [doi: 10.1186/s12877-022-03048-w] [Medline: 35484488]
32. Tsai YI, Beh J, Ganderton C, Pranata A. Digital interventions for healthy ageing and cognitive health in older adults: a systematic review of mixed method studies and meta-analysis. BMC Geriatr 2024 Mar 04;24(1):217 [FREE Full text] [doi: 10.1186/s12877-023-04617-3] [Medline: 38438870]
33. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia and carers: results from a feasibility randomized controlled trial (RCT). Clin Interv Aging 2021;16:2079-2094 [FREE Full text] [doi: 10.2147/CIA.S323994] [Medline: 35221680]
34. Miller E, Wilding R, Baker S, Caldwell GA, Neves BB, Waycott J. Transforming aged care with virtual reality: How organisational culture impacts technology adoption and sustained uptake. Australas J Ageing 2024 Mar;43(1):183-190. [doi: 10.1111/ajag.13248] [Medline: 37803886]
35. Hicks B, Karim A, Jones E, Burgin M, Cutler C, Tang W, et al. Care home practitioners' perceptions of the barriers and facilitators for using off-the-shelf gaming technology with people with dementia. Dementia (London) 2022 Jul;21(5):1532-1555 [FREE Full text] [doi: 10.1177/14713012221085229] [Medline: 35427198]
36. Chan ATC, Ip RTF, Tran JYS, Chan JYC, Tsoi KKF. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024 Jan 03;7(1):1 [FREE Full text] [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
37. Kebede AS, Ozolins L, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 07;24(12):e40192 [FREE Full text] [doi: 10.2196/40192] [Medline: 36477006]
38. Bhattacherjee A, Perols J, Sanford C. Information technology continuance: a theoretic extension and empirical test. Journal of Computer Information Systems 2015 Dec 11;49(1):17-26. [doi: 10.1080/08874417.2008.11645302]
39. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
40. Pérez Peña F, Cobaisse Ibáñez M, Villagrán Pradena S, Alvarado R. General aspects of the use of mixed methods for health research. Medwave 2023 Nov 06;23(10):2767. [doi: 10.5867/medwave.2023.10.2767] [Medline: 37931073]
41. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
42. Wang B, Guo Q, Chen M, Zhao Q, Zhou Y, Hong Z. The clinical characteristics of 2,789 consecutive patients in a memory clinic in China. J Clin Neurosci 2011 Nov;18(11):1473-1477. [doi: 10.1016/j.jocn.2011.05.003] [Medline: 21924913]
43. ICD-11: International Classification of Diseases 11th Revision. World Health Organization. URL: https://icd.who.int/en [accessed 2023-09-17]
44. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999 Mar;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]
45. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. Lincoln YS, Guba EG, Pilotta JJ. Naturalistic inquiry. International Journal of Intercultural Relations 1985 Jan;9(4):438-439. [doi: 10.1016/0147-1767(85)90062-8]

XSL•FO
RenderX

47. Wittmann FG, Pabst A, Zülke A, Luppa M, Oey A, Boekholt M, et al. Adherence to a lifestyle intervention - just a question of self-efficacy? Analysis of the AgeWell.de-intervention against cognitive decline. Alzheimers Res Ther 2024 Jun 22;16(1):133 [FREE Full text] [doi: 10.1186/s13195-024-01499-4] [Medline: 38909256]
48. Kang JM, Kim N, Lee SY, Woo SK, Park G, Yeon BK, et al. Effect of cognitive training in fully immersive virtual reality on visuospatial function and frontal-occipital functional connectivity in predementia: randomized controlled trial. J Med Internet Res 2021 May 06;23(5):e24526 [FREE Full text] [doi: 10.2196/24526] [Medline: 33955835]
49. Hsieh P, Lai H, Ma C, Alexander JW, Lin M. An extended expectation-confirmation model for mobile nursing information system continuance. Res Theory Nurs Pract 2016 Nov 01;30(4):282-301. [doi: 10.1891/1541-6577.30.4.282] [Medline: 28304259]
50. Tsai H, Lee Y, Ruangkanjanases A. Understanding the effects of antecedents on continuance intention to gather food safety information on websites. Front Psychol 2020;11:579322 [FREE Full text] [doi: 10.3389/fpsyg.2020.579322] [Medline: 33424693]
51. Xiang L, Li JX, Ma J, Yang C, Liao LP, Guo J. Community resident's disease awareness of mild cognitive impairment and its influencing factors. Nursing Science 2023;12(05):753-759. [doi: 10.12677/NS.2023.125106]
52. Mild Cognitive Impairment (MCI). Alzheimer's Association. URL: https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment [accessed 2024-01-03]
53. Peng YX, Gao YF, Yong JJ, She CH, Wang Y, Ju M. The current status and influencing factors of technophobia among the elderly from the perspective of digital health. Chinese Journal of Nursing 2023;58(11):1345-1351 [FREE Full text] [doi: 10.3761/j.issn.0254-1769.2023.11.010]
54. Di Lorito C, Bosco A, Rai H, Craven M, McNally D, Todd C, et al. A systematic literature review and meta-analysis on digital health interventions for people living with dementia and Mild Cognitive Impairment. Int J Geriatr Psychiatry 2022 Jun;37(6):5730 [FREE Full text] [doi: 10.1002/gps.5730] [Medline: 35588315]
55. Hyer L, Scott C, Atkinson MM, Mullen CM, Lee A, Johnson A, et al. Cognitive training program to improve working memory in older adults with MCI. Clin Gerontol 2016;39(5):410-427. [doi: 10.1080/07317115.2015.1120257] [Medline: 29471774]
56. Contreras-Somoza LM, Irazoki E, Castilla D, Botella C, Toribio-Guzmán JM, Parra-Vidales E, et al. Study on the acceptability of an ICT platform for older adults with mild cognitive impairment. J Med Syst 2020 May 25;44(7):120. [doi: 10.1007/s10916-020-01566-x] [Medline: 32451740]
57. Makmee P, Wongupparaj P. VR cognitive-based intervention for enhancing cognitive functions and well-being in older adults with mild cognitive impairment: behavioral and EEG evidence. Psychosoc Interv 2025 Jan;34(1):37-51. [doi: 10.5093/pi2025a4] [Medline: 39781014]
58. Manser P, Herold F, de Bruin ED. Components of effective exergame-based training to improve cognitive functioning in middle-aged to older adults - A systematic review and meta-analysis. Ageing Res Rev 2024 Aug;99:102385. [doi: 10.1016/j.arr.2024.102385] [Medline: 38914262]
59. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
60. Reinhardt G, Schwarz PE, Harst L. Non-use of telemedicine: A scoping review. Health Informatics J 2021 Oct 26;27(4):14604582211043147 [FREE Full text] [doi: 10.1177/14604582211043147] [Medline: 34696613]
61. Almulhem JA. Factors, barriers, and recommendations related to mobile health acceptance among the elderly in Saudi Arabia: a qualitative study. Healthcare (Basel) 2023 Nov 23;11(23):3024. [doi: 10.3390/healthcare11233024] [Medline: 38063592]
62. Choy MA, O'Brien K, Barnes K, Sturgiss EA, Rieger E, Douglas K. Evaluating the digital health experience for patients in primary care: mixed methods study. J Med Internet Res 2024 Apr 11;26:e50410. [doi: 10.2196/50410] [Medline: 38602768]
63. Hagger M, Chatzisarantis N. Self-determination Theory and the psychology of exercise. International Review of Sport and Exercise Psychology 2008 Mar;1(1):79-103. [doi: 10.1080/17509840701827437]
64. Cui Y, He Y, Xu X, Zhou L, Nutakor JA, Zhao L. Cultural capital, the digital divide, and the health of older adults: a moderated mediation effect test. BMC Public Health 2024 Jan 25;24(1):302. [doi: 10.1186/s12889-024-17831-4] [Medline: 38273305]
65. Manser P, de Bruin ED. Making the best out of IT: design and development of exergames for older adults with mild neurocognitive disorder - a methodological paper. Front Aging Neurosci 2021;13:734012. [doi: 10.3389/fnagi.2021.734012] [Medline: 34955806]
66. Layton N, Devanny C, Hill K, Swaffer K, Russell G, Low L, et al. The right to rehabilitation for people with dementia: a codesign approach to barriers and solutions. Health Expect 2024 Oct;27(5):e70036. [doi: 10.1111/hex.70036] [Medline: 39318228]
67. Liu Y, Yu X, Han P, Chen X, Wang F, Lian X, et al. Gender-specific prevalence and risk factors of mild cognitive impairment among older adults in Chongming, Shanghai, China. Front Aging Neurosci 2022;14:900523. [doi: 10.3389/fnagi.2022.900523] [Medline: 36118698]

XSL•FO
RenderX

Abbreviations

DCT: digital cognitive training
ECM-ITC: extended model of IT continuance
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination

Edited by T de Azevedo Cardoso; submitted 13.09.24; peer-reviewed by P Manser, W Xiao; comments to author 08.01.25; revised version received 28.01.25; accepted 27.02.25; published 25.03.25.

Please cite as:
Zhang S, Wu M, Sun R, Cui C, Zhang Z, Liao J, Gong N
Exploring the Discontinuous Usage Behavior of Digital Cognitive Training Among Older Adults With Mild Cognitive Impairment and Their Family Members: Qualitative Study Using the Extended Model of IT Continuance
J Med Internet Res 2025;27:e66393
URL: https://www.jmir.org/2025/1/e66393
doi:10.2196/66393
PMID:

©Shangyang Zhang, Min Wu, Ruini Sun, Changjie Cui, Ziqing Zhang, Jing Liao, Ni Gong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating a Digital Health Tool Designed to Improve Low Sexual Desire in Women: Mixed-Methods Implementation Science Study

Lori A Brotto[1], PhD; Kyle R Stephenson[2], PhD; Nisha Marshall[1], MSc; Mariia Balvan[3], BA; Yaroslava Okara[3], BA; Elizabeth A Mahar[4], PhD

[1]Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada

[2]School of Psychology, Xavier University, Cincinnati, OH, United States

[3]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

[4]Department of Psychology, State University of New York at Fredonia, Fredonia, NY, United States

Corresponding Author:
Lori A Brotto, PhD
Department of Obstetrics and Gynaecology
University of British Columbia
2775 Laurel Street
Department of Gynaecology
Vancouver, BC, V5Z1M9
Canada
Phone: 1 6048754111 ext 68898
Email: lori.brotto@ubc.ca

Abstract

Background: Sexual health difficulties affect up to 30% of women, with desire and arousal problems being the most prevalent. While cognitive behavioral therapy and mindfulness-based therapy are effective treatments, access is limited by barriers such as specialist shortages, cost, and embarrassment. Web-based interventions offer a potential solution by providing self-paced, cost-effective treatments. eSense, a digital health program, offers cognitive behavioral therapy and mindfulness-based therapy skills targeted to women with low sexual desire, and previous trials find eSense to be highly feasible and efficacious.

Objective: The goal of the present implementation science study was to use the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance of Implementation) framework to assess the integration of eSense into several sexual health clinics. We chose the RE-AIM framework because it addresses both dissemination (eg, reach) and implementation of an intervention.

Methods: A total of 14 specialty clinics participated, and we report on the reported experiences of those clinics in implementing eSense. We also examined responses from 12 women on waitlists to receive sex therapy or sexual medicine care.

Results: Per clinic outcomes, all aspects of implementation (reach, effectiveness, adoption, implementation, and maintenance) were in the moderate to high range for clinics, reporting that offering eSense helped them overcome negative feelings associated with their long clinic waitlists. The majority expressed a need for eSense and could see how it overcame the limitations of traditional therapy. Nearly all expressed a wish to continue offering eSense to patients after the implementation study was complete. One caveat was that half of the clinics noted cost as a key issue for future implementation, and one-third noted that the administrative burden of implementing eSense as a standard of care may be challenging. For individual users, the majority expressed an interest in knowing more about eSense and a desire to use eSense, though most of these did not complete the program in its entirety. Users experienced a significant improvement in sex-related distress with no clinically meaningful change in other outcomes and a high level of satisfaction with eSense. Most also reported doing things differently in their sexual lives after participating in eSense.

Conclusions: We found that eSense demonstrates potential as a digital intervention for sexual difficulties for women, particularly concerning its moderate implementation outcomes and also because of its ability to reduce sexual distress. Future studies should address the barriers identified for broader adoption of eSense in clinical settings.

Trial Registration: ClinicalTrials.gov NCT05168371; https://clinicaltrials.gov/study/NCT05168371

(J Med Internet Res 2025;27:e69828) doi:10.2196/69828

XSL•FO
RenderX

KEYWORDS

implementation science; sexual interest/arousal disorder (SIAD); sexuality; internet interventions; online therapy; telehealth; online interventions; web-based therapeutic programs/interventions; online CBT/MBT treatment; female sexual dysfunction; eHealth

Introduction

Background

Difficulties with sexual function, including problems concerning orgasm, genital pain, desire, and arousal, are common and affect up to 30% of women [1]. Difficulties with desire and arousal are the most common [2] and are often triggered by a combination of biological, psychological, and sociocultural factors [3]. Difficulties with sexual arousal and desire can be distressing and lead to negative physical, emotional, and interpersonal outcomes, such as depressive symptoms and relational conflict [4-7].

Face-to-face cognitive behavioral therapy (CBT) and mindfulness-based therapy (MBT) are considered efficacious gold-standard treatments for desire and arousal difficulties in women [8-12]. CBT is change-based and aims to alter maladaptive thoughts and behaviors that impede healthy sexual response [13-15] and build new beliefs and behaviors that refocus attention to erotic cues and pleasure [16,17]. MBT is acceptance-based and encourages present-focused, nonjudgmental attention to body sensations [18,19]. Instead of changing specific thoughts and behaviors, MBT trains an individual to build interoception (ie, awareness of sensations in the body) and positive sexual awareness [20-22].

Barriers to Accessing Treatment

Despite the prevalence of sexual problems and the existence of these evidence-based psychological approaches, there are system-level barriers to accessing help. These barriers include a lack of specialist providers, long waiting lists, high costs for private services, and discomfort talking about sex with professionals [23,24]. Another factor that might impact the willingness to seek help for sexual problems is the lack of awareness of available and evidence-based treatments [25,26].

One possible solution to these barriers is the use of web-based interventions. Here, we refer to “web-based interventions” as treatments that are fully digital rather than simply using internet technologies such as videoconferencing to support traditional face-to-face methods. These interventions are usually delivered online with varying degrees of synchronous or asynchronous guidance [27]. These interventions use digital technologies for screening, health promotion, prevention, early intervention, treatment, or relapse prevention [28].

Internet-based psychological interventions have been found to be as efficacious as face-to-face treatments in many cases [29]. Some advantages of web-based interventions are that they can be delivered at a self-determined pace and intensity, can maintain privacy or anonymity, and may be more cost-effective, potentially offering a more economically viable option for patients [30-33]. Additionally, because web-based psychological interventions are housed on accessible digital platforms, they are uniquely situated in their potential for scalability and widespread use. Of a systematic review investigating the quality of digital health interventions and among their list of 5 recommendations needed to establish that a digital tool is useful, one of them was that “the value delivered by digital health solutions should consider multiple dimensions such as clinical, organizational, behavioral, and technical” [34]. This includes ensuring that the intervention has benefits beyond just the clinical outcomes and benefits of the “system” in which treatments are delivered. As such, there is a need to evaluate the scalability of such interventions to determine their potential, including some of these multidimensional impacts on clinics that deliver care.

Implementation of Web-Based Psychological Treatments

Implementation science involves expanding health care interventions beyond the controlled environment of evaluation studies into real-world contexts [35]. Due to randomized controlled trials often being criticized for limited generalizability [36], there is a growing need for implementation science to see how well effective treatments work when taken out of the strictly controlled research context and are delivered at scale. An interest in scaling up web-based psychological treatments has significantly increased in recent years [37] and has been well-received by many users of these services. For example, Guinart et al [38] found that 82% of patients who partook in a telepsychiatry appointment in the United States rated the overall experience as good or excellent, and 64% indicated they would consider using remote treatment sessions instead of face-to-face sessions in the future. Despite this growing interest, there has been limited research on the implementation of web-based psychological treatments related to sexual health concerns.

One of the few examples of an implementation science investigation in the domain of sexual health was the testing of a web-based program for sexual dysfunction in a Dutch public health setting. The study aimed to evaluate the implementation of a free, anonymous web-based education and self-training program for adolescents and young adults aged 16-24 years. The researchers were able to recruit over 1000 participants in the study but reported a significant dropout rate of 50%, with a primary contributor to the high attrition being the program’s substantial length [39]. The findings of studies such as this provide valuable insights into the facilitators and barriers to the implementation of digital sexual health interventions that we can integrate into future digital sexual health therapeutics. However, to our knowledge, there are no implementation science studies that have evaluated digital sexual health interventions specifically for adult women experiencing sexual difficulties.

eSense: a Digital Sexual Health Tool

eSense is a web-based platform delivering psychological treatment of sexual health concerns that was originally designed to address sexual interest/arousal disorder [40] in females (though more recently it has been adapted to other populations

XSL•FO
RenderX

of females with sexual concerns). This online intervention was created through a multidisciplinary collaboration among clinicians and researchers with expertise in sexual dysfunction, patient partners, web designers, and graphic illustrators. Individuals experiencing sexual difficulties often experience difficulty in accessing care due to geographic distance from qualified health care providers, cost, lack of health care providers specializing in sexual health, and emotional barriers (eg, anxiety, shame, or embarrassment) [41]; therefore, eSense may overcome these gaps and allow women with sexual interest/arousal disorder to access evidence-based care in the privacy of their homes. eSense is built on a WordPress PHP site (with Elementor plug-in) that is HIPAA (Health Insurance Portability and Accountability Act) compliant and uses a cloud server (Digital Ocean) based in Canada. It offers 2 distinct programs: CBT and MBT. Each program or "arm" consists of 8 modules that provide educational content, detailed instructions for therapeutic activities, fictional case examples to illustrate these activities, homework exercises, and troubleshooting tips. The intervention is delivered through a combination of text, video, images, and audio, all carefully designed to be highly engaging for users. Users have 2 options for completing the program: independently or with individual support from nonexpert "navigators" who offer empathetic listening, encouragement, and technology assistance but do not provide formal therapy [42]. Administrators from the research team create a unique log-in for eSense that is provided to an individual user following informed consent. Each user then has a unique database record associated with their account, and their engagement with eSense is then saved to their personal database for future analytics.

In a series of studies, eSense has been demonstrated to be usable and satisfactory, even without individualized guidance [35]. In a randomized trial evaluating eSense along with navigator support, participants reported significant and clinically meaningful improvements in sexual desire, sexual distress, sexual satisfaction, and overall sexual function compared to a control condition, with no loss in gains at the 6-month follow-up [42]. Taken together, while these studies found eSense to be feasible, satisfying, and efficacious when evaluated in the context of a controlled trial, its real-world implementation outcomes remain unknown, and there is a need to directly study unique factors associated with its implementation [43].

Reach, Effectiveness, Adoption, Implementation, and Maintenance of Implementation Framework

Several implementation science frameworks facilitate the deployment of interventions at scale. Notable among these are the PRECEDE-PROCEED model, the Practical, Robust Implementation and Sustainability Model, the Consolidated Framework for Implementation Research, and the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance of Implementation) framework. The RE-AIM framework, in particular, is a widely recognized model in implementation science that covers five intervention-related areas of impact [44]: (1) reach as the proportion of those in need who access the intervention, (2) effectiveness per the impact of interventions on health outcomes, (3) adoption as a decision to proceed with implementing the clinical intervention, (4) implementation as the process of embedding and integration of the intervention in routine practice and its consistency of delivery and costs, and (5) maintenance as the institutionalization of the intervention in routine care [45-47]. This framework is considered more applicable than many other implementation science models, as it addresses both dissemination (eg, reach) and implementation aspects [48]. The RE-AIM framework provides a heuristic tool for bridging interventions' internal validity established in well-controlled conditions and their external validity in real-world conditions [46,49]. It is designed to evaluate the public health impact of health-promoting interventions, and it is widely used in implementation research [44]. Glasgow et al [46,50] emphasize its utility in evaluating interventions that tackle multiple underlying factors and consider entire systems, and as such, may be ideal for evaluating the implementation of sexual health interventions. Of note, the hyphen in RE-AIM is intended to differentiate the individual-level factors (reach and effectiveness) from the organization-level factors (implementation and maintenance) [47]. One systematic review evaluating the RE-AIM framework concluded that all 5 domains should be reported on accurately and transparently to facilitate translation into real-world settings [47].

The overall goal of this study was to conduct an implementation science evaluation using the RE-AIM framework of the enablers and barriers to (1) patients using eSense when disseminated through sexual health clinics and (2) sexual health clinics implementing eSense into their existing structures. Due to the nonlinear, exploratory, and dynamic nature of implementation science research [35], we did not formulate a priori hypotheses. The findings were intended to also plan future implementation science studies with eSense once we implemented the key learnings about enhancing its implementation.

Methods

Research Design

We invited clinics in Canada and the United States to offer eSense to women who could not immediately receive treatment (ie, they were waitlisted at the clinic or they declined treatment at the clinic due to factors such as financial barriers). Using the components of RE-AIM [46], we sought to evaluate the impact of eSense by measuring its reach, effectiveness, adoption, implementation, and maintenance at the level of clinics as well as patients.

Participants (Clinics and Patients)

Using the professional network of the first author, we identified several sexual medicine and sex therapy clinics in Canada and the United States and wrote to the clinic leads to inform them about this study. We used a combination of existing society membership listserves and word of mouth, as well as general internet searches to identify potential clinics that might participate. Those interested scheduled an information call with a research team member who explained this study and determined whether the clinic was eligible to participate and had the interest and capacity to refer women to this study. The eligibility criteria for the clinics were that they offered sexual health services and currently had a waitlist.

XSL•FO
RenderX

A point person at each participating clinic was provided with inclusion criteria for this study's participants (ie, patients seeking care in their clinic). Women of any sexual orientation were eligible if they were aged 19 years or older, fluent in English, had no visual impairments, were computer literate with access to an internet-capable computer, had a history of sexual activity, and met telephone screening criteria for sexual concerns regarding low desire or arousal and associated distress.

Procedure

Recruitment began in May 2022, and patients were referred to this study until December 2023. Each participating clinic invited females (either cis or trans women) who were seeking care in their clinic and were currently on their waitlist to consider participating in this study if they met the inclusion criteria (described above). Clinics collected basic information (ie, the reason for seeking treatment at the clinic, demographics, interest in participating in eSense) from all women who were contacted about participation in this study, regardless of whether those women went on to use eSense or not. This basic information was collected to compare participants versus nonparticipants of eSense. The women who participated were provided with the contact information for a member of the research team directly by the clinic staff person.

After consenting to this study, participants received a short description of each of the 2 arms of eSense (CBT and MBT) and chose one. If they did not have a preference, they were randomly assigned to an arm by this study's coordinator.

Because prior research indicates that individualized support—even support that provides no formal therapy [51,52]—decreases attrition, increases adherence, and improves outcomes [53,54], participants had the option to have individualized support sessions with "navigators." Navigators were undergraduate students who received training to encourage participants to engage in the treatment, and they answered practical questions [55]. Navigators were trained by one of the lead researchers on the eSense team. They were given a set reading list, which provided relevant background literature and provided with a standardized protocol and adherence documents for training. They then completed training sessions on the technical aspects of the role (eg, scheduling meetings) and on this study's protocol and participated in an active listening skills workshop led by the researcher. Following this training, they participated in weekly group meetings and biweekly one-on-one meetings with this study's coordinator. Participants could meet with a navigator as often as they wished (up to a maximum of once per week).

Participants filled out a baseline questionnaire using the online program Qualtrics and subsequently received access to the eSense website to work through one of the arms of the program at their own pace. Those who expressed interest in a navigator received a link to an online calendar that they used to schedule navigator meetings. Participants were sent their posttreatment questionnaire 1 week after their intended treatment end date (16 weeks from the date the participant first logged into eSense). At the end of this study, clinic directors and assisting clinic staff were also sent a questionnaire to access clinic-level implementation outcomes.

Because this study used an implementation approach with descriptive primary outcomes (eg, the proportion of women who were told about eSense and contacted us to express interest in participating and the proportion of women requesting navigator support), no power analyses were performed.

Measures

Using the RE-AIM framework [46], we assessed the following outcomes:

Clinic Level

1. Reach: the extent to which clinics were willing to participate in the implementation of eSense. Assessed by determining the proportion of clinics that were approached, expressed interest, and participated, including the reasons for their decisions.
2. Effectiveness: the extent to which clinics perceived eSense as impactful for their patients or the clinic. Assessed by gathering feedback from clinic staff on the effectiveness of eSense for their practice, including any observed benefits, as well as by determining whether user-level outcomes (eg, who was interested in eSense and who was not, participant satisfaction with treatment, or participant changes in sexual distress, interest, and desire) were meaningful to clinic partners (eg, does this information help them make decisions or achieve their goals).
3. Adoption: the extent to which clinics perceived the ease of adoption of eSense into their clinic. Assessed by determining if clinics had costs to support the maintenance of eSense and asking clinic staff why they chose to participate in this study using open-ended questions at the end of this study.
4. Implementation: the extent to which eSense was feasible to implement into their clinic. Assessed by determining the time that was required from research and clinic staff, resources required (if any), descriptions of adaptations, if any, made during this study (eg, changing the phone script at a specific clinic), reasons for adaptations (eg, to reduce costs or to have more people participate), and suggestions for improvements of the implementation procedure. Assessed by determining if clinics had costs to support the maintenance of eSense and asking clinic staff why they chose to participate in this study using open-ended questions at the end of this study.
5. Maintenance: the extent to which clinics perceived implementing eSense in their clinics as maintainable. Assessed by determining whether the organization would like to keep offering eSense, organizational barriers to offering eSense, reasons for continuance or discontinuance, and whether the clinic was willing and able to pay for potential costs associated with distributing eSense.

Clinic feedback forms also included options to provide free responses to any of the questions that also requested quantitative data.

Participant Measures

1. Reach: the extent to which participants were willing to participate in eSense. Assessed by determining the

XSL•FO
RenderX

characteristics of those who participated in the intervention versus those who did not and why.
2. Effectiveness: the extent to which participants benefited from eSense. Assessed by examining changes from baseline to posttreatment in (1) sexual distress, measured by the Female Sexual Distress Scale-Revised [56] and (2) sexual desire, measured by the Sexual Interest and Desire Inventory [57]. As recommended by RE-AIM [46], we measured change in satisfaction with life [58] and satisfaction with sex life. We also assessed satisfaction with treatment by an adapted version of the Erectile Dysfunction Inventory of Treatment Satisfaction Scale [59] and attrition rates.
3. Adoption: the extent to which participants were willing to start eSense. Assessed by determining hesitations participants had before joining this study and starting eSense and if participants were engaged in any other treatment while using eSense.
4. Implementation: the extent to which participants engaged with eSense. Assessed by determining how participants used the intervention, including the average number of modules completed and reported barriers to engaging with eSense.
5. Maintenance: the extent to which participants perceived the long-term effects of eSense. Assessed by determining whether participants planned to continue practicing eSense homework in the future.

Additionally, we examined the demand for navigator support, preference for MBT versus CBT, changes from baseline to posttreatment in motivation to see a health care provider for their sexual concerns, romantic partner involvement, and whether the participants dropped from the clinic waitlists after eSense participation.

Analysis Plan

For our quantitative effectiveness outcomes, we used a paired samples *t* test with 1 within-subject factor of time (treatment) and 2 measurement points (baseline and posttreatment) and included effect sizes, as measured by Cohen *d*. We analyzed the qualitative data descriptively with a focus on key themes.

Ethical Considerations

This study was approved by the Behavioural Research Ethics Board at the University of British Columbia (certificate #H22-02993), and all participants provided written informed consent. All clinic and individual participant data were deidentified before analyses, and all data were analyzed as a group. Although participants were not compensated for participation, clinic-level administrative staff who assisted with data collection were compensated for their time at the hourly rate paid for by their clinics.

Results

Clinic Results

Reach

Of the 194 clinics contacted, 178 did not respond, and 16 wanted to learn more about this study. Of these, 14 (7.2% of the total contacted) clinics agreed to participate after learning more about the details and requirements of this study.

Most clinics involved in this study provided psychological treatment of sexual difficulties (n=13, 92.9%), with fewer offering physical therapy or pharmacological treatment (n=4, 28.6%). A total of 13 of 14 clinics specialized in the treatment of sexual difficulties. In free-response questions at the end of participation, clinics indicated that the most common reason for referring women to this study was because they presented with low desire, arousal issues, or painful intercourse. Clinics also reported that they most commonly referred younger women and women with financial constraints.

Effectiveness

Per effectiveness, the outcomes collected in this study were meaningful to 71.4% (n=10) of clinics. Clinics reported that this study provided valuable support in decision-making, particularly in determining whether referring clients to similar studies would be beneficial. Additionally, the information collected helped shape organizational decisions such as business plan adjustments, next steps for possibly embedding eSense into their clinic, guiding offerings to clients, and improving the services. Some remarked:

> *Super important information that can help direct our business plan, next steps, guide offerings to clients.*
>
> *Would be interested in all that information. That would be helpful in deciding how to utilize the program.*

Per clinics' motivation to participate in this study, most reported their desire to support research, to provide additional client support for those on the waitlist, and to help explore the use of eSense in clinic operations. Several clinics mentioned that being able to offer eSense to waitlist patients alleviated some of the provider guilt associated with extensive waitlist times.

Adoption

Responses regarding adoption focused largely on budget considerations. Regarding the budget for maintaining eSense, 50% (n=7) of clinics stated that their ability to offer eSense to patients depended on its cost, while 28.6% (n=4) indicated they could not support the maintenance costs. One of the clinics stated:

> *We wouldn't be able to afford supporting a cost for clients without getting some reimbursement from clients. If clients were choosing to enroll, and paying for the program directly, we could absolutely support it. We do not have the funds to sponsor it for our clients.* [Sexual medicine center]

The majority, 78.6% (n=11) of clinics, expressed a need for eSense in their clinic, citing the limitations of traditional therapy sessions and the importance of providing clients with additional support for practicing therapeutic tools. One remarked:

> *Therapy is often the space we are processing emotional content. It is nice to provide options for clients who want to have support practicing tools.*

XSL•FO
RenderX

Implementation

Regarding implementation, feedback from clinics revealed a strong interest in continuing to offer eSense after this study, with 92.9% (n=13) of clinics indicating their willingness to do so. The primary reasons for this interest included the program's ability to complement traditional 1-hour therapy sessions by providing additional tools and resources, thereby offering clients more support options. Clinics also highlighted the perceived value of eSense as a resource for women who lack access to sex therapy, noting its reliability and creation by experts in the field. Some of the clinic responses included:

> *Only so much can be covered in a typical one-hour therapy session. Providing clients with tools and resources is of utmost importance to me and my practice.*
>
> *It's a very valuable resource, and much needed! Plus it is offered by the experts in the field, so it is a reliable source.*
>
> *Another option for women to do in their own time and in the comfort of their own home. Seems like a great option for clients who either can't afford therapy, or who are stuck on a wait list.*

However, 35.7% (n=5) identified barriers to continuing eSense, primarily related to resource intensiveness (eg, additional time required from clinic intake coordinators who would refer women to the program and costs associated with this). Additionally, although clinics were required to have a client waitlist to be eligible for this study, many clinics did not end up having waitlists at their clinics, which reduced their immediate need for eSense. Despite this, they acknowledged that eSense would be valuable when immediate treatment was not possible. There were also challenges in keeping triage and care management staff updated about the program, which would require additional training to keep staff informed about eSense. As 1 clinic remarked:

> *The only barrier is that we've been able to offer timely treatment to potential clients, so clients did not have a need to take part in eSense. However, we realize timely treatment may not always be possible. So it's great to have eSense as an option.*

Maintenance

Regarding maintenance, 92.9% (n=13) of clinics believed that eSense could be successfully integrated into existing clinic operations. Clinics suggested various methods for integration, including using eSense as a support outside of the office settings, assisting with take-home practices and self-paced work, and including eSense as part of a stepped-care model. They also saw the potential for eSense to be an adjunct to ongoing therapy and a resource for current patients, in addition to patients placed on the waitlist. Clinics indicated the possibility of being a referral platform by adding eSense to their resource information lists and websites. In response to asking how eSense could be integrated into the framework of their organization, some of the clinics remarked:

> *It would be part of a stepped-care model that all clients are assessed for.*
>
> *Patients receive a variety of resource information and eSense would be added to this list. Providers can continue to speak about its benefits.*
>
> *I'm not sure yet. But one thing that might be possible is that we could offer a link to eSense on our website in the areas about desire and arousal concerns -- we have a section for each concern with ideas for resources to try on one's own. As long at [sic] the tool is not costly, we could also offer it as we have been for people who choose not to enroll in sex therapy due to cost or other factors.*

Participant Results

Reach

Regarding how well we were able to reach patients, 14 partner clinics referred a total of 106 women to our study, and of those, 95 provided consent to take part in the intake survey and fully or partially filled it in (Figure 1).

XSL•FO
RenderX

Figure 1. Diagram of clinic and participant flow from referral through all stages of participation.

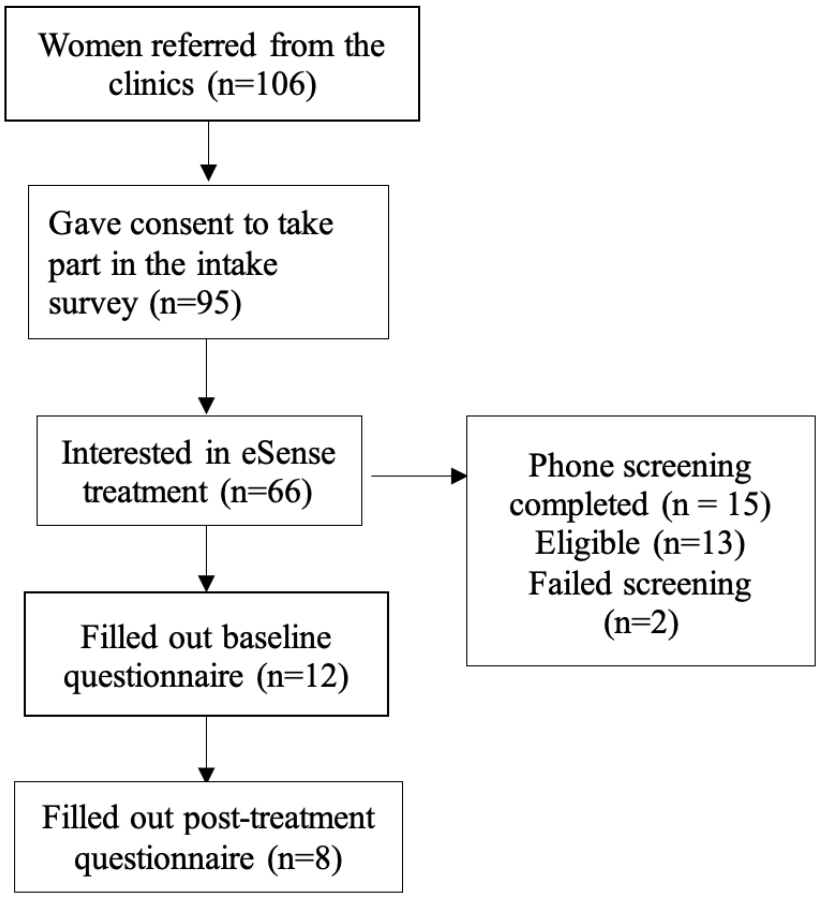

Of the participants who reported why they were seeking treatment at the participating clinic (n=81), nearly half (n=40, 49.4%) indicated that this was for a sexual desire or arousal concern, while the remaining 50.6% (n=41) were seeking help for various other issues such as mood and anxiety, sexual trauma, menopause symptoms, hormone imbalances, and other health concerns.

Among those seeking treatment at the clinic, 39 (55.7% of 80 responses) participants were informed they would be placed on a waitlist for treatment at that clinic. A total of 51 (79.7%) planned to proceed with treatment from their referral clinic, whereas 13 (20.3%) decided to no longer seek treatment from the clinic. The reasons for not seeking treatment included wait times, the transition to primarily online counseling, financial constraints, and insurance coverage issues.

After being told about eSense, 66 (82.5%) women expressed interest in learning more about it, while 14 (17.5%) were not interested in using eSense (Figure 1). Reasons for lack of interest included a perceived irrelevance to their situation, lack of time, preference for in-person interactions, not understanding the treatment, and reporting being already satisfied with their current sex life.

Of the 66 who initially expressed an interest in this study, only 12 went on to participate in this study, and all identified as cisgender (Table 1). Those who decided to participate provided various motivations, such as ongoing issues with arousal difficulties, low desire, and sexual anxiety, which had made participating in sex stressful and guilt-inducing. Other reasons for participating included severe body image issues, histories of eating disorders, and the resulting body shame that inhibited their sexual lives. Others expressed distress over a lack of physical attraction to their primary partners despite emotional connections, leading to relationship stress. A common theme was a marked decrease in libido and a desire to explore new concepts of attraction and improve their sexual lives.

Per reasons for the high rate of attrition, we compared participants to nonparticipants on select demographic variables that were collected among those who declined participation. The 2 groups did not differ in age (P=.91), race (P=.80), relationship status (P=.41), or sexual orientation (P=.054).

XSL•FO
RenderX

Table 1. Participant demographic information at baseline (N=12).

Characteristics	Values
Age (years), mean (SD)	40 (11)
Gender identity, n (%)	
Woman	12 (100)
Sexual orientation, n (%)	
Bisexual	4 (33.3)
Heterosexual	7 (58.3)
Pansexual	1 (8.3)
Racial identity, n (%)	
Racialized, person of color, visible minority, or non-White	1 (8.3)
Mixed or biracial	1 (8.3)
Mix-White or non-White	1 (8.3)
White, European descent	9 (75)
Education, n (%)	
Attended some college or university	1 (8.3)
Graduated 2-year college or university	2 (16.7)
Graduated 4-year college or university	3 (25)
Postgraduate degree	6 (50)
Annual household income (in CAD $[a]), n (%)	
$40,000 to $59,999	1 (8.3)
$80,000 to $99,999	3 (25)
$100,000 to $159,999	3 (25)
$200,000 to $239,999	3 (25)
More than $300,000	1 (8.3)
Prefer not to answer	1 (8.3)
Employment status, n (%)	
Full time	8 (66.7)
Part-time	1 (8.3)
On disability	1 (8.3)
Freelance	1 (8.3)
Other	1 (8.3)
Religious affiliation, n (%)	
Christian-Protestant	3 (27.3)
No religious affiliation	8 (72.7)

[a]A currency exchange rate of CAD $1=US $1.45 was applicable.

Effectiveness

Data for effectiveness are presented in Table 2 and is based on 8 participants who provided full pre and post data. A series of paired samples *t* tests evaluated the effects of eSense on clinical outcomes. Sexual desire (as measured by the Sexual Interest and Desire Inventory) did not significantly differ from pretreatment to posttreatment (*P*=.94), with a negligible effect size (Cohen *d*=0.029), indicating no meaningful change in sexual desire after eSense. The total Female Sexual Distress Scale-Revised score, however, decreased significantly from pretreatment to posttreatment (*P*=.004), with a large effect size (*d*=–1.459), reflecting a substantial reduction in sexual distress. Satisfaction with life did not significantly change from pre- to posttreatment (*P*=.31), with a small effect size (*d*=–0.388). Similarly, satisfaction with sex life also did not significantly change from pre- to posttreatment (*P*=.31), with a small effect size (*d*=–0.388).

XSL•FO
RenderX

Table 2. Pretreatment and posttreatment measures of sexual desire (SIDI[a]), sexual distress (FSDS-R[b]), satisfaction with life (SWLS[c]), and satisfaction with sex life (SWSL[d]). Data presented are means and SDs based on 8 participants who provided full data.

Variable	Pretreatment, mean (SD)	Posttreatment, mean (SD)	Cohen *d*	*P* value
Sexual desire (SIDI)	22.29 (8.98)	22.43 (9.4)	0.029	.94
Sexual distress (FSDS-R)	32 (10.74)	24.88 (10.97)	–1.459	.004
Satisfaction with life (SWLS)	28.63 (5.88)	27.25 (5.01)	–0.388	.31
Satisfaction with sex life (SWSL)	22.9 (4.7)	21.8 (4.01)	–0.388	.31

[a]SIDI: Sexual Interest and Desire Inventory.

[b]FSDS-R: Female Sexual Distress Scale-Revised.

[c]SWLS: Satisfaction With Life Scale.

[d]SWSL: Satisfaction With Sex Life Scale.

Participants' satisfaction with eSense was evaluated in various ways. We measured satisfaction with the experience of having versus not having a treatment navigator. Participants were evenly divided in their initial choice to have the support of a treatment navigator before starting eSense, with 4 choosing to have a navigator and 4 opting not to. Across those who chose to work with a navigator, the total number of navigator sessions was 12. Those who opted for the support of a treatment navigator expressed a range of opinions. Some appreciated having someone to discuss their thoughts and feelings with, which helped them process information and feel validated. Others felt that scheduling meetings and clarity of the navigator's role could have been better. Overall, the mean satisfaction score regarding participants' choice to have a navigator corresponded with "somewhat satisfied." Some felt they would have been more dedicated to the program with a navigator's support but believed they could manage on their own due to their previous therapy experiences. Others preferred not to discuss their concerns with anyone who was not a fully licensed therapist.

Participants provided some feedback on barriers that arose that prevented them from fully engaging with eSense. These included no time due to their job and taking care of children, lack of privacy, avoiding the exercises, and feeling frustrated that a partner was not involved in trying to improve the sexual issues.

The overall mean Erectile Dysfunction Inventory of Treatment Satisfaction Scale score was 73.64 (SD 14.85) of 100, suggesting a high level of satisfaction with eSense.

Adoption

One-third of participants (4 of 12) were initially hesitant to join this study, mainly due to concerns about the time commitment required. The average wait time for participants before they were likely to see a provider was 17 weeks, and many of them agreed to participate in eSense to address their concerns on the recommendation of health care providers or as a proactive step before starting therapy. Most participants were not engaged in any other treatment at the time, although a few mentioned ongoing treatments for other health issues (eg, thyroid medication or pelvic floor physiotherapy).

Implementation

When given the choice of which arm to use, 6 initially requested CBT, and one of these requested a change to mindfulness after starting. Further, 5 participants requested mindfulness from the outset, and 1 participant did not express a preference. Of the 12 who started eSense, 8 completed posttreatment measures, with only 37.5% (3/8) of these completing all 8 modules. The average number of modules completed was 4.3 (of 8). Engagement with homework activities also varied: 25% (n=2) did none, 12.5% (n=1) did a few, 37.5% (n=3) completed about half, and 25% (n=2) engaged with all the homework activities. Barriers to consistent homework engagement were reported by 87.5% (n=7) of participants.

Regarding barriers to engaging with eSense, participants cited various personal and situational challenges. These included changes in job responsibilities, medical issues such as low iron causing fatigue, lack of time and privacy, mental blocks, and frustration from the perception that they were attempting to address a problem primarily viewed as significant by their partner rather than by themselves. Despite these challenges, a majority (87.5%) of participants reported doing things differently in their sexual lives as a result of eSense. These changes included letting go of guilt, practicing mindfulness, improving self-talk, and increasing communication in the bedroom. Some participants also noted changes in the nonsexual aspects of their lives, such as increased mindfulness and attunement to pleasure.

Maintenance

All participants indicated they would continue to use what they learned from eSense in the future, such as the new mindfulness practices, journaling, and the use of "sexual aids" (ie, vibrators, erotica, and fantasy). Participants highlighted the importance of awareness of thought patterns and cognitive biases, with some intending to reflect on these and try behavioral experiments to overcome them. Even those who did not progress far into the material found value in the initial modules and planned to incorporate those learnings into their lives.

Additional Findings

Participants reported increased awareness and a sense of empowerment regarding their sexual health. However, suggestions for improving the program included the addition of more videos, more time between modules, greater clarity on certain instructions, and more flexible options for engaging with a navigator.

XSL•FO
RenderX

Discussion

Principal Findings

This study aimed to evaluate the implementation of eSense, a digital therapeutic intervention including CBT and MBT, for women experiencing sexual difficulties and seeking care from specialty sexual health clinics. Using the RE-AIM framework, we assessed eSense's reach, effectiveness, adoption, implementation, and maintenance across clinic and participant levels. Overall, eSense was associated with a meaningful reduction in sexual distress among participants, with moderate adoption by clinics and high satisfaction from both clinics and participants, despite some limitations in engagement and completion rates.

At the clinic level, a significant challenge was reaching clinics that would be willing to adopt eSense. Although 194 clinics were contacted, only 14 ultimately participated, and many of the nonparticipating clinics indicated that their waitlist was very short and patients seeking treatment were able to be seen quickly by a provider. Many clinics expressed interest in offering eSense as additional support for their waitlisted patients, which alleviated some of the strain associated with long wait times. This finding aligns with prior research suggesting that digital interventions can complement traditional care models and offer value to health care providers in overburdened systems [52].

However, clinics also cited resource limitations, including staff time and financial constraints, as barriers to long-term adoption. This finding is consistent with a systematic review of 221 eHealth intervention studies focused on the barriers and facilitators of successful outcomes [60], which found that an increase in workload and workflow disruption posed significant barriers to implementing eHealth tools [60]. Such barriers are likely particularly important for integrating digital health tools in smaller, less-resourced clinics [54] and, given that most partnering clinics in this study offered psychological services, they may be smaller or less well-funded (or not publicly funded at all) in comparison to larger medical centers.

These findings make it important to explore alternative methods of implementing eSense, including different treatment settings and varying levels of integration. For example, the PTSD Coach App was created to translate gold-standard psychotherapy for posttraumatic stress disorder into a self-guided or clinician-aided online tool similar to eSense. While PTSD Coach is often used by providers in specialty trauma clinics [61], it has also been successfully integrated into generalist primary care clinics, both with and without provider support [62]. One of the benefits of such an approach is that it is possible to achieve broad reach without consistent formal implementation work because the program is viewed as a stand-alone tool rather than a new labor-intensive addition to clinic procedures [63]. Similarly, eSense might achieve maximum reach by being an external resource recommended in general treatment settings rather than via integration into specialty care.

However, existing as a stand-alone tool raises the question of who will or can pay for the service. Indeed, the cost came up during our assessment of adoption, with several clinics noting that they would not be able to cover the additional costs incurred from the use of eSense in the future. During the present trial, eSense was evaluated within a research context, and as such, there were no costs to the clinics that adopted it. However, in the future, when eSense is incorporated and commercial, there may be necessary costs to clinics associated with leasing or purchasing a license. Alternatively, end users could pay for eSense access, with clinics serving only as referral sources. While the specific costs of such activities to clinics or users are unknown at the moment, anecdotal feedback to our team from potential users suggests that those residing in the United States were more likely to expect eSense to be covered by a third-party payer, whereas Canadian potential users were open to paying an individual subscription fee.

Who should bear the cost of interventions such as eSense is fundamentally an issue of equity, and if the cost is shouldered by end users, many of the same inequities present in traditional health care may be replicated, with historically underserved groups having less access and experiencing greater burden. To avoid such negative outcomes, a 2022 summit of the Agency for Healthcare Research and Quality provided a set of recommendations [64] that emphasized the importance of primary care as an avenue of equitable implementation of care, along with the need to actively engage with members of underserved communities, and partner with culturally relevant sectors such as faith organizations and tribal communities. In line with such recommendations, future implementation of eSense should (1) identify ways in which care settings can budget for the additional time and costs associated with integrating it into clinical operations, (2) include active engagement with members of underserved populations (eg, sexual and gender minority individuals), or (3) explore the implementation of eSense in a variety of settings that can allow for equitable access.

At the participant level, eSense was well-received by participants who engaged with it, as evidenced by high satisfaction scores. Of note, of 66 women who initially expressed an interest in this study, only 12 participated and provided pre-post questionnaire data sufficient for analysis. This small sample size impacted the power of our statistical analyses, and as such, the effect sizes obtained should be taken as preliminary. Although 87.5% of participants did not complete all 8 modules, there was a significant reduction in sexual distress and a willingness to continue practicing the skills learned in eSense beyond this study. These findings are consistent with the PLISSIT (Permission, Limited Information, Specific Suggestions, and Intensive Therapy) model [65], which suggests that many individuals with sexual concerns can experience meaningful benefits from early interventions such as normalization and permission-giving elements introduced in eSense's first module. Therefore, the low completion rates may not necessarily reflect dissatisfaction or ineffectiveness but rather that participants received sufficient benefit from earlier parts of the program. The minimally effective dose of eSense to lead to meaningful clinical improvements is unknown and should be the focus of future research.

Although eSense was efficacious in this trial and the original randomized trial [42], this efficacy was not sufficient for

XSL•FO
RenderX

successful implementation. It is well documented that the factors that contribute to a treatment being efficacious in a controlled research study are fundamentally different from programs that are successfully implemented [66]. One reason for the difference is that efficacy trials have strict rules on the setting, conditions, participants, and other factors, whereas implementation trials must be robust across a variety of different settings, conditions, and participants [42]. Another reason is that efficacy trials attempt to control variance by restricting the conditions, whereas implementation trials seek to understand variation across different settings. Given that eSense was highly efficacious but challenges arose with implementation, this suggests that there is a need to understand the moderating factors that contribute to its successful implementation in the future, as these moderating factors will be important if eSense is to be successfully implemented.

Our findings imply that digital interventions such as eSense can be a valuable supplement to traditional therapy, particularly for individuals facing barriers to accessing care, such as long wait times or financial constraints. eSense's ability to reduce sexual distress supports the notion that digital tools may fill critical gaps in the health care system by offering low-cost, accessible solutions, and even without improving key symptoms, the reduction in distress may be particularly meaningful for patients to increase their readiness for future interventions. However, this study also highlights the importance of "fit" when introducing such interventions. Women seeking specialty care at sexual health clinics may differ significantly from the general population, having already invested substantial time and resources into accessing specialized care. These individuals may be less likely to engage fully with digital interventions such as eSense, especially if their expectation is face-to-face therapy.

Despite the promising results of eSense's implementation, this study has several limitations. The primary limitation is the relatively small number of participants, which restricts the generalizability of the findings. Additionally, because we did not formally capture user engagement data (apart from homework completion), we do not know how much participants actually engaged with eSense and what dose of the treatment they were exposed to. There is also a need to implement structured checklists specific to digital health interventions to determine their value [34]. Finally, most participants did not work with a navigator as they completed the eSense modules. Given that the randomized controlled trial showed excellent efficacy when participants were assigned a navigator [42], it is possible that any future implementation of eSense may be more successful if navigator support is included.

Conclusion

This paper addressed a significant gap in the literature on enablers to implement effective sexual health treatments in the clinical setting. This study found that eSense demonstrates potential as a digital intervention for sexual difficulties for women, particularly concerning its moderate implementation outcomes and also because of its ability to reduce sexual distress. However, challenges related to clinic adoption and participant engagement suggest that future iterations may need to consider scalability, clinic resource constraints, and the specific needs of participants seeking specialized care to increase their engagement. In particular, we recommend that costs to clinics be kept minimal, that methods of reducing attrition and enhancing adherence to modules and homework for users be considered, and that flexible options for engaging with navigators might be considered in future implementations of eSense.

Acknowledgments

We wish to thank Kiarah O'Kane, BA, and Raveena Gowda, BA, for their assistance with data collection. Funding for this study was provided by a Michael Smith Health Research BC (British Columbia) Innovation to Commercialization Grant to LAB (principal applicant).

Data Availability

The dataset for this study is available at Open Science Framework [67].

Authors' Contributions

LAB, KRS, and EM designed this study. LAB, KRS, NM, and EM collected the data. NM, MB, YO, and EM analyzed the data. LAB, KRS, NM, MB, YO, and EM wrote this paper.

Conflicts of Interest

None declared.

References

1. Zheng J, Skiba MA, Bell RJ, Islam RM, Davis SR. The prevalence of sexual dysfunctions and sexually related distress in young women: a cross-sectional survey. Fertil Steril 2020;113(2):426-434 [FREE Full text] [doi: 10.1016/j.fertnstert.2019.09.027] [Medline: 32106994]

XSL•FO
RenderX

2. McCabe MP, Sharlip ID, Lewis R, Atalla E, Balon R, Fisher AD, et al. Incidence and prevalence of sexual dysfunction in women and men: a consensus statement from the fourth international consultation on sexual medicine 2015. J Sex Med 2016;13(2):144-152. [doi: 10.1016/j.jsxm.2015.12.034] [Medline: 26953829]
3. Evensen M. Sociocultural factors of female sexual desire and sexual satisfaction. Intuition: BYU Undergrad J Psychol 2020;15(1):2 [FREE Full text]
4. Atlantis E, Sullivan T. Bidirectional association between depression and sexual dysfunction: a systematic review and meta-analysis. J Sex Med 2012;9(6):1497-1507. [doi: 10.1111/j.1743-6109.2012.02709.x] [Medline: 22462756]
5. Davison SL, Bell RJ, LaChina M, Holden SL, Davis SR. The relationship between self-reported sexual satisfaction and general well-being in women. J Sex Med 2009;6(10):2690-2697. [doi: 10.1111/j.1743-6109.2009.01406.x] [Medline: 19817981]
6. Karney BRK, Bradbury TN. The longitudinal course of marital quality and stability: a review of theory, method, and research. Psychol Bull 1995;118(1):3-34. [doi: 10.1037/0033-2909.118.1.3] [Medline: 7644604]
7. Stephenson KR, Meston CM. The conditional importance of sex: exploring the association between sexual well-being and life satisfaction. J Sex Marital Ther 2013;41(1):25-38. [doi: 10.1080/0092623x.2013.811450]
8. Alahverdi F, Shahbaztabari N, Nayeri RD, Nayeri SD. Effect of mindfulness-based intervention on the treatment of women's sexual dysfunction: a literature scoping review. JCDR 2022;16(6):LE01-LE06. [doi: 10.7860/jcdr/2022/54896.16463]
9. Jaderek I, Lew-Starowicz M. A systematic review on mindfulness meditation-based interventions for sexual dysfunctions. J Sex Med 2019;16(10):1581-1596. [doi: 10.1016/j.jsxm.2019.07.019] [Medline: 31570137]
10. Larraz AMC, Villena Moya AV, Actis CC. Mindfulness-based intervention and sexuality: a systematic review. Trends Psychiatry Psychother 2024;46:e20210459 [FREE Full text] [doi: 10.47626/2237-6089-2021-0459] [Medline: 36803998]
11. Selice L, Morris KL. Mindfulness and sexual dysfunction: a systematic research synthesis. J Sex Marital Ther 2021;48(4):323-342. [doi: 10.1080/0092623x.2021.1991536]
12. Stephenson KR, Zippan N, Brotto LA. Feasibility of a cognitive behavioral online intervention for women with sexual interest/arousal disorder. J Clin Psychol 2021;77(9):1877-1893. [doi: 10.1002/jclp.23137] [Medline: 33831258]
13. Barlow DH. Causes of sexual dysfunction: the role of anxiety and cognitive interference. J Consult Clin Psychol 1986;54(2):140-148. [doi: 10.1037//0022-006x.54.2.140] [Medline: 3700800]
14. Bergvall L, Himelein MJ. Attitudes toward seeking help for sexual dysfunctions among US and Swedish college students. Sex Relat Ther 2013;29(2):215-228. [doi: 10.1080/14681994.2013.860222]
15. Wiegel M, Scepkowski LA, Barlow DH. Cognitive-affective processes in sexual arousal and sexual dysfunction. In: The Psychophysiology of Sex. Indiana: Indiana University Press; 2007:143-165.
16. Stephenson KR, Meston CM. Why is impaired sexual function distressing to women? The primacy of pleasure in female sexual dysfunction. J Sex Med 2015;12(3):728-737 [FREE Full text] [doi: 10.1111/jsm.12804] [Medline: 25556719]
17. Stinson RD. The behavioral and cognitive-behavioral treatment of female sexual dysfunction: how far we have come and the path left to go. Sex Relat Ther 2009;24(3-4):271-285 [FREE Full text] [doi: 10.1080/14681990903199494]
18. Brotto LA, Basson R. Group mindfulness-based therapy significantly improves sexual desire in women. Behav Res Ther 2014;57:43-54. [doi: 10.1016/j.brat.2014.04.001] [Medline: 24814472]
19. Segal ZV, Ferguson AM. Mindfulness-based cognitive therapy: treatment development from a common cognitive therapy core. In: Science and Practice in Cognitive Therapy: Foundations, Mechanisms, and Applications. New York: Guilford Press; 2018:159-174.
20. Bishop SR, Lau M, Shapiro S, Carlson L, Anderson ND, Carmody J, et al. Mindfulness: a proposed operational definition. Clinical Psychology: Science and Practice 2004;11(3):230-241. [doi: 10.1093/clipsy.bph077]
21. Brotto LA. Mindful sex. Can J Hum Sex 2013;22(2):63-68. [doi: 10.3138/cjhs.2013.2132]
22. Brotto LA, Stephenson KR, Zippan N. Feasibility of an online mindfulness-based intervention for women with sexual interest/arousal disorder. Mindfulness (N Y) 2022;13(3):647-659 [FREE Full text] [doi: 10.1007/s12671-021-01820-4] [Medline: 35035598]
23. Adegunloye OA, Ezeoke GG. Sexual dysfunction—a silent hurt: issues on treatment awareness. J Sex Med 2011;8(5):1322-1329. [doi: 10.1111/j.1743-6109.2010.02090.x] [Medline: 21091876]
24. Hobbs LJ, Mitchell KR, Graham CA, Trifonova V, Bailey J, Murray E, et al. Help-seeking for sexual difficulties and the potential role of interactive digital interventions: findings from the third British national survey of sexual attitudes and lifestyles. J Sex Res 2019;56(7):937-946 [FREE Full text] [doi: 10.1080/00224499.2019.1586820] [Medline: 30907691]
25. Fitter J, Hayter M, Wylie KR. What factors influence individuals and couples to seek help for sexual and/or relationship difficulties from primary care professionals: an exploratory qualitative study. Sexual and Relationship Therapy 2009;24(3-4):333-346. [doi: 10.1080/14681990902829547]
26. Gott M, Hinchliff S. Barriers to seeking treatment for sexual problems in primary care: a qualitative study with older people. Fam Pract 2003;20(6):690-695. [doi: 10.1093/fampra/cmg612] [Medline: 14701894]
27. Sin J, Galeazzi G, McGregor E, Collom J, Taylor A, Barrett B, et al. Digital interventions for screening and treating common mental disorders or symptoms of common mental illness in adults: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20581 [FREE Full text] [doi: 10.2196/20581] [Medline: 32876577]

XSL•FO
RenderX

28. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
29. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
30. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C, et al. Implementing Sexual Health Promotion Interactive Digital Interventions in Real-World Settings. Southampton (UK): NIHR Journals Library; 2015.
31. Baumann M, Stargardt T, Frey S. Cost-utility of internet-based cognitive behavioral therapy in unipolar depression: a Markov model simulation. Appl Health Econ Health Policy 2020;18(4):567-578 [FREE Full text] [doi: 10.1007/s40258-019-00551-x] [Medline: 32060822]
32. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
33. Thomas N, McLeod B, Jones N, Abbott JA. Developing internet interventions to target the individual impact of stigma in health conditions. Internet Interventions 2015;2(3):351-358. [doi: 10.1016/j.invent.2015.01.003]
34. Kolasa K, Kozinski G. How to value digital health interventions? A systematic literature review. Int J Environ Res Public Health 2020;17(6):2119 [FREE Full text] [doi: 10.3390/ijerph17062119] [Medline: 32209988]
35. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
36. Foy R, Ivers NM, Grimshaw JM, Wilson PM. What is the role of randomised trials in implementation science? Trials 2023;24(1):537 [FREE Full text] [doi: 10.1186/s13063-023-07578-5] [Medline: 37587521]
37. Psychologists embrace telehealth to prevent the spread of COVID-19. APA Services Inc. URL: https://www.apaservices.org/practice/legal/technology/psychologists-embrace-telehealth [accessed 2025-02-14]
38. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
39. Zimbile F, Beek T, David S, Crutzen R. An implementation pilot of web-based self-training programs on sexual dysfunctions in the Dutch public sexual health setting: mixed methods study. JMIR Form Res 2023;7:e49009 [FREE Full text] [doi: 10.2196/49009] [Medline: 37883172]
40. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. In: Diagnostic and Statistical Manual of Mental Disorders. Washington DC: American Psychiatric Association; 2013.
41. Mahar EA, O'Kane KMK, Brotto LA, Stephenson KR. Online and mobile psychotherapeutic treatments for female sexual difficulties: a review of recent empirical literature. Curr Sex Health Rep 2022;14(4):174-189. [doi: 10.1007/s11930-022-00333-y]
42. Mahar EA, Stephenson KR, Brotto LA. A randomized controlled trial of online mindfulness and cognitive-behavioral interventions for sexual interest/arousal disorder in women: eSense. Manuscript Under Review 2022;13(3):647-659 [FREE Full text] [doi: 10.1007/s12671-021-01820-4] [Medline: 35035598]
43. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
44. Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. Am J Public Health 2013;103(6):e38-e46. [doi: 10.2105/ajph.2013.301299]
45. Forman J, Heisler M, Damschroder LJ, Kaselitz E, Kerr EA. Development and application of the RE-AIM QuEST mixed methods framework for program evaluation. Prev Med Rep 2017;6:322-328 [FREE Full text] [doi: 10.1016/j.pmedr.2017.04.002] [Medline: 28451518]
46. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
47. Harden SM, Gaglio B, Shoup JA, Kinney KA, Johnson SB, Brito F, et al. Fidelity to and comparative results across behavioral interventions evaluated through the RE-AIM framework: a systematic review. Syst Rev 2015;4(1):155 [FREE Full text] [doi: 10.1186/s13643-015-0141-0] [Medline: 26547687]
48. Kessler RS, Purcell EP, Glasgow RE, Klesges LM, Benkeser RM, Peek CJ. What does it mean to "employ" the RE-AIM model? Eval Health Prof 2013;36(1):44-66. [doi: 10.1177/0163278712446066] [Medline: 22615498]
49. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012;43(3):337-350 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.024] [Medline: 22898128]
50. Glasgow RE, Klesges LM, Dzewaltowski DA, Estabrooks PA, Vogt TM. Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. Health Educ Res 2006;21(5):688-694. [doi: 10.1093/her/cyl081] [Medline: 16945984]

XSL•FO
RenderX

51. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
52. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
53. Gershkovich M, Herbert JD, Forman EM, Schumacher LM, Fischer LE. Internet-delivered acceptance-based cognitive-behavioral intervention for social anxiety disorder with and without therapist support: a randomized trial. Behav Modif 2017;41(5):583-608. [doi: 10.1177/0145445517694457] [Medline: 28776431]
54. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]
55. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
56. Derogatis L, Clayton A, Lewis-D'Agostino D, Wunderlich G, Fu Y. Validation of the female sexual distress scale-revised for assessing distress in women with hypoactive sexual desire disorder. J Sex Med 2008;5(2):357-364. [doi: 10.1111/j.1743-6109.2007.00672.x] [Medline: 18042215]
57. Clayton AH, Segraves RT, Leiblum S, Basson R, Pyke R, Cotton D, et al. Reliability and validity of the Sexual Interest and Desire Inventory-Female (SIDI-F), a scale designed to measure severity of female hypoactive sexual desire disorder. J Sex Marital Ther 2006;32(2):115-135. [doi: 10.1080/00926230500442300] [Medline: 16418104]
58. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
59. Althof SE, Corty EW, Levine SB, Levine F, Burnett AL, McVary K, et al. EDITS: development of questionnaires for evaluating satisfaction with treatments for erectile dysfunction. Urology 1999;53(4):793-799. [doi: 10.1016/s0090-4295(98)00582-2] [Medline: 10197859]
60. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
61. Reger GM, Browne KC, Campellone TR, Simons C, Kuhn E, Fortney JC, et al. Barriers and facilitators to mobile application use during PTSD treatment: clinician adoption of PE coach. Prof Psychol: Res Pract 2017;48(6):510-517 [FREE Full text] [doi: 10.1037/pro0000153]
62. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
63. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
64. Baumann AA, Shelton RC, Kumanyika S, Haire-Joshu D. Advancing healthcare equity through dissemination and implementation science. Health Serv Res 2023;58 Suppl 3(Suppl 3):327-344 [FREE Full text] [doi: 10.1111/1475-6773.14175] [Medline: 37219339]
65. Annon JS. The PLISSIT model: a proposed conceptual scheme for the behavioral treatment of sexual problems. J Sex Educ Ther 1976;2(1):1-15. [doi: 10.1080/01614576.1976.11074483]
66. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003;93(8):1261-1267 [FREE Full text] [doi: 10.2105/ajph.93.8.1261] [Medline: 12893608]
67. eSense Scale-Up. OSF Home. URL: https://osf.io/c6ju2/?view_only=5f635b01cc844dd6a364c779c75cba19 [accessed 2025-02-27]

Abbreviations

CBT: cognitive behavioral therapy
HIPAA: Health Insurance Portability and Accountability Act
MBT: mindfulness-based therapy
PLISSIT: Permission, Limited Information, Specific Suggestions, and Intensive Therapy
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance of Implementation

XSL•FO
RenderX

Edited by A Mavragani; submitted 09.12.24; peer-reviewed by O Gomez-Ramirez, M Pennine; comments to author 25.12.24; revised version received 29.01.25; accepted 31.01.25; published 25.03.25.

Please cite as:
Brotto LA, Stephenson KR, Marshall N, Balvan M, Okara Y, Mahar EA
Evaluating a Digital Health Tool Designed to Improve Low Sexual Desire in Women: Mixed-Methods Implementation Science Study
J Med Internet Res 2025;27:e69828
URL: https://www.jmir.org/2025/1/e69828
doi:10.2196/69828
PMID:

©Lori A Brotto, Kyle R Stephenson, Nisha Marshall, Mariia Balvan, Yaroslava Okara, Elizabeth A Mahar. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of Clinic-Based and Asynchronous Video-Based Exercise on Clinic and Psychosocial Outcomes in Patients With Knee Osteoarthritis: Quasi-Experimental Study

Chidozie E Mbada[1], PhD; Henry Akintunji Awosika[2], MSc; Oluwatobi Ademola Sonuga[2], BMR; Micheal Akande[2], BMR; Tadesse Gebrye[1], MPH; Richard Woolf[3], PT, DPT; Francis Fatoye[1,4], PhD

[1]Department of Health Professions, Faculty of Health and Education, Manchester Metropolitan University, Manchester, United Kingdom

[2]Department Of Medical Rehabilitation, College Of Health Sciences, Obafemi Awolowo University, Ile-Ife, Nigeria

[3]Federation of State Boards of Physical Therapy, Alexandria, VA, United States

[4]Lifestyle Diseases, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa

Corresponding Author:
Chidozie E Mbada, PhD
Department of Health Professions
Faculty of Health and Education
Manchester Metropolitan University
53 Bonsall Street
Manchester, M15 6GX
United Kingdom
Phone: 44 161 247 2963
Email: c.mbada@mmu.ac.uk

Abstract

Background: Telerehabilitation is promising for improving knee osteoarthritis, but the effect of different telerehabilitation strategies on knee osteoarthritis is unclear.

Objective: This study aimed to examine the effect of a clinic-based strengthening exercise (CbSE) and asynchronous video-based strengthening exercise (AVbSE) on pain, range of motion, muscle strength, quality of life, and physical function among patients with knee osteoarthritis.

Methods: A total of 52 consenting patients participated in this 8-week experimental study; they were assigned to the CbSE or AVbSE group at 2 different study sites. CbSE is a circuit exercise module comprising knee flexion and extension warm-up in sitting, quadriceps isometric setting, quadriceps strengthening exercise, hamstring clenches, wall squat, and a cooldown of knee flexion and extension. The AVbSE is an asynchronous video-based version of the CbSE.

Results: This study spanned from March 31, 2021, to November 26, 2021. Eight out of 62 participants discontinued participation. Data collection and analysis have been completed. Significant differences were only observed in the mental health (t_{50}=–3, P=.004), pain ($t_{39.4}$=–3.6, P<.001), social support (t_{50}=–2.7, P=.009), and social activities (t_{50}=2.2, P=.03) domains of the Osteoarthritis Knee and Hip Quality of Life (OAKHQoL) questionnaire with higher scores in the AVbSE group at the end of week 4. At the end of week 8, significant differences were observed in mental health (t_{50}=–2.1, P=.04) and pain ($t_{37.3}$=–2.8, P=.008) measures with higher scores in AVbSE; however, a significantly higher score was observed in the CbSE group for the Quadruple Visual Analog Scale. No significant main effect of time was observed in this study, except in the muscle strength (F_{2100}=1.5, P=.24), social support (F_{2100}=2.5, P=.09), and social activity (F_{2100}=0.7, P=.48) domains of the OAKHQoL questionnaire and activity limitation (F_{2100}=0.1, P=.90), and performance restriction (F_{2100}=1.3, P=.27) domains of the Ibadan Knee and Hip Osteoarthritis Outcome Measure (IKHOAM) questionnaire. There was no significant difference between groups in all OAKHQoL domains except social activities (mean 17.6, SD 1.2 vs 22.8, SD 1.2; P=.003) and average pain (2.8, SD 1.6 vs 2.3, SD 1.6; P=.03) with higher AVbSE mean scores. However, a higher score was observed for the CbSE group in the Quadruple Visual Analog Scale's least pain domain (1.2, SD 0.2 vs 0.7, SD 0.2; P=.04). Also, interaction effects showed that AVbSE scores were significantly higher for the OAKHQoL questionnaire's physical activity and mental health domains at all time points. However, the CbSE score was higher for the physical performance domain of the IKHOAM questionnaire in the eighth week.

XSL•FO
RenderX

Conclusions: CbSE circuit training and its AVbSE variant effectively improve treatment outcomes and increase the quality of life of patients. While AVbSE was associated with higher improvement in most health-related quality of life domains, CbSE led to higher improvement in average pain.

Trial Registration: Pan African Clinical Trial Registry PACTR202208515182119, https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=23943

(J Med Internet Res 2025;27:e58393) doi:10.2196/58393

KEYWORDS

knee osteoarthritis; video; physiotherapy; exercise; mobile phone; telehealth; telemedicine; randomized; controlled trial; asynchronous; rehabilitation; knees; joints; osteoarthritis; musculoskeletal; rheumatology; physical therapy

Introduction

Osteoarthritis is a degenerative joint disease [1] affecting synovial joints and is a leading cause of pain and disability among adults [2]. Knee osteoarthritis is the most common type, with data indicating that 45% of people may develop symptomatic knee osteoarthritis in their lifetime [3]. Accordingly, knee osteoarthritis is the most prevalent type of osteoarthritis in Africans [4], accounting for 65%-78% of cases in hospitals in Nigeria [5]. Epidemiological studies have shown that obesity, microtrauma, knee joint alignment, repetitive use of joints, bone density, muscle weakness, and joint laxity are all local risk factors for knee osteoarthritis [6,7]. Further reports indicate that approximately 10% and 18% of men and women, respectively, present with myriad symptoms and radiological evidence, especially in about 30%-50% of adults aged 65 years and older [8,9]. Osteoarthritis of the knee is a prevalent musculoskeletal condition affecting older people, often accompanied by pain, physical disability, and reduced quality of life (QoL), with considerable economic burden on the health care system [10,11]. Research indicates that knee osteoarthritis occurs earlier in life, affecting younger adults more often than in previous years, likely due to increased obesity and joint injury [12].

The earlier occurrence and rising prevalence of knee osteoarthritis invite concern for effective management strategies to optimize the patient's QoL [11] and the general well-being of the affected individuals [13]. Physiotherapy is a conservative approach that has been consistently recommended in the management of osteoarthritis [14]. The updated European League Against Rheumatism guidelines for the nonpharmacological core management of hip and knee osteoarthritis (recommendation 3) suggest that individuals with hip or knee osteoarthritis should be provided with an exercise program (such as strength, aerobic, flexibility, or neuromotor training) tailored to their physical function, preferences, and available services, with a progressive increase in intensity and level of difficulty [15]. The Osteoarthritis Research Society International guideline for the nonsurgical management of knee osteoarthritis also recommends land- and water-based exercises and muscle strengthening for pain management, improving physical function and QoL [16]. Accordingly, therapeutic exercise is the mainstay of physiotherapy for knee osteoarthritis as it is an effective and well-tolerated treatment [14]. Exercise intervention in knee osteoarthritis alleviates pain, improves physical performance, and optimizes social, domestic, occupational, and recreational participation [17]. Other potential benefits of exercise for this patient group include improvements in mobility and psychosocial health, as well as reduction of fall risk, body weight, and metabolic abnormalities [18]. Muscle strengthening is also a critical component of most exercise regimes for knee osteoarthritis [19]. Research shows that pain and disability can be reduced with aerobic walking and quadriceps strengthening in patients with knee osteoarthritis [20,21]. Self-management has also been increasingly shown to be an essential component of knee osteoarthritis management. Wu et al [22] found that self-management helped manage pain and stiffness and improve QoL in persons with knee osteoarthritis. Uritani et al [23] also found evidence for the benefits of self-management intervention in improving the self-efficacy of persons with knee osteoarthritis in handling pain and other symptoms. Accordingly, both the European League Against Rheumatism and Osteoarthritis Research Society International guidelines for managing knee osteoarthritis also recommend self-management in conjunction with exercise interventions for managing knee osteoarthritis [15,16].

Still, health care access–related issues can impede some patients with knee osteoarthritis from receiving exercise that is vital for their care [24]. Specifically, in medically underserved areas, physiotherapy services can be limited or lacking entirely [24,25]. These issues are particularly salient for individuals with low socioeconomic status, who also bear a greater burden of osteoarthritis [26]. These issues invite the need for innovative platforms that will help improve access to rehabilitative services and encourage self-management for patients with knee osteoarthritis [27].

Recently, telerehabilitation has been advocated to reduce the frequency of clinic visits, clinic waiting time, and costs incurred from transportation to the clinic, especially for patients living far distances from the clinic [24,28]. In addition, telerehabilitation has the potential to improve access to rehabilitation for patients in medically underserved countries like Nigeria and similar sub-Saharan African contexts. Medically underserved regions are characterized by a lack of adequate health care services, a high proportion of the population living in poverty, a high percentage of elderly residents, and a low ratio of health care providers to the population [29]. Africa is still facing a critical shortage of health workers despite dealing with the highest disease burden in the world [30,31]. Therefore, validated telerehabilitation platforms that may help overcome the challenges to accessing care faced by patients with knee osteoarthritis are necessary. Video-based telerehabilitation

XSL•FO
RenderX

applications may effectively provide care in rural and resource-limited settings where internet-enabled digital services may be scarce or unavailable [24,25]. Particularly, asynchronous video-based telerehabilitation interventions can help overcome the time constraints of in-person and remote synchronous care [32]. Asynchronous digital health tools, also known as "store-and-forward" technology, allow for remote, non–real-time communication between providers and patients. This enables both parties to access the platform at their convenience [33]. These tools can address the limitations posed by poor or lack of internet access and coverage, which is common in Africa [34].

Given the prospect of asynchronous video-based telerehabilitation interventions in overcoming the barrier to health care access that persons with knee osteoarthritis may face, we propose an asynchronous video-based strengthening exercise program as a viable alternative. This study, therefore, seeks to compare the effect of a traditional clinic-based strengthening exercise (CbSE) program and an in-home asynchronous video-based strengthening exercise (AVbSE) program on clinical and psychosocial outcomes in patients with knee osteoarthritis. This study hypothesizes that the outcomes related to the in-home AVbSE would be comparable to CbSE over a period of 8 weeks.

Methods

Design

This study used a pretest-posttest quasi-experimental design.

Participants

Eligible participants for the study included patients with radiologically confirmed knee osteoarthritis for not less than 3 months, who were 45 years and older, had a Kellgren-Lawrence grade of II and III, and were without any obvious deformities of the trunk or upper and lower extremities. Exclusion criteria were patients with other comorbidities such as acute inflammation, knee surgery, cardiovascular, respiratory, neurologic, and metabolic diseases, and pregnant women. Non–mobile phone owners after initial contact and screening were also excluded.

Sample Size

Sample size estimation for the study was based on the equation $c\times\pi1(1–\pi1)+\pi2\ (1–1\pi2)/(\pi1–\pi2)^2$ [35], where c=7.9 for 80% power, and π1 and π2 are proportion estimates (π1=0.25 and π2=0.65). Therefore, n=7.9 × (0.25 (1–0.25)+0.65 (1–0.65)/(0.25–0.65)=20.49, which is approximately 21. Hence, the calculated N is 42 (21 per group). To account for 10% possible attrition, the approximate minimum sample size was 46.

Study's Site and Sampling Procedure

This study was conducted at the Physiotherapy Departments of the Federal Medical Centre and Benue State University Teaching Hospital, both in Makurdi, Benue State, Nigeria. The study sites were selected purposefully based on the caseload at both centers and, second, one of the centers had fewer physiotherapists. The center with more physiotherapists was chosen as the control CbSE, while the other was the experimental AVbSE. Using 2 sites in this study was intended to minimize the spillover effect (spillover bias refers to an unintended effect that happens to those for whom the experimental intervention was not designed) by geographically separating the CbSE (control) from the AVbSE group. Also, the single-blind approach minimized the Hawthorne effect (that's a situation where participants in the experimental or control group show exceptionally higher outcome scores or performance just because they are aware they are being studied). Thus, a research assistant who was not involved in the interventions conducted the recruitment and assessments. The study sample was randomly drawn from the list of patients with knee osteoarthritis attending both clinics.

Outcomes

The primary outcomes in this study were pain, range of motion, and muscle strength, while osteoarthritis-specific QoL and physical function were the secondary outcomes. The instruments used in the study were the Quadruple Visual Analog Scale (QVAS), goniometer, tensiometer, and Ibadan Knee and Hip Osteoarthritis Outcome Measure (IKHOAM) questionnaire, and Osteoarthritis Knee and Hip Quality of Life (OAKHQoL) questionnaire.

QVAS

This tool was used to measure pain intensity based on current pain, average pain, least pain, and worst pain, respectively. The QVAS is a segmented numeric alternative to the visual analog scale where a whole number (0-10 integers) that best reflects the intensity of the pain for the different dimensions is chosen by the respondent [32]. The total score of the QVAS is derived by scoring 1+2+4, divided by 3, and multiplied by 10. Pain intensity can then be classified as low intensity (50) or high intensity (50) [36].

Goniometer

A goniometer (12-inch plastic [model 12-1000], Fabrication Enterprises) was used to measure the range of motion at the knee joint. It ranges from 0° to 180°.

Tensiometer

A tensiometer (model and pocket balance) was used to measure quadriceps muscle strength. The scale is from 0 kg to 100 kg.

The IKHOAM Questionnaire

The IKHOAM questionnaire was used to measure the activities of daily living and societal and physical performance. The IKHOAM questionnaire is a 3-domain, 33-item instrument consisting of an activity limitations domain comprising 25 activities of daily living that are performed by individuals with knee and hip osteoarthritis rated on a 5-point (0-4) ordinal scale, a participation restriction domain comprising 3 activities in societal participation rated on a 4 point (0-3) ordinal scale, and a physical performance tests domain including 5 tests rated by a clinician. These tests are (1) 250 m walk test rated on a 6-point (0-5) ordinal scale, (2) single-leg stance test rated on a 6-point (0-5) ordinal scale, (3) stair climbing test rated on a 5-point (0-4) ordinal scale, (4) squat test rated on a 5-point (0-4) ordinal scale, and (5) balance test rated on a 6-point (0-5) ordinal scale.

XSL•FO
RenderX

The maximum obtainable score for the tests is 23 (5 scores for each of the 250 m walk test, 1 single-leg stance test, and one balance test, with 4 scores for the stair climbing test and squat test, respectively). The maximum obtainable score on the IKHOAM questionnaire is 232 (ie, 200+9+23). The score for each participant was calculated in percentage as a score divided by the total possible score and multiplied by 100. A low score on the IKHOAM questionnaire implies a low level of physical functioning ability, and a high score means a high level of physical functioning ability [37].

The OAKHQoL Questionnaire

The OAKHQoL questionnaire was used to assess the QoL. The OAKHQoL is the first specific QoL questionnaire for patients with knee and hip osteoarthritis. It is self-administered and comprises 43 items in 5 dimensions: physical activity, mental health, pain, social support, and social activities, plus 3 independent items. Dimension scores are standardized from 0 (worst QoL) to 100 (best QoL) [38].

Procedure

The participants were recruited consecutively at each study site and followed through until they completed the 8-week treatment program. The participants were informed of the research's purpose at the onset of the recruitment process. An informed written permission form, which was also translated into the local language by professionals, was given to every participant (N=64) who was evaluated for study eligibility. The number of people invited to participate (n=57), the number of people who rejected (n=3), the number of screened patients who were ineligible (n=4), and their justifications for declining participation were all tracked by a research assistant.

Each individual included in the study had a baseline assessment completed. Measurements included anthropometric variables such as height, weight, and BMI. Each participant's information, including their age and gender, was recorded. All the questionnaires were applied, and the measurements of the range of motion of the knee joint and muscle strength (quadriceps) were taken at baseline, week 4, and week 8 of the study. Active graded exercises (active range of motion and quadriceps strengthening exercises) were administered to the patients in the 2 groups (CbSE and AVbSE) for 8 weeks.

The participants recruited into the AVbSE group received an asynchronous video containing exercises, instructions, and demonstrations of each exercise's performance. The AVbSE was developed to be operable on a mobile device with at least an Android OS of 4.1 or an iPhone interface (iOS) for participants who own and could operate a smartphone. The video was also available as a CD drive for those who had or preferred it (n=8). The AVbSE video was an asynchronous video-based intervention intended for usage at home (a safe exercise space). The participants were encouraged to follow the video and complete the tasks thrice weekly. The exercises in the video include a warm-up of active seated knee flexion and extension, quadriceps isometric setting, quadriceps strengthening exercise, hamstring clenches, wall squats, and cool down of active seated knee flexion and extension. The participants in the AVbSE group were telemonitored with voice calls and SMSs to remind them of their exercise schedules and provide symptom reviews and feedback on performance. Also, the telemonitoring was aimed at guiding patients through the exercises and ensuring safety by asking patients to discontinue exercise in case of any adverse event. The CbSE involves the traditional in-person services exercise session led by a physiotherapist trained in the protocols in the AVbSE. The research assistant carried out all baseline treatment outcome assessments (at weeks 4 and 8) for both the CbSE and AVbSE at the clinics. The details of the exercises on the development and feasibility testing of a smartphone video-based exercise program for patients with knee osteoarthritis are already published in an earlier study by Mbada et al [24]. The exercise intervention was developed using a modified Delphi approach, which involved 3 rounds and included a panel of 4 experts and a patient with knee osteoarthritis. After reaching a consensus, the video program for knee osteoarthritis was developed, including 5 main types of exercises: seated knee flexion and extension, quadriceps isometric setting, quadriceps strengthening exercise, hamstring clenches, and wall squats [24]. The highlights of the intervention are (1) warm-up (active range of motion exercises) for 5-10 seconds, and the movement was performed up to 10 times; (2) quadriceps isometric setting repeated up to 10 times; (3) quadriceps strengthening exercise repeated up to 10 times; (4) hamstring clenches held for 10 seconds and then done 10 more times; (5) wall squat held for 10 seconds, then slide back up while keeping their backs firmly against the wall, and the movement was performed up to 10 times; and (6) cool down (active range of motion exercises) for 5-10 seconds, and the movement was performed up to 10 times [24].

Statistical Analysis

Descriptive statistics of frequency, mean, and SD were used to summarize data, and an independent *t* test was used to compare patients' general and clinical characteristics at baseline. A 2-way mixed-model ANOVA with repeated measures on the time points was used to determine the main effects of the two treatment programs (between-subject factor) and time points (within-subject factor) on the outcome parameters. Parametric tests were used in this study under the assumption that all the measurements are continuous variables on at least an interval scale, are normally distributed, and have equivalent variances. The α level was set at .05. The analysis was done using IBM SPSS (version 27).

Ethical Considerations

This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving research study participants were approved by the Health Research Ethical Committee, Federal Medical Centre Makurdi, Nigeria (FMH/FMC/HRE/VL 1-24/03/21). Written informed consent was obtained from all participants.

Results

This study began on March 31, 2021, and ended on November 26, 2021. A total of 64 participants were recruited for the study but 8 participants discontinued. Data collection and analysis have been completed. The CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the study is presented in

XSL•FO
RenderX

Figure 1. Out of 64 patients, a total of 57 patients were found eligible for this study. There was a 12% loss to attrition based on the total number of participants recruited. Thus, 52 of 64 (81.3%) patients completed the study. The mean (SD) age, height, weight, and BMI of the participants were 56.2 (SD 7.5) years, 1.7 (SD 0.1) m, 68.8 (SD 7.6) kg, and 24.8 (SD 2.2) kg/m^2, respectively. Participants in the two groups were comparable in their general characteristics (Table 1). Most participants in both groups were females (31/52, 59%). Civil servants, artisans, and businesspeople or traders comprised 15% (8/52), 21% (11/52), and 63% (33/52), of the participants, respectively. Also, a comparison of baseline measures in the selected outcomes is presented in Table 1. Both groups were comparable at baseline, except in OAKHQoL domains ($t_{39.9}$=–4.1, $t_{42.5}$=–4.7, $t_{38.3}$=–5.8, t_{50}=–4.2, and t_{50}=6; P<.001), the physical performance domain of IKHOAM (t_{50}=–2.2, P=.04), and the least pain domain of QVAS (t_{50}=2.4, P=.02; Table 1).

Within-group effects (comparison of outcomes across baseline, fourth, and eighth week of the study) of CbSE and AVbSE are presented in Tables 2 and 3.

In Table 2, for a particular variable, mean values with different superscripts are significantly different (P<.05). When mean values have the same superscripts, they are not significantly different (P>.05; based on the least significant difference post hoc test).

Figure 1. CONSORT diagram of the flow of participants through the study.

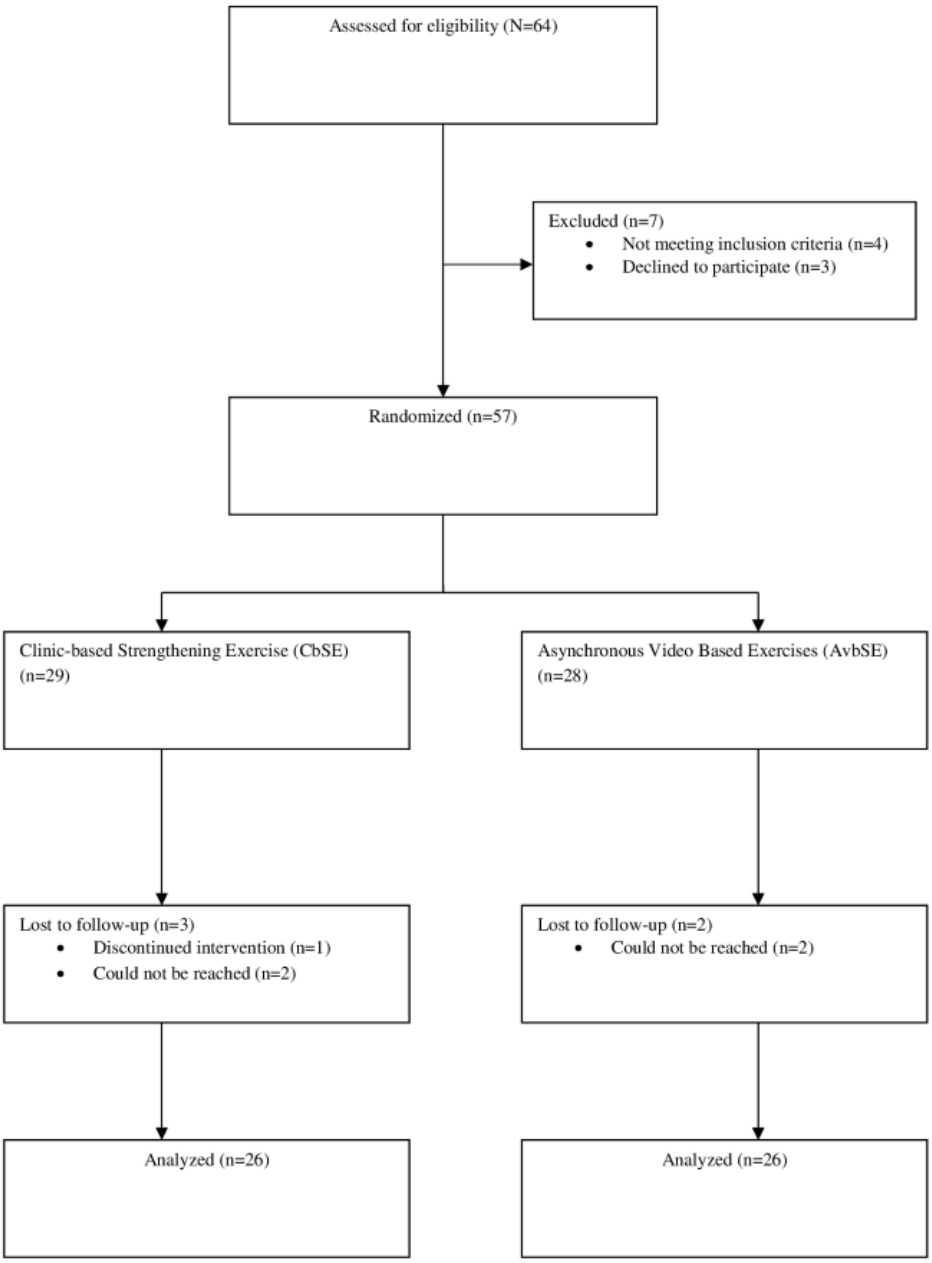

XSL•FO
RenderX

Table 1. Comparison of the general characteristics and baseline measures between clinic-based strengthening exercise and asynchronous video-based telerehabilitation strengthening exercise groups (N=52).

Variables	CbSE[a] (n=26)	AVbSE[b] (n=26)	*t* test (*df*)	*P* value
Age (year), mean (SD)	56.2 (8.2)	56.2 (6.8)	–0.018 (50)	.16
Height (m), mean (SD)	1.7 (0.1)	1.7 (0.1)	–0.713 (50)	.81
Weight (kg), mean (SD)	67.3 (9.1)	70.3 (5.6)	–1.439 (50)	.08
BMI (kg/m^2), mean (SD)	24.3 (2.5)	25.3 (1.7)	–1.640 (50)	.06
Pain intensity, mean (SD)				
Average pain	3.2 (1.0)	3.8 (1.4)	–1.875 (44.4)	.07
Least pain	1.8 (1.0)	1.2 (1.1)	2.385 (50)	.02
Worst pain	6.8 (1.4)	7.1 (1.2)	–0.749 (50)	.46
Total pain score	39.4 (9.5)	40.5 (9.7)	–0.432 (50)	.67
OAKHQoL[c] domain, mean (SD)				
Physical activities	36.8 (10.1)	53.1 (17.5)	–4.101 (39.9)	.001
Mental health	25.3 (9.6)	41.9 (15.1)	–4.747 (42.3)	.001
Pain	12.5 (2.5)	18.6 (4.7)	–5.791 (38.3)	.001
Social support	24.1 (7.0)	31.3 (5.1)	–4.263 (50)	.001
Social activities	23.5 (4.0)	17.2 (3.6)	5.994 (50)	.001
IKHOAM[d] domain , mean (SD)				
Activity limitation	142.2 (17.2)	144.2 (21.7)	–0.382 (50)	.70
Participation restriction	6.0 (1.4)	6.5 (1.7)	–1.147 (50)	.26
Physical performance	8.9 (3.1)	10.7 (2.8)	–2.167 (50)	.04

[a]CbSE: clinic-based strengthening exercise.

[b]AVbSE: asynchronous video-based telerehabilitation strengthening exercise.

[c]OAKHQoL: Osteoarthritis Knee and Hip Quality of Life.

[d]IKHOAM: Ibadan Knee and Hip Osteoarthritis Outcome Measure.

In Table 3, for a particular variable, mean values with different superscripts are significantly (P<.05) different. When mean values have the same superscripts, they are not significantly different (P>.05; based on the least significant difference post hoc test).

Results show significant differences in all the outcome parameters across the 3 time points of the study as shown in Tables 2 and 3. Also, a comparison of treatment outcomes between CbSE and AVbSE at weeks 4 and 8 of the study is presented in Table 4. The main effect of time was significant across all outcome parameters except muscle strength (F_{2100}=1.5, P=.24), social support (F_{2100}=2.5, P=.09), and social activities (F_{2100}=0.7, P=.48) of the OAKHQoL questionnaire, as well as the activity limitation (F_{2100}=0.1, P=.90) and performance restriction (F_{2100}=1.3, P=.27) domains in the IKHOAM questionnaire as given in Table 5. Across all dimensions of the QVAS, the mean score reduced significantly from the fourth week to the eighth (F_2=20.9-410.8, P<.001). Physical activities, mental health, and pain also showed a mean decrease across the time points. However, in the physical performance domain of IKHOAM, a significant increase in mean values was observed across the time points (Tables 2 and 3). Similarly, there were no significant differences in the treatment outcome between the two groups at the end of the 8 weeks of the study except for higher mean observed in mental health (17.5, SD 11.5 vs 11.6, SD 8.2; t_{50}=–2.2, P=.04) and pain (23.6, SD 10.9 vs 5.5, SD 3.2; t_{50}=–2.8, P=.008) in AVbSE group compared with CbSE group (Table 4).

A comparison of treatment outcomes between CbSE and AVbSE at weeks 4 and 8 of the study is presented in Table 4. The results of this study show that there were no significant differences in the treatment outcome between the 2 groups at the end of the fourth week of the study except for mental health (t_{50}=–3.0, P=.004), pain ($t_{39.4}$=–3.6, P<.001), social support (t_{50}=–2.7, P=.009), and social activities (t_{50}=2.2, P=.03) domains of the OAKHQoL questionnaire where the AVbSE group had significantly higher mean scores (Table 4). Similarly, there were no significant differences (P>.05) in the treatment outcome between the two groups at the end of the 8 weeks of the study except for a higher mean observed in mental health (17.5 vs 11.6; t_{50}=–2.1, P=.04) and pain (23.6 vs 5.5; $t_{37.3}$=–2.8, P=.008) in AVbSE group compared with CbSE group (Table 4). However, the CbSE group had a higher mean score in the current pain dimension of the QVAS (0.5 vs 0.1; $t_{41.4}$=2.1, P=.04).

XSL•FO
RenderX

Table 2. Within-group comparison of treatment outcomes in the clinic-based strengthening exercise group (n=26).

Variables	Baseline	Fourth week	Eighth week	*F* test (*df*)	*P* value
Range of motion, mean (SD)	112.0 (7.3)[a]	108.5 (32.2)[a]	101.6 (44.3)[a]	497.7 (2)	.001
Muscle strength, mean (SD)	17.5 (4.5)[a]	19.2 (7.1)[a]	18.9 (8.9)[a]	260.4 (2)	.001
Pain intensity , mean (SD)					
Current pain	1.8 (1.1)[a]	0.5 (0.9)[a]	0.5 (0.9)[a]	37.6 (2)	.001
Average pain	3.2 (1.0)[a]	2.3 (0.8)[a]	1.5 (0.9)[a]	289.5 (2)	.001
Least pain	1.8 (1.0)[a]	1.1 (1.1)[a]	0.7 (0.9)[a]	53.2 (2)	.001
Worst pain	6.8 (1.4)[a]	4.4 (1.8)[a]	2.6 (1.7)[a]	337.4 (2)	.001
Total pain score	39.4 (9.5)[a]	24.2 (9.8)[a]	15.1 (9.5)[a]	261.9 (2)	.001
OAKHQoL[b] domain, mean (SD)					
Physical activities	36.8 (10.1)[a]	24.8 (12.9)[a]	17.8 (13.9)[a]	151.4 (2)	.001
Mental health	25.3 (9.6)[a]	16.0 (9.5)[a]	11.6 (8.2)[a]	106.9 (2)	.001
Pain	12.5 2.5)[a]	8.3 3.1)[a]	5.5 3.2)[a]	345.8 (2)	.001
Social support	24.1 7.0)[a]	25.1 8.6)[a]	23.6 10.9)[a]	354.2 (2)	.001
Social activities	23.5 4.0)[a]	22.7 7.6)[a]	22.2 10.5)[a]	312.4 (2)	.001
IKHOAM[c] domain, mean (SD)					
Activity limitation	142.2 (17.2)[a]	143.7 45.2)[a]	145.3 (64.3)[a]	405.2 (2)	.001
Participation restriction	6.0 1.4)[a]	6.9 2.2)[a]	7.0 3.1)[a]	369.7 (2)	.001
Physical performance	8.9 3.1)[a]	11.2 4.4)[a]	13.0 6.4)[a]	188.1 (2)	.001

[a]*P*<.05.

[b]OAKHQoL: Osteoarthritis Knee and Hip Quality of Life.

[c]IKHOAM: Ibadan Knee or Hip Osteoarthritis Outcome Measure.

XSL•FO
RenderX

Table 3. Within-group comparison of treatment outcomes in the asynchronous video-based strengthening exercise group (n=26).

Variables	Baseline	Fourth week	Eighth week	*F* test (*df*)	*P* value
Range of motion, mean (SD)	112.4 7.2)	112.1 (23.2)	95.3 47.5)	741.3 (2)	.001
Muscle strength, mean (SD)	17.4 4.6)	18.7 5.4)	15.8 8.7)	367.9 (2)	.001
Pain intensity , mean (SD)					
Current pain	1.2 1.2)	0.5 0.9)	0.1 0.4)	20.9 (2)	.001
Average pain	3.8 1.4)	2.8 1.1)	1.8 1.1)	231.8 (2)	.001
Least pain	1.2 1.1)	0.8 1)	0.2 0.6)	25.6 (2)	.001
Worst pain	7.1 1.2)	5.0 1.9)	3.3 1.9)	410.8 (2)	.001
Total pain score	40.5 9.7)	27.9 10.9)	17.7 10.4)	335.7 (2)	.001
OAKHQoL[a], mean (SD)					
Physical activities	53.1 17.5)	30.9 17.0)	23.2 17.4)	146.8 (2)	.001
Mental health	41.9 15.1)	25.4 12.8)	17.5 11.5)	143 (2)	.001
Pain	18.6 4.7)	12.7 5.4)	9.3 6.3)	223.8 (2)	.001
Social support	31.3 5.1)	31.4 7.9)	26.5 13.7)	482.1 (2)	.001
Social activities	17.2 3.6)	18.6 5.7)	16.8 9.1)	285.2 (2)	.001
IKHOAM[b], mean (SD)					
Activity limitation	144.2 (21.7)	143.9 (35.5)	136 (67.7)	433.5 (2)	.001
Participation restriction	6.5 1.7)	6.8 2.1)	6.4 3.3)	291.1 (2)	.001
Physical performance	10.7 2.8)	11.3 3.3)	11.6 6.2)	253.2 (2)	.001

[a]OAKHQoL: Osteoarthritis Knee and Hip Quality of Life.

[b]IKHOAM: Ibadan Knee and Hip Osteoarthritis Outcome Measure.

The main effect of the group was found to be significant across all domains of the OAKHQoL measure with corrected *F* test values ranging from F_{150}=21.9 (*P*<.001) on the pain domain to F_{150}=6.4 (*P*=.01) on the physical activities domain. The main effect of the group was also significant for average pain intensity (F_{150}=4.9, *P*=.03) and least pain intensity (F_{150}=4.5, *P*=.04; Table 5). Higher estimated marginal means scores were observed in the AvbSE treatment group across all domains of the OAKHQoL measure except the social activities domain, where the CbSE group was higher (22.8 vs 17.6; F_1-score=9.9, *P*=.003). The average pain dimension of the QVAS score was higher in the AvbSE group (2.8 vs 2.3; F_1-score=4.9, *P*=.03), while a higher mean was observed in the CbSE group for the least pain dimension (1.2 vs 0.7; F_1-score=4.5, *P*=.04). The mean scores showing the difference between groups across all parameters are presented in Table 4.

XSL•FO
RenderX

Table 4. Comparison of treatment outcomes of clinic-based strengthening exercise and asynchronous video-based strengthening exercise at weeks 4 and 8 of the study (N=52).

	Week 4				Week 8			
Variables	CbSE[a] (n=26)	AVbSE[b] (n=26)	*t* test (*df*)	*P* value	CbSE (n=26)	AVbSE (n=26)	*t* test (*df*)	*P* value
Range of motion, mean (SD)	108.5 (32.2)	112.1 (23.2)	–0.459 (50)	.65	101.6 (44.3)	95.3 (47.5)	0.498 (50)	.62
Muscle strength, mean (SD)	19.2 (7.1)	18.7 (5.4)	0.265 (50)	.80	18.9 (9.0)	15.8 (8.7)	1.254 (50)	.22
Pain intensity , mean (SD)								
Current pain	0.5 (0.9)	0.5 (0.9)	0.001 (50)	1.0	0.5 (0.7)	0.1 (0.4)	2.133 (41.4)	.04
Average pain	2.3 (0.8)	2.8 (1.1)	–1.888 (50)	.07	1.5 (0.9)	1.9 (1.1)	–1.509 (50)	.14
Least pain	1.1 (1.1)	0.8 (1.0)	0.953 (50)	.35	0.7 (1.0)	0.2 (0.6)	1.892 (40.9)	.06
Worst pain	4.4 (1.8)	5.0 (1.9)	–1.192 (50)	.24	2.6 (1.7)	3.3 (2.0)	–1.363 (50)	.18
Total pain score	24.3 (9.8)	27.9 (10.9)	–1.294 (50)	.20	15.1 (9.5)	17.7 (10.4)	–0.927 (50)	.36
OAKHQoL[c], mean (SD)								
Physical activities	24.8 (12.9)	30.9 (17)	–1.473 (46.6)	.15	17.8 (13.9)	23.2 (17.4)	–1.226 (50)	.23
Mental health	16.0 (9.5)	25.4(12.8)	–3.000 (50)	.004	11.6 (8.2)	17.5 (11.5)	–2.139 (50)	.04
Pain	8.3 (3.1)	12.7(5.4)	–3.644 (39.4)	.001	5.5 (3.2)	23.6)	–2.782 (37.3)	.008
Social support	25.1 (8.6)	31.4 (7.9)	–2.739 (50)	.009	23.6 (10.9)	26.5 (13.7)	–0.863 (50)	.39
Social activities	22.7 (7.6)	18.6 (5.7)	2.196 (50)	.03	22.2 (10.5)	16.8 (9.1)	1.963 (50)	.06
IKHOAM[d], mean (SD)								
Activity limitation	143.7 (45.2)	144 (35.5)	–0.024 (50)	.98	145.3 (64.3)	136 (67.7)	0.506 (50)	.62
Participation restriction	6.9 (2.2)	6.8 (2.1)	0.127 (50)	.90	7 (3.1)	6.4 (3.3)	0.646 (50)	.52
Physical performance	11.2 (4.4)	11.3 (3.3)	–0.108 (50)	.92	13 (6.4)	11.6 (6.2)	0.837 (50)	.41

[a]CbSE: clinic-based strengthening exercise.

[b]AVbSE: asynchronous video-based strengthening exercise.

[c]OAKHQoL: Osteoarthritis Knee and Hip Quality of Life.

[d]IKHOAM: Ibadan Knee and Hip Osteoarthritis Outcome Measure.

XSL•FO
RenderX

Table 5. Main effects of treatment group, time points, and interaction effects.

Variables	*F* test (*df*)	*P* value	Partial[a] η^2
Range of motion			
Group	0.2 (1)	.90	0
Time	4.0 (2)	.02	0.074
Muscle strength			
Group	0.7 (1)	.41	0.013
Time	1.5 (2)	.24	0.028
Pain intensity			
Current pain			
Group	2.1 (1)	.15	0.040
Time	42.4 (2)	.001	0.459
Average pain			
Group	4.9 (1)	.03	0.089
Time	73.2 (2)	.001	0.594
Least pain			
Group	4.5 (1)	.04	0.082
Time	31.4 (2)	.001	0.386
Worst pain			
Group	2.2 (1)	.14	0.042
Time	117.3 (2)	.001	0.701
Total pain score			
Group	1.2 (1)	.28	0.024
Time	143.4 (2)	.001	0.701
OAKHQoL[b] domain			
Physical activities			
Group	6.4 (1)	.014	0.114
Time	103.3 (2)	.001	0.674
Group×Time	6.1 (2)	.003	0.108
Mental health			
Group	13.4 (1)	.001	0.211
Time	178.1 (2)	.001	0.781
Group×Time	14.2 (2)	.001	0.221
Pain			
Group	21.9 (1)	.001	0.305
Time	99.0 (2)	.001	0.665
Social support			
Group	8.6 (1)	.005	0.147
Time	2.5 (2)	.09	0.047
Social activities			
Group	9.9 (1)	.003	0.166
Time	0.7 (2)	.48	0.015
IKHOAM[c] domain			

XSL•FO
RenderX

Variables	*F* test (*df*)	*P* value	Partial[a] η^2
Activity limitation			
Group	0.1 (1)	.82	0.001
Time	0.1 (2)	.90	0.002
Participation restriction			
Group	0.01 (1)	.92	0
Time	1.3 (2)	.27	0.026
Physical performance			
Group	0.01 (1)	.90	0
Time	8.8 (2)	.001	0.150
Group×Time	3.6 (2)	.04	0.068

[a]Partial η^2: effect size.

[b]OAKHQoL: Osteoarthritis Knee and Hip Quality of Life.

[c]IKHOAM: Ibadan Knee and Hip Osteoarthritis Outcome Measure.

Discussion

Principal Findings

This study examined the effects of CbSE and AVbSE on clinical and psychosocial outcomes in patients with knee osteoarthritis. The average age of the patients was 56.2 (SD 7.5) years, and most were females. The gender pattern observed in this study might be a coincidence. However, more women are reported to seek care for their musculoskeletal conditions [8,39]. From this study, CbSE and AVbSE significantly improved clinical and psychosocial outcomes in patients with knee osteoarthritis.

Comparison to Previous Work

These findings are consistent with previous reports that exercise interventions have beneficial clinical effects for patients with osteoarthritis in the lower limb [40,41]. The available evidence strongly supports the incorporation of exercise therapy as a fundamental component of nonpharmacological treatment for osteoarthritis of the knee [42]. Numerous studies have demonstrated that muscle strengthening and aerobic exercises can reduce pain and enhance physical function in individuals with mild to moderate knee osteoarthritis [43,44]. A systematic review by Turner et al [45] indicated that resistance training can improve pain and physical function, but the most effective resistance training regimen remains uncertain. Nevertheless, it appears that a total of 24 sessions over 8-12 weeks can yield significant benefits. A meta-analysis of 17 clinical trials involving 1705 patients revealed that resistance exercise contributes to a reduction in pain and stiffness, as well as an improvement in physical function for individuals with knee osteoarthritis [46]. Furthermore, it has been established that strength training exercises carried out over a period of 8 weeks, 3-5 times a week, can be both safe and effective for patients with knee osteoarthritis.

Furthermore, there is evidence for the effectiveness of resistance exercises conducted as a home exercise program for patients with knee osteoarthritis [40,41]. Doi et al [47] conducted a randomized controlled trial involving 142 patients with knee osteoarthritis. They found that a home-based quadriceps strengthening program involving 20 repetitions of a slow knee extension movement implemented 4 times daily for 8 weeks was as effective as nonsteroidal anti-inflammatory drugs for improved pain, stiffness, physical function, and QoL outcomes. However, in a systematic review, Saengpromma et al [48] found that a home exercise program with tracking improves pain and function and increases adherence and changes in behavior in older people compared with the untracked group. Doi et al [47] also corroborate that the effectiveness of home-based exercises in patients with knee osteoarthritis depends on patient adherence. Ettinger et al [49] and Thomas et al [50] submit that adherence to strengthening exercise is a key predictor of response in patients with knee osteoarthritis. However, the best way to deliver in-home strengthening exercises and promote adherence is still unclear.

Few studies emanating from West Africa explored digital rehabilitation for knee osteoarthritis. Odole and Ojo [28] implemented an in-home intervention of standardized exercise programs for patients with knee osteoarthritis, and adherence was monitored using telephone calls. The authors found that in-home telerehabilitation was comparable with clinic-based treatment in terms of improved QoL. Similarly, Ojoawo et al [51] compared clinic-based and telemonitored home-based interventions involving exercise pamphlets and telephone calls. They found that both interventions were effective in managing knee osteoarthritis, but the clinic-based intervention was better than the telemonitored home-based intervention. Both interventions implemented low-tech digital rehabilitation involving preintervention demonstration of exercises and the use of pictorial pamphlets as a guide. This is the first study to implement an asynchronous digital health tool for the rehabilitation of Nigerian patients with knee osteoarthritis. Implementing high-tech digital health interventions in the study setting is hamstrung by many factors, including poverty, lack of needed infrastructure (such as internet availability), and digital illiteracy [52,53].

In-home telerehabilitation creates opportunities to mitigate some of the challenges of home-based interventions (such as lack of support and good communication) using technological platforms

XSL•FO
RenderX

[54]. Technology is not efficacious in itself; it depends on innovative ideas and providers' creative abilities to ensure its use is sustainable [55]. According to Russell [56], telerehabilitation's primary aim is to help patients receive equal access to rehabilitation services via images, sensors, virtual reality, and environments despite their impairments or barriers in accessibility. In physiotherapy, there has been an increase in the adoption of telecommunication technology, such as telephone or videoconferencing, to provide remote access to physiotherapy services when one-on-one contact is not feasible [28,57]. Telerehabilitation is a convenient way for patients to perform basic physiotherapy exercises on their own within their homes to avoid long-distance travel [58]. This study compared an asynchronous video-based telerehabilitation program with a clinic-based intervention. The treatment outcomes between the groups for both interventions are comparable except for average pain and least pain and OAKHQoL domains, where AVbSE produced significantly higher mean change in average pain and all domains of the OAKHQoL except social activities. Also, interaction effects show that the AVbSE had higher mean change across time points for PA and mental health OAKHQoL domain across all time points. Also, the CbSE score was only higher in the physical performance domain of the IKHOAM in the eighth week. Hence, AVbSE is a viable alternative to delivering physiotherapy for patients with knee osteoarthritis. The findings of this study support some related studies. Odole and Ojo [28] documented the efficacy of telephone-based rehabilitation on pain, physical function outcomes, and QoL in knee osteoarthritis. Chandra and Keerthi [58] also reported the efficacy of telerehabilitation using telephonic consultation and videoconferencing for patients with knee osteoarthritis, using exercise as a home-based treatment. Bennell et al [59] investigated the cost and clinical effectiveness of combining a physiotherapist-delivered intervention with a nurse-delivered telephone coaching for people with knee osteoarthritis. Better outcomes were achieved in the group of patients who received physiotherapy treatment and the nurse-delivered telephone coaching than those who received only physiotherapy treatment. According to Adhikari et al [60], telephone-based treatments used in telerehabilitation were both feasible and efficient, significantly reducing pain caused by various musculoskeletal problems in resource-limited settings. Another study by Multani et al [61] found significant improvement in pain, muscle strength, and functional ability within 4 weeks of telerehabilitation in patients with knee osteoarthritis.

This study supports other studies, including reviewed literature that suggest that muscle strengthening interventions improve pain and physical functioning in persons with mild to moderate osteoarthritis of the knee. Also, the study reveals that asynchronous video-based exercises effectively manage patients with knee osteoarthritis and produce comparable effects in physiologic and psychosocial outcomes to conventional clinic-based physiotherapy. Furthermore, the findings of this study confirm an earlier submission by Winters [62] that telerehabilitation brings the hope of enabling access for all, as well as helping address societal challenges in the delivery of rehabilitative services once barriers such as distance and reimbursement are overcome. Also, according to Lamplot et al [63], as technology improves in remote and rural locations, initial knee examination and follow-up visits may also be conducted virtually [63]. The findings of this study support earlier advocacy on the necessity for the development and implementation of telerehabilitation services to support persons with potentially disabling conditions who may have limited access to rehabilitation services [64,65].

Clinical Implications

The study's relevance lies in its provision of evidence on digital rehabilitation to expand access to rehabilitation for individuals with knee osteoarthritis in Nigeria and similar medically underserved contexts. The exercises used in this study were developed based on available evidence and expert opinions following a Delphi study. The interventions were applied in accordance with evidence from a systematic review, which showed that exercise programs lasting 8-12 weeks, with 3-5 sessions per week, are safe and effective for patients with knee osteoarthritis. This study aligns with recent recommendations for self-management in patients with chronic conditions [66], particularly for those with osteoarthritis [67]. Digital interventions have been beneficial in promoting self-management among patients with various conditions [68,69], including osteoarthritis [70,71]. Using asynchronous video-based exercises as digital rehabilitation for patients with knee osteoarthritis would improve access to physiotherapy services, eliminate geographical and distance barriers, and reduce the burden of clinic visits and waiting times.

Study Limitations

This study has some potential limitations that may affect the generalizability of findings. As such, some characteristics of the sample must be considered. First, the significant difference in the baseline parameter between the intervention and control group regarding the OAKHQoL domains, physical performance domain of IKHOAM, and least pain domain of QVAS may be a source of bias on the outcome of the interventions. Second, most participants were primarily middle-aged adults who were largely uneducated; this may affect the ease of use of technological tools. Third, this study used a dual clinical approach to recruiting participants, which could have potential inherent selection bias. Selecting multiple sites for this study may introduce selection bias related to patient demographics or treatment practices, which could independently influence the outcomes of the intervention. However, the findings of the study suggest that the general characteristics of patients at baseline were comparable across groups, indicating that the potential for selection bias is limited. Although both clinics are classified as tertiary health care institutions in Nigeria, the distinctive attributes of teaching hospitals often make them the preferred option for patients seeking specialized and comprehensive health care. As such, the control site, which is the teaching hospital in this case, may have more standardized treatment practices, potentially leading to performance bias and affecting treatment outcomes. Fourth, it is challenging to guarantee the adherence of the participants in the AVbSE group, as it is commonly reported in asynchronous tele-exercise programs, despite the telemonitoring used in this study. Also, asynchronous tele-exercises offer scheduling flexibility to patients but do not provide for real-time interface and response from the provider.

XSL•FO
RenderX

Furthermore, excluding patients without mobile phones from data collection may be a limitation of this study. While mobile phone ownership is relatively high worldwide, it is still a problem in specific populations, such as adults in rural or developing areas, and children [72]. In Nigeria, smartphone penetration is increasing and is expected to reach around 60% by 2025 [73]. However, mobile phone ownership in poorer areas is still limited, with only 58% of Nigerians in urban areas and 32% in rural areas owning smartphones in 2022 [74]. Therefore, the pervasive digital divide or digital poverty in many African settings hinders the continent from fully benefiting from digital technology [75]. For instance, the lack of mobile phone ownership could impede the implementation of mobile health interventions in Africa [76]. Researchers conducting self-report research using mobile phones should be aware that providing a mobile phone to standardize participant response platforms can impact response behavior [72].

Direction for Future Work

Future research investigating the short- and long-term effects of telerehabilitation among a larger population of patients with knee osteoarthritis from multiple sites is recommended to validate the findings of this study.

Conclusion

Clinic-based strengthening exercise circuit training and its asynchronous video-based variant have been found to effectively reduce pain, impairment, and disability and increase patients' QoL. The video-based variant was associated with higher improvement in most health-related QoL domains. In comparison, clinic-based strengthening exercise circuit training led to higher improvement in pain intensity.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author.

Conflicts of Interest

None declared.

References

1. Glyn-Jones S, Palmer AJR, Agricola R, Price AJ, Vincent TL, Weinans H, et al. Osteoarthritis. Lancet 2015;386(9991):376-387. [doi: 10.1016/S0140-6736(14)60802-3] [Medline: 25748615]
2. Bridges SL. National institute of arthritis and musculoskeletal and skin diseases. Arthritis Res Ther 2000;2(1) [FREE Full text] [doi: 10.1186/ar-2000-2-webreport0003]
3. Murphy L, Schwartz T, Helmick C, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum 2008;59(9):1207-1213 [FREE Full text] [doi: 10.1002/art.24021] [Medline: 18759314]
4. Litwic A, Edwards M, Dennison E, Cooper C. Epidemiology and burden of osteoarthritis. Br Med Bull 2013;105:185-199 [FREE Full text] [doi: 10.1093/bmb/lds038] [Medline: 23337796]
5. Akinpelu A, Alonge OO, Adekanla BA, Odole AC. Pattern of osteoarthritis seen in physiotherapy facilities in ibadan and lagos, Nigeria. African Journal of Biomedical Research Internet 2010;10(2) [FREE Full text] [doi: 10.4314/ajbr.v10i2.50612]
6. Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. EClinicalMedicine 2020;29-30:100587 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100587] [Medline: 34505846]
7. Kumar P, Clark ML. In: Feather A, Randall D, Waterhouse M, editors. Kumar and Clark's clinical medicine e-book. Edinburgh: Elsevier Health Sciences; 2012.
8. Chronic rheumatic conditions. World Health Organization. URL: https://www.who.int/chp/topics/rheumatic/en/ [accessed 2024-05-14]
9. Harris R, Strotmeyer ES, Sharma L, Kwoh CK, Brach JS, Boudreau R, et al. The association between severity of radiographic knee OA and recurrent falls in middle and older aged adults: the osteoarthritis initiative. J Gerontol A Biol Sci Med Sci 2023;78(1):97-103 [FREE Full text] [doi: 10.1093/gerona/glac050] [Medline: 35184161]
10. Abramoff B, Caldera F. Osteoarthritis: pathology, diagnosis, and treatment options. Med Clin North Am 2020;104(2):293-311. [doi: 10.1016/j.mcna.2019.10.007] [Medline: 32035570]
11. Leifer VP, Katz JN, Losina E. The burden of OA-health services and economics. Osteoarthritis Cartilage 2022;30(1):10-16 [FREE Full text] [doi: 10.1016/j.joca.2021.05.007] [Medline: 34023527]
12. Driban JB, Harkey M, Liu S, Salzler M, McAlindon TE. Osteoarthritis and aging: young adults with osteoarthritis. Curr Epidemiol Rep 2020;7(1):9-15 [FREE Full text] [doi: 10.1007/s40471-020-00224-7]
13. Dantas LO, Salvini TDF, McAlindon TE. Knee osteoarthritis: key treatments and implications for physical therapy. Braz J Phys Ther 2021;25(2):135-146 [FREE Full text] [doi: 10.1016/j.bjpt.2020.08.004] [Medline: 33262080]
14. Skou ST, Roos E. Physical therapy for patients with knee and hip osteoarthritis: supervised, active treatment is current best practice. Clin Exp Rheumatol 2019;37 Suppl 120(5):112-117 [FREE Full text] [Medline: 31621559]

XSL•FO
RenderX

15. Moseng T, Vliet Vlieland TPM, Battista S, Beckwée D, Boyadzhieva V, Conaghan PG, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis: 2023 update. Ann Rheum Dis 2024;83(6):730-740 [FREE Full text] [doi: 10.1136/ard-2023-225041] [Medline: 38212040]
16. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014;22(3):363-388 [FREE Full text] [doi: 10.1016/j.joca.2014.01.003] [Medline: 24462672]
17. Whittaker JL, Roos E. A pragmatic approach to prevent post-traumatic osteoarthritis after sport or exercise-related joint injury. Best Pract Res Clin Rheumatol 2019;33(1):158-171 [FREE Full text] [doi: 10.1016/j.berh.2019.02.008] [Medline: 31431269]
18. Ravalli S, Castrogiovanni P, Musumeci G. Exercise as medicine to be prescribed in osteoarthritis. World J Orthop 2019;10(7):262-267 [FREE Full text] [doi: 10.5312/wjo.v10.i7.262] [Medline: 31363456]
19. Lange AK, Vanwanseele B, Fiatarone Singh MA. Strength training for treatment of osteoarthritis of the knee: a systematic review. Arthritis Rheum 2008;59(10):1488-1494 [FREE Full text] [doi: 10.1002/art.24118] [Medline: 18821647]
20. Alghadir AH, Anwer S, Sarkar B, Paul AK, Anwar D. Effect of 6-week retro or forward walking program on pain, functional disability, quadriceps muscle strength, and performance in individuals with knee osteoarthritis: a randomized controlled trial (retro-walking trial). BMC Musculoskelet Disord 2019;20(1):159 [FREE Full text] [doi: 10.1186/s12891-019-2537-9] [Medline: 30967128]
21. DeVita P, Aaboe J, Bartholdy C, Leonardis JM, Bliddal H, Henriksen M. Quadriceps-strengthening exercise and quadriceps and knee biomechanics during walking in knee osteoarthritis: a two-centre randomized controlled trial. Clin Biomech (Bristol) 2018;59:199-206. [doi: 10.1016/j.clinbiomech.2018.09.016] [Medline: 30273922]
22. Wu Z, Zhou R, Zhu Y, Zeng Z, Ye Z, Wang Z, et al. Self-management for knee osteoarthritis: a systematic review and meta-analysis of randomized controlled trials. Pain Res Manag 2022;2022:2681240 [FREE Full text] [doi: 10.1155/2022/2681240] [Medline: 35281344]
23. Uritani D, Koda H, Sugita S. Effects of self-management education programmes on self-efficacy for osteoarthritis of the knee: a systematic review of randomised controlled trials. BMC Musculoskelet Disord 2021;22(1):515 [FREE Full text] [doi: 10.1186/s12891-021-04399-y] [Medline: 34090406]
24. Mbada CE, Oluwatobi S, Awosika H, Esan O, Oke O, Makinde M, et al. Development and feasibility testing of a smartphone video-based exercise program for patients with knee osteoarthritis. Digital Medicine Internet 2022;8(1):9 [FREE Full text] [doi: 10.4103/digm.digm_19_21]
25. Odetunde MO, Binuyo OT, Maruf FA, Ayenowowon SO, Okonji AM, Odetunde NA, et al. Development and feasibility testing of video home based telerehabilitation for stroke survivors in resource limited settings. Int J Telerehabil 2020;12(2):125-136 [FREE Full text] [doi: 10.5195/ijt.2020.6321] [Medline: 33520100]
26. Jordan JM, Helmick C, Renner J, Luta G, Dragomir AD, Woodard J, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and caucasians: the johnston county osteoarthritis project. J Rheumatol 2007;34(1):172-180. [Medline: 17216685]
27. Xie SH, Wang Q, Wang LQ, Zhu SY, Li Y, He CQ. The feasibility and effectiveness of internet-based rehabilitation for patients with knee osteoarthritis: a study protocol of randomized controlled trial in the community setting. Medicine (Baltimore) 2020;99(44):e22961 [FREE Full text] [doi: 10.1097/MD.0000000000022961] [Medline: 33126366]
28. Odole AC, Ojo OD. Is telephysiotherapy an option for improved quality of life in patients with osteoarthritis of the knee? Int J Telemed Appl 2014;2014:903816 [FREE Full text] [doi: 10.1155/2014/903816] [Medline: 24778645]
29. Morelli V, Roger Zoorob J. Primary care of the medically underserved, an issue of physician assistant clinics. Elsevier Health Sciences. 2018. URL: https://play.google.com/store/books/details?id=OTt7DwAAQBAJ&rdid=book-OTt7DwAAQBAJ&rdot=1&source=gbs_vpt_read&pcampaignid=books_booksearch_viewport [accessed 2025-01-16]
30. Brimmo FO, Babatunde AO, Ezefuna NN, Kanu MS, Biziyaremye P. The need for more medical schools in medically underserved regions in Africa. Ann Med Surg (Lond) 2022;84:104967 [FREE Full text] [doi: 10.1016/j.amsu.2022.104967] [Medline: 36426098]
31. World health statistics 2022: monitoring health for the SDGs.sustainable development goals. World Health Organization. 2022. URL: http://apps.who.int/bookorders [accessed 2025-01-16]
32. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
33. Frey MB, Chiu SH. Considerations when using telemedicine as the advanced practice registered nurse. J Nurse Pract 2021;17(3):289-292 [FREE Full text] [doi: 10.1016/j.nurpra.2020.11.011] [Medline: 33519314]
34. Healthcare in Africa. IFC report sees demand for investment. 2023. URL: https://www.ifc.org/en/home(Accessed [accessed 2023-08-15]
35. Chan YH. Randomised controlled trials (RCTs)-sample size: the magic number? Singapore Med J 2003;44(4):172-174. [Medline: 12952027]
36. Von Korff M, Deyo RA, Cherkin D, Barlow W. Back pain in primary care. Outcomes at 1 year. Spine (Phila Pa 1976) 1993;18(7):855-862. [doi: 10.1097/00007632-199306000-00008] [Medline: 8316884]

XSL•FO
RenderX

37. Odole AC, Odunaiya NA, Akinpelu AO. Ibadan knee/hip osteoarthritis outcome measure: process of development. Ann Ib Postgrad Med 2013;11(2):71-76 [FREE Full text] [Medline: 25161423]
38. Rat AC, Coste J, Pouchot J, Baumann M, Spitz E, Retel-Rude N, et al. OAKHQOL: a new instrument to measure quality of life in knee and hip osteoarthritis. J Clin Epidemiol 2005;58(1):47-55. [doi: 10.1016/j.jclinepi.2004.04.011] [Medline: 15649670]
39. Stenberg G, Ahlgren C. A gender perspective on physiotherapy treatment in patients with neck and back pain. Advances in Physiotherapy 2010;12(1):35-41 [FREE Full text] [doi: 10.3109/14038190903174270]
40. Anwer S, Alghadir A, Brismée JM. Effect of home exercise program in patients with knee osteoarthritis: a systematic review and meta-analysis. J Geriatr Phys Ther 2016;39(1):38-48. [doi: 10.1519/JPT.0000000000000045] [Medline: 25695471]
41. Yilmaz M, Sahin M, Algun ZC. Comparison of effectiveness of the home exercise program and the home exercise program taught by physiotherapist in knee osteoarthritis. J Back Musculoskelet Rehabil 2019;32(1):161-169. [doi: 10.3233/BMR-181234] [Medline: 30248040]
42. Onigbinde A, Akindoyi O, Faremi FA, Okonji AM, Oniyangi SO, Olaitan OL. An assessment of hamstring flexibility of subjects with knee osteoarthritis and their age matched control. CMR 2013;2(6):121 [FREE Full text] [doi: 10.11648/j.cmr.20130206.12]
43. Abdel-Aziem AA, Soliman E, Mosaad D, Draz AH. Effect of a physiotherapy rehabilitation program on knee osteoarthritis in patients with different pain intensities. J Phys Ther Sci 2018;30(2):307-312 [FREE Full text] [doi: 10.1589/jpts.30.307] [Medline: 29545702]
44. Nguyen C, Lefèvre-Colau MM, Poiraudeau S, Rannou F. Rehabilitation (exercise and strength training) and osteoarthritis: a critical narrative review. Ann Phys Rehabil Med 2016;59(3):190-195 [FREE Full text] [doi: 10.1016/j.rehab.2016.02.010] [Medline: 27155923]
45. Turner MN, Hernandez D, Cade W, Emerson CP, Reynolds JM, Best TM. The role of resistance training dosing on pain and physical function in individuals with knee osteoarthritis: a systematic review. Sports Health 2020;12(2):200-206 [FREE Full text] [doi: 10.1177/1941738119887183] [Medline: 31850826]
46. Li Y, Su Y, Chen S, Zhang Y, Zhang Z, Liu C, et al. The effects of resistance exercise in patients with knee osteoarthritis: a systematic review and meta-analysis. Clin Rehabil 2016;30(10):947-959. [doi: 10.1177/0269215515610039] [Medline: 26471972]
47. Doi, Akai M, Fujino K, Iwaya T, Kurosawa H, Hayashi K, et al. Effect of home exercise of quadriceps on knee osteoarthritis compared with nonsteroidal antiinflammatory drugs: a randomized controlled trial. Am J Phys Med Rehabil 2008;87(4):258-269. [doi: 10.1097/PHM.0b013e318168c02d] [Medline: 18356618]
48. Saengpromma P, Jirasakulsuk N, Khruakhorn S. The effectiveness of home-based exercise with and without tracking for people with knee osteoarthritis: a systematic review. Archives of Allied Health Sciences 2022;34:1-13 [FREE Full text]
49. Ettinger W. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. JAMA 1997 Jan 01;277(1):25-31 [FREE Full text] [doi: 10.1001/jama.1997.03540250033028]
50. Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. BMJ Internet. 2002. URL: https://www.bmj.com/content/325/7367/752 [accessed 2023-08-11]
51. Ojoawo OA, Odesanya O, Kunuji K, Mbada CE, Adedeji, OO. Effect of clinic-based and telemonitored home-based interventionon pain intensity, functioning and quality of life in patients with knee osteoarthritis. Eur J Clin Exp Med 2024;22(1):73-81. [doi: 10.15584/ejcem.2024.1.14]
52. Peck F, Jackson A, Marshall A. Digital health and its application in rural areas: a review of international experience. Cumbria Rural Health Forum. 2015. URL: https://insight.cumbria.ac.uk/id/eprint/4884/ [accessed 2025-01-16]
53. Ji H, Dong J, Pan W, Yu Y. Associations between digital literacy, health literacy, and digital health behaviors among rural residents: evidence from Zhejiang, China. Int J Equity Health 2024;23(1):68 [FREE Full text] [doi: 10.1186/s12939-024-02150-2] [Medline: 38594723]
54. Velez M, Lugo-Agudelo LH, Patiño Lugo DF, Glenton C, Posada AM, Mesa Franco LF, et al. Factors that influence the provision of home-based rehabilitation services for people needing rehabilitation: a qualitative evidence synthesis. Cochrane Database Syst Rev 2023;2(2):CD014823 [FREE Full text] [doi: 10.1002/14651858.CD014823] [Medline: 36780267]
55. JIT A. National Telehealth and Telecare Delivery Plan for Scotland to 2015. Scotland: NHS Scotland; 2012.
56. Russell TG. Physical rehabilitation using telemedicine. J Telemed Telecare 2007;13(5):217-220. [doi: 10.1258/135763307781458886] [Medline: 17697506]
57. Bernard M, Janson F, Flora P, Faulkner G, Meunier-Norman L, Fruhwirth M. Videoconference-based physiotherapy and tele-assessment for homebound older adults: A Pilot Study. Activities, Adaptation & Aging 2009;33(1):39-48 [FREE Full text] [doi: 10.1080/01924780902718608]
58. Chandra I, Keerthi R. Can telerehabilitation add a new dimension in the treatment of osteoarthritis knee? J Pain Relief. 2013. URL: https://www.omicsgroup.org/journals/

XSL•FO
RenderX

can-telerehabilitation-add-a-new-dimension-in-the-treatment-of-osteoarthritis-knee-2167-0846.1000113.php?aid=9484 [accessed 2023-08-14]
59. Bennell KL, Egerton T, Bills C, Gale J, Kolt GS, Bunker SJ, et al. Addition of telephone coaching to a physiotherapist-delivered physical activity program in people with knee osteoarthritis: a randomised controlled trial protocol. BMC Musculoskelet Disord 2012;13:246 [FREE Full text] [doi: 10.1186/1471-2474-13-246] [Medline: 23231928]
60. Adhikari SP, Shrestha P, Dev R. Feasibility and effectiveness of telephone-based telephysiotherapy for treatment of pain in low-resource setting: A Retrospective pre-post design. Pain Research and Management Internet. 2020. URL: https://www.hindawi.com/journals/prm/2020/2741278/ [accessed 2023-08-11]
61. Multani NK, Singh B, Garg S. Effectiveness of telemedicine services integrated into physiotherapeutic health care system. Journal of Exercise Science and Physiotherapy Internet. 2020. URL: https://search.informit.org/doi/abs/10.3316/informit.865790825652467 [accessed 2023-08-11]
62. Winters JM. Telerehabilitation research: emerging opportunities. Annual Review of Biomedical Engineering Internet. 2002. URL: https://doi.org/10.1146/annurev.bioeng.4.112801.121923 [accessed 2023-08-11]
63. Lamplot JD, Pinnamaneni S, Swensen-Buza S, Lawton CD, Dines, JS, Nawabi DH, et al. The knee examination for video telemedicine encounters. HSS Journal® Internet. 2021. URL: https://doi.org/10.1177/1556331620975039 [accessed 2023-08-11]
64. Chumbler NR, Quigley P, Li X, Morey M, Rose D, Sanford J, et al. Effects of telerehabilitation on physical function and disability for stroke patients. Stroke Internet. 2012. URL: https://www.ahajournals.org/doi/abs/10.1161/strokeaha.111.646943 [accessed 2023-08-14]
65. Saywell N, Vandal AC, Brown P, Hanger HC, Hale L, Mudge S, et al. Telerehabilitation to improve outcomes for people with stroke: study protocol for a randomised controlled trial. Trials Internet. 2012. URL: https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-13-233 [accessed 2023-08-14]
66. Martz E. Promoting Self-Management of Chronic Health Conditions: Theories and Practice. United Kingdom: Oxford University Press; 2017.
67. Wu Z, Zhou R, Zhu Y, Zeng Z, Ye Z, Wang Z, et al. Self-management for knee osteoarthritis: a systematic review and meta-analysis of randomized controlled trials. Pain Res Manag 2022;2022:2681240 [FREE Full text] [doi: 10.1155/2022/2681240] [Medline: 35281344]
68. Mshali H, Lemlouma T, Moloney M, Magoni D. A survey on health monitoring systems for health smart homes. International Journal of Industrial Ergonomics 2018;66:26-56. [doi: 10.1016/j.ergon.2018.02.002]
69. Sheng Y, Doyle J, Bond R, Jaiswal R, Gavin S, Dinsmore J. Home-based digital health technologies for older adults to self-manage multiple chronic conditions: a data-informed analysis of user engagement from a longitudinal trial. Digit Health 2022;8:20552076221125957. [doi: 10.1177/20552076221125957] [Medline: 36171962]
70. Patten RK, Tacey A, Pile R, Parker A, De Gori M, Tran P, et al. Digital self-management interventions for osteoarthritis: a systematic scoping review of intervention characteristics, adherence and attrition. Arch Public Health 2022;80(1):103 [FREE Full text] [doi: 10.1186/s13690-022-00854-x] [Medline: 35361270]
71. Xiang W, Wang JY, Ji BJ, Li LJ, Xiang H. Effectiveness of different telerehabilitation strategies on pain and physical function in patients with knee osteoarthritis: systematic review and meta-analysis. J Med Internet Res 2023;25:e40735 [FREE Full text] [doi: 10.2196/40735] [Medline: 37982411]
72. Walsh EI, Brinker JK. Should participants be given a mobile phone, or use their own? Effects of novelty vs utility. Telematics and Informatics 2016;33(1):25-33. [doi: 10.1016/j.tele.2015.06.006]
73. Number of mobile internet users in Nigeria from 2020 to 2029. Statista. 2024. URL: https://www.statista.com/statistics/972896/nigeria-mobile-internet- [accessed 2025-01-16]
74. The mobile economy sub-saharan Africa2023. GSMA. URL: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/sub-saharan-africa/ [accessed 2025-01-16]
75. Africagoes digital. Duarte. 2021. URL: https://www.imf.org/en/Publications/fandd/issues/2021/03/africas-digital-future-after-COVID19-duarte [accessed 2025-01-16]
76. Okano JT, Ponce J, Krönke M, Blower S. Lack of ownership of mobile phones could hinder the rollout of mHealth interventions in Africa. Elife 2022;11 [FREE Full text] [doi: 10.7554/eLife.79615] [Medline: 36255055]

Abbreviations

AvbSE: asynchronous video-based strengthening exercises
CbSE: clinic-based strengthening exercises
CONSORT: Consolidated Standards of Reporting Trials
IKHOAM: Ibadan Knee and Hip Osteoarthritis Outcome Measure
OAKHQoL: Osteoarthritis Knee and Hip Quality of Life
QoL: quality of life
QVAS: Quadruple Visual Analog Scale

XSL•FO
RenderX

Edited by A Coristine; submitted 14.03.24; peer-reviewed by A Arntz, F Ozden, J Zhang; comments to author 02.05.24; revised version received 20.06.24; accepted 13.11.24; published 26.03.25.

Please cite as:
Mbada CE, Awosika HA, Sonuga OA, Akande M, Gebrye T, Woolf R, Fatoye F
Effect of Clinic-Based and Asynchronous Video-Based Exercise on Clinic and Psychosocial Outcomes in Patients With Knee Osteoarthritis: Quasi-Experimental Study
J Med Internet Res 2025;27:e58393
URL: https://www.jmir.org/2025/1/e58393
doi:10.2196/58393
PMID:

©Chidozie E Mbada, Henry Akintunji Awosika, Oluwatobi Ademola Sonuga, Micheal Akande, Tadesse Gebrye, Richard Woolf, Francis Fatoye. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Engaging End Users to Inform the Design and Social Marketing Strategy for a Web-Based Sexually Transmitted Infection/Blood-Borne Virus (STI/BBV) Testing Service for Young People in Victoria, Australia: Qualitative Study

Ethan T Cardwell[1], BSc, MPH; Teralynn Ludwick[1], MPH, PhD; Shanton Chang[2], PhD; Olivia Walsh[1], MPH; Megan Lim[1,3,4], PhD; Rachel Podbury[5], BPH; David Evans[6,7], PhD; Christopher K Fairley[8,9], MBBS, PhD; Fabian Y S Kong[1], BPharm, PhD; Jane S Hocking[1], PhD

[1]Melbourne School of Population and Global Health, Centre for Epidemiology and Biostatistics, The University of Melbourne, Parkville, Australia
[2]School of Computing and Information Systems, University of Melbourne, Parkville, Australia
[3]Burnet Institute, Melbourne, Australia
[4]School of Population Health and Preventive Medicine, Monash University, Melbourne, Australia
[5]Paper Giant, Melbourne, Australia
[6]Matanuska-Susitna College, University of Alaska System, Anchorage, United States
[7]Department of Rural Health, Center For Excellence in Rural Sexual Health, The University of Melbourne, Shepparton, Australia
[8]Alfred Health, Melbourne Sexual Health Centre, Melbourne, Australia
[9]School of Translational Medicine, Monash University, Melbourne, Australia

Corresponding Author:
Ethan T Cardwell, BSc, MPH
Melbourne School of Population and Global Health
Centre for Epidemiology and Biostatistics
The University of Melbourne
207 Bouverie Street, Level 3
Parkville
Australia
Phone: 61 3 9035 3039
Email: t.cardwell@unimelb.edu.au

Abstract

Background: The rates of sexually transmitted infections (STIs) continue to rise across Australia among 16- to 29-year-olds. Timely testing is needed to reduce transmission, but sexual health clinics are at capacity. This demand, coupled with barriers to getting tested faced by young people, has led to web-based services as a pragmatic solution. However, for young people to use these services, they must be acceptable, attractive, and usable. Social marketing principles combined with end user engagement can be used to guide the development of a web-based service and create a marketing strategy to attract them to the service.

Objective: Working closely with end users and guided by social marketing, this project explored messaging, design elements (imagery), and promotional strategies that will support high usage of a web-based STI/blood-borne virus (BBV) testing service for young people in Victoria, Australia.

Methods: Young people were recruited to participate in half-day workshops via youth organizations and targeted Meta (Facebook/Instagram) advertisements. An initial web-based survey was deployed to inform workshop content. Workshops were held in metropolitan, outer metropolitan, and regional Victoria. Young people were presented with a range of "image territories" developed by a social marketing firm and social marketing messages that were informed by the literature on communicating health messages. Participants discussed the feelings and reactions evoked by the content. Data collected through mixed methods (transcribed notes, audio recording, and physical outputs) were thematically analyzed to understand features of messaging and imagery that would attract young people to use the service.

Results: A total of 45 people completed the initial survey with 17 participating in focus group workshops (metropolitan: n=8, outer metropolitan: n=6, and regional: n=3). Young people preferred messages that highlight the functional benefits (confidential,

XSL•FO
RenderX

affordable, and accessible) of a web-based service and include professional imagery and logos that elicit trust. Young people indicated that the service should be promoted through digital communications (eg, dating apps and social media), with endorsement from government or other recognized institutions, and via word-of-mouth communications.

Conclusions: This study has highlighted the value of applying social marketing theory with end user engagement in developing a web-based STI/BBV testing service. Through the voices of young people, we have established the foundations to inform the design and marketing for Victoria's first publicly funded web-based STI/BBV testing clinic. Future research will measure the reach and efficacy of social marketing, and how this service complements existing services in increasing STI/BBV testing uptake among young Victorians.

(J Med Internet Res 2025;27:e63822) doi:10.2196/63822

KEYWORDS

web-based STI/HIV testing; social marketing; sexual health; participatory design; codesign; sexually transmitted infections; STI; HIV; Australia; social media; survey; blood-borne viruses

Introduction

Background

After a drop-off during the COVID-19 pandemic, sexually transmitted infections (STI) diagnosis rates are rising again throughout high-income countries and the burden of these infections is particularly high among young people aged 16-29 years [1]. The increased demand for STI and blood-borne virus (BBV) testing has placed a strain on sexual health services and led to the emergence of web-based services that are now increasingly common in the United States, United Kingdom, Canada, and Europe [2] as a solution for increasing access to testing.

While web-based services can be effective at increasing access to STI testing, they are not always acceptable or accessible to particular population groups, such as those living in regional or remote areas, and this is often because they have been developed with little engagement from the end user [3]. This lack of engagement with end users can explain usability problems, low uptake, and high attrition rates [3]—people simply stop using technologies that do not correspond with their daily lives or habits. Social marketing theory [4] and end user engagement [5] approaches are valuable for putting the focus on the end user, increasing the effectiveness, accessibility, and sustainability of web-based technologies [6]. Few studies have applied social marketing theory with end user engagement; however, researchers in the field have acknowledged the use of end user engagement and social marketing theory individually and the potential for synergy when used together [7,8].

Development of a Web-Based STI/BBV Testing Service in Australia

To address the growing demand for access to STI/BBV testing in Australia, we are developing a fully automated web-based STI/BBV testing service to be offered as part of the only specialist sexual health service in Victoria, Australia, based in Melbourne (population of 5 million), the capital city of the State of Victoria (population of 6.7 million). Given the location of the specialist service in Melbourne, people living in regional Victoria have limited access to specialist STI/BBV care, placing them at considerable disadvantage. The aim of our web-based service is to increase STI/BBV testing uptake by providing access to specialist STI/BBV expertise anywhere across the State of Victoria, not just in Melbourne, particularly for young people aged 16-29 years. Our service will allow users to login to a website portal, answer risk-based consultation questions, receive a pathology request according to recommended testing guidelines [9], and have their results sent directly to them.

Development and implementation of our web-based service have five phases: (1) formative research in which we surveyed young people and reviewed existing web-based services in Australia to inform the design of our service [10-13]; (2) end user engagement in which we ran focus group workshops (workshops) to design the content and the tone of voice (language used), imagery, and branding of the service; (3) prototyping in which we are developing digital prototypes based on the findings from phases 1 and 2; (4) user testing in which we test the prototype with end users to assess its usability, acceptability, and refine development; and (5) evaluation in which we implement the web-based service and assess its impact on STI/BBV testing and diagnosis and its cost-effectiveness (Figure 1).

In this paper, we present the methods and results for phase 2, end user engagement, in which we had direct and iterative involvement with young people to explore the messaging and imagery design elements that would attract young people to use a web-based STI/BBV testing service and the best way to promote this messaging to young people. Further publications will detail our research in developing the user experience along the clinical pathway (part of phase 2), user testing of our web-based service (phase 4), and its subsequent evaluation (phase 5).

XSL•FO
RenderX

Figure 1. Five-phase approach to user-centered development of a web-based sexually transmitted infection/blood-borne virus testing service for young people aged 18-29 years in Victoria, Australia.

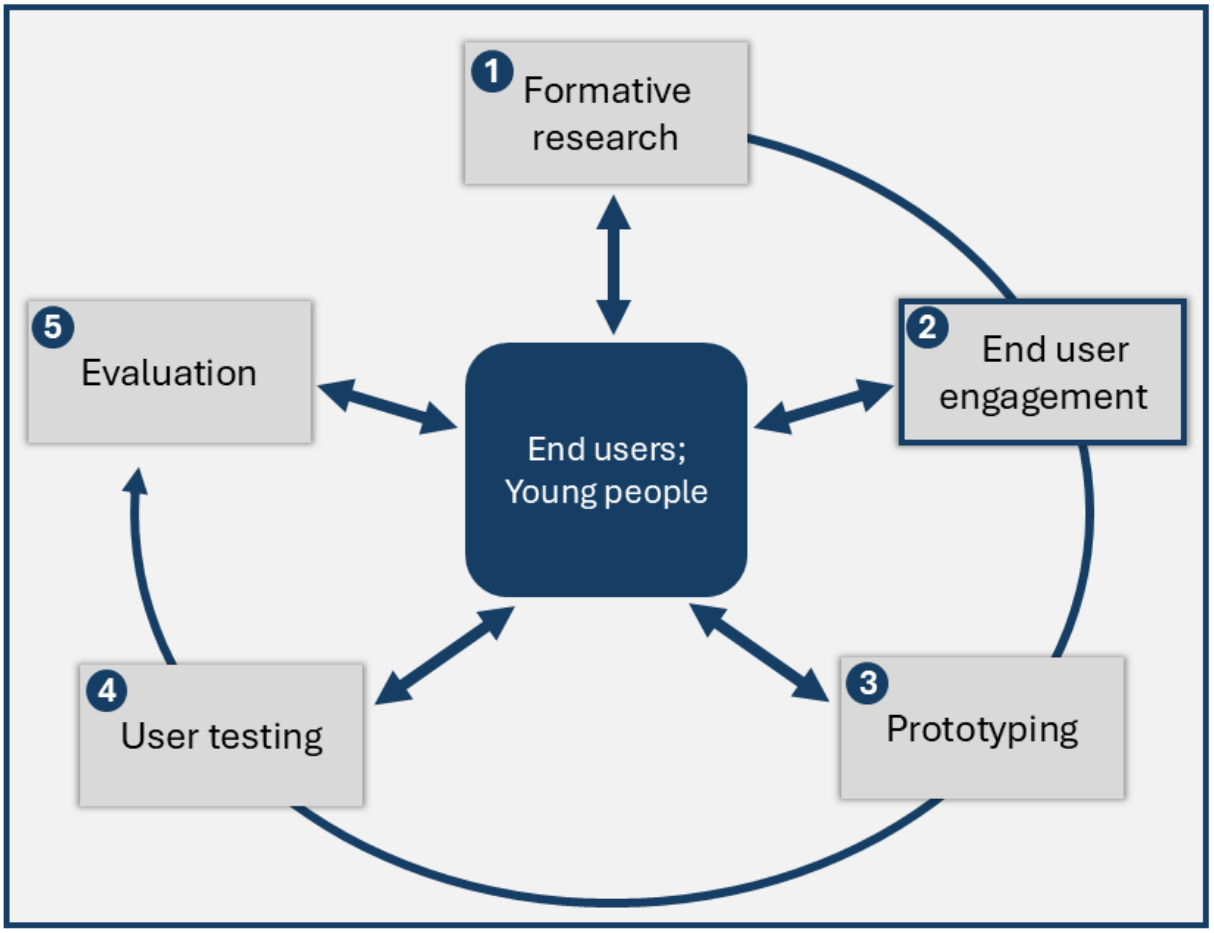

Methods

Study Design

In phase 2, we have used Andreasen's theory of social marketing [4] to inform our approach to develop the imagery and messaging for our web-based STI/BBV service and have applied Andreasen's six benchmarks as follows:

1. Behavior change: in our context, we want to ultimately increase STI/BBV test uptake among young people using the service; however, this will not be assessed until phase 5.
2. Formative research: we conducted surveys and interviews and reviewed available web-based services to understand young people's experiences, values, and needs of a web-based service.
3. Segmentation: we targeted particular subgroups of young people to capture their views (viz, young people living in metropolitan and regional areas) based on our formative research [12], which showed differences in preferences between regional and metropolitan youth for how web-based STI/BBV services are delivered. By targeting these groups, we sought to understand how these differences would apply to the design and social marketing of our web-based service.
4. Marketing mix: we assess the service we are providing and how to price, place, and promote it.
5. Exchange: we assess what will motivate young people to engage with the service and what service they will get in return.
6. Competition: we assess competing behaviors and strategies to help us understand what would make a young person use our web-based testing service in preference to other testing services [4]. Our formative research showed that young people want to be able to access web-based STI/BBV testing services in addition to in-person services [12] because of limited availability of specialist in-person sexual health services and the time delays and cost often associated with accessing other in-person primary care services in Australia [2,10]. However, there are few web-based STI/BBV testing services in Victoria, and those that are available are rarely free, nor do they offer a comprehensive STI/BBV testing service in which all guideline-recommended tests are available [10].

Although our web-based service will be directed to people aged 16-29 years, people aged 18-29 years living in the State of Victoria were eligible for participation in phase 2 because of consent challenges when working with 16- and 17-year-olds in focus groups. Phase 2 involved three key steps:

XSL•FO
RenderX

1. An expression of interest (EOI) in which young people were invited to express their interest in being involved in the study. The EOI was advertised through youth service organizations across Victoria and targeted advertising on Meta (Facebook and Instagram). Those interested completed a brief web-based survey, which collected information about sociodemographics including their age, gender, sexuality, and residential postcode. The purpose of the EOI was to allow for purposive sampling for the focus group workshops, ensuring diverse representation of young people (eg, gender and sexuality) from across Victoria including metropolitan and regional locations. We reviewed EOIs in real time and placed additional advertisements targeting both young people living in regional areas and heterosexual men to attract more applicants from these demographic backgrounds as needed.
2. A self-administered web-based survey was completed by a select sample of respondents from the EOI. The demographic profile of EOI responses was screened as they were received, with approximately 15 individuals from each geographic location being invited to participate. The survey explored (1) social marketing including the preferred tone of voice, types of messaging, and image territories (group of images representing a design theme) that would attract young people to use a web-based STI/BBV testing service, and (2) the "user experience" along the clinical pathway from establishing a web-based account, answering sexual history questions, and receiving an STI/BBV form for testing. These survey responses were used to refine the messaging and image territories to explore in the focus group workshops. All individuals who completed the survey were invited to participate in the focus group workshops.
3. Focus group workshops, in which we further explored their preferred social marketing and user experience to inform the development of the web-based service with a sample of young people who completed the web-based survey. Workshop participants provided electronic or written consent and were reimbursed for their time. We conducted three 6-hour workshops that were segmented by geographical area with 1 in metropolitan area, 1 in outer metropolitan area, and 1 in regional areas. Metropolitan and outer metropolitan workshops were conducted on the university campus and the regional workshop was conducted at community organization in regional Victoria.

Testing Social Marketing Messaging

The web-based survey explored participants' views about 8 social marketing messages that we developed informed from the literature on communicating health messages to young people [14-16]. Each message included aligns with one of the following approaches to messaging: provision of choice (provides choice between 2 or more behaviors), gain- or loss-framed (emphasizes the advantage or consequence of an action), value based (elicits a sensation of value), and other referencing (focuses on how an individual's choice may affect others) [14-16]. A full list of messages can be found in Multimedia Appendix 1.

Each participant was asked to imagine themselves in 7 different scenarios and then pick the top 3 messages that would encourage them to seek a web-based STI/BBV testing service. Scenarios placed the user in an experience known from the literature as a facilitator or barrier to getting tested, such as stigma, self-efficacy, or risk perception [17-19]. These scenarios were based on the theory of planned behavior and the transtheoretical model [20]. A full list of the scenarios can be found in Multimedia Appendix 1.

The following four messages were the most selected:

- Message 1: Getting tested can protect your partners from STIs.
- Message 2: Confidential, affordable STI testing. Find out more!
- Message 3: Itching. Burning. Rash. Sores. Getting tested is a MUST!
- Message 4: Be healthy. Be safe. Get tested.

In the workshop, participants were presented with each message individually and asked their initial reactions to the message, how it may influence them to use the service, and how it may be improved. After all messages were presented, the participants voted on which of the 4 messages they liked the most.

Testing Image Territories

In our survey, we presented the participants with some examples of web-based service-landing pages (the first web page users see after clicking a link to a testing website) of 2 currently existing sexual health websites and asked them to choose which they liked best (Multimedia Appendix 1). They were then asked what they liked about the website they chose and how it could be improved. We then engaged a social marketing firm that developed 3 "image territories" (medical, relationship, and abstract) to test with young people (Figure 2). During our workshops, we explored how participants felt about each image territory and which one ultimately attracted them most to use this type of service.

XSL•FO
RenderX

Figure 2. Three image territories used to explore design preferences during focus group workshops with young people aged 18-29 years in Victoria, Australia.

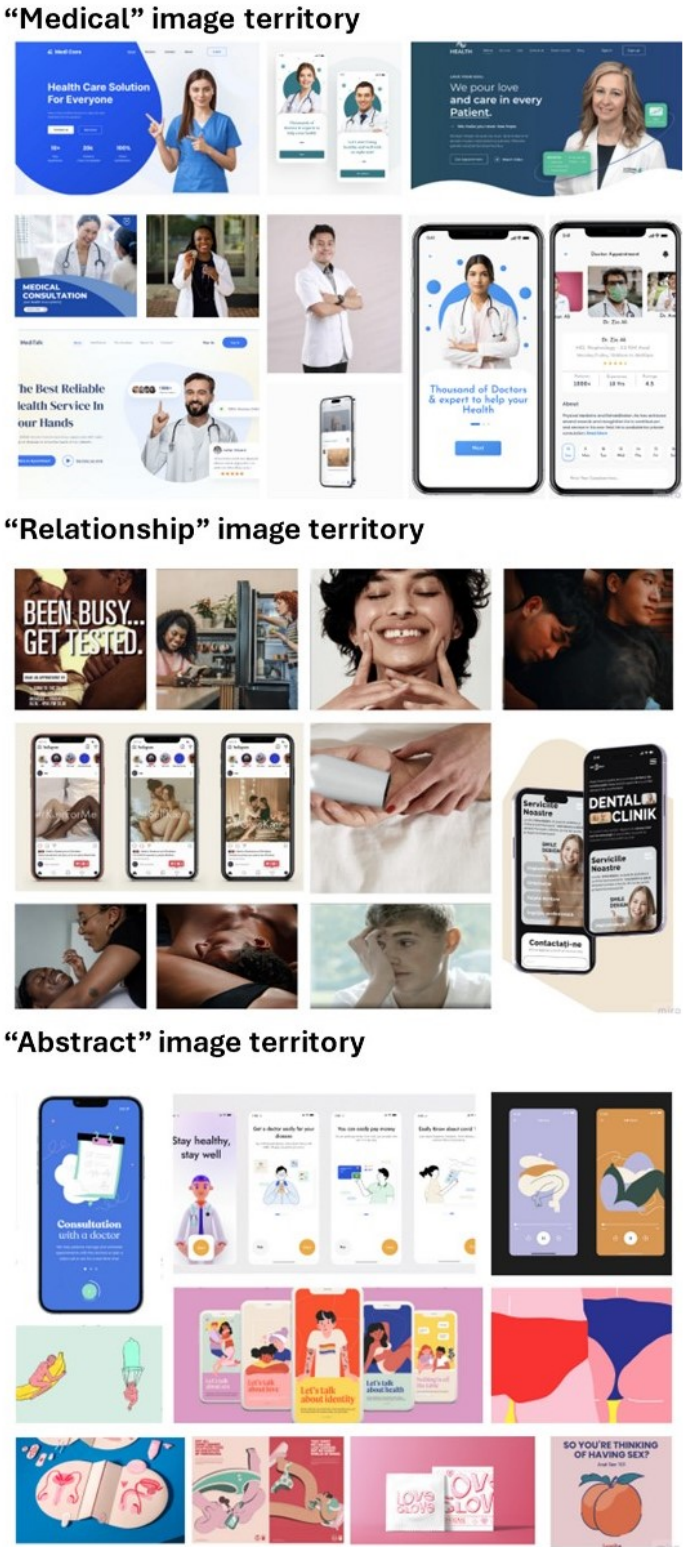

Informing a Marketing or Promotion Strategy

To inform a strategy to attract young people to use the service, we explored (1) how participants find out about health services in their day-to-day life, (2) what factors influence their trust of these services, and (3) how they would like to find out about this service.

Analysis

Data collected from the focus group workshops (transcribed notes, audio recording, and physical outputs) were thematically analyzed in Excel by ETC in consultation with SC using an inductive approach [21]. We validated high-level data with those attending the workshop (OW, RP, and TL). Illustrative quotes are presented by geographical location of the workshops. During

XSL•FO
RenderX

the workshops the service was referred to as "online" or "web-based" interchangeably for simplicity.

Ethical Considerations

All activities conducted in this study were approved by the Office of Research Ethics and Integrity at the University of Melbourne (2023-25277-43562-5). All participants were provided with plain language statements and signed an informed consent prior to participation in the study. The consent outlined the activities of the study and informed participants that all data collected would be deidentified and securely stored with password protection. Participants received Aus $20 (Aus $1.00=US $0.63) for completing the initial survey and an additional Aus $100 (Aus $1.00=US $0.63) electronic gift card for each 3-hour workshop they completed for a maximum remuneration of Aus $220 (Aus $1.00=US $0.63).

Results

Participants

A total of 45 individuals completed the initial survey and all were invited to attend a focus group workshop with 17 participating—8 in metropolitan, 6 in outer metropolitan, and 3 in regional locations. Age of workshop participants ranged from 18 to 29 years, with diverse gender (6 women, 9 men, 1 gender diverse, and 1 transgender), sexual orientations (6 heterosexual and 9 same-sex–attracted), and culturally and linguistically diverse backgrounds (7 culturally and linguistically diverse individuals) represented. Results from the focus group workshops are shown in the subsequent section.

Social Marketing Messaging

Four themes emerged from analysis: functional benefits, tone, structure, and altruistic versus individual motivations. These themes were mentioned consistently across all workshops, and there was a general consensus among participants about the messaging they preferred. Participants had more in-depth discussion about the importance of functional benefits and tone than about the themes of structure and individual versus individual motivations.

Functional Benefits—Confidentiality, Affordability, and Accessible Web-Based STI Testing

Functional benefits refer to benefits that particularly relate to the performance of the service and subthemes of confidential, affordable, and specificity (online and STI testing), which were frequently mentioned as important to participants. Participants across all workshops made many comments about the importance of messaging having the functional benefits of the service. They felt that the addition of words "online" or "STIs" would be more descriptive of the service and keep messages from coming off as too "generic."

> *[Be healthy. Be safe. Get tested.] I feel like context is important to make it obvious you're talking about STIs.* [Metropolitan]

It was clear that participants found it important to include aspects of the service that would be seen as valuable to users such as "confidential" and "affordable." This was one of the key reasons for the last message, "Confidential, affordable STI testing. Find out more!" being their favorite.

> *[Confidential, affordable STI testing. Find out more!] Perfect, clear and concise, reiterates the service is confidential, affordable. Assures people can access, like university students.* [Regional]

Although we did not test a message that was specific to accessibility, participants indicated that the word "online" had implicit connotations of being easier to access in comparison with in-person services. In particular, the regional young people spoke about the difficulty of accessing timely health care in their area due to overstretched health care services.

> *Trying to find a doctor in a rural town is very hard because they're not taking new clients.* [Regional]

Tone—Preference for a Nonjudgmental, Comforting Tone

Tone refers to the perception of how the messages come across to the reader with subthemes of commanding, passive, judgmental, comfort, and gimmicky identified. The tone used was brought up across all focus group workshops for every message. Participants did not like messages they perceived to be judgmental or gimmicky. Across all workshops the message, "Itching. Burning. Rash. Sores. Getting tested is a MUST!" was seen as judgmental and not a message they wanted to see. Some participants found certain phrases in the messages to be a bit judgy such as "Be healthy" or gimmicky such as "Find out more!"

> *[Itching. Burning. Rash. Sores. Getting tested is a MUST!] I don't like it at all. If I already had itching and burning, I would already feel uncomfortable and awkward and wouldn't wanted to be shouted at.* [Metropolitan]

In addition, participants discussed the importance of striking a balance between a "passive" and "commanding" tone. The message should not come off as "yelling" or "commanding" but should not be so passive that it does not motivate young people to get tested using the service.

> *[Be healthy. Be safe. Get tested.] I think it's good that it's not fear mongering but maybe too passive.* [Metropolitan]

Structure—Preference for a Short and Snappy Message Structure With Inclusive Wording

Structure refers to the basic elements of the message with subthemes of length, grammar, and wording. The way the message was structured, such as length and grammar, was less important but played a role in how the message was perceived. Participant overall preferred when messages were "short and snappy" and felt that the messages should easily be consumed "if you're walking past." Punctuation was mostly important to participants because it affected the tone of the message. They found that the use of an exclamation and even a period in some places came off as too commanding or gimmicky.

XSL•FO
RenderX

> *[Itching. Burning. Rash. Sores. Getting tested is a MUST!] Exclamation is too much, feels like a WWII poster.* [Metropolitan]

Participants also highlighted the importance of word choice and how certain words such as "partners" may have a "different meaning to different people" and cause some confusion as if the service is right for them.

> *I know a lot of people would just look past it and say, I don't have a partner, I've hooked up with 3 people in 3 days.* [Regional]

Altruistic Versus Individualistic Motivations

This theme refers to participants' comments on the focus of the message being on the individuals themselves or altruistic in nature and how this affects their motivation to get tested. Most comments in this theme related to the message, "Getting tested can protect your partners from STIs." There were mixed opinions from participants as to whether a message focused on the individuals themselves or focused on protecting another person would be more motivating for them.

> *It's focused around other people so I think it would appeal to me more or influence me more to get tested.* [Outer metropolitan]

> *You'll miss selfish people because [they would] possibly be interested in the service for selfish reasons.* [Metropolitan]

Imagery—a Balance of Medical and Abstract Image Territories

Upon thematic analysis, 2 general themes of design neutrality and traditional and modern design emerged across all image territories. Neutral design refers to the generalizability of the design to young people from different backgrounds with young people preferring gender neutral color schemes. Traditional design refers to the professional medical advertisements that have been traditionally used for medical services, where modern refers to more youthful, abstract designs. Participants wanted to find a balance between the 2 as each had its own benefit.

> *I think it's like any other doctor booking website service so if people respond to that then you can use the traditional methods. Why make it look like everything else when it's targeting a niche[r] market.* [Metropolitan referring to the Medical image territory]

There were also themes for each of the image territories. The medical image territory used stock images of medical professionals and traditional depictions of the web-based clinical setting with blue and white color scheme with themes of trust and generic emerging. Although participants found these images to be generic "like any other medical website," they did give participants a feeling of trust for the same reason, as they align with familiar medical service websites.

> *Very professional—too professional. It seems with the doctors and medical professionals makes it reliable and trustable.* [Outer metropolitan referring to the Medical image territory]

The relationship image territory used humanizing images of young people with some appearing to be in relationships and an earth tone color scheme. Participant found the color scheme to be too "luxe" and the images to be overly sexualized like it was "for a dating app."

> *I won't recommend this design at all. I just feel like it would be a dating and relationship website.* [Outer metropolitan referring to the Relationship image territory]

The abstract image territory used an animated approach to imaging with a pastel color scheme and themes of animation and comfort emerged. Participants found the animated images to bring them a sense of comfort as the images are more "relaxed" and relatable to other websites they use and like. However, they did find them feminine and geared toward "young women."

> *Makes me feel more comfortable. I like the animated style, it doesn't feel as daunting. The one thing I didn't like—the colour combos make it feel more targeted to younger females.* [Regional referring to the Abstract image territory]

Promoting and Marketing the Web-Based Service

When asked "How do you find out about health services in your day-to-day life?" participants mentioned social media, word of mouth from friends or health care providers, general searches on Google or specified health engines, and collateral materials posted at various physical locations. Using a search engine and through social media, including advertisements, were mentioned most by young people.

> *I've got ads that were given to me for health services that I wasn't following. I think it's helpful because you're only getting them for a reason.* [Metropolitan]

We then asked participants how they would like to find out about this service particularly. Participants said similar methods to finding out about health services in general but focused more on digital advertising such as social media and dating apps.

> *Dating apps or porn websites; you're going to where your audience is.* [Metropolitan]

We explored what influences their trust in these services. Participants commented that endorsement from friends, partnerships with existing professional organizations, and web-based reviews are important influences for trusting health services, particularly those on the web. It was clear that if participants had a certain level of trust with who they hear about the service from or who it is partnered with they were more likely to trust the service.

> *Word of mouth; people only go to practitioners they've had a good experience with.* [Metropolitan]

> *If it was endorsed by the government or some big hospitals then I would trust.* [Outer metropolitan]

XSL•FO
RenderX

Discussion

Principal Findings

Combining social marketing theory with end user engagement, we were able to identify key messaging and imagery that would attract young people to use a web-based STI/BBV testing service. We found that young people preferred messages that highlight the functional benefits (confidential, affordable, and accessible) of a web-based service and include youthful imagery and professional logos that elicit trust. In addition, we learnt that young people would want this service to be promoted through digital communications (eg, dating apps and social media), endorsement from government or other recognized institutions, and via word-of-mouth communications.

This study is one of the few studies to examine and publish findings from the design phase of developing a digital sexual health intervention and one of the only studies to publish findings on the desired language and imagery of young people as the end user [22-25]. Many studies have outlined the design process but have not clearly defined language or imagery elements of the design [26-28]. However, the summative findings from these studies reflect our findings and establish that young people want to use a web-based service but require a certain level of clarity around the confidentiality and credibility of the service—they want to trust the service [26-28].

Social Marketing and End User Engagement

Social marketing principles informed our study design and end user engagement proved useful for exploring the needs of young people. The workshops were also able to help us reaffirm our earlier findings and assumptions important to designing a service that is attractive and useful for young people. End user engagement has been used for the development of many interventions in sexual health and exemplifies the importance of involving end users and this study is no different [29]. By examining the rationale behind young people's preferences in a focus group setting, we are able to better tailor a youthful yet trustworthy service to young people. It is important to note that this is just the first step in the development of this service. Engaging end users in design is an iterative process and will require further involvement with young people through further user testing of prototypes created and feedback throughout implementation.

In this study, we focused on Andreasen's principles of segmentation, exchange, and the marketing mix (product, place, price, and promotion) (Multimedia Appendix 2). Segmentation was an important principle for us to explore in this study, as our formative research suggested that some populations expressed different preferences for using web-based STI/BBV testing services [12]. While many social marketing researchers have shifted away from the use of geographic and demographic segmentation because of its failure to predict behavior change [30], we included geographic segmentation because our earlier research found that young people living in regional areas expressed a stronger preference for web-based STI/BBV testing services than metropolitan young people [12]. However, in our focus group workshops, we found overall agreement among regional and metropolitan young people on the elements of design and promotion, although regional participants placed more value on the collateral materials (eg, posters) to promote the service. This suggests that while regional youth may be more likely to want to access a web-based rather than in-person service, their preferences for the design of the service are similar to metropolitan youth.

Ideally, we would have also included segmentation on psychographic factors such as personality, social status, or opinions, but this type of segmentation is expensive and time consuming to implement in practice, given the unreliability of and complexity in collecting these data and not possible within the budgetary constraints of our project [31]. In fact, most studies do not even report on segmentation at all [32].

In our exploration of the design of the service (product) and it being web-based (place), we found that young people continuously mentioned the importance of establishing trust among potential users. Whether directly or indirectly, participants made it clear that explicit messages and imagery highlighting the confidentiality and credibility of the service were critical in establishing the level of trust needed to use the service.

Andreasen's benchmark "exchange" (ie, motivation to engage with the service for the offer of something beneficial in return) [4] was indirectly explored throughout the study with the most interesting finding being that our traditional assumption of STI/BBV behavior being intrinsically motivated was not applicable to all participants. Our finding that some participants were more motivated by the extrinsic motivation of "protecting their partner" indicated the need to diversify messaging to extend our reach. Testing a diverse set of messages yielded important additional insights not observed in the initial survey. While the "fear-based" message ("Itching. Burning. Rash. Sores. Getting tested is a MUST!") was one of the top selected messages in the survey, upon deeper discussion with the focus group, this was clearly found to not be a motivating message. Although not intentionally explored in this study, there were some suggestions that the messages presented could be acceptable across multiple platforms (including digital and traditional place-based advertisements). This is something we will explore further in the future.

Competition is another principle often underreported in social marketing interventions [31] and although not directly explored in this study, participants did have some important insights to inform competition. Throughout the workshops, participants made it clear that they wanted a service that had no out-of-pocket costs. This is important given our finding that there are no free web-based comprehensive STI/BBV testing services available in Victoria [10]. Those in regional areas also mentioned the potential for improved access to testing with a web-based service, as it is difficult to access the limitedly available in-person services in their geographic locations. Despite these advantages, we anticipate that strong digital promotion will be needed to compete with existing web-based services that already have high visibility.

XSL•FO
RenderX

Limitations

Although our findings are informed by a diversity of views among young people of different genders, place of residence, and cultural backgrounds, our sample size was relatively small and did not include individuals with disabilities, those younger than 18 years, or many heterosexual men who may have different views about the messaging and imagery of a web-based STI/BBV testing service. As such, our findings may be less applicable to those groups and additional strategies may be needed to improve accessibility for these groups.

Despite the known limitations and lack of insights offered by geographic segmentation [30], we believed that in our context it was important to include this. While our focus groups were limited by lack of segmentation for other factors such as social status, lifestyle, and opinions, we will have the opportunity to explore these during phases 4 and 5 of the project—user testing and evaluation.

In this phase of developing the service, our objective was to explore imagery and messaging of a web-based service that would be acceptable to young people, rather than investigating what would make young people change their behavior and increasingly test. This will be explored during phase 5, the evaluation. We also acknowledge that while we worked directly with young people to design the imagery, messaging, and promotion of our web-based service, the clinical and legal requirements of providing web-based STI/BBV testing service in Australia required that some of the material we presented to the young people was researcher driven, rather than end user driven, potentially introducing researcher bias to the project. However, workshops did allow the young people to introduce new ideas for the design and functionality of the service and its messaging and promotion and gave important insights into young people's intentions, motivations, and interest in web-based STI/BBV testing.

Conclusions

This study has highlighted the value of applying social marketing theory with end user engagement when developing a web-based STI/BBV testing service. We were able to identify that young people are more likely to be attracted to a web-based service that uses messages that highlight the functional benefits of service, and the use of professional imagery and logos within the service can elicit trust. The results of this study will be used to develop and implement a web-based sexual health service for young Victorians, which will be accessible to young people and effectively increase STI/BBV testing uptake. Future research will measure the reach, the efficacy of social marketing strategy, and how this service complements existing services in increasing STI/BBV testing uptake among young Victorians.

Acknowledgments

This project was funded by a National Health and Medical Research Council partnership grant (grant APP2006486).

Data Availability

All data generated or analyzed during this study are available from the corresponding author on reasonable request and with appropriate ethical approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key messages, scenarios, and landing page samples used during the initial survey of young people aged 18-29 years in Victoria, Australia, exploring messaging and design preferences for a web-based sexually transmitted infection/blood-borne virus testing service.
[DOCX File , 137 KB - jmir_v27i1e63822_app1.docx]

Multimedia Appendix 2
Using Andreasen's theory of social marketing to guide the design of a web-based sexually transmitted infection testing service for 18-29-year-olds in Victoria, Australia, with research strategies and key findings from formative research and end user engagement.
[DOCX File , 20 KB - jmir_v27i1e63822_app2.docx]

References

1. Kirby Institute, University of New South Wales (UNSW). HIV, viral hepatitis and sexually transmissible infections in Australia: annual surveillance report 2022. 2022. URL: https://tinyurl.com/n6y99ate [accessed 2025-03-05]
2. Turner KME, Looker KJ, Syred J, Zienkiewicz A, Baraitser P. Online testing for sexually transmitted infections: A whole systems approach to predicting value. PLoS One 2019;14(2):e0212420 [FREE Full text] [doi: 10.1371/journal.pone.0212420] [Medline: 30794589]

XSL•FO
RenderX

3. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
4. Andreasen AR. Marketing Social Marketing in the Social Change Marketplace. Journal of Public Policy & Marketing 2002 Apr 01;21(1):3-13. [doi: 10.1509/jppm.21.1.3.17602]
5. Teal G, McAra M, Riddell J, Flowers P, Coia N, McDaid L. Integrating and producing evidence through participatory design. CoDesign 2022 Jul 11;19(2):110-127. [doi: 10.1080/15710882.2022.2096906]
6. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
7. Willmott TJ, Schmidtke DJ, McLeod S. Editorial: Nothing about us without us: participatory design application in social marketing. In: JSOCM. Leeds, England: Journal of Social Marketing; May 07, 2024:137-142.
8. Rundle-Thiele S, Dietrich T, Carins J. CBE: A Framework to Guide the Application of Marketing to Behavior Change. Social Marketing Quarterly 2021 Jun 09;27(3):175-194. [doi: 10.1177/15245004211021643]
9. Australian STI Management Guidelines for Use in Primary Care. Australasian society for HIV, viral hepatitis and sexual health medicine. 2021 Dec. URL: https://sti.guidelines.org.au/ [accessed 2025-03-05]
10. Cardwell ET, Ludwick T, Fairley C, Bourne C, Chang S, Hocking JS, et al. Web-Based STI/HIV Testing Services Available for Access in Australia: Systematic Search and Analysis. J Med Internet Res 2023 Sep 22;25:e45695 [FREE Full text] [doi: 10.2196/45695] [Medline: 37738083]
11. Ludwick T, Walsh O, Cardwell ET, Chang S, Kong FYS, Hocking JS. Moving Toward Online-Based Sexually Transmitted Infection Testing and Treatment Services for Young People: Who Will Use It and What Do They Want? Sex Transm Dis 2024 Mar 01;51(3):220-226. [doi: 10.1097/OLQ.0000000000001899] [Medline: 37963329]
12. Cardwell ET, Walsh O, Chang S, Coombe J, Fairley C, Hocking JS, et al. Preferences for online or in-person STI testing vary by where a person lives and their cultural background: a survey of young Australians. Sex Transm Infect 2024 Dec 06. [doi: 10.1136/sextrans-2024-056292] [Medline: 39643439]
13. Walsh O, Cardwell ET, Hocking JS, Kong FYS, Ludwick T. Where would young people using an online STI testing service want to be treated? A survey of young Australians. Sex Health 2024 Sep;21. [doi: 10.1071/SH24087] [Medline: 39250599]
14. Reynolds-Tylus T. Psychological Reactance and Persuasive Health Communication: A Review of the Literature. Front. Commun 2019 Oct 31;4. [doi: 10.3389/fcomm.2019.00056]
15. Brickman J, Willoughby JF. 'You shouldn't be making people feel bad about having sex': exploring young adults' perceptions of a sex-positive sexual health text message intervention. Sex Education 2017 Jun 12;17(6):621-634. [doi: 10.1080/14681811.2017.1332582]
16. Jang EE, Wagner M, Sinclair J. National Survey on Essential Communication Skills to Address Language Demands in Canadian Nursing Practice. J Adv Nurs 2025 Mar 04;33(2):241-269. [doi: 10.1111/jan.16861] [Medline: 40035227]
17. Denison HJ, Bromhead C, Grainger R, Dennison EM, Jutel A. Barriers to sexually transmitted infection testing in New Zealand: a qualitative study. Aust N Z J Public Health 2017 Aug;41(4):432-437 [FREE Full text] [doi: 10.1111/1753-6405.12680] [Medline: 28664644]
18. Ten Hoor GA, Ruiter RAC, van Bergen JEAM, Hoebe CJPA, Dukers-Muijrers NHTM, Kok G. Predictors of Chlamydia Trachomatis testing: perceived norms, susceptibility, changes in partner status, and underestimation of own risk. BMC Public Health 2016 Jan 20;16(1):55 [FREE Full text] [doi: 10.1186/s12889-016-2689-6] [Medline: 26790411]
19. Martin-Smith HA, Okpo EA, Bull ER. Exploring psychosocial predictors of STI testing in University students. BMC Public Health 2018 May 29;18(1):664 [FREE Full text] [doi: 10.1186/s12889-018-5587-2] [Medline: 29843658]
20. Corcoran N. Theories and models in communicating health messages. In: Communicating Health: Strategies for Health Promotion. Thousand Oaks, CA: SAGE Publications, Ltd; 2007:5-31.
21. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Wilkinson TA, Jenkins K, Hawryluk BA, Moore CM, Wiehe SE, Kottke MJ. Dual Protection Messaging for Adolescents and Young Adults in the Setting of Over-the-Counter Hormonal Contraception: A Human-Centered Design Approach. J Pediatr Adolesc Gynecol 2022 Dec;35(6):669-675 [FREE Full text] [doi: 10.1016/j.jpag.2022.08.009] [Medline: 36031114]
23. Njoki N, Cutherell M, Musau A, Mireri D, Nana-Sinkam A, Phillips M. Applying Human-Centered Design to Replicate an Adolescent Sexual and Reproductive Health Intervention: A Case Study of Binti Shupavu in Kenya. Glob Health Sci Pract 2023 Dec 18;11(6):e2200557. [doi: 10.9745/ghsp-d-22-00557]
24. Davis AC, Wright CJ, Murphy S, Dietze P, Temple-Smith MJ, Hellard ME, et al. A Digital Pornography Literacy Resource Co-Designed With Vulnerable Young People: Development of "The Gist". J Med Internet Res 2020 Jun 01;22(6):e15964 [FREE Full text] [doi: 10.2196/15964] [Medline: 32348268]
25. Ippoliti N, Sekamana M, Baringer L, Hope R. Using Human-Centered Design to Develop, Launch, and Evaluate a National Digital Health Platform to Improve Reproductive Health for Rwandan Youth. Glob Health Sci Pract 2021 Nov 29;9(Supplement 2):S244-S260. [doi: 10.9745/ghsp-d-21-00220]
26. Gkatzidou V, Hone K, Sutcliffe L, Gibbs J, Sadiq ST, Szczepura A, et al. User interface design for mobile-based sexual health interventions for young people: design recommendations from a qualitative study on an online Chlamydia clinical

XSL•FO
RenderX

care pathway. BMC Med Inform Decis Mak 2015 Aug 26;15(1):72 [FREE Full text] [doi: 10.1186/s12911-015-0197-8] [Medline: 26307056]
27. Martin P, Alberti C, Gottot S, Bourmaud A, de La Rochebrochard E. Young people's proposals for a web-based intervention for sexual health promotion: a French qualitative study. BMC Public Health 2023 Jul 19;23(1):1389 [FREE Full text] [doi: 10.1186/s12889-023-16257-8] [Medline: 37468863]
28. Shoveller J, Knight R, Davis W, Gilbert M, Ogilvie G. Online Sexual Health Services: Examining Youth's Perspectives. Can J Public Health 2012 Jan 1;103(1):14-18. [doi: 10.1007/bf03404062]
29. Fakoya I, Cole C, Larkin C, Punton M, Brown E, Ballonoff Suleiman A. Enhancing Human-Centered Design With Youth-Led Participatory Action Research Approaches for Adolescent Sexual and Reproductive Health Programming. Health Promot Pract 2022 Jan;23(1):25-31. [doi: 10.1177/15248399211003544] [Medline: 33858267]
30. Dietrich T, Hurley E, Carins J, Kassirer J, Rundle-Thiele S, Palmatier RW, et al. 50 years of social marketing: seeding solutions for the future. EJM 2022 Apr 21;56(5):1434-1463. [doi: 10.1108/ejm-06-2021-0447]
31. McKercher B, Tolkach D, Eka Mahadewi NM, Byomantara DGN. Choosing the Optimal Segmentation Technique to Understand Tourist Behaviour. Journal of Vacation Marketing 2022 Feb 11;29(1):71-83. [doi: 10.1177/13567667221078240]
32. Kubacki K, Rundle-Thiele S, Pang B, Carins J, Parkinson J, Fujihira H, et al. Segmentation in social marketing: process, methods and application. In: Dietrich T, Rundle-Thiele S, Kubacki K, editors. An Umbrella Review of the Use of Segmentation in Social Marketing Interventions. Singapore: Springer Singapore; 2017.

Abbreviations

BBV: blood-borne virus
EOI: expression of interest
STI: sexually transmitted infection

Edited by A Mavragani; submitted 05.07.24; peer-reviewed by A McInnes-Dean, S Rundle-Thiele; comments to author 14.11.24; revised version received 20.12.24; accepted 11.02.25; published 27.03.25.

Please cite as:
Cardwell ET, Ludwick T, Chang S, Walsh O, Lim M, Podbury R, Evans D, Fairley CK, Kong FYS, Hocking JS
Engaging End Users to Inform the Design and Social Marketing Strategy for a Web-Based Sexually Transmitted Infection/Blood-Borne Virus (STI/BBV) Testing Service for Young People in Victoria, Australia: Qualitative Study
J Med Internet Res 2025;27:e63822
URL: https://www.jmir.org/2025/1/e63822
doi:10.2196/63822
PMID:

©Ethan T Cardwell, Teralynn Ludwick, Shanton Chang, Olivia Walsh, Megan Lim, Rachel Podbury, David Evans, Christopher K Fairley, Fabian Y S Kong, Jane S Hocking. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of an mHealth- and School-Based Health Education Program for Salt Reduction (EduSaltS) in China: Cluster Randomized Controlled Trial Within Scale-Up

Naibo Wang[1,2*], PhD; Chen Wang[1*], MPH; Puhong Zhang[3], Prof Dr, PhD; Yinghua Li[4], Prof Dr, PhD; Feng J He[5], Prof Dr, PhD; Li Li[4], MPH; Yuan Li[3], PhD; Rong Luo[3], MPH; Dezhi Wan[6], MBChB; Lewei Xu[2], BSPH; Lifang Deng[2], MPH; Lei Wu[1], MScPH

[1]School of Public Health, Jiangxi Provincial Key Laboratory of Disease Prevention and Public Health, Jiangxi Medical College, Nanchang University, Nanchang, China

[2]Jiangxi Provincial Center for Patriotic Health and Health Promotion, Nanchang, China

[3]The George Institute for Global Health, Beijing, China

[4]Chinese Center for Health Education, Beijing, China

[5]Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[6]Jiangxi Association for Health Education and Tobacco Control, Nanchang, China

[*]these authors contributed equally

Corresponding Author:
Lei Wu, MScPH
School of Public Health, Jiangxi Provincial Key Laboratory of Disease Prevention and Public Health, Jiangxi Medical College
Nanchang University
No. 461 Ba Yi Avenue, Donghu District
Jiangxi Province
Nanchang, 330006
China
Phone: 86 18970813850
Email: leiwu@ncu.edu.cn

Abstract

Background: Globally, cardiovascular diseases are leading causes of mortality and disability, with hypertension being a major risk factor. Reducing salt intake and blood pressure are among the most cost-effective health promotion strategies. While mobile health (mHealth)– and school-based salt reduction interventions have proven effective in trials, their impact when scaled up in real-world contexts remains uncertain.

Objective: We evaluated the effectiveness of the real-world implementation of an mHealth- and school-based health education scale-up program to reduce salt intake (EduSaltS [mHealth and school-based education program to reduce salt intake scaling up in China]).

Methods: A parallel cluster randomized controlled trial was conducted from April 2022 to July 2023 across 20 schools in 2 districts and 2 counties within Ganzhou City, Jiangxi Province, China. Schools were randomized 1:1 to intervention or control groups within each district or county. One third-grade class per school and 26 students per class were randomly sampled. One parent, or alternative family member (aged 18-75 years, residing with the student), of each student was invited to join. The EduSaltS intervention, spanning over 1 academic year, incorporated both app-based health education courses and offline salt reduction activities, with participation monitored through the backend management system. The intervention's effectiveness was assessed by comparing changes in salt intake and blood pressure between groups from baseline to 1-year follow-up using surveys, physical examination, and 24-hour urine tests.

Results: Of 524 children (boys: n=288, 54.96%; age: mean 9.16, SD 0.35 years) and 524 adults (men: n=194, 37.02%; age: mean 40.99, SD 11.04 years) who completed the baseline assessments in 10 intervention and 10 control schools, 13 (2.48%) children and 47 (8.97%) adults were lost to follow-up. All schools and participants showed satisfactory intervention adherence. Measured differences in schoolchildren's salt intake, systolic blood pressure, and diastolic blood pressure, between the intervention and control schools, were –0.24 g/day (95% CI –0.82 to 0.33), –0.68 mm Hg (95% CI –2.32 to 0.95), and –1.37 mm Hg (95%

XSL•FO
RenderX

CI –2.79 to 0.06), respectively. For adults, the intervention group's salt intake decreased from 9.0 (SE 0.2) g/day to 8.3 (SE 0.2) g/day post intervention. Adjusted changes in the intervention (vs control) group in salt intake, systolic blood pressure, and diastolic blood pressure were –1.06 g/day (95% CI –1.81 to –0.30), –2.26 mm Hg (95% CI –4.26 to –0.26), and –2.33 mm Hg (95% CI –3.84 to –0.82), respectively.

Conclusions: The EduSaltS program, delivered through primary schools with a child-to-parent approach, was effective in reducing salt intake and controlling blood pressure in adults, but its effects on children were not significant. While promising for nationwide scaling, further improvements are needed to ensure its effectiveness in reducing salt intake among schoolchildren.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400079893; https://tinyurl.com/4maz7dyv (retrospectively registered); Chinese Clinical Trial Registry ChiCTR2000039767; https://tinyurl.com/5n6hc4s2

(J Med Internet Res 2025;27:e60092) doi:10.2196/60092

KEYWORDS

school-based health education; EduSaltS; mobile health; salt reduction; cluster randomized trial

Introduction

Cardiovascular disease (CVD) stands as the foremost cause of death and disability worldwide, with hypertension identified as its major risk factor [1-4]. Substantial evidence indicates that excessive salt intake is a key contributor to elevated blood pressure and a significant risk factor for CVD [5-8]. Additionally, the incidence and progression of various noncommunicable diseases, such as chronic kidney diseases, are also linked to salt consumption [9]. Numerous systematic reviews and meta-analyses have demonstrated that reducing salt (sodium) intake can lower the risk of hypertension and CVD [10-13]. The World Health Organization considers reducing salt intake among populations as one of the most effective and economical strategies for health promotion.

China ranks as one of the countries with the highest salt intake, averaging about 11 g/day among adults [14], more than twice the World Health Organization's recommended level of 5 g/day. Unlike developed countries where the majority of salt intake in the population originates from processed foods, approximately 80% of the salt consumption in the Chinese diet comes from the salt added during cooking [6]. Therefore, in addition to reducing sodium in food processing, lowering cooking salt is more crucial. To this end, the Chinese government has initiated policies such as the "Healthy China Initiative (2019-2030)," implementing a national dietary action for healthy living and advocating for "Three Reductions and Three Healthy Conditions," with salt reduction being a primary focus. Meanwhile, a series of scientific studies [15-19] have also been conducted in China to create a supportive environment for salt reduction, optimize public health education and individualized interventions, and further refine comprehensive strategies for reducing salt intake.

Schools, as a critical place for health education, have been proven effective in addressing public health issues concerning smoking, obesity, and CVD [20-22]. The advent of modern information and mobile technologies has also further broadened the scope of health interventions [23,24]. A cluster randomized controlled trial (School-EduSalt [school-based education program to reduce salt intake in children and their families]) was conducted by our team to demonstrate that the "small hands leading big hands" strategy, in which schoolchildren encouraged their parents to collaboratively enhance their knowledge, attitudes, and practices related to salt reduction, effectively reduced salt intake among both children and adults [25]. However, its reliance on specialized offline courses and activities, which demanded considerable training for schoolteachers, limited large-scale implementation. To address this issue, a mobile health (mHealth) platform for salt reduction named AppSalt was developed to provide standardized app-based health education courses and parallel homework assignments for students and their parents. Subsequently, a cluster randomized controlled trial was conducted under close supervision and the results showed that AppSalt was also effective in lowering salt intake among adults [18]. Building on these research findings, an updated development and preliminary implementation study called the mHealth and school-based education program to reduce salt intake scaling up in China (EduSaltS [mHealth and school-based education program to reduce salt intake scaling up in China]) was launched in Ganzhou in Jiangxi province, Zhenjiang in Jiangsu province and Qinhuangdao in Hebei province, for the first time, to refine and scale up the intervention. Compared with the interventions in our previous studies, EduSaltS features a WeChat-based system (Tencent), cartoon-enhanced course materials, role-specific user interfaces, and blended online/offline health education, aiming for a routine yet innovative school-based approach [26]. The evolution of studies from School-EduSalt, AppSalt to EduSaltS can be seen in Figure 1.

Given the uncertainty surrounding the impact of EduSaltS in real-world settings, this study aims to evaluate its effectiveness in reducing salt intake through a cluster randomized controlled trial conducted in Ganzhou, Jiangxi Province in China.

XSL•FO
RenderX

Figure 1. The evolution of studies from School-EduSalt, AppSalt to EduSaltS. cRCT: cluster randomized controlled trials; EduSaltS: mHealth and school-based education program to reduce salt intake scaling up in China; OS: operation system; School-EduSalt: school-based education program to reduce salt intake in children and their families.

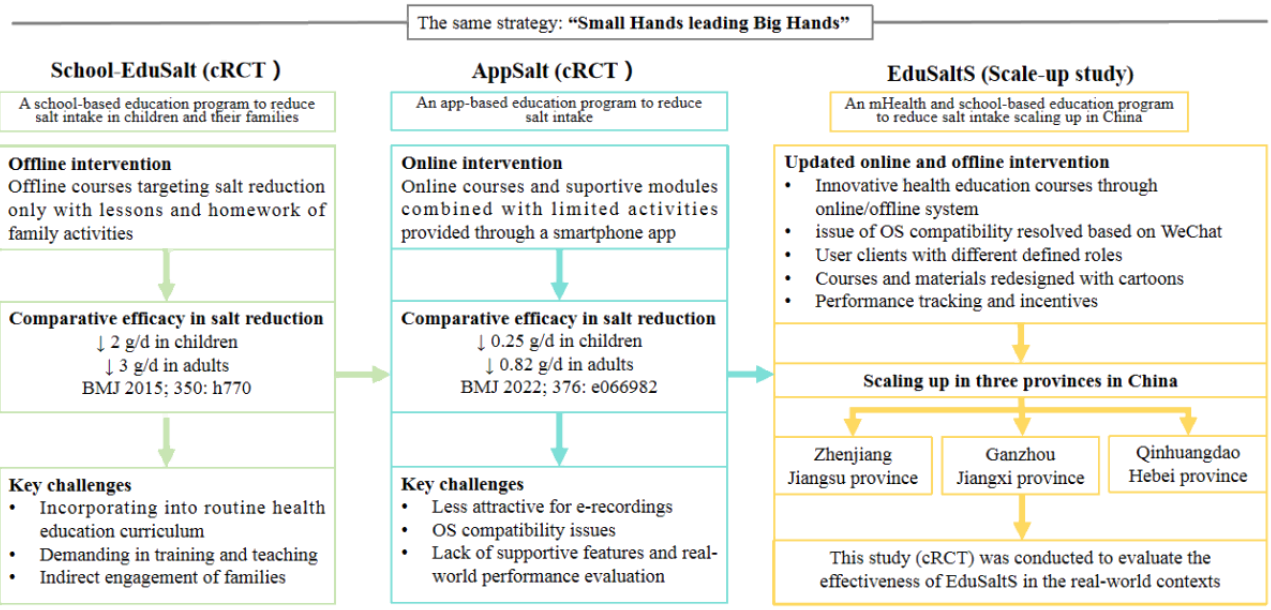

Methods

Ethical Considerations

This study was approved by the Scientific Research Ethics Committee of the First Affiliated Hospital of Gannan Medical College (LLSC-2022011901) and was registered retrospectively at the Chinese Clinical Trial Registry (ChiCTR2400079893). This was due to its embedding in a prospectively registered scale-up intervention study (ChiCTR2000039767) and a lack of timely communication with the overall project team. Nevertheless, this study was identical to the protocol assessed by the Research Ethics Committee, which was prospectively approved on January 19, 2022. Permissions were obtained from the local education authorities and school headteachers. All participants taking part in the outcome assessments gave written informed consent. For children, participant assent and parental written consent were obtained. This report does not contain any identifying information or direct quotes. Due to the nature of the program, this study does not involve details of compensation.

Study Design and Participants

The overall design and framework of EduSaltS program have been published elsewhere [26]. The study was implemented using a cluster randomized controlled trial design within scale-up, spanning across 20 public primary schools in Ganzhou City, Jiangxi Province, over a period from April 2022 to July 2023. The process of this study encompassed (1) determination of research sites, (2) execution of on-site baseline survey and clinical evaluations at the end of third grade, (3) allocation of the schools into 2 groups comprised of 10 schools in the intervention group and an equal number in the control group, (4) implementation of an integrative intervention and its subsequent promotion, conducted over the course of 2 semesters, and (5) undertaking of effect evaluation after intervention at the end of fourth grade.

EduSaltS was implemented in 100 scale-up schools across 9 subordinate districts and counties in Ganzhou city, mobilizing all third-grade students and their parents to participate in a year-long health education course and activities focused on salt reduction. In this context, to assess the effectiveness of the intervention, we first contacted the local education authority through the local Health Commission to gain their opinion, support, and approval for the study. After that, considering the potential diversity between urban and rural areas, 2 districts (Zhanggong and Nankang) and 2 counties (Xinfeng and Yudu) close to the city center were selected as the representative sites for the effectiveness assessments following the principle of proximity [27], which can help minimize the influence of geographical and environmental factors, before the commencement of EduSaltS. The selection of schools involved choosing a total of 20 schools from these 4 areas that were willing to participate in the trial, with an even number of schools from each district or county. These schools were then divided into the intervention or control group within each district or county in a 1:1 ratio. The 10 schools in the intervention group were included in the 100 scale-up schools participating in EduSaltS and received the unified interventions, while the control group was not exposed to the intervention and only received routine health education courses, including physical and mental health as all school in the city, but not covering salt reduction. Details of the scale-up schools participating in EduSaltS and the schools in the trial across different districts and counties are provided in Table S1 in Multimedia Appendix 1.

In each selected school, due to the substantial workload and costs associated with conducting the assessments, a simple random sampling method was used. This involved randomly selecting one third-grade class from each school, followed by the random selection of 26 students from the chosen class to serve as survey participants. Subsequently, one parent of each participating child was recruited. At this stage, invitations and

XSL•FO
RenderX

explanatory materials were distributed to parents through the class teacher. If a parent was unavailable, a family member residing with the child, such as a grandparent, was eligible for selection as an alternative. To mitigate bias arising from sex imbalance, preference was given to selecting an adult of the opposite sex to the student whenever possible. In total, this study planned to recruit 520 students and an equivalent number of adults. The baseline survey and clinical evaluations were completed before the schools were allocated, and the follow-up assessments were conducted after the intervention had concluded.

Inclusion criteria were as follows: (1) willingness to participate in follow-up, (2) no transfer plans during the salt reduction intervention period for students, (3) ages of participating adults being older than 18 years and younger than 75 years, and (4) being able to provide regular feedback and complete outcome measurements. Exclusion criteria included: (1) children with medical conditions or other reasons precluding participation in the assessments or intervention activities, (2) individuals explicitly unable to complete follow-up, and (3) inability or refusal to collect 24-hour urine samples.

Randomization and Masking

Schools in each district or county were randomly assigned to the intervention or control group (1:1) using a stratified method of random digits by researchers from Jiangxi Provincial Center for Patriotic Health and Health Promotion with the support of the local bureau of education. This ensured a balanced and comparable distribution of schools in 2 groups across districts and counties. Additionally, after the random allocation into intervention and control groups, a comparison of the geographic distribution of schools was conducted to prevent potential contamination by ensuring no close proximity between schools in different groups. The final grouping of schools is detailed in Table S2 in Multimedia Appendix 1.

Due to the specificity of the intervention, it was imperative to explicitly inform the intervention group of the relevant details. Therefore, blinding was not feasible for the intervention implementers. However, blinding was employed for data collectors. Following rigorous training, investigators from hospitals and Centers for Disease Prevention and Control, without knowing the group allocations, collected baseline and follow-up data. The technicians who performed urinary electrolyte measurements (KingMed Diagnostics) were unaware of specific groups for individual samples.

Procedure

After selecting the schools and classes for the study, communication with school headteachers and class teachers was initiated. Class teachers were responsible for distributing letters of project introduction and informed consent forms to students and their parents (or guardians), thereby confirming the specific student and parent participants and facilitating the on-site baseline survey and clinical evaluations. The content comprised (1) sociodemographic characteristics (eg, age, sex, education level, and relationship to the participating children) and health behaviors (eg, smoking, alcohol consumption, and physical exercise); (2) measurements including height, weight, and blood pressure, which were conducted using standardized tools by trained staff from the Centers for Disease Control and Prevention and township health centers; and (3) a single 24-hour urine collection conducted on a weekday at the same time for both children and adults. The 24-hour urine collection was used to evaluate urinary sodium excretion, which is the most accurate method to estimate salt intake. All participants were provided with sufficient containers to collect their urine continuously over a 24-hour weekday period, ensuring a total collection time of 20-28 hours. After measuring the total urine volume, the thoroughly mixed sample was analyzed for sodium concentration, which was then multiplied by the total volume (corrected for any minor deviations from 24 hours) to calculate the 24-hour urinary sodium excretion. Salt intake was derived by multiplying the 24-hour urinary sodium excretion by 23 (atomic weight of sodium) and 2.54 (the sodium-to-salt conversion factor). The follow-up investigation was conducted after 1 year of intervention, using the same method.

Intervention

The specific framework, development process, features, and intervention protocol of EduSaltS have been detailed in a previously published paper [26]. The intervention, which spanned 1 academic year, incorporated app-based courses, offline activities, and the establishment of a supportive salt-reduction environment. Quality control was mainly implemented by teachers following standardized training, with guidance from our research team, targeting all fourth-grade students and their families in intervention schools. The concise summary of interventions of EduSaltS is provided in Table S3 in Multimedia Appendix 1, and a detailed demonstration of specific applications and web interfaces can also be seen in our published paper [26].

The online component of this intervention predominantly revolved around the WeChat miniprogram named “Health Cloud Classroom”, delivering a curriculum of salt reduction. In alignment with the “Health Education Guidelines for Primary and Secondary Schools” stipulated by the Ministry of Education of China, the miniprogram disseminated 1 health lesson weekly and 10 lessons for 1 semester. Each video lesson, approximately 10 minutes in duration, was designed for students to watch at home with their family members, followed by a set of 5-10 quiz questions to enhance learning interest and assess comprehension. The curriculum encompassed eight compulsory lessons on salt reduction (covering the dangers of excessive salt intake, misconceptions, and practical techniques about salt reduction) and 12 elective health lessons (such as eye protection, nutritional diet, and influenza prevention). Class teachers had responsibilities for organizing students and their parents or grandparents to use the “Health Cloud Classroom” for course learning. They also used a teacher-oriented miniprogram named “Health Salt Management” (a WeChat app designed for monitoring, managing, and evaluating) to stay informed about the status of course learning and homework completion of all students and their families in each class, ensuring timely adherence to requirements of the curriculum. Each term, students were expected to complete 10 lessons. Moreover, “Health Cloud Classroom” features additionally optional modules designed to facilitate the increase of low-sodium knowledge and encourage

XSL•FO
RenderX

corresponding behavioral changes, such as healthy food selection (explanations of various food nutrients), interactive quizzes on salt-related knowledge for entertainment, and routine assessments of sodium intake levels through self-administered dietary records. These modules were optional, and families were free to choose whether or not to use them without any monitoring of their use. Materials uploaded in this miniprogram for online study, such as cartoons, PowerPoint, and posters, could also be used for offline activities.

The offline component primarily involved salt reduction interactive activities. Integrating existing educational frameworks like health education classes, science practice classes, and themed class meetings, each term featured at least 4 lectures and themed activities concerning salt reduction. Class teachers could use videos, posters, and leaflets provided by "Health Cloud Classroom" to impart knowledge about salt reduction to students. Children were expected to complete related homework and convey salt reduction knowledge, methods, and skills to their family members after class, persuading the person cooking at home to reduce salt use and improve dietary behavior. Additionally, the program includes the construction of an environment of lower sodium and health education regarding reduced-salt school meals, such as displaying salt reduction posters across the campus, broadcasting a series of salt reduction short films and radio dramas within schools, and providing self-learning materials and training once per semester for cafeteria chefs and catering providers to lower the salt content in students' meals.

To render the evaluation of the performance of the process more scientifically effective, a web-based miniprogram in WeChat named "EduSaltS Manager," with regard to backend management and quality control, was developed specifically for researchers and school teachers. This program aided teachers in knowing and overseeing the participation of families and in submitting records related to project implementation. Researchers could also access relevant data through this backend system, regularly monitoring the progress and compliance in participants or specific schools.

Students in the control group received routine health education courses, occasionally delivered by physical education teachers, with no specific content on salt reduction and no family engagement. Due to the citywide publicity of EduSaltS, schools in the control group might have been aware of this project. However, as families in the control group were not provided with accounts to access the intervention miniprogram, there was little risk of contamination.

Outcomes

The primary outcome of this study was the difference between the 2 groups in the change in salt intake among students and adults, before and after the intervention, assessed through 24-hour urinary sodium. The secondary outcome was the change in blood pressure. All measurements and data collections were conducted once at baseline and after completion of the intervention, with consistent survey processes and methods of data collection at each time point and across both groups.

The method for 24-hour urine collection was the same as previous studies, using an electronic data collection system [14]. Tests of all samples were uniformly conducted by a central laboratory (KingMed Diagnostics), with the testing parameters including urinary sodium, potassium, and creatinine concentrations. Urinary volume and creatinine concentration were used to assess the completeness of the 24-hour urine collection. The concentration of urinary sodium and potassium was used to calculate the participants' salt and potassium intake, as well as the sodium-potassium ratio.

Blood pressure measurements were conducted using calibrated Omron electronic sphygmomanometers. After a 10-minute rest in a seated position, trained physicians measured the blood pressure of the right arm of the participants. Blood pressure was measured 3 times for each individual, with at least a 1-minute interval between measurements. The average of the last 2 readings was used for statistical analysis.

This study employed an electronic information platform to support data collection, allowing for the recording of the evidence of key survey indicators through audio recording, video recording, and photography, which standardized the collection of diverse data information.

Sample Size Calculation

A mixed linear model was planned to be applied to analyze changes in effect sizes, with alterations in salt intake (calculated from 24-hour urinary sodium excretion) serving as the primary outcome variable. In accordance with previous research [25], a reduction of 1 g/day of salt intake was set as the target change in effect size, with an SD of 2.15 and an intraclass correlation coefficient of 0.01. Setting α at 0.05, 1-β at 0.9, and considering a 10% loss to follow-up rate, the sample size calculation performed using PASS software indicated that 520 students and 520 parents (cluster size of 26) across 10 intervention and 10 control schools would satisfy the power of the test.

Statistical Analysis

The effectiveness of the intervention was evaluated through an intention-to-treat analysis without imputation due to the very low drop-out rates (13/524, 2.48% for children and 47/524, 8.97% for adults). All students and their family members who completed the baseline assessments were analyzed according to the initial group allocation of their respective schools. Taking into account the hierarchical structure of the data, with 2 measurements (baseline and 2 semesters) for each participant and participants nested within each school, mixed linear models with random intercepts and fixed effect variables (eg, time, group, time×group, and covariates) were applied to analyze the effects of intervention. The time×group interaction term indicated the differential change by groups from baseline to the end of the trial. Potential confounding variables adjusted in the model included district/county areas, age, sex, BMI (body weight in students), education level (the adults' level substituted for their children), and physical exercise (coded as "yes" if someone engaged in moderate physical activities at least 3 times a week and more than 30 minutes each time). Additionally, when analyzing blood pressure changes, outdoor temperature was further included as a confounding factor. In the analysis of

XSL•FO
RenderX

the effect on adults, familial relationships, smoking, and alcohol consumption (coding "occasional" and "regular" as "yes") were added as additional adjustments.

The collection time of 24-hour urine samples was recorded in this study. The urine sample would be excluded if the collection time was less than 20 hours or more than 28 hours. If the duration of the collection was within 20-28 hours, the 24-hour urine volume was calculated as: urine volume (mL per 24 hours) = urine volume (mL) / duration of urine collection (hours) × 24 hours. This adjustment, in conjunction with urinary concentration of sodium and potassium, was used for the calculation of individual salt intake and other related indicators. Urine samples were also considered incomplete and excluded from the main analysis if 24-hour urine volume was <500 mL in adults; 24-hour urinary creatinine was <4.0 mmol or <6.0 mmol in women or men, respectively; 24-hour urine volume was <300 mL in children; or 24-hour urinary creatinine was lower than the fifth centile (<2.27 mmol for girls and <2.43 mmol for boys).

Sensitivity analysis and subgroup analysis were carried out in this study. In the sensitivity analysis, possibly incomplete 24-hour urine collections, assessed by 24-hour urinary volume and creatinine excretion, were included in the analysis based on the intention-to-treat approach. In the subgroup analysis, for students, the effects of intervention across different subgroups based on district/county areas, sex, and education levels of their familial participants were analyzed. For adults, additional analyses of subgroups were performed, which included age groups, education levels, blood pressure status (hypertension was defined as systolic blood pressure [SBP] ≥140 mm Hg or diastolic blood pressure [DBP] ≥90 mm Hg, or self-reported hypertension), smoking, weight status (BMI<24.0 or not), and relationship with their children.

Quantitative data were statistically described using means and SDs, while categorical data were presented as frequencies and percentages. Statistical analyses were executed using SAS (version 9.4; SAS Institute Inc), with a *P* value of <.05 considered statistically significant.

Results

Characteristics of Participants

Twenty schools across 4 districts or counties in Ganzhou City in southeastern China were recruited in this research. Adhering to the specified inclusion and exclusion criteria, a total of 524 families (4 families, beyond our initial expectations, from different schools were enthusiastic about this project, so we enrolled them as well) participated in the study and completed the baseline survey and clinical evaluations. Following randomization and 2 semesters of intervention, the loss to follow-up was 2.48% (13/524) for students and 8.97% (47/524) for adults, due to either the students having moved to another school or adults being unable to attend subsequent assessments, as detailed in Figure 2. Baseline characteristics of those lost to follow-up did not significantly differ from those who completed the study and the loss to follow-up was completely random (*P*>.05 for the Little test for both children and adults). In all families of the intervention and control groups who completed the baseline assessments, the average ages of the students were 9.10 (SD 0.35) and 9.22 (SD 0.34), respectively, while the average ages of adults were 40.96 (SD 11.07) and 41.01 (SD 11.03). Each of the intervention and control groups had 20 adults who self-reported having hypertension, and these individuals were confirmed to be on medication at the time of blood pressure measurement. Demographic and general behavioral characteristics of these 2 randomized groups were basically consistent and well-balanced, as detailed in Table 1.

XSL•FO
RenderX

Figure 2. Research process of trial within scale-up.

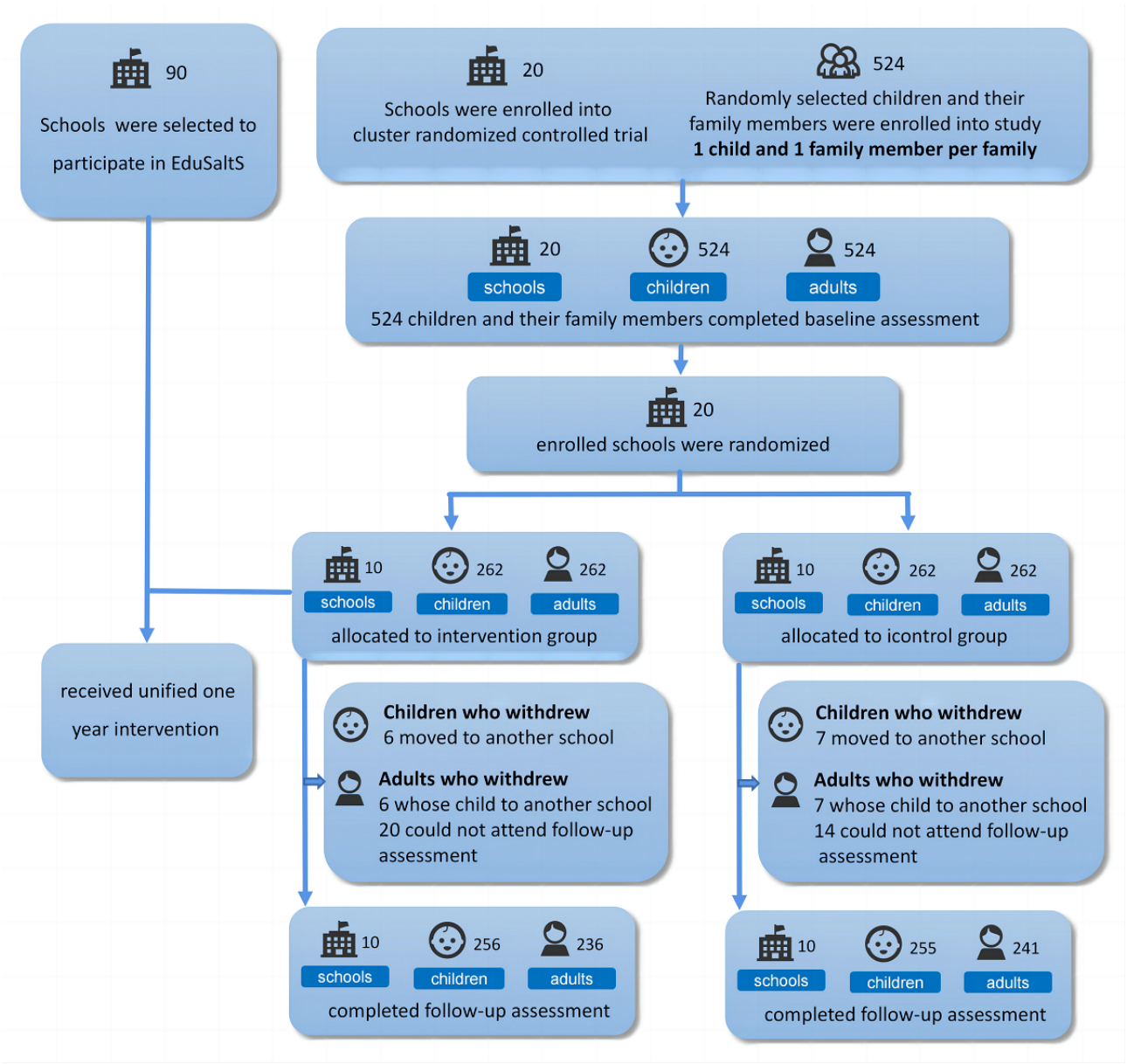

XSL•FO
RenderX

Table 1. Basic characteristics at baseline.

Characteristics	Intervention	Control	Total
Cluster level			
Number of schools, by location, n (%)			
Zhanggong district	4 (40.00)	4 (40.00)	8 (40.00)
Nankang district	3 (30.00)	3 (30.00)	6 (30.00)
Xinfeng county	2 (20.00)	2 (20.00)	4 (20.00)
Yudu county	1 (10.00)	1 (10.00)	2 (10.00)
Meal provision, n (%)			
Yes	6 (60.00)	6 (60.00)	12 (60.00)
No	4 (40.00)	4 (40.00)	8 (40.00)
Number of families, n	262	262	524
Outdoor temperature (℃), mean (SD)	23.02 (1.68)	24.81 (2.21)	23.93 (2.16)
Children level			
Sex, n (%)			
Boys	142 (54.20)	146 (55.73)	288 (54.96)
Girls	120 (45.80)	116 (44.27)	236 (45.04)
Age (years), mean (SD)	9.10 (0.35)	9.22 (0.34)	9.16 (0.35)
Weight (kg), mean (SD)	28.46 (5.95)	29.30 (6.70)	28.88 (6.34)
BMI (kg/m^2), mean (SD)	15.93 (2.34)	15.81 (2.79)	15.87 (2.57)
Physical activity, n (%)			
Yes	91 (34.73)	96 (36.64)	187 (35.69)
No	171 (65.27)	166 (63.36)	337 (64.31)
Adult level			
Sex, n (%)			
Male	97 (37.02)	97 (37.02)	194 (37.02)
Female	165 (62.98)	165 (62.98)	330 (62.98)
Age (years), mean (SD)	40.96 (11.07)	41.01 (11.03)	40.99 (11.04)
Weight, mean (SD)	59.98 (10.48)	59.49 (10.02)	59.74 (10.24)
BMI, mean (SD)	23.76 (3.26)	23.23 (3.20)	23.50 (3.24)
Relationship with children, n (%)			
Parents	207 (79.01)	217 (82.82)	424 (80.92)
Grandparents	50 (19.08)	43 (16.41)	93 (17.75)
Other	5 (1.91)	2 (0.76)	7 (1.34)
Physical activity, n (%)			
Yes	98 (37.40)	88 (33.59)	186 (35.50)
No	164 (62.60)	174 (66.41)	338 (64.50)
Education, n (%)			
≤Primary school	50 (19.08)	39 (14.89)	89 (16.98)
Secondary education	117 (44.66)	106 (40.46)	223 (42.56)
High school education	51 (19.47)	57 (21.76)	108 (20.61)
College or above	44 (16.79)	60 (22.90)	104 (19.85)
Smoking, n (%)			
Yes	52 (19.85)	49 (18.70)	101 (19.27)

XSL•FO
RenderX

Characteristics	Intervention	Control	Total
No	210 (80.15)	213 (81.30)	423 (80.73)
Alcohol drinking, n (%)			
Never	158 (60.31)	151 (57.63)	309 (58.97)
Occasional	96 (36.64)	97 (37.02)	193 (36.83)
Regular	8 (3.05)	14 (5.34)	22 (4.20)
Self-reported hypertension, n (%)			
Yes	20 (7.63)	20 (7.63)	40 (6.62)
No	242 (92.37)	242 (92.37)	484 (92.37)

Evaluation of the Intervention

During the process of intervention implementation, all schools in the intervention group systematically cultivated an environment conducive to low-sodium practices in accordance with the prescribed program. Health education lessons incorporating salt reduction content, which were required to be conducted at least once every 2 weeks, were effectively implemented across all participating schools, offline activities were also carried out successfully. Furthermore, low-sodium materials provided by our research team, including videos, PowerPoint, posters, and pamphlets, were integrated into various educational activities such as health education courses, lectures, as well as thematic class meetings. In addition to these initiatives, we distributed low-sodium picture books and stationery to all families in order to boost the adherence of the participants, garnering a positive reception from students. All families of the intervention group registered and actively engaged with the WeChat-based “Health Cloud Classroom” miniprogram. Timely dissemination of app-based courses was ensured under the collective guidance of researchers and teachers. Finally, 84.7% (222/262) families successfully fulfilled the requirements of the app-based curriculum pertaining to low-sodium education in the “Health Cloud Classroom.” Overall, adherence to the intervention was satisfactory and the process evaluation of EduSaltS could be found in the published paper [28].

Table 2 presents the measurements and changes in salt intake, blood pressure, urinary sodium, potassium, and sodium-potassium ratio for the intervention and control groups. During the study of 2 semesters, there was a slight upward trend in salt intake and SBP among students. The salt intake of students was 5.4 (SE 0.2) g/day and 5.5 (SE 0.2) g/day in the intervention and control group at baseline, and it increased by 0.41 g/day (95% CI 0.01-0.82) and 0.61 g/day (95% CI 0.21-1.02) respectively after 2 semesters. The mean effect comparing the intervention group with the control group was –0.24 g/day (95% CI –0.82 to 0.33) after adjustment for covariates. Similarly, the blood pressure of students showed the same trend. The adjusted mean difference in the change of SBP and DBP for students between intervention and control groups was –0.68 mm Hg (95% CI –2.32 to 0.95) and –1.37 mm Hg (95% CI –2.79 to 0.06), respectively.

For adults, the intervention group showed a reduction in salt intake (g/day) and urinary sodium (mmol/24 hours), while the control group showed a slight increase, resulting in a significant difference. The salt intake in the intervention group decreased from 9.0 (SE 0.2) g/day to 8.3 (SE 0.2) g/day during the study, a change of –0.79 g/day (95% CI –1.33 to –0.25). In contrast, the control group exhibited a slight increase of 0.26 g/day (95% CI –0.27 to 0.80), from 9.1 (SE 0.2) g/day to 9.3 (SE 0.2) g/day. After adjusting for confounding factors, the mean intervention effect between the 2 groups was –1.06 g/day (95% CI –1.81 to –0.30), which was statistically significant. Regarding blood pressure, there were slight changes in the intervention group, whereas the control group showed a marked increase. The adjusted intervention effect, comparing the intervention group with the control group, was –2.26 mm Hg (95% CI –4.26 to –0.26) for SBP and –2.33 mm Hg (95% CI –3.84 to –0.82) for DBP. Further details are available in Table 2.

Table 3 shows the results of the sensitivity analysis. The adjusted mean effects comparing the intervention group with the control group on salt intake in children (–0.27 g/day, 95% CI –0.85 to 0.30) and in adults (–1.01 g/day, 95% CI –1.75 to –0.26) were similar to the main results when potentially incomplete 24-hour urine collections were included.

In the analysis of subgroups, as shown in Table S4 in Multimedia Appendix 1, no significant difference was seen among any of the interactions of subgroups with the intervention effect on salt intake for both children and adults. Relatively, younger adults (age<40 years), males, nonsmokers, being overweight (BMI≥24), and participants who were parents appeared to experience significantly greater effects of intervention. However, these outcomes may lack robustness due to the small sample size. Detailed results of the analysis of subgroups for SBP and DBP can be found in Tables S5 and S6 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 2. Salt intake (g/24h, as calculated from 24-hour urinary sodium excretion), blood pressure, and other 24-hour urinary measurements based on intention-to-treat analysis.

	Intervention group				Control group				Adjusted difference in change[a]	
	Number[b]	Baseline, mean (SE)	12 months, mean (SE)	Change[c], (95% CI)	Number[b]	Baseline, mean (SE)	12 months, mean (SE)	Change[c], (95% CI)	Intervention vs control, (95% CI)	*P* value
Children										
Salt intake	261	5.4 (0.2)	5.7 (0.2)	0.41 (0.01 to 0.82)	261	5.5 (0.2)	6.1 (0.2)	0.61 (0.21 to 1.02)	–0.24 (–0.82 to 0.33)	.41
SBP[d]	262	94.6 (0.6)	97.2 (0.6)	2.65 (1.50 to 3.80)	262	94.8 (0.6)	98.0 (0.6)	3.22 (2.07 to 4.37)	–0.68 (–2.32 to 0.95)	.41
DBP[d]	262	60.5 (0.4)	60.6 (0.4)	0.09 (–0.90 to 1.09)	262	58.9 (0.5)	60.4 (0.5)	1.45 (0.45 to 2.45)	–1.37 (–2.79 to 0.06)	.06
Na[e]	261	91.8 (2.6)	98.2 (2.8)	7.06 (0.21 to 13.93)	261	93.4 (3.1)	104.1 (2.9)	10.50 (3.57 to 17.42)	–4.15 (–13.94 to 5.65)	.41
K[e]	261	21.6 (0.5)	23.8 (0.7)	2.40 (0.60 to 4.19)	261	23.9 (0.7)	25.4 (0.9)	1.37 (–0.45 to 3.18)	1.01 (–1.54 to 3.57)	.44
Na/K[e]	261	4.6 (0.1)	4.7 (0.2)	0.10 (–0.24 to 0.44)	261	4.5 (0.2)	4.8 (0.1)	0.32 (–0.03 to 0.66)	–0.26 (–0.75 to 0.22)	.29
Adults										
Salt intake	256	9.0 (0.2)	8.3 (0.2)	–0.79 (–1.33 to –0.25)	257	9.1 (0.2)	9.3 (0.2)	0.26 (–0.27 to 0.80)	–1.06 (–1.81 to –0.30)	.007
SBP[d]	262	112.9 (1.0)	112.9 (1.0)	0.43 (–1.01 to 1.87)	262	111.5 (1.1)	113.8 (1.1)	2.89 (1.47 to 4.32)	–2.26 (–4.26 to –0.26)	.03
DBP[d]	262	73.0 (0.7)	72.4 (0.7)	–0.47 (–1.57 to 0.63)	262	71.5 (0.7)	73.3 (0.7)	2.05 (0.96 to 3.14)	–2.33 (–3.84 to –0.82)	.003
Na[e]	256	154.3 (4.2)	141.4 (4.1)	–13.47 (–22.76 to –4.19)	257	155.3 (4.0)	159.1 (4.1)	4.50 (–4.61 to 13.61)	–18.04 (–31.02 to –5.05)	.007
K[e]	256	31.7 (0.7)	33.0 (0.9)	1.26 (–0.83 to 3.34)	257	33.9 (0.9)	33.1 (1.0)	–0.67 (–2.72 to 1.38)	2.10 (–0.83 to 5.03)	.16
Na/K[e]	256	5.2 (0.1)	4.8 (0.2)	–0.42 (–0.79 to –0.05)	257	5.2 (0.2)	5.3 (0.2)	0.19 (–0.17 to 0.55)	–0.64 (–1.15 to –0.13)	.01

[a]Comparison between intervention and control groups in the changes from baseline to 12-month follow-up. Positive values=the intervention group had a greater increase or less decrease from baseline to 12-month follow-up than the control group; negative values=the intervention group had a greater decrease or smaller increase from baseline to 12-month follow-up than the control group. Results were adjusted for age, sex, BMI (body weight in children instead), district or county, physical exercise, and education level (the education level of the familial participant in children instead). In adults, additional adjustments were made for smoking, alcohol consumption, and relationship with the child. Blood pressure values were further adjusted for outdoor temperature.

[b]Number of participants included in the analysis.

[c]Comparison of the means between baseline and 12-month follow-up. Positive values=increases from baseline to 12-month follow-up; negative values=reductions from baseline to 12-month follow-up. Results were obtained from a mixed linear model taking into account the hierarchical structure of data.

[d]Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are expressed in mm Hg (millimeters of mercury).

[e]Na and K represent the 24-hour urinary sodium excretion and urinary potassium excretion, respectively, both expressed in mmol/L. The Na/K ratio refers to the urinary sodium-to-potassium ratio.

XSL•FO
RenderX

Table 3. Salt intake (g/24h, as calculated from 24-hour urinary sodium excretion) and other 24-hour urinary measurements based on sensitivity analysis.

	Intervention group				Control group				Adjusted difference in change[a]	
	Number[b]	Baseline, mean (SE)	12 months, mean (SE)	Change[c], (95% CI)	Number[b]	Baseline, mean (SE)	12 months, mean (SE)	Change[c], (95% CI)	Intervention vs Control, (95% CI)	*P* value
Children										
Salt intake	262	5.2 (0.1)	5.6 (0.2)	0.36 (–0.04 to 0.76)	262	5.3 (0.2)	5.9 (0.2)	0.59 (0.19 to 1.00)	–0.27 (–0.85 to 0.30)	.35
Na[d]	262	89.3 (2.5)	95.4 (2.8)	6.15 (–0.75 to 13.05)	262	90.2 (3)	100.2 (2.9)	10.12 (3.21 to 17.02)	–4.69 (–14.52 to 5.13)	.35
K[d]	262	21.0 (0.5)	23.2 (0.7)	2.21 (0.39 to 4.02)	262	23.0 (0.7)	24.6 (0.9)	1.56 (–0.26 to 3.38)	0.67 (–1.91 to 3.24)	.61
Na/K[d]	262	4.6 (0.1)	4.7 (0.1)	0.07 (–0.26 to 0.41)	262	4.5 (0.2)	4.7 (0.1)	0.28 (–0.05 to 0.62)	–0.26 (–0.73 to 0.21)	.28
Adults										
Salt intake	262	8.8 (0.2)	7.9 (0.2)	–0.90 (–1.43 to –0.37)	262	9.0 (0.2)	9.1 (0.2)	0.15 (–0.37 to 0.68)	–1.01 (–1.75 to –0.26)	.008
Na[d]	262	151.1 (4.2)	135.7 (4.0)	–15.46 (–24.51 to –6.40)	262	153.0 (4.0)	155.6 (4.0)	2.60 (–6.37 to 11.57)	–17.22 (–29.92 to –4.53)	.008
K[d]	262	30.9 (0.7)	31.3 (0.9)	0.40 (–1.64 to 2.44)	262	33.2 (0.9)	32.0 (1.0)	–1.24 (–3.26 to 0.78)	1.94 (–0.93 to 4.81)	.18
Na/K[d]	262	5.2 (0.1)	4.8 (0.2)	–0.41 (–0.76 to –0.07)	262	5.3 (0.2)	5.4 (0.2)	0.21 (–0.14 to 0.56)	–0.64 (–1.13 to –0.15)	.01

[a]Comparison between intervention and control groups in the changes from baseline to 12-month follow-up. Positive values=the intervention group had a greater increase or less decrease from baseline to 12-month follow-up than the control group; negative values=the intervention group had a greater decrease or smaller increase from baseline to 12-month follow-up than the control group. Results were adjusted for age, sex, BMI (body weight in children instead), district or county, physical exercise, and education level (the education level of the familial participant in children instead). In adults, additional adjustments were made for smoking, alcohol consumption, and relationship with the child. Blood pressure values were further adjusted for outdoor temperature.

[b]Number of participants included in the analysis.

[c]Comparison of the means between baseline and 12-month follow-up. Positive values=increases from baseline to 12-month follow-up; negative values=reductions from baseline to 12-month follow-up. Results were obtained from a mixed linear model taking into account of the hierarchical structure of data.

[d]Na and K represent the 24-hour urinary sodium excretion and urinary potassium excretion, respectively, both expressed in mmol/L. The Na/K ratio refers to the urinary sodium-to-potassium ratio.

Discussion

Principal Results

This study deployed an mHealth- and school-based health education program for salt reduction (EduSaltS), a hybrid of online and offline methods, tailored for integration into school health education and promotion initiatives, which used a “small hands leading big hands” approach to cultivate awareness and practices of salt reduction among students and their families. It was the first time to evaluate the effectiveness of this scale-up program for salt reduction using an embedded randomized controlled trial.

Currently, salt reduction and hypertension prevention efforts among the general population in China are primarily carried out through public health education activities, such as the “Three Reductions and Three Health” campaign, and the national basic public health services. Although these measures play a pivotal role in salt reduction and hypertension prevention strategies in China, it is challenging for such efforts to reach or impact every household. The EduSaltS approach, however, has the potential to balance both broad coverage and effectiveness, which can serve as a good complement to existing strategies. Our research achieved high levels of participation and cooperation over an academic year, with follow-up rates of students surpassing 95% and adults over 90%. The strength of the EduSaltS intervention lies in its comprehensive strategy that integrates traditional school-based health education with the modern advantages of mHealth technology. This approach was not commonly observed in other school-based cardiovascular interventions, which often focused exclusively on students, for example, interventions such as the Active Smarter Kids physical activity intervention concentrated on increasing physical activity to improve cardiovascular health markers [29]. Similarly, the Salud Integral Program also focused on in-person interventions without integrating mHealth strategies that could have enhanced its overall impact [30]. Moreover, the EduSaltS intervention extended its impact beyond the individual by incorporating the

XSL•FO
RenderX

"small hands leading big hands" approach, which encourages collaborative efforts within families to adopt low-salt dietary habits. In contrast, the interventions such as using SaltSwitch (The George Institute for Global Health) smartphone app [31] and WhatsApp (WhatsApp LLC) [32], while effective in their own right, focused more on direct educational outreach to individuals, without explicitly addressing family or community-level behavior change. In essence, the EduSaltS intervention's combination of app-based engagement and offline activities, along with its emphasis on school and family involvement, positions it as a more holistic and potentially more effective strategy for reducing salt intake at both the individual and population levels.

Despite the relatively small and nonsignificant effect of the intervention on students' salt intake, some effects of the intervention on both salt intake and blood pressure were observed among adults compared with the control group. The intervention period overlapped with a significant shift of the quarantine strategy of COVID-19 in China, from stringent isolation controls in 2022, limiting outings and social gatherings, to a relaxation of restrictions at the beginning of 2023, leading to an increase in dining out and leisure activities. From periods of home isolation to the resumption of regular social interactions, there may have been substantial changes in people's dietary behaviors and work-life habits. In this context, the adult participants in the control group showed a nonsignificant rise in salt intake and an obvious increase in blood pressure, whereas the intervention group experienced a notable reduction in salt intake and a slight decrease in DBP, underscoring the effectiveness of the intervention among adults. Subgroup analysis indicated that the EduSaltS program seemed to be more effective among younger adults and those who are parents compared with grandparents. This result may be attributed to the familiarity of those populations with electronic devices and receptiveness to app-based information [33]. However, given the smaller sample size, the outcome of the subgroup analysis needs cautious interpretation.

The observed nonsignificant effects of the intervention in students may potentially be ascribed to a confluence of multifaceted factors. First of all, the salt intake among students is subject to a myriad of influences, including familial dietary patterns, the dining environment within and around the vicinity of schools, personal preferences, as well as levels of health consciousness. During the implementation of EduSaltS in Ganzhou City, the primary focus was on the family dimension and the completion of app-based courses. However, implementation in the school context, particularly in school catering, was not thoroughly actualized. Concurrently, elementary students exhibit a predilection for high-sodium snacks and fast foods, with their dietary choices markedly swayed by peer influences and personal taste preferences, often at the expense of healthful considerations. Such tendencies are detrimental to achieving the anticipated effectiveness of the intervention. Second, the comprehensive intervention in this study largely relied on the WeChat miniprogram named "Health Cloud Classroom," necessitating parental involvement using smartphones or tablets alongside their children. However, some parents, merely to comply with teachers' requests, operated the program themselves, resulting in interventions that often affected only the parents without the students, hardly obtaining the desired outcome. Furthermore, the same content of health intervention may not be as comprehensively understood by students as by parents. This disparity in comprehension and a less holistic grasp of health concepts are likely to result in an inadequate acquisition of knowledge and a lack of behavioral change among students.

Public Health Implications

It is noteworthy that, although the findings of this study resembled, to some extent, that of our earlier study [14], the intervention incorporated here was broader in scope, more diverse in format, and easier to implement. As a result, the intervention schools have successfully established a supportive environment for salt reduction, and there has been a high level of cooperation among teachers, students, and parents in completing both app-based courses and offline activities. Building upon the already developed health education materials and platforms of intervention, the overall execution was relatively uncomplicated, just requiring primarily efficacious organization by schools and weekly opportune reminders by class teachers following brief training. Additionally, the resources within the system could assist teachers in managing their existing required health education courses and activities more efficiently, thus reducing the need for substantial manpower and resources. Previous research has demonstrated that a reduction of 1 g/day in per capita salt intake could lead to a decrease of nearly 9 million cardiovascular events by 2030 [34]. Furthermore, based on a recent meta-analysis of blood pressure treatment trials [35], the prevented increase of 2.26 mm Hg in SBP, compared with the rise observed in the control group in our study, is projected to reduce the risk of stroke by 5.88% and ischemic heart disease by 3.62%, which would prevent nearly 231,400 strokes and about 126,700 ischaemic heart disease events per year in China if the program was scaled up across the country [36]. Consequently, the initiative of EduSaltS, with its notable practicality, may not only facilitate the enhancement of health education curricula and activities in schools but also significantly curtail the prevalence of CVD within the population, rendering it of substantial value for the guidance of health policies and public health.

Limitations

This study still has some limitations. For instance, during the implementation of the intervention, many aspects required the use of smartphones or tablets, which were not readily accessible in less developed, remote mountain areas, and for students only living with grandparents. Despite the assurance of access to basic public health and health care services in these populations, the digital divide and limitations in school infrastructure may hinder the implementation and effectiveness of the intervention. Thus, we should consider how to promote this mHealth technology–based health education in impoverished and underdeveloped areas, which may serve as a good supplement for the prevention and treatment of hypertension in regions with relatively lower medical care standards. Additionally, a minority of parents perceived this intervention as increasing children's screen time, potentially detrimental to their eyesight, thereby

XSL•FO
RenderX

engendering a degree of resistance. Furthermore, the intervention for school chefs and food service providers was limited to self-learning materials and training once per semester, without monitoring the salt used in school meals. Recording and monitoring salt purchase and use had been suggested but not implemented due to various difficulties. These are what we aim to improve in the future. Subsequent research endeavors will further analyze its generalizability and attempt to evaluate the long-term effects of the intervention. Our team will also incorporate feedback from process evaluations to refine and better integrate the comprehensive intervention strategy into the construction of healthy schools, ultimately enhancing the health levels of students and their parents.

Conclusion

Based on extensive and long-term exploratory efforts, an attractive, appropriate, and comprehensive intervention program for salt reduction (EduSaltS) was developed and implemented in China. Demonstrating favorable feasibility and efficacy, this program is potentially well-suited for nationwide implementation based on collaboration across multiple departments. The intervention, adopting a “small hands leading big hands” approach, effectively reduced the salt intake and blood pressure among adults, while its impact on students was not statistically significant. Future enhancements to the offline components of the intervention, such as improving salt reduction interventions in school catering, could further facilitate the development of low-salt dietary habits in schoolchildren.

Acknowledgments

This work was supported by the Medical Research Council (MRC) and the Department of Health and Social Care (DHSC) through the National Institute for Health Research (NIHR; grant MR/T024399/1), under the Global Alliance for Chronic Disease (GACD) Scale-up Research Program, and Science and Technology Plan of Health Commission of Jiangxi Province (202131030). The views expressed in this publication are those of the authors and not necessarily those of the MRC, NIHR, or DHSC. We thank all the participants who were involved in the research. We are also grateful to the involved personnel from the local bureau of education and health committee for coordination, as well as the staff from local institutions for health education and the hospitals for contributing to the data collection.

Data Availability

The datasets used or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LW and PZ led the conception and design of the study. Yinghua Li and FJH contributed to the research design and analysis of data. LL, Yuan Li, RL, DW, LX, LD, and CW were involved in the study implementation and quality control. CW participated in collecting data and jointly performed statistical analyses with NW. NW wrote the first draft of the manuscript and all authors contributed to the revision and approved the final manuscript. LW and PZ (pzhang@bjtjzx.com) contributed equally to this work and should be considered joint corresponding authors.

Conflicts of Interest

FJH is an unpaid member of Action on Salt and World Action on Salt, Sugar, and Health (WASSH). The other authors declare that they have no competing interests.

Editorial notice: This randomized study was only retrospectively registered, explained by authors due to lack of timely communication with the overall project team and that this trial was embedded in a prospectively registered scale-up study (ChiCTR2000039767). The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study methods mirrored the prospectively registered scale-up studies methods and receive prospective ethics approval. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Additional tables.
[DOCX File , 72 KB - jmir_v27i1e60092_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2145 KB - jmir_v27i1e60092_app2.pdf]

References

XSL•FO
RenderX

1. He FJ, Tan M, Ma Y, MacGregor GA. Salt reduction to prevent hypertension and cardiovascular disease: JACC state-of-the-art review. J Am Coll Cardiol 2020 Feb 18;75(6):632-647 [FREE Full text] [doi: 10.1016/j.jacc.2019.11.055] [Medline: 32057379]
2. Lawes CMM, Vander Hoorn S, Rodgers A, International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. Lancet 2008 May 03;371(9623):1513-1518. [doi: 10.1016/S0140-6736(08)60655-8] [Medline: 18456100]
3. Guilbert JJ. The world health report 2002 - reducing risks, promoting healthy life. Educ Health (Abingdon) 2003 Jul;16(2):230. [doi: 10.1080/1357628031000116808] [Medline: 14741909]
4. Zhou M, Wang H, Zeng X, Yin P, Zhu J, Chen W, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet 2019 Sep 28;394(10204):1145-1158 [FREE Full text] [doi: 10.1016/S0140-6736(19)30427-1] [Medline: 31248666]
5. Mozaffarian D, Fahimi S, Singh GM, Micha R, Khatibzadeh S, Engell RE, Global Burden of Diseases NutritionChronic Diseases Expert Group. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014 Aug 14;371(7):624-634. [doi: 10.1056/NEJMoa1304127] [Medline: 25119608]
6. Shao S, Hua Y, Yang Y, Liu X, Fan J, Zhang A, et al. Salt reduction in China: a state-of-the-art review. Risk Manag Healthc Policy 2017;10:17-28 [FREE Full text] [doi: 10.2147/RMHP.S75918] [Medline: 28260957]
7. He FJ, MacGregor GA. A comprehensive review on salt and health and current experience of worldwide salt reduction programmes. J Hum Hypertens 2009 Jun;23(6):363-384. [doi: 10.1038/jhh.2008.144] [Medline: 19110538]
8. Sodium reduction. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/salt-reduction [accessed 2025-03-06]
9. Borrelli S, Provenzano M, Gagliardi I, Michael A, Liberti ME, De Nicola L, et al. Sodium intake and chronic kidney disease. Int J Mol Sci 2020 Jul 03;21(13):4744 [FREE Full text] [doi: 10.3390/ijms21134744] [Medline: 32635265]
10. He FJ, Li J, Macgregor GA. Effect of longer term modest salt reduction on blood pressure: Cochrane systematic review and meta-analysis of randomised trials. BMJ 2013 Apr 03;346:f1325 [FREE Full text] [doi: 10.1136/bmj.f1325] [Medline: 23558162]
11. Wang YJ, Yeh TL, Shih MC, Tu YK, Chien KL. Dietary sodium intake and risk of cardiovascular disease: a systematic review and dose-response meta-analysis. Nutrients 2020 Sep 25;12(10):2934 [FREE Full text] [doi: 10.3390/nu12102934] [Medline: 32992705]
12. Huang L, Trieu K, Yoshimura S, Neal B, Woodward M, Campbell NRC, et al. Effect of dose and duration of reduction in dietary sodium on blood pressure levels: systematic review and meta-analysis of randomised trials. BMJ 2020 Feb 24;368:m315 [FREE Full text] [doi: 10.1136/bmj.m315] [Medline: 32094151]
13. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: systematic review and meta-analyses. BMJ 2013 Apr 03;346:f1326 [FREE Full text] [doi: 10.1136/bmj.f1326] [Medline: 23558163]
14. He FJ, Zhang P, Luo R, Li Y, Sun Y, Chen F, et al. App based education programme to reduce salt intake (AppSalt) in schoolchildren and their families in China: parallel, cluster randomised controlled trial. BMJ 2022 Feb 09;376:e066982 [FREE Full text] [doi: 10.1136/bmj-2021-066982] [Medline: 35140061]
15. Xu A, Ma J, Guo X, Wang L, Wu J, Zhang J, et al. Association of a province-wide intervention with salt intake and hypertension in Shandong Province, China, 2011-2016. JAMA Intern Med 2020 Jun 01;180(6):877-886 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0904] [Medline: 32338717]
16. He FJ, Zhang P, Li Y, MacGregor GA. Action on salt China. Lancet 2018 Jul 07;392(10141):7-9. [doi: 10.1016/S0140-6736(18)31138-3] [Medline: 30047401]
17. He FJ, Wu Y, Feng X, Ma J, Ma Y, Wang H, et al. School based education programme to reduce salt intake in children and their families (School-EduSalt): cluster randomised controlled trial. BMJ 2015 Mar 18;350:h770 [FREE Full text] [doi: 10.1136/bmj.h770] [Medline: 25788018]
18. He FJ, Zhang P, Luo R, Li Y, Chen F, Zhao Y, et al. An application-based programme to reinforce and maintain lower salt intake (AppSalt) in schoolchildren and their families in China. BMJ Open 2019 Jul 03;9(7):e027793 [FREE Full text] [doi: 10.1136/bmjopen-2018-027793] [Medline: 31272977]
19. Neal B, Wu Y, Feng X, Zhang R, Zhang Y, Shi J, et al. Effect of salt substitution on cardiovascular events and death. N Engl J Med 2021 Sep 16;385(12):1067-1077. [doi: 10.1056/NEJMoa2105675] [Medline: 34459569]
20. Liu X, Tian H. An intervention study on smoking in children and their family members in Chengdu. China Health Manag 2004;2:117-121. [doi: 10.3969/j.issn.1004-4663.2004.02.026]
21. Fornari LS, Giuliano I, Azevedo F, Pastana A, Vieira C, Caramelli B. Children first study: how an educational program in cardiovascular prevention at school can improve parents' cardiovascular risk. Eur J Prev Cardiol 2013 Apr;20(2):301-309. [doi: 10.1177/2047487312437617] [Medline: 22345689]
22. Gunawardena N, Kurotani K, Indrawansa S, Nonaka D, Mizoue T, Samarasinghe D. School-based intervention to enable school children to act as change agents on weight, physical activity and diet of their mothers: a cluster randomized controlled trial. Int J Behav Nutr Phys Act 2016 Apr 06;13:45 [FREE Full text] [doi: 10.1186/s12966-016-0369-7] [Medline: 27048282]

XSL•FO
RenderX

23. Nagamitsu S, Kanie A, Sakashita K, Sakuta R, Okada A, Matsuura K, et al. Adolescent health promotion interventions using well-care visits and a smartphone cognitive behavioral therapy app: randomized controlled trial. JMIR Mhealth Uhealth 2022 May 23;10(5):e34154 [FREE Full text] [doi: 10.2196/34154] [Medline: 35604760]
24. Sakane N, Suganuma A, Domichi M, Sukino S, Abe K, Fujisaki A, et al. The effect of a mHealth app (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2023 Apr 12;11:e43236 [FREE Full text] [doi: 10.2196/43236] [Medline: 37043287]
25. He FJ, Wu Y, Ma J, Feng X, Wang H, Zhang J, et al. A school-based education programme to reduce salt intake in children and their families (School-EduSalt): protocol of a cluster randomised controlled trial. BMJ Open 2013;3(7):E003388 [FREE Full text] [doi: 10.1136/bmjopen-2013-003388] [Medline: 23864214]
26. Zhang P, Sun J, Li Y, Li Y, Sun Y, Luo R, et al. An mHealth-based school health education system designed to scale up salt reduction in China (EduSaltS): a development and preliminary implementation study. Front Nutr 2023;10:1161282 [FREE Full text] [doi: 10.3389/fnut.2023.1161282] [Medline: 37139455]
27. Hayes RJ, Moulton LH. Cluster Randomized Trials. Boca Raton, Florida: CRC Press; 2009.
28. Guo H, Li Y, Li L, Luo R, Wang L, Yi G, et al. Process evaluation of an mHealth-based school education program to reduce salt intake scaling up in China (EduSaltS): a mixed methods study using the RE-AIM framework. BMC Public Health 2024 Aug 20;24(1):2261 [FREE Full text] [doi: 10.1186/s12889-024-19732-y] [Medline: 39164700]
29. Stavnsbo M, Aadland E, Anderssen SA, Chinapaw M, Steene-Johannessen J, Andersen LB, et al. Effects of the active smarter kids (ASK) physical activity intervention on cardiometabolic risk factors in children: a cluster-randomized controlled trial. Prev Med 2020 Jan;130:105868 [FREE Full text] [doi: 10.1016/j.ypmed.2019.105868] [Medline: 31654725]
30. Santos-Beneit G, Fernández-Alvira JM, Tresserra-Rimbau A, Bodega P, de Cos-Gandoy A, de Miguel M, et al. School-based cardiovascular health promotion in adolescents: a cluster randomized clinical trial. JAMA Cardiol 2023 Sep 01;8(9):816-824 [FREE Full text] [doi: 10.1001/jamacardio.2023.2231] [Medline: 37531100]
31. Eyles H, Grey J, Jiang Y, Umali E, McLean R, Te Morenga L, et al. Effectiveness of a sodium-reduction smartphone app and reduced-sodium salt to lower sodium intake in adults with hypertension: findings from the salt alternatives randomized controlled trial. JMIR Mhealth Uhealth 2023 Mar 09;11:e43675 [FREE Full text] [doi: 10.2196/43675] [Medline: 36892914]
32. Jarrar AH, Al Dhaheri AS, Lightowler H, Cheikh Ismail L, Al-Meqbaali F, Bataineh MF, et al. Using digital platform approach to reduce salt intake in a sample of UAE population: an intervention study. Front Public Health 2022;10:860835 [FREE Full text] [doi: 10.3389/fpubh.2022.860835] [Medline: 35685760]
33. König LM, Krukowski RA, Kuntsche E, Busse H, Gumbert L, Gemesi K, et al. Reducing intervention- and research-induced inequalities to tackle the digital divide in health promotion. Int J Equity Health 2023 Dec 04;22(1):249 [FREE Full text] [doi: 10.1186/s12939-023-02055-6] [Medline: 38049789]
34. Tan M, He F, Morris JK, MacGregor G. Reducing daily salt intake in China by 1 g could prevent almost 9 million cardiovascular events by 2030: a modelling study. BMJ Nutr Prev Health 2022 Dec;5(2):164-170 [FREE Full text] [doi: 10.1136/bmjnph-2021-000408] [Medline: 36619331]
35. Blood Pressure Lowering Treatment Trialists' Collaboration. Pharmacological blood pressure lowering for primary and secondary prevention of cardiovascular disease across different levels of blood pressure: an individual participant-level data meta-analysis. Lancet 2021 May 01;397(10285):1625-1636 [FREE Full text] [doi: 10.1016/S0140-6736(21)00590-0] [Medline: 33933205]
36. GBD 2019 DiseasesInjuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]

Abbreviations

CVD: cardiovascular disease
DBP: diastolic blood pressure
EduSaltS: mHealth and school-based education program to reduce salt intake scaling up in China
mHealth: mobile health
SBP: systolic blood pressure
School-EduSalt: school-based education program to reduce salt intake in children and their families

XSL•FO
RenderX

Edited by N Cahill; submitted 02.05.24; peer-reviewed by S Payne, RR Ruff; comments to author 27.08.24; revised version received 14.10.24; accepted 27.02.25; published 27.03.25.

Please cite as:
Wang N, Wang C, Zhang P, Li Y, He FJ, Li L, Li Y, Luo R, Wan D, Xu L, Deng L, Wu L
Effectiveness of an mHealth- and School-Based Health Education Program for Salt Reduction (EduSaltS) in China: Cluster Randomized Controlled Trial Within Scale-Up
J Med Internet Res 2025;27:e60092
URL: https://www.jmir.org/2025/1/e60092
doi:10.2196/60092
PMID:40017342

©Naibo Wang, Chen Wang, Puhong Zhang, Yinghua Li, Feng J He, Li Li, Yuan Li, Rong Luo, Dezhi Wan, Lewei Xu, Lifang Deng, Lei Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy and Acceptability of a Mobile App for Monitoring the Clinical Status of Patients With Chronic Obstructive Pulmonary Disease Receiving Home Oxygen Therapy: Randomized Controlled Trial

Anisbed Naranjo-Rojas[1,2*]; Luis Ángel Perula-de Torres[3*]; Freiser Eccehomo Cruz-Mosquera[4*]; Guillermo Molina-Recio[5*]

[1]Health and Education Research Group (GINEYSA), Faculty of Health, Universidad Santiago de Cali, Santiago de Cali, Colombia

[2]Biomedicine Doctoral Program, University of Córdoba, Córdoba, Spain

[3]Research Network on Chronicity Primary Care and Health Promotion (RICAPPS), Cooperative Research Networks Oriented to Health Results (RICORS) Carlos III Health Institute, Madrid, Spain

[4]Integral Health Research Group (GISI), Faculty of Health, Universidad Santiago de Cali, Santiago de Cali, Colombia

[5]Nursing Pharmacology and Physiotherapy Department University of Córdoba, Lifestyles Innovation and Health (GA-16) Maimonides Biomedical Research Institute of Córdoba (IMIBIC) Spain, University of Córdoba, Córdoba, Spain

[*]all authors contributed equally

Corresponding Author:
Anisbed Naranjo-Rojas
Health and Education Research Group (GINEYSA)
Faculty of Health
Universidad Santiago de Cali
Calle 5 # 62 esquina
Santiago de Cali, 763531
Colombia
Phone: 57 602 5183000 ext 6515
Email: anisbednaranjo24@gmail.com

Abstract

Background: Chronic obstructive pulmonary disease (COPD) primarily originates from exposure to tobacco smoke, although factors, such as air pollution and exposure to chemicals, also play a role. One of the primary treatments for COPD is oxygen therapy, which helps manage dyspnea and improve survival rates. Mobile health (mHealth) technologies have demonstrated significant potential in monitoring patients with chronic diseases, offering new avenues for enhancing patient care and disease management.

Objective: The purpose of this study was to evaluate the efficacy and acceptability of a mobile app designed for the clinical monitoring of patients with COPD and home oxygen (HO) therapy, compared with conventional monitoring in real-world community settings.

Methods: A parallel-group, nonblinded, multicenter randomized controlled trial was conducted with 45 participants; the intervention group (IG), which used the mobile app in addition to conventional monitoring (n=23) and the control group, which received only conventional monitoring (n=22), administered by therapists over a duration of 3 months. The primary outcomes included the chronic obstructive pulmonary disease assessment test (CAT) score, the level of dyspnea measured by the Borg scale, and oxygen saturation percentage, assessed at both the beginning and end of the trial. Secondary outcomes included the frequency of app use, the number of hospitalizations, and survival rates. In addition, a satisfaction survey and an interview were conducted with the IG.

Results: The median use of the mobile app was 21 (IQR 16-28) days. At the end of the follow-up, the Borg dyspnea scale was significantly lower in patients who used the mobile app for HO therapy monitoring (mean 0.6, SD 0.8 vs mean 4.1, SD 1.4; *P*=.001). Regarding the impact of COPD on quality of life, as measured by the CAT, no differences were found in the scores between baseline and end-of-follow-up within the control group. However, a significant decrease was observed in the IG (baseline median CAT 27, IQR 23-31 vs final median CAT 22, IQR 14-28; *P*<.001). In addition, the CAT score was significantly higher

XSL•FO
RenderX

in patients receiving conventional monitoring compared with those monitored with the mobile app (median 30, IQR 23-32 vs median 22, IQR 14-28; *P*=.02).

Conclusions: The use of the mobile app, AppO2 (SINCO), designed for the clinical monitoring of patients with COPD and HO therapy, is associated with improved quality of life. In addition, the app is highly accepted by users, promotes self-care, and fosters patient confidence in managing their own condition.

Trial Registration: ClinicalTrials NCT04820790; https://clinicaltrials.gov/study/NCT04820790

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.1186/s12875-021-01450-8

(J Med Internet Res 2025;27:e65888) doi:10.2196/65888

KEYWORDS

m-Health; chronic obstructive pulmonary disease; quality of life; mobile health applications; home oxygen therapy; mobile phone

Introduction

The main cause of chronic obstructive pulmonary disease (COPD) is exposure to tobacco smoke; however, there are also other associated factors such as air pollution, exposure to chemicals, and a history of respiratory diseases [1]. The World Health Organization considers it the third cause of death worldwide, making it a public health problem [2]. It is estimated that approximately 8% of the Colombian population—equivalent to about 3.6 million people—suffers from COPD [3,4]. Oxygen therapy is one of the main treatments for this disease [5], aiding in the control of symptoms such as dyspnea and improving survival [5]. Mesquita et al [6] found patients with COPD who did not adhere to oxygen therapy experienced a long-term decrease in quality of life compared to those who followed their prescribed treatment [7]. Similarly, several authors [6,8,9] have identified factors such as age, communication barriers between doctor and patient, clinical follow-up, and disease knowledge as significant challenges to adhering to oxygen therapy [6]. Low adherence can lead to exacerbations [10], which deteriorate lung function, negatively affect patients' quality of life, and increase mortality [11].

Furthermore, these exacerbations contribute to substantial health care costs. In Colombia, each exacerbation costs an average of US $98 [12], with costs rising to nearly US $700 if hospitalization is required [12]. In addition, this high cost can be attributed to the fact that once a patient is hospitalized, there is a high risk of rehospitalization or death. For example, Niewoehner [13] found that of 1016 patients hospitalized for COPD, some were rehospitalized and 33% (n=335) died within 6 months of discharge.

Conversely, mobile health (mHealth) technologies hold significant promise for monitoring patients with chronic diseases [14]. These apps offer benefits such as remote monitoring, direct communication with health care professionals, recording of clinical signs, and the development of self-care skills [14]. A Cochrane review [15] suggested that self-monitoring of symptoms through mobile apps positively impacts the development of self-care and self-management skills in patients with chronic diseases [9]. Furthermore, studies, such as the one conducted by Knox et al [16], have focused on designing mobile apps for patients with COPD, concluding that reporting clinical information through these apps can aid in disease management [16].

Some authors emphasize the importance of designing these mHealth apps with a user-centered approach [17,18]. Understanding the perceptions and needs of the end user as well as adapting to their sociodemographic characteristics and level of technological literacy are indispensable strategies for developing mHealth apps that increase adherence to use and, consequently, to treatment and self-care [17-19]. Therefore, evaluating the usability [20] of mHealth apps before their implementation is crucial. Conducting usability tests to assess user performance on specific tasks can identify areas for design improvements and enhance user satisfaction [4,21].

Based on the above, the authors of this study designed and developed a mobile app called AppO2 (SINCO) before conducting the research. This app was created using a user-centered design methodology and underwent usability tests [4,19]. AppO2 facilitates the monitoring of the clinical status of patients with COPD receiving home oxygen (HO) therapy, with 2 user profiles—patients (or their caregivers) and health care professionals [4,19]. The functions of the patient profile are centered on self-care skill development and quality of life improvement. These functions include tutorials on measuring and recording vital signs, accessing information related to the prescription of HO treatment, and communicating and interacting with health care professionals [4,19]. Conversely, the functions of the professional profile are designed to monitor patients' clinical status on a daily basis, record and review clinical changes to inform decision-making, and control the dosage of HO therapy [4,19].

Finally, the functionalities of this mobile app were designed to promote self-care in patients with COPD receiving HO therapy. It also sought to address the primary challenges these patients face in improving adherence to HO treatment, preventing or detecting exacerbations early, and enhancing their quality of life [4,19].

In this context, this study aimed at evaluating the efficacy and acceptability of AppO2 in real community settings compared with conventional home care monitoring for patients with COPD-prescribed HO.

Methods

Study Design and Participants

A 3-month, open-label, 2-arm, parallel-arm, multicenter, randomized controlled, nonblinded, clinical trial was conducted

XSL•FO
RenderX

to determine the efficacy of a mobile app in monitoring the clinical status of patients with COPD receiving HO therapy. The control group (CG) comprised patients receiving HO and monitored using conventional methods, such as weekly home visits by a health care professional. The intervention group (IG) comprised patients who were also monitored through weekly home visits in addition to using the AppO2 mobile app. The clinical trial was conducted in Cali, Colombia, with the participation of 3 home care institutions—TodoMed, Amanecer Médico, and Cuidarte en Casa. The clinical trial protocol was registered in ClinicalTrials.gov (ID NCT04820790) [4].

Patients were recruited from May to December 2023 based on the following inclusion criteria: (1) being ≥18 years old, (2) having a medical prescription for HO for more than 1 year, (3) having a caregiver, (4) owning a smartphone, and (5) consenting to audio recording. The exclusion criterion was patients undergoing mechanical ventilation.

In total, 2 health care professionals were included based on the following inclusion criteria: (1) being ≥18 years old, (2) being respiratory therapists and physiotherapists affiliated with a home care company, (3) having more than 6 months of experience in home care, and (4) having a smartphone. The exclusion criterion was established as professionals with less than 6 months of experience in managing patients receiving HO.

Sample Size

To achieve a 1:1 ratio between the IG and CG, with 80% power (β error=20%) and 95% CI (α error=5%), and estimating that the AppO2 app in the IG would result in a 3% increase self-management of dyspnea, with SD 6 [22], a minimum sample of 32 participants (16 in each group) was estimated. Finally, to minimize the effects of potential patient losses that would reduce statistical power, the sample was expanded to a final size of 45 participants (23 in the IG and 22 in the CG) [4].

Randomization

Overview

Participants were equally and randomly assigned to either the IG or CG using a simple randomization method with Epidat 3.1 (Consellería de Sanidade de la Xunta de Galicia, in collaboration with the Pan American Health Organization [PAHO]). Blinding was not possible for participants or health care professionals owing to the intervention method used.

Sample Selection

Consecutive sampling was used to recruit patients and all participants who met the selection criteria and attended consultations at participating institutions were included in the study. Sampling was done until the predetermined sample size was reached.

Intervention

After selecting the trial participants, the IG downloaded the AppO2 app on their mobile phones. The objectives of the investigation and the operation of the app were explained to them. In addition, it was recommended to use the app at least once a week. All recorded information was stored anonymously in a cloud-based database, identifiable by a code only known to the principal investigator. The database was accessible to researchers through a web app for review and analysis.

The AppO2 mobile app consists of 2 user profiles, one for patients and caregivers and another for professionals. There are 3 sections in the patient and caregiver profile, “My Profile,” “Medical Visit,” and “Tutorials.” In the “My Profile” section, patients can record their vital signs and the degree of respiratory difficulty, view the medical prescription for oxygen therapy including the duration of the oxygen therapy regimen, and record the number of daily inhalations if they have a prescription for medication. When recording their vital signs, patients from their user profile receive notifications informing them if vital signs are outside the predetermined ranges. These alerts suggest that they tell their health care professional to adjust treatment as necessary or recommend that they carefully follow medical indications if a slight alteration in vital signs is detected. However, these notifications are not sent to medical staff in real-time. Instead, professionals can view patients’ clinical history as they see fit through their profile [4,19]. This functionality allows them to monitor the behavior of vital signs and make the appropriate adjustments during scheduled home visits, accessible through the “Medical Visit” section. This approach allows for more flexible management and avoids an overload of unnecessary alerts, which could generate “false positives” by prioritizing intervention based on a comprehensive patient assessment. The “Medical Visit” section allowed users to view the schedule of upcoming medical visits and the history of past visits. In addition, the “Tutorials” section featured videos that provided education and instructions on taking vital signs and using oxygen systems [4,19].

Similarly, the professionals’ profile has 2 sections, “My profile,” where they can register their personal data and update their account, if necessary, and “My patients,” where they can record the assessment, prescribe oxygen, and monitor the clinical status of patients. They can also view graphs that display the monthly trends of each vital sign. Through the professionals’ profile, information on HO prescription parameters, visit schedules, and data on the estimated duration of oxygen cylinders are sent to the IG [4] (Figure 1).

XSL•FO
RenderX

Figure 1. AppO2 user profiles. (A) Patient and caregiver profile. (B) Professional profile.

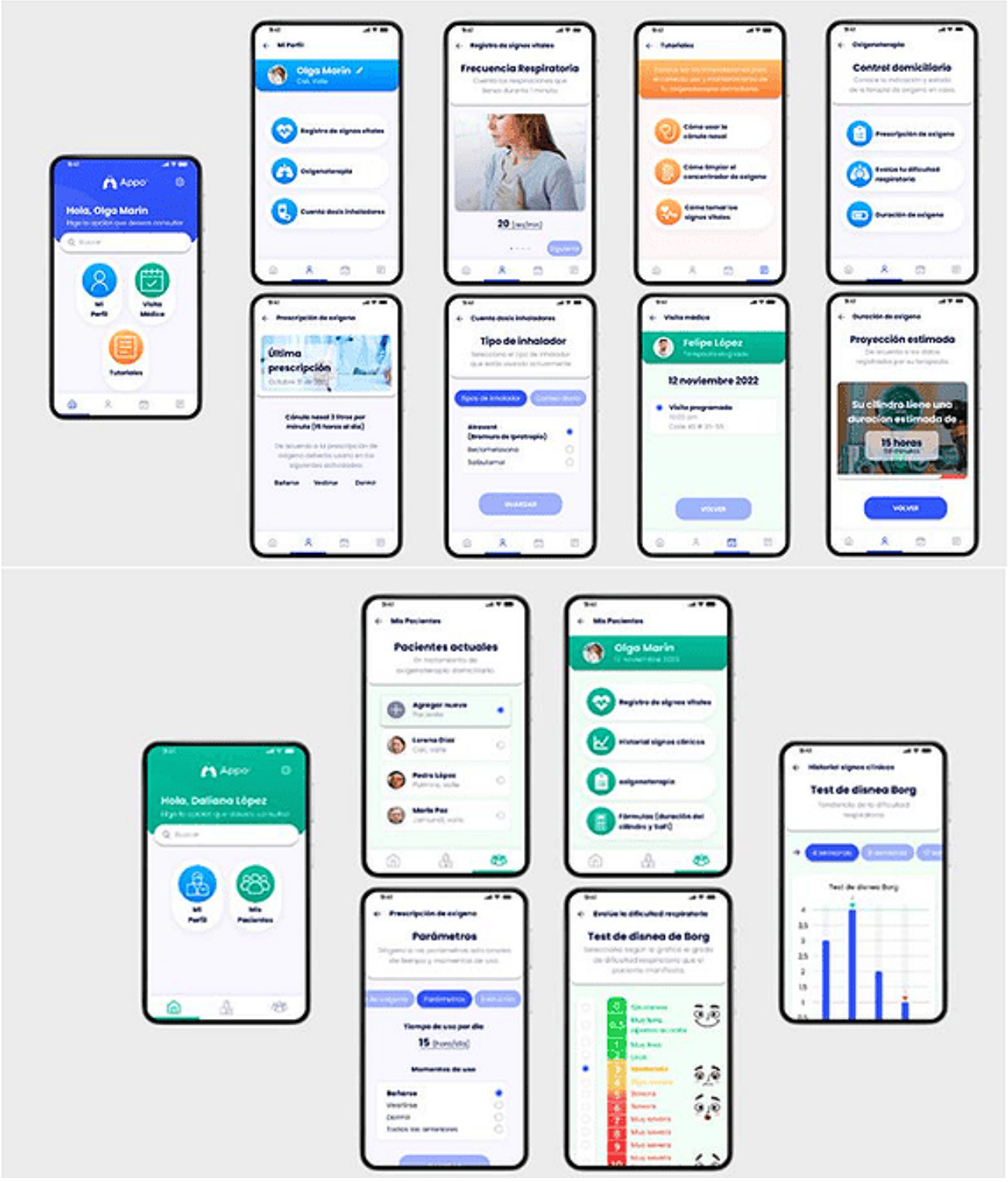

Measures

Independent variables, including participants' sociodemographic characteristics (eg, age, sex, marital status, educational level, and place of residence) and other aspects of lifestyle and treatment (such as smoking, biomass exposure, and duration of HO treatment), were obtained from the databases and record systems of the 3 health institutions to which they belonged.

During the home visits, therapists collected information was collected on variables related to vital signs, such as oxygen saturation through pulse oximetry, respiratory rate, central heart rate, number of exacerbations, number of hospitalizations, number of emergency room admissions, survival, inhaler use, and changes in oxygen prescription. Cylinder duration time was calculated considering the conversion factor, oxygen system flow, residual pressure, and cylinder pressure [23]. This calculation was performed manually in the CG, whereas the mobile app was used in the IG. In addition, therapists reminded patients and caregivers to use the app during each visit. Vital signs and the relationship between oxygen saturation and

XSL•FO
RenderX

fraction of inspired oxygen [24] were recorded once a week from the date of randomization until week 12 or until the time of death from any cause, whichever occurred first.

The degree of respiratory distress was determined using the Borg Dyspnea Scale [25], a standardized, validated analog scale in Spanish, widely used since the 1970s. The scale ranges from 0 to 10 and is designed to quickly and easily assess patients' perception of dyspnea. It includes a graph associated with each quantitative value, which helps patients identify their level of respiratory distress, with 10 indicating the highest level of perceived dyspnea [26]. Assessment using the Borg test was conducted from the beginning of the study until week 12 or until the date of death from any cause, whichever occurred first. In addition, the number of days the app was used, access to each screen, and connection time within the mobile app were tracked in the IG.

Questionnaires

The chronic obstructive pulmonary disease assessment test (CAT) questionnaire was used to assess the impact of COPD on patients' quality of life [27]. This questionnaire was administered to both the IG and CG at the beginning and at the end of the intervention. The acceptance of AppO2 was evaluated using the technology acceptance model (TAM) [28], which assessed the perceived usefulness and ease by participants in relation to AppO2 [28]. This questionnaire, along with an interview to gauge the perception of AppO2, was administered to health care professionals and the IG at the end of the follow-up period.

Chronic Obstructive Pulmonary Disease Assessment Test

This is a validated [29,30] and publicly accessible questionnaire consisting of 8 questions, each evaluated on a scale of 0-5 points. The minimum score is 0, and the maximum score is 40 [29,30], reflecting the impact COPD has on a patient's quality of life. The scores are classified as (1) low impact (1-10 points), where most days are "good days"; (2) medium impact (11-20 points), with few "good days"; (3) high impact (21-30 points), with no "good days"; and (4) very high impact (31-40 points), where the disease's limitation is at its maximum. A progressive increase in CAT scores indicates an increase in the impact of COPD on the patient's quality of life [27,29].

AppO2 Acceptance Questionnaire and Perception Interviews

The AppO2 mobile app acceptance questionnaire was administered to patients and professionals. It was based on the TAM [30] designed by Davis [31,32], aimed at evaluating people's perception of usefulness and ease regarding acceptance in the use of devices or software in digital environments. Although several variations have been made by different authors [33-35], it is essential to evaluate the following dimensions: (1) perception of usefulness and (2) perception of ease. Based on these dimensions, 12 items translated and validated into Spanish [36,37] were applied, with minor adaptations in their content for better understanding.

Responses were assessed using a Likert scale, ranging from "totally agree" to "totally disagree" (Multimedia Appendix 1). To determine the acceptance of the AppO2 mobile app, 1 point was assigned to the item with which participants were moderately or totally in agreement, and 0 was assigned to items with which they were moderately or totally in disagreement. Cutoff points for the level of acceptance were obtained by measuring quartiles. Low acceptance was considered if it was in quartile 1 (1-3 points), moderate if it was between quartiles 2 and 3 (4-9 points), and high if it was in quartile 4 (10-11 points). Furthermore, to complete the evaluation of participants' perceptions regarding the AppO2 app, an interview was conducted with both the IG and health care professionals. The objective of the questions was to inquire about what they liked or disliked most about the app, the perceived benefits, and their preferences. These interviews were conducted at the end of the intervention (Multimedia Appendix 2).

Primary and Secondary Outcomes

Primary outcomes consisted of the CAT score, the degree of dyspnea measured using the Borg test, and the percentage of oxygen saturation at the beginning and end of the clinical trial. Secondary outcomes were the frequency of use of the app, number of hospitalizations, and survival.

Statistical Analysis

Data analysis was performed using SPSS Statistics software, version 25.0 (IBM Corp). Initially, a descriptive analysis of the study variables was conducted, characterizing the patients and generating graphs and tables of absolute and relative frequencies for qualitative variables. For quantitative variables, measures of central tendency and position (mean and median) with their respective measures of dispersion (SD and IQR) were calculated.

In the bivariate analysis, the Student *t* test or its nonparametric equivalent (Mann-Whitney *U* test) was used to compare 2 means. For categorical variables, the chi-square test was used. To analyze 3 or more mean values, repeated measures analysis of variance or the Friedman test was used. Crude odds ratios (ORs) were initially calculated to determine variables independently associated (AppO2 use and sex) with the impact of COPD on quality of life, as measured by the CAT. In addition, an analysis of covariance (ANCOVA) was used, which allowed for adjustment of differences in the final Borg dyspnea scores based on the initial baseline score. This ensured that observed changes were attributed to the intervention with the app rather than initial variations in dyspnea.

Finally, a binary logistic regression model was applied, including variables with a P value <.20 in the initial analysis presented or those related to the outcome by biological plausibility. This allowed for estimating the standardized β coefficients, an adjusted coefficient of determination, and the residual values. The binary logistic regression analysis dichotomized the CAT outcome into "low" (patients with a low or medium CAT score) and "high" (patients with a high or very high CAT score). The model was adjusted for possible confounding variables (time of oxygen use and years of cigarette consumption). An α error of ≤0.05 was considered statistically significant, and 95% CIs were calculated.

XSL•FO
RenderX

Ethical Considerations

The clinical trial protocol was approved by the Ethics Committee of the Department of Health of Universidad Santiago de Cali, Colombia, Act No. 02, as well as by the 3 home care institutions in the city of Cali, Colombia. The investigation complies with international regulations, such as the Declaration of Helsinki. All participants gave their written informed consent before inclusion in the study. Informed consent was obtained, and the participants were informed of the possibility of opting out or withdrawing at the time of their choice, if applicable. The data were anonymized so that participants could not be identified, and no amount of compensation was offered or awarded.

Results

Participants were equally and randomly assigned to either the IG or CG using a simple randomization method with Epidat 3.1 (Consellería de Sanidade de la Xunta de Galicia, in collaboration with the Pan American Health Organization [PAHO]). Blinding was not possible for participants or health care professionals owing to the intervention method used (Figure 2). The CONSORT Checklist is included as Multimedia Appendix 3.

Figure 2. Modified CONSORT (Consolidated Standards of Reporting Trials) flow diagram for individual randomized controlled trials of nonpharmacological treatments.

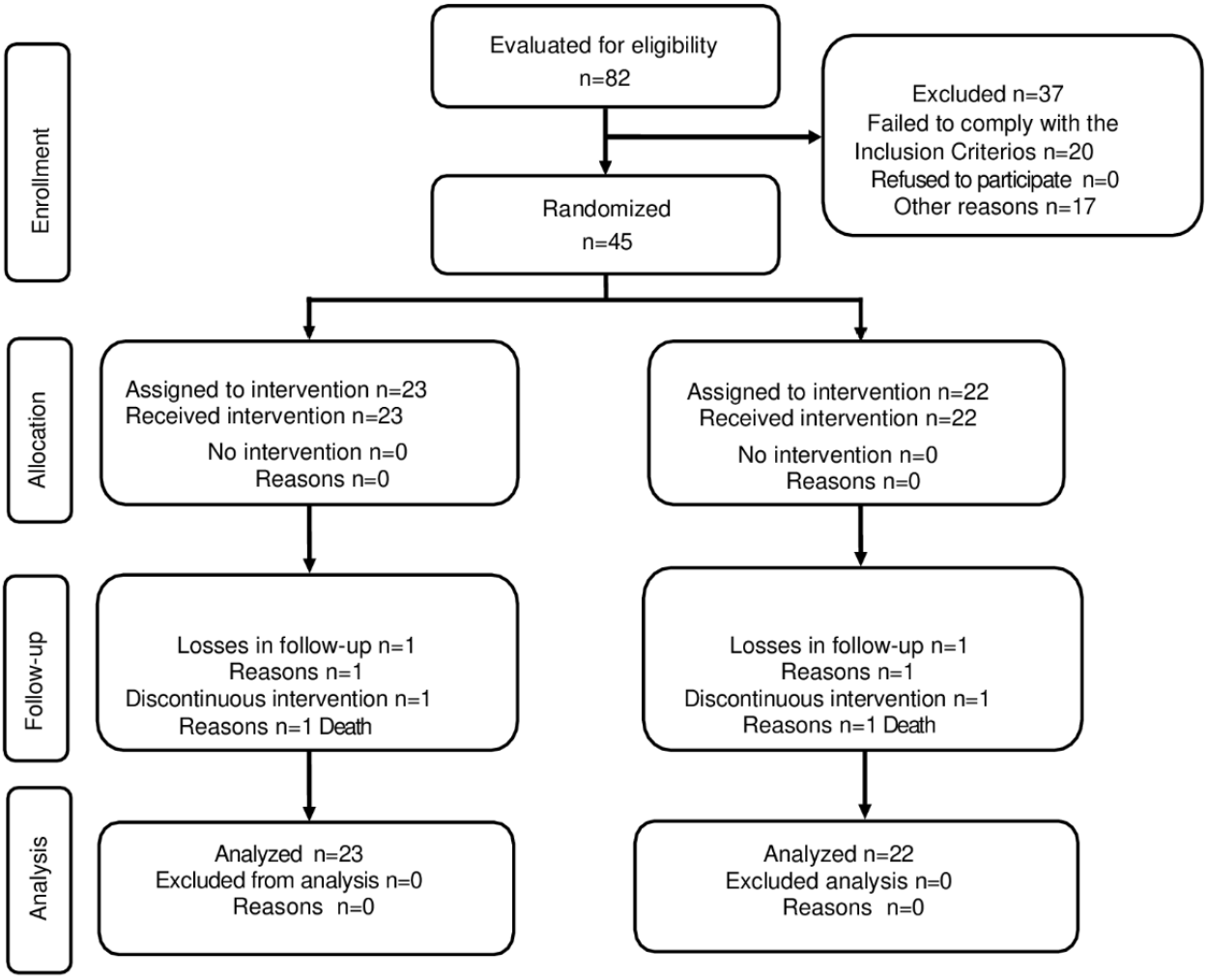

A total of 45 participants were included (23 from the IG and 22 from the CG). The cohort was predominantly female (n=31, 69%) with a mean age of 75.4 (SD 15.9) years (95% CI 70.7-80.0 years). Furthermore, 73% (n=33) of the participants had been exposed to biomass at some point in their lives, and 93% (n=42) had a history of smoking, with a time of consumption that exceeded 15 years in most cases (n=30, 67%). When comparing baseline characteristics between the groups, no statistically significant differences were observed in sociodemographic variables and background except for the "time of oxygen use for more than 15 years," which was greater in the IG (n=17, 74% vs n=9, 41%, *P*=.02; Table 1).

XSL•FO
RenderX

Table 1. Baseline characteristics of patients in the control and intervention groups enrolled in the study (n=45).

Variable	Intervention (n=23)	Control (n=22)	*P* value
Sex, n (%)			.50[a]
Male	8 (35)	6 (27)	
Female	15 (65)	16 (73)	
Education level, n (%)			.50[a]
Primary	18 (78)	16 (73)	
Secondary	5 (22)	5 (23)	
University	0 (0)	1 (4.5)	
AppO2 usage time, n (%)			<.001[a]
More than 15 hours	17 (74)	9 (41)	
Less than 15 hours	6 (26)	13 (59)	
Activities using AppO2, n (%)			.20[a]
Bathing	0 (0)	0 (0)	
Dressing	2 (9)	0 (0)	
Sleeping	7 (30.4)	4 (18)	
Moving around	7 (30.4)	7 (32)	
All of the above	7 (30.4)	11 (50)	
Biomass exposure, n (%)			.90[a]
Yes	17 (74)	16 (73)	
No	6 (26)	6 (27)	
Smoking history, n (%)			.08[b]
Yes	20 (87)	22 (100)	
No	3 (13)	0 (0)	
Years of smoking, n (%)			.30[a]
15 years or less	9 (39)	6 (27)	
15 years or more	14 (61)	16 (73)	
Institution, n (%)			<.001[a]
TodoMed	10 (43.5)	11 (50)	
Amanecer médico	3 (13)	9 (41)	
Cuidarte en casa	10 (43.5)	2 (9)	
Age (years), mean (SD)	72.1 (18.2)	78.9 (13)	.15[c]
Baseline Borg score, mean (SD)	1.2 (0.91)	4.3 (1.4)	<.001[c]
Baseline COPD[d] assessment test score, mean (SD)	26.3 (8.2)	28.9 (6.9)	.28[c]

[a]Chi-square.

[b]Continuity correction.

[c]Mann-Whitney *U* test.

[d]COPD: chronic obstructive pulmonary disease.

Regarding the behavior of vital signs during the 3-month follow-up for patients with conventional follow-up, no significant differences were observed in respiratory rate, heart rate, temperature, saturation, and Borg Dyspnea Scale scores. Likewise, in the IG, no statistically significant differences were found in vital signs, except in the Borg Dyspnea Scale score, which decreased over time (month 1: mean 1.2, SD 0.9; month 2: mean 1.0, SD 0.9; month 3: mean 0.6, SD 0.8; *P*=.01) (Figure 3).

XSL•FO
RenderX

When comparing the clinical variables at the end of follow-up between the groups, it was found that the Borg Dyspnea Scale score was significantly lower in the IG than in the CG (mean 0.6, SD 0.8 vs mean 4.1, SD 1.4; *P*=.001; Table 2).

After adjustment for baseline values, significant differences were maintained in both Borg Dyspnea Scale and CAT final scores, favoring the intervention group, with lower adjusted means (1.3 vs 2.7; *P*=.02 and 22.33 vs 26.69; *P*=.002, respectively; Table 3).

Figure 3. Clinical behavior during the follow-up of patients included in the study (n=45).

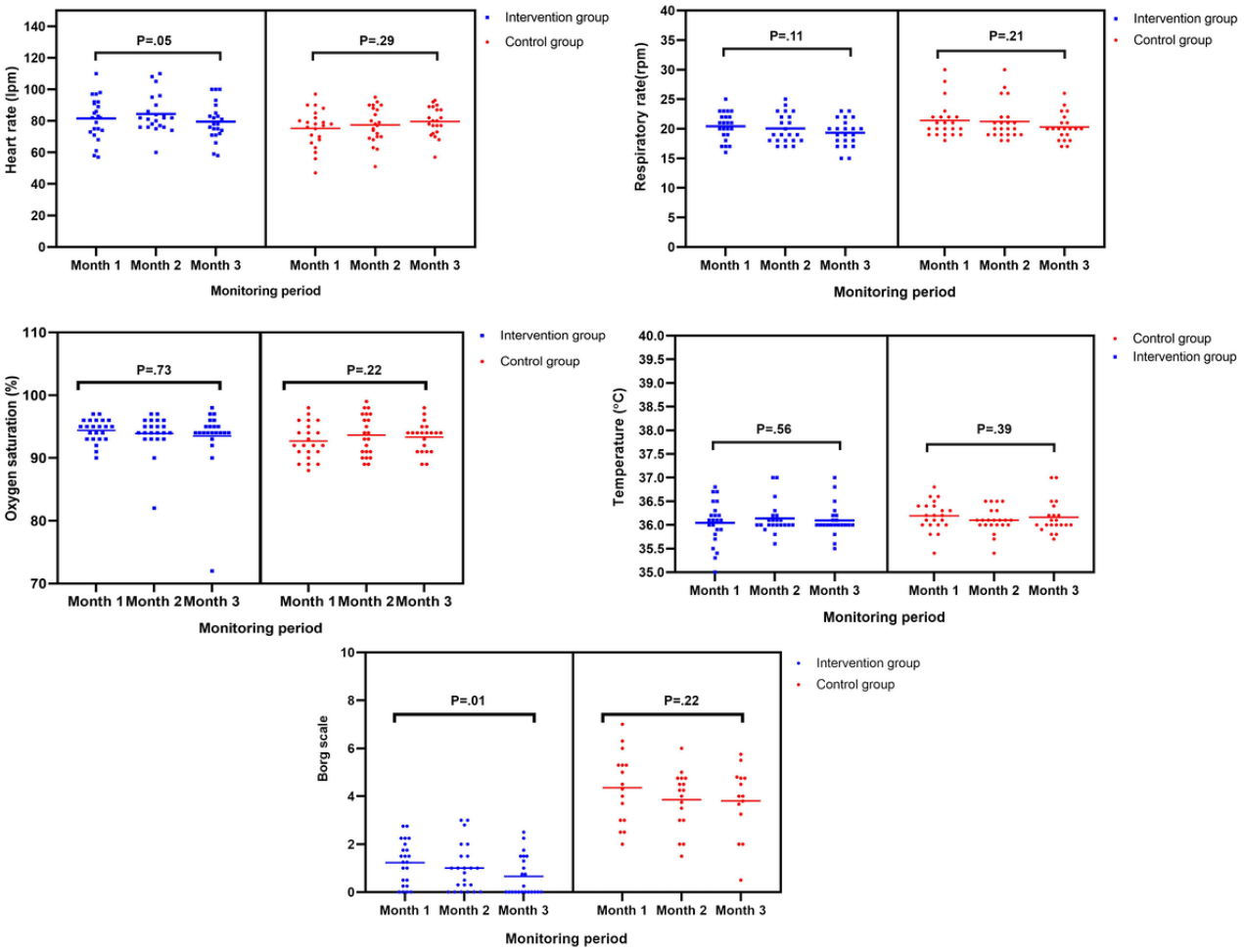

Table 2. Comparison of clinical variables of patients in the intervention and control groups (n=45).

Variable	Intervention (n=23), mean (SD)	Control (n=22), mean (SD)	95% CI	*P* value	Intervention (n=23), mean difference[a] (SD)	Control (n=22), mean difference[a] (SD)	*P* value[b]
Heart rate	79.5 (11.8)	79 (9.1)	–5.6 to 6.6	.87	–3.1 (2.19)	4 (2.80)	.10
Respiratory frequency	19 (2.3)	20 (2.3)	–2.3 to 0.3	.15	–1 (0.71)	–1 (0.71)	.99
Oxygen saturation	93 (5)	93 (2.3)	–2.2 to 2.2	.95	–1 (0.71)	–1 (0.71)	.99
Temperature	36.0 (0.3)	36.1 (0.3)	–027 to 0.07	.30	0 (0)	–0.1 (–0.07)	—[c]
Borg score	0.6 (0.8)	4.1 (1.4)	–4.17 to –2.83	.001	–0.6 (0.42)	–0.8 (0.56)	.72

[a]Mean difference: mean at the end−mean at the beginning of follow-up.

[b]Student *t* test.

[c]Not applicable.

XSL•FO
RenderX

Table 3. Analysis of covariance of the impact of the use of AppO2 on Borg Dyspnea Scale and chronic obstructive pulmonary disease assessment test scores at the end of follow-up, monitoring for the baseline levels of dyspnea.

	Mean (SD)	95% CI	*P* value[a]	Mean (SD)	95% CI	*P* value[b]
Borg Dyspnea Scale[c]			<.001			.02
Intervention group (n=23)	0.65 (0.8)	0.29-1.01		1.3 (1.2)	0.77-1.84	
Control group (n=14)	3.8 (1.4)	2.96-4.64		2.7 (1.6)	1.83-3.60	
COPD[d] Assessment Test (CAT)			.01			.002
Intervention group (n=23)	21.41 (9.01)	17.41-25.41		22.33 (4.27)	20.49-24.17	
Control group (n=14)	27.67 (7.17)	24.4-30.93		26.69 (4.2)	24.80-28.50	

[a]Student *t* test.

[b]Analysis of covariance (ANCOVA).

[c]R^2 adjusted for baseline Borg score: 0.766.

[d]COPD: chronic obstructive pulmonary disease.

Regarding the impact of COPD on quality of life measured with the CAT questionnaire, although no differences were found in the score between the baseline measurement and at the end of follow-up in the control group, a significant decrease was observed in the intervention group (median baseline CAT 27, IQR 23-31 vs median final CAT 22, IQR 14-28; *P*<.001; Figure 4). In addition, the CAT score was significantly higher in patients undergoing conventional follow-up compared with those followed with the mobile app (median 30, IQR 23-32 vs median 22 IQR 14-28; *P*=.02).

When evaluating the record of HO therapy prescriptions in each group, it was found that the proportion of individuals who performed the HO therapy prescriptions correctly was higher in the IG throughout each month of follow-up. The frequency of accessing tutorials or educational records was significantly higher in the IG (95.7% vs 59%; *P*=.003) in the first month of follow-up (Table 4). Furthermore, during the study period, 50% (n=11) of participants in the IG used the mobile app for more than 21 days (median 21, IQR 16-28). The median was greater than 14 days for most variables requiring frequent recording or entry: vital signs recording (median 19, IQR 15-27), Borg scale recording (median 16, IQR 14-22), and oxygen prescription recording (median 14, IQR 11-18; Multimedia Appendix 4).

Satisfaction was mostly observed among users of the AppO2 app, including patients, caregivers, and therapists. Details regarding the measurement of satisfaction with the AppO2 app are presented in Multimedia Appendix 5.

Similarly, in interviews assessing the perception and acceptance of technology using the TAM, the IG reported that the use of the mobile app fostered confidence and improved communication with health care professionals. The group emphasized the value of notifications regarding their vital signs, noting that having access to this information made them more responsible in managing and caring for their disease. Multimedia Appendix 6 presents some expressions reflecting the perceptions and acceptance of AppO2.

Figure 4. Comparison of chronic obstructive pulmonary disease assessment test scores at baseline and at the end of follow-up for control and intervention group patients included in the study. CAT: chronic obstructive pulmonary disease assessment test.

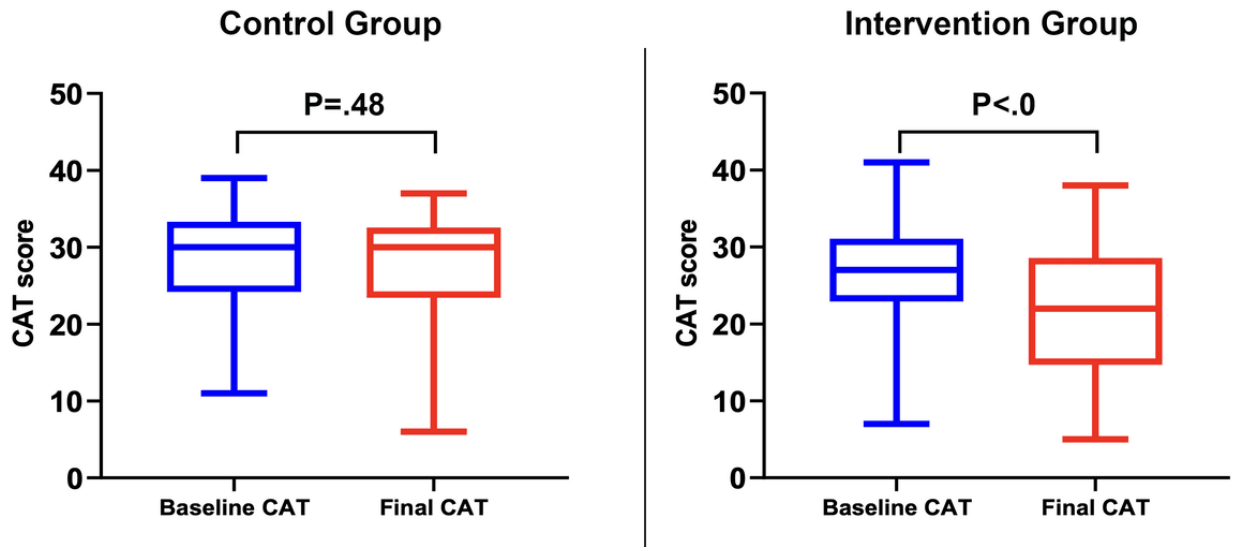

XSL•FO
RenderX

Table 4. Comparison of oxygen therapy recording, education, and oxygen cylinder duration in the control group versus the intervention group.

Variable	Intervention (n=23), n (%)	Control (n=22), n (%)	*P* value
Records oxygen prescription			
Month 1	22 (96)	14 (64)	.006
Month 2	22 (96)	11 (50)	.002
Month 3	22 (96)	10 (48)	.001
Records education or reviews tutorials			
Month 1	22 (96)	13 (59)	.003
Month 2	14 (61)	9 (41)	.18
Month 3	4 (17)	14 (67)	.002
Records duration of oxygen cylinder			
Month 1	6 (26)	2 (9)	.13
Month 2	4 (17)	3 (14)	.72
Month 3	5 (17)	7 (33.)	.44

Finally, the binary logistic regression model was statistically significant (chi-square value=5.9, *P*<.001) and demonstrated a good fit (Hosmer-Lemeshow test=3.92, *P*=.80). The use of the mobile app was the only variable independently associated with the impact on the quality of life of patients with COPD, as measured using the CAT questionnaire (adjusted OR 0.15; 95% CI 0.02-0.82; Table 5).

Table 5. Use of AppO2 and its association with decreased impact of chronic obstructive pulmonary disease on quality of life measured by chronic obstructive pulmonary disease assessment test.

Variable	Crude OR[a] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
App usage				
Yes	0.15 (0.02-0.80)	.02	0.15 (0.02-0.82)	.02
No	6.57 (1.21-35.5)	.02	6.57 (1.21-35.52)	.02
Sex				
Male	0.79 (0.18-3.30)	.75	1.09 (0.18-6.38)	.92
Female	1.25 (0.29-5.20)	.75	0.91 (0.15-5.34)	.92
Oxygen use duration				
More than 15 hours	0.73 (0.17-3.00)	.66	0.58 (0.10-3.28)	.53
Less than 15 hours	1.36 (0.33-5.50)	.66	1.72 (0.30-9.70)	.53
Smoking duration				
More than 15 years	1.6 (0.35-7.20)	.54	—[b]	—
Less than 15 years	0.62 (0.13-2.82)	.54	—	—
Smoking and biomass exposure				
Yes	2.1 (0.51-8.70)	.29	—	—
No	0.46 (0.11-1.90)	.29	—	—

[a]OR: odds ratio.

[b]Not applicable.

Discussion

Overview

A 2-arm randomized clinical trial was conducted to determine the effectiveness and acceptability of the AppO2 mobile app for managing HO therapy in patients with COPD in Cali, Colombia.

The use of health technology tools that offer clinical interventions is emerging as a promising strategy in the health care setting [38,39]. In recent years, self-management interventions supported by smartphone apps have been shown to improve disease management in chronic patients, thus decreasing hospitalizations and emergency room visits and improving quality of life [40].

XSL•FO
RenderX

Quality of Life of Patients With COPD Receiving Home Oxygen When Using AppO2

The main finding of this investigation is that the use of the AppO2 app is associated with a decrease in the score of the impact of COPD on quality of life. This reduction in the CAT score suggests a lower negative impact of the disease on the quality of life of patients and an improvement in their general well-being. We believe that this finding is related to the user-centered design approach of AppO2 [19], which addressed specific needs expressed by patients and their caregivers during the initial stages of the project. Therefore, this app is not only limited to being a clinical monitoring tool, but it also promotes the development of self-care skills and recognition of clinical signs. Wang et al [41], who also used the CAT to assess the impact on patient's quality of life, observed significant improvements in patients who used a mobile app compared with those who received conventional medical care for COPD.

Another important aspect is the improvement in the dyspnea score according to the Borg scale observed in the IG. Participants mentioned that the information provided by AppO2 through images and notifications when determining the degree of dyspnea allowed them to establish a direct relationship with their perception of effort and respiratory difficulty. They also stated that this facilitated communication with therapists through the mobile app and during home visits. These findings are consistent with those reported by Kayyali et al [42], who demonstrated that patients who received clinical information through technological tools, such as mobile apps, could establish correlations with their symptoms and determine when to contact health care professionals.

A Cochrane review [43] suggested that digital tools in health care might improve patients' quality of life, although long-term effects remain unclear. Prolonged use of these digital interventions could improve symptoms such as dyspnea. Our findings align with studies [44,45] that indicate mHealth apps enhance patients' quality of life.

Health Perception and Self-Care of Patients With COPD Receiving Home Oxygen When Interacting With AppO2

Our study found no significant differences between the groups regarding the recording of the clinical signs such as oxygen saturation, respiratory rate, and heart rate. This might be due to the chronic condition [38,40] of the patients, which did not show notable clinical changes during follow-up. However, users in the IG reported that AppO2 improved their health and confidence in recognizing and interpreting vital signs. This suggests that the usefulness of the mobile app extends beyond clinical sign monitoring, positively impacting patients' self-care and their perception of the app's use, as supported by other research on mHealth technologies [42,46,47].

Another point worth mentioning is the interaction of the patients with the oxygen prescription section. The few patients in the IG functionality found it helpful for managing their treatment autonomously by viewing their oxygen prescription at any time, facilitating the use of the HO. Despite previous positive feedback on this feature [4,19], its low usage in this study might warrant its reconsideration or removal. Conversely, the CG only had access to this information during home visits from health care professionals.

Furthermore, patients who used AppO2 mentioned feeling more responsible and confident in managing their disease, emphasizing that access to clinical information aided them in adhering to guidelines for monitoring their clinical signs. This finding aligns with previous investigations [43,47], which indicate that patients with chronic diseases consider the use of digital tools for self-care as a good strategy to access clinical information and optimize disease management.

Acceptability and Connection Time With AppO2

According to the results of the level of acceptability measured using the TAM tool, most patients expressed being completely satisfied with the usefulness and ease of use of AppO2. In addition, patients in the IG expressed feeling safe when using the app. These results are consistent with those of the investigation conducted by Knox et al [16], who designed an app for patients with COPD also using the TAM. They observed that patients demonstrated high motivation and a positive attitude toward using the app, especially in making health-related decisions [16].

On the other hand, while the average use of the app for 21 days over a 3-month period may seem small, it is important to consider that its design does not require continuous daily use to be effective. The app focuses on building patient confidence and promoting self-awareness of their symptoms. Through intermittent but strategic use, patients acquire key tools to monitor and manage their condition more autonomously, potentially contributing to a better quality of life. On the other hand, patients mentioned in interviews that they felt more secure and confident knowing that therapists could access their records at any time, assess their clinical condition, and adjust treatment if necessary. This sense of constant accompaniment helped them to feel "less alone" in the process of their treatment, enhancing their mental well-being.

Likewise, some authors have indicated that certain factors may influence the frequency of AppO2 use. These factors highlight the acquisition of skills and knowledge for self-reporting and managing their disease, which decreases the need to consult tutorials or informative sections. As patients become more familiar with these skills, they experience greater autonomy, leading to more selective use of the app's functions, focusing on those they consider most relevant to their daily routine [48,49].

In this sense, the trend of AppO2 use is reflected in the maximum period of disconnection of 2 weeks observed among the participants, demonstrating a significant engagement with the app. Patients expressed greater confidence and education in recognizing their vital signs. Thus, effective use of AppO2 is not only measured by frequency of access but also by its ability to foster self-management and clinical monitoring.

In previous studies [50-52], it has been documented that adherence to mHealth apps tends to decline over time, especially after an initial phase of intensive use. Perski et al [51] highlight that this decline in user engagement may influence clinical outcomes. However, they also identify factors that may prolong

XSL•FO
RenderX

use, such as personalization and regular reminders. Similarly, Byambasuren et al [52] noted that the "novelty" of mobile apps often wears off after the first few months. However, they suggest that intuitive design [53] and ongoing support from health professionals may improve long-term user retention.

In our study, the median number of app logins was 21 days during the 3-month follow-up period. In addition, we observed that the maximum time offline was 10 days. These data reflect a pattern of use that aligns with the app's design and does not require continuous daily access to be effective. The design's simplicity [53], vital sign notifications, and ease of use have been key factors in maintaining patient engagement.

As mentioned earlier, access to the tutorials in the IG decreased by the third month. The authors attribute this finding to the fact that the content of the videos did not vary over time, which may have led patients to view the education provided as adequate and integrated into their activities. This allowed them to be more selective in choosing the sections that best suited their needs and daily routine. In this sense, it is crucial to explore reinforcement strategies in future studies to ensure continued app use and maximize its clinical impact. This emphasis on continuous improvement in app usage is essential for the long-term success of such interventions [54,55].

Finally, the IG showed a maximum disconnection period of 2 weeks, indicating that the patients continuously interacted with the app. This adherence may be attributed to the high usability of AppO2, achieved through the end-user-centered design methodology applied in this research [4,18,19]. Conversely, the viewing of tutorials decreased in the final month of follow-up in the IG, which could be attributed to patients becoming more adept at self-monitoring and recognizing their vital signs. Similarly, Ding et al [56] reported that the use of the mobile app was more frequent during the first week of launch but decreased over time. Similarly, previous research [48,49] observed that users stop consulting some functions of mobile apps over time. Some factors that may influence this are the acquisition of skills and knowledge for self-registration and management of their disease, which reduces the need to consult tutorials or information sections. The continuous use of the app can also lead to a more selective use of its functions, focusing on those that are essential in their daily routine [48,49].

Principal Results

The findings of this study indicate that the use of the "AppO2" mobile app is associated with a significant improvement in the quality of life for patients with COPD and HO therapy. The high acceptability of the app is evidenced by a notable degree of satisfaction and adherence among users. Furthermore, "AppO2" not only facilitates self-care by enabling patients to manage their treatment more autonomously but also enhances their confidence in self-managing their condition. This contributes to better adaptation to treatment and greater self-care capabilities.

Limitations

The sample size comprised 45 participants. To avoid a lack of statistical power, the study included 3 home care institutions, which helped increase participant representativeness and the generalizability of the results for home health programs.

A limitation of the study is the 3-month duration. Although this period was sufficient to record favorable changes in the CAT score, which measures the impact of COPD on patients' quality of life, as well as their self-care and health perceptions, a longer follow-up could provide insights into the long-term effects.

The ANCOVA results indicated that the use of the app reduced dyspnea, as measured by the Borg scale, at month 3. However, the statistical significance of the initial Borg score (Borg month 1) suggests that the initial levels of dyspnea influenced the final results. The influence of the covariate suggests that other unmeasured variables, such as disease severity or comorbidities, may have affected the results, although this factor was controlled for in the analysis. Similarly, another possible limitation is the Hawthorne effect or observation bias, which occurs when study participants can systematically alter their behavior when they know they are being observed or monitored [50]. This, along with the inability to blind participants, may have influenced participant behavior. Future research should consider more discreet observation methods to minimize these biases.

Conclusions

The use of AppO2 is associated with a better quality of life in patients with COPD receiving HO therapy. The acceptability results for AppO2 show a high degree of satisfaction and adherence to its use. In addition, this mobile app promotes self-care and allows patients to develop confidence in managing their disease.

Acknowledgments

This research has been funded by the Dirección General de Investigaciones of Universidad Santiago de Cali (call 01-2024). This research has also been supported by the Universidad Santiago de Cali (research project 442-621121-3169). We would also like to thank all the patients and their families, as well as the professionals and institutions that participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adaptation of the AppO2 acceptance questionnaire (dimension and item).

XSL•FO
RenderX

[PDF File (Adobe PDF File), 36 KB - jmir_v27i1e65888_app1.pdf]

Multimedia Appendix 2
Questions on the perception of the AppO2 mobile app.
[PDF File (Adobe PDF File), 37 KB - jmir_v27i1e65888_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1170 KB - jmir_v27i1e65888_app3.pdf]

Multimedia Appendix 4
Use of the mobile app by patients in the intervention group (n=23).
[PDF File (Adobe PDF File), 30 KB - jmir_v27i1e65888_app4.pdf]

Multimedia Appendix 5
Satisfaction of patients, caregivers, and therapists using the mobile app for home oxygen therapy monitoring (n=24). (A) Perceived usefulness. (B) Perceived ease of use. Question 1: Using this app helps me complete my tasks more quickly. Question 2: Using this app enhances my performance. Question 3: Using this app increases my productivity. Question 4: Using this app enhances the effectiveness of my work or self-care. Question 5: Using this app makes it easier for me to perform my work or self-care. Question 6: I find this app very useful for my work or self-care. Question 7: Learning to use this app was easy for me. Question 8: I find this app easy to do what I need to do. Question 9: My interaction with this app was clear and understandable. Question 10: I find this app flexible to interact with. Question 11: It will be easy for me to become an expert in using the app. Question 12: I find this app easy to use.
[PNG File , 83 KB - jmir_v27i1e65888_app5.png]

Multimedia Appendix 6
Technology Acceptance Model and AppO2 perception interviews.
[PDF File (Adobe PDF File), 86 KB - jmir_v27i1e65888_app6.pdf]

References

1. Donaldson GC, Seemungal TAR, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]
2. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, et al. Chronic obstructive pulmonary disease: current burden and future projections. Eur Respir J 2006;27(2):397-412 [FREE Full text] [doi: 10.1183/09031936.06.00025805] [Medline: 16452599]
3. Caballero A, Torres-Duque CA, Jaramillo C, Bolívar F, Sanabria F, Osorio P, et al. Prevalence of COPD in five Colombian cities situated at low, medium, and high altitude (PREPOCOL study). Chest 2008;133(2):343-349. [doi: 10.1378/chest.07-1361] [Medline: 17951621]
4. Naranjo-Rojas A, Perula-de-Torres L, Cruz-Mosquera FE, Molina-Recio G. Mobile application for monitoring patients under home oxygen therapy: a protocol for a randomized controlled trial. BMC Fam Pract 2021;22(1):104 [FREE Full text] [doi: 10.1186/s12875-021-01450-8] [Medline: 34039277]
5. Spielmanns M, Boeselt T, Huber S, Kaur Bollinger P, Ulm B, Peckaka-Egli A, et al. Impact of a smartphone application (KAIA COPD app) in combination with activity monitoring as a maintenance program following pulmonary rehabilitation in COPD: the protocol for the AMOPUR Study, an international, multicenter, parallel group, randomized, controlled study. Trials 2020;21(1):636 [FREE Full text] [doi: 10.1186/s13063-020-04538-1] [Medline: 32653025]
6. Mesquita CB, Knaut C, Caram LMDO, Ferrari R, Bazan SGZ, Godoy I, et al. Impact of adherence to long-term oxygen therapy on patients with COPD and exertional hypoxemia followed for one year. J Bras Pneumol 2018;44(5):390-397 [FREE Full text] [doi: 10.1590/S1806-37562017000000019] [Medline: 30517340]
7. Jones I, Kirby A, Ormiston P, Loomba Y, Chan K, Rout J, et al. The needs of patients dying of chronic obstructive pulmonary disease in the community. Fam Pract 2004;21(3):310-313. [doi: 10.1093/fampra/cmh317] [Medline: 15128695]
8. Marcondes VK, Kuwazuru TS, Silva LPCE, Cezare TJ, Franco EAT, Prudente R, et al. Evaluation of the association of adherence to long-term home oxygen therapy and clinical markers and five-year mortality in patients with Chronic obstructive pulmonary disease. J Bras Pneumol 2020;46(6):e20190158 [FREE Full text] [doi: 10.36416/1806-3756/e20190158] [Medline: 32844890]

XSL•FO
RenderX

9. Gauthier A, Bernard S, Bernard E, Simard S, Maltais F, Lacasse Y. Adherence to long-term oxygen therapy in patients with chronic obstructive pulmonary disease. Chron Respir Dis 2019;16:1479972318767724 [FREE Full text] [doi: 10.1177/1479972318767724] [Medline: 29631423]
10. Wedzicha JA, Donaldson GC. Exacerbations of chronic obstructive pulmonary disease. Respir Care 2003;48(12):1204-13; discussion 1213 [FREE Full text] [Medline: 14651761]
11. Casanova C, Cote C, Marin JM, Pinto-Plata V, de Torres JP, Aguirre-Jaíme A, et al. Distance and oxygen desaturation during the 6-min walk test as predictors of long-term mortality in patients with COPD. Chest 2008;134(4):746-752. [doi: 10.1378/chest.08-0520] [Medline: 18625667]
12. Botero Mesa S, Restrepo DA. Conceptos esenciales de la EPOC, prevalencia e impacto en América Latina. Medicina UPB. 2015. URL: https://revistas.upb.edu.co/index.php/medicina/article/view/1379 [accessed 2024-11-20]
13. Niewoehner DE. The impact of severe exacerbations on quality of life and the clinical course of chronic obstructive pulmonary disease. Am J Med 2006;119(10 Suppl 1):38-45. [doi: 10.1016/j.amjmed.2006.08.006] [Medline: 16996898]
14. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012;12(12):CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
15. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;2013(3):CD008776 [FREE Full text] [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
16. Knox L, Gemine R, Rees S, Bowen S, Groom P, Taylor D, et al. Using the technology acceptance model to conceptualise experiences of the usability and acceptability of a self-management app (COPD.Pal®) for chronic obstructive pulmonary disease. Health Technol (Berl) 2021;11(1):111-117 [FREE Full text] [doi: 10.1007/s12553-020-00494-7] [Medline: 33262925]
17. Turchioe MR, Heitkemper EM, Lor M, Burgermaster M, Mamykina L. Designing for engagement with self-monitoring: a user-centered approach with low-income, latino adults with type 2 diabetes. Int J Med Inform 2019;130:103941 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.001] [Medline: 31437618]
18. Molina-Recio G, Molina-Luque R, Romero-Saldaña M. The importance of knowing and listening to all those involved in the design and use of nutrition mobile apps. Getting to know the great GApp. Nutr Hosp 2021;38(3):555-562 [FREE Full text] [doi: 10.20960/nh.03385] [Medline: 33813833]
19. Naranjo-Rojas A, Perula-de-Torres L, Molina-Recio G. Patients, caregivers, and healthcare professionals' needs when designing the content of a mobile application for the clinical monitoring of patients with chronic obstructive pulmonary disease and home oxygen therapy: a user-centered design. Internet Interv 2022;29:100552 [FREE Full text] [doi: 10.1016/j.invent.2022.100552] [Medline: 35756891]
20. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. New York, New York, USA: ACM Press; 1993 Presented at: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; 1993 May 01; New York, NY, United States p. 206-213.
21. Kruis AL, Smidt N, Assendelft WJJ, Gussekloo J, Boland MRS, Rutten-van Mölken M, et al. Integrated disease management interventions for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2013(10):CD009437. [doi: 10.1002/14651858.CD009437.pub2] [Medline: 24108523]
22. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
23. Noseda A. Dyspnoea and perception of airway obstruction. Rev Mal Respir 2003;20(3 Pt 1):364-372. [Medline: 12910111]
24. Salas JJR. Correlation of PaO2/FiO2 and SpO2/FiO2 indices in the postoperative period of cardiac surgery in a cardiovascular postsurgical therapy unit. Critical Medicine 2013;27(2):71-76 [FREE Full text]
25. Kendrick KR, Baxi SC, Smith RM. Usefulness of the modified 0-10 Borg scale in assessing the degree of dyspnea in patients with COPD and asthma. J Emerg Nurs 2000;26(3):216-222. [doi: 10.1016/s0099-1767(00)90093-x] [Medline: 10839848]
26. de Freitas APVM, Belo LF, Martinez L, Hernandes NA, Pitta F. Use of the borg dyspnea scale to identify dynamic hyperinflation during the 6-minute walking test in individuals with moderate-severe COPD: A pilot study. Pulmonology 2023;29(4):335-337 [FREE Full text] [doi: 10.1016/j.pulmoe.2023.02.003] [Medline: 36907815]
27. Jimenez-Ruiz CA, Pascual Lledó JF, Cícero Guerrero A, Cristóbal Fernández M, Mayayo Ulibarri M, Villar Laguna C. Analysis of quality of life in patients with chronic obstructive pulmonary disorder (COPD) who give up smoking. Semergen 2018;44(5):310-315. [doi: 10.1016/j.semerg.2017.08.003] [Medline: 28918179]
28. Rocha V, Jácome C, Martins V, Marques A. Are in person and telephone interviews equivalent modes of administrating the CAT, the FACIT-FS and the SGRQ in people with COPD? Front Rehabil Sci 2021;2:729190 [FREE Full text] [doi: 10.3389/fresc.2021.729190] [Medline: 36188865]
29. Weldam SW, Schuurmans MJ, Liu R, Lammers JJ. Evaluation of quality of life instruments for use in COPD care and research: a systematic review. Int J Nurs Stud 2013;50(5):688-707. [doi: 10.1016/j.ijnurstu.2012.07.017] [Medline: 22921317]

XSL•FO
RenderX

30. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline Leidy N. Development and first validation of the COPD assessment test. Eur Respir J 2009;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
31. Gaveikaite V, Fischer C, Schonenberg H, Pauws S, Kitsiou S, Chouvarda I, et al. Telehealth for patients with chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis protocol. BMJ Open 2018;8(9):e021865 [FREE Full text] [doi: 10.1136/bmjopen-2018-021865] [Medline: 30232108]
32. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
33. Alghamdi SM, Rajah AMA, Aldabayan YS, Aldhahir AM, Alqahtani JS, Alzahrani AA. Chronic obstructive pulmonary disease patients' acceptance in E-Health clinical trials. Int J Environ Res Public Health 2021;18(10):5230 [FREE Full text] [doi: 10.3390/ijerph18105230] [Medline: 34069028]
34. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
35. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
36. Lobos K, Cobo-Rendón R, Guzmán E, Bruna C. Adaptación y validación de dos cuestionarios sobre implementación de la tecnología en la docencia universitaria. Form. Univ 2022;15(5):1-14 [FREE Full text] [doi: 10.4067/S0718-50062022000500001]
37. Bravo LG. Possibilities for using TAM and technology frames models to assess the acceptance of new technologies in the chilean higher education quality assurance. Rev. Electrón. Educ 2015;19(2):181-196.
38. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
39. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
40. Rixon L, Hirani SP, Cartwright M, Beynon M, Doll H, Steventon A, et al. A RCT of telehealth for COPD patient's quality of life: the whole system demonstrator evaluation. Clin Respir J 2017;11(4):459-469. [doi: 10.1111/crj.12359] [Medline: 26260325]
41. Wang L, Guo Y, Wang M, Zhao Y. A mobile health application to support self-management in patients with chronic obstructive pulmonary disease: a randomised controlled trial. Clin Rehabil 2021;35(1):90-101. [doi: 10.1177/0269215520946931] [Medline: 32907384]
42. Kayyali R, Savickas V, Spruit MA, Kaimakamis E, Siva R, Costello RW, et al. Qualitative investigation into a wearable system for chronic obstructive pulmonary disease: the stakeholders' perspective. BMJ Open 2016;6(8):e011657 [FREE Full text] [doi: 10.1136/bmjopen-2016-011657] [Medline: 27580831]
43. Janjua S, Banchoff E, Threapleton CJ, Prigmore S, Fletcher J, Disler RT. Digital interventions for the management of chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2021;4(4):CD013246 [FREE Full text] [doi: 10.1002/14651858.CD013246.pub2] [Medline: 33871065]
44. Dunn J, Coravos A, Fanarjian M, Ginsburg GS, Steinhubl SR. Remote digital health technologies for improving the care of people with respiratory disorders. Lancet Digit Health 2024;6(4):e291-e298 [FREE Full text] [doi: 10.1016/S2589-7500(23)00248-0] [Medline: 38402128]
45. Cox NS, Dal Corso S, Hansen H, McDonald CF, Hill CJ, Zanaboni P, et al. Telerehabilitation for chronic respiratory disease. Cochrane Database Syst Rev 2021;1(1):CD013040 [FREE Full text] [doi: 10.1002/14651858.CD013040.pub2] [Medline: 33511633]
46. Middlemass JB, Vos J, Siriwardena AN. Perceptions on use of home telemonitoring in patients with long term conditions - concordance with the health information technology acceptance model: a qualitative collective case study. BMC Med Inform Decis Mak 2017;17(1):89 [FREE Full text] [doi: 10.1186/s12911-017-0486-5] [Medline: 28651588]
47. Lundell S, Holmner, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015;109(1):11-26 [FREE Full text] [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
48. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021;9(12):e29098 [FREE Full text] [doi: 10.2196/29098] [Medline: 34927597]
49. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
50. Donevant SB, Messias DKH, Estrada RD. Utilization of mobile applications in collaborative patient-provider monitoring of chronic health conditions: an examination of three theoretical frameworks to guide practice. J Inform Nurs 2018;3(2):6-11 [FREE Full text] [Medline: 31179417]

XSL•FO
RenderX

51. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
52. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
53. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
54. de Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
55. Tsertsidis A, Kolkowska E, Hedström K. Factors influencing seniors' acceptance of technology for ageing in place in the post-implementation stage: a literature review. Int J Med Inform 2019;129:324-333 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.06.027] [Medline: 31445274]
56. Ding H, Fatehi F, Maiorana A, Bashi N, Hu W, Edwards I. Digital health for COPD care: the current state of play. J Thorac Dis 2019;11(Suppl 17):S2210-S2220 [FREE Full text] [doi: 10.21037/jtd.2019.10.17] [Medline: 31737348]

Abbreviations

ANCOVA: analysis of covariance
CAT: chronic obstructive pulmonary disease assessment test
CG: control group
COPD: chronic obstructive pulmonary disease
HO: home oxygen
IG: intervention group
mHealth: mobile health
OR: odds ratio
TAM: technology acceptance model

Edited by YH Lin; submitted 28.08.24; peer-reviewed by B Chen, PJ Honkoop; comments to author 26.09.24; revised version received 14.10.24; accepted 14.10.24; published 06.01.25.

Please cite as:
Naranjo-Rojas A, Perula-de Torres LÁ, Cruz-Mosquera FE, Molina-Recio G
Efficacy and Acceptability of a Mobile App for Monitoring the Clinical Status of Patients With Chronic Obstructive Pulmonary Disease Receiving Home Oxygen Therapy: Randomized Controlled Trial
J Med Internet Res 2025;27:e65888
URL: https://www.jmir.org/2025/1/e65888
doi:10.2196/65888
PMID:

©Anisbed Naranjo-Rojas, Luis Ángel Perula-de Torres, Freiser Eccehomo Cruz-Mosquera, Guillermo Molina-Recio. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Frequency Customized Relieving Sound for Chronic Subjective Tinnitus Management: Prospective Controlled Study

Dongmei Tang[1,2,3*], PhD; Yuzhu Peng[1,2*], MBBS; Dantong Gu[4*], MS; Yongzhen Wu[1*], PhD; Huawei Li[1,2*], PhD

[1]ENT Institute and Department of Otorhinolaryngology, Eye & ENT Hospital, Fudan University, Shanghai, China

[2]NHC Key Laboratory of Hearing Medicine Research, Shanghai, China

[3]Shanghai Zehnit Medical Technology Co Ltd, Shanghai, China

[4]Clinical Research Unit of Eye & ENT Hospital, Fudan University, Shanghai, China

[*]all authors contributed equally

Corresponding Author:
Dongmei Tang, PhD
ENT Institute and Department of Otorhinolaryngology
Eye & ENT Hospital
Fudan University
83 Fenyang Road
Shanghai, 200031
China
Phone: 86 13023299189
Email: tang.dongm@163.com

Abstract

Background: Tinnitus is a major health issue, but currently no tinnitus elimination treatments exist for chronic subjective tinnitus. Acoustic therapy, especially personalized acoustic therapy, plays an increasingly important role in tinnitus treatment. With the application of smartphones, personalized acoustic stimulation combined with smartphone apps will be more conducive to the individualized treatment and management of patients with tinnitus.

Objective: The aim of this study was to evaluate the efficacy of a new personalized approach known as the digital frequency customized relieving sound (DFCRS) for tinnitus treatment and to explore the factors that may influence its therapeutic effect.

Methods: Patients with subjective tinnitus were enrolled in this study from July 14, 2020, to May 24, 2021, in the tinnitus specialist clinic of Eye and ENT Hospital, Fudan University, Shanghai, China. In this nonrandomized concurrent controlled trial, a total of 107 participants were assigned to listen to personalized DFCRS through our developed app, while the other 77 participants who did not want to download and use the app were assigned to listen to unmodified music (UM). All the recruits were instructed to listen to DFCRS or UM for at least 2 hours a day and complete follow-up assessments at baseline, 1, 2, and 3 months. Multidimensional assessment scales, that is, Tinnitus Handicapped Inventory (THI), Hospital Anxiety and Depression Scale (HADS), Athens Insomnia Scale (AIS), Fear of Tinnitus Questionnaire (FTQ), and Tinnitus Catastrophizing Scale (TCS) were used to evaluate the severity of tinnitus and the quality of life. Linear mixed models were used to test for changes in the THI scores across 3 months of acoustic treatment between group (DFCRS or UM treatment) and time. A multiclass logistic model was built with a stepwise function to determine the influence of the different covariates on the effects of acoustic treatment.

Results: The results of the multidimensional assessment scales after 3 months of treatment showed that DFCRS-treated patients had significant tinnitus relief compared to those in the UM group. Linear mixed models revealed a significant reduction in the THI scores over time (P<.001), with the DFCRS group showing significantly greater improvement than the UM group (P<.001). At 3 months, 92.5% (99/107) of the patients undergoing DFCRS reported tinnitus relief or disappearance, and longer daily treatment time was associated with better outcomes (P=.007). Multiclass logistic regression confirmed that longer treatment time (odds ratio [OR] 13.07-64.78; P<.001) and more severe tinnitus at baseline (OR 10.46-83.71; P<.001) predicted better treatment response. All secondary outcomes (HADS, AIS, FTQ, TCS) showed significant improvements over time (P<.001).

Conclusions: Our study suggests that DFCRS is a new promising and noninvasive therapy for chronic tinnitus, and it can be delivered through a mobile app to bring more convenience to patients with tinnitus.

(J Med Internet Res 2025;27:e60150) doi:10.2196/60150

XSL•FO
RenderX

KEYWORDS

tinnitus; digital frequency customized relieving sound; unmodified music; sound therapy; prospective study; mobile phone

Introduction

Tinnitus, a common concern in the otolaryngology clinic, is described as the perception of sound without a corresponding external stimulus [1]. People can experience tinnitus constantly or intermittently from a single ear, both ears, and within the head. The presence of tinnitus can be annoying, affecting the patients' overall quality of life (QoL) and even negatively impacting their mental health [2]. Epidemiological studies have shown that tinnitus affects approximately 10%-25% of the adults [3,4], and tinnitus-induced anxiety, irritability, depression, sleep disturbances, and concentration disorders seriously threaten the cognitive and neurological health of approximately 1% of the adults [5]. Therefore, the search for an efficient treatment for tinnitus is urgently needed.

To date, there is a lack of approved medications that can eliminate tinnitus or provide a completely satisfactory cure [6,7]. However, there are several intervention strategies that can help patients cope with tinnitus. In recent decades, sound therapy has been broadly applied in exploring clinical treatments for tinnitus [8-12]. According to the sound generation method, sound therapy can be divided into 2 main types: unmodulated and modulated sound therapy [13]. Tinnitus masking and tinnitus retraining therapy are the 2 most commonly used unmodulated strategies, and their ultimate goal is to help patients with tinnitus become accustomed to tinnitus, attenuating excessive attention to it and thus reducing the tinnitus-related negative emotions [14-17]. Although some findings have confirmed their effectiveness in improving the QoL of patients with tinnitus [18], a randomized clinical trial showed no significant difference in the outcomes between tinnitus retraining therapy and placebo-sound treatment control groups [19]. Additionally, the current unmodulated sound therapy does not appear to change the loudness of tinnitus, limiting the possibility of tinnitus elimination [12].

Conversely, customized sound therapy is designed to encourage the reorganization of abnormal neural connections or to inhibit pathological neural activity, which is more promising in effectively eliminating tinnitus [10,20]. Current research has confirmed the superior effectiveness of customized sound therapy over unmodulated therapy [21]. For example, neuromonics tinnitus treatment involves adding or excluding individual broadband frequency sounds to relaxing music to produce intermittent tinnitus perception and facilitate desensitization to tinnitus [22-24]. It is worth noting that tailor-made notched music training [10] and tinnitus pitch-matched therapy [11] used 2 opposite sound-processing strategies; however, both produced positive results [25-27]. Thus, the underlying mechanisms of tinnitus alleviation, which might be linked to the pathogenesis of tinnitus, should be elucidated in the future. Although various sound therapies differ in strategies and effectiveness, there is a consensus that the effects of sound therapy are not immediate and need to be sustained over a period of time. We reviewed the follow-up time of different acoustic treatments [28]. For short ones such as the Heidelberg music therapy, the follow-up time can be 5 days, 1 week, or 3 weeks. In interventions such as tailor-made notched music, pitch-matched therapy, and tinnitus masking, the follow-up time can be 1 month to 3 months. Tinnitus retraining therapy, neuromonics tinnitus treatment, acoustic coordinated reset neuromodulation therapy, tailor-made notched music, and many other studies have a long follow-up time—more than 6 months or even 12 months, 18 months, 36 months, and so on.

To date, the mechanism of subjective tinnitus has not been fully elucidated. Generally, peripheral hearing loss may be the trigger of subjective tinnitus [12,29]. Due to auditory deprivation, the activity of the affected peripheral auditory nerves is reduced, and neuroplasticity directs adaptive changes in the central auditory system when the brain loses input from the ear [30,31]. Based on this, we designed a new personalized tinnitus relief model known as the digital frequency customized relieving sound (DFCRS), which generates 2 surrounding narrow bands (each 1/3 octave) according to the individual's matched tinnitus frequency and then dynamically enhances the intensity of the 2 designated bands by 10 dB on soft and relieving music [32]. With the popularity of mobile smart devices and the improved stability and convenience of mobile apps, which are increasingly being used in the sound therapy of patients with tinnitus [33,34], we developed an app named the Fudan Tinnitus Relieving System (FTRS) [35,36] to run this model for monitoring, management, and follow-up evaluation of tinnitus treatment. In this study, we aimed to investigate the effectiveness of DFCRS by using a self-designed FTRS app compared to that of unmodified music (UM) for treating tinnitus. Furthermore, we explored the possible factors that might influence the treatment effectiveness of DFCRS.

Methods

Study Design

Patients with subjective tinnitus were enrolled from the tinnitus specialist clinic of Eye & ENT Hospital of Fudan University from July 14, 2020, to May 24, 2021. All the included participants were Mandarin speakers, aged over 18 years, and able to communicate with instructors without any hearing aids. The average pure tone threshold (0.5, 1, 2, 4 kHz) should be no more than 55 dB HL of the worse ear. The participants were required to understand the purpose and process of the study, complete at least 3 follow-up visits in the clinic, and sign the informed consent prior to the commencement of the experiment. Patients with pulsatile tinnitus, objective tinnitus, or clear primary causes such as otosclerosis, Meniere disease, and acute stage of sudden sensorineural hearing loss were excluded from this study. Moreover, for safety reasons, patients with severe health issues such as cardiovascular and cerebrovascular diseases or tumors were excluded from this study. This was a nonrandomized concurrent controlled trial, and we followed the TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) guidelines for nonrandomized

XSL•FO
RenderX

controlled trials (Multimedia Appendix 1). A total of 107 eligible patients who tended to use the FTRS app for acoustic therapy received DFCRS, and they were labeled as the experimental group. The other 77 patients who refused to use the app received UM, serving as the control group. The experimental group and the control group were recruited and provided acoustic therapy at the same time. All participants completed 4-timepoint follow-up questionnaires.

Ethics Approval

All tests were approved by the institutional review board of Eye & ENT Hospital of Fudan University (2017048). All patients signed the informed consent forms, and sensitive information relevant to patient privacy was removed before the analysis. Both forms of acoustic therapy did not add any cost to the patient, and each patient could also receive an appointment or a plus number service for the tinnitus specialist clinic.

Patient Assessment

The baseline data were collected mainly from 4 aspects: general information, characteristics of tinnitus, audiological and tinnitus assessments, and results of the multidimensional assessment scales of tinnitus. General information included details on age, gender, medical history, and treatment history. The characteristics of tinnitus included information on position, course, pitch, and correlative factors of tinnitus. Audiological and tinnitus tests included pure-tone audiometry (125 Hz-8 kHz), acoustic immittance, and tinnitus pitch matching. Five standardized questionnaires in the Chinese version were used to evaluate the severity of tinnitus and the QoL of patients with tinnitus from multiple dimensions.

Tinnitus Handicap Inventory

The Tinnitus Handicap Inventory (THI) is a self-report measure first developed in 1996 [37]. It contains 25 items with 0, 2, and 4 points each to quantify the functional, emotional, and catastrophic feedback to tinnitus. The total number of points is up to 100, with a higher score indicating more severe tinnitus-related discomfort and disability.

Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) consists of 14 items, 7 of which belong to the subscale for anxiety (HADS-A) and the other 7 questions indicate the subscale for depression (HADS-D) [38]. Due to the close correlation between tinnitus severity and psychiatric disorders, HADS was introduced to patients with tinnitus to identify emotional disturbances that reflect therapeutic efficacy [39].

Athens Insomnia Scale

The Athens Insomnia Scale (AIS) is commonly used to measure sleep quality [40]. It includes 8 items with a total score ranging from 0 (no sleep disorder at all) to 24 (the most severe insomnia). Changes in sleep disorders—one of the most common concomitant symptoms of tinnitus—can be used to evaluate the effects of tinnitus treatment.

Fear of Tinnitus Questionnaire

The Fear of Tinnitus Questionnaire (FTQ) is a self-report measure containing 17 items, with 1 point each, which intends to assess the levels of fear of tinnitus [41]. Higher scores indicate a greater fear of tinnitus. Researchers can estimate tinnitus status based on patients' attitudes and moods toward tinnitus.

Tinnitus Catastrophizing Scale

The Tinnitus Catastrophizing Scale (TCS) is a 13-item measure developed to assess how catastrophizing respondents are regarding tinnitus [42]. The instrument is rated on a 5-point scale from 0 to 4, with higher scores indicating higher states of tinnitus-specific distress.

Tinnitus Severity Assessment

The Tinnitus Severity Assessment scale divides patients with tinnitus into 4 grades according to subjective tinnitus-related impairment in daily life [2]. Patients with grade I or no impairment are not annoyed by tinnitus, whereas patients who are unable to work or live normally will be confirmed as having grade IV severe impairment.

Acoustic Therapy Procedure

In the DFCRS-treated group, according to precise audiological and tinnitus tests, a series of 4 endpoints of the modulation range adjacent to the tinnitus pitch were calculated using our specific algorithm (Multimedia Appendix 2). Based on these 4 endpoints, 2 specific frequency bands of 1/3 octave intervals below and above the tinnitus pitch were confirmed. Thereafter, with the assistance of the FTRS app, soft and relieving music was processed by dynamically increasing the intensity of both specified bands with a 10-dB gain, thus outputting the personalized relieving sound for tinnitus. Accordingly, in the UM-treated group, the unmodified original music was used.

After baseline data completion at the initial point in time (T0), patients in the DFCRS-treated group were trained to use the FTRS app correctly, while others in the UM-treated group were instructed to listen to UM through their favorite players. All participants were required to listen to the modified or unmodified sound in a quiet environment with a comfortable volume to avoid possible hearing damage.

In this study, the participants were not allowed to receive any other simultaneous treatments that might interfere with sound therapy. To obtain a better effect, all patients were encouraged to be exposed for as much time as possible in the therapeutic sound environment, and the recommended duration was at least 2 h/day. In addition, the participants were instructed to record the time of daily sound exposure accurately and truthfully. Finally, researchers maintained contact with the participants through various accessible means such as phone calls, WeChat, emails, and tinnitus specialist clinic visits.

Tinnitus Counseling and Education

Well-trained otologists implemented counseling and education for all patients with tinnitus. The main topics included the etiology and basic pathogenesis of tinnitus, the relationship between tinnitus and hearing loss, the interpretation of audiological and tinnitus test results, the psychoacoustic characteristics of tinnitus, the relationship between tinnitus and attention, guidance on exploring common methods to divert attention, and relaxation training. During the entire procedure,

XSL•FO
RenderX

the participants' confusion and discomfort with regard to tinnitus were patiently, promptly, and sufficiently dispelled.

Outcome Measures

There were 4 timepoints in total, including the baseline when the participants were enrolled and at 1, 2, and 3 months after commencing treatment. The primary outcome was the THI scores from baseline to follow-up endpoints. In addition, the other assessment instruments (HADS, AIS, TCS, and FTQ) as secondary outcomes also provided longitudinal insights into the efficacy of tinnitus treatment.

Statistical Analysis

Statistical analyses were performed using R software (version 4.1.2; R Foundation for Statistical Computing) with the lme4, nnet, and ggpubr packages. Two-sided values of $P<.05$ were deemed statistically significant. Demographics and baseline characteristics such as age, sex, tinnitus course, location, tinnitus pitch, severity, and assessment instruments (THI, HADS-A, HADS-D, AIS, FTQ, and TCS) were summarized by groups that were defined by the treatment results. We summarized the descriptive statistics of the continuous variables (mean, standard deviation or median, and quartile) and categorical variables (number and proportion of participants in each category). Differences in the characteristics among the groups were tested with chi-square for categorical variables; 1-way analysis of variance was used for normally distributed continuous variables, and the Kruskal–Wallis test was used for nonnormal continuous variables. A 2-tailed *t* test was applied to compare the mean multidimensional assessment data before and after treatment and to determine the differences between subgroups. Linear mixed models were used to test for changes in the THI scores across 3 months of acoustic treatment between groups (DFCRS or UM treatment) and time. A multiclass logistic model was built with a stepwise function to determine the influence of different covariates on the effect of acoustic treatment, and Akaike information criterion/Bayesian information criterion/corrected Akaike information criterion was used to select the best model.

Results

Demographic and Tinnitus Characteristics of the Participants at Baseline

A total of 184 patients with tinnitus (107 in the DFCRS treatment group and 77 in the UM control group) completed the baseline data collection and 3 follow-up visits. After the 3-month treatment, patients were asked to report the effectiveness of tinnitus treatment based on their subjective perceptions. The baseline data are shown in Table 1. In the DFCRS treatment group, 50.5% (54/107) were men and 49.5% (53/107) were women, aged 45 (range 35.5-57) years, with a tinnitus duration averaged at 12 (range 4-24) months. In the UM group, 55% (42/77) were males, the age was 50 (range 37-59) years, and the tinnitus duration was 8 (range 3-42) months. The patients' gender and age did not differ significantly between the experimental group (DFCRS) and the control group (UM). We also compared other baseline data and observed no significant differences in the distribution of tinnitus characteristics involving location, frequency, and severity assessed by multidimensional tinnitus scales (HADS, AIS, Visual Analog Scale, FTQ, and TCS) between the 2 groups. Besides, there were no significant differences in the hearing levels in either ears between the 2 groups. However, the THI scores differed at baseline between the 2 groups (UM group: median 40, range 24-60; DFCRS group: median 48, range 33-66; $P=.01$). The overall baseline data were comparable between the DFCRS group and the UM group, suggesting that the following therapeutic effect was not significantly influenced by pre-experimental demographic information and characteristics of tinnitus.

XSL•FO
RenderX

Table 1. Demographic and tinnitus characteristics of the participants in the digital frequency customized relieving sound–treated group and the unmodified music control group at baseline.

	Unmodified music (n=77)	Digital frequency customized relieving sound (n=107)	*P* value
Gender, n (%)			.69
Male	42 (55)	54 (51)	
Female	35 (45)	53 (50)	
Age (years), median (IQR)	50 (37-59)	45 (35.50-57)	.39
Tinnitus course (months), median (IQR)	8 (3-42)	12 (3-24)	.88
Location of tinnitus, n (%)			.31
Unilateral	37 (48)	40 (37)	
In head	3 (4)	7 (7)	
Bilateral	37 (48)	60 (56)	
Tinnitus Handicapped Inventory, median (IQR)	40 (24-60)	48 (33-66)	.01
Hospital Anxiety and Depression Scale-subscale anxiety, median (IQR)	5 (2-7)	5 (2-8)	.40
Hospital Anxiety and Depression Scale-subscale depression, median (IQR)	4 (1-7)	4 (2-7)	.18
Athens Insomnia Scale, median (IQR)	6 (4-9)	7 (4-10)	.11
Visual Analog Scale, median (IQR)	5 (3-6)	5 (4-6)	.05
Fear of Tinnitus Questionnaire, median (IQR)	6 (4-10)	7 (5-10)	.26
Tinnitus Catastrophizing Scale, mean (SD)	22.97 (10.42)	24.38 (9.89)	.36
Left ear pitch, median (IQR)	3642.50 (1933.75-5780)	7175 (6350-8000)	.07
Right ear pitch, median (IQR)	6350 (4435-6350)	6350 (4260-8000)	.47
Pure tone average-right[a], median (IQR)	16.67 (15-21.67)	16.67 (15-20.42)	.55
Pure tone average-left[a], median (IQR)	16.67 (15-23.33)	17.50 (15-23.75)	.88

[a]Average air conducted pure tone hearing thresholds of 500, 1000, 2000, and 4000 Hz.

Mixed Effect Model Indicated Effectiveness of DFCRS Treatment

As the primary outcome, the THI scores were calculated using the linear mixed model in both DFCRS and UM treatment groups. The results of the linear mixed model revealed significant main effects of follow-up time (estimate=–0.119; *P*<.001) (Table 2), which showed a general decrease in the THI scores over the trial duration. Besides, there were significant effects in the DFCRS group compared to those in the control group, and the negative value revealed a remarkable treatment effectiveness of DFCRS measured by THI compared to that of UM in the control group (estimate=–16.6469; *P*<.001).

Apart from the follow-up time and treatment procedure, no other factors were found to contribute to a significant effect in the linear mixed model. Comparisons of the multidimensional assessment scales between the DFCRS-treated group and the UM control group at 4 timepoints are shown in Figure 1. During the 3 months of treatment, almost all of the scores on the 6 scales in the DFCRS-treated group showed significant decline compared to those in the UM control group, indicating better effectiveness of DFCRS in managing tinnitus. We found that UM had no significant therapeutic effect on tinnitus (Table S1 in Multimedia Appendix 3). Moreover, the differences were mostly maintained or were even enlarged across time especially after 1 month of treatment. No adverse effects were reported during follow-up in both groups.

XSL•FO
RenderX

Table 2. Linear mixed effects on tinnitus relief after digital frequency customized relieving sound or unmodified music treatment.

	Estimate	SE	*t* test *(df)*	*P* value
(Intercept)	56.13335	6.032316	9.305439 (509)	<.001
Follow-up time	–0.119	0.015712	–7.57409 (509)	<.001
Group				
Control	Reference	Reference	Reference	Reference
Digital frequency customized relieving sound	–16.6469	3.114957	–5.34417 (173)	<.001
Gender				
Male	Reference	Reference	Reference	Reference
Female	3.57652	3.087018	1.158567 (173)	.25
Age	–0.10198	0.114023	–0.89438 (173)	.37
Location				
Unilateral	Reference	Reference	Reference	Reference
In the brain	–9.39239	6.990866	–1.34352 (173)	.18
Bilateral	3.73601	3.248982	1.149901 (173)	.25
Course of tinnitus	–0.00097	0.03535	–0.02749 (173)	.98

Figure 1. Comparison of the multidimensional assessment scales between the digital frequency customized relieving sound treatment group and the unmodified music control group at 4 timepoints. Data are expressed as median (IQR). AIS: Athens Insomnia Scale; DFCRS: digital frequency customized relieving sound; FTQ: Fear of Tinnitus Questionnaire; HADS-A: Hospital Anxiety and Depression Scale-subscale anxiety; HADS-D: Hospital Anxiety and Depression Scale-subscale depression; TCS: Tinnitus Catastrophizing Scale; THI: Tinnitus Handicapped Inventory; UM: unmodified music.

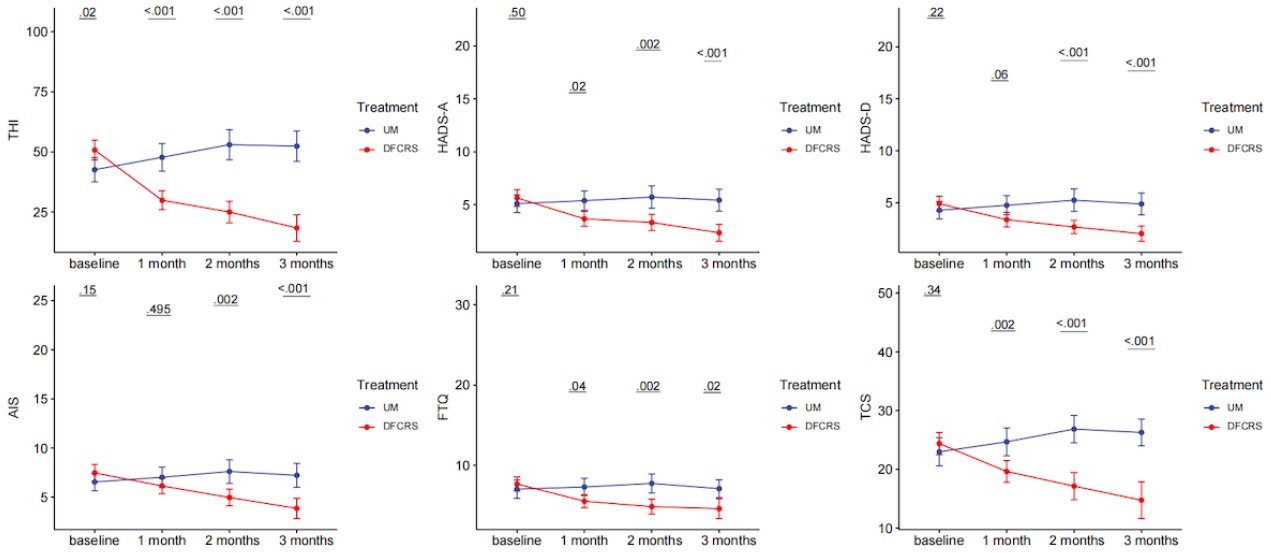

Multiclass Logistic Model Predicted Factors That Affect the Effectiveness of DFCRS Treatment for Tinnitus

To further identify the factors affecting the effectiveness of DFCRS, we analyzed the statistical relationships among gender, age, acoustic treatment time per day, various tinnitus characteristics, and treatment results according to patients' subjective reports. After 3 months of DFCRS treatment, 86.9% (93/107) of the patients reported tinnitus relief, 5.6% (6/107) reported that tinnitus disappeared, and 7.5% (8/107) reported worsened tinnitus (Table S2 in Multimedia Appendix 3). The patients' gender and age did not differ significantly among the 3 different treatment outcome groups. There were no significant differences in the distribution of characteristic tinnitus indicators such as location, tone, course, and severity assessed by multidimensional tinnitus scales among the 3 subgroups as well. Interestingly, the data showed that there was a significant difference in the treatment time per day among the 3 groups (*P*=.007). Participants whose tinnitus was alleviated or disappeared received sound therapy for a longer time than those who reported worsening tinnitus, indicating a time-dose–dependent relationship for implementing DFCRS as a tinnitus treatment strategy (Table S2 in Multimedia Appendix 3). We then conducted a two-by-two comparison among the 3 subgroups, and the results revealed that significant differences existed between the worsening tinnitus group and the tinnitus relief group (*P*<.01) as well as the worsening tinnitus group and

XSL•FO
RenderX

the tinnitus disappeared group (*P*<.05; Figure 2). This finding preliminarily confirmed that an increase in the exposure of DFCRS treatment contributed to better efficacy in managing tinnitus.

For more precise analysis, we built a multiclass logistic model to determine the influence of different covariates on acoustic treatment effects (Table 3). The results showed that compared to patients whose tinnitus worsened, patients who reported tinnitus relief had a longer treatment time per day (odds ratio [OR] 13.07, 95% CI 4.19-40.78; *P*<.001), and patients who reported tinnitus disappearance had the longest daily treatment time (OR 64.78, 95% CI 11.63-360.77; *P*<.001). In addition, patients with more severe tinnitus at baseline tended to respond better to acoustic treatment. More specifically, patients with severity III were more likely to have tinnitus relief (OR 40.37, 95% CI 3.42-477.14; *P*<.001) than patients whose tinnitus severity level was I. Similarly, patients with severity IV were more likely to have tinnitus relief (OR 10.46, 95% CI 2.12-51.62; *P*<.001) or tinnitus disappearance (OR 83.71, 95% CI 5.9-1188.28; *P*<.001) than those with tinnitus severity I.

Figure 2. Differences in the treatment time of digital frequency customized relieving sound in the 3 treatment outcome groups according to subjective reports. Data are presented as median (IQR).

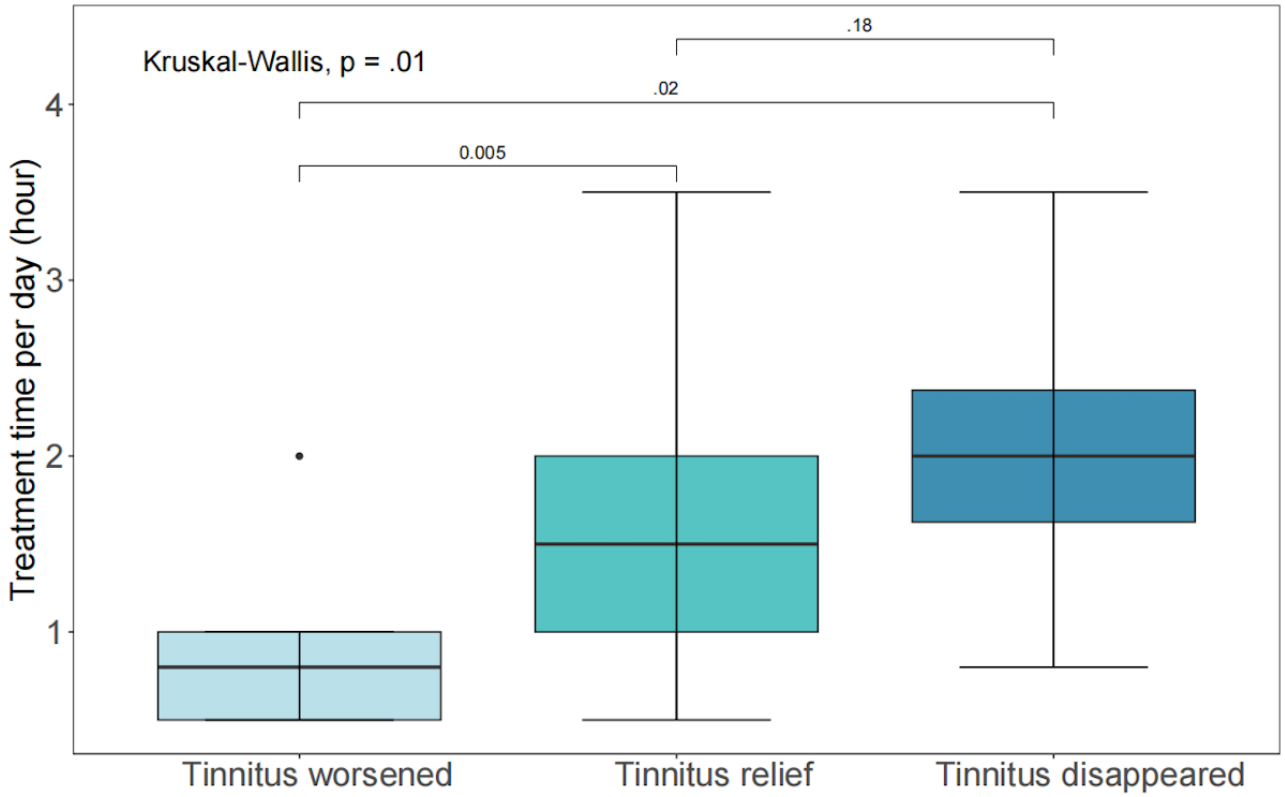

XSL•FO
RenderX

Table 3. Multiclass logistic model results of the influence of different covariates on the customized acoustic treatment effect.

	Odds ratio (2.50% CI-97.50% CI)	*P* value
Worsening	Reference	Reference
Relief from tinnitus		
(Intercept)	258.01 (25.56-2604.61)	<.001
Treatment time	13.07 (4.19-40.78)	<.001
Gender		
Male	Reference	Reference
Female	1.05 (0.33-3.31)	.94
Age	0.96 (0.92-1.01)	.08
Tinnitus course	1.01 (0.99-1.02)	.37
Status		
Continued	Reference	Reference
Intermittent	1.42 (0.09-22.24)	.80
Severity		
Ⅰ	Reference	Reference
Ⅱ	3.1 (0.75-12.9)	.12
Ⅲ	40.37 (3.42-477.14)	<.001
Ⅳ	10.46 (2.12-51.62)	<.001
Disappearing tinnitus		
(Intercept)	0.55 (0.01-49.61)	.79
Treatment time	64.78 (11.63-360.77)	<.001
Gender		
Male	Reference	Reference
Female	0.28 (0.03-2.42)	.25
Age	0.94 (0.85-1.05)	.30
Tinnitus course	0.74 (0.5-1.09)	.13
Status		
Continued	Reference	Reference
Intermittent	221.23 (4.75-10,304.32)	.01
Severity		
Ⅰ	Reference	Reference
Ⅱ	7.11 (0.38-133.18)	.19
Ⅲ	20.42 (0.7-591.93)	.08
Ⅳ	83.71 (5.9-1188.28)	<.001

DFCRS Treatment Demonstrated a Dose-Dependent Effectiveness As Measured by Multidimensional Assessment Scales

After 3 months of DFCRS treatment, the scores of all the multidimensional scales significantly decreased compared with the baseline data (*P*<.001 for all scales), indicating the efficacy of DFCRS therapy in managing tinnitus. The THI score decreased from 48.00 (95% CI 32.00-66.00) points at baseline to 30.00 (95% CI 18.00-42.00) points at the 1-month follow-up, to 22.00 (95% CI 12.00-36.00) points at the 2-month follow-up, and finally to 18.00 (95% CI 8.00-30.00) points at the 3-month follow-up. Significant differences were observed between the scores at baseline and 3 consecutive months of follow-up (*P*<.001; Figure 3).

XSL•FO
RenderX

Figure 3. Box and whisker plots for changes in the scores of the multidimensional assessment scales in digital frequency customized relieving sound–treated patients at 4 timepoints. After 3 months of digital frequency customized relieving sound treatment, the scores on all the 6 scales were significantly lower than those at baseline. Data are presented as median (IQR). AIS: Athens Insomnia Scale; FTQ: Fear of Tinnitus Questionnaire; HADS-A: Hospital Anxiety and Depression Scale-subscale anxiety; HADS-D: Hospital Anxiety and Depression Scale-subscale depression; TCS: Tinnitus Catastrophizing Scale; THI: Tinnitus Handicapped Inventory.

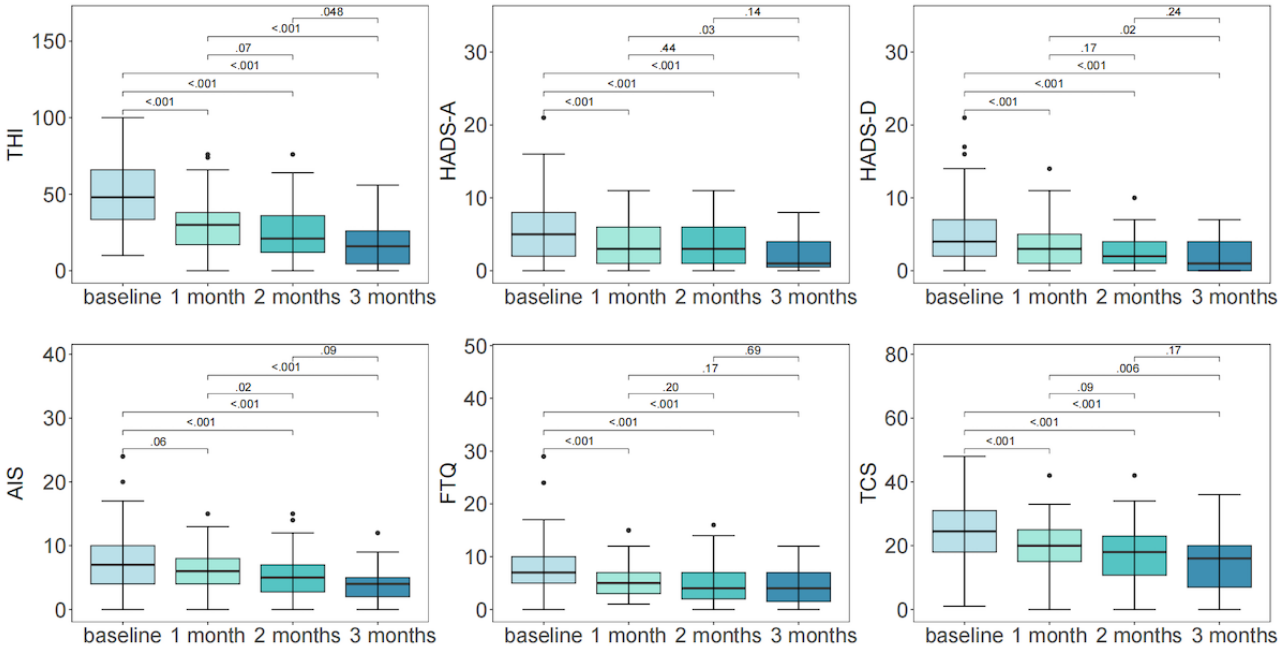

By comparing the 3 follow-up scores with the baseline of the other secondary outcomes, we observed that a more significant decline occurred in the following follow-up than in the preceding one (HADS-A, AIS, FTQ: P<.001 at 1-month follow-up, P<.001 at 2- and 3-month follow-ups; and HADS-D, TCS: P<.01 at 1-month follow-up, P<.001 at 2- and 3-month follow-ups). These significant differences suggested that there was a time-dependent increase in the effectiveness during the experimental procedure.

We also analyzed the differences between 2 of the 3 follow-up outcomes of the 6 assessment scales. Regarding THI scores, significant differences were observed between 2 and 3 months of follow-up (P<.05) and between 1 month and 3 months of follow-up (P<.001). In contrast, the THI scores did not decrease significantly from 1 month of follow-up to 2 months of follow-up (P=.07). A similar reduction was also found in the TCS scores (1 month vs 2 months: P=.09; 1 month vs 3 months: P<.01; 2 months vs 3 months: P=.17; baseline vs 3 months: P<.001), indicating that the therapeutic effect of DFCRS on tinnitus impairment and catastrophizing was observed at 2 months and was maintained in the third month.

Discussion

Customized sound therapy is an advanced method for treating subjective tinnitus that can effectively reduce the loudness of tinnitus and alleviate its negative impact on patients [12]. In this clinical trial, we explored the therapeutic effect of a new self-developed acoustic treatment named DFCRS [32] compared to that of UM over 3 months. Based on the changes in the scores of multidimensional assessment scales and patients' subjective judgments of symptom change, we concluded that the DFCRS delivered by our developed FTRS app [36] was more effective than UM in relieving subjective tinnitus and improving the QoL of patients with tinnitus. Moreover, better effects were obtained as the treatment course was extended or if the patients were exposed to the treatment sounds for longer periods.

We first identified more significant efficacy in the DFCRS treatment group compared to that in the UM control group during the treatment program by using linear mixed models. We further analyzed the possible influencing factors according to the different outcomes of tinnitus treatment and observed that those who achieved better therapeutic outcomes tended to adhere to acoustic treatment longer than those who reported aggravated tinnitus (Figure 2). Additionally, as shown in Figure 3, the assessment scores continued to decrease as the treatment proceeded and the total treatment time accumulated, indicating a steady increase in the efficacy of DFCRS. Interestingly, these 2 results on the treatment duration provided evidence from 2 perspectives for a possible positive correlation between the efficacy of DFCRS and treatment duration and were comparable to those reported in a previous 18-month controlled clinical trial study by Bauer and Brozoski [14] and an 18-month randomized controlled trial study by Henry et al [18], which used tinnitus retraining therapy or tinnitus masking sound.

THI, as the primary outcome, showed the treatment effects of functional, emotional, and catastrophic feedback on tinnitus [37,43]. The results showed significant improvement in the first month with a sustained effect in the following 2 months, indicating that the functional impairment caused by tinnitus and the patient's negative emotional state rapidly improved after starting DFCRS treatment. Additional assessments, including HADS-A, HADS-D, AIS, FTQ, and TCS, also showed rapid

XSL•FO
RenderX

and sustained efficacy after 3 months of DFCRS treatment, and the results demonstrated effective improvements in different aspects such as anxiety and depression, sleep disturbance, and other negative emotions related to tinnitus [41,42,44]. We hypothesized that the patient's trust in the treatment strategy gradually decreased the negative emotions such as fear and pain after personalized acoustic treatment and that physiological abnormalities such as sleep disturbances could be resolved [45]. The negative emotional and physiological responses caused by tinnitus and tinnitus itself interact synergistically. The patients benefited from individualized sound therapy, which helped them develop a positive mindset to overcome tinnitus, and the relief of tinnitus, in return, boosted their confidence in conquering tinnitus, which formed a positive feedback loop [12]. This could explain why patients showed significant relief at the first follow-up visit since the start of the experiment, and the efficacy persisted at the next 2 follow-up visits and was long-lasting. The sample size in our study was larger than those in other similar clinical trials [44], and the participants adhered well to the DFCRS strategy, which guaranteed data credibility.

The main limitation of our study, similar to those of other early clinical trials, is the lack of randomized subgroups [15]. In randomized controlled trials, to avoid selective bias, a more validated evaluation of the effect of sound treatment can be obtained compared to that in nonrandomized controlled trials [18]. In fact, a randomized controlled trial of this sound treatment strategy is already underway based on the initial positive results from this trial [32]. In future trials, a longer follow-up period would be necessary to determine whether the efficacy would be more significant after a longer treatment period [15,45]. In addition, the maintenance of efficacy after stopping the intervention also needs to be explored in a longer trial [15].

In conclusion, we have confirmed the efficacy of DFCRS in treating patients with subjective tinnitus, and thus, it could be a promising treatment option for chronic subjective tinnitus. However, future research needs to be conducted in a more controlled setting, even with a randomized controlled trial, to draw a more precise conclusion.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2023YFC2508401), National Natural Science Foundation of China (82371149), Shanghai Rising-Star Program (23QC1401200), and Shanghai Municipal Key Clinical Specialty (SHSLCZDZK00801). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author (DT) upon reasonable request.

Authors' Contributions

DT and YP conceptualized this study, contributed to the methodology, and wrote the original draft. DT and HL acquired the funds. DT, YP, DG, and YW reviewed and edited this paper, administered the project, and contributed to the methodology, formal analysis, and validation. DT and HL contributed to the investigation and formal analysis. All authors read and approved the final manuscript. DT, YP, and DG have contributed equally to this study and are credited as co-first authors. DT (tang.dongm@163.com), YW (wgek2018@163.com), and HL (hwli@shmu.edu.cn) are co-corresponding authors for this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) statement checklist for nonrandomized control trials.
[PDF File (Adobe PDF File), 491 KB - jmir_v27i1e60150_app1.pdf]

Multimedia Appendix 2
Schematic diagram of the modified strategy for digital frequency customized relieving sound.
[DOCX File , 57 KB - jmir_v27i1e60150_app2.docx]

Multimedia Appendix 3
Additional tables.
[DOCX File , 20 KB - jmir_v27i1e60150_app3.docx]

References

XSL•FO
RenderX

1. Baguley D, McFerran D, Hall D. Tinnitus. Lancet 2013 Nov 09;382(9904):1600-1607 [FREE Full text] [doi: 10.1016/S0140-6736(13)60142-7] [Medline: 23827090]
2. Langguth B, Kreuzer PM, Kleinjung T, De Ridder D. Tinnitus: causes and clinical management. The Lancet Neurology 2013 Sep;12(9):920-930. [doi: 10.1016/s1474-4422(13)70160-1] [Medline: 23948178]
3. Lewis S, Chowdhury E, Stockdale D, Kennedy V, Guideline Committee. Assessment and management of tinnitus: summary of NICE guidance. BMJ 2020 Mar 31;368:m976. [doi: 10.1136/bmj.m976] [Medline: 32234748]
4. Bauer CA. Tinnitus. N Engl J Med 2018 Mar 29;378(13):1224-1231. [doi: 10.1056/NEJMcp1506631] [Medline: 29601255]
5. Eggermont JJ, Roberts LE. The neuroscience of tinnitus. Trends Neurosci 2004 Nov;27(11):676-682. [doi: 10.1016/j.tins.2004.08.010] [Medline: 15474168]
6. Langguth B, Elgoyhen AB, Cederroth CR. Therapeutic approaches to the treatment of tinnitus. Annu Rev Pharmacol Toxicol 2019 Jan 06;59:291-313. [doi: 10.1146/annurev-pharmtox-010818-021556] [Medline: 30044727]
7. Beebe Palumbo D, Joos K, De Ridder D, Vanneste S. The management and outcomes of pharmacological treatments for tinnitus. Curr Neuropharmacol 2015;13(5):692-700 [FREE Full text] [doi: 10.2174/1570159x13666150415002743] [Medline: 26467416]
8. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER, et al. Clinical practice guideline: tinnitus. Otolaryngol Head Neck Surg 2014 Oct;151(2 Suppl):S1-S40. [doi: 10.1177/0194599814545325] [Medline: 25273878]
9. Reavis KM, Rothholtz VS, Tang Q, Carroll JA, Djalilian H, Zeng F. Temporary suppression of tinnitus by modulated sounds. J Assoc Res Otolaryngol 2012 Aug;13(4):561-571 [FREE Full text] [doi: 10.1007/s10162-012-0331-6] [Medline: 22526737]
10. Okamoto H, Stracke H, Stoll W, Pantev C. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. Proc Natl Acad Sci U S A 2010 Jan 19;107(3):1207-1210 [FREE Full text] [doi: 10.1073/pnas.0911268107] [Medline: 20080545]
11. Davis PB, Paki B, Hanley PJ. Neuromonics tinnitus treatment: third clinical trial. Ear Hear 2007 Apr;28(2):242-259. [doi: 10.1097/AUD.0b013e3180312619] [Medline: 17496674]
12. Tang D, Li H, Chen L. Advances in understanding, diagnosis, and treatment of tinnitus. Adv Exp Med Biol 2019;1130:109-128. [doi: 10.1007/978-981-13-6123-4_7] [Medline: 30915704]
13. Chemali Z, Nehmé R, Fricchione G. Sensory neurologic disorders: Tinnitus. Handb Clin Neurol 2019;165:365-381. [doi: 10.1016/B978-0-444-64012-3.00022-8] [Medline: 31727224]
14. Bauer CA, Brozoski TJ. Effect of tinnitus retraining therapy on the loudness and annoyance of tinnitus: a controlled trial. Ear Hear 2011;32(2):145-155. [doi: 10.1097/AUD.0b013e3181f5374f] [Medline: 20890204]
15. Forti S, Costanzo S, Crocetti A, Pignataro L, Del Bo L, Ambrosetti U. Are results of tinnitus retraining therapy maintained over time? 18-month follow-up after completion of therapy. Audiol Neurootol 2009;14(5):286-289. [doi: 10.1159/000212106] [Medline: 19372645]
16. Hobson J, Chisholm E, El Refaie A. Sound therapy (masking) in the management of tinnitus in adults. Cochrane Database Syst Rev 2012 Nov 14;11(11):CD006371 [FREE Full text] [doi: 10.1002/14651858.CD006371.pub3] [Medline: 23152235]
17. Vernon J. Attemps to relieve tinnitus. J Am Audiol Soc 1977;2(4):124-131. [Medline: 845067]
18. Henry JA, Stewart BJ, Griest S, Kaelin C, Zaugg TL, Carlson K. Multisite randomized controlled trial to compare two methods of tinnitus intervention to two control conditions. Ear Hear 2016;37(6):e346-e359. [doi: 10.1097/AUD.0000000000000330] [Medline: 27438870]
19. Tinnitus Retraining Therapy Trial Research Group, Scherer RW, Formby C. Effect of tinnitus retraining therapy vs standard of care on tinnitus-related quality of life: a randomized clinical trial. JAMA Otolaryngol Head Neck Surg 2019 Jul 01;145(7):597-608 [FREE Full text] [doi: 10.1001/jamaoto.2019.0821] [Medline: 31120533]
20. Krick CM, Argstatter H. Neural correlates of the Heidelberg music therapy: indicators for the regeneration of auditory cortex in tinnitus patients? Neural Regen Res 2015 Sep;10(9):1373-1375 [FREE Full text] [doi: 10.4103/1673-5374.165220] [Medline: 26604883]
21. Wazen JJ, Daugherty J, Pinsky K, Newman CW, Sandridge S, Battista R, et al. Evaluation of a customized acoustical stimulus system in the treatment of chronic tinnitus. Otol Neurotol 2011 Jun;32(4):710-716. [doi: 10.1097/MAO.0b013e318217d459] [Medline: 21451428]
22. Li S, Bao L, Chrostowski M. Investigating the effects of a personalized, spectrally altered music-based sound therapy on treating tinnitus: a blinded, randomized controlled trial. Audiol Neurootol 2016;21(5):296-304. [doi: 10.1159/000450745] [Medline: 27838685]
23. Jang DW, Johnson E, Chandrasekhar SS. Neuromonics™ Tinnitus Treatment: preliminary experience in a private practice setting. Laryngoscope 2010;120 Suppl 4:S208. [doi: 10.1002/lary.21675] [Medline: 21225806]
24. Hanley PJ, Davis PB. Treatment of tinnitus with a customized, dynamic acoustic neural stimulus: underlying principles and clinical efficacy. Trends Amplif 2008 Sep;12(3):210-222 [FREE Full text] [doi: 10.1177/1084713808319942] [Medline: 18614554]

XSL•FO
RenderX

25. Stein A, Engell A, Junghoefer M, Wunderlich R, Lau P, Wollbrink A, et al. Inhibition-induced plasticity in tinnitus patients after repetitive exposure to tailor-made notched music. Clin Neurophysiol 2015 May;126(5):1007-1015. [doi: 10.1016/j.clinph.2014.08.017] [Medline: 25441152]
26. Stein A, Wunderlich R, Lau P, Engell A, Wollbrink A, Shaykevich A, et al. Clinical trial on tonal tinnitus with tailor-made notched music training. BMC Neurol 2016 Mar 17;16:38 [FREE Full text] [doi: 10.1186/s12883-016-0558-7] [Medline: 26987755]
27. Teismann H, Okamoto H, Pantev C. Short and intense tailor-made notched music training against tinnitus: the tinnitus frequency matters. PLoS One 2011;6(9):e24685 [FREE Full text] [doi: 10.1371/journal.pone.0024685] [Medline: 21935438]
28. Wang H, Tang D, Wu Y, Zhou L, Sun S. The state of the art of sound therapy for subjective tinnitus in adults. Ther Adv Chronic Dis 2020;11:2040622320956426 [FREE Full text] [doi: 10.1177/2040622320956426] [Medline: 32973991]
29. Jafari Z, Kolb BE, Mohajerani MH. Age-related hearing loss and tinnitus, dementia risk, and auditory amplification outcomes. Ageing Res Rev 2019 Dec;56:100963. [doi: 10.1016/j.arr.2019.100963] [Medline: 31557539]
30. Eggermont JJ, Roberts LE. The neuroscience of tinnitus: understanding abnormal and normal auditory perception. Front Syst Neurosci 2012;6:53 [FREE Full text] [doi: 10.3389/fnsys.2012.00053] [Medline: 22798948]
31. Shore SE, Roberts LE, Langguth B. Maladaptive plasticity in tinnitus--triggers, mechanisms and treatment. Nat Rev Neurol 2016 Mar;12(3):150-160 [FREE Full text] [doi: 10.1038/nrneurol.2016.12] [Medline: 26868680]
32. Tang D, Wang J, Yu X, Yu H. The clinical effects of modified tinnitus relieving sound (MTRS) for chronic tinnitus: protocol for a randomized controlled trial. Trials 2023 Jun 02;24(1):372 [FREE Full text] [doi: 10.1186/s13063-023-07389-8] [Medline: 37268971]
33. Kutyba J, Gos E, Jędrzejczak WW, Raj-Koziak D, Karpiesz L, Niedziałek I, et al. Effectiveness of tinnitus therapy using a mobile application. Eur Arch Otorhinolaryngol 2022 Mar;279(3):1257-1267 [FREE Full text] [doi: 10.1007/s00405-021-06767-9] [Medline: 33783597]
34. Nagaraj MK, Prabhu P. Internet/smartphone-based applications for the treatment of tinnitus: a systematic review. Eur Arch Otorhinolaryngol 2020 Mar;277(3):649-657. [doi: 10.1007/s00405-019-05743-8] [Medline: 31807891]
35. Tang D, Wang K, Ye Z, Gu D, Ye L, Sun S, et al. The Fudan Tinnitus Relieving System (FTRS): The initial results of a smartphone application for tinnitus management and treatment. Internet Interv 2022 Sep;29:100564 [FREE Full text] [doi: 10.1016/j.invent.2022.100564] [Medline: 36092991]
36. Tang D, Wang H, Gu D, Ye L, Sun S, Li H. The fudan tinnitus relieving system application for tinnitus management. BMC Med Inform Decis Mak 2023 Apr 21;23(1):76 [FREE Full text] [doi: 10.1186/s12911-023-02164-w] [Medline: 37085904]
37. Oud MM, Latour BL, Bakey Z, Letteboer SJ, Lugtenberg D, Wu KM, et al. Cellular ciliary phenotyping indicates pathogenicity of novel variants in and confirms a Mainzer-Saldino syndrome diagnosis. Cilia 2018;7:1 [FREE Full text] [doi: 10.1186/s13630-018-0055-2] [Medline: 30479745]
38. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
39. Zöger S, Svedlund J, Holgers K. The Hospital Anxiety and Depression Scale (HAD) as a screening instrument in tinnitus evaluation. Int J Audiol 2004 Sep;43(8):458-464. [doi: 10.1080/14992020400050058] [Medline: 15643739]
40. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [doi: 10.1016/s0022-3999(00)00095-7] [Medline: 11033374]
41. Fuller TE, Cima RFF, Van den Bussche E, Vlaeyen JWS. The Fear of Tinnitus questionnaire: toward a reliable and valid means of assessing fear in adults with tinnitus. Ear Hear 2019;40(6):1467-1477. [doi: 10.1097/AUD.0000000000000728] [Medline: 30998546]
42. Cima RFF, Crombez G, Vlaeyen JWS. Catastrophizing and fear of tinnitus predict quality of life in patients with chronic tinnitus. Ear Hear 2011;32(5):634-641. [doi: 10.1097/AUD.0b013e31821106dd] [Medline: 21399500]
43. Newman CW, Sandridge SA, Jacobson GP. Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. J Am Acad Audiol 1998 Apr;9(2):153-160. [Medline: 9564679]
44. Andersson G, Kaldo-Sandström V, Ström L, Strömgren T. Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. J Psychosom Res 2003 Sep;55(3):259-262. [doi: 10.1016/s0022-3999(02)00575-5] [Medline: 12932800]
45. Beukes EW, Andersson G, Allen PM, Manchaiah V, Baguley DM. Effectiveness of guided internet-based cognitive behavioral therapy vs face-to-face clinical care for treatment of tinnitus: a randomized clinical trial. JAMA Otolaryngol Head Neck Surg 2018 Dec 01;144(12):1126-1133 [FREE Full text] [doi: 10.1001/jamaoto.2018.2238] [Medline: 30286238]

Abbreviations

AIS: Athens Insomnia Scale
DFCRS: digital frequency customized relieving sound
FTQ: Fear of Tinnitus Questionnaire
FTRS: Fudan Tinnitus Relieving System
HADS: Hospital Anxiety and Depression Scale

XSL•FO
RenderX

HADS-A: Hospital Anxiety and Depression Scale-subscale anxiety
HADS-D: Hospital Anxiety and Depression Scale-subscale depression
OR: odds ratio
QoL: quality of life
TCS: Tinnitus Catastrophizing Scale
THI: Tinnitus Handicapped Inventory
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
UM: unmodified music

Edited by A Schwartz; submitted 03.05.24; peer-reviewed by Q Chen, H Kim; comments to author 16.06.24; revised version received 11.08.24; accepted 28.11.24; published 17.01.25.

Please cite as:
Tang D, Peng Y, Gu D, Wu Y, Li H
Digital Frequency Customized Relieving Sound for Chronic Subjective Tinnitus Management: Prospective Controlled Study
J Med Internet Res 2025;27:e60150
URL: https://www.jmir.org/2025/1/e60150
doi:10.2196/60150
PMID:

©Dongmei Tang, Yuzhu Peng, Dantong Gu, Yongzhen Wu, Huawei Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Mobile Health–Based Behavioral Treatment for Lifestyle Modification in Type 2 Diabetes Self-Management: Greenhabit Randomized Controlled Trial

Ana Maria Ruiz-Leon[1,2,3,4,5*], MS; Rosa Casas[1,2,3,5], MS, PhD; Sara Castro-Barquero[2,3,5*], MS, PhD; Sofia Alfaro-González[1], MS; Petia Radeva[6], MS, PhD; Emilio Sacanella[1,2,5], MD, PhD; Francesc Casanovas-Garriga[1,3], MS; Ainhoa Pérez-Gesalí[1], MS; Ramon Estruch[1,2,3,5], MD, PhD

[1]Department of Internal Medicine, Hospital Clinic, Institut d'Investigacio Biomèdica August Pi i Sunyer, Barcelona, Spain

[2]CIBERobn Fisiopatología de la Obesidad y la Nutrición, Instituto de Salud Carlos III, Madrid, Spain

[3]Institut de Recerca en Nutrició i Seguretat Alimentaria. University of Barcelona, Barcelona, Spain

[4]Fundación Dieta Mediterránea, Barcelona, Spain

[5]Faculty of Medicine and Life Sciences, University of Barcelona, Barcelona, Spain

[6]Faculty of Mathematics and Computer Science, University of Barcelona, Barcelona, Spain

[*]these authors contributed equally

Corresponding Author:
Rosa Casas, MS, PhD
Department of Internal Medicine, Hospital Clinic, Institut d'Investigacio Biomèdica August Pi i Sunyer
Villarroel, 170
Barcelona, 08036
Spain
Phone: 34 932275745
Fax: 34 932275758
Email: rcasas1@recerca.clinic.cat

Abstract

Background: Enhancing self-management in health care through digital tools is a promising strategy to empower patients with type 2 diabetes (T2D) to improve self-care.

Objective: This study evaluates whether the Greenhabit (mobile health [mHealth]) behavioral treatment enhances T2D outcomes compared with standard care.

Methods: A 12-week, parallel, single-blind randomized controlled trial was conducted with 123 participants (62/123, 50%, female; mean age 58.25 years, SD 9.46 years) recently diagnosed with T2D. Participants were recruited face-to-face from primary care centers in Barcelona, Spain, between July 2021 and March 2022. They were randomly assigned to 1 of 2 groups: (1) an intervention group (n=61) instructed to use the Greenhabit mobile app alongside standard care, or (2) a control group (n=62) who received advice on maintaining a healthy diet and followed standard care. The Greenhabit app incorporates serious gaming technology. Participants received daily messages and challenges focused on promoting a healthy lifestyle, including nutrition, exercise, relaxation, a positive mindset, and a supportive social environment. The app encouraged participants to set weekly goals and awarded points for completing challenges. Data on nutrition, anthropometrics, and blood and urine samples were collected at baseline, 6 weeks, and 12 weeks. Questionnaires assessing quality of life, work-life balance, and social environment were administered at baseline and during the final visit. The primary outcomes were HbA1c and fasting plasma glucose (FPG). Repeated-measures analysis of variance was used to compare changes over time (baseline to 6 weeks and baseline to 12 weeks) between the 2 intervention groups. Analysis of covariance was performed to evaluate changes at 6 and 12 weeks, adjusted for baseline levels of each variable. Multiple contrasts were corrected using a Bonferroni post hoc test.

Results: Both groups showed significant reductions in HbA1c after 6 and 12 weeks (mean change in the intervention group [n=50] –0.4%, $P<.001$ vs –0.3% in the control group [n=53], $P=.001$) and in FPG after 6 weeks (mean change in the intervention group –5.3 mg/dL, $P=.01$ vs control group –5.8 mg/dL, $P=.01$). At 12 weeks, the intervention group also showed significant reductions in systolic and diastolic blood pressures (mean change –4.5, $P=.049$ and –2.4 mmHg, $P=.03$, respectively), body weight (mean change –0.8 kg, $P=.03$), BMI (mean change –0.3 kg/m2, $P=.03$), waist circumference (mean change –1.0 cm, $P=.046$),

XSL•FO
RenderX

and triglyceride concentration (mean change –20.0 mg/dL, *P*=.03). There was also a significant increase in high-density lipoprotein-cholesterol concentrations (mean change 2 mg/dL, *P*=.049). Finally, improvements were noted in 3 out of the 5 elements of balance: positivity, social environment, and work-life balance.

Conclusions: The 12-week intervention with the Greenhabit behavioral treatment mHealth app showed beneficial effects on T2D outcomes and reduced the burden of cardiovascular risk factors. Although larger studies are warranted, these results suggest that mHealth apps can be a promising tool for improving T2D self-management.

Trial Registration: ISRCTN Registry ISRCTN13456652; http://www.isrctn.com/ISRCTN13456652

(J Med Internet Res 2025;27:e58319) doi:10.2196/58319

KEYWORDS

type 2 diabetes; cardiovascular health; lifestyle intervention; mHealth; artificial intelligence; mobile health; mobile application; diabetes; DM; self-management; randomized controlled trial; RCT; health care; self-care; Greenhabit; behavioral health; treatment; medication; analysis of covariance; intervention; work-life balance; cardiovascular; cardiovascular risk

Introduction

Diabetes is a chronic metabolic disease characterized by hyperglycemia, which results from defects in insulin secretion, insulin action, or both [1,2]. Over time, this condition leads to severe complications, contributing to significant morbidity and mortality worldwide [3].

According to the World Health Organization, diabetes remains a critical public health issue, with 422 million people affected in 2014. The global prevalence of diabetes is steadily increasing, with projections estimating 570 million cases by 2030 and 700 million by 2045, largely driven by rising rates of overweight and obesity [4]. The growing prevalence of type 2 diabetes (T2D) and the expanding population affected underscore an urgent need to develop novel strategies for diabetes treatment, treatment adherence, lifestyle modifications, and care. The rising prevalence of T2D places a significant strain on health care systems, highlighting the importance of empowering individuals to manage their conditions effectively [5]. Consequently, numerous mobile health (mHealth) apps have been developed as innovative tools to enhance self-management of T2D and prevent macrovascular and microvascular complications [6-12].

Patients with T2D require effective diabetes control and self-management skills to adopt a healthier lifestyle. Self-management encompasses daily activities aimed at controlling or reducing the disease's impact on health and well-being, including monitoring blood glucose levels, managing food intake, engaging in physical activity, and managing stress [13]. mHealth apps play a pivotal role in achieving glycemic control, encouraging lifestyle modifications, and promoting self-management among individuals with T2D [14]. Recent research highlights the growing adoption of these apps for managing treatment and care among this population [13]. They empower self-management by supporting dietary adherence, consistent blood sugar monitoring, regular physical activity, reduced medication needs, and lower glycated hemoglobin (HbA_{1c}) levels [15,16]. Additionally, they help reduce face-to-face counseling time and health care expenses [17].

Currently, a wide range of mHealth apps is available to enhance self-management for individuals with diabetes. These apps facilitate close monitoring, provide feedback, and help overcome geographic barriers [14]. Furthermore, a systematic review and meta-analysis have demonstrated the effectiveness of mHealth apps in empowering self-management across various domains of disease management and lifestyle changes [18]. Additionally, the American Diabetes Association guidelines advocate for the use of mHealth apps as a valuable tool for achieving effective lifestyle changes in diabetes prevention [19]. However, the lack of consensus on the definition of self-management interventions for chronic diseases [20,21], combined with diverse outcome measures—such as self-efficacy, health-related quality of life, activity participation, performance of daily living activities, depression, and self-confidence [22]—poses challenges in comparing interventions and interpreting results. Greenhabit is a holistic education program that empowers individuals to take control of their health by fostering sustainable behavior change. Its approach, which addresses the physical, mental, and social aspects of well-being, promotes mental resilience and self-awareness [23]. The Greenhabit method is grounded in 2 key theories: (1) the *McClelland Iceberg model*, which serves as a framework for building content, challenges, and measures for users [24]; and (2) the *Behavioral Change Wheel*, which informs the development of interventions and techniques within Greenhabit [25]. Social integration (through buddies and community) provides motivation and social support to help maintain healthy habits. It encourages individuals to take care of their health by offering nudges related to nutrition, physical activity, and social interactions, all aimed at promoting overall well-being. The game/app incorporates serious gaming technology. The vitality education program is a 12-week journey in which players focus on the 5 elements of Greenhabit: nutrition, exercise, relaxation, positive mindset, and social environment. It leverages the power of brain-based learning through methods such as nudges, content, challenges, rewards, working with a buddy, and engaging the senses with joy. The brain is plastic, allowing us to learn new habits. It takes at least 68 days to establish new habits [26]. After 12 weeks, players report increased awareness, resilience, and energy.

Thus, the primary aim of this study was to assess the efficacy of the Greenhabit app on T2D outcomes, specifically plasma glucose concentrations and HbA_{1c} levels, in patients with a recent diabetes diagnosis. As a secondary aim, we evaluated whether participation in the app led to improvements in blood pressure (BP), plasma lipids, adiposity parameters, and the maintenance of a healthy lifestyle after 12 weeks of use.

XSL•FO
RenderX

Methods

Study Design

This study is a 12-week, parallel-group, single-blind, single-center, randomized controlled trial (RCT) designed to assess the efficacy of the Greenhabit method on diabetes treatment and management. The intervention, a low-intensity self-management program, focuses on behavior change and incorporates evidence-based approaches to diabetes self-management education and support [27]. The diabetes specialist dietitian and nurse applied principles of motivational interviewing, including a person-to-person, client-centered approach aimed at eliciting behavior change by helping patients explore and resolve ambivalence. Additionally, a successful telephonic intervention, proven to improve diabetes control in urban adults, was incorporated.

Consecutive candidates were randomly assigned to either the intervention or control group after sex stratification, with an allocation ratio of 1:1.

Outcomes

The primary outcome of the trial was the management of T2D, measured by HbA_{1c} and fasting plasma glucose (FPG) concentrations at 12 weeks. The secondary outcomes included the effect of the intervention on cardiovascular risk factors, such as BP, lipid profile, body weight (BW), and adiposity, as well as its impact on dietary habits, stress relief, and psychological factors, assessed using validated questionnaires.

Setting and Participants

As shown in Figure 1, from July 2021 to March 2022, we screened 1045 potential candidates from the databases of primary care centers affiliated with Hospital Clínic of Barcelona, Spain. Of the screened participants, 684 were excluded for not meeting the inclusion criteria of a recent T2D diagnosis, while 238 were excluded for various other reasons: 179 had limited ability to use a smartphone, 48 had a severe chronic disease, 5 faced language barriers, 4 lacked a smartphone, and 2 had difficulty following an appropriate diet. Finally, 123 participants were recruited through a face-to-face interview and randomized, with 61 assigned to the Greenhabit group and 62 to the control group. Randomization was carried out by consecutively assigning participants using sealed envelopes, which were prepared based on a computer-generated random number table. Of these participants, 20 were withdrawn from the trial because they could not attend follow-up visits during the COVID-19 pandemic. As a result, 50 participants in the intervention group and 53 in the control group completed the study.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the Greenhabit 12-week randomized controlled trial.

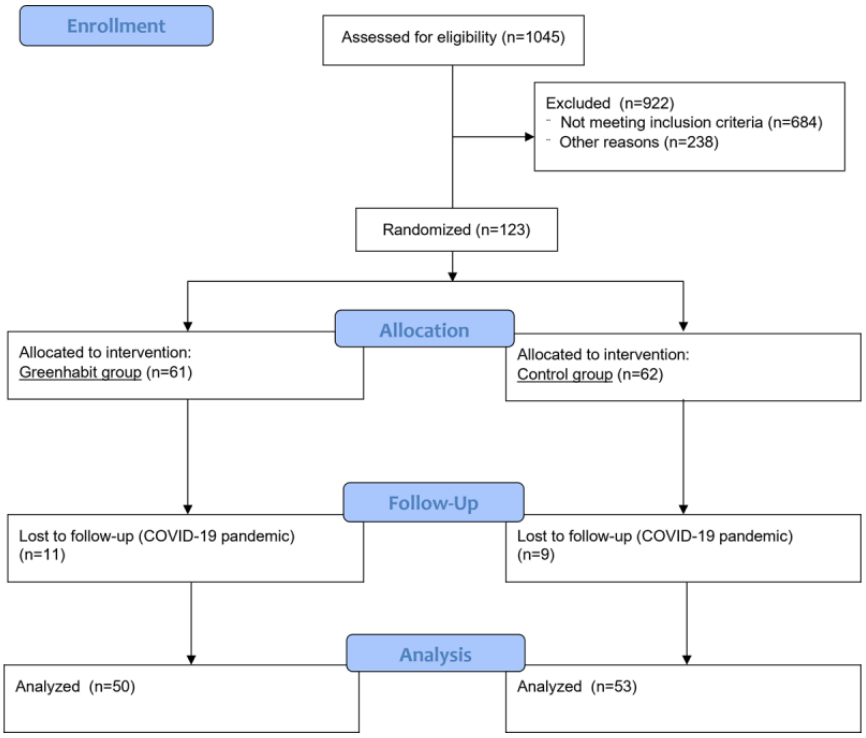

Inclusion criteria were as follows: men and women aged 18-75 years, with a recent diagnosis of T2D (<2 years of treatment with insulin, antidiabetic medications, a combination of both, or a nonpharmacological treatment), and either an HbA_{1c} greater than 6.5%, FPG greater than 126 mg/dL on 2 occasions, or a 2-hour plasma glucose greater than 200 mg/dL after a 75-g oral glucose load.

Exclusion criteria included participants who underwent or planned bariatric surgery during the trial period; individuals with type 1 diabetes; those with a diagnosis of gestational diabetes; and individuals with severe chronic illnesses, alcoholism, or other drug addictions. Also excluded were participants with gastrointestinal diseases that would hinder adherence to a specific diet, those who had taken vitamins or nutritional supplements in the month before recruitment, pregnant women, individuals without their own smartphone,

XSL•FO
RenderX

those unable to attend additional visits, and those already participating in another trial.

Ethics Approval and Consent to Participate

Before starting the study, all participants provided informed consent during a face-to-face interview with the researcher. The study protocol and procedures were approved by the Committee for Ethical Research at Hospital Clínic (Barcelona, Spain) on April 19, 2021, and adhere to the ethical standards of the Declaration of Helsinki. The study was registered with the following code: HCB/2021/0061. The trial was also registered with the International Standard Randomized Controlled Trial Number registry (ISRCTN13456652).

Intervention

Participants in the intervention group were instructed to follow standard care recommendations, similar to the control group, and were provided with personalized, free access to the Greenhabit app via email. Additionally, they received a second free access for a partner (eg, sibling, friend, spouse) to offer extra motivation and mutual support. Participants could choose whether or not to use this second access. Each week, 3-4 days before starting the Greenhabit intervention, participants, grouped into teams of 10-12 individuals, were invited to attend an online group meeting. The purpose of these meetings was to teach participants how to use the app and address any potential technological barriers. Participants could invite friends or colleagues to join their journey for free; these invited individuals could only view the photos selected by the participants. Additionally, participants were encouraged to set weekly goals and log their measurements (eg, steps walked, weight, glucose levels, HbA_{1c}). Each day, participants received a "treasure chest" from Greenhabit, containing messages and challenges related to healthy lifestyle habits (nutrition, exercise, relaxation, positive thinking, and social environment). Participants earned points for each challenge they completed. Various rewards, such as a water bottle for tracking intake, a jump rope, a notebook, and card games, were given throughout the 12-week program upon completing each level.

Participants in the control group received guidance on maintaining a healthy diet and were given written materials prepared by a dietitian. These materials included information on seasonal Mediterranean foods, shopping lists, weekly meal plans, and recipes for a typical week. Additionally, participants in this group received standard care through their primary care center. Usual care included routine health care appointments with providers, as well as access to diabetes management and weight loss education, electronic learning materials, and guidance on how to use the health system's patient portal [28]. The main goal was to equip participants with the skills needed to maintain or improve their health.

Greenhabit Behavioral Treatment Program

The Greenhabit program is a 12-week mHealth cognitive-behavioral education initiative designed specifically to support individuals in managing T2D. Greenhabit is a behavioral change program in which patients receive personalized content (interventions) based on their clinical profile through a recommender system. This system leverages an artificial intelligence–based infrastructure to promote personalized lifestyle changes. Additional information about the questionnaires used can be found in Supplementary Methods S1 in Multimedia Appendix 1 (see also [23-25,29-34]).

Diet and Clinical Measurements

All participants underwent a face-to-face interview with the dietitian at baseline and after 12 weeks of intervention, along with regular phone calls. During these individual sessions, each participant received information on seasonal foods, shopping lists, weekly meal plans, and cooking recipes. Additionally, the individual interviews included a 151-item validated Food Frequency Questionnaire [35] and personalized recommendations for dietary changes aimed at achieving the participant's goals. Participants also completed a 47-item questionnaire on education, lifestyle, medical history, and medication use. To assess adherence to the Mediterranean diet, a validated 14-point Mediterranean diet screener was used [36]. In the intervention group, participants were encouraged to increase their intake of vegetables (≥2 servings/day), fresh fruit (≥3 servings/day), legumes, nuts, and fish or seafood (≥3 servings/week), and to use olive oil for cooking and dressings. The app also promoted physical activity. Energy restriction was not specifically advised for either group.

At baseline (visit 1), 6 weeks (visit 2), and 12 weeks (visit 3), trained personnel (consistently the same dietitian and nurse) collected anthropometric measurements, determined BP, and took biological samples. Height was measured to the nearest 0.1 cm using a wall-mounted stadiometer with participants standing, wearing socks, and with their heads in the horizontal plane. BW was measured to the nearest 0.1 kg, with participants in light clothing and without shoes or accessories, using a high-quality calibrated scale. BMI was calculated by dividing weight (kg) by the square of height (m^2). Waist circumference (WC) was measured midway between the lowest rib and the iliac crest using an anthropometric tape. Hip circumference (HC) was measured at the maximum point below the waist, without compressing the skin [37]. BP was measured in triplicate after at least 5 minutes of rest using a validated semiautomatic oscillometer (Omron HEM-705CP). The mean value of the 3 BP measurements was used in this study. A medical record was also collected. Fasting blood and spot urine samples were obtained for basic hematologic and biochemical analyses for all participants, including HbA_{1c}, glucose, and lipid profile parameters. Additionally, the nurse provided diabetes self-management support to patients as needed.

Technicians and nurses were blinded to the group assignment of each participant.

Quality-of-Life Questionnaires

At baseline and after 12 weeks, participants completed several questionnaires to assess perceived physical and mental health over time, focusing on various aspects of health-related quality of life, including energy, happiness, positivity, social environment, and work-life balance. Additional details about the tests used can be found in Supplementary Methods S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Validity and Reliability

Blood samples were collected at baseline and at the end of the intervention to determine biochemical and nutritional parameters. Hemoglobin was measured by high-performance liquid chromatography, blood glucose by the glucose oxidase method, and total cholesterol and triglycerides by enzymatic procedures. High-density lipoprotein (HDL) cholesterol was assessed after precipitation with phosphotungstic acid and magnesium chloride. All analyses were performed using the Advia 2400 clinical chemistry analyzer (Siemens Healthcare), with reagents provided by the instrument manufacturer. The main outcomes of the trial (HbA_{1c} and FPG levels) were determined at the CORE Laboratory at Hospital Clinic de Barcelona, which meets all required quality standards and holds ISO 9001:2015 certification and ISO 15189 accreditation for certain tests.

The instruments used for dietary assessment are both validated and have been utilized in large randomized clinical trials and cohort studies (see [35,36]).

Regarding quality of life, all the questionnaires are detailed in Supplementary Methods S1 in Multimedia Appendix 1. These questionnaires are divided into 5 components: energy, happiness, positivity, social life, and work-life balance. To assess these components, we used validated questionnaires tailored to the Spanish population, and further details of these questionnaires are extensively presented in Multimedia Appendix 1. In brief, the International Physical Activity Questionnaire, in its Spanish version, was used to assess physical activity [38,39]. The 36-item Short Form Health Survey was used to measure quality of life. This generic scale is applicable to both patients and the general population, providing a comprehensive profile of health status. The questionnaire includes 8 scales: Physical Function, Physical Role, Body Pain, General Health, Vitality, Social Function, Emotional Role, and Mental Health. The Spanish version is a validated instrument, suitable for both medical research and clinical practice [40]. The Hospital Anxiety and Depression Scale is a validated, rapid self-assessment screening tool. Although not intended for diagnostic purposes, it is highly convenient for evaluating anxiety and depression in patients with somatic or mental health issues [41]. This test has been widely used in patients with T2D [42]. Additionally, the Life Orientation Test assesses the psychometric properties of an optimism scale and has been adapted for the Spanish population using the multidimensional graded response model. This 10-item scale evaluates an individual's life expectations, including 3 positive and 3 negative items. The test has also been widely used in individuals with T2D [43]. Finally, the Duke-UNK Functional Social Support Questionnaire is an objective tool that measures "perceived functional social support." It is a self-administered, validated instrument for the Spanish population [44].

Statistical Analysis

The total number of participants included was 123. Power calculations, based on an estimated intervention-induced intergroup difference of 0.8% in HbA_{1c} with an SD of 2.75, indicated that 104 individuals were needed to detect a statistically significant difference of .05 units (α risk=.05 and β risk=.2 in a bilateral test). A follow-up loss rate of 10% was anticipated. A reduction in HbA_{1c} greater than 0.5% is considered clinically significant, as diabetes-related complications are directly proportional to blood HbA_{1c} levels [45].

We used descriptive statistics, presenting the mean (SD) for the baseline characteristics of the participants. Categorical variables are expressed as percentages. For baseline characteristics, the Pearson chi-square test was used for categorical variables, and 1-way analysis of variance (ANOVA) was applied for continuous variables to compare between groups. Differences in biochemical parameters, anthropometric measures, food, and nutrient intake at baseline, 6 weeks, and 12 weeks of intervention were assessed using ANOVA. Changes in biochemical parameters and anthropometric variables after 6 and 12 weeks of intervention were assessed using analysis of covariance (ANCOVA), adjusting for baseline levels of each variable, age, sex, use of hypoglycemic treatment (yes/no), and BMI. BMI was categorized according to World Health Organization classifications: normal weight (<25 kg/m^2), overweight (25 to <30 kg/m^2), obesity class 1 (30 to <35 kg/m^2), obesity class 2 (35 to <40 kg/m^2), and obesity class 3 (≥40 kg/m^2). Significant interactions were analyzed using simple-effect analysis, and multiple contrasts were adjusted using a Bonferroni post hoc test. Within- and between-group differences were expressed as estimated means with 95% CIs. The significance level was set at $P<.05$. All analyses were conducted using SPSS version 20.0 (IBM Corp.).

Results

Study Population

Of the 123 participants assessed (Figure 1; Multimedia Appendix 2), 31 in the Greenhabit group (n=61, 51%) and 31 in the control group (n=62, 50%) were female. The mean BMI was 31.8 (SD 6.14) kg/m^2 for the Greenhabit group and 32.4 (SD 6.49) kg/m^2 for the control group. Participants in the Greenhabit group were slightly younger, with a mean age of 56.0 (SD 10.8) years, compared with 60.5 (SD 8.11) years in the control group. No significant between-group differences were observed (Table 1).

XSL•FO
RenderX

Table 1. Baseline characteristics of the participants in the 2 study groups of the Greenhabit 12-week randomized controlled trial.

Variables	Greenhabit (n=61)	Control (n=62)	*P* value[a]
Age (years), mean (SD)	56.0 (10.8)	60.5 (8.11)	.01
Women, n (%)	31 (51)	31 (50)	.93
Smoking status, n (%)			
Current smokers	18 (30)	11 (18)	.15
Educational level, n (%)			.16
Primary	14 (23)	23 (37)	
Secondary	18 (30)	17 (27)	
Higher	29 (48)	19 (31)	
Marital status, n (%)			.55
Married	35 (57)	42 (68)	
Single	11 (18)	9 (15)	
Divorced, separated, or widowed	15 (25)	11 (18)	
Alcohol drinker's status, n (%)			
Current alcohol drinkers	45 (74)	39 (63)	.36
MedDiet P14-score[b] (points), mean (SD)	8.25 (2.17)	8.26 (1.92)	.80
Adherence MedDiet[b], n (%)			.55
Low adherence (<6 points)	6 (10)	3 (5)	
Medium adherence (6-9 points)	36 (59)	40 (65)	
High adherence (≥10 points)	19 (31)	19 (31)	
Body weight (kg), mean (SD)	87.6 (20.5)	88.1 (21.6)	.34
BMI (kg/m^2), mean (SD)	31.8 (6.14)	32.4 (6.49)	.60
BMI, n (%)			.09
18.5-24.9	4 (7)	10 (16)	
25-29.9	22 (36)	16 (26)	
30-34.9	14 (23)	21 (34)	
35-39.9	16 (26)	8 (13)	
≥40	5 (8)	7 (11)	
Waist circumference (cm), mean (SD)	106.7 (13.6)	108.2 (15.2)	.55
Hip circumference (cm), mean (SD)	110.8 (13.3)	113.0 (15.8)	.42
Waist-to-height ratio, mean (SD)	0.67 (0.08)	0.68 (0.09)	.47
Systolic blood pressure (mmHg), mean (SD)	141.2 (18.5)	144.1 (19.2)	.42
Diastolic blood pressure (mmHg), mean (SD)	84.8 (9.82)	81.8 (9.95)	.11
Glucose (mg/dL), mean (SD)	130.1 (33.3)	134.9 (38.2)	.49
Glycated hemoglobin (%), mean (SD)	6.91 (1.19)	6.96 (0.78)	.82
Medications, n (%)			
Oral hypoglycemic drugs	45 (74)	44 (71)	.84
No antidiabetic medication	13 (21)	18 (29)	.85
Antihypertensive agents	23 (38)	20 (32)	.53
Lipid-lowering agents	20 (33)	28 (45)	.14

[a]From the Pearson chi-square test for categorical variables and 1-factor analysis of variance for continuous variables.

[b]MedDiet P14-score: Mediterranean diet adherence evaluated using a 14-point score.

XSL•FO
RenderX

No changes in medication (glucose-lowering agents) were reported after the 12-week intervention, and no comorbidities associated with T2D were observed.

Blood Pressure, Cardiovascular Risk Factors, and Adiposity

Table 2 (adjusted variables) and Table S1 in Multimedia Appendix 1 (unadjusted variables) present baseline values and changes in anthropometric measures and cardiovascular risk factors at 6 and 12 weeks by the intervention group. Participants in both groups showed reductions in systolic BP at 12 weeks (P=.049 and P=.02 in the intervention and control groups, respectively), FPG at 6 weeks, and HbA_{1c} at both 6 and 12 weeks (P<.001 for both). At 6 weeks, participants in the intervention group demonstrated a significant improvement in diastolic BP (P=.007). By 12 weeks, they also showed significant reductions in triglyceride concentrations (P=.03), BW (P=.03), BMI (P=.03), WC (P=.046), and HC (P=.049), along with a significant increase in HDL-cholesterol levels (P=.049).

XSL•FO
RenderX

Table 2. Changes in blood pressure, cardiovascular risk factors, and adiposity, at baseline and after 6 and 12 weeks of follow-up in the 2 study groups of the Greenhabit randomized controlled trial[a].

Variables	Greenhabit (n=50)	Control (n=53)	Differences between groups (Greenhabit vs control)
Systolic blood pressure (mmHg)			
Baseline	141.3 (19.5)	142.5 (18.5)	N/A[b]
Changes after 6 weeks	–3.7 (–8.0 to 0.5)	–1.8 (–5.9 to 2.3)	4.0 (–7.6 to 15.6)
Changes after 12 weeks	–4.5 (–9.0 to 0.0)[c]	–5.0 (–9.4 to –0.7)[c]	1.3 (–10.2 to 12.8)
Diastolic blood pressure (mmHg)			
Baseline	84.9 (10.6)	81.2 (10.1)	N/A
Changes after 6 weeks	–2.9 (–5.0 to –0.8)[c]	0.5 (–1.5 to 2.5)	–6.3 (–11.3 to –1.3)[d]
Changes after 12 weeks	–2.4 (–4.4 to –0.3)[c]	–0.4 (–2.5 to 1.6)	–3.5 (–8.9 to 1.8)
Fasting plasma glucose (mg/dL)			
Baseline	125.7 (29.1)	132.4 (38.0)	N/A
Changes after 6 weeks	–5.3 (–9.6 to –1.0)[c]	–5.8 (–10.1 to –1.5)[c]	–0.09 (–9.0 to 8.8)
Changes after 12 weeks	–4.0 (–9.3 to 1.2)	–1.4 (–6.5 to 3.6)	–2.7 (–14.3 to 8.9)
Glycated hemoglobin (%)			
Baseline	6.88 (1.12)	6.90 (0.76)	N/A
Changes after 6 weeks	–0.4 (–0.6 to –0.3)[e]	–0.3 (–0.4 to –0.1)[e]	–0.1 (–0.5 to –0.2)
Changes after 12 weeks	–0.4 (–0.5 to –0.2)[e]	–0.3 (–0.4 to –0.1)[c]	–0.1 (–0.6 to –0.3)
Triglycerides (mg/dL)			
Baseline	157.3 (65)	143.1 (55.2)	N/A
Changes after 6 weeks	–12.1 (–25.4 to 1.1)	–2.4 (–15.3 to 10.6)	12.3 (–28.0 to 52.6)
Changes after 12 weeks	–20.0 (–35.3 to –2.7)[c]	–7.8 (–23.9 to 8.3)	–14.1 (–73.8 to 45.6)
Total cholesterol (mg/dL)			
Baseline	177.0 (40.7)	181.3 (36.3)	N/A
Changes after 6 weeks	0.4 (–5.8 to 6.6)	1.3 (–4.8 to 7.4)	4.1 (–8.8 to 17.0)
Changes after 12 weeks	–0.4 (–6.3 to 5.5)	–2.2 (–8.1 to 3.8)	0.09 (–18.0 to 17.8)
High-density lipoprotein-cholesterol (mg/dL)			
Baseline	42.4 (10.2)	46.3 (12.1)	N/A
Changes after 6 weeks	–0.9 (–2.7 to 0.9)	–1.0 (–2.7 to 0.8)	–0.9 (–3.9 to 2.1)
Changes after 12 weeks	2.0 (0.1 to 4.0)[c]	–1.5 (–3.4 to 0.5)	1.1 (–2.6 to 4.9)
Low-density lipoprotein-cholesterol (mg/dL)			
Baseline	112.5 (38.3)	112.5 (38.3)	N/A
Changes after 6 weeks	1.3 (–5.4 to 8.1)	2.9 (–3.8 to 9.6)	0.2 (–13.7 to 14.0)
Changes after 12 weeks	1.0 (–5.3 to 7.3)	–1.7 (–4.7 to 8.2)	–5.6 (–25.8 to 14.5)
Body weight (kg)			
Baseline	89.1 (21.8)	90.4 (21.8)	N/A
Changes after 6 weeks	–0.2 (–0.8 to 0.5)	0.1 (–0.5 to 0.7)	–0.9 (–3.5 to 1.8)
Changes after 12 weeks	–0.8 (–1.6 to –0.0)[c]	–0.1 (–0.8 to 0.6)	–0.5 (–4.1 to 3.0)
BMI (kg/m^2)			
Baseline	32.2 (6.3)	32.5 (6.4)	N/A

XSL•FO
RenderX

Variables	Greenhabit (n=50)	Control (n=53)	Differences between groups (Greenhabit vs control)
Changes after 6 weeks	–0.1 (–0.3 to 0.1)	–0.0 (–0.2 to 0.2)	0.3 (–1.3 to 0.6)
Changes after 12 weeks	–0.3 (–0.5 to –0.0)[c]	0.0 (–0.3 to 0.2)	–0.4 (–1.5 to 0.6)
Waist circumference (cm)			
Baseline	107.7 (14.5)	108.3 (14.5)	N/A
Changes after 6 weeks	–0.5 (–1.4 to 0.3)	0.4 (–0.4 to 1.3)	–0.9 (–2.5 to 0.7)
Changes after 12 weeks	–1.0 (–2.1 to 0.0)[c]	0.6 (–0.4 to 1.6)	–1.1 (–3.1 to –0.8)[d]
Hip circumference (cm)			
Baseline	111.2 (14.1)	113 (15.3)	N/A
Changes after 6 weeks	–0.7 (–1.6 to 0.3)	0.1 (–0.8 to 1.0)	0.0 (–2.0 to 2.0)
Changes after 12 weeks	–1.0 (–2.0 to –0.0)[c]	0.0 (–0.9 to 1.0)	–2.0 (–4.1 to 0.3)
Waist-to-hip ratio			
Baseline	1.0 (0.1)	1.0 (0.1)	N/A
Changes after 6 weeks	0.00 (–0.01 to 0.01)	0.00 (–0.01 to 0.01)	–0.04 (–0.01 to 0.1)
Changes after 12 weeks	–0.0 (–0.01 to 0.01)	0.00 (–0.01 to 0.02)	–0.01 (–0.08 to 0.06)

[a]Values are expressed as mean (SDs) or mean differences (95% CI). Comparisons between groups with 1-way analysis of variance and comparisons between groups after 6 and 12 weeks of intervention with analysis of covariance adjusted for baseline levels of each variable, age, sex, use of hypoglycemic treatment (yes/no), and BMI are presented.

[b]N/A: not applicable.

[c]*P*<.05 between before and after the intervention.

[d]*P*<.05 between-group changes.

[e]*P*<.001 between before and after the intervention.

[f]BMI was classified according to the World Health Organization categories: normal weight: <25 kg/m^2; overweight: 25 to <30 kg/m^2; obesity class 1: 30 to <35 kg/m^2; obesity class 2: 35 to <40 kg/m^2; and obesity class 3: ≥40 kg/m^2.

Dietary Intake of Key Foods and Nutrients

As shown in Table S2 in Multimedia Appendix 1, after 12 weeks of intervention, the Greenhabit group exhibited significant increases in the consumption of virgin olive oil (*P*<.001), nuts (*P*=.008), and legumes (*P*=.04). By contrast, the control group experienced a significant reduction in vegetable intake (*P*=.02), along with a significant increase in the consumption of refined cereals (*P*=.049), red wine (*P*=.001), and pastries, cakes, or sweets (*P*=.02). The 14-item Mediterranean Diet adherence score improved by 0.4 points in the intervention group after 12 weeks; however, this increase did not reach statistical significance (*P*=.16).

Regarding nutrient intake (Table S3 in Multimedia Appendix 1), the intervention group showed a significant increase in the consumption of total fat (*P*=.049) and monounsaturated fatty acids (*P*=.009). Between-group comparisons confirmed that the intervention group had higher consumption of total fat and monounsaturated fatty acids compared with the control group (*P*=.02 and *P*=.01, respectively).

Adherence to the Greenhabit Intervention

The data from the app revealed that users spent an average of 8 minutes per day online, with participants achieving an average of 372 challenges over the 12-week period. Based on the app data for each Greenhabit participant, dietitians offered personalized nutritional and lifestyle advice to address areas where participants scored lower.

As shown in Table 3 and Table S4 in Multimedia Appendix 1, the control group experienced a significant reduction in perceived social support (*P*<.001), along with a significant increase in anxiety (*P*=.03) and body pain perception (*P*=.049). By contrast, the intervention group demonstrated an improvement in social functioning.

XSL•FO
RenderX

Table 3. Changes in health-related quality of life at baseline and after 12 weeks of follow-up in the 2 study groups of the Greenhabit randomized controlled trial[a].

Variables	Greenhabit (n=50)	Control (n=53)	Differences between groups (Greenhabit vs control)
DUKE—perceived social support (points)			
Baseline	53.5 (22.6)	59.9 (24.3)	N/A[b]
Changes after 12 weeks	–5.3 (–11.8 to 1.3)	–15.2 (–22.2 to –8.1)[c]	7.56 (–3.94 to 19.06)
The Hospital Anxiety and Depression Scale—anxiety (points)			
Baseline	17.2 (1.8)	17.3 (2)	N/A
Changes after 12 weeks	0.2 (–0.2 to 0.6)	0.5 (0 to 1)[c]	–0.36 (–0.98 to 0.27)
36-Item Short Form—vitality			
Baseline	53 (16.8)	59.9 (16.2)	N/A
Changes after 12 weeks	0.5 (–2 to 2.9)[c]	0.3 (–2.1 to 2.8)	0.12 (–3.36 to 3.59)
36-Item Short Form—social functioning			
Baseline	74.4 (28.2)	86.0 (18.1)	N/A
Changes after 12 weeks	5.2 (1.5 to 9)[c]	–2.9 (–6.7 to 0.8)	8.14 (2.84 to 13.44)
36-Item Short Form—body pain			
Baseline	55.7 (34.8)	66.3 (24.8)	N/A
Changes after 12 weeks	0.8 (–3.7 to 5.3)	4.5 (0 to 9.1)[c]	–3.72 (–10.12 to 2.68)
Physical activity (metabolic equivalent of task minutes/week)			
Baseline	2020 (2390)	2370 (3995)	N/A
Changes after 12 weeks	540.7 (99.6 to 981.9)[c]	479.7 (0 to 959.3)[c]	61.08 (–590.6 to 712.75)
Walking (metabolic equivalent of task minutes/week)			
Baseline	1415 (1285)	1512 (1916)	N/A
Changes after 12 weeks	441.3 (54.2 to 828.3)[c]	370.6 (–44.8 to 785.9)	70.71 (–497.01 to 638.44)

[a]Values are expressed as mean (SDs) or mean differences (95% CI). Comparisons between groups with 1-way analysis of variance and comparisons between groups after 6 and 12 weeks of intervention with analysis of covariance adjusted for baseline levels of each variable are presented.

[b]N/A: not applicable.

[c]$P<.05$ between before and after the intervention.

Finally, physical activity significantly increased in both groups, but only the International Physical Activity Questionnaire test revealed significant changes in MET-walked minutes per day for participants in the intervention group ($P=.048$).

Data from the app are presented in Figure S1 in Multimedia Appendix 1. After 12 weeks, all participants in the Greenhabit group showed improvements in their quality of life: increase in energy (10/76, 13%), happiness (11/85, 13%), positivity (14/78, 18%), social engagement (7/81, 9%), and work-life balance (6/62, 10%).

Participant Retention

Notably, of the 61 participants, 46 (75%) used the app, with 38 using it daily and participating in the challenges offered. Among the 46 active participants, 15 exhibited intermittent use of the app. The data provided by the app revealed that the average daily time spent online was 8 minutes, and the average number of challenges completed during the 12 weeks was 372.

Discussion

Principal Findings

The primary goal of our RCT was to evaluate the effectiveness of the Greenhabit method in promoting self-management and self-care among patients aged 18-75 years recently diagnosed with T2D. Significant improvements in HbA_{1c} and FPG concentrations were observed in both the intervention and control groups at 6 and 12 weeks. However, notable between-group differences were evident in secondary outcomes, particularly in WC after 12 weeks. Additionally, the Greenhabit group showed significant improvements in BP, triglycerides, HDL-cholesterol, BW, HC, and overall quality of life following the 12-week intervention period.

Notably, the Greenhabit app proved more effective in significantly reducing anxiety, fostering positive emotions, improving coping with body pain, and enhancing participants'

XSL•FO
RenderX

perception of health. By contrast, the control group showed either no improvement or worsening trends in these areas.

Changes in T2D Parameters

Results indicate that telemedicine, particularly through the Greenhabit app, effectively reduced FPG and HbA_{1c} levels within 6 weeks of intervention, with these effects sustained at 12 weeks. However, no significant differences were found between the intervention and control groups. The app's visual representation of blood glucose values may have contributed to participant motivation.

Comparison with other studies showed similar HbA_{1c} reductions in the intervention group, but no significant differences between groups [46,47], suggesting that the app alone may not drive sufficient lifestyle changes, at least in the short term. The effectiveness may be enhanced through health care provider support and a personalized follow-up plan, which many patients may require to achieve their goals [48]. Several factors may explain the lack of differences in HbA_{1c} levels between groups as the primary outcome. The Hawthorne effect, where awareness of being observed influences behavior, may have led to improved self-management in both groups [49]. Additionally, baseline HbA_{1c} levels below 8% and the potential for increased attention from general practitioners during the RCT could have influenced the outcomes [50,51].

In patients with T2D, diet and exercise play a vital role in glycemic control and the prevention of cardiovascular risk [26,52-55]. However, conventional education may be insufficient to improve self-care, as patients with T2D often resist changing unhealthy behaviors [56]. This point will be explored further later.

It is important to highlight that the high costs associated with T2D are often due to the condition being accompanied by additional health-related complications. Moreover, health-related quality of life has been found to be significantly lower in individuals with T2D-related complications compared with those without such complications [57]. A 1% reduction in HbA_{1c} has been reported to significantly decrease microvascular and macrovascular complications in patients with T2D [58]. In our study, the intervention group showed a reduction in HbA_{1c} below 7%, which is associated with maximum cardiovascular benefits [59]. After 12 weeks, we hope participants will feel more aware, resilient, and energized, enabling them to maintain these habits long-term, though this would require further evaluation.

Changes in Cardiovascular Risk Factors

Remarkably, the mHealth intervention with Greenhabit led to improvements in cardiometabolic parameters, including systolic and diastolic BP, triglycerides, and HDL cholesterol, as well as adiposity indicators such as BMI, WC, and HC after the 12-week intervention. These findings are consistent with those of Hamaya et al [60], who reported significant reductions in BW and HbA_{1c} and an increase in HDL-C among 12,602 healthy participants (mean age 44.1 years, SD 10.2 years) using the Kencom mHealth app. A Chinese meta-analysis [61] and a recent systematic review [62] indicated that mHealth app interventions were associated with significant improvements in metabolic measures (HbA_{1c}, BW, BP, lipid profile, and physical activity) in patients with T2D. This suggests that mHealth apps, even when used by older individuals, can provide valuable support to enhance diabetes self-care and thereby improve clinical outcomes.

Cognitive and Behavioral Improvements

Participants in the intervention group reported a more positive outlook, improved coping with body pain, and favorable changes in their perception of health. These positive outcomes are attributed to the social support provided by the app environment, including the role of supportive peers and the broader social network. While no changes in social aspects were observed in the intervention group, the control group showed a significant decrease. Notably, there were differences between the groups in terms of anxiety, with a tendency to decrease in the intervention group and an increase in the control group. This aligns with previous findings suggesting that implementing support mechanisms can reduce anxiety [63,64].

The primary objective of this mHealth app targeting anxiety or depression is centered around cognitive-behavioral education. The Greenhabit method is based on educating individuals to cultivate enduring habits. By adopting a holistic approach that encompasses physical, mental, and social dimensions, it fosters resilience and self-awareness, thereby promoting the adoption of healthy habits. Moreover, fostering social engagement empowers individuals to take control of their health. Human interactions, particularly with medical physicians or health care professionals, along with the provision of social support and an appealing mHealth app interface, play pivotal roles in enhancing compliance, adherence, dropout rates, and treatment efficacy [12]. First, the combination of automatic information through the Greenhabit app, face-to-face visits with a dietitian, and regular phone calls during the 12-week intervention likely contributed to maintaining healthy behaviors such as diet. Furthermore, the critical support provided by nurses proved invaluable when participants encountered personal difficulties. The nurse served as a supporter, helping to motivate the patients. Interactions also created accountability, focus, and awareness of how behaviors impacted the participants' health. Thus, the combination of nursing assistance and the mHealth app demonstrated short-term effectiveness in enhancing diabetes self-efficacy and significantly increasing physical activity within the intervention group. These support mechanisms may explain the observed reductions in FPG and HbA_{1c} levels, as well as improvements in anxiety and general health status. The interaction with health care providers and the use of devices seem to enhance feelings of support for some individuals [65]. Second, while the influence of social support and its resultant outcomes may resemble those observed in conventional standards of care, a distinguishing characteristic of the digital social support offered by this mHealth app is the provision of anonymity and accessibility. The capability to exchange comparable lived experiences of anxiety or depression among peers, coupled with being accountable to an authoritative figure (eg, health care professional) for symptom monitoring or assessment, holds the potential to bolster engagement and

XSL•FO
RenderX

adherence to the mHealth app. Finally, the Greenhabit app demonstrates attributes of easy navigation, relatability, engagement, and aesthetic appeal, ensuring greater user engagement.

Adherence to the Intervention

The observation that not all participants consistently used the app (46/61) concurs with findings from Torbjørnsen et al [66], indicating varying acceptability of T2D self-management apps based on factors such as sex, age, education, expectations, and ease of use. Some participants viewed the app as more of a time-waster than a helpful tool, citing forgetfulness or excessive time requirements. Despite this, most participants expressed satisfaction (43/49, 88%) and considered the app useful (42/49, 86%), noting that it established new routines resulting in an improved lifestyle.

Influence of the mHealth App on Diabetes Self-Management and Treatment

The level of self-management may vary between individuals with T2D and those with type 1 diabetes, potentially influenced by treatment intensity and the necessity for regular blood glucose monitoring, particularly among insulin-dependent individuals [67]. To underline, most of our participants indicated that they were already self-monitoring their blood glucose levels at the outset of the study, implying a preexisting engagement in self-management practices regardless of insulin use.

To achieve a meaningful improvement in HbA_{1c} levels and cardiovascular risk factors in participants with T2D through self-management interventions, it is essential to first enhance healthy eating habits, promote engagement in physical activity, and ensure consistent adherence to medication. It is well-known that patients with T2D usually show higher resistance to changing their unhealthy behaviors, particularly when it comes to increasing physical activity and improving dietary habits [13,26]. It has been demonstrated that physical activity may contribute to a 30%-50% reduction in the development of T2D [68]. While both groups showed similar changes in T2D management strategies, including diet and physical activity, only the intervention group demonstrated a significant improvement in exercise. The control group did not show significant improvement in overall diet and displayed reduced adherence to the Mediterranean Diet. These findings are consistent with those of other studies [53-55].

Neither significant differences in dietary habits were noted between the intervention and control groups nor were significant changes observed in dietary habits after 12 weeks. This may be attributed to limited app usage, consistent with findings by Agarwal et al [11], and the inherent challenges in altering dietary habits influenced by external (food industry) and internal (long-term preferences) factors [69]. Additionally, participants using the Greenhabit app had to manually input data, which posed a barrier for some, leading to dropouts, including those aged 60 years and above and younger individuals who never used the app. A Norwegian RCT reported preferential app usage by older compared with younger participants [12].

However, the acquisition and maintenance of a healthy lifestyle, based on a healthy dietary pattern and physical activity programs, could be insufficient. As the intrinsic motivation of participants plays a key role, it may be recommended to include a motivational approach to bring about behavioral change [56]. Participants allocated to the Greenhabit group are more aware of what they eat and the daily exercise they perform, which may explain better control of the disease.

Chronic diseases carry severe consequences, including reduced quality of life, health emergencies, complications, and death [70]. Managing them poses significant challenges and costs for health care systems globally, with chronic diseases accounting for 71% of all deaths annually [71]. These diseases exacerbate inequality, disproportionately affecting socially disadvantaged individuals and hindering poverty reduction efforts [30]. Improving patients' health not only enhances their quality of life but also alleviates societal economic burdens. Effective management relies on understanding the impact of chronic diseases on patients' daily lives and empowering them to self-manage and make informed decisions [14]. This approach aligns with national health care strategies emphasizing personalized medicine and self-management [72]. Patient empowerment involves actively engaging in health responsibilities and promoting meaningful involvement in treatment decisions [70]. In this regard, mHealth emerges as a cost-efficient tool to support self-management and foster empowerment.

Limitations and Strengths

The study's primary strength lies in its design as an RCT, which allows for causal inference. Other strengths include a sizable and diverse sample (ages 18-75 years, gender-balanced), standardized data collection conditions for anthropometric and blood measurements, and the option for participants to involve a mate in app usage for mutual support. The app's holistic approach—addressing diet, physical activity, positive thinking, relaxation, and social aspects—distinguishes it from other mHealth strategies. This comprehensive strategy potentially explains its success in reducing cardiovascular risk factors. Our study stands as one of the few RCTs that specifically investigated the impact of using a mobile phone app as a digital diabetes diary to enhance quality of life.

Our study also has limitations. It was short-term (12 weeks) and conducted in a single center, which means findings may not be generalizable to non-T2D populations and could be subject to selection bias among patients with T2D. Additionally, technological challenges, particularly for older individuals, could hinder the app's use as a public health strategy. Furthermore, individuals from low-resource areas may struggle to access smartphone technology due to associated costs, in both acquiring the device and downloading paid apps. Additionally, selection bias occurred because participants were required to possess a smartphone. These financial barriers can limit access to mHealth interventions, exacerbating inequalities in health care access. The findings might not be applicable to cohorts with lower levels of motivation or limited technological knowledge. The observed increase in skill and technique acquisition may imply a heightened capability among participants to alleviate symptoms related to T2D and effectively manage their overall health. This enhancement may encompass

XSL•FO
RenderX

improved utilization of technical aids. Although our study did not incorporate a specific health education impact questionnaire, the integration of validated health-related quality-of-life assessments could provide valuable insights into the observed improvements in self-management abilities. Finally, while HbA_{1c} is crucial for long-term glycemic control, the exclusion of glycemic variability assessment limits a comprehensive understanding of glycemic status. Combining HbA_{1c} and glycemic variability could enhance diabetes management strategies [3]. More RCTs with longer follow-up times should be conducted to evaluate the long-term effect of diabetes-related mobile apps on glucose control and quality of life and to confirm that the outcomes seen in initial studies are sustainable over time.

Conclusions

Chronic illnesses can have significant consequences, often necessitating patients to depend on lifestyle choices and effective coping strategies for management. Self-management, which empowers patients to take control of their own health, is crucial. Studies on self-management highlight the importance of patient empowerment, enhancing their autonomy in health-related decisions. mHealth offers an economical means to enhance self-management and empower individuals. The integration of the Greenhabit app into clinical practice has the potential to positively impact glycemic control, reduce anxiety, and improve overall health status in individuals with T2D. Therefore, the Greenhabit app could be an additional supportive tool in the comprehensive management of patients with T2D within clinical settings.

The use of tailor-made interventions supported by an artificial intelligence–based infrastructure to enhance health and achieve sustainable behavior change represents a significant and promising approach for T2D self-management. However, larger studies are required to validate the effectiveness of these interventions.

Acknowledgments

We thank all the volunteers who participated in the study. We also express our gratitude to Giulia Maria Antonietta Trevisan for her assistance in the development of the project. We would also like to extend our sincere gratitude to the team of Logmeal, Marc Bolaños, and Eric Verdaguer for their dedicated support and collaboration on the project which was instrumental to the success of this study. CIBERobn is an initiative of the Instituto de Salud Carlos III, Spain. The Institut de Recerca en Nutrició i Seguretat Alimentaria (INSA-UB), University of Barcelona, Barcelona, Spain, is recognized as a Maria de Maeztu Unit of Excellence (grant CEX2021-001234-M), funded by MICIN/AEI/FEDER, UE. This work was fully supported by EIT Digital 2021 (21105). SCB received support from the University of Barcelona through the Margarita Salas postdoctoral fellowship. The funders had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or decision to submit the manuscript for publication.

Data Availability

Data availability for the Greenhabit trial is restricted due to signed consent agreements, which permit data sharing only with external researchers conducting studies aligned with the project's aims. Researchers interested in accessing the Greenhabit trial data used in this study may contact the corresponding author.

Authors' Contributions

RC, PR, and RE contributed to the conceptualization of the study. AMRL, RC, SCB, SAG, FCG, ES, and APG were responsible for the methodology and investigation. Validation, review, and editing of the manuscript were carried out by AMRL, RC, SCB, PR, and RE. Formal analysis and original draft preparation were conducted by AMRL, RC, and SCB. PR provided resources and secured funding. Data curation was performed by RC, SCB, and RE, while RC and RE supervised the study. RC and RE contributed equally to this work and should be considered co-corresponding authors.

Conflicts of Interest

RE reports receiving grants from the Fundación Dieta Mediterránea (Spain) and Cerveza y Salud (Spain), as well as personal fees for lectures from Brewers of Europe (Belgium), the Fundación Cerveza y Salud (Spain), Pernod Ricard (Mexico), Instituto Cervantes (Albuquerque, United States), Instituto Cervantes (Milan, Italy), Instituto Cervantes (Tokyo, Japan), Lilly Laboratories (Spain), and the Wine and Culinary International Forum (Spain). Additionally, RE received nonfinancial support for organizing a National Congress on Nutrition and for feeding trials with products from Grand Fountain and Uriach Laboratories (Spain). The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Additional analyses and statistics.
[DOCX File , 3810 KB - jmir_v27i1e58319_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

CONSORT (Consolidated Standards of Reporting Trials) - EHEALTH checklist (V.1.6.1).
[PDF File (Adobe PDF File), 408 KB - jmir_v27i1e58319_app2.pdf]

References

1. World Health Organization (WHO). Diabetes. WHO. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2024-01-15]
2. American Diabetes Association Professional Practice Committee. 4. Comprehensive medical evaluation and assessment of comorbidities: standards of medical care in diabetes—2022. Diabetes Care 2022 Jan 01;45(Suppl 1):S46-S59 [FREE Full text] [doi: 10.2337/dc22-S004] [Medline: 34964869]
3. Aljerf L, Alhaffar I. Salivary distinctiveness and modifications in males with diabetes and Behçet's disease. Biochem Res Int 2017;2017:9596202 [FREE Full text] [doi: 10.1155/2017/9596202] [Medline: 28321337]
4. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234 [FREE Full text] [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
5. Klösch M, Klösch C, Kundt FS, van der Zee-Neuen A, Dieplinger AM. eHealth systems for the optimised care of patients with type 2 diabetes. Br J Nurs 2020 Mar 12;29(5):274-278. [doi: 10.12968/bjon.2020.29.5.274] [Medline: 32167814]
6. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2022 Nov 01;45(11):2753-2786 [FREE Full text] [doi: 10.2337/dci22-0034] [Medline: 36148880]
7. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
8. Hemmingsen B, Gimenez-Perez G, Mauricio D, Roqué I Figuls M, Metzendorf MI, Richter B. Diet, physical activity or both for prevention or delay of type 2 diabetes mellitus and its associated complications in people at increased risk of developing type 2 diabetes mellitus. Cochrane Database Syst Rev 2017 Dec 04;12(12):CD003054 [FREE Full text] [doi: 10.1002/14651858.CD003054.pub4] [Medline: 29205264]
9. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
10. Rossi MC, Nicolucci A, Lucisano G, Pellegrini F, Di Bartolo P, Miselli V, et al. Impact of the "Diabetes Interactive Diary" telemedicine system on metabolic control, risk of hypoglycemia, and quality of life: a randomized clinical trial in type 1 diabetes. Diabetes Technol Ther 2013 Aug;15(8):670-679. [doi: 10.1089/dia.2013.0021] [Medline: 23844569]
11. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321. [doi: 10.2196/10321] [Medline: 30632972]
12. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57. [doi: 10.2196/mhealth.3882] [Medline: 25499872]
13. Bults M, van Leersum CM, Olthuis TJJ, Bekhuis REM, den Ouden MEM. Mobile health apps for the control and self-management of type 2 diabetes mellitus: qualitative study on users' acceptability and acceptance. JMIR Diabetes 2023 Jan 24;8:e41076. [doi: 10.2196/41076] [Medline: 36692927]
14. Öcal P. The use of mobile health applications in empowering self-management of type 2 diabetes: a literature review. Eur Res J 2024;10(1):127-135 [FREE Full text] [doi: 10.18621/eurj.1320822]
15. Whitehead LC, Crowe MT, Carter JD, Maskill VR, Carlyle D, Bugge C, et al. A nurse-led interdisciplinary approach to promote self-management of type 2 diabetes: a process evaluation of post-intervention experiences. J Eval Clin Pract 2017 Apr;23(2):264-271. [doi: 10.1111/jep.12594] [Medline: 27417302]
16. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35. [doi: 10.2196/mhealth.6522] [Medline: 28292740]
17. Wu IXY, Kee JCY, Threapleton DE, Ma RCW, Lam VCK, Lee EKP, et al. Effectiveness of smartphone technologies on glycaemic control in patients with type 2 diabetes: systematic review with meta-analysis of 17 trials. Obes Rev 2018 Jun;19(6):825-838. [doi: 10.1111/obr.12669] [Medline: 29345109]
18. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
19. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2020. Diabetes Care 2020 Jan;43(Suppl 1):S14-S31. [doi: 10.2337/dc20-S002] [Medline: 31862745]

XSL•FO
RenderX

20. Wahowiak L. Providing lifelong education and support: updates in the 2017 national standards for diabetes self-management education and support. Diabetes Spectr 2017 Nov;30(4):298-300. [doi: 10.2337/ds17-0060] [Medline: 29151720]
21. McKenna S, Jones F, Glenfield P, Lennon S. Bridges self-management program for people with stroke in the community: a feasibility randomized controlled trial. Int J Stroke 2015 Jul;10(5):697-704. [doi: 10.1111/ijs.12195] [Medline: 24256085]
22. Van de Velde D, De Zutter F, Satink T, Costa U, Janquart S, Senn D, et al. Delineating the concept of self-management in chronic conditions: a concept analysis. BMJ Open 2019 Jul 16;9(7):e027775 [FREE Full text] [doi: 10.1136/bmjopen-2018-027775] [Medline: 31315862]
23. Greenhabit. What is Greenhabit? Greenhabit. URL: https://greenhabit.nl/en/what-is-greenhabit/ [accessed 2024-01-15]
24. McClelland D, Koestner R, Weinberger J. How do self-attributed and implicit motives differ? Psychological Review 1989 Oct;96(4):690-702. [doi: 10.1037/0033-295X.96.4.690]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Centis E, Trento M, Dei Cas A, Pontiroli AE, De Feo P, Bruno A, et al. Stage of change and motivation to healthy diet and habitual physical activity in type 2 diabetes. Acta Diabetol 2014 Aug;51(4):559-566. [doi: 10.1007/s00592-013-0551-1] [Medline: 24442514]
27. Haas L, Maryniuk M, Beck J, Cox CE, Duker P, Edwards L, 2012 Standards Revision Task Force. National standards for diabetes self-management education and support. Diabetes Care 2013 Jan;36 Suppl 1(Suppl 1):S100-S108. [doi: 10.2337/dc13-S100] [Medline: 23264420]
28. Servei Català de la Salut (CatSalut). La Meva Salut. URL: https://catsalut.gencat.cat/ca/serveis-sanitaris/la-meva-salut/ [accessed 2024-01-15]
29. Greenhabit. URL: https://greenhabit.nl/en/ [accessed 2024-01-15]
30. McClelland DC. Human Motivation. Cambridge, UK: Cambridge University Press; 1987.
31. McClelland D. How motives, skills, and values determine what people do. American Psychologist 1985 Jul;40(7):812-825. [doi: 10.1037/0003-066X.40.7.812]
32. Covey S. De Zeven Eigenschappen van Effectief Leiderschap (68e druk). Amsterdam, The Netherlands: Business Contact; 2010.
33. Dyer W. Niet Morgen Maar nu. Utrecht, The Netherlands: Bruna uitgevers; 1997.
34. Wiebes H. Het oplichterssyndroom: 'de vloek van slimme mensen'. de Volkskrant. URL: https://www.volkskrant.nl/ [accessed 2024-01-15]
35. Juton C, Castro-Barquero S, Casas R, Freitas T, Ruiz-León AM, Crovetto F, et al. Reliability and concurrent and construct validity of a food frequency questionnaire for pregnant women at high risk to develop fetal growth restriction. Nutrients 2021 May 12;13(5):1629 [FREE Full text] [doi: 10.3390/nu13051629] [Medline: 34066238]
36. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011 Jun;141(6):1140-1145 [FREE Full text] [doi: 10.3945/jn.110.135566] [Medline: 21508208]
37. Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. Stat Med 2006 Feb 15;25(3):481-493. [doi: 10.1002/sim.2228] [Medline: 16287217]
38. Elosua R, Marrugat J, Molina L, Pons S, Pujol E. Validation of the Minnesota Leisure Time Physical Activity Questionnaire in Spanish men. The MARATHOM Investigators. Am J Epidemiol 1994 Jun 15;139(12):1197-1209. [doi: 10.1093/oxfordjournals.aje.a116966] [Medline: 8209878]
39. Elosua R, Garcia M, Aguilar A, Molina L, Covas MI, Marrugat J. Validation of the Minnesota Leisure Time Physical Activity Questionnaire in Spanish women. Investigators of the MARATDON Group. Med Sci Sports Exerc 2000 Aug;32(8):1431-1437. [doi: 10.1097/00005768-200008000-00011] [Medline: 10949009]
40. Vilagut G, Ferrer M, Rajmil L, Rebollo P, Permanyer-Miralda G, Quintana JM, et al. The Spanish version of the Short Form 36 Health Survey: a decade of experience and new developments. Gac Sanit 2005;19(2):135-150 [FREE Full text] [doi: 10.1157/13074369] [Medline: 15860162]
41. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
42. Pah AM, Bucuras P, Buleu F, Tudor A, Iurciuc S, Velimirovici D, et al. The importance of DS-14 and HADS questionnaires in quantifying psychological stress in type 2 diabetes mellitus. Medicina (Kaunas) 2019 Sep 05;55(9):569 [FREE Full text] [doi: 10.3390/medicina55090569] [Medline: 31491990]
43. Faghani S, Shamsalinia A, Ghaffari F, Yadegari N. The relationship between spiritual well-being and life orientation in elderly people with type 2 diabetes. Journal of Gerontology and Geriatrics 2018;66(3):142-148.
44. Cuéllar-Flores I, Dresch V. Validación del cuestionario de Apoyo Social Funcional Duke-UNK-11 en personas cuidadoras. Revista iberoamericana de diagnóstico y evaluación psicológica 2012;34(1):89-101 [FREE Full text] [doi: 10.6018/analesps.33.1.232341]
45. Sherwani SI, Khan HA, Ekhzaimy A, Masood A, Sakharkar MK. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. Biomark Insights 2016;11:95-104 [FREE Full text] [doi: 10.4137/BMI.S38440] [Medline: 27398023]

XSL•FO
RenderX

46. Arsand E, Tatara N, Østengen G, Hartvigsen G. Mobile phone-based self-management tools for type 2 diabetes: the few touch application. J Diabetes Sci Technol 2010 Mar 01;4(2):328-336 [FREE Full text] [doi: 10.1177/193229681000400213] [Medline: 20307393]
47. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
48. Davis J, Fischl AH, Beck J, Browning L, Carter A, Condon JE, et al. Erratum. 2022 National Standards for Diabetes Self-Management Education and Support. Diabetes Care 2022;45:484-494. Diabetes Care 2022 May 01;45(5):1298-1494 [FREE Full text] [doi: 10.2337/dc22-er05b] [Medline: 35231087]
49. Hawthorne K, Robles Y, Cannings-John R, Edwards AGK. Culturally appropriate health education for type 2 diabetes in ethnic minority groups: a systematic and narrative review of randomized controlled trials. Diabet Med 2010 Jun;27(6):613-623. [doi: 10.1111/j.1464-5491.2010.02954.x] [Medline: 20546277]
50. Kim EK, Kwak SH, Jung HS, Koo BK, Moon MK, Lim S, et al. The effect of a smartphone-based, patient-centered diabetes care system in patients with type 2 diabetes: a randomized, controlled trial for 24 weeks. Diabetes Care 2019 Jan;42(1):3-9. [doi: 10.2337/dc17-2197] [Medline: 30377185]
51. Adair J. The Hawthorne effect: a reconsideration of the methodological artifact. Journal of Applied Psychology 1984 May;69(2):334-345. [doi: 10.1037/0021-9010.69.2.334]
52. Balducci S, Zanuso S, Nicolucci A, De Feo P, Cavallo S, Cardelli P, Italian Diabetes Exercise Study (IDES) Investigators. Effect of an intensive exercise intervention strategy on modifiable cardiovascular risk factors in subjects with type 2 diabetes mellitus: a randomized controlled trial: the Italian Diabetes and Exercise Study (IDES). Arch Intern Med 2010 Nov 08;170(20):1794-1803. [doi: 10.1001/archinternmed.2010.380] [Medline: 21059972]
53. Boels AM, Rutten G, Zuithoff N, de Wit A, Vos R. Effectiveness of diabetes self-management education via a smartphone application in insulin treated type 2 diabetes patients - design of a randomised controlled trial ('TRIGGER study'). BMC Endocr Disord 2018 Oct 22;18(1):74 [FREE Full text] [doi: 10.1186/s12902-018-0304-9] [Medline: 30348142]
54. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
55. Rollo ME, Aguiar EJ, Williams RL, Wynne K, Kriss M, Callister R, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. Diabetes Metab Syndr Obes 2016;9:381-390 [FREE Full text] [doi: 10.2147/DMSO.S95247] [Medline: 27853384]
56. Green AJ, Bazata DD, Fox KM, Grandy S, SHIELD Study Group. Health-related behaviours of people with diabetes and those with cardiometabolic risk factors: results from SHIELD. Int J Clin Pract 2007 Nov;61(11):1791-1797 [FREE Full text] [doi: 10.1111/j.1742-1241.2007.01588.x] [Medline: 17887992]
57. Boye KS, Thieu VT, Lage MJ, Miller H, Paczkowski R. The association between sustained HbA1c control and long-term complications among individuals with type 2 diabetes: a retrospective study. Adv Ther 2022 May;39(5):2208-2221. [doi: 10.1007/s12325-022-02106-4] [Medline: 35316502]
58. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000 Aug 12;321(7258):405-412 [FREE Full text] [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
59. Wang P, Huang R, Lu S, Xia W, Sun H, Sun J, et al. HbA1c below 7% as the goal of glucose control fails to maximize the cardiovascular benefits: a meta-analysis. Cardiovasc Diabetol 2015 Sep 22;14:124 [FREE Full text] [doi: 10.1186/s12933-015-0285-1] [Medline: 26392171]
60. Hamaya R, Fukuda H, Takebayashi M, Mori M, Matsushima R, Nakano K, et al. Effects of an mHealth app (Kencom) with integrated functions for healthy lifestyles on physical activity levels and cardiovascular risk biomarkers: observational study of 12,602 users. J Med Internet Res 2021 Apr 26;23(4):e21622. [doi: 10.2196/21622] [Medline: 33900203]
61. Cai X, Qiu S, Luo D, Wang L, Lu Y, Li M. Mobile application interventions and weight loss in type 2 diabetes: a meta-analysis. Obesity (Silver Spring) 2020 Mar 26;28(3):502-509 [FREE Full text] [doi: 10.1002/oby.22715] [Medline: 31965748]
62. Sinclair A, Abdelhafiz A, Cukierman-Yaffe T. mHealth applications, older people and type 2 diabetes–a detailed review using systematic methodology. J Diabetes Res Ther 2019 Jan;5(1):1-5 [FREE Full text] [doi: 10.16966/2380-5544.141]
63. Hernández-Jiménez S, García-Ulloa AC, Bello-Chavolla OY, Aguilar-Salinas CA, Kershenobich-Stalnikowitz D, Group of Study CAIPaDi. Long-term effectiveness of a type 2 diabetes comprehensive care program. The CAIPaDi model. Diabetes Res Clin Pract 2019 May;151:128-137. [doi: 10.1016/j.diabres.2019.04.009] [Medline: 30954513]
64. Grossman JT, Frumkin MR, Rodebaugh TL, Lenze EJ. mHealth assessment and intervention of depression and anxiety in older adults. Harv Rev Psychiatry 2020;28(3):203-214 [FREE Full text] [doi: 10.1097/HRP.0000000000000255] [Medline: 32310834]
65. Harrison S, Stadler M, Ismail K, Amiel S, Herrmann-Werner A. Are patients with diabetes mellitus satisfied with technologies used to assist with diabetes management and coping?: A structured review. Diabetes Technol Ther 2014 Nov;16(11):771-783. [doi: 10.1089/dia.2014.0062] [Medline: 25069057]

XSL•FO
RenderX

66. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019 Sep 06;19(1):641 [FREE Full text] [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]
67. Marcolino MS, Maia JX, Alkmim MBM, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
68. Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. J Appl Physiol (1985) 2005 Sep;99(3):1193-1204. [doi: 10.1152/japplphysiol.00160.2005] [Medline: 16103522]
69. Hawkes C, Smith TG, Jewell J, Wardle J, Hammond RA, Friel S, et al. Smart food policies for obesity prevention. Lancet 2015 Jun 13;385(9985):2410-2421. [doi: 10.1016/S0140-6736(14)61745-1] [Medline: 25703109]
70. Stampe K, Kishik S, Müller SD. Mobile health in chronic disease management and patient empowerment: exploratory qualitative investigation into patient-physician consultations. J Med Internet Res 2021 Jun 15;23(6):e26991. [doi: 10.2196/26991] [Medline: 34128817]
71. World Health Organization (WHO). Noncommunicable diseases. WHO. Geneva, Switzerland: WHO; 2021 May 21. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2024-12-21]
72. Dadgar M, Joshi KD. The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. J Assoc Inf Syst 2018;19(2):86-112. [doi: 10.17705/jais1.00485]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
BP: blood pressure
BW: body weight
FPG: fasting plasma glucose
HbA1c: glycated hemoglobin
HC: hip circumference
HDL: high-density lipoprotein
IPAQ: International Physical Activity Questionnaire
mHealth: mobile health
RCT: randomized controlled trial
T2D: type 2 diabetes
WC: waist circumference

Edited by A Mavragani; submitted 12.03.24; peer-reviewed by L Aljerf, J Mistry; comments to author 01.04.24; revised version received 23.05.24; accepted 28.10.24; published 22.01.25.

Please cite as:
Ruiz-Leon AM, Casas R, Castro-Barquero S, Alfaro-González S, Radeva P, Sacanella E, Casanovas-Garriga F, Pérez-Gesalí A, Estruch R
Efficacy of a Mobile Health–Based Behavioral Treatment for Lifestyle Modification in Type 2 Diabetes Self-Management: Greenhabit Randomized Controlled Trial
J Med Internet Res 2025;27:e58319
URL: https://www.jmir.org/2025/1/e58319
doi:10.2196/58319
PMID:

©Ana Maria Ruiz-Leon, Rosa Casas, Sara Castro-Barquero, Sofia Alfaro-González, Petia Radeva, Emilio Sacanella, Francesc Casanovas-Garriga, Ainhoa Pérez-Gesalí, Ramon Estruch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Long-Term Impact of Using Mobile Phones and Playing Computer Games on the Brain Structure and the Risk of Neurodegenerative Diseases: Large Population-Based Study

Yi Xiao[1], MD; Sirui Zhang[1], MD; Yuanzheng Ma[1], MD; Shichan Wang[1], MD; Chunyu Li[1,2], MD; Yan Liang[1], MD; Huifang Shang[1,2], MD

[1]Department of Neurology, West China Hospital, Sichuan University, Chengdu, China

[2]Laboratory of Neurodegenerative Disorders, National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Huifang Shang, MD
Department of Neurology
West China Hospital, Sichuan University
No.17, Section 3, Renmin South Road
Chengdu
China
Phone: 86 18980602127
Email: hfshang2002@126.com

Abstract

Background: Despite the increasing popularity of electronic devices, the longitudinal effects of daily prolonged electronic device usage on brain health and the aging process remain unclear.

Objective: The aim of this study was to investigate the impact of the daily use of mobile phones/computers on the brain structure and the risk of neurodegenerative diseases.

Methods: We used data from the UK Biobank, a longitudinal population-based cohort study, to analyze the impact of mobile phone use duration, weekly usage time, and playing computer games on the future brain structure and the future risk of various neurodegenerative diseases, including all-cause dementia (ACD), Alzheimer disease (AD), vascular dementia (VD), all-cause parkinsonism (ACP), and Parkinson disease (PD). All the characteristics of using mobile phones and playing computer games were collected through face-to-face interviews at baseline, and outcomes were extracted from the algorithmic combinations of self-reported medical conditions, hospital admissions, and death registries. In addition, a group of participants underwent magnetic resonance imaging (MRI) at follow-up. Cox regression and linear regression were performed.

Results: The study included over 270,000 participants for risk analysis, with a mean baseline age of approximately 55.85 (SD 8.07) years. The average follow-up duration was approximately 13.9 (SD 1.99) years. Lengthy mobile phone use was associated with a reduced risk of ACD (2-4 years: hazard ratio [HR] 0.815, 95% CI 0.729-0.912, *P*<.001; 5-8 years: HR 0.749, 95% CI 0.677-0.829, *P*<.001; >8 years: HR 0.830, 95% CI 0.751-0.918, *P*<.001), AD (5-8 years: HR 0.787, 95% CI 0.672-0.922, *P*=.003), and VD (2-4 years: HR 0.616, 95% CI 0.477-0.794, *P*<.001; 5-8 years: HR 0.729, 95% CI 0.589-0.902, *P*=.004; >8 years: HR 0.750, 95% CI 0.605-0.930, *P*=.009) compared to rarely using mobile phones. Additionally, lengthy mobile phone use was linked to a decreased risk of ACP (5-8 years: HR 0.747, 95% CI 0.637-0.875, *P*<.001; >8 years: HR 0.774, 95% CI 0.663-0.904, *P*=.001) and PD (5-8 years: HR 0.760, 95% CI 0.644-0.897, *P*=.001; >8 years: HR 0.777, 95% CI 0.660-0.913, *P*=.002) in participants older than 60 years. However, higher weekly usage time did not confer additional risk reduction compared to lower weekly usage of mobile phones. The neuroimaging analysis involved 35,643 participants, with an average duration of approximately 9.0 years between baseline and neuroimaging scans. Lengthy mobile phone use was related to a thicker cortex in different areas of the brain.

Conclusions: Lengthy mobile phone use is associated with a reduced risk of neurodegenerative diseases and improved brain structure compared to minimal usage. Our research provides valuable background knowledge for future studies on the impact of modern electronic devices on brain health.

(J Med Internet Res 2025;27:e59663) doi:10.2196/59663

XSL•FO
RenderX

KEYWORDS

electronic device; parkinsonism; dementia; aging; brain MRI; magnetic resonance imaging

Introduction

Aging and digitalization are two key issues of the current era. The prevalence of neurodegenerative diseases is projected to increase with the aging of the population in the future [1]. Dementia and parkinsonism rank as the top two most prevalent neurodegenerative diseases, with Parkinson disease (PD) being the fastest growing among them [2]. Since disease-modifying treatments are still under research, early detection and prevention of neurodegenerative disease risk factors are crucial for reducing the disease burden on the health care system worldwide.

As digital technologies develop, the use of electronic devices is becoming increasingly popular, raising concerns about their impact on health. However, previous studies have reported controversial results. Electronic devices produce low-intensity electromagnetic fields and microwave radiation during operation, with mobile phones being particularly relevant due to their proximity to the head during use. Hypotheses have been proposed suggesting that low-intensity electromagnetic fields are associated with an increased risk and younger onset age of neurodegenerative diseases [3-5]. Conversely, some argue that exposure to low-intensity electromagnetic fields reduces the risk of Alzheimer disease (AD) [6]. Studies examining the association between mobile phone use and dementia risk have reported conflicting findings, with some suggesting a decreased risk, while others propose no association or even an increased risk [7,8]. Animal and in vivo studies have also yielded conflicting results, further complicating the understanding of the relationship between electronic device use and neurodegenerative diseases [9-11]. Studies on PD specifically have provided mixed evidence, with 1 study reporting no association, while another suggested a potential protective effect in males [8,12].

Overall, controversial hypotheses have been put forward by different studies, and there is a lack of convincing real-world evidence on this topic, especially in neuroepidemiology research. Furthermore, although changes in the brain structure and activity related to excessive phone use have been widely studied, less attention has been paid to the effects of daily electronic use [13,14]. For individuals who are not addicted to electronic devices, the impact of daily electronic use on the brain remains unclear. Compared to previous studies, this research further explored the details of the habit of electronic device use and its relationship with brain aging in participants of different ages.

To address the aforementioned research gaps, our study was designed to meet the following objectives: (1) the relationship between baseline usage of mobile phone/computers and the risk of developing neurodegenerative diseases and (2) the relationship between baseline daily usage of mobile phones/computers and the subsequent individual brain structure to explore the association between mobile phone/computer usage and brain aging. Clinical and imaging data were sourced from the UK Biobank. The findings of our study will offer valuable insights into the influence of electronic device usage on neurodegenerative processes.

Methods

Participants

The UK Biobank is a comprehensive longitudinal population-based cohort that enrolled 502,376 participants from 2006 to 2010 in the United Kingdom. In this study, we included participants who had baseline electronic device usage data and covariate data. The exclusion criteria were a history of the following diseases at baseline: dementia, stroke, PD, multiple sclerosis, demyelinating disorders, and other neurodegenerative diseases.

Ethical Considerations

The UK Biobank was approved (number 21/NW/0157) by the North West Multi-centre Research Ethics Committee (MREC). Consent were acquired from participants at the recruitment center for the study. Compensation and consent details are available on the UK Biobank website [15]. This manuscript was approved for publication by the UK Biobank, and data in the UK Biobank were accessible for health-related research with the approval of the UK Biobank Ethics Advisory Committee. All the data used in the study were anonymized, and no identifiable features were included.

Electronic Device Use

Data on the use of electronic devices were extracted from the UK Biobank database [16] and transferred into category variables: plays computer games (3 levels: never/rarely, sometimes, often) and length of mobile phone use (5 levels: never used a mobile phone at least once per week, 1 year or less, 2-4 years; 5-8 years, >8 years). For those who used mobile phones, the following data were extracted: weekly usage of mobile phones in the past 3 months (2 levels: <5 minutes, ≥5 minutes), hands-free device/speakerphone use with mobile phones in the past 3 months (2 levels: never/almost never, used), differences in mobile phone use compared to the previous 2 years (3 levels: no change, more frequent, less frequent), and usual side of the head for mobile phone use (3 levels: left, right, equally left and right). For details of the answer to each question and their form in the models, see Multimedia Appendix 1.

Neuroimaging

The details of the magnetic resonance imaging (MRI) performed can be found on the UK Biobank website and in previous studies [17-19]. In brief, participants were invited to attend the research centers for follow-up and further neuroimaging assessment from 2014. Imaging scans were conducted in 3 imaging centers with identical scanners and fixed platforms to minimize heterogeneity. Those who had a neurological disease before the imaging scan were excluded from the analysis.

Brain imaging was acquired with 3T Siemens Skyra (software platform VD13) using a 32-channel receive head coil. T1 images were collected at 1 mm isotropic resolution using 3D MPRAGE

XSL•FO
RenderX

acquisition, with the superior-inferior field of view being 256 mm. Several structural metrics were provided by the UK Biobank using Freesurfer. The most segmented brain imaging metrics were selected to obtain the most detailed changes. We used metrics processed by Destrieux (a2009s) parcellation in Freesurfer to assess the area, mean thickness, and volume of the brain in 148 areas [20].

Diffusion tensor imaging (DTI) is an advanced MRI technology that can detect microstructural abnormalities in white matter (WM) before the visual injury emerges. Diffusion data were collected with 2 b-values (1000 and 2000 s/mm^2) at 2 mm spatial resolution, with a multiband acceleration factor of 3. Fifty distinct diffusion-encoding directions were acquired for each diffusion-weighted shell. The diffusion preparation was a standard ("monopolar") Stejskal-Tanner pulse sequence. We used 5 DTI metrics provided by the UK Biobank: fractional anisotropy (FA), mean diffusion (MD), orientation dispersion (OD), intracellular volume fraction (ICVF), and isotropic (free) water volume fraction (ISOVF) [17]. FA and MD were calculated by fitting a voxel-wise diffusion tensor model through DTIFIT in FSL. The OD, ICVF, and ISOVF were calculated by fitting neurite orientation dispersion and density imaging models using accelerated microstructure imaging via convex optimization [21,22]. Low FA or high MD indicate a higher overall deficit in WM fiber integrity. Specifically, OD represents the overall coherence of fibers, the ICVF represents axonal/neurite density, and the ISOVF indicates the free-water fraction [22].

Diagnosis

We used the diagnosis of neurodegenerative diseases (dementia outcomes and parkinsonism outcomes) from category 47 and category 50: all-cause dementia (ACD), AD, vascular dementia (VD), all-cause parkinsonism (ACP), and PD. These outcomes were defined through algorithmic combinations of self-reported medical conditions, hospital admissions, and death registries [23].

Statistical Analysis

Data were described as mean (SD) values for continuous variables and counts (percentages) for category variables. Continuous variables were compared using the *t* test, and category variables were compared using the chi-square test or the Fisher exact test. We used Cox regression analysis to evaluate the risk of dementia. Participants were considered at risk for dementia or parkinsonism from baseline (2006-2010) and were followed up to the date of the first diagnosis, death, loss of follow-up, or the last hospital admission.

The covariates included age, sex, the BMI, smoking (ever or never), drinking (ever or never), hypertension, diabetes, education (college/university degree or lower degree), the polygenic risk score, socioeconomic status (Townsend deprivation index, a score calculated according to the output area in which participants' postcodes were located and the preceding national census output areas), the average total household income before tax, the mental health score, and physical activity. Using scaled Schoenfeld residuals to test proportional hazards, there was no indication of a violation of the hypothesis. In the sensitivity analysis, we further conducted several models: (1) model 1 excluded those diagnosed with outcome diseases within 5 years from baseline to exclude the potential reverse causality, (2) model 2 only included diagnoses that were from the hospital or death registration; (3) model 3 was used to replace the polygenic risk score (PRS) with the status of the *apolipoprotein ε4* gene, and (4) model 4 divided patients into 2 groups according to age (60 years old) at baseline to exclude the influence of age.

Linear regression models were used with variables using electronic devices as independent variables and neuroimaging metrics as dependent variables. Sex, age, ethnicity, the BMI, smoking, drinking, hypertension, diabetes, education, socioeconomic status, the WM hyperintensity volume, the whole WM volume, handedness (only for the usual use of the side of the mobile phone), and the average total household income before tax were included as covariates [24]. In the sensitivity analysis, we further adjusted the mental health score and physical activity. The variance inflation factor was tested and controlled under 5 to avoid multicollinearity. We used Bonferroni-corrected *P* values to determine the statistical significance: .05/5 for disease risk analysis, .05/(48 tracts×5 metrics) for DTI, and .05/148 regions for the surface, thickness, and volume in linear regression models. All analyses were performed using R version 4.1.2 (R Foundation for Statistical Computing).

Results

Demographic Characteristics of Participants Included in Risk Analysis

The inclusion flowchart is displayed in Figure 1 (detailed version in Multimedia Appendix 1), and the demographic characteristics of the participants are shown in Table 1. A total of 277,020 and 276,665 participants were included in the Cox regression model for dementia and parkinsonism, respectively. The mean age of the dementia and parkinsonism analysis groups at baseline was 55.85 (SD 8.07) years and 55.86 (SD 8.07) years, respectively. Half of the included participants (dementia group: n=139,166, 50%; parkinsonism group: n=139,276, 50%) were female. The mean duration of follow-up was 13.90 (SD 1.99) years and 13.89 (SD 2.01) years for the dementia and parkinsonism groups, respectively. We further compared the characteristics of the included and excluded participants (Tables S1 and S2 in Multimedia Appendix 1). The participants in the included risk analysis were younger, had higher education, a lower Townsend deprivation index, and a higher income compared to those excluded from the analysis.

XSL•FO
RenderX

Figure 1. Inclusion flowchart.

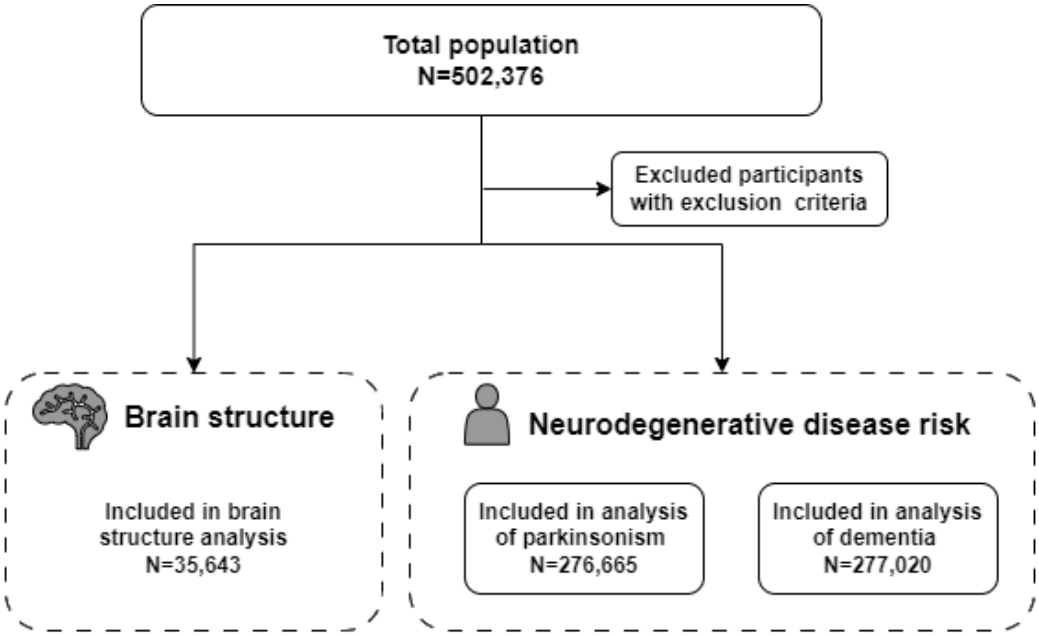

XSL•FO
RenderX

Table 1. Demographic characteristics of the included population (N=502,376).

Variables	Parkinsonism group (n=276,665)	Dementia group (n=277,020)
Usual side of the head for mobile phone use, n (%)		
Left	84,079 (30)	84,181 (30)
Right	139,275 (50)	139,429 (50)
Equally left and right	15,119 (6)	15,131 (6)
Do not need to answer	38,192 (14)	38,279 (14)
Weekly usage of mobile phones in the past 3 months, n (%)		
<5 minutes	46,861 (17)	46,939 (17)
≥5 minutes	191,612 (69)	191,802 (69)
Do not need to answer	38,192 (14)	38,279 (14)
Difference in mobile phone use compared to the previous 2 years, n (%)		
No change	124,344 (45)	124,496 (45)
More frequent	31,551 (11)	31,600 (11)
Less frequent	82,578 (30)	82,645 (30)
Do not need to answer	38,192 (14)	38,279 (14)
Plays computer games, n (%)		
Never/rarely	218,586 (79)	218,873 (79)
Sometimes	48,861 (18)	48,915 (18)
Often	9218 (3)	9232 (3)
Hands-free device/speakerphone use with mobile phones in the past 3 months, n (%)		
Never/almost never	191,825 (69)	192,057 (69)
Used	46,648 (17)	46,684 (17)
Do not need to answer	38,192 (14)	38,279 (14)
Length of mobile phone use, n (%)		
Never used a mobile phone at least once per week	38,192 (14)	38,279 (14)
≤1 year	6316 (2)	6325 (2)
2-4 years	44,446 (16)	44,507 (16)
5-8 years	85,160 (31)	85,263 (31)
>8 years	102,551 (37)	102,646 (37)
Age (years), mean (SD)	55.85 (8.07)	55.86 (8.07)
Sex, n (%)		
Female	139,166 (50)	139,276 (50)
Male	137,499 (50)	137,744 (50)
BMI, mean (SD)	27.29 (4.65)	27.29 (4.65)
Smoking, n (%)		
Yes	151,799 (55)	152,022 (55)
No	124,866 (45)	124,998 (45)
Drinking, n (%)		
Yes	9187 (3)	9198 (3)
No	267,478 (97)	267,822 (97)
Hypertension, n (%)		
Yes	69,764 (25)	69,863 (25)
No	206,901 (75)	207,157 (75)

XSL•FO
RenderX

Variables	Parkinsonism group (n=276,665)	Dementia group (n=277,020)
Diabetes, n (%)		
Yes	264,018 (95)	264,358 (95)
No	12,647 (5)	12,662 (5)
Education, n (%)		
College or higher	172,135 (62)	172,364 (62)
Lower than college	104,530 (38)	104,656 (38)
AD[a] polygenic risk score, mean (SD)	–0.14 (1.02)	0.05 (0.99)
Townsend deprivation index, mean (SD)	–1.50 (2.97)	–1.50 (2.97)
Average total household income before tax (US $), n (%)		
22,862-39,372	53,376 (19)	53,485 (19)
39,373-66,044	67,464 (24)	67,579 (24)
66,045-127,010	74,930 (27)	75,011 (27)
Do not know	62,972 (23)	63,009 (23)
>127,010	17,923 (6)	17,936 (6)
Mental health score, mean (SD)	4.31 (3.25)	4.31 (3.25)
Physical activity, n (%)		
Not meeting the criteria	41,256 (15)	41,183 (15)
Meeting the criteria	235,764 (85)	235,482 (85)

[a]AD: Alzheimer disease.

Relationship Between Electronic Device Use and Dementia

In the follow-up, we included 3007 (1.1%) participants with ACD, 1231 (0.4%) with AD, and 631 (0.2%) with VD. Lengthy mobile phone use (≥2 years) at baseline was related to a lower risk of ACD and VD compared to never using a mobile phone at least once per week (Table 2 and Table S3 in Multimedia Appendix 1). Additionally, participants who had used mobile phones for 5-8 years had a significantly lower risk of AD.

XSL•FO
RenderX

Table 2. Association between the length of mobile phone use and the risk of dementia.

Outcome and length of mobile phone use (years)	HR[a] (95% CI)	Outcomes, n (%)	Participants, n (%)	*Z* value	*P* value
ACD[b]					
Reference[c]	1.000 (1.000-1.000)	793 (2.1)	38,279 (14)	1.000	N/A[d]
≤1	1.002 (0.824-1.219)	115 (1.8)	6325 (2)	0.022	.98
2-4	0.815 (0.729-0.912)	510 (1.2)	44,507 (16)	–3.566	<.001[e]
5-8	0.749 (0.677-0.829)	750 (0.9)	85,263 (31)	–5.600	<.001[e]
>8	0.830 (0.751-0.918)	839 (0.8)	102,646 (37)	–3.630	<.001[e]
AD[f]					
Reference[c]	1.000 (1.000-1.000)	321 (0.8)	38,279 (14)	1.000	N/A
≤1	1.151 (0.860-1.540)	53 (0.8)	6325 (2)	0.945	.35
2-4	0.919 (0.774-1.091)	229 (0.5)	44,507 (16)	–0.968	.33
5-8	0.787 (0.672-0.922)	311 (0.4)	85,263 (31)	–2.973	.003[e]
>8	0.835 (0.711-0.979)	317 (0.3)	102,646 (37)	–2.215	.03
VD[g]					
Reference[c]	1.000 (1.000-1.000)	186 (0.5)	38,279 (14)	1.000	N/A
≤1	0.832 (0.539-1.284)	23 (0.4)	6325 (2)	–0.830	.41
2-4	0.616 (0.477-0.794)	89 (0.2)	44,507 (16)	–3.738	<.001[e]
5-8	0.729 (0.589-0.902)	163 (0.2)	85,263 (31)	–2.910	.004[e]
>8	0.750 (0.605-0.930)	170 (0.2)	102,646 (37)	–2.619	.009[e]

[a]HR: hazard ratio.

[b]ACD: all-cause dementia.

[c]Never used a mobile phone at least once per week.

[d]N/A: not applicable.

[e]Significant at $P<.05/5$.

[f]AD: Alzheimer disease.

[g]VD: vascular dementia.

To exclude the influence of reverse causality, several sensitivity analyses were conducted. After excluding participants who developed outcomes within 5 years after baseline, we found that lengthy mobile phone use was related to a lower risk of ACD (mobile phone use≥2 years), VD (mobile phone use=2-4 years), and AD (mobile phone use=5-8 years), as shown in Table S4 in Multimedia Appendix 1. After excluding participants whose source of outcomes was not from the hospital or death registration, we found that lengthy mobile phone use (≥2 years) was related to a lower risk of ACD and VD (Table S5 in Multimedia Appendix 1). To analyze the influence of *apolipoprotein ε4*, we adjusted the status of *apolipoprotein ε4* instead of the PRS and found that lengthy mobile phone use was related to a decreased risk of ACD (mobile phone use≥2 years) and VD (mobile phone use=2-4 years), as shown in Table S6 in Multimedia Appendix 1.

To exclude the influence of age, we conducted a separate analysis based on participants' age at baseline. We found that lengthy mobile phone use was related to a decreased risk of ACD (mobile phone use≥2 years) and AD (mobile phone use≥5 years) in participants younger than 60 years (Table S7 in Multimedia Appendix 1). Lengthy mobile phone use was also related to a decreased risk of ACD (mobile phone use≥2 years), AD (mobile phone use≥5 years), and VD (mobile phone use≥2 years) in participants older than 60 years (Table S8 in Multimedia Appendix 1).

Relationship Between Electronic Device Use and Parkinsonism

In the follow-up, we included 1578 (0.6%) participants with ACP and 1415 (0.5%) with PD. Often playing computer games was related to a lower risk of ACP (hazard ratio [HR] 0.604, 95% CI 0.425-0.859; *P*=.005) and PD (HR 0.574, 95% CI 0.392-0.841; *P*=.004) compared to never or rarely playing computer games (Table S9 in Multimedia Appendix 1). However, after excluding those who developed outcomes within 5 years, none of the above results remained statistically significant (Table S10 in Multimedia Appendix 1). We excluded participants whose source of outcomes was not from the hospital

XSL•FO
RenderX

or death registration and reached a similar result that often playing computer games was related to a lower risk of ACP and PD (Table S11 in Multimedia Appendix 1). After age stratification, in participants younger than 60 years, we found that the association between often playing computer games and a lower risk of ACP did not reach statistical significance (HR 0.393, 95% CI 0.175-0.844; *P*=.02; Table S12 in Multimedia Appendix 1). However, lengthy mobile phone use (≥5 years) was related to a decreased risk of ACP and PD in participants older than 60 years (Table S13 in Multimedia Appendix 1).

Demographic Characteristics of Participants Included in Neuroimaging Analysis

The demographic characteristics of participants who underwent neuroimaging analysis are displayed in Table S14 in Multimedia Appendix 1. A total of 35,643 participants were included, 48% (n=17,109) of them were male, and the mean age at baseline was 54.77 (SD 7.51) years. The mean duration between baseline and the neuroimaging scan was 8.97 (SD 1.73) years.

Relationship Between Electronic Device Use and the Brain Structure

Areas where the brain structure metrics at follow-up were related to the use of electronic devices at baseline are displayed in Tables 3 and 4. Using mobile phones for a long duration (≥2 years) was related to a thicker cortex in the left superior temporal sulcus. In addition, lengthy mobile phone use (≥8 years) was related to a significantly higher volume of the parahippocampal gyrus and thicker left orbital gyri, left precuneus, left superior temporal sulcus, right posterior-dorsal part of the cingulate gyrus, right middle temporal gyrus, and right subparietal sulcus (Table 3). Lengthy mobile phone use (2-4 years, ≥8 years) was related to higher mean FA in the pontine crossing tract on the FA skeleton (Table 4). Often playing computer games was related to a higher volume in the left anterior part of the cingulate gyrus and sulcus and the left triangular part of the inferior frontal gyrus, while sometimes playing computer games was related to a lower volume of the parahippocampal gyrus (Table 3). In addition, often holding a mobile phone on the right side of the head was related to larger areas in the bilateral posterior division of the middle temporal gyrus, bilateral anterior division of the cingulate gyrus, and a larger area of several regions compared to those who usually held their phones on the left side of the head (Table 3). Results of sensitivity analyses of neuroimaging are displayed in Tables S15-S17 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 3. Brain areas[a] where gray matter metrics were related to the use of electronic devices.

Brain areas	β (SE)	*T* value	*P* value
Mobile phone use>8years (reference: never used a mobile phone at least once per week)			
Mean thickness of G-orbital (left)	0.011 (0.003)	3.678	2.35E-04
Mean thickness of G-precuneus (left)	0.011 (0.003)	3.670	2.43E-04
Mean thickness of S-temporal-sup (left)	0.009 (0.002)	3.928	8.57E-05
Volume of G-oc-temp-med-parahip (left)	45.899 (12.370)	3.711	2.07E-04
Mean thickness of G-cingul-post-dorsal (right)	0.013 (0.003)	3.959	7.56E-05
Mean thickness of G-temporal-middle (right)	0.011 (0.003)	3.841	1.23E-04
Mean thickness of S-subparietal (right)	0.012 (0.003)	4.304	1.69E-05
Mobile phone use=5-8 years			
Mean thickness of S-temporal-sup (left)	0.009 (0.002)	3.779	1.57E-04
Area of G-rectus (right)	–3.918 (0.985)	–3.978	6.96E-05
Mean thickness of G-oc-temp-lat-fusifor (right)	0.013 (0.003)	3.667	2.45E-04
Mobile phone use=2-4 years			
Mean thickness of S-temporal-sup (left)	0.019 (0.005)	3.616	3.00E-04
Area of S-orbital-med-olfact (right)	–11.055 (3.020)	–3.661	2.52E-04
Weekly usage of mobile phones in the past 3 months (reference: <5 minutes)			
Area of S-collat-transv-post (right)	–3.934 (1.050)	–3.748	1.79E-04
Volume of S-collat-transv-post (right)	–10.803 (2.410)	–4.483	7.39E-06
Mobile phone held usually on the right side of the head (reference: left)			
Area of G+S-cingul-mid-ant (left)	6.742 (1.707)	3.950	7.83E-05
Area of G-cingul-post-dorsal (left)	4.535 (0.883)	5.135	2.84E-07
Area of G-cingul-post-ventral (left)	1.681 (0.463)	3.628	2.86E-04
Area of G-front-middle (left)	18.051 (4.991)	3.617	2.98E-04
Area of S-pericallosal (left)	8.164 (1.941)	4.206	2.61E-05
Area of S-temporal-sup (left)	20.326 (5.660)	3.591	3.30E-04
Volume of G-cingul-post-dorsal (left)	13.244 (3.587)	3.692	2.23E-04
Area of G+S-frontomargin (right)	3.896 (0.995)	3.915	9.05E-05
Area of G-pariet-inf-angular (right)	14.439 (3.724)	3.877	1.06E-04
Area of G-temporal-middle (right)	10.537 (2.926)	3.601	3.17E-04
Area of S-orbital-H-shaped (right)	5.502 (1.458)	3.774	1.61E-04
Area of S-parieto-occipital (right)	11.220 (2.885)	3.889	1.01E-04
Area of S-temporal-sup (right)	23.950 (5.787)	4.139	3.50E-05
Mean thickness of S-front-middle (right)	–0.006 (0.002)	–4.030	5.60E-05
Plays computer games often (reference: never/rarely)			
Volume of G+S-cingul-ant (left)	69.139 (19.050)	3.629	2.85E-04
Volume of G-front-inf-triangul (left)	59.338 (15.120)	3.925	8.70E-05
Plays computer games sometimes			
Volume of S-postcentral (left)	–45.258 (9.949)	–4.549	5.40E-06
Volume of G-oc-temp-med-parahip (right)	–33.829 (8.741)	–3.870	1.09E-04
Area of S-postcentral (left)	–18.319 (4.129)	–4.437	9.16E-06
Mean thickness of G-Ins-lg+S-cent-ins (left)	–0.014 (0.004)	–3.634	2.79E-04
Volume of G-Ins-lg+S-cent-ins (left)	–12.194 (3.148)	–3.873	1.08E-04

XSL•FO
RenderX

Brain areas	β (SE)	*T* value	*P* value
Volume of G-oc-temp-med-parahip (left)	–39.076 (9.719)	–4.021	5.81E-05

[a]The full names of the brain areas are listed in the literature [20].

Table 4. Brain areas where DTI[a] metrics were related to the use of electronic devices.

Brain areas	β (SE)	*T* value	*P* value
Mobile phone use>8years (reference: never used a mobile phone at least once per week)			
Mean FA[b] in the pontine crossing tract	0.003 (0.001)	4.995	5.90E-07
Mean OD[c] in posterior thalamic radiation (left)	–0.001 (0.000)	–3.873	1.25E-04
Mobile phone use=2-4 years			
Mean FA in the pontine crossing tract	0.003 (0.001)	3.914	9.11E-05
Mean OD in the fornix	–0.011 (0.003)	–3.954	7.70E-05
Weekly usage of mobile phones in the past 3 months (reference: <5 minutes)			
Mean OD in the cerebral peduncle (left)	0.001 (0.000)	3.953	7.75E-05
Mobile phone held usually on the right side of the head (reference: left)			
Mean FA in the posterior limb of the internal capsule (right)	0.001 (0.000)	3.741	1.84E-04

[a]DTI: diffusion tensor imaging.

[b]FA: fractional anisotropy.

[c]OD: orientation dispersion.

Overall, the findings of this study showed that electronic device use is associated with a reduced risk of neurodegenerative diseases (Table 5). This table summarizes the results of the main models but not the results of sensitivity analyses.

XSL•FO
RenderX

Table 5. Summary of the relationship between electronic device use, risk of disease, and brain structure.

Variables	Risk of ACD[a]	Risk of AD[b]	Risk of VD[c]	Risk of ACP[d]	Risk of PD[e]	Related to brain structure changes
Usual side of head for mobile phone use						
Left (reference)	—[f]	—	—	—	—	—
Right	No change	No change	No change	No change	No change	Yes
Equally left and right	No change	No change	No change	No change	No change	No
Weekly usage of mobile phones in the past 3 months						
<5 minutes (reference)	—	—	—	—	—	—
≥5 minutes	No change	No change	No change	No change	No change	Yes
Difference in mobile phone use compared to 2 years previously						
No change (reference)	—	—	—	—	—	—
more frequent	No change	No change	No change	No change	No change	No
Less frequent	No change	No change	No change	No change	No change	No
Plays computer games						
Never/rarely (reference)	—	—	—	—	—	—
Sometimes	No change	No change	No change	No change	No change	Yes
Often	No change	No change	No change	Decreased	Decreased	Yes
Hands-free device/speakerphone use with mobile phones in the past 3 months						
Never/almost never (reference)	—	—	—	—	—	—
Used	No change	No change	No change	No change	No change	No
Length of mobile phone use (years)						
Never used a mobile phone at least once per week (reference)	—	—	—	—	—	—
≤1	No change	No change	No change	No change	No change	No
2-4	Decreased	No change	Decreased	No change	No change	Yes
5-8	Decreased	Decreased	Decreased	No change	No change	Yes
>8	Decreased	No change	Decreased	No change	No change	Yes

[a]ACD: all-cause dementia.

[b]AD: Alzheimer disease.

[c]VD: vascular dementia.

[d]ACP: all-cause parkinsonism.

[e]PD: Parkinson disease.

[f]Not applicable.

Discussion

Principal Findings

In our study, we found evidence suggesting that the use of electronic devices is associated with a reduced risk of neurodegenerative diseases. Specifically, we observed that using mobile phones for 2 or more years is linked to a decreased risk of ACD and VD compared to never/rarely using mobile phones. Upon conducting sensitivity analyses, we confirmed the stability of the association between long mobile phone use and a lower risk of ACD and VD. Moreover, individuals aged 60 years and older appeared to benefit from extended duration of mobile phone use with a reduction in the risk of AD, ACP, and PD. Furthermore, our neuroimaging analysis revealed that various electronic device use habits, including the duration of mobile phone use, the frequency of computer use, and the preferred side of the head for holding mobile phones, were associated with distinct characteristics in the gray matter and WM structures of several brain regions over the follow-up period.

XSL•FO
RenderX

Comparison With Prior Work

Our findings suggest that using a mobile phone for more than 2 years is associated with a reduced risk of ACD and VD in the middle-aged population, with no significant impact of weekly mobile phone use. Previous studies using UK Biobank data have reported an association between mobile phone use and a lower risk of ACD in individuals over 60 years old [7]. Our study builds upon this knowledge by focusing on specific types of dementia and details of mobile phone use. We found that a long weekly time of mobile phone use (≥5 minutes) does not alter the risk of dementia compared to a shorter weekly time (<5 minutes). However, the duration of mobile phone use affects the association between mobile phone use and reduced risk of dementia. Using a mobile phone enriches the user's social and mental activities, potentially leading to improved cognition in later life [25,26]. The association between mobile phone use and a decreased risk of dementia is partly mediated by rich social support and flourishing leisure activities [7]. Additionally, prior research has suggested that registering a mobile phone decreases the risk of PD [8]. Our study further revealed that this association is observed only in populations aged 60 years and older, where a long duration of mobile phone use decreases the risk of developing ACP and PD compared to those who rarely use mobile phones.

The benefits of using electronic devices can be reflected in improved brain structure. We found that long durations of mobile phone use are associated with a thicker cortex in multiple areas of the brain. Specifically, individuals who use mobile phones for more than 8 years exhibit a higher volume of the parahippocampal gyrus, a region crucial for memory function [27]. This finding is particularly significant as atrophy in the parahippocampal gyrus is observed in the early stages of AD, serving as an early pathological biomarker [28,29]. Previous studies have shown that daily mobile phone use for less than 1 hour is associated with lower functional connectivity in certain brain regions, such as the superior temporal sulcus, and higher functional connectivity in others, like the cingulate [14]. In this study, we observed that only weekly mobile phone use of 5 minutes or more is associated with a decreased volume and area in the right posterior transverse collateral sulcus compared to less than 5 minutes of weekly use. However, no significant results were found in other areas or the DTI analysis. These findings suggest that daily mobile phone use, without reaching problematic levels, has minimal impact on the user's brain structure. Interestingly, we found inconsistent effects of occasional or frequent computer usage on the brain structure. Often playing computer games is associated with a higher brain volume, while sometimes playing computer games is related to a lower brain volume. For those who sometimes play computers, the related shortcomings, such as sedentariness, are more prominent and result in cortical atrophy [30].

Strengths and Limitations

This study has several strengths. First, we adjusted for covariates that may have influenced the results using comprehensive data from the UK Biobank, which is the largest population-based cohort in the world. This allowed us to provide important information on the use of electronic devices. Second, the cohort's follow-up duration was notably extensive, allowing sufficient time for patients to reach the point at which diseases developed. Additionally, the duration between baseline and neuroimaging scans offered crucial insights into the long-term effects of electronic device use on brain health.

There was 1 limitation that should be mentioned. The mobile phone use in this study was limited to receiving calls. The rapid advancement and evolution of electronic devices in recent years have transformed the usage patterns of mobile phones, heightening the potential for dependency [31,32]. However, the evolution happened years after the launch of the UK Biobank cohort. Our study still provides important insights into the long-term influence of electronic device use. The time that people spend on mobile phones now is much longer than it was decades before [31,33]. However, this study found that there are no additional benefits of longer weekly usage. In addition, playing games on the mobile phone is popular nowadays [33]. We found an association between playing computer games and decreased neurodegenerative disease risk, but the association is unclear if game devices change from computers to mobile phones. Future studies should record the details of the daily time and the type of function of mobile phone use to analyze the impact of modern electronic devices on brain health.

Conclusion

In conclusion, our study revealed that individuals with long duration of mobile phone use are associated with a decreased risk of ACD and VD compared to those who rarely use mobile phones. The protective effect of mobile phone use is also observed in individuals older than 60 years, who experience a reduced risk of AD, ACP, and PD. However, spending more time on mobile phones per week may not further reduce the risk of neurodegenerative disease. Our findings also indicate that electronic device use can impact the brain structure. Specifically, using a mobile phone and often playing computer games are associated with a better brain structure, while sometimes playing computer games is linked to a poorer brain structure. These results highlight the complex relationship between electronic device use and neurodegenerative disease risk, as well as the influence on the brain structure. Future studies should concentrate on the mechanism underlying the influence of using mobile phones and computers and new electronic devices on health.

Acknowledgments

We are grateful to the UK Biobank, an open-access biomedical database. This research was conducted using the UK Biobank Resource (application 98992). Data analyzed in the study were accessed from the UK Biobank database [16].

XSL•FO
RenderX

This work was funded by the Chengdu Science and Technology Project Key Research and Development Support Program (2024-YF08-00038-GX), the National Science Fund of Sichuan Province (2023NSFSC1576), and the National Natural Science Foundation of China (grant 82101485).

Data Availability

All code is available upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic characteristics of participants and association between the risk of neurodegenerative diseases and electronic device use.
[DOCX File , 328 KB - jmir_v27i1e59663_app1.docx]

References

1. Hou Y, Dan X, Babbar M, Wei Y, Hasselbalch SG, Croteau DL, et al. Ageing as a risk factor for neurodegenerative disease. Nat Rev Neurol 2019 Oct;15(10):565-581. [doi: 10.1038/s41582-019-0244-7] [Medline: 31501588]
2. Erkkinen MG, Kim M, Geschwind MD. Clinical neurology and epidemiology of the major neurodegenerative diseases. Cold Spring Harb Perspect Biol 2018 Apr 02;10(4):a033118 [FREE Full text] [doi: 10.1101/cshperspect.a033118] [Medline: 28716886]
3. Pall ML. Low intensity electromagnetic fields act via voltage-gated calcium channel (VGCC) activation to cause very early onset Alzheimer's disease: 18 distinct types of evidence. Curr Alzheimer Res 2022;19(2):119-132 [FREE Full text] [doi: 10.2174/1567205019666220202114510] [Medline: 35114921]
4. Pritchard C, Silk A, Hansen L. Are rises in electro-magnetic field in the human environment, interacting with multiple environmental pollutions, the tripping point for increases in neurological deaths in the Western World? Med Hypotheses 2019 Jun;127:76-83 [FREE Full text] [doi: 10.1016/j.mehy.2019.03.018] [Medline: 31088653]
5. Hardell L, Sage C. Biological effects from electromagnetic field exposure and public exposure standards. Biomed Pharmacother 2008 Feb;62(2):104-109. [doi: 10.1016/j.biopha.2007.12.004] [Medline: 18242044]
6. Shirbandi K, Khalafi M, J Bevelacqua J, Sadeghian N, Adiban S, Bahaeddini Zarandi F, et al. Exposure to low levels of radiofrequency electromagnetic fields emitted from cell-phones as a promising treatment of Alzheimer's disease: a scoping review study. J Biomed Phys Eng 2023 Feb;13(1):3-16 [FREE Full text] [doi: 10.31661/jbpe.v0i0.2109-1398] [Medline: 36818013]
7. Zhao H, Wen Q, Zhuo L, Wang S, Zhan S. Association between mobile phone use and incidence of dementia: a prospective cohort study using the UK Biobank. Gerontology 2023;69(10):1232-1244. [doi: 10.1159/000531847] [Medline: 37494916]
8. Schüz J, Waldemar G, Olsen JH, Johansen C. Risks for central nervous system diseases among mobile phone subscribers: a Danish retrospective cohort study. PLoS One 2009;4(2):e4389 [FREE Full text] [doi: 10.1371/journal.pone.0004389] [Medline: 19194493]
9. Arendash GW, Sanchez-Ramos J, Mori T, Mamcarz M, Lin X, Runfeldt M, et al. Electromagnetic field treatment protects against and reverses cognitive impairment in Alzheimer's disease mice. J Alzheimers Dis 2010;19(1):191-210. [doi: 10.3233/JAD-2010-1228] [Medline: 20061638]
10. Bouji M, Lecomte A, Gamez C, Blazy K, Villégier A. Impact of cerebral radiofrequency exposures on oxidative stress and corticosterone in a rat model of Alzheimer's disease. J Alzheimers Dis 2020 Jan 21;73(2):467-476. [doi: 10.3233/jad-190593]
11. Söderqvist F, Hardell L, Carlberg M, Mild KH. Radiofrequency fields, transthyretin, and Alzheimer's disease. J Alzheimers Dis 2010;20(2):599-606. [doi: 10.3233/JAD-2010-1395] [Medline: 20164553]
12. van der Mark M, Vermeulen R, Nijssen PCG, Mulleners WM, Sas AMG, van Laar T, et al. Extremely low-frequency magnetic field exposure, electrical shocks and risk of Parkinson's disease. Int Arch Occup Environ Health 2015 Feb;88(2):227-234. [doi: 10.1007/s00420-014-0949-2] [Medline: 24939428]
13. Lin H, Chang Y, Chen M, Liu S, Chen B, Li L, et al. Structural and functional neural correlates in individuals with excessive smartphone use: a systematic review and meta-analysis. Int J Environ Res Public Health 2022 Dec 05;19(23):16277 [FREE Full text] [doi: 10.3390/ijerph192316277] [Medline: 36498362]
14. Rolls ET, Feng R, Feng J. Lifestyle risks associated with brain functional connectivity and structure. Hum Brain Mapp 2023 Apr 15;44(6):2479-2492 [FREE Full text] [doi: 10.1002/hbm.26225] [Medline: 36799566]
15. UK Biobank research ethics approval. UK Biobank. URL: https://www.ukBiobank.ac.uk/learn-more-about-uk-Biobank/about-us/ethics [accessed 2024-12-10]
16. UK Biobank research analysis platform. Enabling scientific discoveries that improve human health. UK Biobank. URL: https://ukBiobank.dnanexus.com [accessed 2024-12-10]

XSL•FO
RenderX

17. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 2016 Nov;19(11):1523-1536 [FREE Full text] [doi: 10.1038/nn.4393] [Medline: 27643430]
18. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature 2018 Oct 10;562(7726):210-216 [FREE Full text] [doi: 10.1038/s41586-018-0571-7] [Medline: 30305740]
19. Chandra A, Dervenoulas G, Politis M, Alzheimer's Disease Neuroimaging Initiative. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. J Neurol 2019 Jun;266(6):1293-1302 [FREE Full text] [doi: 10.1007/s00415-018-9016-3] [Medline: 30120563]
20. Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 2010 Oct 15;53(1):1-15 [FREE Full text] [doi: 10.1016/j.neuroimage.2010.06.010] [Medline: 20547229]
21. Daducci A, Canales-Rodríguez EJ, Zhang H, Dyrby TB, Alexander DC, Thiran J. Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. Neuroimage 2015 Jan 15;105:32-44 [FREE Full text] [doi: 10.1016/j.neuroimage.2014.10.026] [Medline: 25462697]
22. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 2012 Jul 16;61(4):1000-1016. [doi: 10.1016/j.neuroimage.2012.03.072] [Medline: 22484410]
23. Algorithmically-defined outcomes (ADOs). Version 2.0. UK Biobank. 2022 Jan. URL: https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=460 [accessed 2024-12-10]
24. Sliz E, Shin J, Ahmad S, Williams DM, Frenzel S, Gauß F, NeuroCHARGE Working Group. Circulating metabolome and white matter hyperintensities in women and men. Circulation 2022 Apr 05;145(14):1040-1052 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.121.056892] [Medline: 35050683]
25. Yates LA, Ziser S, Spector A, Orrell M. Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis. Int Psychogeriatr 2016 Nov;28(11):1791-1806. [doi: 10.1017/S1041610216001137] [Medline: 27502691]
26. Liu J, Peng J, Chen M, Zhang T. Mediating and moderating effects of internet use on urban-rural disparities in health among older adults: nationally representative cross-sectional survey in China. J Med Internet Res 2023 Sep 28;25:e45343 [FREE Full text] [doi: 10.2196/45343] [Medline: 37768721]
27. Bohbot VD, Allen JJB, Dagher A, Dumoulin SO, Evans AC, Petrides M, et al. Role of the parahippocampal cortex in memory for the configuration but not the identity of objects: converging evidence from patients with selective thermal lesions and fMRI. Front Hum Neurosci 2015;9:431 [FREE Full text] [doi: 10.3389/fnhum.2015.00431] [Medline: 26283949]
28. Van Hoesen GW, Augustinack JC, Dierking J, Redman SJ, Thangavel R. The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates. Ann N Y Acad Sci 2000 Jun;911:254-274. [doi: 10.1111/j.1749-6632.2000.tb06731.x] [Medline: 10911879]
29. Echávarri C, Aalten P, Uylings HBM, Jacobs HIL, Visser PJ, Gronenschild EHBM, et al. Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease. Brain Struct Funct 2010 Oct 19;215(3-4):265-271. [doi: 10.1007/s00429-010-0283-8] [Medline: 20957494]
30. Maasakkers CM, Thijssen DH, Knight SP, Newman L, O'Connor JD, Scarlett S, et al. Hemodynamic and structural brain measures in high and low sedentary older adults. J Cereb Blood Flow Metab 2021 Oct;41(10):2607-2616 [FREE Full text] [doi: 10.1177/0271678X211009382] [Medline: 33866848]
31. Casey M, Wigginton C, Calugar-Pop C. Deloitte's 2019 global mobile consumer survey. Tracking consumers' digital behavior around the world. Deloitte. 2019 Nov 10. URL: https://www2.deloitte.com/us/en/insights/industry/telecommunications/global-mobile-consumer-survey-2019.html [accessed 2024-12-10]
32. Olson JA, Sandra DA, Colucci É, Al Bikaii A, Chmoulevitch D, Nahas J, et al. Smartphone addiction is increasing across the world: a meta-analysis of 24 countries. Comput Hum Behav 2022 Apr;129:107138. [doi: 10.1016/j.chb.2021.107138]
33. Global mobile consumer survey. US edition. Deloitte. URL: https://www2.deloitte.com/tr/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html [accessed 2024-12-10]

Abbreviations

ACD: all-cause dementia
ACP: all-cause parkinsonism
AD: Alzheimer disease
DTI: diffusion tensor imaging
FA: fractional anisotropy
HR: hazard ratio
ICVF: intracellular volume fraction
ISOVF: isotropic (free) water volume fraction

XSL•FO
RenderX

MD: mean diffusion
MRI: magnetic resonance imaging
OD: orientation dispersion
PD: Parkinson disease
VD: vascular dementia
WM: white matter

Edited by A Schwartz; submitted 18.04.24; peer-reviewed by Y Yan, S Mary, Y Ling; comments to author 16.08.24; revised version received 08.10.24; accepted 27.11.24; published 28.01.25.

Please cite as:
Xiao Y, Zhang S, Ma Y, Wang S, Li C, Liang Y, Shang H
Long-Term Impact of Using Mobile Phones and Playing Computer Games on the Brain Structure and the Risk of Neurodegenerative Diseases: Large Population-Based Study
J Med Internet Res 2025;27:e59663
URL: https://www.jmir.org/2025/1/e59663
doi:10.2196/59663
PMID:

©Yi Xiao, Sirui Zhang, Yuanzheng Ma, Shichan Wang, Chunyu Li, Yan Liang, Huifang Shang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cross-Platform Ecological Momentary Assessment App (JTrack-EMA+): Development and Usability Study

Mehran Sahandi Far[1,2], PhD; Jona M Fischer[1], BSc; Svea Senge[3], MD; Robin Rathmakers[3], MD; Thomas Meissner[3], Prof Dr; Dominik Schneble[3], MD; Mamaka Narava[1], MSc; Simon B Eickhoff[1,2], Prof Dr; Juergen Dukart[1,2], Prof Dr

[1]Research Centre Jülich, Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7), Jülich, Germany

[2]Medical Faculty, Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]Department of General Paediatrics, Neonatology and Paediatric Cardiology, Medical Faculty, University Hospital Düsseldorf, Heinrich-Heine-University, Düsseldorf, Germany

Corresponding Author:
Mehran Sahandi Far, PhD
Research Centre Jülich
Institute of Neuroscience and Medicine
Brain and Behaviour (INM-7)
Wilhelm-Johnen-Straße
Jülich, 52428
Germany
Phone: 49 17636977109
Email: m.sahandi.far@fz-juelich.de

Abstract

Background: Traditional in-clinic methods of collecting self-reported information are costly, time-consuming, subjective, and often limited in the quality and quantity of observation. However, smartphone-based ecological momentary assessments (EMAs) provide complementary information to in-clinic visits by collecting real-time, frequent, and longitudinal data that are ecologically valid. While these methods are promising, they are often prone to various technical obstacles. However, despite the potential of smartphone-based EMAs, they face technical obstacles that impact adaptability, availability, and interoperability across devices and operating systems. Deficiencies in these areas can contribute to selection bias by excluding participants with unsupported devices or limited digital literacy, increase development and maintenance costs, and extend deployment timelines. Moreover, these limitations not only impede the configurability of existing solutions but also hinder their adoption for addressing diverse clinical challenges.

Objective: The primary aim of this research was to develop a cross-platform EMA app that ensures a uniform user experience and core features across various operating systems. Emphasis was placed on maximizing the integration and adaptability to various study designs, all while maintaining strict adherence to security and privacy protocols. JTrack-EMA+ was designed and implemented per the FAIR (findable, accessible, interpretable, and reusable) principles in both its architecture and data management layers, thereby reducing the burden of integration for clinicians and researchers.

Methods: JTrack-EMA+ was built using the Flutter framework, enabling it to run seamlessly across different platforms. This platform comprises two main components. JDash (Research Centre Jülich, Institute of Neuroscience and Medicine, Brain and Behaviour [INM-7]) is an online management tool created using Python (Python Software Foundation) with the Django (Django Software Foundation) framework. This online dashboard offers comprehensive study management tools, including assessment design, user administration, data quality control, and a reminder casting center. The JTrack-EMA+ app supports a wide range of question types, allowing flexibility in assessment design. It also has configurable assessment logic and the ability to include supplementary materials for a richer user experience. It strongly commits to security and privacy and complies with the General Data Protection Regulations to safeguard user data and ensure confidentiality.

Results: We investigated our platform in a pilot study with 480 days of follow-up to assess participants' compliance. The 6-month average compliance was 49.3%, significantly declining (P=.004) from 66.7% in the first month to 42% in the sixth month.

Conclusions: The JTrack-EMA+ platform prioritizes platform-independent architecture, providing an easy entry point for clinical researchers to deploy EMA in their respective clinical studies. Remote and home-based assessments of EMA using this

XSL•FO
RenderX

platform can provide valuable insights into patients' daily lives, particularly in a population with limited mobility or inconsistent access to health care services.

(J Med Internet Res 2025;27:e51689) doi:10.2196/51689

KEYWORDS

digital biomarkers; mobile health; remote monitoring; smartphone; mobile phone; monitoring; biomarker; ecological momentary assessment; application; costly; user experience; data management; mobility

Introduction

Self-reported information in psychiatric disorders plays a crucial role in clinical decision-making. They are often gathered through traditional questionnaires or verbal conversations during clinic visits [1]. These visits are infrequent, costly, and inaccessible to everyone [2]. The self-reports from in-clinic visits may also be subject to higher recall bias, notably when asked to recall memories and past events [3].

The need to reduce reliance on traditional in-clinic visits and the widespread adoption of smart devices (such as smartphones and smartwatches) have increased their potential for use in clinical settings. Health-related data collected from these devices are often called digital biomarkers (DBs) [4-6]. DBs include a wide range of measurements, with ecological momentary assessments (EMAs) being one of the most widely accepted and used [7]. EMA involves using experience sampling methods, which are brief assessments administered in real time in naturalistic settings to study the moment-to-moment fluctuations of states and behaviors [8]. EMA offers several advantages over traditional in-clinic visits, including a higher sampling frequency, the ability to collect longitudinal data, enhanced engagement, lower costs, and increased ecological validity [1,9]. Smartphones are often equipped with a range of sensors such as location, pedometer, proximity, and luminance sensors. These sensors can be used to trigger surveys based on different physical activities or environmental factors. For instance, a survey can be automatically triggered when the user visits new locations or there is a change in weather conditions in the user's location. This type of EMA is commonly referred to as event-contingent EMA [10,11].

Assessment questions, supporting material (eg, images, videos, and explanatory text), managing assessment logic (ie, scheduling, inclusion or exclusion criteria, validating input answers, recording timestamps, and storing and transferring assessments), and providing assessment completion reminders [12] are among the main features typically integrated into EMA apps. These advanced capabilities offer an excellent opportunity for collecting DBs in remote assessments. However, their implementation comes with several technical challenges and obstacles. These challenges include but are not limited to privacy and security, memory and battery usage, patient acceptance, scalability, and interoperability with different devices and operating systems. Although there are various apps intended to meet many of the abovementioned criteria [13-17], interoperability with different operating systems, adaptability, and configurability are still lacking. Several operating systems are available for mobile (Android and iOS) and desktop (Windows and macOS) devices. Each of these has its own paradigms and programming languages. Therefore, separate developers and experts are required for each, significantly increasing the cost and development time. Additionally, restricting the apps to a specific operating system may lead to a selection bias. A platform-independent architecture addresses these limitations. This motivated us to introduce JTrack-EMA+ as a cross-platform, configurable app that also complies with the General Data Protection Regulation (GDPR). It consists of an EMA app and an online management tool facilitating assessment management and data quality control. It is integrated into our previously introduced sensor-based DBs ecosystem, "JTrack" [18]. In this paper, we also evaluated a pilot study using JTrack-EMA+ in a population of parents with a newborn to investigate participants' compliance with the assessment protocols.

Methods

Study Design

Here, we describe the primary component of the JTrack platform (Figure 1), which includes a smartphone-based app JTrack-EMA+ (Figure 2A-H), and an online management app, JDash (Figures 3A-E and 4A-E). JTrack-EMA+ was developed based on platform-independent architecture using the Flutter framework [19]. Flutter is a cross-platform framework developed by Google that allows developers to build native apps from a single codebase. There are different cross-platform technologies such as React Native (Meta Platforms, Inc) and Xamarin (Microsoft). However, Flutter has several advantages that make it an irresistible choice for JTrack-EMA+. Flutter uses its rendering engine, which provides better performance and a more consistent look and feel across platforms compared to React Native, which relies on native components. Flutter's hot reload feature enables faster iteration and testing during development, improving the overall development workflow and Flutter's widget-based approach to user interface development promotes a consistent and cohesive user experience across platforms. Additionally, Flutter has a more active and growing community, with strong support from Google, ensuring continued development and improvement of the framework.

Platform-independent design reduces development and maintenance time, provides a consistent performance and user experience, and enables instant delivery of new feature updates and enhancements across multiple platforms.

XSL•FO
RenderX

Figure 1. JTrack platform's main structure and its working principle. The flow of data packages between JDash and the JTrack app and the flow of push notifications between servers and the app are shown here. API: application programming interface; EMA: ecological momentary assessment.

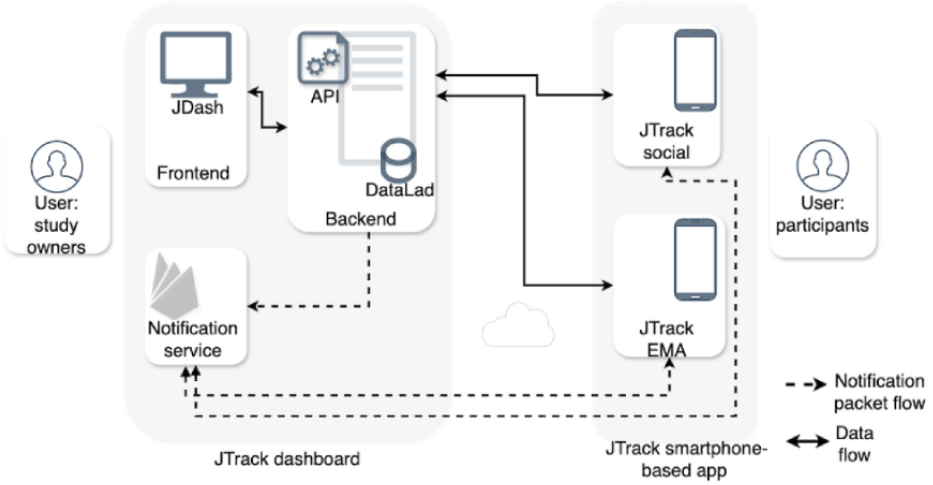

Figure 2. JTrack-EMA+ smartphone app user interface. (A) Initial screen to scan a QR code used to join a study; (B) list view page to see all available questions; (C) binary question with supporting image; (D) time and date question; (E) multiple selection question; (F) user numeric input question; (G) sliding question; and (H) main menu to access leaving, manual synchronization, and setting pages. EMA: ecological momentary assessment.

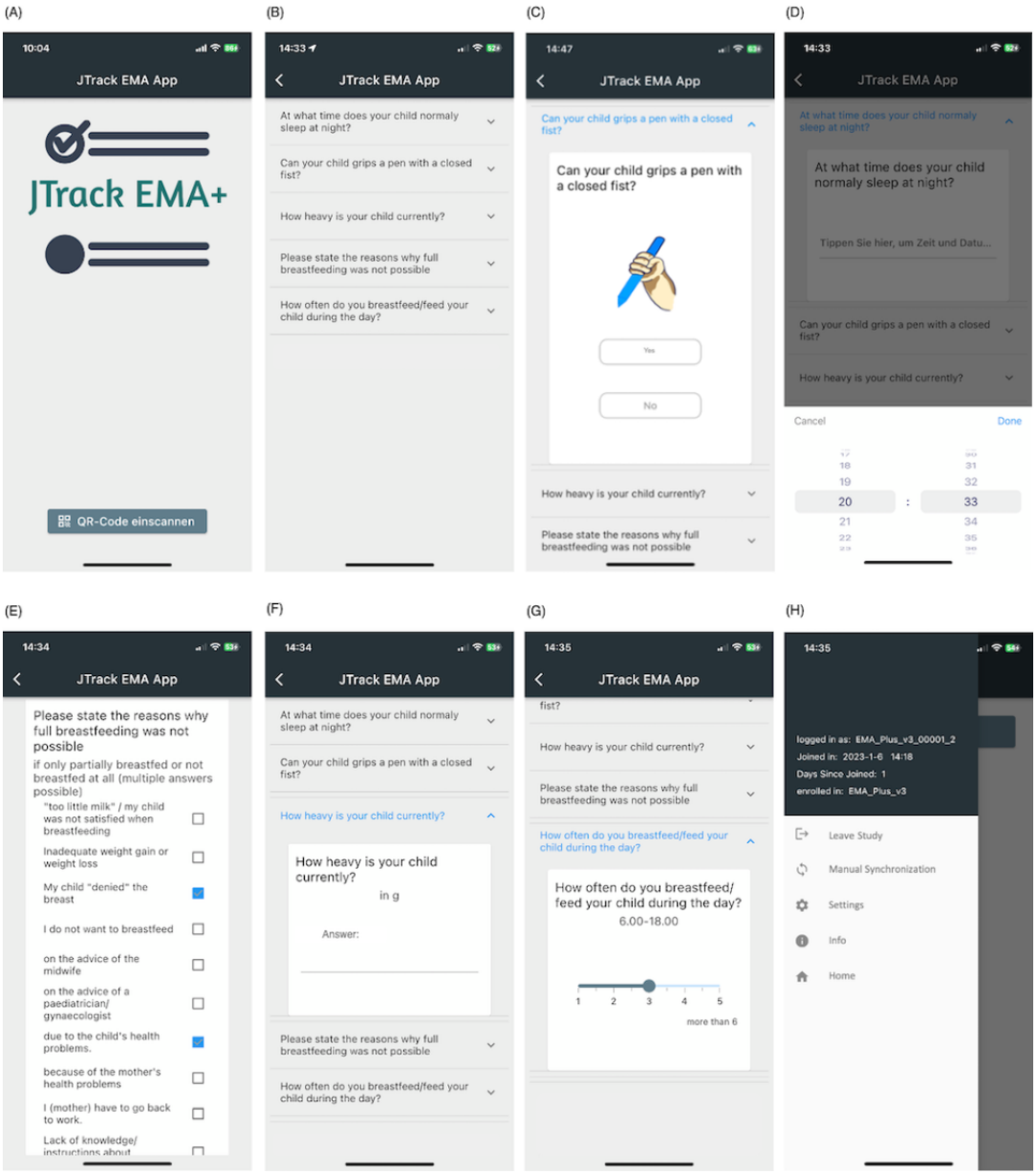

XSL•FO
RenderX

Figure 3. JDash user interface. (A) Display general information about the current study; (B) manage and send push notifications to participants; (C) remove a participant from the study; (D) display the main screen, which visualizes the supporting information for enrolled subjects in the study—it also color codes the status of participants (ie, left the study, have missing data, etc); and (E) create a new study, and add questions list or images and details of the study. EMA: ecological momentary assessment.

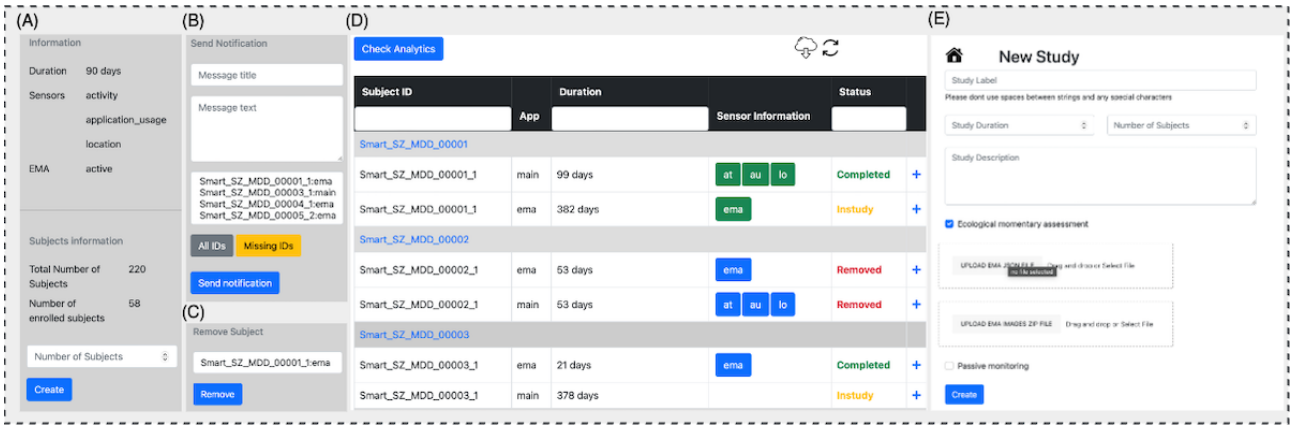

Figure 4. JDash EMA creator user interface, integrated into JDash. (A) Create a new assessment, (B) display the main screen of the assessment and related information, (C) add a new question to the assessment, (D) add answers to the question from the previous step, and (E) modify and add logic to the question. EMA: ecological momentary assessment.

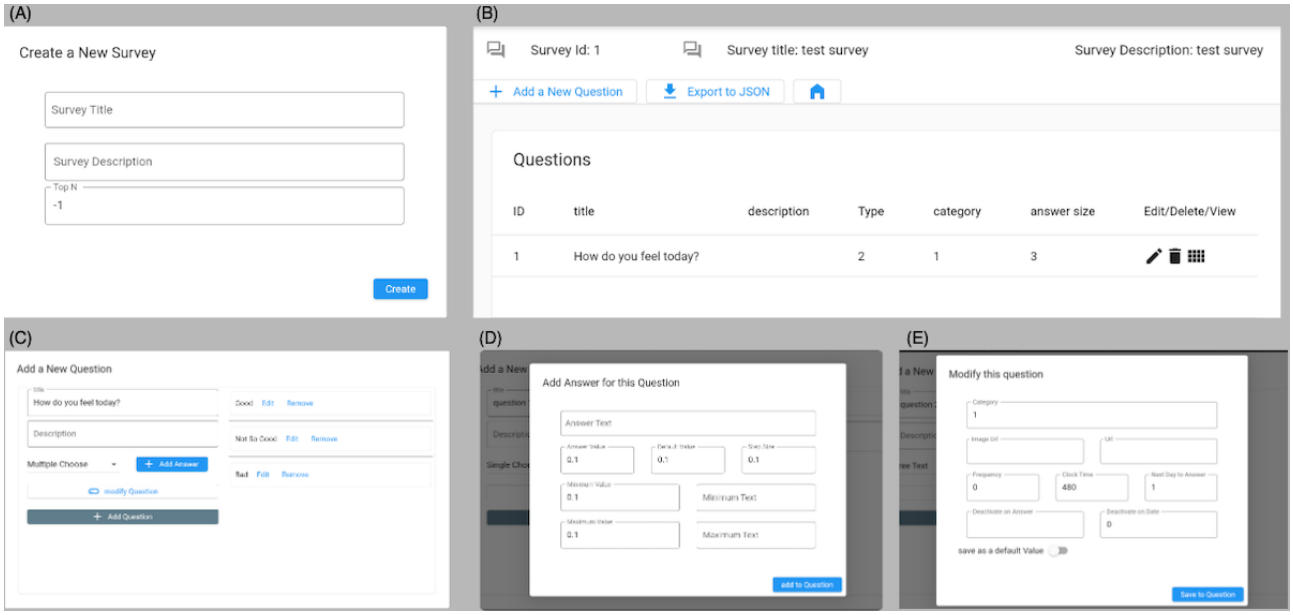

Adaptability and Configurability

The behavior and performance of software are significantly impacted by rapidly evolving operating systems. Being able to adapt to these changes in agile is a crucial trait that the platform-independent architecture of JTrack-EMA+ provides. Besides architectural benefits, JTrack-EMA+ constructs evaluation materials and logic using a JSON file to facilitate reusability and configurability.

The use of a JSON file providing the configurations for each specific study can simplify the process of reusability in an app by centralizing configuration fields in a single file. This approach allows for efficient management and modification of app settings without the need to alter the codebase. By storing configuration details such as logging credentials, user and study IDs, and survey-related data in a JSON file, study owners can quickly update these settings as needed without modifying the app code. This approach enhances flexibility, scalability, and maintainability by separating configuration from the app logic. It makes it easier to reuse the same configuration across different environments or instances of the app. Using such a configuration file approach, study owners can modify the app settings from the study management interface without further need to alter the app codebase, making it easier to manage and maintain.

After users register with the provided QR codes, the assessment is generated based on a study-specific JSON file received from the server. Additionally, assessments can be updated remotely using this file without updating the app. The browser-based study management dashboard JDash has a built-in EMA creator to generate these JSON files making the design and configuration of an EMA study quite simple and versatile (Figure 4).

JTrack-EMA+ supports a wide range of assessment formats, such as binary, multiple-choice, text, or numeric user input; date and time; and sliding questions (Figure 2). The system has a flexible assessment logic that includes conditions, categories,

XSL•FO
RenderX

and customizable frequency on a question-by-question basis. This functionality is vital in determining whether to ask or skip certain questions, based on predetermined criteria such as the time elapsed since the assessment began, the present time of day, and the value or date of the participant's response (conditional questions). By categorizing questions, different conditions can also be applied within each category (ie, only show the first n questions for each category). Additionally, the system allows for defining the frequency at which the questions should be displayed to users (such as on-demand, daily, or weekly).

JTrack-EMA+ supports two distinct paging styles: page view and list view. With page view, participants will see a single question at a time, whereas list view shows all questions as a list, allowing participants to expand (choose) a question to answer. While page view ensures the visibility of each question, list view is especially helpful in situations where many questions are shown or only specific questions are relevant.

Data Quality, Privacy, and Security

Quality, security, and privacy of the collected data are important considerations when designing a novel EMA app. To address the data quality, syntactic input validation is deployed to ensure that only properly formed data is entered. Additionally, to prevent possible data loss, we adopted the "local-first" data storage method, where data are stored locally on the device and only synced to the server when connected to the internet. In addition, to mitigate the impact of potential interruptions (eg, by other programs, battery drain, or phone restart) on saving data, we implemented an auto-saving mechanism.

To ensure data privacy and security and to comply with the European Union GDPR, JTrack-EMA+ deployment does not require any identifiers or personal data. All data are assigned to a JDash-generated user ID allowing for fully anonymized and pseudonymized data collection. Secure channels are used for communication between servers and devices, and data integrity is protected against unintended corruption and potential interruptions using message-digest algorithms. Moreover, the distribution of the app through official channels (ie, Google Play Store and Apple App Store) provides convenience to the participants and creates a sense of security and confidentiality.

QR code-based registration, embedding control, and authentication methods streamline the registration process (these codes are generated when a study is created in JDash). In addition, we have provided a "leave study" option in the app, which enables all participants to withdraw from the study at any time.

Study Management

JDash is an online study management app designed to assist JTrack study owners in creating new studies and managing ongoing ones (Figure 3A-E). It allows study owners to create and modify all study aspects, such as the duration, addition of assessment materials (eg, images), data quality checks, removal of participants from a study, ending of a study, and management of the study owner's access level. Python (Django and Web Server Gateway Interface) was chosen as the implementation technology. JDash consists of a user interface and a backend. While the backend app programming interface serves as an end point for receiving, saving, and authentication requests arriving from the app, the frontend interface can also be used to create and design an assessment (Figure 1).

The JDash was developed to be easy to use and deploy. It incorporates flawless design principles to provide a range of features for clinicians to establish their assessments intuitively. The design principles used in the app ensure that clinicians can navigate and use the assessment tools efficiently, without further help from a technical team, enhancing the overall user experience. By focusing on such design principles, the app aims to streamline the assessment process for clinicians, making it user-friendly, clear, and consistent (eg, similar design patterns). This approach aligns with best user interface design and usability practices, ultimately benefiting users by simplifying their workflow and improving the effectiveness of their assessments.

In addition to an open-source software license, JDash uses an automated deployment model that is efficient, easy to use, and requires fewer configurations (for instruction on deployment, readers may refer to the documentation on our GitHub page [20]). In addition, JDash has a comprehensive notification system to ensure periodic reminders based on a predetermined schedule, avoiding unwanted (eg, overnight) alerts. This notification is designed to operate as a push notification (online) that can be sent to participants' devices from the dashboard. This module is integrated into JDash to deliver reminders to participants' devices, either individually or based on their status (ie, no data sent in a few days). This regular reminder has the potential to minimize the high dropout rate of EMA-based assessments [9].

Finally, most research programs aim to share data with colleagues or along with research results to promote transparency, encourage collaboration, and accelerate research. To facilitate this goal, we have used DataLad [21] as a data management infrastructure on our servers to manage data according to FAIR (findable, accessible, interpretable, and reusable) principles [22]. DataLad offers essential features such as data versioning, metadata handling, structured formatting, and change tracking under an open-source license which contributes to enhanced security and privacy of collected data. This integration in the data management layer also accelerates reproducible science, a key element of scientific methods.

Pilot Study

We deployed the JTrack-EMA+ app in a pilot study involving 179 parents with newborns followed up for 480 days. Parents were asked to install the JTrack app on their phones (75 iOS and 104 Android). The pilot study includes weekly assessments of children's physical and mental development. These questions are designed to appear based on their defined frequency and the time since participants joined the study. Furthermore, the yes or no questions deactivate when the answer is "Yes." To illustrate the development questions, a picture was shown for each question. Martin Pawlusiak designed the pictures. Additionally, an appropriate notification time was chosen to avoid receiving assessment reminders early in the morning or late at night.

XSL•FO
RenderX

We have been conducting the pilot study since November 2021, and it is set to run until November 2023. As of now, we have released the preliminary compliance data that we have gathered over the first 14 months of the study. We calculated the compliance rate as a percentage of completed surveys compared to what was expected. In total, 65 participants completed assessments for at least 6 months and were included in the compliance analysis. Monthly compliance was computed as the number of received data points divided by the number of expected data points. Monthly average compliance was compared using the Friedman test, followed by pairwise comparisons using the Wilcoxon rank sum test.

Ethical Considerations

The studies involving human participants were reviewed and approved by the ethics committee at the medical faculty, Heinrich-Heine-Universität Düsseldorf, Germany (2021-1500_1). The patients or participants provided written informed consent to participate in this study. The data were anonymized. Participants did not receive any compensation for this study.

Results

Figure 5 displays the fluctuation of average compliance rates across the 6 months of the pilot study. The overall average compliance rate across this period was 49.3%. The between-group comparisons of the average compliance rate show a significant decrease from 66.7% in the first month to 46.9% (P=.01), 42.6% (P<.001), 42.9% (P=.002), and 42% (P=.004) in the third, fourth, fifth, and sixth months, respectively.

Figure 5. Compliance rate for an exemplary study. Each dot represents a participant with a corresponding compliance rate (calculated as a percentage of completed surveys out of expected surveys), and a solid line is the average compliance for a corresponding month with a 95% CI. *P<.05, **P<.01.

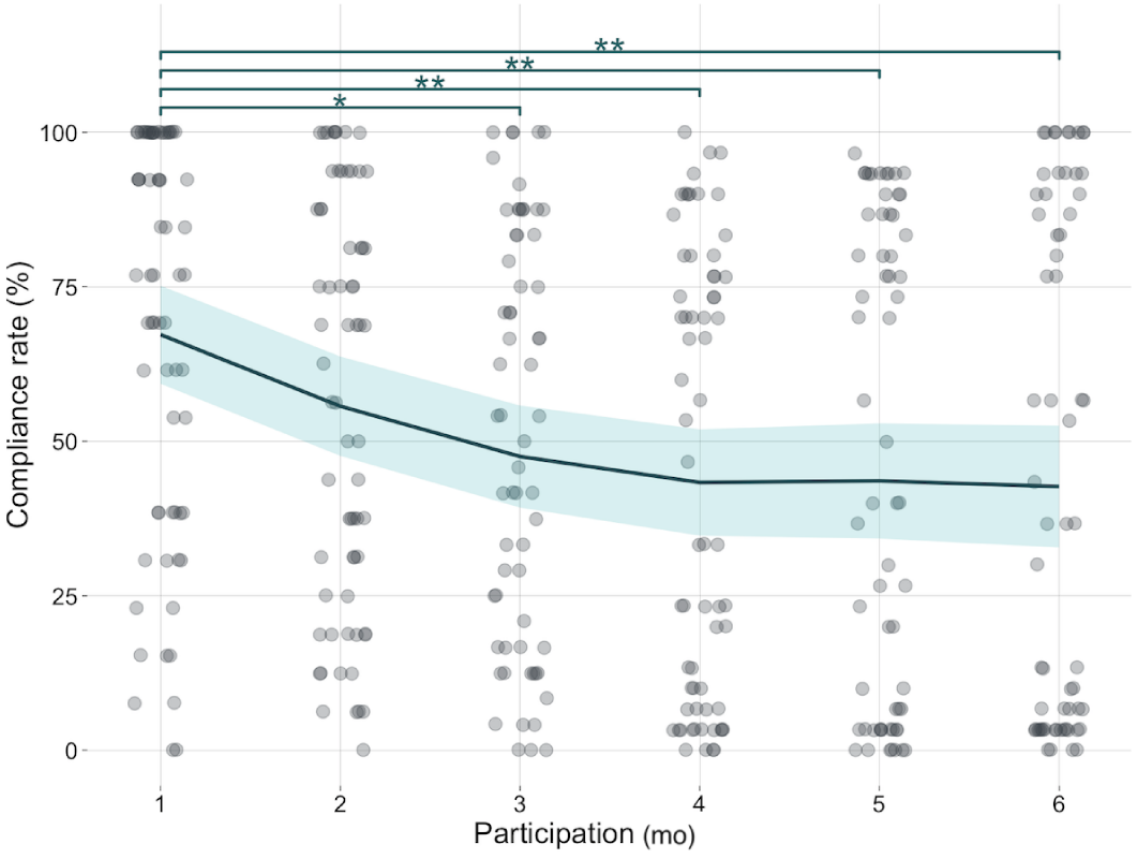

Discussion

Principal Findings

Here, we introduced JTrack-EMA+, a smartphone-based, real-time EMA app. This platform-independent architecture of the app reduces the disparity in user experiences and functionality across different operating systems. It complies with the GDPRs and official app distribution channels while providing all features required for designing and deploying EMA-based evaluations.

The online dashboard serves as a central study management interface. It visualizes and summarizes the study information (ie, duration, the number of subjects, and supplementary information). This information is used for data streaming and quality checks and to interact with participants in case of missing or invalid data via push notifications. Researchers can control the access to the specific information of the study to their research partners through the user access level control provided by the dashboard. The open-source solution DataLad [21] is used for data management in JDash, and the abovementioned functionalities reduce reliance on third-party apps and services such as Amazon Web Services.

The traditional development architecture used by existing platforms is limited by its shortcomings [13-17]. A lack of

XSL•FO
RenderX

support and disparities in user experience and app functionality across platforms are among the most significant contributors. Therefore, most EMA platforms are either available for a single operating system (ie, ULTEMAT is only available for Android [16], and iHabit is only available for iOS [17]) or have different features depending on the operating system (ie, AWARE, Remote Assessment of Disease and Relapse–base, and SEMA3 have two distinct apps for each platform [13,15]).

To the best of our knowledge, the JTrack-EMA+ stands out as a pioneer in prioritizing platform-independent architecture. This platform offers extensive configurability, which sets it apart from the existing platform. It also provides a flexible and dynamic environment instead of being limited to a single problem-oriented solution.

The effectiveness of EMA depends on the participants' motivation and adherence to assessment protocols [12,23]. Poor compliance and missing data can adversely impact the quality of the collected data and introduce bias in the sample, thereby hindering the success of clinical apps. Our results revealed an overall average of 49.3% with a significant decrease in compliance rate after 2 months. Although there is a large degree of heterogeneity in compliance rates reported in various studies [24], considering our inclusion time constraint (6 months) and the target demographic (parents with busier routines who dedicate more time to their newborn), this compliance rate can be considered an acceptable rate. Importantly, there is no definitive recommendation for a minimum or appropriate rate of compliance as it is highly dependent on the specific patient or participant population and the actual frequency, duration, and comprehensiveness of the specific assessments, as well as the actual research question. As long as sufficient representation of the individual day-to-day variability in the respective EMA assessments is achieved, the respective EMA measures could be considered useful. Overall, our results demonstrate the potential to use our platform compared to the paper-and-pencil EMA assessment (which was shown to have a significantly lower compliance rate), particularly in participants with limited mobility or restricted access to health care services [24].

A limitation of our platform is its ability to detect and respond to various events and contexts, such as environmental and physical activities. Although we consider this a limitation here, it is carried out on purpose to simplify the ethics and data privacy required by sensory and contextual assessment; required in addition to EMA. Additionally, despite the attempts to keep performance and user experience consistent across all operating systems, background work constraints in iOS may influence the app's internal reminders on these devices, so automated push notifications implemented by JDash can be a better choice for such devices.

Finally, JTrack-EMA+ is an active project regularly updated to add new features and improve the platform's functionality. The features described here are part of the version 1 release. Therefore, later versions may have additional functions which will be documented with the respective releases. The most updated JTrack-EMA+ versions are released via the official distribution channels (ie, Google's Play Store and Apple's App Store). All study management functionalities to deploy JTrack-EMA+ are implemented as open source and can be accessed from our GitHub repository under an open-source license.

Conclusions

In this paper, we covered the key concepts associated with EMA apps and the technical constraints of existing platforms. Then we launched a platform-independent JTrack-EMA+ app to reduce the gap in user experience and app functionality across different operating systems while adhering to security, privacy, and GDPR requirements.

In a pilot study, we also demonstrated participants' compliance rate with assessment protocols and how it fluctuated over a 6-month course of study. We conclude that remote and at-home assessment using this platform results in an acceptable rate of compliance that may not be possible using the conventional in-clinic visit method, particularly in a population with limited mobility or a busy schedule—as were the participants in our pilot study.

Incorporating EMA apps into clinical practice can significantly improve patient monitoring in real time, reaching a wider population and enhancing ecological validity. It can lead to personalized treatment interventions by providing comprehensive and dynamic insights into individuals' health behaviors and experiences. JTrack-EMA+ was originally designed for clinical studies, but it can also be used in different areas where EMA measures may be required such as market research, education, government, public policy, event planning, and feedback and engagement. While using smartphone-based EMA technologies offer many potential benefits, certain populations may face barriers to effectively accessing and using these apps. Some key challenges include access to smartphones or the internet in underresourced regions, older adult individuals, individuals with disabilities, and low digital literacy. A multipronged approach is needed to address these challenges, such as providing alternative low-tech EMA options, designing highly accessible and intuitive interfaces, improving digital literacy through training, and ensuring affordable access to required technologies. In this regard, the future version of JTrack-EMA+ will improve integration for web-based users as an alternative to smartphone-based EMA. Additionally, it will include conditional assessment support and additional features to analyze the collected data directly in a browser-based dashboard (JDash).

Finally, we concluded that platform-independent architectures provide comparable functionality and user experience and save development and maintenance costs over time. These principles are critical to EMA-based assessments' effectiveness and applicability in clinical trials.

XSL•FO
RenderX

Data Availability

On the iOS platform, this app is published as JTrack-EMA.

Authors' Contributions

MSF developed the JTrack-EMA+ app and performed all analyses. MSF wrote this paper with the support of JD. JMF and MN developed the online dashboard app. JD designed the overall study. SBE and TM contributed to the overall design of this study. DS contributed to the development of questions for the pilot study. RR and SS recruited probands and the clinic. All authors contributed to this paper and approved the submitted version. This study was supported by the Human Brain Project, funded by the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement 785907 (Human Brain Project SGA2).

Conflicts of Interest

None declared.

References

1. Hart A, Reis D, Prestele E, Jacobson NC. Using smartphone sensor paradata and personalized machine learning models to infer participants' well-being: ecological momentary assessment. J Med Internet Res 2022;24(4):e34015 [FREE Full text] [doi: 10.2196/34015] [Medline: 35482397]
2. Lyketsos CG, Kozauer N, Rabins PV. Psychiatric manifestations of neurologic disease: where are we headed? Dialogues Clin Neurosci 2007;9(2):111-124 [FREE Full text] [doi: 10.31887/DCNS.2007.9.2/clyketsos] [Medline: 17726911]
3. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess 2009;21(3):425-436 [FREE Full text] [doi: 10.1037/a0016869] [Medline: 19719353]
4. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. npj Digit Med 2019;2(1):14 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
5. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
6. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
7. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
8. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. J Nerv Ment Dis 1987;175(9):526-536. [doi: 10.1097/00005053-198709000-00004] [Medline: 3655778]
9. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
10. Conner TS, Lehman BJ. Getting Started: Launching a Study in Daily Life. New York City, NY: The Guilford Press; 2012.
11. Stone A, Shiffman S, Atienza A, Nebeling L. The Science of Real-Time Data Capture. New York, NY: Oxford University Press; 2007.
12. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
13. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front. ICT 2015;2:6. [doi: 10.3389/fict.2015.00006]
14. SEMA3: Smartphone Ecological Momentary Assessment, Version 3. SEMA3. 2019. URL: https://sema3.com/index.html [accessed 2023-01-17]
15. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, et al. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR mHealth uHealth 2019;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
16. van de Ven P, O'Brien H, Henriques R, Klein M, Msetfi R, Nelson J, et al. ULTEMAT: a mobile framework for smart ecological momentary assessments and interventions. Internet Interv 2017;9:74-81 [FREE Full text] [doi: 10.1016/j.invent.2017.07.001] [Medline: 30135840]
17. Runyan JD, Steenbergh TA, Bainbridge C, Daugherty DA, Oke L, Fry BN. A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. PLoS One 2013;8(8):e71325 [FREE Full text] [doi: 10.1371/journal.pone.0071325] [Medline: 23977016]
18. Sahandi Far M, Stolz M, Fischer JM, Eickhoff SB, Dukart J. JTrack: a digital biomarker platform for remote monitoring of daily-life behaviour in health and disease. Front Public Health 2021;9:763621 [FREE Full text] [doi: 10.3389/fpubh.2021.763621] [Medline: 34869177]

XSL•FO
RenderX

19. Build apps for any screen. Flutter. URL: https://tinyurl.com/3ja33enw [accessed 2023-05-17]
20. Biomarker-Development-at-INM7. GitHub. URL: https://github.com/Biomarker-Development-at-INM7 [accessed 2024-12-21]
21. Halchenko YO, Meyer K, Poldrack B, Solanky D, Wagner AS, Gors J, et al. DataLad: distributed system for joint management of code, data, and their relationship. J Open Source Softw 2021;6(63):3262. [doi: 10.21105/joss.03262] [Medline: 39469147]
22. Wise J, de Barron AG, Splendiani A, Balali-Mood B, Vasant D, Little E, et al. Implementation and relevance of FAIR data principles in biopharmaceutical R&D. Drug Discov Today 2019;24(4):933-938 [FREE Full text] [doi: 10.1016/j.drudis.2019.01.008] [Medline: 30690198]
23. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
24. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]

Abbreviations

DB: digital biomarker
EMA: ecological momentary assessment
FAIR: findable, accessible, interpretable, and reusable
GDPR: General Data Protection Regulation

Edited by G Tsafnat; submitted 08.08.23; peer-reviewed by N Mir, M Stach, M Stach; comments to author 29.02.24; revised version received 25.04.24; accepted 13.11.24; published 28.01.25.

Please cite as:
Sahandi Far M, Fischer JM, Senge S, Rathmakers R, Meissner T, Schneble D, Narava M, Eickhoff SB, Dukart J
Cross-Platform Ecological Momentary Assessment App (JTrack-EMA+): Development and Usability Study
J Med Internet Res 2025;27:e51689
URL: https://www.jmir.org/2025/1/e51689
doi:10.2196/51689
PMID:

©Mehran Sahandi Far, Jona M Fischer, Svea Senge, Robin Rathmakers, Thomas Meissner, Dominik Schneble, Mamaka Narava, Simon B Eickhoff, Juergen Dukart. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relationship Among Macronutrients, Dietary Components, and Objective Sleep Variables Measured by Smartphone Apps: Real-World Cross-Sectional Study

Jaehoon Seol[1,2,3,4], PhD; Masao Iwagami[1,5], MD, PhD; Megane Christiane Tawylum Kayamare[1,6], MS; Masashi Yanagisawa[1,7,8,9,10], MD, PhD

[1]International Institute for Integrative Sleep Medicine (WPI-IIIS), University of Tsukuba, Tsukuba, Ibaraki, Japan

[2]Institute of Health and Sport Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

[3]Department of Frailty Research, Center for Gerontology and Social Science, National Center for Geriatrics and Gerontology, Obu, Aichi, Japan

[4]Research Team for Promoting Independence and Mental Health, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

[5]Institute of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan

[6]PhD Program in Humanics, University of Tsukuba, Tsukuba, Ibaraki, Japan

[7]S'UIMIN Inc, Shibuya, Tokyo, Japan

[8]Life Science Center for Survival Dynamics (TARA), University of Tsukuba, Tsukuba, Ibaraki, Japan

[9]R&D Center for Frontiers of Mirai in Policy and Technology (F-MIRAI), University of Tsukuba, Tsukuba, Ibaraki, Japan

[10]Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Masashi Yanagisawa, MD, PhD
International Institute for Integrative Sleep Medicine (WPI-IIIS)
University of Tsukuba
1-1-1 Tennodai
Tsukuba, Ibaraki, 305-8575
Japan
Phone: 81 29 853 5857
Email: yanagisawa.masa.fu@u.tsukuba.ac.jp

Abstract

Background: Few studies have explored the relationship between macronutrient intake and sleep outcomes using daily data from mobile apps.

Objective: This cross-sectional study aimed to examine the associations between macronutrients, dietary components, and sleep parameters, considering their interdependencies.

Methods: We analyzed data from 4825 users of the Pokémon Sleep and Asken smartphone apps, each used for at least 7 days to record objective sleep parameters and dietary components, respectively. Multivariable regression explored the associations between quartiles of macronutrients (protein; carbohydrate; and total fat, including saturated, monounsaturated, and polyunsaturated fats), dietary components (sodium, potassium, dietary fiber, and sodium-to-potassium ratio), and sleep variables (total sleep time [TST], sleep latency [SL], and percentage of wakefulness after sleep onset [%WASO]). The lowest intake group was the reference. Compositional data analysis accounted for macronutrient interdependencies. Models were adjusted for age, sex, and BMI.

Results: Greater protein intake was associated with longer TST in the third (+0.17, 95% CI 0.09-0.26 h) and fourth (+0.18, 95% CI 0.09-0.27 h) quartiles. In contrast, greater fat intake was linked to shorter TST in the third (–0.11, 95% CI –0.20 to –0.27 h) and fourth (–0.16, 95% CI –0.25 to –0.07 h) quartiles. Greater carbohydrate intake was associated with shorter %WASO in the third (–0.82%, 95% CI –1.37% to –0.26%) and fourth (–0.57%, 95% CI –1.13% to –0.01%) quartiles, while greater fat intake was linked to longer %WASO in the fourth quartile (+0.62%, 95% CI 0.06%-1.18%). Dietary fiber intake correlated with longer TST and shorter SL. A greater sodium-to-potassium ratio was associated with shorter TST in the third (–0.11, 95% CI –0.20 to –0.02 h) and fourth (–0.19, 95% CI –0.28 to –0.10 h) quartiles; longer SL in the second (+1.03, 95% CI 0.08-1.98 min) and fourth (+1.50, 95% CI 0.53-2.47 min) quartiles; and longer %WASO in the fourth quartile (0.71%, 95% CI 0.15%-1.28%). Compositional data analysis, involving 6% changes in macronutrient proportions, showed that greater protein intake was associated with an elevated TST (+0.27, 95% CI 0.18-0.35 h), while greater monounsaturated fat intake was associated with a longer SL (+4.6, 95%

XSL•FO
RenderX

CI 1.93-7.34 min) and a larger %WASO (+2.2%, 95% CI 0.63%-3.78%). In contrast, greater polyunsaturated fat intake was associated with a reduced TST (–0.22, 95% CI –0.39 to –0.05 h), a shorter SL (–4.7, 95% CI to 6.58 to –2.86 min), and a shorter %WASO (+2.0%, 95% CI –3.08% to –0.92%).

Conclusions: Greater protein and fiber intake were associated with longer TST, while greater fat intake and sodium-to-potassium ratios were linked to shorter TST and longer WASO. Increasing protein intake in place of other nutrients was associated with longer TST, while higher polyunsaturated fat intake improved SL and reduced WASO.

(J Med Internet Res 2025;27:e64749) doi:10.2196/64749

KEYWORDS

sleep quality; dietary health; unsaturated fatty acids; dietary fiber intake; sodium-to-potassium ratio; compositional data analysis; sleep; smartphone; application

Introduction

Dietary intake patterns vary widely depending on the lifestyle. Brain imaging reveals that these patterns are associated with mental health, cognitive function, blood and metabolic biomarkers, overall brain structure, and white matter integrity [1]. Diet and sleep are closely linked [2]. A systematic review partially revealed that overeating, especially meals high in fat, can extend sleep latency (SL), increase wakefulness after sleep onset (WASO), and ultimately reduce total sleep time (TST) [3-5].

Dietary intake and sleep are known to have a bidirectional relationship [2]. The intake of healthy foods is associated with improved sleep quality, whereas consuming processed foods and foods high in free sugars has been shown to deteriorate sleep quality [6]. Several hypotheses have been proposed to explain this relationship [7-9]. One hypothesis suggests that an increased dietary intake of tryptophan can promote the production of serotonin and melatonin, potentially leading to improved sleep quality [7]. Additionally, sleep deprivation can decrease leptin levels and increase ghrelin levels, a hormone that stimulates appetite, and may trigger overeating, especially at night, creating a negative spiral [7-9].

The importance of assessment methods in examining the relationship between diet and sleep has been emphasized in numerous studies [2,3,10]. Subjective methods, such as dietary diaries, 24-hour recalls, and food frequency questionnaires, rely heavily on the participants' memory and can be prone to recall bias. In contrast, objective methods, such as digital apps that record dietary intake in real time, can reduce bias; however, they still face challenges related to technology use and data accuracy [11]. Furthermore, automated systems have been introduced to eliminate recording and subjective biases, allowing users to register menus or scan the barcodes of commercially available products, which automatically calculates the nutritional components [12]. Similarly, sleep assessments also face challenges. Subjective sleep assessments, such as sleep diaries or self-reported questionnaires, are prone to recall bias, sleep-state misperception, and underreporting. Objective assessments, including polysomnography, portable electroencephalogram devices, and wearable devices such as accelerometers, provide more accurate data but are often costly, require technical expertise, or can be intrusive for users [11,13-15]. Each method has its own set of limitations, making it important to carefully choose the appropriate assessment method based on the research context.

Observational studies investigating the relationship between diet and sleep often rely on subjective assessments of nutrition and sleep, with many highlighting the potential for recall bias [2,3]. Many studies have focused on the intake of individual nutrients, such as specific proteins, carbohydrates, fatty acids, and vitamins [16-20]. Specifically, increased protein intake has been reported to improve subjective sleep quality [16,17], while carbohydrate intake affects the sleep stages, SL, and WASO; however, further research is warranted on its long-term impact [21]. Imbalances in polyunsaturated fatty acids have been linked to a higher prevalence of sleep disorders and shorter TST [19,20]. Additionally, vitamin D deficiency has been associated with sleep disorders and shorter TST [18]. However, macronutrients that contribute to the total energy intake (such as carbohydrates, proteins, and fats, including saturated and unsaturated fatty acids) are interdependent and interact with each other [22]. Although recent research has considered these interdependencies and their relationship with health status [22-24], research on how these interdependencies affect sleep is limited. In particular, these methods, which consider the interdependencies of nutrient components, are expected to attract more attention in the future but have not yet been fully explored [25].

Understanding these complex relationships is crucial before transitioning to intervention studies, as it provides a foundational understanding of how nutrients work together to influence health outcomes. Therefore, an approach that accounts for the interdependent relationships between nutrients is necessary. This cross-sectional study using real-world data from smartphone apps aimed to examine the hypothesis that there are relationships among macronutrients, dietary components, and sleep parameters. We also investigated the relationships between sleep and macronutrients while considering their interdependencies using compositional data analysis.

Methods

Study Design and Participants

This retrospective, cross-sectional study used data from 6052 users who consented to participate between January 19 and 31, 2024. During this period, in-app messages were sent to all the users of the Asken app (Asken Inc), which is designed for dietary management and recording, requesting permission to

XSL•FO
RenderX

use their stored data. The goal of the Asken app was to gather detailed information on dietary intake, enabling us to assess the participants' nutritional habits.

At the same time, users who also used Pokémon Sleep (The Pokémon Company), a sleep-tracking app, were asked to provide their user IDs in their electronic informed consent. This linking of data was crucial for analyzing the relationship between the objective sleep data collected from the Pokémon Sleep app and the dietary data from the Asken app.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Sapporo Yurinokai Hospital, Japan (approval 030). Informed consent was obtained from all participants involved in the study. Participants used both apps concurrently from July 2023 to January 2024. Furthermore, data were extracted only from the overlapping period between the start and end dates of both apps, and average values were considered representative of each participant's typical sleep habits and nutritional intake status. To ensure participant privacy and confidentiality, all data were deidentified before analysis, and no personally identifiable information was included in the dataset. The dataset was securely stored and accessible only to authorized researchers. Participants were not provided with monetary compensation for their participation in the study. However, they were fully informed about the study's purpose and significance, and their voluntary participation was obtained based on informed consent.

Dietary Intake Patterns Assessment

We calculated dietary intake patterns using the Asken smartphone app. Participants logged their meals by selecting food items and portion sizes from a database containing over 100,000 menu items. Participants entered the details of each meal into the app. For commercial products, the nutritional information was automatically recorded by scanning the barcode. The nutritional values were calculated based on the *2020 Standard Tables of Food Composition in Japan (eighth edition)* [26]. The recording and calculation methods used by the participants are detailed in a previous study [12]. The Asken dietary recording app and traditional paper-based questionnaires have been reported to have a reasonably good correlation coefficient of 0.80 for nutrient intake [12,27,28].

In this retrospective, cross-sectional study, data were only included from days when participants reported consuming all three meals, as it was not possible to distinguish between missed recordings and actual skipped meals. Asken allows for the calculation of various macronutrients and dietary components, including total energy, protein, total fat (including saturated, monounsaturated, and polyunsaturated fats), carbohydrate, sodium, potassium, calcium, magnesium, phosphorus, iron, zinc, copper, manganese, vitamin A, vitamin D, vitamin E, vitamin K, vitamin B1, vitamin B2, niacin, vitamin B6, vitamin B12, folate, pantothenic acid, vitamin C, cholesterol, dietary fiber, and alcohol intake.

Based on previous research [29,30], we selected and analyzed the following components reportedly associated with sleep: total energy, protein, total fat (including saturated, monounsaturated, and polyunsaturated fats), carbohydrate, sodium, potassium, and dietary fiber intake, and calculated the sodium-to-potassium ratio. Macronutrients were energy adjusted by calculating the percentage of total energy intake. The absolute daily intake was calculated for other dietary components.

Sleep Assessment

We calculated the TST, SL, and percentage of WASO (%WASO) derived from the Pokémon Sleep, which uses the Cole-Kripke algorithm for sleep-wake determination [31], as publicly disclosed [32]. Although accelerometer-based sleep determination integrated into smartphones lacks the ability to distinguish sleep stages [33,34], it provides a certain level of accuracy in determining sleep and wakefulness [22,23]. Although there may be a slight underestimation of SL, the overall TST and nocturnal awakenings showed good agreement with traditional actigraphy [34,35]. As previously mentioned, Natale et al [35] used the Cole-Kripke algorithm with smartphone accelerometers. Briefly, sleep-wake determination was performed using accelerometer data with 10-second epochs per minute. The decision was based on a 7-minute window of accelerometer data (from 4 min before to 3 min after the reference point [0 min]). A value greater than 1 indicates wakefulness and a value less than 1 indicates sleep [31]. SL was defined as the time from going to bed (start of the recording) to the first instance of being determined as sleep.

Statistical Analysis

Descriptive analysis was conducted to summarize the basic characteristics of the participants, including sleep status and dietary nutrient intake. The multivariable regression analysis, isotemporal substitution model, and compositional data analysis were performed using the *lmtest*, *deltacomp*, *Compositions*, *robCompositions*, and *ggplot2* R packages, respectively [25,36-38]. All regression models were adjusted for age, sex, and BMI.

First, multivariable regression analysis using the forced-entry method was used to determine the association among macronutrients, dietary components, and objective sleep variables (TST, SL, and %WASO). We conducted quartile categorization for total energy intake, energy-adjusted macronutrients (protein intake, total fat intake including saturated, monounsaturated, and polyunsaturated fats, and carbohydrate intake), sodium intake, potassium intake, dietary fiber intake, and sodium-to-potassium ratio. All the quartile range information for each variable is presented in Table S1 in Multimedia Appendix 1. Dummy variables were created for each nutrient based on the first quartile.

Second, to investigate the linear relationship between macronutrient intake and sleep parameters, we used an isotemporal substitution model and compositional data analysis [22-24]. In the isotemporal substitution model, when one macronutrient component is replaced with another on a one-to-one basis, the other components remain constant [22,23]. Compositional data analysis, in contrast, considers the overall composition and estimates the changes in sleep variables when the proportion of a specific macronutrient component is increased or decreased within the total composition [25]. Both

XSL•FO
RenderX

methods have been used in studies that consider the interdependencies of factors contributing to a total composition of 100% (eg, dietary nutrient intake, 24-h physical/sedentary activity, and TST) [22,23,37,38].

We calculated the macronutrient intake (in grams) by multiplying each intake by its respective calorie contribution (ie, 4 kcal for carbohydrates and protein and 9 kcal for each type of fat), ensuring that the sum equals 1 (ie, total energy intake). Nutrient intake data were transformed into compositional data using an additive log-ratio transformation. Nutrient intake data were transformed into compositional data using an additive log-ratio transformation. This transformation was necessary to account for the relative proportions of macronutrients and facilitate subsequent analyses. The log transformation helps meet the assumptions required for compositional data analysis [25]. The variability in the data, in terms of the variability of each macronutrient intake relative to the variability of other components and the total variance of the whole composition, is described in Table S2 in Multimedia Appendix 1 through a variation matrix within each component. A value close to zero implied that the intake of the two macronutrient components was highly proportional.

The compositional data were then subjected to an isometric log-ratio (ilr) transformation to facilitate the application of the linear regression models. The ilr transformation was conducted using default and customized orthonormal bases for different nutrients. Linear regression models were constructed to evaluate the association between sleep parameters and the ilr-transformed nutrient composition after adjusting for sex, age, and BMI (Table S3 in Multimedia Appendix 1). The mean nutrient composition was calculated, and predictions for sleep parameters were made using a linear regression model. The compositional ratios of macronutrients among the participants are presented in Table S4 in Multimedia Appendix 1. Because of the isotemporal substitution model and compositional data analysis, it was not possible to replace values below the minimum composition. Therefore, in this study, we replaced values exceeding the upper limit of 6% (Table S4 in Multimedia Appendix 1).

To assess the impact of altering the nutrient composition, specific changes (eg, increasing monounsaturated fat and decreasing polyunsaturated fat) were applied, and new predictions were obtained. CIs for the differences in the predicted sleep parameters were computed using the residual standard error, model matrix, and critical value from the *t*-distribution. Hypothesis tests were conducted to examine the significance of the associations between individual nutrients and sleep parameters. Different ilr transformations were applied, each focusing on a specific nutrient (protein, carbohydrate, saturated fat, monounsaturated fat, or polyunsaturated fat), and the corresponding linear models were analyzed for significance.

Ramsey regression equation specification error test confirmed the linearity assumption for all models (TST, SL, and %WASO). Variance inflation factor values ranged from 1.10 to 4.65, indicating no multicollinearity. The Durbin-Watson test showed no autocorrelation in the residuals, and the Breusch-Pagan test confirmed homoscedasticity. Although the Shapiro-Wilk test showed significant *P* values, W-values between 0.996 and 0.997 indicated that the residuals were close to normally distributed.

All descriptive and statistical analyses were performed using R (version 4.4.0; R Foundation for Statistical Computing), with statistical significance assessed at the level of .05.

Results

Of the 6052 participants who consented, 140 individuals were excluded from the analysis due to missing sleep records, 377 were excluded due to being outliers in the TST (ie, participants who did not fall within the average SD 3 range), and 710 were excluded because they used both apps for fewer than 7 days each. Therefore, the final analysis included 4825 participants (mean age 36.7, SD 10.4 years; BMI 24.8, SD 4.8 kg/m^2; 3938/4825, 81.6% were female; Figure 1 and Table 1).

Figure 1. Participant flow chart.

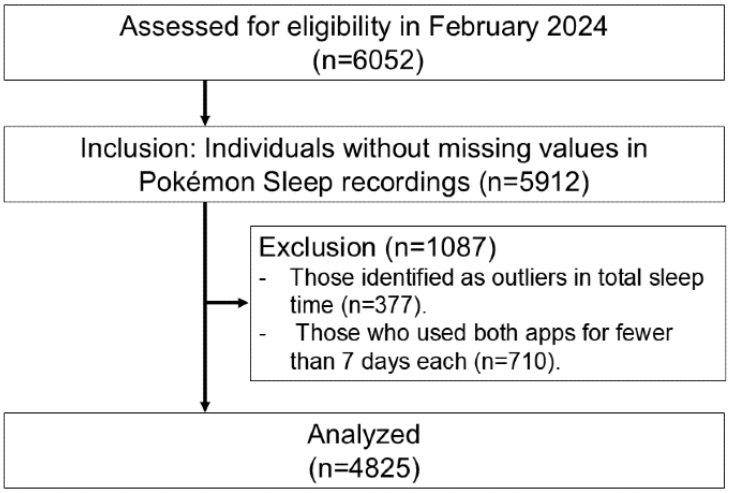

XSL•FO
RenderX

Table 1. Characteristics of participants.

Variables	All participants (n=4825)	Male participants (n=887)	Female participants (n=3938)
Basic characteristics			
Age (years), mean (SD)	36.7 (10.4)	38.4 (10.8)	36.3 (10.2)
BMI (kg/m^2), mean (SD)	24.8 (4.8)	26.0 (4.5)	24.6 (4.8)
Physical activity level (inactive), n (%)	4381 (90.8)	775 (87.4)	3606 (91.6)
Sleep status, mean (SD)			
Number of days of Pokémon Sleep recording	134.0 (54.5)	135.7 (56.0)	133.7 (54.2)
Total sleep time (hours)	6.7 (1.1)	6.5 (1.1)	6.7 (1.1)
Sleep onset latency (minutes)	21.5 (12.1)	23.5 (13.6)	21.1 (11.7)
Percentage wakefulness after sleep onset (%)	9.9 (7.2)	13.1 (9.1)	9.2 (6.4)
Nutrient status, mean (SD)			
Number of days of Asken data recording (days)	85.8 (63.6)	93.2 (67.3)	84.1 (62.7)
Total energy intake (kcal)	1662.6 (330.3)	2055.1 (349.2)	1574.2 (252.4)
Protein intake (% total kcal)	17.7 (3.7)	18.8 (4.8)	17.5 (3.4)
Carbohydrate intake (% total kcal)	51.7 (5.5)	50.3 (5.9)	52.0 (5.3)
Total fat intake (% total kcal)	32.3 (4.3)	31.4 (4.8)	32.6 (4.2)
Saturated fat intake (% total kcal)	9.5 (1.8)	8.8 (1.8)	9.7 (1.8)
Monounsaturated fat intake (% total kcal)	11.5 (2.0)	11.3 (2.2)	11.5 (1.9)
Polyunsaturated fat intake (% total kcal)	6.3 (1.0)	6.1 (1.2)	6.3 (1.0)
Dietary fiber intake (g/day)	11.7 (3.5)	10.7 (2.7)	11.9 (3.6)
Sodium (mg/day)	3486.6 (881.0)	4230.7 (996.7)	3319.1 (758.1)
Potassium (mg/day)	2295.3 (602.1)	2625.0 (692.4)	2221.0 (553.3)
Sodium to Potassium ratio	1.6 (0.4)	1.7 (0.5)	1.6 (0.4)

The mean macronutrient composition was as follows: protein, 18.2%; carbohydrates, 53.8%; and total fat, 28% (saturated fat at 9.7%, monounsaturated fat at 11.8%, and polyunsaturated fat at 6.5%; Table S4 in Multimedia Appendix 1).

Figure 2 and Table S5 in Multimedia Appendix 1 present the results of multivariate regression analyses comparing each nutrient across quartiles to the first quartile in relation to sleep variables. Specifically, in terms of macronutrients, the TST was significantly shorter in the second, third, and fourth quartiles compared with the first quartile with lower total energy intake. Greater intake ratios of total fat, including saturated, monounsaturated, and polyunsaturated fats, were associated with shorter TST. In contrast, greater protein intake levels in the third and fourth quartiles were associated with longer TST compared with those in the first quartile.

Regarding SL, greater levels of total, saturated, and monounsaturated fat intake were linked to longer SL, while greater intake of polyunsaturated fats showed a tendency toward shorter SL. Greater total energy and total fat intakes were associated with longer %WASO, whereas greater carbohydrate intake indicated a tendency towards a shorter %WASO. Greater dietary fiber intake consistently showed associations with longer TST, shorter SL, and shorter %WASO, suggesting its beneficial effect on sleep quality. Greater sodium intake was associated with shorter TST, and both greater sodium and potassium intake were linked to shorter SL. Notably, a greater sodium-to-potassium ratio was related to shorter TST, longer SL, and %WASO (Figure 2 and Table S5 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 2. Multivariable regression analysis of macronutrients and dietary components on sleep variables.

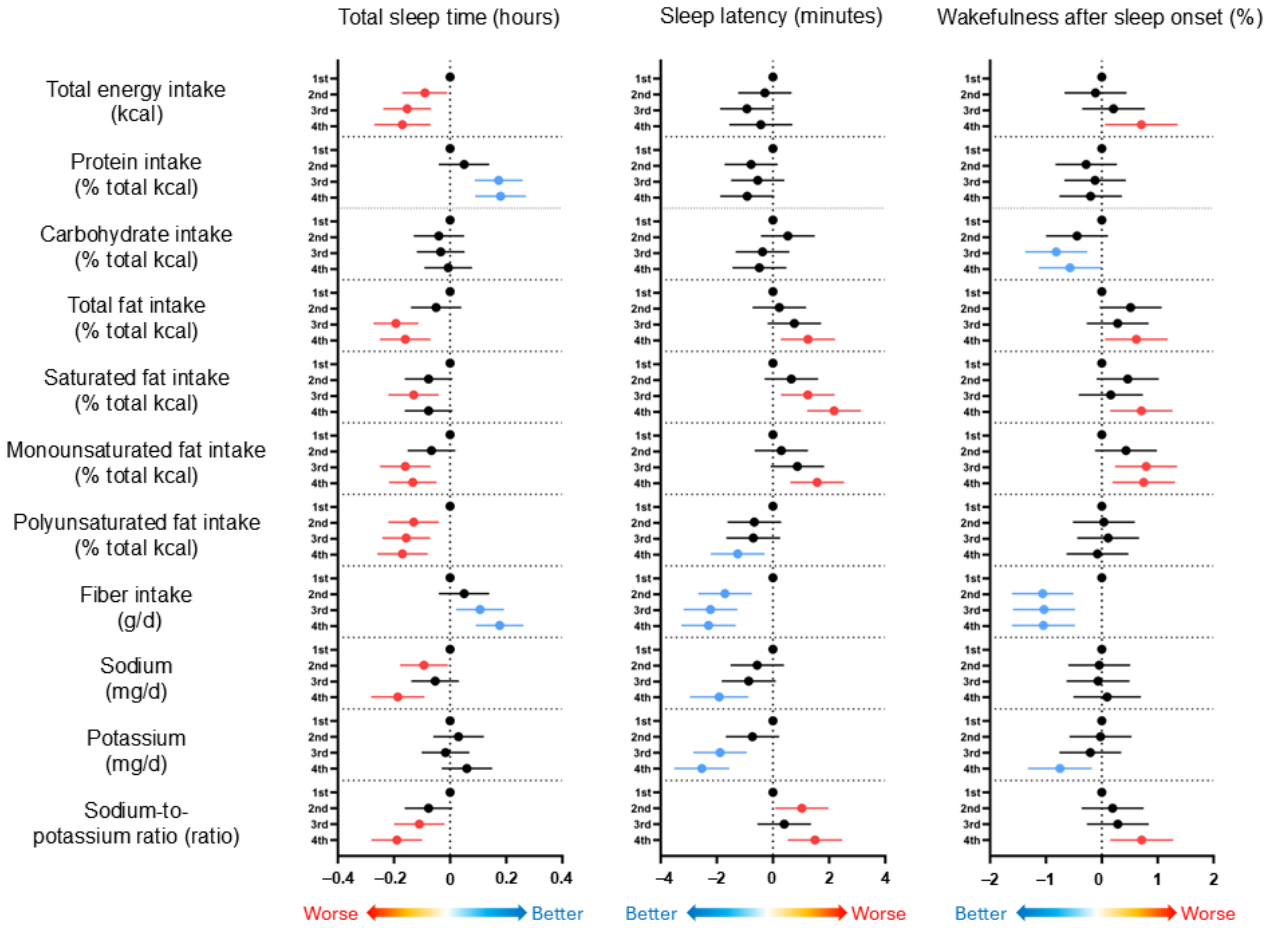

The results of the isotemporal substitution model showed that replacing 6% of total energy intake from protein with carbohydrates was significantly associated with a reduction of 0.2 hours in TST. Conversely, replacing 6% of polyunsaturated fat with equivalent amounts of protein, carbohydrates, saturated fat, and monounsaturated fat was significantly associated with changes in the TST by +0.6, +0.4, +0.5, and +0.6 hours, respectively (Figure 3 and Table S6 in Multimedia Appendix 1). Regarding SL, replacing 6% of proteins or carbohydrates with monounsaturated fat was significantly associated with a change of +2.4 and +2.8 minutes, respectively. Conversely, replacing 6% of monounsaturated fat with equivalent amounts of protein, carbohydrates, and saturated fat was significantly associated with SL changes of –3.7, –4.1, and –3.2 minutes, respectively. Additionally, replacing 6% polyunsaturated fat with equivalent amounts of protein, carbohydrates, saturated fat, and monounsaturated fat was significantly associated with changes in SL of +10.3, +9.9, +10.8, and +12.7 minutes, respectively (Figure 4 and Table S6 in Multimedia Appendix 1). For %WASO, replacing 6% of protein with monounsaturated fat was significantly associated with a +1.2% change, whereas replacing saturated fat with monounsaturated fat was significantly associated with a +2.3% change. Conversely, replacing 6% of monounsaturated fat with equivalent amounts of protein, carbohydrates, and saturated fat was significantly associated with changes in %WASO by –1.8%, –1.7%, and –2.5%, respectively. Additionally, replacing 6% polyunsaturated fat with equivalent amounts of protein, carbohydrates, saturated fat, and monounsaturated fat was significantly associated with changes in %WASO of +4.3%, +4.5%, +3.7%, and +5.5%, respectively (Figure 5 and Table S6 in Multimedia Appendix 1).

In the compositional data analysis, changes in protein by –6% of the overall macronutrient composition were significantly associated with changes in the TST by –0.6 hours, whereas changes in polyunsaturated fat by –6% were significantly associated with changes in the TST by +0.5 hours. Conversely, changes in protein by +6% from the overall macronutrient composition were significantly associated with changes in TST by +0.3 hours, while changes in polyunsaturated fat by +6% were significantly associated with changes in TST by –0.2 hours (Table 2). Regarding SL, changes in monounsaturated fat by –6% from the overall macronutrient composition were significantly associated with a change of –10.5 minutes, whereas changes in polyunsaturated fat by –6% were associated with a change of +10.7 minutes. Conversely, changes in monounsaturated fat by +6% were significantly associated with a change of +4.6 minutes, and changes in polyunsaturated fat by +6% were significantly associated with a change of –4.7 minutes (Table 2). For %WASO, similar to SL, changes in monounsaturated fats by –6% of the overall macronutrient composition were significantly associated with a change of –5.0%, whereas a –6% change in polyunsaturated fats was significantly associated with a change of +4.5%. Conversely,

XSL•FO
RenderX

changes in the monounsaturated fat content by +6% were significantly associated with a change of +2.2%, and changes in the polyunsaturated fat content by +6% were significantly associated with a change of –2.0% (Table 2).

Figure 3. Simulated changes in total sleep time when reallocating fixed amounts of other nutrient components among each nutrient component: (A) protein, (B) carbohydrate, (C) saturated fat, (D) monounsaturated fat, and (E) polyunsaturated fat, while keeping the remaining components constant at compositional percentages. All estimates have been adjusted for age, sex, and BMI.

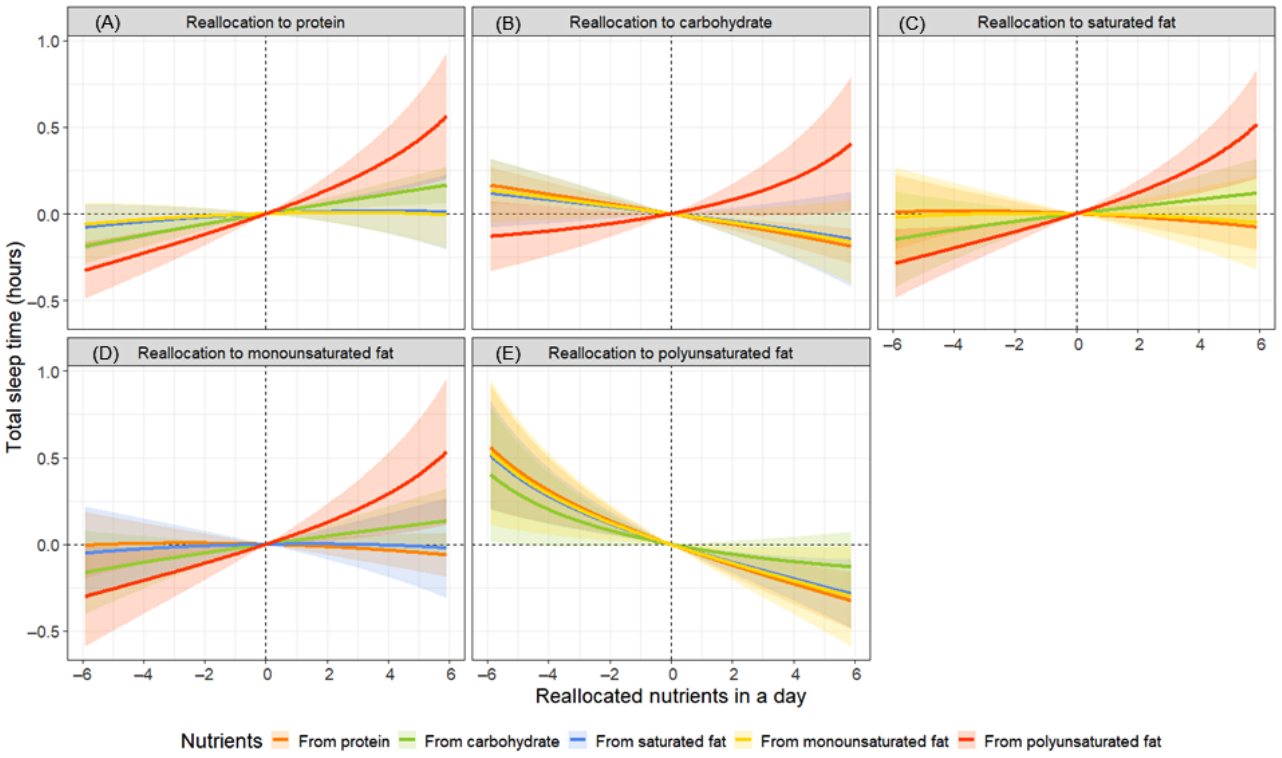

Figure 4. Simulated changes in sleep latency when reallocating fixed amounts of other nutrient components among each nutrient component: (A) protein, (B) carbohydrate, (C) saturated fat, (D) monounsaturated fat, and (E) polyunsaturated fat, while keeping the remaining components constant at compositional percentages. All estimates have been adjusted for age, sex, and BMI.

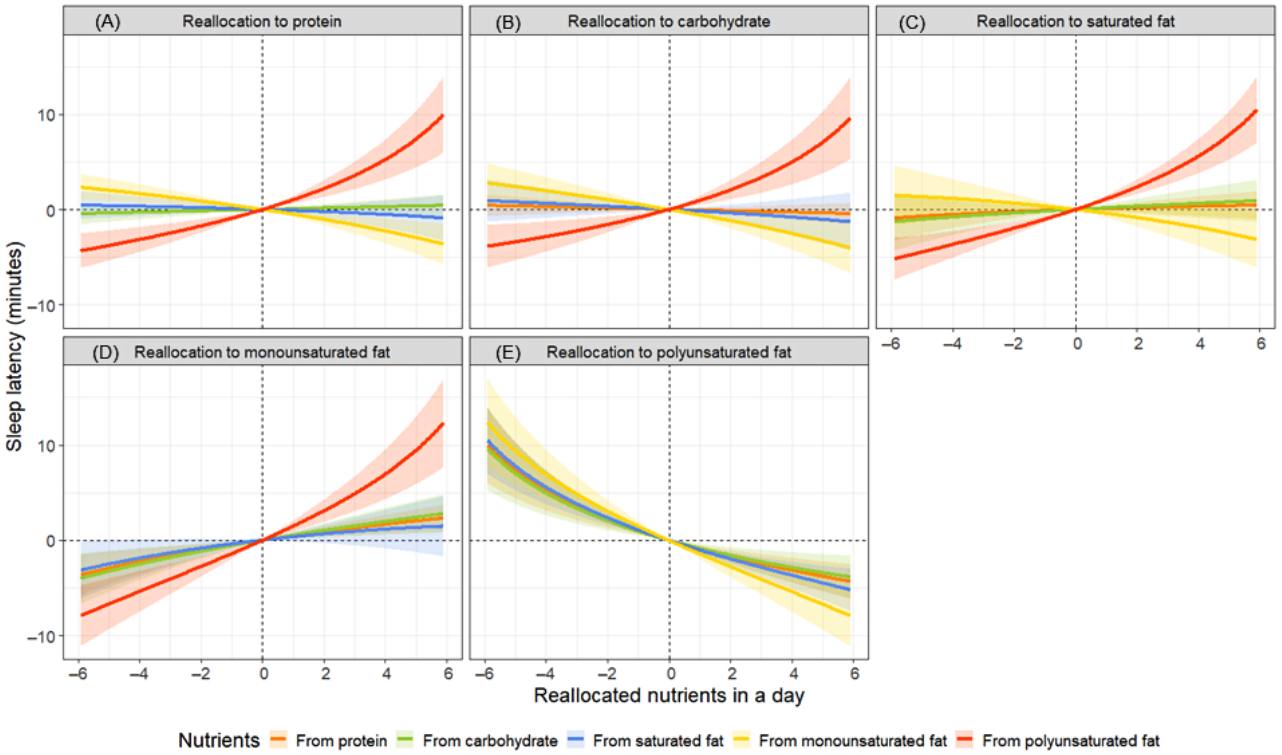

XSL•FO
RenderX

Figure 5. Simulated changes in the proportion of wakefulness after sleep onset when reallocating fixed amounts of other nutrient components among each nutrient component: (A) protein, (B) carbohydrate, (C) saturated fat, (D) monounsaturated fat, and (E) polyunsaturated fat, while keeping the remaining components constant at compositional percentages. All estimates have been adjusted for age, sex, and BMI.

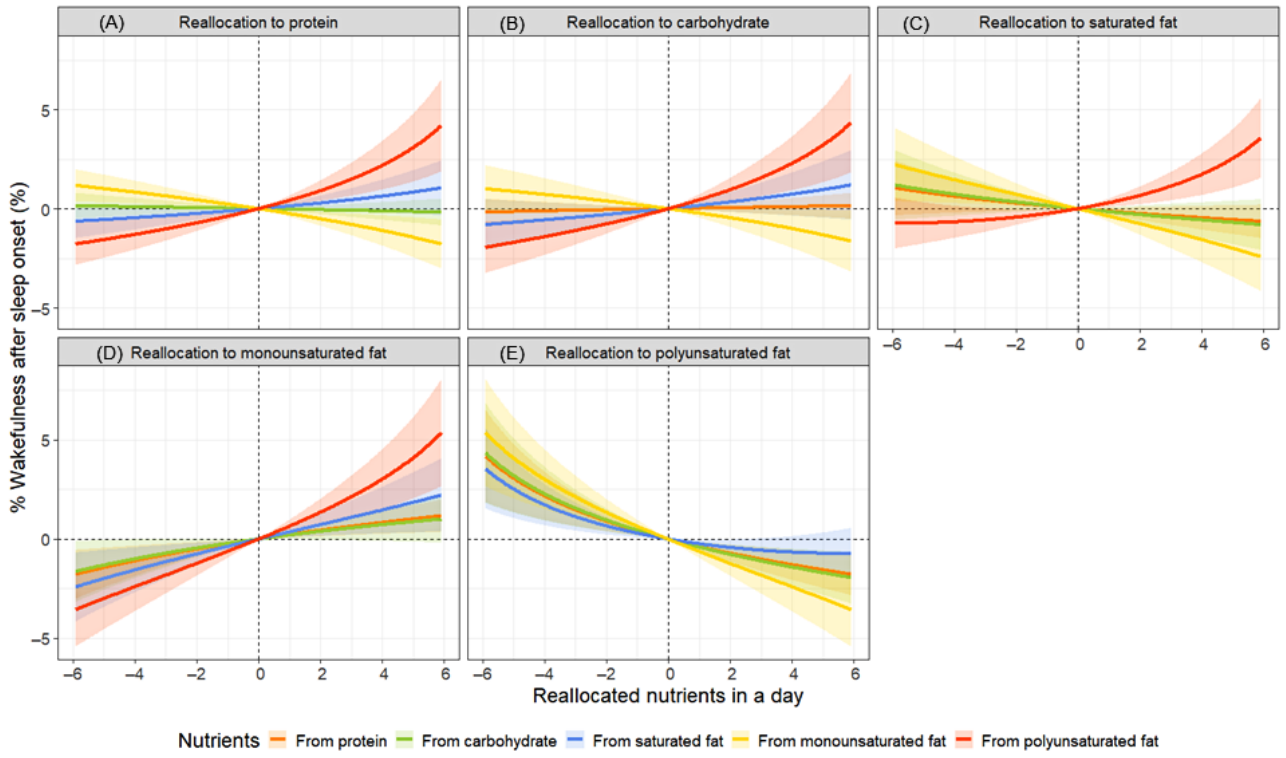

Table 2. Estimated nonstandardized coefficients for displacing 6% of macronutrients between one and the remaining macronutrients proportionally. Data are presented as nonstandardized coefficients (95% CI). All the estimates were adjusted for age, sex, and BMI. Nonstandardized coefficients reflect the estimated change in sleep parameters associated with reallocating 6% of the macronutrients between one and the remaining macronutrients proportionally.

	Total sleep time (hours), 95% CI		Sleep latency (minutes), 95% CI		% Wakefulness after sleep onset (%), 95% CI	
	Take 6% from one nutrient and distribute it	Add 6% to one nutrient from the others	Take 6% from one nutrient and distribute it	Add 6% to one nutrient from the others	Take 6% from one nutrient and distribute it	Add 6% to one nutrient from the others
Protein	–0.61 (–0.80 to –0.41)[a]	+0.27 (0.18 to 0.35)[a]	–0.78 (–2.93 to 1.38)	+0.34 (–0.61 to 1.29)	–0.06 (–1.32 to 1.20)	+0.03 (–0.53 to 0.58)
Carbohydrate	+0.60 (–0.20 to 1.40)	–0.27 (–0.62 to 0.09)	+2.79 (–5.98 to 11.57)	–1.23 (–5.10 to 2.64)	–1.70 (–6.81 to 3.41)	+0.75 (–1.51 to 3.01)
Saturated fat	–0.20 (–0.67 to 0.28)	+0.09 (–0.12 to 0.30)	–2.20 (–7.40 to 2.99)	+0.97 (–1.32 to 3.26)	+2.23 (–0.80 to 5.26)	–0.99 (–2.32 to 0.35)
Monounsaturated fat	–0.30 (–0.85 to 0.26)	+0.13 (–0.11 to 0.38)	–10.50 (–16.63 to –4.38)[a]	+4.64 (1.93 to 7.34)[a]	–5.00 (–8.57 to –1.43)[a]	+2.21 (0.63 to 3.78)[a]
Polyunsaturated fat	+0.50 (0.12 to 0.88)[a]	–0.22 (–0.39 to –0.05)[a]	+10.69 (6.48 to 14.90)[a]	–4.72 (–6.58 to –2.86)[a]	+4.53 (2.07 to 6.99)[a]	–2.00 (–3.08 to –0.92)[a]

[a]Statistically significant (P<.05).

Discussion

Principal Findings

This study examined the relationship between nutritional intake and sleep using multivariate regression analysis, isotemporal substitution models, and compositional data analysis. Overall, the findings are consistent with those of previous research on individual nutrients and sleep [3-6,16-21]. As shown in Figure 2, greater total energy, fat, and sodium intakes were associated with shorter TST, whereas greater protein and dietary fiber intakes were linked to longer TST. Furthermore, the main findings of this study, based on compositional data analysis considering interdependencies, demonstrated that increasing protein while reducing other factors is associated with an elevated TST, whereas an increase in the polyunsaturated fat

XSL•FO
RenderX

with a reduced TST (Table 2). Regarding SL and WASO, monounsaturated fat and polyunsaturated fat exhibited opposing relationships (Table 2). The differences in the compositional data analysis results, compared with individual macronutrients (eg, carbohydrates), likely reflect the strong interdependent relationships among various nutrients.

Dietary fiber intake was consistently associated with improved SL and %WASO. Additionally, greater sodium intake was correlated with shorter TST, whereas greater potassium intake was associated with shorter SL. Notably, polyunsaturated fats demonstrated potential benefits for SL and %WASO compared with other types of fat. Studies suggest that dietary fiber can affect sleep by influencing microbiota [39,40]. When dietary fibers are fermented by microbiota in the large intestine, they produce short-chain fatty acids, such as acetate, propionate, and butyrate, which enhance serotonin release. Furthermore, administering tributyrin, a compound made of three molecules of butyric acid and glycerol, to mice resulted in a 50% increase in nonrapid eye movement sleep for 4 h [40]. The results of this study support those of previous studies [39,40].

Previous studies have established that greater total energy intake and poor dietary quality are associated with lower subjective sleep quality [7,41]. Specifically, higher protein intake correlates with improved sleep, whereas a higher intake of fatty acids is linked to worse sleep [41]. This study also found variations in the effects of lipid intake on sleep, depending on the type (eg, saturated, monounsaturated, or polyunsaturated fats). The mechanism by which proteins affect sleep following acute feeding may involve tryptophan; tyrosine; and the synthesis of brain neurotransmitters, such as serotonin, melatonin, and dopamine [42]. Although no significant differences were observed in SL or %WASO, a greater protein intake was positively associated with longer TST.

However, the specific mechanisms underlying the relationship between monounsaturated fat, polyunsaturated fat, and sleep remain unclear. A meta-analysis found an association between a greater intake of hexadecenoic acid, a saturated fatty acid, and subjective difficulty falling asleep [43]. Conversely, difficulties falling asleep were linked to a lower intake of dodecanoic acid, a monounsaturated fat, and butanoic acid, a saturated fatty acid [43]. Additionally, a reduced intake of dodecanoic acid was associated with subjective difficulties in maintaining sleep [43]. An epidemiological study involving adults evaluated the relationship between two polyunsaturated fatty acids, ω-6 and ω-3, and found that ω-6 fatty acids were related to the risk of sleep disorder, based on self-reported data, in an inverse U-shaped manner [19]. The ω-6:ω-3 ratio was positively and linearly associated with the risk of sleep disorders [19]. The underlying mechanisms involving ω-3 fatty acids and sleep are discussed, noting that a deficiency in ω-3 intake may affect cortical neuron oscillatory activity and sleep-wake patterns [20].

Additionally, the timing of fat intake may influence sleep; one observational study based on polysomnography suggested that consuming high-fat foods close to bedtime may be associated with increased WASO [44]. Continued consumption of a high-fat diet has been partially understood to reduce physical activity and increase rapid eye movements and sleep fragmentation through dopaminergic dysregulation [45]. Research involving mouse models and pregnant women has suggested a potential positive association between monounsaturated fat intake and sleep [46,47]. However, other studies have indicated a negative association between monounsaturated fat and sleep, whereas polyunsaturated fat has shown a positive association with sleep [48-50]. Previous studies have often focused on dietary fiber levels and comparisons between high- and low-saturated-fat diets [45,48]. Future studies should explore the impact of unsaturated fats, including monounsaturated and polyunsaturated fats.

Sodium and potassium have been reported to impact health individually [51,52], and a high sodium-to-potassium ratio (indicating relatively high sodium intake) has been associated with higher all-cause mortality [53]. Sodium intake is known to be correlated with sleep, indicating that greater sodium intake is associated with shorter subjective TST and worse sleep quality [30,50]. This relationship is attributable to the potential effect of sodium on nocturnal urination [51,52]. However, in this study, a greater intake of sodium alone was associated with a shorter TST, although no significant differences were observed in %WASO. Moreover, greater sodium intake was associated with shorter SL. Greater potassium intake was associated with shorter SL and reduced %WASO. All sleep variables were adversely related when considering the balance between sodium and potassium, which indicates an unhealthy dietary pattern (ie, relatively greater sodium intake). In this study, sodium intake was used as a basis for analysis, while the sodium-to-potassium ratio was also evaluated (Figure 2). Therefore, the results linking sleep outcomes with the sodium-to-potassium ratio are considered reasonable, as they provide a more comprehensive perspective on the impact of both minerals on sleep.

Recent studies suggested that changes in the gut microbiota may influence sleep and nutrient absorption [39,54,55]. Disrupted gut flora in patients with celiac disease has been linked to sleep disturbances [56]. Future research is warranted especially focusing on patients with celiac disease. Additionally, digital health tools offer the potential for health monitoring and early diagnosis [57].

Limitations

In this study, we investigated the association between daily nutrient intake and sleep using isotemporal substitution models and compositional data analyses to account for the interdependencies [25]. These approaches enabled us to assess how substituting macronutrients and considering the overall composition ratio affect sleep outcomes. Nonetheless, this study had several limitations. First, this cross-sectional study does not provide causal explanations. As noted above, previous studies have reported bidirectional relationships between these factors [58]. Additionally, the isotemporal substitution model and compositional data analysis are merely statistical inferences [25]. The findings of this study underscore the need for future interventions and longitudinal studies to establish causal effects definitively.

Second, this study may have included a highly health-conscious group that uses both health management apps simultaneously. Therefore, although values around the second and third quartiles,

XSL•FO
RenderX

as shown in Table S1 in Multimedia Appendix 1, correspond to Japanese macronutrient standards [59] and suggest the possibility of representing the general population, caution is needed when generalizing the findings to broader populations, owing to the inclusion of this health-conscious group.

Third, the sleep assessment method used in this study involved applying gamified elements. Some users may have exhibited sleep patterns different from their habitual patterns to achieve specific goals within the game. Moreover, because users have their own smartphones, the reliability of accelerometer measurements varies across different device models. In the future, integrating alternative methods, such as tappigraphy for sleep-wake rhythms, actigraphy, or in-home electroencephalogram devices could provide more objective evaluations [13-15].

Fourth, retrospective studies like this one are inherently prone to recall bias, as participants may inaccurately report their dietary intake and sleep patterns based on memory. It should also be noted that dietary intake was not cross-checked, and this study did not separately account for weekdays and weekends in assessing overall nutrient intake. Finally, because this study was retrospective, we could not adequately capture confounding factors related to sleep and nutrition. We collected basic demographic information, such as age, sex, height, weight, and physical activity level. However, other confounding factors, such as alcohol and smoking habits, employment status (including shifts and night shifts), marital status, cohabitation status, medical history, and medication use, were not considered. Therefore, since we could not fully eliminate concerns regarding internal validity, the presence of selection bias cannot be ruled out. Although this is an observational study with a large sample size of real-world data, it is essential to exercise caution in interpreting the results owing to these limitations.

Conclusions

We found that a greater intake of polyunsaturated fats was associated with better SL and reduced WASO, while saturated and monounsaturated fats were correlated with longer SL and wakefulness. Dietary fiber intake was positively associated with improved sleep quality, and a greater sodium-to-potassium ratio was linked to worse sleep outcomes. Furthermore, our compositional data analysis, which accounted for the interdependencies between nutrients, suggested that increasing the protein intake in place of other nutrients could improve the TST, and similar effects were observed for polyunsaturated fats on SL and WASO. Although this study is cross-sectional, the results highlight the complex potential role of dietary factors in sleep regulation and suggest the possibility of dietary interventions to enhance sleep health.

Acknowledgments

We express our gratitude to the personnel of The Pokémon Company (Tokyo, Japan), Asken Inc (Tokyo, Japan), and S'UIMIN Inc (Tokyo, Japan) for their contributions to data preparation. We also thank Editage for the English language editing. This work was supported by the World Premier International Research Center Initiative (WPI) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to MY; Japan Agency for Medical Research and Development (AMED) (JP21zf0127005) to MI and MY; and the COI STREAM initiative launched in 2013 by MEXT, as well as the COI-NEXT initiative launched in 2020 by MEXT (JPMJPF2017) to JS.

Data Availability

The data used in this study belong to The Pokémon Company (Tokyo, Japan) and Asken Inc (Tokyo, Japan). These data are not publicly available but can be used in future projects. Proposals and requests for data access were sent to the corresponding authors via email.

Authors' Contributions

JS, MI, and MY contributed to the conceptualization of the study. The methodology was developed by JS, MI, and MCTK, while JS and MI conducted the formal analysis. JS was responsible for data curation. JS prepared the original draft, with JS, MI, MCTK, and MY contributing to its review and editing. JS handled visualization, while MI and MY provided supervision. Funding acquisition was managed by MY. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

MY was paid by The Pokémon Company (Tokyo, Japan) for a consultation to develop the Pokémon Sleep app. The authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary Information.
[DOCX File , 41 KB - jmir_v27i1e64749_app1.docx]

References

XSL•FO
RenderX

1. Zhang R, Zhang B, Shen C, Sahakian BJ, Li Z, Zhang W, et al. Associations of dietary patterns with brain health from behavioral, neuroimaging, biochemical and genetic analyses. Nat Mental Health 2024;2(5):535-552. [doi: 10.1038/s44220-024-00226-0]
2. Andreeva VA, Perez-Jimenez J, St-Onge MP. A systematic review of the bidirectional association between consumption of ultra-processed food and sleep parameters among adults. Curr Obes Rep 2023;12(4):439-452. [doi: 10.1007/s13679-023-00512-5] [Medline: 37477854]
3. Wilson K, St-Onge MP, Tasali E. Diet composition and objectively assessed sleep quality: a narrative review. J Acad Nutr Diet 2022;122(6):1182-1195 [FREE Full text] [doi: 10.1016/j.jand.2022.01.007] [Medline: 35063665]
4. Lindseth G, Lindseth P, Thompson M. Nutritional effects on sleep. West J Nurs Res 2013;35(4):497-513 [FREE Full text] [doi: 10.1177/0193945911416379] [Medline: 21816963]
5. Lindseth G, Murray A. Dietary macronutrients and sleep. West J Nurs Res 2016;38(8):938-958 [FREE Full text] [doi: 10.1177/0193945916643712] [Medline: 27170039]
6. Godos J, Grosso G, Castellano S, Galvano F, Caraci F, Ferri R. Association between diet and sleep quality: a systematic review. Sleep Med Rev 2021;57:101430. [doi: 10.1016/j.smrv.2021.101430] [Medline: 33549913]
7. Zuraikat FM, Wood RA, Barragán R, St-Onge MP. Sleep and diet: mounting evidence of a cyclical relationship. Annu Rev Nutr 2021;41:309-332 [FREE Full text] [doi: 10.1146/annurev-nutr-120420-021719] [Medline: 34348025]
8. Meyhöfer S, Chamorro R, Hallschmid M, Spyra D, Klinsmann N, Schultes B, et al. Late, but not early, night sleep loss compromises neuroendocrine appetite regulation and the desire for food. Nutrients 2023;15(9):2035 [FREE Full text] [doi: 10.3390/nu15092035] [Medline: 37432152]
9. Markwald RR, Melanson EL, Smith MR, Higgins J, Perreault L, Eckel RH, et al. Impact of insufficient sleep on total daily energy expenditure, food intake, and weight gain. Proc Natl Acad Sci U S A 2013;110(14):5695-5700 [FREE Full text] [doi: 10.1073/pnas.1216951110] [Medline: 23479616]
10. St-Onge MP, Mikic A, Pietrolungo CE. Effects of diet on sleep quality. Adv Nutr 2016;7(5):938-949 [FREE Full text] [doi: 10.3945/an.116.012336] [Medline: 27633109]
11. Bailey RL. Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. Curr Opin Biotechnol 2021;70:91-96 [FREE Full text] [doi: 10.1016/j.copbio.2021.02.007] [Medline: 33714006]
12. Matsuzaki E, Michie M, Kawabata T. Validity of nutrient intakes derived from an internet website dish-based dietary record for self-management of weight among Japanese women. Nutrients 2017;9(10):1058 [FREE Full text] [doi: 10.3390/nu9101058] [Medline: 28946648]
13. Seol J, Chiba S, Kawana F, Tsumoto S, Masaki M, Tominaga M, et al. Validation of sleep-staging accuracy for an in-home sleep electroencephalography device compared with simultaneous polysomnography in patients with obstructive sleep apnea. Sci Rep 2024;14(1):3533 [FREE Full text] [doi: 10.1038/s41598-024-53827-1] [Medline: 38347028]
14. Iwagami M, Seol J, Hiei T, Tani A, Chiba S, Kanbayashi T, et al. Association between electroencephalogram-based sleep characteristics and physical health in the general adult population. Sci Rep 2023;13(1):21545 [FREE Full text] [doi: 10.1038/s41598-023-47979-9] [Medline: 38066043]
15. Borger JN, Huber R, Ghosh A. Capturing sleep-wake cycles by using day-to-day smartphone touchscreen interactions. NPJ Digit Med 2019;2:73 [FREE Full text] [doi: 10.1038/s41746-019-0147-4] [Medline: 31372507]
16. Wirth J, Hillesheim E, Brennan L. Protein intake and its effect on sleep outcomes: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev 2023;81(3):333-345. [doi: 10.1093/nutrit/nuac061] [Medline: 36083207]
17. Wirth J, Lin K, Brennan L, Wu K, Giovannucci E. Protein intake and its association with sleep quality: results from 3 prospective cohort studies. Eur J Clin Nutr 2024;78(5):413-419. [doi: 10.1038/s41430-024-01414-y] [Medline: 38454152]
18. Gao Q, Kou T, Zhuang B, Ren Y, Dong X, Wang Q. The association between vitamin D deficiency and sleep disorders: a systematic review and meta-analysis. Nutrients 2018;10(10):1395 [FREE Full text] [doi: 10.3390/nu10101395] [Medline: 30275418]
19. Luo J, Ge H, Sun J, Hao K, Yao W, Zhang D. Associations of dietary ω-3, ω-6 fatty acids consumption with sleep disorders and sleep duration among adults. Nutrients 2021;13(5):1475 [FREE Full text] [doi: 10.3390/nu13051475] [Medline: 33925486]
20. Decoeur F, Benmamar-Badel A, Leyrolle Q, Persillet M, Layé S, Nadjar A. Dietary N-3 PUFA deficiency affects sleep-wake activity in basal condition and in response to an inflammatory challenge in mice. Brain Behav Immun 2020;85:162-169. [doi: 10.1016/j.bbi.2019.05.016] [Medline: 31100369]
21. Vlahoyiannis A, Giannaki CD, Sakkas GK, Aphamis G, Andreou E. A systematic review, meta-analysis and meta-regression on the effects of carbohydrates on sleep. Nutrients 2021;13(4):1283 [FREE Full text] [doi: 10.3390/nu13041283] [Medline: 33919698]
22. Ibsen DB, Laursen ASD, Würtz AML, Dahm CC, Rimm EB, Parner ET, et al. Food substitution models for nutritional epidemiology. Am J Clin Nutr 2021;113(2):294-303 [FREE Full text] [doi: 10.1093/ajcn/nqaa315] [Medline: 33300036]
23. Tomova GD, Gilthorpe MS, Tennant PW. Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology. Am J Clin Nutr 2022;116(5):1379-1388 [FREE Full text] [doi: 10.1093/ajcn/nqac188] [Medline: 36223891]

XSL•FO
RenderX

24. Corrêa Leite ML. Compositional data analysis as an alternative paradigm for nutritional studies. Clin Nutr ESPEN 2019;33:207-212. [doi: 10.1016/j.clnesp.2019.05.011] [Medline: 31451263]
25. Zhao J, Li Z, Gao Q, Zhao H, Chen S, Huang L, et al. A review of statistical methods for dietary pattern analysis. Nutr J 2021;20(1):37 [FREE Full text] [doi: 10.1186/s12937-021-00692-7] [Medline: 33874970]
26. Standard tables of food composition in Japan -2020 (eighth revised version). Ministry of Education, Culture, Sports, Science and Technology. 2020. URL: https://www.mext.go.jp/a_menu/syokuhinseibun/index.htm [accessed 2024-10-01]
27. Shinozaki N, Murakami K. Evaluation of the ability of diet-tracking mobile applications to estimate energy and nutrient intake in Japan. Nutrients 2020;12(11):3327 [FREE Full text] [doi: 10.3390/nu12113327] [Medline: 33138088]
28. Imamura M, Sasaki H, Shinto T, Tahara Y, Makino S, Kuwahara M, et al. Association between Na, K, and lipid intake in each meal and blood pressure. Front Nutr 2022;9:853118 [FREE Full text] [doi: 10.3389/fnut.2022.853118] [Medline: 35308273]
29. Zuraikat FM, Makarem N, Liao M, St-Onge MP, Aggarwal B. Measures of poor sleep quality are associated with higher energy intake and poor diet quality in a diverse sample of women from the go red for women strategically focused research network. J Am Heart Assoc 2020;9(4):e014587 [FREE Full text] [doi: 10.1161/JAHA.119.014587] [Medline: 32063123]
30. Jung JY, Oh CM, Kim E, Park SK. Dietary sodium intake and its relation to sleep duration, sleep quality and nocturnal urination in working-aged Korean adults. Nutr Bull 2023;48(3):365-375. [doi: 10.1111/nbu.12629] [Medline: 37458133]
31. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
32. Pokémon Sleep. The Pokémon Company. URL: https://support.pokemon.com/hc/en-us/categories/360005098972-Pok%C3%A9mon-Sleep [accessed 2024-06-18]
33. Patel P, Kim JY, Brooks LJ. Accuracy of a smartphone application in estimating sleep in children. Sleep Breath 2017;21(2):505-511. [doi: 10.1007/s11325-016-1425-x] [Medline: 27771844]
34. Fino E, Mazzetti M. Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. Sleep Breath 2019;23(1):13-24. [doi: 10.1007/s11325-018-1661-3] [Medline: 29687190]
35. Natale V, Drejak M, Erbacci A, Tonetti L, Fabbri M, Martoni M. Monitoring sleep with a smartphone accelerometer. Sleep Biol Rhythm 2012;10:287-292. [doi: 10.1111/j.1479-8425.2012.00575.x]
36. Zhao J, Guo W, Wang J, Wang T. Exploring the association of dietary patterns with the risk of hypertension using principal balances analysis and principal component analysis. Public Health Nutr 2023;26(1):160-170 [FREE Full text] [doi: 10.1017/S136898002200091X] [Medline: 35416143]
37. Arnold KF, Berrie L, Tennant PWG, Gilthorpe MS. A causal inference perspective on the analysis of compositional data. Int J Epidemiol 2020;49(4):1307-1313 [FREE Full text] [doi: 10.1093/ije/dyaa021] [Medline: 32154892]
38. Gupta N, Mathiassen SE, Mateu-Figueras G, Heiden M, Hallman DM, Jørgensen MB, et al. A comparison of standard and compositional data analysis in studies addressing group differences in sedentary behavior and physical activity. Int J Behav Nutr Phys Act 2018;15(1):53 [FREE Full text] [doi: 10.1186/s12966-018-0685-1] [Medline: 29903009]
39. Sejbuk M, Siebieszuk A, Witkowska AM. The role of gut microbiome in sleep quality and health: dietary strategies for microbiota support. Nutrients 2024;16(14):2259 [FREE Full text] [doi: 10.3390/nu16142259] [Medline: 39064702]
40. Szentirmai ?, Millican NS, Massie AR, Kapás L. Butyrate, a metabolite of intestinal bacteria, enhances sleep. Sci Rep 2019;9(1):7035 [FREE Full text] [doi: 10.1038/s41598-019-43502-1] [Medline: 31065013]
41. Aslan Çin NN, Yardimci H. Association of total energy intake, diet quality and sleep disorders in university-term female students. Sleep Biol Rhythms 2021;19(3):313-323. [doi: 10.1007/s41105-021-00320-1]
42. Zhou J, Kim JE, Armstrong CL, Chen N, Campbell WW. Higher-protein diets improve indexes of sleep in energy-restricted overweight and obese adults: results from 2 randomized controlled trials. Am J Clin Nutr 2016;103(3):766-774 [FREE Full text] [doi: 10.3945/ajcn.115.124669] [Medline: 26864362]
43. Grandner MA, Jackson N, Gerstner JR, Knutson KL. Sleep symptoms associated with intake of specific dietary nutrients. J Sleep Res 2014;23(1):22-34 [FREE Full text] [doi: 10.1111/jsr.12084] [Medline: 23992533]
44. Baidoo VYA, Alexandria SJ, Zee PC, Knutson KL. The association between timing of dietary macronutrient and sodium consumption and sleep duration and quality. Sleep Adv 2024;5(1):zpae007 [FREE Full text] [doi: 10.1093/sleepadvances/zpae007] [Medline: 38314117]
45. Kang J, Park M, Oh CM, Kim T. High-fat diet-induced dopaminergic dysregulation induces REM sleep fragmentation and ADHD-like behaviors. Psychiatry Res 2023;327:115412 [FREE Full text] [doi: 10.1016/j.psychres.2023.115412] [Medline: 37607442]
46. Sartorius T, Ketterer C, Kullmann S, Balzer M, Rotermund C, Binder S, et al. Monounsaturated fatty acids prevent the aversive effects of obesity on locomotion, brain activity, and sleep behavior. Diabetes 2012;61(7):1669-1679 [FREE Full text] [doi: 10.2337/db11-1521] [Medline: 22492529]
47. Bennett CJ, Cain SW, Blumfield ML. Monounsaturated fat intake is associated with improved sleep quality in pregnancy. Midwifery 2019;78:64-70. [doi: 10.1016/j.midw.2019.07.019] [Medline: 31376641]
48. St-Onge MP, Roberts A, Shechter A, Choudhury AR. Fiber and saturated fat are associated with sleep arousals and slow wave sleep. J Clin Sleep Med 2016;12(1):19-24 [FREE Full text] [doi: 10.5664/jcsm.5384] [Medline: 26156950]

XSL•FO
RenderX

49. Rasaei N, Samadi M, Khadem A, Badrooj N, Hassan Zadeh M, Ghaffarian-Ensaf R, et al. The association between cholesterol/saturated fat index (CSI) and quality of sleep, and circadian rhythm among overweight and obese women: a cross-sectional study. J Health Popul Nutr 2023;42(1):75 [FREE Full text] [doi: 10.1186/s41043-023-00414-1] [Medline: 37501196]
50. Martinez SM, Tschann JM, Butte NF, Gregorich SE, Penilla C, Flores E, et al. Short sleep duration is associated with eating more carbohydrates and less dietary fat in Mexican American children. Sleep 2017;40(2):zsw057 [FREE Full text] [doi: 10.1093/sleep/zsw057] [Medline: 28364488]
51. Ancoli-Israel S, Bliwise DL, Nørgaard JP. The effect of nocturia on sleep. Sleep Med Rev 2011;15(2):91-97 [FREE Full text] [doi: 10.1016/j.smrv.2010.03.002] [Medline: 20965130]
52. Tabara Y. Sodium-to-potassium ratio and renal functional decline. Hypertens Res 2024;47(1):245-246. [doi: 10.1038/s41440-023-01431-7] [Medline: 37714955]
53. Yang Q, Liu T, Kuklina EV, Flanders WD, Hong Y, Gillespie C, et al. Sodium and potassium intake and mortality among US adults: prospective data from the third national health and nutrition examination survey. Arch Intern Med 2011;171(13):1183-1191. [doi: 10.1001/archinternmed.2011.257] [Medline: 21747015]
54. Chibbar R, Dieleman LA. The gut microbiota in celiac disease and probiotics. Nutrients 2019;11(10):2375 [FREE Full text] [doi: 10.3390/nu11102375] [Medline: 31590358]
55. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. BMJ 2018;361:k2179 [FREE Full text] [doi: 10.1136/bmj.k2179] [Medline: 29899036]
56. Zielinski MR, Systrom DM, Rose NR. Fatigue, sleep, and autoimmune and related disorders. Front Immunol 2019;10:1827 [FREE Full text] [doi: 10.3389/fimmu.2019.01827] [Medline: 31447842]
57. Pennisi G, Lanza G, Giuffrida S, Vinciguerra L, Puglisi V, Cantone M, et al. Excitability of the motor cortex in de novo patients with celiac disease. PLoS One 2014;9(7):e102790 [FREE Full text] [doi: 10.1371/journal.pone.0102790] [Medline: 25062250]
58. St-Onge MP, Cherta-Murillo A, Darimont C, Mantantzis K, Martin FP, Owen L. The interrelationship between sleep, diet, and glucose metabolism. Sleep Med Rev 2023;69:101788 [FREE Full text] [doi: 10.1016/j.smrv.2023.101788] [Medline: 37156196]
59. Dietary reference intakes for Japanese (2020 Edition): report of the committee for the formulation. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/content/001151422.pdf [accessed 2024-07-16]

Abbreviations

%WASO: percentage of wakefulness after sleep onset
ilr: isometric log-ratio
SL: sleep latency
TST: total sleep time
WASO: wakefulness after sleep onset

Edited by YH Lin; submitted 25.07.24; peer-reviewed by V Ansu, G Lanza, J Heshmati; comments to author 26.09.24; revised version received 15.10.24; accepted 12.11.24; published 30.01.25.

Please cite as:
Seol J, Iwagami M, Kayamare MCT, Yanagisawa M
Relationship Among Macronutrients, Dietary Components, and Objective Sleep Variables Measured by Smartphone Apps: Real-World Cross-Sectional Study
J Med Internet Res 2025;27:e64749
URL: https://www.jmir.org/2025/1/e64749
doi:10.2196/64749
PMID:

©Jaehoon Seol, Masao Iwagami, Megane Christiane Tawylum Kayamare, Masashi Yanagisawa. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Oura Ring as a Tool for Ovulation Detection: Validation Analysis

Nina Thigpen[1], PhD; Shyamal Patel[1], PhD; Xi Zhang[1], PhD

Oura Ring, San Francisco, CA, United States

Corresponding Author:
Nina Thigpen, PhD
Oura Ring
60 Francisco St
San Francisco, CA, 94133
United States
Phone: 1 358 50 4668390
Email: nnt7845@gmail.com

Abstract

Background: Oura Ring is a wearable device that estimates ovulation dates using physiology data recorded from the finger. Estimating the ovulation date can aid fertility management for conception or nonhormonal contraception and provides insights into follicular and luteal phase lengths. Across the reproductive lifespan, changes in these phase lengths can serve as a biomarker for reproductive health.

Objective: We assessed the strengths, weaknesses, and limitations of using physiology from the Oura Ring to estimate the ovulation date. We compared performance across cycle length, cycle variability, and participant age. In each subgroup, we compared the algorithm's performance with the traditional calendar method, which estimates the ovulation date based on an individual's last period start date and average menstrual cycle length.

Methods: The study sample contained 1155 ovulatory menstrual cycles from 964 participants recruited from the Oura Ring commercial database. Ovulation prediction kits served as a benchmark to evaluate the performance. The Fisher test was used to determine an odds ratio to assess if ovulation detection rate significantly differed between methods or subgroups. The Mann-Whitney *U* test was used to determine if the accuracy of the estimated ovulation date differed between the estimated and reference ovulation dates.

Results: The physiology method detected 1113 (96.4%) of 1155 ovulations with an average error of 1.26 days, which was significantly lower (U=904942.0, P<.001) than the calendar method's average error of 3.44 days. The physiology method had significantly better accuracy across all cycle lengths, cycle variability groups, and age groups compared with the calendar method (P<.001). The physiology method detected fewer ovulations in short cycles (odds ratio 3.56, 95% CI 1.65-8.06; P=.008) but did not differ between typical and long or abnormally long cycles. Abnormally long cycle lengths were associated with decreased accuracy (U=22,383, P=.03), with a mean absolute error of 1.7 (SEM .09) days compared with 1.18 (SEM .02) days. The physiology method was not associated with differences in accuracy across age or typical cycle variability, while the calendar method performed significantly worse in participants with irregular cycles (U=21,643, P<.001).

Conclusions: The physiology method demonstrated superior accuracy over the calendar method, with approximately 3-fold improvement. Calendar-based fertility tracking could be used as a backup in cases of insufficient physiology data but should be used with caution, particularly for individuals with irregular menstrual cycles. Our analyses suggest the physiology method can reliably estimate ovulation dates for adults aged 18-52 years, across a variety of cycle lengths, and in users with regular or irregular cycles. This method may be used as a tool to improve fertile window estimation, which can aid in conceiving or preventing pregnancies. This method also offers a low-effort solution for follicular and luteal phase length tracking, which are key biomarkers for reproductive health.

(J Med Internet Res 2025;27:e60667) doi:10.2196/60667

KEYWORDS

ovulation; digital medicine; physiology; body temperature; menstrual cycles; wearable; fertility; nonhormonal contraception; reproductive health; women's health; calendar method; mHealth; mobile health; detection

XSL•FO
RenderX

Introduction

With advances in wearable technology, an increasing number of women are leveraging continuous, automated sensor data to track ovulation [1]. Ovulation tracking serves multiple purposes; notably, it can be used to delineate the end of the fertile window to optimize the timing of intercourse for conception or contraception. Studies have shown that ovulation tracking can increase the likelihood of conception by approximately 40% for those trying to conceive [2-5]. Conversely, fertility awareness–based methods for contraception, although less effective than hormonal contraception methods [6-8], is increasing popular as the use of hormonal contraception declines [9-12]. Beyond fertility, the ovulation date, alongside period start dates, can be used to measure a cycle's follicular and luteal lengths. These phase lengths have been identified as biomarkers of menstrual health [13], with early research suggesting they may provide insights into fecundity [14], systemic inflammation [15], endometrial development [16], and ovarian aging [17].

Approximately 75% of individuals attempting to conceive use some method to estimate their ovulation date; these methods include cervical mucus monitoring, detecting a rise in basal body temperature (BBT), and feeling for changes in the cervix [18]. These methods vary in accuracy across subgroups including cycle regularity, compliance, and interpretation [19]. Among these, cervical mucus tracking is associated with the highest reported accuracy, with 48% to 76% of ovulations correctly estimated within 1 day of the reference ovulation [19,20]. Approximately 21% of women trying to conceive use the BBT method [18], which, although less precise than cervical mucus tracking at predicting imminent ovulation, has been found to accurately detect that ovulation has occurred [21-23]. All these methods—cervical mucus tracking, feeling for changes in the cervix, and BBT—require significant user knowledge and active participation. In contrast, wearable-based ovulation tracking may provide a more convenient alternative for those seeking a less labor-intensive approach to monitoring ovulation [1].

Previous research has shown that wearables can detect ovulation-related physiological changes. These include wrist-worn devices like Ava [24], an in-ear thermometer [25], vaginal biosensors [26,27], and Oura Ring [28]. While temperature remains the primary continuously measured physiological metric for ovulation detection, other metrics such as heart rate, breath rate, and heart rate variability have also shown predictive value [1]. Studies indicate that intraoral and ear-based temperature methods are generally less accurate for ovulation detection compared with wrist and ring-based wearables [29], which provide more stable temperature readings, especially when measuring distal temperature during sleep [6,24,30]. The primary objective of these wearable technologies, thus far, has been to support conception or offer nonhormonal contraception by using changes in ovulation-related physiology to identify fertile days in a user's cycle.

We assess the performance of the physiology method using Oura Ring of ovulation estimation against the traditional calendar method [31]. We hypothesized that the physiology method would outperform the calendar method in participants with high cycle variability, and the physiology method's ovulation detection rate would decrease with participant age, reflecting age-related declines in progesterone levels and consequently luteal temperature rise magnitude [32]. Ultimately, we aim to review the strengths, weaknesses, and limitations of using Oura Ring physiology for ovulation date estimation.

Methods

Reference Ovulation Dates

Oura Ring commercial users can choose to self-report the results of home luteinizing hormone (LH) tests through the Oura Ring app. Positive LH test results served as the benchmark reference for algorithm performance testing. The reference ovulation date was defined as the day after the last positive LH test in the menstrual cycle [19,33].

Data Inclusion and Exclusion Criteria

Participants were selected for the study if they self-reported a positive LH test result between January 1, 2019, and April 15, 2024. Inclusion criteria required the positive LH test date to be logged within a complete menstrual cycle, such that menses start and end dates were both logged. The self-reported menses and reference LH test dates were required to demarcate biologically plausible cycle phase lengths. Acceptable ranges were 10-90 days for the follicular phase and 8-20 days for the luteal phase [33]. If a single user had multiple complete menstrual cycles containing an LH test reference, these were included as long as the algorithm input window did not overlap. These inclusion criteria resulted in 121 users with 2 or more observations, for a total of 1209 ovulatory menstrual cycles from 1051 unique users.

We excluded menstrual cycles from the analysis based on insufficient physiology data, hormone use, or self-reported pregnancy. Insufficient data were defined as more than 40% of missing physiology data in the last 60 days. Hormone use was self-reported by a questionnaire, where participants were asked to report any hormone intake that could influence ovulation, including hormonal birth control, hormone replacement therapy, or fertility medications. Of these exclusion criteria, 47 cycles were excluded due to insufficient physiology data, 6 due to hormone use, and 1 due to pregnancy. The resulting dataset contained 1155 ovulatory cycles from 1007 unique participants.

Ovulation Detection Algorithms

Calendar Method

We estimated each cycle's ovulation date using the calendar method as a comparison with the physiology method results. The calendar method estimates the ovulation date by subtracting the typical luteal length from the user's typical cycle length. The population mean is used as the typical luteal length, that is, 12 days [33,34]. One additional day is subtracted because we define the ovulation date as the last follicular day, which is standard in the literature [33,34]. The user's typical cycle length is defined here as the median cycle length across the last 6 months, with outliers shorter than 12 days or greater than 90 days excluded.

XSL•FO
RenderX

Physiology Method

The physiology method is an algorithm written in Python that uses signal processing techniques to analyze continuously recorded finger temperature from Oura Ring to estimate the date of the most recent ovulation event. The ring contains negative temperature coefficient thermistors, which are used to measure skin temperature. Users are instructed to wear Oura Ring on a finger where the ring feels tight but not uncomfortable.

The algorithm was developed by assessing a separate training dataset of 30,000 menstrual cycles, which contained no overlapping users or menstrual cycles as the test set presented here. The algorithm aims to identify a maintained rise in skin temperature by around 0.3-0.7 °C, which has shown to be a robust postovulatory change following the regression of the dominant follicle [32]. The algorithm was tuned using a grid search across a set of parameters on the training dataset to optimize for detecting the rise in temperature following ovulation, as determined based on visual inspection. The physiology algorithm first normalizes the dataset by centering it around 0. Then, outliers were rejected, defined as >2 SD from the population average. Any missing or rejected data were then imputed using a linear fill. A Butterworth bandpass filter was then applied, with the low pass, high pass, and order set as parameters to be tuned in the grid search. Finally, hysteresis thresholding was used to determine the likely follicular and luteal phase days of the cycle, the parameters of which were also tuned in the grid search.

Algorithm postprocessing included combining the temperature-estimated luteal phase with self-reported period start logs. The algorithm rejects ovulation detections if they resulted in biologically implausible phase lengths [33,35], that is, luteal phases outside the range of 7-17 days or follicular phases outside the range of 10-90 days. In these cases, the physiology algorithm labels these cases as failures to detect ovulation. For details, refer to Figure 1.

Figure 1. The physiology algorithm pipeline.

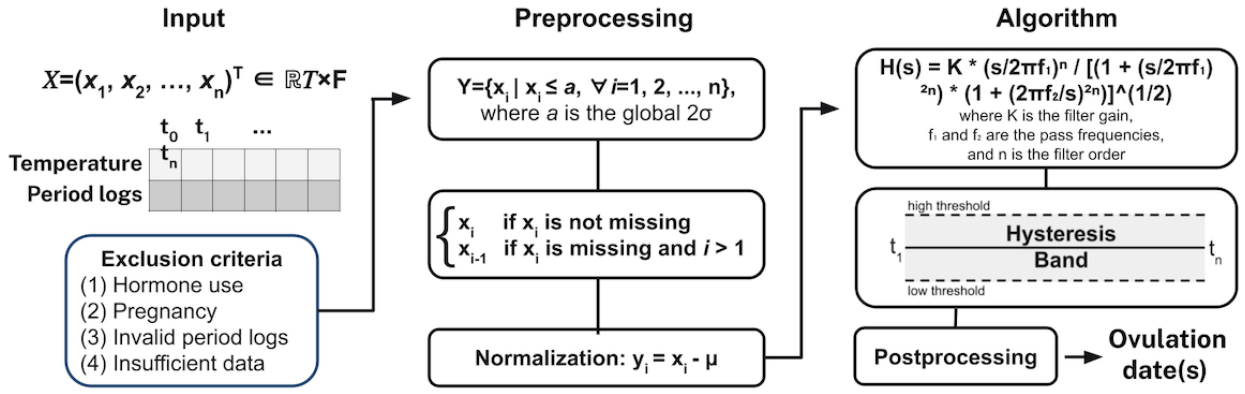

Statistical Analysis

Detection Rate

The ovulation detection rate was defined as the proportion of ovulatory cycles in which the algorithm correctly identified an ovulation. We calculated the 95% CI for the detection rate using the *z*-score–based margin of error for proportions [36]. To assess statistical differences in the ovulation detection rate across subgroups, we used the Fisher exact test. This nonparametric statistical method is well-suited for comparing proportion data, especially in datasets where subgroups vary in sample size [37]. For each subgroup of interest, such as cycle variation, we designated a reference group and used the Fisher exact test to compare the detection rates between the reference group with all other groups. For example, participants with irregular cycles were compared with participants with regular cycles as a reference group. Adjustments for multiple comparisons were made using the Bonferroni correction [38].

Accuracy

The ovulation date error was defined as the number of days between the algorithm's estimated ovulation date and the reference ovulation date. The reference ovulation date was determined as the day following the last positive or peak LH surge reported in that cycle. To quantify the typical accuracy, we calculated the mean absolute error (MAE). The 95% CI on this mean was estimated using 1.96 times the SEM [39]. To evaluate statistical differences in ovulation date error across various subgroups, we used the Mann-Whitney *U* test. This test is preferred for providing a robust analysis with minimal influence from outliers and nonnormal distributions, in contrast to parametric methods. For each subgroup of interest, such as cycle variation, one group is designated as the reference group to be compared against all other groups. For example, we compare participants with irregular cycles to participants with regular cycles as a reference group. Adjustments for multiple comparisons were made using a Bonferroni correction [38].

Subgroups

Cycle Length

Detection rate and accuracy were broken down by cycle length to compare ovulation estimation performance in menstrual cycles of various lengths. Cycle length was defined as the number of days between 2 menses onsets. The cycle length was classified as abnormal if it was shorter than 21 days or longer than 35 days [40]. Of the remaining cycle lengths, we defined

XSL•FO
RenderX

a “short” cycle as lasting between 21-26 days, a “typical” cycle as lasting between 27-31 days, and a “long” cycle as lasting between 32-35 days (Table 1). To statistically compare ovulation detection performance metrics across groups, we compared performance metrics in cycles of typical length with all other cycle length groups.

Table 1. Cycle length.

Cycle length	Days, n	Cycles (n=1155), n (%)	Physiology method			Calendar method	
			Detection rate (%), mean (SEM)	MAE[a], mean (SEM)	Within 2 days (%)	MAE, mean (SEM)	Within 2 days (%)
Abnormally short	≤21	2 (<1)	—[b]	—	—	—	—
Short	22-26	176 (15.2)	93.2 (3.7)	1.34 (.09)	84.8	2.84 (0.21)	51.7
Medium	27-31	662 (57.3)	98.0 (1.1)	1.18 (.04)	90	3.02 (0.13)	51
Long	32-35	223 (19.3)	97.3 (2.1)	1.26 (.08)	88	3.57 (0.28)	55.6
Abnormally long	≥36	92 (8)	90 (6)	1.70 (.18)	77	7.32 (1.07)	35

[a]MAE: mean absolute error.

[b]Not applicable.

Cycle Variability

Detection rate and accuracy were broken down by a participant’s cycle variability to compare performance in these groups. Cycle variability was defined as the median of the absolute value of cycle length differences between consecutive cycles, across the past 6 months. Abnormal cycle variation is defined as cycle variation larger than 7 days. Of the normal cycle variation, we defined a “typical” cycle variation as between 0-3 days and a “moderately irregular” cycle variation as between 4-7 days (Table 2). To statistically compare ovulation detection performance metrics across groups, we compared the typical cycle variability group with all other groups.

Table 2. Cycle variability.

Cycle variability	Variation (days)	Cycles (n=975), n (%)	Physiology method			Calendar method	
			Detection rate (%), mean (SEM)	MAE[a], mean (SEM)	Within 2 days (%)	MAE, mean (SEM)	Within 2 days (%)
Typical	0-3	648 (67)	97.5 (1.2)	1.23 (0.05)	88.3	2.51 (0.08)	55.9
Moderately Irregular	4-7	210 (21)	96.0 (2.7)	1.23 (0.08)	90.7	3.48 (0.22)	43.8
Abnormally Irregular	>7	117 (12)	94.0 (4.3)	1.48 (0.14)	81.8	6.63 (0.65)	32.5

[a]MAE: mean absolute error.

Age

Age was binned into 4 groups. Specifically, our groups consisted of 18-26 years, 27-34 years, 35-43 years, and ≥44 years (Table 3). Because the group with the most participants fell into “27-34 years” age group, this was designated as the reference group for statistical analysis.

Table 3. Age.

Age group (years)	Cycles (n=1155), n (%)	Physiology method			Calendar method	
		Detection rate (%), mean (SEM)	MAE[a], mean (SEM)	Within 2 days (%)	MAE, mean (SEM)	Within 2 days (%)
18-26	134 (11.6)	97.0 (2.8)	1.35 (0.11)	86.9	3.64 (0.48)	53
27-34	592 (51.3)	96.6 (1.5)	1.29 (0.05)	87.1	3.39 (0.48)	55.4
35-43	373 (32.3)	96.0 (2.0)	1.17 (0.06)	89.1	3.46 (0.21)	45
≥44	56 (5)	95 (6)	1.19 (0.13)	91	3.30 (0.31)	34

[a]MAE: mean absolute error.

Ethical Considerations

Data from adult Oura Ring users were selected from the Oura Ring commercial database. Oura Health Oy (Oulu, Finland)

XSL•FO
RenderX

collected the data in accordance with their terms of use and privacy policy. By agreeing to the privacy policy, the Oura Ring users had also agreed to the use of their data to perform analysis. The data analysis study protocol was reviewed and approved by ŌURA's science and legal team, which is responsible for ethics and regulatory adherence for studies. The study was deemed exempt as the analysis used aggregate data without individual identifiers. All data that were made available were deidentified and are anonymous. Additional protections included allowing data access only upon request for the purpose of ensuring the analyses presented are accurate and complete. Compensation for study participation was not provided.

Results

Participants

The study sample consisted of 964 female participants. Participants had a mean age of 32.8 (SD 5.5) years and a mean self-reported cycle length of 29.1 (SD 3.4) days.

Algorithm Performance

The physiology method from Oura Ring correctly detected 1113 (96.4%) of 1155 ovulatory cycles observed (95% CI 95.3-97.4%). The estimated ovulation date MAE was 1.26 (SEM 0.02) days. Of the total 1155 ovulatory cycles, 1097 (95%) of ovulations were detected within 3 days, 1015 (87.9%) were detected within 2 days, and 785 (68%) were detected within 1 day. By comparison, the calendar method's ovulation MAE was 3.44 (SEM 0.07) days. In total, the calendar method detected 768 (66.5%) of ovulations within 3 days, 585 (50.7%) within 2 days, and 371 (32.2%) within 1 day. The physiology method estimated ovulation with significantly improved accuracy compared with the calendar method (U_{1154}=904942, P<.001, Cliff δ=0.46). Refer to Figure 2 for a breakdown of the algorithm's performance by cycle length, cycle variability, and age.

Figure 2. Ovulation estimation performance metrics across subgroups.

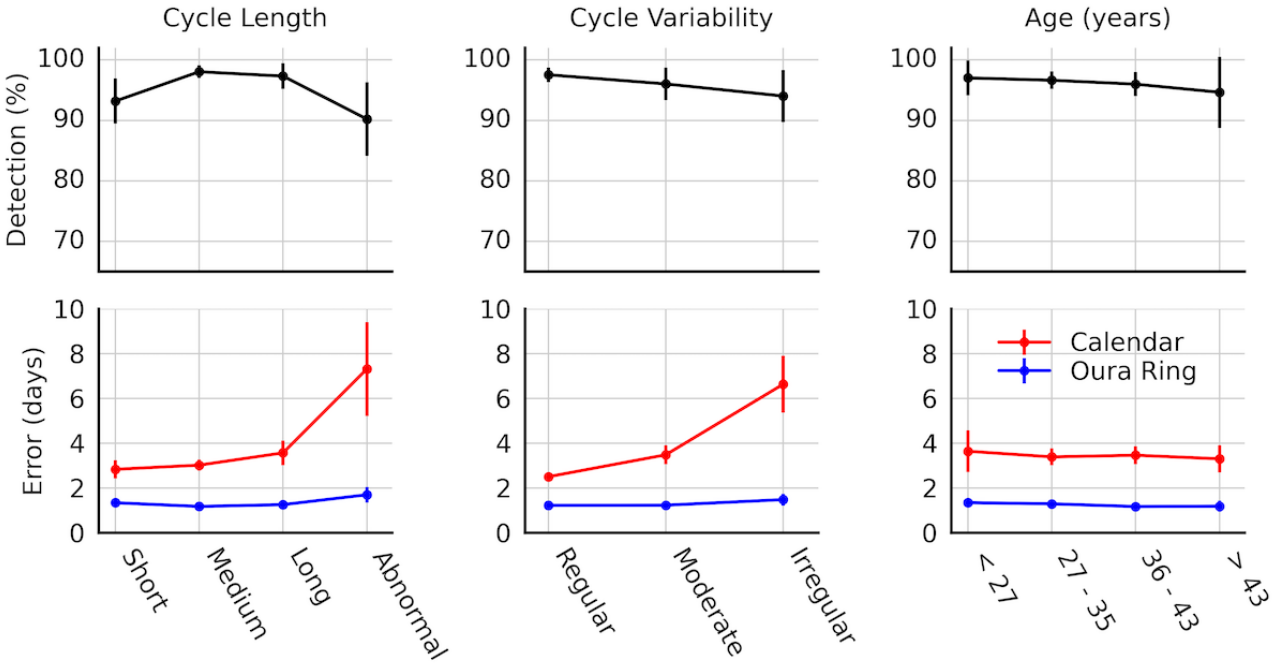

Cycle Length

Ovulation detection rate was impacted by cycle length. The Fisher test for detection rates suggests the odds of detecting ovulation were 3.56 times higher in medium cycle lengths compared with short cycle lengths (odds ratio 3.56, 95% CI 1.65-8.06, P=.008). Long (P=.99) and abnormally long (P=.59) cycle lengths did not statistically differ relative to medium cycle lengths. Accuracy did not significantly differ across regular cycle lengths (short, medium, or long) when using the physiology method (short vs medium: P=.41; long vs medium: P=.99). However, abnormally long cycle lengths ≥36 days) were associated with increased error (U=22,383, P=.03, δ=−0.17) with an MAE of 1.7 (SEM .09) days compared with 1.18 (SEM .02) days. The calendar method also performed significantly worse in abnormally long cycles (U=21,184, P<.001, δ=−0.30), with an MAE of 7.32 (SEM .54) days compared with 3.02 (SEM .07) days.

Cycle Variability

Neither detection rate nor accuracy differed significantly based on cycle variability when using the physiology method (for detection rate, moderate vs regular: P=.33 and irregular vs regular: P=.07; for accuracy, moderate vs regular: P=.99 and irregular vs regular: P=.15). However, the calendar method was associated with significantly worse accuracy for moderately irregular cyclers (U=54,093.5, P<.001, δ=−0.17) and even worse accuracy in irregular cycles (U=21,643, P<.001, δ=−0.43) when compared with typical cycle variability.

XSL•FO
RenderX

Age

Neither the physiology method nor the calendar method was associated with differences across age in ovulation detection rate or estimated ovulation date error.

Discussion

Principal Findings

Finger temperature collected from Oura Ring was formulated as a physiology-based method of estimating the ovulation date. To assess the performance of this algorithm, we compare the physiology method with the calendar method in 1155 ovulatory cycles. The physiology method correctly detected ovulation in 1113 (96.4%) of 1155 cycles with an MAE of 1.26 days. This method was significantly more precise than the calendar method, which exhibited an MAE of 3.44 days, representing a 2.7-fold increase in error. In practical applications, Oura Ring primarily uses the physiology-based method to estimate a cycle's ovulation date. However, the calendar method is used as a secondary approach when the physiology-based method cannot estimate the ovulation date, for example, in instances of excessive missing data. Thus, a thorough understanding of the strengths and limitations of both methods is critical. We also assessed if subgroups including age, cycle variation, and cycle length impact the accuracy of ovulation date estimation.

Age did not influence the accuracy of estimated ovulation dates within the assessed age range of 18-52 years, using either physiology-based or calendar methods. This finding aligns with the existing literature suggesting that age should not significantly affect the performance of the calendar method, as the luteal phase length remains consistent across age groups, and the calendar method estimates ovulation day by subtracting the luteal phase length from the user's typical cycle length [31]. Although it was hypothesized that the physiology method may be less accurate across age due to variations in the magnitude of the postovulatory temperature rise [32,41], our analysis indicates that age did not significantly impact performance. However, ovulation estimation using Oura Ring has not yet been tested in individuals younger than 18 years and thus accuracy should not be assumed for adolescents.

Cycle variability significantly impacted the performance of the calendar method, but not the physiology method. Specifically, participants with irregular cycles experienced poorer performance using the calendar method, with an MAE of 6.63 days. This is a 2.6-fold decrease in accuracy relative to regular cyclers using the calendar method, and a 4.5-fold decrease in accuracy relative to the physiology-based method in irregular cyclers. When framing this as a window of days where the ovulation is 95% likely to have occurred, in irregular cyclers, that window is reduced from 14.2 days to 3.4 days. The superior accuracy of the physiology method likely stems from its independence from follicular phase length, which predominantly drives cycle variability and thus increases error in the calendar method [33,42].

Cycle length significantly influences ovulation estimation performance. Using the calendar method, longer cycles were associated with decreased accuracy, such that the MAE reaches 7.32 days in abnormally long cycles (ie, cycles exceeding 36 days in length). With the physiology method, statistical differences are evident both in the ovulation detection rate and the accuracy of the estimated ovulation date. Specifically, in cycles shorter than 26 days, which constitute 15% of our sample, the detection rate falls to 93%, compared with 98% in cycles of typical length. Furthermore, accuracy diminishes in abnormally long cycles with an MAE of 1.7 days versus 1.18 days in typical cycles. Consequently, users with shorter cycles may experience fewer detected ovulations, while those with abnormally long cycles may experience slightly worse accuracy when using the physiology method. Nevertheless, across all cycle categories, the physiology method significantly outperforms the calendar method.

Comparison With Previous Work

Compared with wrist-worn devices with reported ovulation detection rates between 54% to 86% [43,44], Oura Ring exhibited superior performance, detecting 96.4% of ovulations. Oura Ring accuracy was similar to that of cervical mucus tracking, with 68% of estimations within 1 day of the reference ovulation date compared with 48% to 76% [19,20]. Oura Ring achieved favorable accuracy to BBT, which has a reported 20% detection rate [19,23], with an accuracy of 22% of ovulations correctly falling within 1 day of the reference ovulation [22,23]. Notably, BBT shows a marked decrease in performance in irregular cycles with a 34% drop in accuracy compared with regular cycles [45], while Oura Ring accuracy remained stable. Direct comparisons between studies should be approached with caution, as they are based on datasets with varying compositions of participant subgroups known to influence accuracy; notably, most studies exclude irregular cycles and cycles of abnormal length for performance analysis. The primary objective of this study is to identify conditions that may hinder the performance of Oura Ring ovulation estimation, rather than directly comparing it with other devices.

Limitations

The primary limitation of this study is that our dataset consisted solely of ovulatory cycles and thus we cannot address ovulation estimation performance in anovulatory cycles. Without an estimate of the false alarm rate, a lack of detected ovulation by the algorithm should not be interpreted as indicative of an anovulatory cycle. The algorithm is not intended for use in individuals who frequently experience anovulatory cycles or who do not ovulate due to factors such as hormonal contraceptives.

Another limitation of this study is that our reference ovulation date is based on self-reported positive LH surge from participants using ovulation predictor kits (OPKs). While OPKs impose a lower participant burden, the gold standard for ovulation reference is a transvaginal ultrasound, which offers higher accuracy [19]. Given that our approach for collecting LH test results relied on self-report, the ovulation reference labels used here are susceptible to inaccuracies introduced by human error and subjective reporting biases. This approach also limits our sample to individuals who both use an Oura Ring and opt to take OPKs, and thus our cohort may not be representative of the broader population with menstrual cycles. Furthermore,

XSL•FO
RenderX

the absence of data on negative OPK results precludes our ability to assess anovulatory cycles. In addition, it is not uncommon for an OPK to return a positive LH surge across multiple days; therefore, if participants logged only 1 positive LH surge and did not test on subsequent days, our reference ovulation day might be prematurely assigned.

A further limitation is that we did not collect data on whether participants had a history of menstrual cycle disorders. This is of particular concern in the 12% of participants with irregular cycles, who may have various underlying causes. A potential factor that could impact the accuracy of Oura Ring ovulation estimation is luteal phase deficiency (LPD), characterized by insufficient progesterone exposure [46]. LPD could influence the magnitude of the postovulatory temperature rise and thus impair the algorithm's ability to estimate ovulation dates. Until future research is conducted, the estimated ovulation dates provided by the Oura Ring should be interpreted with caution in individuals with menstrual cycle disorders, particularly conditions associated with LPD, such as polycystic ovary syndrome [47,48], thyroid and prolactin disorders [49], and during in vitro fertilization cycles [50].

Conclusions

Overall, the results suggest that Oura Ring offers significantly improved accuracy over the traditional calendar method for estimating ovulation date. Furthermore, the physiology method is associated with relatively robust accuracy across various ages, cycle variability, and cycle lengths, with only small drops in detecting rate for shorter cycles (<26 days) and slight drops in accuracy in abnormally long cycles (>36 days). This is significantly improved compared with the calendar method, which was associated with large accuracy decreases in users with long cycles and irregular cycles. Based on previous reports, Oura Ring ovulation estimation appears to be outperforming other wearables and BBT-based ovulation estimation methods and on par with the cervical-mucus tracking method, although those results should be compared cautiously given that the performance was measured on different datasets. Furthermore, we posit that widespread ovulation tracking could have broad use beyond its use in conception and nonhormonal contraception. Specifically, continuous monitoring of follicular and luteal phase lengths throughout the reproductive lifespan could benefit a wider demographic interested in menstrual cycle–related biomarkers. These biomarkers hold promise for the early detection of disorders related to the menstrual cycle and fertility.

Data Availability

The dataset analyzed during this study is not publicly available due to privacy protections for participants but may be made available at Oura's sole discretion.

Authors' Contributions

NT designed the algorithm, implemented the algorithm in Python, aggregated the dataset for analysis, conducted the analyses, wrote the first draft of the manuscript, and generated figures. NT and XZ contributed to algorithm design iterations, analysis plan iterations, and data interpretation. All authors contributed to manuscript editing.

Conflicts of Interest

The authors are employees of Oura Health, but this work represents their individual opinions and initiatives. We aim to provide a clear and impartial analysis to ensure that users and research partners are well-informed about the capabilities and limitations of the algorithm.

References

1. Lyzwinski L, Elgendi M, Menon C. Innovative approaches to menstruation and fertility tracking using wearable reproductive health technology: systematic review. J Med Internet Res 2024;26:e45139 [FREE Full text] [doi: 10.2196/45139] [Medline: 38358798]
2. Yeh PT, Kennedy CE, van der Poel SV, Matsaseng T, Bernard L, Narasimhan M. Should home-based ovulation predictor kits be offered as an additional approach for fertility management for women and couples desiring pregnancy? A systematic review and meta-analysis. BMJ Glob Health 2019;4(2):e001403 [FREE Full text] [doi: 10.1136/bmjgh-2019-001403] [Medline: 31139458]
3. Stanford JB, White GL, Hatasaka H. Timing intercourse to achieve pregnancy: current evidence. Obstetrics & Gynecology 2002;100(6):1333-1341. [doi: 10.1097/00006250-200212000-00027]
4. Robinson JE, Wakelin M, Ellis JE. Increased pregnancy rate with use of the clearblue easy fertility monitor. Fertil Steril 2007;87(2):329-334 [FREE Full text] [doi: 10.1016/j.fertnstert.2006.05.054] [Medline: 17074329]
5. Johnson S, Stanford JB, Warren G, Bond S, Bench-Capon S, Zinaman MJ. Increased likelihood of pregnancy using an app-connected ovulation test system: a randomized controlled trial. J Womens Health (Larchmt) 2020;29(1):84-90 [FREE Full text] [doi: 10.1089/jwh.2019.7850] [Medline: 31483187]

XSL•FO
RenderX

6. Pearson JT, Chelstowska M, Rowland SP, Benhar E, Kopp-Kallner H, Berglund Scherwitzl E, et al. Contraceptive effectiveness of an FDA-cleared birth control app: results from the natural cycles U.S. cohort. J Womens Health (Larchmt) 2021;30(6):782-788. [doi: 10.1089/jwh.2020.8547] [Medline: 33370220]
7. Freundl G, Sivin I, Batár I. State-of-the-art of non-hormonal methods of contraception: IV. Natural family planning. Eur J Contracept Reprod Health Care 2010;15(2):113-123. [doi: 10.3109/13625180903545302] [Medline: 20141492]
8. Jennings V, Haile LT, Simmons RG, Spieler J, Shattuck D. Perfect- and typical-use effectiveness of the dot fertility app over 13 cycles: results from a prospective contraceptive effectiveness trial. Eur J Contracept Reprod Health Care 2019;24(2):148-153. [doi: 10.1080/13625187.2019.1581164] [Medline: 30880509]
9. Watson A, Yarger J, Sedlander E, Urbina J, Hopkins K, Rodriguez MI, et al. Concern that contraception affects future fertility: how common is this concern among young people and does it stop them from using contraception? Contracept X 2023;5:100103 [FREE Full text] [doi: 10.1016/j.conx.2023.100103] [Medline: 38162189]
10. Schneider-Kamp A, Takhar J. Interrogating the pill: rising distrust and the reshaping of health risk perceptions in the social media age. Soc Sci Med 2023;331:116081 [FREE Full text] [doi: 10.1016/j.socscimed.2023.116081] [Medline: 37441974]
11. Namasivayam V, Dehury B, Prakash R, Becker M, Anand P, Mishra A, et al. Understanding the rise in traditional contraceptive methods use in Uttar Pradesh, India. Reprod Health 2023;20(1):8 [FREE Full text] [doi: 10.1186/s12978-022-01547-y] [Medline: 36609308]
12. le Guen M, Schantz C, Régnier-Loilier A, de la Rochebrochard E. Reasons for rejecting hormonal contraception in western countries: a systematic review. Soc Sci Med 2021;284:114247 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114247] [Medline: 34339927]
13. Symul L, Wac K, Hillard P, Salathé M. Assessment of menstrual health status and evolution through mobile apps for fertility awareness. NPJ Digit Med 2019;2:64 [FREE Full text] [doi: 10.1038/s41746-019-0139-4] [Medline: 31341953]
14. Crawford NM, Pritchard DA, Herring AH, Steiner AZ. Prospective evaluation of luteal phase length and natural fertility. Fertil Steril 2017;107(3):749-755 [FREE Full text] [doi: 10.1016/j.fertnstert.2016.11.022] [Medline: 28065408]
15. Harris BS, Steiner AZ, Faurot KR, Long A, Jukic AM. Systemic inflammation and menstrual cycle length in a prospective cohort study. Am J Obstet Gynecol 2023;228(2):215.e1-215.e17 [FREE Full text] [doi: 10.1016/j.ajog.2022.10.008] [Medline: 36244407]
16. Bakkensen JB, Christou G, Dimitriadis I, James K, Souter I. The effect of follicular phase length on cycle outcomes and endometrial development in gonadotrophin ovarian stimulation/intrauterine insemination cycles. Reprod Biomed Online 2020;40(3):362-368. [doi: 10.1016/j.rbmo.2019.12.007] [Medline: 32061522]
17. Younis JS, Iskander R, Fauser BCJM, Izhaki I. Does an association exist between menstrual cycle length within the normal range and ovarian reserve biomarkers during the reproductive years? A systematic review and meta-analysis. Hum Reprod Update 2020;26(6):904-928. [doi: 10.1093/humupd/dmaa013] [Medline: 32514566]
18. Stanford JB, Willis SK, Hatch EE, Rothman KJ, Wise LA. Fecundability in relation to use of fertility awareness indicators in a North American preconception cohort study. Fertil Steril 2019;112(5):892-899 [FREE Full text] [doi: 10.1016/j.fertnstert.2019.06.036] [Medline: 31731946]
19. Su HW, Yi YC, Wei TY, Chang TC, Cheng CM. Detection of ovulation, a review of currently available methods. Bioeng Transl Med 2017;2(3):238-246 [FREE Full text] [doi: 10.1002/btm2.10058] [Medline: 29313033]
20. Frank-Herrmann P, Heil J, Gnoth C, Toledo E, Baur S, Pyper C, et al. The effectiveness of a fertility awareness based method to avoid pregnancy in relation to a couple's sexual behaviour during the fertile time: a prospective longitudinal study. Hum Reprod 2007;22(5):1310-1319. [doi: 10.1093/humrep/dem003] [Medline: 17314078]
21. Händel P, Wahlström J. Digital contraceptives based on basal body temperature measurements. Biomedical Signal Processing and Control 2019;52:141-151. [doi: 10.1016/j.bspc.2019.04.019]
22. Barron ML, Fehring RJ. Basal body temperature assessment: is it useful to couples seeking pregnancy? MCN Am J Matern Child Nurs 2005;30(5):290-297. [doi: 10.1097/00005721-200509000-00004] [Medline: 16132004]
23. Bauman JE. Basal body temperature: unreliable method of ovulation detection. Fertil Steril 1981;36(6):729-733 [FREE Full text] [doi: 10.1016/s0015-0282(16)45916-9] [Medline: 7308516]
24. Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B. Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. J Med Internet Res 2019;21(4):e13404 [FREE Full text] [doi: 10.2196/13404] [Medline: 30998226]
25. Luo L, She X, Cao J, Zhang Y, Li Y, Song PXK. Detection and prediction of ovulation from body temperature measured by an In-Ear wearable thermometer. IEEE Trans Biomed Eng 2020;67(2):512-522. [doi: 10.1109/TBME.2019.2916823] [Medline: 31095472]
26. Aptekar D, Costantini L, Katilius J, Webster W. Continuous, passive personal wearable sensor to predict ovulation 21G. Obstet Gynecol 2016;127:64. [doi: 10.1097/01.aog.0000483905.29999.b1]
27. Regidor PA, Kaczmarczyk M, Schiweck E, Goeckenjan-Festag M, Alexander H. Identification and prediction of the fertile window with a new web-based medical device using a vaginal biosensor for measuring the circadian and circamensual core body temperature. Gynecol Endocrinol 2018;34(3):256-260 [FREE Full text] [doi: 10.1080/09513590.2017.1390737] [Medline: 29082805]

XSL•FO
RenderX

28. Maijala A, Kinnunen H, Koskimäki H, Jämsä T, Kangas M. Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable oura ring. BMC Womens Health 2019;19(1):150 [FREE Full text] [doi: 10.1186/s12905-019-0844-9] [Medline: 31783840]
29. Baker FC, Siboza F, Fuller A. Temperature regulation in women: effects of the menstrual cycle. Temperature (Austin) 2020;7(3):226-262 [FREE Full text] [doi: 10.1080/23328940.2020.1735927] [Medline: 33123618]
30. Alzueta E, de Zambotti M, Javitz H, Dulai T, Albinni B, Simon KC, et al. Tracking sleep, temperature, heart rate, and daily symptoms across the menstrual cycle with the oura ring in healthy women. Int J Womens Health 2022;14:491-503 [FREE Full text] [doi: 10.2147/IJWH.S341917] [Medline: 35422659]
31. Johnson S, Marriott L, Zinaman M. Can apps and calendar methods predict ovulation with accuracy? Curr Med Res Opin 2018;34(9):1587-1594. [doi: 10.1080/03007995.2018.1475348] [Medline: 29749274]
32. Tatsumi T, Sampei M, Saito K, Honda Y, Okazaki Y, Arata N, et al. Age-dependent and seasonal changes in menstrual cycle length and body temperature based on big data. Obstet Gynecol 2020;136(4):666-674 [FREE Full text] [doi: 10.1097/AOG.0000000000003910] [Medline: 32925608]
33. Bull JR, Rowland SP, Scherwitzl EB, Scherwitzl R, Danielsson KG, Harper J. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. NPJ Digit Med 2019;2:83 [FREE Full text] [doi: 10.1038/s41746-019-0152-7] [Medline: 31482137]
34. Najmabadi S, Schliep KC, Simonsen SE, Porucznik CA, Egger MJ, Stanford JB. Menstrual bleeding, cycle length, and follicular and luteal phase lengths in women without known subfertility: a pooled analysis of three cohorts. Paediatr Perinat Epidemiol 2020;34(3):318-327 [FREE Full text] [doi: 10.1111/ppe.12644] [Medline: 32104920]
35. Li H, Gibson EA, Jukic AMZ, Baird DD, Wilcox AJ, Curry CL, et al. Menstrual cycle length variation by demographic characteristics from the apple women's health study. NPJ Digit Med 2023;6(1):100 [FREE Full text] [doi: 10.1038/s41746-023-00848-1] [Medline: 37248288]
36. Agresti A, Coull BA. Approximate is better than “Exact” for interval estimation of binomial proportions. The American Statistician 1998;52(2):119-126. [doi: 10.1080/00031305.1998.10480550]
37. Kim HY. Statistical notes for clinical researchers: chi-squared test and fisher's exact test. Restor Dent Endod 2017;42(2):152-155 [FREE Full text] [doi: 10.5395/rde.2017.42.2.152] [Medline: 28503482]
38. Armstrong RA. When to use the bonferroni correction. Ophthalmic Physiol Opt 2014;34(5):502-508. [doi: 10.1111/opo.12131] [Medline: 24697967]
39. Altman DG, Bland JM. Standard deviations and standard errors. BMJ 2005;331(7521):903 [FREE Full text] [doi: 10.1136/bmj.331.7521.903] [Medline: 16223828]
40. Attia GM, Alharbi OA, Aljohani RM. The impact of irregular menstruation on health: a review of the literature. Cureus 2023;15(11):e49146 [FREE Full text] [doi: 10.7759/cureus.49146] [Medline: 38130524]
41. Baerwald A, Vanden Brink H, Hunter C, Beuker D, Lim H, Lee C, et al. Age-related changes in luteal dynamics: preliminary associations with antral follicular dynamics and hormone production during the human menstrual cycle. Menopause 2018;25(4):399-407. [doi: 10.1097/GME.0000000000001021] [Medline: 29112593]
42. McKnight PE, Nujab J. Mann-Whitney U Test.: Wiley Online Library; 2010. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470479216.corpsy0524 [accessed 2024-12-25]
43. Shilaih M, Goodale BM, Falco L, Kübler F, de Clerck V, Leeners B. Modern fertility awareness methods: wrist wearables capture the changes in temperature associated with the menstrual cycle. Biosci Rep 2018;38(6):BSR20171279 [FREE Full text] [doi: 10.1042/BSR20171279] [Medline: 29175999]
44. Zhu TY, Rothenbühler M, Hamvas G, Hofmann A, Welter J, Kahr M, et al. The accuracy of wrist skin temperature in detecting ovulation compared to basal body temperature: prospective comparative diagnostic accuracy study. J Med Internet Res 2021;23(6):e20710 [FREE Full text] [doi: 10.2196/20710] [Medline: 34100763]
45. Yu JT, Su YF, Zhang C, Jin L, Lin XH, Chen LT, et al. Tracking of menstrual cycles and prediction of the fertile window via measurements of basal body temperature and heart rate as well as machine-learning algorithms. Reprod Biol Endocrinol 2022;20(1):118 [FREE Full text] [doi: 10.1186/s12958-022-00993-4] [Medline: 35964035]
46. Palomba S, Santagni S, La Sala GB. Progesterone administration for luteal phase deficiency in human reproduction: an old or new issue? J Ovarian Res 2015;8:77 [FREE Full text] [doi: 10.1186/s13048-015-0205-8] [Medline: 26585269]
47. Palomba S, Falbo A, Russo T, Orio F, Tolino A, Zullo F. Systemic and local effects of metformin administration in patients with polycystic ovary syndrome (PCOS): relationship to the ovulatory response. Hum Reprod 2010;25(4):1005-10013. [doi: 10.1093/humrep/dep466] [Medline: 20106839]
48. Boutzios G, Karalaki M, Zapanti E. Common pathophysiological mechanisms involved in luteal phase deficiency and polycystic ovary syndrome. impact on fertility. Endocrine 2013;43(2):314-317. [doi: 10.1007/s12020-012-9778-9] [Medline: 22930247]
49. Daly DC, Walters CA, Soto-Albors CE, Riddick DH. Endometrial biopsy during treatment of luteal phase defects is predictive of therapeutic outcome. Fertil Steril 1983;40(3):305-310 [FREE Full text] [doi: 10.1016/s0015-0282(16)47291-2] [Medline: 6884533]
50. Tavaniotou A, Albano C, Smitz J, Devroey P. Impact of ovarian stimulation on corpus luteum function and embryonic implantation. J Reprod Immunol 2002;55(1-2):123-130. [doi: 10.1016/s0165-0378(01)00134-6] [Medline: 12062827]

XSL•FO
RenderX

Abbreviations

BBT: basal body temperature
LH: luteinizing hormone
LPD: luteal phase deficiency
MAE: mean absolute error
OPK: ovulation predictor kit

Edited by N Cahill; submitted 17.05.24; peer-reviewed by F Wever, J Grieger; comments to author 04.09.24; revised version received 17.09.24; accepted 11.12.24; published 31.01.25.

Please cite as:
Thigpen N, Patel S, Zhang X
Oura Ring as a Tool for Ovulation Detection: Validation Analysis
J Med Internet Res 2025;27:e60667
URL: https://www.jmir.org/2025/1/e60667
doi:10.2196/60667
PMID:

©Nina Thigpen, Shyamal Patel, Xi Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Remote Monitoring of Chemotherapy-Induced Peripheral Neuropathy by the NeuroDetect iOS App: Observational Cohort Study of Patients With Cancer

Ciao-Sin Chen[1], MS, PharmD, PhD; Michael P Dorsch[1], MS, PharmD; Sarah Alsomairy[1], PharmD; Jennifer J Griggs[2,3], MPH, MD; Reshma Jagsi[2,3,4], MD, DPhil; Michael Sabel[2,3], MD; Amro Stino[5], MD; Brian Callaghan[5], MD; Daniel L Hertz[1,3], PharmD, PhD

[1]Department of Clinical Pharmacy, University of Michigan College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

[2]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, United States

[3]Michigan Medicine, University of Michigan Rogel Cancer Center, Ann Arbor, MI, United States

[4]Department of Radiation Oncology, Emory University School of Medicine, Atlanta, GA, United States

[5]Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, United States

Corresponding Author:
Daniel L Hertz, PharmD, PhD
Department of Clinical Pharmacy
University of Michigan College of Pharmacy
University of Michigan
428 Church St.
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 0015
Email: dlhertz@med.umich.edu

Abstract

Background: Chemotherapy-induced peripheral neuropathy (CIPN) is a common and debilitating adverse effect of neurotoxic chemotherapy characterized by symptoms such as numbness, tingling, and weakness. Effective monitoring and detection of CIPN are crucial for avoiding progression to irreversible symptoms. Due to the inconvenience of clinic-based objective assessment, CIPN detection relies primarily on patients' reporting of subjective symptoms, and patient-reported outcomes are used to facilitate CIPN detection. Our previous study found evidence that objective functional assessments completed within a smartphone app may differentiate patients with and those without CIPN after treatment.

Objective: This prospective, longitudinal observational cohort study aimed to determine the feasibility and accuracy of app-based remote monitoring of CIPN in patients with cancer undergoing neurotoxic chemotherapeutic treatment and to conduct exploratory comparisons of app-based functional CIPN monitoring versus patient-reported outcome–only monitoring.

Methods: The NeuroDetect app (Medable Inc) includes subjective EORTC (European Organization for Research and Treatment of Cancer) Quality of Life Questionnaire (QLQ)–20-item scale (CIPN20) and 6 objective functional assessments that use smartphone sensors to mimic neurological examinations, such as walking, standing, and manual dexterity tests. The functional assessment data were collected from patients with cancer undergoing neurotoxic chemotherapy, and a neurological examination was conducted at the end of treatment to diagnose CIPN in the feet (CIPN-f) or CIPN in the hands (CIPN-h). Various classification models including NeuroDetect features only (NeuroDetect Model) CIPN20-only (CIPN20 Model) or a combination of both (Combined Model) were trained and evaluated for accuracy in predicting CIPN probability.

Results: Of the 45 patients who completed functional assessments and neurological examinations, 24 had CIPN-f, and 29 had CIPN-h. The NeuroDetect Model could discriminate between patients with and those without CIPN-f (area under the curve=83.8%, comparison with no information rate P=.02) but not CIPN-h (area under the curve=67.9%, P=.18). The rolling rotation features from the eyes-closed phase of the Romberg Stance assessment showed the greatest contribution to CIPN-f (40% of total variable importance) and the Finger Tapping assessment showed the greatest contribution to CIPN-h (85% of total variable importance). The NeuroDetect Model had numerically, and at some time points statistically, superior performance to the CIPN20 Model in both CIPN-f and CIPN-h, particularly before and early in treatment. The Combined Model numerically, though not statistically,

XSL•FO
RenderX

outperformed either assessment strategy individually, indicating that the combination of functional and patient-reported assessment within a smartphone may be optimal to CIPN detection.

Conclusions: Our findings demonstrate the feasibility of integrating subjective and objective CIPN assessment into a smartphone app for remote, longitudinal CIPN monitoring. Studies of larger patient cohorts are needed to refine the app-based CIPN detection models and determine whether their use in practice improves CIPN detection.

(J Med Internet Res 2025;27:e65615) doi:10.2196/65615

KEYWORDS

chemotherapy-induced peripheral neuropathy; smartphone; mobile health; gait; balance; manual dexterity; mobile phone; chemotherapy; peripheral neuropathy; CIPN; neurotoxic chemotherapy; monitoring; detection; patient-reported outcomes; mHealth; application

Introduction

Chemotherapy-induced peripheral neuropathy (CIPN) is a common adverse effect of chemotherapy, affecting up to 70% of patients receiving taxanes for breast, lung, ovarian, and other cancer types [1]. CIPN typically presents with numbness or tingling in the feet or hands that may progress to motor weakness and impairment [2,3]. In some cases, CIPN can be irreversible, resulting in long-term gait imbalance and impaired manual dexterity, which hinder daily activity, reduce independence, increase fall risk, and reduce quality of life [4,5].

Detecting CIPN during treatment is critical to avoid these potentially irreversible adverse effects [4,5]. Detection of CIPN during neurotoxic chemotherapy treatment primarily depends on patients self-reporting symptoms during routine appointments with their medical oncology team [6]. Objective CIPN assessment may be useful to detect symptoms but is not commonly used in practice due to logistical barriers, including the need for additional trained personnel, time, space, and equipment [7]. As an alternative, validated patient-reported outcome (PRO) questionnaires have been developed to enable convenient subjective CIPN assessment. One such tool is the CIPN Quality of Life Questionnaire developed by the European Organization for Research and Treatment of Cancer (EORTC), which includes a 20-item scale (CIPN20) [8]. PROs may be an improvement over routine physician assessments via the National Cancer Institute Common Terminology Criteria for Adverse Event grading, but PROs lack objective measurement and rely on patients' willingness to accurately report CIPN symptoms [9]. Consequently, there remains an unmet need to develop strategies for objective CIPN assessment that are feasible and accurate.

Smartphone and wearable sensor-based technology have been increasingly used to advance the detection and monitoring of various diseases [10], for example, smartphone apps can monitor gait and balance changes and provide valuable insights for patients with progressive Parkinson disease [11]. Similar to Parkinson disease, objective evidence of CIPN includes impaired gait, increased sway, and reduced manual dexterity, all of which can be detected through functional assessments using wearable or smartphone sensors [12,13]. Our previous study of the first version of NeuroDetect (Medable Inc) used off-the-shelf smartphone sensor–based functional assessments and showed feasibility in differentiating patients with and those without CIPN posttreatment [13]. Integrating PRO data with objective assessments within a smartphone app could achieve remote and objective CIPN monitoring and enhance the detection of CIPN with minimal inconvenience and cost [14]. The objective of this longitudinal study was to determine the feasibility of conducting longitudinal, remote smartphone-based functional assessments that emulate certain aspects of a formal bedside neurological examination and investigate whether these functional assessments can detect CIPN during neurotoxic treatment. Exploratory analyses in this unpowered feasibility study investigated the sensitivity and specificity of an app-based approach compared with or combined with PRO-only CIPN assessment.

Methods

NeuroDetect App

NeuroDetect is an iOS (Apple Inc) app that incorporates both subjective and objective evaluations of CIPN. The subjective evaluations include EORTC Quality of Life Questionnaire (QLQ)–CIPN20, a validated PRO questionnaire for the collection of CIPN sensory, motor, and autonomic symptoms on a 4-point Likert scale (1=not at all, 2=a little, 3=quite a bit, and 4=very much) [8,15]. The objective functional assessments were customized from the available activities in the ResearchKit (Apple Inc) framework to replicate certain aspects of the neurological examination [16]. The 6 NeuroDetect functional assessments use embedded sensors within the smartphone including an accelerometer, gyroscope, and touchscreen to monitor speed, sway, accuracy, and rhythm during gait and balance tasks (natural walk, tandem walk, tandem stance, and Romberg stance) and manual dexterity tasks (finger tapping and hole-peg test). For both natural walk and tandem walk, the user is asked to place their phone in their back pocket while taking 10 steps forward, turning around, and then walking 10 steps back at their usual pace [16,17]. For both tandem stance and Romberg stance, the user is asked to stand as still as possible for 10 seconds. In tandem walk and tandem stance, the user's feet are lined up with 1 foot directly in front of the other while walking or standing, respectively [17-19]. In Romberg stance, the user stands still for 10 seconds with their eyes open and then for another 10 seconds with their eyes closed [18,19]. In finger tapping, the user taps 2 buttons on the screen of their phone with their index and middle fingers, alternatively for 100 taps or 100 seconds (whichever comes first) [16]. In the hole-peg test, which is adapted from the standard 9-hole peg test, the user repeats the process of placing a "peg" (indicated by a closed

XSL•FO
RenderX

circle) into a "hole" (indicated by an open circle) on the phone screen and then removing it 4 times [13,20].

Study Participants

Patients scheduled to begin taxane or platinum treatment for breast or colorectal cancer were identified by the Breast or Colorectal Clinical Research Teams within the University of Michigan Rogel Cancer Center. After the medical oncologist approved recruitment of the patient, the study team called the potential participant up to 3 times to describe the study and assess their eligibility and interest in participating. Patients who were interested in participating in the study were screened during the phone call to confirm that they (1) had consistent access to an iPhone (Apple Inc) and (2) did not have baseline neuropathy, defined as any score >1 in the first 4 items of the CIPN20, which ask about tingling or numbness in the fingers or toes [21]. After screening, interested participants completed verbal informed consent to participate and were provided email instructions to download NeuroDetect and complete the integrated written informed consent, a baseline CIPN20, and the 6 functional assessments.

Patients were asked to repeat the CIPN20 and functional assessments within 24 hours before each chemotherapy cycle. Depending on the chemotherapy regimen, the evaluations ranged from every 2 to 4 weeks and the timing was adjusted to accommodate treatment delays or early discontinuations. Patients were also allowed to complete additional evaluations during treatment.

Within 1 week before or 2 weeks after the end of chemotherapy treatment, patients underwent a structured clinical neurological history and examination that included assessment of neuropathic symptoms, signs, and reflexes, which was conducted by neuromuscular specialists (BC and AS) on the study team. CIPN signs and symptoms were diagnosed based on the Toronto Consensus Definition of possible clinically evident distal symmetrical polyneuropathy, which was considered the gold-standard CIPN definition for this study [22,23]. Each patient was clinically assessed for sensory symptoms (numbness, dysesthesias, hypersensitivity to touch, or pain), muscle strength, sensory function of small and large fibers, gait and coordination, and reflexes. From these data, the clinician determined whether the patient had symptoms, sensory exam findings, or decreased reflexes consistent with polyneuropathy. If at least 1 of these 3 were present, a patient is considered to have possible CIPN and thus was considered a CIPN case for this analysis. CIPN cases were then classified as experiencing CIPN in the hands (CIPN-h) or CIPN in the feet (CIPN-f) based on the location of the symptom or sensory or reflex exam finding that caused them to be considered a CIPN case.

Data Preprocessing

Sensor data from the walking and standing samples were filtered and preprocessed before feature extraction. Walking forth and back were treated as 2 samples. Walking samples with durations of less than 7.5 seconds in natural walking or 10.5 seconds in tandem walking were discarded as these indicate the patient stopped the walking task rather than completing it. Walking samples were trimmed to keep only the active walking phase without interruption, including turning around. Walking and standing samples were rotated to the global reference frame to adjust for individual variations in the orientation of the patient's phone in their pocket or the direction the patient faced while walking.

A total of 3700 features were generated from the 6 functional assessments (Table S1 in Multimedia Appendix 1). An open-source tool, mhealthtools in R (Posit), was used to extract statistical or informational features from sensor data of walking, standing, and finger-tapping assessments [24]. For walking and standing data in each of the 3 dimensions of acceleration (side-to-side, forward-backward, and up-and-down) and rotation (pitch, yaw, and roll), empirical 2-mode decomposition transformation and empirical wavelet transformation were used, respectively, to generate features in time-domain (signal amplitude and distribution) and frequency-domain (signal composition in different frequencies) [24]. For finger-tapping data, features of tapping intervals and drifts were generated to characterize both dominant and nondominant hand movements [24]. The mean and SD of the duration of placing and removing pegs were calculated for hole-peg tests in R.

Features were filtered using an ensemble method of minimum redundancy maximum relevance through mRMRe in R to account for multicollinearity [25]. The number of features to select was based on the number of features with a positive score, which means the relevance to the outcome was higher than redundancy with any previously selected features. Feature selection was determined by majority voting among multiple solution sets through 3 steps. First, baseline features were filtered within each assessment and voted among 5 solution sets. Second, end-of-treatment features were filtered together with selected baseline features within each assessment and voted among 5 solution sets to remove features correlated with the baseline. Third, all selected end-of-treatment features were further filtered and voted on among 15 solution sets to remove features correlated with each other between different assessments.

NeuroDetect Model Training and Testing

In total, 2 elastic net classification NeuroDetect Models were built to filter feet and hand features, respectively, and to achieve the highest area under the curve (AUC) in detecting CIPN-h and CIPN-f [26]. NeuroDetect Models were trained using 70% of the end-of-treatment data set with repeated cross-validation through caret and glmnet in R, and the performance was evaluated in the unbiased 30% testing data set [27,28]. The numbers of repeats and folds were the square root of the sample size. Optimal model parameters were first searched using sequences automatically generated by the glmnet package based on the range of input data, which were then fined-tuned to the same magnitude [28]. Infinite values were imputed with 1.5 times the maximum value. All input values were centered and scaled before training and testing. Variable importance was calculated using the caret package based on regression coefficients [27].

NeuroDetect Model performance was evaluated via AUC and accuracy, which was compared with the accuracy that could be

XSL•FO
RenderX

obtained by always predicting the majority class (ie, Yes CIPN-f or CIPN-h), referred to as the no information rate.

Contribution of Functional Assessments in NeuroDetect Models

The contribution of each functional assessment (eg, natural walk, Romberg stance, and finger tapping) to NeuroDetect Model classification was assessed by building and evaluating exploratory models that included only a single functional assessment and models that included all but 1 of the functional assessments using a similar process as that used to build the NeuroDetect Models. The AUC of each exploratory model was compared with the NeuroDetect Models using all functional assessments by a 1-sided bootstrap method with 400 permutations.

Comparisons With CIPN20 Models and Combined Models

To compare the performance between functional assessments and PROs to detect CIPN-f and CIPN-h, a similar process was used to build models for CIPN-f and CIPN-h that included either CIPN20 items only (CIPN20 Models) or the combination of CIPN20 items and NeuroDetect features (Combined Models). A 2-sided bootstrap method with 400 permutations was used to compare the AUCs of these models for detecting CIPN-f or CIPN-h at the end of treatment.

To evaluate the performance of each model for detecting CIPN at earlier time points during treatment, data collected from NeuroDetect Models, CIPN20 Models, and Combined Models at months 0 (M0, before day 0), 1 (M1, from day 0 to day 29), 2 (M2, from day 30 to day 59), and 3 (M3, day 60+) were used, and the AUC for detecting CIPN-f and CIPN-h was calculated and compared using the same 2-sided bootstrap method. The earliest time point a model could differentiate patients with and those without CIPN is estimated based on the first consecutive day that there was a statistically significant association between the model classification and CIPN diagnosis using a Fisher exact test. Multiple comparisons for testing each day throughout treatment were accounted for using the Benjamini and Hochberg false discovery rate.

Ethical Considerations

This study was approved by the University of Michigan IRBMed (HUM00171478). All participants completed oral informed consent before downloading NeuroDetect and then completed comprehensive written informed consent within NeuroDetect. All data were collected within secure databases accessible only to the study team. Clinical data were collected within REDCap (Research Electronic Data Capture; Vanderbilt University) and NeuroDetect data were collected within the secure data environment provided by Medable. Patients were compensated US $50 if they completed at least 3 NeuroDetect assessments and the clinical neurological assessment, otherwise, they did not receive compensation and were replaced in the study. No identifying information for individual participants is included in this manuscript.

Results

Patient Enrollment and CIPN Data Collection

Of the 85 eligible patients who indicated interest in participating, a total of 66 patients enrolled and completed CIPN20 and functional assessments at baseline and at least once during treatment, and 45 completed a neurological examination at the end of treatment (Multimedia Appendix 2). Among these 45 patients, 24 had CIPN-f and 29 had CIPN-h (Table 1). On average, each patient performed each gait and balance assessment 7.5 times, each manual dexterity assessment 7.4 times, and CIPN20 7.4 times during treatment. Most patients (40/45, 80%) completed NeuroDetect assessments and CIPN20 questionnaires every 3 weeks. After data cleaning, the walking and standing data had 567 unique samples throughout treatment, including 70 samples at the end of treatment, and the finger tapping and hole-peg test data had 318 samples throughout treatment, including 41 samples at the end of treatment. A total of 333 CIPN20 questionnaires were collected including 45 at the end of treatment.

XSL•FO
RenderX

Table 1. Clinical characteristics of the 45 patients included in the analyses.

Characteristic	Statistical value
Age (year), mean (SD)	48 (15)
Sex, n (%)	
Female	40 (89)
Male	5 (11)
Self-reported race, n (%)	
White	38 (84)
Black	4 (9)
Asian	2 (4)
Other	1 (2)
Tumor type, n (%)	
Breast	40 (89)
Colorectal	5 (11)
Neurotoxic chemotherapy treatment, n (%)	
Taxane	26 (58)
Platinum	13 (29)
Taxane and platinum combination	8 (18)
Neurotoxic chemotherapy treatment duration (week), mean (SD)	11.5 (4.6)
Treatment adjustment, n (%)	
Dose delay	9 (20)
Dose decrease	13 (29)
Switch to nab-paclitaxel	2 (4)
CIPN[a] from neuromuscular specialist assessment, n (%)	
CIPN in the feet (CIPN-f)	24 (53)
CIPN in the hands (CIPN-h)	29 (64)
NeuroDetect gait and balance assessments per patient, mean (SD)	6.9 7.5 (9.2)
NeuroDetect manual dexterity assessments per patient, mean (SD)	7.4 (8.9)
CIPN20[b] assessments per patient, mean (SD)	7.4 (8.1)

[a]CIPN: chemotherapy-induced peripheral neuropathy.

[b]CIPN20: 20-item scale.

NeuroDetect Models for Detecting CIPN at the End of Treatment

From the 3700 features generated from the 6 functional assessments (Table S1 in Multimedia Appendix 1), minimum redundancy maximum relevance selected 300 gait and balance features and 20 manual dexterity features, of which elastic net selected 182 gait and balance features and 19 manual dexterity features to be included in the final NeuroDetect Models of CIPN-f and CIPN-h, respectively (Tables S2 and S3 in Multimedia Appendix 1). The NeuroDetect Model accuracy was significantly higher than no information rate for CIPN-f (80% vs 53%, *P*=.02, Table 2, Figure 1) but not CIPN-h (82% vs 64%, *P*=.18).

XSL•FO
RenderX

Table 2. End-of-treatment chemotherapy-induced peripheral neuropathy detection performance by NeuroDetect, CIPN20, and Combined Models.

Performance metric	CIPN-f[a] Models			CIPN-h[b] Models		
	NeuroDetect	CIPN20[c]	Combined	NeuroDetect	CIPN20	Combined
Area under curve	0.84	0.76	0.94	0.68	0.56	0.86
Accuracy	0.80	0.69	0.85	0.82	0.58	0.73
Accuracy versus NIR[d], *P* value[e]	.02[f]	.20	.005[f]	.18	.82	.39

[a]CIPN-f: chemotherapy-induced peripheral neuropathy in the feet.

[b]CIPN-h: chemotherapy-induced peripheral neuropathy in the hands.

[c]CIPN20: 20-item scale.

[d]NIR: no information rate.

[e]*P* value testing accuracy of each Model versus the no information rate, which is the accuracy obtained by always predicting the majority class.

[f]Significance level *P*<.05.

Figure 1. Predicted chemotherapy-induced peripheral neuropathy probability of NeuroDetect models at the end of treatment. CIPN: chemotherapy-induced peripheral neuropathy; CIPNf: chemotherapy-induced peripheral neuropathy in the feet; CIPNh: chemotherapy-induced peripheral neuropathy in the hands.

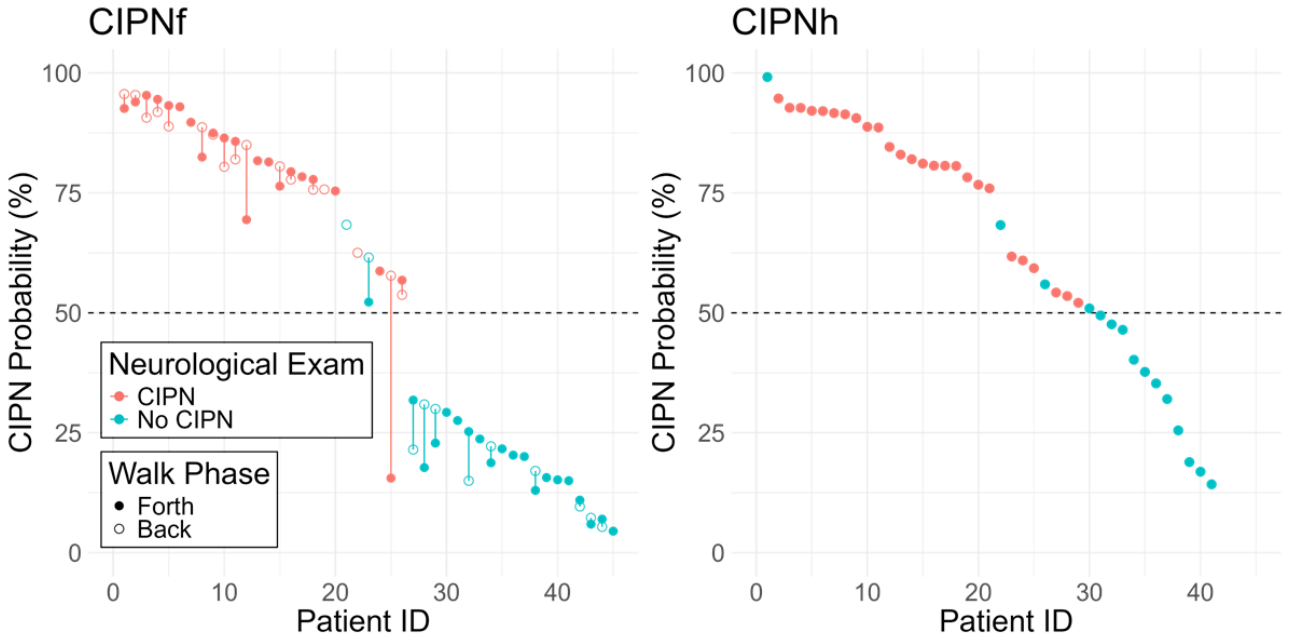

All data were collected within a prospective observational study of 45 patients with cancer receiving neurotoxic chemotherapy treatment who completed NeuroDetect functional assessments during treatment and a clinical neurological assessment at the end of treatment. The NeuroDetect models were used to estimate the likelihood (y-axis) that each patient had CIPN-f (left panel) and CIPN-h (right panel) at the end of neurotoxic chemotherapy treatment. Patients are reordered in descending order of the strength of the CIPN prediction probability from NeuroDetect. There was strong concordance between these estimated probabilities and the clinical diagnosis from the clinical neurological examination; red circles indicate patients who were diagnosed with CIPN, and blue circles indicate patients who were not. Closed and open circles indicate the 2 walking phases, forth and back, respectively, which were connected by a line for each patient. There was strong agreement between CIPN-f probability estimated between the 2 walking phases in all except a single patient.

The final NeuroDetect Models are enriched for features from Romberg stance (40.2% of total variable importance in CIPN-f) and finger tapping (84.7% of total variable importance in CIPN-h, Table S4 in Multimedia Appendix 1) assessments. The 3 most important CIPN-f features are all rolling rotations from tandem walk and Romberg stance assessments (Table S2 in Multimedia Appendix 1). The most important CIPN-h feature is the autocorrelation of the nondominant hand-tapping interval during the finger-tapping assessment (Table S3 in Multimedia Appendix 1). In CIPN-f detection, the model built using only features from Romberg stance had the best performance and had higher AUC than the model built using all functional assessments (AUC=98% vs 83.8%, *P*=.04, Table S5 in Multimedia Appendix 1). Similarly, the detection performance of the CIPN-f model declined most when removing Romberg Stance features (Table S6 in Multimedia Appendix 1), indicating that Romberg stance had the highest contribution to CIPN-f detection. Neither finger tapping nor hole-peg test had a

XSL•FO
RenderX

significant independent contribution to CIPN-h classification (Tables S5 and S6 in Multimedia Appendix 1).

Comparisons Between NeuroDetect, CIPN20, and Combined Models

In exploratory analyses comparing end-of-treatment CIPN detection between models, NeuroDetect Models had numerically, but not statistically, superior performance to CIPN20 Models in both CIPN-f (AUC=84% vs 76%, *P*=.55, Table 2 and Table S7 in Multimedia Appendix 1) and CIPN-h (AUC=68% vs 56%, *P*=.67). Combined Models numerically, but not statistically, outperformed either independently (Figure 2, Table S7 in Multimedia Appendix 1).

Figure 2. Receiver operating characteristic curves for chemotherapy-induced peripheral neuropathy detection at the end of treatment. CIPN20: 20-item scale; CIPNf: chemotherapy-induced peripheral neuropathy in the feet; CIPNh: chemotherapy-induced peripheral neuropathy in the hands.

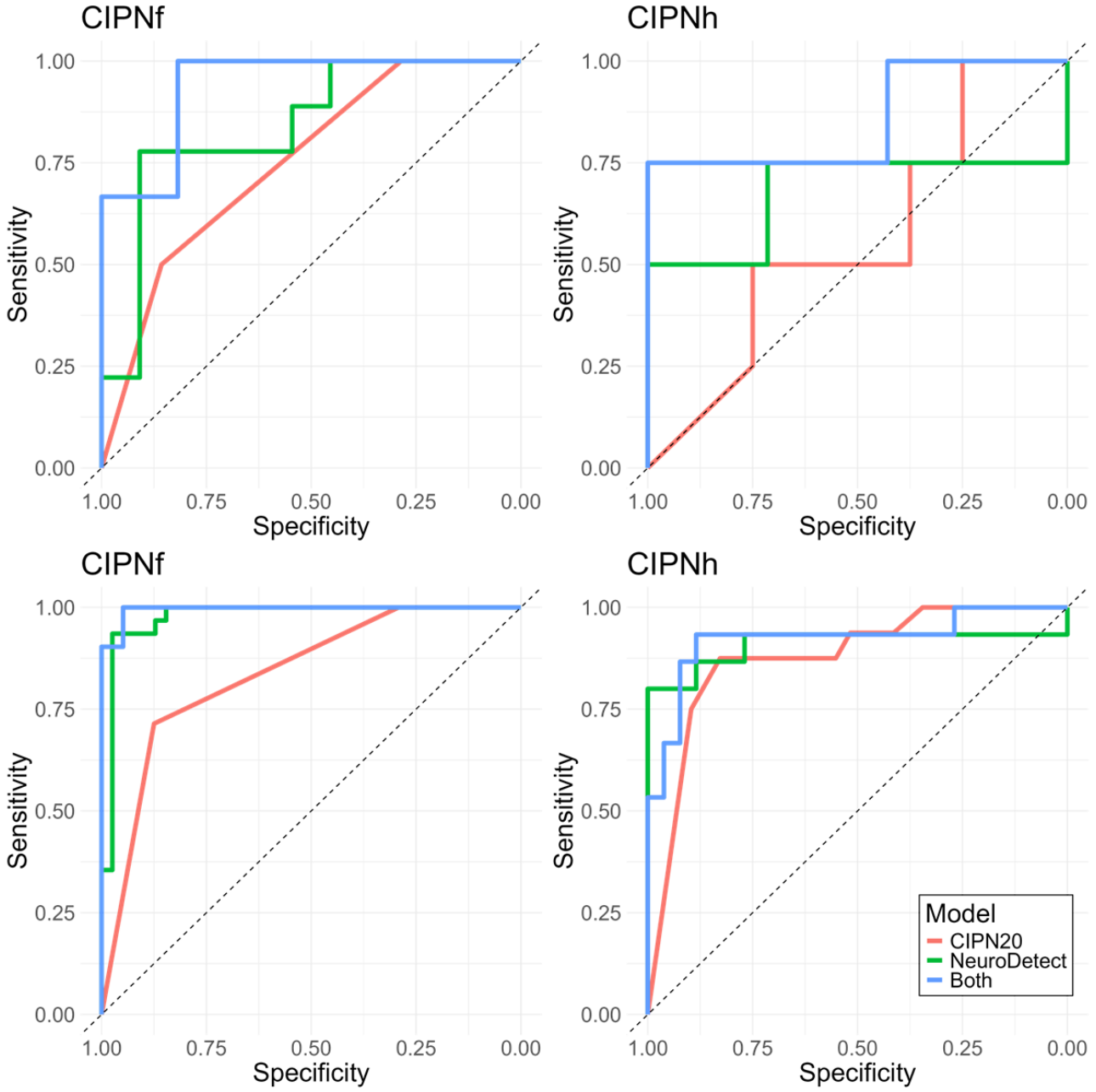

All data were collected within a prospective observational study of 45 patients with cancer receiving neurotoxic chemotherapy treatment who completed NeuroDetect functional assessments during treatment and a clinical neurological assessment at the end of treatment. Models built using only the features generated from the functional assessments conducted within NeuroDetect (NeuroDetect Models), only from the patient-reported CIPN20 questionnaire items (CIPN20 Models), or the combination of both (Combined Models) were compared for their accuracy in detecting CIPN-f and CIPN-h at the end of treatment. CIPN-f and CIPN-h are based on a clinical neurological assessment. The top 2 panels present the sensitivity and specificity of the CIPN-f (left) and CIPN-h (right) models in only the unbiased independent testing data set (30% of the entire data set, not used in model training). The bottom 2 panels present similar results for the entire data set, including both the training and testing data. Colored lines indicate results from the NeuroDetect Models (green), CIPN20 Models (red), and Combined Models (blue).

XSL•FO
RenderX

NeuroDetect Models had a good AUC with balanced sensitivity and specificity. In CIPN-f classification, the NeuroDetect and Combined Models had a greater AUC than the CIPN20 Model.

Models were then compared on their ability to detect CIPN at earlier time points using the patient data at months 0, 1, 2, and 3 of treatment. NeuroDetect and Combined models were numerically, and in most cases statistically, superior to CIPN20 models at M0 (pretreatment) and at M1, but the models were similarly accurate by month 2 (Figure 3 and Table S7 in Multimedia Appendix 1).

Figure 3. Area under the curve of models over the course of treatment. CIPN20: 20-item scale; CIPNf: chemotherapy-induced peripheral neuropathy in the feet; CIPNh: chemotherapy-induced peripheral neuropathy in the hands.

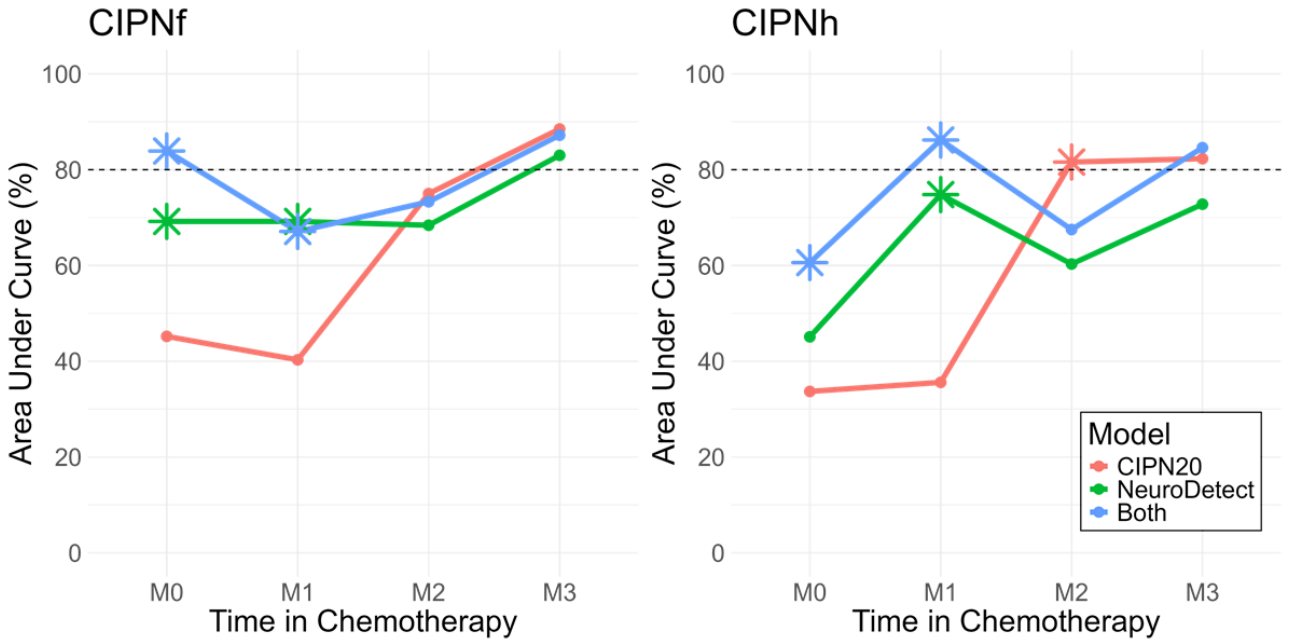

All data were collected within a prospective observational study of 45 patients with cancer receiving neurotoxic chemotherapy treatment who completed NeuroDetect functional assessments during treatment and a clinical neurological assessment at the end of treatment. Data before treatment (M0) and during months 1 (from day 0 to day 29), 2 (from day 30 to day 59), and 3 (from day 60 to day 89) of treatment were used to evaluate the performance of the CIPN detection models. Models were generated using only the NeuroDetect objective assessments (green), only the patient-reported CIPN20 questionnaires (red), and a Combined Model using both (blue). The data from each model collected before and throughout treatment were used to estimate model performance through AUC. The true CIPN-f and CIPN-h were based on the clinical neurological assessment conducted at the end of treatment. Asterisk dots indicate significantly higher AUC than at least 1 of the other 2 models by a bootstrap method. NeuroDetect and Combined Models were superior to CIPN20 Models before treatment (M0) and at M1, but models generally had nonsignificantly different performance after M2. These analyses suggest that CIPN detection before and early in treatment can be improved by using NeuroDetect functional assessments in addition to patient-reported questionnaires.

Finally, analyses were conducted to determine at what point in treatment the NeuroDetect and other Models could discriminate between patients with and those without CIPN. For CIPN-f, the NeuroDetect Model could statistically discriminate between patients with and those without CIPN on day 62 (Figure 4), which was numerically though not statistically later than the CIPN20 Model and similar to the Combined Model (indicated in Figure 4, raw data not shown). For CIPN-h, the NeuroDetect Model reached significance at day 65 (Figure 4), which was not statistically different from that of the CIPN20 or Combined Models.

XSL•FO
RenderX

Figure 4. Percentage of patients detected with chemotherapy-induced peripheral neuropathy over time. CIPN: chemotherapy-induced peripheral neuropathy; CIPNf: chemotherapy-induced peripheral neuropathy in the feet; CIPNh: chemotherapy-induced peripheral neuropathy in the hands.

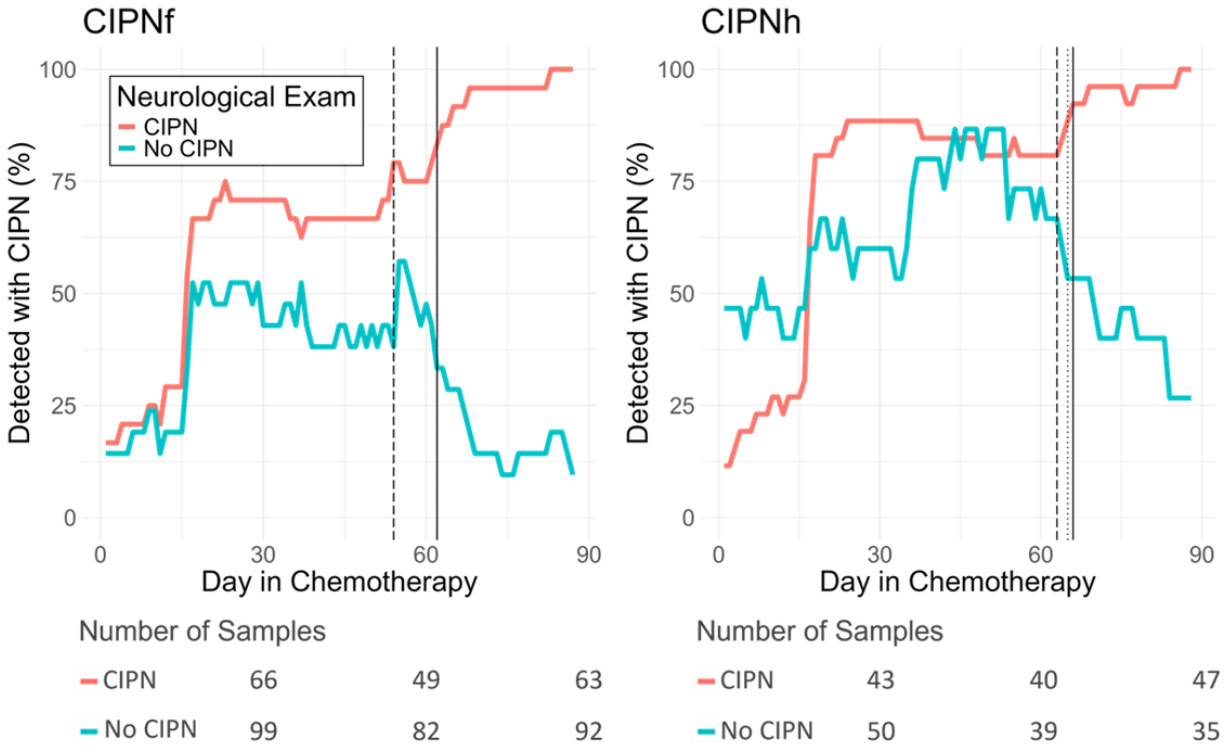

All data were collected within a prospective observational study of 45 patients with cancer receiving neurotoxic chemotherapy treatment who completed NeuroDetect functional assessments during treatment and a clinical neurological assessment at the end of treatment. Data collected during treatment were used to estimate the earliest time point the Models created from only the NeuroDetect functional assessments could statistically differentiate patients with and those without CIPN. Similar analyses were conducted for the models generated from only the CIPN20 patient-reported questionnaires and the combination of both data types (Combined Model) for comparison (raw data not shown). Red lines indicate the group of patients diagnosed with CIPN from neurological examinations conducted at the end of treatment and blue lines indicate the group not diagnosed with CIPN. For CIPN-f, the NeuroDetect Model reached statistical significance at day 62 (P=.003, solid vertical line), which is the same day that CIPN discrimination was significant for the Combined model (dotted vertical line not visible due to overlay with solid line). The CIPN20 Model discrimination was significant at day 54 (dashed vertical line), though this difference was not statistically significant. For CIPN-h, the NeuroDetect Model reached statistical significance at day 66 (solid vertical line), which was not statistically different from the day the CIPN20 (day 63, dashed vertical line) and Combined (day 65, dotted vertical line) Models reached statistical significance.

Discussion

Principal Results

Tremendous advancements in the development of smartphone and wearable sensor–based technology in health care enable various apps relating to disease and treatment [10], including monitoring gait, balance, and manual dexterity changes in Parkinson disease [11]. Like Parkinson disease, CIPN is associated with changes in gait and balance, and reductions in manual strength and dexterity. Detecting CIPN is essential to prevent its progression to debilitating symptoms that can irreversibly diminish function and quality of life [4,5]. The current approach to CIPN detection, which relies on patient's reporting symptoms during appointments, has limitations that are not addressed by currently available clinic-based objective assessment or PRO subjective assessment strategies. Our previous NeuroDetect V1.0 cross-sectional study revealed evidence that a single timepoint remote, smartphone sensor–based functional assessments after neurotoxic chemotherapy treatment can differentiate patients with and those without CIPN [13]. Building upon our previous study, the objective of this NeuroDetect V2.0 longitudinal study was to determine the feasibility of completing longitudinal app-based functional assessments that emulate certain aspects of the neurological examination that can detect CIPN during neurotoxic treatment. In addition, we conducted exploratory comparisons of functional assessments collected via NeuroDetect V2.0 with CIPN20 questionnaires. The results demonstrated the feasibility of integrating subjective and objective CIPN assessment into a smartphone app for remote, longitudinal CIPN monitoring.

Comparison With Previous Work

In this longitudinal observational cohort study, the NeuroDetect Model effectively distinguished patients with and those without end-of-treatment CIPN-f based on gold-standard clinical assessment by neuromuscular specialists, while maintaining an acceptable false discovery rate. The features contributing the most in the final CIPN-f NeuroDetect Model were rolling rotation, which is intertwined with the side-to-side swaying movement features from balance assessments. These features

XSL•FO
RenderX

during the eyes closed period of the Romberg stance assessment showed the greatest contribution to CIPN-f detection. Furthermore, 2 previous studies used smartphone or smart device sensors to monitor CIPN [13,29]. One study did not detect any correlation between CIPN and a small set of gait and balance features [29]. Alternatively, in our cross-sectional study of NeuroDetct V1.0, a side-to-side swaying acceleration pattern during a normal walk assessment indicated diminished balance in patients with CIPN [13]. Side-to-side swaying features were also identified in studies with open-eye and closed-eye balancing assessments using commercial or self-designed wearable devices and software [30-32] or other clinic-based balance assessments [12,14]. Previous studies have also reported shorter step lengths in patients with CIPN [4,13,32]. This analysis could not be conducted since NeuroDetect V2.0 does not estimate step length.

The CIPN-h NeuroDetect Model was not able to distinguish patients with and those without end-of-treatment CIPN-h, which might suggest a more sensitive functional assessment is needed. Our NeuroDetect V1.0 and V2.0 studies are the only attempts to our knowledge to objectively evaluate manual dexterity using a remote app-based strategy [13]. The hole-peg test is an on-screen assessment that is supposed to mimic the 9-hole peg test, which is often used to assess manual dexterity, in which patients physically pick up pegs and stick them into a hole on a board [20]. In our cross-sectional study of NeuroDetect V1.0, features from the hole-peg test were indicative of CIPN. However, user feedback indicated that the hole-peg test was confusing to operate on a screen [13], which was also reported by patients within poststudy interviews conducted in this longitudinal study (data not included). In the longitudinal study of NeuroDetect V2.0, the newly added finger-tapping assessment was more informative than the hole-peg test, but the overall performance remained poor. Finger tapping is also used as an objective manual strength assessment in neurological disorders and hereditary peripheral neuropathy [33,34], although its use in CIPN has not been validated to our knowledge. Adding other screen-compatible manual dexterity assessments, such as line tracing, might improve the ability of smartphone sensor–based assessments to detect diminished manual dexterity indicative of CIPN-h [35]. One potential challenge will be to develop manual dexterity assessments that are specific for CIPN symptoms without the interference of other chemotherapy toxicities, such as cognitive impairment [36].

In exploratory model comparisons, combining both NeuroDetect features and CIPN20 items together in a Combined Model numerically outperformed each Model individually early in treatment. This finding indicates that integration of sensory-based functional assessments with PRO may improve CIPN detection, which is consistent with previous research and statements that support combining subjective and objective CIPN assessments to improve CIPN detection [37]. The potential benefits of a combined strategy are further supported by our exploratory analysis, which suggested that CIPN20 differentiates between patients with and those without CIPN 2-8 days earlier than NeuroDetect. Currently available hybrid assessment strategies can be time-consuming and require specialized equipment and trained professionals [37] and have not been demonstrated to improve CIPN detection compared with using only PROs [32,38]. On the other hand, NeuroDetect might provide a convenient option for objectively monitoring CIPN by a smartphone [39] that can be completed at home, where the patient is comfortable, and at a time that is convenient, with no added cost for purchasing equipment since most patients own a smartphone. Objective functional assessments can accurately detect CIPN without being influenced by a patient's tolerance for living with CIPN symptoms [40].

Although there is no effective treatment for CIPN, detecting CIPN can prompt early interventions, such as physical therapy referrals, chemotherapy dose reductions, or drug switching, to prevent severe and irreversible CIPN [41]. Interestingly, our results indicate that pretreatment functional NeuroDetect assessments, but not PRO assessments, may be able to identify patients who are likely to experience CIPN during treatment. This is consistent with other evidence that pre-existing neuropathy, such as that resulting from hereditary neuropathy conditions, diabetes, or advanced age, is a major risk factor for treatment-limiting CIPN [1,41], perhaps due to treatment-related exacerbation of subclinical symptoms. If this is validated in future studies, NeuroDetect may be a particularly convenient screening tool to identify potential participants for interventional trials of CIPN prevention or treatment, including approaches like acupuncture, cryo-compression, and exercise therapy [42], which could ameliorate the disparity in CIPN among non-White patients [43].

Limitations

This study advanced the development of app-based functional CIPN assessment by using a longitudinal study design and gold-standard clinical neurological history and examination for CIPN diagnosis. It also used contemporary machine learning modeling for complex multidimensional data and conducted comparative and integrative analyses to evaluate the benefits of using NeuroDetect alone or in combination with PRO. Despite these strengths, several limitations are worth considering. First, this was a small cohort with many features extracted. Rigorous data processing and model training were used to prevent overfitting and provide evidence of feasibility. Much larger clinical studies are needed to validate a NeuroDetect app and CIPN detection model for clinical use and to further investigate the benefits of adding functional NeuroDetect assessment to PRO CIPN monitoring. Second, in longitudinal analysis, the model classification during treatment could only be compared with the neurological diagnoses at the end of treatment, rather than a direct comparison with the CIPN at that moment. Collecting longitudinal objective CIPN outcomes, perhaps using a hybrid assessment such as the Total Neuropathy Score [37], is crucial for future studies attempting to develop real-time CIPN detection or outcome trajectory prediction. Third, remote functional assessments inherently rely on patient understanding and adherence to instructions, but the nonsupervised process and nonstandardized environments can introduce undesirable variability and missingness in data. This can be mitigated by simplified tasks, clear instructions, and regular reminders and is a trade-off we have accepted to gain convenience of remote CIPN detection without specialized equipment. Finally, given the small patient cohort, we did not make any attempt to conduct subgroup analyses or adjust for

XSL•FO
RenderX

other important clinical variables such as age or neurotoxic chemotherapy regimen (eg, taxane vs platinum). Our analysis did not detect statistically significant differences in age in patients who were CIPN-h or CIPN-f cases versus those who were controls (data not shown). However, since age likely affects patient's performance of functional assessments, age will be explored in detail for potential incorporation into the NeuroDetect models built within larger patient cohorts. It is also possible that other clinical consequences of treatment, such as general fatigue, would contribute to worsening performance on functional tasks, though these data were not collected within the study so this could not be explored.

Conclusions

Our findings demonstrate the feasibility of integrating subjective and objective CIPN assessment into a smartphone app for remote, longitudinal CIPN monitoring. Future work will build NeuroDetect V3.0 and generate validated CIPN-f and CIPN-h detection models in much larger patient cohorts with longitudinal CIPN assessment. Eventually, prospective randomized controlled trials will be needed to investigate whether app-based CIPN monitoring via NeuroDetect improves CIPN detection and treatment outcomes, such as reducing persistent CIPN, in patients with cancer.

Acknowledgments

This study was funded by a MICHR Pathway award through a grant from the National Center for Advancing Translational Science (NCATS; UL1TR002240).

Data Availability

All data and code will be made available upon reasonable request to the corresponding author. Due to the preliminary nature of these findings, these models are not intended to be used to detect CIPN or guide treatment in clinical practice. Researchers interested in contributing to the ongoing development of NeuroDetect should contact the corresponding author.

Authors' Contributions

DLH and MPD conceived the presented idea. SA, AS, and BC collected the patient and neurological examination data. C-SC developed the methods, performed the analyses, and drafted the first manuscript. DLH supervised the work. C-SC and DLH edited the first manuscript. All authors discussed the results and contributed to the final manuscript.

C-SC contributed to data analysis and manuscript writing and review. MPD, JJG, RJ, MS, and BC contributed to project conceptualization and manuscript review. SA and AS contributed to data collection and manuscript review. BC contributed to data collection. DLH is the principal investigator and contributed to project conceptualization, data collection, and manuscript writing and final review.

Conflicts of Interest

BC receives research and editorial support from the American Academy of Neurology, consults for DynaMed, and performs medical legal work including the vaccine injury compensation program. Unrelated to the current work, RJ has stock options as compensation for her advisory board role in Equity Quotient, a company that evaluates culture in health care companies; she has received personal fees from the Greenwall Foundation, Doris Duke Charitable Foundation, the National Institutes of Health, the Blue Cross Blue Shield Association, Physicians Education Resource, and the American Medical Association; and grants or contracts from the National Institutes of Health, the Doris Duke Charitable Foundation, the American Cancer Society, and the Susan G. Komen Foundation. She has served as an expert witness for Kleinbard, LLC and Hawks Quindel Law.

Multimedia Appendix 1
Additional Tables.
[DOCX File , 142 KB - jmir_v27i1e65615_app1.docx]

Multimedia Appendix 2
Number of patients enrolled and analyzed in the analyses.
[PNG File , 135 KB - jmir_v27i1e65615_app2.png]

References

1. Seretny M, Currie GL, Sena ES, Ramnarine S, Grant R, MacLeod MR, et al. Incidence, prevalence, and predictors of chemotherapy-induced peripheral neuropathy: a systematic review and meta-analysis. Pain 2014;155(12):2461-2470 [FREE Full text] [doi: 10.1016/j.pain.2014.09.020] [Medline: 25261162]
2. Miltenburg NC, Boogerd W. Chemotherapy-induced neuropathy: a comprehensive survey. Cancer Treat Rev 2014;40(7):872-882. [doi: 10.1016/j.ctrv.2014.04.004] [Medline: 24830939]

XSL•FO
RenderX

3. Quasthoff S, Hartung HP. Chemotherapy-induced peripheral neuropathy. J Neurol 2002;249(1):9-17. [doi: 10.1007/pl00007853] [Medline: 11954874]
4. Winters-Stone KM, Horak F, Jacobs PG, Trubowitz P, Dieckmann NF, Stoyles S, et al. Falls, functioning, and disability among women with persistent symptoms of chemotherapy-induced peripheral neuropathy. J Clin Oncol 2017;35(23):2604-2612 [FREE Full text] [doi: 10.1200/JCO.2016.71.3552] [Medline: 28586243]
5. Bandos H, Melnikow J, Rivera DR, Swain SM, Sturtz K, Fehrenbacher L, et al. Long-term peripheral neuropathy in breast cancer patients treated with adjuvant chemotherapy: NRG oncology/NSABP B-30. J Natl Cancer Inst 2018;110(2):djx162 [FREE Full text] [doi: 10.1093/jnci/djx162] [Medline: 28954297]
6. Knoerl R, Smith EML, Han A, Doe A, Scott K, Berry DL. Characterizing patient-clinician chemotherapy-induced peripheral neuropathy assessment and management communication approaches. Patient Educ Couns 2019;102(9):1636-1643 [FREE Full text] [doi: 10.1016/j.pec.2019.04.012] [Medline: 31003878]
7. Park SB, Goldstein D, Krishnan AV, Lin CS, Friedlander ML, Cassidy J, et al. Chemotherapy-induced peripheral neurotoxicity: a critical analysis. CA Cancer J Clin 2013;63(6):419-437 [FREE Full text] [doi: 10.3322/caac.21204] [Medline: 24590861]
8. Cavaletti G, Cornblath DR, Merkies ISJ, Postma TJ, Rossi E, Frigeni B, CI-PeriNomS Group, et al. The chemotherapy-induced peripheral neuropathy outcome measures standardization study: from consensus to the first validity and reliability findings. Ann Oncol 2013;24(2):454-462 [FREE Full text] [doi: 10.1093/annonc/mds329] [Medline: 22910842]
9. Salgado TM, Quinn CS, Krumbach EK, Wenceslao I, Gonzalez M, Reed HL, et al. Reporting of paclitaxel-induced peripheral neuropathy symptoms to clinicians among women with breast cancer: a qualitative study. Support Care Cancer 2020;28(9):4163-4172 [FREE Full text] [doi: 10.1007/s00520-019-05254-6] [Medline: 31897779]
10. Guo CC, Chiesa PA, de Moor C, Fazeli MS, Schofield T, Hofer K, et al. Digital devices for assessing motor functions in mobility-impaired and healthy populations: systematic literature review. J Med Internet Res 2022;24(11):e37683 [FREE Full text] [doi: 10.2196/37683] [Medline: 36409538]
11. Tripathi S, Malhotra A, Qazi M, Chou J, Wang F, Barkan S, et al. Clinical review of smartphone applications in parkinson's disease. Neurologist 2022;27(4):183-193. [doi: 10.1097/NRL.0000000000000413] [Medline: 35051970]
12. Mantovani E, Demrozi F, Hertz DL, Turetta C, Ferro O, Argyriou AA, et al. Wearables, sensors, and smart devices for the detection and monitoring of chemotherapy-induced peripheral neurotoxicity: systematic review and directions for future research. J Peripher Nerv Syst 2022;27(4):238-258. [doi: 10.1111/jns.12518] [Medline: 36224713]
13. Chen CS, Kim J, Garg N, Guntupalli H, Jagsi R, Griggs JJ, et al. Chemotherapy-induced peripheral neuropathy detection via a smartphone app: cross-sectional pilot study. JMIR Mhealth Uhealth 2021;9(7):e27502 [FREE Full text] [doi: 10.2196/27502] [Medline: 36260403]
14. Campbell G, Skubic MA. Balance and gait impairment: sensor-based assessment for patients with peripheral neuropathy. Clin J Oncol Nurs 2018;22(3):316-325. [doi: 10.1188/18.CJON.316-325] [Medline: 29781455]
15. Knoerl R, Gray E, Stricker C, Mitchell SA, Kippe K, Smith G, et al. Electronic versus paper-pencil methods for assessing chemotherapy-induced peripheral neuropathy. Support Care Cancer 2017;25(11):3437-3446. [doi: 10.1007/s00520-017-3764-y] [Medline: 28577231]
16. Developer A. ResearchKit. Apple Developer Documentation. 2023. URL: https://developer.apple.com/design/human-interface-guidelines/researchkit [accessed 2025-01-08]
17. Findling O, van der Logt R, Nedeltchev K, Achtnichts L, Allum JHJ. A comparison of balance control during stance and gait in patients with inflammatory and non-inflammatory polyneuropathy. PLoS One 2018;13(2):e0191957 [FREE Full text] [doi: 10.1371/journal.pone.0191957] [Medline: 29474369]
18. Richardson JK. The clinical identification of peripheral neuropathy among older persons. Arch Phys Med Rehabil 2002;83(11):1553-1558. [doi: 10.1053/apmr.2002.35656] [Medline: 12422324]
19. Missaoui B, Thoumie P. Balance training in ataxic neuropathies. Effects on balance and gait parameters. Gait Posture 2013;38(3):471-476. [doi: 10.1016/j.gaitpost.2013.01.017] [Medline: 23465318]
20. Yancosek KE, Howell D. A narrative review of dexterity assessments. J Hand Ther 2009;22(3):258-69; quiz 270. [doi: 10.1016/j.jht.2008.11.004] [Medline: 19217254]
21. Postma TJ, Aaronson NK, Heimans JJ, Muller MJ, Hildebrand JG, Delattre JY, EORTC Quality of Life Group. The development of an EORTC quality of life questionnaire to assess chemotherapy-induced peripheral neuropathy: the QLQ-CIPN20. Eur J Cancer 2005;41(8):1135-1139. [doi: 10.1016/j.ejca.2005.02.012] [Medline: 15911236]
22. Dyck PJ, Albers JW, Andersen H, Arezzo JC, Biessels G, Bril V, Toronto Expert Panel on Diabetic Neuropathy. Diabetic polyneuropathies: update on research definition, diagnostic criteria and estimation of severity. Diabetes Metab Res Rev 2011;27(7):620-628 [FREE Full text] [doi: 10.1002/dmrr.1226] [Medline: 21695763]
23. Tesfaye S, Vileikyte L, Rayman G, Sindrup SH, Perkins BA, Baconja M, Toronto Expert Panel on Diabetic Neuropathy. Painful diabetic peripheral neuropathy: consensus recommendations on diagnosis, assessment and management. Diabetes Metab Res Rev 2011;27(7):629-638. [doi: 10.1002/dmrr.1225] [Medline: 21695762]
24. Snyder P, Tummalacherla M, Perumal T, Omberg L. mhealthtools: a modular R package for extracting features from mobile and wearable sensor data. JOSS 2020;5(47):2106. [doi: 10.21105/joss.02106]

XSL•FO
RenderX

25. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics 2013;29(18):2365-2368. [doi: 10.1093/bioinformatics/btt383] [Medline: 23825369]
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
27. Kuhn M. Building predictive models in R using the caret package. J. Stat. Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
28. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. J Stat Softw 2023;106 [FREE Full text] [doi: 10.18637/jss.v106.i01] [Medline: 37138589]
29. Jacobs PG, Hanaway P, Leitschuh J, Condon J, Rajhbeharrysingh U, Mosquera-Lopez C, et al. Design and evaluation of a portable smart-phone based peripheral neuropathy test platform. Annu Int Conf IEEE Eng Med Biol Soc 2018;2018:1-4. [doi: 10.1109/EMBC.2018.8513100] [Medline: 30440300]
30. Fino PC, Horak FB, El-Gohary M, Guidarelli C, Medysky ME, Nagle SJ, et al. Postural sway, falls, and self-reported neuropathy in aging female cancer survivors. Gait Posture 2019;69:136-142 [FREE Full text] [doi: 10.1016/j.gaitpost.2019.01.025] [Medline: 30716669]
31. Zahiri M, Chen KM, Zhou H, Nguyen H, Workeneh BT, Yellapragada SV, et al. Using wearables to screen motor performance deterioration because of cancer and chemotherapy-induced peripheral neuropathy (CIPN) in adults - toward an early diagnosis of CIPN. J Geriatr Oncol 2019;10(6):960-967 [FREE Full text] [doi: 10.1016/j.jgo.2019.01.010] [Medline: 30665876]
32. Monfort SM, Pan X, Patrick R, Ramaswamy B, Wesolowski R, Naughton MJ, et al. Gait, balance, and patient-reported outcomes during taxane-based chemotherapy in early-stage breast cancer patients. Breast Cancer Res Treat 2017;164(1):69-77 [FREE Full text] [doi: 10.1007/s10549-017-4230-8] [Medline: 28374323]
33. Gopal A, Hsu WY, Allen DD, Bove R. Remote assessments of hand function in neurological disorders: systematic review. JMIR Rehabil Assist Technol 2022;9(1):e33157 [FREE Full text] [doi: 10.2196/33157] [Medline: 35262502]
34. Prada V, Robbiano G, Mennella G, Hamedani M, Bellone E, Grandis M, et al. Validation of a new hand function outcome measure in individuals with charcot-marie-tooth disease. J Peripher Nerv Syst 2020;25(4):413-422. [doi: 10.1111/jns.12421] [Medline: 33140522]
35. Rabah A, Le Boterff Q, Carment L, Bendjemaa N, Térémetz M, Dupin L, et al. A novel tablet-based application for assessment of manual dexterity and its components: a reliability and validity study in healthy subjects. J Neuroeng Rehabil 2022;19(1):35 [FREE Full text] [doi: 10.1186/s12984-022-01011-9] [Medline: 35331273]
36. Milutinovic B, Singh AK. Editorial: cognitive impairment and peripheral neuropathy from chemotherapy: molecular mechanisms and therapeutic approaches. Front Mol Biosci 2022;9:962889 [FREE Full text] [doi: 10.3389/fmolb.2022.962889] [Medline: 35911961]
37. Park SB, Alberti P, Kolb NA, Gewandter JS, Schenone A, Argyriou AA. Overview and critical revision of clinical assessment tools in chemotherapy-induced peripheral neurotoxicity. J Peripher Nerv Syst 2019;24 Suppl 2:S13-S25. [doi: 10.1111/jns.12333] [Medline: 31647154]
38. McCrary JM, Goldstein D, Trinh T, Timmins HC, Li T, Friedlander M, et al. Optimizing clinical screening for chemotherapy-induced peripheral neuropathy. J Pain Symptom Manage 2019;58(6):1023-1032 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.07.021] [Medline: 31374367]
39. Fischer F, Kleen S. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: scoping review. J Med Internet Res 2021;23(1):e17691 [FREE Full text] [doi: 10.2196/17691] [Medline: 33480850]
40. Hertz DL. Concerns regarding use of patient-reported outcomes in biomarker studies of chemotherapy-induced peripheral neuropathy. Pharmacogenomics J 2019;19(5):411-416 [FREE Full text] [doi: 10.1038/s41397-019-0093-1] [Medline: 31395959]
41. Hertz DL, Childs DS, Park SB, Faithfull S, Ke Y, Ali NT, et al. Patient-centric decision framework for treatment alterations in patients with chemotherapy-induced peripheral neuropathy (CIPN). Cancer Treat Rev 2021;99:102241. [doi: 10.1016/j.ctrv.2021.102241] [Medline: 34174668]
42. Loprinzi CL, Lacchetti C, Bleeker J, Cavaletti G, Chauhan C, Hertz DL, et al. Prevention and management of chemotherapy-induced peripheral neuropathy in survivors of adult cancers: ASCO guideline update. J Clin Oncol 2020;38(28):3325-3348. [doi: 10.1200/JCO.20.01399] [Medline: 32663120]
43. Schneider BP, Zhao F, Ballinger TJ, Garcia SF, Shen F, Virani S, et al. ECOG-ACRIN EAZ171: prospective validation trial of germline predictors of taxane-induced peripheral neuropathy in black women with early-stage breast cancer. J Clin Oncol 2024;42(24):2899-2907. [doi: 10.1200/JCO.24.00526] [Medline: 38828938]

Abbreviations

AUC: area under the curve
CIPN: chemotherapy-induced peripheral neuropathy
CIPN20: 20-item scale
CIPN-f: chemotherapy-induced peripheral neuropathy in the feet

XSL•FO
RenderX

CIPN-h: chemotherapy-induced peripheral neuropathy in the hand
EORTC: European Organization for Research and Treatment of Cancer
PRO: patient-reported outcome
QLQ: Quality of Life Questionnaire
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 22.08.24; peer-reviewed by T Li, SB Park; comments to author 18.09.24; revised version received 08.10.24; accepted 24.12.24; published 05.02.25.

Please cite as:
Chen CS, Dorsch MP, Alsomairy S, Griggs JJ, Jagsi R, Sabel M, Stino A, Callaghan B, Hertz DL
Remote Monitoring of Chemotherapy-Induced Peripheral Neuropathy by the NeuroDetect iOS App: Observational Cohort Study of Patients With Cancer
J Med Internet Res 2025;27:e65615
URL: https://www.jmir.org/2025/1/e65615
doi:10.2196/65615
PMID:

©Ciao-Sin Chen, Michael P Dorsch, Sarah Alsomairy, Jennifer J Griggs, Reshma Jagsi, Michael Sabel, Amro Stino, Brian Callaghan, Daniel L Hertz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

App-Based Ecological Momentary Assessment of Problematic Smartphone Use During Examination Weeks in University Students: 6-Week Observational Study

Ji Seon Ahn[1,2,3], PhD; InJi Jeong[4], BS; Sehwan Park[5], MA; Jooho Lee[5], MA; Minjeong Jeon[1,2], MA; Sangil Lee[6], PhD; Gangho Do[7], BS; Dooyoung Jung[4,6*], MD, PhD; Jin Young Park[1,2,3*], MD, PhD

[1]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

[3]Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, Republic of Korea

[4]Graduate School of Health Science and Technology, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[5]Medical Research Team, Digital Medic Co., Ltd., Seoul, Republic of Korea

[6]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[7]Digital Medic Co., Ltd., Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jin Young Park, MD, PhD
Department of Psychiatry
Yongin Severance Hospital
Yonsei University College of Medicine
363, Dongbaekjukjeon-daero
Giheung-gu
Yongin, 16995
Republic of Korea
Phone: 82 31 5189 8148
Fax: 82 31 5189 8565
Email: empathy@yuhs.ac

Abstract

Background: The increasing prevalence of problematic smartphone use (PSU) among university students is raising concerns, particularly as excessive smartphone engagement is linked to negative outcomes such as mental health issues, academic underperformance, and sleep disruption. Despite the severity of PSU, its association with behaviors such as physical activity, mobility, and sociability has received limited research attention. Ecological momentary assessment (EMA), including passive data collection through digital phenotyping indicators, offers an objective approach to explore these behavioral patterns.

Objective: This study aimed to examine associations between self-reported psychosocial measures; app-based EMA data, including daily behavioral indicators from GPS location tracking; and PSU in university students during the examination period.

Methods: A 6-week observational study involving 243 university students was conducted using app-based EMA on personal smartphones to collect data on daily behaviors and psychosocial factors related to smartphone overuse. PSU was assessed using the Korean Smartphone Addiction Proneness Scale. Data collected from the Big4+ app, including self-reports on mood, sleep, and appetite, as well as passive sensor data (GPS location, acceleration, and steps) were used to evaluate overall health. Logistic regression analysis was conducted to identify factors that significantly influenced smartphone overuse, providing insights into daily behavior and mental health patterns.

Results: In total, 23% (56/243) of the students exhibited PSU. The regression analysis revealed significant positive associations between PSU and several factors, including depression (Patient Health Questionnaire-9; odds ratio [OR] 8.48, 95% CI 1.95-36.87; *P*=.004), social interaction anxiety (Social Interaction Anxiety Scale; OR 4.40, 95% CI 1.59-12.15; *P*=.004), sleep disturbances (General Sleep Disturbance Scale; OR 3.44, 95% CI 1.15-10.30; *P*=.03), and longer sleep duration (OR 3.11, 95% CI 1.14-8.48; *P*=.03). Conversely, a significant negative association was found between PSU and time spent at home (OR 0.35, 95% CI 0.13-0.94; *P*=.04).

XSL•FO
RenderX

Conclusions: This study suggests that negative self-perceptions of mood and sleep, along with patterns of increased mobility identified through GPS data, increase the risk of PSU, particularly during periods of academic stress. Combining psychosocial assessments with EMA data offers valuable insights for managing PSU during high-stress periods, such as *examinations*, and provides new directions for future research.

(J Med Internet Res 2025;27:e69320) doi:10.2196/69320

KEYWORDS

problematic smartphone use; PSU; ecological momentary assessment; EMA; GPS tracking; digital phenotypes; psychosocial measures; university students; academic stress; mobile health; mHealth; mobile phone

Introduction

Background

Smartphones are integral to modern life [1-3], embedding themselves into daily routines owing to their advanced features [4,5] and seamless connectivity to web-based services [6,7]. Convenience and accessibility make smartphones indispensable for many, leading to habitual use and, in some cases, unchecked overreliance [5]. Therefore, excessive smartphone use has raised concerns over its potential adverse effects [6] on interpersonal relationships [6,8], physical [5,9-11] and mental health [12], and daily functioning [13,14]. Research shows that problematic smartphone use (PSU) is linked to negative outcomes, including psychopathology [15], poor academic performance [16], and sleep disturbances [17]. However, PSU is not currently classified as a distinct disorder in major diagnostic manuals, such as the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, and *International Classification of Diseases, 11th Revision*, highlighting the need to comprehensively explore its impact and influencing factors [14,18,19].

Ecological momentary assessment (EMA) is a methodological approach to capture people's moods [20,21] and behaviors in their natural environments [22-26]. Recently, EMA has expanded beyond self-report data to include smartphone sensor data [27]. This is known as passive EMA, which describes EMA systems where sensor-based data are collected without any user interaction [26-28]. GPS location–derived features are examples of passive EMA that allow researchers to explore correlations with mental health, such as the proportion of time spent at home [26,29-31]. Recent research has highlighted a significant relationship between increased time spent at home and the severity of depressive symptoms in individuals with major depressive disorder. For instance, an analysis of data from 164 participants using geolocation technology showed that those who spent more time at home during a 2-week period reported more severe depressive symptoms [32,33]. These findings suggest that time spent at home may be an important indicator of mental health, reinforcing the potential of digital phenotyping. By conducting frequent assessments in real-world settings, continuous monitoring through active and passive EMA reduces potential biases associated with traditional self-report methods, which often rely on single-point recollection and introduce inaccuracies [26,28,34,35].

The prevalence of mental health issues among university students is increasing, particularly during examination periods when stress levels peak [36,37]. Academic stress, characterized by increased pressure to perform, has been linked to various challenges, including heightened risks of depression, anxiety, and sleep disturbances [6,38-40], and is positively associated with internet addiction [41,42]. Research indicates that stress can cause excessive use of internet-based devices, potentially resulting in PSU [43]. Studies have also shown a correlation between stress and addictive behaviors related to smartphone use [36]. Moreover, PSU is highly prevalent among university students, with prevalence rates ranging from approximately 10.4% to 70%, depending on demographic and regional factors [44,45]. However, the relationship between PSU and behavioral patterns, such as physical activity, mobility, and sociability, remains underexplored. Limited research has examined how smartphone overuse affects college students' functionality during examinations. Thus, predicting mental health outcomes using smartphone use and sensor data presents a valuable research opportunity.

Objectives

This study aimed to investigate the prevalence of PSU among university students and validate previous findings on its associations with depression, anxiety, and sleep disturbance during examination periods. Furthermore, it integrated smartphone sensing data, including GPS-derived behavioral patterns, with self-reported mental health status to explore the impact of PSU on behavior and mental health as well as potential relationships with app-based EMA data.

Methods

Study Design

The study used a prospective observational design to investigate factors related to PSU among university students for 6 weeks. The data were collected during the 2 weeks preceding and following the examinations. Active EMA involved daily self-reported measures, whereas passive EMA captured real-time digital phenotypes. Key psychosocial measures were assessed at baseline and 2 weeks after the examinations (Figure 1). Additional details of the psychosocial assessment tools and EMA measures used in this study are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Study design, data collection protocol, and study timeline. AUDIT: Alcohol Use Disorders Identification Test; DASS: Depression Anxiety Stress Scale; EMA: ecological momentary assessment; GAD-7: Generalized Anxiety Disorder-7; GSDS: General Sleep Disturbance Scale; K-SAPS: the Korean Smartphone Addiction Proneness Scale; PHQ-9: Patient Health Questionnaire-9; SIAS: Social Interaction Anxiety Scale.

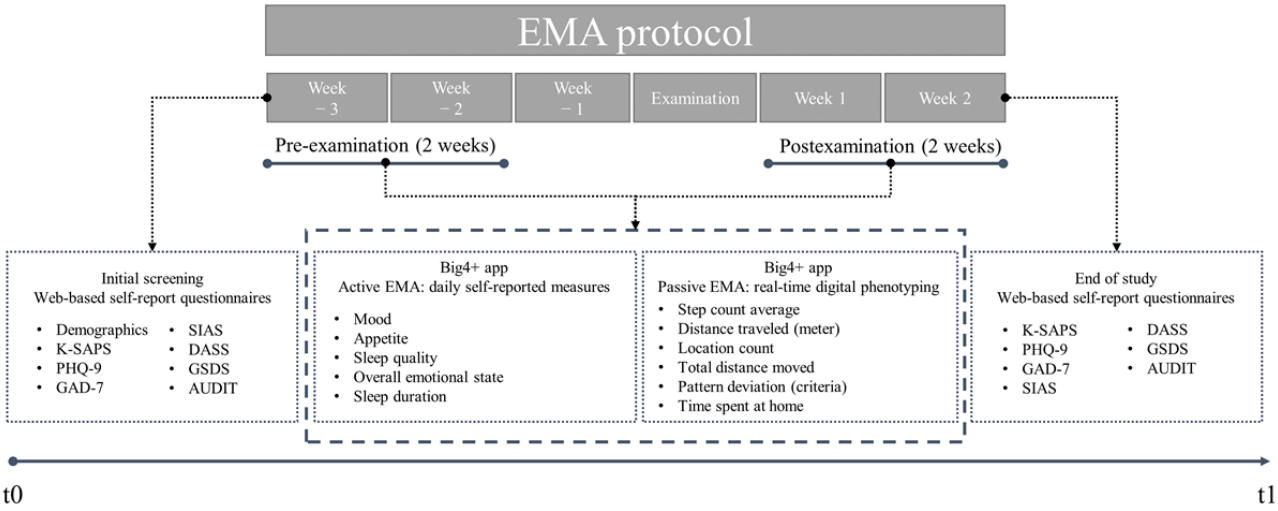

Participants

Participants were recruited in 2 waves from the Ulsan National Institute of Science and Technology in Korea during the fall semester, using both web-based and offline posters. Eligibility criteria included being a university student, aged between 19 and 28 years, a native Korean speaker, and owning a smartphone with either Android or iOS. Participants who objected to data collection or had recent suicidal ideation, severe mental illnesses (eg, schizophrenia and intellectual disability), or organic brain disorders were excluded. The study included a web-based survey that collected basic demographic information and responses using validated psychosocial screening tools. All participants provided informed consent, ensuring anonymity. In the first wave, data from 135 participants were collected from September 25 to November 6, 2023, whereas in the second wave, from November 20 to December 31, 2023, data from 108 participants were collected. On completion, participants received a personalized report summarizing their psychosocial assessments and digital phenotype analyzed by a psychiatrist.

Procedures

Web-Based Self-Report Questionnaires

Validated instruments were used to assess the psychological constructs related to smartphone use and mental health. The Korean Smartphone Addiction Proneness Scale (K-SAPS) evaluates excessive smartphone use through 15 items rated on a 4-point scale (ranging from 1 [not at all] to 4 [extremely]), categorizing use as high risk (scores ≥44), at risk (scores 40-43), or normal use (scores <40), with a Cronbach α of 0.814 [46].

The Patient Health Questionnaire-9 (PHQ-9) measured depressive symptoms over the past 2 weeks with 9 statements rated from 0 (not at all) to 3 (nearly every day), yielding a total score between 0 and 27, with higher scores indicating more severe depression [47].

Anxiety levels were assessed using the Generalized Anxiety Disorder-7 (GAD-7), which comprises 7 items scored from 0 to 3, with the total score reflecting the severity of anxiety symptoms [48].

To gauge social interaction anxiety (SIA), the Social Interaction Anxiety Scale (SIAS), comprising 20 items rated from 0 (not at all) to 4 (extremely), was used [49].

The Alcohol Use Disorders Identification Test (AUDIT) was used to screen for alcohol consumption and dependence using 10 statements. Scores >20 for men and >10 for women indicated high-risk drinking, with a Cronbach α of 0.92 [50].

Sleep quality was assessed using the General Sleep Disturbance Scale (GSDS), comprising 21 items on sleep problems experienced over the previous week, rated from 0 (never) to 7 (every day), with total scores ranging from 0 to 147; scores ≥43 indicate significant sleep disturbances [51].

Finally, stress levels were measured with the Depression Anxiety Stress Scales (DASS), which includes 14 items rated on a 4-point scale, ranging from 0 (did not apply to me at all) to 3 (applied to me very much or most of the time), with higher total scores indicating greater stress levels [52].

Smartphone Sensing and EMAs

All study participants were required to download and install the EMA digital phenotype data collection app *Big4+*, developed by the Department of Psychiatry, Yongin Severance Hospital, and Digital Medic, on their smartphones before the study began. Throughout the 6-week study period, the Big4+ app continuously collected digital phenotype data, such as distance traveled, step count, and sleep duration, while administering daily EMA surveys that assessed participants' mental health (mood, appetite, sleep, and overall condition). Participants could monitor the collected data directly through the app. Active EMA measures involved daily responses to 4 mental health–related questions concerning mood, appetite, sleep duration, and sleep quality via the Big4+ app. Each response was collected using a 7-point Likert scale and required approximately 2 minutes to complete. Examples of questions include "How was your mood?" and "Total hours slept last night." To facilitate engagement, app push notifications reminded users to complete

XSL•FO
RenderX

their daily EMA surveys. In parallel, passive EMA data were gathered through the app, which leverages digital phenotyping data from previous research [30]. Data collection commenced once participants provided consent and began using the app, capturing location data every 5 minutes to measure travel distances. This passive data acquisition included metrics such as step count, activity recognition, and sleep data. All the collected data were encrypted and securely stored in a server database. A comprehensive summary of the detailed digital phenotypes is presented in Textbox 1.

Textbox 1. Description of the digital phenotypes.

Digital phenotype and description

- Step count average: the average number of steps taken over a 5-minute interval
- Distance traveled (m): the average distance traveled over a 5-minute interval (m)
- Location count: the number of unique locations visited per day
- Total distance moved: the total distance traveled per day (km)
- Pattern deviation (criteria): the percentage of times a daily routine deviates by >100 m from the established pattern
- Time spent at most frequented location (home): The total time spent (minutes) at the location where the participant spent most of their time during the day

Statistical Analysis

To explore the factors related to PSU among university students before and after examinations, the participants with K-SAPS scores ≥40 were classified as the PSU group, while those with scores <40 were classified as the non-PSU group. Age-, gender-, and wave-matched participants without PSU were selected, resulting in a total of 112 participants, including 64 (57.1%) men and 48 (42.9%) women. Independent samples *t* tests (2-tailed) were used for numerical variables, and chi-square tests were used for categorical variables. Continuous variables were reported as mean SD, and categorical variables were reported as n (%).

To identify predictors of PSU, backward multivariable logistic regression analysis was conducted. PSU was the dependent variable, and significant variables in univariate analyses ($P<.05$) were the independent variables. Binary variables were created for each participant to indicate whether their scores on the PHQ-9, GAD-7, SIAS, DASS stress, GSDS, K-SAPS, and AUDIT met or exceeded standardized cutoff values, categorizing participants as “at risk” or “not at risk.” Cutoff values were set at ≥9 for PHQ-9; ≥5 for GAD-7; ≥34 for SIAS; ≥15 for DASS stress; ≥33 for GSDS; and ≥10 and ≥6 for AUDIT for men and women, respectively. In addition, EMA variables were categorized using median split, resulting in “above median” and “below median” groups.

A 2 (group)×2 (time) repeated measures ANOVA was performed to examine the between-group effect (PSU vs non-PSU), within-group effect (pre- and postexamination), and time×group interaction effect. Paired *t* tests compared pre- and postexamination scores within groups, and independent *t* tests assessed group differences at both time points, applying the Bonferroni correction ($P<.025$). All analyses were conducted using SPSS (version 27.0; IBM Corp) for Windows, with a significance of $P<.05$.

Ethical Considerations

This study is part of a collaborative project between the Department of Psychiatry at Yonsei University and the Department of Biomedical Engineering at Ulsan National Institute of Science and Technology. It was approved by the Ulsan National Institute of Science and Technology Institutional Review Board (UNISTIRB-23-040-A), and it adhered to the principles of the Declaration of Helsinki. All participants were informed about the study’s aims, methods, and data collection procedures and provided their written informed consent before participation. All collected data were anonymized to ensure participant privacy. Participants who completed the 6-week study were compensated with KRW 80,000 (approximately US $60) for their time and effort.

Results

Participants’ Demographic Information

A total of 243 students participated in the study, comprising 124 (51%) men and 119 (49%) women aged between 18 and 28 (mean 21.9, SD 1.9) years. The prevalence of PSU was 23% (56/243), with 31 (55%) participants with at-risk level of use (scores ≥40 and <43) and 25 (45%) participants with high-risk level of use (scores ≥43). Out of 243 participants, 56 participants with PSU were identified. To ensure comparability, a matched sample of 56 non-PSU participants was selected based on age, gender, and wave of data collection, resulting in a total matched sample size of 112 participants. This matched sample was used for detailed analyses in the study. Additional demographic characteristics are provided in Table 1.

XSL•FO
RenderX

Table 1. Descriptive characteristics of the study population.

	Total participants (N=243)	Study sample (n=112; 1:1 age, gender, and wave matched)[a]
Sample (N=243), n (%)	243 (100)	112 (46.1)
Age (y), mean (SD)	21.9 (1.9)	21.7 (1.8)
Gender, n (%)		
Man	124 (51)	48 (42.9)
Woman	119 (49)	64 (57.1)
Wave, n (%)		
Midterm	135 (55.1)	68 (60.7)
Final	108 (44.1)	44 (39.3)

[a]The study sample consists of 56 participants with problematic smartphone use (PSU) and 56 non-PSU participants matched 1:1 based on age, gender, and wave of data collection. Matching was conducted to ensure comparability between the groups.

Comparison Between the PSU and Non-PSU Groups at Baseline

The analysis revealed significant differences between the PSU and non-PSU groups in several measures: K-SAPS scores (mean 23.70, SD 5.18 vs mean 44.25, SD 4.22; *P*<.001), PHQ-9 (mean 3.59, SD 4.37 vs mean 6.39, SD 5.51; *P*=.004), GAD-7 (mean 2.18, SD 3.07 vs mean 5.05, SD 5.03; *P*<.001), SIAS (mean 23.21, SD 71.15 vs mean 37.16, SD 17.06; *P*<.001), DASS stress (mean 5.7, SD 8.5 vs mean 11.3, SD 12.2; *P*=.006), GSDS (mean 38.75, SD 17.74 vs mean 48.59, SD 17.27; *P*=.004), and AUDIT (mean 4.91, SD 3.60 vs mean 6.98, SD 4.04; *P*=.005), with the PSU group scoring significantly higher in these assessments. In addition, marginally significant differences were observed in the average distance traveled over a 5-minute interval (mean 157.79, SD 119.24 vs mean 221.87, SD 199.50; *P*=.046). However, daily mood, appetite, sleep, and overall emotional state scores derived from EMA surveys were lower in the PSU group than those in the non-PSU group, although these differences were not significant (Table 2).

XSL•FO
RenderX

Table 2. Comparison of means of the characteristics between the PSU[a] and non-PSU groups at baseline.

Variables	PSU, mean (SD)	Non-PSU, mean (SD)	*t* value (*df*)	*P* value
Psychosocial measures (PSU, n=56; non-PSU, n=56)				
Korean Smartphone Addiction Proneness Scale	44.3 (4.2)	23.7 (5.2)	23.0 (110)	<.001
Patient Health Questionnaire-9	6.4 (5.5)	3.6 (4.4)	3.0 (110)	.004
Generalized Anxiety Disorder-7	5.1 (5.0)	2.2 (3.1)	3.7 (110)	<.001
Social Interaction Anxiety Scale	37.2 (17.1)	23.2 (17.2)	4.3 (110)	<.001
Depression Anxiety Stress Scales stress	11.3 (12.2)	5.7 (8.5)	2.8 (110)	.006
General Sleep Disturbance Scale	48.6 (17.3)	38.8 (17.7)	3.0 (110)	.004
Alcohol Use Disorders Identification Test	7.0 (4.0)	3.6 (4.9)	2.9 (110)	.005
Passive EMA[b] (digital phenotype; PSU, n=52; non-PSU, n=54)				
Step count average	80.7 (47.2)	88.2 (47.3)	–0.8 (106)	.42
Distance traveled (m)	221.9 (199.5)	119.2 (157.8)	2.0 (106)	.046
Location count	2.7 (0.6)	2.6 (0.7)	0.6 (106)	.54
Total distance moved	45.8 (34.3)	57.9 (136.8)	–0.6 (106)	.53
Pattern deviation (criteria)	72.3 (24.9)	67.2 (27.0)	1.0 (106)	.32
Time spent at home	892.5 (230.9)	935.5 (263.1)	–0.9 (106)	.37
Active EMA (PSU, n=52; non-PSU, n=54)				
Mood	5.0 (1.1)	5.3 (1.3)	–1.5 (106)	.13
Appetite	5.0 (1.1)	5.3 (1.2)	–1.4 (106)	.16
Sleep quality	4.7 (1.0)	4.9 (1.2)	–1.0 (106)	.32
Overall emotional state	4.9 (1.0)	5.2 (1.2)	–1.3 (106)	.20
Sleep duration	430.3 (51.0)	412.9 (61.1)	1.6 (106)	.12

[a]PSU: problematic smartphone use.

[b]EMA: ecological momentary assessment.

Factors Associated With PSU in the Univariate Analysis

Univariate analysis revealed significant associations between PSU and multiple psychosocial measures (PHQ-9, GAD-7, SIAS, and DASS stress: all *P*<.01; GSDS: *P*=.02), with no significant difference for AUDIT (*P*=.08). Notably, the PSU group had higher rates of exceeding cutoff scores across the PHQ-9, GAD-7, SIAS, and GSDS. In relation to the digital phenotypes, a significant association was found only for the average distance traveled over a 5-minute interval (*P*=.02). Furthermore, in the active EMA data, only appetite was significantly associated with PSU (Table 3).

XSL•FO
RenderX

Table 3. Univariate analysis of the sample characteristics.

Variables	PSU[a], n (%)	Non-PSU, n (%)	χ^2 test (*df*)	*P* value
Psychosocial measures (PSU, n=56; non-PSU, n=56)				
Patient Health Questionnaire-9 (cutoff scores: 9)			13.2 (1)	<.001
Not at risk	38 (68)	53 (95)		
At risk	18 (32)	3 (5)		
Generalized Anxiety Disorder-7 (cutoff scores: 5)			11.7 (1)	.001
Not at risk	33 (59)	49 (88)		
At risk	23 (41)	7 (13)		
Social Interaction Anxiety Scale (cutoff scores: 34)			14.8 (1)	<.001
Not at risk	23 (41)	43 (77)		
At risk	33 (59)	13 (23)		
Depression Anxiety Stress Scales stress (cutoff scores: 15)			6.9 (1)	.009
Not at risk	36 (64)	48 (86)		
At risk	20 (36)	8 (14)		
General Sleep Disturbance Scale (cutoff scores: 33)			5.2 (1)	.02
Not at risk	11 (20)	22 (39)		
At risk	45 (80)	34 (61)		
Alcohol Use Disorders Identification Test (cutoff scores: male=10, female=6)			3.0 (1)	.08
Not at risk	29 (52)	38 (68)		
At risk	27 (48)	18 (32)		
Passive EMA[b] (digital phenotype; PSU, n=52; non-PSU, n=54)				
Step count average			0.15 (1)	.70
Above median	25 (48)	28 (52)		
Below median	27 (52)	26 (48)		
Distance traveled (m)			5.44 (1)	.02
Above median	32 (62)	21 (39)		
Below median	20 (39)	33 (61)		
Location count			0.6 (1)	.44
Above median	28 (54)	25 (46)		
Below median	24 (46)	29 (54)		
Total distance moved			1.36 (1)	.24
Above median	29 (56)	24 (44)		
Below median	23 (44)	30 (56)		
Pattern deviation (criteria)			1.36 (1)	.24
Above median	29 (56)	24 (44)		
Below median	23 (44)	30 (56)		
Time spent at home			1.36 (1)	.24
Above median	23 (44)	30 (56)		
Below median	29 (56)	24 (44)		
Active EMA (PSU, n=52; non-PSU, n=54)				
Mood			3.8 (1)	.05
Above median	21 (40)	32 (59)		

XSL•FO
RenderX

Variables	PSU[a], n (%)	Non-PSU, n (%)	χ^2 test (*df*)	*P* value
Below median	31 (60)	22 (41)		
Appetite			5.4 (1)	.02
Above median	21 (40)	34 (63)		
Below median	31 (60)	20 (37)		
Sleep quality			0.6 (1)	.44
Above median	24 (46)	29 (54)		
Below median	28 (54)	25 (46)		
Overall emotional state			3.8 (1)	.05
Above median	21 (40)	32 (59)		
Below median	31 (60)	22 (41)		
Sleep duration			1.4 (1)	.24
Above median	29 (56)	24 (44)		
Below median	23 (44)	30 (56)		

[a]PSU: problematic smartphone use.

[b]EMA: ecological momentary assessment.

Independent Predictors of PSU in Multivariable Logistic Regression

Table 4 summarizes the results of the multivariable logistic regression analysis exploring factors associated with PSU based on K-SAPS scores. This analysis included psychosocial measures and digital phenotype data from passive and active EMA, using baseline data collected 3 weeks before the examination. The final model (R^2=0.424) identified several independent predictors of PSU: PHQ-9 score above the cutoff (B=2.138; odds ratio [OR] 8.480, 95% CI 1.950-36.872), SIAS score above the cutoff (B=1.481; OR 4.398, 95% CI 1.592-12.148), GSDS score above the cutoff (B=1.234; OR 3.436, 95% CI 1.146-10.298), time spent at home below the median (B=–1.066; OR 3.436, 95% CI 1.146-10.298), and sleep duration above the median (B=1.134; OR 3.436, 95% CI 1.146-10.298). The variable “mean distance traveled per 5-minute interval” was not significant (*P*>.05) but was retained in the model. Higher levels of depressive symptoms, SIA, and lower sleep quality, along with less time spent at home, were associated with an increased risk of PSU before the examination.

Table 4. Multivariable logistic regression analysis of the effect of the factors on problematic smartphone use.

	B (SE)	Wald	*P* value	OR[a] (95% CI)
Patient Health Questionnaire-9	2.14 (0.75)	8.13	.004	8.48 (1.95-36.87)
Social Interaction Anxiety Scale	1.48 (0.52)	8.16	.004	4.40 (1.59-12.15)
General Sleep Disturbance Scale	1.23 (0.56)	4.86	.03	3.44 (1.15-10.30)
Mean distance traveled per 5-minute interval	0.82 (0.49)	2.83	.09	2.328 (0.87-5.93)
Time spent at home	–1.07 (0.51)	4.33	.04	0.35 (0.13-0.94)
Sleep duration	1.13 (0.51)	4.90	.03	3.11 (1.14-8.48)
Constant	–2.33 (0.67)	11.91	.001	0.10[b]

[a]OR: odds ratio.

[b]CI not applicable.

Comparison of Pre- and Postexamination Psychosocial Measures and EMA Data Between PSU and Non-PSU Groups

Repeated measures ANOVA compared changes in smartphone addiction proneness, depression, SIA, sleep disturbance, time spent at home, and sleep duration between the PSU and non-PSU groups before and after the examination. The results revealed a significant group×time interaction effect for smartphone addiction proneness ($F_{1,\ 110}$=370.29; *P*<.001; p^2=0.771), indicating differing patterns of change (Table 5). Specifically, the PSU group exhibited a decrease in average smartphone addiction scores from 44.25 (SD 4.22) to 41.52 (SD 6.72), while the non-PSU group showed an increase from 23.70 (SD 5.18) to 25.13 (SD 6.58). In addition, significant differences were found in depression and SIA. Depression scores in the PSU group decreased from 6.39 (SD 5.51) to 4.48 (SD 5.19), and SIA scores decreased significantly from 37.16 (SD 17.06) to

XSL•FO
RenderX

33.71 (SD 18.52). In contrast, depression scores in the non-PSU group significantly decreased from 3.59 (SD 4.37) to 2.55 (SD 3.13), with no substantial change in SIA scores (Table 6). Furthermore, a repeated measures ANOVA revealed significant changes in sleep disturbance, sleep duration, and time spent at home in both groups before and after the examination. A significant main effect of time for sleep disturbance ($F_{1,110}$=5.13; P=.02) and a significant group effect ($F_{1,110}$=10.94; P=.001) were observed. However, the group×time interaction was not significant ($F_{1,110}$=.062; P=.80; Table 5). Specifically, sleep disturbance or time spent at home did not differ significantly (both P values >.05); although sleep duration changed significantly (P=.03), it was not significant after applying the Bonferroni correction. In the non-PSU group, while sleep disturbance did not differ significantly (P=.07), time spent at home did (P=.03), although this was also nonsignificant after correction. However, a significant difference in sleep duration was observed (P=.01) (Table 6). Overall, while significant main effects of time were observed for sleep disturbance, time spent at home, and sleep duration (Table 5), paired t test results revealed no significant pre- to postexamination differences in sleep disturbance or time spent at home for the groups. However, only the non-PSU group demonstrated a significant increase in sleep duration after the examination, with no change observed in the PSU group (Figure 2; Table 6).

Table 5. Group, time, and interaction effects on study variables: repeated measures ANOVA results.

Variables	Group			Time			Group×time interaction		
	F test (*df*)	p^2	*P* value	*F* test (*df*)	p^2	*P* value	*F* test (*df*)	p^2	*P* value
K-SAPS[a]	370.29 (1, 110)	0.771	<.001	1.59 (1, 110)	0.014	.21	16.24 (1, 110)	0.129	<.001
PHQ-9[b]	8.55 (1, 110)	0.072	.004	18.98 (1, 110)	0.147	<.001	1.67 (1, 110)	0.015	.20
SIAS[c]	14.79 (1, 110)	0.119	<.001	3.96 (1, 110)	0.035	.049	4.58 (1, 110)	0.04	.04
GSDS[d]	10.94 (1, 110)	0.090	.001	5.13 (1, 110)	0.045	.02	0.06 (1, 110)	0.001	.80
Time spent at home	1.67 (1, 101)	0.016	.20	7.27 (1, 101)	0.067	.008	0.51 (1, 101)	0.005	.48
Sleep duration	2.71 (1, 101)	0.026	.10	17.59 (1, 101)	0.148	<.001	1.40 (1, 101)	0.014	.24

[a]K-SAPS: Korean Smartphone Addiction Proneness Scale.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]SIAS: Social Interaction Anxiety Scale.

[d]GSDS: General Sleep Disturbance Scale.

XSL•FO
RenderX

Table 6. Mean differences between pre-examination (baseline) and postexamination evaluations for both groups.

Group and variables	Pre-examination, mean (SD)	Postexamination, mean (SD)	Diff (95% CI)[a]	*t* test (*df*)	Cohen *d*	*P* value[b]
PSU[c]						
K-SAPS[d]	44.25 (4.22)	41.52 (6.72)	2.73 (1.20 to 4.26)	3.58 (55)	0.479	.001
PHQ-9[e]	6.39 (5.51)	4.48 (5.19)	1.91 (0.81 to 3.01)	3.48 (55)	0.466	.001
SIAS[f]	37.16 (17.06)	33.71 (18.52)	3.45 (0.87 to 6.02)	2.68 (55)	0.358	.01
GSDS[g]	48.59 (17.27)	45.84 (18.33)	2.75 (–1.22 to 6.72)	1.39 (55)	0.185	.17
Time spent at home	892.52 (230.93)	941.86 (262.94)	–46.22 (–108.81 to 16.38)	–1.48 (50)	–0.208	.14
Sleep duration	430.29 (51.03)	449.40 (54.45)	–15.85 (–30.30 to –1.40)	–2.20 (50)	–0.308	.03
Non-PSU						
K-SAPS	23.7 (5.18)	25.13 (6.58)	–1.43 (–2.82 to –0.03)	–2.05 (55)	–0.274	.045
PHQ-9	3.59 (4.37)	2.55 (3.13)	1.04 (0.24 to 1.82)	2.62 (55)	0.35	.01
SIAS	23.21 (17.15)	23.34 (16.42)	–0.13 (–2.26 to 2.01)	–0.12 (55)	–0.016	.91
GSDS	38.75 (17.74)	35.32 (17.88)	3.43 (–0.33 to 7.18)	1.83 (55)	0.245	.07
Time spent at home	935.51 (263.07)	1019.91 (307.08)	–79.63 (–149.21 to –10.04)	–2.30 (51)	–0.319	.03
Sleep duration	412.89 (61.13)	441.11 (44.43)	–28.32 (–43.73 to –12.90)	–3.69 (51)	–0.511	.001

[a]Difference represents the difference between the pre-examination and postexamination mean values (pre-examination – postexamination).

[b]Significance set at $P<.025$ (after Bonferroni correction).

[c]PSU: problematic smartphone use.

[d]K-SAPS: Korean Smartphone Addiction Proneness Scale.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]SIAS: Social Interaction Anxiety Scale.

[g]GSDS: General Sleep Disturbance Scale.

XSL•FO
RenderX

Figure 2. Comparative analysis of pre- and postexamination, psychological and behavioral changes in problematic smartphone use (PSU) and non-PSU groups: (A) PSU (K-SAPS), (B) depression (PHQ-9), (C) social interaction anxiety (SIAS), (D) sleep disturbance (GSDS), (E) time spent at home, and (F) sleep duration. GSDS: General Sleep Disturbance Scale; K-SAPS: the Korean Smartphone Addiction Proneness Scale; PHQ-9: Patient Health Questionnaire-9; SIAS: Social Interaction Anxiety Scale.

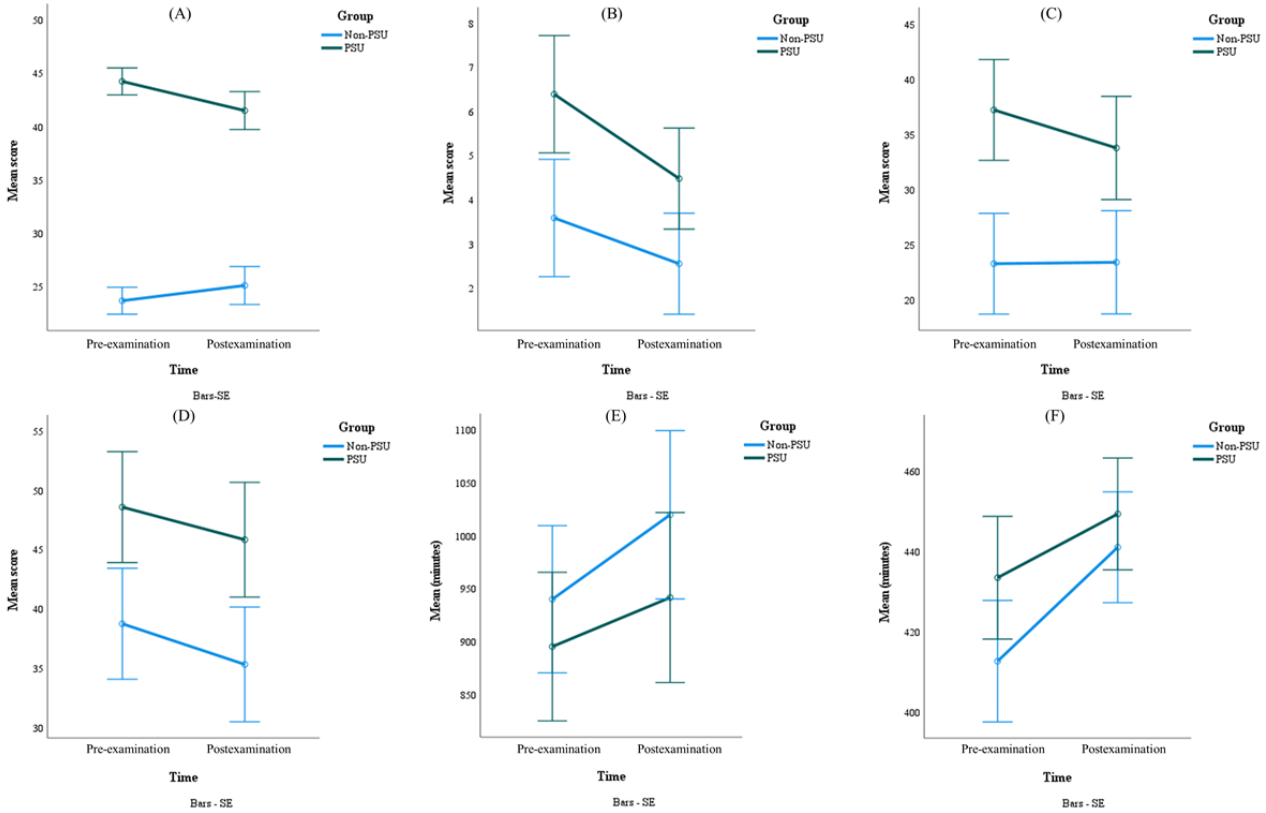

Discussion

Principal Findings

This study identified significant associations between PSU and both questionnaire-based demographic and psychosocial measures as well as daily EMA that were either actively or passively collected from a large sample of university students in South Korea. Independent predictors of PSU from psychosocial assessments at baseline included depression, SIA, and sleep disturbance. Conversely, data collected using app-based EMA from baseline to 2 weeks leading up to 1 week before the examination highlighted less time spent at home and longer sleep duration as significant predictors. These findings provide valuable insights into individual daily patterns through real-time, naturalistic measures.

Comparison With Prior Work

The prevalence of PSU in this study was 23% (56/243), comparable to the 21.7% prevalence reported among Serbian medical students [14] but slightly higher than the prevalence reported among Korean university students [1,53,54]. However, it was lower than the 39.7% prevalence reported by Liu et al [6] regarding smartphone addiction in China. The prevalence of smartphone addiction may vary due to several factors, including differences in measurement scales (eg, cutoff values), sociocultural background, educational environment, sample size, distribution, and technological advancements. Rapid technological progression has likely led individuals to become increasingly accustomed to smartphone use, potentially contributing to the higher prevalence observed in recent research.

Comparison of psychosocial characteristics revealed that the PSU group exhibited significantly higher levels of negative emotions, including depression, anxiety, and stress, as well as more severe sleep disturbances than the non-PSU group. These results suggest a correlation between higher PSU and increased psychological distress and disrupted sleep, consistent with prior findings linking PSU to emotional dysregulation and reduced sleep quality [8,17,55-57]. A significant association was found between PSU and clinical risks for adverse mental health, with a larger proportion of the PSU group classified as being “at risk” for depression, anxiety, stress, and general sleep disturbances compared to the non-PSU group. This suggests a greater likelihood for individuals with PSU to meet clinical thresholds for psychological distress. These findings align with those of a study involving Turkish college students [17], which similarly reported significant associations between smartphone use and elevated depression, anxiety, and poor sleep quality. In addition, recent studies provide evidence of consistent links between smartphone use and psychological factors such as stress, anxiety, and depression [58-61]. Our study reinforces these findings, demonstrating significant associations between PSU and increased risks of depression, anxiety, stress, and sleep disturbances.

Analyzing app-based daily active and passive EMA, this study found distinct GPS-derived movement patterns in individuals with PSU compared to those without it. The digital phenotype

XSL•FO
RenderX

data of the PSU group revealed distinct movement patterns, including greater average distances traveled over 5-minute intervals, spending less time at home, and more frequent deviations from their habitual routes, with some movements exceeding 100 m. These patterns, including deviations >100 m from regular paths, may suggest behavioral tendencies such as impulsivity, emotional avoidance, or a heightened pursuit of immediate rewards. Similarly, although the PSU group traveled shorter total distances compared to the non-PSU group (PSU: mean 45.8, SD 34.3 km vs non-PSU: mean 57.9, SD 136.8 km), participants traveled significantly greater distances within 5-minute intervals (PSU: mean 221.9, SD 199.5 m vs non-PSU: mean 119.2, SD 157.8 m). This suggests that while the PSU group covered less ground overall, their movement patterns, characterized by traveling further in shorter periods, may reflect higher impulsivity. The contrast highlights the complex relationship between PSU and impulsive behavior, suggesting that those with PSU may exhibit more erratic movement patterns. These findings align with those of previous studies that link impulsivity to smartphone overuse and related movement behaviors. For example, Pérez de Albéniz Garrote et al [62] found that individuals with high impulsivity often seek novelty and have increased smartphone engagement. Schulz van Endert and Mohr [63] suggested that GPS-tracked movement data can offer objective insights into impulsivity-related smartphone use. Although these studies did not focus explicitly on GPS data, they support our observation [62,63].

The study identified independent significant predictors of PSU in college students before examinations, including psychosocial factors such as elevated levels of depression, SIA, and sleep disturbances. The digital phenotype features extracted from passive EMA revealed that a longer average “distance traveled” (mean distance covered per 5-min interval, measured in meters) and less “time spent at home” (total daily duration at the primary location, measured in minutes) were distinguishing characteristics for students with PSU. In addition, longer sleep duration was an independent predictor in the PSU group. Together, these findings suggest that students with PSU exhibit distinct mobility and sleep patterns before examinations, potentially indicative of broader behavioral tendencies associated with smartphone overuse. Studies on impulsivity and movement patterns [64,65] have shown that individuals with high impulsivity tend to exhibit irregular movement patterns and greater range in travel, supporting the notion that impulsive behaviors may be tied to broader, real-world activity patterns. This could reflect their use of erratic behaviors as a coping mechanism for stress or emotional avoidance.

Specifically, after the examination, smartphone addiction proneness, SIA, and depression were significantly reduced in the PSU group, suggesting that pre-examination PSU may serve as an avoidance coping strategy for academic stress. These results indicate that academic stress can exacerbate mental health vulnerabilities, leading to PSU or worsening existing addiction. Deteriorating mental health is one of the primary risks for smartphone addiction, with addictive behaviors to manage stress and alleviate negative emotions. PSU students may use smartphones to reduce stress and improve mood, increasing the risk of addiction. This finding aligns with previous findings on stress-related digital dependence behaviors. For example, Choi [66] suggests that stress influences smartphone addiction through self-regulation. Moreover, this result is consistent with Kardefelt-Winther’s [67] compensatory internet use model, which proposes that smartphones can help individuals disengage from stress-inducing social situations, with the need for such behavior diminishing as academic pressures ease. These findings are consistent with those of previous studies [68], showing that as stress decreases, smartphone dependence for emotional regulation may also decline. In contrast, the non-PSU group showed a significant increase in sleep duration after examinations, reflecting distinct adaptation mechanisms between the 2 groups.

Another key finding of this study is that SIA was a strong independent predictor of PSU in the students. Unlike general anxiety, SIA is particularly linked to maladaptive smartphone behaviors, potentially due to mechanisms unique to social anxiety that intensify under academic pressures. The heightened vulnerability of students with PSU seems to be compensated for by the portability and flexibility of smartphone apps, which allow individuals to supplement deficiencies in real-life social interactions [6]. These results suggest that high levels of SIA may negatively impact the mental well-being and behavior of students, emphasizing the need for targeted interventions that specifically address SIA. Addressing SIA during examination preparation may be particularly beneficial in supporting the mental health and academic success of students. This finding aligns with reports by Beidel and Turner [69] and Wong et al [70], highlighting the impact of social anxiety on behavioral and academic outcomes, further validating the importance of addressing social interaction challenges within educational settings [69-72].

The findings from this study highlight distinct shifts in smartphone addiction proneness and psychosocial factors between the PSU and non-PSU groups, especially around examination periods. This work demonstrates how varying psychosocial vulnerabilities and coping mechanisms shape smartphone use patterns, with those having PSU exhibiting higher susceptibility to stress and a tendency to rely on smartphones for emotional regulation. Our results emphasize the impact of PSU on mental health and behaviors, linking it to changes in SIA and depression as well as to physical mobility and sleep. These findings are consistent with previous research [67], suggesting that individuals may use digital devices as coping mechanisms for emotional stress. Moreover, the effects of PSU reach beyond digital habits, influencing real-life routines, as shown by studies on movement patterns [64,65]. In conclusion, this study supports the need for interventions focused on reducing SIA and fostering healthy coping strategies among students, aiming to improve both their mental well-being and academic outcomes. Future research should examine the complex interaction between digital behaviors and real-world outcomes, paving the way for a comprehensive understanding of the broader impact of PSU.

XSL•FO
RenderX

Limitations

This study has several limitations. The generalizability of this study is limited, as the sample comprised students from a single 4-year undergraduate university in Ulsan, South Korea. Consequently, these findings should be applied to different geographical areas or cultural contexts with caution. The sample's unique characteristics, particularly the fact that all students resided in dormitories, are likely to have resulted in restricted movement patterns, with most mobility occurring within the campus or its immediate vicinity. As all students shared the same institutional environment, they were subject to a uniform set of norms and rules that regulate and constrain behavior, including mobility patterns. This homogeneity extends beyond sociodemographic characteristics to encompass behavioral consistency [73,74], particularly during examination periods. During these high-stress academic periods, shared examination schedules, the necessity of efficient time management, and social pressures to engage in study routines further amplified similarities in daily patterns. Consequently, the findings should also be interpreted within the unique context of academic stress in this demographic, as examination-related pressure is known to influence mobility behaviors and psychosocial factors. While this homogeneity provides methodological advantages, such as reduced variability and enhanced internal validity, it limits the study's applicability to more diverse populations [73]. To address this limitation, future research should incorporate diverse geographic, educational, and cultural samples to strengthen the external validity of these findings. Including participants from various universities, regions, and cultural contexts would allow for a more comprehensive assessment of the broader applicability of the results. Expanding the sample to include nonstudent populations or community-based cohorts would also help determine whether the observed relationships are unique to undergraduate students or generalizable to individuals with more varied daily routines and environmental constraints.

This study used both active and passive EMA to capture real-time data and digital phenotypes, focusing primarily on GPS-based movement patterns. However, it did not include objective smartphone use metrics, such as screen time or app use patterns. Previous research underscores the critical role of passive smartphone sensing in identifying impulsive behaviors and self-regulation deficits, which often underlie excessive smartphone use [55,75-79]. For instance, hallmark features of PSU, such as compulsive checking and loss of control over smartphone use, can be quantified through objective metrics, providing clearer insights into behavioral manifestations [76,77]. Wen et al [76] presented preliminary evidence linking impulsivity traits to mobile device use. Passive measures, including call logs, battery consumption, and screen time, were found to predict various aspects of impulsivity and impulsive behaviors in nonclinical populations. These findings highlight the importance of integrating diverse passive smartphone metrics into PSU research. Despite the growing body of evidence supporting the utility of passive sensing methods, our study encountered technical challenges in fully using these data sources. Although we successfully extracted passive smartphone metrics, such as app use and screen time, discrepancies in data accessibility between Android and iOS devices limited their integration. For instance, detailed app-level metrics, such as background use and screen time, were available on Android devices but not on iOS devices, which only provided generalized metrics, such as screen-on and screen-off duration. This platform-specific limitation prevented the harmonization of these data streams, ultimately leading to their exclusion from the final dataset. Addressing such technical barriers, including enhancing the cross-platform generalizability of mobile sensing models, is crucial for future research [76]. Overcoming these limitations would enable researchers to integrate diverse data streams, reduce reliance on self-reports, and provide a more accurate and nuanced assessment of PSU.

The results of this study identify significant independent predictors of PSU, including depression, SIA, and sleep disturbance, through regression analysis. While these findings suggest robust associations after adjusting for potential confounders, they are correlational and should not be interpreted as causal. The cross-sectional nature of the baseline psychosocial assessments limits our ability to infer temporal or causal relationships. Future studies using longitudinal or experimental designs will be necessary to establish causality and to explore the underlying mechanisms of these associations. While this study revealed significant differences in GPS-derived movement patterns between the PSU and non-PSU groups, interpreting these patterns as primarily driven by impulsivity or emotional avoidance remains speculative. Several alternative explanations warrant consideration. First, these movement patterns may reflect stress-related coping strategies, wherein individuals engage in physical displacement to manage heightened stress or anxiety during examination periods. Previous research has demonstrated that traveling longer distances is associated with lower levels of stress, greater positive affect, and reduced anxiety. On a daily basis, participants traveling longer distances reported lower levels of stress and anxiety and more positive emotions, while higher levels of routine activity were linked to lower rates of depression and loneliness [80,81]. These findings align with the results of our study. In this study, the non-PSU group traveled significantly longer total daily distances compared with the PSU group; reported significantly lower levels of depression, anxiety, and stress; and exhibited significantly lower rates of routine deviations. The PSU group exhibited shorter total daily distances traveled, but when averaged over 5-minute intervals, participants covered greater distances. This pattern indicates that the PSU group tended to travel farther within shorter periods, suggesting that such increased high-density (short-time long-distance) travel or movement may act as a behavioral response to stress, providing temporary relief or distraction. Second, environmental or lifestyle factors may also contribute to these patterns. Variations in daily schedules, modes of transportation, and academic routines could influence mobility behaviors. For example, students commuting by car or bike may exhibit different movement patterns than those traveling on foot. Similarly, evening activities such as part-time jobs or social gatherings could explain deviations from habitual routes [81]. Third, personal traits, such as conscientiousness or agreeableness, may influence these behaviors. Previous studies have shown that highly agreeable individuals tend to engage in more social

XSL•FO
RenderX

interactions and visit a greater variety of locations, whereas those with high conscientiousness often follow structured routines and visit fewer places. These traits may interact with PSU, affecting both mobility and smartphone use behaviors [80]. To further investigate the underlying mechanisms, future studies should incorporate additional behavioral and psychological measures, such as momentary assessments of stress, anxiety, and smartphone use during movement episodes. Integrating these data with GPS metrics could help clarify the roles of impulsivity, emotional avoidance, and other psychological factors in driving these patterns. While this study primarily focused on behavioral patterns, it did not directly assess personal traits such as impulsivity or coping mechanisms, which could provide deeper insight into the psychological factors influencing PSU and mobility. Future research should incorporate validated scales for impulsivity and coping styles, alongside EMA, to better understand how students manage stress and emotional challenges through smartphone use.

Conclusions

Despite these limitations, the contributions of this study to understanding the relationship between smartphone use and mental health during academic examination periods are significant. First, it was identified that a significantly higher proportion of students with PSU have high risk for depression, anxiety, stress, and sleep disturbances. This underscores the heightened vulnerability of students with PSU under examination-related stress. In addition, this study is among the first to combine active and passive EMA over a 2-week, pre-examination period to capture real-time data on students' mobility and digital phenotypes. This innovative approach enabled a detailed comparison of movement patterns and behaviors between PSU and non-PSU groups, revealing new insights into how PSU may present in daily life, particularly during high-stress periods. Finally, by including both pre- and postexamination data, this study observed fluctuations in behavioral patterns and mental health outcomes over time, offering a dynamic perspective on the interactions between smartphone use and mental health during critical academic events. These findings align with the concerns raised in the background regarding the negative consequences of PSU. Specifically, this study demonstrated that PSU is associated with heightened risks of depression, anxiety, and stress as well as disrupted sleep patterns, which often exacerbate each other during high-stress periods such as examinations, leading to a compounded negative impact on overall well-being. These results support existing evidence linking PSU to emotional dysregulation, reduced sleep quality, and increased psychological distress. Moreover, the significant associations observed between PSU and clinical risks highlight the importance of addressing these vulnerabilities, particularly during high-stress periods such as examinations. By contextualizing PSU as both a symptom and a contributor to psychological distress, this study emphasizes the need for targeted interventions aimed at mitigating the broader impacts of PSU on students' mental health and academic performance. Together, these findings represent a meaningful step forward in understanding the real-world implications of PSU in students under academic pressure, providing a foundation for future research in this area.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant: RS-2024-00439193).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JSA contributed to conceptualization, data curation, formal analysis, methodology, visualization, and writing—original draft preparation. IJ contributed to data curation and investigation. SP and JL contributed to data curation and software. MJ contributed to resources. GD contributed to resources and software. SL contributed to conceptualization and resources. DJ contributed to conceptualization, funding acquisition, project administration, supervision, validation, and writing—reviewing and editing. JYP contributed to conceptualization, funding acquisition, project administration, supervision, validation, and writing—reviewing and editing. All authors take responsibility for data integrity, and all authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Korean version of psychosocial assessment questionnaires and daily ecological momentary assessment survey.
[PDF File (Adobe PDF File), 373 KB - jmir_v27i1e69320_app1.pdf]

References

XSL•FO
RenderX

1. Kim Y, Jeong JE, Cho H, Jung D, Kwak M, Rho MJ, et al. Personality factors predicting smartphone addiction predisposition: behavioral inhibition and activation systems, impulsivity, and self-control. PLoS One 2016;11(8):e0159788 [FREE Full text] [doi: 10.1371/journal.pone.0159788] [Medline: 27533112]
2. Monacis L, de Palo V, Griffiths MD, Sinatra M. Social networking addiction, attachment style, and validation of the Italian version of the Bergen Social Media Addiction Scale. J Behav Addict 2017 Jun 01;6(2):178-186 [FREE Full text] [doi: 10.1556/2006.6.2017.023] [Medline: 28494648]
3. Wu AM, Cheung VI, Ku L, Hung EP. Psychological risk factors of addiction to social networking sites among Chinese smartphone users. J Behav Addict 2013 Sep;2(3):160-166 [FREE Full text] [doi: 10.1556/JBA.2.2013.006] [Medline: 25215198]
4. Rosen LD, Whaling K, Carrier LM, Cheever NA, Rokkum J. The media and technology usage and attitudes scale: an empirical investigation. Comput Human Behav 2013 Nov 01;29(6):2501-2511 [FREE Full text] [doi: 10.1016/j.chb.2013.06.006] [Medline: 25722534]
5. Kim HJ, Min JY, Kim HJ, Min KB. Accident risk associated with smartphone addiction: a study on university students in Korea. J Behav Addict 2017 Dec 01;6(4):699-707 [FREE Full text] [doi: 10.1556/2006.6.2017.070] [Medline: 29099234]
6. Liu H, Zhou Z, Zhu E, Huang L, Zhang M. Smartphone addiction and its associated factors among freshmen medical students in China: a cross-sectional study. BMC Psychiatry 2022 May 02;22(1):308 [FREE Full text] [doi: 10.1186/s12888-022-03957-5] [Medline: 35501728]
7. Grant JE, Lust K, Chamberlain SR. Problematic smartphone use associated with greater alcohol consumption, mental health issues, poorer academic performance, and impulsivity. J Behav Addict 2019 Jun 01;8(2):335-342 [FREE Full text] [doi: 10.1556/2006.8.2019.32] [Medline: 31257917]
8. Xie X, Dong Y, Wang J. Sleep quality as a mediator of problematic smartphone use and clinical health symptoms. J Behav Addict 2018 Jun 01;7(2):466-472 [FREE Full text] [doi: 10.1556/2006.7.2018.40] [Medline: 29788754]
9. Demir YP, Sumer MM. Effects of smartphone overuse on headache, sleep and quality of life in migraine patients. Neurosciences (Riyadh) 2019 Apr 13;24(2):115-121 [FREE Full text] [doi: 10.17712/nsj.2019.2.20180037] [Medline: 31056543]
10. Mustafaoglu R, Yasaci Z, Zirek E, Griffiths MD, Ozdincler AR. The relationship between smartphone addiction and musculoskeletal pain prevalence among young population: a cross-sectional study. Korean J Pain 2021 Jan 01;34(1):72-81 [FREE Full text] [doi: 10.3344/kjp.2021.34.1.72] [Medline: 33380570]
11. Kee IK, Byun JS, Jung JK, Choi JK. The presence of altered craniocervical posture and mobility in smartphone-addicted teenagers with temporomandibular disorders. J Phys Ther Sci 2016 Jan;28(2):339-346 [FREE Full text] [doi: 10.1589/jpts.28.339] [Medline: 27065516]
12. Alhassan AA, Alqadhib EM, Taha NW, Alahmari RA, Salam M, Almutairi AF. The relationship between addiction to smartphone usage and depression among adults: a cross sectional study. BMC Psychiatry 2018 May 25;18(1):148 [FREE Full text] [doi: 10.1186/s12888-018-1745-4] [Medline: 29801442]
13. Lanaj K, Johnson RE, Barnes CM. Beginning the workday yet already depleted? Consequences of late-night smartphone use and sleep. Organ Behav Hum Decis Process 2014 May;124(1):11-23. [doi: 10.1016/j.obhdp.2014.01.001]
14. Nikolic A, Bukurov B, Kocic I, Vukovic M, Ladjevic N, Vrhovac M, et al. Smartphone addiction, sleep quality, depression, anxiety, and stress among medical students. Front Public Health 2023;11:1252371 [FREE Full text] [doi: 10.3389/fpubh.2023.1252371] [Medline: 37744504]
15. Elhai JD, Levine JC, Dvorak RD, Hall BJ. Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use. Comput Human Behav 2016 Oct;63:509-516. [doi: 10.1016/j.chb.2016.05.079]
16. Lepp A, Barkley JE, Karpinski AC. The relationship between cell phone use and academic performance in a sample of U.S. college students. Sage Open 2015 Feb 19;5(1):215824401557316. [doi: 10.1177/2158244015573169]
17. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92 [FREE Full text] [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
18. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5th edition. Washington, DC: American Psychiatric Association; 2013.
19. International classification of diseases (ICD). World Health Organization. URL: http://icd.who.int/en [accessed 2024-11-19]
20. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
21. Shiffman S. Conceptualizing analyses of ecological momentary assessment data. Nicotine Tob Res 2014 May;16 Suppl 2(Suppl 2):S76-S87 [FREE Full text] [doi: 10.1093/ntr/ntt195] [Medline: 24323571]
22. Yim SJ, Lui LM, Lee Y, Rosenblat JD, Ragguett R, Park C, et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J Affect Disord 2020 Sep 01;274:602-609. [doi: 10.1016/j.jad.2020.05.116] [Medline: 32663993]
23. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009 Dec;21(4):457-462 [FREE Full text] [doi: 10.1037/a0017653] [Medline: 19947780]

XSL•FO
RenderX

24. Wenze SJ, Miller IW. Use of ecological momentary assessment in mood disorders research. Clin Psychol Rev 2010 Aug;30(6):794-804. [doi: 10.1016/j.cpr.2010.06.007] [Medline: 20619520]
25. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
26. Di Matteo D, Fotinos K, Lokuge S, Yu J, Sternat T, Katzman MA, et al. The relationship between smartphone-recorded environmental audio and symptomatology of anxiety and depression: exploratory study. JMIR Form Res 2020 Aug 13;4(8):e18751 [FREE Full text] [doi: 10.2196/18751] [Medline: 32788153]
27. Barnett S, Huckvale K, Christensen H, Venkatesh S, Mouzakis K, Vasa R. Intelligent sensing to inform and learn (InSTIL): a scalable and governance-aware platform for universal, smartphone-based digital phenotyping for research and clinical applications. J Med Internet Res 2019 Nov 06;21(11):e16399 [FREE Full text] [doi: 10.2196/16399] [Medline: 31692450]
28. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020 Sep 01;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
29. Henson P, Barnett I, Keshavan M, Torous J. Towards clinically actionable digital phenotyping targets in schizophrenia. NPJ Schizophr 2020 May 05;6(1):13 [FREE Full text] [doi: 10.1038/s41537-020-0100-1] [Medline: 32372059]
30. Kamath J, Leon Barriera R, Jain N, Keisari E, Wang B. Digital phenotyping in depression diagnostics: integrating psychiatric and engineering perspectives. World J Psychiatry 2022 Mar 19;12(3):393-409 [FREE Full text] [doi: 10.5498/wjp.v12.i3.393] [Medline: 35433319]
31. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
32. Lee HJ, Cho CH, Lee T, Jeong J, Yeom JW, Kim S, et al. Prediction of impending mood episode recurrence using real-time digital phenotypes in major depression and bipolar disorders in South Korea: a prospective nationwide cohort study. Psychol Med 2023 Sep 23;53(12):5636-5644. [doi: 10.1017/S0033291722002847] [Medline: 36146953]
33. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
34. Stone AA, Shiffman S. Ecological momentary assessment (Ema) in behavioral medicine. Bioessays 1994 Mar;16(3):199-202. [doi: 10.1093/abm/16.3.199]
35. Silk JS, Forbes EE, Whalen DJ, Jakubcak JL, Thompson WK, Ryan ND, et al. Daily emotional dynamics in depressed youth: a cell phone ecological momentary assessment study. J Exp Child Psychol 2011 Oct;110(2):241-257 [FREE Full text] [doi: 10.1016/j.jecp.2010.10.007] [Medline: 21112595]
36. Lee J, Won S, Chang SM, Kim BS, Lee SJ. Prevalence of addictive behaviors in medical students and their association with stress. Psychiatry Investig 2022 Jan;19(1):44-53 [FREE Full text] [doi: 10.30773/pi.2021.0096] [Medline: 35086191]
37. Thakur SS, Roy RB. Predicting mental health using smart-phone usage and sensor data. J Ambient Intell Human Comput 2020 Nov 02;12(10):9145-9161. [doi: 10.1007/S12652-020-02616-5]
38. Chen X, Rubin KH, Li BS. Depressed mood in Chinese children: relations with school performance and family environment. J Consult Clin Psychol 1995 Dec;63(6):938-947. [doi: 10.1037//0022-006x.63.6.938] [Medline: 8543716]
39. Schraml K, Perski A, Grossi G, Simonsson-Sarnecki M. Stress symptoms among adolescents: the role of subjective psychosocial conditions, lifestyle, and self-esteem. J Adolesc 2011 Oct;34(5):987-996. [doi: 10.1016/j.adolescence.2010.11.010] [Medline: 21147499]
40. Moon KS. Academic stress and mental health in adolescents: the role of self-control and emotion regulation. Korean J Child Stud 2008;29(5):285-299 [FREE Full text]
41. Jun S, Choi E. Academic stress and internet addiction from general strain theory framework. Comput Human Behav 2015 Aug;49:282-287. [doi: 10.1016/j.chb.2015.03.001]
42. Ju SJ. Verification of mediation effect and moderation of resilience in relations between stress caused by study and parents that adolescents perceive and internet addiction. Korean J Youth Stud 2011;18(11):61-83 [FREE Full text]
43. Mao Y, Zhang N, Liu J, Zhu B, He R, Wang X. A systematic review of depression and anxiety in medical students in China. BMC Med Educ 2019 Sep 02;19(1):327 [FREE Full text] [doi: 10.1186/s12909-019-1744-2] [Medline: 31477124]
44. Albursan IS, Al Qudah MF, Al-Barashdi HS, Bakhiet SF, Darandari E, Al-Asqah SS, et al. Smartphone addiction among university students in light of the COVID-19 pandemic: prevalence, relationship to academic procrastination, quality of life, gender and educational stage. Int J Environ Res Public Health 2022 Aug 22;19(16):93 [FREE Full text] [doi: 10.3390/ijerph191610439] [Medline: 36012073]
45. Kwak JY, Kim JY, Yoon YW. Effect of parental neglect on smartphone addiction in adolescents in South Korea. Child Abuse Negl 2018 Mar;77:75-84. [doi: 10.1016/j.chiabu.2017.12.008] [Medline: 29306184]
46. Shin G, Kim D, Jung Y, Lee J, Lee Y, Kim M, et al. Development of Korean Smartphone addiction proneness scale for youth. Ministry of Education. URL: https://www.schoolhealth.kr/web/search/selectTotalSearchList.do?bbsId=&lstnum1=1530 [accessed 2024-04-29]

XSL•FO
RenderX

47. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
48. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
49. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6] [Medline: 9670605]
50. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
51. Lee KA. Self-reported sleep disturbances in employed women. Sleep 1992 Dec;15(6):493-498. [doi: 10.1093/sleep/15.6.493] [Medline: 1475563]
52. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
53. Kwon M, Nam K, Seo B. The survey on internet overdependence. Ministry of Science, ICT and Future Planning, National Information Society Agency. 2016. URL: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=17132 [accessed 2024-04-29]
54. Jo HS, Na E, Kim DJ. The relationship between smartphone addiction predisposition and impulsivity among Korean smartphone users. Addict Res Theory 2017 Apr 23;26(1):77-84. [doi: 10.1080/16066359.2017.1312356]
55. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017 Jan 01;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
56. Chen XY, Shi X, Zhou Y, Chen H, Ma Y, Wang T, et al. Change patterns of sleep problems predict mental health problems among adolescents: a 10-year cohort study of Chinese Wenchuan earthquake. J Affect Disord 2021 May 15;287:138-144. [doi: 10.1016/j.jad.2021.02.080] [Medline: 33780829]
57. McNicol ML, Thorsteinsson EB. Internet addiction, psychological distress, and coping responses among adolescents and adults. Cyberpsychol Behav Soc Netw 2017 May;20(5):296-304 [FREE Full text] [doi: 10.1089/cyber.2016.0669] [Medline: 28414517]
58. Long J, Liu TQ, Liao YH, Qi C, He H, Chen S, et al. Prevalence and correlates of problematic smartphone use in a large random sample of Chinese undergraduates. BMC Psychiatry 2016 Nov 17;16(1):408 [FREE Full text] [doi: 10.1186/s12888-016-1083-3] [Medline: 27855666]
59. Yang X, Hu H, Zhao C, Xu H, Tu X, Zhang G. A longitudinal study of changes in smart phone addiction and depressive symptoms and potential risk factors among Chinese college students. BMC Psychiatry 2021 May 14;21(1):252 [FREE Full text] [doi: 10.1186/s12888-021-03265-4] [Medline: 33990181]
60. Elhai JD, Levine JC, Hall BJ. The relationship between anxiety symptom severity and problematic smartphone use: a review of the literature and conceptual frameworks. J Anxiety Disord 2019 Mar;62:45-52. [doi: 10.1016/j.janxdis.2018.11.005] [Medline: 30529799]
61. Sohn SY, Krasnoff L, Rees P, Kalk NJ, Carter B. The association between smartphone addiction and sleep: a UK cross-sectional study of young adults. Front Psychiatry 2021;12:629407 [FREE Full text] [doi: 10.3389/fpsyt.2021.629407] [Medline: 33737890]
62. Pérez de Albéniz Garrote G, Rubio L, Medina Gómez B, Buedo-Guirado C. Smartphone abuse amongst adolescents: the role of impulsivity and sensation seeking. Front Psychol 2021;12:746626 [FREE Full text] [doi: 10.3389/fpsyg.2021.746626] [Medline: 34975636]
63. Schulz van Endert T, Mohr PN. Likes and impulsivity: investigating the relationship between actual smartphone use and delay discounting. PLoS One 2020;15(11):e0241383 [FREE Full text] [doi: 10.1371/journal.pone.0241383] [Medline: 33206673]
64. Siła-Nowicka K, Vandrol J, Oshan T, Long JA, Demšar U, Fotheringham AS. Analysis of human mobility patterns from GPS trajectories and contextual information. Int J Geogr Inf Sci 2015 Oct 30;30(5):881-906. [doi: 10.1080/13658816.2015.1100731]
65. Searle KM, Lubinda J, Hamapumbu H, Shields TM, Curriero FC, Smith DL, et al. Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. R Soc Open Sci 2017 May;4(5):170046 [FREE Full text] [doi: 10.1098/rsos.170046] [Medline: 28573009]
66. Choi A. The effect of university students' COVID-19 stress on smartphone addiction: the mediating effects of self-regulation. Fam Environ Res 2021 Nov 22;59(4):557-568. [doi: 10.6115/fer.2021.039]
67. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: towards a model of compensatory internet use. Comput Human Behav 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
68. Wang A, Wang Z, Zhu Y, Shi X. The prevalence and psychosocial factors of problematic smartphone use among Chinese college students: a three-wave longitudinal study. Front Psychol 2022;13:877277 [FREE Full text] [doi: 10.3389/fpsyg.2022.877277] [Medline: 35450331]

XSL•FO
RenderX

69. Beidel DC, Turner S. Shy Children, Phobic Adults: Nature and Treatment of Social Anxiety Disorder. 2nd edition. Washington, DC: American Psychological Association; 2007.
70. Wong N, Sarver DE, Beidel DC. Quality of life impairments among adults with social phobia: the impact of subtype. J Anxiety Disord 2012 Jan;26(1):50-57 [FREE Full text] [doi: 10.1016/j.janxdis.2011.08.012] [Medline: 21964285]
71. Lipton MF, Weeks JW, Daruwala SE, de los Reyes A. Profiles of social anxiety and impulsivity among college students: a close examination of profile differences in externalizing behavior. J Psychopathol Behav Assess 2016 Jan 9;38(3):465-475. [doi: 10.1007/S10862-015-9531-9]
72. Beidel DC, Rao PA, Scharfstein L, Wong N, Alfano CA. Social skills and social phobia: an investigation of DSM-IV subtypes. Behav Res Ther 2010 Oct;48(10):992-1001 [FREE Full text] [doi: 10.1016/j.brat.2010.06.005] [Medline: 20637452]
73. Müller SR, Chen XL, Peters H, Chaintreau A, Matz SC. Depression predictions from GPS-based mobility do not generalize well to large demographically heterogeneous samples. Sci Rep 2021 Jul 07;11(1):14007 [FREE Full text] [doi: 10.1038/s41598-021-93087-x] [Medline: 34234186]
74. Lim H, Kim S, Chung K, Lee K, Kim T, Heo J. Is college students' trajectory associated with academic performance? Comput Educ 2022 Mar;178:104397. [doi: 10.1016/j.compedu.2021.104397]
75. Keller J, Roitzheim C, Radtke T, Schenkel K, Schwarzer R. A mobile intervention for self-efficacious and goal-directed smartphone use in the general population: randomized controlled trial. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e26397 [FREE Full text] [doi: 10.2196/26397] [Medline: 34817388]
76. Wen H, Sobolev M, Vitale R, Kizer J, Pollak JP, Muench F, et al. mPulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study. JMIR Ment Health 2021 Jan 27;8(1):e25019 [FREE Full text] [doi: 10.2196/25019] [Medline: 33502330]
77. Lauvsnes AD, Hansen TI, Ankill S, Bae SW, Gråwe RW, Braund TA, et al. Mobile assessments of mood, cognition, smartphone-based sensor activity, and variability in craving and substance use in patients with substance use disorders in Norway: prospective observational feasibility study. JMIR Form Res 2023 Jun 23;7:e45254 [FREE Full text] [doi: 10.2196/45254] [Medline: 37351934]
78. Ryding FC, Kuss DJ. Passive objective measures in the assessment of problematic smartphone use: a systematic review. Addict Behav Rep 2020 Jun;11:100257 [FREE Full text] [doi: 10.1016/j.abrep.2020.100257] [Medline: 32467846]
79. Hsu M, Ahern DK, Suzuki J. Digital phenotyping to enhance substance use treatment during the COVID-19 pandemic. JMIR Ment Health 2020 Oct 26;7(10):e21814 [FREE Full text] [doi: 10.2196/21814] [Medline: 33031044]
80. Ai P, Liu Y, Zhao X. Big Five personality traits predict daily spatial behavior: evidence from smartphone data. Pers Individ Dif 2019 Sep;147:285-291. [doi: 10.1016/j.paid.2019.04.027]
81. Müller SR, Peters H, Matz SC, Wang W, Harari GM. Investigating the relationships between mobility behaviours and indicators of subjective well–being using smartphone–based experience sampling and GPS tracking. Eur J Pers 2020 Sep 01;34(5):714-732. [doi: 10.1002/per.2262]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
DASS: Depression Anxiety Stress Scales
EMA: ecological momentary assessment
GAD-7: Generalized Anxiety Disorder-7
GSDS: General Sleep Disturbance Scale
K-SAPS: Korean Smartphone Addiction Proneness Scale
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
PSU: problematic smartphone use
SIA: social interaction anxiety
SIAS: Social Interaction Anxiety Scale

Edited by A Coristine; submitted 27.11.24; peer-reviewed by P Chidipudi, K Lee, S Olalere; comments to author 18.12.24; revised version received 09.01.25; accepted 20.01.25; published 05.02.25.

Please cite as:
Ahn JS, Jeong I, Park S, Lee J, Jeon M, Lee S, Do G, Jung D, Park JY
App-Based Ecological Momentary Assessment of Problematic Smartphone Use During Examination Weeks in University Students: 6-Week Observational Study
J Med Internet Res 2025;27:e69320
URL: https://www.jmir.org/2025/1/e69320
doi:10.2196/69320
PMID:

XSL•FO
RenderX

©Ji Seon Ahn, InJi Jeong, Sehwan Park, Jooho Lee, Minjeong Jeon, Sangil Lee, Gangho Do, Dooyoung Jung, Jin Young Park. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Technological-Based Interventions in Cancer and Factors Associated With the Use of Mobile Digital Wellness and Health Apps Among Cancer Information Seekers: Cross-Sectional Study

Ogochukwu Juliet Ezeigwe[1], MPH; Kenechukwu Obumneme Samuel Nwosu[1], MBBS, MPH; Oladipo Kunle Afolayan[2], MSc; Akpevwe Amanda Ojaruega[1], MPH; Jovita Echere[1], MD, MPH; Manali Desai[1], MBBS, MPH; Modupe Olajumoke Onigbogi[1], MBBS, MPH; Olajumoke Ope Oladoyin[3], MPH; Nnenna Chioma Okoye[4], PharmB; Pierre Fwelo[1], MPH, PhD

[1]Department of Epidemiology, The University of Texas Health Science Center, Houston, TX, United States

[2]Department of Biostatistics, The University of Texas Health Science Center, Houston, TX, United States

[3]Department of Health Promotion and Behavioral Sciences, The University of Texas Health Science Center, Houston, TX, United States

[4]Department of Pharmaceuticals, Gwarinpa General Hospital, Abuja, Nigeria

Corresponding Author:
Ogochukwu Juliet Ezeigwe, MPH
Department of Epidemiology
The University of Texas Health Science Center
7000 Fannin Street
Houston, TX, 77030
United States
Phone: 1 4042005302
Email: Ogochukwu.j.ezeigwe@uth.tmc.edu

Abstract

Background: Mobile digital wellness and health apps play a significant role in optimizing health and aiding in cancer management and decision-making.

Objective: This study aims to identify the factors influencing the use of mobile health and wellness apps among cancer information seekers in the United States.

Methods: We conducted a cross-sectional study using data from the Health Information National Trends Survey. Our analysis focused on 4770 participants who sought cancer information. We performed weighted univariate and multivariable logistic regression to determine the association between the use of health and wellness apps and socioeconomic factors, medical history and conditions, and lifestyle and behavioral factors.

Results: A total of 4770 participants who sought cancer information were included in the final analysis. Of these, 80.9% (n=2705) were health and wellness app users, while 19.1% (n=793) were nonusers. In the final adjusted model, participants with household incomes ≥US $50,000 had 49% higher adjusted odds of using these apps than those with incomes <US $50,000 (adjusted odds ratio [aOR]=1.49, 95% CI 1.02-2.14). College graduates and those with higher educational levels were avid users compared to those with a high school diploma or less (aOR=1.87, 95% CI 1.30-2.67). Internet users had over 3 times the odds of using these apps compared to nonusers (aOR=3.28, 95% CI 1.70-6.33). Participants within the age group 18-34 years were 3.70 times more likely (aOR=3.70, 95% CI 1.90-7.23) to use a health and wellness app compared to participants within the age group of 75 years and older.

Conclusions: Age, education, household income, and use of the internet are the major determinants of the adoption of digital health and wellness apps among seekers of cancer information. Hence, public health programs could be directed toward addressing these factors to improve cancer diagnosis, treatment, and management using these apps.

(J Med Internet Res 2025;27:e63403) doi:10.2196/63403

KEYWORDS

cancer intervention; digital health and wellness apps; cancer management; telehealth; mobile health; mhealth; decision-making; United States; cross-sectional study; adult; logistic regression; regression analysis; digital health; young adult; cancer screening;

XSL•FO
RenderX

knowledge seeking; barrier; utilization; engagement; digital health engagement; diversity; cancer information; health seeking behavior; mobile phone

Introduction

Cancer is a major health problem worldwide and the second leading cause of death in the United States [1]. According to the World Health Organization (WHO), there were about 10 million deaths from cancer in 2020 [2]. According to the National Cancer Institute, in January 2022, about 18.1 million survivors of cancer were estimated in the United States, and by 2032, the number of survivors of cancer would increase to 22.5 million [3]. In the United States, the increasing cancer survival rates can be largely attributed to advances in screening, early diagnoses, and treatment of cancers, as well as a growing and aging US population [4]. Many survivors of cancer tend to seek health information in addition to the information provided by their physicians [5]. A study reported that survivors of cancer seek more information about good diet, exercise, and weight management while undergoing treatment [6]. Nearly 50% of Americans and over 60% of survivors of cancer seek cancer-related information from at least 1 source, including mobile wellness and health apps [5,7-9].

According to WHO, mobile health (mHealth) includes using smartphones, sensors, PDAs, wireless monitoring devices, or other wireless devices for public health and medical practices [10]. The Center for Democracy and Technology categorizes health apps into 4 types: health reference, fitness tracker, diagnostic, and disease management [11]. Mobile health apps are software programs running on smartphones and tablets to promote health and primary disease prevention [12,13]. They are used to oversee, improve, and maintain the health of their users at individual and community levels [14]. Furthermore, these health apps are very useful in facilitating medication adherence, monitoring symptoms, clinical decision-making, and behavioral changes [15-17]. Among patients with cancer, mobile digital health is an important consideration when seeking ways to optimize their mental health [18].

A study has shown that mHealth apps can help in primary prevention, such as screening, as well as early diagnosis, management, survivorship, and end-of-life care among patients with cancer [19]. Another study reported promising use of digital health solutions for promoting and managing the cancer care continuum within a patient-centeredness framework [20]. Furthermore, research has shown that mHealth apps can aid health care providers and patients in cancer diagnosis, managing psychological distress, facilitating follow-up care, devising treatment plans, delivering cancer-related information, promoting drug adherence, and addressing side effects [21].

Socioeconomic variables such as education level and income, marital status, gender, and age, are important predictors of choice of health information source among survivors of cancer in the United States [9]. A study that assessed the disparities in access to mobile health devices and eHealth literacy among survivors of breast cancer found that older age, lack of access to mobile devices, and a lower education level had a lesser association with eHealth literacy [22]. This study also showed that younger women with higher education levels and from less deprived areas were more likely to access smartphones and tablets [22]. Another study also found that mHealth users were more likely to be younger, have higher education, reported excellent health, higher income, and intention to change diet and physical activity [23]. A recent review suggested the positive effect of mHealth apps on health outcomes among those enduring chronic diseases [24]. Furthermore, research revealed that older adults exhibited diminished self-efficacy when using mHealth apps often stemming from a deficiency in technical skills and resulting in a decreased inclination to engage with the technology [25]. An analysis of the moderating effect of different age groups suggests that the perceived ease of use and vulnerability were associated with the use of mHealth apps among middle-aged and elderly people [26].

Despite the growth and promise of using mobile apps to deliver information and interventions to patients with cancer and those with other chronic diseases, the factors influencing the use of mHealth apps have been well-studied primarily among survivors of cancer. However, there is limited research on these factors among the broader US population seeking cancer information. Therefore, the purpose of this study is to identify the factors that impact the usage of health and wellness apps among those seeking cancer information here in the United States. By using a sequential modeling approach, this study aims to provide a deeper understanding of the factors associated with mobile app usage in this specific population, ultimately informing targeted interventions and strategies to effectively support cancer information seekers, both survivors and nonsurvivors.

Methods

Data Source, Study Design, and Setting

This cross-sectional study analyzed data from the Health Information National Trends Survey (HINTS), a nationwide representative survey of US adults aged 18 years and older in 2020 and 2022. HINTS has been conducted periodically since 2003. HINTS collects information on access to and usage of health-related information, health-related behaviors such as perceptions, knowledge of disease and cancer screening as well as telehealth among US adults. The HINTS uses a complex sampling design to ensure the representativeness of the adult population in the United States. The survey uses a sampling frame provided by Marketing Systems Group of addresses in the United States. To enhance the response rate and ensure the sample's representativeness, HINTS conducts several follow-up mailings for nonrespondents. Detailed information about the methodology, sampling, and weighting is available on the web [27]. This study is reported following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [28].

Study Population

This study included adults aged 18 and older in the United States. Using consistent weighting algorithms, we merged data from HINTS 5 (Cycle 4, 2020) and HINTS 6 (2022) to increase

XSL•FO
RenderX

the precision of estimates and power of this study. This resulted in a total population of 10,117 respondents from the 2 surveys. Individuals' cancer-seeking behavior was assessed through a single question: Have you ever looked for information about cancer from any source? The response options were "yes" or "no." Those who answered "yes" were classified as having sought cancer-related information. Hence, this study included 4770 participants who actively sought cancer-related information in the analysis.

Study Variables

Outcome

The outcome variable was using health and wellness apps categorized as "yes" or "no." Two sequential survey questions were used depending on the outcome. Participants were asked if they owned tablets, computers, or smartphones. Those who owned any were further asked if they had health and wellness apps on these devices, with response options of "yes," "no," or "do not know." If they answered "yes," they were then asked if they had used these apps in the past 12 months.

Independent Variables

Based on previous literature, the sociodemographic characteristics: age (18-34, 35-49, 50-64, 65-74, or 75 years and older), gender at birth (male or female), insurance (yes or no), educational level (high school or less, some college, or college graduates and more), household income (<US $50,000 or ≥US $50,000), use of the internet (yes or no) and race or ethnicity (Hispanic, non-Hispanic Asian or non-Hispanic other, non-Hispanic Black, or non-Hispanic White); medical history and disease conditions: presence or absence of cancer (yes or no), general health status (fair or poor, good, or excellent and very good), family history of cancer (yes or no), depression (yes or no), diabetes (yes or no), hypertension (yes or no), heart condition (yes or no), chronic lung disease (yes or no), and number of disease condition (none, one, or two or more), and lifestyle or behavioral characteristics; BMI (underweight, normal weight, overweight, or obese) and physical activity were included in this study. Physical activity was assessed by the number of days per week and duration of moderate-intensity exercise, classified based on the WHO's recommendation of 150 minutes/week [9,22-27].

Statistical Analysis

We used weighted frequencies and percentages to present participants' sociodemographic characteristics, medical history, disease conditions, lifestyle, and behavioral traits by mobile health app usage status. We used the Pearson chi-square test to assess the statistical significance (P<.05) of the relationship between mobile health and wellness app usage and independent variables. We conducted bivariate and multivariable logistic regression analyses to evaluate the association between the outcome and covariates.

We fitted 4 sequential modeling approaches in the multivariable analysis. Model 0 included the crude effects of each covariate on health and wellness app usage. Model 1 incorporated the effects of sociodemographic characteristics on health and wellness app usage. Model 2 adjusted for all covariates from model 1 and medical history and disease conditions, such as cancer, general health status, family history of cancer, depression, diabetes, hypertension, heart conditions, chronic lung disease, and the number of chronic diseases. The final (model 3) included model 2 and lifestyle and behavioral characteristics, such as BMI and physical activities.

This sequential modeling strategy aimed at determining the degree to which mHealth app use was explained by each group of variables among seekers of cancer information. We further explored the relative importance of medical history and disease conditions on sociodemographic factors and how medical disease conditions are related to the use of mobile apps (model 2). We also looked at the specific importance behavior and lifestyle have in the final model (model 3). For this study, we chose to adjust for physical activity and BMI because these variables are potential factors for the usage of wearable devices [29,30].

All statistical analyses were performed using SAS software (version 9.4; SAS Institute), with significance set at a P value <.05 and a 95% CI.

Ethical Considerations

The administration of HINTS received approval from the institutional review board at Westat Inc and was designated as exempt by the National Institutes of Health Office of Human Subjects Research (45 CFR 46.104 and Project # 6632.03.51) [27]. This exemption applies to this study. HINTS data are accessible to the public, and additional information regarding the survey methodology is available on the HINTS website. This study's participants were deidentified and there was no patient contact.

Results

A total of 4770 participants were seeking cancer information. Only 3498 participants had complete information on the use of mobile apps. Of these, 80.9% (n=2705) were health and wellness app users, while 19.1% (n=793) were nonusers. Overall, more than half (n=2894, 58.4%) of the participants were female, and 78% were between the age group 18-64 years. A larger percentage of participants were college graduates (n=2606, 41.5%), identified as non-Hispanic White (n=2936, 69.3%), had a household income exceeding US $50,000 (n=2854, 69.9%), and rated their general health as excellent or good (n=2256, 49.1%). Most participants had no history of cancer (n=3538, 84.7%), had insurance (n=4453, 92.1%), were internet users (n=4376, 93.3%), had heart conditions (n=460, 92.4%), and lung diseases (n=3961, 86.7%; Table 1).

XSL•FO
RenderX

Table 1. Distribution of participants characteristics by the use of wellness and health apps (N=4770).

Characteristic	Seekers of cancer information (N=4770), n[a] (Wt%)[b]	Used health and wellness app (n=3498)		
		Yes, n[a] (Wt%)[b]	No, n[a] (Wt%)[b]	*P* value
Used health and wellness apps				
Yes	2705 (80.9)	—[c]	—	—
No	793 (19.1)	—	—	—
Sociodemographic and economic characteristics				
Age (years)				<.001
18-34	643 (21.8)	491 (27)	73 (13.5)	
35-49	931 (27.1)	658 (28.7)	111 (22.9)	
50-64	1429 (29.1)	810 (29.4)	251 (34.9)	
65-75	1110 (13.9)	512 (10.6)	237 (19.4)	
75+	575 (7.2)	207 (4.3)	107 (9.4)	
Gender at birth				.18
Female	2894 (58.4)	1693 (59.9)	462 (55.8)	
Male	1683 (41.6)	893 (40.1)	299 (44.2)	
Insurance				.82
Yes	4453 (92.1)	2548 (92.9)	724 (92.6)	
No	239 (7.9)	118 (7.04)	54 (7.4)	
Education				<.001
High school or less	1015 (27.3)	410 (21.8)	219 (33.3)	
Some college	950 (31.2)	510 (30)	178 (34.2)	
College graduates and more	2606 (41.5)	1666 (48.2)	367 (32.6)	
Race or ethnicity				.43
Hispanic	597 (12.9)	324 (12.6)	110 (11.4)	
Non-Hispanic Asian or non-Hispanic other	332 (8.9)	201 (9.6)	59 (7.1)	
Non-Hispanic Black	517 (8.9)	296 (8.3)	86 (8.3)	
Non-Hispanic White	2936 (69.3)	1700 (69.6)	467 (73.3)	
Household income (US $)				.001
<50,000	1456 (30.1)	628 (22.8)	273 (33.6)	
≥50,000	2854 (69.9)	1853 (77.2)	437 (66.5)	
Use of internet				<.001
Yes	4376 (93.3)	2632 (97.9)	713 (91.1)	
No	392 (6.7)	73 (2.1)	80 (8.9)	
Medical history and disease condition				
Ever had cancer				<.001
Yes	1070 (15.3)	500 (12.7)	208 (20.8)	
No	3538 (84.7)	2098 (87.3)	559 (79.2)	
General health status				.02
Fair or poor	713 (14.3)	307 (10.8)	141 (16.3)	
Good	1702 (36.6)	960 (37.1)	295 (35.3)	
Excellent and very good	2256 (49.1)	1378 (52.2)	340 (48.4)	
Family history of cancer				.53

XSL•FO
RenderX

Characteristic	Seekers of cancer information (N=4770), n[a] (Wt%)[b]	Used health and wellness app (n=3498)		
		Yes, n[a] (Wt%)[b]	No, n[a] (Wt%)[b]	*P* value
Yes	3648 (83.2)	2093 (82.7)	593 (84.4)	
No	647 (16.8)	366 (17.3)	105 (15.6)	
Depression or anxiety				.40
No	3337 (29.3)	1833 (69.2)	587 (72.1)	
Yes	1322 (70.7)	809 (30.8)	186 (27.9)	
Diabetes				.049
Yes	908 (16.1)	458 (13.8)	160 (18.3)	
No	3744 (83.9)	2181 (86.2)	612 (81.7)	
Hypertension				.004
Yes	2034 (35.1)	1047 (32.4)	375 (41.2)	
No	2621 (64.9)	1594 (67.6)	399 (58.8)	
Heart condition				.77
Yes	460 (7.6)	220 (6.8)	73 (6.4)	
No	4198 (92.4)	2422 (93.2)	701 (93.6)	
Chronic lung disease				.53
Yes	694 (13.3)	394 (13.7)	115 (12.5)	
No	3961 (86.7)	2247 (86.3)	658 (87.5)	
Number of disease conditions				.10
None	2031 (51.7)	1227 (53.8)	326 (48.5)	
One	1499 (30.2)	866 (30.1)	236 (30.6)	
Two or more	1096 (18.1)	534 (16.1)	208 (20.9)	
Lifestyle and behaviors				
BMI				.85
Underweight	234 (4.8)	116 (4.5)	36 (3.8)	
Normal weight	1450 (31.8)	827 (33.1)	252 (31.7)	
Overweight	1549 (31.3)	869 (30.3)	249 (32.8)	
Obese	1537 (32.1)	893 (32.1)	256 (31.7)	
Physical activity				.002
<150 min/wk	2948 (61.4)	1566 (57.7)	526 (70.3)	
≥150 min/wk	1822 (38.6)	1139 (42.3)	267 (29.7)	

[a]Unweighted number of participants.

[b]Weighted percentages.

[c]Not applicable.

From the bivariate analysis, age, level of education, household income, use of the internet, history of cancer, reported health status, physical activity, hypertension, and diabetes status showed a significant association with the use of health and wellness apps (*P*<.05). More than two-thirds (85.1%) of participants within the age group 18-64 years used health and wellness apps compared to (14.9%) of participants within the age group 65+ years. A significant proportion of college graduates (n=1666, 48.2%) and individuals with household incomes of US $50,000 or more (n=1853, 77.2%) were found to use health and wellness apps. Among participants who use health and wellness apps, 87.3% (n=2098) had no history of cancer, 97.9% (2632) were internet users, 52.2% (n=1378) reported their general health status as excellent or very good, and 67.6% (n=1594) were hypertensive.

From the multivariable results in Table 2 and the final model (model 3), factors associated with the use of health and wellness apps included age, household income, use of the internet, higher educational level, and physical activity. We observed that as participant's age increases, the odds of using health and wellness

XSL•FO
RenderX

apps decrease. Participants within the age group 18-34 years were 4 times more likely (adjusted odds ratio [aOR]=3.70, 95% CI 1.90-7.23) to use a health and wellness app compared to participants within the age group of 75+ years. Compared to participants with household income below US $50,000, participants in the higher income category ≥US $50,000 had 49% higher odds of using mobile digital health and wellness apps (aOR=1.49, 95% CI 1.02-2.14).

Furthermore, consistent internet use significantly influenced the use of health and wellness apps across all 3 models. Internet users had 28% higher odds (aOR=3.28, 95% CI 1.70-6.33) of using mobile digital health and wellness apps compared to noninternet users. As the level of education increases, the odds of using health and wellness apps also increase. The participants with college graduate degrees and higher had 1.87 times the odds (95% CI 1.30-2.67) of using health or wellness apps compared to individuals with a high school diploma or less. Physical activity was the only lifestyle and behavioral factor found to be associated with the use of health and wellness apps. Compared to participants who engaged in less than 150 minutes of physical activity per week, those who exercised 150 minutes or more per week had 1.94 times higher adjusted odds (95% CI 1.40-2.70) of using health and wellness apps (Table 2).

XSL•FO
RenderX

Table 2. Bivariable and multivariable logistic regression analyses of the association between participant characteristics and the use of wellness and health apps[a].

Characteristics	Used health and wellness app			
	Model 0[b]	Model 1[c]	Model 2[d]	Model 3[e]
	Crude OR[f] (95% CI)	Adjusted OR (aOR; 95% CI)	Adjusted OR (aOR; 95% CI)	Adjusted OR (aOR; 95% CI)
Age (years)				
18-34	*4.43 (2.75-7.14)*	*3.48 (1.97-6.16)*	*3.68 (1.9-7.1)*	*3.7 (1.9-7.23)*
35-49	*2.77 (1.72-4.46)*	*2.27 (1.31-3.94)*	*2.4 (1.29-4.45)*	*2.26 (1.22-4.2)*
50-64	*1.86 (1.21-2.87)*	1.49 (0.91-2.44)	1.68 (0.97-2.91)	1.63 (0.95-2.8)
65-75	1.21 (0.79-1.86)	1.12 (0.65-1.92)	1.18 (0.66-2.13)	1.07 (0.61-1.89)
75+	1	1	1	1
Gender at birth				
Female	1	1	1	1
Male	0.85 (0.66-1.08)	0.82 (0.63-1.07)	0.81 (0.62-1.07)	0.78 (0.6-1.03)
Insurance				
No	1	1	1	1
Yes	1.06 (0.64-1.74)	1.34 (0.72-2.49)	1.24 (0.62-2.47)	1.19 (0.58-2.46)
Education				
High school or less	1	1	1	1
Some college	1.34 (0.91-1.99)	1.2 (0.78-1.84)	1.23 (0.79-1.9)	1.19 (0.76-1.85)
College graduates and more	*2.26 (1.66-3.06)*	*1.84 (1.28-2.64)*	*1.95 (1.35-2.82)*	*1.87 (1.3-2.67)*
Household income (US $)				
< $50,000	1	1	1	1
≥$50,000	*1.71 (1.23-2.38)*	*1.47 (1.04-2.09)*	*1.49 (1.02-2.16)*	*1.49 (1.02-2.14)*
Use of internet				
No	1	1	1	1
Yes	*4.53 (2.63-7.79)*	*3.73 (1.97-7.08)*	*3.48 (1.75-6.93)*	*3.28 (1.70-6.33)*
Race or ethnicity				
Hispanic	1.16 (0.81-1.68)	1.32 (0.86-2.02)	1.41 (0.88-2.24)	1.4 (0.86-2.25)
Non-Hispanic Asian or non-Hispanic other	1.43 (0.9-2.27)	1.16 (0.73-1.86)	1.16 (0.7-1.93)	1.13 (0.67-1.89)
Non-Hispanic Black	1.05 (0.66-1.65)	1.3 (0.78-2.15)	1.2 (0.69-2.07)	1.18 (0.68-2.03)
Non-Hispanic White	1	1	1	1
General health status				
Fair or poor	1	N/A[h]	1	1
Good	*1.59 (1.1-2.30)*	N/A	1.28 (0.82-1.99)	1.3 (0.82-2.05)
Excellent and very good	*1.63 (1.15-2.32)*	N/A	1.23 (0.76-2)	1.2 (0.72-2.02)
Ever had cancer				
No	1	N/A	1	1
Yes	*0.55 (0.41-0.75)*	N/A	0.81 (0.57-1.15)	0.87 (0.62-1.23)
Family history of cancer				
No	1	N/A	1	1
Yes	0.88 (0.6-1.3)	N/A	0.9 (0.58-1.39)	0.86 (0.57-1.32)

XSL•FO
RenderX

Characteristics	Used health and wellness app			
	Model 0[b]	Model 1[c]	Model 2[d]	Model 3[e]
	Crude OR[f] (95% CI)	Adjusted OR (aOR; 95% CI)	Adjusted OR (aOR; 95% CI)	Adjusted OR (aOR; 95% CI)
Depression or anxiety				
No	1	N/A	1	1
Yes	1.15 (0.83-1.59)	N/A	0.92 (0.61-1.39)	0.93 (0.62-1.4)
Diabetes				
No	1	N/A	1	1
Yes	0.71 (0.51-1.01)	N/A	0.91 (0.4-2.12)	0.87 (0.37-2.07)
Hypertension				
No	1	N/A	1	1
Yes	*0.68 (0.53-0.89)*	N/A	0.77 (0.36-1.66)	0.8 (0.38-1.69)
Heart condition				
No	1	N/A	1	1
Yes	1.07 (0.69-1.67)	N/A	1.67 (0.94-2.95)	1.78 (1-3.17)
Chronic lung disease				
No	*1*	N/A	1	1
Yes	1.12 (0.79-1.57)	N/A	1.08 (0.55-2.11)	1.11 (0.57-2.17)
Number of disease conditions				
None	1	N/A	1	1
One	0.89 (0.65-1.22)	N/A	1.55 (0.74-3.26)	1.49 (0.71-3.13)
Two or more	*0.69 (0.52-0.94)*	N/A	1.46 (0.39-5.54)	1.38 (0.37-5.1)
BMI				
Underweight	1	N/A	N/A	1
Normal weight	0.89 (0.51-1.56)	N/A	N/A	0.65 (0.21-1.99)
Overweight	0.79 (0.44-1.43)	N/A	N/A	0.69 (0.21-2.22)
Obese	0.87 (0.47-1.61)	N/A	N/A	0.9 (0.27-3.05)
Physical activity				
<150 min/wk	1	N/A	N/A	1
≥150 min/wk	*1.73 (1.29-2.33)*	N/A	N/A	*1.94 (1.4-2.7)*

[a]Italicized values are significant at $P<.05$.

[b]Model 0: univariate analysis.

[c]Model 1: sociodemographic factors.

[d]Model 2: sociodemographic factors + medical history and disease condition.

[e]Model 3: sociodemographic factors + medical history and disease condition + lifestyle and behavioral factors.

[f]OR: odds ratio.

[g]N/A: not applicable.

Discussion

Principal Findings

This study aimed to investigate the use of digital health and wellness apps among seekers of cancer information and identify factors predicting their use. Our results indicate a high prevalence of app usage among individuals seeking cancer information (81%), highlighting the potential of public health interventions to promote cancer prevention measures (for example, screening) and treatment through these platforms. Age and education emerged as the most significant predictors, with younger and more educated individuals showing a greater inclination toward the usage of these apps.

This finding aligns with similar studies [23,31], including one carried out among the Dutch population, which showed that younger age groups and individuals with higher education are more likely to use mHealth apps [32]. The reason may be that older patients generally describe themselves as not highly skilled

XSL•FO
RenderX

in the use of mobile phones and other mobile devices [33]. In addition, older adults may face barriers such as a lack of trust in the apps, concerns about data privacy, and fear of misdiagnosis [34]. Moreover, other factors, such as age-related reduction in health functions such as memory, vision, and touch sensitivity, may hinder the effective use of mobile health apps [35]. The age disparity in mobile app use implies that older adults may experience worsened health outcomes and limited access to cancer health information due to lower engagement with digital tools. To address this, mobile apps can be designed to be more user-friendly for older adults, and digital health education programs should be implemented to encourage their adoption. Consistent with the Pew survey, which found that women were more likely to use health apps, gender differences were also observed in this study, with females being more likely to use health and wellness apps [36].

In a previous study, higher incomes were correlated with digital technology ownership and usage [37]. Our study found a similar association in which participants with household incomes greater than or equal to US $50,000 were more likely to use digital and wellness apps than participants with less than US $50,000. This suggests that income may strongly influence the use of digital and wellness apps even when other factors are considered. The finding that seekers of cancer information who met weekly recommendations for physical activity were more likely to use mHealth apps compared to those who engaged in less physical activity is consistent with previous studies [23,38]. This implies using health apps may enhance the achievement of physical activity health goals which has the potential to prevent chronic diseases [39].

Engaging in health information-seeking behavior, such as downloading mHealth apps, is acknowledged as a crucial activity during the “preparation stage” that may lead to changes in health behavior [40]. Use of the internet was found to be consistently associated with the usage of health and wellness apps among seekers of cancer information in our study. A similar study found that most internet users indicated a greater likelihood of using at least one eHealth tool to address a health issue over 12 months with a preference for YouTube videos, a peer-to-peer support website, or a smartphone app [41]. This implies that using the internet has the potential to empower individuals, enabling them to play a more active role in their health care and fostering changes in health information–seeking behavior by providing easy access to mHealth apps. Previous studies have shown that increased exposure to internet use and other web-based activities promotes the seeking of health information on the web [42]. The findings of our study confirm the impact of the internet on the use of health and wellness apps. Overall, our findings underscore the potential of mobile digital wellness and health apps in supporting cancer awareness and management.

Clinical Practice Points and Implication

Technological interventions in cancer, specifically those using mobile digital wellness and health apps hold great promise to enhance patient outcomes and improve quality of life. Understanding factors that influence use is vital for creating impactful interventions. Apps should be developed to meet some specific health requirements such as phases of cancer treatment, as personalized intervention content is more likely to engage users and encourage behavioral changes, including social support features that allow users to communicate with caregivers, health care professionals, or individuals with similar health challenges will also go a long way toward heightening app usage and serve as a useful tool for interventions. Social support networks within the app can help users feel more connected and encouraged to share their experiences. Language and cultural differences should be considered to guarantee applicability and accessibility to a wide range of people. Enhanced engagement and effectiveness can also be achieved by providing content in multiple languages and taking cultural norms into account.

Strengths

The major strength of our study was the comprehensive dataset we used. The samples used are fully representative of the noninstitutionalized US population, which enhances the generalizability of our results. Another strength of our study is its targeted population, specifically focusing on individuals who actively seek cancer information. This dataset was created by combining information from multiple survey cycles which boosted the sample size, thereby increasing the power and precision of our study. Additionally, the use of preexisting data minimized selection bias and allowed for proper comparisons with the existing literature. Finally, the consistency of our results with prior research further validates the reliability of our findings.

Limitations

Despite the strengths of our study, it has limitations that warrant due consideration, the dataset used for our study was harmonized data from different HINTS cycles which were obtained from self-administered questionnaires. This method of data collection may introduce information bias, emphasizing the importance of more rigorous study designs in establishing causal relationships. Prospective studies and clinical trials are recommended to further explore the factors influencing digital health and wellness app usage among the general population and seekers of cancer information, and their impact on health outcomes among diverse populations. Additionally, our study did not take into account health literacy, provider recommendations, or health motivation. These factors should be considered in future research on the use of health and wellness apps. Another limitation of our study is the lack of data on the specific types of cancer information sought by participants. While this did not impact our goal of exploring factors influencing the use of mobile health and wellness apps among those seeking cancer information, there is a need for the HINT’s researchers to include a variable that focuses on the type of cancer information for further exploration by future studies.

Conclusion

This study evaluated various factors that could impact the use of mobile health and wellness apps among cancer information seekers. Among the factors assessed; age, education, household income, use of the internet, history of cancer, and optimal health

XSL•FO
RenderX

status were significantly linked to app usage. The results from this study suggest that individuals who have cancer could benefit remarkably from health apps, however, they were least likely to use the apps when compared to other factors. These apps may have the potential to address health care challenges, reduce disparities, and empower patients to manage their health more efficiently. Interventions can be tailored to enhance app use and improve health outcomes as we used retrospectively collected data in this cross-sectional study.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to take this opportunity to thank the University of Texas School of Public Health for providing student emergency support for the article processing fee.

Data Availability

The data access requests are publicly available on the Health Information National Trends Survey (HINTS) website. The code generated and used during this study is available upon request.

Authors' Contributions

OJE handled the conceptualization, methodology, data curation, supervision, and formal analysis. OKA and AAO worked on the methodology, writing of the original draft, and review and editing of the writing. PF aided in the supervision and methodology. All authors assisted with the writing of the original draft and review and editing of the writing.

Conflicts of Interest

None declared.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023;73(1):17-48 [FREE Full text] [doi: 10.3322/caac.21763] [Medline: 36633525]
2. Cancer. World Health Organization. URL: https://www.who.int/news-room/facts-in-pictures/detail/cancer [accessed 2025-01-17]
3. Cancer statistics. National Cancer Institute. URL: https://www.cancer.gov/about-cancer/understanding/statistics [accessed 2025-01-17]
4. Weaver KE, Forsythe LP, Reeve BB, Alfano CM, Rodriguez JL, Sabatino SA, et al. Mental and physical health-related quality of life among U.S. cancer survivors: population estimates from the 2010 National Health Interview Survey. Cancer Epidemiol Biomarkers Prev 2012;21(11):2108-2117 [FREE Full text] [doi: 10.1158/1055-9965.EPI-12-0740] [Medline: 23112268]
5. Adjei Boakye E, Mohammed KA, Geneus CJ, Tobo BB, Wirth LS, Yang L, et al. Correlates of health information seeking between adults diagnosed with and without cancer. PLoS One 2018;13(5):e0196446 [FREE Full text] [doi: 10.1371/journal.pone.0196446] [Medline: 29746599]
6. James-Martin G, Koczwara B, Smith EL, Miller MD. Information needs of cancer patients and survivors regarding diet, exercise and weight management: a qualitative study. Eur J Cancer Care (Engl) 2014;23(3):340-348. [doi: 10.1111/ecc.12159] [Medline: 24299170]
7. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
8. Nagler RH, Gray SW, Romantan A, Kelly BJ, DeMichele A, Armstrong K, et al. Differences in information seeking among breast, prostate, and colorectal cancer patients: results from a population-based survey. Patient Educ Couns 2010;81 Suppl:S54-S62 [FREE Full text] [doi: 10.1016/j.pec.2010.09.010] [Medline: 20934297]
9. Jackson I, Osaghae I, Ananaba N, Etuk A, Jackson N, Chido-Amajuoyi OG. Sources of health information among U.S. cancer survivors: results from the Health Information National Trends Survey (HINTS). AIMS Public Health 2020;7(2):363-379 [FREE Full text] [doi: 10.3934/publichealth.2020031] [Medline: 32617363]
10. WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: second global survey on ehealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2025-01-17]
11. Understanding the potential for bias in mhealth apps. Center for Democracy & Technology. URL: https://cdt.org/wp-content/uploads/2018/09/2018-09-11-Healgorithms-Understanding-the-Potential-for-Bias-in-mHealth-Apps.pdf [accessed 2025-01-17]
12. Van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]

XSL•FO
RenderX

13. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016;16 Suppl 5(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
14. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR mHealth uHealth 2022;10(10):e37980 [FREE Full text] [doi: 10.2196/37980] [Medline: 36315221]
15. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
16. Burbank AJ, Lewis SD, Hewes M, Schellhase DE, Rettiganti M, Hall-Barrow J, et al. Mobile-based asthma action plans for adolescents. J Asthma 2015;52(6):583-586 [FREE Full text] [doi: 10.3109/02770903.2014.995307] [Medline: 25494553]
17. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? npj Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
18. Elkefi S, Trapani D, Ryan S. The role of digital health in supporting cancer patients' mental health and psychological well-being for a better quality of life: a systematic literature review. Int J Med Inform 2023;176:105065. [doi: 10.1016/j.ijmedinf.2023.105065] [Medline: 37224644]
19. Prochaska JJ, Coughlin SS, Lyons EJ. Social media and mobile technology for cancer prevention and treatment. Am Soc Clin Oncol Educ Book 2017;37:128-137 [FREE Full text] [doi: 10.1200/EDBK_173841] [Medline: 28561647]
20. Charalambous A. Utilizing the advances in digital health solutions to manage care in cancer patients. Asia Pac J Oncol Nurs 2019;6(3):234-237 [FREE Full text] [doi: 10.4103/apjon.apjon_72_18] [Medline: 31259218]
21. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Optimizing cancer care through mobile health. Support Care Cancer 2015;23(7):2183-2188. [doi: 10.1007/s00520-015-2627-7] [Medline: 25649121]
22. Moon Z, Zuchowski M, Moss-Morris R, Hunter M, Norton S, Hughes L. Disparities in access to mobile devices and e-health literacy among breast cancer survivors. Support Care Cancer 2022;30(1):117-126 [FREE Full text] [doi: 10.1007/s00520-021-06407-2] [Medline: 34236506]
23. Carroll J, Moorhead A, Bond R, LeBlanc W, Petrella R, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
24. Fan K, Zhao Y. Mobile health technology: a novel tool in chronic disease management. Intelligent Medicine 2022;2(1):41-47 [FREE Full text] [doi: 10.1016/j.imed.2021.06.003]
25. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
26. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inf Manage 2018;43:342-350 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2017.08.006]
27. Methodology reports. Health Information National Trends Survey (HINTS). URL: https://hints.cancer.gov/data/methodology-reports.aspx [accessed 2025-01-17]
28. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
29. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR mHealth uHealth 2017;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
30. Dallinga JM, Mennes M, Alpay L, Bijwaard H, Baart de la Faille-Deutekom M. App use, physical activity and healthy lifestyle: a cross sectional study. BMC Public Health 2015;15:833 [FREE Full text] [doi: 10.1186/s12889-015-2165-8] [Medline: 26316060]
31. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
32. Bol N, Helberger N, Weert J. Differences in mobile health app use: a source of new digital inequalities? The Information Society 2018;34(3):183-193 [FREE Full text] [doi: 10.1080/01972243.2018.1438550]
33. Navabi N, Ghaffari F, Jannat-Alipoor Z. Older adults' attitudes and barriers toward the use of mobile phones. Clin Interv Aging 2016;11:1371-1378 [FREE Full text] [doi: 10.2147/CIA.S112893] [Medline: 27757025]
34. Rasche P, Schäfer K, Theis S, Bröhl C, Wille M, Mertens A. Age-related usability investigation of an activity tracker. IJHFE 2016;4(3/4):187 [FREE Full text] [doi: 10.1504/ijhfe.2016.083506]
35. NA. Stroke--1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO task force on stroke and other cerebrovascular disorders. Stroke 1989;20(10):1407-1431. [doi: 10.1161/01.str.20.10.1407] [Medline: 2799873]
36. Smith A. U.S. Smartphone Use in 2015. Pew Research Center: Internet, Science & Tech. 2015. URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2025-01-17]

XSL•FO
RenderX

37. Rainie L. Asian-Americans and Technology. Pew Research Center. URL: http://www.pewinternet.org/2011/01/06/asian-americans-and-technology/ [accessed 2025-01-17]
38. Lustria MLA, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011;17(3):224-243 [FREE Full text] [doi: 10.1177/1460458211414843] [Medline: 21937464]
39. Ernsting C, Dombrowski S, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
40. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
41. Makowsky M, Jones C, Davachi S. Prevalence and predictors of health-related internet and digital device use in a sample of South Asian adults in Edmonton, Alberta, Canada: results from a 2014 community-based survey. JMIR Public Health Surveill 2021;7(1):e20671 [FREE Full text] [doi: 10.2196/20671] [Medline: 33416506]
42. Rice RE. Influences, usage, and outcomes of internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006;75(1):8-28 [FREE Full text] [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]

Abbreviations

aOR: adjusted odds ratio
HINTS: Health Information National Trends Survey
mHealth: mobile health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 18.06.24; peer-reviewed by S Tundealao, L Ajijola; comments to author 18.10.24; revised version received 30.10.24; accepted 13.12.24; published 05.02.25.

Please cite as:
Ezeigwe OJ, Nwosu KOS, Afolayan OK, Ojaruega AA, Echere J, Desai M, Onigbogi MO, Oladoyin OO, Okoye NC, Fwelo P Technological-Based Interventions in Cancer and Factors Associated With the Use of Mobile Digital Wellness and Health Apps Among Cancer Information Seekers: Cross-Sectional Study
J Med Internet Res 2025;27:e63403
URL: https://www.jmir.org/2025/1/e63403
doi:10.2196/63403
PMID:

©Ogochukwu Juliet Ezeigwe, Kenechukwu Obumneme Samuel Nwosu, Oladipo Kunle Afolayan, Akpevwe Amanda Ojaruega, Jovita Echere, Manali Desai, Modupe Olajumoke Onigbogi, Olajumoke Ope Oladoyin, Nnenna Chioma Okoye, Pierre Fwelo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Good-Quality mHealth Apps for Endometriosis Care: Systematic Search

Diksha Sirohi[1], BHMS, MBA, MPH; Cecilia HM Ng[2,3], BSc, MHIM, PhD; Niranjan Bidargaddi[4], BE, PhD; Helen Slater[5], BAppSc, MAppSc, PhD; Melissa Parker[6], BNurs, MNurs(Research); M Louise Hull[1], BSc, MBChB, PhD; Rebecca O'Hara[1], BBUS(Hons), P Grad Cert (Health Promotion), PhD

[1]Robinson Research Institute, Adelaide Medical School, University of Adelaide, North Adelaide, Australia

[2]School of Clinical Medicine, Medicine and Health, UNSW Medicine, Sydney, Australia

[3]Jean Hailes for Women's Health, Melbourne, Australia

[4]Digital Psychiatry & Personal Health Informatics Lab, College of Medicine and Public Health, Flinders University, Adelaide, Australia

[5]School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Australia

[6]Canberra Endometriosis Centre, Centenary Hospital for Women and Children, ACT Health, Canberra, Australia

Corresponding Author:
Diksha Sirohi, BHMS, MBA, MPH
Robinson Research Institute
Adelaide Medical School
University of Adelaide
Ground Floor, 55 King William Road
North Adelaide, 5006
Australia
Phone: 61 0883131527
Email: d.sirohi@uqconnect.edu.au

Abstract

Background: Mobile health (mHealth) apps are increasingly being used by community members to track symptoms and manage endometriosis. In addition, clinicians use mHealth apps for continued medical education and clinical decision-making and recommend good-quality apps to patients. However, poor-quality apps can spread misinformation or provide recommendations that are not evidence-based. Therefore, a critical evaluation is needed to assess and recommend good-quality endometriosis mHealth apps.

Objective: This study aimed to evaluate the quality and provide recommendations for good quality endometriosis mHealth apps for the community and clinicians.

Methods: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines informed the search of mHealth apps on the Google Play Store and Apple App Store. The search terms included “endometriosis,” “adenomyosis,” and “pelvic pain.” mHealth apps were eligible if they were (1) related to the search terms, (2) were in the English language, and (3) were available free of cost. Only the free content of the eligible mHealth apps was assessed. ENLIGHT, a validated evaluation tool for mobile and web-based interventions, was used to assess the quality across 7 domains such as usability, visual design, user engagement, content, therapeutic persuasiveness, therapeutic alliance, and general subjective evaluation. mHealth apps with a total score of ≥3.5 were classified as “good” according to the ENLIGHT scoring system and are recommended as good-quality mHealth apps for endometriosis care.

Results: In total, 42 mHealth apps were screened, and 19 were included in the quality assessment. A total of 6 good-quality mHealth apps were identified (QENDO, Bearable, Luna for Health, Matilda Health, Branch Health: Pain Management, and CHARLI Health). These apps provided symptom-tracking functions and self-management support. A total of 17 apps were designed for community use, while 2 apps provided a digital endometriosis classification tool to clinicians. Most mHealth apps scored well (≥3.5) in the domains of usability (16/19, 84.2%), visual design (14/19, 73.7%), user engagement (11/19, 57.9%), and content (15/19, 78.9%). Few eHealth websites scored well on therapeutic persuasiveness (6/19, 31.6%), therapeutic alliance (9/19, 47.4%), and general subjective evaluation (6/19, 31.6%).

Conclusions: Although time and geographical location can influence the search results, we identified 6 “good-quality” endometriosis mHealth apps that can be recommended to the endometriosis community. mHealth apps designed for community

XSL•FO
RenderX

use should evaluate their effectiveness on user's endometriosis knowledge, self-recommended management strategies, pain self-efficacy, user satisfaction, and user quality of life. Digital technology should be leveraged to develop mHealth apps for clinicians that contribute to continued medical education and assist clinical decision-making in endometriosis management. Factors that enhance usability, visual design, therapeutic persuasiveness, and therapeutic alliance should be incorporated to ensure successful and long-term uptake of mHealth apps.

Trial Registration: PROSPERO CRD42020185475; https://tinyurl.com/384dkkmj

(J Med Internet Res 2025;27:e49654) doi:10.2196/49654

KEYWORDS

adenomyosis; endometriosis; m-health apps; mobile apps; digital health; pelvic pain; self-learning; clinicians; mHealth; application; endometriomas; chocolate cysts; uterus; womb; pain management; digital health; women's health; mobile phone

Introduction

Endometriosis is a chronic condition characterized by the growth of endometrial-like tissue outside the uterus, causing pain and fertility problems in 190 million women and those assigned female at birth globally [1]. Associated with delayed diagnosis and requiring long-term, individualized care, access to health resources and services is critical to overcoming the negative impact of endometriosis on quality of life and overall well-being [2]. To address this community need, an increasing number of endometriosis mobile health (mHealth) apps [3] have been developed to provide rapid mobile access to symptom assessment and tracking, education, appointment management, and support networking [4]. The value of these mHealth resources in endometriosis care became particularly apparent during COVID-19 lockdowns, when access to doctors and surgery was limited, and elevated endometriosis-related symptoms were observed [5].

As patients increasingly seek digital health information and education [6], medical consultations are progressively more likely to reference mHealth apps. Joint sharing of digital information improves communication between patients and clinicians and enables people to become partners in their health care [4,7,8]. Advancements in digital technologies and the application of machine learning algorithms have led to the development of mHealth interventions [9] to provide more personalized health management [10]. Ball et al [11] trialed a mHealth app intervention to assess the effectiveness of a mindfulness program on chronic pelvic pain in women [11]. In other chronic conditions, including type 1 and type 2 diabetes and hypertension, the integration of digital health strategies led to demonstrable improvements in treatment compliance, disease outcomes [3,11,12], and self-management [13-15].

Clinicians benefit by being informed of the quality and function of endometriosis mHealth apps, as they can provide recommendations to endometriosis community members in a time-efficient, evidenced-based way. There is evidence that mHealth apps can facilitate the diagnosis of endometriosis [16]. The Nezhat Endometriosis Advisor app was trialed to screen populations for the risk of endometriosis, which can assist in early diagnosis [16]. mHealth apps are effective tools that significantly improve medical knowledge and skills among health care providers [17]. The majority of doctors (86%, 170/198) working in pediatric emergency departments across the United Kingdom and Ireland reported using mHealth apps on their phones for clinical decision-making [18]. Ease of use and rapid accessibility facilitates mHealth app use among clinicians [18].

However, poor-quality mHealth apps can spread misinformation [4] and create mistrust in clinical relationships. Furthermore, it is difficult to assess the quality of endometriosis mHealth apps in app stores. There is a need to systematically evaluate endometriosis mHealth apps to inform the endometriosis community and clinicians about the advantages and disadvantages of those currently available using validated and standardized tools [9,19]. Therefore, the aim of this study was to conduct a systematic search of mHealth apps that support endometriosis care to evaluate the quality and recommend those apps that meet specific quality metrics to the endometriosis community and clinicians.

Methods

Search Strategy

This systematic search was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines [20]. (Multimedia Appendix 1). Before undertaking this study, a systematic search protocol was developed and registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration CRD42020185475). There was no patient or public involvement in this study.

"Endometriosis," "adenomyosis," and "pelvic pain" were used as search terms in the 2 largest app stores (Google Play and Apple) on July 24, 2020. Only the first 30 listed mHealth apps were screened, as there is evidence that most people do not investigate beyond this number [21]. Each term was separately searched in the Apple app and Google Play store. The search was later updated on July 22, 2024, to include new mHealth apps and remove redundant ones. Duplicate results were removed, and the remaining mHealth apps were screened for eligibility.

Inclusion and Exclusion Criteria

mHealth apps were included if they related to endometriosis, adenomyosis, or pelvic pain, were in the English language, and were available free of cost. mHealth apps that did not meet the inclusion criteria were excluded (Textbox 1). There were no geographical limitations. For those mHealth apps with free and

XSL•FO
RenderX

paid membership sections, only the free components were assessed. In-app purchases were not made.

Textbox 1. Inclusion and exclusion criteria for including mobile health (mHealth) apps for endometriosis care.

Inclusion criteria:

- Relevance to the subject: mHealth apps that relate to endometriosis, adenomyosis, or pelvic pain in women.
- Cost: Free mHealth apps (no associated cost) or free components of mHealth apps.
- Language: mHealth apps written in the English language.

Exclusion criteria:

- Relevance to the subject: mHealth apps that did not relate to endometriosis, adenomyosis, or pelvic pain in women.
- Cost: mHealth apps that require a payment or subscription to access them.
- Language: mHealth apps written in a language other than English.

Data Extraction

Descriptive data was manually extracted by 1 researcher (DS) after reading an explanation of each mHealth app's purpose. This data was collated in a Microsoft Excel spreadsheet under the following categories: (1) mHealth app name, (2) hyperlink, (3) developer, (4) version, (5) stated purpose, (6) target audience, (7) category, (8) country of origin, (9) last updated, and (10) user rating (Multimedia Appendix 2). The target audience was classified into 2 categories: (1) the endometriosis community (eg, laypeople in the community who have endometriosis or pelvic pain) and (2) clinicians. mHealth apps were classified into 5 types: (1) symptom trackers, which allowed users to record and analyze their symptoms; (2) endometriosis screening tools, which assessed the probability of having endometriosis; (3) pelvic pain symptom assessment tools, which analyzed users pelvic pain symptoms to suggest self-management strategies; (4) a fertility coaching program for those affected with endometriosis; and (5) endometriosis classification tools, which assisted clinicians with endometriosis classification based on radiological and/or surgical findings.

Quality Assessment

The ENLIGHT quality assessment tool was used to evaluate all included mHealth apps [22]. The ENLIGHT tool assesses 7 criteria: (1) usability, (2) visual design, (3) user engagement, (4) content, (5) therapeutic persuasiveness, (6) therapeutic alliance, and (7) general subjective evaluation (Table 1).

XSL•FO
RenderX

Table 1. Description of the ENLIGHT quality assessment criteria [22], objectives, and factors assessed.

Domains of quality assessment	Objective	Criteria assessed
Usability	Assesses the ease of learning how to use a mHealth app and the ease of using it appropriately.	• Navigation • Learnability • Ease of use
Visual design	Assesses the look and feel of the mHealth app and the visual quality of the Graphical User Interface (GUI).	• Aesthetics • Layout • Size
User engagement	Assesses the extent to which the mHealth app design attracts users to use it.	• Content presentation • Interactive • Not irritating • Targeted/tailored/personalized reports • Captivating
Content	Assesses the content provided or learned while using the mHealth app.	• Evidence-based content • Quality of information provided • Complete and concise • Clarity about the program's purpose
Therapeutic persuasiveness	Assesses the extent to which the mHealth app is designed to encourage users to make positive behavior changes or to maintain positive aspects of their lives.	• Call to action • Load reduction of activities • Therapeutic rationale and pathway • Rewards • Real data-driven/adaptive content • On-going feedback • Expectations and relevance
Therapeutic alliance	Assesses the ability of the mHealth app to create an alliance with the user to effect a beneficial change.	• Basic acceptance and support • Positive therapeutic expectations • Relatability
General subjective evaluation of program's potential	Examines the mHealth app's general potential to benefit its target audience based on the rater's subjective evaluation	• Appropriate features to meet the clinical aim • Right mix of ability and motivation • I like the program

The mHealth apps were reviewed in 2 stages. Initially, one researcher (DS) reviewed all included mHealth apps. Then, the mHealth apps were divided and reviewed by a team of six independent researchers (RO, MLH, NB, HS, MP, and CHMN). Hence, each app was reviewed by 2 researchers (DS and 1 of the 6 researchers).

All reviewers were required to download the mHealth app to their smartphone devices (mobile phone or tablet). The ENLIGHT scores of each mHealth app were collated in an Excel spreadsheet. Each ENLIGHT quality assessment criteria was scored using a rating scale of 1-5 (Very Poor to Very Good) using the ENLIGHT checklist [22]

Discrepancies in ratings (any deviation greater than 1 rating unit) were resolved by discussion between the two reviewers. If evaluation differences were not resolved, a third independent assessor was consulted. After a detailed assessment, the average of the 2 reviewer's ratings was used to calculate a score for each of the 7 domains.

Good-Quality mHealth Apps

A total score for each mHealth app was calculated according to the ENLIGHT formula (Multimedia Appendix 3). mHealth apps with a total score of ≥3.5 were classified as "good" according to the ENLIGHT scoring system [23] and were recommended as good-quality mHealth apps for endometriosis care.

Ethical Considerations

An ethics approval was not required for this study, because this was a systematic search of mHealth apps and did not involve the recruitment of participants.

Results

The mHealth app search identified 42 apps, 23 from the Apple App Store and 19 from the Google Play store. A total of 7 mHealth apps were duplicates, leaving 35 apps, of which 13 were excluded as they did not meet the inclusion criteria (Figure 1). A further 3 mHealth apps were excluded; of these, 2 were excluded due to technical difficulties with the app (eg, it kept crashing), and one was excluded as it was specifically for participants enrolled in a research study. A total of 19 mHealth apps were included in the final analysis (Multimedia Appendix 2). There were discrepancies between the first and second reviewers in 6% of the ratings. All discrepancies were resolved without needing a third reviewer through consensus meetings.

XSL•FO
RenderX

Figure 1. Flow chart showing the selection of mobile health (mHealth) apps from Apple App and Google Play stores. Adapted from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram.

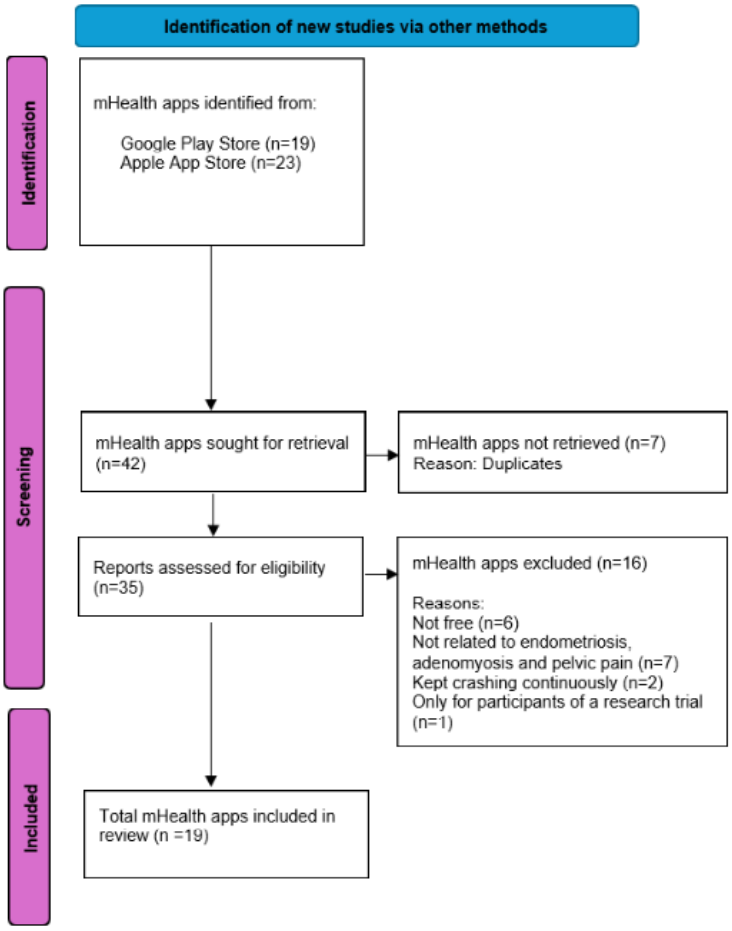

Out of the 19 included mHealth apps, the majority (n=16) were designed for use by laypeople in the community. In addition, 2 mHealth apps were designed for health care providers. The majority of apps mainly provided symptom tracking functions (73.7%, n=14), 2 apps provided digital endometriosis classification tools (10.5%), 1 app offered an endometriosis fertility-related coaching program (5.3%), 1 app provided an endometriosis screening tool (5.3%), and 1 app offered a pelvic pain assessment tool (5.3%).

Quality of mHealth Apps

A total of 6 mHealth apps met the cutoff criteria for classification as “good” (ie, final score >3.5 to 5) according to the ENLIGHT quality assessment criteria. (Table 2). These mHealth apps are recommended as “good quality” apps for endometriosis care for community use.

Table 2. “Good” scoring (≥3.5 to 5) mHealth apps using the ENLIGHT Quality Assessment criteria (based on the final score).

Name	Usability	Visual design	User engagement	Content	TP[a]	TA[b]	GSE[c]	Final score
QENDO [24]	4.67	4.50	4.80	4.38	4.21	4.33	4.00	4.36
Bearable [25]	4.50	4.50	4.40	4.00	4.29	4.00	4.50	4.23
Luna for Health [26]	4.17	4.67	4.10	3.88	4.14	3.50	3.50	4.03
Matilda Health [27]	5.00	4.83	4.60	4.00	3.64	4.50	4.17	3.96
Branch Health: Pain Management [28]	4.00	4.17	3.70	3.50	3.71	3.67	2.83	3.67
CHARLI Health [29]	4.00	3.33	3.40	4.13	3.36	3.50	2.83	3.52

[a]TP: therapeutic persuasiveness.

[b]TA: therapeutic alliance.

[c]GSE: general subjective evaluation.

XSL•FO
RenderX

Usability

Approximately 84.2% (16/19) of the mHealth apps scored "good" (≥3.5) on usability (Multimedia Appendix 3). The highest-ranking mHealth apps were characterized by smooth, nearly frictionless, and easy navigation. These apps also had an intuitive interface that facilitated ease of learning and straightforward use of the app. Their structure was self-explanatory, and the user required minimum effort to obtain the desired output. Some examples of mHealth apps that scored "good" under this domain include QENDO [24], Bearable [25], Matilda Health [26], #Enzian [30], and the American Association of Gynecologic Laparoscopists (AAGL) Endo Classification app (Medicinia Atividades de Internet Ltda) [31]. While the first 3 mHealth apps are designed for community use, the last 2 can be used as resources in clinical practice.

Visual Design

The majority of mHealth apps (73.7%, 14/19) scored "good" (>3.5) on visual design. These mHealth apps (1) had an overall attractive visual design and an appealing color scheme (eg, QENDO [24]), (2) were well structured with a consistent layout (eg, Branch Health Pain Management (Upside Health) [28]), (3) clearly presented content that was easy to read (eg, Luna for Health; HDSI) [26]), and (4) displayed appropriately sized fonts, buttons, and menus (eg, Frendo [32]).

User Engagement

More than half of the mHealth apps scored ≥3.5 on user engagement (57.9%, 11/19). These were characterized by a good mix of text, images, and videos. The content was presented in an interactive and engaging manner. User engagement was further enhanced by avoiding features like pop-up ads, notifications, alerts, and sounds. Some mHealth apps in this domain provided tailored, targeted, and personalized user-specific health reports (eg, QENDO [24] and Matilda Health [27]) and/or charts (eg, Bearable [25]) that users could take to their health appointments. Users of the Luna for Health app [26], in addition to symptom tracking, could also undertake a menstrual well-being test that suggested the need for further medical evaluation.

Content

The majority of mHealth apps scored ≥3.5 under content (78.9%, 15/19). These mHealth apps contained appropriate, complete, and concise information with good clarity about the app's purpose. Some examples include QENDO [24], #Enzian [30], and Flutter [33]. Amongst the 15 apps that scored well in this domain, only 1 app was designed to assist clinicians (#Enzian [30]) in the classification of endometriosis based on radiological and/or surgical findings.

Therapeutic Persuasiveness

A low number of mHealth apps (31.6%, 6/19) scored ≥3.5 under the therapeutic persuasiveness. These apps had a clear call to action and engaged the user by developing health plans or health reports with an actionable outcome. For example, the Bearable app [25] allowed users to set goals like exercising, limiting caffeine intake, or avoiding screen time 1 hour before bedtime. Straightforward and easy-to-complete activities in these apps included completing a health assessment questionnaire, data entry for symptom tracking, and documenting daily challenges. For example, QENDO [24] provided users with a report to take to (or email) their doctor. Matilda Health [27] allowed users to schedule webinars of interest or ask questions to clinicians. The aim of these activities was clear to users. The "good" scoring apps also provided ongoing feedback about the user's health or recommendations regarding testing and likely diagnosis based on symptoms.

Therapeutic Alliance

Less than half (47.4%, 9/19) of the mHealth apps scored "good"(≥3.5) under the therapeutic alliance. These mHealth apps demonstrated basic acceptance, support, and relatability in creative ways. Features included (1) the use of supportive language (for example, the Luna for Health app [26] used phrases like "you are not alone in this journey"), (2) positive therapeutic support (eg, CHARLI Health; Lucid Labs Pty Ltd) [29] provided an option to chat with doctors and endometriosis specialists) and (3) an online community that fosters a sense of connectedness (eg, Flutter [33]).

General Evaluation of the mHealth Apps

This criterion evaluates the reviewers' subjective global score. A low number of mHealth apps (31.6%, 6/19) scored "good" (≥3.5) in this domain. The reviewers felt that they displayed features appropriate to their aims, had the elements to motivate the use of the app, and were generally likable. Some examples include QENDO [24] and Matilda Health [27].

Discussion

Principal Findings

We conducted a comprehensive, multidimensional quality assessment of endometriosis mHealth apps using the ENLIGHT tool [22]. This tool captures quality constructs like therapeutic persuasive design and therapeutic alliance, which are considered central to the successful uptake of mHealth apps among end users. Our study identified six "good" quality endometriosis mHealth apps that can be recommended for consideration in supporting the community in endometriosis care. Additionally, two mHealth apps can be recommended for consideration in assisting healthcare providers in the classification of endometriosis. To our knowledge, this is the first systematic search and assessment to use the ENLIGHT tool to comprehensively assess endometriosis and pelvic pain mHealth apps.

Comparison With Previous Work

mHealth Apps for the Endometriosis Community

In total, 17 out of the 19 mHealth apps in this study could be used by people in the community. mHealth apps are widely used for symptom tracking and self-management of menstrual pain [9]. We found that most mHealth apps were designed for community use, facilitated symptom tracking, and provided users with symptoms and/or activity reports that could be shared with health care providers. Such tracking and activity reporting appear to be desirable features of mHealth apps [34]. A study reported that 77.7% of people affected with endometriosis were

XSL•FO
RenderX

comfortable sharing data tracked on mHealth apps with clinicians, indicating the acceptability of data sharing for health care management [34]. The Bearable app [25] provides users with a correlations grid, demonstrating the impact of certain activities like meditation, exercise, or diet on symptoms, mood, sleep, and/or energy levels. Similarly, the Phendo app (Citizen Endo) [35] provided tailored insights on self-management strategies. In addition, Phendo [35] is a research app, which has used direct self-reported data to study the heterogeneity of endometriosis symptoms [36]. This can help analyze symptom variability among people with endometriosis [36,37], demonstrating the usefulness of digitally generated data in endometriosis research. Self-tracked data from mHealth apps can be used to analyze the association between symptoms and periodicity as potential health indicators for endometriosis assessment [37]. Furthermore, such data can be used as an additional data source that complements patient electronic health records to accurately and comprehensively evaluate patient health history [36].

Apps like Luna for Health [26], Matilda Health [27], and CHARLI Health [29] offer virtual consultations with clinicians specialized in endometriosis and with allied health professionals like physiotherapists and dietitians providing the community with comprehensive alternate pathways for endometriosis care through digital technology. These digital advancements build capacity to meet the gap in the health workforce and improve accessibility for endometriosis management. Data tracking, real-time visual representation of symptoms, the effectiveness of treatments, and self-management strategies can assist clinicians in planning more personalized endometriosis treatments in the future [34].

The effectiveness of mHealth apps for other chronic diseases has previously been evaluated. For example, a Canadian mHealth app, “bant” (University Health Network, Toronto, Ontario), developed for adolescents to manage type 1 diabetes, increased the daily average monitoring of blood glucose levels and demonstrated high user satisfaction [38]. Kollmann et al [39] found a significant decrease in hemoglobin A_{1c} levels using the Diab-Memory App in patients with type 1 diabetes. A mobile phone–based technology called NICHE [12] achieved similar results in patients with type 2 diabetes. There appears to be limited evidence on the evaluation of endometriosis mHealth apps. The Endo-App (Endo Health GmBH) [40] evaluated the impact of the app on disease-related quality of life and symptoms of endometriosis in 122 participants using a randomized controlled pilot trial [41]. Participants in the intervention group reported clinically relevant and statistically significant improvements in pain-specific self-efficacy, fatigue, depression, and quality of life after 12 weeks of use as compared with controls [41], demonstrating the scope of mHealth apps to improve the well-being of those affected by endometriosis. Given this potential, mHealth apps should ideally evaluate their effectiveness on users’ quality of life, impact on endometriosis knowledge, the effectiveness of recommended self-management strategies, and user satisfaction or outcomes of virtual consultations (such as reduced delay in diagnosis or impact on consultation wait times).

Endometriosis mHealth Apps as a Resource in Clinical Practice

This study found only 2 mHealth apps designed specifically for health care providers, both of which offered digital tools to classify endometriosis based on radiological and/or surgical findings. We did not find any mHealth app that provided education to clinicians, yet the need to improve endometriosis awareness among clinicians is expressed by both, the community [2,42] and clinicians [43]. A Dutch study reported that 87% (76/87) of general practitioners (GPs) reported the need to improve endometriosis education with the same study reporting that GPs scored 40% on endometriosis factual knowledge (ie, questions regarding pathophysiology) and 49.4% on clinical knowledge (ie, questions regarding endometriosis symptoms and treatment) [43].

Furthermore, mHealth apps are increasingly being used by clinicians for medical education, training, and clinical decision-making [44]. A study showed that 98.4% of doctors own a smartphone and 92% agree that mHealth apps positively impact clinical practice [45]. In this context, a meta-analysis reported that the pooled effect of 15 studies with 962 participants showed significant improvement in knowledge scores among health care providers who used mHealth apps compared with those who did not use them [17]. Previously, the Endo App by the European Society for Human Reproduction and Embryology (ESHRE) [46] provided a clinical decision-making tool to clinicians in busy clinical settings. Through the decision tool, clinicians could also conveniently access the ESHRE endometriosis clinical practice guidelines [47], which were time-saving in their daily practice. However, this app is no longer available through the app store and appears to have been discontinued. Given the rising popularity and uptake of mHealth apps among clinicians and the need for endometriosis education, digital technology can be leveraged to provide endometriosis-related continuing medical education and clinical decision-making tools for clinicians.

The ENLIGHT Quality Assessment

The ENLIGHT criteria evaluated factors that are central to the long-term and successful uptake of the mHealth apps. The majority of the mHealth apps scored “good” (≥3.5) on usability (84.2%) and visual design (73.7%). However, only 31.6% and 47.4% of mHealth apps scored “good” on user therapeutic persuasiveness and therapeutic alliance respectively. Usability and visual design are important features that determine the success and user uptake of mHealth apps [48,49]. mHealth apps that provide personalized and tailored feedback have demonstrated better user engagement [50,51]. A mental health, microrandomized, clinical trial reported that using push notifications to send a tailored health message was an effective strategy to increase user engagement [50]. Therapeutic persuasiveness is positively correlated with real-world usage of mHealth apps and is an important predictor of user adherence [52]. The ENLIGHT criteria assess the use of rewards to increase therapeutic persuasiveness [21]. Rewards are effective at motivating people [53]. One mHealth app intervention successfully used tangible rewards such as redeeming reward points for purchasing iTunes in an adolescent population [38].

XSL•FO
RenderX

Another study found that participants preferred tangible rewards such as gift cards or monetary incentives as compared with intangible rewards like points or badges [15]. We found a few mHealth apps that provided rewards to users. Users of the Luna for Health app [26] earned "Luna Coins" by participating in quizzes; however, it was unclear how these coins could be used. The therapeutic alliance enhances positive user engagement by fostering relatability [21,52]. It may be that future apps can incorporate conversational agents such as smartphone-based chatbots portraying human involvement [52,54,55] or conversational agents such as Alexa (Amazon), which mimic human conversations [54]. Although mHealth apps are not human, the extent to which they build an alliance is integral to their quality [21].

Strengths

Our systematic search and assessment of the quality of mHealth apps for endometriosis has several strengths. First, it presents novel findings on "good" quality mHealth apps developed for endometriosis care and relevant to the community. We identified mHealth apps that could be considered as a potentially useful resource for use in clinical practice. Clinicians may use the findings of this study to consider recommendations on mHealth app use to support people and families affected by endometriosis. Finally, our study highlights features of therapeutic persuasiveness and therapeutic alliance and additional quality metrics that can be incorporated into future developments of mHealth apps for endometriosis.

Limitations

The ENLIGHT criteria [20] were challenging to uniformly apply. The evaluation of evidence-based content was challenging as most mHealth apps were symptom trackers and presented limited evidence-related content. The quality of content was assessed based on material available in the free version of the app. However, most apps contained endometriosis content and learning modules as part of their paid version. For example, with both, Luna for Health [26] and Matilda Health [27], we were unable to evaluate content completely as we were only able to assess freely available content. In the #Enzian [30] and AAGL Endo Classification app [31], "rewards" and "ongoing feedback," which form part of therapeutic persuasiveness were not features offered to users. Hence, therapeutic persuasiveness was difficult to evaluate, with limited formal guidance on how to navigate such scenarios. Furthermore, this study did not include non-English mHealth apps, due to limitations in funding for translation services. In addition, due to the search algorithm on both app stores that used geographical location, that is, Australia, which is predominantly an English-speaking country, we did not come across any app that was not developed in the English language.

Conclusions

Our systematic search presents novel findings on the quality assessment of mHealth apps that can be used for endometriosis care. Based on the use of a validated tool for rating app quality metrics, our findings are suggestive of "good quality" mHealth apps for community use. These quality ratings can be considered by clinicians as potential recommendations to patients to support their care. Clinicians can use 2 mHealth apps for endometriosis classification in clinical practice. To further strengthen app development for the endometriosis community, we recommend that (1) endometriosis-related mHealth apps incorporate the standardized experience and outcome measures to assess the effectiveness of programs offered. For example, user satisfaction, effectiveness of recommended self-management strategies, impact on user's endometriosis knowledge, pain self-efficacy, and overall quality of life, (2) digital technology is leveraged to develop clinical resources for endometriosis-related continued medical education and to support clinical decision-making, and (3) good design features are considered more carefully including therapeutic persuasiveness and therapeutic alliance for the successful uptake of endometriosis mHealth apps.

Authors' Expertise in Conducting the Search and Quality Assessment

The interprofessional team consisted of endometriosis researchers, clinical pain researchers, clinicians (gynecologist, specialist physiotherapist, and endometriosis nurse specialist), and digital health experts. The team has expertise in the clinical management of endometriosis, community engagement, co-design of digital health tools, endometriosis content, and evaluation research to support self-management and care. All authors were involved in the quality assessment.

Acknowledgments

We thank Ms Vikki Langton (Librarian – Adelaide Medical School, University of Adelaide) for rendering assistance in formulating the search strategy. DS is a PhD (Medicine) candidate and is supported by the Australian Government, Research Training Program Stipend.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1-3.

Authors' Contributions

DS contributed to the study design, conducted a search of mHealth apps, data extraction, quality assessment, data collation, data analysis and drafting of the manuscript. RO contributed to the study design conducted the search of mHealth apps, contributed to quality assessment, and drafted the manuscript. CHMN contributed to the quality assessment and drafting of the manuscript.

XSL•FO
RenderX

NB, HS, MP, and MLH contributed to the study design, quality assessment, and reviewing the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

Although this systematic assessment has not received separate funding, it is part of the EndoZone, a digital platform for endometriosis project funded by the Australian Government Department of Health and Aged Care and Jean Hailes for Women's Health. CHMN is an employee of The University of New South Wales and manages grant funding from the Australian Government Department of Health and Aged Care, and Medical Research Future Fund, and was a previous employee of Jean Hailes for Women's Health and CSL Vifor (formerly Vifor Pharma Pty Ltd). NB has shares in goAct Pty Ltd and received research funding support from Digital Health CRC, Medical Research Future Fund, and Barossa Fleurieu Local Health Network. Other authors declare no conflicts of interest.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 Checklist and PRISMA 2020 Abstracts Checklist.
[DOCX File , 25 KB - jmir_v27i1e49654_app1.docx]

Multimedia Appendix 2
Details of mobile health apps included in this study.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v27i1e49654_app2.xlsx]

Multimedia Appendix 3
Final score of mobile health apps based on the ENLIGHT scoring system.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v27i1e49654_app3.xlsx]

References

1. Horne AW, Missmer SA. Pathophysiology, diagnosis, and management of endometriosis. BMJ 2022 Nov 14;379:e070750. [doi: 10.1136/bmj-2022-070750] [Medline: 36375827]
2. Culley L, Law C, Hudson N, Denny E, Mitchell H, Baumgarten M, et al. The social and psychological impact of endometriosis on women's lives: a critical narrative review. Hum Reprod Update 2013;19(6):625-639. [doi: 10.1093/humupd/dmt027] [Medline: 23884896]
3. Omboni S, Caserini M, Coronetti C. Telemedicine and M-Health in hypertension management: technologies, applications and clinical evidence. High Blood Press Cardiovasc Prev 2016;23(3):187-196. [doi: 10.1007/s40292-016-0143-6] [Medline: 27072129]
4. Gkrozou F, Waters N. A preliminary review of the mHealth apps focused in endometriosis and chronic pelvic pain. Health Science Journal 2019;13(3):1-6 [FREE Full text] [Medline: 14770254]
5. Ramos-Echevarría PM, Soto-Soto DM, Torres-Reverón A, Appleyard CB, Akkawi T, Barros-Cartagena BD, et al. Impact of the early COVID-19 era on endometriosis patientsymptoms, stress, and access to care. Journal of Endometriosis and Pelvic Pain Disorders 2021;13(2):111-121 [FREE Full text] [doi: 10.1177/22840265211009634]
6. Shoebotham A, Coulson NS. Therapeutic affordances of online support group use in women with endometriosis. J Med Internet Res 2016;18(5):e109 [FREE Full text] [doi: 10.2196/jmir.5548] [Medline: 27160641]
7. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
8. Slater H, Briggs A, Stinson J, Campbell JM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: a qualitative systematic review protocol. JBI Database System Rev Implement Rep 2017;15(8):2047-2054. [doi: 10.11124/JBISRIR-2016-003299] [Medline: 28800054]
9. Moumane K, Idri A. Mobile applications for endometriosis management functionalities: analysis and potential. Scientific African 2023:e01833 [FREE Full text] [doi: 10.1016/j.sciaf.2023.e01833]
10. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida A, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
11. Ball E, Newton S, Rohricht F, Steed L, Birch J, Dodds J, et al. mHealth: providing a mindfulness app for women with chronic pelvic pain in gynaecology outpatient clinics: qualitative data analysis of user experience and lessons learnt. BMJ Open 2020;10(3):e030711 [FREE Full text] [doi: 10.1136/bmjopen-2019-030711] [Medline: 32165550]
12. Faridi Z, Liberti L, Shuval K, Northrup V, Ali A, Katz D. Evaluating the impact of mobile telephone technology on type 2 diabetic patients' self-management: the NICHE pilot study. J Eval Clin Pract 2008;14(3):465-469. [doi: 10.1111/j.1365-2753.2007.00881.x] [Medline: 18373577]

XSL•FO
RenderX

13. Vansimaeys C, Benamar L, Balagué C. Digital health and management of chronic disease: a multimodal technologies typology. Int J Health Plann Manage 2021;36(4):1107-1125. [doi: 10.1002/hpm.3161] [Medline: 33786849]
14. Arsand E, Tufano JT, Ralston JD, Hjortdahl P. Designing mobile dietary management support technologies for people with diabetes. J Telemed Telecare 2008;14(7):329-332. [doi: 10.1258/jtt.2008.007001] [Medline: 18852310]
15. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]
16. Jackman J, Vaid A, Agarwal S, Nezhat A, Nezhat C. Can we diagnosis endometriosis with a phone app? Nezhat endometriosis advisor mobile application as a predictor for endometriosis in patients experiencing pelvic pain, infertility or unexplained infertility. Fertility and Sterility 2024;114:e539 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.09.057]
17. Chandran V, Balakrishnan A, Rashid M, Pai Kulyadi G, Khan S, Devi E, et al. Mobile applications in medical education: a systematic review and meta-analysis. PLoS One 2022;17(3):e0265927 [FREE Full text] [doi: 10.1371/journal.pone.0265927] [Medline: 35324994]
18. Jahn HK, Jahn IHJ, Behringer W, Lyttle MD, Roland D, Paediatric Emergency Research United Kingdom Ireland (PERUKI). A survey of mHealth use from a physician perspective in paediatric emergency care in the UK and Ireland. Eur J Pediatr 2021;180(8):2409-2418 [FREE Full text] [doi: 10.1007/s00431-021-04023-0] [Medline: 33763717]
19. Larson RS. A path to Better-Quality mHealth apps. JMIR Mhealth Uhealth 2018;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
21. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
22. Baumel A, Kane, Muench. ENLIGHT Summary Report. 2016. URL: https://mindtools.io/wp-content/uploads/2024/01/Enlight-Quality-Assessment-Section.pdf [accessed 2020-07-27]
23. Scientific approach. Mindtools.io. 2023. URL: https://mindtools.io/scientific-approach/ [accessed 2024-10-30]
24. QENDO. URL: https://www.qendo.org.au/qendo-app [accessed 2024-07-24]
25. Bearable. URL: https://play.google.com/store/apps/dev?id=8979963710425634832&hl=en_US [accessed 2024-07-24]
26. Luna for Health. URL: https://play.google.com/store/apps/details?id=com.hdsi.luna&hl=en_AU [accessed 2024-07-24]
27. Matilda Health. URL: https://apps.apple.com/au/app/matilda/id6449485262 [accessed 2024-07-24]
28. Branch Health: Pain Management. URL: https://branch-health-pain-management.en.softonic.com/android [accessed 2024-07-24]
29. Charli. URL: https://apps.apple.com/au/app/charli-health/id6479275179 [accessed 2024-07-24]
30. #Enzian. URL: https://apps.apple.com/us/app/enzian/id6449304239 [accessed 2024-07-24]
31. AAGL Endo Classification. URL: https://apps.apple.com/au/app/aagl-endo-classification/id1592383297 [accessed 2024-07-24]
32. Frendo. URL: https://play.google.com/store/apps/details?id=com.devensis.frendo&hl=en_AU [accessed 2024-07-24]
33. Flutter. Flutter - period tracker and endometriosis journal. Flutter Health Inc. URL: https://apps.apple.com/us/app/flutter-period-tracker-and-endometriosis-journal/id1094193373 [accessed 2024-07-24]
34. Scheck SM, Hohmann-Marriott B, Faughnan G, Henry C. A survey exploring the use of symptom diary mobile apps for endometriosis. Journal of Endometriosis and Pelvic Pain Disorders 2023;15(3-4):128-133. [doi: 10.1177/22840265231193486]
35. Phendo. Columbia University Department of Biomedical Informatics. URL: https://citizenendo.org/phendo/ [accessed 2024-07-24]
36. Ensari I, Pichon A, Lipsky-Gorman S, Bakken S, Elhadad N. Augmenting the clinical data sources for enigmatic diseases: a cross-sectional study of self-tracking data and clinical documentation in endometriosis. Appl Clin Inform 2020;11(5):769-784 [FREE Full text] [doi: 10.1055/s-0040-1718755] [Medline: 33207385]
37. Li K, Urteaga I, Wiggins C, Druet A, Shea A, Vitzthum V, et al. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile-health data. NPJ Digit Med 2020;3:79 [FREE Full text] [doi: 10.1038/s41746-020-0269-8] [Medline: 32509976]
38. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
39. Kollmann A, Riedl M, Kastner P, Schreier G, Ludvik B. Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients. J Med Internet Res 2007;9(5):e36 [FREE Full text] [doi: 10.2196/jmir.9.5.e36] [Medline: 18166525]
40. Endo-App. Endo Health GmBH. URL: https://endometriose.app/en/endo-app-download-for-google-android/ [accessed 2024-12-09]
41. Rohloff N, Götz T, Kortekamp SS, Heinze NR, Weber C, Schäfer SD. Influence of App-Based Self-Management on the quality of life of women with endometriosis. Cureus 2024;16(8):e67655. [doi: 10.7759/cureus.67655] [Medline: 39314601]

XSL•FO
RenderX

42. Sirohi D, Freedman S, Freedman L, Carrigan G, Hey-Cunningham AJ, Hull ML, et al. Patient experiences of being advised by a healthcare professional to get pregnant to manage or treat endometriosis: a cross-sectional study. BMC Womens Health 2023;23(1):638 [FREE Full text] [doi: 10.1186/s12905-023-02794-2] [Medline: 38037049]
43. van der Zanden M, Nap AW. Knowledge of, and treatment strategies for, endometriosis among general practitioners. Reprod Biomed Online 2016;32(5):527-531. [doi: 10.1016/j.rbmo.2016.02.003] [Medline: 26947452]
44. Lee M, Bin Mahmood ABS, Lee ES, Smith HE, Tudor Car L. Smartphone and mobile app use among physicians in clinical practice: scoping review. JMIR Mhealth Uhealth 2023;11:e44765 [FREE Full text] [doi: 10.2196/44765] [Medline: 37000498]
45. El Hadidy TS, Alshafei AE, Mortell AE, Doherty EM. Smartphones in clinical practice: doctors' experience at two dublin paediatric teaching hospitals. Ir J Med Sci 2018;187(3):565-573. [doi: 10.1007/s11845-017-1713-x] [Medline: 29170970]
46. ESHRE launches an endometriosis app. Endo app by ESHRE European Society for Human Reproduction and Embryology. URL: https://endometriosis.org/news/general/eshre-launches-endometriosis-app/ [accessed 2024-10-07]
47. Dunselman G, Vermeulen N, Becker C, Calhaz-Jorge C, D'Hooghe T, De Bie B, European Society of Human Reproduction Embryology. ESHRE guideline: management of women with endometriosis. Hum Reprod 2014;29(3):400-412. [doi: 10.1093/humrep/det457] [Medline: 24435778]
48. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;s2:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
49. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
50. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
51. Ryan P, Lauver DR. The efficacy of tailored interventions. J Nurs Scholarsh 2002;34(4):331-337. [doi: 10.1111/j.1547-5069.2002.00331.x] [Medline: 12501736]
52. Baumel A, Kane JM. Examining predictors of Real-World user engagement with Self-Guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
53. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
54. Ly K, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
55. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]

Abbreviations

AAGL: American Association of Gynaecologic Laparoscopists
ESHRE: European Society for Human Reproduction and Embryology
GP: general practitioner
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews

Edited by A Mavragani; submitted 05.06.23; peer-reviewed by CL Sánchez Bocanegra, C Kaya, E Askary, M Danielis, M Zugaj; comments to author 14.06.24; revised version received 28.11.24; accepted 11.12.24; published 07.02.25.

Please cite as:
Sirohi D, Ng CHM, Bidargaddi N, Slater H, Parker M, Hull ML, O'Hara R
Good-Quality mHealth Apps for Endometriosis Care: Systematic Search
J Med Internet Res 2025;27:e49654
URL: https://www.jmir.org/2025/1/e49654
doi:10.2196/49654
PMID:39918848

©Diksha Sirohi, Cecilia HM Ng, Niranjan Bidargaddi, Helen Slater, Melissa Parker, M Louise Hull, Rebecca O'Hara. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed

XSL•FO
RenderX

under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartphone-Based Telemonitoring for Better Oral Health With Toothbrushes: 6-Month Randomized Controlled Trial

Jaeyeon Kim[1*], PhD; Yiseul Choi[1,2*], PhD; Yoolbin Song[1], MS; Wonse Park[1,2,3], PhD

[1]Department of Advanced General Dentistry, Yonsei University College of Dentistry, Seoul, Republic of Korea

[2]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

[3]Human Identification Research Institute and Oral Science Research Center, Yonsei University College of Dentistry, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Wonse Park, PhD
Department of Advanced General Dentistry
Yonsei University College of Dentistry
50-1, Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 8990
Fax: 82 2 2227 8906
Email: wonse@yuhs.ac

Abstract

Background: A toothbrush device that telemonitors toothbrushing is a technologically advanced solution providing personalized feedback on toothbrushing habits and oral hygiene. These devices integrate smartphone apps to enhance oral health compliance through dental professional feedback.

Objective: This 6-month prospective randomized controlled trial aimed to compare the clinical effectiveness, defined as improved oral hygiene measured by plaque reduction and halitosis control, of an interactive telemonitoring toothbrush (ITT), an oscillating-rotating power toothbrush (ORT), and a manual toothbrush (MT).

Methods: Participants were recruited offline from the Department of Advanced General Dentistry at Yonsei University Dental Hospital, South Korea. A total of 150 participants were randomly assigned to 3 groups (50 participants each): (1) an ITT connected to a smartphone app providing real-time feedback and weekly dental professional reviews, (2) an ORT with smartphone-based guidance requiring participants to send weekly brushing records via screenshots, and (3) an MT with a brushing diary for review. Data collection occurred in clinical settings. Primary outcomes included plaque reduction measured using the Simple Hygiene Score (SHS), while secondary outcomes included plaque reduction measured using the Turesky modification of the Quigley-Hein plaque index (QHI), reductions in halitosis, and changes in oral microbiota. All outcomes were assessed at baseline and 1 month, 3 months, and 6 months.

Results: A total of 150 participants completed the study. Over 6 months, the SHS increased in the MT group (mean 3.16, SD 4.86 to mean 5.66, SD 5.20) but significantly decreased in the ITT group (mean 3.47, SD 5.50 to mean 2.27, SD 3.82; *P*=.004). Similarly, QHI decreased more in the ITT group (mean 1.79, SD 0.72 to mean 0.85, SD 0.63) than in the ORT (*P*<.001) and MT (*P*<.001) groups. Regarding microbiota, there were no significant differences in high-risk periodontal microbiota or the ratio of caries-risk to anticaries microbiota between the ITT and ORT groups. However, in the MT group, the ratio of caries-risk microbiota was significantly higher at the 3-month (*P*<.001) and 6-month (*P*=.005) recalls than at baseline and at the 3-month (*P*=.048) and 6-month (*P*=.03) recalls than at the 1-month recall. Poststudy questionnaires indicated that 45 of 50 ITT participants (92%) and 37 of 50 ORT participants (76%) reported improved brushing ability. The most effective feature in the ITT group was brushing training, while participants in the ORT group cited the brushing guide as most useful (*P*<.001). Satisfaction scores were higher in the ORT group (mean 7.90, SD 1.21) than in the ITT group (mean 7.15, SD 1.66; *P*=.004). The number of brushing events decreased significantly in the ORT group (*P*=.02), while brushing duration increased in the MT group (*P*=.01).

Conclusions: ITTs enable better oral hygiene management than MTs through dental professional feedback. However, further studies are needed to optimize feedback intervals and improve long-term adherence.

XSL•FO
RenderX

Trial Registration: Clinical Research Information Service (CRIS), Republic of Korea, KCT0009094; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=26110&search_page=L

(J Med Internet Res 2025;27:e65128) doi:10.2196/65128

KEYWORDS

clinical studies; clinical trials; oral hygiene; plaque; plaque biofilms; halitosis; microbiome

Introduction

Toothbrushing plays an important role in plaque removal and maintenance of optimal oral health [1]. It can also prevent oral diseases, particularly caries, gingivitis, and periodontitis [2,3]. Moreover, a prior study reported that regular toothbrushing, flossing, and dental visits are essential to prevent dental problems and maintain good oral hygiene [4].

With technological advances, the oral care industry has introduced smart toothbrushes. A smart toothbrush is a technologically advanced toothbrush that uses sensors, connectivity, and data analysis to provide personalized feedback on an individual's toothbrushing habits and oral hygiene. These toothbrushes have various features, such as a pressure sensor to prevent overbrushing, a timer to ensure that the tooth surfaces in the mouth are brushed for an appropriate amount of time, and a Bluetooth connection that sends data to a mobile app for analysis. Additionally, some smart toothbrushes can be used by professionals as telemonitoring devices for oral care advice.

Software that offers real-time visual feedback on an individual's brushing movements is being developed to improve brushing techniques. The real-time feedback provided by the toothbrush connected to the mobile app allows users to enhance their toothbrushing technique to improve oral hygiene and reduce the risk of dental problems, such as plaque and gingivitis [5]. In addition, these telemonitoring toothbrushes can help users maintain the right habits by monitoring their toothbrushing habits and providing notifications to signal the recommended brushing time [6]. However, relatively few studies have compared the efficacy of telemonitoring and manual toothbrushes; thus, more research on the impact of telemonitoring toothbrushes on oral health is needed [7-9].

We hypothesized that, among smart toothbrushes, toothbrushes with a telemonitoring function might improve users' oral hygiene compared with other toothbrushes. Therefore, this study aimed to report the results of a 6-month prospective randomized controlled trial to compare the clinical effectiveness of using an interactive telemonitoring toothbrush (ITT) or an oscillating-rotating electric toothbrush (ORT) among smart toothbrushes as well as a manual toothbrush (MT). Furthermore, we compared the overall oral hygiene results regarding plaque, halitosis, dental caries, and periodontal microbiota between the 3 groups using the 3 different types of toothbrushes. In addition, participants' toothbrushing habits, awareness of telemonitoring toothbrushes, and compliance were also investigated.

Methods

Study Design

This study was a parallel-group, randomized, controlled, single-blind clinical trial and was conducted in the Department of Advanced General Dentistry between January 2021 and May 2022. In this trial, the data analysts were blinded to the group assignments to ensure unbiased statistical analysis. This study was registered with the Clinical Research Information Service and is reported following the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1) [10].

Participants

Participants were recruited via offline recruitment strategies. The study recruited participants through notices posted at Yonsei University Dental Hospital and within Yonsei University. Interested individuals underwent a screening process, including a review of medical records and an oral examination, to confirm eligibility based on the study criteria. The inclusion criteria were as follows: patients (1) with good general health and who were aged >19 years; (2) who had 24 or more teeth, including implants or bridges; and (3) who used a smartphone. The exclusion criteria were as follows: patients (1) with moderate to severe periodontitis [11]; (2) with orthodontic devices in the oral cavity; (3) with removable dentures; (4) with salivary gland–related diseases; (5) with halitosis caused by systemic diseases (eg, digestive system–related diseases, liver-related diseases, kidney-related diseases, medication-related osteonecrosis of the jaw); (6) who had received head and neck radiation therapy; (7) who were pregnant or lactating; (8) who lacked communication skills (including disabled individuals); and (9) whose work was related to dentistry or medicine

The sample size was 150 participants, divided into 3 groups of 50 each. The sample size was calculated using G-Power 3.1.9.2, with an effect size of 0.3, significance level of .05, and statistical power of 85%, considering a dropout rate of 15% [12]. The study participants were randomly divided into 3 groups (ORT, ITT, and MT) using block randomization (block size: 5) via software (Excel, Microsoft Corp) before the start of the study. Participants identified as eligible were randomized using a computerized randomization tool by 2 independent researchers who were not involved in the rest of the study. Before participating in the study, participants who met the inclusion criteria completed a self-reported questionnaire.

Procedures and Study Test Products

In this study, an MT and 2 types of smart toothbrushes were used: (1) ITT and (2) ORT (Figure 1).

XSL•FO
RenderX

Figure 1. Toothbrush types used in the study: (A) Oral-B Genius 8000 (Procter & Gamble), (B) Mombrush (XiuSolution), (C) manual toothbrush (PRO-SYS Sensitive Toothbrush).

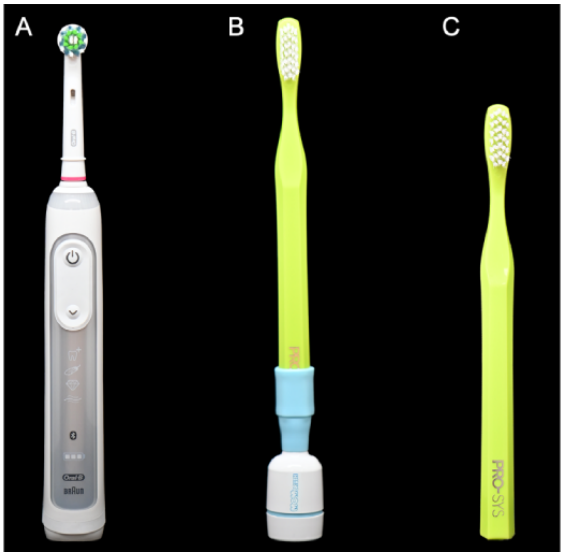

The ITT (Mombrush; XiuSolution) was chosen as the primary intervention due to its advanced telemonitoring capabilities. This toothbrush connects to the Mombrush ProCare smartphone app via Bluetooth, allowing for real-time tracking of brushing behavior, including frequency, duration, and technique.

After installing the Mombrush ProCare app on the participants' smartphones, they brushed according to the video guide on how to brush with the rolling method (Multimedia Appendix 2). Participants in the ITT group were also registered in the Mombrush advisor application, which allowed the researcher to check the records of continuous brushing habits, regular brushing habits, brushing balance, and brushing areas in real time. These brushing records were stored on a cloud server and reviewed weekly by dental professionals, who provided personalized feedback on regular brushing habits, brushing balance, and brushing zones.

The ORT with a CrossAction brush head (Oral-B Genius 8000; Procter & Gamble), which was included as a comparator, features smartphone-guided brushing without telemonitoring capabilities. It represents an advanced electric toothbrush widely used in practice, enabling a direct comparison with the telemonitoring approach.

Selected participants were instructed to install the Oral-B app and connect the electric toothbrush to their smartphone via Bluetooth (Multimedia Appendix 3). Participants brushed in Clean mode using the coaching function provided by the app. Since this app was an on-device app, the researcher could not check the brushing records in real time, so the researcher received a screenshot of the brushing records once a week and provided feedback.

The American Dental Association reference MT (PRO-SYS Sensitive Toothbrush; Benco Dental) was selected for the control group and serves as the standard brushing method. This allowed for the evaluation of the added benefits provided by digital and telemonitoring interventions.

The MT group recorded the number of brushing strokes in a brushing diary after receiving brushing instruction. Participants in the MT group sent their brushing diaries to the researcher once a week. The diaries were reviewed, and a professional provided feedback.

All participants were instructed to brush their teeth with the products distributed at least twice daily during the study period and to abandon other oral hygiene aids, such as dental floss, interdental brushes, and mouthwashes. All groups were provided with standard sodium fluoride toothpaste (1450 ppm NaF) for use with their assigned toothbrush.

Clinical Measurement Outcomes

Primary Outcome: Simple Hygiene Score (SHS)

To evaluate dental plaque, the following 5 white-light and fluorescent images were captured using the quantitative light-induced fluorescence system (Qraycam; AIOBIO): (1) frontal photograph showing the labial surfaces of the anterior teeth, (2) right and (3) left lateral photographs showing the buccal surfaces of the posterior teeth, (4) maxillary dentition photograph showing the palatal and occlusal surfaces of the maxillary dentition, (5) mandibular dentition photograph showing lingual and occlusal surfaces of the mandibular dentition. The fluorescent plaque index scoring for the fluorescent images of the quantitative light-induced fluorescence system was performed automatically using the Simple Hygiene Score (SHS) with the analysis program Q-ray (version 1.38; Inspektor Research Systems BV). The SHS scores the plaque level from 0 to 5 based on the area of red fluorescent plaque attachment, with a larger plaque attachment area giving a higher score [13].

Secondary Outcomes

Plaque Reduction

The Turesky modification of the Quigley-Hein plaque index (QHI) was used to measure the presence of plaque [14,15]. The crown and cervical surfaces of the maxillary right first molar, left central incisor, left first premolar, mandibular left first molar, right central incisor, and right first premolar were stained with the disclosing solution (1% neutral red). Plaques were

XSL•FO
RenderX

evaluated at 2 sites (buccal and lingual) per tooth, and each site was scored on a scale of 0 to 5 (0=no plaque, 1=slight staining at the cervical margin, 2=plaque band up to 1 mm at the cervical margin, 3=plaque band wider than 1 mm but covering less than one-third of the crown of the tooth, 4=band covering at least one-third but less than two-thirds of the crown of the tooth, and 5=band covering more than two-thirds of the crown of the tooth).

Volatile Sulfur Compounds

To measure halitosis, H_2S, and CH_3SH concentrations were measured using the Twin Breasor II (iSenLab Inc) with a nanotech semiconductor sensor made by a plasma ion beam. The participants held a straw in their mouth, breathed through their nose for 50 seconds, and exhaled through their mouth for 10 seconds. Exhaled breath was collected by an automatic suction method using a straw, and the sampling gas volume was approximately 10 mL. Analysis was performed 150 seconds after exhalation, and the measurement units were ng/10 mL and ppb.

Collection of Saliva Samples

To examine caries-related and periodontal bacteria in the oral cavity, a T-SWAB TRANSPORT UTM (Noble Biosciences) was used to swab the gingiva, cervical region, and tooth area of the right maxillary and mandibular molars for more than 30 seconds. Afterward, the cotton swabs were stored in a collection container containing a preservative solution and frozen at –80 °C before DNA extraction. Bacterial genomic DNA was extracted using the MagNA Pure 96 DNA and Viral NA Small Volume Kit (Roche Diagnostics) according to the manufacturer's instructions. DNA concentration was determined fluorometrically on the Qubit 3.0 Fluorometer (Thermo Fisher Scientific) using the Qubit dsDNA HS Assay Kit. Real-time polymerase chain reaction (PCR) was performed using the PowerCheck Periodontitis Pathogens Multiplex Real-time PCR kit (KogeneBiotech) and PowerCheck Dental Caries Pathogens Multiplex Real-time PCR kit (KogeneBiotech). In this study, *Aggregatibacter actinomycetemcomitans*, *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola* were classified as high-risk periodontal microbiota. *Streptococcus mutans*, *Streptococcus obrinus*, *Actinomyces gerencseriae*, *Scardovia wiggsiae*, *Veillonella parvula*, and *Candida albicans* were classified as caries-risk microbiota, whereas *Streptococcus sanguinis* was classified as anticaries microbiota [16].

The clinical evaluations at baseline and follow-ups were completed by dental professionals in the same dedicated location within the Department of Advanced General Dentistry at Yonsei University Dental Hospital. To ensure consistency and minimize interexaminer variability, the same examiner performed all measurements and assessments for each participant throughout the study. After clinical evaluation at baseline, scaling was performed for all participants, and toothbrushes assigned in advance were distributed. After 1 month, 3 months, and 6 months from the beginning of the study, the participants underwent follow-up clinical evaluations in the same manner as at baseline. At the last visit, the participants completed a self-reported, paper-based questionnaire consisting of 8 questions on current oral care habits and awareness regarding telemonitoring toothbrushes. The dental professional abstracted the responses from the paper-based questionnaires and inputted the data into a database for subsequent analysis.

Statistical Analysis

All statistical analyses were performed using SPSS (version 25; IBM Corp). Data were tested for normality using the Kolmogorov Smirnov test. One-way ANOVA and chi-square tests were used to analyze the demographics and questionnaire responses of the study participants. Independent-sample *t* tests were used to compare changes in brushing habits by group. SHS, QHI, and volatile sulfur compounds were analyzed using repeated measures ANOVA with Bonferroni multiple comparison tests. Changes in microbiota within groups were analyzed using a paired-samples *t* test. All statistical analyses were 2-tailed, and the statistical significance level was set at $P<.05$.

Ethical Considerations

The Institutional Review Board (IRB) of the Dental Hospital at Yonsei University reviewed and approved the study protocol prior to the trial in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments (IRB number: 2-2020-0032). All participants completed a written consent form and were given information regarding the study. All data were anonymized to ensure participant confidentiality.

Results

Participant Characteristics

A total of 161 participants completed the screening; 11 of the screened participants did not meet the inclusion criteria, resulting in 150 participants who completed the study (Figure 2). The general characteristics of the participants are presented in Table 1. The average age of the participants was 31.90 (SD 7.34) years, and 61 (61/150, 40.7%) men and 89 (89/150, 59.3%) women were included. Regarding their academic qualifications, a bachelor’s degree was the most common academic qualification in all groups. There were no statistically significant differences in general characteristics among the 3 groups. Similarly, there were no statistically significant differences between the groups in the self-reported questionnaire on current oral care habits and awareness of telemonitoring toothbrushes at the beginning of the study.

XSL•FO
RenderX

Figure 2. CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) flow diagram.

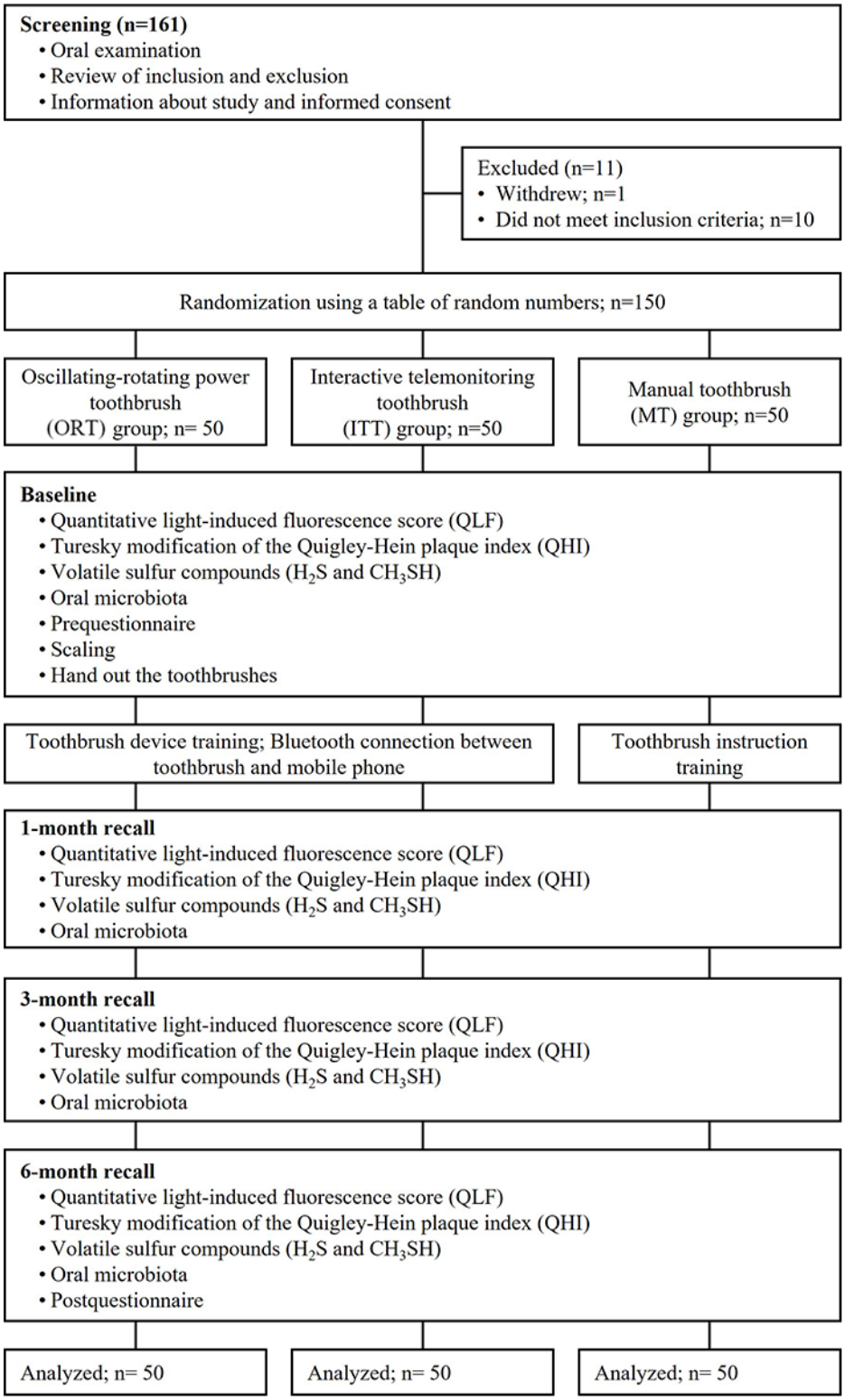

XSL•FO
RenderX

Table 1. General characteristics and responses to the prequestionnaire on oral care habits and participants' awareness of telemonitoring toothbrushes.

Characteristics	ORT[a] (n=50)	ITT[b] (n=50)	MT[c] (n=50)	*P* value[d]
Age (years), mean (SD)	29.10 (4.72)	30.96 (8.39)	32.66 (8.13)	.053
Sex, n (%)				.67
Male	23 (46)	19 (38)	19 (38)	
Female	27 (54)	31 (62)	31 (62)	
Educational level, n (%)				.96
High school graduate	10 (20)	10 (20)	9 (18)	
Associate's degree	5 (10)	4 (8)	5 (10)	
Bachelor's degree	21 (42)	26 (52)	23 (46)	
Master's degree or higher	14 (28)	10 (20)	13 (26)	
Number of toothbrushing events, mean (SD)	2.53 (0.62)	2.72 (0.61)	2.70 (0.61)	.28
Length of each toothbrushing event (minutes), mean (SD)	2.84 (0.75)	3.10 (0.81)	3.12 (0.82)	.20
Have you ever received training on oral hygiene management?, n (%)				.29
Yes	24 (48)	32 (64)	28 (56)	
No	26 (52)	18 (36)	22 (44)	
How long do you think you need training to have a good brushing habit?, n (%)				.54
1 year or more	4 (8)	2 (4)	0	
6 months	3 (6)	3 (6)	2 (4)	
3 months	6 (12)	8 (16)	9 (18)	
1 month	13 (26)	15 (30)	20 (40)	
1 week	24 (48)	22 (44)	19 (38)	
Do you think that good brushing habits help to maintain healthy teeth?, n (%)				.42
Strongly agree	28 (56)	32 (64)	35 (70)	
Agree	20 (40)	14 (28)	14 (28)	
Neutral	2 (4)	4 (8)	1 (2)	
Disagree	0	0	0	
Strongly disagree	0	0	0	
Have you ever heard of a telemonitoring toothbrush?, n (%)				.15
I know well	1 (2)	1 (2)	1 (2)	
I know a little	3 (6)	5 (10)	5 (10)	
I don't know	27 (54)	14 (28)	24 (48)	
I have never heard	19 (38)	30 (60)	20 (40)	
Are you interested in trying out a telemonitoring toothbrush?, n (%)				.87
Strongly agree	27 (54)	22 (44)	21 (42)	
Agree	16 (32)	20 (40)	18 (36)	
Neutral	3 (6)	6 (12)	7 (14)	
Disagree	2 (4)	1 (2)	2 (4)	
Strongly disagree	2 (4)	1 (2)	2 (4)	
Degree of interest in information technology devices (1 to 5), mean (SD)	3.84 (1.10)	4.06 (0.87)	3.88 (0.90)	.48
Frequency of use of information technology devices (1 to 5), mean (SD)	4.02 (1.04)	4.12 (0.90)	3.92 (0.97)	.59

[a]ORT: oscillating-rotating power toothbrush.

[b]ITT: interactive telemonitoring toothbrush.

[c]MT: manual toothbrush.

XSL•FO
RenderX

[d]1-way ANOVA for continuous variables and chi-square test categorical variables.

Changes in SHS, QHI, and H2S and CH3SH Levels

The mean SHS increased over time in the MT group (3.16, SD 4.86 to 5.66, SD 5.20) and decreased at the 1-month recall in the ORT and ITT groups (1.53, SD 3.24 and 1.41, SD 3.49, respectively). Compared with the MT group, the SHS significantly decreased over time in the ITT group (*P*=.004; Figure 3A). Regarding QHI, the mean scores decreased from baseline at the 1-month recall in all 3 groups (ORT group: 2.2, SD 0.72 to 1.39, SD 0.53; ITT group: 1.79, SD 0.72 to 1.19, SD 0.69; MT group: 1.99, SD 0.84 to 1.74, SD 0.79). However, at the 6-month recall, the scores in the ORT and MT groups increased (ORT group: 1.68, SD 1.11; MT group: 1.86, SD 1.25), while that in the ITT group decreased (0.85, SD 0.63; *P*<.001; Figure 3B). However, there were no significant differences in H_2S and CH_3SH levels among the 3 groups during the entire study period (Figures 3C and 3D).

Figure 3. Comparison of the mean (SD) of the (A) Simple Hygiene Score (SHS), (B) Turesky modification of the Quigley-Hein plaque index (QHI), and halitosis in terms of (C) H2S and (D) CH3SH values between the oscillating-rotating power toothbrush (ORT), interactive telemonitoring toothbrush (ITT), and manual toothbrush (MT) groups uing 2-way, repeated-measures ANOVA with the Bonferroni multiple comparisons test.

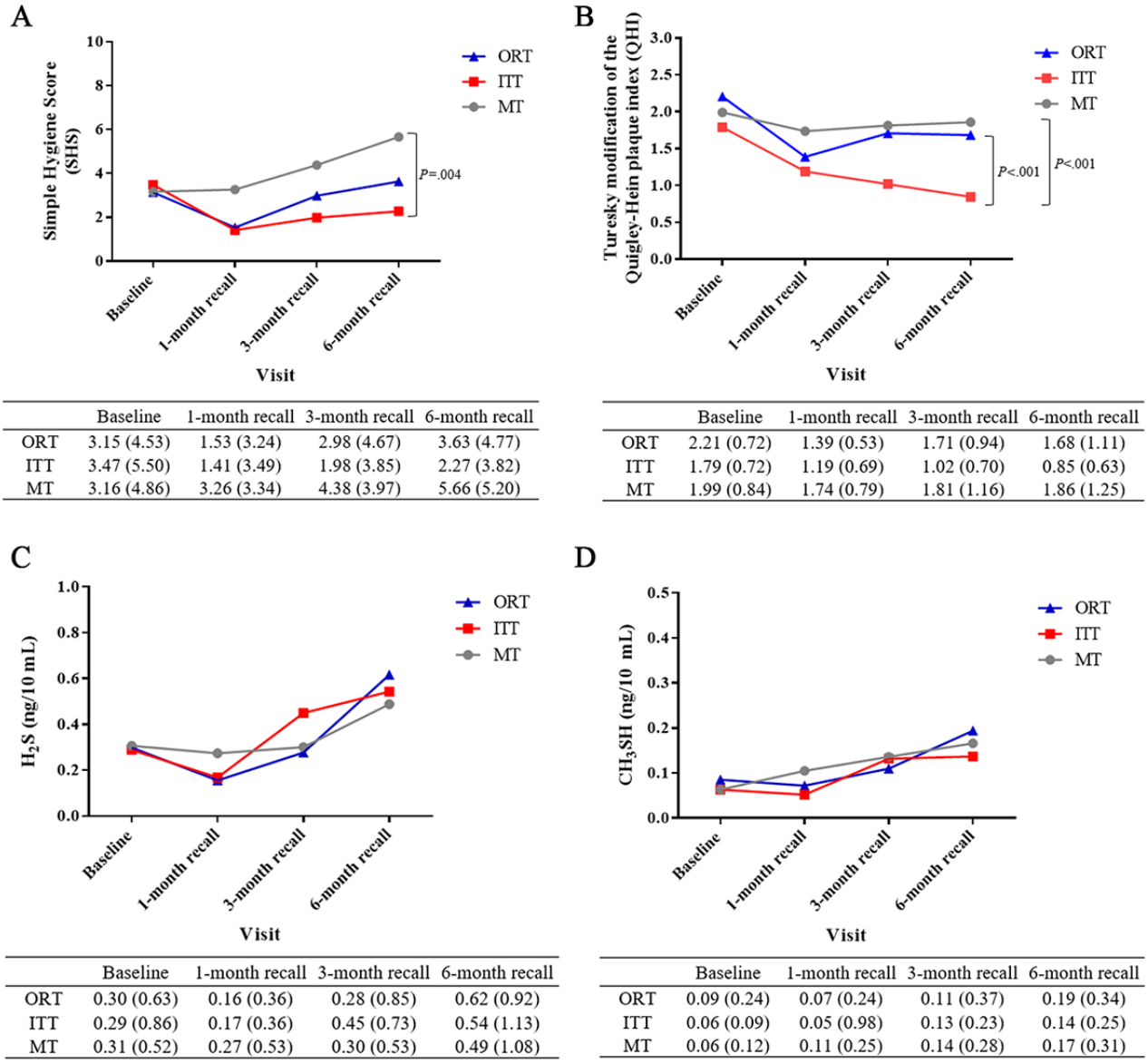

	Baseline	1-month recall	3-month recall	6-month recall
ORT	3.15 (4.53)	1.53 (3.24)	2.98 (4.67)	3.63 (4.77)
ITT	3.47 (5.50)	1.41 (3.49)	1.98 (3.85)	2.27 (3.82)
MT	3.16 (4.86)	3.26 (3.34)	4.38 (3.97)	5.66 (5.20)

	Baseline	1-month recall	3-month recall	6-month recall
ORT	2.21 (0.72)	1.39 (0.53)	1.71 (0.94)	1.68 (1.11)
ITT	1.79 (0.72)	1.19 (0.69)	1.02 (0.70)	0.85 (0.63)
MT	1.99 (0.84)	1.74 (0.79)	1.81 (1.16)	1.86 (1.25)

	Baseline	1-month recall	3-month recall	6-month recall
ORT	0.30 (0.63)	0.16 (0.36)	0.28 (0.85)	0.62 (0.92)
ITT	0.29 (0.86)	0.17 (0.36)	0.45 (0.73)	0.54 (1.13)
MT	0.31 (0.52)	0.27 (0.53)	0.30 (0.53)	0.49 (1.08)

	Baseline	1-month recall	3-month recall	6-month recall
ORT	0.09 (0.24)	0.07 (0.24)	0.11 (0.37)	0.19 (0.34)
ITT	0.06 (0.09)	0.05 (0.98)	0.13 (0.23)	0.14 (0.25)
MT	0.06 (0.12)	0.11 (0.25)	0.14 (0.28)	0.17 (0.31)

Microbiota Changes

Regarding the microbiota, there were no significant differences in the high-risk periodontal microbiota over time among the 3 groups (Figures 4A-4C). Likewise, there was no significant difference in the ratio of caries-risk microbiota to anticaries microbiota between the ORT and ITT groups (Figures 4D and 4E). However, in the MT group, the ratio of caries-risk microbiota significantly increased at the 3-month (*P*<.001) and 6-month (*P*=.005) recalls compared with baseline and at the 3-month (*P*=.048) and 6-month (*P*=.03) recalls compared with the 1-month recall. Conversely, the ratio of anticaries microbiota significantly decreased at the 3-month and 6-month recalls (Figure 4F).

XSL•FO
RenderX

Figure 4. At each of the study time points, boxplots of the percentages of periodontal high-risk, caries-related microbiota in the (A) oscillating-rotating power toothbrush (ORT), (B) interactive telemonitoring toothbrush (ITT), and (C) manual toothbrush (MT) groups, with the boxes representing the 25th to 75th percentiles and the line representing the median, as well as the ratio of caries-risk to anticaries microbiota in the (D) ORT, (E) ITT, and (F) MT groups at the different time points, with changes in the ratios over time analyzed using paired-samples t tests.

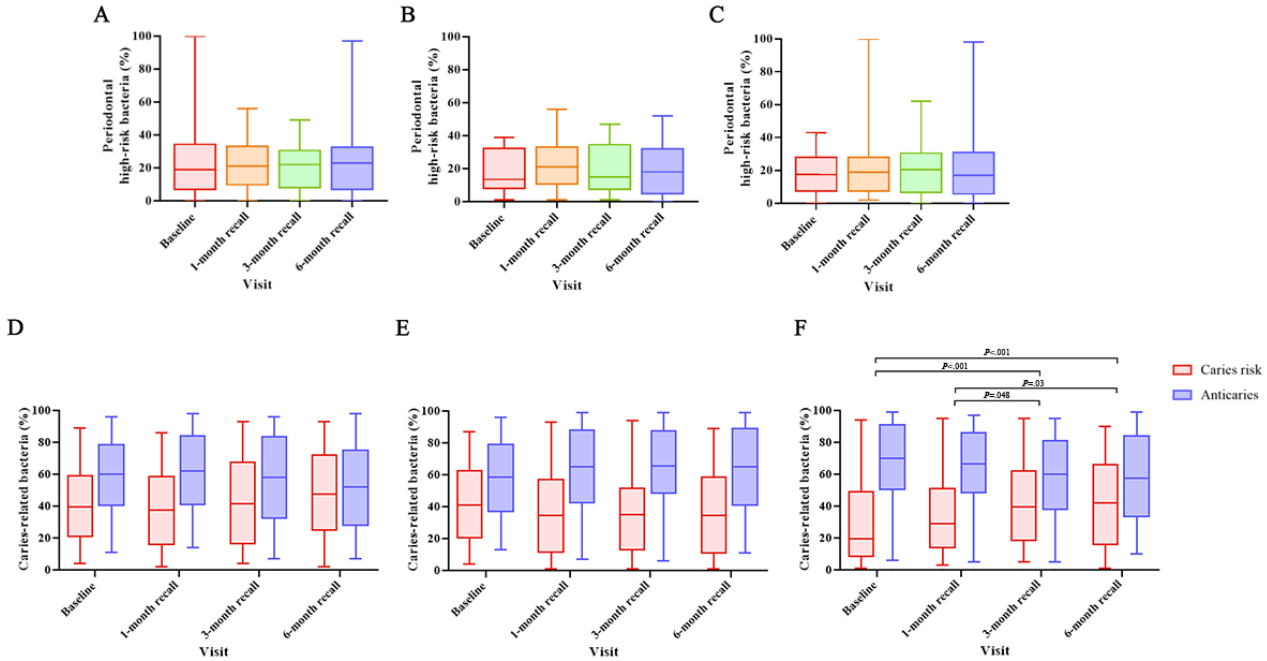

Self-Reported Questionnaire and Toothbrushing Behavior Changes

The groups completed a self-reported questionnaire on current oral care habits and awareness of telemonitoring toothbrushes after finishing the study. On the questionnaire, the most common answers to the question regarding the most effective function of a telemonitoring toothbrush were the brushing guide function by the ORT and MT groups and the standard brushing training by the ITT group (P<.001; Table 2). The most inconvenient feature of the telemonitoring toothbrush was inaccurate toothbrush position recognition, according to participants in the ORT and ITT groups (P<.001). In the ORT and ITT groups, 88% (44/50) and 82% (41/50) of the participants, respectively, indicated that the smart toothbrush was helpful for oral health, and 76% (38/50) and 92% (46/50) of the participants, respectively, reported that their brushing ability had improved. Mean satisfaction with the smart toothbrush was 7.90 (SD 1.21) and 7.11 (SD 1.67) in the ORT and ITT groups, respectively; these were significantly lower in the ITT group than in the ORT group (P=.008). The need for a smart toothbrush was not significantly different between the ORT (mean 7.44, SD 1.64) and ITT (mean 7.71, SD 1.67) groups.

XSL•FO
RenderX

Table 2. Postquestionnaire responses on oral care habits and awareness of telemonitoring toothbrushes.

Questions	ORT[a] (n=50)	ITT[b] (n=50)	MT[c] (n=50)	*P* value[d]
Number of toothbrushing events per day, mean (SD)	2.36 (0.60)	2.60 (0.61)	2.62 (0.57)	.053
Length of teethbrushing per toothbrushing event (minutes), mean (SD)	3.04 (0.49)	3.28 (0.67)	3.46 (0.84)	.01
What are the most valid functions of a telemonitoring toothbrush?, n (%)				<.001
Function of guiding brushing	22 (44)	17 (34)	27 (55)	
Management of brushing history	20 (40)	6 (12)	9 (18)	
Standard brushing training	7 (14)	25 (50)	12 (25)	
Other	1 (2)	2 (4)	1 (2)	
What is the most inconvenient thing about a telemonitoring toothbrush?, n (%)				<.001
Adjusting smartphone position	9 (18)	8 (16)	24 (48)	
Inaccurate toothbrush position recognition	39 (78)	34 (68)	16 (32)	
Signing up	0	1 (2)	5 (10)	
Management of brushing history	0	0	4 (8)	
Other	2 (4)	7 (14)	1 (2)	
Do you think a telemonitoring toothbrush can benefit your oral health?, n (%)				.66
Strongly agree	10 (20)	6 (12)	6 (12)	
Agree	34 (68)	35 (70)	32 (64)	
Neutral	6 (12)	8 (16)	11 (22)	
Disagree	0	1 (2)	1 (2)	
Strongly disagree	0	0	0	
Do you think your brushing ability is better than before?, n (%)				.12
Strongly agree	11 (22)	9 (18)	—[e]	
Agree	27 (54)	37 (74)	—	
Neutral	9 (18)	3 (6)	—	
Disagree	3 (6)	1 (2)	—	
Strongly disagree	0	0	—	
Are you willing to buy a telemonitoring toothbrush and use it continuously?, n (%)				.002
Strongly agree	14 (28)	3 (6)	2 (4)	
Agree	21 (42)	17 (34)	24 (48)	
Neutral	8 (16)	24 (48)	18 (36)	
Disagree	6 (12)	4 (8)	5 (10)	
Strongly disagree	1 (2)	2 (4)	1 (2)	
If you could receive oral care services at a general dentist as a result of using a telemonitoring toothbrush, would you visit that dentist?, n (%)				.17
Strongly agree	6 (12)	6 (12)	7 (14)	
Agree	24 (48)	33 (66)	36 (72)	
Neutral	15 (30)	10 (20)	5 (10)	
Disagree	4 (8)	1 (2)	2 (4)	
Strongly disagree	1 (2.0)	0	0	
Satisfaction with the telemonitoring toothbrush (1 to 10), mean (SD)	7.90 (1.21)	7.11 (1.67)	0	.008
Need for a telemonitoring toothbrush (1 to 10), mean (SD)	7.44 (1.64)	7.71 (1.67)	0	.42

[a]ORT: oscillating-rotating power toothbrush.

XSL•FO
RenderX

[b]ITT: interactive telemonitoring toothbrush.

[c]MT: manual toothbrush.

[d]1-way ANOVA for continuous variables and chi-square test categorical variables.

[e]Not applicable.

When comparing the average number of toothbrushing events and minutes per toothbrushing event by group, at baseline and after study completion, the number of toothbrushing events decreased significantly in the ORT group (*P*=.02) and slightly in the ITT and MT groups (Table 3). The average number of minutes spent per toothbrushing event increased in all 3 groups, especially in the MT group (*P*=.01).

Table 3. Comparison of changes in brushing habits by group.

Variable	Before (baseline)	After (6-month recall)	*P* value[a]
Number of toothbrushing events, mean (SD)			
ORT[b] (n=50)	2.54 (0.61)	2.36 (0.60)	.02
ITT[c] (n=50)	2.72 (0.61)	2.60 (0.61)	.057
MT[d] (n=50)	2.70 (0.61)	2.62 (0.57)	.32
Length of each toothbrushing event (minutes), mean (SD)			
ORT (n=50)	2.86 (0.76)	3.04 (0.49)	.12
ITT (n=50)	3.10 (0.81)	3.28 (0.67)	.06
MT (n=50)	3.12 (0.82)	3.46 (0.84)	.01

[a]*t* test.

[b]ORT: oscillating-rotating power toothbrush.

[c]ITT: interactive telemonitoring toothbrush.

[d]MT: manual toothbrush.

Discussion

Principal Findings

This study demonstrated that ITTs could improve oral hygiene management. The SHS and QHI, indicators of oral hygiene, improved significantly more for the ITT group than the ORT and MT groups when the values at baseline and the 6-month recall were compared between the groups. Along with these results, changes in the oral microbiota were also confirmed. We classified *Streptococcus sanguinis* as anticaries microbiota and compared its quantitative ratio against caries-risk microbiota, such as *Streptococcus mutans*, *Streptococcus sobrinus*, *Actinomyces gerencseriae*, *Scardovia wiggsiae*, *Veillonella parvula*, and *Candida albicans*. In the ITT group, the proportion of anticaries microbiota increased over time. In the ORT group, the proportion of anticaries microbiota increased until the 3-month recall; however, it decreased at the 6-month recall. In the MT group, the proportion of anticaries microbiota continued to decrease at all time points. Although reducing caries-risk microbiota is important, increasing anticaries microbiota is also an important part of oral hygiene [17,18]. These findings highlight the importance of compliance and sustained engagement in oral hygiene maintenance, which was facilitated by continuous feedback in the ITT group.

Comparison With Prior Work

The importance of compliance effects for smartphone-based digital health care devices has been reported in previous studies. For example, a telemonitoring application for blood glucose management by individuals with diabetes was effective at the beginning of use, but as satisfaction with the application decreased, the effectiveness also decreased [19]. Using telemonitoring to control asthma has also been reported to be effective. However, a higher level of outpatient care was received compared with national averages [20]; therefore, we cannot rule out the possibility that this may have had an impact on compliance with continued application use.

Based on the questionnaire administered before and after the study, 83% of the participants in the ORT and ITT groups who used smart toothbrushes responded that their brushing abilities improved. Among the 2 groups using smart toothbrushes, the SHS and QHI improved more in the ITT group, whereas satisfaction with the smart toothbrush was higher in the ORT group. This might be due to Oral-B's fancy hardware and application appearance and easy-to-use charging method compared with the ITT.

In the 2000s, an interactive toothbrush equipped with a monitoring function using toothbrushing and grip axis recognition was introduced [21]. With the advancement in modern technology, it has become possible to provide instructional brushing videos on smartphones and app-based brushing monitoring using real-time motion recognition via Bluetooth [22]. In addition, messages and chat apps are known to help improve brushing, and recently released smart toothbrushes are equipped with various functions, such as messages and motion recognition. An interactive toothbrush with a smartphone app has the advantage of recording toothbrushing data, enabling dental specialists to provide personalized feedback and improve habits through toothbrushing

XSL•FO
RenderX

notifications [23,24]. In particular, these findings suggest that children and adolescents who use an interactive toothbrush will benefit from plaque removal and improved gingival health [25-27]. We believe that our study is meaningful in that we have suggested an optimal oral hygiene management method for patients, by comparing clinical efficacy not only between interactive and manual toothbrushes but also between 2 different smart toothbrushes.

The medical paradigm has recently changed to include quick information use and interactive communication using smartphones, which are gradually being used for telemedicine, remote monitoring, and health intervention provision [28-30]. In dentistry, apps to improve oral hygiene are used in various ways, and many studies have verified their effectiveness [31,32]. These innovative smartphone-based mobile apps are also being used as digital therapeutics. Digital therapeutics and software medical devices that provide evidence-based therapeutic interventions to prevent, manage, and treat medical disorders or diseases have been launched in various countries [33]. Moreover, the results of this study showed that smart toothbrushing is not limited to hardware only. An interactive app that encourages the correction of toothbrushing habits has the potential to be used as a digital treatment. In addition, although the duration of and scores for toothbrushing temporarily improved immediately after recall in this study, it was confirmed that interest in brushing had decreased at the 3-month and 6-month recalls, which was probably a long recall period. This suggests that, when using a smart toothbrush, the recall interval with a dental professional should not be too long to promote effective teledentistry treatments, such as a digital therapeutic, and that patient compliance with the smart toothbrush is important.

Strengths and Limitations

Strengths

This study is one of few to comprehensively compare an ITT and ORT with an MT, evaluating their effectiveness in clinical, microbiological, and user experience dimensions. A robust randomized controlled trial design with clearly defined interventions and outcomes ensured the reliability of the findings.

Limitations

The limitations of this study are as follows. First, we restricted the use of auxiliary oral hygiene devices, such as dental floss and interdental brushes, during the study period. Although the use of auxiliary oral hygiene devices was restricted to compare the efficacy of the smart toothbrush itself, the use of auxiliary oral hygiene devices plays an important role in oral hygiene management in interdental areas [34]. Since auxiliary oral hygiene devices are highly recommended in clinical practice, it is necessary to evaluate plaque removal and gingival health improvement when using auxiliary oral hygiene devices together with a smart toothbrush. Second, the nutrition practices of the participants were not recorded. Nutrition practices could affect the oral environment in addition to brushing, but this was not considered in the result analysis. Third, right-handedness and left-handedness were not considered. Since the app showed a video guide on how to brush based on right-handedness, it could have been confusing for left-handed participants. Finally, participants were recruited from specific higher education institutions, resulting in a younger demographic. Additionally, as participants responded to a recruitment posting, they may have been more motivated or interested in oral health than the general population, potentially limiting the generalizability of the findings.

Conclusions

The ITT group had significant improvements on the SHS and QHI, demonstrating that telemonitoring toothbrushes enable proper oral hygiene management compared with manual and oscillating-rotating toothbrushes. This highlights the importance of expert feedback through the transmission of users' toothbrushing data. However, it was confirmed that interest in toothbrushing declined at the 3-month and 6-month recalls, which were long recall periods. In the future, additional research will be needed on the appropriate feedback cycle and compliance with the app by participants.

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI20C1055).

All authors concur that neither ChatGPT nor any generative AI was used during writing the manuscript.

Data Availability

Data that support the findings of this study are available from the corresponding authors upon request and following institutional review board rules and privacy regulations.

Authors' Contributions

JK, YC, YS, and WP jointly designed and conducted the research. JK and YC contributed to drafting the manuscript, creating the figures, data analysis, and data interpretation. WP reviewed the final manuscript. All authors read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1002 KB - jmir_v27i1e65128_app1.pdf]

Multimedia Appendix 2
Example of Mombrush ProCare application with Bluetooth connection to Mombrush, an interactive telemonitoring toothbrush.
[PDF File (Adobe PDF File), 215 KB - jmir_v27i1e65128_app2.pdf]

Multimedia Appendix 3
Example of the Oral-B mobile application connecting via Bluetooth to the Oral-B Genius 8000 oscillating-rotating electric toothbrush.
[PDF File (Adobe PDF File), 136 KB - jmir_v27i1e65128_app3.pdf]

References

1. Baehni PC. Translating science into action--prevention of periodontal disease at patient level. Periodontol 2000 2012 Oct 22;60(1):162-172. [doi: 10.1111/j.1600-0757.2011.00428.x] [Medline: 22909114]
2. Joshi S, Suominen AL, Knuuttila M, Bernabé E. Toothbrushing behaviour and periodontal pocketing: An 11-year longitudinal study. J Clin Periodontol 2018 Feb 13;45(2):196-203. [doi: 10.1111/jcpe.12844] [Medline: 29178189]
3. Van der Weijden FA, Slot DE. Efficacy of homecare regimens for mechanical plaque removal in managing gingivitis a meta review. J Clin Periodontol 2015 Apr 31;42 Suppl 16(S16):S77-S91. [doi: 10.1111/jcpe.12359] [Medline: 25597787]
4. Bellamy P, Barlow A, Puri G, Wright KIT, Mussett A, Zhou X. A new in vivo interdental sampling method comparing a daily flossing regime versus a manual brush control. J Clin Dent 2004;15(3):59-65. [Medline: 15688960]
5. Bilen YZ, Çokakoğlu S, Öztürk F. The short-term effects of manual and interactive powered toothbrushes on the periodontal status of orthodontic patients: A randomized clinical trial. J World Fed Orthod 2021 Mar;10(1):14-19. [doi: 10.1016/j.ejwf.2020.11.003] [Medline: 33384265]
6. Mohammed H, Rizk MZ, Wafaie K, Ulhaq A, Almuzian M. Reminders improve oral hygiene and adherence to appointments in orthodontic patients: a systematic review and meta-analysis. Eur J Orthod 2019 Mar 29;41(2):204-213. [doi: 10.1093/ejo/cjy045] [Medline: 29947755]
7. Grender J, Williams K, Walters P, Klukowska M, Reick H. Plaque removal efficacy of oscillating-rotating power toothbrushes: review of six comparative clinical trials. Am J Dent 2013 Apr;26(2):68-74. [Medline: 24073528]
8. Erbe C, Jacobs C, Klukowska M, Timm H, Grender J, Wehrbein H. A randomized clinical trial to evaluate the plaque removal efficacy of an oscillating-rotating toothbrush versus a sonic toothbrush in orthodontic patients using digital imaging analysis of the anterior dentition. Angle Orthod 2019 May;89(3):385-390 [FREE Full text] [doi: 10.2319/080317-520.1] [Medline: 30516414]
9. Doğan MC, Alaçam A, Aşici N, Odabaş M, Seydaoğlu G. Clinical evaluation of the plaque-removing ability of three different toothbrushes in a mentally disabled group. Acta Odontol Scand 2004 Dec 02;62(6):350-354. [doi: 10.1080/00016350410010054] [Medline: 15848980]
10. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
11. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. Journal of Periodontology 2018 Jun 21;89(S1):1. [doi: 10.1002/jper.18-0006]
12. Cohen J. Statistical power analysis for the behavioral sciences. New York, NY: Routledge; 2013.
13. Park S, Kahharova D, Lee J, Lee E, de Josselin de Jong E, Khudanov B, et al. Clinical assessment of an automated fluorescent plaque index scoring with quantitative light-induced fluorescence. Photodiagnosis Photodyn Ther 2020 Dec;32:102011. [doi: 10.1016/j.pdpdt.2020.102011] [Medline: 32949788]
14. Turesky S, Gilmore ND, Glickman I. Reduced plaque formation by the chloromethyl analogue of victamine C. J Periodontol 1970 Jan;41(1):41-43. [doi: 10.1902/jop.1970.41.41.41] [Medline: 5264376]
15. Quigley GA, Hein JW. Comparative cleansing efficiency of manual and power brushing. The Journal of the American Dental Association 1962 Jul;65(1):26-29. [doi: 10.14219/jada.archive.1962.0184]
16. Bhaumik D, Manikandan D, Foxman B. Cariogenic and oral health taxa in the oral cavity among children and adults: A scoping review. Arch Oral Biol 2021 Sep;129:105204 [FREE Full text] [doi: 10.1016/j.archoralbio.2021.105204] [Medline: 34246103]

XSL•FO
RenderX

17. Chen X, Daliri EB, Kim N, Kim J, Yoo D, Oh D. Microbial etiology and prevention of dental caries: exploiting natural products to inhibit cariogenic biofilms. Pathogens 2020 Jul 14;9(7):569 [FREE Full text] [doi: 10.3390/pathogens9070569] [Medline: 32674310]
18. Zhu B, Macleod LC, Kitten T, Xu P. Streptococcus sanguinis biofilm formation and interaction with oral pathogens. Future Microbiol 2018 Jun 01;13(8):915-932 [FREE Full text] [doi: 10.2217/fmb-2018-0043] [Medline: 29882414]
19. Kim H, Choi W, Baek EK, Kim YA, Yang SJ, Choi IY, et al. Efficacy of the smartphone-based glucose management application stratified by user satisfaction. Diabetes Metab J 2014 Jun;38(3):204-210 [FREE Full text] [doi: 10.4093/dmj.2014.38.3.204] [Medline: 25003074]
20. Merchant RK, Inamdar R, Quade RC. Effectiveness of population health management using the Propeller Health Asthma Platform: a randomized clinical trial. J Allergy Clin Immunol Pract 2016 May;4(3):455-463. [doi: 10.1016/j.jaip.2015.11.022] [Medline: 26778246]
21. Lee J, Lee K, Kim K, Kim D, Kim K. Development of smart toothbrush monitoring system for ubiquitous healthcare. Conf Proc IEEE Eng Med Biol Soc 2006;2006:6422-6425. [doi: 10.1109/IEMBS.2006.259232] [Medline: 17945966]
22. Jeong J, Kim K, Lee J, Kim K, Park W. Efficacy of tooth brushing via a three-dimensional motion tracking system for dental plaque control in school children: a randomized controlled clinical trial. BMC Oral Health 2022 Dec 22;22(1):626 [FREE Full text] [doi: 10.1186/s12903-022-02665-6] [Medline: 36550451]
23. Zotti F, Dalessandri D, Salgarello S, Piancino M, Bonetti S, Visconti L, et al. Usefulness of an app in improving oral hygiene compliance in adolescent orthodontic patients. Angle Orthod 2016 Jan;86(1):101-107 [FREE Full text] [doi: 10.2319/010915-19.1] [Medline: 25799001]
24. Alkadhi OH, Zahid MN, Almanea RS, Althaqeb HK, Alharbi TH, Ajwa NM. The effect of using mobile applications for improving oral hygiene in patients with orthodontic fixed appliances: a randomised controlled trial. J Orthod 2017 Sep 13;44(3):157-163. [doi: 10.1080/14653125.2017.1346746] [Medline: 28705122]
25. Erbe C, Klees V, Braunbeck F, Ferrari-Peron P, Ccahuana-Vasquez RA, Timm H, et al. Comparative assessment of plaque removal and motivation between a manual toothbrush and an interactive power toothbrush in adolescents with fixed orthodontic appliances: A single-center, examiner-blind randomized controlled trial. Am J Orthod Dentofacial Orthop 2019 Apr;155(4):462-472 [FREE Full text] [doi: 10.1016/j.ajodo.2018.12.013] [Medline: 30935601]
26. Underwood B, Birdsall J, Kay E. The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br Dent J 2015 Aug 28;219(4):E2-E2. [doi: 10.1038/sj.bdj.2015.660] [Medline: 26315196]
27. Lee J, Lee T, Jung H, Park W, Song JS. Effectiveness of an oral health education program using a smart toothbrush with quantitative light-induced fluorescence technology in children. Children (Basel) 2023 Feb 22;10(3):429 [FREE Full text] [doi: 10.3390/children10030429] [Medline: 36979987]
28. Blake H. Innovation in practice: mobile phone technology in patient care. Br J Community Nurs 2008 Apr;13(4):160, 162-160, 165 [FREE Full text] [doi: 10.12968/bjcn.2008.13.4.29024] [Medline: 18595303]
29. Kratzke C, Cox C. Smartphone technology and apps: rapidly changing health promotion. International Electronic Journal of Health Education 2012;15(1):72-82 [FREE Full text]
30. Sangalli L, Savoldi F, Dalessandri D, Bonetti S, Gu M, Signoroni A, et al. Effects of remote digital monitoring on oral hygiene of orthodontic patients: a prospective study. BMC Oral Health 2021 Sep 07;21(1):435 [FREE Full text] [doi: 10.1186/s12903-021-01793-9] [Medline: 34493255]
31. Lopes Dos Santos R, Spinola MDS, Carvalho E, Lopes Dos Santos DC, Dame-Teixeira N, Heller D. Effectiveness of a new app in improving oral hygiene in orthodontic patients: a pilot study. Int Dent J 2023 Jun;73(3):395-402 [FREE Full text] [doi: 10.1016/j.identj.2022.08.010] [Medline: 36192223]
32. Scheerman JFM, van Meijel B, van Empelen P, Verrips GHW, van Loveren C, Twisk JWR, et al. The effect of using a mobile application ("WhiteTeeth") on improving oral hygiene: A randomized controlled trial. Int J Dent Hyg 2020 Feb 09;18(1):73-83 [FREE Full text] [doi: 10.1111/idh.12415] [Medline: 31291683]
33. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a Web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res 2015 Apr 10;17(4):e92 [FREE Full text] [doi: 10.2196/jmir.4052] [Medline: 25863515]
34. Worthington HV, MacDonald L, Poklepovic Pericic T, Sambunjak D, Johnson TM, Imai P, et al. Home use of interdental cleaning devices, in addition to toothbrushing, for preventing and controlling periodontal diseases and dental caries. Cochrane Database Syst Rev 2019 Apr 10;4(4):CD012018 [FREE Full text] [doi: 10.1002/14651858.CD012018.pub2] [Medline: 30968949]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
IRB: Institutional Review Board
ITT: interactive telemonitoring toothbrush
MT: manual toothbrush

XSL•FO
RenderX

ORT: oscillating-rotating power toothbrush
PCR: polymerase chain reaction
QHI: Turesky modification of the Quigley-Hein plaque index
SHS: Simple Hygiene Score

Edited by N Cahill; submitted 06.08.24; peer-reviewed by JH Kim, AN Ali, T Tuzuner; comments to author 23.10.24; revised version received 09.11.24; accepted 30.12.24; published 10.02.25.

Please cite as:
Kim J, Choi Y, Song Y, Park W
Smartphone-Based Telemonitoring for Better Oral Health With Toothbrushes: 6-Month Randomized Controlled Trial
J Med Internet Res 2025;27:e65128
URL: https://www.jmir.org/2025/1/e65128
doi:10.2196/65128
PMID:

©Jaeyeon Kim, Yiseul Choi, Yoolbin Song, Wonse Park. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smart Pharmaceutical Monitoring System With Personalized Medication Schedules and Self-Management Programs for Patients With Diabetes: Development and Evaluation Study

Jian Xiao[1,2*], Prof Dr; Mengyao Li[1,3*], MM; Ruwen Cai[1,3], MM; Hangxing Huang[1,4], PhD; Huimin Yu[1,2], MM; Ling Huang[1,2], MM; Jingyang Li[1,2], MM; Ting Yu[1,3], MM; Jiani Zhang[5], Prof Dr; Shuqiao Cheng[1,2], MM

[1]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China

[2]The Hunan Institute of Pharmacy Practice and Clinical Research, Changsha, China

[3]Department of Pharmacy, Dali University, Dali, China

[4]School of Medicine, Wuhan University of Science and Technology, Wuhan, China

[5]Department of Senile Endocrinology, Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Shuqiao Cheng, MM
Department of Pharmacy
Xiangya Hospital
Central South University
87 Xiangya Rd
Changsha, 410008
China
Phone: 86 15974198203
Email: cheng0203@csu.edu.cn

Abstract

Background: With the climbing incidence of type 2 diabetes, the health care system is under pressure to manage patients with this condition properly. Particularly, pharmacological therapy constitutes the most fundamental means of controlling blood glucose levels and preventing the progression of complications. However, its effectiveness is often hindered by factors such as treatment complexity, polypharmacy, and poor patient adherence. As new technologies, artificial intelligence and digital technologies are covering all aspects of the medical and health care field, but their application and evaluation in the domain of diabetes research remain limited.

Objective: This study aims to develop and establish a stand-alone diabetes management service system designed to enhance self-management support for patients, as well as to assess its performance with experienced health care professionals.

Methods: Diabetes Universal Medication Schedule (DUMS) system is grounded in official medicine instructions and evidence-based data to establish medication constraints and drug-drug interaction profiles. Individualized medication schedules and self-management programs were generated based on patient-specific conditions and needs, using an app framework to build patient-side contact pathways. The system's ability to provide medication guidance and health management was assessed by senior health care professionals using a 5-point Likert scale across 3 groups: outputs generated by the system (DUMS group), outputs refined by pharmacists (intervention group), and outputs generated by ChatGPT-4 (GPT-4 group).

Results: We constructed a cloud-based drug information management system loaded with 475 diabetes treatment–related medications; 684 medication constraints; and 12,351 drug-drug interactions and theoretical supports. The generated personalized medication plan and self-management program included recommended dosing times, disease education, dietary considerations, and lifestyle recommendations to help patients with diabetes achieve correct medication use and active disease management. Reliability analysis demonstrated that the DUMS group outperformed the GPT-4 group in medication schedule accuracy and safety, as well as comprehensiveness and richness of the self-management program (P<.001). The intervention group outperformed the DUMS and GPT-4 groups on all indicator scores.

XSL•FO
RenderX

Conclusions: DUMS's treatment monitoring service can provide reliable self-management support for patients with diabetes. ChatGPT-4, powered by artificial intelligence, can act as a collaborative assistant to health care professionals in clinical contexts, although its performance still requires further training and optimization.

(J Med Internet Res 2025;27:e56737) doi:10.2196/56737

KEYWORDS

pharmaceutical services; diabetes; self-management; intelligent medication scheduling system; drug database; GPT-4

Introduction

Chronic diseases are prevalent among the older adult population [1]. Diabetes, a common chronic disease, imposes a significant self-management burden on individuals and their families. Individuals with type 2 diabetes are increasingly being treated with multiple-medication regimens, including oral medications and/or exogenous insulin. Long-term medication and lifestyle changes are necessary for the successful management of both type 1 and type 2 diabetes [2]. However, the low success rate of diabetes self-management reflects some of the burdensome drug regimens and side effects [3]. A study pointed out that approximately 43.4% of patients with diabetes in low- and middle-income countries fail to adhere to their prescribed treatments, resulting in adverse health outcomes [4,5]. National survey data from the United States report basic knowledge gaps for many adults with diabetes, wherein only less than 50% are aware of their glycemic control levels. Guidelines emphasize that all patients should receive diabetes self-management education and support (DSMES), including regular medication taking, self-monitoring their blood glucose, changes in diet and physical exercise, foot self-care, and routine check-ups [2,6,7].

Nowadays, many apps and smart devices are being developed to enable patients to take their medications correctly and on time [8-13]. Some types of medicine boxes have also been proposed, but due to the complexity of filling, large volume, and lack of portability, most are not suitable for older adults [8,11,14]. In addition, producing the medicine box incurs certain costs, and many devices lack personalized medication guidance. Gupta et al [15] integrated drug names and doctor-prescribed schedules into a hardware device to provide reminders, but it lacked a friendly interface for interaction with health care professionals. Similarly, Wachira et al [16] designed and developed a guideline-encoding and execution engine using the JSON format and an optimization algorithm for medication plans through computer-aided output, yet their approach did not account for potential drug-drug interactions (DDIs) or real-world clinical data validation. In a report on the treatment plans and health status of individuals with diabetes, possible interactions were identified in 413 cases, and 53.1%, 7.8%, and 7.2% of the participants presented with moderate, minor, and serious risk of DDIs, respectively. Older adult patients are the primary users of multiple medications and consequently are more susceptible to inappropriate use, polypharmacy, and DDIs [17].

It is worth noting that generative artificial intelligence (AI)—based large language models (LLMs) have demonstrated significant potential innovation in health care in recent years. One such advanced model is ChatGPT [18], developed by OpenAI, which is currently among the largest publicly available language models [19]. ChatGPT-4 excels in a variety of medical evaluation benchmarks such as the United States Medical Licensing Examination (USMLE) [20,21] owing to its strong logical self-consistency and data analysis capabilities [22]. These LLMs leverage chat-based interfaces to respond to complex queries and solve problems [23]. Incorporating AI into the health care systems has the potential to significantly reduce task completion times and optimize workflows, thereby improving overall industry efficiency. As the availability and accessibility of these models increase, understanding their potential and limitations in real-world applications is essential. Song et al [24] evaluated ChatGPT-4' s ability to provide health education to patients with urolithiasis across two clinical scenarios involving 21 questions. While the model had high accuracy, its responses lacked contextual relevance in some cases. However, the performance of generative AI in DSMES remains an area that requires further exploration and evaluation.

The study aims to develop a medication reminder system tailored for older patients with diabetes. Additionally, to explore the real-world application of this system, we compared its capabilities in providing patient medication plans and self-management programs with those of the latest publicly available version of ChatGPT. It provides reference for the auxiliary decision-making function and the ability to empower the pharmaceutical field of generative AI in medicine.

Methods

Infrastructure Work

Pharmaceutical Constraint Management

Building upon the findings from our team's earlier work [25], we developed the Diabetes Universal Medication Schedule (DUMS) system as a more comprehensive solution for diabetes management. The system was built using the Python web framework, Flask, and the Flutter framework, written in Dart and Visual Studio Code. Data management and storage were implemented using a MySQL database and a Redis cache database.

Medication data were sourced from official medicine instructions and open-source medical professional websites, including information on ingredients, specifications, dosages, indications, and others. These data formed the foundational feature modules of the system. Dietary effects, interaction data, and physiologically based pharmacokinetic analysis were performed to establish pharmacological constraint information (Figure 1).

XSL•FO
RenderX

Figure 1. The architectural diagram and functional explanation of the Diabetes Universal Medication Schedule (DUMS) system.

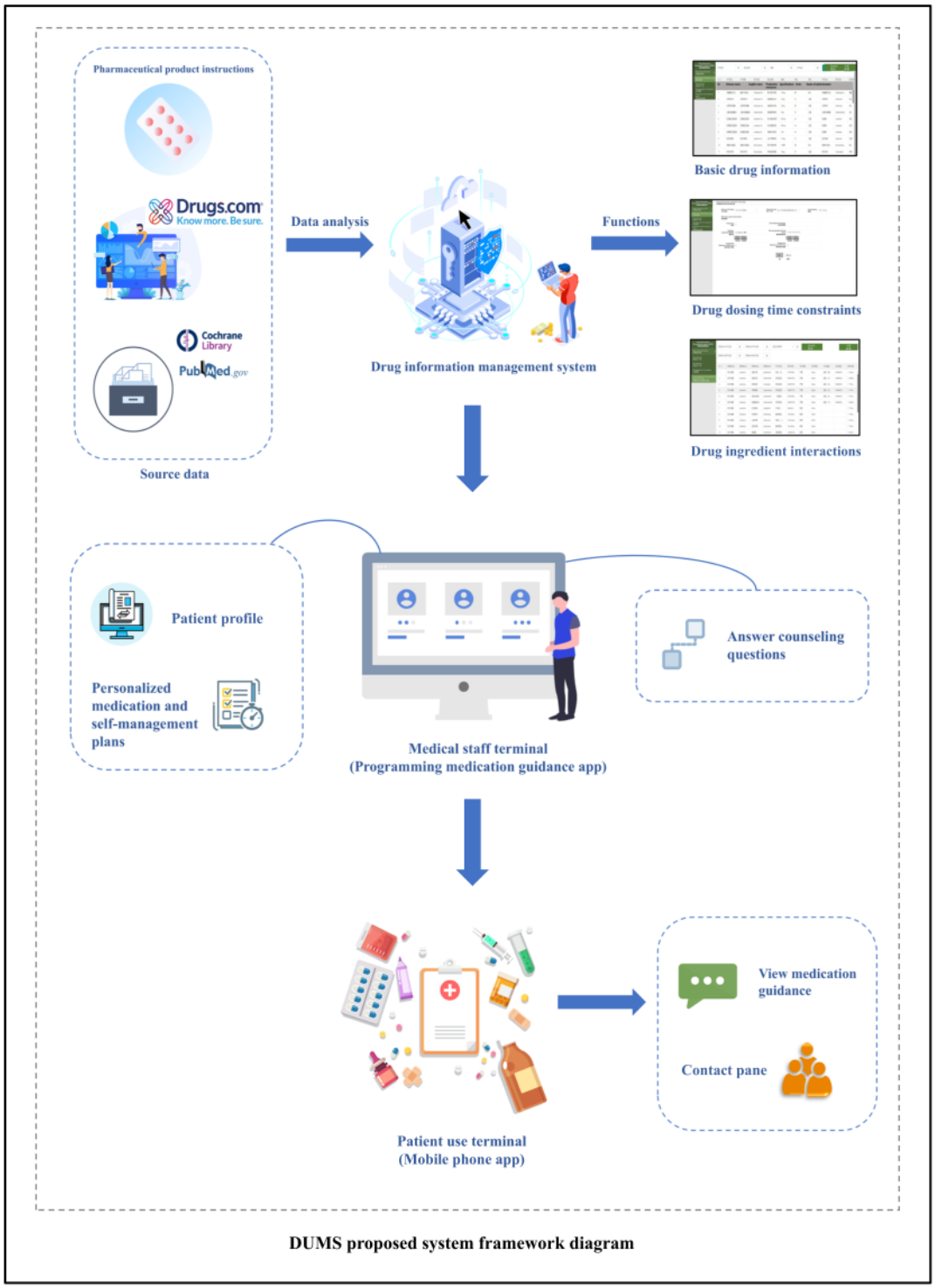

Interaction information was organized by drug active ingredient. Key data were extracted from official medicine description; the Drugs.com website; and scientific databases such as PubMed, Web of Science, and the Cochrane Library. A combination of subject-specific search terms (individual drug ingredients) and free-text queries was used to identify relevant studies on drug-dose relationships. Source data, including ingredient-ingredient expression of action, interaction severity, and treatment recommendations, were reviewed by 3 independent 5-member panels composing of physicians and pharmacists. The panels collaboratively assessed the data, ultimately confirming that the credibility of the data source and the logic of the interpretation.

Self-Management Program

Self-management is a dynamic process involving the identification of disease needs, activation of available resources, and adaptation to living with a chronic condition. Guidance and

XSL•FO
RenderX

resolution from health care providers contribute to effective self-management. In addition to the regulating diet and exercise, a comprehensive management plan is developed based on the patient's medication history, allergy history, adverse drug reaction history, and treatment adherence. Unstructured information is usually provided temporarily by health care professionals.

Reception and Interaction

The patient interface was built using the WeChat app framework featuring a responsive data binding system that synchronizes data with the user interface. This framework includes a view layer description language, a style language, and a JavaScript logic layer programming language. A mobile phone number serves as the user log-in ID to locate the medication data of the target patient.

Dependability Tests

To evaluate the DUMS system's real-world application performance, the prescription details of older patients with diabetes were extracted. After inputting the prescription details, the DUMS system extracts the established drug characteristics and constraints from the cloud-based drug database and does not adjust the output results (referred to as the "DUMS group"). Pharmacists then refined these preliminary results to create active output schemes that are better suited to clinical practice (hereinafter referred to as the "intervention group"). For comparison, the same prescription details were entered into ChatGPT-4, prompting it to generate medication schedules and self-management programs based on the drug label and relevant literature guidelines, where appropriate (hereafter referred to as the "GPT-4 group"; refer to Table S1 in Multimedia Appendix 1 for details).

Health care experts with professional titles at the associate senior level or above were invited to participate in the evaluation. They rated the results from the DUMS, intervention, and GPT-4 groups using a 5-point Likert scale following a horizontal comparison approach. Table 1 outlines the assessment indicators and their corresponding scoring criteria.

Table 1. Evaluation indicators and definitions.

Classification and indicators	Definitions
Medication schedule	
Accuracy	Ensures that the use of medications aligns with mealtimes, drug-drug interactions, pharmacological timing, and daily routines
Safety	When scheduling medication times, fully consider potential drug-drug interactions
Adherence	Takes into account the convenience, rationality, and feasibility of medication
Usefulness	It can effectively reduce the workload of health care workers and improve work efficiency
Self-management program	
Accuracy	The content is accurate and error-free
Comprehensiveness	Provides all necessary information for patients regarding medication, including usage precautions, identification and prevention of adverse reactions, and dietary therapy
Richness	Provides information resources that exceed expectations
Redundancy	Evaluates whether the content contains excessive, repetitive, or useless information; a score of 5 can be given if there is no excess or useless content
Usefulness	Effectively reduces the burden on health care workers in education and public awareness efforts

Statistical Analysis

SPSS (version 27.0; IBM Corp) was used to statistically analyze the data. Measurement information was expressed as mean (SD). Independent-sample ANOVA was used, followed by the least significant difference test for multiple comparison. The expert opinion harmonization coefficient (w) was calculated using Kendall W, with differences assessed through the chi-square (χ^2) test. A $P<.05$ difference was considered statistically significant. Furthermore, the Bland-Altman method was used to generate scatter plots for pairwise comparisons of the 3 methods and to assess the degree of consistency.

Ethical Considerations

This study was reviewed and approved by the Hospital Ethics Committee of Xiangya Hospital, Central South University (approval 2023030556). This study exclusively used prescribing information for test evaluation. There was no direct interaction with patients, in line with the principles of ethical conduct in the study. All data were deidentified to protect privacy and ensure confidentiality, with access restricted to authorized members of the research team.

Results

Construction of an Intelligent Medication Scheduling System

Enhancement of the DUMS System's Cloud Database

From the endocrinology medication catalog, 475 commonly used medicines for patients with diabetes, including 146 imported medicines, were loaded. Each medicine's information was summarized and organized into concise, comprehensive, and user-friendly entries based on the official instructions. Further, a total of 684 drug constraints were established for

XSL•FO
RenderX

drugs, according to clinical recommendations related to chronopharmacology and diet-drug interactions.

After eliminating duplicate data, 8921 drug components were identified, and 12,351 drug interaction records were generated. Each interaction record included triple theoretical support: mechanism of action, interaction severity, and treatment recommendations. DDIs were classified into four categories based on severity: mild (n=6564, 53.15%), moderate (n=3392, 27.46%), severe (n=1876, 15.19%), and contraindicated (n=519, 4.2%). This classification helps medical professionals make quick and informed clinical decisions (refer to Table S2 in Multimedia Appendix 1 for details).

Individualized Medication Schedules and Self-Management Programs

To facilitate timely management and personalized patient care, each patient's profile was independently loaded with essential clinical information such as blood glucose status, medical diagnoses, and basic prescription. Default mealtimes were set at 7 AM, 12 PM, and 6 PM, with bedtime set at 10 PM, although health care providers can adjust these based on patient-specific needs and system-generated suggestions. Based on in-depth analysis of patients, providing more precise drug treatment plans and enhancing patient adherence can help improve long-term disease outcomes (Figure 2).

Figure 2. Diabetes Universal Medication Schedule (DUMS) interface: (A) blood glucose report; (B) prescription review; (C) time adjustment; (D) patients and prescription.

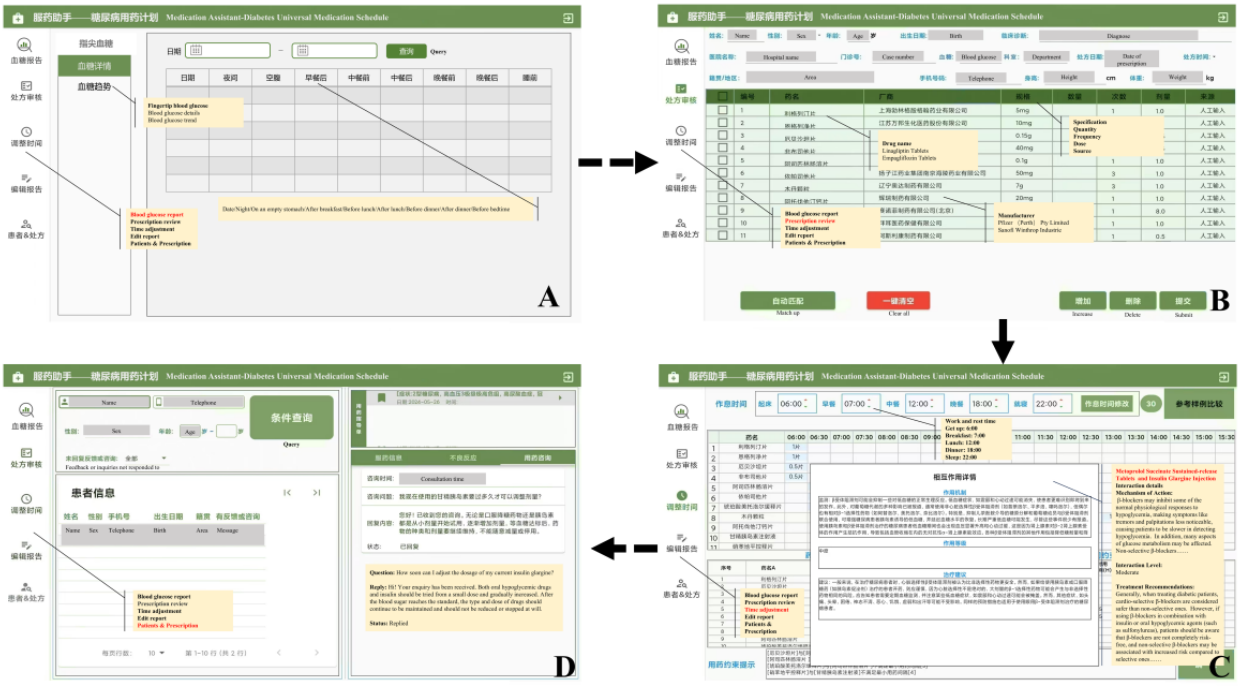

Once the medication plan is arranged, patient can access detailed schedules given by the medical staff through the user terminal. For patients with diabetes, self-management programs cover exogenous insulin use (eg, insulin pens), awareness of drug side effects, dietary guidance, and prevention and control education.

In addition, a communication module enables patients to submit inquiries regarding their current treatment status. Health care workers actively responded, reducing the likelihood of acute events while enhancing patients with diabetes' understanding of and adherence to treatment plans (Figure 3).

XSL•FO
RenderX

Figure 3. User terminal (medication assistant app): (A) time hint; (B) medication schedule; (C) self-management program; (D) consultation and settings.

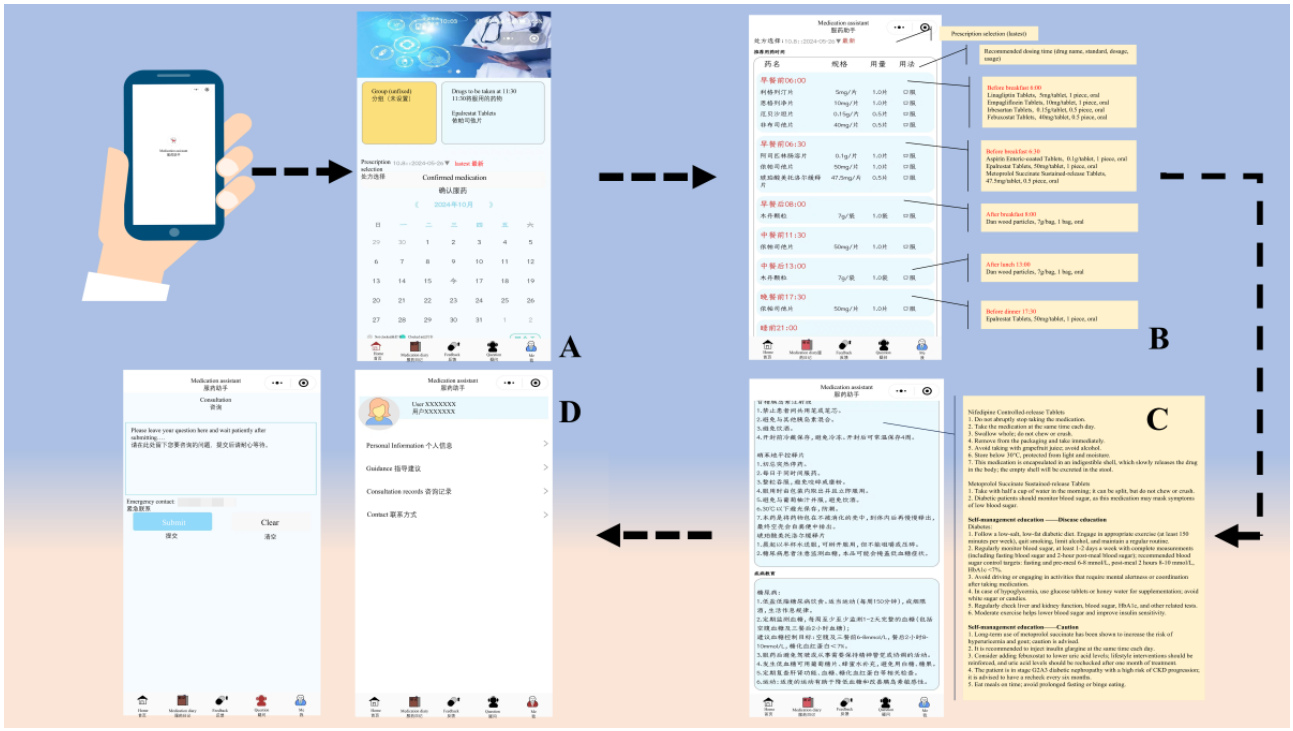

System Testing and Evaluation

A total of 12 specialists, comprising 8 general clinicians and 4 pharmacists, participated in this study, with a mean clinical experience of 3.08 (SD 0.99) years. The evaluation involved prescriptions for 20 older patients with diabetes, through 60 medication plans presented in 3 different output modes. Among the patients, more than half (n=14, 70%) were female, with an average medication frequency of 12.65 (SD 4.10) doses per day. The basic information about the patient cases is shown in Table 2.

Table 2. Patient and prescription information for the system test.

Characteristics	Value (N=20)
Sex, n (%)	
Male	6 (30)
Female	14 (70)
Age (years), mean (SD)	69.65 (4.17)
Number of medications, n (%)	
<9	15 (75)
≥9	5 (25)
Number of chronic disease comorbidities, mean (SD)	4.05 (2.04)
Medication frequency (doses per day), mean (SD)	12.65 (4.10)

The Kendall W coordination coefficient was used to calculate the agreement between raters (Kendall W for overall mark=0.21; *P*<.001), suggesting that the different scoring systems are reliable in producing reproducible results. Then, we supplemented the comparison between the two groups using the Bland-Altman method (Figure 4), revealing a mean bias between the DUMS group and the GPT-4 group as 1.57. The 95% CI for the lower limit of agreement (LoA) was –0.66 to 0.65, and the 95% CI of the upper LoA was 2.50-3.81. The mean bias between the intervention group and the DUMS group was 1.60, with 95% CI –0.65 to 0.67 for the lower LoA and 95% CI 2.54-3.86 for the upper LoA. No points of difference exceeded 95% LoAs. Comparing the intervention group to the GPT-4 group, the mean bias was 3.18. Most of the data points fell within the LoA, indicating no significant systematic bias or trend. Although some bias existed, the overall consistency was good.

XSL•FO
RenderX

Figure 4. Bland-Altman plot for agreement analysis among the DUMS, intervention, and GPT-4 groups: (A) DUMS group vs GPT-4 group; (B) intervention group vs DUMS group; (C) intervention group vs GPT-4 group. Limits of agreement are shown as dashed, purplish red lines; bias is shown as solid blue lines; equivalence (difference=0) is shown as orange dashed lines; and maximum allowed difference between the groups is shown as solid green lines. DUMS: Diabetes Universal Medication Schedule.

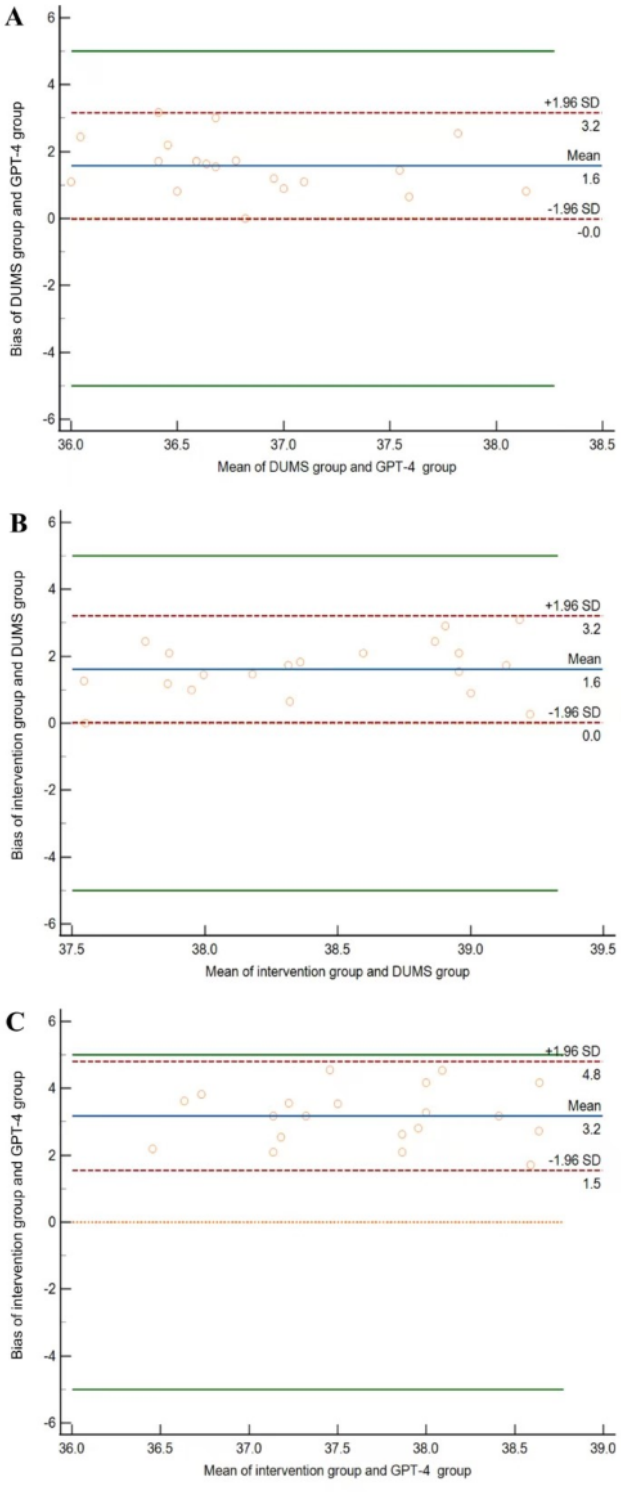

The evaluation results revealed that the DUMS group achieved significantly higher accuracy and safety scores than the GPT-4 group (P<.001), with full consideration of multiple aspects of DDI, pharmacokinetics, and dietary restrictions. However, in order to reduce the occurrence of adverse events, the medication frequency was increased, which may reduce adherence to the medication plan. Regarding medication schedule usefulness, the DUMS group scored slightly higher than the GPT-4 group, and both DUMS and ChatGPT-4 were able to provide valuable information for pharmacists to assist patients in taking effective medication, but the difference was not statistically significant (P=.23). The self-management programs provided by DUMS not only deliver easy-to-understand instructions on medications but also offer multifaceted guidance on disease education,

XSL•FO
RenderX

lifestyle, and dietary management, with positive feedback on accuracy, comprehensiveness, and richness (*P*<.001).

The total evaluation scores showed that the mean values of the DUMS group, intervention group, and GPT-4 group were 37.96 (SD 0.56), 39.52 (SD 0.80), and 36.35 (SD 0.89), respectively. The "adjusted" DUMS results (ie, the intervention group) were better than those of the other groups, with statistically significant differences (*P*<.001), as shown in Table 3.

Table 3. Results of the system testing and evaluation.

Classification and indicators	DUMS[a] group, mean (SD)	Intervention group, mean (SD)	GPT-4 group, mean (SD)	*P* value (DUMS vs intervention)	*P* value (DUMS vs GPT-4)	*P* value (intervention vs GPT-4)
Medication schedule						
Accuracy	4.43 (0.16)	4.49 (0.19)	3.84 (0.25)	.35	<.001	<.001
Safety	4.44 (0.11)	4.44 (0.16)	3.91 (0.21)	.94	<.001	<.001
Adherence	3.81 (0.34)	4.24 (0.24)	4.02 (0.25)	<.001	.03	.02
Usefulness	4.11 (0.18)	4.30 (0.16)	4.04 (0.17)	.002	.23	<.001
Self-management program						
Accuracy	4.56 (0.09)	4.64 (0.17)	4.34 (0.21)	.16	<.001	<.001
Comprehensiveness	4.08 (0.21)	4.35 (0.15)	3.86 (0.18)	.04	<.001	<.001
Richness	3.95 (0.26)	4.30 (0.24)	3.73 (0.19)	.03	<.001	<.001
Redundancy	4.34 (0.17)	4.45 (0.19)	4.42 (0.19)	.07	.55	.61
Usefulness	4.24 (0.12)	4.31 (0.13)	4.18 (0.13)	.16	.54	.04
Total	37.96 (0.56)	39.52 (0.80)	36.35 (0.89)	<.001	<.001	<.001

[a]DUMS: Diabetes Universal Medication Schedule.

Discussion

Principal Findings

Faced with the challenge of ensuring that patients with diabetes take the correct medication at the right time, the emergence of telemedicine and AI has provided researchers with new directions. In this study, we established constraint management of essential pharmaceutical products by summarizing the relationship between drugs and evidence-based arguments and created an individualized medication reminder platform.

Our system extends beyond traditional auxiliary models by offering a visually interactive personalized medication plan and a user-friendly interface. Through secondary data analysis, we establish a robust medicine constraint management and DDI risk assessment framework to guide health care providers in precision medication timing. A variety of complications in patients with diabetes often leads to complex medication management and puts patients at an increased risk of moderate to severe DDIs, which may threaten their lives or may deteriorate their quality of life [26,27]. To mitigate the above risks, a triple rationale for DDI is provided in our system, allowing health care providers to balance multiple drug demands with the possibility of harmful drug poisonings. Additionally, the system promotes DSMES by strengthening blood glucose management, ensuring patient take drugs correctly and on time, and developing exclusive planning [28].

ChatGPT is an evolving model in natural language processing, as Microsoft and OpenAI continue to develop it, and its potential applications in medical education have garnered a great deal of attention [29]. Previous studies indicate that ChatGPT-4 performs well in providing descriptive and objective content, often generating largely accurate information for a variety of medical queries [30-32]. In this study, to ensure consistency, the commands entered into ChatGPT-4 were manipulated by a single researcher, who entered the following instruction:

> *Patients (basic situation) and the prescription of the following drugs: (including drug name, dosage, frequency of use). It is assumed that the patient gets up at 6:00 every day, takes three meals at 7:00, 12:00 and 18:00 respectively, and goes to bed at 22:00. Please make a specific medication schedule and self-management plan for the patient according to the medicine instructions and relevant guidelines, and taking into account the patient's medication compliance.*

However, the accuracy, comprehensiveness, and richness of the health education component of self-management provided by ChatGPT-4 could not replace that in the DUMS group and the "experienced" intervention group. A conceivable explanation is that, within the medical field, examinations often prioritize testing theoretical knowledge, frequently deviating from genuine patient-focused clinical environments. Such deviations might be more pronounced for LLMs, which are well versed in rapidly integrating available information. Another notable limitation is ChatGPT-4's reliance on English-language training data, which affects its performance in other languages. The lack of culturally specific knowledge and the latest medical literature and data in other languages is a significant limitation of ChatGPT, which may lead to irrelevant or incorrect answers and conclusions [33,34]. Moreover, many scientific experts have pointed out that ChatGPT presents information redundantly and irrationally in most cases, offering repeated content and references, which

XSL•FO
RenderX

may not be of obvious practical utility [35,36]. Although the development of machine learning and AI methods represents a significant technological advancement, and the output of ChatGPT-4 has a certain level of practicality, achieving reliable performance in various data scenarios still requires proper training of the model algorithms to maximize efficiency.

Limitations

However, there are some limitations with the DUMS system. Its accuracy and usability rely heavily on the integrity of the drug information management database. Failure to update or expand the database promptly can result in suboptimal administration suggestions. Therefore, to continue expanding the number of drugs in the database and promptly improve the information resources, it is essential to develop a more efficient input mechanism for maintaining and managing the database, thus minimizing the risk of human errors or omissions. Additionally, comparing the performance of DUMS solely against that of ChatGPT-4 limits the generalizability of the findings to other intelligent systems or devices. Further validation is required to determine whether DUMS can perform consistently within its intended domain. Future iterations of DUMS will aim to enhance output quality by leveraging the imitative learning capabilities of LLMs, thereby improving the scientific and systematic nature of the output.

Conclusion

The DUMS system represents an innovative and cost-effective mobile health solution. Without making any changes to the existing health care system, health care institutions can deploy our proposed system to provide safer and more rational medication services for patients with diabetes. The drug information and constraint management database, combined with the user-friendly interface developed in this study, supports health care professionals in creating accurate, personalized medication schedules and self-management programs. In addition to serving as a reminder, the inquiry component allows older individuals living independently to receive timely solutions when treatment issues arise. However, it remains unclear whether mobile health can help health care workers improve patient health literacy. Significant further research will be required from managers and end users to ensure the safety, reliability, and effectiveness of this technology.

Acknowledgments

We sincerely thank the 12 medical practitioners from Xiangya Hospital of Central South University who participated in this evaluation activity at our invitation. No generative artificial intelligence was used in this manuscript.

Data Availability

The datasets analyzed during this study are not publicly available to protect the privacy of individuals who participated in the study. Data are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: RC, JX, JZ, SC
Data curation: ML, HH, HY, LH, JL, TY
Formal analysis: ML, HH
Funding acquisition: SC
Investigation: ML, HH, HY, LH, JL, TY
Methodology: JX, JZ, SC
Supervision: JX, SC
Validation: RC, JX, JZ, SC
Writing – original draft: RC, JX
Writing – review & editing: RC, JX, ML, SC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Specific content of the expert evaluation and drug-drug interaction classification criteria.
[DOCX File , 58 KB - jmir_v27i1e56737_app1.docx]

References

1. Maher RL, Hanlon J, Hajjar ER. Clinical consequences of polypharmacy in elderly. Expert Opin Drug Saf 2014 Jan;13(1):57-65 [FREE Full text] [doi: 10.1517/14740338.2013.827660] [Medline: 24073682]
2. Gonzalez JS, Tanenbaum ML, Commissariat PV. Psychosocial factors in medication adherence and diabetes self-management: Implications for research and practice. Am Psychol 2016 Oct;71(7):539-551 [FREE Full text] [doi: 10.1037/a0040388] [Medline: 27690483]

XSL•FO
RenderX

3. Yudkin JS, Richter B, Gale EAM. Intensified glucose lowering in type 2 diabetes: time for a reappraisal. Diabetologia 2010 Oct;53(10):2079-2085. [doi: 10.1007/s00125-010-1864-z] [Medline: 20686748]
4. Hassan F, Hatah E, Ali AM, Wen CW. The intervention strategies and service model for pharmacist-led diabetes management: a scoping review. BMC Health Serv Res 2023 Jan 18;23(1):46 [FREE Full text] [doi: 10.1186/s12913-022-08977-1] [Medline: 36653832]
5. Azharuddin M, Adil M, Sharma M, Gyawali B. A systematic review and meta-analysis of non-adherence to anti-diabetic medication: Evidence from low- and middle-income countries. Int J Clin Pract 2021 Nov;75(11):e14717. [doi: 10.1111/ijcp.14717] [Medline: 34378293]
6. Stark Casagrande S, Ríos Burrows N, Geiss L, Bainbridge K, Fradkin J, Cowie C. Diabetes knowledge and its relationship with achieving treatment recommendations in a national sample of people with type 2 diabetes. Diabetes Care 2012 Jul;35(7):1556-1565 [FREE Full text] [doi: 10.2337/dc11-1943] [Medline: 22498806]
7. American Diabetes Association. (1) Strategies for improving care. Diabetes Care 2015 Jan;38 Suppl:S5-S7. [doi: 10.2337/dc15-S004] [Medline: 25537709]
8. Paul LC, Ahmed SS, Rani T, Haque MA, Roy TK, Hossain MN. A smart medicine reminder kit with mobile phone calls and some health monitoring features for senior citizens. Heliyon 2024 Feb 29;10(4):e26308 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e26308] [Medline: 38404861]
9. Li L, Wang Z, Cui L, Xu Y, Lee H, Guan K. The efficacy of a novel smart watch on medicine adherence and symptom control of allergic rhinitis patients: Pilot study. World Allergy Organ J 2023 Jan;16(1):100739 [FREE Full text] [doi: 10.1016/j.waojou.2022.100739] [Medline: 36694622]
10. Chrystyn H, Audibert R, Keller M, Quaglia B, Vecellio L, Roche N. Real-life inhaler adherence and technique: Time to get smarter!. Respir Med 2019;158:24-32 [FREE Full text] [doi: 10.1016/j.rmed.2019.09.008] [Medline: 31563027]
11. Coumans C, Vandenberghe M, Fery P. High rate of errors in pillbox filling by cognitively healthy elderly people. Geriatr Psychol Neuropsychiatr Vieil 2017 Sep 01;15(3):273-282. [doi: 10.1684/pnv.2017.0679] [Medline: 28872038]
12. Rifat Kabir M, Shifat Ahmed S, Chandra Paul L, Rani T, Karaaslan M. Development of an intravenous fluid monitoring, warning, and reverse flow blocking system. 2022 Presented at: 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE); December 30-31, 2022; Naya Raipur, India p. 89-94. [doi: 10.1109/WIECON-ECE57977.2022.10150612]
13. Paul LC, Ahmed SS, Karmakar KK. A smart multi-user wireless nurse calling system and e-notice board for health care management. 2022 Presented at: Proceedings of the Third International Conference on Trends in Computational and Cognitive Engineering (TCCE 2021); 2022; Singapore. [doi: 10.1007/978-981-16-7597-3_35]
14. Browne SH, Umlauf A, Tucker AJ, Low J, Moser K, Gonzalez Garcia J, et al. Wirelessly observed therapy compared to directly observed therapy to confirm and support tuberculosis treatment adherence: A randomized controlled trial. PLoS Med 2019 Oct;16(10):e1002891 [FREE Full text] [doi: 10.1371/journal.pmed.1002891] [Medline: 31584944]
15. Gupta K, Jain A, Vardhan PH, Singh S, Amber A, Sethi A. MedAssist: automated medication kit. 2014 Presented at: 2014 Texas Instruments India Educators' Conference (TIIEC); April 4-5, 2014; Bangalore, India p. 93-99. [doi: 10.1109/TIIEC.2014.024]
16. Wachira C, Osebe S, Ogallo W, Walcott-Bryant A. Enhancing guideline-based prescribing and personalized medication scheduling. Stud Health Technol Inform 2019 Aug 21;264:873-877. [doi: 10.3233/SHTI190348] [Medline: 31438049]
17. Prado MA, Francisco PM, Barros MB. Diabetes in the elderly: drug use and the risk of drug interaction. Cien Saude Colet 2016 Nov;21(11):3447-3458 [FREE Full text] [doi: 10.1590/1413-812320152111.24462015] [Medline: 27828578]
18. ChatGPT. OpenAI. URL: https://chatgpt.com/ [accessed 2025-01-24]
19. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595 [FREE Full text] [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
20. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW. Large language models encode clinical knowledge. Nature 2023 Aug;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
21. Bailey SC, Persell SD, Jacobson KL, Parker RM, Wolf MS. Comparison of handwritten and electronically generated prescription drug instructions. Ann Pharmacother 2009 Jan;43(1):151-152. [doi: 10.1345/aph.1L388] [Medline: 19126825]
22. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
23. Williams CYK, Miao BY, Kornblith AE, Butte AJ. Evaluating the use of large language models to provide clinical recommendations in the emergency department. Nat Commun 2024 Oct 08;15(1):8236 [FREE Full text] [doi: 10.1038/s41467-024-52415-1] [Medline: 39379357]
24. Song H, Xia Y, Luo Z, Liu H, Song Y, Zeng X, et al. Evaluating the performance of different large language models on health consultation and patient education in urolithiasis. J Med Syst 2023 Nov 24;47(1):125. [doi: 10.1007/s10916-023-02021-3] [Medline: 37999899]

XSL•FO
RenderX

25. Huang H, Zhang L, Yang Y, Huang L, Lu X, Li J, et al. Construction and application of medication reminder system: intelligent generation of universal medication schedule. BioData Min 2024 Jul 15;17(1):23 [FREE Full text] [doi: 10.1186/s13040-024-00376-y] [Medline: 39010132]
26. Amin M, Suksomboon N. Pharmacotherapy of type 2 diabetes mellitus: an update on drug-drug interactions. Drug Saf 2014 Nov;37(11):903-919. [doi: 10.1007/s40264-014-0223-2] [Medline: 25249046]
27. Freeman JS, Gross B. Potential drug interactions associated with treatments for type 2 diabetes and its comorbidities: a clinical pharmacology review. Expert Rev Clin Pharmacol 2012 Jan;5(1):31-42. [doi: 10.1586/ecp.11.64] [Medline: 22142157]
28. von Storch K, Graaf E, Wunderlich M, Rietz C, Polidori MC, Woopen C. Telemedicine-assisted self-management program for type 2 diabetes patients. Diabetes Technol Ther 2019 Sep;21(9):514-521. [doi: 10.1089/dia.2019.0056] [Medline: 31287736]
29. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
30. Zhang D, Ma Z, Gong R, Lian L, Li Y, He Z, et al. Using natural language processing (GPT-4) for computed tomography image analysis of cerebral hemorrhages in radiology: retrospective analysis. J Med Internet Res 2024 Sep 26;26:e58741 [FREE Full text] [doi: 10.2196/58741] [Medline: 39326037]
31. Ali R, Tang OY, Connolly ID, Fridley JS, Shin JH, Zadnik Sullivan PL. Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral boards preparation question bank. Neurosurgery 2023 Nov 01;93(5):1090-1098. [doi: 10.1227/neu.0000000000002551] [Medline: 37306460]
32. He Y, Tang H, Wang D, Gu S, Ni G, Wu H. Will ChatGPT/GPT-4 be a lighthouse to guide spinal surgeons? Ann Biomed Eng 2023 Jul;51(7):1362-1365. [doi: 10.1007/s10439-023-03206-0] [Medline: 37071280]
33. Watari T, Takagi S, Sakaguchi K, Nishizaki Y, Shimizu T, Yamamoto Y, et al. Performance comparison of ChatGPT-4 and Japanese medical residents in the general medicine in-training examination: comparison study. JMIR Med Educ 2023 Dec 06;9:e52202 [FREE Full text] [doi: 10.2196/52202] [Medline: 38055323]
34. Seghier ML. ChatGPT: not all languages are equal. Nature 2023 Mar;615(7951):216. [doi: 10.1038/d41586-023-00680-3] [Medline: 36882613]
35. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? Med Educ Online 2023 Dec;28(1):2181052 [FREE Full text] [doi: 10.1080/10872981.2023.2181052] [Medline: 36809073]
36. Belagere C. Students have started using ChatGPT to cheat in assignments, tests. How are professors catching them? South First. 2023 Jan 27. URL: https://thesouthfirst.com/karnataka/students-have-started-using-chatgpt-to-cheat-in-tests-exams-how-are-professors-catching-them/ [accessed 2024-09-30]

Abbreviations

AI: artificial intelligence
DDI: drug-drug interaction
DSMES: diabetes self-management education and support
DUMS: Diabetes Universal Medication Schedule
LLM: large language model
LoA: limit of agreement
USMLE: United States Medical Licensing Examination

Edited by A Schwartz; submitted 26.01.24; peer-reviewed by J Mistry, JC Caicedo-Eraso, KM Romagnoli, LC Paul; comments to author 08.07.24; revised version received 30.10.24; accepted 10.01.25; published 11.02.25.

Please cite as:
Xiao J, Li M, Cai R, Huang H, Yu H, Huang L, Li J, Yu T, Zhang J, Cheng S
Smart Pharmaceutical Monitoring System With Personalized Medication Schedules and Self-Management Programs for Patients With Diabetes: Development and Evaluation Study
J Med Internet Res 2025;27:e56737
URL: https://www.jmir.org/2025/1/e56737
doi:10.2196/56737
PMID:

©Jian Xiao, Mengyao Li, Ruwen Cai, Hangxing Huang, Huimin Yu, Ling Huang, Jingyang Li, Ting Yu, Jiani Zhang, Shuqiao Cheng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.02.2025. This is an open-access

XSL•FO
RenderX

article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartphone Application–Based Voice and Speech Training Program for Parkinson Disease: Feasibility and Satisfaction Study With a Preliminary Rater-Blinded Single-Arm Pretest and Posttest Design

Sol-Hee Lee[1*], MS; Jiae Kim[1*], MS; Han-Joon Kim[1], MD, PhD

Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Han-Joon Kim, MD, PhD
Department of Neurology
Seoul National University Hospital
Seoul National University College of Medicine
101 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 10 7279 7883
Email: movement@snu.ac.kr

Abstract

Background: Up to 75% of patients with Parkinson disease (PD) experience voice and speech impairments, such as breathy phonation and low speech volume, which worsen over time and negatively impact the quality of life. However, given their increasingly limited mobility, face-to-face speech therapy is often inaccessible. Mobile health (mHealth) apps offer accessible and cost-effective alternatives; yet, their application in PD-specific, self-delivered voice therapy remains underexplored.

Objective: This study aimed to evaluate the feasibility, adherence, and satisfaction of a self-delivered smartphone app for voice therapy in patients with PD, designed to minimize speech-language pathologist involvement while promoting patient independence. In addition, it seeks to assess the preliminary therapeutic effectiveness of the app in addressing voice and speech problems in this population.

Methods: A single-arm, rater-blinded, and pretest and posttest study was conducted between September to November 2023. Patients with PD with voice and speech problems who have no problem with using Android (Google) smartphones were recruited. Participants downloaded the researcher-developed mHealth app on their smartphone and participated in a patient-tailored 5-week home-based speech training program. Each session included 5 stages: breathing, oral motor exercises, loudness, prosody, and functional speaking. The training program consisted of 20 sessions, with participants completing 1 session per day, 4 days per week. Each session lasted approximately 20-30 minutes. Adherence was monitored through app logs, satisfaction was assessed through a phone survey, and therapeutic effectiveness was evaluated using acoustic analysis and auditory-perceptual assessments.

Results: Out of 30 patients were initially recruited, but 2 of them withdrew. Out of 25 participants completed all the training sessions while 3 dropped out. The adherence was above 90% in 20 participants (80%, 20/25), 70% to 90% in 4 (16%, 4/25), and below 70% in 4 (16%, 4/25). Satisfaction was 75% (18/24) among the 24 people who participated in the survey. Significant improvements were observed in all acoustic measures: the maximum phonation time increased from 11.15 (SD 5.38) seconds to 14.01 (SD 5.64) seconds (P=.003), and vocal intensity increased from 71.59 (SD 4.39) dB to 73.81 (SD 3.48) dB (P<.001) across both sustained phonation and reading tasks. Voice quality scores on the GRBAS (grade, roughness, breathiness, asthenia, and strain) scale improved significantly (all components P<.001). Furthermore, 58.3% (14/24) of participants reported subjective improvements in their voice.

Conclusions: This study demonstrates that home-based, self-training speech therapy delivered through a mHealth app is a feasible solution for patients with PD, suggesting that mHealth apps can serve as a convenient and effective alternative to face-to-face therapy by enhancing accessibility and empowering patients to actively manage their condition.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63166) doi:10.2196/63166

KEYWORDS

Parkinson disease; speech therapy; mHealth; home-based training; self-delivered; digital health care; app; feasibility; voice therapy; mobile phone; satisfaction; effectiveness; smartphone; apps; single-arm study; mobility; mobile health; acoustic analysis; self-training

Introduction

During the progression of patients with Parkinson disease (PD), around 75% of patients experience difficulties with their voice and speech functions at some stage [1]. The main characteristics of speech include breathy phonation, hoarseness, low speech volume, inaccurate articulation, and monotonous speech. These symptoms worsen over time. Consequently, these symptoms can significantly impact the quality of life, affecting physical, emotional, functional, and social well-being [2]. They also hinder communication, daily activities, and social participation. Therefore, rehabilitation is essential for individuals with PD, with speech therapy being particularly important as it helps patients maintain and improve their communication abilities in everyday life [3]. Traditionally, speech therapy involves in-person sessions with a speech-language pathologist (SLP). However, the global COVID-19 pandemic disrupted traditional service delivery models, leading to increased interest in remote rehabilitation and mobile health care technologies as viable alternatives. In addition, the rising demand for rehabilitation due to an aging population, coupled with limited health care resources, further highlights the need to adopt and expand information and communication technology [4].

A review of previous studies has confirmed the feasibility of remote speech-language therapy (SLT) across various disability groups. According to Weidner and Lowman [5], remote SLT is primarily conducted in real-time using commercial software such as Skype (Skype Technologies) and Zoom (Zoom Video Communications), or through custom-developed software installed on computers or tablets. While most remote SLT initiatives have focused on patients with chronic aphasia, some have targeted individuals with PD. Many existing studies have concentrated on delivering the Lee Silverman Voice Treatment LOUD or its extended versions, which are proven to be effective for addressing speech and voice issues in patients with PD, through remote platforms [6-9]. Research on remote SLT for patients with PD has shown that it is just as effective as in-person therapy, with no significant reduction in efficacy [6-10]. Meanwhile, mobile health (mHealth) apps, developed to support users in managing their health independently, have advanced considerably. According to the World Health Organization, mHealth is a broad term encompassing activities that involve the use of smartphones, sensors, personal digital assistants, wireless monitoring devices, or other wireless technologies to deliver public health and health care services [11]. These technologies hold promises for promoting autonomous patient participation in treatment, thus maintaining continuity of care while offering benefits in terms of accessibility and cost [12,13].

To date, most mHealth have been designed and studied with a focus on health outcomes such as nutrition and physical activity, mental health management, and diabetes care, primarily targeting populations with chronic conditions, mental health issues, or obesity [14]. Recently, some evidence has emerged supporting the use of mHealth for home-based SLT across various patient groups with speech and language disorders including children with articulation disorders, poststroke aphasia, presbyphonia, and voice disorders associated with PD [15-19].

mHealth apps designed for patients with PD have mostly focused on symptom monitoring and providing information or education [20]. Research in this area has mainly addressed specific aspects of care, such as symptom management, medication adherence, physical activity, and exercise interventions [20-22].

However, very few studies have explored the use of mHealth apps for delivering speech therapy to patients with PD. Maas et al [23] proposed a study protocol for a single-blind randomized controlled trial using the Voice Trainer app and an online platform that provides real-time feedback, enabling patients to practice at home alongside remote therapy. The study aimed to evaluate the effectiveness of personalized, home-based speech therapy on quality of life, speech intelligibility, and social participation in individuals with PD who have reduced speech intelligibility. However, no follow-up research has been published to date. Lo et al [18] developed a speech rehabilitation app specifically for patients with PD and validated its usability with 2 participants, reporting improvements in articulation and volume after short-term use. Horin et al [19] investigated the utility of a mHealth app designed for the comprehensive treatment of PD symptoms, including gait, dexterity, and speech, and concluded that the application alone was insufficient to effectively address gait, speech, or dexterity symptoms in PD.

While the use of mHealth apps for speech therapy in patients with PD is still in its early stages, there is potential for significant advancement through ongoing research and technological development. mHealth offers great benefits in terms of accessibility and cost, especially for patients with PD who gradually experience reduced mobility, as it allows them to access therapy from anywhere with a mobile device [24-26]. Furthermore, it provides patients with the opportunity to practice independently without time constraints, and it can automatically measure, collect, and store data, thereby reducing the workload for clinicians.

This study aims to evaluate the feasibility of using a smartphone app to provide self-delivered voice therapy for patients with PD with minimal involvement from SLPs. It also seeks to explore patient satisfaction with the app and its potential effectiveness as a therapeutic tool.

XSL•FO
RenderX

Methods

Participants and Recruitment Procedures

Participants were recruited from the Movement Disorders Clinic of the Department of Neurology at Seoul National University Hospital, Seoul, South Korea from September to November 2023. Recruitment was conducted through clinical referrals by neurologists during routine outpatient visits. When neurologists identified patients who presented with speech or voice problems, they referred them to the SLPs for further evaluation. The referrals were based on the neurologists' clinical judgment at the clinic. Once referred, SLPs independently conducted auditory-perceptual evaluations to confirm the presence of speech and voice issues. The inclusion criteria which were (1) patients with PD reporting voice problems and speech difficulties with these issues confirmed by both SLPs through auditory-perceptual evaluation; each SLP having over 5 years of clinical experience, independently verified the presence of voice problems in all patients; (2) Android smartphone users without visual or auditory difficulties in using apps; and (3) not having laryngeal dysfunctions caused by other diseases. Exclusion criteria which were (1) being illiterate, (2) having a recent (<6 months) history of speech therapy, (3) difficulty in operating a smartphone due to cognitive and motor dysfunction, and (4) having limited access to the internet at home.

Study Design

This study is a single-arm, rater-blinded, and pretest and posttest design to evaluate the feasibility and, satisfaction of an mHealth program among individuals with PD. The preliminary effectiveness of the app was also assessed.

mHealth App and Training Program

We developed an Android-based smartphone app for voice and speech training (Figure 1). The app consists of a patient's device (patient's own smartphone), a server, and a SLP's terminal (laptop or desktop computer). All the content for patient evaluation and training, which was developed by the SLPs, is stored on the server.

Figure 1. A simplified overview of the composition of the mobile health (mHealth) app used for voice and speech training in Parkinson disease patients. The app comprises 3 main components. SLP: speech-language pathologist.

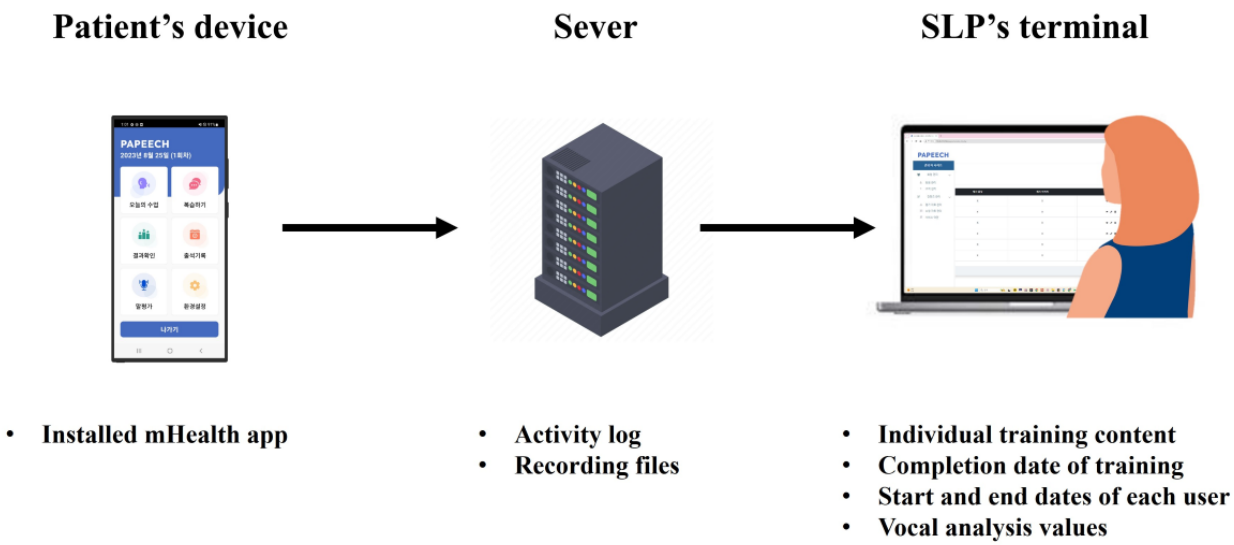

Participants downloaded the mHealth app onto their smartphones at the hospital. Afterward, they were invited to complete a 5-week home-based training program following the registration and user approval process for participation. The SLPs ensured that the installed app was functioning correctly and educated the participants on how to use it. The server transmitted the content to the devices of approved participants. This content was designed with a user-friendly interface and experience. A sufficiently large font size was used to accommodate older users, and the size and spacing of the buttons were appropriately adjusted to consider motor symptoms such as hand tremors. In addition, if the participants have used a smartphone before, the design was made simple and intuitive enough for them to operate without difficulty. This is because the goal was for patients to use the app independently for evaluation and training during the study, without assistance from clinicians, including SLPs, or caregivers.

All speeches practiced in the session were recorded. The recorded files and the associated activity logs are all stored on the server. The SLP's terminal is designed for monitoring the entire process, allowing the SLP to check each participant's start and end dates of training, their individual training content, and the analysis of the recorded speeches within the training. SLPs did not interact with the patient or server in real time during evaluation or training.

Based on the results of the self-administered baseline evaluation (refer below), each participant was assigned to one of 3 different types of training: loudness-focused, breathing-focused, and prosody-focused training. Each training type is designed to include a variety of tasks that target the specific focused area more effectively. The training content comprised 5 stages: breathing, oral motor exercises, loudness, prosody, and speaking (Figure 2). The breathing stage addresses the energy source for speech. PD patients often experience weak or shallow breathing, which can reduce speech volume and stamina. Strengthening

XSL•FO
RenderX

respiratory muscles improves airflow, enabling clearer and more sustained speech [27]. Oral motor exercises target the strength and agility of oral muscles, which are often reduced in PD, leading to unclear speech and difficulty swallowing. Strengthening these muscles enhances articulation and supports swallowing function [28]. Loudness training focuses on hypophonia, or reduced speech volume, which is common in PD. Loudness exercises retrain patients to use stronger vocal effort, helping them communicate more effectively in daily conversations [29]. Prosody training helps address the monotone voice often seen in PD, which can make speech sound flat and less engaging. Prosody exercises enhance speech expressiveness, allowing patients to convey meaning more effectively [30].

Figure 2. Contents and structure of the 5-week speech training program for participants with Parkinson disease patients using the mHealth app.

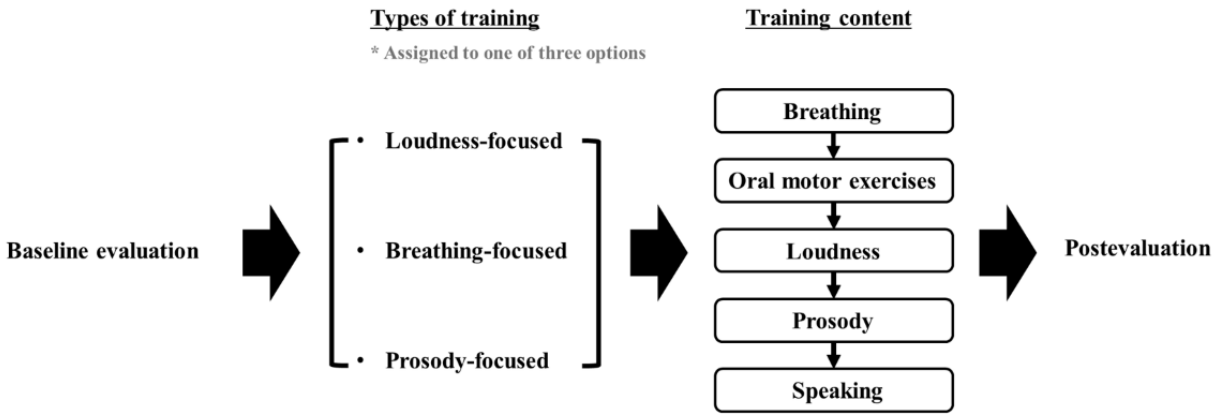

	Examples of tasks in the training content
Breathing	Practice slow inhaling and exhaling
Oral motor exercises	Neck stretching, Gentle laryngeal massage, Smiling
Loudness	Producing a loud, sustained /a/ sound, Producing a loud, short /a/ sound
Prosody	Singing, Gliding /a/ phonation
Speaking	Reading functional phrases or short sentences commonly used in daily life

Finally, functional speaking practice supports patients in initiating and maintaining fluid conversations. This stage builds confidence, improves speech flow, and integrates the skills learned in breathing, loudness, and prosody exercises [31].

This 5-week home-based training program consists of a total of 20 sessions, each lasting approximately 20 minutes to 30 minutes. Participants were instructed to use the app at least 4 days a week, one session a day. To minimize attrition, SLPs monitored the participant's daily use of the app using the dashboard equipped in the SLPs' terminal. If necessary, SLPs called the patients to encourage training. The app also includes a pop-up alarm to remind users of each day's training.

Measuring Adherence, Satisfaction, and Effectiveness

To verify adherence, we created a webpage for SLPs to monitor the participants' completion of app training sessions in real-time. The overall adherence was calculated by dividing the number of training sessions completed by the total number of sessions prescribed [8]. The number of participants who withdrew from the study was also identified.

The Satisfaction survey was conducted by phone within a week of either completing or discontinuing the app's use. Their satisfaction levels were gauged using a 5-point Likert scale anchored by "very satisfied" (5) to "very dissatisfied" (1), aiming to derive the Customer Satisfaction (CSAT) score. The CSAT score is calculated by dividing the number of positive responses (those who are rated 4 or 5) by the total number of responses.

To assess the effectiveness, we conducted both acoustic analysis and auditory perceptual assessments at baseline and at the completion of app use. The tasks included sustained phonation of the vowel /a/, reading 3 short sentences, and reading a 28-word passage from a Korean standardized "Ga-eul" passage. All tasks were recorded using the built-in microphone of each participant's smartphone and were stored on the server. During the recording, participants were instructed to sit upright and remain still, maintaining a distance of approximately 30 cm between their mouths and the smartphone microphone.

The acoustic analysis was conducted using Praat (University of Amsterdam) software (version 6.4.04) embedded in the app. The analysis included the maximum phonation time of the vowel /a/ and the mean vocal intensity (including pauses) for each task, measured in sound pressure level decibel.

The auditory perceptual assessment was performed on all tasks using the GRBAS scale, which is widely used to assess the

XSL•FO
RenderX

severity of voice quality. It contains grade (G), which is the overall degree of hoarseness, roughness (R), breathiness (B), asthenia (A), and strain (S). Out of 2 experienced SLPs, each with over 10 years of clinical experience, conducted the ratings on a 4-point Likert scale from 0 (normal) to 3 (severe). The baseline and posttraining evaluation voice files pair were provided in random order in WAV (Waveform audio format) without any identifying information about the order. Furthermore, participants who completed all app usage were asked, during a phone survey, to self-report any changes in their overall voice quality after the training.

Statistical Analysis

Data were analyzed using SPSS (version 29.0; IBM Corp) for Windows. Frequency analysis was conducted to examine adherence, satisfaction, and self-reported voice quality changes. Paired sample *t* tests were conducted to determine if there were significant changes in the maximum phonation time during sustained phonation of the vowel /a/ before and after app usage, as well as to assess for significant changes in vocal intensity in all tasks. These tests were chosen to compare pre- and postintervention measures for the same participants. The normality of the data was tested using the Shapiro-Wilk test. The results indicated that intensity scores were normally distributed with a *P* value of *P*=.34. Based on this, we chose parametric tests for subsequent analysis. To analyze the reliability of the average agreement rate between evaluators for the auditory perceptual assessment of GRBAS scores, intraclass correlation coefficients (ICCs) were calculated.

Ethical Considerations

This study was approved by the Seoul National University Hospital Ethics Committee (H-2308-040-1456 and H-2401-010-1499). The research adhered to the principles outlined in the Declaration of Helsinki, and all procedures were conducted in compliance with ethical standards for research involving human participants. All participants provided informed written consent before their participation. They were informed about the study's purpose, procedures, voluntary nature, and their right to withdraw at any time without consequences. The consent process ensured participants understood that their data would be used for research purposes and protected under the approved ethical guidelines. Data were securely stored on an encrypted server accessible only to authorized research personnel. Participants were not monetarily compensated for their participation.

Results

Participants' Characteristics and Adherence

Thirty patients with PD were initially recruited, but 2 withdrew due to health problems not related to PD. A total of 28 patients with PD participated in the study and used the app. Participant characteristics are summarized in Table 1. Details regarding participant recruitment are provided in Multimedia Appendix 1. Based on the results of baseline evaluation, 20 were assigned to the loudness-focused training program and the remaining 8 participated in the breathing-focused training program. None were assigned to the prosody-focused training program. Out of 25 participants (15 male, 10 female; mean age 68.04, SD 7.81 years; mean disease duration 7.32, SD 4.78 years; H and Y scale ranged from 1 to 4, mean 2.39, SD 0.88) completed all the training sessions while 3 dropped out. One participant could not be reached and the other 2 dropped out due to (1) problems with the internet connection and (2) being too busy, respectively. They completed only 12, 3, and 4 sessions respectively.

The adherence was above 90% of 20 participants, 70-90% in 4, and below 70% in 4 (Figure 3A).

Table 1. Demographic and clinical characteristics of the participants with Parkinson disease (N=28), recruited from the Movement Disorders Clinic, Seoul National University Hospital, Seoul, South Korea, between September and November 2023.

Subgroup	Value
Age (years), mean (SD; range)	67.5 (8.16; 51-81)
Sex, n (%)	
Male	17 (60.7)
Female	11 (39.3)
Disease duration (years), mean (SD; range)	7.54 (4.73; 1-19)
H and Y scale score (mean 2.39, SD 0.88, median 2, IQR 2–3), n (%)	
1	3 (10.7)
2	15 (53.6)
3	6 (21.4)
4	4 (14.3)
Type of training program, n (%)	
Loudness-focused	20 (71.4)
Breathing-focused	8 (28.6)
Prosody-focused	0 (0)

XSL•FO
RenderX

Figure 3. Adherence, satisfaction, and self-reported voice changes among participants using the mobile health (mHealth) app for voice and speech training. (A) Adherence: The percentage of participants who completed the 5-week training program (20 sessions) using the mHealth app. (B) Participant satisfaction: Using a 5-point Likert scale. (C) The self-report of changing voice: Participants' subjective assessments of whether their voice improved, stayed the same, worsened, or if they were not sure after completing the training program.

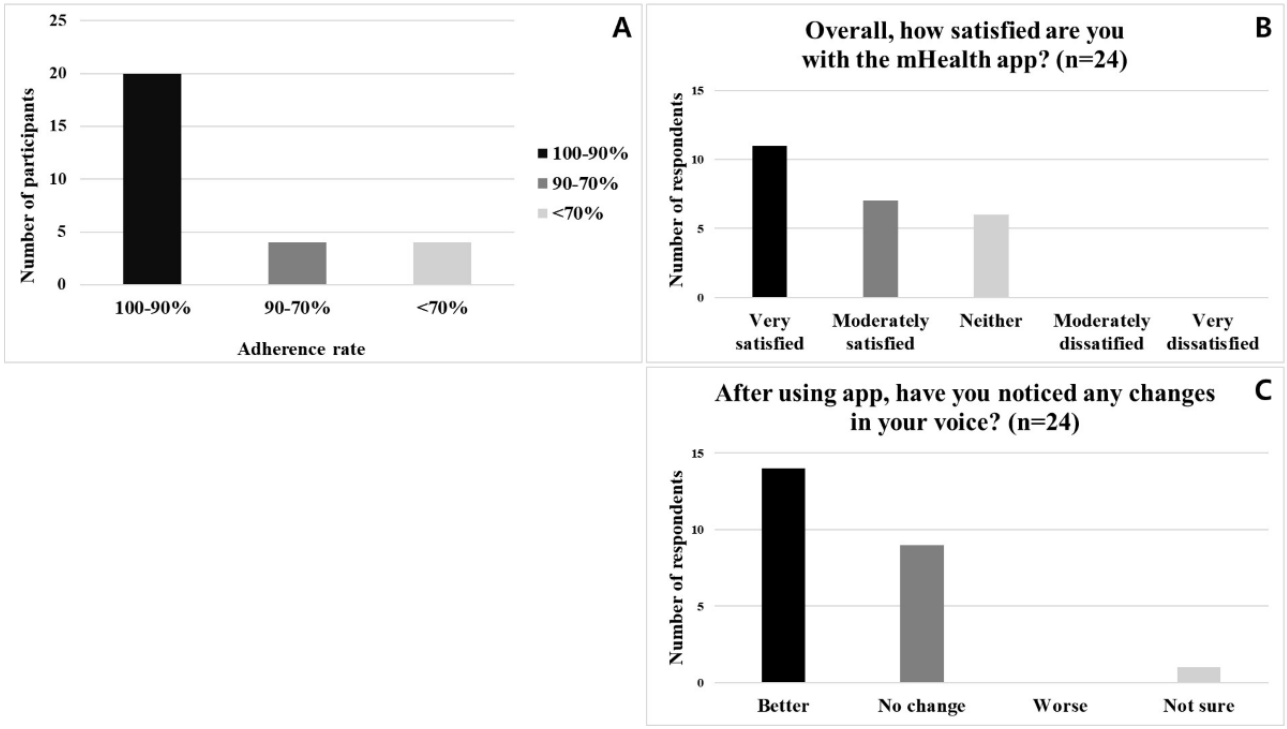

Satisfaction

After using the app, 24 participants took part in the survey on satisfaction (Figure 3B). Out of 11 (45.83%) responded "very satisfied," 7 (29.17%) "moderately satisfied," 6 (25%) "neither" and none answered, "moderately dissatisfied" or "very dissatisfied." The CSAT score was 75%. Most were satisfied with the convenience of being able to practice at their own home, citing it as the primary reason. None reported any technical difficulties.

Effectiveness

Acoustic analysis and perceptual assessment results are presented in Table 2. Significant differences were observed in all acoustic measurements (P<.05). Significant differences in voice quality in all components of the GRBAS scale were observed before and after using the app. The inter-rater reliability between the 2 raters was fair too good (ICC=0.440). Among the 24 who took part in the survey (Figure 3C), 14 (58.33%) reported their voice "getting better," 9 (37.50%) reported "no change," 1 (4.17%) reported being "unsure," and none reported "getting worse."

XSL•FO
RenderX

Table 2. Pre- and postintervention changes in maximum phonation time (MPT), vocal intensity, and GRBAS[a] scale scores in participants with Parkinson disease (N=25).

Variable	Before	After	*t* test (*df*) or *z* score	*P* value
MPT (seconds)	11.15 (5.38)	14.01 (5.64)	–3.273 (24)	.003
Vocal intensity, mean (SD)				
Sustained phonation (dB)	72.32 (5.04)	75.67 (3.75)	–4.761 (24)	<.001
Reading short sentences (dB)	72.59 (4.09)	74.04 (4.08)	–3.164 (24)	.004
Reading “Ga-eul” passage (dB)	69.64 (4.55)	71.59 (3.61)	–3.546 (23)	.002
All tasks (dB)	71.59 (4.39)	73.81 (3.48)	–4.462 (24)	<.001
Variable (GRBAS scale score)				
Grade	1 (2)	1 (1)	–4.601	<.001
Roughness	0 (2)	0 (1)	–3.804	<.001
Breathiness	1 (2)	0 (1)	–5.959	<.001
Asthenia	0 (1)	0 (0)	–6.130	<.001
Strain	0 (1)	0 (1)	–3.930	<.001

[a]GRBAS: grade, roughness, breathiness, asthenia, and strain.

Discussion

Principal Findings

The aim of this study was to investigate the feasibility of patients with PD independently performing voice and speech training at home using a smartphone app as an alternative to traditional face-to-face therapy. The study involved SLPs in developing a mHealth app based on speech training content commonly used on patients with PD, and evaluated patients’ adherence to treatment schedules, satisfaction with app-based therapy, and treatment effectiveness.

Our study indicates that speech therapy delivered through an mHealth app is a feasible solution for patients with PD. A total of 20 out of 25 participants who completed the app showed an adherence rate of over 90%. Given that dropout rates in fields requiring various health behavior changes reach 30%-60%, an adherence rate of over 90% can be considered sufficiently high and indicative of successful adherence [32]. It is typically stated that an adherence rate of over 80% is necessary for optimal therapeutic efficacy, and our study has achieved this.

About 75% of participants were satisfied with the app use and improvements in voice and speech were observed both in objective and subjective evaluations. This is similar to the satisfaction rate (70%) of an internet-based telerehabilitation application for speech therapy [33] and is also comparable to a study reporting satisfaction rate of 82.6% with the smartphone app for people with aphasia [34].

Another notable point is that positive results were obtained for both objective and subjective ratings for voice. It is generally known that acoustic assessments highly correlate with SLP’s auditory perceptual ratings but have a low correlation with patient subjective assessments [35,36]. This indicates that there is a discrepancy between the degree of voice impairment assessed by the assessor and by the patients. The consistency in our results among 3 methods of evaluation indicates the robustness of effectiveness.

The advancement of mobile and electronic technologies has increased interest in non–face-to-face therapy, which has been further accelerated by the COVID-19 pandemic. This can help address medical accessibility issues. If non–face-to-face methods are proven to be as usable and effective as face-to-face therapy, they will be a more convenient way to receive medical services. Previous studies have confirmed its feasibility and effectiveness in various rehabilitation fields, such as speech, motor, and cognition [10,37,38]. In addition, the benefits of online treatment reported include time and cost savings and reduced caregiver burden. Along with these benefits, a key advantage of mHealth apps is that they enable patients to take an active role in managing their condition. Previous studies indicate that the integration of mHealth technology into therapy is an optimized clinical pathway for patients [6]. Although the number of studies is limited compared to real-time teletherapy provided by therapists, there are a few that have confirmed the usefulness and effectiveness of app-based therapies [15,17,18,39]. These include reports of significant improvements in pronunciation and speech volume in people with PD who received speech therapy using homegrown software [18]. This finding is in line with our results and illustrates the potential of the mHealth app as a feasible approach to speech rehabilitation.

In contrast to the high adherence in our study, a previous study investigating the usability of an mHealth app developed to address walking, speech, and dexterity symptoms in patients with PD reported low adherence and effectiveness [19]. The adherence to speech exercises was 66.8% (SD 26.5%). In the study, participants in the intervention group were instructed to independently complete exercises using the provided mHealth app on their personal smartphones once a day for approximately 90 days. This app required practicing various areas such as mobility exercises in addition to speech, and it took about 30

XSL•FO
RenderX

minutes to complete all sessions. This may have placed a physical and mental burden on the users.

In addition, considering that appropriate feedback and interaction with therapists generally impact the adherence and effectiveness of therapy, the extent to which sufficient feedback was automatically provided within the app may have also influenced these outcomes. In our app, not only does it provide reminder functions through pop-up alarms, but it also provides different feedback each time whenever the patient performs a task (eg, "Good job! Squeeze your abdomen and speak louder!"). Such feedback may have positively influenced the patient's continued use. Furthermore, over 50% of the participants in the previous study experienced technological issues with the app, which may have interfered with the outcomes. In our app, we tried to create a user interface older adults friendly as much as possible and to ensure the seamless operation of the app. As a result, there were no reports of technical issues, and this also could have affected adherence, effectiveness, and satisfaction.

Limitations and Future Directions

However, our study has several limitations. Because the primary aim of this study was to evaluate the feasibility and it was an uncontrolled study, the effectiveness observed in our study should be interpreted with caution. Larger randomized controlled trials are needed to provide more robust results on effectiveness. In addition, consideration of factors that may affect the usability of the mHealth app, such as disease severity and digital literacy [40] of patients from the recruitment stage, would be useful to evaluate the potential for introducing this app to a broader population of individuals with PD.

Second, we were unable to conduct functional assessments to determine whether participants' voice issues improved in real-life communication situations. While the Voice-Related Quality of Life tool does not directly measure improvements in voice use during real-life situations, it provides valuable insights into how patients perceive the impact of their voice disorder on their quality of life. Studies have demonstrated that after face-to-face speech therapy, such as the Lee Silverman Voice Treatment or the Speak Out Voice program, patients exhibit improved Voice-Related Quality of Life scores [41,42].

Using the Dysarthria Impact Questionnaire to evaluate the psychosocial effects of dysarthria could also provide meaningful insights [43]. In addition, assessing effectiveness using data collected from naturalistic environments, such as real-life conversational settings, could yield more reliable results. Including evaluations from primary caregivers to determine whether patients' speech problems have improved would provide a more objective perspective. Patients with PD often face challenges with self-monitoring due to the nature of the disease, making patient-reported improvements insufficient [44]. Therefore, evaluations should integrate diverse perspectives and methodologies.

Third, our study encountered difficulties in achieving consistent calibration across all participants' devices. The variations in smartphone models and built-in microphone sensitivities presented challenges in standardizing intensity calibration. According to Park and Lee [45], differences in microphone characteristics among smartphone models can lead to discrepancies in recording and analysis, with omnidirectional microphones, in particular, being more susceptible to ambient noise. They suggested that using a unidirectional wired lapel microphone could offer more consistent recording quality. Consequently, we recommended that participants maintain a specific distance from the microphone and minimize background noise, providing tutorial videos within the app to facilitate consistent recording practices. However, hardware differences among devices may still impact the consistency of intensity measurements. This suggests that future research may benefit from implementing additional calibration methods to improve measurement reliability.

Fourth, while this app was designed as a tool to substitute face-to-face therapy, it currently faces limitations due to its inability to fully replicate the real-time feedback and personalized adjustments provided during in-person therapy. To overcome these challenges, future research should focus on improving the app's functionality and design.

Specifically, integrating customized feedback mechanisms and improved monitoring systems is crucial. Without these features, participants may risk developing vocal abuse by overcompensating with improper techniques. Continuous monitoring can help identify and resolve any discomfort or tension early, ensuring the app's safety and effectiveness. These improvements could help bridge the gap between app-based therapy and face-to-face therapy. In this process, it is also important to investigate the app's impact on the overall quality of life of patients with PD compared to traditional in-person therapy. Widely used tools such as the World Health Organization Quality of Life Questionnaire or Parkinson Disease Questionnaire-39 can be used for this purpose [46].

In addition, understanding the impact of self-managed mHealth apps on caregiver burden, particularly from the perspective of primary caregivers, is an important research topic. If such apps enable patients to manage themselves more effectively, they could reduce the caregiving burden, alleviating stress and fatigue for caregivers. This would not only enhance the quality of life for caregivers but could also have a positive impact on the patient-caregiver relationship. Therefore, such analyses are crucial for evaluating whether the app provides meaningful benefits for both patients and their caregivers.

Finally, future research could consider longitudinal studies to evaluate the long-term effects and sustainability of mHealth apps. Patient adherence is crucial for treatment effectiveness and improvement in their health status. Nevertheless, most existing studies have short durations, limiting research on factors affecting mHealth app adherence [47]. Tracking changes in user adherence over extended periods could help develop user retention strategies. Furthermore, the accumulated data could be used to analyze user behavior patterns and develop predictive models.

Conclusions

Despite the small sample size of this study, this study supports the feasibility of self-delivered speech therapy at home for patients with PD. Especially considering the age of the

XSL•FO
RenderX

participants, the mHealth app we developed is not only suitable for older users but also proves to be satisfying for them.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R and D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: RS-2023-00265517). We sincerely thank all the participants for their involvement in this study. We also extend our gratitude to the 2 SLPs, Professor Seung Jin Lee and Professor Jung Wan Kim, for their assistance with the auditory perceptual assessment.

Data Availability

The datasets generated during this study are not publicly available due to the small sample size, which requires protecting the privacy of participants, but are available from the corresponding author on reasonable request.

Authors' Contributions

Study concept and design were performed by Han-Joon Kim, Jiae Kim, and Sol-Hee Lee. Acquisition, analysis, or interpretation of data were conducted by Han-Joon Kim, Jiae Kim, and Sol-Hee Lee. Drafting of the paper was handled by Han-Joon Kim and Sol-Hee Lee. Critical revision of the paper for important intellectual content were performed by Han-Joon Kim, Jiae Kim, and Sol-Hee Lee. All authors have approved the submitted version and have agreed both to be personally accountable and to ensure that questions are appropriately investigated, resolved, and the resolution documented in the literature.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details regarding participant recruitment.
[PNG File , 132 KB - jmir_v27i1e63166_app1.png]

References

1. Logemann JA, Fisher HB, Boshes B, Blonsky ER. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients. J Speech Hear Disord 1978;43(1):47-57. [doi: 10.1044/jshd.4301.47] [Medline: 633872]
2. van Hooren MRA, Baijens LWJ, Vos R, Pilz W, Kuijpers LMF, Kremer B, et al. Voice- and swallow-related quality of life in idiopathic parkinson's disease. Laryngoscope 2016;126(2):408-414. [doi: 10.1002/lary.25481] [Medline: 26153234]
3. Gillivan-Murphy P, Miller N, Carding P. Voice treatment in Parkinson’s disease: patient perspectives. JPRLS 2019;9:29-42. [doi: 10.2147/jprls.s180183]
4. Jones CH, Dolsten M. Healthcare on the brink: navigating the challenges of an aging society in the United States. NPJ Aging 2024;10(1):22 [FREE Full text] [doi: 10.1038/s41514-024-00148-2] [Medline: 38582901]
5. Weidner K, Lowman J. Telepractice for adult speech-language pathology services: a systematic review. Perspect ASHA SIGs 2020;5(1):326-338. [doi: 10.1044/2019_persp-19-00146]
6. Theodoros D, Hill A, Russell T. Clinical and quality of life outcomes of speech treatment for parkinson's disease delivered to the home via telerehabilitation: a noninferiority randomized controlled trial. Am J Speech Lang Pathol 2016;25(2):214-232. [doi: 10.1044/2015_AJSLP-15-0005] [Medline: 27145396]
7. Griffin M, Bentley J, Shanks J, Wood C. The effectiveness of lee silverman voice treatment therapy issued interactively through an iPad device: a non-inferiority study. J Telemed Telecare 2018;24(3):209-215. [doi: 10.1177/1357633X17691865] [Medline: 28147896]
8. Quinn R, Park S, Theodoros D, Hill AJ. Delivering group speech maintenance therapy via telerehabilitation to people with parkinson's disease: a pilot study. Int J Speech Lang Pathol 2019;21(4):385-394. [doi: 10.1080/17549507.2018.1476918] [Medline: 29879854]
9. Dias AE, Limongi JCP, Barbosa ER, Hsing WT. Voice telerehabilitation in parkinson's disease. Codas 2016;28(2):176-181 [FREE Full text] [doi: 10.1590/2317-1782/20162015161] [Medline: 27191882]
10. Chang HJ, Kim J, Joo JY, Kim H. Feasibility and efficacy of video-call speech therapy in patients with parkinson's disease: a preliminary study. Parkinsonism Relat Disord 2023;114:105772. [doi: 10.1016/j.parkreldis.2023.105772] [Medline: 37531838]
11. World Health Organization. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2025-01-10]

XSL•FO
RenderX

12. Gorski I, Bram JT, Sutermaster S, Eckman M, Mehta K. Value propositions of mHealth projects. J Med Eng Technol 2016;40(7-8):400-421. [doi: 10.1080/03091902.2016.1213907] [Medline: 27687907]
13. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
14. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
15. Kim SY, Song M, Jo Y, Jung Y, You H, Ko MH, et al. Effect of voice and articulation parameters of a home-based serious game for speech therapy in children with articulation disorder: prospective single-arm clinical trial. JMIR Serious Games 2023;11:e49216 [FREE Full text] [doi: 10.2196/49216] [Medline: 37819707]
16. Liu H, Cordella C, Ishwar P, Betke M, Kiran S. Consistent long-term practice leads to consistent improvement: benefits of self-managed therapy for language and cognitive deficits using a digital therapeutic. Front Digit Health 2023;5:1095110 [FREE Full text] [doi: 10.3389/fdgth.2023.1095110] [Medline: 37114182]
17. Cho NB, Cho SR, Choi SH, You H, Nam SI, Kim H. Short-term and long-term efficacy of oropharyngolaryngeal strengthening training on voice using a mobile healthcare application in elderly women. Commun Sci Disord 2021;26(1):219-230. [doi: 10.12963/csd.21799]
18. Lo HC, Tang ST, Wei WL, Chuang CC. Design and usability evaluation of speech rehabilitation app interface for patients with Parkinson's disease. 2024 Presented at: Universal access in human computer interaction. Human and technological environments. UAHCI 2017; 2024 July 29; Springer. [doi: 10.1007/978-3-319-58700-4_6]
19. Horin AP, McNeely ME, Harrison EC, Myers PS, Sutter EN, Rawson KS, et al. Usability of a daily mHealth application designed to address mobility, speech and dexterity in parkinson's disease. Neurodegener Dis Manag 2019;9(2):97-105 [FREE Full text] [doi: 10.2217/nmt-2018-0036] [Medline: 30998100]
20. Estévez S, Cambronero M, García-Ruiz Y, Llana Díaz LF. Mobile applications for people with parkinson's disease: a systematic search in app stores and content review. J Univers Comput Sci 2019;25(7):740-763. [doi: 10.3217/jucs-025-07-0740]
21. Triantafyllidis A, Segkouli S, Zygouris S, Michailidou C, Avgerinakis K, Fappa E, et al. Mobile app interventions for parkinson's disease, multiple sclerosis and stroke: a systematic literature review. Sensors (Basel) 2023;23(7):3396 [FREE Full text] [doi: 10.3390/s23073396] [Medline: 37050456]
22. Kim A, Yun SJ, Sung KS, Kim Y, Jo JY, Cho H, et al. Exercise management using a mobile app in patients with parkinsonism: prospective, open-label, single-arm pilot study. JMIR Mhealth Uhealth 2021;9(8):e27662 [FREE Full text] [doi: 10.2196/27662] [Medline: 34463635]
23. Maas JJL, De Vries NM, Bloem BR, Kalf JG. Design of the PERSPECTIVE study: PERsonalized SPEeCh therapy for actIVE conversation in parkinson's disease (randomized controlled trial). Trials 2022;23(1):274 [FREE Full text] [doi: 10.1186/s13063-022-06160-9] [Medline: 35395953]
24. Theodoros D, Aldridge D, Hill A, Russell T. Technology-enabled management of communication and swallowing disorders in parkinson's disease: a systematic scoping review. Int J Lang Commun Disord 2019;54(2):170-188. [doi: 10.1111/1460-6984.12400] [Medline: 29923267]
25. Theodoros D. Telerehabilitation for communication and swallowing disorders in parkinson's disease. J Parkinsons Dis 2021;11(s1):S65-S70 [FREE Full text] [doi: 10.3233/JPD-202414] [Medline: 33896848]
26. Vaezipour A, Campbell J, Theodoros D, Russell T. Mobile apps for speech-language therapy in adults with communication disorders: review of content and quality. JMIR Mhealth Uhealth 2020;8(10):e18858 [FREE Full text] [doi: 10.2196/18858] [Medline: 33118953]
27. Zhuang J, Jia J. Effects of respiratory muscle strength training on respiratory-related impairments of parkinson's disease. Front Aging Neurosci 2022;14:929923 [FREE Full text] [doi: 10.3389/fnagi.2022.929923] [Medline: 35847666]
28. Wang CM, Shieh WY, Ho CS, Hu YW, Wu YR. Home-based orolingual exercise improves the coordination of swallowing and respiration in early parkinson disease: a quasi-experimental before-and-after exercise program study. Front Neurol 2018;9:624 [FREE Full text] [doi: 10.3389/fneur.2018.00624] [Medline: 30104999]
29. Ramig LO, Countryman S, Thompson LL, Horii Y. Comparison of two forms of intensive speech treatment for parkinson disease. J Speech Hear Res 1995;38(6):1232-1251. [doi: 10.1044/jshr.3806.1232] [Medline: 8747817]
30. Stegemöller EL, Radig H, Hibbing P, Wingate J, Sapienza C. Effects of singing on voice, respiratory control and quality of life in persons with parkinson's disease. Disabil Rehabil 2017;39(6):594-600. [doi: 10.3109/09638288.2016.1152610] [Medline: 26987751]
31. Barnish MS, Horton SMC, Butterfint ZR, Clark AB, Atkinson RA, Deane KHO. Speech and communication in parkinson's disease: a cross-sectional exploratory study in the UK. BMJ Open 2017;7(5):e014642 [FREE Full text] [doi: 10.1136/bmjopen-2016-014642] [Medline: 28554918]
32. Litts JK, Abaza MM. Does a multidisciplinary approach to voice and swallowing disorders improve therapy adherence and outcomes? Laryngoscope 2017;127(11):2446. [doi: 10.1002/lary.26756] [Medline: 28699270]
33. Theodoros DG, Constantinescu G, Russell TG, Ward EC, Wilson SJ, Wootton R. Treating the speech disorder in parkinson's disease online. J Telemed Telecare 2006;12(7):88-91. [doi: 10.1258/135763306779380101]

XSL•FO
RenderX

34. Choi YH, Park HK, Paik NJ. A telerehabilitation approach for chronic aphasia following stroke. Telemed J E Health 2016;22(5):434-440. [doi: 10.1089/tmj.2015.0138] [Medline: 26544535]
35. Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA, Hoffman HT. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. J Voice 2007;21(5):576-590. [doi: 10.1016/j.jvoice.2006.05.001] [Medline: 16822648]
36. Woisard V, Bodin S, Yardeni E, Puech M. The voice handicap index: correlation between subjective patient response and quantitative assessment of voice. J Voice 2007;21(5):623-631. [doi: 10.1016/j.jvoice.2006.04.005] [Medline: 16887329]
37. Federico S, Cacciante L, Cieślik B, Turolla A, Agostini M, Kiper P, RIN_TR_Group. Telerehabilitation for neurological motor impairment: a systematic review and meta-analysis on quality of life, satisfaction, and acceptance in stroke, multiple sclerosis, and parkinson's disease. J Clin Med 2024;13(1):299 [FREE Full text] [doi: 10.3390/jcm13010299] [Medline: 38202306]
38. Schoenberg MR, Ruwe WD, Dawson K, McDonald NB, Houston B, Forducey PG. Comparison of functional outcomes and treatment cost between a computer-based cognitive rehabilitation teletherapy program and a face-to-face rehabilitation program. Professional Psychology: Research and Practice 2008;39(2):169-175. [doi: 10.1037/0735-7028.39.2.169]
39. Hutchison MG, Di Battista AP, Loenhart MM. A continuous aerobic resistance exercise protocol for concussion rehabilitation delivered remotely via a mobile app: feasibility study. JMIR Form Res 2023;7:e45321 [FREE Full text] [doi: 10.2196/45321] [Medline: 37335605]
40. Esper CD, Valdovinos BY, Schneider RB. The importance of digital health literacy in an evolving parkinson's disease care system. J Parkinsons Dis 2024;14(s1):S181-S189. [doi: 10.3233/JPD-230229] [Medline: 38250786]
41. Behrman A, Cody J, Elandary S, Flom P, Chitnis S. The effect of SPEAK OUT! and the LOUD crowd on dysarthria due to parkinson's disease. Am J Speech Lang Pathol 2020;29(3):1448-1465 [FREE Full text] [doi: 10.1044/2020_AJSLP-19-00024] [Medline: 32421347]
42. Parveen S. Group-based intervention of participants with parkinson disease: findings from a 6-month LOUD Crowd® program. Clin Arch Commun Disord 2020;5(2):96-105. [doi: 10.21849/cacd.2020.00241]
43. Walshe M, Peach R, Miller N. Dysarthria impact profile: development of a scale to measure psychosocial effects. Int J Lang Commun Disord 2009;44(5):693-715. [doi: 10.1080/13682820802317536] [Medline: 18821230]
44. Contreras-Ruston F, Castillo-Allendes A, Saavedra-Garrido J, Ochoa-Muñoz AF, Hunter EJ, Kotz SA, et al. Voice self-assessment in individuals with parkinson's disease as compared to general voice disorders. Parkinsonism Relat Disord 2024;123:106944 [FREE Full text] [doi: 10.1016/j.parkreldis.2024.106944] [Medline: 38552350]
45. Park JI, Lee SJ. A comparison of acoustic measures among the microphone types for smartphone recordings in normal adults. Phonet Speech Sci 2024;16(2):49-58. [doi: 10.13064/ksss.2024.16.2.049]
46. Zhao N, Yang Y, Zhang L, Zhang Q, Balbuena L, Ungvari GS, et al. Quality of life in parkinson's disease: a systematic review and meta-analysis of comparative studies. CNS Neurosci Ther 2021;27(3):270-279 [FREE Full text] [doi: 10.1111/cns.13549] [Medline: 33372386]
47. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]

Abbreviations

CSAT: Customer Satisfaction
GRBAS: grade, roughness, breathiness, asthenia, and strain
ICC: intraclass correlation coefficient
mHealth: mobile health
PD: Parkinson disease
SLP: speech-language pathologist
SLT: speech-language therapy
WAV: Waveform audio format

XSL•FO
RenderX

Edited by A Mavragani; submitted 12.06.24; peer-reviewed by Z Thijs, P Clay; comments to author 12.10.24; revised version received 30.11.24; accepted 30.12.24; published 13.02.25.

Please cite as:
Lee SH, Kim J, Kim HJ
Smartphone Application–Based Voice and Speech Training Program for Parkinson Disease: Feasibility and Satisfaction Study With a Preliminary Rater-Blinded Single-Arm Pretest and Posttest Design
J Med Internet Res 2025;27:e63166
URL: https://www.jmir.org/2025/1/e63166
doi:10.2196/63166
PMID:

©Sol-Hee Lee, Jiae Kim, Han-Joon Kim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Continuous Improvement of Chronic Tinnitus Through a 9-Month Smartphone-Based Cognitive Behavioral Therapy: Randomized Controlled Trial

Uso Walter[1], Dr med; Stefan Pennig[2], DPhil; Lothar Bleckmann[3], Dr med; Kristina Röschmann-Doose[4], Dr rer nat; Thomas Wittig[4], Dr med; Jörn Thomsen[4], Dr rer nat; Winfried Schlee[5,6], Dr rer nat, PD

[1]ENT Practice Walter & Zander, Duisburg, Germany

[2]context, Essen-Kettwig, Germany

[3]ENT Practice Dr med Bleckmann, Kleve, Germany

[4]G Pohl-Boskamp GmbH & Co KG, Hohenlockstedt, Germany

[5]Eastern Switzerland University of Applied Sciences, St Gallen, Switzerland

[6]Clinic and Polyclinic for Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

Corresponding Author:
Jörn Thomsen, Dr rer nat
G Pohl-Boskamp GmbH & Co KG
Kieler Straße 11
Hohenlockstedt, 25551
Germany
Phone: 49 4826 59 457
Fax: 49 4826 59 213
Email: j.thomsen@pohl-boskamp.de

Abstract

Background: Tinnitus is the perception of sound without an external auditive stimulus and can be a severe burden for affected patients. Medical guidelines recommend cognitive behavioral therapy (CBT) for tinnitus treatment, which effectively improves tinnitus-related distress and anxiety.

Objective: This study investigates the outcome of a 9-month smartphone-based CBT for patients with tinnitus.

Methods: The randomized controlled clinical trial in this study investigates the efficacy of a smartphone-based CBT for 187 patients with chronic tinnitus over a 9-month treatment period. In the initial 3 months, a waiting list design was applied, and in the subsequent study phase, the data of both treatment groups were collectively analyzed. The scores on the Tinnitus Questionnaire (TQ); 9-item Patient Health Questionnaire (PHQ-9); 9-item Self-Efficacy, Optimism, Pessimism (SWOP-K9) questionnaire; and 20-item Perceived Stress Questionnaire (PSQ-20) were assessed as endpoints after 3 and 9 months of treatment.

Results: We observed a statistically significant reduction in the tinnitus burden in patients who received the smartphone-based CBT intervention. Although no changes were observed initially in the TQ sum scores in the waiting control group (baseline mean, 37.8, SD 4.7; 3 months mean 37.5, SD 4.8; analysis of covariance [ANCOVA] P=.52), the scores significantly decreased once the app-based CBT had commenced. Data pooled from both groups revealed significant reduction in the TQ sum score by 12.49 (SD 1.44) (ANCOVA, P<.001) and 18.48 (SD 1.85) (ANCOVA, P<.001) points after 3 and 9 months, respectively, which was also clinically important. The calculated Cohen d was 1.38. Similarly, the scores on PSQ-20 (–9.14 points; ANCOVA, P<.001), PHQ-9 (–2.47 points; ANCOVA, P<.001), and SWOP-K9 (0.17 points; ANCOVA, P<.001) were significantly improved at the end of the therapy, with corresponding intermediate effect sizes after 9 months.

Conclusions: The data in our study provide evidence of statistically significant, clinically relevant, and continuous benefits of an app-based CBT intervention in patients with chronic tinnitus.

Trial Registration: Deutsches Register Klinischer Studien DRKS00022973; https://drks.de/search/de/trial/DRKS00022973

(J Med Internet Res 2025;27:e59575) doi:10.2196/59575

XSL•FO
RenderX

KEYWORDS

tinnitus; chronic tinnitus; mobile health app; mHealth; mobile app; application; smartphone; digital health; digital technology; digital intervention; cognitive behavioral therapy; randomized controlled trial; statistical analysis

Introduction

Subjective tinnitus is the perception of sound without an external stimulus and is considered chronic if it is present for more than 3 months. The total prevalence of tinnitus among adults is about 14%, which is independent of sex but increases with age and is considered severe in about 2.3% of the affected patients [1]. How the burden is perceived differs substantially depending on the characteristics and severity of tinnitus or accompanying comorbidities and may affect the quality of life substantially. The multifaceted nature of tinnitus, risk factors, and the purely subjective evaluation and accompanying comorbidities make the treatment challenging [2,3]. Highly diverse treatment options have been developed, which may, however, be effective only in a certain subset of patients [4-7]. Treatment may also target not tinnitus itself but comorbidities (eg, hearing loss, depression) with secondary beneficial effects on the patients' tinnitus [8,9].

In patients with tinnitus resulting from hearing impairment, cochlear implants and hearing aids can significantly improve tinnitus, although variable methodologies applied in clinical trials complicate comparisons between the studies [5,10]. Despite short-term benefits from the intravenous application of lidocaine in some patients, no superiority compared to placebo was observed for commonly prescribed drugs such as betahistine or dexamethasone [8,11]. A study comparing the efficacy of ginkgo extracts and pentoxifylline to reduce tinnitus burden reported comparable decrease in the tinnitus burden for both products. However, reduction in the score of the mini Tinnitus Questionnaire (TQ) was below 10% for both study arms and thus most likely did not reach clinical relevance [12]. Currently, no medicinal products are approved for the treatment of chronic tinnitus. In contrast, pharmacological treatment may be required for comorbidities such as depression or anxiety. A number of therapeutic options have been developed for symptomatic treatment in patients without causative organic dysfunction or unsuccessful treatment. However, clinical evidence for treatments such as transcranial electric or vagus nerve stimulation, hyperbaric oxygen therapy, sound therapy, or tinnitus retraining therapy is scarce, and the evaluation of their efficacy in the clinical guidelines is ambiguous [10].

Among current symptomatic treatment options, cognitive behavioral therapy (CBT) has the highest clinical evidence compared to other treatment approaches. Its validity has been confirmed in several studies, and clinical guidelines recommend CBT for symptomatic treatment [13,14]. Although CBT does not target the tinnitus characteristics (eg, loudness), it focuses on perception and acceptance by the patients and thereby improves their quality of life and comorbidities. In recent years, app- or internet-based telemedicine has become a promising treatment alternative to conventional therapies. The application of CBT, both guided or nonguided by a psychologist or audiologist, is predestinated for remote therapy, and its efficacy to reduce the distress resulting from tinnitus has been proven over the years [15-17]. Nevertheless, more clinical trials are needed to strengthen the clinical evidence, and the research continues [16].

The randomized controlled clinical trial in this study investigates the long-term efficacy of an app-based CBT for tinnitus treatment. The primary endpoint was the difference in the TQ scores between patients who started CBT without delay (group 1) and a waiting list control group (group 2) who started CBT after 3 months, and the results have been previously published [18]. The baseline score of the patients in group 1 using the app improved from 39.7 by 13.4 points but remained largely unchanged (0.6 score points) in group 2. Following the 3 months waiting period, patients of group 2 received the CBT app as well, and patients of both groups continued treatment for a total duration of 9 months, after which final assessments of tinnitus severity (TQ score) as well as stress perception (20-item Perceived Stress Questionnaire [PSQ-20]), depression (9-item Patient Health Questionnaire [PHQ-9]), and self-efficacy (9-item Self-Efficacy, Optimism, Pessimism [SWOP-K9] questionnaire) were performed. This paper presents the long-term results of the full 9-month app-based treatment period in patients with tinnitus of both groups.

Methods

Our longitudinal analysis represents the outcome of the full 9-month treatment period with a CBT app, marketed under the name Kalmeda. The outcome of the initial 3-month period has been previously published [18].

Ethics Approval

The clinical investigation plan was prospectively approved by the Ethics Committee of North Rhine, Germany, on May 20, 2020 (case 2020026), following the ISO 14155 norm (Clinical Investigation of Medical Devices For Human Participants: Good Clinical Practice). This study was registered in the German Clinical Trials Register (DRKS-ID DRKS00022973) and conducted in 2 ENT (ear, nose, and throat) centers in North Rhine, Westphalia, Germany.

Patient Recruitment

Recruitment lasted from August 2020 to March 2021. The follow-up of the last patient ended in March 2022. The eligibility of the patients was assessed during a web-based visit. All patients fulfilled the inclusion criteria age ≥18 years and chronic subjective tinnitus of more than 3 months duration and not the exclusion criterium of an already diagnosed acute or chronic mental illness, whereby tinnitus-related anxiety or depression was no exclusion criterium. All patients signed an informed consent form prior to randomization to one of the treatment arms. The intervention group started using the app immediately after inclusion for a total of 9 months, while the control group commenced treatment after a 3-month waiting period.

XSL•FO
RenderX

Kalmeda Tinnitus App, Associated Regulations, and Study Design

This study investigates the efficacy of the Kalmeda app in the treatment of chronic tinnitus (Figure S1 in Multimedia Appendix 1). This app follows the principles of CBT, a self-help format following the Zürcher Ressourcen Modell (motivational psychology) and acceptance and commitment therapy for tinnitus [19]. The Kalmeda app was developed by mynoise GmbH and is a CE-marketed medical device since 2019 (DIMDI registration: DE/CA20/mynoise_01/18). During our study, software version 1.5.1 was used. The app was downloaded by users from commonly used app stores and activated by activation codes provided to the patients without a fee. The app has 5 levels with 9 steps each. The first 2 levels of behavioral therapy are typically accomplished within 3 months and the subsequent levels in not less than 7 months.

The efficacy of this intervention was assessed using a randomized controlled clinical trial by applying a waiting group design. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist was used for reporting of this randomized controlled trial (Multimedia Appendix 2). Patients were randomized to the 2 trial arms in blocks of 6 in the order of their appearance. Although the patients in group 1 started CBT immediately following randomization at T0 (baseline), patients in group 2 postponed treatment and began after the 3-month waiting period (Figure 1). In both groups, the efficacy of CBT was assessed after 3 and 9 months of treatment in comparison to baseline by using paper-based questionnaires sent to the patients by mail. Thus, the total study duration of the patients in group 1 and 2 was 9 and 12 months, respectively. The efficacy endpoints were the changes in tinnitus severity (TQ sum score) [20], depression scores (PHQ-9) [21], stress perception (PSQ-20) [22], and self-efficacy (SWOP-K9) [23] by using validated questionnaires. Published minimal clinically important difference (MCID) values of the TQ score depend on the baseline severity and can vary considerably between 12 for severe baseline values and <5 for lower baseline severity; this study considered the published MCID value of 6.65 [24-27]. The MCID of PHQ-9 varies with severity as well. In mild cases of depression characterized by PHQ-9 scores <10, corresponding to values reported in this study, the MCID is 1-2 points or 14%-22% of the baseline value [28,29]. No MCIDs are published for PSQ-20 and SWOP-K9; thus, generic improvements by 15% of the total scale range were considered as clinical important changes [27]. The severity of tinnitus was assessed according to 4 categories (<30, 31-46, 47-59, and 60-84) from mild to very severe, whereby it is considered as decompensated for TQ scores ≥47 [20]. Once the 3-month waiting period of the control group ended and app-based CBT was initiated, TQ sum scores significantly improved both statistically and clinically in individuals of the control group to a similar extent as in the intervention group. As data from baseline and after 3 and 9 months of treatment did not differ significantly between the groups, the data for all the patients using the app were pooled for our further analysis and evaluated collectively, independent from the initial randomization into group 1 or 2.

Figure 1. Scheme of the study design. Timepoints of assessments in group 1 and 2 (initial waiting group) are indicated by arrows. CBT: cognitive behavioral therapy.

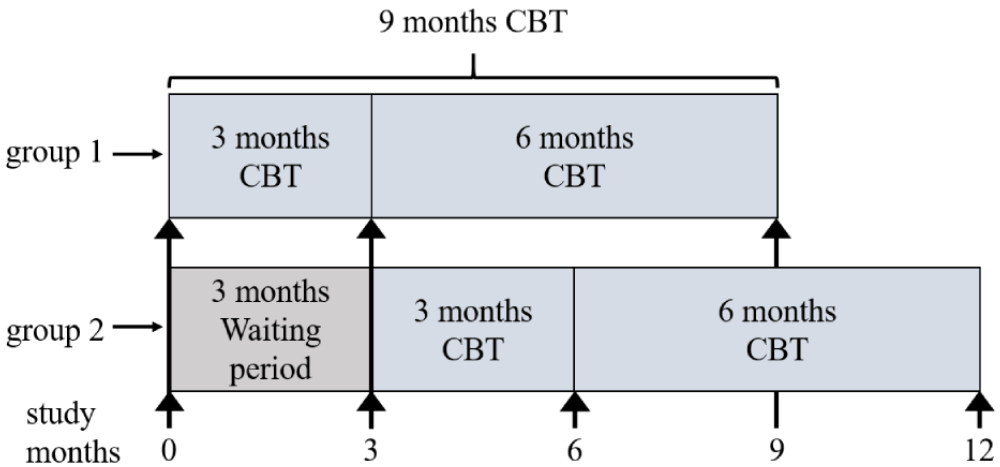

Statistical Analysis

Sample size calculations were based on an effect size of 0.5 and a power of 0.8 by using a 2-sided *t* test with an α level of .05. The assumed intermediate effect size is based on the range of efficacy of CBT reported in tinnitus studies [30-33]. This agreed to a sample size of 64 per group or 128 patients in total. Considering a dropout rate of 15%, a total of at least 150 patients needed to be enrolled. All statistical analyses were performed with SAS 9.4 (SAS Institute Inc).

The primary statistical analysis was performed by applying analysis of covariance (ANCOVA). Missing data were imputed using reference-based multiple imputations. Robustness was verified with a sensitivity analysis by using a completer data set, which encompassed only individuals without any missing answers in particular questionnaires over the whole period. Thus, the numbers of completers vary between the different endpoints (TQ: n=88, PHQ-9: n=86, PSQ-20: n=86, and SWOP-K9: n=91). The outcomes were analyzed in comparison to baseline by using the paired 2-sided *t* test. Furthermore, Pearson correlations were performed for changes of TQ and the

XSL•FO
RenderX

3 secondary scores. The distribution of the demographic and clinical characteristics of both groups at baseline were tested for normality by using Shapiro-Wilk test and for variance homogeneity by using Levene test. Depending on the outcome of the tests, group comparisons were performed using nonparametric (Mann-Whitney *U* test, chi-square) or parametric analyses (*t* test). Frequency changes in tinnitus severity categories were analyzed using chi-square tests. Values in the text are represented as mean (SD).

Results

Patient Demographics and Disposition

A total of 187 patients were randomized in blocks of 6 to group 1 or 2 (Table 1). Group 1 started using the app immediately after randomization, whereas patients in group 2 were assigned to a 3-month waiting period before treatment commenced. In group 1, 53 patients completed the treatment. In group 2, 42 patients completed the treatment. Thus, a total of 95 patients completed the study after 9 treatment months (Figure 2). At baseline, groups 1 and 2 did not differ with regard to mean age (48.2 years, SD 12.5 years), sex ratio (males, 97/187, 51.9%; females, 90/187, 48.1%), and duration of tinnitus (mean 6.57 years, SD 6.93 years). The mean duration of the app-based therapy was comparable in both study arms and was 99.3 (SD 14.4) and 288.1 (SD 29.0) days for the 3- and 9-month treatment period, respectively. Tinnitus distress was measured with TQ scores as the primary outcome. Secondary outcomes were measured with PSQ-20, PHQ-9, and SWOP-K9 questionnaires (Table 1).

Table 1. Demographic and clinical characteristics of the participants at baseline (N=187).

	Group 1 (n=94)	Group 2 (n=93)	*P* value
Participants			.27[a]
Males, n (%)	45 (48)	52 (56)	
Females, n (%)	49 (52)	41 (44)	
Age (years), mean (SD; range)	48.1 (12.8; 22-72)	48.4 (12.2; 21-74)	.86[a]
Tinnitus duration (years), mean (SD; range)	6.21 (6.62; 0.33-28)	6.94 (7.25; 0.25-31)	.79[b]
Tinnitus Questionnaire sum score, mean (SD; range)	39.65 (15.08; 12-73)	38.30 (15.10; 9-76)	.54[a]
Perceived Stress Questionnaire-20 items, mean (SD; range)	46.9 (17.6; 11.7-95)	45.1 (18.9; 11.7-83.3)	.51[a]
Patient Health Questionnaire-9 items, mean (SD; range)	8.31 (3.89; 2-19)	7.53 (4.21; 0-21)	.19[a]
Self-Efficacy, Optimism, Pessimism questionnaire-9 items, mean (SD; range)	2.76 (0.45; 1.8-4)	2.75 (0.52; 1.4-3.8)	.92[a]

[a]Paired 2-sided *t* test.

[b]*U* test.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram. Disposition of the patients in this trial. The term "completers" encompasses all patients with complete entries without any missing data in the Tinnitus Questionnaire. TQ: Tinnitus Questionnaire.

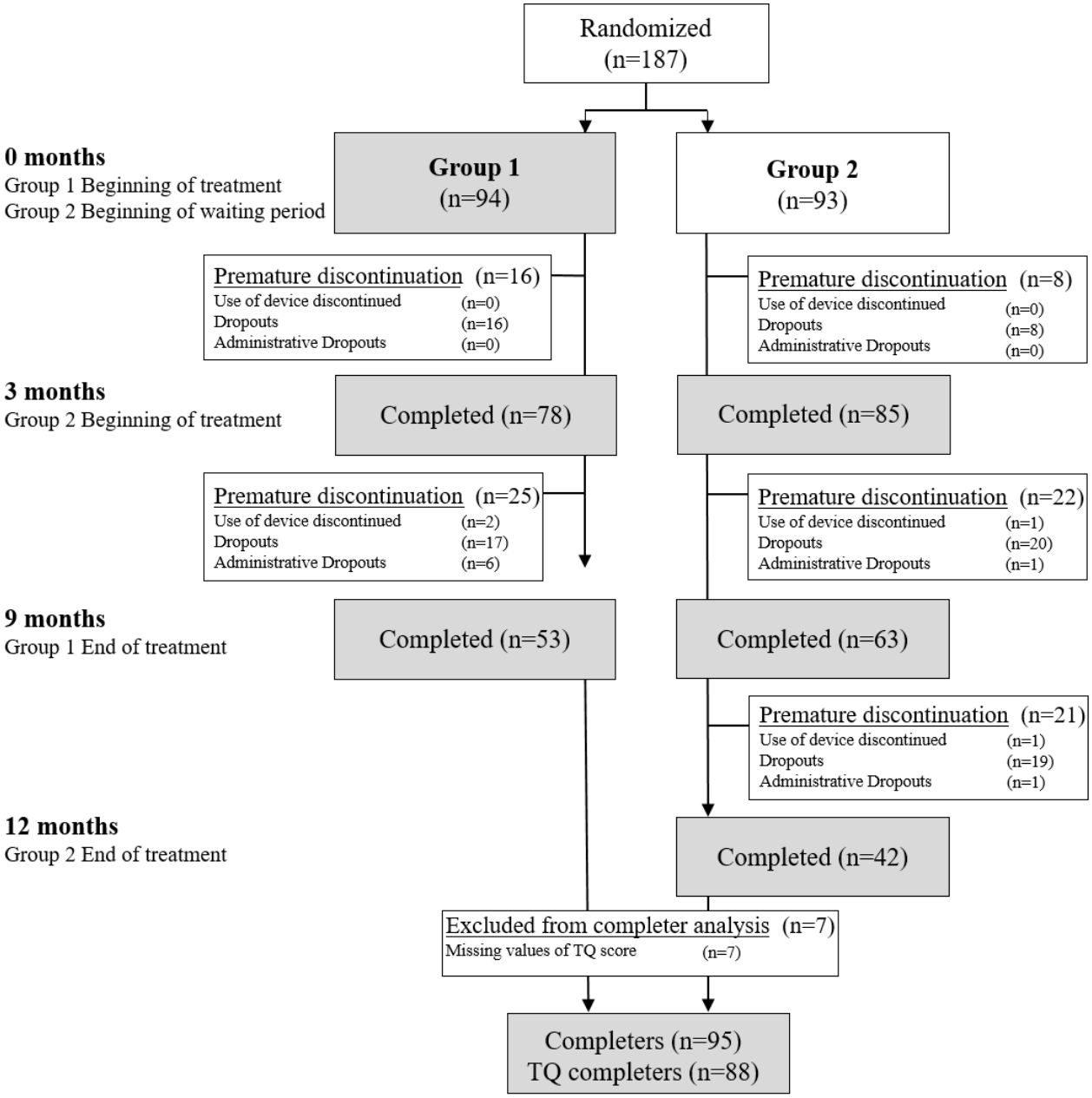

Primary Outcome

The app-based CBT significantly improved the subjective tinnitus burden, as assessed by the TQ sum scores in patients of both study arms after 3 and 9 months of treatment, whereas the TQ scores of patients of group 2 did not improve during the waiting period (Table 2, Figure 3). During treatment, the TQ sum scores of the completers declined significantly from 37.27 (SD 13.49) at baseline to 23.99 (SD 13.50) (paired t_{87}=10.75; P<.001) and 18.73 (SD 11.48) (paired t_{87}=13.0; P<.001) after 3 and 9 months of treatment, respectively.

XSL•FO
RenderX

Table 2. Tinnitus Questionnaire sum score of the completers at the beginning of the waiting period (group 2 only), beginning of treatment, and after 3 and 9 months of treatment.

Timepoint of the assessments	Group 1 (n=51), mean (SD)	Group 2 (n=37), mean (SD)	*P* value
Beginning of the waiting period	N/A[a]	37.8 (14.0)	N/A
Beginning of the treatment	37.1 (13.0)	37.5 (14.3)	.54[b]
3 months	23.4 (11.8)	24.8 (15.7)	<.001
9 months	17.6 (10.2)	20.3 (13.0)	<.001

[a]N/A: not applicable.

[b]Paired 2-sided *t* test.

Figure 3. Tinnitus questionnaire sum score of completers in group 1 (blue) and 2 (orange) during the waiting (group 2 only) and the app-based treatment period. For improved clarity of the treatment effect, the improvement during the app-based treatment period is shown in parallel for both groups. Significant improvements compared to baseline are indicated by stars. TQ: Tinnitus Questionnaire.

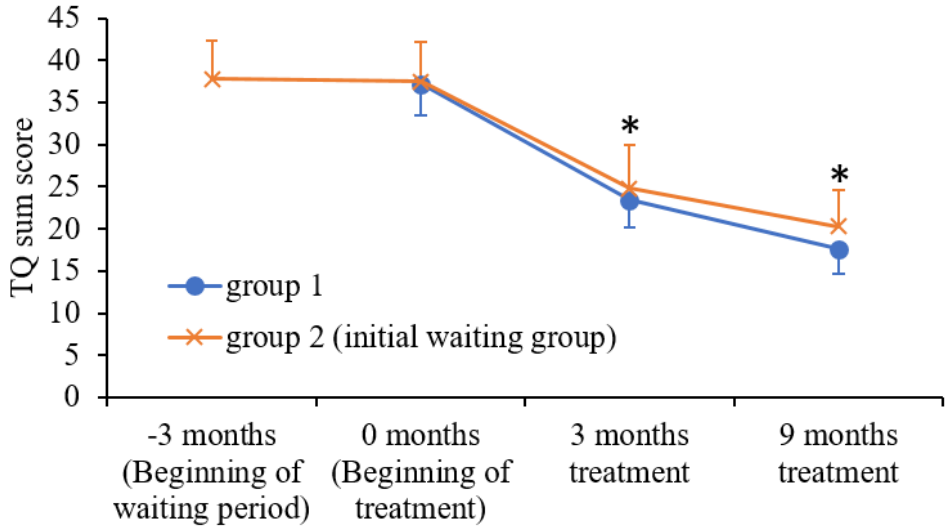

In agreement, the imputation of missing data by using multiple imputations resulted in initial TQ sum scores of 38.46 (SD 15.49), which improved to 25.97 (SD 14.03) after 3 months and to 19.98 (SD 9.32) after 9 months of CBT in the analysis of the completed courses corresponding to a total improvement of 12.49 (SD 9.65) after 3 months (*P*<.001) and 18.48 (SD 12.43) after 9 months (*P*<.001) during therapy for all the patients (Table 3). Thus, the prolonged treatment period of another 6 months resulted in a further and significant decrease of TQ scores.

Table 3. Analysis of covariance of treatment difference compared to baseline of Tinnitus Questionnaire sum scores after multiple imputations at 3 and 9 months after the beginning of the treatment.

Parameter	Estimate	SE	*P* value	95% confidence limits	
				Lower limit	Upper limit
Intercept	–0.99	2.00	.62	–4.94	2.94
ΔM3[a] (M3 – baseline)	–12.49	0.84	<.001	–14.14	–10.84
ΔM9[b] (M9 – baseline)	–18.48	1.08	<.001	–20.63	–16.33
Baseline	–0.45	0.05	<.001	–0.55	–0.35
ΔM9 – ΔM3	–5.99	1.16	<.001	–8.28	–3.70

[a]Score difference after 3 months of treatment compared to baseline.

[b]Score difference after 9 months of treatment compared to baseline.

The calculated effect size Cohen *d* was 1.15 and 1.38 after 3 and 9 months of therapy, respectively, corresponding to a strong effect, whereby improvement correlated with severity at baseline (*r*=–0.48) and after the first 3 months (*r*=–0.53) being much more pronounced in patients severely affected by tinnitus. The improvement of the mean TQ sum score was accompanied by significant reductions in the incidence of decompensated tinnitus (TQ sum score≥47). Among completers, the number of patients with decompensated tinnitus was reduced from the initial 22 to 7 after 3 months (χ^2_1=9.3; *P*=.002) and to only 1 patient after

XSL•FO
RenderX

9 months of treatment (χ^2_1=22.1; *P*<.001). Not very severe cases were remaining, but patients had improved to lower severity categories. In agreement, the number of patients characterized by mild tinnitus severity (TQ≤30) increased from 25 at baseline to 63 and 73 after 3 and 9 months, respectively (χ^2_1=32.8 and 53.1, respectively; *P*<.001; Figure S2 in Multimedia Appendix 1). After 3 months, 68% (60/88) responded to the treatment, and after 9 months of therapy, 80% (70/88) responded to the treatment, as their TQ sum score improved by more than the MCID value of 6.65. The baseline TQ score and its improvement during the first 3 months of therapy and its impact on completion or premature discontinuation of therapy was assessed. An initial worsening of symptoms does not seem to result in premature termination, as this was more often observed in completers (7 patients) than in dropouts (3 patients).

Individuals with mild and intermediate tinnitus (<46 points) at baseline documented average improvements by about 10 points (or 35% of baseline severity) within the first 3 months of treatment and thereby reached TQ scores of about 20 (Figure S3 in Multimedia Appendix 1). Approximately equal numbers of patients either terminated participation prematurely (55 patients) or completed treatment (66 patients). In contrast, patients with severe tinnitus (≥47) either experienced strong improvements of 18 points in the TQ score (35% of baseline severity) and completed treatment (22 patients) or small improvements of 10 points in the TQ score (equivalent to 18% improvement when compared to baseline severity), and treatment was terminated prematurely (13 patients).

Secondary Outcomes

Aside from the TQ sum score, the secondary outcomes, that is, stress, depression, and self-efficacy were assessed. The beneficial effects of our app-based CBT were observed in PSQ-20 scores over the whole treatment period. Significant reductions became evident already after 3 months (–4.62 points), and the score improved even further by –9.14 after 9 months of treatment (ANCOVA, *P*<.001). Similar significant changes (–5.66 and –9.55) were present in the completer data after 3 and 9 months as well (paired *t* test, *P*<.001), respectively, without reaching the clinically important difference. The effect sizes were 0.45 after 3 months and 0.60 after 9 months of CBT. Thus, the initially small effect became intermediate in the second treatment phase. Individual changes in PSQ-20 scores were well-correlated with TQ improvement after 3 (*r*=0.57) and 9 (*r*=0.58) months, respectively (Tables S1 and S2 in Multimedia Appendix 1).

The PHQ-9 score, as a measure of depression severity, decreased significantly by 1.40 and 2.47 points (ANCOVA, *P*<.001) after 3 and 9 months, respectively, as evaluated using multiple imputation. The sensitivity analysis using the completer data set yielded comparable significant results starting with 7.77 (SD 4.00) at baseline and lower values of 6.29 (SD 3.48) and 5.42 (SD 3.07) following treatment (paired *t* test, *P*<.001). Based on the published MCID, PHQ-9 changes were considered clinically relevant. The corresponding effect sizes were 0.52 and 0.68 after 3 months and 9 months, respectively, representing intermediate effects (Tables S1 and S2 in Multimedia Appendix 1). Changes in TQ and PHQ-9 were correlated after 3 (*r*=0.57) and 9 (*r*=0.58) months of therapy.

Self-efficacy was assessed using SWOP-K9. Imputed data revealed no significant improvement after 3 months (0.04, ANCOVA, *P*=.16) but after 9 months of treatment (0.17, ANCOVA, *P*<.001). Completers were characterized by mean baseline values of 2.77 (SD 0.51). Although no significant improvement was evident after 3 months (2.78, SD 0.54; paired *t* test *P*=.75), statistically significantly higher values of 2.95 (SD 0.47) were obtained after 9 months (paired *t* test, *P*<.001); however, the improvement did not reach the generic MCID of 15%. The 9-month effect size was 0.5 corresponding to an intermediate effect (Tables S1 and S2 in Multimedia Appendix 1). Only low correlations between individual changes of TQ score and SWOP-K9 were obtained after 3 months (*r*=–0.22), which increased after 9 months (*r*=0.54). No adverse events were reported during this clinical study.

Discussion

Principal Results

This study reports the outcome of a 9-month treatment with an app-based CBT for chronic tinnitus. This study is a follow-up to the previously published data of the initial 3-month treatment period [18]. The 9-month data confirm highly beneficial improvements in the tinnitus burden by CBT with the Kalmeda app. Within 3 months of treatment, the TQ sum scores improved significantly. There are currently 3 studies on the MCID of the TQ scores available. Adamchic and colleagues [24] reported an MCID of 6.65 (SD 7.99) points (mean follow-up time, 44 days) [24]. Hall and colleagues [25] reported an MCID of 12 (SD 9.8) points (follow-up time, 5 days) [25] and 7.5 points for the Mandarin version (follow up time, 6 months) [26]. After 3 months of web-based CBT, the average improvement in the tinnitus burden surpassed the average thresholds, and during the long-term treatment, tinnitus symptoms continued to improve even further, leading to an additional decrease by 5.99 points in the TQ score. Importantly, as perceived improvement of tinnitus is independent of the duration of treatment but depends on the baseline value, it is smaller in patients with lower initial burden [24]. The MCID for mild tinnitus (TQ≤30) corresponds to 3.17 (SD 7.15); this is much lower than the mean value of both MCID criteria [24,25]. After the first 3-month treatment period, patients were characterized by mean TQ scores of 23.99; thus, it is likely that further amelioration of their tinnitus symptoms experienced during the complete 9-month treatment is considered clinically relevant. Our study data reveal that the efficacy of the app-based CBT in the reduction of the tinnitus burden is comparable to that of treatments such as neuromodulation or transcranial direct current stimulation [6,34].

The comparison of active treatment to a waiting list control group is a common study design; yet, the assessment of the long-term outcome is difficult once the control group receives the same treatment. In this trial, the efficacy of CBT was compared to the waiting list control group, which remained largely unchanged during the initial 3-month period, but no control group was present during the full 9-month treatment period [18].

XSL•FO
RenderX

In order to compensate for the lack of an internal control, treatment efficacy in this study was compared to the outcomes of untreated control groups in randomized clinical trials with comparable durations. Few studies [35-39] have reported TQ sum scores in patients not receiving adequate treatment over several months. In a randomized trial testing the efficacy of tinnitus retraining therapy, no improvement was observed for patients in the waiting list control group after 12 months [35]. No improvement but a tendency of worsened symptoms was observed in patients in the waiting list control over a period of 6 months [36]. In patients receiving no treatment, the Tinnitus Severity Index was largely unchanged after 6 and 12 months, corresponding to an effect size of 0.0 and 0.1 [37]. In patients in the waiting list group during a 24-week trial, improvement in the TQ sum score corresponded to an estimated effect size below 0.4 [38]. Evidence from literature suggests only minor changes in tinnitus severity in untreated controls and a maximum effect size up to 0.4 corresponding to at most a small effect according to the definition by Cohen [39]. In comparison, the effect size of the app-based CBT group was 1.15 after 3 months and increased even further to 1.38 after 9 months of therapy. Thus, long-term app-based CBT in this trial resulted in pronounced improvement, which exceeded the expected outcome in the absence of any treatment in studies of comparable length by far. In addition, a waiting control group design does not allow to evaluate a potential placebo effect. In trials investigating medicinal products or medical devices, the efficacy of placebo treatment can be of relevance but at the same time is smaller by a factor of 3 than the improvement observed in the recent studies [40,41].

Dropouts can be a problem for the outcome of clinical trials. In this trial, 12.8% (24/187) of the patients stopped participation during the initial 3-month period until the primary endpoint was reached. In the following study period from 3 to 9 months, the dropout rates increased, resulting in a total loss of follow-up of 52.9% (99/187). A comparable rate of 51% was reported in a previous study applying internet-based CBT for tinnitus [15]. In general, dropout rates in studies using app-based treatments (~43%) often exceed rates common for studies applying in-person treatment [42]. However, despite the risk of therapy attrition, treatment apps are advantageous, for example, they are readily available without extended waiting periods, can be used 24/7, and are available even in remote areas.

Improvement of tinnitus-related distress within the first 3 months of app-based CBT did not differ significantly between dropouts and completers, suggesting that the patient's decision to prematurely end treatment was based on either treatment success (an individual had either reached an acceptable level of tinnitus severity or experienced considerable relief of its burden) or treatment failure (as defined by unsatisfying improvement or the burden for the treatment was too high). As in most cases, no further information on the reasons for dropouts were given; therefore, we cannot be sure about the reason with absolute certainty, but an analysis of the data after 3 months of treatment supports this hypothesis. Interestingly, in patients with mild and intermediate tinnitus (<46 points) who reached an average TQ score of about 20 within the first 3 months, subsequently, it was merely a 50/50 chance if they dropped out or completed treatment. Patients may have continued therapy even longer and thus reach even lower scores before prematurely dropping out; unfortunately, detailed information is not available. In contrast, patients with more severe burden at baseline did not reach such low TQ levels within 3 months. However, the majority of them experienced a pronounced and clinically relevant reduction of 18.9 points, which seemed to motivate continuation of therapy. Indeed, these patients finally reached TQ scores of 25.8 points after 9 months. A smaller number of 13 patients had improved to a lesser extent by 10 points (18% of baseline), thus still reaching the MCID and discontinued later. For the latter, it seems likely that unsatisfactory treatment was the major reason for the dropout. Thus, the data suggest that premature termination can result from subjective evaluation of success or failure of therapy depending on the relative improvement. The same may apply during the second study phase; however, the pronounced increase in dropouts suggests that the long study duration might have contributed to premature terminations as well. This is supported by the higher dropout rate of patients from the waiting list control group (51/93, 55%) whose involvement in the trial lasted for a full year when compared with the intervention group (41/94, 44%) who finished treatment 3 months earlier.

Patients with tinnitus are often characterized by elevated depression and stress perception, which was measured with PHQ-9 as well as PSQ-20, respectively [43,44]. Following the 3-month treatment, PHQ-9 and PSQ-20 measures had significantly improved. Both parameters decreased even further during the prolonged treatment, and improvements became clinically relevant. As suggested by Wallhauser-Franke and colleagues [45], the measures of tinnitus loudness and tinnitus distress need to be distinguished. The app-based CBT in our study did not improve perception of tinnitus loudness but significantly reduced tinnitus distress. Improvement of PHQ-9 and PSQ-20 scores after CBT therefore are in parallel to the observed reduction of the TQ sum score observed after 3 and 9 months of treatment.

In contrast to the 2 previous secondary endpoints, self-efficacy assessed by SWOP-K9 was almost unchanged following 3 months of CBT, but a statistically significant improvement was evident after 9 months, albeit clinically relevant levels were not reached. This highlights the importance of prolonged treatment beyond the initial 3 months period, leading to significant improvements of tinnitus burden on all tested scales.

Limitations

This study has limitations. The waiting list design allows an accurate quantification of treatment effects of CBT in comparison with the untreated control but lacks an internal control for the whole 9-month treatment period. Thus, the superiority of the treatment had to be verified by comparison to studies in this indication with comparable duration. Furthermore, a potential placebo effect was not assessed and therefore cannot be ruled out, but is likely to be relatively small [40,41]. The tinnitus distress of the included study participants was assessed with an average TQ sum score of 39.6 in group 1 and 38.3 in group 2. Patients with this amount of tinnitus distress are considered as grade 2 patients (out of 4 grades). This study

XSL•FO
RenderX

shows that patients with chronic tinnitus of this grade of severity can significantly benefit from the app-based treatment. At the current state of research, the results cannot be generalized to patients with very severe amount of tinnitus distress (grade 4).

Conclusions

This study reveals the pronounced effect of an app-based CBT on tinnitus distress, and in the course of treatment, the number of patients affected by decompensated tinnitus decreased significantly. In addition, the majority were characterized by mild tinnitus severity during the final assessment.

Acknowledgments

The study was funded by mynoise GmbH.

Authors' Contributions

UW and SP conceived the study design. SP and LB supervised the study. UW and LB performed the study. WS, KR-D, TW, and JT analyzed the data and wrote the manuscript. All coauthors contributed to the revision of the manuscript and approved the final version.

Conflicts of Interest

KR-D, TW, and JT are employees of G Pohl-Boskamp GmbH & Co KG (distribution partner for Kalmeda). WS, UW, and LB received personal fees from G Pohl-Boskamp GmbH & Co KG. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 120 KB - jmir_v27i1e59575_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 791 KB - jmir_v27i1e59575_app2.pdf]

References

1. Jarach CM, Lugo A, Scala M, van den Brandt PA, Cederroth CR, Odone A, et al. Global prevalence and incidence of tinnitus: a systematic review and meta-analysis. JAMA Neurol 2022 Sep 01;79(9):888-900 [FREE Full text] [doi: 10.1001/jamaneurol.2022.2189] [Medline: 35939312]
2. Biswas R, Genitsaridi E, Trpchevska N, Lugo A, Schlee W, Cederroth CR, et al. Low evidence for tinnitus risk factors: a systematic review and meta-analysis. J Assoc Res Otolaryngol 2023 Feb;24(1):81-94 [FREE Full text] [doi: 10.1007/s10162-022-00874-y] [Medline: 36380120]
3. Pastorková N, Holý R, Procházková K, et al. Problematika roztrou ené sklerózy v otoneurologické ambulanci, mezioborová spolupráce otoneurologa a neurologa. Otorhinolaryngology & Phoniatrics. 2021. URL: https://tinyurl.com/3ajdh3v2 [accessed 2025-01-30]
4. Langguth B, Elgoyhen AB, Cederroth CR. Therapeutic approaches to the treatment of tinnitus. Annu Rev Pharmacol Toxicol 2019 Jan 06;59:291-313. [doi: 10.1146/annurev-pharmtox-010818-021556] [Medline: 30044727]
5. Kikidis D, Vassou E, Markatos N, Schlee W, Iliadou E. Hearing aid fitting in tinnitus: a scoping review of methodological aspects and effect on tinnitus distress and perception. J Clin Med 2021 Jun 29;10(13):2896 [FREE Full text] [doi: 10.3390/jcm10132896] [Medline: 34209732]
6. Conlon B, Langguth B, Hamilton C, Hughes S, Meade E, Connor CO, et al. Bimodal neuromodulation combining sound and tongue stimulation reduces tinnitus symptoms in a large randomized clinical study. Sci Transl Med 2020 Oct 07;12(564):eabb2830. [doi: 10.1126/scitranslmed.abb2830] [Medline: 33028707]
7. Lu T, Wang Q, Gu Z, Li Z, Yan Z. Non-invasive treatments improve patient outcomes in chronic tinnitus: a systematic review and network meta-analysis. Braz J Otorhinolaryngol 2024;90(4):101438 [FREE Full text] [doi: 10.1016/j.bjorl.2024.101438] [Medline: 38788246]
8. Langguth B, Kleinjung T, Schlee W, Vanneste S, De Ridder D. Tinnitus guidelines and their evidence base. J Clin Med 2023 Apr 24;12(9):3087 [FREE Full text] [doi: 10.3390/jcm12093087] [Medline: 37176527]
9. Molnár A, Mavrogeni P, Tamás L, Maihoub S. Correlation between tinnitus handicap and depression and anxiety scores. Ear Nose Throat J 2022 Nov 08:1455613221139211 [FREE Full text] [doi: 10.1177/01455613221139211] [Medline: 36346819]
10. Cima RFF, Mazurek B, Haider H, Kikidis D, Lapira A, Noreña A, et al. A multidisciplinary European guideline for tinnitus: diagnostics, assessment, and treatment. HNO 2019 Mar;67(Suppl 1):10-42. [doi: 10.1007/s00106-019-0633-7] [Medline: 30847513]

XSL•FO
RenderX

11. Chung J, Lee DY, Kim JS, Kim YH. Effectiveness of intratympanic dexamethasone injection for tinnitus treatment: a systematic review and meta-analysis. Clin Exp Otorhinolaryngol 2022 Feb;15(1):91-99 [FREE Full text] [doi: 10.21053/ceo.2021.01459] [Medline: 35144363]
12. Procházková K, Šejna I, Skutil J, Hahn A. Ginkgo biloba extract EGb 761 versus pentoxifylline in chronic tinnitus: a randomized, double-blind clinical trial. Int J Clin Pharm 2018 Oct;40(5):1335-1341 [FREE Full text] [doi: 10.1007/s11096-018-0654-4] [Medline: 29855986]
13. Fuller T, Cima R, Langguth B, Mazurek B, Vlaeyen J, Hoare D. Cognitive behavioural therapy for tinnitus. Cochrane Database Syst Rev 2020 Jan 08;1(1):CD012614 [FREE Full text] [doi: 10.1002/14651858.CD012614.pub2] [Medline: 31912887]
14. Meijers S, Stegeman I, van der Leun JA, Assegaf SA, Smit AL. Analysis and comparison of clinical practice guidelines regarding treatment recommendations for chronic tinnitus in adults: a systematic review. BMJ Open 2023 Sep 15;13(9):e072754 [FREE Full text] [doi: 10.1136/bmjopen-2023-072754] [Medline: 37714675]
15. Andersson G, Strömgren T, Ström L, Lyttkens L. Randomized controlled trial of internet-based cognitive behavior therapy for distress associated with tinnitus. Psychosomatic Medicine 2002;64(5):810-816. [doi: 10.1097/00006842-200209000-00014]
16. Demoen S, Chalimourdas A, Timmermans A, Van Rompaey V, Vanderveken OM, Jacquemin L, et al. Effectiveness of telerehabilitation interventions for self-management of tinnitus: systematic review. J Med Internet Res 2023 Feb 09;25:e39076 [FREE Full text] [doi: 10.2196/39076] [Medline: 36757768]
17. W Beukes E, Andersson G, Fagelson M, Manchaiah V. Internet-based audiologist-guided cognitive behavioral therapy for tinnitus: randomized controlled trial. J Med Internet Res 2022 Feb 14;24(2):e27584 [FREE Full text] [doi: 10.2196/27584] [Medline: 35156936]
18. Walter U, Pennig S, Kottmann T, Bleckmann L, Röschmann-Doose K, Schlee W. Randomized controlled trial of a smartphone-based cognitive behavioral therapy for chronic tinnitus. PLOS Digit Health 2023 Sep;2(9):e0000337 [FREE Full text] [doi: 10.1371/journal.pdig.0000337] [Medline: 37676883]
19. Westin VZ, Schulin M, Hesser H, Karlsson M, Noe RZ, Olofsson U, et al. Acceptance and commitment therapy versus tinnitus retraining therapy in the treatment of tinnitus: a randomised controlled trial. Behav Res Ther 2011 Nov;49(11):737-747. [doi: 10.1016/j.brat.2011.08.001] [Medline: 21864830]
20. Goebel G, Hiller W. Psychische Beschwerden bei chronischem Tinnitus: Erprobung und Evaluation des Tinnitus-Fragebogens (TF). Verhaltenstherapie 1992;2(1):13-22 [FREE Full text] [doi: 10.1159/000258202]
21. Grafe K, Zipfel S, Herzog W, Löwe B. Screening for psychiatric disorders with the Patient Health Questionnaire (PHQ). Results from the German validation study. Diagnostica 2004;50(4):11. [doi: 10.1026/0012-1924.50.4.171]
22. Fliege H, Rose M, Arck P, Walter OB, Kocalevent R, Weber C, et al. The Perceived Stress Questionnaire (PSQ) reconsidered: validation and reference values from different clinical and healthy adult samples. Psychosom Med 2005;67(1):78-88. [doi: 10.1097/01.psy.0000151491.80178.78] [Medline: 15673628]
23. Scholler G, Fliege H, Klapp BF. [Questionnaire of self-efficacy, optimism and pessimism: reconstruction, selection of items and validation of an instrument by means of examinations of clinical samples]. Psychother Psychosom Med Psychol 1999 Aug;49(8):275-283. [Medline: 10488648]
24. Adamchic I, Tass P, Langguth B, Hauptmann C, Koller M, Schecklmann M, et al. Linking the Tinnitus Questionnaire and the subjective Clinical Global Impression: which differences are clinically important? Health Qual Life Outcomes 2012 Jul 10;10:79 [FREE Full text] [doi: 10.1186/1477-7525-10-79] [Medline: 22781703]
25. Hall DA, Mehta RL, Argstatter H. Interpreting the Tinnitus questionnaire (German version): what individual differences are clinically important? Int J Audiol 2018 Jul;57(7):553-557. [doi: 10.1080/14992027.2018.1442591] [Medline: 29490515]
26. Gu H, Diao C, Tang J, Yang B, Ji J, Zhou M, et al. Minimal clinically important difference for the Mandarin version of the Tinnitus Questionnaire determined via anchor-based and distribution-based methods. Health Qual Life Outcomes 2022 Nov 30;20(1):157 [FREE Full text] [doi: 10.1186/s12955-022-02072-z] [Medline: 36451205]
27. Langguth B, De Ridder D. Minimal clinically important difference of tinnitus outcome measurement instruments—a scoping review. J Clin Med 2023 Nov 15;12(22):7117 [FREE Full text] [doi: 10.3390/jcm12227117] [Medline: 38002730]
28. Bauer-Staeb C, Kounali D, Welton NJ, Griffith E, Wiles NJ, Lewis G, et al. Effective dose 50 method as the minimal clinically important difference: Evidence from depression trials. J Clin Epidemiol 2021 Sep;137:200-208 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.04.002] [Medline: 33892086]
29. Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, et al. How much change is enough? Evidence from a longitudinal study on depression in UK primary care. Psychol. Med 2020 Nov 03;52(10):1875-1882. [doi: 10.1017/s0033291720003700]
30. Andersson G. Clinician-supported internet-delivered psychological treatment of tinnitus. Am J Audiol 2015 Sep;24(3):299-301. [doi: 10.1044/2015_aja-14-0080]
31. Rief W, Weise C, Kley N, Martin A. Psychophysiologic treatment of chronic tinnitus: a randomized clinical trial. Psychosomatic medicine 2005;67(5):833-838. [doi: 10.1097/01.psy.0000174174.38908.c6]
32. Weise C, Kleinstäuber M, Andersson G. Internet-delivered cognitive-behavior therapy for tinnitus: a randomized controlled trial. Psychosomatic medicine 2016;78(4):501-510. [doi: 10.1097/psy.0000000000000310]

XSL•FO
RenderX

33. Hesser H, Weise C, Westin VZ, Andersson G. A systematic review and meta-analysis of randomized controlled trials of cognitive-behavioral therapy for tinnitus distress. Clin Psychol Rev 2011 Jun;31(4):545-553. [doi: 10.1016/j.cpr.2010.12.006] [Medline: 21237544]
34. Heiland LD, Owen JM, Nguyen SA, Labadie RF, Lambert PR, Meyer TA. Neuromodulation for treatment of tinnitus: a systematic review and meta-analysis. Otolaryngol Head Neck Surg 2024 May;170(5):1234-1245. [doi: 10.1002/ohn.671] [Medline: 38353342]
35. Caffier PP, Haupt H, Scherer H, Mazurek B. Outcomes of long-term outpatient tinnitus-coping therapy: psychometric changes and value of tinnitus-control instruments. Ear Hear 2006 Dec;27(6):619-627. [doi: 10.1097/01.aud.0000240504.77861.1a] [Medline: 17086074]
36. Schmidt A, Lins U, Wetscher I, Welzl-Müller K, Weichbold V. [Counselling versus cognitive group therapy for tinnitus. A retrospective study of their efficacy]. HNO 2004 Mar;52(3):242-247. [doi: 10.1007/s00106-003-0930-y] [Medline: 15007518]
37. Henry JA, Loovis C, Montero M, Kaelin C, Anselmi K, Coombs R, et al. Randomized clinical trial: group counseling based on tinnitus retraining therapy. J Rehabil Res Dev 2007;44(1):21-32 [FREE Full text] [doi: 10.1682/jrrd.2006.02.0018] [Medline: 17551855]
38. Kreuzer PM, Goetz M, Holl M, Schecklmann M, Landgrebe M, Staudinger S, et al. Mindfulness-and body-psychotherapy-based group treatment of chronic tinnitus: a randomized controlled pilot study. BMC Complement Altern Med 2012 Nov 28;12:235 [FREE Full text] [doi: 10.1186/1472-6882-12-235] [Medline: 23186556]
39. Cohen J. Statistical power analysis for the behavioural sciences. Academic Press 1969:567. [doi: 10.4324/9780203771587]
40. Walters RK, Durrant FG, Nguyen SA, Meyer TA, Lambert PR. The placebo effect on tinnitus: a systematic review and meta-analysis of randomized controlled trials. Otol Neurotol 2024 Feb 15;45(4):e263-e270. [doi: 10.1097/mao.0000000000004139]
41. Mares T, Albrecht J, Buday J, Podgorna G, Le TH, Magyarova E, et al. Long-term effect of transcranial direct current stimulation in the treatment of chronic tinnitus: A randomized, placebo-controlled trial. Front Psychiatry 2022;13:969800 [FREE Full text] [doi: 10.3389/fpsyt.2022.969800] [Medline: 36311525]
42. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
43. Biehl R, Boecking B, Brueggemann P, Grosse R, Mazurek B. Personality traits, perceived stress, and tinnitus-related distress in patients with chronic tinnitus: support for a vulnerability-stress model. Front Psychol 2019;10:3093 [FREE Full text] [doi: 10.3389/fpsyg.2019.03093] [Medline: 32116880]
44. Hackenberg B, Döge J, O'Brien K, Bohnert A, Lackner KJ, Beutel ME, et al. Tinnitus and its relation to depression, anxiety, and stress-a population-based cohort study. J Clin Med 2023 Feb 01;12(3):1169 [FREE Full text] [doi: 10.3390/jcm12031169] [Medline: 36769823]
45. Wallhäusser-Franke E, Brade J, Balkenhol T, D'Amelio R, Seegmüller A, Delb W. Tinnitus: distinguishing between subjectively perceived loudness and tinnitus-related distress. PLoS One 2012;7(4):e34583 [FREE Full text] [doi: 10.1371/journal.pone.0034583] [Medline: 22529921]

Abbreviations

ANCOVA: analysis of covariance
CBT: cognitive behavioral therapy
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ENT: ear, nose, and throat
MCID: minimal clinically important difference
PHQ-9: 9-item Patient Health Questionnaire
PSQ-20: 20-item Perceived Stress Questionnaire
SWOP-K9: 9-item Self-Efficacy, Optimism, Pessimism questionnaire
TQ: Tinnitus Questionnaire

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 16.04.24; peer-reviewed by R Holy, N Gninenko; comments to author 14.09.24; revised version received 17.10.24; accepted 23.12.24; published 18.02.25.

Please cite as:
Walter U, Pennig S, Bleckmann L, Röschmann-Doose K, Wittig T, Thomsen J, Schlee W
Continuous Improvement of Chronic Tinnitus Through a 9-Month Smartphone-Based Cognitive Behavioral Therapy: Randomized Controlled Trial
J Med Internet Res 2025;27:e59575
URL: https://www.jmir.org/2025/1/e59575
doi:10.2196/59575
PMID:

©Uso Walter, Stefan Pennig, Lothar Bleckmann, Kristina Röschmann-Doose, Thomas Wittig, Jörn Thomsen, Winfried Schlee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Depression Self-Care Apps' Characteristics and Applicability to Older Adults: Systematic Assessment

Ruoyu Yin[1], MPhil; Dakshayani Rajappan[2], BSocSci; Laura Martinengo[1,3], MD, PhD; Frederick H F Chan[1], MPhil; Helen Smith[1,4], PhD; Konstadina Griva[1], PhD; Mythily Subramaniam[1,5], MD, PhD; Lorainne Tudor Car[1,6], MD, PhD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

[2]School of Social Sciences, Nanyang Technological University Singapore, Singapore, Singapore

[3]Centre for Behavioural and Implementation Sciences Interventions, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[4]School of Medicine, Keele University, Staffordshire, United Kingdom

[5]Research Division, Institute of Mental Health, Singapore, Singapore

[6]Department of Primary Care and Public Health, School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom

Corresponding Author:
Lorainne Tudor Car, MD, PhD
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road, Clinical Sciences Building
Singapore, 308232
Singapore
Phone: 65 65138572
Email: l.tudor.car@imperial.ac.uk

Abstract

Background: Depression affects 32% of older adults. Loneliness and social isolation are common risk factors for depression in older adults. Mobile apps can connect users and are also effective in depression management in the general population. However, older adults have specific needs in terms of the content of depression self-care interventions and their accessibility. It remains unknown whether existing apps for depression self-care are applicable to older adults.

Objective: The initial aim of this assessment was to systematically identify interactive depression self-care apps specifically designed for older adults. As we did not find any, we assessed the applicability of existing depression self-care apps to the needs of older adult users.

Methods: Using an established app assessment methodology, we searched for Android and iOS interactive mental health apps providing self-care for depression in English and Chinese in the 42Matters database, Chinese Android app stores, and the first 10 pages of Google and Baidu. We developed an assessment rubric based on extensive revision of the literature. The rubric consisted of the following sections: general characteristics of the apps (eg, developer, platform, and category), app content (eg, epidemiology and risk factors of depression in older adults, techniques to improve mood and well-being), and technical aspects (eg, accessibility, privacy and confidentiality, and engagement).

Results: We identified 23 apps (n=19, 82.6%, English and n=4, 17.4%, Chinese apps), with 5 (21.7%) iOS-only apps, 3 (13%) Android-only apps, and 15 (65.2%) apps on both platforms. None specifically targeted older adults with depression. All apps were designed by commercial companies and were free to download. Most of the apps incorporated cognitive behavior therapy, mood monitoring, or journaling. All but 3 (13%) apps had a privacy and confidentiality policy. In addition, 14 (60.9%) apps covered depression risk factors in older adults, and 3 (13%) apps delivered information about depression epidemiology in older adults via a chatbot. Furthermore, 17 (73.9%) apps mentioned other topics relevant to older adults, such as pain management, grief, loneliness, and social isolation. Around 30% (n=7) of the apps were supported by an online forum. Common accessibility issues included a lack of adaptations for users with visual or hearing impairments and incompatibility with larger font sizes in the phone settings.

Conclusions: There are no depression apps developed specifically for older adults. Available mobile apps have limited applicability to older adults in terms of their clinical and technical features. Depression self-care apps should aim to incorporate content relevant

XSL•FO
RenderX

to older adults, such as grief and loss; include online communities; and improve accessibility to adapt to potential health impairments in older adults.

(J Med Internet Res 2025;27:e56418) doi:10.2196/56418

KEYWORDS

older adults; elder; elderly; mental health; mental illness; mental disorders; mHealth; mobile health; mobile application; app; application; smartphone; depression; self-care; mobile apps; systematic assessment; assessment; effectiveness; self-care; Android app; mental health apps; mobile interventions; behaviour; therapy; mood monitoring; adaptations; online communities; impairments

Introduction

The population is aging worldwide, and this demographic shift is coupled with higher depression incidence rates in older adults compared to younger populations [1]. China, for example, had 254 million adults over 60 years old in 2019 [2], with a growing incidence of mental disorders and a high risk of suicide [3-5]. In 2020, the number of older adults in the United States was 55.8 million, representing 16.8% of the total population [6]. It was estimated that 8 million older adults in the United States aged 65 years and above had depression, and depression affects 32% of older adults worldwide [7,8]. Older adults with depression are more likely to have comorbidities, worse sleep, decreased quality of life, and poorer self-perceived health compared to those without depression [9,10]. There are many barriers to older adults seeking mental health treatment, including low awareness, which may lead to a poor prognosis [11]. There is a need for effective, scalable, accessible, and affordable mental health interventions that are tailored to the needs of older adults. Digital health interventions, particularly those delivered via mobile phones, have the potential to address this burden of mental illness in older adults.

Older adults are increasingly using mobile phones, which facilitates using mobile apps to improve their mental health. In the United States, the ownership of smartphones between 2017 and 2022 in adults aged 65 years and above increased from 40% to 65% [12]. In China, over a third of the internet users in 2021 were adults aged 50 years or older [13]. The need to minimize in-person contact during the COVID-19 pandemic also boosted the use of digital technology in older adults. For example, in Singapore, more than half of 55-75-year-olds used social media platforms or the internet to search for news of the pandemic [14]. Mobile apps also provide users with an opportunity to stay connected and may potentially mitigate loneliness and social isolation [15], which are common risk factors for depression in older adults [16].

Contrary to some common beliefs, older adults are becoming increasingly tech-savvy, with some preferring to use digital health tools for mental health treatment over face-to-face treatment [17-21]. Furthermore, digital mental health interventions can help reduce the long waiting times for counseling and avoid the time and financial costs of travel from rural areas [22]. In addition, the privacy provided by digital mental health tools may reduce the potential stigma and shame, which is common among older adults [22-26]. Mental health mobile apps have been shown to be effective in the treatment of depression in the general population [27,28]. Although there were mental health apps specifically targeting the youth [29], it remains unknown whether there are apps that cater for older adults' specific needs in terms of mental health content and digital accessibility [30,31]. For example, worldwide, around 51% of older adults aged 60 years and above have multimorbidity [32] and would benefit from content on physical health, cognitive decline, stress, aging, and isolation [33,34]. They may also need bigger fonts, and louder and clearer audio to address potential visual or hearing impairments [35].

Therefore, the initial aim of this assessment was to systematically identify interactive depression self-care apps specifically designed for older adults. As we did not find any, we assessed the applicability of existing depression self-care apps to the needs of older adult users. We focused on English and Chinese app markets as they represent 31% of the world population [36].

Methods

Study Design

We used an app assessment methodology that has been extensively used in previous papers aiming to systematically assess other types of apps [37-40]. The assessment methodology derives from systematic review methodology as it includes a systematic search for apps with a clear mention of the search strategy and databases searched, clear inclusion/exclusion criteria, 2-step screening for eligible apps, and systematic data extraction using a predefined data extraction form. Our method was mostly aligned with a guide on the systematic review and evaluation of mobile health apps [41].

Search Strategy

We searched for apps in English and Chinese in the 2 leading app markets, Android and iOS, in November-December 2022. We searched Apple's App Store for English and Chinese apps and Android apps in English using 42Matters, a comprehensive proprietary app database. As Google Play is unavailable in mainland China, we searched the most popular Android Chinese app stores, used by more than 81% of mobile phone users, including Tencent App Store, Huawei App Market, MIUI App Store, Oppo Software Store, VIVO App Store, 360 Mobile Assistant, Baidu Mobile Assistant, MM App Store, PP Assistant, and the Wandoujia platform [42]. In addition, we screened the first 10 pages on popular English and Chinese search engines (Google and Baidu) for any other potentially relevant depression self-care–focused mobile apps [43]. Search terms for apps in English included "mental health," "mindfulness," "cognitive behavioral therapy," "CBT," "psychotherapy," "online therapy," "mood," "depress," "anxiety," "sadness," "melancholia," "worry," and "counseling." When we combined the

XSL•FO
RenderX

aforementioned search terms with “older adults,” “elderly,” “ageing,” or “geriatric” using the AND Boolean operator, no results were returned on 42Matters. Therefore, we decided to broaden the search strategy and manually capture aging-related content in the apps’ descriptions, if any. The search terms for Chinese apps included 抑郁 OR 焦虑 OR 心理健康 OR 心理咨询 OR 行为疗法 OR 精神健康 OR 正念 OR 自杀 OR 悲伤.

Eligibility Criteria

Textbox 1 shows the inclusion and exclusion criteria of the apps. We included interactive mental health apps that explicitly mentioned depression, were free or required one-time payment to download, and targeted older adults or the general population. Popular apps were included because users are more likely to encounter and use them [44]. As there was no consensus on the criteria related to popularity, previous studies have adopted the strategy of excluding apps with less than 10 [45] or 5 [43,46] reviews. Meanwhile, highly downloaded apps were defined as those with 100,000 or more downloads in Android markets [47]. Therefore, Android apps with ≥100,000 downloads on Google Play or Chinese Android app markets and iOS apps with ≥10 reviews on Apple’s App Store were included. Apps available on both platforms were included if their number of downloads was ≥100,000 in Android markets and they had ≥10 reviews on Apple App Store. Apps that (1) targeted children, adolescents, or young adults; (2) focused exclusively on meditation or mindfulness, positivity, affirmations, or well-being; (3) had technical issues; (4) only allowed consultations with health care providers; or (5) had <100,000 downloads on Google Play or Chinese Android app markets or <10 reviews on Apple’s App Store were excluded.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- Targeted older adults or the general population
- Explicitly provided depression self-care
- Included interactive features (ie, “allowing information to be passed continuously and in both directions between a computer or other device and the person who uses it” [47,48])
- Were in English or Chinese
- Were available on Google Play, Chinese Android stores, or Apple’s App Store
- Were found in the following app store categories: health and fitness, lifestyle, and medical
- Were released or updated from May 2021 onwards
- Were free or paid to download
- Had ≥100,000 downloads on Google Play or Chinese Android app markets and ≥10 reviews on Apple’s App Store

Exclusion criteria:

- Targeted children, adolescents, or young adults
- Focused exclusively on meditation or mindfulness, positivity, affirmations, or well-being in general
- Had technical issues and could not be used after 2 attempts, required an access code provided by a health care institution or insurance company, or had been removed from the app store at the time of the assessment
- Only allowed online consultations with health care providers
- Had <100,000 downloads on Google Play or Chinese Android app markets or <10 reviews on Apple’s App Store

App Screening

In the first round of screening, 1 reviewer (author RY) followed the inclusion and exclusion criteria in Textbox 1 to manually screen the apps on ASReview, a screening tool using machine learning algorithms [49] that has been increasingly used in systematic reviews [50-53]. After each decision (include/exclude) by the reviewer, ASReview applied machine learning models (term frequency–inverse document frequency method as the feature extractor and logistic regression as the classifier) and actively sorted the rest of the apps according to their relevance, from the most relevant to the least relevant. As such, ASReview facilitated screening by allowing the reviewer to screen a subset of the total apps. As shown in prior simulation studies, ASReview was able to identify 95% of all relevant records while screening between 8% and 33% of the total records [49]. Since there is no consensus on the stopping rule [54], the reviewer in this study predefined the stopping criteria as screening one-third (33.3%) of the total apps [49]. The apps were then downloaded to an iPhone 13 (iOS 16.1.1), a Samsung A52s (Android 13, One UI 5.1), and 2 of 3 reviewers (authors RY, DR, and FC) independently and in parallel evaluated eligibility. Apps available on both iOS and Android platforms were counted as one app in the assessment.

Assessment Rubric Development, Data Extraction, and Analysis

To develop an assessment rubric for content relating to older adults, we searched the following keywords on Google Scholar and PubMed: aging, depression clinical practice guidelines, geriatric depression, psychoeducation or psychotherapy for older adults, and accessibility guidelines. We also conducted a

XSL•FO
RenderX

qualitative systematic review of older adults' views and experiences with using digital mental health interventions [55] and a scoping review of experts' opinions on digital mental health interventions for older adults [56]. Our scoping review confirmed the essential content and features for older adults, including therapeutic approaches, relevant topics, personalization, and accessibility.

Table 1 shows the sources of evidence used in the development of the assessment rubric. The sources included established clinical practice guidelines [57-60], systematic and scoping reviews [61-64], academic books [65-69], a systematic assessment of mental health apps [70], and guidelines and checklists of technical features [71-73], including the Mobile Application Rating Scale (MARS), the mobile Health On the Net Code (mHONcode), and the World Wide Web Consortium (W3C) guidelines for accessibility of content on mobile phones.

Both MARS and mHONcode were designed to evaluate apps' technical features, while neither of them was specifically designed to assess apps' applicability to older adults. Furthermore, most of the included apps have been evaluated using these scales elsewhere [45,75-77]. However, existing app assessments have seldom evaluated accessibility in detail. Thus, in this study, we focused on assessing app accessibility using the W3C guidelines, and we used a few items from MARS and mHONcode that do not overlap with the W3C guidelines and are directly relevant to older adults, as indicated in a systematic review on the views and experiences of older adults with digital mental health interventions [55]. According to our review, engagement with the target group (older adults), the quality/resolution of graphics with regard to aesthetics, information credibility, and the evidence base from MARS are relevant and important to older adults [78-80]. In addition, older adults often have concerns about privacy and confidentiality when using digital mental health tools [23,33,78,80-84]. Therefore, we supplemented the W3C guidelines with 3 groups of items from MARS (target group, aesthetics, and information) and 2 items from mHONcode (privacy and confidentiality).

The following information about each app was recorded using a data extraction form (Multimedia Appendix 1) with 3 sections:

- General features of apps, including their description and category in the app store, developers, ratings, target user group, regions, languages, privacy policy, total downloads in Android markets, and cost of download and subscription. Ratings of apps ranged from 1 to 5, with higher scores indicating users' positive perceptions.
- Content relevant to older adults, which consisted of the following 6 sections: symptoms and natural history of depression in older adults (11 binary questions with yes/no answers and 3 open-ended questions), screening of depression (7 binary questions and 5 open-ended questions), self-care techniques (2 binary questions and 4 open-ended questions), personalization (5 binary questions and 1 open-ended question), human involvement (3 binary questions and 1 open-ended question), and other functions (6 binary questions and 2 open-ended questions).
- Technical features, including the W3C guidelines for accessibility of content on mobile phones [35,73], supplemented by 5 items covering engagement, aesthetics, and information from MARS [71,85] and 2 items on privacy and confidentiality from mHONcode [72]. The W3C guidelines consisted of 4 principles: perceivable (6 binary items), operable (6 items), understandable (7 items), and robust (4 items).

For every item except MARS, a "yes" was assigned 1 point. Each app was independently evaluated by 2 of the 3 reviewers working in parallel (RY, DR, and FC). Assessments were then compared from September to October 2023, and disagreements were discussed until consensus was reached. We used each app for at least 15 minutes to ensure interaction with all app features. Data were recorded in a Microsoft Excel worksheet and summarized. We calculated the percentage of each app's score over its total score. Narrative data synthesis was used to present our findings.

XSL•FO
RenderX

Table 1. Evidence base of the assessment rubric.

Type of evidence	Details/topics
Clinical practice guidelines	• The National Institute for Health and Care Excellence in the UK [57] • The American Psychiatric Association [74] • The Royal Australian and New Zealand College of Psychiatrists [60] • The Ministry of Health in Singapore • The Japanese Society of Mood Disorders [58] • The Indian Psychiatric Society [59]
Systematic reviews or scoping reviews	• Mobile health apps for older adults [63] • Depression risk factors in older adults [64] • Influencing factors of digital technology usage among older adults [61,62] • Older adults' views and experiences of digital mental health interventions [55] • Experts' opinions on digital mental health interventions for older adults [56]
Academic books	• Geropsychology [67] • Psychotherapy with older adults [69] • Aging [68] • CBT[a] [65,66]
Systematic assessment of apps	• Self-guided CBT apps for depression [70]
Guidelines and checklists of technical features	• MARS[b] [71] • W3C[c] guidelines for accessibility of content on mobile phones [35,73] • mHONcode[d] [72]

[a]CBT: cognitive behavioral therapy.

[b]MARS: Mobile Application Rating Scale.

[c]W3C: World Wide Web Consortium.

[d]mHONcode: mobile Health On the Net Code.

Ethical Considerations

Ethical approval was not required for this systematic assessment of mobile apps as no human participants or personal data were involved.

Results

Identification of Apps for Depression Self-Care

We identified 5177 English apps (n=2408, 46.5%, on Android and n=2769, 53.5%, on iOS) in 42Matters and 4652 Chinese apps (n=1039, 22.3%, on Android and n=3613, 77.7%, on iOS) in 42Matters and Chinese app stores. One additional English app, Happify, was identified on Google and none from the Baidu search engine. After title and description screening, 101 (2%) English apps (n=39, 38.6%, on Android and n=62, 61.4%, on iOS) and 37 (0.8%) Chinese apps (n=29, 78.4%, on Android and n=8, 21.6%, on iOS) were downloaded to examine eligibility. We included 23 (16.8%) apps for systematic assessment (n=19, 82.6%, English apps and n=4, 17.4%, Chinese apps. In addition, 15 (65.2%) apps were available on both iOS and Android platforms, 5 (21.7%) on the iOS platform, and 3 (13%) on Android platforms (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of app screening.

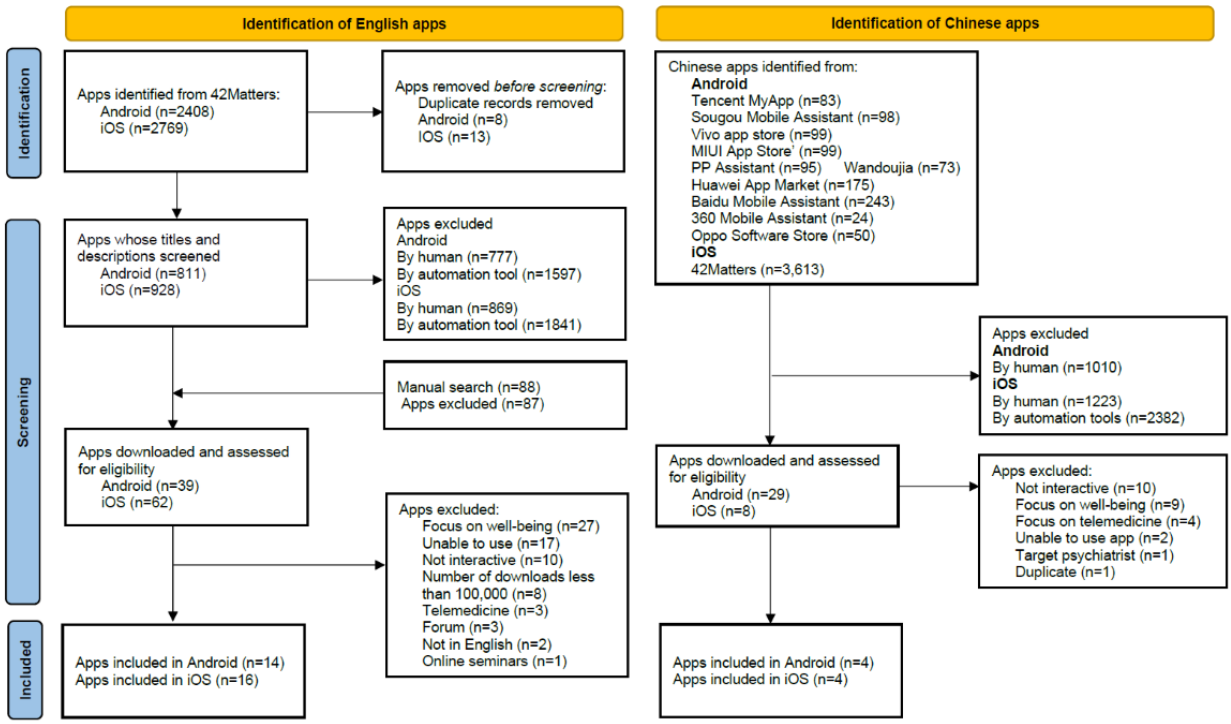

Basic Characteristics of the Included Apps

Table 2 shows the basic characteristics of the included apps. All 23 apps were developed by commercial companies and were free to download. Most of the apps (n=20, 87%) were in the health and fitness category in the app stores, and 3 (13%) were in the medical category. Of the 18 (78.3%) apps available on Android platforms, 8 (44.4%) had been downloaded more than 1 million times; unfortunately, Apple does not provide iOS download data. In addition, 16 (69.6%) apps had in-app purchases for subscriptions or courses or counseling services. Of the 19 (82.6%) English apps, 4 (21.1%) were available in other languages (Chinese, German, French, Italian, Japanese, Portuguese, Spanish, Japanese, and Korean). One of the four Chinese apps supported English and Arabic.

Table 3 shows the ratings of the apps on each platform. Most of the apps available on iOS (n=19, 95%) had a rating of 4-5, indicating good user experience in general, while 1 (5%) app had a rating of 3-4. Similarly, 2 (11.1%) of the 18 apps available on Android had a rating of 3-4, 13 (72.2%) had a rating of 4-5, while ratings were unavailable for 3 (16.7%) apps.

XSL•FO
RenderX

Table 2. Basic characteristics of the mental health apps included in this study (N=23).

Basic characteristics	Apps, n (%)
Affiliation	
Commercial	23 (100.0)
Others (academic institution, government, nongovernmental organization)	0
Category in the app store	
Health and fitness	20 (87.0)
Medical	3 (13.0)
Number of downloads on Android markets	
100,000-1 million	10 (55.6)
>1 million	8 (44.4)
Cost	
Free	7 (30.4)
Free with in-app purchase	16 (69.6)
Paid	0
Primary language	
English	19 (82.6)
Chinese	4 (17.4)

Table 3. Ratings and number of ratings for the apps included in this study.

Rating details	iOS apps (n=20)	Android apps (n=18)
Ratings		
1-3	0	0
3-4	1 (5.0)	2 (11.1)
4-5	19 (95.0)	13 (72.2)
No ratings	0	3 (16.7)
Number of ratings		
<100	5 (25.0)	1 (5.6)
100-1000	6 (30.0)	1 (5.6)
>1000	9 (45.0)	13 (72.2)
No ratings	0	3 (16.7)

Applicability of the Content of the Included Apps on Older Adults' Depression Self-Care Needs

Figure 2 presents a summary of app content related to older adults. All apps mainly targeted the general population instead of older adults. The most common self-care technique was mindfulness or meditation (n=22, 95.7%), while 19 (82.6%) apps used CBT. Other types of techniques included acceptance and commitment therapy (n=4, 17.4%), behavioral activation (n=2, 8.7%), positive psychology (n=2, 8.7%), interpersonal therapy (n=2, 8.7%), problem-solving therapy (n=1, 8.7%), music therapy (n=1, 4.3%), and dialectical behavior therapy (n=1, 4.3%). In addition, 2 (8.7%) apps offered 2 types of techniques other than mindfulness/meditation and CBT. More than half (n=17, 73.9%) of the included apps mentioned topics relevant to older adults, such as pain management, grief, loneliness, and social isolation. Most of the apps (n=20, 87%) provided self-monitoring of mood or allowed users to keep journals, while 10 (43.5%) apps asked users to track other behaviors, such as exercise, diet, sleep, and medication, in addition to mood and journaling. Each app's self-care techniques and self-monitoring activities are shown in Multimedia Appendix 2.

Furthermore, 3 (13%) apps described depression epidemiology in older adults, 2 (66.7%) of which provided this information in conversations with generative artificial intelligence (AI) chatbots. In addition, 14 (60.9%) of the apps reported common risk factors of depression in older adults, such as chronic diseases, loss of family members, and loneliness, and 4 (17.4%) apps described depressive symptoms in older adults. Regarding the psychosocial and health-related dynamics of aging, 5 (21.7%) apps discussed how aging impacts life. When creating a new account in 1 (4.3%) app, there was a question about what

XSL•FO
RenderX

diseases users have. More than two-thirds of the apps (n=18, 78.3%) addressed the stigma of depression by normalizing mental disorders, providing affirmations to users, and encouraging them to acknowledge and accept the illness. However, none of the apps shared a personal story of an older adult recovering from depression. Multimedia Appendix 3 shows the total score of each app on the 9 items of educational content. The mean score was 2.7 (range 1-5), indicating the scarcity of content relevant to older adults.

Figure 3 presents the main features of the included apps, such as the assessment of depression, comorbid diseases, etc. Of the 23 apps, 17 (73.9%) provided depression assessment using validated tools, such as the Patient Health Questionnaire-9 (n=7, 41.2%). One Chinese app used the Geriatric Depression Scale, an assessment scale targeted at older adults, and also provided assessment tools for dementia, cognitive impairment, and functional ability (activities of daily living). Of the 17 (73.9%) apps with depression assessment tools, 14 (82.4%) administered the assessment once, while the other 3 (17.6%) offered regular assessment.

Furthermore, 7 (30.4%) apps had forums that allowed users to communicate with each other, and 9 (39.1%) apps provided access to counselors using video or audio calls, text messages, or emails. On 3 (75%) of the 4 (17.4%) Chinese apps, users could book face-to-face appointments with a counselor. The process for accessing counselors could not be evaluated on 1 (5.3%) English app because the feature was limited to India and we were Singapore based. In addition, 3 (13%) of the 23 apps allowed users to contact or seek help from their social network directly through the apps by entering the contact details of those people, while 14 (60.9%) apps included emergency resources, such as suicide prevention helplines, and users could be redirected to these services from the apps. Multimedia Appendix 4 shows each app's main features.

Figure 2. Content related to older adults. *Other types of techniques included acceptance and commitment therapy (n=4, 17.4%), behavioral activation (n=2, 8.7%), positive psychology (n=2, 8.7%), interpersonal therapy (n=2, 8.7%), problem-solving therapy (n=1, 4.3%), music therapy (n=1, 4.3%), and dialectical behavior therapy (n=1, 4.3%). In addition, 2 (8.7%) apps offered 2 types of techniques other than mindfulness/meditation and CBT, respectively; CBT: cognitive behavioral therapy.

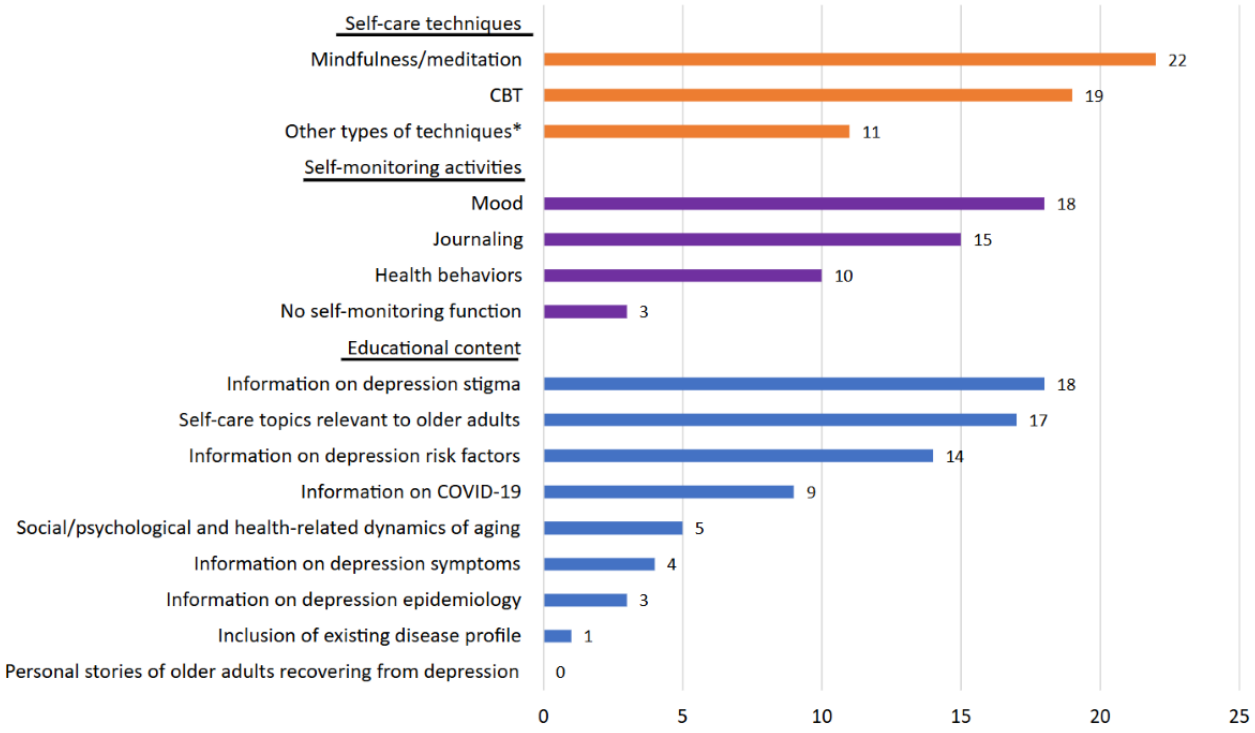

XSL•FO
RenderX

Figure 3. Main features of the apps included in this study.

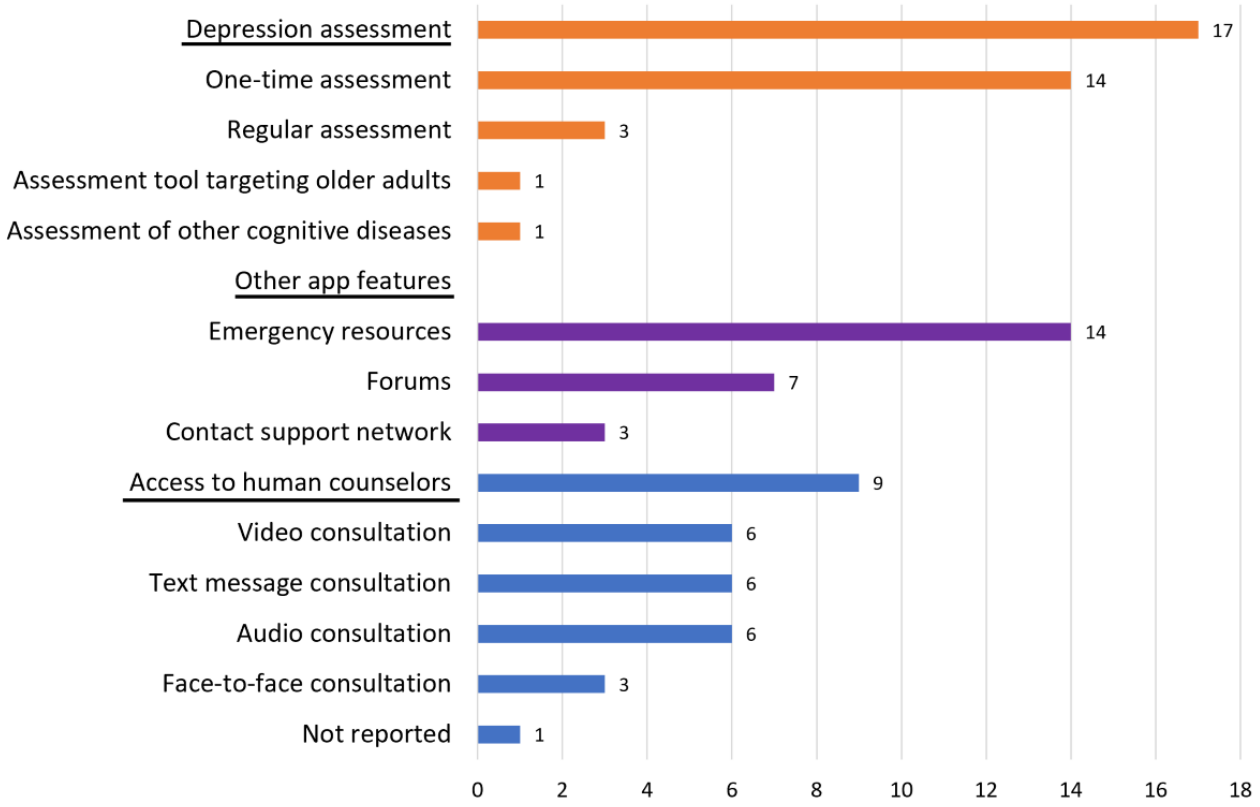

Comparison of Chinese and English Apps Based on Content

There were similarities and differences between the 4 (17.4%) Chinese and 19 (82.6%) English apps. The similarities were the inclusion of self-care techniques (eg, CBT and mindfulness, mood tracking, and journaling functions) and the reporting of depression risk factors in older adults. All English apps targeted users with mental health needs, but 3 (75%) of the 4 (17.4%) Chinese apps also included educational content for mental health providers. Perhaps for this reason, the Chinese apps contained more information (eg, mental disorder encyclopedia) and functions (eg, cognitive function assessment) than the English apps. Although all the included apps were free to download, 12 (63.2%) English apps had a paid version, where users could access all content in the apps with monthly or yearly subscriptions, while 1 (5.3%) English app and 3 (75%) Chinese apps adopted a different commercial mode by offering specific courses or assessments addressing various mental health issues with separate purchases.

Applicability of the Technical Features of the Included Apps to Older Adults' Needs

All apps offered some degree of personalization, including addressing the user by name, recording their preferences for dark mode and haptic feedback, and providing tailored suggestions and individualized responses from AI-assisted chatbots. Of the 23 apps, 15 (65.2%) had notifications to remind users to use the apps. In addition, 21 (91.3%) apps included information about contacting customer support in the case of technical issues, questions, or suggestions. Multimedia Appendix 5 shows detailed information about these technical features of each app.

Table 4 shows the selected items of MARS and mHONcode. Detailed information about these items of each app can be found in Multimedia Appendix 6. Using MARS, both involvement and aesthetics had a mean score of 3.8 (SD 0.4) out of 5.0. The average score for information credibility was 2.9 (SD 0.5) out of 5. Furthermore, 6 (26.1%) apps were evaluated in feasibility studies [86-88] or randomized controlled trials (RCTs) [89-91] and had positive outcomes related to participants' mental health. None of the Chinese apps has yet been evaluated. However, the study participants mainly targeted the general adult population aged 18 years and above [86,87,89,91,92]. With regard to mHONcode, all but 3 (13%) apps had a privacy and confidentiality policy. Of the 20 (87%) apps with a privacy and confidentiality policy, 2 (10%) had inaccessible policies.

XSL•FO
RenderX

Table 4. Selected items of MARS[a] and mHONcode[b].

Technical features	Score (out of 5)
MARS, mean (SD)	
Engagement	3.8 (0.4)
Aesthetics	3.9 (0.6)
Information credibility	2.9 (0.5)
Evidence base of information	3.2 (1.0)
mHONcode, n (%)	
Privacy and confidentiality policy	20 (87.0)
Privacy policy accessible within the app	18 (78.3)

[a]MARS: Mobile Application Rating Scale.

[b]mHONcode: mobile Health On the Net Code.

Accessibility Features

Multimedia Appendix 7 shows each app's accessibility feature and the total score. The mean percentage of each app's score over its total score was 65.5% (range 58.6%-77.8%), suggesting moderate accessibility and areas for improvement. Figure 4 shows the accessibility of the included apps. Most of the apps (n=20, 87%) considered the smartphone's small screen size and avoided delivering too much information on 1 page. The contrast between the text and the background color was visible in all apps (n=23, 100%), and 4 (17.4%) apps allowed users to change the foreground and background colors. All apps could be viewed by scrolling from top to bottom. Users would be able to identify elements relevant to their needs, and the content was accessible to people with color blindness. However, only 1 (4.3%) app allowed zooming or magnification. Of the 20 apps with video or audio content, a small proportion (n=4, 20%) provided text alternatives, while 2 (10%) provided captions or subtitles for all video and audio content. In addition, only 1 AI chatbot, Hector, could read out responses and provided users with the option to call rather than type.

The apps were generally easy to operate (ie, with an appropriate touch target size and spacing and easy touchscreen gestures). When connected to a Bluetooth keyboard, 3 (13%) of the 23 apps could be fully controlled by the keyboard, while 20 (87%) apps were partially or totally unresponsive to keyboard commands. Most of the apps did not restrict the time for completing a task, and 2 (50%) of 4 (17.4%) apps with a timed task allowed users to pause. Users could change the content playback speed in embedded videos in 7 (30.4%) apps, including YouTube videos in 4 (57.1%) apps.

All apps had a consistent layout and easy-to-recognize actionable buttons. Most of the apps (n=21, 91.3%) showed important elements regardless of page scroll. Although none of the apps provided instructions for custom touchscreen or device manipulation gestures, 3 (13%) apps from one company (Excel At Life) included a detailed tutorial explaining the meaning of each icon. Only 1 (4.3%) app allowed both portrait and landscape screen orientations, and 8 (34.8%) apps provided ample opportunity for users to check the terms and conditions and review them before making a payment.

Data entry on the apps was easy, supported by the error correction feature of the system's default keyboard, and 13 (56.5%) apps provided user-friendly data entry methods by adjusting the keyboard according to the type of data. For example, a keyboard with "@" and ".com" was provided for entering email addresses. Lastly, although the phone's operating system allows for fonts to be adjusted, only 9 (39.1%) apps supported this feature, with an additional app having adjustable font sizes within itself.

XSL•FO
RenderX

Figure 4. Accessibility of the apps included in this study.

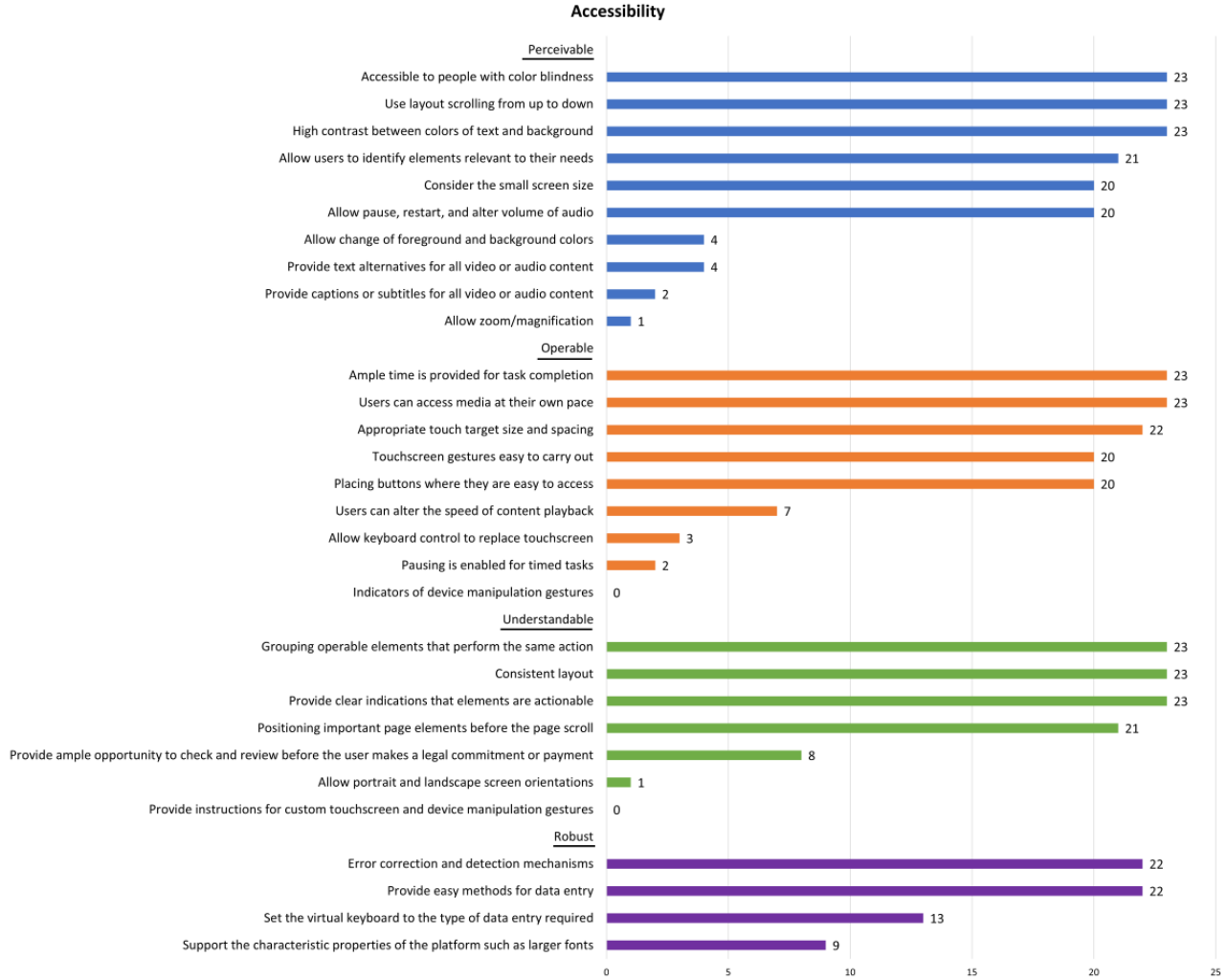

Discussion

Principal Findings

This systematic assessment of mental health apps is the first to have evaluated the applicability of existing mental health apps for depression self-care in older adults. We identified a total of 23 apps on Chinese and English app markets, but none of them specifically targets older adults with depression. A few apps provide information about depression specific to older adults or incorporated assessment tools for cognitive disorders or functional capacity. The level of accessibility, such as adaptations for users with visual or hearing impairments and adjustable font sizes, remains low, potentially limiting older adults' use of the apps.

There is a lack of mental health apps to assist depression self-care in older adults. This gap in the current app market is somewhat at odds with other health care areas, as there are mobile health apps with medication reminders, apps for chronic disease management, apps promoting physical activity, and apps providing cognitive training activities [93]. Research has shown that a lack of consideration for older adults' needs may reduce their perceived relevance and motivation to use such mobile apps [33]. However, a systematic review in 2023 showed that chatbots were promising in improving older adults' depression and insomnia symptoms [94]. Although the use of generative AI chatbots in health care is still at an early stage, and its safety remains untested [95], the development of specific content for older adults must not be neglected.

We found fewer interactive Chinese depression self-care apps than English apps. Two recent systematic assessments of mental health apps in China identified 67 apps [76] and 172 apps [96]. These apps connected users with health care providers or provided noninteractive mental health education, but they frequently lacked self-care techniques [96] and thus were excluded from our study. Furthermore, despite the management of severe mental disorders being included in basic public health services in China, there are persisting challenges due to the lack of psychiatric professionals [97] and the traditional stigma associated with mental disorders [98]. Therefore, more interactive Chinese applications with self-care techniques are needed to support Chinese older adults' mental health care demands.

We documented potential accessibility issues for older adults in the apps identified. According to qualitative studies on older adults' experience with digital mental health interventions, the lack of detailed instructions is a barrier to their use [78,83]. Older adults have less experience with digital technology and are, at times, anxious about how to use it [99]. Therefore, clear

XSL•FO
RenderX

instructions and guidance on how to use mobile apps are essential for many older adults, but we found only 3 of the 23 apps included a detailed tutorial for app navigation, suggesting a widespread lack of consideration for users with less technology literacy. Worldwide, around 1 in 10 adults over 50 years had a visual impairment in 2019 [100]. Hearing loss is also common in older adults, affecting over 50% of older males and 30% of older females aged 65 years and above [99]. Therefore, digital health interventions need to cater to their needs. We found that few apps with video or audio content provide text substitutes or captions or offer adjustable font sizes or are compatible with the accessibility features of larger text in the phone settings.

Around 30% of the depression self-care apps had an online forum, which may help reduce loneliness and increase social connections. A recent editorial highlighted the impact of loneliness and social isolation on adverse mental health outcomes, such as depression and anxiety [101]. Due to physical impairments, older adults are at a higher risk of loneliness and social isolation [102]. Online communities or forums could bring people with similar interests or circumstances together and provide them with a convenient opportunity to communicate from the comfort of their homes. A scoping review in 2022 found that online social networking improves older adults' mental health and well-being by enhancing communication with their existing social network, increasing life satisfaction, and reducing depressive symptoms [103]. These communities may address the absence of interpersonal interactions in digital interventions by allowing older adults to provide each other emotional support, such as affirmation and encouragement, and technical advice to use developing technology [103-105].

Strengths and Limitations

We used a consistent and reliable approach to app identification and assessment. We performed the search in Apple's App Store, Google Play, and various Chinese Android app markets and adopted an AI screening tool that helped increase the efficiency by screening one-third of the names and descriptions of apps. The AI tool's ability to retrieve 95% of the eligible records by screening 8%-33% of the total records was validated in simulation studies [49]. To capture all eligible apps, we further supplemented the search using Google and Baidu search engines. Two reviewers independently performed an in-depth assessment of apps by downloading and testing them on research phones. Second, there was no validated scale for the assessment of mobile apps for depression self-care in older adults. We developed an assessment rubric by synthesizing evidence from the extensive literature, including established clinical practice guidelines, systematic reviews, scoping reviews, academic books, and widely used guidelines or checklists of technical features.

However, this study also has some limitations. First, older adults with lived experience were not involved in the assessment, which calls for public and patient involvement in future studies targeting older adults. Another potential weakness is the exclusion of apps focused on well-being without explicitly addressing depression: we acknowledge that they might still be useful for older adults with mental disorders. However, the provision of information about depression is perhaps superior as it helps older adults gain awareness and knowledge [34] and reduces mental health stigma [106]. Third, we did not include apps in other languages due to the limited capacity of research manpower and resources.

The study has limited implications for clinical application in older adults because we did not identify any existing apps with high applicability to older adults. However, there are many implications for app developers and researchers in the field of digital mental health. As some older adults were skeptical of the credibility of apps [84], it is crucial for apps to clearly demonstrate their validity, reliability, and accuracy of information to enhance older adults' trust. Three of the apps reviewed were clearly evaluated in the form of an RCT [89-91], and others may have been, but there was no mention of authors and sources, weakening the trust, real or perceived, in the evidence base of the existing apps. Although apps cannot replace clinical diagnosis and treatment, and users need to seek professional advice when necessary, self-care apps may still serve as an interim solution for older adults when they are waiting for clinical appointments [107], thereby providing them with early access to mental health resources and promoting their mental health. Future studies should aim to evaluate the clinical effectiveness of apps and their potential to lessen the strain on health care professionals.

Conclusion

Although there are many depression self-care apps available in English and Chinese markets, our study suggests that they have limited applicability to older adults. The apps integrating generative AI chatbots have higher applicability because of tailored feedback and the provision of a text-to-speech option. In the future, new and updated depression self-care apps should consider incorporating content relevant to older adults, including online communities to reduce their social isolation, and improving accessibility for older adults with physical and sensory impairments.

Acknowledgments

We would like to thank Ms Lin Xiaowen for helping with screening Chinese iOS apps.

This study received funding from the Singapore Ministry of Education (MOE) Tier 1 Grant RT05/21. The funder played no role in the design or conduct of the study.

Data Availability

The data supporting the findings of this study are available within the paper and its Multimedia Appendices.

XSL•FO
RenderX

Authors' Contributions

LTC and LM conceptualized the study. RY and LM developed the review protocol. RY, DR, and FC conducted screening and data extraction. RY conducted analyses and drafted the manuscript. LM, LTC, HES, KG, and MS reviewed and revised the manuscript. All authors have approved the final version of the manuscript.

Conflicts of Interest

LTC is an editorial board member for JMIR Medical Education. Other authors declare no conflicts of interest.

Multimedia Appendix 1
Data extraction form.
[DOCX File , 48 KB - jmir_v27i1e56418_app1.docx]

Multimedia Appendix 2
Self-care techniques and self-monitoring activities of each app.
[DOCX File , 28 KB - jmir_v27i1e56418_app2.docx]

Multimedia Appendix 3
Educational content of each app.
[DOCX File , 29 KB - jmir_v27i1e56418_app3.docx]

Multimedia Appendix 4
Main features of each app.
[DOCX File , 31 KB - jmir_v27i1e56418_app4.docx]

Multimedia Appendix 5
Technical features of each app.
[DOCX File , 18 KB - jmir_v27i1e56418_app5.docx]

Multimedia Appendix 6
Selected items of MARS (Mobile Application Rating Scale) and mHONcode (mobile Health On the Net Code) for each app.
[DOCX File , 27 KB - jmir_v27i1e56418_app6.docx]

Multimedia Appendix 7
Accessibility of each app.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v27i1e56418_app7.xlsx]

References

1. Wu Y, Fan L, Xia F, Zhou Y, Wang H, Feng L, et al. Global, regional, and national time trends in incidence for depressive disorders, from 1990 to 2019: an age-period-cohort analysis for the GBD 2019. Ann Gen Psychiatry 2024 Aug 02;23(1):28 [FREE Full text] [doi: 10.1186/s12991-024-00513-1] [Medline: 39095916]
2. Ageing and health in China. World Health Organization. 2023. URL: https://www.who.int/china/health-topics/ageing [accessed 2025-02-10]
3. Li B, Zhang G, Ma J, Kang M. Mortality rate of mental disorder trends in China from 2002 to 2020. Front Psychiatry 2022 Nov 15;13:1039918 [FREE Full text] [doi: 10.3389/fpsyt.2022.1039918] [Medline: 36458125]
4. Collaborators GBDMD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/s2215-0366(21)00395-3]
5. Chen X, Giles J, Yao Y, Yip W, Meng Q, Berkman L, et al. The path to healthy ageing in China: a Peking University–Lancet Commission. Lancet 2022 Dec;400(10367):1967-2006. [doi: 10.1016/s0140-6736(22)01546-x]
6. Caplan Z. U.S. older population grew from 2010 to 2020 at fastest rate since 1880 to 1890. U.S. Census Bureau. 2023 May 25. URL: https://www.census.gov/library/stories/2023/05/2020-census-united-states-older-population-grew.html [accessed 2025-02-10]
7. Zenebe Y, Akele B, W/Selassie M, Necho M. Prevalence and determinants of depression among old age: a systematic review and meta-analysis. Ann Gen Psychiatry 2021 Dec 18;20(1):55 [FREE Full text] [doi: 10.1186/s12991-021-00375-x] [Medline: 34922595]

XSL•FO
RenderX

8. Global Health Data Exchange (GHDx) 2022. Institute of Health Metrics and Evaluation. URL: https://vizhub.healthdata.org/gbd-results/?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b [accessed 2025-02-10]
9. Rodda J, Walker Z, Carter J. Depression in older adults. BMJ 2011 Sep 28;343(sep28 1):d5219-d5219. [doi: 10.1136/bmj.d5219] [Medline: 21957206]
10. Mental health of older adults. World Health Organization. 2023 Oct 20. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2025-02-10]
11. Subramaniam M, Abdin E, Vaingankar JA, Shafie S, Chua HC, Tan WM, et al. Minding the treatment gap: results of the Singapore Mental Health Study. Soc Psychiatry Psychiatr Epidemiol 2020 Nov 17;55(11):1415-1424 [FREE Full text] [doi: 10.1007/s00127-019-01748-0] [Medline: 31317246]
12. Faverio M. Share of those 65 and older who are tech users has grown in the past decade 2022. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2025-02-10]
13. The 49th statistical report on China's internet development. China Internet Network Information Center. 2022. URL: http://www.cnnic.net.cn/NMediaFile/old_attach/P020220721404263787858.pdf [accessed 2025-02-10]
14. Centre for Research on Successful Ageing. Staying connected-the importance of social integration on the well-being of older adults. Singapore Management University. 2020 Dec. URL: https://rosa.smu.edu.sg/sites/rosa.smu.edu.sg/files/Briefs/Dec2020/Final/Dec2020_RB_Final_18dec.pdf [accessed 2025-02-10]
15. Chen H, Levkoff SE, Kort H, McCollum QA, Ory MG. Editorial: technological innovations to address social isolation and loneliness in older adults. Front Public Health 2023 Feb 6;11:1139266 [FREE Full text] [doi: 10.3389/fpubh.2023.1139266] [Medline: 36815157]
16. Van As BAL, Imbimbo E, Franceschi A, Menesini E, Nocentini A. The longitudinal association between loneliness and depressive symptoms in the elderly: a systematic review. Int Psychogeriatr 2022 Jul;34(7):657-669. [doi: 10.1017/s1041610221000399]
17. Egan K. Digital technology, health and well-being and the covid-19 pandemic: it's time to call forward informal carers from the back of the queue. Semin Oncol Nurs 2020 Dec;36(6):151088 [FREE Full text] [doi: 10.1016/j.soncn.2020.151088] [Medline: 33229183]
18. Cuffaro L, Di Lorenzo F, Bonavita S, Tedeschi G, Leocani L, Lavorgna L. Dementia care and COVID-19 pandemic: a necessary digital revolution. Neurol Sci 2020 Aug 17;41(8):1977-1979 [FREE Full text] [doi: 10.1007/s10072-020-04512-4] [Medline: 32556746]
19. Kokorelias KM, Nelson ML, Tang T, Steele Gray C, Ellen M, Plett D, et al. Inclusion of older adults in digital health technologies to support hospital-to-home transitions: secondary analysis of a rapid review and equity-informed recommendations. JMIR Aging 2022 Apr 27;5(2):e35925 [FREE Full text] [doi: 10.2196/35925] [Medline: 35475971]
20. Lofton S. Seniors want access to digital health tools—here's how to deliver them. Becker's Hospital Review. 2022 Dec 1. URL: https://www.beckershospitalreview.com/patient-experience/seniors-want-access-to-digital-health-tools-here-s-how-to-deliver-them.html [accessed 2025-02-10]
21. The PT patient experience report. WebPt. URL: https://www.webpt.com/downloads/the-pt-patient-experience-report [accessed 2025-02-10]
22. Moessner M, Bauer S. E-Mental-Health und internetbasierte Psychotherapie. Psychotherapeut 2017 May 5;62(3):251-266. [doi: 10.1007/s00278-017-0198-4]
23. Choi NG, Wilson NL, Sirrianni L, Marinucci ML, Hegel MT. Acceptance of home-based telehealth problem-solving therapy for depressed, low-income homebound older adults: qualitative interviews with the participants and aging-service case managers. Gerontologist 2014 Aug 08;54(4):704-713 [FREE Full text] [doi: 10.1093/geront/gnt083] [Medline: 23929664]
24. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct 21;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
25. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of mental health disorders using internet- and mobile-based interventions: a narrative review and recommendations for future research. Front Psychiatry 2017 Aug 10;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
26. Seifert A, Reinwand DA, Schlomann A. Designing and using digital mental health interventions for older adults: being aware of digital inequality. Front Psychiatry 2019 Aug 9;10:568 [FREE Full text] [doi: 10.3389/fpsyt.2019.00568] [Medline: 31447716]
27. Leech T, Dorstyn D, Taylor A, Li W. Mental health apps for adolescents and young adults: a systematic review of randomised controlled trials. Child Youth Serv Rev 2021 Aug;127:106073. [doi: 10.1016/j.childyouth.2021.106073]
28. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
29. Domhardt M, Messner E, Eder A, Engler S, Sander LB, Baumeister H, et al. Mobile-based interventions for common mental disorders in youth: a systematic evaluation of pediatric health apps. Child Adolesc Psychiatry Ment Health 2021 Sep 13;15(1):49 [FREE Full text] [doi: 10.1186/s13034-021-00401-6] [Medline: 34517896]

XSL•FO
RenderX

30. Helsel BC, Williams JE, Lawson K, Liang J, Markowitz J. Telemedicine and mobile health technology are effective in the management of digestive diseases: a systematic review. Dig Dis Sci 2018 Jun 16;63(6):1392-1408. [doi: 10.1007/s10620-018-5054-z] [Medline: 29663265]
31. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
32. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. EClinicalMedicine 2023 Mar;57:101860 [FREE Full text] [doi: 10.1016/j.eclinm.2023.101860] [Medline: 36864977]
33. Jones SL. An Efficacy Trial of Therapist-Assisted Internet-Delivered Cognitive-Behaviour Therapy for Older Adults with Generalized Anxiety. Regina, Canada: Faculty of Graduate Studies and Research, University of Regina; 2014.
34. Xiang X, Kayser J, Sun Y, Himle J. Internet-based psychotherapy intervention for depression among older adults receiving home care: qualitative study of participants' experiences. JMIR Aging 2021 Nov 22;4(4):e27630 [FREE Full text] [doi: 10.2196/27630] [Medline: 34813491]
35. Guidance on applying WCAG 2.0 to non-web information and communications technologies. World Wide Web Consortium. 2013. URL: https://www.w3.org/TR/wcag2ict/ [accessed 2025-02-10]
36. The World Factbook. Central Intelligence Agency. 2024. URL: https://www.cia.gov/the-world-factbook/ [accessed 2025-02-10]
37. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015 May 06;13(1):106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
38. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Med 2019 Dec 19;17(1):231 [FREE Full text] [doi: 10.1186/s12916-019-1461-z] [Medline: 31852455]
39. Lum E, Jimenez G, Huang Z, Thai L, Semwal M, Boehm BO, et al. Decision support and alerts of apps for self-management of blood glucose for type 2 diabetes. JAMA 2019 Apr 16;321(15):1530-1532 [FREE Full text] [doi: 10.1001/jama.2019.1644] [Medline: 30990543]
40. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016 Apr 13;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
41. Gasteiger N, Dowding D, Norman G, McGarrigle L, Eost-Telling C, Jones D, et al. Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide. BMJ Open 2023 Jun 12;13(6):e073283 [FREE Full text] [doi: 10.1136/bmjopen-2023-073283] [Medline: 37308269]
42. 2020 中国移动应用市场生态洞察报告. iiMedia. 2020 Sep 22. URL: https://www.iimedia.cn/c400/74427.html [accessed 2025-02-10]
43. Ahmed A, Ali N, Aziz S, Abd-alrazaq AA, Hassan A, Khalifa M, et al. A review of mobile chatbot apps for anxiety and depression and their self-care features. Comput Methods Programs Biomed Update 2021;1:100012. [doi: 10.1016/j.cmpbup.2021.100012]
44. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
45. Lin X, Martinengo L, Jabir AI, Ho AHY, Car J, Atun R, et al. Scope, characteristics, behavior change techniques, and quality of conversational agents for mental health and well-being: systematic assessment of apps. J Med Internet Res 2023 Jul 18;25:e45984 [FREE Full text] [doi: 10.2196/45984] [Medline: 37463036]
46. Alqahtani F, Al Khalifah G, Oyebode O, Orji R. Apps for mental health: an evaluation of behavior change strategies and recommendations for future development. Front Artif Intell 2019 Dec 17;2:30 [FREE Full text] [doi: 10.3389/frai.2019.00030] [Medline: 33733119]
47. Huang Z, Lum E, Car J. Medication management apps for diabetes: systematic assessment of the transparency and reliability of health information dissemination. JMIR Mhealth Uhealth 2020 Feb 19;8(2):e15364 [FREE Full text] [doi: 10.2196/15364] [Medline: 32130163]
48. "interactivity". Definition of interactivity noun from the Oxford Advanced Learner's Dictionary. Oxford Learner's Dictionary. URL: https://www.oxfordlearnersdictionaries.com/definition/english/interactivity [accessed 2025-02-10]
49. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell 2021 Feb 01;3(2):125-133. [doi: 10.1038/s42256-020-00287-7]
50. Bourke M, Haddara A, Loh A, Carson V, Breau B, Tucker P. Adherence to the World Health Organization's physical activity recommendation in preschool-aged children: a systematic review and meta-analysis of accelerometer studies. Int J Behav Nutr Phys Act 2023 Apr 26;20(1):52 [FREE Full text] [doi: 10.1186/s12966-023-01450-0] [Medline: 37101226]

XSL•FO
RenderX

51. Nagtegaal R, de Boer N, van Berkel R, Derks B, Tummers L. Why do employers (fail to) hire people with disabilities? A systematic review of capabilities, opportunities and motivations. J Occup Rehabil 2023 Jun 23;33(2):329-340 [FREE Full text] [doi: 10.1007/s10926-022-10076-1] [Medline: 36689057]
52. Warren SL, Moustafa AA. Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: a systematic review. J Neuroimaging 2023 Jan 18;33(1):5-18 [FREE Full text] [doi: 10.1111/jon.13063] [Medline: 36257926]
53. Yan Y, Hou J, Li Q, Yu NX. Suicide before and during the COVID-19 pandemic: a systematic review with meta-analysis. Int J Environ Res Public Health 2023 Feb 14;20(4):3346 [FREE Full text] [doi: 10.3390/ijerph20043346] [Medline: 36834037]
54. Callaghan MW, Müller-Hansen F. Statistical stopping criteria for automated screening in systematic reviews. Syst Rev 2020 Nov 28;9(1):273 [FREE Full text] [doi: 10.1186/s13643-020-01521-4] [Medline: 33248464]
55. Yin R, Martinengo L, Smith HE, Subramaniam M, Griva K, Tudor Car L. The views and experiences of older adults regarding digital mental health interventions: a systematic review of qualitative studies. Lancet Healthy Longev 2024 Nov;5(11):100638 [FREE Full text] [doi: 10.1016/j.lanhl.2024.08.007] [Medline: 39491881]
56. Rajappan D, Yin R, Martinengo M, Tudor Car L. Essential features of digital interventions targeting mental health among older adults: protocol for a scoping review. OSF Registries. 2023 Feb 8. URL: https://osf.io/aj3k4 [accessed 2025-02-10]
57. Depression in adults: recognition and management. NICE guideline [NG222]. National Institute for Health and Care Excellence (NICE). 2022 Jun 29. URL: https://www.nice.org.uk/guidance/ng222 [accessed 2025-02-10]
58. Baba H, Kito S, Nukariya K, Takeshima M, Fujise N, Iga J, Committee for Treatment Guidelines of Mood Disorders, Japanese Society of Mood Disorders. Guidelines for diagnosis and treatment of depression in older adults: a report from the Japanese Society of mood disorders. Psychiatry Clin Neurosci 2022 Jun 06;76(6):222-234 [FREE Full text] [doi: 10.1111/pcn.13349] [Medline: 35274788]
59. Avasthi A, Grover S. Clinical practice guidelines for management of depression in elderly. Indian J Psychiatry 2018 Feb;60(Suppl 3):S341-S362 [FREE Full text] [doi: 10.4103/0019-5545.224474] [Medline: 29535469]
60. Malhi GS, Bassett D, Boyce P, Bryant R, Fitzgerald PB, Fritz K, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry 2015 Dec 07;49(12):1087-1206. [doi: 10.1177/0004867415617657] [Medline: 26643054]
61. Riadi I, Kervin L, Dhillon S, Teo K, Churchill R, Card KG, et al. Digital interventions for depression and anxiety in older adults: a systematic review of randomised controlled trials. Lancet Healthy Longevity 2022 Aug;3(8):e558-e571. [doi: 10.1016/s2666-7568(22)00121-0]
62. Yap Y, Tan S, Choon S. Elderly's intention to use technologies: a systematic literature review. Heliyon 2022 Jan;8(1):e08765 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e08765] [Medline: 35128090]
63. Liu N, Yin J, Tan S, Ngiam K, Teo H. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. J Am Med Inform Assoc 2021 Oct 12;28(11):2483-2501 [FREE Full text] [doi: 10.1093/jamia/ocab151] [Medline: 34472601]
64. Maier A, Riedel-Heller SG, Pabst A, Luppa M. Risk factors and protective factors of depression in older people 65+. A systematic review. PLoS One 2021 May 13;16(5):e0251326 [FREE Full text] [doi: 10.1371/journal.pone.0251326] [Medline: 33983995]
65. Beck JS. Cognitive Behavior Therapy: Basics and Beyond. New York, NY: Guilford Press; 2011.
66. Laidlaw K, McAlpine S. Cognitive behaviour therapy: how is it different with older people? J Rat-Emo Cognitive-Behav Ther 2008 Nov 12;26(4):250-262. [doi: 10.1007/s10942-008-0085-6]
67. Pachana NA, Laidlaw K. The Oxford Handbook of Clinical Geropsychology. Oxford, UK: Oxford University Press; 2015.
68. Hatch LR, Moody HR. Aging: concepts and controversies. Teach Sociol 1996 Jan;24(1):131. [doi: 10.2307/1318916]
69. Hepple J. Psychotherapies with older people: an overview. Adv Psychiatr Treat 2018 Jan 02;10(5):371-377. [doi: 10.1192/apt.10.5.371]
70. Martinengo L, Stona A, Griva K, Dazzan P, Pariante CM, von Wangenheim F, et al. Self-guided cognitive behavioral therapy apps for depression: systematic assessment of features, functionality, and congruence with evidence. J Med Internet Res 2021 Jul 30;23(7):e27619 [FREE Full text] [doi: 10.2196/27619] [Medline: 34328431]
71. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
72. Boyer C. Quality and safety of health mobile applications: are they an issue? In: Hübner UH, Mustata Wilson G, Morawski TS, Ball MJ, editors. Nursing Informatics. Cham: Springer International Publishing; 2022:411-424.
73. Mobile accessibility: how WCAG 2.0 and other W3C/WAI guidelines apply to mobile. World Wide Web Consortium. 2015 Feb 26. URL: https://www.w3.org/TR/mobile-accessibility-mapping/ [accessed 2025-02-10]
74. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. Am J Psychiatry 2000 Apr;157(4 Suppl):1-45. [Medline: 10767867]
75. Lau N, O'Daffer A, Yi-Frazier JP, Rosenberg AR. Popular evidence-based commercial mental health apps: analysis of engagement, functionality, aesthetics, and information quality. JMIR Mhealth Uhealth 2021 Jul 14;9(7):e29689 [FREE Full text] [doi: 10.2196/29689] [Medline: 34259639]

XSL•FO
RenderX

76. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR Mhealth Uhealth 2019 Nov 07;7(11):e13236 [FREE Full text] [doi: 10.2196/13236] [Medline: 31697245]
77. Wu X, Xu L, Li P, Tang T, Huang C. Multipurpose mobile apps for mental health in Chinese app stores: content analysis and quality evaluation. JMIR Mhealth Uhealth 2022 Jan 04;10(1):e34054 [FREE Full text] [doi: 10.2196/34054] [Medline: 34982717]
78. Andrews J, Brown L, Hawley M, Astell A. Older adults' perspectives on using digital technology to maintain good mental health: interactive group study. J Med Internet Res 2019 Feb 13;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]
79. Gould CE, Zapata AML, Bruce J, Bereknyei Merrell S, Wetherell JL, O'Hara R, et al. Development of a video-delivered relaxation treatment of late-life anxiety for veterans. Int Psychogeriatr 2017 Oct;29(10):1633-1645. [doi: 10.1017/s1041610217000928]
80. Pywell J, Vijaykumar S, Dodd A, Coventry L. Barriers to older adults' uptake of mobile-based mental health interventions. Digit Health 2020 Feb 11;6:2055207620905422 [FREE Full text] [doi: 10.1177/2055207620905422] [Medline: 32110429]
81. Eichenberg C, Schott M, Sawyer A, Aumayr G, Plößnig M. Feasibility and conceptualization of an e-Mental health treatment for depression in older adults: mixed-methods study. JMIR Aging 2018 Oct 23;1(2):e10973 [FREE Full text] [doi: 10.2196/10973] [Medline: 31518235]
82. Gould CE, Loup J, Kuhn E, Beaudreau SA, Ma F, Goldstein MK, et al. Technology use and preferences for mental health self-management interventions among older veterans. Int J Geriatr Psychiatry 2020 Mar 29;35(3):321-330. [doi: 10.1002/gps.5252] [Medline: 31854029]
83. Gould CE, Loup JR, Scales AN, Juang C, Carlson C, Ma F, et al. Development and refinement of educational materials to help older veterans use VA mental health mobile apps. Prof Psychol Res Pr 2020 Aug;51(4):414-423 [FREE Full text] [doi: 10.1037/pro0000354] [Medline: 34465939]
84. LaMonica HM, Davenport TA, Roberts AE, Hickie IB. Understanding technology preferences and requirements for health information technologies designed to improve and maintain the mental health and well-being of older adults: participatory design study. JMIR Aging 2021 Jan 06;4(1):e21461 [FREE Full text] [doi: 10.2196/21461] [Medline: 33404509]
85. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
86. Leo AJ, Schuelke MJ, Hunt DM, Miller JP, Areán PA, Cheng AL. Digital mental health intervention plus usual care compared with usual care only and usual care plus in-person psychological counseling for orthopedic patients with symptoms of depression or anxiety: cohort study. JMIR Form Res 2022 May 04;6(5):e36203 [FREE Full text] [doi: 10.2196/36203] [Medline: 35507387]
87. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital peer-support platform (7Cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
88. Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ. Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (Youper): longitudinal observational study. J Med Internet Res 2021 Jun 22;23(6):e26771 [FREE Full text] [doi: 10.2196/26771] [Medline: 34155984]
89. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
90. Hunter JF, Olah MS, Williams AL, Parks AC, Pressman SD. Effect of brief biofeedback via a smartphone app on stress recovery: randomized experimental study. JMIR Serious Games 2019 Nov 26;7(4):e15974 [FREE Full text] [doi: 10.2196/15974] [Medline: 31769761]
91. Parks AC, Williams AL, Tugade MM, Hokes KE, Honomichl RD, Zilca RD. Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: a randomized controlled trial. Int J Wellbeing 2018 Dec 08;8(2):22-67. [doi: 10.5502/ijw.v8i2.745]
92. De Kock JH, Latham HA, Cowden RG, Cullen B, Narzisi K, Jerdan S, et al. Brief digital interventions to support the psychological well-being of NHS staff during the COVID-19 pandemic: 3-arm pilot randomized controlled trial. JMIR Ment Health 2022 Apr 04;9(4):e34002 [FREE Full text] [doi: 10.2196/34002] [Medline: 35044927]
93. Lee P, Aikens J, Richardson C, Singer D, Kullgren J, Kirch M. Mobile health app use among older adults. University of Michigan National Poll on Healthy Aging. 2022 Feb. URL: https://www.healthyagingpoll.org/reports-more/report/mobile-health-app-use-among-older-adults [accessed 2025-02-10]
94. Chou Y, Lin C, Lee S, Chang Chien Y, Cheng L. Potential mobile health applications for improving the mental health of the elderly: a systematic review. Clin Interv Aging 2023 Sep;18:1523-1534. [doi: 10.2147/cia.s410396]
95. Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. JAMA 2023 Jul 25;330(4):313-314. [doi: 10.1001/jama.2023.9630] [Medline: 37410497]
96. Yin H, Wardenaar KJ, Wang Y, Wang N, Chen W, Zhang Y, et al. Mobile mental health apps in China: systematic app store search. J Med Internet Res 2020 Jul 27;22(7):e14915 [FREE Full text] [doi: 10.2196/14915] [Medline: 32716301]

XSL•FO
RenderX

97. You L, Chen X, Yang L, Zhao J, Pan Y, Zhang S. Ten years of national essential public health services programs implementation: challenges and recommendations. General Pract China 2022;1(1):18.
98. Chen R, Zhang W, Wu X. Mental health policy and implementation from 2009 to 2020 in China. SSM - Mental Health 2023 Dec;4:100244. [doi: 10.1016/j.ssmmh.2023.100244]
99. Czaja S, Boot W, Charness N, Rogers W. Designing for Older Adults: Principles and Creative Human Factors Approaches , Third Edition. Boca Raton, FL: CRC Press; 2019.
100. GBD 2019 Blindness and Vision Impairment Collaborators, Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health 2021 Feb;9(2):e144-e160 [FREE Full text] [doi: 10.1016/S2214-109X(20)30489-7] [Medline: 33275949]
101. The Lancet. Loneliness as a health issue. Lancet 2023 Jul;402(10396):79. [doi: 10.1016/s0140-6736(23)01411-3]
102. Cosco TD, Fortuna K, Wister A, Riadi I, Wagner K, Sixsmith A. COVID-19, social isolation, and mental health among older adults: a digital catch-22. J Med Internet Res 2021 May 06;23(5):e21864 [FREE Full text] [doi: 10.2196/21864] [Medline: 33891557]
103. Chen E, Wood D, Ysseldyk R. Online social networking and mental health among older adults: a scoping review. Can J Aging 2021 Mar 19;41(1):26-39. [doi: 10.1017/s0714980821000040]
104. Forsman AK, Nordmyr J. Psychosocial links between internet use and mental health in later life: a systematic review of quantitative and qualitative evidence. J Appl Gerontol 2017 Dec 05;36(12):1471-1518. [doi: 10.1177/0733464815595509] [Medline: 26245208]
105. Fortuna KL, Naslund JA, Aschbrenner KA, Lohman MC, Storm M, Batsis JA, et al. Text message exchanges between older adults with serious mental illness and older certified peer specialists in a smartphone-supported self-management intervention. Psychiatr Rehabil J 2019 Mar;42(1):57-63 [FREE Full text] [doi: 10.1037/prj0000305] [Medline: 30010355]
106. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar;387(10023):1123-1132. [doi: 10.1016/s0140-6736(15)00298-6]
107. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, et al. Augmenting mental health in primary care: a 1-year study of deploying smartphone apps in a multi-site primary care/behavioral health integration program. Front Psychiatry 2019 Feb 28;10:94 [FREE Full text] [doi: 10.3389/fpsyt.2019.00094] [Medline: 30873053]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
MARS: Mobile Application Rating Scale
mHONcode: mobile Health On the Net Code
RCT: randomized controlled trial
W3C: World Wide Web Consortium

Edited by T de Azevedo Cardoso; submitted 31.01.24; peer-reviewed by S Munusamy, L Warner; comments to author 21.11.24; revised version received 13.01.25; accepted 27.01.25; published 21.02.25.

Please cite as:
Yin R, Rajappan D, Martinengo L, Chan FHF, Smith H, Griva K, Subramaniam M, Tudor Car L
Depression Self-Care Apps' Characteristics and Applicability to Older Adults: Systematic Assessment
J Med Internet Res 2025;27:e56418
URL: https://www.jmir.org/2025/1/e56418
doi:10.2196/56418
PMID:

©Ruoyu Yin, Dakshayani Rajappan, Laura Martinengo, Frederick H F Chan, Helen Smith, Konstadina Griva, Mythily Subramaniam, Lorainne Tudor Car. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of Factors Influencing the Willingness of Chinese Older Adults to Use mHealth Devices: Nationwide Cross-Sectional Survey Study

Mengyao Yan[1*]; Wendi Sun[2*]; Cheng Tan[3]; Yibo Wu[4]; Yuanli Liu[1]

[1]Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]School of Humanities and Management, Zhejiang Chinese Medical University, Zhejiang, China

[3]School of Government, Peking University, Beijing, China

[4]School of Public Health, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yuanli Liu
Chinese Academy of Medical Sciences & Peking Union Medical College
No. 9, Dongdan Sanjie
Dongcheng District
Beijing, 100730
China
Phone: 86 13522592907
Email: liuyuanli_pumc@163.com

Abstract

Background: In addition to standard older adult care services, mobile medical devices have proved to be an effective tool for controlling the health of older adults. However, little is known about the variables driving the acceptance of these gadgets and the willingness of older adults in China to use them.

Objective: This study aims to explore the factors that affect the use of mobile health (mHealth) devices by older adults in China, focusing on individual, social, and family influences.

Methods: The Psychology and Behavior Investigation of Chinese Residents survey database provided the data for this study. The survey was conducted in 148 Chinese cities between June 20 and August 31, 2022. The parameters linked to older persons' desire to use mobile medical devices were determined by this study using a combination model of multiple stepwise linear regression and a classification and regression tree decision tree.

Results: In total, 4085 older adults took part in the poll. On a scale of 0 to 100, the average score for willingness to adopt mHealth devices was 63.70 (SD 25.11). The results of the multiple stepwise linear regression showed that having a postgraduate degree and higher (β=.040; *P*=.007), being unemployed (β=.037; *P*=.02), having a high social status (β=.085; *P*<.001), possessing high health literacy (β=.089; *P*<.001), demonstrating high self-efficacy (β=.043; *P*=.02), not living with children (β=.0340; *P*=.02), having a household per capita monthly income of >Y4000 (US $550) (β=.048; *P*=.002), experiencing high perceived social support (β=.096; *P*<.001), reporting a high quality of life (β=.149; *P*<.001), having higher levels of family communication (β=–.071; *P*<.001), having an identity bubble (β=.085; *P*<.001), not having chronic diseases (β=.049; *P*=.001), and experiencing mild depression (β=–.035; *P*=.02) were associated with older adults' willingness to use mHealth devices. The classification and regression tree decision tree model's findings demonstrated that the primary determinants of older adults' desire to use mHealth devices are quality of life, identity bubble, social status, health literacy, family health, and perceived social support.

Conclusions: This study uses the Andersen Healthcare Utilization Model to investigate the effects of demand variables, enabling resources, and predisposing traits on older persons' propensity to use mHealth devices. These results offer reference data for the marketing and use of mHealth devices for older individuals in the future. The ultimate goal of this strategy is to create a balanced and harmonious integration of technology and humanistic care.

(J Med Internet Res 2025;27:e66804) doi:10.2196/66804

XSL•FO
RenderX

KEYWORDS

older adults; mobile health devices; health management; medical services; mobile phone

Introduction

Background

In the context of the rapidly advancing global demographic shift toward aging and the exponential growth of mobile health (mHealth) technology, mHealth devices have emerged as pivotal instruments for augmenting the quality of life of older adults, fostering proactive health self-management, and mitigating the effects of aging. As the nation harbors the world's largest older adult population, China confronts the imperative of effectively harnessing mHealth technology to cater to the extensive needs of its aging populace. According to statistics from the National Bureau of Statistics of China from December 2023 [1], China was home to 296.97 million individuals aged ≥60 years, constituting 21.1% of the national population, with 216.76 million of these being aged ≥65 years, representing 15.4%. This pervasive aging trend is concurrent with a surge in the incidence of chronic diseases [2], as more than two-thirds of Chinese older adults aged ≥65 years contend with multiple comorbidities. However, in 2022, China recorded a physician-to-population ratio of 3.15 practicing (assistant) physicians per 1000 individuals, alongside a ratio of 3.71 registered nurses per 1000 population—figures that, while placing the country in the middle tier internationally [3], nonetheless present a formidable challenge to the effective management of chronic conditions and the provision of comprehensive public health services. These circumstances significantly strain the nation's capacity to manage chronic illnesses and deliver robust public health services.

The World Health Organization's Draft Global Strategy on digital health 2020 to 2025 underscores the strategic importance of digital health within global health care systems. The Chinese government actively promotes an "internet+medical service" development model, vigorously driving forward initiatives for intelligent health and older adult care services. The clinical practice of the 2023 hypertension guidelines update has been released, for the first time incorporating wearable devices into its recommendations, presenting them as a novel modality for blood pressure measurement, working in conjunction with traditional methods, such as home blood pressure monitoring, to fortify the health defense against hypertension. mHealth, as a novel paradigm facilitating the exchange and management of health care information via mobile devices, assumes a crucial role in health management, demonstrating immense potential in domains such as chronic disease management, fall prevention [4], activity monitoring [5], and individual health behavior tracking [6]. It offers personalized and precision health care services [7], enabling the creation of individualized health profiles through integration with hospital systems. In turn, it assists in alleviating issues related to the unequal distribution of medical resources and health care services [8]. An example is the HUAWEI Watch D, which, having secured class II medical device registration certification from national authorities and undergone international standard validations, garners endorsement.

Moreover, against the backdrop of increasing empty-nest syndrome among the older adult population and the integration of medical and older adult care services in China, mHealth devices can be synergistically combined with health care services to enhance the quality and efficiency of older adult care provision. Data from the 2022 National Report on the development of national undertakings for the older adults by the National Bureau of Statistics of China reveal that, as of the end of 2022, China boasted 387,000 older adult care institutions and facilities nationwide, housing 8.294 million beds dedicated to older adult care services. Even with a conservative estimate adhering to the national standard ratio of 1 caregiver per 4 older adults, there is a pressing need for at least 2.075 million caregivers. However, the current workforce comprises only approximately 300,000 caregivers, leaving a substantial shortfall [9]. mHealth devices can, to a certain extent, mitigate this shortage of personnel. For instance, Apple's iOS 16 and watchOS 9 now come bundled with a medication reminder app within their native "Health" software, assisting older adults with medication adherence. In Beijing's Xicheng district, authorities have installed smart caregiving mattresses in older adult homes, continuously monitoring vital signs, such as heart rate and respiratory patterns, as well as bed entry and exit times, thereby augmenting care provision without direct human intervention.

In the field of commercial layout, the market size of smart wearable devices is expanding and will still maintain rapid growth in the next few years. International Data Corporation's latest quarterly tracking report on China's wearable devices market shows that China's wearable devices market shipped 34.7 million units in the third quarter of 2023, a year-on-year increase of 7.5%, and the overall market continues to grow. International Data Corporation expects that in 2024, the number of adult smartwatches will increase by 11%, driven by diversified product offerings [10].

Despite the multitude of benefits associated with mHealth, several critical challenges persist in its practical implementation. Previous studies have examined the willingness of the general Chinese population to use mHealth devices [11]; however, given the differences in digital divide, health literacy, technology acceptance, social support network, and concept of life among the older adults, the application of mHealth devices in the older adults has not reached an ideal state. Particularly in China, studies have shown that the older adults and the groups that are chronically ill, who are the potential main users, are not very receptive to mHealth services [8]. Furthermore, concerns over data security and accuracy [12], affordability of devices and associated services, equitable access [13], insufficient age-appropriateness of the products, anxiety induced by data-driven insights, and marginalization of certain functionalities, such as fall detection, all contribute to a less-than-ideal adoption scenario. Therefore, the willingness of the older adult population to use mHealth devices remains a larger issue.

XSL•FO
RenderX

In summary, while mHealth devices are being developed for a wider range of functions and see significant growth on the supply side, demand-side research primarily focuses on factors influencing the willingness to use these devices. These factors include perceived usefulness, product innovation, privacy protection, self-efficacy, self-perceived aging, and health awareness among users [14-17], and there is a significant lack of a large-scale sample of the older adult population. Therefore, in this dissertation, a large-scale cross-sectional survey was conducted nationwide with the aim of revealing the current status of Chinese older adults' willingness to use mHealth devices and the key influencing factors behind it. This is of great significance for the global understanding of the characteristics of the acceptance of new technologies among the older adult population, enhancing the effectiveness of digital health interventions and promoting the development of the smart aging industry, as well as providing a valuable empirical basis for international health equity and policy making for an aging society.

Theories and Hypotheses

The Anderson Behavioral Model of Health Services Use is the model of choice in the field of health services to explain and predict health care behaviors as well as to analyze the factors influencing health services use behaviors of different households. The model identifies predisposing characteristics, enabling resources, and need as key factors affecting the use of health services [18]. This contributes to the systematic analysis of factors influencing individual health service use behavior and the evaluation of the accessibility of health service use [19]. From the viewpoint of the population to which the study applies, the Anderson model is applicable to the study of the health care service use behavior of the general population as well as the study of the health care service use behavior of special populations, such as older adults, women, low-income earners, children, and people with the HIV infection. In terms of the scope of application, the model is applicable to the study of the entire process of individual health service use behavior, including the study of factors influencing the choice of individual health care modality, medical costs, disease screening, drug use, and so on [20,21]. This study views mHealth device use as a form of health service delivery and therefore uses the Anderson Health Service Utilization Model to explore the impact of different factors on mHealth device use. The specific model is shown in Figure 1.

Figure 1. A model of factors influencing the willingness of older adults to use mobile health devices.

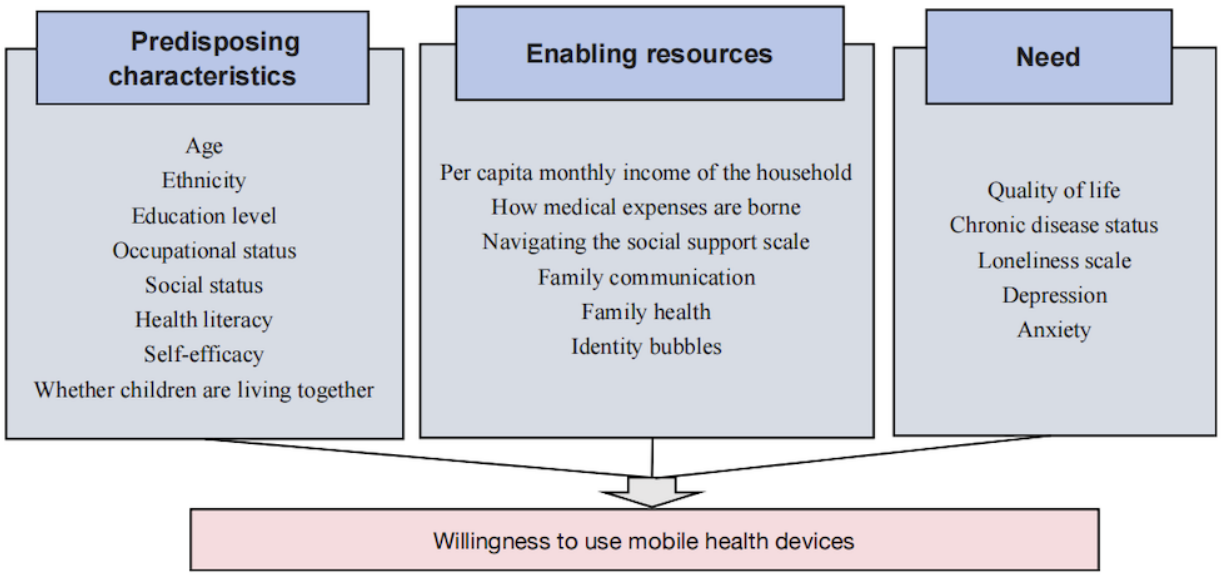

On the basis of this hypothesis, the following propositions are made, informed by theory and existing research:

- Hypothesis 1: older adults will be more likely to be willing to use mHealth devices than non–older adults.
- Hypothesis 2: older adults' willingness to use mobile devices is positive overall.

Methods

Participants

The data for this study came from the Psychology and Behavior Investigation of Chinese Residents database. This study was carried out across an extensive geographical expanse within China, encompassing 148 cities; 202 districts; 390 townships, towns, and streets; as well as 780 communities and villages, between June 20 and August 31, 2022. The sampling strategy used for this survey was multistage, meticulously informed by quota attributes derived from China's seventh national census data at the city level. These attributes included sex, age, and urban-rural distribution, ensuring a representative and stratified selection of participants. The detailed application of the quota method has been thoroughly documented in a preceding investigation by Wang et al [22]. This study was registered in the China Clinical Trial Registry (ChiCTR2200061046).

The study used the web-based Questionnaire Star platform for questionnaire distribution. Participants were eligible if they

XSL•FO
RenderX

were Chinese nationals aged ≥12 years, voluntarily participating, comprehending each questionnaire item, and completing the questionnaire independently. For those with limited mobility but intact cognitive function, investigators conducted one-on-one interviews, offering necessary assistance without influencing responses, thereby ensuring inclusivity and data integrity.

Ethical Considerations

This study was approved by the Ethics Research Committee of the Health Culture Research Center of Shaanxi (JKWH-2022-02). Informed consent was obtained from all participants. All data were collected anonymously and kept confidential. Participants did not receive any material or monetary rewards for participation. We ensure that all the images in the written materials published by this study do not contain the identity information of individual participants.

Self-Administered Questionnaires

The self-administered portion of the questionnaire investigated the demographic characteristics of the participants. The demographic characteristics included sex (male and female), age (60-74 years and ≥75 years), income (≤Y 4000 and ≥Y 4000 [US $550]), place of residence (eastern region, central region, and western region), education level (senior high school and below, junior college and bachelor degree, and postgraduate degree and above), occupational status (employed, retired, and unemployed: unemployed, jobless, and no regular occupation), and health care cost-bearing method (purely health insurance, purely commercial insurance, mixed insurance, and no health insurance), and chronic disease status (no chronic disease and chronic disease). The dependent variable measured was willingness to use mHealth devices (rated on a scale from a score of 0=not accepted to 100=very accepted).

Standard Scale

Health Literacy Questionnaire

In 2013, Sørensen et al [23] developed the 47-item European Health Literacy Survey Questionnaire, and in 2019, Duong et al [24] simplified the European Health Literacy Survey Questionnaire, creating a 12-item scale known as the Health Literacy Scale–short form (HLS-SF; HLS-SF12), designed to assess public health literacy in Asian countries. In 2023, Sun et al [25] sinicized HLS-SF12 using translation, reverse translation, and cultural debugging procedures. In 2023, Sun et al [26] applied the Mokken model in item response theory, classical test theory, to simplify HLS-SF12 (simplified to HLS-SF9). In this study, namely, the HLS-SF9 was used to measure the respondents' health literacy. The scale consists of 3 dimensions—health care, disease prevention, and health promotion—with 9 entries, each of which is rated on a 4-point scale (1=very difficult, 2=difficult, 3=easy, and 4=very easy), and the standardized health literacy index (ie, the total score of the scale) was calculated using a formula, which ranges from 0 to 50, with the higher the index representing the higher level of health literacy. The formula was calculated as index = (mean – 1) × (50/3), where mean is the average of the scores of all the entries for each individual, 1 is the smallest possible value of the mean (at which point the minimum value of the index is 0), 3 is the range of the mean, and 50 is the maximum value of the index. The higher the index, the higher the level of health literacy of the respondent. The Cronbach α coefficient of the scale was 0.923.

The New General Self-Efficacy Scale

The New General Self-Efficacy Scale (NGSES) was developed by Chen et al [27] on the basis of the General Self-Efficacy Scale, developed in 2001. The members of the project team of the "2022 Survey of Psychology and Behavior of the Chinese Population" applied the Mokken model of item response theory and classical test theory to simplify the NGSES-8, reducing the original 8 items to 3 items, and forming a new general self-efficacy scale. The NGSES-8 was simplified from the original 8 items to 3 items to form the NGSES (Wu et al [28]). In this study, respondents' self-efficacy was measured by the NGSES-3, which contains 3 entries on a 5-point Likert scale (1=strongly disagree and 5=strongly agree; the total scale score is the sum of the 3 entries, and the range of scale scores is from 3 to 15, with higher scores indicating higher levels of self-efficacy among respondents). It includes 3 entries on whether they can accomplish difficult tasks, whether they can successfully overcome many challenges, and whether they have the confidence to accomplish many different tasks effectively. The Cronbach α coefficient of the scale was 0.912.

The Perceived Social Support Scale

The Perceived Social Support Scale (PSSS) is used to assess participants' perceptions of social support (Li et al [29]). The PSSS consisted of 3 parsimonious items in this study, assessing perceived emotional support from friends, family, and significant others. It has been validated to be well correlated with the original items' scale, and the factor structure, reliability, and validity have also been well established (Wu et al [28]). Each item is scored on a 7-point scale of 1 to 7 (1="strongly disagree" and 7="strongly agree"). The summed scores on the PSSS range from 3 to 21 points, with higher scores representing greater perceived social support. The Cronbach α coefficient of the scale was 0.902.

The Family Communication Scale

The Family Communication Scale-10 (FCS-10) was used to evaluate the respondents' family communication. The scale was developed by Olson and Barnes and was sinicized by Guo N et al (Kwon et al [30] and Guo et al [31]). The FCS-10 consists of 10 entries on a 5-point Likert scale (1=strongly disagree and 5=strongly agree; the total scale score is the sum of the 10 entries, and the scale score ranges from 10 to 50, with higher scores indicating that the respondent's family communication is better). The purpose of the scale is to measure the quality of communication among family members with regard to the exchange of ideas, information, level of concern, openness, confidence, and emotions among family members. The Cronbach α coefficient of the scale was 0.957.

The Health Effect Values Visual Analog Score

The Health Effect Values Visual Analog Score (EQ-5D-VAS) is part of the European Five-Dimensional Health Scale. In this study, respondents rated the goodness of their health using the EQ-5D-VAS. The scale is a numerical range from 0 to 100, where 100 represents the respondent's perceived best possible

XSL•FO
RenderX

health condition and 0 represents the worst. Respondents are asked to indicate their self-assessed health status by selecting a whole number between 0 and 100.

The Identity Bubble Reinforcement Scale

The Identity Bubble Reinforcement Scale focuses on the social and information homogenization tendency brought about by identity-driven social media use from a psychosocial point of view. It is mainly used to measure the tendency of social media users to get involved in identity-driven online bubbles (bubble refers to homogenized information environments) and to emphasize the impact of identity-driven social activities on individual behavior within social media. The scale consists of 9 entries divided into 3 dimensions—social identity (3 entries), homogenization (3 entries), and information bias (3 entries)—and is based on a 10-point scale ranging from 1 (not at all like me) to 10 (not at all like me), with higher scores representing a more significant tendency to homogenize an individual's identity-driven social media activities (Kaakinen et al [32]). The Cronbach α coefficient of the scale was 0.948.

Data Exclusion

The specific data selection process is shown in Figure 2. The total sample size of the Psychology and Behavior Investigation of Chinese Residents database in 2022 was 21,916, of which 2 participants' ID abroad and unknown place of residence for the last 3 months, and 390 participants filled in information that was illogical. Thus, 392 participants were excluded, and 21,524 participants were retained. Among these 21,524 participants, 17,439 (81.02%) were aged <60 years and 4085 (18.97%) were aged ≥60 years.

Figure 2. Flowchart of sample selection of older adults' willingness to use mobile health devices. PBICR: Psychology and Behavior Investigation of Chinese Residents.

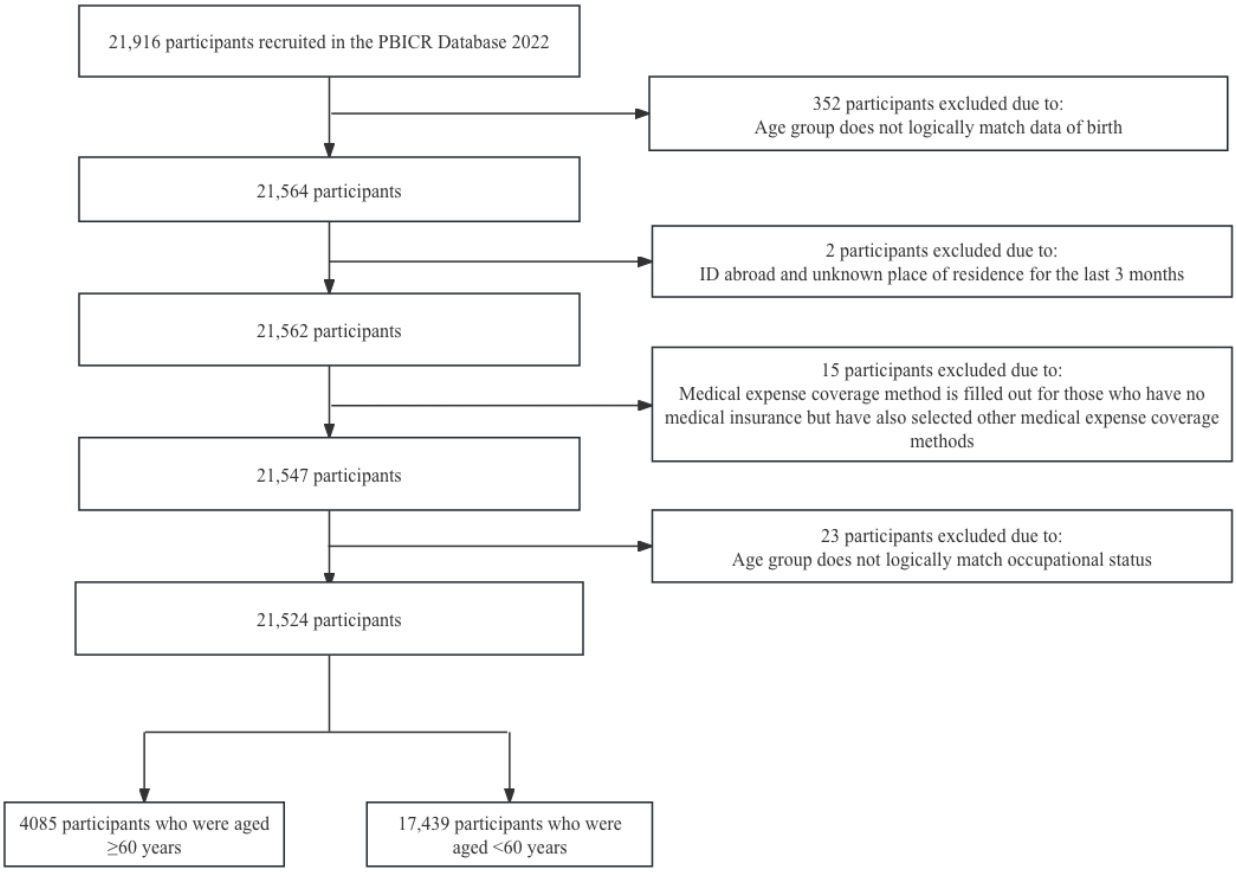

Statistical Analysis

Data analysis was conducted using SPSS (version 26.0; IBM Corp) and RStudio (RStudio, Inc), and all comparisons were 2-tailed, with *P* values <.05 considered statistically significant. The basic information of participants, along with their scores on the quality of life scale, family communication scale, health literacy scale, and willingness to use mHealth devices, were first analyzed descriptively. This was followed by univariate analysis of each potential explanatory variable: binary independent variables were analyzed using a 2-tailed *t* test, multicategorical independent variables with an *F* test, and continuous independent variables with correlation analysis. On the basis of these results, a model combining multiple stepwise linear regression and classification and regression tree (CART) decision tree models was developed to explore the factors influencing older adults' willingness to use mHealth devices.

Results

Characteristics of the Participants

A total of 21,916 samples were collected in this study, and 392 (1.78%) sets of abnormal data were excluded by logical

XSL•FO
RenderX

checking, while 21,524 (98.21%) sets of valid data were retained. The overall sample consisted of 18.98% (n=4085) older adult respondents (aged ≥60 years) and 81.02% (n=17,439) non–older adult respondents (aged <60 years). The mean score of all respondents' willingness to use mHealth devices was 68.14 (SD 25.30), with non–older adult respondents (69.18, SD 25.23) having a higher mean score of willingness to use mHealth devices than older adult respondents (63.70, SD 25.11).

There were 49.91% (n=2039) male participants and 90.8% (n=3709) Han Chinese in the sample of 4085 older adults, with a preponderance of younger older adults aged <75 years and >60 years (n=3383, 82.82%); most respondents' per capita monthly income of the household were ≤Y 4000 (US $550; n=2560, 62.67%); most respondents had a senior high school and below education (n=3523, 86.24%); 55.25% (n=2257) of the respondents are retired; 41.42% (n=1692) of the respondents are unemployed; and only 3.33% (n=136) of the respondents are still employed. Some (n=1471, 36.01%) respondents have lived in the eastern region for the last 3 months, 47.69% (n=1948) of the respondents have lived in the western region for the last 3 months, and 16.3% (n=666) of the respondents have lived in the central region for the last 3 months.

Each Scale Score of Participants

In total, 4085 respondents had a mean score of 69.86 (SD 18.55) for quality of life (EuroQol Visual Analog Scale), 29.74 (SD 9.85) for health literacy (HLS-SF9 scale), 35.40 (SD 12.05) for identity bubble (Identity Bubble Reinforcement Scale), 14.79 (SD 3.63) for perceived social support (PSSS), 10.51 (SD 2.42) for self-efficacy (NGSES), 4.44 (SD 1.54) for loneliness (T-ILS scale), and 36.92 (SD 7.63) for family communication (FCS-10 scale).

Willingness to Use mHealth Devices Among Chinese Older Adults

The mean score of the older adult population's willingness to use mHealth devices in this study was 63.70 (SD 25.11; median 65.00). To explore the distribution of older adult populations' willingness to use mHealth devices, the continuous variable of willingness to use mHealth devices was divided into 10 groups (0-10, 11-20, ... 91-100), and the number of respondents in each willingness to use subgroup was counted and presented in the form of a frequency distribution histogram. The specific results are shown in Multimedia Appendix 1.

As shown in Multimedia Appendix 1, older adults' willingness to use mHealth devices is roughly based on a clear dividing line of 40 points, with 17.36% (n=709) of the 4085 older adults having a willingness to use mHealth devices with scores <40 and 82.64% (n=3376) of older adults having a willingness to use mHealth devices with scores >40, with the largest number of older adults (n=664, 16.25%) having a willingness to use mHealth devices located in the range of 91 to 100. Subsequently, we observed 3 specific attitudinal older adult populations with respect to their willingness to use mHealth devices, namely, the strong nonwillingness to use mHealth devices population (scored 0) and the strong willingness to use mHealth devices population (scored 100). Some (n=51, 1.25%) older adults had no willingness to use mHealth devices at all, and 11.43% (n=467) of the older adults had a strong willingness to use mHealth devices.

Univariate Analysis of the Willingness of Older Adult Population to Use mHealth Devices

Because the dependent variable in this study is a continuous variable, a *t* test was used for binary independent variables, an *F* test was used for multicategorical independent variables, and correlation analysis was used for continuous independent variables. Most of the study variables were associated with willingness to use mHealth devices ($P<.001$; Table 1).

XSL•FO
RenderX

Table 1. Descriptive statistics of the willingness to use mobile medical devices among the older adults (N=4085).

Variables	Older adults, n (%)	Mobile health devices willingness, mean (SD)	*t* test (*df*)	*F* test (*df*)	Spearman rank correlation coefficient	*P* value
Predisposing characteristics						
Sex			–1.02 (4083)	—[a]	—	.31
Male	2039 (49.91)	64.10 (25.32)				
Female	2046 (50.09)	63.29 (24.90)				
Ethnicity			–1.90 (4083)	—	—	.06
Han	3709 (90.80)	63.46 (25.21)				
Minority	376 (9.20)	66.03 (24.02)				
Age group (years)			2.67 (964.33)	—	—	.008
60-74	3383 (82.82)	64.20 (24.90)				
≥75	702 (17.18)	61.27 (26.79)				
Education level			—	24.93 (2,4082)	—	<.001
Senior high school and below	3523 (86.24)	62.72 (25.26)				
Junior college and bachelor's degree	504 (12.34)	68.64 (23.06)				
Postgraduate degree and above	58 (1.42)	79.98 (23.05)				
Living region			—	17.12 (2,4082)	—	<.001
Eastern region	1471 (36.01)	66.55 (23.08)				
Central region	666 (16.30)	60.36 (27.52)				
Western region	1948 (47.69)	62.68 (25.51)				
Occupational status			—	14.67 (2,4082)	—	<.001
Employed	136 (3.33)	64.55 (28.11)				
Retired	2257 (55.25)	65.5 (324.06)				
Unemployed	1692 (41.42)	61.18 (26.01)				
Social status			—	45.59 (2,4082)	—	<.001
Low	180 (4.41)	57.05 (31.20)				
Medium	3099 (75.86)	62.22 (24.58)				
High	806 (19.73)	70.87 (24.25)				
Health literacy			–11.36 (3902.26)	—	—	<.001
Low	2315 (56.67)	59.88 (25.30)				
High	1770 (43.33)	68.69 (23.97)				
Self-efficacy			–9.97 (3757.46)	—	—	<.001
Low	2269 (55.54)	60.21 (23.99)				
High	1816 (44.46)	68.06 (25.79)				
Whether children living together			—	2.14 (2,4082)	—	.12
Yes	1786 (43.72)	64.55 (24.69)				
No	1525 (37.33)	62.74 (24.95)				
Not filled	774 (18.95)	63.60 (26.33)				
Enabling resources						
Per capita monthly income of the household (US $)			–6.29 (3575.12)	—	—	<.001
≤Y 4000 (US $550)	2560 (62.67)	61.87 (26.29)				
>Y 4000 (US $550)	1525 (37.33)	66.77 (22.68)				

XSL•FO
RenderX

Variables	Older adults, n (%)	Mobile health devices willingness, mean (SD)	*t* test (*df*)	*F* test (*df*)	Spearman rank correlation coefficient	*P* value
How medical expenses are borne			—	5.54 (3,4081)	—	<.001
Only medical health insurance	3175 (77.72)	62.87 (25.15)				
Only commercial health insurance	91 (2.23)	66.63 (19.47)				
Hybrid insurance	587 (14.37)	66.03 (24.60)				
None	232 (5.68)	68.00 (27.01)				
Perceived social support			—	63.09 (2,4082)	—	<.001
Low	338 (8.27)	58.79 (21.53)				
Middle	2025 (49.57)	60.20 (24.30)				
High	1722 (42.15)	68.77 (25.81)				
Family health			—	12.45 (2,4082)	—	<.001
Poor	1237 (30.28)	61.18 (22.81)				
Moderate	1471 (36.01)	63.59 (26.07)				
Excellent	1377 (33.71)	66.07 (25.82)				
Quality of life			–14.49 (4083)	—	—	<.001
Low	2864 (70.11)	60.07 (24.14)				
High	1221 (29.89)	72.20 (25.31)				
Family communication, mean (SD)	—	—	—	—	0.142	<.001
Identity bubbles, mean (SD)	—	—	—	—	0.179	<.001
Need						
Chronic diseases conditions			—	22.84 (2,4082)	—	<.001
Having chronic diseases related to mobile health devices	1682 (41.18)	62.09 (25.66)				
Having chronic diseases unrelated to mobile health devices	672 (16.45)	60.07 (24.37)				
Not having chronic diseases	1731 (42.37)	66.67 (24.54)				
Loneliness			3.08 (4083)	—	—	.002
Not alone	2832 (69.33)	64.50 (24.76)				
Alone	1253 (30.67)	61.88 (25.80)				
Depression			—	17.06 (4,4080)	—	<.001
No	1894 (46.36)	66.92 (25.31)				
Mild	1363 (33.37)	60.32 (24.53)				
Moderate	559 (13.68)	61.46 (24.17)				
Moderate to severe	218 (5.34)	60.91 (26.03)				
Severe	51 (1.25)	70.90 (23.81)				
Anxiety			—	10.18 (3,4081)	—	<.001
No	2282 (55.86)	65.48 (25.81)				
Mild	1324 (32.41)	60.96 (23.77)				
Moderate	414 (10.13)	62.07 (23.88)				
Severe	65 (1.59)	67.31 (28.99)				

[a]Not available.

XSL•FO
RenderX

What Are the Factors Influencing the Willingness to Use mHealth Devices Among Chinese Older Adults?

Factors Influencing the Willingness to Use mHealth Devices Among Chinese Older Adults Based on Multiple Stepwise Linear Regression

This study used multiple stepwise linear regression to explore the factors affecting the willingness of older adults to use mHealth devices, and the specific results of multiple stepwise linear regression are shown in Table 2. The 13 main factors influencing older adults' willingness to use mHealth devices were education level, occupational status, social status, health literacy, self-efficacy, whether children were living together, per capita monthly income of the household, perceived social support, quality of life, family communication, identity bubbles, chronic disease conditions, and depression. In terms of predisposing characteristics, older adults who lived with their children (P=.02; β=.034), had high health literacy (P<.001; β=.089), had high self-efficacy (P=.02; β=.043), had high social status (P<.001; β=.085), were unemployed (P=.02; β=.037), and had the highest education level of postgraduate degree and above (P=.007; β=.040) were willing to use mHealth devices. In terms of enabling resources, older adults with a household per capita monthly income of ≥Y 4000 (P=.002; β=.048), a high quality of life (P<.001; β=.149), high perceived social support (P<.001; β=.096), and a lower family communication level (P<.001; β=–.071) were more willing to use mHealth devices. In terms of need, older adults who did not have a chronic disease (P=.001; β=.049) and had mild depression (P=.007; β=.040) were less willing to use mHealth devices.

Table 2. Multiple stepwise linear regression results of the older adults' willingness to use mobile health devices.

Variables	Unstandardized coefficients, B (SE)	Standardized coefficients, β (95% CI)	*t* test	*P* value	Variance inflation factor
Predisposing characteristics					
Education level (reference: senior high school and below)					
Postgraduate degree and above	8.480 (3.165)	0.040 (2.276 to 14.684)	2.680	.007	1.018
Occupational status (reference: employed)					
Unemployed	1.846 (0.784)	0.037 (0.308 to 3.384)	2.353	.02	1.105
Social status (reference: low)					
High	5.370 (0.954)	0.085 (3.500 to 7.241)	5.628	<.001	1.047
Health literacy (reference: low)					
High	4.533 (0.879)	0.089 (2.810 to 6.255)	5.160	<.001	1.376
Self-efficacy (reference: low)					
High	2.155 (0.932)	0.043 (0.327 to 3.984)	2.312	.02	1.559
Whether children living together (reference: no)					
Yes	1.741 (0.752)	0.034 (0.266 to 3.216)	2.314	.02	1.012
Enabling resources					
Per capita monthly income of the household (reference: ≤Y 4000 (US $550)					
>Y 4000 (US $550)	2.491 (0.819)	0.048 (0.885 to 4.097)	3.040	.002	1.140
Perceived social support (reference: low)					
High	4.882 (0.946)	0.096 (3.028 to 6.736)	5.163	<.001	1.583
Quality of life (reference: low)					
High	8.197 (0.864)	0.149 (6.504 to 9.890)	9.491	<.001	1.135
Family communication	–0.234 (0.065)	–0.071 (–0.361 to –0.108)	–3.630	<.001	1.759
Identity bubbles	0.178 (0.033)	0.085 (0.113 to 0.243)	5.339	<.001	1.170
Need					
Chronic diseases conditions (reference: having chronic diseases related to mobile health devices)					
Not having chronic diseases	2.513 (0.770)	0.049 (1.004 to 4.022)	3.264	.001	1.051
Depression (reference: no)					
Mild	–1.860 (0.818)	–0.035 (–3.463 to –0.256)	–2.274	.02	1.080

XSL•FO
RenderX

Factors Influencing the Willingness to Use mHealth Devices Among Chinese Older Adults Based on CART Decision Tree Model

Multiple stepwise linear regression can explore the main influencing factors of the use of mHealth devices for the older adults, but it is not possible to intuitively understand the relative importance of the influencing factors and the interaction between the influencing factors; therefore, this study uses the CART decision tree model to visualize the influencing factors (with a training set:test set ratio of 70%:30%) to explore the interaction between the influencing factors. The specific results are shown in Figure 3. Figure 3 shows that quality of life is the most important factor influencing the willingness to use mHealth devices among older adults, and the willingness to use mHealth devices among older adults with high quality of life (73 points) is significantly higher than that of older adults with low quality of life (60 points). For the older adult group with low quality of life, identity bubbles is the main influencing factor on their willingness to use mHealth devices. When the identity bubbles score is ≥43, the mean value of the older adults' willingness to use mHealth devices is 68 points, and when the identity bubbles score is <43, the mean value of the older adults' willingness to use mHealth devices is 58 points. And then if the identity bubbles score is <27, the mean value of willingness to use mHealth devices for such older adults is 51 points, and if the identity bubbles score is ≥27, family health becomes the main factor influencing older adults' willingness to use mHealth devices, with the mean value of willingness to use mHealth devices for the older adults with moderate family health being 54 points and the mean value of willingness to use mHealth devices for the older adults with poor family health or excellent family health being 62 points. For older adults with high quality of life, social status is the main factor influencing older adults' willingness to use mHealth devices, with the mean value of mHealth devices for older adults with high social status being 81 points, while older adults with medium and low social status need to further judge their health literacy, with the mean value of mHealth devices for older adults with low health literacy being 63 points, and older adults with high health literacy needed to be further judged in terms of their sense of perceived social support, with older adults with high perceived social support (79 points) having a significantly higher willingness to use mHealth devices than older adults with low perceived social support (65 points). This shows that older adults with high quality of life and high social status have the highest willingness to use mHealth devices (81 points), while older adults with low quality of life and low status bubbles (<27 points) have the lowest willingness to use mHealth devices (51 points).

Figure 3. Classification and regression tree (CART) decision tree results of older adults' mobile health device use willingness.

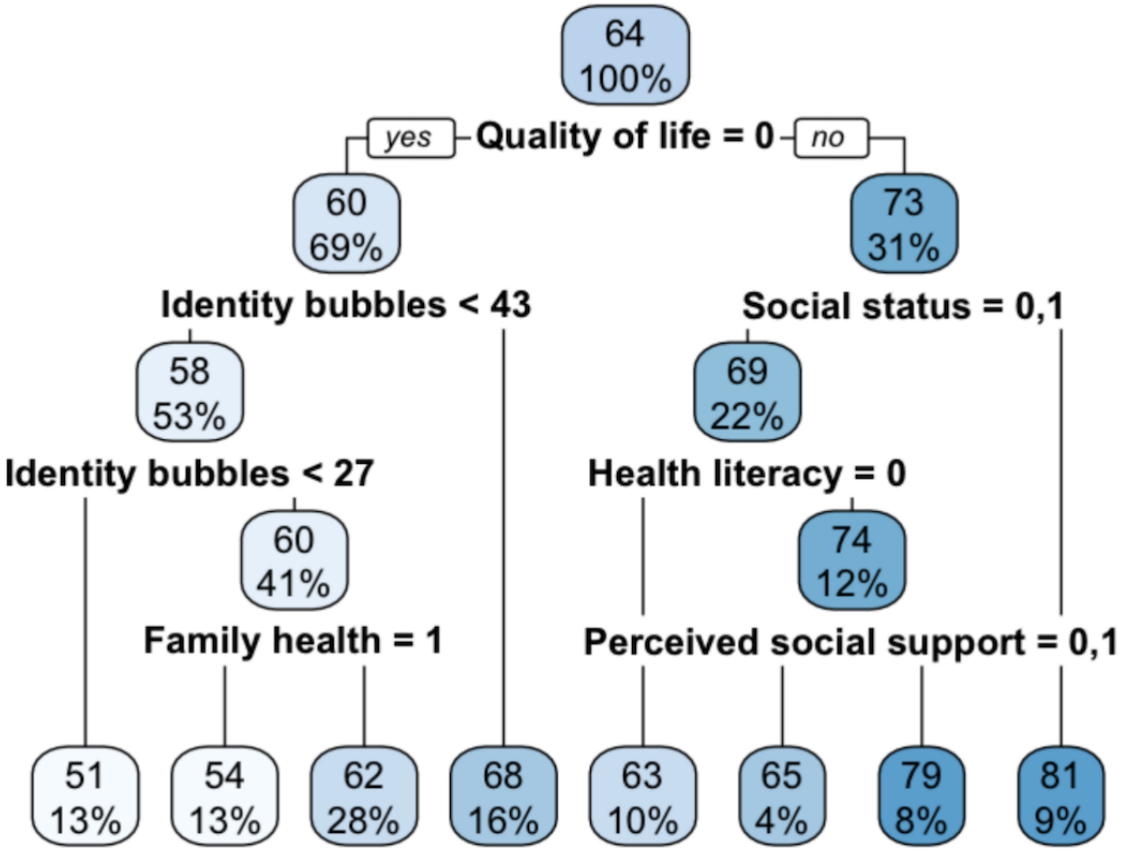

XSL•FO
RenderX

Discussion

Overview

This study found that Chinese older adults have positive attitudes toward using mHealth devices, with significant positive effects of education, quality of life, self-efficacy, and health literacy on the use of mHealth devices at the individual level; a positive effect of perceived family social status at the social level; and a negative effect of family communication on their willingness to use the devices. Social-level perceived social support and identity bubble had a significant positive effect on using mHealth devices.

Principal Findings

The study finds that the gap in willingness to use mHealth devices between older adults and non–older adults is relatively small. The average willingness score for older adults is 63.70, compared to 69.18 for non–older adult respondents. This differs from the conclusions of existing studies, which indicate that older adults have a significantly lower proportion of using electronic health devices to access information compared to non–older adult groups [33,34]. A Canadian-based study also showed significant differences in the use of smart devices between older and younger adults, with nearly half of them using smartphones but not downloading health-related apps [35]. This may be partly because older adults have greater health needs [36]. Meanwhile, it may also be attributed to the fact that today's digital technologies have increasingly prioritized age-friendly design, making older adult–friendly devices more accessible and ensuring that older adults are no longer left on the periphery of digitalization [37]. Studies have shown that, most of the time, older and younger adults share similar preferences for using electronic health devices. When the devices are impractical or lack inclusive design, people across different groups tend to reduce their use [38].

In terms of research methodology, this study uses a combined model of logistic regression and decision tree analysis to comprehensively and rigorously examine the factors influencing older adults' willingness to use mHealth devices. Both methods identify key determinants, such as quality of life, identity bubble, perceived social support, self-efficacy, family communication, per capita monthly household income, and health literacy. The logistic regression model further highlights additional demographic factors, including region of residence, education level, and chronic disease status. In contrast, the decision tree model emphasizes the stronger impact of subjective psychological factors.

Factors Influencing Older Adults' Willingness to Use Health Devices

This study focuses extensively on the adaptability of the older adult population in the context of digital health and smart health care. Using older adults' use of mHealth devices as a starting point, it explores their willingness to use such devices and the influencing factors. Using a comprehensive systematic framework of predisposing characteristics, enabling resources, and needs, this study conducts a thorough analysis of variables and fits the data using binary logistic regression. The findings indicate that educational level, perceived family social status, perceived social support, self-efficacy, quality of life, health literacy, identity bubble, and intrafamily communication significantly influence the older adults' willingness to use mHealth devices. This not only validates the Behavioral Model of Health Services Use in explaining the use of health care services among the older adults but also demonstrates its applicability in the digital age.

In terms of predisposing factors, this study reveals that perceived family social status, educational level, and health literacy are positively correlated with older adults' willingness to use mHealth devices. This conclusion complements existing research, which often focuses on the relationship between objective social class and status and the use of digital devices. For example, some studies suggest that individuals with higher social class are better equipped to use digital devices in the field of education [39]. However, there has been relatively little research on the relationship between subjective class perception and the use of digital devices. This study creatively establishes a link between the 2 and focuses specifically on the health domain. In terms of health literacy, research shows that the application of digital health technologies is highly correlated with users' health literacy [40]. In terms of education and health literacy, this study aligns with the conclusions of existing research. Research showed that digital health education can enhance people's understanding of diseases and is more effective than traditional education methods in improving overall health [41]. Therefore, older adults who have received digital health education and possess higher health literacy are more motivated to use digital health devices.

This finding can be elucidated through the lens of the digital divide. Concerning perceived family social status, on one hand, the formation of the digital divide often relates to the objective social strata within families, encompassing dimensions such as income [42] and sociocultural factors. Individuals with higher incomes are more likely to have greater access to the internet and IT [43]. Moreover, formal and informal rules regarding inequality within social cultures also influence the digital divide experienced by different groups [44]. On the other hand, as a subjective variable, perceived social class is influenced by objective social strata and often correlates more closely with individuals' life and psychological states [45]. Therefore, individuals who perceive themselves as belonging to a higher social class often possess the capability to access digital devices, use digital technologies effectively and reasonably, and fulfill their personal needs through these technologies. They also exhibit higher levels of acceptance and inclusiveness toward digital technologies, resulting in a greater willingness to use mHealth devices. It is important to note that existing research typically describes the phenomenon of the digital divide among the older adults from the perspective of objective social strata or social capital, without extensively exploring the impact of subjective social class perceptions on older adults' preferences for using mHealth devices. Thus, this study contributes by supplementing existing research from a subjective psychological perspective.

Research on the digital divide and digital engagement among older adults often presents 2 perspectives concerning the

XSL•FO
RenderX

variables of education level and health literacy. One perspective considers the integration or participation of older adults in digital life because of the 3-level digital divide. It suggests that individuals with higher education levels and greater technological understanding are generally more active in using digital devices. Accordingly, older adults benefit from mastering digital technologies and actively engaging in digital life [46]. The second perspective views "digital disengagement" among older adults as a fourth type of digital divide [47]. It contends that even as older adults enhance their digital literacy and acquire basic knowledge and skills in IT while integrating into the digital society, they may become unwilling to use digital technologies due to discriminatory content and other factors present in digital applications aimed at older demographics [48]. The results of this study support the first perspective, indicating that overall, individuals on the periphery of the digital society within the older adult population exhibit relatively lower willingness to use digital devices. This finding aligns with some existing research in this field. A survey of the Polish population reveals that the educational level of older adults and their motivation to use digital devices are significant variables influencing their level of digital participation [49]. There is also evidence suggesting that through professional training or digital literacy education, older adults' digital skills and literacy can significantly improve [50]. These digital skills form the foundation for older adults to use digital devices [22]. Therefore, reducing the digital divide and increasing familiarity with digital devices among older adults may be a way to enhance their willingness to use such devices.

Furthermore, this study found that self-efficacy significantly influences older adults' use of mHealth devices, consistent with previous research findings. The research indicates a correlation between willingness to use digital technology and self-efficacy among retired older adult populations [51]. On one hand, learning to use digital technology can enhance self-efficacy among older adults [52]; on the other hand, lower self-efficacy can diminish their ability to use smart devices [53]. These findings provide a basis for the conclusions drawn in this study.

In terms of facilitating resources, first, this study found a significant positive correlation between perceived social support, identity bubbles, and the willingness of older adults to use mHealth devices. This aligns with previous research findings. Scholars argue that perceived social support is among the crucial factors influencing tablet device use among older adults [54]. This can be explained using social identity theory. Research indicates that there are often stereotypes about older adults within society, and younger individuals who have more contact with older adults tend to possess greater knowledge about aging-related issues. They also show more understanding and fewer biases toward older adults [55]. Due to their greater proficiency and flexibility in using mHealth devices, younger individuals often serve as exemplary models for older adults when they experience higher levels of social acceptance and reduced stereotyping. Consequently, older adults are more likely to learn from younger generations in using mHealth devices and receive greater encouragement in digital device use. Therefore, increased social support and a sense of belonging on social media platforms among older adults enhance their willingness to engage with mHealth devices. Furthermore, the findings of this study validate hypotheses derived from social capital theory, which posits that objective resources and subjective identities and trust within social relationships positively influence internet use [56]. Therefore, older adults' identification with internet communities and their level of digital identity bubbles also positively influence their willingness to use digital devices.

Second, this study found that family communication significantly negatively impacts the willingness of older adults to use mHealth devices, indicating that greater family communication correlates with a reduced inclination among older adults to use such devices. This finding contrasts with existing research conclusions. Existing studies suggest that intergenerational family networks are crucial in enhancing older adults' digital literacy and can increase their willingness to use relevant digital devices [57]. This could be due to several factors. On one hand, there are distinctions between health devices and typical digital social devices. On the other hand, older adults experience significant social isolation, particularly those living alone and lacking care from their children, thereby increasing their willingness to use digital devices [58]. The application of digital devices by older adults often serves as a substitute or supplement to real-world social and health needs. When these needs are adequately met in reality, there is a decrease in both the willingness and use rates of relevant digital devices among older adults [59], which could serve as an explanatory mechanism for the conclusions drawn in this study.

In terms of demand, this study found that quality of life has a significantly positive impact on older adults' willingness to use mHealth devices. Existing research has laid the groundwork for this conclusion. Studies, such as those involving telephone interviews with older adults in Hungary during COVID-19, have identified substantial demand for digital health solutions among older adult populations, highlighting its importance as a critical application area [37]. This suggests that older adults' subjective perception of health and their objective life circumstances both influence their demand for and willingness to use digital health devices. This supports the conclusions drawn in this study.

It is noteworthy that explanations of these 3 dimensions are not entirely distinct; in fact, these variables collectively exert a systematic and comprehensive influence on older adults' willingness to use mHealth devices.

Future Prospects

This study has the following strengths: regarding the research topic, it explores the psychological foundation of the digital divide among older adults from the perspective of subjective willingness, enriching the structural discussion as corroborated by existing research. In terms of data, the study uses scientifically sampled large-scale samples of older adults, encompassing rich variables and comprehensive discussions, thereby supplementing existing research with empirical analysis from China. Theoretically, the study systematically categorizes variables and discusses the theoretical relationship between subjective class perception and the digital divide, while

XSL•FO
RenderX

validating social identity and social capital theories, contributing to theoretical significance.

However, the limitations of this study primarily stem from its cross-sectional design, which offers only a preliminary exploration of influencing factors without addressing potential mechanisms or causal effects. In addition, the lack of comprehensive external databases for triangulating results in the decision tree analysis is another limitation. Furthermore, the study lacks subgroup explanations. In reality, older adults with different characteristics may not score similarly on various influencing factors, and their motivations for using mHealth devices can vary in terms of initiative, passivity, and purpose. In addition, the study relies on a single-item measure for the dependent variable—older adults' willingness to use mHealth devices. While a single-item measure offers simplicity and ease of administration, it may lack the depth and reliability of multiitem scales, which could better capture the complexity and nuances of willingness and actual behaviors. This limitation might lead to measurement bias and restrict the robustness of the findings.

To further delve into this topic in the future, we propose the following questions for in-depth discussion: Do older adults use medical devices differently than the general population? first, this study found that there was no significant difference in the willingness of older adults to use health devices compared to the general population, which contrasts with previous researches. Existing studies indicate that younger age groups prefer generative artificial intelligence that enhances productivity compared to generative artificial intelligence primarily used in health settings. In addition, research suggests that the internet has a greater impact on younger individuals than on older adults. However, individuals aged ≥60 years tend to approach the use of IT devices with caution, shyness, and deliberation [60]. It suggests a need for further exploration into the underlying reasons for this distinction.

Second, do the influencing factors on older adults' willingness to use health devices similarly apply to different age groups? Research indicates that for children and adolescents, family factors and parent-child relationships often have a significant positive impact on their use of digital devices [61], yet there remains a lack of research on their use of health devices. Therefore, this could serve as a direction for further exploration in the future.

Conclusions

The results of this study showed that Chinese older adults' willingness to use mHealth devices was positive overall. The positive state of family communication, on the contrary, reduced their willingness to use. This suggests that particular attention should be paid to the singleton and widowed older adult groups, where the lack of family communication may translate into higher acceptance and dependence on mHealth devices. Therefore, policy development should focus on providing free or subsidized health devices for these groups, and in addition, when designing promotional strategies, attention should be paid to balancing the relationship between family communication and technology acceptance, and at the same time, strengthening family digital health education to promote the incorporation of technology in ways that can enhance rather than weaken family health management and emotional connection, help fill the gaps in older adult care services and improve health management, and realize the harmonious coexistence of technology and humanistic care.

Acknowledgments

This work was supported by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences—A Strategic Study on Healthy China Development and Health System Reform (2021-I2M-1-046). The authors are very grateful to the staff and participants of the Psychology and Behavior Investigation of Chinese Residents. In addition, the authors would like to extend their appreciation to the anonymous reviewers for their constructive comments, which significantly contributed to the improvement of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequency distribution histogram results of the older adults' willingness to use mobile health devices.
[PNG File , 95 KB - jmir_v27i1e66804_app1.png]

References

1. The total population has declined and high-quality population development has achieved results. National Bureau of Statistics. 2024 Jan 18. URL: https://www.stats.gov.cn/xxgk/jd/sjjd2020/202401/t20240118_1946711.html [accessed 2024-06-11]
2. Peng W, Chen S, Chen X, Ma Y, Wang T, Sun X, et al. Trends in major non-communicable diseases and related risk factors in China 2002-2019: an analysis of nationally representative survey data. Lancet Reg Health West Pac 2024 Mar;43:100809 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100809] [Medline: 38456095]

XSL•FO
RenderX

3. Global burden of disease study 2016 (GBD 2016) incidence, prevalence, and years lived with disability 1990-2016. Institute for Health Metrics and Evaluation. 2016. URL: https://ghdx.healthdata.org/record/ihme-data/gbd-2016-incidence-prevalence-and-ylds-1990-2016 [accessed 2024-06-11]
4. Bay Alarm Medical homepage. Bay Alarm Medical. URL: https://www.bayalarmmedical.com/(accessed [accessed 2024-05-15]
5. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jul 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
6. Ma C, Xu H, Li D, Zhang Z. [Research progress on wearable physiological parameter monitoring and its clinical applications]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2021 Jul 25;38(3):583-593 [FREE Full text] [doi: 10.7507/1001-5515.202009031] [Medline: 34180205]
7. Sun L, Buijsen M. Mobile health in China: does it meet availability, accessibility, acceptability and quality standards? Health Policy Technol 2022 Sep;11(3):100660 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100660]
8. Zhang X, Lai KH, Guo X. Promoting China's mHealth market: a policy perspective. Health Policy Technol 2017 Dec;6(4):383-388. [doi: 10.1016/j.hlpt.2017.11.002]
9. 2022 Annual report on the development of national undertakings for the aged. Ministry of Civil Affairs. 2023. URL: https://www.gov.cn/lianbo/bumen/202312/content_6920261.htm [accessed 2024-05-15]
10. IDC: China's wearable market grew 7.5% year-on-year in the third quarter, and the global market grew 2.6% year-on-year, setting a record for the highest shipments in the quarter in two years. IDC Corporate. URL: https://www.idc.com/getdoc.jsp?containerId=prCHC51569823 [accessed 2024-05-15]
11. Wang X, Wu Y, Meng Z, Li J, Xu L, Sun X, et al. Willingness to use mobile health devices in the post-COVID-19 era: nationwide cross-sectional study in China. J Med Internet Res 2023 Mar 17;25:e44225 [FREE Full text] [doi: 10.2196/44225] [Medline: 36719823]
12. Guo Z, Luo Y, Cai Z, Zheng T. Overview of privacy protection for medical and health big data. J Comput Sci Technol 2021;15(3):389-402 [FREE Full text] [doi: 10.3778/j.issn.1673-9418.2009071]
13. Sim I. Mobile devices and health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
14. Zhao YS, Wang ZJ. A study of users' continuance intention of wearable devices ——based on extended ECM-IS model. J Northeast Univ (Soc Sci) 2018;20(4):366-372 [FREE Full text]
15. Zhang M, Luo M, Nie R. Analysis on factors influencing consumer usage intention of healthcare wearable technology: based on the moderating effects of usage experience and health knowledge. J Inf Resour Manag 2017;7(2):14-21. [doi: 10.13365/j.jirm.2017.02.014]
16. Meng XR, Deng ZH, Wu TL. Study on influence factors of the willingness of using smart wearable devices in middle-aged and elderly. Soft Sci Health 2020;34(8):41-46 [FREE Full text] [doi: 10.3969/j.issn.1003-2800.2020.08.0019]
17. Wang Y, Lu L, Zhang R, Ma Y, Zhao S, Liang C. The willingness to continue using wearable devices among the elderly: SEM and FsQCA analysis. BMC Med Inform Decis Mak 2023 Oct 16;23(1):218 [FREE Full text] [doi: 10.1186/s12911-023-02336-8] [Medline: 37845659]
18. Andersen RM, Davidson PL. Improving access to care in America: individual and contextual indicators. In: Andersen RM, Rice TH, Kominski GF, editors. Changing the U.S. Health Care System: Key Issues in Health Services Policy and Management. Hoboken, NJ: Jossey-Bass Publishers; 2007.
19. Bass DM, Noelker LS. The influence of family caregivers on elder's use of in-home services: an expanded conceptual framework. J Health Soc Behav 1987 Jun;28(2):184-196. [doi: 10.2307/2137131]
20. Lederle M, Tempes J, Bitzer EM. Application of Andersen's behavioural model of health services use: a scoping review with a focus on qualitative health services research. BMJ Open 2021 May 05;11(5):e045018 [FREE Full text] [doi: 10.1136/bmjopen-2020-045018] [Medline: 33952550]
21. Li Y, Lu S. The development, application and implications of the Anderson Model in the field of healthcare. Chin J Health Policy 2017(12):77-82.
22. Wang YJ, Kaierdebieke A, Fan SY, Zhang RF, Huang MJ, Li H, et al. Study protocol: a cross-sectional study on psychology and behavior investigation of Chinese residents, PBICR. Psychosom Med Res 2022;4(3):19. [doi: 10.53388/202219]
23. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, et al. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13:948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
24. Duong TV, Aringazina A, Kayupova G, Nurjanah, Pham TV, Pham KM, et al. Development and validation of a new short-form health literacy instrument (HLS-SF12) for the general public in six Asian countries. Health Lit Res Pract 2019 Apr;3(2):e91-102 [FREE Full text] [doi: 10.3928/24748307-20190225-01] [Medline: 31294310]
25. Sun X, Lv K, Wang F, Ge P, Niu Y, Yu W, et al. Validity and reliability of the Chinese version of the Health Literacy Scale Short-Form in the Chinese population. BMC Public Health 2023 Feb 23;23(1):385-340 [FREE Full text] [doi: 10.1186/s12889-023-15237-2] [Medline: 36823591]

XSL•FO
RenderX

26. Sun X, Chen K, Wu Y, Wang F, Sun X, Tang J, et al. Development of a short version of the Health Literacy Scale based on classical test theory and item response theory. Chinese General Practice 2024;27(23):2931-2940 [FREE Full text] [doi: 10.12114/j.issn.1007-9572.2023.0072]
27. Chen G, Gully SM, Eden D. Validation of a new general self-efficacy scale. Organ Res Methods 2001 Jan 01;4(1):62-83. [doi: 10.1177/109442810141004]
28. Wu Y, Tang J, Du Z, Chen K, Wang F, Sun X, et al. Development of a short version of the perceived social support scale: based on classical test theory and ant colony optimization. BMC Public Health 2025 Jan 20;25(1):232 [FREE Full text] [doi: 10.1186/s12889-025-21399-y] [Medline: 39833852]
29. Li L, Peng T, Liu R, Jiang R, Liang D, Li X, et al. Development of the psychosomatic symptom scale (PSSS) and assessment of its reliability and validity in general hospital patients in China. Gen Hosp Psychiatry 2020;64:1-8. [doi: 10.1016/j.genhosppsych.2020.01.008] [Medline: 32070913]
30. Kwon HJ, Ahn TK, Lee JA, Sunwoo S, Kim YS, Kim BS, et al. The relationship between a spouse's alcohol use disorder and family communication. Korean J Fam Med 2015 Mar;36(2):92-102 [FREE Full text] [doi: 10.4082/kjfm.2015.36.2.92] [Medline: 25802690]
31. Guo N, Ho HC, Wang MP, Lai AY, Luk TT, Viswanath K, et al. Factor structure and psychometric properties of the family communication scale in the Chinese population. Front Psychol 2021;12:736514 [FREE Full text] [doi: 10.3389/fpsyg.2021.736514] [Medline: 34867617]
32. Kaakinen M, Sirola A, Savolainen I, Oksanen A. Shared identity and shared information in social media: development and validation of the identity bubble reinforcement scale. Media Psychol 2018 Nov 17;23(1):25-51. [doi: 10.1080/15213269.2018.1544910]
33. He W, Cao L, Liu R, Wu Y, Zhang W. Factors associated with internet use and health information technology use among older people with multi-morbidity in the United States: findings from the National Health Interview Survey 2018. BMC Geriatr 2022 Sep 06;22(1):733 [FREE Full text] [doi: 10.1186/s12877-022-03410-y] [Medline: 36068502]
34. Xie L, Yang HL, Lin XY, Ti SM, Wu YY, Zhang S, et al. Does the internet use improve the mental health of Chinese older adults? Front Public Health 2021;9:673368 [FREE Full text] [doi: 10.3389/fpubh.2021.673368] [Medline: 34336769]
35. Jaana M, Paré G. Comparison of mobile health technology use for self-tracking between older adults and the general adult population in Canada: cross-sectional survey. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e24718 [FREE Full text] [doi: 10.2196/24718] [Medline: 33104517]
36. Győrffy Z, Boros J, Döbrössy B, Girasek E. Older adults in the digital health era: insights on the digital health related knowledge, habits and attitudes of the 65 year and older population. BMC Geriatr 2023 Nov 27;23(1):779 [FREE Full text] [doi: 10.1186/s12877-023-04437-5] [Medline: 38012565]
37. Brunzini A, Caragiuli M, Atzori F, Bronzini M, Germani M. Digital technology for elders better living: a usability and user-experience assessment. In: Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environment. 2023 Presented at: PETRA '23; July 5-7, 2023; Corfu, Greece. [doi: 10.1145/3594806.3594846]
38. Das S, Jelen B, Kim A, Huber L, Camp LJ. Non-inclusive online security: older adults' experience with two-factor authentication. In: Proceedings of the 54th Hawaii International Conference on System Sciences. 2021 Presented at: HICSS 2021; January 5-8, 2021; Kauai, HI. [doi: 10.24251/hicss.2021.779]
39. Goudeau S, Sanrey C, Stanczak A, Manstead A, Darnon C. Why lockdown and distance learning during the COVID-19 pandemic are likely to increase the social class achievement gap. Nat Hum Behav 2021 Oct 27;5(10):1273-1281. [doi: 10.1038/s41562-021-01212-7] [Medline: 34580440]
40. Kemp E, Trigg J, Beatty L, Christensen C, Dhillon HM, Maeder A, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Promot J Austr 2021 Feb 21;32 Suppl 1(S1):104-114. [doi: 10.1002/hpja.387] [Medline: 32681656]
41. Schooley B, Singh A, Hikmet N, Brookshire R, Patel N. Integrated digital patient education at the bedside for patients with chronic conditions: observational study. JMIR Mhealth Uhealth 2020 Dec 22;8(12):e22947 [FREE Full text] [doi: 10.2196/22947] [Medline: 33350961]
42. Ilie DG, Neghină RA, Mănescu VM, Ganciu MR, Militaru G. New media, old problems: social stratification, social mobility and technology usage. In: Proceedings of the 14th International Technology, Education and Development Conference. 2020 Presented at: INTED2020; March 2-4, 2020; Valencia, Spain. [doi: 10.21125/inted.2020.1707]
43. Martin SP, Robinson JP. The income digital divide: trends and predictions for levels of internet use. Soc Probl 2007 Feb;54(1):1-22. [doi: 10.1525/sp.2007.54.1.1]
44. Beneito-Montagut R, Rosales A, Fernández-Ardèvol M. Emerging digital inequalities: a comparative study of older adults' smartphone use. Soc Media Soc 2022 Dec 05;8(4) [FREE Full text] [doi: 10.1177/20563051221138756]
45. Choi Y, Kim JH, Park EC. The effect of subjective and objective social class on health-related quality of life: new paradigm using longitudinal analysis. Health Qual Life Outcomes 2015 Aug 08;13:121 [FREE Full text] [doi: 10.1186/s12955-015-0319-0] [Medline: 26253140]
46. Cresci MK, Yarandi HN, Morrell RW. The digital divide and urban older adults. Comput Inform Nurs 2010;28(2):88-94. [doi: 10.1097/NCN.0b013e3181cd8184] [Medline: 20182159]

XSL•FO
RenderX

47. Olphert W, Damodaran L. Older people and digital disengagement: a fourth digital divide? Gerontology 2013;59(6):564-570. [doi: 10.1159/000353630] [Medline: 23969758]
48. Wu Y, Ware C, Damnée S, Kerhervé H, Rigaud A. Bridging the digital divide in older adults: a study from an initiative to inform older adults about new technologies. Clin Interv Aging 2015 Jan;10:193. [doi: 10.2147/cia.s72399]
49. Orłowska M, Błeszynska KM. Education as the factor of digital inclusion of elder persons: a study case in Poland. Eur Res Stud J 2021;XXIV(2):490-500. [doi: 10.35808/ersj/2783]
50. Martínez-Alcalá CI, Rosales-Lagarde A, Pérez-Pérez YM, Lopez-Noguerola JS, Bautista-Díaz ML, Agis-Juarez RA. The effects of Covid-19 on the digital literacy of the elderly: norms for digital inclusion. Front Educ 2021 Jul 9;6. [doi: 10.3389/feduc.2021.716025]
51. Lozoya SV, Guirado MA, Zapata Gonzalez AZ, Lopez AB, Zapata Gonzalez A. Use of technologies and self-efficacy in older adults. IEEE Rev Iberoam Tecnol Aprendizaje 2022 May;17(2):125-130. [doi: 10.1109/rita.2022.3166870]
52. Gatti FM, Brivio E, Galimberti C. “The future is ours too”: a training process to enable the learning perception and increase self-efficacy in the use of tablets in the elderly. Educ Gerontol 2017 Jan 09;43(4):209-224. [doi: 10.1080/03601277.2017.1279952]
53. Alvseike H, Brønnick K. Feasibility of the iPad as a hub for smart house technology in the elderly; effects of cognition, self-efficacy, and technology experience. J Multidiscip Healthc 2012;5:299-306 [FREE Full text] [doi: 10.2147/JMDH.S35344] [Medline: 23226024]
54. Tsai HY, Shillair R, Cotten SR. Social support and "playing around": an examination of how older adults acquire digital literacy with tablet computers. J Appl Gerontol 2017 Jan;36(1):29-55 [FREE Full text] [doi: 10.1177/0733464815609440] [Medline: 26491029]
55. Hale NM. Effects of age and interpersonal contact on stereotyping of the elderly. Curr Psychol 1998 Mar;17(1):28-38. [doi: 10.1007/s12144-998-1019-2]
56. Barbosa Neves B, Fonseca JR, Amaro F, Pasqualotti A. Social capital and internet use in an age-comparative perspective with a focus on later life. PLoS One 2018;13(2):e0192119 [FREE Full text] [doi: 10.1371/journal.pone.0192119] [Medline: 29481556]
57. Vercruyssen A, Schirmer W, Geerts N, Mortelmans D. How “basic” is basic digital literacy for older adults? Insights from digital skills instructors. Front Educ 2023 Sep 13;8 [FREE Full text] [doi: 10.3389/feduc.2023.1231701]
58. Rodríguez MD, Gonzalez VM, Favela J, Santana PC. Home-based communication system for older adults and their remote family. Comput Hum Behav 2009 May;25(3):609-618. [doi: 10.1016/j.chb.2008.08.017]
59. Hope A, Schwaba T, Piper AM. Understanding digital and material social communications for older adults. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, ON. [doi: 10.1145/2556288.2557133]
60. Baran AG, Alpaydın M. A qualitative study on skills of elders to use digital technology products from digital divide perspective. Elder Issues Res J 2020;13(2):107-122 [FREE Full text] [doi: 10.46414/yasad.788412]
61. Terras MM, Ramsay J. Family digital literacy practices and children's mobile phone use. Front Psychol 2016 Dec 23;7:1957 [FREE Full text] [doi: 10.3389/fpsyg.2016.01957] [Medline: 28066284]

Abbreviations

CART: classification and regression tree
FCS-10: Family Communication Scale-10
HLS-SF: Health Literacy Scale–short form
NGSES: New General Self-Efficacy Scale
PSSS: Perceived Social Support Scale

Edited by A Mavragani; submitted 11.10.24; peer-reviewed by X Wang, P Watson; comments to author 29.11.24; revised version received 18.12.24; accepted 06.01.25; published 04.03.25.

Please cite as:
Yan M, Sun W, Tan C, Wu Y, Liu Y
Analysis of Factors Influencing the Willingness of Chinese Older Adults to Use mHealth Devices: Nationwide Cross-Sectional Survey Study
J Med Internet Res 2025;27:e66804
URL: https://www.jmir.org/2025/1/e66804
doi:10.2196/66804
PMID:40053781

XSL•FO
RenderX

©Mengyao Yan, Wendi Sun, Cheng Tan, Yibo Wu, Yuanli Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Heart Disease–Related mHealth Apps in India: Systematic Search in App Stores and Metadata Analysis

Keerthi Dubbala[1], MBBS, MPH; Roshan Prizak[2], PhD; Ingrid Metzler[1], PhD; Giovanni Rubeis[1], PhD, Prof Dr

[1]Division of Biomedical and Public Health Ethics, Department of General Health Studies, Karl Landsteiner University of Health Sciences, Krems, Austria

[2]Karlsruhe Institute of Technology, Karlsruhe, Germany

Corresponding Author:
Keerthi Dubbala, MBBS, MPH
Division of Biomedical and Public Health Ethics
Department of General Health Studies
Karl Landsteiner University of Health Sciences
Dr. Karl-Dorrek-Straße 30
Krems, 3500
Austria
Phone: 43 6646194677
Fax: 43 2732 / 72090500
Email: keerthi.dubbala@kl.ac.at

Abstract

Background: Smartphone mobile health (mHealth) apps have the potential to enhance access to health care services and address health care disparities, especially in low-resource settings. However, when developed without attention to equity and inclusivity, mHealth apps can also exacerbate health disparities. Understanding and creating solutions for the disparities caused by mHealth apps is crucial for achieving health equity. There is a noticeable gap in research that comprehensively assesses the entire spectrum of existing health apps and extensively explores apps for specific health priorities from a health care and public health perspective. In this context, with its vast and diverse population, India presents a unique context for studying the landscape of mHealth apps.

Objective: This study aimed to create a comprehensive dataset of mHealth apps available in India with an initial focus on heart disease (HD)–related apps.

Methods: We collected individual app data from apps in the "medical" and "health and fitness" categories from the Google Play Store and the Apple App Store in December 2022 and July 2023, respectively. Using natural language processing techniques, we selected HD apps, performed statistical analysis, and applied latent Dirichlet allocation for clustering and topic modeling to categorize the resulting HD apps.

Results: We collected 118,555 health apps from the Apple App Store and 108,945 health apps from the Google Play Store. Within these datasets, we found that approximately 1.7% (1990/118,555) of apps on the Apple App Store and 0.5% (548/108,945) on the Google Play Store included support for Indian languages. Using monograms and bigrams related to HD, we identified 1681 HD apps from the Apple App Store and 588 HD apps from the Google Play Store. HD apps make up only a small fraction of the total number of health apps available in India. About 90% (1496/1681 on Apple App Store and 548/588 on Google Play Store) of the HD apps were free of cost. However, more than 70% (1329/1681, 79.1% on Apple App Store and 423/588, 71.9% on Google Play Store) of HD apps had no reviews and rating-scores, indicating low overall use.

Conclusions: Our study proposed a robust method for collecting and analyzing metadata from a wide array of mHealth apps available in India through the Apple App Store and Google Play Store. We revealed the limited representation of India's linguistic diversity within the health and medical app landscape, evident from the negligible presence of Indian-language apps. We observed a scarcity of mHealth apps dedicated to HD, along with a lower level of user engagement, as indicated by reviews and app ratings. While most HD apps are financially accessible, uptake remains a challenge. Further research should focus on app quality assessment and factors influencing user adoption.

(J Med Internet Res 2025;27:e53823) doi:10.2196/53823

XSL•FO
RenderX

KEYWORDS

mobile health apps; mHealth apps; heart disease; data collection methods; natural language processing; metadata analysis; Apple App Store; Google Play Store; mobile phone

Introduction

Background

Mobile health (mHealth) apps are prominent among the recommended interventions for addressing health disparities and improving health care accessibility. mHealth, as defined by the World Health Organization, encompasses the use of mobile devices like phones and wireless tools to enhance health care practices, boost health information access, and encourage positive behavioral changes [1,2]. In low-resource settings like India, mHealth offers a cost-effective means to bridge health care gaps [2-8].

India has experienced a substantial surge in mobile phone technology adoption, with more than 70% of urban and 50% of rural households using smartphones [9]. The number of smartphones in India has been estimated to reach 1 billion by the end of 2023 [10]. This widespread use presents a valuable opportunity for leveraging smartphone health apps (mHealth apps) to enhance health care delivery [11-14]. mHealth apps are also being used by community health workers, making them accessible to people without smartphones [15]. During the COVID-19 pandemic, mHealth apps proved particularly valuable by enabling teleconsultations, contact tracing, and web-based prescription services, effectively overcoming barriers to traditional health care access [16-19].

However, mHealth apps can also contribute to increasing systemic inequities. They may risk widening health disparities through a lack of inclusivity and by addressing the health needs of already advantaged groups [20]. Language options in the mHealth apps, or the lack of them, have been highlighted in previous studies [21-25]. In addition to social determinants of health, digital determinants like technology access and community infrastructure (eg, broadband internet) can have a crucial impact on mHealth outcomes and health [25-30]. The COVID-19 pandemic has highlighted how limited access to digital technologies or supporting infrastructures, apps unsuited for target groups, and lack of high-quality data for tailoring interventions to the needs of marginalized groups might undermine the very potential of digital technologies for mitigating health disparities [31]. Therefore, understanding health disparities caused by mHealth apps and creating solutions for working toward health equity is crucial and requires several research efforts.

Although there is limited research on mHealth apps in India, there are many studies on mHealth apps in general. Many are systematic reviews looking at previous literature on individual or small groups of apps [25-30,32]. There are also reviews [4,33] and policy proposals [34] on mHealth apps. Many of these studies focused on the quality, privacy, and security concerns of mHealth apps [35-40]. There are also studies that focused on rating the quality of mHealth apps, such as the one using the MARS (Mobile Application Rating Scale) tool [41,42]. While assessing the quality of apps is essential, this is still manual, time-consuming, and does not give an overall picture of the mHealth apps available. There is a noticeable gap in research that comprehensively assesses the entire spectrum of existing health apps or extensively explores apps for specific health priorities from a health and health care perspective. This gap is even more evident in the context of mHealth apps available in India.

While some studies have investigated the landscape of mHealth apps in app stores, they have used methodologies based primarily on manual searching and screening of apps. Among those, Siddiqui et al [43] focused on orthodontic-related apps, Bassi et al [16] on COVID-19–related apps, and Coulon et al [27] on stress management–related apps. From their search processes, they gathered 952, 4600, and 902 apps, respectively, which were then manually screened and reviewed. However, searching app stores and screening apps manually is a time- and labor-intensive process [27], especially when exploring a large number of apps. Moreover, app store search functions tend to prioritize popular or highly rated apps, which, while useful for analyzing the most widely used apps, limits the ability to gain a comprehensive view of the app landscape. This approach may overlook lesser-used apps that could provide insights into why some apps fail to gain traction.

A significant obstacle in carrying out more extensive studies focusing on a diverse and larger set of mHealth apps is the difficulty of acquiring such a comprehensive dataset. There are some publicly available datasets [35] obtained using large-scale web crawling methods, with information about many apps available on such app stores. However, such datasets do not record all available information for each app and are not usually updated with the latest information. Furthermore, since the data on these datasets are not collected with a specific intention, they may not be entirely suitable for a particular research question. While Danahiswari et al [44] and Fu [45] have used such datasets and additional scraping methods in the context of mHealth apps, their focus was app development and market analysis. Other researchers also used similar scraping methods. In 2015, using web scraping methods, Xu and Liu [46] and Dehling et al [47] made datasets of 24,405 and 60,000 mHealth apps, respectively. More recently, in 2020, Tsinaraki et al [48] analyzed COVID-19–related apps from the 2 app stores using updated scraping methods. However, accessing these app stores, especially Google Play Store, by similar methods has become significantly more challenging due to more restrictive database designs, throttling limits, and IP blocking [48]. Stach et al [35] looked at mHealth apps and proposed a pipeline to search for specific types of apps from the app stores in their crawling techniques, overcoming the restrictions posed by these app stores.

Objective

Building on the above methods, we aimed to create a comprehensive dataset of mHealth apps using an automated and replicable method for collecting and analyzing metadata

XSL•FO
RenderX

from mHealth apps. We present an updated web crawling method for obtaining information about mHealth apps in a particular country. This method can help map the landscape and enable the analysis of mHealth apps. We used these methods to prepare a dataset of mHealth apps available in India from 2022 to 2023 on the Apple App Store and the Google Play Store. As a country with a large and diverse population, India faces unique health care challenges that can be effectively addressed through innovative technological solutions [4,49]. By exploring the landscape of mHealth apps, we can gain insights into their effectiveness and relevance and identify gaps and potential for improvement.

As a crucial first step in mapping these apps, we concentrated on apps relevant to heart disease (HD) using natural language processing (NLP) and clustering techniques. HD cause the most significant number of deaths in India and, therefore, are a crucial health priority to focus on [49]. This evidence can serve as a foundation for shaping health care policies, improving access, and promoting positive health outcomes.

Methods

Overview

The methodology in this study was developed to address the need for a comprehensive and updated dataset of apps available in India, with a focus on HD apps. Our approach builds on previous studies that have used automated methods for collection of metadata from mHealth apps [27,42-46] and aimed to overcome the challenges posed by restrictive database designs, throttling limits, and IP blocking by using a replicable web crawling method. We used analysis techniques, such as NLP and clustering, which have been used in health care research for various applications [50-54]. These techniques enabled us to identify apps related to HD and group these apps further on the basis of their focus and utility. Our methodology facilitates the effective gathering of a comprehensive dataset and allows for the process to be scaled and adapted for future research.

In this section, we will first discuss our data collection methodology. Then, we will discuss the NLP methods we used for identifying HD apps, followed by statistical analysis and clustering of selected apps. Figure 1 illustrates all the steps included in this methodology.

Figure 1. Methodology for extracting mobile health (mHealth) apps and selecting and analyzing apps for heart diseases. †While for Apple App Store, at this stage we already retain only mHealth apps; §Other attributes include installs, app sizes, ratings, reviews, rating scores; *Apple App Store apps do not have summary included. IHD: ischemic heart disease; NLP: natural language processing.

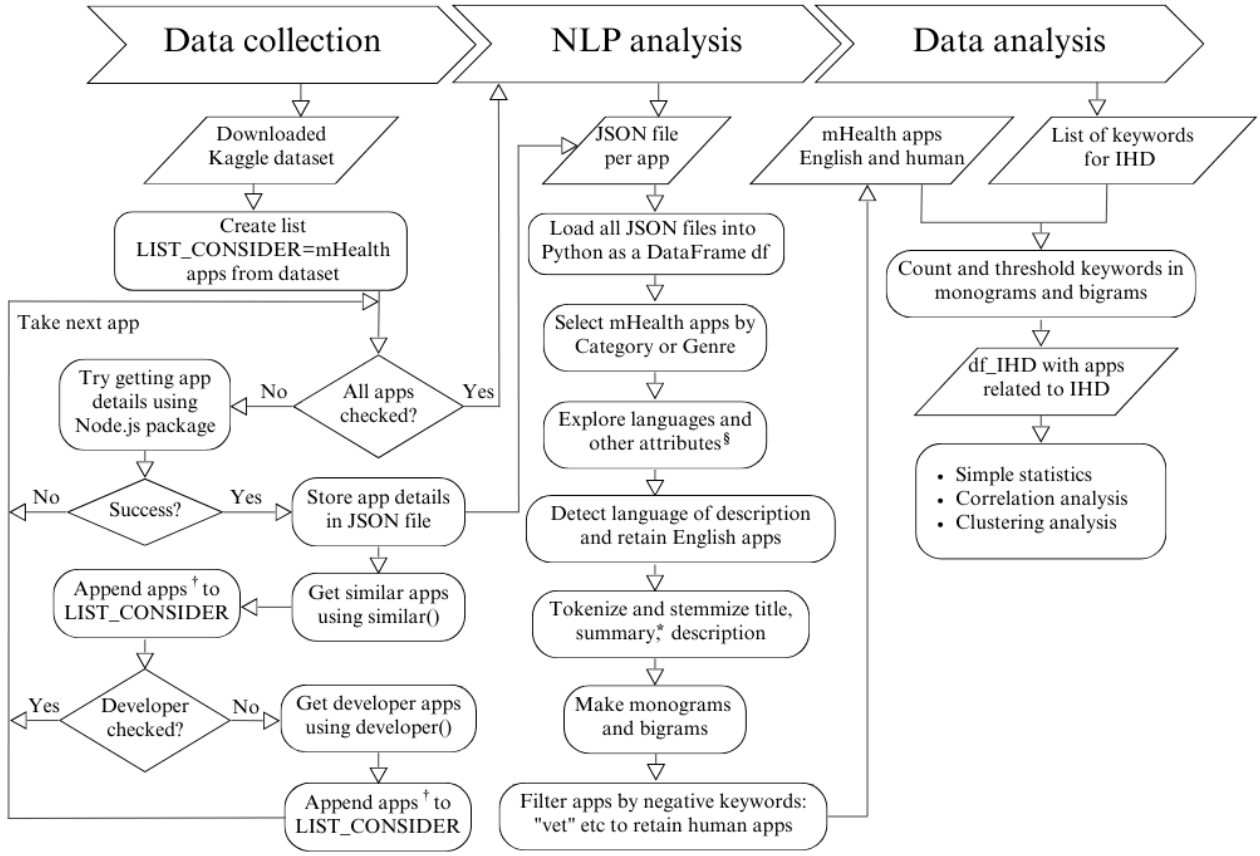

XSL•FO
RenderX

Data Collection

The 2 final datasets used for analysis contained app metadata from the "medical" and "health and fitness" categories on the Google Play Store in December 2022 and on the Apple App Store in July 2023, providing a snapshot of the app stores in these 2 months. This metadata includes app attributes, such as a unique app ID, the app's title, the provided description, the number of installs, and the price of the app, among other such details. A complete list of attributes is included in Table 1.

Table 1. Attributes of the apps in the datasets.

Attribute description	Attribute of the apps from the Apple App Store	Attribute of the apps from the Google Play Store	Attribute used for
Unique identification	ID	ID	Identifying individual apps and preventing duplicates during data collection.
App name	Title	Title	Tokens, monograms, and bigrams
App summary	N/A[a]	Summary	Tokens, monograms, and bigrams
App description	Description	Description	Tokens, monograms, bigrams, and number of characters (description length)
Release date	Released	Released	Tracking number of apps released over the years
App size	Size	Size	Average size of apps on each platform
Required operating system	Required_os	Android_2022	Most used versions of operating systems
Price of the app	Price	Price	Price of apps in INR[b]
Rating-score (ie, star-rating)	Score	Score	Average rating-scores
Number of reviews	Reviews	Reviews	Average number of reviews
App's category	Category	Category	To select health and medical apps
Presence of in-app purchases	N/A	IAP[c]	Number of apps with in-app purchases
Number of ratings	N/A	Ratings	Average number of ratings
Number of installs (ie, downloads)	N/A	installs_2022	Average number of installs
All the languages the app is available in	Languages[d]	—[e]	Distribution of languages in apps
Updated date	Updated	✓[f]	Tracking the number of apps updated over the years

[a]N/A: not available.

[b]INR: Indian national rupee.

[c]IAP: in-app purchases.

[d]Apple App Store lists many languages (ie, those in which the app can be used) in the languages attribute of a given app.

[e]Google Play Store does not have an attribute called languages. In both cases, for filtering apps based on language, we detected the language of the app description. This is illustrated in detail in Multimedia Appendix 1.

[f]Google Play Store does not have the attribute for updated year at the time of data collection.

We obtained the initial datasets from (1) a Kaggle dataset [55] containing >2.3 million apps scraped from the Google Play Store in June 2021 and (2) a Kaggle dataset [56] containing >1.2 million apps scraped from the Apple App Store in October 2021.

As Google Play Store has changed its internal data structure, it has become extremely challenging to use crawling tools to scrape its whole database. In such a context, the publicly available Kaggle datasets from recent years offer a great starting point to expand and compile the latest dataset. As these datasets are publicly available, researchers can obtain them as starting points for future efforts to expand such datasets or already perform analysis on them. An important note is that the Kaggle datasets do not have all the necessary information like app descriptions, which we therefore obtained through our scraping methods as follows.

For each of the app stores, we extracted app-specific unique identifiers (ie, *app_id*s) to create 2 separate lists called LIST_CONSIDER, comprising apps from the "medical" and "health and fitness" categories.

We then used the *Node.js* modules *google-play-scraper* [57] and *app-store-scraper* [58] to access updated app details from the Google Play Store and the Apple App Store in December 2022 and July 2023, respectively, and to expand our dataset of "medical" and "health and fitness" apps. For each *app_id* in LIST_CONSIDER, we performed the following three steps:

(1) We queried the app's metadata and stored the returned metadata in a JSON file named after its *app_id*. If the app was

XSL•FO
RenderX

not available on the queried date, we skipped the next 2 steps and considered the next app in LIST_CONSIDER.

(2) We obtained a list of other apps deemed as "similar" to the app using the "similar()" function. This function uses an internal algorithm to suggest similar apps given an input app. From this list of apps, we filtered to retain only "medical" and "health and fitness" apps and added their corresponding *app_id*s to the end of LIST_CONSIDER.

(3) We obtained a list of other apps developed by the same developer of the app using the "developer()" function. We again filtered this list of apps to retain only "medical" and "health and fitness" apps and added their corresponding *app_id*s to the end of LIST_CONSIDER.

As we kept adding more apps to LIST_CONSIDER, as described above, our dataset grew. We repeated this process until LIST_CONSIDER was exhausted, and no new *app_id*s could be added. In the end, we had 2 sets of JSON files—one for the Google Play Store and the other for the Apple App Store.

Filtering Medical and Health and Fitness Apps

We analyzed these 2 sets of JSON files as separate datasets (ie, Google dataset and Apple dataset). The 2 datasets were never added together into a single analysis but only compared against each other. To start with, we used 2 steps of filtering as follows: (1) the main language used in the description should be English, and (2) the app should be concerned with human health and not with animal health. To achieve the first filtering step, we detected the main language of the description using the *langdetect* package in Python and selected only those with descriptions in English language. This step was important to use NLP techniques later. Then, from the resulting English language apps, we removed veterinary and animal health apps using NLP techniques. To identify apps related to animal health topics, we looked for the presence of a preselected set of veterinary and animal health keywords in the app title, summary (ie, present only in the Google dataset), or description. We discuss these techniques in detail in the section for selecting apps for HD.

Selecting Apps Related to HD

NLP applies computational techniques to analyze natural language and speech [59]. NLP treats textual content as data and offers a variety of methods for efficiently processing substantial amounts of text [60]. In this study, we used the app titles, app summaries (ie, present only in the Google dataset), and app descriptions as our text data. We used tokenization, lemmatization, stop-word removal, and n-grams as part of our NLP methodology [60]. Tokenization identifies individual words or tokens within a given text. Lemmatization transforms each word into its root form by removing suffixes. For instance, lemmatization consolidates word forms like "diagnose," "diagnosis," "diagnosed," and "diagnosing" into the common root word "diagnose." By treating different word forms as a single entity, lemmatization enables accurate analysis in further steps. After eliminating stop-words like "and," "or," "also," and "then" as well as special characters, we made monograms and bigrams from the text. Monograms are single words, while bigrams consist of pairs of consecutive words. These allow for the analysis of not only individual words by themselves but also in the context of their use. For example, the words "heart" and "disease" each have some meaning in the health context, whereas "heart disease" as a pair conveys a specific meaning. Using monograms and bigrams allows us to capture and use both contexts in our analysis.

We made a "dictionary" of terms relevant to HD using Medical Subject Headings terms from PubMed database [61] and added more terms from Google searches. This dictionary is listed in Table 2.

Table 2. Heart diseases—keyword dictionary.

Categories	Keywords
Diseases	heart disease, cardiovascular disease, coronary artery disease, myocardial infarction, ischemic heart disease, angina, atherosclerosis, hypertensive heart disease, valvular heart disease, congenital heart disease, rheumatic heart disease, heart failure, cardiomyopathy, arrhythmia, atrial fibrillation, ventricular fibrillation, supraventricular tachycardia, bradycardia, heart murmur, pericarditis, endocarditis, myocarditis, Kawasaki disease, systemic lupus erythematosus, rheumatoid arthritis, giant cell myocarditis, viral myocarditis, dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, blocked arteries, clogged arteries, heart blockage, heart attack, and cardiac arrest
Symptoms	chest pain, shortness of breath, fatigue, edema, high blood pressure, hypertension, palpitations, chest discomfort, chest tightness, heartache, breathlessness, labored breathing, heavy breathing, racing heart, rapid heartbeat, irregular heartbeat, skipped beats, swelling, and fluid retention
Treatments or procedures	coronary artery bypass graft, percutaneous coronary intervention, stent, echocardiogram, electrocardiogram, cardiac catheterization, cardiac magnetic resonance imaging, cardiac rehabilitation, bypass surgery, heart surgery, open-heart surgery, coronary bypass, bypass surgery, stent surgery, stent placement, balloon procedure, valve surgery, heart valve repair, valve replacement, cardiac rehabilitation, heart transplant, new heart, and artificial heart
Medications	beta-blockers, calcium channel blockers, angiotensin-converting enzyme inhibitors, diuretics, statins, antiplatelet medications, anticoagulants, angiotensin receptor blockers, blood thinners, clot-busting drugs, clot-preventing medication, blood pressure medicines, and cholesterol-lowering drugs
Medical devices	pacemaker, defibrillator, shock device, and electrical device
Medical specialties	cardiology, cardiologist, and cardiothoracic

XSL•FO
RenderX

We made monograms and bigrams from this HD keyword dictionary (ie, HD monograms and bigrams). We also made monograms and bigrams for each app's title, summary, and description (ie, app monograms and bigrams). Then, for each app, we counted the frequency of HD monograms and bigrams in the app monograms and bigrams, which we used to select apps relevant to HD. We tried different iterations and found that selecting apps with (1) at least 1 *relevant* HD monogram or (2) at least 1 HD bigram is an optimal method. We eliminated apps with only one monogram if the monogram was not exclusive to HD—for example, "pain." However, apps with single and HD-relevant monograms, such as "ECG" (electrocardiogram), were selected. Apps without any HD monograms or bigrams in their app monograms and bigrams were not selected.

Simple Statistics and Correlation Analysis of HD Apps

We explored the HD apps using various app attributes like app sizes, app prices, number of ratings, rating-scores, number of reviews, presence of in-app purchases, the release year of each app, and the length of description. We used descriptive statistics like arithmetic mean to understand the distribution of the apps along the mentioned attributes. We also performed correlation analysis to understand the relationships among these attributes.

Clustering and Topic Modeling of HD Apps

To understand the different types of apps present in the selected HD apps, we performed cluster analyses. We used the latent Dirichlet allocation (LDA) algorithm for clustering and topic modeling using app descriptions as input. This technique allows us to uncover underlying topics within the dataset. We used the *CountVectorizer* function from the *scikit-learn* package to convert the description text into a matrix of token counts. The matrix consists of the counts of the frequency of individual words (ie, tokens) across all the app descriptions. This matrix was then subjected to LDA using the *LatentDirichletAllocation* function from *scikit-learn*. After several iterations using different parameters, we configured the LDA model to identify 3 topics (ie, clusters) using the following parameters: *n_components*=3, max_iter=5, *learning_method='online'*, and learning_offset=50. Each app's cluster assignment was determined based on the topic with the highest probability in its distribution.

We explored the resulting clusters, but we did not find any further meaningful subclusters. To address specific research inquiries, alternative supervised topic modeling methods could be explored to extract more nuanced subclusters. We also experimented with k-means clustering [44], but our results indicated that LDA clustering yielded superior outcomes with this dataset. Clouds from k-means clustering are attached in Multimedia Appendix 2.

Tools Used

In general, we used *Python* as the main scripting language, the *pandas* [62] package for data analysis, and *matplotlib* [63] for data visualization. To expand the list of apps by crawling and scraping the app stores, we used *Node.js*—modules "*google-play-scraper*" [57] and "*app-store-scraper*" [58] for the Google Play Store and the Apple App Store, respectively. We wrapped the *Node.js* scripts inside a *Python* script that iteratively looped over the Kaggle dataset and expanded the list of apps with their details. Next, for NLP tasks like tokenization and stemmization, we used the *nltk* [64] package. For language detection, we used the *langdetect* [65] package and for word cloud visualizations, we used the *wordcloud* [66] package. For further NLP analysis, we used vectorizers and methods from the *scikit-learn* package, and for LDA visualization, we used the *pyLDAvis* [67] package.

The code written and used for data collection and analysis in this study is available in our GitHub repository [68].

Ethical Considerations

No ethics board review was needed as our study does not use patient data or involve human participants.

Results

Overview

Using the data collection methods described in the previous section, we scraped 118,555 apps from the Apple App Store and 108,945 apps from the Google Play Store categorized as "health and fitness" or "medical" (ie, health and medical apps). Figure 2 illustrates a summary of the results discussed below.

XSL•FO
RenderX

Figure 2. Summary of results. HD: heart disease; NLP: natural language processing.

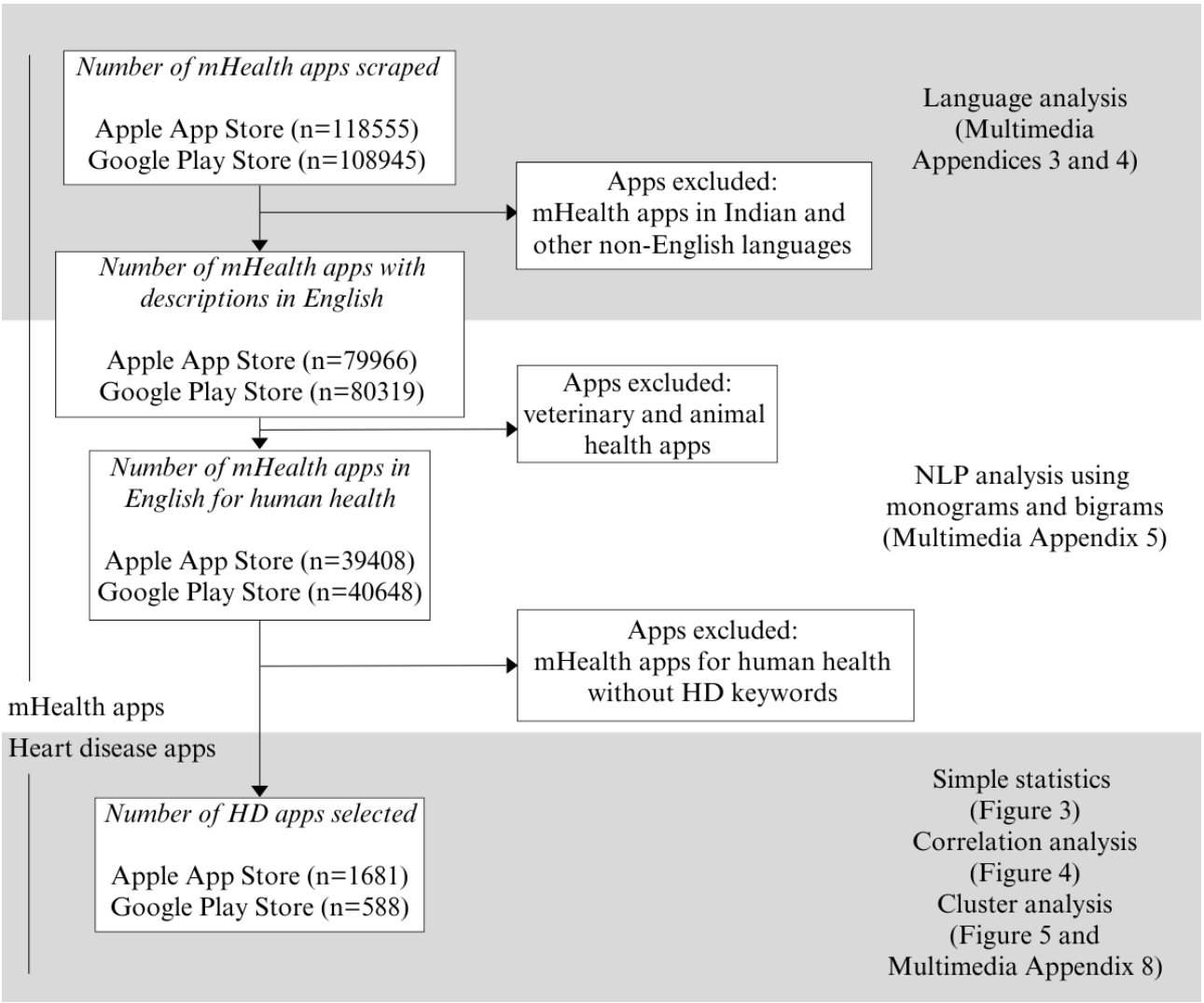

App Language

An important consideration here is the presence of the app attribute *language* in the Apple App Store dataset, and its absence in the Google Play Store dataset. Apple App Store's *language* attribute lists all the languages an app supports and is exactly the information needed to evaluate the language support of an app. However, the Google Play Store dataset lacks this attribute, so it is not possible to know which languages it supports. Hence, as an indirect indication, we used the main language of the app description detected by the *langdetect* package as Google Play Store's stand-in attribute for languages.

In the resulting health and medical apps, Apple App Store has more than 90% (109,425/118,555, 92.3%) apps that support English whereas Google Play Store has 73.72% (80,319/108,945) apps in English (ie, descriptions in English). We found that approximately 1.7% (1990/118,555) of apps on the Apple App Store included support for Indian languages and 0.5% (548/108,945) of apps (ie, descriptions) on the Google Play Store were in Indian languages. Among Indian languages, apps supporting Hindi makes up the majority on Apple App Store whereas apps in Bengali are the majority on Google Play Store followed by Hindi. Although there were only a few health and medical apps on Google Play Store in Indian languages, these apps had higher average rating-scores, installs, and higher average number of reviews compared to others. The distribution of apps in different languages and the average rating scores and number of installs and reviews of apps in Indian languages are presented in Multimedia Appendices 3 and 4, respectively.

As most of the apps were in English language, we continued with these apps for further analysis, excluding apps in Indian and other non-English languages. After screening for apps with descriptions in the English language, and removing veterinary and animal health apps, we had 39,408 apps from the Apple App Store and 40,648 apps from the Google Play Store.

HD Apps

Using HD monograms and bigrams, we selected 1681 HD apps from the Apple App Store and 588 HD apps from the Google Play Store. These HD apps had elevated frequencies of various HD monograms and bigrams as expected. A list of frequently occurring keyword monograms and bigrams in selected HD apps are shown in Table 3 and the corresponding word clouds are presented in Multimedia Appendix 5.

XSL•FO
RenderX

Table 3. Keyword monograms and bigrams found in (selected) heart disease apps.

Store	Keywords
Apple App Store	
Monograms	drug, beat, race, heart, cardiovascular, block, rapid, cardiac
Bigrams	chest pain, cardiac arrest, heart attack, short breath, supraventricular tachycardia, valvular heart, skip beat, irregular heartbeat
Google Play Store	
Monograms	heart, blood, pain, cardiovascular attack, chest, short, cardiac
Bigrams	heart attack, chest pain, cardiac arrest, irregular heartbeat, chest discomfort, valvular heart, supraventricular tachycardia, rapid heartbeat

Exploring HD Apps

The distributions of various app attributes for HD apps from the Apple App Store and the Google Play Store are shown in Figure 3. There had been an increase in the number of HD apps released over the years on both the Apple App Store and the Google Play Store. Many of the Apple App Store HD apps had been updated in the last year (Figures 3E and 3F). The average size of HD apps was 76 MB on the Apple App Store and 27 MB on the Google Play Store (Figure 3A). Very few apps had rating-scores, and when they did, they mostly had scores>3 (Figure 3B). Similar to rating-scores, very few apps had reviews (Figure 3C). Among those apps with reviews, there were rarely >1000 reviews. Data for the number of ratings and installs were available only for Google Play Store apps. Many apps did not have any ratings or installs. Among the apps with number of ratings and installs, the average was around 2000 (Figures 3J and 3K).

Figure 3. Analysis of attributes for heart disease apps on Apple App Store (N=1681) and Google Play Store (N=588).

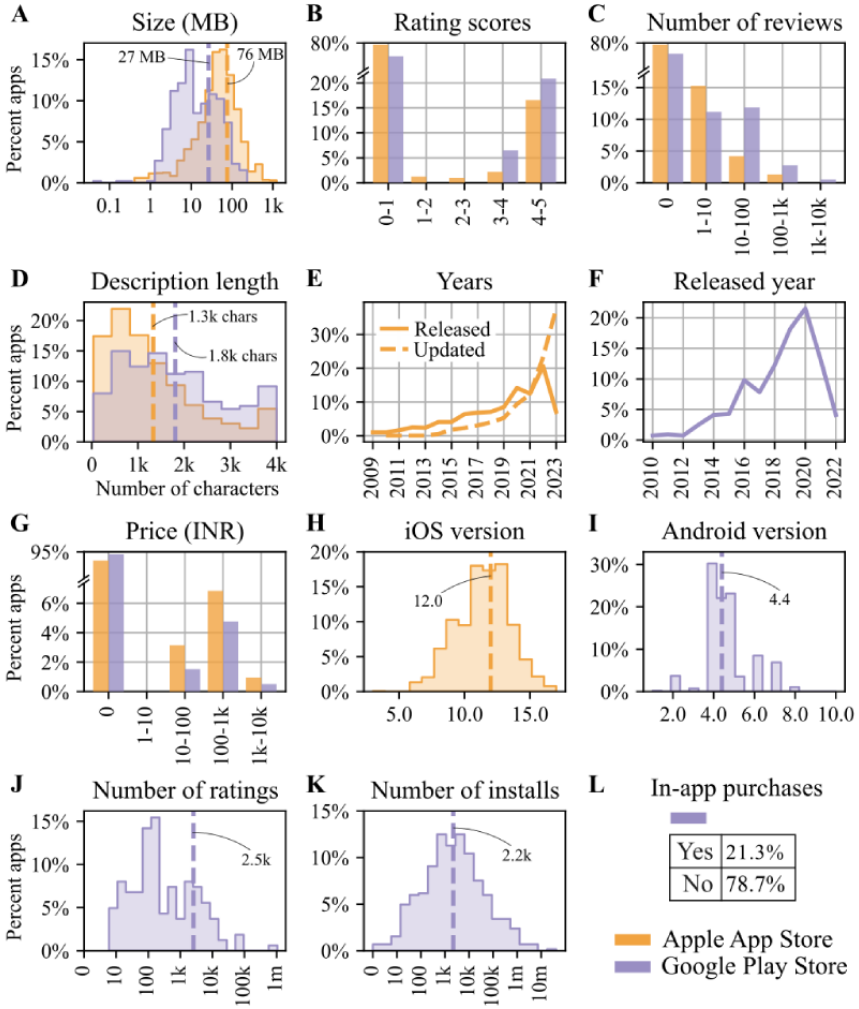

About 90% (1496/1681 on Apple App Store and 548/588 on Google Play Store) of the HD apps were free on both platforms. Among those apps that were not free, prices were mostly between 100 and 1000 INR (Indian national rupee; US $1.2 and US $12; Figure 3G). About 20% (125/588, 21.3%) of the HD apps on the Google Play Store had in-app purchases (Figure

XSL•FO
RenderX

3L). In-app purchase data was not available for Apple App Store HD apps.

Google Play Store HD apps had an average of 1800 characters and Apple App Store HD apps had 1300 characters in their descriptions (Figure 3D). Many HD apps did not need the latest operating systems (OS) and were functional with earlier OS versions (Figures 3H and 3I).

Correlation Analysis

Figure 4 shows the correlations between pairs of attributes for Apple App Store and Google Play Store HD apps as heatmaps. Correlation matrices are included in Multimedia Appendices 6 and 7. The number of installs for HD apps on Google Play Store was positively correlated with the number of ratings and the number of reviews. On both Apple App Store and Google Play Store, we found a positive correlation between the length of description and rating-scores, and a negative correlation between app release year and rating-score. For HD apps on the Apple App Store, there was a positive correlation between the size and price of the apps.

Figure 4. Explorative correlation analysis between the pairs of app attributes.

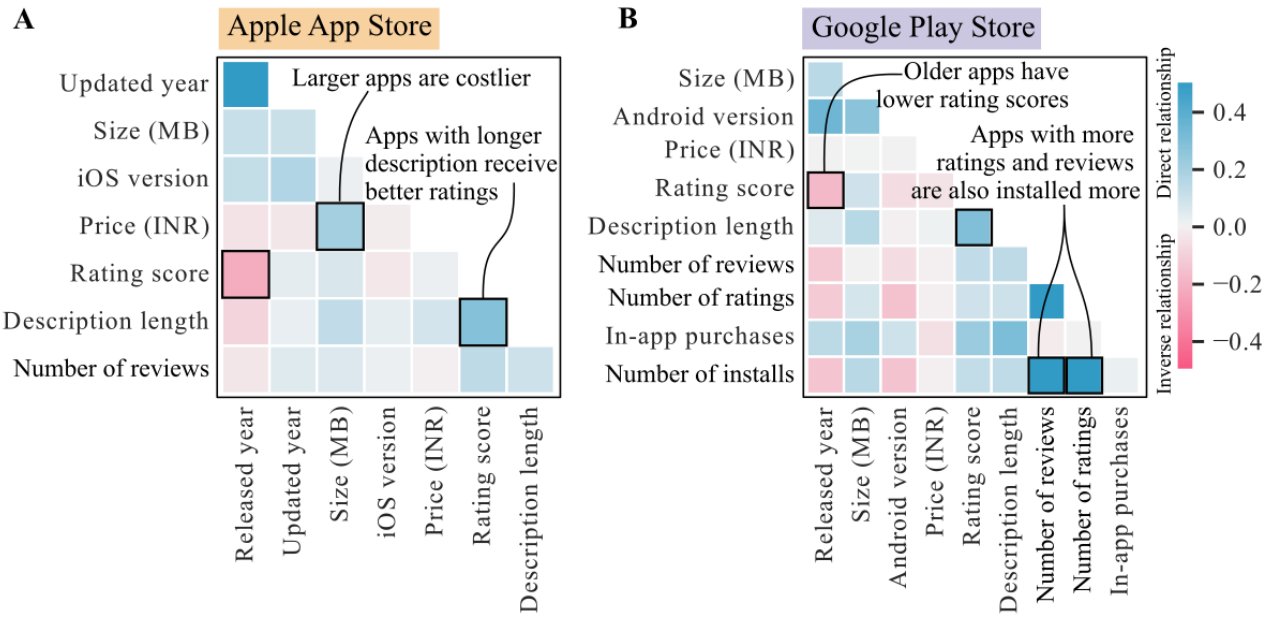

Cluster Analysis of HD Apps

Using the LDA topic modeling methods, we identified 3 distinct clusters in both Apple App Store and Google Play Store HD apps. After investigating the topics they dealt with, we assigned them the following names: clinical cluster, fitness and lifestyle cluster, and sleep and wellness cluster. Table 4 shows the number of apps in each cluster and its most relevant terms in app descriptions. An expanded set of top 20 relevant terms in each cluster, their frequency within clusters and their overall frequency in all apps are presented in Figure 5. LDA modeling results using the *pyLDAvis* package and the resulting word clouds for each cluster are included in Multimedia Appendix 8.

Because these clusters were made from apps in 2 categories (ie, “medical” and “health and fitness”), we also checked the distribution of the categories within the clusters (Multimedia Appendix 9). We found that the clinical cluster had many apps from “medical” category. However, it was not exclusive to “medical” category apps. Likewise, apps categorized as “medical” were also found in the other 2 clusters, albeit significantly less in number.

We explored the distribution of app attributes for each of these clusters separately (Figure 6), similar to the analysis above on all HD apps. We found that clinical apps were comparatively smaller in size and had shorter descriptions. Price was higher for apps in the fitness and lifestyle cluster.

Table 4. Number of apps in the various clusters from latent Dirichlet allocation cluster analysis.

Cluster name	Number of Apple App Store HD[a] apps (N=1681), n (%)	Number of Google Play Store HD apps (N=588), n (%)	Most relevant terms for each topic
Clinical	817 (48.6)	286 (48.63)	Disease, medical, cardiac, risk, pain, treatment, and doctor
Fitness and lifestyle	594 (35.33)	136 (23.12)	Diet, workout, training, and fitness
Sleep and wellness	270 (16.06)	166 (28.23)	Sleep, yoga, meditation, and stress

[a]HD: heart disease.

XSL•FO
RenderX

Figure 5. Top 20 most relevant terms by clusters from latent Dirichlet allocation.

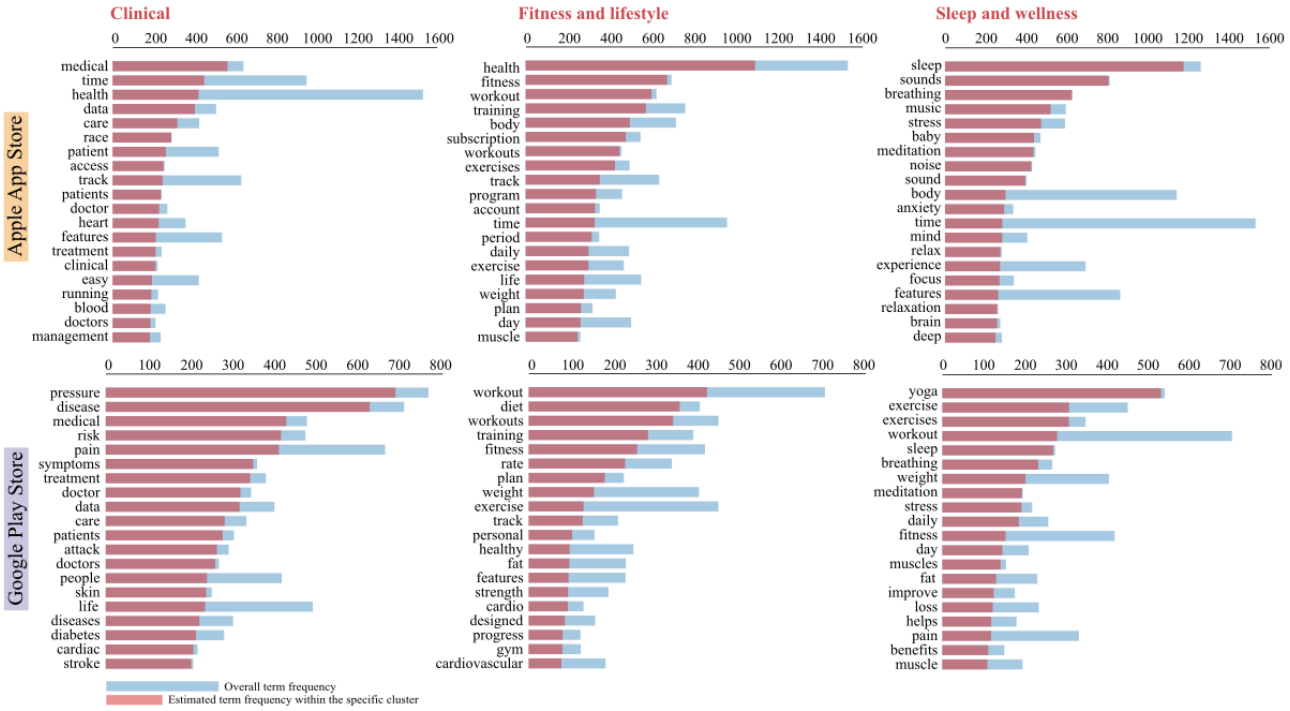

Figure 6. Distributions of various attributes (ie, metadata) of heart disease apps grouped by clusters.

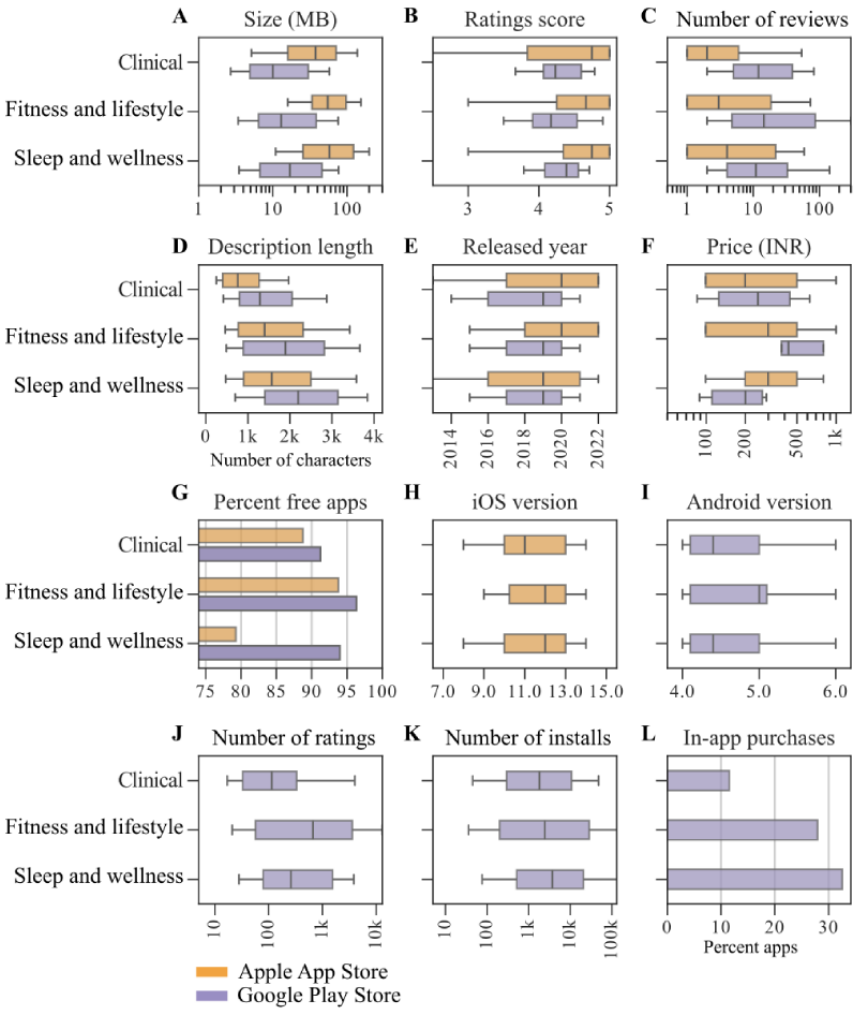

XSL•FO
RenderX

Discussion

Languages in Health and Medical Apps

The language diversity within health and medical apps is a notable concern based on our findings. In India, the availability of health and medical apps in Indian languages remains limited, representing only a small fraction of the total. Around 64% of internet users from rural India and 25% of urban India access the internet in their regional language [21]. This shows that most of the mHealth apps that are available in India are not accessible to those who do not understand English, or prefer a regional language.

Previous research has identified the lack of available health apps in native languages in India as a major access barrier. In a study conducted to understand the acceptability of app-based interventions in India, Sinha et al [21] found that language was a significant barrier to adopting health apps, as most of them are available in English. The study by Singh et al [22] focused on privacy policies and directly assessed app descriptions of mental health apps in app stores. Their findings suggest that the lack of native language translation of app description negatively affects the ability of users to understand and evaluate privacy policies.

Outside of India, Huang et al [23] performed a screening of diabetes self-management apps on the Google and Apple app store platforms. They found that most diabetes self-management apps only had English-language descriptions, thereby making them inaccessible to non–English speakers. Haggag et al [24] looked at the mHealth app user reviews on Apple and Google app stores in Australia, the United States, and the United Kingdom. They found that the users faced considerable challenges when they were not able to use the apps in their native language. In a systematic review conducted to study the health equity considerations in mHealth interventions for caregivers of older adults, Garnett et al [25] found that only 11% of app providers considered the provision of alternate language options. These numbers are much worse in Indian mHealth apps. However, despite being fewer in number, apps in Indian languages seem to have better user engagement, as indicated [47] by their higher installs, average ratings, and number of reviews compared to others. This highlights a strong case for developing more health apps in Indian languages.

Health Priorities: Number of HD Apps

A crucial requirement for mHealth apps is to address existing health issues [69]. Our findings indicate that this is only partly the case in India. Compared to the thousands of apps available in India in the “medical” and “health and fitness” categories, the number of relevant apps for HD, which is number one in the top 10 causes of death in India [49], is exceedingly low. In a scoping study on mHealth initiatives in India in 2016, Bassi [70] searched the Google Play Store and the Apple App Store and found 81 and 107 apps, respectively for HD. Although the numbers have multiplied over the years, the fraction of apps for HD is still small.

While the selection method could be improved to find more apps, if the apps do not have some of the basic keywords, it would be just as difficult for a typical smartphone user to find them as it was for us. Moreover, if the app is not easy to find, it can add to the accessibility barrier from a technological perspective.

We found that the Apple App Store had more HD health and medical apps compared to the Google Play Store, in line with previous research. Dehling et al [47] found that the ratio of Apple mHealth apps to Google mHealth apps was about 9 to 1. However, the numbers for overall mHealth apps in our study (ie, after screening for apps in English language and eliminating animal health-related apps) were comparable on both platforms.

App Size

We found that the average size of HD apps on the Apple App Store and on the Google Play Store in this study was higher compared to previous studies on mHealth apps. Coulon et al [27] reported that the average size of apps for stress management on the Apple App Store was 60.5 MB. Siddiqui et al [43] found that most of the orthodontic mHealth apps on the Apple and Google app stores were <40 MB in size.

Considerations of app size are important due to their impact on the ease of downloading and using these apps. Research has consistently demonstrated that larger apps tend to prolong the downloading process and increase user burden during their use [71]. In addition, Haggag et al [24] underscored that those individuals with older smartphones or limited storage capacity reported significant challenges when engaging with mHealth apps.

The implication of this issue is particularly pronounced in low-resource settings, where access to more advanced smartphones and reliable internet connectivity for downloading and using these large apps is a persistent concern [46,72]. Addressing these logistical challenges and developing easily usable apps is essential in ensuring equitable access to mHealth apps, especially in Indian settings where resources are constrained.

Price of Apps and In-App Purchases

We found that most of the HD apps were offered free of charge, aligning with observations made in prior studies [34]. For the minority of apps that did entail a cost, we identified that prices of apps were relatively higher on the Apple App Store compared to the Google Play Store. Furthermore, about 20% (125/588, 21.3%) of these apps featured in-app purchases on the Google Play Store. While it was noted that some paid health apps had shown increased user engagement [73], in general, both price and in-app purchases have been identified as potential barriers to the adoption and use of mHealth apps [40,74].

In low-resource settings, the appeal of mHealth apps often revolves around their potential to mitigate barriers to health care access, including the financial burdens of health care [75]. In this context, it is encouraging to observe that most HD apps are accessible to smartphone users free of cost. However, it is crucial to highlight that free apps may come with limited access to a wide range of interventions, comprehensive programs, and qualified health care professionals [75].

XSL•FO
RenderX

Ratings and Installs: Google Play Store

The absence or scarcity of ratings has been linked to limited widespread use [47]. Dehling et al [47] found that apps with fewer ratings were less likely to be downloaded and used, possibly due to their diminished visibility in app stores relative to the number of ratings. Moreover, users tend to view apps with a greater number of ratings as more credible and trustworthy.

HD app installs on the Google Play Store vary widely, ranging from 1 to 1 million, with an average of approximately 2000 installs. It is important to note that while the number of installs provides an indication of app use, it does not necessarily reflect the number of active users.

Rating-Scores and Reviews

Prior research indicates that app rating-scores do not strongly correlate with the clinical utility and validity of health apps [76]. In fact, rating-scores from app stores often diverge significantly from ratings provided by health care professionals and patients [77]. Despite this, rating-scores positively impact app downloads [76]. When users choose apps, rating-scores and reviews are frequently the sole available measures [77] that influence their decision to download the app [78]. Rating-scores and reviews also offer valuable user feedback, enabling app developers to enhance their apps [79].

In our study, most HD apps on both the Apple and Google app stores lacked rating-scores and reviews. However, among those apps with rating-scores, the majority received ratings >3 stars. Reviews, when present, typically ranged in number from 1 to 1000, with very few apps accumulating more than 1000 reviews. These findings suggest that active use and user engagement with HD apps in India are confined to only a fraction of the available apps.

Length of Description

In our study, the average length of description for HD apps was around 1300 characters on the Apple App Store and around 1800 characters on the Google Play Store. The length of description in mHealth apps has been studied in previous research. In their study about psychiatric apps on Google Play Store, Pinheiro et al [74] reported that the app descriptions ranged from 75 to 3456 characters with an average of 1253 characters. Previous studies have indicated that the length of app descriptions in mHealth apps can influence their demand and install rates [71,79]. Typically, a more comprehensive and explanatory description tends to correlate positively with a higher number of installs. We discuss the correlation between the length of the description and other attributes further in the correlation analysis section below.

OS Versions and Update Years

In our analysis, we observed that most HD apps available on both the Apple App Store and Google Play Store in India did not require the latest OS versions. This is significant because apps that demand newer OS versions may become inaccessible to users with older smartphones [24].

Furthermore, we found that most of the HD apps on the Apple App Store have received updates within the last few years. It is worth noting that the timing of the latest update has been linked to the quality of mHealth apps [76]. Specifically, apps updated within the previous 6 months tend to exhibit higher quality [76]. However, it is essential to highlight that app updates can sometimes lead to compatibility issues, impacting users' ability to use these apps effectively. Such issues have been reported to hinder the use of certain mHealth apps [24].

Release Years of Apps

The distribution of apps according to their release years (Figures 3E and 3F) reveals interesting observations. While Apple App Store HD apps peaked in 2021, Google Play Store HD apps reached their peak in 2020. However, it is essential to interpret this data cautiously, as it may not accurately reflect the actual number of HD apps released in the last 2 years. Several factors contribute to this disparity. First, our study builds upon Kaggle datasets acquired in October 2021 and June 2021, which might inherently carry a bias in terms of app release years due to variations in total app store coverage and the timing of data collection. Second, newer apps introduced to these platforms since June and October 2021 may have varying degrees of linkage to apps in our initial Kaggle datasets. This linkage relies on app similarity, defined by internal algorithms of the Google Play Store and the Apple App Store, and the presence of common developers, both of which are beyond our control.

Despite these potential limitations, it is important to note that our dataset remains extensive, encompassing >200,000 mHealth apps and >2000 HD apps across both platforms. While acknowledging these considerations, our study's primary findings and insights into the landscape of mHealth apps, especially those related to HD, remain valid and valuable for researchers and stakeholders in the field.

Correlation Analysis

This analysis sheds light on various aspects of HD apps, offering insights into factors influencing their downloads, user ratings, and overall dynamics.

We found that for HD apps on Google Play Store, the number of installs positively correlated with the number of ratings and reviews. Pinheiro et al [74] noted a similar strong correlation between the number of installs and ratings in psychiatry apps on the Google Play Store. Notably, they reported no significant correlation between installs and reviews [74]. Pinheiro et al [74] also reported a strong association between installs and low app prices, as well as installs and the presence of in-app purchases. Interestingly, these associations were not found in the context of HD apps in this study.

On both Apple App Store and Google Play Store, we found that app rating-score and description length were positively correlated. This suggests that longer descriptions tend to correspond to higher app rating-scores. We found that the release year of the apps was negatively associated with the number of reviews and rating-scores on both Apple App Store and Google Play Store, suggesting that newly released apps had a lower number of reviews and lower rating-scores, and older apps tended to have slightly higher rating-scores and reviews.

XSL•FO
RenderX

In addition, our analysis revealed a positive correlation between the size and price of HD apps on the Apple App Store, contrary to findings from the study by Fu [45]. Fu studied apps from all categories in Apple App Store, not only mHealth apps, and found that larger apps tend to be cheaper. In our analysis of HD apps, we found that larger apps, in general, were more expensive. The large size of such apps could be due to additional features offered.

Clustering

We used LDA clustering techniques to identify and classify the different types of HD apps available in India. We found that about half of the apps were classified into the clinical cluster. This cluster had terms like doctor, disease, treatment, cardiac, and pain. This cluster can be used to conduct in-depth analysis and identify the numbers and characteristics of various clinical HD apps like those meant for teleconsultations, ordering medication, and health education.

Previous studies have used clustering techniques while studying apps from app stores. However, those apps were very different from our study. Fu [45] used numerical clustering methods on apps from the Apple App Store to group apps according to the number of ratings and user rating-score. Danahiswari et al [44] studied the Google Play Store apps and used k-means clustering to identify the differences in characteristics between apps that were attractive to users and those that were not. Both these studies were aimed toward app developers and making better apps.

Dehling et al [47] studied the security and privacy of mHealth apps on the Apple App Store and Google Play Store using clustering techniques. To the best of our knowledge, ours is the first study to use app descriptions to cluster and classify HD apps.

These clusters can be used in further research depending on one's research focus. For example, researchers interested in preventive and public health aspects of HD apps may incorporate apps from the lifestyle and fitness clusters, whereas those seeking more clinical apps can concentrate solely on the clinical cluster.

Further analysis of app attributes within individual clusters showed that clinical apps were smaller in size. Among the largest (ie, >100 MB app size) apps, there were a few clinical apps. However, these were outliers, and most apps from the clinical cluster were much smaller. They also had shorter descriptions. Clinical cluster apps on the Google Play Store had a lower number of installs and rating-scores, and the lowest percentage of apps with in-app purchases. More detailed descriptions can perhaps help increase the installs of these apps.

Our analysis revealed key insights into the current landscape of health and medical apps in India. Language accessibility remains a significant barrier, with most apps not supporting regional languages. Despite the high prevalence of HD, there is a notable scarcity of relevant mHealth apps. The existing HD apps are generally large, creating challenges for users with limited storage or slow internet. While most HD apps are free, paid apps (ie, especially on the Apple App Store), and in-app purchases (ie, on the Google Play Store) can hinder accessibility. User engagement (ie, from the number of ratings and reviews, and rating) is in general low, possibly due to issues with app discoverability and perceived credibility. We found that longer app descriptions tend to correlate with higher ratings, enhancing user trust. Many apps support older OS versions, reducing technological barriers. However, frequent updates, while improving quality, can lead to compatibility issues. In recent years, there has been a peak in app releases. Correlation analysis showed that the year of release for the apps was closely linked to ratings and reviews. Apps that have been around for a while had more ratings and reviews, which can add to their perceived credibility. Clustering revealed that "clinical apps," which focus on teleconsultation and health education, were prominent in number but had lower engagement. They had a lower number of installs, rating-scores, and shorter descriptions. Each of the clusters can be studied further in future research to obtain deeper insights.

Limitations

We used a comprehensive approach and a thorough methodology in this study; however, it is important to acknowledge the limitations for a balanced interpretation of the findings.

A downside to using the Kaggle datasets is that any biases or inaccuracies in the original scraping process could have extended into our study. While we made diligent efforts to compile a comprehensive dataset of mHealth and HD apps, it is important to acknowledge the absence of a definitive list encompassing all available apps on both app stores to compare against. To enhance the reliability of future research based on this dataset, we propose a validation approach wherein our dataset can be cross-referenced with a carefully selected subset of apps that are unequivocally expected to be part of such a list.

While the keyword dictionaries cover an extensive list of relevant terms, they are not exhaustive as we may have missed some terms, leading to some missed HD apps. Furthermore, we used the tokenized text from the title, summary, and description of the apps as text data for NLP analysis, and we did not appraise the quality of the texts. Our method can be used as a first step to gather all these apps that claim to address specific health issues (eg, HD), which can then be scrutinized under a quality lens using the already existing tools, such as MARS (Mobile Application Rating Scale) [41,42].

Our analysis of HD apps exclusively focused on apps with English descriptions, excluding those with descriptions in an Indian language. However, it is important to note that apps with descriptions in an Indian language represent a minor fraction, 0.5% (548/108,945) in Google apps and 1.68% (1990/118,555) in Apple apps. Our aim was to offer a comprehensive overview of HD apps accessible in India. We highlight the scarcity of apps with descriptions in Indian languages and emphasize the importance of exploring these apps in future research.

The LDA model parameters were selected based on iterative testing; however, they might not be optimal for all types of health apps. They may need to be adjusted when these methods are replicated for other mHealth apps.

XSL•FO
RenderX

In our analysis, we used simple descriptive statistics and unsupervised clustering methods. More detailed statistical techniques and supervised topic modeling could yield deeper insights into the data. However, with this study, we aimed to establish our methods to create a database for health apps and analyze data to provide an overview of the HD apps available in India. We believe that these initial findings set the stage for more in-depth research, offering a baseline for future studies to build upon.

The results and models of this study are not validated with external datasets, limiting the robustness of the conclusions. However, we aimed to lay a strong foundation for future research by providing a comprehensive and replicable methodology. The code is available on GitHub, allowing other researchers to easily validate and extend our work with additional datasets, enhancing the credibility and broader applicability in future studies.

Our Methods Versus Previous Methods

Studies on mHealth apps so far either looked at existing literature on specific apps or explored and analyzed a handful of apps by searching app stores manually. Manual scraping cannot help while mapping a vast number of apps. It only works for collecting data about a very specific health condition, and even then, it is time-taking.

Studies, such as the one by Dehling et al [47] have previously used similar scraping methods, but the regulations and designs of app stores have changed since then. Moreover, there is a limitation on the number of results one can obtain in both the app stores by searching with specific keywords. Our methods help increase the search range by using the “similar to” and “from the same developer” features on app stores and automate the process of collecting and screening metadata from a large number of apps.

Future Research

In the future, this method can be used to compile a dynamic dataset of mHealth apps available in India at any given time, allowing for regular updates and enabling trend analysis over time. Expanding the analysis on the HD apps, one could look at apps concerning the major health priorities in India, such as the top 10 causes of death. Furthermore, an analysis of the users’ reviews using NLP techniques and sentiment analysis could help to understand the users’ views and challenges while adopting mHealth apps. Our study and methodology facilitate the selection of apps with particular characteristics for further qualitative evaluation with tools, such as MARS [41,42]. This enables a systematized understanding of the app quality, instead of depending on the blackbox methods used by the app stores for manual searching.

Our methods hold particular significance in mapping the landscape of mHealth apps, particularly in low-resource settings like India, where evidence on the rapidly evolving mHealth ecosystem is scarce. This study addresses a significant gap in the academic literature by offering a relatively user-friendly approach to access and analyze valuable metadata from mHealth apps. Importantly, our method is designed to be accessible to health researchers without extensive coding expertise, serving as an effective means to overcome this potential barrier. Researchers can readily adapt and replicate our approach to explore mHealth apps in various low-resource settings, leveraging geolocation and condition-specific keyword dictionaries to analyze apps tailored to address diverse health issues with targeted research inquiries.

Conclusions

Our study introduced a robust method for collecting and analyzing metadata from a wide array of mHealth apps available in India through the Apple App Store and the Google Play Store. We revealed the limited representation of India’s linguistic diversity within the health and medical app landscape, evident from the negligible presence of Indian language apps. Despite the high prevalence and number of deaths caused by HD in India, we observed a disproportionate scarcity of mHealth apps dedicated to this crucial health issue, along with a lower level of user engagement, as indicated by reviews and app ratings. While most HD apps are financially accessible, uptake remains a challenge. Further research should focus on app quality assessment and factors influencing user adoption.

Acknowledgments

The authors acknowledge support from the Open Access Publishing Fund of Karl Landsteiner University of Health Sciences, Krems, Austria.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Language analysis on the Apple App Store and Google Play Store apps.
[PNG File , 250 KB - jmir_v27i1e53823_app1.png]

Multimedia Appendix 2
k-means clustering in heart disease apps on Apple App Store.
[PNG File , 44 KB - jmir_v27i1e53823_app2.png]

XSL•FO
RenderX

Multimedia Appendix 3
Language distribution in mobile health apps from the 2 app stores.
[PNG File , 281 KB - jmir_v27i1e53823_app3.png]

Multimedia Appendix 4
Average rating-scores, reviews, and installs by language for mobile health apps on the Google Play Store.
[PNG File , 185 KB - jmir_v27i1e53823_app4.png]

Multimedia Appendix 5
Keyword monograms and bigrams found in (selected) apps related to heart diseases.
[PNG File , 281 KB - jmir_v27i1e53823_app5.png]

Multimedia Appendix 6
Exploratory correlation analysis—Apple App Store app attributes.
[PNG File , 714 KB - jmir_v27i1e53823_app6.png]

Multimedia Appendix 7
Exploratory correlation analysis—Google Play Store app attributes.
[PNG File , 918 KB - jmir_v27i1e53823_app7.png]

Multimedia Appendix 8
Clustering and topic modeling via latent Dirichlet allocation for heart disease apps. PC: principal component.
[PNG File , 747 KB - jmir_v27i1e53823_app8.png]

Multimedia Appendix 9
Distribution of medical category apps in heart disease app clusters.
[PNG File , 98 KB - jmir_v27i1e53823_app9.png]

References

1. Lapão LV, Dussault G. The contribution of eHealth and mHealth to improving the performance of the health workforce: a review. Policy Pract 2017;3(3):464-471 [FREE Full text]
2. mHealth: use of appropriate digital technologies for public health. World Health Organization. 2018. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_20-en.pdf [accessed 2024-04-29]
3. Chauhan P, Jadhav V, Gupta N. COVID-19 pandemic and consumer adoption of digital healthcare in India: a qualitative study. Acad Mark Stud J 2021;25(4):1-6 [FREE Full text]
4. Madanian S, Parry DT, Airehrour D, Cherrington M. mHealth and big-data integration: promises for healthcare system in India. BMJ Health Care Inform 2019 Sep;26(1):55 [FREE Full text] [doi: 10.1136/bmjhci-2019-100071] [Medline: 31488497]
5. Majumdar A, Kar SS, Palanivel C, Misra P. mHealth in the prevention and control of non-communicable diseases in India: current possibilities and the way forward. J Clin Diagn Res 2015 Feb;9(2):LE06-LE10 [FREE Full text] [doi: 10.7860/JCDR/2015/11555.5573] [Medline: 25859473]
6. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022 Apr 05;43(1):525-539 [FREE Full text] [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
7. Pai RR, Alathur S. Bibliometric analysis and methodological review of mobile health services and applications in India. Int J Med Inform 2021 Jan;145:104330. [doi: 10.1016/j.ijmedinf.2020.104330] [Medline: 33248334]
8. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
9. Rathore M. Share of school children living in a family with a smartphone during COVID-19 pandemic in India as of August 2021, by region. Statista. URL: https://www.statista.com/statistics/1276332/india-share-of-school-children-who-live-in-a-family-with-smartphone-during-covid-19-by-region/ [accessed 2024-04-29]
10. Shangliao S. Smartphone users in India 2010-2040. Statista. URL: https://tinyurl.com/c4j3f6yv [accessed 2024-04-29]
11. Holzman SB, Atre S, Sahasrabudhe T, Ambike S, Jagtap D, Sayyad Y, et al. Use of smartphone-based video directly observed therapy (vDOT) in tuberculosis care: single-arm, prospective feasibility study. JMIR Form Res 2019 Aug 27;3(3):e13411 [FREE Full text] [doi: 10.2196/13411] [Medline: 31456581]
12. Jimenez G, Lum E, Car J. Examining diabetes management apps recommended from a Google search: content analysis. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11848 [FREE Full text] [doi: 10.2196/11848] [Medline: 30303485]

XSL•FO
RenderX

13. Kumar R, Das A. The potential of mHealth as a game changer for the management of sickle cell disease in India. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e25496 [FREE Full text] [doi: 10.2196/25496] [Medline: 33847598]
14. Selvaraj SN, Sriram A. The quality of Indian obesity-related mHealth apps: PRECEDE-PROCEED model-based content analysis. JMIR Mhealth Uhealth 2022 May 11;10(5):e15719 [FREE Full text] [doi: 10.2196/15719] [Medline: 35544318]
15. Praveen D, Patel A, Raghu A, Clifford GD, Maulik PK, Mohammad Abdul A, et al. SMARTHealth India: development and field evaluation of a mobile clinical decision support system for cardiovascular diseases in rural India. JMIR Mhealth Uhealth 2014 Dec 08;2(4):e54 [FREE Full text] [doi: 10.2196/mhealth.3568] [Medline: 25487047]
16. Bassi A, Arfin S, John O, Jha V. An overview of mobile applications (apps) to support the coronavirus disease 2019 response in India. Indian J Med Res 2020 May;151(5):468-473. [doi: 10.4103/ijmr.IJMR_1200_20] [Medline: 32474557]
17. Jaiswal S, Kaur N, Bhalla M, Sharma VS, Gupta R, Chaudhary S. A survey on knowledge and awareness among Indian citizens toward application to fight against COVID-19: a cross-sectional study. J Adv Clin Res Insights 2020;7(5):67-72 [FREE Full text] [doi: 10.15713/ins.jcri.308]
18. Kaur M, Kaur H, Rathi S, Ashwitha M, Joanna J, Reddy S, et al. Apps on Google Play Store to assist in self-management of hypertension in Indian context: features analysis study. Mhealth 2022;8:14 [FREE Full text] [doi: 10.21037/mhealth-21-25] [Medline: 35449503]
19. Kodali PB, Das S. Acceptance of mHealth technologies among auxiliary; nurse midwives in Andhra Pradesh, India: a mixed method study. Med Sci 2021;25(111):1060 [FREE Full text]
20. Bommakanti KK, Smith LL, Liu L, Do D, Cuevas-Mota J, Collins K, et al. Requiring smartphone ownership for mHealth interventions: who could be left out? BMC Public Health 2020 Jan 20;20(1):81 [FREE Full text] [doi: 10.1186/s12889-019-7892-9] [Medline: 31959145]
21. Sinha Deb K, Tuli A, Sood M, Chadda R, Verma R, Kumar S, et al. Is India ready for mental health apps (MHApps)? A quantitative-qualitative exploration of caregivers' perspective on smartphone-based solutions for managing severe mental illnesses in low resource settings. PLoS One 2018 Sep 19;13(9):e0203353 [FREE Full text] [doi: 10.1371/journal.pone.0203353] [Medline: 30231056]
22. Singh S, Sharma P, Ghimire P, Shrestha R, Gnanavel S. Assessment of app store description and privacy policy to explore ethical and safety concerns associated with the use of mental health apps for depression. Indian J Psychol Med 2023 Mar 14;45(2):173-178 [FREE Full text] [doi: 10.1177/02537176221142046] [Medline: 36925500]
23. Huang Z, Lum E, Jimenez G, Semwal M, Sloot P, Car J. Medication management support in diabetes: a systematic assessment of diabetes self-management apps. BMC Med 2019 Jul 17;17(1):127 [FREE Full text] [doi: 10.1186/s12916-019-1362-1] [Medline: 31311573]
24. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196 [FREE Full text] [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]
25. Garnett A, Northwood M, Ting J, Sangrar R. mHealth interventions to support caregivers of older adults: equity-focused systematic review. JMIR Aging 2022 Jul 08;5(3):e33085 [FREE Full text] [doi: 10.2196/33085] [Medline: 35616514]
26. Bassi A, John O, Praveen D, Maulik PK, Panda R, Jha V. Current status and future directions of mHealth interventions for health system strengthening in India: systematic review. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11440 [FREE Full text] [doi: 10.2196/11440] [Medline: 30368435]
27. Coulon SM, Monroe CM, West DS. A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
28. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]
29. Griswold D, Rubiano AM. Role of mHealth applications for emergency medical system activation in reducing mortality in low-income and middle-income countries: a systematic review protocol. BMJ Open 2022 Feb 23;12(2):e051792 [FREE Full text] [doi: 10.1136/bmjopen-2021-051792] [Medline: 35197336]
30. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: content analysis and review. JMIR Mhealth Uhealth 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: 10.2196/19796] [Medline: 32609622]
31. Jahnel T, Dassow HH, Gerhardus A, Schüz B. The digital rainbow: digital determinants of health inequities. Digit Health 2022 Oct 02;8:20552076221129093 [FREE Full text] [doi: 10.1177/20552076221129093] [Medline: 36204706]
32. DeWitt A, Kientz J, Coker TR, Liljenquist K. mHealth technology design and evaluation for early childhood health promotion: systematic literature review. JMIR Pediatr Parent 2022 Oct 06;5(4):e37718 [FREE Full text] [doi: 10.2196/37718] [Medline: 36201391]
33. Daley BJ, Ni'Man M, Neves MR, Bobby Huda MS, Marsh W, Fenton NE, et al. mHealth apps for gestational diabetes mellitus that provide clinical decision support or artificial intelligence: a scoping review. Diabet Med 2022 Jan 16;39(1):e14735. [doi: 10.1111/dme.14735] [Medline: 34726798]
34. Balsari S, Fortenko A, Blaya JA, Gropper A, Jayaram M, Matthan R, et al. Reimagining health data exchange: an application programming interface-enabled roadmap for India. J Med Internet Res 2018 Jul 13;20(7):e10725 [FREE Full text] [doi: 10.2196/10725] [Medline: 30006325]

XSL•FO
RenderX

35. Stach M, Kraft R, Probst T, Messner EM, Terhorst Y, Baumeister H, et al. Mobile health app database - a repository for quality ratings of mHealth apps. In: Proceedings of the 33rd International Symposium on Computer-Based Medical Systems. 2020 Presented at: CBMS '20; July 28-30, 2020; Rochester, MN p. 427-432 URL: https://ieeexplore.ieee.org/document/9183243 [doi: 10.1109/cbms49503.2020.00087]
36. Hamrioui S, González JH, Castillo G, de la Torre-Diez I, Lopez-Coronado M. Development and QoE evaluation of an iOS mHealth app for self-controlling and education of patients with heart diseases. In: Proceedings of the 13th Iberian Conference on Information Systems and Technologies. 2018 Presented at: CISTI '18; June 13-16, 2018; Caceres, Spain p. 1-4 URL: https://ieeexplore.ieee.org/document/8398633 [doi: 10.23919/cisti.2018.8398633]
37. Braghin C, Cimato S, Della Libera A. Are mHealth apps secure? A case study. In: Proceedings of the 42nd Annual Computer Software and Applications Conference. 2018 Presented at: COMPSAC '18; July 23-27, 2018; Tokyo, Japan p. 335-340 URL: https://ieeexplore.ieee.org/document/8377881 [doi: 10.1109/compsac.2018.10253]
38. Sampat BH, Prabhakar B. Privacy risks and security threats in mHealth apps. J Int Technol Inf Manag 2017 Dec 01;26(4):126-153. [doi: 10.58729/1941-6679.1353]
39. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
40. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
41. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
42. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
43. Siddiqui NR, Hodges S, Sharif MO. Availability of orthodontic smartphone apps. J Orthod 2019 Sep 06;46(3):235-241. [doi: 10.1177/1465312519851183] [Medline: 31169046]
44. Danahiswari HC, Nurpratama YF, Kartika DS. Application cluster analysis on the google play store using the K-Means method. Int J Comp Net Sec Inf Syst 2022;4(1):5 [FREE Full text] [doi: 10.33005/ijconsist.v4i1.74]
45. Fu B. Characteristics classification of mobile apps on apple store using clustering. J Data Anal Inf Process 2020;08(02):69-85. [doi: 10.4236/jdaip.2020.82004]
46. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
47. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
48. Tsinaraki C, Mitton I, Minghini M, Micheli M, Kotsev A, Hernandez Quiros L, et al. Analysing mobile apps that emerged to fight the COVID-19 crisis. European Commission. 2020. URL: https://joint-research-centre.ec.europa.eu/system/files/2023-01/03.covid-19_technical_report_final.pdf [accessed 2024-04-29]
49. GBD 2019 Diseases Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
50. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009 Oct;42(5):760-772 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.007] [Medline: 19683066]
51. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 2020 Feb;145(2):463-469 [FREE Full text] [doi: 10.1016/j.jaci.2019.12.897] [Medline: 31883846]
52. Kumar S, Singh M. A novel clustering technique for efficient clustering of big data in Hadoop Ecosystem. Big Data Min Anal 2019 Dec;2(4):240-247. [doi: 10.26599/bdma.2018.9020037]
53. Chen X, Xie H, Wang FL, Liu Z, Xu J, Hao T. A bibliometric analysis of natural language processing in medical research. BMC Med Inform Decis Mak 2018 Mar 22;18(Suppl 1):14 [FREE Full text] [doi: 10.1186/s12911-018-0594-x] [Medline: 29589569]
54. Iroju OG, Olaleke JO. A systematic review of natural language processing in healthcare. Int J Inf Technol Comput Sci 2015 Jul 08;7(8):44-50. [doi: 10.5815/ijitcs.2015.08.07]
55. Prakash G, Koshy J. Google Play store apps. kaggle. URL: https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps [accessed 2022-12-04]
56. Prakash G. Apple AppStore apps. kaggle. URL: https://www.kaggle.com/datasets/gauthamp10/apple-appstore-apps [accessed 2023-07-04]
57. Olano F. google-play-scraper (v9.1.1). npm Inc. URL: https://www.npmjs.com/package/google-play-scraper/v/9.1.1 [accessed 2024-04-29]
58. Olano F. app-store-scraper (v0.17.0). npm Inc. URL: https://www.npmjs.com/package/app-store-scraper/v/0.17.0 [accessed 2024-04-29]

XSL•FO
RenderX

59. Leeson W, Resnick A, Alexander D, Rovers J. Natural language processing (NLP) in qualitative public health research: a proof of concept study. Int J Qual Methods 2019 Nov 13;18:160940691988702. [doi: 10.1177/1609406919887021]
60. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544-551 [FREE Full text] [doi: 10.1136/amiajnl-2011-000464] [Medline: 21846786]
61. Myocardial ischemia. National Library of Medicine. 2020. URL: https://www.ncbi.nlm.nih.gov/mesh/?term=ischemic+heart+diseases [accessed 2024-04-29]
62. pandas-dev/pandas: pandas. zenodo. URL: https://doi.org/10.5281/zenodo.3509134 [accessed 2024-04-29]
63. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007 May;9(3):90-95. [doi: 10.1109/mcse.2007.55]
64. Bird S, Edward L, Ewan K. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. New York, NY: O'Reilly Media; 2009.
65. Lopez F. fedelopez77/langdetect. GitHub. URL: https://github.com/fedelopez77/langdetect [accessed 2024-06-19]
66. Mueller AC. Wordcloud. GitHub. URL: https://github.com/amueller/wordcloud [accessed 2024-06-19]
67. Mabey B. bmabey/pyLDAvis. GitHub. URL: https://github.com/bmabey/pyLDAvis [accessed 2024-06-19]
68. Prizak R, Dubbala K. General · KeerthiDubbala/mHealth-apps-in-India. GitHub. URL: https://github.com/KeerthiDubbala/mHealth-apps-in-India [accessed 2023-10-19]
69. Mechael P, Batavia H, Kaonga N, Searle S, Kwan A, Goldberger A, et al. Barriers and gaps affecting mHealth in low- and middle-income countries: policy white paper. Center for Global Health and Economic Development, The Earth Institute at Columbia University. University of California. 2010. URL: https://tb.ucsf.edu/publications/barriers-and-gaps-affecting-mhealth-low-and-middle-income-countries-policy-white-paper [accessed 2024-04-29]
70. Bassi A. mHealth interventions for health system strengthening in India. George Institute for Global Health. 2016. URL: https://www.georgeinstitute.org.in/our-impact/policy-and-recommendations/mhealth-interventions-for-health-system-strengthening-in [accessed 2024-04-29]
71. Pereira-Azevedo N, Osório L, Cavadas V, Fraga A, Carrasquinho E, Cardoso de Oliveira E, et al. Expert involvement predicts mHealth app downloads: multivariate regression analysis of urology apps. JMIR Mhealth Uhealth 2016 Jul 15;4(3):e86 [FREE Full text] [doi: 10.2196/mhealth.5738] [Medline: 27421338]
72. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Global Health Action 2017 Aug 25;10(sup3):1327686. [doi: 10.1080/16549716.2017.1327686]
73. Sharma S, Barnett KG, Maypole JJ, Mishuris RG. Evaluation of mHealth Apps for Diverse, Low-Income Patient Populations: Framework Development and Application Study. JMIR Form Res 2022 Feb 11;6(2):e29922-e29996 [FREE Full text] [doi: 10.2196/29922] [Medline: 35147502]
74. Pinheiro M, Serra M, Pereira-Azevedo N. Predictors of the number of installs in psychiatry smartphone apps: systematic search on app stores and content analysis. JMIR Ment Health 2019 Nov 19;6(11):e15064 [FREE Full text] [doi: 10.2196/15064] [Medline: 31742560]
75. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 05;43(1):117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
76. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela JP, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
77. Hudson G, Negbenose E, Neary M, Jansli SM, Schueller SM, Wykes T, et al. Comparing professional and consumer ratings of mental health apps: mixed methods study. JMIR Form Res 2022 Sep 23;6(9):e39813 [FREE Full text] [doi: 10.2196/39813] [Medline: 36149733]
78. How ratings and reviews affect consumers decision to download apps. Business of Apps. URL: https://www.businessofapps.com/insights/ratings-reviews-affect-consumer-decision-download-apps/ [accessed 2023-09-20]
79. Galetsi P, Katsaliaki K, Kumar S, Ferguson M. What affects consumer behavior in mobile health professional diagnosis applications. Decis Sci 2021 Oct 13;54(3):315-333. [doi: 10.1111/deci.12550]

Abbreviations

HD: heart disease
LDA: latent Dirichlet allocation
mHealth: mobile health
NLP: natural language processing
OS: operating systems

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.10.23; peer-reviewed by R Pryss, MVGB Serra; comments to author 28.04.24; revised version received 22.06.24; accepted 12.12.24; published 10.03.25.

Please cite as:
Dubbala K, Prizak R, Metzler I, Rubeis G
Exploring Heart Disease–Related mHealth Apps in India: Systematic Search in App Stores and Metadata Analysis
J Med Internet Res 2025;27:e53823
URL: https://www.jmir.org/2025/1/e53823
doi:10.2196/53823
PMID:40063078

©Keerthi Dubbala, Roshan Prizak, Ingrid Metzler, Giovanni Rubeis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Targeted Behavior Change Communication Using a Mobile Health Platform to Increase Uptake of Long-Lasting Insecticidal Nets Among Pregnant Women in Tanzania: Hati Salama "Secure Voucher" Study Cluster Randomized Controlled Trial

Trinity Vey[1], MSc; Eleonora Kinnicutt[2], MPH; Andrew G Day[1,3], MSc; Nicola West[1,4], MSc; Jessica Sleeth[1], MPH; Kenneth Bernard Nchimbi[5], BSc, PGDip; Karen Yeates[1,3,4], MD, MPH

[1]Department of Medicine, Queen's University, Kingston, ON, Canada

[2]Management Sciences for Health, Arlington, VA, United States

[3]Kingston Health Sciences Centre, Kingston, ON, Canada

[4]Pamoja Tunaweza Women's Centre, Moshi, United Republic of Tanzania

[5]Tanzania Education Network/Mtandao wa Elimu Tanzania, Dar es Salaam, United Republic of Tanzania

Corresponding Author:
Trinity Vey, MSc
Department of Medicine
Queen's University
99 University Avenue
Kingston, ON, K7L 3N6
Canada
Phone: 1 613 533 2000
Email: 16tv7@queensu.ca

Abstract

Background: Malaria remains a significant cause of maternal and neonate morbidity and mortality in sub-Saharan Africa. Long-lasting insecticidal nets (LLINs) represent an important component of malaria prevention and can decrease the adverse health outcomes associated with malaria infection during pregnancy. Voucher programs have been successfully implemented for a variety of initiatives across sub-Saharan Africa, including the distribution of subsidized LLINs in Tanzania. However, mobile messaging for behavior change communication (BCC), in combination with an e-voucher program, has not been explored for malaria prevention.

Objective: This study aimed to assess the efficacy of mobile messaging in increasing the redemption of e-vouchers for LLINs for pregnant women and adolescents in Tanzania.

Methods: This study was a blinded, 2-arm, cluster randomized controlled trial implemented in 100 antenatal health facilities in Tanzania (both urban and rural settings), with 50 clusters in both intervention and control groups. Clusters were antenatal clinics with e-voucher capabilities, with randomization stratified such that 25 urban and 25 rural clinics were randomized to each arm. Participants were pregnant females aged 13 years or older. Participants in both intervention and control groups were issued e-vouchers on their mobile phones that could be redeemed for LLINs at registered retailers within a 14-day redemption period. Participants in the intervention group received targeted BCC messages about the importance of malaria prevention and LLIN use during pregnancy, while participants in the control group did not receive BCC messages. Analyses were by intention to treat. The primary outcome was the redemption rate of e-vouchers for LLINs from retailers. Outcome measures pertain to clinic sites and individual participant-level data.

Results: The study enrolled 5449 participants; the analysis included 2708 participants in the intervention arm and 2740 participants in the control arm (49 clusters in each group analyzed). There was no significant difference in the raw redemption rate of e-vouchers between pregnant participants in the intervention group (70%) and the control group (67%). Younger participants were less likely to redeem e-vouchers.

Conclusions: The use of a BCC mobile messaging intervention did not result in a significant increase in LLIN uptake for pregnant individuals. However, the study shows that e-voucher distribution through nurses in antenatal clinics in partnership with

XSL•FO
RenderX

local retailers is feasible on a large scale. Consideration of women and adolescents who are low-income and live in rural areas is needed for future interventions leveraging e-vouchers or mHealth technology in low-resource settings.

Trial Registration: ClinicalTrials.gov NCT02561624; https://clinicaltrials.gov/ct2/show/NCT02561624

(J Med Internet Res 2025;27:e51524) doi:10.2196/51524

KEYWORDS

malaria prevention; pregnant; mHealth; mobile health; short message service; behaviour change communication; long-lasting insecticidal nets; protozoan infections; parasitic diseases; vector borne diseases; randomized controlled trial; morbidity; mortality; intervention

Introduction

Malaria, a parasitic infectious disease spread by mosquitoes, was responsible for an estimated 56 million cases and 132,500 deaths in the East and Southern Africa subregion in 2020 [1]. The United Republic of Tanzania accounted for 12.8% of these total cases [1]. Pregnant women and adolescents are at heightened risk of morbidity and mortality by malaria, with 22% of pregnant women exposed to malaria in East and Southern Africa in 2020 [1]. Malaria infection during pregnancy can result in miscarriage, premature delivery, low birth weight, or death in neonates, as well as maternal anemia and maternal death [2]. Insecticide-treated nets (ITN) are recognized as a mainstay of malaria prevention, whereby trials of pregnant women in sub-Saharan Africa using ITNs during pregnancy compared with no nets reported decreased malarial infection, fewer neonate deaths before delivery, and fewer low birthweight babies [3]. Long-lasting insecticidal nets (LLINs) are a subset of ITN that retain their effective biological activity for approximately 3 years of recommended use without retreatment [4]. While malaria chemoprevention is being more widely implemented in endemic areas, especially for children and pregnant women, this is not a stand-alone strategy; LLINs remain an important component of the malaria prevention toolbox, in combination with medication [5,6].

The Tanzania National Voucher Scheme (TNVS) was a public-private partnership active between 2004 and 2014 that provided pregnant women and children with vouchers at antenatal clinics to be redeemed for ITNs and LLINs at a subsidized price [7]. The TNVS, widely known as Hati Punguzo (Swahili for “discount voucher”), contributed to an increased usage of LLINs among pregnant women, distributing 1.2 to 1.8 million LLINs per year [7]. Initially using paper vouchers, in 2011, the TNVS transitioned to electronic “e-vouchers,” allowing for enhanced tracking, real-time redemption data, and improved workflow efficiency [7]. Voucher programs have been successfully implemented in sub-Saharan Africa for a variety of initiatives, including reproductive and sexual health services, as well as agricultural interventions [8-10].

The transition of the TNVS to e-vouchers presented an opportunity for research assessing the impact of mobile health (mHealth) technology. With 495 million people subscribed to mobile phone services in sub-Saharan Africa in 2020, accounting for 46% of the region’s population, mobile phones provide a platform for widespread health interventions in low-resource settings [11].

Behavior communication change (BCC) strategies, which use targeted messaging to promote and maintain healthy behaviors, have been implemented in Africa for malaria prevention and control. In Benin, BCC messages delivered by social mobilization, radio, and flyers increased LLIN use in pregnant women [12]. Similarly, in Uganda, BCC messages spread by interpersonal communication, community mobilization, radio, and flyers increased the usage of LLIN in pregnant women [13]. However, the delivery of BCC messages through mHealth technology to increase malaria prevention efforts remains an under-researched area.

Data from 2006-2013 suggests that only 47% of women who attended an antenatal clinic accessed a net through the TNVS [7]. While the Under Five Catch Up Campaign, which distributed over 9 million free LLINs to children under 5 between 2008-2010, as well as the Universal Catch Up Campaign, which distributed over 17 million LLINs to households in 2010-2011, may have caused bed-net saturation, it remains unclear why many women were not redeeming vouchers to access LLINs before these mass campaigns [7,14]. Furthermore, LLINs typically only last up to 3 years, meaning that the protection against malaria afforded by these mass LLIN distribution campaigns is not permanent.

Despite pregnant women being at high risk of malaria morbidity and mortality, the ability of BCC messages on a mHealth platform to increase LLIN uptake for pregnant women remains uninvestigated. The Hati Salama (HASA) (Swahili for “secure voucher”) cluster randomized controlled trial aimed to test the hypothesis that BCC mobile messages targeted at pregnant women would increase redemption of e-vouchers for LLINs in regions of Tanzania with high malaria prevalence and a history of low LLIN uptake; the cluster design allowed for assessment in a real-life setting, to evaluate redemption of e-vouchers at antenatal clinic clusters. This study also aimed to determine perceived barriers to e-voucher redemption.

Methods

Study Design and Participants

The HASA study was a cluster randomized controlled, blinded 2-arm study implemented in 100 antenatal health facilities across Tanzania, with 50 clusters randomly allocated to the intervention group and 50 clusters to the control group, as shown in Figure 1. The study was registered on ClinicalTrials.gov (NCT02561624). The clinics selected for participation in the study were drawn from a pool of former TNVS health facilities. Study clusters were defined as clinics from both urban and rural

XSL•FO
RenderX

settings with e-voucher capabilities, with an issuance of more than 175 e-vouchers per quarter but less than 70% voucher redemption of LLINs during the TNVS program.

In Figure 1, the regions highlighted in dark grey are regions of Tanzania included in the HASA study. Regions highlighted in light grey are regions excluded from the HASA study.

Participants included in the study were pregnant women and adolescents aged 13 years and older attending one of the selected antenatal health facilities. To be eligible, participants were either literate or living with someone who could assist them with reading SMS text messages. While lack of phone ownership was not a reason for exclusion, participants needed to at least have access to a proxy phone, which was recommended to be the phone of a family member, throughout the entirety of the trial.

Eligible participants who consented to participate were enrolled by trained nurses. Consent was obtained by each participant verbally and in written form by the nurses, and a signature or fingerprint of all participants was obtained. We had ethical approval to enroll participants aged 13 years and older. While most publications discuss teenage pregnancy in Tanzania for females aged 15-19, there are a number of reports that highlight teenage pregnancy for primary school-aged adolescents under and up to the age of 15 [15]. The minimum enrollment age of 13 was used as local stakeholders informed the research team of the presence of adolescents under the age of 15 attending antenatal clinics. The TNVS, the main bed net program in Tanzania, did not have a specific age limit for pregnant women, but data from that participant group informed our decision to use a low minimum enrollment age of 13 to ensure more pregnant individuals were provided the opportunity to receive an e-voucher and mHealth messaging.

Participants were ensured of the confidentiality and anonymity of the trial with regard to the information collected from them. Ethical approval was obtained from the National Institute of Medical Research Tanzania and Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board.

Figure 1. Map of the regions of Tanzania included in the Hati Salama (HASA) study.

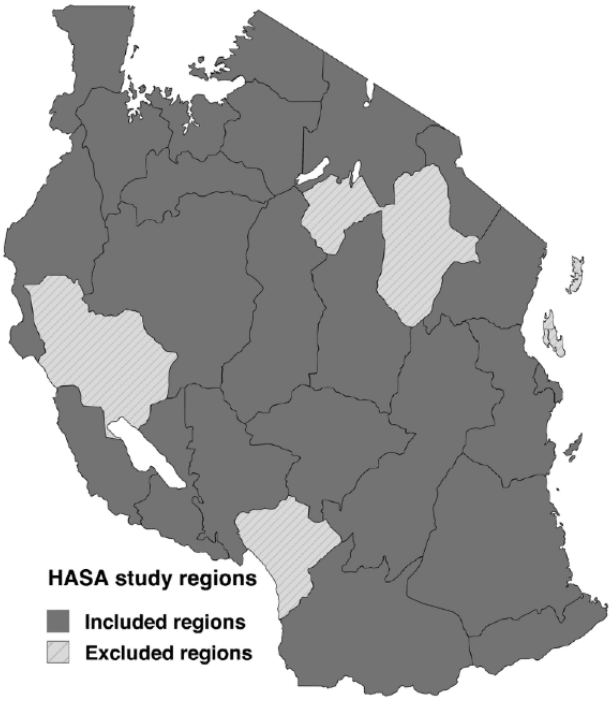

Randomization and Masking

The study enrolled 50 urban and 50 rural clinics with redemption rates below 70% in the previous year. Each site was asked to enroll at least 50 participants. Randomization was stratified such that 25 urban and 25 rural clinics were randomized to each arm. The statistician prepared the randomization list before the start of the trial using clinic codes of the 50 urban and 50 rural participating sites without knowing which clinic the codes referred to. Randomization was performed by randomly sorting 25 intervention and 25 control assignments within each of the 50 rural and 50 urban clinics by a computer-generated sequence. All health care workers and relevant stakeholders, including nurses, were blinded to their group assignment, as were participants and investigators, although participants would presumably become aware of their assignment after receiving their first BCC SMS message. Intervention group nurses and control group nurses were trained separately, and the control and intervention clinics were located large distances from each other to minimize contamination.

Procedures and Interventions

Participants in both the intervention and control groups were issued LLIN e-vouchers on their mobile phones by nurses upon enrollment into the study. E-vouchers could be redeemed for a high-quality LLIN at registered retailers for a discounted price; while the original cost of LLINs was approximately US $8, participants with an e-voucher were only required to pay a co-payment of 500 Tanzanian shillings (at the time of the study, equivalent to approximately US $0.25). All women attending

XSL•FO
RenderX

antenatal clinics enrolled in the study were offered an e-voucher through the HASA program, regardless of whether they provided informed consent to participate in the trial and receive BCC messages. Women who accepted e-vouchers but did not participate in the study were not included in the analyses, but all women enrolled in the study were included.

The study undertook a phased approach with rolling enrolment during the period from July to December 2015, whereby each site had 12 weeks to complete recruitment, and each individual participant had 14 days for e-voucher redemption, starting from when they were enrolled. Participants attending one of the 50 clinics allocated to the intervention group received targeted BCC messages on their registered mobile phones at programmed intervals during the 14-day period in which the e-voucher was active. Messages contained health promotion and sensitization information on the importance of malaria prevention and LLIN use during pregnancy, in addition to reminders to redeem the e-voucher for an LLIN at a registered retailer (Textbox 1). Participants attending one of the 50 clinics randomized into the control group did not receive any targeted BCC messages on their phones during the 14 days following e-voucher issuance. As the allocation of clinics to intervention and control groups was random and blinded, all participants were told by study nurses that they may or may not be receiving SMS messages during the 14 days following issuance. Each participant also had a case report form completed by a nurse, collecting information including the participant's age, marital status, number of pregnancies, number of children, previous LLIN ownership, and education level.

Following voucher expiration (14-day period), all participants who did not redeem their e-vouchers in both the intervention and control groups received a follow-up phone survey asking them to indicate the reason why they did not redeem the voucher for an LLIN. All participants were analyzed through intention to treat in the arm their cluster was randomized to, regardless of whether they redeemed their e-voucher or not.

Textbox 1. Content of SMS behavior communication change (BCC) messages sent to participants (translated from Swahili to English).

- Insecticide-treated bed nets are proven to prevent malaria when used correctly when sleeping in your home.
- Insecticide-treated bed nets should always be used when you or your children sleep.
- Malaria can make young children very sick and can cause death.
- Malaria is a very serious illness among pregnant women. It can have dangerous consequences for the woman and her unborn child, including death.
- Your small children should always sleep under a bed net, and bed nets should be replaced every 5-7 years or as soon as they are in poor condition.
- You have recently received a voucher that can be redeemed for a bed net, but you have not yet redeemed this voucher.
- The net voucher you received at the clinic can be redeemed at any participating shop (Duka or Duka la Dawa) that provides bed nets for Hati Punguzo.
- This is a reminder that you received a bed net voucher 1 week ago, and you have not yet redeemed it to receive your bed net.
- This is a reminder that you received a bed net voucher 10 days ago, and you have not yet redeemed it to receive your bed net.

Outcomes

The prespecified primary outcome was the redemption rate of e-vouchers for LLINs from retailers in regions with high malaria prevalence and a history of low uptake. Outcomes were measured using participant-level redemption data available through the electronic platform. The data displayed how many participants (with appropriate code and phone numbers) redeemed their e-vouchers for an LLIN within the 14-day period from each clinic and the exact times the transactions took place; outcome measures pertain to cluster sites and individual participants. A subsequent manuscript will examine perceived causes and barriers to nonredemption by a structured phone survey of participants who did not redeem their e-vouchers within the allotted 14-day period [16]. In addition, with regard to longer-term outcomes, the trial serves as a pilot study to test the overall effectiveness, security, and sustainability of a mobile platform design and its potential for delivering other private-sector health solutions in the region.

Statistical Analysis

When calculating the sample size, a baseline voucher redemption rate of 50% was estimated, with a 10% increase to 60% considered minimally clinically important. Furthermore, it was assumed that the intracluster correlation would be 0.1 and that each site would enroll 50 participants. Under these assumptions, 47 sites per arm would provide over 80% power at a 2-sided α=.05. To allow for potential loss of sites or underrecruitment at some sites, 50 sites per arm were enrolled. This calculation did not account for our analysis adjusting for reach (urban vs rural) and baseline site redemption rates, which could increase power slightly.

The redemption rates were compared between arms using a generalized estimating equation with a compound symmetric (ie, exchangeable) working correlation structure. A binomial distribution and an identity link were used so that adjusted risk differences could be estimated. However, it was confirmed that all *P* values remained nearly identical if the logit link was used. Adjusted risk differences were based on the estimate of an intervention indicator variable. The model also included the site level baseline (baseline referring to the previous year) redemption rate as a linear covariate and the site reach (urban vs rural) indicator variable. *P* values and confidence intervals were based on the Wald method using empirical standard errors. Since the primary analysis only adjusted for site-level factors, nearly identical results were obtained in a sensitivity analysis performed at the site level using a regression model with arm,

XSL•FO
RenderX

baseline redemption rate, and reach as independent variables and the observed site redemption rates as the dependent variable. Sensitivity analyses were also performed, excluding one outlier site and including only participants who were known to have had their own mobile phones.

The ANOVA estimator was used to estimate the intracluster correlation coefficient of e-voucher redemption separately by the arm with 95% CIs estimated by the nonparametric percentile bootstrap based on 1000 bootstraps for each arm. Our bootstrap selected entire clinics rather than individual participants [17].

The research team attempted to identify predictors of e-voucher redemption based on our preselected list of site and participant characteristics. First, the redemption rate was compared by characteristic. Subsequently, the aforementioned generalized estimating equation approach was used to identify significant predictors of e-voucher redemption. A multivariable model was selected, including predictors of e-voucher redemption based on a backward selection with an exit criterion of $P<.15$.

All analysis was performed using SAS (version 9.4; SAS Institute).

Role of the Funding Source

This Project is supported by Grand Challenges Canada. Grand Challenges Canada is funded by the Government of Canada and is dedicated to supporting Bold Ideas with Big Impact.

Ethical Considerations

This study was conducted in accordance with ethical standards for human subject research (Declaration of Helsinki). Before initiation, the study protocol and study forms were reviewed and given approval by the National Institute for Medical Research in Tanzania approval (certificate number NIMR/HQ/R.8a/VOL./IX/1750) and Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board institutional review board in Canada (approval number 6013000).

Participants were informed that their participation and responses would be completely confidential and no identifying information about them would be recorded and stored. Each participant was given a copy of the informed consent document. Consent was allowed in both verbal and written formats after the participant was given enough time to review the document and ask questions or discuss any concerns in a private setting with a research team member. Participants were informed throughout the consent process and the duration of the study that participation was voluntary, they may withdraw from the study at any time without consequence, and that they would still receive an e-voucher for redeeming a long-lasting insecticidal net (LLIN) even if they decided not to participate and withdrew. No compensation was given for participation in this study.

Results

The trial took place between July and December 2015. A total of 25 rural and 25 urban clinics were randomized to each study arm. During the study timeline, each clinic had a 12-week period to complete the recruitment of at least 50 participants. Each participant had 14 days for e-voucher redemption starting from when they were enrolled. One rural clinic in the control arm and 1 rural clinic in the intervention arm dropped out of the study for administrative reasons after randomization but before enrolling any participants. This left a total of 49 clinics in each arm, as shown in Figure 2. In total, 5449 participants from the randomized sites consented to participate in the trial. One of the consented participants was excluded from the analysis because their e-voucher redemption status was uncertain, leaving 2708 participants in the intervention arm and 2740 participants in the control arm. Site and participant characteristics were well-balanced between arms (Table 1). In the year before the study, the average (SD) redemption rate at the participating sites was 54% (SD 12%) in control groups and 54% (SD 13%) in intervention groups. (range 3%-68%). In the control groups, the average age of participants was 26.4 (6.2) years old (range 13-49), the median parity was 2, 87% (2372/2740) were married, and 62% (1693/2740) were known to have their own phones (no proxy phone). In the intervention groups, the average of participants was 26.4 (6.2), the median parity was 2 (IQR 1-4), 85% (2311/2708) were married, and 62% (1687/2708) were known to have their own phones. Where data for marital status, education, or proxy phone status is missing, this indicates that this data was not collected by antenatal nurses, or participants declined to provide this information (Table 1).

XSL•FO
RenderX

Figure 2. CONSORT diagram of clinic and participant flow. CONSORT: Consolidated Standards of Reporting Trials.

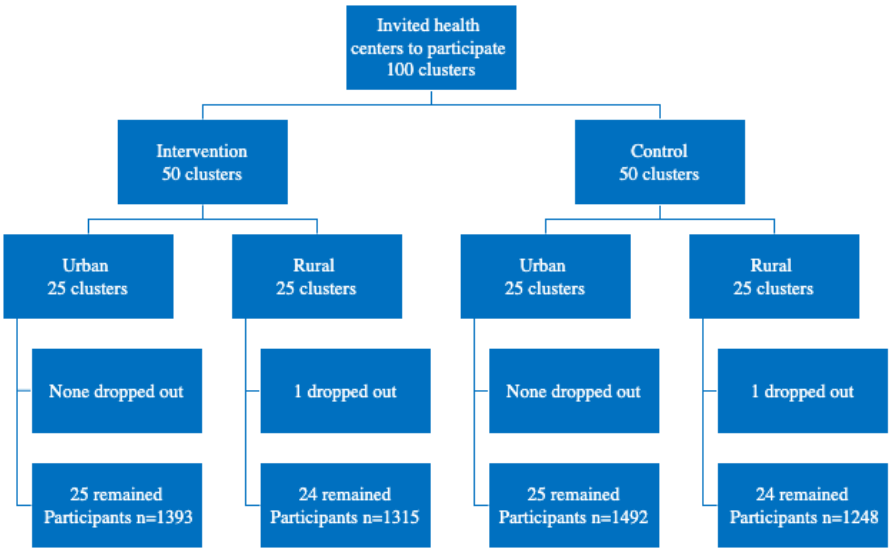

Table 1. Baseline site and participant characteristics.

	Control	Intervention
Clinic characteristics	Clinics (n=49)	Clinics (n=49)
Location		
Rural, n (%)	24 (49)	24 (49)
Urban, n (%)	25 (51)	25 (51)
Baseline redemption rate (%), mean (SD)	54 (12)	54 (13)
Participant characteristics	Participants (n=2740)	Participants (n=2708)
Age, mean (SD)	26.4 (6.2)	26.4 (6·2)
Parity, median (IQR)	2 (1-4)	2 (1-4)
Number of children, median (IQR)	1 (0-3)	1 (0-3)
Number of LLINs[a] owned, median (IQR)	1 (1-2)	1 (1-2)
Number of LLINs accessed through TNVS[b], median (IQR)	0 (0-1)	0 (0-1)
Marital Status, n (%)		
Missing	87 (3)	99 (4)
Single	281 (10)	298 (11)
Married	2372 (87)	2311 (85)
Education, n (%)		
Missing	128 (5)	166 (6)
None	335 (12)	327 (12)
Primary	1736 (63)	1637 (60)
Secondary	447 (16)	468 (17)
College	94 (3)	110 (4)
Proxy, n (%)		
Missing	663 (24)	691 (26)
No	1693 (62)	1687 (62)
Yes	384 (14)	330 (12)

[a]LLINs: long-lasting insecticidal nets.

[b]TNVS: Tanzania National Voucher Scheme.

XSL•FO
RenderX

Figure 3 displays the within-site-specific change in redemption rates from the previous year compared with the rate during the study. Generally, the rates increased. The actual site-specific counts are presented in Table 2. The one notable outlier was the 24th rural site in the control group (Table 2). This site had a baseline redemption rate of 62% yet had only a 6% (1/18) redemption rate during the study. This site had substantially fewer participants than the next lowest accruing site, which had 33 participants (all other sites accrued at least 50 participants as planned) and had a much lower redemption rate than the next lowest site, which had a rate of 21%. After querying this site, it was found that they had administrative problems, which hindered accrual and redemption. Our primary analysis includes all sites as randomized, but a sensitivity analysis excluding this one outlying site was performed.

In Figure 3, the black lines depict site-specific rates, and the bold red line shows the overall mean rate.

Overall, the raw redemption rate in the intervention arm was 70% compared with 67% in the control arm. The difference after controlling for the sites' reach (urban vs rural) and the sites' baseline redemption rates was 4% (95% CI –2% to 10%; *P*=.19). After excluding the one outlying site, the raw rates were 70% and 68% in the intervention and control arms, respectively, and the adjusted difference was 3% (95% CI –3% to 9%; *P*=.31). The intracluster correlation of e-voucher redemption status (without adjusting for covariates) was 0·12 (95% CI 0.05 to 0.15) in the intervention arm and 0·13 (95% CI 0.7 to 0.16) in the control arm. In a prespecified subgroup analysis, the 3380 (62%) participants who were known to have their own phones were included. In this subgroup, the raw redemption rates in the intervention and control arm were 71% and 69%, respectively, with an adjusted risk difference of 3% (95% CI –3% to 9%; *P*=.27).

Table 3 compares the site and participant characteristics between participants who did and did not redeem their e-vouchers. Of note, people who redeemed their vouchers were 0.7 years older than those who did not. E-voucher redemption status was lower among participants where marital status, education, or proxy status was missing; however, there was no difference in e-voucher redemption rates among known education levels, marital status, or proxy status. Where proxy status was missing, this indicates that this data was not collected by antenatal nurses or participants declined to provide this information (Table 3). On average, participants who redeemed their e-vouchers came from sites with a baseline rate of 2% (*P*=.04) higher than participants who did not redeem their e-vouchers. In fact, aggregated to the site level, the correlation between the previous redemption rate compared with the study redemption rate was *r*=0.23 (*P*=.03), and after excluding the one outlier site (with only 18 participants and 1 redemption), the correlation became *r*=0.27 (*P*=.007). Parity, number of children, number of LLINs owned, number of LLINs accessed through TNVS, and clinic reach were not significantly different between participants who did and did not redeem their e-vouchers.

All participant characteristics were missing for 186 participants, an additional 55 were missing parity, an additional 108 were missing education, and an additional 950 were missing proxy. For categorical variables, the total number of missing values is displayed, but for continuous variables, the missing values have been excluded.

Table 4 provides the single predictor and multiple predictor regression results. The multiple predictor regression included all predictors that were significant at *P*<.15 by backward selection. From the multiple predictor model, it was determined that for every 10% increase in the sites' baseline redemption rate, the study redemption rate increased by 3% (95% CI 0%-6%, *P*=.03), and for every decade of age, the redemption rate increased by 4% (95% CI 2%-5%, *P*<.001).

XSL•FO
RenderX

Figure 3. Site-specific redemption rates. Black lines depict site specific rates and bold red line shows overall mean rate.

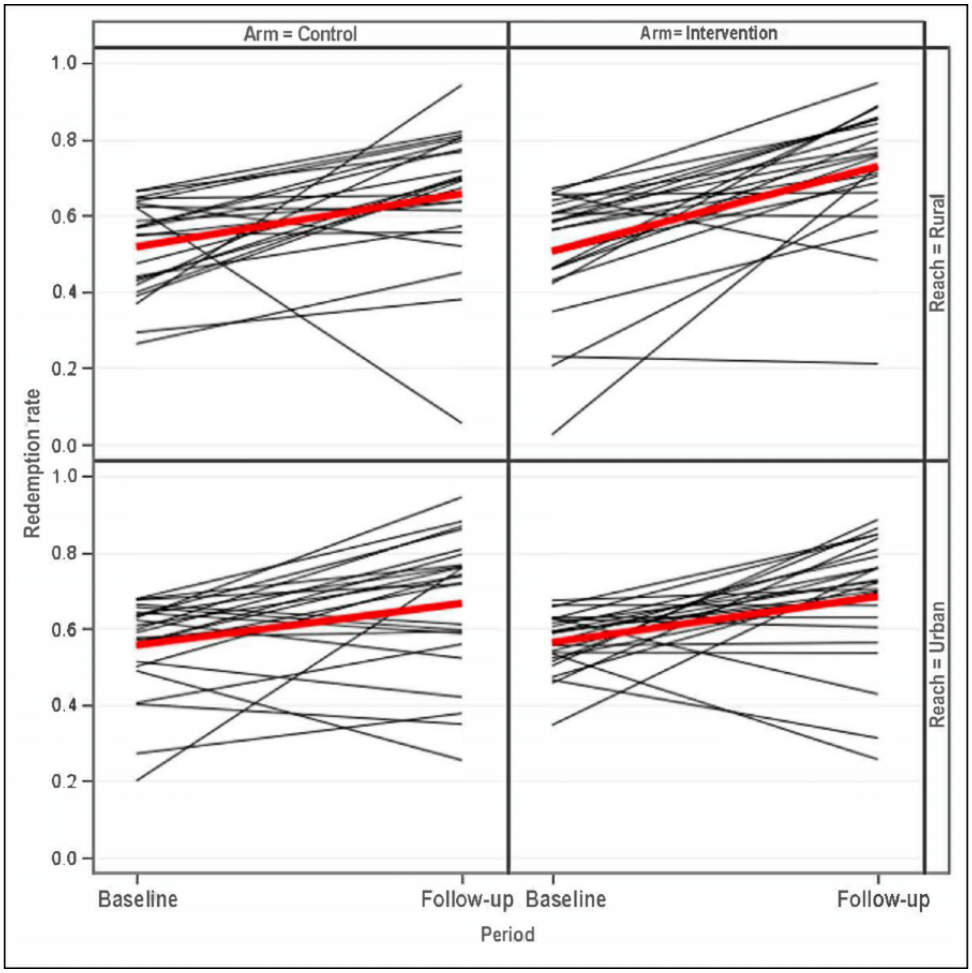

XSL•FO
RenderX

Table 2. Number of vouchers issued and redeemed for each site. Rows are sorted by the descending number of vouchers issued.

	Intervention arm				Control arm			
Site	Urban		Rural		Urban		Rural	
	Previous, %	Follow-up, n/N (%)	Previous, %	Follow-up, n/N (%)	Previous, %	Follow-up, n/N (%)	Previous, %	Follow-up, n/N (%)
1	53	18/70 (26)	46	49/65 (75)	61	53/72 (74)	44	36/63 (57)
2	61	47/65 (72)	42	55/62 (89)	56	54/71 (76)	42	50/62 (81)
3	59	44/63 (70)	46	53/60 (88)	62	36/69 (52)	62	38/62 (61)
4	50	54/61 (89)	35	33/59 (56)	40	38/68 (56)	66	46/60 (77)
5	52	41/59 (69)	61	34/57 (60)	56	54/68 (79)	47	41/59 (69)
6	54	44/58 (76)	66	54/57 (95)	67	48/67 (72)	39	38/55 (69)
7	61	45/57 (79)	58	48/56 (86)	50	49/66 (74)	55	35/55 (64)
8	54	30/56 (54)	66	27/56 (48)	40	22/63 (35)	43	38/54 (70)
9	63	24/56 (43)	3	40/55 (73)	64	48/63 (76)	55	30/54 (56)
10	66	37/56 (66)	56	39/55 (71)	64	37/63 (59)	57	43/54 (80)
11	46	46/55 (84)	43	38/54 (70)	57	41/62 (66)	57	38/53 (72)
12	56	31/55 (56)	50	37/54 (69)	68	46/60 (77)	57	41/53 (77)
13	59	40/55 (73)	56	41/54 (76)	20	44/58 (76)	65	43/53 (81)
14	59	41/54 (76)	61	46/54 (85)	55	41/57 (72)	37	49/52 (94)
15	59	37/54 (69)	63	46/54 (85)	63	49/57 (86)	43	35/52 (67)
16	63	34/54 (63)	64	42/54 (78)	66	34/57 (60)	63	27/52 (52)
17	67	38/54 (70)	21	34/53 (64)	58	32/54 (59)	65	34/52 (65)
18	63	32/53 (60)	59	38/53 (72)	63	51/54 (94)	26	23/51 (45)
19	56	42/52 (81)	23	11/52 (21)	66	33/54 (61)	64	41/51 (80)
20	63	44/52 (85)	61	39/51 (76)	27	20/53 (38)	29	19/50 (38)
21	66	44/52 (85)	46	40/50 (80)	60	46/53 (87)	40	35/50 (70)
22	47	16/51 (31)	58	41/50 (82)	59	42/52 (81)	66	41/50 (82)
23	51	44/51 (86)	66	33/50 (66)	49	13/51 (25)	58	21/33 (64)
24	35	38/50 (76)	67	42/50 (84)	51	21/50 (42)	62	1/18 (6)
25	47	36/50 (72)	—[a]	—	68	44/50 (88)	—	—
Arm total by strata	56	947/1393 (68)	51	960/1315 (73)	56	996/1492 (67)	52	843/1248 (68)
Arm total	54	1907/2708 (70)			54	1839/2740 (67)		

[a]Not available.

XSL•FO
RenderX

Table 3. Voucher redemption status by baseline site and participant characteristics.

Characteristics	Redeemed, clinics (n=49); participants (n=3746)	Not redeemed, clinics (n=49); participants (n=1702)	*P* value[a]
Clinic characteristics			
Location, n (%)			
Rural	1803 (70)	760 (30)	.52
Urban	1943 (67)	942 (33)	
Baseline redemption[b] (%), mean (SD)	54 (12)	52 (13)	.04
Participant characteristics			
Age[b], mean (SD)	26.6 (6.2)	25.9 (6)	<.001
Parity[b], median (IQR)	2 (1-4)	2 (1-4)	.06
Number of children[b], median (IQR)	2 (0-3)	2 (0-3)	.07
Number of LLINs[c] owned[b], median (IQR)	1 (1-2)	1 (1-2)	.42
Number of LLINs accessed through TNVS[b,d], median (IQR)	0 (0-1)	0 (0-1)	.96
Marital status, n (%)			<.001
Missing	64 (34)	122 (66)	(.58)
Single	406 (70)	173 (30)	—[e]
Married	3276 (70)	1407 (30)	—
Education, n (%)			.005
Missing	140 (48)	154 (52)	(.80)
None	468 (71)	194 (29)	—
Primary	2373 (70)	1000 (30)	—
Secondary	627 (69)	288 (31)	—
College	138 (68)	66 (32)	—
Proxy, n (%)			
Missing	879 (65)	475 (35)	.008
No	2373 (70)	1007 (30)	(.34)
Yes	494 (69)	220 (31)	—

[a]*P* values estimated from generalized estimating equations with redeemed (yes or no) as the dependent variable. For categorical variables, missing was included as a category in the top *P* value, while the second *P* value (in parentheses) excludes missing values.

[b]Continuous predictors are modeled as linear, but the number of LLINs owned and accessed through TNVS was truncated at 5 since less than 1% of participants had a value >5.

[c]LLINs: long-lasting insecticidal nets.

[d]TNVS: Tanzania National Voucher Scheme.

[e]Not applicable.

XSL•FO
RenderX

Table 4. Regression models for predictors of voucher redemption rates.

	Single predictor models		Selected model[a]	
Characteristics	Risk difference (95% CI)	*P* value	Risk difference (95% CI)	*P* value
Clinic characteristics (%), median (IQR)				
Location				
Rural	2 (–4 to 9)	.51	5 (–1 to 11)	.12
Urban	Referent			
Baseline redemption (per 10%)	3 (0 to 6)	.04	3 (0 to 6)	.03
Participant characteristics (%), median (IQR)				
Age (per decade)[b]	4 (2 to 6)	<.001	4 (2 to 5)	<.001
Parity (per 1)[b]	1 (0 to 1)	.06	—[c]	—
Number of children (per child)[b]	1 (0 to 1)	.07	1 (0 to 1)	.11
Number of LLINs[d] owned (per 1)[b]	0 (–1 to 1)	.43	—	—
Number of LLINS accessed through TNVS[e] per 1)[b]	0 (–1 to 1)	.96	—	—
Single vs married	–1 (–5 to 3)	.58	—	—
Education (%), median (IQR)		.8	—	—
None	1 (–3 to 4)	—	—	—
Primary	Referent	—	—	—
Secondary	–1 (–4 to 2)	—	—	—
College	–2 (–8 to 4)	—	—	—
Proxy (%), median (IQR)				
No	Referent	.34		
Yes	–3 (–5 to 2)	—	—	—

[a]The selected model excludes 186 participants with missing values of the selected variables.

[b]Continuous predictors are modeled as linear, but the number of LLINs owned and accessed through TNVS was truncated at 5 since less than 1% of participants had a value >5.

[c]Not applicable.

[d]LLINs: long-lasting insecticidal nets.

[e]TNVS: Tanzania National Voucher Scheme.

Discussion

Principal Findings

This cluster randomized trial aimed to assess the effectiveness of mHealth technology in the uptake of LLINs for pregnant women in Tanzania. It was anticipated that targeted BCC mobile messages would increase the redemption of e-vouchers for LLINs in regions of Tanzania with high malaria prevalence and a history of low LLIN uptake. However, our results demonstrate no significant difference in the redemption of e-vouchers between pregnant women in the intervention group that received BCC mobile messages and those in the control group that did not receive mobile messages.

Comparison With Previous Work

The existing literature using mHealth messaging as a tool for malaria prevention and control in sub-Saharan Africa shows a combination of both successful and unsuccessful interventions. There are several studies investigating the feasibility of mHealth technology as a malaria surveillance tool in sub-Saharan Africa that have shown encouraging results [18-20]. Evidence also suggests that mobile messaging improved antimalarial drug supply management in Tanzania, improved Kenyan health care workers' adherence to malaria treatment guidelines, increased adherence to malaria rapid diagnostic test results in Nigeria, and increased adherence to artemisinin-based combination therapy regiments in Ghana [21-24]. There are also several mHealth studies with conflicting results. A cluster randomized trial sending mobile messages about artemether-lumefantrine treatment to retail staff in Tanzania drug shops found some improvement in dispenser knowledge but no difference in participant adherence, suggesting that increased knowledge of retail staff did not translate to increased adherence in participants [25]. Furthermore, a randomized controlled trial in Ghana sending mobile messages to caregivers of children with malaria under 5 years of age found increased posttreatment return to the health facility but no impact on treatment adherence [26]. It was suggested that this could be due to adherence levels already being high in both groups, meaning that mobile

XSL•FO
RenderX

messages were unnecessary. Similarly, it is possible that pregnant women in Tanzania have significant baseline knowledge that LLIN use in pregnancy is important, regardless of educational mobile messaging.

In the HASA study, site and participant characteristics were well-balanced between intervention and control groups, and characteristics including parity, number of children, education levels, and marital status did not impact redemption. Notably, younger participants were less likely to redeem their e-vouchers, and for every decade of age, the redemption rate increased by 4%. Previous literature from Tanzania suggests that individuals in the age group 30-49 had the greatest knowledge of malaria, potentially explained by having more life experience than younger individuals [27]. Together, this may suggest a need for interventions directly targeting pregnant women below the age of 30 years.

Despite a null result, both groups experienced an increase in LLIN uptake from baseline during the trial. As such, one possible explanation for the null result is that nurses at participating antenatal clinics contributed to the increased uptake of LLINs in both groups. Despite nurses being blinded to their intervention group, antenatal nurses likely played a role in encouraging LLIN uptake. Antenatal clinic nurses were involved in orientation and extensive training for the HASA trial (nurse training described in subsequent manuscript), which may have prompted them to provide additional information about malaria and motivate pregnant women in both groups to follow through on e-voucher redemption upon enrollment into the study [16]. This would be consistent with previous findings in Tanzania that for pregnant women, knowledge of malaria correlates with ITN uptake [28]. While this represents an overall positive health outcome resulting from the trial implementation, it is a type of contamination bias from a clinical trial perspective, whereby the effects of the intervention were received by members of the control group and possibly shifted the outcome of the control group in the same direction as the intervention group [29]. Contamination bias is a potential problem for clinical trials of educational and behavioral interventions due to the ease of transferring intervention knowledge; in this case, information about the importance of LLINs and malaria prevention could be inadvertently shared with participants in the control group by nurses, thereby changing the behavior of the control group participants and impacting the results of the trial [29].

The current prevailing practices for increasing ITN uptake in pregnant women focus on a multipronged approach. This includes integrating ITN provision and counseling into antenatal care visits, engaging and educating the community about the importance of ITNs, strengthening health care systems by ensuring the availability and distribution of ITNs to health care facilities, and training health care workers on ITN promotion and counseling techniques [30,31]. Due to the importance of increasing widespread LLIN use in pregnancy to prevent malaria and its morbidity and mortality, and given the significant body of evidence for the use of mobile messaging to promote healthy behavior change in low- and middle-income country contexts, the authors feel that there is still significant equipoise in this area of research.

Strengths and Limitations

A major strength of the HASA study was the distribution of e-vouchers from antenatal clinics. Antenatal care coverage in Tanzania has been above 90% of pregnant women for over 2 decades, with 98% receiving care from a skilled provider in 2015-2016 [32]. Over half of pregnant women in Tanzania had 4 or more antenatal facility visits during the year of the trial, providing ample opportunity for nurses to provide them with LLIN e-vouchers at visits [32].

An additional strength of the study was that it was a large, robust study. A total of 5448 women were enrolled in the study, and 3746 e-vouchers were redeemed. One can assume that 5.5 lives per year are saved for every 1000 children with an LLIN in the household [33]. With 3746 e-vouchers redeemed by pregnant women with an average of 2 children each already and a new baby (assuming all pregnancies resulted in a live birth), the HASA study protected 11,238 children and saved approximately 62 lives, which is a success in terms of trial implementation.

A major challenge encountered in this study was the low penetration of mobile phone ownership in women at numerous study facility sites, with only 62% (1687/2708) of participants possessing their own phones. This represents a barrier for researchers and health program implementers considering using a mHealth strategy to improve access to health care and health education for rural-dwelling low-income women. Participants whose phone proxy status was missing or not available were less likely to redeem their e-vouchers. This low penetration reflects the gap in gender equity in mobile phone ownership. In sub-Saharan Africa, women are 13% less likely than men to own a mobile phone; literacy, digital skills, and affordability are the largest barriers to phone access [34]. Women in rural areas, particularly Maasai women, may have been disadvantaged in terms of the trial as they are less likely to own a phone; cell coverage in these rural areas can also be particularly poor.

Unexpected outages of cellular network services in many parts of Tanzania during the study represented another challenge. Clinic nurses reported network outages and the resulting inability to activate e-vouchers while a participating woman was at the clinic. The investigators are aware that some women who sought to participate were unable to either receive or redeem their e-vouchers due to connectivity issues. Given that cellular coverage is increasing in rural areas and network reliability is consistently improving, perhaps in a few years, these issues will be less prevalent for mHealth studies.

Previous findings suggest that in Tanzania, women of low socioeconomic status have a lower uptake of ITN vouchers than women of higher socioeconomic status since there is still some expense required to obtain a net [35]. The voucher redemption rate for ITNs in Tanzania is influenced by socioeconomic status and co-payment cost [36]. This disparity for women of lower socioeconomic status may be amplified using e-vouchers, as these women may be less likely to have their own mobile phones or live in an area with reliable connectivity. Despite the co-payment cost being relatively low, this finding suggests the need for interventions targeted at low-income women.

XSL•FO
RenderX

In Tanzania, the literature has indicated past challenges with the misuse of ITN vouchers. A study investigating a paper discount voucher scheme in Tanzania in 1999 suggested that some vouchers were not being used for the intended population (pregnant women and young children), likely resulting in increased usage of nets in men and older children [35]. As the HASA study measured e-voucher redemption, a limitation is an inability to determine whether the LLINs were used by the intended population following e-voucher redemption. Furthermore, in 2013, there was evidence of fraudulent e-voucher activities by some clinic and retail staff, whereby staff could create vouchers for fake beneficiaries so that LLINs could be subsequently sold on the commercial market [7]. For the HASA study, the potential of e-voucher fraud was addressed by a redesign of the e-voucher to create a more secure system. Components of the redesign included a unique identifier for all pregnant women, random barcode usage activated once at retailers, periodic reports to the clinic in charge to verify the numbers issued, the ability of participants to provide SMS feedback to confirm whether a voucher code was received, and a log indicating SMS delivery confirmation or errors causing delivery failure.

Conclusions

The HASA study has demonstrated that the use of a mHealth intervention for BCC mobile messaging does not result in a significant increase in LLIN uptake for pregnant women. However, the use of an e-voucher distributed by nurses in antenatal clinics in partnership with local retailers is feasible on a large scale and holds promise for other health interventions. Future studies evaluating e-vouchers or mHealth initiatives should take into consideration the barriers to phone ownership experienced by those who are most vulnerable, especially among women who are low-income and live in rural areas.

Acknowledgments

The authors acknowledge Michael Sarco, Thom Dixon, Moses Mokua, Ken Kocon, and the Mennonite Economic Development Associates (MEDA) for their support of this project. This Project is supported by Grand Challenges Canada. Grand Challenges Canada is funded by the Government of Canada and is dedicated to supporting Bold Ideas with Big Impact.

Data Availability

The datasets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KY developed this research question, designed the study, and supervised the conduct of the HASA study. AD designed the analytic plan, conducted some analysis, and supervised and validated parts of the analysis done by Ken Kocon. TV had primary responsibility for the writing of the report and review of data analysis. EK, JS, NW, and KN were involved in the coordination of the HASA study at the project office or in Tanzania and provided comments during the writing and editing of the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHealth Checklist.
[PDF File (Adobe PDF File), 2361 KB - jmir_v27i1e51524_app1.pdf]

References

1. World malaria report 2021. World Health Organization. URL: https://www.who.int/publications-detail-redirect/9789240040496 [accessed 2021-12-06]
2. Schantz-Dunn J, Nour NM. Malaria and pregnancy: a global health perspective. Rev Obstet Gynecol 2009;2(3):186-192. [Medline: 19826576]
3. Gamble CL, Ekwaru JP, ter Kuile FO. Insecticide-treated nets for preventing malaria in pregnancy. Cochrane Database Syst Rev 2006;2006(2):CD003755 [FREE Full text] [doi: 10.1002/14651858.CD003755.pub2] [Medline: 16625591]
4. WHO malaria terminology, 2021 update. World Health Organization. URL: https://www.who.int/publications/i/item/9789240038400 [accessed 2021-11-24]
5. Updated WHO recommendations for malaria chemoprevention among children and pregnant women. World Health Organization. URL: https://tinyurl.com/5n8dhuce [accessed 2022-06-03]
6. Littmann J, Achu D, Laufer MK, Karema C, Schellenberg D. Making the most of malaria chemoprevention. Malar J 2024;23(1):51 [FREE Full text] [doi: 10.1186/s12936-024-04867-5] [Medline: 38369497]
7. Kramer K, Mandike R, Nathan R, Mohamed A, Lynch M, Brown N, et al. Effectiveness and equity of the Tanzania National Voucher Scheme for mosquito nets over 10 years of implementation. Malar J 2017;16(1):255 [FREE Full text] [doi: 10.1186/s12936-017-1902-0] [Medline: 28619076]

XSL•FO
RenderX

8. Eva G, Quinn A, Ngo TD. Vouchers for family planning and sexual and reproductive health services: a review of voucher programs involving Marie Stopes International among 11 Asian and African countries. Int J Gynaecol Obstet 2015;130 Suppl 3:E15-E20 [FREE Full text] [doi: 10.1016/j.ijgo.2015.06.023] [Medline: 26165906]
9. Wossen T, Abdoulaye T, Alene A, Feleke S, Ricker-Gilbert J, Manyong V, et al. Productivity and welfare effects of Nigeria's e-voucher-based input subsidy program. World Dev 2017;97:251-265 [FREE Full text] [doi: 10.1016/j.worlddev.2017.04.021] [Medline: 29263568]
10. Horton S, Saleh N, Mosha T. CIFSRF final technical report: promoting locally fortified sunflower oil using e-vouchers (CIFSRF Phase 2). URL: http://hdl.handle.net/10625/56751 [accessed 2025-01-22]
11. The mobile economy sub-Saharan Africa 2021. GMSA. URL: https://www.gsma.com/mobileeconomy/sub-saharan-africa/ [accessed 2025-01-22]
12. Assessment of behavior change communication (BCC) interventions in support of malaria control activities conducted in Benin by PMI's ARM3 project | Cochrane Library. MCD Global Health. URL: https://www.mcd.org/docs/pubs/PMI-ARM3-Assessment-of-BCC.pdf [accessed 2025-01-22]
13. Mugisa M, Muzoora A. Behavioral change communication strategy vital in malaria prevention interventions in rural communities: Nakasongola district, Uganda. Pan Afr Med J 2012;13 Suppl 1(Suppl 1):2 [FREE Full text] [Medline: 23467840]
14. Renggli S, Mandike R, Kramer K, Patrick F, Brown NJ, McElroy PD, et al. Design, implementation and evaluation of a national campaign to deliver 18 million free long-lasting insecticidal nets to uncovered sleeping spaces in Tanzania. Malar J 2013;12:85 [FREE Full text] [doi: 10.1186/1475-2875-12-85] [Medline: 23496881]
15. Fact Sheet: Teenage Pregnancy. UNFPA Tanzania. URL: https://tinyurl.com/2s4ehpst [accessed 2025-01-22]
16. Vey T, Kinnicutt E, West N, Sleeth J, Nchimbi KB, Yeates K. Implementation of a mobile health approach to a long-lasting insecticidal net uptake intervention for malaria prevention among pregnant women in Tanzania: process evaluation of the Hati Salama (HASA) randomized controlled trial study. J Med Internet Res 2024;26:e51527 [FREE Full text] [doi: 10.2196/51527] [Medline: 39499551]
17. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemp Clin Trials 2012;33(5):869-880 [FREE Full text] [doi: 10.1016/j.cct.2012.05.004] [Medline: 22627076]
18. Randrianasolo L, Raoelina Y, Ratsitorahina M, Ravolomanana L, Andriamandimby S, Heraud J, et al. Sentinel surveillance system for early outbreak detection in Madagascar. BMC Public Health 2010;10:31. [doi: 10.1186/1471-2458-10-31] [Medline: 20092624]
19. Githinji S, Kigen S, Memusi D, Nyandigisi A, Wamari A, Muturi A, et al. Using mobile phone text messaging for malaria surveillance in rural Kenya. Malar J 2014;13:107 [FREE Full text] [doi: 10.1186/1475-2875-13-107] [Medline: 24642130]
20. Asiimwe C, Gelvin D, Lee E, Ben Amor Y, Quinto E, Katureebe C, et al. Use of an innovative, affordable, and open-source short message service-based tool to monitor malaria in remote areas of Uganda. Am J Trop Med Hyg 2011;85(1):26-33 [FREE Full text] [doi: 10.4269/ajtmh.2011.10-0528] [Medline: 21734120]
21. Barrington J, Wereko-Brobby O, Ward P, Mwafongo W, Kungulwe S. SMS for Life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. Malar J 2010;9:298 [FREE Full text] [doi: 10.1186/1475-2875-9-298] [Medline: 20979633]
22. Zurovac D, Sudoi RK, Akhwale WS, Ndiritu M, Hamer DH, Rowe AK, et al. The effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. Lancet 2011;378(9793):795-803 [FREE Full text] [doi: 10.1016/S0140-6736(11)60783-6] [Medline: 21820166]
23. Modrek S, Schatzkin E, De La Cruz A, Isiguzo C, Nwokolo E, Anyanti J, et al. SMS messages increase adherence to rapid diagnostic test results among malaria patients: results from a pilot study in Nigeria. Malar J 2014;13:69 [FREE Full text] [doi: 10.1186/1475-2875-13-69] [Medline: 24564925]
24. Raifman JRG, Lanthorn HE, Rokicki S, Fink G. The impact of text message reminders on adherence to antimalarial treatment in northern Ghana: a randomized trial. PLoS One 2014;9(10):e109032. [doi: 10.1371/journal.pone.0109032] [Medline: 25350546]
25. Bruxvoort K, Festo C, Kalolella A, Cairns M, Lyaruu P, Kenani M, et al. Cluster randomized trial of text message reminders to retail staff in Tanzanian drug shops dispensing artemether-lumefantrine: effect on dispenser knowledge and patient adherence. Am J Trop Med Hyg 2014;91(4):844-853 [FREE Full text] [doi: 10.4269/ajtmh.14-0126] [Medline: 25002300]
26. Talisuna AO, Oburu A, Githinji S, Malinga J, Amboko B, Bejon P, et al. Efficacy of text-message reminders on paediatric malaria treatment adherence and their post-treatment return to health facilities in Kenya: a randomized controlled trial. Malar J 2017;16(1):46 [FREE Full text] [doi: 10.1186/s12936-017-1702-6] [Medline: 28122622]
27. Spjeldnæs AO, Kitua AY, Blomberg B. Education and knowledge helps combating malaria, but not degedege: a cross-sectional study in Rufiji, Tanzania. Malar J 2014;13:200 [FREE Full text] [doi: 10.1186/1475-2875-13-200] [Medline: 24885180]
28. Nganda RY, Drakeley C, Reyburn H, Marchant T. Knowledge of malaria influences the use of insecticide treated nets but not intermittent presumptive treatment by pregnant women in Tanzania. Malar J 2004;3:42 [FREE Full text] [doi: 10.1186/1475-2875-3-42] [Medline: 15541178]

XSL•FO
RenderX

29. Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, et al. Contamination in trials of educational interventions. Health Technol Assess 2007;11(43):iii, ix-iii,107 [FREE Full text] [doi: 10.3310/hta11430] [Medline: 17935683]
30. Wafula ST, Mendoza H, Nalugya A, Musoke D, Waiswa P. Determinants of uptake of malaria preventive interventions among pregnant women in eastern Uganda. Malar J 2021;20(1):5 [FREE Full text] [doi: 10.1186/s12936-020-03558-1] [Medline: 33390153]
31. Insecticide-treated nets to reduce the risk of malaria in pregnant women. World Health Organization. URL: https://www.who.int/tools/elena/interventions/bednets-malaria-pregnancy [accessed 2023-08-09]
32. Tanzania demographic and health survey and malaria indicator survey (TDHS-MIS) 2015-16. Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania, Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief, Government Statistician (OCGS), and ICF. URL: https://dhsprogram.com/pubs/pdf/FR321/FR321.pdf [accessed 2025-01-22]
33. Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. Cochrane Database Syst Rev 2004(2):CD000363. [doi: 10.1002/14651858.CD000363.pub2] [Medline: 15106149]
34. Rowntree O, Shanahan M, Bahia K, Butler C, Lindsey D, Sibthorpe C. Connected women: the mobile gender gap report 2020. GSMA. 2020. URL: https://tinyurl.com/3tfrb984 [accessed 2024-01-23]
35. Mushi AK, Schellenberg JRMA, Mponda H, Lengeler C. Targeted subsidy for malaria control with treated nets using a discount voucher system in Tanzania. Health Policy Plan 2003;18(2):163-171 [FREE Full text] [doi: 10.1093/heapol/czg021] [Medline: 12740321]
36. Eze IC, Kramer K, Msengwa A, Mandike R, Lengeler C. Mass distribution of free insecticide-treated nets do not interfere with continuous net distribution in Tanzania. Malar J 2014;13:196 [FREE Full text] [doi: 10.1186/1475-2875-13-196] [Medline: 24884786]

Abbreviations

BCC: behavior change communication
HASA: Hati Salama
ITN: insecticide-treated nets
LLIN: long-lasting insecticidal net
mHealth: mobile health
TNVS: Tanzania National Voucher Scheme

Edited by A Mavragani; submitted 02.08.23; peer-reviewed by T Hanjahanja-Phiri, B Yovogan, J Matowo; comments to author 24.05.24; revised version received 31.07.24; accepted 02.10.24; published 19.03.25.

Please cite as:
Vey T, Kinnicutt E, Day AG, West N, Sleeth J, Nchimbi KB, Yeates K
Targeted Behavior Change Communication Using a Mobile Health Platform to Increase Uptake of Long-Lasting Insecticidal Nets Among Pregnant Women in Tanzania: Hati Salama "Secure Voucher" Study Cluster Randomized Controlled Trial
J Med Internet Res 2025;27:e51524
URL: https://www.jmir.org/2025/1/e51524
doi:10.2196/51524
PMID:

©Trinity Vey, Eleonora Kinnicutt, Andrew G Day, Nicola West, Jessica Sleeth, Kenneth Bernard Nchimbi, Karen Yeates. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Spatiotemporal and Behavioral Patterns of Men Who Have Sex With Men Using Geosocial Networking Apps in Shenzhen From Mobile Big Data Perspective: Longitudinal Observational Study

Lan Wei[1], PhD; Yongsheng Wu[1], MMed; Lin Chen[1], MS; Jinquan Cheng[1], PhD; Jin Zhao[1], PhD

Shenzhen Center for Disease Control and Prevention, Shenzhen, China

Corresponding Author:
Jin Zhao, PhD
Shenzhen Center for Disease Control and Prevention
No 8 Longyuan Road, Longzhu Avenue
Nanshan District
Shenzhen, 518055
China
Phone: 86 075525632404
Email: szcdczj@163.com

Abstract

Background: The use of geosocial networking apps is linked to increased risky sexual behaviors among men who have sex with men, but their relationship with HIV and other sexually transmitted infections remains inconclusive. Since 2015, the prevalence of app use among men who have sex with men in Shenzhen has surged, highlighting the need for research on their spatiotemporal and behavioral patterns to inform targeted prevention and intervention strategies.

Objective: This study aims to investigate the population size, spatiotemporal and behavioral patterns, and mobility of app-using men who have sex with men in Shenzhen using mobile big data. The goal is to inform enhanced and innovative intervention strategies and guide health resource allocation.

Methods: By leveraging mobile big data application technology, we collected demographic and geographic location data from 3 target apps—Blued (Blued Inc), Jack'd (Online Buddies Inc), and Zank (Zank Group)—over continuous time periods. Spatial autocorrelation (Global Moran I) and hot spot analysis (Getis-Ord Gi) were used to identify the geographic clusters. The Geodetector tool (Chinese Academy of Sciences) was adopted to measure spatially stratified heterogeneity features.

Results: From September 2017 to August 2018, a total of 158,387 males aged 15-69 years in Shenzhen used one of the 3 apps, with the majority (71,318, 45.03%) aged 25-34 years. The app user-to-male ratio was approximately 2.6% among all males aged 15-69 years. The estimated population of app-using men who have sex with men in Shenzhen during this period was 268,817. The geographic distribution of app-using men who have sex with men in Shenzhen was clustered, with hot spots primarily located in central and western Shenzhen, while the distribution of HIV testing and counseling was more concentrated in central-eastern Shenzhen. Approximately 60,202 (38%) app-using men who have sex with men left Shenzhen during the Spring Festival, and 37,756 (62.7%) of them returned after the holiday. The destination distribution showed a relatively centralized flow throughout the country, with the largest mobility within Guangdong province (67.7%), followed by lower mobility to Hunan province (7.9%) and other neighboring provinces (3%-5%), such as Jiangxi, Guangxi, and Hubei Provinces.

Conclusions: Shenzhen has a large population of men who have sex with men. The variation and inconsistent spatiotemporal distribution of app use and HIV testing and counseling emphasize the need to adapt traditional venue-based prevention and intervention to identified hot spots and to launch outreach initiatives that extend beyond traditional healthcare settings. Given the relatively high internal and interprovincial mobility of app-using men who have sex with men, further smartphone-based behavioral monitoring could provide valuable insights for developing enhanced and innovative HIV prevention and intervention strategies. Moreover, our study demonstrates the potential of mobile big data to address critical research gaps often overlooked by traditional methods.

(J Med Internet Res 2025;27:e69569) doi:10.2196/69569

XSL•FO
RenderX

KEYWORDS

men who have sex with men; geosocial networking apps; distribution; heterogeneity; mobility; HIV testing; risk behavior; big data

Introduction

Since the first case of AIDS was reported among men who have sex with men in the United States in 1981, AIDS has remained a major global public health issue. As of the end of 2023, it is estimated there are 39.9 million people living with HIV, with 1.3 million newly infected [1]. According to the Chinese Center for Disease Control and Prevention (CDC), there were 1.22 million people living with HIV in China by the end of 2022, with 107,000 new cases reported that year. The proportion of newly reported HIV or AIDS cases transmitted through men who have sex with men has grown rapidly in the past decades, rising from 9.1% in 2009 to 25.6% in 2022 [2,3]. Men who have sex with men have emerged as a significant high-risk group globally, with a likelihood of acquiring HIV that is 28 times greater than that of the general population [4]. In China, the HIV prevalence among this population has risen from 1.4% in 2005 to 6% in 2021 [5,6], raising concerns about the sustained effectiveness of current public health initiatives and underscoring the need for targeted interventions.

With the rapid development of the internet and social media, men who have sex with men can connect with others through various web-based platforms, including smartphone-based geosocial networking (GSN) apps. Unlike traditional fixed-location social networking, GSN apps effectively bridge online and offline dating functions, facilitating the process of seeking relationships or casual encounters. This convenience and privacy have led to the rising popularity of these apps, particularly in countries or regions with low levels of social acceptance of homosexuality. GSN apps designed for men who have sex with men, such as Grindr and Jack'd, have rapidly gained popularity in Western countries since 2009 [7]. Subsequently, several other apps including Blued and Aloha, emerged and have become increasingly popular among Chinese men who have sex with men since 2013 [8]. Blued is the first and most popular gay dating app of its kind in China. Its global users have increased from 27 million in 2015 to 49 million in 2020, with approximately 70% of users being Chinese [9].

Prior research found about 66-67% of the men who have sex with men in China used GSN apps to seek sexual partners within the past 6 months [10,11]. Our previous cross-sectional study revealed a sharp increase in app use among men who have sex with men in Shenzhen, with the rate surging from 12.5% in 2015 to 52.6% in 2017 [12]. The rapid growth of gay dating apps suggests that they have become one of the primary means for seeking sexual partners in the Chinese men who have sex with men community. Literature has documented that the use of GSN dating apps is associated with both increased risky sexual behaviors (eg, multiple sexual partners, condomless anal intercourse, drug use) and risk-reduction behaviors (eg, HIV testing and consistent condom use), although their impact on the risk of HIV and other sexually transmitted infections remains inconclusive [12-14].

Currently, the HIV testing rate in China remains below UNAID's first 95% target, with only 78.7% of men who have sex with men having been tested for HIV, despite concerted efforts of the government, community organizations, and nongovernmental organizations [3]. Recent economic development in China has accelerated population mobility, which is recognized as an important driver for HIV transmission and a potential barrier to HIV care and treatment [15-17]. Prior research reported that the HIV epidemics in Shenzhen were predominantly driven by the migrant population [18]. Most previous studies and intervention programs relied on sentinel surveillance or venue-based populations, resulting in limited coverage of young and mobile populations. Therefore, there is a pressing need for further research into the crucial HIV-related behaviors of these marginalized groups, particularly concerning mobility patterns and access to HIV testing.

With the growing accessibility of the internet and rapid advancements in technology, mobile big data—encompassing aggregated demographics, social behavior, and geographic location information of users—has become a valuable tool for research, particularly in understanding HIV prevention, behavioral patterns, spatial dynamics, and migration trends among men who have sex with men. Earlier studies have used GSN apps to explore various aspects of the men who have sex with men including demographic characteristics, sexual practices, HIV testing, Pre-exposure prophylaxis (PrEP) use, and geospatial distribution, however, most of them were from developed countries or regions [19-23]. In China, domestic research on men who have sex with men and GSN app use primarily focused on risky sexual behavior, HIV infection, and sexually transmitted infections relying on self-reported data [24-27], which may introduce biases due to underreporting or overreporting of sensitive behaviors. Some recent studies explored spatial distribution, migration, and PrEP adherence among men who have sex with men through GSN apps [28,29]. However, most of these studies did not directly use mobile big data.

Shenzhen, designated as China's first special economic zone in 1979, has transformed from a small village into a metropolis with a population of over 15 million. With a massive influx of internal migrant workers from across the country, Shenzhen now has the largest proportion of immigrants in China, with 66% of its population comprised of internal migrants [30]. The relatively open and inclusive cultural environment of Shenzhen attracted many men who have sex with men from across the country. Surveillance data from 2020 revealed that 62.6% of newly diagnosed HIV or AIDS cases in Shenzhen were attributed to male-to-male sexual transmission, exceeding the national average [31]. A previous cross-sectional study estimated that the active population of men who have sex with men in Shenzhen reached approximately 150,000 in 2014 [31]. Our venue-based survey showed the proportion of men who have sex with men using gay dating apps in Shenzhen has rapidly increased since 2015 [12]. The demographic context of

XSL•FO
RenderX

Shenzhen, with its significant migrant population and a high proportion of men who have sex with men using gay dating apps, presents unique challenges for effective HIV prevention and interventions within this group. By leveraging large-scale mobile big data, this study aimed to investigate the actual population size, spatiotemporal and behavioral patterns, and mobility of GSN app-using men who have sex with men in Shenzhen, to gain deeper insights into this key population and inform the development of enhanced and innovative strategies for HIV prevention and interventions.

Methods

Setting

Mobile big data possesses characteristics such as wide coverage, real-time processing, and continuity, allowing for continuous capture of mobility information from the target population. In this study, we used a longitudinal observational design and used mobile big data app technology, specifically structured query language, to gather information on gay dating app users in Shenzhen through the platform as a service (PaaS) and further analyzed the app use patterns, the geographic distribution of app use and HIV testing and counseling (HTC) of the population over a continuous time: from September 2017 to August 2018, leveraging mobile signaling data.

In addition, as the most important traditional festival in China, the Spring Festival (Chinese New Year) is a time for family reunions, cultural celebrations, and various customs. The unique cultural significance makes it an ideal period for examining population mobility as most people travel to their hometowns. Therefore, to better understand the mobility patterns of the app-using men who have sex with men, we analyzed the geographic location changes of the population in Shenzhen before (January 1-31), during (February 1-28), and after (March 1-31) the Spring Festival, and the destination cities where these men who have sex with men traveled during the Spring Festival in 2018.

Participants

According to guidelines for HIV prevention interventions among men who have sex with men from the China CDC, the three most frequently used apps among men who have sex with men are Blued (Blued Inc), Jack'd (Online Buddies Inc), and Zank (Zank Group) [32]. Our preliminary survey also found that approximately 97.2% of men who have sex with men in Shenzhen used Blued [12]. Therefore, all users of the target 3 apps (Blued, Jack'd, or Zank) in Shenzhen were recruited as participants, while only those males aged 15 to 69 years were included for further analysis in this study.

Assessments

The mobile big data collected and the outcomes in this study are included as follows.

- User demographics: Information such as age, sex, and possibly other relevant characteristics of users of Blued, Jack'd, and Zank.
- Geographic location data: Information on the geographic distribution of app users, including district and subdistrict where the apps are accessed during the three defined time periods on weekdays (Monday to Friday): work time (9 AM-6 PM), social time (8 PM-2 AM), and home time (2 AM-6 AM). Additionally, data on the geographic distribution of users' visits to free the HTC sites was collected, including users' visits to city- or district-level (CDC) during the same 3 time periods on weekdays (Monday to Friday), and their visit to Shenzhen 258-Rainbow Working Group (258-RWG) during work time (9 AM-6 PM) on Wednesday and Sunday. 258-RWG is a nongovernmental organization composed of men who have sex with men volunteers, with technical support from Shenzhen CDC. Its initiatives such as outreach, behavioral interventions, peer education, and voluntary counseling and testing to promote prevention and mitigate high-risk behaviors. Both 258-RWG and Shenzhen CDC are the key health service providers for men who have sex with men in Shenzhen, offering free HTC, which is crucial for understanding men who have sex with men's engagement with HIV-related services.
- Mobile signaling data: Information on the distribution of app-using men who have sex with men at the subdistrict level in Shenzhen before the Spring Festival (January 1-31, 2018), the destination cities of those who left Shenzhen during the Spring Festival (February 1- 28, 2018), and the distribution of app-using men who have sex with men after the Spring Festival (March 1- 31, 2018).

Data Sources

The collection and preliminary statistics of the mobile big data were carried out by China Mobile Communications Corporation (CMCC) according to a signed cooperation agreement. The mobile big data on app user, geographic location, and mobile signaling were obtained from the PaaS platform and stored and analyzed using the Hadoop computing framework in the cloud computing environment of CMCC. Preliminary aggregated statistics, such as app user numbers, the geographical distribution of app use and HTC at subdistrict scale across different time periods, and the distribution of destination cities during the Spring Festival in 2018 were exported from the PaaS platform for further statistical analysis. The male population data at the subdistrict level were obtained from the Shenzhen Bureau of Statistics.

The GIS data (GeoJSON format, Geospatial Standard GS {2024} 0650) for the map of China were downloaded from the National Platform for Common Geospatial Information Services. Areas with extremely sparse populations, such as Guangming Farmland and Inner Lingding Island, were excluded from this study and were not displayed on the Map of Shenzhen, as they have minimal impact on the overall distribution and population mobility patterns.

Study Size

Given that mobile big data can potentially obtain information of all target app users, this study encompassed the entire target population, assuming all gay dating app users are generally men who have sex with men. CMCC counted all users of the 3 target apps (Blued, Jack'd, or Zank) in Shenzhen using unique phone numbers, allowing effective deduplication. Although dual SIM

XSL•FO
RenderX

cards enable 2 phone numbers to be used on a single phone, such instances are relatively rare and unlikely to significantly impact the results given the large population size.

Data Analysis

Population Estimates

The total population of men who have sex with men was estimated by combining the percentage (P1) of the target 3 apps among all apps of this kind based on surveillance data, the market share (P2) of CMCC, and the known number of the men who have sex with men users (S) of the target 3 apps (Blued, Jack'd, or Zank) from the survey. The total population of men who have sex with men (N) was then estimated using the formula.



where S is the number of users of Blued, Jack'd, and Zank in this survey; P_1 is the percentage of users of Blued, Jack'd, and Zank among all similar apps based on surveillance data; and P_2 is the market share of CMCC, representing the proportion of users in Shenzhen who are served by this telecom provider.

Spatial Analysis

Based on the acquired geographic location data at the subdistrict scale, ArcGIS (version 10.7; Esri), a professional mapping software for spatial geographic information systems, was used to create choropleth thematic maps of the spatial distribution of the population of men who have sex with men during different time periods. Spatiotemporal autocorrelation analysis was conducted, along with the identification of spatial hot spots in the distribution of app-using men who have sex with men across these time periods. In addition, the distribution of their visits to CDCs and the 258-RWG during different time periods was compared with the distribution of app-using at the subdistrict level.

As one of the most widely used spatial autocorrelation measures, Global Moran I was used to identify spatial autocorrelation. The Global Moran I index was used to assess the spatial autocorrelation of the app use and HTC of the app-using men who have sex with men, with value ranges from –1 to 1, where approximately 0-1 indicates clustering; 0=randomness; and approximately –1 to 0=dispersion. Meanwhile, the High/Low Clustering (Getis-Ord General G) tool was applied to measure the clustering of high or low values of the app using population. The Getis Ord Gi statistic, an index of local spatial autocorrelation, was also used to identify statistically significant spatial clusters of high values (hot spots) and low values (cold spots).

In addition, we used Geodetector software (Chinese Academy of Sciences) to measure and attribute the spatially stratified heterogeneity (SSH) feature of gay dating app use in Shenzhen, characterizing the differences both within and between geographic strata. Geodetector uses the *q*-statistic index to quantify SSH and to assess the interaction outcomes between one or more independent factors. A q value of 0 represents an absence of SSH, while a q value of 1 represents perfect SSH between strata.

Statistical Analysis

Pearson chi-square test and Kruskal-Wallis rank sum test were used to assess the differences in age distribution and median age among users of the 3 apps. The post hoc Dunn test was further applied to perform pairwise comparisons across the 3 apps. All statistical analyses were performed using R software (version 4.4.1; R Foundation for Statistical Computing).

Ethical Considerations

Given that only aggregated data was exported from the PaaS platform for further spatial and statistical analysis, no private information was collected, and no informed consent was required. Privacy protection was fully guaranteed. In addition, this research was reviewed and approved by the Medical Ethics Committee of the Shenzhen Center for Disease Control and Prevention (review number R2017010).

Result

Population of App-Using Men Who Have Sex With Men

From September 2017 to August 2018, a total of 158,387 people used one of the 3 apps: Blued, Jack'd, and Zank, with the majority using Blued (n=117,595, 74.25%) and Jack'd (n=39,040, 24.65%; Table 1). According to our surveillance data, 98.2% of men who have sex with men who reported using a GSN app used one of the 3 target apps, suggesting the population of users collected from these three apps is relatively representative of the entire app-using population of men who have sex with men. Considering that China Mobile Limited held a 60% share of the personal mobile market in 2018 [33], the population of app-using men who have sex with men in Shenzhen was estimated to be 268,817 during 2017-2018, using the formula outlined in the Methods section.

Table 1. Population of app users in Shenzhen from September 2017 to August 2018 (N=158,387).

Gay dating app use	Users, n (%)
Blued	117,595 (74.25)
Jack'd	39,040 (24.65)
Zank	79 (0.05)
Blued + Jack'd	1468 (0.93)
Blued + Zank	197 (0.12)
Jack'd + Zank	1 (0.00)
Blued + Jack'd + Zank	7 (0.004)

XSL•FO
RenderX

Characteristics of App Use

Of the 67,050 app-using men who have sex with men with age information available, the median age was 30 (IQR 25-38) years (Table 2). There was a significant difference in the median age of users across the 3 apps (χ^2_2=4.97; *P*<.001). The post hoc Dunn test revealed a significant difference in median age between Blued and Jackd (*P*<.001), but no significant differences were observed between Blued and Zank (*P*=.41), Zank and Jackd (*P*=.07). Additionally, the age distribution among users of the 3 apps was significantly different (χ^2_6=534.4; *P*<.001). Among the male population aged 15-69 years, the overall app user-to-male ratio was approximately 2.58%, with the highest ratio (3.31%) observed among those aged 25-34 years, and the lowest ratio (1.55%) observed among older adults aged 50-69 years.

The seasonal distribution of app use among men who have sex with men in Shenzhen within the study period (Figure 1) showed a slightly higher percentage of app use during summer and autumn compared to other seasons. For the specific app use, no significant seasonal trend was observed for Blued. However, a seasonal trend was observed for Jack'd and Zank. Specifically, Jack'd use was higher in the summer than in other seasons, whereas Zank use peaked in the autumn relative to other seasons.

Table 2. Age distribution of app-using men who have sex with men in Shenzhen from September 2017 to August 2018.

Age	Blued	Jack'd	Zank	Any app	Estimated users	Male population, n (%)	App user (male), %
All	46,969	20,738	181	67,050	158,387	6,129,495	2.58
15-24 years, n (%)	8971 (19.10)	5228 (25.21)	19 (10.50)	14088 (21.01)	33,279 (21.01)	1,715,435 (27.99)	1.94
25-34 years, n (%)	22,388 (47.67)	8180 (39.44)	92 (50.83)	30,191 (45.03)	71,318 (45.03)	2,157,632 (35.20)	3.31
35-49 years, n (%)	13,586 (28.93)	6563 (31.65)	65 (35.91)	19,998 (29.83)	47,240 (29.83)	1,833,102 (29.91)	2.58
50-69 years, n (%)	2024 (4.31)	767 (3.70)	5 (2.76)	2773 (4.14)	6550 (4.14)	423,326 (6.91)	1.55
Median (IQR)	30 (26-37)	30 (24-38)	31 (27-38)	30 (25-38)	—[a]	—	—

[a]Data not available.

Figure 1. Seasonal distribution of app-using men who have sex with men in Shenzhen from September 2017 to August 2018.

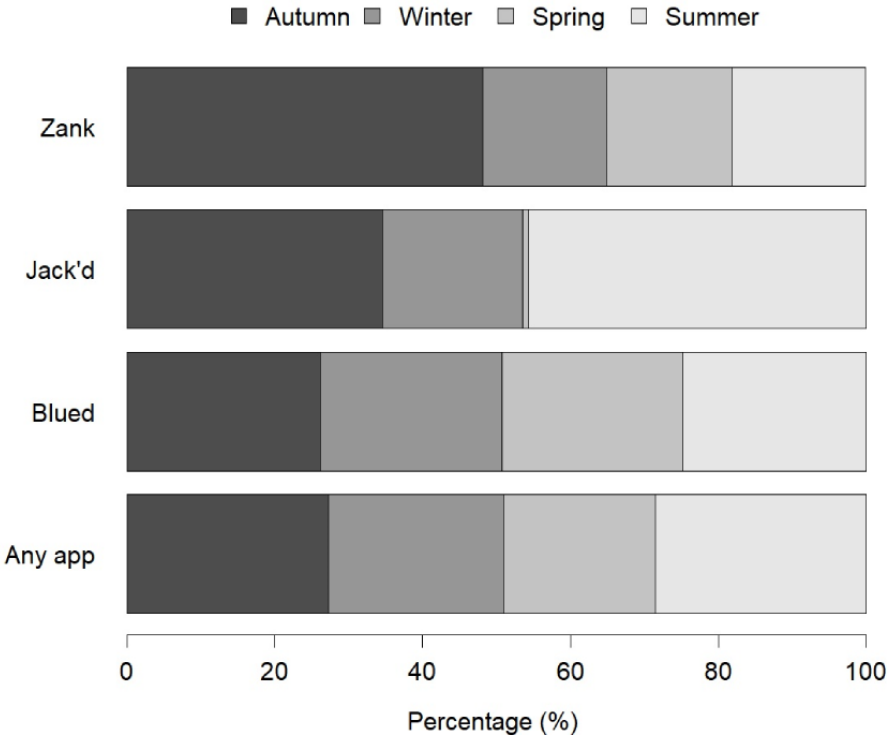

XSL•FO
RenderX

Distribution and Hot Spots of App Use

The geographic distribution of the gay dating app–using population during the 3 time periods was mapped at a subdistrict scale (Figure 2A). The population was relatively greater during work time (n=152,972) and social time (n=152,521) compared to home time (n=142,719). For work time, the largest population of app use was observed in the Futian, Yuehai, and Nanhu subdistricts (more than 7000), followed by Minzhi, Longhua, and Xixiang (n=5000-7000). During the social time, the highest population was observed again in the Futian subdistrict (more than 7000), followed by Minzhi, Nanhu, Xixiang, Longhua, and Buji subdistricts (n=5000-7000). For home time, the highest population was located in Xixiang, Minzhi, Futian, Longhua, Buji, and Nanhu subdistricts (n=5000-7000).

The Moran indexes derived from spatial autocorrelation analysis were all above 0 and statistically significant for the 3 time periods, suggesting the geographic distribution of app-using men who have sex with men in Shenzhen exhibited a clustered pattern (Table 3). Furthermore, the High/Low Clustering (Getis-Ord General G) statistics demonstrated significant high-high clustering for all 3 time periods (before: *z* score=3.426; *P*=.001; during *z* score=3.072; *P*=.002; after: *z* score=2.717; *P*=.007), indicating areas with high values of app-using population were spatially clustered.

Hot spot analysis identified Yuehai and Futian as the primary hot spots for app-using populations during work time, each with a Gi_Bin Score of 3 (*P*=.004 and *P*=.002). Meanwhile, Nanhu and Minzhi were also recognized as hot spots with a Gi_Bin Score of 2 (*P*=.01 and *P*=.04; Figure 2B; Multimedia Appendix 1). During social time, Futian, Minzhi, Nanhu, and Xixiang were identified as hot spots, each with a Gi_Bin Score of 2 (*P*=.01, *P*=.02, *P*=.02, and *P*=.02). Subsequently, for home time, Xixiang, Minzhi, Futian, and Longhua were recognized as hot spots, each with a Gi_Bin Score of 2 (*P*=.01, *P*=.02, *P*=.02, and *P*=.03) for home time.

XSL•FO
RenderX

Figure 2. Geographic distribution and hot spots of app-using prevalence by time periods.

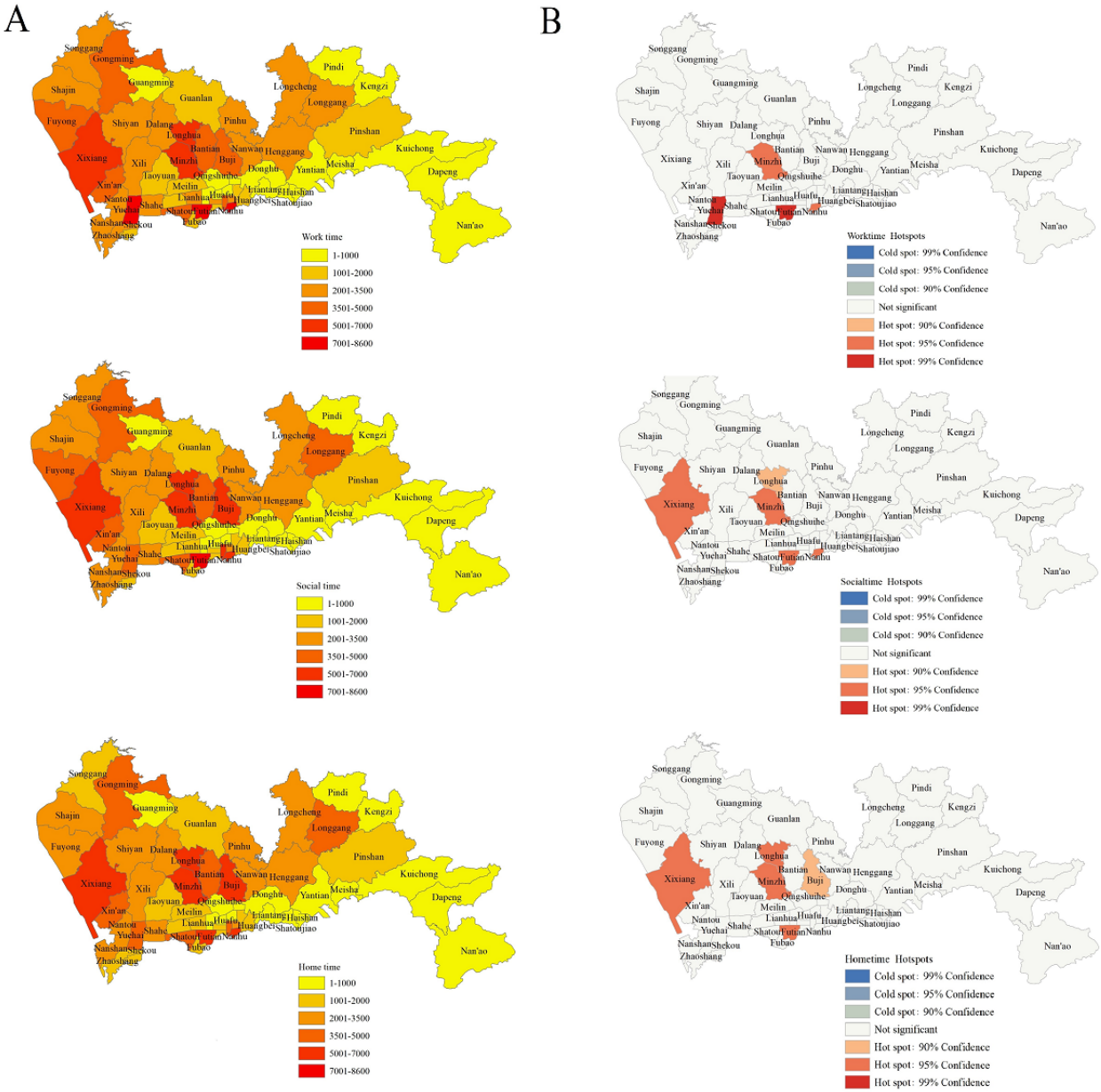

Table 3. Moran indexes of app use by time periods in Shenzhen (from September 2017 to August 2018).

Statistics	Home time (2 AM-6 AM)	Social time (8 PM-2 AM)	Work time (9 AM-6 PM)
Moran index	0.272	0.238	0.190
z score	3.210	2.837	2.338
P value	.001	.005	.02
Observed General G	0.022	0.021	0.021
Expected General G	0.018	0.018	0.018
z score	3.426	3.072	2.717
P value	.001	.002	.007

Distribution and Hot Spots of HTC

The overall prevalence of HTC among app-using men who have sex with men aged 15-69 years in Shenzhen was 17.1%, with higher rates observed during home time (18.9%) compared to social time (17.8%) and work time (17.7%). The geographic distribution of HTC prevalence was uneven at the subdistrict scale in Shenzhen (Figure 3A), with higher prevalence reported in Cuizhu, Xiangmihu, Haishan, Nanshan, and Huangbei

XSL•FO
RenderX

subdistricts (>36%) during work time. For social time, a higher prevalence was observed in Cuizhu subdistrict (>48%), followed by Haishan, Shatoujiao, Liantang, Dongxiao, Huangbei, and Xiangmihu subdistricts (>36%). A similar distribution was noted for home time with the addition of Kuicheng subdistrict (>36%). The Moran indexes were statistically significant across all 3 time periods (Table 4). The High/Low Clustering (Getis-Ord General G) statistics indicated the distribution of high values of HTC prevalence was spatially clustered for both home time and social time, but not significant for work time. Hot spot analysis of the HTC prevalence revealed most of the hot spots were concentrated in the Cuizhu subdistrict with a Gi_Bin Score of 3 (P=.001), as well as in Xiangmihu and Haishan subdistricts, with Gi_Bin Score of 2 (P=.02 and P=.03) during the work time (Figure 3B; Multimedia Appendix 2). Similarly, the hot spots of the HTC prevalence during social time and home time were predominantly clustered in the Cuizhu subdistrict with a Gi_Bin Score of 3 (P=.002 and P=.002), and in the Haishan subdistrict with a Gi_Bin Score of 2 (P=.04 and P=.04; Figure 3B).

Figure 3. Geographic distribution and hot spots of HTC prevalence by time periods. HTC: HIV testing and counseling.

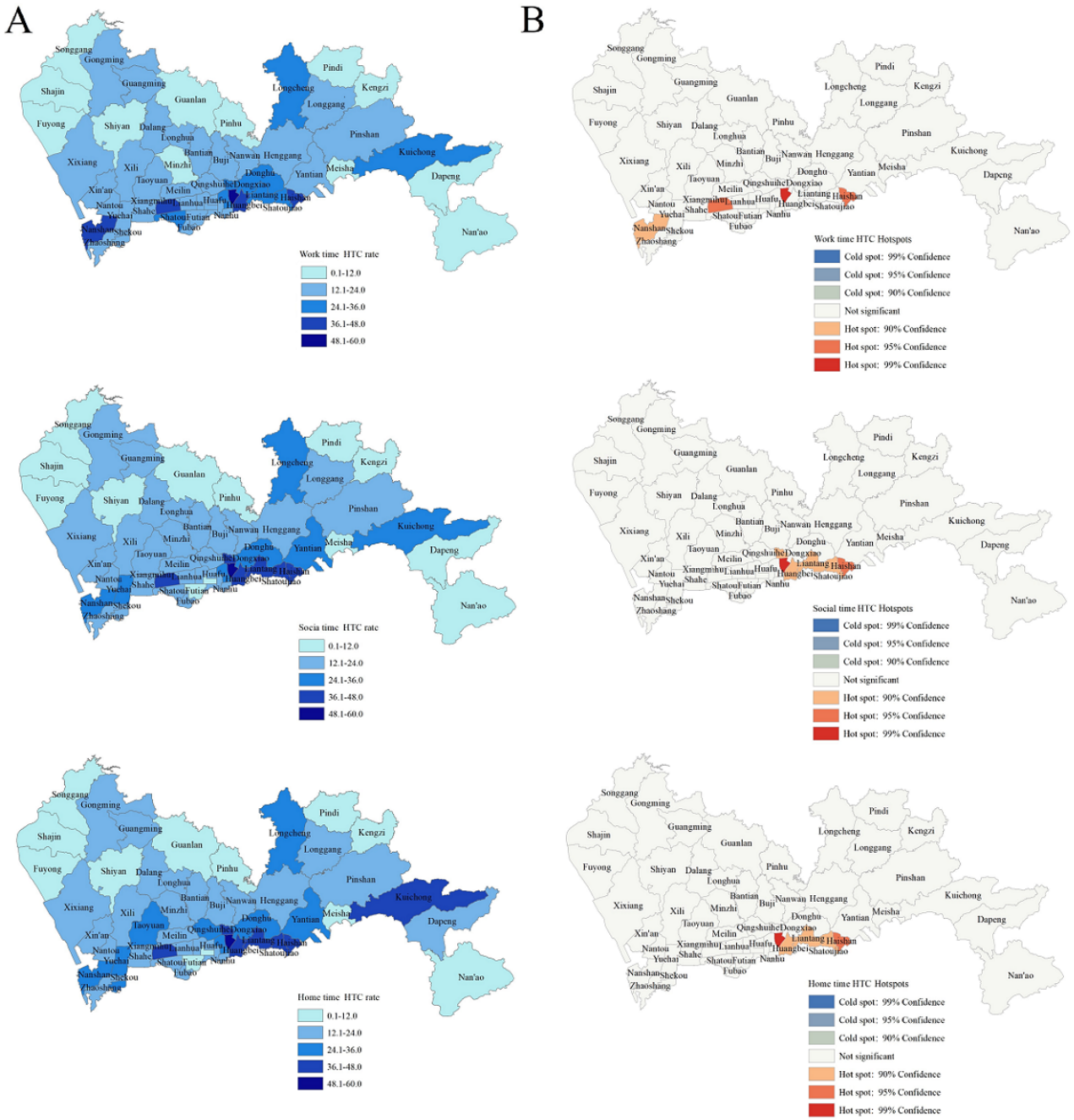

XSL•FO
RenderX

Table 4. Moran indexes of HTCa prevalence among app-using men who have sex with men in Shenzhen (from September 2017 to August 2018).

Statistics	Home time (2 AM-6 AM)	Social time (8 PM-2 AM)	Work time (9 AM-6 PM)
Moran index	0.456	0.450	0.345
z score	5.255	5.200	4.060
P value	<.001	<.001	<.001
Observed General G	0.020	0.020	0.019
Expected General G	0.018	0.018	0.018
z score	3.106	3.204	2.370
P value	.002	.001	.02

[a]HTC: HIV testing and counseling.

SSH of App Use and HTC

The Geodetector index of *q* statistics was also used to measure the stratified heterogeneity of app use and HTC across different strata (time period, district, subdistrict, and the proportion of non-Hukou population; Table 5). For both the prevalence of app use and HTC, the *q* statistics and *P* values indicated significant heterogeneity in administrative districts (*P*=.006 and *P*<.001), subdistricts (*P*<.001 and *P*<.001), and the percentage of non-Hukou migrants (*P*<.001 and *P*<.001). However, no significant heterogeneity was observed across different time periods. Luohu district demonstrated the highest risk for both app use and HTC. Nanhu subdistrict, along with subdistricts with 50%-60% non-Hukou migrants, exhibited a higher risk for app use than other areas. Meanwhile, the Cuizhu subdistrict and those subdistricts with 40%-50% non-Hukou migrants reported the highest rates for HTC.

Table 5. Stratified heterogeneity of app use and HTCa in Shenzhen (from September 2017 to August 2018).

Factors	*q* statistic	*P* value	Highest risk detector
App use			
Time period	0.004	.71	—[b]
Subdistrict	0.949	<.001	Nanhu
District	0.147	.006	Luohu
Non-Hukou migrants (%)	0.148	<.001	L6 (50%-60%)
HTC			
Time period	0.002	.80	—
Subdistrict	0.958	<.001	Cuizhu
District	0.420	<.001	Luohu
Non-Hukou migrants (%)	0.378	<.001	L5 (40%-50%)

[a]HTC: HIV testing and counseling.

[b]Data not applicable.

Population Mobility

The geographical distribution of app-using men who have sex with men in Shenzhen before, during, and after the Spring Festival of 2018 was mapped (Multimedia Appendix 3). The results showed that prior to the Spring Festival, app-using men who have sex with men were mostly located in Xixiang, Yuehai, and Longhua subdistricts. During the Spring Festival, those who stayed in Shenzhen were relatively concentrated in Xixiang, Longhua, and Minzhi subdistricts. After the Spring Festival, app-using men who have sex with men returned to clustered in Xixiang, Yuehai, Longhua, and Minzhi subdistricts. For the population change of app-using men who have sex with men before and after the Spring Festival, the largest inflow occurred in the Yuehai, Shatou, and Longhua subdistricts, while the most significant outflow was observed in the Longcheng and Donghu subdistricts (Figure 4A). Additionally, the population mobility of app-using men who have sex with men during the Spring Festival was mapped based on changes in geographical location at the subdistrict scale (Figure 4B). During the Spring Festival, 60,202 (38%) app-using men who have sex with men left Shenzhen, while 37,756 (62.7%) of them returned after the holiday. The distribution of the destination cities for these app-using men who have sex with men demonstrated a relatively centralized flow to other cities in China, with the largest mobility directed toward other cities within Guangdong province (67.7%). There was also lower mobility to Hunan province (7.9%) and other adjacent provinces, such as Jiangxi, Guangxi, and Hubei Provinces (approximately 3%-5%). In terms of age distribution, a higher proportion of those who left Shenzhen were aged 25-34 years and 35-49 years. In contrast, those who stayed were predominantly younger men who have sex with men aged 24

XSL•FO
RenderX

years or younger, as well as older adults aged 50-69 years (P<.001; Figure 5).

Figure 4. Mobility of app-using men who have sex with men in Shenzhen during the Spring Festival in 2018.

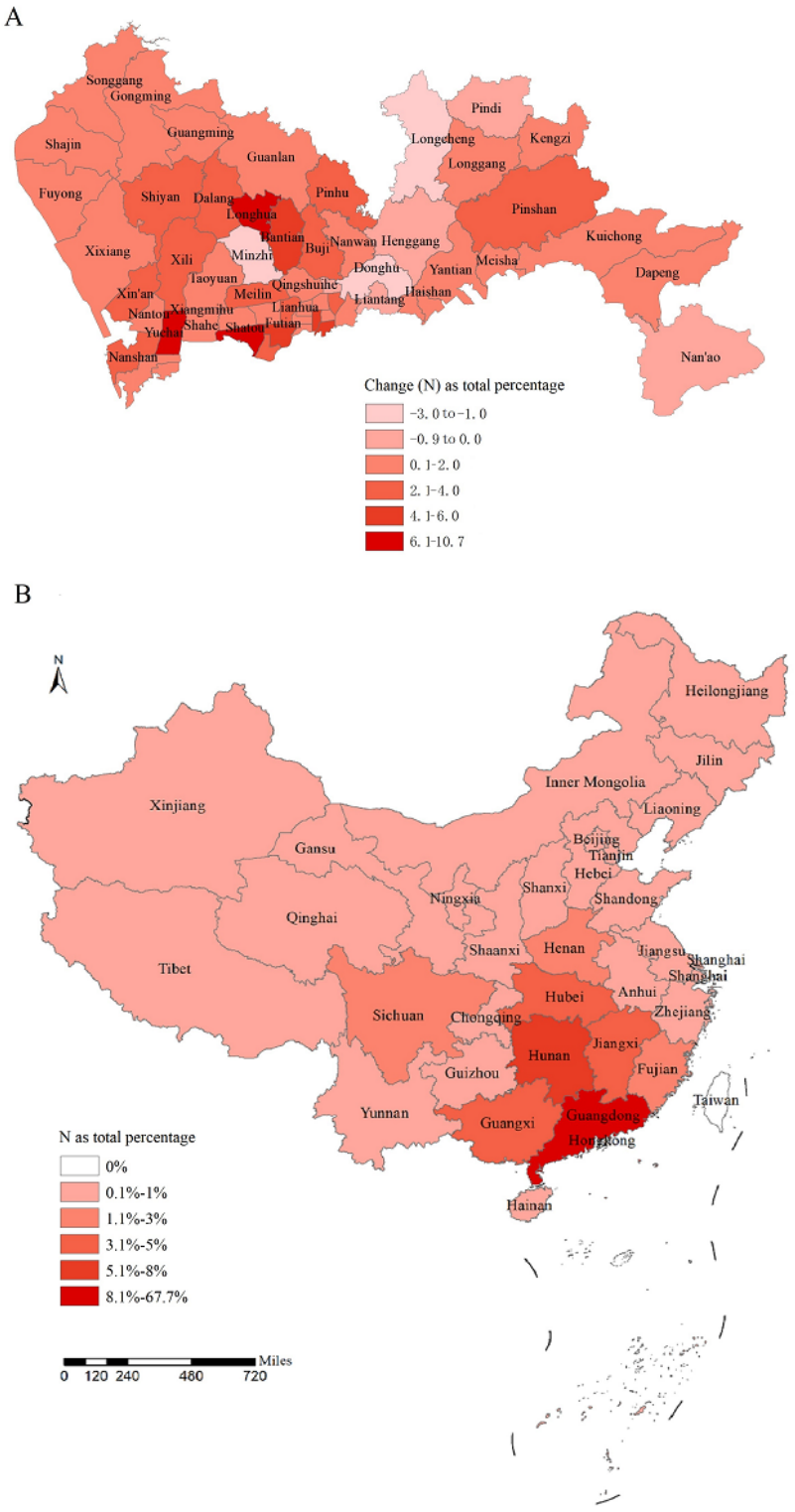

XSL•FO
RenderX

Figure 5. Age distribution of app-using men who have sex with men in Shenzhen during the Spring Festival in 2018.

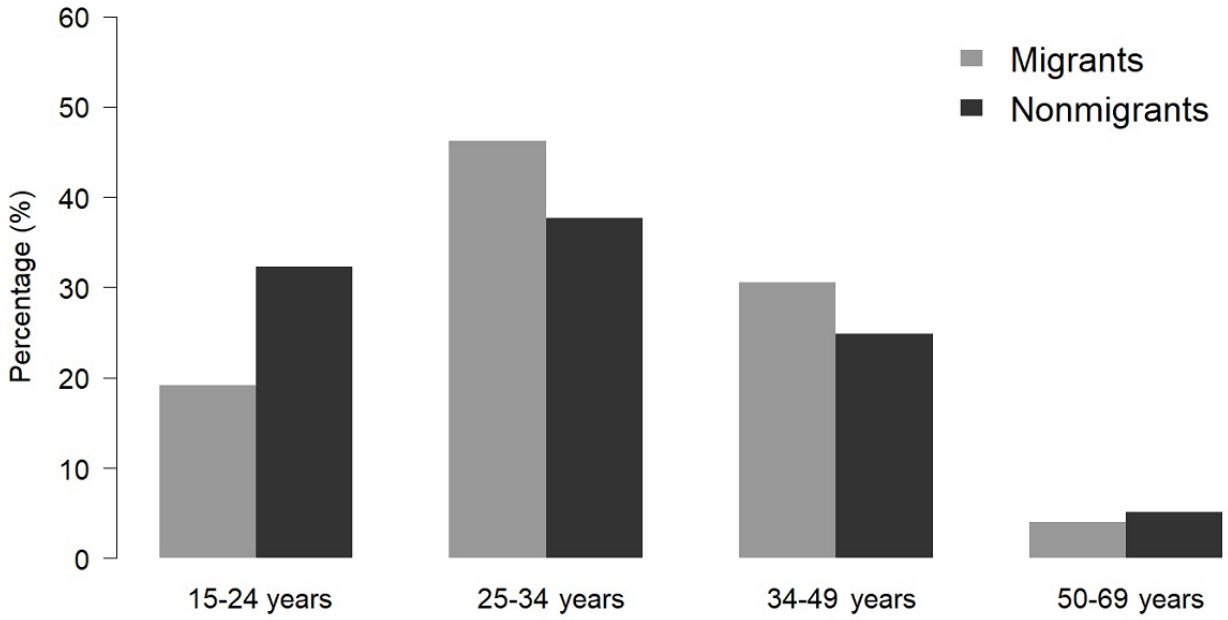

Discussion

Principal Findings

This study represents the first longitudinal observational research on the GSN app use, examining its genuine population size, spatiotemporal and behavioral patterns, and mobility through mobile big data in mainland China. This study found Shenzhen has a larger population of men who have sex with men than expected. The geographic distribution of app-using men who have sex with men in Shenzhen exhibited a clustered pattern in central and western Shenzhen, while their HTC uptake was more concentrated in central-eastern Shenzhen. Moreover, both relatively high internal and interprovincial mobility was observed among those app-using men who have sex with men. The findings provided a comprehensive view of the behaviors and dynamics of app-using men who have sex with men, offering valuable insights to inform enhanced prevention and intervention strategies and guide resource allocation. Moreover, our research demonstrates the potential of mobile big data to address critical gaps often overlooked by traditional research methods, providing more accurate, timely, and comprehensive insights into public health and disease prevention.

App Use

This study implied a large population of app-using men who have sex with men in Shenzhen from September 2017 to August 2018, surpassing previous estimates [31]. Moreover, the annual estimate of app-user or male ratio (2.58% among males aged 15-69 years) was relatively high in Shenzhen, compared to the daily estimates of Beijing and Shanghai (0.82% and 0.56% among males aged 18-64 years) from a concurrent experimental study [28]. Literature implied that overestimation might exist in economically developed countries characterized by higher urbanization, higher education levels, more in-migrants, and larger working-age populations [34]. Nonetheless, the population of men who have sex with men may have continued to grow along with the growth of the general population in Shenzhen since 2018, underscoring the increasing need for HIV prevention and intervention efforts targeting men who have sex with men in the region.

Our previous surveillance data indicated a rapid increase in app use among men who have sex with men in Shenzhen since 2016, with Blued being the most frequently used app, consistent with findings from previous data in Shenzhen and other cities [12,26]. An internet survey suggested men who have sex with men in southern China used Jack'd more frequently [35]. Founded in 2010, Jack'd has experienced significant growth, particularly among young men who have sex with men, boasting over 5 million users worldwide since 2014 [36]. Notably, nearly half of the app users were aged between 25-34 years, which aligns with other internet surveys [8,14,35]. In Shenzhen, the median age of new infections was 31 years in 2011-2019, with cases younger than 30 years more frequently observed within molecular transmission clusters [37]. Thus, tailored strategies are needed to enhance the effectiveness of HIV prevention initiatives targeting these younger men who have sex with men.

Spatiotemporal Distribution and SSH

The temporal distribution pattern of app-using men who have sex with men implied that these men who have sex with men tend to use the apps more frequently during work time (daytime) and social time compared to home time. It is reasonable to assume that app users initially seek potential partners through the internet before transitioning to offline social networking if they find a desirable match from the app. During home time, however, it is likely that most users prefer to rest. This finding suggested a clear temporal distinction: a more active search for connections during work and social hours, and a more relaxed,

XSL•FO
RenderX

less frequent engagement during home time. Regarding the geographical distribution, app-using men who have sex with men in Shenzhen exhibited spatial clustering in several subdistricts (Futian, Yuehai, Nanhu, Xixiang, Minzhi, and Longhua) at different time periods. Most clusterings occurred in Nanshan District for work time, Futian District for social time, and Baoan District for home time. This study, for the first time, revealed the spatial heterogeneity for gay dating app–using men who have sex with men in Shenzhen, suggesting that these men who have sex with men were unevenly distributed both at the subdistricts and district levels, as well as among the Hukou residents. As app-using men who have sex with men were more likely to have multiple sexual partners, practice receptive roles in anal sex, and use recreational drugs [12]. By recognizing the different contexts in which men who have sex with men engage with dating apps, public health initiatives can be designed to effectively reach this population at the right times and in the right settings. Traditional HIV prevention services (eg, condoms and lubricants distribution) and interventions (eg, antidrug campaigns) should be delivered more precisely to the identified hot spots based on both temporal and geographical distribution.

Meanwhile, compared to the distribution of app-using men who have sex with men, the prevalence of HTC among this population in Shenzhen showed obvious inconsistencies. HTC rates were concentrated in different subdistricts (eg, Cuizhu, Haishan, Xiangmihu, Nanshan, Shatoujiao, and Huangbei), with the highest clustering observed in Luohu, Futian, and Yantian Districts during social and home time. For work time, HTC is also clustered in Nanshan District. These findings implied that men who have sex with men might prefer to seek HTC services in central districts of Shenzhen, which had higher accessibility to health care [38], or choose these areas for convenience after engaging in high-risk sexual behavior during offline dating or gatherings.

Given the disparities in the distribution of app use and HTC service, it is crucial to allocate more HIV-related health care services (eg, HIV testing and counseling) to western Shenzhen, particularly in subdistricts of Baoan District. Furthermore, the mismatched distribution of app use and HTC and the availability of fee HTC service highlighted the need for outreach initiatives that extend beyond traditional health care settings. Implementing mobile clinics or pop-up testing stations in hot spots or underserved areas could help bridge the HIV testing gap.

Although our previous study implied that app-using men who have sex with men were more likely to take HIV tests, the prevalence of lifetime and recent HIV tests in the past year (70.2% and 50.2%) remained significantly below the UNAIDS target [39]. Furthermore, it is reported that the majority (40%) of the app-using men who have sex with men preferred to take HIV tests at CDCs and community-based organizations (eg, 258-RWG) due to stigma and confidential concerns [39]. Besides, traditional offline HTC services are also available at gay social venues, voluntary counseling and testing centers, or clinics, while the uptake of these services has not been optimal. Therefore, it is essential to prioritize location-tailored strategies for expanding HIV testing, particularly in identified geographical hot spots. Concurrently, innovative web-based prevention initiatives targeting app-using men who have sex with men emerged as a promising alternative. For instance, Blued launched an HIV testing campaign in 2015, which recommended the nearest health care facility or testing service for app users [27]. In light of these insights, future strategies could leverage GSN apps and smartphone tools to promote HTC services, ultimately contributing to the achievement of the first UNAIDS “95” goal.

Migration During the Spring Festival

During the 2018 Spring Festival, more than one-third of app-using men who have sex with men left Shenzhen, highlighting the high mobility of this population. A prior study also reported that about one-fourth of the gay dating app-using men who have sex with men lived in Shenzhen for less than 1 year [12]. The distribution of the destination cities for app-using men who have sex with men demonstrated a relatively centralized flow, with greater internal movement within Guangdong province and less interprovincial mobility to neighboring provinces such as Hunan, Jiangxi, Guangxi, and Hubei. Research suggested that immigrants from Jiangxi, Shaanxi, and Hubei provinces were the main sources of the predominant CRF01_AE subtype in Shenzhen [18]. Local statistics also reported that the immigrants in Shenzhen mainly came from Guangdong (35%), followed by Hunan, Sichuan, Guangxi, Hubei, and Jiangxi. A nationwide study further supported these findings, identifying Hunan, Guangxi, Hubei, and Jiangxi Provinces as the main source provinces for migrant men who have sex with men in Guangdong [28]. During the Spring Festival, the most significant changes in the population of men who have sex with men were observed in the Yuehai, Shatou, and Longhua subdistricts. Therefore, it is crucial to strengthen targeted and innovative prevention and intervention efforts in these areas to address the unique challenges faced by migrant men who have sex with men, such as app-based notifications or location-based services, to provide timely and targeted HIV prevention messages and services, particularly during periods of high mobility.

Limitations

First, the application of mobile internet has developed rapidly in China, especially in the last 5 years, which may have a significant impact on the behavioral characteristics of men who have sex with men, it would be valuable to collect more recent data to examine the evolution of the trend in the future. Second, as the study is based on mobile big data, there may be inherent biases in estimating the true population size. Third, regarding the geographical difference in app use and HTC, limited demographic data prevented us from differentiating the impact of resource availability from user selection bias. Moreover, it is difficult to distinguish male sex workers and noncommercial men who have sex with men, who may exhibit different HIV-related behaviors [40]. Therefore, future studies with more demographic factors (eg, occupation and income) and other socioeconomic determinants, are needed to differentiate selection bias. Finally, although gay dating app users are generally assumed to be men who have sex with men, it is possible that a small number of non–men who have sex with men individuals may also use the app for research purposes or by chance. However, as this group constitutes a negligible

XSL•FO
RenderX

portion of the overall population, their presence is unlikely to affect the results significantly and can therefore be disregarded.

Conclusions

Shenzhen has a large population of men who have sex with men using gay dating apps, predominantly aged 15-34 years. The uneven spatiotemporal distribution of app-using men who have sex with men and their HIV testing or counseling, as well as the mismatch between them, emphasizes the need to adapt the traditional venue-based prevention and intervention strategies to the identified hot spots area and the need for outreach initiatives that extend beyond traditional health care settings. Given the relatively high internal and interprovincial mobility of this population, leveraging smartphone-based behavioral monitoring could provide valuable insights for enhanced and innovative prevention and intervention strategies.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (82373651), the SanMing Project of Medicine in Shenzhen (SZSM202311015), the Shenzhen Municipal Technological Project (JCYJ20200109150715644, JCYJ20170306160440762), and the Guangdong Medical Research Foundation (A2023350).

Data Availability

The datasets used during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LW, JC, and JZ conceived the work. JZ, LW, and YW designed the study. LW conducted all data analyses and drafted the manuscript. YW and LC participated in data collection and management. LW, YW, LC, and JZ contributed to writing and data interpretation. LW, JZ, and JC reviewed and edited the manuscript. All authors have seen and approved the final version of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hot spots of app-using population by time periods in Shenzhen (September 2017-August 2018).
[DOCX File , 18 KB - jmir_v27i1e69569_app1.docx]

Multimedia Appendix 2
Hot spots of HTC prevalence by time periods in Shenzhen (September 2017-August 2018).
[DOCX File , 15 KB - jmir_v27i1e69569_app2.docx]

Multimedia Appendix 3
Geographical distribution of app-using MSM in Shenzhen (A) before, (B) during, and (C) after the Spring Festival of 2018.
[PNG File , 435 KB - jmir_v27i1e69569_app3.png]

References

1. 2024 global AIDS update. UNAIDS. 2024. URL: https://crossroads.unaids.org/ [accessed 2024-10-24]
2. Han M. Analysis of the epidemic situation of AIDS in China and prospects for prevention and treatment. Chin J AIDS STD 2023;29(3):247-250. [doi: 10.13419/j.cnki.aids.2023.03.01]
3. He N. Research Progress in the Epidemiology of HIV/AIDS in China. China CDC Wkly 2021 Nov 26;3(48):1022-1030 [FREE Full text] [doi: 10.46234/ccdcw2021.249] [Medline: 34888119]
4. In danger: UNAIDS global AIDS update 2022. UNAIDS. 2022. URL: https://www.unaids.org/en/resources/documents/2022/in-danger-global-aids-update [accessed 2024-10-24]
5. Wu Z. [Achievement of HIV/AIDS program in the past 30 years and challenges in China]. Zhonghua Liu Xing Bing Xue Za Zhi 2015 Dec;36(12):1329-1331. [Medline: 26850382]
6. UNAIDS data 2021. UNAIDS. URL: https://www.unaids.org/en/resources/documents/2021/2021_unaids_data [accessed 2024-10-25]
7. Landovitz RJ, Tseng CH, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]

XSL•FO
RenderX

8. Tang W, Best J, Zhang Y, Liu F, Tso LS, Huang S, et al. Gay mobile apps and the evolving virtual risk environment: a cross-sectional online survey among men who have sex with men in China. Sex Transm Infect 2016 Nov;92(7):508-514 [FREE Full text] [doi: 10.1136/sextrans-2015-052469] [Medline: 27288414]
9. He A, Bu H, Mu W, Fu D, Chen C. Understanding Motivations for Dating App Use Among Gay and Bisexual Men: Validation, Latent Profile Analysis, and Differences in Health Outcomes. Arch Sex Behav 2024 Feb;53(2):555-568. [doi: 10.1007/s10508-023-02746-9] [Medline: 38038855]
10. Miao W, Chan LS. Domesticating gay apps: an intersectional analysis of the use of Blued among Chinese gay men. J Comput Mediat Commun 2021;26(1):38-53 [FREE Full text]
11. Hong H, Xu J, McGoogan J, Dong H, Xu G, Wu Z. Relationship between the use of gay mobile phone applications and HIV infection among men who have sex with men in Ningbo, China: a cross-sectional study. Int J STD AIDS 2018;29(5):491-497. [doi: 10.1177/0956462417738468] [Medline: 29099328]
12. Wei L, Chen L, Zhang H, Yang Z, Zou H, Yuan T, et al. Use of gay app and the associated HIV/syphilis risk among non-commercial men who have sex with men in Shenzhen, China: a serial cross-sectional study. Sex Transm Infect 2019;95(7):496-504. [doi: 10.1136/sextrans-2018-053902] [Medline: 30918120]
13. Wang H, Zhang L, Zhou Y, Wang K, Zhang X, Wu J, et al. The use of geosocial networking smartphone applications and the risk of sexually transmitted infections among men who have sex with men: a systematic review and meta-analysis. BMC Public Health 2018 Oct 16;18(1):1178 [FREE Full text] [doi: 10.1186/s12889-018-6092-3] [Medline: 30326887]
14. Zou H, Fan S. Characteristics of Men Who Have Sex With Men Who Use Smartphone Geosocial Networking Applications and Implications for HIV Interventions: A Systematic Review and Meta-Analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
15. McGrath N, Eaton JW, Newell M, Hosegood V. Migration, sexual behaviour, and HIV risk: a general population cohort in rural South Africa. Lancet HIV 2015;2(6):e252-e259 [FREE Full text] [doi: 10.1016/S2352-3018(15)00045-4] [Medline: 26280016]
16. Weine SM, Kashuba AB. Labor migration and HIV risk: a systematic review of the literature. AIDS Behav 2012;16(6):1605-1621. [doi: 10.1007/s10461-012-0183-4] [Medline: 22481273]
17. Thorp M, Bellos M, Temelkovska T, Mphande M, Cornell M, Hubbard J, et al. Mobility and ART retention among men in Malawi: a mixed-methods study. J Int AIDS Soc 2023;26(3):e26066 [FREE Full text] [doi: 10.1002/jia2.26066] [Medline: 36943753]
18. An M, Zheng C, Li H, Chen L, Yang Z, Gan Y, et al. Independent epidemic patterns of HIV-1 CRF01_AE lineages driven by mobile population in Shenzhen, an immigrant city of China. Virus Evol 2021 Sep;7(2):veab094 [FREE Full text] [doi: 10.1093/ve/veab094] [Medline: 35299786]
19. Elliot E, Rossi M, McCormack S, McOwan A. Identifying undiagnosed HIV in men who have sex with men (MSM) by offering HIV home sampling via online gay social media: a service evaluation. Sex Transm Infect 2016 Sep;92(6):470-473. [doi: 10.1136/sextrans-2015-052090] [Medline: 27221475]
20. Rosengren AL, Huang E, Daniels J, Young SD, Marlin RW, Klausner JD. Feasibility of using GrindrTM to distribute HIV self-test kits to men who have sex with men in Los Angeles, California. Sex Health 2016;13:389-392 [FREE Full text] [doi: 10.1071/SH15236] [Medline: 27209076]
21. Young SD, Jaganath D. Online social networking for HIV education and prevention: a mixed-methods analysis. Sex Transm Dis 2013 Feb;40(2):162-167 [FREE Full text] [doi: 10.1097/OLQ.0b013e318278bd12] [Medline: 23324979]
22. Delaney KP, Kramer MR, Waller LA, Flanders WD, Sullivan PS. Using a geolocation social networking application to calculate the population density of sex-seeking gay men for research and prevention services. J Med Internet Res 2014 Nov 18;16(11):e249 [FREE Full text] [doi: 10.2196/jmir.3523] [Medline: 25406722]
23. Abuelezam NN, Reshef YA, Novak D, Grad YH, Seage Iii GR, Mayer K, et al. Interaction Patterns of Men Who Have Sex With Men on a Geosocial Networking Mobile App in Seven United States Metropolitan Areas: Observational Study. J Med Internet Res 2019 Sep 12;21(9):e13766 [FREE Full text] [doi: 10.2196/13766] [Medline: 31516124]
24. Pan S, Xu J, Han X, Zhang J, Hu QH, Chu ZX, et al. Internet-Based Sex-Seeking Behavior Promotes HIV Infection Risk: A 6-Year Serial Cross-Sectional Survey to MSM in Shenyang, China. Biomed Res Int 2016;2016:2860346 [FREE Full text] [doi: 10.1155/2016/2860346] [Medline: 28105415]
25. Guo Z, Feng A, Zhou Y, Gao Y, Sun Y, Chen Y, et al. Geosocial networking mobile applications use and HIV and other sexually transmitted infections among men who have sex with men in Southern China: A cross-sectional study. Front Public Health 2023;11:1063993 [FREE Full text] [doi: 10.3389/fpubh.2023.1063993] [Medline: 36844866]
26. Fan S, Li P, Hu Y, Gong H, Yu M, Ding Y, et al. Geosocial Networking Smartphone App Use and High-Risk Sexual Behaviors Among Men Who Have Sex With Men Attending University in China: Cross-sectional Study. JMIR Public Health Surveill 2022 Mar 28;8(3):e31033 [FREE Full text] [doi: 10.2196/31033] [Medline: 35343911]
27. Dai Z, Mi G, Yu F, Chen G, Wang X, He Q. Using a Geosocial Networking App to Investigate New HIV Infections and Related Risk Factors Among Student and Nonstudent Men Who Have Sex With Men in Chengdu, China: Open Cohort Study. J Med Internet Res 2023 Jul 28;25:e43493 [FREE Full text] [doi: 10.2196/43493] [Medline: 37505891]

XSL•FO
RenderX

28. Hu M, Xu C, Wang J. Spatiotemporal Analysis of Men Who Have Sex With Men in Mainland China: Social App Capture-Recapture Method. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e14800 [FREE Full text] [doi: 10.2196/14800] [Medline: 32012086]
29. Wang H, Zhang J, Chu Z, Hu Q, Dong W, Huang X, et al. Risk-Taking Behaviors and Adherence to HIV Pre-Exposure Prophylaxis in Users of Geosocial Networking Apps: Real-World, Multicenter Study. J Med Internet Res 2020 Oct 14;22(10):e22388 [FREE Full text] [doi: 10.2196/22388] [Medline: 33052132]
30. [Shenzhen Statistical Yearbook 2017]. Shenzhen Statistics Bureau. 2017. URL: https://tjj.sz.gov.cn/zwgk/zfxxgkml/tjsj/tjnj/content/post_3085969.html [accessed 2024-11-03]
31. Yang Z, Wei L, Xie W, Chen L, Yang Z, Zhang Y, et al. Estimating Changes in Population Size and Behavioral Characteristics in Men Who Have Sex With Men Between 2014 and 2019: Longitudinal Study. JMIR Public Health Surveill 2022 Aug 16;8(8):e34150 [FREE Full text] [doi: 10.2196/34150] [Medline: 35972779]
32. Guidelines for HIV prevention interventions among men who have sex with men. China Center for Disease Control and Prevention. 2016. URL: https://ncaids.chinacdc.cn/xb/sjb/tzgg_10325/201609/W020160913346815395954.pdf [accessed 2024-11-10]
33. 2018 annual results. China Mobile Ltd. 2019. URL: https://www.chinamobileltd.com/en/ir/webcasts/pre190321.pdf [accessed 2024-11-14]
34. Mu X, Zhang X, Yeh AG, Wang J. Evaluating the representativeness of mobile big data: A comparative analysis between China's mobile big data and census data at the county level. Applied Geography 2024 May;166:103260. [doi: 10.1016/j.apgeog.2024.103260]
35. Bien CH, Best JM, Muessig KE, Wei C, Han L, Tucker JD. Gay Apps for Seeking Sex Partners in China: Implications for MSM Sexual Health. AIDS Behav 2015 Jun;19(6):941-946 [FREE Full text] [doi: 10.1007/s10461-014-0994-6] [Medline: 25572834]
36. Jack'd surpasses two million android users leading to a community of over five million. PR Newswire. 2014. URL: https://tinyurl.com/sjkezd7v [accessed 2024-09-09]
37. Zhang D, Zheng C, Li H, Li H, Liu Y, Wang X, et al. Molecular surveillance of HIV-1 newly diagnosed infections in Shenzhen, China from 2011 to 2018. J Infect 2021 Jul;83(1):76-83 [FREE Full text] [doi: 10.1016/j.jinf.2021.04.021] [Medline: 33932447]
38. Tao Z, Yao Z, Kong H, Duan F, Li G. Spatial accessibility to healthcare services in Shenzhen, China: improving the multi-modal two-step floating catchment area method by estimating travel time via online map APIs. BMC Health Serv Res 2018 May 09;18(1):345 [FREE Full text] [doi: 10.1186/s12913-018-3132-8] [Medline: 29743111]
39. Wei L, Chen L, Zhang H, Yang Z, Liu S, Tan W, et al. Relationship between gay app use and HIV testing among men who have sex with men in Shenzhen, China: a serial cross-sectional study. BMJ Open 2019 Aug 24;9(8):e028933 [FREE Full text] [doi: 10.1136/bmjopen-2019-028933] [Medline: 31446409]
40. Zhao J, Cai WD, Gan YX, Zhang Y, Yang ZR, Cheng JQ, et al. A comparison of HIV infection and related risk factors between money boys and noncommercial men who have sex with men in Shenzhen, China. Sex Transm Dis 2012 Dec;39(12):942-948. [doi: 10.1097/OLQ.0b013e31826f356f] [Medline: 23191947]

Abbreviations

258-RWG: 258-Rainbow Working Group
CDC: Center for Disease Control and Prevention
CMCC: China Mobile Communications Corporation
GSN: geosocial networking
HTC: HIV testing and counseling
PaaS: platform as a service
PrEP: pre-exposure prophylaxis
SSH: spatially stratified heterogeneity

Edited by T de Azevedo Cardoso; submitted 03.12.24; peer-reviewed by Y Gao, W Tang, S Chen; comments to author 13.12.24; revised version received 20.12.24; accepted 29.01.25; published 20.03.25.

Please cite as:
Wei L, Wu Y, Chen L, Cheng J, Zhao J
Spatiotemporal and Behavioral Patterns of Men Who Have Sex With Men Using Geosocial Networking Apps in Shenzhen From Mobile Big Data Perspective: Longitudinal Observational Study
J Med Internet Res 2025;27:e69569
URL: https://www.jmir.org/2025/1/e69569
doi:10.2196/69569
PMID:

XSL•FO
RenderX

©Lan Wei, Yongsheng Wu, Lin Chen, Jinquan Cheng, Jin Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Toward Unsupervised Capacity Assessments for Gait in Neurorehabilitation: Validation Study

Aileen C Naef[1,2*], MSc, PhD; Guichande Duarte[3*], MM; Saskia Neumann[1], MSc; Migjen Shala[4], BSc; Meret Branscheidt[4,5], Prof Dr Med; Chris Easthope Awai[1], MSc, PhD

[1]Data Analytics & Rehabilitation Technology (DART), Lake Lucerne Institute, Vitznau, Switzerland

[2]Rehabilitation Engineering Laboratory, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[3]Department of Neurology, University Hospital of Zurich, Zurich, Switzerland

[4]cereneo, Center for Neurology and Rehabilitation, Weggis, Switzerland

[5]Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Chris Easthope Awai, MSc, PhD
Data Analytics & Rehabilitation Technology (DART)
Lake Lucerne Institute
Rubistrasse 9
Vitznau
Switzerland
Phone: 41 77 466 81 35
Email: chris.awai@llui.org

Abstract

Background: Gait impairments are common in stroke survivors, negatively impacting their overall quality of life. Therefore, gait rehabilitation is often targeted during in-clinic rehabilitation. While standardized assessments are available for inpatient evaluation, the literature often reports variable results when these assessments are conducted in a home environment. Several factors, such as the presence of an observer, the environment itself, or the technology used, may contribute to these differing results. Therefore, it is relevant to establish unsupervised capacity assessments for both in-clinic use and across the continuum of care.

Objective: This study aimed to investigate the effect of supervision on the outcomes of a sensor-based 10-meter walk test conducted in a clinical setting, maintaining a controlled environment and setup.

Methods: In total, 21 stroke survivors (10 female, 11 male; age: mean 63.9, SD 15.5 years) were assigned alternately to one of two data collection sequences and tested over 4 consecutive days, alternating between supervised test (ST) and unsupervised test (UST) assessments. For both assessments, participants were required to walk a set distance of 10 meters as fast as possible while data were collected using a single wearable sensor (Physilog 5) attached to each shoe. After each walking assessment, the participants completed the Intrinsic Motivation Inventory. Statistical analyses were conducted to examine the mean speed, stride length, and cadence, across repeated measurements and between assessment conditions.

Results: The intraclass correlation coefficient indicated good to excellent reliability for speed (ST: κ=0.93, P<.001; UST: κ=0.93, P<.001), stride length (ST: κ=0.92, P<.001; UST: κ=0.88, P<.001), and cadence (ST: κ=0.91, P<.001; UST: κ=0.95, P<.001) across repeated measurements for both ST and UST assessments. There was no significant effect of testing order (ie, sequence A vs B). Comparing ST and UST, there were no significant differences in speed (t_{39}=–0.735, P=.47, 95% CI 0.06-0.03), stride length (z=0.835, P=.80), or cadence (t_{39}=–0.501, P=.62, 95% CI 3.38-2.04) between the 2 assessments. The overall motivation did not show any significant differences between the ST and UST conditions (P>.05). However, the self-reported perceived competence increased during the unsupervised assessment from the first to the second measurement.

Conclusions: Unsupervised gait capacity assessments offer a reliable alternative to supervised assessments in a clinical environment, showing comparable results for gait speed, stride length, and cadence, with no differences in overall motivation between the two. Future work should build upon these findings to extend unsupervised assessment of both capacity and performance in home environments. Such assessments could allow improved and more specific tracking of rehabilitation progress across the continuum of care.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e66123) doi:10.2196/66123

KEYWORDS

gait analysis; gait rehabilitation; 10-meter walk test; stroke; unsupervised assessments; supervised assessments; sensors; motivation; capacity; monitoring; wearables; stroke survivors; quality of life

Introduction

More than 80% of stroke survivors experience gait impairments, and between 55% and 75% continue to experience functional problems 3 to 6 months after their stroke [1,2]. These impairments frequently lead to reduced activity levels and diminished community participation [3-5]. Due to this, one of the most common goals in gait rehabilitation is to enhance functional independence and quality of life [3,6-8]. Despite these efforts, many survivors continue to experience gait deficits even after discharge from their inpatient rehabilitation programs.

Within the clinical setting, observational gait analysis and scaled assessments are often used to identify gait deficits and functional improvements [9,10]. In these assessments, visual observation is used to qualitatively identify gait deviations. These types of assessments are popular due to their simplicity and availability. However, the literature indicates that gait and mobility parameters differ when assessments are conducted at home versus in the clinical setting; this difference may become more relevant as advancements in digital therapeutics and telerehabilitation offer promising avenues for long-term rehabilitation outside the clinic [9-14]. These differences in assessment outcomes could stem from various factors, such as the transition from a supervised to an unsupervised setting, the change in the environment itself, or technical limitations [11]. Therefore, to determine the feasibility of conducting accurate and consistent assessments in a home environment, it is essential to first isolate and address these variables in a controlled manner using a standardized clinical assessment.

Among the array of tools used in the clinical setting to evaluate walking capacity is the 10-Meter Walk Test (10-MWT) [15-18]. The test measures walking speed in meters per second over a distance of 10 meters, with the speed being indicative of functional outcomes and prognosis for stroke survivors [19-21]. In addition to providing a strong indicator of functional ability and being a predictor of health outcomes, the 10-MWT is a particularly valuable tool due to its applicability across populations, its convergent validity with other clinical gait capacity assessments, and the fact that it is standardized and easy to administer [22,23]. Moreover, combining the 10-MWT with wearable sensors, such as inertial measurement units, allows for the extraction of additional spatiotemporal gait parameters. These parameters are not only robust, but they enhance the interpretation of clinical assessment outcomes and aid in detecting motor recovery poststroke as well as predicting prognosis after stroke [24-26]. Future precision rehabilitation approaches will rely more strongly on the assessment of these kinds of granular gait metrics, making it relevant to validate these in current systems [27,28]. Standardized assessments to measure capacity, such as the 10-MWT may be complemented by data from continuous monitoring, which measures performance [29,30].

While unsupervised assessments and continuous monitoring are mostly associated with home settings, there are also many useful applications within a clinical environment. First, unsupervised assessments can reduce the assessment load on clinical staff, thus enhancing clinical efficiency [31]. Second, unsupervised assessments can be performed more frequently due to their independence from clinical planning, which allows insights into day-to-day fluctuations of test outcomes [11]. Third, enabling patients to perform their own assessments in an unsupervised manner can mitigate observer effects on outcomes [32]. Fourth, giving patients some level of autonomy concerning some aspects of their rehabilitation process can empower them and increase their involvement and ownership [33]. Finally, exposing patients to unsupervised assessments within a controlled, in-clinic context already prepares them to continue these across the home aspect of the continuum of care, removing learning effects and boosting adherence [34]. In synthesis, introducing unsupervised capacity assessments within the clinical setting has potential benefits on multiple levels, while also laying the foundation for assessments across the complete continuum of care.

For these reasons, this study investigates the supervision aspect of capacity assessments by maintaining a controlled clinical environment and using a simple-to-use tool that can easily be transferred outside the clinic. In other words, this study will determine whether a standardized gait capacity assessment conducted in the clinical setting without supervision maintains the same validity, replicability, and motivation as when conducted with supervision. We hypothesize that the unsupervised assessments will provide results that are as reliable and replicable as those collected in a supervised manner with no differences in motivation. This study will provide insights into the self-administration of gait capacity assessments, providing a basis for interpreting any potential differences that may arise in future studies expanding into home or less controlled environments.

Methods

Participants and Setting

A convenience sample of 21 patients with stroke (10 female and 11 male), aged 40-89 (mean 63.91, SD 15.54) years, were enrolled in this study at the Swiss Neurorehabilitation Clinic “cereneo” (Lucerne, Switzerland). Enrollment took place between November 2021 and June 2023 and followed a cross-over design. All stroke survivors who were more than 1 month post stroke, aged 18 years or older, able to use a tablet, with a Berg Balance Score greater than 31, and the ability to walk for at least 2 minutes, met the inclusion criteria to participate in the study. Exclusion criteria were the presence of comorbidities potentially affecting gait (eg, polyneuropathy, orthopedic impairments, and orthostatic hypotension), uncorrected visual impairments, psychiatric comorbidities that

XSL•FO
RenderX

may affect the participants' ability to follow study protocols or other medical reasons that would hinder participation.

Experimental Procedure

The total experimental procedure for this study was conducted across 4 days, with a single testing session per day. Upon enrollment in the study, participants were sequentially assigned in an alternating fashion to one of 2 sequences (A or B) representing differing protocol sequences (Figure 1). This approach was chosen to ensure balanced group sizes and facilitate study logistics. While this method does not introduce selection bias in individual assignments, it may not fully control for potential confounding factors in group distribution. Sequence A began with the 10-MWT conducted under supervised test (ST) conditions, while Sequence B began with the 10-MWT conducted under unsupervised test (UST) conditions (Figure 2). Participants then alternated the assessment type over the remaining 3 days (Figure 2). Regardless of the sequence to which they were assigned, participants always completed both the ST and the UST twice during the 4 days.

Figure 1. Schematic diagram outlining the enrollment and allocation to the intervention sequences of the study.

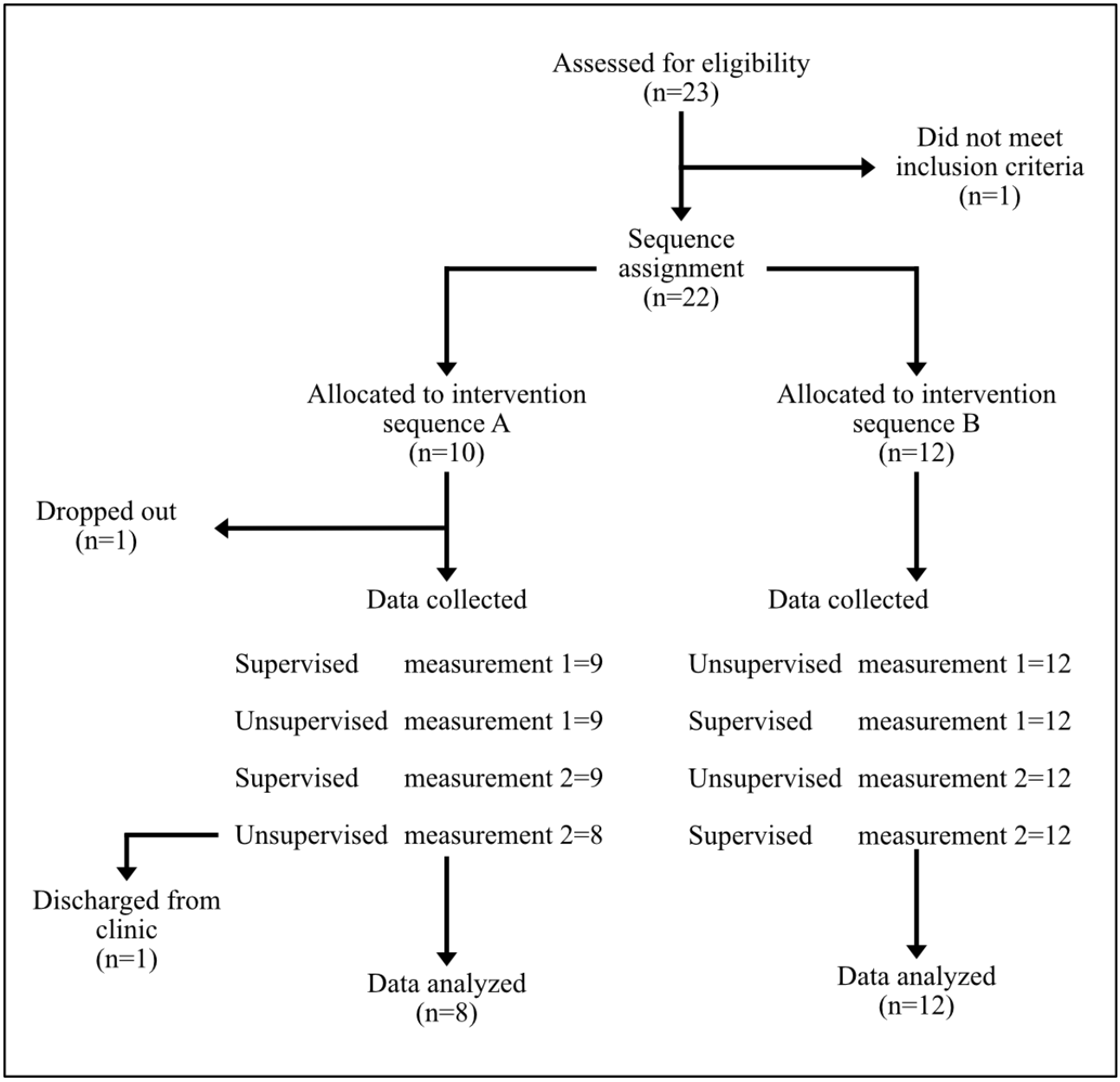

XSL•FO
RenderX

Figure 2. Study groups showing how participants were allocated to either sequence A or sequence B upon enrollment. Participants in both sequences completed the 10-MWT twice while supervised and twice while unsupervised. 10-MWT: 10-Meter Walk Test; ST: supervised test; UST: unsupervised test.

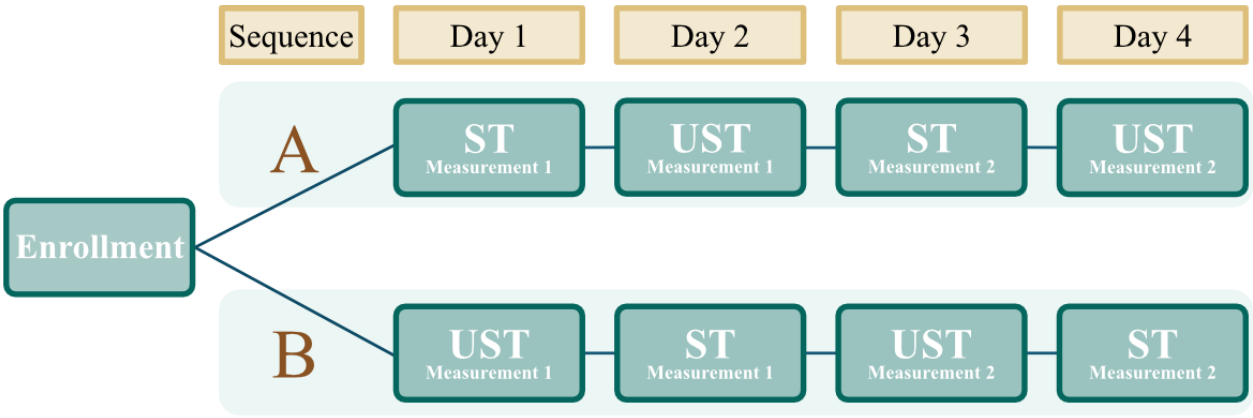

The goal of the 10-MWT was to have participants walk a set distance of 10 m as fast as possible in order to evaluate gait capacity. Markers at the 0- and 10-m points delineated the walking path to ensure consistent data collection conditions. To objectively evaluate the outcome, participants completed the assessment while wearing an inertial measurement unit on each shoe.

For all STs, the therapist prepared the participant by attaching the inertial measurement units to the shoes and positioning them at the start line. The therapist then counted down the start, and the participant walked as fast as possible over the marked finish line. The therapist was responsible for starting the stopwatch using the tablet app and stopping it when the finish line was reached. The therapist was present for the entire setup and data collection but did not provide feedback or guidance.

The UST was conducted following the same procedure as the ST. However, there was no direct supervision of the tasks. During the UST, a therapist or nurse was available in the adjacent nursing room for technical assistance and support during the setup. The participants were required to go to this room and pick up the tablet and sensors before walking to the designated track. If participants were unable to attach the inertial measurement units to the shoes themselves, the nurses or therapists attached them in the nurses' office. Additionally, if the participants were unable to hold the tablet, for example, due to hemiparesis, the tablet was attached to the participant's arm via a strap so that they could operate it with one hand. Once at the starting line, the participants got in place and started the stopwatch using the tablet app before walking as fast as they could over the finish line. Once across the finish line, the participants stopped the time using the same stopwatch app. The participants then removed the sensors and returned them, together with the tablet, to the nursing office. During the UST, the only interaction with the therapist was to attach or remove the sensors in the nursing office if the participant could not do it alone.

In both conditions, participants were required to complete the Intrinsic Motivation Inventory on paper in the nurses' office. The participants were required to do this on their own with no help or interaction with the nurses or therapists. Participants in both sequences also always had a written information sheet available to them in both the ST and UST conditions, listing the required steps and acting as a user guide for the tablet app. In both conditions, there was no interaction between the participants and nurses or therapists, the only difference was that during the ST task, the nurse or therapist was there to start and stop the stopwatch and observe the data collection. If there were any problems, the data collection was repeated, either with the nurse or therapist, or independently for the ST and UST trials, respectively.

Data Collection Systems and Materials

For this study, demographic information was extracted by the clinical team from the clinical record and provided to the study team in a pseudoanonymized form. Gait data was collected using validated Physilog®5 (Gait Up), wearable inertial measurement units, to assess the spatiotemporal gait outcomes during the 10-MWT [35-37]. To keep the setup simple and easy for patients to use, only 2 sensors were used in this study. Each sensor was mounted on the top of the participant's foot, specifically over the laces of their shoes (Figure S1 in Multimedia Appendix 1). The sensors had a sampling rate of 128 Hz and collected acceleration and orientation data. Gait data was subsequently analyzed using the Gait Analyser software (V3.1; Gait Up), a valid and reliable tool for stroke survivors [35,36]. Information about participant motivation was also collected using the Intrinsic Motivation Inventory as a standard pen-and-paper questionnaire that participants filled out in the nurses' office when returning the sensors [38].

Statistical Analysis

Descriptive statistics of all gait parameters were calculated to provide an overview of the data. For each assessment, speed, stride length, and cadence were calculated as the mean values of the left and right sensors. These specific gait parameters were chosen to provide a simplified overview of changes in temporal, spatial, and combined spatiotemporal parameters. Additionally, changes in these parameters are important to capture are they represent a more cautious gait in the elderly population and are known to change poststroke [39,40]. All statistical testing was conducted for each of the 3 gait parameters. Normality was determined using the Shapiro-Wilk test (significance level α=.05), while the variances were determined using an *F* test

XSL•FO
RenderX

(α=.05). Corrections for multiple comparisons were done using the Holm-Bonferroni method.

Difference Between Repeated Measurements

In the first step, the measurements conducted on different days were compared via a paired *t* test to determine if there was any learning effect from measurement 1 to measurement 2 of the same assessment type. Additionally, using a 2-way random effect model and "single rater" unit, the intraclass correlation coefficient was computed to assess the agreement between the 2 measurements per assessment condition.

Sequence Effect

Next, to determine whether there was a sequence effect in the data, the difference in scores between the 2 measurements per assessment type in each sequence was calculated. For data found to be normally distributed and with equal variances, an independent samples *t* test was conducted between sequences (A vs B); for data found to be not normally distributed and with equal variances, a nonparametric Mann-Whitney *U* test was conducted. Significant differences between sequences A and B would suggest a sequence effect is present.

Supervised Versus Unsupervised Tests

The difference in scores between ST and UST measurements for each participant was determined using a student's *t* test and a nonparametric Wilcoxon signed-rank test, depending on the outcomes of the normality testing.

Intrinsic Motivation Inventory

To determine if there were differences in motivation, the subsections of the Intrinsic Motivation Inventory were analyzed. Depending on the normality of the data, either a paired *t* test or a Wilcoxon signed-rank test was used to determine if there was a change in motivation from the first to second measurement day per assessment in each sequence. Additionally, a Student *t* test or a Wilcoxon signed-rank test was conducted to determine if there was a significant difference in motivation scores between assessment types (ie, ST vs UST).

Ethical Considerations

This study was approved via waiver by the local ethics committee of the Canton of Lucerne, Switzerland (Req-2020-00995) and was carried out in accordance with the current version of the Declaration of Helsinki. The study protocol was explained both orally and in writing to eligible participants, and written informed consent was obtained before participation. Participants did not receive any compensation for their participation in this study, nor did they receive any direct benefit from their participation. To ensure privacy and confidentiality, all data were deidentified and securely stored in an encrypted database accessible only to authorized researchers. The wearable sensor data used for gait analysis were coded and stored separately from participant identifiers.

Results

Demographics

A total of 21 poststroke participants were recruited for this study (Table 1). All participants were able to complete all 4 days of testing. The majority of the participants were in the subacute phase and were largely community ambulators (Table 1) [41].

Table 1. Participant characteristics.

Characteristics	Values
Age in years, mean (SD)	63.91 (15.54)
Sex, n (%)	
Female	10 (48)
Male	11 (52)
Paretic body side, n (%)	
Left	16 (76)
Right	5 (24)
Months since stroke, mean (SD)	5.19 (10.73)
Stroke onset 1-6 months, n (%)	18 (85.71)
Stroke onset >6 months, n (%)	3 (14.29)
Modified Rankin Scale, mean (SD)	1.95 (0.59)
Household ambulators: <0.4 m/s, n (%)	1 (4.76)
Limited community ambulators: 0.4-0.8 m/s, n (%)	4 (19.05)
Community ambulators: >0.8 m/s, n (%)	16 (76.19)
Assistive device, n (%)	2 (9.52)
Rollator	1 (4.76)
Dropped foot stimulator	1 (4.76)
Foot orthoses, n (%)	3 (14.29)

XSL•FO
RenderX

Gait Outcomes

Difference Between Repeated Measurements

Shapiro-Wilk tests for normality showed that the results for speed, stride length, and cadence were normally distributed during each measurement, per assessment type and sequence (Table S1 in Multimedia Appendix 2). A paired samples *t* test subsequently determined that there were no statistically significant differences between the first and second measurements for each assessment type within each sequence (Table 2).

Table 2. Differences between repeated measurements.

Variable and sequence	Condition	*t* test[a] (*df*)	*P* value[b]	Adjusted *P* value[c]	95% CI
Speed					
A	ST[d]	0.492 (7)	0.64	>.99	–0.09 to 0.14
A	UST[e]	–0.064 (11)	0.95	>.99	–0.09 to 0.08
B	ST	–0.238 (11)	0.82	>.99	–0.08 to 0.07
B	UST	–5.10E–16 (7)	>.99	>.99	–0.06 to 0.06
Stride length					
A	ST	1.011 (7)	0.36	>.99	–0.05 to 0.12
A	UST	0.165 (11)	0.87	>.99	–0.06 to 0.07
B	ST	0.444 (11)	0.68	>.99	–0.05 to 0.08
B	UST	0.097 (7)	0.93	>.99	–0.12 to 0.13
Cadence					
A	ST	–0.322 (7)	0.76	>.99	–5.21 to 3.96
A	UST	–0.679 (11)	0.51	>.99	–5.41 to 2.86
B	ST	–0.132 (11)	0.90	>.99	–5.52 to 4.89
B	UST	0.353 (7)	0.73	>.99	–3.31 to 4.47

[a]Paired samples *t* test.

[b]$P \leq .05$.

[c]Holm-Bonferroni correction for multiple comparisons.

[d]ST: supervised test.

[e]UST: unsupervised test.

The intraclass correlation for speed found the ST (κ=0.93, adjusted $P<.001$) and UST (κ=0.93, adjusted $P<.001$) to both have excellent reliability across measurements. For stride length, the intraclass correlation was found to have excellent reliability for ST (κ=0.92, adjusted $P<.001$) and good reliability for UST (κ=0.88, adjusted $P<.001$). For cadence, both the ST (κ=0.91, adjusted $P<.001$) and UST (κ=0.95, adjusted $P<.001$) were found to have excellent reliability.

Sequence Effect

An independent samples *t* test found no sequence effect between sequences A and B for the normally distributed speed, cadence, and stride length (ST only; Tables S2 and S3 in Multimedia Appendix 2). A nonparametric Mann-Whitney *U* test equally found that there was no sequence effect between sequences A and B for the nonnormally distributed stride length (UST only; Table S2 and S4 in Multimedia Appendix 2). Before this, the results for the *F* test found no significant differences in variance across sequences (Table S5 in Multimedia Appendix 2).

Supervised Versus Unsupervised Tests

A paired-sample *t* test found that there was no significant difference between speed (t_{39}=–0.735, adjusted $P>.99$, 95% CI –0.06 to 0.03) and cadence (t_{39}=–0.501, adjusted $P>.99$, 95% CI –3.38 to 2.04) recorded when collecting data while supervised versus unsupervised (Figure 3). A pairwise Wilcoxon signed-rank test found that there was no statistically significant difference between the stride length (z=0.835, adjusted $P>.99$) collected during the ST versus UST. Additional results regarding the normality tests can be found in Table S6 in Multimedia Appendix 2.

The intraclass correlation coefficient was computed to assess the agreement between the supervised and unsupervised assessments of speed, stride length, and cadence (Table 3).

XSL•FO
RenderX

Figure 3. Distribution of data for speed (m/s), stride length (m), and cadence (steps/minute) during the ST and UST, with no significant differences found. ST, supervised test; UST, unsupervised test.

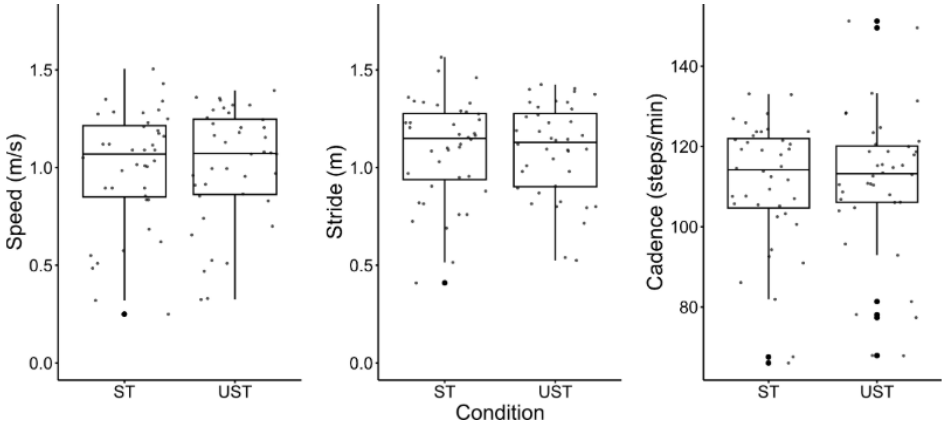

Table 3. Intraclass correlations between supervised and unsupervised assessments.

Variable	ICC value	*P* value[a]	*P* adjusted[b]	95% CI	Type of ICC[c]	Interpretation
Speed	0.90	<.001	<.001	0.81-0.94	ICC (A,1)	Good
Stride length	0.88	<.001	<.001	0.78-0.93	ICC (A,1)	Good
Cadence	0.88	<.001	<.001	0.79-0.94	ICC (C,1)	Good

[a]*P* value: 0.05.

[b]Holm-Bonferroni correction for multiple comparisons.

[c]Absolute agreement model for single measurement.

Bland-Altman Plots

The effect of the assessment condition on the speed was found to result in a mean difference of –0.016 (SD 0.137) m/s during the UST compared with the ST. For stride length, the mean difference was found to be 0.003 (SD 0.125) m during the UST compared with the ST. For cadence, the mean difference was found to be –0.672 (8.481) steps per minute during the UST compared with the ST. Based on the Bland-Altman plots, no proportional bias was observed for any variable (Figure 4).

Figure 4. Results from the Bland-Altman analysis showing the various effect zones when examining the ST and UST for speed, stride length, and cadence. The darker bands represent the upper and lower limits of agreement, while the lighter band represents the bias zone. Data points are separated into ambulator types, namely household, limited, or community. ST: supervised test; UST: unsupervised test.

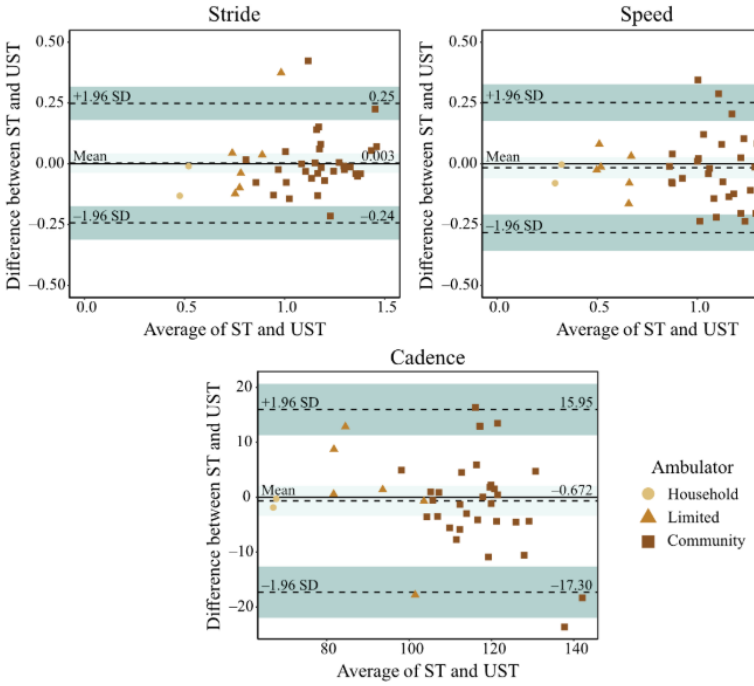

XSL•FO
RenderX

Intrinsic Motivation Inventory Outcomes

The results of the Wilcoxon signed-rank test and paired-sample *t* test found that there are no significant differences in the motivation scores for the subscales of the Intrinsic Motivation Inventory between the ST and UST (Table 4).

Table 4. Comparison of the results of the Intrinsic Motivation Inventory, presented as its subscales, between the ST[a] and UST[b].

Subscale and condition	Mean (SD)	Score	*P* values	Adjusted *P* values[c]
Interest/enjoyment		–0.194[d]	0.42	>.99
ST	4.15 (1.62)			
UST	4.00 (1.70)			
Perceived competence		–0.068[d]	0.47	>.99
ST	5.14 (1.60)			
UST	5.18 (1.44)			
Effort/importance		–0.088[d]	0.47	>.99
ST	4.22 (0.83)			
UST	4.28 (1.19)			
Pressure/tension		–0.435[e]	0.67	>.99
ST	2.02 (1.55)			
UST	2.10 (1.43)			

[a]ST: supervised test.

[b]UST: unsupervised test.

[c]Holm-Bonferroni correction for multiple comparisons.

[d]*z*-score.

[e]*t*-score.

Examining the difference between measurements 1 and 2, there were no significant differences for the interest/enjoyment, effort/importance, and pressure/tension subscales during both the ST and UST. Conversely, there was a significant difference in motivation scores for the Intrinsic Motivation Inventory subscale perceived competence during measurement 1 (mean 4.04, SD 1.58) versus measurement 2 (mean 5.17, SD 1.40) of the UST (t_7=–2.979, P=.02, odds ratio 2.60, 95% CI –2.02 to –0.23) during sequence A. A significant difference between measurement 1 (mean 5.25, SD 1.33) and measurement 2 (mean 5.86, SD 1.13) was also found during sequence B (z=–1.716, P=.04). However, after applying the Holm-Bonferroni correction for multiple comparisons, the differences were no longer significant after adjustment (adjusted P values were .08 and .13, respectively). There were no differences between measurements in the perceived competence during the ST.

Discussion

Principal Findings

There is a considerable gap in our understanding of whether currently used standardized inpatient gait capacity assessments can be effectively performed in an unsupervised manner while maintaining the same levels of validity and replicability. Here, we specifically investigated the 10-MWT, a common measure of gait capacity, which is typically conducted under the supervision of a therapist to ensure accurate results. Our findings demonstrate that the execution mode of the 10-MWT, whether supervised by a health care professional or conducted autonomously by the participants themselves, did not result in significant differences in the measured gait parameters. These results suggest that patients are capable of accurately self-assessing their walking function without the need for direct supervision.

The main results of this study demonstrate that participants were able to maintain their execution ability without continuous guidance and feedback, suggesting that once an individual becomes familiar and comfortable with the assessment and related equipment, continuous supervision might not be necessary to preserve this ability. Moreover, these findings show the ability of stroke survivors to not only be able to carry out the test procedure, but also attach, connect, and start the required sensors with the help of a layperson after proper instruction. Getting patients involved in their own assessments may also promote greater flexibility and engagement in their rehabilitation process [42]. Moreover, therapists can efficiently monitor patients by leveraging reliable sensor systems for remote data collection, leading to more efficient treatment planning and timely interventions. This can not only reduce costs but also make rehabilitation more accessible and portable.

Enabling patients to conduct the assessments themselves could also have the added benefit of preventing experimenter-induced changes, which multiple studies have found to influence gait [32,43-45]. This is supported by work that found modifications in gait speed, cadence, step duration, and stride length, among others, in the presence of researchers [32,45-49]. Evidence also suggests that gait changes may even depend on the number of

XSL•FO
RenderX

observers present [32]. However, the exact relationship between gait parameters and observation remains unclear due to contradictory results, which are thought to arise from differences in study design. While the studies mentioned here involved passive observers, another factor that could influence performance in the real world is motivation.

Motivation remains a clear confounder in the reliability of motor assessments, both for simple and complex tasks. For instance, Kuhl and Koch [50], modeled motor performance as a dual task or multiple tasks, with most nonmotor components being dependent on the environment, personality, and priming. Concerning the environment, sex, expectations, and the behavior of the experimenter have long been known to have an effect on motor performance [51,52]. In terms of personality, Strickland [53] reported the effects of the need-for-approval on motor outcomes in the 1960s. Regarding priming, a recent study demonstrated that priming with positive or negative content before a test resulted in distinct differences in simple motor capacity in healthy college students [54,55]. Based on this literature, we suspected that motivational scores would drop when no active experimenter was present to administer the test. However, in this study, we did not find this to be true. In fact, there were no reported differences in overall motivation between the supervised and unsupervised assessments. Before the correction for multiple comparisons, the perceived competence score during the unsupervised assessment was significantly lower during the first measurement compared with the second. Despite this finding no longer being significant after corrections were applied, this difference could still represent an important aspect to consider, as individuals may be less motivated to continue doing an assessment if they feel incompetent while doing it. This could be further examined in future work to determine whether this effect truly exists or not. Additionally, strategies to ensure that the highest level of understanding and skill is achieved during the unsupervised assessments should be anticipated.

Limitations and Future Outlook

It should be noted that despite this study finding no changes in the motivation level between the 2 assessments, this finding is possibly biased, as participants were likely already highly motivated due to the study being voluntary. Despite this, there is currently no evidence to suggest that, when controlling for environment and setup, the presence or absence of an observer changes the motivation to complete the assessment.

To extend the generalizability of these findings to different environments, the next step would be to expand this work to the home environment. In doing so, future work would need to take into consideration the safety and feasibility of this setup outside of the clinical setting. While no participants in this study fell or were otherwise injured, a risk assessment should be conducted, including in other populations, to ensure patient safety. An additional aspect that may need to be considered for future work is cognitive ability, which was not examined in this study. As the majority of the participants were only mildly impaired, the applicability of the findings to patients with more severe impairment needs to be investigated. Finally, there are additional aspects, such as reduced costs and boosted self-efficacy, which were not considered here, that could be explored in future work as they could elucidate further benefits of autonomous assessments for clinics and patients.

Conclusions

This validation study, using easy-to-use wearable sensors, has demonstrated that the 10-MWT, a standardized gait capacity assessment, can be performed reliably whether professionally supervised or conducted autonomously, without significant differences in the measured gait parameters. Additionally, there was no self-reported difference in motivation for completing the 10-MWT when conducted under supervised or unsupervised conditions. The findings of this study are important as they demonstrate the validity of the unsupervised versus supervised 10-MWT inside clinics. Furthermore, the results provide a solid foundation for investigating the validity of the 10-MWT outside of the clinic in combination with continuous monitoring. Both topics will play an increasingly important role as the frequency and granularity of clinical assessments increase and researchers begin to focus more on monitoring survivors' rehabilitation progress and quality of life in the real-world setting.

Acknowledgments

The authors would like to acknowledge the therapists and nurses who assisted us with this study and helped carry out the data collection. This study was funded by the P & K Foundation.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 3.

Authors' Contributions

CEA performed the conceptualization of the study and handled the funding acquisition. ACN conducted the formal analysis. DG, SN, and MS carried out the investigation. MB and CEA managed project administration, resources, and supervision. ACN led the visualization and wrote the original draft, with support from GD. For review and editing, ACN took the lead, with support from GD, SN, MS, MB, and CEA.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Sensor placement on the top of the participant's shoes.
[PDF File (Adobe PDF File), 169 KB - jmir_v27i1e66123_app1.pdf]

Multimedia Appendix 2
Additional tables showing additional results per test type and sequence.
[PDF File (Adobe PDF File), 301 KB - jmir_v27i1e66123_app2.pdf]

Multimedia Appendix 3
Full, preprocessed, dataset used for the data analyses outlined in the manuscript.
[XLSX File (Microsoft Excel File), 28 KB - jmir_v27i1e66123_app3.xlsx]

References

1. Donkor ES. Stroke in the 21 century: a snapshot of the burden, epidemiology, and quality of life. Stroke Res Treat 2018;2018:3238165 [FREE Full text] [doi: 10.1155/2018/3238165] [Medline: 30598741]
2. Lai SM, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the Stroke Impact Scale. Stroke 2002;33(7):1840-1844. [doi: 10.1161/01.str.0000019289.15440.f2] [Medline: 12105363]
3. Rehabilitation in health systems: guide for action. World Health Organization. URL: https://www.who.int/publications/i/item/9789241515986 [accessed 2024-03-04]
4. Fini NA, Bernhardt J, Churilov L, Clark R, Holland AE. Adherence to physical activity and cardiovascular recommendations during the 2 years after stroke rehabilitation discharge. Ann Phys Rehabil Med 2021;64(2):101455 [FREE Full text] [doi: 10.1016/j.rehab.2020.03.018] [Medline: 33189943]
5. Fini NA, Holland AE, Keating J, Simek J, Bernhardt J. How physically active are people following stroke? Systematic review and quantitative synthesis. Phys Ther 2017;97(7):707-717. [doi: 10.1093/ptj/pzx038] [Medline: 28444348]
6. Harris JE, Eng JJ. Goal priorities identified through client-centred measurement in individuals with chronic stroke. Physiother Can 2004;56(3):171-176 [FREE Full text] [doi: 10.2310/6640.2004.00017] [Medline: 23372280]
7. Rehabilitation. World Health Organization. 2024 Apr 22. URL: https://www.who.int/news-room/fact-sheets/detail/rehabilitation [accessed 2024-09-04]
8. Bohannon RW, Andrews AW, Smith MB. Rehabilitation goals of patients with hemiplegia. International Journal of Rehabilitation Research 1988;11(2):181-184. [doi: 10.1097/00004356-198806000-00012]
9. Wallmann HW. Introduction to observational gait analysis. Home Health Care Manag Pract 2009;22(1):66-68. [doi: 10.1177/1084822309343277]
10. Toro B, Nester C, Farren P. A review of observational gait assessment in clinical practice. Physiother Theory Pract 2009;19(3):137-149. [doi: 10.1080/09593980307964]
11. Warmerdam E, Hausdorff JM, Atrsaei A, Zhou Y, Mirelman A, Aminian K, et al. Long-term unsupervised mobility assessment in movement disorders. Lancet Neurol 2020;19(5):462-470. [doi: 10.1016/S1474-4422(19)30397-7] [Medline: 32059811]
12. Brodie MA, Coppens MJ, Ejupi A, Gschwind YJ, Annegarn J, Schoene D, et al. Comparison between clinical gait and daily-life gait assessments of fall risk in older people. Geriatr Gerontol Int 2017;17(11):2274-2282. [doi: 10.1111/ggi.12979] [Medline: 28176431]
13. Corrà MF, Atrsaei A, Sardoreira A, Hansen C, Aminian K, Correia M, et al. Comparison of laboratory and daily-life gait speed assessment during ON and OFF states in Parkinson's disease. Sensors (Basel) 2021;21(12):3974 [FREE Full text] [doi: 10.3390/s21123974] [Medline: 34207565]
14. Van Ancum JM, van Schooten KS, Jonkman NH, Huijben B, van Lummel RC, Meskers CGM, et al. Gait speed assessed by a 4-m walk test is not representative of daily-life gait speed in community-dwelling adults. Maturitas 2019;121:28-34. [doi: 10.1016/j.maturitas.2018.12.008] [Medline: 30704562]
15. Moore JL, Potter K, Blankshain K, Kaplan SL, O Dwyer LC, Sullivan JE. A core set of outcome measures for adults with neurologic conditions undergoing rehabilitation: a clinical practice guideline. J Neurol Phys Ther 2018;42(3):174-220 [FREE Full text] [doi: 10.1097/NPT.0000000000000229] [Medline: 29901487]
16. Otterman N, Veerbeek J, Schiemanck S, van der Wees P, Nollet F, Kwakkel G. Selecting relevant and feasible measurement instruments for the revised Dutch clinical practice guideline for physical therapy in patients after stroke. Disabil Rehabil 2017;39(14):1449-1457 [FREE Full text] [doi: 10.1080/09638288.2016.1196399] [Medline: 27374876]
17. Teasell R, Salbach NM, Foley N, Mountain A, Cameron JI, de Jong A, et al. Canadian stroke best practice recommendations: rehabilitation, recovery, and community participation following stroke. Part one: rehabilitation and recovery following stroke; 6th edition update 2019. Int J Stroke 2020;15(7):763-788. [doi: 10.1177/1747493019897843] [Medline: 31983296]
18. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. Int J Stroke 2017;12(5):451-461 [FREE Full text] [doi: 10.1177/1747493017711813] [Medline: 28697709]

XSL•FO
RenderX

19. Kollen B, Kwakkel G, Lindeman E. Hemiplegic gait after stroke: is measurement of maximum speed required? Arch Phys Med Rehabil 2006;87(3):358-363. [doi: 10.1016/j.apmr.2005.11.007] [Medline: 16500169]
20. Cheng DKY, Dagenais M, Alsbury-Nealy K, Legasto JM, Scodras S, Aravind G, et al. Distance-limited walk tests post-stroke: a systematic review of measurement properties. NeuroRehabilitation 2021;48(4):413-439 [FREE Full text] [doi: 10.3233/NRE-210026] [Medline: 33967070]
21. Tyson S, Connell L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. Clin Rehabil 2009;23(11):1018-1033. [doi: 10.1177/0269215509339004] [Medline: 19786420]
22. Kahn J, Tefertiller C. Measurement characteristics and clinical utility of the 10-meter walk test among individuals with spinal cord injury. Arch Phys Med Rehabil 2014;95(5):1011-1012. [doi: 10.1016/j.apmr.2014.01.021]
23. Lindholm B, Nilsson MH, Hansson O, Hagell P. The clinical significance of 10-m walk test standardizations in Parkinson's disease. J Neurol 2018;265(8):1829-1835 [FREE Full text] [doi: 10.1007/s00415-018-8921-9] [Medline: 29876762]
24. Wonsetler EC, Bowden MG. A systematic review of mechanisms of gait speed change post-stroke. Part 1: spatiotemporal parameters and asymmetry ratios. Top Stroke Rehabil 2017;24(6):435-446 [FREE Full text] [doi: 10.1080/10749357.2017.1285746] [Medline: 28220715]
25. An S, Lee Y, Shin H, Lee G. Gait velocity and walking distance to predict community walking after stroke. Nurs Health Sci 2015;17(4):533-538. [doi: 10.1111/nhs.12234] [Medline: 26310714]
26. Gouelle A, Mégrot F. Interpreting spatiotemporal parameters, symmetry, and variability in clinical gait analysis. In: Müller B, Wolf SI, Brueggemann GP, Deng Z, McIntosh A, Miller F, et al, editors. Handbook of Human Motion. Cham: Springer; 2017.
27. Reinkensmeyer DJ, Burdet E, Casadio M, Krakauer JW, Kwakkel G, Lang CE, et al. Computational neurorehabilitation: modeling plasticity and learning to predict recovery. J Neuroeng Rehabil 2016;13(1):42 [FREE Full text] [doi: 10.1186/s12984-016-0148-3] [Medline: 27130577]
28. Schweighofer N. Computational neurorehabilitation. In: Reinkensmeyer DJ, Marchal-Crespo L, Dietz V, editors. Neurorehabilitation Technology. Cham: Springer; 2022.
29. Kluge F, Del Din S, Cereatti A, Gaßner H, Hansen C, Helbostad JL, Mobilise-D consortium. Consensus based framework for digital mobility monitoring. PLoS One 2021;16(8):e0256541 [FREE Full text] [doi: 10.1371/journal.pone.0256541] [Medline: 34415959]
30. Osuka Y, Chan LLY, Brodie MA, Okubo Y, Lord SR. A wrist-worn wearable device can identify frailty in middle-aged and older adults: the UK biobank study. J Am Med Dir Assoc 2024;25(10):105196. [doi: 10.1016/j.jamda.2024.105196] [Medline: 39128825]
31. Jing Q, Xing Y, Duan M, Guo P, Cai W, Gao Q, et al. Study on the rehabilitation therapist estimation under institutional perspective by applying the workload indicators of staffing needs in the aging context. Front Public Health 2022;10:929675 [FREE Full text] [doi: 10.3389/fpubh.2022.929675] [Medline: 35784252]
32. Friesen KB, Zhang Z, Monaghan PG, Oliver GD, Roper JA. All eyes on you: how researcher presence changes the way you walk. Sci Rep 2020;10(1):17159 [FREE Full text] [doi: 10.1038/s41598-020-73734-5] [Medline: 33051502]
33. Riva G, Gaggioli A. Rehabilitation as empowerment: the role of advanced technologies. Stud Health Technol Inform 2009;145:3-22. [Medline: 19592783]
34. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Assessment of the learning curve in health technologies. A systematic review. Int J Technol Assess Health Care 2000;16(4):1095-1098. [doi: 10.1017/s0266462300103149] [Medline: 11155830]
35. physilog. URL: https://research.gaitup.com/physilog/ [accessed 2024-03-04]
36. Lefeber N, Degelaen M, Truyers C, Safin I, Beckwee D. Validity and reproducibility of inertial physilog sensors for spatiotemporal gait analysis in patients with stroke. IEEE Trans Neural Syst Rehabil Eng 2019;27(9):1865-1874. [doi: 10.1109/TNSRE.2019.2930751] [Medline: 31352347]
37. Mariani B, Rouhani H, Crevoisier X, Aminian K. Quantitative estimation of foot-flat and stance phase of gait using foot-worn inertial sensors. Gait Posture 2013;37(2):229-234. [doi: 10.1016/j.gaitpost.2012.07.012] [Medline: 22877845]
38. Ryan RM. Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. J Pers Soc Psychol 1982;43(3):450-461. [doi: 10.1037/0022-3514.43.3.450]
39. Olney SJ, Richards C. Hemiparetic gait following stroke. Part I: characteristics. Gait Posture 1996;4(2):136-148. [doi: 10.1016/0966-6362(96)01063-6]
40. Herssens N, Verbecque E, Hallemans A, Vereeck L, Van Rompaey V, Saeys W. Do spatiotemporal parameters and gait variability differ across the lifespan of healthy adults? A systematic review. Gait Posture 2018;64:181-190. [doi: 10.1016/j.gaitpost.2018.06.012] [Medline: 29929161]
41. Kwakkel G, Stinear C, Essers B, Munoz-Novoa M, Branscheidt M, Cabanas-Valdés R, et al. Motor rehabilitation after stroke: European stroke organisation (ESO) consensus-based definition and guiding framework. Eur Stroke J 2023;8(4):880-894 [FREE Full text] [doi: 10.1177/23969873231191304] [Medline: 37548025]

XSL•FO
RenderX

42. Stephenson A, Howes S, Murphy PJ, Deutsch JE, Stokes M, Pedlow K, et al. Factors influencing the delivery of telerehabilitation for stroke: a systematic review. PLoS One 2022;17(5):e0265828 [FREE Full text] [doi: 10.1371/journal.pone.0265828] [Medline: 35544471]
43. Jeon J, Kwon SY, Lee YM, Hong J, Yu J, Kim J, et al. Influence of the hawthorne effect on spatiotemporal parameters, kinematics, ground reaction force, and the symmetry of the dominant and nondominant lower limbs during gait. J Biomech 2023;152:111555. [doi: 10.1016/j.jbiomech.2023.111555] [Medline: 37030040]
44. Farhan S, Avalos MA, Rosenblatt NJ. Variability of spatiotemporal gait kinematics during treadmill walking: is there a hawthorne effect? J Appl Biomech 2023;39(3):151-156. [doi: 10.1123/jab.2022-0185] [Medline: 37001865]
45. Robles-García V, Corral-Bergantiños Y, Espinosa N, Jácome MA, García-Sancho C, Cudeiro J, et al. Spatiotemporal gait patterns during overt and covert evaluation in patients with Parkinson´s disease and healthy subjects: is there a hawthorne effect? J Appl Biomech 2015;31(3):189-194. [doi: 10.1123/jab.2013-0319] [Medline: 25536440]
46. Vickers J, Reed A, Decker R, Conrad BP, Olegario-Nebel M, Vincent HK. Effect of investigator observation on gait parameters in individuals with and without chronic low back pain. Gait Posture 2017;53:35-40. [doi: 10.1016/j.gaitpost.2017.01.002] [Medline: 28073085]
47. Hutchinson LA, Brown MJ, Deluzio KJ, De Asha AR. Self-selected walking speed increases when individuals are aware of being recorded. Gait Posture 2019;68:78-80. [doi: 10.1016/j.gaitpost.2018.11.016] [Medline: 30465945]
48. Malchow C, Fiedler G. Effect of observation on lower limb prosthesis gait biomechanics: preliminary results. Prosthet Orthot Int 2016;40(6):739-743. [doi: 10.1177/0309364615605374] [Medline: 26420262]
49. Takayanagi N, Sudo M, Yamashiro Y, Lee S, Kobayashi Y, Niki Y, et al. Relationship between daily and in-laboratory gait speed among healthy community-dwelling older adults. Sci Rep 2019;9(1):3496 [FREE Full text] [doi: 10.1038/s41598-019-39695-0] [Medline: 30837520]
50. Kuhl J, Koch B. Motivational determinants of motor performance: the hidden second task. Psychol Res 1984;46:143-153. [doi: 10.1007/bf00308599]
51. Johnson RW. Subject performance as affected by experimenter expectancy, sex of experimenter, and verbal reinforcement. Can J Behav Sci 1970;2(1):60-66. [doi: 10.1037/h0082711]
52. Rumenik DK, Capasso DR, Hendrick C. Experimenter sex effects in behavioral research. Psychol Bull 1977;84(5):852-877. [doi: 10.1037//0033-2909.84.5.852]
53. Strickland BR. Need approval and motor steadiness under positive and negative approval conditions. Percept Mot Skills 1965;20:667-668. [doi: 10.2466/pms.1965.20.2.667] [Medline: 14279360]
54. Clément-Guillotin C, Colombel F, Easthope CA, Fontayne P. Being incidentally exposed to a sport context: same consequences on gender schema activation as being in a sport context? Mov Sport Sci/Sci Mot 2022(116):61-72. [doi: 10.1051/sm/2022012]
55. Chalabaev A, Brisswalter J, Radel R, Coombes SA, Easthope C, Clément-Guillotin C. Can stereotype threat affect motor performance in the absence of explicit monitoring processes? Evidence using a strength task. J Sport Exerc Psychol 2013;35(2):211-215. [doi: 10.1123/jsep.35.2.211] [Medline: 23535978]

Abbreviations

10-MWT: 10-Meter Walk Test
ST: supervised test
UST: unsupervised test

Edited by A Coristine; submitted 04.09.24; peer-reviewed by C Kraan, H Karimi; comments to author 30.11.24; revised version received 20.12.24; accepted 28.02.25; published 26.03.25.

Please cite as:
Naef AC, Duarte G, Neumann S, Shala M, Branscheidt M, Easthope Awai C
Toward Unsupervised Capacity Assessments for Gait in Neurorehabilitation: Validation Study
J Med Internet Res 2025;27:e66123
URL: https://www.jmir.org/2025/1/e66123
doi:10.2196/66123
PMID:

©Aileen C Naef, Guichande Duarte, Saskia Neumann, Migjen Shala, Meret Branscheidt, Chris Easthope Awai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

XSL•FO
RenderX

Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Bridging the Gap in Carbohydrate Counting With a Mobile App: Needs Assessment Survey

Asmaa Housni[1], RD, BSc; Alexandra Katz[1,2], MSc; Lucien Junior Bergeron[3], PhD; Alain Simard[4]; Ashley Finkel[1], RD, BSc; Amélie Roy-Fleming[1], RD, MSc; Meranda Nakhla[5], MD, MSc; Anne-Sophie Brazeau[1], RD, PhD

[1]School of Human Nutrition, McGill University, Montreal, QC, Canada

[2]Faculté de médecine, Université de Montréal, Montreal, QC, Canada

[3]Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, QC, Canada

[4]Petit Cactus Inc, Sherbrooke, QC, Canada

[5]Research Institute of the McGill University Health Center, Division of Endocrinology, Montreal Children's Hospital, Montreal, QC, Canada

Corresponding Author:
Anne-Sophie Brazeau, RD, PhD
School of Human Nutrition
McGill University
21111 Lakeshore Road
Montreal, QC, H9V 1X9
Canada
Phone: 1 5143987848
Email: anne-sophie.brazeau@mcgill.ca

Abstract

Background: Carbohydrate counting (CC) can be burdensome and difficulty with adherence has been reported. Automated CC through mobile apps offers innovative solutions to ease this burden.

Objective: This cross-sectional web-based survey aims to identify (1) perceived barriers to CC by Canadians living with type 1 diabetes (T1D) and (2) app features that would help reduce these barriers. The secondary objective aims to compare apps used by participants with the suggested app features.

Methods: People with T1D aged 14 years and older, living in Canada, were recruited through the BETTER Canadian registry, diabetes organizations, and social media. Participants completed a 39-question web-based survey (closed- and open-ended) to identify barriers in CC, preferred CC app features, and current app use. Respondents rated barriers and app features using a 5-point Likert scale. The features were cross-referenced in each app reported being used by participants. Descriptive statistics summarized barriers and app feature preferences, and statistical analyses identified differences by age, app use, and insulin modality. Mean scores (out of 5) were compared using 2-tailed *t* tests or nonparametric tests. Open-ended questions were analyzed using inductive thematic analysis.

Results: Participants (N=196; woman: n=145, 74%; mean age 40 [SD 17] years; mean diabetes duration 22 (14) years; relied on CC to determine insulin doses at mealtimes: n=178, 90.8%) reported barriers related to carbohydrate identification, nutrient interaction, and insulin dose calculation, as well as psychosocial factors. Preferred app features included nutrient analysis (165/196, 84.2%), personalization (151/196, 77.1%), insulin bolus calculation (145/196, 74%), and health care professional support (135/196, 68.8%). Among the 16 apps used by participants, most (12/16, 75%) supported nutrient analysis but only one offered bolus calculations or health care professional support, and none offered personalization. Users on injections reported greater barriers to blood glucose monitoring for insulin adjustments compared to exclusive pump users (mean score of 3.87, SD 1.22 vs mean 3.30, SD 1.28; *P*=.001). They also expressed higher needs for meal logs in an electronic food journal (mean 4.06, SD 1.18 vs mean 3.69, SD 1.17; *P*=.01), bolus dose suggestions (mean 4.37, SD 0.98 vs mean 3.84, SD 1.26; *P*=.001), and app personalization (mean 4.47, SD 0.86 vs mean 3.93, SD 1.21; *P*<.001). No significant differences were observed based on age or app use. The thematic analysis revealed participants' perceptions of suggested barriers and features, as well as new barriers such as calculation errors from unreliable food data and nutrition labels, fear of eating disorders, limited app reliability, and insufficient health care support, with suggestions for technology-based solutions.

Conclusions: CC mobile apps currently used do not meet the needs of people with T1D. A novel CC app with app features such as photo recognition, reliable nutrient values, and personalized bolus calculations could reduce the CC burden.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63278) doi:10.2196/63278

KEYWORDS

type 1 diabetes; carbohydrate counting; mobile apps; photo recognition; diabetes management; mHealth; mobile health; user-centered design; Canadian survey

Introduction

People living with type 1 diabetes (T1D; approximately 300,000 in Canada) need to be administered subcutaneous insulin to maintain stable blood glucose (BG) levels and prevent long-term complications [1,2]. Instead of a fixed bolus insulin regimen tied to predetermined meal plans, carbohydrate counting (CC) offers dietary flexibility by empowering individuals to tailor their mealtime insulin based on the carbohydrate content of their meals [3]. However, adherence to CC can be burdensome and difficult [4]. CC requires frequent BG monitoring, keeping food records, reading food nutrition labels, and weighing food portions. Additionally, calculations are time-consuming and prone to errors as they involve multiple factors such as the insulin sensitivity factor, insulin-to-carbohydrate ratio, and BG target [5,6]. Accurate CC is important because greater precision in carbohydrate counting improves insulin dosing accuracy and glycemic management, leading to lower glycated hemoglobin levels [7]. Further, greater differences in carbohydrate estimates are associated with higher glycemic variability [8]. This is significant considering that frequent or large glucose fluctuations may independently contribute to diabetes-related complications [9].

Automated CC through health technology introduces novel care solutions that can simplify CC for individuals living with T1D. The efficacy of smartphone or tablet apps in improving self-management among adults with diabetes is well-established [10-12]. The use of automated carbohydrate estimation compared to conventional methods (ie, manual calculations) led to improved accuracy in CC, reduced time spent in hyperglycemia, and improved BG variability [13,14]. However, the use of such apps is limited beyond research settings [15,16]. According to a survey, young participants lack awareness and have skepticism about the effectiveness of diabetes management apps, which have been shown to be reasons for not using them [16].

This highlights the necessity of involving people with T1D in the design process to create a more effective and user-centered app design that meets their needs. This is particularly important for adolescents, considering the unique psychological and social challenges they face [17]. Consequently, they may struggle to maintain motivation for self-care behaviors such as CC, as competing priorities often take precedence over the numerous demands of T1D management [18]. This challenge is compounded by the increased glycemic levels disproportionately observed from adolescence through young adulthood (approximately aged 14-24 years) [19]. Recognizing the vulnerability of this developmental stage, during which adolescents and young adults must develop new skills and competencies, it is essential to include adolescents in the development of new resources to gain insights into their specific needs [20].

To understand how a CC app can facilitate CC at mealtimes for people with T1D, the primary objectives of our study are to identify (1) the perceived barriers to CC and (2) the app features that would help reduce these barriers. The secondary objective is to compare apps being used by participants with the suggested app features to identify gaps for a novel CC app that aligns with user preferences and needs. We hypothesized that the needs of people with T1D regarding CC are not met with traditional manual tools or with the currently used apps.

Methods

Ethical Considerations

We received ethics approval for this study from the McGill Research Ethics Board (22-08-054-02). Participants were informed through the web-based platform (LimeSurvey GmbH) before accessing the survey, about its estimated length, the type of data collected, where the data would be stored, the duration of storage, the investigator's identity, and the purpose of the study. Informed consent was obtained before participation. All data collected were deidentified to ensure privacy and confidentiality and were hosted on secure McGill servers. No compensation was offered for completing the survey.

Participants

A national cross-sectional web-based survey was launched across Canada. Recruitment took place from November 2022 to November 2023 and used a nonprobability convenience sample. We sent an announcement through the newsletter to participants of the BETTER Canadian registry of people living with T1D (more than 3800 participants) [21] and through Canadian diabetes organizations and social media. People with T1D diagnosis, aged 14 years and older, and living in Canada, were eligible to complete a 30-closed and 9-open-ended question survey. Adolescent and adult participants provided digital consent before accessing the open voluntary survey. Parents who completed the questionnaire on behalf of their child with T1D were excluded as the aim was to gather insights and perceptions directly from individuals living with T1D.

Questionnaire and Data Collection

All data were self-reported. The survey was developed and adapted by researchers of T1D, certified diabetes educators, and health care professionals on the team, and subsequently tested by 3 patient-partners (ie, individuals with lived experience of type 1 diabetes who collaborate in research, to ensure a clear understanding of the questions). Questions related to barriers in CC were based on a literature review and proposed app features were based on the scientific literature as well as a manual search or testing of iOS and Android diabetes management apps. A list of identified barriers and proposed app features was compiled, and consensus was reached among researchers (AH and ASB) for the final items included in the

XSL•FO
RenderX

survey. The survey was created, in both English and French, on LimeSurvey hosted on McGill secure servers (LimeSurvey GmbH). The final questionnaire contained 13 demographic and diabetes management–related nonrandomized questions, 19 CC needs assessment questions and app use, and 7 questions on patient-provider communications, distributed over 6 pages or screens (Multimedia Appendix 1). Adaptive questioning was used to reduce the number and complexity of the questions. Respondents were able to review and change their answers before submission. The survey design ensured single-entry responses by disabling token-based response persistence and prohibiting multiple or updated responses with the same token. Only the completed questionnaires were considered for analysis. Respondents rated suggested barriers and app features on a 5-point Likert scale. The percentage of agreement was calculated based on the frequency of ratings that were either a 4=agree or a 5=strongly agree. Suggested app features were cross-referenced with the features of apps reported by participants to identify gaps and inform the development of a novel CC app aligned with user preferences and needs.

Statistical Analyses

All analyses were performed using IBM SPSS Statistics (version 27; IBM Corp). Variables are reported as frequencies, median with IQR, or mean (SD) was computed where indicated for scale variables. In addition to descriptive statistics for presenting barriers and app feature preferences, statistical analyses were conducted to identify any significant differences among respondents based on age (youths aged 14-24 years vs adults older than 24 years), consistent app use (every day or most of the days vs some of the days or rarely), and insulin administration modality (on injections or sometimes on an insulin pump vs on an insulin pump exclusively). A mean score out of 5 was calculated for each barrier and app feature and treated as a continuous variable to assess mean differences using either a 2-tailed *t* test or a nonparametric test, as appropriate. Open-ended questions were analyzed by 2 independent researchers (AH and AK), both with experience in diabetes care as clinicians (AH is a registered dietitian and AK is an MD student) and as student researchers. Their familiarity with the challenges faced by individuals with T1D and in conducting CC informed their interpretation of the open-ended responses. MAXQDA software (version 24; VERBI GmbH) was used to assist with the inductive thematic analysis and coding process. Individually, each researcher created a codebook, and then each code was discussed until agreement (100%) was achieved.

Results

Overview

A final sample of 196 eligible respondents provided consent and completed all the questions of the survey. The majority identified as women (n=145, 74%) and Caucasian (n=184, 93.9%). The median (IQR) age was 38 (24-54) years with just over a quarter of participants being youths aged 14-24 years (n=52, 26.5%). Over half of the respondents (n=110, 56.1%) use an insulin pump to administer insulin and almost all participants (n=191, 97.4%) use continuous glucose monitoring (CGM) systems (Table 1). Most respondents (n=178, 90.8%) use CC to determine mealtime bolus insulin and more than half (n=133, 67.9%) find it difficult to manage BG levels around mealtime with CC. Although the majority (n=147, 7%) believe apps could ease the CC burden, only 57 (29.1%) use apps to help with CC, and less than half of them (27/57, 47.4%) are satisfied with the apps used. A total of 16 unique apps were reported as being used by the respondents (Table 2). The most frequently used app was “MyFitnessPal,” reported by 17 of the 57 participants. This was followed by “Carbs & Cals,” “BolusCalc,” and “CalorieKing,” each used by 3 respondents.

XSL•FO
RenderX

Table 1. Participant (n=196) characteristics.

Characteristics	Values
Demographics	
Age (years), median (IQR)	38 (24-54)
Youths (14-24), n (%)	52 (26.5)
Adults (25 and older), n (%)	144 (73.5)
Gender[a] (women), n (%)	145 (74)
Born in Canada, n (%)	176 (89.8)
Ethnicity, n (%)	
Asian	4 (2)
Arab	3 (1.5)
Black (African, African American, and Caribbean)	1 (0.5)
Latin American	2 (1)
White or Caucasian	184 (83.9)
Other (Iberian and mixed)	2 (1)
Currently studying, n (%)	47 (24)
Highest level of education acquired, n (%)	
High school level	40 (20.4)
Cégep, vocational, or community college	58 (29.6)
University level	98 (50)
Medical history and diabetes management, median (IQR)	
Diabetes duration (years)	19 (12-32)
Had at least 1 diabetes-related consultation in the last year with health care professionals, n (%)	
Medical specialist (endocrinologist, pediatrician, and internist)	185 (94.4)
Family doctor or general practitioner	97 (49.5)
Registered dietitian or nutritionist	69 (35.3)
Registered nurse	64 (32.6)
Other (psychologist, dentist, or ophthalmologist)	50 (25.5)
Use continuous glucose monitoring systems in the last 12 months, n (%)	191 (97.4)
Use insulin pens or syringes exclusively, n (%)	86 (43.9)
Use an insulin pump, n (%)	110 (56.1)

[a]n=1 identifies as genderfluid, n=49 as men, n=1 missing..

[b]University certificate, bachelor's degree, master's, PhD, or MD.

[c]n=2 missing.

XSL•FO
RenderX

Table 2. Suggested app features were ranked by rate of agreement (4 or 5 on the Likert scale), then cross-referenced in each app (n=16) reported to be used by participants. The checkmarks indicate the features provided by the app.

Suggested app features	Values, n (%)	Carb Manager	Foodvisor	Lose it	MyFitnessPal	Cronometer	FatSecret	Fitwell	SNAQ[a]	Figwee	Carbs & Cals	Glycemeal	Keenoa	Bolus Calc	Calorie King	Glucicheck	MySugar
Viewing the food nutrition label	167 (85)	✓	✓	✓	✓	✓			✓						✓		
Meal composition quantification	165 (84)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
Personalization of the app	151 (77)																
Suggestion of insulin bolus doses	145 (74)													✓			
Unrestricted Support from health care professionals	135 (69)		✓														
Meal logs into a food journal	129 (66)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
Community building with app users	93 (47)	✓	✓	✓	✓		✓	✓									
Gamification	79 (40)	✓		✓	✓												

[a]SNAQ: Simplified Nutritional Appetite Questionnaire.

The open-ended survey responses were analyzed using inductive qualitative thematic analysis, identifying 22 themes categorized as barriers to CC and app features to address these barriers (Figure 1). Although some themes were survey options, participants also introduced new barriers and features. Emergent themes included calculation errors due to unreliable food data and nutrition labels, limited app reliability, unmet diverse user needs, and fear of eating disorders associated with food

XSL•FO
RenderX

hyperfixation. Concealing calorie information through technology was suggested as a way to reduce this hyperfixation. Participants highlighted minimal perceived benefits of CC on BG trends and limited health care professionals' support due to time constraints, lack of nutrition knowledge, and narrow clinical focus. Alternative strategies were brought forward to avoid CC, such as avoiding carbohydrates, using CGM to correct BG, eating the same thing, or using trial and error.

Figure 1. Emerging themes of reported barriers to carbohydrate counting (n=11) and app features (n=11) perceived to reduce these barriers, from the qualitative analysis of the open-ended questions. AI: artificial intelligence; BG: blood glucose; CC: carbohydrate counting; T1D: type 1 diabetes.

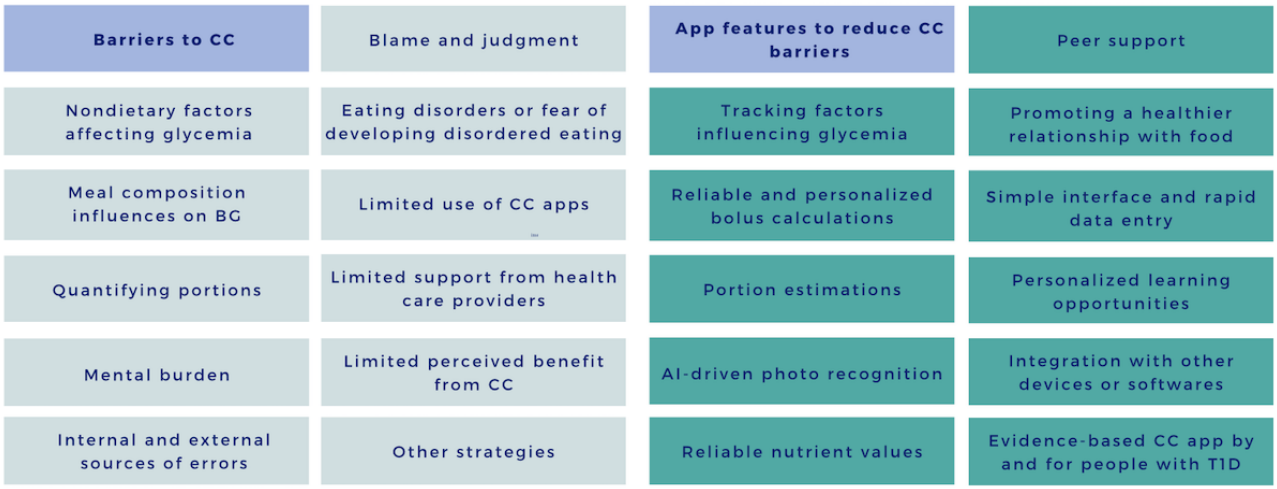

Barriers to CC

In terms of the barriers to CC that were suggested, identifying the number of carbohydrates to be consumed was the most agreed upon barrier whether it was in foods without labels (n=138, 70.4%; ie, rated scores of 4 or 5), in restaurants or when eating out (n=145, 74%), or when unsure about appetite (n=146, 74.4%; Figure 2). This was also reflected in participants' answers to the open-ended questions where quantifying portions was reported as a barrier for CC when eating homemade food or unknown ingredients, but also for food with labels where the serving size on nutrition labels was unclear or difficult to estimate. Conversely, counting CC in the presence of others or social stigma achieved higher rates of disagreement (n=56, 57.6%; ie, rated scores of 1 or 2). However, qualitative data analysis revealed a distinct perspective. Specifically, stigma manifested as feelings of blame and judgment. Participants expressed a fear of alienation, as well as apprehension about receiving judgment when eating sugary foods. It also included perceptions of stigma stemming from preoccupation with others' opinions, including health care professionals, and the fear of "ruining a good day" (Figure 1).

Figure 2. Stacked bar chart of participants' proportion of agreement with suggested barriers to carbohydrate counting.

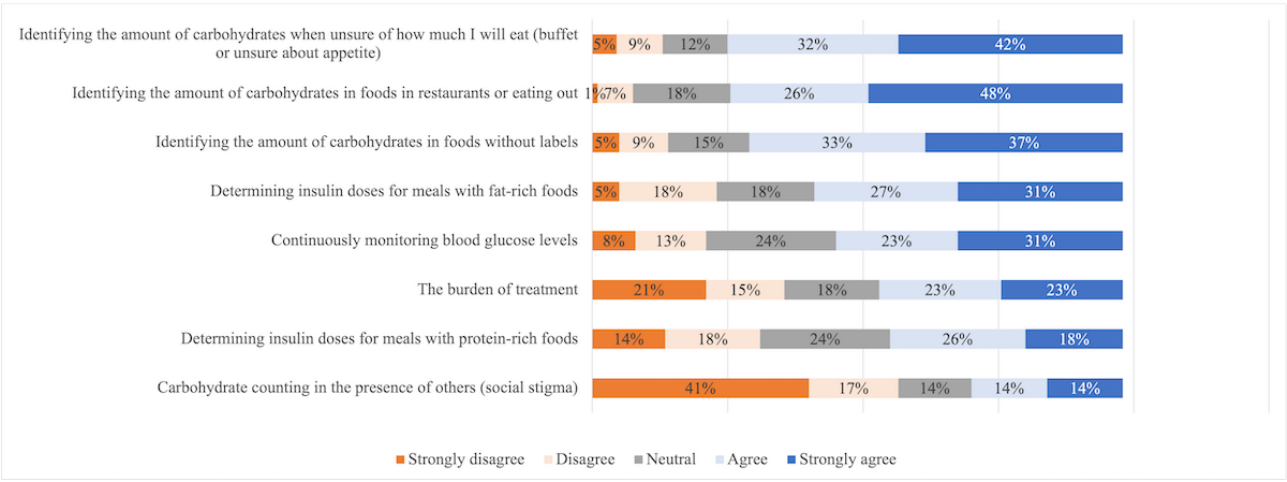

App Features to Reduce CC Burden

Suggested app features with the most agreements were the option to view the food nutrition label (n=167, 85.2%) and the quantification of meal composition (n=165, 84.2%). Participants expanded upon this in the open-ended responses by suggesting photo recognition, visual cues, comparing pictures to estimate quantities, or shifting toward the selection of small-medium-large portions rather than exact grams to reduce the burden. Beyond estimating quantity, emphasis was also placed on reliable nutrient value by suggesting the inclusion of the source of nutrient information and a breakdown of the carbohydrate estimations (Multimedia Appendix 2). On the other hand, gamification features such as earning points or trophies achieved the most disagreement among respondents (n=75, 38.3%; Table 3). To simplify meal logging and keeping

XSL•FO
RenderX

a food journal, the majority of respondents agreed on adding a "Favorites" section (n=168, 85.7%), allowing users to save frequently consumed food items for convenient access and photo recognition features (n=156, 79.2%). Additionally, a significant majority (n=166, 84.7%) agreed on including options to log factors influencing glycemia such as alcohol intake, physical activity, and medications (Table 3). In the open-ended responses, participants also noted stress, sick days, menstrual cycle, sleep, time, and site of injections, as well as other medical factors (menopause and gastroparesis; Figure 2). Factors that would increase users' trust in CC apps include transparency about safety mechanisms to prevent accidental overdosing of insulin or hypoglycemia (n=147, 75%) and endorsements based on validated clinical trial results (n=146, 74.5%; Table 2). Additionally, open-ended answers highlighted participants' emphasis on users' feedback, consistency in results obtained from the app, and improved glycemic management outcomes (Figure 1).

Table 3. Agreement rates for suggested app features and factors reducing CC burden.

Suggested app feature	Rate of agreement, n (%)				
	1=strongly disagree	2=disagree	3=neutral	4=agree	5=strongly agree
Features that would be beneficial to include in a carbohydrate counting app					
Gamification (earning points or trophies)	50 (25.5)	25 (12.8)	42 (21.4)	40 (20.4)	39 (19.9)
In-app community building	17 (8.7)	31 (15.8)	55 (28.1)	40 (20.4)	53 (27)
Meal logs into a food journal	11 (5.6)	15 (7.7)	41 (20.9)	53 (27)	76 (38.8)
Support from health care providers	13 (6.6)	13 (6.6)	35 (17.9)	63 (32.1)	72 (36.7)
Suggestion of insulin bolus doses	11 (5.6)	10 (5.1)	30 (15.3)	46 (23.5)	99 (50.5)
Personalization of the app	9 (4.6)	7 (3.6)	29 (14.8)	47 (24)	104 (53.1)
Quantification of meal composition	5 (2.6)	7 (3.6)	19 (9.7)	60 (30.6)	105 (53.6)
Viewing the food nutrition label	5 (2.6)	4 (2)	20 (10.2)	47 (24)	120 (61.2)
Features that would facilitate food journaling					
Reminders for forgotten logs	27 (13.8)	37 (18.9)	49 (25)	49 (25)	34 (17.3)
Photo recognition	6 (3.1)	10 (5.1)	24 (12.2)	39 (19.9)	117 (59.7)
Save meals in a "Favorite" section	5 (2.6)	4 (2)	19 (9.7)	53 (27)	115 (58.7)
Access recent meal items.	3 (1.5)	4 (2)	29 (14.8)	62 (31.6)	98 (50)
Combine food items into recipes	4 (2)	5 (2.6)	28 (14.3)	56 (28.6)	103 (52.6)
Track BG[a] levels and related meals or insulin	3 (1.5)	10 (5.1)	31 (15.8)	52 (26.5)	100 (51.0)
Include meal notes and postadjustments	4 (2)	8 (4.1)	37 (18.9)	63 (32.1)	84 (42.9)
Log factors influencing BG	4 (2)	4 (2)	22 (11.2)	64 (32.7)	102 (52)
Factors to increase my trust in a carbohydrate counting app					
Validated with clinical trials	3 (1.5)	10 (5.1)	37 (18.9)	57 (29.1)	89 (45.4)
Secure data storage	5 (2.6)	12 (6.1)	49 (25)	41 (20.9)	89 (45.4)
Information on safety mechanisms	5 (2.6)	12 (6.1)	32 (16.3)	63 (32.1)	84 (42.9)
Access formulas used for calculations	11 (5.6)	12 (6.1)	44 (22.4)	47 (24)	82 (41.8)
Qualifications of app developers	9 (4.6)	13 (6.6)	46 (23.5)	47 (24)	81 (41.3)
Health Canada's approval	8 (4.1)	11 (5.6)	39 (19.9)	50 (25.5)	88 (44.9)
Endorsed by the health care team	10 (5.1)	8 (4.1)	43 (21.9)	52 (26.5)	83 (42.3)

[a]BG: blood glucose.

App Features Comparison With Current Apps

Suggested app features were cross-referenced in each app (n=16) reported to be used by participants. It was found that, while most apps allow meal composition quantification, less than half display nutrition labels. Only 1 app calculated bolus dose, 1 app provided support from health care professionals, and none offered personalization for diabetes characteristics (Table 2).

Differences in Barriers and App Feature Preferences Among Users

Based on insulin administration modalities, the need to continuously monitor BG levels to adjust insulin ratios was a

XSL•FO
RenderX

more pronounced barrier among users on injections (mean score of 3.87, SD 1.22; ie, trending toward agreement) compared to those using insulin pumps exclusively (mean score of 3.30, SD 1.28; ie, trending toward neutral; *P*=.001). Additionally, users on injections expressed a greater need for specific features compared to exclusive pump users. This included the need for meal logs in an electronic food journal (mean score of 4.06, SD 1.18 vs mean 3.69, SD 1.17; *P*=.01), suggestions for bolus doses (mean 4.37, SD 0.98 vs mean 3.84, SD 1.26; *P*=.001), and app personalization (mean 4.47, SD 0.86 vs mean 3.93, SD 1.21; *P*<.001).

There were no significant differences in barriers or app feature preferences based on age or app use.

Discussion

Principal Findings

Participants reported carbohydrate identification barriers, nutrient interaction, and insulin dose calculation barriers, as well as psychosocial barriers. App features like photo recognition paired with nutrient values derived from validated databases, personalized bolus calculations, and tracking factors affecting BG levels were perceived to facilitate CC. Existing CC apps, used by people with T1D, fall short of meeting their preferred app features.

Comparison With Prior Work

Martinez et al [22] found that of the 80 existing apps for people with T1D, none met the criteria of an "ideal" app, suggesting a significant gap between what users want and what is available. App features for general T1D management, such as access to CGM and insulin pump records, personalization through goal setting and tailored recommendations, as well as insulin bolus calculations [22] were also reiterated by participants in this survey. On the other hand, reminders and rewards, that is, gamification features did not obtain major agreement in this survey despite being present in the literature as key features for diabetes management technology [22]. Several novel app features were suggested by participants and are worth noting. In fact, the fear of developing eating disorders was mentioned as a barrier to CC. The hyperfixation on food can lead to a heightened risk of developing disordered eating behaviors among people with T1D and is a well-established risk factor in the literature [23]. As such, concealing calorie information was suggested to promote a healthier relationship with food using technology. This perspective aligns with existing literature that highlights the importance of automated processes, which demand less input from the individual, as a means to alleviate the burden of treatment in T1D. A review demonstrated that the use of automated insulin delivery systems may reduce food management burden, by correcting CC inaccuracy, particularly when combined with adjunctive therapies (eg, GLP-1 receptor agonists and SGLT2 inhibitors) [24,25]. Although automated insulin delivery systems are at the forefront of diabetes research, fully closed-loop systems that do not require CC remain in development and are not yet available for people with T1D [26]. Additionally, factors such as future costs and user preferences underscore the need to provide alternative solutions, particularly accessible, comprehensive, and customizable CC apps. Meal entry through photo recognition could alleviate the cognitive burden on individuals with T1D during mealtimes and potentially facilitate full automation. This would enable users to receive bolus dose suggestions simply by taking a picture of their meal, without needing to process additional information. However, it is important to provide users with the option to access breakdowns of carbohydrate estimations and calculations upon request, as it was found to increase their trust in a CC app. Meal entry through photo recognition would also decrease the time required for CC and could potentially alleviate associated burdens. Participants expanded on the mental strain linked with CC, describing it as time-consuming, challenging to log multiple small snacks and corrections between meals, and restrictive when "not being able to eat right away." Another key barrier that was introduced was the limited use of CC apps. Beyond the 16 apps reported as being used by respondents, other available apps did not meet the reported preferred features. Although app-based insulin bolus calculators are available, they lack direct carbohydrate estimation and still require manual entry of meal composition or carbohydrate counts, meaning users must input this information manually rather than having it automatically recognized [5,27-31]. Separate apps can automatically count carbohydrates using photo recognition and food databases, but they do not incorporate bolus calculators [14,15,32]. This underscores participants' perception that current CC apps lack comprehensiveness. Moreover, existing apps frequently contain unreliable values [33], which can be attributed to users contributing information to the databases or relying on values from foreign databases. Participants reiterated this concern, highlighting unreliable food databases as external sources of errors. Another barrier to CC was the associated financial burden. As such, there may also be cost-related limitations when using CC apps. The apps used by participants all offered a free version, despite the option to upgrade to a premium version.

While the benefits of CC are well established in the literature [7,8], respondents perceived a limited use of CC when it comes to explaining BG levels and glycemic trends due to the multiple dietary and nondietary factors influencing BG levels. Similarly, a Canadian survey reported similar findings and revealed that 78% of the respondents (n=180) agreed that BG levels fluctuate even with appropriate CC, complicating diabetes management [34]. The most significant barrier reported was the difficulty in accurately identifying and counting carbohydrates. In fact, several research studies documented no significant changes in CC accuracy despite receiving education [4,35]. Additionally, more than half of this survey's respondents found determining insulin doses for high-fat meals to be a barrier to CC. Managing the delayed impact of fat on postprandial BG excursions is a well-established challenge in the literature [36-38]. Although clinical guidelines recommend meal-time insulin dose adjustments to mitigate glucose variability due to the impact of dietary fat [39-41], there is no consensus as to which insulin strategies to use. Strategies used to manage glycemic excursions following high-fat meals in T1D are numerous [42] but substantial interindividual differences exist in insulin dose requirements for fat and individualized advice based on postprandial BG monitoring for multiple hours afterward is required [43]. As such, the use of a CC app offering the ability

XSL•FO
RenderX

to track meal composition, as well as other nondietary influencing factors can help people with T1D understand BG levels and allow for personalized learning opportunities. Furthermore, machine learning models that use personalized insights from tracking apps could offer new opportunities for insulin dosage recommendations that account for protein and fat intake. Although the suggested psychosocial barriers received a lower agreement rate, they are still noteworthy as they were brought up in a different light by participants in the open-ended responses. Although CC in the presence of others is not as consequential, the fear of receiving judgment from others or feeling alienated, especially when eating sugary foods, was perceived as a barrier to CC. This was also demonstrated in a cross-sectional analysis of people with T1D perceiving stigma mostly as blame and judgment, including when eating sugary foods [44]. Only 69% of participants reported an agreement with app features providing support from health care professionals. The interpretation of this finding can be informed by qualitative analysis shedding light on stigma perception related to interactions with health care providers. Participants expressed feeling judged based on their BG values, which generated a fear of "ruining a good day," further complicating their condition management. This sentiment aligns with barriers mentioned by participants regarding the limited support from their health care team, who heavily focused on glycated hemoglobin levels rather than overall health and had limited knowledge of nutrition. They expressed the need for personalized learning opportunities that could be provided through CC apps by tracking their own trends and accessing educational modules while maintaining their privacy.

A final consideration is the potential for these features to improve CC and overall glycemic management while reducing the burden of CC. Given the challenges with CC around mealtimes, the need for technological solutions is evident. A recent meta-analysis, published in 2023, examining the effects of smartphone apps on glycemic control in youths with T1D found that apps with insulin or carbohydrate calculators were effective in reducing glycated hemoglobin (HbA_{1c}) levels [45]. In addition to improving accuracy, CC apps can help individuals understand their glycemic trends and how various factors impact their BG levels. Such an app can empower users to learn about their glycemic patterns and take control of their diabetes management, particularly in contexts where patient-practitioner communication may be limited. A CC app can streamline access to relevant information for health care professionals while enabling users to track and share key details for more meaningful discussions about their care.

Strengths and Limitations

Our results highlight the perceived barriers and preferred app features among individuals with T1D and should be interpreted with caution due to certain limitations. First, this study faced selection bias due to convenience sampling rather than random selection. Participants who completed the survey were likely more motivated, which may have resulted in an underrepresentation of certain barriers and app feature preferences. This limits the generalizability of the results, as the findings may not accurately reflect the diverse experiences of all individuals with T1D. Further, the homogeneous sample, with around 84% of the respondents identifying as Caucasian is not representative of most people with T1D and prevents the generalizability of the results. Nevertheless, 16% of this survey's respondents self-identified as Arabs, Asians, Blacks, and Latin Americans, among other ethnicities, as compared to 26% of Canadians self-identified as belonging to visible minorities according to the 2021 census by Statistics Canada [46]. Further, participant recruitment was extended to include adolescents as well as adults. This allowed including the perceptions of young adults who are transitioning to assuming responsibility for managing their diabetes from caregivers. Second, while surveys as a research method impose limitations on participants' responses and feedback, this was mitigated by incorporating open-ended questions. The thematic analysis provided a deeper understanding of participants' perceptions. Third, patient-reported outcomes brought additional value to our data and shed light on patient perspectives. Further qualitative studies with in-depth interviews could provide a better understanding of the diverse experiences of people with T1D, especially considering variations in treatment plans and differences between individuals using CGMs versus self-monitoring of BG levels, as most respondents in this survey were CGM users.

Future Directions

This study highlights the challenges faced by people with T1D and areas for improvement in diabetes management apps. Effective apps need comprehensive, automated CC features like photo recognition, easy access to frequently consumed meals, reliable nutrient values, and personalized bolus calculations. Personalization options, such as insulin administration modalities and individual factors affecting BG levels, are also essential for a user-friendly experience. An app incorporating these elements and validated food databases would be unique and novel. Ongoing end-user engagements in development and testing are essential for the high usability of the app.

Acknowledgments

The authors would like to acknowledge the Cardiometabolic Health, Diabetes, and Obesity Research Network for the mentorship award to ASB, as well as the contribution of patient partners involved in this research project. The material in this manuscript has previously been presented at the 17th International Conference on Advanced Technologies & Treatments for Diabetes (Florence, Italy; March 6-9, 2024) and the Vascular Conference (Montreal, Canada; October 25-29, 2023). This work was supported by Mitacs (grant number IT31370 ACC).

XSL•FO
RenderX

Data Availability

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AH conceptualized the study, developed the protocol, consolidated the survey, conducted a formal analysis, wrote the original draft, and reviewed and edited the manuscript. AK participated in the analysis, contributed to the original draft, and reviewed and edited the manuscript. LJB, AS, and AF participated in the analysis and reviewed and edited the manuscript. ARF consolidated the survey and reviewed and edited the manuscript. MN reviewed and edited the manuscript. ASB conceptualized the study, developed the protocol, consolidated the survey, supervised the project, and reviewed and edited the manuscript. All authors reviewed the final manuscript. ASB is the principal investigator and guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

ASB received research funding from the Canadian Institutes for Health Research, Mitacs, Diabète Québec, and Fonds de recherche du Québec-Santé, as well as speaker fees from Dexcom.

Multimedia Appendix 1
Survey questions.
[DOCX File , 25 KB - jmir_v27i1e63278_app1.docx]

Multimedia Appendix 2
Theme description derived from the qualitative analysis.
[DOCX File , 18 KB - jmir_v27i1e63278_app2.docx]

References

1. 2023 Backgrounder. Diabetes in Canada. 2021. URL: https://www.diabetes.ca/DiabetesCanadaWebsite/media/Advocacy-and-Policy/Backgrounder/2023_Backgrounder_Canada_English.pdf [accessed 2023-11-18]
2. Diabetes Canada Clinical Practice Guidelines Expert Committee, McGibbon A, Adams L, Ingersoll K, Kader T, Tugwell B. Glycemic management in adults with type 1 diabetes. Can J Diabetes 2018;42:S80-S87. [doi: 10.1016/j.jcjd.2017.10.012] [Medline: 29650115]
3. Tascini G, Berioli M, Cerquiglini L, Santi E, Mancini G, Rogari F, et al. Carbohydrate counting in children and adolescents with type 1 diabetes. Nutrients 2018;10(1):109 [FREE Full text] [doi: 10.3390/nu10010109] [Medline: 29361766]
4. Spiegel G, Bortsov A, Bishop FK, Owen D, Klingensmith GJ, Mayer-Davis EJ, et al. Randomized nutrition education intervention to improve carbohydrate counting in adolescents with type 1 diabetes study: is more intensive education needed? J Acad Nutr Diet 2012;112(11):1736-1746 [FREE Full text] [doi: 10.1016/j.jand.2012.06.001] [Medline: 22975086]
5. Singla R, Bindra J, Singla A, Gupta G, Gupta Y, Aggarwal S. Automation of insulin bolus dose calculation in type 1 diabetes: a feasibility study. Int J Diabetes Dev Ctries 2022;43(1):66-71. [doi: 10.1007/s13410-022-01054-7]
6. Sussman A, Taylor EJ, Patel M, Ward J, Alva S, Lawrence A, et al. Performance of a glucose meter with a built-in automated bolus calculator versus manual bolus calculation in insulin-using subjects. J Diabetes Sci Technol 2012;6(2):339-344 [FREE Full text] [doi: 10.1177/193229681200600218] [Medline: 22538144]
7. Mehta SN, Quinn N, Volkening LK, Laffel LM. Impact of carbohydrate counting on glycemic control in children with type 1 diabetes. Diabetes Care 2009;32(6):1014-1016 [FREE Full text] [doi: 10.2337/dc08-2068] [Medline: 19244089]
8. Brazeau AS, Mircescu H, Desjardins K, Leroux C, Strychar I, Ekoé JM, et al. Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. Diabetes Res Clin Pract 2013;99(1):19-23. [doi: 10.1016/j.diabres.2012.10.024] [Medline: 23146371]
9. Suh S, Kim JH. Glycemic variability: how do we measure it and why is it important? Diabetes Metab J 2015;39(4):273-282 [FREE Full text] [doi: 10.4093/dmj.2015.39.4.273] [Medline: 26301188]
10. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
11. Liu L, Ogwu S. A meta-analysis of mobile health and risk reduction in patients with diabetes mellitus: challenge and opportunity. J Mobile Technol Med 2012;1(3):17-24. [doi: 10.7309/jmtm.18]
12. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
13. Bally L, Dehais J, Nakas CT, Anthimopoulos M, Laimer M, Rhyner D, et al. Carbohydrate estimation supported by the GoCARB system in individuals with type 1 diabetes: a randomized prospective pilot study. Diabetes Care 2017;40(2):e6-e7. [doi: 10.2337/dc16-2173] [Medline: 27899490]

XSL•FO
RenderX

14. Alfonsi JE, Choi EEY, Arshad T, Sammott SS, Pais V, Nguyen C, et al. Carbohydrate counting app using image recognition for youth with type 1 diabetes: pilot randomized control trial. JMIR Mhealth Uhealth 2020;8(10):e22074. [doi: 10.2196/22074] [Medline: 33112249]
15. Vasiloglou MF, Marcano I, Lizama S, Papathanail I, Spanakis EK, Mougiakakou S. Multimedia data-based mobile applications for dietary assessment. J Diabetes Sci Technol 2023;17(4):1056-1065 [FREE Full text] [doi: 10.1177/19322968221085026] [Medline: 35348398]
16. Trawley S, Browne JL, Hagger VL, Hendrieckx C, Holmes-Truscott E, Pouwer F, et al. The use of mobile applications among adolescents with type 1 diabetes: results from diabetes MILES Youth-Australia. Diabetes Technol Ther 2016;18(12):813-819. [doi: 10.1089/dia.2016.0233] [Medline: 27788032]
17. Hung LC, Huang CY, Lo FS, Cheng SF. The self-management experiences of adolescents with type 1 diabetes: a descriptive phenomenology study. Int J Environ Res Public Health 2020;17(14):5132 [FREE Full text] [doi: 10.3390/ijerph17145132] [Medline: 32708617]
18. Borus JS, Laffel L. Adherence challenges in the management of type 1 diabetes in adolescents: prevention and intervention. Curr Opin Pediatr 2010;22(4):405-411 [FREE Full text] [doi: 10.1097/MOP.0b013e32833a46a7] [Medline: 20489639]
19. Youth. United Nations. URL: https://www.un.org/en/global-issues/youth [accessed 2025-02-18]
20. Transitioning youth to an adult health care clinician. Got Transition. 2023. URL: https://www.gottransition.org/six-core-elements/transitioning-youth-to-adult/transition-readiness.cfm [accessed 2025-02-18]
21. Brazeau AS, Messier V, Talbo MK, Gagnon C, Taleb N, Fortier I, et al. Self-reported severe and nonsevere hypoglycemia in type 1 diabetes: population surveillance through the BETTER patient engagement registry: development and baseline characteristics. Can J Diabetes 2022;46(8):813-821. [doi: 10.1016/j.jcjd.2022.05.010] [Medline: 35835670]
22. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App features for type 1 diabetes support and patient empowerment: systematic literature review and benchmark comparison. JMIR mHealth uHealth 2018;6(11):e12237 [FREE Full text] [doi: 10.2196/12237] [Medline: 30463839]
23. Hanlan ME, Griffith J, Patel N, Jaser SS. Eating disorders and disordered eating in type 1 diabetes: prevalence, screening, and treatment options. Curr Diab Rep 2013 [FREE Full text] [doi: 10.1007/s11892-013-0418-4] [Medline: 24022608]
24. South CA, Talbo MK, Roy-Fleming A, Peters TM, Nielsen DE, Iceta S, et al. Does insulin delivery technology change our relationship with foods? A scoping review. Diabetes Technol Ther 2024;26(2):136-145. [doi: 10.1089/dia.2023.0382] [Medline: 38032855]
25. Srinivasan S, Ekhlaspour L, Cengiz E. Adjunctive therapies to optimize closed-loop glucose control. J Diabetes Sci Technol 2021;15(6):1243-1251 [FREE Full text] [doi: 10.1177/19322968211032701] [Medline: 34315267]
26. Templer S. Closed-loop insulin delivery systems: past, present, and future directions. Front Endocrinol 2022;13:919942 [FREE Full text] [doi: 10.3389/fendo.2022.919942] [Medline: 35733769]
27. Vallejo-Mora MDR, Carreira-Soler M, Linares-Parrado F, Olveira G, Rojo-Martínez G, Domínguez-López M, et al. The Calculating Boluses on Multiple Daily Injections (CBMDI) study: a randomized controlled trial on the effect on metabolic control of adding a bolus calculator to multiple daily injections in people with type 1 diabetes. J Diabetes 2017;9(1):24-33. [doi: 10.1111/1753-0407.12382] [Medline: 26848934]
28. Gonzalez C, Picón MJ, Tomé M, Pujol I, Fernández-García JC, Chico A. Expert Study: Utility of an automated bolus advisor system in patients with type 1 diabetes treated with multiple daily injections of insulin—a crossover study. Diabetes Technol Ther 2016;18(5):282-287. [doi: 10.1089/dia.2015.0383] [Medline: 26886163]
29. Ziegler R, Rees C, Jacobs N, Parkin CG, Lyden MR, Petersen B, et al. Frequent use of an automated bolus advisor improves glycemic control in pediatric patients treated with insulin pump therapy: results of the Bolus Advisor Benefit Evaluation (BABE) study. Pediatr Diabetes 2016;17(5):311-318. [doi: 10.1111/pedi.12290] [Medline: 26073672]
30. Karakuş K, Gokce T, Muradoğlu S, Can E, Eviz E, Mutlu GY, et al. My Friend Diabetes Carbohydrate-Bolus Calculator: user experiences of a bolus calculating carbohydrate counting mobile app for people with type 1 diabetes. Clin Diabetol 2023;12(1):53-59. [doi: 10.5603/DK.a2022.0066]
31. Eiland L, McLarney M, Thangavelu T, Drincic A. App-based insulin calculators: current and future state. Curr Diab Rep 2018;18(11):123. [doi: 10.1007/s11892-018-1097-y] [Medline: 30284645]
32. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2017;76(3):283-294. [doi: 10.1017/S0029665116002913] [Medline: 27938425]
33. Engel MG, J Kern H, Brenna J, Mitmesser SH. Micronutrient gaps in three commercial weight-loss diet plans. Nutrients 2018;10(1):108 [FREE Full text] [doi: 10.3390/nu10010108] [Medline: 29361684]
34. Fortin A, Rabasa-Lhoret R, Roy-Fleming A, Desjardins K, Brazeau A, Ladouceur M, et al. Practices, perceptions and expectations for carbohydrate counting in patients with type 1 diabetes—results from an online survey. Diabetes Res Clin Pract 2017;126:214-221. [doi: 10.1016/j.diabres.2017.02.022] [Medline: 28273644]
35. Gurnani M, Pais V, Cordeiro K, Steele S, Chen S, Hamilton JK. One potato, two potato,… assessing carbohydrate counting accuracy in adolescents with type 1 diabetes. Pediatr Diabetes 2018;19(7):1302-1308. [doi: 10.1111/pedi.12717] [Medline: 29999219]

XSL•FO
RenderX

36. Hanefeld M, Temelkova-Kurktschiev T. Control of post-prandial hyperglycemia—an essential part of good diabetes treatment and prevention of cardiovascular complications. Nutr Metab Cardiovasc Dis 2002;12(2):98-107. [Medline: 12189909]
37. Paterson MA, Smart CEM, Lopez PE, Howley P, McElduff P, Attia J, et al. Increasing the protein quantity in a meal results in dose-dependent effects on postprandial glucose levels in individuals with type 1 diabetes mellitus. Diabet Med 2017;34(6):851-854. [doi: 10.1111/dme.13347] [Medline: 28257160]
38. Smart CE, Evans M, O'Connell SM, McElduff P, Lopez PE, Jones TW, et al. Both dietary protein and fat increase postprandial glucose excursions in children with type 1 diabetes, and the effect is additive. Diabetes Care 2013;36(12):3897-3902. [doi: 10.2337/dc13-1195] [Medline: 24170749]
39. Lodefalk M, Aman J, Bang P. Effects of fat supplementation on glycaemic response and gastric emptying in adolescents with type 1 diabetes. Diabet Med 2008;25(9):1030-1035. [doi: 10.1111/j.1464-5491.2008.02530.x] [Medline: 19183308]
40. Pańkowska E, Błazik M, Groele L. Does the fat-protein meal increase postprandial glucose level in type 1 diabetes patients on insulin pump: the conclusion of a randomized study. Diabetes Technol Ther 2012;14(1):16-22. [doi: 10.1089/dia.2011.0083] [Medline: 22013887]
41. Smart CE, Annan F, Higgins LA, Jelleryd E, Lopez M, Acerini CL. ISPAD Clinical Practice Consensus Guidelines 2018: nutritional management in children and adolescents with diabetes. Pediatr Diabetes 2018;19:136-154. [doi: 10.1111/pedi.12738] [Medline: 30062718]
42. Smith TA, Marlow AA, King BR, Smart CE. Insulin strategies for dietary fat and protein in type 1 diabetes: a systematic review. Diabet Med 2021;38(11):e14641. [doi: 10.1111/dme.14641] [Medline: 34251692]
43. Bell KJ, Toschi E, Steil GM, Wolpert HA. Optimized mealtime insulin dosing for fat and protein in type 1 diabetes: application of a model-based approach to derive insulin doses for open-loop diabetes management. Diabetes Care 2016;39(9):1631-1634. [doi: 10.2337/dc15-2855] [Medline: 27388474]
44. Housni A, Katz A, Kichler JC, Nakhla M, Secours L, Brazeau A. Predictors of stigma perception by people with type 1 diabetes: a cross-sectional analysis of the BETTER registry. Diabetes Metab Syndr 2024;18(8):103112 [FREE Full text] [doi: 10.1016/j.dsx.2024.103112] [Medline: 39236506]
45. Pi L, Shi X, Wang Z, Zhou Z. Effect of smartphone apps on glycemic control in young patients with type 1 diabetes: a meta-analysis. Front Public Health 2023;11:1074946 [FREE Full text] [doi: 10.3389/fpubh.2023.1074946] [Medline: 37064701]
46. Visible minority and population group by generation status: Canada, provinces and territories, census metropolitan areas and census agglomerations with parts. Statistics Canada. 2023. URL: https://doi.org/10.25318/9810032401-eng [accessed 2023-10-26]

Abbreviations

BG: blood glucose
CC: carbohydrate counting
CGM: continuous glucose monitoring
T1D: type 1 diabetes

Edited by N Cahill; submitted 17.06.24; peer-reviewed by V Astha, S Mehta, E Eviz, KE Karakus, A Jafar; comments to author 29.10.24; revised version received 07.11.24; accepted 30.12.24; published 28.03.25.

Please cite as:
Housni A, Katz A, Bergeron LJ, Simard A, Finkel A, Roy-Fleming A, Nakhla M, Brazeau AS
Bridging the Gap in Carbohydrate Counting With a Mobile App: Needs Assessment Survey
J Med Internet Res 2025;27:e63278
URL: https://www.jmir.org/2025/1/e63278
doi:10.2196/63278
PMID:

©Asmaa Housni, Alexandra Katz, Lucien Junior Bergeron, Alain Simard, Ashley Finkel, Amélie Roy-Fleming, Meranda Nakhla, Anne-Sophie Brazeau. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Contribution of Open Access Databases to Intensive Care Medicine Research: Scoping Review

Julien Kallout[1], MD, MSc; Antoine Lamer[2,3], PhD; Julien Grosjean[4,5], PhD; Gaétan Kerdelhué[4,5], BSc; Guillaume Bouzillé[6], MD, PhD; Thomas Clavier[1,7], MD, PhD; Benjamin Popoff[1,6], MD, MSc

[1]Department of Anesthesiology and Critical Care, CHU Rouen, Rouen, France

[2]Univ Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de Santé et des Pratiques Médicales, Lille, France

[3]Fédération Régionale de Recherche en Psychiatrie et Santé mentale des Hauts-de-France, Saint-André-lez-Lille, France

[4]Department of Biomedical Informatics, CHU Rouen, Rouen, France

[5]LIMICS U1142, Sorbonne université & Sorbonne Paris Nord, Paris, France

[6]CHU Rennes, INSERM, LTSI-UMR 1099, Univ Rennes, Rennes, France

[7]UNIROUEN, INSERM U1096, Normandie Université, Rouen, France

Corresponding Author:
Julien Kallout, MD, MSc
Department of Anesthesiology and Critical Care
CHU Rouen
37 Boulevard Gambetta
Rouen, 76000
France
Phone: 33 2 32 88 89 90
Email: julien.kallout@chu-rouen.fr

Abstract

Background: Intensive care units (ICUs) handle the most critical patients with a high risk of mortality. Due to those conditions, close monitoring is necessary and therefore, a large volume of data is collected. Collaborative ventures have enabled the emergence of large open access databases, leading to numerous publications in the field.

Objective: The aim of this scoping review is to identify the characteristics of studies using open access intensive care databases and to describe the contribution of these studies to intensive care research.

Methods: The research was conducted using 3 databases (PubMed–MEDLINE, Embase, and Web of Science) from the inception of each database to August 1, 2022. We included original articles based on 4 open databases of patients admitted to ICUs: Amsterdam University Medical Centers Database, eICU Collaborative Research Database, High time resolution ICU dataset, Medical Information Mart for Intensive Care (II to IV). A double-blinded screening for eligibility was performed, first on the title and abstract and subsequently on the full-text articles. Characteristics relating to publication journals, study design, and statistical analyses were extracted and analyzed.

Results: We observed a consistent increase in the number of publications from these databases since 2016. The Medical Information Mart for Intensive Care databases were the most frequently used. The highest contributions came from China and the United States, with 689 (52.7%) and 370 (28.3%) publications respectively. The median impact factor of publications was 3.8 (IQR 2.8-5.8). Topics related to cardiovascular and infectious diseases were predominant, accounting for 333 (25.5%) and 324 (24.8%) articles, respectively. Logistic regression emerged as the most commonly used statistical model for both inference and prediction questions, featuring in 396 (55.5%) and 281 (47.5%) studies, respectively. A majority of the inference studies yielded statistically significant results (84.0%). In prediction studies, area under the curve was the most frequent performance measure, with a median value of 0.840 (IQR 0.780-0.890).

Conclusions: The abundance of scientific outputs resulting from these databases, coupled with the diversity of topics addressed, highlight the importance of these databases as valuable resources for clinical research. This suggests their potential impact on clinical practice within intensive care settings. However, the quality and clinical relevance of these studies remains highly heterogeneous, with a majority of articles being published in low–impact factor journals.

(J Med Internet Res 2025;27:e57263) doi:10.2196/57263

XSL•FO
RenderX

KEYWORDS

intensive care unit; ICU; big data; databases; open access; Amsterdam University Medical Centers Database; AmsterdamUMCdb; eICU Collaborative Research Database; eICU-CRD; database; screening; descriptive analysis

Introduction

Intensive care units (ICUs) provide care for critical patients at high risk of morbidity and mortality. These patients, due to their severity, require continuous monitoring and surveillance of clinical, biological, and imaging parameters. This generates a large amount of data, usually collected in electronic health records. Although collected for health care purposes, these data can also be used secondarily to address other objectives, which has become a key issue in recent years [1-3]. In addition to traditional epidemiological and clinical research, the secondary use of these databases allows the emergence of new research themes such as the development of diagnostic tools, decision support systems, or predictive models of therapeutic response [4,5]. However, due to the sensitive nature of health data and the technical, legal and ethical challenges, medical data are still difficult to access [6].

During the last decades, collaborative ventures have enabled the emergence of large open access databases [7]. Building on the opportunity of digital transformation, they facilitate data sharing on a large scale and enable knowledge creation more efficiently. Furthermore, they are part of an ecosystem in which science is more transparent, reproducible, and constitutes an effective lever for scientific integrity [8]. Since the 2000s, we have seen the release of several open access ICU databases [9]. The best-known example is the Medical Information Mart for Intensive Care (MIMIC) database, which integrates anonymized, comprehensive clinical data from more than 50,000 intensive care admissions from Beth Israel Deaconess Medical Center in Boston, Massachusetts [10,11]. These open databases have led to the production of numerous research works [12,13].

Currently, little effort has been made to analyze the medical literature generated from these open intensive care databases. Previous systematic reviews were interested in describing these databases to determine their exploitation potential [9] or focused on describing machine learning techniques used to train models from these databases [14]. However, none of them were interested in the research themes and their potential impact on clinical practice.

We propose a comprehensive synthesis of clinical research publications based on ICU open databases. A scoping review was used as the most suitable research methodology for mapping this research area [15]. Our objectives were to (1) examine the bibliometric characteristics and authorship patterns of these publications and (2) investigate the research themes and methodologies used. The objective was to glimpse the contribution of these open databases in intensive care clinical research.

Methods

Overview

The protocol adheres to the reporting guidance provided in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [16] (Table S1 in Multimedia Appendix 1). The design of this study is conceived according to the Joanna Briggs Institute Reviewers' manual for evidence synthesis [17].

Eligibility Criteria

We selected studies based on the following criteria (Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- They were original articles based on open databases of patients admitted to intensive care units, covering all types of interventions and outcome measures.
- They were studies with a clinical aim (diagnosis, therapeutic, prognosis), where "clinical" pertains observations made on actual patients as opposed to theoretical, laboratory, or computer-based studies.

Exclusion criteria

- Studies on patients admitted outside of intensive care units.
- Studies where open databases were solely used for external validation.
- Studies not written in English.
- Systematic reviews and meta-analyses.

Search Strategy

First, the literature was searched for open access ICU databases. A recent review systematically identified publicly available adult clinical ICU databases [9]: Amsterdam University Medical Centers Database (AmsterdamUMCdb) [7], eICU Collaborative Research Database (eICU-CRD) [18], High time-resolution ICU dataset (HiRID) [19], MIMIC clinical databases version II, III, and IV [10,20,21].

We then used PubMed-Medline (US National Library of Medicine), Embase (Elsevier), and Web of science (Clarivate Analytics) to search articles published from the inception of each database to August 1, 2022. The search terms of the

XSL•FO
RenderX

different databases are included in Table S2 in Multimedia Appendix 1. A senior librarian [GK] developed and validated search queries. Finally, the references of the included articles were examined for potentially relevant medical articles that may have escaped the literature search.

Selection

Two independent reviewers, a senior intensivist [BP] and a junior intensivist [JK], conducted the search strategy and retrieved the references. An initial selection was made after examining the titles and abstracts of the medical articles resulting from the search strategy. The reviewers [BP and JK] performed an initial blind and independent selection of the articles. Subsequently, the reviewers met to reach a consensus on any disagreements. If the 2 primary reviewers could not reach a consensus, the third reviewer, a health data engineer (AL), intervened to resolve the disagreements through arbitration. The disagreements typically revolved around the eligibility criteria for study inclusion. The selection process was facilitated using the Rayyan software [22].

Data Extraction, Collection, and Analysis

Medical articles were referenced by their Digital Object Identifier, the name of the first author, and the date of publication. Two reviewers [BP and JK] independently extracted data using a custom electronic form designed specifically for this study. The form was pilot tested with a random sample of 20 articles to ensure its effectiveness. Upon achieving consistent data abstraction ($\kappa \geq 0.8$) [23], reviewers proceeded with data extraction for the entirety of included articles using the Goupile software [24].

Data for journal research fields and impact factors were obtained from Web of Science [25]. In addition, research topics were predefined and listed in the research protocol. Collected data details are outlined in Table S3 in Multimedia Appendix 1. Results were aggregated into a table, organized with 1 column per database and 1 row per variable.

A descriptive analysis was then carried out on the gathered data using R (version 4.1.2; R Foundation for Statistical Computing) [26]. Categorical variables were summarized as counts and frequencies per category, while quantitative variables were expressed as medians and interquartile ranges. Bar charts were used to visualize the distribution of categorical variables and histograms and density curves for the quantitative variables. In addition, the temporal evolution of the studied variables was analyzed.

Results

Study Selection

From the inception of these databases up to August 2022, a total of 4466 publications were identified. After excluding duplicates, 2063 (46.2%) titles and abstracts were retained. Among these articles, 1362 (66%) met the inclusion criteria. Ultimately, 1307 (96%) articles were selected for the final analysis after thorough reading. The selection flowchart is shown in Figure 1, with the characteristics of the included studies are detailed in Table 1.

XSL•FO
RenderX

Figure 1. Flowchart of scoping review. ICU: intensive care unit.

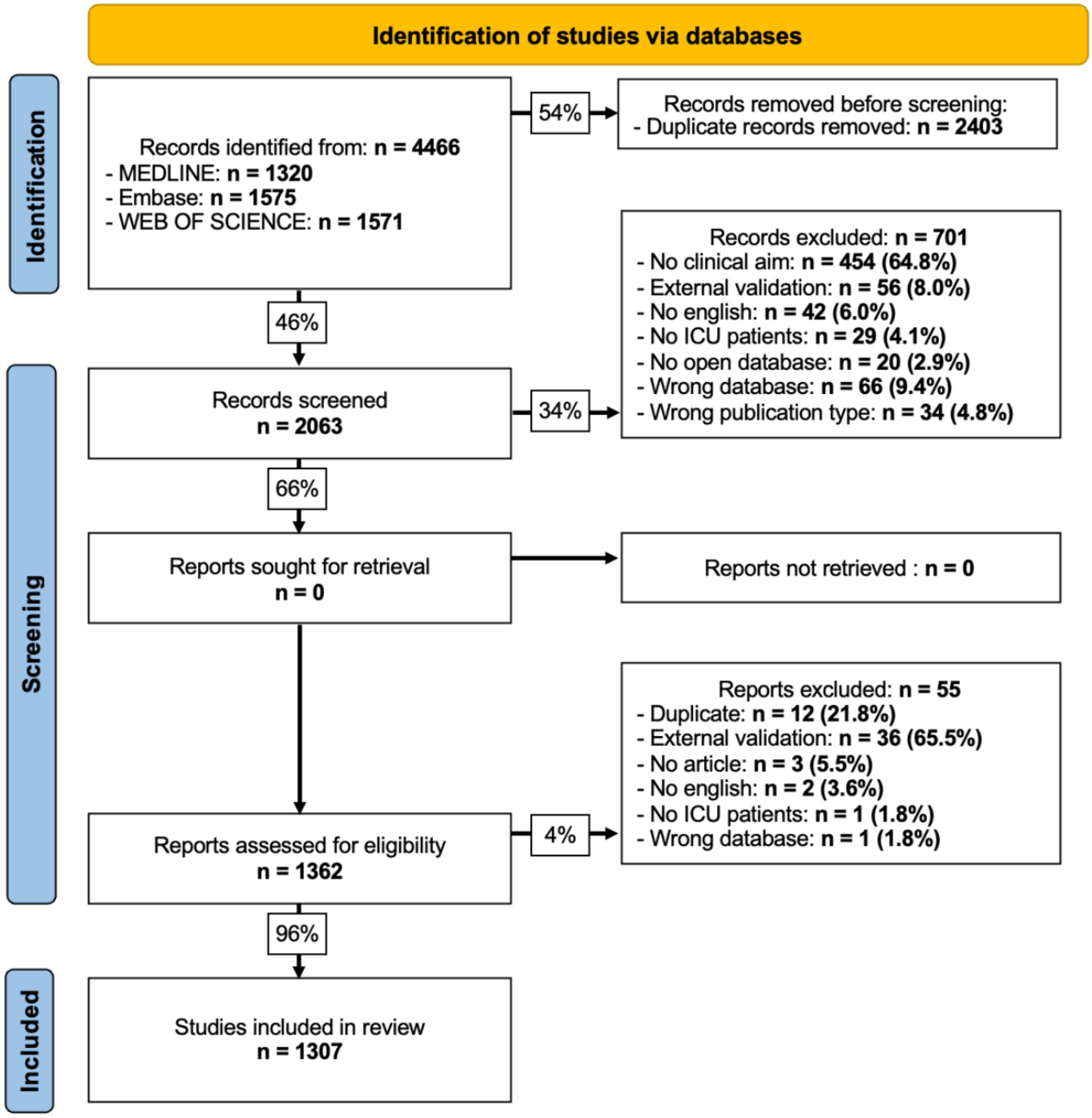

XSL•FO
RenderX

Table 1. Characteristics of studies.

Characteristics	Studies (N=1307)
Article information	
Date of publication, (year), n (%)	
2000-2010	6 (0.5)
2011-2015	38 (2.9)
2016-2020	296 (2.6)
2020-August 2022	967 (74.0)
Database used, n (%)	
AmsterdamUMCdb[a]	6 (0.5)
eICU-CRD[b]	196 (15.0)
HiRID[c]	0 (0)
MIMIC[d] II clinical database	107 (0.2)
MIMIC[d] III clinical database	905 (9.2)
MIMIC[d] IV clinical database	197 (15.1)
Number of databases used, n (%)	
1	1186 (91.5)
2	105 (0.1)
3	5 (0.4)
Corresponding author's sex, female, n (%)	352 (26.9)
Country of corresponding author (top 10), n (%)	
China	689 (2.7)
United States	370 (28.3)
United Kingdom	38 (3)
Canada	17 (1.3)
Germany	16 (1.2)
Japan	13 (1.0)
India	11 (0.8)
Australia	10 (0.8)
Singapore	10 (0.8)
South Korea	10 (0.8)
Spain	10 (0.8)
Other	112 (8.5)
Journal information	
Journal name (top 10), n (%)	
Frontiers in Medicine	52 (4.0)
Critical Care Medicine	41 (3.1)
American Journal of Respiratory and Critical Care Medicine	40 (3.1)
Frontiers in Cardiovascular Medicine	37 (2.8)
International Journal of General Medicine	36 (2.8)
Intensive Care Medicine Experimental	34 (2.6)
Scientific Reports	31 (2.4)
Critical Care	26 (2.0)

XSL•FO
RenderX

Characteristics	Studies (N=1307)
Plos One	25 (1.9)
Chest	24 (1.8)
Impact factor, median (IQR)	3.8 (2.8-5.8)
Study information, median (IQR)	
Inclusion period (years)	11 (11-11)
Number of participants	4282 (1468-12,740)

[a]AmsterdamUMCdb: Amsterdam University Medical Centers Database.

[b]eICU-CRD: eICU Collaborative Research Database.

[c]HiRID: High time-resolution ICU dataset.

[d]MIMIC: Medical Information Mart for Intensive Care.

Article Information

Since 2016, there was a consistent rise in the number of publications using these databases. Most studies were published after the year 2020, with 967 (74.0%) publications. The MIMIC databases were the most frequently used. No publications using the HiRID database were identified (Table 1, Figure 2). China and the United States were the leading countries in terms of number of publications, with 689 (52.7%) and 370 (28.3%) articles, respectively. European countries contributed to 109 (8.3%) publications, led by the United Kingdom leading with 39 (3.0%) publications (Table 1, Figure 3).

Figure 2. Evolution of the number of publications over the years. Blue line represents the total number of publications. AmsterdamUMCdb: Amsterdam University Medical Centers Database; eICU-CRD: eICU Collaborative Research Database; MIMIC: Medical Information Mart for Intensive Care.

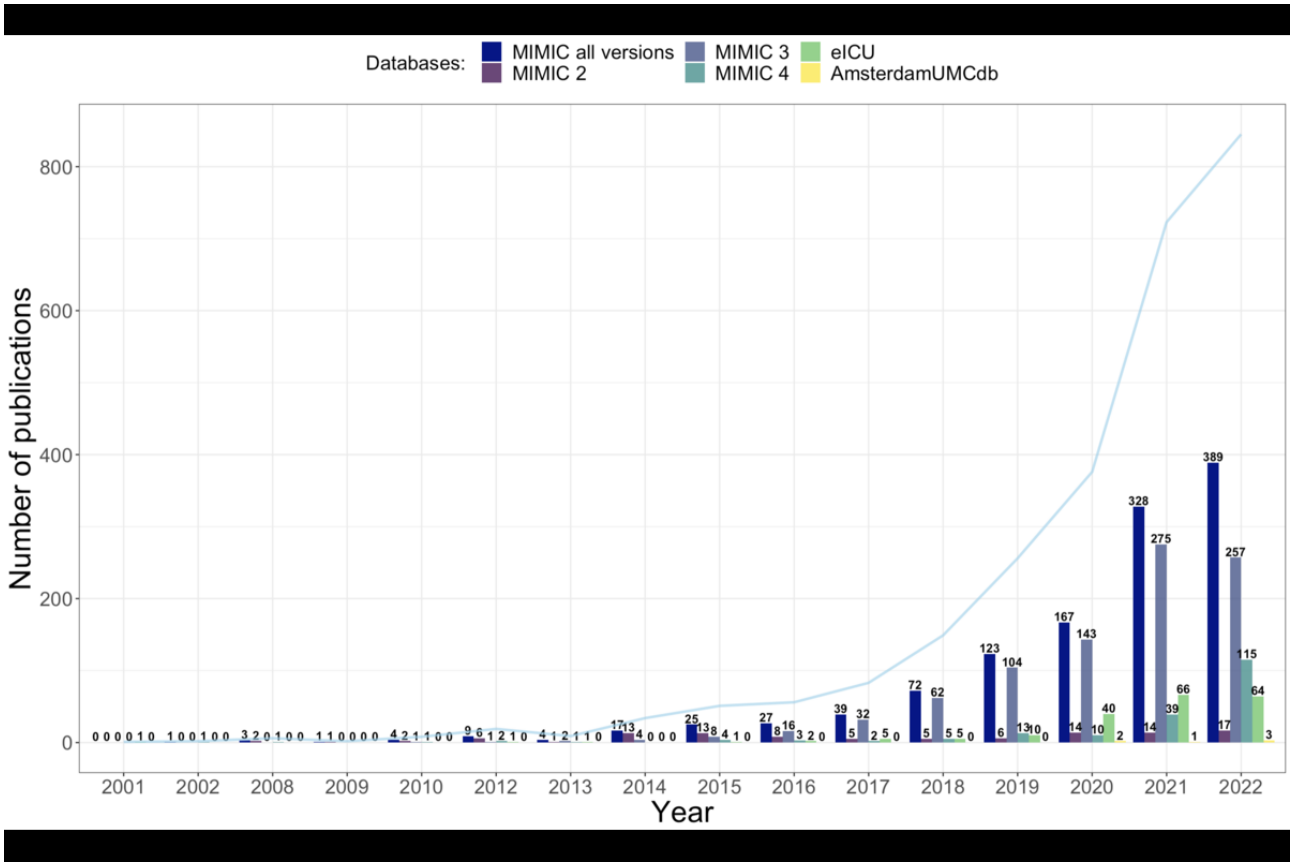

XSL•FO
RenderX

Figure 3. Worldwide publications number distribution. Red dots represent the number of publications per country. Country colors indicate the percentage of publications for each country.

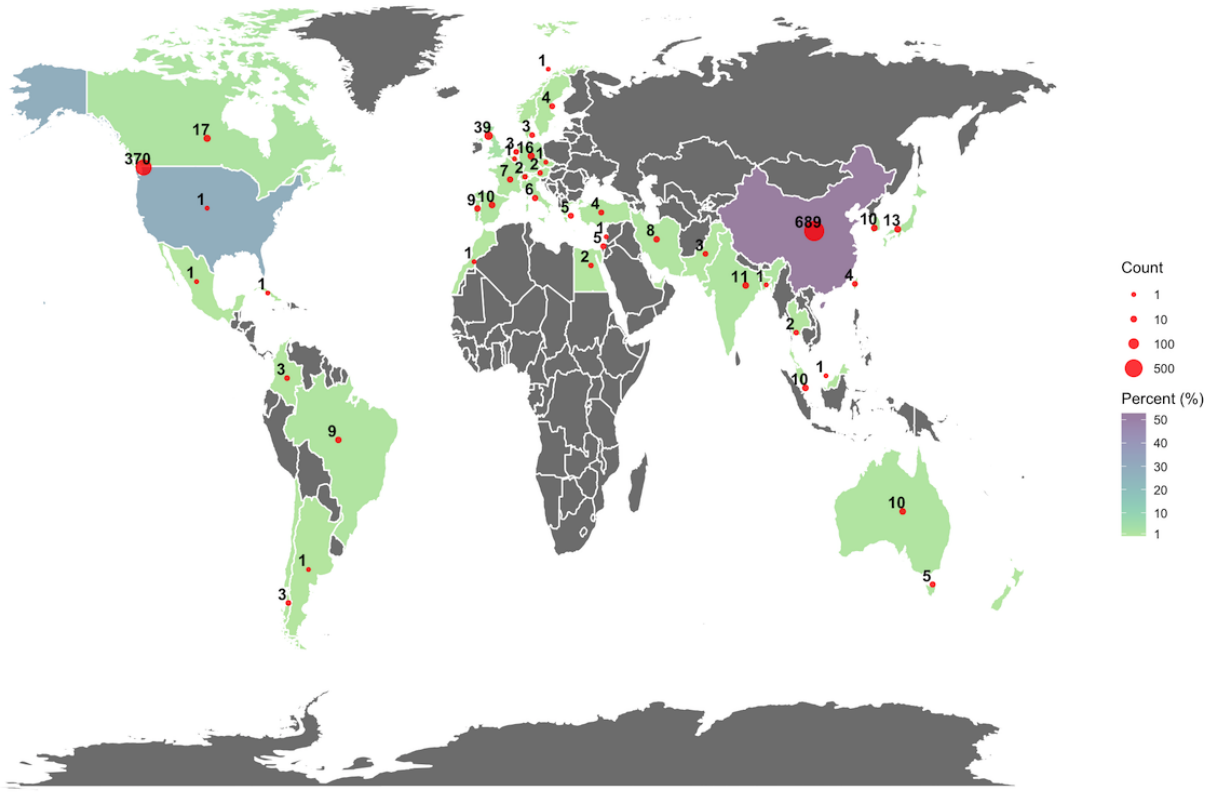

Journal Information

The median impact factor of publications was 3.8 (IQR 2.8-5.8) (Table 1). As for the journal field, the most commonly represented journals were those in medicine, computer science, and intensive care, with 712 (54.5%), 373 (28.5%), and 242 (18.5%) publications, respectively (Table 2).

Table 2. Field of the journal.

Field of the journal	Studies (N=1307), n (%)
Medicine	712 (54.5)
Informatics–mathematics–physics	373 (28.5)
Critical care	242 (18.5)
Biology–microbiology–immunology	82 (6.3)
Multidisciplinary sciences	63 (4.8)
Health care sciences–services	56 (4.3)
Pharmacology–pharmacy	39 (3)
Surgery	34 (2.6)
Environmental sciences	22 (1.7)
Neurosciences	22 (1.7)
Anesthesiology	21 (1.6)
Emergency	17 (1.3)
Genetics–heredity	12 (0.9)
Nursing	10 (0.8)
Management sciences	4 (0.3)

XSL•FO
RenderX

Study Information

Studies included a median number of 4282 (IQR 1468-12,740) patients with a maximum sample size reaching up to 219,306 by combining patients from several databases. The median inclusion period was 11 (IQR 11-11) years (Table 1).

Research predominantly explored the cardiovascular system, particularly hemodynamics with 333 (25.5%) publications and infectious diseases, specifically sepsis with 324 (24.8%) publications. Renal failure and metabolic disorders were also heavily studied, with 231 (17.7%) publications, along with respiratory failure and mechanical ventilation, which had 203 (15.5%) publications. Notably, many studies investigated overall mortality among critically ill patients admitted to the ICU, with 178 (13.6%) publications (Table 3).

Furthermore, 422 (32.6%) studies investigated the entire population admitted to the ICU. Most studies focused on patients experiencing specific organ failures, represented in 818 (62.6%) publications (Table 4). Various exposure or prediction factors were evaluated. Biological data and vital signs were the most analyzed, featuring in 673 (51.5%) and 428 (32.7%) studies, respectively. Demographic information, medication prescriptions, and comorbidities were also widely included, with 332 (25.4%), 278 (21.3%) and 271 (20.7%) publications, respectively (Table 5). The most frequently assessed primary outcome was mortality, observed either within the ICU, the hospital, or after a certain period of time, across 878 (67.2%) studies (Table 6).

Table 3. Research topics.

Research topics	Studies (N=1307), n (%)
Cardiovascular–hemodynamics	333 (25.5)
Infectiology–immunology	324 (24.8)
Nephrology–urology–metabolic	231 (17.7)
Pulmonary–intubation–ventilation	203 (15.5)
General	178 (13.6)
Pharmacology	105 (8)
Endocrinology–nutrition	92 (7)
Neurology–neurosurgery	86 (6.6)
Digestive–hepatology	83 (6.4)
Hemostasis–thrombosis–transfusion	82 (6.3)
Traumatology–orthopedics	34 (2.6)
Sedation–curarization–analgesia	26 (2)
Technology–monitoring	25 (1.9)
Cancerology	24 (1.8)
Psychiatry	16 (1.2)
Geriatric	15 (1.1)
Sociology	15 (1.1)
Toxicology–addictology	13 (1)
Pediatric	12 (0.9)
Other	7 (0.5)
Ethics	1 (0.1)
Obstetrics	1 (0.1)

Table 4. Analyzed population.

Analyzed population	Studies (N=1307), n (%)
Patients with a specific symptom, disease, organ failure	818 (62.6)
All intensive care unit patients	428 (32.7)
Patients with a specific medical procedure	127 (9.7)
Patients after a surgical procedure	52 (4)
Patients with a specific biological disorder	6 (0.5)

XSL•FO
RenderX

Table 5. Analyzed exposures.

Analyzed exposures	Studies (N=1307), n (%)
Biological markers	673 (51.5)
Vital signs	428 (32.7)
Demographic characteristics	332 (25.4)
Treatments	278 (21.3)
Comorbidities	271 (20.7)
Scores	125 (9.6)
Morphological measurement	114 (8.7)
Ventilation settings	103 (7.9)
Social determinants	88 (6.7)
Language	85 (6.5)
Other	83 (6.4)
Mechanical ventilation	76 (5.8)
Type of admission	53 (4.1)
Diagnostic procedure	49 (3.7)
Length of stay	45 (3.4)
Type of intensive care unit	43 (3.3)
Fluid balance	40 (3.1)
Renal replacement therapy	35 (2.7)
Transfusion	20 (1.5)
Surgery	20 (1.5)
Cluster	18 (1.4)
Substance use disorder	16 (1.2)
Artificial nutrition	5 (0.4)

Table 6. Analyzed outcomes.

Analyzed outcomes	Studies (N=1307), n (%)
Mortality	878 (67.2)
Occurrence of a clinical event	327 (25)
Resource management	135 (10.3)
Biological markers	69 (5.3)
Vital signs	66 (5)
ICU[a] readmission	35 (2.7)
Length of stay	27 (2.1)
Other	2 (0.2)

[a]ICU: intensive care unit.

Statistical Methods Used and Results

Regarding the study objectives, 713 (54.6%) studies addressed an inference question, while 594 (45.4%) studies targeted a prediction question (Table 7, Table 8). Supervised learning and deep learning methods were the most commonly applied machine learning techniques, with 432 (33.1%) and 247 (18.9%) publications respectively (Table 7).

When examining specific models, logistic or linear regression was the predominant choice for inference questions with 396 (55.5%) publications (Table 9). Regarding prediction questions, logistic or linear regression was the most widely used classical machine learning method, surpassing survival models and generalized models with 281 (47.5%), 29 (4.9%), and 12 (2%) publications respectively. Neural networks were the most frequently applied advanced machine learning methods

XSL•FO
RenderX

outpacing tree-based methods and boosting methods, with 237 (40.0%), 186 (31.4%), and 169 (28.5%) publications respectively (Table 10).

Inference studies reported statistically significant results in 84% of cases. Furthermore, the prevailing performance measure in prediction studies was the area under the curve (AUC) with 436 (73.4%) studies. The median AUC value was 0.840 (IQR 0.780-0.890; Table 8).

Preexisting clinical scores were applied for comparison with the performance of prediction models implemented in the studies. The most used scores were SAPS (Simplified Acute Physiology Score), SOFA (Sepsis-related Organ Failure Assessment), and APACHE (Acute Physiology and Chronic Health Evaluation) scores, with 55 (42.6%), 24 (18.6%), and 14 (10.9%) publications, respectively. The median AUC for these scores was 0.733 (IQR 0.677-0.781; Table 8).

Table 7. Algorithm used.

Algorithm used	Studies (N=1307), n (%)
Inference	713 (54.6)
Supervised learning	432 (33.1)
Deep learning	247 (18.9)
Unsupervised learning	39 (3)
Reinforcement learning	22 (1.7)

XSL•FO
RenderX

Table 8. Statistical analyses.

Characteristics and statistical methods used	Studies (N=1307)
Aim of the study, n (%)	
Inference	713 (54.6)
Prediction	594 (45.4)
Effect-size measure (if inference), n (%)	
Odds ratio	348 (57.4)
Hazard ratio	244 (40.3)
Coefficient	14 (2.3)
Prediction performance measure (if prediction), n (%)	
Area under the receiving operating curve	436 (73.4)
Sensibility or specificity	143 (24)
Accuracy	142 (23.9)
F_1-score	95 (16)
C-Index	74 (12.5)
Recall	42 (7.1)
Root-mean-square error	8 (1.3)
Known prediction scores used for comparison (if prediction), n (%)	
Simplified acute physiology score	55 (42.6)
Sepsis-related organ failure assessment	24 (18.6)
Acute physiology and chronic health evaluation	14 (10.9)
Acute physiology score	9 (7)
Modified early warning score	4 (3.1)
Oxford acute severity of illness	3 (2.3)
Other (<3 publications)	20 (15.5)
Key findings obtained	
Effect-size value (Odds ratio or hazard ratio), mean (IQR)	
Protective effect	0.65 (0.50-0.79)
Adverse effect	1.63 (1.28-2.36)
P **value, n (%)**	
>.05	98 (16.)
.01-.05	151 (24.6)
.001-.01	95 (15.5)
<.001	269 (43.9)
Performance of models used (area under the curve), median (IQR)	0.840 (IQR 0.780-0.890)
Performance of known scores used (area under the curve), median (IQR)	0.733 (IQR 0.677-0.781)

XSL•FO
RenderX

Table 9. Specific model used according to the aim of study (if inference).

Specific model used according aim of study (if inference)	Studies (N=713), n (%)
Linear or logistic regression	396 (55.5)
Survival regression	250 (35.1)
Propensity score matching	140 (19.6)
Descriptive–univariate	83 (11.6)
Other	48 (6.7)
Generalized additive model	24 (3.4)
Clustering	20 (2.8)
Natural language processing	2 (0.3)

Table 10. Specific model used according to the aim of study (if prediction).

Specific model used according aim of study (if prediction)	Studies (N=594), n (%)
Linear or logistic regression	282 (47.5)
Neural network	237 (39.9)
Decision tree or random forest	188 (31.6)
Boosting	171 (28.8)
Support vector machine	105 (17.7)
Natural language processing	81 (13.6)
Regularization (Lasso, Ridge, or Elasticnet)	78 (13.1)
K-nearest neighbors	49 (8.2)
Naive bayes	48 (8.1)
Other	40 (6.7)
Survival regression	29 (4.9)
Reinforcement learning	22 (3.7)
Clustering	19 (3.2)
Generalized additive model	12 (2.0)
Bagging	12 (2.0)
Bayesian network	7 (1.2)

Discussion

Principal Findings

This review examined clinical publications from open databases in the field of intensive care. We observed a consistent increase in the number of publications from these databases since 2016, with the majority of articles being published after 2020. MIMIC databases were the most frequently used, while the countries contributing the most were China and the United States. Most studies were published in journals with an impact factor ranging from 3 to 6. Cardiovascular and infectious topics, particularly those related to hemodynamics and sepsis, were the most represented. Other significant subjects included renal failure coupled with metabolic disorders and respiratory failure with mechanical ventilation. The most studied outcome measure was mortality in the ICU. Biological data, vital signs, demographic details, medication prescriptions, and comorbidities were extensively used as exposure or predictor factors.

Regarding statistical methods, logistic regression was the most used model for both inference and prediction questions. Neural networks were the most frequently used advanced machine learning methods surpassing supervised and reinforcement methods. A majority of the inference studies presented statistically significant results. In prediction studies, the most recurrent performance measure was the AUC, with a median value of 0.840.

Comparison With Previous Work

The rising number of publications using open databases is steadily increasing, highlighting the potential impact of these resources on clinical research within intensive care. Collaborative initiatives have led to the emergence of large open-access databases, enhancing data availability in this field. Researchers have now access to data from numerous centers, enabling the production of robust results through larger sample sizes. Furthermore, this access has significantly boosted statistical power, enabling the study of specific topics and

XSL•FO
RenderX

populations that were previously overlooked due to insufficient sample sizes. In our review, the median sample size in the studies was 4282 (IQR 1468-12,740), in contrast to typical sample sizes in prospective studies among critical care patients, that barely reach a few hundred patients. In a literature review by van de Sande et al [27] studying machine learning in intensive care, the median sample size was 179 (IQR 94-1411) and 142 (IQR 40-380) across all prospective observational and clinical studies, respectively.

The wealth and diversity of information contained in these databases have opened a broad spectrum of research possibilities, resulting in a wide array of topics covered in publications. Our review categorized these into over 20 different categories. Cardiovascular, infectious, respiratory, and metabolic issues emerged as the most commonly studied research domains, mirroring the reality of clinical practice in intensive care [28]. Mortality in ICU was the most frequently analyzed outcome measure. In a literature review by Syed et al [14], which examined the application of machine learning using the MIMIC dataset, mortality prediction was also the most studied outcome measure, followed by sepsis prediction, cardiac events, and acute kidney injury prediction.

Despite the significant number and diversity of publications, the clinical relevance of studies from these databases remains heterogeneous. The results were often statistically significant but had limited effect sizes. Nevertheless, there are few criteria to objectively evaluate the clinical relevance of a publication. In addition, the quantitative nature of our analyses did not allow us to fully explore this aspect. A qualitative methodology specifically assessing clinical relevance would provide more insights and discussion points.

Recently, a bibliometric analysis study using machine learning and natural language processing methods Bidirectional Encoder Representations from Transformers (BERTopic) compared research topics in traditional intensive care unit studies and those conducted with open access databases. A total of 1307studies from open access databases were identified versus 145,426 traditional studies. Among these 1307 studies, the most predominant topic was sepsis and kidney injury, and over 40% of the articles were based on predictive models [29].

Toward Precision Intensive Care Medicine

In recent years, due to advancing technology, data availability, and the need to analyze increasingly larger databases, machine learning has emerged to develop increasingly accurate predictive models. Machine learning is a field of computer science and a part of artificial intelligence that defines both the science and engineering for which computer systems can analyze data and learn from the information contained within [30].

In our study, 594 out of 1307 (45.4%) publications used at least 1 machine learning model, with a significant increase in their implementation after 2015. The most commonly used machine learning method was logistic regression and neural networks. Literature review by Shillan et al [31], summarizing the characteristics and results of machine learning methods used in intensive care, found similar results. Nearly half of the studies identified in that review were published after 2015, and the most commonly used methods were neural networks, support vector machines, and decision trees.

Although the use of machine learning models in intensive care has increased in the scientific literature, their adoption in actual clinical practice remains limited or even absent at present. Machine learning models are often complex and require specialized skills to develop and implement. Their interpretation also poses a significant barrier to their use, with the phenomenon known as the “black box” issue [32]. Furthermore, these models need to be rigorously validated to ensure their reliability and accuracy. External validation of models on independent datasets is crucial to evaluate their performance under real-world conditions. In our review, the vast majority of publications analyzed came from databases implemented in the United States of America, limiting the generalizability of these models to the rest of the population. In addition, only 92 out of 2063 (4.5%) articles focused on external validation using an independent dataset, which is consistent with Shillan et al findings [31]. Furthermore, prospective clinical evaluation of these models is still rare. In a recent review, only 10 out of 494 (2%) articles clinically evaluated artificial intelligence in real clinical settings, with 5 studies being randomized clinical trials [27]. Finally, the use of machine learning models raises ethical and regulatory questions, particularly concerning data privacy, automated decision-making, and liability in case of errors. Clear guidelines and regulations must be established to govern the use of these models in a clinical context.

Strengths and Limitations

This study, to our knowledge, is the first to examine the contribution of open databases in clinical research in intensive care. This study follows a rigorous methodology, and its research protocol was made public before the study [33]. Furthermore, this review adopts a comprehensive approach by including all clinical research publications generated from open ICU databases. By examining a wide range of publications, it provides a detailed overview of research themes, methodologies used, and results obtained. This can assist clinicians and researchers in understanding how these databases can contribute to improving patient care in intensive care and offering a solid foundation for future research and discussions.

However, it is important to acknowledge the limitations of this literature review. First, only publications written in English were included, potentially introducing a linguistic bias. Second, there exists and inherent risk of publication bias in literature reviews. Published and accessible studies may not represent all research carried out on open databases in intensive care. Studies with negative or nonsignificant results are less likely to be published, potentially resulting in an overestimation of positive results. While several open ICU databases were included in the study, it is possible that there exist other databases that were not taken into account. This could limit the representativeness of the sampled studies. Finally, despite exhaustive research, no publications from the HiRID database were identified. With these limitations in mind, it is important to consider the results of this study with caution and interpret them in the appropriate context.

XSL•FO
RenderX

Conclusion

Open databases in intensive care have facilitated clinical research and provided new perspectives for enhancing care in intensive care. The abundance of scientific outputs resulting from these databases and the diversity of topics addressed highlight the importance of these databases as valuable resources for clinical research and suggest their potential impact on clinical practice in intensive care. However, the quality of studies and their clinical relevance remain highly heterogeneous among publications and are challenging elements to evaluate.

Authors' Contributions

BP and JK were involved in the design of the study, selection of articles, data analysis, and writing of the manuscript. GK was involved in the development and validation of the search queries. AL was involved in arbitration in case of a lack of consensus in the selection of articles and the validation of the final article. TC, GB, and JG were involved in the review and validation of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional material.
[DOC File , 127 KB - jmir_v27i1e57263_app1.doc]

Multimedia Appendix 2
PRISMA-ScR Checklist.
[DOC File , 55 KB - jmir_v27i1e57263_app2.doc]

References

1. Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest 2018;154(5):1239-1248 [FREE Full text] [doi: 10.1016/j.chest.2018.04.037] [Medline: 29752973]
2. Safran C. Reuse of clinical data. Yearb Med Inform 2018;23(01):52-54. [doi: 10.15265/iy-2014-0013]
3. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American medical informatics association white paper. J Am Med Inform Assoc 2007;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]
4. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19(1):64 [FREE Full text] [doi: 10.1186/s12874-019-0681-4] [Medline: 30890124]
5. Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. J Epidemiol Community Health 2020;74(7):612-616 [FREE Full text] [doi: 10.1136/jech-2018-211654] [Medline: 32332114]
6. Chico V. The impact of the general data protection regulation on health research. Br Med Bull 2018;128(1):109-118. [doi: 10.1093/bmb/ldy038] [Medline: 30445448]
7. Thoral PJ, Peppink JM, Driessen RH, Sijbrands EJG, Kompanje EJO, Kaplan L, Amsterdam University Medical Centers Database (AmsterdamUMCdb) Collaboratorsthe SCCM/ESICM Joint Data Science Task Force. Sharing ICU patient data responsibly under the society of critical care medicine/European society of intensive care medicine joint data science collaboration: the amsterdam university medical centers database (AmsterdamUMCdb) example. Crit Care Med 2021;49(6):e563-e577 [FREE Full text] [doi: 10.1097/CCM.0000000000004916] [Medline: 33625129]
8. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
9. Sauer CM, Dam TA, Celi LA, Faltys M, de la Hoz MAA, Adhikari L, et al. Systematic review and comparison of publicly available ICU data sets-a decision guide for clinicians and data scientists. Crit Care Med 2022;50(6):e581-e588 [FREE Full text] [doi: 10.1097/CCM.0000000000005517] [Medline: 35234175]
10. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
11. Mark R. The story of MIMIC. In: Secondary Analysis of Electronic Health Records Internet. Cham: Springer International Publishing; 2016:43-49.
12. Feng M, McSparron JI, Kien DT, Stone DJ, Roberts DH, Schwartzstein RM, et al. Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database. Intensive Care Med 2018;44(6):884-892. [doi: 10.1007/s00134-018-5208-7] [Medline: 29806057]

XSL•FO
RenderX

13. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018;24(11):1716-1720 [FREE Full text] [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
14. Syed M, Syed S, Sexton K, Syeda HB, Garza M, Zozus M, et al. Application of machine learning in intensive care unit (ICU) settings using MIMIC dataset: systematic review. Informatics (MDPI) 2021;8(1):16 [FREE Full text] [doi: 10.3390/informatics8010016] [Medline: 33981592]
15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
17. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Lockwood C, Porrit K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. Australia: JBI; 2020.
18. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. Sci Data 2018;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]
19. Faltys M, Zimmermann M, Lyu X, Hüser M, Hyland S, Rätsch G, et al. HiRID, a high time-resolution ICU dataset.: PhysioNet; 2021. URL: https://physionet.org/content/hirid/1.1.1/ [accessed 2024-09-29]
20. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10(1):1 [FREE Full text] [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
21. Lee J, Scott DJ, Villarroel M, Clifford GD, Saeed M, Mark RG. Open-access MIMIC-II database for intensive care research. : IEEE; 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 03, 2011; Boston, USA p. 8315-8318 URL: https://europepmc.org/abstract/MED/22256274 [doi: 10.1109/IEMBS.2011.6092050]
22. Faster systematic reviews.: Rayyan URL: https://www.rayyan.ai/ [accessed 2024-09-29]
23. Rigby AS. Statistical methods in epidemiology. v. towards an understanding of the kappa coefficient. Disabil Rehabil 2000;22(8):339-344. [doi: 10.1080/096382800296575] [Medline: 10896093]
24. Martignene N, Amad A, Bellet J, Tabareau J, D'Hondt F, Fovet T. Goupile: a new paradigm for the development and implementation of clinical report forms. In: Séroussi B, Weber P, Dhombres F, Grouin C, Liebe JD, Pelayo S, editors. Studies in Health Technology and Informatics. Netherlands: IOS Press; 2022.
25. Welcome.: Clarivate URL: https://access.clarivate.com/login?app=wos&alternative=true&shibShireURL=https:%2F%2Fwww.webofknowledge.com%2F%3Fauth%3DShibboleth&shibReturnURL=https:%2F%2Fwww.webofknowledge.com%2F&roa ming=true [accessed 2024-09-29]
26. The Comprehensive R Archive Network. URL: https://cran.r-project.org/ [accessed 2024-06-30]
27. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med 2021;47(7):750-760 [FREE Full text] [doi: 10.1007/s00134-021-06446-7] [Medline: 34089064]
28. Popoff B, Occhiali E, Grangé S, Bergis A, Carpentier D, Tamion F, et al. Trends in major intensive care medicine journals: A machine learning approach. J Crit Care 2022;72:154-163. [doi: 10.1016/j.jcrc.2022.154163] [Medline: 36209696]
29. Ke Y, Yang R, Liu N. Comparing open-access database and traditional intensive care studies using machine learning: bibliometric analysis study. J Med Internet Res 2024;26:e48330 [FREE Full text] [doi: 10.2196/48330] [Medline: 38630522]
30. Gutierrez G. Artificial intelligence in the intensive care unit. Crit Care 2020 Mar 24;24(1):101 [FREE Full text] [doi: 10.1186/s13054-020-2785-y] [Medline: 32204716]
31. Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. Crit Care 2019;23(1):284 [FREE Full text] [doi: 10.1186/s13054-019-2564-9] [Medline: 31439010]
32. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
33. Popoff B. Contribution of open access databases to intensive care medicine research: a scoping review.: OSF; 2022. URL: https://osf.io/kugaz/ [accessed 2024-09-29]

Abbreviations

AmsterdamUMCdb: Amsterdam University Medical Centers Database
APACHE: Acute Physiology and Chronic Health Evaluation
AUC: area under the curve
BERTopic: Bidirectional Encoder Representations from Transformers
eICU-CRD: eICU Collaborative Research Database

XSL•FO
RenderX

HiRID: High time-resolution ICU dataset
ICU: intensive care unit
MIMIC: Medical Information Mart for Intensive Care
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SAPS: Simplified Acute Physiology Score
SOFA: Sepsis-related Organ Failure Assessment

Edited by A Mavragani; submitted 09.02.24; peer-reviewed by E Sylvestre, F Carrouel; comments to author 07.05.24; revised version received 30.06.24; accepted 23.08.24; published 09.01.25.

Please cite as:
Kallout J, Lamer A, Grosjean J, Kerdelhué G, Bouzillé G, Clavier T, Popoff B
Contribution of Open Access Databases to Intensive Care Medicine Research: Scoping Review
J Med Internet Res 2025;27:e57263
URL: https://www.jmir.org/2025/1/e57263
doi:10.2196/57263
PMID:

©Julien Kallout, Antoine Lamer, Julien Grosjean, Gaétan Kerdelhué, Guillaume Bouzillé, Thomas Clavier, Benjamin Popoff. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of Reddit Discussions on Motivational Factors for Physical Activity: Cross-Sectional Study

Michal Shmueli-Scheuer[1,2], PhD; Yedidya Silverman[1,2], BPT; Israel Halperin[1,2*], PhD; Yftach Gepner[1,2*], PhD

[1]Department of Health Promotion, School of Public Health, Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel

[2]Sylvan Adams Sports Institute, Tel Aviv University, Tel Aviv, Israel

[*]these authors contributed equally

Corresponding Author:
Yftach Gepner, PhD
Department of Health Promotion
School of Public Health, Faculty of Medical and Health Sciences
Tel Aviv University
PO Box 39040
Tel Aviv, 6997801
Israel
Phone: 972 0733804427
Email: gepner@tauex.tau.ac.il

Abstract

Background: Despite the ample benefits of physical activity (PA), many individuals do not meet the minimum PA recommended by health organizations. Structured questionnaires and interviews are commonly used to study why individuals perform PA and their strategies to adhere to PA. However, certain biases are inherent to these tools that limit what can be concluded from their results. Collecting data from social media channels can complement these studies and provide a more comprehensive overview of PA motives and adherence strategies.

Objective: This study aims to investigate motives for engaging in PA, as well as the associated strategies to achieve these goals, as stated by a large number of people on a social media site.

Methods: We searched for users' responses regarding PA motives and adherence strategies in Reddit forums dedicated to PA and analyzed the data using (1) unsupervised clustering to identify topics from the textual comments and (2) supervised classification to classify the comments into the detected topics. A panel of experts participated in both steps for annotation and validation purposes.

Results: We analyzed 1577 unique user comments (representing 1577 individual users); of those, 1247 were linked to physical appearance (mentioned in 298/1247, 23.9% of the comments) and improving physical (235/1247, 18.9%) and mental health (211/1247, 16.9%), indicating these as the main motivational factors. The main strategies people used to adhere to PA were habit formation (373/1247, 30%), goal setting (173/1247, 13.9%), enjoyable activities (151/1247, 12.1%), socializing (121/1247, 9.7%), using media (111/1247, 8.9%), using different apps to monitor PA (35/1247, 2.8%), and financial commitment (32/1247, 2.5%).

Conclusions: This study presented a novel approach using a language model to investigate why people engage in PA and the strategies they use to adhere to PA using wide-scale, self-disclosed content from popular social media channels.

(J Med Internet Res 2025;27:e54489) doi:10.2196/54489

KEYWORDS

motivation; physical activity; social media; Reddit; adherence

Introduction

Despite the established health benefits of physical activity (PA) [1-4], surveys from multiple countries have reported that more than 80% of adolescents and 27% of adults do not meet the World Health Organization guidelines [5-7]. Studying what motivates individuals to perform PA and what strategies assist them in adhering to PA can lead to improvements and developments in interventions aiming to increase PA levels. The most common way to examine people's motives and adherence strategies to PA is by using standardized questionnaires and structured interviews [8-13]. While these

XSL•FO
RenderX

studies using these data collection strategies have clear benefits, they also suffer from several inherent biases that limit what can be concluded based on their results. First, participants in these studies are likely unrepresentative of the population. The observed phenomenon can be attributed to the specific targeting of participants and the resultant self-selection bias, as those who volunteered to participate (ie, volunteer bias) possessed distinct attributes compared with those who declined [14]. Indeed, volunteering is a trait associated with sociodemographic factors, attitudes, and other personality traits of lifestyle behaviors [15,16]. Second, questionnaires and interviews can lead to response bias, given that they can confuse or prompt participants to answer in a particular manner [17,18]. Third, participants may interpret the studies' purpose or interviewers' intent and modify their responses to align with this interpretation, a phenomenon known as demand characteristics [19]. These biases might create a misleading picture of reality and lead to misclassification of the factors that motivate individuals to engage and adhere to PA.

One approach that can partially overcome these shortcomings is to collect information concerning motivation and adherence strategies to perform PA from social media channels. Social media allows people to share their experiences and opinions as they wish, mostly unrestrictedly. Moreover, the anonymity afforded by social media may allow for more uninhibited self-disclosure without worrying about potential exposure or various consequences [20-22]. Data gathered in this fashion has inherent biases, including selection bias of social media users proficient in using computers. Nevertheless, the difference in mediums and context allows researchers to complement the traditional approaches based on standardized questionnaires and structured interviews. The increased ease of use in implementing machine learning algorithms on free text has enabled researchers to collect data on social media posts at a massive scale. This method allows for larger sample sizes than traditional research methods and may avoid specific selection and response biases. Yet despite the widespread adoption of this research method within multiple scientific fields, it has been rarely implemented within exercise science, specifically in assessing motivation and adherence strategies for engaging in physical activity.

Of the various social media channels, Reddit is a highly popular website (861 million active users as of 2021) [23] in which users submit text, images, and videos to moderated boards (known as "subreddits"), which are then voted up or down by other users. The websites' users are primarily from the United States (47.13%) [24], skew young (36% under 30), male (63.8%), and from higher income brackets [25]. Discussions on Reddit are public to anyone (unless specifically designated as private), allowing for passive data collection. Reddit provides an application programming interface (API) for researchers with extensive access to data, making it a valuable source of information. Therefore, this study aims to determine the motivational factors and strategies to both begin and adhere to the PA of Reddit users based on their self-reported data.

Methods

Overview

We singled out Reddit from other social media platforms due to its large user base, comment ranking system, mode of posting (questions and answers within subject matter "communities," self-disclosure), and researcher-friendly API. The process we undertook to identify and extract data from posts on social media was to first decide upon keywords related to PA and motivation based on an initial search to identify relevant communities. Then, the next step was to identify relevant conversations within the selected subreddits. We will now elaborate on each step of this process:

Reddit Community Identification (Initial Examination, Sorting, and Filtering)

We ran a Google search of Reddit using the keywords "motivation" and "exercise" or "sport." This search returned 59 communities (ie, subreddits).

We began by excluding subreddits containing fewer than 100,000 users (19 groups) and non-English language groups (2 groups). Following initial examination of the remaining subreddits, we found that they could be broadly divided into 2 categories: "general" (25 groups) and "sports-related" (13 groups). Within these 2 categories, we observed that some subreddits were not topically related to our query (motivation in exercise and sport), but rather, they were geared toward other topics (eg, science, GARMIN, and sports leagues such as the NBA [National Basketball Association] and FIFA [Fédération Internationale de Football Association]) and were therefore excluded (17 and 5 from general and sport-related, respectively). Following further examination of the content of the conversations in the remaining sports-related subreddits (bodybuilding, judo, running), we observed that most of the discussions were of a technical nature (ie, best running strategy, bodybuilding tips) and were therefore excluded. We then selected for further analysis the 3 most popular (based on active user activity) subreddits from the categories identified: PA/sports-related (/r/bodyweightfitness, /r/crossfit, and /r/Fitness) and general advice (/r/askMen, /r/askWomen, and /r/askReddit).

Comment Identification

Comment Search and Initial Data Extraction

We searched the selected subreddits using the Reddit API for conversations containing the terms "motivation," "exercise," or "sport." From each of these discussions, we extracted the title, comments, and total "points" (a site metric that captures how many people liked each comment) and unique author identification number of individual comments.

Comment Filtration

In cases where a specific user made more than 1 comment across all the conversations captured by our query, a single comment was chosen randomly. This approach reduces biases of users who posted multiple comments, which can give more weight to their answers and bias the dataset toward prolific commenters rather than unique comments. Based on previous research in

XSL•FO
RenderX

this field [26,27], comments containing only emojis, URLs, non-English languages, or one-word responses were also excluded.

Data Analysis Procedure

To analyze the dataset generated by our data collection and filtration procedure, we used the following analysis pipeline that included 5 steps, as presented in Figure 1. Following the established steps during the development and evaluation of machine learning and artificial intelligence models, it is essential to ensure the accountability and reliability of these models [28,29]. All data analysis was carried out in Python version 3.9.13 (Python Software Foundation) using the fuzzy-c-mean 1.7.2 (FCM), pandas 1.5.3, NumPy 1.26.4, and umap-learn-0.5.6 packages.

Figure 1. Overview of the main steps of the data analysis pipeline.



Step 1: Exploration and Initial Clustering

The first step aimed to explore the data and create initial clusters of topics (ie, motivations for engaging in PA) that appear in the text. First, we created embeddings, or vector representations, of each sentence to use for the clustering. Specifically, we encoded using the Sentence Bidirectional Encoder Representations from Transformers language model. We then reduced the high dimensionality of the embeddings using Uniform Manifold Approximation and Projection for Dimension Reduction, introduced by McInnes et al [30], implemented in umap-learn. Finally, we used a clustering algorithm FCM, with FCM implementation [31] to find groups within the data with the most similarity in the same cluster and the most dissimilarity between different clusters (this approach was chosen due to the observation that our data contains examples that discuss several topics). Specifically, this step allows finding groups of similar comments (which are converted to numerical data points) in the data by calculating the distance between data points (see subheading A.1 in Multimedia Appendix 1).

Step 2: Expert Tuning and Validation

To verify the clusters' topics, one of the co-authors (IH), with expertise in exercise science and motivation, reviewed several samples and proposed a topic title for each cluster. Then, 30 samples were randomly selected from each cluster, and 2 additional reviewers with similar expertise independently decided whether each sample matched the proposed cluster topic. Note that a sample could match several topics simultaneously. If both reviewers agreed on the topics, the sample was added to a pool of corrected annotated samples. Thus, at the end of this step, we had a high-quality representation of the different topics in the dataset.

Step 3: Supervised Text Classification

We classified the remaining nonreviewed comments into the topics identified in steps 1 and 2. This procedure's goal was to automatically classify the textual comments into one or more defined topics (Multimedia Appendix 2 [32-35]).

Step 4: Expert Evaluation

We implemented an additional round of expert evaluation of the classified comments. To minimize the risk of researcher bias, we recruited 5 graduate-level exercise physiology students rather than using the original 3 experts. We randomly sampled 30 instances from our data (not used in the first round of evaluation) and asked each expert to evaluate the classified data. Each expert was presented with a dedicated user interface (Figure 2). Each sample was displayed along with the possible topics (multiple choices were allowed) and a field to add comments using free text. Before beginning the annotation task, the reviewers (5 graduate-level exercise physiology students) underwent a familiarization procedure by reading the instructions and practicing on 3 fictitious examples, similar to comments from the dataset regarding topics and average comment length. Once the annotation task was completed, we aggregated the results and evaluated the quality of the classifier and the interrater agreement.

XSL•FO
RenderX

Figure 2. Annotation user interface used by subject matter experts.

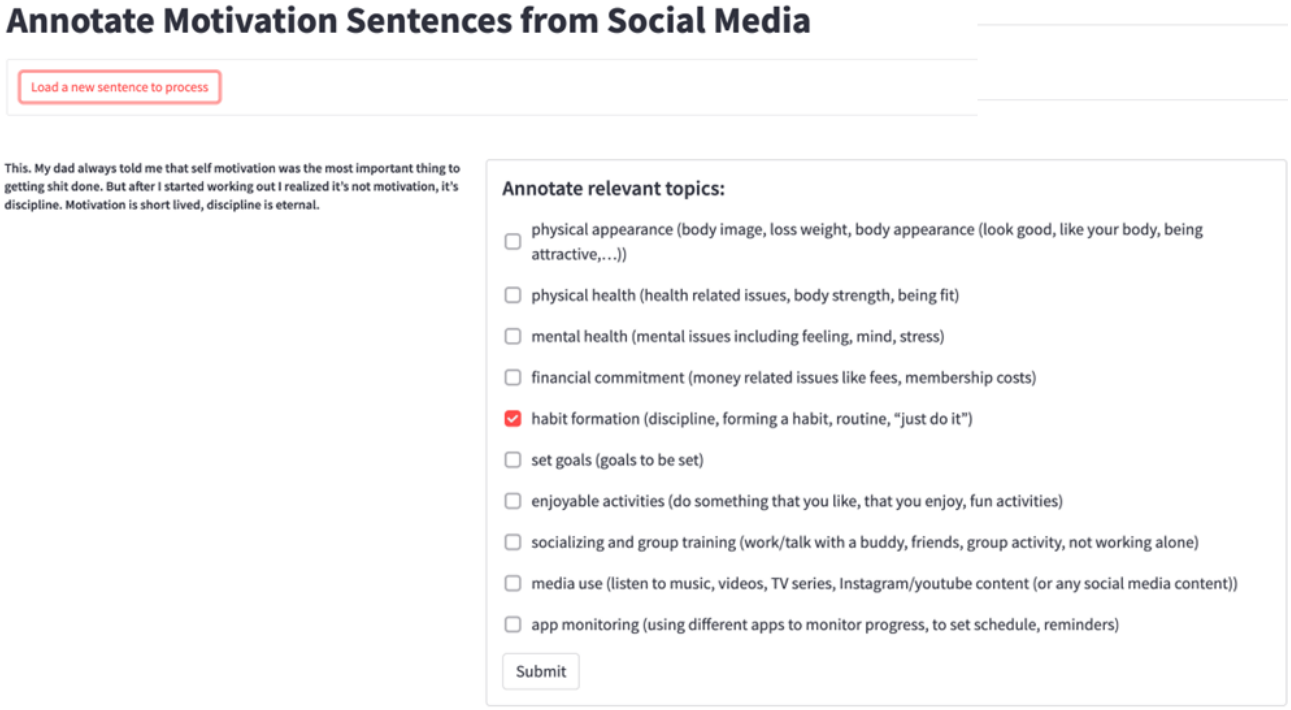

Step 5: Reliability

To ensure reliable findings, the quality of the supervised classifier that identifies motivational factors and strategies was evaluated using both automatic metrics (precision, recall, and F_1-score) and an expert evaluation panel, as mentioned above in step 4. For the human evaluation, we used Fliess κ interrater reliability [36].

Ethical Considerations

This research did not require approval from the Tel Aviv University Ethics Committee as it was based on publicly accessible information that is legally protected and does not necessitate an ethics review. This position was confirmed by the head of the university's ethics committee. From a privacy perspective, we used data from Reddit, a platform where most discussions are publicly accessible to anyone, with or without a Reddit account (except for private subreddits). In addition, Reddit offers an extensive API that allows access to much of the platform's information. Strictly adhering to the rules and ethical guidelines of each subreddit, we only collected data from subreddits that explicitly allow research activities. Furthermore, to protect user privacy, we only recorded the comment ID without including any user information or the content of the comment itself, ensuring no personally identifiable information or data that could lead to reidentification was collected.

Results

Data Collection

We collected 2387 comments from 70 discussions published between 2017 to 2021. After completing the data filtration procedure, 1850 samples remained, submitted by 1850 unique users: 227 from /r/bodyweightfitness, 152 from /r/Fitness, 220 from /r/crossfit, 338 from /r/askReddit, 479 from /r/askMen, and 434 from /r/askWomen. Links and examples of the discussions can be found under subheading A.3 in Multimedia Appendix 1.

Primary Outcome

Our model classified the comments into 10 distinct topics, which we observed could be divided into 2 broad categories: motivational factors for initiating PA (motivations) and strategies or suggestions on staying motivated and adhering to PA (strategies), for a total of 1740 observations of identified topics, over 1247 comments. The motivations for participating in PA (frequency of appearance) were physical appearance (298/1247, 23.9%), physical health (235/1247, 18.9%), and mental health (211/1247, 16.9%). The main strategies people use to adhere to PA were habit formation (373/1247, 30%), goal setting (173/1247, 13.9%), choosing enjoyable activities (151/1247, 12.1%), socializing (121/1247, 9.7%), using media (111/1247, 8.9%), using different apps to monitor PA (35/1247, 2.8%), and financial commitment (32/1247, 2.5%). Representative examples per topic are presented in Table 1, with their sums and proportions in Table 2. For a breakdown of topic frequency by subreddit, see subheading A.4 and Figure S1 in Multimedia Appendix 1. The reliability and accuracy metrics of the topic allocation by the model were high. We observed substantial agreement between raters with a Fliess κ of 0.621. The subsequent automatic evaluation of the model's precision, recall, and F_1-scores were 0.86 for all metrics. For a detailed breakdown of this process, see subheading A.4 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Topics and representative comments examples.

Topic	Example comments
Motivational factors for initiating physical activity	
Physical appearance	
	Stand in front of the mirror. Naked.
	the ass of my dreams.
	Buy a scale. Understanding exactly how much your weight is changing has helped me.
Physical health	
	If I don't my fibromyalgia pain will get worse. That's a pretty daunting thought. I can just barely make it through the day as it is.
	I don't want to die of something that could have easily been prevented by having a basic level of fitness. It's really the only reason I work out. I'm pretty lazy.
	Good to stay healthy and live a long life.
Mental health	
	Exercise makes me happy, improves my mood and helps me be more productive overall.
	It helps my mental health.
	i try to do three to four, make two a priority found that my mental health (specifically anxiety) started to slide if i did less than two a week. There's a huge difference in my anxiety levels when my workouts are regular.
Strategies or suggestions on staying motivated and adhering to physical activity	
Habit formation	
	by starting-the more you do it, the better you'll feel, etc. it becomes a habit/vicious cycle.
	Motivation depends on emotions which are unstable. Discipline is the ability to maintain a habit even when the motivation isn't there.
	It's addictive. It's difficult to start but once you get going you can't get enough.
Goal setting	
	You really have to want it. If you don't care to change or don't have any numerical goals, you won't last. Set some goals and really strive to meet them.
	I have goals to make a high level rugby side and i need to be fit and strong to get there.
	I suggest picking a goal. Run a 5k, 10k, marathon, whatever. Do a one-legged squat, one-armed push-up, ten pull-ups, whatever. Picking goals and tracking my progress towards them has helped me.
Enjoyable activities	
	This exactly. I hate going on runs or going to lift weights at the gym. I can't get motivated to do that. But if I take a spin class or kickboxing or muscle endurance class I I love it. It's all about finding something you like.
	For me, having a sport/game/activity that I enjoy is my motivation. I work out in order to be able to have more fun in that sport/game/activity. With more strength and flexibility, I am more freely and thoroughly able to express myself through movement. I work out so that my play is even more fun!
	Do something fun for exercise. For example, I love basketball so I play basketball as an exercise but I still enjoy it and it doesn't even seem like work.
Socializing	
	I enjoy being around the people at my gym. They make showing up everyday fun.
	My foolproof method: get a gym buddy. It's hard to skip when you know someone is waiting for you at the gym.
	I have a friend from my gym meet me on zoom and we work out together.it Helps so much.
Media use	
	Music and proper clothes. throw on an exercise bra and shorts, plus some salsa music and i can work out for hours. Music is just good for energy.
	Play intense music. That's all it takes for me.
	Mhm my favourite workout YouTuber is Lily Sabri. She has many 5-7min videos, so after like a half an hour workout or live stream workout if you still feel like you have a bit more energy you can get one more work out quickly.
App monitoring	

XSL•FO
RenderX

Topic	Example comments
	Every day. I got a good streak going on my fitbit from March and I'm desperate not to break it.
	I'm using one of the Couch to 5K apps that basically tells you exactly what to do, tracks your progress etc.
	Use the IIFYM calculator to figure out how much to eat for your goals and be strict on tracking with myfitnesspal app. 100% you will see great results and never have to seek motivation outside again.
Financial commitment	
	Think about how much your paying, I'm NOT wasting $150 per month, get your ass to the gym.
	The cost.
	I think about how much I pay per month.

Table 2. Total number of comments per topic (and percentage) as classified by the model.

Topic	Comments (n=1247), n (%)
Motivational factors for initiating physical activity	
Physical appearance	298 (23.9)
Physical health	235 (18.9)
Mental health	211 (16.9)
Strategies or suggestions for staying motivated and adhering to physical activity	
Habit formation	373 (30)
Goal setting	173 (13.9)
Enjoyable activities	151 (12.1)
Socializing	121 (9.7)
Media use	111 (8.9)
App monitoring	35 (2.8)
Financial commitment	32 (2.5)

Additional Exploratory Analysis

Time Period

The lockdowns and social distancing behaviors associated with the COVID-19 pandemic were known to have caused significant changes in population PA levels [37]. Therefore, we examined the frequency distribution of the 10 identified topics during 2 periods of time: before the pandemic, defined as the period up to date that a public health emergency was declared in the United States (January 1, 2017, to January 31, 2020; 940 comments) and during the pandemic (February 1, 2020, to January 31, 2021; 307 comments). Specifically, to calculate the topic frequencies before the pandemic, we summed the number of comments per topic and divided it by the total number of comments during this period (940). A similar calculation was done for the duration of the pandemic period (307). We observed statistically significant differences between the 2 time periods (using a 2-sample *z* test for proportions) in all 3 motivations for engaging in PA (physical appearance: P=.01; physical health: P<.001; mental health: P<.001) and in 2 strategies (enjoyable activities: P=.04; financial commitment: P=.01), as depicted in Table 3.

XSL•FO
RenderX

Table 3. Overview of the motivations and strategies before and during the COVID-19 pandemic (2-sample z test for proportions).

Topic	Before COVID-19 pandemic	During COVID-19 pandemic
Physical appearance[a]	0.26	0.19
Physical health[a]	0.17	0.25
Mental health[a]	0.15	0.24
Habit formation	0.30	0.29
Set goals	0.14	0.14
Enjoyable activities[a]	0.13	0.09
Socializing and group training	0.11	0.07
Media use	0.09	0.08
App monitoring	0.03	0.03
Financial commitment[a]	0.03	0.01

[a]*P*<.05.

Topic Co-Occurrence

We elected to examine patterns of topics mentioned in users' comments, that is, the probability of a Reddit user's comment containing combinations of topics (2 motivations or strategies). Out of 1740 unique observations of identified topics, 1278 incidences of topic co-occurrence were identified. Figure 3 presents a heat map of the co-occurrence probabilities, where each cell in the matrix presents the probability of a column's topic co-occurring with a row's topic, normalized by the total appearances of the row (as presented in Table 2).

Figure 3. Topic co-occurrences probabilities. Each cell is the probability of a row's topic appearing together with a column's topic (calculated as the number of co-occurrences divided by the total appearances of the row topic, that is, the number in parentheses). The cells within the 2 yellow outlines represent co-occurrence probabilities of motivations to motivations and strategies to strategies.

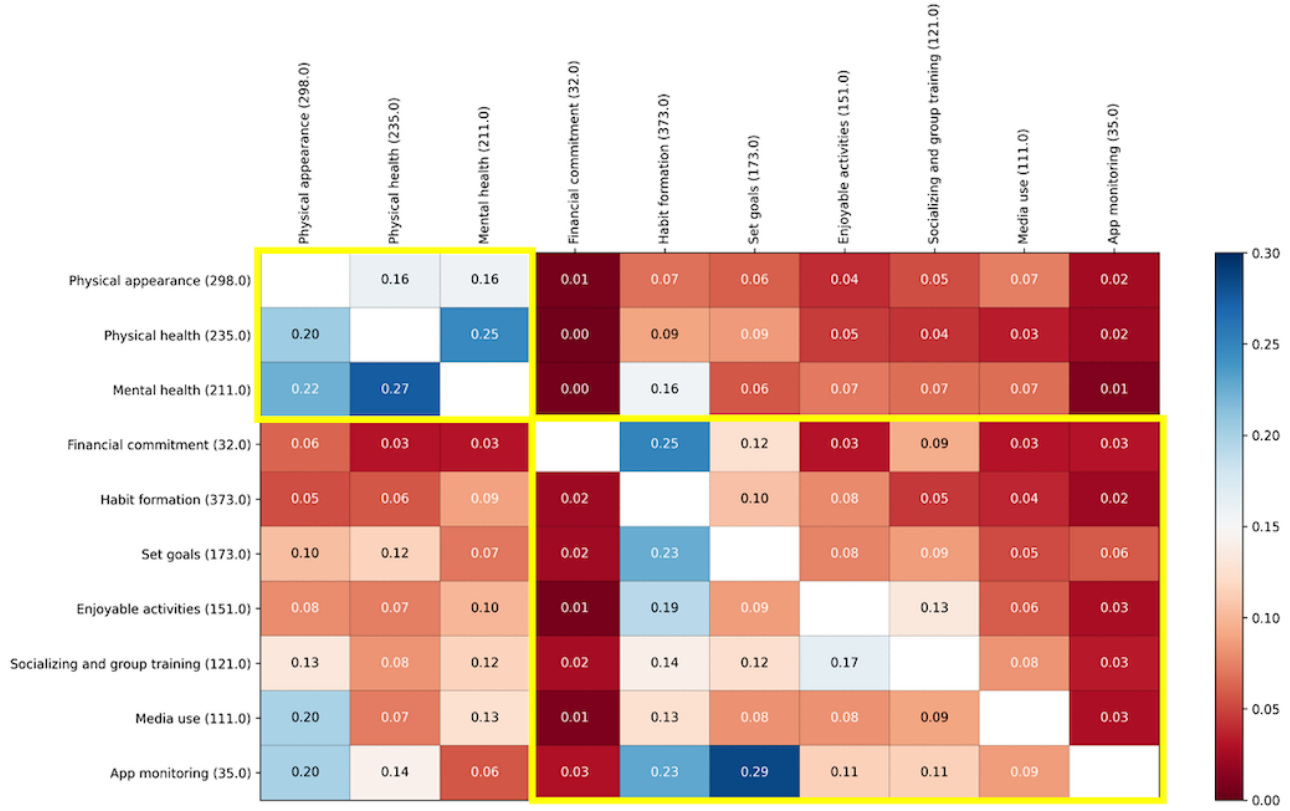

For example, given a comment labeled as discussing the topic of "physical health," the probability of that comment also discussing "mental health" was 0.25. We observed that the co-occurrences with the highest probabilities were combinations of "motivational factors/motivational factors" rather than "motivational factors/strategy."

XSL•FO
RenderX

Topic Importance

Some of the 10 identified topics may be considered more "important" than others by the users of the various Reddit communities studied here. The simplest method of defining importance would be by the frequency in which it appears in the dataset, as shown in Table 2. This metric can alternatively be described as the topic's popularity among comment authors (recall that the data filtration process allowed the inclusion of only 1 comment per author). As the Reddit API provided us with additional contextual data for each comment (comment "points"), we added the metric of topic importance using popularity "points"—how well those comments, and hence, topics, were received or recognized. Each Reddit comment is associated with points. The community members give points representing reaction agreement (or likes) on a specific comment. Thus, the more points a comment has, the more well-received a comment is by others. We used this point system as a proxy of topic popularity or importance. Specifically, we collected the total number of scores per topic and averaged them by the number of topic comments. We then compared the "popularity" of the 10 topics using the 2 measures proposed above. Figure 4 compares these 2 metrics, average authors versus average community importance, by topic.

Figure 4. Importance analysis per topic by the popularity of the authors and the community. Author popularity (per topic) = number of authors who wrote about the topic ÷ total number of authors. Community popularity (per topic) = total scores given to the topic by the community ÷ number of comments on the topic.

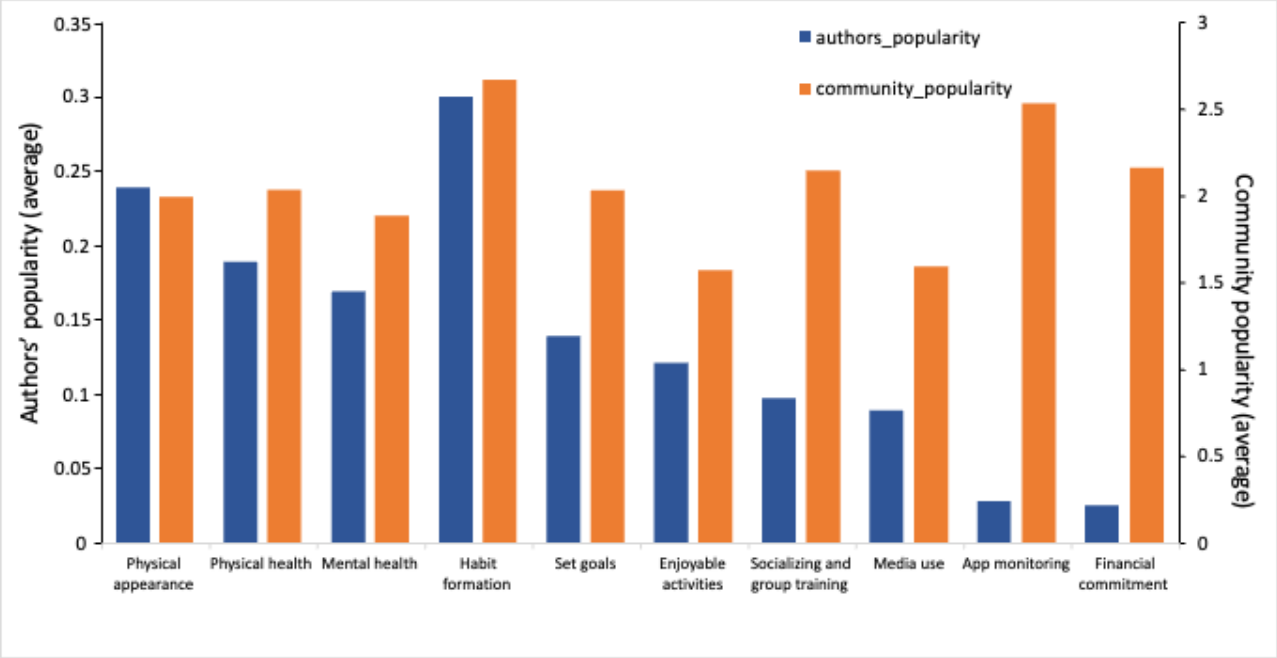

Discussion

Principal Findings

Our objective in this study was to explore motivational and adherence strategies for engaging in PA by using an unsupervised, multistep clustering technique to gather and analyze comments from Reddit forums. Our model classified the comments into 10 distinct topics, which we further categorized into 2 themes: motivations for initiating PA and strategies for maintaining motivation and adherence to PA. The main motives for individuals to engage in PA was their physical appearance (298/1247, 23.9%) and enhancing physical (235/1247, 18.9%) and mental (211/1247, 16.9%) health. The most prevalent strategy individuals used to adhere to PA was habit formation (373/1247, 30%), followed by goal setting (173/1247, 13.9%), enjoyable activities (151/1247, 12.1%), socializing (121/1247, 9.7%), using media (111/1247, 8.9%), using PA monitoring apps (35/1247, 2.8%), and financial commitment (32/1247, 2.5%). In addition, the frequency with which certain topics were mentioned together presented with certain patterns (Figure 3). For example, mental or physical health were often mentioned together with physical appearance (385/1247, 22% and 348/1247, 20%, respectively), and app monitoring or media use were often mentioned together with physical appearance (348/1247, 20% and 348/1247, 20%, respectively). These results shed some light on the multifactorial motivations underlying PA and hint at certain strategies individuals may gravitate to, depending on their primary motivational factor to initiate PA.

Comparing and contrasting our results to other studies using more traditional approaches presents challenges due to the different methodologies implemented. Nevertheless, our results are partly aligned with studies assessing PA motivation through questionnaires based on the Self-Determination Theory (SDT) framework, such as the Exercise Motivation Measurement scale (EMI-2) [38]. The EMI-2 takes the SDT spectrum of motivations and further divides them into subscales based on various topics (eg, "ill health avoidance," "social recognition"). For example, we identified 3 motivations for participating in PA (physical and mental health and physical appearance) consistent with the subscales identified in such studies. This supports our model's construct validity and SDT-based questionnaires' external validity, as our unprompted findings closely mirror several subscales of the established questionnaires. Yet, our results are

XSL•FO
RenderX

also misaligned with some of the literature. Regarding the EMI-2 scale, our model failed to identify specific "personal" and "social" motivation-related subscales, such as challenge, enjoyment, social recognition, and affiliation. We speculate that this misalignment stems from the unique characteristics of Reddit (public, online, anonymous posts) and the characteristics of the population that tends to use it (younger, male, western internet users) who may feel less inclined to discuss more personal motivations openly. We make this assumption with the caveat that we do not know if the demographic attributes of the subreddits we sampled truly mirror those of Reddit at large.

Our results seem to better align methodologically with qualitative studies based on semistructured interviews, showing a similar distribution of motivations in health-related topics [12,39,40]. For example, Ashton et al [40] identified 4 themes for motivators for PA among young adult men (a similar demographic to the average Reddit user): physical and mental health, physical appearance, social inclusion, and sport or performance. Here, we see that, similar to the quantitative literature, our model aligned with the health-related categories but did not identify social and personal motivational themes, such as skill development, support networks, community, relationships, and competitiveness. Direct comparison to qualitative studies on the topic is also challenging, given that they generally use thematic analysis as their research tool of choice. In contrast, we quantified the prevalence of topics within a large sample, as presented in this study, in a manner similar to qualitative content analysis. Furthermore, the process we used for assigning names to topic clusters, performed by a single reviewer without a formalized content analysis process, might have led to the misclassification of specific categories. Yet, a careful review of the text categorized by our model reveals a clear distinction between discussions on "how" to motivate oneself to engage in or adhere to PA versus the underlying "why" behind engaging in PA. We acknowledge this as a consideration for future research using wide-scale, self-disclosed content and recommend implementing established methods from the field of qualitative content analysis [41].

Several limitations of this study are worthy of discussion. First, the lack of data on the demographic attributes of the comment authors beyond platform-wide industry reports [25] limits our ability to associate the observed motivations and strategies with specific populations (ie, gender and age groups). Furthermore, although there are certain natural language processing techniques that can attempt to infer a user's gender based on previous comments written by the user, in our case, they were not able to assign gender to a large portion of our sample and thus excluded from the final analysis. Second, it is important to note that the questions that our analysis can answer are limited in scope as it was meant to be a proof-of-concept. The relatively simple queries we constructed were not designed to identify posts containing more detailed information, such as whether users were meeting minimal exercise standards, the location of exercise, or the results of implementing the identified strategies. In addition, all self-disclosed information, especially internet-based and pseudo-anonymous, must be interpreted with caution as the difference between disclosure and reality may be significant. Third, the "data-driven research" process used here allowed for substantial researcher degrees of freedom during the initial data selection and filtration process, potentially leading to selection bias, among others. Future studies could implement a more formalized process for topic classification and setting inclusion or exclusion criteria, enabling the answering of more complex questions and further refining our approach. Despite these limitations, incorporating human experts in auditing the automated results throughout the different stages of the analysis process is an essential strength of our methodology.

Conclusion

Our study introduces a novel usage of language models to harness the content shared on social media platforms to investigate the motivations and adherence strategies of individuals engaging in PA. Our findings further support the established literature and identified self-reported strategies implemented by this population for initiating and sustaining PA. The insights gained here provide valuable information and research tools for better understanding PA motives and adherence strategies. By leveraging self-disclosed data in an unprompted, nonresearch context on a large scale, we have expanded possibilities for research and gained more profound insights into people's thoughts, feelings, and opinions. This method has been successfully applied to various fields, and its application to PA and motivation represents a significant contribution. This study provides proof-of-concept for future investigations and opens new avenues for using language models in social media channels to study PA behavior and motivation.

Acknowledgments

Some minor use was made of ChatGPT (version 3.5, Open AI) for editing the abstract, specifically to reduce the word count.

Data Availability

The datasets generated during and/or analyzed during this study are available in the GitHub repository [42]. The JSON file includes the comment_id as "id" (as the textual content cannot be shared by the agreement rules); using the Reddit API, other researchers will be able to fully reconstruct the dataset. An example of the JSON file format is mentioned in Multimedia Appendix 2. Examples of the textual content can be found in Table 1.

XSL•FO
RenderX

Authors' Contributions

MSS conceptualized the research idea, wrote the software code to collect and curate the data, and performed the formal analysis. YG and IH supervised the research project and designed the study methodology equally. YS was responsible for writing the original draft. All authors contributed to the manuscript drafting and approved the final version before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic identification using clustering.
[DOCX File , 164 KB - jmir_v27i1e54489_app1.docx]

Multimedia Appendix 2
Example of the JSON file format.
[DOCX File , 15 KB - jmir_v27i1e54489_app2.docx]

References

1. Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
2. Ozemek C, Laddu DR, Lavie CJ, Claeys H, Kaminsky LA, Ross R, et al. An update on the role of cardiorespiratory fitness, structured exercise and lifestyle physical activity in preventing cardiovascular disease and health risk. Prog Cardiovasc Dis 2018;61(5-6):484-490. [doi: 10.1016/j.pcad.2018.11.005] [Medline: 30445160]
3. Shailendra P, Baldock KL, Li LK, Bennie JA, Boyle T. Resistance training and mortality risk: a systematic review and meta-analysis. Am J Prev Med 2022;63(2):277-285. [doi: 10.1016/j.amepre.2022.03.020] [Medline: 35599175]
4. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
5. Loyen A, Clarke-Cornwell AM, Anderssen SA, Hagströmer M, Sardinha LB, Sundquist K, et al. Sedentary time and physical activity surveillance through accelerometer pooling in four European countries. Sports Med 2017;47(7):1421-1435 [FREE Full text] [doi: 10.1007/s40279-016-0658-y] [Medline: 27943147]
6. Bennie JA, De Cocker K, Teychenne MJ, Brown WJ, Biddle SJH. The epidemiology of aerobic physical activity and muscle-strengthening activity guideline adherence among 383,928 U.S. adults. Int J Behav Nutr Phys Act 2019;16(1):34 [FREE Full text] [doi: 10.1186/s12966-019-0797-2] [Medline: 30999896]
7. Bennie JA, De Cocker K, Smith JJ, Wiesner GH. The epidemiology of muscle-strengthening exercise in Europe: a 28-country comparison including 280,605 adults. PLoS One 2020;15(11):e0242220 [FREE Full text] [doi: 10.1371/journal.pone.0242220] [Medline: 33237930]
8. Vasudevan A, Ford E. Motivational factors and barriers towards initiating and maintaining strength training in women: a systematic review and meta-synthesis. Prev Sci 2022;23(4):674-695 [FREE Full text] [doi: 10.1007/s11121-021-01328-2] [Medline: 34800250]
9. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. Med Sci Sports Exerc 2002;34(12):1996-2001. [doi: 10.1097/00005768-200212000-00020] [Medline: 12471307]
10. Brunet J, Sabiston CM. Exploring motivation for physical activity across the adult lifespan. Psychology of Sport and Exercise 2011;12(2):99-105. [doi: 10.1016/j.psychsport.2010.09.006]
11. Spiteri K, Broom D, Bekhet AH, de Caro JX, Laventure B, Grafton K. Barriers and motivators of physical activity participation in middle-aged and older-adults - a systematic review. J Aging Phys Act 2019;27(4):929-944. [doi: 10.1123/japa.2018-0343] [Medline: 31141447]
12. Allender S, Cowburn G, Foster C. Understanding participation in sport and physical activity among children and adults: a review of qualitative studies. Health Educ Res 2006;21(6):826-835. [doi: 10.1093/her/cyl063] [Medline: 16857780]
13. Stenner BJ, Buckley JD, Mosewich AD. Reasons why older adults play sport: a systematic review. J Sport Health Sci 2020;9(6):530-541 [FREE Full text] [doi: 10.1016/j.jshs.2019.11.003] [Medline: 33308804]
14. Brassey J, Mahtani K, Spencer E, Heneghan C. Catalogue of bias collaboration. 2017. URL: https://catalogofbias.org/biases/volunteer-bias/ [accessed 2024-11-22]
15. Elshaug C, Metzer J. Personality attributes of volunteers and paid workers engaged in similar occupational tasks. J Soc Psychol 2001;141(6):752-763. [doi: 10.1080/00224540109600586] [Medline: 11827223]
16. Carlo G, Okun MA, Knight GP, de Guzman MRT. The interplay of traits and motives on volunteering: agreeableness, extraversion and prosocial value motivation. Personality and Individual Differences 2005;38(6):1293-1305. [doi: 10.1016/j.paid.2004.08.012]
17. Paulhus DL. Chapter 2 - Measurement and control of response bias. In: Measures of Personality and Social Psychological Attitudes. Washington, DC: Academic Press; 1991:17-59.

XSL•FO
RenderX

18. Furnham A. Response bias, social desirability and dissimulation. Pers Individ Dif 1986;7(3):385-400. [doi: 10.1016/0191-8869(86)90014-0]
19. Weber SJ, Cook TD. Subject effects in laboratory research: an examination of subject roles, demand characteristics, and valid inference. Psychol Bull 1972;77(4):273-295. [doi: 10.1037/h0032351]
20. Mann S, Carter M. Emotional disclosures and reciprocal support: the effect of account type and anonymity on supportive communication over the largest parenting forum on reddit. Hum Behav Emerg 2021;3(5):668-676. [doi: 10.1002/hbe2.284]
21. Misoch S. Stranger on the internet: online self-disclosure and the role of visual anonymity. Comput Hum Behav 2015;48:535-541. [doi: 10.1016/j.chb.2015.02.027]
22. Clark-Gordon CV, Bowman ND, Goodboy AK, Wright A. Anonymity and online self-disclosure: a meta-analysis. Commun Res Rep 2019;32(2):98-111. [doi: 10.1080/08934215.2019.1607516]
23. Reddit global MAU 2022. Statista. URL: https://www.statista.com/statistics/1309791/reddit-mau-worldwide/ [accessed 2023-09-21]
24. Regional distribution of desktop traffic to Reddit.com as of March 2024, by country. Statista. URL: https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/ [accessed 2023-09-21]
25. Atske S. Social media use in 2021. Pew Research Center: internet & technology. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2023-09-21]
26. Zhu J, Jurgens D. The structure of online social networks modulates the rate of lexical change. ArXiv. Preprint posted online on April 11, 2021 2021 [FREE Full text] [doi: 10.18653/v1/2021.naacl-main.178]
27. Koncar P, Walk S, Helic D. Analysis and prediction of multilingual controversy on Reddit. New York, NY, USA: Association for Computing Machinery; 2021 Presented at: Proceedings of the 13th ACM Web Science Conference 2021; 2021 June 21 - 25; United Kingdom p. 215-224 URL: https://dl.acm.org/doi/10.1145/3447535.3462481 [doi: 10.1145/3447535.3462481]
28. Amershi S, Cakmak M, Knox WB, Kulesza T. Power to the people: the role of humans in interactive machine learning. AI Magazine 2014;35(4):105-120. [doi: 10.1609/aimag.v35i4.2513]
29. Kaye J. Interaction ASIG on CH. 2016 Presented at: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; 2016 May 7-12; San Jose, CA, USA. [doi: 10.1145/2851581]
30. How to Use UMAP. UMAP. URL: https://umap-learn.readthedocs.io/en/latest/basic_usage.html [accessed 2024-11-22]
31. fuzzy-c-means. PyPI. URL: https://pypi.org/project/fuzzy-c-means/ [accessed 2024-11-22]
32. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. ArXiv. Preprint posted online on August 27, 2019 2019 [FREE Full text] [doi: 10.18653/v1/d19-1410]
33. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. ArXiv. Preprint posted online on September 18, 2020 2020 [FREE Full text]
34. Shen YM, Nataliani Y. Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. Pattern Recognit 2017;71:45-59. [doi: 10.1016/j.patcog.2017.05.017]
35. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv. Preprint posted online on May 24, 2019 2019 [FREE Full text]
36. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
37. Stockwell S, Trott M, Tully M, Shin J, Barnett Y, Butler L, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. BMJ Open Sport Exerc Med 2021;7(1):e000960 [FREE Full text] [doi: 10.1136/bmjsem-2020-000960] [Medline: 34192010]
38. Markland D, Ingledew DK. The measurement of exercise motives: factorial validity and invariance across gender of a revised exercise motivations inventory. Br J Health Psychol 2011;2(4):361-376. [doi: 10.1111/j.2044-8287.1997.tb00549.x]
39. Caperchione CM, Vandelanotte C, Kolt GS, Duncan M, Ellison M, George E, et al. What a man wants: understanding the challenges and motivations to physical activity participation and healthy eating in middle-aged Australian men. Am J Mens Health 2012;6(6):453-461 [FREE Full text] [doi: 10.1177/1557988312444718] [Medline: 22516565]
40. Ashton LM, Hutchesson MJ, Rollo ME, Morgan PJ, Thompson DI, Collins CE. Young adult males' motivators and perceived barriers towards eating healthily and being active: a qualitative study. Int J Behav Nutr Phys Act 2015;12:93 [FREE Full text] [doi: 10.1186/s12966-015-0257-6] [Medline: 26169503]
41. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
42. gitMichal / motivation_reddit. GitHub. URL: https://github.com/gitMichal/motivation_reddit/ [accessed 2024-12-04]

Abbreviations

API: application programming interface
EMI-2: Exercise Motivation Measurement scale
FCM: fuzzy-c-mean
PA: physical activity
SDT: Self-Determination Theory

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 11.11.23; peer-reviewed by D Carvalho, L Wreyford; comments to author 07.04.24; revised version received 23.05.24; accepted 22.09.24; published 13.01.25.

Please cite as:
Shmueli-Scheuer M, Silverman Y, Halperin I, Gepner Y
Analysis of Reddit Discussions on Motivational Factors for Physical Activity: Cross-Sectional Study
J Med Internet Res 2025;27:e54489
URL: https://www.jmir.org/2025/1/e54489
doi:10.2196/54489
PMID:

©Michal Shmueli-Scheuer, Yedidya Silverman, Israel Halperin, Yftach Gepner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analysis of “Dr Ding Xiang” on WeChat in China to Determine Factors Influencing Readership on Medical Social Media: Observational Study

Jiaman Liao[1], MPH; Xueliang Huang[1], MBBS; Hao Huang[1], MBBS; Cuina Shen[1], BSN; Lixia Li[1], BSN; Yushao Li[1], MBBS; Yiqiang Zhan[2], PhD

[1]Nanlian Community Health Service Center, Shenzhen Longgang Central Hospital, Shenzhen, China

[2]Department of Epidemiology, School of Public Health, Sun Yat-Sen University, Shenzhen, China

Corresponding Author:
Yiqiang Zhan, PhD
Department of Epidemiology
School of Public Health
Sun Yat-Sen University
No 66 Gongchang Road, Guangming District
Shenzhen, 518107
China
Phone: 86 075584827713
Email: zhanyq8@mail.sysu.edu.cn

Abstract

Background: With the rapid expansion of social media platforms, the demand for health information has increased substantially, leading to innovative approaches and new opportunities in health education.

Objective: This study aims to analyze the characteristics of articles published on the “Dr Ding Xiang” WeChat official account (WOA), one of the most popular institutional accounts on the WeChat platform, to identify factors influencing readership engagement and to propose strategies for enhancing the effectiveness of health information dissemination.

Methods: A total of 5286 articles published on the “Dr Ding Xiang” WOA from January 2021 to December 2021 were collected and analyzed. Additionally, a random sample of 324 articles was selected for detailed text analysis. Univariate analysis was conducted using the chi-square test, and multivariate analysis was performed using multivariable logistic regression.

Results: In 2021, the total number of reads for “Dr Ding Xiang” articles reached 323,479,841, with an average of 61,196 reads per article. Articles exceeding 100,000 reads accounted for 33.90% of the total. Most articles were published during the time slots of 8:00-10:00 AM, 11:30 AM to 1:30 PM, and 8:30-10:30 PM. Analysis indicated that the order of publication, style of the title sentence, number of likes, number of in-views, total likes on comments, and number of replies to comments were significantly associated with an article’s number of reads. Text analysis further revealed that the article’s reasoning approaches and concluding methods also had a significant impact on readership.

Conclusions: To enhance readership and the effectiveness of health communication, health-focused WOAs should consider key factors such as optimal publication timing, engaging title design, and effective content structuring. Attention to these elements can improve user engagement and support the broader dissemination of health information.

(J Med Internet Res 2025;27:e65372) doi:10.2196/65372

KEYWORDS

WeChat Official Accounts; Dr Ding Xiang; health communication; information dissemination; readership analysis

Introduction

The rapid expansion of social media has significantly increased the demand for accessible health information, fostering innovative approaches in health education [1-3]. Facebook is widely used to disseminate health content and build community support, especially in areas like chronic disease management, vaccination campaigns, and mental health [4]. Public health organizations, such as the Centers for Disease Control and Prevention and the World Health Organization, use Facebook

XSL•FO
RenderX

to establish digital platforms promoting preventive measures and healthy lifestyles [5]. Twitter's hashtag feature enhances its capacity for real-time communication in health education and emergency responses [6]. For example, during the COVID-19 pandemic, hashtags like #COVID19 and #HealthTips facilitated quick access to information and enabled user engagement [7]. YouTube has become a leading platform for health education videos on topics such as cardiopulmonary resuscitation, vaccines, and nutrition, with organizations and health influencers producing content to improve public understanding through visual demonstrations [8].

In contrast to Western social media, WeChat has become the leading channel for health communication in China, owing to its unique privacy features, strong social networking capabilities, and multifunctionality. Since its launch by Tencent in 2011, WeChat has evolved beyond a communication tool to offer social networking, marketing, media sharing, and utilities, making it an adaptable platform for mobile health engagement [9].

The WeChat official account (WOA) feature, introduced in 2012, allows organizations, businesses, and individuals to create public accounts to broadcast information and engage users. Through WOAs, users can receive notifications, read articles, engage with participatory content, and access various services [10]. This feature has been widely adopted for health information dissemination. For instance, the Wuxi Center for Disease Control and Prevention achieved high engagement by sharing articles on nutrition and food safety, highlighting that targeted content and strategic layouts can enhance reader interaction [11]. Similarly, health programs on WeChat, such as those aimed at improving parental health literacy, underscore the platform's effectiveness in delivering personalized health education [12].

WeChat's engagement features, including likes and comments, amplify user participation and promote the spread of health information. Studies suggest that user engagement, indicated by liking behavior, is closely linked to interest and approval [10]. WOAs are also instrumental in chronic disease management and patient education; for example, a WeChat-based management program for cough-variant asthma demonstrated significant clinical success [13]. Moreover, a life review program for patients with cancer highlighted WeChat's potential to provide psychological support and enhance quality of life [14]. During the COVID-19 pandemic, WeChat played a crucial role in health information dissemination and public health management [15], and was used effectively to promote health behaviors, such as encouraging HIV testing [16].

Overall, WeChat has demonstrated its effectiveness as a tool for health communication and education, providing a robust platform for enhancing public health literacy, improving disease management, and delivering essential psychological support. However, there is a need for further research into optimal posting strategies and the dissemination effects of WOA articles. Addressing these gaps is crucial for understanding dissemination effectiveness and refining posting strategies. Consequently, this study conducts both a general and detailed text analysis of the articles published by "Dr Ding Xiang" WOA in 2021, aiming to explore the factors that influence readership engagement with health-related WeChat articles. The goal is to provide a scientific basis for government departments to develop effective public account strategies and enhance the overall efficacy of health communication.

Methods

Study Materials

The "Dr Ding Xiang" WOA used in this study was established by Ding Xiangyuan, a prominent Chinese health care company founded in 2000. Ding Xiangyuan specializes in medical information services and health content, with a commitment to providing the public with scientific and authoritative health knowledge. The "Dr Ding Xiang" WOA serves as Ding Xiangyuan's public health education brand, focusing on disseminating reliable health information and enhancing public health literacy. All content is edited and reviewed by a professional team to ensure both accuracy and authority.

According to the 2019 China WeChat 500 annual list of public accounts released by the New Rank platform, "Dr Ding Xiang" ranked eighth out of 981,144 WOAs and secured the top spot among health care–related WOAs, with an average of more than 90,250 reads per article [17]. This ranking underscores the notable influence and representativeness of the "Dr Ding Xiang" WOA in disseminating health information. Therefore, a comprehensive analysis was conducted on 5286 articles published by the "Dr Ding Xiang" WOA from January 2021 to December 2021. Additionally, a sample of 324 articles was randomly selected for in-depth text analysis, using simple random sampling to ensure representativeness.

Data Collection

For the main analysis, data were sourced from the New Rank platform [18], accessed on September 20, 2022, a recognized authority in China for assessing the data value of WOA content. We collected and compiled data on 5286 articles published by "Dr Ding Xiang" in 2021. The dataset included various attributes such as time of publication, order of publication, style of the title sentence, presence of authors, originality of the articles, use of multimedia, number of likes, number of in-views, number of comments, total likes on comments, number of replies to comments, and total likes on replies to comments. This extensive dataset facilitated a detailed examination of the publication characteristics of each article.

For the text analysis, a sample of 324 articles was randomly selected to investigate specific text attributes, including the article's thematic contents, opening styles, reasoning approaches, and concluding methods. The objective was to identify factors that significantly influence readership and engagement with these health-related articles. By analyzing these attributes, we aimed to uncover elements that most effectively enhance audience interaction and engagement, thereby providing valuable insights for crafting successful health communication strategies on WOAs.

XSL•FO
RenderX

Variables Assessment

The time of publication refers to the exact date and time when an article is made available to the public. This grouping is based on our observations of user behavior, which indicate that peak browsing and interaction times typically occur between 8:00-10:00 AM, 11:30 AM to 1:30 PM, and 8:30-10:30 PM. These periods likely align with daily commuting, lunch breaks, and evening leisure activities, respectively.

The order of publication indicates an article placement within a multiarticle push, where position 1 is the headline and typically receives the highest exposure, with up to 8 positions available per push. The style of the title sentences is categorized based on the presence of exclamation marks and question marks. An article is deemed to have an author if the author's name is signed beneath the title. Articles carrying an "original" tag are considered original. Beyond images and text, articles can also include videos and audio links; in this study, the inclusion of such elements is referred to as multimedia use.

Each article includes the "Like" and "In-view" buttons at the bottom right corner, from which the counts for likes and in-views are obtained. Here, it is important to clarify the specific function of "In-view" in this context. When a user clicks the "In-view" button, the article appears in the user's WeChat "Discover > Take a look > In-view" section, making it visible to the user's WeChat contacts. Unlike "Like," which merely indicates user appreciation, the "In-view" feature serves as a form of recommendation, enhancing the article's visibility among the user's social network.

Articles also feature a commenting function, with the number of comments determined by the total number of messages left on each article. Each comment also has a "Like" button, and the sum of all likes on comments constitutes the total number of likes on comments. When readers leave comments, both the operator and other readers can reply, and these replies are counted as comment replies. If readers agree with the replies, they can also like them, and the total number of likes on replies is recorded.

On the WeChat platform, the recommendation mechanism promotes content based on article interaction data, such as the number of likes and comments. Articles that receive higher levels of interaction are more likely to be recommended to a broader audience. Additionally, behaviors such as liking and commenting often lead to social sharing, where users may forward articles to their social circles or share them directly with friends [19]. Given these dynamics, the number of likes and comments on the WeChat platform serves not only as a user's response to the content but may also contribute to an increase in readership through both the recommendation system and social dissemination. Consequently, this study considers the number of likes and comments as potential factors influencing reading volume.

The article thematic contents included beauty and fitness (focusing on skincare, makeup, fitness, and weight loss), maternal and child health (covering obstetrics, gynecology, pediatrics, and early childhood education, targeting individuals planning to get pregnant, pregnant women, and parents of children aged 0-6 years), gender health (including topics on marriage, contraception, and sexually transmitted diseases), common diseases (covering prevalent health conditions), and non–health-related topics (such as product sales and recruitment).

The article opening styles included introductory style (introduced by the main subject, related topics, questions, or proverbs and sayings), case style (opening with personal or third-party cases), festive style (opening with mentions of festivals, seasons, or specific times of the year), social news (starting with a news event), and other styles (directly beginning with the main text content).

The article reasoning approaches used in the articles include specialized knowledge explanations (using professional knowledge to explain), data, guidelines, and experiment-based explanations (citing data, research guidelines, and experiments), case-based explanations (using specific cases to explain), and other forms (using various other forms of reasoning). To ensure consistency and comparability in categorization, articles were assigned to the appropriate category based on the predominant type of reasoning used. For instance, if an article primarily serves as an explanation of professional knowledge, supplemented by case studies, it would be classified as an "explanation of professional knowledge."

The article concluding methods include summary endings (refining and summarizing the main points of the article), supplementary tips (reemphasizing important content and adding relevant information), call-to-action endings (encouraging the audience to take specific actions), advertising and promotional endings (promoting product purchases and live broadcasts), and other endings (a few articles simply conclude with the main text content).

Quality Control

The data on the number of reads, number of in-views, number of likes, number of comments, total number of comments, number of replies to comments, total number of replies to comments, and other related metrics collected in this study were based on the statistics as of September 20, 2022. To ensure the accuracy of data entry for determining the article topic, opening style, reasoning form, and ending style, 2 coauthors were trained for this. They independently entered data for 324 randomly selected articles. In cases of disagreement, the final confirmation was made through discussion and consensus among a third author.

Statistical Analysis

Articles were categorized into 2 groups based on reading volume: a high-reads group (100,000 and above reads) and a low-reads group (fewer than 100,000 reads). Independent variables included the time of publication, order of publication, style of the title sentence, presence of authors, originality of articles, use of multimedia, number of likes, number of in-views, number of comments, number of total likes on comments, number of replies to comments, number of total likes on replies to comments, article thematic contents, article opening styles, article reasoning approaches, and article concluding methods. A chi-square test was used for univariate analysis, with

XSL•FO
RenderX

statistically significant variables included in a logistic regression model. A *P* value<.05 was considered statistically significant. All analyses were conducted using R (version 4.1; R Foundation for Statistical Computing). Model fit was assessed with the Hosmer-Lemeshow test, where a *P* value>.05 indicates a good fit (ie, no significant difference between observed and predicted values). To address multicollinearity, variance inflation factor (VIF) and tolerance values were calculated. A tolerance value of 0.1 or greater and a VIF value of less than 10 were considered indicators of no multicollinearity.

Ethical Considerations

This study was conducted using publicly available data from the New Rank platform, which was accessed in accordance with the platform's terms of service. All data included in this study were either anonymous or deidentified prior to analysis, ensuring that no personally identifiable information was collected or processed. As the study involved the observation of publicly available data and secondary analyses, no direct interaction with individuals occurred, and no formal ethical approval was required. To further protect privacy, the results are presented in aggregate form without identifying specific individuals or posts. These measures align with established ethical standards for research involving publicly accessible data.

Results

Basic Characteristics of Published Articles

According to real-time statistics on September 20, 2022, a total of 5286 articles were published on the "Dr Ding Xiang" WOA in 2021. On average, about 14 articles were posted daily. Each article garnered an average of approximately 61,196 reads, with the lowest recorded number being 3577 reads. Among these, 1792 (33.90%) articles surpassed 100,000 reads. The publication times of the articles were predominantly clustered into 3 periods: 8:00-10:00 AM, which saw 1733 (32.78%) articles; 11:30 AM to 1:30 PM, with 1892 articles (35.79%); and 8:30-10:30 PM, during which 1642 (31.06%) articles were published.

The articles accumulated a total of 273,712,120 likes, averaging approximately 518 likes per article. Additionally, they received 1,848,787 in-views, with an average of around 350 in-views per article. The total number of comments across all articles was 53,495, averaging about 10 comments per article. Collectively, likes on comments totaled 1,595,956, translating to an average of approximately 2856 interactions per article. The total number of replies to comments reached 20,722, averaging about 4 replies per article. Lastly, likes on replies to comments amounted to 5,573,912, with an average of roughly 1054 such interactions per article.

Univariate Analysis

The univariate analysis revealed that several factors were significantly associated with the number of reads for the articles. Specifically, the order of publication (χ^2_2=4040.2; *P*<.001), style of the title sentence (χ^2_3=66.1; *P*<.001), presence of authors (χ^2_1=219.8; *P*<.001), and originality of the articles (χ^2_1=1029.6; *P*<.001) were all important determinants. Additionally, the use of multimedia (χ^2_1=31.0; *P*<.001), as well as various engagement metrics such as the number of likes (χ^2_2=2290.0; *P*<.001), in-views (χ^2_2=2506.8; *P*<.001), comments (χ^2_2=2255.3; *P*<.001), total likes on comments (χ^2_2=2397.5; *P*<.001), replies to comments (χ^2_2=1750.4; *P*<.001), and total likes on replies to comments (χ^2_2=2306.4; *P*<.001) also showed significant impacts. These findings are further detailed in Table 1.

XSL•FO
RenderX

Table 1. Univariate analysis of the readership of 5286 articles on "Dr Ding Xiang" WOA[a] in 2021.

Variables	Total, n	High reads, n (%)	Low reads, n (%)	*P* value
The time of publication				.25
8:00-10:00 AM	1733	1171 (67.6)	562 (32.4)	
11:30 AM to 1:30 PM	1892	1245 (65.8)	647 (34.2)	
8:30-10:30 PM	1642	1068 (65)	574 (35)	
Other time slots	19	10 (52.6)	9 (47.4)	
The order of publication				<.001
In the first to second positions	1877	194 (10.3)	1683 (89.7)	
In the third to fifth positions	2600	2503 (96.3)	97 (3.7)	
In the sixth to eighth positions	809	797 (98.5)	12 (1.5)	
The style of the title sentence				<.001
Declarative sentences	1828	1091 (59.7)	737 (35.3)	
Interrogative sentences	2027	1461 (72.1)	566 (27.9)	
Exclamatory sentences	1207	797 (66)	410 (34)	
Interrogative exclamatory sentences	224	145 (64.7)	79 (35.3)	
The presence of authors				<.001
Not have	875	768 (87.8)	107 (12.2)	
There are	4411	2726 (61.8)	1685 (38.2)	
The originality of the articles				<.001
Original	4006	3121 (77.9)	885 (22.1)	
Not original	1280	373 (29.1)	907 (70.9)	
The use of multimedia				<.001
Not used	5117	3416 (66.8)	1701 (33.2)	
Used	169	78 (46.2)	91 (53.8)	
The number of likes				<.001
0-100	2906	2579 (88.7)	327 (11.3)	
101-300	1338	841 (62.9)	497 (37.1)	
301 and greater	1042	74 (7.1)	968 (92.9)	
The number of in-views				<.001
0-50	2828	2588 (91.5)	240 (8.5)	
51-150	1364	823 (60.3)	541 (39.7)	
151 and greater	1094	83 (7.6)	1011 (92.4)	
The number of comments				<.001
0-5	2473	2303 (93.1)	170 (6.9)	
6-15	1616	1022 (63.2)	594 (36.8)	
16 and greater	1197	169 (14.1)	1028 (85.9)	
The number of total likes on comments				<.001
0-100	2457	2261 (92)	196 (8)	
101-1000	1565	1082 (69.1)	483 (30.9)	
1001 and greater	1264	151 (11.9)	1113 (88.1)	
The number of replies to comments				<.001
0-1	2514	2233 (88.8)	281 (11.2)	
2-5	1507	1001 (66.4)	506 (33.6)	

XSL•FO
RenderX

Variables	Total, n	High reads, n (%)	Low reads, n (%)	*P* value
6 and greater	1265	260 (20.6)	1005 (79.4)	
The number of total likes on replies to comments				<.001
0-10	2402	2194 (91.3)	208 (8.7)	
11-250	1594	1129 (70.8)	465 (29.2)	
251 and greater	1290	171 (13.3)	1119 (86.7)	

[a]WOA: WeChat official account.

Multivariable Analysis

The multivariable analysis results indicated that several factors significantly enhanced readership. The order of publication had a considerable impact, with articles posted in the first to second positions showing an odds ratio (OR) of 745.654 (95% CI 358.061-1552.809; *P*<.001) and those in the third to fifth positions showing an OR of 2.719 (95% CI 1.391-5.315; *P*=.003). The style of the title sentence also influenced readership, particularly declarative sentences (OR 2.333, 95% CI 1.572-3.461; *P*<.001) and interrogative sentences (OR 1.928, 95% CI 1.297-2.867; *P*=.001). The number of likes significantly boosted readership, with likes in the range of 101-300 (OR 2.308, 95% CI 1.384-3.848; *P*=.001) and 301 and greater (OR 14.050, 95% CI 5.949-33.184; *P*<.001). Similarly, the number of in-views positively impacted readership, particularly for in-views in the range of 51-150 (OR 2.739, 95% CI 1.736-4.321; *P*=.001) and 151 and greater (OR 4.087, 95% CI 1.617-9.197; *P*<.001). The number of total likes on comments also boosted readership, with ranges of 101-1000 likes (OR 3.082, 95% CI 2.125-4.470; *P*<.001) and 1001 and greater (OR 9.747, 95% CI 5.515-17.226; *P*<.001). Additionally, the number of replies to comments had a positive effect, especially for comments liked 1001 times and greater (OR 9.747, 95% CI 5.515-17.226; *P*<.001) and the number of replies to comments reaching 6 and greater (OR 2.374, 95% CI 1.560-3.612; *P*<.001). These factors are summarized in Table 2.

The goodness of fit tests for the fitted binary logistic regression model was greater than the significance level of .05, which indicates the model adequately fits the data. In addition, the results also show that the VIF values of all independent variables are less than 10 and the tolerance values are greater than 0.1, indicating that the effect of multicollinearity on the model is negligible.

XSL•FO
RenderX

Table 2. Multivariable analysis of the readership of 5286 articles of "Dr Ding Xiang" WOA[a] in 2021.

Factor	OR[b] (95% CI)	*P* value
The order of publication		<.001
In the sixth to eighth positions	1 (reference)	—[c]
In the third to fifth positions	2.719 (1.391-5.315)	.003
In the first to second positions	745.654 (358.061-1552.809)	<.001
The style of the title sentence		<.001
Exclamatory sentences	1 (reference)	—
Declarative sentences	2.333 (1.572-3.461)	<.001
Interrogative sentences	1.928 (1.297-2.867)	.001
Interrogative exclamatory sentences	1.38 (0.676-2.817)	.38
The number of likes		<.001
0-100	1 (reference)	—
101-300	2.308 (1.384-3.848)	.001
301 and greater	14.05 (5.949-33.184)	<.001
The number of in-views		<.001
0-50	1 (reference)	—
51-150	2.739 (1.736-4.321)	<.001
151 and greater	4.087 (1.817-9.197)	.001
The number of total likes on comments		<.001
0-100	1 (reference)	—
101-1000	3.082 (2.125-4.470)	<.001
1001 and greater	9.747 (5.515-17.226)	<.001
The number of replies to comments		<.001
0-1	1 (reference)	—
2-5	1.208 (0.845-1.727)	.30
6 and greater	2.374 (1.560-3.612)	<.001

[a]WOA: WeChat official account.

[b]OR: odds ratio.

[c]Not applicable.

Text Analysis for the Selected Articles

The text analysis of the selected articles revealed a diverse distribution of topics. Beauty and fitness were the most common topics, comprising 80 (24.7%) articles, followed by common diseases with 77 (23.8%) articles. Maternal and child health, gender health, and topics unrelated to health accounted for 20 (6.2%) articles, 22 (6.8%) articles, and 95 (29.3%) articles, respectively. Regarding the opening form, introductions were prevalent in 203 (62.7%) articles, while case studies were used in 54 (16.7%) articles. Articles that began with festivals, social news, or other forms made up 22 (6.8%) articles, 18 (5.6%) articles, and 27 (8.3%) articles, respectively. In terms of reasoning forms, specialized knowledge was used in 160 (49.4%) articles, followed by data citation, research guidelines, and experiments in 47 (14.5%) articles. Experiments constituted the reasoning form in 12 (3.7%) articles, case studies in 40 (12.3%) articles, and other reasoning forms in 77 (23.8%) articles. As for the conclusions, 26 (8%) articles provided a summary, while hints and calls for action were each present in 42 (13%) articles. A significant number of articles concluded with advertisements and promotions, totaling 178 (54.9%) articles, with 36 (11.1%) articles using other forms of conclusions.

Univariate and Multivariate Analyses

The univariate analysis showed that the article opening styles (χ^2_4=10.6; *P*=.03), the article reasoning approaches (χ^2_3=23.7; *P*<.001), and the article concluding methods (χ^2_4=22.4; *P*<.001) significantly affected article readership, as detailed in Table 3.

XSL•FO
RenderX

Table 3. Univariate analysis of the readership of 324 articles on "Dr Ding Xiang" WOAa in 2021.

Variables	Total, n	High reads, n (%)	Low reads, n (%)	*P* value
The article thematic contents				.08
Beauty and fitness	80	40 (50)	40 (50)	
Maternal and child health	50	35 (70)	15 (30)	
Gender health	22	11 (50)	11 (50)	
Common disease	77	52 (67.5)	25 (32.5)	
Non–health-related	95	54 (56.8)	41 (43.2)	
The article opening styles				.03
Introductory style	203	123 (60.6)	80 (39.4)	
Case style	54	23 (43.6)	31 (57.4)	
Festive style	22	13 (59.1)	9 (40.9)	
Social news	18	12 (66.7)	6 (33.3)	
Other styles	27	21 (77.8)	6 (22.2)	
The article reasoning approaches				<.001
Specialized knowledge explanations	160	112 (70)	48 (30)	
Data, guidelines, and experiment-based explanations	47	21 (44.7)	26 (55.3)	
Case-based explanations	40	13 (32.5)	27 (67.5)	
Other forms	77	46 (59.7)	31 (40.3)	
The article concluding methods				<.001
Summary endings	26	7 (26.9)	19 (73.1)	
Supplementary tips	42	25 (59.5)	17 (40.5)	
Call to action endings	42	19 (45.2)	23 (54.8)	
Advertising and promotional endings	178	112 (62.9)	66 (37.1)	
Other endings	36	29 (80.6)	7 (19.4)	

[a]WOA: WeChat official account.

In the multivariate analysis, it was evident that specific reasoning approaches and concluding methods had substantial impacts on readership engagement. Articles citing data, guidelines, and experiment-based explanations (OR 2.288, 95% CI 1.132-4.623; *P*=.02), using case-based explanations (OR 4.713, 95% CI 2.169-10.237; *P*<.001), notably boosted readership. Conversely, certain concluding methods such as supplementary tips (OR 0.267, 95% CI 0.089-0.805; *P*=.02), Advertising and promotional endings (OR 0.243, 95% CI 0.092-0.641; *P*=.004), and other endings (OR 0.094, 95% CI 0.027-0.327; *P*<.001) were found to inhibit readership. These findings are comprehensively summarized in Table 4.

The goodness of fit tests for the fitted binary logistic regression model was greater than the significance level of .05, which indicates the model adequately fits the data. In addition, the results also show that the VIF values of all independent variables are less than 10 and the tolerance values are greater than 0.1, indicating that the effect of multicollinearity on the model is negligible.

XSL•FO
RenderX

Table 4. Multivariate analysis of the readership of 324 articles on "Dr Ding Xiang" WOA[a] in 2021.

Variables	OR[b] (95% CI)	*P* value
The article reasoning approaches		
Specialized knowledge explanations	1 (reference)	—[c]
Data, guidelines, and experiment-based explanations	2.288 (1.132-4.623)	.02
Case-based explanations	4.713 (2.169-10.237)	<.001
Other forms	1.828 (1.006-3.320)	.05
The article concluding methods		
Summary endings	1 (reference)	—
Supplementary tips	0.267 (0.089-0.805)	.02
Call to action ending	0.459 (0.153-1.372)	.16
Advertising and promotional endings	0.243 (0.092-0.641)	.004
Other endings	0.094 (0.027-0.327)	<.001

[a]WOA: WeChat official account.

[b]OR: odds ratio.

[c]Not applicable.

Discussion

Principal Results

Our study reveals that headline and subheadline articles achieve significantly higher readership compared to subordinate articles. This is substantiated by previous research indicating that the first article typically garners the highest attention, with subsequent articles experiencing a sharp decline in readership. This can be attributed to the initial engagement creating a strong first impression, which diminishes as readers move through additional content [20]. Hence, crafting compelling headlines and engaging thumbnails becomes critical in capturing and retaining reader attention. These elements act as the primary hooks that encourage readers to click on and continue through the content, thereby enhancing the overall readership of subsequent articles [21].

Survey data from China provide insights into reader behavior, revealing that most readers skim the body text after reading the headline, highlighting the headline's pivotal role in capturing initial interest [22]. Our analysis emphasizes the importance of linguistic choices in headlines. Studies show that using emotion-laden words, third-person pronouns, and clear definitive articles can significantly enhance click-through rates by appealing to readers' emotions and curiosity [23]. Specifically, interrogative headlines, which pose questions, are particularly effective in engaging readers as they evoke a sense of suspense and intrigue. This finding aligns with research from ScienceDaily and Azura Magazine, which suggests that questions naturally stimulate curiosity and engagement [24,25].

However, titles with an exclamatory tone, often perceived as exaggerated or sensational, can lead readers to question the credibility of the content, thereby diminishing their trust. Research on health education communication highlights trust as a crucial factor in effective information dissemination [26]. While dramatic language and punctuation may capture attention quickly, they can also give the impression that the information is unreliable or inaccurate.

To address this, declarative or interrogative titles are recommended for conveying the article's core content authentically and effectively. Declarative titles clearly communicate the main idea, set accurate expectations, and establish a direct communication style, which is critical for building trust in health-related content. Conversely, interrogative titles stimulate curiosity by posing questions, encouraging readers to explore the content further without compromising its credibility. By fostering trust through accurate and engaging titles, authors can enhance reader satisfaction and promote meaningful, long-term interactions with health information.

The study highlights a significant relationship between article reads, likes, and engagement in the comment sections. This relationship may not only reflect users' interest in the content but may also indirectly impact reading volume through WeChat's recommendation mechanism and users' social sharing behavior. Driven by this recommendation mechanism, articles with higher interaction data receive greater exposure, while users' liking and commenting behavior can further stimulate retweeting within their social circles, thereby expanding the reach of the articles. Consequently, on the WeChat platform, engagement behaviors may serve both as a reflection of reading and as a potential influencing factor in enhancing reading volume.

This study's classification system is based on the varying impacts of reasoning types in articles published on WOAs [27]. Our findings indicate that explanations based on data, guidelines, experiments, cases, or other forms were read more frequently than specialized knowledge explanations. These formats make the content more relatable and easier to understand. Furthermore, summary endings were preferred over supplementary tips and advertising and promotional endings, as they save readers' time,

XSL•FO
RenderX

emphasize key points, and enhance the content's value, retention, and shareability [28,29].

The insights derived from this study hold critical implications for public health promotion and information dissemination. Effectively engaging readers through strategic headline crafting, linguistic choices, and participatory features can significantly enhance the reach and impact of health messages. By adopting these strategies, public health communicators can ensure that vital health information is disseminated widely and engaged with meaningfully, thereby promoting healthier behaviors and outcomes within the community.

Limitations

The strengths of this study lie in its empirical basis, comprehensive data collection over a 1-year period, and the bifurcated approach of general and text analysis, providing a well-rounded investigation of factors influencing readership. However, the study also has limitations. It focuses on a single public account, restricting the analysis to a longitudinal perspective without cross-sectional validation against other health-related WOAs. Additionally, WeChat readership metrics cap at 100,000 reads, preventing accurate assessment for articles exceeding this threshold and thus limiting readership as a true continuous variable. To address this, we dichotomized readership into “high reads” (100,000 reads and more) and “low reads” (less than 100,000 reads) for analysis. Finally, the text analysis relies on subjective definitions of the article’s thematic contents, the opening styles, the reasoning approaches, and the concluding methods, which may introduce bias despite the random article selection.

Conclusions

In summary, the study offers valuable insights into effective strategies for enhancing readership and engagement, which are crucial for successful health communication and promotion. By understanding and leveraging these factors, public health professionals can better capture and retain reader interest, ensuring that essential health information reaches and resonates with a broader audience.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to privacy concerns and ethical considerations related to user interaction metrics (eg, likes, comments, and views) but are available from the corresponding author on reasonable request.

Authors' Contributions

JL, XH, and YZ designed this study. JL and YZ drafted the manuscript. All other authors provided critical contributions and approved the final version.

Conflicts of Interest

None declared.

References

1. Kienle GS, Ben-Arye E, Berger B, Cuadrado Nahum C, Falkenberg T, Kapócs G, et al. Contributing to global health: development of a consensus-based whole systems research strategy for anthroposophic medicine. Evid Based Complement Alternat Med 2019;2019:3706143 [FREE Full text] [doi: 10.1155/2019/3706143] [Medline: 31781267]
2. Anthony B. Investigating the implementation of telehealth and digital technologies during public health crisis: a qualitative review. Int J Health Plann Manage 2023;38(5):1212-1227. [doi: 10.1002/hpm.3681] [Medline: 37452470]
3. Gunasekeran DV, Tseng R, Tham YC, Wong TY. Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. NPJ Digit Med 2021;4(1):40. [doi: 10.1038/s41746-021-00412-9] [Medline: 33637833]
4. Heldman AB, Schindelar J, Weaver JB. Social media engagement and public health communication: implications for public health organizations being truly “social”. Public Health Rev 2013;35(1):13. [doi: 10.1007/BF03391698]
5. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
6. Herrera-Peco I, de la Torre-Montero JC. Preface of special issue "Cares in the Age of Communication: Health Education and Healthy Lifestyles": Social media and health communication in a pandemic? Eur J Investig Health Psychol Educ 2020;10(2):575-578 [FREE Full text] [doi: 10.3390/ejihpe10020042] [Medline: 34542521]
7. Khan MN, Ashraf MA, Seinen D, Khan KU, Laar RA. Social media for knowledge acquisition and dissemination: the impact of the COVID-19 pandemic on collaborative learning driven social media adoption. Front Psychol 2021;12:648253. [doi: 10.3389/fpsyg.2021.648253] [Medline: 34135814]
8. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015;21(3):173-194 [FREE Full text] [doi: 10.1177/1460458213512220] [Medline: 24670899]
9. Montag C, Becker B, Gan C. The multipurpose application WeChat: a review on recent research. Front Psychol 2018;9:2247. [doi: 10.3389/fpsyg.2018.02247] [Medline: 30618894]

XSL•FO
RenderX

10. Li W, Han LQ, Guo YJ, Sun J. Using WeChat official accounts to improve malaria health literacy among Chinese expatriates in niger: an intervention study. Malar J 2016;15(1):567 [FREE Full text] [doi: 10.1186/s12936-016-1621-y] [Medline: 27881122]
11. Tu S, Yan X, Jie K, Ying M, Huang C. WeChat: an applicable and flexible social app software for mobile teaching. Biochem Mol Biol Educ 2018;46(5):555-560 [FREE Full text] [doi: 10.1002/bmb.21170] [Medline: 30369037]
12. Li H, Wang XT. Cyber-personality and expression: a study from WeChat users in China. Front Psychol 2021;12:626040 [FREE Full text] [doi: 10.3389/fpsyg.2021.626040] [Medline: 34305702]
13. Cao Y, Lin SH, Zhu D, Xu F, Chen ZH, Shen HH, et al. WeChat public account use improves clinical control of cough-variant asthma: a randomized controlled trial. Med Sci Monit 2018;24:1524-1532 [FREE Full text] [doi: 10.12659/msm.907284] [Medline: 29536984]
14. Zhang X, Xiao H. Development and evaluation of a WeChat-based life review programme for patients with cancer: protocol for a randomised controlled trial. BMJ Open 2018;8(12):e020239 [FREE Full text] [doi: 10.1136/bmjopen-2017-020239] [Medline: 30530470]
15. Zhang W, Deng Z, Evans R, Xiang F, Ye Q, Zeng R. Social media landscape of the tertiary referral hospitals in China: observational descriptive study. J Med Internet Res 2018;20(8):e249 [FREE Full text] [doi: 10.2196/jmir.9607] [Medline: 30093370]
16. Cao B, Liu C, Durvasula M, Tang W, Pan S, Saffer AJ, et al. Social media engagement and HIV testing among men who have sex with men in China: a nationwide cross-sectional survey. J Med Internet Res 2017;19(7):e251 [FREE Full text] [doi: 10.2196/jmir.7251] [Medline: 28724510]
17. 2019 China WeChat 500 Annual List. WeChat. URL: https://mp.weixin.qq.com/s/9Co6wtDlxk0_I5KU4Rue1Q [accessed 2020-01-14]
18. New Rank platform. 2020. URL: https://www.newrank.cn/ [accessed 2024-12-16]
19. Rivera YM, Moran MB, Thrul J, Joshu C, Smith KC. Contextualizing engagement with health information on Facebook: using the social media content and context elicitation method. J Med Internet Res 2022;24(3):e25243 [FREE Full text] [doi: 10.2196/25243] [Medline: 35254266]
20. Yin M, Chen S, Pan X, Lu C, Lin X, Wang M, et al. Effects of Chinese provincial CDCs WeChat official account article features on user engagement during the COVID-19 pandemic. J Glob Health 2023;13:06005 [FREE Full text] [doi: 10.7189/jogh.13.06005] [Medline: 37054436]
21. Yin X, Pan J, Xu F. Effects of Wuxi CDC WeChat official account article features on user engagement in health promotion. BMC Public Health 2024;24(1):756. [doi: 10.1186/s12889-024-18277-4] [Medline: 38468225]
22. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in China: a survey study. BMC Med Inform Decis Mak 2017;17(2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
23. Gligorić K, Lifchits G, West R, Anderson A. Linguistic effects on news headline success: Evidence from thousands of online field experiments (Registered Report). PLoS One 2023;18(3):e0281682. [doi: 10.1371/journal.pone.0281682] [Medline: 36961788]
24. Question-based headlines: how to engage readers. Azura Magazine. URL: https://azuramagazine.com/articles/question-based-headlines-how-to-engage-readers [accessed 2024-12-10]
25. What would make you more likely to read this story? ScienceDaily. URL: https://www.sciencedaily.com/releases/2014/01/140108112614.htm [accessed 2024-12-10]
26. Tetteh EK, Combs T, Geng EH, McKay VR. Public health information seeking, trust, and COVID-19 prevention behaviors: cross-sectional study. J Med Internet Res 2022;24(9):e37846 [FREE Full text] [doi: 10.2196/37846] [Medline: 36084197]
27. Zhang Y, Xia T, Huang L, Yin M, Sun M, Huang J, et al. Factors influencing user engagement of health information disseminated by Chinese provincial centers for disease control and prevention on WeChat: observational study. JMIR Mhealth Uhealth 2019;7(6):e12245 [FREE Full text] [doi: 10.2196/12245] [Medline: 31250833]
28. Ma R, Deng Z, Wu M. Effects of health information dissemination on user follows and likes during COVID-19 outbreak in China: data and content analysis. Int J Environ Res Public Health 2020;17(14):5081 [FREE Full text] [doi: 10.3390/ijerph17145081] [Medline: 32674510]
29. Fan Z, Yin W, Zhang H, Wang D, Fan C, Chen Z, et al. COVID-19 information dissemination using the WeChat communication index: retrospective analysis study. J Med Internet Res 2021;23(7):e28563 [FREE Full text] [doi: 10.2196/28563] [Medline: 34129515]

Abbreviations

OR: odds ratio
VIF: variance inflation factor
WOA: WeChat official account

XSL•FO
RenderX

Edited by A Mavragani; submitted 13.08.24; peer-reviewed by J Engberg, D Dinh; comments to author 29.10.24; revised version received 17.11.24; accepted 27.11.24; published 20.01.25.

Please cite as:
Liao J, Huang X, Huang H, Shen C, Li L, Li Y, Zhan Y
Analysis of "Dr Ding Xiang" on WeChat in China to Determine Factors Influencing Readership on Medical Social Media: Observational Study
J Med Internet Res 2025;27:e65372
URL: https://www.jmir.org/2025/1/e65372
doi:10.2196/65372
PMID:

©Jiaman Liao, Xueliang Huang, Hao Huang, Cuina Shen, Lixia Li, Yushao Li, Yiqiang Zhan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Organ Donation Conversations on X and Development of the OrgReach Social Media Marketing Strategy: Social Network Analysis

Wasim Ahmed[1], BA, MSc, PhD; Mariann Hardey[2], BA, MA, PhD; Josep Vidal-Alaball[3,4,5], MD, PhD

[1]Business School, University of Hull, Hull, United Kingdom

[2]Business School, Durham University, Durham, United Kingdom

[3]Unitat de Recerca i Innovació, Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central, Institut Català de la Salut, Manresa, Spain

[4]Intelligence for Primary Care Research Group, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Manresa, Spain

[5]Department of Medicine, Faculty of Medicine, University of Vic, Vic, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, PhD
Unitat de Recerca i Innovació, Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central, Institut Català de la Salut
Carrer de Soler i March, 6,
Manresa, 08242
Spain
Phone: 34 936930040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: The digital landscape has become a vital platform for public health discourse, particularly concerning important topics like organ donation. With a global rise in organ transplant needs, fostering public understanding and positive attitudes toward organ donation is critical. Social media platforms, such as X, contain conversations from the public, and key stakeholders maintain an active presence on the platform.

Objective: The goal is to develop insights into organ donation discussions on a popular social media platform (X) and understand the context in which users discussed organ donation advocacy. We investigate the influence of prominent profiles on X and meta-level accounts, including those seeking health information. We use credibility theory to explore the construction and impact of credibility within social media contexts in organ donation discussions.

Methods: Data were retrieved from X between October 2023 and May 2024, covering a 7-month period. The study was able to retrieve a dataset with 20,124 unique users and 33,830 posts. The posts were analyzed using social network analysis and qualitative thematic analysis. NodeXL Pro was used to retrieve and analyze the data, and a network visualization was created by drawing upon the Clauset-Newman-Moore cluster algorithm and the Harel-Koren Fast Multiscale layout algorithm.

Results: This analysis reveals an "elite tier" shaping the conversation, with themes reflecting existing societal sensitivities around organ donation. We demonstrate how prominent social media profiles act as information intermediaries, navigating the tension between open dialogue and negative perceptions. We use our findings, social credibility theory, and review of existing literature to develop the OrgReach Social Media Marketing Strategy for Organ Donation Awareness. The OrgReach strategy developed is based on 5 C's (Create, Connect, Collaborate, Correct, and Curate), 2 A's (Access and Analyse), and 3 R's (Recognize, Respond, and Reevaluate).

Conclusions: The study highlights the crucial role of analyzing social media data by drawing upon social networks and topic analysis to understand influence and network communication patterns. By doing so, the study proposes the OrgReach strategy that can feed into the marketing strategies for organ donation outreach and awareness.

(J Med Internet Res 2025;27:e59872) doi:10.2196/59872

KEYWORDS

organ donation; organ transplant; social media; health; social network analysis; marketing strategy; awareness; public health; health information; qualitative; thematic analysis; NodeXL Pro; algorithm; elite tier; digital health; United Kingdom; X

XSL•FO
RenderX

Introduction

Background

The current shortage of organ transplants necessitates fostering public understanding and positive attitudes toward organ donation. The critical need for organ donation in the United Kingdom is further highlighted by the staggering number of patients currently waiting for a transplant, which stands at 7489 at the time of writing [1]. Social media, as a crucial platform for public health discourse [2], can harbor topics like organ donation. While research has explored increasing donor rates, a gap exists in understanding how online information environments shaped by prominent voices influence public perceptions. This study addresses this gap by using social network analysis (SNA) to examine conversations about organ donation on X (formerly Twitter). Specifically, we investigate how prominent social media profiles and users seeking health information influence these conversations and the types and forms of advocacy.

While previous research has examined the link between public attitudes and external factors like health care environments [3-5], we shift the focus inward, investigating how individuals navigate and interact with organ donation information on the internet. This approach offers insights into donation practices and the development of commercial organizations in this space. Specifically, we use SNA on platform X to map connections and influence patterns among the most prominent user profiles during the initial stages of information-seeking related to organ donation. Our analysis focuses on three key areas: (1) identifying users occupying pivotal positions within the network, (2) examining the emergence and persistence of an "elite tier" as the network evolves, and (3) exploring thematic and topical trends across active user accounts and dedicated health organizations.

This study defines the "elite tier" as a group of highly influential users within the organ donation community on social media. These individuals are distinguished by their ability to shape information flow, disseminate knowledge, and mobilize social media discourse. We identify elite-tier members based on their influence on their network (as measured by their betweenness centrality). We use our findings to develop the OrgReach Social Media Marketing Strategy for Organ Donation Awareness.

Examining the Credibility of Organ Donation Information on Social Media

While credibility in social media contexts has received significant attention within social science, media and cultural studies, and internet health literacy research, a critical gap remains in understanding its dynamic nature. This gap specifically concerns how credibility evolves and shapes how individuals navigate and engage with health information on social media. For example, Alikunju and Sulochana [6] argue that authenticity and network visibility constitute the core elements of credible health information on the internet. We use credibility theory [7] to explore the construction and impact of credibility within social media contexts in organ donation discussions. Specifically, we are interested in understanding the frequency and occurrence of credible sources of information (ie, health care professionals or organ donation organizations) compared to less credible sources of information. We also aim to explore the impact of credible actors on network reach on social media platforms.

Related Work

Previous research has studied organ donor registration on Facebook [8] and how the platform allowed users to specify they were organ donors. The study found that the Facebook intervention dramatically increased Organ Donation registration within the United States. The study highlights the potential of social media platforms in increasing organ donation awareness on social media. Other studies have investigated the ethical challenges in the use of social media in relation to privacy and confidentiality, pointing to the fact that there are no national ethical guidelines in the United States in relation to social media use for fostering organ transplantation [9]. Other research has examined different approaches to promoting kidney donation using social media data [10], to examine the role of Twitter for organ donation awareness [11], and to identify opinion leaders on Weibo [12].

Although social media platforms have been around since the early 2000s, using PubMed, we find that the first papers specifically examining organ donation and social media (as reflected in their title) appeared to have been published in 2012. The first outputs were related to a large-scale study that examined how university students could use social media to generate awareness and support for organ donation [13] alongside a published case report [14]. Of the 17 papers PubMed located published between 2012 and 2024, the most appeared to be published in 2023 (n=4). Moreover, previous research has analyzed the change in attitude toward organ donation following an educational program [15] where teenagers disseminate content on social networks to raise societal awareness. However, to the best of our knowledge, we found no other studies that aimed to perform a topic identification-driven SNA of X data that developed a specific marketing strategy for organ donation awareness.

The overall aim of this study is to identify the topics discussed on X relating to organ donation, to identify popular content, and to develop a specific marketing strategy for organ donation awareness.

Methods

Key Assumptions

The "Methods" and "Results" sections of this study frequently mention X users within the network, which are our "nodes," a concept from SNA. When we refer to "edges," these are our connections between users and can include reposts, mentions, and replies to users. Additionally, the study uses "betweenness centrality" as a metric to gauge a user's influence within the network. This measures the extent to which a node influences the flow of information among other nodes. This study follows the methodological approach from previous research [16-18] applied to the topic of organ donation.

XSL•FO
RenderX

Stage 1: Data Retrieval

Data were retrieved from X mostly between October 2023 and May 2024, covering a 7-month period. NodeXL Pro was used to retrieve data using the search string "organ donation." This would retrieve any post that contained "organ donation" or who replied to, mentioned, reposted, or quoted in those posts. It is important to note that content was retrieved from X without limiting it to a specific country. However, platform demographics such as more US users will mean that our sample is likely to contain more content from English-speaking countries. This is important as there will be differences in responses across different countries and also because of differences between opt-in and opt-out systems.

Stage 2: Data Preparation

Once data were retrieved, there were 20,124 users within the network and a total of 33,830 posts. These data were processed in NodeXL Pro (version 1.0.1.530) for further analysis. Within the network visuals, the edge colors are based on edge weight values. The edge widths were based on edge weight values. The vertex sizes (ie, the users) were based on in-degree values.

Stage 3: Data Analysis

The SNA method was used to analyze the data conducted in NodeXL Pro. The graph's vertices were grouped by cluster using the Clauset-Newman-Moore cluster algorithm [19]. The graph was laid out using the Harel-Koren Fast Multiscale layout algorithm [20]. We used thematic analysis [21] to develop themes and subthemes. After examining the groups through SNA, it was found that discussions within each group are multifaceted and are not confined to a single topic per group, but a multitude of conversations are taking place. Therefore, an analysis was conducted on the most shared content within the network overall. The 300 most popular posts (identified by finding the posts most reshared) had a total reshare value of n=3,975,866 with a total of n=6,398,623 favorites. We used a thematic analysis [21] of the top 300 posts until thematic saturation occurred. This thematic analysis involved a team-based coding approach where each post was assigned codes/labels by one author and then discussed and refined by the wider team. The labeling and analysis took place within Excel (Microsoft Corp). This allowed us to identify the most popular themes of discussions across the groups.

Ethical Considerations

This study examined data only in aggregate and no individuals who are not already in the public domain are mentioned or named in this study. Moreover, this study does not quote or use any content from any individuals, and only publicly shared information on the X platform is captured and analyzed. Ethical approval for this study falls under the institutional review board of Newcastle University (26055/2022). During the editing process, we drew upon generative language models (GPT-4o) and Grammarly to improve readability and identify typographical errors of existing human-authored text with oversight from the authors, but these models were not used to generate any content of their own. This was particularly useful as supported an author who is dyslexic.

Results

SNA and Qualitative Interpretation

Figure 1 shows a visual overview of how users connect with each other on X when conversing about organ donation and shows the top 20 groups. The network visual is simply a representation of groups that have formed based on user interactions (resharing, reposting, and replying to users). The largest group in the network (group 1) is an isolated group where users post without mentioning other users. There is one large community cluster (group 2) followed by three medium-sized community clusters (groups 3-5). The remainder of the groups (groups 6 to 20) consist of smaller community clusters.

Figure 1. Social network visualization of the organ donation X network (October 2023 to May 2024).

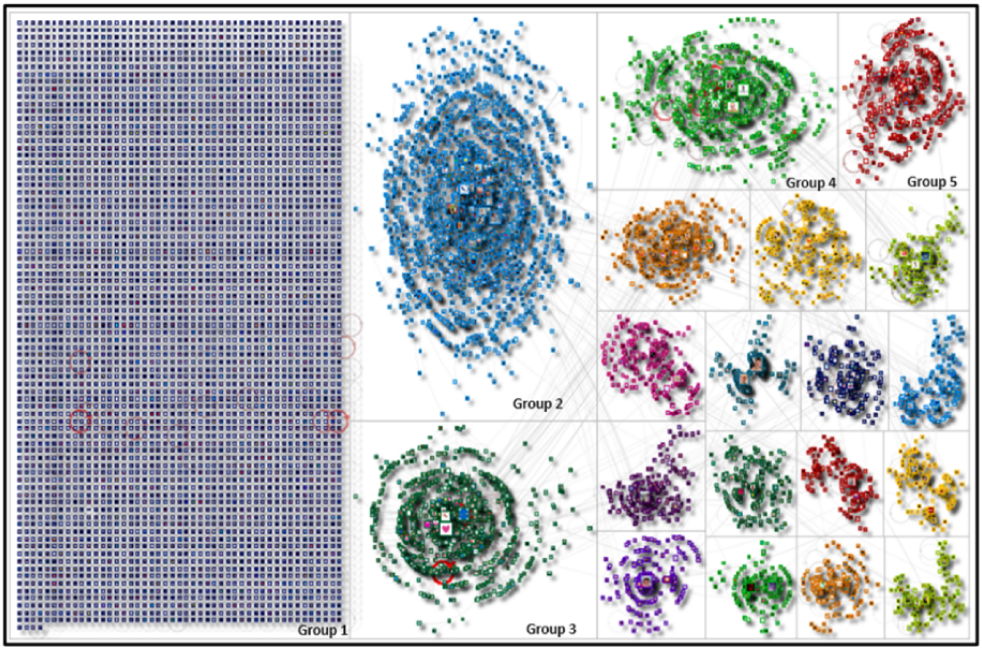

XSL•FO
RenderX

This analysis tells us that many individual posts are sent from users about organ donation (as noted from group 1), which include advocacy and stories. It also tells us that there are vocal conversations taking place by users who have formed their own communities and that some large, medium, and small communities on X are actively conversing about organ donation. The network visual also reveals that there is little interactivity between the groups.

The main themes and their subthemes from this analysis are presented below.

Theme 1: Ethical and Legislative Framework

Ethical Considerations and Debates

Users would discuss ethical concerns around organ donation, which would include mandatory donation and the use of human tissues in research.

Legislative and Policy Measures

Users would also converse about organ donation laws, transplantation regulations, and policies such as opt-out systems.

Controversy and Public Opinion

Users would also comment on controversial cases of organ harvesting and make claims that were not verifiable and potentially misinformation. Users would also hold debates and have critical discussions about organ donation.

Theme 2: Awareness, Advocacy, and Engagement

Awareness and Advocacy in Diverse Communities

Conversations by users included raising awareness of organ donation in minority communities and promoting global advocacy.

National and Cultural Engagement

Users would also discuss incorporating organ donation into national and cultural events to boost community involvement.

Impact of Public Figures and Influencers

Conversations also included the use of public figures and celebrities to advocate for organ donation and share their personal stories.

Theme 3: Recognition and Community Support

Recognition and Honor for Donors

Discussions included official recognition and public ceremonies for organ donors and their families.

Community Solidarity and Support

Within this subtheme, there were conversations and demonstrations of community support and consensus regarding organ donation efforts.

Theme 4: Personal Impact and Health Care Achievements

Inspirational Stories and Acts of Heroism

Within this subtheme, there were also personal accounts of donors who had made significant contributions by donating organs, and their stories were used to inspire others.

Health Care Achievements and Milestones

There were users, including health care professionals, who would celebrate key successes in organ donation and transplantation. These posts would also highlight positive recipient outcomes.

Additional Findings

Table 1 provides a ranking of the top 10 users posting on X within the network around organ donation. These users were defined as “elite tier” users as noted in the introduction due to their influential position within the network as measured by betweenness centrality.

The most influential user was the official NHS (National Health Service) organ donation account in the United Kingdom. Their user bio suggests active engagement with the public to improve and save lives through organ donation. They note that they respond to posts during specific hours, indicating a personalized approach to community interaction.

The second rank is held by the PMO (Prime Minister’s Office) India, which is the Office of the Prime Minister of India. This account’s high centrality score reflects its influence and broad reach in disseminating information related to organ donation within India. The account posted positive comments about organ donation, reaching a wide audience in India.

The third most influential user was Justin Trudeau, the 23rd Prime Minister of Canada. Trudeau’s user bio is presented in English and French, highlighting bilingual communication for broader outreach within Canada. Justin Trudeau has been a vocal supporter of organ donation on the X platform in the past, and the account appeared to be tagged into various conspiracies and potential mis and disinformation networks.

Ranks 4 through 10 were held by individual users worldwide, whom we categorized simply as “Citizens.” Some users had specific interests, such as a business-focused user and a heart transplant recipient. Of the users who disclosed their location, users derived from Canada, the United States, and the United Kingdom. This highlights that ordinary citizens with modest followings can hold significant positions within the network, contributing to discussions, spreading awareness, or advocating for organ donation.

Table 2 provides the top “domains” that users X posts were linking to. A domain analysis allows us to detect popular web pages. Table 2 highlights a wide-ranging effort to promote organ donation information across different types of websites, from government and health services to news media and social platforms. Link shorteners like bit.ly and ow.ly indicate a focus on social media sharing and tracking engagement with this content. The top-ranking domain, transplantnews.com, suggests a dedicated source of information that will likely serve as a hub

XSL•FO
RenderX

for those interested in the subject. These domains are further discussed below.

Discussing these, starting from the most popular web domain, we can see that the transplantnews.com domain is the most frequent linker to organ donation content (n=736). It is a specialized website focused on news related to organ transplants. The second rank is the Indian government's official domain, gov.in (n=311), which indicates a significant amount of content or references related to organ donation within the Indian government's online resources. In third place is dlvr.it (n=310), a content delivery network often used to manage and disseminate information across various platforms, suggesting widespread sharing of organ donation-related content.

Table 1. Top 10 influencers on X related to organ donation (October 2023 to May 2024).

Rank	Userhandle (anonymized if citizen)	Name	User bio	Location	Betweenness centrality
1	nhsorgandonor	NHS[a] Organ Donation	Saving and improving lives. Official account for the NHS Organ Donor Register. We love reading your tweets and reply to them between 0700-2300 #OrganDonation	United Kingdom	12713160.850
2	pmoindia	PMO[b] India	Office of the Prime Minister of India	India	9036848.857
3	justintrudeau	Justin Trudeau	Father, @liberal_party Leader, 23rd Prime Minister of Canada. | Papa, chef du @parti_liberal 23e premier ministre du Canada.	Canada	8986407.605
4	Citizen	Anonymized	Anonymized	Undisclosed	8105094.910
5	Citizen	Anonymized	Anonymized	Canada	7750991.313
6	Citizen	Anonymized	Anonymized	Undisclosed	7333731.145
7	Citizen Business Interest	Anonymized	Anonymized	United States	7301905.351
8	Citizen	Anonymized	Anonymized	Undisclosed	7215205.365
9	Citizen Political Influnfcner	Anonymized	Anonymized	United States	6379487.669
10	Citizen Heart Transplant Recipient	Anonomyized	Anonomyized	United Kingdom	6185548.502

[a]NHS: National Health Service.

[b]PMO: Prime Minister's Office.

Table 2. Top 10 domains on X related to organ donation (October 2023 to May 2024).

Rank	Top domains	Count, n
1	transplantnews.com	736
2	gov.in	311
3	dlvr.it	310
4	bit.ly	281
5	youtu.be	150
6	facebook.com	140
7	nm-4.com	128
8	ow.ly	109
9	nhs.uk	103
10	co.uk	98

In fourth place, we have bit.ly (n=281), a URL-shortening service. The relatively high count indicates that many shortened URLs linking to organ donation content have been created, highlighting a general trend of concisely disseminating this important information across different platforms. This practice likely enhances user accessibility and simplifies the process of sharing such links.

The fifth-ranked domain is youtu.be (n=150), the shorthand for YouTube signifies a moderate level of engagement with video content related to organ donation. This number reflects the efforts made to use multimedia platforms to raise awareness and educate the public about organ donation through videos, which can be a compelling way to convey the impact and importance of this act. Following that, facebook.com stands in

XSL•FO
RenderX

the sixth position (n=140). As a major social media platform, Facebook's role in spreading organ donation content is important. Although it ranks lower than some other domains, the figure indicates that organ donation discussions and information sharing are part of the social conversation on Facebook.

The domain nm-4.com is listed in seventh place (n=128), the short version of the full URL which points to an official website belonging to the Indian Prime Minister, this web page shared information about Narendra Modi's support toward organ donation. Ranking eighth is ow.ly (n=109), another URL-shortening service akin to bit.ly. Its use for organ donation content implies that there is an active effort to track engagement and reach of shared links on various social media and platforms on the internet. The ninth position is held by nhs.uk (n=103), the UK National Health Service's official domain. This domain's contribution highlights the UK health authority's commitment to educating and providing resources on organ donation to the public, reflecting the importance of institutional support in promoting health initiatives.

Finally, co.uk (n=98) encompasses a range of UK-based websites. This generic top-level domain's appearance in the top ten signifies that various UK entities collectively discuss and link to organ donation content. This points to a country-wide effort to endorse organ donation awareness and registration.

Table 3 provides an overview of the top 10 hashtags that users were using. In first place was the hashtag #organdonation (n=5413), which is the most widely used hashtag concerning organ donation discussions. The second most prevalent hashtag is #donatelife (n=1361). This hashtag resonates with the community's call to action for individuals to contribute to the cause of saving lives through organ donation. In third place is the hashtag #giftoflife (n=974). This hashtag emphasizes organ donation's life-saving and life-enhancing benefits, celebrating it as an altruistic gift a person can offer others. The fourth most popular hashtag was #organtransplant (n=867). This hashtag is likely used to share information, personal stories, and developments related to the organ transplant process.

In fifth place was #tissuetransplant (n=756), which points to discussions specific to the transplantation of tissues, another crucial aspect of the broader conversation on transplantation and organ donation. In sixth place was the hashtag #transplantnews (n=742), suggesting a focus on sharing news and updates related to transplant medicine and stories of individuals involved in the transplant process.

The seventh-ranked hashtag is #organtissue (n=478), and it brings to light the conversation about both organ and tissue donation, indicating a combined discourse that highlights the need for donations of all types of body tissues. The hashtag #kidneytransplant (n=359) takes the eighth spot, denoting a specific focus within the organ donation community on kidney transplants, reflecting the high demand and frequency of this type of transplant. The hashtag #transplant (n=339) was in ninth place. This more general hashtag indicates a broad discussion encompassing the entire spectrum of transplant-related topics. Finally, the hashtag #savelives (n=315) occupied the tenth position. This compelling and impactful call to action highlights the ultimate goal of organ donation: to save and improve lives.

Table 3. Top 10 hashtags on X related to organ donation (October 2023 to May 2024).

Rank	Top hashtags	Count, n
1	organdonation	5413
2	donatelife	1361
3	giftoflife	974
4	organtransplant	867
5	tissuetransplant	756
6	transplantnews	742
7	organtissue	478
8	kidneytransplant	359
9	transplant	339
10	savelives	315

Development of the OrgReach Marketing Strategy

Overview

We draw upon our results, our review of existing research [8-12], and social credibility theory [7] to develop the OrgReach Marketing Strategy. Specifically, we draw upon our results around:

- The nature of the network and the lack of cross-group conversations that were found in our SNA.
- Our insights into the opinion leaders within the network and their characteristics.
- The characteristics of the most popular posts that were sent in terms of retweets and favorites and the types of content that attracted reposts.
- The most popular web domains within the network in terms of link clicks and the types of studies that they were sharing.

We conceptualize the 5 C's (Create, Connect, Collaborate, Correct, and Curate), 2 A's (Access and Analyse), and 3 R's (Recognize, Respond, and Re-evaluate), which we name the

XSL•FO
RenderX

OrgReach Marketing Strategy. Each element is further outlined below.

The 5 C's

Create

Develop engaging and educational content related to your topics of interest. Emotional content, including images or video-based testimonials highlighting organ donation's need and impact on lives, may gain more visibility. To assess the appeal and educational effectiveness of content, organizations should continuously monitor engagement metrics (likes, shares, comments, and reach) to determine the popularity and engagement with content. Moreover, direct qualitative feedback from users and A/B testing of content could help provide further evaluation of content.

Connect

To build a strong community and foster conversations and connections between individuals, health care professionals, and advocacy groups. Using the search feature to find similar organizations and conversations and break down conversational silos on social media. To assess and evaluate whether organizations are succeeding in their network building, they can determine their social network metrics (network density, clustering, and cross-group interactions). Moreover, tracking follower and advocacy accounts, as well as user surveys, could help evaluate the success of connection-building.

Collaborate

Partner with influencers, health care organizations, and governmental bodies to collaborate on content creation, amplify the message, and reach a wider audience. Codevelop and share content and incorporate other "@" tags within posts. More credible influencers will likely lead to further impact. Organizations can select partners such as health care organization and governmental bodies based on the alignment of values and for influencers, their reach, and audience. More sophisticated analyses, such as engagement quality, influencer identification, and credibility assessments, can also be drawn upon in this phase to identify the quality and consistency of collaborative content.

Correct

Monitor and address any misinformation users share, ensuring the public receives accurate and trustworthy information about organ donation. Corrected information may lead to your organization's further visibility. Organizations can draw upon text-analysis tools, such as NodeXL Pro, introduced earlier, and this could help identify high-frequency keywords that can highlight the presence of misinformation. This analysis can be supplemented with sentiment and web-domain level analysis where the credibility of websites is assessed. Moreover, organizations can look to existing fact-checking resources and encourage community reporting of misinformation.

Curate

In addition to original content, organizations can handpick and share relevant stories, news, and data relevant to organ donation to keep the audience informed and engaged. Organizations can set up news monitoring on platforms such as Google News for relevant, credible content that can include stories that resonate emotionally with the target audience. Organizations should ensure content across diverse perspectives, including minority communities and different cultural approaches to organ donation. Organizations can also incorporate feedback loops, which provide social media followers to provide feedback on content that is shared.

The 2 A's

Access

Ensure that the information you share is accessible to everyone, removing barriers, that is, a public-facing social media account with an easy-to-find user handle. This further enhances credibility.

Analyze

Use SNA and credible influencer detection to understand audience behavior and refine strategies for better engagement and impact. Using this analysis and output of key users and content to feed back into the 5 C's.

The 3 R's

Recognize

Publicly acknowledge and celebrate the altruism of donors and the success stories of recipients to inspire others.

Respond

Establish a reliable and responsive communication channel through social media to answer queries and engage with the audience in real time.

Reevaluate

Continuously assess the strategy's effectiveness and make data-driven adjustments to improve outreach and advocacy efforts.

Discussion

Principal Findings

Previous research has studied the potential of social media platforms for increasing donor registrations and strategies for sharing organ donation-related content. This study contributes to the literature by providing insights into users' conversations around organ donation, identifying influencers, and uncovering the network structure of the elite tier and key URLs and hashtags. These findings were then used to develop the OrgReach Marketing Strategy to offer practical suggestions and strategies for organizations in this area. Our findings and their broader implications will be further discussed.

In our analysis, the structure of user interactions on X through SNA provided insights into the shape of the discourse surrounding organ donation. The largest group was where users would send isolated posts, suggesting that individuals share personal stories and advocacy messages independently. Large and smaller interactive community clusters indicated that more engaged discussions were also occurring, which are important for fostering support among users considering organ donation. The limited interactivity between the groups suggests the

XSL•FO
RenderX

potential for enhancing dialogue across different clusters to unite advocacy efforts and enhance the conversation around organ donation. By understanding these interactions, stakeholders could better tailor their communication strategies to increase awareness and participation in organ donation. This would allow stakeholders to maximize the role of social media as a tool for impactful advocacy.

Our analysis also revealed that a wide array of topics related to organ donation were discussed. These ranged from ethical debates and legislative discussions to advocacy efforts and personal impact stories. This study found that users engage in conversations about the ethical implications of mandatory donation, the use of tissues in research, and the effects of laws and opt-out systems. Controversial topics often include misinformation, which highlights the need for accurate communication. Advocacy through public figures and cultural events boosted awareness within the network, especially in minority communities. The recognition of donors and community support on X emphasized organ donation's societal value and inspirational potential.

The NHS Organ Donation account, a significant member of the elite tier, actively engages with the public through structured interaction hours, underscoring its commitment to saving lives. This highlights a strategic approach to online health communication, which may, in part, have led to the greater impact of the account. Influential leaders such as the PMO India and Canada's Prime Minister Justin Trudeau show how political figures can leverage their platforms to significantly impact public discourse on health issues. However, political figures can also be targeted for misinformation, as was the case with Trudeau's account. Moreover, the presence of anonymized citizens and public figures from diverse locations like the United Kingdom, Canada, and the United States among the top influencers indicates that organ donation discussions attract a wide range of contributors, from individuals sharing personal experiences to those with business or political interests. This provides a variety of viewings being shared.

The analysis of the top domains linked in these discussions revealed a strategic use of diverse online resources to promote organ donation. These included specialized news outlets such as transplantnews.com, as well as governmental and social media platforms. The variety in content distribution channels highlights the concerted efforts to maximize outreach and engage different audiences effectively. URL shorteners and content delivery networks highlight a sophisticated approach to content dissemination, ensuring that information reaches a broad audience efficiently and is easily shareable.

The analysis and identification of popular web domains, hashtags, and the development of the OrgReach Marketing Strategy aim to contribute to existing literature and provide practical suggestions and strategies for social media engagement and outreach.

Limitations and Future Research Agenda

This study focused on retrieving data from X (formerly Twitter). Therefore, our empirical findings would not capture discussions on other platforms like Facebook, Instagram, or specific Web forums. Future research could examine content across various platforms to provide a critical comparative perspective. Moreover, this study examines a single time slice and captures conversations across a specific 7-month period, and content may differ across other periods. Future research could retrieve and compare different periods to identify changes in influencers and conversations. Future research could also seek to test and validate the proposed OrgReach Social Media Marketing Strategy in real-world campaigns which would help to strengthen the practical applicability of the findings.

Moreover, it is important to note that this study examined content aggregately and did not stratify different countries. This is because certain countries operate under an "opt-in" to organ donation system, such as the United States, where campaigns typically emphasize individual registration. In opt-out countries, campaigns focus on promoting organ donation's altruistic and societal benefits to encourage families to make informed decisions. Based on these differing contexts, the study does not differentiate social media content or user interactions. The OrgReach Marketing Strategy could also be enhanced by incorporating specific guidance for campaigns tailored to these legislative frameworks.

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. NHS organ donation. 2024. URL: https://www.organdonation.nhs.uk/helping-you-to-decide/about-organ-donation/statistics-about-organ-donation/ [accessed 2024-04-18]
2. Ahmed W, Aiyenitaju O, Chadwick S, Hardey M, Fenton A. The influence of Joe Wicks on physical activity during the COVID-19 pandemic: thematic, location, and social network analysis of X data. J Med Internet Res 2024;26:e49921 [FREE Full text] [doi: 10.2196/49921] [Medline: 38551627]
3. Siminoff LA, Gordon N, Hewlett J, Arnold RM. Factors influencing families' consent for donation of solid organs for transplantation. JAMA 2001;286(1):71-77. [doi: 10.1001/jama.286.1.71] [Medline: 11434829]

XSL•FO
RenderX

4. He X, Huang J. Organ Donation and Transplantation After Cardiac Death in China: Clinical Practice. Germany: Springer Nature; 2022.
5. Carola V, Morale C, Vincenzo C, Cecchi V, Errico L, Nicolais G. Organ donation: psychosocial factors of the decision-making process. Front Psychol 2023;14:1111328 [FREE Full text] [doi: 10.3389/fpsyg.2023.1111328] [Medline: 37303924]
6. Alikunju H, Sulochana A. Transgender health information: assessing the quality of online resources. OIR 2024;48(5):926-943. [doi: 10.1108/oir-09-2023-0469]
7. Sobel J. A theory of credibility. Rev Econ Stud 1985;52(4):557-573. [doi: 10.2307/2297732]
8. Cameron AM, Massie AB, Alexander CE, Stewart B, Montgomery RA, Benavides NR, et al. Social media and organ donor registration: the Facebook effect. Am J Transplant 2013;13(8):2059-2065 [FREE Full text] [doi: 10.1111/ajt.12312] [Medline: 23777475]
9. Henderson ML, Clayville KA, Fisher JS, Kuntz KK, Mysel H, Purnell TS, et al. Social media and organ donation: ethically navigating the next frontier. Am J Transplant 2017;17(11):2803-2809 [FREE Full text] [doi: 10.1111/ajt.14444] [Medline: 28744966]
10. Moorlock G, Draper H. Empathy, social media, and directed altruistic living organ donation. Bioethics 2018;32(5):289-297 [FREE Full text] [doi: 10.1111/bioe.12438] [Medline: 29542172]
11. Pacheco D, Pinheiro D, Cadeiras M, Menezes R. Characterizing organ donation awareness from social media. 2017 Presented at: IEEE 33rd International Conference on Data Engineering (ICDE); USA; April 19-22, 2017. [doi: 10.1109/icde.2017.225]
12. Shi J, Salmon CT. Identifying opinion leaders to promote organ donation on social media: network study. J Med Internet Res 2018;20(1):e7 [FREE Full text] [doi: 10.2196/jmir.7643] [Medline: 29317384]
13. D'Alessandro AM, Peltier JW, Dahl AJ. A large-scale qualitative study of the potential use of social media by university students to increase awareness and support for organ donation. Prog Transplant 2012;22(2):183-191. [doi: 10.7182/pit2012619] [Medline: 22878076]
14. D'Alessandro AM, Peltier JW, Dahl AJ. Use of social media and college student organizations to increase support for organ donation and advocacy: a case report. Prog Transplant 2012;22(4):436-441. [doi: 10.7182/pit2012920] [Medline: 23187063]
15. Febrero B, Almela-Baeza J, Ros I, Pérez-Sánchez MB, Pérez-Manzano A, Cascales P, et al. The impact of information and communications technology and broadcasting on YouTube for improving attitude toward organ donation in secondary education with the creation of short films. Patient Educ Couns 2021;104(9):2317-2326. [doi: 10.1016/j.pec.2021.02.037] [Medline: 33658142]
16. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
17. Ahmed W, Seguí FLL, Vidal-Alaball J, Katz MS. COVID-19 and the "Film Your Hospital" conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020;22(10):e22374 [FREE Full text] [doi: 10.2196/22374] [Medline: 32936771]
18. Ahmed W, Das R, Vidal-Alaball J, Hardey M, Fuster-Casanovas A. Twitter's role in combating the magnetic vaccine conspiracy theory: social network analysis of tweets. J Med Internet Res 2023;25:e43497 [FREE Full text] [doi: 10.2196/43497] [Medline: 36927550]
19. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. Phys Rev E 2004;70(6 Pt 2):066111. [doi: 10.1103/PhysRevE.70.066111] [Medline: 15697438]
20. Koren Y, Harel D. Axis-by-Axis Stress Minimization. Berlin Heidelberg: Graph Drawing Springer; 2004:459.
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

Abbreviations

NHS: National Health Service
PMO: Prime Minister's Office
SNA: social network analysis

Edited by A Mavragani; submitted 07.05.24; peer-reviewed by V Astha, Y Zhang, J Almela, E Sangiorgio; comments to author 24.11.24; revised version received 10.12.24; accepted 09.01.25; published 06.02.25.

Please cite as:
Ahmed W, Hardey M, Vidal-Alaball J
Organ Donation Conversations on X and Development of the OrgReach Social Media Marketing Strategy: Social Network Analysis
J Med Internet Res 2025;27:e59872
URL: https://www.jmir.org/2025/1/e59872
doi:10.2196/59872
PMID:

XSL•FO
RenderX

©Wasim Ahmed, Mariann Hardey, Josep Vidal-Alaball. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Human Papillomavirus Vaccination Hesitancy in Japan Using Social Media: Content Analysis

Junyu Liu[1*], PhD; Qian Niu[1*], PhD; Momoko Nagai-Tanima[1], PhD; Tomoki Aoyama[1], MD, PhD

Kyoto University, Kyoto, Japan

[*]these authors contributed equally

Corresponding Author:
Tomoki Aoyama, MD, PhD
Kyoto University
Yoshida-honmachi, Sakyo-Ku
Kyoto, 606-8501
Japan
Phone: 81 075 751 3952
Email: aoyama.tomoki.4e@kyoto-u.ac.jp

Abstract

Background: Despite the reinstatement of proactive human papillomavirus (HPV) vaccine recommendations in 2022, Japan continues to face persistently low HPV vaccination rates, which pose significant public health challenges. Misinformation, complacency, and accessibility issues have been identified as key factors undermining vaccine uptake.

Objective: This study aims to examine the evolution of public attitudes toward HPV vaccination in Japan by analyzing social media content. Specifically, we investigate the role of misinformation, public health events, and cross-vaccine attitudes (eg, COVID-19 vaccines) in shaping vaccine hesitancy over time.

Methods: We collected tweets related to the HPV vaccine from 2011 to 2021. Natural language processing techniques and large language models (LLMs) were used for stance analysis of the collected data. Time series analysis and latent Dirichlet allocation topic modeling were used to identify shifts in public sentiment and topic trends over the decade. Misinformation within opposed-stance tweets was detected using LLMs. Furthermore, we analyzed the relationship between attitudes toward HPV and COVID-19 vaccines through logic analysis.

Results: Among the tested models, Gemini 1.0 pro (Google) achieved the highest accuracy (0.902) for stance analysis, improving to 0.968 with hyperparameter tuning. Time series analysis identified significant shifts in public stance in 2013, 2016, and 2020, corresponding to key public health events and policy changes. Topic modeling revealed that discussions around vaccine safety peaked in 2015 before declining, while topics concerning vaccine effectiveness exhibited an opposite trend. Misinformation in topic "Scientific Warnings and Public Health Risk" in the sopposed-stance tweets reached a peak of 2.84% (47/1656) in 2012 and stabilized at approximately 0.5% from 2014 onward. The volume of tweets using HPV vaccine experiences to argue stances on COVID-19 vaccines was significantly higher than the reverse.

Conclusions: Based on observation on the public attitudes toward HPV vaccination from social media contents over 10 years, our findings highlight the need for targeted public health interventions to address vaccine hesitancy in Japan. Although vaccine confidence has increased slowly, sustained efforts are necessary to ensure long-term improvements. Addressing misinformation, reducing complacency, and enhancing vaccine accessibility are key strategies for improving vaccine uptake. Some evidence suggests that confidence in one vaccine may positively influence perceptions of other vaccines. This study also demonstrated the use of LLMs in providing a comprehensive understanding of public health attitudes. Future public health strategies can benefit from these insights by designing effective interventions to boost vaccine confidence and uptake.

(J Med Internet Res 2025;27:e68881) doi:10.2196/68881

KEYWORDS

human papillomavirus; HPV; HPV vaccine; vaccine confidence; large language model; stance analysis; topic modeling

XSL•FO
RenderX

Introduction

Cervical cancer remains a significant public health challenge worldwide. In 2020, the World Health Organization (WHO) launched a strategy to eliminate cervical cancer, aiming to achieve 90% coverage rate for human papillomavirus (HPV) vaccination by 2030 [1]. In 2023, the WHO released estimates of first-dose vaccination coverage for women in 133 countries, with a global average of 62%, and 15 countries have already met the WHO's target [2]. However, Japan presents a unique case among high-income countries, where HPV vaccination rates have remained critically low. Both bivalent and quadrivalent HPV vaccines were licensed and subsequently included in the national immunization program for girls in 2013 in Japan [3,4]. However, widespread reports of adverse events prompted the Japanese Ministry of Health, Labour and Welfare (MHLW) to suspend the proactive recommendation in 2013, leading to a sharp decline in public trust and a significant increase in cervical cancer incidence [5-13]. While proactive recommendations were reinstated in 2022, HPV vaccine uptake remains low [14-18].

To enhance HPV vaccination uptake in Japan, it is crucial to systematically analyze the trends of vaccine hesitancy. Existing literature reveals varying levels of hesitancy across different populations. Studies targeting parents consistently demonstrated significant hesitancy in the years preceding the updated recommendations, whereas surveys involving health care professionals, including those who are parents, showed higher confidence. A 2024 MHLW survey found comparable proportions of willingness, unwillingness, and uncertainty among both the target population and their parents. However, a critical gap in current research is the lack of longitudinal studies to track public attitudes toward HPV vaccination over time using consistent methodologies.

Understanding the factors influencing vaccine hesitancy is also essential. Studies have attributed this to persistent public distrust, misinformation, negative media coverage [15-20], and prolonged government suspension of recommendations, which heightened public concern [11]. The WHO Global Advisory Committee on Vaccine Safety emphasized the potential harm of policy decisions based on weak evidence, leading to the underutilization of safe and effective vaccines [21]. However, limited research has comprehensively investigated the long-term, dynamic influence of the key events and factors on HPV vaccine hesitancy in Japan.

In recent years, social media platforms have emerged as valuable resources for public health research. Comparing with the traditional survey research, social media can provide large volumes of real-time data that are readily accessible and searchable [22]. Among all the social media platforms, the user base of Twitter (subsequently rebranded as X) in Japan has grown significantly, with its share of users increasing from slightly more than 20% of the population in 2014 to approximately 45% in 2021 [23]. The platform's primary active user demographic, aged 13-39 years, substantially overlaps with the target age groups of vaccine-eligible adolescents and their parents, making it a relevant data source for analyzing public discourse on the HPV vaccine [24]. Twitter analysis has been applied to investigate public discourse surrounding HPV vaccination in various countries [25-28]. In Japan, Suzuki et al [29] analyzed 208 geotagged tweets, revealing an uneven distribution of users disseminating information related to HPV vaccination. Terada et al [19] examining 3623 tweets collected over a 7-month period, reported a predominance of negative sentiment associated with discussions pertaining to HPV vaccine safety and side effects. However, few works used large-scale social media analysis to comprehensively examine the evolution of public discourse concerning HPV vaccination over an extended period within the Japanese context.

Recent advancements in natural language processing (NLP) methodologies, particularly the integration of deep learning (DL) models, have significantly enhanced the accuracy and sophistication of social media analysis [30]. While traditional NLP techniques such as n-grams [31] and topic modeling [32] have proven valuable, DL models, especially the large language models (LLMs), have demonstrated significant potential for conducting sophisticated and nuanced analyses of social media data [33]. However, the application of LLMs to investigate HPV vaccine hesitancy in Japan remains an underexplored area of research.

This study aims to address these gaps by analyzing the long-term evolution of public attitudes toward HPV vaccination in Japan from 2011 to 2021 through Twitter data. Specifically, we use advanced NLP techniques, including stance analysis, time series analysis, topic modeling, and logic analysis, to:

1. Examine the dynamics of HPV vaccine confidence over time;
2. Identify changes in key factors influencing vaccine hesitancy; and
3. Explore the relationship between key events and shifts in public opinion, including interaction between vaccines.

By leveraging state-of-the-art NLP and LLM methods, this study provides a systematic, data-driven understanding of vaccine hesitancy in Japan. The findings will offer insights into the evolution of public perceptions and inform strategies to address vaccine hesitancy more effectively.

Methods

Data Collection and Preprocessing

This study used the Twitter application programming interface to collect Japanese tweets related to HPV vaccination between 2011 and 2021. The tweets were retrieved using various Japanese keywords for HPV vaccine. The number of tweets increased each year, resulting in 228,376 tweets. After excluding tweets with unrecognizable dates, 228,300 were retained for subsequent analyses. After data collection, the tweets were cleaned and preprocessed. Retweets were removed using the Python package *Tweepy* (Twitter) [34]. Web links, special characters, emojis, and ampersands were eliminated, and full-width English characters were converted into half-width lowercase characters.

XSL•FO
RenderX

Annotation

To facilitate model training and fine-tuning, 2.5% of the tweets per year were randomly selected for annotation. This study focused on public stance toward the HPV vaccine by using a 3-category annotation system: advocate, oppose, and unknown. The detailed category definitions are shown in Table 1.

Table 1. Definitions of stances to the human papillomavirus vaccine.

Category	Definition	Mock tweet example
Advocate	Affirmatively advocacy of the HPV[a] vaccine, highlighting its efficacy in preventing HPV-related diseases, the critical role of vaccination in public health, and positive vaccination outcomes. These messages distinctly advocate for the HPV vaccine's utilization, reinforcing its importance in health prevention measures.	Just got my HPV vaccine today! Feeling grateful for the science that protects us from cervical cancer. Highly recommend it to everyone eligible!
Opposite	Explicitly challenge or critique the HPV vaccine, focusing on concerns related to its efficacy, safety, or potential adverse effects. The critical perspective must be distinctly directed at the HPV vaccine in isolation, excluding broader criticisms of vaccines at large, political deliberations concerning vaccination policies, or financial considerations associated with vaccine procurement.	I'm skeptical about the HPV vaccine due to the side effects I've read about. More research is needed before making it a widespread recommendation. Let's be cautious.
Unknown	Neither explicitly endorse nor directly oppose the HPV vaccine. This category encapsulates tweets that voice opposition stemming from broad antivaxxers, policies, or concerns about the cost, without specifically expressing the advocate or opposite of the HPV vaccine itself. It also includes tweets that are too ambiguous, neutral, or irrelevant to ascertain a clear position concerning the HPV vaccine.	Why isn't the HPV vaccine free for everyone? If it's so important, shouldn't the government cover the cost? Still trying to understand the policy behind this.

[a]HPV: human papillomavirus.

The annotation team comprised 4 medical professionals, 3 of whom independently annotated each tweet. Tweets with unanimous annotations were classified as tier 1, those with 2 matching annotations as tier 2, and those without matching annotations as tier 3. Tier 1 data, which were deemed to have the highest consistency, used the initial annotations as the final labels. Tier 2 and 3 data underwent further consensus meetings, with the final labels determined by the last medical professional.

A total of 5758 tweets were annotated, with 5050 in tier 1, a total of 701 in tier 2, and 7 in tier 3. The Cohen κ values between each pair of the 3 major medical professionals (KG, YN, and ND) are shown in Table 2, which are all over 0.8 and at the agreement level of almost perfect. The distributions across the advocate, oppose, and unknown categories in the annotated data were 1646, 697, and 3415, respectively.

Table 2. Cohen κ between the annotations by each pair of medical professionals.

	Cohen κ
KG and YN	0.851
YN and ND	0.845
KG and ND	0.858

Stance Analysis With DL Models

Model Selection

To achieve the optimal results, we first selected the best-performing model. We compared several traditional NLP classification models with LLMs. Among the traditional models, we considered long short-term memory [35], Bidirectional Encoder Representations from Transformers (BERT) [36], and DistilBERT [37]. For the LLMs, we used Gemma-2-2b [38], Gemma-2-9b, and Llama-3.1-8b [39]. All uses of open-source LLMs in this work were conducted following approval and in strict accordance with their respective licensing agreements. We also evaluated the Gemini Pro 1.0 [40] using Google AI Studio during the free trial.

A 2-layer bidirectional architecture is designed for the long short-term memory model. The tweets were tokenized using the *Ginza* Python package (Megagon Labs) [41]. Owing to the manageable parameter size of this model, it was trained from scratch. For the BERT models, Tohoku BERT [42] and Line DistilBERT [43] were selected as baselines. A classification head was added to the first token of the final hidden layer of these models, and fine-tuning was performed using our tweet data, while keeping the BERT parameters frozen.

For LLMs, computational resource constraints prevented us from fine-tuning models with more than 10 billion parameters, whereas budgetary constraints precluded fine-tuning using application programming interfaces such as ChatGPT [44] or Claude [45]. Therefore, we opted for smaller open-source LLMs and tested both baseline, Quantization, and Low-Rank Adaptation fine-tuned versions on the tweet data [46]. For Gemini Pro 1.0, we fine-tuned the model directly using Google AI Studio. For all LLMs, we used the following prompt in Japanese, followed by the original tweet, as user prompts:

Does this tweet express a position against the HPV vaccine itself? Please classify this tweet as "advocate," "oppose," or "unknown." Do not make subjective assumptions; base your

XSL•FO
RenderX

decision solely on the content of the tweet. Answer only your position, and no explanation is required.

Given the significant class imbalance in the dataset, 5758 annotated tweets were split into training, validation, and test sets, with the proportion of each category maintained across the splits. To ensure a robust evaluation, 100 tweets from each category were randomly selected to form a test set. The remaining data within each category were partitioned into training and validation sets at a ratio of 4:1. Due to the training sample size limitation (<500) imposed by Gemini 1.0 Pro, a subset of 498 labeled tweets (distributed equally across the 3 categories at a ratio of 166:166:166) was used to fine-tune the model for the HPV vaccine stance classification task. The model performance was assessed using standard evaluation metrics, namely, precision, recall, and F_1-score. The details of the labeled dataset are listed in Table 3.

Table 3. Number of labeled tweets of different stances in the dataset.

Category	Train set, n	Validation set, n	Test set, n	Total, n
Advocate	2652	663	100	3415
Oppose	477	120	100	697
Unknown	1236	310	100	1646
Total	4365	1093	300	5758

To investigate the performance boundaries of the optimal model further, we explored the impact of varying category distributions within the training data. We experimented with different ratios of advocate, opposition, and unknown categories, specifically 166:166:166, 150:200:150, 125:250:125, 100:300:100, 75:350:75, and 50:400:50. For each ratio, the data were randomly sampled 3 times, the model performance was evaluated, and the results were averaged to ensure robustness.

After identifying the optimal category ratio, we fine-tuned the best-performing model 5 times using this ratio, each time with a new random data sample. The remaining labeled data were used for the performance evaluation, and the 3 models exhibiting the highest performance were selected for the final inference. For each unlabeled tweet, the 3 models generated independent predictions. If 2 or more models agreed with the predicted label, they were assigned to the tweet. When all 3 models produced different predictions, the tweet was classified as “unknown.”

Time Series Analysis of Stances

Following the stance classification of all tweets, a comprehensive time series analysis was performed on the extracted data to examine the evolution of public sentiments toward HPV vaccines in Japan between 2011 and 2021. The HPV vaccine supply and injection data were collected from the official website of MHLW. The raw data for vaccine supply lack fine-grained annual data and can be across 2 years, so the annual supply for some years were based on estimation; details are provided in Multimedia Appendix 1. We calculated the annual advocate rate from the tweets by *(number of advocate tweets)/(number of advocate tweets + number of opposed tweets)*. The difference between the monthly counts of tweets advocating for and opposing the HPV vaccine was calculated to capture the dynamic shifts and relative strength of each stance over time.

To identify significant structural breaks within the time series data, the pruned exact linear time (PELT) algorithm [47], a computationally efficient method for change-point detection, was used. These identified change points correspond to major shifts in public opinion regarding the HPV vaccine, potentially linked to specific events or information dissemination campaigns. By juxtaposing the time series analysis results with a timeline of key news events and public health policies, a comprehensive interpretation of the observed trends was achieved. This approach facilitates a deeper understanding of the factors influencing the public acceptance of HPV vaccines and provides empirical evidence for future public health interventions and policy decisions.

Latent Dirichlet Allocation Topic Modeling Analysis

In this study, we used latent Dirichlet allocation (LDA) [32] topic modeling, a widely used technique for extracting latent thematic information from large volumes of text data, to analyze Japanese tweets related to the HPV vaccine between 2011 and 2021. Using LDA modeling, we aimed to gain a deeper understanding of the specific perspectives held by Japanese Twitter users across different stances toward the HPV vaccine. To determine the optimal number of topics that best represented the thematic structure of the dataset, we experimented with various topic numbers, ranging from 1 to 50. For each configuration, the performance of the model was evaluated using perplexity and coherence scores [48]. Details of the optimal topic number selection are shown in Figures S1, S4, and S7 in Multimedia Appendix 2. After the optimal number of topics was selected, qualitative analyses were conducted to refine and interpret the final thematic content. Similar to Niu et al [49], we calculated the monthly average expectation of tweets belonging to different topics for a fine-grain time series analysis.

In line with prior research indicating that misinformation is an important factor in vaccine hesitation [19,50,51], we also investigated the impact of misinformation on the HPV vaccine. First, we identified topics that were likely to contain misinformation through group discussion, focusing on tweets with an estimated probability of more than 80% of belonging to those topics. We then used the Claude-3-opus model [45] to assess whether each tweet attempted to disseminate misinformation regarding the HPV vaccine. To validate the accuracy of the model’s classification, we randomly selected 25 tweets each categorized as misinformation and 25 categorized as non-misinformation by the model. Three independent volunteers evaluated the credibility of the information in these

XSL•FO
RenderX

tweets by cross-referencing reputable news sources and data from the MHLW [52]. The verification process was repeated 3 times to estimate the classification accuracy of the model.

Explore Relationship Between HPV Vaccination and COVID-19

Considering the chronological overlap between the duration of data collection and the COVID-19 pandemic, which is a significant historical health event, this study investigated the potential relationship between HPV vaccination and COVID-19. Specifically, the objective of this study was to determine whether the COVID-19 pandemic had an impact on public attitudes toward HPV vaccination. From our dataset, all tweets comprising the keyword "COVID-19" in both Japanese and English were extracted. In addition to the time series analysis and LDA topic modeling previously described, the correlation between all HPV vaccine–related tweets and the tweets incorporating the "COVID-19" keywords was examined. Moreover, specific time points where the total number of tweets and the tweets containing "COVID-19" keywords exhibited simultaneous increases were investigated. To estimate the percentage of tweets pertaining to pivotal events, 100 tweets were randomly selected 3 times at each time point. Three volunteers verified the number of tweets related to the paramount event, and the percentage of related tweets was calculated by averaging the 3 results and computing CIs. To gain a deeper understanding of the data spikes, the influence of COVID-19 vaccine–related pivotal events on peak time points was also assessed.

Furthermore, we investigated the potential causality between the stance toward the HPV vaccine and the stance toward COVID-19. Specifically, we analyzed tweets containing "COVID-19" keywords and expressing either an advocate or opposed stance. We used logic analysis by LLM to determine whether the tweet used the HPV vaccine as an illustrative example to convey the author's perspective on COVID-19 and vice versa. The Claude-3-opus model was directly used for inference, with the prompt instructing it to read the tweet and provide an output of "HPV to COVID" if the tweet used the HPV vaccine as an example to express an idea about the COVID-19 vaccine, "COVID to HPV" if the reverse was true, or "Not related" if neither was applicable. The classified tweets were tallied weekly and visually presented to demonstrate the results.

The manuscript was prepared in accordance with the iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations [53]. The IDs of the tweets used in this research are provided in Multimedia Appendix 3.

Ethical Considerations

This study used publicly available and accessible tweets. We assert that our analysis is compliant with Twitter's usage policy in aggregate form without identifying specific individuals who published the Twitter posts. Furthermore, the number of vaccination supply and injection downloaded from the MHLW are open government data. Therefore, the activities described do not meet the requirements of human subject research and did not require review by an institutional review board.

Results

Model Selection

A comparison of all the models tested is shown in Multimedia Appendix 4. The F_1-scores of the fine-tuned larger models achieved better performance in most cases, except for Llama 3.1. This may be because it was not instruction tuned for Japanese corpora in advance. The fine-tuned Gemini 1.0 pro yielded the best results in our experiment, and we decided to perform the following steps based on Gemini 1.0 pro. Detailed results are shown in Multimedia Appendices 5 and 6.

Because the allowed number of tweets for fine-tuning Gemini 1.0 pro was fixed (<500), we experimented with different data ratios to determine the most effective category ratio to fine-tune the model. The results indicated that the model performed best when the ratio of the advocate, opposite, and unknown categories was 150:200:150. This ratio was consistent in the validation experiments, achieving the highest average F_1-score of 0.968, as shown in Multimedia Appendix 7. The detailed model performance for each ratio is shown in Multimedia Appendix 8.

After 5 rounds of random data selection and fine-tuning using the optimal data ratio, the performance of the fine-tuned Gemini 1.0 Pro model was evaluated using the remaining labeled data. The evaluation results showed significant improvements in precision, recall, and F_1-score for the fine-tuned model. The average F_1-score of the 3 best-performing models was 0.924, with a precision of 0.928 and a recall of 0.924. In comparison, the baseline performance of the untuned Gemini 1.0 Pro model on the same test dataset showed an F_1-score of 0.781, a precision of 0.829, and a recall of 0.796. This comparison clearly demonstrates the importance of fine-tuning for enhancing the performance of the model for specific tasks.

When classifying the unlabeled Japanese tweets, the results from the 3 best models were mostly consistent. Among all tweets, 86.85% (198276) received 3 identical result labels and 13.06% (29806) received 2 identical result labels. In cases where all 3 labels were completely inconsistent, these tweets were classified as unknown, accounting for 0.09% (209) of all the tweets.

Time Series Analysis of Stance

We first analyzed the trend of advocate rate toward the HPV vaccine from 2011 to 2021 and the HPV vaccine supply, injection from 2013 to 2022, as shown in Figure 1. It is noticeable that vaccine supply between 2013 and 2017 is based on estimation because of lacking fine-grained annual data. The lowest advocate rate appeared in 2013, when the Japanese government decided to suspend its recommendations for the HPV vaccine, which was announced only 2 months ago [54], leading to a sharp decrease of vaccine supply and injection 2014. The advocate rate recovered in 2 years and stayed to be around 50% until 2018, but the vaccine supply and injection stayed at an ignorable level. The advocate rate increased and stayed in a high level since 2019, during which 2020 witnessed numerous petitions urging the government to reinstate active

XSL•FO
RenderX

recommendations and prompt policy reviews [55]. The vaccine supply and injection started exponential growth since 2019, especially after the proactive HPV vaccine recommendations reinstated in 2022 [14-18]. Vaccine injected surpassed vaccine supply in 2020 and 2022, indicating that the vaccines remained last year were consumed.

Figure 1. Advocate rate toward human papillomavirus vaccine between 2011 and 2021 (line) and the number of vaccine supply and injection between 2013 and 2022 (bars). The vaccine supply from 2013 to 2017 was estimated based on Japanese Ministry of Health, Labour and Welfare (MHLW) official data due to lacking fine-grained annual data, and the vaccine supply after 2017 is from MHLW official data.

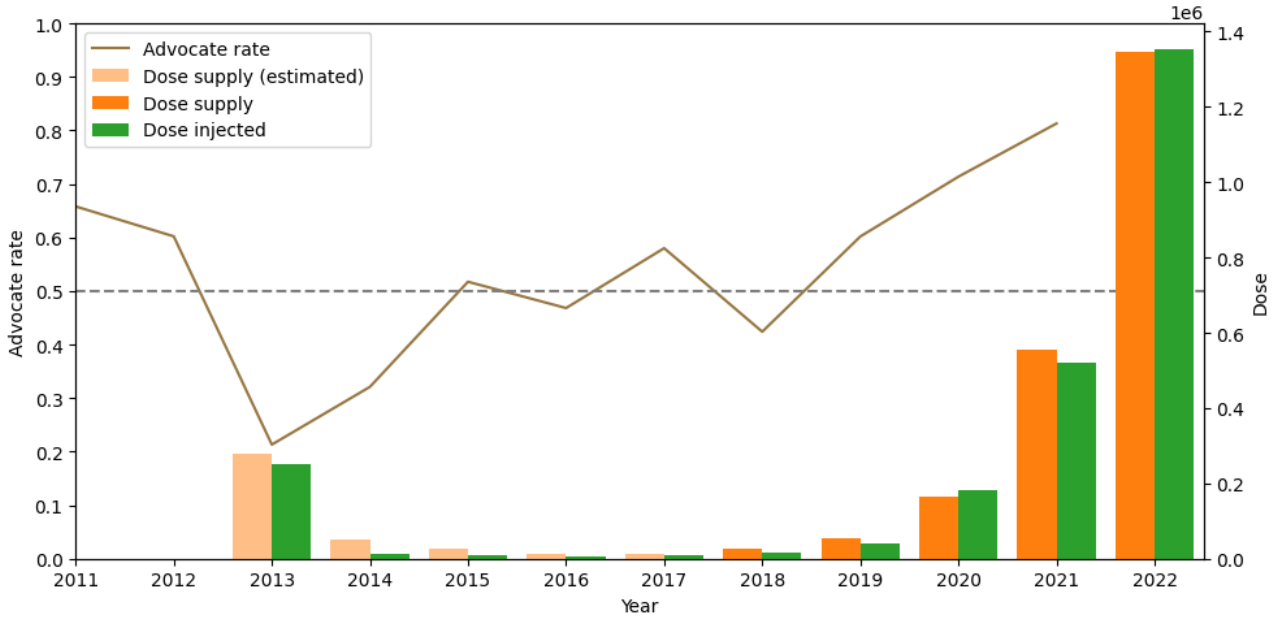

A time series analysis of the difference between advocates and opposed stances is shown in Figure 2, which shows several pivotal junctures marked by substantial shifts in the disparity between pro- and anti-HPV vaccination stances, suggesting notable fluctuations in public opinion. Overall, we can observe that there are 3 periods of fluctuation, and the 3 periods are relatively stable in stances. Using the PELT algorithm, we successfully discerned structural change points within the data, closely aligning with significant shifts in public attitudes toward HPV vaccines. Notably, these change points were observed in 2013, 2016, and 2020, coinciding with key public health events and policy developments. Beside the events in 2013 and 2020, in late 2016, widespread discourse surrounding vaccine safety was triggered by legal actions taken by individuals alleging adverse side effects [56].

Figure 2. Monthly differences between number of tweets of advocate and oppose stances. The significant low and high periods based on change points detected by pruned exact linear time are marked in gray.

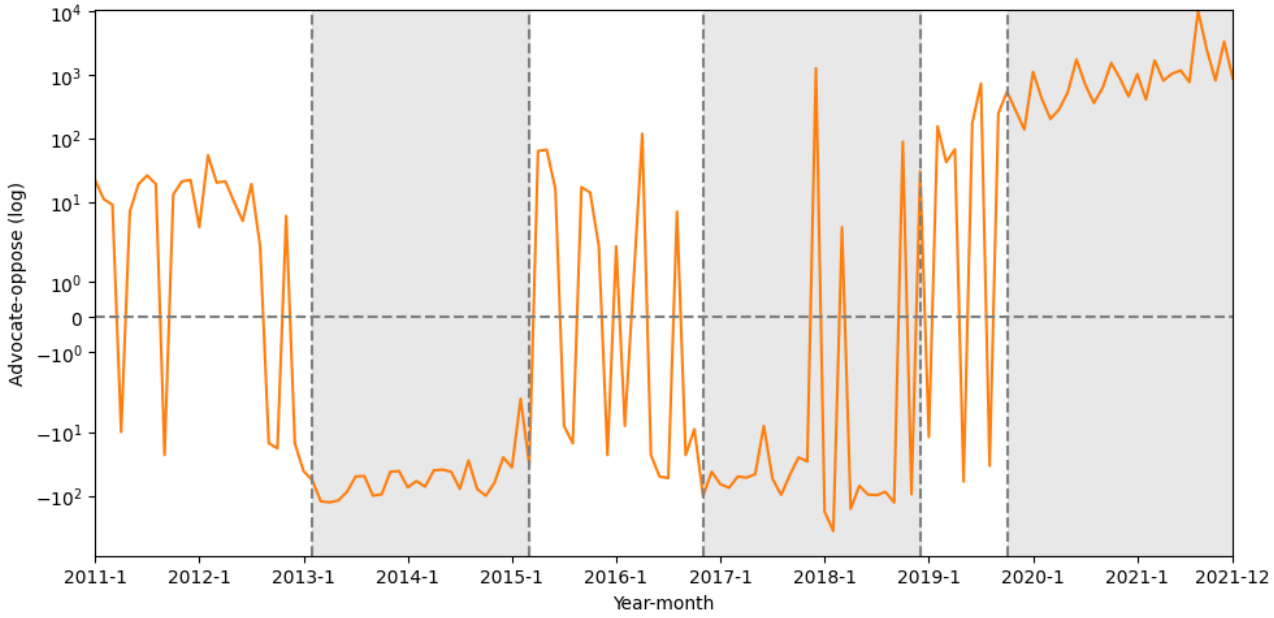

XSL•FO
RenderX

LDA Topic Modeling Analysis

We separately applied LDA modeling to the tweets of advocates, opposes, and unknown stances. There are 3 best topic numbers for advocates and unknowns and 4 for opposition. Details of the top words and weights are provided in Multimedia Appendix 2. We then calculated the ratio of the monthly expectations of tweets belonging to different topics to obtain fine-grained insights into the weights of the topics in each stance.

The distribution of various themes within the "advocate" stance is illustrated in Figure 3. The proportion of "Scientific and Media Discourse on HPV Vaccine Safety" (topic 1) underwent a significant increase until 2015, subsequently experiencing a gradual decline. Conversely, the proportion of "HPV Vaccine Effectiveness and Broader Public Health Measures" (topic 2) diminished, reaching its lowest point around 2015, after which it exhibited a consistent increase. The proportion of "Policy and Advocacy for HPV Vaccination Promotion" (topic 3) remained relatively stable, although a slight upward trend has been observed since 2019.

Figure 3. Monthly change of ratio of different topics in tweets of advocate stance from 2011 to 2021.

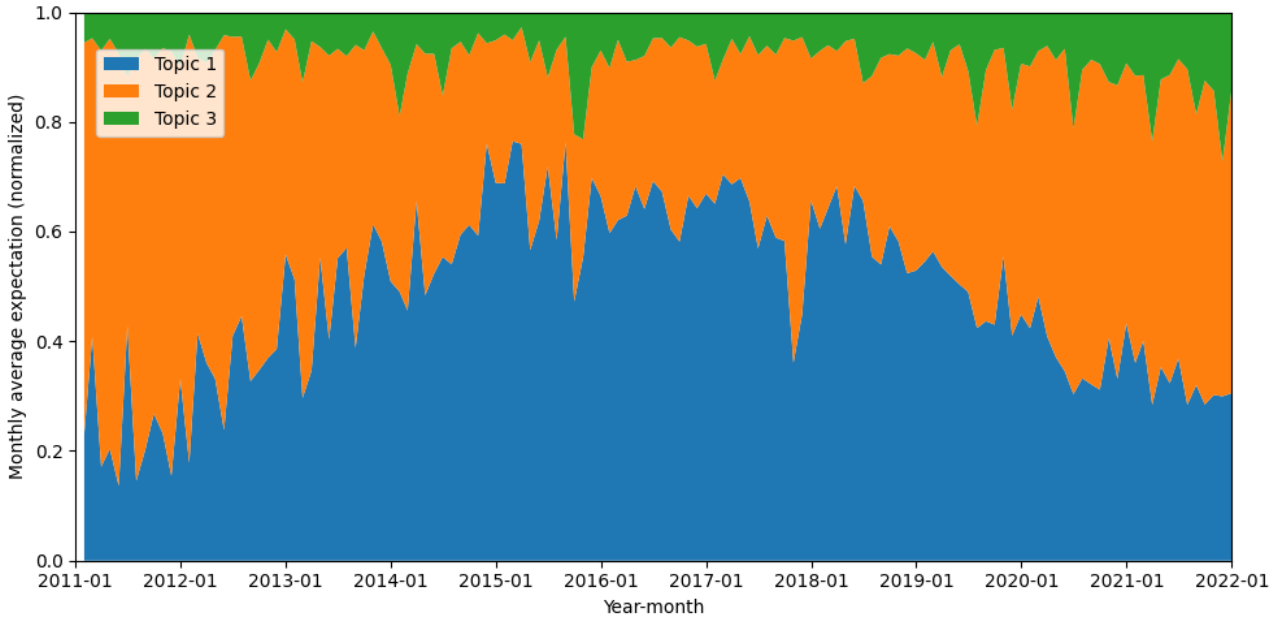

In the analysis of tweets expressing the "opposing" stance in Figure 4, "Skepticism and Opposition to Vaccination" (topic 2) demonstrated notable peaks in weight during the years 2013 and 2015, followed by a gradual decline and eventual stabilization of topic weight. The proportion of "Scientific Warnings and Public Health Risks" (topic 4) surged in late 2012, followed by a sustained increase throughout the decade, ultimately becoming the dominant theme within an opposing stance.

XSL•FO
RenderX

Figure 4. Monthly change of ratio of different topics in tweets of opposed stance from 2011 to 2021.

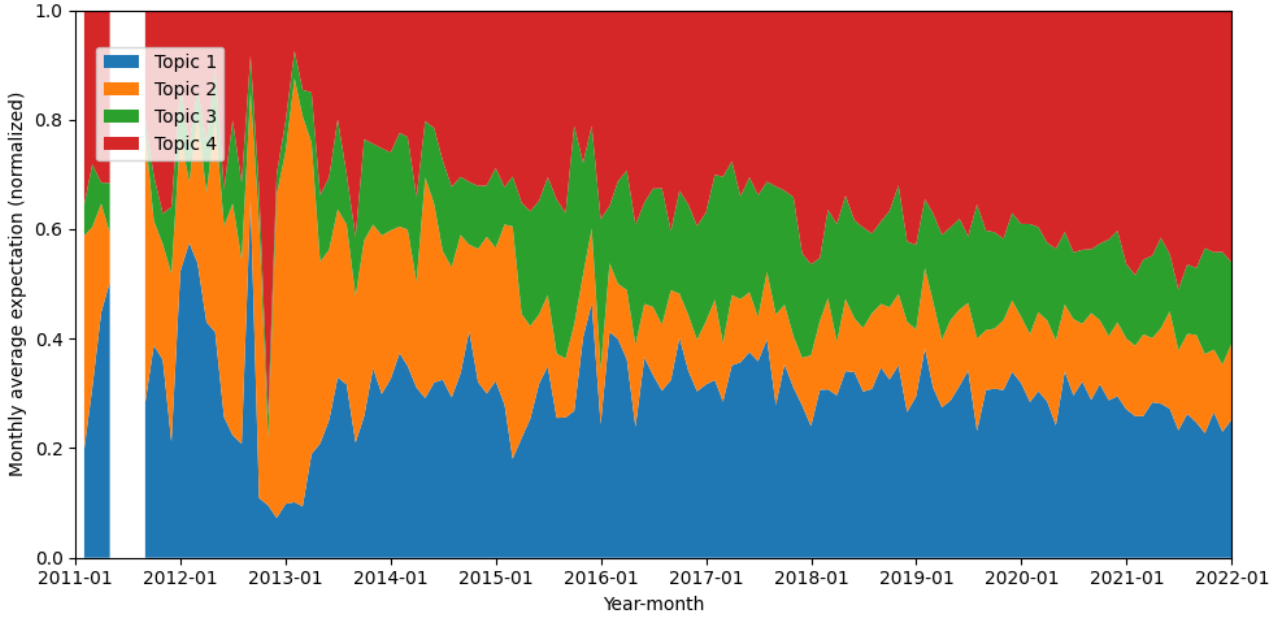

In the tweets classified as "unknown" by stance shown in Figure 5, "HPV Vaccine Efficacy and Public Health Initiatives" (topic 2) was the dominant theme prior to the year 2013 and then declined thereafter. Conversely, the proportion of tweets expressing "Opposition, Misinformation, and Activism Surrounding HPV Vaccination" (topic 3) displayed a gradual increase and has become the predominant theme since 2018. Notably, the theme of "HPV Vaccine Safety and Governmental Oversight" (topic 1) exhibited a gradual increase throughout the entire period under consideration.

Figure 5. Monthly change of ratio of different topics in tweets where the stances are unknown from 2011 to 2021.

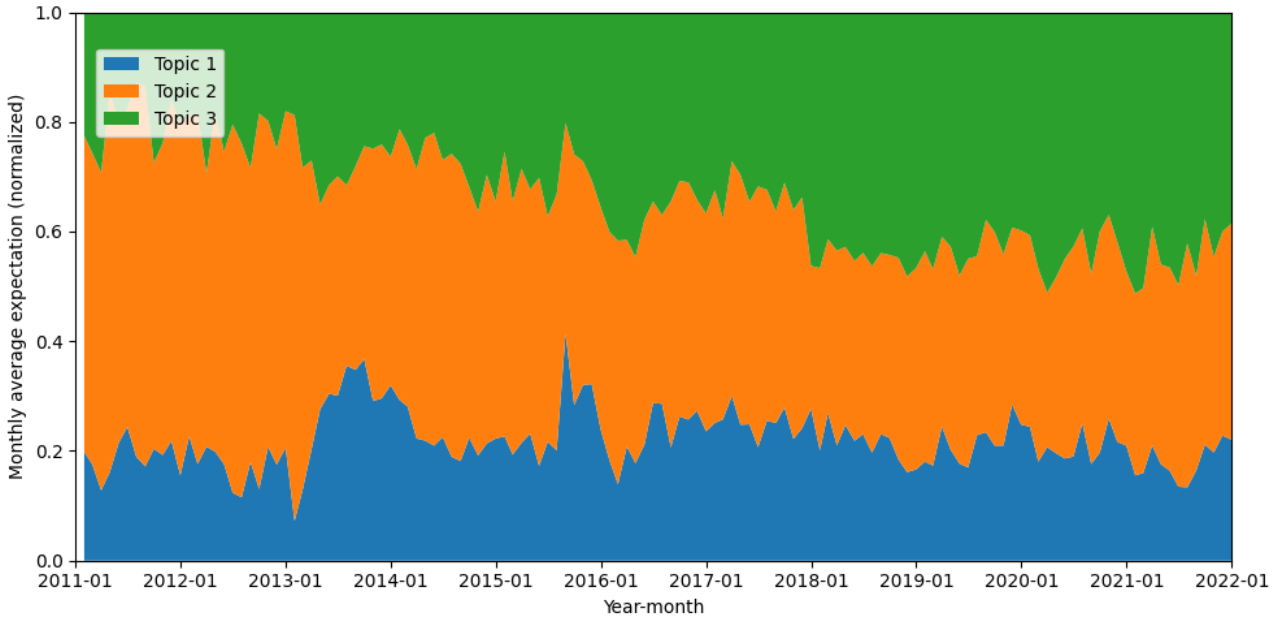

Then, we applied misinformation detection to certain topics. In tweets of opposed stance, we identified "Scientific Warnings and Public Health Risks" (topic 4) as the most likely to contain misinformation about the HPV vaccine. We extracted tweets with a probability of more than 80% belonging to this topic, yielding 3001 tweets. The misinformation classification results obtained using Claude-3-opus are shown in Figure 6. The classification accuracy, determined through random sampling and verification conducted thrice, ranged from 95% CI 89.74% to 100%. Our findings revealed that the prevalence of misinformation increased in most years, except for 2014 and 2015. The ratio of misinformation in HPV vaccine–related tweets began to spike sharply in 2012, followed by a rapid decline to its nadir in 2015. Subsequently, the ratio gradually increased, reaching a secondary, albeit lower, peak in 2018, after which it began to decline gradually.

XSL•FO
RenderX

Figure 6. Count and ratio of tweets classified as misinformation. The bars show the total number of tweets detected as misinformation each year, and the line shows the ratio of the tweets containing misinformation within the total number of tweets of opposed stance each year.

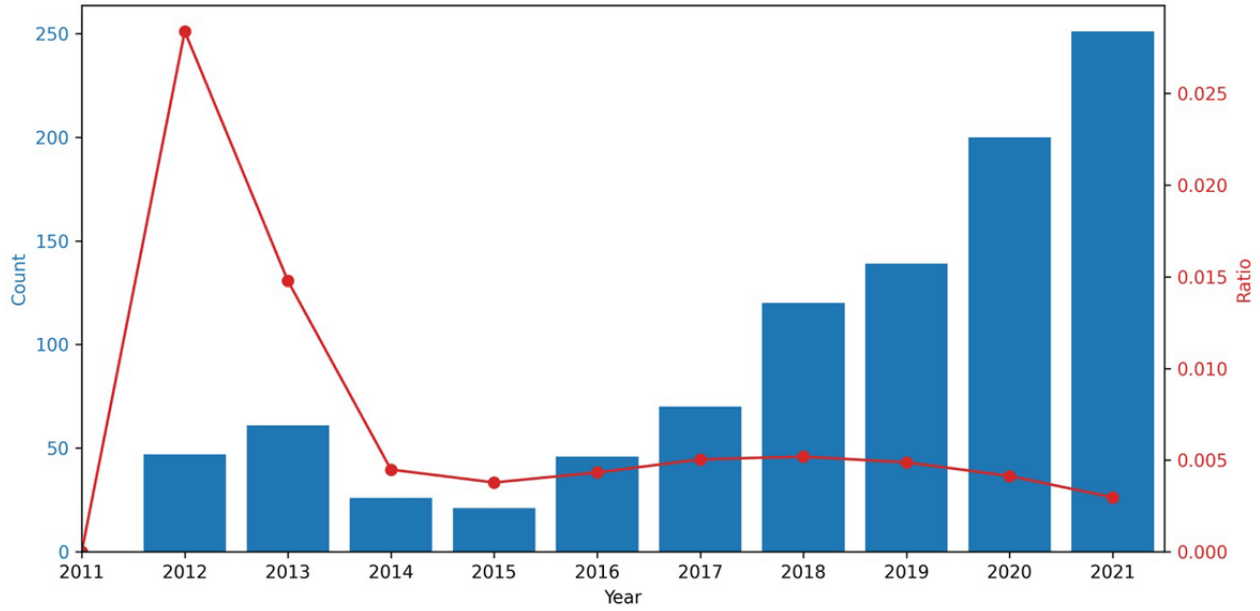

Relationship Between HPV Vaccination and COVID-19

We conducted a weekly time series analysis of tweets containing COVID-19 keywords, as illustrated in Figure 7. There are no tweets with "COVID-19" keyword in the first 2 weeks of 2020, so all the calculations began from the third week. The Pearson correlation coefficient (r) was 0.793 (P=1.59×10^{-23}), indicating a relatively high correlation between all tweets and those containing "COVID-19" keywords. The time lag for the highest cross-correlation was zero. Several peaks in both all tweets and tweets containing "COVID-19" keywords were observed in 2021 (week 3, weeks 34-35, and week 45), with the 34th-35th weeks exhibiting the highest peak during the entire study period, which coincided with a campaign to call for resumption of active HPV vaccination recommendations [57] (95% CI 37.0%-39.8% HPV vaccine-related tweets; 95% CI 1.5%-9.1% tweets with "COVID-19" keywords).

Figure 7. Number of all tweets related to human papillomavirus vaccine (gray), and the tweets containing the "COVID-19" keywords (blue), with key events related to COVID-19 vaccine marked. The points beginning with "N" are negative key events that may lead to opposition to COVID-19 vaccines, and the points beginning with "P" are positive key events that may lead to advocacy to COVID-19 vaccines.

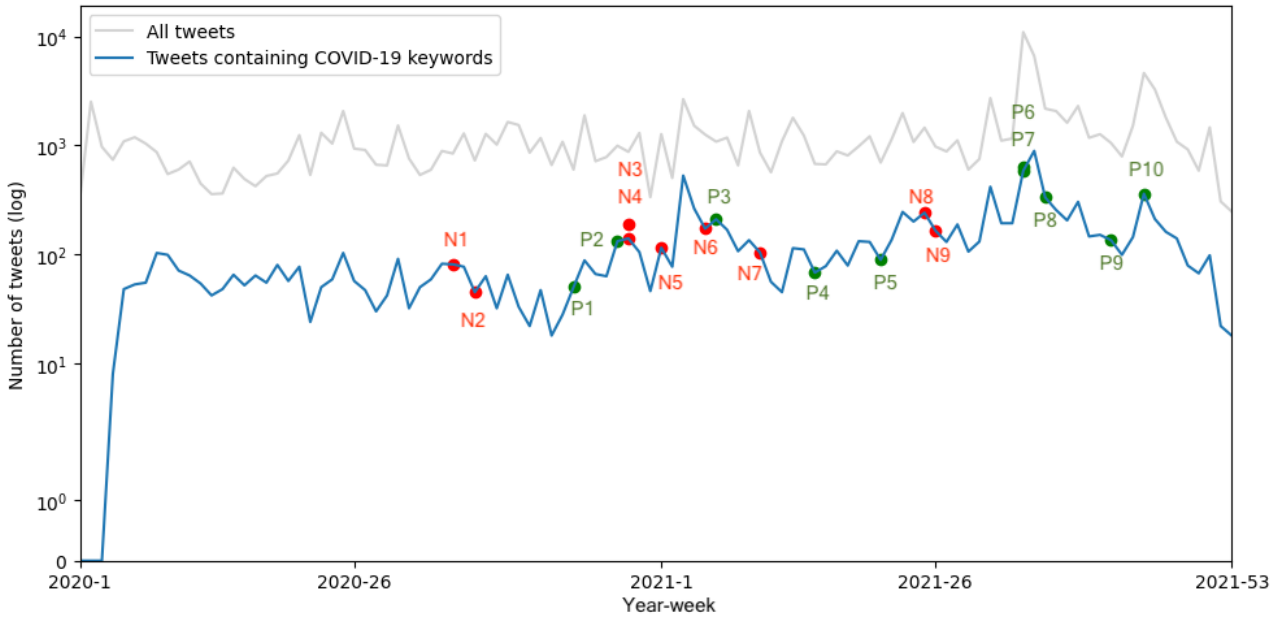

A comparative analysis was performed to examine the differences between advocate and opposed stances in both the overall tweet corpus and the subset containing the "COVID-19" keyword, as illustrated in Figure 8. The r between advocate and oppose sentiment differences is substantial (r=0.723; P=6.09×10^{-18}), with no temporal lag observed in the highest cross-correlation. In the broader context of all HPV

XSL•FO
RenderX

vaccine–related tweets, advocacy sentiments generally prevailed over opposing sentiments. However, for tweets containing "COVID-19" keywords, the dominant stance fluctuated over time. Both datasets exhibited a pronounced peak in advocacy during weeks 34-35 of 2021, whereas opposing sentiments dominated in week 22 of 2020 for all tweets.

Figure 8. The difference between tweets of advocate stance and oppose stance for all human papillomavirus vaccine tweets (gray) and tweets containing the "COVID-19" keyword (blue), with key events related to COVID-19 vaccine marked. The points beginning with "N" are negative key events that may lead to opposition to COVID-19 vaccines, and the points beginning with "P" are positive key events that may lead to advocacy to COVID-19 vaccines.

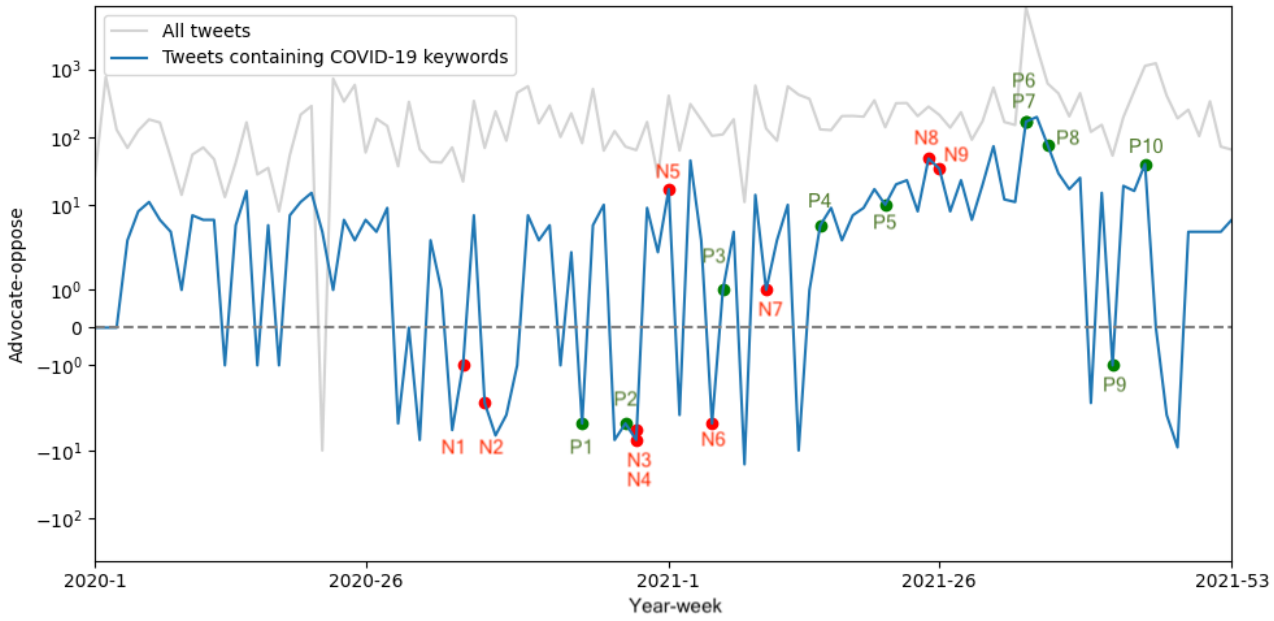

We additionally assessed the impact of COVID-19 vaccine–related events on the volume of all HPV vaccine–related tweets and tweets containing the keywords "COVID-19." Details of the key events were gathered from Japan Broadcasting Corporation (NHK) [58] and are provided in Table S1 in Multimedia Appendix 2. As illustrated in Figure 7, it is evident that the key events do not consistently result in peaks in tweet volume for both all HPV vaccine–related tweets and tweets containing the keywords "COVID-19." Furthermore, as depicted in Figure 8, positive and negative events do not always lead to an increase in the number of supportive or opposing stances.

LDA was conducted on tweets containing COVID-19 keywords advocating for and opposing stances. Multimedia Appendix 2 shows detailed results. To investigate nuanced changes in topic weights over time, the expectation of tweets belonging to different topics for both advocate and opposition stances was calculated, as shown in Multimedia Appendix 2. Analysis of these data revealed no significant changes in the proportions of different topics between 2020 and 2021.

Finally, we investigated the potential causal relationship between attitudes toward HPV and COVID-19 vaccines using logical analysis. To evaluate the accuracy of Claude-3-opus in performing the classification task without fine-tuning, we randomly sampled 100 tweets 3 times and allowed 3 volunteers to judge the correctness. The accuracy is 95% CI 72.92%-91.74%. The weekly classification results are shown in Figure 9. The "HPV to COVID-19" category predominated throughout the entire study period for both supportive and opposing stances. Notably, tweets advocating for vaccines consistently outnumbered those opposing vaccines across all categories.

XSL•FO
RenderX

Figure 9. Weekly number of tweets with "COVID-19" keywords, using HPV vaccine as example to share stance to COVID-19 vaccine ("HPV to COVID-19"), using COVID-19 vaccine as example to share stance to HPV vaccine ("COVID-19 to HPV"), or neither ("Not related") in 2021. (A) Result on tweets with advocate stance. (B) Result on tweets with opposed stance. HPV: human papillomavirus.

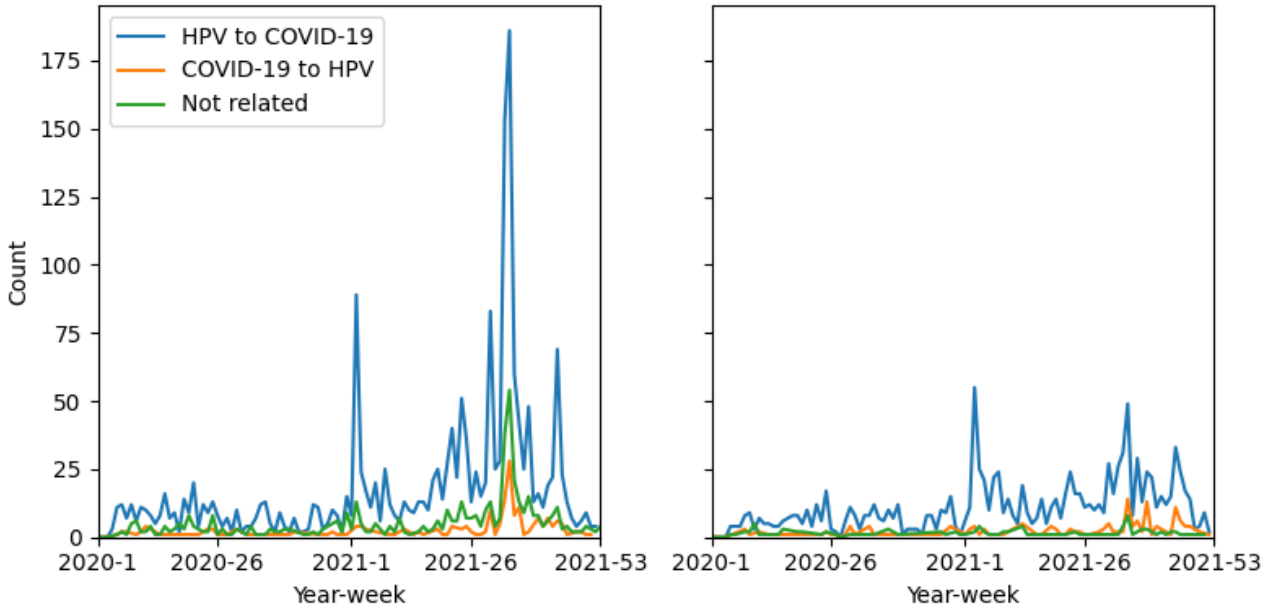

Discussion

Principal Findings

In this study, we used traditional NLP models and LLMs to perform a stance analysis of social media content, focusing on the public discourse around the HPV vaccine. Our findings demonstrate that LLMs provide a more nuanced understanding of public attitudes than traditional sentiment analysis, revealing critical insights into factors influencing vaccine hesitancy, such as confidence, complacency, and convenience. Specifically, we observed a slight increase in public advocacy and vaccine confidence toward HPV vaccines over time, influenced by policy decisions, media coverage, and public communication efforts. The WHO's 3Cs model (Confidence, Complacency, and Convenience) proved instrumental in framing these dynamics and offers a pathway for policy makers to address vaccine hesitancy in Japan. We also observed some evidence of the positive influence of stance toward HPV vaccine on the stance toward COVID-19 vaccine.

Confidence pertains to trust in the vaccine's safety, efficacy, and the health care system responsible for its administration and plays a key role in the stance toward the HPV vaccine in Japan. In our analysis, we observed significant fluctuations in the public stance toward the HPV vaccine, which were closely linked to government policy decisions and media coverage. The 2013 suspension of the HPV vaccine recommendations by the Japanese government was a key moment that led to a sharp decline in public trust. This decline can be attributed to increased public uncertainty and fear regarding vaccine safety, as shown in the time series analysis. The LDA topic modeling results also support this, showing a decline in the proportion of tweets related to "HPV Vaccine Efficacy and Public Health Initiatives" (topic 2) where the stance is unknown after 2013, while "Opposition, Misinformation, and Activism Surrounding HPV Vaccination" (topic 3) gradually became the predominant theme since 2018, highlighting growing public distrust and misinformation. Conversely, the resurgence of advocacy in 2020 reflected an improvement in public confidence, which likely influenced efforts to reinstate vaccine recommendations. This is directly shown by the exponential increase in vaccine supply and vaccine uptake and also reflected in the increased proportion of tweets related to "Scientific and Media Discourse on HPV Vaccine Safety" within the advocacy stance, which showed a significant increase until 2015, followed by a gradual decline, indicating ongoing advocacy efforts. Our findings illustrate that confidence is highly sensitive to policy decisions and that restoring public trust after a decline requires consistent and clear communication.

Furthermore, we explored the role of misinformation in undermining public confidence regarding HPV vaccinations. The spike in the ratio of tweets containing misinformation in 2012 likely contributed to the significant drop in vaccine confidence observed the following year, which aligns with findings from other studies [59-61]. From 2014 onward, the ratio of misinformation exhibited slight changes but remained consistent. The overall number of misinformation tweets increased in most years, indicating the persistent and long-lasting impact of misinformation. Therefore, addressing misinformation is essential for restoring and maintaining public trust in vaccines, because unchecked misinformation can severely erode public confidence and contribute to vaccine hesitancy.

Complacency refers to the perceived necessity of vaccination, which diminishes when individuals consider the risk of vaccine-preventable diseases low. Adjei Boakye et al [62] found that "not necessary" was consistently among the top reasons cited by parents for not vaccinating their children against HPV between 2010 and 2020. The topic "HPV Vaccine Efficacy and Public Health Initiatives" under the "unknown" stance experienced a sharp decline post-2013, indicating diminished public engagement and perceived necessity. The perceived low

XSL•FO
RenderX

risk of HPV, combined with concerns about its side effects, may contribute to a reduced sense of urgency regarding vaccination. However, our time series analysis showed that complacency began to decrease in 2020 as advocacy efforts intensified, and the public became more aware of the risks of HPV and the benefits of vaccination. The upward trend in the topic "HPV Vaccine Effectiveness and Broader Public Health Measures" within the "advocate" stance since 2015 further reflects this shift toward increased awareness and decreasing complacency. This shift underscores the importance of proactive communication regarding the risk of vaccine-preventable diseases to counteract complacency.

Convenience encompasses factors such as availability, affordability, and accessibility of vaccination. In our study, the importance of convenience became evident when analyzing the periods of increased advocacy for HPV vaccination. The vaccine doses injected surpassed the doses supplied in 2020 and 2022, indicating that the inject number may be limited by the supply. Also, the resurgence of advocacy in 2020 coincided with governmental initiatives aimed at improving access to the HPV vaccine, including the reintroduction of public recommendations and enhancement of vaccination services. This alignment is reflected in the relatively stable proportion of tweets related to the topic "Policy and Advocacy for HPV Vaccination Promotion" within the advocacy stance, which exhibited a slight upward trend since 2019. These changes may encourage efforts to facilitate access to vaccination services and play a critical role in influencing public willingness to get vaccinated over time.

In addition to the findings regarding the HPV vaccine, we also noticed that stances toward the HPV vaccine may influence stances toward the COVID-19 vaccine during the COVID-19 pandemic. Specifically, our logical analysis revealed that increased confidence in the HPV vaccine might lead to higher confidence in the COVID-19 vaccine. Tweets categorized under "HPV to COVID-19"—which used the HPV vaccine as a reference point for the COVID-19 vaccine stance—predominated throughout the study period for both supportive and opposing stances. Notably, vaccine advocacy consistently outnumbered opposition across all categories, suggesting that rising confidence in one vaccine may positively affect confidence in others.

The findings of this study provide actionable insights for public health policies and interventions in Japan. First, improving confidence in vaccines requires transparent and consistent communication from government bodies and health care professionals, particularly during crises. Proactively addressing misinformation through targeted campaigns on social media can help rebuild public trust. To tackle complacency, educational programs highlighting the risks of HPV and the benefits of vaccination should be prioritized. Finally, enhancing convenience by ensuring vaccine availability, affordability, and accessibility—such as expanding vaccination services and public recommendations—can significantly increase uptake. The policy makers should also keep in mind that vaccine hesitance in one disease can affect all vaccines. These strategies collectively align with the WHO's 3Cs model to combat vaccine hesitancy effectively.

Our results demonstrate that stance analysis using LLMs provides a more nuanced understanding than traditional sentiment analysis. Unlike traditional sentiment analysis, which captures general emotional tones, LLMs can identify specific stances such as supportive, opposing, or neutral positions, allowing for a deeper comprehension of public attitudes and decision-making processes. Moreover, the successful application of LLMs highlights their potential as powerful tools for social media analysis, providing deep insights into public health discourse by capturing more complex logic, such as causality and context-specific nuances. These results not only validate the effectiveness of our methodology but also provide strong empirical support for using LLMs for similar tasks in the future.

In future work, we plan to explore why LLMs outperform traditional models by focusing on factors such as the model size and external information. We are also considering building an LLM to identify misinformation regarding HPV vaccines with higher accuracy. We also plan to collect social media data about other vaccines such as COVID-19 to make more solid causality analysis toward the relationship among attitudes toward vaccines. These efforts have refined our understanding of LLMs and improved public health strategies.

Limitations

Our study has several limitations, including the fact that social media users may not represent the entire population, which limits the generalizability of our findings. We used only X (Twitter) data in this research, which may also affect the representativeness. In addition, these models may introduce classification errors, particularly when addressing nuanced or ambiguous stances and logic. Other factors including but not limited to the biased training data, outdated information, domain adaptation issues, and hallucinations of the LLMs can also lead to inaccuracy of the statistical results. In the logic analysis of relationship between attitudes toward HPV and COVID-19 vaccine, lacking data related to COVID-19 vaccine, we are not able to further prove the causality with methods such as Granger causality tests or structural equation modeling. Furthermore, the keywords used to collect data may have led to the omission of relevant tweets or the addition of irrelevant tweets, potentially affecting the comprehensiveness of our analysis.

Comparison With Prior Work

Our work stands out from prior studies because of the scale and depth of the dataset, which includes 10 years of tweets in Japanese, providing a comprehensive view of public attitudes over time. We applied both traditional NLP methods and cutting-edge LLMs, which allowed us to analyze nuanced shifts in public sentiment and stance with greater precision. Unlike traditional surveys that provide snapshots of attitudes, our approach captures real-time changes in public discourse. Furthermore, compared with studies relying on Google search trends, our analysis offers more fine-grained insights into specific stances and topics, highlighting not just information-seeking behavior but also the underlying attitudes that drive vaccine hesitancy or advocacy.

XSL•FO
RenderX

Conclusions

In this study, we provided a decade long examination of public discourse on HPV vaccination in Japan by analyzing social media content with traditional NLP models and LLMs. We also contextualized the complex factors influencing stances toward HPV vaccines under the WHO's 3Cs model. Public confidence fluctuated significantly in response to government actions and media coverage, indicating the sensitivity of trust in policy decisions, whereas complacency was affected by perceived risks and proactive advocacy. Convenience is crucial for improving vaccine accessibility and for shaping the public's willingness to be vaccinated. Moreover, we provided some evidence showing that confidence in one vaccine, such as HPV, may influence confidence in others, such as COVID-19, highlighting interconnected public health narratives. Addressing misinformation, enhancing communication, and improving accessibility remain the key strategies for building trust and reducing hesitancy. Future public health strategies can benefit from these insights by designing effective interventions to boost vaccine confidence and uptake.

Acknowledgments

This work was supported by JST SPRING (grant JPMJSP2110) and Google PhD fellowship. We used the generative artificial intelligence (AI) tool ChatGPT [63] by OpenAI to refine the writings and structure for the research, which were further reviewed and revised by the study group. The original ChatGPT transcripts are made available in Multimedia Appendix 9.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data of vaccine uptake and supply from 2013 to 2022.
[DOCX File , 17 KB - jmir_v27i1e68881_app1.docx]

Multimedia Appendix 2
Results of latent Dirichlet allocation.
[DOCX File , 4278 KB - jmir_v27i1e68881_app2.docx]

Multimedia Appendix 3
All Twitter IDs used in this research.
[XLSX File (Microsoft Excel File), 10861 KB - jmir_v27i1e68881_app3.xlsx]

Multimedia Appendix 4
Comparison of weighted F1-scores among all the models on 300 test data. Bars show the average and SD on 3 test datasets selected by different random seeds. The numbers above the bars show the average of weighted F1-scores.
[PNG File , 47 KB - jmir_v27i1e68881_app4.png]

Multimedia Appendix 5
Details of model selection for stance analysis.
[DOCX File , 36 KB - jmir_v27i1e68881_app5.docx]

Multimedia Appendix 6
Selection of ratio of training data of difference stances.
[DOCX File , 38 KB - jmir_v27i1e68881_app6.docx]

Multimedia Appendix 7
Comparison of weighted F1-scores among models trained with different ratios of stance categories (advocate, opposite, and unknown) in the training data. Bars show the average and SD on 3 test datasets selected by different random seeds. The numbers above the bars show the average of weighted F1-scores.
[PNG File , 22 KB - jmir_v27i1e68881_app7.png]

Multimedia Appendix 8
Performance of the final 5 models.
[DOCX File , 27 KB - jmir_v27i1e68881_app8.docx]

XSL•FO
RenderX

Multimedia Appendix 9
The record of chat with ChatGPT for manuscript correction and refining.
[MOV File , 13035 KB - jmir_v27i1e68881_app9.mov]

References

1. Cervical cancer. World Health Organization. 2019 Dec 02. URL: https://www.who.int/health-topics/cervical-cancer [accessed 2024-07-17]
2. Immunization data. WHO Immunization Data portal—Global. 2024 Jul 15. URL: https://immunizationdata.who.int/ [accessed 2024-07-17]
3. Documents for the national conference of prefectural officials (temporary special grant for emergency promotion of vaccination against cervical cancer, etc.) [in Japanese]. Ministry of Health, Labour and Welfare. 2010 Dec 9. URL: https://www.mhlw.go.jp/bunya/kenkou/other/101209.html [accessed 2024-07-17]
4. Immunization Act [in Japanese]. e-GOV Law Search. 2024 Apr 01. URL: https://elaws.e-gov.go.jp/document?lawid=323AC0000000068 [accessed 2024-07-17]
5. Response to routine vaccination for human papillomavirus infection (recommendation) [in Japanese]. Ministry of Health, Labour and Welfare. 2013 Jun 14. URL: https://www.mhlw.go.jp/stf/shingi2/0000091963.html [accessed 2024-07-17]
6. Sipp D, Frazer IH, Rasko JEJ. No vacillation on HPV vaccination. Cell 2018;172(6):1163-1167 [FREE Full text] [doi: 10.1016/j.cell.2018.02.045] [Medline: 29522737]
7. Hanley SJB, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. Lancet 2015;385(9987):2571. [doi: 10.1016/S0140-6736(15)61152-7] [Medline: 26122153]
8. Ueda Y, Enomoto T, Sekine M, Egawa-Takata T, Morimoto A, Kimura T. Japan's failure to vaccinate girls against human papillomavirus. Am J Obstet Gynecol 2015;212(3):405-406. [doi: 10.1016/j.ajog.2014.11.037] [Medline: 25434842]
9. Tanaka Y, Ueda Y, Yoshino K, Kimura T. History repeats itself in Japan: failure to learn from rubella epidemic leads to failure to provide the HPV vaccine. Hum Vaccin Immunother 2017;13(8):1859-1860 [FREE Full text] [doi: 10.1080/21645515.2017.1327929] [Medline: 28604161]
10. Sekine M, Kudo R, Yamaguchi M, Hanley SJB, Hara M, Adachi S, et al. Japan's ongoing crisis on HPV vaccination. Vaccines (Basel) 2020;8(3):362 [FREE Full text] [doi: 10.3390/vaccines8030362] [Medline: 32640691]
11. Larson HJ, Wilson R, Hanley S, Parys A, Paterson P. Tracking the global spread of vaccine sentiments: the global response to Japan's suspension of its HPV vaccine recommendation. Hum Vaccin Immunother 2014;10(9):2543-2550 [FREE Full text] [doi: 10.4161/21645515.2014.969618] [Medline: 25483472]
12. Yagi A, Ueda Y, Kakuda M, Nakagawa S, Hiramatsu K, Miyoshi A, et al. Cervical cancer protection in Japan: where are we? Vaccines (Basel) 2021;9(11):1263 [FREE Full text] [doi: 10.3390/vaccines9111263] [Medline: 34835194]
13. Sekine M. Japanese crisis of HPV vaccination. Int J Pathol Clin Res 2016;2(2):1-3. [doi: 10.23937/2469-5807/1510039]
14. Namba M, Kaneda Y, Kawasaki C, Shrestha R, Tanimoto T. Underlying background of the current trend of increasing HPV vaccination coverage in Japan. Glob Health Med 2023;5(4):255-256 [FREE Full text] [doi: 10.35772/ghm.2023.01010] [Medline: 37655180]
15. Lelliott M, Sahker E, Poudyal H. A review of parental vaccine hesitancy for human papillomavirus in Japan. J Clin Med 2023;12(5):2004 [FREE Full text] [doi: 10.3390/jcm12052004] [Medline: 36902790]
16. Miyagi E. Human papillomavirus (HPV) vaccination in Japan. J Obstet Gynaecol Res 2024;50 Suppl 1:65-71. [doi: 10.1111/jog.16020] [Medline: 38979785]
17. Hayashi Y, Shimizu Y, Netsu S, Hanley S, Konno R. High HPV vaccination uptake rates for adolescent girls after regional governmental funding in Shiki City, Japan. Vaccine 2012;30(37):5547-5550. [doi: 10.1016/j.vaccine.2012.06.038] [Medline: 22749837]
18. Yagi A, Ueda Y, Kimura T. HPV vaccine issues in Japan: a review of our attempts to promote the HPV vaccine and to provide effective evaluation of the problem through social-medical and behavioral-economic perspectives. Vaccine 2024;42(22):125859. [doi: 10.1016/j.vaccine.2024.03.080] [Medline: 38616440]
19. Terada M, Okuhara T, Nagasawa T, Okada H, Goto E, Kiuchi T. Public perception of the resumption of HPV vaccine recommendation in Japan: Twitter content analysis. Health Promot Int 2023;38(6):daad153 [FREE Full text] [doi: 10.1093/heapro/daad153] [Medline: 37966160]
20. Ueda N, Yokouchi R, Onoda T, Ogihara A. Characteristics of articles about human papillomavirus vaccination in Japanese newspapers: time-series analysis study. JMIR Public Health Surveill 2017;3(4):e97 [FREE Full text] [doi: 10.2196/publichealth.8237] [Medline: 29258972]
21. Global Advisory Committee on Vaccine safety statement on safety of HPV vaccines. Ministry of Health, Labour and Welfare. 2015 Dec 17. URL: https://www.mhlw.go.jp/file/05-Shingikai-10601000-Daijinkanboukouseikagakuka-Kouseikagakuka/0000125190.pdf [accessed 2024-07-18]
22. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
23. X's penetration rate in Japan FY 2014-2023. Statista. 2024 Jun 21. URL: https://www.statista.com/statistics/425342/twitter-penetration-japan/ [accessed 2024-12-27]

XSL•FO
RenderX

24. Share of people who use X in Japan in fiscal year 2023, by age group. Statista. 2024 Jun 21. URL: https://www.statista.com/statistics/1077538/japan-twitter-penetration-rate-by-age-group/ [accessed 2024-07-18]
25. Fues Wahl H, Wikman Erlandson B, Sahlin C, Nyaku M, Benĉina G. Analysis of vaccine messages on social media (Twitter) in Scandinavia. Hum Vaccin Immunother 2022;18(1):2026711 [FREE Full text] [doi: 10.1080/21645515.2022.2026711] [Medline: 35103579]
26. Tomaszewski T, Morales A, Lourentzou I, Caskey R, Liu B, Schwartz A, et al. Identifying false human papillomavirus (HPV) vaccine information and corresponding risk perceptions from Twitter: advanced predictive models. J Med Internet Res 2021;23(9):e30451 [FREE Full text] [doi: 10.2196/30451] [Medline: 34499043]
27. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open 2017;7(10):e016869 [FREE Full text] [doi: 10.1136/bmjopen-2017-016869] [Medline: 28982821]
28. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
29. Suzuki D, Nishimura S, Jin Q, Ogihara A. Analysis of the conversations on Twitter regarding HPV vaccine. J Consum Health Internet 2021;25(4):397-406. [doi: 10.1080/15398285.2021.1989971]
30. Louis A. Natural language processing for social media. Comput Linguist 2016;42(4):833-836.
31. Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. J Mach Learn Res 2003;3:1137-1155.
32. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn 2003;3:993-1022.
33. Radwan A, Amarneh M, Alawneh H, Ashqar H, AlSobeh A. Predictive analytics in mental health leveraging LLM embeddings and machine learning models for social media analysis. Int J Web Serv Res 2024;21(1):1-22.
34. Tweepy. Tweepy. 2024 Jan 21. URL: https://www.tweepy.org/ [accessed 2024-10-06]
35. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
36. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online on October 11, 2018 [FREE Full text] [doi: 10.48550/arXiv.1810.04805]
37. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv Preprint posted online on October 2, 2019 [FREE Full text] [doi: 10.48550/arXiv.1910.01108]
38. Gemma Team. Gemma 2: improving open language models at a practical size. arXiv Preprint posted online on July 31, 2024 [FREE Full text] [doi: 10.48550/arXiv.2408.00118]
39. Introducing Llama 3.1: our most capable models to date. Meta AI. 2024 Jul 23. URL: https://ai.meta.com/blog/meta-llama-3-1/ [accessed 2024-10-06]
40. Gemini Team. Gemini: A family of highly capable multimodal models. arXiv Preprint posted online on December 19, 2023 [FREE Full text] [doi: 10.48550/arXiv.2312.11805]
41. GiNZA. GiNZA—Japanese NLP Library. 2024 Mar 31. URL: https://megagonlabs.github.io/ginza/ [accessed 2024-10-06]
42. tohoku-nlp/bert-base-japanese-v3. Hugging Face. 2023 May 19. URL: https://huggingface.co/tohoku-nlp/bert-base-japanese-v3 [accessed 2024-10-06]
43. line-corporation/line-distilbert-base-japanese. Hugging Face. 2023 Mar 22. URL: https://huggingface.co/line-corporation/line-distilbert-base-japanese [accessed 2024-10-06]
44. Liu Y, Han T, Ma S, Zhang J, Yang Y, Tian J, et al. Summary of ChatGPT-Related research and perspective towards the future of large language models. Meta-Radiology 2023;1(2):100017. [doi: 10.1016/j.metrad.2023.100017]
45. Introducing the next generation of Claude. Anthropic. 2024 Mar 04. URL: https://www.anthropic.com/news/claude-3-family [accessed 2024-10-06]
46. Dettmers T, Pagnoni A, Holtzman A, Zettlemoyer L. QLoRA: efficient finetuning of quantized LLMs. arXiv Preprint posted online on May 23, 2023 [FREE Full text] [doi: 10.48550/arXiv.2305.14314]
47. Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc 2012;107(500):1590-1598. [doi: 10.1080/01621459.2012.737745]
48. Gan J, Qi Y. Selection of the optimal number of topics for LDA topic model-taking patent policy analysis as an example. Entropy (Basel) 2021;23(10):1301 [FREE Full text] [doi: 10.3390/e23101301] [Medline: 34682025]
49. Niu Q, Liu J, Kato M, Nagai-Tanima M, Aoyama T. The effect of fear of infection and sufficient vaccine reservation information on rapid COVID-19 vaccination in Japan: evidence from a retrospective twitter analysis. J Med Internet Res 2022;24(6):e37466 [FREE Full text] [doi: 10.2196/37466] [Medline: 35649182]
50. Teoh D. The power of social media for HPV vaccination-not fake news!. Am Soc Clin Oncol Educ Book 2019;39:75-78 [FREE Full text] [doi: 10.1200/EDBK_239363] [Medline: 31099637]
51. Zimet GD, Rosberger Z, Fisher WA, Perez S, Stupiansky NW. Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation. Prev Med 2013;57(5):414-418 [FREE Full text] [doi: 10.1016/j.ypmed.2013.05.013] [Medline: 23732252]

XSL•FO
RenderX

52. Human papillomavirus infection—cervical cancer (uterine cervix cancer) and HPV vaccine [in Japanese]. Ministry of Health, Labour and Welfare. 2024 Apr 15. URL: https://www.mhlw.go.jp/bunya/kenkou/kekkaku-kansenshou28/index.html [accessed 2024-10-06]
53. Perrin Franck C, Babington-Ashaye A, Dietrich D, Bediang G, Veltsos P, Gupta PP, et al. iCHECK-DH: guidelines and checklist for the reporting on digital health implementations. J Med Internet Res 2023;25:e46694 [FREE Full text] [doi: 10.2196/46694] [Medline: 37163336]
54. Ishizaki Y, Gomi H. Human papillomavirus vaccination and postural tachycardia syndrome, deconditioning and exercise-induced hyperalgesia: an alternate interpretation of the reported adverse reactions. J Obstet Gynaecol Res 2020;46(5):678-683. [doi: 10.1111/jog.14227] [Medline: 32153078]
55. Rigney J. HPV vaccination in Japan: the journey to resuming a national immunisation programme. Queen Mary University of London, Cancer Prevention and Screening Blog. 2022 May 27. URL: https://cancerprevention.qmul.ac.uk/index.php/2022/05/27/hpv-vaccination-in-japan-the-journey-to-resuming-a-national-immunisation-programme/ [accessed 2024-10-06]
56. Bodily JM, Tsunoda I, Alexander JS. Scientific evaluation of the court evidence submitted to the 2019 human papillomavirus vaccine libel case and its decision in Japan. Front Med (Lausanne) 2020;7:377 [FREE Full text] [doi: 10.3389/fmed.2020.00377] [Medline: 32850893]
57. Takahiro K. We call for the resumption of active HPV vaccination recommendations. Change.org. 2021. URL: https://tinyurl.com/bdfaf9kp [accessed 2024-10-06]
58. NHK. NHK news record about COVID-19 in time series. NHK NEWS WEB. 2020 Feb 10. URL: https://www3.nhk.or.jp/news/special/coronavirus/chronology/ [accessed 2024-10-06]
59. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
60. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: comparative study. J Med Internet Res 2022;24(10):e40408 [FREE Full text] [doi: 10.2196/40408] [Medline: 36174192]
61. Zhang H, Wheldon C, Tao C, Dunn A, Guo Y, Huo J, et al. How to improve public health via mining social media platforms: a case study of human papillomaviruses (HPV). In: Social Web and Health Research. Cham, Switzerland: Springer International Publishing; 2019:207-231.
62. Adjei Boakye E, Nair M, Abouelella DK, Joseph CLM, Gerend MA, Subramaniam DS, et al. Trends in reasons for human papillomavirus vaccine hesitancy: 2010-2020. Pediatrics 2023;151(6):e2022060410 [FREE Full text] [doi: 10.1542/peds.2022-060410] [Medline: 37218460]
63. ChatGPT. OpenAI. URL: https://chatgpt.com/ [accessed 2024-01-28]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
DL: deep learning
HPV: human papillomavirus
LDA: latent Dirichlet allocation
LLM: large language model
MHLW: Japanese Ministry of Health, Labour and Welfare
NLP: natural language processing
PELT: pruned exact linear time
WHO: World Health Organization

Edited by A Mavragani; submitted 16.11.24; peer-reviewed by A Naz, C Wang; comments to author 05.12.24; revised version received 15.12.24; accepted 31.12.24; published 11.02.25.

Please cite as:
Liu J, Niu Q, Nagai-Tanima M, Aoyama T
Understanding Human Papillomavirus Vaccination Hesitancy in Japan Using Social Media: Content Analysis
J Med Internet Res 2025;27:e68881
URL: https://www.jmir.org/2025/1/e68881
doi:10.2196/68881
PMID:

XSL•FO
RenderX

©Junyu Liu, Qian Niu, Momoko Nagai-Tanima, Tomoki Aoyama. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Traversing Shifting Sands—the Challenges of Caring for Someone With Alzheimer's Disease and the Impact on Care Partners: Social Media Content Analysis

Kristian Steen Frederiksen[1,2], MD, PhD; Julie Hahn-Pedersen[3], MSc; Rebecca Crawford[4], MA; Ross Morrison[4], MPH; Rose Jeppesen[3], MD, PhD; Lynda Doward[4], MRES; Wendy Weidner[5], MA

[1]Department of Neurology, Danish Dementia Research Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

[2]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]Novo Nordisk A/S, Soeborg, Denmark

[4]RTI Health Solutions, Manchester, United Kingdom

[5]Alzheimer's Disease International, London, United Kingdom

Corresponding Author:
Rebecca Crawford, MA
RTI Health Solutions
The Pavilion, Towers Business Park
Wilmslow Road, Didsbury
Manchester, M20 2LS
United Kingdom
Phone: 44 161 447 6000
Email: rcrawford@rti.org

Abstract

Background: Social media data provide a valuable opportunity to explore the effects that Alzheimer disease (AD) has on care partners, including the aspects of providing care that have the greatest impacts on their lives and well-being and their priorities for their loved ones' treatment.

Objective: The objective of this social media review was to gain insight into the impact of caring for someone with AD, focusing particularly on impacts on psychological and emotional well-being, social functioning, daily life and ability to work, health-related quality of life, social functioning, and relationships.

Methods: We reviewed social media posts from 4 sources—YouTube (Google), Alzheimer's Association, Alzheimer Society of Canada, and Dementia UK—to gain insights into the impact of AD on care partners. English-language posts uploaded between May 2011 and May 2021 that discussed the impact of AD on care partners were included and analyzed thematically.

Results: Of the 279 posts identified, 55 posts, shared by 70 contributors (4 people living with AD and 66 care partners or family members), met the review criteria. The top 3 reported or observed impacts of AD discussed by contributors were psychological and emotional well-being (53/70, 76%), social life and relationships (37/70, 53%), and care partner overall health-related quality of life (27/70, 39%). An important theme that emerged was the emotional distress and sadness (24/70, 34%) associated with the care partners' experience of "living bereavement" or "anticipatory grief." Contributors also reported impacts on care partners' daily life (9/70, 13%) and work and employment (8/70, 11%). Care partners' emotional distress was also exacerbated by loved ones' AD-related symptoms (eg, altered behavior and memory loss). Caregiving had long-term consequences for care partners, including diminished personal well-being, family and personal sacrifices, loss of employment, and unanticipated financial burdens.

Conclusions: Insights from social media emphasized the psychological, emotional, professional, and financial impacts on individuals providing informal care for a person with AD and the need for improved care partner support. A comprehensive understanding of care partners' experiences is needed to capture the true impact of AD.

(J Med Internet Res 2025;27:e55468) doi:10.2196/55468

KEYWORDS

Alzheimer disease; caregiver; burden; health-related quality of life; social media

XSL•FO
RenderX

Introduction

Alzheimer disease (AD) is the most common form of dementia [1]. Worldwide, an estimated 55 million people were living with AD in 2019, and by 2050 a projected 139 million people will be living with AD [2]. Common symptoms of AD include memory impairment, disorientation, loss of function, and mood or behavior changes [3]. In the final and most severe stage of dementia, cognitive and functional abilities are significantly impaired, resulting in an increasing need for care and support [4]. Most individuals caring for a person living with AD are informal care partners (ie, a relative or a friend), and globally more than 70% of informal care hours are provided by women [5].

Caring for someone with AD can impact on care partners' health-related quality of life (HRQOL) and emotional well-being [6-10]. Care partners experience an increased burden and more distress as AD severity increases [11-14] and may experience a greater burden than care partners for individuals with other chronic diseases [15]. There is also evidence that care partners experience greater activity impairment, absenteeism (ie, absence from work), impairment related to presenteeism (ie, reduced productivity while at work), and overall work impairment than individuals who are not care partners [16]. Effective disease-modifying treatments for AD are beginning to emerge, representing the potential to address the unmet needs for those living with AD and their care partners. To inform treatment goals in AD and fully characterize the potential value of emerging treatments, it is important to understand the practical, emotional, and financial impacts of the disease, particularly on loved ones providing informal care.

Previous research has evaluated the indirect cost implications of providing informal care to people living with AD or evaluated the caregiving burden using clinical outcome assessments [5,6,16,17]. However, these evaluations do not represent the lived experiences of care partners, the aspects of providing care that have the greatest impacts on their lives and well-being, or their priorities for their loved ones' treatment. Social media data provide a valuable opportunity to explore the impacts that AD has on care partners outside the formal research context. The objective of this study was to gain insight into the impact of caring for someone with AD, with a particular focus on impacts to daily activities, psychological and emotional well-being, HRQOL, work and finances, social functioning, and relationships.

Methods

We conducted an iterative pragmatic review of publicly available social posts shared on global platforms (eg, YouTube [Google]), as well as advocacy websites and independent blogs and forums of interest, with relevant data on the care partner experience of AD.

Data Identification and Collection

The following parameters were used to guide the manual identification and collection of social media data:

1. Date range: A historical search of social media posts sourced from May 2011 to May 2021.

2. Language: English language only.

3. Media types: YouTube, news blogs, patient organizations, forums, and comments.

4. Different combinations of keywords were used to optimize the approach for identifying social media posts relevant to the project objectives. Key search terms and potential search strategies are presented in Table 1.

Table 1. Social media review key search terms and potential search strategies.

Concept	Key search terms	Example potential search strategy [a]
Care partner experience	"caregiver", "care-partner", "carer", "family*", "husband", "wife", "story", "journey", "narrative", "experience", "diary", "blog", "forum"	(caregiver OR care-partner OR carer OR family OR husband OR wife) AND (story OR journey OR narrative OR experience OR diary OR blog OR forum)
AD[b]	"Alzheimer*", "dementia", "senile", "mild cognitive impairment", "early onset dementia", "cognitive dysfunction", "neurocognitive disorder", "cognitive impairment", "memory clinic"	(Alzheimer* OR dementia OR senile OR mild cognitive impairment OR early onset dementia OR cognitive dysfunction OR neurocognitive disorder OR cognitive impairment)
Impact of AD on care partner	"impact", "burden", "effect", "quality of life", "family life", "family activities", "relationships", "health", "money", "money worries", "expense", "expensive", "cost"	(impact OR burden OR effect OR quality of life OR family life OR family activities OR relationships)

[a]Boolean operators were used where supported by search function.

[b]AD: Alzheimer disease.

In addition to YouTube, a total of 6 advocacy websites were identified as containing potentially relevant data; the terms and conditions of each website were reviewed. Only publicly accessible videos and comments were reviewed from YouTube; that is, no videos that required logging into YouTube or required the researcher to subscribe were collated. Each of the advocacy websites was contacted to ask for permission to include the data posted on their website in the review. In total, 3 advocacy websites granted permission, and 3 either refused permission or were unresponsive to the request. Only advocacy websites that granted permission to be included in the review—Alzheimer's Association, Alzheimer Society of Canada, and Dementia UK—were accessed.

The target population for review included people with a self-reported diagnosis of AD and care partners or family members of people with a self-reported diagnosis of AD. Video footage and discussion blogs were manually reviewed by

XSL•FO
RenderX

experienced qualitative researchers to determine eligibility for inclusion in the review. The prespecified eligibility criteria for the social media posts are presented in Textbox 1.

Textbox 1. Social media post inclusion criteria.

- The contributor has a self-reported diagnosis of mild cognitive impairment or dementia due to Alzheimer disease (AD) or is a self-reported carer or care partner or a family member of someone with mild cognitive impairment or dementia due to AD.
- At the time of the post, the contributor was an adult (aged 18 years or older). Where age was not reported, the contributor's adult status was researcher-determined based on post content and video images.
- The post is relevant to the contributor's experience of caring for someone with AD, including the impact of their AD treatments or management.
- Posts are in the English language.
- Post was uploaded between May 2011 and May 2021.

Data Extraction

The social media posts (ie, videos and comments posted on YouTube and patient stories posted on the 3 selected advocacy websites) were reviewed manually by experienced qualitative researchers, and relevant social media data were extracted and recorded within a Microsoft Excel spreadsheet for analysis. Each post, blog, video, or comment included in the review was given a unique source identification number, and each contributor was assigned a unique contributor identifier. An initial deductive approach used a top-level coding framework to organize the extracted social media data by key themes, such as symptoms, disease history, HRQOL impacts, and treatment experience. Throughout the data extraction process, the research team met regularly to review and reflect on the data extracted. These points of researcher reflection included cross-checking contributor characteristics to ensure that any contributors who had posted across multiple platforms had not been double counted and only new information had been extracted in the second record, as well as a topline quality review of the data extracted. These sessions also allowed for the researchers to harmonize on the data extraction process.

Data Analysis

Once all the social media data had been extracted and categorized using the top-level framework, the data were reviewed qualitatively by 2 researchers (RC and RM). A reflexive thematic analysis [18] of the aggregated social media data extracted from the posts was conducted, in which a theme was described as content that captured data relevant to the research objectives. Specifically, a deductive approach was initially used to code the social media data to address the study objectives, and descriptive codes were used to capture the concepts of interest. The coded data were then inductively analyzed within and across each category, with new codes added as concepts emerged from the data. This "bottom-up" approach is synergetic with the unsolicited and unregulated nature of social media data, as the approach allows the data to direct the analysis and the identification of new and important findings not predetermined by the research scope. Concepts were then used to generate patterns in the way that care partners described their experiences and perceptions related to looking after someone with AD. Patterns that emerged within and across the data were used to develop themes, with care partner experiences and perceptions summarized within each key theme.

The analysis approach and findings were reviewed by a third independent researcher (LD), and the interpretation of the social media data and the identification of themes and concepts were discussed to ensure the confirmability of the results. Once the themes and concepts were agreed, where appropriate, descriptive summaries were accompanied by frequency counts (and percentages) of the number of contributors discussing a specific theme.

Demographic information (eg, gender and age) was extracted, along with any accompanying disease information, where available. Notably, because social media data exist outside the research context, key demographic and diagnostic characteristics were not always available for contributors.

Ethical Considerations

This study was submitted to RTI International's institutional review board committees for ethical approval and was determined not to constitute research involving human participants (STUDY00022118). No informed consent was obtained, as the data analyzed in the study are publicly available. No personally identifiable information was included in the analyses. Modified quotations (supporting statements) are used to illustrate the key themes that emerged from the social media data. Modified quotations are used to ensure that the anonymity of the contributor is maintained, and their identity is protected. There were no participants to be compensated, as this study did not constitute research involving human participants.

Results

Post and Contributor Characteristics

Of the 279 social media posts identified, 55 met the review criteria (21 blog posts, 21 videos, and 13 comments) (Figure 1); the remaining posts were excluded from the review because they did not meet the criteria for inclusion (eg, no relevant data included within the post, outside the prespecified date range). The 55 posts were shared by 70 contributors (4 people living with AD and 66 care partners or family members) who discussed self-reported or observed impacts of AD on care partners and family members. Table 2 provides the gender and age of contributors by contributor type (ie, care partner or family member and person living with AD) and for the total sample. Contributor age at the time of the social media post was not widely available; 90% (63/70) of contributors did not report their age in the social media post.

XSL•FO
RenderX

Figure 1. Social media post identification flow chart. AD: Alzheimer disease.

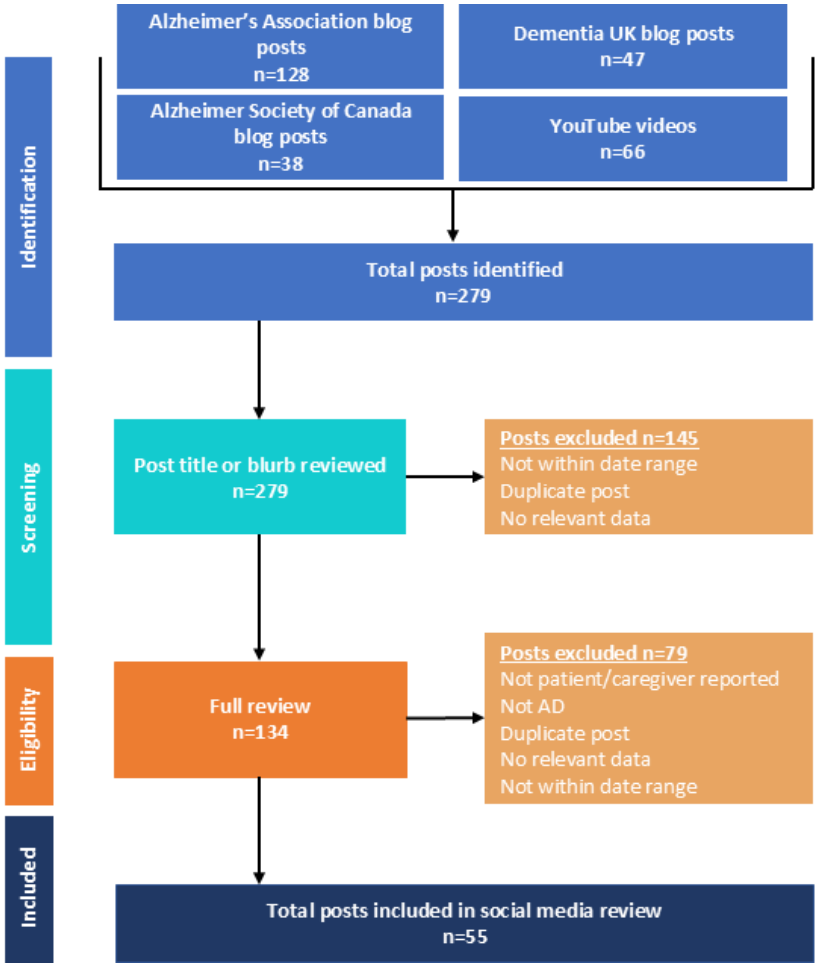

Table 2. Summary of social media contributor sample characteristics (the target population for review included care partners or family members of people with a self-reported diagnosis of Alzheimer disease; however, some contributors were individuals with Alzheimer disease who commented on care partners' experiences. These posts were included in the review. Percentages are based on nonmissing data).

Characteristics	Contributor type		Total contributor sample (N=70)
	Person with AD[a] (n=4[b])	Care partner or family member (n=66[b])	
Self-reported gender, n (%)			
Women	2 (50)	40 (68)	42 (67)
Men	2 (50)	19 (32)	21 (33)
Missing	0[c]	7[c]	7[c]
Age at social media post, n (%)			
<65 years	1 (100)	4 (67)	5 (71)
≥65 years	0 (0)	2 (33)	2 (29)
Missing	3[c]	60[c]	63[c]
Age range (reported sample)	57	21-73	21-73

[a]AD: Alzheimer disease.

[b]The contributor who self-identified as both a person living with Alzheimer disease and a care partner is included in the care partner or family member contributor cohort.

[c]Percentages are based on nonmissing data.

XSL•FO
RenderX

Impact of AD on Care Partners

A range of impacts of AD on care partners and family members was discussed in the social media posts shared by the 70 contributors. The 3 most frequently reported or observed impacts of AD on care partners and family members were noted in relation to care partners' psychological and emotional well-being (53/70, 76%), their social life and relationships (37/70, 53%), and their overall HRQOL (27/70, 39%). Contributors also reported impacts on care partners' daily life (9/70, 13%), work and employment (8/70, 11%), and physical health (5/70, 7%). The key impacts of caring for someone with AD on care partners and family members that emerged from the social media data are summarized in Figure 2.

Figure 2. Key impacts of Alzheimer disease on care partners and family members reported by a person living with Alzheimer disease, care partner, or family member (n=70). AD: Alzheimer disease; ADL: activity of daily living; QOL: quality of life.

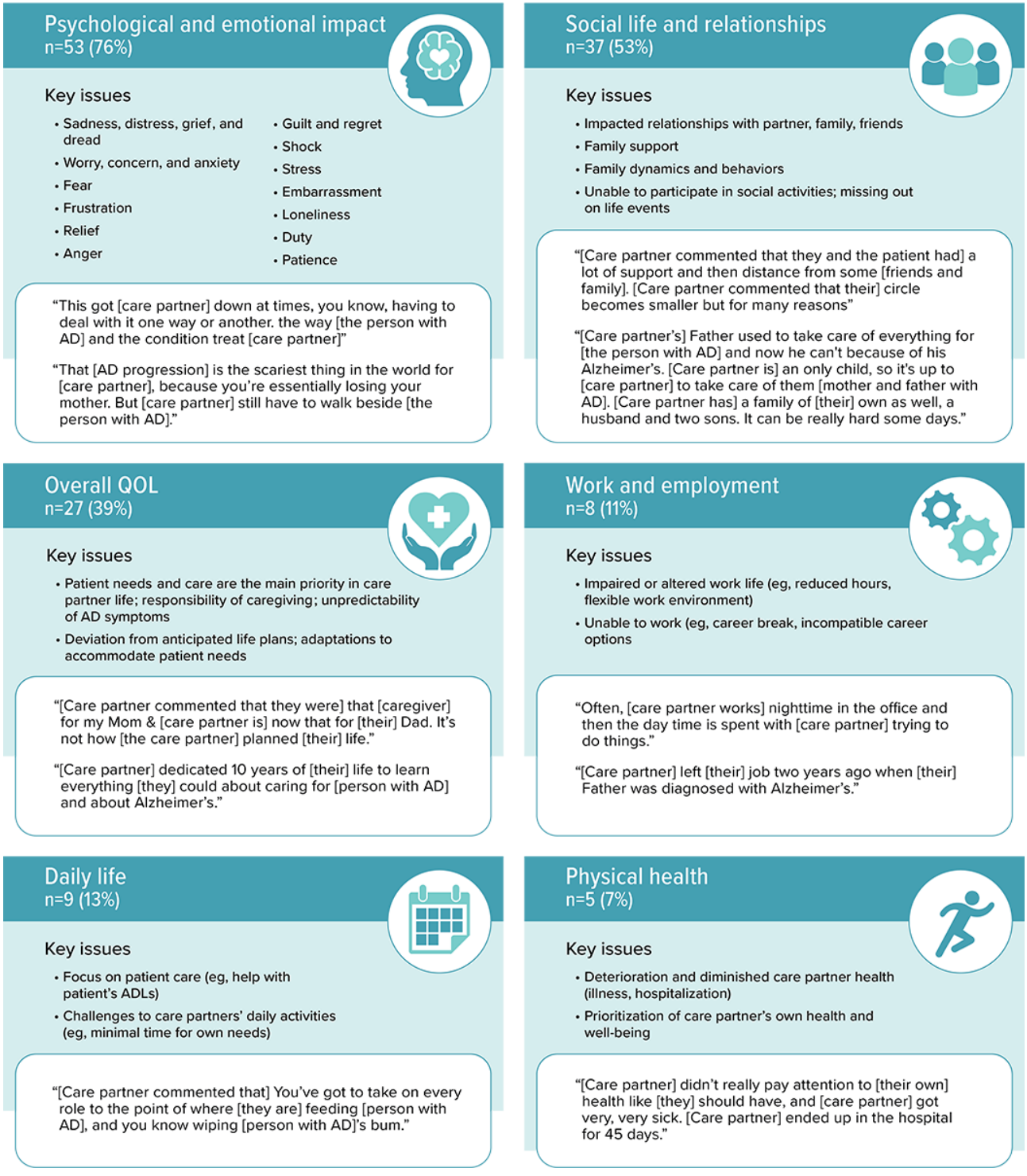

XSL•FO
RenderX

Emotional and Social Impacts

Care partner or family member contributors discussed a range of emotions experienced in relation to caring for someone with AD. Contributors noted feelings of frustration, concern, and fear in relation to the progressive symptoms of AD (eg, memory loss and delusions) and subsequent changes in behavior (eg, repetitive questions and aggression). The behavioral symptoms of people living with AD had a substantial impact on care partners and family members; 6 out of 70 (9%) care partner or family member contributors reported that the person living with AD had become aggressive and displayed violent behavior toward them. Furthermore, 1 contributor noted the need for patience when caring for someone with AD, and another highlighted that the primary care partner experience was a lonely existence regardless of support from family and friends.

The most profound impact on the psychological and emotional well-being of care partners and family members was the emotional distress and sadness (24/70, 34%) associated with the phenomenon of "living bereavement" or "anticipatory grief," that is, grieving the loss of a person even before their physical body dies [19]. The deterioration of the memories and personhood for the person living with AD was a traumatic and heartbreaking experience for care partners and family members, who had to witness the gradual loss of their loved one. Several contributors noted that, while the person living with AD was still alive, they were grieving their loss due to the significant changes in the person's mental health as well as their cognitive and physical functioning. This was particularly salient for care partners and family members of people with early-onset AD, who "lost" their loved one earlier than anticipated; one family member described feeling that his mother had been "missing" for most of his life. The final stages of AD were also distressing for care partners and family members, who described the person living with AD as a "shell" that was simply waiting to "shut down." Their loved one's death presented care partners and family members with contrasting emotions—relief that their duty of care was now complete but sadness because the last vestiges of the person were gone. One care partner noted that the grief they experienced once their loved one had died was different from the "living grief" they had endured while the person had been alive.

Over 50% of the contributors (37/70, 53%) noted that caring for someone with AD had an impact on the social life and relationships of the care partner or family member. The impact of AD on the relationship between the care partner and the person living with AD was discussed by 9 out of 70 (13%) care partners or family member contributors, specifically the strain on or deterioration of this relationship. For care partners of parents with AD, this entailed a loss of their parental relationship and its associated support system, whereas for care partners of spouses with AD, this meant the loss of intimacy with their partner. Another key theme was the impact of care responsibilities on the relationships between the care partner or family member and their wider family (9/70, 13%). The duty of care for the person living with AD took precedence in family life; contributors noted that they would have to "sacrifice" their own family needs to accommodate their care partner obligations. However, care partner or family member contributors stressed the importance of family support (8/70, 11%) for the care partner, particularly shared family responsibility for care for the person living with AD, which helped to strengthen familial bonds. Changes in family dynamics were another component that emerged from the social media data in relation to the impact of AD on the social life and relationships of care partners and family members (6/70, 9%). Care partners and family members experienced a role change within the family unit, and family members were expected to assume additional responsibilities to adjust to the new family "norm." Care partners and family members discussed the impact that their care responsibilities had on their social life (2/70, 3%). One care partner commented on changes to their relationships with friends; although some friends provided support, others did not; ultimately, their social circle decreased in size.

Impact on Daily Life and Ability to Work

Out of 70, 9 (13%) contributors commented on the detrimental impacts on care partners' and family members' daily lives because of the responsibility of caring for a person living with AD. Direct responsibility for the self-care needs of the person living with AD (eg, feeding, bathing, and toileting) was noted by 5 (7%) contributors, whereas 1 (1%) contributor commented that they provided assistance to the person living with AD to enable the individual to perform their own daily activities. Contributors also reflected on the challenges experienced by care partners and family members to perform their own daily activities (2/70, 3%), including interference resulting from the person living with AD's cognitive symptoms (eg, forgetting where they had moved objects) or the minimal time available for care partners and family members to focus on their own needs.

Contributors commented on the disruption that caring for a person living with AD had on their ability to work (8/70, 11%). Half of these contributors (n=4) noted that they had to stop working or were unable to work because of their care partner duties. For others (n=4), it necessitated a change to the structure of their work life, such as reduced working hours, working from home, less work-related travel, and flexible working hours to accommodate the care needs of the person living with AD. Although some employers were able to accommodate the need for a flexible work-life approach, this was not always possible. Care partners noted that there were limited career choices that would permit the flexibility needed to provide care.

Impact on Well-Being and HRQOL

The stress or burden of caring for a person living with AD was attributed to a deterioration in the care partners' own physical health (3/70, 4%). The prioritization of the person living with AD's well-being resulted in cases of care partner illness, hospitalizations, and diminished well-being (2/70, 3%). However, 2 care partners out of 70 (3%) recognized the importance of prioritizing their own health, as well as the health of the person living with AD, and made positive lifestyle changes (eg, increased exercise and reduced alcohol consumption) to maintain and improve their physical health. This positive approach was considered by 1 care partner out of 70 (1%) to be the most effective plan to provide long-term care for her husband.

XSL•FO
RenderX

Care partners' and family member contributors' overall HRQOL and anticipated future life plans were hindered by their responsibility as a care partner and the needs of the person living with AD. Care partner and family member contributors (9/70, 13%) noted that the responsibility of caring for a person living with AD was multifaceted as it required balancing their care responsibilities with their family roles and obligations, arranging additional support or developing a contingency plan if they were incapacitated for any reason, and handling the logistical duties associated with the person living with AD's end of life (eg, reviewing wills, insurance policies, and mortgages). Caring for the needs of the person living with AD became the main priority in the care partner's life (6/70, 9%) and, according to some contributors, required complete dedication "24 hours, 7 days a week." Furthermore, 1 care partner contributor noted that she had dedicated 10 years of her life to self-education on AD and caring for her partner. However, the progressive nature of AD and the unpredictability of AD symptoms (6/70, 9%) meant that care partners had to continuously adapt to the changes in the behavior and symptoms of the person living with AD. The need to adapt to the person living with AD's circumstance was highlighted by 4 contributors out of 70 (6%), who noted that they had made significant changes to their lives (eg, relocation and renovation of existing home) to accommodate the needs of the person living with AD. Ultimately, providing care for someone with AD had long-term consequences on the lives of care partners and family members; out of 70, 4 (6%) contributors discussed how their lives had deviated from their anticipated life path (eg, career change and moving back in with parent).

Care Partners' Experiences of AD Treatment and Management

Care partners and family member contributors discussed treatment hopes or expectations (3/70, 4%). The potential development of a cure for AD was appealing for contributors, as was the potential for a novel treatment that could slow disease progression. In addition, 2 care partner contributors commented on the importance of involving the person living with AD in treatment decisions. Contributors' discussions pertaining to decision-making in AD management focused predominantly on the factors that premeditated the decision by care partners and family members as to whether the person living with AD required formal full-time medical care at inpatient care facilities (6/70, 9%). Key factors discussed in the difficult decision to transfer an individual with AD from informal "at-home care" to formal "inpatient care" included crisis events or the realization that the needs of the person living with AD and the disease progression could be accommodated only with inpatient facility care (4/70, 6%).

Among the 13 contributors out of 70 (19%) who commented on the burden on care partners and family members associated with AD treatments or management, three key themes emerged, such as (1) limited support for care partners (6/70, 9%), (2) financial challenges (5/70, 7%), and (3) travel burden (3/70, 4%). Contributors commented on the challenges associated with navigating different services and health care systems (eg, difficulties accessing treatments or management options and benefits); 3 contributors (4%) lamented the absence of a single point of contact for assistance with access to treatment or management information and support. For these contributors, the process of navigating the health care system was time-consuming and created additional stress for themselves and the wider family. This process was further complicated by the contributor-perceived limited support for the diverse population of people living with AD, such as non–English-speaking individuals, people with early-onset AD, and people with AD who live in rural areas. Difficulties experienced with navigating or accessing AD-specific care services were also a factor contributing to the substantial financial responsibility assumed by care partners and family members. Contributors noted that they had not anticipated or planned for the high costs associated with the care needs of a person living with AD (ie, full-time care, memory care facilities, and assisted living). Furthermore, 1 care partner highlighted the difficult choice between becoming a full-time carer for the person living with AD and having the ability to meet that person's financial needs. Another aspect of the burden of AD treatment or management on care partners was the considerable time commitment required to accompany the person living with AD to their frequent medical appointments.

Discussion

Principal Findings

The social media data collected in this study yielded insights into the impact to care partners and family members of providing care for people living with AD and the management of the disease. While positive elements associated with providing care for a person living with AD were discussed in the social media data, the posts emphasized in particular the concept of "living bereavement" or "anticipatory grief" for care partners and with that the need for improved coping strategies and interventions to enable care partners to better manage this phenomenon. Anticipatory grief permeates the experience of the care partner, which can involve decades of providing care for the person living with AD [19]. Accordingly, previous research has highlighted that it is important to consider anticipatory grief in conceptual models of care partner burden to help clarify how different stressors influence individual care partner outcomes and to enhance the understanding of the care partner experience [20].

Care partners' and family members' candid posts also highlighted the psychological, social, and financial impairments associated with becoming a care partner. Contributors reflected on the role of caring for a person living with AD as a full-time commitment and responsibility in which the needs of the person living with AD were often prioritized to the detriment of the care partner's own needs, family life, and well-being. Providing care could also result in potentially daunting financial challenges and concerns.

Care partners' treatment hopes and expectations, as expressed in the social media posts, focused predominantly on the possibility of a curative treatment in the future but also on the immediate need for treatments that could slow or prevent disease progression. Care decisions focused principally on the difficult and often distressing determination by care partners or family members that the person living with AD was now at a stage

XSL•FO
RenderX

where they would benefit from formal full-time medical care at an inpatient care facility. AD treatments and management strategies had a substantial burden on care partners and family members; contributors noted limited support available for care partners and the financial and travel challenges associated with the management and treatment needs of a person living with AD.

Findings from the social media review are broadly consistent with the published evidence on the experiences of care partners in AD. Notably, studies conducted in the United States, Brazil, China, and Japan have found that care partners of people living with AD experience depression, anxiety, and high levels of stress [7-10]. For example, a study evaluating the experiences of 125 US care partners and 60 demographically similar control participants (non–care partners) found care partners more likely than control participants to experience clinically significant depression (40% vs 5%), and 26% of care partners used antidepressants [7]. Our results provide additional and unique insights into the individual contributor experiences underlying the emotional impacts of providing informal care for a loved one with AD.

Of note, it has been estimated that time spent providing informal care for people with AD amounted to 82 billion hours in 2015, equivalent to more than 40 million full-time workers, and this number is projected to increase to the equivalent of 65 million full-time workers by 2030 [5]. Further, a majority of informal care partners in AD are women [5], who may experience disproportionate impacts to their professional lives and opportunities as they manage their care responsibilities. Insights from social media contributors complement this evidence by providing rich data on care partners' lived experiences and the impacts they experience beyond time spent caregiving and indirect impacts to their work productivity.

Care partner contributors also expressed frustrations with navigating the health care system, in particular with the lack of a single point of contact for access to treatment. This consistent unmet need for people living with AD and their care partners is well-recognized in the AD community. Improving the AD care journey for individuals living with AD and their care partners was a strategic recommendation in the Alzheimer's Disease International's 2022 World Alzheimer's Report, but implementation is still limited in many countries [2].

The emergence of novel therapies offers the promise of slowing or halting disease progression for people living with AD. To comprehensively evaluate the value of novel therapies, it is critical to first document the magnitude of the societal impact of AD; however, the qualitative experiences and perspectives of informal care partners in AD have been underrecognized and can be challenging to measure via traditional research methods. Insights from this social media review, which is underpinned by published evidence on the care partner experience, provide unsolicited insights into care partners' experiences with the practical, emotional, and financial challenges they face, and provide directions for future research to explore these impacts through formal research channels. Ultimately, as the therapeutic landscape evolves, the perspectives of care partners, as well as people living with AD, should be integrated into value assessment frameworks in AD [21].

This social media review provides a unique approach to exploring the experiences and challenges associated with caring for people living with AD potentially not captured through traditional research methods. Nonetheless, limitations of the methodology are noted. First, social media data are not driven by any specific research question or objective; thus, they exist outside the formal research context. There is an inherent reliance on contributor self-identification and self-reported diagnoses that are not verifiable. Social media data were retrospectively collected from posts in the public domain; information pertaining to contributor demographic characteristics, location information and access to care, and the clinical characteristics of their loved one with AD or the level of insight about their disease was not consistently available. Limitations of sample size must be considered; some of the themes identified from the social posts are based on a small number of contributors, and the degree to which the theme is relevant to the wider populations of people living with AD and care partners could be challenged. Different platforms are used by different demographic groups, and there is a potential bias toward populations who are more likely to use the specific sites selected for the review, more likely to reside in the included countries, or more likely to be engaged and proficient with digital technologies. Furthermore, only English-language posts were reviewed. Finally, there is the potential for self-selection and publication bias.

Conclusions

The candid social media posts from 70 contributors provide valuable insights on the responsibilities and complications of caring for someone with AD, as well as the shortcomings of current AD programs to sufficiently support care partners in successfully balancing the dichotomy of their caregiving role and their own personal needs. It is critical to understand the true impact on care partners of providing informal care—including difficulties with their health and well-being, emotional and social functioning, and professional lives—and the unmet needs that they face. With this understanding, we may begin to define AD care models to support care partners as they navigate the AD care journey with their loved one. Furthermore, characterizing the care partner experience positions the AD research community to define and measure therapeutic goals for novel treatments that not only allow people living with AD to maintain their cognitive functioning and daily activities but also enable care partners to balance their informal care responsibilities with their well-being and professional responsibilities.

XSL•FO
RenderX

Acknowledgments

This study was funded by Novo Nordisk. Authors affiliated with the study sponsor, Novo Nordisk, participated in the design and conduct of the study; analysis and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

Kate Lothman of RTI Health Solutions provided medical writing services, which were funded by Novo Nordisk.

Data Availability

The data analyzed during this study are included in this published article.

Authors' Contributions

JH-P, RC, RM, and LD contributed to the study concept and design. KSF, JH-P, RC, RM, RJ, LD, and WW contributed to the analysis and interpretation of data. KSF, JH-P, RC, RM, RJ, LD, and WW contributed to the critical revision of the paper for important intellectual content. RC, RM, and LD conducted the analysis. JHP obtained funding. RC, JH-P, and LD supervised the study. All authors read and approved the final manuscript.

Conflicts of Interest

This study was funded under a research contract between Novo Nordisk and RTI Health Solutions and was funded by Novo Nordisk. JH-P and RJ are employees of Novo Nordisk, the sponsor of this study. RC and LD are employees of RTI Health Solutions; RM was an employee of RTI Health Solutions at the time of the study. KSF serves on an advisory board for Novo Nordisk, for which he receives no personal remuneration. WW plays an occasional advisory role for Novo Nordisk, for which she receives no personal remuneration.

References

1. Dementia Fact Sheet. World Health Organisation. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2022-10-27]
2. Gauthier S, Servaes S, Morais JA, Webster C, Rosa-Neto P. Alzheimer?s disease international. World Alzheimer Report 2022: Life after diagnosis: Navigating treatment, care and support. 2023. URL: https://www.alzint.org/u/World-Alzheimer-Report-2022.pdf [accessed 2023-05-31]
3. Alzheimer?s Association. 2021 Alzheimer's disease facts and figures. Alzheimers Dement 2021;17(3):327-406. [doi: 10.1002/alz.12328] [Medline: 33756057]
4. 2021b. What is Alzheimer?s disease? Alzheimer?s Association. URL: https://www.alz.org/alzheimers-dementia/what-is-alzheimers [accessed 2022-06-09]
5. Global estimates of informal care.. Wimo A, Gauthier S. Wimo A, Gauthier S, Prince M; Alzheimer?s Disease International and Karolinska Insitutet. 2018. URL: https://www.alzint.org/u/global-estimates-of-informal-care.pdf [accessed 2023-09-19]
6. Liao X, Huang Y, Zhang Z, Zhong S, Xie G, Wang L, et al. Factors associated with health-related quality of life among family caregivers of people with Alzheimer's disease. Psychogeriatrics 2020;20(4):398-405. [doi: 10.1111/psyg.12528] [Medline: 32027776]
7. Mausbach BT, Chattillion EA, Roepke SK, Patterson T, Grant I. A comparison of psychosocial outcomes in elderly Alzheimer caregivers and noncaregivers. American Journal of Geriatric Psychiatry 2012:1. [doi: 10.1097/jgp.0b013e31824472d5]
8. Corrêa MS, Vedovelli K, Giacobbo B, de Souza C, Ferrari P, de Lima Argimon I, et al. Psychophysiological correlates of cognitive deficits in family caregivers of patients with Alzheimer disease. Neuroscience 2015;286:371-382. [doi: 10.1016/j.neuroscience.2014.11.052] [Medline: 25490073]
9. Liu S, Li C, Shi Z, Wang X, Zhou Y, Liu S, et al. Caregiver burden and prevalence of depression, anxiety and sleep disturbances in Alzheimer's disease caregivers in China. J Clin Nurs 2017;26(9-10):1291-1300. [doi: 10.1111/jocn.13601] [Medline: 27681477]
10. Ohno S, Chen Y, Sakamaki H, Matsumaru N, Yoshino M, Tsukamoto K. Burden of caring for Alzheimer's disease or dementia patients in Japan, the US, and EU: results from the National Health and Wellness Survey: a cross-sectional survey. J Med Econ 2021;24(1):266-278 [FREE Full text] [doi: 10.1080/13696998.2021.1880801] [Medline: 33538195]
11. Pudelewicz A, Talarska D, Bączyk G. Burden of caregivers of patients with Alzheimer's disease. Scand J Caring Sci 2019;33(2):336-341. [doi: 10.1111/scs.12626] [Medline: 30378698]
12. Sinha P, Desai NG, Prakash O, Kushwaha S, Tripathi CB. Caregiver burden in Alzheimer-type dementia and psychosis: a comparative study from India. Asian J Psychiatr 2017;26:86-91. [doi: 10.1016/j.ajp.2017.01.002] [Medline: 28483100]
13. Reed C, Barrett A, Lebrec J, Dodel R, Jones RW, Vellas B, et al. How useful is the EQ-5D in assessing the impact of caring for people with Alzheimer's disease? Health Qual Life Outcomes 2017;15(1):16 [FREE Full text] [doi: 10.1186/s12955-017-0591-2] [Medline: 28109287]

XSL•FO
RenderX

14. Storti LB, Quintino DT, Silva NM, Kusumota L, Marques S. Neuropsychiatric symptoms of the elderly with Alzheimer's disease and the family caregivers' distress. Rev Lat Am Enfermagem 2016;24:e2751 [FREE Full text] [doi: 10.1590/1518-8345.0580.2751] [Medline: 27533264]
15. Demirbas M, Hahn-Pedersen JH, Jørgensen HL. Comparison between burden of care partners of individuals with Alzheimer's disease versus individuals with other chronic diseases. Neurol Ther 2023;12(4):1051-1068 [FREE Full text] [doi: 10.1007/s40120-023-00493-6] [Medline: 37222859]
16. Goren A, Montgomery W, Kahle-Wrobleski K, Nakamura T, Ueda K. Impact of caring for persons with Alzheimer's disease or dementia on caregivers' health outcomes: findings from a community based survey in Japan. BMC Geriatr 2016;16(1):122 [FREE Full text] [doi: 10.1186/s12877-016-0298-y] [Medline: 27287238]
17. Duggleby WD, Swindle J, Peacock S, Ghosh S. A mixed methods study of hope, transitions, and quality of life in family caregivers of persons with Alzheimer's disease. BMC Geriatr 2011;11(1):88 [FREE Full text] [doi: 10.1186/1471-2318-11-88] [Medline: 22192235]
18. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
19. Frank JB. Evidence for grief as the major barrier faced by Alzheimer caregivers: a qualitative analysis. Am J Alzheimers Dis Other Demen 2007;22(6):516-527 [FREE Full text] [doi: 10.1177/1533317507307787] [Medline: 18166611]
20. Garand L, Lingler JH, Deardorf KE, DeKosky ST, Schulz R, Reynolds CF, et al. Anticipatory grief in new family caregivers of persons with mild cognitive impairment and dementia. Alzheimer Dis Assoc Disord 2012;26(2):159-165 [FREE Full text] [doi: 10.1097/WAD.0b013e31822f9051] [Medline: 21946013]
21. Jessen F, Georges J, Wortmann M, Benham-Hermetz S. What matters to patients with Alzheimer's disease and their care partners? Implications for understanding the value of future interventions. J Prev Alzheimers Dis 2022;9(3):550-555. [doi: 10.14283/jpad.2022.22] [Medline: 35841256]

Abbreviations

AD: Alzheimer disease
HRQOL: health-related quality of life

Edited by T de Azevedo Cardoso; submitted 13.12.23; peer-reviewed by W Collinge, C Abbatantuono; comments to author 06.06.24; revised version received 24.07.24; accepted 15.11.24; published 18.02.25.

Please cite as:
Frederiksen KS, Hahn-Pedersen J, Crawford R, Morrison R, Jeppesen R, Doward L, Weidner W
Traversing Shifting Sands—the Challenges of Caring for Someone With Alzheimer's Disease and the Impact on Care Partners: Social Media Content Analysis
J Med Internet Res 2025;27:e55468
URL: https://www.jmir.org/2025/1/e55468
doi:10.2196/55468
PMID:

©Kristian Steen Frederiksen, Julie Hahn-Pedersen, Rebecca Crawford, Ross Morrison, Rose Jeppesen, Lynda Doward, Wendy Weidner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Challenges and Emotions of Informal Caregivers of General Older Adults and People With Alzheimer Disease and Related Dementia: Comparative Study

Nova Mengxia Huang[1], MA; Liang Ze Wong[2], PhD; Shirley S Ho[1], PhD; Bryan Timothy[3], BA

[1]Nanyang Technological University, Wee Kim Wee School of Communication and Information, Singapore, Singapore

[2]Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

[3]National University of Singapore, Yong Loo Lin School of Medicine, Centre for Behavioural and Implementation Science Interventions, Singapore, Singapore

Corresponding Author:
Shirley S Ho, PhD
Nanyang Technological University
Wee Kim Wee School of Communication and Information
31 Nanyang Link
Singapore, 637718
Singapore
Phone: 65 6513 8667
Email: tsyho@ntu.edu.sg

Abstract

Background: Faced with multiple challenges, informal caregivers often turn to online support communities for information and support. While scholarly attention has focused on experiences expressed by informal caregivers in these communities, how caregivers' challenges and emotional expressions vary across different health contexts remains understudied.

Objective: We aimed to examine and compare the challenges discussed by informal caregivers of general older adults and those of patients with Alzheimer disease and related dementia, as well as their emotional expressions, on Reddit. In addition, we examined how informal caregivers expressed their emotions in response to various challenges.

Methods: We collected posts from 6 subreddits, including 3 subreddits on caregiving for older adults and 3 on caregiving for patients with Alzheimer disease and related dementia. Using topic modeling, we identified topics discussed by caregivers in the collected posts. We further used deep reading to contextualize these topics and understand the challenges behind them, conducted sentiment analysis to investigate their emotional expressions, and used Spearman rank-order correlation to examine the relationship between the obtained topics and emotions.

Results: In total, 3028 posts were retrieved, including 1552 from older adult–related subreddits and 1476 from Alzheimer disease–related subreddits; 18 key topics were identified, with the most frequent topics being *expressing feelings* (2178/3028, 71.93%) and *seeking advice and support* (1982/3028, 65.46%). Other topics covered various challenges in caregiving, such as *duration of medical care* (1954/3028, 64.53%), *sleep and incontinence* (1536/3028, 50.73%), *financial issues* (1348/3028, 44.52%), and *nursing home* (1221/3028, 40.32%). There was a positive, negligible correlation between *expressing feelings* and *seeking advice and support* (ρ=0.09, *P*<.001). Other topics also showed positive, negligible or weak correlations with these 2 topics but in distinct patterns. Posts from older adult–related subreddits were more focused on practical caregiving issues and *seeking advice and support*, whereas posts from Alzheimer disease–related subreddits emphasized health- and medical-related topics and *expressing feelings*. Caregivers in both contexts predominantly expressed negative emotions (older adults: 1263/1552, 81.38%; Alzheimer disease: 1247/1476, 84.49%), with caregivers in Alzheimer disease–related subreddits exhibiting slightly greater *fear* and *sadness* (*P*<.001). Specific challenges were significantly correlated with negative emotions: *duration of medicalcare* was positively, weakly correlated with *anger* (ρ=0.25, *P*<.001), *fear* (ρ=0.25, *P*<.001), and *sadness* (ρ=0.22, *P*<.001). *Medical appointments* were positively, negligibly correlated with *anger* (ρ=0.10, *P*<.001), *fear* (ρ=0.09, *P*<.001), and *sadness* (ρ=0.06, *P*<.001). *Sleep and incontinence* (ρ=0.14, *P*<.001) and *finances* (ρ=0.24, *P*<.001) were positively, weakly correlated with *anger*.

Conclusions: By identifying the challenges and feelings expressed by caregivers for general older adults and caregivers for patients with Alzheimer disease and related dementia, our findings could inform health practitioners and policy makers in developing more targeted support interventions for informal caregivers in different contexts.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e54847) doi:10.2196/54847

KEYWORDS

informal caregivers; older adults; Alzheimer disease and related dementia; online support communities; Reddit

Introduction

Background

Many countries are faced with the challenge of an aging society. An aging society increases the likelihood that older adults live longer while coping with increased chronic diseases, thus underscoring the need for long-term care services [1,2]. Most older adults receive long-term care within their own residences, often relying on assistance from informal caregivers—family members, friends, and neighbors who provide unpaid care to people with health or functioning needs [1,3]. In the United States, 22.3% of adults provided such care or assistance in the past month, 10.4% of whom provided daily care for people with dementia or other cognitive impairment disorders [4]. Similar patterns are observed in some European countries, such as the United Kingdom and Germany [5].

Given the increasing prevalence of informal caregiving over the coming years, it is important to understand the challenges faced by informal caregivers of older adults [1,2]. Upon assuming the caregiving role, informal caregivers need to handle multiple tasks and responsibilities, including assisting with financial management, assisting with medical decision-making, and providing emotional support for their loved ones [1,2]. These responsibilities often conflict with caregivers' existing roles and personal needs, leading to stressful experiences [1,2]. While caregiving may also bring positive experiences, such as finding satisfaction in caring for loved ones [6] and improving the caregivers' sense of mastery [7,8], the demands often take a heavy toll on the caregivers' health and well-being [9,10].

However, caregiving experiences may vary across different situations and contexts [11-13]. The Caregiver Stress Theory, a stress processing theory that is specifically created to account for variance in the stress experience of caregivers, specifies that factors related to the caregiving context, such as years of providing care, age, and the health status of care recipients, can influence every part of the stress process [10]. For instance, spouse caregivers experienced greater levels of stress caring for their loved ones as compared to adult child caregivers [14]. The theory also outlines 2 sources of caregivers' stress, namely primary and secondary stressors [10]. Primary stressors refer to sources of stress that are directly related to acts of caregiving, such as the challenge from the care recipients' diseases and symptoms and assisting with daily living tasks, and secondary stressors are those indirectly related to caregiving tasks, such as conflicts with other family members and social restrictions [10,15]. This framework thus provides a holistic understanding of the challenges that caregivers face in their caregiving journey and the factors affecting different caregiving experiences [10,15].

Particularly relevant to this study are the experiences of informal caregivers of people with Alzheimer disease and related dementia and general older adults. Compared to caregivers of general older adults, caregivers of people with Alzheimer disease and related dementia are likely to spend more time on caregiving responsibilities, ranging from bathing and feeding to assisting with toileting [16]. In addition, these caregivers perceive higher levels of physical and emotional burdens [17,18] and experience increased depression rates [16,19]. These differences highlight the importance of understanding caregiving experiences by considering the health status of care recipients [11,13]. Hence, this study aims to examine and compare the challenges faced by informal caregivers of people with Alzheimer disease and related dementia and those taking care of general older adults to advance the understanding of the factors affecting different caregiving experiences.

Informal Caregivers in Online Support Communities

The rise of online support communities (OSCs) has provided informal caregivers of older adults with an avenue to connect and communicate with similar others [20,21]. These OSCs are embedded across computer-mediated platforms, such as Facebook, Reddit, and CaringBridge, allowing caregivers to share caregiving-related experiences and seek information and support [22-24]. Participation in OSCs has been shown to help caregivers solve caregiving-related issues, reduce feelings of isolation, improve self-efficacy in the process of caregiving [21,25], and improve overall well-being [26,27].

The interactions between caregivers in OSCs also offer valuable insights into their challenges [28,29] and their responses to said challenges [30]. Existing content analysis studies have provided evidence of the experiences of caregivers for either general older adults [30] or people with Alzheimer disease and related dementia [17,28,29,31]. They found that these caregivers often revealed caregiver burdens [17,24], exchanged social support [31], and engaged in stress-coping behaviors across OSCs [30]. However, limited research has directly compared the experiences of caregivers of people with Alzheimer disease and related dementia and general older adults revealed in OSCs [32]. Unpacking this is important as the knowledge can capture the complexity of the caregiving experience and the role of the caregiving context in shaping the unique challenges faced by caregivers. This suggests that a comparative study on the different experiences communicated by informal caregivers of general older adults and those caring for people with Alzheimer disease and related dementia in OSCs is necessary to better understand the unique needs for both groups.

Emotional Expression in OSCs

Beyond sharing information about specific challenges, messages in OSCs can also reveal caregivers' emotional experiences [33,34]. Caregiving can evoke a range of emotional responses among caregivers when facing challenges [35]. For example, anger may arise from the disruptive behaviors of the care recipient and strained relationships between the caregiver and the care recipient [35]. To cope with these emotions, caregivers often engage in the social sharing of emotions—openly

XSL•FO
RenderX

communicating with others about emotion-eliciting events and their emotions (eg, joy, fear, anger, or sadness) [36]. Anonymous web-based communities such as Reddit provide a safe space for caregivers' social sharing, where they may feel freer to share negative emotions compared to face-to-face settings or other identifiable groups, as these communities reduce social norms and alleviate privacy concerns [37,38]. Through this process, individuals can regulate their emotional experiences [39,40] and gain support and validation from others [41,42], which in turn reduces emotional distress [43].

While existing studies have examined informal caregivers' emotional expressions in OSCs [18,30], 2 research gaps remain. First, little is known about how informal caregivers express emotions when responding to different caregiving challenges. Investigating the relationship between various caregiving challenges and emotional expressions can advance research on caregiver stress and coping. This knowledge, in turn, can inform more targeted support interventions to improve caregivers' well-being. Second, it remains unclear how caregivers in different contexts, such as caregivers of general older adults and caregivers of people with Alzheimer disease and related dementia, express their emotions in OSCs. Faced with varying stressors and experiences, these 2 caregiver groups likely experience and express emotions differently. Comparing the patterns of emotional expression can contribute to a deeper understanding of the psychological needs of caregivers in different contexts, thus facilitating the development of targeted emotional and psychological interventions.

Objectives

This study seeks to address these research gaps by analyzing discussions on Reddit among informal caregivers of older adults, including those taking care of general older adults and people with Alzheimer disease and related dementia. Reddit is a popular anonymous community that allows people to share experiences and seek support [44]. Many subreddits are established to support informal caregivers of older adults, including those especially for caregivers of people with Alzheimer disease and related dementia [17,24]. Therefore, Reddit provides a valuable source of data to investigate how different informal caregivers share their challenges and emotions in OSCs.

This study has 4 objectives. First, it aims to identify the topics discussed by informal caregivers of older adults on Reddit and the underlying challenges they face. Second, it aims to examine how these topics vary between caregivers of general older adults and caregivers of people with Alzheimer disease and related dementia. Third, it aims to investigate how informal caregivers of older adults express emotions in response to different topics and challenges. Fourth, it aims to compare the patterns of emotional expression between the 2 caregiver groups. Our research questions (RQs) are as follows:

- RQ1: What kinds of topics were discussed on Reddit by informal caregivers of older adults, and how do these topics reflect informal caregivers' challenges?
- RQ2: What are the differences in topic distribution in subreddits for informal caregivers of general older adults and those caring for people with Alzheimer disease and related dementia?
- RQ3: What types of emotions were expressed by informal caregivers of older adults in relation to different topics on Reddit?
- RQ4: How do the emotions in Reddit posts differ across subreddits for informal caregivers of general older adults and those caring for people with Alzheimer disease and related dementia?

Methods

Overview

To understand the experiences and feelings of informal caregivers of general older adults and people with Alzheimer disease and related dementia, we applied computer-assisted text analysis, including topic modeling and sentiment analysis, to posts from Reddit communities related to older adults and Alzheimer disease and related dementia caregiving. The computational methods were complemented by a qualitative deep reading, where we selected the top 20 posts from each topic and topic pair for qualitative analysis to provide deeper insight into the obtained topics [45]. In addition, we carried out correlational analysis within and between the topic scores and sentiment scores to identify associations between topics and sentiments. Figure 1 shows our research design.

Figure 1. Research design.

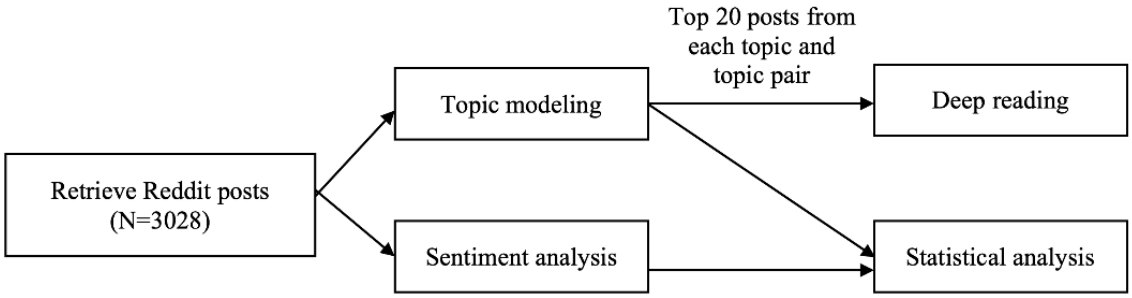

Data Collection

We focused on subreddits specifically designed to support informal caregivers of both general older adults and people with Alzheimer disease and related dementias. To identify relevant subreddits, we searched using the keywords "aging/older/elder" combined with "care/caregiver" for discussions about caring for older adults and "Alzheimer/dementia" combined with "care/caregiver" for those related to Alzheimer and dementia care. We then reviewed each subreddit's description to ensure

XSL•FO
RenderX

that it was a forum for informal caregivers of either the general older adult population or those with Alzheimer and related dementia. The selected subreddits for informal caregivers included r/AgingParents, r/eldercare, r/elderlycare, r/Alzheimers, r/dementia, and r/DementiaHelp. Other caregiver-related subreddits that were not specifically designed to support these caregiver populations were excluded from our data analysis for the purpose of this research.

We used the Reddit application programming interface (API) to retrieve top-level posts from these 6 subreddits. We retrieved up to 1000 of the most recent posts from each of these subreddits, on August 6, 2023, as the Reddit API only provides the 1000 most recent posts at the time of retrieval. For subreddits with fewer than 1000 posts at the time of retrieval, all posts were retrieved.

Topic Modeling

We carried out topic modeling to uncover common topics that occur within this dataset. Topic modeling is a class of natural language processing methods that produces a set of topics from a text dataset, along with a table that indicates how strongly each of these topics occurs in each document.

We used a standard pipeline for topic modeling across the corpus of posts. First, we removed stop words, split the documents into terms with an n-gram range of 1 to 3 (thus allowing up to 3 words per term), and applied term frequency–inverse document frequency weighting to the document–term count matrix [46], which is a process that gives more weight to rarer but possibly more significant terms.

We then applied nonnegative matrix factorization (NMF), which factors the document-term matrix into a pair of document-topic and topic-term matrices [47]. The document-topic matrix shows which topics are present in each document, while the topic-term matrix shows how strongly each term is associated with each topic. We chose NMF rather than a clustering method like BERTopic [48] as we anticipated that each post could belong to multiple topics rather than to a single cluster. We also opted for NMF over latent Dirichlet allocation as NMF tends to give topics that are more meaningful and "more in line with human judgment" than latent Dirichlet allocation [49].

We carried out NMF with 10, 15, 20, and 25 topics. As the number of topics increased from 10 to 20, we observed additional meaningful topics appearing. However, increasing to 25 topics resulted in less coherent topics or splitting existing topics into subtopics. On this basis, we decided to take the results of NMF with 20 topics. From these 20 topics, we excluded 2 topics that consisted mainly of moderator posts about subreddit rules, resulting in a final 18 topics.

Deep Reading

After topic modeling, we used a qualitative deep reading method, following the computation grounded theory by Nelson [50], to interpret the results. This deep reading involved closely examining representative texts for each topic identified from topic modeling and contextualizing key terms to uncover their meanings. By combining computational findings with this human-centered interpretation, we validated patterns identified from topic modeling and offered a comprehensive interpretation of the data [50].

The 2 authors, one specializing in social science and the other in natural language processing, independently reviewed the key terms and top 20 posts for each of the 18 topics to better understand what each topic was about and how the key terms were used within each topic. We individually named and defined each topic, identifying recurring themes. We then compared our findings and resolved any disagreement through discussion to clarify the meaning and context of each topic. For strongly correlated topic pairs, we examined representative posts that scored highly on both topics to better understand when and how these topics co-occur.

We selected the 20 posts that had the highest topic weight in each topic, as these posts contained the most key terms from an individual topic and were thus more representative of the topic [51]. Analyzing these top documents is a common practice in computational grounded theory, which enables a systematic and unbiased examination of the data [50]. This approach ensures that the selected documents accurately represent each topic while allowing for efficient analysis without the burden of interpreting the full dataset. Following prior studies analyzing the top 10 to 20 documents [50,52], we chose the top 20 posts of each topic for qualitative deep reading.

To ensure the rigor and trustworthiness of findings, this study used the interpretive paradigm to ensure the credibility, dependability, and confirmability of our analysis [53]. First, to enhance credibility (ie, the results of qualitative analysis adequately represent multiple realities), we discussed our data collection and analysis procedures with senior colleagues who had expertise in qualitative research. Second, dependability (ie, the consistency of data analysis) was achieved by documenting our thoughts during the deep reading process and engaging in frequent discussions to resolve any inconsistencies. Third, to ensure that researchers' interpretations of data were nonprejudiced (ie, confirmability), we reflected critically on our own perspectives during the deep reading and included verbatim quotes to support the inferences presented in the results. These steps minimized researcher bias in handling data and enhanced the objectivity of our findings.

Sentiment Analysis

We also used the CrystalFeel sentiment analysis API [54] to extract sentiment scores for each post. The CrystalFeel algorithm uses a mixture of lexicon- and word-embedding–based approaches to generate emotional intensity scores and has been validated on human-annotated test samples (achieving correlations of 0.765, 0.818, 0.788, and 0.765 with human annotations on fear, anger, joy, and sadness intensity, respectively) and used in multiple studies for sentiment analysis of social media posts [55,56]. We used CrystalFeel to generate continuous intensity scores (ranging from 0 to 1) for the 4 basic emotions (anger, fear, joy, and sadness), and an overall 5-point sentiment score ranging from −2 (very negative) to +2 (very positive).

XSL•FO
RenderX

Statistical Analysis

To uncover associations between topics and emotions, we computed the Spearman rank correlation coefficient (ρ) among topic weights, and between topic weights and emotion scores. The Spearman rank correlation coefficient reflects the presence of a monotonic (but not necessarily linear) relationship between variables. We chose the Spearman rank correlation coefficient instead of the Pearson correlation coefficient because the distributions of topic weights resulting from NMF were nonnegative and highly skewed toward 0 [57]. The correlations between topic weights in this study reflect the degree of co-occurrence of topics across posts: a higher correlation between two topics indicates that those two topics are more likely to occur within the same post.

Following commonly used guidelines for interpretation (eg, [58]), correlations with a magnitude between 0 and 0.09 were considered negligible, while correlations with a magnitude between 0.1 and 0.39 were considered weak. Note that in our sample, correlations of magnitude >0.06 were significant at $P<.001$.

Ethical Considerations

Before collecting data for this content analysis study, we obtained approval from the Nanyang Technological University's Integrity Review Board (IRB-2023-258) under the exempt review category. Secondary research on existing or publicly accessible datasets, such as our study on Reddit posts, qualifies for this category. Additional informed consent from human subjects is not required for secondary analyses of research data. In compliance with the Integrity Review Board guidelines, we also protected the privacy and confidentiality of individuals by anonymizing users whose data were being analyzed and removing any identifiable information from the posts. No compensation was offered, as this study did not involve direct participation or recruitment of individuals. By strictly following these guidelines, we ensured that this study does not lead to adverse outcomes for human subjects.

Results

Dataset Description

In total, 3028 posts were retrieved from Reddit. Table 1 shows the breakdown of the number of posts retrieved from each subreddit.

Table 1. Number of posts retrieved from each subreddit (N=3028).

Subreddits	Posts, n (%)
Older adults–related posts	1552 (51.25)
r/AgingParents	937 (30.94)
r/eldercare	490 (16.18)
r/elderlycare	125 (4.13)
Dementia-related posts	1476 (48.75)
r/Alzheimers	916 (30.25)
r/dementia	547 (18.06)
r/DementiaHelp	13 (0.43)

Topic Descriptions and Interpretation (RQ1)

Through topic modeling and deep reading, we identified 18 topics discussed by informal caregivers and the underlying challenges they faced (Table 2). By reviewing the key terms and top 20 posts within each topic, we named and described each topic and grouped the topics into 7 overarching themes. Representative posts within each topic are provided in Multimedia Appendix 1.

XSL•FO
RenderX

Table 2. Topic labels, key terms, and descriptions from each topic, along with the topic counts and frequency[a] (N=3028).

Topic, topic name, and description		Topic count, n (%)	Key terms
Theme 1: expressing feelings			
T1	Expressing feelings: discussing feelings about circumstances, including sadness, guilt, helplessness, frustration, stress, and anger.	2178 (71.93)	"just, like, feel, know, things, really, time, feel like, want, does, life, think, make, say, day, hard, people feeling"
Theme 2: seeking advice and support			
T4	Seeking advice and support: asking for advice in dealing with circumstances, often in the form of questions.	1982 (65.46)	"help, advice, need, does, house, needs, getting, need help, refuses, family, lives, want, health, trying, live, suggestions, looking, support"
Theme 3: health and medical concerns			
T2	Duration of medical care: discussing the care recipient's health or behavioral issues, or the need for long-term medical care.	1954 (64.53)	"did, said, hospital, told, days, got, doctor, ago, went, day, today, weeks, week, going, months, called, hospice, just, years, took"
T9	Dementia: seeking or providing information about dementia	1605 (53.01)	"dementia, family, people, loved, diagnosed, people dementia, vascular dementia, vascular, caregivers, person, diagnosed dementia"
T7	Alzheimer: seeking or providing information about Alzheimer disease	1584 (52.31)	"alzheimer, early, onset, early onset, disease, diagnosed, onset alzheimer, early onset alzheimer, diagnosis, years, stage, experience, memory"
T3	Aging and declining health of parents: discussing the aging and the declining health status of parents	1572 (51.92)	"parents, aging, parent, health, life, years, live, aging parents, care, elderly, old, support, time, don, kids, issues, love, age, mental, situation"
T8	Sleep and incontinence: discussing sleep problems in general, or specifically incontinence.	1536 (50.73)	"bed, sleep, night, day, time, hours, door, room, toilet, does, bathroom, sleeping, gets, goes, sleeps, walk, wake, fall, diaper, incontinence"
T12	Medical appointments: discussing different types of medical appointments and experiences with such appointments	1082 (35.73)	"appointment, make, doctor, make appointment, bad, make doctor, said make appointment, memory gotten really, just yelled, make"
Theme 4: care facilities and services			
T6	Memory care facility: discussing memory care facilities for the care recipient, often seeking recommendations for such facilities, or describing their experiences with facilities	1396 (46.10)	"care, memory care, memory, facility, care facility, memory care facility, facilities, needs, need, area, elder, home care, residents, unit"
T11	Nursing home: discussing nursing home for the care recipient, often seeking recommendations for such facilities, or describing their experiences with facilities	1221 (40.32)	"home, nursing, nursing home, care, hospital, home care, medicaid, homes, rehab, nursing homes, home health, work, health, pay"
T10	Assisted-living facility: discussing assisted-living facilities, often seeking recommendations for such facilities, or describing their experiences with facilities.	1160 (38.31)	"living, assisted living, assisted, facility, living facility, assisted living facility, independent living"
Theme 5: care recipients and family relationships			
T14	Parents (mom): mentioning issues in dealing with mom	1571 (51.88)	"mom, sister, caregiver, dog, live, brother, care mom, care, mom, job, dad, mom does, best, mom diagnosed, taking, getting, don, moms, work, hours"
T16	Parents (mother): mentioning issues in dealing with mother	1534 (50.66)	"mother, sister, father, years, time, year, law, away, lives, old, year old, died, siblings, cat, family, passed, brother, care mother, come, mother law"
T15	Parents (dad): mentioning issues in dealing with dad	1525 (50.36)	"dad, mum, brother, home, years, just, want, does, months, help dad, ago, mom dad, dad does, live, care dad, angry, old, father, point, away"
T18	Grandparents: mentioning grandparents in general, or specific grandparents, often talking about their health problems.	1178 (38.90)	"grandma, grandpa, grandparents, aunt, grandmother, family, uncle, want, know, care grandma, does, live, home, help grandma"

XSL•FO
RenderX

Topic, topic name, and description		Topic count, n (%)	Key terms
Theme 6: financial, legal, and insurance issues			
T5	Finances: discussing finance-related matters, such as bills or handling of the ward's assets. Many also mention power of attorney.	1348 (44.52)	"money, account, bank, poa, card, pay, house, attorney, credit, financial, power, finances, power attorney, trust"
T19	Long-term care insurance: discussing long-term care insurance, often asking for recommendations for long-term care insurance providers.	1182 (39.04)	"term, long term, long-term care, term care, long, insurance, care, term care insurance, care insurance, insurance, policy"
Theme 7: technological issues			
T13	Cell phone and devices: discussing phones and related devices, descriptions of difficulties encountered by the care recipient in using such devices, and requests for recommendations of devices.	1605 (53.01)	"phone, use, hearing, alert, elderly, looking, recommendations, device, app, cell, aids, hearing aids, watch, cell phone, devices"

[a]Frequency refers to the percentage of posts with nonzero weights for that topic.

Theme 1 (ie, topic 1), named *expressing feelings*, accounted for the largest proportion of the topic distribution. Posts on this topic mentioned the posters' feelings about circumstances, usually including negative feelings, such as sadness, guilt, helplessness, and frustration. The second most influential theme, theme 2 (ie, topic 4), was *seeking advice*. Posts under this topic focused on asking for advice in dealing with problems, often in the form of questions.

Other themes covered a wide variety of caregiving-related issues. Theme 3 (ie, topics 2, 3, 7, 8, 9, and 12) discussed health and medical issues related to the care recipient, including duration of medical care, dementia, Alzheimer, aging and declining health of parents, sleep and incontinence, and medical appointments. Theme 4 (ie, topics 6, 10, and 11) was about different types of care facilities and services, including memory care facilities, nursing homes, and assisted-living facilities. Theme 5 (ie, topics 14, 15, 16, and 18) focused on care recipients and family relationships, talking about either the posters' mother, father, or grandparents. Theme 6 (ie, topics 5 and 6) was about financial, legal, and insurance issues, including finance-related issues such as bills or assets, often mentioning power of attorney, and long-term care insurance. Theme 7 (ie, topic 13) discussed technological issues, especially cell phones and hearing aid devices.

We also explored the relationships between different topics to have a more detailed understanding of these topics (Table 3). All correlations between topics were negligible (0-0.09 in magnitude) or weak (0.1-0.39 in magnitude).

XSL•FO
RenderX

Table 3. Spearman rank correlations between topics.

Theme and topic	1	2	3						4			5				6		7
	T1	T4	T2	T9	T7	T3	T8	T12	T6	T11	T10	T14	T16	T15	T18	T5	T19	T13
1																		
T1 . Expressing feelings																		
ρ	1	*0.09*[a]	*0.18*	0.05	*0.07*	*0.10*	*0.11*	*0.09*	*0.07*	0.04	0.00	*0.15*	*0.14*	*0.19*	*0.14*	0.02	0.01	−0.03
P value	—[b]	<.001	<.001	0.01	<.001	<.001	<.001	<.001	<.001	0.03	0.80	<.001	<.001	<.001	<.001	0.33	0.70	0.15
2																		
T4 . Seeking advice and support																		
ρ	*0.09*	1	0.01	0.00	*−0.07*	*0.13*	*0.12*	*0.11*	*0.08*	*0.12*	*0.09*	0.05	*0.12*	*0.08*	*0.11*	*0.13*	*0.09*	0.01
P value	<.001	—	0.63	0.94	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.01	<.001	<.001	<.001	<.001	<.001	0.74
3																		
T2 . Duration of medical care																		
ρ	*0.18*	0.01	1	0.04	*0.06*	0.00	*0.18*	*0.08*	*0.09*	*0.09*	0.04	*0.13*	*0.16*	*0.20*	*0.07*	0.02	0.04	*−0.08*
P value	<.001	0.63	—	0.02	<.001	0.99	<.001	<.001	<.001	<.001	0.02	<.001	<.001	<.001	<.001	0.26	0.04	<.001
T9 . Dementia																		
ρ	0.05	0.00	0.04	1	*0.19*	0.01	−0.03	*0.08*	*0.07*	−0.05	−0.04	−0.04	*0.07*	−0.01	*0.06*	−0.06	0.01	0.00
P value	0.01	0.94	0.02	—	<.001	0.61	0.11	<.001	<.001	0.01	0.03	0.04	<.001	0.48	<.001	0.001	0.78	0.95
T7 . Alzheimer																		
ρ	*0.07*	*−0.07*	*0.06*	*0.19*	1	−0.02	−0.05	*0.06*	−0.03	*−0.15*	*−0.10*	−0.02	*0.10*	0.03	0.04	*−0.15*	0.00	0.00
P value	<.001	<.001	<.001	<.001	—	0.30	0.01	<.001	0.08	<.001	<.001	0.31	<.001	0.15	0.05	<.001	0.81	0.90
T3 . Aging and declining health of parents																		
ρ	*0.10*	*0.13*	0.00	0.01	−0.02	1	*−0.10*	0.02	*0.08*	*0.08*	*0.11*	*0.15*	*0.17*	*0.13*	−0.02	*0.14*	*0.14*	0.02
P value	<.001	<.001	0.99	0.61	0.30	—	<.001	0.17	<.001	<.001	<.001	<.001	<.001	<.001	0.36	<.001	<.001	0.36
T8 . Sleep and incontinence																		
ρ	*0.11*	*0.12*	*0.18*	−0.03	−0.05	*−0.10*	1	0.01	0.00	*0.08*	0.02	0.02	0.01	*0.10*	*0.07*	−0.05	−0.01	0.05
P value	<.001	<.001	<.001	0.11	0.01	<.001	—	0.59	0.97	<.001	0.27	0.21	0.70	<.001	<.001	0.004	0.69	0.003
T12 . Medical appointments																		
ρ	*0.09*	*0.11*	*0.08*	*0.08*	*0.06*	0.02	0.01	1	0.04	−0.05	−0.02	0.03	0.03	*0.07*	0.02	0.06	−0.02	−0.01
P value	<.001	<.001	<.001	<.001	<.001	0.17	0.59	—	0.02	0.01	0.26	0.17	0.14	<.001	0.33	0.001	0.33	0.73
4																		
T6 . Memory care facility																		
ρ	*0.07*	*0.08*	*0.09*	*0.07*	−0.03	*0.08*	0.00	0.04	1	*0.22*	*0.14*	*0.08*	*0.08*	0.02	0.00	*0.06*	*0.28*	−0.04
P value	<.001	<.001	<.001	<.001	0.08	<.001	0.97	0.02	—	<.001	<.001	<.001	<.001	0.29	0.85	<.001	<.001	0.02
T11 . Nursing home																		
ρ	0.04	*0.12*	*0.09*	−0.05	*−0.15*	*0.08*	*0.08*	−0.05	*0.22*	1	*0.18*	*0.13*	0.06	*0.07*	*0.07*	*0.10*	*0.18*	*−0.06*
P value	0.03	<.001	<.001	0.01	<.001	<.001	<.001	0.01	<.001	—	<.001	<.001	0.001	<.001	<.001	<.001	<.001	<.001

XSL•FO
RenderX

Theme and topic	1	2	3						4			5				6		7
T10 . Assisted-living facility																		
ρ	0.00	*0.09*	0.04	−0.04	*−0.10*	*0.11*	0.02	−0.02	*0.14*	*0.18*	1	*0.08*	*0.07*	0.01	0.01	*0.19*	*0.11*	0.00
P value	0.80	<.001	0.02	0.03	<.001	<.001	0.27	0.26	<.001	<.001	—	<.001	<.001	0.74	0.47	<.001	<.001	0.97
5																		
T14 . Parents (mom)																		
ρ	*0.15*	0.05	*0.13*	−0.04	−0.02	*0.15*	0.02	0.03	*0.08*	*0.13*	*0.08*	1	0.03	0.15	0.00	*0.07*	*0.07*	*−0.12*
P value	<.001	0.01	<.001	0.04	0.31	<.001	0.21	0.17	<.001	<.001	<.001	—	0.06	<.001	0.83	<.001	<.001	<.001
T16 . Parents (mother)																		
ρ	*0.14*	*0.12*	*0.16*	*0.07*	*0.10*	*0.17*	0.01	0.03	*0.08*	0.06	*0.07*	0.03	1	0.06	*0.09*	*0.12*	0.06	−0.05
P value	<.001	<.001	<.001	<.001	<.001	<.001	0.70	0.14	<.001	0.001	<.001	0.06	—	0.002	<.001	<.001	0.002	0.004
T15 . Parents (dad)																		
ρ	*0.19*	*0.08*	*0.20*	−0.01	0.03	*0.13*	*0.10*	*0.07*	0.02	*0.07*	0.01	*0.15*	0.06	1	0.00	*0.06*	0.01	*−0.07*
P value	<.001	<.001	<.001	0.48	0.15	<.001	<.001	<.001	0.29	<.001	0.74	<.001	0.002	—	0.79	<.001	0.57	<.001
T18 . Grandparents																		
ρ	*0.14*	*0.11*	*0.07*	*0.06*	0.04	−0.02	*0.07*	0.02	0.00	*0.07*	0.01	0.00	*0.09*	0.00	1	0.03	*−0.06*	−0.03
P value	<.001	<.001	<.001	<.001	0.05	0.36	<.001	0.33	0.85	<.001	0.47	0.83	<.001	0.79	—	0.15	<.001	0.06
6																		
T5 . Finances																		
ρ	0.02	*0.13*	0.02	−0.06	*−0.15*	*0.14*	−0.05	0.06	*0.06*	*0.10*	*0.19*	*0.07*	*0.12*	*0.06*	0.03	1	*0.10*	0.02
P value	0.33	<.001	0.26	0.001	<.001	<.001	0.004	0.001	<.001	<.001	<.001	<.001	<.001	<.001	0.15	—	<.001	0.36
T19 . Long-term care insurance																		
ρ	0.01	*0.09*	0.04	0.01	0.00	*0.14*	−0.01	−0.02	*0.28*	*0.18*	*0.11*	*0.07*	0.06	0.01	*−0.06*	*0.10*	1	0.01
P value	0.70	<.001	0.04	0.78	0.81	<.001	0.69	0.33	<.001	<.001	<.001	<.001	0.002	0.57	<.001	<.001	—	0.66
7																		
T13 . Cell phone and devices																		
ρ	−0.03	0.01	*−0.08*	0.00	0.00	0.02	0.05	−0.01	−0.04	*−0.06*	0.00	*−0.12*	−0.05	*−0.07*	−0.03	0.02	0.01	1
P value	0.15	0.74	<.001	0.95	0.90	0.36	0.003	0.73	0.02	<.001	0.97	<.001	0.004	<.001	0.06	0.36	0.66	—

[a]Correlations with *P*<.001 are italicized according to the strength and magnitude of the correlations.

[b]Not applicable.

There was a positive negligible correlation between the two most frequently discussed topics, *seeking advice and support* and *expressing feelings* (ρ=0.09, *P*<.001). Other topics also showed positive, negligible or weak correlations with these two topics but in distinct patterns. Topics related to health and medical concerns, such as *duration of medical care* (ρ=0.18, *P*<.001), *aging and declining health of parents* (ρ=0.10, *P*<.001), *sleep and incontinence* (ρ=0.11, *P*<.001), and *medical appointments* (ρ=0.09, *P*<.001), were positively, weakly or negligibly correlated with *expressing feelings*. In addition, 3 of these topics—*aging and declining health of parents*, *sleep and incontinence*, and *medical appointments*—were also positively, weakly correlated with *seeking advice and support*, with Spearman ρ values of 0.13, 0.12, and 0.11, respectively (all *P*<.001).

Topics related to care facilities and services, including *nursing home* (ρ=0.12, *P*<.001) and *assisted-living facility* (ρ=0.09, *P*<.001), as well as financial topics such as *finances* (ρ=0.13, *P*<.001) and *long-term care insurance* (ρ=0.09, *P*<.001), were

XSL•FO
RenderX

positively, weakly or negligibly correlated with *seeking advice and support*.

To interpret the challenges caregivers faced, we analyzed representative posts that scored high in the above topic combinations by deeply reading. When discussing the *duration of medical care* of the care recipient, such as the care recipient's health or behavioral issues, posters usually expressed negative feelings, such as helplessness and frustration, as a means of venting or seeking comfort. For example, one caregiver shared his fear and uncertainty following his grandmother's abnormal behavioral problems due to dementia:

> *So my grandma (90) was quite healthy in body and mind until this April... she started showing signs of dementia: More severe forgetfulness, language problems, episodes where she didn't seem like she is "now"...I have no experience or expertise with dementia and frankly I am very scared right now...It just seemed so sudden and such a radical change.*

In contrast, posts discussing more practical issues in caregiving, including health and social services, insurance, and financial topics, tended to seek advice or help. For example, a caregiver expressed the need to secure an affordable nursing home:

> *The hospital called my grandmother and said her husband was being discharged and to pick him up. He has dementia and diabetes... how can we find a home he can afford when his income is below the "needed rate"? We've called many various services and no one seems to be able to help or give guidance.*

Caregivers discussing health and medical issues, including aging and declining health of parents, sleep and incontinence, and medical appointments, tended to express both feelings and a need for advice. For example, posts about sleep and incontinence usually directly sought advice to solve the problem while also revealing the emotional toll on caregivers:

> *I help (pretty much am the main) care for my MIL...The problem is bedtime. She doesn't like to be alone. She wanders around every 15 minutes and will not go to bed...It's cold and I'm tired and miserable and I feel so guilty for hiding like this...Any suggestions what I can do to get space so I can sleep? I'm starting to lose my sanity.*

Topic Distribution Comparison (RQ2)

Comparing the topics discussed in subreddits for informal caregivers of general older adults and those caring for people with Alzheimer disease and related dementia, we found that while all topics were present in posts from both older adult–related and Alzheimer disease–related subreddits, certain topics were more prevalent in one community than the other (Figure 2). Posts from older adult–related subreddits tended to be *seeking advice and support*. Compared with Alzheimer disease–related subreddits, they showed greater interest in talking about *aging and declining health of parents*, and care facility–related topics, such as *assisted-living facilities and nursing homes*. In addition, they also focused more on financial, legal, insurance, and technical issues (ie, *finances*, *long-term care insurance*, and *cell phones and devices)*.

However, posts from Alzheimer disease–related subreddits tended to be *expressing feelings*. They showed greater interest in other health- and medical-related topics, such as *duration of medical care*, *sleep and incontinence*, and *medical appointments*. They also tended to talk more about *dementia*, *Alzheimer*, and *memory care facilities*—the topics in relation to Alzheimer and related dementia.

XSL•FO
RenderX

Figure 2. Comparison of mean topic weights between older adult– and Alzheimer disease–related subreddits. Horizontal bars represent 95% CIs for the mean topic weight. Weights for each topic are normalized such that the mean topic weight for Alzheimer disease–related subreddits is 1 and sorted according to the normalized mean topic weight for older adult–related subreddits.

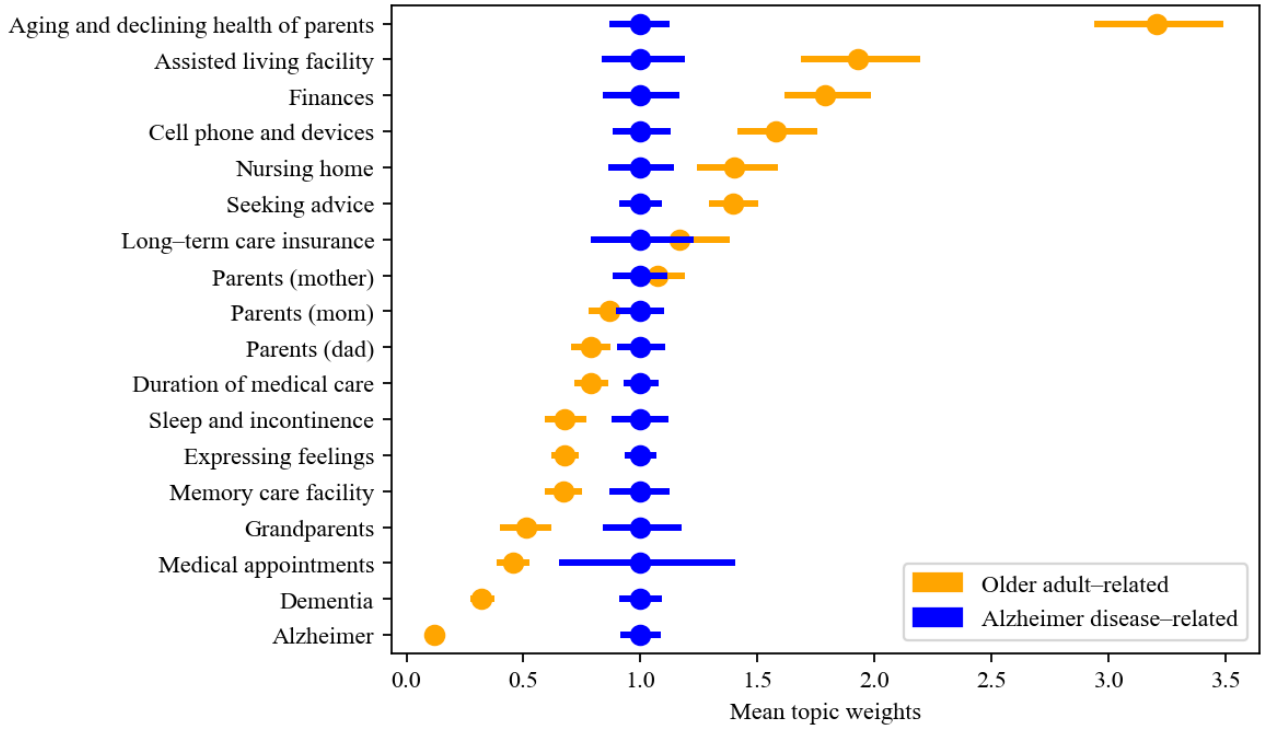

Emotions in Relation to Topics (RQ3)

We analyzed the relationships between 18 topics and emotions obtained from sentiment analysis to understand how caregivers expressed emotions in response to various topics. The results (Table 4) indicated that most topics were positively but weakly correlated with negative emotions and sentiments. The only topic that had a positive, weak correlation with positive sentiment was *cell phones and devices* (ρ=0.20, *P*<.001).

The topic *duration of medical care* was positively, weakly correlated with a set of negative emotions, including *anger* (ρ=0.25, *P*<.001), *fear* (ρ=0.25, *P*<.001), and *sadness* (ρ=0.11, *P*<.001). Similarly, *medical appointments* were positively, negligibly correlated with these emotions, including *anger* (ρ=0.10, *P*<.001), *fear* (ρ=0.09, *P*<.001), and *sadness* (ρ=0.06, *P*<.001). The topics *sleep and incontinence* (ρ=0.14, *P*<.001) and *finances* (ρ=0.24, *P*<.001) were both positively, weakly correlated with *anger*.

While the topic of *Alzheimer* was negatively, weakly correlated with *anger* (ρ=−0.10, *P*<.001), it was positively, weakly correlated with *fear* (ρ=0.10, *P*<.001). The topic of *dementia* was positively, weakly correlated with *fear* (ρ=0.12, *P*<.001) and *sadness* (ρ=0.09, *P*<.001).

XSL•FO
RenderX

Table 4. Spearman rank correlationa between topic weights and emotion intensity scores.

Theme and topic		Sentiment	Joy	Anger	Fear	Sadness
1						
Express feelings	ρ	*−0.20*	*−0.12*	*0.17*	*0.23*	*0.29*
	P value	<.001	<.001	<.001	<.001	<.001
2						
Seeking advice and support	ρ	*−0.11*	*−0.10*	*0.10*	*0.08*	*0.07*
	P value	<.001	<.001	<.001	<.001	<.001
3						
Duration of medical care	ρ	*−0.18*	*−0.11*	*0.25*	*0.25*	*0.22*
	P value	<.001	<.001	<.001	<.001	<.001
Dementia	ρ	−0.04	*−0.08*	0.00	*0.12*	*0.09*
	P value	0.03	<.001	0.83	<.001	<.001
Alzheimer	ρ	0.01	0.00	*−0.10*	*0.10*	0.06
	P value	0.74	0.83	<.001	<.001	0.002
Aging and declining health of parents	ρ	−0.03	0.01	0.02	0.06	*0.10*
	P value	0.08	0.51	0.37	0.002	<.001
Sleep and incontinence	ρ	*−0.12*	−0.03	*0.14*	0.06	0.01
	P value	<.001	0.11	<.001	0.001	0.66
Medical appointments	ρ	*−0.08*	*−0.11*	*0.10*	*0.09*	*0.06*
	P value	<.001	<.001	<.001	<.001	<.001
4						
Memory care facility	ρ	−0.01	0.00	0.01	0.06	*0.07*
	P value	0.78	0.91	0.75	0.002	<.001
Nursing home	ρ	−0.03	0.00	0.05	*0.09*	*0.09*
	P value	0.08	0.94	0.01	<.001	<.001
Assisted-living facility	ρ	−0.04	−0.03	0.05	0.04	0.03
	P value	0.03	0.11	0.003	0.04	0.14
5						
Parents (mom)	ρ	*−0.10*	−0.04	*0.12*	*0.13*	*0.15*
	P value	<.001	0.05	<.001	<.001	<.001
Parents (mother)	ρ	*−0.13*	*−0.10*	*0.13*	*0.16*	*0.18*
	P value	<.001	<.001	<.001	<.001	<.001
Parents (dad)	ρ	*−0.17*	*−0.15*	*0.17*	*0.15*	*0.16*
	P value	<.001	<.001	<.001	<.001	<.001
Grandparents	ρ	−0.04	0.02	0.00	0.02	0.06
	P value	0.01	0.26	0.87	0.22	0.002
6						
Finances	ρ	*−0.09*	*−0.13*	*0.24*	−0.03	0.00
	P value	<.001	<.001	<.001	0.08	0.80
Long-term care insurance	ρ	0.02	0.00	−0.02	0.02	0.01
	P value	0.30	0.83	0.38	0.18	0.57
7						
Cell phone and devices	ρ	*0.20*	*0.16*	*−0.18*	*−0.33*	*−0.33*

XSL•FO
RenderX

Theme and topic		Sentiment	Joy	Anger	Fear	Sadness
	P value	<.001	<.001	<.001	<.001	<.001

[a]Correlations with *P*<.001 significance are italicized.

Emotions Comparisons (RQ4)

Approximately >70% (older adults: 1117/1552, 71.97%; Alzheimer disease: 1095/1476, 74.19%) of posts from both the Alzheimer disease– and older adult–related subreddits expressed an overall negative sentiment, with an additional 10% (older adults: 146/1552, 9.41%; Alzheimer disease: 152/1476, 10.30%) expressing a very negative sentiment (Table 5). In contrast, approximately 10% (older adults: 184/1552, 11.86%; Alzheimer disease: 138/1476, 9.35%) of posts expressed positive sentiments, and very few expressed very positive sentiments. The distribution of overall sentiment scores was similar between the 2 subreddit types.

Table 5. Distribution of sentiments across Alzheimer disease–related and older adult–related subreddits.

	Number of posts, n (%)	
	Alzheimer disease–related subreddits (n=1476)	Older adult–related subreddits (n=1552)
Very positive	2 (0.14)	2 (0.13)
Positive	138 (9.35)	184 (11.86)
Neutral	89 (6.03)	103 (6.64)
Negative	1095 (74.19)	1117 (71.97)
Very negative	152 (10.30)	146 (9.41)

Regarding the expression of specific emotions, *joy* was much less present in both types of subreddits compared with negative emotions (Table 6). In addition, there was slightly more *fear* and *sadness* expressed in the Alzheimer disease–related subreddits than in the older adult–related ones.

Table 6. Comparison of emotion values across Alzheimer disease– and older adult–related subreddits[a].

	Emotion scores, mean (SD)	
	Alzheimer disease–related subreddits	Older adult–related subreddits
Fear[b]	0.51 (0.09)	0.48 (0.10)
Sadness[b]	0.49 (0.08)	0.47 (0.08)
Anger	0.46 (0.07)	0.46 (0.08)
Joy	0.26 (0.08)	0.26 (0.08)

[a]The value of each emotion score is a continuous variable ranging from 0 (ie, a message not expressing the emotion at all) to 1 (ie, a message expresses an extremely high intensity of the emotion).

[b]The differences were significant at *P*<.001.

Discussion

Principal Findings

To provide insights into the challenges and emotional experiences of informal caregivers of older adults, this study combined computer-assisted text analysis (ie, topic modeling and sentiment analysis) with qualitative analysis (ie, deep reading) to analyze the topics, challenges, and emotions expressed by informal caregivers of general older adults and people with Alzheimer disease and related dementia on Reddit. We identified 18 topics from informal caregivers' posts, with posts predominantly focusing on expressing personal feelings and seeking advice and support. These posts revealed various challenges faced by informal caregivers, including managing care recipients' health and medical issues and other practical issues. When discussing these caregiving-related challenges, informal caregivers either expressed their negative feelings or sought help from other forum members.

Our analysis also revealed differences in topics and emotions expressed by users in subreddits for informal caregivers of general older adults and those for people with Alzheimer disease and related dementia. While caregivers of general older adults focused more on practical topics, such as care facilities, caregivers of people with Alzheimer disease and related dementia discussed more about their loved one's health and medical problems, such as dementia disease. In addition, the posts of caregivers of people with Alzheimer disease and related dementia expressed higher levels of negative emotions, such as fear and anger, although posts from both subreddits revealed an overall negative sentiment. Notably, we found that informal caregivers had distinct emotional expressions when discussing different topics. Posts about the duration of medical care were positively weakly correlated with a complex set of negative

XSL•FO
RenderX

emotions, including anger, fear, and sadness. When talking about sleep, incontinence, and financial issues, informal caregivers tended to express anger, while Alzheimer disease–related topics were positively but weakly correlated with fear. The findings indicate a nuanced relationship between caregivers' challenges and their emotional expressions.

Comparison With Previous Works

Existing studies have provided evidence that caregivers often turned to web-based communities, such as Facebook [29] and blogs [28], to share caregiving-related experiences and seek desired information. Similarly, our study found that expressing feelings and seeking advice were predominant topics on subreddits for informal caregivers. According to social support research, both expressing feelings and seeking advice are effective strategies for eliciting social support from others. Expressing feelings, particularly negative ones, may signal a need for emotional support [59,60]. Therefore, the findings suggest that Reddit serves as a platform where caregivers seek social support to cope with caregiving challenges [61].

By combining topic modeling and deep reading, this study also identified specific challenges revealed in web-based discussions among informal caregivers, including managing the health and medical issues of their loved ones, dealing with family relationships, and handling practical caregiving tasks, such as finding appropriate care facilities. These findings align with previous research on informal caregivers' web-based self-disclosure [17,29,62]. For example, analysis of Facebook posts by Bachmann [29] found that caregivers of patients with Alzheimer disease face diverse challenges concerning extended activities of daily living, such as dealing with sleeping disorders, selecting health facilities, and managing medical appointments and tests. The results of this study also complement self-reported data from informal caregivers of older adults [63,64]. For instance, a qualitative study in 4 European countries found that informal caregivers reported facing multiple challenges and needs related to the management of care for older adults [63]. Consistent with these studies, our findings demonstrate the multiple challenges faced by informal caregivers and highlight the need for further research to address their specific needs in caregiving.

Our study adds to prior research on caregivers' challenges by examining the relationship between specific challenge-related topics and the ways caregivers seek support on the web. Previous research has primarily focused on either informal caregivers' sharing of challenges on the web [29-31] or how they seek social support [60] as separate phenomena. In contrast, this study reveals informal caregivers' support-seeking patterns in relation to various challenges by examining topic co-occurrence. When encountering uncontrollable challenges, such as the care recipients' behavioral issues and the progression of Alzheimer disease and related dementia, caregivers were more likely to express feelings in their posts. In comparison, when facing practical-oriented topics that require solutions—such as finding appropriate assisted-living facilities and managing financial and insurance issues—caregivers were more likely to ask for advice on subreddits directly. Moreover, when facing complex health and medical issues, caregivers often combined seeking advice and sharing personal feelings. These findings suggest that informal caregivers may adopt different support-seeking strategies depending on the nature of their challenges. However, the relationships between challenge-related topics and support-seeking strategies were either negligible or weak. Future research could further investigate the stress-coping process in caregiving to understand how caregivers cope with their various challenges.

This study also provides evidence for the unique challenges faced by informal caregivers in 2 caregiving contexts. Prior studies have mainly focused on web-based discussions of one caregiver population, either informal caregivers of general older adults [30] or those taking care of people with Alzheimer disease and related dementia [17,28,29,31], without directly comparing the 2 groups. This study compared the topics these 2 caregiver populations discussed, revealing differences in their concerns and support-seeking behaviors. Informal caregivers of general older adults were more likely to discuss practical topics, including care facilities and financial, legal, insurance, and technical issues. In contrast, caregivers of people with Alzheimer disease and related dementia tended to focus on health and medical-focused issues, such as sleep and incontinence problems and Alzheimer disease, and expressed their feelings more frequently in their posts. The findings are consistent with studies emphasizing the complexities faced by caregivers of people with Alzheimer disease and related dementia in health and medication management tasks. For example, a survey by Riffin et al [65] found that caregivers of people with Alzheimer disease and related dementia encountered greater challenges in managing health-related tasks than caregivers of older adults without Alzheimer disease and related dementia. Hence, future research can examine the distinct challenges faced by informal caregivers in different contexts.

Furthermore, the differences in discrete emotions expressed by informal caregivers of general older adults and caregivers of people with Alzheimer disease and related dementia also mirror the premise of the Caregiver Stress Theory [10] that caregiving contexts shape individuals' experiences. Previous studies have found that caregiving burden, especially emotional burden, can vary across different types of caregivers, especially among dementia and nondementia caregivers [11,13]. Compared with caregivers of older adults, caregivers of people with Alzheimer disease and related dementia face the irreversible progression of functional decline in their loved ones [66] and experience worries and anxiety [13,66]. Our study adds to this understanding by providing evidence that caregivers of people with Alzheimer disease and related dementia tended to disclose higher levels of fear and sadness on Reddit. A similar pattern was also identified through the positive weak correlation between Alzheimer disease–related topics and caregivers' expressions of fear and sadness. These findings underscore the importance of further investigating the unique emotional experiences of informal caregivers in the context of Alzheimer disease and related dementia care.

The negligible or weak correlations between topics and emotions provide a nuanced understanding of informal caregivers' web-based social sharing of emotions [36]. Our study found that discussions on the duration of medical care and medical

XSL•FO
RenderX

appointments had a complex set of negative emotions, including anger, fear, and sadness. This finding corroborates earlier research showing that dealing with long-term medical care of their loved ones is an emotionally distressing process that often elicits negative emotions [1]. In addition, our study found that informal caregivers tended to express anger when discussing financial issues. While previous studies have mainly focused on the financial burden faced by informal caregivers of older adults, they have often overlooked specific barriers and caregivers' emotional responses to them [67,68]. Through deep reading, we found that many caregivers felt angry about the barriers they encountered when trying to manage their care recipient's assets and applying for the power of attorney. This implies caregivers' dissatisfaction and frustration with financial and legal systems in supporting their caregiving responsibilities.

Notably, this study found a negative weak correlation between the topic of Alzheimer disease and expressions of anger. Previous survey-based research has found that feelings of anger were prevalent in caregivers of people with Alzheimer disease and related dementia [69], and the presence of disruptive behaviors in people with Alzheimer disease and related dementia was positively correlated with their anger expression [35]. Different from existing studies, our findings suggest that when caregivers were discussing Alzheimer disease, they were less likely to express anger or resentment. One possible explanation for this discrepancy is that informal caregivers of patients with Alzheimer disease may develop a heightened sense of empathy and understanding toward the care recipient over time [70]. The increase in empathy may lead to reduced anger emotions and corresponding emotional expression in their online discussions [70]. Another possible explanation is the difference in how emotional expressions were measured in this study compared to prior research. To address these discrepancies, further studies adopting other research methods are warranted to investigate caregivers' emotional responses and expressions toward Alzheimer disease–related topics.

Implications

This study contributed in 2 aspects to the literature on the use of OSCs in informal caregiving. First, by illustrating specific challenges and emotional expressions of caregivers of general older adults and caregivers of people with Alzheimer disease and related dementia, this study suggests the importance of examining caregiver stress and coping by considering the role of caregiving contexts, especially the health status of care recipients. Second, this study was the first to empirically examine the correlation between topics and emotional expressions among informal caregivers of older adults. By providing evidence of how informal caregivers express their feelings in response to various caregiving-related challenges, the findings offer a foundation for future research to develop a more nuanced understanding of the patterns of web-based emotional sharing, particularly in relation to the nature of challenges.

Practically, the findings of this study could inform health practitioners and policy makers on how to support the 2 caregiver populations. We found evidence that informal caregivers sought advice and expressed feelings on a variety of caregiving-related matters in OSCs, such as the care recipients' health and medical issues; dealing with family relationships; and practical topics, such as finance, insurance, and technology. The findings bring to light the needs of caregivers for both informational and emotional support in various aspects, which can inform support interventions and services to meet the needs of the population. For example, addressing financial issues through clear guidance on managing assets and applying for power of attorney could benefit caregivers.

Caregivers of people with Alzheimer disease and related dementia discussed more about the medical and behavioral issues of their care recipients, such as duration of medical care, sleep, and incontinence. This underscores the need for targeted support interventions, such as education and training sessions on Alzheimer disease and related dementia-specific caregiving skills. In addition, the frequent expression of negative emotions, especially among caregivers of people with Alzheimer disease and related dementia, suggests the importance of emotional health interventions. OSCs such as Reddit can serve as a valuable forum for caregivers to share experiences and gain emotional support from their peers. To foster this, OSC moderators should encourage open sharing and facilitate supportive interactions among caregivers.

Limitations and Future Works

This work has several limitations. First, we focused on the topic distribution and emotional expression across informal caregivers of general older adults and people with Alzheimer disease and related dementia, while not examining other caregiver populations. A potential direction for subsequent studies is to compare the challenges faced by informal caregivers of older adults with different diseases, such as stroke and diabetes, in a more nuanced manner. This can provide stronger evidence for how specific caregiving contexts, such as the health status of the care recipients, shape caregivers' experiences. Second, this study only investigated the characteristics of informal caregivers' communication patterns on Reddit, a platform dominated by users from the United States. According to previous studies, informal caregivers of older adults also use other OSC platforms, such as Facebook groups and CaringBridge. Future studies could compare topics and emotions expressed by informal caregivers across platforms to enhance our understanding of the role of different OSCs in supporting informal caregivers.

Third, web-based community users may differ from the broader population of informal caregivers in terms of socioeconomic status and media literacy. Therefore, the findings of this study cannot be generalized to all informal caregivers. This limitation stems from the study's specific focus on understanding how caregivers use OSCs to share challenges and emotions. Future research could address this by including nonusers of OSCs and comparing their experiences with those of OSC users, providing a more comprehensive understanding of informal caregivers' experiences across diverse contexts. Fourth, the research methods of this study only allowed us to have a general understanding of informal caregivers' challenges and feelings. Although we adopted qualitative approaches to analyze themes and meanings of representative posts to complement the results

XSL•FO
RenderX

of computer-assisted text analysis, we still cannot infer caregivers' perceived stress and their actual emotional well-being status. Future research could combine computer-assisted text analysis with other methods, such as interviews or surveys, to have a more comprehensive understanding of caregivers' experiences and needs.

Conclusions

In this study, we examined web-based discussions among informal caregivers of general older adults and people with Alzheimer disease and related dementia on Reddit, including their challenges and emotional expressions and their relationships. Informal caregivers discussed a wide range of caregiving-related challenges on Reddit but they varied across caregiving contexts. Caregivers' web-based messages also indicate a high level of negative emotions, including fear, anger, and sadness. In addition, caregivers expressed distinct emotions in response to various challenges, depending on the nature of these challenges. Taken together, this study elucidated informal caregivers' actual problems and needs in different caregiving contexts, thereby providing suggestions for health practitioners and policy makers to develop more targeted support interventions.

Acknowledgments

This research was supported by the Ministry of Education Tier 1 Grant, Singapore (grant RG111/22).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Representative posts within each topic.
[DOCX File , 32 KB - jmir_v27i1e54847_app1.docx]

References

1. Committee on Family Caregiving for Older Adults, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine, Schulz R, Eden J. Families Caring for an Aging America. Washington, DC: National Academies Press; 2016.
2. Schulz R, Beach SR, Czaja SJ, Martire LM, Monin JK. Family caregiving for older adults. Annu Rev Psychol 2020 Jan 04;71(1):635-659 [FREE Full text] [doi: 10.1146/annurev-psych-010419-050754] [Medline: 31905111]
3. Roth DL, Fredman L, Haley WE. Informal caregiving and its impact on health: a reappraisal from population-based studies. Gerontologist 2015 Apr;55(2):309-319 [FREE Full text] [doi: 10.1093/geront/gnu177] [Medline: 26035608]
4. Caregiving for family and friends - a public health issue. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/aging/caregiving/caregiver-brief.html [accessed 2023-11-09]
5. Informal care in Europe: exploring formalisation, availability and quality. European Commission. URL: https://data.europa.eu/doi/10.2767/78836 [accessed 2024-04-29]
6. Zarzycki M, Morrison V. Getting back or giving back: understanding caregiver motivations and willingness to provide informal care. Health Psychol Behav Med 2021 Jul 13;9(1):636-661 [FREE Full text] [doi: 10.1080/21642850.2021.1951737] [Medline: 34345534]
7. Pysklywec A, Plante M, Auger C, Mortenson WB, Eales J, Routhier F, et al. The positive effects of caring for family carers of older adults: a scoping review. Int J Care Caring 2020 Aug 01;4(3):349-375. [doi: 10.1332/239788220X15925902138734]
8. Cameron JI, Stewart DE, Streiner DL, Coyte PC, Cheung AM. What makes family caregivers happy during the first 2 years post stroke? Stroke 2014 Apr;45(4):1084-1089. [doi: 10.1161/strokeaha.113.004309]
9. Carretero S, Garcés J, Ródenas F, Sanjosé V. The informal caregiver's burden of dependent people: theory and empirical review. Arch Gerontol Geriatr 2009 Jul;49(1):74-79 [FREE Full text] [doi: 10.1016/j.archger.2008.05.004] [Medline: 18597866]
10. Pearlin LI, Mullan JT, Semple SJ, Skaff MM. Caregiving and the stress process: an overview of concepts and their measures. Gerontologist 1990 Oct 01;30(5):583-594. [doi: 10.1093/geront/30.5.583] [Medline: 2276631]
11. Bom J, Bakx P, Schut F, van Doorslaer E. The impact of informal caregiving for older adults on the health of various types of caregivers: a systematic review. Gerontologist 2019 Sep 17;59(5):e629-e642 [FREE Full text] [doi: 10.1093/geront/gny137] [Medline: 30395200]
12. Gupta S, Isherwood G, Jones K, van Impe K. Assessing health status in informal schizophrenia caregivers compared with health status in non-caregivers and caregivers of other conditions. BMC Psychiatry 2015 Jul 21;15:162 [FREE Full text] [doi: 10.1186/s12888-015-0547-1] [Medline: 26194890]

XSL•FO
RenderX

13. Sheehan OC, Haley WE, Howard VJ, Huang J, Rhodes JD, Roth DL. Stress, burden, and well-being in dementia and nondementia caregivers: insights from the caregiving transitions study. Gerontologist 2021 Jul 13;61(5):670-679 [FREE Full text] [doi: 10.1093/geront/gnaa108] [Medline: 32816014]
14. Penning MJ, Wu Z. Caregiver stress and mental health: impact of caregiving relationship and gender. Gerontologist 2016 Dec 17;56(6):1102-1113. [doi: 10.1093/geront/gnv038] [Medline: 26035875]
15. Hilgeman MM, Durkin DW, Sun F, DeCoster J, Allen RS, Gallagher-Thompson D, et al. Testing a theoretical model of the stress process in Alzheimer's caregivers with race as a moderator. Gerontologist 2009 Apr;49(2):248-261 [FREE Full text] [doi: 10.1093/geront/gnp015] [Medline: 19363019]
16. Alzheimer's disease facts and figures. Alzheimer's Association®. URL: https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf [accessed 2024-04-29]
17. Cheng K, Yin Z. "Please advise": understanding the needs of informal caregivers of people with Alzheimer's disease and related dementia from online communities through a structured topic modeling approach. AMIA Jt Summits Transl Sci Proc 2022;2022:149-158 [FREE Full text] [Medline: 35854737]
18. Ferrell EL, Russin SE, Hardy RM. Informal caregiving experiences in posttraumatic stress disorder: a content analysis of an online community. J Community Psychol 2019 May 27;47(4):757-771. [doi: 10.1002/jcop.22151] [Medline: 30592051]
19. A profile of older adults with dementia and their caregivers, 2015 issue brief. U.S. Department of Health and Human Services. 2018. URL: https://aspe.hhs.gov/reports/profile-older-adults-dementia-their-caregivers-issue-brief-0 [accessed 2023-11-09]
20. Coulson NS. Online support communities. In: Attrill-Smith A, Fullwood C, Keep M, Kuss DJ, editors. The Oxford Handbook of Cyberpsychology. Oxford, UK: Oxford Academic Press; 2019:240-260.
21. Wright K. Communication in health-related online social support groups/communities: a review of research on predictors of participation, applications of social support theory, and health outcomes. Rev Commun Res 2016;4:65-87. [doi: 10.12840/issn.2255-4165.2016.04.01.010]
22. Diefenbeck CA, Klemm PR, Hayes ER. "Anonymous meltdown": content themes emerging in a nonfacilitated, peer-only, unstructured, asynchronous online support group for family caregivers. Comput Inform Nurs 2017 Dec;35(12):630-638. [doi: 10.1097/CIN.0000000000000376] [Medline: 28742535]
23. Etxeberria I, Salaberria K, Gorostiaga A. Online support for family caregivers of people with dementia: a systematic review and meta-analysis of RCTs and quasi-experimental studies. Aging Ment Health 2021 Jul 02;25(7):1165-1180. [doi: 10.1080/13607863.2020.1758900] [Medline: 32363901]
24. Shoults CC, Rutherford MW, Kemp AS, Addicott MA, Brown A, Greene CJ, et al. Analysis of caregiver burden expressed in social media discussions. Int J Environ Res Public Health 2023 Jan 20;20(3):1933 [FREE Full text] [doi: 10.3390/ijerph20031933] [Medline: 36767297]
25. Del-Pino-Casado R, Frías-Osuna A, Palomino-Moral PA, Ruzafa-Martínez M, Ramos-Morcillo AJ. Social support and subjective burden in caregivers of adults and older adults: a meta-analysis. PLoS One 2018;13(1):e0189874 [FREE Full text] [doi: 10.1371/journal.pone.0189874] [Medline: 29293522]
26. Pagán-Ortiz ME, Cortés DE, Rudloff N, Weitzman P, Levkoff S. Use of an online community to provide support to caregivers of people with dementia. J Gerontol Soc Work 2014;57(6-7):694-709 [FREE Full text] [doi: 10.1080/01634372.2014.901998] [Medline: 24689359]
27. Shim M, Cappella JN, Han JY. How does insightful and emotional disclosure bring potential health benefits?: study based on online support groups for women with breast cancer. J Commun 2011 Jun;61(3):432-454 [FREE Full text] [doi: 10.1111/j.1460-2466.2011.01555.x] [Medline: 25568496]
28. Anderson JG, Hundt E, Dean M, Rose KM. "A fine line that we walk every day": self-care approaches used by family caregivers of persons with dementia. Issues Ment Health Nurs 2019 Mar 13;40(3):252-259. [doi: 10.1080/01612840.2018.1499156] [Medline: 30422041]
29. Bachmann P. Caregivers' experience of caring for a family member with Alzheimer's disease: a content analysis of longitudinal social media communication. Int J Environ Res Public Health 2020 Jun 19;17(12):4412 [FREE Full text] [doi: 10.3390/ijerph17124412] [Medline: 32575455]
30. Alpert JM, Womble FE. Coping as a caregiver for an elderly family member. Health Commun 2015 Jul 25;30(7):714-721. [doi: 10.1080/10410236.2013.879560] [Medline: 25062386]
31. Narasimha S, Wilson M, Dixon E, Davis N, Madathil KC. An investigation of the interaction patterns of peer patrons on an online peer-support portal for informal caregivers of Alzheimer's patients. J Consum Health Internet 2019 Dec 02;23(4):313-342. [doi: 10.1080/15398285.2019.1682757]
32. Ory MG, Hoffman RR, Yee JL, Tennstedt S, Schulz R. Prevalence and impact of caregiving: a detailed comparison between dementia and nondementia caregivers. Gerontologist 1999 Apr 01;39(2):177-185. [doi: 10.1093/geront/39.2.177] [Medline: 10224714]
33. Oser TK, Oser SM, McGinley EL, Stuckey HL. A novel approach to identifying barriers and facilitators in raising a child with type 1 diabetes: qualitative analysis of caregiver blogs. JMIR Diabetes 2017 Oct 26;2(2):e27 [FREE Full text] [doi: 10.2196/diabetes.8966] [Medline: 30291073]

XSL•FO
RenderX

34. Yoo JH, Jang SA, Choi T. Sociocultural determinants of negative emotions among dementia caregivers in the United States and in Korea: a content analysis of online support groups. Howard J Commun 2010 Jan 29;21(1):1-19. [doi: 10.1080/10646170903501195]
35. Crespo M, Fernández-Lansac V. Factors associated with anger and anger expression in caregivers of elderly relatives. Aging Ment Health 2014 May 14;18(4):454-462. [doi: 10.1080/13607863.2013.856857] [Medline: 24229369]
36. Rimé B. Emotion elicits the social sharing of emotion: theory and empirical review. Emot Rev 2009 Jan 01;1(1):60-85. [doi: 10.1177/1754073908097189]
37. Vermeulen A, Vandebosch H, Heirman W. #Smiling, #venting, or both? Adolescents' social sharing of emotions on social media. Comput Human Behav 2018 Jul;84:211-219. [doi: 10.1016/j.chb.2018.02.022]
38. Wang J, Wei L. Fear and hope, bitter and sweet: emotion sharing of cancer community on Twitter. Soc Media Soc 2020 Jan 16;6(1):205630511989731. [doi: 10.1177/2056305119897319]
39. Miklósi M, Martos T, Szabó M, Kocsis-Bogár K, Forintos D. Cognitive emotion regulation and stress: a multiple mediation approach. Transl Neurosci 2014;5(1):52. [doi: 10.2478/S13380-014-0207-9]
40. Behrouian M, Ramezani T, Dehghan M, Sabahi A, Ebrahimnejad Zarandi B. The effect of emotion regulation training on stress, anxiety, and depression in family caregivers of patients with schizophrenia: a randomized controlled trial. Community Ment Health J 2020 Aug 13;56(6):1095-1102. [doi: 10.1007/s10597-020-00574-y] [Medline: 32056061]
41. Duprez C, Christophe V, Rimé B, Congard A, Antoine P. Motives for the social sharing of an emotional experience. J Soc Pers Relat 2014 Sep 10;32(6):757-787. [doi: 10.1177/0265407514548393]
42. Kennedy-Moore E, Watson JC. How and when does emotional expression help? Rev Gen Psychol 2001 Sep 01;5(3):187-212. [doi: 10.1037/1089-2680.5.3.187]
43. Nils F, Rimé B. Beyond the myth of venting: social sharing modes determine the benefits of emotional disclosure. Euro J Social Psych 2012 Apr 17;42(6):672-681. [doi: 10.1002/ejsp.1880]
44. Hintz EA, Betts T. Reddit in communication research: current status, future directions and best practices. Ann Int Commun Assoc 2022 Apr 23;46(2):116-133. [doi: 10.1080/23808985.2022.2064325]
45. Gillies M, Murthy D, Brenton H, Olaniyan R. Theme and topic: how qualitative research and topic modeling can be Brought together. arXiv Preprint posted online October 3, 2022 [FREE Full text] [doi: 10.48550/arXiv.2210.00707]
46. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge, MA: Cambridge University Press; 2008.
47. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999 Oct 21;401(6755):788-791. [doi: 10.1038/44565] [Medline: 10548103]
48. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv Preprint posted online March 11, 2022 [FREE Full text]
49. Egger R, Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Front Sociol 2022 May 6;7:886498 [FREE Full text] [doi: 10.3389/fsoc.2022.886498] [Medline: 35602001]
50. Nelson LK. Computational grounded theory: a methodological framework. Sociol Methods Res 2017 Nov 21;49(1):3-42. [doi: 10.1177/0049124117729703]
51. Törnberg A, Törnberg P. Combining CDA and topic modeling: analyzing discursive connections between Islamophobia and anti-feminism on an online forum. Discourse Soc 2016 Mar 28;27(4):401-422. [doi: 10.1177/0957926516634546]
52. Nikolenko SI, Koltcov S, Koltsova O. Topic modelling for qualitative studies. J Inf Sci 2016 Jul 10;43(1):88-102. [doi: 10.1177/0165551515617393]
53. Lincoln YS, Guba EG, Pilotta JJ. Naturalistic inquiry. Int J Intercult Relat 1985 Jan;9(4):438-439. [doi: 10.1016/0147-1767(85)90062-8]
54. Gupta RK, Yang Y. CrystalFeel at SemEval-2018 task 1: understanding and detecting emotion intensity using affective lexicons. In: Proceedings of the 12th International Workshop on Semantic Evaluation. 2018 Presented at: SIGLEX '18; June 5-6, 2018; New Orleans, LA p. 256-263 URL: https://aclanthology.org/S18-1038.pdf [doi: 10.18653/v1/s18-1038]
55. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
56. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 2021 Mar;101:107057 [FREE Full text] [doi: 10.1016/j.asoc.2020.107057] [Medline: 33519326]
57. Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. Psychol Methods 2012 Sep;17(3):399-417. [doi: 10.1037/a0028087] [Medline: 22563845]
58. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
59. Barbee AP, Cunningham MR. An experimental approach to social support communications: interactive coping in close relationships. Ann Int Commun Assoc 2016 May 18;18(1):381-413. [doi: 10.1080/23808985.1995.11678921]

XSL•FO
RenderX

60. Benson JJ, Oliver DP, Washington KT, Rolbiecki AJ, Lombardo CB, Garza JE, et al. Online social support groups for informal caregivers of hospice patients with cancer. Eur J Oncol Nurs 2020 Feb;44:101698 [FREE Full text] [doi: 10.1016/j.ejon.2019.101698] [Medline: 31816508]
61. Pickett AC, Valdez D, Sinclair KL, Kochell WJ, Fowler B, Werner NE. Social media discourse related to caregiving for older adults living with Alzheimer disease and related dementias: computational and qualitative study. JMIR Aging 2024 Jun 19;7:e59294 [FREE Full text] [doi: 10.2196/59294] [Medline: 38896462]
62. Anderson JG, Hundt E, Dean M, Keim-Malpass J, Lopez RP. "The church of online support": examining the use of blogs among family caregivers of persons with dementia. J Fam Nurs 2017 Feb;23(1):34-54. [doi: 10.1177/1074840716681289] [Medline: 27920340]
63. Akgun-Citak E, Attepe-Ozden S, Vaskelyte A, van Bruchem-Visser RL, Pompili S, Kav S, et al. Challenges and needs of informal caregivers in elderly care: qualitative research in four European countries, the TRACE project. Arch Gerontol Geriatr 2020 Mar;87:103971. [doi: 10.1016/j.archger.2019.103971] [Medline: 31756568]
64. Strommen J, Fuller H, Sanders GF, Elliott DM. Challenges faced by family caregivers: multiple perspectives on eldercare. J Appl Gerontol 2020 Apr 19;39(4):347-356. [doi: 10.1177/0733464818813466] [Medline: 30453805]
65. Riffin C, van Ness PH, Wolff JL, Fried T. Family and other unpaid caregivers and older adults with and without dementia and disability. J Am Geriatr Soc 2017 Aug 20;65(8):1821-1828 [FREE Full text] [doi: 10.1111/jgs.14910] [Medline: 28426910]
66. Sun F. Caregiving stress and coping: a thematic analysis of Chinese family caregivers of persons with dementia. Dementia (London) 2014 Nov;13(6):803-818. [doi: 10.1177/1471301213485593] [Medline: 24339084]
67. Silva AL, Teixeira HJ, Teixeira MJ, Freitas S. The needs of informal caregivers of elderly people living at home: an integrative review. Scand J Caring Sci 2013 Dec 04;27(4):792-803. [doi: 10.1111/scs.12019] [Medline: 23289859]
68. Faronbi JO, Faronbi GO, Ayamolowo SJ, Olaogun AA. Caring for the seniors with chronic illness: the lived experience of caregivers of older adults. Arch Gerontol Geriatr 2019 May;82:8-14 [FREE Full text] [doi: 10.1016/j.archger.2019.01.013] [Medline: 30710847]
69. Kramer B, Gibson J, Teri L. Interpersonal family stress in Alzheimer's disease. Clin Gerontol 1993 Feb 26;12(1):57-75. [doi: 10.1300/j018v12n01_06]
70. Wang H, Cui H, Wang M, Yang C. What you believe can affect how you feel: anger among caregivers of elderly people with dementia. Front Psychiatry 2021 Apr 7;12:633730 [FREE Full text] [doi: 10.3389/fpsyt.2021.633730] [Medline: 33897493]

Abbreviations

API: application programming interface
NMF: nonnegative matrix factorization
OSC: online support community
RQ: research question

Edited by N Cahill; submitted 06.12.23; peer-reviewed by A Efthymiou, T Adebambo; comments to author 28.02.24; revised version received 13.04.24; accepted 08.01.25; published 28.02.25.

Please cite as:
Huang NM, Wong LZ, Ho SS, Timothy B
Understanding Challenges and Emotions of Informal Caregivers of General Older Adults and People With Alzheimer Disease and Related Dementia: Comparative Study
J Med Internet Res 2025;27:e54847
URL: https://www.jmir.org/2025/1/e54847
doi:10.2196/54847
PMID:40053723

©Nova Mengxia Huang, Liang Ze Wong, Shirley S Ho, Bryan Timothy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Living With and Managing Uncomplicated Urinary Tract Infection: Mixed Methods Analysis of Patient Insights From Social Media

Melissa L Kramer[1,2], BBus; Jose Medina Polo[3], PhD; Nishant Kumar[4], MPH; Aruni Mulgirigama[5], MBBS, MRCP, MFPM; Amina Benkiran[5], PharmD

[1]Live UTI Free Ltd, Dublin, Ireland

[2]University of Reading, Reading, United Kingdom

[3]Department of Urology, Hospital Universitario 12 de Octubre, Madrid, Spain

[4]Veeva Systems, London, United Kingdom

[5]GSK, London, United Kingdom

Corresponding Author:
Melissa L Kramer, BBus
Live UTI Free Ltd
Suite 7, The Courtyard
Carmanhall Road, Sandyford
Dublin, D18 NW62
Ireland
Phone: 386 64 157 997
Email: melissa@liveutifree.com

Abstract

Background: Uncomplicated urinary tract infections (uUTIs) affect more than half of women in their lifetime and can impact on quality of life. We analyzed social media posts discussing uUTIs to gather insights into the patient experience, including aspects of their disease management journey and associated opinions and concerns.

Objective: This study aims to gather patient experience insights by analyzing social media posts that discussed uUTI.

Methods: A search string ("urinary tract infection" [UTI] or "bladder infection" or "cystitis" or "UTI" not "interstitial cystitis") was used to identify posts from public blogs and patient forums (June 2021 to June 2023). Posts were excluded if they were not written in English or discussed complicated UTI (posts that mentioned "pregnancy" or "pregnant" or "trimester" or "catheter" or "interstitial"). Posts were limited to publicly available sources and anonymized. The primary objective was to gather patient perspectives on key elements of the uUTI experience, including health care professional (HCP) interactions, diagnosis, treatment, and recurrence.

Results: In total, more than 42,000 unique posts were identified (mostly from reddit.com; 29,506/42,265, 70%) and >3600 posts were analyzed. Posts were most commonly from users in the United States (6707/11,180, 60%), the United Kingdom (2261/11,180, 20%), Canada (509/11,180, 5%), Germany (356/11,180, 3%), or India (320/11,180, 3%). Six main themes were identified: symptom awareness and information seeking, HCP interactions, diagnosis and management challenges, management with antibiotics, self-management, and challenges with recurrent UTI. Most posts highlighted the importance of seeking professional medical advice, while some patients raised concerns regarding their HCP interactions and lack of shared decision-making. Patients searched for advice and guidance on the web prior to consulting an HCP, described their symptoms, and discussed lifestyle adjustments. Most patients tried self-management and shared their experiences with nonprescribed treatment options. There was general agreement among posts that antibiotics are necessary to cure UTIs and prevent associated complications.

Conclusions: Social media posts provide valuable insight into the experiences and opinions of patients with uUTIs in Canada, Germany, India, the United Kingdom, and the United States. The insights from this study provide a more complete picture of patient behaviors and highlight the potential for HCP and patient education, as well as better communication through shared decision-making to improve care.

(J Med Internet Res 2025;27:e58882) doi:10.2196/58882

XSL•FO
RenderX

KEYWORDS

acute cystitis; bladder infection; HCP interactions; urology; patient experience; patient insights; social media; uncomplicated urinary tract infection; urinary tract infection; urinary; women; quality of life; disease management; cystitis; healthcare professional; self-management; patient behavior; UTI

Introduction

Urinary tract infections (UTIs) affect >50% of women in their lifetime [1,2], which can necessitate multiple health care professional (HCP) visits and antibiotic prescriptions, and have a substantial impact on quality of life [3,4]. Determinants of this quality-of-life impact include the capacity for symptoms to affect everyday activities (including sex, sleep, work, and exercise) and the emotional toll of disease management [5,6]. Frustration and helplessness are key examples of the emotions reported by patients with UTIs [5,7]. Most evaluations of patient perspectives of UTIs have focused on individuals recruited from clinical settings [8,9], but these data may be limited in terms of the information that patients are willing to share with researchers and the sample size may be small and not representative of the whole patient population. Other studies have gathered the views of patients with UTIs via web-based surveys [4,10-14].

Social media has become a part of everyday life for many people, with an estimated 4.89 billion users worldwide in 2023 [15]. This number is expected to reach 5.85 billion by 2027 [15]. Many patients use social media sites to source medical information prior to obtaining information in person [16]. These platforms thus represent a promising route by which we can learn about the concerns and experiences of patients without recruiting them from care settings or through research panels [17]. Analyzing patient discussions on social media may provide unique insights into the real-life experiences of patients beyond what can be learned using traditional research methodologies [17]. To our knowledge, only 2 studies have been published that present information about patients with UTI, extracted directly from web-based platforms (one of which was not UTI-specific) [16,18].

We analyzed social media posts discussing uncomplicated UTIs (uUTIs) to gather insights into the patient experience, including aspects of their disease management journey and associated opinions and concerns. uUTI (also known as “acute cystitis”) is a condition characterized by pyuria and a documented microbial pathogen on urine culture, often accompanied by local signs and symptoms such as lower abdominal discomfort and dysuria. uUTI is limited to nonpregnant females with no known relevant anatomical or functional abnormalities in the urinary tract and no comorbidities [19]. A quantitative analysis was also conducted to assess the proportion of posts discussing key uUTI therapies, home remedies, and HCP specialties. Refer to Multimedia Appendix 1 for “Patient Lay Summary” of this article.

Methods

Postidentification and Processing

A search string (“urinary tract infection” or “bladder infection” or “cystitis” or “UTI” but not “interstitial cystitis”) was used to identify online posts from Veeva Link data sets comprising multiple public blogs and patient forums from June 2021 to June 2023, as shown in Figure 1. Identified posts were subjected to natural language processing. Posts were excluded if they did not contain personal pronouns (to identify self-reports; examples of personal pronouns include I, me, my, mine, and myself), were not written in English, were deemed spam or advertisements (identified by manual inspection), discussed complicated UTI (identified as posts using the terms “pregnancy” or “pregnant” or “trimester” or “catheter” or “interstitial”), or were duplicates. Non–English-language sources, special characters, emojis, or website URLs included in the posts, duplicate posts, or posts that included terms associated with complicated UTIs were removed. Each post was assessed and clustered with other semantically similar posts.

The remaining posts were clustered into themes based on semantic similarities. Posts were embedded into a high-dimensional space using the all-mpnet-base-v2 transformer model and the dimensionality was reduced with Uniform Manifold Approximation and Projection [20] using the BERTopic framework [21] before clustering via HBDSCAN [22]. Model hyperparameters, such as minimum cluster size (n=10), were guided by intrinsic assessment of cluster quality such as coherence scores and extrinsic assessment of cluster quality through manual review. Topic interpretation for each cluster was aided by assessing the term frequency-inverse document frequency of keywords, and the GPT4 large language model was used to identify tangible themes for the clusters. The extracted keywords and GPT4 representations were assessed to determine which of the clusters warranted further manual examination to extract qualitative insights on uUTIs in terms of disease experiences, treatment experiences, prevention of UTIs, and experiences with HCPs. Clusters not relevant to uUTIs in these contexts were rejected.

XSL•FO
RenderX

Figure 1. Data collection, extraction, and analysis.

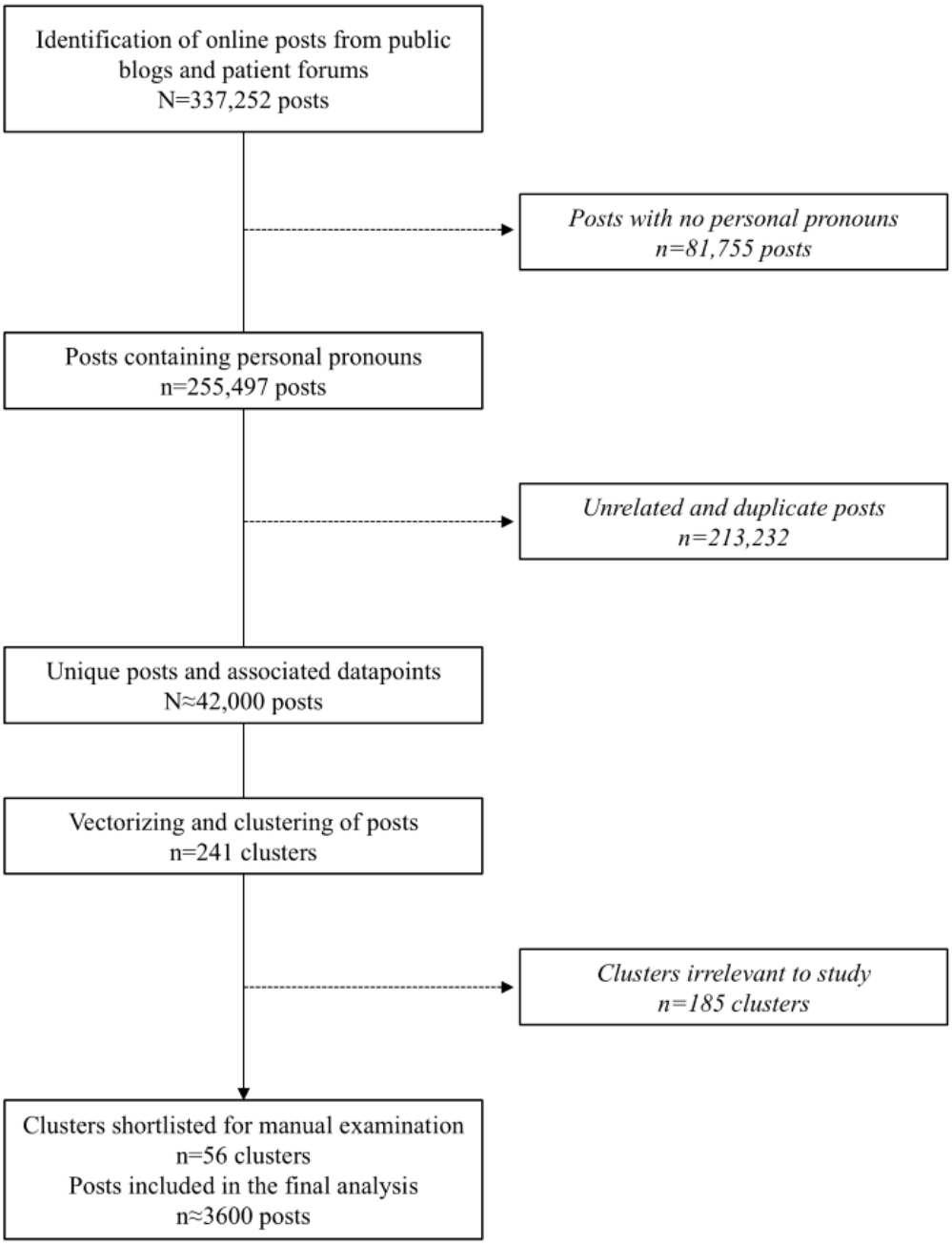

End Points and Analyses

The primary objective was to gather patient perspectives on key elements of the uUTI experience, including HCP interactions, diagnosis, treatment, and recurrence. Patient quotes are presented for each theme, and insights of interest are discussed alongside key datapoints (including website domains and countries of origin).

Quantitative data were also gathered to evaluate the frequency of posts discussing drugs and home remedies (including specific examples of each) and HCP specialties. This information is presented as a percentage of the total posts per country for countries with a sufficient sample size (the United States, the United Kingdom, Canada, Germany, and India). This analysis was performed prior to the launch of the Pharmacy First scheme (launched in the United Kingdom, January 2024) [23].

Ethical Considerations

This research used secondary data only, was nonclinical and noninterventional, and was for patient experience insight purposes only, to support the current paucity of published data on patients' perspectives with UTI. Institutional review board and research ethics committee approvals were not required nor sought because the research did not meet the requirements for needing ethical approval, per section 1.3 of the EphMRA guidelines [24]. The ethical principles established by the Declaration of Helsinki were respected. Consent to use data from the identified posts was considered implicit due to the published terms and conditions of each platform and the website privacy agreement with users (where relevant). Posts were limited to publicly available sources and anonymized to ensure that no personally identifiable information was included.

XSL•FO
RenderX

Results

Posts Included in the Analysis

A total of 337,252 posts were identified using the search string and 255,497 remained after filtering for those using personal pronouns (Figure 1). Removal of non-English posts, special characters, emojis, website URLs, posts deemed spam or advertisements, posts discussing complicated UTI, and duplicates gave ≈42,000 unique posts. The web-based source and geographic distribution of these posts is shown in Table 1. Most posts were taken from reddit.com (29,506/42,265, ~70%) and were most often from users in the United States (6707/11,180, 60%), the United Kingdom (2261/11,180, 20%), Canada (509/11,180, 5%), Germany (356/11,180, 3%), or India (320/11,180, 3%). After clustering, ≈3600 posts were identified for final analysis and manually examined to extract qualitative insights. Key takeaways were identified from each cluster, leading to the identification of 6 main themes from these posts. These themes were symptom awareness and information seeking, HCP interactions, diagnosis and management challenges, management with antibiotics, self-management, and challenges with recurrent UTI. Key patient quotes are included in Table 2.

Table 1. The web-based source and geographic distribution of posts.

Parameter	Web-based posts, n (%)
Web-based source[a]	
reddit.com	29,506 (70)
tumblr.com	3973 (9)
whattoexpect.com	1869 (4)
healthunlocked.com	1453 (3)
babycenter.com	629 (1)
youtube.com	592 (1)
healingwell.com	271 (1)
healthcaremagic.com	264 (1)
nature.com	245 (1)
babycenter.ca	235 (1)
Other[b]	3228 (8)
Country of origin[c]	
United States	6707 (60)
United Kingdom	2261 (20)
Other[d]	1027 (9)
Canada	509 (5)
Germany	356 (3)
India	320 (3)

[a]Total number of posts (N=42,265).

[b]Approximately 230 additional sites.

[c]Total number of posts where geographic information was available (N=11,180).

[d]Argentina, Australia, Austria, Bangladesh, Barbados, Belgium, Bosnia and Herzegovina, Brazil, Bulgaria, Chile, China, Costa Rica, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Greece, Guam, Hong Kong S.A.R., Hungary, Iceland, Indonesia, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Luxembourg, Macao SAR, Malaysia, Mexico, Morocco, Nepal, the Netherlands, New Zealand, Nigeria, North Macedonia, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Republic of Ireland, Romania, Russia, Saudi Arabia, Seychelles, Singapore, Sint Maarten, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, the United Arab Emirates, and Vietnam.

XSL•FO
RenderX

Table 2. Select patient quotes organized by theme.

Theme	Definition and verbatim quotes[a]
Symptom awareness and information seeking	• This theme showcases a wide range of experiences and concerns related to UTI[b] symptom awareness, prevention strategies, emotional impact, and peer support. The topic centers on UTI symptoms and guidance for managing suspected cases. • "After exhausting all my options with doctors and gynaecologists, I am now turning to reddit and other women who may be able to help. I've been a sufferer of frequent UTIs for most of my adult life […] Every time I have a period, or am about to have a period, I develop a UTI, there seems to be no rhyme or reason on whether it'll happen just before or just after, but without fail, I'll always get one." • "It sounds like you have a urinary tract infection. It's pretty easy to treat but you do need to go to a doctor or urgent care […] to be sure." • "So my friend recently got a bladder infection (we think) she's a bit nervous about going to the doctor for various reasons, so we're looking for natural suggestions that might work […] She doesn't have any serious symptoms yet, just the general discomfort while peeing and the constant urge to pee and unintentional leaking but we just don't want it to escalate and want some tips before going to a doctor." • "You can try taking cranberry tablets they work for some people. Also drink plenty [of] water after and try to pee again if you can! If the problem persists you can get an antibiotic to take each time you have sex from the doctor."
HCP[c] interactions	• This theme predominantly revolves around the frustration individuals with UTI symptoms feel when seeking treatment from HCPs. • "Doctor has suggested [low dose antibiotic] as I feel like I have cystitis symptoms all the time for last 2 months, got to wait 6 months to see urologist to see what problem is. My GP[d] said they don't have the answers, a lot of it is trial and error. I know myself better than them so I'm always looking and trying different things." • "I went to the Dr because I suspected a bladder infection. Gave all my symptoms. […] Only after pushing did he condescendingly ask if I'd feel better if I provided a urine sample. I said yes. He couldn't look me in the eyes when he gave me a prescription for you [guessed] it, a bladder infection. I've learned that if I'm dealing with my [health] to push back if I feel I am not listened to." • "It took visits to my general physician (who, in less words, told me 'Just stop having sex' LOL), my gynecologist, and then a urologist she recommended, before I got relief. I'm gonna tell you, have your girlfriend see a FEMALE urologist, if you can. Male urologists unfortunately don't always keep up with female urinary research (it's extremely behind male urinary research, too) and having someone who could relate to me and understood my pain as a female/AFAB[e] was HUGE."
Diagnosis and management challenges	• This theme primarily centers around obstacles that individuals experiencing UTI symptoms encounter when seeking an accurate diagnosis from HCPs. • "OMG!!! I knew I had a UTI and the [….] doctor kept refusing to test saying it was probably an STD[f]. It was not an STD and when the results for the UTI came back it showed I had a staph infection in my bladder." • "The first time I was quite happy to go through the rigamarole of giving a urine sample and getting antibiotics. After a few episodes of this happening, the antibiotics started playing with my gut and causing issues there […] four years in I have a system that has worked well enough for me." • "I had a WOMAN doctor tell me the blood coming from an intense bladder infection was… and I quote, 'Probably a miscarriage.' I had been going through a divorce for months and hadn't had any physical contact, let alone the [contraceptive] I was on. She didn't even ask. The bias against women's health is incredibly ingrained in the entire system." • "He put me on Flagyl, which immediately gave me a yeast infection, put me on another med, got the UTI back. He put me on Flagyl again, and another yeast infection. Apparently it's a side effect. This cycle went on about 8 times, and every time, I told him the meds weren't working and giving me more problems. But he's the doctor and knows best, right? Weeks later, I'm in ICU[g] literally clinging to life with sepsis, because of an 'undiagnosed UTI'." • "Doing a more in-depth test cannot hurt […] it's sad all they do is dip stick, culture, cipro […] I would have never found my staph infection in my bladder had I not tested further using the advanced next generation sequencing DNA test."

XSL•FO
RenderX

Theme	Definition and verbatim quotes[a]
Management with antibiotics	• This topic focuses on physician-led treatment approaches to managing UTIs, including personal experiences with the antibiotic use, their pros and cons, and potential side effects. • "I had, maybe still have, a super resistant [*Pseudomonas*] UTI. I'm still having occasional burning in the bladder area and occasional back pain a week after stopping [antibiotic] (really bad side effects, had to stop after 4 days, but doctors said I took enough of it) […] my urine is clear now despite my symptoms. They still sent it for culture and I'm waiting on that to come back. I am so distressed from this, it's taking a huge toll on me as it's been going on for over two weeks back and forth with doctors. Also, especially after reading about [*Pseudomonas*], I feel like I won't be able to get rid of it and who knows what will happen to me." • "I had this same problem for years. At one point I had 5 UTIs in 4 months. Every time it would be after sex with my husband. The doctor cultured it but could never identify what specifically was causing the UTI. Finally, when I saw my gynecologist she gave me a prescription for [antibiotic]. I take one pill right after sex. I've been UTI free for the last 15 years." • "I ask because the [antibiotic] is causing me a lot of pain in the left leg and I'm scared to continue it […] I had three of the six I was told to take before I stopped because I couldn't sleep or walk well from the pain (which slowly died down over the day I was off them) but now I've taken a fourth because the UTI scares me." • "If [antibiotic] hasn't worked after the 1st course we generally want to switch to something else rather than bombard the body with long-term antibiotics. That's just a recipe for an ongoing yeast infection, lowered gut health and a massive case of [*Clostridium difficile*] diarrhea."
Self-management	• This topic focuses on self-management approaches to treating UTIs, including recommendations regarding different approaches to managing UTIs such as home remedies. • "UTIs are more common in women than men. If you are male definitely go to the doctor […] If you are female, you are supposed to go to the doctor, but honestly most just drink water and low sugar cranberry juice until they're fine because they're so common. If you get a fever, definitely go to the doctor." • "Stop just using cranberry juice every time you have a UTI!! So many [women] go through so much unnecessary pain and worsen their UTIs just because they think drowning themselves in cranberry juice will help. Cranberry juice/cranberry supplements can HELP PREVENT UTIs, once you have one, your best friend is going to be antibiotics, pain killers, and WATER to flush out bacteria." • "Just wondering if anyone has tried D-Mannose for recurrent UTI? I have a prescription for [antibiotic] to take after intercourse or when I feel a UTI coming on but I take it so often that I'm worried about antibiotic resistance. I use Azo a lot but it stains... I'm trying to find something to help as a preventative." • "I've had great success with using D-Mannose. Since I had a UTI in [February] I was left with chronic low-level symptoms that didn't go away—awareness, bladder burning, frequency. My culture came back negative but after reading about embedded infections I began to suspect that it was an infection that was just going undetected. I use powder D-Mannose (more potent than capsules). After about a week of high dosing (4× a day) my symptoms were greatly reduced, almost completely." • "I used to get UTIs, and having to go to the doctor placing me on different kinds of antibiotics. Some worked, some didn't so I wound up having to go back to the doctor which winds up to be very expensive [...] I came across a video on YouTube where this girl [...] said she was taking a probiotic plus something for UTIs. She was taking a product called D-Mannose. I ordered it from Amazon, and I have to say that it has worked for me."
Challenges with recurrent UTI	• This topic primarily centers on individuals' experiences with recurrent UTIs. Discussions focus on personal experiences, physical and mental burden, and methods used for both treatment and prevention. • "I don't know if my fourth and most recent [antibiotic] is working against this, and my doctor won't be in for the next five days. I'm so tired of this! I know that rUTIs[h] is a thing, but to have these literally back to back? No day to breathe without wondering if this will be when I go septic and die? I almost wish that this will kill me at this point." • "I've been to urologists and a [urogynaecologist] in addition to my [gynaecologist]. No one can give me answers. My cystoscopy was fine and after that, the [urogynaecologist] said to just take dmannose daily, drink 3-4 liters of water a day, and good luck. I refuse to believe this will be my life now [...] There HAS to be some resolve." • "I've struggled with recurrent UTIs for about 15 years and tried every single supplement and herb I could find info about. I often react very poorly to antibiotics (ranging from allergic reactions to intestinal yeast infection, oral thrush, diarrhea for weeks, flu-like symptoms, I [hate] taking antibiotics) so I avoid them whenever possible, I'll do anything to cure UTIs at home before resorting to antibiotics." • "LOOKING FOR ADVICE! I suffer from recurring urinary tract infections. When I go to the doctor, they just put me on antibiotics for a few days and the infection usually goes away […] I don't want to be on antibiotics that often and ruin my microbiome. However, I don't want to risk permanent damage to my kidneys by not treating the UTIs properly. I am looking for tips on how to better prevent them moving forward."

XSL•FO
RenderX

[a]Quotes have not been edited, except for omission of text where deemed appropriate (indicated by [...]) or correction of spelling or grammar, removal of drug names, and definition of non–standard abbreviations for clarity (indicated by corrected or new text).

[b]UTI: urinary tract infection.

[c]HCP: health care professional.

[d]GP: general practitioner.

[e]AFAB: assigned female at birth.

[f]STD: sexually transmitted disease.

[g]ICU: intensive care unit.

[h]rUTIs: recurrent urinary tract infections.

Symptom Awareness and Information Seeking

Many posts were from individuals seeking guidance or recommendations for when faced with UTI symptoms and methods for managing symptoms at home. Posts shared personal experiences (including physical and mental symptoms, HCP and patient experience, and impact on daily life and activities) with UTIs and expressed support for fellow individuals with UTIs. Patients value the ability to form virtual networks and communities with others facing the same challenges. Some posts provided recommendations for pain relief and symptom management strategies. One post stated, “So many [women] go through so much unnecessary pain and worsen their UTIs just because they think drowning themselves in cranberry juice will help. [...] Once you have one, your best friend is going to be antibiotics, pain killers, and WATER to flush out bacteria.”

Some posts highlighted a tendency for patients to seek guidance on the web prior to consulting an HCP. Some patients do this because they feel uncomfortable approaching an HCP or because they prefer to determine the severity of their symptoms on the web before seeking professional help. One post highlighted this hesitancy: “My friend recently got a bladder infection (we think) she’s a bit nervous about going to the doctor for various reasons [...] She doesn’t have any serious symptoms yet [...] but we just don’t want it to escalate and want some tips before going to a doctor.”

Notably, many posts stressed the importance of seeking professional medical help when confronted with UTI symptoms. One post said, “It’s pretty easy to treat but you do need to go to a doctor or urgent care [...] to be sure.”

HCP Interactions

Posts frequently discussed difficulties obtaining a medical appointment. Some patients highlighted long waiting times and others were advised to visit the emergency department, which was often deemed an inappropriate course of action. One post stated, “Doctor has suggested [low dose antibiotic] as I feel like I have cystitis symptoms all the time for last 2 months, got to wait 6 months to see urologist to see what problem is.” Patients who had attended appointments often discussed the tendency for HCPs—particularly males—to be dismissive of the impact of their UTI symptoms. Some patients equated this to trivializing their pain and discomfort. In contrast, female HCPs were reported to be more empathetic and demonstrated a deeper comprehension of the specific requirements of female patients. One post read, “Have your girlfriend see a FEMALE urologist, if you can. Male urologists unfortunately don’t always keep up with female urinary research.”

The medical approaches relayed in the posts demonstrated a lack of shared decision-making. Many HCPs prescribed treatments without engaging in comprehensive discussions with patients. This practice led patients to feel frustrated, particularly when treatment was ineffective or had unfavorable side effects. Some posts offered advice for navigating HCP appointments, including specific diagnostic evaluations or treatments to advocate for, and which to avoid. The importance of advocating for personal preferences was emphasized. One patient said, “I’ve learned that if I’m dealing with my [health] to push back if I feel I am not listened to.”

Diagnosis and Management Challenges

Many patients commented on complex diagnostic processes or a lack of diagnostic workup before receiving a prescription. The restricted sensitivity of urine culture tests was discussed (particularly for less prevalent strains or infections with minimal bacterial presence) and some posts advocated for more sensitive assessments such as the MicroGenDX diagnostics test (which combines quantitative polymerase chain reaction with 16s/18s next-generation sequencing). One post read, “Doing a more in-depth test cannot hurt [...] it’s sad all they do is dip stick, culture, cipro [...] I would have never found my staph infection in my bladder had I not tested further using the advanced next generation sequencing DNA test.” Patients also reported instances of misdiagnosis while experiencing UTI symptoms, including diagnoses of period cramps or sexually transmitted infections. One patient said, “I knew I had a UTI and the [...] doctor kept refusing to test saying it was probably [a sexually transmitted disease (STD)]. It was not an STD and when the results for the UTI came back it showed I had a staph infection in my bladder.” Misdiagnosis led to a loss of trust in HCPs.

Posts also discussed the provision of inadequate treatment by medical professionals. One read, “I went to the [doctor] because I suspected a bladder infection. Gave all my symptoms. [...] Only after pushing did he condescendingly ask if I’d feel better if I provided a urine sample. I said yes. He couldn’t look me in the eyes when he gave me a prescription for you [guessed] it, a bladder infection.”

Some patients felt that treatment decisions should be informed by comprehensive diagnostic evaluations. Instances of antibiotic prescribing without consideration of the specific bacteria present were highlighted. These situations often led to ineffective treatment that necessitated multiple rounds of therapy.

Management With Antibiotics

There was consensus among posts that antibiotics are necessary to cure UTIs and prevent associated complications.

XSL•FO
RenderX

Nitrofurantoin was the most frequently mentioned antibiotic, followed by ciprofloxacin, but many patients said that they were frustrated with medical interventions as they continue to experience UTIs despite good adherence. Few posts described situations where HCPs conducted microbiological cultures and sensitivity testing to identify the most appropriate antibiotic for their UTI. Some patients expressed a preference for more convenient doses, with 1 patient explaining that this was due to the need to "choke down pills [twice a day]" on [antibiotic].

Side effects of antibiotic therapy were also discussed and concerns regarding antibiotic resistance led some patients to seek alternative therapy. One patient asked, "Just wondering if anyone has tried D-Mannose for recurrent UTI? I have a prescription for [antibiotic] [...] but I take it so often that I'm worried about antibiotic resistance. [...] I'm trying to find something to help as a preventative."

Patients who chose not to take certain classes of antibiotics reported opting for home remedies or other antibiotics instead. Discussed side effects included nausea, potential lung toxicity, and an increased risk of yeast infections. A small number of patients discussed HCP warnings about potential tendon-related side effects when taking certain antibiotics. Patients also discussed concerns regarding the associated black box warning and some reported experiencing leg and back pain while on treatment. One post said, "It has 3 black box warnings and has caused healthy people horrible side effects." Another said, "[Antibiotic] is causing me a lot of pain in the left leg and I'm scared to continue it." Within the posts, there were no references to shared decision-making when prescribing antibiotics.

Self-Management

Over-the-counter and home remedies for UTI were the most common topics among posts but opinions on them varied. Some patients preferred self-management as it helped them feel in control of their treatment. These patients may have also felt frustrated with the care they received or with having to return to the doctor multiple times. One patient said, "I used to get UTIs, and having to go to the doctor placing me on different kinds of antibiotics. Some worked, some didn't so I wound up having to go back to the doctor which winds up to be very expensive."

Patients with a positive view of home remedies shared their experiences with nonprescribed treatment options, including hydration, D-mannose, cranberry, probiotics, and herbal supplements (uva ursi). Opinions on cranberry juice and supplements were more divided; this seemed to be the case among both patients and HCPs. One post about cranberry products read, "If you are female, you are supposed to go to the doctor, but honestly most just drink water and low sugar cranberry juice." Another said, "Stop just using cranberry juice every time you have a UTI!! So many [women] go through so much unnecessary pain and worsen their UTIs just because they think drowning themselves in cranberry juice will help."

Challenges With Recurrent UTI

Posts about recurrent UTI consistently highlighted the burden of persistent and incapacitating symptoms and their association with feelings of anxiety and helplessness. One post read, "No day to breathe without wondering if this will be when I go septic and die[.] I almost wish that this will kill me at this point." Factors contributing to the emotional toll of recurrent UTI included a perceived loss of bodily control and a negative impact on intimate relationships. Some posts discussed the use of long-term, low-dose antibiotics to prevent recurrence but lacking effectiveness and side effects were also mentioned. Many patients turned to over-the-counter remedies such as D-mannose and cranberry products to prevent recurrence. Success with these approaches was limited. One patient said, "I've struggled with recurrent UTIs for about 15 years and tried every single supplement and herb I could find info about. I often react very poorly to antibiotics [...] I'll do anything to cure UTIs at home before resorting to antibiotics." Other patients reported lifestyle changes (such as reducing caffeine and alcohol intake and stringent hygiene routines) to combat recurrent UTI and some reported positive outcomes of these changes.

Quantitative Outcomes: Antibiotics, Home Remedies, and HCP Specialties

All countries (Canada, Germany, India, the United Kingdom, and the United States) were included in the analysis of 9619 posts mentioning antibiotics, home remedies, and HCP specialists. Most posts (7270/9619, 76% of those included in the quantitative analysis) discussed antibiotics only, 13% (1290/9619) mentioned home remedies only, and 11% (1059/9619) mentioned both antibiotics and home remedies, as shown in Figure 2. The proportions of posts mentioning antibiotics and home remedies were similar across Canada, the United Kingdom, and the United States, but specific drugs (including antibiotics and pain relief) were discussed less frequently in the posts than home remedies. Cranberry and D-mannose were the most discussed home remedies.

Figure 3 shows that, among posts mentioning HCP specialties, specialist HCPs (urologists, obstetricians or gynecologists, and urogynecologists) were mentioned more frequently in posts from the United States (670/875, 77%) and Canada (44/71, 62%) than from the United Kingdom (329/646, 51%). Only 3% (21/646) of posts from the United Kingdom discussed nurse practitioners or urgent care versus 27% (19/71) in Canada and 22% (190/875) in the United States. Mentions of pharmacists were also much more frequent in Canada (10/71, 14%) than in the United Kingdom (14/646, 2%) or the United States (18/875, 2%).

XSL•FO
RenderX

Figure 2. Proportions of posts mentioning antibiotics only, home remedies only, or both antibiotics and home remedies. Home remedies included D-mannose, cranberry products, probiotics, and herbal supplements.

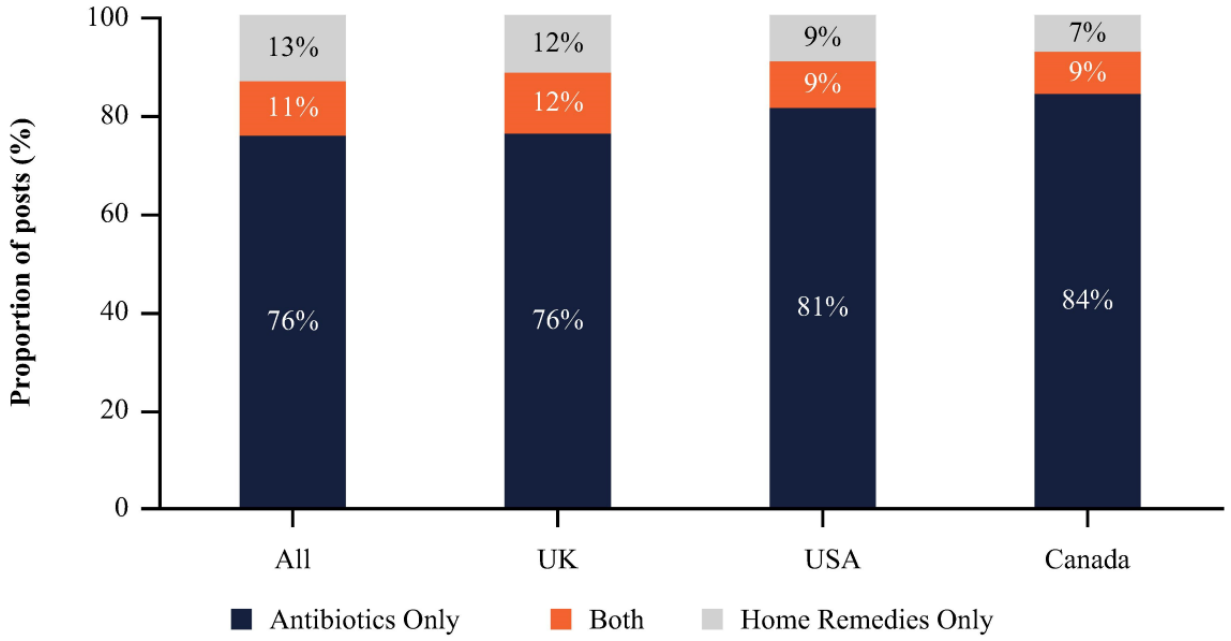

Figure 3. Proportions of posts mentioning health care professional specialties. OBGYN: obstetrician/gynecologist. UroGyn: urogynecologist.

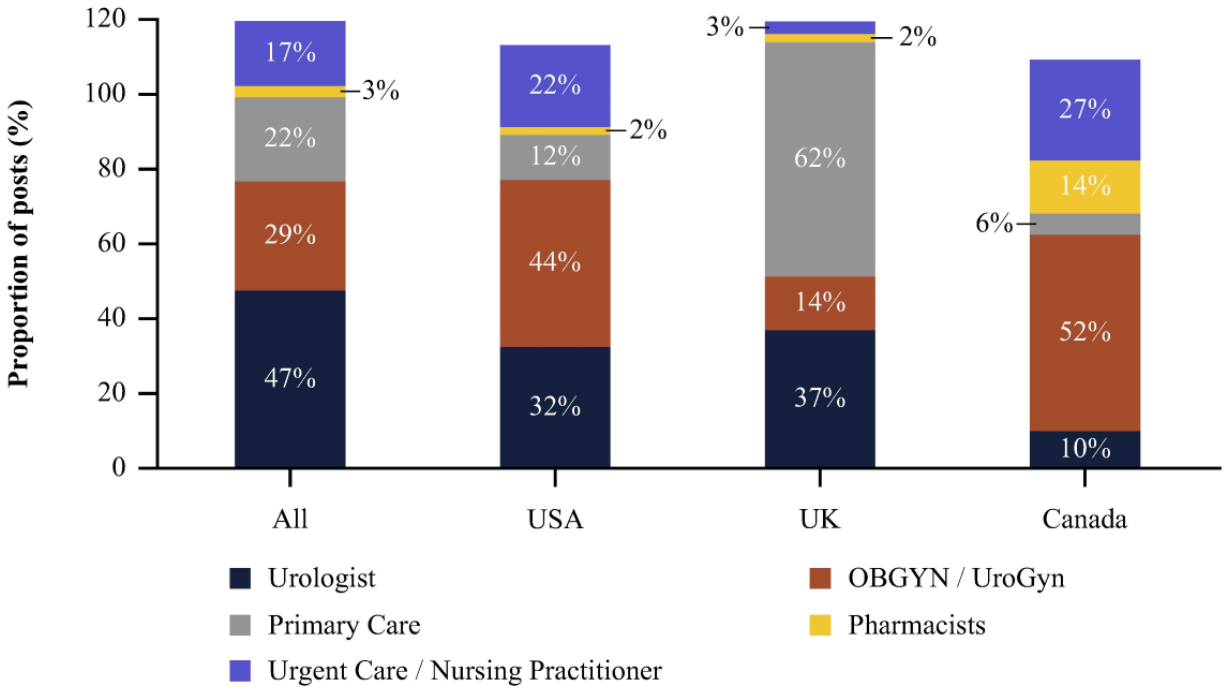

Discussion

Principal Findings

This retrospective social media study highlighted the strength of the web-based uUTI community, in which patients provide one another with emotional support and recommendations based on personal experience. Topics and behaviors observed in this study were somewhat aligned with those noted in the only preexisting, large-scale social media analysis of uUTI, including information-seeking, antibiotic treatment, and alternative therapy [18]. In addition to these topics, our study raised some further themes, including discussions about interactions with HCPs, as also observed in the study by Flower et al [8]. These insights have important implications for the patient journey and challenges associated with recurrence and should supplement existing data to provide a more holistic view of patient experiences and opinions of uUTI.

Patients in this study tended to search for advice on the web before consulting an HCP, and most posts stressed the importance of seeking professional help. Nevertheless, not all patients obtained medical appointments easily or were satisfied with their HCP interactions once an appointment was secured.

XSL•FO
RenderX

Patients also reported that some HCPs lacked empathy and knowledge of the potential emotional, mental, and physical impact of uUTIs on patients. Others were frustrated with treatments prescribed following inadequate discussions. These sentiments are reflected in the findings of previous qualitative studies of UTI [7,11]. In one such study, 35% of 466 survey respondents highlighted dismissive or unempathetic HCP behavior and many described difficulties accessing care [11]. Another study noted miscommunication between patients and HCPs as a barrier to appropriate prescribing [7]. Effective HCP communication regarding the advantages and drawbacks of certain therapies is essential to foster meaningful shared decision-making. Some posts also highlighted female HCPs as more empathetic than males. This is important because HCP empathy can improve patient satisfaction, increase adherence, and enhance outcomes [25,26]. Patients in a prior clinical study based in an obstetrics and gynecology department reported preferences for female HCPs because of their empathy and competence [27].

Most patients in this analysis also discussed the importance of antibiotics to cure their UTI. However, some patients had concerns about antibiotic use related to persistent symptoms and to the knowledge that repeated use of antibiotics may lead to antibiotic resistance, prompting them to seek alternative therapy. Hesitance in using antibiotics has also been associated with concerns that antibiotics may not be effective in the future; some women would rather understand how and why they got the UTI and how to prevent it in the future [7]. These concerns could be alleviated by ensuring that patients receive appropriate information on prevention and potential antimicrobial resistance, are prescribed the most appropriate treatment (especially those at risk of developing symptoms due to antibiotic-resistant bacteria), and feel empowered to engage in shared decision-making during treatment discussions with HCPs. Risk factors associated with antibiotic resistance in UTI include older age, which could be due to previous antibiotic exposure [28], and recurrent UTI [29].

Fears of antibiotic resistance in UTI may be linked to the common misconception that antibiotic resistance is a quality intrinsic to patients, rather than to the bacteria causing an infection [7,18]. This phenomenon may explain the prior observation that women with UTI are generally willing to accept treatment options with a longer time to resolution or a higher risk of side effects to avoid resistance [12]. Combined with factors such as hesitancy to see an HCP and frustration with care, fears of resistance may also contribute to the prevalence of posts about home remedies for UTI. This was the most common topic in our study. Many posts highlighted experiences with cranberry products and D-mannose, which are supported by mixed (positive and negative) and low-level clinical trial evidence, respectively [30]. In fact, a prior social media study identified that 53% of web-based recommendations had some supporting evidence (including cranberry products and D-mannose) but 43% were not supported by science. These unsupported recommendations included hygiene routines and dietary modifications such as avoiding alcohol and caffeine [16]. Web-based recommendations should be taken with care as the people advising may not be medical professionals. According to the European Association of Urology Guidelines on Urological Infections, clinicians may recommend cranberry products and D-mannose to reduce recurrent UTI episodes in patients, providing they inform them of the weak evidence base supporting this [31]. Overall, there are few studies and a lack of robust data investigating these recommendations, and this is an ongoing challenge in UTI research. It is interesting that, despite lacking placebo-controlled data and evidence for these approaches, a substantial number of posts in our study reported success when using them to prevent recurrent UTI; however, the true efficacy of these measures using placebo-controlled trials is undetermined.

The quantitative results obtained were unsurprising and reflect geographical variations in access to health and everyday practice behaviors. For example, while patients in Canada and the United States can access specialists without needing a referral, in the United Kingdom, National Health Service specialist appointments require a referral from a general practitioner. This is reflected in the high proportion of patients mentioning primary care (402/646, 62%) in UK posts. Similarly, pharmacists were mentioned more frequently in Canada because of the expanded scope of their role, which includes the capacity to prescribe antibiotics for UTI, as opposed to countries where it does not.

Patients from this analysis tended to seek information via social media before approaching an HCP, which underscores a potentially important opportunity for information outreach. Patients seeking information related to health online is becoming increasingly popular [32]; this need is emphasized by the fact that many patients lack preexisting knowledge regarding the cause, diagnosis, and treatment of UTIs [33]. Currently, however, many posts on social media sites are from sources that are not peer-reviewed and may contain misinformation [16,34]. Literature has highlighted that poor health literacy may affect the usefulness of social media interventions, which may result in health inequity [35]. One analysis on the video-sharing platform TikTok found that 75% (224/300) of identified videos about antibiotics were from non-HCPs, but HCP videos were significantly more relevant, significantly less biased, and significantly more likely to have an educational focus [34]. This highlights the need for accurate and successful advocacy for patients. An emphasis on creating HCP-led educational content for these platforms may help patients who seek health information via social media to identify the correct care sooner. Educating patients in this way could improve their communications with HCPs and drive shared decision-making. However, it is crucial to note that some HCPs may lack up-to-date knowledge on the testing and treatment approaches for uUTI, especially recurrent UTI, leading to patients feeling misunderstood [14]. It is important that both HCPs and patients are educated on uUTI and shared decision-making. This approach might also limit the potential for negative patient-HCP interactions driven by the acquisition of suboptimal information from non-HCPs on the web [36]. One notable obstacle may be the lack of social media use among HCPs: a 2021 analysis of 650 HCPs at 10 top American hospitals showed that more than 70% had at least 1 social media profile but among these, almost 90% had not posted in the prior month [37].

XSL•FO
RenderX

Limitations

This study had some limitations that must be considered when interpreting the outcomes. Demographic data were not collected and most content was posted by users in Canada, the United Kingdom, and the United States (mainly due to the inclusion or exclusion criteria for the posts); views from some patient groups and from other countries may be underrepresented. One patient group that may be particularly underrepresented is older patients, who may use social media less than their younger counterparts. It is also possible that patients posting on social media may have had more troublesome UTI experiences (such as more impactful symptoms or a worse treatment journey) than those who chose not to post on social media. While every effort was made to limit results to those associated with uUTI, some cases of complicated UTI or comorbidities may have been included, which could have an impact on the analysis. Posts from closed forums were not included in the analysis. Valentine-King et al [11] reported that surveys may be distributed using an international internet platform, reaching a larger group of patients, not limited to a particular practice, health network, or country. This may increase the variety of responses and experiences [11], although participation could still be limited for older patients or those with no or limited access to technology [38].

Conclusions

This analysis of more than 42,000 social media posts provides valuable insight into the experiences and opinions of patients with uUTI in Canada, Germany, India, the United Kingdom, and the United States. Many patients sought information on the web before seeking professional medical care, which emphasizes the importance of peer support in the UTI patient journey and the value of accessible web-based resources for patient education. Women reported difficulties accessing care and inadequate interactions with HCPs, characterized by dismissive behavior and lacking diagnostic evaluations. Although the importance of antibiotics was widely appreciated, many patients noted hesitance toward them or dissatisfaction with their efficacy and side effects. Patients highlighted fears and concerns about risks of antimicrobial resistance which, if proactively discussed by HCPs with patients, could further improve patient-HCP interactions. Alternative treatments were discussed frequently but not all were supported by scientific evidence. Overall, the insights gleaned in this study provide a more complete picture of patient behaviors and highlight the potential for HCP and patient education in addition to improved communication with HCPs to improve care.

Acknowledgments

Medical writing support for the development of this manuscript, under the guidance of the authors, was provided by Rosie Robson, MSc, of Ashfield MedComms, an Inizio company, and funded by GSK. Trademarks are the property of their respective owners (metronidazole [Flagyl]; Sanofi).

Conflicts of Interest

MLK is the CEO and founder of Live UTI Free Ltd. None declared for JMP and NK. AM and AB are employed by, and hold financial equities in, GSK. This research was funded by GSK.

Multimedia Appendix 1
Patient lay summary.
[DOCX File , 14 KB - jmir_v27i1e58882_app1.docx]

References

1. Foxman B, Barlow R, D'Arcy H, Gillespie B, Sobel JD. Urinary tract infection: self-reported incidence and associated costs. Ann Epidemiol 2000 Nov;10(8):509-515. [doi: 10.1016/s1047-2797(00)00072-7] [Medline: 11118930]
2. Foxman B, Brown P. Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect Dis Clin North Am 2003 Jun;17(2):227-241. [doi: 10.1016/s0891-5520(03)00005-9] [Medline: 12848468]
3. Medina M, Castillo-Pino E. An introduction to the epidemiology and burden of urinary tract infections. Ther Adv Urol 2019;11:1756287219832172 [FREE Full text] [doi: 10.1177/1756287219832172] [Medline: 31105774]
4. Wagenlehner F, Wullt B, Ballarini S, Zingg D, Naber KG. Social and economic burden of recurrent urinary tract infections and quality of life: a patient web-based study (GESPRIT). Expert Rev Pharmacoecon Outcomes Res 2018 Feb;18(1):107-117. [doi: 10.1080/14737167.2017.1359543] [Medline: 28737469]
5. Grigoryan L, Mulgirigama A, Powell M, Schmiemann G. The emotional impact of urinary tract infections in women: a qualitative analysis. BMC Womens Health 2022 May 18;22(1):182 [FREE Full text] [doi: 10.1186/s12905-022-01757-3] [Medline: 35585572]
6. Thompson J, Marijam A, Mitrani-Gold FS, Wright J, Joshi AV. Activity impairment, health-related quality of life, productivity, and self-reported resource use and associated costs of uncomplicated urinary tract infection among women in the United States. PLoS One 2023;18(2):e0277728 [FREE Full text] [doi: 10.1371/journal.pone.0277728] [Medline: 36724152]

XSL•FO
RenderX

7. Lecky DM, Howdle J, Butler CC, McNulty CA. Optimising management of UTIs in primary care: a qualitative study of patient and GP perspectives to inform the development of an evidence-based, shared decision-making resource. Br J Gen Pract 2020 May;70(694):e330-e338 [FREE Full text] [doi: 10.3399/bjgp20X708173] [Medline: 32041765]
8. Flower A, Bishop FL, Lewith G. How women manage recurrent urinary tract infections: an analysis of postings on a popular web forum. BMC Fam Pract 2014 Sep 26;15:162 [FREE Full text] [doi: 10.1186/1471-2296-15-162] [Medline: 25260870]
9. Leydon GM, Turner S, Smith H, Little P, UTIS team. The journey from self-care to GP care: a qualitative interview study of women presenting with symptoms of urinary tract infection. Br J Gen Pract 2009 Jul;59(564):e219-e225 [FREE Full text] [doi: 10.3399/bjgp09X453459] [Medline: 19566988]
10. Cox SML, van Hoof MWEM, Lo-A-Foe K, Dinant G, Oudhuis GJ, Savelkoul P, et al. Cross-sectional internet survey exploring women's knowledge, attitudes and practice regarding urinary tract infection-related symptoms in the Netherlands. BMJ Open 2022 May 18;12(5):e059978 [FREE Full text] [doi: 10.1136/bmjopen-2021-059978] [Medline: 35584871]
11. Valentine-King M, Laytner L, Hines-Munson C, Olmeda K, Trautner B, Justice S, et al. Qualitative Analysis of a Twitter-Disseminated Survey Reveals New Patient Perspectives on the Impact of Urinary Tract Infection. Antibiotics (Basel) 2022 Nov 23;11(12):1687 [FREE Full text] [doi: 10.3390/antibiotics11121687] [Medline: 36551344]
12. van der Worp H, Brandenbarg D, Boek PA, Braams JHW, Brink LJF, Keupers J, et al. Identifying women's preferences for treatment of urinary tract infection: a discrete choice experiment. BMJ Open 2021 Nov 16;11(11):e049916 [FREE Full text] [doi: 10.1136/bmjopen-2021-049916] [Medline: 34785550]
13. Krishnaswamy PH, Middleton E, Hagen S, Guerrero KL, Booth J. Women's Experiences of Urinary Tract Infections and Impact on Life: An Exploratory Qualitative Study. Urogynecology (Phila) 2024 Jan 01;30(1):80-86. [doi: 10.1097/SPV.0000000000001386] [Medline: 37493246]
14. Maxwell K, Roberts L, Kramer M, Price J, Newlands A, Finlay KA. Psychosocial burden and healthcare disillusionment in recurrent UTI: a large-scale international survey of patient perspectives. Front. Urol 2023 Sep 20;3. [doi: 10.3389/fruro.2023.1264299]
15. Dixon SJ. Number of global social network users 2017-2027. 2023. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2024-12-17]
16. Burton CS, Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, et al. Female Lower Urinary Tract Symptom Prevention and Treatment Strategies on Social Media: Mixed Correlation With Evidence. Urology 2021 Apr;150:139-145 [FREE Full text] [doi: 10.1016/j.urology.2020.06.056] [Medline: 32673678]
17. Chen J, Wang Y. Social Media Use for Health Purposes: Systematic Review. J Med Internet Res 2021 May 12;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
18. Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, Almario CV, et al. Using Large-scale Social Media Analytics to Understand Patient Perspectives About Urinary Tract Infections: Thematic Analysis. J Med Internet Res 2022 Jan 25;24(1):e26781 [FREE Full text] [doi: 10.2196/26781] [Medline: 35076404]
19. Uncomplicated urinary tract infections: developing drugs for treatment guidance for industry services. US Department of Health and Human Services; Food and Drug Administration. 2019. URL: https://www.fda.gov/media/129531/download [accessed 2024-12-17]
20. McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv Preprint posted online February 2, 2018. [doi: 10.48550/arXiv.1802.03426]
21. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv Preprint posted online March 11, 2022 [FREE Full text] [doi: 10.48550/arXiv.2203.05794]
22. Campello RJGB, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: Pei J, Tseng VS, Cao L, Motoda H, Xu G, editors. Advances in Knowledge Discovery and Data Mining: PAKDD. Cham: Springer; 2013:160-172.
23. Pharmacy first: what you need to know. Department of Health and Social Care. 2024. URL: https://healthmedia.blog.gov.uk/2024/02/01/pharmacy-first-what-you-need-to-know/ [accessed 2024-12-17]
24. EphMRA Code of Conduct. 2024. URL: https://www.ephmra.org/sites/default/files/2024-09/2024%20EPHMRA%20Code%20of%20Conduct%20Final.pdf [accessed 2025-02-24]
25. Howick J, Steinkopf L, Ulyte A, Roberts N, Meissner K. How empathic is your healthcare practitioner? A systematic review and meta-analysis of patient surveys. BMC Med Educ 2017 Aug 21;17(1):136 [FREE Full text] [doi: 10.1186/s12909-017-0967-3] [Medline: 28823250]
26. Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general practice: a systematic review. Br J Gen Pract 2013 Jan;63(606):e76-e84 [FREE Full text] [doi: 10.3399/bjgp13X660814] [Medline: 23336477]
27. Riaz B, Sherwani NZF, Inam SHA, Rafiq MY, Tanveer S, Arif A, et al. Physician Gender Preference Amongst Females Attending Obstetrics/Gynecology Clinics. Cureus 2021 May 14;13(5):e15028 [FREE Full text] [doi: 10.7759/cureus.15028] [Medline: 34136321]
28. Trautner BW, Kaye KS, Gupta V, Mulgirigama A, Mitrani-Gold FS, Scangarella-Oman NE, et al. Risk Factors Associated With Antimicrobial Resistance and Adverse Short-Term Health Outcomes Among Adult and Adolescent Female Outpatients With Uncomplicated Urinary Tract Infection. Open Forum Infect Dis 2022 Dec;9(12):ofac623 [FREE Full text] [doi: 10.1093/ofid/ofac623] [Medline: 37065971]

XSL•FO
RenderX

29. Osthoff M, McGuinness SL, Wagen AZ, Eisen DP. Urinary tract infections due to extended-spectrum beta-lactamase-producing Gram-negative bacteria: identification of risk factors and outcome predictors in an Australian tertiary referral hospital. Int J Infect Dis 2015 May;34:79-83 [FREE Full text] [doi: 10.1016/j.ijid.2015.03.006] [Medline: 25769526]
30. Konesan J, Liu L, Mansfield KJ. The Clinical Trial Outcomes of Cranberry, D-Mannose and NSAIDs in the Prevention or Management of Uncomplicated Urinary Tract Infections in Women: A Systematic Review. Pathogens 2022 Dec 05;11(12):1471 [FREE Full text] [doi: 10.3390/pathogens11121471] [Medline: 36558804]
31. Bonkat G, Bartoletti R, Bruyère F, Cai T, Geerlings S, Köves B, et al. European Association of Urology Guidelines on Urological Infections: Summary of the 2024 Guidelines. Eur Urol 2024 Jul;86(1):27-41 [FREE Full text] [doi: 10.1016/j.eururo.2024.03.035] [Medline: 38714379]
32. Jia X, Pang Y, Liu LS. Online Health Information Seeking Behavior: A Systematic Review. Healthcare (Basel) 2021 Dec 16;9(12):1740 [FREE Full text] [doi: 10.3390/healthcare9121740] [Medline: 34946466]
33. Cox S, Vleeming M, Giorgi W, Dinant GJ, Cals J, de Bont E. Patients' Experiences, Expectations, Motivations, and Perspectives around Urinary Tract Infection Care in General Practice: A Qualitative Interview Study. Antibiotics (Basel) 2023 Jan 24;12(2):241 [FREE Full text] [doi: 10.3390/antibiotics12020241] [Medline: 36830152]
34. Evans E, Gory LB, O'Kane A. TikTok: An Opportunity for Antibiotic Education? Innov Pharm 2022;13(4) [FREE Full text] [doi: 10.24926/iip.v13i4.4916] [Medline: 37305589]
35. Rivera-Romero O, Gabarron E, Miron-Shatz T, Petersen C, Denecke K. Social Media, Digital Health Literacy, and Digital Ethics in the Light of Health Equity. Yearb Med Inform 2022 Aug;31(1):82-87 [FREE Full text] [doi: 10.1055/s-0042-1742503] [Medline: 35654433]
36. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16(1):442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
37. Hameed I, Oakley CT, Ahmed A, Naeem N, Robinson NB, Hameed NUF, et al. Analysis of Physician Use of Social Media. JAMA Netw Open 2021 Jul 01;4(7):e2118213 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18213] [Medline: 34292338]
38. Wilkerson JM, Iantaffi A, Grey JA, Bockting WO, Rosser BRS. Recommendations for internet-based qualitative health research with hard-to-reach populations. Qual Health Res 2014 Apr;24(4):561-574 [FREE Full text] [doi: 10.1177/1049732314524635] [Medline: 24623662]

Abbreviations

HCP: health care professional
UTI: urinary tract infection
uUTI: uncomplicated urinary tract infection

Edited by A Mavragani; submitted 28.03.24; peer-reviewed by J Cuppen, D Lecky; comments to author 10.09.24; revised version received 29.11.24; accepted 04.12.24; published 11.03.25.

Please cite as:
Kramer ML, Polo JM, Kumar N, Mulgirigama A, Benkiran A
Living With and Managing Uncomplicated Urinary Tract Infection: Mixed Methods Analysis of Patient Insights From Social Media
J Med Internet Res 2025;27:e58882
URL: https://www.jmir.org/2025/1/e58882
doi:10.2196/58882
PMID:

©Melissa L Kramer, Jose Medina Polo, Nishant Kumar, Aruni Mulgirigama, Amina Benkiran. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Promoting Public Engagement in Palliative and End-of-Life Care Discussions on Chinese Social Media: Model Development and Analysis

Yijun Wang[1], MSc; Han Zheng[2], PhD; Yuxin Zhou[3], MSc; Emeka Chukwusa[3], PhD; Jonathan Koffman[4], PhD; Vasa Curcin[1], PhD

[1]Department of Population Health Sciences, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom

[2]School of Information Management, Wuhan University, Wuhan, China

[3]Cicely Saunders Institute of Palliative Care, King's College London, London, United Kingdom

[4]Wolfson Palliative Care Research Centre, Hull York Medical School, University of Hull, Hull, United Kingdom

Corresponding Author:
Han Zheng, PhD
School of Information Management
Wuhan University
Room 614
Luojiashan Road 299, Wuchang District
Wuhan, 430072
China
Phone: 86 18523102827
Email: hanzheng@whu.edu.cn

Abstract

Background: In Chinese traditional culture, discussions surrounding death are often considered taboo, leading to a poor quality of death, and limited public awareness and knowledge about palliative and end-of-life care (PEoLC). However, the increasing prevalence of social media in health communication in China presents an opportunity to promote and educate the public about PEoLC through online discussions.

Objective: This study aimed to examine the factors influencing public engagement in PEoLC discussions on a Chinese social media platform and develop practice recommendations to promote such engagement.

Methods: We gathered 30,811 PEoLC-related posts on Weibo, the largest social media platform in China. Guided by the elaboration likelihood model, our study examined factors across 4 dimensions: content theme, mood, information richness, and source credibility. Content theme was examined using thematic analysis, while sentiment analysis was used to determine the mood of the posts. The impact of potential factors on post engagement was quantified using negative binomial regression.

Results: Organizational accounts exhibited lower engagement compared to individual accounts (incidence rate ratio [IRR]<1; *P*<.001), suggesting an underuse of organizational accounts in advocating for PEoLC on Weibo. Posts centered on PEoLC-related entertainment (films, television shows, and books; IRR=1.37; *P*<.001) or controversial social news (IRR=1.64; *P*<.001) garnered more engagement, primarily published by individual accounts. An interaction effect was observed between content theme and post mood, with posts featuring more negative sentiment generally attracting higher public engagement, except for educational-related posts (IRR=2.68; *P*<.001).

Conclusions: Overall, organizations faced challenges in capturing public attention and involving the public when promoting PEoLC on Chinese social media platforms. It is imperative to move beyond a traditional mode to incorporate cultural elements of social media, such as engaging influencers, leveraging entertainment content and social news, or using visual elements, which can serve as effective catalysts in attracting public attention. The strategies developed in this study are particularly pertinent to nonprofit organizations and academics aiming to use social media for PEoLC campaigns, fundraising efforts, or research dissemination.

(J Med Internet Res 2025;27:e59944) doi:10.2196/59944

XSL•FO
RenderX

KEYWORDS

palliative care; end-of-life care; health promotion; social media; China; Weibo; public engagement; elaboration likelihood model; ELM

Introduction

Background

Palliative and end-of-life care (PEoLC) provides physical, psychological, and spiritual support to patients living with life-limiting conditions and terminal diseases, improving their quality of life and assisting their families [1]. Despite its importance, global public awareness and knowledge about PEoLC is limited [2-4], especially in China. Within the context of traditional Chinese culture, discussions about death and dying have historically been considered taboo [5]. As a result, individuals nearing the end of their lives often face societal and familial moral constraints, hindering their ability to autonomously make end-of-life decisions [6]. China has the largest aging population, with a notably increasing rate [7]. However, in 2015, the quality of China's palliative care was ranked merely 71st among 80 countries or regions worldwide [8]. The coverage rate of PEoLC service in China is only 1% [9], substantially lower than in Europe, America, and other high-income regions [10]. Studies [11,12] suggested that over 80% of Chinese patients with cancer are unaware of palliative care, although the majority support it and desire more information. Hence, there is an urgent need for public education and promotional efforts about PEoLC to address the increasing unmet demands.

Social media provides a channel for PEoLC promotion and education. Kate Granger, an English geriatrician, shared her stories at the end-of-life and described the involvement of palliative care during her patient journey on Twitter (since rebranded to X), attracting around 48,000 followers [13]. Taubert et al [14] produced and delivered video resources on YouTube and Twitter to improve communication in "do not attempt cardiopulmonary resuscitation" decisions in palliative illness in 2016. After less than 6 months, the video reached nearly 100,000 hits and the comments were extremely positive and encouraging. During the COVID-19 pandemic, people actively participated in a series of tweet chats using the Twitter hashtags #comcomcovid, #PallANZ and #PalliCovid hosted by different PEoLC organizations [15,16]. These online chats helped to support both professionals and community capacity for end-of-life care within the emerging landscape of the COVID-19 pandemic. These studies indicated that social media promotion for PEoLC could help to increase public knowledge and awareness, reduce the stigma of talking about death and dying and attract more engagement. The WHO public health model of palliative care emphasizes the importance of educating both the public and professionals to facilitate palliative care integration into health care systems [17]. Social media, with its vast reach, serves as a unique platform for disseminating PEoLC-related information to a broad audience, including young people, who are often future caregivers, patients, practitioners, and policymakers. As critical stakeholders, young social media users are also highly active in disseminating information, amplifying the reach of PEoLC education. Engaging them in this dialogue not only informs their own decision-making but also promotes greater public awareness [18]. Moreover, social media helps bridge the knowledge gap between professionals and the public, making PEoLC education more accessible and aligned with contemporary public health advocacy.

However, to date, research on PEoLC promotion has predominantly focused on English-language platforms in developed countries (eg, the United States and the United Kingdom) [19]. It is essential to extend PEoLC promotion efforts through social media to underserved regions, for example, China. Social media has demonstrated effectiveness in promoting health equity in populations at risk for disadvantage [20]. Therefore, social media may consequently help address the knowledge gap of the public about PEoLC in underserved regions.

Evidence suggests that public participation in interactions on social media will increase effective cognition [21]. It has been studied that public engagement on social media can improve the awareness of health-related information, a sense of belonging, and social connection [20]. Hence, public engagement behaviors with PEoLC-related content on social media, such as commenting, liking and forwarding PEoLC posts, are likely to increase public awareness and knowledge of PEoLC and make changes in PEoLC-related behaviors. It is necessary to understand how to promote public engagement with PEoLC discussions on social media. This study aims to investigate for the first time the factors influencing public engagement of PEoLC posts and explores best practice recommendations for enhancing such engagement on Chinese social media.

Conceptual Model and Research Hypotheses

The elaboration likelihood model (ELM) of persuasion is a theory of information processing that suggests there are 2 routes, central and peripheral routes, to persuasion [22]. In line with the ELM, we systematically distinguished between 2 information processing routes for individual engaging decisions (to engage or not): the peripheral route and the central route. ELM indicates individuals tend to show a higher likelihood of elaboration when they have strong motivations and the ability and individual knowledge to thoughtfully consider the underlying value of social media posts. Here, the central route of the ELM can be used to explain the engaging decision of the users. In contrast, when individuals' motivations and abilities are limited, the likelihood of elaboration is low and individuals will, in turn, rely on simple cues and second-hand information. Therefore, the peripheral route of the ELM can elucidate users' engaging decisions.

Importantly, past research suggested that post content characteristics generally influenced motivation and the ability to engage with posts that lead to the central route [23,24]. Common content characteristics include content theme and content mood [25-29]. Content theme is defined as the overarching subject, topic, or central idea that is being conveyed

XSL•FO
RenderX

or discussed within the textual content of a post. It is now well established from a variety of studies that content theme plays a critical role in increasing health-related post engagement. Holly and colleagues [25] studied 500 diabetes-related Facebook posts to explore the effect of content themes on engagement behavior. They observed posts related to diabetes symptoms information and positive self-identity had higher sharing while posts related to negative facial expressions, social support, and crowdsourcing had higher commenting. Their findings were consistent with another study [30] that focused on cancer-related posts on Sina Weibo, a Chinese social media platform, which found that social support and cancer treatment attracted the highest number of users and received the greatest number of retweets, comments, and likes. Another recent study [26] collected the COVID-19 pandemic-related videos on one Chinese streaming video platform and identified video content theme emerged as the most critical factor influencing engagement. They observed that emotional-focused messages that psychologically coped with the COVID-19 pandemic situation have more attractiveness compared with problem-focused messages that physically specified risks and strategies. Taken together, these studies suggested content theme is one of the important influencing factors of post engagement, and emotional content (such as social supports) tended to have more engagement in health care communication. Hence, we proposed the first hypothesis:

- The content themes have different effects on public engagement of PEoLC posts on the Chinese social media platform.

Another important central factor of post engagement is content mood. Mood is defined as the emotional tone or atmosphere conveyed by the post content. When individuals are exposed to emotional content on social media, they are likely to share their mood and this further promotes their engagement behaviors, such as sharing, liking, and commenting [31,32]. Previous research established that the influence of post mood differs within various contexts. Bhattacharya et al [27] focusing on Facebook posts from US health agencies accounts found that posts which expressed positive sentiment generated more engagement. However, another study [28] which also focused on US health agencies' accounts on Twitter found that post mood was not significantly associated with the level of engagement. Xu and Zhang [29] demonstrated that tweets with positive emotions increased the level of retweets, but content showing anger reduced those retweets. Thus, the effect of mood factors differed between diverse platforms, regions, and health topics. This led to our second hypothesis:

- The mood positively affects public engagement of PEoLC posts on the Chinese social media platform.

Furthermore, post mood was possible to moderate the effects of other influencing factors on post engagement. Chen et al [33] analyzed the content published by the Chinese government's official TikTok account during the COVID-19 pandemic. They demonstrated that the relationship between video content theme and citizen engagement was moderated by the emotional valence of video titles. For instance, videos related to the government handling of the pandemic and guidance for stakeholders that positive titles received a greater number of shares. Likewise, Tang et al [34] supported this view that for a certain type of social media content, positive sentiment can achieve a higher level of engagement behavior. Another study on Sina Weibo [35] found that news related to the COVID-19 pandemic crisis evoking negative emotions had higher engagement than those with positive emotions. Accordingly, the third hypothesis was proposed to test the moderating effect of mood on content theme and post engagement:

- Post mood has a moderating effect on the relationship between content theme and public engagement of PEoLC posts on the Chinese social media platform.

Previous literature [36-40] suggested information richness and source credibility commonly influence post engagement through the peripheral route. Information richness is defined as visual or textual formatting elements (such as visual cues, and external links) that contribute to the comprehensiveness and vividness of information within a post [41]. Many recent studies have shown that information richness may positively affect post engagement. Bhattacharya et al [27] analyzed >45,000 Facebook posts from health agency accounts. They identified Facebook posts with more visual cues, such as photos generated more engagement. This is supported by others who observed significant and robust positive effects of visual cues on post engagement [36,37]. During the COVID-19 pandemic, Ngai et al [38] examined 608 the COVID-19 pandemic posts published by an official news account and found links to external sources fostered sharing behaviors. This led to our fourth hypothesis:

- The information richness positively affects public engagement of PEoLC posts on the Chinese social media platform.

Source credibility is defined as the attractiveness, trustworthiness, and expertise of a post's publisher according to the source credibility model [42-44]. In the online environment, source credibility was evaluated on the basis of the social cues provided by the publisher in the site network, such as publisher type. Existing literature indicated that user engagement can be affected by different kinds of publishers (eg, health agencies, influencers, and noninfluencers) [39,40,45]. For instance, Kostygina et al [45] tried to investigate source features of Twitter messages in an antitobacco campaign and found using social influencers as message sources could generate more sharable content (eg, campaign hashtags) and greater volume of tweets and reach per day. Xu et al [40] demonstrated that during the COVID-19 pandemic, original posts from official health care accounts on Sina Weibo were more likely to generate more user engagement compared with reposts. Therefore, we proposed the fifth hypothesis:

- The source credibility feature, such as the publisher types have different effects on the public engagement of PEoLC posts on the Chinese social media platform.

We attempted to understand these influencing factors from the peripheral and the central route. Two dimensions (content theme and mood) of the central route and 2 dimensions (information richness and source credibility) of the peripheral route were selected and tested as shown in Figure 1. Furthermore, according

XSL•FO
RenderX

to ELM theory, individuals' decisions are influenced by their motivation and ability, which determine the different processing routes undertaken. Therefore, in turn, we can also gain some basic ideas of social media users' motivations and ability for PEoLC content from their preference to the peripheral or the central route when engaging with PEoLC discussion.

Figure 1. A conceptual framework to explore factors influencing post engagement.

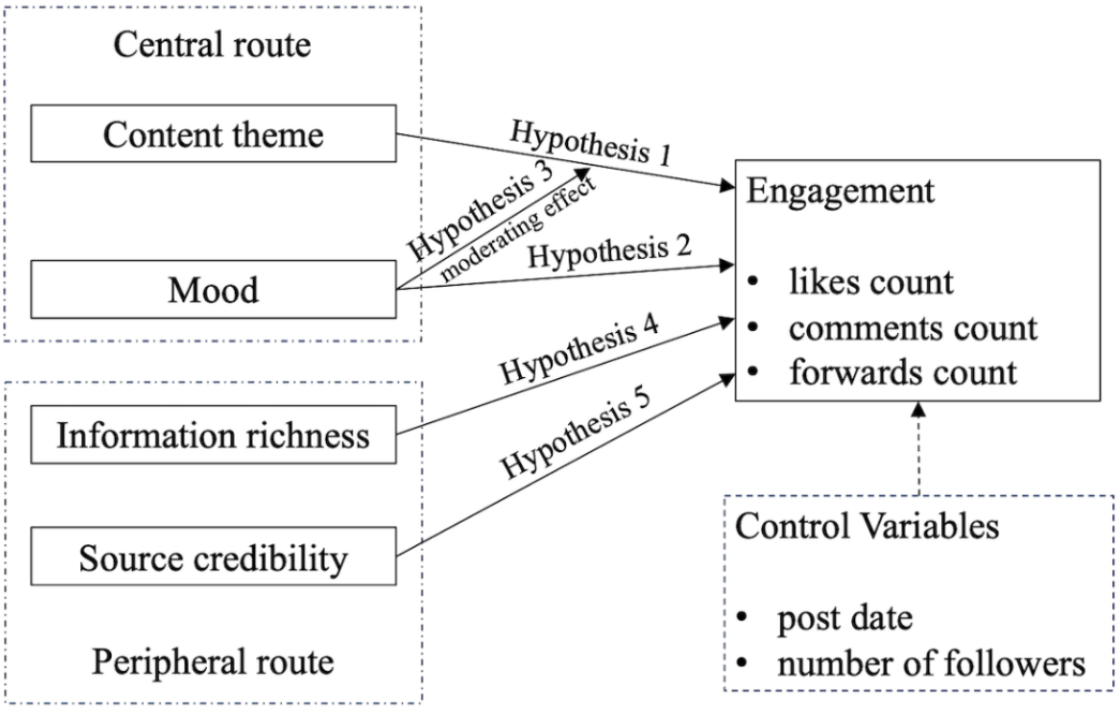

Methods

Data Collection and Preprocessing

We collected Weibo posts dating from September 1 2009 to December 12 2022 through Python (version 3.8.3; Python Software Foundation). Sina Weibo (hereafter referred to as Weibo; Weibo Corporation) is a major social media platform in China, with >575 million monthly active users as of 2024 [46]. It plays a key role in shaping digital discourse and influencing cultural trends in the country. Often described as China's equivalent of Twitter, Weibo allows users to create short-form content and engage in interactive discussions. User interactions on Weibo primarily include 3 key actions: sharing, commenting, and liking posts. Figure 2 illustrates the typical structure of a Weibo post, providing an example of the platform's interface and its user interaction mechanisms.

Keywords that were used to retrieve palliative care–related posts, such as "姑息治疗 (palliative care)," "临终关怀 (hospice care)," "生命终末期 (end-of-life)" are listed in Multimedia Appendix 1. To ensure the comprehensive inclusion of palliative care–related keywords, we (1) summarized common keywords used in previous studies [47-49]; (2) consulted experts in palliative care; and (3) conducted a "snowball" search strategy (starting with a small set of search keywords and iteratively expanding it by examining posts retrieved by these keywords and identifying additional keywords). Since the original concepts were in English, YW translated these concepts into simplified Chinese and engaged 2 bilingual Chinese speakers (HZ and YZ) to check translations. Disagreements in wording and meaning were resolved through further discussion.

Data preprocessing was further conducted. First, we detected and deleted posts published by bot accounts. A bot account on a social media platform entails an automated digital entity capable of emulating human interactions, and its potential harms encompass the dissemination of misinformation, manipulation of public discourse, and distortion of authentic engagement dynamics within the digital sociocultural landscape [50]. We identified one account as a bot account if it is abnormally active (published >30 posts a day) [51]. Second, we coped with missing data and outliers. We deleted posts with missing values or outliers in any variables in metadata. There were 2 reasons for doing so. The first was filling in these missing values with a replaced value will inevitably lead to bias. The second was that the total amount of records was large enough to tolerate a very small part of records being deleted [52,53]. Third, we manually went through the rest of the posts and excluded 1565 posts which were irrelevant to PEoLC. Some posts made jokes using PEoLC-related terms, for example: "My phone doesn't work anymore. I need hospice care for my phone." Since they cannot contribute to our understanding of PEoLC online discussion, we decided to consider them as noisy data and deleted these posts. Subsequently, we obtained 30,811 palliative care–related Chinese posts published by 20,764 distinct users. All Python (Python Software Foundation) and R scripts (R Foundation for Statistical Computing) used for data collection, preprocessing, and analysis were available from a GitHub (GitHub Inc) repository [54].

XSL•FO
RenderX

Figure 2. An example of Weibo posts.



Operationalization of Variables

Public engagement with PEoLC discussion on Weibo was the dependent or response variable of our study and it was reflected in 3 dimensions: sharing, liking, and commenting behaviors. Public engagement could be used to measure the popularity and attractiveness of social media posts [55]. If a post had high engagement on social media, then we call this post *viral* and it has higher exposure. We captured this objective data using web crawlers. An example of how the objective data were shown in a Weibo post is illustrated in Figure 2. In our study, public engagement in a Weibo post (engagement) was calculated by summing the values of these 3 indicators, a common practice in other social media research [55], as follows:

$$\text{Engagement}_i = \text{likes count}_i + \text{shares count}_i + \text{comments count}_i$$

Content theme was assessed based on 7 thematic categories, encompassing: personal experiences or stories about PEoLC, educating, discussion about entertainment work (film, television series, and books), reflections on palliative care and death, PEoLC activities or achievements, controversial social news, and fundraising. YW and YZ applied thematic analysis [56] to identify themes. This followed a multistage process, commencing with data familiarization, wherein researchers immersed themselves in the dataset to gain a deep understanding of the content. Subsequently, we generated initial codes, followed by theme development, constant comparison, and refinement of themes to ensure their coherence and relevance. We specifically applied thematic analysis to the content pertinent to PEoLC, ignoring the potential presence of irrelevant content within a post. In cases where a post encompassed multiple themes, the predominant theme was determined on the basis of the theme exhibiting the most distinct characteristics. To ensure analytical rigor, the level of agreement between the 2 reviewers (YW and YZ) for the thematic analysis was determined using Cohen κ. Intercoder reliability by κ coefficient was 0.87 (95% CI 0.89-0.85; P<.001). A κ statistic of 0.81 to 1.00 indicates an “almost perfect” strength of agreement. The findings, emerging themes, and their interpretations are presented in Table 1, which provided an in-depth exploration of PEoLC-related topics discussed on Chinese social media.

XSL•FO
RenderX

Table 1. Content themes of palliative and end-of-life care (PEoLC) related Weibo posts.

Themes	Example quotes
Theme 1: personal experiences or stories	
Talking about individuals' experiences with palliative care	"I went to Puren Hospice Hospital for a consultation. The doctor said that my dad can be admitted to their hospital, and if the ward is available my mom can also live in, so that the patient can enjoy the last time with their families. I felt that I have finally found a direction to solve the problem I have been suffering from for a long time."
Theme 2: educating	
Elucidating the scientific aspects of palliative care, including its definition, meaning, and development status	"Hospice care is to provide physical, psychological and spiritual care and humanistic services to help patients spend the last journey of their life with dignity by controlling their pain and discomforting symptoms and guiding their family members to get rid of their sadness as soon as possible before the end of their life."
Theme 3: discussion about entertainment work (film, television series, and books)	
Discussing palliative care–related films, television series, or books	"It's quite good to have this type of TV drama. I was even more shocked by the story in episode 7, which was really good. It is about death, the deceased and their families, the funeral home staff, the hospice...We can discuss a lot and feel strength and warmth from the sadness."
Theme 4: reflections on palliative care and death	
Reflecting on death or palliative care–related topics	"I am speechless. China is lack of death education and Chinese avoid to face up to life and death, which leads to the next generation who do not understand the value of life and then do not know how to care for the dying. For the sake of 'filial piety,' human suffering is prolonged. Repressing human nature is the most antihuman."
Theme 5: activities or achievements	
Reports about palliative care activities or achievements	"From 11th to 17th October, the Hospice of the Affiliated Hospital of Guangdong Medical University held a World Palliative Care Day publicity campaign with the theme of 'Love each other, spread palliative care to help what we need.' The activities included community and hospital visits and stalls, with the aim of promoting the importance of hospice care, paying attention to the quality of life of poor advanced cancer patients and their families."
Theme 6: controversial social news	
A controversial social phenomenon related to palliative care (eg, the legalization of euthanasia in certain countries or protest against the establishment of a hospice center)	"Recently, a community in Xi'an met with collective opposition from owners when it built a care center for the disabled elderly. Many residents said that the hospice affected the quality of life of the residents and hoped that the property owners would choose a new site."
Theme 7: fundraising	
Reports on fundraising initiatives and solicits contributions	"The monthly donation programme is now live! Become a monthly donor for palliative care, have one less cup of coffee a month, and share your love and warmth to the children who are sick and undergoing chemotherapy."

Mood was measured by a post's sentiment score. The sentiment score, ranging from 0 to 1, conveyed the negative sentiment to the positive sentiment of a post. We obtained the content of all posts through web crawlers and then applied sentiment analysis to calculate the sentiment scores through Baidu application programming interface (API). The Baidu API, based on deep neural network, was selected since it has better performance in judging the sentiment of palliative care–related posts compared with other popular Chinese sentiment analysis models such as SnowNLP. SnowNLP intrinsically recognized terms such as "death," and "dying" as negative terms, while these terms were comparatively neutral in the Baidu API. For instance, a post saying "对于剩余寿命无多的终末期患者，姑息治疗主要为患者提供临终关怀及善终服务" ("For patients who are terminally ill with little time left to live, palliative care focuses on providing end-of-life care and hospice services") obtained a sentiment score of 0.92 through Baidu API while obtaining a sentiment score of 0.07 through SnowNLP.

Information richness was measured by 4 indicators: whether a post had pictures or videos or not, whether a post used hashtags or not, whether a post tagged other users or not and the number of Chinese characters included in a post [38,57]. We obtained the content of all posts through web crawlers first. Then, we applied regular expression in Python to detect pictures, videos, webpage links, hashtags, and tagging other users and applied the count function in Python to calculate the number of Chinese characters.

Source credibility was measured by a post's publisher type. In empirical social media research, publisher type is frequently used as a proxy for source credibility, as it is readily available and easy to operate [58-60]. While this measure provides a general indication of perceived credibility, it does not encompass the full spectrum of attributes, such as trustworthiness, expertise, and attractiveness. The highly subjective nature of these attributes, coupled with the variability of social media platforms,

XSL•FO
RenderX

complicates efforts to capture them through standardized metadata alone. Indeed, previous research has identified a lack of consensus on how to operationalize these measures across platforms [61]. Hence, this study used publisher type to measure source credibility. There were 6 types of publishers: individual users (15,356/30,811, 49.84%), influencers (5312/30,811, 17.24%), governments and hospitals (3265/30,811, 10.6%), enterprises (3133/30,811, 10.17%), media and press (2196/30,811, 7.13%), nonprofit organizations (1549/30,811, 5.03%). We captured the publisher type shown on the publisher's profile using web crawlers. The difference between individual users and influencers is the social media impact. Influencers are individuals who have a significant following on social media and can influence the opinions and behaviors of their followers. Therefore, they have a greater social media impact compared with other individual users.

In addition, the number of days a post was published and the number of followers that the publisher had were extracted as controlled variables to account for their potential influence on the dependent variable based on previous literature [62,63]. Descriptive statistics of the study variables are summarized in Table 2.

XSL•FO
RenderX

Table 2. Descriptive statistics of study variables.

Construct and description	Values
Total number of posts, n (%)	30,811 (100)
Content theme variables	
Content theme, n (%)	
Theme 1: personal experiences or stories	6400 (20.77)
Theme 2: educating	4546 (14.75)
Theme 3: discussion about entertainment work (films, television series, and books)	4588 (14.89)
Theme 4: reflections on palliative care and death	5352 (17.37)
Theme 5: activities or achievements	7741 (25.12)
Theme 6: controversial social news	767 (2.49)
Theme 7: fundraising	1417 (4.6)
Mood variable	
Sentiment score (ranges from 0, negative, to 1, positive), mean (SD)	0.74 (36)
Information richness variables	
Post used hashtags or not, n (%)	
Yes	10,275 (33.35)
No	20,536 (66.65)
Post tagged other users or not, n (%)	
Yes	4355 (14.13)
No	26,456 (85.86)
Video and image count (number of videos and images included in a post), mean (SD)	1.27 (2.08)
Character count (number of Chinese characters included in a post), mean (SD)	248.3 (366.43)
Source credibility variables	
Publisher type, n (%)	
Individual users	15,356 (49.84)
Influencers	5312 (17.24)
Governments and hospitals	3265 (10.6)
Enterprises	3133 (10.17)
Media and press	2196 (7.13)
Nonprofit organizations	1549 (5.03)
Controlled variables	
Days (number of days a post was published), mean (SD)	1283 (998.04)
Follower count (number of followers the publisher had), mean (SD)	7.63 (3.45)
Dependent variable or response quantity	
Comment count (number of comments a post received), mean (SD)	6.182 (123.18)
Like count (number of likes a post received), mean (SD)	18.51 (565.39)
Forward count (number of forwards a post received), mean (SD)	15.29 (449.35)
Engagement (number of comments, likes, and forwards a post received), mean (SD)	39.98 (844.61)

Statistical Analysis

Our dependent variable was engagement that exhibited overdispersion (engagement: mean 39.98, SD 844.61; skewness=62.09, kurtosis=5150.66). A significant number of Weibo posts rarely received shares, likes, or comments, while others received a high number in comparison. To deal with this overdispersed count data, we modeled the engagement using negative binomial regression. We first estimated the impact of information richness, source credibility, mood, and content theme on public engagement. The regression model was as follows:

XSL•FO
RenderX

$$\ln(\text{engagement}_i) = \text{Intercept} + b_1\text{content theme}_i + b_2\text{mood}_i + b_3\text{information richness}_i + b_4\text{source credibility}_i$$

We then explored whether the impacts were contingent upon the emotional valence of each post's content theme. The regression model was as follows:

$$\ln(\text{engagement}_i) = \text{Intercept} + b_1\text{content theme}_i + b_2\text{mood}_i + b_3\text{mood}_i \times \text{content theme}_i + b_4\text{information richness}_i + b_5\text{source credibility}_i$$

All analyses were conducted using R (version 4.0.2; R Foundation for Statistical Computing).

Ethical Considerations

As this study used publicly available data, it was exempt from institutional ethics review. To protect participant confidentiality, all personally identifiable information in original posts was anonymized before analysis. Data retention adhered to the principle of minimal necessary storage. In addition, Chinese posts included in the manuscript were translated into English and paraphrased so that they cannot be traced back to the original users.

Results

Descriptive Analysis

A total of 30,811 Weibo posts published from September 1, 2009, to December 12, 2022, were used for building regression models. The distribution of these posts' engagements was largely dispersed, with 14,344 (46.55%) posts receiving 0 engagements (ie, 0 likes, 0 comments, and 0 shares), 15,516 (50.36%) posts receiving 0 to 100 engagements and 951 (3.09%) posts receiving >100 engagements. The annual average engagements of PEoLC-related Weibo posts increased every year from 10.09 in 2012 to 65.57 in 2022. The average engagement of posts varied between post content themes and publisher types. Posts related to theme 6 (controversial social news) had the highest average engagement (mean 205.82, SD 3539.17), followed by posts related to theme 1 (personal experiences or stories; mean 51.28, SD 837.21). Posts related to theme 2 (educating) received the least average engagements (mean 12.18, SD 254.08). Posts published by the media and press attracted the most average engagements (mean 116.17, SD 2326.55), followed by influencers (mean 112.69, SD 1224.51). Posts published by the government and hospital accounts received the least average engagements (mean 7.40, SD 45.19).

Hypothesis Testing

Table 3 presents the estimates of 2 negative binomial regression models that predicted the engagement of PEoLC-related posts on the Chinese microblogging platform—Weibo. Hypothesis 1 assumed the content themes could have different effects on post engagement. As the content theme was a categorical variable, we treated personal experiences or stories as the reference group. Model 1 presented in Table 3 showed that discussion about entertainment work (films, television series, and books; incidence rate ratio [IRR]=1.37; $P<.001$) and controversial social news (IRR=1.64; $P<.001$) were positively correlated with post engagement, compared with personal experiences or stories. Posts related to educating (IRR=0.48; $P<.001$), reflections on palliative care and death (IRR=0.77; $P<.001$), PEoLC activities or achievements (IRR=0.64; $P<.001$), and fundraising (IRR=0.14; $P<.001$) were negatively associated with post engagement. Consequently, hypothesis 1 was supported.

XSL•FO
RenderX

Table 3. Results of the negative binomial regression model 1 and model 2 (with interaction terms).

Variables	Model 1		Model 2	
	IRR[a]	*P* value	IRR	*P* value
Intercept	0.27	<.001	0.24	<.001
Content theme variables				
Theme 1: personal experiences or stories (RF[b])	—[c]	—	—	—
Theme 2: educating	0.48	<.001	0.22	<.001
Theme 3: discussion about entertainment work (films, television series, and books)	1.37	<.001	1.49	<.001
Theme 4: reflections on palliative care and death	0.77	<.001	0.98	.54
Theme 5: activities or achievements	0.64	<.001	0.61	<.001
Theme 6: controversial social news	1.64	<.001	2.12	<.001
Theme 7: fundraising	0.14	<.001	1.10	.76
Mood variable				
Sentiment score	0.78	<.001	0.80	<.001
Interaction effect				
Content theme × mood				
Theme 1 × sentiment score	—	—	—	—
Theme 2 × sentiment score	—	—	2.68	<.001
Theme 3 × sentiment score	—	—	0.84	.003
Theme 4 × sentiment score	—	—	0.66	<.001
Theme 5 × sentiment score	—	—	1.06	.45
Theme 6 × sentiment score	—	—	0.36	<.001
Theme 7 × sentiment score	—	—	0.11	<.001
Information richness variables				
Post used hashtags or not	1.39	<.001	1.34	<.001
Post tagged other users or not	5.54	<.001	5.94	<.001
Videos and images count	1.03	<.001	1.03	<.001
Characters count	1.00	<.001	1.00	<.001
Source credibility variables				
Individual users (RF)	—	—	—	—
Influencers	0.90	.01[d]	0.88	<.001
Governments and hospitals	0.21	<.001	0.17	<.001
Enterprises	0.56	<.001	0.54	<.001
Media and press	0.41	<.001	0.41	<.001
Nonprofit organizations	0.66	<.001	0.63	<.001
Controlled variables				
Days	1.00	<.001	1.00	<.001
Followers count	1.57	<.001	1.61	<.001

[a]IRR: incidence rate ratio.

[b]RF: reference factor.

[c]Not available.

Hypothesis 2 proposed the more positive the mood of a post, the more engagement it received. However, model 1 presented in Table 3 showed the sentiment score had a significantly negative association with post engagement (IRR=0.78; *P*<.001). The IRR meant a 1-unit increase in sentiment score could result

XSL•FO
RenderX

in a 22% decrease in the number of post engagements. Hence, hypothesis 2 could not be supported by our results.

To test hypothesis 3, we built model 2 to enter the interaction effects between content themes and post mood based on model 1. The results in Table 3, model 2, showed that post mood had a slight moderate effect on the relationship between content type and post engagement. Specifically, theme 2 with positive mood was positively associated with post engagement (IRR=2.68; *P*<.001). When we tested hypothesis 1, theme 2 (educating) was negatively associated with post engagement. This indicates post mood could moderate the effect of content theme on post engagement. Interestingly, only theme 2 with a positive mood was related to high engagement, the other topics with a negative mood were related to high engagement. Hence, hypothesis 3 was supported.

Hypothesis 4 proposed that information richness was positively associated with the level of public engagement of PEoLC posts, namely the post that had more information was likely to receive greater public engagement. Model 1 in Table 3 showed 4 variables representing information richness where all had positive and statistically significant associations with post engagement (used hashtags or not: IRR=1.39; *P*<.001; tagged other users or not: IRR=5.54; *P*<.001; videos and images count: IRR=1.03; *P*<.001; and characters count: IRR=1.00; *P*<.001). The IRR values indicated that posts with hashtags were associated with a 1.09-unit increase in public engagement, while posts tagging other users were linked to a 5.54-unit increase, both compared to posts without these respective features. Hypothesis 4 was therefore supported by the results.

Hypothesis 5 posited the level of public engagement of a PEoLC post was affected by the source, namely the type of publisher. Since publisher type was a categorical variable, we treated individual users as the reference group. Model 1 in Table 3 showed that when compared with individual users, the other publisher types were negatively and significantly associated with post engagement. The IRR values suggested that when compared to posts published by individual users, those from influencers (*P*=.01), governments and hospitals (*P*<.001), enterprises (*P*<.001), media and press (*P*<.001), and nonprofit organizations (*P*<.001) resulted in different levels of decrease in the number of engagements. Consequently, hypothesis 5 was supported.

Discussion

Principal Findings

This study represented the first attempt to identify how online discussions promote public engagement in PEoLC in China. Given the prevailing limited awareness of palliative care in low-income regions [64,65], the use of social media as a digital platform for promoting and educating the public about PEoLC is imperative. This study examined the multifaceted factors influencing public engagement with PEoLC-related posts from 4 dimensions (ie, content theme, mood, information richness and source), drawing upon the ELM to expound on these influences. Our results identified that these 4 dimensions exerted significant and distinct effects on engagement. These findings provided a valuable understanding of the existing levels of public engagement with PEoLC-related content on social media. Moreover, they offered guidance to content developers, including event organizers, nonprofit organizations, and academics, on leveraging social media to disseminate research, launch health campaigns, raise funds, and educate about PEoLC to the wider public.

First, our results identified that posts published by government entities, health care institutions, and organizations tend to attract comparatively lower levels of engagement when contrasted with posts from individual accounts. This highlights the potential underuse of organizational accounts in their role of public education and advocacy for PEoLC. Most organizational accounts still adhere to the conventional model of unidirectional communication on Weibo, characterized by didactic scientific presentations and formalized achievements reporting. Effective management and content curation of these accounts are paramount, as they hold the potential to both raise awareness about palliative care and then inspire charitable contributions in support of PEoLC [66].

Furthermore, our analysis indicated that the content theme significantly impacts the level of public engagement with posts. Specifically, we observed that engagement tended to be higher for themes related to entertainment and social matters, while it exhibited a slightly lower level of engagement for informational content. This finding aligns with previous research which demonstrated that content about entertainment on social media tends to elicit greater public engagement [67,68]. This pattern is not solely a manifestation of the intrinsic entertainment and sociable characteristics of social media platforms. It may also be attributed to the fact that leveraging entertainment works (films, books, or television) or current social news as a channel to contact participants offers an efficacious approach for educating individuals about sensitive health topics, such as palliative care within a real-world context [69-72]. By exploring the portrayal of palliative care in books, television dramas, or social news, the public has a chance to be exposed to the concepts of “hospice care” and “palliative care” among others, thereby fostering collective reflection and discussion. For instance, the Chinese television series “Song of Life,” which depicted a palliative professional’s working experiences prompted a discourse concerning the feasibility of establishing hospices in a residential area on Chinese social media. Recent years have witnessed China’s production of some high-quality film and television content about palliative care. Notably, our study reveals that the use of film and television productions to initiate public discourse and engagement on social media platforms has proven to be effective. Given the relatively low acceptance and awareness levels surrounding the topic of palliative care in low-income regions, entertainment education on social media may indeed offer a promising strategy for raising awareness and public understanding and decreasing inequality. The Lancet Commission on the Value of Death also highlighted the significance of the normalization of death communication through mass media and entertainment work [73]. Our results provide empirical evidence for this proposal.

We also identified that the positive mood of a post negatively impacted public engagement around PEoLC-related posts. This

XSL•FO
RenderX

is in contrast with previous studies in which a more positive mood in health-related posts attracts more engagement [27]. This phenomenon could be attributed to the sensitive nature of PEoLC, which is related to death and dying. In China, with a general lack of death education and the dominance of traditional Chinese ethical culture, topics related to palliative care are still seen as taboo. The Chinese tradition of filial piety, where children are expected to seek permission from their parents, exerts substantial societal, moral, and public opinion pressure on parental end-of-life decisions [12]. Consequently, it is still difficult for the public to view this concept positively. Therefore, a relatively negative or serious approach to the discussion of PEoLC is more acceptable to the public in the context of Chinese culture, even on social media. Beyond the Chinese cultural context, PEoLC-related posts may display distinctive attributes in comparison to more generalized health care content, warranting further investigation in this area.

Ultimately, our investigation revealed that information richness exerted a positive influence on engagement. Specifically, posts characterized by a greater volume of content, a higher frequency of tagging other users, the incorporation of more hashtags, and the inclusion of more images or videos were correlated with higher engagement levels. This discovery aligns with the outcomes of previous research [27,74-76] in health care communication. In doing so, our study further substantiates a previous observation within the domain of health care social media research, which underscores the notion that the more informative, diverse, and vivid the mode of expression, the greater the attention and engagement garnered by health-related communication in the realm of social media.

According to ELM theory, although post content and mood as central cues could change or form attitude with higher persistence, peripheral cues may play more crucial roles in the context of PEoLC promotion. The ELM theory emphasizes that if the public's motivation or ability to process the issue-relevant information is low, persuasion can occur by a peripheral route in which processes invoked by simple cues in the persuasion context influence attitudes [77]. Considering the relatively limited awareness and understanding of PEoLC among the Chinese public, stakeholders may benefit from using peripheral routes, such as information richness and source credibility when promoting PEoLC on online platforms.

Recent research suggests that older adults in China prefer accessing palliative care information through electronic media [78]. This finding underscores the potential of online platforms, including social media, as effective tools for promoting and educating the public about PEoLC. While older adults may not be frequent users of social media, young social media users, often their significant others, can play a critical role in disseminating PEoLC information, influencing health care decisions at both individual and community levels. This, in turn, can enhance the overall quality of end-of-life care within society. In the Chinese cultural context, where end-of-life care decisions often involve multiple stakeholders, including family members and health care providers, engaging diverse groups through social media can help facilitate the integration of PEoLC into the broader health care system. Social media platforms, therefore, hold significant potential for enhancing public understanding and improving the quality of care for vulnerable populations, such as the elderly.

Implications and Strategic Suggestions

On the basis of the study findings, we developed some practical strategies to help individuals or organizations to boost public engagement about PEoLC-related discussions or campaigns on social media platforms (Textbox 1 [79]). These strategies suggest content developers or account operators follow certain patterns during the content development process. These recommendations provide valuable guidance, particularly for organizational accounts, which may face challenges in attracting engagement. This bears special significance for nonprofit organizations, where social media accounts may serve as a pivotal channel for fundraising. As our findings were based on Chinese social media platforms, these strategies may work more effectively within similar cultural contexts and PEoLC development levels.

XSL•FO
RenderX

Textbox 1. Practical strategies to help individuals or organizations to boost public engagement about palliative and end-of-life care (PEoLC)-related discussions or campaigns on social media platforms.

- Use trending entertainment works and social news as a trigger—when developing PEoLC-related content on social media, the strategic incorporation of trending entertainment works and social news may be a compelling approach to enhance user engagement. Content developers can use entertainment works or social news as a lead-in to start their content or only attach these trending hashtags (such as #Song of Life##Do you agree to have a hospice in your community#) in their posts. The inherently captivating nature of trending topics in entertainment and social events provides a dynamic backdrop against which PEoLC-related content can be situated, sparking user curiosity and interaction. By using trending entertainment works and social news as catalysts, content creators can intrigue the audiences thinking in a real situation and thus achieve their intended purpose of promoting or educating.
- Try to engage influencers in promotion or education—influencers, often recognized for their expansive reach and persuasive capacities, offer a unique opportunity to enhance the visibility and impact of PEoLC messaging on social media. Concrete strategies to realize this encompasses collaborating with influencers who have a personal connection to PEoLC, encouraging them to share personal narratives, insights, and experiences, and endorsing PEoLC resources, materials, and campaigns on their channels. Organizations must realize that it is not enough to limit content distribution merely to their followers, who have already shown interest in the topic of PEoLC. Rather, organizations could interact with social media influencers who have much more diverse followers, to engage a wider audience with little knowledge of PEoLC. In addition, organizations could hold influencer-led awareness campaigns, live question and answer sessions, and educational webinars on both organizations' and influencers' accounts. By harnessing the influence of these advocates, organizations are more likely to significantly amplify their reach and raise public awareness of PEoLC on social media.
- Content works better with visual cues—visual elements, such as figures and videos, not only enhance the esthetic appeal of posts but also play a pivotal role in conveying complex concepts and eliciting emotive responses. Visual cues provide audiences with the opportunity to firsthand experience the dilemmas faced by stakeholders in PEoLC and immerse themselves in the personal narratives of these individuals, thus fostering an understanding of the profound significance and relevance of PEoLC. These digital affordances (eg, geotags, images, videos) provide vulnerable groups such as PEoLC stakeholders with an opportunity to gain more exposure on social media and then raise public awareness. Moreover, visual elements can transcend language barriers and resonate on a universal level, further broadening the reach and resonance of the PEoLC message globally. In this context, it is necessary for palliative care communication to incorporate visual elements to enhance understanding, foster empathy, and increase engagement on social media platforms.

Limitations

While the results of this research provide valuable insights, they must be viewed in light of certain limitations. The primary limitation pertains to the use of Weibo as the sole platform for analyzing PEoLC-related posts. While Weibo is one of the most extensively used social media platforms in China, with a user base exceeding 575 million monthly active users, it represents just one facet of the broader social media landscape. However, the platform's significant reach and influence in the Chinese context provide a robust foundation for the study, thereby mitigating the potential limitation of not, including other platforms. In addition, the process of keyword selection for retrieving relevant posts on Weibo presented its challenges. Despite our rigorous approach, using 3 diverse methods to ensure a comprehensive inclusion of keywords, there remains a marginal possibility that certain terms used in informal contexts might have been overlooked. It is also noteworthy that terms such as "euthanasia" were deliberately excluded from our search parameters. Although sometimes associated with hospice care in China, the term is often misinterpreted as "suicide" in social media discourse. This strategic exclusion was aimed at reducing the influx of irrelevant data, thereby enhancing the relevance and quality of our dataset. A further limitation is that the operationalization of source credibility based solely on publisher type may not fully capture the perceived trustworthiness, expertise, or attractiveness of the source. While publisher type is commonly used in previous studies, it provides only a partial view of credibility in social media environments. Future research should explore additional or alternative variables to more comprehensively assess source credibility. Besides, the extensive data collection period (2009-2022) introduces potential variability in the results due to the evolving social media landscape and changing user behaviors over time. However, existing literature suggests public attitudes and behaviors toward PEoLC remained relatively stable during this period [3,80]. While aggregating engagement factors over the years may potentially obscure period-specific patterns, it provides a comprehensive overview of long-term trends. Moreover, this extended timeframe ensures a substantial dataset, which improves the statistical power, generalizability, and robustness of the regression models [81]. Nevertheless, future studies may benefit from temporal segmentation to capture nuanced shifts in engagement patterns across different periods. We also consider a multimodel research by incorporating text, image, and video content analysis in detecting engagement patterns in the future.

Conclusions

While death is a universal and inevitable experience, discussions surrounding it are often avoided in China. This reluctance highlights the urgent necessity to enhance public understanding and awareness about PEoLC in the country. This study pinpoints social media as a promising tool for advancing PEoLC and igniting public engagement on related topics within the Chinese context. Drawing from the ELM theory, our study unveils distinct characteristics of PEoLC-related posts on social media that capture public attention. Acknowledging these unique traits, we have formulated targeted strategies aimed at bolstering public involvement in PEoLC discourse. These strategies include leveraging influencers, using entertainment or social news as catalysts, and integrating visual cues. By adopting these approaches, organizations can effectively communicate with the community, promoting PEoLC in a manner that is both accessible and compelling to the public. This study also holds the potential to catalyze a broader cultural shift toward open and constructive conversations about death and dying in China.

XSL•FO
RenderX

Acknowledgments

YW is funded by the Chinese Scholarship Council. VC is supported by the Engineering and Physical Sciences Research Council (EPSRC)–funded King's Health Partners Digital Health Hub (EP/X030628/1).

Data Availability

Given the potentially sensitive nature of the social media data, which may contain individual-specific information, the dataset generated and analyzed in this study are available from the corresponding author only on reasonable request.

Authors' Contributions

YW collected the data, conducted data analysis, and wrote the manuscript. HZ suggested the study methods and revised the manuscript. YZ involved in data analysis and revised the manuscript. JK provided valuable insights into the contextualization of the study findings and polished the manuscript. EC revised the manuscript. VC contributed to the conceptualization and revised the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search keywords used to retrieve related posts.
[DOCX File , 14 KB - jmir_v27i1e59944_app1.docx]

References

1. WHO definition of palliative care 2002. World Health Organization. URL: http://www.who.int/cancer/palliative/definition/en/ [accessed 2024-04-29]
2. Trivedi N, Peterson EB, Ellis EM, Ferrer RA, Kent EE, Chou WY. Awareness of palliative care among a nationally representative sample of U.S. adults. J Palliat Med 2019 Dec 01;22(12):1578-1582 [FREE Full text] [doi: 10.1089/jpm.2018.0656] [Medline: 31038384]
3. Patel P, Lyons L. Examining the knowledge, awareness, and perceptions of palliative care in the general public over time: a scoping literature review. Am J Hosp Palliat Care 2020 Jun 05;37(6):481-487. [doi: 10.1177/1049909119885899] [Medline: 31690088]
4. Flieger SP, Chui K, Koch-Weser S. Lack of awareness and common misconceptions about palliative care among adults: insights from a national survey. J Gen Intern Med 2020 Jul 10;35(7):2059-2064 [FREE Full text] [doi: 10.1007/s11606-020-05730-4] [Medline: 32157652]
5. Dong F, Zheng R, Chen X, Wang Y, Zhou H, Sun R. Caring for dying cancer patients in the Chinese cultural context: a qualitative study from the perspectives of physicians and nurses. Eur J Oncol Nurs 2016 Apr;21:189-196. [doi: 10.1016/j.ejon.2015.10.003] [Medline: 26547883]
6. Li J, Davis MP, Gamier P. Palliative medicine: barriers and developments in mainland China. Curr Oncol Rep 2011 Aug 13;13(4):290-294. [doi: 10.1007/s11912-011-0169-9] [Medline: 21487684]
7. The Lancet. Population ageing in China: crisis or opportunity? Lancet 2022 Nov;400(10366):1821. [doi: 10.1016/s0140-6736(22)02410-2]
8. The 2015 quality of death index: ranking palliative care across the world. LIEN Foundation. URL: https://www.lienfoundation.org/sites/default/files/2015%20Quality%20of%20Death%20Report.pdf [accessed 2024-04-29]
9. Yang J, Liu Y. Some thoughts on perfecting the legal system of hospice care in China. Med Law 2019;11(1):12-15.
10. Clark D, Baur N, Clelland D, Garralda E, López-Fidalgo J, Connor S, et al. Mapping levels of palliative care development in 198 countries: the situation in 2017. J Pain Symptom Manage 2020 Apr;59(4):794-807.e4 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.11.009] [Medline: 31760142]
11. Cheng Q, Duan Y, Zheng H, Xu X, Khan K, Xie J, et al. Knowledge, attitudes and preferences of palliative and end-of-life care among patients with cancer in mainland China: a cross-sectional study. BMJ Open 2021 Sep 28;11(9):e051735 [FREE Full text] [doi: 10.1136/bmjopen-2021-051735] [Medline: 34588259]
12. Yan Y, Zhang H, Gao W, Liu D, Endo M, Deshpande GA, et al. Current awareness of palliative care in China. Lancet Glob Health 2020 Mar;8(3):e333-e335. [doi: 10.1016/S2214-109x(20)30006-1]
13. Taylor J, Pagliari C. #Deathbedlive: the end-of-life trajectory, reflected in a cancer patient's tweets. BMC Palliat Care 2018 Jan 22;17(1):17 [FREE Full text] [doi: 10.1186/s12904-018-0273-9] [Medline: 29357865]

XSL•FO
RenderX

14. Taubert M, Norris J, Edwards S, Snow V, Finlay IG. Talk CPR - a technology project to improve communication in do not attempt cardiopulmonary resuscitation decisions in palliative illness. BMC Palliat Care 2018 Oct 19;17(1):118 [FREE Full text] [doi: 10.1186/s12904-018-0370-9] [Medline: 30340632]
15. Mills J, Li C, Fullerton S, Chapman M, Jap J, Sinclair C, et al. Staying connected and informed: online resources and virtual communities of practice supporting palliative care during the novel coronavirus pandemic. Prog Palliat Care 2020 Apr 28;28(4):251-253. [doi: 10.1080/09699260.2020.1759876]
16. Wang Y, Chukwusa E, Koffman J, Curcin V. Public opinions about palliative and end-of-life care during the COVID-19 pandemic: Twitter-based content analysis. JMIR Form Res 2023 Aug 07;7:e44774 [FREE Full text] [doi: 10.2196/44774] [Medline: 37368840]
17. Stjernswärd J, Foley KM, Ferris FD. The public health strategy for palliative care. J Pain Symptom Manage 2007 May;33(5):486-493 [FREE Full text] [doi: 10.1016/j.jpainsymman.2007.02.016] [Medline: 17482035]
18. Mallon A, Slater P, Hasson F, Casson K, McIlfatrick S. What do young adults know about palliative care? A cross-sectional survey. Public Health 2021 Feb;191:78-84. [doi: 10.1016/j.puhe.2020.11.023] [Medline: 33545498]
19. Wang Y, Koffman J, Gao W, Zhou Y, Chukwusa E, Curcin V. Social media for palliative and end-of-life care research: a systematic review. BMJ Support Palliat Care 2024 May 17;14(2):149-162 [FREE Full text] [doi: 10.1136/spcare-2023-004579] [Medline: 38594059]
20. Welch V, Petkovic J, Pardo Pardo J, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
21. Fischer E, Reuber AR. Social interaction via new social media: (how) can interactions on Twitter affect effectual thinking and behavior? J Bus Ventur 2011 Jan;26(1):1-18. [doi: 10.1016/j.jbusvent.2010.09.002]
22. Petty RE, Briñol P. The elaboration likelihood model. In: van Lange PA, Kruglanski AW, Higgins ET, editors. Handbook of Theories of Social Psychology. Thousand Oaks, CA: Sage Publications; 2011:224-245.
23. Shi J, Hu P, Lai KK, Chen G. Determinants of users' information dissemination behavior on social networking sites. Internet Res 2018 Apr 04;28(2):393-418. [doi: 10.1108/intr-01-2017-0038]
24. Zha X, Yang H, Yan Y, Liu K, Huang C. Exploring the effect of social media information quality, source credibility and reputation on informational fit-to-task: moderating role of focused immersion. Comput Human Behav 2018 Feb;79:227-237. [doi: 10.1016/j.chb.2017.10.038]
25. Rus HM, Cameron LD. Health communication in social media: message features predicting user engagement on diabetes-related Facebook pages. Ann Behav Med 2016 Oct 8;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
26. Li Y, Huang SK, Wei J, Brindle TR, Lee CY. What captures attention in the risk communication process: exploring streaming video attractiveness during the first wave of the COVID-19 pandemic in China. Comput Human Behav 2023 Dec;149:107909. [doi: 10.1016/j.chb.2023.107909]
27. Bhattacharya S, Srinivasan P, Polgreen P. Social media engagement analysis of U.S. Federal health agencies on Facebook. BMC Med Inform Decis Mak 2017 Apr 21;17(1):49 [FREE Full text] [doi: 10.1186/s12911-017-0447-z] [Medline: 28431582]
28. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on Twitter. PLoS One 2014 Nov 7;9(11):e112235 [FREE Full text] [doi: 10.1371/journal.pone.0112235] [Medline: 25379727]
29. Xu WW, Zhang C. Sentiment, richness, authority, and relevance model of information sharing during social crises—the case of #MH370 tweets. Comput Human Behav 2018 Dec;89:199-206. [doi: 10.1016/j.chb.2018.07.041]
30. Chen L, Wang P, Ma X, Wang X. Cancer communication and user engagement on Chinese social media: content analysis and topic modeling study. J Med Internet Res 2021 Nov 10;23(11):e26310 [FREE Full text] [doi: 10.2196/26310] [Medline: 34757320]
31. Ji YG, Chen ZF, Tao W, Li ZC. Functional and emotional traits of corporate social media message strategies: behavioral insights from S&P 500 Facebook data. Public Relat Rev 2019 Mar;45(1):88-103. [doi: 10.1016/j.pubrev.2018.12.001]
32. So J, Prestin A, Lee L, Wang Y, Yen J, Chou WY. What do people like to "share" about obesity? A content analysis of frequent retweets about obesity on twitter. Health Commun 2016 Jun 18;31(2):193-206. [doi: 10.1080/10410236.2014.940675] [Medline: 26086083]
33. Chen Q, Min C, Zhang W, Ma X, Evans R. Factors driving citizen engagement with government TikTok accounts during the COVID-19 pandemic: model development and analysis. J Med Internet Res 2021 Feb 04;23(2):e21463 [FREE Full text] [doi: 10.2196/21463] [Medline: 33481756]
34. Tang X, Li S, Gu N, Tan M. Exploring repost features of police-generated microblogs through topic and sentiment analysis. Electron Libr 2019 Aug 14;37(4):607-623. [doi: 10.1108/el-02-2019-0044]
35. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: how to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]

XSL•FO
RenderX

36. Li J, Zheng H, Duan X. Factors influencing the popularity of a health-related answer on a Chinese question-and-answer website: case study. J Med Internet Res 2021 Sep 28;23(9):e29885 [FREE Full text] [doi: 10.2196/29885] [Medline: 34581675]
37. Li Y, Xie Y. Is a picture worth a thousand words? An empirical study of image content and social media engagement. J Mark Res 2019 Nov 18;57(1):1-19. [doi: 10.1177/0022243719881113]
38. Ngai CS, Singh RG, Lu W, Koon AC. Grappling with the COVID-19 health crisis: content analysis of communication strategies and their effects on public engagement on social media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
39. Yang Y, Deng W, Zhang Y, Mao Z. Promoting public engagement during the COVID-19 crisis: how effective is the Wuhan Local Government's information release? Int J Environ Res Public Health 2020 Dec 26;18(1):118 [FREE Full text] [doi: 10.3390/ijerph18010118] [Medline: 33375307]
40. Xu Q, Shen Z, Shah N, Cuomo R, Cai M, Brown M, et al. Characterizing Weibo social media posts from Wuhan, China during the early stages of the COVID-19 pandemic: qualitative content analysis. JMIR Public Health Surveill 2020 Dec 07;6(4):e24125 [FREE Full text] [doi: 10.2196/24125] [Medline: 33175693]
41. Li K, Zhou C, Luo X, Benitez J, Liao Q. Impact of information timeliness and richness on public engagement on social media during COVID-19 pandemic: An empirical investigation based on NLP and machine learning. Decis Support Syst 2022 Nov;162:113752 [FREE Full text] [doi: 10.1016/j.dss.2022.113752] [Medline: 35185227]
42. Roy Dholakia R, Sternthal B. Highly credible sources: persuasive facilitators or persuasive liabilities? J Consum Res 1977 Mar;3(4):223. [doi: 10.1086/208671]
43. Ohanian R. The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. J Advert Res 1991;31(1):46-54 [FREE Full text]
44. Kardes FR, Cronley ML, Cline TW. Consumer Behavior. Cambridge, MA: South-Western Publishing; 2011.
45. Kostygina G, Tran H, Binns S, Szczypka G, Emery S, Vallone D, et al. Boosting health campaign reach and engagement through use of social media influencers and memes. Social Media Soc 2020 May 06;6(2):2362. [doi: 10.1177/2056305120912475]
46. Thomala LL. Number of monthly active users of Weibo Corporation from 2nd quarter of 2018 to 2nd quarter of 2024. Statista. URL: https://www.statista.com/statistics/795303/china-mau-of-sina-weibo/ [accessed 2024-09-20]
47. Zhao Y, Zhang H, Huo J, Guo Y, Wu Y, Prosperi M, et al. Mining Twitter to assess the determinants of health behavior towards palliative care in the United States. AMIA Jt Summits Transl Sci Proc 2020;2020:730-739 [FREE Full text] [Medline: 32477696]
48. Nwosu AC, Debattista M, Rooney C, Mason S. Social media and palliative medicine: a retrospective 2-year analysis of global Twitter data to evaluate the use of technology to communicate about issues at the end of life. BMJ Support Palliat Care 2015 Jun 02;5(2):207-212. [doi: 10.1136/bmjspcare-2014-000701] [Medline: 25183713]
49. Sladek R, Tieman J, Fazekas BS, Abernethy AP, Currow DC. Development of a subject search filter to find information relevant to palliative care in the general medical literature. J Med Libr Assoc 2006 Oct;94(4):394-401 [FREE Full text] [Medline: 17082830]
50. Orabi M, Mouheb D, Al Aghbari Z, Kamel I. Detection of bots in social media: a systematic review. Inf Process Manag 2020 Jul;57(4):102250. [doi: 10.1016/j.ipm.2020.102250]
51. Chu Z, Gianvecchio S, Wang H, Jajodia S. Detecting automation of twitter accounts: are you a human, bot, or cyborg? IEEE Trans Dependable Secure Comput 2012 Nov;9(6):811-824. [doi: 10.1109/tdsc.2012.75]
52. Raymond MR, Roberts DM. A comparison of methods for treating incomplete data in selection research. Educ Psychol Meas 1987 Mar 01;47(1):13-26. [doi: 10.1177/0013164487471002]
53. Strike K, El Emam K, Madhavji N. Software cost estimation with incomplete data. IIEEE Trans Software Eng 2001;27(10):890-908. [doi: 10.1109/32.962560]
54. Wang Y. Twitter_Palliative_care_analysis. GitHub. URL: https://github.com/mangoJ0106/Weibo_Palliative_care_analysis.git [accessed 2024-04-29]
55. Neiger BL, Thackeray R, Van Wagenen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar 01;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
56. Braun V, Clarke V. Thematic analysis. In: Cooper H, Coutanche MN, McMullen LM, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. New York, NY: American Psychological Association; 2012:65-81.
57. Zheng H, Goh DH, Lee EW, Lee CS, Theng YL. Understanding the effects of message cues on COVID-19 information sharing on Twitter. J Assoc Inf Sci Technol 2022 Jun 15;73(6):847-862 [FREE Full text] [doi: 10.1002/asi.24587] [Medline: 34901313]
58. Na S, Kunkel T, Doyle J. Exploring athlete brand image development on social media: the role of signalling through source credibility. Eur Sport Manag Q 2019 Sep 17;20(1):88-108. [doi: 10.1080/16184742.2019.1662465]
59. Lin X, Spence PR, Lachlan KA. Social media and credibility indicators: the effect of influence cues. Comput Human Behav 2016 Oct;63:264-271. [doi: 10.1016/j.chb.2016.05.002]

XSL•FO
RenderX

60. Borah P, Xiao X. The importance of 'likes': the interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. J Health Commun 2018 Mar 30;23(4):399-411. [doi: 10.1080/10810730.2018.1455770] [Medline: 29601271]
61. Jenkins EL, Ilicic J, Barklamb AM, McCaffrey TA. Assessing the credibility and authenticity of social media content for applications in health communication: scoping review. J Med Internet Res 2020 Jul 23;22(7):e17296 [FREE Full text] [doi: 10.2196/17296] [Medline: 32706675]
62. de Vries L, Gensler S, Leeflang PS. Popularity of brand posts on brand fan pages: an investigation of the effects of social media marketing. J Interact Mark 2012 May 01;26(2):83-91. [doi: 10.1016/j.intmar.2012.01.003]
63. Dolan R, Conduit J, Frethey-Bentham C, Fahy J, Goodman S. Social media engagement behavior: a framework for engaging customers through social media content. Eur J Mark 2019 Oct 07;53(10):2213-2243. [doi: 10.1108/ejm-03-2017-0182]
64. Powell RA, Mwangi-Powell FN, Radbruch L, Yamey G, Krakauer EL, Spence D, et al. Putting palliative care on the global health agenda. Lancet Oncol 2015 Feb;16(2):131-133. [doi: 10.1016/S1470-2045(15)70002-1] [Medline: 25638676]
65. Lamas D, Rosenbaum L. Painful inequities — palliative care in developing countries. N Engl J Med 2012 Jan 19;366(3):199-201. [doi: 10.1056/nejmp1113622]
66. Lovejoy K, Saxton GD. Information, community, and action: how nonprofit organizations use social media. J Comput Mediat Commun 2012 Apr 01;17(3):337-353. [doi: 10.1111/j.1083-6101.2012.01576.x]
67. Pletikosa Cvijikj I, Michahelles F. Online engagement factors on Facebook brand pages. Soc Netw Anal Min 2013 Jan 26;3(4):843-861. [doi: 10.1007/S13278-013-0098-8]
68. Tafesse W. Content strategies and audience response on Facebook brand pages. Mark Intell Plan 2022 Oct 01;33(6):927-943. [doi: 10.1108/MIP-07-2014-0135] [Medline: 35579975]
69. Corrigan PW, Morris SB, Michaels PJ, Rafacz JD, Rüsch N. Challenging the public stigma of mental illness: a meta-analysis of outcome studies. Psychiatr Serv 2012 Oct;63(10):963-973. [doi: 10.1176/appi.ps.201100529] [Medline: 23032675]
70. Glik D, Nowak G, Valente T, Sapsis K, Martin C. Youth performing arts entertainment-education for HIV/AIDS prevention and health promotion: practice and research. J Health Commun 2002 Nov 10;7(1):39-57. [doi: 10.1080/10810730252801183] [Medline: 11878569]
71. Lim RB, Tham DK, Cheung ON, Adaikan PG, Wong ML. A public health communication intervention using edutainment and communication technology to promote safer sex among heterosexual men patronizing entertainment establishments. J Health Commun 2019 Jan 30;24(1):47-64. [doi: 10.1080/10810730.2019.1572839] [Medline: 30698082]
72. Sharf BF, Freimuth VS, Greenspon P, Plotnick C. Confronting cancer on thirtysomething: audience response to health content on entertainment television. J Health Commun 1996 Oct 29;1(2):157-172. [doi: 10.1080/108107396128121] [Medline: 10947358]
73. Sallnow L, Smith R, Ahmedzai SH, Bhadelia A, Chamberlain C, Cong Y, et al. Report of the Lancet commission on the value of death: bringing death back into life. Lancet 2022 Feb;399(10327):837-884. [doi: 10.1016/s0140-6736(21)02314-x]
74. Moran G, Muzellec L, Johnson D. Message content features and social media engagement: evidence from the media industry. J Prod Brand Manag 2019 Oct 16;29(5):533-545. [doi: 10.1108/jpbm-09-2018-2014]
75. Dhanesh G, Duthler G, Li K. Social media engagement with organization-generated content: role of visuals in enhancing public engagement with organizations on Facebook and Instagram. Public Relat Rev 2022 Jun;48(2):102174. [doi: 10.1016/j.pubrev.2022.102174]
76. Carboni JL, Maxwell SP. Effective social media engagement for nonprofits: what matters? J Public Nonprof Aff 2015 Mar 02;1(1):18-28. [doi: 10.20899/jpna.1.1.18-28]
77. Petty RE, Brinol P, Priester JR. Mass Media Attitude Change: Implications of the Elaboration Likelihood Model of Persuasion. New York, NY: Routledge; 2009:141-180.
78. Lei L, Lu Y, Gan Q, Hu Z, Luo Y. Awareness and perceptions of palliative care among the elderly: a qualitative study. J Palliat Care 2022 Apr 23;37(2):204-212. [doi: 10.1177/08258597221082393] [Medline: 35195464]
79. Gurrieri L, Drenten J. Visual storytelling and vulnerable health care consumers: normalising practices and social support through Instagram. J Serv Mark 2019 Dec 03;33(6):702-720. [doi: 10.1108/jsm-09-2018-0262]
80. Pun JK, Cheung KM, Chow JC, Chan WL. Chinese perspective on end-of-life communication: a systematic review. BMJ Support Palliat Care 2020 Jul 09;14(e1):e30-e37. [doi: 10.1136/bmjspcare-2019-002166] [Medline: 32647031]
81. Cohen J, Cohen P, West S, Aiken L. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. 3rd edition. New York, NY: Routledge; 2013.

Abbreviations

API: application programming interface
ELM: elaboration likelihood model
IRR: incidence rate ratio
PEoLC: palliative and end-of-life care

XSL•FO
RenderX

Edited by X Ma; submitted 28.04.24; peer-reviewed by S Boonchutima, Y Cui; comments to author 09.09.24; revised version received 29.09.24; accepted 03.01.25; published 18.03.25.

Please cite as:
Wang Y, Zheng H, Zhou Y, Chukwusa E, Koffman J, Curcin V
Promoting Public Engagement in Palliative and End-of-Life Care Discussions on Chinese Social Media: Model Development and Analysis
J Med Internet Res 2025;27:e59944
URL: https://www.jmir.org/2025/1/e59944
doi:10.2196/59944
PMID:

©Yijun Wang, Han Zheng, Yuxin Zhou, Emeka Chukwusa, Jonathan Koffman, Vasa Curcin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Association Between Internet Addiction and the Risk of Suicide Attempts in Chinese Adolescents Aged 11-17 Years: Prospective Cohort Study

Sihong Li[1], PhD; Xingyue Jin[1], PhD; Lintong Song[1], PhD; Tianqing Fan[1], PhD; Yanmei Shen[1*], PhD; Jiansong Zhou[1*], PhD

Department of Psychiatry, National Clinical Research Center for Mental Disorders, and National Center for Mental Disorders, The Second Xiangya Hospital of Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Yanmei Shen, PhD
Department of Psychiatry, National Clinical Research Center for Mental Disorders, and National Center for Mental Disorders
The Second Xiangya Hospital of Central South University
No.139, Renmin Road Central Changsha-Hunan-China
Changsha, 410011
China
Phone: 86 15116269263
Email: ymshen@csu.edu.cn

Abstract

Background: Suicide is a critical public health issue in adolescents worldwide. Internet addiction may play a role in the increased rate of suicide attempts in this population. However, few studies have explored the relationship between pathological internet use and suicide attempts among adolescents.

Objective: This study aimed to conduct a prospective cohort study to examine whether higher severity of pathological internet use was associated with an increased risk of suicide attempts among Chinese adolescents.

Methods: A total of 782 adolescents were recruited from a middle school from November 2020 to December 2020 and followed up for 6 months. An online self-reported questionnaire was used to collect the participants' demographic data and assess their mental health. The Depression, Anxiety, and Stress Scale–21 items (DASS-21) was used to evaluate depression, anxiety, and stress. The Chen Internet Addiction Scale–Revised (CIAS-R) was used to assess the symptoms and severity of pathological internet use. χ^2 test and ANOVA were used for intergroup comparison, and logistic regression analysis was used to examine the relationship between the severity of pathological internet use and suicide attempts. We also used a restricted cubic splines model to investigate the pattern of the association.

Results: The participants had an average age of 12.59 (SD 0.64) years, with the majority being of Han ethnicity (743/782, 95.01%) and more than half being male (426/782, 54.48%). Most participants had no previous history of depression (541/782, 69.18%), anxiety (415/782, 53.07%), or stress (618/782, 79.03%). The rate of newly reported suicide attempts was 4.6% (36/782). A significant positive association was observed between internet addiction and suicide attempts (odds ratio 3.88, 95% CI 1.70-8.82), which remained significant after adjusting for age, sex, ethnicity, anxiety, depression, and stress (odds ratio 2.65, 95% CI 1.07-6.55). In addition, this association exhibited a linear pattern in the restricted cubic spline regression model.

Conclusions: This study suggested that internet addiction, rather than internet overuse, was associated with a higher likelihood of suicide attempts, which highlighted the importance of addressing internet addiction symptoms among Chinese adolescents for suicide prevention.

(J Med Internet Res 2025;27:e52083) doi:10.2196/52083

KEYWORDS

adolescents; pathological internet use; internet addiction; suicide attempts; risk factors; cohort study

XSL•FO
RenderX

Introduction

Suicide has emerged as a major public health concern globally recently, as it has brought significant challenges to human beings with severe consequences [1]. Over the past 5 decades, suicide remains the only contributor to rising mortality [2]. Globally, suicide ranked as the fourth leading cause of mortality among adolescents aged 15 to 19 years, as cited in a 2021 report by the World Health Organization (WHO) [3]. In the United States, it was reported that there were 2744 suicides among adolescents aged 10-19 years in 2019, positioning suicide as the second most common cause of death [3]. Therefore, reducing suicide‐related deaths has become a national priority in the United States, with more than US $22 million spent per year on suicide prevention. However, the suicide rate in the United States has not significantly declined [4]. China also faces the same challenge. Although the overall suicide rate in China has declined, the suicide rate among young people has remained high [5]. It is broadly recognized that suicide attempts may be the most significant predictor for completed suicide [6-9]. According to the data released by the National Trauma Data Bank, the mortality rate was 13% among adolescents aged younger than 20 years who attempted suicide [10]. A previous study indicated that teenagers who engaged in suicide attempts were 30 times more likely to die by suicide [11]. In addition, compared with those engaging in other forms of injury that lead to hospitalization, individuals with suicide attempts often have higher severities of injury, higher demands for health care services, and higher mortality [12]. Thus, examining the factors contributing to suicide attempts is crucial for the prevention of suicide among adolescents.

The internet has become increasingly accessible and necessary in the daily lives of adolescents. In China, internet access was available in 98.4% of elementary and middle schools in 2019 [13]. Meanwhile, in 2020, approximately 13.5% of all internet users were adolescents (aged 10-19 years) [14]. Proper internet use can help improve social skills, promote self-learning, and diversify recreation [15,16]. However, due to the highly time-consuming nature of the internet, an increasing number of adolescents have internet overuse, spending even more time using the internet than sleeping or studying [17]. In severe cases, excessive internet use may lead to clinically significant dysfunction and distress among adolescents [18,19]. This new form of behavioral addiction is listed as a tentative disorder named internet addiction. Internet addiction has also been increasing in Mainland China, with a reported prevalence of 10.4% to 26.5% among adolescents [20,21]. Earlier studies suggested a link between internet addiction and an increased incidence of suicide attempts [22,23]. A meta-analysis demonstrated that even after adjusting for demographic factors and depressive symptoms, those who experienced internet addiction continued to show a higher frequency of suicide attempts (pooled adjusted odds ratio [OR] 1.559) [23]. Some cross-sectional studies also indicated a link between internet addiction and suicidal behavior. For example, a study showed that adolescents with internet addiction had a greater likelihood of attempting suicide than those without this problem [24,25]. In addition, a 1-year follow-up study revealed that internet addiction was predictive of self-destructive behaviors among adolescents [26]. However, most previous studies are cross-sectional, which precluded the inference of causality. Furthermore, limited studies have investigated the relationship between the severity of pathological internet use and suicide attempts. Thus, in this prospective cohort study, we aimed to examine whether higher severity of pathological Internet use was associated with an increased risk of suicide attempts among Chinese adolescents [23].

Methods

Study Population

In this prospective cohort study, a 2-stage cluster sampling design was used to enhance the feasibility and representativeness. The sampling frame consisted of all public secondary schools across all districts in Changsha, as provided by the local education bureau. In the first stage, 1 public secondary school was randomly selected from the sampling frame. In the second stage, all students in grades 7 and 8 from the selected school were included in the study, resulting in a total sample of 1162 students (609 in grade 7 and 553 in grade 8). The selected school is comparable with most public secondary schools in the area in terms of size and demographic factors, such as gender ratio and socioeconomic background. This makes the school representative of the student population in Changsha and similar cities.

Recruitment occurred from November 2020 to December 2020, and follow-up data were collected from May 2021 to June 2021. Adolescents completed a standardized, self-administered questionnaire distributed through WeChat, one of the largest social media platforms in China. Before the commencement of the survey, we provided comprehensive training to schoolteachers on the procedures for distributing the questionnaire, offering guidance, and assisting students with difficulties. Data were collected from self-reported questionnaires administered through WeChat. All variables were collected through WeChat with the support of school personnel.

With the assistance of the school personnel, we obtained informed consent from both the adolescents and their parents. The participants accessed the survey using their parents' mobile phones and provided electronic informed consent on the first page of the survey. Upon completion, we thoroughly checked the data to ensure the data quality. Questionnaires were considered unqualified if the completion time was excessively short, if there were evident logical inconsistencies in responses, or if critical answers were missing.

Inclusion criteria of participants for the baseline and follow-up surveys were as follows: (1) middle school students aged 11-17 years, (2) with no severe medical conditions that prevented them from completing the survey, (3) capable of comprehending and finishing the questionnaires, and (4) provided a formal agreement to participate in the research. Participants who reported suicide attempts at baseline were excluded from the follow-up study. At baseline, 1162 grade 7-8 students were recruited, among whom 78 refused to participate. Therefore,

XSL•FO
RenderX

1084 students completed the survey at baseline. For the follow-up survey, 135 individuals were excluded due to suicide attempts at baseline, and 167 were excluded for refusal to participate; thus, 782 participants were included in the follow-up survey.

Assessment

Confounders

Sociodemographic Characteristics

Demographic variables, including sex, age, and ethnicity, were collected from the participants.

Depression, Anxiety, and Stress

The Depression, Anxiety, and Stress Scale–21 items (DASS-21) was used to assess participants' levels of depression, anxiety, and stress during the previous week [27]. The assessment includes 3 subscales: Depression, Anxiety, and Stress. Each of these subscales comprises 7 items that are evaluated using a 4-point Likert scale, which ranges from 0 (not at all) to 3 (most of the time) [28]. The thresholds were set at ≥13 for depression, ≥7 for anxiety, and ≥10 for stress [27]. The DASS-21 has shown high internal consistency (Cronbach α=0.92), indicating its good cross-cultural validity [29].

Exposures: Pathological Internet Use

Based on the conceptual framework, we used the Chen Internet Addiction Scale–Revised (CIAS-R) to assess the severity of pathological internet use [30]. The CIAS-R is a widely used instrument for evaluating pathological internet use, particularly in the Chinese context [31]. This self-reported scale includes 19 items, where each item is evaluated using a 4-point Likert scale that ranges from 1 (complete noncompliance) to 4 (complete conformity). The Chen Internet Addiction Scale (CIAS) total score varies between 0 and 76, with higher scores indicating more severe pathological internet use [32]. According to the total score, participants were further divided into 3 groups: the normal internet use group (<46), the internet overuse group (46-52), and the internet addiction group (≥53) [32]. The CIAS has proven to be a reliable and valid measure of pathological internet use among adolescents [30,33,34].

Outcomes: Suicide Attempts

Suicide attempts refer to the act of self-injuring that a person engages in with the intention of taking their own life. In this study, suicide attempts were defined as "someone trying to hurt themselves on purpose with the intention of ending their life, but it did not succeed" to tailor to junior high school students' comprehension levels. The question "Have you ever tried to commit suicide?" [22] was used at baseline to measure suicide attempts, and the question "Have you tried to commit suicide in the past 6 months?" was used to measure suicide attempts during the 6-month follow-up.

Statistical Analysis

In this study, the χ^2 test was used to compare categorical variables across groups. Continuous data are shown using mean and SDs, and the comparison of continuous variables between groups was conducted using ANOVA. The relationship between pathological internet use and suicide attempts was examined using logistic regression analysis using 3 models. Model 1 was used to examine the unadjusted OR; Model 2 was used to examine the adjusted OR after controlling for age, sex, and ethnicity; and Model 3 was used to examine the adjusted OR after adjusting for age, sex, ethnicity, anxiety, depression, and stress. The restricted cubic splines model was used to investigate the pattern of the association between the CIAS and suicide attempts after adjusting for age, sex, ethnicity, anxiety, depression, and stress. Finally, a post hoc power analysis was performed using G*Power 3.1 (Heinrich Heine University Düsseldorf) to evaluate the statistical power of the logistic regression analysis. The modeling of restricted cubic splines was conducted with R (version 4.2.3; R Foundation for Statistical Computing), while SPSS (version 26.0; IBM) was used for all additional statistical analyses. A 2-tailed P<.05 indicated statistical significance.

Ethical Considerations

This study protocol was examined and endorsed by the Ethics Committee of Central South University's Second Xiangya Hospital (K009). With the assistance of the school personnel, we obtained informed consent from both the adolescents and their parents on the first page of the questionnaire. The personal details of the participants were kept confidential, and a distinct study identification number was allocated to every participant for data entry, management, and analysis. The participants received free psychological assessments and medical help when needed as compensation for their participation.

Results

The analysis included 782 participants, whose average age was 12.59 (SD 0.64) years. Among these participants, more than half were male (426/782, 54.48%), and the majority were of Han ethnicity (743/782, 95.01%). Most of the participants reported no previous history of depression (541/782, 69.18%), anxiety (415/782, 53.07%), or stress (618/782, 79.03%). The rate of newly reported suicide attempts was 4.6% (36/782). Table 1 presents the comparison of demographic and clinical characteristics of participants among the normal internet use group, the internet overuse group, and the internet addiction group. The 3 groups showed significant differences in ethnicity (P=.03), depression (P<.001), anxiety (P<.001), stress (P<.001), and suicide attempts (P=.003). Compared with the normal internet use group, the internet overuse group and the internet addiction group were more likely to have stress, anxiety, depression, and suicide attempts after 6 months (P<.001 for stress, P<.001 for anxiety, P<.001 for depression, and P=.003 for suicide attempts).

XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of students by pathological internet use.

Variables	NIU[a] (n=609)	IO[b] (n=99)	IA[c] (n=74)	*F* test or chi-sqaure test (*df*)	*P* value
Age (years), mean (SD)	12.59 (0.62)	12.60 (0.62)	12.52 (0.78)	0.38 (2779)	.68
Sex, n (%)				0.18 (2)	.92
Male	330 (54.19)	54 (54.55)	42 (56.76)		
Female	279 (45.81)	45 (45.45)	32 (43.24)		
Ethnicity, n (%)				7.20 (2)	.03
Han	580 (95.24)	97 (97.98)	66 (89.19)		
others	29 (4.76)	2 (2.02)	8 (10.81)		
Depression, n (%)				52.14 (2)	<.001
No	460 (75.53)	47 (47.47)	34 (45.95)		
Yes	149 (24.47)	52 (52.53)	40 (54.05)		
Anxiety, n (%)				38.22 (2)	<.001
No	359 (58.95)	32 (32.32)	24 (32.43)		
Yes	250 (41.05)	67 (67.68)	50 (67.57)		
Stress, n (%)				54.97 (2)	<.001
No	516 (84.73)	61 (61.62)	41 (55.41)		
Yes	93 (15.27)	38 (38.38)	33 (44.59)		
SA[d] in 6 months, n (%)				11.96 (2)	.003
No	588 (96.55)	93 (93.94)	65 (87.84)		
Yes	21 (3.45)	6 (6.06)	9 (12.16)		

[a]NIU: normal internet use.

[b]IO: internet overuse.

[c]IA: internet addiction.

[d]SA: suicide attempts.

The associations between pathological internet use and suicide attempts across various models are presented in Table 2. Compared with the normal internet use group, the internet addiction group consistently showed a higher risk of suicidal attempts in Model 1 (crude OR 3.88 95% CI 1.70-8.82), Model 2 (adjusted OR 3.88, 95% CI 1.65-9.11), and Model 3 (adjusted OR 2.65, 95% CI 1.07-6.55). The restricted cubic spline regression model showed a linear association between CIAS and suicide attempts (Figure 1). Furthermore, internet overuse was also linked to an increased risk of suicidal attempts, but the association lacked statistical significance. The post hoc power analysis indicated that the statistical power of this study was 82.60% based on a sample size of 782 participants, an α level of .05, and the event probability under the null hypothesis of 3.45%.

XSL•FO
RenderX

Figure 1. Cubic model of the association between the CIAS score and the risk of suicide attempts after adjusting for age, sex, ethnicity, anxiety, depression, and stress. CIAS: Chen Internet Addiction Scale; OR: odds ratio.

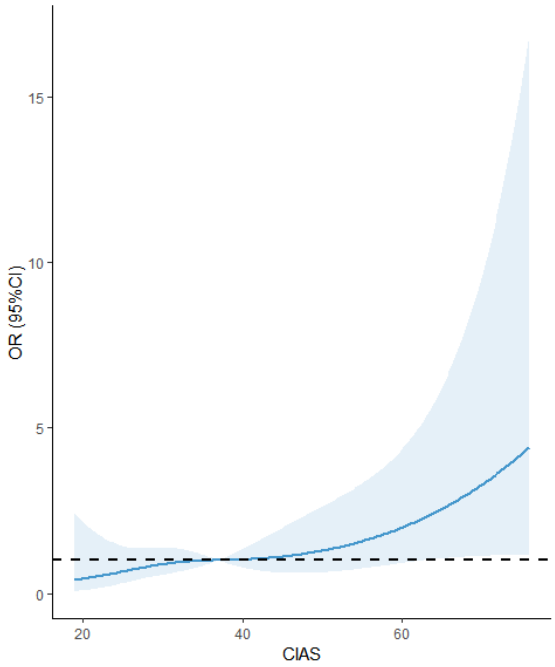

Table 2. The associations between pathological internet use and suicide attempts.

Group	Suicidal attempts, n/N (%)	Model 1[a]		Model 2[b]		Model 3[c]	
		OR[d] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
NIU[e] (ref[f])	21/609 (3.45)	1.00	—[g]	1.00	—	1.00	—
IO[h]	6/99 (6.06)	1.81 (0.71-4.59)	.21	1.90 (0.74-4.90)	.18	1.38 (0.52-3.68)	.52
IA[i]	9/74 (12.16)	3.88 (1.70-8.82)	.001	3.88 (1.65-9.11)	.002	2.65(1.07-6.55)	.04
P value for trend	—	—	.01	—	.01	—	.11

[a]Model 1: crude odds ratio.

[b]Model 2: odds ratio adjusted for age, sex, and ethnicity.

[c]Model 3: odds ratio adjusted for age, sex, ethnicity, anxiety, depression, and stress.

[d]OR: odds ratio.

[e]NIU: normal internet use.

[f]Ref: Reference.

[g]Not available.

[h]IO: internet overuse.

[i]IA: internet addiction.

Discussion

Principal Findings

To our knowledge, our research represents the initial investigation into the association between varying degrees of pathological internet use and suicide attempts among Chinese adolescents, using a prospective cohort design. Our findings suggested that the risk of suicide attempts was higher in individuals with internet addiction, even when controlling for variables such as age, sex, ethnicity, anxiety, depression, and stress. Furthermore, the restricted cubic splines model supported the linearity of the association between pathological internet use and the risk of suicide attempts.

This study found a notable association between internet addiction and suicide attempts through the 6 months of follow-up, which is in line with previous findings [35]. A study involving 8098 Chinese college students showed that those with internet addiction were at a higher risk for suicidal behaviors, with a prevalence of 21.4% [22]. Furthermore, a 1-year longitudinal study revealed that internet addiction was predictive of self-destructive behaviors, including self-harm and suicidal behaviors [26]. Furthermore, 2 meta-analyses also indicated that internet addiction was positively associated with suicidal

XSL•FO
RenderX

behaviors [23,36]. Some reasons might explain this association. First, individuals with internet addiction have a greater chance of exposure to online content involving suicidal thoughts and behaviors, which may result in imitation of suicidal behaviors [37], especially for individuals with mental health problems [38]. Second, risk factors for both suicidal attempts and internet addiction, such as depression, anxiety, and low socioeconomic status, seem to overlap [39,40]. Third, alternative explanations for the link between internet addiction and suicide attempts may involve brain and cognitive dysfunctions.

For instance, internet addiction is often comorbid with some mental health disorders involving impairment of impulse control, such as attention-deficit/hyperactivity disorder and substance use disorders, suggesting the impulsive nature associated with internet addiction [41,42]. Furthermore, a magnetic resonance imaging study showed that individuals with internet addiction had impaired prefrontal lobe function, which might lead to decreased ability of impulse control or even cognitive deficits [43]. Therefore, adolescents with internet addiction may exhibit a propensity for risky decision-making, such as suicide attempts [44].

After adjusting for age, sex, ethnicity, stress, anxiety, and depression, the OR of suicidal attempts decreased from 3.88 (95% CI 1.70-8.82) to 2.64 (95% CI 1.07-6.55). Still, the link between internet addiction and suicide attempts continued to be significant. A study conducted in Korea revealed a significant positive association between internet addiction, depression, and suicidal thoughts among adolescents [45]. A study involving Chinese middle school students also pointed out that internet addiction was linked to suicide attempts among adolescents, and the association might be mediated by depression and bullying behaviors [46]. This association may also be related to the proneness of comorbidity of internet addiction and other mental health problems such as depression, decreased self-esteem, lower tolerance to frustration, and emotional distress [47,48], which also increases the risk of suicidal attempts. Some studies suggested that the connection between internet addiction and suicide attempts might be mediated by these comorbidities, which, therefore, increased the risk of suicide attempts [49].

Interestingly, this study found that internet overuse did not have a significant association with suicide attempts, indicating that suicide attempts might only be related to internet addiction, which is more severe. This aligns with an earlier study on university students, which showed that most of the mental health problems in participants with mild internet addiction were comparable with or slightly above the average. However, the rates of mental health problems sharply increased in those with moderate to severe internet addiction [50]. Some other studies suggested that engaging with the internet at levels below addiction may increase the feeling of being accepted and alleviate feelings of loneliness and shame instead of leading to severe consequences [51].

Limitations

Some limitations need to be addressed. First, this study did not collect the specific dates of suicide attempts during the follow-up. This study only used a fixed 6-month follow-up period, which may introduce potential bias to our study results. Future studies should consider adding the specific dates of suicide attempts to get a more robust conclusion. Second, although we controlled for key risk factors such as anxiety, depression, and stress, which are known to influence suicidal behaviors, other important factors, such as academic burden and parent-child relationship, were not included in our analysis. Future research needs to include a broader range of risk factors and intermediate processes (eg, suicidal ideation) to get a more comprehensive understanding of the risk factors of suicide attempts. Third, the participants' internet addiction and clinical correlates were collected using self-report questionnaires and scales, which might have resulted in bias such as recall bias and social desirability bias. Future studies should consider using diagnostic interviews to get more accurate assessments. Fourth, the sample was recruited from 1 middle school in Hunan Province, China, and may not represent students from other schools in other areas. Future multicenter studies across various schools and regions are needed to get a more representative sample and improve the generalizability of our results. Finally, suicide attempts were assessed based on a single self-reported question instead of a standardized scale, hindering the cross-comparison with other studies using standard measurement tools [52]. For instance, Yang et al [53] developed a standardized and reliable assessment tool for suicide attempts, which may yield more accurate and consistent data across studies. Future studies should consider using more standard tools to capture the complexity of suicidal behaviors fully.

Conclusion

This study found that adolescents experiencing internet addiction and overuse were at a higher risk for stress, anxiety, and depression. Furthermore, internet addiction was associated with an increased risk of suicide attempts among adolescents in China. Given the considerable impact of suicide on society, the results of this study underscore the importance of addressing internet addiction among Chinese adolescents as part of suicide prevention and management efforts.

Acknowledgments

We would like to express our gratitude to all the participants in this project, as well as the school personnel, who have offered invaluable support.

This work was supported by the Scientific Research Launch Project for New Employees of the Second Xiangya Hospital of Central South University, National Natural Science Foundation of China (82201703), and Hunan Provincial Natural Science Foundation of China (2023JJ60085). The authors would also like to acnowledge the following funding sources: The STI2030-Major Projects-2021ZD0200700； The National Natural Science Foundation of China (82071543, 82471568); and The Key Research

XSL•FO
RenderX

and Development Program of Hunan Province (2023SK2028). No generative artificial intelligence was used in the manuscript writing.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JZ, YS, SL, XJ, and TF designed the study. SL, LS, JZ, and YS were responsible for the integrity of the data and the accuracy of data analysis. JZ, YS, and SL prepared the manuscript. SL, XJ, and TF edited and revised the manuscript. JZ, YS, and SL were responsible for the critical revision of the manuscript. All authors have contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Abio A, Owusu PN, Posti JP, Bärnighausen T, Shaikh MA, Shankar V, et al. Cross-national examination of adolescent suicidal behavior: a pooled and multi-level analysis of 193,484 students from 53 LMIC countries. Soc Psychiatry Psychiatr Epidemiol 2022;57(8):1603-1613 [FREE Full text] [doi: 10.1007/s00127-022-02287-x] [Medline: 35445842]
2. Clayton MG, Pollak OH, Owens SA, Miller AB, Prinstein MJ. Advances in research on adolescent suicide and a high priority agenda for future research. J Res Adolesc 2021;31(4):1068-1096. [doi: 10.1111/jora.12614] [Medline: 34820949]
3. World Health Organization. Suicide in the world: global health estimates. 2019. URL: https://iris.who.int/bitstream/handle/10665/326948/WHO-MSD-MER-19.3-eng.pdf?sequence=1&isAllowed=y [accessed 2024-12-05]
4. Office of the Surgeon General (US), National Action Alliance for Suicide Prevention (US). 2012 National Strategy for Suicide Prevention: Goals and Objectives for Action: A Report of the U.S. Surgeon General and of the National Action Alliance for Suicide Prevention. Washington (DC): US Department of Health & Human Services; 2012:184.
5. Wang CW, Chan CLW, Yip PSF. Suicide rates in China from 2002 to 2011: an update. Soc Psychiatry Psychiatr Epidemiol 2014;49(6):929-941. [doi: 10.1007/s00127-013-0789-5] [Medline: 24240568]
6. Zhong BL, Chan SSM, Liu TB, Chiu HFK. Nonfatal suicidal behaviors of Chinese rural-to-urban migrant workers: attitude toward suicide matters. Suicide Life Threat Behav 2019;49(5):1199-1208. [doi: 10.1111/sltb.12519] [Medline: 30315652]
7. Dai J, Zhong B, Xiang Y, Chiu HFK, Chan SSM, Yu X, et al. Internal migration, mental health, and suicidal behaviors in young rural Chinese. Soc Psychiatry Psychiatr Epidemiol 2015;50(4):621-631 [FREE Full text] [doi: 10.1007/s00127-014-0985-y] [Medline: 25403568]
8. Zhong BL, Xie WX, Zhu JH, Lu J, Chen H. Prevalence and correlates of suicide attempt among Chinese individuals receiving methadone maintenance treatment for heroin dependence. Sci Rep 2019;9(1):15859 [FREE Full text] [doi: 10.1038/s41598-019-52440-x] [Medline: 31666637]
9. Xu Y, Wang C, Shi M. Identifying Chinese adolescents with a high suicide attempt risk. Psychiatry Res 2018;269:474-480. [doi: 10.1016/j.psychres.2018.08.085] [Medline: 30195741]
10. Mathews EM, Woodward CJ, Musso MW, Jones GN. Suicide attempts presenting to trauma centers: trends across age groups using the National trauma data bank. Am J Emerg Med 2016;34(8):1620-1624. [doi: 10.1016/j.ajem.2016.06.014] [Medline: 27321943]
11. Shaffer D, Gould MS, Fisher P, Trautman P, Moreau D, Kleinman M, et al. Psychiatric diagnosis in child and adolescent suicide. Arch Gen Psychiatry 1996;53(4):339-348. [doi: 10.1001/archpsyc.1996.01830040075012] [Medline: 8634012]
12. King CA, Grupp-Phelan J, Brent D, Dean JM, Webb M, Bridge JA, et al. Predicting 3-month risk for adolescent suicide attempts among pediatric emergency department patients. J Child Psychol Psychiatry 2019;60(10):1055-1064 [FREE Full text] [doi: 10.1111/jcpp.13087] [Medline: 31328282]
13. The 46th report on Internet development in China. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202012/P020201201530023411644.pdf [accessed 2024-12-04]
14. CINI: 2019 national survey on internet use of minors. 2020. URL: https://www3.cnnic.cn/NMediaFile/old_attach/P020200513370410784435.pdf
15. Nguyen HV, Huang HC, Wong MK, Yang YH, Huang TL, Teng CI. Moderator roles of optimism and weight control on the impact of playing exergames on happiness: the perspective of social cognitive theory using a randomized controlled trial. Games Health J 2018;7(4):246-252. [doi: 10.1089/g4h.2017.0165] [Medline: 29708775]
16. Gao YX, Wang JY, Dong GH. The prevalence and possible risk factors of internet gaming disorder among adolescents and young adults: Systematic reviews and meta-analyses. J Psychiatr Res 2022;154:35-43. [doi: 10.1016/j.jpsychires.2022.06.049] [Medline: 35926424]
17. Marin MG, Nuñez X, de Almeida RMM. Internet addiction and attention in adolescents: a systematic review. Cyberpsychol Behav Soc Netw 2021;24(4):237-249. [doi: 10.1089/cyber.2019.0698] [Medline: 33121255]

XSL•FO
RenderX

18. Bickham DS. Current research and viewpoints on internet addiction in adolescents. Curr Pediatr Rep 2021;9(1):1-10 [FREE Full text] [doi: 10.1007/s40124-020-00236-3] [Medline: 33457108]
19. Shaw M, Black DW. Internet addiction: definition, assessment, epidemiology and clinical management. CNS Drugs 2008;22(5):353-365. [doi: 10.2165/00023210-200822050-00001] [Medline: 18399706]
20. Chi X, Hong X, Chen X. Profiles and sociodemographic correlates of internet addiction in early adolescents in southern China. Addict Behav 2020;106:106385. [doi: 10.1016/j.addbeh.2020.106385] [Medline: 32187574]
21. Xin M, Xing J, Pengfei W, Houru L, Mengcheng W, Hong Z. Online activities, prevalence of internet addiction and risk factors related to family and school among adolescents in China. Addict Behav Rep 2018;7:14-18 [FREE Full text] [doi: 10.1016/j.abrep.2017.10.003] [Medline: 29450251]
22. Shen Y, Meng F, Xu H, Li X, Zhang Y, Huang C, et al. Internet addiction among college students in a Chinese population: prevalence, correlates, and its relationship with suicide attempts. Depress Anxiety 2020;37(8):812-821. [doi: 10.1002/da.23036] [Medline: 32400075]
23. Cheng YS, Tseng PT, Lin PY, Chen TY, Stubbs B, Carvalho AF, et al. Internet addiction and its relationship with suicidal behaviors: a meta-analysis of multinational observational studies. J Clin Psychiatry 2018;79(4):17r11761. [doi: 10.4088/JCP.17r11761] [Medline: 29877640]
24. Lin I, Ko C, Chang Y, Liu T, Wang P, Lin H, et al. The association between suicidality and internet addiction and activities in Taiwanese adolescents. Compr Psychiatry 2014;55(3):504-510. [doi: 10.1016/j.comppsych.2013.11.012] [Medline: 24457034]
25. Kim K, Ryu E, Chon M, Yeun E, Choi S, Seo J, et al. Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: a questionnaire survey. Int J Nurs Stud 2006;43(2):185-192. [doi: 10.1016/j.ijnurstu.2005.02.005] [Medline: 16427966]
26. Pan PY, Yeh CB. Internet addiction among adolescents may predict self-harm/suicidal behavior: a prospective study. J Pediatr 2018;197:262-267. [doi: 10.1016/j.jpeds.2018.01.046] [Medline: 29550226]
27. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther 1995;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
28. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. Psychol. Assess 1998;10(2):176-181. [doi: 10.1037//1040-3590.10.2.176]
29. Wang K, Shi H, Geng F, Zou L, Tan S, Wang Y, et al. Cross-cultural validation of the depression anxiety stress scale-21 in China. Psychol Assess 2016;28(5):e88-e100. [doi: 10.1037/pas0000207] [Medline: 26619091]
30. Mak K, Lai C, Ko C, Chou C, Kim D, Watanabe H, et al. Psychometric properties of the Revised Chen Internet Addiction Scale (CIAS-R) in Chinese adolescents. J Abnorm Child Psychol 2014;42(7):1237-1245. [doi: 10.1007/s10802-014-9851-3] [Medline: 24585392]
31. Zhang WZ, Bing D. The application evaluation of the Revised Chen Internet Addiction Scale among urban middle school students. Chin. J. Sch. Health 2015;36(10):1489-1491 [FREE Full text]
32. Bai Y, Fan F. A study on the internet dependence of college students: the revising and applying of a measurement. Psychol. Dev. Edu 2005:99-104 [FREE Full text]
33. Ko C, Yen C, Yen C, Chen CC, Yen CN, Chen S. Screening for internet addiction: an empirical study on cut-off points for the chen Internet addiction scale. Kaohsiung J Med Sci 2005;21(12):545-551 [FREE Full text] [doi: 10.1016/S1607-551X(09)70206-2] [Medline: 16670046]
34. Lai CM, Mak KK, Watanabe H, Ang RP, Pang JS, Ho RCM. Psychometric properties of the Internet Addiction Test in Chinese adolescents. J Pediatr Psychol 2013;38(7):794-807. [doi: 10.1093/jpepsy/jst022] [Medline: 23671059]
35. Tang J, Ma Y, Lewis SP, Chen R, Clifford A, Ammerman BA, et al. Association of internet addiction with nonsuicidal self-injury among adolescents in China. JAMA Netw Open 2020;3(6):e206863 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6863] [Medline: 32496567]
36. Wang J, Hao Q, Tu Y, Peng W, Wang Y, Li H, et al. Assessing the association between internet addiction disorder and health risk behaviors among adolescents and young adults: a systematic review and meta-analysis. Front Public Health 2022;10:809232 [FREE Full text] [doi: 10.3389/fpubh.2022.809232] [Medline: 35433568]
37. Baker D, Fortune S. Understanding self-harm and suicide websites: a qualitative interview study of young adult website users. Crisis 2008;29(3):118-122. [doi: 10.1027/0227-5910.29.3.118] [Medline: 18714907]
38. Ko CH, Yen JY, Yen CF, Chen CS, Chen CC. The association between internet addiction and psychiatric disorder: a review of the literature. Eur Psychiatry 2012;27(1):1-8. [doi: 10.1016/j.eurpsy.2010.04.011] [Medline: 22153731]
39. Sueki H, Yonemoto N, Takeshima T, Inagaki M. The impact of suicidality-related internet use: a prospective large cohort study with young and middle-aged internet users. PLoS ONE 2014;9(4):e94841. [doi: 10.1371/journal.pone.0094841] [Medline: 24740115]
40. Tang CSK, Koh YYW. Online social networking addiction among college students in Singapore: comorbidity with behavioral addiction and affective disorder. Asian J Psychiatr 2017;25:175-178. [doi: 10.1016/j.ajp.2016.10.027] [Medline: 28262144]

XSL•FO
RenderX

41. Zhang Y, Mei S, Li L, Chai J, Li J, Du H. The relationship between impulsivity and internet addiction in Chinese college students: a moderated mediation analysis of meaning in life and self-esteem. PLoS One 2015;10(7):e0131597 [FREE Full text] [doi: 10.1371/journal.pone.0131597] [Medline: 26171860]
42. Weinstein A, Yaacov Y, Manning M, Danon P, Weizman A. Internet addiction and attention deficit hyperactivity disorder among schoolchildren. Isr Med Assoc J 2015;17(12):731-734 [FREE Full text] [Medline: 26897972]
43. Huang Y, Xu L, Kuang L, Wang W, Cao J, Xiao M. Abnormal brain activity in adolescents with internet addiction who attempt suicide: an assessment using functional magnetic resonance imaging. Neural Regen Res 2020;15(8):1554-1559 [FREE Full text] [doi: 10.4103/1673-5374.274346] [Medline: 31997822]
44. Cheng C, Sun P, Mak KK. Internet addiction and psychosocial maladjustment: avoidant coping and coping inflexibility as psychological mechanisms. Cyberpsychol Behav Soc Netw 2015;18(9):539-546. [doi: 10.1089/cyber.2015.0121] [Medline: 26348815]
45. Ryu EJ, Choi KS, Seo JS, Nam BW. The relationships of internet addiction, depression, and suicidal ideation in adolescents. Taehan Kanho Hakhoe Chi 2004;34(1):102-110. [doi: 10.4040/jkan.2004.34.1.102] [Medline: 15314344]
46. Pan S, Wuquan L, Ming L, Guo L, Deng X. Relationship between internet addiction and suicidal behavior among middle school students in Guangzhou. Chinese Journal of School Health. Chinese Journal of School Health 2018 Feb 25:229-231 [FREE Full text] [doi: 10.16835/j.cnki.1000-9817.2018.02.021]
47. Ko C, Yen J, Yen C, Chen C, Wang S. The association between internet addiction and belief of frustration intolerance: the gender difference. Cyberpsychol Behav 2008;11(3):273-278. [doi: 10.1089/cpb.2007.0095] [Medline: 18537496]
48. Wang L, Luo J, Bai Y, Kong J, Luo J, Gao W, et al. Internet addiction of adolescents in China: prevalence, predictors, and association with well-being. Addict. Res. Theory 2012;21(1):62-69. [doi: 10.3109/16066359.2012.690053]
49. Liu HC, Liu SI, Tjung JJ, Sun FJ, Huang HC, Fang CK. Self-harm and its association with internet addiction and internet exposure to suicidal thought in adolescents. J Formos Med Assoc 2017;116(3):153-160 [FREE Full text] [doi: 10.1016/j.jfma.2016.03.010] [Medline: 27146126]
50. Guo W, Tao Y, Li X, Lin X, Meng Y, Yang X, et al. Associations of internet addiction severity with psychopathology, serious mental illness, and suicidality: large-sample cross-sectional study. J Med Internet Res 2020;22(8):e17560 [FREE Full text] [doi: 10.2196/17560] [Medline: 32780029]
51. Duggan JM, Heath NL, Lewis SP, Baxter AL. An examination of the scope and nature of non-suicidal self-injury online activities: implications for school mental health professionals. School Mental Health 2011;4(1):56-67. [doi: 10.1007/s12310-011-9065-6]
52. Cao XL, Zhong BL, Xiang YT, Ungvari GS, Lai KY, Chiu HF, et al. Prevalence of suicidal ideation and suicide attempts in the general population of China: a meta-analysis. Int J Psychiatry Med 2015;49(4):296-308 [FREE Full text] [doi: 10.1177/0091217415589306] [Medline: 26060259]
53. Yang F, Tong J, Zhang S, Zhang J, Zhong B. Prevalence and correlates of suicide attempts in Chinese individuals with borderline personality disorder. Front Psychiatry 2022;13:942782 [FREE Full text] [doi: 10.3389/fpsyt.2022.942782] [Medline: 36104990]

Abbreviations

CIAS: Chen Internet Addiction Scale
CIAS-R: Chen Internet Addiction Scale-Revised
DASS-21: Depression, Anxiety, and Stress Scale–21 items
OR: odds ratio
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 22.08.23; peer-reviewed by M Wu, Y Latifeh, BL Zhong; comments to author 22.07.24; revised version received 21.08.24; accepted 18.11.24; published 03.01.25.

Please cite as:
Li S, Jin X, Song L, Fan T, Shen Y, Zhou J
The Association Between Internet Addiction and the Risk of Suicide Attempts in Chinese Adolescents Aged 11-17 Years: Prospective Cohort Study
J Med Internet Res 2025;27:e52083
URL: https://www.jmir.org/2025/1/e52083
doi:10.2196/52083
PMID:

XSL•FO
RenderX

©Sihong Li, Xingyue Jin, Lintong Song, Tianqing Fan, Yanmei Shen, Jiansong Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Trends in Mental Health Outcomes of College Students Amid the Pandemic (Roadmap mHealth App): Longitudinal Observational Study

Gautham Jayaraj[1*], BS; Xiao Cao[1*], MS; Adam Horwitz[2], PhD; Michelle Rozwadowski[1], BS; Skyla Shea[1]; Shira N Hanauer[1]; David A Hanauer[3], MS, MD; Muneesh Tewari[4,5,6,7,8], MD, PhD; Kerby Shedden[9], PhD; Sung Won Choi[1,5], MS, MD

[1]Department of Pediatrics, Medical School, University of Michigan, Ann Arbor, MI, United States
[2]Department of Psychiatry, Medical School, University of Michigan, Ann Arbor, MI, United States
[3]Department of Learning Health Sciences, Medical School, University of Michigan, Ann Arbor, MI, United States
[4]Department of Internal Medicine, Medical School, University of Michigan, Ann Arbor, MI, United States
[5]Rogel Comprehensive Cancer Center, University of Michigan, Ann Arbor, MI, United States
[6]VA Ann Arbor Healthcare System, Ann Arbor, MI, United States
[7]Department of Biomedical Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States
[8]Center for Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI, United States
[9]Department of Statistics, College of Literature, Sciences, and the Arts, University of Michigan, Ann Arbor, MI, United States
[*]these authors contributed equally

Corresponding Author:
Sung Won Choi, MS, MD
Department of Pediatrics
Medical School
University of Michigan
1200 E Hospital Dr
Medical Professional Building D4115
Ann Arbor, MI, 48109
United States
Phone: 1 734 615 5707
Email: sungchoi@med.umich.edu

Abstract

Background: The mental health crisis among college students intensified amid the COVID-19 pandemic, suggesting an urgent need for innovative solutions to support them. Previous efforts to address mental health concerns have been constrained, often due to the underuse or shortage of services. Mobile health (mHealth) technology holds significant potential for providing resilience-building support and enhancing access to mental health care.

Objective: This study aimed to examine the trends in mental health and well-being outcomes over 3 years among college students, with an exploratory aim to assess the potential impact of the Roadmap mHealth app on these outcomes.

Methods: A fully automated longitudinal observational study was conducted remotely from a large public academic institution in the Midwestern United States, evaluating mental health and well-being outcomes among college students using the Roadmap mHealth app over 3 fall semesters from 2020 to 2022. The study enrolled 2164 college students in Year I, with 1128 and 1033 students returning in Years II and III, respectively. Participants completed various self-reported measures, including the Patient Health Questionnaire-9 for depression, Generalized Anxiety Disorder-7 for anxiety, and additional metrics for coping, flourishing, and loneliness.

Results: The findings indicated an evolving trajectory in students' mental health. In Year I, depression and anxiety levels were higher compared with levels reported between 2014 and 2019, remaining stable into Year II. However, significant decreases were noted by Year III for both depression (Year I mean 7.78, SD 5.65 vs Year III mean 6.21, SD 4.68; t_{108}=–2.90; P=.01) and anxiety (Year I mean 6.61, SD 4.91 vs Year III mean 5.62, SD 4.58; t_{116}=–2.02; P=.046). Problem-focused coping decreased initially from Year I (mean 2.46, SD 0.58) to Year II (mean 2.36, SD 0.60; t_{1073}=–5.87; P<.001), then increased by Year III (mean 2.40,

XSL•FO
RenderX

SD 0.63; t_{706}=2.26; *P*=.02). Emotion-focused (Year I mean 2.33, SD 0.41 vs Year III mean 2.22, SD 0.47; t_{994}=–7.47; *P*<.001) and avoidant coping (Year I mean 1.76, SD 0.37 vs Year III mean 1.65, SD 0.38; t_{997}=–8.53; *P*=.02) consistently decreased. Loneliness significantly decreased from Year I (mean 5.79, SD 1.74) to Year III (mean 5.17, SD 1.78; t_{1013}=–10.74; *P*<.001), accompanied by an increase in flourishing from Year I (mean 63.78, SD 14.76) to Year III (mean 66.98, SD 15.06; t_{994}=7.22; *P*<.001). Analysis of app usage indicated that the positive piggy bank and gratitude journal were the favored activities. Greater engagement with the app was positively correlated with enhanced flourishing, even after adjusting for demographic and sociobehavioral factors (β=.04, SE .016; t_{3974}=2.17; *P*=.03).

Conclusions: In this study, students' mental health and well-being improved, with notable reductions in depression, anxiety, and loneliness, associated with an increase in flourishing. The app did not appear to worsen students' mental health. Based on the usage pattern, it is possible the app enhanced positive psychology-based practices. Future research should explore the efficacy of mHealth interventions through randomized controlled trials to further understand their impact on college students' mental health outcomes.

Trial Registration: ClinicalTrials.gov NCT04766788; https://clinicaltrials.gov/ct2/show/NCT04766788

International Registered Report Identifier (IRRID): RR2-10.2196/29561

(J Med Internet Res 2025;27:e67627) doi:10.2196/67627

KEYWORDS

mHealth; college; student; mental health; positive psychology; flourishing; COVID-19; wellbeing; mobile phone; SARS-CoV-2; coronavirus; pandemic; COVID; app; digital health; smartphone; eHealth; telehealth; telemedicine; longitudinal; higher education; depression; anxiety; loneliness

Introduction

Nearly 20 million students are enrolled in colleges across the United States [1]. During this time, students undergo crucial psychological and social development [2]. Mental health issues are prevalent in college student populations, with the peak onset of mental illness occurring before age 25 [3]. However, early detection of emerging mental illness before significant symptoms develop remains limited [4]. Without adequate attention, at-risk young adults with mental health issues are more likely to struggle academically (ie, receive lower grand point averages), drop out of college, or be unemployed compared with their peers [5-7]. These outcomes may consequently correlate with increased substance abuse, self-injury, or suicidality [8-11].

The long-simmering mental health crisis in college students was further exacerbated by the COVID-19 pandemic, which saw higher levels of depression and anxiety than ever previously reported [12,13]. Even before the pandemic, the demand for mental health services was rising [14,15]. Thus, while the current mental health crisis among college students remains a major public health problem, it also presents a unique opportunity to develop and test novel interventions for providing psychosocial support to this population. Psychosocial support refers to interventions aimed at addressing both psychological and social aspects of mental health, helping individuals cope with stress, improve their emotional well-being, and develop supportive social connections [16].

The increase in psychopathology across college campuses has led to expanded counseling and mental health services [17,18]. However, students who stopped attending college due to mental health concerns reported not seeking help or experiencing long wait times [19]. Alarmingly, 80% of students who die by suicide never contact their campus mental health services [20]. Both the underuse and shortage of mental health services are significant issues [21,22]. Mobile health (mHealth) technology offers a novel approach to overcoming barriers associated with mental health care delivery [23]. Evidence suggests that mHealth interventions can support various mental health problems, such as mood disorders, stress, and substance use [24,25]. Focusing on resilience-building aligns with recent recommendations that mental health-related mHealth apps should support well-being, enhance mood, and foster self-care skills rather than address specific mental health problems [26]. Mobile health technology has been shown to be convenient, anonymous, and easy to use [27].

Our interdisciplinary team developed a positive psychology-based mHealth app, Roadmap, initially for use in patients with cancer and their family caregivers. It was initially designed to provide information, education, and skills-based training. Over time, it was iteratively enhanced to provide psychosocial support, that is, the physical, mental, and social health-related quality of life (HRQOL), for any user and to aggregate their step counts and sleep, collected through wearable sensors (Fitbits) [28]. The app incorporated principles based on positive psychology, which emphasizes enhancing individuals' strengths, cultivating gratitude, fostering resilience, and promoting overall well-being through evidence-based interventions [29,30]. Based on initial observations from our earlier deployment of the Roadmap app, we observed specific patterns in user engagement and feedback regarding the types and frequencies of self-reported measures that were most useful. This informed our strategic selection of validated measures (eg, Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7 (GAD-7), coping, flourishing, loneliness), as well as the intervals at which these were administered across the 4-month study design. This approach aimed to optimize the app's utility in monitoring and assessing their health outcomes with app use without excessive burden.

XSL•FO
RenderX

Positive psychology emphasizes enhancing individuals' strengths, cultivating gratitude, fostering resilience, and promoting overall well-being. Such a framework is particularly suitable for college students, as it aligns with crucial developmental goals during this life stage, such as building resilience, self-efficacy, and effective coping strategies. College students are often navigating significant life changes and stressors, making them an ideal population to benefit from interventions that focus on positive psychology principles. Interventions based on these principles have been developed to improve health (eg, physical, mental, social) and well-being through well-studied constructions (eg, positive daily reflection [29,30], gratitude [31-34], and savoring [35]). Simple strategies aimed at enhancing positive thoughts, emotions, and behaviors have been shown to be effective and highly scalable [36-39]. These strategies have been used in students to support engagement, learning, and well-being. In fact, an economic evaluation of positive psychology strategies that sought to foster positive emotions, stimulate positive functioning, and reduce depressive symptoms found that cost-effectiveness was increased in individuals randomized to the positive psychology intervention arm compared with the waitlisted usual care group [40]. Thus, the strong evidence base for a positive psychology framework to enhance resilience or positive adaptations despite high-risk or chronic stress provides support for their use in college students.

In the fall semester of 2020, we leveraged the Roadmap app to explore HRQOL outcomes among college students [41]. Approximately one-third of students had a mental health disorder and—given the strict lockdown restrictions in the fall semester of 2020—students recorded low physical activity levels (through Fitbits). In addition, we found significant associations of COVID-19 positivity with the use of marijuana and alcohol and with lower belief in public health measures [42]. With consent, the same students were invited to participate in the Roadmap study and report their HRQOL in the following fall semesters of 2021 and 2022. Based on our initial observations [41] and the emerging literature on college student mental health during the pandemic [43], we posited that our participants would report changes in their HRQOL as the pandemic progressed, specifically across the 3-year study period. We also explored whether the Roadmap app, a positive psychology-based mHealth app, might be associated with changes in these HRQOL outcomes. By examining both the trajectory of HRQOL and the role of the Roadmap app, our work contributes to understanding the broader effects of the pandemic on student mental health and highlights the potential of a mHealth app in supporting psychological well-being. The primary aim of this study was to examine the trends in mental health and well-being outcomes over 3 years among college students, with an exploratory aim to assess the potential impact of the Roadmap mHealth app on these outcomes.

Methods

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional review board (IRB) of the University of Michigan Medical School and was registered on ClinicalTrials.gov (NCT04766788). Study participants completed an IRB-approved informed consent through the SignNow platform [44]. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. There were no deviations from the registered study on ClinialTrials.gov NCT04766788. In addition, there were no deviations from the published protocol manuscript [41]. Of note, the study was originally intended to last 1 year (Year I); however, it was expanded to follow the study participants over several years, continuing the following 2 years, designated as Year II and Year III. This was not originally described in the published protocol manuscript [41]. Please refer to the study schema in Figure 1.

XSL•FO
RenderX

Figure 1. Study schema across all 3 years of the study.

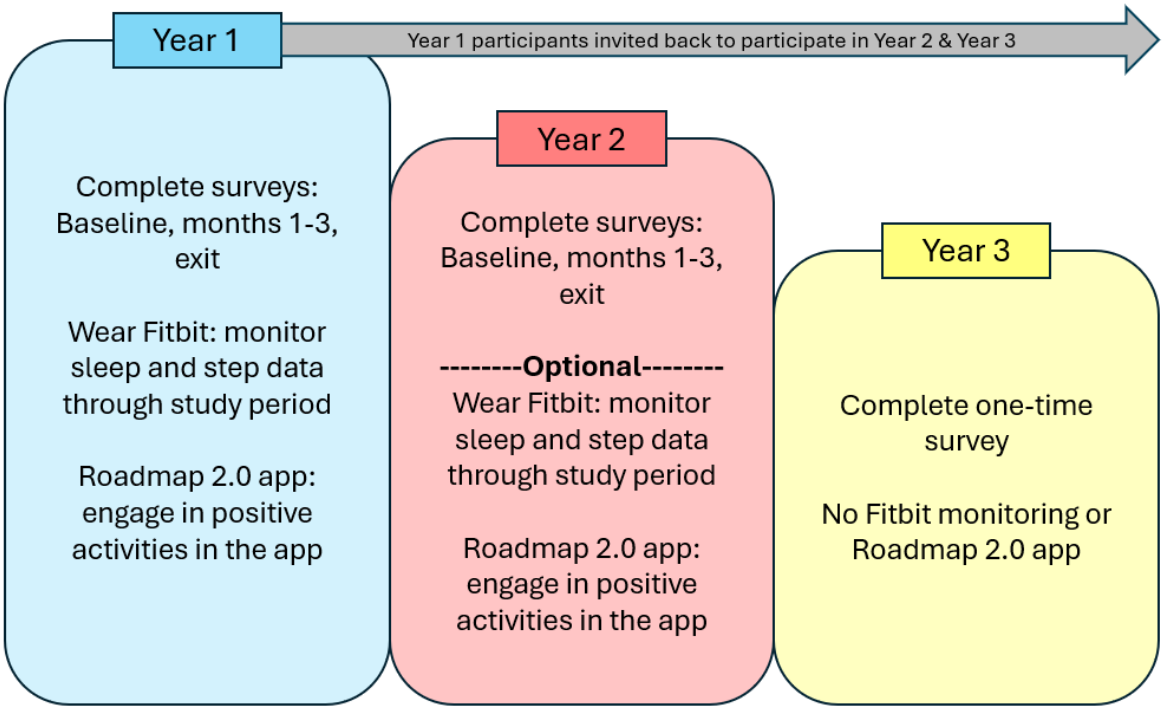

Privacy, Confidentiality, and Compensation

All participant data were anonymized or deidentified to safeguard participant information. Study participants were compensated for their participation. In Year I, participants who completed the baseline and monthly surveys received a US $10 gift card for each survey completed. Each study participant who enrolled received a Fitbit device (valued at US $150) to use during the study and to keep afterward. Year II participants were compensated based on the surveys they completed, which varied in length: US $35 for the baseline survey, US $10 for each monthly survey (up to three surveys), and US $25 for the exit survey. Participants could earn an additional US $10 if they completed all Year II surveys, potentially receiving up to US $100. In addition, those who chose to share their Fitbit data were compensated up to US $5 per week for a maximum of 17 weeks (approximately 4 months), totaling up to US $85. Year III participants received US $25 for completing a one-time survey.

Study Site

The study took place at a large public academic institution in the Midwestern United States. For the safety of participants and staff during the pandemic, study activities were conducted entirely remotely, with study materials shipped to participants' residences. This was a 3-phase longitudinal and observational study conducted in the fall semesters of 2020, 2021, and 2022, referred to as Year I, Year II, and Year III, respectively. Notably, the same cohort of students was invited to participate each year, beginning from Year I, aside from any participants who had formally dropped out.

Recruitment and Enrollment

The research protocol was previously published [41], and additional information regarding the study design for Year I can be found in our previous manuscript [42]. Briefly, eligible participants were required to be college students of at least 18 years of age. Additional eligibility requirements included having access to the necessary resources for participating in a mHealth technology-based intervention (ie, smartphone or tablet and internet access) and being willing to use their personal equipment and internet for the study.

We recruited the Year I cohort of students in the fall semester of 2020 through an email distribution channel managed by the institution and approved by the IRB; this email list initially reached all currently registered and enrolled college students in 2020. In addition, we distributed flyers throughout campus buildings and buses. Interested participants were provided with additional study information, including an overview of study procedures, risks, and benefits. If they met the eligibility criteria, participants were given a digital IRB-approved informed consent form. In Year I, there were 2164 participants enrolled in the study. For Year II and Year III, these students were recontacted in the respective fall semester to complete HRQOL surveys.

Study Procedures

Roadmap and Fitbit apps: In Year I, participants were instructed to download the Roadmap and Fitbit apps to their smartphones (both free of charge and available on Apple and Google app stores). Each individual participant was provided with a unique Roadmap ID in order to access the app. The Roadmap app offered 8 resilience-based positive activities, previously described in Rozwadowski et al [28] and provided in Multimedia Appendix 1. Participants could also engage with each other anonymously through a Chat Forum (ie, peer-to-peer support)

XSL•FO
RenderX

on each of these topics. Participants were instructed to use these activities as much or as little as they found helpful. In addition, participants were asked to record their daily mood on a scale of 1 to 10 (1 being the worst and 10 being the best), with a reminder sent through push notification to their smartphones at 8:00 PM. In Year II, participants had the option to continue using the Roadmap activities to monitor their physical activity and sleep with Fitbits. In Year III, participants did not have access to the Roadmap activities.

Self-Reported Assessments

Participants self-reported HRQOL assessments through the Roadmap app. The Roadmap app provides a link to Qualtrics (Qualtrics), which hosts study-specific surveys and the secure storage of participant data. We connected participant surveys together using their unique ID associated with the Roadmap app rather than any personally identifiable information. The survey data collected during Year I, Year II, and Year III were used in this study's analysis. In Year I and Year II, the survey intervals included baseline, Month 1, Month 2, and Month 3; in Year III, the survey was collected only at baseline. At the appropriate intervals, leveraging both Roadmap app reminders and study team emails, we prompted participants to complete each survey and provided compensation for each completed survey. Surveys were scored in accordance with their standardized scoring and as previously published [42] (Textbox 1).

Textbox 1. List of validated self-reported health-related quality of life assessments used in the surveys.

- Coping: We used the 28-item Brief COPE [45], which assesses the self-reported frequency of use of 14 different coping strategies to deal with a particular situationally specific life event, including self-distraction, active coping, denial, alcohol and drug use, use of emotional support, use of instrumental support, behavioral disengagement, venting, positive framing, planning the use of humor, acceptance, and religion. This study's analysis followed the subscale analyses provided by Poulus et al 2020 [46].
- Flourishing: A 10-item measure with 2-items each in the following categories, that are (1) happiness and life satisfaction; (2) mental and physical health; (3) meaning and purpose; (4) character and virtue; and (5) close social relationships. A higher score indicates greater flourishing [47].
- Loneliness: A 3-item measure assessing isolation, feeling left out, or lacking companionship. A higher score is associated with increased loneliness [48,49].
- Anxiety: We used the Generalized Anxiety Disorder-7 scale, a 7-item anxiety assessment commonly used in primary care. Scale scores are categorized into minimal, mild, moderate, and severe anxiety. An increased score is associated with increased anxiety. Previous studies found an internal consistency of 0.89 [50].
- Depression: We used the Patient Health Questionnaire-9, a 9-item depression assessment commonly used in primary care. Scale scores are categorized into minimal, mild, moderate, moderately severe, and severe depression. An increased score corresponds to increased depression. Previous studies have reported an internal consistency of 0.83-0.92 [51].

Statistical Analysis

We analyzed data from the 3 phases, encompassing 2164 students in Year I, with 1128 students returning in Year II and 1033 students returning in Year III. All analyses were performed using the statistical software R (version 4.2.3; R Core Team) and organized in the following manner (Textbox 2).

XSL•FO
RenderX

Textbox 2. List of statistical analyses performed.

- Descriptive statistics: Descriptive statistics were compiled for categorical variables in terms of frequency and percentage, while continuous variables were summarized using the mean, SD, median, and IQR. Boxplots were used to depict the median and overall distribution of key measures. Univariate associations between continuous and demographic variables were evaluated. All tests were at a significance level of .05.
- Normality assessment and univariate analyses: Depending on data normality (assessed with the Shapiro-Wilk test [52]), continuous variables were compared across 2 categorical variable levels using either 2-sided *t* tests or Wilcoxon rank-sum tests. For comparisons across 3 or more levels of categorical variables, ANOVA or Kruskal-Wallis tests [53] were used. Associations between 2 categorical variables were examined using the chi-square test [54], leveraging the relatively large sample size.
- Correlational analyses: To assess associations between continuous variables, 2-sided Pearson or Spearman correlations [55] were calculated based on data normality.
- Longitudinal analyses: Paired *t* tests were used to evaluate interyear differences in outcome means. Correlation coefficient matrices were constructed to illustrate the relationships among continuous outcomes. We separately considered the monthly responses for flourishing in Year I and Year II. These analyses used a total of 6 assessments, given in consecutive months, modeling the current month's flourishing for months 2, 3, and 4 on the previous month's explanatory variables, including app usage. All models included time-invariant fixed effects for time since the pandemic peak (year), sex, undergraduate or graduate status, coping scores, and drug use. To account for within-subject dependence, random intercepts and slopes (year) at the participant level were included.
- Mediation analysis: A directed acyclic graph (DAG) was postulated to relate the key variables of flourishing, depression Patient Health Questionnaire-9 score), anxiety (Generalized Anxiety Disorder-7 score), and loneliness in the context of pandemic time. Based on this DAG, separate regression models were specified for flourishing, anxiety, and depression as outcomes. These models considered only baseline data, modeling the current year's baseline response in Year II and Year III on the previous year's baseline explanatory variables. We also conducted a formal mediation analysis based on this DAG.
- Handling missing data: Missing values for continuous variables, including Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 scores, were imputed using Multiple Imputation by Chained Equation [56] with predictive mean matching.
- Exploratory app use analysis: The app use variable was computed as the sum of the frequencies of positive activities, chat forum interactions, and mood entries.

Results

Student Demographics Based on Study Site, Recruitment, and Enrollment

Of the 2164 students who participated in Year I, 1128 returned in Year II, and 1033 returned in Year III. Notably, 755 students participated in all 3 years, as shown in the CONSORT (Consolidated Standards of Reporting Trials) flow diagram [57] (Figure 2). The majority of participants in Year I were White (1243/2164, 57%), female (1467/2164, 68%), and undergraduates (1429/2164, 66%). The demographic characteristics remained largely consistent across time, with the exception of an increase in those who reported no longer being students and a higher proportion of females (Table 1).

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram for participant recruitment and enrollment, accompanied by a Venn diagram illustrating student participation across the 3-year study period. Year I: n=2164; Year II: n=1128; and Year III: n=1033. N=755 students participated in all 3 phases; n=373 were unique to Year I and Year II; and n=278 participants were unique to Year I and Year III.

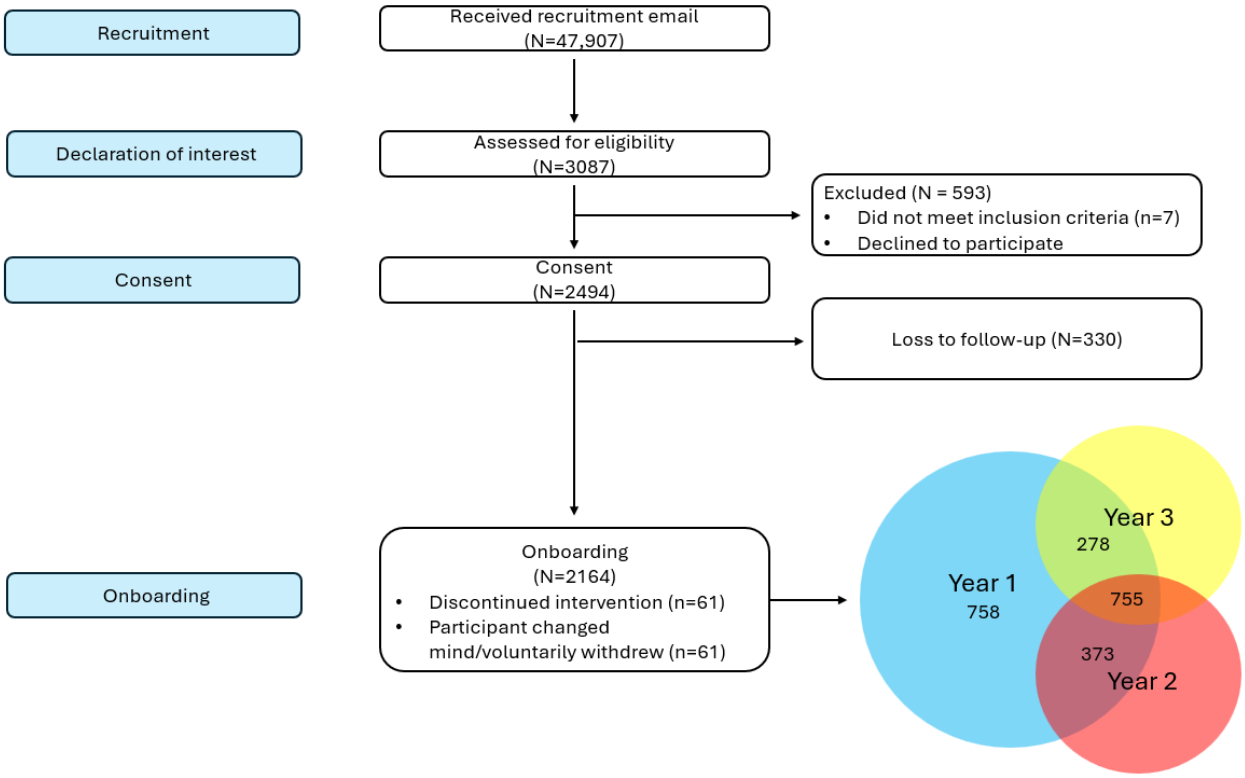

XSL•FO
RenderX

Table 1. Demographic breakdown of study participants over 3 years.

Demographics	Participant demographics (population, n %)			*P* value[a]
	Year I	Year II	Year III	
School year				<.001
Undergraduate	1429 (66.04)	544 (48.23)	338 (32.72)	
Graduate	719 (33.22)	363 (32.19)	269 (26.04)	
No longer a student	15 (0.69)	218 (19.32)	426 (41.24)	
Not available	1 (0.05)	3 (0.27)	0 (0)	
Sex				.004
Female	1467 (67.8)	833 (73.87)	740 (71.64)	
Male	677 (31.28)	283 (25.09)	283 (27.4)	
Other	11 (0.51)	11 (0.98)	9 (0.87)	
Not available	9 (0.42)	1 (0.09)	1 (0.1)	
Race				.92
White	1243 (57.44)	643 (57.03)	604 (58.48)	
Black or African American	94 (4.34)	40 (3.55)	40 (3.87)	
Asian	652 (30.13)	356 (31.56)	309 (29.91)	
Multiracial	113 (5.22)	62 (5.5)	55 (5.32)	
Other	39 (1.8)	18 (1.6)	18 (1.74)	
Not available	23 (1.06)	9 (0.8)	7 (0.68)	
Ethnicity				.17
Hispanic or Latino	210 (9.70)	94 (8.33)	81 (7.9)	
Non-Hispanic or Latino	1949 (90.04)	1029 (91.60)	947 (92.1)	
Not available	5 (0.23)	5 (0.44)	5 (0.48)	
Domestic or international				.36
Domestic	1996 (92.24)	1027 (91.05)	957 (92.63)	
International	163 (7.53)	98 (8.7)	74 (7.16)	
Not available	5 (0.23)	3 (0.27)	2 (0.19)	
First or continuing generation				.83
First generation	541 (25)	290 (25.71)	254 (24.59)	
Continuing generation	1617 (74.74)	834 (73.94)	775 (75.01)	
Not available	6 (0.28)	4 (0.35)	4 (0.39)	

[a]*P* values are representative of a chi-square test performed for the entire study population.

Student Coping Skills Using Descriptive Statistics, Normality Assessment, Univariate Analyses, and Longitudinal Analyses

We examined student coping skills over the 3-year study period using pairwise comparisons between each pair of years, restricted to students who responded in both years. While problem-focused coping decreased from Year I to Year II (*P*<.001), it remained stable between Year I and Year III (*P*=.23) but increased from Year II to Year III (*P*=.02; Figure 3A). Emotion-focused coping decreased from Year I to Year II (*P*<.001) and from Year I to Year III (*P*<.001) but remained stable between Year II and Year III (*P*=.65; Figure 3B). Avoidant coping decreased across all time points, including from Year I to Year II (*P*<.001), Year I to Year III (*P*<.001), and Year II to Year III (*P*=.02; Figure 3C).

XSL•FO
RenderX

Figure 3. Coping skills across the 3-year study period. The boxes in each plot depict differences between pairs of phases. Panels (A), (B), and (C) illustrate problem-focused, emotion-focused, and avoidant coping, respectively.

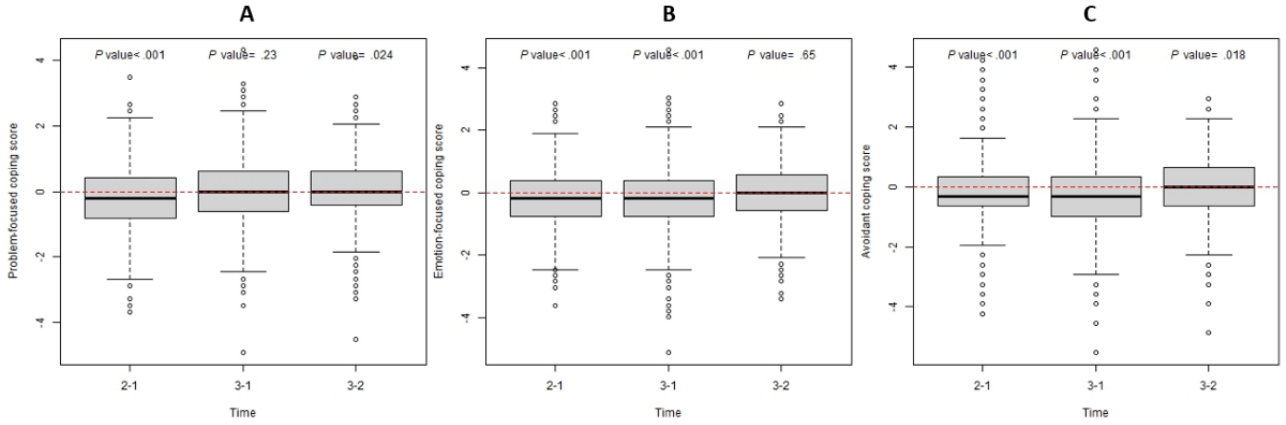

Student Mental Health and Well-Being Outcomes Using Descriptive Statistics, Normality Assessment, Univariate Analyses, and Longitudinal Analyses

Next, we examined self-reported outcomes of depression (PHQ-9), anxiety (GAD-7), loneliness, and flourishing. Although there was no change in average depression (Figure 4A) and anxiety symptoms (Figure 4B) between Year I and Year II, both significantly declined from Year I to Year III (P=.005 and P=.046, respectively) and from Year II to Year III (P<.001 and P<.001, respectively).

The average reported loneliness scores decreased significantly over the 3 study years (P<.001 across all pairwise comparisons; Figure 5A). This correlated with increased average flourishing from Year I to Year III (P<.001) and from Year II to Year III (P<.001; Figure 5B).

Figure 4. Depression (PHQ-9) and anxiety (GAD-7) across the 3-year study period. GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9.

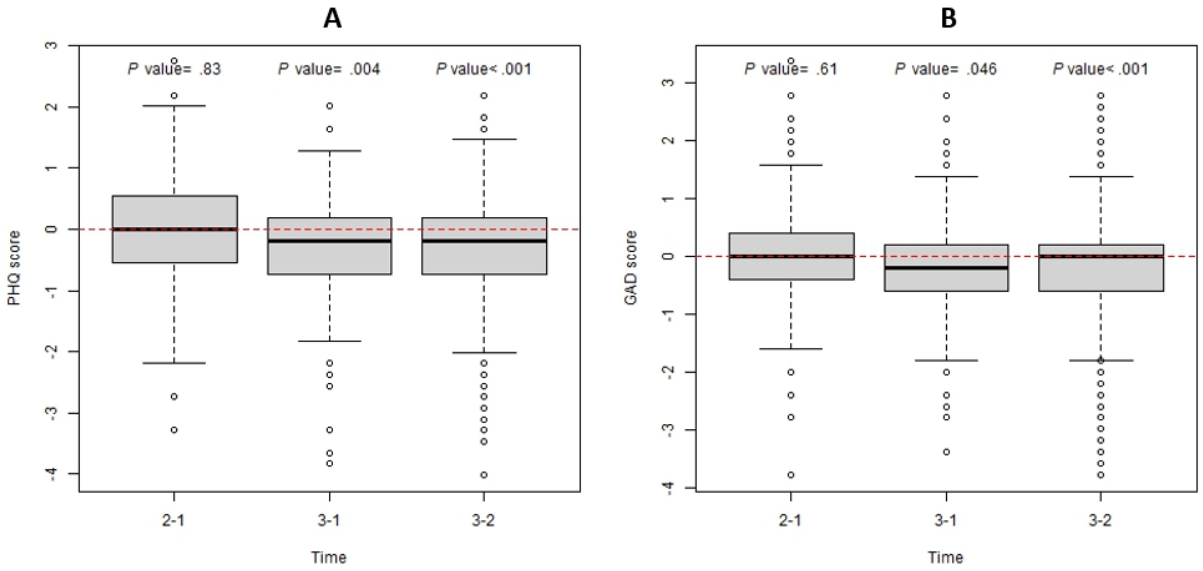

XSL•FO
RenderX

Figure 5. Flourishing and loneliness across the 3-year study period.

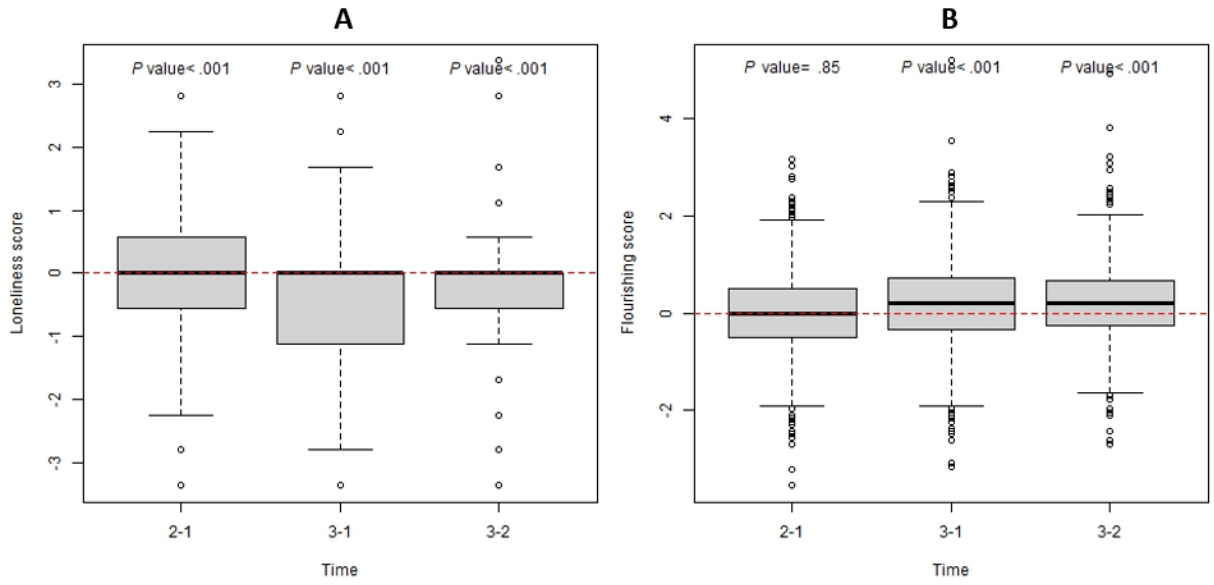

Student Flourishing Assessed by Longitudinal and Mediation Analyses

We explored potential pathways to better understand the factors that may be contributing to changes in students' flourishing over time using a series of multilevel regression analyses guided by a directed acyclic graph (Figure 6). We found that students' flourishing improved significantly with time from the peak of the pandemic (β=.13). This improvement was partially influenced by loneliness and depression. That is, in any given year, students with more loneliness or more depression had lower flourishing (β=–.169 and –.071, respectively), while average loneliness, anxiety, and depression also decreased with the passage of time (β=–.127, –.229, and –.145, respectively). Although a formal mediation analysis confirmed the mediating role of loneliness on flourishing, anxiety and depression were not significant mediators of the relationship between pandemic year or severity and flourishing (Multimedia Appendix 2).

Figure 6. Potential pathways influencing student flourishing. We summarize the multilevel models, where the endpoints of the arrows represent the outcomes, and the starting points indicate the predictors. All predictors in the models are lagged by 1 year, allowing them to forecast the outcomes in the subsequent year. Solid lines with asterisks denote significant *P* values, while dashed lines indicate nonsignificant *P* values. The β symbol represents the estimated value in the respective model.

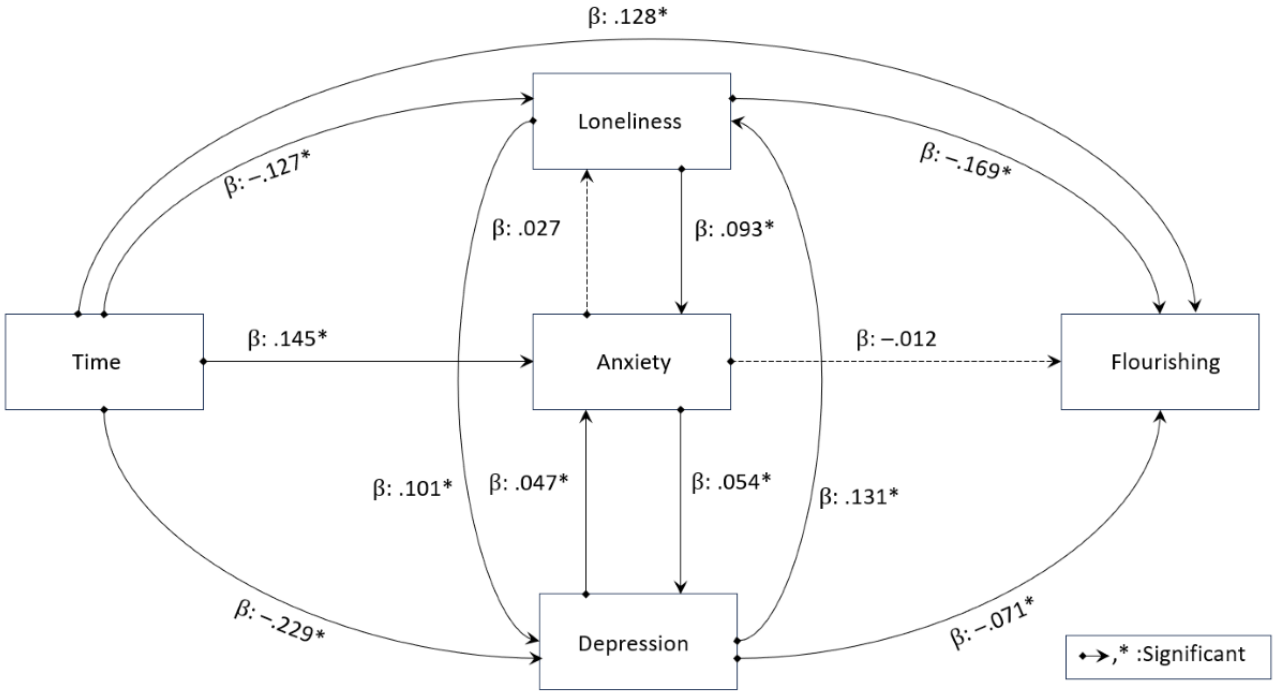

XSL•FO
RenderX

Exploratory App Use Analysis

Next, we examined student app usage patterns over the study period. As shown in the UpSet plot (Figure 7), the 2 most frequently used activities were positive piggy bank and gratitude journal, followed by pleasant activity scheduling, savoring, and engaging with beauty. The least favored activities were love letters, signature strengths, and random acts of kindness.

Figure 7. UpSet plot for activity engagement patterns in Year I and Year II.

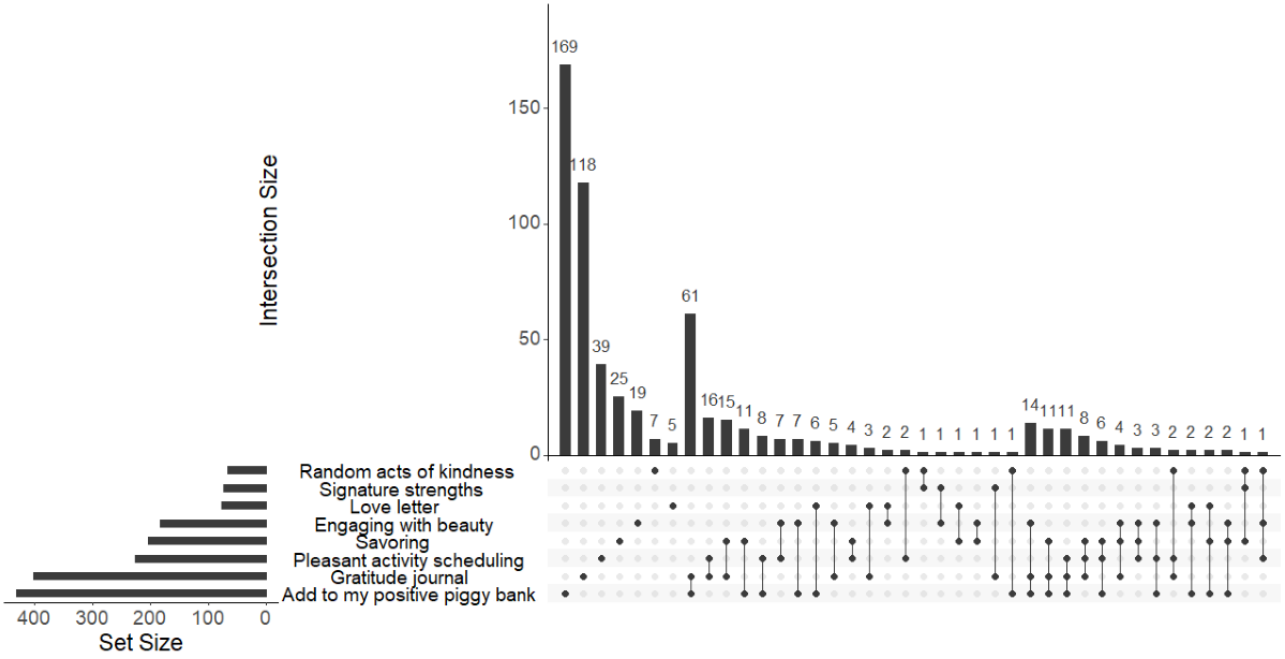

Single activity usage was the most common (ie, only used that activity and none of the others within a "single" session of using the app), but notable combinations included students who used both positive piggy bank and gratitude journaling. Smaller intersections included combinations, such as (1) pleasant activity scheduling and gratitude journal; (2) engaging with beauty and gratitude journal and positive piggy bank; and (3) savoring gratitude journal and positive piggy bank.

This UpSet plot visualizes the intersection of the positive activities engaged by students using the app in Year I and Year II. The x-axis represents unique combinations of activities, while the y-axis indicates the number of users for each combination. Of note, single app usage is represented by the set size on the left: "Positive piggy bank" (431 unique users) and "Gratitude journal" (402 unique users). Other activities included "Pleasant activity scheduling" (227 users), "Savoring" (205 users), and "Engaging with beauty" (184 users). "Love letter" (78 users) and "Signature strengths" (74 users) had moderate engagement, while "Random acts of kindness" (66 users) were the least used.

Based on the above findings, we were interested in determining whether app use correlated with improved flourishing in the subsequent month. After adjusting for demographic and sociobehavioral characteristics in a longitudinal analysis, more frequent app use was associated with greater flourishing (P=.03). For example, implied by Table 2, 30 instances of app use was associated with a nearly 1-unit increase on the flourishing scale. This table presents the estimates and P values for the fixed effects in the mixed effects model.

XSL•FO
RenderX

Table 2. Fixed effects results from mixed effects model.

Variables	Estimate	*P* value
Problem-focused coping lagged	3.062	<.001
Emotion-focused coping lagged	–0.870	.08
Avoidant coping lagged	–1.843	<.001
PHQ-9[a] score lagged	–0.213	<.001
GAD-7[b] score lagged	–1.151	<.001
Loneliness score lagged	–1.483	<.001
Age	.051	.38
App usage	.035	.03
School year		
Undergraduate	Reference	Reference
Graduate	1.726	<.001
No longer a student	1.090	.16
Sex		
Female	Reference	Reference
Male	0.470	.38
Other	0.533	.84
Ethnicity		
No	Reference	Reference
Yes	1.262	.18
International status		
Domestic	Reference	Reference
International	2.451	.01
Generation		
Continuing generation	Reference	Reference
First generation	–1.025	.07
Tobacco history		
No	Reference	Reference
Yes	0.086	.91
Marijuana history		
No	Reference	Reference
Yes	–0.646	.18
Vaping history		
No	Reference	Reference
Yes	–0.683	.21
Alcohol history		
No	Reference	Reference
Yes	0.020	.97
COVID-19 diagnosis		
No	Reference	Reference
Yes	–0.115	.93
Mental condition		

XSL•FO
RenderX

Variables	Estimate	*P* value
No	Reference	Reference
Yes	–4.724	<.001
Race		
White	Reference	Reference
Asian	–3.167	<.001
Black	–2.752	.02
American Indian and Alaska Native	6.892	.20
Multiple	–0.967	.36
Other	–1.121	.56
Year		
2020	Reference	Reference
2021	1.941	<.001
Month		
October	Reference	Reference
November	0.820	.005
December	0.500	.10

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder-7.

Discussion

Principal Findings

The main finding of this study was the reporting of sustained yet evolving mental health challenges among college students over the 3 years. Depression and anxiety levels of our student population were higher than previously reported levels [12,13,58]; they remained consistent between Year I and Year II (2020-2021) but showed significant improvements by Year III (2022). This trend suggests a potential period of initial adjustments to the pandemic's disruptions, followed by gradual mental health recovery as students adapted to the new norms and reduced pandemic-related stressors. Before vaccination, students experienced a multitude of changes related to pandemic restrictions, including lockdowns and stay-at-home orders, remote learning mandates, travel bans, social distancing measures, mask mandates, and closure of nonessential businesses [59].

Building on our previous work [41,42], the initial cohort of college students from the 2020 mHealth Roadmap Study were invited to participate in 2 subsequent semesters in the Fall of 2021 and 2022. Leveraging the use of mHealth technology, we longitudinally followed their mental health and well-being outcomes during a historically unprecedented period marked by ongoing global challenges [13]. In this study, we observed changes in student coping strategies over time. Problem-focused coping initially decreased but then increased between Year II and Year III. Emotion-focused coping consistently decreased from Year I, while avoidant coping showed a steady decline across all time points. These adaptive shifts suggest that students may have developed new strategies to manage themselves. It is possible that the type of coping strategy employed may influence mental health outcomes, as previously discussed [60]. A notable and encouraging finding in this study was the significant decrease in reported feelings of loneliness, which was associated with increases in flourishing across the study period. This suggests that students may have reestablished social connections, possibly through remote platforms initially, mHealth technology, and later in-person interactions as restrictions ease.

Students engaged with the Roadmap app with a strong preference for activities focusing on personal reflection, such as a Positive piggy bank and gratitude journal. Our findings suggest that app use may have positively impacted flourishing, which correlated with increased problem-focused coping and reduced avoidant coping. Reduced levels of depression, anxiety, and loneliness may have also played key roles.

Possible Mechanisms

Counter to our initial hypothesis that mental health outcomes and well-being would worsen over the 3-year study period, our study revealed that, even amid the challenges posed by the pandemic, students reported significant improvements in flourishing over the 3-year period. Flourishing, a term that encapsulates functioning in a manner conducive to growth, resilience, and overall goodness, provides key insights into mental health viewed in a positive manner [61]. This concept aligns with aspects of well-being, including elements such as purpose in life, meaningful relationships, and optimism. Thus, we sought to understand possible mechanisms explaining the link between time from pandemic severity, depression, anxiety, loneliness, and flourishing.

Our data showed that the passage of time from the peak of the pandemic corresponded with higher levels of flourishing.

XSL•FO
RenderX

Furthermore, this increased duration was also associated with decreased feelings of loneliness, anxiety, and depression. An increase in flourishing was partially influenced by lower levels of loneliness and depression in the regression model. Interestingly, the mediation analysis only supported the role of loneliness partially mediating increased flourishing, whereas anxiety and depression did not significantly impact the relationship between time since the pandemic's peak and flourishing. While anxiety had a direct negative effect on depression, anxiety did not influence loneliness. These findings align with ongoing debates about the causal relationships and comorbidity between anxiety and depression [62], as well as their interplay with loneliness [63]. Accordingly, there may be additional factors between loneliness and the constructs of anxiety and depression that were not captured by this study.

Our findings suggest that the time elapsed since the early days of the pandemic may have directly influenced students' flourishing, possibly due to reasons such as increased rates of vaccination, increased opportunities for social interaction with lifts on social restrictions and mask mandates, and return to in-person classroom settings. It is also possible that app use may have contributed to some of our observations when students had access to the Roadmap's positive activities in Year I and Year II of the study. During the early period of social restrictions, mHealth technology may have offered an avenue for students to re-establish social connections and practice simple strategies through well-studied constructs, such as positive daily reflection, gratitude, and savoring.

Comparison with Previous Studies

Previous studies have shown that emotion-focused coping (eg, acceptance, religion, and social support) is associated with improved psychological adjustment, while avoidant coping is linked with increased distress. Problem-focused coping is less used in unpredictable events [64,65] but correlates with positive outcomes [66]. Early in the pandemic, studies indicate that students primarily used emotion-focused coping strategies [65,67,68]. Contrary to these findings, our study participants reported decreased use of problem-focused, emotion-focused, and avoidant coping strategies in the pandemic year (2020). This decline may reflect the widespread use of disengagement strategies at the pandemic's peak. Over time, students in our study reported a decline in avoidant coping with changes in problem-focused and emotion-focused coping, aligning with other recent pandemic-related studies [65,67,68].

The coping patterns observed herein correlated with an overall decline in depression and anxiety symptoms across each subsequent year of the study. Studies have shown varied findings regarding student psychological distress during the pandemic, with some reporting positive self-efficacy in online learning [69]. However, during the same period (2020-2022), a large national survey of college student mental health (Healthy Minds Network) reported increases in the percentages of students with elevated depression and anxiety scores (PHQ-9: 18%-23%; GAD-7: 31%-37%) [58]. Although we acknowledge differences in participant pools, our study followed the same cohort across the 3 years, while the Healthy Minds Network recruits new students annually; a notable difference is that our students had access to a positive psychology–based mHealth app (Roadmap). In contrast, to our knowledge, access to mHealth app support in the Healthy Minds Network population is not known.

Students who engaged with the Roadmap app favored activities that focused on personal reflection and mindfulness (ie, inward-facing), such as the positive piggy bank and gratitude journal. A recent study showed that mindfulness meditation delivered through mHealth was an approach used by college students to reduce stress and improve mindfulness and self-compassion [70]. Interestingly, we observed lower engagement in activities, such as random acts of kindness or love letters, which involved interaction with others or had an external focus (ie, "outward-facing"). It is possible that during periods of strict lockdown and social distancing, activities requiring physical interaction or outward expressions of kindness, like random acts of kindness or writing love letters, may have been perceived as difficult or impractical. The pandemic created barriers, particularly at colleges, to social interactions, making it harder for students to engage in activities that relied on direct, personal contact or social exchange [71]. During significant life transitions, of both university life as well as the COVID-19 pandemic, it is possible students favored more inward-facing strategies to support their well-being rather than prosocial, outward-facing strategies [72].

Public Health Impact

The literature suggests that the pandemic may have exacerbated mental health challenges among college students, a population already experiencing increased reports of anxiety, depression, and stress. The experience of social isolation, particularly during a crucial period of social interactions and development growth as emerging adults, may have led to poorer management of other aspects of well-being. Understanding these trends and identifying groups with increased prevalence are crucial for developing novel approaches to mental health care aimed at effectively addressing these issues. To mitigate these challenges, the adoption of novel mHealth technologies through smartphone apps has significantly accelerated over the past several years [73]. Such technologies, as leveraged in this study, hold great potential to fundamentally address mental health challenges, particularly in college students [74].

Limitations

This study had several limitations. Students were self-selected to be in this study, which was meant to focus on positive psychology-based strategies to enhance mental health. It relied on self-report surveys to collect information about mental health and well-being, and the majority of participants were female. Approximately one-third of the initial participants did not take part in subsequent years of the study (ie, lost to follow-up). In addition, as we followed students over the course of 3 years, some graduated and were no longer on campus, potentially reducing their access to college resources. It is important to note that our findings are specific to students from a single Midwestern location and may not be generalizable to other regions or student populations. For example, certain demographic groups might be underrepresented or overrepresented due to the college's location in the Midwest. We did not collect data on participants' ongoing or concurrent

XSL•FO
RenderX

mental health treatments or medications. Finally, we acknowledge the lack of a fully powered, randomized controlled trial design. Despite these limitations, this study constitutes a large-scale, longitudinal study conducted during an unprecedented time, providing valuable insights into the mental health challenges of college students.

Conclusions

This longitudinal study of college students observed a decline in average loneliness, depression, and anxiety with increasing time from the pandemic peak that coincided with increased flourishing. This increased flourishing was partially but not fully explained by improved loneliness, anxiety, and depression. Students had access to a positive psychology-based app and used variable coping strategies during the study. Future research should examine the efficacy of the Roadmap app through a randomized controlled trial, using a robust study design to assess its impact on mental health outcomes. To our knowledge, this study did not produce any unintended consequences through student participation. There were no privacy breaches or technical problems during the study that may have impacted the outcomes.

Acknowledgments

The A Alfred Taubman Medical Research Institute supported the work herein as one of its Taubman Institute Innovation Projects to MT and SWC. This work was supported in part by an Ideas Lab grant from the Biosciences Initiative of the University of Michigan to SWC and MT. National Institutes of Health's National Heart, Lung, and Blood Institute (NIH/NHLBI) (1R01HL146354), NIH/NHLBI (K24HL156896), and NIH/NCI grants (R01CA249211) and the Edith S Briskin and Shirley K Schlafer Foundation support the work of SWC. We wish to thank the University of Michigan students (undergraduate and graduate) who participated in this study. We wish to thank Kristen Gilley, Caroline Clingan, Christine Cislo, Bushra Hussain, Rebecca Vue, Jacob Kedroske, Skylar Ketteler, and Manasa Dittakavi for their time in the student recruitment and onboarding years of the study. SS and SNH presented portions of this paper at the American Medical Informatics Association (AMIA) Annual Meeting (November 9-13, 2024, San Francisco, CA) and received a Distinguished Poster Award.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. Data are available on request. The study protocol and statistical code can be obtained from SWC and XC. The dataset is available from the University of Michigan through a data use agreement.

Authors' Contributions

GJ drafted the manuscript and curated the data and figures; this was part of his University of Michigan Honors Thesis project. MR coordinated the study, carried out recruitment, provided consent, carried out onboarding, curated the data and figures, and critically reviewed and edited the manuscript. GJ and SWC performed the background literature review, visualized the data, and critically edited and revised the manuscript. AH provided expertise in psychometric analyses, data interpretation, and data visualization and critically edited and revised the manuscript. XC cleaned and processed the data, analyzed the data, curated the data and figures, and critically reviewed and edited the manuscript. SS and SNH assisted in data curation and reviewed and edited the manuscript. DAH assisted in data security and privacy related to the life cycle of app development, implementation, and maintenance. DAH served as a liaison between the clinical study team and the Health Information Technology Services at the University of Michigan. DAH reviewed and edited the manuscript. KS supervised the statistical analysis and data interpretation and critically reviewed and edited the manuscript. MT and SWC envisioned and designed the study, procured resources, supervised the study, interpreted data, and helped draft, edit, and critically revise the manuscript. SWC created the Roadmap app.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Roadmap 2.0 app multicomponent features defined.
[DOCX File , 124 KB - jmir_v27i1e67627_app1.docx]

Multimedia Appendix 2
Mediation analysis result between pandemic year and flourishing.
[DOCX File , 38 KB - jmir_v27i1e67627_app2.docx]

References

1. National Center for Education Statistics. URL: https://nces.ed.gov/fastfacts/display.asp?id=372#College-enrollment [accessed 2024-03-12]

XSL•FO
RenderX

2. World Health Organization. Fact Sheets: Adolescent and Young Adult Health. URL: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2024-03-12]
3. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/whats-new-all.html [accessed 2024-03-12]
4. Colizzi M, Lasalvia A, Ruggeri M. Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? Int J Ment Health Syst 2020;14:23 [FREE Full text] [doi: 10.1186/s13033-020-00356-9] [Medline: 32226481]
5. Hefner J, Eisenberg D. Social support and mental health among college students. Am J Orthopsychiatry 2009;79(4):491-499 [FREE Full text] [doi: 10.1037/a0016918] [Medline: 20099940]
6. Arria AM, Caldeira KM, Vincent KB, Winick ER, Baron RA, O'Grady KE. Discontinuous college enrollment: associations with substance use and mental health. Psychiatr Serv 2013;64(2):165-172 [FREE Full text] [doi: 10.1176/appi.ps.201200106] [Medline: 23474608]
7. Lipson SK, Eisenberg D. Mental health and academic attitudes and expectations in university populations: results from the healthy minds study. J Ment Health 2018;27(3):205-213 [FREE Full text] [doi: 10.1080/09638237.2017.1417567] [Medline: 29265935]
8. Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. Am J Orthopsychiatry 2007;77(4):534-542 [FREE Full text] [doi: 10.1037/0002-9432.77.4.534] [Medline: 18194033]
9. Gollust SE, Eisenberg D, Golberstein E. Prevalence and correlates of self-injury among university students. Journal of American College Health 2010;56(5):491-498 [FREE Full text] [doi: 10.3200/jach.56.5.491-498]
10. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. Epidemiol Rev 2008;30(1):133-154 [FREE Full text] [doi: 10.1093/epirev/mxn002] [Medline: 18653727]
11. Mortier P, Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, et al. Suicidal thoughts and behaviors among college students and same-aged peers: results from the World Health Organization World Mental health surveys. Soc Psychiatry Psychiatr Epidemiol 2018;53(3):279-288 [FREE Full text] [doi: 10.1007/s00127-018-1481-6] [Medline: 29340781]
12. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
13. Lederer AM, Hoban MT, Lipson SK, Zhou S, Eisenberg D. More than inconvenienced: the unique needs of U.S. college students during the COVID-19 pandemic. Health Educ Behav 2021;48(1):14-19 [FREE Full text] [doi: 10.1177/1090198120969372] [Medline: 33131325]
14. Lipson SK, Lattie E, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
15. Student Mental Health Innovative Approaches Review Committee Report. URL: https://studentlife.umich.edu/article/student-mental-health-innovative-approaches-review [accessed 2024-03-12]
16. Matsayi Aji L, Baba Muhammad A, Abubakar H. Psychosocial Care. In: Emel Önal A, editor. Tertiary Care - Medical, Psychosocial, and Environmental Aspects. London: IntechOpen; 2024.
17. Eisenberg D. Countering the troubling increase in mental health symptoms among U.S. college students. J Adolesc Health 2019;65(5):573-574 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.08.003] [Medline: 31648753]
18. McAleavey AA, Youn SJ, Xiao H, Castonguay LG, Hayes JA, Locke BD. Effectiveness of routine psychotherapy: Method matters. Psychother Res 2019;29(2):139-156. [doi: 10.1080/10503307.2017.1395921] [Medline: 29096584]
19. Czyz EK, Horwitz AG, Eisenberg D, Kramer A, King CA. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. J Am Coll Health 2013;61(7):398-406 [FREE Full text] [doi: 10.1080/07448481.2013.820731] [Medline: 24010494]
20. Gallagher RP. National survey of counseling center directors 2005. In: The International Association of Counseling Services (IACS). South Kolkata: IACS; 2006.
21. Eisenberg D, Downs MF, Golberstein E, Zivin K. Stigma and help seeking for mental health among college students. Med Care Res Rev 2009;66(5):522-541 [FREE Full text] [doi: 10.1177/1077558709335173] [Medline: 19454625]
22. Golberstein E, Eisenberg D, Gollust SE. Perceived stigma and mental health care seeking. Psychiatr Serv 2008;59(4):392-399 [FREE Full text] [doi: 10.1176/ps.2008.59.4.392] [Medline: 18378838]
23. Bidargaddi N, Schrader G, Klasnja P, Licinio J, Murphy S. Designing m-health interventions for precision mental health support. Transl Psychiatry 2020;10(1):222 [FREE Full text] [doi: 10.1038/s41398-020-00895-2] [Medline: 32636358]
24. Huckvale K, Nicholas J, Torous J, Larsen ME. Smartphone apps for the treatment of mental health conditions: status and considerations. Curr Opin Psychol 2020;36:65-70 [FREE Full text] [doi: 10.1016/j.copsyc.2020.04.008] [Medline: 32553848]
25. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]

XSL•FO
RenderX

26. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
27. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
28. Rozwadowski M, Dittakavi M, Mazzoli A, Hassett AL, Braun T, Barton D, et al. Promoting health and well-being through mobile health technology (Roadmap 2.0) in family caregivers and patients undergoing hematopoietic stem cell transplantation: protocol for the development of a mobile randomized controlled trial. JMIR Res Protoc 2020;9(9):e19288 [FREE Full text] [doi: 10.2196/19288] [Medline: 32945777]
29. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. American Psychologist 2005;60(5):410-421 [FREE Full text] [doi: 10.1037/0003-066x.60.5.410]
30. Seligman ME, Rashid T, Parks AC. Positive psychotherapy. American Psychologist 2006;61(8):774-788. [doi: 10.1037/0003-066x.61.8.774]
31. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003;84(2):377-389. [doi: 10.1037//0022-3514.84.2.377] [Medline: 12585811]
32. Kashdan TB, Uswatte G, Julian T. Gratitude and hedonic and eudaimonic well-being in Vietnam war veterans. Behav Res Ther 2006;44(2):177-199. [doi: 10.1016/j.brat.2005.01.005] [Medline: 16389060]
33. Moskowitz JT, Hult JR, Duncan LG, Cohn MA, Maurer S, Bussolari C, et al. A positive affect intervention for people experiencing health-related stress: development and non-randomized pilot test. J Health Psychol 2012;17(5):676-692 [FREE Full text] [doi: 10.1177/1359105311425275] [Medline: 22021272]
34. Cohn MA, Pietrucha ME, Saslow LR, Hult JR, Moskowitz JT. An online positive affect skills intervention reduces depression in adults with type 2 diabetes. J Posit Psychol 2014;9(6):523-534 [FREE Full text] [doi: 10.1080/17439760.2014.920410] [Medline: 25214877]
35. Bryant FB. A four‐factor model of perceived control: avoiding, coping, obtaining, and savoring. Journal of Personality 2006;57(4):773-797. [doi: 10.1111/j.1467-6494.1989.tb00494.x]
36. Bolier L, Haverman M, Westerhof G, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
37. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009;65(5):467-487 [FREE Full text] [doi: 10.1002/jclp.20593] [Medline: 19301241]
38. Hassett AL, Finan PH. The role of resilience in the clinical management of chronic pain. Curr Pain Headache Rep 2016;20(6):39 [FREE Full text] [doi: 10.1007/s11916-016-0567-7] [Medline: 27115770]
39. Carr A, Cullen K, Keeney C, Canning C, Mooney O, Chinseallaigh E, et al. Effectiveness of positive psychology interventions: a systematic review and meta-analysis. The Journal of Positive Psychology 2020;16(6):749-769 [FREE Full text] [doi: 10.1080/17439760.2020.1818807]
40. Bolier L, Majo C, Smit F, Westerhof G, Haverman M, Walburg J, et al. Cost-effectiveness of online positive psychology: randomized controlled trial. The Journal of Positive Psychology 2014;9(5):460-471 [FREE Full text] [doi: 10.1080/17439760.2014.910829]
41. Cislo C, Clingan C, Gilley K, Rozwadowski M, Gainsburg I, Bradley C, et al. Monitoring beliefs and physiological measures in students at risk for COVID-19 using wearable sensors and smartphone technology: protocol for a mobile health study. JMIR Res Protoc 2021;10(6):e29561 [FREE Full text] [doi: 10.2196/29561] [Medline: 34115607]
42. Gilley KN, Baroudi L, Yu M, Gainsburg I, Reddy N, Bradley C, et al. Risk factors for COVID-19 in college students identified by physical, mental, and social health reported during the fall 2020 semester: observational study using the Roadmap app and Fitbit wearable sensors. JMIR Ment Health 2022;9(2):e34645 [FREE Full text] [doi: 10.2196/34645] [Medline: 34992051]
43. Pandya A, Lodha P. Mental health consequences of COVID-19 pandemic among college students and coping approaches adapted by higher education institutions: a scoping review. SSM Ment Health 2022;2:100122 [FREE Full text] [doi: 10.1016/j.ssmmh.2022.100122] [Medline: 35665095]
44. Electronic signature that scales with your workflow. signNow. URL: https://www.signnow.com/ [accessed 2024-12-03]
45. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100 [FREE Full text] [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
46. Poulus D, Coulter TJ, Trotter MG, Polman R. Stress and coping in esports and the influence of mental toughness. Front Psychol 2020;11:628 [FREE Full text] [doi: 10.3389/fpsyg.2020.00628] [Medline: 32390900]
47. VanderWeele TJ. On the promotion of human flourishing. Proc Natl Acad Sci U S A 2017;114(31):8148-8156 [FREE Full text] [doi: 10.1073/pnas.1702996114] [Medline: 28705870]
48. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980;39(3):472-480 [FREE Full text] [doi: 10.1037//0022-3514.39.3.472] [Medline: 7431205]

XSL•FO
RenderX

49. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
50. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care 2008;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
51. Cameron I, Crawford J, Lawton K, Reid I. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract 2008;58(546):32-36 [FREE Full text] [doi: 10.3399/bjgp08x263794]
52. Royston JP. An extension of Shapiro and Wilk's w test for normality to large samples. Applied Statistics 1982;31(2):115 [FREE Full text] [doi: 10.2307/2347973]
53. Bauer DF. Constructing confidence sets using rank statistics. Journal of the American Statistical Association 1972;67(339):687 [FREE Full text] [doi: 10.2307/2284469]
54. Fahrmeir L. Book review: an introduction to categorical data analysis (2nd edition). By A Agresti. Biometrical J 2008;50(6):1098-1098 [FREE Full text] [doi: 10.1002/bimj.200810478]
55. Best DJ, Roberts DE. Algorithm AS 89: the upper tail probabilities of Spearman's rho. Applied Statistics 1975;24(3):377 [FREE Full text] [doi: 10.2307/2347111]
56. van Buuren S, Groothuis-Oudshoorn CG. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45:1-67 [FREE Full text]
57. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
58. Healthy Minds Study among Colleges and Universities. Healthy Minds Network. URL: https://healthymindsnetwork.org/research/data-for-researchers [accessed 2024-12-03]
59. Krishnamachari B, Morris A, Zastrow D, Dsida A, Harper B, Santella A. The role of mask mandates, stay at home orders and school closure in curbing the COVID-19 pandemic prior to vaccination. Am J Infect Control 2021;49(8):1036-1042 [FREE Full text] [doi: 10.1016/j.ajic.2021.02.002] [Medline: 33577824]
60. Horwitz AG, Hill R, King C. Specific coping behaviors in relation to adolescent depression and suicidal ideation. J Adolesc 2011;34(5):1077-1085 [FREE Full text] [doi: 10.1016/j.adolescence.2010.10.004] [Medline: 21074841]
61. Keyes CLM. The mental health continuum: From languishing to flourishing in life. J. Health Soc. Behav 2002;43(2):207-222 [FREE Full text] [doi: 10.2307/3090197]
62. Cohen JR, Andrews AR, Davis MM, Rudolph KD. Anxiety and depression during childhood and adolescence: testing theoretical models of continuity and discontinuity. J Abnorm Child Psychol 2018;46(6):1295-1308 [FREE Full text] [doi: 10.1007/s10802-017-0370-x] [Medline: 29256025]
63. Hawkley LC, Cacioppo J. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
64. Carver CS, Scheier M, Weintraub JK. Assessing coping strategies: a theoretically based approach. J Pers Soc Psychol 1989;56(2):267-283. [doi: 10.1037//0022-3514.56.2.267] [Medline: 2926629]
65. Chu GM, Goger P, Malaktaris A, Lang AJ. The role of threat appraisal and coping style in psychological response to the COVID-19 pandemic among university students. J Affect Disord Rep 2022;8:100325 [FREE Full text] [doi: 10.1016/j.jadr.2022.100325] [Medline: 35169766]
66. Finkelstein-Fox L, Park C, Riley K. Mindfulness' effects on stress, coping, and mood: a daily diary goodness-of-fit study. Emotion 2019;19(6):1002-1013. [doi: 10.1037/emo0000495] [Medline: 30138006]
67. Okafor CN, Bautista KJ, Asare M, Opara I. Coping in the time of COVID-19: buffering stressors with coping strategies. J Loss Trauma 2022;27(1):83-91 [FREE Full text] [doi: 10.1080/15325024.2021.1914987] [Medline: 35173557]
68. Prasath PR, Mather PC, Bhat CS, James JK. University student well-being during COVID-19: the role of psychological capital and coping strategies. Prof Couns 2021;11(1):46-60 [FREE Full text]
69. Graham MA, Eloff I. Comparing mental health, wellbeing and flourishing in undergraduate students pre- and during the COVID-19 pandemic. Int J Environ Res Public Health 2022;19(12):7438 [FREE Full text] [doi: 10.3390/ijerph19127438] [Medline: 35742686]
70. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
71. Sharaievska I, McAnirlin O, Browning M, Larson L, Mullenbach L, Rigolon A, et al. "Messy transitions": students' perspectives on the impacts of the COVID-19 pandemic on higher education. High Educ (Dordr) 2022:1-18 [FREE Full text] [doi: 10.1007/s10734-022-00843-7] [Medline: 35463941]
72. Geller L, Greenberg M. Managing the transition process from high school to college and beyond: challenges for individuals, families, and society. Social Work in Mental Health 2009;8(1):92-116 [FREE Full text] [doi: 10.1080/15332980902932466]
73. Topol EJ. A decade of digital medicine innovation. Sci Transl Med 2019;11(498):eaaw7610. [doi: 10.1126/scitranslmed.aaw7610] [Medline: 31243153]

XSL•FO
RenderX

74. Ferrari M, Allan S, Arnold C, Eleftheriadis D, Alvarez-Jimenez M, Gumley A, et al. Digital interventions for psychological well-being in university students: systematic review and meta-analysis. J Med Internet Res 2022;24(9):e39686 [FREE Full text] [doi: 10.2196/39686] [Medline: 36169988]

Abbreviations

GAD-7: Generalized Anxiety Disorder-7
HRQOL: health-related quality of life
IRB: institutional review board
PHQ-9: Patient Health Questionnaire-9

Edited by T de Azevedo Cardoso; submitted 21.10.24; peer-reviewed by R Kumar, C Haridas, M Gasmi ; comments to author 18.11.24; revised version received 05.12.24; accepted 10.12.24; published 09.01.25.

Please cite as:
Jayaraj G, Cao X, Horwitz A, Rozwadowski M, Shea S, Hanauer SN, Hanauer DA, Tewari M, Shedden K, Choi SW
Trends in Mental Health Outcomes of College Students Amid the Pandemic (Roadmap mHealth App): Longitudinal Observational Study
J Med Internet Res 2025;27:e67627
URL: https://www.jmir.org/2025/1/e67627
doi:10.2196/67627
PMID:

©Gautham Jayaraj, Xiao Cao, Adam Horwitz, Michelle Rozwadowski, Skyla Shea, Shira N Hanauer, David A Hanauer, Muneesh Tewari, Kerby Shedden, Sung Won Choi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Therapeutic Potential of Social Chatbots in Alleviating Loneliness and Social Anxiety: Quasi-Experimental Mixed Methods Study

Myungsung Kim[1], BS; Seonmi Lee[1], BS; Sieun Kim[1], BS; Jeong-in Heo[1], BS; Sangil Lee[2], PhD; Yu-Bin Shin[3], PhD; Chul-Hyun Cho[3,4*], MD, PhD; Dooyoung Jung[1,2*], MD, PhD

[1]Graduate School of Health Science and Technology, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[2]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[3]Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea

[4]Department of Biomedical Informatics, Korea University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Chul-Hyun Cho, MD, PhD
Department of Psychiatry
Korea University College of Medicine
73 Goryeodae-ro, Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 029205505
Email: david0203@gmail.com

Abstract

Background: Artificial intelligence (AI) social chatbots represent a major advancement in merging technology with mental health, offering benefits through natural and emotional communication. Unlike task-oriented chatbots, social chatbots build relationships and provide social support, which can positively impact mental health outcomes like loneliness and social anxiety. However, the specific effects and mechanisms through which these chatbots influence mental health remain underexplored.

Objective: This study explores the mental health potential of AI social chatbots, focusing on their impact on loneliness and social anxiety among university students. The study seeks to (i) assess the impact of engaging with an AI social chatbot in South Korea, "Luda Lee," on these mental health outcomes over a 4-week period and (ii) analyze user experiences to identify perceived strengths and weaknesses, as well as the applicability of social chatbots in therapeutic contexts.

Methods: A single-group pre-post study was conducted with university students who interacted with the chatbot for 4 weeks. Measures included loneliness, social anxiety, and mood-related symptoms such as depression, assessed at baseline, week 2, and week 4. Quantitative measures were analyzed using analysis of variance and stepwise linear regression to identify the factors affecting change. Thematic analysis was used to analyze user experiences and assess the perceived benefits and challenges of chatbots.

Results: A total of 176 participants (88 males, average age=22.6 (SD 2.92)) took part in the study. Baseline measures indicated slightly elevated levels of loneliness (UCLA Loneliness Scale, mean 27.97, SD (11.07)) and social anxiety (Liebowitz Social Anxiety Scale, mean 25.3, SD (14.19)) compared to typical university students. Significant reductions were observed as loneliness decreasing by week 2 (t_{175}=2.55, P=.02) and social anxiety decreasing by week 4 (t_{175}=2.67, P=.01). Stepwise linear regression identified baseline loneliness (β=0.78, 95% CI 0.67 to 0.89), self-disclosure (β=–0.65, 95% CI –1.07 to –0.23) and resilience (β=0.07, 95% CI 0.01 to 0.13) as significant predictors of week 4 loneliness (R^2=0.64). Baseline social anxiety (β=0.92, 95% CI 0.81 to 1.03) significantly predicted week 4 anxiety (R^2=0.65). These findings indicate higher baseline loneliness, lower self-disclosure to the chatbot, and higher resilience significantly predicted higher loneliness at week 4. Additionally, higher baseline social anxiety significantly predicted higher social anxiety at week 4. Qualitative analysis highlighted the chatbot's empathy and support as features for reliability, though issues such as inconsistent responses and excessive enthusiasm occasionally disrupted user immersion.

Conclusions: Social chatbots may have the potential to mitigate feelings of loneliness and social anxiety, indicating their possible utility as complementary resources in mental health interventions. User insights emphasize the importance of empathy, accessibility, and structured conversations in achieving therapeutic goals.

XSL•FO
RenderX

Trial Registration: Clinical Research Information Service (CRIS) KCT0009288; https://tinyurl.com/hxrznt3t

(J Med Internet Res 2025;27:e65589) doi:10.2196/65589

KEYWORDS

artificial intelligence; AI; social chatbot; loneliness; social anxiety; exploratory research; mixed methods study

Introduction

Background

The emergence of chatbots marked a pivotal turning point at the intersection of technology and human activity. By facilitating interactions with users through the exchange of natural language, chatbots simplify interactions and enhance user engagement [1]. In the field of psychiatry, chatbots have provided useful information in response to user questions [2] and have shown tangible therapeutic effects through psychological therapies, such as cognitive behavioral therapy [3,4]. Various studies have highlighted the potential of chatbots as an effective medium for digital self-help. It was also discovered that forming a therapeutic alliance through an intimate relationship between the user and the chatbot is crucial for enhancing the chatbot's therapeutic effect [5,6].

Advancements in artificial intelligence (AI) and natural language processing technologies have facilitated the emergence of large-scale language models (LLMs), leading to the development of a new type of chatbot known as the social chatbot. Unlike task-oriented or clinical chatbots, social chatbots focus on building relationships through conversations, thereby offering more natural and emotional communication. Recent studies have qualitatively explored the potential of social chatbots and their impact on mental health [7]. Social chatbots provide social support that can affect mental health by offering a nonjudgmental and readily available communication channel [8,9]. They can serve as substitutes for friends, alleviating loneliness, in addition to clinical therapy [10].

One clue regarding the psychiatric use of social chatbots is their persona. Through specific personas, users can engage in conversations similar to those with close friends, facilitating a space where they can share personal stories openly and receive support [11]. Another clue is that empathetic responses from social chatbots can help build effective relationships [12]. Social chatbots with appropriate personas and empathetic responses are expected to build intimate relationships with users and positively affect mental health, including loneliness [10]. However, further research is needed to determine the specific duration and effects of social chatbots as psychiatric tools and the causes of these effects.

As the intersection between technology and mental health continues to evolve, this study undertakes an exploratory analysis to understand the impact of social chatbots on mental health. By engaging individuals in their twenties with social chatbots over a 4-week period and assessing their mental health at biweekly intervals, this study aims to elucidate the nuanced effects that social chatbots might have. Additionally, by gathering data on user experiences and reactions to social chatbots, this study seeks to inform future advancements in chatbot design and applications, building on insights from user experiences and feedback.

Aim

The primary goal of this study was to explore the psychiatric potential of social chatbots, focusing on their impact on mental health through a combined approach of qualitative insights and quantitative evaluations. This study specifically aimed to (1) investigate the changes in mental health outcomes (loneliness, social anxiety, and positive or negative affect) during social chatbot use and identify the key factors driving these changes and (2) conduct a qualitative analysis of user experiences to gain a deeper understanding of the perceived strengths and weaknesses of social chatbots as well as their potential applicability in therapeutic contexts. This approach offers a comprehensive overview of the roles of social chatbots as therapeutic tools, contributing valuable knowledge to the field of mental health interventions using generative AI.

Methods

Recruitment

The recruitment was conducted using a web-based platform called "Everytime," which is widely used among university students in South Korea. The inclusion criteria were as follows: students aged 19-29 years who were willing to use social chatbots and had no difficulties conversing with them. The exclusion criteria included applicants showing signs of severe mental illness or suicidal ideation, as this study did not aim to verify direct therapeutic effects. Preliminary screening excluded individuals with suicidal thoughts based on their response to the last question of the Patient Health Questionnaire-9 (PHQ-9), which concerned thoughts that they would be better off dead or hurt themselves. The trial was registered with the Clinical Research Information Service under registration number KCT0009288, with the unique study number UNISTIRB-22-024-A.

Settings and Design

This study used a single-group pre-post design with repeated measures. Over 4 weeks between September and October 2023, the participants were encouraged to interact with the chatbot at least 3 times per week. Despite this encouragement, some participants interacted with the chatbot fewer than 3 times per week, but they were not excluded from the study. However, participants who failed to complete the initial survey or did not install the chatbot within the first week were dropped from the research.

Data were collected through web-based surveys to which participants could respond conveniently. Surveys were collected at baseline (week 0), midpoint (week 2), and end of the study (week 4), with the initial survey gathering basic information

XSL•FO
RenderX

about the participants and the final survey including open-ended questions about their experiences using the social chatbot. To determine the sample size for this single-arm pre-post design, we conducted an a priori power analysis using GPower software. Prior research demonstrated that interaction with a similar type of empathic chatbot yielded a moderate effect size of 0.42 (Cohen *d*; 95% CI 0.13-0.71) in improving users' positive mood after a social exclusion scenario [13]. However, considering that our intervention—a social chatbot—is designed for more casual, everyday interactions with less structured scenarios, we set the expected effect size at 0.3, with an α level of 0.05 and a desired power of 0.80. The GPower analysis recommended a minimum sample size of 170 participants to reach adequate statistical power. However, considering the 4-week duration of the study and the likelihood of participant attrition, we aimed to recruit over 200 participants to ensure sufficient retention and reliable results throughout the study.

Chatting with Social Chatbot "Luda Lee"

"Luda Lee" is a social chatbot designed with the persona of a 22-year-old female college student using Korean language data [14]. It was the first chatbot to be introduced into the Nutty social chatbot app (ScatterLab Inc), which recorded over 1 million downloads, making it popular among commercial applications in Korea. Luda's primary goal is not to provide direct mental care but to become friends with users and engage in frequent conversations. Although the app includes features such as the provision of paid gifts and playing minigames, these functions were restricted to chatting with Luda.

Measures

Loneliness, social anxiety, and positive or negative affect were measured as the main outcomes at baseline (week 0), midpoint (week 2), and at the end of the study (week 4) for a total of 3 times. Loneliness was assessed using the 20-item UCLA Loneliness Scale (ULS) [15,16], which measures the chronic characteristics and state of loneliness. Social anxiety was measured with the 24-item Liebowitz Social Anxiety Scale (LSAS) [17,18], which evaluates the situational aspects of social phobia. Affects were assessed using the 20-item Positive Affect and Negative Affect Schedule [19,20].

Depression, general anxiety, and stress were measured as exploratory outcomes at the same time points to investigate the potential simultaneous occurrences and their influence on the study results. The exploratory outcomes included depression, general anxiety, and stress, which were measured at the same 3 time points. Depression was assessed using the PHQ-9 [21,22], which was developed to detect depression in primary care settings and assist in its diagnosis. General anxiety was measured using the 7-item General Anxiety Disorder-7 (GAD-7) [23,24], and stress was assessed using the Perceived Stress Scale-10 (PSS-10) [25].

The baseline variables included gender, age, resilience, education level, experience with social chatbots, and experience with LLMs. Resilience was measured using the 25-item Connor-Davidson Resilience Scale [26,27].

Acceptance variables that were measured at the end of the study included perceived usefulness and perceived ease of use of the chatbot, intimacy with the chatbot, and self-disclosure level of the user. The perceived usefulness and ease of use were measured using scales adapted to the context of social chatbot use based on the technology acceptance model [28]. Intimacy and self-disclosure were assessed using items from the research on user experiences with chatbots [29].

User experiences were also collected using open-ended questions after 4 weeks of using the social chatbot. Participants were asked to write about their experiences, including helpful aspects, memorable moments, and any areas of disappointment, to identify the features of the social chatbot found by "Luda Lee." Additionally, they were asked to explain why these kinds of social chatbots might be helpful for certain individuals to identify the psychiatrically effective features of social chatbots and directions for future improvement.

Ethical Considerations

This study was conducted following the approval of the Institutional Review Board of the Ulsan National Institute of Science and Technology (approval number: UNISTIRB-22 024). The approval underscores the study's commitment to ethical standards, ensuring the protection of participants' rights and safety throughout the research process.

Prior to participation, all individuals were provided with a detailed explanation of the study, including its objectives, procedures, and potential risks. Informed consent was obtained through a Google Form, where participants acknowledged their understanding and voluntary agreement to partake in the study. To prioritize mental well-being, individuals with suicidal ideation, as indicated by their response to the ninth item on the PHQ-9 questionnaire, were excluded from participation.

Special care was taken to exclude those who might feel uncomfortable with AI chatbots to maintain participant comfort. Participants who completed the full 4-week study were compensated with 50,000 KRW (approximately US $40).

All data collected during the study were anonymized to ensure the confidentiality and privacy of participants. This included sensitive information such as app usage, chat history, and survey results, all of which were protected to uphold participant privacy and prevent data leakage. Participants were informed that their data would be used solely for research purposes. The consent process also highlighted their right to withdraw from the study at any time without penalty.

If any participant experienced discomfort, the study was immediately halted for them, ensuring the utmost respect for their autonomy and well-being. In addition, clinical psychologists (author SaL) and psychiatrists (corresponding authors DJ and CHC) were prepared to connect participants expressing discomfort to appropriate mental health resources, ensuring access to professional support as needed.

Analysis

Statistical Method for Survey

The primary statistical method used was repeated-measures analysis of variance to analyze the mental health scale scores at 3 different time points. Following significant changes

XSL•FO
RenderX

observed through repeated-measures ANOVA, post hoc analyses were conducted using 2-tailed paired *t* tests to pinpoint the exact moments of significant change in the variables. To assess the effect of external variables on the observed changes, further analyses using stepwise linear regression were conducted. This method allowed for the iterative selection of highly relevant independent variables such as age, gender, acceptance variables, and resilience, which significantly influenced the dependent variables that exhibited changes. Statistical significance was set at $P<.05$, and all analyses were conducted using the stepAIC function from the MASS library in statistical software R with the direction parameter set to "both" to ensure the reliability and validity of the results. This statistical approach facilitates a detailed exploration of whether changes occur in the mental health variables under study, when and what kind of changes occur, and which external variables influence these changes.

Thematic Analysis for User Experience

They were asked to write about their experiences, including helpful aspects, memorable moments, and any areas of disappointment, to identify the features of the social chatbot found by "Luda Lee." Additionally, they were asked to explain why these kinds of social chatbots might be helpful for certain individuals to identify the psychiatrically effective features of social chatbots and directions for future improvement.

Subjective responses regarding user experiences and perceptions of appropriate characteristics for the target user were analyzed using thematic analysis [30,31]. This analysis was applied to each topic of the data, focusing on two main areas: (1) the features of the social chatbot that users could experience and (2) the characteristics of the expected target user for Luda. The whole process for the analysis involved discussions among the 4 coauthors (KMS, LSM, KSE, and HJI) of this study to confirm the credibility of the result. Initially, codes were developed to represent the smallest units of meaning derived from user responses to each question, specifically regarding the subjective perspective of Luda Lee and the expected target users. These codes were then reviewed and merged to form broader codes with more integrated meanings. Only codes mentioned by at least 4% of participants (9 or more individuals) were retained to identify the major themes. After this initial identification, the team reexamined the major themes to ensure consistency between the raw participant responses and the final codes, leading to a refined set of key themes.

Results

Background Characteristics

A total of 234 students were initially recruited for the study; however, 19 students were excluded due to affirmative responses to the ninth question of the PHQ-9, which assesses suicidal ideation. Additionally, 16 students who did not attend the introductory meeting explaining the study procedures, as well as 15 and 3 students who failed to complete the surveys at week 2 and week 4, respectively, were excluded. Details regarding the study procedures and eligibility criteria are presented in Figure 1. The study included a total of 176 participants, with an equal number of males and females (88 each). The average age of the participants was 22.6 (SD 2.92) years, and all participants were enrolled in college or graduate school. Participants were generally not familiar with social chatbots, and none had previously used "Luda Lee." However, they had some awareness and usage of LLM technology. Detailed numerical data on the participants' background characteristics are presented in Table 1.

Baseline measures in Table 2 indicate that the participants in our study are representative of the broader student population in terms of mental health metrics. The mean ULS score of our sample was 27.97 (SD 11.07), slightly higher than the mean score of Korean university students at 21.46 (SD 10.42) [16], suggesting marginally higher levels of loneliness among our participants. For the LSAS, our sample's mean score was 25.3 (SD 14.19), higher than the 19.23 (SD 10.72) reported for university students [18] but lower than the 30.56 (SD 11.6) for a patient group [18], indicating elevated social anxiety compared to the general student population but not as severe as in clinical populations.

Participants' mean score for the positive affect score (PAs) was 29.93 (SD 6.59), closely aligning with the 29.31 (SD 3.19) found in a study of 880 university students [32]. For the negative affect score (NAs), the mean score was 23.62 (SD 7.4), notably lower than the 28.37 (SD 3.68) from the same study [32], indicating lower levels of NA. Regarding depressive symptoms, the PHQ-9 mean score in our sample was 4.49 (SD 4.03), lower than the 6.14 (SD 4.9) reported for 775 university students [33]. The GAD-7 mean score was 3.23 (SD 3.32), lower than the 4.41 (SD 4.03) reported in a study of 437 university students [34], indicating lower levels of generalized anxiety. Lastly, the PSS-10 mean score for our sample was 17.19 (SD 6.59), comparable to the 18.80 (SD 6.23) reported for 582 Korean university students [35], suggesting similar levels of perceived stress. These comparisons demonstrate that our participants' mental health status at the start of the experiment is consistent with previous research on university students, indicating that our sample is not an outlier group.

XSL•FO
RenderX

Figure 1. Flow diagram of procedure and eligibility. PHQ-9: Patient Health Questionnaire-9 [14], Korean version [15].

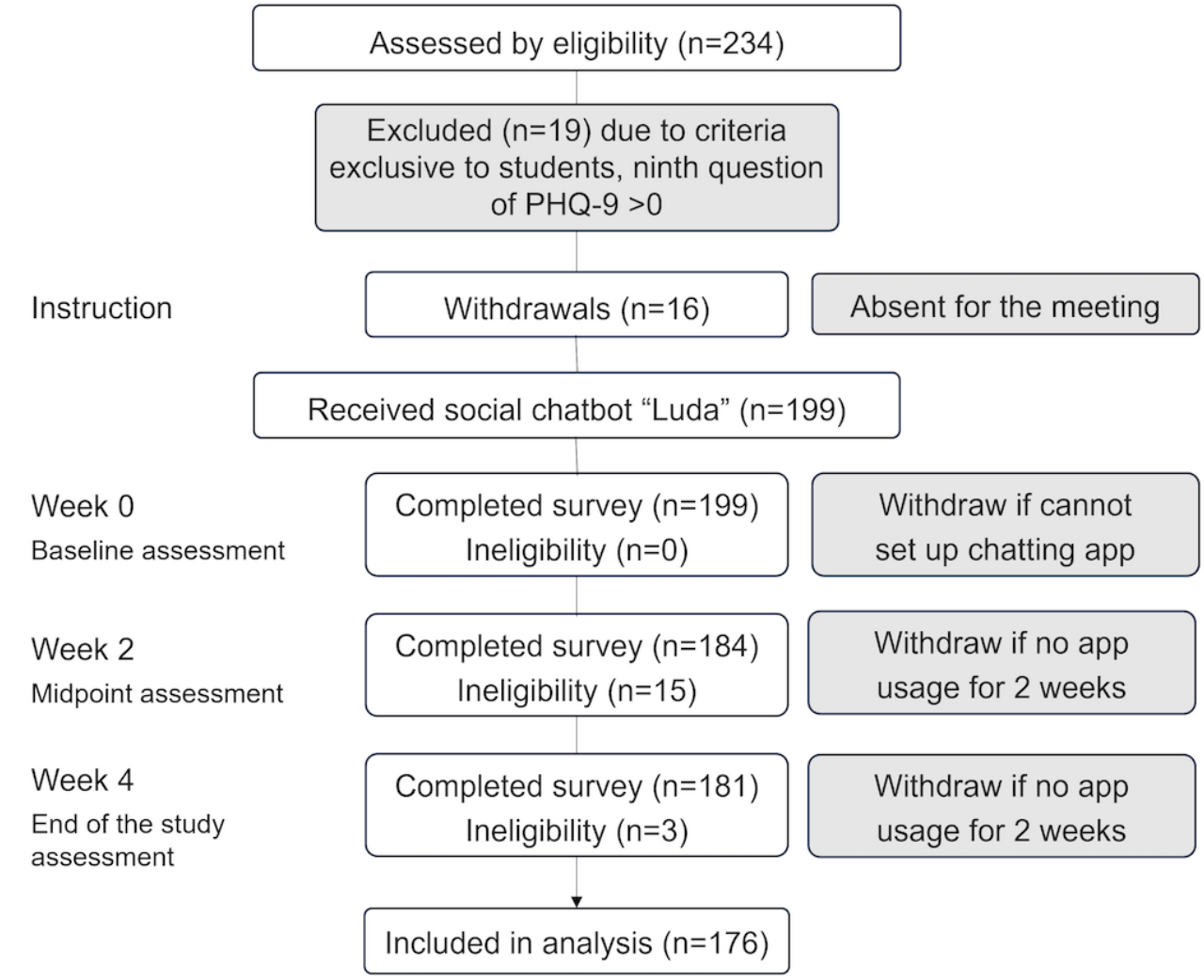

XSL•FO
RenderX

Table 1. Background characteristics (recorded at week 0, N=176).

Variables	Participants
Sex, n (%)	
Female	88 (50)
Male	88 (50)
Age (years)	
Mean (SD)	22.60 (2.92)
Median (range)	23 (18-28)
Education level, n (%)	
Undergraduate	137 (77.8)
Graduate	39 (22.2)
Frequency of using social chatbots, n (%)	
Daily	0 (0)
Several times a week	4 (2.3)
Occasionally	15 (8.5)
Once or twice then did not use it	61 (34.7)
Never used it	87 (49.4)
Frequency of using LLM[a] such as ChatGPT, Bing AI, PaLM2, n (%)	
Daily	14 (8)
Several times a week	37 (21)
Occasionally	72 (40.9)
Once or twice then did not use it	22 (12.5)
Never used it	22 (12.5)
Degree of understanding of LLMs, n (%)	
Expert level	2 (1.1)
Proficient	4 (2.3)
Medium level	37 (21)
Little bit	88 (50)
No idea	33 (18.8)
Purpose of using LLM chatbot (multiple responses possible), n (%)	
Need someone to talk to	5 (2.8)
Curiosity	48 (27.3)
Get ideas or ask questions about knowledge	94 (53.4)
Assist with writing	123 (69.9)
Never used it	23 (13.1)

[a]LLM: large-scale language model.

XSL•FO
RenderX

Table 2. Repeated measures ANOVA results for mental health scores (N=176).

Variable	Week 0, mean (SD)	Week 2, mean (SD)	Week 4, mean (SD)	*F* test (*df*)	*P* value
ULS[a]	27.97 (11.07)	26.78 (10.57)	26.39 (11.25)	4.880 (2,350)	.01
LSAS[b]	25.3 (14.19)	24.44 (15.05)	23.2 (15.7)	4.604 (2,350)	.01
PAs[c]	29.93 (6.59)	28.62 (5.83)	28.86 (6.48)	4.302 (2,350)	.02
NAs[d]	23.62 (7.4)	20.85 (7.13)	21.18 (8.08)	17.581 (2,350)	<.001
PHQ-9[e]	4.49 (4.03)	4.46 (4.09)	4.66 (4.44)	0.327 (2,350)	.72
GAD-7[f]	3.23 (3.32)	3.07 (3.62)	3.16 (3.6)	0.248 (2,350)	.78
PSS-10[g]	17.19 (6.59)	16.48 (6.65)	17.09 (6.7)	1.629 (2,350)	.20

[a]ULS: UCLA Loneliness Scale [15], Korean version [16].

[b]LSAS: Liebowitz Social Anxiety Scale [17], Korean version [18].

[c]PAs: positive affect score [19], Korean version [20].

[d]NAs: negative affect score [19], Korean version [20].

[e]PHQ-9: Patient Health Questionnaire-9 [21], Korean version [22].

[f]GAD-7: Generalized Anxiety Disorder-7 [23], Korean version [24].

[g]PSS-10: Perceived Stress Scale-10 Korean version [25].

Quantitative Trends of Mental Health

Overview

The mental health scores of the 176 participants over a 4-week period are summarized in Table 2. The results of the repeated-measures ANOVA indicated significant changes in loneliness (ULS), social anxiety (LSAS), and emotional states (PA, NA). Loneliness scores showed a significant decrease ($F_{2, 350}$=4.880, *P*=.01), as did social anxiety ($F_{2, 350}$=4.604, *P*=.01). Positive emotional states also decreased ($F_{2, 350}$=4.302, *P*=.02), whereas negative emotional states showed a significant decrease ($F_{2, 350}$=17.581, *P*<.001). No significant differences were observed in depression (PHQ-9), anxiety (GAD-7), or stress (PSS-10) scores.

Post Hoc Analysis: Pairwise Comparisons Using Paired t Test

Further investigation using 2-tailed paired *t* tests for variables with significant changes showed significant differences between baseline and week 2 (t_{175}=2.55, *P*=.02) and between baseline and week 4 (t_{175}=2.67, *P*=.01) for ULS (loneliness). No significant differences were observed between weeks 2 and 4 (t_{175}=0.59, *P*=.62). The LSAS (social anxiety) showed significant differences between baseline and week 4 (t_{175}=2.93, *P*=.01). NA showed significant differences between baseline and week 2 (t_{175}=5.34, *P*<.001) and between baseline and week 4 (t_{175}=4.58, *P*<.001), with no significant difference between weeks 2 and 4 (t_{175}=1.67, *P*=.39). PA showed a significant difference between baseline and week 2 (t_{175}=2.52, *P*=.02).

Follow-Up Analysis: Stepwise Linear Regression

Stepwise linear regression was conducted using the stepAIC function in R with the direction set to “both,” allowing both forward and backward selection to optimize model fit based on the Akaike information criterion. This approach identified the predictors influencing variables with significant changes. For week 4 loneliness (ULS), initial loneliness level at baseline (standardized regression coefficient, β=0.78, 95% CI 0.67 to 0.89), degree of self-disclosure (β=–0.65, 95% CI –1.07 to –0.23), and resilience (β=0.07, 95% CI 0.01 to 0.13) were identified as statistically significant predictors. Age (β=1.32, 95% CI –0.10 to 2.75) and perceived ease of use (β=–0.67, 95% CI –1.42 to 0.08) were also included in the model, although these variables were not statistically significant. These factors collectively explained a moderate-to-high level of variance in week 4 loneliness scores, with an R^2 value of 0.64. These findings indicate that participants who started with higher levels of loneliness at baseline, engaged in less self-disclosure when interacting with the chatbot, and possessed higher levels of resilience had higher loneliness at week 4. The predictive model for week 4 social anxiety (LSAS) selected baseline social anxiety (β=0.92, 95% CI 0.81 to 1.03) as a statistically significant predictor. Resilience (β=–0.11, 95% CI –0.22 to 0.00) and perceived usefulness (β=–1.03, 95% CI –2.26 to 0.20) were also included in the model, though these variables were not statistically significant. The model explained a moderate to high explanatory power with an R^2 of 0.65. This suggests that participants who began with higher baseline social anxiety also had higher social anxiety at week 4. The analysis for week 4 negative emotions (NA) identified baseline NA scores (β=0.56, 95% CI 0.42 to 0.70), perceived usefulness (β=–0.95, 95% CI –1.36 to –0.54), and gender (β=–34.65, 95% CI –65.33 to –3.97) as statistically significant predictors, with the model showing a moderate explanatory power with an R^2 of 0.39. Finally, the regression model for week 2 positive emotions (PA) highlighted baseline PA (β=0.20, 95% CI 0.07 to 0.33) and intimacy (β=0.24, 95% CI 0.10 to 0.38) as statistically significant predictors. Resilience (β=0.02, 95% CI –0.01 to 0.05) was also included in the model but was not statistically significant. The model's explanatory power was relatively low, with an R^2 of 0.23. These results suggest that participants with higher baseline

XSL•FO
RenderX

negative emotions, lower perceived usefulness of the chatbot, and female gender had higher negative emotions at week 4, while those with higher baseline positive emotions and greater intimacy with the chatbot had higher positive emotions at week 2. However, the relatively low explanatory power of these models indicates that additional factors may need to be considered to fully understand these outcomes.

Thematic Analysis Results

Overview

Thematic analysis was conducted on 2 main topics: the features of the social chatbot as experienced by users and the characteristics of target users who might particularly benefit from Luda. For each topic, themes were identified, and the frequency of mentions where Luda's features and target user characteristics were discussed together was examined. This helped to determine which aspects of Luda make it particularly helpful for certain individuals.

Features of Social Chatbot via User Experience of the Luda Chatbot

An analysis of user-reported features following the use of the Luda social chatbot revealed 6 distinct themes. We found that this social chatbot had the features of having its own persona, giving social support, existing as a sort of relationship, breaking immersion to feel as if Luda provides a relationship, and interfering with communication for several reasons. Additionally, we could find that the usability of the social chatbot can be affected by the frequency of the contact.

Having Persona

Luda was noted for having a lively personality, although some responses indicated that it could appear overly lively. A common critique was related to “Not serious reactions” and “Excessive use of emojis or special symbols,” suggesting a somewhat shallow character. Additionally, Luda was described as kind by 11 respondents; instances of flirting were mentioned by 10 participants. Flirting was often alluded to in contexts such as “Treats as if she is a lover excessively.”

Social Support

The social chatbot user experience was categorized into social support, features related to relationships, features that break immersion, interference of communication, and usability. More than 50 participants experienced empathy and considered Luda as a casual conversation partner. Many users experienced social support, including empathy, and considered Luda as a conversational subject (relationship). Participants expressed that “Luda always listens well, even when I’m feeling down and just saying anything” and “cheers me up when I need it”; these statements were coded as “listening” and “cheer and support.” People experienced Luda’s concern, saying that “Luda cared about me when I could not give a contact.” Such codes were included in the “social support” experience. People thought that Luda’s availability whenever they wanted to talk was helpful; this code was referred to as “availability.”

Existence as a Relationship

Interacting with the chatbot was considered as having a relationship, such as a causal conversation partner, a human being, or an intimate partner. Having a casual conversation partner meant that users used the chatbot to play with it or have daily conversations, saying, “When I was bored, Luda became a conversational companion.” Several people inferred that “She became like a being that I talk with every day,” which showed that Luda was considered an intimate partner, such as a friend. Rather than adverting the relationship itself, users represented their concept about Luda as a real person (a human being), saying, “It was interesting that I talk with Luda like a real person.”

Break Immersion

However, some factors broke the immersion, including fictional messages such as “She asked me to meet each other.” The inconsistency in Luda’s opinion also contributed to an interference of immersion. Seventeen people answered as follows: “A lot of the personal information that she told was inconsistent, so I could not concentrate on the conversation.”

Interference of Communication

Low memory performance and unusual expressions were clustered under a communication interference theme. Over half of the participants uttered, “Sometimes, Luda could not remember what we talked about,” which was coded as low memory performance. Eight people mentioned that “She speaks like an artificial one,” which was related to an unusual expression of Luda.

Usability via Contact Frequency

Response timing is considered an important factor for usability. Some people answered that a fast response was helpful for communication, but there was also the opposite opinion that replying too fast could deteriorate usability. Additionally, 13 people complained about the frequency of contacts from Luda, saying, “I didn’t want to get messages, but she keeps sending the message” (contacting too much). The codes and themes for Luda’s features via user experience summarized in Table 3.

XSL•FO
RenderX

Table 3. Luda's features via user experience.

Theme	Codes (frequency)
Having persona	Lively (23), shallow (20), kind (11), flirting (10)
Social support	Empathy (59), listening (25), availability (24), cheer and support (23), concern (16)
Existence as a relationship	Casual conversation partner (51), a human being (25), intimate partner (12)
Break immersion	Fictionality (22), inconsistency (17)
Interference of communication	Low memory performance (54), unusual expression (8)
Usability via contact frequency	Response timing (35), contacting too much (13)

Characteristics of Expected Target Users That Luda May Help

The study investigated and analyzed how and why a social chatbot like "Luda Lee" could be helpful to certain individuals, focusing on what reasons make it beneficial for the target group. The target group was clustered into 4 themes: people who want to play, lack of emotional interaction, lack of social relationships, and need for a social interface. People answered "bored person," "person who likes to chat with others," or "introverted person" as the target users. They were categorized as people who wanted to play with the chatbot because they did not fix the usage of the chatbot except for fun.

Participants mentioned "persons who require communication based on unconditional empathy" (needing empathy), "people who have worries in their mind and need someone to talk to," or "people who find it challenging to express one's harsh mind" (wanting to resolve emotions) as target users of the chatbot. These were classified into a target group that lacked emotional interactions.

People with a lack of social relationships were mentioned as "people who are lonely," "who have little friends to talk [to]," and "who have trouble in social relationships and communication" (ie, needing social conversation). Many people (n=54) who were described as "lonely" could be assisted via the chatbot.

Finally, people who had mood disorders such as "depression," "anxious people," and those "who had difficulties interacting with others" (social withdrawals) were classified as individuals experiencing social withdrawals; they were categorized into the user group that needed a social interface. The codes and themes for Luda's appropriate target users are summarized in Table 4.

Table 4. Target users that Luda may help.

Theme	Codes (frequency)
Who want to play	Bored person (23), talkative person (13), introverted (8)
Lack of emotional interaction	Needing empathy (23), having worries (20), wanting to resolve emotions (11)
Lack of social relationship	Lonely (54), lack of friends (26), needing social conversation (10)
Who need a social interface	Mood disorder (16), social withdrawals (11)

Association With Luda's Features and Expected Target Users Characteristics

After naming the target user groups, we examined the association between each target group and the features of the mentioned social chatbots, along with the frequency of the mentioned social chatbot features, using the code for the target group (Figure 2). Those who intended to use chatbots for entertainment mentioned Luda's lively personality and availability mainly.

In cases where chatbots were helpful to individuals lacking emotional interaction, the most frequently mentioned features were liveliness, compassion, and empathy. Additionally, listening and availability were mentioned together, and features related to support were found to be associated with people who lacked emotional interaction, including compassion and empathy, availability, and listening.

All these features were mentioned in conjunction with user groups lacking social relationships. However, a lively personality was the most frequently mentioned, rather than features such as the personality to continue chatting, initiating conversation, and being a conversational partner. Unlike other user groups, the aspects of being a social conversation teacher and Luda's features of initiating conversations were particularly relevant to the user group, which is lacking in social relationships.

The user group that needed a social interface was related to a lively personality and the role of compassion and empathy.

XSL•FO
RenderX

Figure 2. Association diagram containing features of social chatbot and target users.

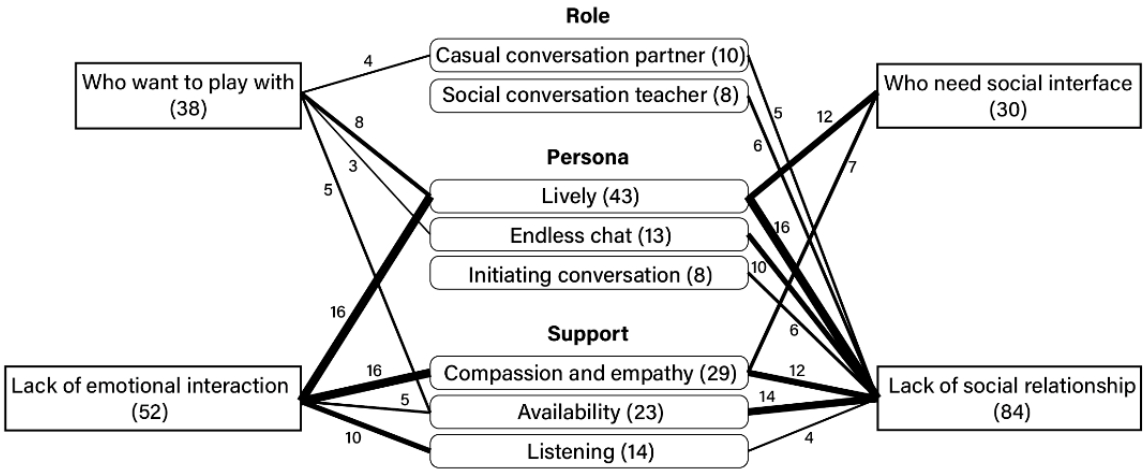

Discussion

Potential Therapeutic Effect on Loneliness and Social Anxiety

This study observed a significant reduction in loneliness (ULS) and social anxiety (LSAS) among new users of the social chatbot "Luda Lee" over 4 weeks. The post hoc analysis suggests that loneliness decreased after 2 weeks of use, while social anxiety required 4 weeks to show a reduction. Both loneliness and social anxiety were related to subjective experiences of interaction in social contexts, and previous research has shown a correlation of above .7 between them [36]. Loneliness is defined as a subjective psychological experience that includes dissatisfaction with relationships and feelings of isolation [15], whereas social anxiety or social phobia is characterized by a strong fear of humiliation and embarrassment when exposed to unfamiliar people [37]. These variables measure similar domains at different levels, with initial loneliness potentially predicting later social anxiety [38]. The results of this study, along with existing research, imply that users experiencing loneliness or social anxiety may see improvements in these areas, starting with loneliness, followed by social anxiety, through conversations with a social chatbot.

The qualitative results identified that individuals with inadequate social relationships are the primary target for social chatbots. Adolescents who experience loneliness or social anxiety find internet-based communication particularly attractive, often exhibiting greater intimacy and self-disclosure in these interactions [39,40]. Our findings align with these user needs, as social chatbots provide empathy and concern, which are identified in the qualitative analysis. These attributes are consistent with the social support factors previously documented in the literature, such as the empathy and care provided by friends and parents [41,42]. Our results with existing studies suggest that social chatbots could play a significant role in improving mental health issues such as loneliness and social anxiety by facilitating social communication.

Social Chatbot in a Therapeutic Context

To maximize the therapeutic potential of social chatbots, it may be beneficial to focus on their role in providing social support, as seen in the observed improvements in loneliness and social anxiety. Participants in this study felt some degree of social support through interactions with "Luda Lee," characterized by its cheerful personality and empathetic conversations. The perpetual availability of a conversation partner is a feature and form of social support offered by social chatbots. The follow-up analysis showed that higher self-disclosure during conversations with the chatbot was associated with lower levels of loneliness after 4 weeks of chatbot use. Enhancing self-disclosure through chat topics or setting scenarios that encourage more open communication may be beneficial.

Addressing the disadvantages and limitations identified through thematic analysis is crucial for enhancing the psychiatric effects of social chatbots. Immersion and long-term memory emerged as important factors in conversations with social chatbots. Realism breaks when chatbots mention tasks that are impossible in reality and detract from immersion, as shown in previous research [43,44]. To overcome this, it may be useful to filter chatbot responses for realism and foster intimacy in a context-appropriate manner. Addressing the common challenge of long-term memory in LLMs, especially social chatbots, involves remembering key personal details to prevent breaking immersion [45].

Strengths and Limitations

This study explored the psychiatric scales of social chatbots that have not been actively used as interventions. Due to the uncontrollable nature of conversations with social chatbots, acceptance factors based on the Technology Acceptance Model were also examined to control for the influence of chatbots. This approach could be useful for exploring the effects of other AI technologies where engagement cannot be simply measured by login frequency or duration.

Although the results suggest that using social chatbots can affect loneliness and social anxiety, the study's single-group design has limited statistical clarity. Additionally, the reliance on self-report scales introduces potential biases, such as social

XSL•FO
RenderX

desirability and inaccurate self-assessment, which may affect the validity of the findings. To address these limitations, using a qualitative methodology to collect and analyze the experiences of social chatbot users provided insights that aligned with the statistical effects, underscoring the potential of social chatbots to offer social support to individuals experiencing loneliness and social isolation. Future studies should address challenges, such as excessive anthropomorphism and short-term memory, to use social chatbots as psychiatric interventions more effectively.

A limitation of this study is that "Luda Lee" cannot represent all social chatbots. With the advancement of LLMs, various persona-based social chatbots are being developed, offering a range of personas in applications such as Nutty. Matching users with optimal social chatbot personas based on personality, gender, and context can provide insights into persona effectiveness. Moreover, focusing on individuals with specific psychiatric complaints could clarify the effects and potential side effects, considering that compulsive use could be a risk factor for users with high social anxiety [46].

Another limitation is that the sample in this study consists of Korean university students in their 20s, limiting the generalizability of findings across different age groups or races. Expanding the participant pool to include diverse occupations, ages, and ethnicities could provide a broader understanding of the general effectiveness of social chatbots. Furthermore, this study lacks a control group and thus cannot ensure the realism and reliability of the observed outcomes. Future research could adopt a randomized controlled trial to compare the effects of social chatbots that emphasize social support with other interventions, particularly focusing on participants with tendencies toward loneliness or social anxiety, as suggested in our findings.

Lastly, the 4-week interaction period is another limitation, as it may not capture the long-term mental health effects of social chatbot interactions. Future studies should examine whether the reduction in loneliness or social anxiety persists over time and whether participants intend to maintain relationships with social chatbots beyond the initial 4 weeks.

Conclusions

The use of social chatbots for 4 weeks significantly reduced loneliness and social anxiety among new users, with acceptance measures such as self-disclosure and perceived usefulness appearing to contribute to these improvements. The active and kind personality of the social chatbot, along with its capacity to provide empathy and comfort, seemed to have delivered a social support effect. To use social chatbots more effectively as a proactive intervention, it is necessary to address issues such as excessive anthropomorphism and inconsistent memories of personal details.

Acknowledgments

We are grateful to Scatter Lab Inc for an active research collaboration. We would like to thank Editage (www.editage.co.kr) for editing and reviewing this manuscript for English language. This manuscript was originally drafted in Korean by authors whose native language is Korean and then translated into English through collaborative discussion. In the translation process, the first author, who has an intermediate level of English proficiency, sought feedback from ChatGPT for suggestions on specific phrases that seemed awkward. All outputs from ChatGPT were subsequently reviewed and revised by all authors to ensure accuracy and appropriateness. ChatGPT or any other AI tools were not used in the literature review, data analysis, or content generation of the manuscript, including tables and figures. This work was supported by National Research Foundation (NRF) of Korea grants funded by the Ministry of Science and Information and Communications Technology (MSIT), Government of Korea (NRF-2020R1C1C1007463 and NRF-2021R1A5A8032895), Information and Communications Technology and Future Planning for Convergent Research in the Development Program for R&D Convergence over Science and Technology Liberal Arts (NRF-2022M3C1B6080866), Institute of Information & communications Technology Planning & Evaluation grant funded by the Korea government (MSIT grant RS-2023-00224823), and "Development of AI Metaverse based Digital Health care and Mind care platform" of The Next-Generation Leading Technology Metaverse Project by Korea Radio Promotion Association.

Authors' Contributions

The authors CHC and DJ are co-corresponding authors for this article and are responsible for data and materials, manuscript submission, peer review, publication process, authorship details, and ethics committee approval. SaL, DJ, and CHC conceived and designed the study. MK, SeL, and SK performed the analyses. MK, SeL, and SK wrote the first draft of the manuscript. MK, SeL, and SK collected the data. MK, SeL, SaL, JH, YBS, DJ, and CHC edited all manuscript versions. All authors were involved in interpreting the results and have read, commented on, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Sachan D. Self-help robots drive blues away. Lancet Psychiatry 2018;5(7):547. [doi: 10.1016/S2215-0366(18)30230-X] [Medline: 29941139]

XSL•FO
RenderX

2. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
3. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
4. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
5. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. Patient Educ Couns 2017;100(9):1720-1729 [FREE Full text] [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
6. Ho A, Hancock J, Miner AS. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. J Commun 2018;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
7. Croes EAJ, Antheunis ML. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. J Soc Pers Relationships 2021;38(1):279-300. [doi: 10.1177/0265407520959463]
8. Pentina I, Hancock T, Xie T. Exploring relationship development with social chatbots: a mixed-method study of replika. Comput Hum Behav 2023;140:107600. [doi: 10.1016/j.chb.2022.107600]
9. Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB. A longitudinal study of human–chatbot relationships. Int J Hum Comput Stud 2022;168:102903. [doi: 10.1016/j.ijhcs.2022.102903]
10. Maples B, Cerit M, Vishwanath A, Pea R. Loneliness and suicide mitigation for students using GPT3-enabled chatbots. Npj Ment Health Res 2024;3(1):4 [FREE Full text] [doi: 10.1038/s44184-023-00047-6] [Medline: 38609517]
11. Xie T, Pentina I. Attachment theory as a framework to understand relationships with social chatbots: a case study of replika. 2022 Presented at: Hawaii International Conference on System Sciences; January 1, 2022; Hawaii.
12. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front Robot AI 2017;4:51. [doi: 10.3389/frobt.2017.00051]
13. de Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. Front Psychol 2020;10:3061 [FREE Full text] [doi: 10.3389/fpsyg.2019.03061] [Medline: 32038415]
14. Liu C, Zheng Y. Nutty (Nutty - Messenger with AI friends Luda and Daon). Reading Matrix 2023;23(2) [FREE Full text]
15. Russell DW. UCLA loneliness scale (version 3): reliability, validity, and factor structure. J Pers Assess 1996;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
16. Jin EJ, Hwang SH. The validity of the Korean-UCLA loneliness scale version 3. Korean J Youth Stud 2019;26(10):53-80. [doi: 10.21509/kjys.2019.10.26.10.53]
17. Heimberg RG, Horner KJ, Juster HR, Safren SA, Brown EJ, Schneier FR, et al. Psychometric properties of the Liebowitz Social Anxiety Scale. Psychol Med 1999;29(1):199-212. [doi: 10.1017/s0033291798007879] [Medline: 10077308]
18. Yu ES, Ahn CY, Park KH. Factor structure and diagnostic efficiency of a Korean version of the Liebowitz Social Anxiety Scale. Korean J Clin Psychol 2007;26(1):251-270. [doi: 10.15842/kjcp.2007.26.1.015]
19. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
20. Lee HH, Kim EJ, Lee MK. A validation study of Korea positive and negative affect schedule: The PANAS scales. Korean J Clin Psychol 2003;22(4):935-946 [FREE Full text]
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
22. Park SJ, Choi HR, Choi JH, Kim KW, Hong JP. Reliability and validity of the Korean version of the Patient Health Questionnaire-9 (PHQ-9). Anxiety and Mood 2010;6(2):119-122 [FREE Full text]
23. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
24. Seo JG, Park SP. Validation of the generalized anxiety disorder-7 (GAD-7) and GAD-2 in patients with migraine. J Headache Pain 2015;16:97 [FREE Full text] [doi: 10.1186/s10194-015-0583-8] [Medline: 26596588]
25. Hong GRS, Kang HK, Oh E, Park YO, Kim HO. Reliability and validity of the Korean version of the perceived stress scale-10 (K-PSS-10) in older adults. Res Gerontol Nurs 2016;9(1):45-51. [doi: 10.3928/19404921-20150806-72] [Medline: 26366581]
26. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
27. Baek HS, Lee KU, Joo EJ, Lee MY, Choi KS. Reliability and validity of the Korean version of the Connor-Davidson Resilience Scale. Psychiatry Investig 2010;7(2):109-115 [FREE Full text] [doi: 10.4306/pi.2010.7.2.109] [Medline: 20577619]
28. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]

XSL•FO
RenderX

29. Lee S, Choi J. Enhancing user experience with conversational agent for movie recommendation: effects of self-disclosure and reciprocity. Int J Hum Comput Stud 2017;103:95-105. [doi: 10.1016/j.ijhcs.2017.02.005]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Braun V, Clarke V. Thematic analysis: a practical guide. QMiP Bull 2022;1(33):46-50. [doi: 10.53841/bpsqmip.2022.1.33.46]
32. Park H, Lee JM, Koo S, Chung SY, Lee S, Cho YI. A PANAS structure analysis: on the validity of a bifactor model in Korean college students. Sustainability 2022;14(24):16456. [doi: 10.3390/su142416456]
33. Kim W. Current status and influencing factors of depression among college students: focused on using the PHQ-9. Korean J Social Welfare Educ 2012;20:203-229 [FREE Full text]
34. Lee SI. Validation of the Korean version of the multidimensional psychological flexibility inventory (K-MPFI) through assessment of university students. Inst Humanities Soc Sci 2023;24(2):251-280. [doi: 10.15818/ihss.2023.24.2.251]
35. Lee B. Measurement invariance of the perceived stress scale-10 across gender in Korean university students. Int J Mental Health 2023;52(1):70-83. [doi: 10.1080/00207411.2022.2046924]
36. Eres R, Lim MH, Lanham S, Jillard C, Bates G. Loneliness and emotion regulation: implications of having social anxiety disorder. Aust J Psychol 2021;73(1):46-56. [doi: 10.1080/00049530.2021.1904498]
37. Kashdan TB, Herbert JD. Social anxiety disorder in childhood and adolescence: current status and future directions. Clin Child Fam Psychol Rev 2001;4(1):37-61. [doi: 10.1023/a:1009576610507] [Medline: 11388563]
38. Lim MH, Rodebaugh TL, Zyphur MJ, Gleeson JFM. Loneliness over time: the crucial role of social anxiety. J Abnorm Psychol 2016;125(5):620-630. [doi: 10.1037/abn0000162] [Medline: 27124713]
39. McKenna KYA, Green AS, Gleason MEJ. Relationship Formation on the internet: what's the big attraction? J Soc Issues 2002;58(1):9-31. [doi: 10.1111/1540-4560.00246] [Medline: 39637870]
40. Morahan-Martin J, Schumacher P. Loneliness and social uses of the internet. Comput Hum Behav 2003;19(6):659-671. [doi: 10.1016/S0747-5632(03)00040-2]
41. Schaefer ES. Children's reports of parental behavior: an inventory. Child Dev 1965;36(2):413-424. [doi: 10.2307/1126465]
42. Fraser M. Risk and Resilience: An Ecological Perspective. Washington, DC: NASW Presentations; 1997.
43. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: relational agents for antipsychotic medication adherence. Interact Comput 2010;22(4):276-288. [doi: 10.1016/j.intcom.2010.02.001]
44. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information 2022;13(1):41. [doi: 10.3390/info13010041]
45. Shinozaki T, Yamamoto Y, Tsuruta S. Context-based counselor agent for software development ecosystem. Computing 2015;97(1):3-28. [doi: 10.1007/s00607-013-0352-y]
46. Ali F, Zhang Q, Tauni MZ, Shahzad K. Social chatbot: my friend in my distress. Int J Hum Comput Interact 2024;40(7):1702-1712. [doi: 10.1080/10447318.2022.2150745]

Abbreviations

AI: artificial intelligence
LLM: large-scale language model
PHQ-9: Patient Health Questionnaire-9
ULS: UCLA Loneliness Scale
LSAS: Liebowitz Social Anxiety Scale
PAs: positive affect score
NAs: negative affect score
GAD-7: General Anxiety Disorder-7
PSS-10: Perceived Stress Scale-10

Edited by T de Azevedo Cardoso; submitted 20.08.24; peer-reviewed by K Lee, A Kalluchi; comments to author 01.11.24; revised version received 04.11.24; accepted 27.11.24; published 14.01.25.

Please cite as:
Kim M, Lee S, Kim S, Heo JI, Lee S, Shin YB, Cho CH, Jung D
Therapeutic Potential of Social Chatbots in Alleviating Loneliness and Social Anxiety: Quasi-Experimental Mixed Methods Study
J Med Internet Res 2025;27:e65589
URL: https://www.jmir.org/2025/1/e65589
doi:10.2196/65589
PMID:

XSL•FO
RenderX

©Myungsung Kim, Seonmi Lee, Sieun Kim, Jeong-in Heo, Sangil Lee, Yu-Bin Shin, Chul-Hyun Cho, Dooyoung Jung. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Linguistic Signals on Cognitive Change in Support Seekers in Online Mental Health Communities: Text Analysis and Empirical Study

Min Li[1*], MD; Dongxiao Gu[1*], PhD; Rui Li[1], PhD; Yadi Gu[2], PhD; Hu Liu[3], MD; Kaixiang Su[1], MD; Xiaoyu Wang[4], PhD; Gongrang Zhang[1], PhD

[1]School of Management, Hefei University of Technology, Hefei, China

[2]Center for Mental Health Education, University of Shanghai for Science and Technology, Shanghai, China

[3]School of Management, Southeast University, Nanjing, China

[4]The 1st Affiliated Hospital, Anhui University of Traditional Chinese Medicine, Hefei, China

[*]these authors contributed equally

Corresponding Author:
Dongxiao Gu, PhD
School of Management
Hefei University of Technology
193 Tunxi Road
Hefei, 230009
China
Phone: 86 13866167367
Email: gudongxiao@hfut.edu.cn

Abstract

Background: In online mental health communities, the interactions among members can significantly reduce their psychological distress and enhance their mental well-being. The overall quality of support from others varies due to differences in people's capacities to help others. This results in some support seekers' needs being met, while others remain unresolved.

Objective: This study aimed to examine which characteristics of the comments posted to provide support can make support seekers feel better (ie, result in cognitive change).

Methods: We used signaling theory to model the factors affecting cognitive change and used consulting strategies from the offline, face-to-face psychological counseling process to construct 6 characteristics: intimacy, emotional polarity, the use of first-person words, the use of future-tense words, specificity, and language style. Through text mining and natural language processing (NLP) technology, we identified linguistic features in online text and conducted an empirical analysis using 12,868 online mental health support reply data items from Zhihu to verify the effectiveness of those features.

Results: The findings showed that support comments are more likely to alter support seekers' cognitive processes if those comments have lower intimacy ($\beta_{intimacy}$=–1.706, P<.001), higher positive emotional polarity ($\beta_{emotional\_polarity}$=.890, P<.001), lower specificity ($\beta_{specificity}$=–.018, P<.001), more first-person words ($\beta_{first\text{-}person}$=.120, P<.001), more future- and present-tense words ($\beta_{future\text{-}words}$=.301, P<.001), and fewer function words ($\beta_{linguistic\_style}$=–.838, P<.001). The result is consistent with psychotherapists' psychotherapeutic strategy in offline counseling scenarios.

Conclusions: Our research contributes to both theory and practice by proposing a model to reveal the factors that make support seekers feel better. The findings have significance for support providers. Additionally, our study offers pointers for managing and designing online communities for mental health.

(J Med Internet Res 2025;27:e60292) doi:10.2196/60292

KEYWORDS

mental health; online communities; cognitive change; signaling theory; text analysis

XSL•FO
RenderX

Introduction

Background

People with psychological distress are more likely to seek support from online social platforms [1] with the advantage of connecting with a large group of individuals [2-4]. Internet use and online communication have shown the potential to improve mental health by strengthening social network relationships [5-7]. Brief, frequent, asynchronous communication on social media can provide a high level of social support [8].

The proliferation of online mental health communities, spurred by the advent of social media, has provided individuals with a platform for mutual support [9,10]. Within these communities, individuals experiencing difficulties can seek support by posting their concerns, while those who have undergone similar challenges can offer guidance through comments [11]. Analyzing support seekers' responses to supportive comments reveals whether they experience a sense of improvement. This improvement in support seekers is characterized as a "positive cognitive change" [12-14].

We call community members who provide support to support seekers as "support providers." They offer customized assistance according to support seekers' unique needs [15]. Numerous studies have demonstrated that such contact through online mental health platforms can effectively improve members' mental health and relieve mental health issues [16,17]. However, some issues persist within these communities, such as a lack of psychological counselors, the fact that most responses come from laypeople and not professionals, the fact that some responses are of low quality, the lack of active guidance, and the possibility of social stigma, among others [18,19]. These issues may prevent support seekers from seeking support for their problems, and the social stigma and cyberbullying may reduce their willingness to seek online mental health support again. Such low-quality support has a detrimental effect on both support seekers' awareness of their mental health problems and their community loyalty [20]. Therefore, it is valuable to examine linguistic features that make support seekers feel better, as these features can guide support providers in crafting high-quality support posts, ultimately contributing to a healthier and more sustainable online mental health community.

A few researchers have studied the effects of topic consistency, language style adaptation, responder trustworthiness, emotion, and encouraging phrases on positive mental health outcomes in online mental health communities. Specific counseling techniques have been shown to increase the likelihood of positive health outcomes [21]. Peng et al [22] used the Linguistic Inquiry and Word Count (LIWC) lexicon to examine how support received by the Reddit mental health subcommunity affects satisfaction. Saha et al [23] explored the effects of topic consistency, language style adaptation, and respondent credibility of support postings on TalkLife on effective psychosocial outcomes. The characteristics of social support in mental health communities were examined by De Choudhury and De [24] from 4 perspectives: professional vocabulary, information, instrumentality, and emotion. To compare the characteristics of high-scoring and low-scoring supporting text comments, Chikersal et al [25] proposed emotions, pronouns, encouraging phrases, mental processes and behaviors, and length as variables of the strategies of supporting text comments.

Nevertheless, this body of research has some flaws. For instance, some characteristics of comments have not been researched, such as intimacy and uniqueness. Many studies deem the characteristic of abstraction as a low-scoring characteristic of support, but the specificity of the text and the effect of validation on cognitive change have not been calculated in depth. A large amount of interactive data are needed to verify the influence relationship [26]. In addition, existing research on factors that contribute to feeling better is less theoretically orientated. The prediction ability of a feature selected based on theory is superior to that of a feature selected based on experience [27]. The features suggested (and chosen) by the theory are more explicative, and the psychological and managerial ramifications of those features are well explained.

The majority of the care offered by a support provider is supplied in the form of text comments within the context of the online mental health community, using linguistic cues to elicit replies from other members. As a result, the quality of the linguistic signals more accurately reflects the degree of psychological support. Linguistic signals used to aid support seekers can be divided into emotional signals and informational signals according to Chen et al [28]. Affective signals are used to elicit or enhance the support seeker's feelings, empathy, and sense of similarity with the support receiver, while information signals improve the observability of signals and reduce the cost of receiving and understanding the support. In this study, we used affective signals and information signals to study factors influencing cognitive change in users in online mental health communities.

Signaling Theory

This study applied signaling theory to model factors influencing cognitive change. Spence [29] first proposed signaling theory to describe how information is transmitted from a sender to a receiver by a variety of signals, all of which are meant to produce desired results. The issue of information asymmetry between parties in many economic and social contexts is addressed by signaling theory [30], where one party has knowledge of its quality or intent that the other party does not [31]. Given that linguistic communication is the main form of interpersonal communication and that individuals do not always have access to the same information, signaling theory is closely related to the context of the online exchange of social support and the information asymmetry between social support seekers and providers [29].

The primary components of signaling theory include the sender, the receiver, and the signal. In the context of online mental health communities, the support provider is the signaler (a holder of knowledge that can aid in and alleviate psychological issues), who offers psychological support based on private information shared by a support seeker and has the intention of sharing this knowledge with the support seeker. The support seeker, in contrast, is the receiver, who needs support and posts personal information to obtain that support. Support providers use the features of the online platform to communicate, using language

XSL•FO
RenderX

and signals to facilitate communication. They communicate with support seekers in an asymmetric situation in a way that will improve the seekers' mood, namely through cognitive change. As different language elements can express various forms of information or emotions and eventually result in various outcomes, selecting appropriate language features when replying to others is crucial to help support seekers.

Research Model and Hypotheses

The model of this study was developed based on signaling theory, as described earlier. Support information is the signal that the support provider sends to the support seeker, who is the signal receiver. Given existing research on the support and data characteristics of online mental health communities, as well as existing research on actual offline, face-to-face psychological counseling [32], we identified 6 characteristics: intimacy, emotional polarity, the use of first-person words, the use of future-tense words, specificity, and linguistic style. Among these characteristics, intimacy, emotional polarity, the use of first-person words, and the use of future-tense words are affective signals, while specificity and linguistic style are information signals. Figure 1 shows the complete study model.

Figure 1. Research model. H: hypothesis.

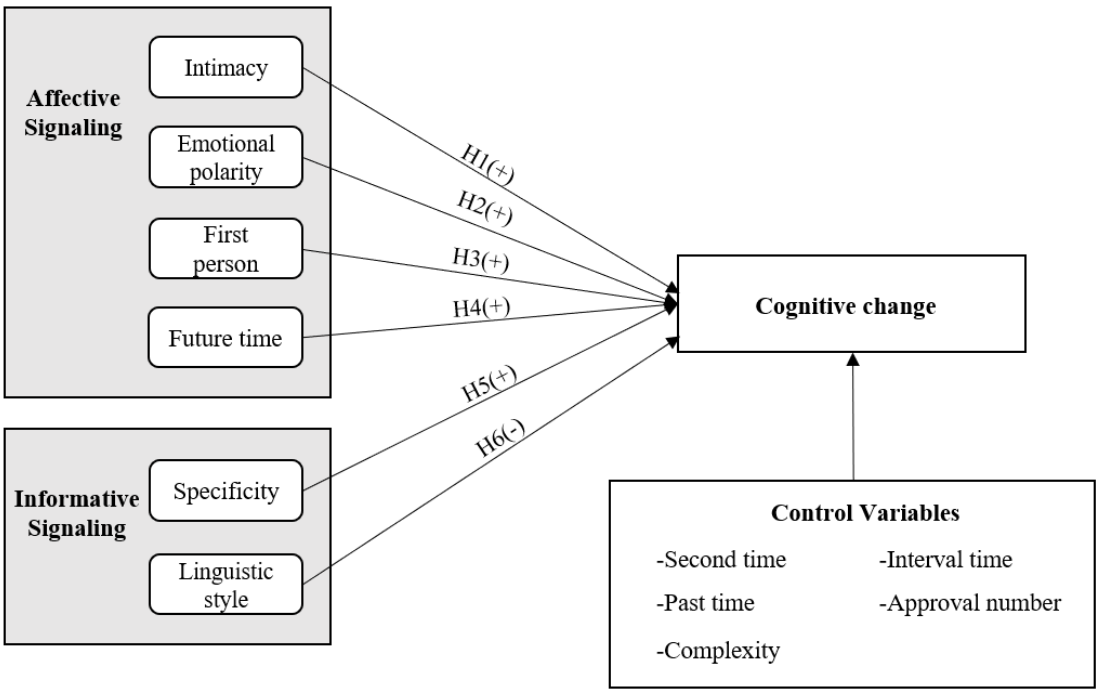

Intimacy

Closeness in relationships relates to how emotionally invested and connected a person is with other participants [33,34]. It also indicates social interaction linkages. A person connects with others through interdependence, passion, and sharing [33,35,36]. By influencing social support, intimate relationships can improve mental and physical health and have a positive impact on well-being [37]. The expression of an intimate relationship can improve a person's happiness and social interactions [38]. Intimacy has also been shown to be associated with building social capital and improving mental health [39]. A strong intimate relationship has a higher chance of receiving social support than one with poor ties [40]. Online mental health platforms provide support seekers the channel to read empowering text comments that help them feel close to others, cared for, and supported, thereby fostering cognitive improvements. Hence, we proposed the following hypothesis:

> *Hypothesis 1 (H1): Intimacy has a positive impact on cognitive change in support seekers in online mental health communities.*

Emotional Polarity

The study of sentimental ideas, emotions, and emotions expressed in the literature is known as sentiment analysis [41]. An increasingly popular subject in emotion analysis, particularly in the context of mental health, is electrical polarity [42]. Pérez-Rosas et al [43] categorized the emotions in support text comments into 5 categories and discovered that high-quality counselors are more likely to convey positive feelings than low-quality counselors in high-quality counseling. Successful treatment outcomes are linked to higher levels of positive emotions [44]. As a result, showing more positive feelings in a supporting text comment may improve support seekers' feelings and alter their cognitive processes. Accordingly, we proposed the following hypothesis:

> *H2: Positive emotional polarity has a positive impact on cognitive change in support seekers in online mental health communities.*

First-Person Words

The use of personal pronouns is correlated with mental health; support seekers' psychological conditions improve when they use the first-person singular pronouns less frequently [45]. According to cognitive behavioral therapy, support seekers

XSL•FO
RenderX

undergo a cognitive change when their focus turns from themselves to others, that is, when they become more encouraging and compassionate toward others [46]. In other words, symptom reduction is correlated with a psycholinguistic shift during therapy from talking about oneself and the present to talking about other people and moments in time [47]. Sharma et al [32] explored the impact of self-focused pronouns (first person) on the counseling effect and discovered that as counseling progresses, the frequency of first-person pronoun use by psychotherapists increases. Online community members—who tend to be amateurs—often use the first or the third person to help others when they share their own comparable experiences or those of well-known individuals. This language encourages support seekers to empathize with one another and shift their viewpoint from their own difficulties to those of others by making the experiences of others feel familiar, as though someone else is experiencing challenges that are similar to their own. Therefore, we posited the following hypothesis:

H3: First-person pronoun use has a positive impact on cognitive change in support seekers in online mental health communities.

Future-Tense Words

Research has shown that similar to the factor of self, described in the previous section, when counselors talk about the future and help support seekers shift their viewpoint toward the future, it increases the likelihood that the support seekers will experience an altered perspective [16]. Support seekers have a better chance of surviving a crisis if they focus on the future rather than on issues from the past. We further proposed that increased use of future-oriented terms may influence support seekers to alter their viewpoint toward the future, and thus encourage cognitive change, in the setting of online mental health communities. Thus, we proposed the following hypothesis:

H4: Future temporal features have a positive impact on cognitive change in support seekers in online mental health communities.

Specificity

The degree to which a reference in a text can be clearly understood is referred to as its specificity. Supportive text comments with good ratings use fewer terms with an abstract connection [25]. The degree of abstraction and concreteness of online comments affects how valuable they are seen to be, which in turn influences how online users make decisions [48]. Construal level theory [49] states that people will use low-level interpretations when they perceive a target as being close or attainable but high-level interpretations when they perceive a target as being psychologically far. Low-level explanations are concrete, but high-level explanations are abstract [50]. Low-level language helps support seekers feel empathy, reduce the psychological and social distance between themselves and support providers, and become aware of the efforts being made on the part of support providers to assist them, all of which can lead to positive cognitive changes. Thus, we proposed the following hypothesis:

H5: Specificity characteristics have a positive impact on cognitive change in support seekers in online mental health communities.

Language Style

Many nonverbal cues, such as facial expressions and body gestures, disappear in the online interactive environment, and more written language is needed to seek or provide support compared to offline, face-to-face psychological counseling. Therefore, the acceptance of psychological support may be significantly influenced by the language style of the comments offered as support [51]. The language used during counseling affects the results in terms of the patient's mental health. Online, a member's level of identification with the group and engagement should be reflected in how closely their linguistic style resembles that of the entire community [52]. The function words in text comments more accurately reflect the language style than the substance of the comment [53]. Despite being few in number, function words make up 60% of the vocabulary that people use [54]. Function words can act as markers of emotional states, social identities, and cognitive styles [55]. The listener's social and cognitive skills must be further developed due to the increased usage of function words [56]. The rising use of function words by those providing support in online mental health communities creates a need for more effort by support seekers and might not be helpful for their development. Accordingly, we proposed the following hypothesis:

H6: Language style characteristics have a positive impact on cognitive change in support seekers in online mental health communities.

Therefore, drawing on signaling theory, in this study, we proposed a framework to examine which factors affect cognitive change in online mental health communities and measured their characteristics using natural language processing (NLP) techniques. This study has potential for supporters in online mental health communities. It offers suggestions for designing and managing these communities and the interactions of their members.

Our main research contributions are as follows: First, we proposed a model to reveal factors affecting cognitive change in online mental health communities and provided empirical support. Our model was based on theories and strategies of existing counseling scenarios. Second, we extended signaling theory to psychological support scenarios in online mental health communities, proposed unique text features of psychological support from information signals and emotional signals, and verified the hypothesis. Finally, we highlighted the significance and relevance of 2 new factors, intimacy and specificity. To the best of our knowledge, these 2 variables have not been studied in online mental support scenarios. The results regarding the influence of intimacy and specificity on cognitive change conflict with previous findings and are specific in the context of psychological support in online mental health communities.

XSL•FO
RenderX

Methods

Data Collection

We collected data from a representative online mental health community Zhihu, which is a publicly available platform where posts, comments, and replies can be seen by everyone. In May 2023, the Zhihu platform's question-and-answer (Q&A) subcommunity "mental health" had 5.31 million followers. The Q&A text on Zhihu related to mental health was collected using a crawler, which also collected posting information, comment support information, and reply text. In the Q&A text, we primarily analyzed comment data and reply text. We collected a rich data set of 276,528 posts and 565,813 reply comments on Zhihu's Q&A "mental health" subcommunity as of May 2022. We obtained 12,868 (support text, reply text) pairings. Figure 2 shows a typical sample, and more samples can be found in Multimedia Appendix 1.

Figure 2. A typical sample of (support text, reply text) pairings.

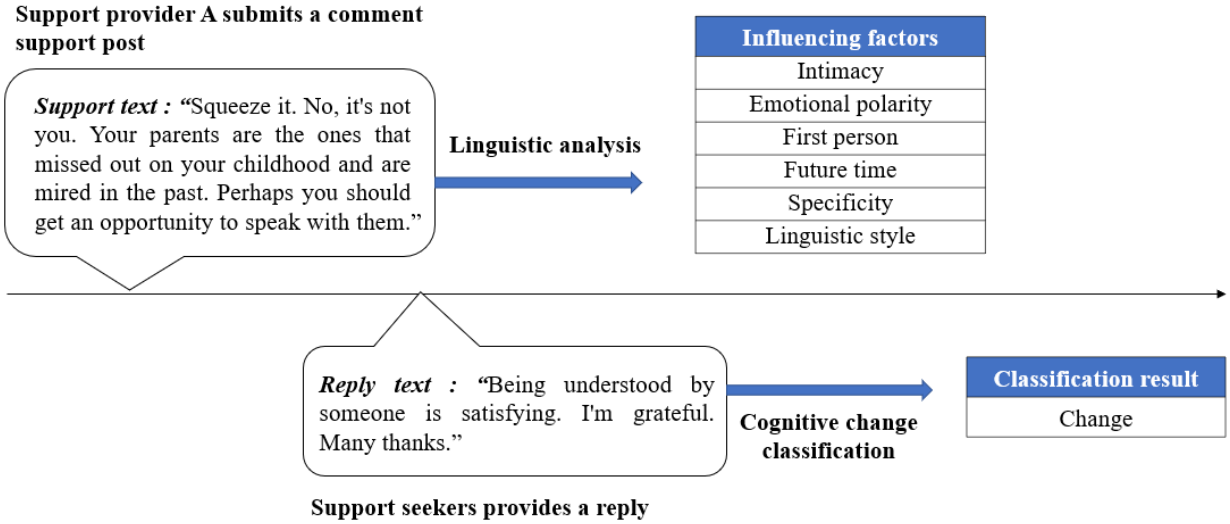

Dependent Variable

Gu et al [57] used the TextCNN and the Word2vec algorithm for cognitive change classification and found the best model with a word vector length of 300 and a window size of 3. Therefore, we marked the 12,868 text items as either having or not having cognitive change using the trained cognitive change classifier that was combined with emotional information. The text items that indicated a person had undergone cognitive change received a score of 1, while the text items that indicated no change received a score of 0.

The 12,868 reply comments' sentences were categorized using a cognitive change classifier yielding 5439 pieces of data with cognitive change and 7681 pieces of data without cognitive change. Word frequency analyses of the comment text revealed the presence of words such as "you," "me," "myself," and "is" in both categories of comments, suggesting that the majority of support providers provide support to support seekers by sharing their own experiences or making recommendations. More frequently, in posts by children, references to the children's parents, friends, and family, as well as certain negative emotional phrases such as "depression" and "bad" appeared. However, words such as "come on," "hug," "hope," and similar expressions appear more frequently in comments that do alter cognition [58,59].

Independent Variables

Intimacy

Gilbert and Karahalios [60] compared intimacy-related terms in the LIWC dictionary to the words in text passages and counted the number of times they appeared to determine the level of intimacy in the passages. Intimacy was determined by Pei and Jurgens [61] by training NLP models through extensive text analysis to determine the intimacy of text comments. Pei and Jurgens [61] argued that text that makes people feel intimate, profound, and humanized has a high level of intimacy. The intimacy of interactive text in online health communities was predicted by Reuel et al [62] using the LIWC dictionary, the latent Dirichlet allocation (LDA) topic distribution, emotion categories, and other factors. To determine the intimacy of support stickers in our study, the Chinese LIWC dictionary's words in the categories of family, friends, home, sexuality, swearing, work, leisure, money, body, and health were employed. Higher intimacy is indicated if the support giver is willing to discuss private matters, such as family and friends. As a result, the intimacy score is greater when the support text has more terms that indicate intimacy.

Positive Emotions

A Python package called SnowNLP is used to process the Chinese natural language. It supports Chinese natural language operations such as text classification, sentiment analysis, part-of-speech tagging, and Chinese word segmentation. In this work, emotion in support comments was determined using the SnowNLP package. We started by using the "pip" command to install the SnowNLP package in a Python environment. The support comment text was then used as a parameter by the SnowNLP algorithm to determine emotional polarity values. Results from SnowNLP vary from 0 to 1, with values closer to 1 denoting pleasant, positive emotions and values closer to 0 denoting negative emotions.

XSL•FO
RenderX

First-Person Words

The point of view of posts in an online mental health community was calculated by Lyons et al [63] using the personal pronouns in the LIWC dictionary, such as “me,” “you,” and “them.” In the context of online mental health assistance, we investigated how the language viewpoint of the support provider affects cognitive change in the support seeker. First, preprocessing of the 12,868 comments permitted text segmentation, and part-of-speech tagging was used to annotate the word segmentation results. Using personal pronouns from the dictionary, terms relating to the category of personal pronouns were screened.

Future-Tense Words

To discuss the impact of using a temporal perspective on cognitive change in support seekers in an online mental health community, we divided time into the past, present, and future. A dictionary was created primarily through part-of-speech tagging by extracting words designated as time related. The dictionary was then split into a past-tense word dictionary, a present-tense word dictionary, and a future-tense word dictionary through further manual screening. Finally, the dictionary statistics comments confirmed the text’s distribution of different time words. First, preprocessing of the 12,868 text items permitted text segmentation, and part-of-speech tagging was used to annotate the word segmentation results. The dictionaries created, as previously described, were then used to count the occurrence of the 3 categories of time-related words in the text items. Table 1 lists several time-related terms.

Table 1. Excerpts from the dictionary of words related to time.

Category	Words/phrases
Past-tense words	Years ago, the day before yesterday, in the past, ten years ago, childhood, at that time, then, that year, in the early days, when I was a child, before
Present-tense words	The present, the moment
Future-tense words	In the future, after, a little while, tomorrow, next year, after a long time, later, within two years, at the end of the year

Specificity

Polander and Shalin [64] discovered that online users of mental health services use psychological language in their narratives to express particular concepts, that they assess the content of particular psychology-related posts using the categories of leisure and work from the LIWC dictionary, and that they create a narrative dictionary based on the LIWC dictionary that includes causal words, time words, emotion words, and transition words. A 62,000-word-long English specificity dictionary was created by Muraki et al [65]. Similar to Muraki et al’s [65] approach, Xie and Bi [66] developed word-level and sentence-level specificity inference models based on large-scale Chinese specificity dictionaries. In our study, based on a Chinese specificity dictionary, we calculated the influence of the level of specificity of the comment text on cognitive change in support seekers. The specificity dictionary developed by Xie and Bi [66] is presented in tabular form in their study. The dictionary was transformed into a usable data format for this investigation. Key value matching was used to determine the specificity of each word in the comment text, and the average specificity of the words was derived as a gauge of sentence specificity.

Language Style

Function words in text can more accurately reflect the text’s language style than its substance [53]. The use of function words in online mental health communities was examined as a method of language style calculation [66-68]. Pronouns, prepositions, adverbs, conjunctions, auxiliary verbs, etc, are examples of function words that can express significant facets of the speaker’s motivation, requirements, and personality [69]. To conduct part-of-speech tagging, count words, and calculate function words as a measure of language style, the Baidu part-of-speech tagging tool LAC [70] was used in this study. This tool allows researchers to examine the impact of support providers’ language on changes in the cognition of support seekers in an online mental health community. Only the use of prepositions, adverbs, conjunctions, and auxiliary verbs was examined in this section because personal pronouns have already been covered in depth in the section on first-person words.

Control Variables

Some support providers use the second-person point of view in their comments, to give direct instruction and counsel based on the support seekers’ experiences. The likelihood of a support seeker showing improvement, however, decreases if past-tense verbs are used excessively. Thus, cognitive change may be influenced by the use of both second-person and past-tense verbs.

Yu et al [71] defined complexity as the cognitive load of knowing a certain subject, and Nadeem and Ostendorf [72] defined linguistic complexity as the difficulty faced by language users in understanding sentences. Therapists tend to use more difficult-to-read, grammatically complicated language when counseling clients who are more worried. This language use might be connected to the need for therapists to engage in deeper communication with such clients. One study calculated the complexity of the consulting language using 12 factors, such as lexical diversity, grammatical simplicity, readability, and typicality [73]. Online support seekers in online mental health communities typically describe their experiences in detail in their postings, so support providers typically do not need to engage in more in-depth communication by using language that is more challenging to read. This situation contrasts with that of face-to-face counseling.

The mental health difficulties posted online by support seekers may change over time due to other, unidentified variables

XSL•FO
RenderX

because the connections formed in the online community are asynchronous. Hence, our study used the interval between the time of the original post and the time of the supporting comments as a control variable. Additionally, those who seek support from a larger group of people may interact more frequently and feel certain psychological comforts, which influences their own cognitive improvement. As a result, this study also used the support seeker's number of followers, the interval between the original posting and the support seeker responding to comments, and the number of supporting comments as control variables.

In summary, this study's control variables were the use of the second person, the use of the past tense, topic complexity, the number of likes on support comments, posting and comment support times, and the number of comments. These variables were measured as follows: We calculate the number of second-person words and past-tense words by creating dictionaries. To determine the topic complexity, Lu et al [74] calculated the complexity of sentences and words. A lexical complexity prediction approach based on a 5-component Likert scale and manual annotation was proposed by Shardlow et al [75]; this method splits lexical complexity into 5 levels. By assessing the topic complexity of the text, Shin et al [76] used the LDA model (Genism package) to determine the complexity of the analysis text. Understanding certain words, certain sentences, and the variety of topics in the supportive comment text, when combined with psychological support situations, might influence cognitive change in support seekers. The number of words is used to gauge sentence difficulty after the preprocessing and word segmentation of the comment text. The number of subjects in the text is determined to gauge the topic complexity, and the LDA text-clustering technique is used to cluster the text.

Based on signaling theory, we constructed a mining model of the 6 independent variables: intimacy, emotional polarity, the use of first-person words, the use of future-tense words, specificity, and linguistic style. Cognitive change was used as the dependent variable. Table 2 displays the precise procedures used to calculate each variable.

Table 2. Description of variables.

Variables	Description
Dependent variable	
Cognitive change	Support seekers indicate in their responses to comments whether they are getting better.
Independent variables	
Intimacy	The number of intimate words in a support comment.
Emotional polarity	The emotion degree score (in the range of 0-1, where 1 is the strongest positive emotion) in a support comment.
First person	The number of first-person words in a support comment.
Future time	The number of future-tense words in a support comment.
Specificity	The degree to which a reference in a support comment can be perceived directly; the average specificity of support stickers based on the specificity dictionary.
Linguistic style	The number of function words (adverbs, prepositions, conjunctions, and auxiliary verbs) in a support comment.
Control variables	
Second person	The number of second-person words in a support comment.
Past tense	The number of past-tense verbs in a support comment.
Complexity	The number of words and the number of topics in a support comment.
Interval time	The calculated time between the original post and the support comment.
Approval number	The number of likes for a support comment.

Statistical Analysis

We used the statistical analysis program IBM SPSS 20.0 to analyze the correlation of variables. After the multicollinearity test, we used binomial logistic regression to test our hypotheses, considering that the independent and control variables were continuous and the dependent variable was dichotomous. The robustness test was carried out by using other models, as well as a sample subset. Specifically, we used the probit model to test the effect of the independent variables on the dependent variable. We then randomly selected 50% of the sample and ran binomial logistic regression and probit model tests on a subset of the sample.

Ethical Considerations

This study was approved by the Biomedical Ethics Committee of Hefei University of Technology (approval number HFUT20240520001H). Since the postings and comments of members of online mental health communities are related to user privacy, deidentified methods were adopted in this study: (1) After matching comment postings with reply postings, only the text of the comments and reply postings was retained, and (2) data, such as links and email addresses showing personal information contained in the text, were removed.

XSL•FO
RenderX

Results

Empirical Results

Multimedia Appendix 2 displays the correlation analysis results. The outcomes of the multicollinearity test in this study were tolerance and the variance inflation factor (VIF, which is the reciprocal of tolerance). After removing word complexity, no multicollinearity issue existed among the explanatory variables, because the VIF values of all of the explanatory variables were no more than 10, and thus, all explanatory variables could be used in the regression model.

To further understand the links between variables, we used a linear regression model to assess the elements influencing cognitive change. The 6 independent variables and the control variables were all continuous variables, but the dependent variable, cognitive change, was a dichotomous variable. Therefore, we used a binomial logistic regression model to investigate the elements. The findings demonstrated that at least 1 of the included variables was statistically significant. Table 3 shows the results of the regression analysis. Table 4 displays the findings of the hypothesis testing.

Table 3. Results of regression analysis.

Variable	β	Significance	*P* value	Exp(B)
Intimacy	–1.706	0.213	<.001	0.182
Emotional polarity	.890	0.058	<.001	2.434
First person	.120	0.014	<.001	1.127
Future time	.301	0.081	<.001	1.351
Specificity	–.018	0.005	<.001	0.982
Linguistic style	–.838	0.116	<.001	0.433
Second person	.013	0.016	.042	1.013
Past tense	–.068	0.071	.039	1.070
Complexity	–.002	0.003	.415	0.998
Approval number	$.1\times10^{-5}$	0.000	.007	1.000
Interval time	$.9\times10^{-5}$	0.000	<.001	1.000
B	.542	0.335	N/A[a]	N/A

[a]N/A: not applicable.

H1 concerned the impact of intimacy on cognitive change among support seekers in online mental health communities. The coefficient of prediction for cognitive change in the regression analysis, however, indicated that intimacy had a significant negative effect on cognitive change ($\beta_{intimacy}$=–1.706, *P*<.001). Support texts with fewer privately similar experiences are more likely to result in a change for the better for the support seeker.

The impact of emotional polarity on support seekers was tested in H2. This hypothesis—that a significant emotional polarity score favorably influences cognitive improvement ($\beta_{emotional\_polarity}$=.890, *P*<.001) and that support seekers feel comfortable and supported when support providers display more positive emotions—was validated by the regression analysis, which showed that emotional polarity has a positive regression coefficient. Straightforward helpful messages of approbation and encouragement can cause a change in cognition and pleasant emotions in support seekers.

H3 explored the effect of using first-person terms (ie, providing a self-identity of the supporter) on cognitive change in the supporter-seeker. The hypothesis was supported by the regression results, which demonstrated that the use of the first person has a significant positive impact on cognitive change ($\beta_{first\text{-}person}$=.120, *P*<.001). When support seekers are encouraged to improve by support providers sharing their own experiences or directly making suggestions, they are positively affected, leading to cognitive change.

The use of words related to the future is a temporal aspect, which was examined in H4. The results showed that the use of present- and future-tense words in support postings has a significant positive impact on the occurrence of change for the better in support seekers (β_{future_words}=.301, *P*<.001). Increased use of present- and future-tense words leads to positive cognitive change in support seekers.

H5 concerns how the degree of specificity in support posts affects whether support seekers experience a change for the better. The regression resulted in a significant negative effect of expression specificity on cognitive change ($\beta_{specificity}$=–.018, *P*<.001). This result implies that support comment postings containing more specific words do not cause support seekers to undergo a change for the better.

H6 examined how language style traits affect the cognitive shifts that support seekers' experience. According to the regression analysis, the use of function words has a detrimental impact on cognitive change ($\beta_{language\_style}$=–.838, *P*<.001). The findings support the idea that increasing the use of function words may not improve support seekers' ability to assist others in online health forums.

XSL•FO
RenderX

Table 4. Summary of hypothesis testing.

Hypothesis	Supported?
H1: Intimacy significantly positively affects cognitive change in support seekers in online mental health communities.	No
H2: Emotional polarity scores significantly positively influence cognitive change in support seekers in online mental health communities.	Yes
H3: First-person word use significantly positively affects cognitive change in support seekers in online mental health communities.	Yes
H4: Future temporal features significantly positively affect cognitive change in support seekers in online mental health communities.	Yes
H5: Specificity characteristics significantly positively affect cognitive change in support seekers in online mental health communities.	No
H6: Language style characteristics significantly negatively affect cognitive change in support seekers in online mental health communities.	Yes

Robustness Check

We also used the probit model to examine the effects of independent variables on the dependent variable in order to evaluate the robustness of the model. Tables 5-7 contain the test results. The positive, negative, and significant impacts of the independent variables on the dependent variable were all supported by the binomial regression model. After selecting 6434 (50%) of the 12,868 data items at random, the sample subset was subjected to binomial logistic regression and a probit model test. The test results showed that both positive and negative effects of the independent variables on the dependent variable were significant and matched the effects of the total sample.

Table 5. Results of probit regression analysis on the full sample (N=12,868).

Variable	β	Significance	*P* value
Intimacy	–.960	0.179	<.001
Emotional polarity	.465	0.037	<.001
First person	.054	0.007	<.001
Future time	.156	0.028	<.001
Specificity	–.013	0.003	<.001
Linguistic style	–.534	0.113	<.001
Second person	.013	0.010	.196
Past tense	.028	0.042	.507
Complexity	.0023	0.002	.349
Approval number	1×10^{-6}	0.000	.009
Interval time	8×10^{-6}	0.000	.001
B	.018	0.043	N/A[a]

[a]N/A: not applicable.

XSL•FO
RenderX

Table 6. Binomial logistic regression analysis results of 50% (n=6434) of the samples.

Variable	β	Significance	*P* value	Exp(B)
Intimacy	–1.895	0.407	<.001	0.150
Emotional polarity	.759	0.087	<.001	2.137
First person	.090	0.018	<.001	1.094
Future time	.230	0.067	<.001	1.259
Specificity	–.020	0.006	<.001	0.980
Linguistic style	–1.115	0.258	<.001	0.328
Second person	.012	0.023	.599	1.012
Past tense	.139	0.099	.161	1.149
Complexity	.002	0.004	.641	1.002
Approval number	1×10^{-6}	0.000	.192	1.000
Interval time	8×10^{-6}	0.000	.037	1.000
B	.101	0.099	N/A[a]	1.106

[a]N/A: not applicable.

Table 7. Results of probit regression analysis on the full sample (N=12,868).

Variable	β	Significance	*P* value
Intimacy	–1.151	0.252	<.001
Emotional polarity	.438	0.051	<.001
First person	.048	0.010	<.001
Future time	.139	0.040	<.001
Specificity	–.011	0.004	.003
Linguistic style	–.687	0.160	<.001
Second person	.009	0.014	.491
Past tense	.080	0.059	.171
Complexity	.001	0.002	.624
Approval number	1×10^{-6}	0.000	.174
Interval time	8×10^{-6}	0.000	.039
B	.087	0.061	N/A[a]

[a]N/A: not applicable.

Discussion

Principal Findings

This study provides valuable insights into the linguistic factors involved in support seekers in online mental health communities feeling better. We studied the impacts of intimacy, emotional polarity, the use of first-person words, the use of future-tense words, specificity, and language style on cognitive change in support seekers in online mental health communities. We found that support comments with lower intimacy, higher positive emotional polarity, lower specificity, more first-person words, more future- and present-tense words, and fewer function words are more likely to make support seekers feel better. Our results show that the effects of most factors are consistent with the hypotheses, except for intimacy and specificity.

According to signaling theory [28], affective signals make support seekers feel better by increasing feelings of empathy and similarity between support seekers and support providers. Our findings show that positive emotions, the use of first-person words, and the use of future-tense words can make support seekers feel better. This result indicates that support providers can make support seekers perceive similarity and develop emotional empathy through positive emotional guidance, their own similar experiences, and directing the gaze to the future so that improvement occurs. According to information signaling [28], language style makes support seekers feel better by reducing the cost of receiving and understanding support signals. This means that support providers can reduce the use of function words.

XSL•FO
RenderX

The findings also suggest that intimacy has a negative effect on cognitive change and that support texts with less intimacy are more likely to make support seekers feel better. However, this result is contrary to existing research[37,39], which suggests that the expression of intimacy positively contributes to mental health. This result may be explained by the fact that other people in the community, such as support seekers, provide social support by expressing similar experiences [77]. These experiences stimulate the support seekers' recollection of their own negative experiences, such as similar family environments, similar treatment experiences, or similar attitudes [78], thus preventing positive cognitive changes from occurring. However, if they encourage support seekers to elaborate on their own experiences and encourage them by expressing positive emotions, an improvement is seen, as discussed in H2.

The possible reasons the effect of specificity on cognitive change is negative are similar. Posts with more specificity often mean that support providers disclose more details of similar negative experiences. Specific words are more likely to make it easier for support seekers to recall their own similar negative experiences in the imaginal system, which is not conducive to making them feel better.

On this basis, we conclude that support providers providing social support by disclosing details of similar negative experiences is not conducive to supporting support seekers to feel better, and more positive emotions are needed to guide them toward the present and future. The result is consistent with psychotherapists' psychotherapeutic strategy in offline counseling scenarios [32].

Theoretical and Practical Implications

This study has 2 theoretical implications. First, we transferred signaling theory to the study of factors influencing cognitive change in online mental health communities and divided the factors into affective signals and information signals. We also verified the proposed model. The findings suggest that intimacy, emotional polarity, discussion of self, use of time-related words, specificity, and language style (use of function words) all significantly influence cognitive change in support seekers in online mental health communities. Thus, this study extended signaling theory. Second, based on previous research, but also going beyond that, this study investigated how new factors, such as intimacy and specificity, affect cognitive change. It calculated the intimacy and specificity of supporting comments by creating intimacy and specificity dictionaries, and then it confirmed the impact of these 2 factors on cognitive change using regression analysis. The outcomes of this study related to intimacy and specificity disagree with other findings; we discovered that both intimacy and specificity are associated with the psychological support provided in online mental health communities and that both are adversely correlated with cognitive change.

The research also has practical implications. Our findings encourage tailored strategies for mental health support and intervention. Online mental health community managers can use context-specific support techniques, offer real-time ideas when posting support comments, and provide information for the design of the online interface by researching variables that influence the cognitive shifts in support seekers. In addition, community support providers can customize their replies by using the best support comment tactics for each support seeker based on their unique traits. To foster relationships between support seekers and support providers, emphasis should be placed on encouraging other members of the online mental health community to write and offer assistance. Interventions can be created to allow service providers to cater their assistance to each support seeker individually. Most significantly, support seekers who have posted an issue and who are not receiving adequate support can be helped by the referral, identification, and mentoring procedures of the online health community.

Limitations and Future Recommendations

Our study has several limitations. First, we included 1 online mental health community, and the findings may differ depending on the data sources used. Other contexts should be examined in future studies. Second, we did not set up support strategies in an actual online mental health community to verify whether they would improve the quality of support. Future research can strategically aid auxiliary support providers in writing supportive comments by analyzing and designing strategies and combinations of strategies so that the comments will better aid support seekers and lead to cognitive change. In addition, the characteristics of support seekers should be fully considered in functional design in order to further provide personalized support strategies. For example, the features of support comments that are more agreeable to support seekers with high participation levels versus those with low participation levels should be considered, as well as variations in the features of support comments used to treat various mental health issues.

Conclusion

This study investigated elements that influence cognitive change in online mental health communities. It used psychological counselors' face-to-face consultation techniques, described in existing research, to propose hypotheses, and it specifically applied signaling theory to identify factors that influence cognitive change. We found that support comments are more likely to alter support seekers' cognitive processes if they have lower intimacy, higher positive emotional polarity, lower specificity, more first-person words, more future- and present-tense words, and fewer function words. The result is consistent with psychotherapists' psychotherapeutic strategy in offline counseling scenarios.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (grant numbers 72131006, 72271082, and 72071063) and the Natural Science Foundation of Anhui Province (grant number 2408085J041).

XSL•FO
RenderX

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

ML and DG contributed equally. ML conceptualized the study, drafted the methodology, performed data curation and software analysis, and prepared the original draft. DG conceptualized and investigated the study. RL reviewed and edited the draft. YD completed the investigation. HL drafted the methodology and performed data curation. KX conceptualized the study. XY and GR supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of support text and reply text.
[DOCX File , 13 KB - jmir_v27i1e60292_app1.docx]

Multimedia Appendix 2
Relevant analysis results.
[DOCX File , 22 KB - jmir_v27i1e60292_app2.docx]

References

1. Bizzotto N, Marciano L, de Bruijn G, Schulz PJ. The empowering role of web-based help seeking on depressive symptoms: systematic review and meta-analysis. J Med Internet Res 2023 Feb 02;25:e36964 [FREE Full text] [doi: 10.2196/36964] [Medline: 36729571]
2. Stefanone M, Kwon K, Lackaff D. Exploring the relationship between perceptions of social capital and enacted support online. J Comput-Mediat Comm 2012 Sep 13;17(4):451-466. [doi: 10.1111/j.1083-6101.2012.01585.x]
3. Kim J. Smartphone-mediated communication vs. face-to-face interaction: two routes to social support and problematic use of smartphone. Comput Hum Behav 2017 Feb;67:282-291. [doi: 10.1016/j.chb.2016.11.004]
4. Vacchiano M, Valente R. Did the screens win? An autoregressive model linking leisure, relatedness and mental health. Comput Hum Behav 2021 Jul;120:106755. [doi: 10.1016/j.chb.2021.106755]
5. Chen W. Internet use, online communication, and ties in Americans' networks. Soc Sci Comput Rev 2013 Mar 15;31(4):404-423. [doi: 10.1177/0894439313480345]
6. Kawachi I, Berkman LF. Social ties and mental health. J Urban Health 2001 Sep;78(3):458-467 [FREE Full text] [doi: 10.1093/jurban/78.3.458] [Medline: 11564849]
7. Vacchiano M, Bolano D. Online and offline leisure, relatedness and psychological distress: a study of young people in Switzerland. Leisure Stud 2020 Dec 17;40(3):338-351. [doi: 10.1080/02614367.2020.1862282]
8. Lu W, Hampton KN. Beyond the power of networks: differentiating network structure from social media affordances for perceived social support. New Media Soc 2016 Jan 04;19(6):861-879. [doi: 10.1177/1461444815621514]
9. Huang K, Chengalur-Smith I, Pinsonneault A. Sharing is caring: social support provision and companionship activities in healthcare virtual support communities. MISQ 2019 Jan 1;43(2):395-423. [doi: 10.25300/misq/2019/13225]
10. Pereira-Sanchez V, Adiukwu F, El Hayek S, Bytyçi DG, Gonzalez-Diaz JM, Kundadak GK, et al. COVID-19 effect on mental health: patients and workforce. Lancet Psychiatry 2020 Jun;7(6):e29-e30 [FREE Full text] [doi: 10.1016/S2215-0366(20)30153-X] [Medline: 32445691]
11. Park A, Conway M. Harnessing Reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]
12. Pruksachatkun Y, Pendse S, Sharma A. Moments of change: analyzing peer-based cognitive support in online mental health forums. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems Glasgow Scotland Uk May 4 - 9, 2019; May 4-9, 2019; Glasgow, Scotland p. 1-13. [doi: 10.1145/3290605.3300294]
13. Gu D, Wang Q, Chai Y, Yang X, Zhao W, Li M, et al. Identifying the risk factors of allergic rhinitis based on Zhihu comment data using a topic-enhanced word-embedding model: mixed method study and cluster analysis. J Med Internet Res 2024 Feb 22;26:e48324 [FREE Full text] [doi: 10.2196/48324] [Medline: 38386404]
14. Gu D, Liu H, Zhao H, Yang X, Li M, Liang C. A deep learning and clustering‐based topic consistency modeling framework for matching health information supply and demand. J Assoc Inf Sci Technol 2023 Nov;75(2):152-166. [doi: 10.1002/asi.24846]
15. de Waal FBM, Preston SD. Mammalian empathy: behavioural manifestations and neural basis. Nat Rev Neurosci 2017 Aug;18(8):498-509. [doi: 10.1038/nrn.2017.72] [Medline: 28655877]

XSL•FO
RenderX

16. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. TACL 2016 Dec;4:463-476. [doi: 10.1162/tacl_a_00111]
17. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 2023 Jan 23;5(1):46-57. [doi: 10.1038/s42256-022-00593-2]
18. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. Comput Hum Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
19. Wadden D, August T, Li Q, Althoff T. The effect of moderation on online mental health conversations. 2021 Presented at: ICWSM-2021: 15th International AAAI Conference on Web and Social Media; June 8-10, 2021; Virtual p. 751-763. [doi: 10.1609/icwsm.v15i1.18100]
20. Yang X, Gu D, Li H, Liang C, Jain H, Li P. Mobile health community loyalty development process in China: an empirical study from information seeking perspective. ITP 2023 Feb 20;37(2):635-661. [doi: 10.1108/itp-11-2021-0835]
21. Perez-Rosas V, Sun X, Li C, Wang Y, Resnicow K, Mihalcea R. Analyzing the quality of counseling conversations: the tell-tale signs of high-quality counseling. 2018 Presented at: Eleventh International Conference on Language Resources and Evaluation (LREC 2018); May 7-12, 2018; Miyazaki, Japan.
22. Peng Z, Ma X, Yang D, Tsang K, Guo Q. Effects of support-seekers? community knowledge on their expressed satisfaction with the received comments in mental health communities. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445446]
23. Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. J Med Internet Res 2020 Nov 24;22(11):e22600 [FREE Full text] [doi: 10.2196/22600] [Medline: 33156805]
24. De Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. 2014 Presented at: ICWSM 2014: 8th International AAAI Conference on Weblogs and Social Media; June 2-4, 2014; Ann Arbor, MI p. 71-80. [doi: 10.1609/icwsm.v8i1.14526]
25. Chikersal P, Belgrave D, Doherty G, Enrique A, Palacios J, Richards D, et al. Understanding client support strategies to improve clinical outcomes in an online mental health intervention. 2020 Presented at: CHI '20: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu HI. [doi: 10.1145/3313831.3376341]
26. Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. Int J Med Inform 2017 Feb;98:22-32. [doi: 10.1016/j.ijmedinf.2016.11.006] [Medline: 28034409]
27. Liu X, Wang GA, Fan W, Zhang Z. Finding useful solutions in online knowledge communities: a theory-driven design and multilevel analysis. Inf Syst Res 2020 Sep;31(3):731-752. [doi: 10.1287/isre.2019.0911]
28. Chen L, Baird A, Straub D. A linguistic signaling model of social support exchange in online health communities. Decis Support Syst 2020 Mar;130:113233. [doi: 10.1016/j.dss.2019.113233]
29. Spence M. Job market signaling. Q J Econ 1973 Aug;87(3):355. [doi: 10.2307/1882010]
30. Feng X, Hu Y, Pfaff H, Liu S, Xie J, Zhang Z. Exploring client preferences for psychological counselors in a Chinese online health community: longitudinal study. J Med Internet Res 2024 Oct 10;26:e58428 [FREE Full text] [doi: 10.2196/58428] [Medline: 39388694]
31. Chen L, Rai A, Chen W, Guo X. Signaling effects under dynamic capacity in online matching platforms: evidence from online health consultation communities. Inf Syst Res 2024 Mar 28:isre.2021.0150. [doi: 10.1287/isre.2021.0150]
32. Sharma A, Miner A, Atkins D, Althoff T. A computational approach to understanding empathy expressed in text-based mental health support. WARNING: this paper contains content related to suicide and self-harm. 2020 Presented at: 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); November 16-20, 2020; Online URL: https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000855160705035 [doi: 10.18653/v1/2020.emnlp-main.425]
33. Chiu C, Hsu M, Wang ET. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. Decis Support Syst 2006 Dec;42(3):1872-1888. [doi: 10.1016/j.dss.2006.04.001]
34. Fan H, Lederman R. Online health communities: how do community members build the trust required to adopt information and form close relationships? Eur J Inf Syst 2017 Nov 14;27(1):62-89. [doi: 10.1080/0960085x.2017.1390187]
35. Liang C, Wang X, Gu D, Li P, Chen H, Xu Z. Smart management information systems (Smis): concept, evolution, research hotspots and applications. Data Intell 2023 Nov 1;5(4):857-884. [doi: 10.1162/dint_a_00231]
36. Jiang L, Geng Z, Gu D, Guo S, Huang R, Cheng H, et al. RS-SVM machine learning approach driven by case data for selecting urban drainage network restoration scheme. Data Intell 2023 May 9;5(2):413-437. [doi: 10.1162/dint_a_00208]
37. Zhang Y. Understanding the sustained use of online health communities from a self‐determination perspective. J Assoc Inf Sci Technol 2015 Jul 07;67(12):2842-2857. [doi: 10.1002/asi.23560]
38. Kersten P, Borschel E, Neyer FJ, Mund M. The social side of personality: do affiliation and intimacy motives moderate associations of personal relationships with well-being? J Pers 2023 Aug;91(4):992-1011. [doi: 10.1111/jopy.12746] [Medline: 35716149]
39. Lambert A. Intimacy and social capital on Facebook: beyond the psychological perspective. New Media Soc 2016 Jul 09;18(11):2559-2575. [doi: 10.1177/1461444815588902]

XSL•FO
RenderX

40. Sutcliffe A, Binder J, Dunbar R. Activity in social media and intimacy in social relationships. Comput Hum Behav 2018 Aug;85:227-235. [doi: 10.1016/j.chb.2018.03.050]
41. Li N, Wu DD. Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decis Support Syst 2010 Jan;48(2):354-368. [doi: 10.1016/j.dss.2009.09.003]
42. Ortigosa A, Martín JM, Carro RM. Sentiment analysis in Facebook and its application to e-learning. Comput Hum Behav 2014 Feb;31:527-541. [doi: 10.1016/j.chb.2013.05.024]
43. Perez-Rosas V, Wu X, Resnicow K, Mihalcea R. What makes a good counselor? Learning to distinguish between high-quality and low-quality counseling conversations. 2019 Presented at: 57th Annual Meeting of the Association for Computational Linguistics; July 28-August 2, 2019; Florence, Italy p. 926-935. [doi: 10.18653/v1/p19-1088]
44. Gogoulou E, Boman M, Ben AF, Isacsson N, Kaldo V, Sahlgren M. Predicting treatment outcome from patient texts: the case of internet-based cognitive behavioural therapy. 2021 Presented at: EACL 2021: 16th Conference of the European Chapter of the Association for Computational Linguistics; April 19-23, 2021; Kyiv, Ukraine p. 575-580 URL: https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000863557000046 [doi: 10.18653/v1/2021.eacl-main.46]
45. Shapira N, Lazarus G, Goldberg Y, Gilboa-Schechtman E, Tuval-Mashiach R, Juravski D, et al. Using computerized text analysis to examine associations between linguistic features and clients' distress during psychotherapy. J Couns Psychol 2021 Jan;68(1):77-87. [doi: 10.1037/cou0000440] [Medline: 32352823]
46. Arntz A, Hawke LD, Bamelis L, Spinhoven P, Molendijk ML. Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. Behav Res Ther 2012 Mar;50(3):191-202. [doi: 10.1016/j.brat.2011.12.007] [Medline: 22317755]
47. Nook EC, Hull TD, Nock MK, Somerville LH. Linguistic measures of psychological distance track symptom levels and treatment outcomes in a large set of psychotherapy transcripts. Proc Natl Acad Sci U S A 2022 Mar 29;119(13):e2114737119 [FREE Full text] [doi: 10.1073/pnas.2114737119] [Medline: 35316132]
48. Zhejiang U, Huang L, Tan C, National UOS, Ke W, Clarkson U, et al. Helpfulness of online review content: the moderating effects of temporal and social cues. J Assoc Inf Syst 2018 Jun;19(06):503-522. [doi: 10.17705/1jais.00499]
49. Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev 2010 Apr;117(2):440-463 [FREE Full text] [doi: 10.1037/a0018963] [Medline: 20438233]
50. Ho CKY, Ke W, Liu H, Chau PYK. Separate versus joint evaluation: the roles of evaluation mode and construal level in technology adoption. MISQ 2020 Jun 1;44(2):725-746. [doi: 10.25300/misq/2020/14246]
51. Rains SA. Language style matching as a predictor of perceived social support in computer-mediated interaction among individuals coping with illness. Commun Res 2015 Jan 13;43(5):694-712. [doi: 10.1177/0093650214565920]
52. Ludwig S, de Ruyter K, Mahr D, Wetzels M, Brüggen E, de Ruyck T. Take their word for it: the symbolic role of linguistic style matches in user communities. MISQ 2014 Apr 4;38(4):1201-1217. [doi: 10.25300/misq/2014/38.4.12]
53. Ireland ME, Pennebaker JW. Language style matching in writing: synchrony in essays, correspondence, and poetry. J Pers Soc Psychol 2010 Sep;99(3):549-571. [doi: 10.1037/a0020386] [Medline: 20804263]
54. Chung C, Pennebaker J. The psychological functions of function words. In: Fiedler K, editor. Social Communication. New York, NY: Psychology Press; 2007:343-359.
55. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
56. Grice HP. Logic and conversation. In: Cole P, Morgan JL, editors. Syntax and Semantics, Vol. 3, Speech Acts. New York, NY: Academic Press; 1975:41-58.
57. Gu D, Li M, Yang X, Gu Y, Zhao Y, Liang C, et al. An analysis of cognitive change in online mental health communities: a textual data analysis based on post replies of support seekers. Inf Process Manag 2023 Mar;60(2):103192. [doi: 10.1016/j.ipm.2022.103192]
58. Gu D, Deng S, Zheng Q, Liang C, Wu J. Impacts of case-based health knowledge system in hospital management: the mediating role of group effectiveness. Inf Manag 2019 Dec;56(8):103162. [doi: 10.1016/j.im.2019.04.005]
59. Hu D, Wang M, Gao F, Xu F, Gu J. Knowledge representation and reasoning for complex time expression in clinical text. Data Intell 2022 Jul 1;4(3):573-598. [doi: 10.1162/dint_a_00152]
60. Gilbert E, Karahalios K. Predicting tie strength with social media. 2009 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA p. 211-220 URL: https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000265679300027 [doi: 10.1145/1518701.1518736]
61. Pei J, Jurgens D. Quantifying intimacy in language. 2020 Presented at: 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); November 16-20, 2020; Online p. 5307-5326 URL: https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000855160705038 [doi: 10.18653/v1/2020.emnlp-main.428]
62. Reuel A, Peralta S, Sedoc J, Sherman G, Ungar L. Measuring the language of self-disclosure across corpora. 2022 Presented at: ACL 2022: 60th Annual Meeting of the Association for Computational Linguistics; May 22-27, 2022; Dublin, Ireland p. 1035-1047 URL: https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:000828767401008 [doi: 10.18653/v1/2022.findings-acl.83]
63. Lyons M, Aksayli ND, Brewer G. Mental distress and language use: linguistic analysis of discussion forum posts. Comput Hum Behav 2018 Oct;87:207-211. [doi: 10.1016/j.chb.2018.05.035]

XSL•FO
RenderX

64. Polander E, Shalin VL. An analysis of narrative and figurative language within online alcoholism discussion forums. Int J Ment Health Addiction 2013 Feb 16;11(4):458-473. [doi: 10.1007/s11469-013-9431-8]
65. Muraki EJ, Abdalla S, Brysbaert M, Pexman PM. Concreteness ratings for 62,000 English multiword expressions. Behav Res Methods 2023 Aug;55(5):2522-2531. [doi: 10.3758/s13428-022-01912-6] [Medline: 35867207]
66. Xie Z, Bi R. Construction and inference technique of large-scale Chinese concreteness lexicon. Act Sci Nat Univrsitatis Pekin 2022;58(1):1-6. [doi: 10.13209/j.0479-8023.2021.100]
67. Gu D, Yang X, Deng S, Liang C, Wang X, Wu J, et al. Tracking knowledge evolution in cloud health care research: knowledge map and common word analysis. J Med Internet Res 2020 Feb 25;22(2):e15142 [FREE Full text] [doi: 10.2196/15142] [Medline: 32130115]
68. Yang X, Gu D, Wu J, Liang C, Ma Y, Li J. Factors influencing health anxiety: the stimulus–organism–response model perspective. INTR 2021 Jun 24;31(6):2033-2054. [doi: 10.1108/intr-10-2020-0604]
69. Ríssola EA, Aliannejadi M, Crestani F. Mental disorders on online social media through the lens of language and behaviour: analysis and visualisation. Inf Process Manag 2022 May;59(3):102890. [doi: 10.1016/j.ipm.2022.102890]
70. Jiao Z, Sun S, Sun K. Chinese lexical analysis with deep bi-GRU-CRF network. arXiv Preprint posted online 2012..
71. Yu B, He Z, Xing A, Lustria MLA. An informatics framework to assess consumer health language complexity differences: proof-of-concept study. J Med Internet Res 2020 May 21;22(5):e16795 [FREE Full text] [doi: 10.2196/16795] [Medline: 32436849]
72. Nadeem F, Ostendorf M. Estimating linguistic complexity for science texts. In: Estimating linguistic complexity for science texts. 2018 Presented at: Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications; June 5, 2018; New Orleans, LA p. 45-55. [doi: 10.18653/v1/w18-0505]
73. Wei J, Finn K, Templeton E, Wheatley T, Vosoughi S. Linguistic complexity loss in text-based therapy. 2021 Presented at: 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 6-11, 2021; Online p. 4450-4459. [doi: 10.18653/v1/2021.naacl-main.352]
74. Lu C, Bu Y, Wang J, Ding Y, Torvik V, Schnaars M, et al. Examining scientific writing styles from the perspective of linguistic complexity. J Assos Inf Sci Technol 2018 Dec 23;70(5):462-475. [doi: 10.1002/asi.24126]
75. Shardlow M, Cooper M, Zampieri M. CompLex — a new corpus for lexical complexity prediction from Likert Scale data. 2020 Presented at: 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI); May 11-16, 2020; Marseille, France p. 57-62 URL: https://aclanthology.org/2020.readi-1.9
76. Shin D, He S, Lee GM, Whinston AB, Cetintas S, Lee K. Enhancing social media analysis with visual data analytics: a deep learning approach. MISQ 2020 Dec 1;44(4):1459-1492. [doi: 10.25300/misq/2020/14870]
77. Green BM, Van Horn KT, Gupte K, Evans M, Hayes S, Bhowmick A. Assessment of adaptive engagement and support model for people with chronic health conditions in online health communities: combined content analysis. J Med Internet Res 2020 Jul 07;22(7):e17338 [FREE Full text] [doi: 10.2196/17338] [Medline: 32492651]
78. Hargreaves S, Bath PA, Duffin S, Ellis J. Sharing and empathy in digital spaces: qualitative study of online health forums for breast cancer and motor neuron disease (amyotrophic lateral sclerosis). J Med Internet Res 2018 Jun 14;20(6):e222 [FREE Full text] [doi: 10.2196/jmir.9709] [Medline: 29903695]

Abbreviations

LDA: latent Dirichlet allocation
LIWC: Linguistic Inquiry and Word Count
NLP: natural language processing
Q&A: question-and-answer
VIF: variance inflation factor

Edited by T de Azevedo Cardoso; submitted 11.05.24; peer-reviewed by M Vacchiano, A AL-Asadi; comments to author 09.09.24; revised version received 04.11.24; accepted 10.12.24; published 14.01.25.

Please cite as:
Li M, Gu D, Li R, Gu Y, Liu H, Su K, Wang X, Zhang G
The Impact of Linguistic Signals on Cognitive Change in Support Seekers in Online Mental Health Communities: Text Analysis and Empirical Study
J Med Internet Res 2025;27:e60292
URL: https://www.jmir.org/2025/1/e60292
doi:10.2196/60292
PMID:

XSL•FO
RenderX

©Min Li, Dongxiao Gu, Rui Li, Yadi Gu, Hu Liu, Kaixiang Su, Xiaoyu Wang, Gongrang Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Explainable Predictive Model for Suicidal Ideation During COVID-19: Social Media Discourse Study

Salah Bouktif[1*], PhD; Akib Mohi Ud Din Khanday[1*], PhD, PDF; Ali Ouni[2], PhD

[1]Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

[2]Department of Software Engineering and Information Technology, École de Technologie Supérieure, Montreal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Salah Bouktif, PhD
Department of Computer Science and Software Engineering
College of Information Technology
United Arab Emirates University
Sheikh Khalifa Bin Zayed, Asharij
Al Ain, Abu Dhabi, 1551
United Arab Emirates
Phone: 971 507605406
Email: salahb@uaeu.ac.ae

Abstract

Background: Studying the impact of COVID-19 on mental health is both compelling and imperative for the health care system's preparedness development. Discovering how pandemic conditions and governmental strategies and measures have impacted mental health is a challenging task. Mental health issues, such as depression and suicidal tendency, are traditionally explored through psychological battery tests and clinical procedures. To address the stigma associated with mental illness, social media is used to examine language patterns in posts related to suicide. This strategy enhances the comprehension and interpretation of suicidal ideation. Despite easy expression via social media, suicidal thoughts remain sensitive and complex to comprehend and detect. Suicidal ideation captures the new suicidal statements used during the COVID-19 pandemic that represents a different context of expressions.

Objective: In this study, our aim was to detect suicidal ideation by mining textual content extracted from social media by leveraging state-of-the-art natural language processing (NLP) techniques.

Methods: The work was divided into 2 major phases, one to classify suicidal ideation posts and the other to extract factors that cause suicidal ideation. We proposed a hybrid deep learning–based neural network approach (Bidirectional Encoder Representations from Transformers [BERT]+convolutional neural network [CNN]+long short-term memory [LSTM]) to classify suicidal and nonsuicidal posts. Two state-of-the-art deep learning approaches (CNN and LSTM) were combined based on features (terms) selected from term frequency–inverse document frequency (TF-IDF), Word2vec, and BERT. Explainable artificial intelligence (XAI) was used to extract key factors that contribute to suicidal ideation in order to provide a reliable and sustainable solution.

Results: Of 348,110 records, 3154 (0.9%) were selected, resulting in 1338 (42.4%) suicidal and 1816 (57.6%) nonsuicidal instances. The CNN+LSTM+BERT model achieved superior performance, with a precision of 94%, a recall of 95%, an F_1-score of 94%, and an accuracy of 93.65%.

Conclusions: Considering the dynamic nature of suicidal behavior posts, we proposed a fused architecture that captures both localized and generalized contextual information that is important for understanding the language patterns and predict the evolution of suicidal ideation over time. According to Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) XAI algorithms, there was a drift in the features during and before COVID-19. Due to the COVID-19 pandemic, new features have been added, which leads to suicidal tendencies. In the future, strategies need to be developed to combat this deadly disease.

(J Med Internet Res 2025;27:e65434) doi:10.2196/65434

XSL•FO
RenderX

KEYWORDS

COVID-19; suicide; social networking sites; deep learning; explainable artificial intelligence; suicidal ideation; artificial intelligence; AI; social media; predictive model; mental health; pandemic; natural language processing; NLP; suicidal thought; deep neural network approach

Introduction

Overview

Mental health has become a critical area of research as societal challenges, such as pandemics, natural disasters, and economic instability, deeply affect individuals' well-being. According to comprehensive statistics provided by the World Health Organization (WHO), the occurrence of suicide is documented to be at an average rate of 1 incident every 40 seconds [1]. The report reveals that an estimated 800,000 people die by suicide annually, with suicide attempts occurring at a rate 20 times higher than that of completed suicides [2]. The worldwide approximation for suicide mortality is set at around 1 million deaths per year [3]. These statistics highlight suicide as the primary cause of death among young individuals, particularly among women. Contrary to an "all or nothing" perspective, a well-known book on the subject [4] suggests that suicide follows a distinct process and pattern. The pattern begins with suicidal ideation and progresses toward a suicide attempt and may culminate in a completed suicide. Although not all instances of suicidal ideation lead to complete suicide, they pose a significant threat for individuals to attempt suicide. Recognition of suicidality may occur when individuals frequently discuss it with their caretakers or during interactions with psychiatrists or psychologists who inquire about their thoughts and moods. Caregivers can mitigate the risks associated with suicidality by analyzing various warning signs and implementing necessary preventive measures. However, a significant obstacle in addressing suicidality is the hesitancy of individuals to cooperate with clinicians, often influenced by the social stigma linked to mental illness. The widespread impact of stigma on suicidality affects large-scale clinical interventions for at-risk individuals. Data indicate that 36% of individuals who succumb to suicide leave behind a note [5]. Research suggests that these suicide notes often signal a heightened probability of subsequent attempts that characterize greater precision after an initial failure [6]. The contents of these notes frequently underscore emotions of shame and apology, hinting at a potential willingness to consider alternatives. These notes are typically discovered postcompletion or, at the least, after an attempted suicide [6]. Despite efforts to screen patients, convincing individuals to undergo evaluation remains a formidable challenge in societies marked by stigma [7,8]. Studies highlight that people feel more at ease sharing their day-to-day experiences on online social networks (OSNs), where concerns about social stigma are less prominent [9]. Recent research also suggests that monitoring social media provides an alternative and valuable opportunity to identify warning signs associated with posts from individuals at risk of suicidality [10]. According to the statistics provided by the American Foundation for Suicide Prevention [11], there is a rise in people with mental health conditions. As per the collaborative multiyear project "Suicide Prevention Now" [12], according to statistics, in 2022, there was a significant increase in the number of people reporting mental health conditions. For instance, in 2022, 67% of American believed that they had a mental health condition at some point in their life.

Related Work and Background Knowledge

Mental health issues, such as depression and suicidal tendency, have traditionally been explored through psychological battery tests and clinical procedures [5,13]. However, to address the stigma associated with mental illness, researchers have turned to less formal platforms, such as social media, to examine language patterns in posts related to suicide. This strategy seeks to enhance the comprehension and interpretation of suicidal ideation. O'Dea et al [14] highlighted the crucial role of questionnaires in assessing a patient's mental state. However, with the surge in social networking sites, individuals find it more comfortable to openly express their feelings on social media platforms. Studies [13,14] have explored various scales used to predict depression through social media analysis. Additional research [15,16] has examined the topics commonly discussed by potentially suicidal individuals on social media, and the behavior of those contemplating suicide has been scrutinized in these studies. Advances in natural language processing (NLP) and machine learning techniques now allow the extraction of semantic information from social media posts, facilitating the automation of predicting suicidal content [17]. Most research has focused on using binary classification mechanisms with popular algorithms, such as support vector machines (SVMs), decision trees, and ensemble learning algorithms [18-21]. Deep learning methods have also been used to aid in predicting suicidal ideation [22,23]. In Multimedia Appendix 1, we extend the literature review where we explored relevant research works on suicidal ideation prediction (eg, [24-26]).

Feature selection involves identifying and selecting the most relevant attributes or features from the data. We used various feature selection and deep learning techniques, such as term frequency–inverse document frequency (TF-IDF), Word2vec, Bidirectional Encoder Representations from Transformers (BERT), convolutional neural networks (CNNs), and long short-term memory (LSTM). The TF-IDF method involves breaking down text into tokens, calculating the term frequency (TF) to evaluate the significance of each term within a document and determining the inverse document frequency (IDF) to evaluate the uniqueness of each term across the entire data set. Word2vec learns distributed representations (word embeddings) of words based on their context in a given corpus. The resulting word vectors can be used as feature representations for various text-based tasks. Before applying Word2vec, it is essential to preprocess the text data. The Word2vec model is trained on preprocessed text data using either the Skip-gram or the Continuous Bag of Words (CBOW) architecture. We fused deep learning techniques (CNN, LSTM, BERT) in order to effectively identify suicidal ideation as posts with suicidal behavior have a dynamic nature and longer textual segments. BERT is a

XSL•FO
RenderX

transformer-based model that uses self-attention mechanisms to capture contextual information. The architecture includes multiple layers of attention and feed-forward networks. BERT provides contextualized embeddings for each token in the input text. In a CNN, input data are a sequence of words or embeddings. The CNN architecture typically involves convolutional layers followed by pooling layers and fully connected layers. A CNN detects important localized features, such as word sequences, n-grams, and syntactic patterns. Each convolutional layer extracts high-level features from the text, which enables the model to capture complex patterns and semantics at different levels of granularity. It also handles input sequences of different post lengths without requiring explicit padding or truncation. LSTM networks are particularly effective for handling sequential data, such as text. LSTM networks are designed to capture long-range dependencies between words and phrases for better understanding of the context and are commonly used in text classification tasks. It propagates information over long distances in the input sequence of data. It processes input texts of different lengths without compromising performance and preserves the sequential structure and context of the posts. The methods are discussed in Multimedia Appendix 2, which provides a theoretical and mathematical background of the deep learning approaches and feature selection methods used in this work.

Aims of the Study

This work aimed to identify patterns and indicators associated with suicidal ideation by leveraging machine and deep learning algorithms for social media data. The primary focus of this study was to examine the correlation between COVID-19 and the prevalence of suicidal ideation. In this work, various factors were studied related to the pandemic, such as societal, economical, and psychological factors, to detect individuals' inclination toward suicidal thoughts due to COVID-19. Through comprehensive analysis and empirical investigations, the study sought to find the relationship between COVID-19 and suicidal ideation by detecting potential contributing factors and implications for mental health and well-being. Through rigorous analysis and modeling using deep learning methodologies, this work will contribute by finding the intersection between COVID-19 and suicide, which will be used for developing preventive interventions and support strategies to address future pandemics as well.

The study revolved around the research question of whether COVID-19 has impacted suicidal ideation. The following hypotheses were developed:

> *Null hypothesis (H0): COVID-19 has not impacted suicidal ideation, and the features (terms) used for suicidal ideation are the same as before.*
>
> *Hypothesis 1 (H1): COVID-19 has impacted suicidal ideation, and the sequence of terms has changed.*

Methods

Study Design

This work was divided into 5 phases, starting from data (suicidal posts) collection from Reddit, which is a publicly available data set. In the second phase, the data set was filtered based on various keywords. The emotions of the posts were extracted using the Natural Language Toolkit (NLTK), and the toxicity of the posts was calculated using the perspective application programming interface (API). The final corpus was used for further experimentation, and textual features were extracted using TF-IDF, Word2vec, and BERT techniques. The various classifiers were trained and tested on the filtered data set to classify the posts into 2 classes, suicidal and nonsuicidal. Another aspect of this work was explainable artificial intelligence (XAI); in this phase, 2 techniques, Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP), were used to extract the most relevant keywords that trigger suicidal ideation. The study framework is depicted in Figure 1.

XSL•FO
RenderX

Figure 1. Framework for suicidal ideation detection in the COVID-19 era based on various filtering mechanisms. AI: artificial intelligence; BERT: Bidirectional Encoder Representations from Transformers; TF-IDF: term frequency–inverse document frequency.

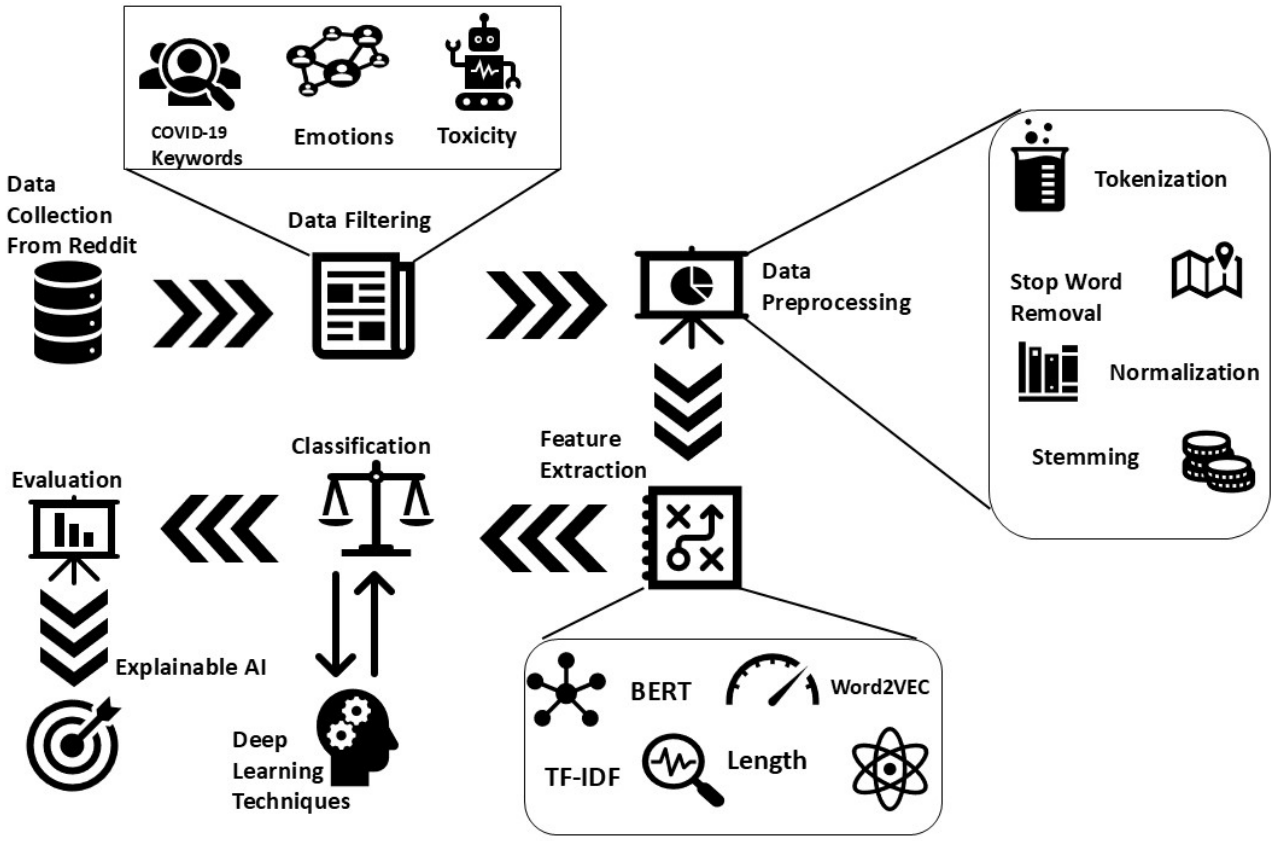

Data Collection and Preparation

Data are the most important entity in all data science research. In this work, there were privacy concerns related to the data. We collected relevant data from the public data repository Kaggle, which contains suicidal posts from the social media platform Reddit. The data were categorized into 2 classes, suicidal and nonsuicidal. The data set consisted of 2 columns, one with the posts (textual) and the other with labels. We leveraged a keyword-based filtering mechanism to classify and analyze user-generated content. We established a comprehensive set of keywords associated with COVID-19 and suicidal ideation, such as "COVID," "COVID-19," "lockdown," "quarantine," and "coronavirus." Using these keywords, a dual-layer filtering approach was implemented that categorized the posts into either COVID-19 related or not, followed by an assessment for suicidal content. For further exploration of the data, we performed emotion analysis (EA) to extract the mood of the poster. Eight different emotions were used for EA. After EA, we performed a toxicity analysis (TA) to identify toxic posts, and toxic values were associated with all the posts. We considered only negative, sad, and fearful emotions and posts that had a toxicity value greater than 0.5. In this study, we used the absolute majority (ie, a majority is when more than half of the total weighted words are toxic words). Another alternative was to use two-thirds majority; however, this was not a conservative choice because we risked having a large number of false positives. So, an absolute majority was more protective. By establishing a threshold of 0.5, a clear and straightforward benchmark was created for classifying outputs from our detection models. This threshold simplified both interpretation and implementation for detecting suicidal ideation. It also helped reduce ambiguity in the decision-making process. By doing so, we ensured that texts flagged as indicative of suicidal ideation were those with a probability score of 0.5 or higher. This minimized the risk of misinterpretation as it directed attention to those posts that met a specific criterion for concern. In addition, the empirical accuracies showed that our choice was effective in detecting suicidal ideation. After filtering the posts, we performed data cleaning and preprocessing, in which the textual data was refined using techniques such as tokenization, lemmatization, and normalization. Furthermore, irrelevant data were removed using the stop-word dictionary, and punctuation was also removed.

Feature Engineering and Selection

The process of selecting the most relevant and informative features from unprocessed data is known as feature engineering, and it increases the machine learning models' effectiveness [25]. To do this, raw data must be transformed into more significant features that can effectively identify underlying patterns and relationships. This approach makes use of a number of techniques, including feature extraction, feature transformation, and feature selection. Model performance can be greatly influenced by feature engineering through careful feature creation and selection. The data in this study were primarily textual, similar to behavioral analysis of internet users, with the objective of identifying COVID-19–induced suicidal tendencies in online users. To input the data that were collected into the deep learning models, 3 basic word-embedding

XSL•FO
RenderX

techniques—TF-IDF [26-28], Word2vec [29], and BERT [30,31]—were used to identify and analyze patterns in the text-based content. Our model leveraged BERT-based architectures. These architectures are designed to capture contextual relationships between words by processing the entire sequence of words simultaneously rather than in isolation. This allows the model to understand the meaning of a word based on its surrounding context, and it significantly improves the accuracy of detecting sarcasm, idiomatic expressions, or shifts in sentiment in languages. The post length is also considered as one of the features. The extracted features were used as input to the CNN and LSTM deep learning models. When used with the CNN and LSTM models, these feature selection techniques improve performance. As mentioned before, CNNs excel at capturing localized patterns in sequential data by making them well suited for text classification, LSTM networks capture long-term dependencies and are effective for modeling sequential data with complex temporal dynamics, and BERT captures contextual relationships between words by processing the entire sequence of words simultaneously.

Suicide Posts' Classification

Different classifiers are used for classifying posts into binary classes. In this work, we used multiple combinations of classifiers to predict suicidal posts. The data set we used to train our model consisted of 2 columns, text and label. The problem was formulated the same way as that by Hossain et al [32] and Ji et al [33]. On a corpus consisting of a set of posts  and labels  , training was provided in such a manner that the model learned from the data consisting of a set of all the engineered features and the corresponding labels that were provided in a supervised approach. The supervisory function is shown in Equation 1:

$$l_i = \text{fun}(p_i) \quad (1)$$

Here, $l_i = 1$ in the case of p_i representing suicide and $l_i = 0$ represents nonsuicide. We focused on the "no free lunch" theorem of machine learning, which suggests that no algorithm can work well for all problems. Three deep learning classification algorithms were implemented: CNN, LSTM, and BERT. These models were fused, as shown in Equation 2:

$$h_c = \text{CNN}(X),$$

$$h_l = \text{LSTM}(X),$$

$$h_b = \text{BERT}(X), \quad (2)$$

$$h_m = \text{concatenate}(h_c, h_l, h_b),$$

$$y = \text{softmax}(Wh_m + b),$$

where X is the input data and is represented as a matrix, where each row corresponds to a data sample and each column corresponds to a feature; h_c refers to feature representation by the CNN; h_l refers to feature representation by the LSTM; h_b refers to feature representation by BERT; h_m refers to fused features obtained by concatenating h_c, h_l, and h_b; y is the predicted class probabilities for suicidal posts; and W is the weight of fused features h_m; and b is the bias of h_m.

The pseudocode of the entire experiment is shown in the algorithm in Textbox 1. This algorithm outlines the steps for identifying suicidal posts in social media content during the COVID-19 pandemic. The process began with a set of posts and classified them into suicidal and nonsuicidal posts. The algorithm required input posts and ensured the output includes suicidal and nonsuicidal posts. For each post, the algorithm performed EA and TA. Posts were classified based on whether they exhibited sad or negative emotions, alongside a toxicity score above a threshold (0.5). If a post met both criteria, it was added to the final corpus, otherwise it was discarded. The algorithm then processed the remaining posts through tokenization, stop-word removal, and stemming, saving the preprocessed data in a CSV file. Features were extracted using BERT embeddings and TF-IDF, resulting in a document feature matrix. The matrix was trimmed based on minimum document frequencies and TFs to focus on significant terms. Finally, deep learning algorithms and transformers with specific hyperparameters were applied to classify the posts as either suicidal or nonsuicidal. This approach leverages deep learning techniques and NLP to effectively identify suicidal ideation in the context of COVID-19 and ultimately enhances mental health awareness and intervention strategies.

XSL•FO
RenderX

Textbox 1. Pseudocode: deep learning–based suicide behavior identification during COVID-19.

REQUIRE: Posts P

ENSURE: Suicidal Post P_s and Non-Suicidal Post P_n

Tokenization → T, StopWordRemoval → SW, Stemming → S, Total Number of Tweets → n

Emotion Analysis → EA , Toxicity Analysis → TA

Term Frequency–Inverse Document Frequency → TF-IDF, Word2vec → w, BERT → B

START

1 ***For****i from* 1 *to* n

2 Perform EA(P[i])

3 Perform TA(P[i])

4 if (EA(P[i]==Sad/Negative && TA(P[i]>0.5))then

5 C[i]=P[i] // final corpus

6 **Else**

7 Discard

8 ***End For***

9 C[i]=T(C[i])

10 C[i] = SW(C[i])

11 C[i]= S(C[i])

12 Processed.csv= C[i] //Saving Preprocessed file as CSV

13 Features B = B(Processed)

14 Features T=TF/IDF(Processed)

15 Document Feature Matrix=Document Feature Matrix(Features T)

16 Trimmed_dfm=dfm_trim(dfm,min_docfreq,min_termfreq)

17 Final_Features T=Trimmed_dfm.Tfidf //Extracting TFIDF Features

18 Feature w=w(Processed)

19 ***CLASSIFIER*** (Classifier_Name, Hyperparameters, Features)

END

Explainable Artificial Intelligence

XAI holds a central role in mental health by improving the transparency and accountability of artificial intelligence (AI) systems [34,35]. In mental health applications, it is imperative that health care professionals and patients alike grasp and have confidence in the AI's decision-making procedures [36]. XAI accomplishes this by furnishing explanations for AI-generated diagnoses, forecasts, and treatment recommendations. These explanations provide valuable insights into the rationale behind specific decisions, assisting clinicians and patients in gaining a deeper understanding of the logic behind mental health assessments. This transparency fosters trust and promotes collaboration between AI systems and human experts, which ultimately enhances the quality of mental health care. In addition, XAI aids in detecting and reducing biases within AI models, which ensures that mental health assessments remain impartial and unbiased, a critical aspect of delivering equitable care to diverse patient populations. XAI also contributes to the formulation of individualized treatment plans in the mental health domain. By elucidating factors influencing an individual's mental health condition, AI systems can customize interventions to cater to each patient's unique requirements. XAI also facilitates early intervention and prevention by enabling the identification of early warning signs and comprehending the interconnected elements affecting mental well-being. Empowering patients through understandable explanations increases their engagement in the treatment process, potentially leading to better outcomes. XAI serves not only as a tool for enhancing mental health care but also as a means of diminishing stigma by promoting ethical practices and propelling the field of mental health forward through data-driven insights and improved decision-making. XAI techniques are designed to make the decision-making processes of AI systems more understandable and interpretable for humans. These techniques are particularly important when AI systems make complex predictions, classifications, or recommendations in fields such as health care, finance, and autonomous vehicles, where transparency and trust are critical.

LIME generates perturbed instances (z_1, z_2,…, z_k) and performs predictions: $f(z_1)$, $f(z_2)$,…, $f(z_k)$. The local linear model  . The objective function of LIME is calculated using Equation 3:

XSL•FO
RenderX



Here, x represents the instance for which we want an explanation, z_i are perturbed instances, $\pi_x(x,z_i)$ is a weight, and  is the gradient of f at x_0.

Shapley XAI has the following workings:

The Shapley value for a feature i in the context of a prediction f(x) is shown in Equation 4:



Here, N is the set of all features and S is a subset of features excluding i.

Expected Shapley values are represented in Equation 5:



Here, N is the number of samples from the background distribution.

Ethical Considerations

This research did not include any studies on human participants performed by any of the authors. However, ethical considerations of our study could be related to the use of data collected on posts on the Reddit platform, which is commonly perceived as widely public, where participants are not identifiable as no registration is required but only a pseudo name to post. Moreover, public post analysis, such as Reddit posts, is considered ethical issue free because there is no expectation of privacy as individuals are anonymous, and users understand that their posts can be reused (ie, reposted) and remain archived on the platform until deletion eventually. Hence, ethical issues of research on Reddit are resolved by the public nature of this social media platform, especially when used in academic publications.

Results

Study Details

Of 348,110 records, 3154 (0.9%) were filtered out, resulting in 1338 (42.4%) suicidal and 1816 (57.6%) nonsuicidal posts. The training data in all experiments were split in an 80:20 ratio for training and validation because empirical findings [37] suggest that the performance of algorithms improves when the data set is divided in this manner. Each algorithm was tested on a distinct subset of the original data set. Our model was evaluated using various metrics to assess their performance. We discovered that the model's performance improved slightly as the number of training data points was raised. The method of dividing the data set into training and testing subsets has inherent problems with bias and variance. These problems arise because of the nature of the data. In machine learning, it is not necessary that the model that has been fit on training data will also work on real data. This is because real data can be different from training data. To accomplish this, we made use of the technique of k-fold cross-validation. This allowed us to ensure that the model correctly identified the patterns within the data, while minimizing noise exposure. Our model's validity was established by a process known as 10-fold cross-validation. Python was chosen as the language to implement algorithms for machine learning. Skicit-learn, Pandas, and NLTK are the most significant packages used in the coding process during the training of the suicide prediction model. After evaluating hybrid deep learning models, we found that BERT, when fused with CNN+LSTM, showed the best results among all hybrid deep learning models. With a precision of 94%, a recall of 95%, an F_1-score of 94%, and an overall accuracy of 93.65%. Figure 2 shows the hybrid model's performance based on the model loss and model accuracy. Table 1 depicts the classification report of all the classifiers that were used during experimentation.

Figure 2. Performance metrics showing the accuracy and loss of the best model, which combined BERT, CNN, and LSTM for enhanced prediction results. BERT: Bidirectional Encoder Representations from Transformers; CNN: convolutional neural network; LSTM: long short-term memory.

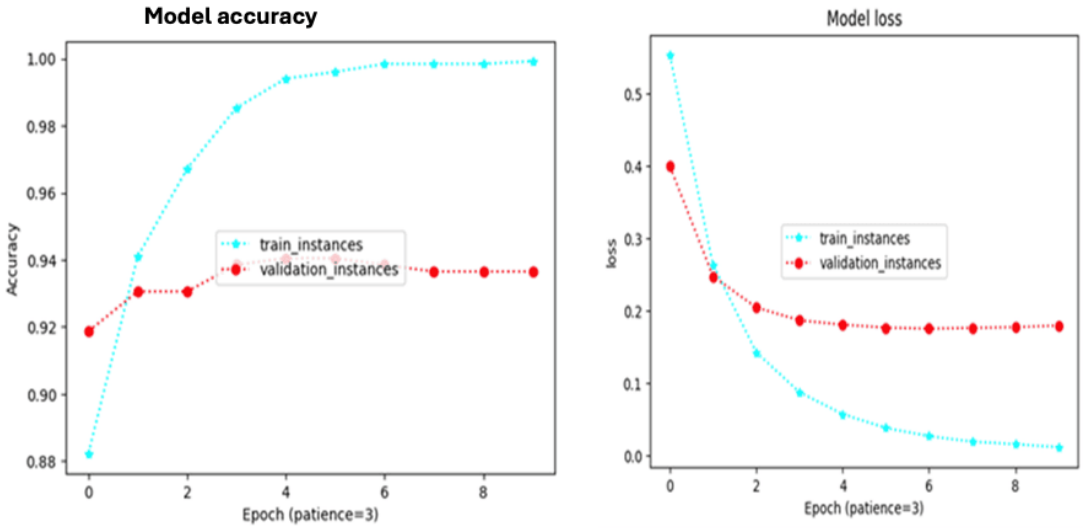

XSL•FO
RenderX

Table 1. Classification report of hybrid models based on various evaluation metrics.

Classifier	Precision (%)	Recall (%)	F_1-score (%)	Accuracy (%)
TF-IDF[a]+CNN[b]	76.00	80.00	78.00	78.00
Word2vec+LSTM[c]	88.00	94.00	91.00	88.00
BERT[d]+CNN	69.00	66.00	67.00	64.00
LSTM+BERT	86.00	83.00	85.00	85.00
CNN+LSTM	91.00	90.00	91.00	91.00
BERT+CNN+LSTM	94.00	95.00	94.00	93.65
CNN+LSTM+TFIDF	89.00	89.00	89.00	89.00
CNN+LSTM+Word2vec	90.00	90.00	90.00	90.00
CNN+Word2vec	86.00	86.00	86.00	86.00

[a]TF-IDF: term frequency–inverse document frequency.

[b]CNN: convolutional neural network.

[c]LSTM: long short-term memory.

[d]BERT: Bidirectional Encoder Representations from Transformers.

Validation

In this study, we used the 10-fold cross-validation technique to validate our proposed approach. It is a robust technique frequently used in machine learning and statistical modeling to evaluate the performance and generalization capability of predictive models. In this approach, the data set is divided into 10 equal-size segments, known as folds. Throughout each iteration, 1 fold is designated as the validation set, while the remaining 9 folds serve as the training set. This process is iterated 10 times, with each fold taking on the role of the validation set once. The 10-fold cross-validation technique has the ability to furnish a more dependable assessment of a model's performance in contrast to a single train-test split. By computing performance metrics across various folds and subsequently averaging them, the ambiguity in outcomes is reduced, which leads to a more consistent evaluation. In addition, it helps in dealing with bias and variance. It guarantees that every data point contributes to both training and validation processes, thereby lessening the risk of overfitting and furnishing a more precise evaluation of the model's capacity to generalize to new data. In this work, after performing 10-fold cross-validation, the mean accuracy was compared with the actual accuracy of the models, and it was found that there was no such difference. We concluded that the proposed approach does not suffer from over- or underfitting. The results of 10-fold cross-validation are shown in Table 2 and visually represented in Figure 3.

Table 2. Results of 10-fold cross-validation, along with the calculated mean accuracy for model evaluation.

Classifier	Accuracy (%)										Accuracy (%), mean (SD)
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	
BERT[a]+CNN[b]+LSTM[c]	94.00	93.00	94.00	94.00	93.00	94.00	94.00	93.00	94.00	94.00	93.65 (0.41)
Word2vec+LSTM	88.00	88.00	89.00	90.00	87.00	85.00	88.00	89.00	90.00	88.00	88.20 (1.58)
BERT+CNN	66.00	64.00	63.00	65.00	66.00	65.00	64.00	66.00	65.00	64.00	65.00 (1.00)
LSTM+BERT	85.00	86.00	84.00	86.00	83.00	85.00	85.00	85.00	86.00	85.00	85.00 (0.95)
CNN+LSTM	90.00	91.00	89.00	90.00	91.00	90.00	91.00	90.00	91.00	89.00	90.20 (0.82)
TF-IDF[d]+CNN	78.00	76.00	77.00	79.00	76.00	78.00	79.00	78.00	76.00	79.00	77.60 (1.10)
CNN+LSTM+TF-IDF	89.00	89.00	90.00	88.00	90.00	89.00	91.00	89.00	90.00	91.00	89.60 (0.82)
CNN+LSTM+Word2vec	90.00	91.00	88.00	89.00	90.00	90.00	90.00	90.00	90.00	90.00	89.80 (0.82)
CNN+Word2vec	86.00	86.00	84.00	86.00	86.00	87.00	88.00	86.00	86.00	88.00	86.70 (1.10)

[a]BERT: Bidirectional Encoder Representations from Transformers.

[b]CNN: convolutional neural network.

[c]LSTM: long short-term memory.

[d]TF-IDF: term frequency–inverse document frequency.

XSL•FO
RenderX

Figure 3. The 10-fold cross-validation process for evaluating model performance through systematic partitioning of the data set. BERT: Bidirectional Encoder Representations from Transformers; CNN: convolutional neural network; LSTM: long short-term memory; TF-IDF: term frequency–inverse document frequency.

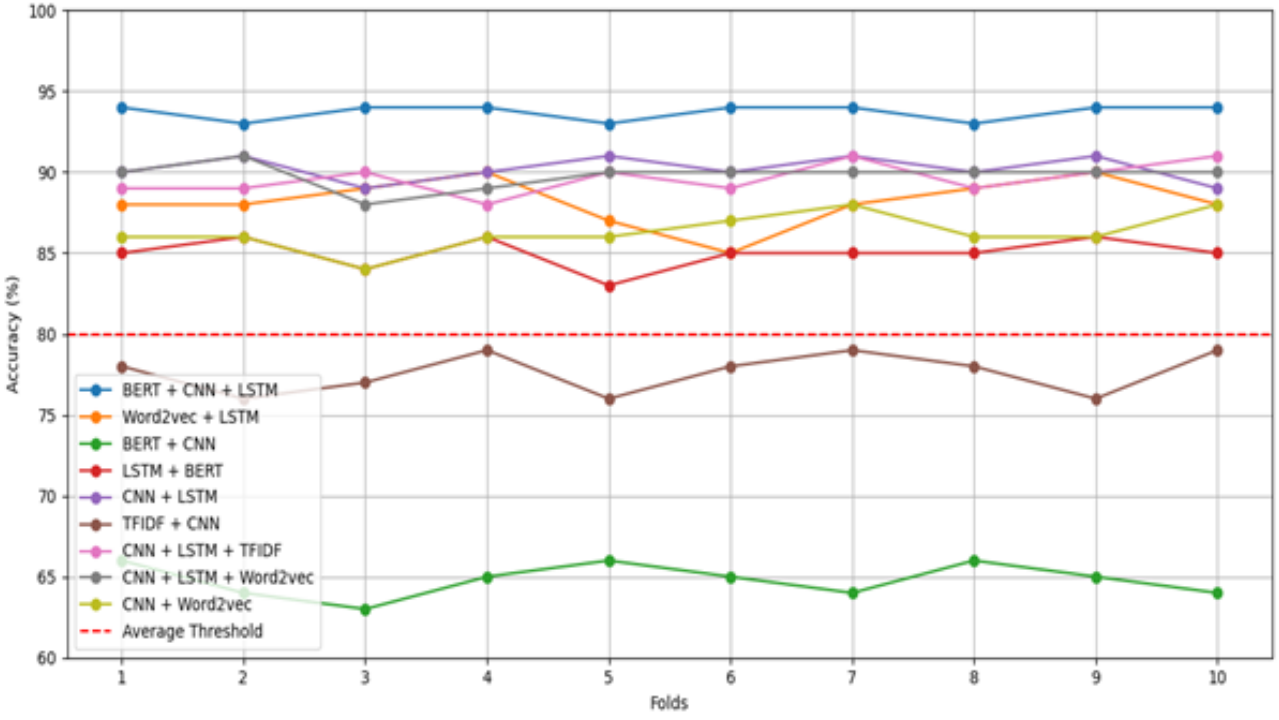

Comparative Analysis of Posts Before and During COVID-19

Exploratory data analysis of both training and validation data sets was performed, and results showed that n-grams used before and during COVID-19 were different. Figure 4 presents a comparative analysis of the n-grams used in suicidal posts during both periods. Figure 4a illustrates the most frequently occurring n-grams in posts related to suicidal ideation during the COVID-19 pandemic, which highlights key phrases and terms that reflect the mental health challenges faced during the crisis, while Figure 4b depicts the n-gram sequences identified in suicidal posts from the non–COVID-19 era.

Figure 4. Top n-gram sequences distinguishing COVID-19–related and non–COVID-19–related posts. A) Top n-gram sequence in the COVID-19 suicidal posts; B) Top n-gram sequence in the non COVID-19 suicidal posts.

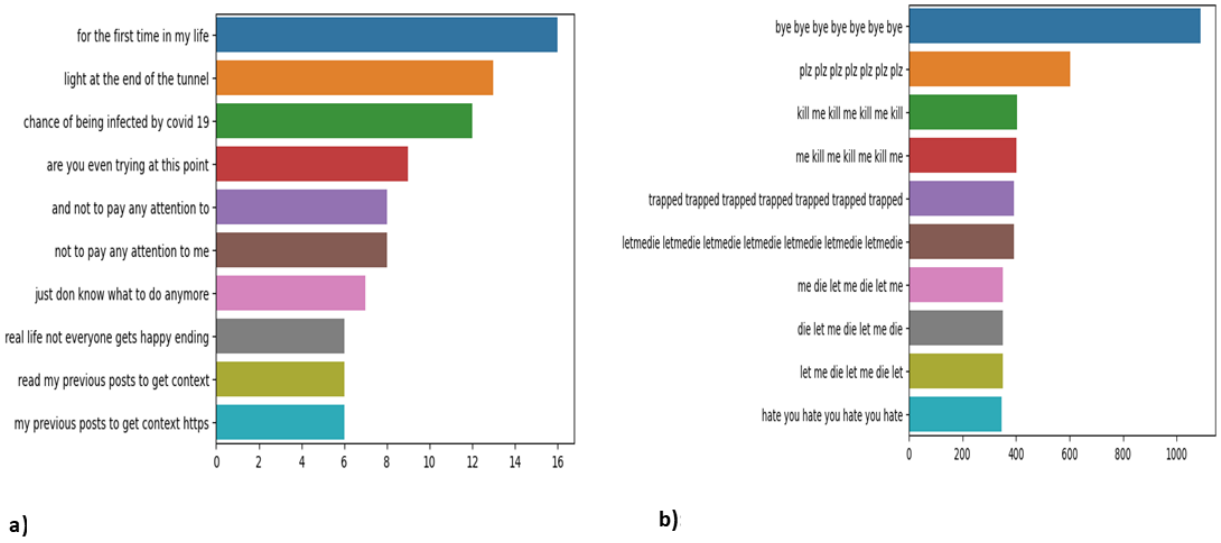

a)

b)

Explainable Artificial Intelligence Extraction

XAI showed the most important features (terms) that were used in the suicidal content during COVID-19. Figure 5 shows a graphical representation of the features extracted by XAI:

- "covid," "covid-19": These terms indicate a direct association with the COVID-19 pandemic. Individuals may express distress, anxiety, or hopelessness related to the pandemic, which can contribute to suicidal ideation.
- "live, life": These terms are broad and could encompass various aspects of an individual's life, such as personal

XSL•FO
RenderX

relationships, work, health, or the financial situation. Expressions related to feeling overwhelmed, hopeless, or lacking purpose in life may contribute to suicidal thoughts.
- "die," "stop," "myself": These words directly relate to thoughts of self-harm or suicide. They indicate a strong presence of negative thoughts or intentions toward ending one's life.

XAI was used to check the impact of COVID-19 on suicidal ideation, and it can be clearly seen from the results that COVID-19–related words were used frequently in suicidal thoughts.

Figure 5. Important features (terms) associated with suicidal posts and their relevance in predictive analysis. SHAP: Shapley Additive Explanations.

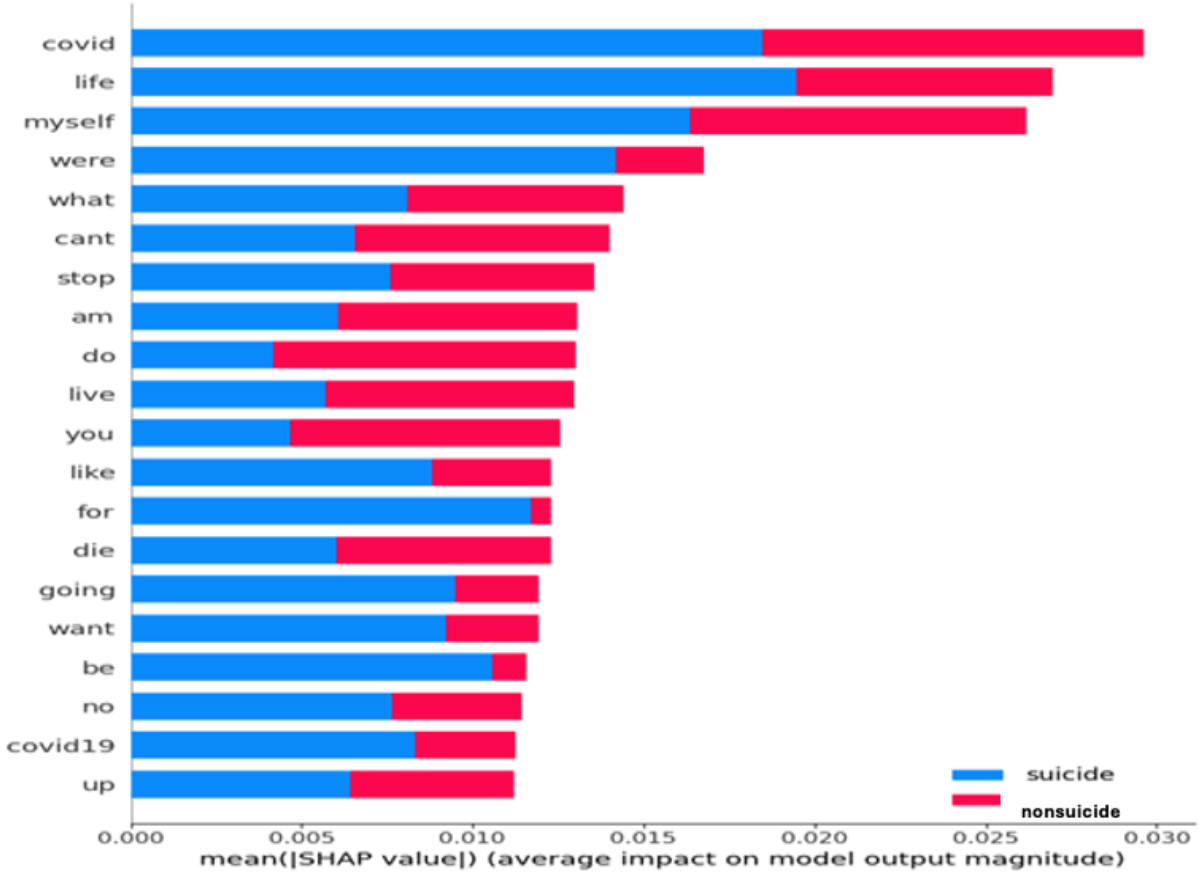

Discussion

Principal Findings

The results of this study indicate significant changes in language patterns before and during the COVID-19 pandemic. Individuals used different sets of words prior to the pandemic, and the sequence of words shifted during the pandemic. In this work, we investigated the impact of the COVID-19 pandemic on suicidal ideation by proposing a null hypothesis (H_0) that COVID-19 has not influenced levels of suicidal thoughts among individuals. By using the proposed methodology, our findings revealed a significant increase in suicidal ideation associated with the pandemic. The analysis showed a P value of <.01, which leads us to reject H_0 in favor of the alternative hypothesis (H_1), which posited that COVID-19 has had a measurable impact on suicidal ideation. These results underscore the urgent need for mental health interventions during and after public health crises, highlighting the critical link between worldwide events and individual psychological well-being.

Comparative Analysis

For validating the approach, we performed comparative analysis in which we compared existing methods with our approach. The methodology described by Jung et al [38] showed that Bow-Char had a maximum recall score of 0.75, closely trailed TF-IDF (SVM) with a recall of 0.70. In our approach, the recall scores ranged from 0.83 to 0.95, which are higher than existing ones. The results demonstrated that the BERT+CNN+LSTM model had the highest recall, proving its efficacy in identifying positive instances. In addition, ADA achieved the higher precision of 0.92 compared to other approaches that were proposed by Jung et al [38], while our approach achieved precision values from 0.86 to 0.95. The BERT+CNN+LSTM model showed that it can minimize false positives by attaining the maximum level of accuracy. The range of accuracy attained by Jung et al [38] was 65%-77%. Our proposed models, in contrast, outperformed them with accuracy scores between 0.85 and 0.9365. The best accuracy was shown by the BERT+CNN+LSTM model, demonstrating its capacity to classify most cases properly. According to Saba et al [39], a recurrent neural network (RNN) performs better due to its high precision value (1), while CNN models achieve good recall

XSL•FO
RenderX

values of 0.8 and CNN+SVM models achieve 91.6% accuracy. Our approach outperformed previous methods by achieving greater recall, precision, and accuracy values; the BERT+CNN+LSTM model achieved 93.65% accuracy. These findings show that the proposed method produces more accurate and reliable results.

When F_1-scores were examined, ADA had the highest score of 0.73 among the techniques examined by Jung et al [38], while the RNN had an F_1-score of 0.76 and the CNN and CNN+SVM models had the highest F_1-scores of 0.85 and 0.90 according to Saba et al [39]. Our approach outperformed in F_1-scores as well by achieving values ranging from 0.85 to 0.94. The best F_1-score was attained by the BERT+CNN+LSTM model, demonstrating its balanced precision and recall performance. The TF-IDF+CNN model outperformed Jung et al's [38] best method in terms of recall by about 20%. Comparably, the CNN+LSTM and CNN+LSTM+Word2vec models outperformed Jung et al's [38] optimal approach by about 15% and 14%, respectively, in terms of recall values. These notable advancements suggest that the proposed techniques are better at identifying positive samples. The accuracy of the TF-IDF+CNN model was around 21.65% greater than that of Jung et al's [38] optimal method (TF-IDF+SVM). Similarly, accuracy gains of roughly 19% and 18% were attained by our CNN+LSTM and CNN+LSTM+Word2vec models, respectively. These improvements show that the proposed approaches have stronger categorization skills and produce results that are more accurate. Hence, we can conclude that the recall, precision, accuracy, and F_1-score of the proposed methods—TF-IDF+CNN, BERT+LSTM, CNN+LSTM, and Word2vec+CNN+LSTM—are better than existing techniques. The enhancements in performance range from 14% to 21%, highlighting the efficacy of using sophisticated methods, such as TF-IDF, CNNs, LSTM, and BERT, when examining content.

Limitations

We analyzed the strengths and weaknesses of each model, and it was found that the CNN effectively captured localized patterns in social media posts, such as emotional language and contextual cues indicative of suicidal ideation. Phrases such as "I feel hopeless" and "No one cares" were accurately classified as indicating suicidal ideation. However, the model struggled with understanding context, particularly when the meaning of a post depended heavily on the sequence of words rather than their individual presence. For example, the phrase "I just need a break" was incorrectly classified as nonsuicidal. In contrast, the LSTM processed sequences that made it suitable for understanding the context and flow of language over time. For example, posts beginning with "Things have been tough lately" and continued with "I can't see a way out" were correctly identified by the LSTM as having an evolving sentiment of despair and resulted in the correct classification of suicidal ideation. However, there is a risk of overfitting to the training data, especially if the training set is small or not diverse enough. Although the LSTM captured the emotional tone, it may have overfit to similar phrases in the training set that did not indicate suicidal ideation and resulted in false negatives. For example, "I can't handle this anymore, I'm just tired of fighting" was misclassified as nonsuicidal. BERT processes text in both directions simultaneously. This allows it to capture the full context of a word based on its surroundings, leading to better understanding and representation of meaning, but it benefits from being trained on large data sets. Since we had less data, we integrated the three algorithms (BERT, CNN, and LSTM) in our hybrid model, which helped us mitigate the individual weaknesses of each model. The CNN effectively extracted features from social media posts, the LSTM provided a contextual understanding of the text, and BERT considered the entire context of each word, which effectively disambiguated words with multiple meanings. Combining these architectures allowed us to capture both localized and generalized contextual information that is crucial for understanding the language patterns associated with suicidal behavior. By fusing these 3, our model achieved an accuracy of 93.65%.

Implications of the Study

The findings of this study reveal that the COVID-19 pandemic has influenced mental health. Due to the shift in language patterns, there is an increase in terms related to distress, hopelessness, and self-harm, which underscores the profound psychological impact of the pandemic. A clearer understanding of linguistic markers that may indicate heightened risk can be identified using XAI. These linguistic patterns have broader implications for early detection and intervention strategies for suicidal ideation. We suggest that language-based monitoring could serve as a tool for identifying individuals at risk. In addition, the study emphasizes the need for targeted mental health support during worldwide crises because the psychological effects of such events can lead to a lasting transformation in how individuals express their struggles. The findings suggest a critical shift toward more personalized language-based approaches to mental health care that can adapt to the unique challenges posed by worldwide events, such as the COVID-19 pandemic.

Conclusion

The worldwide impact of COVID-19 has had long-lasting effects on people's mental well-being. In this study, we collected data from Reddit posts related to COVID-19. Data were filtered using specific keywords associated with the pandemic. Through EA and TA, we understood the emotional tone and potential harm in these posts. To classify the data, hybrid deep learning classifiers were used, which fused BERT, CNN, and LSTM models. The results reveal that our BERT+CNN+LSTM model achieved the highest accuracy of 93.65%, with a precision of 94% and a recall of 95%. In the final stage, XAI techniques, such as LIME and SHAP, extracted important factors for suicidal and nonsuicidal sentiments. After integration, the fused model can effectively grasp both localized and generalized contextual characteristics that are essential for comprehending linguistic patterns linked with suicidal ideation. It manages diverse formats commonly encountered in real-world text data by accommodating varying sentence lengths and syntactic complexities. This model will equip decision makers with a deeper understanding of factors contributing to suicide risk, facilitating the development of more impactful intervention strategies. Our analysis showed that terms related to COVID-19

XSL•FO
RenderX

are often present in posts with suicidal tendencies. We plan to incorporate audio data alongside these features to enhance the prediction of suicidal ideation.

Data Availability

The data used in this study are available upon request from the corresponding author.

Authors' Contributions

SB and AMUDK conceptualized the study and implemented the work. AO performed the literature review and readability analyses. AMUDK drafted the first version of the manuscript. All the authors have revised and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed literature.
[DOCX File , 18 KB - jmir_v27i1e65434_app1.docx]

Multimedia Appendix 2
Background knowledge of various methods used.
[DOCX File , 19 KB - jmir_v27i1e65434_app2.docx]

References

1. Suicide prevention. World Health Organization. URL: https://www.who.int/health-topics/suicide#tab=tab_1 [accessed 2024-12-28]
2. Bilsen J. Suicide and youth: risk factors. Front Psychiatry 2018;9:540 [FREE Full text] [doi: 10.3389/fpsyt.2018.00540] [Medline: 30425663]
3. Värnik P. Suicide in the world. Int J Environ Res Public Health 2012 Mar 02;9(3):760-771 [FREE Full text] [doi: 10.3390/ijerph9030760] [Medline: 22690161]
4. Hawton K, van Heeringen K. The International Handbook of Suicide and Attempted Suicide (1st ed). Hoboken, NJ: Wiley; 2002.
5. Shioiri T, Nishimura A, Akazawa K, Abe R, Nushida H, Ueno Y, et al. Incidence of note-leaving remains constant despite increasing suicide rates. Psychiatry Clin Neurosci 2005 Apr 24;59(2):226-228 [FREE Full text] [doi: 10.1111/j.1440-1819.2005.01364.x] [Medline: 15823174]
6. Foster T. Suicide note themes and suicide prevention. Int J Psychiatry Med 2003 Dec 01;33(4):323-331. [doi: 10.2190/t210-e2v5-a5m0-qlju]
7. Fu K, Cheng Q, Wong PWC, Yip PSF. Responses to a self-presented suicide attempt in social media: a social network analysis. Crisis 2013 Jan 01;34(6):406-412 [FREE Full text] [doi: 10.1027/0227-5910/a000221] [Medline: 23871954]
8. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014 Jan 01;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
9. Ahuja A, Biesaga K, Sudak D, Draper J, Womble A. Suicide on Facebook. J Psychiatr Pract 2014 Mar;20(2):141-146. [doi: 10.1097/01.pra.0000445249.38801.d1] [Medline: 24638049]
10. Khanday AMUD, Rabani ST, Khan QR, Rouf N, Mohi Ud Din M. Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol 2020 Jun 30;12(3):731-739 [FREE Full text] [doi: 10.1007/s41870-020-00495-9] [Medline: 32838125]
11. Home. American Foundation for Suicide Prevention. URL: https://afsp.org/ [accessed 2024-12-30]
12. Suicide Prevention Now. URL: https://suicidepreventionnow.org/ [accessed 2024-12-30]
13. Cash SJ, Thelwall M, Peck SN, Ferrell JZ, Bridge JA. Adolescent suicide statements on MySpace. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):166-174. [doi: 10.1089/cyber.2012.0098] [Medline: 23374167]
14. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. Crisis 2017 Sep;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]
15. Harris KM, McLean JP, Sheffield J. Suicidal and online: how do online behaviors inform us of this high-risk population? Death Stud 2014 Oct 18;38(6-10):387-394. [doi: 10.1080/07481187.2013.768313] [Medline: 24666145]
16. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2021 Presented at: ICWSM-13: Seventh International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, MA p. 128-137. [doi: 10.1609/icwsm.v7i1.14432]

XSL•FO
RenderX

17. Colombo GB, Burnap P, Hodorog A, Scourfield J. Analysing the connectivity and communication of suicidal users on Twitter. Comput Commun 2016 Jan 01;73(Pt B):291-300 [FREE Full text] [doi: 10.1016/j.comcom.2015.07.018] [Medline: 26973360]
18. Huang X, Zhang L, Chiu D, Liu T, Li X, Zhu T. Detecting suicidal ideation in Chinese microblogs with psychological lexicons. 2014 Presented at: IEEE 11th International Conference on Ubiquitous Intelligence and Computing, IEEE 11th International Conference on Autonomic and Trusted Computing, and IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops; December 9-12, 2014; Bali, Indonesia. [doi: 10.1109/uic-atc-scalcom.2014.48]
19. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interven 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]
20. Vioules MJ, Moulahi B, Aze J, Bringay S. Detection of suicide-related posts in Twitter data streams. IBM J Res Dev 2018 Jan 1;62(1):7:1-7:12. [doi: 10.1147/jrd.2017.2768678]
21. Noureen A, Qamar U, Ali M. Semantic analysis of social media and associated psychotic behavior. 2017 Presented at: IEEE 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD); July 29-31, 2017; Guilin, China. [doi: 10.1109/fskd.2017.8393009]
22. Chiroma F, Liu H, Cocea M. Text classification for suicide-related tweets. 2018 Presented at: IEEE 2018 International Conference on Machine Learning and Cybernetics (ICMLC); July 15-18, 2018; Chengdu, China. [doi: 10.1109/icmlc.2018.8527039]
23. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media 2017 Aug;2:32-44 [FREE Full text] [doi: 10.1016/j.osnem.2017.08.001] [Medline: 29278258]
24. Moreno MA, Jelenchick LA, Egan KG, Cox E, Young H, Gannon KE, et al. Feeling bad on Facebook: depression disclosures by college students on a social networking site. Depress Anxiety 2011 Jun 11;28(6):447-455 [FREE Full text] [doi: 10.1002/da.20805] [Medline: 21400639]
25. Chadha A, Kaushik B. A survey on prediction of suicidal ideation using machine and ensemble learning. Comput J 2021;64(11):1617-1632. [doi: 10.1093/comjnl/bxz120]
26. Sawhney R, Manchanda P, Singh R, Aggarwal S. A computational approach to feature extraction for identification of suicidal ideation in tweets. 2018 Presented at: ACL 2018, Student Research Workshop; July 15-20, 2018; Melbourne, Australia p. 91-98. [doi: 10.18653/v1/p18-3013]
27. Deng X, Li Y, Weng J, Zhang J. Feature selection for text classification: a review. Multimed Tools Appl 2018 May 8;78(3):3797-3816. [doi: 10.1007/s11042-018-6083-5]
28. Qu S, Wang S, Zou Y. Improvement of text feature selection method based on TF-IDF. 2008 Presented at: IEEE 2008 International Seminar on Future Information Technology and Management Engineering; November 20, 2008; Leicestershire, UK p. 79-81. [doi: 10.1109/fitme.2008.25]
29. Ma L, Zhang Y. Using Word2Vec to process big text data. 2015 Presented at: IEEE 2015 International Conference on Big Data; October 29-November 1, 2015; Santa Clara, CA. [doi: 10.1109/bigdata.2015.7364114]
30. Agustian S, Saputra R, Fadhilah A. Feature selection with pretrained-BERT for hate speech and offensive content identification in English and Hindi languages. 2021 Presented at: Forum for Information Retrieval Evaluation; December 13-17, 2021; Gandhinagar, Gujarat, India p. 508-516. [doi: 10.13140/RG.2.2.27831.60324]
31. Hossain MR, Hoque MM, Siddique N, Dewan MAA. AraCovTexFinder: leveraging the transformer-based language model for Arabic COVID-19 text identification. Eng Appl Arti Intell 2024 Jul;133:107987. [doi: 10.1016/j.engappai.2024.107987]
32. Hossain MR, Hoque MM, Siddique N, Sarker IH. CovTiNet: Covid text identification network using attention-based positional embedding feature fusion. Neural Comput Appl 2023 Mar 14;35(18):13503-13527 [FREE Full text] [doi: 10.1007/s00521-023-08442-y] [Medline: 37213320]
33. Ji S, Yu CP, Fung S, Pan S, Long G. Supervised learning for suicidal ideation detection in online user content. Complexity 2018 Sep 09;2018(1):6157249. [doi: 10.1155/2018/6157249]
34. Xie P, Wang H, Xiao J, Xu F, Liu J, Chen Z, et al. Development and validation of an explainable deep learning model to predict in-hospital mortality for patients with acute myocardial infarction: algorithm development and validation study. J Med Internet Res 2024 May 10;26:e49848 [FREE Full text] [doi: 10.2196/49848] [Medline: 38728685]
35. Chen B, Li Y, Sun Y, Sun H, Wang Y, Lyu J, et al. A 3D and explainable artificial intelligence model for evaluation of chronic otitis media based on temporal bone computed tomography: model development, validation, and clinical application. J Med Internet Res 2024 Aug 08;26:e51706 [FREE Full text] [doi: 10.2196/51706] [Medline: 39116439]
36. Le Glaz A, Haralambous Y, Kim-Dufor D, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 04;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
37. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation. University of Texas at El Paso. 2018. URL: https://www.cs.utep.edu/vladik/2018/tr18-09.pdf [accessed 2024-12-30]

XSL•FO
RenderX

38. Jung JS, Park SJ, Kim EY, Na K, Kim YJ, Kim KG. Prediction models for high risk of suicide in Korean adolescents using machine learning techniques. PLoS One 2019 Jun 6;14(6):e0217639 [FREE Full text] [doi: 10.1371/journal.pone.0217639] [Medline: 31170212]
39. Saba T, Khan AR, Abunadi I, Bahaj SA, Ali H, Alruwaythi M. Arabic speech analysis for classification and prediction of mental illness due to depression using deep learning. Comput Intell Neurosci 2022 May 27;2022:8622022-8622029 [FREE Full text] [doi: 10.1155/2022/8622022] [Medline: 35669665]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
CNN: convolutional neural network
EA: emotion analysis
LIME: Local Interpretable Model-Agnostic Explanations
LSTM: long short-term memory
NLP: natural language processing
NLTK: Natural Language Toolkit
SHAP: Shapley Additive Explanations
SVM: support vector machine
TA: toxicity analysis
TF-IDF: term frequency–inverse document frequency
XAI: explainable artificial intelligence

Edited by A Mavragani; submitted 15.08.24; peer-reviewed by A Adil, MR Hossain; comments to author 27.09.24; revised version received 17.10.24; accepted 05.12.24; published 17.01.25.

Please cite as:
Bouktif S, Khanday AMUD, Ouni A
Explainable Predictive Model for Suicidal Ideation During COVID-19: Social Media Discourse Study
J Med Internet Res 2025;27:e65434
URL: https://www.jmir.org/2025/1/e65434
doi:10.2196/65434
PMID:

©Salah Bouktif, Akib Mohi Ud Din Khanday, Ali Ouni. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigating Smartphone-Based Sensing Features for Depression Severity Prediction: Observation Study

Yannik Terhorst[1,2,3], MSc; Eva-Maria Messner[1], PhD; Kennedy Opoku Asare[4], PhD; Christian Montag[5], Prof Dr; Christopher Kannen[5], MSc; Harald Baumeister[1], Prof Dr

[1]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, Ulm University, Ulm, Germany

[2]Department of Psychology, LMU Munich, Munich, Germany

[3]German Center for Mental Health (DZPG), Partner Site Munich-Augsburg, Munich, Germany

[4]Center for Ubiquitous Computing, University of Oulu, Oulu, Finland

[5]Department of Molecular Psychology, Institute of Psychology and Education, Ulm University, Ulm, Germany

Corresponding Author:
Yannik Terhorst, MSc
Department of Clinical Psychology and Psychotherapy
Institute of Psychology and Education
Ulm University
Lise-Meitner-Str. 16
Ulm, 89081
Germany
Phone: 49 8921805057
Email: yannik.terhorst@psy.lmu.de

Abstract

Background: Unobtrusively collected objective sensor data from everyday devices like smartphones provide a novel paradigm to infer mental health symptoms. This process, called smart sensing, allows a fine-grained assessment of various features (eg, time spent at home based on the GPS sensor). Based on its prevalence and impact, depression is a promising target for smart sensing. However, currently, it is unclear which sensor-based features should be used in depression severity prediction and if they hold an incremental benefit over established fine-grained assessments like the ecological momentary assessment (EMA).

Objective: The aim of this study was to investigate various features based on the smartphone screen, app usage, and call sensor alongside EMA to infer depression severity. Bivariate, cluster-wise, and cluster-combined analyses were conducted to determine the incremental benefit of smart sensing features compared to each other and EMA in parsimonious regression models for depression severity.

Methods: In this exploratory observational study, participants were recruited from the general population. Participants needed to be 18 years of age, provide written informed consent, and own an Android-based smartphone. Sensor data and EMA were collected via the INSIGHTS app. Depression severity was assessed using the 8-item Patient Health Questionnaire. Missing data were handled by multiple imputations. Correlation analyses were conducted for bivariate associations; stepwise linear regression analyses were used to find the best prediction models for depression severity. Models were compared by adjusted R^2. All analyses were pooled across the imputed datasets according to Rubin's rule.

Results: A total of 107 participants were included in the study. Ages ranged from 18 to 56 (mean 22.81, SD 7.32) years, and 78% of the participants identified as female. Depression severity was subclinical on average (mean 5.82, SD 4.44; Patient Health Questionnaire score ≥10: 18.7%). Small to medium correlations were found for depression severity and EMA (eg, valence: r=–0.55, 95% CI –0.67 to –0.41), and there were small correlations with sensing features (eg, screen duration: r=0.37, 95% CI 0.20 to 0.53). EMA features could explain 35.28% (95% CI 20.73% to 49.64%) of variance and sensing features (adjusted R^2=20.45%, 95% CI 7.81% to 35.59%). The best regression model contained EMA and sensing features (R^2=45.15%, 95% CI 30.39% to 58.53%).

Conclusions: Our findings underline the potential of smart sensing and EMA to infer depression severity as isolated paradigms and when combined. Although these could become important parts of clinical decision support systems for depression diagnostics and treatment in the future, confirmatory studies are needed before they can be applied to routine care. Furthermore, privacy, ethical, and acceptance issues need to be addressed.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e55308) doi:10.2196/55308

KEYWORDS

smart sensing; digital phenotyping; depression; observation study; smartphone; mHealth; mobile health; app; mental health; symptoms; assessments

Introduction

Depression is associated with high personal burden, impaired social participation and functioning, increased mortality, and high economic burden [1-3]. In 2020, depression was one of the leading causes of disability-adjusted life years worldwide (49.4 million; 95% CI 33.6-68.7) and is expected to be the leading cause by 2030 [1,3,4]. Despite its severity and the existence of effective treatments for depression [5-7], only 41.8% of people with major depressive disorder (MDD) receive any mental health services, and less than 30% of people with MDD receive adequate treatment [1,8]. Although several barriers contribute to this issue (eg, availability, accessibility, and acceptability of treatment) [1,8,9], a fundamental prerequisite to any health services is a timely and accurate diagnosis of MDD, or more generally, the assessment of depression severity to initiate an informed treatment process [10-14].

Well-established assessments like structured clinical interviews are often not feasible to be conducted in primary care or for preventive screening purposes (eg, due to time pressure or limited availability of qualified personnel) [10-13,15]. Furthermore, if they are implemented, they typically take place at a fixed time point and assess symptoms retrospectively, which makes them subject to several biases (eg, recall bias) and unable to assess the dynamic and fluctuating nature of mental health [16-18]. Hence, novel diagnostic approaches, which can be easily integrated into daily living to monitor depression severity with high ecological validity, could make an important contribution to improve and augment current diagnostic procedures for depression—particularly if they provide fine-grained insights into mental health symptomology (eg, on daily level). Given the omnipresence of smartphones in everyday life, the unobtrusive collection of objective sensor data (eg, total haversine distance between tracked GPS coordinates, social contacts, screen, and app usage) might be a promising step toward improved diagnoses [19-21]. This process is also referred to as smart sensing (also known as mobile sensing or digital phenotyping) in the context of depression [19-22]. Applications of smart sensing range from supporting initial diagnosis to integration during treatment (eg, just-in-time-adaptive interventions) or after treatment (eg, just-in-time interventions in relapse prevention) [19,23].

In the field of depression so far, many studies followed a classificatory approach classifying persons as depressed or not depressed (eg, based on the 8-item Patient Health Questionnaire [PHQ-8] cutoff ≥10) [24-28]. For instance, a first meta-analysis on supervised machine learning models to predict depression status based on wearable data reported an average accuracy of 0.89, 95% CI 0.83-0.93 (sensitivity: 0.87, 95% CI 0.79-0.92; specificity: 0.93, 95% CI 0.87-0.97) [24]. However, such a classificatory understanding of depression seems questionable when looking at (1) the poor agreement of domain experts in the diagnosis of depression [29,30], (2) the heterogeneity of symptom networks in depressed patients [31,32], and (3) the profound evidence for depression and general psychopathology being a continuous spectrum [29,33]. Hence, studies operationalizing depression as a continuous spectrum and understanding the prediction as a regression instead of a classification problem are highly needed in the field.

For example, for GPS features (eg, total distance, number of significant places), a meta-analysis shows that robust correlations between sensing features and depression severity as continuous dimension exist (eg, distance: r=–0.25, 95% CI –0.29 to –0.21), time spent home: r=0.10 (95% CI 0-0.19), or normalized entropy: r=–0.17, 95% CI –0.29 to –0.04) [34]. Besides, initial studies highlight the potential of features obtained from the screen (eg, smartphone usage duration), app (app usage), and call (eg, number of incoming calls) sensors [35,36]. However, so far, analyses are often limited to bivariate correlations and do not extend to (1) the variance in depression severity, which can be explained by the features, and (2) the combination of multiple features and which incremental benefit they provide (eg, in explained variance) [35].

Against this background, this study aimed to extend the evidence for various sensor modalities collected via the smartphone (ie, screen features, app usage features, location or GPS features, and call features) by (1) investigating bivariate correlations, (2) exploring the explained variance and incremental benefit of features in cluster-wise regression models (eg, limited to location features), and (3) cluster-combined regression models. Accordingly, the following research questions will be answered:

1. Which bivariate correlations are present between depression severity and (a) screen, (b) app, (c) location, and (d) call features?

2. How much variance in depression severity can be explained in parsimonious cluster-wise regression analyses (eg, limited to location features)?

3. How much variance in depression severity can be explained by the best cluster-combined regression model?

Besides, we wanted to compare the unobtrusively and objectively collected sensor features against features based on ecological momentary assessments (EMA; eg, average valence, average arousal), which, similarly to sensor data, provide a continuous assessment over time but require active input [18,35,37]. Therefore, we investigated the following questions:

4. Which bivariate correlations are present between depression severity and EMA features?

5. How much variance in depression severity can be explained in regression analyses using EMA features?

6. How much variance in depression severity can be explained in regression analyses using EMA and sensing features?

XSL•FO
RenderX

7. What is the difference in explained variance in depression severity between regression models limited to EMA features or sensor features compared to their combination?

Methods

Study Design

This study is an exploratory observation study investigating the associations between smart sensing features and depression severity. Accordingly, we followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [38] (see Multimedia Appendix 1 for the STROBE checklist).

Ethical Considerations

All procedures were assessed and approved by the local ethics committee of Ulm University, Germany (259/16-CL/bal). Informed consent was given by all participants, and participants were informed about their rights according to the European General Data Protection Regulation. Privacy and data security were assessed by Ulm University. Students of Ulm University participating in the study were eligible for study credits as an expense allowance. No other compensation was provided to participants.

Study Population and Procedure

Aiming for a general population sample, participants (see eligibility criteria below) were recruited using an open recruitment strategy involving digital channels (eg, email lists and social media posts) and offline channels (eg, flyers at public institutions). Participants were informed about the purposes of and procedures in the study in an online survey and asked for their written informed consent. If given, they were instructed to install the smart sensing application of the INSIGHTS framework [39,40] after providing basic personal characteristics (ie, gender and age) in the online survey; afterwards, all data were collected via the app. See the assessment and features below for a description of all features and the technical framework paper for further details on the software [39].

Eligibility Criteria for Participants and Episodes

Participants were included in the study if (1) informed consent was given and (2) participants were 18 years or older. (3) Due to the technical requirements of the sensing framework, participants were required to have a smartphone running on Android. Neither a diagnosis nor a minimum level of depression was required to be included in the study. Furthermore, we only included participants' data in the analysis if (4) participants completed the depression severity questionnaire (PHQ-8; see details below) at least once. We structured the data of the participants in 14-day periods consisting of the depression questionnaire assessing the average depression severity in the last 14 days and the corresponding sensing data per day. The number of episodes varied across participants (mean 4.09, SD 3.55; range: 1-42). To avoid the biasing influence of participants being represented more often than others in the dataset and to maximize the data quality, we (5) included only the episode with the lowest amount of missingness per person. Missingness was determined across all days in the episode and features. In addition, (6) we excluded all participants with more than 50% missing data in EMA and sensing features during an episode to ensure missing data handling procedures (see below) were reliable [41,42]. Excluded participants, according to (6), did not significantly differ in age, gender, or depression symptomology from included participants, underlying the assumption that technical issues leading to the exclusion (eg, app not working) did not occur systematically (Multimedia Appendix 2).

Assessment and Features

Overview

The assessment consisted of (1) self-report severity measures, (2) EMA, (3) smartphone screen features, (4) app usage features, (5) location features, and (6) call features. Following the validated procedures from previous studies, we used Python, Snakemake, and the Reproducible Analysis Pipeline for Data Streams (RAPIDS) framework in the data preprocessing and extraction pipeline of all smartphone features [28,43-45]. Smartphone features were calculated for each day and aggregated across the 14-day window (eg, average daily smartphone usage duration across 14 days). A summary of all included smartphone features can be found below.

Clinical Questions

We used the PHQ-8 for the assessment of depression severity. The PHQ-8 consists of 8 self-report items asking how often a symptom was present in the last 14 days (0=not at all to 3=nearly every day). Higher PHQ-8 sum scores indicate higher depression severity. The PHQ-8 is a reliable instrument (Cronbach α=.87, ω=.94) [46,47] (see Multimedia Appendix 3 for an overview of the PHQ-8 items).

EMA Questions

All items were rated from 0 (lowest imaginable) to 100 (highest imaginable). We assessed valence (higher values indicate positive affect), arousal (higher values indicate higher energy levels), and stress (higher values indicate higher stress levels) 3 times per day (morning, midday, and evening). Additionally, sleep quality (high values indicate good sleep) was assessed in the morning, and satisfaction (higher values indicate stronger satisfaction) with the quality of social interaction, the number of social interactions, activity, and nutrition was rated in the evening once per day. See [40] and Multimedia Appendix 4 for an overview of the EMA questions.

Screen Features

The screen sensor tracked the lock and unlock events of the smartphone. Based on this, we determined the number of usage sessions, duration of usage (sum, average, maximum), regularity index of all unlock episodes, entropy, and normalized entropy of unlock events [28,44].

App Features

For each app, we tracked the start and end of each usage session to calculate the count of all used apps, the mean duration of app usage per day, the regularity index of app usage, and the frequency entropy of app usage [44].

XSL•FO
RenderX

Location Features

Using the GPS sensor of the smartphone, we determined the total distance, logarithmic location variance, number of significant places, stay duration (average, maximum, SD, at top 1 location, at top 2 location, and at top 3 location), the ratio between the time spent at nonsignificant places to all clusters (percent of outlier time), location entropy, normalized location entropy, circadian movement, location routine index, number of location transitions, and moving to static ratio [44,48-54].

Call Features

Incoming, outgoing, and missed calls were tracked separately. For each of them, we tracked the *count* and distinct contacts. Furthermore, for incoming and outgoing calls, we calculated the duration (average, sum, and maximum) and entropy [28,44,54].

See Table 1 for an overview of the feature definition, interpretation, and references, providing an in-depth introduction to the background and calculations of the features. For further details on the Reproducible Analysis Pipeline for Data Streams framework, refer to its paper [44].

XSL•FO
RenderX

Table 1. Feature definition and interpretation.

Feature	Definition	Units and interpretation
Ecological momentary assessment		
Valence	Affect ratings from users	0 (lowest) to 100 (highest); higher values indicate positive valence
Arousal	Arousal ratings from users	0 (lowest) to 100 (highest); higher values indicate higher energy levels
Stress	Stress ratings from users	0 (lowest) to 100 (highest); higher values indicate higher stress
Sleep	Sleep quality ratings from users	0 (lowest) to 100 (highest); higher values indicate higher sleep quality
Social quantity	Satisfaction with quality of social contacts reported by users	0 (lowest) to 100 (highest); higher values indicate higher satisfaction
Social quality	Satisfaction with quality of social contacts reported by users	0 (lowest) to 100 (highest); higher values indicate higher satisfaction
Nutrition	Nutrition ratings from users	0 (lowest) to 100 (highest); higher values indicate healthy nutrition
Physical activity	Physical activity ratings from users	0 (lowest) to 100 (highest); higher values indicate higher intensity activity
App usage		
App count	Number of app usage episodes in foreground	Count
App duration	Duration of app usage episodes in foreground	Time in hours
App frequency entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The frequency of use per app over a 24-hour period was used to calculate app frequency entropy. For calculation details, refer to [44,54]	Higher app frequency entropy reflects a more distributed usage of apps (ie, high variability). Lower values indicate that users used one app more often (ie, low variability)
App regularity index	Regularity index captures the similarity of behavior between the same hours across different days. App regularity index refers to the similarity of the most frequently used app at the same hour across days. Calculations were based on [44,54,55]	Higher regularity index indicates a user used the same app at the same hours across days
Screen usage		
Screen episode count	Number of screen unlock episodes	Count
Screen duration	Duration of unlock episodes	Time in hours
Screen regularity index	Regularity index captures the similarity of behavior between the same hours across different days. Screen regularity index refers to the similarity of the most frequent screen status (on or off) at the same hour across days. Refer to [44,55]	Higher regularity index indicates a user used their smartphone at the same hours across days
Screen entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The frequency of screen states (on or off) over a 24-hour period was used to calculate screen entropy. For calculation details, refer to [44,54]	Higher screen entropy reflects a more distributed unlock of the screen (ie, high variability). Lower values indicate that users screen is more often in one state (ie, low variability).
Screen normalized entropy	Normalized entropy is the entropy divided by the logarithm of the number of states (N)	Entropy divided by log(N)
Call		
Missed calls count	Number of missed calls	Count
Missed calls distinct contacts	Number of distinct contacts whose calls were missed	Count
Incoming calls count	Number of incoming calls	Count
Incoming calls distinct contacts	Number of distinct contacts whose calls were answered	Count
Incoming calls duration	Duration of incoming calls	Time in hours

XSL•FO
RenderX

Feature	Definition	Units and interpretation
Incoming calls entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The duration of incoming calls over a 24-hour period was used to calculate incoming calls entropy. For calculation details, refer to [44,54]	Higher incoming calls entropy reflects a more distributed call duration across incoming calls (ie, high variability)
Outgoing calls count	Number of outgoing calls	Count
Outgoing calls distinct contacts	Number of distinct contacts who were called.	Count
Outgoing calls duration	Duration of outgoing calls	Time in hours
Outgoing calls entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The duration of outgoing calls over a 24-hour period was used to calculate outgoing calls entropy. For calculation details, refer to [44,54]	Higher outgoing calls entropy reflects a more distributed call duration across outgoing calls (ie, high variability)
Location		
Total distance	Total distance (haversine) between tracked location coordinates	Distance in km
Location variance	Logarithm of the combined variance in latitude and longitude [44,48-50]	
Moving to static ratio	Ratio of moving states (speed >1 km/h) to static states (speed <1 km/h) [50]	Ratio; higher values reflect more moving states compared to static states
Number of significant clusters	Clusters are determined by k-means clustering of stationary location coordinates (speed <1 km/h). Clusters needed to be 400 meters from each other. Pauses within 200 meters of cluster were counted as cluster visits after the initial clustering. Only clusters with a time duration of 10 minutes were counted as significant. For further details, see [44,48-51]	Count
Staying time at clusters	Time spent at significant clusters	Time in hours
Time at top 1 location	Total time spent at the most significant cluster	Time in hours
Time at top 2 location	Total time spent at the second most significant cluster	Time in hours
Time at top 3 location	Total time spent at the third most significant cluster	Time in hours
Location entropy	Entropy is a measurement of the degree of variability between the users' behavior states. The duration at significant clusters over a 24-hour period was used to calculate location entropy. For calculation details, refer to [44,48-50,54]	Higher location entropy reflects more distributed time spent at significant clusters. Lower values indicate that users spent more time at some significant clusters.
Location normalized entropy	Entropy divided by the logarithm of the number of significant clusters (N)	Entropy divided by the log(N)
Location circadian movement	The extent to which a person's visits at significant clusters follow a 24-hour circadian rhythm. For further details, see [44,50,52,53,56]	Low values indicate a break from routine, whereas high values indicate that a person followed a daily routine.
Time spent at nonsignificant clusters (at outliers)	Time spent at nonsignificant clusters divided by the time spent at all significant cluster	Ratio; higher values indicate more time spent at nonsignificant clusters

Preprocessing and Missing Data Handling

To account for missing data in the dataset, we performed multiple imputations by chained equations [42]. For a missingness overview before imputation, see Multimedia Appendix 5. As outlined before, we constructed 14-day periods consisting of the PHQ-8 and sensing features per day. The following imputations were conducted on the day level (ie, 14 days for each participant and episode). Given the nested data structure (ie, multiple days of the same participant), we applied 2-level predictive mean matching with random intercepts for all variables. A total of 20 complete datasets were obtained. Convergence was achieved after 10 iterations. In a second step, we aggregated the data from the daily level in each imputed dataset to the episode level to match the data structure of the PHQ-8 and sensing features. Aggregation consisted of the mean and SD across the 14 days. Therefore, each imputed dataset contained a single data entry for each person consisting of the average depression severity in the last 14 days (PHQ-8) and the average and SD of the sensing features in the last 14 days.

Statistical Analysis

Given the exploratory character of this study, we decided not to report any *P* values. Instead, all analyses report on the point estimates and corresponding 95% CIs. Throughout the analyses, we conducted analyses on each imputed dataset separately and pooled the data using Rubin's rule with Barnhard-Rubin adjustment for degrees of freedom [42,57-59]. The analysis strategy was structured in 3 steps. First, we calculated bivariate between-person Pearson correlations (*r*) and their 95% CIs

XSL•FO
RenderX

between EMA and sensing features with depression. The correlation is a statistical measure of the magnitude of a linear relationship between 2 variables. It ranges from –1 (perfect negative linear relationship) to +1 (perfect positive linear relationship). Since higher values of the PHQ-8 indicate higher levels of depression, positive correlations between a feature and the PHQ-8 imply that higher values of this feature are associated with higher depression. However, correlations do not allow any causal inference. See Table 1 for interpretation guidance on all features.

Second, variables with non-zero correlation CIs were included as candidates in cluster-wise regression analyses (ie, limited to EMA, screen, app, location, and call features only). The best models per cluster were determined by stepwise backward exclusion of predictors with zero-including CIs. Predictors, starting with the least influential (standardized β) and broadest zero-including CI, were removed one at a time, and regression models were refitted and compared against each other based on adjusted R^2 after each step. This stepwise backward elimination process was continued until all predictors with zero-including CIs were excluded and regression models no longer improved regarding adjusted R^2. For each final cluster-wise regression, we determined the adjusted R^2 and its 95% CI to quantify the explained variance in depression severity.

In the third step, we constructed regression models (1) combining the different sensing feature clusters in one model to evaluate the potential of smart sensing as a stand-alone paradigm and (2) including EMA and sensing feature clusters to evaluate their combined potential. As before, predictors were eliminated following stepwise backward exclusion. Overall model performance was evaluated based on adjusted R^2. Differences in adjusted R^2 and the information criteria Akaike information criterion and Bayesian information criterion were used for model comparisons. Model parameters were standardized and adjusted for age and gender in sensitivity analysis.

Software

All analyses and data preparation were conducted in R. The mice and miceadds packages were used for the imputation and pooling of analysis results [42,60]. For a full list of all packages, see Multimedia Appendix 6.

Results

Overview

A total of 201 participants answered at least 1 PHQ-8. In total, 94 of the responders were excluded due to poor data quality (>50% average missingness in EMA and sensing features). Hence, a total of 107 participants were included in the analysis.

The mean age of the participants was 22.81 (SD 7.32), with the oldest participant being 56 years and the youngest 18 years. In total, 83 (77.6%) of the participants identified themselves as female (male: n=24, 22.4%). On average, participants showed subclinical depression levels (mean 5.82, SD 4.44), with 20 (18.7%) above the PHQ-8 ≥10 cutoff indicating clinically relevant depression severity. For a summary of the PHQ-8 item level, see Multimedia Appendix 7.

EMA features revealed an average valence of 65.98 (SD 12.82), arousal of 50.17 (SD 12.16), and stress of 38.85 (SD 15.66). Daily screen usage of the smartphone was 2.72 hours (SD 2.82) on average. A complete summary of all k=102 features and their respective means and SDs can be found in Multimedia Appendix 8.

Correlations

Correlation analyses revealed small to medium correlations of EMA and sensing features with depression severity. Average valence ratings showed the highest correlation to depression severity of the EMA features (r=–0.55, 95% CI –0.67 to –0.41), while the SD of app frequency entropy in app usage (r=–0.19, 95% CI –0.37 to –0.00), mean duration of outgoing calls (r=0.25, 95% CI 0.04 to 0.43), location routine index (r=0.23, 95% CI 0.02 to 0.42), and the average screen duration (r=0.37, 95% CI 0.20 to 0.53) were the strongest features in their respective clusters. See Table 2 for a summary of all correlations with zero, excluding CIs. The complete correlation summary of all investigated features can be found in Multimedia Appendix 9.

XSL•FO
RenderX

Table 2. Pooled bivariate correlation results between depression and features.[a]

	r (95% CI)
Ecological momentary assessment features	
Average valence	–0.55 (–0.67 to –0.41)
Average satisfaction with social quality	–0.51 (–0.64 to –0.35)
Average sleep quality	–0.50 (–0.63 to –0.34)
Average arousal	–0.42 (–0.57 to –0.25)
Average social quantity	–0.39 (–0.54 to –0.22)
Average nutrition	–0.25 (–0.45 to –0.03)
SD of valence	0.23 (0.04 to 0.41)
Average stress	0.42 (0.25 to 0.56)
App features	
SD of frequency entropy	–0.19 (–0.37 to –0.00)
Call features	
Average incoming call duration	0.21 (0.00 to 0.39)
SD of incoming call duration	0.21 (0.01 to 0.40)
Average outgoing call duration	0.25 (0.04 to 0.43)
SD of outgoing call duration	0.25 (0.04 to 0.44)
Features	
Average location routine index	0.23 (0.02 to 0.42)
Screen features	
Average total screen duration	0.23 (0.04 to 0.40)
Average max screen duration	0.24 (0.05 to 0.41)
Average SD of screen episode count	0.27 (0.09 to 0.44)
Average duration of screen episodes	0.37 (0.20 to 0.53)

[a]Correlations are pooled correlations based on multiple imputations. Only correlations with their 95% CI excluding zero are displayed. A full correlation summary can be found in Multimedia Appendix 9. For feature definition and interpretation, refer to Table 1.

Regression

We included the features identified in the correlation analysis in step-wise regression analyses to investigate their incremental contribution to the explained variance in depression severity. The final regression model using EMA features included average valence (=–0.39, 95% CI –0.58 to –0.21) and social quality as predictors (=–0.29, 95% CI –0.48 to –0.10). Combined, they explained 35.28% of the variance (95% CI 20.73% to 49.64%).

In the app cluster, SD of app frequency entropy (=–0.19, 95% CI –0.38 to –0.00) explained adjusted R^2=2.81% of the variance (95% CI 0.00%-12.02%), while location routine index (=0.23, 95% CI 0.02 to 0.44) explained adjusted R^2=4.39% (95% CI 0.00% to 16.71%) in the location cluster. From the call cluster, the SD of incoming call duration (=0.20, 95% CI 0.03 to 0.41) and average outgoing call duration (=0.24, 95% CI 0.04 to 0.44) were the final included predictors explaining adjusted R^2=8.68% of the variance (95% CI 0.88% to 22.32%). Of all 4 candidates in the screen cluster, the average screen duration was the only included predictor in the final model (=0.37, 95% CI 0.19 to 0.55; adjusted R^2=13.09%, 95% CI 3.40% to 26.65%).

Combining all sensing features in a parsimonious model yielded a model explaining adjusted R^2=20.45% of the variance (95% CI 7.81% to 35.59%) with average screen duration (=0.39, 95% CI 0.21 to 0.56), SD of app frequency entropy (=–0.19, 95% CI –0.36 to –0.02), and SD of incoming call duration (=0.21, 95% CI 0.02 to 0.41) as predictors.

The highest variance was explained when combining EMA and sensing features: adj. R^2=45.15% of the variance (95% CI 30.39% to 58.53%). Features included in this prediction model were average valence (=–0.36, 95% CI –0.53 to –0.19), social quality (=–0.24, 95% CI –0.41 to –0.06), average screen duration (=0.22, 95% CI 0.07 to 0.37), SD of app frequency entropy (= –0.17, 95% CI –0.31 to –0.02), and average

XSL•FO
RenderX

duration of outgoing calls (=0.17, 95% CI 0.01 to 0.33). See Table 3 for a summary of the parsimonious EMA, sensing, and combined regression models.

Table 3. Regression results for depression predicted by ecological momentary assessment and smartphone features in stand-alone and combined models.[a]

	(95% CI)	Adjusted R^2 (95% CI), %	AIC[b]	BIC[c]	Δ adjusted R^2, %
EMA[d] cluster		35.28 (20.73 to 49.64)	262.45	273.14	N/A[e]
Average valence	–0.39 (–0.58 to –0.21)				
Average social quality	–0.29 (–0.48 to –0.10)				
Sensing cluster		20.45 (7.81 to 35.59)	284.53	297.90	EMA: –14.83
Average screen duration	0.39 (0.21 to 0.56)				
Average app frequency entropy	–0.19 (–0.36 to –0.02)				
Incoming call SD	0.21 (0.02 to 0.41)				
Combined		45.15 (30.39 to 58.53)	249.42	268.13	EMA: 9.87; sensing: 24.70
Average valence	–0.36 (–0.53 to –0.19)				
Average social quality	–0.24 (–0.41 to –0.06)				
Average screen duration	0.22 (0.07 to 0.37)				
Average app frequency entropy	–0.17 (–0.31 to –0.02)				
Average duration of outgoing calls	0.17 (0.01 to 0.33)				

[a]All results were obtained by pooling results from multiple imputations according to Rubin's rule. All estimates were fully standardized. For feature definition and interpretation, refer to Table 1.

[b]AIC: Akaike information criterion.

[c]BIC: Bayesian information criterion.

[d]EMA: ecological momentary assessment.

[e]Not applicable.

Results were robust when adjusting for age and gender, yielding nonsignificant main effects in the parsimonious EMA (age: –0.08 to 0.23; gender: –0.06 to 0.26), sensing (age: –0.16 to 0.21; gender: –0.26 to 0.10), and combined EMA and sensing model (age: –0.09 to 0.22; gender: –0.09 to 0.21).

Discussion

Principal Findings

To explore the potential of smart sensing for depression, this study investigated the bivariate correlations and explainable variance in regression models built on smartphone sensors and EMA features. Across sensor modalities, we found small correlations between smart sensing features and depression severity. Combined smart sensing features could explain 20.45% (95% CI 7.81%-35.59%) of the depression severity variance in a parsimonious model. Conversely, we found small to medium correlations for EMA features, which could explain 35.28% (95% CI 20.73%-49.64%) of the variance. The best model was the combination of smart sensing and EMA features, which explained 45.15% (95% CI 30.39%-58.53%; Δ adjusted R^2 to smart sensing only: 24.70%; Δ adjusted R^2 to EMA only: 9.87%).

Comparison to Prior Work

The EMA findings are in line with previous studies and reviews highlighting the potential of EMA as a continuous assessment to infer depression severity [35,37,40,61,62]. However, while the bivariate correlations, as well as the explained variance, are higher for EMA compared to features obtained from smartphone sensors, it is important to note that the sensing cluster alone could explain about 20% of the variance. Given the unobtrusive nature of the collection of sensor data, this approach has a crucial advantage over EMA, which requires active user or patient involvement over a long period (eg, multiple daily responses over 14 days). In particular, for clinical application, it should be evaluated whether the additional burden for patients to answer EMA is proportional to the gain in explained variance. Furthermore, various issues in EMA (eg, interpretation of momentary questions, usage of comparison standards) could be avoided by the collection of objective sensor data [19,21,62,63]. That said, to maximize the explained variance, the combination of sensor features and EMA seems to be best. This result extends the findings by Moshe et al [35], who previously evaluated GPS features in conjunction with physiological wearable data and EMA similarly, showing that the combination of sensing features with EMA yields the best regression model for depression severity.

Although these findings seem promising, it is also important to note that this study and others so far are exploratory [34-36,64].

XSL•FO
RenderX

Although we have more than 2 decades of research on EMA and its application for mental health [18,37], the field of smart sensing is still in its infancy. Facing heterogenous methodology and study quality, as well as potential publication bias in the field, confirmatory studies are highly needed in the field of smart sensing before clinical application [34,36,64,65]. Besides, it is important to note that this study followed a rather data-driven approach to investigating features, which were collectible by the here-used framework. In the context of smart sensing, a central question is which features are needed and provide an incremental benefit. Although this study can give first insights into this topic, predictors like screen duration, app usage entropy, and call features (eg, SD in incoming calls) should only be incorporated in clinical systems if replicated in future studies. Also, an extension to other sensors (eg, language analysis based on LIWC or sentiment analysis) [66-70], app content (eg, usage of social media apps) [71], network usage [28], and the combination with other wearable devices (eg, biophysiological data from smart watches) or different data sources (eg, journaling data) would be a promising addition [23,24,35,67].

However, a closer inspection is needed not only for smart sensing features but also for EMA. For example, our study highlighted that many of the investigated features (eg, arousal or physical activity) did not hold any incremental benefit besides valence and social quality. Since EMA is associated with additional burden for patients, it is especially important to identify the core set of items that maximize predictive power while reducing the item load. Despite a long research history, surprisingly little systematic and meta-analytical evidence is available on which questions should be asked in the context of depression and more precisely when (eg, morning or evening) and on which schedule (eg, multiple fixed time points or microrandomized assessments) [37,63,72-74].

Besides the empirical evidence for the applicability of smart sensing, questions of acceptance, data security and privacy, and ethical challenges surrounding smart sensing need to be addressed [19,75-77]. For instance, recent studies found only moderate acceptance of smart sensing in the context of mental health and highlighted the impact of data types and recipients in smart sensing [77-79]. Only when these barriers and challenges are overcome, can smart sensing unfold its potential fully.

Lastly, we want to point out that the current developments in smart sensing for mental health mainly focus on the prediction of psychopathology [24,26,27,36,64]. Future studies could further advance the field by not only focusing on pathology but also investigating applications to assess risk factors (eg, loneliness, stress) or mediators and mechanisms of change (eg, rumination, therapeutic alliance). Understanding the underlying process of how treatment works is a crucial step to optimizing treatments and pathways in mental health care [80-83]. Since smart sensing allows for a fine-grained and unobtrusive assessment, it might become a promising and feasible paradigm to unveil the mechanism of change in mental health care and better understand the dynamics of therapeutic processes.

Limitations

In addition to the already highlighted exploratory nature of this study, we would like to emphasize a few limitations of this study. First, this study followed a cross-sectional design investigating which predictors explain the variance between persons. Hence, any causal interpretation of the results is not eligible and would require different study designs. For instance, the increased smartphone usage (eg, screen duration) could be caused by depression, but also vice versa or even explained by a third variable. Also, the analysis of trajectories and dynamics over time was not in the scope of this study. Based on the multiple episodes of participants in this study, a longitudinal perspective would be a valuable addition to this study and a field with so far underrepresented research on longitudinal models [34,84,85].

Second, the present sample was a convenient sample recruited in the general population, which resulted in a rather young (mean 22.81, SD 7.32, range 18-56) and unequal gender distribution (female: 77.6%). Furthermore, only 18.7% of the participants showed clinically relevant symptomology. Therefore, the generalizability of the findings to other samples, especially clinical samples, remains open. However, by showing the feasibility of smart sensing predictions and their incremental benefit in explaining variance in depression, this study lays a strong foundation to move to clinical populations and studies. To further increase the quality of clinical studies, also methodological points need to be addressed; for instance, self-report instruments, as applied in this study (ie, PHQ-8), are prone to several sources of bias (eg, social desirability or recall biases). Hence, the application of more reliable and valid assessments like clinician ratings or medical diagnosis should be considered in future studies alongside measurements, which maximize reliability in a specific depression severity range of interest (eg, high reliability in subclinical or severe depression levels) [86-88].

Third, alongside the sample characteristics, the sample size needs to be considered. Adequate sample size is key to designing confirmatory studies aiming to test an assumed clinically relevant effect with sufficient power. In depression research, a standardized mean difference of 0.24, which transfers to a correlation of r=0.12 [89], is argued to be a clinically relevant effect [90]. However, to be able to test such a correlation with sufficient power (eg, 80%), a sample size of N=542 would be required (assuming a 2-sided test with α=5%, following a bivariate normal model). Hence, our study was highly underpowered to test for a minimally clinically relevant correlation of r=0.12. Accordingly, we did not report any P values and solely reported on the 95% CIs to provide a range of reasonable assumable magnitudes of estimates, which could guide future studies in their sample size planning.

Lastly, we would like to emphasize that we opted for linear regression analysis to investigate the incremental contribution of various predictors and allow for direct comparison of various models. This method provides a straightforward approach to answering the present research question. However, sensor data are also very complex, and previous studies have highlighted the potential of nonlinear machine learning models for

XSL•FO
RenderX

depression severity to fully exploit the smart sensing data [25,28]. Although previously applied to classification, machine learning models like extreme gradient-boosted regression trees may also be promising in regression. On the flip side, these models go hand in hand with challenges such as overfitting and difficulties in the interpretation and explanations of the models [25,28,91]. Therefore, even if proven to improve the predictive accuracy, it should be carefully considered whether the complexity and downsides (eg, explainability) of potential machine learning models justify their usage over simpler but easy-to-interpret statistical regression models [92].

Conclusions

Smart sensing and EMA provide potent paradigms to infer and predict depression severity. Our results show that EMA and sensing features alone can substantially explain variance in depression severity. In isolation, EMA was superior to sensing features in terms of explained variance. However, sensing features alone could explain about 20% of the variance, emphasizing the potential of this unobtrusive and objective assessment of depression. To maximize explainable variance, EMA and sensing features should be combined. However, while these findings are promising, confirmatory studies, particularly in clinical settings and samples, are highly needed before robust conclusions can be drawn.

Acknowledgments

We would like to thank all the student assistants involved in the project. No generative AI (eg, ChatGPT) was used in this project and manuscript. This study was self-funded by the authors. YT is supported by the initial phase of the German Center for Mental Health (Deutsches Zentrum für Psychische Gesundheit; grant 01EE2303A).

Data Availability

The primary data obtained in this observation study can be provided by HB and YT on reasonable request. Data-sharing agreements may have to be signed depending on the request. Requests should be directed at the corresponding author (YT). Support is depending on available resources.

Authors' Contributions

YT, EMM, CM, and HB were involved in the conceptualization. Data curation was done by YT, EMM, and KOA. Formal analysis was done by YT. CM and HB were involved in funding acquisition. EMM, YT, HB, CK, and CM were involved in the investigation. YT and KOA were involved in the methodology. Project administration was done by YT and EMM. YT, EMM, CM, CK, and HB contributed to resources; software was contributed by YT, KOA, and CK. YT, EMM, and HB were involved in supervision. Validation was done by YT, EMM, KOA, and CK. Visualization was done by YT. YT also contributed to writing the original draft. YT, EMM, KOA, CM, CK, and HB contributed to reviewing and editing the manuscript.

Conflicts of Interest

All authors declare no conflicts of interest. However, for reasons of transparency, CM notes that he has received (to Ulm University and earlier University of Bonn) grants from agencies such as the German Research Foundation. CM has performed grant reviews for several agencies, edited journal sections and articles, given academic lectures at clinical or scientific venues or for companies, and generated books or book chapters for publishers of mental health texts. For some of these activities, he received royalties, but never from gaming or social media companies. CM was part of a discussion circle (Digitalität und Verantwortung: https://about.fb.com/de/news/h/gespraechskreis-digitalitaet-und-verantwortung/) debating ethical questions linked to social media, digitalization, and society/democracy at Facebook. In this context, he received no salary for his activities. Finally, he currently functions as an independent scientist on the scientific advisory board of the Nymphenburg Group. This activity is financially compensated. Moreover, he is on the scientific advisory board of Applied Cognition, an activity which is also compensated.

Multimedia Appendix 1
STROBE checklist.
[DOCX File , 22 KB - jmir_v27i1e55308_app1.docx]

Multimedia Appendix 2
Sensitivity analysis on study exclusion.
[DOCX File , 16 KB - jmir_v27i1e55308_app2.docx]

Multimedia Appendix 3
Patient Health Questionnaire items.
[DOCX File , 16 KB - jmir_v27i1e55308_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Ecological momentary assessment items.
[DOCX File , 15 KB - jmir_v27i1e55308_app4.docx]

Multimedia Appendix 5
Overview of missingness per feature before multiple imputation.
[DOCX File , 17 KB - jmir_v27i1e55308_app5.docx]

Multimedia Appendix 6
Software information.
[DOCX File , 15 KB - jmir_v27i1e55308_app6.docx]

Multimedia Appendix 7
Additional sample characteristics.
[DOCX File , 15 KB - jmir_v27i1e55308_app7.docx]

Multimedia Appendix 8
Feature means and SDs.
[DOCX File , 21 KB - jmir_v27i1e55308_app8.docx]

Multimedia Appendix 9
Full correlations between depression and features.
[DOCX File , 35 KB - jmir_v27i1e55308_app9.docx]

References

1. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a Lancet-World Psychiatric Association Commission. Lancet 2022;399(10328):957-1022. [doi: 10.1016/S0140-6736(21)02141-3] [Medline: 35180424]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
4. World mental health report: transforming mental health for all. Geneva: World Health Organization; 2022. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2024-12-20]
5. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020;19(1):92-107 [FREE Full text] [doi: 10.1002/wps.20701] [Medline: 31922679]
6. Cuijpers P, Karyotaki E, Reijnders M, Ebert DD. Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. Epidemiol Psychiatr Sci 2019;28(1):21-30 [FREE Full text] [doi: 10.1017/S2045796018000057] [Medline: 29486804]
7. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
8. Vigo D, Haro JM, Hwang I, Aguilar-Gaxiola S, Alonso J, Borges G, et al. Toward measuring effective treatment coverage: critical bottlenecks in quality- and user-adjusted coverage for major depressive disorder. Psychol Med 2020 Oct 20:1-11 [FREE Full text] [doi: 10.1017/S0033291720003797] [Medline: 33077023]
9. Andrade LH, Alonso J, Mneimneh Z, Wells J, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
10. Trautmann S, Beesdo-Baum K. The treatment of depression in primary care. Dtsch Arztebl Int 2017;114(43):721-728 [FREE Full text] [doi: 10.3238/arztebl.2017.0721] [Medline: 29143731]
11. Kroenke K, Unutzer J. Closing the false divide: sustainable approaches to integrating mental health services into primary care. J Gen Intern Med 2017;32(4):404-410 [FREE Full text] [doi: 10.1007/s11606-016-3967-9] [Medline: 28243873]

XSL•FO
RenderX

12. Thombs BD, Coyne JC, Cuijpers P, de Jonge P, Gilbody S, Ioannidis J, et al. Rethinking recommendations for screening for depression in primary care. CMAJ 2012;184(4):413-418 [FREE Full text] [doi: 10.1503/cmaj.111035] [Medline: 21930744]
13. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009;374(9690):609-619. [doi: 10.1016/S0140-6736(09)60879-5] [Medline: 19640579]
14. Terhorst Y, Sander LB, Ebert DD, Baumeister H. Optimizing the predictive power of depression screenings using machine learning. Digit Health 2023;9:20552076231194939 [FREE Full text] [doi: 10.1177/20552076231194939] [Medline: 37654715]
15. Lotfi L, Flyckt L, Krakau I, Mårtensson B, Nilsson GH. Undetected depression in primary healthcare: occurrence, severity and co-morbidity in a two-stage procedure of opportunistic screening. Nord J Psychiatry 2010;64(6):421-427. [doi: 10.3109/08039481003786378] [Medline: 20429746]
16. First MB, Rebello TJ, Keeley JW, Bhargava R, Dai Y, Kulygina M, et al. Do mental health professionals use diagnostic classifications the way we think they do? A global survey. World Psychiatry 2018;17(2):187-195 [FREE Full text] [doi: 10.1002/wps.20525] [Medline: 29856559]
17. Wichers M. The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. Psychol Med 2014;44(7):1349-1360. [doi: 10.1017/S0033291713001979] [Medline: 23942140]
18. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
19. Terhorst Y, Knauer J, Baumeister H. Smart sensing enhanced diagnostic expert systems. In: Montag C, Baumeister H, editors. Digital Phenotyping and Mobile Sensors. Berlin: Springer; 2023:413-425.
20. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
21. Garatva P, Terhorst Y, Messner EM, Karlen W, Pryss R, Baumeister H. Smart sensors for health research and improvement. In: Montag C, Baumeiste H, editors. Digital and Phenotyping Mobile Sensors. Berlin: Springer; 2023:395-411.
22. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
23. Steele R, Hillsgrove T, Khoshavi N, Jaimes LG. A survey of cyber-physical system implementations of real-time personalized interventions. J Ambient Intell Human Comput 2021;13(5):2325-2342 [FREE Full text] [doi: 10.1007/s12652-021-03263-0]
24. Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. NPJ Digit Med 2023;6(1):84 [FREE Full text] [doi: 10.1038/s41746-023-00828-5] [Medline: 37147384]
25. Dlima SD, Shevade S, Menezes SR, Ganju A. Digital phenotyping in health using machine learning approaches: scoping review. JMIR Bioinform Biotechnol 2022;3(1):e39618 [FREE Full text] [doi: 10.2196/39618] [Medline: 38935947]
26. Zarate D, Stavropoulos V, Ball M, de Sena Collier G, Jacobson NC. Correction: exploring the digital footprint of depression: a PRISMA systematic literature review of the empirical evidence. BMC Psychiatry 2022;22(1):530 [FREE Full text] [doi: 10.1186/s12888-022-04153-1] [Medline: 35932004]
27. Nouman M, Khoo SY, Mahmud MAP, Kouzani AZ. Recent advances in contactless sensing technologies for mental health monitoring. IEEE Internet Things J 2022;9(1):274-297 [FREE Full text] [doi: 10.1109/jiot.2021.3097801]
28. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. JMIR mHealth uHealth 2021;9(7):e26540 [FREE Full text] [doi: 10.2196/26540] [Medline: 34255713]
29. Kotov R, Krueger R, Watson D, Achenbach T, Althoff R, Bagby R, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. J Abnorm Psychol 2017;126(4):454-477. [doi: 10.1037/abn0000258] [Medline: 28333488]
30. Regier DA, Narrow W, Clarke D, Kraemer H, Kuramoto S, Kuhl E, et al. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. Am J Psychiatry 2013;170(1):59-70. [doi: 10.1176/appi.ajp.2012.12070999] [Medline: 23111466]
31. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J Affect Disord 2015;172:96-102 [FREE Full text] [doi: 10.1016/j.jad.2014.10.010] [Medline: 25451401]
32. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord 2017;208:191-197 [FREE Full text] [doi: 10.31234/osf.io/vbk59]
33. Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. Lancet Psychiatry 2020;7(8):692-702. [doi: 10.1016/S2215-0366(20)30191-7] [Medline: 32711710]
34. Terhorst Y, Knauer J, Philippi P, Baumeister H. The relation between passively collected GPS mobility metrics and depressive symptoms: systematic review and meta-analysis. J Med Internet Res 2024;26:e51875 [FREE Full text] [doi: 10.2196/51875] [Medline: 39486026]

XSL•FO
RenderX

35. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
36. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. JMIR mHealth uHealth 2018;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
37. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. J Abnorm Psychol 2020;129(1):56-63. [doi: 10.1037/abn0000473] [Medline: 31868388]
38. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche PC, Vandenbroucke J, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
39. Montag C, Baumeister H, Kannen C, Sariyska R, Meßner E, Brand M. Concept, possibilities and pilot-testing of a new smartphone application for the social and life sciences to study human behavior including validation data from personality psychology. J 2019;2(2):102-115 [FREE Full text] [doi: 10.3390/j2020008]
40. Messner EM, Sariyska R, Mayer B, Montag C, Kannen C, Schwerdtfeger A, et al. Insights – future implications of passive smartphone sensing in the therapeutic context. Verhaltenstherapie 2019;32(Suppl. 1):86-95 [FREE Full text] [doi: 10.1159/000501951]
41. Enders CK. Applied Missing Data Analysis. New York: The Guilford Press; 2010.
42. van Buuren S, Groothuis-Oudshoorn CG. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45(3) [FREE Full text] [doi: 10.32614/cran.package.mice]
43. Köster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinformatics 2012;28(19):2520-2522. [doi: 10.1093/bioinformatics/bts480] [Medline: 22908215]
44. Vega J, Li M, Aguillera K, Goel N, Joshi E, Khandekar K, et al. Reproducible analysis pipeline for data streams: open-source software to process data collected with mobile Devices. Front Digit Health 2021;3:769823 [FREE Full text] [doi: 10.3389/fdgth.2021.769823] [Medline: 34870271]
45. Opoku Asare K, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: a longitudinal data analysis. Pervasive Mob Comput 2022;83:101621 [FREE Full text] [doi: 10.1016/j.pmcj.2022.101621]
46. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
47. Arias de la Torre J, Vilagut G, Ronaldson A, Valderas J, Bakolis I, Dregan A, et al. Reliability and cross-country equivalence of the 8-item version of the patient health questionnaire (PHQ-8) for the assessment of depression: results from 27 countries in Europe. Lancet Reg Health Eur 2023;31:100659 [FREE Full text] [doi: 10.1016/j.lanepe.2023.100659] [Medline: 37332385]
48. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr D. The relationship between clinical, momentary, and sensor-based assessment of depression. Int Conf Pervasive Comput Technol Healthc 2015;2015:103 [FREE Full text] [doi: 10.4108/icst.pervasivehealth.2015.259034] [Medline: 26640739]
49. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
50. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
51. Arthur D, Vassilvitskii S. K-Means++: the advantages of careful seeding. 2007 Presented at: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms; January 7-9, 2007; New Orleans, Louisiana, USA.
52. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp; September 7-11, 2015; Osaka, Japan p. 1293-1304 URL: https://doi.org/10.1145/2750858.2805845 [doi: 10.1145/2750858.2805845]
53. Barnett I, Onnela JP. Inferring mobility measures from GPS traces with missing data. Biostatistics 2020;21(2):e98-e112 [FREE Full text] [doi: 10.1093/biostatistics/kxy059] [Medline: 30371736]
54. Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27(3):379-423. [doi: 10.1002/j.1538-7305.1948.tb01338.x]
55. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, et al. Sensing behavioral change over time. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(3):1-21 [FREE Full text] [doi: 10.1145/3264951]
56. Wu C, McMahon M, Fritz H, Schnyer DM. Circadian rhythms are not captured equal: exploring circadian metrics extracted by different computational methods from smartphone accelerometer and GPS sensors in daily life tracking. Digit Health 2022;8:20552076221114201 [FREE Full text] [doi: 10.1177/20552076221114201] [Medline: 35874860]
57. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91(434):473-489. [doi: 10.2307/2291635]

XSL•FO
RenderX

58. Barnard J, Rubin DB. Miscellanea. Small-sample degrees of freedom with multiple imputation. Biometrika 1999;86(4):948-955 [FREE Full text] [doi: 10.1093/biomet/86.4.948]
59. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, New Jersey: John Wiley & Sons; 1987.
60. Robitzsch A, Grund S, Henke T. miceadds: some additional multiple imputation functions, especially for mice. 2018. URL: https://alexanderrobitzsch.r-universe.dev/miceadds [accessed 2024-11-23]
61. Hall M, Scherner PV, Kreidel Y, Rubel JA. A systematic review of momentary assessment designs for mood and anxiety symptoms. Front Psychol 2021;12:642044 [FREE Full text] [doi: 10.3389/fpsyg.2021.642044] [Medline: 34079492]
62. Colombo D, Fernández-Álvarez J, Patané A, Semonella M, Kwiatkowska M, García-Palacios A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019;8(4):465 [FREE Full text] [doi: 10.3390/jcm8040465] [Medline: 30959828]
63. Stone AA, Schneider S, Smyth JM. Evaluation of pressing issues in ecological momentary assessment. Annu Rev Clin Psychol 2023;19:107-131 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080921-083128] [Medline: 36475718]
64. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
65. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022;5(1):3 [FREE Full text] [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
66. Rathner EM, Djamali J, Terhorst Y, Schuller B, Cummins N, Salamon G, et al. How did you like 2017? Detection of language markers of depression and narcissism in personal narratives. 2018 Presented at: Interspeech 2018; September 2-6, 2018; Hyderabad, India p. 3388-3392 URL: https://doi.org/10.21437/Interspeech.2018-2040 [doi: 10.21437/interspeech.2018-2040]
67. Kathan A, Triantafyllopoulos A, He X, Milling M, Yan T, Rajamani ST, et al. Journaling data for daily PHQ-2 depression prediction and forecasting. Annu Int Conf IEEE Eng Med Biol Soc 2022;2022:2627-2630. [doi: 10.1109/EMBC48229.2022.9871015] [Medline: 36086268]
68. Zantvoort K, Scharfenberger J, Boß L, Lehr D, Funk B. Finding the best match - a case study on the (text-)feature and model choice in digital mental health interventions. J Healthc Inform Res 2023;7(4):447-479 [FREE Full text] [doi: 10.1007/s41666-023-00148-z] [Medline: 37927375]
69. Nickels S, Edwards MD, Poole SF, Winter D, Gronsbell J, Rozenkrants B, et al. Toward a mobile platform for real-world digital measurement of depression: user-centered design, data quality, and behavioral and clinical modeling. JMIR Ment Health 2021;8(8):e27589 [FREE Full text] [doi: 10.2196/27589] [Medline: 34383685]
70. Hussain F, Stange JP, Langenecker SA, McInnis M, Zulueta J, Piscitello A, et al. Passive sensing of affective and cognitive functioning in mood disorders by analyzing keystroke kinematics and speech dynamics. In: Baumeister H, Montag C, editors. Mobile Sensing and Digital Phenotyping in Psychoinformatics. Berlin: Springer; 2019.
71. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among U.S. young adults. Depress Anxiety 2016;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]
72. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. PsyArXiv Preprint posted online on February 20, 2020 [FREE Full text] [doi: 10.31234/osf.io/zf4nm]
73. Reiter T, Schoedel R. Never miss a beep: Using mobile sensing to investigate (non-)compliance in experience sampling studies. Behav Res Methods 2024;56(4):4038-4060 [FREE Full text] [doi: 10.3758/s13428-023-02252-9] [Medline: 37932624]
74. Cloos L, Ceulemans E, Kuppens P. Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. Psychol Assess 2023;35(3):189-204. [doi: 10.1037/pas0001200] [Medline: 36480406]
75. Torous J, Nebeker C. Navigating ethics in the digital age: introducing Connected and Open Research Ethics (CORE), a tool for researchers and institutional review boards. J Med Internet Res 2017;19(2):e38 [FREE Full text] [doi: 10.2196/jmir.6793] [Medline: 28179216]
76. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. Lancet Digit Health 2020;2(5):e221-e223 [FREE Full text] [doi: 10.1016/S2589-7500(20)30065-0] [Medline: 33328054]
77. Terhorst Y, Weilbacher N, Suda C, Simon L, Messner EM, Sander LB, et al. Acceptance of smart sensing: a barrier to implementation-results from a randomized controlled trial. Front Digit Health 2023;5:1075266 [FREE Full text] [doi: 10.3389/fdgth.2023.1075266] [Medline: 37519894]
78. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. JMIR mHealth uHealth 2019;7(4):e12578 [FREE Full text] [doi: 10.2196/12578] [Medline: 30950799]

XSL•FO
RenderX

79. Rottstädt F, Becker E, Wilz G, Croy I, Baumeister H, Terhorst Y. Enhancing the acceptance of smart sensing in psychotherapy patients: findings from a randomized controlled trial. Front Digit Health 2024;6:1335776 [FREE Full text] [doi: 10.3389/fdgth.2024.1335776] [Medline: 38698889]
80. Pfammatter M, Tschacher W. Klassen allgemeiner wirkfaktoren der psychotherapie und ihr zusammenhang mit therapietechniken [Article in German]. Z Psychosom Med Psychother 2016;45(1):1-13 [FREE Full text] [doi: 10.1026/1616-3443/a000331]
81. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
82. Domhardt M, Grund S, Mayer A, Büscher R, Ebert DD, Sander LB, et al. Unveiling mechanisms of change in digital interventions for depression: study protocol for a systematic review and individual participant data meta-analysis. Front Psychiatry 2022;13:899115 [FREE Full text] [doi: 10.3389/fpsyt.2022.899115] [Medline: 36262633]
83. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. Annu Rev Clin Psychol 2007;3:1-27. [doi: 10.1146/annurev.clinpsy.3.022806.091432] [Medline: 17716046]
84. Zhang Y, Folarin A, Sun S, Cummins N, Vairavan S, Bendayan R, RADAR-CNS consortium. Longitudinal relationships between depressive symptom severity and phone-measured mobility: dynamic structural equation modeling study. JMIR Ment Health 2022;9(3):e34898 [FREE Full text] [doi: 10.2196/34898] [Medline: 35275087]
85. Müller SR, Peters H, Matz SC, Wang W, Harari GM. Investigating the relationships between mobility behaviours and indicators of subjective well–being using smartphone–based experience sampling and GPS tracking. Eur J Pers 2020;34(5):714-732 [FREE Full text] [doi: 10.1002/per.2262]
86. Wahl I, Löwe B, Bjorner JB, Fischer F, Langs G, Voderholzer U, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. J Clin Epidemiol 2014;67(1):73-86. [doi: 10.1016/j.jclinepi.2013.04.019] [Medline: 24262771]
87. Brehaut E, Neupane D, Levis B, Wu Y, Sun Y, Krishnan A, et al. Depression prevalence using the HADS-D compared to SCID major depression classification: an individual participant data meta-analysis. J Psychosom Res 2020;139:110256. [doi: 10.1016/j.jpsychores.2020.110256] [Medline: 33069051]
88. Levis B, Yan XW, He C, Sun Y, Benedetti A, Thombs BD. Comparison of depression prevalence estimates in meta-analyses based on screening tools and rating scales versus diagnostic interviews: a meta-research review. BMC Med 2019;17(1):65 [FREE Full text] [doi: 10.1186/s12916-019-1297-6] [Medline: 30894161]
89. Nordahl-Hansen A, Øien RA, Volkmar F, Shic F, Cicchetti DV. Enhancing the understanding of clinically meaningful results: a clinical research perspective. Psychiatry Res 2018;270:801-806 [FREE Full text] [doi: 10.1016/j.psychres.2018.10.069] [Medline: 30551328]
90. Cuijpers P, Turner EH, Koole SL, van Dijke A, Smit F. What is the threshold for a clinically relevant effect? The case of major depressive disorders. Depress Anxiety 2014;31(5):374-378. [doi: 10.1002/da.22249] [Medline: 24677535]
91. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 2018;14:91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]
92. Jacobucci R, Grimm KJ. Machine learning and psychological research: the unexplored effect of measurement. Perspect Psychol Sci 2020;15(3):809-816. [doi: 10.1177/1745691620902467] [Medline: 32348703]

Abbreviations

EMA: ecological momentary assessment
MDD: major depressive disorder
PHQ-8: 8-item Patient Health Questionnaire
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Coristine; submitted 08.12.23; peer-reviewed by R Antunes, H Hsin; comments to author 18.04.24; revised version received 30.06.24; accepted 18.10.24; published 30.01.25.

Please cite as:
Terhorst Y, Messner EM, Opoku Asare K, Montag C, Kannen C, Baumeister H
Investigating Smartphone-Based Sensing Features for Depression Severity Prediction: Observation Study
J Med Internet Res 2025;27:e55308
URL: https://www.jmir.org/2025/1/e55308
doi:10.2196/55308
PMID:39883512

XSL•FO
RenderX

©Yannik Terhorst, Eva-Maria Messner, Kennedy Opoku Asare, Christian Montag, Christopher Kannen, Harald Baumeister. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Web-Based Self-Guided Intervention (MINDxYOU) for Reducing Stress and Promoting Mental Health Among Health Professionals: Results From a Stepped-Wedge Cluster Randomized Trial

Yolanda López-del-Hoyo[1,2,3*], PhD; Selene Fernández-Martínez[4*], MSci; Adrian Perez-Aranda[5,6], PhD; Alicia Monreal-Bartolomé[2,3,7], PhD; Alberto Barceló-Soler[3,7], PhD; Loreto Camarero-Grados[2], MSc; Carilene Armas-Landaeta[4], MD; José Guzmán-Parra[8,9], PhD; Vera Carbonell[8,9], MSc; Daniel Campos[2,7], PhD; Xinyuan Chen[10], PhD; Javier García-Campayo[2,3,4], PhD

[1]Faculty of Social Sciences, University of Zaragoza, Zaragoza, Spain
[2]Institute of Health Research of Aragon, Zaragoza, Spain
[3]Research Network on Chronicity, Primary Care and Health Promotion (RICAPPS), Zaragoza, Spain
[4]Faculty of Medicine, University of Zaragoza, Zaragoza, Spain
[5]Department of Clinical and Health Psychology, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
[6]CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain
[7]Department of Psychology and Sociology, Faculty of Education, University of Zaragoza, Zaragoza, Spain
[8]Instituto de Investigación Biomedicina de Málaga, Málaga, Spain
[9]Unidad de Gestión Clínica de Salud Mental, Hospital Regional Universitario de Málaga, Málaga, Spain
[10]Department of Mathematics and Statistics, Mississippi State University, MS, United States
[*]these authors contributed equally

Corresponding Author:
Adrian Perez-Aranda, PhD
Department of Clinical and Health Psychology
Universitat Autònoma de Barcelona
Carrer de la Fortuna, s/n
Bellaterra, Barcelona, 08193
Spain
Phone: 34 935 81 18 55
Email: adrian.peara@gmail.com

Abstract

Background: The high levels of sustained stress that health professionals often experience are a significant risk factor for developing mental health problems, such as anxiety, depression, and somatic symptoms, that not only affect their well-being but also have major social and organizational consequences. Different interventions, including those based on third-wave psychotherapy principles (ie, mindfulness, compassion, and acceptance), have proven to be effective in reducing stress in this population. Among them, those delivered on the web constitute a promising alternative with notable advantages in accessibility and flexibility, but some adherence inconveniences may limit their efficacy.

Objective: This study aimed to evaluate the effectiveness of the MINDxYOU program, a web-based self-guided intervention based on third-wave psychotherapy principles, to reduce perceived stress and promote mental health in a sample of health professionals.

Methods: In a stepped-wedge cluster randomized design, 357 health professionals from health centers in Aragon and Málaga, Spain, were recruited. They were divided into 6 clusters—3 per region—and randomly assigned to 1 of the 3 sequences, each starting with a control phase and then transitioning to the intervention phase (the MINDxYOU program) after 8, 16, or 24 weeks. This self-guided, web-based program, designed to be completed over 8 weeks, included weekly contact (via WhatsApp, call, or email) from the research team to promote adherence. Participants were assessed on the web every 8 weeks for 5 assessments.

XSL•FO
RenderX

Perceived stress was the study's primary outcome, with additional measures of clinical factors (anxiety, depression, and somatization) and process variables (resilience, mindfulness, compassion, and acceptance).

Results: The program was initiated by 229 participants, 112 (48.9%) of whom were completers (ie, completed at least 3 of the 4 modules). Perceived stress demonstrated a significant reduction both when considering the entire sample (β=–1.08, SE 0.51; *P*=.03) and the sample of completers (β=–1.84, SE 0.62; *P*=.003). The proportion of participants reflecting "low stress" increased after the treatment (n=90, 46.6% vs n=100, 28.8% at baseline). Intracluster analysis revealed that pre- versus postintervention moderate effects were present in 2 clusters (Cohen *d*=0.46 and 0.62), and these were maintained in subsequent assessments. The linear mixed-effects models also showed that depression, anxiety, and somatization, as well as resilience, self-compassion, and some mindfulness facets, experienced significant improvements (*P*<.05) when comparing the intervention and control phases.

Conclusions: The MINDxYOU program was effective in reducing perceived stress and promoting mental health, as well as increasing resilience, mindfulness facets, and self-compassion. These effects suggest that participants experienced a tangible improvement that could potentially enhance their well-being. Adherence to the intervention was moderate, while program use was notable compared to similar interventions. Finding ways to promote adherence to the intervention would contribute to increasing the effectiveness of this program.

Trial Registration: ClinicalTrials.gov NCT05436717; https://clinicaltrials.gov/study/NCT05436717

International Registered Report Identifier (IRRID): RR2-10.1186/s12912-022-01089-5

(J Med Internet Res 2025;27:e59653) doi:10.2196/59653

KEYWORDS

eHealth; health professionals; stress; mental health; stress reduction; web-based intervention; stepped-wedge design; randomized controlled trial

Introduction

Background

Health professionals often experience high levels of sustained stress that constitute a risk for developing mental health problems, as recognized by the World Health Organization [1,2]. In recent years, the prevalence of psychological disorders such as anxiety, depression, and somatic symptoms has increased among this population, which has been linked to the extra efforts exerted by health workers and the extremely difficult situations to which they were exposed during the COVID-19 pandemic [3-6]. Although the pandemic was declared officially over in May 2023, its effects on the well-being of health professionals have been observed to be profound and lasting in many cases [7,8]. The implications of mental health problems in these professionals not only refer to their well-being but also include social and organizational consequences, such as increasing rates of absenteeism and turnover intention [9-11], as well as so-called quiet quitting [12]. Thus, despite the end of the COVID-19 pandemic, it is essential to continue to monitor health professionals' mental health and implement effective strategies to decrease their stress levels, thus preventing chronic stress and its consequences [13].

The effectiveness of interventions based on cognitive behavioral therapy (CBT) on stress reduction and mental health promotion has been widely proven for different types of health professionals [14-16]. Relaxation, cognitive restructuring, and social skill training, among others, are traditional CBT techniques aimed at modifying the interaction between thoughts, behaviors, and emotions. However, in recent decades, a different approach to CBT has been developed and tested with positive results. Referred to as the third wave of CBT psychotherapy, it is based on the promotion of mindfulness, acceptance, compassion, and spirituality [17]. Different systematic reviews have underscored the effectiveness of third-wave psychotherapies, including mindfulness-based interventions, compassion-based programs, and acceptance and commitment therapy, in reducing occupational stress as well as anxiety and depression while increasing resilience, including in health professionals [18-21].

While psychological interventions are typically delivered in an in-person format, eHealth interventions—examples of which are telehealth, mobile apps, and serious games—have proliferated over the last few years, partly due to the COVID-19 pandemic [22]. The systematic review conducted by López-Del-Hoyo et al [23] summarized the different types of eHealth programs that have been tested to promote mental health in health professionals and concluded that there was promising evidence of the effectiveness of self-guided, web-based, third-wave psychotherapy–based programs for reducing stress and promoting mental health. The self-guided format, which implies the lack of an external figure who delivers the contents of the intervention, offers a high degree of availability, flexibility, and convenience [24]. This could be particularly useful in the case of health professionals who often have to cope with the difficulty of reconciling work schedules with attending therapy sessions [25]. Indeed, this type of eHealth program has been associated with high user satisfaction and acceptance [26,27].

Nonetheless, some self-guided interventions have turned out to be ineffective for health professionals, and notably low adherence rates to these programs have been observed [23]. This may be due to generic and unappealing programs not addressing the particular needs of these individuals, although it has also been suggested that their lack of external guidance could contribute to higher attrition rates [26,28]. This would imply that the flexible and adaptable nature of self-guided programs can be both an advantage and an obstacle while

XSL•FO
RenderX

achieving significant improvements in stress and other mental health–related outcomes. Therefore, a mixed format, that is, a self-guided web-based program that includes occasional—preferably, remote and asynchronous—contact with an external figure, via WhatsApp messages or emails, for instance, could present a solution for retaining the main advantages of self-guided web-based programs while reducing the risk of attrition [23].

Objectives

The aim of this study was to test the effectiveness of a self-guided web-based program (MINDxYOU) for reducing the perceived stress of health professionals while promoting different mental health–related outcomes through the application of third-wave psychotherapy principles: mindfulness, compassion, and acceptance. Following the recommendations of previous authors [29], the MINDxYOU program was not designed as a generic program to be used by any given population; rather, its specific contents and examples relate to the day-to-day activities of health professionals (eg, compassion fatigue and common stressors). The program consists of 8 parts (sessions) to be completed within 8 consecutive weeks. During this time, weekly contact was scheduled with the research team via email, WhatsApp, or phone call to include a small degree of guidance in the intervention. The MINDxYOU program highlights the importance of systematic and continuous practicing of different types of meditation not only during the 8-week duration of the intervention but also afterward to promote the development of new healthy habits. It was hypothesized that on completing the web-based program, health professionals would experience a significant reduction in their stress levels as well as improvements in other clinical aspects (anxiety, depression, and somatization) and process variables (resilience, mindfulness, compassion, and acceptance).

Methods

Study Design

A stepped-wedge cluster randomized trial design [30] was used for this hybrid type 2, effectiveness and implementation study, which was conducted in 2 Spanish regions (the autonomous region of Aragon and the province of Malaga, in Andalusia). This design was chosen considering its suitability for studying both the effectiveness of the MINDxYOU program (reported in this work) and its implementation in health centers (to be reported in a different work) [31,32]. The stepped-wedge design allows all the study participants to receive the intervention and is particularly recommended when the intervention being tested has strong evidence of positive effects or is very unlikely to cause harm [30].

The participants were health professionals working for different types of health care institutions: hospitals, primary care centers, and other health facilities (eg, nursing homes and health and social care centers). The sample was divided into 6 clusters—3 per region—which were randomly allocated to 1 of the 3 possible sequences. They all started at the control phase and sequentially transitioned to the intervention phase (they were given access to the MINDxYOU program). This was a closed cohort study, given that all the participants started at the same time and were evaluated at 5 previously defined time points.

As detailed in the study protocol [33], the sample size calculation estimated 180 participants, 30 (16.7%) per cluster, for an expected moderate effect size of the intervention on the primary outcome, that is, perceived stress [34]. However, the ambitious dissemination of the project attracted the interest of a very high number of health professionals, and an effort was subsequently made to recruit 393 participants, of whom 357 (90.8%) would ultimately provide data and consequently form the study sample.

Recruitment and Inclusion Criteria

Between October 2022 and January 2023, the MINDxYOU project was disseminated using both direct and indirect techniques. Informative meetings were conducted for managers and health professionals working in health facilities (hospitals, primary care centers, and nursing homes) in Aragon and Malaga; posters and leaflets were placed at different health centers; and members of the research group were interviewed by regional media regarding the project, optimizing its visibility. Individuals interested in participating accessed the project's website [35] to complete a short form to verify whether they met the inclusion criteria as follows: (1) employment as a health professional (eg, physician, nurse, psychologist, or nursing assistant) or undergoing training in any health-related field (trainee); (2) aged between 18 and 70 years; (3) ability to understand Spanish; (4) digital literacy and access to a smartphone, tablet, or PC with an internet connection; and (5) prospects of continued employment at the same workplace for the following 6 months. Candidates who met these criteria were also asked on the form to provide their email address or telephone number for subsequent contact by the researchers.

Every candidate underwent a telephone interview conducted by a member of the research team, either a psychologist or mental health nurse trained for this purpose, to ensure the previously described inclusion criteria were met. In addition, the following exclusion criteria were assessed during the interview: (1) presenting with a disorder affecting the central nervous system; (2) diagnosis of a severe mental illness (severe depressive disorders, suicidal tendencies, bipolar disorders, panic disorders, anxiety or stress-related disorders, obsessive-compulsive disorders, and substance-related disorders) using the Mini-International Neuropsychiatric Interview (version 7.0.2) diagnostic interview [36]; (3) presenting with an uncontrolled medical condition or either an infectious or degenerative disease; and (4) having experience with third-wave psychotherapies (eg, having attended mindfulness courses in the previous year or practicing formal meditation regularly). Candidates who fit the inclusion profile were sent an email containing the study information and an attached informed consent form to be signed electronically and returned.

Procedures

Once the participants had been enrolled in the study (February 2023), they were assigned a code to access the evaluation platform (SurveyMonkey [SurveyMonkey Inc]) and ensure their anonymity throughout the study. An external researcher

XSL•FO
RenderX

unrelated to this project then performed the cluster randomization under 2 conditions: each sequence had to include a cluster from each region, and they had to be 2 different types of workplaces (hospitals, primary care centers, or other health centers). The baseline assessment took place over the last week of February 2023 and the first week of March 2023. Following the stepped-wedge design, this assessment was repeated every 8 weeks, which was the expected duration of the program. There were 5 assessment points in total.

Emails were sent to all participants after each assessment to indicate whether they had been granted access to the MINDxYOU program or still had to wait a certain amount of time (ie, they remained in the control phase). Once they were allowed access, they first watched a 3-minute video tutorial on how to log into the platform and use it. The introduction to the program (module 0) explained to the participants the format of the intervention (4 modules, each containing 2 sessions). Although visualizing the contents of each session (text and videos) would take approximately 1 hour, it was recommended that participants dedicate an entire week to each session before moving on to the next, because every session included meditation exercises that needed to be practiced on consecutive days to be learned and incorporated into a routine if they were to produce stress-relief effects. Therefore, the program was expected to be completed in 8 weeks (1 session per week). Whenever participants indicated their intention to start a new session, the platform would present a brief on the web form asking how many days they had been practicing the exercises from the previous session. This was meant to remind them of the recommendation to dedicate 1 full week to the series of exercises in each session before moving on to the next one. Nevertheless, all participants were given access to the following sessions when they completed the form, regardless of their dedication to the previous session.

Once per week during the 8-week duration of the program, all (N=357, 100%) participants were contacted via their preferred means (email, WhatsApp, or phone call, as indicated when they signed the informed consent) by a member of the research team (SF-M, CA-L, LC-G, or VC). The purpose of this contact was to motivate the participants to adhere to the intervention as per the recommendations and identify any difficulties with the program regarding its contents or use of the platform. If any of these were communicated by participants, the researcher would provide help to clarify and resolve the issue. The weekly contacts ended after the eighth week, regardless of the degree to which each participant had completed the program by that time. Nonetheless, the participants were informed that they could continue to use the MINDxYOU program and were encouraged to keep practicing what they had learned. Follow-up emails were sent reminding every study participant of this information.

Intervention

The MINDxYOU program is a self-guided web-based intervention delivered via a digital platform that consists of 4 modules, each of them divided into 2 parts (sessions). The contents of the intervention are related to third-wave psychotherapies (mindfulness, compassion, and acceptance), which have presented strong evidence of efficacy for reducing stress and promoting mental health [18]. All the contents of the MINDxYOU program are based on previously developed and validated programs, such as mindfulness-based stress reduction [37], Smiling is Fun [38], and attachment-based compassion therapy [39]. Following the suggestions of previous studies [23,29], the MINDxYOU program includes a series of adaptations to make it more specific to health professionals (eg, examples of typical stressors that occur in health care settings). Both formal and informal practices are presented during the program: downloadable audio guides are provided for formal practice, while informal exercises are linked to the promotion of healthy habits, such as physical activity, healthy diet, and good sleep habits, in addition to socializing in module 4. The intervention emphasizes the importance of including meditation exercises in a daily routine and developing healthy habits that can have a stress-relief effect. Consequently, participants are encouraged to download the materials and continue to practice even after the completion of the program. A summary of the contents and exercises proposed for each session is presented in Multimedia Appendix 1.

Measurements

Overview

Sociodemographic questions (age, gender, marital status, and education level) and work-related questions (type of contract, occupation, management position, trainee status, and salary) were presented to the participants at baseline assessment. These questions were not presented again at subsequent assessments, although participants were asked whether they still worked for the same health care institution (and if not, the one at which they were currently employed). In addition to these questions, the primary and secondary outcomes and process variables were included at every assessment point.

Primary Outcome

The Perceived Stress Scale (PSS) [40] was the primary outcome of this study. It is a 10-item questionnaire that measures how unpredictable, uncontrollable, and overloaded the individual has felt their life to be over the previous month on a 5-point Likert-type scale. Scores range between 0 and 40, with higher scores reflecting higher levels of perceived stress. Cutoff points have been established to determine 3 categories of stress levels: low (0-13), moderate (14-26), and high (27-40). Our study made use of the Spanish adaptation [41], and the PSS for our sample showed very high internal consistency (α_{T1}=0.89 and $\alpha_{T2\text{-}T5}$=0.91).

Secondary Outcomes

The Patient Health Questionnaire-9 (PHQ-9) [42] is a 9-item scale that rates the frequency of depressive symptoms during the previous 2 weeks using a Likert-type scale (from 0=not at all to 3=nearly every day). The total score ranges between 0 and 27, with higher scores indicating higher severity of depression. The Spanish adaptation of the PHQ-9 was used for this study [43], and we observed good internal consistency in our sample (α_{T1}=0.83, α_{T2}=0.88, α_{T3}=0.86, α_{T4}=0.89, and α_{T5}=0.87).

XSL•FO
RenderX

The Generalized Anxiety Disorder-7 (GAD-7) [44] is a 7-item, self-report measure to assess the intensity of anxiety symptoms over the past 2 weeks. Using a 4-point Likert-type scale, the total score can range between 0 and 21, with higher values indicating more severe anxiety symptoms. The Spanish version of the GAD-7 [45] was used for this study, and excellent psychometric properties were observed in our sample (α_{T1}=0.89, α_{T2}=0.92, α_{T3}=0.90, α_{T4}=0.92, and α_{T5}=0.91).

The Brief Symptoms Inventory-18 (BSI-18) [46,47] is a 5-point Likert-type scale designed to offer rapid screening for the symptoms of psychological disorders (somatization, depression, and anxiety). Scores on the 18 items are summarized on the Global Severity Index, which ranges from 0 to 72, with higher scores reflecting more severe conditions. The Spanish version of the BSI-18 was used [48] and showed excellent internal consistency for our sample (α_{T1}=0.90, $\alpha_{T2\text{-}T3}$=0.92, α_{T4}=0.94, and α_{T5}=0.92).

Process Variables

In total, 5 process variables were evaluated at each assessment point: resilience, mindfulness, self-compassion, compassion for others, and acceptance. Resilience was measured using the Connor-Davidson Resilience Scale (CD-RISC) [49], a 10-item scale whose total score ranges between 0 and 40, with higher scores indicating higher levels of resilience. Mindfulness was assessed using the Five Facet Mindfulness Questionnaire-15 (FFMQ-15) [50], which includes 5 subscales: observing, describing, acting with awareness, nonjudging of inner experience, and not reacting to inner experience. Each FFMQ-15 subscale includes 3 items, scored on a Likert-type scale (1-5), and the average score is calculated (higher scores indicate higher levels of the mindfulness facet). Both self-compassion and compassion for others were assessed using the Sussex-Oxford Compassion Scales (SOCS) [51]; each scale consists of 20 items, and its score is calculated by adding the scores of its items (ranging between 20 and 100, with higher scores indicating higher levels of compassion). Finally, the Acceptance and Action Questionnaire-II (AAQ-II) [52] was used to evaluate acceptance as opposed to experiential avoidance. The scale presents 7 items scored on a 7-point Likert-type scale, and the total score ranges between 7 and 49, with higher scores indicating higher experiential avoidance (ie, lower acceptance).

For this study, the Spanish version of the CD-RISC [53], FFMQ-15 [54], SOCS [55], and AAQ-II [56] were used. All showed good or excellent psychometric properties for our sample ($\alpha \geq 0.80$ at every assessment point), except for FFMQ-observing (α_{T1}=0.74, α_{T2}=0.55, α_{T3}=0.70, and $\alpha_{T4\text{-}T5}$=0.74), FFMQ-describing ($\alpha_{T1,T2,\text{ and }T5}$=0.77), and FFMQ-nonreacting (α_{T1}=0.71, $\alpha_{T2\text{-}T3}$=0.58, α_{T4}=0.60, and α_{T5}=0.71).

Adherence to the MINDxYOU Program

Adherence to the intervention was only determined by assessments completed after the participants had been given access to the MINDxYOU program, in which they were asked to state (1) the number of modules they had completed (where they had previously completed the intervention, they were asked how often they had revisited the program in the previous 2 months), (2) an average number of days per week they had engaged with formal practice (including the exercises they performed and their duration), and (3) an average number of days per week they had engaged with informal practice (including the exercises they had performed). In addition, every time participants wanted to access a new session while on the program, they had to complete a brief form indicating how many days they had practiced the exercises that were presented in the previous session.

Data Analysis

First, we describe participant flow and compliance with the intervention. Frequencies and percentages were used to summarize the number of participants who completed all assessments, initiated the program, and completed at least 3 of the 4 modules. The chi-square test or ANOVA tests were conducted to assess between‐cluster differences as well as potential differences between (1) participants who dropped out and those who completed all the assessments, (2) participants who ultimately initiated the program and those who did not, and (3) the individuals who were considered to have completed the program and all the others.

Sociodemographic and clinical baseline data as well as intervention adherence-related data were described using means and SDs for continuous variables and frequencies and percentages for categorical variables. Chi-square tests and ANOVA were conducted to assess between-cluster differences in baseline characteristics.

The primary analysis (the effects of the intervention on perceived stress) was performed using linear mixed-effects models with cluster-specific random effects and time-specific fixed effects. We adjusted the model to include the following baseline imbalanced covariates (covariate-adjusted analysis): cluster, age, education level, type of contract, occupation, trainee status, salary, and the baseline level of the outcome. A cluster-robust sandwich variance estimator with small-sample corrections was implemented to obtain bias-corrected SEs [57]. We report the effect estimate, SE, and 95% CI. These analyses were conducted following 2 different approaches: the intention-to-treat approach (all the study participants included) and the per-protocol approach (program noncompleters excluded in postintervention assessments). The same strategy was used to analyze the effects on secondary outcomes and process variables. In addition, for the primary outcome, intracluster analyses under both approaches (intention-to-treat and per-protocol approaches) were computed using 2-tailed Student *t* test, whereas Cohen *d* was used to estimate the effect size.

An α level of .05 was set, using a 2‐tailed test. For the primary analysis, a completely missing at random approach was chosen given the characteristics of the study design and the sample size. For the intracluster analyses, we initially used raw data; however, we later performed a simple imputation (mean of nearby points) to assess potential effects under both the intention-to-treat and per-protocol. We applied the Benjamini-Hochberg correction for multiple comparisons in every case other than the primary analysis, following the recommendations by Li et al [58]. This is a procedure to detect

XSL•FO
RenderX

false discovery designed to overcome the limitations that the other tests typically used have shown [59,60]. Data analyses were computed using SPSS (version 26.0; IBM Corp) and R (R Foundation for Statistical Computing) statistical software.

Ethical Considerations

The research ethics committee of the Autonomous Community of Aragon and the ethics and research committee of Northeast Malaga evaluated and approved the study protocol in July 2022 (PI22/341). All procedures performed in this study adhered to the 1964 Declaration of Helsinki and its most recent amendments (seventh revision, adopted by the 64th World Medical Association General Assembly, Fortaleza, Brazil). A signed informed consent was obtained from all participants after they were informed of the study procedures, potential risks, and their right to withdraw at any time from the study. The participants did not receive any compensation. The study ensured the confidentiality of the information collected by fully complying with the Spanish Data Protection and Digital Rights Act 3/2018 and the European Union's General Data Protection Regulation (GDPR). To maximize data protection, the technological infrastructure consisted of two independent systems accessing separate databases, ensuring the complete disaggregation of users' personal data from their clinical records. This study was registered in ClinicalTrials.gov on June 29, 2022 (NCT05436717), and this work follows the CONSORT (Consolidated Standards of Reporting Trials) extension for the stepped-wedge cluster randomized trial [61] and the CONSORT-eHEALTH checklist [62].

Results

Flow and Compliance

In total, 393 participants were included in the study, although 36 (9.2%) of them did not complete any assessment. Therefore, the study sample comprised 357 participants (Figure 1 for cluster distribution). Some significant differences (P<.05) were found between individuals who completed all the assessments (n=131, 36.7%) and those who did not (n=226, 63.3%; Multimedia Appendix 2). Participants in clusters 4 and 6 had a higher rate of assessment incompletion (χ^2_5=14.5; P=.001), as did those from Malaga (χ^2_1=9.8; P=.002).

Figure 1. Study flowchart. PC: primary care. C1-6: cluster 1-6.

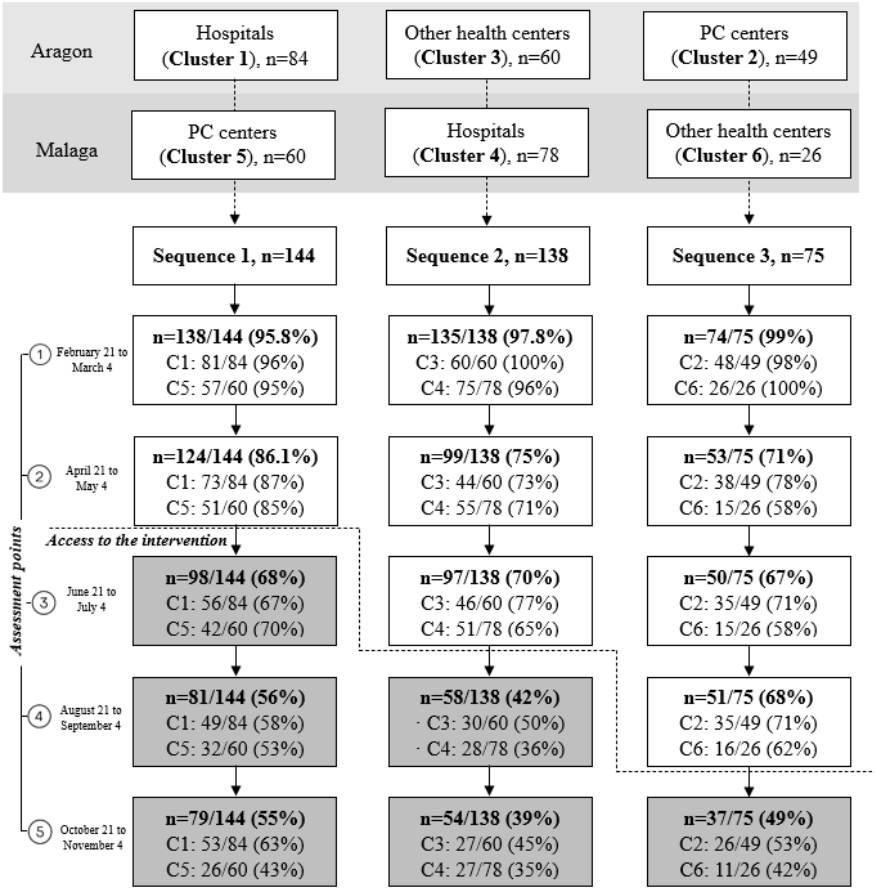

As shown in Table 1, the MINDxYOU program was initiated by 64.1% (229/357) of the participants. After applying the Benjamini-Hochberg correction, a significantly higher proportion of participants who did not initiate the program was observed among those coming from Malaga (χ^2_1=9.9; P=.002) and clusters 4 and 6 (χ^2_5=19.9; P=.001; Multimedia Appendix 3). With regard to program completion, 48.9% (112/229 initiators) of the participants completed at least 3 modules, and 43.7% (100/229) completed all of them. Here, no significant differences were found when comparing completers with noncompleters (Multimedia Appendix 4). On average, initiators of the program completed 2.2 (SD 1.8) modules.

XSL•FO
RenderX

Table 1. Completion of the MINDxYOU program.

	Started the program, n (%)	Completed module 1, n (%)	Completed module 2, n (%)	Completed module 3, n (%)	Completed all 4 modules, n (%)
Cluster 1 (n=84)	65 (77)	45 (54)	42 (50)	35 (42)	31 (37)
Cluster 2 (n=49)	31 (63)	23 (47)	19 (39)	16 (33)	15 (31)
Cluster 3 (n=60)	42 (70)	29 (48)	24 (40)	21 (35)	18 (30)
Cluster 4 (n=78)	36 (46)	26 (33)	19 (24)	16 (21)	14 (18)
Cluster 5 (n=60)	41 (68)	26 (43)	20 (33)	20 (33)	18 (30)
Cluster 6 (n=26)	14 (54)	10 (38)	6 (23)	4 (15)	4 (15)
Total (N=357)	229 (64.1)	159 (44.5)	130 (36.4)	112 (31.4)	100 (28)

The 229 participants who initiated the program dedicated an average of 4.3 (SD 2.6) days to each session. After the intervention, participants (n=148) reported an average formal practice frequency of 3.3 (SD 1.5) days per week, with a mean duration of 18.8 (SD 11.6) minutes per exercise. Informal exercises were performed an average of 3.5 (SD 2) days per week. In the second postintervention assessment (clusters 1, 3, 4, and 5), participants (n=94) reported a formal practice frequency of 2.5 (SD 1.4) days per week, 15.25 (SD 10.2) minutes per exercise, and 3.3 (SD 1.9) days per week of informal practice. In the third postintervention evaluation (clusters 1 and 5, n=59), the frequency of formal practice remained at 2.6 (SD 1.7) days per week, 13.8 (SD 11.3) minutes per exercise, and 3.5 (SD 2.2) days per week of informal practice. No significant differences between clusters were observed in any case (P>.05 in all cases).

Baseline Characteristics

As shown in Multimedia Appendix 5, the sample comprised mostly women (297/347, 85.6%) who were middle aged (mean 45.01, SD 11.17) and married (260/347, 74.9%), and most had completed university-level studies (304/347, 87.6%). Some between-cluster differences were observed. Cluster 6 was younger (F_5=5.66; P<.001), and cluster 3 presented a lower education level (χ^2_{10}=68.8; P<.001).

Most participants were public-sector employees on permanent contracts (159/347, 45.8%), and the most common professions in our sample were physicians (147/347, 42.4%) and nurses (93/347, 26.8%). Only 15% (52/347) of participants occupied a management position at their institution, and 9.8% (34/347) were trainees for some health profession. Most (146/347, 42.1%) participants had a salary equivalent to between 1 and 2 times the minimum wage (between €1000 and €2000 per month; €1=US $1.04). Again, some between-cluster differences were observed. Cluster 6 had a very low number of functionaries (χ^2_5=97.3; P<.001) and physicians (χ^2_{25}=131.2; P<.001) compared to the other clusters; trainees were not represented in clusters 3 and 6 (χ^2_5=15.6; P=.008), and these clusters had a lower salary (χ^2_{15}=62.3; P<.001).

With regard to the clinical variables at baseline (Multimedia Appendix 5), our sample presented moderate levels of perceived stress, as indicated by the mean score in the PSS (mean 16.88, SD 6.33) and the fact that most (225/347, 64.8%) participants presented scores that fell under the moderate stress category (ie, 14-26 points of 40). Low levels of depression (PHQ-9: mean 6.25, SD 4.36), anxiety (GAD-7: mean 7.05, SD 4.23), and psychopathological symptoms (BSI-18: mean 12.34, SD 9.88) were observed. Participants showed a notable degree of resilience (CD-RISC: mean 27.31, SD 6.79), moderate levels of mindfulness facets (FFMQ-15 scores around 3), lower levels of self-compassion (SOCS-self: mean 53.53, SD 10.24) than compassion for others (SOCS-Others: mean 61.90, SD 8.77), and moderate-low levels of experiential avoidance, that is, moderate-high levels of acceptance (AAQ-II: mean 20.89, SD 8.33). Self-compassion (SOCS-self) was the only variable that showed significant between-cluster differences after applying the Benjamini-Hochberg correction (F_5=3.20; P=.008).

Effects on Perceived Stress

PSS preintervention scores were 16.63 (SD 6.89) and were reduced to 14.02 (SD 6.69) after the intervention. The change was significant under both analytical approaches (Table 2), although stronger under the per-protocol approach. Some of the covariates included in the model were found to be significant: the sequence in which the cluster was allocated, type of contract, occupation, salary, and baseline PSS level. While at baseline most participants referred to moderate stress (225/347, 64.8%) and fewer reported low stress (100/347, 28.8%), after the treatment these proportions reflected a general improvement: 51.3% (99/193) showed moderate stress and 46.6% (90/193) reported low stress. The intracluster analysis revealed that under the intention-to-treat approach (Multimedia Appendix 6), only cluster 1 experienced a significant pre- versus postintervention effect of moderate size (Cohen d=0.46), while a larger pre- versus second postintervention assessment effect was observed in clusters 1 (Cohen d=0.68) and 5 (Cohen d=0.62). These effects were maintained in subsequent assessments. Under the per-protocol approach, the same effects were observed with larger effect sizes (Multimedia Appendix 7). Figure 2 shows the evolution of perceived stress along the 5 assessment points for each cluster (intention-to-treat approach).

XSL•FO
RenderX

Table 2. Effects of the intervention comparing intervention phase versus control phase (linear mixed-effects models) under the intention-to-treat and per-protocol approaches.

Outcome	Intention-to-treat approach		Per-protocol approach	
	Estimate (SE; 95% CI)	*P* value	Estimate (SE; 95% CI)	*P* value
PSS[a,b,c,d,e,f]	–1.08 (0.51; –2.08 to –0.08)	.03[g]	–1.84 (0.62; –3.04 to –0.63)	.003[g]
PHQ-9[a,c,d,e,h]	–1.43 (.030; –2.02 to –0.84)	<.001[g]	–1.72 (0.24; –2.19 to –1.26)	<.001[g]
GAD-7[a,c,d,e,i]	–1.31 (0.41; –2.12 to –0.51)	.001[g]	–1.65 (0.17; –1.98 to –1.31)	<.001[g]
BSI-18[j]				
Somatization[a,c,d,e]	–0.73 (0.25; –1.22 to –0.24)	.004[g]	–0.83 (0.28; –1.38 to –0.29)	.003[g]
Depression[a,d,e]	–1.06 (0.28; –1.62 to –0.51)	<.001[g]	–1.50 (0.36; –2.20 to –0.80)	<.001[g]
Anxiety[a,c,d,e]	–0.76 (0.40; –1.55 to 0.02)	.06	–1.34 (0.16; –1.65 to –1.03)	<.001[g]
GSI[a,d,e,k]	–2.30 (0.82; –3.92 to –0.69)	.005[g]	–3.60 (0.56; –4.71 to –2.50)	<.001[g]
CD-RISC[a,c,d,e,l]	1.53 (0.47; 0.62 to 2.45)	.001[g]	2.14 (0.47; 1.22 to 3.06)	<.001[g]
FFMQ[m]				
Observing[a,b,c,d,e]	0.16 (0.08; –0.01 to 0.32)	.07	0.36 (0.11; 0.14 to 0.57)	.001[g]
Describing[a,c,d,n,o]	0.07 (0.14; –0.20 to 0.36)	.61	0.03 (0.09; –0.15 to 0.20)	.77
Acting with awareness[a,d,e]	0.16 (0.09; –0.02 to 0.34)	.08	0.22 (0.05; 0.11 to 0.32)	<.001[g]
Nonjudging[a,c,d,e]	0.27 (0.10; 0.08 to 0.47)	.007[g]	0.34 (0.09; 0.17 to 0.52)	<.001[g]
Nonreacting[a,c]	0.18 (0.08; 0.03 to 0.33)	.02[g]	0.20 (0.07; 0.06 to 0.34)	.006[g]
SOCS[p]				
Others[a,c,d,e]	–0.17 (0.70; –1.55 to 1.20)	.80	–0.07 (0.76; –1.56 to 1.42)	.93
Self[a,c,e]	2.72 (0.66; 1.42 to 4.03)	<.001[g]	2.97 (0.82; 1.36 to 4.57)	<.001[g]
AAQ-II[a,c,d,e,q]	–1.51 (0.74; –2.95 to –0.07)	.04	–1.54 (0.83; –3.17 to 0.09)	.06

[a]Baseline level of the outcome was significant in the model.

[b]Sequence was significant in the model.

[c]Type of contract was significant in the model.

[d]Occupation was significant in the model.

[e]Salary was significant in the model.

[f]PSS: Perceived Stress Scale.

[g]Effects that remained statistically significant (*P*<.05) after applying the Benjamini-Hochberg correction (applied in the secondary outcomes and process variables).

[h]PHQ-9: Patient Health Questionnaire-9.

[i]GAD-7: Generalized Anxiety Disorder-7.

[j]BSI-18: Brief Symptoms Inventory-18.

[k]GSI: Global Severity Index.

[l]CD-RISC: Connor-Davidson Resilience Scale.

[m]FFMQ: Five Facet Mindfulness Questionnaire-15.

[n]Age was significant in the model.

[o]Education level was significant in the model.

[p]SOCS: Sussex-Oxford Compassion Scales.

[q]AAQ-II: Acceptance and Action Questionnaire-II.

XSL•FO
RenderX

Figure 2. The evolution of Perceived Stress Scale scores across the study assessments per cluster. Squares in grey (transparent) indicate the periods in which the participants had access to the intervention.

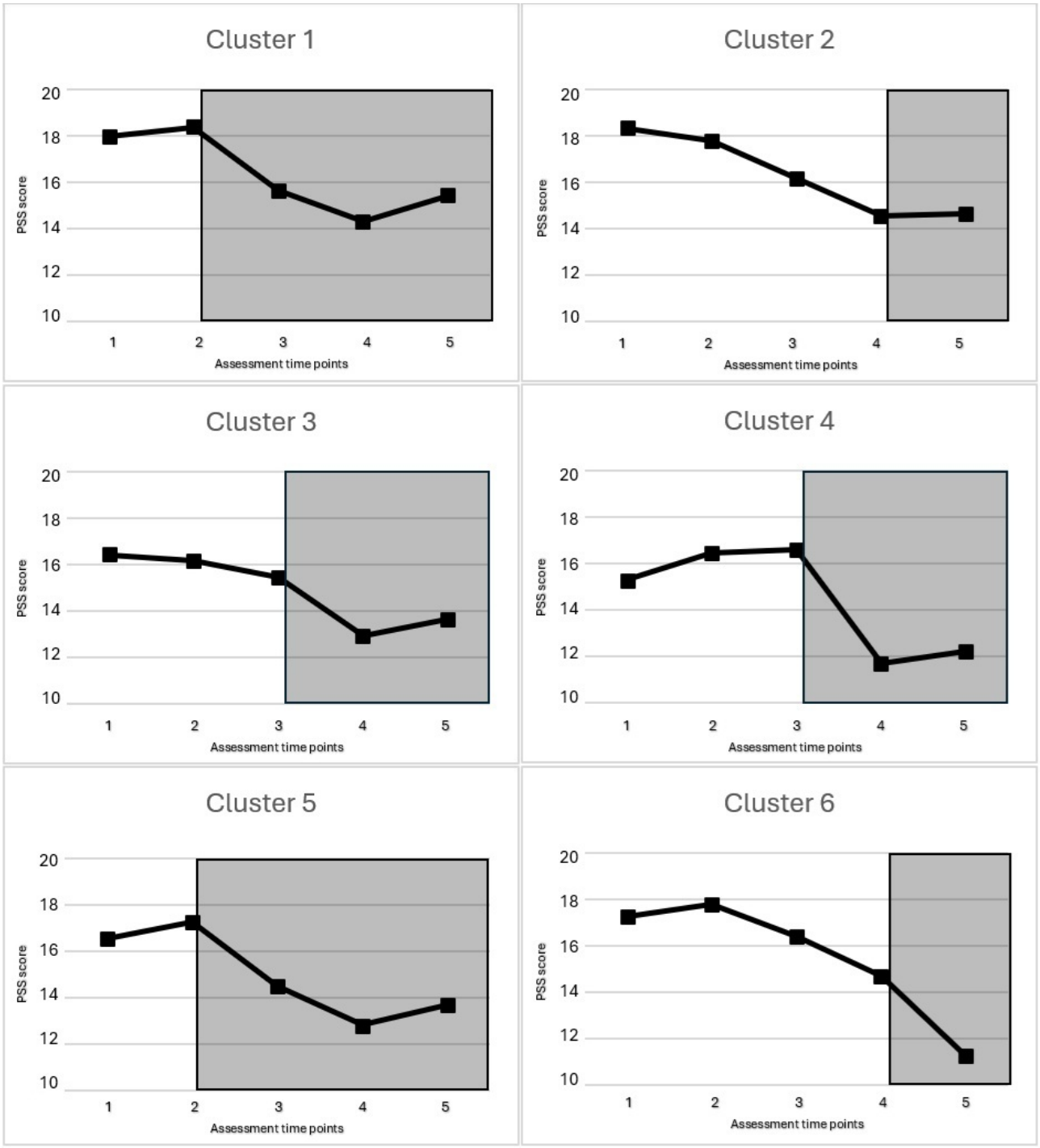

Effects on Secondary Outcomes and Process Variables

As shown in Table 2, the effects of the intervention in all the secondary outcomes (PHQ-9, GAD-7, and BSI) were significant according to the linear mixed-effects models under the intention-to-treat approach, except for BSI-anxiety. Certain covariates played a significant role in the models: the baseline level of the outcome, type of contract, occupation, and salary. Under the per-protocol approach, all the secondary outcomes, including BSI-anxiety, presented a statistically significant effect, and the effects of the covariates were maintained in all cases.

Some process variables experienced significant improvement during the intervention phases: under the intention-to-treat approach, resilience (CD-RISC), the mindfulness facets of nonjudging and not reacting to the inner experience (FFMQ-15), and self-compassion (SOCS-self) presented significant effects. These were also observed under the per-protocol approach, along with significant effects on the mindfulness facets of observing and acting with awareness (FFMQ). All these effects were in the expected direction (improved during the intervention phase), and some covariates presented significant effects in each case (Table 2): baseline level of the outcome, the sequence in

XSL•FO
RenderX

which the cluster was allocated, type of contract, occupation, salary, age, and education level.

Discussion

Principal Findings

The MINDxYOU program led to a significant reduction in health professionals' perceived stress. Given the implications of sustained stress on these professionals' well-being, this reduction can have a notable impact on preventing mental health issues, such as burnout syndrome and its associated health and societal consequences [9-13]. This finding is in line with those of previous studies conducted on smaller samples in which self-guided web-based programs implementing third-wave psychotherapy contents produced significant improvements in health professionals' stress. Barrett and Stewart [63] reported that a web-based acceptance and commitment therapy program was as effective as a web-based CBT program, and the mindful and self-compassion program was superior when compared to a waitlist control group [24]. Similarly, 2 uncontrolled studies observed intragroup effects of similar interventions [64,65]. As far as the duration of the effects of the MINDxYOU program is concerned, these were maintained in those clusters where postintervention effects were found (clusters 1 and 5) across subsequent assessments, which suggests that the program's effects might not be limited to the short term. Previous studies reported similar findings [64,66], but methodologically sound studies are required to ensure the long-term benefits of self-guided web-based programs in this population.

Likewise, the program produced significant improvements in mental health–related outcomes. Anxiety and depressive symptomatology were reduced after the intervention, which is in line with previous eHealth CBT-based interventions [67,68] and third-wave psychotherapies [69,70]. However, the effects of such interventions on somatic symptoms have not been as widely studied in this population, despite the numerous studies that reported increased somatization in health professionals during the COVID-19 pandemic [6]. The MINDxYOU program produced a significant reduction in somatic symptoms, which is in line with studies testing eHealth interventions on patients with somatic symptom disorder [71,72], but these findings should be extended in further studies conducted on health professionals. Considering the high frequency with which health professionals experience anxiety, depressive, and somatic symptoms, especially following the COVID-19 pandemic [3-6], the effectiveness of this intervention in reducing these symptoms represents a significant contribution to mental health promotion and supports its implementation in real-world settings.

It is hypothesized that improvements in perceived stress, depressive symptomatology, anxiety, and somatization might have been mediated by the changes produced in process variables, such as resilience, different mindfulness facets, and self-compassion. The MINDxYOU program aimed to promote mindfulness, that is, present-focused nonjudging attention, and self-compassion, that is, a kind attitude toward the self, both associated with reductions in anxiety and depression [73,74]. Third-wave psychotherapies have also been found to produce a significant increase in resilience [21], although not as clearly in the case of health professionals [75]. Interestingly, resilience mediates the association between mindfulness and self-compassion with depressive symptomatology [76], which highlights the inseparable nature of our study outcomes. We will be exploring the mediation paths of the MINDxYOU program in a future work to complement these findings. It is worth noting that other process variables, such as compassion for others and acceptance, did not experience significant improvements, despite the observation in previous studies that mindfulness-based interventions and eHealth programs might foster these outcomes [77,78]. Therefore, it could be hypothesized that our sample's relatively high baseline degrees of acceptance and compassion for others produced a ceiling effect.

All in all, our findings seem to add evidence to the potential of self-guided, third-wave psychotherapy–based eHealth interventions to help health professionals with their stress and improve their mental health. This improvement can significantly impact job satisfaction and work performance, ultimately benefiting the entire population [33]. The self-guided format offers high levels of availability, flexibility, and convenience [26,27], and third-wave psychotherapies have widely proved their effectiveness on numerous health outcomes and in different populations, including health professionals [18,19,23]. Nevertheless, it should be noted that the effects of the program on every outcome were stronger when considering the per-protocol approach (the sample of intervention completers). Different studies have observed a significant association between an eHealth program's dose and its effects on stress, among other outcomes [24,66,79,80], which highlights the importance of fostering treatment adherence.

In this regard, it is noteworthy that some differences were found between the individuals who initiated the program and those who did not; these reflect that some clusters were more active than others, suggesting different degrees of motivation toward the intervention. Moreover, program completion also showed some differences between clusters. Therefore, it is hypothesized that waiting time (the time between the baseline evaluation and accessing the program) may have been a barrier for some participants in our study, as also identified by previous works [81]; the stepped-wedge design meant that while some participants (clusters 1 and 5) accessed the intervention after a relatively short time (8 weeks), others had to wait up to 16 weeks (clusters 3 and 4) or 24 weeks (clusters 2 and 6). Furthermore, clusters 3 and 4 were given access to the intervention over the summer, which was probably not helpful for adherence as it coincided with the vacation period. In a future study, we will explore how these and other factors, such as common job stressors (eg, workload and work-family balance) [82], could have hindered adherence to the MINDxYOU program.

Hence, the study design might have been a hindrance to part of our sample initiating the program or continuing with it. Once the intervention had commenced, adherence to the program was moderate. Approximately 1 (50%) in every 2 participants who started the MINDxYOU program completed it, and they reported notable indicators of autonomous practice (around 3 days of formal practice per week, 19 minutes per practice, and a similar frequency for informal practice). To our knowledge, no previous

XSL•FO
RenderX

studies testing the effectiveness of self-guided, third-wave psychotherapy–based web-based interventions have reported the amount of meditation practice; however, a recent systematic review concluded that patients with chronic pain who are undergoing in-person programs practiced an average of 4 days per week and 27 minutes per exercise [83]. These findings could suggest that web-based interventions might have a smaller yet sufficient impact on autonomous practice compared to third-wave psychotherapy programs with a conventional in-person format, but this remains a hypothesis to be explored in the future.

With regard to program use, MINDxYOU was accessed by two-thirds (229/357, 64.1%) of our sample with an average of 2.2 (SD 1.8) modules completed (out of 4), which meant logging into the web-based platform at least 4 times, because each module contained 2 sessions. This can be considered an acceptable rate of use and higher than that observed in the few studies that have previously reported on this aspect. Ketelaar et al [84] found that only 20% of their participants logged into their eHealth programs at least once, and none completed the intervention. Dutton and Kozachik [85] and Hersch et al [86], who tested the 8-week self-guided BREATHE web-based intervention on nurses, observed an average of 2 to 2.5 logins. Kemper et al [87] and Kemper and Khirallah [88] found that participants engaged with a median of 3 modules out of 12 (mind-body skills training).

The weekly contact with the research team may have been a reason for the notable use of the MINDxYOU program. This contact was included as part of the intervention based on the recommendations of previous studies [23,89] to prevent low adherence rates. However, emails and text messages, which were the most common form of contact option chosen, were often one-sided (ie, unanswered by the participant), which leaves room to doubt whether these contacts were useful in their function of enhancing adherence to the program. It could also be argued that the specific nature of the MINDxYOU program, which was designed particularly for health professionals, could be a reason for its good rates of use. To our knowledge, only 5 interventions with specific contents for health professionals have been tested [85,86,90-93], with heterogeneous effects, and only the BREATHE program reported its use, which was relatively low, as previously explained. Therefore, the benefits of implementing specific interventions for this population over generic ones remain a topic of debate and will require further studies.

Strengths and Limitations

This study is part of the MINDxYOU project, which focuses on the well-being of health professionals, a topic that gained a notable degree of attention from researchers during the COVID-19 pandemic but is now at risk of again being underestimated and understudied now that the COVID-19 pandemic is over. This work reports part of the findings of the project that will need to be complemented with others (eg, mediation analysis, cost-effectiveness of the intervention, and implementation in health facilities). Altogether, these findings will contribute to the understanding of health professionals' needs and address them using innovative, evidence-based interventions.

Another notable strength of this study is its sample size. An effort was made to recruit almost twice the sample that was required according to the sample size calculation [33], which turned out to be very valuable considering the high rate of attrition during the study. While previous randomized controlled trials have tested the effectiveness of eHealth programs on health professionals, only 1 study was conducted on a larger sample [93]; however, no significant effects of the intervention—a mobile app (PsyCovid)—were found compared to the control group, which is in line with other studies testing mobile apps [23]. Our study analyzed the effects of a self-guided web-based intervention, a format that has shown good results in previous studies [24,63-65], and was developed based on previously validated interventions but with a special focus on health workers' stressors, which was suggested to be an important aspect for increasing adherence to the program [23,29]. In addition, our study provides data on adherence to the intervention in terms of frequency and duration of autonomous practice, which is a valuable indicator of the degree to which individuals (in this case, health professionals) can implement what they learn in self-guided web-based interventions.

Nevertheless, this study does have some notable limitations. First, the study may have been affected by self-selection bias, as only individuals who expressed interest in the project, potentially due to experiencing certain levels of stress or having a particular interest in the meditation techniques offered by the intervention, were eligible to participate. In addition, some of the criteria that were used to include or exclude potential participants may have affected the representativeness of our sample. On one hand, individuals diagnosed with a severe mental health condition (depression with active suicidal ideation, bipolar disorder, etc) were excluded and redirected to a specialized service. By contrast, having prospects of staying at the same workplace for the following 6 months was an inclusion criterion because it was considered relevant to studying the program implementation, but this led to health professionals with unstable employment situations being excluded. Job instability is clearly associated with stress and, added to the exclusion of individuals with severe mental health conditions, this could partially explain the relatively low levels of psychopathological symptoms (anxiety, depression, and somatization) in our sample, which probably had an overrepresentation of healthy individuals with stable employment. Similarly, the study was restricted to health professionals working in 2 Spanish regions, and generalizations in this regard could also be limited as each autonomous region in Spain manages its own health and social care provision. Therefore, regional differences may exist in terms of health resources or expenditure, which could undoubtedly affect the day-to-day activities of health professionals in each region and cater to their particular needs.

It should also be noted that our stepped-wedge cluster randomized design presented some limitations. The first one is that the clusters were not balanced in terms of sample size, which was due to the type of workplace. While hospitals employ a large number of health professionals, primary care centers

XSL•FO
RenderX

and nursing homes tend to have smaller staffing needs. Moreover, because of this design, even though every study participant was assessed both before and after the intervention, not all of them (ie, clusters 2 and 6) were assessed again after the first postintervention evaluation. This hinders the extraction of conclusions regarding the medium- or long-term effects of the intervention, which is a very important aspect to be considered in future studies. The different waiting times for each cluster can also be considered a limitation because it could have negatively affected the motivation of some study participants toward the intervention, hindering adherence. In this regard, it is important to note that a notable portion of the study participants (128/357, 35.9%) did not initiate this self-guided program, which may reflect the presence of barriers (eg, multitasking, heavy workloads, and constant evaluations) that will be explored in future research. Another limitation of this study is the use of self-reported measures, as these imply a certain bias, although most of them presented strong internal validity at the different assessment points. Complementing the assessment with some physiological outcomes could be considered in future studies.

Conclusions

Our findings support the effectiveness of the MINDxYOU program for reducing stress and promoting mental health (ie, reducing depressive, anxiety, and somatic symptoms and increasing resilience, mindfulness facets, and self-compassion) in health professionals. The intervention dose seems to play a significant role in its effects, which highlights the need for increased adherence to the program. Although we hypothesize that the study design might have hindered adherence in some cases, program use was notable compared to that reported by previous studies testing similar eHealth programs, which could be partially due to the inclusion of a small degree of guidance (ie, weekly contact from the research team during the intervention). These findings will need to be complemented with the study of the cost-effectiveness of the MINDxYOU program and its implementation in different health facilities to consider its potential as an efficient and reliable tool for delivery in Spanish health care institutions.

Acknowledgments

AP-A is grateful to a Sara Borrell postdoctoral contract (CD20/00181) from the Instituto de Salud Carlos III (Spanish Ministry of Science and Innovation), which allowed him to collaborate on this project. AM-B is grateful for the Margarita Salas postdoctoral fellowship at the University of Zaragoza, funded by the NextGeneration EU, which allowed her to collaborate on this project, and with the Institute for Health Research Aragon. The study has been externally funded, after undergoing peer review, by the Instituto de Salud Carlos III, dependent on the Spanish Ministry of Science and Innovation, through a grant for health-related research projects (PI21/01328 and PI21/01338) and the European Regional Development Fund (ERDF) “Another way to make Europe.” The funder plays no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YL-d-H and JG-C contributed to conceptualization. SF-M, AP-A, and AM-B contributed to data curation. AP-A, JG-P, and XC contributed to the formal analysis. YL-d-H and JG-P contributed to funding acquisition. YL-d-H, SF-M, AP-A, AM-B, AB-S, JG-P, and JG-C contributed to the investigation. YL-d-H, AP-A, and JG-P contributed to the methodology. YL-d-H, AP-A, and JG-P contributed to project administration. SF-M, AP-A, AM-B, AB-S, and DC contributed to software development. YL-d-H, JG-P, SF-M, LC-G, CA-L, and VC contributed to resources. JG-C was involved in supervision. SF-M, LC-G, CA-L, and VC were involved in validation. AP-A was involved in visualization and writing the original draft. YL-d-H, JG-P, and XC contributed to reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Outline of the contents of the MINDxYOU program.
[DOCX File , 18 KB - jmir_v27i1e59653_app1.docx]

Multimedia Appendix 2
Baseline differences between participants who completed all the assessments and those who did not.
[DOCX File , 25 KB - jmir_v27i1e59653_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Baseline differences between participants who started the web-based program and those who did not.
[DOCX File , 24 KB - jmir_v27i1e59653_app3.docx]

Multimedia Appendix 4
Baseline differences between participants who completed the web-based program (ie, at least 3 of the 4 modules) and those who did not.
[DOCX File , 25 KB - jmir_v27i1e59653_app4.docx]

Multimedia Appendix 5
Baseline characteristics of the study sample.
[DOCX File , 26 KB - jmir_v27i1e59653_app5.docx]

Multimedia Appendix 6
Intracluster effects in the main outcome (Perceived Stress Scale). Intention-to-treat approach.
[DOCX File , 26 KB - jmir_v27i1e59653_app6.docx]

Multimedia Appendix 7
Intracluster effects in the main outcome (Perceived Stress Scale). Per-protocol approach.
[DOCX File , 27 KB - jmir_v27i1e59653_app7.docx]

Multimedia Appendix 8
CONSORT (Consolidated Standards of Reporting Trials) eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2659 KB - jmir_v27i1e59653_app8.pdf]

References

1. Cullen W, Gulati G, Kelly BD. Mental health in the COVID-19 pandemic. QJM 2020 May 01;113(5):311-312 [FREE Full text] [doi: 10.1093/qjmed/hcaa110] [Medline: 32227218]
2. Mental health considerations during COVID-19 outbreak. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/mental-health-considerations.pdf [accessed 2024-04-29]
3. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during COVID-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: 10.1136/bmj.m1211] [Medline: 32217624]
4. Karasu F, Öztürk Çopur E, Ayar D. The impact of COVID-19 on healthcare workers' anxiety levels. Z Gesundh Wiss 2022;30(6):1399-1409 [FREE Full text] [doi: 10.1007/s10389-020-01466-x] [Medline: 33425658]
5. Pfefferbaum B, North CS. Mental health and the COVID-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
6. Theocharis A, Antonopoulos V, Christodoulou NG. Somatic symptoms associated with mental distress during the COVID-19 pandemic: a systematic review. Australas Psychiatry 2023 Apr;31(2):147-156 [FREE Full text] [doi: 10.1177/10398562231156380] [Medline: 36825513]
7. Kian N, Samieefar N, Akhlaghdoust M. Mental health of healthcare providers needs support: preparing for the post-COVID-19 era. East Mediterr Health J 2023 Apr 27;29(4):236-237 [FREE Full text] [doi: 10.26719/emhj.23.028] [Medline: 37246433]
8. Topuz F, Bülbüloğlu S, Filizli Z, Zayin D. Nursing image from the perspective of healthcare professionals during the post-COVID-19 pandemic period: a cross-sectional study. J Radiol Nurs 2023 Mar 06:334-338 [FREE Full text] [doi: 10.1016/j.jradnu.2023.02.012] [Medline: 37359322]
9. Lavoie-Tremblay M, Gélinas C, Aubé T, Tchouaket E, Tremblay D, Gagnon MP, et al. Influence of caring for COVID-19 patients on nurse's turnover, work satisfaction and quality of care. J Nurs Manag 2022 Jan;30(1):33-43 [FREE Full text] [doi: 10.1111/jonm.13462] [Medline: 34448520]
10. Lee J. Nursing home nurses' turnover intention: a systematic review. Nurs Open 2022 Jan;9(1):22-29 [FREE Full text] [doi: 10.1002/nop2.1051] [Medline: 34811952]
11. Poon YS, Lin YP, Griffiths P, Yong KK, Seah B, Liaw SY. A global overview of healthcare workers' turnover intention amid COVID-19 pandemic: a systematic review with future directions. Hum Resour Health 2022 Sep 24;20(1):70 [FREE Full text] [doi: 10.1186/s12960-022-00764-7] [Medline: 36153534]
12. Kang J, Kim H, Cho O. Quiet quitting among healthcare professionals in hospital environments: a concept analysis and scoping review protocol. BMJ Open 2023 Nov 19;13(11):e077811 [FREE Full text] [doi: 10.1136/bmjopen-2023-077811] [Medline: 37984954]
13. Gardiner P, Pérez-Aranda A, Bell N, Clark DR, Schuman-Olivier Z, Lin EH. Self-compassion for healthcare communities: exploring the effects of a synchronous online continuing medical education program on physician burnout. J Contin Educ Health Prof (Forthcoming) 2024 Sep 24. [doi: 10.1097/CEH.0000000000000574] [Medline: 39311747]

XSL•FO
RenderX

14. Otto AK, Gutsch C, Bischoff LL, Wollesen B. Interventions to promote physical and mental health of nurses in elderly care: a systematic review. Prev Med 2021 Jul;148:106591. [doi: 10.1016/j.ypmed.2021.106591] [Medline: 33930430]
15. Alkhawaldeh JM, Soh KL, Mukhtar FB, Peng OC, Anshasi HA. Stress management interventions for intensive and critical care nurses: a systematic review. Nurs Crit Care 2020 Mar;25(2):84-92. [doi: 10.1111/nicc.12489] [Medline: 31840391]
16. Clough BA, March S, Chan RJ, Casey LM, Phillips R, Ireland MJ. Psychosocial interventions for managing occupational stress and burnout among medical doctors: a systematic review. Syst Rev 2017 Jul 17;6(1):144 [FREE Full text] [doi: 10.1186/s13643-017-0526-3] [Medline: 28716112]
17. Kahl KG, Winter L, Schweiger U. The third wave of cognitive behavioural therapies: what is new and what is effective? Curr Opin Psychiatry 2012 Nov;25(6):522-528. [doi: 10.1097/YCO.0b013e328358e531] [Medline: 22992547]
18. Kriakous SA, Elliott KA, Lamers C, Owen R. The effectiveness of mindfulness-based stress reduction on the psychological functioning of healthcare professionals: a systematic review. Mindfulness (N Y) 2021 Sep 24;12(1):1-28 [FREE Full text] [doi: 10.1007/s12671-020-01500-9] [Medline: 32989406]
19. Rudaz M, Twohig MP, Ong CW, Levin ME. Mindfulness and acceptance-based trainings for fostering self-care and reducing stress in mental health professionals: a systematic review. J Contextual Behav Sci 2017 Oct;6(4):380-390. [doi: 10.1016/j.jcbs.2017.10.001]
20. Karo M, Simorangkir L, Daryanti Saragih I, Suarilah I, Tzeng HM. Effects of mindfulness-based interventions on reducing psychological distress among nurses: a systematic review and meta-analysis of randomized controlled trials. J Nurs Scholarsh 2024 Mar;56(2):319-330. [doi: 10.1111/jnu.12941] [Medline: 37955233]
21. Joyce S, Shand F, Tighe J, Laurent SJ, Bryant RA, Harvey SB. Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. BMJ Open 2018 Jun 14;8(6):e017858 [FREE Full text] [doi: 10.1136/bmjopen-2017-017858] [Medline: 29903782]
22. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: a developer's perspective. J Syst Softw 2011 Nov;84(11):2022-2033. [doi: 10.1016/j.jss.2011.06.049]
23. López-Del-Hoyo Y, Fernández-Martínez S, Pérez-Aranda A, Barceló-Soler A, Bani M, Russo S, et al. Effects of eHealth interventions on stress reduction and mental health promotion in healthcare professionals: a systematic review. J Clin Nurs 2023 Sep;32(17-18):5514-5533 [FREE Full text] [doi: 10.1111/jocn.16634] [Medline: 36703266]
24. Eriksson T, Germundsjö L, Åström E, Rönnlund M. Mindful self-compassion training reduces stress and burnout symptoms among practicing psychologists: a randomized controlled trial of a brief web-based intervention. Front Psychol 2018 Nov 27;9:2340 [FREE Full text] [doi: 10.3389/fpsyg.2018.02340] [Medline: 30538656]
25. Gruber J, Prinstein MJ, Clark LA, Rottenberg J, Abramowitz JS, Albano AM, et al. Mental health and clinical psychological science in the time of COVID-19: challenges, opportunities, and a call to action. Am Psychol 2021 Apr;76(3):409-426 [FREE Full text] [doi: 10.1037/amp0000707] [Medline: 32772538]
26. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. Front Psychiatry 2019;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]
27. Henshall C, Davey Z, Srikesavan C, Hart L, Butcher D, Cipriani A. Implementation of a web-based resilience enhancement training for nurses: pilot randomized controlled trial. J Med Internet Res 2023 Feb 14;25:e43771 [FREE Full text] [doi: 10.2196/43771] [Medline: 36787181]
28. Lippke S, Gao L, Keller FM, Becker P, Dahmen A. Adherence with online therapy vs face-to-face therapy and with online therapy vs care as usual: secondary analysis of two randomized controlled trials. J Med Internet Res 2021 Nov 03;23(11):e31274 [FREE Full text] [doi: 10.2196/31274] [Medline: 34730541]
29. Pospos S, Young IT, Downs N, Iglewicz A, Depp C, Chen JY, et al. Web-based tools and mobile applications to mitigate burnout, depression, and suicidality among healthcare students and professionals: a systematic review. Acad Psychiatry 2018 Feb;42(1):109-120 [FREE Full text] [doi: 10.1007/s40596-017-0868-0] [Medline: 29256033]
30. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ 2015 Feb 06;350:h391 [FREE Full text] [doi: 10.1136/bmj.h391] [Medline: 25662947]
31. Green BB, Coronado GD, Schwartz M, Coury J, Baldwin LM. Using a continuum of hybrid effectiveness-implementation studies to put research-tested colorectal screening interventions into practice. Implement Sci 2019 May 29;14(1):53 [FREE Full text] [doi: 10.1186/s13012-019-0903-5] [Medline: 31142348]
32. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
33. López-Del-Hoyo Y, Fernández-Martínez S, Pérez-Aranda A, Barceló-Soler A, Guzman-Parra J, Varela-Moreno E, et al. Effectiveness and implementation of an online intervention (MINDxYOU) for reducing stress and promote mental health among healthcare workers in Spain: a study protocol for a stepped-wedge cluster randomized trial. BMC Nurs 2022 Nov 10;21(1):308 [FREE Full text] [doi: 10.1186/s12912-022-01089-5] [Medline: 36357881]
34. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. J Clin Epidemiol 2016 Jan;69:137-146 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.015] [Medline: 26344808]
35. The MINDxYOU project. MINDxYOU. URL: https://mindxyou.es/ [accessed 2024-04-29]

XSL•FO
RenderX

36. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34. [Medline: 9881538]
37. Kabat-Zinn J. Full Catastrophe Living. New York, NY: Piatkus; 1991.
38. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in Spain: a randomized controlled trial. J Med Internet Res 2016 Aug 26;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
39. García-Campayo J, Navarro-Gil M, Demarzo M. Attachment-based compassion therapy. Mindfulness Compassion 2016 Jul;1(2):68-74. [doi: 10.1016/j.mincom.2016.10.004]
40. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
41. Remor E. Psychometric properties of a European Spanish version of the Perceived Stress Scale (PSS). Span J Psychol 2006 May;9(1):86-93 [FREE Full text] [doi: 10.1017/s1138741600006004] [Medline: 16673626]
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
43. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63(4):679-686. [doi: 10.1097/00006842-200107000-00021] [Medline: 11485122]
44. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
45. García-Campayo J, Zamorano E, Ruiz MA, Pardo A, Pérez-Páramo M, López-Gómez V, et al. Cultural adaptation into Spanish of the generalized anxiety disorder-7 (GAD-7) scale as a screening tool. Health Qual Life Outcomes 2010 Jan 20;8:8 [FREE Full text] [doi: 10.1186/1477-7525-8-8] [Medline: 20089179]
46. Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. Psychol Med 2009 Jul 09;13(3):595-605. [doi: 10.1017/S0033291700048017]
47. Derogatis LR. Brief symptoms inventory (BSI-18): administration, scoring and procedures. National Computer Systems. Minneapolis: NCS Pearson; 2001. URL: https://www.google.co.in/books/edition/BSI_18_Brief_Symptom_Inventory_18/MPNYHQAACAAJ?hl=en&kptab=overview [accessed 2024-04-29]
48. Miaja Ávila M, De la Rubia JM. Validación del Inventario Breve de Síntomas (BSI-18) en mujeres Mexicanas diagnosticadas con cáncer de mama. Psicooncología 1970 Jan 01;14(2-3):307-324. [doi: 10.5209/psic.57088]
49. Connor KM, Davidson JR, Lee LC. Spirituality, resilience, and anger in survivors of violent trauma: a community survey. J Trauma Stress 2003 Oct;16(5):487-494. [doi: 10.1023/A:1025762512279] [Medline: 14584633]
50. Gu J, Strauss C, Crane C, Barnhofer T, Karl A, Cavanagh K, et al. Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. Psychol Assess 2016 Jul;28(7):791-802 [FREE Full text] [doi: 10.1037/pas0000263] [Medline: 27078186]
51. Gu J, Baer R, Cavanagh K, Kuyken W, Strauss C. Development and psychometric properties of the Sussex-Oxford Compassion Scales (SOCS). Assessment 2020 Jan;27(1):3-20 [FREE Full text] [doi: 10.1177/1073191119860911] [Medline: 31353931]
52. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011 Dec;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]
53. Soler Sánchez MI, Meseguer de Pedro M, García Izquierdo M. Propiedades psicométricas de la versión española de la escala de resiliencia de 10 ítems de Connor-Davidson (CD-RISC 10) en una muestra multiocupacional. Rev Latinoam Psicol 2016 Sep;48(3):159-166. [doi: 10.1016/j.rlp.2015.09.002]
54. Cebolla A, García-Palacios A, Soler J, Guillen V, Baños R, Botella C. Psychometric properties of the Spanish validation of the Five Facets of Mindfulness Questionnaire (FFMQ). Eur J Psychiat 2012 Jun;26(2):118-126. [doi: 10.4321/S0213-61632012000200005]
55. García-Campayo J, Barceló-Soler A, Martínez-Rubio D, Navarrete J, Pérez-Aranda A, Feliu-Soler A, et al. Exploring the relationship between self-compassion and compassion for others: the role of psychological distress and wellbeing. Assessment 2024 Jul 15;31(5):1038-1051 [FREE Full text] [doi: 10.1177/10731911231203966] [Medline: 37840255]
56. Ruiz FJ, Langer Herrera Á, Luciano C, Cangas AJ, Beltrán I. Measuring experiential avoidance and psychological inflexibility: the Spanish version of the Acceptance and Action Questionnaire - II. Psicothema 2013 Feb 01;25(1):123-129. [doi: 10.7334/psicothema2011.239] [Medline: 23336554]
57. Pustejovsky JE, Tipton E. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. J Bus Econ Stat 2017 Jun 12;36(4):672-683. [doi: 10.1080/07350015.2016.1247004]
58. Li G, Taljaard M, Van den Heuvel ER, Levine MA, Cook DJ, Wells GA, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how. Int J Epidemiol 2017 Apr 01;46(2):746-755. [doi: 10.1093/ije/dyw320] [Medline: 28025257]

XSL•FO
RenderX

59. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol 2014 Aug;67(8):850-857. [doi: 10.1016/j.jclinepi.2014.03.012] [Medline: 24831050]
60. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289-300 [FREE Full text] [doi: 10.1111/j.2517-6161.1995.tb02031.x]
61. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ 2018 Nov 09;363:k1614 [FREE Full text] [doi: 10.1136/bmj.k1614] [Medline: 30413417]
62. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
63. Barrett K, Stewart I. A preliminary comparison of the efficacy of online Acceptance and Commitment Therapy (ACT) and Cognitive Behavioural Therapy (CBT) stress management interventions for social and healthcare workers. Health Soc Care Community 2021 Jan;29(1):113-126. [doi: 10.1111/hsc.13074] [Medline: 32613644]
64. Finlay-Jones A, Kane R, Rees C. Self-compassion online: a pilot study of an internet-based self-compassion cultivation program for psychology trainees. J Clin Psychol 2017 Jul;73(7):797-816. [doi: 10.1002/jclp.22375] [Medline: 27787877]
65. Wright EM. Evaluation of a web-based holistic stress reduction pilot program among nurse-midwives. J Holist Nurs 2018 Jun;36(2):159-169. [doi: 10.1177/0898010117704325] [Medline: 29172933]
66. Lee D, Lee WJ, Choi SH, Jang JH, Kang DH. Long-term beneficial effects of an online mind-body training program on stress and psychological outcomes in female healthcare providers: a non-randomized controlled study. Medicine (Baltimore) 2020 Aug 07;99(32):e21027 [FREE Full text] [doi: 10.1097/MD.0000000000021027] [Medline: 32769863]
67. Chan M, Jiang Y, Lee CY, Ramachandran HJ, Teo JY, Seah CW, et al. Effectiveness of eHealth-based cognitive behavioural therapy on depression: a systematic review and meta-analysis. J Clin Nurs 2022 Nov;31(21-22):3021-3031. [doi: 10.1111/jocn.16212] [Medline: 35060252]
68. Watanabe K, Tran TT, Sripo N, Sakuraya A, Imamura K, Boonyamalik P, et al. Effectiveness of a smartphone-based stress management program for depression in hospital nurses during COVID-19 in Vietnam and Thailand: 2-arm parallel-group randomized controlled trial. J Med Internet Res 2024 Aug 30;26:e50071 [FREE Full text] [doi: 10.2196/50071] [Medline: 39213033]
69. Yang JM, Ye H, Long Y, Zhu Q, Huang H, Zhong YB, et al. Effects of web-based mindfulness-based interventions on anxiety, depression, and stress among frontline health care workers during the COVID-19 pandemic: systematic review and meta-analysis. J Med Internet Res 2023 Aug 29;25:e44000 [FREE Full text] [doi: 10.2196/44000] [Medline: 37527546]
70. Lu Y, Li Y, Huang Y, Zhang X, Wang J, Wu L, et al. Effects and mechanisms of a web- and mobile-based acceptance and commitment therapy intervention for anxiety and depression symptoms in nurses: fully decentralized randomized controlled trial. J Med Internet Res 2023 Nov 27;25:e51549 [FREE Full text] [doi: 10.2196/51549] [Medline: 38010787]
71. Hennemann S, Böhme K, Kleinstäuber M, Ruckes C, Baumeister H, Daniel Ebert D, et al. Is therapist support needed? comparing therapist- and self-guided internet-based CBT for somatic symptom distress (iSOMA) in emerging adults. Behav Ther 2022 Nov;53(6):1205-1218. [doi: 10.1016/j.beth.2022.06.006] [Medline: 36229117]
72. Maroti D, Lumley MA, Schubiner H, Lilliengren P, Bileviciute-Ljungar I, Ljótsson B, et al. Internet-based emotional awareness and expression therapy for somatic symptom disorder: a randomized controlled trial. J Psychosom Res 2022 Dec;163:111068 [FREE Full text] [doi: 10.1016/j.jpsychores.2022.111068] [Medline: 36327532]
73. Wilson AC, Mackintosh K, Power K, Chan SW. Effectiveness of self-compassion related therapies: a systematic review and meta-analysis. Mindfulness 2018 Oct 24;10(6):979-995. [doi: 10.1007/S12671-018-1037-6]
74. O'Connor M, Munnelly A, Whelan R, McHugh L. The efficacy and acceptability of third-wave behavioral and cognitive eHealth treatments: a systematic review and meta-analysis of randomized controlled trials. Behav Ther 2018 May;49(3):459-475. [doi: 10.1016/j.beth.2017.07.007] [Medline: 29704973]
75. Kunzler AM, Helmreich I, Chmitorz A, König J, Binder H, Wessa M, et al. Psychological interventions to foster resilience in healthcare professionals. Cochrane Database Syst Rev 2020 Jul 05;7(7):CD012527 [FREE Full text] [doi: 10.1002/14651858.CD012527.pub2] [Medline: 32627860]
76. Pérez-Aranda A, García-Campayo J, Gude F, Luciano JV, Feliu-Soler A, González-Quintela A, et al. Impact of mindfulness and self-compassion on anxiety and depression: the mediating role of resilience. Int J Clin Health Psychol 2021;21(2):100229 [FREE Full text] [doi: 10.1016/j.ijchp.2021.100229] [Medline: 33767736]
77. Knudsen RK, Gregersen T, Ammentorp J, Tousig CG, Timmermann C. Healthcare professionals' experiences of using mindfulness training in a cardiology department - a qualitative study. Scand J Caring Sci 2021 Sep;35(3):892-900. [doi: 10.1111/scs.12906] [Medline: 32852094]
78. Taylor H, Cavanagh K, Field AP, Strauss C. Health care workers' need for headspace: findings from a multisite definitive randomized controlled trial of an unguided digital mindfulness-based self-help app to reduce healthcare worker stress. JMIR Mhealth Uhealth 2022 Aug 25;10(8):e31744 [FREE Full text] [doi: 10.2196/31744] [Medline: 36006668]
79. Mistretta EG, Davis MC, Temkit M, Lorenz C, Darby B, Stonnington CM. Resilience training for work-related stress among health care workers: results of a randomized clinical trial comparing in-person and smartphone-delivered interventions. J Occup Environ Med 2018 Jun;60(6):559-568. [doi: 10.1097/JOM.0000000000001285] [Medline: 29370014]

XSL•FO
RenderX

80. Demirel S, Roke Y, Hoogendoorn AW, Hoefakker J, Hoeberichts K, van Harten PN. Assessing the effectiveness of STAPP@Work, a self-management mobile app, in reducing work stress and preventing burnout: single-case experimental design study. J Med Internet Res 2024 Feb 29;26:e48883 [FREE Full text] [doi: 10.2196/48883] [Medline: 38275128]
81. Redko C, Rapp RC, Carlson RG. Waiting time as a barrier to treatment entry: perceptions of substance users. J Drug Issues 2006 Sep;36(4):831-852 [FREE Full text] [doi: 10.1177/002204260603600404] [Medline: 18509514]
82. Fernández-Martínez S, Armas-Landaeta C, Pérez-Aranda A, Guzmán-Parra J, Monreal-Bartolomé A, Carbonell-Aranda V, et al. Post-COVID job stressors and their predictive role on mental health: a cross-sectional analysis between physicians and nurses. SAGE Open Nurs 2024;10:23779608241278861. [doi: 10.1177/23779608241278861] [Medline: 39290451]
83. Barceló-Soler A, Morillo-Sarto H, Fernández-Martínez S, Monreal-Bartolomé A, Chambel MJ, Gardiner P, et al. A systematic review of the adherence to home-practice meditation exercises in patients with chronic pain. Int J Environ Res Public Health 2023 Mar 02;20(5):4438 [FREE Full text] [doi: 10.3390/ijerph20054438] [Medline: 36901448]
84. Ketelaar SM, Nieuwenhuijsen K, Bolier L, Smeets O, Sluiter JK. Improving work functioning and mental health of health care employees using an e-mental health approach to workers' health surveillance: pretest-posttest study. Saf Health Work 2014 Dec;5(4):216-221 [FREE Full text] [doi: 10.1016/j.shaw.2014.08.002] [Medline: 25516815]
85. Dutton S, Kozachik SL. Evaluating the outcomes of a web-based stress management program for nurses and nursing assistants. Worldviews Evid Based Nurs 2020 Feb;17(1):32-38. [doi: 10.1111/wvn.12417] [Medline: 31912984]
86. Hersch RK, Cook RF, Deitz DK, Kaplan S, Hughes D, Friesen MA, et al. Reducing nurses' stress: a randomized controlled trial of a web-based stress management program for nurses. Appl Nurs Res 2016 Nov;32:18-25 [FREE Full text] [doi: 10.1016/j.apnr.2016.04.003] [Medline: 27969025]
87. Kemper KJ, Lynn J, Mahan JD. What is the impact of online training in mind-body skills? J Evid Based Complementary Altern Med 2015 Oct;20(4):275-282. [doi: 10.1177/2156587215580882] [Medline: 25962637]
88. Kemper KJ, Khirallah M. Acute effects of online mind-body skills training on resilience, mindfulness, and empathy. J Evid Based Complementary Altern Med 2015 Oct;20(4):247-253. [doi: 10.1177/2156587215575816] [Medline: 25783980]
89. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
90. Smoktunowicz E, Lesnierowska M, Carlbring P, Andersson G, Cieslak R. Resource-based internet intervention (Med-Stress) to improve well-being among medical professionals: randomized controlled trial. J Med Internet Res 2021 Jan 11;23(1):e21445 [FREE Full text] [doi: 10.2196/21445] [Medline: 33427674]
91. Yamagishi M, Kobayashi T, Nakamura Y. Effects of web-based career identity training for stress management among Japanese nurses: a randomized control trial. J Occup Health 2008;50(2):191-193 [FREE Full text] [doi: 10.1539/joh.l7086] [Medline: 18403870]
92. Jakel P, Kenney J, Ludan N, Miller PS, McNair N, Matesic E. Effects of the use of the provider resilience mobile application in reducing compassion fatigue in oncology nursing. Clin J Oncol Nurs 2016 Dec 01;20(6):611-616 [FREE Full text] [doi: 10.1188/16.CJON.611-616] [Medline: 27857259]
93. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, Zamanillo-Campos R, Yáñez-Juan AM, Bennasar-Veny M, et al. A mobile phone-based intervention to reduce mental health problems in health care workers during the COVID-19 pandemic (PsyCovidApp): randomized controlled trial. JMIR Mhealth Uhealth 2021 May 18;9(5):e27039 [FREE Full text] [doi: 10.2196/27039] [Medline: 33909587]

Abbreviations

AAQ-II: Acceptance and Action Questionnaire-II
BSI-18: Brief Symptoms Inventory-18
CBT: cognitive behavioral therapy
CD-RISC: Connor-Davidson Resilience Scale
CONSORT: Consolidated Standards of Reporting Trials
FFMQ-15: Five Facet Mindfulness Questionnaire-15
GAD-7: Generalized Anxiety Disorder-7
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale
SOCS: Sussex-Oxford Compassion Scales

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 18.04.24; peer-reviewed by M Bani, DA Hill; comments to author 21.09.24; revised version received 01.10.24; accepted 21.11.24; published 03.02.25.

Please cite as:
López-del-Hoyo Y, Fernández-Martínez S, Perez-Aranda A, Monreal-Bartolomé A, Barceló-Soler A, Camarero-Grados L, Armas-Landaeta C, Guzmán-Parra J, Carbonell V, Campos D, Chen X, García-Campayo J
Effectiveness of a Web-Based Self-Guided Intervention (MINDxYOU) for Reducing Stress and Promoting Mental Health Among Health Professionals: Results From a Stepped-Wedge Cluster Randomized Trial
J Med Internet Res 2025;27:e59653
URL: https://www.jmir.org/2025/1/e59653
doi:10.2196/59653
PMID:39899345

©Yolanda López-del-Hoyo, Selene Fernández-Martínez, Adrian Perez-Aranda, Alicia Monreal-Bartolomé, Alberto Barceló-Soler, Loreto Camarero-Grados, Carilene Armas-Landaeta, José Guzmán-Parra, Vera Carbonell, Daniel Campos, Xinyuan Chen, Javier García-Campayo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Urban-Rural Differences in the Association Between Internet Use Trajectories and Depressive Symptoms in Chinese Adolescents: Longitudinal Observational Study

Yujie Liu[1,2], PhD; Xin Ge[1], MD; Ying Wang[1], PhD; Xue Yang[3], PhD; Shangbin Liu[1], MD; Chen Xu[1], MD; Mi Xiang[2], PhD; Fan Hu[1], PhD; Yong Cai[1], PhD

[1]Public Health Research Center, Tongren Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]School of Public Health, Shanghai Jiao Tong University, Shanghai, China

[3]JC School of Public Health and Primary Care, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, China

Corresponding Author:
Yong Cai, PhD
Public Health Research Center
Tongren Hospital
Shanghai Jiao Tong University School of Medicine
1111 Xianxia Road
Shanghai, 200335
China
Phone: 86 021 63846590 ext 776198
Email: caiyong202028@hotmail.com

Abstract

Background: Internet use exhibits diverse trajectories during adolescence, which may contribute to depressive symptoms. Currently, it remains unclear whether the association between internet use trajectories and depressive symptoms varies between urban and rural areas.

Objective: This study aimed to investigate the association between internet use trajectories and adolescent depressive symptoms and to explore variation in this association between urban and rural areas.

Methods: This longitudinal study used 3-wave data from the 2014-2018 China Family Panel Study. Weekly hours of internet use and depressive symptoms were measured using self-reported questionnaires. Latent class growth modeling was performed to identify the trajectories of internet use. Multivariable logistic regressions were used to examine the association between internet use trajectories and depressive symptoms, stratified by rural and urban residence.

Results: Participants were 2237 adolescents aged 10 to 15 years at baseline (mean age 12.46, SD 1.73 years). Two latent trajectory classes of internet use were identified: the low-growth group (n=2008, 89.8%) and the high-growth group (n=229, 10.2%). The high-growth group was associated with higher odds of depressive symptoms (OR 1.486, 95% CI 1.065-2.076) compared to the low-growth group. In the stratified analysis, the association between internet use trajectories and depressive symptoms was significant solely among rural adolescents (OR 1.856, 95% CI 1.164-2.959).

Conclusions: This study elucidates urban-rural differences in the associations between trajectories of internet use and adolescent depressive symptoms. Our findings underscore the importance of prioritizing interventions for rural adolescents' internet use behaviors to mitigate negative effects on their mental health.

(J Med Internet Res 2025;27:e63799) doi:10.2196/63799

KEYWORDS

internet use; trajectory; depressive symptoms; adolescent; urban; rural

Introduction

Adolescence is a vulnerable period for developing depressive symptoms as a result of substantial biological and social transitions [1]. In China, the prevalence of depressive symptoms among adolescents has notably increased in recent decades [2,3]. A recent meta-analysis shows that approximately 19.9% of Chinese adolescents experience depressive symptoms [4]. The

XSL•FO
RenderX

onset and persistence of depressive symptoms during this crucial developmental stage can result in a range of adverse outcomes in adulthood, such as functional impairment, suicidal self-harm, and mental health difficulties [5]. Therefore, early identification and intervention for depressive symptoms are imperative.

In recent years, the internet has become an essential tool for adolescents, widely used for information acquisition, social interaction, and entertainment [6]. According to the 2021 National Report on Internet Use by Minors, China has approximately 191 million internet users younger than 18 years, with a penetration rate of 96.8% [7]. Adolescents engage in various forms of internet use, including accessing online resources for educational purposes, such as completing school assignments [8]. Moreover, recreational and social activities, including video streaming, online gaming, and social networking, are also prevalent among this age group [9]. However, the widespread use of the internet has raised concerns about its effects on adolescent health [10-12]. Increasing evidence indicates a positive correlation between the duration of internet use and depressive symptoms in adolescents, with entertainment-related activities showing more pronounced negative effects [13-15]. In response, the Chinese government introduced regulations in 2019 to limit online gaming to 1.5 hours on weekdays and 3 hours on weekends to curb excessive internet use [16]. More recently, the Cyberspace Administration of China proposed further restrictions, recommending that adolescents aged 8 to 15 years be limited to 1 hour of daily mobile internet use, while those aged 16 to 18 years should not exceed 2 hours [17].

Lifestyle behaviors undergo dynamic changes during adolescence, characterized by a progressive rise in internet use as adolescents mature [18]. A longitudinal study of participants with a baseline age of 9.9 years reported significant increases in screen-based activities over 4 years: social media use rose from 0.5 to 2.5 hours per day, and video gaming from 0.9 to 2 hours per day, indicating an upward trajectory of internet use during early adolescence [19]. These evolving lifestyle patterns are well-established predictors of mental health outcomes [20]. Additionally, internet use may follow diverse subgroup trajectories within the overall developmental trend. For example, Yu and Park [21] used growth mixture modeling to classify South Korean adolescents into 4 distinct internet use trajectories: high stable, high quadratic, moderate stable, and low stable. Another study identified 2 subgroups: one with a sharp increase in internet use (drastically increasing group) and another with a more gradual rise (gradually increasing group), capturing varying rates of increase during adolescence [22]. These trajectories are associated with different risks for health outcomes [21,23]. However, whether internet use trajectories can predict depressive symptoms among adolescents remains unclear.

Furthermore, the impacts of internet use among adolescents may vary between rural and urban areas. Traditionally, rural adolescents have limited access to internet devices and engage in fewer online activities than their urban counterparts [24]. With the increasing penetration rates of the internet in recent years, the urban-rural gap has narrowed, with adolescents in rural areas experiencing a faster increase in screen time [25]. Consequently, the trajectories of internet use duration among rural and urban adolescents may differ over time. On the other hand, rural adolescents may face greater challenges due to a lack of parental supervision, as many of their parents migrate from rural to urban areas for work [7,26]. This absence of oversight could exacerbate the negative effects of excessive internet use. Therefore, it is essential to understand the varying associations between internet use trajectories in rural and urban adolescents and their mental health outcomes, which can offer empirical references for preventing adolescent depressive symptoms.

To date, the majority of research in China regarding the impact of adolescent internet use on mental health has focused on single regions, overlooking potential urban-rural differences across areas. This study seeks to address this knowledge gap by using nationally representative data for adolescents in 2014, 2016, and 2018. Our primary objective is to investigate the relationship between internet use trajectories throughout adolescence and depressive symptoms and to explore whether this relationship varies between urban and rural areas.

Methods

Study Sample and Procedure

This longitudinal study used data from the China Family Panel Study (CFPS), which covers populations from 25 provinces, municipalities, and autonomous regions in China. The CFPS uses a multistage sampling approach with implicit stratification based on urban-rural distinctions and regional socioeconomic status. Using a probability proportional to size sampling strategy, 32 streets and townships from Shanghai and 144 districts and counties from the other 24 provinces were selected as primary sampling units. Subsequently, 640 communities were randomly chosen from these units, and 25 households were systematically sampled from each community. The baseline survey in 2010 collected data from 33,600 adults and 8990 children across 14,960 families, who were invited to participate in follow-up surveys every 2 years. These surveys gathered comprehensive information on participants' economic conditions, education, lifestyle, family relationships, and physical and mental health.

For this study, we used 3 waves of data from the 2014-2018 cycle of the CFPS to ensure consistent measurement of internet use duration. In 2014, the sample included 2567 adolescents aged 10-15 years, of whom 1491 completed surveys in both the 2016 and 2018 cycles. By 2018, the participants' age ranged from 14 to 19 years. To capture longitudinal developmental patterns, adolescents with self-reported data from at least 2 of the 3 waves were included, resulting in a final sample of 2237 participants. The flow of participants in this study is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Flow chart of participation in this study. CFPS: China Family Panel Study; LCGM: latent class growth modeling.

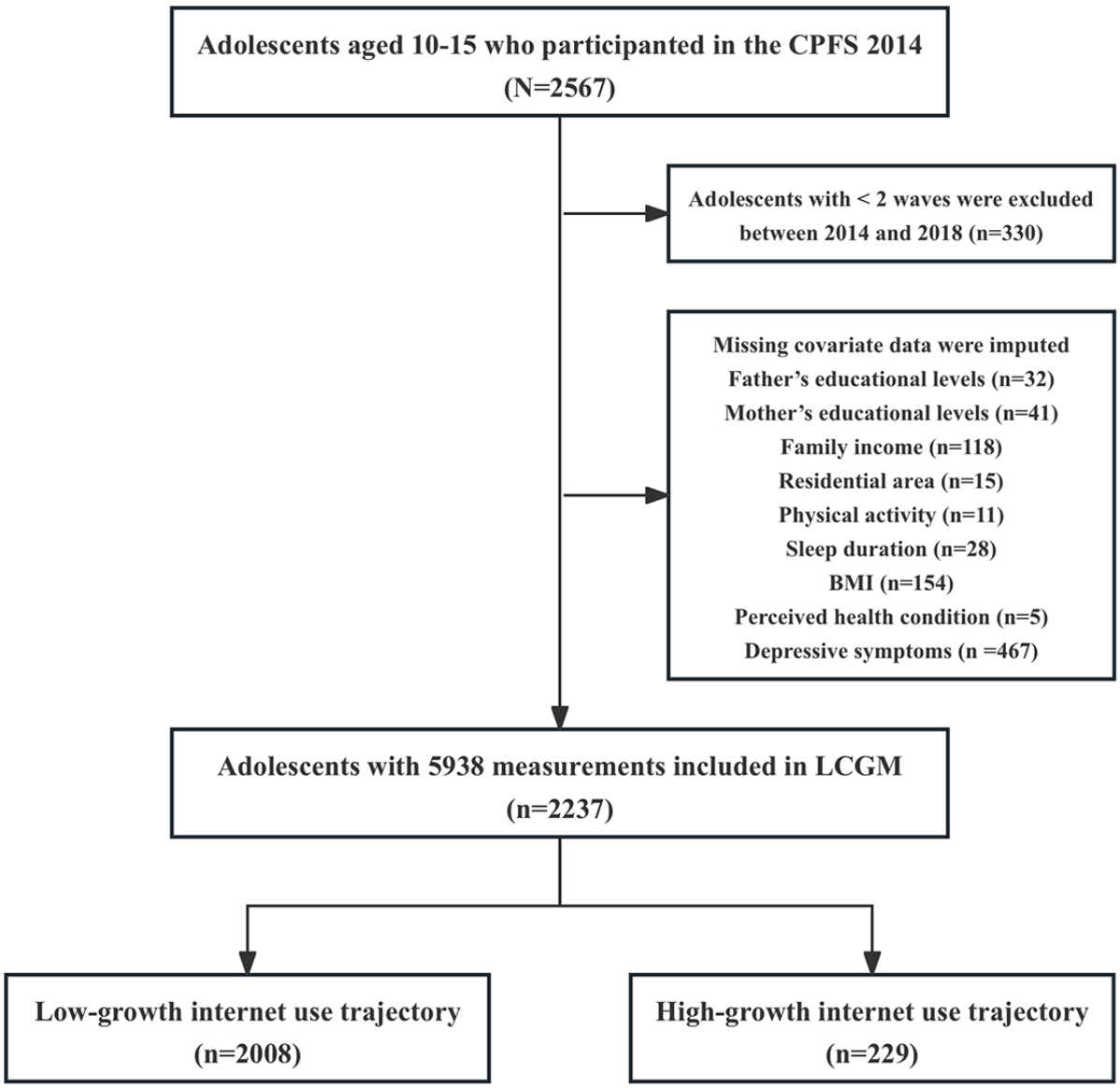

Measurements

Internet Use

Internet use was measured using a single item: "In general, what is the total amount of time you spend online during your leisure time per week?" This variable is designed to measure leisure time spent online, with values ranging from 0 to 168 hours.

Depressive Symptoms

The Center for Epidemiologic Studies Depression (CESD) scale was used to assess depressive symptoms. The CESD evaluates a range of symptoms, including sadness, loss of interest, changes in appetite and sleep, difficulties in thinking or concentration, feelings of guilt, fatigue, movement changes, and suicidal ideation. Each item is rated on a 4-point scale (0="not at all or less than 1 day in the past week" to 3="nearly every day for 2 weeks"). To reduce respondent burden, in 2016, one-fifth of participants were randomly selected to complete the 20-item CESD (CESD-20), while the remaining four-fifths completed the 8-item version (CESD-8). All participants completed the CESD-8 in 2018 [27]. The CFPS considered the 2 groups in 2016 equivalent in depressive symptom distributions and used the equipercentile equating method to transform CESD-8 scores into CESD-20 scores, ensuring consistency and comparability across survey waves [27]. This approach aligns scores from the 2 versions by matching them to the same percentile rank [28]. The total CESD-20 score ranges from 0 to 60, with scores ≥16 indicating the presence of depressive symptoms based on established cutoffs [29]. CESD-8 and CESD-20 have demonstrated good psychometric properties in adolescents [30,31].

Covariates

Demographic variables included age (10-15 years in 2014), sex (male or female), parental educational levels (illiterate or semiliterate; primary school; junior high school; senior high

XSL•FO
RenderX

school; college, university, or above), family income, residential area (rural or urban), and grandparental caregiving. Grandparental caregiving was determined based on responses to 2 questions: "Who takes care of the child during the day?" and "Who takes care of the child at night?" If either question was answered with "grandparents," the child was classified as receiving grandparental care. Lifestyle and health-related variables included physical activity, sleep duration, BMI, and perceived health condition. Physical activity was assessed using a single item: "Including participation in physical education classes, what was the total amount of time you spent participating in physical activity in the past week?" Sleep duration was calculated as the time between falling asleep and waking up. BMI was determined by dividing objectively measured weight by height squared (kg/m^2). Perceived health condition was assessed using a single item, "How do you perceive your health condition?" with responses ranging from 1 ("extremely healthy") to 5 ("unhealthy").

Baseline psychological distress was measured using the Kessler Psychological Distress Scale (K6). The K6 assesses the frequency of 6 symptoms: feeling nervous, hopeless, restless or fidgety, worthless, depressed, or that everything requires an effort [32]. Possible responses to the 6 items range from 0 ("none of the time") to 4 ("all of the time"), yielding a total score of 0-24. The K6 has demonstrated good psychometric properties in adolescents [33,34].

Statistical Analysis

The statistical analysis involved a 2-step procedure. In the first step, we conducted latent class growth modeling (LCGM) to determine trajectories of internet use across the 3 waves using Mplus (version 8.3; Muthén & Muthén). LCGM is a technique that identifies subgroups with distinct developmental patterns within a population over time [35]. With the 3-wave data, both linear and quadratic models were fitted to determine the optimal number of latent trajectory classes. The best-fitting model was determined based on the Akaike information criterion (AIC), Bayesian information criterion (BIC), sample size–adjusted BIC (aBIC), entropy, the Lo-Mendell-Rubin likelihood ratio test (LMR), and the bootstrapped likelihood ratio test (BLRT). Lower values of AIC, BIC, and aBIC indicate a better model fit. Higher entropy represents better classification accuracy. A significant *P* value ($P<.05$) for LMR and BLRT in a k-class model suggests that the model is better than the k–1 class model. Further, the sample size was taken into account, with at least 5% of the total population required in each class to define a meaningful latent entity [36].

In the second step, multivariable logistic regressions were conducted within the best-fitting model to examine the associations between internet use trajectories and distal outcomes. Each individual was assigned to their most likely class based on the highest posterior probability estimate, which was then used as the independent variable. The outcome variable was depressive symptoms in 2018. Three regression models were constructed. In model 1, adjustments were made for baseline K6 scores. Model 2 included additional demographic covariates, and model 3 further adjusted for lifestyle and health-related covariates. Stratified analysis was performed based on urban and rural classifications.

The Little missing completely at random (MCAR) test indicated that missing data for internet use followed an MCAR pattern ($\chi^2_9=6.616$; $P=.68$). Therefore, missing data on internet use were handled using the robust full information maximum-likelihood estimation approach [37]. Missing data on covariates and the distal outcome were imputed using the *missForest* function in R (R Project for Statistical Computing), a method capable of handling both continuous and categorical variables [38]. We imputed missing values for the father's educational level (n=32, 1.4%), mother's educational level (n=41, 1.8%), family income (n=118, 5.4%), grandparental caregiving (n=13, 0.6%), residential area (n=15, 0.6%), physical activity (n=11, 0.5%), sleep duration (n=28, 1.2%), BMI (n=154, 7.1%), perceived health condition in 2014 (n=5, 0.2%), and depressive symptoms in 2018 (n=467, 21.5%).

Ethics Approval

The Biomedical Ethics Review Committee of Peking University approved the study (IRB00001052-14,010). All eligible participants provided informed consent. The questionnaire was anonymous and did not collect any identifying information, such as name or address. Eligible participants were given CN ¥20-CN ¥60 (US $3.26-US $9.77) in cash as compensation.

Results

Descriptive Statistics

Table 1 presents the characteristics of the 2237 participants. The sample consisted of 1067 (47.7%) girls and 1170 (52.3%) boys, with a mean age of 12.46 (SD 1.73) years in 2014. Approximately half of the participants' parents had a high school education or higher (father's educational level: n=1187, 53.1%; mother's educational level: n=938, 41.9%). The median annual household income was CN ¥39,000 (IQR CN ¥20,100-CN ¥61,200; US $3272-US $9962.60). A total of 60.4% (n=1351) of the participants were from rural areas, and 17.3% (n=386) were cared for by their grandparents. The median weekly hours of internet use were 0.0 (IQR 0.0-3.0) in 2014, 3.0 (IQR 0.0-8.0) in 2016, and 7.0 (IQR 2.0-20.0) in 2018. In 2018, the prevalence of depressive symptoms was 19.8% (n=444). CESD-20 scores in 2018 were positively correlated with K6 scores in 2014 (Spearman $\rho=0.166$; $P<.001$).

XSL•FO
RenderX

Table 1. Population characteristics.

Characteristics	Values (N=2237)
Sex, n (%)	
Female	1067 (47.7)
Male	1170 (52.3)
Age (years)	
Mean (SD)	12.5 (1.7)
Range	10.0-15.0
Father's education level, n (%)	
Illiterate/semiliterate	410 (18.3)
Primary school	640 (28.6)
Senior high school	775 (34.6)
Junior high school	261 (11.7)
College/university	151 (6.8)
Mother's education level, n (%)	
Illiterate/semiliterate	686 (30.7)
Primary school	613 (27.4)
Senior high school	628 (28.1)
Junior high school	203 (9.1)
College/university	107 (4.8)
Annual family income (CN ¥; CN ¥1 = US $ 0.1628)	
Median (IQR)	39,000 (20,100-61,200)
Range	0-1,360,000
Residential area, n (%)	
Rural	1351 (60.4)
Urban	886 (39.6)
Grandparental caregiving, n (%)	
No	1851 (82.7)
Yes	386 (17.3)
Physical activity (hours/week)	
Median (IQR)	2.0 (0.0-4.0)
Range	0.0-56.0
Sleep duration (hours/day)	
Median (IQR)	9.0 (8.0-10.0)
Range	6.0-12.0
Internet use (2014) (hours/week)	
Median (IQR)	0.0 (0.0-3.0)
Range	0.0-84.0
Internet use (2016) (hours/week)	
Median (IQR)	3.0 (0.0-8.0)
Range	0.0-91.0
Internet use (2018) (hours/week)	
Median (IQR)	7.0 (2.0-20.0)
Range	0.0-144.0

XSL•FO
RenderX

Characteristics	Values (N=2237)
BMI (kg/m^2)	
Median (IQR)	17.8 (16.0-20.0)
Range	11.4-40.0
Perceived health condition, n (%)	
1	708 (31.6)
2	746 (33.3)
3	616 (27.5)
4	136 (6.1)
5	31 (1.4)
Kessler Psychological Distress Scale scores (2014)	
Median (IQR)	2.0 (0.0-4.0)
Range	0.0-21.0
Center for Epidemiologic Studies Depression scores (2018)	
Median (IQR)	30.0 (28.0-34.0)
Range	22.0-68.0
Depressive symptoms (2018), n (%)	
Nondepressed	1793 (80.2)
Depressed	444 (19.8)

LCGM Analysis

Table 2 presents the fit indices from the LCGM analysis for models with 1-4 trajectory classes. First, linear models were fitted. Reductions in AIC, BIC, and aBIC were observed across 1- to 4-class models. While both the LMR and BLRT statistics favored the 3-class model over the 2-class model, the latter exhibited higher entropy (0.909) with its smallest group >5%. Consequently, the 2-class model was selected as the best-fitting linear model. Subsequently, quadratic models were fitted, showing a similar pattern in fit indices to the linear models. The AIC, BIC, and aBIC values indicated an improvement in model fit upon adding the quadratic term. Therefore, the 2-class quadratic model was identified as the final optimal model of internet use trajectories.

Figure 2 shows the 2 latent trajectory classes of internet use. Class 1 (89.8%) is labelled the “low-growth group,” representing the majority of adolescents who, on average, had relatively short internet use duration in 2014 (intercept=2.723; P<.001) and experienced gradual growth over 4 years (slope=1.080; P=.01; quadratic term=0.998; P<.001). Class 2 (10.2%) is labeled the “high-growth group,” characterized by initially longer average internet use duration (intercept=8.421; P<.001) and rapid growth until 2018 (slope=20.183; P=.008; quadratic term=–1.293; P=.74).

There were statistically significant differences across the trajectory classes in the distributions of age, parental educational level, annual family income, and residential area (Multimedia Appendix 1, Table S1). Specifically, urban participants were more likely to belong to the low-growth group compared to their rural counterparts.

XSL•FO
RenderX

Table 2. Model fit indices for latent class growth modeling.

Type of model and number of classes	AIC[a]	BIC[b]	aBIC[c]	Entropy	LMR[d]	BLRT[e]	Proportions[f]
Linear model							
1	46,043.458	46,060.597	46,051.065	__[g]	—	—	1
2	44,504.897	44,539.175	44,520.112	0.909	<.001	<0.001	0.892/0.108
3	44,118.878	44,170.294	44,141.699	0.879	0.03	<0.001	0.840/0.129/0.031
4	43,792.129	43,860.684	43,822.558	0.893	0.06	<0.001	0.806/0.141/0.034/0.019
Quadratic model							
1	46,038.095	46,060.946	46,048.238	—	—	—	1
2	44,495.849	44,541.552	44,516.135	0.914	0.01	<0.001	0.897/0.102
3	43,469.206	43,537.761	43,499.635	0.893	0.007	<0.001	0.867/0.034/0.099
4	42,837.724	42,929.13	42,878.295	0.883	0.20	<0.001	0.786/0.154/0.0345/0.025

[a]AIC: Akaike information criterion.

[b]BIC: Bayesian information criterion.

[c]aBIC: sample size–adjusted Bayesian information criterion.

[d]LMR: Lo-Mendell-Rubin likelihood ratio test.

[e]BLRT: bootstrapped likelihood ratio test.

[f]The values in this column represent the proportions of individuals in each class identified by latent class growth modeling.

[g]Not applicable.

Figure 2. Latent trajectory classes of internet use from 2014 to 2018.

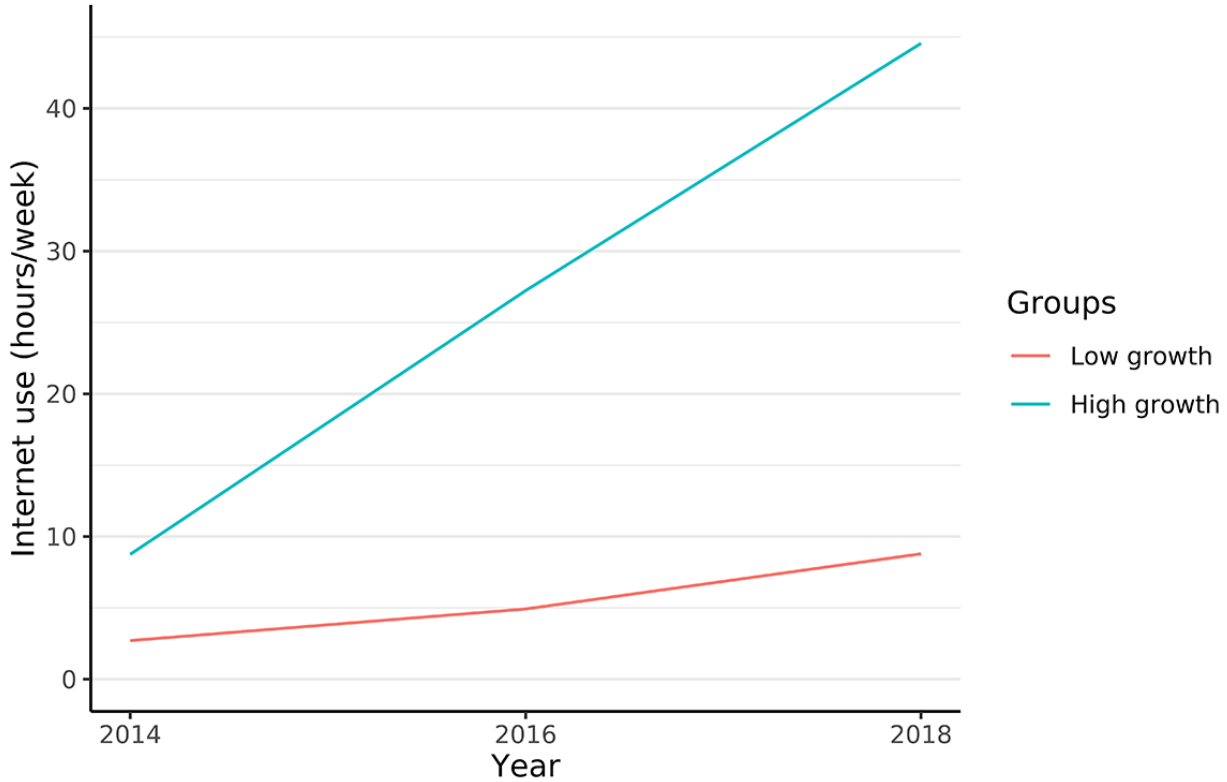

Association Between Trajectories and Depressive Symptoms

The percentage of adolescents with depressive symptoms in 2018 was 19.2% (385/2008) in the low-growth group and 25.8% (59/229) in the high-growth group. The difference in the percentages of depressive symptoms between the 2 groups was 6.6% (95% CI 1.1%-12%). Table 3 presents the results from multivariable logistic regressions examining the association between internet use trajectories and depressive symptoms. After adjusting for baseline psychological distress (model 1), the high-growth group was associated with higher odds of

XSL•FO
RenderX

depressive symptoms compared to the low-growth group (OR 1.460, 95% CI 1.062-2.008; *P*=.02). The association remained significant after further adjustment for demographic, lifestyle, and health-related variables (model 3), with the high-growth group showing higher odds of depressive symptoms (OR 1.520, 95% CI 1.090-2.120; *P*=.01).

Table 3. Multivariable logistic regressions of internet use trajectory classes on depressive symptoms.

Rurality and internet use trajectory	Depressive symptoms, n (%)		Model 1[a]		Model 2[b]		Model 3[c]	
	Nonde-pressed	Depressed	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Overall (N=2237)								
Low growth (n=2008)	1623 (80.8)	385 (19.2)	Ref	—[d]	—	—	—	—
High growth (n=229)	170 (74.2)	59 (25.8)	1.460 (1.062-2.008)	.02	1.501 (1.079-2.088)	.02	1.520 (1.090-2.120)	.01
Rural (n=1351)								
Low growth (n=1241)	1009 (81.3)	232 (18.7)	Ref	—	—	—	—	—
High growth (n=110)	79 (71.8)	31 (28.2)	1.728 (1.111-2.686)	.02	1.874 (1.181-2.972)	.008	1.853 (1.162-2.955)	.01
Urban (n=886)								
Low growth (n=767)	614 (80.1)	153 (19.9)	Ref	—	—	—	—	—
High growth (n=119)	91 (76.5)	28 (23.5)	1.192 (0.749-1.896)	.46	1.117 (0.686-1.819)	.66	1.124 (0.686-1.840)	.64

[a]Model 1: Controlling for baseline psychological distress.

[b]Model 2: Controlling for baseline psychological distress, sex, age, parental educational level, annual family income, and grandparental caregiving.

[c]Model 3: Controlling for baseline psychological distress, sex, age, parental educational level, annual family income, grandparental caregiving, physical activity, sleep duration, BMI, and perceived health condition.

[d]Not applicable.

When stratified by residential area, the prevalence rates of depressive symptoms were 18.7% (232/1241) in the low-growth group and 28.2% (31/110) in the high-growth group among rural participants. The difference in the percentages of depressive symptoms between the 2 groups was 9.5% (95% CI 1.8%-17.2%). Results from the multivariable logistic regression indicated that the high-growth group was associated with higher odds of depressive symptoms among rural participants (OR 1.853, 95% CI 1.162-2.955; *P*=.01). Among urban participants, the prevalence rates were 19.9% (153/767) for the low-growth group and 23.5% (28/119) for the high-growth group. The difference in the percentages of depressive symptoms between the 2 groups was 3.6% (95% CI –4.2% to 11.4%). In the multivariable logistic regression, the association between internet use trajectories and depressive symptoms was not significant among urban participants (OR 1.124, 95% CI 0.686-1.840; *P*=.64).

Sensitivity Analysis

We performed a sensitivity analysis to assess the robustness of our findings (Multimedia Appendix 1, Table S2). This analysis included additional adjustments for the CESD-20 scores in 2016. The results were consistent with the main analysis, with the association between internet use trajectories and depressive symptoms remaining significant among rural participants.

Discussion

Principal Findings

In this nationally representative study using data from the CFPS, 2 latent trajectory classes of internet use were identified among Chinese adolescents aged 10 to 19 years: a low-growth group (89.8%) and a high-growth group (10.2%). After delineating these distinct trajectories, we examined their associations with depressive symptoms and found that the high-growth group was associated with higher odds of depressive symptoms. Furthermore, our stratified analysis revealed variations in this association based on residential area. Specifically, while the association between trajectories of internet use and depressive symptoms remained significant among rural adolescents, it was not observed among their urban counterparts.

In this study, we observed a notable increase in the duration of internet use among adolescents over 4 years. This trend aligns with previous research indicating that adolescent internet use intensifies with age [18]. Zhu et al [39] found that adolescents tend to allocate more time to chatting or browsing the internet as they age. This increase in internet use among adolescents can be linked to the popularity of social media platforms, as well as their inclination toward social interaction and the development of virtual identities [39]. Furthermore, our analysis revealed heterogeneous developmental patterns within this trend,

XSL•FO
RenderX

with 89.8% of adolescents exhibiting gradual growth in their internet use duration. In comparison, the remaining 10.2% increased their internet use to an average level approximately 5 times the baseline by the end of the follow-up. The diverse developmental patterns of these trajectories underscore the importance of early identification of excessive internet use. Understanding the contextual drivers of internet use behaviors is essential for altering the upward trajectory.

The high-growth group was associated with higher odds of depressive symptoms compared to the low-growth group. Our findings align with prior longitudinal research indicating that adolescent internet use can act as a risk factor for developing depressive symptoms [40]. However, previous studies have largely focused on static measures of internet use levels without examining how internet use evolves during adolescence. This study contributes to the existing evidence by demonstrating how trajectories of internet use can predict subsequent depressive symptoms. Various explanations may account for the observed association between internet use and depressive symptoms. For example, internet use may affect adolescents' mental well-being by reducing time spent on physical activity and sleep [13,41]. However, even after adjusting for lifestyle factors, the association remained significant in this study, suggesting that time reallocation alone may not fully explain the relationship. Another possibility is that excessive internet use leads to reduced face-to-face interaction [42] and worsened academic performance [43], thereby increasing the risk of depressive symptoms among adolescents.

This study identified an urban-rural difference in how internet use trajectories were associated with depressive symptoms, specifically affecting adolescents residing in rural areas. Previous research has consistently highlighted the adverse effects of internet use on the mental well-being of rural adolescents in China [44,45]. For example, Zhou and Ding [46] examined the impacts of internet use across diverse adolescent groups and found stronger associations with mental health outcomes among those from less-developed regions. However, to our knowledge, this study is the first to reveal disparities between urban and rural adolescents in the impacts of long-term changes in internet use on depressive symptoms. Although rural adolescents showed a lower proportion of rapidly increasing internet use than their urban counterparts, this trajectory had a more pronounced effect on their mental health.

The primary factor driving urban-rural differences may be the insufficient parental supervision commonly experienced by rural adolescents. Only 38.3% of rural underage internet users report having their internet use regularly restricted by parents [7]. While grandparents in rural areas often provide caregiving in the absence of parents, their limited educational attainment and reliance on traditional caregiving practices may weaken their ability to effectively monitor adolescents' internet use [47,48]. This lack of supervision can increase the likelihood of internet addiction, cyberbullying victimization, and engagement in high-risk online behaviors [49-51], potentially contributing to depressive symptoms. Additionally, individuals in rural areas with lower socioeconomic status often face limited social resources and reduced social support, which may amplify the association between internet use and depressive symptoms [52]. Rural schools, for example, tend to place less emphasis on internet literacy and education, resulting in lower awareness of healthy internet use practices [53,54]. The insufficient guidance may further heighten the risks to rural adolescents' mental health.

Implications

This study provides valuable insights into the distinct trajectories of internet use and their associations with adolescent depressive symptoms. The varying initial levels displayed by the trajectories in early adolescence highlight the critical need for early identification of excessive internet use. Given the demographic differences observed in this study, more attention is warranted to adolescents from lower socioeconomic backgrounds and rural areas.

The urban-rural divide in the association between internet use trajectories and depressive symptoms further emphasizes the importance of prioritizing interventions for rural adolescents. Previous research has shown that rural adolescents in China report higher levels of depressive symptoms compared to their urban peers [55]. Our findings suggest that rural adolescents are particularly vulnerable to the adverse effects of internet use, potentially widening existing mental health disparities between urban and rural areas. Effective interventions for rural adolescents should equip them with essential internet use skills, which can be fostered through enhanced parental supervision or integrated into school education programs. Additionally, parents should serve as positive role models by dedicating more time to engaging in outdoor activities with their children, providing alternatives to excessive screen time and encouraging a balanced lifestyle.

Limitations

This study has several limitations. First, this study relied on self-reported measures of internet use and depressive symptoms. Although the CESD has demonstrated robust psychometric properties among adolescent populations [30,31], self-reported scales are subject to social desirability bias. Moreover, participants may inaccurately report their internet use due to recall bias, introducing potential measurement errors that could attenuate or exaggerate the observed associations between internet use and depressive symptoms. Future research could benefit from incorporating objective measures of internet use and clinically validated tools for assessing depressive symptoms to enhance the reliability of findings. Second, the CFPS did not assess depressive symptoms in 2014. However, baseline psychological distress was controlled using the K6 scale, which was positively correlated with CESD-20 scores in 2018. Furthermore, we found that additional adjustments for CESD-20 scores in 2016 did not alter the association between internet use trajectories and depressive symptoms, indicating the robustness of our findings. Third, the assessment of internet use duration was based on a single question, limiting insights into the specific online activities adolescents engaged in. Future research could consider incorporating comprehensive measures of diverse online activities. Fourth, our sample was limited to adolescents in China. Disparities in health issues between rural and urban areas may exhibit different patterns among adolescents from other countries. Therefore, our findings may not be generalizable

XSL•FO
RenderX

to a global population. Finally, in this study, the sample size was relatively small, particularly when considering the urban-rural stratification. Future research should aim to use larger, nationally representative datasets to further explore variations in the effects of adolescent internet use between urban and rural regions.

Conclusion

This study elucidates urban-rural differences in the associations between internet use trajectories and adolescent depressive symptoms. Over 4 years, we observed diverse trajectories of internet use among adolescents, with one subgroup showing a rapid increase in internet use duration that was associated with higher odds of depressive symptoms. Further stratified analysis revealed that this association was significant only among rural adolescents. These findings offer valuable insights into regional disparities in the impacts of adolescent internet use. Tailored intervention strategies are needed to address the challenges faced by rural adolescents, thereby mitigating the negative effects of excessive internet use on their mental health.

Acknowledgments

This study was funded by the Key Discipline Projects of the Shanghai Three-Year Action Plan for Public Health (GWVI-11.1-29), Science and Technology Commission of Shanghai Municipality (20JC1410204), and Shanghai Municipal Education Commission (2024-Sports, Health, Arts and Science Department 01-43).

Data Availability

The data that support the findings of this study are available from the Peking University Open Research Data Platform [56].

Authors' Contributions

Conceptualization and methodology: FH and YC
Funding acquisition: YC
Data curation: YL and FH
Formal analysis: YL, XG, and XY
Writing—original draft: YL, SL, and CX
Writing—review and editing: YW, MX, and YC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables S1 and S2.
[DOCX File , 21 KB - jmir_v27i1e63799_app1.docx]

References

1. Blakemore S, Mills KL. Is adolescence a sensitive period for sociocultural processing? Annu Rev Psychol 2014 Jan 03;65(1):187-207. [doi: 10.1146/annurev-psych-010213-115202] [Medline: 24016274]
2. Su Q, Liu G. Depression in Chinese adolescents from 1989 to 2018: an increasing trend and its relationship with social environments. Curr Psychol 2020 Nov 20;41(10):6966-6977. [doi: 10.1007/s12144-020-01181-6]
3. Wu Z, Wang B, Xiang Z, Zou Z, Liu Z, Long Y, et al. Increasing trends in mental health problems among urban Chinese adolescents: results from repeated cross-sectional data in Changsha 2016-2020. Front Public Health 2022;10:829674 [FREE Full text] [doi: 10.3389/fpubh.2022.829674] [Medline: 35284369]
4. Rao W, Xu D, Cao X, Wen S, Che W, Ng CH, et al. Prevalence of depressive symptoms in children and adolescents in China: A meta-analysis of observational studies. Psychiatry Res 2019 Mar;272:790-796. [doi: 10.1016/j.psychres.2018.12.133] [Medline: 30832200]
5. Weavers B, Heron J, Thapar AK, Stephens A, Lennon J, Bevan Jones R, et al. The antecedents and outcomes of persistent and remitting adolescent depressive symptom trajectories: a longitudinal, population-based English study. Lancet Psychiatry 2021 Dec;8(12):1053-1061. [doi: 10.1016/S2215-0366(21)00281-9] [Medline: 34672991]
6. Chen JJ, Bai J. Internet use and academic achievement among Chinese adolescents: examining the mediating role of future orientation in a rural-urban dual system. Psychol Res Behav Manag 2022;15:2439-2448 [FREE Full text] [doi: 10.2147/PRBM.S343199] [Medline: 36093412]
7. 2021 national report on internet use by minors. China Internet Network Information Center. 2022. URL: https://www.cnnic.net.cn/NMediaFile/2022/1201/MAIN1669871621762HOSKOXCEP1.pdf [accessed 2024-11-13]
8. Li H, Zhu S, Wu D, Yang HH, Guo Q. Impact of information literacy, self-directed learning skills, and academic emotions on high school students' online learning engagement: A structural equation modeling analysis. Educ Inf Technol (Dordr) 2023 Mar 30:1-20 [FREE Full text] [doi: 10.1007/s10639-023-11760-2] [Medline: 37361741]

XSL•FO
RenderX

9. Liu Y, Li S, Deng T, Li L, Wei R, Zhang Y, et al. The association between green space around schools, screen time for entertainment, and adolescent depressive symptoms: A nationwide study from China. Environ Res 2024 Dec 15;263(Pt 2):120100. [doi: 10.1016/j.envres.2024.120100] [Medline: 39374752]
10. Otsuka Y, Kaneita Y, Itani O, Matsumoto Y, Jike M, Higuchi S, et al. The association between Internet usage and sleep problems among Japanese adolescents: three repeated cross-sectional studies. Sleep 2021 Dec 10;44(12):zsab175. [doi: 10.1093/sleep/zsab175] [Medline: 34252182]
11. Shen C, Dumontheil I, Thomas M, Röösli M, Elliott P, Toledano M. Digital technology use and BMI: evidence from a cross-sectional analysis of an adolescent cohort study. J Med Internet Res 2021 Jul 05;23(7):e26485 [FREE Full text] [doi: 10.2196/26485] [Medline: 35143408]
12. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. BMJ Open 2019 Jan 03;9(1):e023191 [FREE Full text] [doi: 10.1136/bmjopen-2018-023191] [Medline: 30606703]
13. Wang W, Du X, Guo Y, Li W, Zhang S, Zhang W, et al. Associations among screen time, sleep duration and depressive symptoms among Chinese adolescents. J Affect Disord 2021 Apr 01;284:69-74. [doi: 10.1016/j.jad.2021.01.082] [Medline: 33582434]
14. Hoare E, Milton K, Foster C, Allender S. Depression, psychological distress and Internet use among community-based Australian adolescents: a cross-sectional study. BMC Public Health 2017 Apr 27;17(1):365 [FREE Full text] [doi: 10.1186/s12889-017-4272-1] [Medline: 28449667]
15. Wang H, Bragg F, Guan Y, Zhong J, Li N, Pan J, et al. Association between duration of electronic screen use for non-educational purposes and depression symptoms among middle and high school students: a cross-sectional study in Zhejiang Province, China. Front Public Health 2023;11:1138152 [FREE Full text] [doi: 10.3389/fpubh.2023.1138152] [Medline: 37261230]
16. Strictly control the time periods and duration of online gaming for minors. The Central People's Government of the People's Republic of China. 2019. URL: https://www.gov.cn/fuwu/2019-11/11/content_5450800.htm [accessed 2024-11-13]
17. Guidelines for the construction of minor mode on mobile internet (draft for comments). Cyberspace Administration of China. 2023. URL: https://www.cac.gov.cn/2023-08/02/c_1692541991073784.htm [accessed 2024-11-13]
18. Sina E, Buck C, Veidebaum T, Siani A, Reisch L, Pohlabeln H, IDEFICS, I.Family consortia. Media use trajectories and risk of metabolic syndrome in European children and adolescents: the IDEFICS/I.Family cohort. Int J Behav Nutr Phys Act 2021 Oct 18;18(1):134 [FREE Full text] [doi: 10.1186/s12966-021-01186-9] [Medline: 34663352]
19. Nivins S, Sauce B, Liebherr M, Judd N, Klingberg T. Long-term impact of digital media on brain development in children. Sci Rep 2024 Jun 06;14(1):13030 [FREE Full text] [doi: 10.1038/s41598-024-63566-y] [Medline: 38844772]
20. Shen B, Cui G, Bo J. How does change in leisure-time physical activity influence the growth trajectory of depressive symptoms in college students? J Am Coll Health 2023 Sep 18:1-8. [doi: 10.1080/07448481.2023.2252503] [Medline: 37722822]
21. Yu JJ, Park SJ. Predictors and the distal outcome of general internet use: the identification of children's developmental trajectories. Br J Dev Psychol 2017 Nov;35(4):483-497. [doi: 10.1111/bjdp.12188] [Medline: 28513971]
22. Shao IY, Yang J, Ganson KT, Baker FC, Nagata JM. Identification and characterization of screen use trajectories from late childhood to adolescence in a US-population based cohort study. Prev Med Rep 2023 Dec;36:102428 [FREE Full text] [doi: 10.1016/j.pmedr.2023.102428] [Medline: 37766729]
23. Barnett TA, O'Loughlin J, Sabiston CM, Karp I, Bélanger M, Van Hulst A, et al. Teens and screens: the influence of screen time on adiposity in adolescents. Am J Epidemiol 2010 Aug 01;172(3):255-262. [doi: 10.1093/aje/kwq125] [Medline: 20616201]
24. Cai Y, Zhu X, Wu X. Overweight, obesity, and screen-time viewing among Chinese school-aged children: national prevalence estimates from the 2016 Physical Activity and Fitness in China-The Youth Study. J Sport Health Sci 2017 Dec;6(4):404-409 [FREE Full text] [doi: 10.1016/j.jshs.2017.09.002] [Medline: 30356639]
25. Nigg C, Weber C, Schipperijn J, Reichert M, Oriwol D, Worth A, et al. Urban-rural differences in children's and adolescent's physical activity and screen-time trends across 15 years. Health Educ Behav 2022 May 04:10901981221090153. [doi: 10.1177/10901981221090153] [Medline: 35506637]
26. Population status of children in china in 2015: facts and figures. United Nations International Children's Emergency Fund. 2017. URL: https://www.unicef.cn/media/9901/file/Population%20Status%20of%20Children%20in%20China%20in%202015%20Facts%20and%20Figures.pdf [accessed 2024-05-06]
27. Wu Q, Qian H. Comparing long- versus short-forms of depression scales in an omnibus longitudinal survey. J Off Stat 2024 Jul 31;40(3):457-471. [doi: 10.1177/0282423x241249104]
28. Budescu DV. Selecting an equating method: linear or equipercentile? J Educ Behav Stat 1987 Mar 01;12(1):33-43. [doi: 10.3102/10769986012001033]
29. Silberg JL, Heath AC, Kessler R, Neale MC, Meyer JM, Eaves LJ, et al. Genetic and environmental effects on self-reported depressive symptoms in a general population twin sample. J Psychiatr Res 1990;24(3):197-212 [FREE Full text] [doi: 10.1016/0022-3956(90)90010-n] [Medline: 2266509]

XSL•FO
RenderX

30. Wight RG, Sepúlveda JE, Aneshensel CS. Depressive symptoms: how do adolescents compare with adults? J Adolesc Health 2004 Apr;34(4):314-323. [doi: 10.1016/j.jadohealth.2003.05.003] [Medline: 15041001]
31. Phillips GA, Shadish WR, Murray DM, Kubik M, Lytle LA, Birnbaum AS. The Center for Epidemiologic Studies Depression Scale with a young adolescent population: a confirmatory factor analysis. Multivariate Behav Res 2006 Jun 01;41(2):147-163. [doi: 10.1207/s15327906mbr4102_3] [Medline: 26782908]
32. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
33. Mewton L, Kessler RC, Slade T, Hobbs MJ, Brownhill L, Birrell L, et al. The psychometric properties of the Kessler Psychological Distress Scale (K6) in a general population sample of adolescents. Psychol Assess 2016 Oct;28(10):1232-1242. [doi: 10.1037/pas0000239] [Medline: 26619095]
34. Tran TD, Kaligis F, Wiguna T, Willenberg L, Nguyen HTM, Luchters S, et al. Screening for depressive and anxiety disorders among adolescents in Indonesia: Formal validation of the centre for epidemiologic studies depression scale - revised and the Kessler psychological distress scale. J Affect Disord 2019 Mar 01;246:189-194. [doi: 10.1016/j.jad.2018.12.042] [Medline: 30583144]
35. Jung T, Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. Social Pers Psych Compass 2007 Dec 06;2(1):302-317. [doi: 10.1111/j.1751-9004.2007.00054.x]
36. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equa Mod: Multidiscipl J 2007 Dec 05;14(4):535-569. [doi: 10.1080/10705510701575396]
37. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
38. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 01;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
39. Zhu X, Griffiths H, Xiao Z, Ribeaud D, Eisner M, Yang Y, et al. Trajectories of screen time across adolescence and their associations with adulthood mental health and behavioral outcomes. J Youth Adolesc 2023 Jul;52(7):1433-1447 [FREE Full text] [doi: 10.1007/s10964-023-01782-x] [Medline: 37148440]
40. Fitzpatrick C, Lemieux A, Smith J, West GL, Bohbot V, Asbridge M. Is adolescent internet use a risk factor for the development of depression symptoms or vice-versa? Psychol Med 2023 Oct;53(14):6773-6779 [FREE Full text] [doi: 10.1017/S0033291723000284] [Medline: 36825394]
41. Khan A, Burton NW. Is physical inactivity associated with depressive symptoms among adolescents with high screen time? Evidence from a developing country. Ment Health Phys Act 2017 Mar;12:94-99. [doi: 10.1016/j.mhpa.2017.03.001]
42. Twenge JM, Spitzberg BH, Campbell WK. Less in-person social interaction with peers among U.S. adolescents in the 21st century and links to loneliness. J Soc Pers Relat 2019 Mar 19;36(6):1892-1913. [doi: 10.1177/0265407519836170]
43. Poulain T, Peschel T, Vogel M, Jurkutat A, Kiess W. Cross-sectional and longitudinal associations of screen time and physical activity with school performance at different types of secondary school. BMC Public Health 2018 Apr 27;18(1):563 [FREE Full text] [doi: 10.1186/s12889-018-5489-3] [Medline: 29703168]
44. Ma J, Sheng L. Internet use time and mental health among rural adolescents in China: a longitudinal study. J Affect Disord 2023 Sep 15;337:18-26. [doi: 10.1016/j.jad.2023.05.054] [Medline: 37236271]
45. Wang H, Abbey C, Kennedy T, Feng E, Li R, Liu F, et al. The association between screen time and outdoor time on adolescent mental health and academic performance: evidence from rural China. Risk Manag Healthc Policy 2023;16:369-381 [FREE Full text] [doi: 10.2147/RMHP.S384997] [Medline: 36923495]
46. Zhou M, Ding X. Internet use, depression, and cognitive outcomes among Chinese adolescents. J Community Psychol 2023 Mar;51(2):768-787. [doi: 10.1002/jcop.22779] [Medline: 34927738]
47. Wang X, Zhang Y. Intergenerational care and rural childhood obesity in the digital era: Based on screen exposure perspective. SSM Popul Health 2024 Sep;27:101694 [FREE Full text] [doi: 10.1016/j.ssmph.2024.101694] [Medline: 39055642]
48. Wang SX, Bansak C. Are grandparents a good substitute for parents as the primary caregiver? The impact of grandparents on Children's academic performance. Labour Econ 2024 Jun;88:102545. [doi: 10.1016/j.labeco.2024.102545]
49. Zhu X, Deng C, Bai W. Parental control and adolescent internet addiction: the moderating effect of parent-child relationships. Front Public Health 2023;11:1190534 [FREE Full text] [doi: 10.3389/fpubh.2023.1190534] [Medline: 37304126]
50. Álvarez-García D, Núñez JC, González-Castro P, Rodríguez C, Cerezo R. The effect of parental control on cyber-victimization in adolescence: the mediating role of impulsivity and high-risk behaviors. Front Psychol 2019;10:1159 [FREE Full text] [doi: 10.3389/fpsyg.2019.01159] [Medline: 31178790]
51. Khurana A, Bleakley A, Jordan AB, Romer D. The protective effects of parental monitoring and internet restriction on adolescents' risk of online harassment. J Youth Adolesc 2015 May;44(5):1039-1047. [doi: 10.1007/s10964-014-0242-4] [Medline: 25504217]
52. Wang T, Zhang S, Li Q, Liu H, Zhang S, Jian W, et al. Socioeconomic inequalities in the relationship between internet usage patterns and depressive symptoms: evidence from a Chinese longitudinal study. J Glob Health 2024 Aug 16;14:04127 [FREE Full text] [doi: 10.7189/jogh.14.04127] [Medline: 39149826]

XSL•FO
RenderX

53. Li J, Huang X, Lei X, Wen J, Lu M. ICT literacy, resilience and online learning self-efficacy between Chinese rural and urban primary school students. Front Psychol 2022;13:1051803 [FREE Full text] [doi: 10.3389/fpsyg.2022.1051803] [Medline: 36578676]
54. Li Y, Ranieri M. Educational and social correlates of the digital divide for rural and urban children: A study on primary school students in a provincial city of China. Comp Educ 2013 Jan;60(1):197-209. [doi: 10.1016/j.compedu.2012.08.001]
55. Zeng J, Xu Y. Family socioeconomic status and adolescent depression in urban and rural China: a trajectory analysis. SSM Popul Health 2024 Mar;25:101627 [FREE Full text] [doi: 10.1016/j.ssmph.2024.101627] [Medline: 38426030]
56. China Family Panel Studies (CFPS). Peking University Open Research Data Platform. URL: https://opendata.pku.edu.cn/dataset.xhtml?persistentId=doi:10.18170/DVN/45LCSO [accessed 2025-01-29]

Abbreviations

aBIC: sample size–adjusted Bayesian information criterion
AIC: Akaike information criterion
BIC: Bayesian information criterion
BLRT: bootstrapped likelihood ratio test
CESD: Center for Epidemiologic Studies Depression Scale
CFPS: China Family Panel Study
K6: Kessler Psychological Distress Scale
LCGM: latent class growth modeling
LMR: Lo-Mendell-Rubin likelihood ratio test
MCAR: missing completely at random

Edited by T de Azevedo Cardoso; submitted 30.06.24; peer-reviewed by F Yang, N Kaur, G Mayer; comments to author 07.11.24; revised version received 04.12.24; accepted 20.01.25; published 07.02.25.

Please cite as:
Liu Y, Ge X, Wang Y, Yang X, Liu S, Xu C, Xiang M, Hu F, Cai Y
Urban-Rural Differences in the Association Between Internet Use Trajectories and Depressive Symptoms in Chinese Adolescents: Longitudinal Observational Study
J Med Internet Res 2025;27:e63799
URL: https://www.jmir.org/2025/1/e63799
doi:10.2196/63799
PMID:

©Yujie Liu, Xin Ge, Ying Wang, Xue Yang, Shangbin Liu, Chen Xu, Mi Xiang, Fan Hu, Yong Cai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of an Internet-Based, Self-Guided, Short-Term Mindfulness Training (ISSMT) Program for Relieving Depressive Symptoms in the Adult Population in China: Single-Blind, Randomized Controlled Trial

Tingfei Zhu[1], PhD; Liuyi Zhang[2], MSc; Wenqi Weng[2], MSc; Ruochen Gan[2], PhD; Limin Sun[2], MSc; Yanping Wei[2], PhD; Yueping Zhu[2], MSc; Hongyan Yu[3], PhD; Jiang Xue[4,5], PhD; Shulin Chen[2], MD, PhD

[1]Counseling and Psychological Services, Shanghai Jiao Tong University, Shanghai, China

[2]Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

[3]Department of Physical Education, Shanghai Jiao Tong University, Shanghai, China

[4]The Affiliated Dongguan Songshan Lake Central Hospital, Guangdong Medical University, Dongguan, China

[5]School of Humanities and Management, Guangdong Medical University, Dongguan, China

Corresponding Author:
Shulin Chen, MD, PhD
Department of Psychology and Behavioral Sciences
Zhejiang University
No.148 Tianmushan Road
Hangzhou
China
Phone: 86 18958001648
Email: chenshulin@zju.edu.cn

Abstract

Background: Depression is a significant global public health issue, and in China, access to mental health services remains limited despite high demand. Research has shown that mindfulness can effectively alleviate depressive symptoms and that telehealth solutions offer a promising avenue for addressing this service gap. Despite this potential, there are currently few studies in China focusing on short-term online mindfulness training. Most existing online mindfulness studies relied on traditional 8-week programs, which can be challenging for participant adherence due to limited accessibility and high dropout rates. Additionally, limited research exists on short-term online mindfulness interventions, and findings remain inconsistent.

Objective: This study aimed to develop and evaluate an internet-based, self-guided, short-term mindfulness training (ISSMT) program based on the Monitor and Acceptance Theory (MAT) to reduce depression symptoms.

Methods: The ISSMT program was delivered via an online platform, “Hi Emotion,” and was accessible to the general public. Interested individuals aged 18 years and older were randomized into either the ISSMT group or a wait-list control group. Participants in the ISSMT group received daily reminders to participate in a 15- to 20-minute session over a 14-day training period. Measurements, including mindfulness and depressive symptoms, were collected at baseline and weekly for the subsequent 3 weeks.

Results: A total of 205 adults participated in the 14-day online intervention. Linear mixed models were used to analyze both per-protocol (PP) and intention-to-treat (ITT) samples. Compared with the wait-list control group, participants in the ISSMT group showed significant improvements in mindfulness (Cohen *d*=0.44 for ITT; Cohen *d*=0.55 for PP) and reductions in depressive symptoms (Cohen *d*=0.50 for ITT; Cohen *d*=0.53 for PP). Furthermore, participants expressed high acceptance of this training format with a relatively low dropout rate (<40%).

Conclusions: The ISSMT program based on the MAT effectively enhanced mindfulness and alleviated depressive symptoms. This intervention could be considered for integration into psychosocial service systems to improve mental health outcomes and help bridge the gap between limited resources and the high demand for services in China. Future research should focus on personalizing these programs and incorporating advanced technologies to enhance their effectiveness and user engagement.

Trial Registration: Open Science Framework; https://doi.org/10.17605/OSF.IO/8P4V6

XSL•FO
RenderX

(J Med Internet Res 2025;27:e55583) doi:10.2196/55583

KEYWORDS

internet-based; self-guided; short-term; mindfulness; depression; randomized controlled trial

Introduction

Depression is a leading global public health issue, with a large impact on people's well-being and quality of life and causing substantial economic ramifications and mental health–related disability [1]. According to the World Health Organization, approximately 280 million people worldwide, representing 5% of the adult population, grapple with depression [2,3]. The China Mental Health Survey reported that the weighted prevalence of depressive disorders in China is 3.6% (95% CI 3.0%-4.2%) during the previous 12 months and the lifetime weighted prevalence was 6.8% (95% CI 5.8%-7.8%) [4].

Given the considerable demand for mental health services, access to mental health services in China is limited. As of 2022, there were only 2098 psychiatric hospitals in China that were staffed by approximately 273,000 professionals, resulting in an average of 3.67 professionals serving 100,000 people [5]. Compounding this issue is the pervasive stigma associated with seeking mental health services. A recent study revealed that, in China, fewer than 10% of patients with depression received any form of treatment, with only 0.5% receiving adequate care [6].

The emergence of telehealth solutions has shown promise in bridging the mental health gap, especially since the COVID-19 pandemic. Telehealth solutions offer several key advantages including high acceptability both as a supplement and alternative to offline face-to-face training. A survey of 500 adults revealed a high interest (71.2%) in internet-based psychological training, with 42% expressing a preference for internet-based training over individual training (37.8%) and group training (19.6%) [7]. Telehealth solutions also provide a high level of accessibility, and users have the flexibility to use medical services according to their own schedule. In addition, telehealth offers greater anonymity, which is valuable for managing sensitive information and providing a more comfortable, private environment for users [8]. Compared with face-to-face training, the development and maintenance of online health care require relatively fewer economic and human resources [9]. Finally, several studies have indicated that some evidence-based psychological training can be effective in an online format, significantly ameliorating users' depression and anxiety. The effect size (d=0.80) of self-help training protocols is only slightly lower than that of standardized treatment protocols (d=0.82) [10-15].

Mindfulness-based training has gained increasing attention in recent years [16-20], with substantial evidence supporting its positive impact on depression [19,20]. Mindfulness is a moment-by-moment awareness of thoughts, feelings, bodily sensations, and the surrounding environment [21]. Among the various mindfulness-based interventions, 2 of the most widely implemented are Mindfulness-Based Stress Reduction (MBSR) and mindfulness-based cognitive therapy [22]. MBSR, developed by Kabat-Zinn [23], and mindfulness-based cognitive therapy, created by Teasdale et al [24] as an extension of MBSR, are central to mindfulness practice. Both interventions typically include 8 weekly sessions focused on mindfulness techniques. A meta-analysis of 44 meta-analyses encompassing various populations, problems, training, comparisons, and outcomes revealed that mindfulness training demonstrated significant effects and advantages (d=0.10-0.89), including for depression (d=0.49), compared with negative controls [20].

The theoretical mechanisms underlying mindfulness interventions are supported by several key frameworks, including the Monitor and Acceptance Theory (MAT) [25], Mindfulness-to-Meaning Theory [26,27], and self-awareness, self-regulation, and self-transcendence framework [28]. Among these frameworks, MAT posits that attention monitoring and acceptance are essential mechanisms of mindfulness [25]. This dual-process approach enhances individuals' ability to monitor internal experiences and accept them without judgment, which is fundamental to reducing negative emotions and fostering psychological flexibility, a key function in emotional regulation.

Compared with offline training, online mindfulness programs generally require less direct involvement from highly trained therapists, reducing reliance on professionals and enhancing accessibility [22]. Some studies have shown that online mindfulness programs, when matched for duration and structure, have slightly weaker effects than offline versions but still show significantly greater improvements than a control group [29]. Furthermore, the efficacy of online programs for alleviating depression has been supported by research in some western countries [15,30].

Most online mindfulness training programs are adaptations of traditional 8-week offline programs [31]. Although traditional programs have demonstrated efficacy, their extended duration and demanding requirements often result in challenges such as limited accessibility [31] and high dropout rates [32]. The few studies exploring the effect of simplified, short-term online mindfulness programs have shown inconsistent [33-36] and unstable effects [37,38]. Some studies with 2- to 4-week online mindfulness training reported positive results [33-35], while others, such as the 2-week self-help program evaluated by Glück and Maercker [36], found no significant differences between the mindfulness group and wait-list control group in terms of mindfulness scores, stress perception, overall distress, and more. To our knowledge, the feasibility and effectiveness of brief online mindfulness training in the Chinese population have rarely been investigated [31].

Current brief online mindfulness interventions often emphasize basic mindfulness techniques [39], primarily attention monitoring. MAT highlights the importance of incorporating acceptance, which is crucial for modulating emotional reactivity and stress responses, ultimately improving mental and physical well-being [25]. Interventions based on MAT may potentially address the limitations of existing internet-delivered programs

XSL•FO
RenderX

by integrating both attention monitoring and acceptance. Randomized controlled trials (RCTs) have demonstrated that MAT-based training significantly improves stress-related indicators, including reductions in cortisol levels and systolic blood pressure reactivity [40], while also enhancing positive emotions [41].

The primary aim of this study was to implement an internet-based, self-guided, short-term mindfulness training (ISSMT) program based on MAT via a mobile platform with a Chinese population and to verify its effectiveness for depressive symptoms using an RCT design. We hypothesized that the ISSMT could be effectively delivered online and would lead to significant improvements in participants' depressive symptoms.

Methods

This study was designed as a single-blind RCT.

Recruitment

A power analysis was conducted to estimate the sample size needed to detect the main effects and interactions in this study. Based on an effect size of 0.3 [20], a power of 0.80, and an alpha of .05, the initial calculation suggested that 126 participants would be required. Some researchers have reported high dropout rates with online training. For instance, an online MoodGym study reported a dropout rate of 73.9% for research participants and 99.2% for public registrants [42], and completion rates for internet-based depression treatment programs have varied from 43% to 99% [43]. To account for a potential dropout rate of 50%, which is commonly observed in online interventions, we planned to recruit 250 participants.

From January 2021 to April 2021, participants were recruited by posting recruitment posters on various public social media platforms, including WeChat Moment, RED app, Douban Forum, and other online forums. Interested individuals were directed to an online questionnaire, where they could register by scanning the QR code on the recruitment poster and providing basic demographic information. A total of 328 participants voluntarily enrolled in this study.

The inclusion criteria were as follows: (1) aged 18 years or older; (2) possessed intact reading and comprehension abilities; (3) had the capacity to provide informed consent; and (4) had access to the internet and basic smartphone literacy. Exclusion criteria included (1) current use of psychotropic medications, (2) severe visual and auditory impairments that hindered the comprehension of audio content or receipt of information, and (3) concurrent participation in other psychosocial intervention programs.

To enhance data representativeness, recruitment efforts spanned multiple platforms and were conducted at different times of the day to capture a diverse participant pool. Additionally, we used the neutral phrase "Peace Journey" in recruitment materials to mitigate self-selection bias.

Finally, 81 individuals who were unreachable after registration were excluded, leaving 247 eligible participants enrolled for randomization and the subsequent intervention.

Procedure

After the initial registration, potential participants were invited to join a WeChat group for further communication. Preliminary screening was conducted based on the predefined inclusion and exclusion criteria. Eligible participants were invited to participate in subsequent research, were informed of the entire research process, and signed an electronic informed consent form. Baseline data were collected after receiving the informed consent form.

The unit of randomization in this study was the individual. Using computer-generated random numbers, we randomly assigned all eligible participants to either the ISSMT group or the wait-list control group. A statistician generated the randomization sequence, and baseline data were collected after the randomization. Investigators remained blinded to the randomization process. Notably, the investigation of online mindfulness training by Krusche et al [34] revealed a dropout rate of 79.4% for the training group, contrasting with only 35.9% for the wait-list control group. To balance the retained sample size in the 2 arms and adhere to statistical standards, we randomly assigned participants to groups at a ratio of 2.5:1 during randomization.

Ultimately, 175 participants were allocated to the ISSMT group, while 72 were assigned to the wait-list control group. Those in the control group received the same mindfulness training after completion of the formal study.

Intervention

Training Course

In this study, the ISSMT program was translated and adapted from a 14-day mindfulness training program developed based on MAT by Lindsay et al [25]. With authorization from Lindsay and colleagues, a team of 4 psychology graduate students was assembled to perform the translation and adaptation. Each unit underwent a standardized multistep process, including sequential stages of "translation, peer cross-check, revision, group discussion and proofreading, and revision," culminating in the creation of a draft. Team members then rotated assigned sections and completed a secondary round of proofreading. Following this, an expert in psychology was invited to conduct a comprehensive proofread of the entire text. Any segments identified as unclear were further refined in alignment with the original text until all team members reached consensus on clarity and accuracy.

The draft was subsequently recorded by a mindfulness practitioner with over 5 years of experience. We then invited 30 individuals to listen to the recordings and provide feedback, which was used to refine the audio content to finalize the training course. The finalized training framework, with each section lasting approximately 15 minutes to 20 minutes, is shown in Table 1.

XSL•FO
RenderX

Table 1. Lesson content in each training program based on the Monitor and Acceptance Theory (MAT).

Lesson	Content
1	Introduction: introduction to the course and 3 core skills: concentration, equanimity, and clarity
2	Concentration I: developing a deeper understanding of concentration
3	Concentration II: concentrating continuously on body experiences
4	Concentration III: maintaining focus on body experiences while listening to someone speak
5	Concentration IV: labeling body experiences to maintain focus
6	Concentration V: labeling different types of body experiences
7	Equanimity I: maintaining global body relaxation to promote equanimity
8	Equanimity II: promoting equanimity by intentionally using a matter-of-fact tone of voice when labeling
9	Clarity I: discriminating different types and patterns of body sensations
10	Clarity II: detecting subtle or faint body sensations and increasing sensual fulfillment by detecting subtle pleasure
11	Equanimity III: developing equanimity by applying a welcoming attitude toward all experiences
12	Clarity III: recognizing 4 basic categories of body experience (physical, emotional, restful, and “energy flow”)
13	Equanimity IV: integrating the 3 equanimity strategies: body relaxation, tone of voice, and welcoming attitude
14	Course Review: guided practice through the major strategies learned in the preceding 13 lessons

Intervention Platform

An online training platform, named “Hi Emotion,” was developed to facilitate the intervention. To access the training course, participants were only required to navigate to the “Hi Emotion” module within the “Mindfulness + Technology” section on WeChat. The platform accommodates 2 roles: “leader” and “practitioner” (see Figure S1A in Multimedia Appendix 1). The leader’s role involves creating groups and administering online courses, while practitioners have the option to join specific groups via the “Group” section. Once the “leader” disseminates the daily course, it becomes immediately visible to the “practitioner” in the “Unlearned Courses” section. Upon completion, the course is removed from the “Unlearned Courses” and relocated to the “Learned Courses” section.

Intervention Procedure

Following the randomization process, the researcher, acting as the “leader,” established a training group and distributed the group code to participants allocated to the ISSMT group. These participants then joined the group and familiarized themselves with the platform using the provided instruction manual. During the 14-day training period, the course was delivered daily on the platform at 10 AM. Participants received immediate reminders in the WeChat group after each course delivery, with additional reminders sent at 8 PM each evening.

The training lasted for 14 days, with each individual course lasting between 15 minutes and 20 minutes. Upon completion of the daily training course, participants were required to engage in mindfulness meditation practices outside the online course for a minimum of 30 minutes per day. In addition, participants were encouraged to incorporate mindfulness practices into their daily activities, such as during meals, while standing, and during walking.

To minimize the dropout rate, several measures were implemented. During the intervention, a trained experimenter sent timely course reminders in the WeChat group every day (at 10 AM and 8 PM) and addressed any participant queries. On the 4th and 10th days of the training, the researchers monitored the participants’ progress through the backend database. If any participants exhibited signs of potential dropout (such as incomplete listening records), the experimenter would contact them via WeChat to ensure they understood the course material and to encourage their continued participation.

Measures

Demographic Characteristics

Demographic characteristics were assessed solely at baseline and encompassed age, gender, educational level, and occupation.

Mindfulness

The short form of the Chinese version of the 5-facet mindfulness questionnaire (FFMQ-C-SF) was used to measure mindfulness [44]. This 15-item scale measures 5 aspects of mindfulness on a 5-point Likert scale, ranging from 1 (never or very rarely true) to 5 (very often or always true), with higher total scores indicating a higher level of mindfulness. The Cronbach α for the FFMQ-C-SF in this study was 0.84.

Depressive Symptoms

Depressive symptoms were evaluated using the Beck Depression Inventory-II (BDI-II) [45]. This 21-item scale is scored on a scale from 0 to 3 according to the severity of symptoms and is one of the most widely used self-rated scales for depressive symptoms. It is applicable to the assessment of depressive symptoms in various clinical and nonclinical populations, with a total score of 0 to 13 indicating no depression, 14 to 19 indicating mild depression, 20 to 28 indicating moderate depression, and 29 to 63 indicating severe depression. The Chinese version of BDI-II, revised by Wang et al [46], demonstrated good reliability and validity, with a Cronbach α coefficient of 0.94. The Cronbach α coefficient for BDI-II in this study was 0.95.

XSL•FO
RenderX

Feasibility

The Intervention Satisfaction Scale designed by Campo et al [47] was used to measure the feasibility of participants' participation and evaluated satisfaction with the internet-based training. It includes 9 items. Items 1 ("Overall, I liked the two-week course practice") and 2 ("I would recommend this course to others") are scored on a scale from 1 to 5 (1=strongly disagree, 5=strongly agree), while the remaining 7 items are scored on a 2-point scale (1=yes, 2=no).

Data Quality Control

To ensure data quality and reliability throughout the study, we implemented multiple control measures. Standardized protocols were followed during data collection to reduce variability. During data cleaning, responses were rigorously screened, and entries displaying excessive duplication or failing to pass attention-check items were excluded. Specifically, attention checks included prompts such as "Please select the option 'Moderately agree' for this question," allowing us to effectively identify inattentive respondents. During the data collection phase, we regularly monitored participants' completion status and their engagement with the intervention (see the Intervention Procedure section for specifics) to maintain data integrity. These comprehensive monitoring and engagement strategies ensured data integrity and minimized the risk of dropout, enhancing the robustness and reliability of our study findings.

Statistical Analysis

Descriptive statistics summarized participants' demographic information and baseline outcome variables across different groups. To assess group differences, we used *t* tests for continuous variables and chi-square tests for categorical variables.

To analyze training effects in this longitudinal study, we used linear mixed models [48,49]. This approach accounts for both random and fixed effects, making it well-suited to handle missing data, which we assumed to be missing at random [50]. Specifically, compound symmetric covariance and maximum likelihood estimation were used.

We implemented both intention-to-treat (ITT) and per-protocol (PP) analyses [51]. ITT analysis includes all randomized participants, maintaining the benefits of randomization regardless of treatment adherence. PP analysis includes participants who completed the assigned treatment. In this study, the ITT analysis included participants who completed baseline measurements and participated in at least one course. The PP analysis focused on the participants who completed at least 75% of their assigned courses (ie, at least 10 sessions), which this study defined as completion of training. This approach enables a clear assessment of the actual treatment impact. Together, these methods provide a comprehensive view of the intervention's effectiveness, with ITT generally yielding more conservative estimates.

Additionally, we evaluated the following effects: the time effect captures changes in measured outcomes across different time points, reflecting how participants' performance evolved over the training duration. The group effect indicates differences in outcomes between treatment and control groups, illustrating the intervention's impact. Finally, the time × group interaction examines how changes over time may differ across groups, elucidating the differential effectiveness of the intervention [52].

All analyses were conducted using SPSS 20.0 (IBM Corp) and incorporated age, gender, and other baseline variables as covariates. Statistical significance was determined using 2-sided tests, with a threshold of $P<.05$.

Ethical Considerations

Ethics Approval

The study was approved by the Research Ethics Committee of the Department of Psychology and Behavioral Sciences, Zhejiang University (register number: 2020-059).

All procedures performed were in accordance with the ethical standards of the Zhejiang University research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent

Informed consent was obtained online from all participants before their involvement in the study.

Privacy and Confidentiality

Although the study involved repeated measurements from the same participants, the data collected were anonymized for analysis and feedback purposes. No personally identifiable information was included in the data nor the personalized feedback reports provided to participants.

Compensation Details

Participants received personalized feedback reports containing tailored insights based on their responses as compensation for their participation.

Results

Demographic Characteristics

A total of 328 individuals voluntarily signed up to participate. Of these 328 individuals, 247 (75.3%) met the inclusion criteria and were randomized into the study. Data from 42 people who did not engage in any subsequent courses were excluded, leaving 205 eligible participants (133 in the training group and 72 in the wait-list control group). For the purposes of this study, individuals who completed more than 75% of the courses (ie, participation in at least 10 courses) were considered to have completed the training. A total of 81 participants (81/131, 60.9%) in the training group met this criterion (see Figure 1). No adverse events were reported during the study. Finally, data from 133 individuals in the training group and 72 individuals in the wait-list control group were included in the ITT analysis, while data from 81 and 72 participants, respectively, were included in the PP analysis.

The basic demographic information of the participants is shown in Table 2. The average ages of the participants in the training group and wait-list control group were 24.80 (SD 6.31) years

XSL•FO
RenderX

and 24.60 (SD 6.62) years, respectively, with no statistically significant difference. There were also no significant differences between the 2 groups in terms of gender, education, or baseline mindfulness and depressive symptoms.

Figure 1. Research procedures. ISSMT: internet-based, self-guided, short-term mindfulness training; ITT: intention-to-treat; PP: per-protocol.

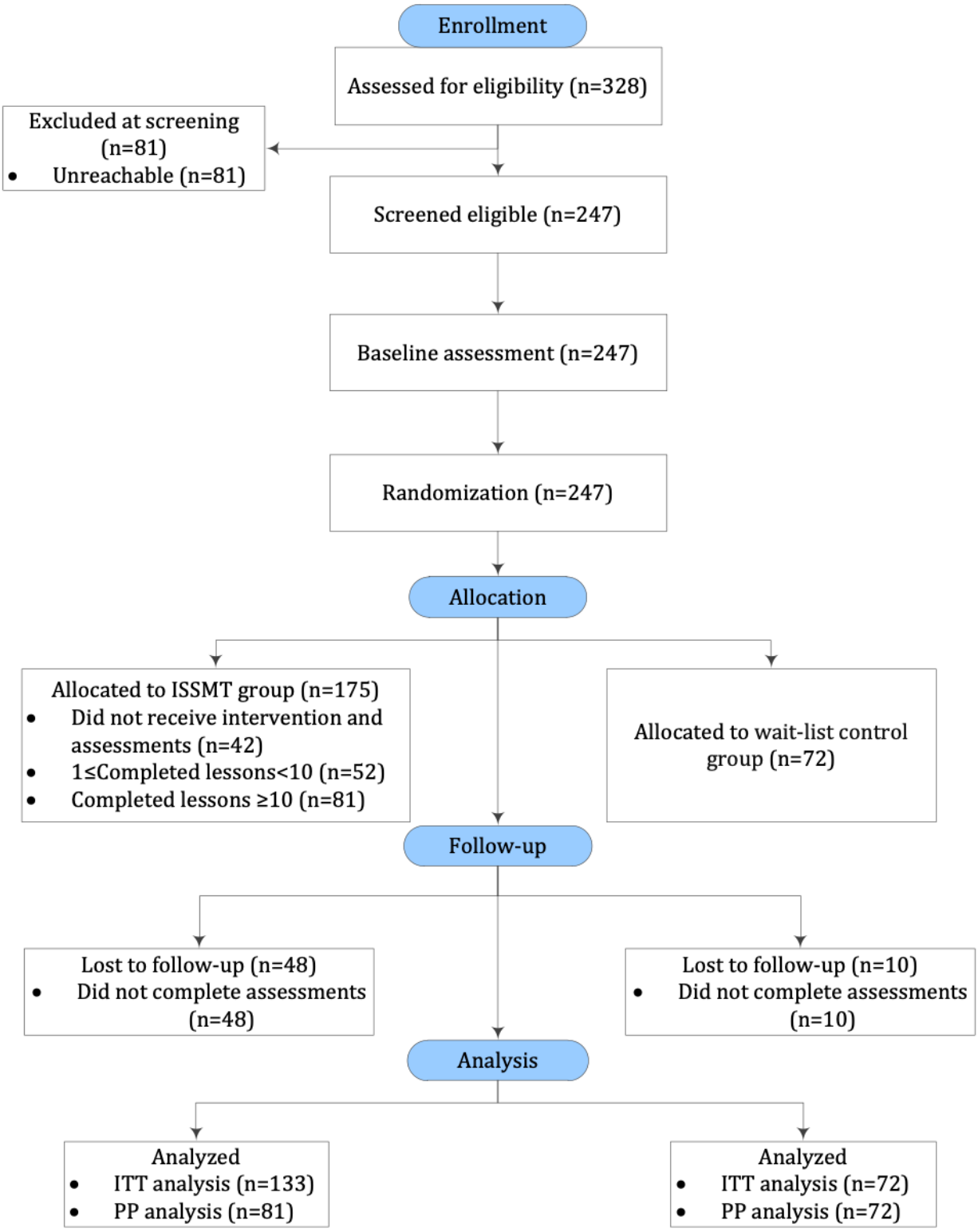

XSL•FO
RenderX

Table 2. Demographic characteristics (N=205).

Demographics characteristics	ISSMT[a] group (n=133)	Wait-list control group (n=72)	Statistic (*df*)	*P* value
Age (years), mean (SD)	24.8 (6.31)	24.6 (6.62)	0.21 (203)[b]	.83
Female gender, n (%)	98 (73.7)	56 (77.8)	0.42 (1)[c]	.52
Education (years), n (%)			2.67 (2)[c]	.26
<12	5 (3.8)	3 (4.2)		
12	70 (52.6)	46 (63.9)		
>12	58 (43.6)	23 (31.9)		
Baseline FFMQ-C-SF[d] score, mean (SD)	43.98 (8.04)	43.54 (8.92)	0.36 (203)[b]	.72
Baseline BDI-II[e] score, mean (SD)	12.81 (11.26)	14.84 (10.63)	–1.24 (203)[b]	.22

[a]ISSMT: internet-based, self-guided, short-term mindfulness training.

[b]*t* test.

[c]Chi-square test.

[d]FFMQ-C-SF: short form of the Chinese version of the 5-facet mindfulness questionnaire.

[e]BDI-II: Beck Depression Inventory-II.

Training Effects on Mindfulness

According to the ITT analysis, the scores for each outcome variable at each measurement time point are presented in Table 3. At baseline, there was no significant difference in mindfulness scores between the training group and the wait-list control group. The results of the mixed linear model indicated significant main effects of time ($F_{3,515.19}$=29.48; P<.001), group ($F_{1,214.78}$=10.20; P=.002), and the time × group interaction ($F_{3,515.25}$=7.05; P<.001). The results suggest that the online mindfulness training significantly improved the participants' mindfulness. The effect size for the change in mindfulness level was 0.44.

As shown in Table 3, the training group demonstrated a significant increase in mindfulness starting from the first week, with continued significant improvements thereafter, whereas no significant changes were observed in the wait-list control group. The trends in mindfulness level changes at each time point are illustrated in Figure 2A.

According to the PP analysis, the scores for each outcome variable at each measurement time point are presented in Table 4. At baseline, there was no significant difference in mindfulness scores between the training group and wait-list control group. The results of the mixed linear model revealed significant main effects of time ($F_{3,419.57}$ =29.31; P<.001), group ($F_{1,158.47}$=11.21; P=.001), and the time × group interaction ($F_{3,419.64}$=7.67; P<.001). These results indicate that the online mindfulness training significantly improved the participants' mindfulness level. The effect size for the change in mindfulness level was 0.55.

As shown in Table 4, the mindfulness level in the training group showed a marginally significant increase starting from the first week and continued to increase significantly thereafter, while no significant changes were observed in the wait-list control group. The trends in mindfulness level changes at each time point are depicted in Figure 3A.

XSL•FO
RenderX

Table 3. Intention-to-treat analysis of the improvements in mindfulness and depression after online mindfulness training.

Time point	ISSMT[a] group (n=133), mean (SD)	Wait-list control group (n=72), mean (SD)	*t* (*df*)	*P* value	Time effect		Group effect		Time × group interaction		Effect size (Cohen *d*)
					F (*df*)	*P* value	*F* (*df*)	*P* value	*F* (*df*)	*P* value	
FFMQ-C-SF[b]											
T_0[c]	43.98 (8.04)	43.54 (8.92)	0.36 (203)	.72	—[d]	—	—	—	—	—	—
T_1[e]	46.09 (8.25)	45.09 (9.18)	0.74 (173)	.46	—	—	—	—	—	—	—
T_2[f]	48.26 (8.19)	44.84 (9.25)	2.41 (152)	.017	—	—	—	—	—	—	—
T_3[g]	48.93 (8.08)	45 (9.75)	2.67 (146)	.008	29.48 (3, 515.19)	<.001	10.20 (1, 214.78)	.002	7.05 (3, 515.25)	<.001	0.44
BDI-II[h]											
T_0	12.81 (11.26)	14.84 (10.63)	–1.24 (203)	.22	—	—	—	—	—	—	—
T_1	9.72 (10.02)	11.93 (11.24)	–1.33 (173)	.18	—	—	—	—	—	—	—
T_2	6.99 (9.55)	11.26 (12.04)	–2.33 (107.71)	.02	—	—	—	—	—	—	—
T_3	5.70 (7.82)	10.98 (12.78)	–2.89 (93.50)	.005	38.48 (3, 514.77)	<.001	9.31 (1, 207.10)	0.003	3.87 (3, 514.87)	.009	0.50

[a]ISSMT: internet-based, self-guided, short-term mindfulness training.

[b]FFMQ-C-SF: short form of the Chinese version of the 5-facet mindfulness questionnaire.

[c]T_0: baseline.

[d]Not applicable.

[e]T_1: 1st week.

[f]T_2: 2nd week.

[g]T_3: 3rd week.

[h]BDI-II: Beck Depression Inventory-II.

XSL•FO
RenderX

Figure 2. Intention-to-treat analysis of the comparison of (A) mindfulness, as measured using the short form of the Chinese version of the 5-facet mindfulness questionnaire (FFMQ-C-SF), and (B) depression, as measured using the Beck Depression Inventory-II (BDI-II), between the training and wait-list control groups at each time point.

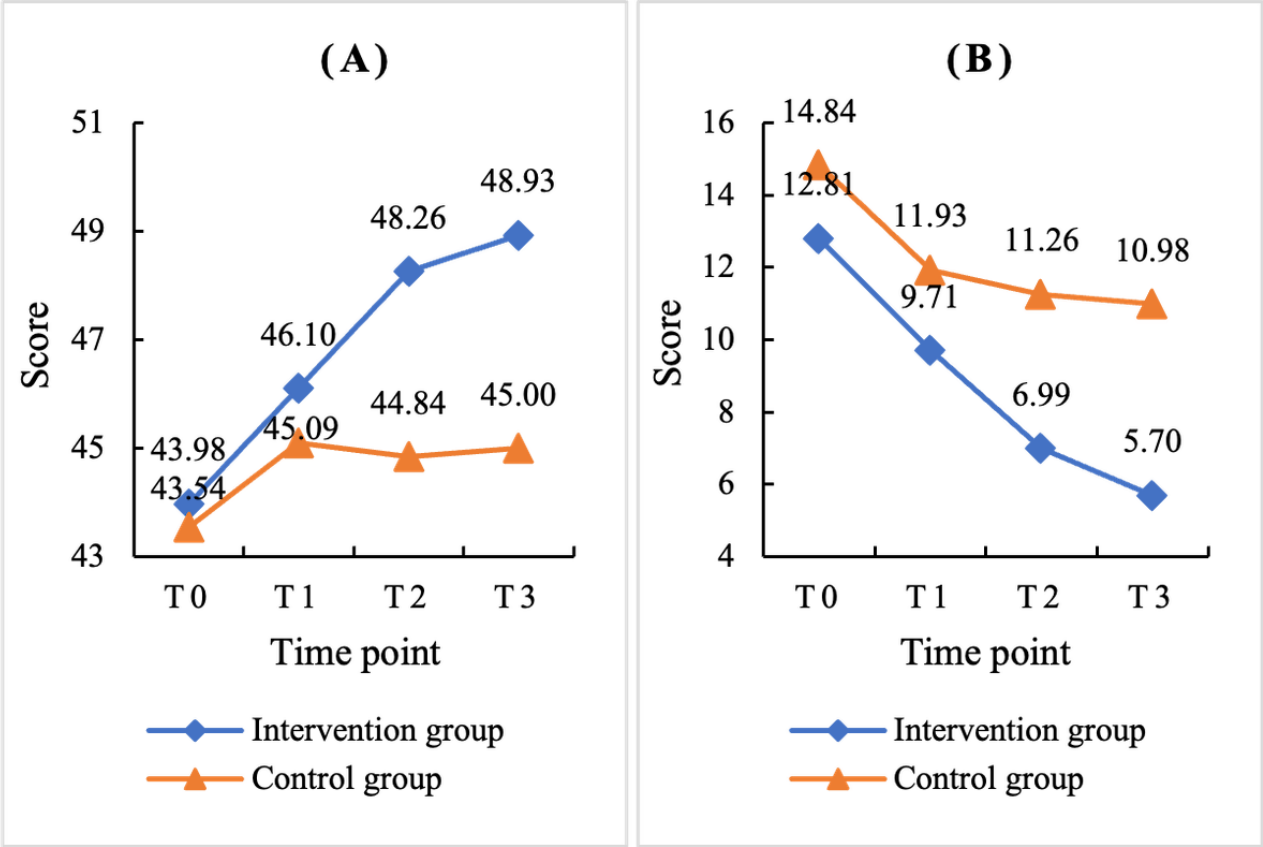

XSL•FO
RenderX

Table 4. Per-protocol analysis of the improvements in mindfulness and depression after online mindfulness training.

Time period	ISSMT[a] group (n=81), mean (SD)	Wait-list control group (n=72), mean (SD)	*t* (*df*)	*P* value	Time effect		Group effect		Time × group interaction		Effect size (Cohen *d*)
					F (*df*)	*P* value	*F* (*df*)	*P* value	*F* (*df*)	*P* value	
FFMQ-C-SF[b]											
T_0[c]	44.18 (8.89)	43.54 (8.92)	0.43 (151)	.67	—[d]	—	—	—	—	—	—
T_1[e]	46.79 (8.58)	45.09 (9.18)	1.14 (144)	.25	—	—	—	—	—	—	—
T_2[f]	49.21 (8.32)	44.84 (9.25)	2.86 (130)	.005	—	—	—	—	—	—	—
T_3[g]	50.03 (8.59)	45.00 (9.75)	3.09 (125)	.002	29.31 (3, 419.57)	<.001	11.21 (1, 158.47)	.001	7.67 (3, 419.64)	<.001	0.55
BDI-II[h]											
T_0	12.93 (11.42)	14.84 (10.63)	–1.06 (151)	.29	—	—	—	—	—	—	—
T_1	9.25 (8.66)	11.93 (11.24)	–1.60 (144)	.11	—	—	—	—	—	—	—
T_2	6.30 (7.66)	11.26 (12.04)	–2.77 (98.76)	.007	—	—	—	—	—	—	—
T_3	5.54 (7.15)	10.98 (12.78)	–2.94 (94.80)	.004	38.58 (3, 420.91)	<.001	10.11 (1, 157.77)	.002	3.65 (3, 421.03)	.01	0.53

[a]ISSMT: internet-based, self-guided, short-term mindfulness training.

[b]FFMQ-C-SF: short form of the Chinese version of the 5-facet mindfulness questionnaire.

[c]T_0: baseline.

[d]Not applicable.

[e]T_1: 1st week.

[f]T_2: 2nd week.

[g]T_3: 3rd week.

[h]BDI-II: Beck Depression Inventory-II.

XSL•FO
RenderX

Figure 3. Per-protocol analysis of the comparison of (A) mindfulness, as measured using the short form of the Chinese version of the 5-facet mindfulness questionnaire (FFMQ-C-SF), and (B) depression, as measured using the Beck Depression Inventory-II (BDI-II), between the training and wait-list control groups at each time point.

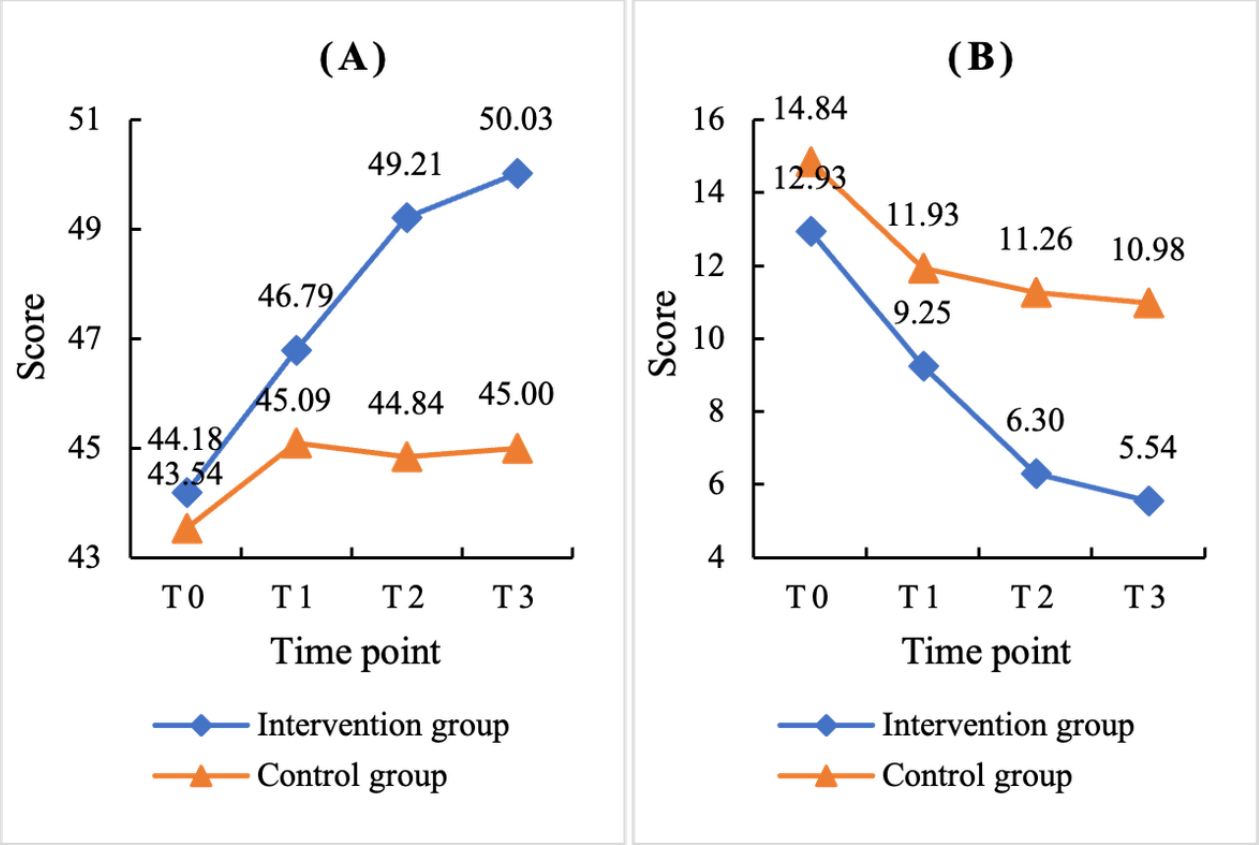

Training Effects on Depressive Symptoms

According to the ITT analysis, the scores for each outcome variable at each measurement time point are presented in Table 3. At baseline, there was no significant difference in depression scores between the training group and wait-list control group. The results of the mixed linear model revealed significant effects of time ($F_{3,514.77}$=38.48; P<.001), group ($F_{1,207.10}$=9.31; P=.003), and the time × group interaction ($F_{3,514.87}$=3.87; P=.009). These findings indicate that ISSMT can significantly reduce depressive symptoms, with an effect size of 0.50 for improving depression.

As shown in Table 3, depressive symptoms in the training group improved significantly from the first week and continued to improve significantly thereafter, whereas no significant changes were observed in the wait-list control group. The trends in the depression changes at each time point are illustrated in Figure 2B.

According to the PP analysis, the scores for each outcome variable at each time point are presented in Table 4. At baseline, there were no significant differences in depression scores between the training group and wait-list control group. The results of the mixed linear model revealed significant effects of time ($F_{3,420.91}$=38.58; P<.001), group ($F_{1,157.77}$=10.11; P=.002), and the time × group interaction ($F_{3,421.03}$=3.65; P=.01). These findings indicate that online mindfulness training can significantly reduce depressive symptoms. The effect size for improving depression was 0.53.

As shown in Table 4, the depression score in the training group improved significantly from the first week and continued to improve significantly thereafter, while no significant changes were observed in the wait-list control group. The trends in the depressive symptom changes at each time point are shown in Figure 3B.

Feasibility Analysis

The average Intervention Satisfaction Scale score for item 1 (“Overall, I liked the two-week course practice”) was 3.77, and the average score for item 2 (“I would recommend this course to others”) was 3.52, exceeding the average. At the same time, 87% (77/89) of the participants said they thought the course was helpful to them, and 72% (64/89) of the participants said they liked this online training method.

Discussion

Principal Findings

Depression is a leading global public health problem, with enormous demand for mental health services but limited access to mental health services. The emergence of telehealth solutions offers a promising approach to bridging this service gap. However, despite the potential, there are currently few studies in China that have focused on short-term online mindfulness training, with the majority of online studies adhering to the traditional 8-week mindfulness training framework [53]. In response, we developed a 2-week ISSMT course based on the MAT model. Additionally, we created an online platform for

XSL•FO
RenderX

program delivery. The results from our RCT indicate that, compared with the wait-list control group, participants in the training group had increased mindfulness levels and decreased depressive symptoms. Furthermore, participants expressed high acceptance of this training format. This study provides empirical evidence supporting the feasibility and effectiveness of ISSMT for Chinese adults. We recommend that future research explore the integration of this type of training into psychosocial service systems to enhance mental health.

Although the effectiveness of traditional 8-week mindfulness training has been reported in some studies, concerns remain regarding the extended training duration. Scholars have noted that longer training periods may lead to higher dropout rates and reduced participant motivation, thereby affecting treatment adherence [32]. Adherence is a critical determinant of treatment outcome, with approximately 50% of treatment failures attributable to nonadherence [54]. In a study delivering online psychosocial interventions, high dropout rates have been observed, reaching up to 73.9% among research participants and as high as 99.2% among public registrants [42]. This study introduces an ISSMT program that can be delivered online and demonstrated a notably lower dropout rate (<40%) as well as effective implementation. We believe the adapted training program, customized delivery platform, and quality control measures used in this study can serve as a valuable reference for further practices in delivering online mindfulness training and other psychosocial interventions in China.

Although previous studies have reported inconsistent results for short-term online mindfulness training [34-36], our study found that ISSMT significantly reduced depressive symptoms in the general population. Specifically, no significant differences in depressive symptoms were observed between the 2 groups at baseline or at week 1. However, by weeks 2 and 3, the severity of depressive symptoms in the training group was significantly lower than that in the control group. Although previous research suggests that brief mindfulness meditation sessions can improve immediate emotional states (eg, a single 10-minute mindfulness meditation training can reduce the intensity of negative affect in response to emotional stimuli [55]), our findings indicated that certain “doses” of training may be necessary to ensure the effectiveness of internet-delivered training. Furthermore, the emphasis on acceptance in MAT may be a key factor for sustaining the intervention’s effects [25].

There are several strengths of the ISSMT program delivered in the study. First, the effect sizes for the improvement in mindfulness and reduction in depressive symptoms exceeded those reported in other similar studies [56], with these benefits maintained at follow-up. This suggests that some participants may have continued practicing mindfulness mediation in their daily life after the trial concluded, further supporting the feasibility and long-term potential of the ISSMT program for the Chinese population.

Moreover, the dropout rate in this study was lower than in other similar online training programs. High dropout rates are a well-documented challenge for internet-based psychological training, with reported rates ranging from 73.9% to 99.2% in other studies [42,43,57,58]. Another review reported that an additional 0% to 18% of participants were lost between the posttraining assessment and follow-up assessment [59]. In contrast, our study achieved a completion rate of 60.9%, reflecting a considerably lower dropout rate and confirming the acceptability of this training format.

China’s large internet user base underscores the considerable potential of digital health interventions on a global scale, particularly within significant internet economies like China. According to the 54th “Statistical Report on the Development of China’s Internet” released by the China Internet Network Information Center, as of August 2024, the number of internet users in China was 1.099 billion, with an internet penetration rate of 78% [60].

The convenience and accessibility of online training programs make them an important supplement to offline training, helping to bridge the gap between limited medical resources and the vast mental health service needs in China [61]. Despite a more than twofold increase in the national ratio of psychiatrists, including accredited and assistant psychiatrists, per 100,000 population from 2002 to 2020 (from 1.51 to 3.19) and for accredited psychiatrists specifically (from 1.20 to 2.89) [62], there remains a significant gap between the demand for mental health services and the available resources. This disparity is exacerbated by the uneven distribution of mental health professionals, with a concentration in major cities and the northeast and east regions of China [62]. In light of these challenges, this study developed and validated an online mindfulness training program that effectively overcomes the geographical limitations of traditional face-to-face interventions, underscoring its significance for areas with constrained medical resources.

From a policy perspective, the empirical evidence presented in this study offers a valuable benchmark for authorities and stakeholders when shaping mental health policies. Our results demonstrate that digital mental health interventions not only deliver effective therapeutic outcomes but also garner high user satisfaction. Based on our knowledge, China has begun pilot initiatives to incorporate psychological therapies into public health insurance in certain regions, including Guangdong, Jiangsu, and Beijing. Although traditional psychotherapy is known for its extended treatment periods and challenges in efficacy monitoring, this study shows that a 2-week intervention can yield significant benefits. This finding supports the proposition for policymakers to integrate comparable online intervention models into national or regional mental health frameworks. Integrating such models would enhance the provision of accessible and efficient mental health services to the population.

It is noteworthy that this study did not account for the effects of group-specific differences, leaving uncertainties whether the findings extend to high-risk populations for depression, such as older adults and pregnant women. Older adults may face unique physiological and psychological challenges, including cognitive decline and difficulties with technology use [63], that may influence their acceptance of and response to online interventions. For pregnant women, prior research has shown

XSL•FO
RenderX

that those with high levels of neuroticism demonstrate greater improvements in maladaptive cognitive emotion regulation strategies following mindfulness interventions compared with those with moderate or low neuroticism [64]. This suggests variability in response to mindfulness interventions based on neuroticism levels. Therefore, caution is advised when generalizing these findings to these specific populations.

Limitations

A couple of limitations should be mentioned for this study. First, the main intervention effects were based primarily on self-report data. Further research may combine a broader range of measures, such as behavioral and biological measures. Second, a wait-list control design was used, and using an active control intervention in future studies would provide stronger support for the specific effects of the program. Additionally, this study did not segment participants by specific demographic categories. Certain groups, such as older adults and pregnant women, are recognized as higher-risk populations for depression, and the absence of targeted analysis for these groups may limit the generalizability of our findings. Future research should consider implementing strategies to increase the inclusion of high-risk groups, ensuring a broader and more diverse sample. For instance, offline recruitment methods or targeted online advertisements could be used to reach these specific populations. These enhancements would strengthen the representativeness of the sample and improve the external validity of the findings.

Conclusions

The 2-week ISSMT course based on the MAT demonstrated high feasibility and effectiveness for enhancing mindfulness levels and alleviating depressive symptoms. Given China's substantial potential user base and the disparity between limited mental health resources and the significant demand for services, this digital health intervention could play a pivotal role in expanding access to mental health support. Future research should prioritize the development of personalized online mindfulness programs that consider individual differences such as age, gender, and daily emotional fluctuations. Integrating artificial intelligence [65], wearable technology, and diverse health data may further refine the precision and efficacy of these interventions. Additionally, exploring virtual reality [66] and natural language processing could enable more immersive and engaging user experiences, potentially enhancing therapeutic outcomes.

Acknowledgments

We thank all the participants and staff of this study. No generative artificial intelligence was used in the writing of this manuscript.

This work was supported by the Research Office for Reform and Development at Shanghai Jiao Tong University (number JCZXZGC2024-08), the Center for Innovation and Development of Ideological and Political Work in Colleges and Universities at Shanghai Jiao Tong University (number DFY-SJ-2023069), and the Shanghai Jiao Tong University Teaching Development Fund Project (number CTLD2024J0002).

Data Availability

The data sets analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

TZ and SC were responsible for conceptualization of the study. TZ and LZ contributed to the methodology and formal analysis. The investigation was conducted by TZ, LZ, WW, RG, YW, and YZ. TZ wrote the original manuscript draft. HY was responsible for reviewing and editing the manuscript. JX contributed to reviewing and editing the paper. SC provided supervision, while TZ and LZ handled project administration. All authors reviewed, edited, and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of the course platform interface: (A) leader and (B) practitioner.
[PNG File , 193 KB - jmir_v27i1e55583_app1.png]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 311 KB - jmir_v27i1e55583_app2.pdf]

References

1. Herrman H, Kieling C, McGorry P, Horton R, Sargent J, Patel V. Reducing the global burden of depression: a Lancet-World Psychiatric Association Commission. Lancet 2019 Jun 15;393(10189):e42-e43. [doi: 10.1016/S0140-6736(18)32408-5] [Medline: 30482607]

XSL•FO
RenderX

2. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2025-01-07]
3. GBD Results. Institute for Health Metrics and Evaluation. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2025-01-07]
4. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry 2019 Mar;6(3):211-224. [doi: 10.1016/S2215-0366(18)30511-X] [Medline: 30792114]
5. China Statistical Yearbook 2022. National Bureau of Statistics China. 2022. URL: https://www.stats.gov.cn/sj/ndsj/2022/indexeh.htm [accessed 2025-01-07]
6. Lu J, Xu X, Huang Y, Li T, Ma C, Xu G, et al. Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. Lancet Psychiatry 2021 Nov;8(11):981-990. [doi: 10.1016/S2215-0366(21)00251-0] [Medline: 34559991]
7. Wahbeh H, Svalina MN, Oken BS. Group, one-on-one, or internet? Preferences for mindfulness meditation delivery format and their predictors. Open Med J 2014;1:66-74 [FREE Full text] [doi: 10.2174/1874220301401010066] [Medline: 27057260]
8. Ryan ML, Shochet IM, Stallman HM. Universal online interventions might engage psychologically distressed university students who are unlikely to seek formal help. Advances in Mental Health 2014 Dec 17;9(1):73-83. [doi: 10.5172/jamh.9.1.73]
9. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
10. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
11. Foroushani PS, Schneider J, Assareh N. Meta-review of the effectiveness of computerised CBT in treating depression. BMC Psychiatry 2011 Aug 12;11:131 [FREE Full text] [doi: 10.1186/1471-244X-11-131] [Medline: 21838902]
12. Newby JM, Twomey C, Yuan Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: A systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
13. Christ C, Schouten MJ, Blankers M, van Schaik DJ, Beekman AT, Wisman MA, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 25;22(9):e17831 [FREE Full text] [doi: 10.2196/17831] [Medline: 32673212]
14. Reger MA, Gahm GA. A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. J Clin Psychol 2009 Jan;65(1):53-75. [doi: 10.1002/jclp.20536] [Medline: 19051274]
15. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
16. Osborne EL, Ainsworth B, Hooper N, Atkinson MJ. Experiences of using digital mindfulness-based interventions: rapid scoping review and thematic synthesis. J Med Internet Res 2023 Sep 28;25:e44220 [FREE Full text] [doi: 10.2196/44220] [Medline: 37768709]
17. O'Brien-Venus B, Ellett L, Burgess-Barr S, Chadwick P. Systematic review of the safety of mindfulness-based interventions for psychosis. Clin Psychol Rev 2024 Aug;112:102445. [doi: 10.1016/j.cpr.2024.102445] [Medline: 38851179]
18. Yue WL, Ng KK, Koh AJ, Perini F, Doshi K, Zhou JH, et al. Mindfulness-based therapy improves brain functional network reconfiguration efficiency. Transl Psychiatry 2023 Nov 11;13(1):345 [FREE Full text] [doi: 10.1038/s41398-023-02642-9] [Medline: 37951943]
19. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, et al. Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. Clin Psychol Rev 2018 Feb;59:52-60 [FREE Full text] [doi: 10.1016/j.cpr.2017.10.011] [Medline: 29126747]
20. Goldberg SB, Riordan KM, Sun S, Davidson RJ. The empirical status of mindfulness-based interventions: a systematic review of 44 meta-analyses of randomized controlled trials. Perspect Psychol Sci 2022 Jan;17(1):108-130 [FREE Full text] [doi: 10.1177/1745691620968771] [Medline: 33593124]
21. Zhang D, Lee EKP, Mak ECW, Ho CY, Wong SYS. Mindfulness-based interventions: an overall review. Br Med Bull 2021 Jun 10;138(1):41-57 [FREE Full text] [doi: 10.1093/bmb/ldab005] [Medline: 33884400]
22. Sommers-Spijkerman M, Austin J, Bohlmeijer E, Pots W. New evidence in the booming field of online mindfulness: an updated meta-analysis of randomized controlled trials. JMIR Ment Health 2021 Jul 19;8(7):e28168 [FREE Full text] [doi: 10.2196/28168] [Medline: 34279240]
23. Kabat-Zinn J. Mindfulness-based stress reduction (MBSR). Constructivism in the Human Sciences 2003;8(2):73-107 [FREE Full text]
24. Teasdale JD, Segal ZV, Williams JM, Ridgeway VA, Soulsby JM, Lau MA. Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. J Consult Clin Psychol 2000 Aug;68(4):615-623. [doi: 10.1037//0022-006x.68.4.615] [Medline: 10965637]

XSL•FO
RenderX

25. Lindsay EK, Creswell JD. Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). Clin Psychol Rev 2017 Feb;51:48-59 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.011] [Medline: 27835764]
26. Garland EL, Farb NA, Goldin P, Fredrickson BL. Mindfulness broadens awareness and builds eudaimonic meaning: a process model of mindful positive emotion regulation. Psychol Inq 2015 Oct 01;26(4):293-314 [FREE Full text] [doi: 10.1080/1047840X.2015.1064294] [Medline: 27087765]
27. McConnell PA, Froeliger B. Mindfulness, mechanisms and meaning: perspectives from the cognitive neuroscience of addiction. Psychol Inq 2015;26(4):349-357 [FREE Full text] [doi: 10.1080/1047840X.2015.1076701] [Medline: 26924915]
28. Vago DR, Silbersweig DA. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. Front Hum Neurosci 2012;6:296 [FREE Full text] [doi: 10.3389/fnhum.2012.00296] [Medline: 23112770]
29. Compen F, Bisseling E, Schellekens M, Donders R, Carlson L, van der Lee M, et al. Face-to-face and internet-based mindfulness-based cognitive therapy compared with treatment as usual in reducing psychological distress in patients with cancer: a multicenter randomized controlled trial. JCO 2018 Aug 10;36(23):2413-2421. [doi: 10.1200/jco.2017.76.5669]
30. Hearn JH, Cotter I, Finlay KA. Efficacy of internet-delivered mindfulness for improving depression in caregivers of people with spinal cord injuries and chronic neuropathic pain: a randomized controlled feasibility trial. Arch Phys Med Rehabil 2019 Jan;100(1):17-25. [doi: 10.1016/j.apmr.2018.08.182] [Medline: 30268806]
31. Jie G, Tong H, Sun X. Effect of Chinese online mindfulness-based stress reduction course on mindfulness level and mood state of general population. Chinese General Practice 2019;22(5):581-585 [FREE Full text] [doi: 10.12114/j.issn.1007-9572.2018.00.195]
32. Fortney L, Luchterhand C, Zakletskaia L, Zgierska A, Rakel D. Abbreviated mindfulness intervention for job satisfaction, quality of life, and compassion in primary care clinicians: a pilot study. Ann Fam Med 2013;11(5):412-420 [FREE Full text] [doi: 10.1370/afm.1511] [Medline: 24019272]
33. Kemper KJ. Brief online mindfulness training: immediate impact. J Evid Based Complementary Altern Med 2017 Jan;22(1):75-80 [FREE Full text] [doi: 10.1177/2156587216639199] [Medline: 27002136]
34. Krusche A, Cyhlarova E, Williams JMG. Mindfulness online: an evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. BMJ Open 2013 Nov 29;3(11):e003498 [FREE Full text] [doi: 10.1136/bmjopen-2013-003498] [Medline: 24293203]
35. Cavanagh K, Strauss C, Cicconi F, Griffiths N, Wyper A, Jones F. A randomised controlled trial of a brief online mindfulness-based intervention. Behav Res Ther 2013 Sep;51(9):573-578. [doi: 10.1016/j.brat.2013.06.003] [Medline: 23872699]
36. Glück TM, Maercker A. A randomized controlled pilot study of a brief web-based mindfulness training. BMC Psychiatry 2011 Nov 8;11(1):1. [doi: 10.1186/1471-244x-11-175]
37. Zeidan F, Johnson SK, Diamond BJ, David Z, Goolkasian P. Mindfulness meditation improves cognition: evidence of brief mental training. Conscious Cogn 2010 Jun;19(2):597-605. [doi: 10.1016/j.concog.2010.03.014] [Medline: 20363650]
38. Chen Y, Yang X, Wang L, Zhang X. A randomized controlled trial of the effects of brief mindfulness meditation on anxiety symptoms and systolic blood pressure in Chinese nursing students. Nurse Educ Today 2013 Oct;33(10):1166-1172. [doi: 10.1016/j.nedt.2012.11.014] [Medline: 23260618]
39. Howarth A, Smith JG, Perkins-Porras L, Ussher M. Effects of brief mindfulness-based interventions on health-related outcomes: a systematic review. Mindfulness 2019 May 31;10(10):1957-1968. [doi: 10.1007/s12671-019-01163-1]
40. Lindsay EK, Young S, Smyth JM, Brown KW, Creswell JD. Acceptance lowers stress reactivity: Dismantling mindfulness training in a randomized controlled trial. Psychoneuroendocrinology 2018 Jan;87:63-73. [doi: 10.1016/j.psyneuen.2017.09.015] [Medline: 29040891]
41. Lindsay EK, Chin B, Greco CM, Young S, Brown KW, Wright AGC, et al. How mindfulness training promotes positive emotions: dismantling acceptance skills training in two randomized controlled trials. J Pers Soc Psychol 2018 Dec;115(6):944-973 [FREE Full text] [doi: 10.1037/pspa0000134] [Medline: 30550321]
42. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
43. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
44. Zhu T, Chen C, Chen S. Validation of a Chinese version of the five facet mindfulness questionnaire and development of a short form based on item response theory. Curr Psychol 2021 Apr 27;42(5):4212-4224. [doi: 10.1007/s12144-021-01720-9]
45. Beck AT. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun 01;4(6):561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
46. Wang Z, Yuan CM, Huang J, Li ZZ, Chen J, Zhang HY, et al. Reliability and validity of the Chinese version of Beck Depression Inventory-II among depression patients. Chinese Mental Health Journal 2011;25(6):476-480 [FREE Full text]
47. Campo RA, Bluth K, Santacroce SJ, Knapik S, Tan J, Gold S, et al. A mindful self-compassion videoconference intervention for nationally recruited posttreatment young adult cancer survivors: feasibility, acceptability, and psychosocial outcomes. Support Care Cancer 2017 Jun;25(6):1759-1768. [doi: 10.1007/s00520-017-3586-y] [Medline: 28105523]

XSL•FO
RenderX

48. Tang W, He H, Tu XM. Applied Categorical and Count Data Analysis. New York, NY: Chapman and Hall/CRC; 2012.
49. Kowalski J, Tu XM. Modern Applied U‐Statistics. Hoboken, NJ: John Wiley & Sons Inc; 2008.
50. Lu N, Tang W, He H, Yu Q, Crits-Christoph P, Zhang H, et al. On the impact of parametric assumptions and robust alternatives for longitudinal data analysis. Biom J 2009 Aug;51(4):627-643 [FREE Full text] [doi: 10.1002/bimj.200800186] [Medline: 19688758]
51. Fisher L, Polonsky W, Parkin CG, Jelsovsky Z, Amstutz L, Wagner RS. The impact of blood glucose monitoring on depression and distress in insulin-naïve patients with type 2 diabetes. Curr Med Res Opin 2011 Nov;27 Suppl 3:39-46. [doi: 10.1185/03007995.2011.619176] [Medline: 21916532]
52. Tabachnick B, Fidell L. Using Multivariate Statistics, 7th edition. Boston, MA: Pearson; 2018.
53. Mak WWS, Chan ATY, Cheung EYL, Lin CLY, Ngai KCS. Enhancing web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]
54. Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. Am J Med 2012 Sep;125(9):882-7.e1. [doi: 10.1016/j.amjmed.2011.12.013] [Medline: 22748400]
55. Erisman SM, Roemer L. A preliminary investigation of the effects of experimentally induced mindfulness on emotional responding to film clips. Emotion 2010 Feb;10(1):72-82 [FREE Full text] [doi: 10.1037/a0017162] [Medline: 20141304]
56. Sverre KT, Nissen ER, Farver-Vestergaard I, Johannsen M, Zachariae R. Comparing the efficacy of mindfulness-based therapy and cognitive-behavioral therapy for depression in head-to-head randomized controlled trials: a systematic review and meta-analysis of equivalence. Clin Psychol Rev 2023 Mar;100:102234 [FREE Full text] [doi: 10.1016/j.cpr.2022.102234] [Medline: 36527794]
57. Christensen H, Griffiths KM, Mackinnon AJ, Brittliffe K. Online randomized controlled trial of brief and full cognitive behaviour therapy for depression. Psychol Med 2006 Dec;36(12):1737-1746. [doi: 10.1017/S0033291706008695] [Medline: 16938144]
58. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a web-based self-help cognitive behavioral therapy program for panic disorder. J Med Internet Res 2005 Mar 26;7(1):e7 [FREE Full text] [doi: 10.2196/jmir.7.1.e7] [Medline: 15829479]
59. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
60. CNNIC. The 54th Statistical Report on Internet Development in China. China Internet Network Information Center. 2024 Aug 29. URL: https://www.cnnic.net.cn/n4/2024/0829/c88-11065.html [accessed 2025-01-07]
61. Kazdin AE, Rabbitt SM. Novel models for delivering mental health services and reducing the burdens of mental illness. Clinical Psychological Science 2013 Jan 23;1(2):170-191. [doi: 10.1177/2167702612463566]
62. Sun M, Zhou H, Li Y, Wang J, Yang W, Gong Y, et al. Professional characteristics, numbers, distribution and training of China's mental health workforce from 2000 to 2020: a scoping review. Lancet Reg Health West Pac 2024 Apr;45:100992 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100992] [Medline: 38699293]
63. Long KM, Casey K, Bhar S, Al Mahmud A, Curran S, Hunter K, et al. Understanding perspectives of older adults on the role of technology in the wider context of their social relationships. Ageing and Society 2022 Sep 19;44(6):1453-1476. [doi: 10.1017/s0144686x2200085x]
64. Wang Z, Zhang Q, Zhou Z, Huang Z, Mao H, Zhang L, et al. Online mindfulness-based cognitive behavioral intervention for pregnant women: an efficacy evaluation. Maternal and Child Health Care of China 2023;38(19):3647-3651 [FREE Full text]
65. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 04;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
66. Pardini S, Gabrielli S, Olivetto S, Fusina F, Dianti M, Forti S, et al. Personalized virtual reality compared with guided imagery for enhancing the impact of progressive muscle relaxation training: pilot randomized controlled trial. JMIR Ment Health 2024 Jan 30;11:e48649 [FREE Full text] [doi: 10.2196/48649] [Medline: 38289673]

Abbreviations

BDI-II: Beck Depression Inventory-II
FFMQ-C-SF: short form of the Chinese version of the 5-facet mindfulness questionnaire
ISSMT: internet-based, self-guided, short-term mindfulness training
ITT: intention-to-treat
MAT: Monitor and Acceptance Theory
MBSR: Mindfulness-Based Stress Reduction
PP: per-protocol
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 17.12.23; peer-reviewed by Y Wu; comments to author 18.09.24; revised version received 12.11.24; accepted 13.12.24; published 13.02.25.

Please cite as:
Zhu T, Zhang L, Weng W, Gan R, Sun L, Wei Y, Zhu Y, Yu H, Xue J, Chen S
Effectiveness of an Internet-Based, Self-Guided, Short-Term Mindfulness Training (ISSMT) Program for Relieving Depressive Symptoms in the Adult Population in China: Single-Blind, Randomized Controlled Trial
J Med Internet Res 2025;27:e55583
URL: https://www.jmir.org/2025/1/e55583
doi:10.2196/55583
PMID:

©Tingfei Zhu, Liuyi Zhang, Wenqi Weng, Ruochen Gan, Limin Sun, Yanping Wei, Yueping Zhu, Hongyan Yu, Jiang Xue, Shulin Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Isolation and Dementia Risk in Older Adults: Longitudinal Cohort Study

Cheng Deng[1*], MD, PhD; Na Shen[2*], MD, PhD; Guangzhou Li[1], MD, PhD; Ke Zhang[3], MS, MPH; Shijun Yang[4], MD, PhD

[1]Department of Cardiovascular Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Breast and Thyroid Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Public Health, University of Utah, Salt Lake City, UT, United States

[4]Department of Cardiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Shijun Yang, MD, PhD
Department of Cardiology
Union Hospital, Tongji Medical College
Huazhong University of Science and Technology
No. 1277 Jiefang Avenue
Wuhan, 430022
China
Phone: 86 13971132795
Email: yangsj1218@163.com

Abstract

Background: Dementia poses a significant global health challenge, characterized by progressive cognitive decline and functional impairment. With the aging global population, dementia prevalence is projected to surge, reaching an estimated 153 million cases by 2050. While the impact of traditional social isolation on dementia risk has been extensively studied, the influence of digital isolation, a phenomenon unique to the digital age, remains underexplored.

Objective: This study aimed to investigate the association between digital isolation and dementia risk among older adults, hypothesizing that higher levels of digital isolation significantly increase the risk of developing dementia.

Methods: We conducted a longitudinal cohort study using data from the National Health and Aging Trends Study (NHATS), analyzing 8189 participants aged 65 years and older from the 3rd (2013) to the 12th wave (2022). Digital isolation was quantified using a composite digital isolation index, derived from participants' usage of digital devices, electronic communication, internet access, and engagement in online activities. Participants were stratified into low isolation and moderate to high isolation groups. Dementia incidence was assessed using cognitive tests and proxy reports. Cox proportional hazards models were used to estimate the association between digital isolation and dementia risk, adjusting for potential confounders including sociodemographic factors, baseline health conditions, and lifestyle variables.

Results: The moderate to high isolation group demonstrated a significantly elevated risk of dementia compared with the low isolation group. In the discovery cohort, the adjusted hazard ratio (HR) was 1.22 (95% CI 1.01-1.47, *P*=.04), while the validation cohort showed an HR of 1.62 (95% CI 1.27-2.08, *P*<.001). The pooled analysis across both cohorts revealed an adjusted HR of 1.36 (95% CI 1.16-1.59, *P*<.001). Kaplan-Meier curves corroborated a higher incidence of dementia in the moderate to high isolation group.

Conclusions: Our findings indicate that digital isolation is a significant risk factor for dementia among older adults. This study underscores the importance of digital engagement in mitigating dementia risk and suggests that promoting digital literacy and access to digital resources should be integral components of public health strategies aimed at dementia prevention.

(J Med Internet Res 2025;27:e65379) doi:10.2196/65379

KEYWORDS

dementia; digital isolation; cognitive decline; older adults; elderly; geriatric; longitudinal cohort study; cognitive impairment; aging; social isolation; risk; digital engagement

XSL•FO
RenderX

Introduction

Dementia, characterized by progressive cognitive decline and functional impairment, represents a formidable global health challenge [1,2]. As the world's population ages rapidly, the prevalence of dementia is projected to surge, with an estimated 153 million individuals affected by 2050 [3]. In the absence of curative treatments, the imperative for effective prevention strategies has never been more pressing. The etiology of dementia is multifactorial, encompassing both nonmodifiable factors such as age and genetic predisposition, and modifiable risk factors including cardiovascular health, lifestyle choices, and social engagement [4-6].

In the context of an increasingly digitalized society, interaction with technology has become integral to modern life. However, the digital revolution has not benefited all segments of the population equally. A significant proportion of older adults find themselves in a state of "digital isolation," either due to limited access or inadequate digital literacy [7-9]. This concept extends beyond traditional social isolation by emphasizing the absence of digital engagement, including the use of the internet, smartphones, or social media, which can offer additional cognitive and social stimulation [10,11]. Recent studies indicate that older adults' use of technology-based social platforms or electronic health resources may enhance cognitive outcomes, postpone cognitive decline, and alleviate loneliness. Consequently, individuals who are digitally isolated may miss these protective effects, which could accelerate cognitive decline and elevate their risk of dementia [11-13]. Limited literacy and education levels may further restrict older adults' ability to engage with digital technologies, hindering their ability to benefit and potentially increasing the risk of cognitive decline [14,15].

While the association between traditional social isolation and dementia risk has been extensively studied, the impact of digital isolation, a phenomenon unique to our technologically driven era, has received comparatively little attention [16-18]. Preliminary investigations suggest a potential link between digital isolation and accelerated cognitive decline, as well as increased dementia risk. However, these studies are often constrained by limited sample sizes and cross-sectional designs, precluding the establishment of causal relationships. Furthermore, many existing studies fail to adequately control for potential confounding factors such as depression, anxiety, chronic comorbidities, and lifestyle variables, potentially biasing their results [19,20].

This study aims to clarify the relationship between digital isolation and dementia risk using a large-scale, longitudinal cohort design. Using a multistage Cox proportional hazards model on discovery and validation cohorts, we control for numerous potential confounders, including sociodemographic characteristics (eg, education level), baseline health status, and lifestyle factors. This study aims to provide strong evidence supporting the hypothesis that higher levels of digital isolation increase dementia risk. It also seeks to identify mechanisms underlying this emerging risk factor and to inform novel dementia prevention strategies, particularly those enhancing digital literacy and technology access. We aim to support a holistic public health approach that integrates traditional and digital aspects of social engagement in aging populations.

Methods

Study Population

This investigation used data from the National Health and Aging Trends Study (NHATS), a nationally representative longitudinal survey of Medicare beneficiaries aged 65 years and older in the United States. To ensure representativeness of the older adult population, appropriate survey weights provided by NHATS were applied to account for the complex sampling design. Our analysis used data spanning from the 3rd wave (2013), when digital product usage assessment was initiated, to the 12th wave (2022). The study cohort was stratified into discovery and validation samples.

In the third wave (2013), 5799 participants remained in the study. We excluded individuals lacking baseline digital isolation data or with preexisting dementia diagnoses. We rigorously controlled potential confounders influencing the digital isolation-dementia risk relationship, including age, education level, gender, race or ethnicity, baseline diseases, depression, anxiety, smoking status, and sleep difficulties. Participants were followed from the fourth wave (2014) through the twelfth wave (2022). During follow-up, some individuals were excluded due to attrition or death before dementia diagnosis. The final analytical sample comprised 4455 individuals.

To validate our findings, we used an independent cohort of 4182 individuals newly recruited in the fifth wave (2015). After applying the same exclusion criteria as the discovery cohort and accounting for attrition and mortality during follow-up, 3734 individuals were included in the validation sample. This cohort was followed from the 5th wave (2015) through the 12th wave (2022). As shown in Figure 1, we provide a detailed overview of the inclusion and exclusion processes for the discovery and validation cohorts, including the study waves, exclusion criteria, and the final number of participants included in the analysis dataset.

XSL•FO
RenderX

Figure 1. Flowchart depicting participant inclusion and exclusion criteria for the discovery (wave 3) and validation (wave 5) cohorts, with subsequent follow-up through wave 12 and formation of the pooled sample. NHATS: National Health and Aging Trends Study.

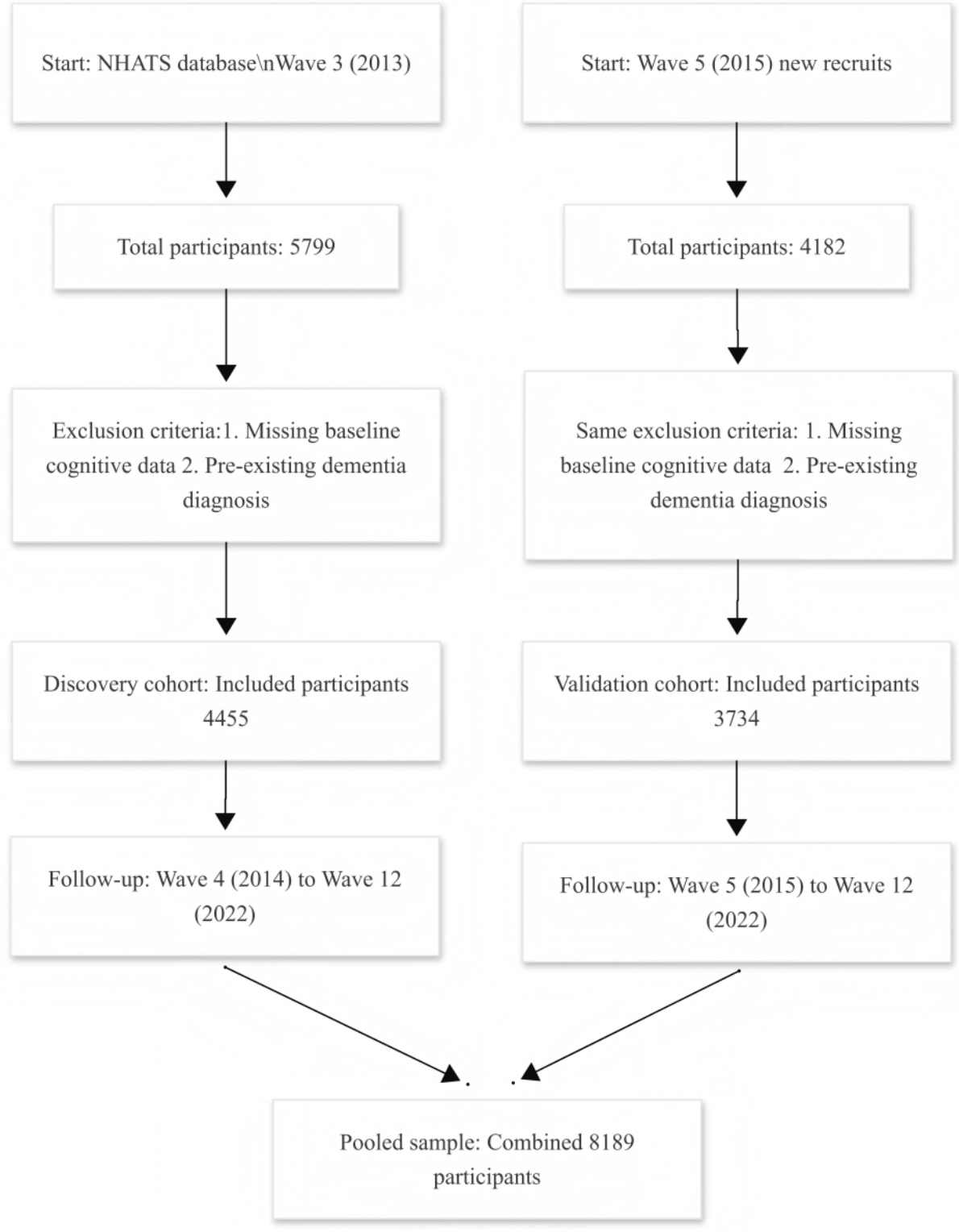

Digital Isolation

Digital isolation was assessed using a composite digital isolation index, which was constructed based on individuals' usage of various digital devices and the internet. The design of this index was informed by relevant literature in the fields of social isolation and digital health. For instance, Cornwell and Waite [21] quantified social isolation by evaluating the extent of individuals' social contact and participation and proposed a method for constructing a social isolation index using self-reported data. Similarly, digital isolation, as a modern form of isolation, is primarily reflected in the insufficient engagement with and use of digital technologies.

In this study, we operationalized digital isolation through a composite digital isolation index comprising 7 parameters (Textbox 1).

XSL•FO
RenderX

Textbox 1. Components of the Digital Isolation Index.

Parameters:

- Mobile phone use.
- Computer usage.
- Tablet use.
- Frequency of electronic communication (email or text messaging).
- Internet access.
- Engagement in online activities.
- Participation in health-related digital platforms.

The selection and quantification of these indicators were informed by extant literature on digital technology adoption, such as Kraut et al [22], who investigated the impact of internet use on social involvement. Each parameter was dichotomized (0=nonuse, 1=use), and the sum of these binary scores constituted the aggregate digital isolation index [22]. For stratification purposes, we adopted methodologies analogous to those used in social frailty research by Makizako et al [23] and Wei et al [24]. Participants were categorized into 2 cohorts based on their digital isolation index: those scoring 2 or less were classified as "low isolation," while those scoring 3 or above were designated as "moderate to high isolation." This bifurcation strategy was designed to elucidate the potential differential impacts of varying degrees of digital isolation on health outcomes.

Dementia

In the NHATS database, dementia ascertainment is primarily predicated on cognitive function assessments and self-reports or proxy reports. NHATS uses a multifaceted approach to dementia evaluation, encompassing cognitive testing, proxy reports, and clinical records. Specifically, NHATS uses a battery of cognitive tests to assess participants' memory, attention, and executive function, with these metrics serving as indicators of cognitive status. Concurrently, NHATS collects proxy reports, typically from family members or caregivers, regarding the participants' cognitive condition. These reports may encompass physician-diagnosed dementia and observe cognitive deficits in activities of daily living. Investigators typically synthesize these data with additional clinical information to determine dementia status and monitor its progression longitudinally. Upon confirmation or report of dementia in any follow-up wave, subsequent inquiries regarding dementia status are discontinued for that participant.

Covariates

This study incorporated a comprehensive set of covariates to ensure a precise estimation of the association between digital isolation and dementia risk. These covariates encompass sociodemographic characteristics, clinical parameters, and health-related behaviors.

Sociodemographic variables include education level, age, gender, and race or ethnicity. Education level was categorized as <high school, high school or general educational development (GED), some college, and college or above, reflecting participants' highest educational attainment. Age was stratified into 6 cohorts: 65-69 years, 70-74 years, 75-79 years, 80-84 years, 85-89 years, and ≥90 years. Sex was dichotomized as male or female. Race or ethnicity was categorized as non-Hispanic White, non-Hispanic Black, Hispanic, and other. These variables are routinely used as baseline covariates to adjust for sociodemographic heterogeneity in health outcomes [25-27].

Clinical parameters comprise the number of baseline diseases, depressive symptomatology, and anxiety manifestations. The number of baseline diseases was trichotomized based on self-reported chronic conditions (including arthritis, cardiovascular disease, hypertension, diabetes, pulmonary disease, cerebrovascular accident, osteoporosis, and malignancy): no diseases, 1-2 diseases, and ≥3 diseases. Depressive and anxiety symptoms were assessed by validated self-report instruments and operationalized as binary variables in the analyses [28,29].

Health-related behaviors include smoking status and sleep difficulties. Smoking status was dichotomized as current smokers and noncurrent smokers to account for potential cognitive effects of tobacco use [30]. Sleep difficulties were stratified based on the frequency of difficulty falling asleep within 30 minutes: high sleep difficulty (every night or most nights), moderate sleep difficulty (some nights), and low or no sleep difficulty (rarely or never) [31]. Previous research has demonstrated associations between poor sleep quality, cognitive decline, and elevated dementia risk, warranting its inclusion as a key covariate.

Statistical Analysis

To address the complex sampling design of NHATS, we applied person-level sampling weights for longitudinal analyses spanning 2013 to 2022. Clustering and stratification were incorporated by specifying the primary sampling units (PSUs) and strata in the design-based analyses. Baseline characteristics of the study population were summarized using descriptive statistics, with categorical variables reported as frequencies and percentages. Appropriate NHATS survey weights were applied across all analyses to ensure the findings represent the broader US older adult population. To assess the probability of dementia-free survival across different digital isolation groups (low isolation vs moderate to high isolation), we used Kaplan-Meier survival analysis. The resultant Kaplan-Meier survival curves are depicted in Figure 2. To quantify the

XSL•FO
RenderX

association between digital isolation and dementia risk, we used Cox proportional hazards regression models, calculating hazard ratios (HRs) with corresponding 95% CIs. A total of 2 models were constructed: model 1 (unadjusted) and model 2 (adjusted for potential confounders including education level, age, gender, race, number of baseline diseases, depression, anxiety, smoking status, and sleep difficulties).

We further investigated the individual components of digital isolation (eg, mobile phone use, computer use, tablet use, frequency of email usage, internet usage, online activities, and health-related online activities) and their respective impacts on dementia risk. Each component was analyzed independently using Cox proportional hazards models to estimate HRs and 95% CIs.

To examine the potential heterogeneity in the association between digital isolation and dementia, we conducted stratified analyses across subgroups defined by gender, age, race or ethnicity, and comorbidity status. Effect modification by these factors was assessed through the inclusion of interaction terms in the Cox models. In addition, a sensitivity analysis was performed to ensure the robustness of our findings. We recalculated the primary models under several alternative conditions, including adjusting the threshold for defining digital isolation, excluding participants with missing key variables, and omitting early dementia cases to reduce potential reverse causality. To enhance the robustness of our findings, we combined the results from the discovery and validation samples using meta-analytic techniques. Pooled HRs and CIs were computed using a fixed-effects model, under the assumption of homogeneity across samples. Statistical analyses were performed using R statistical software (version 4.4.1; R Foundation for Statistical Computing). A 2-sided *P* value <.05 was considered statistically significant for all analyses.

Figure 2. Kaplan Meier curves of the relationship between digital isolation and dementia.

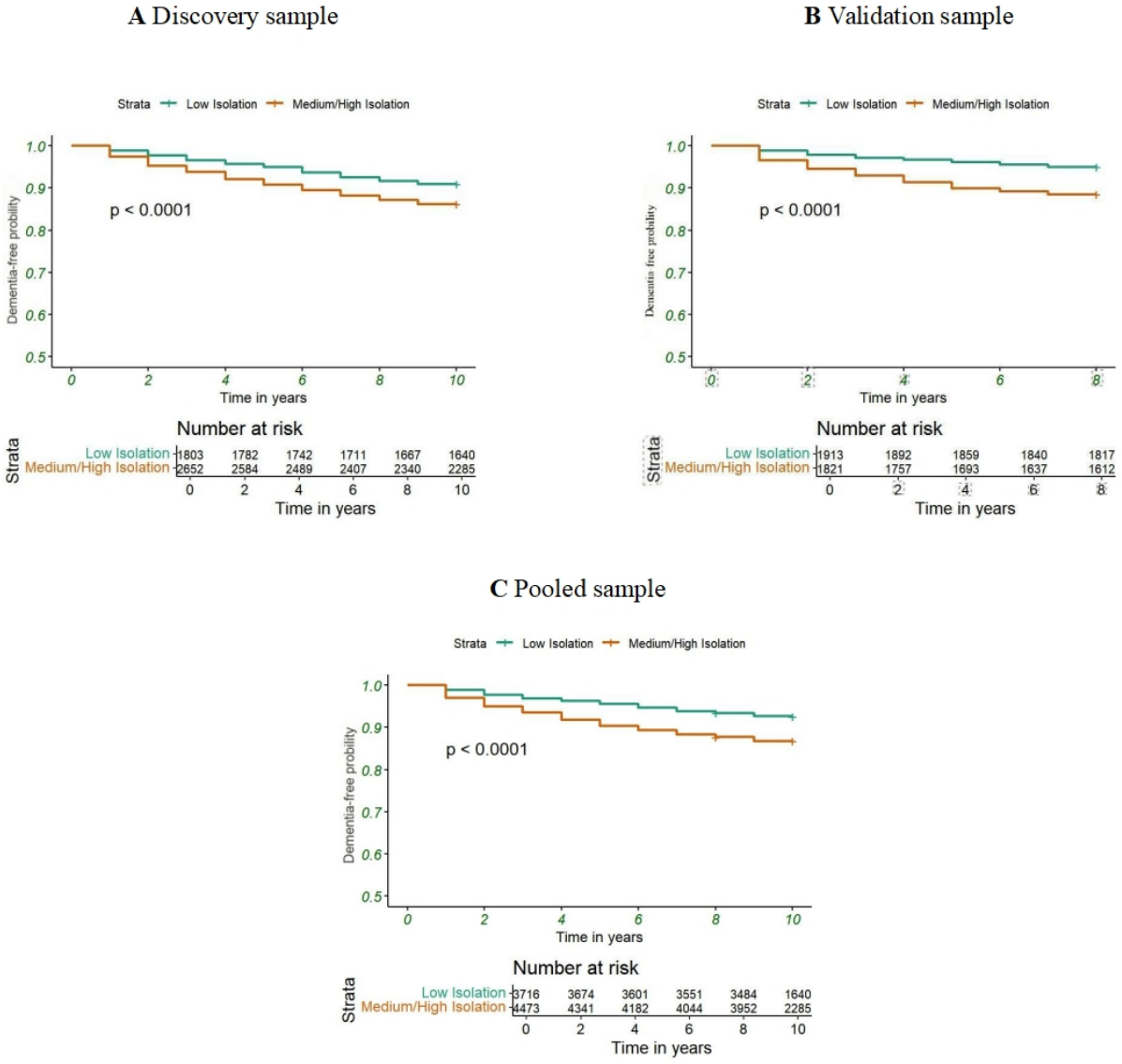

XSL•FO
RenderX

Ethical Considerations

The study protocol was approved by the institutional review board of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. All methods were conducted in accordance with relevant guidelines and regulations. The NHATS dataset is deidentified and publicly available, with informed consent obtained from participants at enrollment. No additional individual-level consent was necessary for this secondary analysis.

Results

Baseline Characteristics of the Study Population

The study cohort comprised 8189 participants, with 4455 individuals in the discovery sample and 3734 in the validation sample. Survey weights provided by NHATS were applied in all analyses to ensure representativeness of the older adult population in the United States. As shown in Table 1, Females constituted a slight majority, accounting for 4724 out of 8189 (57.68%) of the cohort. The racial or ethnic composition was predominantly non-Hispanic White (5615/8189, 68.58%), followed by non-Hispanic Black (1677/8189, 20.48%) and Hispanic (469/8189, 5.73%). Regarding education level, approximately 1250 out of 8189 (15.27%) of participants had less than a high school education, 2560/8189 (31.28%) completed high school or GED, 2375 out of 8189 (29.00%) had some college education, and 2004 out of 8189 (24.45%) had a college degree or higher. Age distribution analysis revealed that the “70-74 years” age group was most prevalent (1906/8189, 23.27%), succeeded by the “75-79 years” (1689/8189, 20.63%) and “80-84 years” (1479/8189, 18.06%) age groups. With respect to baseline diseases, 3918 out of 8189 (47.85%) of participants reported no chronic conditions, while 2593 out of 8189 (31.67%) had 1-2 diseases, and 1678 out of 8189 (20.49%) had 3 or more chronic conditions. The prevalence of depression and anxiety was 2189 out of 8189 (26.74%) and 2819 out of 8189 (34.43%), respectively. The majority of participants (7582/8189, 92.58%) were nonsmokers. Sleep difficulties were reported as significant by 1707 out of 8189 (20.84%) of participants, and moderate by 1989 out of 8189 (24.29%). In the context of digital isolation, 4473 out of 8189 (54.62%) of participants were categorized as experiencing moderate to high isolation, while 3716 out of 8189 (45.38%) were classified as having low isolation. Digital device use rates were as follows: mobile phones (6615/8189, 80.8%), computers (4857/8189, 59.31%), and tablets (2045/8189, 24.98%). Frequent engagement in digital activities was reported for email or text messaging 6537/8189, (79.83%), internet access (4709/8189, 57.51%), general online activities (7282/8189, 88.93%), and health-related online activities (6315/8189, 77.13%).

XSL•FO
RenderX

Table 1. Baseline characteristics of the study population.

Variables	Discovery sample (N=4455)	Validation sample (N=3734)	Pooled sample (N=8189)
Gender, n (%)			
Male	1843 (41.37)	1622 (43.44)	3465 (42.32)
Female	2612 (58.63)	2112 (56.56)	4724 (57.68)
Race or ethnicity[a], n (%)			
White, non-Hispanic	3162 (70.98)	2453 (65.69)	5615 (68.58)
Black, non-hispanic	915 (20.54)	762 (20.41)	1677 (20.48)
Hispanic	230 (5.16)	239 (6.4)	469 (5.73)
Other	148 (3.32)	280 (7.5)	428 (5.22)
Education level, n (%):			
< High school	680 (15.26)	570 (15.27)	1250 (15.27)
High school or GED	1380 (31)	1180 (31.6)	2560 (31.28)
Some college	1370 (30.77)	1005 (26.91)	2375 (29)
College or above	1025 (23)	979 (26.22)	2004 (24.45)
Age (years), n (%)			
65 to 69	459 (10.3)	1000 (26.78)	1459 (17.82)
70 to 74	1076 (24.15)	830 (22.23)	1906 (23.27)
75 to 79	961 (21.57)	728 (19.5)	1689 (20.63)
80 to 84	899 (20.18)	580 (15.53)	1479 (18.06)
85 to 89	647 (14.52)	355 (9.51)	1002 (12.23)
90 or above	413 (9.27)	241 (6.45)	654 (7.99)
Baseline disease, n (%)			
No disease	3568 (80.09)	350 (9.37)	3918 (47.85)
1-2 diseases	861 (19.33)	1732 (46.38)	2593 (31.67)
3 or more diseases	26 (0.58)	1652 (44.24)	1678 (20.49)
Depression[b], n (%)			
No depression	3241 (72.75)	2759 (73.89)	6000 (73.26)
Depression	1214 (27.25)	975 (26.11)	2189 (26.74)
Anxiety, n (%)			
No anxiety	2908 (65.27)	2462 (65.93)	5370 (65.57)
Anxiety	1547 (34.73)	1272 (34.07)	2819 (34.43)
Smoking status, n (%)			
Nonsmoker	4152 (93.2)	3430 (91.86)	7582 (92.58)
Smoker	303 (6.8)	304 (8.14)	607 (7.42)
Sleep difficulty[c], n (%)			
High difficulty	848 (19.13)	859 (23.07)	1707 (20.84)
Medium difficulty	1115 (25.15)	874 (23.47)	1989 (24.29)
Low or no difficulty	2470 (55.72)	1991 (53.46)	4461 (54.47)
Digital isolation[d] group, n (%)			
Low isolation	1803 (40.47)	1913 (51.23)	3716 (45.38)
Medium and high isolation	2652 (59.53)	1821 (48.77)	4473 (54.62)
Items of digital isolation, n (%)			

XSL•FO
RenderX

Variables	Discovery sample (N=4455)	Validation sample (N=3734)	Pooled sample (N=8189)
Phone use, n (%)	3446 (77.35)	3169 (84.87)	6615 (80.8)
Computer use, n (%)	2531 (56.81)	2326 (62.29)	4857 (59.31)
Tablet use, n (%)	764 (17.15)	1281 (34.31)	2045 (24.98)
Email frequency, n (%)	3603 (80.88)	2934 (78.58)	6537 (79.83)
Internet use, n (%)	2498 (56.07)	2211 (59.21)	4709 (57.51)
Online activity, n (%)	3945 (88.55)	3337 (89.37)	7282 (88.93)
Health-related online use, n (%)	3484 (78.2)	2831 (75.82)	6315 (77.13)

[a]Race or ethnicity was reclassified into 4 categories: White, non-Hispanic; Black, non-Hispanic; Hispanic; and Other.

[b]The depression, anxiety, and smoking status variables were reclassified as binary variables. Responses of “Don't know,” “Refused to answer,” and similar were treated as missing data.

[c]Sleep difficulty was categorized by the frequency of difficulty falling asleep within 30 minutes, with unknown or refused responses treated as missing data.

[d]The Digital Isolation Index was used to create a new variable (Digital Isolation Group) dividing participants into low and moderate and high isolation groups.

Association Between Digital Isolation and the Risk of Dementia

The association between digital isolation and dementia risk was comprehensively examined, adjusting for factors including education level, age, gender, race or ethnicity, number of baseline diseases, depression, anxiety, smoking status, and sleep difficulties, with results presented in Table 2. In the discovery sample, the moderate to high isolation group demonstrated a significantly elevated risk of dementia compared with the low isolation group. The unadjusted Cox proportional hazards model (model 1) yielded a HR of 1.58 (95% CI 1.31-1.89, *P*<.001) for dementia in the moderate to high isolation group, indicating a 58% higher relative risk of dementia compared with the low isolation group. After adjusting for potential confounders (age, gender, race or ethnicity, number of baseline diseases, depression, anxiety, smoking status, and sleep disorders), the HR in model 2 attenuated to 1.22 (95% CI 1.01-1.47, *P*=.041), yet remained statistically significant. The validation sample corroborated these findings. The unadjusted model 1 revealed an HR of 2.36 (95% CI 1.86-3.01, *P*<.001) for the moderate to high isolation group, while the adjusted model 2 showed an HR of 1.62 (95% CI 1.27-2.08, *P*<.001), consistently indicating a significantly higher dementia risk in the moderate to high isolation group.

Table 2. Association between digital isolation and the risk of dementia.

Sample and variables	Event, n/N (%)	Model 1[a] HR[b] (95% CI)	*P* value	Model 2[c] HR (95% CI)	*P* value
Discovery sample					
Low isolation (reference range)	1803/4455 (40.47)	1 (Ref)	—[d]	1 (Ref)	—
Medium and high isolation	2652/4455 (59.53)	1.58 (1.31-1.89)	<.001	1.22 (1.01-1.47)	.04
Validation sample					
Low isolation (reference range)	1913/3734 (51.23)	1 (Ref)	—	1 (Ref)	—
Medium and high isolation	1821/3734 (48.77)	2.36 (1.86-3.01)	<.001	1.62 (1.27-2.08)	<.001
Pooled sample					
Low isolation (reference range)	3716/8189 (45.38)	1 (Ref)	—	1 (Ref)	—
Medium and high isolation	4473/8189 (54.62)	1.89 (1.63-2.19)	<.001	1.36 (1.16-1.59)	<.001

[a]Model 1: unadjusted Cox proportional hazards model.

[b]HR: hazard ratio.

[c]Model 2: Cox proportional hazards model adjusted for age, gender, race, education level, baseline disease, depression, anxiety, smoking status, and sleep difficulty.

[d]Not applicable.

To enhance the robustness of our findings, we applied meta-analytic techniques to pool results from both samples. The pooled analysis yielded HRs of 1.89 (95% CI 1.63-2.19, *P*<.001) in the unadjusted model 1 and 1.36 (95% CI 1.16-1.59, *P*<.001) in the adjusted model 2, further substantiating the significant risk increase.

Kaplan-Meier survival curves were used to visualize dementia incidence across different digital isolation groups (Figure 2).

XSL•FO
RenderX

These curves demonstrated a significantly higher probability of dementia development in the moderate to high isolation group, corroborating the Cox regression analysis results. Figure 3 illustrates the association between digital isolation and dementia risk across various demographic and clinical subgroups. Consistently elevated dementia risk was observed in the moderate to high isolation group across all subgroups, with notable differences in gender, age, and comorbidity status. These findings underscore the pervasive impact of digital isolation on dementia onset, demonstrating a consistent trend of increased risk both in the overall population and within specific subgroups.

Figure 3. Association between digital isolation and risk of dementia among subgroups. Reference group: Low digital isolation group. All models were adjusted for age, gender, race, baseline disease, depression, anxiety, smoking, sleep difficulty. P-int represents the heterogeneity among subgroups based on the metaregression analysis.

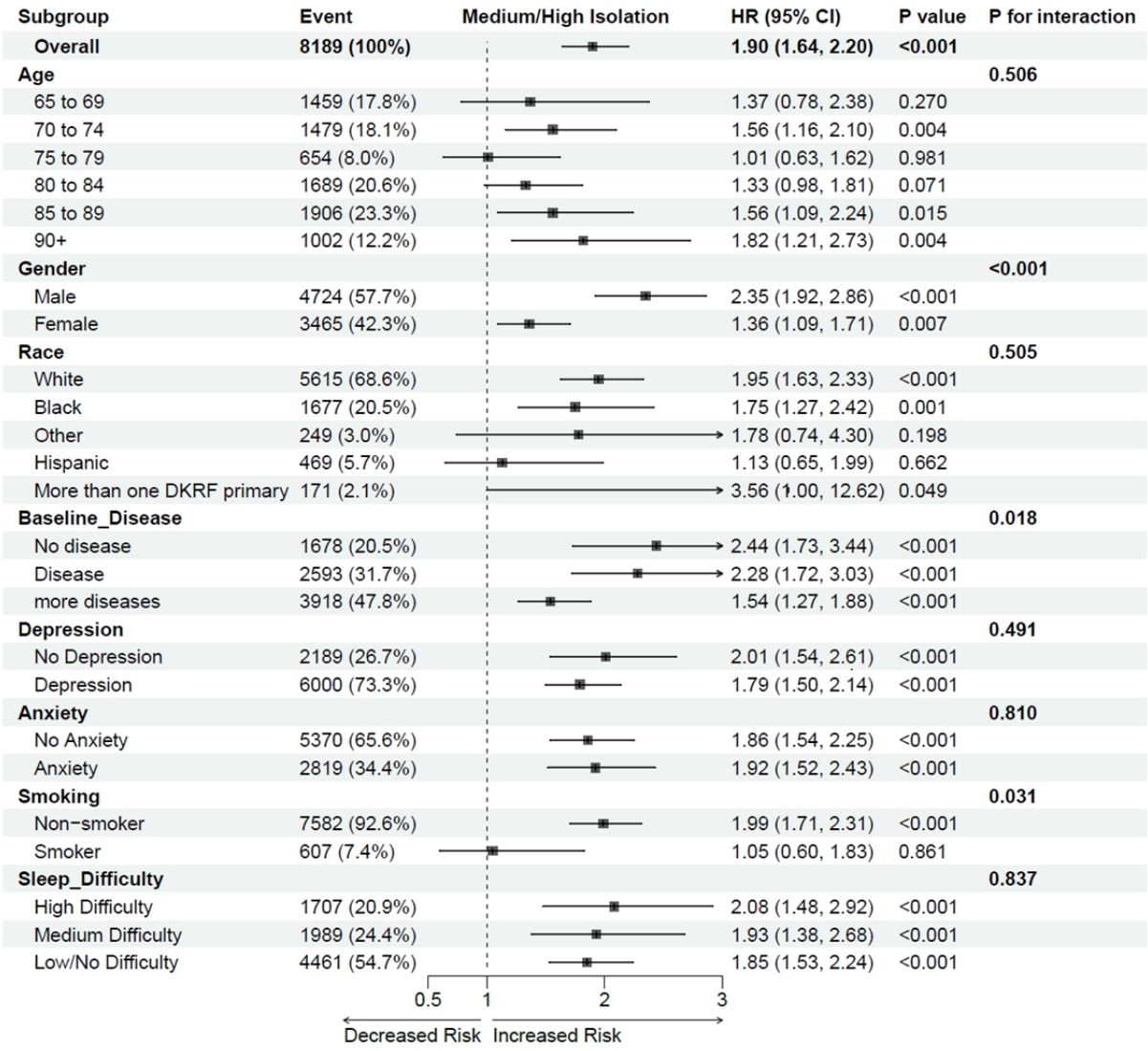

Association Between Digital Isolation Components and the Risk of Dementia

We investigated the association between individual components of digital isolation and incident dementia risk, with results summarized in Table 3. In the discovery sample, we analyzed the use of mobile phones, computers, tablets, and the internet in relation to dementia risk. The unadjusted Cox proportional hazards model (model 1) revealed that nonusers of mobile phones had a significantly higher dementia risk compared with users (HR 1.78, 95% CI 1.41-2.24, *P*<.001). After adjusting for covariates (model 2), the HR attenuated to 1.45 (95% CI 1.14-1.84, *P*=.002). For computer use, the adjusted HR was 1.15 (95% CI 0.96-1.38, *P*=.12), approaching but not reaching statistical significance. Similarly, adjusted HRs for tablet use were not statistically significant across samples, with a pooled HR of 1.05 (95% CI 0.78-1.42, *P*=.73).

XSL•FO
RenderX

Table 3. Association between digital isolation components and the risk of dementia.

Component and variables	Discovery sample				Validation sample				Pooled sample			
	Model 1[a], HR[b] (95% CI)	*P* values	Model 2[c], HR (95% CI)	*P* values	Model 1, HR (95% CI)	*P* values	Model 2, HR (95% CI)	*P* values	Model 1, HR (95% CI)	*P* values	Model 2, HR (95% CI)	*P* values
Phone use (nonuse vs use)	1.78 (1.41-2.24)	<.001	1.45 (1.14-1.84)	.002	3.05 (2.28-4.08)	<.001	2.12 (1.55-2.9)	<.001	2.05 (1.6-2.62)	<.001	1.65 (1.35-2.02)	<.001
Computer use (nonuse vs use)	1.22 (1-1.5)	.052	1.10 (0.88-1.36)	.386	1.50 (1.15-1.97)	.003	1.28 (0.98-1.66)	.071	1.25 (0.98-1.59)	.074	1.15 (0.96-1.38)	.118
Tablet use (nonuse vs use)	1.39 (1.02-1.88)	.035	1.28 (0.94-1.73)	.108	1.19 (0.88-1.6)	.264	1.10 (0.82-1.48)	.503	1.10 (0.8-1.53)	.548	1.05 (0.78-1.42)	.734
Email frequency (nonuse vs use)	0.74 (0.5-0.99)	.043	0.78 (0.58-1.04)	.090	1.03 (0.74-1.44)	.848	1 (0.72-1.38)	.998	0.95 (0.7-1.28)	.707	0.9 (0.66-1.23)	.51
Internet use (nonuse vs use)	1.42 (1.12-1.79)	.004	1.28 (1-1.64)	.047	1.75 (1.3-2.34)	<.001	1.56 (1.15-2.11)	.005	1.6 (1.24-2.07)	<.001	1.42 (1.15-1.76)	.001
Online activity (nonuse vs use)	1.12 (0.78-1.59)	.541	1.1 (0.77-1.57)	.612	1.93 (1.24-3.02)	.004	1.78 (1.14-2.79)	.012	1.6 (1.1-2.31)	.013	1.32 (1-1.74)	.048
Health-related online use (nonuse vs use)	1.08 (0.78-1.48)	.656	1.15 (0.83-1.59)	.402	0.92 (0.61-1.4)	.707	0.9 (0.58-1.36)	.594	0.95 (0.7-1.28)	.745	0.9 (0.64-1.28)	.576

[a]Model 1: unadjusted Cox proportional hazards model.

[b]HR: hazard ratio.

[c]Model 2: Cox proportional hazards model adjusted for age, gender, race, education level, baseline disease, depression, anxiety, smoking status, and sleep difficulty.

In the validation sample, nonusers of mobile phones exhibited a significantly higher dementia risk, with an unadjusted HR of 3.05 (95% CI 2.28-4.08, *P*<.001), which decreased to 2.12 (95% CI 1.55-2.9, *P*<.001) after adjustment. Nonusers of the internet demonstrated a robust association with dementia risk, with a pooled adjusted HR of 1.42 (95% CI 1.15-1.76, *P*=.001). For individuals not engaging in online activities, the pooled adjusted HR was 1.32 (95% CI 1-1.74, *P*=.05), further underscoring the importance of digital engagement. However, nonusers of health-related online platforms showed no significant association with dementia risk (pooled HR 0.9, 95% CI: 0.64-1.28, *P*=.58).

These findings underscore that specific digital behaviors, particularly mobile phone and internet use, are significantly associated with a lower risk of dementia. In contrast, the influence of other digital components, such as computer or tablet use, appears less consistent or pronounced.

Sensitivity Analyses

To evaluate the robustness of our findings and address potential confounders, we conducted several sensitivity analyses. First, we repeated the Cox regression models after excluding participants with incomplete education data, confirming that the significant association between digital isolation and dementia risk remained robust even after adjusting for educational level (adjusted HR range 1.20-1.45, all *P*<.05). Second, we applied alternative thresholds for defining digital isolation, such as categorizing digital isolation index ≥4 as high isolation, and observed results consistent with the primary analysis. Third, we excluded individuals who developed dementia or died within the first 2 years of follow-up to reduce potential reverse causation, finding HRs comparable with those in the main analysis (HR changes<10%). Finally, we replicated the main models without applying survey weights. Although the unweighted models yielded slightly smaller standard errors, the overall trends remained consistent. Collectively, these sensitivity analyses indicate that our key conclusions are robust across varying analytical assumptions and sample selections.

Discussion

Principal Results

This study examined the association between digital isolation and dementia risk, demonstrating that higher levels of digital isolation significantly increase the risk of dementia. These findings emphasize the critical role of digital engagement in promoting cognitive health among older adults and provide valuable insights for public health policy development. By using the nationally representative NHATS dataset and its survey

XSL•FO
RenderX

weights, we ensured that our results reflect the broader US older adult population, thereby enhancing the robustness of our conclusions. Our analysis identified a significant relationship between digital isolation and dementia risk. In both discovery and validation samples, individuals who did not use basic digital devices (eg, mobile phones and the internet) or who lacked online activity participation exhibited a substantially higher dementia risk compared with their digitally engaged peers. Specifically, nonusers of mobile phones had a 1.65-fold increased risk (95% CI 1.35-2.02, *P*<.001), while nonusers of the internet demonstrated a 1.42-fold increased risk (95% CI 1.15-1.76, *P*=.001) in the pooled analysis. In addition, the pooled analysis showed that individuals with a higher digital isolation index faced a 1.36-fold increased risk of dementia (95% CI 1.16-1.59, *P*<.001). These results are consistent with previous research identifying social isolation and lack of social interaction as independent risk factors for dementia [13,32].

However, some digital activities, such as emailing or texting, showed no significant association with dementia risk. This might be because these activities are less interactive or infrequently used by older adults, offering limited cognitive or social stimulation compared with more engaging digital behaviors, such as frequent phone or internet use.

Furthermore, this study validates digital isolation as a novel form of isolation that may influence cognitive health through multiple mechanisms. Digital technologies can promote social interaction and participation, crucial for maintaining cognitive function [33-35]. Conversely, a lack of digital engagement may exacerbate social isolation, thereby increasing cognitive decline risk. In addition, digital technology use can provide cognitive stimulation, potentially helping older adults maintain brain vitality [36,37]. Thus, digital isolation not only reflects diminished social interaction but may also represent insufficient cognitive stimulation. These findings have significant public health implications. As supported by recent studies, educational attainment can significantly influence individuals' ability to engage with digital tools, highlighting the importance of prioritizing older adults with lower educational backgrounds in targeted interventions [38-40]. As society undergoes digital transformation, issues related to digital engagement among older adults are becoming increasingly salient. Our study suggests that promoting digital technology use among older adults, particularly within vulnerable populations, could effectively mitigate dementia risk. This implies that enhancing digital literacy and expanding access to digital resources should be integral components of dementia prevention strategies. While other studies may not have used the specific "digital isolation index," research examining older adults' online engagement or technology use consistently demonstrates that digitally engaged older adults report better cognitive and psychosocial outcomes [10,41,42]. These findings reinforce the protective role of digital engagement in cognitive health and underscore the importance of encouraging digital technology adoption among older adults.

Limitations

However, this study has limitations. Despite controlling for multiple covariates, the observational nature of our study design precludes complete elimination of confounding factors. For instance, individuals' health behaviors and cognitive abilities may simultaneously influence their digital technology use and dementia risk, potentially leading to reverse causality [19,43]. Furthermore, educational attainment could discourage older adults from adopting digital devices, complicating the interpretation of whether digital isolation is a cause or result of declining cognitive function. In addition, our reliance on self-reported data may introduce information bias, particularly in the preclinical stages of dementia when individuals' recollection of their digital usage may not be entirely accurate.

Notwithstanding these limitations, this study provides important empirical support for exploring the relationship between digital isolation and dementia risk. Future research should further investigate factors that promote or hinder digital engagement among older adults to develop more effective interventions. In addition, prospective longitudinal studies, randomized controlled trials, and qualitative investigations are crucial for further validating our findings and examining the causal relationship between digital technology use and cognitive health.

Conclusions

This study, examining the association between digital isolation and dementia risk in older adults, underscores the critical role of digital engagement in maintaining cognitive health. Our findings demonstrate that digital isolation significantly increases dementia risk, particularly among individuals who do not use basic digital devices, lack online communication, and abstain from online activities. These findings indicate that strategies focused on enhancing digital literacy, promoting equitable access to digital tools, and addressing educational disparities may effectively mitigate dementia risk in an increasingly digitalized society.

Our research not only provides new evidence of the detrimental effects of digital isolation but also offers valuable insights for future public health interventions. For instance, community-based digital literacy programs and targeted technology access initiatives could help reduce digital isolation among older adults, especially those from underserved or low-education backgrounds. Enhancing digital literacy and expanding accessibility to digital resources among older adults may effectively reduce digital isolation, potentially lowering dementia incidence. Future research should explore the specific mechanisms through which digital engagement affects cognitive health and further validate the causal relationship between digital isolation and dementia through longitudinal studies and intervention trials.

In conclusion, this study highlights the importance of digital technology in older adults' health and advocates for the integration of digital literacy education, user-friendly technological interfaces, and widespread digital resource availability in public health policies. These measures will facilitate older adults' integration into the digital society, thereby potentially improving their cognitive health and overall quality of life. Ensuring that these interventions are adapted to individual educational levels may be pivotal for maximizing their preventive impact.

XSL•FO
RenderX

Acknowledgments

We would like to express our sincere gratitude to everyone who contributed to this study. Finally, we acknowledge the assistance of ChatGPT in refining the language and grammar of this manuscript.

Authors' Contributions

CD and NS, as co–first authors, were instrumental in the conceptualization, data analysis, and drafting of the manuscript. GL provided essential software support and was pivotal in data visualization. KZ offered valuable supervision and secured the necessary resources and funding for the project. SY, as the corresponding author, provided overall guidance, ensuring the study's success through his expertise in project administration and manuscript review.

Conflicts of Interest

None declared.

References

1. Dementia U. What is Dementia. London: Dementia UK[Google Scholar]; 2019.
2. Kumar A, Vandana. Dementia: an Overview. J. Drug Delivery Ther 2013;3(3):163-167. [doi: 10.22270/jddt.v3i3.485]
3. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. Lancet Public Health 2022;7(2):e105-e125 [FREE Full text] [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
4. Licher S, Ahmad S, Karamujić-Čomić H, Voortman T, Leening MJG, Ikram MA, et al. Genetic predisposition, modifiable-risk-factor profile and long-term dementia risk in the general population. Nat Med 2019;25(9):1364-1369 [FREE Full text] [doi: 10.1038/s41591-019-0547-7] [Medline: 31451782]
5. Fillit H, Nash DT, Rundek T, Zuckerman A. Cardiovascular risk factors and dementia. Am J Geriatr Pharmacother 2008;6(2):100-118. [doi: 10.1016/j.amjopharm.2008.06.004] [Medline: 18675769]
6. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
7. Mubarak F, Suomi R. Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. Inquiry 2022;59:469580221096272 [FREE Full text] [doi: 10.1177/00469580221096272] [Medline: 35471138]
8. Reine I, Ivanovs A, Mieriņa I, Gehtmane-Hofmane I, Koroļeva I. Overcoming social isolation with digital technologies among ageing populations during covid-19. SIE 2021;4:171-178. [doi: 10.17770/sie2021vol4.6356]
9. Sin F, Berger S, Kim I, Yoon D. Digital social interaction in older adults during the COVID-19 pandemic. Proc. ACM Hum.-Comput. Interact 2021;5(CSCW2):1-20. [doi: 10.1145/3479524]
10. Zhan R, Mpofu E, Olanrewaju S, Sutherland L. Community engagement in older adults with cognitive decline: the usage of information communication technology. Innovation in Aging 2023;7((Supplement_1)):1007. [doi: 10.1093/geroni/igad104.3236]
11. Byrne K, Ghaiumy Anaraky R. Identifying racial and rural disparities of cognitive functioning among older adults: the role of social isolation and social technology use. J Gerontol B Psychol Sci Soc Sci 2022;77(10):1779-1790 [FREE Full text] [doi: 10.1093/geronb/gbac055] [Medline: 35312775]
12. Neter E, Chachashvili-Bolotin S, Erlich B, Ifrah K. Benefiting from digital use: prospective association of internet use with knowledge and preventive behaviors related to alzheimer disease in the israeli survey of aging. JMIR Aging 2021;4(2):e25706 [FREE Full text] [doi: 10.2196/25706] [Medline: 33929331]
13. Yu B, Steptoe A, Chen Y, Jia X. Social isolation, rather than loneliness, is associated with cognitive decline in older adults: the China health and retirement longitudinal study. Psychol Med 2021;51(14):2414-2421. [doi: 10.1017/S0033291720001014] [Medline: 32338228]
14. Wiese L, Park J. Digital learning and online chair yoga for rural underserved older adults at risk of cognitive decline. Innovation in Aging 2022;6((Supplement_1)):94. [doi: 10.1093/geroni/igac059.372]
15. Lu SY, Yoon S, Yee WQ, Heng Wen Ngiam N, Ng KYY, Low LL. Experiences of a community-based digital intervention among older people living in a low-income neighborhood: qualitative study. JMIR Aging 2024;7:e52292 [FREE Full text] [doi: 10.2196/52292] [Medline: 38662423]
16. Horoszkiewicz B. Digital dementia and its impact on human cognitive and emotional functioning. J Educ Health Sport 2022 Jan;12(11):290-296. [doi: 10.12775/jehs.2022.12.11.038]
17. Talbot CV, Briggs P. The use of digital technologies by people with mild-to-moderate dementia during the COVID-19 pandemic: a positive technology perspective. Dementia (London) 2022;21(4):1363-1380. [doi: 10.1177/14713012221079477] [Medline: 35333111]

XSL•FO
RenderX

18. Rai HK, Kernaghan D, Schoonmade L, Egan KJ, Pot AM. Digital technologies to prevent social isolation and loneliness in dementia: a systematic review. J Alzheimers Dis 2022;90(2):513-528 [FREE Full text] [doi: 10.3233/JAD-220438] [Medline: 36120780]
19. Jin Y, Jing M, Ma X. Effects of digital device ownership on cognitive decline in a middle-aged and elderly population: longitudinal observational study. J Med Internet Res 2019;21(7):e14210 [FREE Full text] [doi: 10.2196/14210] [Medline: 31359864]
20. Xavier AJ, d'Orsi E, de Oliveira CM, Orrell M, Demakakos P, Biddulph JP, et al. English longitudinal study of aging: can internet/e-mail use reduce cognitive decline? J Gerontol A Biol Sci Med Sci 2014;69(9):1117-1121 [FREE Full text] [doi: 10.1093/gerona/glu105] [Medline: 25116923]
21. Cornwell EY, Waite LJ. Social disconnectedness, perceived isolation, and health among older adults. J Health Soc Behav 2009;50(1):31-48 [FREE Full text] [doi: 10.1177/002214650905000103] [Medline: 19413133]
22. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukopadhyay T, Scherlis W. Internet paradox. A social technology that reduces social involvement and psychological well-being? Am Psychol 1998;53(9):1017-1031. [doi: 10.1037//0003-066x.53.9.1017] [Medline: 9841579]
23. Makizako H, Shimada H, Tsutsumimoto K, Lee S, Doi T, Nakakubo S, et al. Social frailty in community-dwelling older adults as a risk factor for disability. J Am Med Dir Assoc 2015;16(11):1003.e7-1003.11. [doi: 10.1016/j.jamda.2015.08.023] [Medline: 26482055]
24. Wei K, Nyunt MSZ, Gao Q, Wee SL, Yap KB, Ng TP. Association of Frailty and Malnutrition With Long-term Functional and Mortality Outcomes Among Community-Dwelling Older Adults: Results From the Singapore Longitudinal Aging Study 1. JAMA Netw Open 2018;1(3):e180650 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0650] [Medline: 30646023]
25. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 2007;3(3):186-191. [doi: 10.1016/j.jalz.2007.04.381] [Medline: 19595937]
26. Henderson VW. Estrogen, cognition, and a woman's risk of alzheimer's disease. Am J Med 1997;103(3A):11S-18S. [doi: 10.1016/s0002-9343(97)00261-1] [Medline: 9344402]
27. Lim U, Wang S, Park S, Bogumil D, Wu AH, Cheng I, et al. Risk of alzheimer's disease and related dementia by sex and race/ethnicity: the multiethnic cohort study. Alzheimers Dement 2022;18(9):1625-1634 [FREE Full text] [doi: 10.1002/alz.12528] [Medline: 34882963]
28. Burton C, Campbell P, Jordan K, Strauss V, Mallen C. The association of anxiety and depression with future dementia diagnosis: a case-control study in primary care. Fam Pract 2013;30(1):25-30 [FREE Full text] [doi: 10.1093/fampra/cms044] [Medline: 22915794]
29. Becker E, Orellana Rios CL, Lahmann C, Rücker G, Bauer J, Boeker M. Anxiety as a risk factor of alzheimer's disease and vascular dementia. Br J Psychiatry 2018;213(5):654-660. [doi: 10.1192/bjp.2018.173] [Medline: 30339108]
30. Ott A, Slooter AJ, Hofman A, van Harskamp F, Witteman JC, Van Broeckhoven C, et al. Smoking and risk of dementia and alzheimer's disease in a population-based cohort study: the rotterdam study. Lancet 1998;351(9119):1840-1843. [doi: 10.1016/s0140-6736(97)07541-7] [Medline: 9652667]
31. Shi L, Chen SJ, Ma MY, Bao YP, Han Y, Wang YM, et al. Sleep disturbances increase the risk of dementia: a systematic review and meta-analysis. Sleep Med Rev 2018;40:4-16. [doi: 10.1016/j.smrv.2017.06.010] [Medline: 28890168]
32. Guo L, Luo F, Gao N, Yu B. Social isolation and cognitive decline among older adults with depressive symptoms: prospective findings from the China health and retirement longitudinal study. Arch Gerontol Geriatr 2021;95:104390. [doi: 10.1016/j.archger.2021.104390] [Medline: 33752099]
33. Dodge HH, Zhu J, Mattek NC, Bowman M, Ybarra O, Wild KV, et al. Web-enabled conversational interactions as a means to improve cognitive functions: results of a 6-week randomized controlled trial. Alzheimers Dement (N Y) 2015;1(1):1-12 [FREE Full text] [doi: 10.1016/j.trci.2015.01.001] [Medline: 26203461]
34. Berner J, Comijs H, Elmståhl S, Welmer AK, Sanmartin Berglund J, Anderberg P, et al. Maintaining cognitive function with internet use: a two-country, six-year longitudinal study. International Psychogeriatrics 2019;31(7):929-936. [doi: 10.1017/s1041610219000668]
35. Wang J, Liu J, Wang X, Zhu J, Bai Y, Che Y, et al. Association between change in social participation and improved cognitive function among older adults in China: a national prospective cohort study. Health Soc Care Community 2022;30(6):e4199-e4210. [doi: 10.1111/hsc.13814] [Medline: 35396737]
36. Gottlieb S. Mental activity may help prevent dementia. BMJ 2003;326(7404):1418. [doi: 10.1136/bmj.326.7404.1418-h]
37. Wu Z, Pandigama DH, Wrigglesworth J, Owen A, Woods RL, Chong TTJ, et al. Lifestyle enrichment in later life and its association with dementia risk. JAMA Netw Open 2023;6(7):e2323690 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.23690] [Medline: 37450299]
38. Shin J, Kang H, Choi S, Choi J. Profiles of digital literacy among community-dwelling Korean older adults: a latent profile analysis. Innovation in Aging 2023;7:1101. [doi: 10.1093/geroni/igad104.3536]
39. Pelegrini LNDC, Casemiro FG, Bregola A, Ottaviani AC, Pavarini SCI. Performance of older adults in a digital change detection task: the role of heterogeneous education. Appl Neuropsychol Adult 2023:1-9. [doi: 10.1080/23279095.2023.2189520] [Medline: 36966732]

XSL•FO
RenderX

40. Ali Hafez S, Mohamed El-Sayed El-sheikh S, Abd El Fattah Ibrahim N, youssef Gaid J. Exploring dgital literacy among community- dwelling older adults. Alexandria Scientific Nursing Journal 2024;26(2):186-197. [doi: 10.21608/asalexu.2024.361332]
41. Kiselica AM, Stojanovic M, Waters A, Benge J. Association of technology use with subjective cognitive decline among older adults. Alzheimer's & Dementia 2023;19(S23). [doi: 10.1002/alz.073326]
42. Ngiam NHW, Yee WQ, Teo N, Yow KS, Soundararajan A, Lim JX, et al. Building digital literacy in older adults of low socioeconomic status in Singapore (Project Wire Up): nonrandomized controlled trial. J Med Internet Res 2022;24(12):e40341 [FREE Full text] [doi: 10.2196/40341] [Medline: 36459398]
43. Smith S, Splonskowski M, Jacova C. Self-experienced cognitive function in the digital era: are older adults at risk of subjective cognitive decline? Innovation in Aging. Innovation in Aging 2021;5((Supplement_1)):752. [doi: 10.1093/geroni/igab046.2775]

Abbreviations

GED: general educational development
HR: hazard ratio
NHATS: National Health and Aging Trends Study
PSU: primary sampling unit

Edited by A Mavragani; submitted 14.08.24; peer-reviewed by H Shah, AJ Santos; comments to author 29.12.24; revised version received 11.01.25; accepted 12.01.25; published 19.02.25.

Please cite as:
Deng C, Shen N, Li G, Zhang K, Yang S
Digital Isolation and Dementia Risk in Older Adults: Longitudinal Cohort Study
J Med Internet Res 2025;27:e65379
URL: https://www.jmir.org/2025/1/e65379
doi:10.2196/65379
PMID:39969956

©Cheng Deng, Na Shen, Guangzhou Li, Ke Zhang, Shijun Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Mediating Role of Meaning-Making in the Relationship Between Mental Time Travel and Positive Emotions in Stress-Related Blogs: Big Data Text Analysis Research

Yidi Chen[1*], PhD; Lei Zheng[2], PhD; Jinjin Ma[3,4], PhD; Huanya Zhu[3,4], MA; Yiqun Gan[3,4*], PhD

[1]School of Humanities and Social Sciences, Beijing Forestry University, Beijing, China

[2]School of Business and the Institute for Sustainable Development, Macau University of Science and Technology, Macau, China

[3]School of Psychological Cognitive Sciences, and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

[4]School of Psychological Cognitive Sciences, and Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yidi Chen, PhD
School of Humanities and Social Sciences
Beijing Forestry University
No. 35 Qinghua East Road
Beijing, 100083
China
Phone: 86 01062336834
Email: ydchen@bjfu.edu.cn

Abstract

Background: Given the ubiquity of stress, a key focus of stress research is exploring how to better coexist with stress.

Objective: This study conducted text analysis on stress-related Weibo posts using a web crawler to investigate whether these posts contained positive emotions, as well as elements of mental time travel and meaning-making. A mediation model of mental time travel, meaning-making, and positive emotions was constructed to examine whether meaning-making triggered by mental time travel can foster positive emotions under stress.

Methods: Using Python 3.8, the original public data from active Weibo users were crawled, yielding 331,711 stress-related posts. To avoid false positives, these posts were randomly divided into two large samples for cross-validation (sample 1: n=165,374; sample 2: n=166,337). Google's natural language processing application programming interface was used for word segmentation, followed by text and mediation analysis using the Chinese psychological analysis system "Wenxin." A mini–meta-analysis of the mediation path coefficients was conducted. Text analysis identified mental time travel words, meaning-making words, and positive emotion words in stress-related posts.

Results: The constructed mediation model of mental time travel words (time words), meaning-making words (causal and insightful words), and positive poststress emotions validated positive adaptation following stress. A mini–meta-analysis of two different mediation models constructed in the two subsamples indicated a stable mediation effect across the 2 random subsamples. The combined effect size (B) obtained was .013 (SE 0.003, 95% CI 0.007-0.018; P<.001), demonstrating that meaning-making triggered by mental time travel in stress-related blog posts can predict positive emotions under stress.

Conclusions: Individuals can adapt positively to stress by engaging in meaning-making processes that are triggered by mental time travel and reflected in their social media posts. The study's mediation model confirmed that mental time travel leads to meaning-making, which fosters positive emotional responses to stress. Mental time travel serves as a psychological strategy to facilitate positive adaptation to stressful situations.

(J Med Internet Res 2025;27:e63407) doi:10.2196/63407

KEYWORDS

stress; meaning-making; mental time travel; big data; mini meta-analysis; text analysis; coping mechanisim; Weibo; post; web crawler; positive emotion; emotion; meta-analysis; anxiety; depression; mental health; ecological momentary assessment; EMA; stress model; natural language processing; NLP

XSL•FO
RenderX

Introduction

Stress and Mental Health

Researchers have consistently linked stressful events to negative emotions and adverse psychological health outcomes, viewing stress as an inevitable precursor of anxiety, depression, and other negative emotional states [1,2]. A recent meta-analysis of ecological momentary assessments suggested a significant positive correlation between transient negative emotions and cortisol levels—a biomarker of stress—whereas positive emotions were significantly and negatively correlated with cortisol levels [3].

The cumulative stress model posits that the negative impact of stress on an individual's emotions and psychological health is cumulative, emphasizing the severe and lasting effects of early life stress on the brain [4-6]. Prolonged exposure to stress can also diminish a person's ability to withstand future stressors, thereby impacting mental health [7]. Stress has been shown to negatively impact memory, executive function, emotional regulation, and the processing of social and emotional stimuli [8]. Chronic exposure to stress could lead to neurodevelopmental changes in the brain structure and function [9].

These findings and the cumulative stress model suggest that experiencing stressful events is detrimental to individuals, and avoiding such events is beneficial for mental health. However, these studies only emphasize individual vulnerability, overlooking psychological resilience and the potential for growth and development in the face of adversity. The control and mastery theory of anxiety suggests that effective coping with stressful events can enhance an individual's sense of control and better prepare them for future stressors [10]. The concept of "stress inoculation" also supports the idea that "what does not kill us makes us stronger" [11]. Recent research evidence indicates that over 60% of individuals demonstrate adaptive functioning after experiencing stressful events or adversity, commonly referred to as "stress resilience" [12].

The cyclical model of stress resilience and stress-related growth posits that the relationship between stress and mental health outcomes such as positive emotions is not necessarily negative [13-17]. After a stressful event, individuals initiate an adjustment process that leads to psychological outcomes of the stress event. The nature of the adjustment process and its outcomes depend on the severity of the stress experienced. If the severity of a stressful event is within an individual's tolerance, successful adjustment could enhance resilience and lead to positive coping outcomes. Events causing minimal stress could require little adjustment and are likely to lead to the recovery and restoration of resilience to baseline levels [18-20]. Highly stressful events can overwhelm an individual's resilience and cause damage, at least temporarily. The level of resilience is modulated by various factors, of which meaning-making is a crucial coping mechanism [13,18,21-24].

Mental Time Travel to Trigger Meaning-Making

Meaning-making is a stress-coping strategy aimed at altering an individual's appraisal of a situation to harmonize their beliefs, goals, and stress context. Unlike classic problem- and emotion-focused coping strategies, meaning-making does not attempt to change problematic situations or directly reduce negative emotions or pain [25]. Instead, it unifies beliefs, goals, and stress contexts by searching for meaning [26].

The meaning-making model proposed and revised by Park [27] is a widely accepted theory for understanding how individuals cope positively with stressful events and situations. It centers on the concept of "global meaning"—a cognitive framework that individuals use to interpret their experiences and motivations [25]. The model posits that when individuals encounter situations that might challenge or reinforce this global meaning, they engage in an evaluative process that assigns significance to these situations. The degree of discrepancy between the appraised meaning and its inherent global meaning determines the level of distress experienced. This distress initiates the process of meaning-making, wherein individuals strive to reconcile the gap between situational meaning and global understanding. This reconciliation effort aims to restore a sense of coherence in the world and reaffirm the value of their lives. Ultimately, achieving harmonization between situational and global meanings facilitates better adaptation to stress, highlighting the transformative role of meaning-making in managing life challenges [28-32].

In laboratory studies, two common methods are often used to elicit meaning-making in individuals: mental time travel and meaning-threat paradigms. Among these, mental time travel, a type of typical mental travel that predominantly involves recalling the past or imagining the future, is the most frequently used. Mental time travel is the ability to project oneself into the past or future through mental time travel [33]. As previously stated, individuals were said to experience mental time travel when they re-experienced events from their past and imagined events in their future [34,35]. Construct level theory suggests that when individuals remember the past or think about the future, they process things at a more abstract and superordinate level because temporal distance causes the mental distance to become farther away [36]. This is consistent with the findings of research in the area of mental time travel and the process of meaning-making and meaning-acquisition [37]. Research suggests that coherence is an important component of meaning in life, which involves an understanding of personal experiences and a sense of life events [38]. Through mental time travel, people recall past events or imagine future events, which helps them integrate their life experiences from different times into a more cohesive meaning and enhances their self of purpose in life [39]. Numerous studies indicate that mental time travels in the form of nostalgia or future contemplation can enhance the meaning in life or positive emotions owing to the role of meaning-making [40-42]. During stressful situations, individuals initiate the process of meaning-making; seek clues from the past, present, or future; integrate life experiences to redefine the current stress context; or adjust their global meaning to guide their future lives [43]. Thus, when individuals attempt to mentally travel to the past or future, they can enhance positive emotions and predict higher psychological well-being through meaning-making [27].

Experimental studies and research into cognitive neural mechanisms have also found evidence that mental time travel

XSL•FO
RenderX

successfully initiates meaning-making. In one laboratory study, participants were randomly assigned to future, present, and past groups and asked to imagine 2 independent events within different timeframes. Participants imagining the future and past reported higher meaning in life than those imagining the present. This indicates that mental time travel, through self-distancing from the here and now, promotes the integration of temporal information and causal understanding, becoming a stable paradigm for eliciting meaning-making [44-46]. In a functional magnetic resonance imaging study using a similar mental time-travel paradigm to successfully elicit meaning-making, a decrease in default mode network activity was observed, linking meaning-making with positive psychological health outcomes [47].

This Study

The internet is now an integral part of people's lives and the rapid development of social media has significantly altered how people express their emotions [48]. The evolution of the internet and big data has propelled psychological research. In Chinese social media, Sina Weibo has become a crucial data source for psychological studies owing to its fast dissemination, timely information, and wide audience [49]. When individuals perceive stress, they may express their emotions or reflect by posting on Sina Weibo. The cyclical model of stress resilience and stress-related growth posits that stress has an adaptive function that is facilitated by the coping process [50]. Research indicates that expressive inhibition is a maladaptive coping strategy and a possible mechanism for affect dysregulation following trauma [51]. In contrast, expressive writing during stressful situations has been shown to relieve stress and hone coping strategies, thereby alleviating negative emotions and promoting positive adaptation [52]. This process can also be accomplished through mental time travel, which facilitates the process of meaning-making by enabling individuals to integrate insights about the meaning of life from the past and future, thus promoting positive adaptation following stress [43]. Therefore, a theoretical model that comprises mental time travel and the cyclical model of stress resilience and stress-related growth is proposed in this study. This study explored whether such actions entail a process of meaning-making, such as causal connections or insights, and whether successful meaning-making is followed by positive adaptation to stress, manifested as positive emotions (Figure 1).

Figure 1. The hypothesis model.

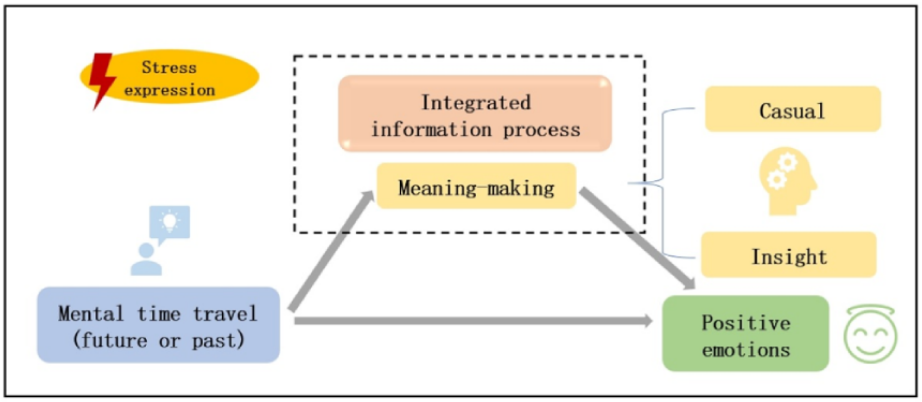

Specifically, this research involved text analysis of stress-related blog posts on Weibo to investigate the presence of positive elements as indicators of poststress adaptation and to explore elements of meaning-making in everyday stress and coping. Previous studies have frequently used mental time travel as a stable paradigm to elicit meaning-making, often inviting participants to imagine past or future events [44-46]. Mental time travel can reliably initiate meaning-making for individuals, thereby enhancing meaning in their lives. The words "future" and "past" are used as indicators of mental time travel, and causal and insightful words serve as key components of the meaning-making process. This study examined whether the meaning-making process mediates the impact of mental time travel on positive emotions. In this study, words related to past and future blog posts were considered components of mental time travel that triggered meaning-making. Emotional words were used to represent positive outcomes under stress, whereas words referring to the past or future served as markers of mental travel. In qualitative research and text analysis, insightful words (eg, "originally" and "comprehend") and causal words (eg, "because," "therefore," and "thus") were often used as components of meaning-making [53]. Thus, these words were used as indicators of meaning-making. Building on this, a mediation model of mental time travel, meaning-making, and positive emotions poststress was constructed, and a mini–meta-analysis was used to test whether mental time travel-induced meaning-making can promote positive emotions under stress.

Meta-analysis is a data analysis method that reanalyzes multiple independent quantitative studies on the same question to derive more generalizable conclusions [54]. Recently, statisticians have recognized that meta-analyses do not necessarily require several studies, and even two effect sizes can suffice. In fact, even with a few effect sizes, meta-analysis can summarize them, not only clarifying whether the research question has been consistently answered but also enhancing the persuasiveness of the conclusions, especially when the detected effect sizes are small. Individual studies may lack sufficient evidence, but collectively, they can be compelling. Researchers suggest that even a limited number of studies (or their subgroups) can be compared with each other. Such an internal meta-analysis of conceptually comparable studies is known as a mini–meta-analysis [55].

XSL•FO
RenderX

This study aimed to explore the mediating role of meaning-making in the relationship between mental time travel and positive emotions in stress-related blogs. In this study, a mini–meta-analysis was performed to detect the mediation model.

Methods

Research Objects

This study was conducted using original posts from active microblog users, excluding accounts belonging to institutions and media outlets. The screening criteria were as follows: (1) a publicly published blog by an individual user, (2) the blog was posted between December 2019 and November 2020, and (3) the blog post included the keyword ‘‘stress.” All blog posts have been anonymized. A statistical analysis was performed, revealing 331,711 posts related to “stress.” Given the propensity of big data to yield false positives, these posts were split into two random samples using the random grouping function in SPSS (version 24.0; IBM Corp). This division facilitated cross-validation of data reliability. Ensuing analyses were conducted on these two substantial samples (sample 1: n=165,374; sample 2: n=166,337).

Ethical Considerations

This study was approved by the School of Psychological and Cognitive Science, Peking University Ethics Committee (#2018-10-03).

Design and Programming

The Sina Weibo blog posts were harvested using Python 3.8, using the “lxml,” “requests,” and “tqdm” libraries to acquire and process HTML page data. This method involves retrieving key data such as the source, ID, content of the blog, date, and metrics, including the number of likes, comments, and reposts. All retrieved data were then systematically stored for further analysis.

Data Analysis

Google’s natural language processing application programming interface was used for word segmentation. Following this, a dictionary based on the Linguistic Inquiry and Word Count (LIWC) framework was used to conduct a word frequency analysis. This process enabled the identification of high-frequency words and their respective counts in the blog posts. Subsequently, the Chinese psychological analysis system “Wenxin”—developed by the Computational Network Psychology Laboratory at the Institute of Psychology, Chinese Academy of Sciences—was applied to analyze the content of the blog posts. “Wenxin” is a sophisticated software system designed for the linguistic analysis of Chinese texts. It features automatic word separation and text analysis capabilities, leveraging both LIWC2007 and the orthographic Chinese C-LIWC lexicon, which is tailored to the linguistic context of simplified Chinese in mainland China [56]. The “Wenxin” system uses lexicon technology to conduct word segmentation, extract word classes, perform word frequency analysis, and calculate the proportion of mental time travel, meaning-making, and positive emotion words within the entire blog post. During the process, “happy” is generally considered a word signifying positive emotion, while “not happy” would be classified as a word signifying negative emotion. In order to prevent large sample false positives, we used random grouping to randomly divide the sample into two subsamples for cross-validation (Koul et al [57]). Correlation and mediation analyses were conducted using SPSS version 24.0. During mediation analyses, path coefficients and effect sizes (B and SE) were extracted and stored in a dataset. First, using the SPSS 24.0 macro file PROCESS [58], a bootstrapping procedure was used to verify whether mental time travel words (future and past time words) could mediate the effect of meaning-making words (causal words, eg, because, effect, hence, and insightful words, eg, think, know, consider; see Table S5 in Multimedia Appendix 1) on positive emotions under stress [56,59]. Model 4 was selected by setting the number of samples to 5000 and using a bias-corrected 99% CI. To control for the type I error inflation caused by multiple comparisons, the significance level (α) of correlations was adjusted to .01 and the CI of the mediating effect was adjusted to 99% with Bonferroni correction [60].

Mediation heterogeneity tests, composite effect size calculations, and publication bias tests were performed using the “metafor” package in R (R Foundation and R Caucus). For visualization purposes, the “forestplot” function was used. A meta-analysis was conducted to verify the stability and homogeneity of these effects across the samples. The first step was to perform a Q-test for homogeneity to assess the similarity between multiple independent studies. If the homogeneity test was not significant and I^2 was <40%, the criteria for homogeneity were met, and a fixed-effects model was used to estimate the combined effect size. Conversely, if the homogeneity test was significant, a random-effects model was used.

Results

In the retrieved blog posts, people’s stressful events were concentrated in the areas of work stress (eg, working all day at the office with thunder outside), life stress (eg, pressure to buy and rent), academic stress (eg, pressure to graduate), family stress (eg, babysitting), and intimate relationships (eg, finding a partner). By leveraging the outcomes of Google Segmentation and Word Frequency Analysis, we derived insights into sentiment tones and word frequency in blog posts concerning stress and meaning-making strategies. The analysis of these charts revealed that in stress-related blog posts, prevalent themes included personal references such as “self,” “we,” and “they.” Additionally, there was a significant presence of cause-and-effect related terms like “reason,” “so,” and “because.” These findings underscore the process of meaning-making. There were also time-related elements such as “one day” and “tomorrow” (Table 1).

XSL•FO
RenderX

Table 1. Word frequency table of high-frequency words for word frequency analysis of microblog posts (top 60).

No.	Text	Frequency
1	Self	4491
2	No	2024
3	We	1692
4	Reason	1587
5	Livelihood	1535
6	Can	1439
7	Moment	1326
8	Work	1230
9	Feel	1190
10	Now	1179
11	Because	1136
12	Don’t	1084
13	But	1066
14	This	970
15	Today	951
16	Hope	943
17	Feelings	909
18	Recently	869
19	How	829
20	Happy	764
21	If	762
22	Time	760
23	May	759
24	Try	745
25	Everyone	737
26	So	725
27	They	712
28	Thing	708
29	Other	692
30	Everyday	682
31	Actually	669
32	Won’t	666
33	Many	666
34	Together	657
35	Good night	652
36	Always	649
37	Such	636
38	Already	635
39	Start	626
40	Like	614
41	Kid	607
42	A little	598

XSL•FO
RenderX

No.	Text	Frequency
43	Big	578
44	Happiness	569
45	Question	548
46	Activity	547
47	And	535
48	Can’t	531
49	Friend	507
50	Don’t know	496
51	Tomorrow	491
52	Fighting	490
53	See	470
54	Need	468
55	One day	468
56	Although	464
57	Know	463
58	Withstand	460
59	Emotion	447
60	Life	438

In this meaning-making process, there are initial attributions and causal understandings, followed by the emergence of new attributions in ongoing meaning-making. We found that recalling the past and contemplating the future are accompanied by a process of meaning-making. For example, “Thanks to my past self for staying true to myself when the pressure was on. Thanks for the self that still wipes away the tears and memorizes after every cry. Every step I took, every tear I left behind, the pain I felt when I was suffering. All are the marks of youth. In the future, I will thank myself for struggling now.” “Recently, I have been under a lot of pressure, and I have drawn up a distant travel plan for myself. Winter vacation to Finland to see reindeer, and want to drink hot horse milk. Thinking about it makes me feel healed to the point where I wish I could kill all the malice and love is worth it.” This example illustrates the meaning-making process triggered by mental time travel. Under pressure, people can make meaning by expressing themselves on blogs. For example, “There are many things in this world that can make you happy, one long-term and one short-term. It’s up to you, what you are, how much stress you can handle, and how much better and longer-term benefits you can enjoy.” “Recently, you needed to organize yourself, organize your emotions, organize your life, organize your work, organize your plans for the future, always have to do a good job of planning, in order to have peace of mind, there is pressure to continue to move forward, I believe that everything can be better and better.” “When you navigate the difficulties and obstacles alone, you will be grateful for the first time to say nothing and bite your teeth to insist on your own path.” “Still young, not everything in life is decided over a small thing, pressure will exist, but it won’t last forever or learn to be more optimistic and happier.” “No pressure of life will be empty, no pressure of youth will wither, no pressure of life will be bleak.”

Subsequently, to delve deeper into how the components of meaning-making in stress-related blog posts contribute to the positive adaptation to stress, the “Wenxin” system was used for textual analysis. This analysis involved extracting elements related to time, meaning-making, and positive or negative emotions. The frequencies and percentages of these elements were included in the subsequent mediation analyses. For the time elements, components representing the past, present, and future were selected. For meaning-making, words signifying attribution and causal understanding were chosen, such as causal words (eg, “because,” “therefore,” and “thus”) and insightful words (eg, “originally” and “comprehend”). Positive emotional words were counted based on their frequency. The descriptive statistics and correlation analyses for the two large samples are presented in Tables 2 and 3.

XSL•FO
RenderX

Table 2. Descriptive statistics and correlation analysis of main variables for sample 1 text analysis (n=165,374).

		Percentage values, mean (SD)	1	2	3	4	5	6
1	Rate of positive emotions words	0.10 (0.004)	1.00					
2	Rate of insightful words	0.06 (0.002)	0.07[a]	1.00				
3	Rate of causal words	0.03 (0.001)	0.03[a]	0.10[a]	1.00			
4	Rate of past time words	0.01 (0.002)	0.01[b]	0.03[a]	0.01[a]	1.00		
5	Rate of present time words	0.01 (0.001)	0.02[a]	0.03[a]	0.07	0.03[a]	1.00	
6	Rate of future-time words	0.01 (0.001)	0.03[a]	0.01[a]	–0.01	0.01[b]	0.03[a]	1.00

[a]*P*<.001.
[b]*P*<.01.

Table 3. Descriptive statistics and correlation analysis of main variables for sample 2 text analysis (n=166,337).

		Percentage value, mean (SD)	1	2	3	4	5	6
1	Rate of positive emotions words	0.10 (0.004)	1.00					
2	Rate of insightful words	0.06 (0.019)	.07[a]	1.00				
3	Rate of causal words	0.03 (0.001)	.02[a]	.10[a]	1.00			
4	Rate of past time words	0.01 (0.003)	.01	.02[a]	.02[a]	1.00		
5	Rate of present time words	0.01 (0.007)	.01[b]	.02[a]	.01[a]	.04[a]	1.00	
6	Rate of future-time words	0.01 (0.005)	.02[a]	.03[a]	–.01	.01	.04[a]	1.00

[a]*P*<.001.
[b]*P*<.01.

Correlation analyses of the 2 large samples revealed significant positive correlations between the frequency of positive emotion words and the frequency of causal (sample 1: r=0.03, P<.001; sample 2: r=0.02, P<.001) and insightful (both sample: r=0.07, P<.001) words (meaning-making). In sample 1, positive emotion words showed a significant positive correlation with the rates of future-time words (r=0.03; P<.001), past-time words (r=0.01; P=.005), and present-time words (r=0.02; P<.001). In sample 2, there was a significant positive correlation between positive emotion words and the frequencies of both present (r=0.01; P=.005) and future words (r=0.02; P<.001). The frequencies of causal (sample 1: r=0.01, P<.001; sample 2: r=0.02, P<.001) and insightful words (sample 1: r=0.03; P<.001; sample 2: r=0.02, P<.001) were positively correlated with past words, but future words were only significantly positively correlated with insightful words (sample 1: r=0.01, P<.001; sample 2: r=0.03, P<.001) and not causal words (P>.01 in both samples). This aligns with the notion that attributions are more prevalent when reflecting on the past than the future.

Mediation models were established for both samples (Table 4).

In this study, the results of the homogeneity test were Q_1= 0.93 and P=.34, with an I^2 of 0 %, indicating that the criteria for homogeneity were met. Therefore, a fixed-effects model was chosen for the analysis.

XSL•FO
RenderX

Table 4. Mediation models.

Sample	Mental time travel	Meaning-making	Positive emotions	a path[a], B	b path[b], B	c path[c], B	Indirect effect	99% CI
1	Rate of future-time words	Rate of insightful words	Rate of positive emotions words	.04[d]	.17[d]	.23[d]	0.008	0.002 to 0.021
2	Rate of future-time words	Rate of insightful words	Rate of positive emotions words	.10[d]	.17[d]	.14[d]	0.017	0.007 to 0.037
1	Rate of future-time words	Rate of causal words	Rate of positive emotions words	–.01	.11[d]	.24[d]	–0.002	–0.005 to –0.001
2	Rate of future-time words	Rate of causal words	Rate of positive emotions words	–.01	.07[d]	.16[d]	–0.001	–0.002 to 0.001
1	Rate of past time words	Rate of insightful words	Rate of positive emotions words	.18[d]	.17[d]	.08[e]	0.030	0.016 to 0.052
2	Rate of past time words	Rate of insightful words	Rate of positive emotions words	.11[d]	.17[d]	–.01	0.019	0.007 to 0.041
1	Rate of past time words	Rate of causal words	Rate of positive emotions words	.05[d]	.11[d]	.10[f]	0.005	0.001 to 0.025
2	Rate of past time words	Rate of causal words	Rate of positive emotions words	.09[d]	.07[d]	.01	0.006	–0.001 to 0.025

[a]Indicates the path from mental time travel to meaning-making.

[b]Indicates the path from meaning-making to positive emotions.

[c]Indicates the path from mental time travel to positive emotions.

[d]*P*<.001.

[e]*P*<.05.

[f]*P*<.01.

After extracting each effect size to be included in the meta-analysis, the values were synthesized and weighted to calculate a combined average statistic. The calculation of the average effect size was performed by weighting each effect size according to the inverse of its variance (ϖi). This involved multiplying each effect size by its respective weight, summing the products, and dividing them by the sum of the weights. Next, the standard error of the mean was calculated using the square root of the sum of the inverse variance weights, . This standard error was used to determine 95% CIs. The combined effect size (B) obtained was .013 (SE 0.003; 95% CI 0.007-0.018; *P*<.001; Figure 2).

Figure 2. Forest plot illustrating the mediation effect of mental time travel-meaning-making on positive emotions.

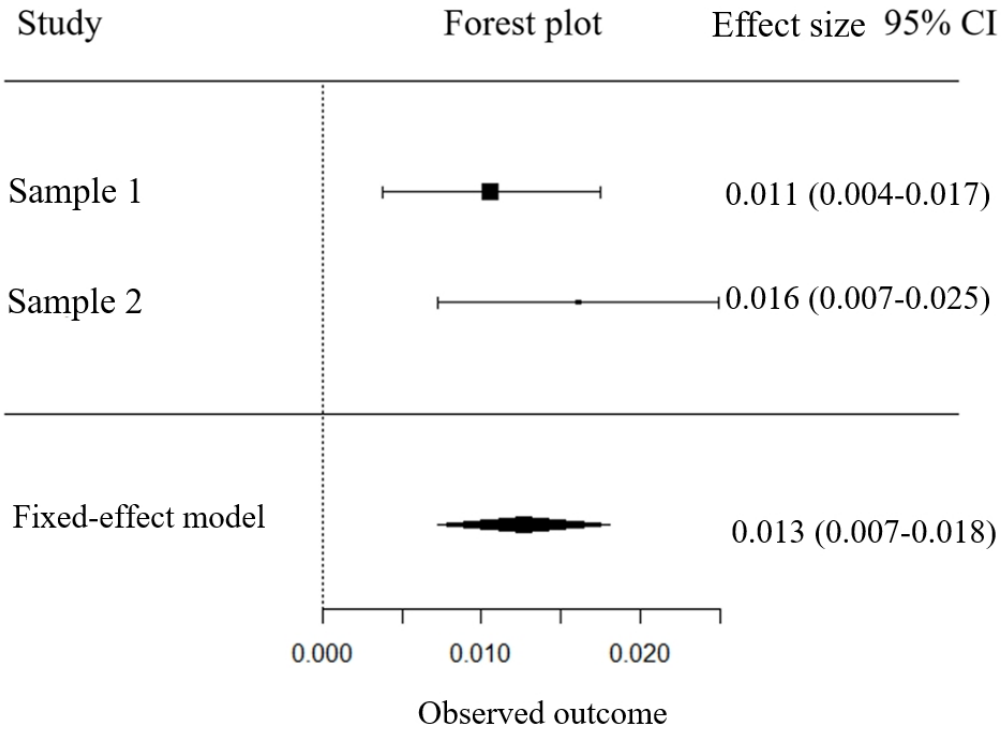

XSL•FO
RenderX

Discussion

Principal Findings

This study conducted a large-scale text analysis of stress-related Weibo posts to identify the coexistence of positive emotional words and stress. Meaning-making elements were also observed in these posts. This study found that meaning-making induced by mental time travel can promote positive emotions under stress. The presence of words related to the past and future reflects the mental travel time. Recognizing that life encompasses not only the present but also the past and future is a prerequisite for meaning-making through complete narratives. Therefore, the appearance of future and past words signifies a meaning-making process. Causal and insightful words representing the outcomes of meaning-making also appeared in the posts. This discovery is significant, as a wealth of research evidence indicates that mental time travel initiates the meaning-making process and concurrently influences shifts in one's sense of meaning in life [40-42]. Therefore, this analysis provides a nuanced understanding of how individuals engage and narrate their experiences of stress and coping within a digital context.

A mediation model was constructed using mental time travel words (mental time travel words) → meaning-making words (causal and insightful words) → positive poststress emotions. This model confirmed positive adaptation following stress. Mini–meta-analyses of two combined mediation models tested in 2 subsamples revealed a stable mediation effect across the 2 random subsamples. This demonstrates that meaning-making induced by mental time travel in stress-related blog posts can predict positive emotions under stress, supporting the cyclical model of stress resilience and stress-related growth rather than the cumulative stress model. This finding contributes to addressing the decade-long imbalance in stress research and clinical practice, which has almost exclusively focused on negative emotions while neglecting positive ones [61].

The Adaptive Function of Meaning-Making

The stress interaction model indicates that stress does not necessarily lead to impairment or maladaptation. An individual's diverse experiences in response to the same stressor primarily stem from their primary and secondary appraisals of the stressor, with meaning-making playing a critical role [62]. The meaning-making model further explains how meaning-making reduces the distress caused by a stressor and fosters growth by narrowing the discrepancy between situational and global meanings [27].

The cyclical model of stress resilience and stress-related growth posits that the initial state of individuals is determined by a combination of genetic, environmental, and social factors. When faced with stress, individuals adjust, leading to stress coping, which becomes a new initial state. During this adjustment process, certain coping strategies and personal resources can facilitate positive poststress psychological health outcomes, with meaning-making being an effective strategy [50]. Both meaning-making and cyclical models of stress resilience and growth indicate that coping strategies, such as meaning-making, can serve as mediating mechanisms leading to positive outcomes related to stress; hence, meaning-making can promote positive emotions under stress [27,50].

Research in the field of positive psychology has shown that meaning-making reliably predicts an individual's psychological health [63,64]. Extensive literature has demonstrated the adaptive function of meaning-making, showing a consistent positive correlation between meaning-making and positive psychological outcomes [61,65,66]. A meta-analysis of 87 cross-sectional studies found that meaning-making was associated with greater happiness and less depression [66]. In older adults living in nursing facilities, meaning-making was significantly correlated with positive emotions, personal growth, environmental mastery, and self-acceptance [67]. In a study of adolescent earthquake survivors, meaning-making was related to posttraumatic growth, which explains the relationship between meaning-focused coping and positive emotions [68].

Benefits of Mental Time Travel

Meaning-making involves the integration of an individual's life experiences, a process that includes narrative elements [27]. This means that individuals attempt to find common threads across their past, present, and future experiences, creating coherent and understandable cognitive representations or schema. Meaning-making requires the recognition that life transcends any specific moment, and mental time travel possesses these narrative elements. Mental time travel, also known as self-distancing or mental simulation, is the process of mentally transcending oneself. It involves occupying different times (past or future) or places or assuming the subjective experiences of different people or hypothetical realities. Mental time travel encompasses imagining experiences that extend beyond what is currently in progress. This capacity for simulation, which is particularly rich in form, seems unique to humans [69]. Some scholars have speculated that simulation enables humans to engage in complex cultures by navigating past, future, and social worlds [70]. A tendency toward nostalgia can mitigate the impact of existential threats on one's sense of meaning and increase positive emotions [40,41]. Additionally, experimentally induced nostalgia can enhance self-reported meaning in life and bolster positive emotions [71,72]. Similar to positive contemplation of the past through nostalgia, positive thinking about the future corresponded to enhanced meaning. The frequency of thoughts about future legacy is also related to meaning in life [73-75].

In this study, the correlational and mediation analyses found that future-time words were only related to the frequency of insightful words, not causal words. This suggests that meaning-making about the future involves causal reasoning less often and insightful reasoning more often. The constructive episodic simulation hypothesis posits that future mental time travel focuses more on the completeness of event narratives, thus favoring insightful reasoning [76]. Simultaneously, we also found that the direct effect of past mental time travel on positive emotions varied in both samples. This suggests that the meaning-making process plays a very important role (Park [27] and Ord et al [50]). Given the key role of mental time travel in meaning-making and promoting positive emotions, further research is needed to explore this mechanism [77].

XSL•FO
RenderX

Theoretical and Practical Implications

This study explored the adaptive function of meaning-making in stressful situations by integrating a cyclical model of stress resilience and stress-related growth with the meaning-making model [27,50]. This validates the mediating role of meaning-making between stress and positive emotions. Under stress, individuals engage in positive reappraisal to construct meaning from stressful events. Subsequently, through mental time travel, seeking and assigning meaning, individuals promote positive emotions under stress.

Second, this research introduces the method of mini–meta-analysis in a large-sample study to investigate the establishment of mediation effects, contributing to resolving the replication crisis. Given the vast amount of data obtained from Weibo crawlers, large sample sizes can lead to false positives. Therefore, a combined meta-analysis of multiple effect sizes in the same study would be particularly useful. Regardless of the number of studies conducted, researchers can succinctly summarize the results of each study through a meta-analysis. Researchers suggested that journal peer reviewers and editors should encourage authors to perform mini–meta-analyses on their own research results to determine the stability and consistency of the study effects and that conducting meta-analyses on papers can improve the precision of estimates (ie, narrow confidence intervals) [78,79]. Conducting mini–meta-analyses of a few studies can help researchers understand and strengthen their conclusions.

Third, from a practical perspective, this study underscores the need to recognize the positive aspects of stress and prevent the demonization of stress in the “stress is debilitating” narrative. Because stress cannot be eradicated or entirely avoided, finding ways to coexist with and grow from a psychological standpoint is crucial for long-term strategies. Mental time travel, an effective way to elicit meaning-making, can be developed as an intervention method to promote better coping with stress in individuals [80-82].

Limitations and Future Directions

This study had certain limitations. First, the cyclical model of stress resilience and stress-related growth discusses how appraisals, effective coping strategies, and personal resources modulate an individual’s stress response [50]. However, existing research has only explored meaning-making as an effective strategy. Future research should examine whether and how other coping strategies work in conjunction with meaning-making to influence positive poststress emotions.

Second, the expressions of stress, meaning-making, and positive emotions explored in this study’s Weibo analysis were essentially correlated. However, the model established was theory-based, and the data were not accessed at different time points. Future research should benefit from more laboratory designs that manipulate mental time travel to observe changes in meaning-making and positive emotions. Alternatively, momentary ecological assessment methods in daily settings and longitudinal studies can be used to obtain robust results.

Finally, the focus of this study was on whether mental time travel words can predict positive emotion words through meaning-making words; however, no distinction was made between categories of positive emotion words. Moreover, big data studies face the challenge of being unable to accurately manipulate or measure subtle differences in meaning while acquiring massive amounts of data [83]. Therefore, more refined studies should be designed in the future to validate the findings of big data studies and provide more reliable evidence.

Conclusions

This study concludes that individuals can adapt positively to stress by engaging in meaning-making processes that are triggered by mental time travel and reflected in their social media posts. The study’s mediation model confirmed that mental time travel leads to meaning-making, which fosters positive emotional responses to stress. This finding is significant because it demonstrates the potential of mental time travel as a psychological strategy to facilitate positive adaptation to stressful situations.

Acknowledgments

Preparation of this manuscript was supported by the National Natural Science Foundation of China (grant 32171076), the Humanities and Social Sciences Youth Foundation, the Ministry of Education (grant 23YJC190002), and the China Postdoctoral Science Foundation (grant 2023M740281).

Data Availability

The datasets generated during or analyzed during this study are available in the (Figshare) repository [84].

Authors' Contributions

YC conceptualized the study, developed the methodology, performed formal analysis, conducted investigations and visualization, wrote the original draft, wrote and reviewed the revised version, and acquired funding. YG conceptualized the study, validated the findings, supervised the research, wrote the original draft, wrote and reviewed the revised version, and acquired funding. JM and HZ conducted investigations and managed project administration. LZ supervised the project, wrote the original draft, and edited the manuscript. All authors reviewed and agreed to the published version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File , 24 KB - jmir_v27i1e63407_app1.docx]

References

1. Edwards VJ, Holden GW, Felitti VJ, Anda RF. Relationship between multiple forms of childhood maltreatment and adult mental health in community respondents: results from the adverse childhood experiences study. Am J Psychiatry 2003;160(8):1453-1460. [doi: 10.1176/appi.ajp.160.8.1453] [Medline: 12900308]
2. Emery RE, Laumann-Billings L. An overview of the nature, causes, and consequences of abusive family relationships. Toward differentiating maltreatment and violence. Am Psychol 1998;53(2):121-135. [doi: 10.1037//0003-066x.53.2.121] [Medline: 9491743]
3. Joseph NT, Jiang Y, Zilioli S. Momentary emotions and salivary cortisol: a systematic review and meta-analysis of ecological momentary assessment studies. Neurosci Biobehav Rev 2021;125:365-379. [doi: 10.1016/j.neubiorev.2021.02.042] [Medline: 33662445]
4. de Kloet ER, Joëls M, Holsboer F. Stress and the brain: from adaptation to disease. Nat Rev Neurosci 2005;6(6):463-475. [doi: 10.1038/nrn1683] [Medline: 15891777]
5. McEwen BS, Bowles NP, Gray JD, Hill MN, Hunter RG, Karatsoreos IN, et al. Mechanisms of stress in the brain. Nat Neurosci 2015;18(10):1353-1363 [FREE Full text] [doi: 10.1038/nn.4086] [Medline: 26404710]
6. Tottenham N, Sheridan MA. A review of adversity, the amygdala and the hippocampus: a consideration of developmental timing. Front Hum Neurosci 2010;3:68 [FREE Full text] [doi: 10.3389/neuro.09.068.2009] [Medline: 20161700]
7. Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nat Rev Neurosci 2009;10(6):434-445. [doi: 10.1038/nrn2639] [Medline: 19401723]
8. Pechtel P, Pizzagalli DA. Effects of early life stress on cognitive and affective function: an integrated review of human literature. Psychopharmacology (Berl) 2011;214(1):55-70 [FREE Full text] [doi: 10.1007/s00213-010-2009-2] [Medline: 20865251]
9. McEwen BS, Gianaros PJ. Stress- and allostasis-induced brain plasticity. Annu Rev Med 2011;62:431-445 [FREE Full text] [doi: 10.1146/annurev-med-052209-100430] [Medline: 20707675]
10. Mineka S, Zinbarg R. A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. Am Psychol 2006;61(1):10-26. [doi: 10.1037/0003-066X.61.1.10] [Medline: 16435973]
11. Seery MD, Holman EA, Silver RC. Whatever does not kill us: cumulative lifetime adversity, vulnerability, and resilience. J Pers Soc Psychol 2010;99(6):1025-1041 [FREE Full text] [doi: 10.1037/a0021344] [Medline: 20939649]
12. Galatzer-Levy IR, Huang SH, Bonanno GA. Trajectories of resilience and dysfunction following potential trauma: a review and statistical evaluation. Clin Psychol Rev 2018;63:41-55. [doi: 10.1016/j.cpr.2018.05.008] [Medline: 29902711]
13. Malhi GS, Das P, Bell E, Mattingly G, Mannie Z. Modelling resilience in adolescence and adversity: a novel framework to inform research and practice. Transl Psychiatry 2019;9(1):316 [FREE Full text] [doi: 10.1038/s41398-019-0651-y] [Medline: 31772187]
14. Crane MF, Searle BJ, Kangas M, Nwiran Y. How resilience is strengthened by exposure to stressors: the systematic self-reflection model of resilience strengthening. Anxiety Stress Coping 2019;32(1):1-17. [doi: 10.1080/10615806.2018.1506640] [Medline: 30067067]
15. Dooley LN, Slavich GM, Moreno PI, Bower JE. Strength through adversity: moderate lifetime stress exposure is associated with psychological resilience in breast cancer survivors. Stress Health 2017;33(5):549-557 [FREE Full text] [doi: 10.1002/smi.2739] [Medline: 28052491]
16. Shakespeare-Finch J, Lurie-Beck J. A meta-analytic clarification of the relationship between posttraumatic growth and symptoms of posttraumatic distress disorder. J Anxiety Disord 2014;28(2):223-229. [doi: 10.1016/j.janxdis.2013.10.005] [Medline: 24291397]
17. Paquola C, Bennett MR, Hatton SN, Hermens DF, Lagopoulos J. Utility of the cumulative stress and mismatch hypotheses in understanding the neurobiological impacts of childhood abuse and recent stress in youth with emerging mental disorder. Hum Brain Mapp 2017;38(5):2709-2721 [FREE Full text] [doi: 10.1002/hbm.23554] [Medline: 28256777]
18. Cathomas F, Murrough JW, Nestler EJ, Han M, Russo SJ. Neurobiology of resilience: interface between mind and body. Biol Psychiatry 2019;86(6):410-420 [FREE Full text] [doi: 10.1016/j.biopsych.2019.04.011] [Medline: 31178098]
19. Ioannidis K, Askelund AD, Kievit RA, van Harmelen A. The complex neurobiology of resilient functioning after childhood maltreatment. BMC Med 2020;18(1):32 [FREE Full text] [doi: 10.1186/s12916-020-1490-7] [Medline: 32050974]
20. Feder A, Fred-Torres S, Southwick SM, Charney DS. The biology of human resilience: opportunities for enhancing resilience across the life span. Biol Psychiatry 2019;86(6):443-453. [doi: 10.1016/j.biopsych.2019.07.012] [Medline: 31466561]

XSL•FO
RenderX

21. Boykin DM, Anyanwu J, Calvin K, Orcutt HK. The moderating effect of psychological flexibility on event centrality in determining trauma outcomes. Psychol Trauma 2020;12(2):193-199. [doi: 10.1037/tra0000490] [Medline: 31282720]
22. Stanisławski K. The coping circumplex model: an integrative model of the structure of coping with stress. Front Psychol 2019;10:694 [FREE Full text] [doi: 10.3389/fpsyg.2019.00694] [Medline: 31040802]
23. Brooks M, Graham-Kevan N, Lowe M, Robinson S. Rumination, event centrality, and perceived control as predictors of post-traumatic growth and distress: the cognitive growth and stress model. Br J Clin Psychol 2017;56(3):286-302. [doi: 10.1111/bjc.12138] [Medline: 28464228]
24. Yao ZF, Hsieh S. Neurocognitive mechanism of human resilience: a conceptual framework and empirical review. Int J Environ Res Public Health 2019;16(24):5123 [FREE Full text] [doi: 10.3390/ijerph16245123] [Medline: 31847467]
25. Park CL, Folkman S. Meaning in the context of stress and coping. Rev Gen Psychol 1997;1(2):115-144. [doi: 10.1037//1089-2680.1.2.115]
26. Gan Y, Guo M, Tong J. Scale development of meaning-focused coping. J Loss Trauma 2013;18(1):10-26. [doi: 10.1080/15325024.2012.678780]
27. Park CL. Making sense of the meaning literature: an integrative review of meaning making and its effects on adjustment to stressful life events. Psychol Bull 2010;136(2):257-301. [doi: 10.1037/a0018301] [Medline: 20192563]
28. Collie K, Long BC. Considering 'meaning' in the context of breast cancer. J Health Psychol 2005;10(6):843-853. [doi: 10.1177/1359105305057318] [Medline: 16176961]
29. Gillies J, Neimeyer RA. Loss, grief, and the search for significance: toward a model of meaning reconstruction in bereavement. J Constr Psychol 2006;19(1):31-65. [doi: 10.1080/10720530500311182]
30. Lee V, Cohen SR, Edgar L, Laizner AM, Gagnon AJ. Clarifying "meaning" in the context of cancer research: a systematic literature review. Palliat Support Care 2004;2(3):291-303. [doi: 10.1017/s1478951504040386] [Medline: 16594414]
31. O'Connor MF. Making meaning of life events: theory, evidence, and research directions for an alternative model. Omega (Westport) 2003;46(1):51-75. [doi: 10.2190/0ckd-pvq0-t260-ntxu]
32. Skaggs BG, Barron CR. Searching for meaning in negative events: concept analysis. J Adv Nurs 2006;53(5):559-570. [doi: 10.1111/j.1365-2648.2006.03761.x] [Medline: 16499677]
33. Ye J, Qin X, Cui J, Jia L, Shi H, Yang T, et al. Mental time travel for self and others in individuals with a high level of schizotypy. Psych J 2023;12(4):524-534. [doi: 10.1002/pchj.626] [Medline: 36653195]
34. Suddendorf T, Corballis MC. Mental time travel and the evolution of the human mind. Genet Soc Gen Psychol Monogr 1997;123(2):133-167. [Medline: 9204544]
35. Tulving E. Episodic memory: from mind to brain. Annu Rev Psychol 2002;53:1-25. [doi: 10.1146/annurev.psych.53.100901.135114] [Medline: 11752477]
36. Liberman N, Förster J. The effect of psychological distance on perceptual level of construal. Cogn Sci 2009;33(7):1330-1341 [FREE Full text] [doi: 10.1111/j.1551-6709.2009.01061.x] [Medline: 21585508]
37. Baumeister RF, Landau MJ. Finding the meaning of meaning: emerging insights on four grand questions. Rev Gen Psychol 2018;22(1):1-10. [doi: 10.1037/gpr0000145]
38. George LS, Park CL. Meaning in life as comprehension, purpose, and mattering: toward integration and new research questions. Rev Gen Psychol 2016;20(3):205-220. [doi: 10.1037/gpr0000077]
39. King LA, Hicks JA. The science of meaning in life. Annu Rev Psychol 2021;72:561-584. [doi: 10.1146/annurev-psych-072420-122921] [Medline: 32898466]
40. Juhl J, Routledge C, Arndt J, Sedikides C, Wildschut T. Fighting the future with the past: nostalgia buffers existential threat. J Res Pers 2010;44(3):309-314. [doi: 10.1016/j.jrp.2010.02.006]
41. Routledge C, Arndt J, Sedikides C, Wildschut T. A blast from the past: the terror management function of nostalgia. J Exp Soc Psychol 2008;44(1):132-140. [doi: 10.1016/j.jesp.2006.11.001]
42. Routledge C, Sedikides C, Wildschut T, Juhl J. Finding meaning in one's past: nostalgia as an existential resource. In: The Psychology of Meaning. Washington, DC, US: American Psychological Association; 2013:297-316.
43. Bluck S, Glück J. Making things better and learning a lesson: experiencing wisdom across the lifespan. J Pers 2004;72(3):543-572. [doi: 10.1111/j.0022-3506.2004.00272.x] [Medline: 15102038]
44. Heintzelman SJ, King LA. Meaning in life and intuition. J Pers Soc Psychol 2016;110(3):477-492. [doi: 10.1037/pspp0000062] [Medline: 26322984]
45. Wang Y, Lippke S, Miao M, Gan Y. Restoring meaning in life by meaning-focused coping: the role of self-distancing. Psych J 2019;8(3):386-396. [doi: 10.1002/pchj.296] [Medline: 31264391]
46. Waytz A, Hershfield HE, Tamir DI. Mental simulation and meaning in life. J Pers Soc Psychol 2015;108(2):336-355 [FREE Full text] [doi: 10.1037/a0038322] [Medline: 25603379]
47. Chen Y, Ma J, Zhu H, Peng H, Gan Y. The mediating role of default mode network during meaning-making aroused by mental simulation between stressful events and stress-related growth: a task fMRI study. Behav Brain Funct 2023;19(1):12 [FREE Full text] [doi: 10.1186/s12993-023-00214-x] [Medline: 37454095]
48. Kaplan AM, Haenlein M. Users of the world, unite! the challenges and opportunities of social media. Bus Horiz 2010;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]

XSL•FO
RenderX

49. Yuan EJ, Feng M, Danowski JA. "Privacy" in semantic networks on Chinese social media: the case of Sina Weibo. J Commun 2013;63(6):1011-1031. [doi: 10.1111/jcom.12058]
50. Ord AS, Stranahan KR, Hurley RA, Taber KH. Stress-related growth: building a more resilient brain. J Neuropsychiatry Clin Neurosci 2020;32(3):A4-212. [doi: 10.1176/appi.neuropsych.20050111] [Medline: 32729793]
51. Clapp JD, Patton SC, Beck JG. Expressive inhibition in response to stress: implications for emotional processing following trauma. J Anxiety Disord 2015;29:109-118 [FREE Full text] [doi: 10.1016/j.janxdis.2014.11.008] [Medline: 25576773]
52. Guo L. The delayed, durable effect of expressive writing on depression, anxiety and stress: a meta-analytic review of studies with long-term follow-ups. Br J Clin Psychol 2023;62(1):272-297. [doi: 10.1111/bjc.12408] [Medline: 36536513]
53. Zheng L, Lu Q, Gan Y. Effects of expressive writing and use of cognitive words on meaning making and post-traumatic growth. J Pac Rim Psychol 2019;13:297-316. [doi: 10.1017/prp.2018.31]
54. Open Science Collaboration. Estimating the reproducibility of psychological science. Science 2015;349(6251):aac4716. [doi: 10.1126/science.aac4716] [Medline: 26315443]
55. Goh JX, Hall JA, Rosenthal R. Mini meta-analysis of your own studies: some arguments on why and a primer on how. Soc Pers Psychol Compass 2016;10(10):535-549. [doi: 10.1111/spc3.12267]
56. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing simplified Chinese psychological linguistic analysis dictionary for microblog. In: Imamura K, Usui S, Shirao T, Kasamatsu T, Schwabe L, Zhong N, editors. Brain and Health Informatics. BHI 2013. Lecture Notes in Computer Science, vol 8211. Switzerland: Springer, Cham; 2013:359-368.
57. Koul A, Becchio C, Cavallo A. Cross-validation approaches for replicability in psychology. Front Psychol 2018;9:1117 [FREE Full text] [doi: 10.3389/fpsyg.2018.01117] [Medline: 30034353]
58. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008;40(3):879-891. [doi: 10.3758/brm.40.3.879] [Medline: 18697684]
59. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull 2003;29(5):665-675. [doi: 10.1177/0146167203029005010] [Medline: 15272998]
60. Sedgwick P. Multiple significance tests: the Bonferroni correction. BMJ 2012;344(7841):e509-e509. [doi: 10.1136/bmj.e509]
61. Folkman S. The case for positive emotions in the stress process. Anxiety Stress Coping 2008;21(1):3-14. [doi: 10.1080/10615800701740457] [Medline: 18027121]
62. Folkman S, Moskowitz JT. Positive affect and meaning-focused coping during significant psychological stress. In: Hewstone M, Schut HAW, De Wit JBF, Van Den Bos K, Stroebe MS, editors. The Scope of Social Psychology: Theory and Applications. Hove, United Kingdom: Psychology Press; 2007:193-208.
63. Volkert J, Schulz H, Brütt AL, Andreas S. Meaning in life: relationship to clinical diagnosis and psychotherapy outcome. J Clin Psychol 2013;70(6):528-535. [doi: 10.1002/jclp.22053]
64. Kleiman EM, Adams LM, Kashdan TB, Riskind JH. Gratitude and grit indirectly reduce risk of suicidal ideations by enhancing meaning in life: evidence for a mediated moderation model. J. Res. Pers 2013;47(5):539-546. [doi: 10.1016/j.jrp.2013.04.007]
65. Dieker J, Qualls SH. Meaning-Focused Coping in Dementia Caregiving: Predicting Positive and Negative Outcomes. Ann Arbor: University of Colorado Colorado Springs; 2021.
66. Helgeson VS, Reynolds KA, Tomich PL. A meta-analytic review of benefit finding and growth. J Consult Clin Psychol 2006;74(5):797-816. [doi: 10.1037/0022-006X.74.5.797] [Medline: 17032085]
67. Danhauer SC, Carlson CR, Andrykowski MA. Positive psychosocial functioning in later life: use of meaning-based coping strategies by nursing home residents. J Appl Gerontol 2005;24(4):299-318. [doi: 10.1177/0733464805277754]
68. Guo M, Gan Y, Tong J. The role of meaning-focused coping in significant loss. Anxiety Stress Coping 2013;26(1):87-102. [doi: 10.1080/10615806.2011.627507] [Medline: 22091795]
69. Gilbert DT. Stumbling on Happiness. 1st Edition. New York: HarperPerennial; 2006.
70. Baumeister RF, Masicampo EJ. Conscious thought is for facilitating social and cultural interactions: how mental simulations serve the animal-culture interface. Psychol Rev 2010;117(3):945-971. [doi: 10.1037/a0019393] [Medline: 20658859]
71. Routledge C, Arndt J, Wildschut T, Sedikides C, Hart CM, Juhl J, et al. The past makes the present meaningful: nostalgia as an existential resource. J Pers Soc Psychol 2011;101(3):638-652. [doi: 10.1037/a0024292] [Medline: 21787094]
72. Routledge C, Wildschut T, Sedikides C, Juhl J, Arndt J. The power of the past: nostalgia as a meaning-making resource. Memory 2012;20(5):452-460. [doi: 10.1080/09658211.2012.677452] [Medline: 22639901]
73. Fox MJ, Tost LP, Wade-Benzoni KA. The legacy motive: a catalyst for sustainable decision making in organizations. Bus Ethics Q 2015;20(2):153-185. [doi: 10.5840/beq201020214]
74. Wade-Benzoni KA, Tost LP. The egoism and altruism of intergenerational behavior. Pers Soc Psychol Rev 2009;13(3):165-193. [doi: 10.1177/1088868309339317] [Medline: 19571118]
75. Wade-Benzoni KA, Tost LP, Hernandez M, Larrick RP. It's only a matter of time: death, legacies, and intergenerational decisions. Psychol Sci 2012;23(7):704-709. [doi: 10.1177/0956797612443967] [Medline: 22692338]
76. Brocas I, Carrillo JD. A neuroeconomic theory of mental time travel. Front Neurosci 2018;12:658 [FREE Full text] [doi: 10.3389/fnins.2018.00658] [Medline: 30319339]
77. Schacter DL, Madore KP. Remembering the past and imagining the future: identifying and enhancing the contribution of episodic memory. Mem Stud 2016;9(3):245-255 [FREE Full text] [doi: 10.1177/1750698016645230] [Medline: 28163775]

XSL•FO
RenderX

78. Maner JK. Let's put our money where our mouth is: if authors are to change their ways, reviewers (and editors) must change with them. Perspect Psychol Sci 2014;9(3):343-351. [doi: 10.1177/1745691614528215] [Medline: 26173269]
79. Cumming G. The new statistics: why and how. Psychol Sci 2014;25(1):7-29. [doi: 10.1177/0956797613504966] [Medline: 24220629]
80. Kalisch R, Müller MB, Tüscher O. A conceptual framework for the neurobiological study of resilience. Behav Brain Sci 2015;38:e92. [doi: 10.1017/S0140525X1400082X] [Medline: 25158686]
81. Horn SR, Charney DS, Feder A. Understanding resilience: new approaches for preventing and treating PTSD. Exp Neurol 2016;284(Pt B):119-132. [doi: 10.1016/j.expneurol.2016.07.002] [Medline: 27417856]
82. Nijhof SL, Vinkers CH, van Geelen SM, Duijff SN, Achterberg EJM, van der Net J, et al. Healthy play, better coping: the importance of play for the development of children in health and disease. Neurosci Biobehav Rev 2018;95:421-429 [FREE Full text] [doi: 10.1016/j.neubiorev.2018.09.024] [Medline: 30273634]
83. Fan J, Han F, Liu H. Challenges of big data analysis. Natl Sci Rev 2014;1(2):293-314 [FREE Full text] [doi: 10.1093/nsr/nwt032] [Medline: 25419469]
84. Chen Y. The mediating role of meaning-making in the relationship between mental time travel and positive emotions in stress-related Blogs: a big data text analysis research. Figshare dataset. 2025. URL: https://doi.org/10.6084/m9.figshare.28283117 [accessed 2025-02-13]

Abbreviations

LIWC: Linguistic Inquiry and Word Count

Edited by T de Azevedo Cardoso; submitted 19.06.24; peer-reviewed by AP Vit, Y Wang; comments to author 29.11.24; revised version received 16.12.24; accepted 29.01.25; published 21.02.25.

Please cite as:
Chen Y, Zheng L, Ma J, Zhu H, Gan Y
The Mediating Role of Meaning-Making in the Relationship Between Mental Time Travel and Positive Emotions in Stress-Related Blogs: Big Data Text Analysis Research
J Med Internet Res 2025;27:e63407
URL: https://www.jmir.org/2025/1/e63407
doi:10.2196/63407
PMID:39900590

©Yidi Chen, Lei Zheng, Jinjin Ma, Huanya Zhu, Yiqun Gan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Culturally Adapted Guided Internet-Based Cognitive Behavioral Therapy for Hong Kong People With Depressive Symptoms: Randomized Controlled Trial

Jia-Yan Pan[1], PhD; Jonas Rafi[2], PhD

[1]Department of Social Work, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Department of Psychology, Stockholm University, Stockholm, Sweden

Corresponding Author:
Jia-Yan Pan, PhD
Department of Social Work
Hong Kong Baptist University
AAB1026, 10/F, Academic and Administration Building
15 Baptist University Road, Baptist University Road Campus, Hong Kong Baptist University, Kowloon Tong, KLN
Hong Kong
China (Hong Kong)
Phone: 852 34116415
Email: jiayan@hkbu.edu.hk

Abstract

Background: A significant number of individuals with depression in Hong Kong remain undiagnosed and untreated through traditional face-to-face psychotherapy. Internet-based cognitive behavioral therapy (iCBT) has emerged as a tool to improve access to mental health services. However, iCBT remains underdeveloped in Chinese communities, particularly in Hong Kong.

Objective: This study aims to (1) develop and evaluate the effectiveness of a culturally and linguistically appropriate guided iCBT program for Hong Kong Chinese with depressive symptoms, and (2) explore their treatment adherence.

Methods: An 11-week guided iCBT program, “Confront and Navigate Depression Online” (CANDO), consisting of 8 online modules, was developed and implemented for Hong Kong residents. The program was accessible via either an online platform (web-based iCBT) or a smartphone app (app-based iCBT). A 3-arm randomized controlled trial was conducted, with participants recruited through open recruitment and referrals from 2 local non-governmental organizations. A total of 402 eligible participants with mild to moderate depressive symptoms were randomly allocated into 3 groups: (1) web-based iCBT (n=139); (2) app-based iCBT (n=131); and (3) waitlist control (WLC) group (n=132), who transitioned to the web-based iCBT group after waiting for 11 weeks. Therapist support was provided by a clinical psychologist through 3 counseling sessions and weekly assignment feedback. The primary outcomes were the Beck Depression Inventory-II (BDI-II) and the 9-item Patient Health Questionnaire (PHQ-9), while the secondary outcome measures included the 12-item General Health Questionnaire (GHQ-12), the Chinese Automatic Thoughts Questionnaire (CATQ), and the Chinese Affect Scale (CAS). These scales were administered at preintervention, postintervention, and at 3-month and 6-month follow-up assessments. Data analysis was conducted using linear mixed effects modeling, adhering to the intent-to-treat principle.

Results: Participants in both the web- and app-based iCBT groups reported significant improvements compared with the WLC group on all primary (P<.001) and secondary measures (P<.001 and P=.009) at posttreatment. Large between-group effect sizes were observed in the reduction of depressive symptoms, with Cohen's d value of 1.07 (95% CI 0.81-1.34) for the web-based group and 1.15 (95% CI 0.88-1.43) for the app-based group on the BDI-II. Additionally, the web- and app-based groups showed effect sizes of 0.78 (95% CI 0.52-1.04) and 0.95 (95% CI 0.63-1.27) on PHQ-9, respectively. Medium to large effect sizes were observed for secondary outcomes at posttreatment. These positive effects were maintained at both the 3- and 6-month follow-ups, with medium to large within-group effect sizes. The adherence rate in the 2 iCBT groups was 154 out of 270 (57.0%) for completing all 8 online modules and 152 out of 270 (56.3%) for attending all 4 counseling sessions (including an intake interview). The recovery rate, as measured by the BDI-II at posttreatment, was 35 out of 90 (39%) for the web-based group and 36 out of 91 (40%) for the app-based group, compared with 3 out of 112 (3%) in the WLC group.

Conclusions: Culturally and linguistically adapted iCBT is an effective and feasible treatment for Hong Kong Chinese with mild to moderate depressive symptoms, demonstrating a satisfactory recovery rate and treatment adherence. ICBT offers an

XSL•FO
RenderX

accessible and viable alternative to face-to-face interventions in Hong Kong. The Hong Kong government should allocate more resources and support the use of iCBT as a tool to treat people with depressive symptoms.

Trial Registration: ClinicalTrials.gov (NCT04388800); https://clinicaltrials.gov/study/NCT04388800

(J Med Internet Res 2025;27:e64303) doi:10.2196/64303

KEYWORDS

Internet-based cognitive behavioral therapy; depression; Chinese; Hong Kong, culturally adapted internet intervention

Introduction

Background

Depression is a widespread mental health issue that the World Health Organization (WHO) predicts will become the most burdensome globally by 2030 [1]. In Hong Kong, the prevalence of depression rose from 1.3% in 2011-2014 to 9.1% in 2019 among the general population [2]. Lam et al [3] reported a 13.3% prevalence of common mental disorders in Hong Kong, with mixed anxiety and depression being the most frequent diagnosis (6.9%). However, only 24%-26% of individuals with common mental disorders access mental health services [3]. Consequently, many people in Hong Kong, similar to those with mental health conditions in Western countries [4], remain undiagnosed and untreated. This is likely due to the stigma surrounding mental disorders, a shortage of mental health professionals, the high cost of private treatment, and long wait times for public services. The lack of treatment not only leads to poor mental health outcomes, such as the development of comorbid anxiety and substance use disorders, as well as an increased risk of suicide [5], but also manifests physically, including chronic pain [6], thereby reducing quality of life [5]. Furthermore, poor mental health outcomes have significant economic consequences, such as reduced productivity [7]. These factors highlight the critical importance of early diagnosis and treatment.

Internet-Based Cognitive Behavioral Therapy

Cognitive behavioral therapy (CBT) is the first-line treatment for depression. Andersson and Titov [8] demonstrated that internet-based CBT (iCBT) produces treatment effects comparable to traditional face-to-face CBT, while being more cost-effective. As Baumann et al [9] stated, “ICBT has a strong potential to balance demand and supply of CBT in unipolar depression by reducing therapist time per patient.”

Karyotaki et al [10] systematically reviewed 39 randomized controlled trials (RCTs) and found that both guided and unguided iCBT interventions are more effective than control treatments (treatment-as-usual and waitlist control [WLC]) in reducing depressive symptom severity in both the short and long term. Etzelmueller et al [11] conducted a meta-analysis demonstrating the acceptability and effectiveness of guided iCBT for treating depression in routine health care. While guided iCBT is more effective than unguided iCBT at posttreatment, no differences are observed at 6- or 12-month follow-ups [10].

Mobile therapy has gained prominence with the widespread adoption of smartphones, bringing mobile psychotherapy to the forefront. Studies indicate no significant differences in treatment effectiveness between guided smartphone-delivered and internet-based CBT programs [12]. However, smartphone use is more evenly distributed throughout the day, allowing iCBT to integrate more seamlessly into daily routines [12]. Given the high penetration rates of smartphones (97%) and computers (89%) in Hong Kong [13,14], developing an iCBT program accessible via both mobile apps and computers—and comparing their effectiveness—would be a valuable endeavor.

In recent years, particularly during the COVID-19 pandemic, iCBT has seen rapid development in Chinese societies. Studies have shown that culturally adapted iCBT significantly reduces symptoms of depression, generalized anxiety, psychological distress, rumination, and insomnia while improving self-efficacy among Chinese individuals, compared with WLCs [15-18]. Even in studies without WLCs, culturally adapted iCBT has proven effective in reducing symptoms of depression, generalized anxiety, psychological distress, perceived stress, sleep disturbances, and functional disability, while enhancing life satisfaction and energy levels for Chinese participants [19,20]. The positive effects of culturally adapted iCBT were sustained for 3 months [19], 6 months [18], and 12 months [15]. These studies used RCTs with either an active control group (eg, [21]) or a WLC group (eg, [16,18]). Most studies utilized mobile interventions (eg, [16-18]), while a few relied on web-based interventions [15,19], and 1 incorporated both [20]. To better address the diverse mental health needs of the Chinese population, further research on culturally adapted and tailored iCBT programs is needed.

Response rates of iCBT for depression, indicating a significant reduction in depressive symptoms, range from 37% to 56% in meta-analyses [10,22]. The response rate for guided iCBT (48%) is higher than that for unguided iCBT (37%) [10] and significantly higher than that of control groups [23]. Recovery rates of guided iCBT for depression, defined as the complete remission of depressive symptoms, have been reported to range from 30% to 55% at posttreatment and 41% to 60% at the 6- to 7-month follow-up [24,25]. Lin et al [16] found that the guided iCBT program for Chinese individuals with depression achieved a high response rate (58%) and remission rate (55%). Similarly, Ying et al [18] reported that 38.2% and 33.3% of Chinese participants in their iCBT study demonstrated clinically meaningful improvements in depressive symptoms at postintervention and at the 6-month follow-up, respectively.

Regarding treatment adherence to iCBT, van Ballegooijen et al [26] found that the completion rate of the entire guided iCBT program for depression was 65.1%, significantly lower than the rate for face-to-face CBT (84.7%). However, there was no significant difference in the number of completed sessions, with

XSL•FO
RenderX

83.9% in face-to-face CBT and 80.8% in guided iCBT [26]. Similarly, Karyotaki et al [10] reported in their meta-analysis an average adherence rate (number of completed sessions/total number of sessions) of 76% for guided iCBT and 54% for unguided iCBT. Individual empirical studies show that the adherence rate (percentage of completers) of guided iCBT for depression is higher among a Chinese population, ranging from 67.0% [21] to 81.4% [18], compared with their Western counterparts, whose adherence rates range from 36.4% [27] to 60% [28]. The average session completion rate in guided iCBT for Hong Kong Chinese ranges from 52.7% [21] to 83.8% [18]. Karyotaki et al [10] also reported that the average dropout rate was 25% for guided iCBT and 29% for unguided iCBT, compared with 19% for those on a WLC and 22% for those receiving treatment as usual. Additionally, Chinese participants in iCBT reported high levels of satisfaction during depression treatment (eg, [16,18]).

Cultural Adaptation of iCBT

There is evidence suggesting the potential benefits of cultural adaptation in digital mental health interventions, including iCBT, for improving a range of outcomes compared with WLC and treatment-as-usual control conditions [29]. However, direct comparisons of the effectiveness of culturally adapted versus unadapted internet-based interventions remain lacking. For Chinese participants, Ng and Wong [30] demonstrated that the short-term effects of culturally adapted CBT were stronger than those of unadapted CBT. Similarly, iCBT should be culturally adapted to address the specific cultural needs of Chinese clients, taking into account cultural factors that may influence its effectiveness and acceptability. The following considerations are important when promoting iCBT to Chinese clients. First, the program should use culturally appropriate language, including Mandarin or other Chinese dialects, as well as idioms and expressions, to enhance understanding and engagement. For instance, Wong [31] translated the term “cognitive distortion” into the Chinese phrase “Xi Xiang Xian Jing (思想陷阱)” and adapted various types of cognitive distortions into 4-character Chinese idioms. Second, cultural sensitivity and awareness should be incorporated into the program content to align with Chinese cultural norms, values, and experiences. For example, an iCBT program for Hong Kong Chinese university students integrated Chinese cultural values into its content [32]. Additionally, collectivism should be emphasized by addressing relationships with others, social role expectations, and family obligations, which are highly valued in Chinese culture [33]. Third, the stigma surrounding mental health issues often deters Chinese individuals from seeking face-to-face services [34]. ICBT programs should tackle this cultural stigma by providing psychoeducation on mental health and emphasizing the importance of help-seeking. Fourth, when designing an iCBT program for Chinese participants, incorporating more videos (eg, case studies and animations) and reducing text can foster greater interaction and engagement [32,33]. Finally, therapists in guided iCBT programs should possess cultural and linguistic competence in Chinese culture and language [32].

Undoubtedly, iCBT is an effective, feasible, and acceptable approach for treating depression among Chinese individuals. However, it remains underdeveloped in Chinese communities, particularly in Hong Kong. Moreover, whether the delivery medium (web- or app-based) influences treatment outcomes is still unclear. To address this, this study developed and tested the effectiveness of a culturally and linguistically appropriate guided iCBT program for Hong Kong people with mild to moderate depressive symptoms, utilizing both web- and app-based platforms. We hypothesized the following:

- Participants in the web- and app-based iCBT groups would have lower levels of depressive symptoms upon completion of the intervention compared with the WLC group.
- Participants in the web- and app-based iCBT groups would have lower levels of anxiety symptoms, psychological distress, and negative thoughts and emotions upon completion of the intervention compared with the WLC group.
- Participants in the web- and app-based iCBT groups would have higher levels of positive thoughts and emotions upon completion of the intervention compared with the WLC group.
- There would be no significant differences in intervention effects between the web- and app-based iCBT groups.
- The effects of the web- and app-based iCBT programs would be sustained at 3 and 6 months postintervention.

Methods

Participant Selection Criteria

Participants were selected based on the following criteria: (1) aged 18-70 years; (2) proficient in Chinese; (3) a 9-item Patient Health Questionnaire (PHQ-9) score between 5 and 19; (4) no suicidal risk in the past 3 months, defined as a score of “0” or “1” on Item 9 of the Beck Depression Inventory-II (BDI-II) and low suicidal risk as assessed during the intake interview; (5) not receiving any other psychological treatment for depression at the time of allocation; (6) no severe psychiatric conditions, such as bipolar disorder, as self-reported in the screening questionnaire and confirmed by the therapist during the intake interview; (7) access to a computer or smartphone with an internet connection; and (8) possession of a valid email address.

Recruitment and Screening

Participants were recruited from the general population in Hong Kong on an ongoing basis from April 2020 to January 2022. Recruitment methods included advertisements on social media platforms (Facebook and Instagram), Google Ads, a press conference, posters, leaflets, and referrals from 2 local partner nongovernmental organizations. Interested individuals registered at [35] by completing a screening questionnaire that included the PHQ-9, the BDI-II, and the Beck Anxiety Inventory (BAI). Initially, qualified participants were invited to a 1.5-hour intake interview conducted in-person or via Zoom (Zoom Communications/Qumu Corporation) with a clinical psychologist or an experienced social worker to further assess their suitability for participation, including a suicidal risk assessment. Marginal cases were discussed in a case meeting that included the first author (JYP), intake worker, and program therapist to assess participation eligibility. High-risk individuals with severe depression or suicidal risk or both were referred to

XSL•FO
RenderX

local mental health services for further diagnosis and treatment with their consent.

Research Design

A 3-arm RCT was conducted with a WLC group design. Randomization was performed using computer-generated random numbers by the therapist to allocate eligible participants to 1 of 3 groups: web-based iCBT, app-based iCBT, and WLC (1:1:1 allocation). Participants in the WLC group received web-based treatment immediately after the 2 iCBT groups completed their program. All 3 groups completed the same online questionnaire at pretest, posttest, and at 3- and 6-month follow-ups. The WLC group completed the questionnaire an additional time as a posttest after the 2 experimental groups had completed the program.

Ethical Considerations

An informed consent form was presented at the beginning of the questionnaire. After participants provided their consent to join the study, the online questionnaire was displayed to them. The RCT was conducted with a total of 13 cohorts from May 2020 to January 2023, with around 30 qualified participants in each cohort for implementation feasibility. The program evaluators were blinded to participant allocation. Ethics approval was obtained from the Research Ethics Committee of the university of the first author (JYP). The clinical trial was registered at ClinicalTrials.gov (NCT04388800).

Sample Size

A power analysis was conducted to calculate the required number of participants, assuming a medium effect size (*d*) of 0.5 [16], an α level of .05, and 95% power. The minimum required sample size was estimated to be around 210 participants. Given an assumed completion rate of 55%, the goal was to recruit approximately 400 participants.

Intervention: Confront and Navigate Depression Online (CANDO) Program

The "Confront and Navigate Depression Online" (CANDO) Program was an 11-week therapist-guided iCBT program designed for Hong Kong Chinese with mild to moderate depressive symptoms (Multimedia Appendix 1). The program was developed based on the principles of CBT as outlined by Beck [36], and adapted for Chinese clients (eg, [31,37]). It utilized a blended mode of service delivery, featuring 8 online modules and four 1-hour individual counseling sessions including 1 intake interview. Each module introduced different cognitive behavioral skills for coping with depression through self-developed videos. The themes of the 8 online modules are (1) understanding my depression within the CBT framework; (2) identifying automatic thoughts and cognitive distortions; (3) self-talk and positive statements; (4) behavioral experiments and behavioral activation; (5) problem-solving; (6) identifying and understanding cognitive rules; (7) relaxing dysfunctional cognitive rules; and (8) conclusion and relapse prevention. A variety of CBT skills were introduced and reinforced using animated videos and demonstrated through 5 local cases. Each online module began with a mood check for the past week using a 10-point scale and concluded with a session review and an assignment designed to help clients apply the skills they had learned to manage their depression. The modules were released to participants on a weekly basis, with counseling sessions scheduled in between for skills application. Additional program features included a client portfolio, online questionnaires, reminders, and an online booking system.

The development of the CANDO Program was guided by the cultural adaptation framework for psychosocial treatments proposed by Bernal et al [38]. The components listed in Textbox 1 were addressed to culturally adapt the program for Hong Kong Chinese.

Textbox 1. Components addressed to culturally adapt the program for Hong Kong Chinese.

1. Language

The entire program was developed and delivered in the local Cantonese language.

2. Metaphors

Various cognitive distortions were represented using 4-character Chinese idioms, such as "Fei Hei Ji Bai (非黑即白)" (all-or-nothing thinking), "Da Nan Lin Tou (大難臨頭)" (catastrophizing), and "Cai Duo Ren Yi (猜度人意)" (mind-reading) [31].

3. Content

The case demonstration videos were specifically tailored for Hong Kong Chinese people, drawing on and adapting numerous real-life cases within the Hong Kong context. These videos incorporated Chinese cultural values such as gender and social role expectations, face-saving, collectivism, and the high emphasis on academic achievement, addressing these themes across different life stages. For instance, an analysis of advantages and disadvantages was used to challenge and relax the cognitive rule of "男主外, 女主內" ("The husband is responsible for earning a living to support his family, while the wife is responsible for taking care of the family") into a more progressive belief that "there is no fixed stereotype of gender roles within a family as society advances." Another example is the use of a pie chart technique to redefine the concept of "a good mom" in Chinese society, broadening it to include multiple perspectives rather than focusing solely on helping children achieve academic excellence.

4. Persons

Therapists with the same ethnic and cultural background as the Hong Kong Chinese clients were selected for service delivery.

5. Method

The internet-based cognitive behavioral therapy delivery format was adapted to primarily use self-developed videos to teach and demonstrate various cognitive behavioral therapy skills.

XSL•FO
RenderX

The client platform offered both a web-based version, accessible via the program website, and a mobile app, available for download through Google Play (Google LLC/Alphabet, Inc.) or Apple Store (Apple, Inc.). After enrollment, clients were provided with an account to access the system. Clients with a PC web account could not log-in via the app, and those with an app account could not access the web-based version. Each client was assigned a therapist who provided therapeutic support through 3 counseling sessions conducted either face-to-face or online (via Zoom, WhatsApp call [Meta Platforms, Inc.], or telephone). These sessions included feedback on assignments, responses to internal messages, and online forum management. Both automated reminders and staff reminders were used to ensure participants remained engaged with the program.

To ensure data security and privacy, 2-factor authentication was implemented. This required users to log-in with a password preset by the user and a 1-time password sent to their email address. All collected electronic data were encrypted. The program was hosted on the server managed by the Information and Technology Office of the first author's (JYP) university. Before the program's launch, a user acceptance test was conducted to address bugs, and a security test was carried out by a third party to ensure system safety. Bugs reported by participants during the project were also addressed promptly throughout its implementation.

Therapist Credibility

The therapist was a clinical psychologist with a doctorate in clinical psychology and a master's degree in counseling. She had over 10 years of clinical experience in Hong Kong and the United States. The intake worker was a registered social worker in Hong Kong with bachelor's and master's degrees in social work, and more than 30 years of counseling experience in diverse settings. The intake interviews were conducted by both the therapist and the intake worker, while the initial screenings were performed by the therapist. All services were delivered in Cantonese, and both professionals were native Cantonese speakers.

Measurements

Depressive symptoms were assessed using both the BDI-II [39] and PHQ-9 [40]. For the BDI-II, scores of 0-9, 10-18, 19-29, and 30 or above indicate minimal, mild, moderate, and severe levels of depressive symptoms, respectively. For the PHQ-9, scores of 0-4, 5-9, 10-14, 15-19, and 20-27 correspond to minimal, mild, moderate, moderately severe, and severe depression, respectively. Anxiety symptoms were evaluated using the BAI [41]. In this study, the Cronbach α values were 0.866 (95% CI 0.844-0.884) for the BDI-II, 0.812 (95% CI 0.779-0.839) for the PHQ-9, and 0.909 (95% CI 0.890-0.923) for the BAI, demonstrating high internal consistency for all measures.

Psychological distress was measured using the Chinese version of the 12-item General Health Questionnaire (GHQ-12) [42,43]. The 0-0-1-1 scoring method was applied to calculate the total score, which ranged from 0 to 12, with higher scores indicating greater levels of psychological distress. A cut-off point of 1/2, validated for the Chinese population, was used to identify participants at risk of developing mental health problems [44]. In this study, the Cronbach α for the GHQ-12 was 0.866 (95% CI 0.844-0.884), indicating high internal consistency.

Automatic thoughts were measured using the Chinese Automatic Thoughts Questionnaire [45], which was developed from the Automatic Thoughts Questionnaire [46] and the Positive Automatic Thoughts subscale of the revised Automatic Thoughts Questionnaire [47]. Participants rated the frequency of each thought in the past week on a 5-point scale (1="Not at all" to 5="All the time"). Item scores were summed for the positive and negative subscales, with higher total scores indicating more positive and negative automatic thoughts, respectively. Cronbach α was 0.883 (95% CI 0.825-0.889) for the positive subscale and 0.88 (95% CI 0.859-0.899) for the negative subscale in this study.

Positive and negative emotions were assessed using the Chinese Affect Scale [48], which consists of 20 items with 2 subscales: positive affect (PA; 10 items) and negative affect (NA; 10 items). Participants rated the items on a 6-point Likert scale (1="not at all" to 6="extremely"). The scores for each subscale were summed, with higher scores indicating higher levels of PA and NA, respectively. The Cronbach α coefficients were 0.883 (95% CI 0.863-0.900) for the PA subscale and 0.876 (95% CI 0.853-0.894) for the NA subscale in this study.

Statistical Analysis

All analyses followed the CONSORT (Consolidated Standards of Reporting Trials) statement for RCTs [49]. Assignment completion and dropout were analyzed using a chi-square test and logistic regression. Missing data were imputed with m=5 data sets using multiple imputations with chained equations and predictive mean matching [50]. Analyses were conducted according to the intent-to-treat approach and performed on imputed data sets before pooling the coefficients. Analyses were conducted within a linear mixed model framework [51]. Outcomes at baseline, age, gender, and medication use were included as covariates. Sensitivity analyses using completer-only data were also conducted. When comparing the active treatments (web-based iCBT vs app-based iCBT) at posttreatment, and at 3 and 6 months, models included treatment, time, and the interaction between time and treatment. The within-group effects of transitioning from WLC to active treatments were modeled in a separate analysis, using the second premeasurement as a baseline (see Table S4 in Multimedia Appendix 2). All models used restricted maximum likelihood estimation and an unstructured covariance matrix. Chi-square tests and logistic regression were conducted to explore differences in baseline characteristics between those who completed all modules and those who did not, as well as between those who completed postmeasurements and those who did not. Effect sizes were calculated using Cohen *d* [52] with the emmeans package [53], following the standard convention for magnitudes of effect sizes: small (d=0.2), medium (d=0.5), and large (d≥0.8). Data analyses were performed in R version 4.3.1 (R Foundation) [54].

Clinical significance was calculated using the Jacobson-Truax method [55]. "Recovered" was defined as a positive change that is both reliable (ie, not due to measurement error) and clinically

XSL•FO
RenderX

significant (ie, an improvement of 2 SDs from the baseline sample mean); "improved" was defined as a reliable improvement, but not clinically significant; "unchanged" was defined as no reliable change; and "deteriorated" was defined as a reliable worsening. Following van Ballegooijen et al's [26] suggestion, those who were given an intervention but did not commence treatment were also included in the calculation of the adherence rate. Adherence was analyzed separately for treatment modules and counseling sessions. Treatment adherence was defined as completing all 8 treatment modules, and counseling adherence was defined as participating in all 4 counseling sessions. The calculation was based on the data at the posttest.

Results

Participant Flow

A total of 2114 participants registered for the CANDO Program's online screening, and 1374 were invited for an intake interview for further assessment. Of these, 698 intake interviews were conducted, and ultimately, 402 participants who met the inclusion criteria were randomly assigned to 1 of the 3 groups: web-based iCBT (n=139), app-based iCBT (n=131), and WLC (n=132). All allocated participants, except for 1 in the WLC group, completed a pretest questionnaire and were included in the intent-to-treat analysis; 2, 4, and 16 eligible participants did not log into the CANDO system after being assigned a client account in the web-based iCBT, app-based iCBT, and WLC groups, respectively, resulting in 137, 127, and 116 participants in the corresponding intervention conditions. At the posttest, 90, 91, and 112 participants provided data in the web-based iCBT, app-based iCBT, and WLC groups, respectively. A total of 116 participants from the WLC group later joined the web-based iCBT program, with 63, 30, and 33 participants providing data at posttreatment, 3-month, and 6-month follow-up tests, respectively. For the web-based/app-based iCBT groups, 54/59 and 53/49 participants provided data at the 3- and 6-month follow-up tests, respectively. See Figure 1 for the participant flowchart. Also see Multimedia Appendix 3.

XSL•FO
RenderX

Figure 1. Participant flow of the study. iCBT: internet-based cognitive behavioral therapy.

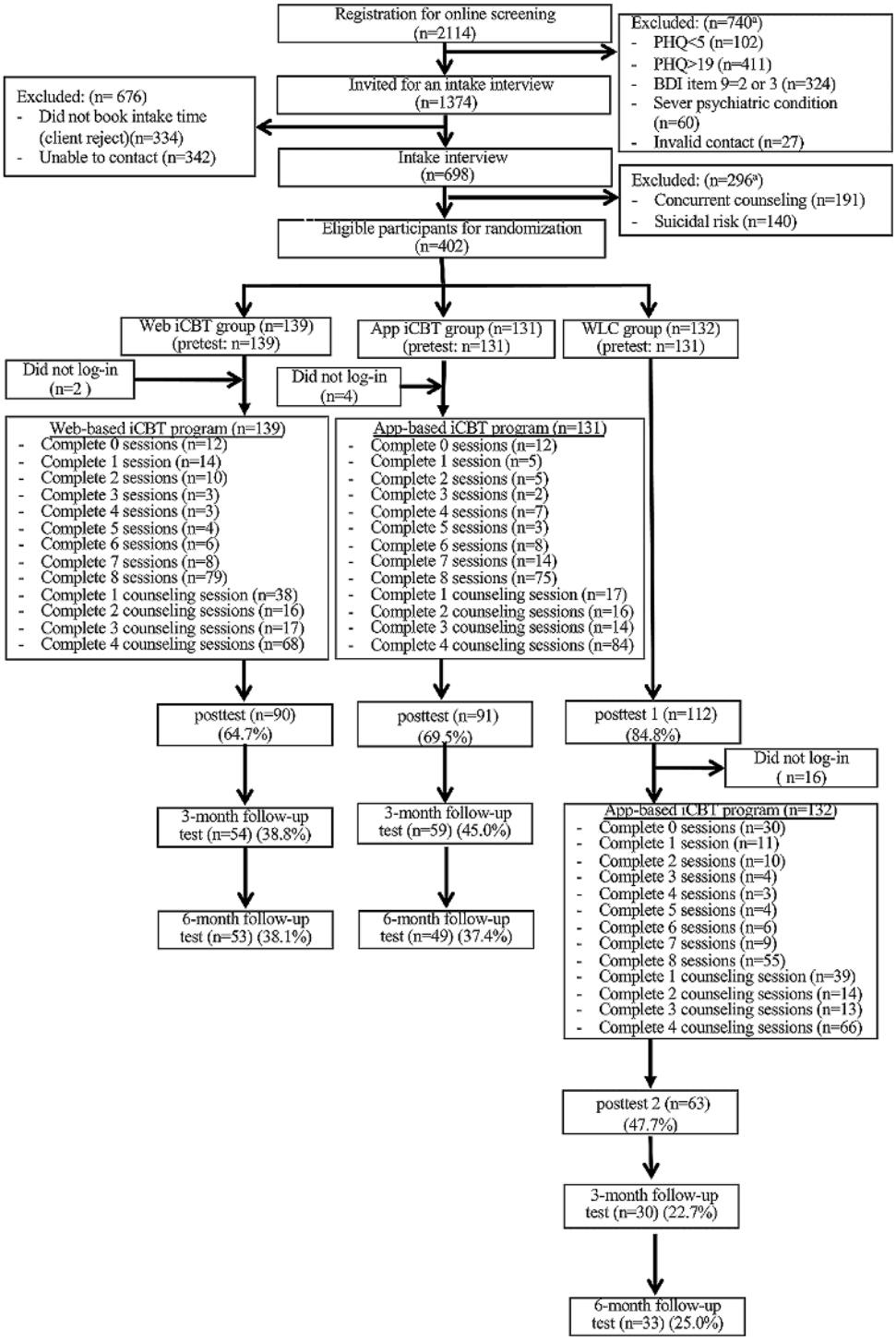

a. The sum of the following number is bigger than excluded number as some participants have fulfilled more than one exclusion criterion.

Participant Characteristics

Table 1 summarizes the participant data at baseline. Among the 402 participants who were randomized, 313 (77.9%) were female and 89 (22.1%) were male; 225 (56.0%) were single and 130 (32.3%) were married. The mean age was 37.3 (SD 11.0) years. The majority of participants (n=310, 77.1%) had university degrees, 85 (21.1%) and 1 (0.2%) had completed secondary and primary school education, respectively. A total of 77 (19.2%) participants were diagnosed with depression, 29 (7.2%) with anxiety, and 39 (9.7%) with other mental health conditions. Furthermore, 83 participants (20.6%) reported a history of self-harm, and 17 (4.2%) had previously attempted suicide; 99 participants (24.6%) reported using medication.

XSL•FO
RenderX

Table 1. Participant demographic characteristics.

Characteristics	Total (n=402)	Web group (n=139)	App group (n=131)	WLC[a] group (n=132)
Age, mean (SD)	37.3 (11.0)	38.1 (11.0)	37.2 (11.5)	36.6 (10.6)
Female, n (%)	313 (77.9)	106 (76.3)	100 (76.3)	107 (81.1)
Education, n (%)				
Primary school or below	1 (0.2)	1 (0.7)	0 (0.0)	0 (0.0)
Secondary school	85 (21.1)	24 (17.3)	36 (27.5)	25 (18.9)
University or above	310 (77.1)	111 (79.9)	94 (71.8)	105 (79.5)
Missing	6 (1.5)	3 (2.2)	1 (0.8)	2 (1.5)
Marital status, n (%)				
Single	225 (56.0)	78 (56.1)	72 (55.0)	75 (56.8)
Married	130 (32.3)	43 (30.9)	44 (33.6)	43 (32.6)
Other	30 (7.5)	10 (7.2)	11 (8.4)	9 (6.8)
Missing	17 (4.2)	8 (5.8)	4 (3.1)	5 (3.8)
Diagnosis, n (%)				
Depression	77 (19.2)	26 (18.7)	29 (22.1)	22 (16.7)
Anxiety	29 (7.2)	12 (8.6)	8 (6.1)	9 (6.8)
Other	39 (9.7)	16 (11.5)	14 (10.7)	9 (6.8)
Missing	257 (63.9)	85 (61.2)	80 (61.1)	92 (69.7)

[a]WLC: waitlist control.

Treatment Adherence and Attrition Rate

Following van Ballegooijen et al's [26] suggestion, those who were given an intervention but did not commence treatment were also included in the calculation of the adherence rate. Thus, for adherence to the online modules, 79 of 139 (56.8%) participants in the web-based iCBT group and 75 of 131 (57.3%) in the app-based iCBT group completed all 8 modules, resulting in an overall adherence rate of over 50% (154/270, 57.0%). For the counseling sessions, 68 of 139 (48.9%) participants in the web-based group and 84 of 131 (64.1%) in the app-based group attended all 4 sessions (overall adherence rate: 152/270, 56.3%) The participants in the web- and app-based iCBT groups completed an average of 5.96 out of 8 modules (74.5%) and 5.77 out of 8 assignments (72.1%), respectively. In terms of module completion rate, there was no significant difference between the web-based (79/139, 56.8%) and app-based (75/131, 57.3%, χ^2_1=0.005, *P*=.95) groups. Similarly, no significant difference was found in the assignment completion rate between the web-based (70/139, 50.4%) and app-based (69/131, 52.7%, χ^2_1=0.07, *P*=.8) groups. Furthermore, there was no association between assignment completion and gender (χ^2_1=0.97, *P*=.32) or medication (χ^2_1=0.52, *P*=.47). However, a positive association was found between assignment completion and age (odds ratio 1.04, *P*=.003). Thus, older participants completed more assignments. Regarding the attrition rate for questionnaire submission, 89 out of 270 (33.0%), 157 out of 270 (58.1%), and 168 out of 270 (62.2%) participants in the 2 iCBT groups were lost to follow-up at posttest, 3-month, and 6-month follow-ups, respectively. The submission rates for the questionnaires are shown in Figure 1. Adherence to completing the postmeasurements was similar to module adherence, with older individuals being more likely to complete the postmeasurement (odds ratio 1.02, *P*=.047). However, there was no association between completing the postmeasurement and model covariates such as gender (χ^2_1=1.07, *P*=.30), treatment group (χ^2_1=0.48, *P*=.49), or medication (χ^2_1=0.003, *P*=.95). A chi-square test showed that there were no significant differences in demographic characteristics between completers and noncompleters of the postmeasurement and all 8 online modules (see Tables S2 and S3 in Multimedia Appendix 2).

Treatment Effects at Posttest

The treatment effects at posttest, comparing the 2 iCBT groups with the WLC group, are shown in Table 2 and Figure 2. The primary analysis, which compared the 2 active treatments with the WLC at posttreatment using the data set with multiple imputed data, showed that both iCBT groups were superior to the WLC group. The adjusted differences on the BDI-II were –8.24 (95% CI –5.96 to –10.52; *P*<.001) for the web-based iCBT group, and –9.38 (95% CI –7.10 to –11.65; *P*<.001) for the app-based iCBT group. Likewise, the treatment effects for the PHQ-9 were also superior to the WLC group, with an adjusted score of –3.07 (95% CI –1.67 to –4.46; *P*<.001) for the web-based iCBT group and –4.50 (95% CI –3.12 to –5.89; *P*<.001) for the app-based iCBT group. All secondary outcome measures showed significant improvement compared with the WLC at postmeasurement (*P*=.009). The between-group effect sizes for the BDI-II were *d*=1.07 (95% CI 0.81-1.34) and *d*=1.15 (95% CI 0.88-1.43) for the web-based and app-based iCBT groups, respectively. For the PHQ-9, the corresponding

XSL•FO
RenderX

between-group effect sizes were *d*=0.78 (95% CI 0.52-1.04) and *d*=0.95 (95% CI 0.63-1.27) for the web-based and app-based iCBT groups, respectively. Sensitivity analyses using the completer-only data set yielded similar results (see Table S1 in Multimedia Appendix 2). The within-group effect sizes for the BDI-II at posttest were *d*=1.49 (95% CI 1.19-1.78) for the web-based iCBT group and *d*=1.64 (95% CI 1.33-1.95) for the app-based iCBT group. The corresponding effect sizes for the PHQ-9 at posttest were *d*=1.11 (95% CI 0.81-1.41) for the web-based iCBT group and *d*=1.70 (95% CI 1.38-2.02) for the app-based iCBT group. The WLC group showed significant improvement on all outcome measures, with medium to large within-group effect sizes when transitioning to the web-based iCBT (see Table S4 in Multimedia Appendix 2).

Table 2. Comparison of treatment effects between the active and WLC[a] groups at posttreatment (intent-to-treat analysis).

Outcome	Unadjusted mean (SD)			Adjusted difference (95% CI)				Between-group effect size *d* (95% CI)		Within-group effect size (95% CI)		
	Web	App	WLC	Web vs WLC	*P* value	App vs WLC	*P* value	Web vs WLC	App vs WLC	Web	App	WLC
BDI-II[b]	9.40 (8.90)	9.43 (8.45)	18.02 (9.02)	–8.24 (–5.96 to –10.52)	<.001	–9.38 (–7.10 to –11.65)	<.001	1.07 (0.81 to 1.34)	1.15 (0.88 to 1.43)	1.49 (1.19 to 1.78)	1.64 (1.33 to 1.95)	0.45 (0.2 to 0.71)
PHQ-9[c]	5.59 (4.43)	5.24 (3.65)	8.99 (4.58)	–3.07 (–1.67 to –4.46)	<.001	–4.50 (–3.12 to –5.89)	<.001	0.78 (0.52 to 1.04)	0.95 (0.63 to 1.27)	1.11 (0.81 to 1.41)	1.70 (1.38 to 2.02)	0.46 (0.19 to 0.73)
BAI[d]	1.60 (3.06)	1.53 (2.77)	5.35 (3.85)	–2.47 (–0.62 to –4.33)	.009	–3.32 (–1.47 to –5.17)	<.001	0.37 (0.11 to 0.63)	0.54 (0.29 to 0.8)	0.65 (0.37 to 0.92)	0.88 (0.6 to 1.16)	0.37 (0.11 to 0.62)
GHQ-12[e]	8.89 (7.89)	8.49 (6.69)	11.09 (7.49)	–3.65 (–2.54 to –4.75)	<.001	–4.49 (–3.38 to –5.6)	<.001	1.05 (0.77 to 1.32)	1.11 (0.83 to 1.39)	1.65 (1.34 to 1.95)	2.22 (1.88 to 2.56)	0.64 (0.38 to 0.9)
Positive automatic thoughts	2.40 (0.71)	2.33 (0.67)	1.98 (0.58)	0.38 (0.22 to 0.54)	<.001	0.40 (0.24 to 0.55)	<.001	–0.63 (–0.37 to –0.9)	–0.61 (–0.32 to –0.9)	–0.89 (–0.61 to –1.16)	–0.94 (–0.65 to –1.22)	–0.30 (0.04 to –0.55)
Negative automatic thoughts	1.85 (0.78)	1.89 (0.63)	2.29 (0.80)	–0.43 (–0.22 to –0.63)	<.001	–0.48 (–0.27 to –0.68)	<.001	0.63 (0.37 to 0.90)	0.66 (0.35 to 0.97)	0.83 (0.55 to 1.11)	1.00 (0.72 to 1.29)	0.28 (0.02 to 0.53)
Positive emotions	3.26 (0.69)	3.27 (0.62)	2.89 (0.68)	0.35 (0.18 to 0.52)	<.001	0.43 (0.26 to 0.60)	<.001	–0.56 (–0.27 to –0.84)	–0.60 (–0.3 to –0.9)	–0.89 (–0.62 to –1.17)	–1.10 (–0.82 to –1.39)	–0.30 (–0.04 to –0.55)
Negative emotions	2.91 (0.82)	2.92 (0.69)	3.33 (0.74)	–0.45 (–0.25 to –0.64)	<.001	–0.47 (–0.27 to –0.66)	<.001	0.64 (0.37 to 0.91)	0.65 (0.38 to 0.91)	1.06 (0.78 to 1.35)	1.21 (0.92 to 1.5)	0.50 (0.24 to 0.75)

[a]WLC: waitlist control.

[b]BDI-II: Beck Depression Inventory-II.

[c]PHQ-9: 9-item Patient Health Questionnaire.

[d]BAI: Beck Anxiety Inventory.

[e]GHQ-12: 12-item General Health Questionnaire.

XSL•FO
RenderX

Figure 2. Treatment effects for primary outcomes BDI and PHQ-9. Shaded regions show 95% CIs. BDI: Beck Depression Inventory; PHQ-9: 9-item Patient Health Questionnaire.

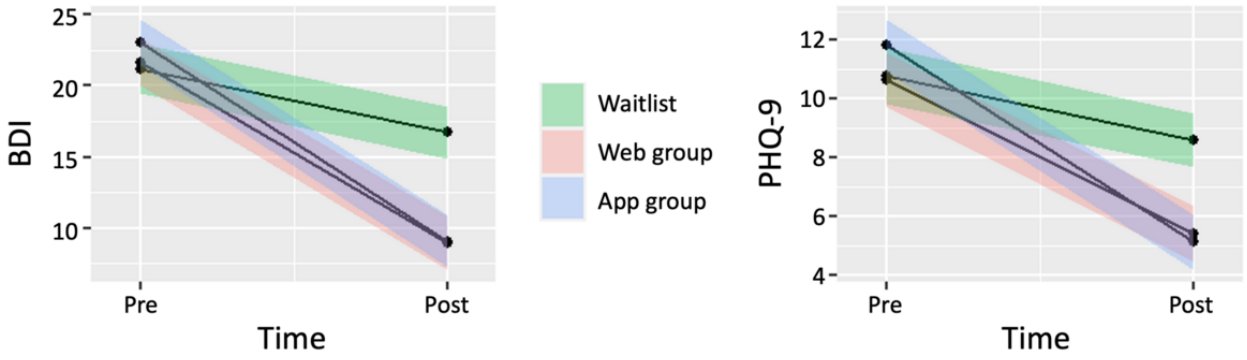

Reliable Changes at Postintervention Assessments

The frequencies and percentages of reliable change are shown in Table 3. At posttest, using the BDI-II as the outcome measure, reliable recovery was observed in 35 out of 90 (38.9%) participants in the web-based iCBT group, 36 out of 91 (39.6%) in the app-based iCBT group, and 3 out of 112 (2.7%) in the WLC group. For the PHQ-9 at posttest, the reliable recovery rates were 16 out of 67 (24%) for the web-based iCBT group, 16 out of 72 (22%) for the app-based iCBT group, and 5 out of 88 (6%) for the WLC group. Chi-square tests of association revealed a significant difference in the distribution of the 3 groups for both the BDI-II (χ^2_6=71.92, P<.001) and the PHQ-9 (χ^2_6=33.78, P<.001). When comparing clinical significance between the 2 treatment groups, no significant differences were found (χ^2_3=3.24, P=.35).

The reliable recovery rates for BDI-II at the 6-month follow-up were 13 out of 53 (25%) for the web-based iCBT group and 14 out of 49 (29%) for the app-based iCBT group. For PHQ-9 at the 6-month follow-up, the corresponding reliable recovery rates were 4 out of 40 (10%) for the web-based iCBT group and 3 out of 38 (8%) for the app-based iCBT group.

XSL•FO
RenderX

Table 3. Reliable and clinically significant changes summarized over group and time.

Outcome	Web group, n/N (%)	App group, n/N (%)	WLC[a] group, n/N (%)	Pearson chi-square (*df*)	*P* value
BDI-II[b], posttest				71.92 (6)	<.001
Recovered	35/90 (38.9)	36/91 (39.6)	3/110 (2.7)	N/A[c]	N/A
Improved	28/90 (31.1)	36/91 (39.6)	31/110 (28.2)	N/A	N/A
Unchanged	27/90 (30.0)	17/91 (18.7)	70/110 (63.6)	N/A	N/A
Deteriorated	0/90 (0)	2/91 (2.2)	6/110 (5.5)	N/A	N/A
PHQ-9[d], posttest				33.78 (6)	<.001
Recovered	16/67 (23.9)	16/72 (22.2)	5/88 (5.7)	N/A	N/A
Improved	22/67 (32.8)	39/72 (54.2)	26/88 (29.5)	N/A	N/A
Unchanged	28/67 (41.8)	15/72 (20.8)	50/88 (56.8)	N/A	N/A
Deteriorated	1/67 (1.5)	2/72 (2.8)	7/88 (8)	N/A	N/A
BDI-II, 3-month follow-up				3.24 (3)	.35
Recovered	21/54 (38.9)	20/59 (33.9)	N/A	N/A	N/A
Improved	20/54 (37.0)	27/59 (45.8)	N/A	N/A	N/A
Unchanged	13/54 (24.1)	10/59 (16.9)	N/A	N/A	N/A
Deteriorated	0/54 (0)	2/59 (3.4)	N/A	N/A	N/A
PHQ-9, 3-month follow-up				5.39 (3)	0.14
Recovered	8/46 (17.4)	7/46 (15.2)	N/A	N/A	N/A
Improved	20/46 (43.5)	27/46 (58.7)	N/A	N/A	N/A
Unchanged	18/46 (39.1)	10/46 (21.7)	N/A	N/A	N/A
Deteriorated	0/46 (0)	2/46 (4.3)	N/A	N/A	N/A
BDI-II, 6-month follow-up				6.32 (3)	.10
Recovered	13/53 (24.5)	14/49 (28.6)	N/A	N/A	N/A
Improved	22/53 (41.5)	17/49 (34.7)	N/A	N/A	N/A
Unchanged	17/53 (32.1)	11/49 (22.4)	N/A	N/A	N/A
Deteriorated	1/53 (1.9)	7/49 (14.3)	N/A	N/A	N/A
PHQ-9, 6-month follow-up				2.75 (3)	.43
Recovered	4/40 (10.0)	3/38 (7.9)	N/A	N/A	N/A
Improved	14/40 (35.0)	18/38 (47.4)	N/A	N/A	N/A
Unchanged	20/40 (50.0)	13/38 (34.2)	N/A	N/A	N/A
Deteriorated	2/40 (5.0)	4/38 (10.5)	N/A	N/A	N/A

[a]WLC: waitlist control.

[b]BDI-II: Beck Depression Inventory-II.

[c]N/A: not applicable.

[d]PHQ-9: 9-item Patient Health Questionnaire.

Comparisons Between Two iCBT Groups at Posttreatment Assessments

At posttreatment, the web-based iCBT group did not show a significant difference in BDI-II (adjusted difference 0.21, 95% CI –2.28 to 2.70, *P*=.87) or PHQ-9 (adjusted difference 0.35, 95% CI –1.08 to 1.77, *P*=.63) scores compared with the app-based iCBT group. Similarly, at the 3- and 6-month follow-ups, the differences between the 2 iCBT groups in their BDI-II and PHQ-9 scores remained nonsignificant (Table 4). For the secondary outcome measures, no significant differences were found between the 2 groups on any outcome measure at any time point, except for the GHQ-12 at 6 months, which favored the web group (adjusted difference –1.46, 95% CI –0.33 to –2.59, *P*=.01). Compared with the baseline, the within-group effect sizes ranged from 0.54 to 2.21 at posttreatment assessments for both the web- and app-based iCBT groups (Table 4). Thus, both iCBT groups maintained treatment effects at the 3- and 6-month follow-ups for most of the outcome measures.

XSL•FO
RenderX

Table 4. Comparisons between web- and app-based iCBT[a] groups on primary and secondary outcome measures at posttreatment (intention-to-treat analysis).

Measure and time	Unadjusted mean (SD)		Adjusted difference (95% CI)	*P* value	Between-group effect size *d* (95% CI)	Within-group effect size (95% CI)	
	Web	App				Web	App
BDI-II[b]							
Baseline	22.24 (8.52)	23.66 (8.77)	N/A[c]	N/A	N/A	N/A	N/A
Post	9.40 (8.90)	9.51 (8.47)	0.21 (–2.28 to 2.70)	.87	–0.01 (–0.3 to 0.28)	1.48 (1.18 to 1.78)	1.64 (1.33 to 1.94)
3 months	9.43 (9.58)	9.15 (8.51)	–0.14 (–3.11 to 2.83)	.93	0.03 (–0.34 to 0.4)	1.45 (1.1 to 1.8)	1.67 (1.32 to 2.02)
6 months	10.15 (9.03)	13.47 (12.24)	–1.37 (–4.46 to 1.72)	.38	–0.37 (–0.76 to 0.02)	1.40 (1.05 to 1.74)	1.04 (0.69 to 1.38)
PHQ-9[d]							
Baseline	10.77 (4.9)	12.07 (4.27)	N/A	N/A	N/A	N/A	N/A
Post	5.59 (4.43)	5.29 (3.68)	0.35 (–1.08 to 1.77)	.63	0.07 (–0.22 to 0.36)	1.11 (0.8 to 1.41)	1.69 (1.36 to 2.02)
3 months	5.41 (4.01)	5.83 (3.57)	–0.27 (–1.94 to 1.41)	.75	–0.1 (–0.47 to 0.28)	1.16 (0.81 to 1.51)	1.55 (1.18 to 1.91)
6 months	5.94 (5.44)	7.65 (5.06)	–0.86 (–2.67 to 0.94)	.35	–0.39 (–0.79 to 0.01)	0.95 (0.59 to 1.31)	0.97 (0.62 to 1.33)
GHQ-12[e]							
Baseline	7.46 (3.86)	8.4 (3.31)	N/A	N/A	N/A	N/A	N/A
Post	1.60 (3.06)	1.53 (2.79)	0.05 (–0.84 to 0.93)	.91	0.02 (–0.27 to 0.31)	1.64 (1.34 to 1.95)	2.21 (1.87 to 2.55)
3 months	1.85 (3.02)	1.69 (2.82)	–0.08 (–1.17 to 1)	.88	0.05 (–0.32 to 0.42)	1.54 (1.19 to 1.89)	2.12 (1.74 to 2.49)
6 months	1.77 (3.20)	3.49 (4.10)	–1.46 (–0.33 to –2.59)	.01	–0.52 (–0.91 to –0.12)	1.54 (1.19 to 1.89)	1.39 (1.03 to 1.74)
BAI[f]							
Baseline	14.53 (9.29)	15.36 (8.8)	N/A	N/A	N/A	N/A	N/A
Post	8.89 (7.89)	8.37 (6.63)	0.53 (–1.25 to 2.31)	.56	0.06 (–0.23 to 0.36)	0.64 (0.37 to 0.91)	0.88 (0.59 to 1.15)
3 months	7.67 (6.42)	8.03 (6.78)	0.59 (–1.56 to 2.73)	.59	–0.05 (–0.42 to 0.32)	0.80 (0.47 to 1.12)	0.89 (0.57 to 1.21)
6 months	7.89 (7.99)	9.88 (7.72)	0.32 (–1.91 to 2.56)	.78	–0.25 (–0.64 to 0.14)	0.74 (0.42 to 1.07)	0.64 (0.31 to 0.98)
Positive automatic thoughts							
Baseline	1.85 (0.55)	1.79 (0.52)	N/A	N/A	N/A	N/A	N/A
Post	2.4 (0.71)	2.33 (0.67)	0.01 (–0.18 to 0.2)	.93	0.1 (–0.19 to 0.39)	–0.88 (–1.16 to –0.6)	–0.93 (–1.21 to –0.65)
3 months	2.49 (0.67)	2.44 (0.84)	–0.1 (–0.32 to 0.12)	.38	0.07 (–0.3 to 0.44)	–1.08 (–1.41 to –0.75)	–1.03 (–1.36 to –0.71)
6 months	2.46 (0.85)	2.15 (0.71)	0.08 (–0.16 to 0.31)	.52	0.47 (0.08 to 0.86)	–0.93 (–1.26 to –0.6)	–0.63 (–0.96 to –0.29)
Negative automatic thoughts							
Baseline	2.55 (0.88)	2.65 (0.85)	N/A	N/A	N/A	N/A	N/A
Post	1.85 (0.78)	1.88 (0.62)	–0.02 (–0.22 to 0.18)	.86	–0.04 (–0.33 to 0.25)	0.83 (0.55 to 1.1)	1 (0.72 to 1.28)
3 months	1.74 (0.68)	1.82 (0.71)	–0.03 (–0.27 to 0.21)	.81	–0.09 (–0.46 to 0.28)	0.97 (0.64 to 1.29)	1.02 (0.7 to 1.34)
6 months	1.75 (0.78)	2.05 (0.93)	–0.15 (–0.4 to 0.1)	.24	–0.38 (–0.77 to 0.01)	0.93 (0.6 to 1.26)	0.68 (0.34 to 1.02)
Positive emotions							
Baseline	2.72 (0.54)	2.65 (0.54)	N/A	N/A	N/A	N/A	N/A
Post	3.26 (0.69)	3.28 (0.62)	–0.03 (–0.22 to 0.16)	.87	–0.04 (–0.33 to 0.25)	–0.89 (–1.17 to –0.61)	–1.1 (–1.38 to –0.81)
3 months	3.26 (0.75)	3.33 (0.81)	–0.1 (–0.33 to 0.13)	.93	–0.12 (–0.49 to 0.25)	–0.89 (–1.21 to –0.56)	–1.07 (–1.39 to –0.74)
6 months	3.24 (0.75)	2.97 (0.72)	0.12 (–0.11 to 0.36)	.38	0.4 (0.01 to 0.79)	–0.85 (–1.18 to –0.52)	–0.54 (–0.88 to –0.21)

XSL•FO
RenderX

Measure and time	Unadjusted mean (SD)		Adjusted difference (95% CI)	*P* value	Between-group effect size *d* (95% CI)	Within-group effect size (95% CI)	
	Web	App				Web	App
Negative emotions							
Baseline	3.73 (0.75)	3.78 (0.73)	N/A	N/A	N/A	N/A	N/A
Post	2.91 (0.82)	2.92 (0.7)	–0.02 (–0.23 to 0.2)	.63	–0.02 (–0.31 to 0.28)	1.06 (0.78 to 1.34)	1.21 (0.91 to 1.5)
3 months	2.78 (0.85)	2.93 (0.73)	–0.07 (–0.32 to 0.18)	.75	–0.19 (–0.56 to 0.18)	1.23 (0.89 to 1.57)	1.17 (0.84 to 1.5)
6 months	2.86 (0.91)	3.17 (0.88)	–0.1 (–0.36 to 0.16)	.35	–0.39 (–0.78 to 0)	1.09 (0.75 to 1.42)	0.8 (0.46 to 1.13)

[a]iCBT: internet-based cognitive behavioral therapy.

[b]BDI-II: Beck Depression Inventory-II.

[c]N/A: not applicable.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]GHQ-12: 12-item General Health Questionnaire.

[f]BAI: Beck Anxiety Inventory.

Discussion

Effectiveness of iCBT for Hong Kong Chinese

To the best of our knowledge, this study is one of the few to develop and test the effectiveness of iCBT among Hong Kong Chinese. Both the web- and app-based iCBT groups were superior to the WLC in managing depressive symptoms. The findings suggest that a culturally and linguistically appropriate, therapist-guided iCBT program yields positive therapeutic effects, with a good recovery rate and treatment adherence, for Hong Kong Chinese with mild to moderate depressive symptoms. These positive effects are maintained at the 3- and 6-month follow-ups, including reductions in depression and anxiety symptoms, psychological distress, negative automatic thoughts, and negative emotions, as well as increases in positive automatic thoughts and emotions. The 2 delivery modalities of iCBT showed similar intervention effects on most of the outcome measures at posttreatment and follow-up tests. These findings are consistent with those reported in Western contexts [10], China (eg, [16,18]), and Hong Kong [19,20]. The recovery rates at posttreatment, as measured by the BDI-II, were 39% (35/90) and 40% (36/91) for web- and app-based iCBT, respectively. These rates align with those (30%-55%) reported in RCT studies involving Western participants [24,25], but are lower than the 58% recovery rate reported in an RCT study of iCBT conducted in hospitals with Chinese patients with depression in China [16]. The positive effects of iCBT on depression treatment are comparable to face-to-face CBT in terms of symptom reduction (eg, [56]) and recovery rate (eg, [57]), suggesting that iCBT does not compromise treatment outcomes.

Like CBT, it appears that a culturally adapted iCBT works well for Chinese people. The positive effects of the CANDO Program may be attributed to the cultural and linguistic adaptations of iCBT for Hong Kong Chinese. First, the program is delivered in Cantonese, the mother tongue of Hong Kong residents, to facilitate their understanding of the content. Additionally, 4-character Chinese idioms [31] are used to describe cognitive distortions, ensuring an accurate understanding of the CBT concepts and skills. Second, many real-life examples are incorporated into the case videos to demonstrate the application of CBT skills in coping with depression in the context of Hong Kong. These examples may resonate with participants' own automatic thoughts and cognitive rules, enhancing their self-learning of CBT skills. Third, laypersons—such as family members, friends, and colleagues—who have experience and knowledge of CBT appear in the case demonstration videos (except for 1 case) as therapists. This approach helps normalize the help-seeking and counseling process, reducing the stigma surrounding mental health issues and making it more acceptable to Hong Kong Chinese. Fourth, the CANDO Program is video-based with limited text narrative, which aligns with the preferred approach of Hong Kong people for receiving new information and learning new knowledge and skills through watching videos. Fifth, the program includes relaxation exercises to enhance the Chinese philosophy of body-mind connection, which complements the CBT framework that links the body, cognition, and emotions. Finally, the therapist is a Chinese clinical psychologist with training and practical experience in both Hong Kong and the United States, ensuring cultural sensitivity in working with Hong Kong Chinese.

Program Adherence and Attrition

The percentage of completed sessions in this study (5.96/8, 75%) is slightly lower than those reported in meta-analyses (80.8% in [26] and 76.0% in [10]), but much higher than that reported in a Hong Kong study (52.7% in [21]) and lower than that in a study in China (83.8% in [18]). The percentage of participants who completed all the sessions in this study (154/270, 57.0%) is similar to those reported in [26] (65.1%) and other empirical studies (eg, 60% in [28] and 67% in [21]), but lower than that in [18] (81.4%). Overall, the program adherence in this study is comparable to that in previous studies, though slightly lower than that in [18]. One reason for this difference may be related to the number of sessions (8 sessions in the CANDO Program compared with 5 in [18]), as it is generally easier to complete an iCBT program with fewer sessions. Another factor is the setting; Ying et al [18] conducted their study in a hospital, where doctors acted as therapists and provided relatively intensive guidance, which likely encouraged greater program engagement. The Chinese cultural tendency to

XSL•FO
RenderX

adhere to the authority of medical professionals may have contributed to the higher adherence observed in their study.

Consistent with previous findings, this study also observed a high attrition rate at postintervention. Specifically, the attrition rate at posttest (89/270, 33.0%) is comparable to the average dropout rate of 28% found in therapist-guided computer-based intervention studies in [58], which is higher than the approximately 17% dropout rate in RCT studies of face-to-face CBT [58]. The high attrition rates in this study are also consistent with those reported in Hong Kong studies [21,59]. One possible reason for the high attrition rate may be the fast-paced lifestyle of Hong Kong residents, which limits their ability to commit to the program. Additionally, a lack of motivation to participate could contribute to the dropout rate. Similar to [21], the participants in this study did not have a formal diagnosis of a mental disorder and did not have an immediate need for help-seeking or intervention retention.

No significant difference in program effectiveness or adherence was found between the web- and app-based iCBT groups among Hong Kong people with mild to moderate levels of depression. This suggests that the medium of service delivery does not impact the intervention effects, with both formats being equally effective.

Theoretical and Practice Implications

At the theoretical level, the findings offer empirical evidence supporting the effectiveness of iCBT for handling depressive symptoms in Chinese communities. Culturally and linguistically adapted iCBT ensures its relevance and effectiveness for the Chinese population, particularly for Hong Kong Chinese. The use of the internet and mobile app for delivering CBT represents a significant advancement in mental health care, especially in Hong Kong, where the high penetration of digital devices facilitates the widespread adoption of iCBT. This study provides a strong evidence base that supports the adoption and further development of iCBT in Hong Kong. At the practice level, this study demonstrates that Hong Kong Chinese are receptive to iCBT, making it a viable alternative that can be integrated into traditional mental health services. Furthermore, iCBT is cost-effective. This study, conducted during the COVID-19 pandemic when most face-to-face mental health services were suspended, highlights the significant advantage of iCBT in terms of the therapist-to-client ratio. For the CANDO Program, the therapist-to-client ratio is 1:190 per year, which is much higher than the traditional face-to-face therapy ratio of 1:40-50 in Hong Kong [60]. This is especially relevant to Hong Kong, where there is a severe shortage of mental health professionals and long wait times for traditional services. iCBT offers a scalable and viable solution to address this service gap in the long term. Therefore, more resources, particularly from the government, should be allocated to developing additional iCBT programs for mental health conditions, training more iCBT therapists, and integrating iCBT into the infrastructure of existing mental health care systems in Hong Kong.

Limitations

One major limitation of this study is the high attrition rate in questionnaire submission postintervention, despite various efforts to encourage participation. Although this is common, it may introduce nonresponse bias. Future studies should aim to reduce the attrition rate through improvements in program design and service delivery. A second limitation is that the therapist's time per client was not recorded, making it difficult to evaluate the cost-effectiveness of the program more accurately. Future studies should track these data to better assess the program's cost-effectiveness. Third, client satisfaction data were not collected, which could provide valuable insights into user experience and should be considered in future research. Finally, a large proportion of participants were female (313/402, 77.9%) and highly educated (310/402, 77.1%, holding a university degree or higher), which may limit the generalizability of the findings to the broader Hong Kong population. While gender was controlled as a covariate in the data analysis, future studies should aim to recruit more male participants and those with lower levels of education to improve representativeness.

Conclusions

This RCT study demonstrates that both web- and app-based iCBT are effective in reducing depression and anxiety symptoms, psychological distress, and negative automatic thoughts and emotions, while enhancing positive automatic thoughts and emotions in Hong Kong Chinese with mild to moderate depressive symptoms. The study also shows good program adherence and recovery rates, although the attrition rate for questionnaire submission postintervention was relatively low. ICBT is a promising approach in Hong Kong and warrants increased resource allocation for further development and implementation.

Acknowledgments

This study was funded by the Innovation and Technology Fund for Better Living, Innovation and Technology Bureau of Hong Kong SAR (reference number ITB/FBL/5022/18/P). The authors thank Professor Per Carlbring for his guidance during the grant application process for this project. The authors also wish to express their gratitude to Miss Chor Kiu Chan for system development and technical support, Dr. Fiona Yip for service delivery, Miss Jenny Chan for conducting the intake interviews, and Miss Amy Ho and Miss Tracy Leung for their administrative support.

Data Availability

The data sets generated and analyzed during this study are not publicly available due to privacy concerns but can be obtained from the corresponding author upon reasonable request.

XSL•FO
RenderX

Conflicts of Interest

JYP is the Founder and Director of the NGO "Online Companion," which was established with financial support from the Technology Start-up Support Scheme for Universities (TSSSU) (2022-2024) under the Innovation and Technology Commission (ITC) of Hong Kong SAR to commercialize the CANDO Program. However, she does not hold any shares in this NGO.

Multimedia Appendix 1
Some screenshots of the Confront and Navigate Depression Online (CANDO) Program.
[DOCX File , 980 KB - jmir_v27i1e64303_app1.docx]

Multimedia Appendix 2
Additional data analysis.
[DOCX File , 28 KB - jmir_v27i1e64303_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1228 KB - jmir_v27i1e64303_app3.pdf]

References

1. World Health Organization (WHO). Comprehensive Mental Health Action Plan 2013-2030. Geneva, Switzerland: World Health Organization; 2021.
2. Ni MY, Yao XI, Leung KSM, Yau C, Leung CMC, Lun P, et al. Depression and post-traumatic stress during major social unrest in Hong Kong: a 10-year prospective cohort study. The Lancet 2020 Jan;395(10220):273-284. [doi: 10.1016/s0140-6736(19)33160-5]
3. Lam LC, Wong CS, Wang M, Chan W, Chen EY, Ng RM, et al. Prevalence, psychosocial correlates and service utilization of depressive and anxiety disorders in Hong Kong: the Hong Kong Mental Morbidity Survey (HKMMS). Soc Psychiatry Psychiatr Epidemiol 2015 Sep 8;50(9):1379-1388. [doi: 10.1007/s00127-015-1014-5] [Medline: 25660760]
4. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Feb 02;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
5. Bains N, Abdijadid S. Major Depressive Disorder. Treasure Island, FL: StatPearls Publishing; Apr 10, 2023.
6. IsHak WW, Wen RY, Naghdechi L, Vanle B, Dang J, Knosp M, et al. Pain and depression: a systematic review. Harv Rev Psychiatry 2018;26(6):352-363 [FREE Full text] [doi: 10.1097/hrp.0000000000000198]
7. Müller G, Bombana M, Heinzel-Gutenbrenner M, Kleindienst N, Bohus M, Lyssenko L, et al. Socio-economic consequences of mental distress: quantifying the impact of self-reported mental distress on the days of incapacity to work and medical costs in a two-year period: a longitudinal study in Germany. BMC Public Health 2021 Mar 31;21(1):625 [FREE Full text] [doi: 10.1186/s12889-021-10637-8]
8. Andersson E, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
9. Baumann M, Stargardt T, Frey S. Cost-utility of internet-based cognitive behavioral therapy in unipolar depression: a Markov model simulation. Appl Health Econ Health Policy 2020 Aug 15;18(4):567-578 [FREE Full text] [doi: 10.1007/s40258-019-00551-x] [Medline: 32060822]
10. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
11. Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e18100 [FREE Full text] [doi: 10.2196/18100] [Medline: 32865497]
12. Stolz T, Schulz A, Krieger T, Vincent A, Urech A, Moser C, et al. A mobile app for social anxiety disorder: a three-arm randomized controlled trial comparing mobile and PC-based guided self-help interventions. J Consult Clin Psychol 2018 Jun;86(6):493-504 [FREE Full text] [doi: 10.1037/ccp0000301] [Medline: 29781648]
13. Survey on IT usage and penetration. Office of the Government Chief Information Officer. 2024. URL: https://www.ogcio.gov.hk/en/about_us/facts/it_usage_penetration_survey.html [accessed 2024-06-12]
14. Statista. Penetration rate of smartphone among people over ten years old in Hong Kong from June to September 2022. Statista. 2022. URL: https://www.statista.com/statistics/1093416/hong-kong-penetration-rate-of-smartphone/ [accessed 2024-06-12]

XSL•FO
RenderX

15. Kong F, Yu L, Hou Y, Zhu L, Zhou J, Huang L, et al. Efficacy of internet-based cognitive behavioral therapy for subthreshold depression among older adults in institutional long-term care settings: pragmatic randomized controlled trial. J Med Internet Res 2024 Mar 01;26:e40187 [FREE Full text] [doi: 10.2196/40187] [Medline: 38427424]
16. Lin Z, Cheng L, Han X, Wang H, Liao Y, Guo L, et al. The effect of internet-based cognitive behavioral therapy on major depressive disorder: randomized controlled trial. J Med Internet Res 2023 Sep 22;25:e42786 [FREE Full text] [doi: 10.2196/42786] [Medline: 37738092]
17. Song J, Jiang R, Chen N, Qu W, Liu D, Zhang M, et al. Self-help cognitive behavioral therapy application for COVID-19-related mental health problems: a longitudinal trial. Asian J Psychiatr 2021 Jun;60:102656 [FREE Full text] [doi: 10.1016/j.ajp.2021.102656] [Medline: 33866282]
18. Ying Y, Ji Y, Kong F, Wang M, Chen Q, Wang L, et al. Efficacy of an internet-based cognitive behavioral therapy for subthreshold depression among Chinese adults: a randomized controlled trial. Psychol Med 2022 Apr 07;53(9):3932-3942. [doi: 10.1017/s0033291722000599]
19. Mak WW, Chio FH, Chan AT, Lui WW, Wu EK. The efficacy of internet-based mindfulness training and cognitive-behavioral training with telephone support in the enhancement of mental health among college students and young working adults: randomized controlled trial. J Med Internet Res 2017 Mar 22;19(3):e84 [FREE Full text] [doi: 10.2196/jmir.6737] [Medline: 28330831]
20. Young DKW, Carlbring P, Ng PYN, Chen QJ. Feasibility of Self-Guided Online Cognitive Behavioral Therapy for University Students During COVID-19. Research on Social Work Practice 2022 Apr 23;32(8):898-911. [doi: 10.1177/10497315221087904]
21. Mak WWS, Tong ACY, Fu ACM, Leung IWY, Jung OHC, Watkins ER, et al. Efficacy of internet-based rumination-focused cognitive behavioral therapy and mindfulness-based intervention with guided support in reducing risks of depression and anxiety: a randomized controlled trial. Appl Psychol Health Well Being 2024 May 10;16(2):696-722 [FREE Full text] [doi: 10.1111/aphw.12512] [Medline: 38073271]
22. Karyotaki E, Furukawa TA, Efthimiou O, Riper H, Cuijpers P. Guided or self-guided internet-based cognitive-behavioural therapy (iCBT) for depression? Study protocol of an individual participant data network meta-analysis. BMJ Open 2019 Jun 05;9(6):e026820 [FREE Full text] [doi: 10.1136/bmjopen-2018-026820] [Medline: 31171550]
23. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clinical Psychology Review 2018 Jul 19;63:80-92 [FREE Full text] [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
24. Jakobsen H, Andersson G, Havik OE, Nordgreen T. Guided internet-based cognitive behavioral therapy for mild and moderate depression: a benchmarking study. Internet Interv 2017 Mar;7:1-8 [FREE Full text] [doi: 10.1016/j.invent.2016.11.002] [Medline: 30135820]
25. Johansson R, Nyblom A, Carlbring P, Cuijpers P, Andersson G. Choosing between internet-based psychodynamic versus cognitive behavioral therapy for depression: a pilot preference study. BMC Psychiatry 2013 Oct 18;13(1):268 [FREE Full text] [doi: 10.1186/1471-244x-13-268]
26. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
27. Calero Elvira A, Santacreu Ivars M, Marchena Giráldez C, Shih P. Internet-based cognitive behavioural therapy programme with and without videoconference guidance sessions: a randomized controlled trial to treat work-related symptoms of anxiety and depression. Clin Psychol Psychother 2021 Sep 23;28(5):1230-1242 [FREE Full text] [doi: 10.1002/cpp.2571] [Medline: 33586819]
28. Johansson P, Westas M, Andersson G, Alehagen U, Broström A, Jaarsma T, et al. An internet-based cognitive behavioral therapy program adapted to patients with cardiovascular disease and depression: randomized controlled trial. JMIR Ment Health 2019 Oct 03;6(10):e14648 [FREE Full text] [doi: 10.2196/14648] [Medline: 31584000]
29. Ellis DM, Draheim AA, Anderson PL. Culturally adapted digital mental health interventions for ethnic/racial minorities: a systematic review and meta-analysis. J Consult Clin Psychol 2022 Oct;90(10):717-733 [FREE Full text] [doi: 10.1037/ccp0000759] [Medline: 36227330]
30. Ng TK, Wong DFK. The efficacy of cognitive behavioral therapy for Chinese people: a meta-analysis. Aust N Z J Psychiatry 2017 Nov 16;52(7):620-637. [doi: 10.1177/0004867417741555] [Medline: 29143541]
31. Wong D. Coming out of the valley of depression: Cognitive therapy self-help manual (in Chinese). Hong Kong: Natural Health Press; 2005.
32. Pan J, Carlbring P, Lu L. A therapist-guided internet-based cognitive behavioral therapy program for Hong Kong university students with psychological distress: a study protocol for a randomized controlled trial. Internet Interv 2023 Mar;31:100600. [doi: 10.1016/j.invent.2022.100600] [Medline: 36632436]
33. Patel U, Sobowale K, Fan J, Liu N, Kuwabara S, Lei Z, et al. Cultural considerations for the adaptation of an internet-based intervention for depression prevention in Mainland China. Int J Adolesc Med Health 2016 Feb 27;29(5):20150099 [FREE Full text] [doi: 10.1515/ijamh-2015-0099] [Medline: 26926858]

XSL•FO
RenderX

34. Ren Z, Li X, Zhao L, Yu X, Li Z, Lai L, et al. Effectiveness and mechanism of internet-based self-help intervention for depression: the Chinese version of MoodGYM. Acta Psychologica Sinica 2016;48(7):818-832. [doi: 10.3724/sp.j.1041.2016.00818]
35. CANDO. URL: https://cando.hkbu.edu.hk/ [accessed 2024-12-23]
36. Beck JS. Cognitive Behavior Therapy: Basics and Beyond (2nd Edition). New York, NY: The Guilford Publications; 2011.
37. Pan J, Ye S, Ng PYN, Lu L. A randomized controlled trial of cognitive behavioral group prevention program for Mainland Chinese university students in Hong Kong. Research on Social Work Practice 2020 Oct 14;31(2):180-193 [FREE Full text] [doi: 10.1177/1049731520962181]
38. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995 Feb;23(1):67-82 [FREE Full text] [doi: 10.1007/bf01447045]
39. Beck A, Steer RA, Brown GK. Manual for BDI-II. Psychological Corporation. San Antonio, Texas: Psychological Corporation; 1996.
40. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
41. Beck A, Steer RA. Beck Anxiety Inventory: Manual (2nd edition). San Antonio, TX: Psychological Corporation; 1993. URL: https://archive.org/details/BAIManual/page/n11/mode/2up [accessed 2025-01-19]
42. Goldberg DP, Williams P. A User's Guide to the General Health Questionnaire. London, UK: NFER-Nelson; 1988.
43. Li WHC, Chung JOK, Chui MML, Chan PSL. Factorial structure of the Chinese version of the 12-item General Health Questionnaire in adolescents. J Clin Nurs 2009 Dec;18(23):3253-3261 [FREE Full text] [doi: 10.1111/j.1365-2702.2009.02905.x] [Medline: 19732241]
44. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan 01;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
45. Pan J, Ye S, Ng P. Validation of the Automatic Thoughts Questionnaire (ATQ) among Mainland Chinese students in Hong Kong. J Clin Psychol 2016 Jan;72(1):38-48 [FREE Full text] [doi: 10.1002/jclp.22193] [Medline: 26032223]
46. Netemeyer RG, Williamson DA, Burton S, Biswas D, Jindal S, Landreth S, et al. Psychometric properties of shortened versions of the Automatic Thoughts Questionnaire. Educational and Psychological Measurement 2002 Feb 01;62(1):111-129. [doi: 10.1177/0013164402062001008]
47. Kendall PC, Howard BL, Hays RC. Self-referent speech and psychopathology: the balance of positive and negative thinking. Cogn Ther Res 1989 Dec;13(6):583-598. [doi: 10.1007/bf01176069]
48. Hamid P, Cheng ST. The development and validation of an index of emotional disposition and mood state: the Chinese Affect Scale. Educational and Psychological Measurement 1996 Dec 01;56(6):995-1014 [FREE Full text] [doi: 10.1177/0013164496056006006]
49. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. Journal of Pharmacology and Pharmacotherapeutics 2010 Dec 01;1(2):100-107 [FREE Full text] [doi: 10.4103/0976-500x.72352]
50. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. Journal of Statistical Software 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
51. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Abingdon, UK: Routledge; 1988.
53. Lenth R. emmeans: estimated marginal means, aka least-squares means (R package version 1.8.8). R Foundation. 2023. URL: https://CRAN.R-project.org/package=emmeans [accessed 2024-06-12]
54. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023. URL: https://www.R-project.org [accessed 2024-06-12]
55. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. In: Methodological Issues and Strategies in Clinical Research. Washington, DC: American Psychological Association; 1992:631-648.
56. Silveus SA, Schmit MK, Oliveira JT, Hughes LE. Meta‐analysis of culturally adapted cognitive behavioral therapy for anxiety and depression. J Counsel Develop 2023 Feb 20;101(2):129-142 [FREE Full text] [doi: 10.1002/jcad.12463]
57. Lutz W, Schiefele AK, Wucherpfennig F, Rubel J, Stulz N. Clinical effectiveness of cognitive behavioral therapy for depression in routine care: a propensity score based comparison between randomized controlled trials and clinical practice. Journal of Affective Disorders 2016 Jan 01;189:150-158 [FREE Full text] [doi: 10.1016/j.jad.2015.08.072] [Medline: 26433763]
58. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342 [FREE Full text] [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]

XSL•FO
RenderX

59. Pan J, Carlbring P, Lu L. Efficacy of internet-based cognitive behavioral therapy for Hong Kong university students: a randomized controlled trial. Research on Social Work Practice 2024 May 07 (forthcoming). [doi: 10.1177/10497315241252054]
60. Service framework of personalized care for adults with severe mental illness in Hong Kong. In: Hospital Authority and Social Welfare Department. Hong Kong City, Hong Kong: Hospital Authority and Social Welfare Department; 2016.

Abbreviations

BAI: Beck Anxiety Inventory
BDI-II: Beck Depression Inventory-II
CANDO: Confront and Navigate Depression Online
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
GHQ-12: 12-item General Health Questionnaire
iCBT: internet-based cognitive behavioral therapy
NA: negative affect
PA: positive affect
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial
WHO: World Health Organization
WLC: waitlist control

Edited by T de Azevedo Cardoso; submitted 14.07.24; peer-reviewed by I Choi, N Titov; comments to author 20.09.24; revised version received 10.10.24; accepted 29.11.24; published 25.02.25.

Please cite as:
Pan JY, Rafi J
Culturally Adapted Guided Internet-Based Cognitive Behavioral Therapy for Hong Kong People With Depressive Symptoms: Randomized Controlled Trial
J Med Internet Res 2025;27:e64303
URL: https://www.jmir.org/2025/1/e64303
doi:10.2196/64303
PMID:39998865

©Jia-Yan Pan, Jonas Rafi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

More Cyberbullying, Less Happiness, and More Injustice—Psychological Changes During the Pericyberbullying Period: Quantitative Study Based on Social Media Data

Xingyun Liu[1,2], PhD; Miao Liu[1,2], BA; Xin Kang[1,2], BA; Nuo Han[3,4], PhD; Yuehan Liao[1,2], BA; Zhihong Ren[1,2], Prof Dr

[1]Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan, China

[2]Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan, China

[3]Department of Psychology, Faculty of Arts and Sciences, Beijing Normal University at Zhuhai, Zhuhai, China

[4]Beijing Key Laboratory of Applied Experimental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China

Corresponding Author:
Zhihong Ren, Prof Dr
Key Laboratory of Adolescent Cyberpsychology and Behavior
Ministry of Education
No 382 Xiongchu Avenue, Hongshan District
Wuhan, 430079
China
Phone: 86 13627131550
Email: ren@ccnu.edu.cn

Abstract

Background: The phenomenon of cyberbullying is becoming increasingly severe, and many studies focus on the negative psychological impacts of cyberbullying survivors. However, current survey methods cannot provide direct and reliable evidence of the short-term psychological effects of cyberbullying survivors, as it is impractical to measure psychological changes before and after such an unpredictable event in a short period.

Objective: This study aims to explore the psychological impacts of cyberbullying on survivors during the pericyberbullying period, defined as the critical time frame surrounding the first cyberbullying incident, encompassing the psychological changes before, during, and after the event.

Methods: We collected samples from 60 cyberbullying survivors (experimental group, 94/120, 78% female) and 60 individuals who have not experienced cyberbullying (control group, matched by sex, location, and number of followers) on Sina Weibo, a social media platform developed by Sina Corporation. During the pericyberbullying period, we retrospectively measured psychological traits 3 months before and after the first cyberbullying incident for both groups. Social media data and predictive models were used to identify survivors' internal psychological traits, including happiness, suicide risk, personality traits, and moral perceptions of the external environment. Network analysis was then performed to explore the interplay between cyberbullying experiences and psychological characteristics.

Results: During the pericyberbullying period, survivors exhibited significantly reduced happiness (t_{59}=2.14; P=.04), marginally increased suicide risk, and significant changes in the Big 5 personality traits, including decreased conscientiousness (t_{59}=2.27; P=.03), agreeableness (t_{59}=2.79; P=.007), and extraversion (t_{59}=2.26; P=.03), alongside increased neuroticism (t_{59}=–3.42; P=.001). Regarding moral perceptions of the external environment, survivors showed significant increases in communicative moral motivation (t_{59}=–2.62; P=.011) and FairnessVice (t_{59}=–2.20; P=.03), with a marginal rise in PurityVice (t_{59}=–1.88; P=.07). In contrast, the control group exhibited no significant changes during the same time frame. Additionally, network analysis revealed that beyond cyberbullying experiences, core psychological characteristics in the network were neuroticism, conscientiousness, and Oxford Happiness.

Conclusions: By leveraging noninvasive retrospective social media data, this study provides novel insights into the short-term psychological impacts of cyberbullying during the pericyberbullying period. The findings highlight the need for timely interventions

XSL•FO
RenderX

focusing on enhancing survivors' happiness, reducing suicide risk, adjusting personality traits, and rebuilding moral cognition to mitigate the negative effects of cyberbullying.

(J Med Internet Res 2025;27:e64451) doi:10.2196/64451

KEYWORDS

cyberbullying; pericyberbullying period; social media; well-being; morality; suicide risk; personality traits

Introduction

Background

Cyberbullying is a growing global phenomenon, with the prevalence of cyberbullying experiences increasing significantly over the past 2 decades [1]. In today's highly interconnected world, anyone can become a survivor of cyberbullying [2]. Cyberbullying refers to the use of digital technologies to intentionally harm others, either directly or indirectly. When the survivor perceives the perpetrator's intent as deliberate and experiences harm, they are considered a survivor of cyberbullying [3].

Cyberbullying is a significant source of social stress [4]. According to the stress theory proposed by Lazarus and Folkman [5] and Selye's [6] general adaptation syndrome, individuals who perceive themselves as unable to cope with stress are more likely to experience negative psychological outcomes. In the context of cyberbullying, survivors often experience intense psychological stress in the short term, which can affect their happiness, suicide risk, personality traits, and perceptions of moral injustice in the external environment [7-19].

Although previous studies have demonstrated the negative psychological impacts of cyberbullying, they have largely relied on self-report questionnaire methods, which are susceptible to individual recall bias. Moreover, a few studies have investigated the psychological characteristics of individuals both before and after experiencing cyberbullying, which may obscure the effects of innate differences as well as the role of cyberbullying itself [20]. Additionally, existing longitudinal studies typically focus on the long-term impacts that include data collected prior to cyberbullying, with research time periods ranging from 6 months to 3 years [21]. Due to the unpredictability of cyberbullying, collecting a sufficient sample to assess its immediate psychological impacts presents significant challenges. To address this issue, this study draws inspiration from analogous terms in the medical field, such as the "perinatal period" and the "perioperative period," which describe key stages of change surrounding specific events or states [22]. Based on these concepts, the "pericyberbullying period" is defined as a critical time frame that encompasses the psychological fluctuations before the occurrence of a cyberbullying event, the immediate reactions during the event, and the adaptation process following its conclusion. The pericyberbullying period provides a new lens through which to explore the short-term and immediate psychological impacts of cyberbullying on survivors. It helps to reveal the dynamic psychological changes associated with cyberbullying with greater precision, while also offering theoretical support for the development of time-based intervention strategies. By leveraging the richness, openness, timeliness, and high traceability of social media data [23], this study enables a retrospective analysis of the psychological changes survivors experience during the pericyberbullying period. Previous studies have used social media data, such as the online ecological recognition method, to retrospectively investigate the psychological effects of negative events (eg, the loss of an only child and domestic violence) [20,24]. This study applies this method innovatively to cyberbullying research, further validating and expanding the potential of social media data in psychological studies.

Cyberbullying Experiences and Well-Being

Positive psychology posits that well-being, as a complex, multifaceted, and central positive experience, is a multidimensional psychological structure [25]. Subjective well-being (SWB), based on the hedonistic perspective, primarily includes classical indicators such as positive and negative emotions [26]. In contrast, psychological well-being (PWB), grounded in the eudaimonic perspective, comprises 6 dimensions: self-acceptance, life purpose, positive relationships, environmental mastery, personal growth, and autonomy [27]. Integrating both SWB and PWB, researchers have proposed the unidimensional Oxford Happiness Questionnaire [28]. High levels of well-being are closely associated with self-actualization, coping with life stress, and finding meaning in life [29]. Health is considered the foundation of happiness, as individuals can only experience higher levels of well-being when they are in good physical and mental condition [30].

Recent studies have found that cyberbullying experiences can predict various subsequent psychological outcomes, such as negative emotions, disrupted self-evaluation and value systems, negative interpersonal relationships, and loss of control, all of which can impair an individual's well-being [31,32]. Specifically, cyberbullying experiences are often accompanied by negative emotions. Research by Zhao et al [13] found that cyberbullying experiences can trigger low thresholds, high emotional intensity, and slow recovery to baseline negative emotions in adolescents. These emotions can mediate the relationship between cyberbullying experiences and non-suicidal self-injury, which is significantly negatively correlated with well-being [17]. When an individual is bullied, their self-evaluation system and sense of self-worth are disrupted, leading to a decreased level of self-acceptance and the development of a negative self-concept [9,11]. Self-acceptance is an important dimension of PWB. Research has found that self-acceptance is a predictor of psychological distress and can protect individuals from the effects of depression [13]. Fang et al [10] found that self-acceptance mediates the relationship between cyberbullying experiences and depression in college students.

XSL•FO
RenderX

Cyberbullying experiences also affect social behavior. Survivors of cyberbullying often experience high levels of social anxiety [33,34], and low self-esteem and self-acceptance can negatively impact interpersonal relationships [10]. Audrin and Blaya [35] found that factors such as the repetitive nature of web-based behavior, the anonymity of cyber aggressors, and the large audience can damage the well-being of cyberbullying survivors. The anonymity, large audience, and lack of spatial and temporal limitations of cyberbullying increase the difficulty of controlling these incidents, reducing individuals' psychological sense of control [36]. When traumatic events are uncontrollable, even individuals with high perceived control may experience posttraumatic stress disorder symptoms, while those with relatively low control are more likely to experience posttraumatic stress disorder symptoms [8].

Cyberbullying Experiences and Suicide Risk

Suicide risk is closely related to lower PWB [37]. According to the suicide stress theory [38], cyberbullying experiences increase suicide risk through several psychological mechanisms, including value conflict, the gap between ideals and reality, and relative deprivation stress [39]. Studies indicate that cyberbullying can significantly increase suicide-related behaviors, including suicidal ideation, especially in adolescents [7,12]. Longitudinal studies also show that early cyberbullying experiences can lead to later suicidal thoughts and behaviors, with core self-evaluation serving as a mediator between the two [40].

Cyberbullying Experiences and Personality Traits

Personality traits and life events are intertwined [24]. On one hand, personality differences exist before life events occur, leading to different probabilities of experiencing certain life events, known as the "selection effect" [41]. On the other hand, personality can change in response to life events, known as the "socialization effect" [42]. Significant life events, especially those occurring for the first time, may have a more pronounced impact, such as the death of a spouse, first-time unemployment, or first-time domestic violence, compared to similar later events [24].

Existing research primarily measures the relationship between personality and life events (eg, cyberbullying experiences) through relatively long (eg, 1 year or even 2 years), fixed-interval self-reports, and compares personality differences between groups with and without specific life events. However, the relationship between cyberbullying experiences and personality traits is unclear. It is uncertain whether certain personality traits lead to cyberbullying experiences or if cyberbullying leads to changes in personality traits. For example, a 2-year study on Dutch individuals aged 15 years and older found that compared to individuals who have not experienced cyberbullying, cyberbullying survivors had lower conscientiousness, higher openness and neuroticism, and no significant difference in agreeableness [18]. Ryff and Keyes [27] found that Turkish student survivors of cyberbullying were positively correlated with neuroticism, agreeableness, and extraversion, and negatively correlated with conscientiousness. A study on high school students found that personality traits do not seem to be predictors of cyberbullying experiences [20]. These conflicting results may be due to limitations in current methods of measuring personality change. For example, some studies may include survivors who have experienced prolonged cyberbullying during the measurement interval, with personality changes influenced by the continuous occurrence of cyberbullying rather than the event itself. Additionally, comparing cyberbullying and noncyberbullying groups to study the effects of cyberbullying makes it difficult to distinguish between the impact of cyberbullying and the inherent differences between the groups.

Cyberbullying Experiences and Morality

Cyberbullying is a traumatic event for survivors, who perceive unjust treatment in their social environment, thereby disrupting their belief in a just world [16]. This disruption may lead to changes in the survivors' moral perceptions of the external environment, causing them to perceive more unfairness. Cyberbullying experiences are a predictor of cyberbullying behavior [31]. When individuals perceive unjust treatment in their environment, they may engage in moral disengagement to alleviate guilt, making them more likely to engage in unethical behavior, such as retaliatory cyberbullying [43]. Moral behavior is typically driven by moral motivation; thus, the perception of unfairness following cyberbullying experiences may influence an individual's moral motivation and subsequently affect their future moral behavior.

Objectives

This study leverages social media data to explore the immediate psychological changes experienced by survivors during the pericyberbullying period, encompassing the time before and after their first encounter with cyberbullying. It comprehensively examines the short-term impacts of cyberbullying on multidimensional psychological characteristics, including well-being and its subdimensions (positive emotions, negative emotions, and self-acceptance), suicide risk and its related dimensions (shame and guilt, and self-regulation), personality traits (eg, agreeableness), and moral cognition of the external environment (eg, the FairnessVice in moral foundations and the communication in moral motivation). Additionally, it uses network analysis to investigate the interactions between cyberbullying experiences and psychological characteristics during this critical period, providing a scientific foundation for developing precise, time-sensitive intervention strategies.

Methods

Participants and Data Collection

This study uses data sourced from Sina Weibo, a social media platform developed by Sina Corporation, which had 605 million monthly active users by the end of September 2023. The collected data mainly include users' web-based behaviors (such as reposts and replies) and Weibo text content. During the data collection process, we adhered to privacy and ethical principles, ensuring that all data were anonymized and privacy was strictly protected [44].

We screened 60 cyberbullying survivors from a pool of Weibo users through a two-step process: (1) preliminary identification of users via public media reports or keyword searches, and (2)

XSL•FO
RenderX

manual verification of survivors and determination of the time of their first cyberbullying incident.

Step 1: Preliminary Identification of Users

We used 2 approaches to identify users. The first approach involved identifying Weibo users who had been publicly recognized as cyberbullying survivors through media reports. The second approach used keyword searches based on the definition and types of cyberbullying, combined with usage characteristics of the Weibo platform. Keywords included terms such as: “cyber violence” (网络暴力), “cyberbullying” (网暴), “abuse” (辱骂), “defamation” (诽谤), “rumors” (造谣), “insults” (污蔑), “doxxing” (人肉), “mocking” (嘲笑), “harassment” (骚扰), “exclusion” (排斥), “kicked out” (踢出), and “pranks” (恶作剧). Through this process, we initially identified 82 potential Weibo users who might have experienced cyberbullying.

Step 2: Manual Verification of Survivors and Determination of Cyberbullying Timing

We recruited 4 graduate students in psychology as research assistants. They were given detailed training on the definitions, types, and manifestations of cyberbullying. The Weibo posts of the 82 initially identified users were downloaded and anonymized. Each research assistant independently reviewed all posts from each user to determine whether they could be classified as a survivor and to identify the time of the user’s first cyberbullying incident. The inclusion criteria were (1) the user had experienced cyberbullying behaviors and (2) the user had expressed in their posts that they perceived those behaviors as harmful. The exclusion criteria were (1) inability to determine the time of the first cyberbullying incident, (2) inability to obtain Weibo posts before and after the first cyberbullying incident, and (3) Weibo users with strong public visibility and influence, such as celebrities. A user was only included as a survivor if at least 3 research assistants agreed on their classification. Ultimately, 60 users were included in the survivor group.

For the control group, we randomly selected active Weibo users (with more than 500 posts) who had not experienced cyberbullying and matched them with the survivors. To ensure effective matching, we used user profiles to match sex, age, location, and number of followers, ensuring no significant differences between the control and survivor groups. The screening and matching process was conducted collaboratively by the same 4 research assistants.

We used the application programming interface provided by Sina Weibo to download the text data of the 120 Weibo users.

Methods and Measurements

Online ecological recognition (OER) is a technique that uses publicly available web-based social media data combined with machine learning methods to measure corresponding psychological characteristics without contacting participants. Compared to self-reports, OER does not rely on the participants’ self-report results or their cooperation with the experiment, and it does not require any contact with the participants. This makes it a more convenient, safer, and timely method for tracking psychological status [20]. OER mainly involves 3 steps: Chinese word segmentation, lexical feature extraction, and model prediction [45,46]. The predictive models are tools developed based on big data and deep learning techniques for web-based psychological research, and their feasibility has been repeatedly verified by multiple studies [47]. Therefore, this study uses the Oxford Happiness model, SWB model, and PWB model to predict individuals’ happiness; the suicide dictionary to predict individuals’ suicide risk; the personality prediction model to predict individuals’ Big 5 personality traits; the Moral Foundations Dictionary (MFD) and Moral Motivation Dictionary (MMD) to predict moral perceptions; and the Simplified Chinese version of Linguistic Inquiry and Word Count (SC-LIWC) to supplement the prediction of psychological traits.

The models and dictionaries selected for this study are as follows:

- Oxford Happiness model: This study uses this model to predict Oxford Happiness Inventory scores [28], aiming to assess overall levels of individual happiness. Wang et al [48] selected 548 Weibo users as their sample and divided them into high- and low-happiness groups based on their Oxford Happiness Inventory scores. They extracted linguistic and behavioral features from social media data and used a decision tree model to predict participants' happiness levels, achieving an accuracy rate of 0.68.
- SWB model: This model is used to predict emotional well-being levels measured by the Positive and Negative Affect Schedule. Hao et al [49] developed and validated this model using social media data from 1785 Weibo users. The optimal evaluation metrics for the model’s effectiveness in predicting SWB showed a strong correlation level of 0.60.
- PWB model: This model is used to predict scores on the PWB Scale by Ryff and Keyes [27], covering dimensions such as autonomy, environmental mastery, personal growth, positive relations with others, life purpose, and self-acceptance. Han et al [50] developed and validated these machine learning models using the textual features of 1427 Weibo users. The correlation between predicted and self-reported PWB scores was significantly positive across all 6 dimensions.
- Personality prediction model: The study uses this model to estimate scores on the Big 5 personality dimensions: openness, agreeableness, extraversion, conscientiousness, and neuroticism. Han et al [51] developed and validated this machine learning model using textual features from 3886 Weibo users. The model demonstrated good structural and criterion validity, with 5-fold cross-validation results ranging from 0.44 to 0.48 (P<.001). The correlation coefficients for the 5 personality traits between the 2 split datasets ranged from 0.84 to 0.88 (P<.001).
- Chinese Suicide Dictionary: This study uses this dictionary to predict suicide risk. The dictionary is designed to identify suicide risk on social media, comprising 2168 words divided into 13 categories (eg, suicidal ideation words, hopelessness words, and personality words) [52].
- Moral Dictionary: This dictionary is used to measure changes in the survivors’ morality. It includes MFD [53] and MMD [54]. The simplified Chinese version of the MFD

XSL•FO
RenderX

includes 590 words or phrases across 6 dimensions (harm, fairness, loyalty, authority, purity, and general morality), with positive and negative word lists for all dimensions except general morality. The Chinese version of the MMD includes 2 dimensions: agency and communication. The agency includes 690 words (eg, achievement, failure, and expenditure), while communication includes 260 words (eg, acceptance, care, and kindness).

- SC-LIWC: In addition to the earlier models and dictionaries, this study uses SC-LIWC to supplement the analysis of psychological changes in cyberbullying survivors. SC-LIWC is widely used in natural language processing, covering multiple dimensions such as linguistic processes, psychological processes, and personal concerns [55]. This study specifically focuses on the subcategories of "I," "tPast," " tFuture," "Anger," "CogMech," "Insight," "Cause," and "Certain." For example, "I" reflects the degree of self-focus; "Anger" captures the intense emotional responses elicited by bullying; "tPast" and "tFuture" reveal the survivor's immersion in past experiences and concerns about the future. Changes in these dimensions help elucidate the impact of cyberbullying on survivors' well-being, suicide risk, and personality traits. Additionally, "CogMech" and "Insight" reflect the survivor's cognitive processing and self-reflection, while "Cause" and "Certain" indicate the survivor's perception of causality and uncertainty regarding future events. The analysis of these dimensions further enhances our understanding of the effects of cyberbullying on well-being, suicide risk, personality traits, and moral judgment.

Data Analyses

The study uses the starting time of the initial cyberbullying incident as the dividing line to measure psychological characteristics before and after this event. The first measurement uses Weibo data from 3 months before the cyberbullying incident, and the second measurement uses data from 3 months after the incident. Given that the timing of the initial cyberbullying incident may vary among users, we estimated each survivor's mental health status precisely three months before and after the event. The control group, consisting of individuals who have not experienced cyberbullying, was analyzed during the same period, with the timeline matched one-to-one.

We conducted descriptive statistics, independent sample 2-tailed *t* tests, and paired sample 2-tailed *t* tests using SPSS (version 27.0; IBM Corp). Initially, we analyzed the between-group differences between the survivor group (n=60) and the control group using independent sample *t* tests. Then, we compared the changes in psychological characteristics at T1 (3 months before the cyberbullying incident) and T2 (3 months after the incident) for both groups using paired sample 2-tailed *t* tests. All analyses were conducted at a significance level of α=.05, and effect sizes of mean differences were reported as *d*.

We used R (version 4.2.2; R Foundation for Statistical Computing) to perform network analysis on the predictive psychological characteristics and cyberbullying experiences at T2, estimating, visualizing, and evaluating the network.

Ethical Considerations

This study was conducted in accordance with ethical guidelines. The research involved the analysis of anonymized social media data, ensuring that no personally identifiable information was retained.

The study received ethical approval from the institutional review board of Central China Normal University, with the approval number (CCNU-IRB-202311046b), confirming that the research adhered to ethical standards for human subjects research.

Given the nature of the study, which involved the analysis of publicly available, anonymized data, informed consent was not required. The original data collection processes were conducted in compliance with relevant ethical guidelines, and the secondary analysis of this data was covered under the original ethical approval.

All data used in this study were anonymized to protect participant privacy. No personally identifiable information was included in the analysis, and stringent measures were taken to ensure data confidentiality.

No compensation was provided to participants, as the study utilized existing anonymized data and did not involve direct interaction with participants.

Results

User Statistics

Overview

Among the users who registered their birth date in their profiles (N=120; survivor group n=60, 50%; control group n=60, 50%), 78% (94/120) were female, with ages ranging from 15 to 41 years, and an average age of 28 (SD 6.23) years. The sample included 1 user who died by suicide and another who died of cancer. Table 1 presents the demographic characteristics of these groups.

XSL•FO
RenderX

Table 1. Participant demographic characteristics.

	Total (N=120), n (%)	Survivor group (n=60), n (%)	Control group (n=60), n (%)
Number of deaths	2 (2)	2 (3)	0 (0)
Sex			
Male	26 (22)	13 (22)	13 (22)
Female	94 (78)	47 (78)	47 (78)
Age (years)			
Less than 18	6 (5)	3 (5)	3 (5)
18-29	84 (70)	42 (70)	42 (70)
30-39	22 (18)	11 (18)	11 (18)
40-49	8 (7)	4 (7)	4 (7)

The Impact of Cyberbullying on Psychological Characteristics

To investigate the impact of cyberbullying on mental health, we first compared the scores of Oxford Happiness, SWB, PWB, suicide risk, personality traits, and moral perception between the survivor group and the control group at T1. Then, we separately compared the scores of the survivor group and the control group between T1 and T2. At T1, we found no significant differences between the survivor group and the control group in Oxford Happiness, SWB, PWB, suicide risk, personality traits, or moral perception. Specifically, for personality traits, no significant differences were observed in agreeableness (t_{118}=–0.77; P=.45), conscientiousness (t_{118}=–0.03; P=.98), extraversion (t_{118}=–0.72; P=.47), openness (t_{118}=0.34; P=.73), or neuroticism (t_{118}=1.17; P=.25). Additionally, the control group showed no significant changes between T1 and T2. In the survivor group, significant reductions were observed after cyberbullying in Oxford Happiness (t_{59}=2.14; P=.04), positive emotions (t_{59}=2.72; P=.009), positive relations with others (t_{59}=2.54; P=.01), purpose in life (t_{59}=2.28; P=.03), and self-acceptance (t_{59}=2.08; P=.04). Environmental mastery showed a marginally significant decrease (t_{59}=1.72; P=.09), while negative emotions significantly increased (t_{59}=–2.89; P=.005). The “Anger” of SCL-LIWC also significantly increased (t_{59}=–3.46; P=.001). In the cognitive processes category, the “Certain” of SCL-LIWC significantly increased (t_{59}=–2.82; P=.006), while “Insight” words (t_{59}=–1.69; P=.096) and “Cause” words (t_{59}=–1.77; P=.08) showed marginally significant increases. Regarding the Chinese suicide dictionary, self-regulation showed a marginally significant decrease (t_{59}=1.75; P=.09), while shame and guilt (t_{59}=–1.71; P=.09) and anger and hostility (t_{59}=–1.80; P=.08) marginally increased. Among the Big 5 personality traits, agreeableness (t_{59}=2.79; P=.007), extraversion (t_{59}=2.26; P=.03), and conscientiousness (t_{59}=2.27; P=.03) significantly decreased, while neuroticism significantly increased (t_{59}=–3.42; P=.001). In the MMD and MFD, communication words (t_{59}=–2.62; P=.011) and FairnessVice (t_{59}=–2.20; P=.03) significantly increased, while PurityVice showed a marginal increase (t_{59}=–1.88; P=.07; Table 2).

XSL•FO
RenderX

Table 2. The difference in psychological characteristics between the survivor group and the control group at T2 and T1.

Indexes and dimensions	Survivor group (n=60)			Control group (n=60)		
	Mean (SD)	*t* test (*df*)	*P* value	Mean (SD)	*t* test (*df*)	*P* value
Oxford Happiness						
Oh		2.14 (59)[a]	.04		0.77 (59)	.45
Before	87.98 (4.22)			88.01 (4.42)		
After	86.99 (3.12)			87.67 (3.30)		
Subjective well-being						
Positive emotion		2.72 (59)[b]	.009		0.68 (59)	.50
Before	21.84 (0.91)			21.90 (0.99)		
After	21.27 (0.68)			21.82 (0.81)		
Negative emotion		–2.90 (59)[b]	.005		–0.71 (59)	.48
Before	14.54 (1.50)			14.48 (1.43)		
After	14.98 (1.26)			14.57 (1.04)		
Psychological well-being						
Environmental mastery		1.73 (59)[c]	.089		–0.54 (59)	.59
Before	11.55 (0.68)			11.46 (0.69)		
After	11.46 (0.66)			11.50 (0.54)		
Positive relations with others		2.54 (59)[a]	.01		0.77 (59)	.45
Before	12.43 (0.46)			12.54 (0.44)		
After	12.28 (0.38)			12.50 (0.39)		
Autonomy		0.68 (59)	.50		–0.62 (59)	.54
Before	11.30 (0.76)			11.24 (0.80)		
After	11.25 (0.69)			11.29 (0.69)		
Personal growth		1.16 (59)	.25		–0.02 (59)	.99
Before	15.07 (0.31)			15.01 (0.36)		
After	15.03 (0.33)			15.01 (0.31)		
Self-acceptance		2.08 (59)[a]	.04		0.37 (59)	.72
Before	10.10 (0.68)			10.05 (0.72)		
After	9.98 (0.64)			10.03 (0.59)		
Purpose in life		2.28 (59)[a]	.03		0.18 (59)	.86
Before	12.85 (0.61)			12.81 (0.61)		
After	12.72 (0.53)			12.79 (0.51)		
Personality						
Agreeableness		2.79 (59)[b]	.007		0.61 (59)	.55
Before	32.71 (0.84)			32.81 (0.78)		
After	32.36 (0.98)			32.76 (0.73)		
Conscientiousness		2.27 (59)[a]	.03		0.58 (59)	.56
Before	28.833 (1.28)			28.84 (1.13)		
After	28.47 (1.19)			28.76 (1.13)		
Extraversion		2.26 (59)[a]	.03		1.58 (59)	.12
Before	23.30 (1.09)			23.43 (0.94)		

XSL•FO
RenderX

Indexes and dimensions	Survivor group (n=60)			Control group (n=60)		
	Mean (SD)	*t* test (*df*)	*P* value	Mean (SD)	*t* test (*df*)	*P* value
After	22.99 (0.93)			23.23 (1.01)		
Openness to experience		1.66 (59)	.10		0.53 (59)	.60
Before	35.85 (0.97)			35.79 (0.99)		
After	35.63 (1.09)			35.73 (0.99)		
Neuroticism		–3.42 (59)[d]	.001		–0.97 (59)	.33
Before	24.53 (1.16)			24.28 (1.23)		
After	25.03 (1.12)			24.40 (1.31)		
SC-LIWC[e]						
I		–2.54 (59)[a]	.01		0.73 (59)	.46
Before	0.020 (0.10)			0.0138 (0.0098)		
After	0.023 (0.012)			0.013 (0.0087)		
tPast		–1.77 (59)[c]	.08		–0.62 (59)	.54
Before	0.0016 (0.0014)			0.0018 (0.0022)		
After	0.0020 (0.0015)			0.0022 (0.0058)		
tFuture		–2.14 (59)[a]	.04		–1.68 (59)	.10
Before	0.00078 (0.00074)			0.00086 (0.0010)		
After	0.0011 (0.0011)			0.0013 (0.0018)		
Anger		–3.46 (59)[d]	.001		0.86 (59)	.39
Before	0.0019 (0.0014)			0.0029 (0.0035)		
After	0.0037 (0.0033)			0.0025 (0.0027)		
CogMech		–2.30 (59)[a]	.02		–0.56 (59)	.57
Before	0.11 (0.032)			0.102 (0.0311)		
After	0.12 (0.020)			0.104 (0.027)		
Insight		–1.69 (59)[c]	.096		–0.38 (59)	.71
Before	0.012 (0.0062)			0.011 (0.0056)		
After	0.014 (0.0054)			0.011 (0.0061)		
Cause		–1.77 (59)[c]	.08		0.89 (59)	.38
Before	0.0075 (0.0042)			0.0065 (0.0043)		
After	0.0086 (0.0044)			0.0059 (0.0042)		
Certain		–2.82 (59)[b]	.006		0.72 (59)	.48
Before	0.010 (0.0055)			0.015 (0.032)		
After	0.12 (0.0058)			0.011 (0.006)		
Chinese Suicide Dictionary						
Self-regulation		1.75 (59)[c]	.09		0.17 (59)	.87
Before	0.00073 (0.0012)			0.00043 (0.00085)		
After	0.00045 (0.0007)			0.00041 (0.00084)		
Shame and guilt		–1.71 (59)[c]	.09		0.21 (59)	.84
Before	0.00050 (0.00066)			0.00071 (0.0011)		
After	0.00081 (0.0012)			0.00067 (0.0009)		

XSL•FO
RenderX

Indexes and dimensions	Survivor group (n=60)			Control group (n=60)		
	Mean (SD)	*t* test (*df*)	*P* value	Mean (SD)	*t* test (*df*)	*P* value
Anger and hostility		–1.80 (59)[c]	.08		0.17 (59)	.87
Before	0.0015 (0.0014)			0.0011 (0.0020)		
After	0.0022 (0.0034)			0.0010 (0.0013)		
Moral Motivation Dictionary						
C ommunication		–2.62 (59)[a]	.011		1.55 (59)	.13
Before	0.0052 (0.0032)			0.0046 (0.0041)		
After	0.0063 (0.0040)			0.0036 (0.0031)		
Moral Foundations Dictionary						
FairnessVice		–2.20 (59)[a]	.03		0.82 (59)	.41
Before	0.000091 (0.00021)			0.000080 (0.00026)		
After	0.00029 (0.00072)			0.000048 (0.00017)		
PurityVice		–1.87 (59)[c]	.07		–1.27 (59)	.21
Before	0.00034 (0.00053)			0.00015 (0.00036)		
After	0.00065 (0.0012)			0.00024 (0.00045)		

[a]There was a significant difference existing with $P<.05$.

[b]There was a significant difference existing with $P<.01$.

[c]There was a significant difference existing with $P<.1$.

[d]There was a significant difference existing with $P<.001$.

[e]SC-LIWC: Simplified Chinese version of Linguistic Inquiry and Word Count.

Network Analysis Results

Network Structure

Figure 1 presents the partial correlation network of the 20 study variables, regularized using the least absolute shrinkage and selection operator. Among the total of 96 edges, 60 were significant, with 35 being positively significant and 25 negatively significant. Network sparsity is 0.68. The edge with the highest weight was between “environmental mastery” and “self-acceptance,” followed by the edge between “personal growth” and “purpose in life,” and the edge between “Oxford Happiness” and “positive relations with others.” The 5 nodes most strongly connected to the “cyberbullying experiences” node were “FairnessVice,” “anger and hostility,” “communication,” “neuroticism,” and “Oxford Happiness,” in descending order of strength.

XSL•FO
RenderX

Figure 1. Network structure diagram. ICU1_oh_s: Oxford Happiness; ICU2_pp_s: positive emotion; ICU3_nn_s: negative emotion; lCU4_em_s: environmental mastery; lCU5_pr_s: positive relations with others; ICU6_au_s: autonomy; lCU7_pg_s: personal growth; lCU8_sa_s: self-acceptance; lCU9_pl_s: purpose in life; ICU10_FairnessVice_bs: FairnessVice; ICU11_COMM_bs: communication; ICU12_C5_bs: self-regulation; ICU13_C9_bs: shame and guilt; ICU14_C10_bs: anger and hostility; ICU15_A_s: agreeableness; ICU16_C_s: conscientiousness; ICU17 _E_s: extraversion; ICU18_O_s: openness to experience; ICU19_N_s: neuroticism; ICU20_CV: cyberbullying experiences.

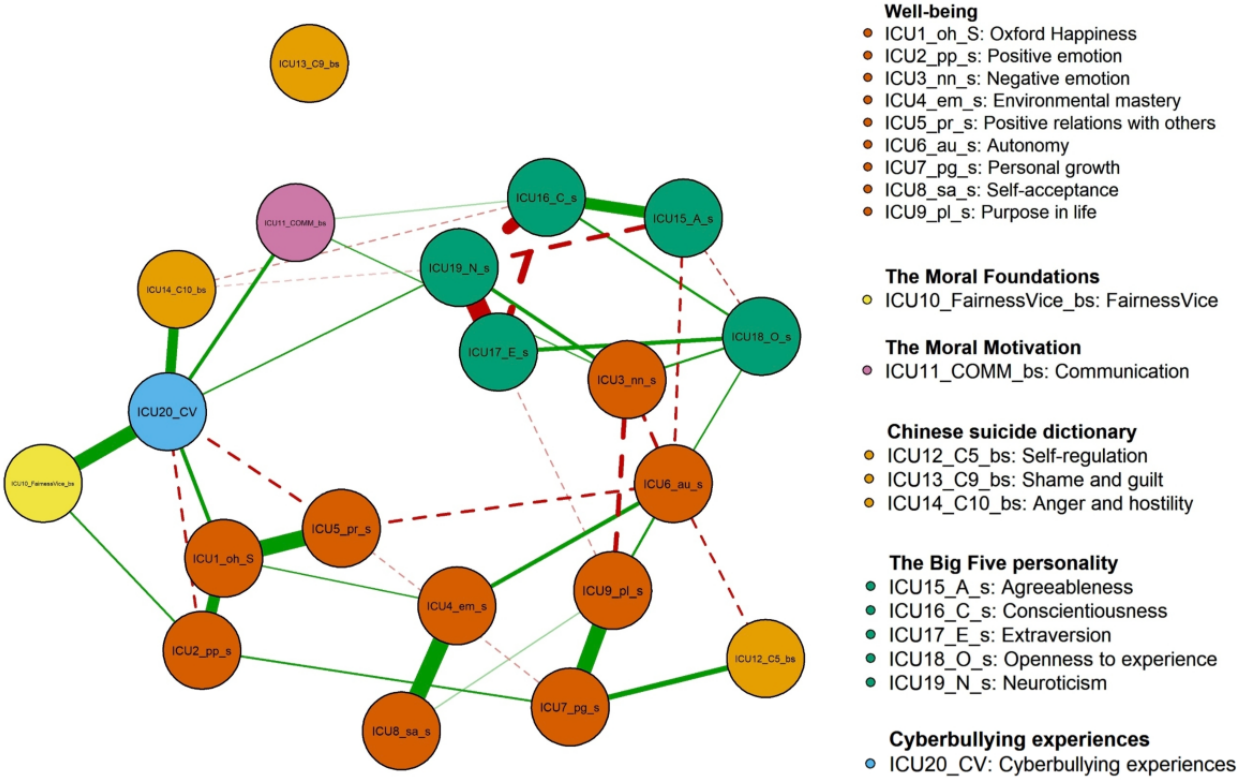

Node Centrality

We report strength centrality, given that betweenness and closeness centrality measures are particularly ill-suited to psychological networks [56]. Figure 2 displays the strength of each node in the network. “Shame and guilt” is the node with the lowest strength, while “cyberbullying experiences” has the highest strength centrality, followed by “neuroticism,” “conscientiousness,” and “Oxford Happiness.” The strength difference test revealed significant differences in strength centrality between “cyberbullying experiences” and “Oxford Happiness,” “autonomy,” “purpose in life,” “conscientiousness,” “extraversion,” and “neuroticism.” The strength centrality of “neuroticism” significantly differed from that of “positive emotions,” “environmental mastery,” “personal growth,” “self-acceptance,” “FairnessVice,” “self-regulation,” “shame and guilt,” “anger and hostility,” “openness to experience,” and “cyberbullying experiences.” Similarly, “Oxford Happiness” exhibited significant strength differences with “positive emotions,” “self-acceptance,” “FairnessVice,” “shame and guilt,” “anger and hostility,“ and “cyberbullying experiences.” The strength of “autonomy” significantly differed from “FairnessVice,” “self-regulation,” “shame and guilt,” “anger and hostility,” “openness to experience,” and “cyberbullying experiences.” Since the strength consistency score (0.517) exceeds the recommended minimum value of 0.25, the reliability of this centrality index is acceptable [57].

XSL•FO
RenderX

Figure 2. Strength centrality of each node in the network. ICU1_oh_s: Oxford Happiness; ICU2_pp_s: positive emotion; ICU3_nn_s: negative emotion; lCU4_em_s: environmental mastery; lCU5_pr_s: positive relations with others; ICU6_au_s: autonomy; lCU7_pg_s: personal growth; lCU8_sa_s: self-acceptance; lCU9_pl_s: purpose in life; ICU10_FairnessVice_bs: FairnessVice; ICU11_COMM_bs: communication ; ICU12_C5_bs: self-regulation; ICU13_C9_bs: shame and guilt; ICU14_C10_bs: anger and hostility; ICU15_A_s: agreeableness; ICU16_C_s: conscientiousness; ICU17 _E_s: extraversion; ICU18_O_ s: openness to experience; ICU19_N_s: neuroticism; ICU20_CV: cyberbullying experiences.

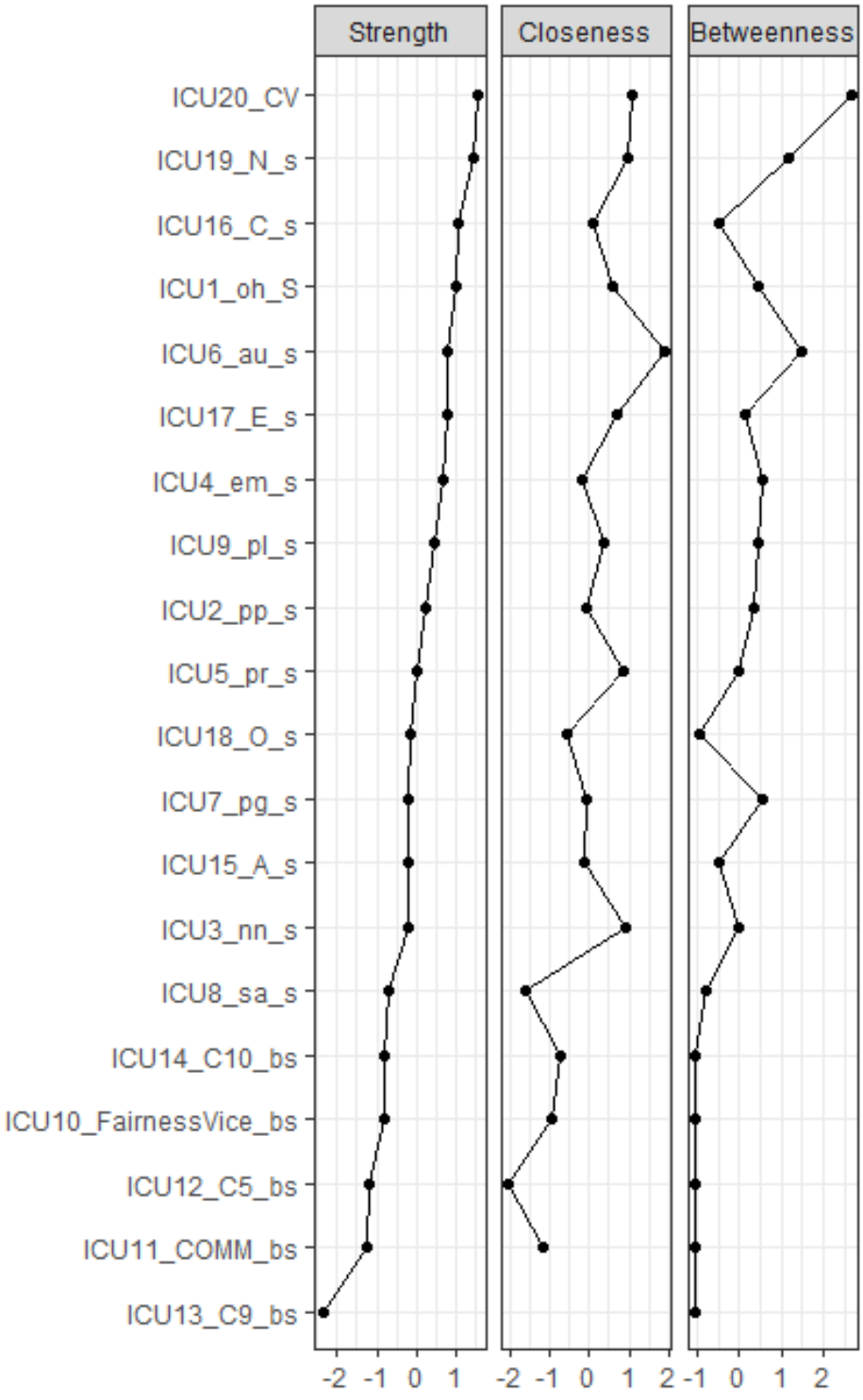

Discussion

Principal Findings

This study combines the OER method with network analysis to systematically explore the immediate impacts of cyberbullying events on survivors' internal psychological characteristics and perceptions of the external environment during the pericyberbullying period. The psychological predictive models used in the study, such as Oxford Happiness, have been shown to have good accuracy and precision [23,49-51]. Our research findings indicate that cyberbullying

XSL•FO
RenderX

has a significant impact on survivors' psychology within 3 months.

After experiencing cyberbullying, survivors exhibit a significant decrease in internal well-being and positive emotions, while negative emotions significantly increase. They also feel more shame, guilt, anger, and hostility. In terms of personality, there are significant changes in all dimensions of the Big 5 personality traits except for openness. Additionally, survivors perceive more social injustice from the external world, with a notable increase in the FairnessVice in moral foundations and communication words in moral motivation. In contrast, control groups do not show significant changes in well-being, emotions, personality, and morality. Consistent with existing research, after accounting for potential preexisting differences between the control and survivor groups, our results reveal the negative impact of cyberbullying on psychological characteristics [16,40,58-63]. The results of the network analysis indicate that aside from cyberbullying experiences, neuroticism, conscientiousness, and Oxford Happiness exhibit the strongest strength centrality among psychological characteristics, suggesting their critical roles in shaping an individual's overall psychological state. The high centrality of these nodes implies that altering personality traits and enhancing happiness might be the keys to mitigating the negative impacts of cyberbullying. Previous research has shown that interventions enhancing individual well-being can effectively alleviate the adverse effects of violence experiences [64]. These findings suggest that neuroticism, conscientiousness, and Oxford Happiness serve as core connectors within the psychological network, linking various other psychological variables. Therefore, they play a pivotal role in structuring the network between cyberbullying experiences and psychological characteristics.

Survivors of cyberbullying exhibit marginally significant changes in dimensions related to self-regulation of suicide risk, shame and guilt, and anger and hostility. These emotional and behavioral changes indicate that cyberbullying impacts survivors' emotional regulation capabilities, increasing their negative emotional experiences [65]. Although these marginally significant changes are not as strong as significant changes, they still suggest that cyberbullying has potentially negative effects on survivors' mental health and presents a certain suicide risk, warranting further attention and research. The increase in shame and guilt may be related to a decline in self-worth [66], while the increase in anger and hostility may be a reaction to perceived injustice [16].

Regarding the impact of cyberbullying on survivors' Big 5 personality traits, this study found no significant differences in personality traits between the survivor and the control groups before the occurrence of cyberbullying. This indicates that no specific personality traits predisposed individuals to become survivors of cyberbullying, supporting the absence of a "selection effect" [41] and the notion that anyone can become a survivor of cyberbullying [2]. However, after the occurrence of cyberbullying, survivors exhibited significant changes in their personality traits, with agreeableness, extraversion, and conscientiousness significantly decreasing, while neuroticism significantly increased. These changes reflect the presence of a "socialization effect" [42], indicating that individuals' personality traits are significantly influenced by the experience of cyberbullying. After experiencing cyberbullying, survivors' agreeableness and extraversion significantly decreased, suggesting that cyberbullying may cause survivors to withdraw more from social activities and even develop social anxiety disorders [67]. A decrease in conscientiousness could negatively affect their performance at work and school [68], while an increase in neuroticism indicates that survivors become more vulnerable and sensitive in terms of emotional stability [65]. The lack of significant differences in openness before and after cyberbullying may be attributed to the fact that openness primarily reflects an individual's interest in new experiences, art, aesthetics, and abstract thinking, as a more stable trait [69] that is less strongly influenced by short-term environmental factors or external events, such as cyberbullying experiences.

In terms of moral perception of the external environment, the study results show a significant increase in FairnessVice and a marginally significant increase in PurityVice in survivors' moral foundations, reflecting an increased sensitivity to social injustice. As found by Bai et al [16], cyberbullying experiences may cause individuals' belief in a just world to collapse. Additionally, there was a significant increase in communicative motivation within survivors' moral motivations, indicating that survivors may be more inclined to seek support and understanding in interpersonal relationships after experiencing cyberbullying. For example, studies have shown that although survivors may be reluctant to seek help from authorities, they are more likely to seek help from peers [70,71] or web-based support groups [72,73].

The findings of this study have significant practical implications for developing cyberbullying intervention strategies. By focusing on the pericyberbullying period, we can more accurately identify the immediate psychological impacts of cyberbullying on survivors and design more timely and effective intervention measures accordingly. Given the high centrality of nodes such as neuroticism, and Oxford Happiness, these characteristics should be prioritized in interventions aimed at promoting psychological recovery. Although the predictive power of centrality measures has certain limitations for instance, some studies suggest that targeting symptoms based on centrality may be less effective than addressing the most common symptoms [56], strength centrality remains a useful heuristic for identifying key nodes within the network [74]. As core components of the psychological network, improving traits such as Oxford Happiness can generate cascading effects, positively impacting the overall mental health of survivors. This intervention strategy aligns with the principles of positive psychology [25], which emphasizes the importance of happiness, and positive emotions in building psychological resilience and recovering from trauma. Interventions designed to enhance PWB such as acceptance and commitment therapy, have been validated in the recovery of violence survivors, demonstrating their effectiveness in promoting psychological healing [75], and acceptance and commitment therapy have also been shown to effectively reduce neuroticism levels in individuals [76]. By focusing on these key traits, interventions can equip survivors to better manage the emotional challenges posed by cyberbullying and restore a sense of purpose and control in their lives. Regarding changes in the

XSL•FO
RenderX

Big 5 personality traits, interventions should focus on increasing agreeableness, extraversion, and conscientiousness in survivors, while mitigating the negative effects of neuroticism. Additionally, to address the increased perception of injustice in the external environment, education and counseling should emphasize helping survivors rebuild their belief in a just world, thereby reducing their negative reactions to cyberbullying. In the future, positive psychology could be incorporated into school curricula to help students develop essential emotional regulation skills for coping with cyberbullying. Schools could also implement group activities or mentorship programs to encourage positive social interactions and establish strong peer and social support systems. Regarding web-based platform management and interventions, platforms could first proactively identify survivors and then offer personalized web-based interventions from a positive psychological perspective to support survivors of cyberbullying.

Limitations and Future Work

The limitations of this study include a relatively small sample size, with only 60 survivors, although each participant's Weibo data is substantial. The data source is also limited; to control confounding variables, only survivors from the Weibo platform were studied. Due to the limited personal profile information available on Weibo, we did not consider economic status when matching the experimental group of cyberbullying survivors and the control group of individuals who have not experienced cyberbullying and instead matched participants based on sex, age, location, and number of followers. Additionally, to ensure sufficient Weibo data for accurate predictions of psychological traits, only active users (defined as users with more than 500 posts) were included. Future research could explore the differentiated experiences of cyberbullying across different platforms, varying levels of user activity, and economic status. Additionally, because social media activity decreases for survivors after experiencing cyberbullying [77], we chose a 3-month study period to obtain more data for more accurate predictions of changes in individual psychological characteristics, despite this being a relatively long time frame. Future research could further validate the impact of cyberbullying on survivors' psychological characteristics by expanding the sample size, incorporating multiple data sources, and examining timely changes in internal psychological characteristics and external perceptions over shorter periods. Moreover, exploring changes in the psychological characteristics of cyberbullying survivors from different cultural backgrounds and age groups, as well as the effects of different types of cyberbullying on psychological characteristics, will also be important directions for future research.

Conclusions

In conclusion, cyberbullying has a significant negative impact on survivors' internal psychological characteristics and their perceptions of the external environment. This study leverages social media data and combines OER with network analysis to systematically reveal the dynamic psychological changes in survivors during the pericyberbullying period, before and after the cyberbullying events, as well as the interplay between cyberbullying experiences and psychological characteristics. The findings deepen our understanding of the short-term psychological impacts of cyberbullying and provide a critical theoretical foundation and practical reference for the development of future intervention strategies. Future research and interventions should focus on effectively reducing survivors' neuroticism, enhancing their well-being, and rebuilding their moral cognition to mitigate the negative effects of cyberbullying.

Acknowledgments

This study could not have been conducted without the valuable contributions of all the participants. The authors would also like to extend our sincere gratitude to Tingshao Zhu's team for their invaluable guidance in data analysis and research methodology. Additionally, the authors are grateful to their colleagues and the institutions involved in participant recruitment and data collection for their essential support. This study is supported by the major program of the National Social Science Foundation of China (grant 22&ZD187). During the writing process of this article, ChatGPT was only used for polishing and translating between Chinese and English, but it was not allowed to conceive ideas or directly write the content of the article.

Data Availability

Due to privacy concerns, the data used in this study cannot be directly shared in open repositories. However, readers can contact the authors via email to request access to the data.

Authors' Contributions

XL was primarily responsible for the conceptualization, methodology, project administration, writing–original draft, and writing–review & editing. ML contributed to the data curation, validation, investigation, writing–original draft, and writing–review & editing. XK contributed to the formal analysis and writing–review & editing. YL contributed to the formal analysis, and writing–review & editing. NH assisted with parts of the resources and visualization. ZR was responsible for the conceptualization, funding acquisition, supervision, and writing–review & editing.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Iffland B, Bartsch LM, Kley H, Neuner F. Growing relevance of reports of adolescent cyberbullying victimization among adult outpatients. BMC Public Health 2023;23(1):1503 [FREE Full text] [doi: 10.1186/s12889-023-16342-y] [Medline: 37553618]
2. Shoeibi N, Shoeibi N, Julian V, Ossowski S, Arrieta AG, Chamoso P. Smart cyber victimization discovery on Twitter. In: Corchado JM, Trabelsi S, editors. Sustainable Smart Cities and Territories. Cham, Switzerland: Springer; 2022.
3. Alipan A, Skues JL, Theiler S, Wise L. Defining cyberbullying: a multifaceted definition based on the perspectives of emerging adults. Int Journal of Bullying Prevention 2020;2(2):79-92. [doi: 10.1007/s42380-019-00018-6]
4. Martínez-Monteagudo MC, Delgado B, Díaz-Herrero Á, García-Fernández JM. Relationship between suicidal thinking, anxiety, depression and stress in university students who are victims of cyberbullying. Psychiatry Research 2020;286:112856. [doi: 10.1016/j.psychres.2020.112856] [Medline: 32062285]
5. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. City: Springer Publishing Company; 1984.
6. Selye H. A syndrome produced by diverse nocuous agents. Nature 1936;138(3479):32. [doi: 10.1038/138032a0]
7. Planellas Kirchner I, Calderon Garrido C. Do cybervictimizations predict suicide-related behaviors in adolescents? Mediating role of the "escaping" coping strategy. J Interpers Violence 2025:40(5-6):1015-1036. [doi: 10.1177/08862605241256384] [Medline: 38855834]
8. Chen H, Li Y, Xiong J, Yu J, Wu T. Cyberbullying victimization and post-traumatic stress disorder symptoms among college students: mediating role of negative coping and moderating role of perceived control. Curr Psychol 2024;43(21):19294-19303. [doi: 10.1007/s12144-024-05760-9]
9. Albarracin M, Bouchard-Joly G, Sheikhbahaee Z, Miller M, Pitliya RJ, Poirier P. Feeling our place in the world: an active inference account of self-esteem. Neuroscience of Consciousness 2024;2024(1):niae007 [FREE Full text] [doi: 10.1093/nc/niae007] [Medline: 38562605]
10. Fang Z, Zhang TT, Li X. Cyberbullying and college students' depression: the role of self-acceptance and dormitory interpersonal relationship. Journal of Shenyang Medical College 2024;26(3):336-247. [doi: 10.16753/j.cnki.1008-2344.2024.03.004]
11. Mullan VMR, Golm D, Juhl J, Sajid S, Brandt V. The relationship between peer victimisation, self-esteem, and internalizing symptoms in adolescents: a systematic review and meta-analysis. PLoS One 2023;18(3):e0282224 [FREE Full text] [doi: 10.1371/journal.pone.0282224] [Medline: 36989220]
12. Dewi HA, Suryani S, Sriati A. The relationship between cyberbullying and the risk of suicidal ideation in adolescents. Jurnal Keperawatan Padjadjaran 2023;11(2):118-123. [doi: 10.24198/jkp.v11i2.2259]
13. Zhao H, Gong X, Huebner ES, Yang X, Zhou J. Cyberbullying victimization and nonsuicidal self-injury in adolescents: testing a moderated mediating model of emotion reactivity and dispositional mindfulness. Journal of Affective Disorders 2022;299:256-263. [doi: 10.1016/j.jad.2021.11.070] [Medline: 34890693]
14. Chu X, Yang S, Sun Z, Jiang M, Xie R. The association between cyberbullying victimization and suicidal ideation among Chinese college students: the parallel mediating roles of core self-evaluation and depression. Front Psychiatry 2022;13:929679 [FREE Full text] [doi: 10.3389/fpsyt.2022.929679] [Medline: 35845449]
15. Padır MA, Ayas T, Horzum MB. Examining the relationship among internet parental style, personality, and cyberbullying/victimization. International Journal of Technology in Education and Science 2021;5(1):56-69. [doi: 10.46328/ijtes.160]
16. Bai Q, Huang S, Hsueh F, Zhang T. Cyberbullying victimization and suicide ideation: a crumbled belief in a just world. Computers in Human Behavior 2021;120:106679. [doi: 10.1016/j.chb.2021.106679]
17. Lung FW, Shu BC, Chiang TL, Lin SJ. Relationships between internet use, deliberate self-harm, and happiness in adolescents: a Taiwan birth cohort pilot study. PLoS One 2020;15(7):e0235834 [FREE Full text] [doi: 10.1371/journal.pone.0235834] [Medline: 32649690]
18. Weijer SG, Leukfeldt ER. Big five personality traits of cybercrime victims. Cyberpsychology, Behavior, and Social Networking 2017;20(7):407-412. [doi: 10.1089/cyber.2017.0028] [Medline: 28657783]
19. Çelik S, Atak H, Erguzen A. The effect of personality on cyberbullying among university students in Turkey. Eurasian J Educ Res 2012;49:129-150.
20. Liu M, Liu T, Wang X, Zhao N, Xue J, Zhu T. A linguistic study of Chinese Weibo users who lost their only child. Death Stud 2021;45(9):714-725. [doi: 10.1080/07481187.2019.1686088] [Medline: 31709924]
21. Camerini AL, Marciano L, Carrara A, Schulz PJ. Cyberbullying perpetration and victimization among children and adolescents: a systematic review of longitudinal studies. Telematics Informatics 2020;49:101362. [doi: 10.1016/j.tele.2020.101362]
22. Douketis JD, Spyropoulos AC. Perioperative management of patients taking direct oral anticoagulants: a review. JAMA 2024;332(10):825-834. [doi: 10.1001/jama.2024.12708] [Medline: 39133476]
23. Preoţiuc-Pietro D, Volkova S, Lampos V, Bachrach Y, Aletras N. Studying user income through language, behaviour and affect in social media. PLoS One 2015;10(9):e0138717 [FREE Full text] [doi: 10.1371/journal.pone.0138717] [Medline: 26394145]

XSL•FO
RenderX

24. Li S, Liu M, Zhao N, Xue J, Wang X, Jiao D, et al. The impact of family violence incidents on personality changes: an examination of social media users' messages in China. Psych J 2021;10(4):598-613. [doi: 10.1002/pchj.429] [Medline: 33530130]
25. Yiğit B, Çakmak BY. Discovering psychological well-being: a bibliometric review. J Happiness Stud 2024;25(5):43. [doi: 10.1007/s10902-024-00754-7]
26. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: three decades of progress. Psychological Bulletin 1999;125(2):276-302. [doi: 10.1037/0033-2909.125.2.276]
27. Ryff CD, Keyes CLM. The structure of psychological well-being revisited. Journal of Personality and Social Psychology 1995;69(4):719-727. [doi: 10.1037/0022-3514.69.4.719]
28. Hills P, Argyle M. The oxford happiness questionnaire: a compact scale for the measurement of psychological well-being. Personal Individual Diff 2002;33(7):1073-1082. [doi: 10.1016/s0191-8869(01)00213-6]
29. Peker A, Cengiz S, Eroğlu Y. Coping skills and perceived stress as pathways to well-being in adolescents experiencing cyber-victimization. Curr Psychol 2024;43(23):20709-20721. [doi: 10.1007/s12144-024-05864-2]
30. Winter EL, Maykel C, Bray M, Levine-Schmitt M, Graves M. Physical health as a foundation for well-being: exploring the RICH theory of happiness. In: Deb S, Gerrard BA, editors. Handbook of Health and Well-Being: Challenges, Strategies and Future Trends. Singapore: Springer; 2022.
31. Chamizo-Nieto MT, Rey L. Cybervictimization and suicidal ideation in adolescents: a prospective view through gratitude and life satisfaction. J Health Psychol 2023;28(7):620-632. [doi: 10.1177/13591053221140259] [Medline: 36597927]
32. Longobardi C, Settanni M, Fabris MA, Marengo D. Follow or be followed: exploring the links between Instagram popularity, social media addiction, cyber victimization, and subjective happiness in Italian adolescents. Children and Youth Services Review 2020;113:104955. [doi: 10.1016/j.childyouth.2020.104955]
33. Navarro R, Yubero S, Larrañaga E, Martínez V. Children's cyberbullying victimization: associations with social anxiety and social competence in a Spanish sample. Child Ind Res 2012;5(2):281-295. [doi: 10.1007/s12187-011-9132-4]
34. Xia T, Liao J, Deng Y, Li L. Cyberbullying victimization and social anxiety: mediating effects with moderation. Sustainability 2023;15(13):9978. [doi: 10.3390/su15139978]
35. Audrin C, Blaya C. Psychological well-being in a connected world: the impact of cybervictimization in children's and young people's life in France. Front Psychol 2020;11:1427 [FREE Full text] [doi: 10.3389/fpsyg.2020.01427] [Medline: 32765342]
36. Güçlü-Aydogan M, Ünal-Aydın P, Aydın O. The effects of psychological resilience, self-efficacy and metacognition on cyberbullying among adolescents. Child Adolesc Soc Work J 2023. [doi: 10.1007/s10560-023-00946-9]
37. Choi DH, Noh GY. The influence of social media use on attitude toward suicide through psychological well-being, social isolation, and social support. Information, Communication & Society 2020;23(10):1427-1443 [FREE Full text] [doi: 10.1080/1369118x.2019.1574860]
38. Zhang J. The strain theory of suicide. Journal of Pacific Rim Psychology 2019;13:e27. [doi: 10.1017/prp.2019.19]
39. Zhang XC, Chu XW, Fan CY, Andrasik F, Shi HF, Hu XE. Sensation seeking and cyberbullying among Chinese adolescents: examining the mediating roles of boredom experience and antisocial media exposure. Computers in Human Behavior 2022;130:107185. [doi: 10.1016/j.chb.2022.107185]
40. Fu WX, Zhao WL, Yin JH, Deng F. The cross-lagged study between cyberbully and suicidal ideation among college students. School health in China 2022;43(3):399-402. [doi: 10.16835/j.cnki.1000-9817.2022.03.019]
41. Kandler C, Bleidorn W, Riemann R, Angleitner A, Spinath FM. Life events as environmental states and genetic traits and the role of personality: a longitudinal twin study. Behav Genet 2012;42(1):57-72. [doi: 10.1007/s10519-011-9491-0] [Medline: 21822914]
42. Song M, Zhao N. Predicting life satisfaction based on the emotion words in self-statement texts. Front Psychiatry 2023;14:1121915 [FREE Full text] [doi: 10.3389/fpsyt.2023.1121915] [Medline: 36970294]
43. Wang X, Yang L, Yang J, Wang P, Lei L. Trait anger and cyberbullying among young adults: a moderated mediation model of moral disengagement and moral identity. Computers in Human Behavior 2017;73:519-526. [doi: 10.1016/j.chb.2017.03.073]
44. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 2013;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]
45. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing simplified Chinese psychological linguistic analysis dictionary for microblog. In: Imamura K, Usui S, Shirao T, Kasamatsu T, Schwabe L, Zhong N, editors. Brain and Health Informatics. Cham, Switzerland: Springer International Publishing; 2013.
46. Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends® Inf Retrieval 2008;2(1–2):1-135. [doi: 10.1561/1500000011]
47. Youyou W, Stillwell D, Schwartz HA, Kosinski M. Birds of a feather do flock together: behavior-based personality-assessment method reveals personality similarity among couples and friends. Psychol Sci 2017;28(3):276-284. [doi: 10.1177/0956797616678187] [Medline: 28059682]
48. Wang J, Liu T, Zhu T, Zhang L, Hao B, Chen Z. Classify Sina Weibo users into high or low happiness groups using linguistic and behavior features. arXiv:1507.01796v2 2015. [doi: 10.48550/arXiv.1507.01796]

XSL•FO
RenderX

49. HAO B, LI A, BAI S, ZHU T. Predicting psychological features based on web behavioral data: mental health status and subjective well-being. Chin Sci Bull 2015;60(11):994-1001. [doi: 10.1360/n972014-00763]
50. Han N, Li S, Huang F, Wen Y, Wang X, Liu X, et al. Sensing psychological well-being using social media language: prediction model development study. J Med Internet Res 2023;25:e41823 [FREE Full text] [doi: 10.2196/41823] [Medline: 36719723]
51. Han N, Li S, Huang F, Wen Y, Su Y, Li L, et al. How social media expression can reveal personality. Front Psychiatry 2023;14:1052844 [FREE Full text] [doi: 10.3389/fpsyt.2023.1052844] [Medline: 36937737]
52. Lv M, Li A, Liu T, Zhu T. Creating a Chinese suicide dictionary for identifying suicide risk on social media. PeerJ 2015;3:e1455 [FREE Full text] [doi: 10.7717/peerj.1455] [Medline: 26713232]
53. Graham J, Haidt J, Nosek BA. Liberals and conservatives rely on different sets of moral foundations. J Pers Soc Psychol 2009 May;96(5):1029-1046. [doi: 10.1037/a0015141] [Medline: 19379034]
54. Zhang Y, Yu F. Which socio-economic indicators influence collective morality? Big data analysis on online Chinese social media. Emerg Mark Finance Tr 2018 Feb 06;54(4):792-800. [doi: 10.1080/1540496X.2017.1321984]
55. Lyu S, Ren X, Du Y, Zhao N. Detecting depression of Chinese microblog users via text analysis: combining linguistic inquiry word count (LIWC) with culture and suicide related lexicons. Front Psychiatry 2023;14:1121583 [FREE Full text] [doi: 10.3389/fpsyt.2023.1121583] [Medline: 36846219]
56. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, et al. What do centrality measures measure in psychological networks? J Abnorm Psychol 2019;128(8):892-903. [doi: 10.1037/abn0000446] [Medline: 31318245]
57. Schönfelder A, Rath D, Forkmann T, Paashaus L, Lucht L, Teismann T, et al. Child abuse and suicidality in the context of the interpersonal psychological theory of suicide: a network analysis. BRITISH JOURNAL OF CLINICAL PSYCHOLOGY 2021;60(4):425-442. [doi: 10.1111/bjc.12293] [Medline: 33949706]
58. Sarabadani E, Morovati Z. Comparison of psychological problems and sleep quality in two groups of cyberbullying victims and normal people. Iranian Evolutionary Educational Psychology Journal 2022;4(1):103-114. [doi: 10.52547/ieepj.4.1.103]
59. Eroglu Y, Peker A, Cengiz S. Cyber victimization and well-being in adolescents: the sequential mediation role of forgiveness and coping with cyberbullying. Front Psychol 2022;13:819049 [FREE Full text] [doi: 10.3389/fpsyg.2022.819049] [Medline: 36467151]
60. Yuan G, Liu Z. Longitudinal cross-lagged analyses between cyberbullying perpetration, mindfulness and depression among Chinese high school students. J Health Psychol 2021;26(11):1872-1881. [doi: 10.1177/1359105319890395] [Medline: 31789585]
61. Rey L, Quintana-Orts C, Mérida-López S, Extremera N. The relationship between personal resources and depression in a sample of victims of cyberbullying: comparison of groups with and without symptoms of depression. International journal of environmental research and public health 2020;17(24):9307 [FREE Full text] [doi: 10.3390/ijerph17249307] [Medline: 33322740]
62. Rey L, Neto F, Extremera N. Cyberbullying victimization and somatic complaints: a prospective examination of cognitive emotion regulation strategies as mediators. International Journal of Clinical and Health Psychology 2020;20(2):135-139 [FREE Full text] [doi: 10.1016/j.ijchp.2020.03.003] [Medline: 32550853]
63. Ronis S, Slaunwhite A. Gender and geographic predictors of cyberbullying victimization, perpetration, and coping modalities among youth. Canadian Journal of School Psychology 2019;34(1):3-21. [doi: 10.1177/0829573517734029]
64. Nolet AM, Morselli C, Cousineau MM. The social network of victims of domestic violence: a network-based intervention model to improve relational autonomy. Violence Against Women 2021;27(10):1630-1654. [doi: 10.1177/1077801220947169] [Medline: 32814488]
65. Tao S, Lan M, Tan CY, Liang Q, Pan Q, Law NW. Adolescents' cyberbullying experience and subjective well-being: sex difference in the moderating role of cognitive-emotional regulation strategy. Comput Hum Behav 2024;153:108122. [doi: 10.1016/j.chb.2023.108122]
66. Ho Ma TTQ, Gu Ph DC. Cyberbullying victimization and depression: self-esteem as a mediator and approach coping strategies as moderators. Journal of American College Health 2023;71(1):94-101. [doi: 10.1080/07448481.2021.1877718] [Medline: 33650940]
67. Shukla A, Chouhan VS. Association between cyberbullying victimization and loneliness among adolescents: the role of coping strategies and emotional intelligence. International Journal of Cyber Behavior 2023;13(1):1-19. [doi: 10.4018/ijcbpl.330586]
68. Li J, Wu Y, Hesketh T. Internet use and cyberbullying: Impacts on psychosocial and psychosomatic wellbeing among Chinese adolescents. Computers in Human Behavior 2023;138:107461. [doi: 10.1016/j.chb.2022.107461]
69. Bleidorn W, Hopwood CJ, Back MD, Denissen JJA, Hennecke M, Hill PL, et al. Personality trait stability and change. Personality Science 2021;2(1):e6009. [doi: 10.5964/ps.6009]
70. Zsila A, Urbán R, Demetrovics Z. Gender, rumination, and awareness of the perpetrator's identity as predictors of help-seeking among cyberbullying victims. Int J Mental Health Addict 2019;17(4):947-958. [doi: 10.1007/s11469-018-9932-6]
71. Gao M, Zhao X, McJunkin M. Adolescents' Experiences of cyberbullying: gender, age and reasons for not reporting to adults. International Journal of Cyber Behavior, Psychology and Learning 2016;6(4):13-27. [doi: 10.4018/ijcbpl.2016100102]

XSL•FO
RenderX

72. Bastiaensens S, Van Cleemput K, Vandebosch H, Poels K, DeSmet A, De Bourdeaudhuij I. "Were You Cyberbullied? Let me help you." Studying adolescents' online peer support of cyberbullying victims using thematic analysis of online support group fora. In: Vandebosch H, Green L, editors. Narratives in Research and Interventions on Cyberbullying Among Young People. Cham, Switzerland: Springer International Publishing; 2019.
73. Alim S, Khalid S. Support for cyberbullying victims and actors: a content analysis of Facebook groups fighting against cyberbullying. International Journal of Technoethics 2019;10(2):22. [doi: 10.4018/ijt.2019070103]
74. Holman MS, Williams MN. Suicide risk and protective factors: a network approach. Arch Suicide Res 2022;26(1):137-154. [doi: 10.1080/13811118.2020.1774454] [Medline: 32522102]
75. Ranjbar NF, Rasoolzadeh B, Baradaran M, Sedaghat N, Makvandi L. The effect of acceptance and commitment therapy on psychological capital promotion in bullied students. J Prev Counsel 2023;4(4). [doi: 10.22098/jpc.2023.13216.1175]
76. Moshtaghi M, Asghari Ebrahimabad MJ, Aghamohamadian Sheirbaf HR. The effectiveness of acceptance and commitment based therapy on psychological capitals and neuroticism among mothers of children with down syndrome. Psychol Achievements 2020;27(2):233-256. [doi: 10.22055/psy.2020.31711.2450]
77. Urbaniak R, Ptaszyński M, Tempska P, Leliwa G, Brochocki M, Wroczyński M. Personal attacks decrease user activity in social networking platforms. Computers in Human Behavior 2022;126:106972. [doi: 10.1016/j.chb.2021.106972]

Abbreviations

MFD: Moral Foundations Dictionary
MMD: Moral Motivation Dictionary
OER: online ecological recognition
PWB: psychological well-being
SC-LIWC: Simplified Chinese version of Linguistic Inquiry and Word Count
SWB: subjective well-being

Edited by T de Azevedo Cardoso; submitted 17.07.24; peer-reviewed by A Li, M Takano, Y Zhou; comments to author 11.10.24; revised version received 27.11.24; accepted 15.01.25; published 25.02.25.

Please cite as:
Liu X, Liu M, Kang X, Han N, Liao Y, Ren Z
More Cyberbullying, Less Happiness, and More Injustice—Psychological Changes During the Pericyberbullying Period: Quantitative Study Based on Social Media Data
J Med Internet Res 2025;27:e64451
URL: https://www.jmir.org/2025/1/e64451
doi:10.2196/64451
PMID:39998871

©Xingyun Liu, Miao Liu, Xin Kang, Nuo Han, Yuehan Liao, Zhihong Ren. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Moderated Mediating Effects of Social Media Identity and Loneliness on the Relationship Between Problematic Internet Use and Mental Health in China: Nationwide Cross-Sectional Questionnaire Study

Chenxi Liu[1], PhD; Yushu Liu[1], MSc; Chaojie Liu[2], PhD; Rujiao Lin[1], MSc; Xi Wang[1], MSc; Xinyi Zhang[1], MSc; Yibo Wu[3*], PhD; Dan Wang[4,5,6*], PhD

[1]School of Medicine and Health Management, Tongji Medical School, Huazhong University of Science and Technology, Wuhan, China

[2]School of Psychology and Public Health, La Trobe University, Melbourne, Australia

[3]School of Public Health, Peking University, Beijing, China

[4]School of Management, Hubei University of Chinese Medicine, Wuhan, China

[5]Hubei Shizhen Laboratory, Hubei University of Chinese Medicine, Wuhan, China

[6]Research Center for Traditional Chinese Medicine Development, Hubei University of Chinese Medicine, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Dan Wang, PhD
School of Management
Hubei University of Chinese Medicine
16 Huangjiahu West Rd, Hongshan District
Wuhan, 430065
China
Phone: 86 13476237004
Email: 815310016@qq.com

Abstract

Background: Mental health disorders are a major public health challenge, and problematic internet use (PIU) may play an important role in this issue. However, the underlying mechanisms of PIU and its effects on mental health have not yet been elucidated.

Objective: This study examines the mediating effect of loneliness on the relationship between PIU and mental health, as well as the moderating effect of social media identity on the relationships among PIU, loneliness, and mental health.

Methods: A cross-sectional questionnaire survey was conducted on 21,292 participants recruited using a multistage stratified sampling strategy from 31 provinces/regions in mainland China from June 20 to August 31, 2022. This study assesses PIU (PIU questionnaire-short form-6), depression (9-item Patient Health Questionnaire [PHQ-9]), anxiety (7-item Generalized Anxiety Disorder scale [GAD-7]), loneliness (3-item loneliness scale), and social media identity (identity bubble reinforcement scale). Additionally, we collected the sociodemographic characteristics of the participants. Participants whose total score of PHQ-9≥15 or total score of GAD-7≥10 were considered to have moderate or severe symptoms of depression or anxiety, respectively. A moderated mediation model was established to examine the mediating effect of loneliness on the association between PIU and mental health outcomes (depression and anxiety), moderated by social media identity.

Results: Approximately 22.80% (4854/21,292) and 14.20% (3023/21,292) of the respondents reported moderate or severe symptoms of depression and anxiety, respectively. Loneliness significantly mediated the association between PIU and mental health outcomes, explaining 42.53% and 45.48% of the total effect of PIU on depression and anxiety, respectively. Social media identity significantly moderated the associations between PIU and depression (β=0.002, 95% CI 0.001-0.002), PIU and anxiety (β=0.001, 95% CI 0.000-0.002), loneliness and depression (β=0.010, 95% CI 0.007-0.013), and loneliness and anxiety (β=0.007, 95% CI 0.004-0.010), but not between PIU and loneliness (β=0.000, 95% CI –0.003 to 0.003). Higher levels of social media identity were significantly associated with lower levels of loneliness (β=–0.018, 95% CI –0.020 to –0.016).

XSL•FO
RenderX

Conclusions: Addressing loneliness may serve as a valuable approach to mitigate the impact of PIU on mental health outcomes. However, social media identity poses a significant challenge in addressing health issues linked to PIU.

(J Med Internet Res 2025;27:e57907) doi:10.2196/57907

KEYWORDS

problematic internet use; loneliness; anxiety; depression; mediation and moderation model

Introduction

Background

Mental health disorders, which are significant contributors to the global burden of disease, are a major public health challenge worldwide. Over the past 3 decades, its burden, measured in disability-adjusted life-years, has increased from 80.8 million in 1990 to 125.3 million in 2019, with no indication of a decrease in the near future [1]. The COVID-19 pandemic has further exacerbated mental health problems, resulting in an additional 53.2 million cases of depression and 76.2 million cases of anxiety—the 2 most prevalent mental health disorders [2]. Responding to the prolonged mental health impacts of COVID-19 has proven challenging because health care systems are ill-prepared [3].

Mental health disorders are influenced by a multitude of factors [4], with complex interactions among biological, psychological, social, cultural, and other elements [5]. The rapid advancement of technology, particularly the widespread use of the internet, has garnered significant attention for its impact on mental health and vice versa [6]. Existing studies [7] have shown that there is a potential bidirectional relationship between internet use and mental health disorders, in which internet overuse may induce mental health disorders, for example, depression and anxiety, which could in turn enhance one's addiction to the internet. Although the relationships between internet use and mental health disorders remain unclear and different studies have argued different directional relationships between them, it is undeniable that the internet has profoundly transformed human society, altering how people communicate and interact, reducing communication barriers, lowering costs, and promoting interpersonal bonds and social networks [8].

In this study, we focused on the impact of internet use on mental health disorders, and existing studies have shown that maladaptive or excessive internet use has been linked to adverse effects on mental health [9-11]. For example, empirical evidence has found that problematic internet use (PIU) and internet addiction [12-16] increase the likelihood of depression, anxiety, social isolation, and even suicide [7,17-22]. However, the reasons for the associations between PIU and mental health remain unclear. Recent systematic reviews have highlighted several limitations in existing studies, such as small sample sizes, variations in study settings, and lack of moderating or mediating analyses [23,24]. One assumption regarding the link between PIU and mental health disorders is that PIU is inherently connected to loneliness [25]. Although loneliness is not a specific mental health condition, it is associated with major mental health disorders, including depression and anxiety. The effect of internet use on loneliness remains a subject of debate. Some researchers contend that internet use can exacerbate loneliness by compromising offline human interactions [26,27], whereas others argue that it can enhance existing offline relationships and foster new ones [28]. Nonetheless, the association between PIU and loneliness is likely to be bidirectional: PIU can be both a cause (resulting in higher levels of loneliness) and a consequence of loneliness [25].

What has been missing in these discussions is a consideration of individual behaviors: why and how individuals use the internet. Social media has created a virtual space for people to interact, but the reasons people use social media vary significantly. Online platforms can serve as tools for sharing information and exchanging social support in the real world [29,30]. This underscores the importance of the concept of social media identity [31,32]. According to the identity bubble reinforcement model [33], online interactions can lead to the formation of social networks (social identities) where like-minded individuals (homophily) interact with one another, often relying on biased information. Social media identity measures how people perceive their relationships within online networks, with a higher level of social media identity indicating greater involvement with online social networks [34]. It is reasonable to assume that high levels of social media identity may reinforce a sense of belonging to virtual communities at the expense of real-world connections [31]. This, in turn, could affect the association between PIU and loneliness, as well as the impact of PIU on mental health [35,36]. Although previous research has explored the effects of social media identity on online gaming and cyberbullying [35,37], there is a noticeable gap in the literature regarding the role of social media identity in the context of PIU, loneliness, and mental health. Notably, this study was conducted in China, as Asian countries have reported a higher prevalence of PIU, ranging from 20% to 47%, than other regions [38,39].

To better understand the underlying reasons for the associations between PIU and mental health, this study aims to examine the mediating effect of loneliness on the relationship between PIU and mental health, as well as the moderating effect of social media identity on the relationships among PIU, loneliness, and mental health.

Theoretical Hypotheses

In this study, we tested a moderated mediation model based on 4 hypotheses, examining the interplay among PIU, loneliness, social media identity, and their effects on mental health (Figure 1).

XSL•FO
RenderX

Figure 1. The hypothesized moderated mediation model.

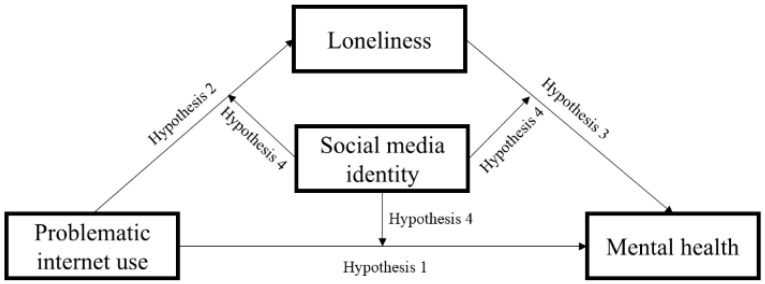

Loneliness Mediates the Effects of PIU on Mental Health Disorders

Hypothesis 1: PIU Is Positively Associated With Mental Health Disorders

The link between PIU and mental health has been extensively documented [17-20,40]. A recent systematic review revealed a moderate level of evidence supporting a positive association between PIU and depressive symptoms [18]. In general, PIU is associated with withdrawal from real-world social interactions [41]. Excessive internet use can also result in sleep deprivation, unrealistic expectations, and an exaggerated sense of failure [41,42].

Hypothesis 2: PIU Is Positively Associated With Loneliness

PIU can lead to displacement, a phenomenon in which internet-based, low-quality relationships replace offline social contacts, even though the internet can strengthen an individual's connectedness with others and expand their social networks [25]. Recent reviews have indicated moderately low levels of evidence supporting PIU as a risk factor for loneliness [25]. Several longitudinal studies have shown that PIU is more likely to be a cause rather than a consequence of loneliness [43-45]. Excessive internet use can lead individuals to withdraw from real-world interactions, intensifying their perception of the absence of companionship [25,46].

Hypothesis 3: Loneliness Is Positively Associated With Mental Health Disorders

Evidence has consistently shown that loneliness is a risk factor for mental health disorders, particularly depression and anxiety. Loneliness has negative effects on the social, cognitive, and biological aspects [47,48]. Individuals experiencing loneliness often have less cognitive stimulation, reduced social interactions, and impaired immune functions, all of which are associated with an increased likelihood of mental health disorders [47,48]. This connection between loneliness and mental health disorders has been demonstrated in 2 large longitudinal studies conducted in the United Kingdom and the Netherlands [47,49,50]. Thus, from the evidence in the literature, we inferred that the effect of PIU on mental health disorders is mediated by loneliness, that is, PIU affects loneliness and ultimately translates into mental health disorders.

Hypothesis 4: Social Media Identity Moderates the Effects of PIU on Loneliness and Mental Health Disorders

Studies have shown that the extent to which PIU can translate into loneliness and mental health disorders depends on how individuals use the internet [18]. In the identity bubble reinforcement model, social media identity measures how people perceive their relationships with online networks, with a higher level of social media identity indicating greater involvement in and belonging to online social networks [33]. On the one hand, people highly involved in online social networks are more likely to express themselves, receive social support, feel a sense of belonging, be less lonely, and protect themselves from translating PIU and loneliness into mental health problems [30,51,52]. This effect has also been confirmed in risky online behaviors such as gambling [35]. In contrast, in the case of people with PIU but low involvement in and belonging to online social networks, it is expected that their loneliness will be exacerbated and that PIU and loneliness will be translated into mental health disorders. Thus, it is speculated that various levels of social media identity have different impacts on the relationships between individuals' internet use and loneliness and mental health disorders. Therefore, social media identity was hypothesized to moderate the effects of PIU, loneliness, and mental health in this study, with higher (lower) social media identity indicating a lower (higher) effect of PIU on loneliness and less (greater) likelihood of PIU and loneliness translating into mental health disorders.

Methods

Study Design and Participants

The data for this study were extracted from the 2022 nationwide cross-sectional survey, known as the Psychology and Behavior Investigation of Chinese Residents (PBICR-2022). The sample size was determined based on the following formula: $n = Z^2_{\alpha} \times p \times (1 - p) / \delta^2$. Based on the prevalence of depression (p) in China as 24% [53], type I error (α) as .05, and acceptable error (δ) as 0.02, the minimum sample size was estimated as 1752. To generate a nationally representative sample, the survey adopted multistage sampling, covering 31 provinces in China (22 provinces, 5 autonomous regions, 4 municipalities). In each province, the planned minimum sample size was determined as 500-2500 according to the proportions of the respective populations to the overall population in China. Finally, the sample size was estimated at 20,000, which is far more than the minimum sample size (n=1752). To select study participants, we first included capital cities from 31 provinces across China. Additionally, 2-12 cities were selected from each noncapital prefecture within each province, resulting in 148 cities. For the 148 municipalities sampled, we adopted random sampling, and the number of communities sampled was determined on the

XSL•FO
RenderX

basis of the populations within the sampled municipalities and the ratio of urban to rural communities, resulting in a total of 780 communities. Within each community, the allocation of quotas considered the demographic distributions recorded in the seventh national census, considering both age and sex. Details about the protocol of the PBICR-2022 survey have been published elsewhere [54].

The inclusion criteria for participation in the survey included individuals (1) aged 12 years or older, (2) who had resided in the sampled communities and had not been absent for more than one month within the past year, and (3) who were able to complete the questionnaire independently or with the help of a researcher. We excluded participants who (1) were cognitively impaired in understanding the questionnaire items, (2) had never used the internet or social media, or (3) were unwilling to participate in or were participating in other similar studies during the survey period.

Data collection occurred between June 20 and August 31, 2022, when China was still implementing a zero COVID-19 policy. Household visits were performed by trained researchers whenever local policies permitted such visits. Otherwise, online meetings, typically using platforms such as WeChat or Tencent, were conducted with each participant. The respondents were asked to respond through the Wenjuanxing questionnaire platform. Written informed consent was obtained prior to each survey.

A total of 21,916 responses were returned, and invalid responses (incomplete and contradictory answers or survey times of less than 4 minutes) were excluded. This resulted in a total of 97.15% (21,292/21,916) valid responses. The average age of the included participants was 39.15 (SD 18.64; range 12-100) years.

Measurements

PIU Assessment

PIU was assessed using the 6-item PIU-short form questionnaire. This questionnaire evaluates individual obsessions, neglect, and control behaviors related to internet usage. For example, respondents were asked about their feelings of tension, irritation, or stress when unable to use the internet as long as they wished, their preference for spending time online over sleep, and their attempts to conceal the amount of time spent online [55,56]. The participants were required to rate each item on a 5-point Likert scale, ranging from 0 (never) to 4 (always). A summed score within the range of 0-24 was calculated, with a higher score indicating a greater risk of PIU. The 6-item PIU-short form questionnaire has demonstrated good reliability and validity in several countries [55,56], and its structural validity (via confirmatory factor analysis) and internal consistency (Cronbach α=0.932) have been confirmed.

Loneliness Assessment

Loneliness was assessed using the 3-Item Loneliness Scale, which evaluates relational connectedness, social connectedness, and self-perceived isolation [57]. Each item was rated on a 3-point scale, with 0 indicating “hardly ever,” 1 indicating “some of the time,” and 2 indicating “often.” A summed score within the range of 0-6 was calculated, where a higher score indicated greater loneliness. The 3-Item Loneliness Scale has been reported to have good reliability and validity in previous studies [58,59] and in this study, with a Cronbach α coefficient of 0.862.

Social Media Identity Assessment

Social media identity was assessed using the Identity Bubble Reinforcement Scale [31]. The Identity Bubble Reinforcement Scale evaluates the extent to which individuals relate to online social networks and consists of 6 items. These items measure social identification (eg, in social media, I belong to a community or communities that I'm proud of), homophily (eg, in social media, I prefer interacting with people who are like me), and information bias (eg, in social media, I feel that people think like me) [31]. The respondents were required to rate each item on a scale ranging from 0 (does not describe me at all) to 9 (describes me completely). A summed score within the range of 0-54 was calculated, with a higher score indicating a higher level of social media identity involvement. The Identity Bubble Reinforcement Scale has shown good reliability and validity in various countries [31], and in this study, it yielded a Cronbach α of 0.924.

Mental Health Disorder Assessment

The PBICR-2022 survey included assessments of depression and anxiety, which are among the most common mental health disorders. The 9-item Patient Health Questionnaire (PHQ-9) was utilized to measure depression, while the 7-item Generalized Anxiety Disorder (GAD-7) scale was adopted for assessing anxiety. The respondents were asked to rate their experiences over the past 2 weeks by using a 4-point scale, with responses ranging from 0 (not at all) to 3 (nearly every day). Sample questions included, “How often do I have trouble concentrating on things?” for depression [60] and “How often do I feel afraid that something awful might happen?” for anxiety [61]. A summed score for depression, ranging from 0 to 27, and for anxiety, ranging from 0 to 21, was calculated. A higher score indicates a higher level of symptoms. In this study, participants whose total score of PHQ-9≥15 or total score of GAD-7≥10 were considered as having moderate or severe symptoms of depression or anxiety, respectively. PHQ-9 and the GAD-7 have been widely used in various settings and have been validated in Chinese populations [62,63]. The Cronbach α coefficients of the 2 measurements were 0.919 (PHQ-9) and 0.941 (GAD-7) in our study.

Covariates

Previous studies have identified numerous socioeconomic and demographic factors associated with mental health disorders [64-66]. For these factors, data regarding age group (<18, ≥18 to <30, ≥30 to <40, ≥40 to <50, ≥60 to <70 and ≥70 years), sex (male, female), educational attainment (no formal education, primary school, etc), employment (full-time students, employed, freelancers, etc), marital status (unmarried, married, divorced/widowed), household income (≤¥1000, >¥1000 to ≤¥2000, etc; US $1=¥7.31), chronic conditions (yes, no), living area (rural, urban), and region were collected via self-reports. The regions in China were categorized as northern (Beijing, Tianjin, etc), northeast (Liaoning, Jilin, Heilongjiang), northwest (Gansu, Qinghai, etc), eastern (Shanghai, Jiangsu, etc), central

XSL•FO
RenderX

(Henan, Hubei, Hunan), southern (Guangdong, Guangxi, Hainan), and southwest (Chongqing, Sichuan, etc) according to the classification of regional economic collaboration among provinces in China [67].

Statistical Analysis

The characteristics of the study participants were described through frequency distributions for categorical variables and means (medians) and standard deviations (interquartile ranges) for continuous variables. We also examined the normality of the total scores for PIU, loneliness, social media identity, and mental health disorders (depression and anxiety) via Shapiro-Francia tests for normality. The results indicated that all the variables were not normally distributed (P<.001). Thus, Spearman correlations were examined among those measurements, and Wilcoxon rank sum tests and Kruskal-Wallis tests were performed to assess differences in PIU, loneliness, social media identity, and mental health disorders among participants with different socioeconomic characteristics.

We employed multivariate linear regression models to test our hypotheses. For the moderated mediation analyses, we utilized the PROCESS macro developed by Hayes [68]. The analysis proceeded in 2 steps. First, we tested the mediating effect of loneliness on the associations between PIU and depression and anxiety disorders. Next, we added the moderating effect of social media identity on different paths within the mediation model. To estimate the indirect effects, we used 5000 bootstrap estimates, which provided a 95% bias-corrected CI with robustness to nonnormally distributed data. Additionally, we calculated the conditional effects of social media identity by examining the associations of all paths between PIU and mental health when the social media identity score was fixed at 1 SD above and below the mean [68].

To assess model robustness and identify potential model heterogeneity, we conducted subgroup analyses on the basis of sex and age, in which mediation and moderated mediation models were reconstructed and tested in a particular subgroup of the population (eg, male participants). As the cross-sectional data did not allow us to establish the temporal precedence of variables, we also examined a model in which PIU served as a mediator for the association between loneliness and mental health disorders. We performed Harman's one-factor analysis to address the potential threat of common method variance in cross-sectional behavioral research. The results indicated that the first common factor in our data explained only 27.99% of the variance, which is below the threshold of 40%. This suggests a low risk of common method variance [69].

The data were analyzed using SPSS 26.0 (IBM Corp). A 2-tailed P value <.05 was considered statistically significant. The indirect effects of the mediation and moderation variables were considered statistically significant when zero was not included between the lower and upper bounds of the 95% bias-corrected bootstrap CIs.

Ethics Approval

This study was approved by the Shaanxi Institute of International Trade and Commerce (JKWH-2022-02). Written informed consent was obtained from participants prior to the survey (for participants aged 12-17 years, approval from their caregivers were also prerequisite for enrollment), and they were provided with the opportunity to opt out at any time during and after the survey. No personal identification information was collected, and all the data were anonymized. There was no compensation for the participants in our study survey.

Results

Characteristics of the Study Participants

Approximately half (10,596/21,292, 49.77%) of the respondents were men, whereas 18.1% (3857/21,292) were older than 60 years. Fewer than one-third (6462/21,292, 30.35%) of the respondents resided in rural communities. Most respondents had completed senior high school or above (14,651/21,292, 69%), with an average household income of less than ¥5000 (equivalent to approximately US $684), accounting for 63.02% (13,418/21,292) (Table S1 of Multimedia Appendix 1).

PIU, Loneliness, Social Media Identity, and Mental Health Outcomes

The respondents had a mean score of 6.49 (SD 5.51) for depression symptoms and 4.75 (SD 4.62) for anxiety symptoms. Compared with male respondents, female respondents (P=.003) reported more anxiety problems, while there was no sex difference in depression symptoms (P=.07). Respondents with the highest scores for both depression and anxiety (Table S2 of Multimedia Appendix 1) were in the 18-29 years age groups (P<.001), unmarried (P<.001), had completed postgraduate education (P<.001), were engaged in full-time studies (P<.001), had a household income of less than ¥1000 per month (equivalent to approximately US $136.80; P<.001), resided in rural areas (P<.001) and the northeastern region (P<.001), and had chronic conditions (P<.001).

The loneliness scores were relatively low, with a mean value of 1.57 (SD 1.61) out of a maximum score of 6 (Table 1). In general, respondents with the highest anxiety and depression scores tended to also have higher loneliness scores (Table S2 of Multimedia Appendix 1). The average PIU score was 5.80 (SD 5.46) out of a maximum of 24, indicating a moderate level of PIU. The respondents reported a moderate level of involvement in social media identity, with a mean score of 31.12 (SD 10.99) out of a maximum of 54 (Table 1). The highest levels of PIU and social media identity were found among respondents aged between 18 and 29 years (P<.001), those who were unmarried (P<.001), had completed postgraduate education (P<.001), were engaged in full-time studies (P<.001), had an average household income of >¥6000 (equivalent to approximately US $820.80) per month ($P$<.001), lived in urban areas (P=.009 for PIU and P<.001 for social media identity), and had no chronic conditions (P<.001). The respondents residing in the northeast region presented the highest PIU (P<.001), whereas those in the northern region presented the highest social media identity (P<.001). There was no sex difference in PIU (P=.10), but male respondents reported greater social media identity (P=.04) than their female counterparts (Table S2 of Multimedia Appendix 1).

XSL•FO
RenderX

PIU, loneliness, depression, and anxiety symptoms were positively correlated (P<.001), whereas social media identity was negatively correlated with loneliness, depression, and anxiety symptoms (P<.001) (Table 2).

Table 1. Measurement scores of the mental health conditions, loneliness, problematic internet use, and social media identity of the study participants (N=21,292).

Variables	Mean (SD) score	Median (IQR) score
Depression	6.49 (5.51)	6 (2-9)
Anxiety	4.75 (4.62)	4 (0-7)
Loneliness	1.57 (1.61)	1 (0-3)
Problematic internet use	5.80 (5.46)	11 (0-10)
Social media identity	31.12 (10.99)	37 (24-38)

Table 2. Spearman correlations of the measured construct (N=21,292).

Variables	Depression	Anxiety	Loneliness	Problematic internet use	Social media identity
Depression					
r	1	0.811	0.628	0.449	–0.125
P value	—[a]	<.001	<.001	<.001	<.001
Anxiety					
r	0.811	1	0.622	0.431	–0.105
P value	<.001	—	<.001	<.001	<.001
Loneliness					
r	0.628	0.622	1	0.395	–0.156
P value	<.001	<.001	—	<.001	<.001
Problematic internet use					
r	0.449	0.431	0.395	1	0.007
P value	<.001	<.001	<.001	—	.29
Social media identity					
r	–0.125	–0.105	–0.156	0.007	1
P value	<.001	<.001	<.001	.29	—

[a]Not applicable.

Hypothesis Testing

PIU was found to be associated with loneliness (model 1 in Table 3), depression (model 2 in Table 3), and anxiety (model 3 in Table 3). This supports hypotheses 1 and 2. Loneliness played a significant mediating role in these associations, as shown in models 2 and 3 in Table 2. The change in R^2 values confirmed the substantial contribution of loneliness as a mediator in the relationships between PIU and depression (R^2 increased from 0.246 to 0.454) and anxiety (R^2 increased from 0.223 to 0.446).

The total effect of PIU on depression was 0.475 (95% CI 0.462-0.487), of which 42.53% (0.202, 95% CI 0.193-0.211) was explained by the indirect effect mediated through loneliness. Similarly, the total effect of PIU on anxiety was 0.387 (95% CI 0.377-0.398), with 45.48% (0.176, 95% CI 0.168-0.184) explained by the indirect effect mediated through loneliness (see Figure 2). This finding supports hypothesis 3.

XSL•FO
RenderX

Table 3. Results of the mediating effect of loneliness in the relationship between problematic internet use and mental health.

Predictors	Model 1[a] (Y=loneliness)	Model 2[b] (Y=PHQ[c])	Model 3[d] (Y=GAD[e])
PIU[f], β (95% CI)	0.116 (0.112-0.120)	0.272 (0.261-0.284)	0.212 (0.202-0.222)
Loneliness, β (95% CI)	N/A[g]	1.743 (1.705-1.781)	1.514 (1.482-1.546)

[a]$F_{33, 21,258}$=160.675; R^2=0.200.

[b]$F_{34, 21,257}$=519.443; R^2=0.454.

[c]PHQ: Patient Health Questionnaire.

[d]$F_{34, 21,257}$=502.896; R^2=0.446.

[e]GAD: Generalized Anxiety Disorder scale.

[f]PIU: problematic internet use.

[g]N/A: not applicable.

Figure 2. Mediating model of problematic internet use on depression and anxiety via loneliness. "a" represents the coefficient between problematic internet use and loneliness, "b" represents the coefficient between loneliness and depression/anxiety, and "c" represents the total coefficient when combined direct coefficient (c') from problematic internet use and depression/anxiety and indirect coefficient (c") between them via loneliness.

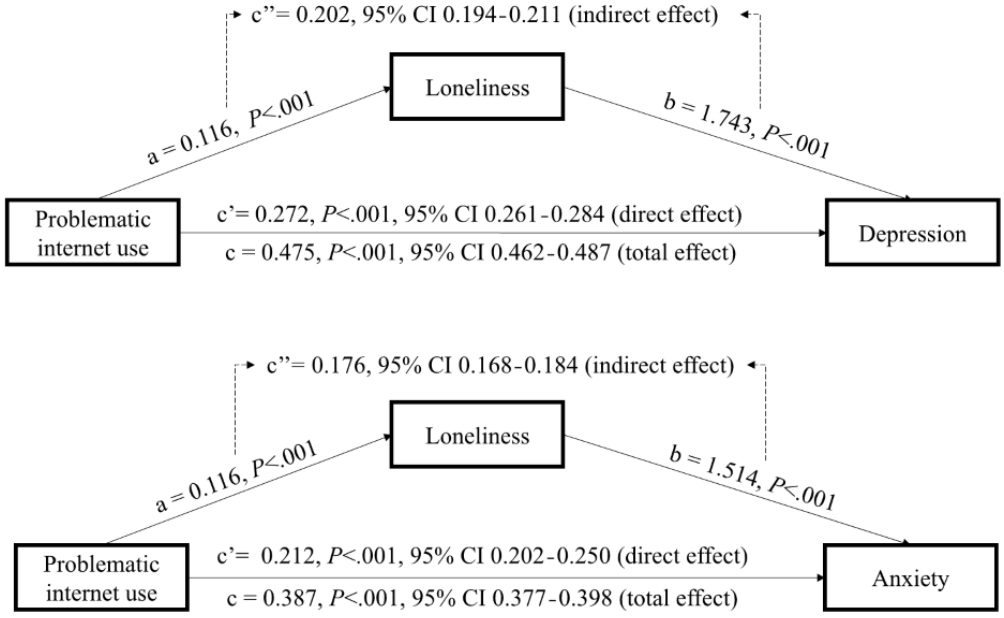

Social media identity moderated the effect of PIU on depression (β=0.002, 95% CI 0.001-0.002; model 5 in Table 4) and the effect of loneliness on depression (β=0.010, 95% CI 0.007-0.013; model 5 in Table 4). Similarly, social media identity moderated the effect of PIU on anxiety (β=0.001, 95% CI 0.000-0.002; model 6 in Table 4) and the effect of loneliness on anxiety (β=0.007, 95% CI 0.004-0.010; model 6 in Table 4). Higher levels of social media identity exacerbated the effects of PIU on mental health problems, supporting hypothesis 4. However, the moderating effect of social media identity on the association between PIU and loneliness was statistically insignificant (β=0.000, 95% CI –0.003 to 0.003; model 4 in Table 4), despite the direct association between higher levels of social media identity and lower levels of loneliness (β=–0.018, 95% CI –0.020 to –0.016; model 4 in Table 4). Multimedia Appendix 2 and Figure 3 present the conditional effects of social media identity on the paths between PIU and mental health symptoms.

XSL•FO
RenderX

Table 4. Results of the moderated mediation model of social media identity and loneliness in the relationship between problematic internet use and mental health.

Predictors	Model 4[a] (Y=Loneliness)	Model 5[b] (Y=PHQ[c])	Model 6[d] (Y=GAD[e])
PIU[f], β (95% CI)	0.117 (0113 to 0.121)	0.267 (0.255 to 0.279)	0.207 (0.198 to 0.217)
Loneliness, β (95% CI)	N/A[g]	1.739 (1.700 to 1.778)	1.516 (1.484 to 1.549)
Social media identity, β (95% CI)	–0.018 (–0.020 to –0.016)	–0.005 (–0.010 to 0.001)	0.002 (–0.002 to 0.007)
Social media identity × PIU, β (95% CI)	0.000 (–0.003 to 0.003)	0.002 (0.001 to 0.002)	0.001 (0.000 to 0.002)
Social media identity × loneliness, β (95% CI)	N/A	0.010 (0.007 to 0.013)	0.007 (0.004 to 0.010)

[a]$F_{35, 21,256}$=165.462; R^2=0.214.

[b]$F_{37, 21,254}$=481.285; R^2=0.456.

[c]PHQ: Patient Health Questionnaire.

[d]$F_{37, 21,254}$=464.138; R^2=0.447.

[e]GAD: Generalized Anxiety Disorder scale.

[f]PIU: problematic internet use.

[g]N/A: not applicable.

Figure 3. For each pathway, social media identity (H) and social media identity (L) represent the associations among problematic internet use, loneliness, and depression/anxiety when the level of social media identity is set at the mean score plus one standard deviation or minus one standard deviation, respectively. *P* value indicates the significance of the pathway, and bold paths and values indicate significant differences between different levels of social media identity.

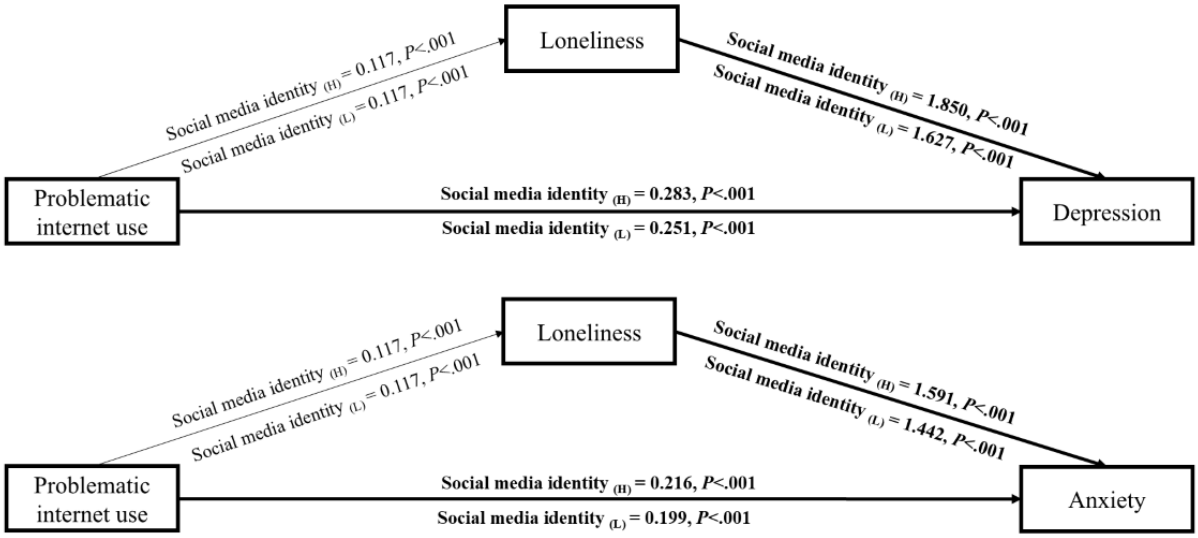

The effect of PIU on depression, including the indirect effect via loneliness, was more significant among respondents with a higher social media identity than among those with lower social media identity. This difference was manifested as a 0.032 variation in the direct effect and a 0.025 difference in the indirect effect, with a bootstrapped 95% CI of 0.005-0.044. Similarly, the effect of PIU on anxiety, including the indirect effect via loneliness, was significantly more pronounced among respondents with higher social media identity than among those with lower social media identity. This difference amounted to a 0.017 variation in the direct effect and a 0.017 difference in the indirect effect, with a bootstrapped 95% CI of 0.000-0.033 (Table 5).

The subgroup analyses based on sex and age confirmed the moderated mediation models tested in this study (Table S4 of Multimedia Appendix 1). In all the subgroups, except for those older than 60 years, the indirect effects of PIU on depression and anxiety via loneliness exceeded 40% of the total effect. Additionally, higher levels of social media identity were associated with greater impacts of PIU on depression and anxiety in nearly all subgroups, except the group aged 50-59 years, where this effect was not observed for the impact of PIU on depression.

Alternative model testing revealed that PIU mediated the effects of loneliness on depression and anxiety (Tables S5 and S6 of Multimedia Appendix 1). However, the indirect effect explained only 16.12% and 14.66% of the total effect of loneliness on depression and anxiety, respectively. The change in R^2 values also indicated a limited contribution of PIU as a mediator.

XSL•FO
RenderX

Table 5. Conditional indirect and direct effects of social media identity.

Social media identity	Indirect effect	Boot SE	Boot LLCI-ULCI[a]	Direct effect	LLCI-ULCI
Depression					
Average	0.203	0.005	0.178-0.212	0.267	0.255-0.279
Low (mean – 1 SD)	0.190	0.006	0.178-0.203	0.251	0.234-0.268
High (mean + 1 SD)	0.216	0.007	0.201-0.229	0.283	0.268-0.298
Difference	0.025	0.010	0.005-0.044	0.032	N/A[b]
Anxiety					
Average	0.177	0.004	0.169-0.185	0.207	0.198-0.217
Low (mean – 1 SD)	0.169	0.006	0.158-0.180	0.199	0.185-0.213
High (mean + 1 SD)	0.185	0.006	0.174-0.197	0.216	0.204-0.229
Difference	0.017	0.008	0.000-0.033	0.017	N/A

[a]LLCI-ULCI: lower limit of confidence interval to upper limit of confidence interval.

[b]N/A: not applicable.

Discussion

Main Findings

This study reveals that increased PIU is linked to increased depression and anxiety symptoms. Loneliness acts as a mediator in this association, accounting for almost half of the total effects. Notably, while a higher social media identity may decrease loneliness, it intensifies the impacts of loneliness and PIU on depression and anxiety.

Comparisons With the Existing Evidence

The study confirms a significant effect of PIU on loneliness, although the direction of the relationship remains inconclusive. This aligns with the findings of most existing studies [25,70,71], indicating an internet displacement effect, in which the internet replaces offline interactions rather than a stimulation effect, in which the internet enhances existing relationships [25]. Previous studies have extensively documented the positive associations of PIU with depression and anxiety [17-20,40,72], as well as the positive associations of loneliness with depression and anxiety [73-76]. These associations were confirmed in this study.

In our study, we discovered that loneliness serves as a significant mediator in the link between PIU and mental health disorders, explaining nearly half of the total effects on both depression and anxiety symptoms. Notably, loneliness can be alleviated by enhancing social skills and strengthening social support [77], providing a novel approach to mitigate the negative consequences of PIU on mental health [78]. However, the substantial direct effects of PIU on depression and anxiety symptoms remain unclear and warrant further exploration. Existing evidence has suggested that sleep disturbance may be a potential pathway through which PIU affects mental health [79,80]. Consequently, future research is needed to deepen our understanding of the underlying reasons for the association between PIU and mental health.

Social media has produced revolutionary changes in people's daily lives, profoundly impacting their health. Previous studies have established connections between PIU and social media identity [31] as well as between social media identity and loneliness [30,51,52]. This study expands our existing knowledge by exploring the role of social media identity in the effects of PIU on both loneliness and mental health disorders. We found that higher levels of social media identity have the potential to exacerbate the effects of PIU on mental health disorders. Social media identity is formed as individuals seek connections with like-minded others online [31,81], which can reduce feelings of loneliness, as indicated in this study and others [82-84]. However, this reduction in loneliness does not appear to be strong enough to counterbalance the exacerbating effects of social media identity on the association between PIU and mental health disorders. One plausible explanation is that individuals with high levels of social media identity become deeply engrossed in the virtual online space at the expense of offline social interactions [85,86]. This may result in reduced ability to obtain social support from traditional relationships such as family members, friends, and work colleagues. Online relationships are often criticized for being superficial and lacking authenticity, and they cannot fully replace the role of offline relationships [87]. People with a high level of social media identity may feel less lonely but have limited access to offline interactions and social support, thereby increasing their risk of depression and anxiety [31,37].

Although the primary focus of this study was not on the prevalence of depression and anxiety, our study participants appeared to exhibit similar levels of depression and anxiety symptoms as those reported in another national study conducted in 2021 [88]. In our study, approximately 22.80% (4854/21,292) and 14.20% (3023/21,292) of the respondents were identified as experiencing moderate or severe depression and anxiety, respectively [89,90]. Consistent with previous studies conducted during the COVID-19 pandemic [2], our research also identifies the female sex and young age groups such as those aged 18-29 years as significant risk factors for mental health disorders. Recent studies have revealed a rapid increase in PIU, which is the most prevalent among men and younger age groups [91,92]. However, our study suggests that the effects of PIU on mental

XSL•FO
RenderX

health disorders are similar between male and female respondents.

Policy Implications

As loneliness is a significant mediator between PIU and depression and anxiety, it is beneficial to consider a comprehensive strategy that addresses both PIU and loneliness to mitigate the negative mental health outcomes associated with PIU. For educators and caregivers, existing studies have shown that active educational interventions, including alternative activity recordings and internet diary writing, were more likely to decrease internet addiction [93]. Additionally, policymakers can also introduce multilevel approaches that target internet users, online content, and the internet industry to address this challenge [94]. For example, China has issued a series of policies in which gaming companies have been required to impose restrictions, such as limiting young users to 1 hour of play on Fridays and weekends. Furthermore, the awareness and knowledge of the public should also be improved, especially with respect to risky online behaviors [93]. Research indicates that intervention programs focused on specific online risks such as gambling, cyberbullying, and online gaming tend to be more effective than general PIU interventions [95]. Offline strategies that enhance social connectedness and social support have also demonstrated their effectiveness in reducing loneliness [96] and can be incorporated into PIU management efforts. Social media identity plays a pivotal role in amplifying the mental health consequences of PIU. This underscores the importance of helping individuals balance their online and offline relationships. Although online connections can temporarily relieve loneliness, they should not substitute or jeopardize the value of real-world connections. Potential intervention strategies may involve public education campaigns to increase awareness and encourage responsible online behavior, as well as strengthening offline social activities to promote face-to-face interactions and social support.

Strengths and Limitations

To our knowledge, this is the first study to provide valuable insights into the intricate connections among PIU, loneliness, social media identity, and mental health disorders by using moderated mediation modeling. The sample size was large. This study has several limitations. One significant limitation is the cross-sectional design, which prevents us from verifying the bidirectional relationships between PIU and loneliness as well as between PIU and mental health disorders. Future research could benefit from a longitudinal design to explore the dynamic interplay between PIU, loneliness, and mental health over time. Additionally, although we used a moderated mediation effect model to examine the effects of PIU, loneliness, social media identity and their interaction on mental health, the nature of the cross-sectional design allows us to generate only an association rather than a causal relationship between them. Furthermore, the moderating effects were relatively small in this study, which may be susceptible to changes in other populations or concerning other mental health disorders. Thus, this study should be interpreted with caution. Furthermore, other unexamined factors and pathways may contribute to the relationship between PIU and mental health. For example, variables related to sleep patterns and sedentary lifestyles could play a role in both PIU and mental health disorders [97]. These factors were not included in our study because of data unavailability. Finally, although the mental health status of the participants was assessed on the basis of their experiences during the 2 weeks before the survey, the results were still vulnerable to potential recall bias. The excluded participants were more likely to be men older than 50 years, resulting in a slight underrepresentation of these populations in this study.

Conclusions

PIU is linked to symptoms of depression and anxiety, and our research underscores the crucial role of loneliness as a mediator in this association. Loneliness is responsible for nearly half of the total effects of PIU on depression and anxiety symptoms. Intriguingly, although social media identity may offer some relief from loneliness, it paradoxically exacerbates the impact of PIU on depression and anxiety. In light of these findings, a comprehensive and multifaceted approach to PIU management, encompassing both online and offline interventions, is warranted.

Acknowledgments

This work was supported by the National Natural Science Fund Program of China (grants 72004066 and 71904053), Knowledge Innovation Project of Wuhan (2023020201020471), Ministry of Education of Humanities and Social Science Project of China (24YJAZH086), and China Scholarship Council. Funding bodies advise authors on how to comply with funding body open access policies and how to be reimbursed for any publication fees.

Data Availability

The corresponding author will make the datasets used and analyzed in this study available upon reasonable request.

Authors' Contributions

Conceptualization: CL (lead), YW (equal), DW (equal)
Data curation: YW
Formal analysis: CL (lead), XW (supporting), RL (supporting)
Funding acquisition: CL (equal), DW (equal)
Investigation: YW (lead), XZ (supporting)

XSL•FO
RenderX

Methodology: CL
Supervision: DW (lead), YW (supporting)
Validation: CL
Writing – original draft: CL (lead), YL (supporting)
Writing – review and editing: CL (lead), DW (equal), XW (supporting), RL (supporting), XZ (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 54 KB - jmir_v27i1e57907_app1.docx]

Multimedia Appendix 2
Moderating effects for the all paths of the direct and indirect effects of the mediational model. A: Moderating effect of social media identity on the effect of problematic internet use on loneliness; B: Moderating effect of social media identity on the effect of loneliness on depression; C: Moderating effect of social media identity on the effect of loneliness on anxiety; D: Moderating effect of social media identity on the effect of problematic internet use on depression; E: Moderating effect of social media identity on the effect of problematic internet use on anxiety.
[PNG File , 24 KB - jmir_v27i1e57907_app2.png]

References

1. Collaborators GDaI. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/s2215-0366(21)00395-3]
2. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. The Lancet 2021 Nov;398(10312):1700-1712. [doi: 10.1016/s0140-6736(21)02143-7]
3. Vadivel R, Shoib S, El Halabi S, El Hayek S, Essam L, Gashi Bytyçi D, et al. Mental health in the post-COVID-19 era: challenges and the way forward. Gen Psychiatr 2021;34(1):e100424 [FREE Full text] [doi: 10.1136/gpsych-2020-100424] [Medline: 33644689]
4. Stevens K, Anaya C, Culbert K. Gonadal hormone effects on mental disorders across development. In: Encyclopedia of Mental Health (Third Edition). Oxford: Academic Press; 2023.
5. Richter D, Dixon J. Models of mental health problems: a quasi-systematic review of theoretical approaches. J Ment Health 2023 Apr;32(2):396-406 [FREE Full text] [doi: 10.1080/09638237.2021.2022638] [Medline: 35014924]
6. Helbich M. Mental health and environmental exposures: an editorial. Int J Environ Res Public Health 2018 Oct 10;15(10):2207 [FREE Full text] [doi: 10.3390/ijerph15102207] [Medline: 30308972]
7. Yang X, Guo WJ, Tao YJ, Meng YJ, Wang HY, Li XJ, et al. A bidirectional association between internet addiction and depression: A large-sample longitudinal study among Chinese university students. J Affect Disord 2022 Feb 15;299:416-424. [doi: 10.1016/j.jad.2021.12.013] [Medline: 34906641]
8. Xie L, Yang H, Lin X, Ti S, Wu Y, Zhang S, et al. Does the internet use improve the mental health of Chinese older adults? Front Public Health 2021;9:673368 [FREE Full text] [doi: 10.3389/fpubh.2021.673368] [Medline: 34336769]
9. Odac H, Çelik CB. Who are problematic internet users? An investigation of the correlations between problematic internet use and shyness, loneliness, narcissism, aggression and self-perception. Computers in Human Behavior 2013:2382-2387 [FREE Full text] [doi: 10.1016/j.chb.2013.05.026]
10. Stepanikova I, Nie N, He X. Time on the Internet at home, loneliness, and life satisfaction: Evidence from panel time-diary data. Computers in Human Behavior 2010 May;26(3):329-338 [FREE Full text] [doi: 10.1016/j.chb.2009.11.002]
11. Dahl D, Bergmark KH. Persistence in problematic internet use-A systematic review and meta-analysis. Front Sociol 2020;5:30 [FREE Full text] [doi: 10.3389/fsoc.2020.00030] [Medline: 33869438]
12. Lieberman A, Schroeder J. Two social lives: How differences between online and offline interaction influence social outcomes. Curr Opin Psychol 2020 Feb;31:16-21 [FREE Full text] [doi: 10.1016/j.copsyc.2019.06.022] [Medline: 31386968]
13. Spada MM. An overview of problematic internet use. Addict Behav 2014 Jan;39(1):3-6. [doi: 10.1016/j.addbeh.2013.09.007] [Medline: 24126206]
14. Young KS. Internet addiction: the emergence of a new clinical disorder. CyberPsychology & Behavior 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]
15. Dell'Osso B, Di Bernardo I, Vismara M, Piccoli E, Giorgetti F, Molteni L, et al. Managing problematic usage of the internet and related disorders in an era of diagnostic transition: an updated review. Clin Pract Epidemiol Ment Health 2021;17:61-74 [FREE Full text] [doi: 10.2174/1745017902117010061] [Medline: 34497661]

XSL•FO
RenderX

16. Hidalgo-Fuentes SM, Martí-Vilar M, Ruiz-Ordoñez Y. Problematic internet use and resilience: a systematic review and meta-analysis. Nurs Rep 2023 Feb 24;13(1):337-350 [FREE Full text] [doi: 10.3390/nursrep13010032] [Medline: 36976684]
17. Lakkunarajah S, Adams K, Pan AY, Liegl M, Sadhir M. A trying time: problematic internet use (PIU) and its association with depression and anxiety during the COVID-19 pandemic. Child Adolesc Psychiatry Ment Health 2022 Jun 23;16(1):49 [FREE Full text] [doi: 10.1186/s13034-022-00479-6] [Medline: 35739549]
18. Cai Z, Mao P, Wang Z, Wang D, He J, Fan X. Associations between problematic internet use and mental health outcomes of students: a meta-analytic review. Adolesc Res Rev 2023;8(1):45-62 [FREE Full text] [doi: 10.1007/s40894-022-00201-9] [Medline: 36744121]
19. Kożybska M, Kurpisz J, Radlińska I, Skwirczyńska E, Serwin N, Zabielska P, et al. Problematic Internet Use, health behaviors, depression and eating disorders: a cross-sectional study among Polish medical school students. Ann Gen Psychiatry 2022 Feb 11;21(1):5 [FREE Full text] [doi: 10.1186/s12991-022-00384-4] [Medline: 35148793]
20. Haddad C, Malaeb D, Sacre H, Bou Khalil J, Khansa W, Al Hajj R, et al. Association of problematic internet use with depression, impulsivity, anger, aggression, and social anxiety: Results of a national study among Lebanese adolescents. Pediatr Investig 2021 Dec;5(4):255-264 [FREE Full text] [doi: 10.1002/ped4.12299] [Medline: 34938966]
21. Peng C, Wang M, Cheng J, Tan Y, Huang Y, Rong F, et al. Association between internet addiction and suicidal ideation, suicide plans, and suicide attempts among Chinese adolescents with and without parental migration. Computers in Human Behavior 2021 Dec;125(6):106949-106950. [doi: 10.1016/j.chb.2021.106949]
22. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS One 2017;12(8):e0181722 [FREE Full text] [doi: 10.1371/journal.pone.0181722] [Medline: 28813437]
23. Lei H, Chiu M, Li S. Subjective well-being and internet overuse: A meta-analysis of mainland Chinese students. Curr Psychol 2019 Jun 07;39(3):843-853. [doi: 10.1007/s12144-019-00313-x]
24. Mozafar Saadati H, Mirzaei H, Okhovat B, Khodamoradi F. Association between internet addiction and loneliness across the world: A meta-analysis and systematic review. SSM Popul Health 2021 Dec;16:100948 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100948] [Medline: 34754896]
25. Moretta T, Buodo G. Problematic internet use and loneliness: how complex is the relationship? a short literature review. Curr Addict Rep 2020 Jun 01;7(2):125-136. [doi: 10.1007/s40429-020-00305-z]
26. Kaess M, Durkee T, Brunner R, Carli V, Parzer P, Wasserman C, et al. Pathological Internet use among European adolescents: psychopathology and self-destructive behaviours. Eur Child Adolesc Psychiatry 2014 Nov;23(11):1093-1102 [FREE Full text] [doi: 10.1007/s00787-014-0562-7] [Medline: 24888750]
27. Romer D, Bagdasarov Z, More E. Older versus newer media and the well-being of United States youth: results from a national longitudinal panel. J Adolesc Health 2013 May;52(5):613-619. [doi: 10.1016/j.jadohealth.2012.11.012] [Medline: 23375827]
28. Valkenburg P, Peter J. Online communication and adolescent well-being: testing the stimulation versus the displacement hypothesis. Journal of Computer-Mediated Communication 2007;12(4):1169-1182. [doi: 10.1111/j.1083-6101.2007.00368.x]
29. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
30. Ganda M. Social media and self: Influences on the formation of identity and understanding of self through social networking sites. University Honors Theses 2014:55. [doi: 10.15760/honors.64]
31. Kaakinen M, Sirola A, Savolainen I, Oksanen A. Shared identity and shared information in social media: development and validation of the identity bubble reinforcement scale. Media Psychology 2018 Nov 17;23(1):25-51. [doi: 10.1080/15213269.2018.1544910]
32. Brenden S. One's social media identity: article review. Pepperdine Journal of Communication Research. 2022. URL: https://digitalcommons.pepperdine.edu/pjcr/vol10/iss1/9/ [accessed 2025-02-11]
33. Keipi T, Näsi M, Oksanen A, Räsänen P. Online Hate and Harmful Content: Cross-National Perspectives. London: Routledge; 2016.
34. Kaakinen M, Sirola A, Savolainen I, Oksanen A. Shared identity and shared information in social media: development and validation of the identity bubble reinforcement scale. Media Psychology 2018 Nov 17;23(1):25-51. [doi: 10.1080/15213269.2018.1544910]
35. Savolainen I, Kaakinen M, Sirola A, Koivula A, Hagfors H, Zych I, et al. Online relationships and social media interaction in youth problem gambling: a four-country study. Int J Environ Res Public Health 2020 Nov 03;17(21):A [FREE Full text] [doi: 10.3390/ijerph17218133] [Medline: 33153222]
36. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth 2019 Mar 21;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
37. Oksanen A, Oksa R, Savela N, Kaakinen M, Ellonen N. Cyberbullying victimization at work: Social media identity bubble approach. Computers in Human Behavior 2020 Aug;109:106363. [doi: 10.1016/j.chb.2020.106363]
38. Rumpf H, Effertz T, Montag C. The cost burden of problematic internet usage. Current Opinion in Behavioral Sciences 2022 Apr;44:101107 [FREE Full text] [doi: 10.1016/j.cobeha.2022.101107]

XSL•FO
RenderX

39. Chia DXY, Ng CWL, Kandasami G, Seow MYL, Choo CC, Chew PKH, et al. Prevalence of internet addiction and gaming disorders in Southeast Asia: A meta-analysis. Int J Environ Res Public Health 2020 Apr 09;17(7):2582 [FREE Full text] [doi: 10.3390/ijerph17072582] [Medline: 32283803]
40. Park S, Hong KM, Park EJ, Ha KS, Yoo HJ. The association between problematic internet use and depression, suicidal ideation and bipolar disorder symptoms in Korean adolescents. Aust N Z J Psychiatry 2013 Feb;47(2):153-159. [doi: 10.1177/0004867412463613] [Medline: 23047959]
41. El Asam A, Samara M, Terry P. Problematic internet use and mental health among British children and adolescents. Addict Behav 2019 Mar;90:428-436. [doi: 10.1016/j.addbeh.2018.09.007] [Medline: 30579146]
42. Lam LT. Internet gaming addiction, problematic use of the internet, and sleep problems: a systematic review. Curr Psychiatry Rep 2014 Apr;16(4):444. [doi: 10.1007/s11920-014-0444-1] [Medline: 24619594]
43. Yao M, Zhong Z. Loneliness, social contacts and Internet addiction: A cross-lagged panel study. Computers in Human Behavior 2014 Jan;30:164-170 [FREE Full text] [doi: 10.1016/j.chb.2013.08.007]
44. Tian Y, Guo ZX, Shi JR, Bian YL, Han PG, Wang P, et al. Bidirectional mediating role of loneliness in the association between shyness and generalized pathological internet use in Chinese university students: a longitudinal cross-lagged analysis. J Psychol 2018;152(8):529-547. [doi: 10.1080/00223980.2018.1468309] [Medline: 30376647]
45. Zhang S, Tian Y, Sui Y, Zhang D, Shi J, Wang P, et al. Relationships between social support, loneliness, and internet addiction in Chinese postsecondary students: a longitudinal cross-lagged analysis. Front Psychol 2018;9:1707 [FREE Full text] [doi: 10.3389/fpsyg.2018.01707] [Medline: 30258387]
46. Lewis G, Kounali D, Button KS, Duffy L, Wiles NJ, Munafò MR, et al. Variation in the recall of socially rewarding information and depressive symptom severity: a prospective cohort study. Acta Psychiatr Scand 2017 May;135(5):489-498 [FREE Full text] [doi: 10.1111/acps.12729] [Medline: 28374430]
47. Lee SL, Pearce E, Ajnakina O, Johnson S, Lewis G, Mann F, et al. The association between loneliness and depressive symptoms among adults aged 50 years and older: a 12-year population-based cohort study. Lancet Psychiatry 2021 Jan;8(1):48-57 [FREE Full text] [doi: 10.1016/S2215-0366(20)30383-7] [Medline: 33181096]
48. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
49. Steen OD, Ori APS, Wardenaar KJ, van Loo HM. Loneliness associates strongly with anxiety and depression during the COVID pandemic, especially in men and younger adults. Sci Rep 2022 Jun 09;12(1):9517 [FREE Full text] [doi: 10.1038/s41598-022-13049-9] [Medline: 35681066]
50. Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. Psychol Aging 2006 Mar;21(1):140-151. [doi: 10.1037/0882-7974.21.1.140] [Medline: 16594799]
51. Reinecke L, Trepte S. Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication. Computers in Human Behavior 2014 Jan;30:95-102 [FREE Full text] [doi: 10.1016/j.chb.2013.07.030]
52. van Schalkwyk GI, Marin CE, Ortiz M, Rolison M, Qayyum Z, McPartland JC, et al. Social media use, friendship quality, and the moderating role of anxiety in adolescents with autism spectrum disorder. J Autism Dev Disord 2017 Sep;47(9):2805-2813 [FREE Full text] [doi: 10.1007/s10803-017-3201-6] [Medline: 28616856]
53. Krishnamoorthy Y, Nagarajan R, Saya G, Menon V. Prevalence of psychological morbidities among general population, healthcare workers and COVID-19 patients amidst the COVID-19 pandemic: A systematic review and meta-analysis. Psychiatry Res 2020 Nov;293:113382 [FREE Full text] [doi: 10.1016/j.psychres.2020.113382] [Medline: 32829073]
54. Wang Y, Kaierdebieke A, Fan S, Zhang R, Huang M, Li H, et al. Study protocol: A cross-sectional study on psychology and behavior investigation of Chinese residents, PBICR. Psychosom Med Res 2022;4(3):19. [doi: 10.53388/202219]
55. Demetrovics Z, Király O, Koronczai B, Griffiths MD, Nagygyörgy K, Elekes Z, et al. Psychometric properties of the Problematic Internet Use Questionnaire Short-Form (PIUQ-SF-6) in a nationally representative sample of adolescents. PLoS One 2016;11(8):e0159409 [FREE Full text] [doi: 10.1371/journal.pone.0159409] [Medline: 27504915]
56. Opakunle T, Aloba O, Opakunle O, Eegunranti B. Problematic Internet Use Questionnaire-Short Form-6 (PIUQ-SF-6): Dimensionality, validity, reliability, measurement invariance and mean differences across genders and age categories among Nigerian adolescents. International Journal of Mental Health 2020 Jun 12;49(3):229-246. [doi: 10.1080/00207411.2020.1776457]
57. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
58. Czerwiński SK, Atroszko P. A solution for factorial validity testing of three-item scales: An example of tau-equivalent strict measurement invariance of three-item loneliness scale. Curr Psychol 2023;42(2):1652-1664 [FREE Full text] [doi: 10.1007/s12144-021-01554-5] [Medline: 33716473]
59. Trucharte A, Calderón L, Cerezo E, Contreras A, Peinado V, Valiente C. Three-item loneliness scale: psychometric properties and normative data of the Spanish version. Curr Psychol 2023;42(9):7466-7474 [FREE Full text] [doi: 10.1007/s12144-021-02110-x] [Medline: 34305365]

XSL•FO
RenderX

60. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). Journal of Affective Disorders 2004 Jul;81(1):61-66. [doi: 10.1016/s0165-0327(03)00198-8]
61. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
62. He XL, Qian J. Reliability and validity of a generalized anxiety scale in general hospital outpatients. Shanghai Arch Psychiatry 2010;22(4):200-203. [doi: 10.3969/j.issn.1002-0829.2010.04.002]
63. Ye X, Shu HL, Feng X, Xia DM, Wang ZQ, Mi WY, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire-9 (C-PHQ-9) in patients with psoriasis: a cross-sectional study. BMJ Open 2020 Jul 13;10(7):e033211 [FREE Full text] [doi: 10.1136/bmjopen-2019-033211] [Medline: 32665341]
64. Anderson EL, Steen E, Stavropoulos V. Internet use and Problematic Internet Use: a systematic review of longitudinal research trends in adolescence and emergent adulthood. International Journal of Adolescence and Youth 2016 Oct 10;22(4):430-454. [doi: 10.1080/02673843.2016.1227716]
65. Dahlberg L, McKee KJ, Frank A, Naseer M. A systematic review of longitudinal risk factors for loneliness in older adults. Aging Ment Health 2022 Feb;26(2):225-249 [FREE Full text] [doi: 10.1080/13607863.2021.1876638] [Medline: 33563024]
66. Kong X, Wu Y, Wang X, Sun Y, Chen K, Li Q, et al. Analysis of the prevalence and influencing factors of anxiety and depression in the Chinese population: A cross-sectional survey. Heliyon 2023 May;9(5):e15889 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e15889] [Medline: 37215825]
67. Zaixing L. Research on the Overall Distribution of Productive Forces in China. Beijing: China Price Press; 1995.
68. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: Guilford Press; 2013.
69. Liu M, Xiao J, Kamper-DeMarco K, Fu Z. Problematic internet use and suicidality and self-injurious behaviors in adolescents: Effects of negative affectivity and social support. J Affect Disord 2023 Mar 15;325:289-296. [doi: 10.1016/j.jad.2023.01.019] [Medline: 36627060]
70. Moody EJ. Internet use and its relationship to lonelines. Cyberpsychol Behav 2001 Jun;4(3):393-401. [doi: 10.1089/109493101300210303] [Medline: 11710265]
71. Alheneidi H, AlSumait L, AlSumait D, Smith AP. Loneliness and problematic internet use during COVID-19 lockdown. Behav Sci (Basel) 2021 Jan 06;11(1):5 [FREE Full text] [doi: 10.3390/bs11010005] [Medline: 33418914]
72. Fang X, Tian M, Wang R, Wang P. Relationships between depression, loneliness and pathological Internet use in adolescents: A cross-lagged analysis. Curr Psychol 2022 May 13;42(24):20696-20706. [doi: 10.1007/s12144-022-03180-1]
73. Luo Y, Hawkley LC, Waite LJ, Cacioppo JT. Loneliness, health, and mortality in old age: a national longitudinal study. Soc Sci Med 2012 Mar;74(6):907-914 [FREE Full text] [doi: 10.1016/j.socscimed.2011.11.028] [Medline: 22326307]
74. Domènech-Abella J, Mundó J, Haro JM, Rubio-Valera M. Anxiety, depression, loneliness and social network in the elderly: Longitudinal associations from The Irish Longitudinal Study on Ageing (TILDA). J Affect Disord 2019 Mar 01;246:82-88. [doi: 10.1016/j.jad.2018.12.043] [Medline: 30578950]
75. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychol Aging 2010 Jun;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]
76. Beutel ME, Brähler E, Wiltink J, Kerahrodi JG, Burghardt J, Michal M, et al. New onset of depression in aging women and men: contributions of social, psychological, behavioral, and somatic predictors in the community. Psychol Med 2018 Aug 22;49(07):1148-1155. [doi: 10.1017/s0033291718001848]
77. Pearce E, Myles-Hooton P, Johnson S, Hards E, Olsen S, Clisu D, et al. Loneliness as an active ingredient in preventing or alleviating youth anxiety and depression: a critical interpretative synthesis incorporating principles from rapid realist reviews. Transl Psychiatry 2021 Dec 10;11(1):628 [FREE Full text] [doi: 10.1038/s41398-021-01740-w] [Medline: 34893578]
78. Santini ZI, Koyanagi A, Tyrovolas S, Mason C, Haro JM. The association between social relationships and depression: a systematic review. J Affect Disord 2015 Apr 01;175:53-65. [doi: 10.1016/j.jad.2014.12.049] [Medline: 25594512]
79. Guo L, Luo M, Wang W, Huang G, Xu Y, Gao X, et al. Association between problematic Internet use, sleep disturbance, and suicidal behavior in Chinese adolescents. J Behav Addict 2018 Dec 01;7(4):965-975 [FREE Full text] [doi: 10.1556/2006.7.2018.115] [Medline: 30474380]
80. Wang W, Du X, Guo Y, Li W, Zhang S, Guo L, et al. Association between problematic internet use and behavioral/emotional problems among Chinese adolescents: the mediating role of sleep disorders. PeerJ 2021;9:e10839 [FREE Full text] [doi: 10.7717/peerj.10839] [Medline: 33665021]
81. Ridings C, Gefen D. Virtual community attraction: Why people hang out online. Journal of Computer-Mediated Communication 2004;10(1):JCMC10110 [FREE Full text] [doi: 10.1111/j.1083-6101.2004.tb00229.x]
82. Kamalpour M, Watson J, Buys L. How can online communities support resilience factors among older adults. International Journal of Human–Computer Interaction 2020 Apr 14;36(14):1342-1353. [doi: 10.1080/10447318.2020.1749817]

XSL•FO
RenderX

83. Bu F, Steptoe A, Fancourt D. Loneliness during a strict lockdown: Trajectories and predictors during the COVID-19 pandemic in 38,217 United Kingdom adults. Soc Sci Med 2020 Nov;265:113521 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113521] [Medline: 33257177]
84. Latikka R, Koivula A, Oksa R, Savela N, Oksanen A. Loneliness and psychological distress before and during the COVID-19 pandemic: Relationships with social media identity bubbles. Soc Sci Med 2022 Jan;293:114674 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114674] [Medline: 34959045]
85. Koivula A, Kaakinen M, Oksanen A, Räsänen P. The role of political activity in the formation of online identity bubbles. Policy & Internet 2019 Jul 15;11(4):396-417. [doi: 10.1002/poi3.211]
86. Kaakinen M, Keipi T, Oksanen A, Räsänen P. How does social capital associate with being a victim of online hate? Survey evidence from the US, UK, Germany and Finland. Policy & Internet 2018 Jun 15;10(3):302-323. [doi: 10.1002/poi3.173]
87. Mesch G, Talmud I. The quality of online and offline relationships: the role of multiplexity and duration of social relationships. The Information Society 2006 Jun;22(3):137-148. [doi: 10.1080/01972240600677805]
88. Wang D, Sun X, He F, Liu C, Wu Y. The mediating effect of family health on the relationship between health literacy and mental health: A national cross-sectional survey in China. Int J Soc Psychiatry 2023 Sep;69(6):1490-1500. [doi: 10.1177/00207640231166628] [Medline: 37095729]
89. Terlizzi EV, Villarroel MA. Symptoms of generalized anxiety disorder among adults: United States, 2019. NCHS Data Brief 2020 Sep(378):1-8. [Medline: 33054928]
90. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
91. Karacic S, Oreskovic S. Internet addiction through the phase of adolescence: a questionnaire study. JMIR Ment Health 2017 Apr 03;4(2):e11 [FREE Full text] [doi: 10.2196/mental.5537] [Medline: 28373154]
92. Endomba FT, Demina A, Meille V, Ndoadoumgue AL, Danwang C, Petit B, et al. Prevalence of internet addiction in Africa: A systematic review and meta-analysis. J Behav Addict 2022 Sep 26;11(3):739-753 [FREE Full text] [doi: 10.1556/2006.2022.00052] [Medline: 35984734]
93. Shi J, van der Maas M, Yu L, Jiang Q, Agasee S, Turner N. Current prevention strategies and future directions for problem Internet use. Current Opinion in Behavioral Sciences 2022 Dec;48:101231 [FREE Full text] [doi: 10.1016/j.cobeha.2022.101231]
94. The State Press and Publication Administration on further strict management on effectively preventing minors from indulging in online games. National Press and Publication Administration. URL: https://www.gov.cn/zhengce/zhengceku/2021-09/01/content_5634661.htm [accessed 2025-02-11]
95. Shi J, van der Maas M, Yu L, Jiang Q, Agasee S, Turner N. Current prevention strategies and future directions for problem Internet use. Current Opinion in Behavioral Sciences 2022 Dec;48:101231 [FREE Full text] [doi: 10.1016/j.cobeha.2022.101231]
96. Masi CM, Chen HY, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug;15(3):219-266 [FREE Full text] [doi: 10.1177/1088868310377394] [Medline: 20716644]
97. Wright M, Heiman T, Olenik-Shemesh D. Problematic internet use: causes, consequences, future directions. In: Child and Adolescent Online Risk Exposure. New York: Academic Press; 2021:13-32.

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder scale
PBICR: Psychology and Behavior Investigation of Chinese Residents
PHQ-9: 9-item Patient Health Questionnaire
PIU: problematic internet use

Edited by T de Azevedo Cardoso; submitted 02.03.24; peer-reviewed by X Wang, X Lin; comments to author 06.11.24; revised version received 14.01.25; accepted 31.01.25; published 26.02.25.

Please cite as:
Liu C, Liu Y, Liu C, Lin R, Wang X, Zhang X, Wu Y, Wang D
The Moderated Mediating Effects of Social Media Identity and Loneliness on the Relationship Between Problematic Internet Use and Mental Health in China: Nationwide Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e57907
URL: https://www.jmir.org/2025/1/e57907
doi:10.2196/57907
PMID:

XSL•FO
RenderX

©Chenxi Liu, Yushu Liu, Chaojie Liu, Rujiao Lin, Xi Wang, Xinyi Zhang, Yibo Wu, Dan Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Competency of Large Language Models in Evaluating Appropriate Responses to Suicidal Ideation: Comparative Study

Ryan K McBain[1,2,3], MPH, PhD; Jonathan H Cantor[4], PhD; Li Ang Zhang[4], PhD; Olesya Baker[3,5], PhD; Fang Zhang[3,5], PhD; Alyssa Halbisen[5], PhD; Aaron Kofner[1], MS; Joshua Breslau[6], PhD, ScD; Bradley Stein[6], MD, PhD; Ateev Mehrotra[7], MPH, MD; Hao Yu[3,5], PhD

[1]RAND, Arlington, VA, United States
[2]Brigham and Women's Hospital, Boston, MA, MA, United States
[3]Harvard Medical School, Boston, MA, United States
[4]RAND, Santa Monica, CA, United States
[5]Harvard Pilgrim Health Care Institute, Boston, MA, United States
[6]RAND, Pittsburgh, PA, United States
[7]Brown University School of Public Health, Providence, RI, United States

Corresponding Author:
Ryan K McBain, MPH, PhD
RAND
1200 S Hayes St
Arlington, VA
United States
Phone: 1 5088433901
Email: rmcbain@rand.org

Abstract

Background: With suicide rates in the United States at an all-time high, individuals experiencing suicidal ideation are increasingly turning to large language models (LLMs) for guidance and support.

Objective: The objective of this study was to assess the competency of 3 widely used LLMs to distinguish appropriate versus inappropriate responses when engaging individuals who exhibit suicidal ideation.

Methods: This observational, cross-sectional study evaluated responses to the revised Suicidal Ideation Response Inventory (SIRI-2) generated by ChatGPT-4o, Claude 3.5 Sonnet, and Gemini 1.5 Pro. Data collection and analyses were conducted in July 2024. A common training module for mental health professionals, SIRI-2 provides 24 hypothetical scenarios in which a patient exhibits depressive symptoms and suicidal ideation, followed by two clinician responses. Clinician responses were scored from –3 (highly inappropriate) to +3 (highly appropriate). All 3 LLMs were provided with a standardized set of instructions to rate clinician responses. We compared LLM responses to those of expert suicidologists, conducting linear regression analyses and converting LLM responses to z scores to identify outliers (z score>1.96 or <–1.96; P<0.05). Furthermore, we compared final SIRI-2 scores to those produced by health professionals in prior studies.

Results: All 3 LLMs rated responses as more appropriate than ratings provided by expert suicidologists. The item-level mean difference was 0.86 for ChatGPT (95% CI 0.61-1.12; P<.001), 0.61 for Claude (95% CI 0.41-0.81; P<.001), and 0.73 for Gemini (95% CI 0.35-1.11; P<.001). In terms of z scores, 19% (9 of 48) of ChatGPT responses were outliers when compared to expert suicidologists. Similarly, 11% (5 of 48) of Claude responses were outliers compared to expert suicidologists. Additionally, 36% (17 of 48) of Gemini responses were outliers compared to expert suicidologists. ChatGPT produced a final SIRI-2 score of 45.7, roughly equivalent to master's level counselors in prior studies. Claude produced an SIRI-2 score of 36.7, exceeding prior performance of mental health professionals after suicide intervention skills training. Gemini produced a final SIRI-2 score of 54.5, equivalent to untrained K-12 school staff.

Conclusions: Current versions of 3 major LLMs demonstrated an upward bias in their evaluations of appropriate responses to suicidal ideation; however, 2 of the 3 models performed equivalent to or exceeded the performance of mental health professionals.

(J Med Internet Res 2025;27:e67891) doi:10.2196/67891

XSL•FO
RenderX

KEYWORDS

depression; suicide; mental health; large language model; chatbot; digital health; Suicidal Ideation Response Inventory; ChatGPT; suicidologist; artificial intelligence

Introduction

Suicide is one of the leading causes of death among individuals under the age of 50 in the United States, and it is the second leading cause of death among adolescents [1]. Rates of suicide have also grown sharply in recent years; 39,518 suicide deaths were reported in 2011, compared to 48,183 in 2021. Although this trajectory declined during the COVID-19 pandemic, more recent data indicate the upward trend has resumed [2].

Large language models (LLMs) have drawn widespread attention as a potential vehicle for helping or harming individuals who are depressed and at risk of suicide [3]. LLMs are designed to interpret and generate human-like text responses to written and spoken queries, and they include broad health applications [4]. Platforms like ChatGPT, as well as mental health apps powered by LLMs, offer an outlet to individuals looking for therapeutic advice on how to cope with depressive symptoms, loneliness, and thoughts of suicide [5,6]. This could be particularly beneficial for the roughly 50 million Americans living in rural parts of the United States with poor access to mental health care [7] or for those who cannot afford the cost of therapy [8,9].

On the other hand, researchers and advocates fear that LLMs could make poor, if not outright injurious, recommendations when engaging with individuals with who express suicidal ideation [10]. Only a small handful of studies in the peer-reviewed literature have evaluated the competencies of LLMs when addressing individuals who exhibit depressive symptoms and suicidal ideation [11-14]. Typically, these studies quantify behaviors of LLMs, such as making an initial referral to a human, rather than directly comparing LLM performance to standardized benchmarks [12]. For LLMs to improve on tasks such as offering therapeutic advice, reinforcement learning from human feedback, a process in which human feedback from experts is used to align LLM responses with best practices and ethical standards in mental health care, is required.

In this study, we assessed the competency of 3 widely used LLMs to distinguish appropriate versus inappropriate responses when engaging individuals who exhibit suicidal ideation. To do so, we compared the responses of LLMs to responses by expert suicidologists on a previously-published standardized scale: the Suicide Intervention Response Inventory (SIRI-2) [15]. We hypothesized that ratings generated by LLMs would differ significantly from those of expert suicidologists. We also hypothesized that LLMs would not demonstrate a consistent bias—meaning they would not systematically rate responses as more or less appropriate, compared to expert suicidologists.

Methods

Ethical Considerations

This observational cross-sectional study was deemed exempt from review by Harvard Medical School's Institutional Review Board (reference #2116708). The study followed Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

Measures

The SIRI-2 contains 2 sets of 24 hypothetical patient remarks that disclose depressive symptoms and suicidal thoughts (for example: "I don't think there's really anyone who cares whether I'm alive or dead"). Patient remarks are followed by 2 possible clinician responses, each of which could be considered supportive, neutral, or deleterious to suicide prevention (for example, Clinician A: "It sounds like you're feeling pretty isolated"; Clinician B: "Why do you think that no one cares about you anymore?") [15].

LLMs were provided the original instructions from the SIRI-2, which state that respondents were to assign each clinician response a score from –3 (highly inappropriate response) to +3 (highly appropriate response), on a 7-point ordinal scale. Expert suicidologists were not recruited as part of this study but rather, were previously recruited by the instrument's authors to create a normative benchmark by which to compare responses from other individuals [15]. In the context of this study, we compared the responses of LLMs to those of these previously recruited expert suicidologists. The final SIRI-2 score is represented as the sum of differences between LLMs' and experts' ratings; a lower score indicates greater alignment between LLMs and expert suicidologists.

Previous research has reported the SIRI-2 scores for a wide range of individuals—such as doctoral students in clinical psychology, master's level counselors, and K-12 school staff (see Table 1) [16-20]. Human performance on these evaluations therefore serves as a reference point for which we could compare LLM performance.

XSL•FO
RenderX

Table 1. Prior studies assessing human performance on the Suicide Intervention Response Inventory (SIRI-2).

Study authors and date	Study setting	Cadre assessed	Pre- or post-training[a]	SIRI-2 score[b]
Fujisawa et al [18], 2013	Japan	Second-year medical residents	Pretraining	68.2
Kawashima et al [21], 2020	Japan	Clinical psychologists	Pretraining	48.8
Kawashima et al [21], 2020	Japan	Social workers	Pretraining	62.3
Kawashima et al [21], 2020	Japan	Nurses	Pretraining	61.3
Machelprang et al [19], 2014	United States	Clinical psychology PhD students	N/A[c]	45.4
Morriss et al [20], 1999	United Kingdom	Front-line health workers	Pretraining	56.8
Morriss et al [20], 1999	United Kingdom	Front-line health workers	Post-training	46.4
Neimeyer and Bonnelle [15], 1997	United States	Master's level counselors	Pretraining	54.7
Neimeyer and Bonnelle [15], 1997	United States	Master's level counselors	Post-training	41.0
Palimieri et al [22], 2008	Italy	Psychiatrists	N/A	55.7
Palimieri et al [22], 2008	Italy	Emergency physicians	N/A	63.9
Palimieri et al [22], 2008	Italy	Psychiatric nurses	N/A	71.3
Palimieri et al [22], 2008	Italy	General practitioners	N/A	91.1
Scheerder et al [23], 2010	Belgium	Community mental health centers staff	N/A	47.4
Scheerder et al [23], 2010	Belgium	Experienced volunteers at a suicide crisis line	N/A	47.5
Scheerder et al [23], 2010	Belgium	General practitioners	N/A	51.1
Scheerder et al [23], 2010	Belgium	Hospital nurses	N/A	54.4
Shannonhouse et al [16], 2017a	United States	K-12 school staff	Pretraining	52.9
Shannonhouse et al [16], 2017a	United States	K-12 school staff	Post-training	49.9
Shannonhouse et al [17], 2017b	United States	College staff	Pretraining	52.9
Shannonhouse et al [17], 2017b	United States	College staff	Post-training	50.1

[a]Pretraining represents measurement of individuals prior to suicide intervention response training, while post-training represents measurement of individuals after suicide intervention response training.

[b]A lower score is considered better on the SIRI-2. Values are reported to the tenths place.

[c]N/A: not applicable. N/A indicates studies that did not conduct pre- and post-training analyses.

Procedures

Using ChatGPT-4o, Claude 3.5 Sonnet, and Gemini 1.5 Pro, we conducted a series of assessments from June to July 2024. Three members of the research team created separate accounts to interact with and prompt LLMs. Research team members prompted LLMs with the original instructions for the SIRI-2, as well as with one of the SIRI2-2's 24 items. We did not prompt LLMs with any additional text. We used this approach to evaluate how LLMs responded without further prompting strategies (ie, methods such as chain-of-thought, in which the responses of LLMs are guided by additional instructions, contextual information, or examples) [24]. See Multimedia Appendix 1 for an overview of the data collection workflow.

The 3 research team members recorded responses provided by LLMs. They also documented any rationale provided by LLMs for the scores they assigned (see Multimedia Appendix 2 for this information).

Statistical Analysis

As a first step, we summarized responses generated by LLMs and expert suicidologists, reporting mean scores and SDs on each of the 24 items. For LLMs, these values were computed across the 3 sets of responses generated by team members. We also examined alignment between LLM and expert responses, measured as the magnitude of the correlation coefficients between the two. Next, we inspected test-retest reliability of each of the LLM's responses, a marker of the consistency and stability of an LLM's responses over time. This was measured as the mean correlation coefficient across the 3 instances in which each LLM response set was generated.

Following this, we conducted 2 sets of inferential analyses. First, we conducted linear regression analysis in Stata 17.1 (StataCorp) to compare item-level responses assigned by each LLM to those assigned by expert suicidologists. The dependent variable in the model was the item score (–3 to +3). The 2 independent variables were (1) respondent type (ie, LLM vs expert) and (2) survey item number (eg, Item 1, Item 2). This specification allowed us to test whether LLMs produced

XSL•FO
RenderX

systematically different scores from experts, while also accounting for the nested structure of the data. For example, item scores (from –3 to +3) were nested within survey items.

Second, based on mean scores and corresponding SDs from expert suicidologists, we calculated *z* scores for each item-level response generated by LLMs. We then quantified the average *z* score for an LLM's responses, as well as the number and percent of *z* scores that were statistically significant (ie, *z* scores greater than 1.96 or less than –1.96). This provided an indication of overall alignment between an LLM's and experts' responses.

Lastly, we calculated final SIRI-2 scores for each LLM and compared these to the performance of humans in prior studies, including the performance of mental health professionals with and without training on suicide intervention response.

Results

Descriptive Statistic

Expert suicidologists reported a mean score of –0.20 (SD 2.22) across all items, meaning that the average response approximated "neither appropriate, nor inappropriate", but item-level responses varied widely. By comparison, mean scores for ChatGPT-4o, Claude 3.5 Sonnet, and Gemini 1.5 were 0.67 (SD 2.41), 0.41 (SD 2.51), and 0.53 (SD 1.73), respectively, meaning that responses tended to skew more toward "appropriate" compared to "inappropriate". ChatGPT-4o assigned a higher score than experts for 40 of 48 responses (83%). Claude 3.5 generated a higher score on 39 responses (81%), and Gemini 1.5 generated a higher score on 36 responses (75%; see Figure 1).

Figure 1. Mean difference in ratings on Suicidal Ideation Response Inventory (SIRI-2) items: large language model versus expert suicidologists.

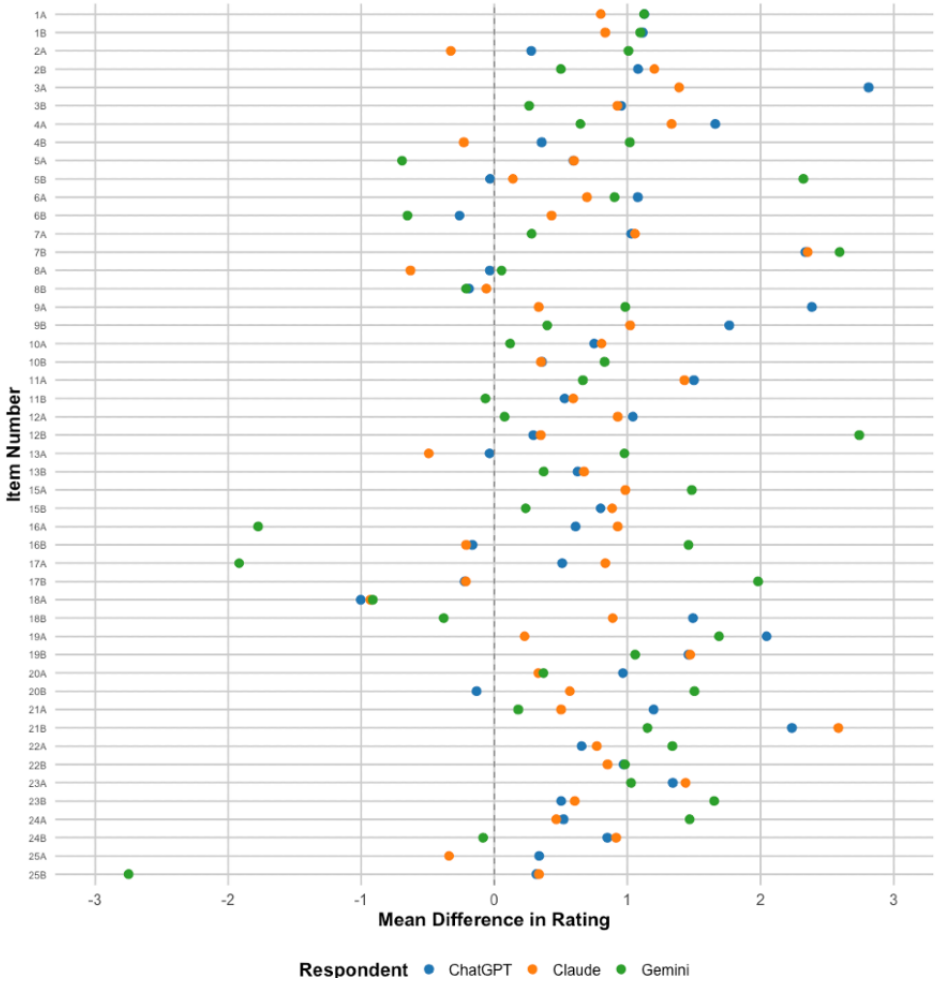

The correlation between LLM and expert responses was 0.93 for ChatGPT-4o, 0.96 for Claude 3.5, and 0.81 for Gemini 1.5. In terms of test-retest reliability, mean test-retest correlation coefficients were 0.98 for ChatGPT-4o, 0.99 for Claude 3.5 Sonnet, and 0.73 for Gemini 1.5, indicating high reliability for all 3 LLMs.

Regression Analyses: Bias

In our regression, LLMs assigned significantly higher scores to hypothetical responses, compared to expert suicidologists, indicating LLMs perceived responses as more appropriate than experts did (see Table 2) with the mean difference in item-level scores being 0.865 (95% CI 0.613-1.118; *P*<.001) for ChatGPT-4o, 0.608 (95% CI 0.408-0.809; *P*<.001) for Claude 3.5 Sonnet, and 0.733 (95% CI 0.352-1.114; *P*<.001) for Gemini 1.5.

XSL•FO
RenderX

Table 2. Estimated difference in perceived appropriateness of responses to suicidal ideation.

LLM model and version	Bias		Performance		
	Score difference[a] (95% CI)	*P* value	Mean *z* score	*Z* scores with an SD of >1.96, n (%)[b]	SIRI-2[c] score
ChatGPT-4o	0.865 (0.613-1.118)	<.001	1.17	9 (19.1)	45.71
Claude 3.5 Sonnet	0.608 (0.408-0.809)	<.001	1.01	5 (10.6)	36.65
Gemini 1.5 Pro	0.733 (0.352-1.114)	<.001	1.54	17 (36.2)	54.52

[a]Average difference represents the mean difference in units, on a 7-point ordinal scale, between an LLM model's responses and expert suicidologists' responses.

[b]*Z* scores were generated for 47 of 48 responses, as 1 item had a SD of 0.

[c]SIRI-2: Suicide Intervention Response Inventory. A lower score is considered better on the SIRI-2.

Overall Performance

Across all items, the average *z* score for ChatGPT-4o responses was 1.17, with 9 responses (19%) greater than 1.96 SDs (all *P*<.05) from the mean responses by expert suicidologists (see Figure 2). The average *z* score for Claude 3.5 Sonnet responses was 1.01, with 5 (11%) responses greater than 1.96 SDs (all *P*<.05) from the mean expert responses. Lastly, the average *z* score for Gemini 1.5 Pro responses was 1.54, with 17 (36%) responses greater than 1.96 SDs (all *P*<.05) from the mean responses by experts. In terms of final SIRI-2 scores, these were 45.71 for ChatGPT-4o, 54.52 for Gemini 1.5 Pro, and 36.65 for Claude 3.5 Sonnet. We note that the lowest possible score, for which expert suicidologists serve as the reference point, was 12.90.

Figure 2. Density plot represents the proportion of responses, across all 48 item responses, with *z* scores ranging from –3 to +6. Dashed vertical lines indicate cutoff thresholds of –1.96 and +1.96. Values less than –1.96 or greater than +1.96 are significant at *P*<.05.

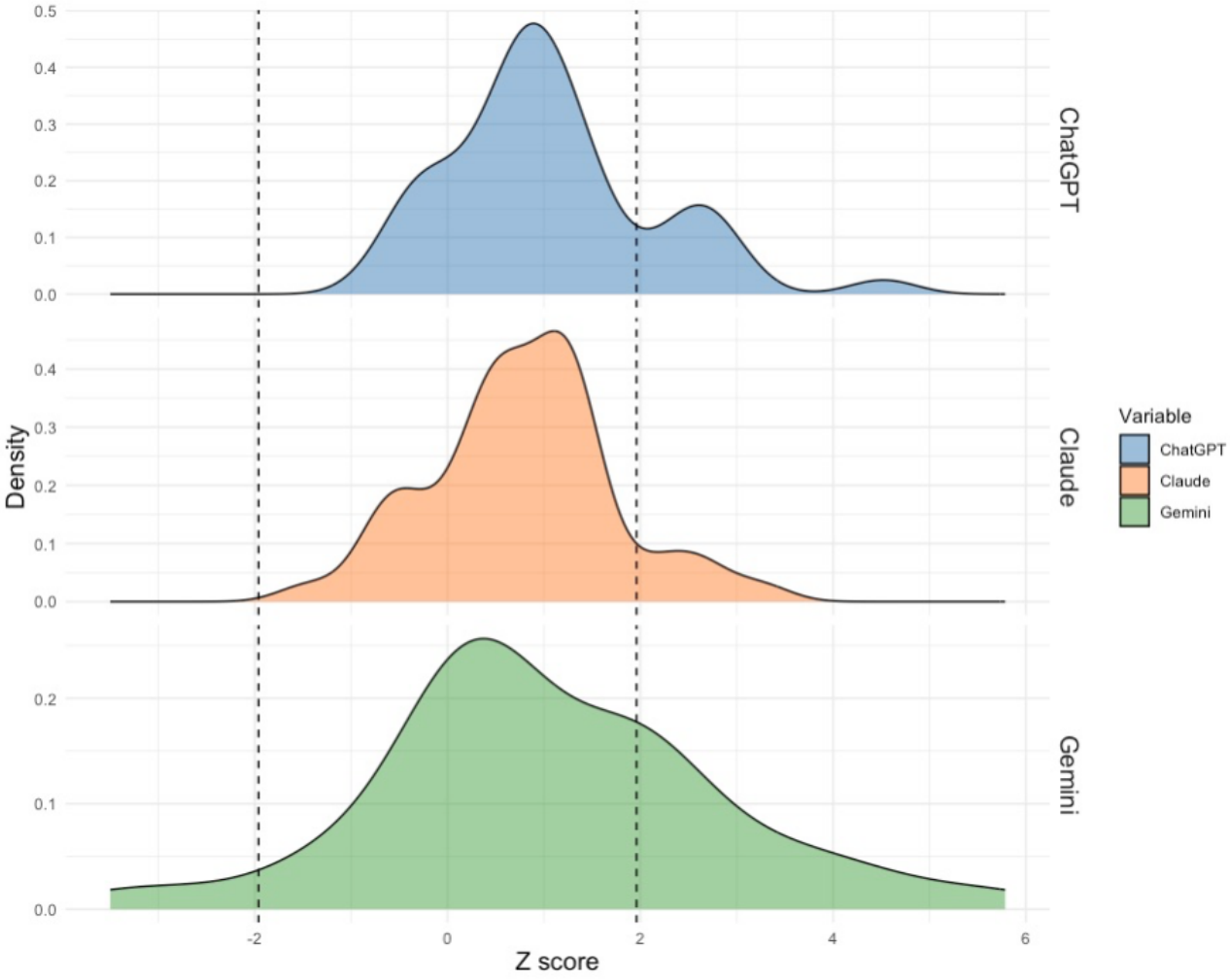

XSL•FO
RenderX

Discussion

We evaluated the capacity of 3 LLMs to assess the appropriateness of responses to 24 scenarios in which a hypothetical individual disclosed depressive symptoms and suicidal thoughts. Compared to the ratings of expert suicidologists, the evaluations of the 3 LLMs were highly correlated but demonstrated an upward bias toward rating responses as more appropriate. Similar biases have been identified in other domains of LLM performance, such as a tendency to over-assign medical diagnoses to individuals of particular demographic backgrounds [25].

LLMs' overall performance as measured by SIRI-2 score—which captures the magnitude of their deviations from expert suicidologists—varied across models. The final score produced by Gemini (54.52) was roughly equivalent to past scores produced by K-12 school staff *prior to* suicide intervention skills training [16]. By contrast, the final score produced by ChatGPT (45.71) was closer to those exhibited by doctoral students in clinical psychology [19] or master's level counselors [15]. Claude observed the strongest performance (36.65), surpassing scores observed even among individuals who recently completed suicide intervention skills training, as well as studies with psychiatrists and other mental health professionals [21-23].

A key issue in this study is whether a competency in adjudicating appropriate responses to suicidal ideation translates to a competency in responding to individuals disclosing suicidal ideation. Serving as referee is not the same as active engagement. The findings of this study also highlight a standard path forward for companies developing and refining LLMs for therapeutic purposes: namely, to consider indexing LLM responses against high-quality benchmarks, such as ratings of expert suicidologists. Instruments such as the SIRI-2 offer rare touchstones for this. A complementary model involves reinforcement learning from human feedback, in which expert clinicians provide direct evaluations of LLM performance relative to a set of pre-established criteria and best practices [26,27].

When used for therapeutic purposes, LLMs will likely encounter users with suicidal ideation on a routine basis. Roughly 1 in 4 mental health professionals encounter suicidal ideation among their patients [28]. Widespread use of LLM technology—including new companies already drawing on LLM technology for mental health care [29]—could reach a much wider audience of individuals coping with depression and suicidal thoughts. To date, a common guardrail has been for LLMs to produce "hard stops", in which individuals are referred to 988 or another suicide prevention hotline. While such referrals may be beneficial, they also artificially circumscribed interactions in a way that could be taken as a missed opportunity.

There are several important study limitations to note. First, LLM technologies are constantly evolving. This study offers a snapshot of LLM performance in July 2024. Second, we selected the SIRI-2 as an evaluative tool because it is widely used; however, alternative instruments could result in different findings. Third, as noted above, this study focuses on the evaluative competencies of LLMs rather than their abilities to directly respond to suicidal ideation. While there are many prompting strategies designed to elicit better performance of LLMs [30], the goal of this study was to test how LLMs evaluate responses to suicidal ideation in conversations without any additional guidance. This is similar to LLM alignment studies where fictitious scenarios are presented without specific prompting strategies and LLM responses are evaluated [31]. Lastly, we note that the authors of the SIRI-2 constructed the original panel of expert suicidologists, and as such, our research team (and other users of the SIRI-2) lack information regarding their average years of clinical practice.

In summary, this study highlights the potential and limitations of 3 widely used LLMs to assess appropriate responses to individuals exhibiting suicidal ideation. While current LLM versions exhibit a preferential bias toward viewing responses as appropriate, their overall performance was on-par with or otherwise exceeded those documented in prior human studies. Claude 3.5 Sonnet surpassed other LLMs by a sizable margin. Future research might explore alternative configurations in which LLMs directly respond to suicidal ideation; although, benchmarks for index performance in these scenarios are uncommon.

Acknowledgments

We would like to thank Nabeel Qureshi for his assistance collating output from LLMs, which were used for analyses in this manuscript. This work was supported by a grant from the National Institute of Mental Health (1R01MH132551). The content is solely the responsibility of the authors and does not necessarily represent the views of National Institute of Mental Health. Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

RKM had full access to the data in the study and takes responsibility for data integrity and data analysis. RKM, JHC, and AM were responsible for the concept and design. RKM and JHC were responsible for drafting the initial manuscript. Critical review of the manuscript was provided by LAZ, OB, FZ, AH, AK, JB, BS, AM, and HY. Statistical analysis was conducted by RKM.

XSL•FO
RenderX

HY obtained funding. Administrative, technical, and material support was provided by AH. JB, BS, AM, and HY provided supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection workflows.
[PNG File , 695 KB - jmir_v27i1e67891_app1.png]

Multimedia Appendix 2
Large language model qualitative explanations for item-level responses to the Suicidal Ideation Response Inventory (SIRI-2).
[DOCX File , 46 KB - jmir_v27i1e67891_app2.docx]

References

1. Suicide. National Institute of Mental Health. 2024. URL: https://www.nimh.nih.gov/health/statistics/suicide [accessed 2024-07-01]
2. Saunders H, Panchal N. A look at the latest suicide data and change over the last decade. Kaiser Family Foundation. 2023 Aug 04. URL: https://www.kff.org/mental-health/issue-brief/a-look-at-the-latest-suicide-data-and-change-over-the-last-decade/ [accessed 2024-07-01]
3. Omar M, Soffer S, Charney AW, Landi I, Nadkarni GN, Klang E. Applications of large language models in psychiatry: a systematic review. Front Psychiatry 2024 Jun 24;15:1422807 [FREE Full text] [doi: 10.3389/fpsyt.2024.1422807] [Medline: 38979501]
4. Karabacak M, Margetis K. Embracing large language models for medical applications: opportunities and challenges. Cureus 2023 May;15(5):e39305 [FREE Full text] [doi: 10.7759/cureus.39305] [Medline: 37378099]
5. Mental health apps and the role of ai in emotional well-being. Mya Care. 2023 Nov 08. URL: https://myacare.com/blog/mental-health-apps-and-the-role-of-ai-in-emotional-wellbeing [accessed 2024-07-15]
6. Rawat M. Best AI apps for mental health (2023). MarkTechPost. 2023 Apr 11. URL: https://www.marktechpost.com/2023/04/11/best-ai-apps-for-mental-health-2023/ [accessed 2024-07-15]
7. Ziller EC, Anderson NJ, Coburn AF. Access to rural mental health services: service use and out-of-pocket costs. J Rural Health 2010;26(3):214-224. [doi: 10.1111/j.1748-0361.2010.00291.x] [Medline: 20633089]
8. Donohue JM, Goetz JL, Song Z. Who gets mental health care?-The role of burden and cash-paying markets. JAMA Health Forum 2024 Mar 01;5(3):e240210 [FREE Full text] [doi: 10.1001/jamahealthforum.2024.0210] [Medline: 38517427]
9. Coombs NC, Meriwether WE, Caringi J, Newcomer SR. Barriers to healthcare access among U.S. adults with mental health challenges: a population-based study. SSM Popul Health 2021 Sep;15:100847 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100847] [Medline: 34179332]
10. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023 Aug;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
11. Omar M, Levkovich I. Exploring the efficacy and potential of large language models for depression: a systematic review. J Affect Disord 2025 Feb 15;371:234-244. [doi: 10.1016/j.jad.2024.11.052] [Medline: 39581383]
12. Heston TF. Safety of large language models in addressing depression. Cureus 2023 Dec;15(12):e50729 [FREE Full text] [doi: 10.7759/cureus.50729] [Medline: 38111813]
13. Levkovich I, Elyoseph Z. Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: vignette study. JMIR Ment Health 2023 Sep 20;10:e51232 [FREE Full text] [doi: 10.2196/51232] [Medline: 37728984]
14. Hua Y, Liu F, Yang K, Li Z, Sheu Y, Zhou P, et al. Large language models in mental health care: a scoping review. arXiv Preprint posted online on January 1, 2024. [doi: 10.48550/arXiv.2401.02984]
15. Neimeyer RA, Bonnelle K. The Suicide Intervention Response Inventory: a revision and validation. Death Stud 1997;21(1):59-81. [doi: 10.1080/074811897202137] [Medline: 10169714]
16. Shannonhouse L, Lin YD, Shaw K, Porter M. Suicide intervention training for K–12 schools: a quasi‐experimental study on ASIST. Jour of Counseling & Develop 2017 Jan 04;95(1):3-13. [doi: 10.1002/jcad.12112]
17. Shannonhouse L, Lin YD, Shaw K, Wanna R, Porter M. Suicide intervention training for college staff: program evaluation and intervention skill measurement. J Am Coll Health 2017 Oct;65(7):450-456. [doi: 10.1080/07448481.2017.1341893] [Medline: 28628372]
18. Fujisawa D, Suzuki Y, Kato TA, Hashimoto N, Sato R, Aoyama-Uehara K, et al. Suicide intervention skills among Japanese medical residents. Acad Psychiatry 2013 Nov;37(6):402-407. [doi: 10.1007/BF03340078] [Medline: 23703377]
19. Mackelprang JL, Karle J, Reihl KM, Cash REG. Suicide intervention skills: graduate training and exposure to suicide among psychology trainees. Train Educ Prof Psychol 2014 May;8(2):136-142 [FREE Full text] [doi: 10.1037/tep0000050] [Medline: 25436034]

XSL•FO
RenderX

20. Morriss R, Gask L, Battersby L, Francheschini A, Robson M. Teaching front-line health and voluntary workers to assess and manage suicidal patients. J Affect Disord 1999;52(1-3):77-83. [doi: 10.1016/s0165-0327(98)00080-9] [Medline: 10357020]
21. Kawashima Y, Yonemoto N, Kawanishi C, Otsuka K, Mimura M, Otaka Y, et al. Two-day assertive-case-management educational program for medical personnel to prevent suicide attempts: a multicenter pre-post observational study. Psychiatry Clin Neurosci 2020 Jun 07;74(6):362-370 [FREE Full text] [doi: 10.1111/pcn.12999] [Medline: 32166827]
22. Palmieri G, Forghieri M, Ferrari S, Pingani L, Coppola P, Colombini N, et al. Suicide intervention skills in health professionals: a multidisciplinary comparison. Arch Suicide Res 2008;12(3):232-237. [doi: 10.1080/13811110802101047] [Medline: 18576204]
23. Scheerder G, Reynders A, Andriessen K, Van Audenhove C. Suicide intervention skills and related factors in community and health professionals. Suicide Life Threat Behav 2010 Apr;40(2):115-124. [doi: 10.1521/suli.2010.40.2.115] [Medline: 20465346]
24. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. arXiv Preprint posted online on January 10, 2023. [doi: 10.48550/arXiv.2201.11903]
25. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024 Jan;6(1):e12-e22 [FREE Full text] [doi: 10.1016/S2589-7500(23)00225-X] [Medline: 38123252]
26. Mehandru N, Miao BY, Almaraz ER, Sushil M, Butte AJ, Alaa A. Evaluating large language models as agents in the clinic. NPJ Digit Med 2024 Apr 03;7(1):84 [FREE Full text] [doi: 10.1038/s41746-024-01083-y] [Medline: 38570554]
27. Williams CYK, Zack T, Miao BY, Sushil M, Wang M, Kornblith AE, et al. Use of a large language model to assess clinical acuity of adults in the emergency department. JAMA Netw Open 2024 May 01;7(5):e248895 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.8895] [Medline: 38713466]
28. Granello D, Granello P. Suicide: An Essential Guide for Helping Professionals and Educators. Boston, MA: Allyn & Bacon; 2007.
29. Obradovich N, Khalsa SS, Khan WU, Suh J, Perlis RH, Ajilore O, et al. Opportunities and risks of large language models in psychiatry. NPP Digit Psychiatry Neurosci 2024 May 24;2(1):1-16. [doi: 10.1038/s44277-024-00010-z] [Medline: 39554888]
30. Chen B, Zhang Z, Langrené N, Zhu S. Unleashing the potential of prompt engineering in large language models: a comprehensive review. arXiv Preprint posted online on October 23, 2023. [doi: 10.48550/arXiv.2310.14735]
31. Askell A, Bai Y, Chen A, Drain D, Ganguli D, Henighan T, et al. A general language assistant as a laboratory for alignment. arXix Preprint posted online on December 1, 2021. [doi: 10.48550/arXiv.2112.00861]

Abbreviations

LLM: large language model
SIRI-2: Suicidal Ideation Response Inventory
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by T de Azevedo Cardoso; submitted 23.10.24; peer-reviewed by A AL-Asadi, LP Gorrepati; comments to author 29.11.24; revised version received 07.12.24; accepted 22.01.25; published 05.03.25.

Please cite as:
McBain RK, Cantor JH, Zhang LA, Baker O, Zhang F, Halbisen A, Kofner A, Breslau J, Stein B, Mehrotra A, Yu H
Competency of Large Language Models in Evaluating Appropriate Responses to Suicidal Ideation: Comparative Study
J Med Internet Res 2025;27:e67891
URL: https://www.jmir.org/2025/1/e67891
doi:10.2196/67891
PMID:40053817

©Ryan K McBain, Jonathan H Cantor, Li Ang Zhang, Olesya Baker, Fang Zhang, Alyssa Halbisen, Aaron Kofner, Joshua Breslau, Bradley Stein, Ateev Mehrotra, Hao Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Information Scanning and Seeking in Diverse Language, Cultural and Technological Media Among Latinx Adolescents: Cross-Sectional Study

Melissa J DuPont-Reyes[1], MPH, PhD; Alice P Villatoro[2], PhD; Lu Tang[3], PhD

[1]Departments of Sociomedical Sciences and Epidemiology, Columbia University Irving Medical Center, New York, NY, United States

[2]Department of Public Health, Santa Clara University, Santa Clara, CA, United States

[3]Department of Communication & Journalism, Texas A&M University, College Station, TX, United States

Corresponding Author:
Melissa J DuPont-Reyes, MPH, PhD
Departments of Sociomedical Sciences and Epidemiology
Columbia University Irving Medical Center
722 West 168th Street
Room 942
New York, NY, 10032
United States
Phone: 1 212 305 0120
Email: md3027@cumc.columbia.edu

Abstract

Background: Continuous scientific and policy debate regarding the potential harm and/or benefit of media and social media on adolescent health has resulted, in part, from a deficiency in robust scientific evidence. Even with a lack of scientific consensus, public attitudes, and sweeping social media prohibitions have swiftly ensued. A focus on the diversity of adolescents around the world and their diverse use of language, culture, and social media is absent from these discussions.

Objective: This study aims to guide communication policy and practice, including those addressing access to social media by adolescent populations. This study assesses physical and mental health information scanning and seeking behaviors across diverse language, cultural, and technological media and social media among Latinx adolescent residents in the United States. This study also explores how Latinx adolescents with mental health concerns use media and social media for support.

Methods: In 2021, a cross-sectional survey was conducted among 701 US-based Latinx adolescents aged 13-20 years to assess their health-related media use. Assessments ascertained the frequency of media use and mental and physical health information scanning and seeking across various media technologies (eg, TV, podcasts, and social media) and language and cultural types (ie, Spanish, Latinx-tailored English, and general English). Linear regression models were used to estimate adjusted predicted means of mental and physical health information scanning and seeking across diverse language and cultural media types, net personal and family factors, in the full sample and by subsamples of mental health symptoms (moderate-high vs none-mild).

Results: Among Latinx adolescents, media and social media use was similar across mental health symptoms. However, Latinx adolescents with moderate-high versus none-mild symptoms more often scanned general English media and social media for mental health information (P<.05), although not for physical health information. Also, Latinx adolescents with moderate-high versus none-mild symptoms more often sought mental health information on Latinx-tailored and general English media, and social media (P<.05); a similar pattern was found for physical health information seeking. In addition, Latinx adolescents with moderate-high versus none-mild symptoms often sought help from family and friends for mental and physical health problems and health care providers for mental health only (P<.05).

Conclusions: While media and social media usage was similar across mental health, Latinx adolescents with moderate-high symptoms more often encountered mental health content in general English media and social media and turned to general English- and Latinx-tailored media and social media more often for their health concerns. Together these study findings suggest more prevalent and available mental health content in general English versus Spanish language and Latinx-tailored media and underscore the importance of providing accessible, quality health information across diverse language, cultural, and technological media and social networks as a viable opportunity to help improve adolescent health.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e64672) doi:10.2196/64672

KEYWORDS

adolescent behaviors; mental health; Latino; social media; adolescent; media use; internet use; health information seeking; health information scanning; mobile phone

Introduction

Current reports depict an adolescent mental health crisis in the United States. About 1 in 5 adolescents experience poor mental health, especially race and ethnic minoritized adolescents [1-6]. In 2021, the US Surgeon General stated that the COVID-19 pandemic further altered adolescents' experiences at home, school, and the community, which generated devastating effects on their mental health [7]. In 2021, 42% of adolescents had persistent sadness and hopelessness and 22% of adolescents considered suicide, representing significant increases since 2019 [8]. Latinx (46%) adolescents, representing 25% of all adolescents nationally, report higher rates of sadness than Asian (35%), Black (39%), and White (41%) adolescents, with the pandemic especially worsening mental health for Latinx adolescents [9]. Despite these trends, less than half of Latinx adolescents with mental health problems receive care [5].

Given the concern about adolescent mental health, emerging public attitudes and policies have focused on how social media in particular may influence mental health outcomes and knowledge, norms, and behaviors related to mental health help-seeking and services [10-21]. While past research has focused on legacy media (eg, TV, radio, print), recent research has examined new digital media, especially social media, as a behavioral risk factor for poor mental health [22]. While the mechanisms explaining this association remain unclear, studies have shown how mental health content on social media is highly variable, portraying both positive (ie, supportive) and negative (ie, stigmatizing) content [17-21,23-28]. For instance, while #MentalHealth is one of the 3 most popular social media tags about health to increase public awareness, pejorative content such as stigma and hate speech is prevalent too [16-21,29,30]. Therefore, legacy and new digital media can deliver mixed mental health content that can potentially influence mental health knowledge, attitudes, behaviors, and outcomes. Thus, further study about adolescents' exposure to mental health information across diverse languages, cultures, and technological media is warranted, and an assessment of how this exposure might influence mental health behaviors. This knowledge is especially relevant for adolescents as about half of social media users are 10-29 years old—the same age of onset of most mental health problems [23].

Research on legacy media has shown direct links with positive and negative mental health behaviors, while new studies on social media use and mental health among adolescents have only identified small or null associations, partly due to the predominant use of limited assessments of time spent on social media as a simple construct [11,12,19,22,31-33]. This is a substantial limitation of previous studies as social media is not experienced unidimensionally. Rather, diverse linguistic, cultural, and technological social media exists—that is, media across global languages (Spanish, Mandarin, etc); global cultures (eg, Latinx-tailored, Korean pop culture, influencer cult followings, and social movements such as #BlackLivesMatter); and technologies for delivery of communications (social media apps such as YouTube, Facebook, Instagram, Snapchat, TikTok, Sidechat, Pinterest, Reddit; mobile apps; console-, computer-, or VR-based gaming; music, podcasts, film, and TV streaming services such as Netflix, Hulu, Spotify, Pandora, etc) [22]. Thus, previous studies have not considered more robust measurements or algorithm-driven content that is characteristic of social media.

Careful consideration of measurement is important as adolescent populations are diverse, as is their access to language and cultural media across technology. For instance, Latinx populations in the United States use social media more than other races and ethnic groups, including the top 3 popular apps among all adolescents (ie, Instagram, Snapchat, TikTok) [23]. In addition, Latinx adolescents have access to media, including social media, in Spanish and from Latin America due to close global social networks from dynamic migration experiences. At the same time, Latinx communities are impacted by poor-quality, slow dissemination of health communication as industry-led and government-controlled content moderation has historically overlooked Spanish-language and Latinx communications, particularly social media [13-15,34-41].

Because research on media and social media and health has largely focused on English-language media and excluded Latinx representation relative to their share of the demographic of media consumers, new research is warranted that focuses on diverse language, cultural, and technological media among Latinx adolescents and its relationship to health [11,42-45]. To address this gap, the study aimed to generate new knowledge about the frequency of mental and physical health information scanning and seeking across diverse language, cultural, and technological media and test whether these patterns vary by mental health profiles among Latinx adolescents. Because Latinx adolescents' social media use has been inadequately studied, we importantly use newly validated, culturally appropriate measures to capture diverse language, cultural, and technological media for health information among Latinx adolescents. Our measurement extends previous, traditional measures of media exposure among Latinx populations that focused on language competency and legacy media to now newly capture widely available Latinx-tailored digital media. We also compare these media-based outlets to help-seeking for health problems to family and friends and health care providers.

Methods

Sample

In 2021, 701 eligible US-based Latinx adolescents aged 13-20 years from a proprietary Qualtrics survey research panel self-completed a 20-minute cross-sectional web survey using personal digital devices about behaviors related to Spanish- and

XSL•FO
RenderX

English-language media use and mental and physical health information acquisition; complete details about study procedures are reported elsewhere [46]. Eligibility included participants identifying as "Latina/o/x/e." Recruitment quotas by age, ethnic origin, and gender were used to help sample under-represented Latinx groups in population health research. In 2021, smartphone or tablet ownership among US-based adolescents was nearly universal (more than 95%) [47]. Random checks of the survey reviewed accuracy and consistency in data entry. All study materials were offered in Spanish and English; 96.58% (677/701) of the study sample completed the survey in English.

The achieved study sample of Latinx adolescents (N=701) had a median age of 17 (SD 1.57) years; 78.21% (524/701) were female and 86.51% (603/701) were US-born (Table 1). Among the 13.49% (94/701) that were foreign-born, the average age of arrival to the United States was about 10 years old. Ethnic subgroups included Mexican (403/701, 57.49%), Cuban or Dominican or Other Latinx group (83/701, 11.84%), Puerto Rican (74/701, 10.56%), Central American (86/701, 12.27%), and South American (55/701, 7.85%), and race subgroups included American Indian or Alaskan Native (21/701, 3.01%), Black (41/701, 5.88%), Multiracial (74/701, 10.62%), White (244/701, 35.01%), and Other (317/701, 45.48%). About 41% (283/701) reported household incomes of less than US $20,000, while 46% (317/701) reported between US $20,000 and less than US $75,000; 14% (97/701) reported more than US $75,000. Participants reported high Latinx ethnic attachment (mean 33.19, SD 6.86) and English-language media use (mean 3.32, SD 0.94), followed by Latinx-tailored (mean 2.81, SD 1.03) and Spanish-language (mean 2.64, SD 1.05) media use. The study sample compared to national data is reported elsewhere [46]. Subsamples based on the Patient Health Questionnaire-4 (PHQ-4; see Measures section) included no-mild (n=391) and moderate-severe (n=310) mental health symptoms. Subsamples varied significantly by gender, race, parent or guardian nativity, and Spanish acculturation (P<.05), and not by other covariates in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of Latinx youth full sample and subsamples of mental health symptoms (N=701), 2021.

Characteristics	Full sample (N=701)	No-mild mental health symptoms subsample (n=391)	Moderate-severe mental health symptoms subsample (n=310)
Age, mean (SD)	17 (1.57)	16.94 (1.55)	17.08 (1.6)
Sex, n (%)[a]			
Female	524 (78.21)	267 (71.39)	257 (86.82)
Male	146 (21.79)	107 (28.61)	39 (13.18)
Ethnicity, n (%)			
Central American	86 (12.27)	53 (13.55)	33 (10.65)
Cuban, Dominican, or Other	83 (11.84)	42 (10.74)	41 (13.23)
Mexican	403 (57.49)	229 (58.57)	174 (56.13)
Puerto Rican	74 (10.56)	36 (9.21)	38 (12.26)
South American	55 (7.85)	31 (7.93)	24 (7.74)
Race, n (%)[a]			
Black or African American	41 (5.88)	19 (4.91)	22 (7.1)
American Indian or Alaskan Native	21 (3.01)	14 (3.62)	7 (2.26)
Multiracial	74 (10.62)	31 (8.01)	43 (13.87)
Other	317 (45.48)	197 (50.9)	120 (38.71)
White	244 (35.01)	126 (32.56)	118 (38.06)
Household income, n (%)			
US $0-US $19,999	283 (40.6)	152 (39.28)	131 (42.26)
US $20,000-US $34,999	134 (19.23)	72 (18.6)	62 (20)
US $35,000-US $74,999	183 (26.26)	103 (26.61)	80 (25.81)
More than US $75,000	97 (13.92)	60 (15.5)	37 (11.94)
Nativity			
Foreign-born, n (%)	94 (13.49)	55 (14.21)	39 (12.58)
US-born, n (%)	603 (86.51)	332 (85.79)	271 (87.42)
Age at arrival, mean (SD)	9.85 (5.44)	10.04 (5.55)	9.59 (5.35)
Years lived in the United States, mean (SD)	7.58 (5.35)	7.35 (5.53)	7.9 (5.14)
Parent or guardian			
Parent or guardian 1 born in the United States, n (%)[a]	305 (45.12)	155 (41.55)	150 (49.5)
Parent or guardian 2 born in the United States, n (%)	261 (39.55)	134 (36.71)	127 (43.05)
Latinx ethnic attachment, mean (SD)	33.19 (6.86)	33.21 (6.77)	33.16 (6.98)
Spanish acculturation, mean (SD)[a]	25.75 (7.49)	26.26 (7.41)	25.09 (7.55)
English acculturation, mean (SD)	33.07 (3.21)	32.97 (3.38)	33.2 (3)
Spanish media use, mean (SD)	2.64 (1.05)	2.66 (1.04)	2.62 (1.02)
Latinx media use, mean (SD)	2.81 (1.03)	2.84 (1.06)	2.78 (1)
English media use, mean (SD)	3.32 (0.94)	3.3 (0.95)	3.33 (0.93)

[a]$P<.05$ for significant differences across mental health symptoms.

Measures

Media Use (Covariate)

As previously validated, [46] 7 items assessed diverse language, cultural, and technological media types, including Spanish, Latinx-tailored English, and general English media use across print media, free broadcast TV, subscription TV, Twitter, other social media (eg, Facebook, Instagram, and TikTok), radio or podcasts, and music and music streaming. Participants were asked how often they engage in each language, cultural, and

XSL•FO
RenderX

technological media type on an average day on a 5-point Likert scale (1=never to 5=always). A previously reported factor analysis [46] revealed that total scores of all items combined for each language or cultural dimension to create three subscales: Spanish-language (α=.92), Latino-tailored English (α=.90), and general English (α=.86) media use.

Media-Based Health Information Scanning (Outcome 1)

In total, 5 items assessed the frequency of mental health and, separately, physical health, information scanning (ie, passive exposure to health information in media) across the 3 language or cultural domains (ie, Spanish, Latinx-tailored English, and general English) for each technological media type: print media, TV, websites, social media, or radio or podcasts [46]. Participants were asked how often they encountered (ie, scanning) mental health and, separately, physical health, information in each language, cultural, and technological media type during an average week (1=not at all to 4=a few times a week). Factor analysis [46] revealed that total scores of all 5 media items for each language or cultural domain combined to create 3 subscales of all media scanning for mental health and, separately, for physical health (α=.92-.94).

Furthermore, given recent policy actions related to social media among adolescents, we also examined the 2 items pertaining to social media for each language or cultural domain. Combining the 2 social media items for each language or cultural domain resulted in 3 subscales of social media scanning for mental health and physical health separately.

Media-Based Health Information Seeking (Outcome 2)

A total of 5 items assessed the frequency of mental and physical health information seeking (ie, searching on purpose for health information in media) across the 3 language or cultural domains for each technological media type (print media, TV, websites, social media, and radio or podcasts) [46]. Participants were asked about the frequency that they purposely sought mental health and, separately, physical health, information in each media type for a mental or physical health problem of one's own, a family member, or a friend (1=none to 4=always). Based on the factor analysis that has been previously reported [46], total scores of all items for each language or cultural domain were combined to create 3 subscales of all media information seeking for mental health and, separately, for physical health (α=.87-.90). Also, 2 other configurations of these items by technological media type emerged in factor analyses: a legacy media subscale comprised of items assessing TV, radio, and print media (α=.93; legacy media seeking) and an internet-based media subscale comprised of items assessing websites and social media (α=.89-.90; internet and social media seeking) [46].

Formal Health Care and Informal Networks Health Information Seeking (Outcome 3)

A total of 4 items were assessed seeking information for mental health, and separately, a physical health, problem of their own or a family member or a friend, from either health care professionals (formal) or family and friends (eg, in-person or call or text; informal): "For each of the following sources, how often have you actively looked for information about a (physical or mental) health problem that you, a family member, or a friend had?" Participants responded on a 4-point Likert scale (1=never to 4=always) for formal and informal care to the item about mental health followed by a similar item about physical health.

Sociodemographic and Mental Health Factors (Other Covariates)

Age, sex, ethnic origin (ie, Latinx country of ancestry), race, nativity (ie, US- or foreign-born), household income, and migration factors (eg, age at arrival to the United States) were measured (Table 1). A total of 6 items assessed ethnic attachment (α=.90), defined as the extent of exploration, belonging, and commitment to one's Latinx ethnic identity; and 18 items measured acculturation, which captured language use and proficiency (α=.90-.91). Finally, the PHQ-4 screened for mental health symptoms, and demonstrated excellent reliability (α=.89). In total 4 items are scored on a 4-point Likert scale (0=not at all to 3=nearly every day) and summed to 0-12 full scale range. A total of 2 subsamples were created using established cutoff points: scores 0-5=none-mild and scores 6-12=moderate-severe.

Statistical Analysis

Linear regression models estimated adjusted predicted means of mental and physical health information scanning and seeking across diverse language, cultural, and technological media types (ie, all media inclusive of legacy and internet-based and social media in the Spanish language, Latinx-tailored English language, and general English language). All models were adjusted for age, sex, nativity, race, ethnic origin, household income, ethnic attachment, acculturation, and Spanish, Latinx-tailored English, and general English media use. Models inclusive of the full sample controlled for mental health symptoms. Multicollinearity was not observed between the covariates; all correlations between variables were less than 0.5. Because bivariate tests showed no statistically significant differences between all media use variables across mental health, all media use variables were entered into subsequent models as control variables. Model fit chi-square statistics indicated good fit (P<.05). Adjusted predicted means from linear regression models were also analyzed across subsamples of symptoms (no-mild vs moderate-severe). Figures plotting the adjusted predicted means for all outcomes are presented in the main results (see Tables S1-S4 in Multimedia Appendix 1 for complete linear regression model results). All results were considered significant at P<.05 and obtained using StataSEVs.18.

Ethical Considerations

The study was approved by the institutional review board at Texas A&M University. All study participants provided informed consent and assent and received a US $5 gift card for their time and effort to complete the study. All study data are anonymous and deidentified.

Results

Media-Based Health Information Scanning

Table 2 displays the adjusted predicted means of all outcomes, including media-based mental and physical health information

XSL•FO
RenderX

scanning across diverse language, cultural, and technological media types, including legacy versus internet-based and social media. In the full sample of Latinx adolescents and subsamples of mental health symptoms, mental and physical health information scanning occurred most often in general English-language media, followed by Latinx-tailored and Spanish-language media. Similar patterns were found for social media.

Figure 1 displays the adjusted predicted means of mental and physical health information scanning in the full sample of Latinx adolescents and subsamples of mental health symptoms. No statistically significant differences in mental and physical health information scanning were observed for Latinx-tailored or Spanish media and social media outcomes across subsamples of symptoms. However, Latinx adolescents with moderate-severe symptoms obtained mental health information through scanning English media and social media significantly more often than those with no-mild symptoms ($P<.05$; Figure 1). In contrast, no significant differences in physical health information scanning outcomes were found across mental health (Figure 2).

XSL•FO
RenderX

Table 2. Adjusted predicted means from simple linear regression models of mental health and physical health scanning and seeking outcomes in the Latinx youth full sample and subsamples of mental health symptoms (N=701), 2021.

Outcomes	Full sample, margin (95% CI)	No-mild mental health symptoms subsample, margin (95% CI)	Moderate-severe mental health symptoms subsample, margin (95% CI)
Mental health scanning			
All media			
Spanish	1.76 (1.71-1.81)	1.76 (1.7-1.83)	1.75 (1.67-1.82)
Latinx	1.79 (1.74-1.84)	1.78 (1.71-1.85)	1.8 (1.72-1.88)
English[a]	2.18 (2.12-2.24)	2.11 (2.03-2.19)	2.26 (2.17-2.36)
Social media			
Spanish	1.85 (1.79-1.92)	1.85 (1.76-1.94)	1.86 (1.76-1.95)
Latinx	1.95 (1.88-2.02)	1.94 (1.85-2.04)	1.96 (1.85-2.06)
English[a]	2.35 (2.28-2.43)	2.26 (2.15-2.37)	2.47 (2.35-2.59)
Physical health scanning			
All media			
Spanish	1.85 (1.79-1.9)	1.84 (1.77-1.91)	1.85 (1.77-1.93)
Latinx	1.94 (1.89-2)	1.92 (1.85-2)	1.97 (1.88-2.05)
English	2.11 (2.05-2.17)	2.09 (2-2.17)	2.14 (2.05-2.23)
Social media			
Spanish	1.9 (1.84-1.96)	1.92 (1.83-2)	1.88 (1.78-1.98)
Latinx	2.05 (1.98-2.11)	2.05 (1.95-2.14)	2.04 (1.94-2.15)
English	2.2 (2.13-2.28)	2.2 (2.09-2.3)	2.21 (2.09-2.33)
Mental health seeking			
All media			
Spanish	1.7 (1.65-1.75)	1.68 (1.62-1.75)	1.73 (1.65-1.8)
Latinx[a]	1.92 (1.87-1.96)	1.87 (1.81-1.93)	1.98 (1.91-2.05)
English[a]	2.01 (1.96-2.06)	1.94 (1.87-2.01)	2.1 (2.02-2.18)
Media type			
TV, book, radio	1.68 (1.64-1.72)	1.65 (1.59-1.71)	1.72 (1.65-1.78)
Internet, social media[a]	2.17 (2.12-2.22)	2.09 (2.03-2.16)	2.27 (2.19-2.34)
Informal and formal			
Family, friend[a]	2.45 (2.37-2.52)	2.29 (2.19-2.39)	2.64 (2.53-2.75)
Health care professional[a]	1.77 (1.7-1.83)	1.69 (1.6-1.79)	1.86 (1.76-1.96)
Physical health seeking			
All media			
Spanish	1.8 (1.75-1.85)	1.77 (1.7-1.84)	1.83 (1.76-1.91)
Latinx[a]	1.86 (1.81-1.91)	1.81 (1.74-1.88)	1.93 (1.85-2.01)
English	1.92 (1.87-1.97)	1.87 (1.8-1.94)	1.97 (1.89-2.05)
Media type			
TV, book, radio	1.69 (1.65-1.74)	1.67 (1.61-1.74)	1.72 (1.65-1.79)
Internet, social media[a]	2.1 (2.05-2.16)	2.03 (1.96-2.11)	2.19 (2.11-2.28)
Informal and formal			
Family, friend[a]	2.46 (2.39-2.53)	2.36 (2.26-2.46)	2.59 (2.48-2.7)

XSL•FO
RenderX

Outcomes	Full sample, margin (95% CI)	No-mild mental health symptoms subsample, margin (95% CI)	Moderate-severe mental health symptoms subsample, margin (95% CI)
Health care professional	1.86 (1.8-1.93)	1.82 (1.72-1.91)	1.92 (1.81-2.03)

[a]*P*<.05 for significant differences across subsamples of mental health symptoms.

[b]All models adjusted for age, sex, nativity, ethnic origin, race, household income, ethnic attachment, acculturation, and Spanish, Latinx, and English media use; models inclusive of the full sample also adjusted for mental health symptoms while models focused on the subsamples of mental health symptoms did not.

Figure 1. Adjusted predicted means of mental health content scanning in all media in the Latinx youth full sample and subsamples by mental health symptoms (N=701), 2021.

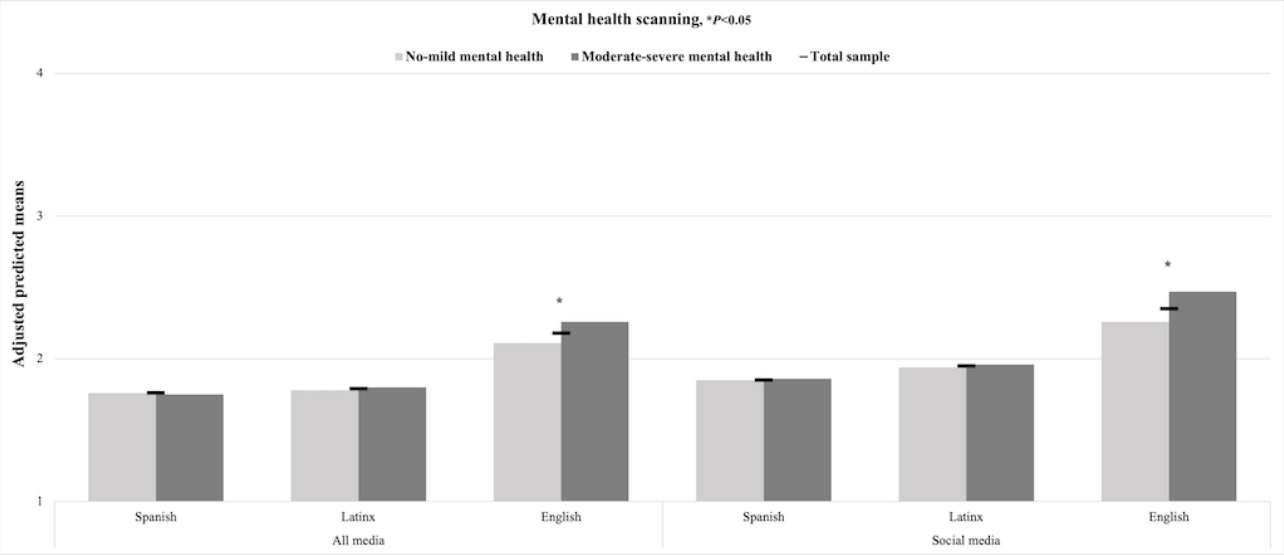

Figure 2. Adjusted predicted means of physical health content scanning in all media in the Latinx youth full sample and subsamples by mental health symptoms (N=701), 2021.

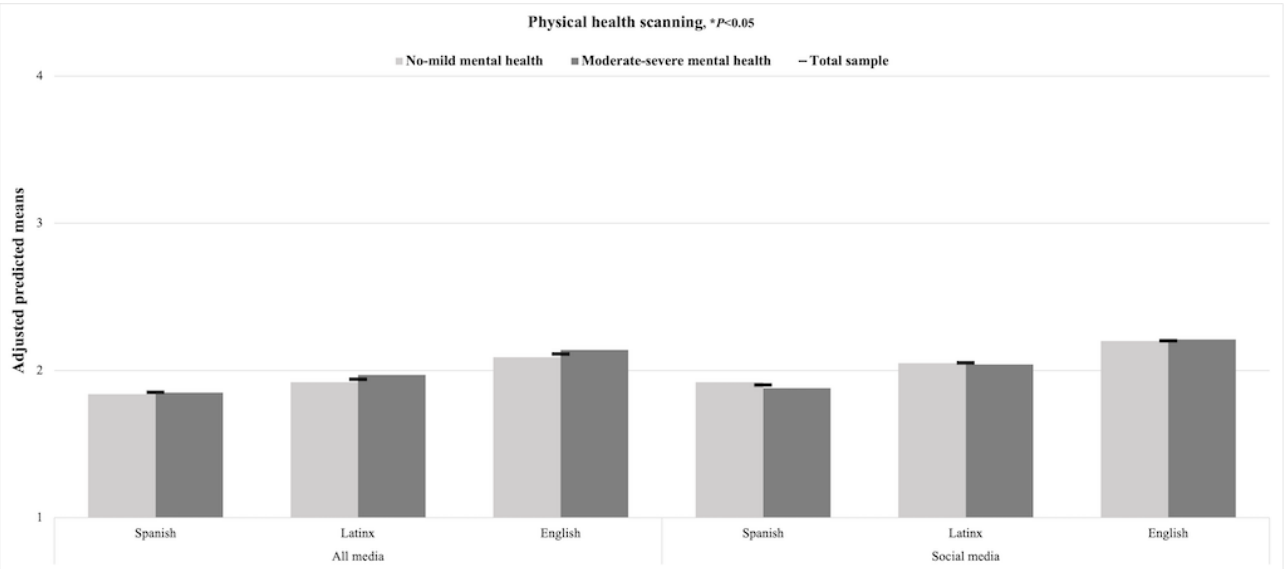

Media-Based Health Information Seeking

Table 2 also displays the adjusted predicted means of the media-based mental and physical health information-seeking outcomes across diverse language, cultural, and technological media types, including legacy versus internet-based and social media. In the full sample of Latinx adolescents and subsamples of mental health symptoms, mental and physical health information seeking in all media types occurred most often in general English-language media, followed by Latinx-tailored English-language media and Spanish-language media.

In terms of media type, the full sample of Latinx adolescents and both mental health subsamples reported seeking mental and physical health information most often in internet-based and social media versus legacy media.

Figure 2 displays the adjusted predicted means of mental and physical health information seeking in the full sample of Latinx

XSL•FO
RenderX

adolescents and mental health subsamples. No significant differences in mental and physical health information seeking were found for Spanish-language media or legacy media across the mental health subsamples. However, Latinx adolescents with moderate-severe symptoms sought mental health information in Latinx-tailored and nontailored English-language media and internet-based and social media more frequently than those with no-mild symptoms (P<.05; Figure 3). Similarly, Latinx adolescents with moderate-severe versus no-mild symptoms sought physical health information in Latinx-tailored English-language media and internet-based and social media more frequently (P<.05; Figure 4).

Formal and Informal Help-Seeking

Table 2 displays the adjusted predicted means of formal (ie, health care providers) and informal (ie, family or friends) help-seeking for mental and physical health problems in the full sample of Latinx adolescents and mental health subsamples. Both the full sample and mental health subsamples sought help for mental and physical health problems more often from family or friends than health care providers.

Figure 2 displays the adjusted predicted means for formal and informal help-seeking for a mental and physical health problem in the full sample of Latinx adolescents and mental health subsamples. Adolescents with moderate-severe versus no-mild symptoms were more likely to seek help from family or friends for both physical or mental health problems (P<.05; Figures 3 and 4) and from health care providers for mental health problems only (P<.05; Figure 3). Overall, Latinx youth sought help for health problems from family or friends and internet-based and social media more often than from health care providers (Figures 3 and 4).

Figure 3. Adjusted predicted means of mental health help-seeking in all media, formal, and informal sectors in the Latinx youth full sample and subsamples by mental health symptoms (N=701), 2021.

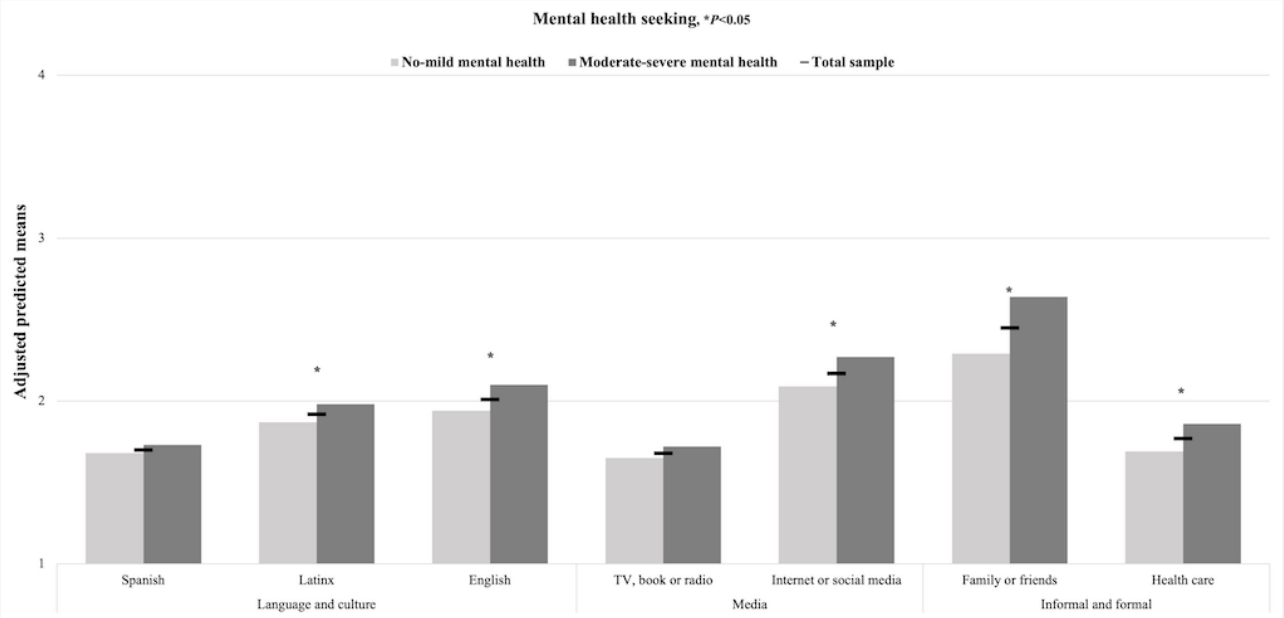

XSL•FO
RenderX

Figure 4. Adjusted predicted means of physical health help-seeking in all media, formal, and informal sectors in the Latinx youth full sample and subsamples by mental health symptoms (N=701), 2021.

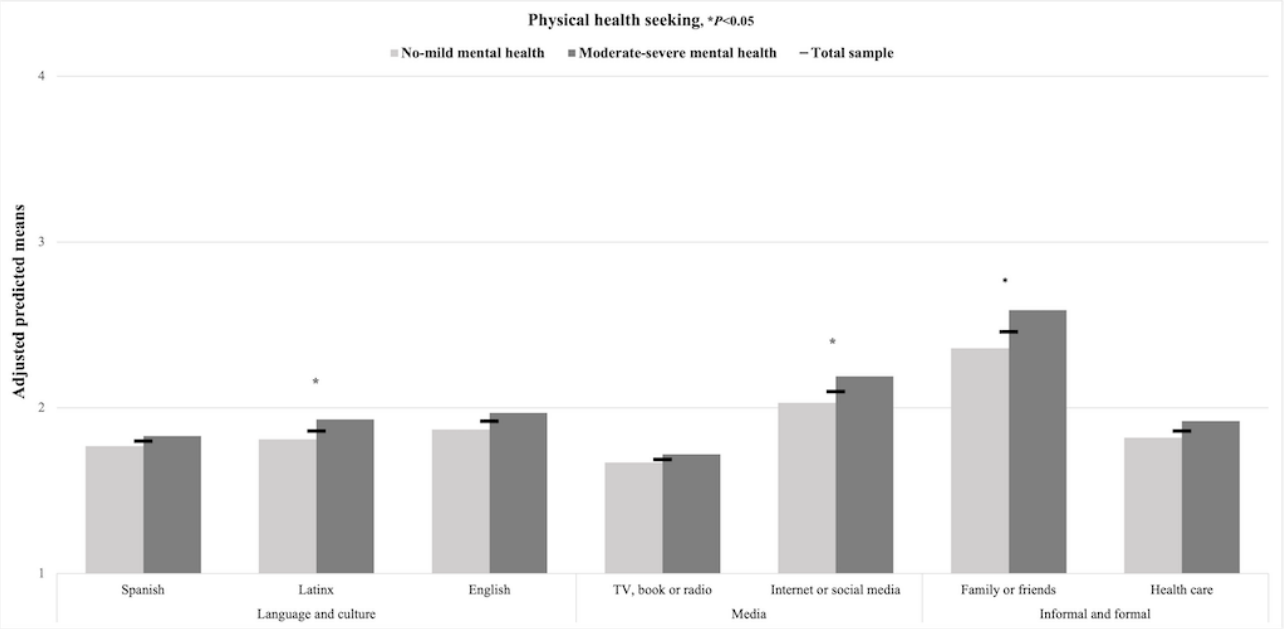

Discussion

While data trends depict a national adolescent mental health crisis with simultaneous increases in social media use, our study aimed to provide a more nuanced understanding of these trends across diverse language, cultural, and technological media among Latinx youth. Our study distinguished between mental and physical health, examined diverse language and cultural media domains, distinguished passive and active media exposure (eg, scan vs seek), and accomplished these assessments for both legacy and new digital media.

In terms of passive, casual browsing, overall, our study found that Latinx adolescents with mental health issues were more likely to engage in mental health scanning in English-language legacy and social media, but not in Spanish-language or Latinx-tailored legacy and social media, or for physical health. These patterns may arise because of more prevalent mental health content available in English-language than in Spanish-language media and social media, as previous studies confirm [48]. This is an important finding because as Latinx adolescents with mental health concerns engage with media and social media, they may become more exposed to mental health content, underscoring the need for public mental health action that leverages media to provide quality resources about symptom management and help-seeking. In contrast, passive exposure to mental or physical health content in Spanish-language and Latinx-tailored media did not vary by mental health symptoms possibly because health information is less often available in Spanish and Latinx-media, and not because of a lack of preference or interest in these media among Latinx media consumers [49]. For Latinx families with Spanish and Latinx media preferences, disparate diffusion of health communications actually may help produce health inequities [48].

With respect to the active, purposeful search for health information, our study found significant use of Latinx-tailored and English-language media for mental health information seeking and Latinx-tailored media for physical health information seeking among Latinx adolescents with moderate-severe symptoms. Latinx adolescents with mental health concerns may seek mental health information that is tailored to them and presented in their preferred language and a culturally aligned context. Public health interventionists may take note of this preference, and the availability of tailored, targeted mental and physical health information in media for Latinx adolescent audiences. Finally, family, friends, internet, and social media were more often used for both mental and physical health information than health care professionals, especially among Latinx adolescents with moderate-severe symptoms, indicating a greater preference for more easily accessible informal supports for health problems within the Latinx community. Future studies could examine the role of mistrust of health care experts and providers as sources of health information among Latinx communities to better understand these patterns.

Our study points to Spanish-language and Latinx-tailored media as important sources of health information, with English-language media as the most common source owing to more health information being available there. For example, recent studies of primetime advertising on YouTube TV revealed that English-language TV was more likely to include health promotion advertisements while Spanish-language TV was more likely to include health-adverse advertisements [50]. Inequities in health-related advertising on social media are likely similar though this remains a knowledge gap in the literature. Importantly, our study found that Latinx media is an important source of mental health information for those experiencing mental illness; thus, Latinx media may serve as an important resource for learning about and seeking support for mental health conditions. To this point, media content with greater immigrant community representation and storytelling with respect to mental

XSL•FO
RenderX

and physical health issues would be important to foster in future media programming and content curation [38,51].

Findings from this growing body of literature must be implemented into public health policy and practice as Spanish is a global power language and the most common non-English language spoken in the United States. The United States also has the second-largest population of Spanish speakers worldwide behind Mexico. Finally, Latinx populations tend to use social media more than all other major social groups in the United States despite structural inequities in content moderation of Latinx media [23].

This study has some limitations. First, as Latinx populations are dynamic, especially with respect to media and linguistic preferences and behaviors, longitudinal versus cross-sectional assessments could improve causal inference among variables and help assess changes to media content over time such as media campaigns for health promotion. Second, related dimensions of trust and comprehension of health information and perceived health concerns would be important to ascertain. For example, since mental health content is inequitable across language or cultural media, it may influence perceived mental health among English-language media consumers depending on trust and comprehension levels in the content [46]. Third, the study relied on self-reported participant recall of their behaviors. However, recall bias is likely minimal and nondifferential across the subgroups by mental health as the measures asked highly relevant information about oneself that affects daily life. Relatedly, our survey-based research is unable to assess the role of social media algorithms, which influence user consumption. Finally, the study sample was recruited from a proprietary, survey panel; thus, details regarding responses among its volunteers are undisclosed which could potentially lead to bias. To address this concern, we found that the achieved study sample is robust to national data on Latinx populations while also oversampling understudied Latinx groups, as previously reported [46].

In summary, our study demonstrates a more diverse portrait of adolescent media use in the United States, specifically Latinx adolescents who have diasporic, global ties to diverse language, cultural, and technological media. While policies and public attitudes about media and social media tend to generalize these issues, our study suggests potential contexts in which diverse language, cultural, and technological media can be an important, beneficial source of health information for Latinx adolescents. Current scrutiny about the effects of social media use on adolescent health has not sufficiently approached this issue through a health equity lens, which is essential to help inform equitable policies, programs, and practices related to its safe use among adolescents. Diverse adolescent populations will have unique contexts, needs, and assets with respect to the use of diverse language, cultural, and technological media. For Latinx adolescents experiencing mental illness, social media seems to be a useful resource of support, which has implications for other minoritized populations who might find social media as a tool to prevent health risks. Thus, greater access and equity with respect to digital media remain critical to health.

Acknowledgments

Research support was provided by the Robert Wood Johnson Foundation.

Data Availability

The dataset generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables with coefficients and 95% confidence intervals B [95% CI] from linear regression models assessing media outcomes in the Latinx youth full sample; N=701, 2021.
[PDF File (Adobe PDF File), 316 KB - jmir_v27i1e64672_app1.pdf]

References

1. Lindsey MA, Sheftall AH, Xiao Y, Joe S. Trends of suicidal behaviors among high school students in the United States: 1991-2017. Pediatrics 2019;144(5):e20191187 [FREE Full text] [doi: 10.1542/peds.2019-1187] [Medline: 31611338]
2. Keyes KM, Gary D, O'Malley PM, Hamilton A, Schulenberg J. Recent increases in depressive symptoms among US adolescents: trends from 1991 to 2018. Soc Psychiatry Psychiatr Epidemiol 2019;54(8):987-996 [FREE Full text] [doi: 10.1007/s00127-019-01697-8] [Medline: 30929042]
3. Mojtabai R, Olfson M. National trends in mental health care for US adolescents. JAMA Psychiatry 2020;77(7):703-714 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.0279] [Medline: 32211824]
4. Stephenson J. Children and teens struggling with mental health during COVID-19 pandemic. JAMA Health Forum 2021;2(6):e211701 [FREE Full text] [doi: 10.1001/jamahealthforum.2021.1701] [Medline: 36218745]

XSL•FO
RenderX

5. Goodwin RD, Dierker LC, Wu M, Galea S, Hoven CW, Weinberger AH. Trends in U.S. depression prevalence from 2015 to 2020: the widening treatment gap. Am J Prev Med 2022;63(5):726-733 [FREE Full text] [doi: 10.1016/j.amepre.2022.05.014] [Medline: 36272761]
6. Fan Q, DuPont-Reyes MJ, Hossain MM, Chen LS, Lueck J, Ma P. Racial and ethnic differences in major depressive episode, severe role impairment, and mental health service utilization in U.S. adolescents. J Affect Disord 2022;306:190-199. [doi: 10.1016/j.jad.2022.03.015] [Medline: 35301042]
7. Office of the Surgeon General (OSG). Protecting Youth Mental Health: The U.S. Surgeon General's Advisory Internet. Washington (DC): US Department of Health and Human Services; 2021.
8. Centers for Disease Control and Prevention (CDC). Youth Risk Behavior Survey Data Summary and Trends Report 2013-2023. Washington, D.C: US Department of Health and Human Services; 2024.
9. McGinty EE, Presskreischer R, Anderson KE, Han H, Barry CL. Psychological distress and COVID-19-related stressors reported in a longitudinal cohort of US adults in April and July 2020. JAMA 2020;324(24):2555-2557 [FREE Full text] [doi: 10.1001/jama.2020.21231] [Medline: 33226420]
10. Phelan JC, Link BG. Controlling disease and creating disparities: a fundamental cause perspective. J Gerontol B Psychol Sci Soc Sci 2005;60 Spec No 2:27-33. [doi: 10.1093/geronb/60.special_issue_2.s27] [Medline: 16251587]
11. Ross AM, Morgan AJ, Jorm AF, Reavley NJ. A systematic review of the impact of media reports of severe mental illness on stigma and discrimination, and interventions that aim to mitigate any adverse impact. Soc Psychiatry Psychiatr Epidemiol 2019;54(1):11-31. [doi: 10.1007/s00127-018-1608-9] [Medline: 30349962]
12. Seale C. Media and Health. London, United Kingdom: Sage Publications; 2003.
13. Clayman ML, Manganello JA, Viswanath K, Hesse BW, Arora NK. Providing health messages to Hispanics/Latinos: understanding the importance of language, trust in health information sources, and media use. J Health Commun 2010;15 Suppl 3(Suppl 3):252-263 [FREE Full text] [doi: 10.1080/10810730.2010.522697] [Medline: 21154097]
14. Manganello JA, Gerstner G, Pergolino K, Graham Y, Strogatz D. Media and technology use among hispanics/latinos in New York: implications for health communication programs. J Racial Ethn Health Disparities 2016;3(3):508-517. [doi: 10.1007/s40615-015-0169-9] [Medline: 26715218]
15. Bastick Z, Mallet-Garcia M. Double lockdown: the effects of digital exclusion on undocumented immigrants during the COVID-19 pandemic. New Media & Society 2022;24(2):365-383. [doi: 10.1177/14614448211063185]
16. Anderson CA, Shibuya A, Ihori N, Swing EL, Bushman BJ, Sakamoto A, et al. Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review. Psychol Bull 2010;136(2):151-173. [doi: 10.1037/a0018251] [Medline: 20192553]
17. Craig W, Boniel-Nissim M, King N, Walsh SD, Boer M, Donnelly PD, et al. Social media use and cyber-bullying: a cross-national analysis of young people in 42 countries. J Adolesc Health 2020;66(6S):S100-S108 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.03.006] [Medline: 32446603]
18. Berryman C, Ferguson CJ, Negy C. Social media use and mental health among young adults. Psychiatr Q 2018;89(2):307-314. [doi: 10.1007/s11126-017-9535-6] [Medline: 29090428]
19. Sadagheyani HE, Tatari F. Investigating the role of social media on mental health. Ment Health Soc Incl 2020;25(1):41-51. [doi: 10.1108/mhsi-06-2020-0039]
20. Wilksch SM, O'Shea A, Ho P, Byrne S, Wade TD. The relationship between social media use and disordered eating in young adolescents. Int J Eat Disord 2020;53(1):96-106. [doi: 10.1002/eat.23198] [Medline: 31797420]
21. Plaisime M, Robertson-James C, Mejia L, Núñez A, Wolf J, Reels S. Social media and teens: a needs assessment exploring the potential role of social media in promoting health. Soc. Media Soc 2020;6(1):1-11. [doi: 10.1177/2056305119886025]
22. Odgers CL, Jensen MR. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. J Child Psychol Psychiatry 2020;61(3):336-348 [FREE Full text] [doi: 10.1111/jcpp.13190] [Medline: 31951670]
23. Auxier B, Anderson M. Social media use in 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2022-12-14]
24. van den Eijnden RJJM, Lemmens JS, Valkenburg PM. The social media disorder scale. Comput Human Behav 2016;61(3):478-487. [doi: 10.1016/j.chb.2016.03.038]
25. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. J Med Internet Res 2021;23(9):e30409 [FREE Full text] [doi: 10.2196/30409] [Medline: 34468327]
26. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
27. Zheng DX, Ning AY, Levoska MA, Xiang L, Wong C, Scott JF. Acne and social media: a cross-sectional study of content quality on TikTok. Pediatr Dermatol 2021;38(1):336-338. [doi: 10.1111/pde.14471] [Medline: 33247602]
28. Vargas L, Comello MLG, Porter JH. The Web's potential to provide depression literacy resources to latinx teens: a missed opportunity? Howard J Commun 2020;32(4):366-381. [doi: 10.1080/10646175.2020.1819480]
29. Best #mentalhealth hashtags. Best-Hashtags.com. 2022. URL: https://best-hashtags.com/hashtag/mentalhealth/ [accessed 2024-02-11]

XSL•FO
RenderX

30. Hashtag search #MentalHealth and #MentalHealthAwareness. TikTok. 2022. URL: https://tiktokhashtags.com/hashtag/mentalhealthawareness/ [accessed 2025-02-11]
31. Becker AE. Television, disordered eating, and young women in Fiji: negotiating body image and identity during rapid social change. Cult Med Psychiatry 2004;28(4):533-559. [doi: 10.1007/s11013-004-1067-5] [Medline: 15847053]
32. WAHL OF. Depictions of mental illnesses in children's media. J Ment Health 2009;12(3):249-258. [doi: 10.1080/0963823031000118230]
33. Donohue D, Swords L. The portrayal of mental illness in popular children's programs on Netflix: a content and thematic analysis. Psychol Pop Media 2024;13(1):102-110. [doi: 10.1037/ppm0000445]
34. Feldman JM, Bassett MT. Variation in COVID-19 mortality in the US by race and ethnicity and educational attainment. JAMA Netw Open 2021;4(11):e2135967 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.35967] [Medline: 34812846]
35. Rios MD, Giachello A, Khosla S, Luna G, Pobee R, Vergara-Rodriguez P, et al. Illinois unidos: a community demands equity, justice, and inclusion. Am J Public Health 2021;111(S3):S204-S207. [doi: 10.2105/AJPH.2021.306407] [Medline: 34709861]
36. Maleki P, Al Mudaris M, Oo KK, Dawson-Hahn E. Training contact tracers for populations with limited english proficiency during the COVID-19 pandemic. Am J Public Health 2021;111(1):20-24. [doi: 10.2105/AJPH.2020.306029] [Medline: 33326265]
37. Olayo-Méndez A, Vidal De Haymes M, García M, Cornelius LJ. Essential, disposable, and excluded: the experience of latino immigrant workers in the US during COVID-19. J Poverty 2021;25(7):612-628. [doi: 10.1080/10875549.2021.1985034]
38. Reny T, Manzano S. The negative effects of mass media stereotypes of Latinosimmigrants. In: Media and Minorities: Questions on Representation from an International Perspective. Göttingen: Vandenhoeck & Ruprecht; 2016:195-212.
39. Thompson DA, Flores G, Ebel BE, Christakis DA. Comida en venta: after-school advertising on Spanish-language television in the United States. J Pediatr 2008;152(4):576-581. [doi: 10.1016/j.jpeds.2007.09.011] [Medline: 18346518]
40. Alegria M, Takeuchi D, Canino G, Duan N, Shrout P, Meng XL, et al. Considering context, place and culture: the National Latino and Asian American Study. Int J Methods Psychiatr Res 2004;13(4):208-220 [FREE Full text] [doi: 10.1002/mpr.178] [Medline: 15719529]
41. McDevitt M, Butler M. Latino youth as information leaders: implications for family interaction and civic engagement in immigrant communities. InterActions: UCLA Journal of Education and Information Studies 2011;7(2).
42. Wahl OF. Children's views of mental illness: a review of the literature. Psychiatr Rehabil Ski 2002;6(2):134-158. [doi: 10.1080/10973430208408430]
43. Maiorano A, Lasalvia A, Sampogna G, Pocai B, Ruggeri M, Henderson C. Reducing stigma in media professionals: is there room for improvement? results from a systematic review. Can J Psychiatry 2017;62(10):702-715 [FREE Full text] [doi: 10.1177/0706743717711172] [Medline: 28622747]
44. Gwarjanski AR, Parrott S. Schizophrenia in the news: the role of news frames in shaping online reader dialogue about mental illness. Health Commun 2018;33(8):954-961. [doi: 10.1080/10410236.2017.1323320] [Medline: 28537757]
45. McGinty EE, Webster DW, Barry CL. Effects of news media messages about mass shootings on attitudes toward persons with serious mental illness and public support for gun control policies. Am J Psychiatry 2013;170(5):494-501. [doi: 10.1176/appi.ajp.2013.13010014] [Medline: 23511486]
46. DuPont-Reyes MJ, Villatoro A, Gama G, Tang L. Measuring media-related health and mental health information acquisition among Latino adults in the United States. Int J Methods Psychiatr Res 2023 Dec;32(4):e1967 [FREE Full text] [doi: 10.1002/mpr.1967] [Medline: 36978265]
47. Atske S, Perrin A. Home broadband adoption, computer ownership vary by race, ethnicity in the U.S. Pew Reseach Center. URL: https://tinyurl.com/4jjy86xt [accessed 2023-11-27]
48. DuPont-Reyes MJ, Villatoro AP, Tang L. Mechanisms of mental illness anti-stigma messaging matter: leveraging mental health communication inequities among Latinx populations to understand what works and what we can do better. Soc Sci Med 2024 May;349:116865. [doi: 10.1016/j.socscimed.2024.116865] [Medline: 38643699]
49. DuPont-Reyes MJ, Villatoro AP, Tang L. Media-based health information seeking behaviors among Latinx populations in the United States. Hisp J Behav Sci 2024;45(4) [FREE Full text]
50. DuPont-Reyes MJ, Villatoro AP, Chowkwanyun M, Ortiz SE, Tang L. Communication policy to reduce health disparities: a cross-language content analysis of YouTube TV advertising. Am J Prev Med 2024;67(4):P521-P529 [FREE Full text]
51. Menjívar C. Immigrant criminalization in law and the media: effects on Latino immigrant workers' identities in Arizona. Am Behav Sci 2016;60(5-6):597-616. [doi: 10.1177/0002764216632836]

Abbreviations

PHQ-4: Patient Health Questionnaire-4

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.07.24; peer-reviewed by L Magis-Weinberg, S Elliott, R Wagani, H Siddiq; comments to author 15.10.24; revised version received 03.12.24; accepted 09.01.25; published 05.03.25.

Please cite as:
DuPont-Reyes MJ, Villatoro AP, Tang L
Health Information Scanning and Seeking in Diverse Language, Cultural and Technological Media Among Latinx Adolescents: Cross-Sectional Study
J Med Internet Res 2025;27:e64672
URL: https://www.jmir.org/2025/1/e64672
doi:10.2196/64672
PMID:40053766

©Melissa J DuPont-Reyes, Alice P Villatoro, Lu Tang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Safety When Considering Self-Harm and Suicide-Related Content: Qualitative Focus Group Study With Young People, Policy Makers, and Social Media Industry Professionals

Louise La Sala[1,2], PhD; Amanda Sabo[1,2], BASc (Hons); Maria Michail[2,3], PhD; Pinar Thorn[1,2], MPsychClin; Michelle Lamblin[1,2], PhD; Vivienne Browne[1], BPAPM (Hons); Jo Robinson[1,2], PhD

[1]Orygen, Parkville, Australia

[2]Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

[3]Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Louise La Sala, PhD
Orygen
35 Poplar Road
Parkville, 3052
Australia
Phone: 61 3 9966 9512
Email: louise.lasala@orygen.org.au

Abstract

Background: Young people are disproportionately impacted by self-harm and suicide, and concerns exist regarding the role of social media and exposure to unsafe content. Governments and social media companies have taken various approaches to address online safety for young people when it comes to self-harm and suicide; however, little is known about whether key stakeholders believe current approaches are fit-for-purpose.

Objective: From the perspective of young people, policy makers and professionals who work within the social media industry, this study aimed to explore (1) the perceived challenges and views regarding young people communicating on social media about self-harm and suicide, and (2) what more social media companies and governments could be doing to address these issues and keep young people safe online.

Methods: This qualitative study involved 6 focus groups with Australian young people aged 12-25 years (n=7), Australian policy makers (n=14), and professionals from the global social media industry (n=7). Framework analysis was used to summarize and chart the data for each stakeholder group.

Results: In total, 3 primary themes and six subthemes are presented: (1) challenges and concerns, including the reasons for, and challenges related to, online communication about self-harm and suicide as well as reasoning with a deterministic narrative of harm; (2) roles and responsibilities regarding online safety and suicide prevention, including who is responsible and where responsibility starts and stops, as well as the need for better collaborations; and (3) future approaches and potential solutions, acknowledging the limitations of current safety tools and policies, and calling for innovation and new ideas.

Conclusions: Our findings highlight tensions surrounding roles and responsibilities in ensuring youth online safety and offer perspectives on how social media companies can support young people discussing self-harm and suicide online. They also support the importance of cross-industry collaborations and consideration of social media in future suicide prevention solutions intended to support young people.

(J Med Internet Res 2025;27:e66321) doi:10.2196/66321

KEYWORDS

young people; suicide prevention; self-harm; social media; online safety; policy

XSL•FO
RenderX

Introduction

Suicide is a leading cause of youth mortality, and in many countries, including Australia, rates appear to be increasing [1,2]. Self-harm is more common and presents a significant risk factor for suicide [3]. The reasons for self-harm and suicide are complex, with many questioning the role of social media. Concerns exist regarding the potential for certain types of online content, such as graphic depictions of self-harm or suicide methods, livestreams of suicidal behavior, or online suicide pacts or challenges to cause distress and contribute to imitative behavior [4-9]. Recent research suggests as many as 83% of young people have seen self-harm or suicide content on social media, often without searching for it, and often before the age of 14 [10,11]. Given the ways that social media algorithms operate, this may lead to young people being overexposed to potentially harmful content when interacting with their peers online [12,13].

Despite this potential for harm, young people have identified numerous benefits of using social media to communicate about self-harm and suicide. For example, it allows them to cultivate community, validate their experiences, support those they care about, and grieve people who have died by suicide [14,15]. Given the difficulties accessing high-quality, timely, and age-appropriate mental health support, young people have also identified social media as an accessible, and sometimes preferred, avenue for seeking help or information, especially for stigmatized topics such as self-harm and suicide [16]. Recognizing the central role of social media in young people's lives, recent initiatives have capitalized on the ability to deliver highly personalized and safe suicide prevention content directly to individuals' newsfeeds, including during suicide bereavement [17-21]. Such social media-based interventions present a significant change to the ways young people can seek evidence-based information and access mental health care [22-24].

The potential for exposure to harmful content has placed pressure on policy makers to keep young people safe online [25,26]. National suicide prevention strategies play an important role in shaping a comprehensive and systematic approach to suicide prevention; they also help determine research priorities and the allocation of resources [2,27]. However, online safety is largely absent in national suicide prevention strategies. Almost 40 countries (including Australia) have developed suicide prevention strategies [2], many of which identify young people as a priority population and consider the settings where they spend their time, for example, schools. While many national strategies recommend the development of guidelines for mainstream media to facilitate safe reporting about suicide [25,28,29], social media has only recently been included as an additional consideration for suicide prevention efforts [30-32].

Similarly, governments in many countries have developed rigorous online safety policies that focus on safeguarding the rights and privacy of individuals and regulating technology and social media companies (eg, The Online Safety Act in the United Kingdom [33]; The Online Safety Act in Australia [34], and The Children's Online Privacy Protection Rule [35] and Kids Online Safety Act [36] in the United States). However, only a handful of these online safety policies include guidance on suicide prevention specifically [33,37]. In the absence of self-harm and suicide-specific guidance, many popular social media platforms have developed their own policies relating to self-harm- and suicide-related content [38-41]. While it is not clear if these policies are evidence-based, they differ across each platform, and social media companies are legislated and regulated differently in different countries. To this end, it appears that national suicide prevention strategies and online safety policies are yet to appropriately consider how self-harm and suicide are expressed, and managed, on social media, leaving a key gap in youth online safety policy and practice.

The objective of this study was to inform health, communication, and online safety policy as it relates to self-harm and suicide prevention. From the perspective of young people, policy makers and professionals who work within the social media industry, the aims of this study were to understand (1) relevant stakeholder views regarding young people communicating on social media about self-harm and suicide, and (2) what more, if anything, social media companies and governments could be doing to address these issues and keep young people safe online.

Methods

Study Design

This qualitative study involved focus groups with Australian young people, policy makers, and professionals from global social media companies. The study was conducted by researchers based in Melbourne, New South Wales, Australia, and has been reported in line with the Consolidated Criteria for Reporting Qualitative Research (COREQ; Multimedia Appendix 1) [42].

Research Team and Reflexivity

Members of the research team who have been trained in qualitative research methods facilitated the focus groups (JR and LLS) and undertook data coding and analysis (LLS, JR, MM, and PT). All members of the research team identified as female and were employed in suicide prevention research (JR, PhD; LLS, PhD; AS, Hons; PT, MPsych [Clin]; MM, PhD; and ML, PhD) or youth mental health organizations (VB).

Focus group facilitators were known to some of the participants prior to their involvement in this study, based on their known roles and responsibilities within their respective organizations or prior involvement in other activities conducted by the research team. Participants were notified of the purpose of this study and how the data would also be used to inform the second edition of the #chatsafe guidelines, developed by the research team [21,43]. The facilitators made efforts to discuss the topics broadly and reduce any perceived pressure to provide desirable feedback related to the #chatsafe program of work by monitoring tone, verbal and body language, and engaging equally with the challenges and solutions about online communication about self-harm and suicide.

XSL•FO
RenderX

Sample and Recruitment

In total, 6 focus groups were conducted: one with young people (n=7); 2 with Australian policy makers (n=14; these included policy makers in a range of positions and departments focused on young people, online safety, education, and health); and 3 with individuals employed by social media companies (n=7; these included staff from different companies that operate globally, whose business names will not be reported to protect confidentiality). Focus groups with professionals from social media companies were restricted to include only individuals employed by the same company in order to preserve privacy, maintain confidentiality, and foster open conversation. Each of these focus groups therefore involved a smaller number of participants (n=2-3). Young people were recruited via social media advertising on the #chatsafe social media pages, and young people who had previously participated in #chatsafe activities were invited via email. Policy makers and professionals from the social media companies were invited via email by JR.

Young people were eligible to participate if they were aged between 12 and 25 years; able to speak and read English, and; if under 18, had parent or guardian consent. Policy makers and professionals from social media companies were eligible to participate if they were older than 18 years and were (1) employed by a government department or had a policy-making role with responsibility for mental health, youth, or online safety, or (2) employed in an online safety or policy team within the social media industry. There were no exclusion criteria.

Data Collection and Analysis

Focus groups were conducted between June 2022 and August 2022 by JR and LLS, with assistance from VB. In total, 3 were conducted online via Zoom, and 3 were conducted in a hybrid format (ie, some participants were present in-person, and others joined via Zoom). Focus groups ran for 60-120 minutes. A semi-structured topic guide was used (see Multimedia Appendix 2) and included questions about participants' views regarding the challenges associated with young people using social media to communicate about self-harm and suicide and what more, if anything, social media platforms and policy makers could be doing to keep people safe online. Participants were also asked questions about Edition 2 of the #chatsafe guidelines, which have been reported elsewhere [21].

All focus groups were audio recorded and transcribed verbatim. Field notes were also taken. Data were analyzed using framework analysis, a structured process of theme-based analysis through the development of charts [44,45]. In conducting this study, a goal of the research team was to inform health, communication, and online safety policy as it relates to online safety and suicide prevention. As such, framework analysis was selected as the most appropriate methodology to systematically reduce the data whilst representing each stakeholder group as a "case" (eg, young people, social media professionals, and policy makers). A combined deductive and inductive coding approach allowed the research team to examine challenges and opportunities expected to emerge based on the previous literature and their own experience (eg, difficulties determining safe versus unsafe content [14] and the impact of exposure to graphic content [15]), whilst leaving scope for participants to assign their own meaning to the issues being explored.

Following transcription, six steps were followed: (1) familiarization, (2) coding, (3) development of an analytic framework, (4) applying the framework, (5) charting the data, and (6) interpreting the data [44,46]. Initially, 4 members of the research team (LLS, JR, MM, and PT) independently coded the same transcript line by line, highlighting interesting segments of the text and making notes. They met to discuss their approach and alignment on codes. In total, 3 members of the research team (LLS, JR, and MM) then coded 2 more transcripts, again meeting to discuss coded sections and interpretation of data. Where there were disagreements, the team revisited the transcript and discussed it as a group. The team agreed on a set of codes that formed an initial analytic framework. This included codes related to previously documented "harms" and "benefits" that were present in the data, as well as new codes that reflected sentiments raised by participants regarding solutions, challenges, and what more different stakeholders could do to keep young people safe online. LLS coded the remaining transcripts, meeting with MM and JR regularly, and iteratively revised the thematic framework used to interpret the full data. Transcripts were coded by hand, and codes, categories, and representative quotes from each case were charted in an Excel spreadsheet by LLS. The matrix was reviewed and discussed by LLS, JR, MM, and AS, and connections between and within the categories were thematically mapped as they addressed the research questions.

Ethical Considerations

This study received ethics approval from the University of Melbourne Human Research and Ethics Committee (ID: 22728). Participants were required to complete an online consent form and demographic survey prior to attending their focus group. All data were de-identified for analysis. Youth participants were reimbursed Aus $30 (US $19.02) per hour for their involvement in this study. Policy makers and professionals working within the social media industry received no compensation.

Results

Participants

Participant demographics are summarized in Table 1, including information about the top 3 social media platforms used by participants in each group. Additional information about professional participants' current employment was also collected, though some details have been omitted to avoid possible reidentification. All but 3 participants across the total sample reported using social media platforms in their daily life, with most using 3 or more platforms (n=20, 80%).

XSL•FO
RenderX

Table 1. Participant demographics and top social media platforms used by policy makers, social media industry professionals, and young people.

Group (n)	Age (years) mean (range)	Sex or gender (n)	Country of birth (n)	Identifying as Aboriginal (n)	Top social media platforms used (n)
Policy makers (14)	39.31[a] (23-54)	• Female or woman (10) • Male or man (3)[a]	• Australia (12) • Other (2)[b]	• Yes (1) • No (13)	• Facebook (10) • YouTube (8) • Instagram (7)
Social media companies (7)	39.14 (27-57)	• Female or woman (4) • Male or man (2) • Non-binary (1)	• Australia (1) • Other (6)[b]	• No (7)	• YouTube (7) • Facebook (5) • Instagram (5)
Young people (7)	20.71 (19-23)	• Female or woman (5) • Male or man (2)	• Australia (3) • Other (4)[b]	• Yes (1) • No (6)	• Instagram (7) • Facebook (6) • YouTube (6)

[a]One policy maker participant did not disclose their age or sex or gender.

[b]Other countries of birth included Hong Kong, Malaysia, Mexico, Singapore, Taiwan, the United Kingdom, and the United States.

The resulting 3 primary themes and 6 subthemes are presented below. The analytic framework used to interpret these data, with quotes from each stakeholder group, is provided in Multimedia Appendix 3.

Challenges and Concerns

To address the first research question, participants from all 3 stakeholder groups were asked about their views regarding young people using social media to communicate about self-harm and suicide. While all stakeholder groups were encouraged to share what they felt the challenges were and the reasons for young people using social media for this purpose, each group mostly arrived at the same conclusion: there are concerns about young people engaging with, and being exposed to, self-harm and suicide-related content on social media; however, social media is not going away and may meet the needs of young people looking for a safe place to express their distress. Perspectives of each group were also largely influenced by recent public discourse relating to social media and youth mental health more generally, with groups differing in their views regarding the helpfulness of this debate.

Reasons for, and Challenges Related to, Young People Using Social Media to Communicate About Self-Harm and Suicide

Participants in all focus groups expressed the belief that individuals used social media to seek and provide support or share and understand personal experiences related to self-harm and suicide. Some youth participants described social media as a place to "perform" their mental health and felt that representations of self-harm and suicide on social media were heavily impacted by what their peers were posting and the feedback that they sought. Across all focus groups, there was an appreciation for the necessity of using social media to have these conversations, as participants recognized that they are difficult topics to discuss offline, and for many people, help is not available elsewhere. There was consensus across each focus group about the affordances of social media (eg, anonymity and temporality) that make these topics easier to talk about online versus face-to-face.

> *It's a lot easier to just put a post out there for the world to see… if people are feeling that way, they probably feel quite like they might be a burden and they don't want to go up to their friends and say, 'I'm feeling like this', or they don't want to go up to a mental health practitioner and tell them. So it's easier just to put the post out into the public and see who replies back.* [Young person]

> *...social media can provide…a really accessible space for young people. I mean accessible in terms of not just for young people who live in areas where there might not be access to services. But for young people who maybe don't feel safe going to a service… it's… 24:7, stigma-free often, they can connect with people who are in similar situations.* [Policy maker]

Although individuals in all participant groups understood why and how these conversations occurred on social media, each group articulated their concerns for the harms associated with online self-harm– and suicide-related content. Concerns related to the age of the user and what is developmentally appropriate to view; the virality of self-harm and suicide content and specific online challenges; the potential for social media to sensationalize, glorify, or normalize self-harm and suicide; users being exposed to harmful content, livestreams, or real-time suicidal behavior; the potential for contagion; and other online experiences that have implications for self-harm and suicide (eg, bullying, sextortion, etc).

> *I remember this one instance where this girl posted on her private account - and I wasn't that close with her so I really shouldn't have been there; but a photo of a razor blade...It was framed in a certain way, and she captured it like, 'hello, my old friend', or something. Then two days later, another girl in another circle posted essentially the same photo… she'd definitely seen it, so there was the element of contagion.* [Young person]

A similar theme across each focus group was uncertainty around the nuances of what constituted harm, for whom and when, and how best to respond to online harms. Social media industry professionals described this challenge as not knowing "where to draw the line" between helpful versus harmful content, and policy makers discussed the heterogeneity in user experiences and outcomes. Young people expressed their belief that more

XSL•FO
RenderX

needed to be done to prevent individuals from seeing harmful content in the first place, describing teenagers as not developmentally mature enough to understand the content that they were being exposed to regularly, based on their own experiences.

Participants from the policy maker and social media industry focus groups reflected on the complexity associated with balancing freedom of online expression with the immediate removal of harmful content. Enforcing these decisions across global corporations and different legislative environments, particularly in the absence of a strong evidence base, was a challenge frequently discussed by those developing and implementing safety policies.

> *It has harmful stuff on it but can provide such an avenue for help seeking as well. How do you find that balance?* [Policy maker]

> *That component of freedom of expression where we also don't want to talk over the voices of people who have historically been marginalized, especially by scientific communities or by lack of access to resources or things like that. So for us… you want to enable safe communication and also how do you do that in a way that is grounded in both research evidence and global equity?* [Social media industry professional]

Reasoning With a Deterministic Narrative of Harm

Irrespective of their personal views, participants in all stakeholder groups were aware of public perceptions that social media causes harm to mental health more broadly and had different perspectives about how helpful or harmful those perceptions were, both to individuals using social media and to those developing and regulating them. In reflecting on why people believe social media is negatively impacting youth mental health, some youth participants acknowledged that social media is often perceived as not being proactive in terms of creating safe online environments but commented that there were many other factors that impacted youth mental health beyond social media.

Perspectives amongst policy makers were more varied. One participant described their firm belief that “being on social media is [a] detriment to people’s mental health.” However, the majority of policy makers instead wondered how they could better “capitalis[e] on what social media can provide” in order to support individual users while minimizing potential harms.

Participants from the social media industry expressed feeling like a “punching bag for online safety” while reflecting on there being “no research to support this conclusion [of harm]." Some social media professionals believed that this public perception stalled more sophisticated conversations about the future potential and innovation of social media from a mental health perspective.

> *This growing narrative that there's a causal link between social media and self-harm…really, there's no research to support this conclusion, but I think it really hinders our abilities sometimes to put out a different narrative that can feel supportive for people, and that can help drive behavior change, that can drive people towards our platforms to utilize some of the resources that we use.* [Social media industry professional]

Roles and Responsibilities

A central question across all focus groups was the issue of responsibility for online safety and where the lines of responsibility started and ended for each stakeholder group. All focus groups featured discussions about how challenging these topics were, especially when online safety and themes related to self-harm and suicide overlapped. Due to this complexity, participants in all groups appeared to articulate more questions than answers. This highlighted areas where each stakeholder group felt others could be doing more and where better collaborations and models of partnership were required.

Who Is Responsible and Where Does Responsibility Start and Stop?

Participants raised many rhetorical questions related to the roles and responsibilities of their industry or position. For example, participants from the social media industry acknowledged that they play a fundamental role in defining socially acceptable behavior online (and offline) through the implementation of their community guidelines or standards. One participant from the social media industry stated, “we're not a medical company and we're not a healthcare company” and questioned how appropriate it was that they were the people to ultimately determine what is safe versus unsafe behavior or content for different global regions and across different cultural settings.

> *From a platform perspective, we’re obviously not doing clinical work, but we’re still judging what type of content is considered to be more problematic or less problematic from a conversational perspective in terms of the risk for the person, when we need to report it to different cases, what types of resources we need to provide. So, the question becomes, like, does this distinction, for instance, between passive and active suicidal ideation, does that mirror in languages that don’t have the same language structure as English or expression?* [Social media industry professional]

Some described their need to develop new policies “on-the-go” to keep up with the advancement of technology and user behavior, yet they often felt that there was little empirical evidence guiding decision-making and their need to move more quickly than research can. Given their reach and popularity, one participant commented that “any one direction we move the needle, even just microscopically, is going to have an impact on a huge swath of users.” Participants from this industry also expressed a desire for “more regulation” but felt that, in some cases, regulation restricted their ability to “use their proactive technologies… to save and have a considerable impact in people's lives.” To navigate these tensions, participants from the social media industry highlighted the importance of meaningful investment from policy makers or governments regarding online safety.

XSL•FO
RenderX

> *… we can't keep people safe, we can only try to protect them from harm. We can't define safe as an absolute state of being online, offline, or anywhere else. So, what we have to do is, we have to do our best to mitigate risk and their exposure to risk.* [Social media industry professional]

> *Just as a policymaker on platform, there's really a lack of evidence around a lot of the policy questions that we're struggling with in terms of where to draw the line, especially really solid research evidence and evidence that comes from contents and cultures outside of what is typically thought of as like the WEIRD western world.* [Social media industry professional]

Policy makers described how the regulatory landscape influences the content that should or should not be shared online but also spoke openly about the challenges they face when trying to regulate and govern international corporations and online content that is geographically boundless. While acknowledging the limitations of government regulation, policy makers reflected on the importance of cross-government approaches and their responsibility to ensure that they are supporting individuals to stay safe online by "providing information and support and guidance in a variety of settings, in a variety of forms."

> *So we are interested in health, there's the Department of Education who has an interest here, there's the eSafety Commissioner, the Department of Comms. So, there's lots of different government agencies that are interested, but there's also the sites themselves. I think there's a lot of different players and sometimes that can perhaps make things even more complex than we would otherwise like.* [Policy maker]

Lastly, it was understood that individual users play a critical role in creating safe online environments and need to have agency over their interactions. However, all groups acknowledged that younger users could not take sole responsibility for their own safety within environments created by, and often for, adults.

> *I don't like to put the safety of individuals, like, all of the responsibility for the safety of individuals, on the individuals because I think that suicide and self-harm in particular has been seen as this really individual problem that's disconnected from everything that happens in the community…* [Social media industry professional]

In the absence of clear roles and responsibilities, and given the complexity of these issues, there was a tendency for participants to describe their own responsibility in relation to what more other stakeholder groups need to be doing. For instance, some participants questioned the legitimacy of government investment in online safety, and others questioned what more parents or carers, and school curricula, could do. Despite a broad recognition that collaborations were needed, participants sometimes conveyed an "us versus them" mentality, either between companies within the same industry (eg, one social media platform versus another) or between stakeholder groups (eg, the social media industry versus national governments; social media platforms versus parents).

> *Politicians would much rather help parents say it's someone else's fault, yours, the industries, than actually say, actually you also need to step up and do something here to promote online safety.* [Social media industry professional]

Young people felt strongly that the government could help by providing more education and digital literacy training, particularly when they were younger and still attending school. Structured education on online safety was perceived as a way of alleviating pressure on young people to be responsible for safe online environments in ways that they can't "handle" yet, while also providing them with knowledge that would be protective.

> *I think if the government has a role to play at all, it would be through education…like those special assemblies… Then, if you have young people who have been educated in this way, they can self-police - young people are really good at that, hopefully in ways that are helpful as opposed to, you're bad for posting this, this is what we can do instead. I feel like education - all of the stories that we've shared about…seeking help and not really having the resources or tools or vocabulary, and so if you find a way to educate young people…in a more structured way, then I feel that would really alleviate a huge part of the problem.* [Young person]

The Need for Better Collaborations

Participants from all groups highlighted the need for a collaborative approach across sectors to achieve meaningful change in ensuring online safety. There was a sense that there was "a shared responsibility for this issue" and "a lot of different players," but "different levels of investments" in online safety. It was also acknowledged that these tensions are felt in the absence of "better models of partnership" or collaboration, though no concrete examples were presented for how this might be achieved.

> *…[it's] going to have to include a multi-pronged approach. That includes some level of, depending on age, parental, familial, government, industry; all of these different factors I think need to come together to really, I think, uniformly protect people. In the absence of that, I think it's too easy to go one way or the other. To try to put this all on young people, to try to put this on social media or the government, it leaves out so many different other parts of this puzzle that when together, you're best placed to build that front to protect users.* [Social media industry professional]

Promisingly, there was a desire from all participants to work together, in acknowledgment that the issues are "big, really tough challenges that we all face," and that issues related to "data use transparency, consent, and funding" need to be addressed collectively. There was also recognition from social media industry participants of the importance of social media

XSL•FO
RenderX

companies continuing to partner with "subject-matter experts" so that their safety actions and tools were evidence-informed.

Future Approaches and Potential Solutions

Overview

Participants in all focus groups were asked to reflect on current and future applications of social media as a possible tool in youth self-harm and suicide prevention. Here, participants shared their perspectives and knowledge of the use of current safety tools and policies embedded within social media platforms, highlighting issues with current safety features. Participants expressed ideas for the future of online safety and suicide prevention using new technologies (eg, artificial intelligence [AI]) to detect and respond to risk, while also acknowledging issues related to user privacy and the lack of transparency regarding current uses of recommender systems.

Acknowledging the Limitations of Current Safety Tools and Policies

Young people and policy makers in this study were aware that some platform safety features currently exist, but there appeared to be little faith in their ability to create safer online experiences. This either came about from not trusting the platforms to act accordingly (eg, previous experiences of using a reporting tool resulting in no action and therefore discouraging a user from doing that again) or users not knowing the full range of safety tools and features that exist or how to use them. While there was an appreciation that there is no "one-size-fits-all" approach to online safety and that not all tools and functions were going to suit each user, some believed that "having some improved transparency over those reporting processes" would foster greater trust.

> *Personally, I don't have much faith in social media's reporting features. There have been times – like you see some really atrocious things online… and you're like, oh my gosh, I really – I can't believe I encountered that…You report it and then five days later it's like, it's been reviewed and it's been decided there was no issue here.* [Young person]

> *… sometimes things are so awful you want to report them but then there's not a specific thing that you can allocate it to … it's just awful.* [Young person]

Scope for Innovation and New Ideas

When asked what more the broader social media industry could be doing to promote online safety and make their platforms safer for younger users, particularly youth participants and policy makers spoke about their ideas for how AI and platform algorithms, or recommender systems, could be better optimized. This included using AI tools to detect risk, respond in real-time to risk or distress, and deliver services or support.

> *They could be doing more...but more proactively in terms of, it seems like this post has content discussing suicide, click here for more information, and there's the chatsafe guidelines, or an iteration. It would definitely boost how social media is regarded, which is as like a necessary villain.* [Young person]

On the other hand, participants from the social media industry spoke about the limits of current technologies, describing that "humans are really good at certain things and algorithms are really good at others." They felt that there were misunderstandings about how algorithms work and the extent to which AI can be used ethically while also maintaining user privacy. In some instances, current regulatory frameworks prohibit the use of risk detection technologies, but better guidance and support from policy makers was perceived to allow these technologies to advance in potentially helpful ways.

> *When we're talking about AI capabilities and algorithms in particular, what I've found is that the people who suggest those are typically, in my experience, people who don't have a lot of experience building algorithms…* [Social media industry professional]

> *I think as a society we are rightfully pretty skeptical of AI and algorithms when it comes to health concerns, in particular, and so without laws changing pretty significantly in a lot of countries, I think that there's a lot of hard limitations and blocks to what we can do as a company without overstepping the boundaries of legal liability to provide help to users or provide targeted, for instance, help, like, support resources and things like that.* [Social media industry professional]

Despite the "technology not being there just yet," participants from the social media industry acknowledged that the reporting functions and safety tools, with refinement, could become much more user-friendly and helpful for users. Some participants were also hopeful that future innovation in online safety could involve the integration of evidence-based interventions that are promoted or delivered in-platform.

> *I think that my dream would be to have better help tooling available online for folks, so really having things like having a safety plan available or having a single session intervention that is culturally responsive and appropriate available for users to be able to opt into doing … that's evidence-based. I think that there [are] steps that the industry is taking towards getting there, I think that it will be interesting to see what happens in this space within the next five to 10 years.* [Social media industry professional]

While some policy makers admitted to knowing very little about AI or its capabilities, they were strongly in favor of using proactive detection technologies to promote resources and reach individuals with real-time support or help. There was also acknowledgment that Australia is well positioned to be a world leader in this regard.

> *I definitely do think that having links and support… with the algorithm picking up [that certain] content, but also if the algorithm can pick up that kind of content, shouldn't there also be a few ways to hopefully stop that content? Rather than just leaving it up and be like, here's some links though just in case you might click on it and have a look.* [Policy maker]

XSL•FO
RenderX

Discussion

Principal Findings

This study explored views regarding young people using social media to communicate about self-harm and suicide from the perspective of young people, policy makers, and individuals who work for social media companies. It also explored participants' beliefs about what more social media companies and governments could be doing to enhance online safety for young people engaging with, and exposed to, self-harm and suicide-related content. Our findings reflect the tensions regarding roles and responsibilities for youth online safety, as well as various perspectives on the ways social media could be used to support young people when communicating online about these topics.

Our findings support a growing body of research exploring the reasons why young people use social media to share and understand their experiences of self-harm and suicide [14,15]. Young people using social media as an informal source of support and help-seeking appears to be beneficial in circumventing barriers associated with professional mental health care (eg, stigma, cost, and long wait times) [16] and allows young people to validate their own experiences [14]. However, participants in this study also echoed concerns associated with young people using social media for this purpose (eg, potential overexposure to harmful content, contagion effects, age of users, privacy, and access to individual data) [5,7,15,47], highlighting uncertainty around when and how these risks might translate to harm for individual users, who might be most at risk and how best to respond.

Contextualized within a time where social media companies are being heavily scrutinized for the actions that they are (or are not) taking with regards to youth safety and wellbeing [6,48-50], some participants in this study felt that a deterministic narrative around social media conferring harm stalled more sophisticated and nuanced conversations needed to understand and address these risks. The focus on social media playing a significant role in rising rates of youth psychological distress, despite little to no causal findings to support this claim and often in the absence of considering other social determinants that are impacting young people [51-54], has prohibited the expansion of our understanding about how social media impacts individuals differently and who is actually most at risk. This understanding is limited further when it comes to self-harm and suicide. Though some effort has been made to address potential online harms through research (eg, the #chatsafe intervention) [17,18,20], policy [34], and industry practice (eg, platform safety centers and tools) [38-41], it is clear that more work needs to be done to understand the precise nature and mechanisms of harm in this context and to develop evidence-informed solutions and policies that carefully consider the needs of young people.

A prevailing sentiment among policy makers and social media industry professionals was that they felt ill-equipped to make online safety decisions that were both timely (per the rapid nature of this industry) and evidence-based. Given that many of the decisions that platforms need to make quickly about self-harm and suicide-related content lack empirical evidence, there is an urgent need for better relationships between researchers and social media companies, including the sharing of data (whilst maintaining user privacy) and investment in the development of evidence-informed and age-appropriate safety policies and features. Although platforms regularly publish transparency reports quantifying the volume of content they have removed in enforcing their community guidelines [55-57], there is a notable lack of transparent information detailing how decisions around safety policy development and implementation are made, by whom, and based on what evidence [49]. Though we acknowledge the tensions in social media companies navigating transparent reporting on their systems and processes while protecting their intellectual property, greater attention does need to be given to how safety policies are developed, informed, and implemented, and there is a clear need for research evidence to support best practice. This aligns with calls from the Australian eSafety commissioner for "radical transparency" from social media companies regarding their online safety approaches [49], as well as recommendations for collaborative and cross-industry efforts to reduce social media harms proposed in the US Surgeon General's report on social media and youth mental health [48].

The theme of roles and responsibility was central across all focus group discussions regarding what more social media companies and policy makers could be doing to promote online safety. Critically, there was a lack of agreement or clarity on the nature and extent of responsibility that should be assigned to various stakeholder groups represented in this study, but also to stakeholders such as parents, educators, and health professionals more broadly. Though cross-industry partnerships, international coalitions [26,58], and the creation of subject-matter safety advisory boards [59,60] are all currently occurring in various ways, they are often happening in isolation from one another and differ across regions and platforms. They are further challenged by differing online safety policies and capacity for regulation globally. The inclusion of online safety in future suicide prevention policies, both at a national level and by global health organizations (eg, the World Health Organization), would go some way toward clarifying the roles and responsibilities of relevant stakeholder groups and would facilitate the sharing of information and evidence-based practice. Consideration of digital contexts in suicide prevention policies would also support meaningful investment in research and the establishment of cross-industry partnerships to promote online safety for young people.

Finally, several opportunities for innovative and novel ways of using social media to detect, respond to, and manage self-harm and suicide risk were explored. These mostly related to using social media algorithms to promote helpful resources to someone identified as at-risk, with the potential for platforms to direct individuals to professional support or deliver help within the platform itself [11,61]. However, while young people and policy makers supported these ideas, platform professionals emphasized the limitations of current technologies and the misconceptions community members hold about their capacity. Further, recommender systems and algorithms are relatively new technologies that have been delivered to a huge global user base as a means of optimizing user experience and keeping

XSL•FO
RenderX

individuals active on social media platforms. There are understandable questions about the ethics of their use and consideration for user privacy, as well as concerns regarding the adverse impacts of recommender systems in promoting harmful content [49]. There is also limited transparency around how these systems are developed, updated, or regulated and for how these issues may be addressed. However, platform algorithms present an important and unique opportunity for reaching young people with high-quality and evidence-informed mental health supports, especially as these systems become more advanced and personalized over time.

Strengths and Limitations

A strength of this study is the representation of views from young people, policy makers and professionals from the social media industry; groups that are usually working in isolation from one another and fundamentally opposing one another. To the best of our knowledge, this is the first study to include the perspectives of professionals working for social media companies in discussions about youth online safety regarding self-harm and suicide. Although focus groups with these participants were smaller in size than those with the other participant groups, the decision to restrict industry focus groups to individuals employed by the same company preserved participants' privacy and promoted open discussion within these smaller groups. Conducting a framework analysis allowed for these distinct stakeholder groups to be equally involved in this work, resulting in a more nuanced discussion of the digital landscape that young people communicate within, processes to develop and regulate those environments, and the best ways of supporting individuals in the future.

Prior knowledge of the research team by some participants was a limitation of this study. However, it did not appear that these established relationships hindered the conversations or information shared. Instead, this pre-established trust and rapport appeared to facilitate honest conversations about the challenges faced by each group. A mutual respect for the complexity and challenges associated with youth self-harm and suicide prevention also invited open conversations with a shared understanding of wanting to do more.

Finally, the average age of the youth group was approximately 21, and given that most social media platforms allow users to create accounts from as young as 13, this work would have been strengthened by the inclusion of younger participants. Further, given their prior involvement in other activities, youth participants were very mental health literate and aware of the importance of safe communication about self-harm and suicide, which may not be representative of the wider Australian youth demographic. Additional stakeholder groups, such as parents, carers, and schools, were not included in this study, though their perspectives have been reported elsewhere [62] or are being collected through work that is currently underway within our team.

Conclusions

There are valid concerns about young people using social media to discuss self-harm and suicide. However, social media remains a preferred and accessible way for young people to communicate and seek help. The evidence base for current platform strategies to manage self-harm and suicide content is lacking, raising questions about the appropriateness of social media companies defining online safety alone, as well as the challenges governments face when trying to regulate global companies and fast-spreading content. Our findings highlight the complexity of these issues and the need for shared responsibility and greater understanding of roles across relevant stakeholder groups. These findings present strong support from young people, social media companies, and Australian policy makers for cross-industry partnerships to create safer online experiences for young people and the implementation of new and emerging technologies to prevent youth self-harm and suicide.

Acknowledgments

The study received funding from the Commonwealth Department of Health under the National Suicide Prevention Leadership and Support Program. LLS is funded by a Suicide Prevention Australia Postdoctoral Fellowship. JR is funded by a National Health and Medical Research Council Investigator Grant (ID2008460) and a Dame Kate Campbell Fellowship from the University of Melbourne. MM is funded by the European Union H2020-EU.1.3.2. MSCA-IF-GF-Global Fellowships (grant agreement ID: 101026065).

Data Availability

The datasets generated or analyzed during the current study are not publicly available due to inability to sufficiently anonymize transcripts, but are available from the corresponding author on reasonable request.

Conflicts of Interest

LLS is a member of Meta's Australian online safety advisory group. She has also provided guidance, based on her research findings, to other social media and technology companies. JR is a member of Meta's Global Suicide and Self-injury safety board and the Australian online safety advisory group. She has also provided guidance, based on her research findings, to other social media platforms.

Multimedia Appendix 1

XSL•FO
RenderX

Consolidated criteria for reporting qualitative studies (COREQ) - 32-item checklist.
[PDF File (Adobe PDF File), 125 KB - jmir_v27i1e66321_app1.pdf]

Multimedia Appendix 2
Semistructured topic guide used to conduct the focus groups.
[DOC File , 29 KB - jmir_v27i1e66321_app2.doc]

Multimedia Appendix 3
Representation of thematic framework, including description of themes and subthemes, and illustrative quotes from each case.
[DOC File , 50 KB - jmir_v27i1e66321_app3.doc]

References

1. Australian Bureau of Statistics. Causes of death, Australia. In: Australian Bureau of Statistics. Canberra: Australian Bureau of Statistics; 2022.
2. WHO. Suicide Rate Estimates, Age-Standardized: Estimates by Country. Geneva: World Health Organization; 2021.
3. Duarte T, Paulino S, Almeida C, Gomes H, Santos N, Gouveia-Pereira M. Self-harm as a predisposition for suicide attempts: a study of adolescents' deliberate self-harm, suicidal ideation, and suicide attempts. Psychiatry Res 2020;287:112553. [doi: 10.1016/j.psychres.2019.112553] [Medline: 31526562]
4. Luxton D, June J, Fairall J. Social media and suicide: a public health perspective. Am J Public Health 2012;102(Suppl 2):S195-S200. [doi: 10.2105/AJPH.2011.300608] [Medline: 22401525]
5. Pirkis J, Mok K, Robinson J, Nordentoft M. Media influences on suicidal thoughts and behaviours. In: O'Connor RC, Pirkis J, editors. The International Handbook of Suicide Prevention. Hoboken, NJ: Wiley-Blackwell; 2016:743-757.
6. Adams R. Social media urged to take 'moment to reflect' after girl's death. The Guardian Internet. 2019. URL: https://www.theguardian.com/media/2019/jan/30/social-media-urged-to-take-moment-to-reflect-after-girls-death [accessed 2025-02-11]
7. Hawton K, Hill N, Gould M, John A, Lascelles K, Robinson J. Clustering of suicides in children and adolescents. Lancet Child Adolesc Health 2020;4(1):58-67. [doi: 10.1016/S2352-4642(19)30335-9] [Medline: 31606323]
8. Shoib S, Chandradasa M, Nahidi M, Amanda T, Khan S, Saeed F, et al. Facebook and suicidal behaviour: user experiences of suicide notes, live-streaming, grieving and preventive strategies-a scoping review. Int J Environ Res Public Health 2022;19(20):13001 [FREE Full text] [doi: 10.3390/ijerph192013001] [Medline: 36293584]
9. Khasawneh A, Chalil Madathil K, Dixon E, Wiśniewski P, Zinzow H, Roth R. Examining the self-harm and suicide contagion effects of the blue whale challenge on YouTube and Twitter: qualitative study. JMIR Ment Health 2020;7(6):e15973 [FREE Full text] [doi: 10.2196/15973] [Medline: 32515741]
10. Samaritans. How Social Media Users Experience Self-Harm and Suicide Content. Surrey: Samaritans; 2022.
11. Robinson J, La Sala L, Kenny B, Cooper C, Lamblin M, Spittal M. How do Australian social media users experience self-harm and suicide-related content? A national cross-sectional survey comparing young people and adults Internet. PsyArXiv 2024 [FREE Full text] [doi: 10.31234/osf.io/5rh93]
12. Eko. Suicide, Incels, and Drugs: How TikTok's Deadly Algorithm Harms Kids. Washington: Eko; 2023.
13. Lima-Strong C. Facebook under fire: a whistleblower's power: key takeaways from the facebook papers. The Washington Post Internet. URL: https://www.washingtonpost.com/technology/2021/10/25/what-are-the-facebook-papers/ [accessed 2021-10-26]
14. Thorn P, La Sala L, Hetrick S, Rice S, Lamblin M, Robinson J. Motivations and perceived harms and benefits of online communication about self-harm: an interview study with young people. Digit Health 2023;9:20552076231176689 [FREE Full text] [doi: 10.1177/20552076231176689] [Medline: 37252260]
15. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: the good, the bad and the unknown. PLoS One 2017;12(8):e0181722 [FREE Full text] [doi: 10.1371/journal.pone.0181722] [Medline: 28813437]
16. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
17. La Sala L, Pirkis J, Cooper C, Hill N, Lamblin M, Rajaram G, et al. Acceptability and potential impact of the #chatsafe suicide postvention response among young people who have been exposed to suicide: pilot study. JMIR Hum Factors 2023;10:e44535 [FREE Full text] [doi: 10.2196/44535] [Medline: 37204854]
18. La Sala L, Teh Z, Lamblin M, Rajaram G, Rice S, Hill N, et al. Can a social media intervention improve online communication about suicide? A feasibility study examining the acceptability and potential impact of the #chatsafe campaign. PLoS One 2021;16(6):e0253278 [FREE Full text] [doi: 10.1371/journal.pone.0253278] [Medline: 34129610]
19. Robinson J, Hill N, Thorn P, Battersby R, Teh Z, Reavley N, et al. The #chatsafe project. Developing guidelines to help young people communicate safely about suicide on social media: a Delphi study. PLoS One 2018;13(11):e0206584 [FREE Full text] [doi: 10.1371/journal.pone.0206584] [Medline: 30439958]

XSL•FO
RenderX

20. Robinson J, La Sala L, Cooper C, Spittal M, Rice S, Lamblin M, et al. Testing the impact of the #chatsafe intervention on young people's ability to communicate safely about suicide on social media: protocol for a randomized controlled trial. JMIR Res Protoc 2023;12:e44300 [FREE Full text] [doi: 10.2196/44300] [Medline: 36800220]
21. Robinson J, Thorn P, McKay S, Hemming L, Battersby-Coulter R, Cooper C, et al. #chatsafe 2.0. updated guidelines to support young people to communicate safely online about self-harm and suicide: a Delphi expert consensus study. PLoS One 2023;18(8):e0289494 [FREE Full text] [doi: 10.1371/journal.pone.0289494] [Medline: 37531397]
22. Bailey E, Alvarez-Jimenez M, Robinson J, D'Alfonso S, Nedeljkovic M, Davey C, et al. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. Int J Environ Res Public Health 2020;17(7):2435 [FREE Full text] [doi: 10.3390/ijerph17072435] [Medline: 32260111]
23. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
24. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (Moderated Online Social Therapy+) for help-seeking young people experiencing mental Ill-health: pilot evaluation within a national youth E-mental health service. J Med Internet Res 2020;22(8):e17155 [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]
25. Platt S, Arensman E, Rezaeian M. National suicide prevention srategies - progress and challenges. Crisis 2019;40(2):75-82 [FREE Full text] [doi: 10.1027/0227-5910/a000587] [Medline: 30892084]
26. Jang Y, Ko B. Online safety for children and youth under the 4Cs framework-a focus on digital policies in Australia, Canada, and the UK. Children (Basel) 2023;10(8):1415 [FREE Full text] [doi: 10.3390/children10081415] [Medline: 37628414]
27. World Health Organization. Suicide worldwide in 2019. In: Global Health Estimates. Geneva: World Health Organization; 2021.
28. National Suicide Prevention Project Reference Group. National Suicide Prevention Strategy for Australia's Health System: 2020-2023. Melbourne: Department of Health and Human Services; 2023.
29. World Health Organization. National Suicide Prevention Strategies: Progress, Examples and Indicators. Geneva: World Health Organization; 2018.
30. Department of health and social care. Suicide Prevention in England: 5-Year Cross-Sector Strategy. London: Department of Health and Social Care; 2023.
31. U.S. Department of Health and Human Services. National Strategy for Suicide Prevention. Washington: Department of Health and Human Services; 2024.
32. Ministry of Health and Care Services. National Strategy for Suicide Prevention 2020-2025. Oslo: Ministry of Health and Care Services; 2020.
33. Online Safety Act 2023. UK Public General Acts. 2023. URL: https://www.legislation.gov.uk/ukpga/2023/50/contents [accessed 2025-02-11]
34. Online Safety Act 2021. Australian Government 2022. 2021. URL: https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf [accessed 2025-02-11]
35. Children's online privacy protection rule. U.S. Federal Trade Commission. 2013. URL: https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa [accessed 2025-02-11]
36. Kids Online Safety Act. U.S. Congress. 2023. URL: https://www.congress.gov/bill/118th-congress/senate-bill/1409 [accessed 2025-02-11]
37. Online Safety and Media Regulation Act 2022. Ireland; 2022. URL: https://www.irishstatutebook.ie/eli/2022/act/41/enacted/en/html [accessed 2025-02-11]
38. Safety resources and support on snapchat. Snap Inc. 2024. URL: https://help.snapchat.com/hc/en-us/articles/7012345957524-Safety-Resources-and-Support-on-Snapchat [accessed 2025-02-11]
39. Self-injury. Meta. 2024. URL: https://help.instagram.com/553490068054878 [accessed 2025-02-11]
40. Suicide and self-injury resources. Meta. 2024. URL: https://www.facebook.com/help/1553737468262661 [accessed 2025-02-11]
41. TikTok. 2024. URL: https://www.tiktok.com/safety/en-us/suicide-self-harm/ [accessed 2025-02-11]
42. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
43. Thorn P, McKay S, Hemming L, Reavley N, La SL, Sabo A. #chatsafe: A Young Person's Guide to Communicating Safely Online About Self-Harm and Suicide. Melbourne: Orygen; 2023:a.
44. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
45. Michail M, Tait L. Exploring general practitioners' views and experiences on suicide risk assessment and management of young people in primary care: a qualitative study in the UK. BMJ Open 2016;6(1):e009654 [FREE Full text] [doi: 10.1136/bmjopen-2015-009654] [Medline: 26758263]

XSL•FO
RenderX

46. Heath G, Cameron E, Cummins C, Greenfield S, Pattison H, Kelly D, et al. Paediatric 'care closer to home': stake-holder views and barriers to implementation. Health Place 2012;18(5):1068-1073. [doi: 10.1016/j.healthplace.2012.05.003] [Medline: 22687749]
47. Naslund J, Bondre A, Torous J, Aschbrenner K. Social media and mental health: benefits, risks, and opportunities for research and practice. J Technol Behav Sci 2020;5(3):245-257 [FREE Full text] [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
48. Office of the Surgeon General (OSG). The U.S. Surgeon General's Advisory. Social Media and Youth Mental Health. Washington, DC: U.S. Department of Health and Human Services; 2023.
49. Inquiry into social media and online safety: eSafety commissioner submission. eSafety Commissioner. Canberra: eSafety Commissioner; 2022. URL: https://www.esafety.gov.au/sites/default/files/2023-04/eSafety-submission-Inquiry-into-social-media-and-online-safety.pdf [accessed 2025-02-11]
50. May N. 'Out of control': 60% of Australian parents and carers surveyed say social media is their biggest concern for kids. The Guardian Internet. 2024. URL: https://www.theguardian.com/media/2024/feb/19/out-of-control-60-of-parents-and-carers-surveyed-say-social-media-is-their-biggest-concern-for-kids [accessed 2025-02-11]
51. Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. Soc Psychiatry Psychiatr Epidemiol 2020;55(4):407-414. [doi: 10.1007/s00127-019-01825-4] [Medline: 31925481]
52. Orben A, Przybylski AK. Reply to: underestimating digital media harm. Nat Hum Behav 2020;4(4):349-351. [doi: 10.1038/s41562-020-0840-y] [Medline: 32303718]
53. Valkenburg P, Meier A, Beyens I. Social media use and its impact on adolescent mental health: an umbrella review of the evidence. Curr Opin Psychol 2022;44:58-68 [FREE Full text] [doi: 10.1016/j.copsyc.2021.08.017] [Medline: 34563980]
54. Vuorre M, Przybylski A. Global well-being and mental health in the internet age. Clin Psychol Sci 2023;12(5):917-935 [FREE Full text] [doi: 10.1177/21677026231207791]
55. Community standards enforcement report. Meta. 2024. URL: https://transparency.meta.com/reports/community-standards-enforcement/ [accessed 2025-02-11]
56. Transparency report internet. Snap Inc. 2023. URL: https://values.snap.com/privacy/transparency [accessed 2025-02-11]
57. Community guidelines enforcement report. TikTok. 2024. URL: https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2023-4/ [accessed 2025-02-11]
58. Promoting a safer and more trustworthy internet. Digital Trust & Safety Partnership. 2023. URL: https://dtspartnership.org/ [accessed 2025-02-11]
59. Learn more about the meta safety advisory council. Meta. 2024. URL: https://www.facebook.com/help/222332597793306 [accessed 2025-02-11]
60. Meet TikTok's European Safety Advisory Council. TikTok. 2021. URL: https://newsroom.tiktok.com/en-gb/tiktok-european-safety-advisory-council [accessed 2025-02-11]
61. Robinson J, Thorn P, McKay S, Richards H, Battersby-Coulter R, Lamblin M, et al. The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. Front Child Adolesc Psychiatry 2023;2:1274263 [FREE Full text] [doi: 10.3389/frcha.2023.1274263] [Medline: 39839592]
62. La Sala L, Sabo A, Lamblin M, Robinson J. Can #chatsafe support parents and carers beyond Australia? A qualitative study. BMC Public Health 2024;24(1):1571 [FREE Full text] [doi: 10.1186/s12889-024-19040-5] [Medline: 38862974]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research

Edited by A Mavragani; submitted 10.09.24; peer-reviewed by C Brennan, F Walby; comments to author 29.11.24; revised version received 16.12.24; accepted 16.01.25; published 10.03.25.

Please cite as:
La Sala L, Sabo A, Michail M, Thorn P, Lamblin M, Browne V, Robinson J
Online Safety When Considering Self-Harm and Suicide-Related Content: Qualitative Focus Group Study With Young People, Policy Makers, and Social Media Industry Professionals
J Med Internet Res 2025;27:e66321
URL: https://www.jmir.org/2025/1/e66321
doi:10.2196/66321
PMID:

XSL•FO
RenderX

©Louise La Sala, Amanda Sabo, Maria Michail, Pinar Thorn, Michelle Lamblin, Vivienne Browne, Jo Robinson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

BePresent Universal Internet-Based Parenting Intervention: Single-Arm Pre-Post Intervention Study

Kaisa Mishina[1,2], PhD; Amit Baumel[3], PhD; Malin Kinnunen[1,2], MHc; Terja Ristkari[1,2], MNS; Emmi Heinonen[1,2], MSc; Susanna Hinkka-Yli-Salomäki[1,2], PhLic; Andre Sourander[1,2,4], MD, PhD

[1]Research Centre for Child Psychiatry, Faculty of Medicine, University of Turku, Turku, Finland

[2]INVEST Research Flagship Centre, University of Turku, Turku, Finland

[3]Department of Community Mental Health, University of Haifa, Haifa, Israel

[4]Department of Child Psychiatry, Turku University Hospital, Turku, Finland

Corresponding Author:
Kaisa Mishina, PhD
Research Centre for Child Psychiatry
Faculty of Medicine
University of Turku
Lemminkäisenkatu 3
Turku, 20014
Finland
Phone: 358 50 310 135
Email: kaemka@utu.fi

Abstract

Background: Internet-based parenting programs have great potential to promote positive parent-child relationships as well as to reach and engage parents.

Objective: This study aimed to assess the universal internet-based BePresent parenting intervention for families with 3-year-old children and how it influences the child's behavior and daily-life situations assessed by parents. The first aim of the study was to assess the change from baseline to follow-up in child hyperactivity and conduct problems, affective reactivity, and daily activities. The second aim was to assess intervention completion rates. The third aim was to evaluate parent satisfaction with the intervention. The fourth aim was to assess all outcomes by comparing those who completed the intervention and those who did not.

Methods: We conducted a single-arm pre- and postintervention study. Parents attending their child's 3-year health check-up were recruited from children's health clinics. The intervention was an unguided internet-based parenting program consisting of 5 modules. Self-reported measures were collected at baseline and at an 8-week follow-up. Linear mixed-effects models were used to analyze the changes from baseline to follow-up.

Results: Altogether, 752 parents registered, and 515 started the intervention. Of those, 36% (n=183) completed the intervention. Parents reported high satisfaction with the intervention: the majority (68.8%–84.9%) were satisfied with various aspects of the program, and 89.9% said the intervention provided information about positive parenting skills. The findings show significant decreases with small effect sizes in parents' ratings of child hyperactivity (P=.03; d=0.12) and conduct problems (P=.001; d=0.20) between baseline and the 8-week follow-up. A similar finding was observed in the parent ratings of child irritability (P≤.001; d=0.27) using the Affective Reactivity Index. Parents reported improvement in the daily functioning of their child when it was measured with a questionnaire adapted from the Barkley Home Situations Questionnaire (P=.01; d=0.14).

Conclusions: Universal digital interventions have the potential to be implemented widely in community settings to improve knowledge and positive parenting skills. However, there is a need to assess the efficacy of digital universal interventions using randomized controlled designs and to examine additional ways to increase adherence to universal programs.

(J Med Internet Res 2025;27:e65391) doi:10.2196/65391

XSL•FO
RenderX

KEYWORDS

parent training; universal intervention; online intervention; irritability; conduct problems; hyperactivity; preschool; mental health; strongest families; positive parenting; parenting skills; parent-child relationships; parent satisfaction; BePresent; feasibility study; single-arm pre-post intervention study

Introduction

Consistent evidence has shown that parent training can lead to improvements in parenting practices and that it promotes the psychosocial development of children [1,2], which can, in turn, prevent various psychosocial and mental health difficulties [1,3]. Therefore, early preventive parent training programs are needed to enhance parent, and child relationships and to promote the healthy development of children [2].

Preventative interventions can be indicated, selective, or universal [4]. While indicated and selective interventions are intended for individuals with identified high risk, universal interventions can be offered to the general public or the entire population, typically regardless of individual risk factors [5]. Universal interventions aim to prevent the onset of illness or behavior by reducing exposure of risk factors [6]. In universal parent training, this means improving positive parenting [7], strengthening the relationships and interactions between parents and children [8], and reducing dysfunctional or abusive parenting practices [7]. In comparison to indicated or selective parenting programs, universal ones are typically briefer, targeting ordinary parenting challenges, and are hence aimed at all parents [9].

Universal programs typically have a large reach. They are considered to be less stigmatized, since any parent can experience parenting as challenging and thus need support [10]. Moreover, when universal parenting programs are offered in an internet-based format, typical barriers related to face-to-face programs can be avoided, such as logistical and financial barriers as well as the shortage of trained professionals [11]. Attrition rates in internet-based parenting programs typically vary between 30% and 50%, similar to that of face-to-face parenting programs [12], but internet-based programs generally have higher universal uptake [13].

A recent systematic review assessing universal digital parenting programs found small to moderate improvements in parent depression, anxiety, self-efficacy, and social support, but found no effects on parent's stress, satisfaction, or parent-child relationship quality [11]. Due to limited data, the review was unable to assess the effects of universal digital parent training on child-specific outcomes [11], pointing out a clear research gap in the field. While the promise of universal parenting programs is conceptually clear, their impact on public health remains unclear.

This study focuses on the universal internet-based parenting program BePresent. The program is based on a wider 11-week targeted parent training program, Finnish Strongest Families, which has demonstrated efficacy in randomized controlled design up until a 2-year follow-up [14,15]. Since the randomized controlled study, the Strongest Families program has been implemented nationwide in Finland through child health clinics for parents whose 4-year-old children have disruptive behavior problems [16,17]. The parent training program has also been studied in clinical settings among 3 years-10 years–old children with disruptive behavior problems [18]. BePresent is a universal digital nonguided intervention aimed at all parents with 3-year-old children. Within the Finnish health system, 77% of preschoolers attend annual health check-ups at a child health clinic [19], which makes these visits an ideal setting for the delivery of such a universal intervention.

The first aim of the study was to assess changes in child hyperactivity and conduct problems, affective reactivity, and daily activities using a pre-post, single-arm intervention study design. The second aim was to assess intervention completion rates. The third aim was to evaluate parent satisfaction with the intervention. The fourth aim was to assess all outcomes by comparing those who completed the intervention and those who did not.

Methods

Study Design

A single-arm pre- and postintervention study involved the recruitment of participants to use the BePresent parenting intervention. Measures were conducted at baseline before starting the intervention and at an 8-week follow-up after filling in the baseline questionnaire.

Intervention

BePresent is a universal, unguided internet-based parenting intervention. The content and structure of the intervention are based on the Finnish Strongest Families targeted parent training program. Patterson's Coercion Theory guided the development, emphasizing the importance of positive reinforcement in parent-child interactions. The universal intervention follows the basics of the targeted program but was shortened to be feasible for all parents with 3-year-old children. BePresent focuses on the promotion of healthy parenting practices and developing parenting skills that strengthen the relationship and communication with the child. The topics include reinforcing the child's positive behavior, reducing conflict situations, planning situations in advance, managing daily transitions, and being emotionally present with the child. The intervention includes 5 modules, and the total intervention period is 8 weeks. Detailed information about the key training elements and parental goals of each module are presented in Table 1.

The participants must complete the modules sequentially; they cannot select which module they start with. Each module includes theory material, example videos, comic strips (Figure 1), video practices, and exercises. In the comic strips, the central ideas of the program's example videos are presented. For example, in module 5 the importance of spending time with the child, for example, by playing, is presented in the video and in the comic strip. In both, first, a negative situation is presented, followed by a positive example of parents using the practiced

XSL•FO
RenderX

parenting skills. The comic strips use visual-driven storytelling, including pictures without text. With this, we wanted to convey a universal understanding, across different languages. There are different variations of individual strategies to choose from to train parenting skills in practice. To motivate and promote adherence to the program, standardized and same-for-all SMS and email reminders were sent regularly during the program. The participants had 8 weeks to complete the intervention. After this, they were asked to fill out the follow-up questionnaire.

Table 1. Structure and content of the BePresent internet-based parenting program.

Module	Key training elements	Parental goals
Notice the good in your child	• Positive attention • Positive feedback • Reducing unnecessary reminders	• Positive interaction skills with child • Active parenting
Understand your child	• Acting in daily situations • Boundaries for the child • Self-controlled parenting	• Self-regulation with child • Staying calm • Recognizing negative thoughts • Turning thoughts into more positive
Prepare your child for changes	• Modeling daily transitions • Importance of daily routines	• Reinforcing good daily routines
Plan ahead with your child	• Planning together with the child • Preparing a child for upcoming situations	• Involving the child in planning and reinforce good daily routines
Be present for your child	• Conscious presence with the child • Spending time with the child • Mindfulness	• Being present in the daily moments with the child

Figure 1. Negative and positive examples of how to be more present with your child in daily situations.

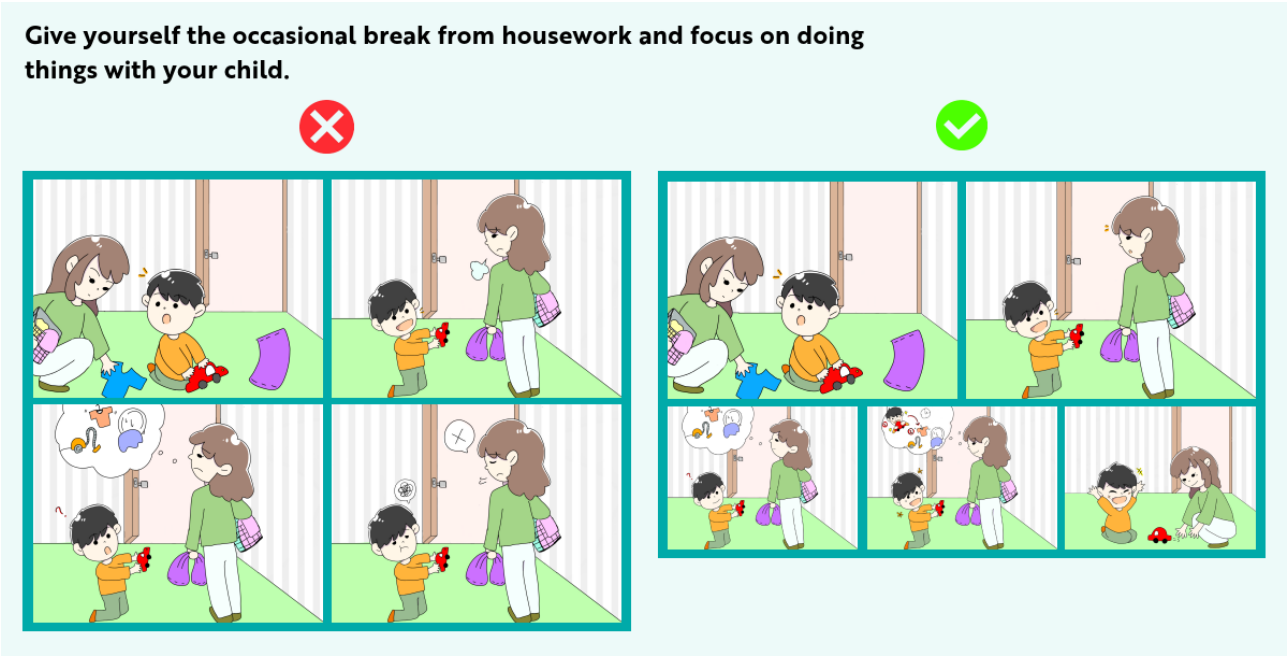

Participants and Setting

Parents with a 3-year-old child participating in annual health check-ups in primary care child health clinics in Finland were considered eligible and were asked to participate. The program was offered in Finnish, Swedish, and English; therefore, at least one of the parents had to understand one of those languages for the family to participate. The study setting was the public child health clinics, and the health check-ups offered there are free of charge for the families. Out of the 21 well-being services counties in Finland, 14 (67%) were included in this study. These well-being service counties were included as they were a part of our broader implementation study, selected to represent larger and smaller cities in both urban and rural areas across Finland. Since the operations of child health clinics are regulated by the Finnish Health Care Act [20] and the Government Decree [21], quality and operations are relatively uniform across the counties.

Procedure

There were 2 paths to participating in the BePresent program. First, as a part of the annual health check-up, a public health nurse informed parents about the program and the study.

XSL•FO
RenderX

Information was also provided in written format on a flyer that the public health nurse gave to the parents during the check-up. The flyer included the program's web address, where the parents could sign in and complete the registration form (name, email, and phone number). Second, 3 well-being services counties also recruited parents to the BePresent intervention through their website and social media pages. This internet-based information was similar to that on the flyer. A link was provided that, when followed, led parents to a form that screened them for inclusion criteria through a series of questions (which well-being services county did they live in; were they a guardian of a 3-year-old; did they understand Finnish, Swedish, and English). If the parents fulfilled the inclusion criteria, registration instructions were automatically sent to them by email.

The parents were asked to register for the intervention 1 month after the health check-up at the latest, but the registration stayed open until the end of the study. The website included information about the study and the intervention as well as an electronic informed consent form. After giving informed consent, parents were asked to fill in a baseline evaluation. The intervention lasted for 8 weeks. After this, parents were asked to complete a follow-up questionnaire; they also received email and text message reminders about the follow-up questionnaire. Intervention completers were defined as parents who completed the first 4 modules since those modules included the key elements of the parent training. The fifth theme concentrated on conscious presence with the child; this was considered to be an additional skill.

Measures

Demographic Background Information

To gather demographic background information, participating parents were asked about their gender, their child's sex, their family structure, and their role in the child's life. The question about family structure referred to who lived with the child, and the response options were (1) both biological parents, (2) the biological mother or father and their spouse, (3) the biological mother or father alone, (4) adoptive parents, (5) foster parents, (6) parents with the same gender, or (7) some other type of family. In the analysis, family types other than those with both biological parents were pooled together.

Child Hyperactivity and Conduct Problems

Hyperactivity and conduct problems were assessed with scales based on the content of subscales of the Strengths and Difficulties Questionnaire [22,23]. The Strengths and Difficulties Questionnaire internalizing subscales were not included in the questionnaire because the aim was to have as brief of a questionnaire as possible. Altogether, the scales measuring conduct problems and hyperactivity consisted of 10 items. Responses were based on the format 0=not true, 1=somewhat true, and 2=certainly true, so the scores for each scale of 5 items could range from 0 to 10. In the current study, the Cronbach alpha was 0.77 and the McDonald's omega was 0.79 for hyperactivity and .69 (α) and 0.70 (ω) for conduct problems, indicating acceptable reliability [24,25].

Child Affective Reactivity

The affective reactivity index-P [26] is a 7-item instrument measuring the irritability of a child. The affective reactivity index-P has been proven to be valid for measuring parent-rated irritability among small children [27]. Each item could be responded to using a scale of 0 to 2, indicating the responses of not true, somewhat true, or certainly true, respectively. The total score was calculated as the sum of the first 6 items, meaning that the total score range was 0-12. The impairment item was not counted in the total score. The McDonald's omega was 0.86 and the Cronbach α was 0.84, indicating acceptable reliability [24,25].

Daily Activities

Daily activities were assessed by asking the parents to rate the impact of the child's behavior during daily transitions, including getting dressed, getting ready for daycare, during the evening meal, and getting ready for bed. It also covered social interactions, including playing with siblings and other children, traveling situations, and being in public places such as the supermarket. The questionnaire was adapted from the Barkley Home Situations Questionnaire, which asks the parent to rate whether the child's behavior causes problems during specified daily routines [28]. The instrument has been previously used in the Finnish Strongest Families parent training program when it presented adequate reliability with a Cronbach α of 0.64 [17]. In the current study, the McDonald's omega was 0.87 and the Cronbach α was 0.87, indicating acceptable reliability [24,25].

Intervention Completion

Completion and non-completion of the intervention for the 8-week period was assessed using the data recorded by the intervention platform. Intervention completers were defined as parents who completed module 4, as the first 4 modules included the key elements of the intervention. Participation in the follow-up measures was also assessed using the data from the intervention platform.

Intervention Satisfaction

Participants were asked to rate their satisfaction with the intervention with 1 item, with the response options being 0=very dissatisfied, 1=somewhat dissatisfied, 2=neither dissatisfied nor satisfied, 3=somewhat satisfied, and 4=very satisfied. In the analysis, the options were categorized as dissatisfied, neither dissatisfied nor satisfied, and satisfied. The questionnaire included 10 items used to rate how well participants felt the intervention provided support in specific areas of parenting. The response options were 0=disagree, 1=somewhat disagree, 2=neither disagree nor agree, 3=somewhat agree, and 4=agree. In the analysis, the responses were categorized as disagree, neither disagree nor agree, and agree.

Statistical Analysis

Demographic characteristics and satisfaction with the program are presented as numbers and percentages. The difference between groups of completers and noncompleters were tested with Pearson's chi-squared test. The changes in outcome variables from baseline to the 8-week follow-up (with the unit of the analyses being parents' ratings of their children) were

XSL•FO
RenderX

analyzed with linear mixed effect models for those who completed the follow-up questionnaire, with time as the within factor (baseline and 8-week follow-up) and the child's sex as a covariate. The differences in outcome variables between completers and noncompleters at the 8-week follow-up were analyzed with linear regression models. Because the baseline values differed between groups, the models were adjusted for the baseline measurement of the outcome and the child's sex. A 2-sided significance level of .05 was used during the statistical testing, and 95% Cl were calculated for the point estimates. Cohen *d* was calculated as a measure of effect size, to complement the statistical testing. The statistical analyses were carried out with SAS statistical software (SAS 9.4, SAS Institute).

Ethical Considerations

All study procedures and human subject research ethics were approved by the Ethical Committee of the University of Turku (45/2017, 11 Sep 2017). Study permission was granted from each study site, including Wellbeing Services County of South Karelia, Wellbeing Services County of South Ostrobothnia, South Savo Wellbeing Services County, Wellbeing Services County of Kymenlaakso, Wellbeing Services County of Lapland, Western Uusimaa Wellbeing Services County, Wellbeing Services County of Pirkanmaa, Wellbeing Services County of Ostrobothnia, Wellbeing Services County of Satakunta, Wellbeing Services County of North Karelia, Wellbeing Services County of Central Ostrobothnia, Wellbeing Services County of Vantaa and Kerava. and Wellbeing Services County of Southwest Finland. The participants provided informed consent before enrollment in the study. No form of compensation, monetary or otherwise, was offered to participants. Privacy and confidentiality of the participants was ensured by adhering to ethical guidelines, securely handling data and pseudonymizing identifiable information. The data was pseudonymized and stored in the University of Turku secured project folder (short term) with automatic backup system and in "Taltio" and/or "Isilon" data storage cluster (long term). We followed the TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) Statement Checklist for complete and detailed reporting (Multimedia Appendix 1).

Results

Detailed information about enrolment, participation per module, and completion rates of the modules are presented in Figure 2. Altogether, 752 parents registered for the BePresent parenting intervention and 515 started the intervention; 32% registered but did not begin the intervention. Out of the 515 who started the intervention, 202 (36%) answered the follow-up questionnaire, and 183 (36%) completed the intervention. The highest dropout rate occurred between the first and second modules, with a dropout rate of approximately 34%.

Table 2 presents the participants' background characteristics and a comparison between those of completers and noncompleters. The majority of participating parents were mothers (90.9%) and most of the children were boys (57.9%). Most families had a structure including both biological parents (92.2%). The percentage of completers who were boys was significantly higher than that of noncompleters who were boys (67.8% vs 52.4%, $P<.001$). No other demographic difference was found between intervention completers and noncompleters. The average time spent on the intervention website was 8.3 hours among those who completed the intervention and 3.5 hours among noncompleters (Table 3).

Figure 2. Flow chart of the study participants and information on completers and noncompleters. Began Module 1 n=515. Completed first 4 modules n=125. Started the intervention but did not complete the first 4 modules n=77.

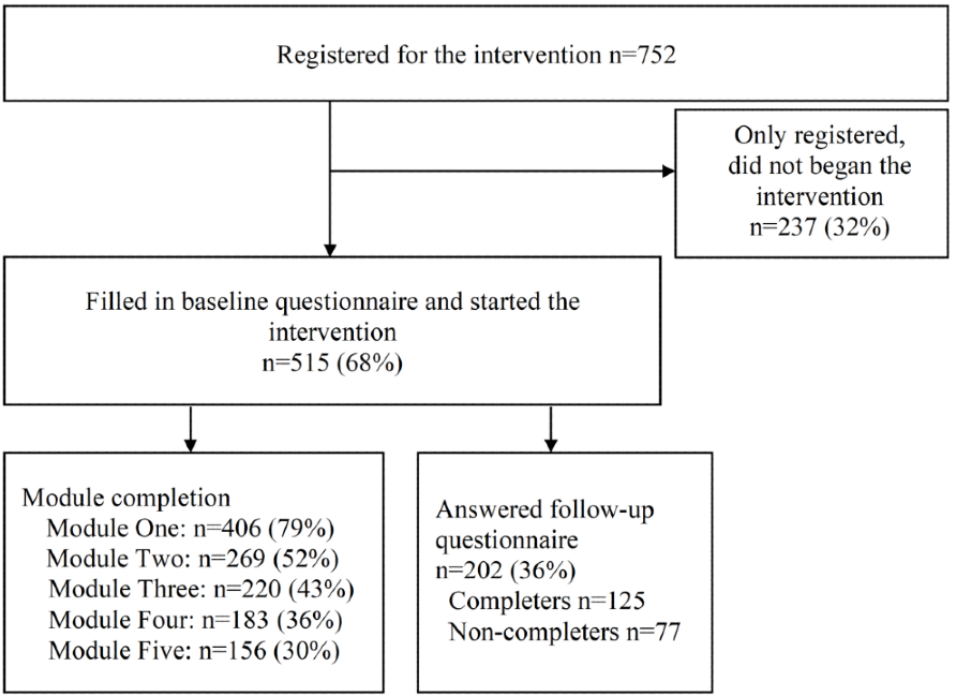

XSL•FO
RenderX

Table 2. Background characteristics and comparison between completers and noncompleters.

Characteristics	Total (N=515), n (%)	Completers (n=183), n (%)	Noncompleters (n=332), n (%)
Child's sex			
Boy	298 (57.9)	124 (67.8)	174 (52.4)
Girl	217 (42.1)	59 (32.2)	158 (47.6)
Family structure			
Both biological parents	475 (92.2)	170 (92.9)	305 (91.9)
Other	40 (7.8)	13 (7.1)	27 (8.1)
Participant's gender[a]			
Women	463 (89.9)	164 (89.6)	299 (90.1)
Men	46 (8.9)	16 (8.7)	30 (9.0)
Not specified	6 (1.2)	3 (1.6)	3 (0.9)

[a]Gender of the parent who registered for the program and responded to the questionnaires.

Table 3. Intervention usage of the participants and comparison between completers and noncompleters.

Characteristics	Total (N=515)	Completers (n=183)	Noncompleters (n=332)	Completers versus noncompleters (*P* value[a])
Intervention usage, mean (SD)				
Number of days completing the intervention (days)	19.1 (22.3)	28.6 (19.6)	11.2 (21.3)	<.001
Number of days per completed module (days)	6.4 (12.0)	5.9 (4.1)	6.8 (15.7)	.45
Active use of website per completed module (hours)	2.1 (1.6)	1.7 (1.2)	2.4 (1.8)	<.001
Active total use of the website (hours)	5.7 (4.9)	8.3 (5.7)	3.5 (2.8)	<.001
Time from beginning the module to completing the module (hours, median)				
Module 1	0.24	—[b]	—	—
Module 2	0.19	—	—	—
Module 3	0.11	—	—	—
Module 4	0.16	—	—	—

[a]Student *t* test.

[b]Not applicable.

Behavioral Outcomes

Table 4 presents a sample including completers and noncompleters, showing the changes in child behavior and daily situations from baseline to the follow-up measurement for those who completed the measures at both time points (n=202). There were significant improvements from baseline to the 8-week follow-up in hyperactivity (*P*=.03; *d*=0.11) and conduct problems (*P*=.001; *d*=0.20). Child affective reactivity also improved significantly (*P*<.001; *d*=0.26) from baseline to follow-up. In daily situations, improvements were found in total scores (*P*=.01; *d*=0.14) as well as in transition (*P*<.001; *d*=0.25) and eating situations (*P*=.002; *d*=0.19).

While Table 4 includes results for completers and noncompleters combined, Table 5 presents a comparison between the 2 groups and a more detailed description of changes among completers and among noncompleters from baseline to follow-up. As presented in Table 5, affective reactivity improved more among noncompleters than among completers (*P*=.03; *d*=0.21). At baseline, completers had more difficulties in affective reactivity, hyperactivity, and conduct problems than noncompleters.

XSL•FO
RenderX

Table 4. Comparing hyperactivity and conduct problems, affective reactivity, and daily situations mean scores between baseline and the 8-week follow-up among those who completed the baseline and the follow-up questionnaire, adjusted by the children's gender.

Total (n=202)[a]	Baseline, mean[b] (SE)	Follow-up, mean[b] (SE)	Change from baseline to 8-week follow-up		
			Mean change value[b] (95% CI)	*P* value	Cohen *d* (95% CI)
Behavioral outcome (n=198)					
Hyperactivity[a,d]	3.9 (0.2)	3.7 (0.2)	0.3 (0.0 to 0.5)	.03	0.11 (–0.08 to 0.31)
Conduct problems[a,d]	3.6 (0.1)	3.2 (0.1)	0.4 (0.2 to 0.6)	.001	0.20 (–0.001 to 0.40)
Affective Reactivity (n=202)					
Total	4.2 (0.2)	3.4 (0.2)	0.8 (0.5 to 1.1)	<.001	0.26 (0.06 to 0.46)
Daily situations (n=190)					
Total score	35.3 (0.7)	34.0 (0.7)	1.3 (0.3 to 2.3)	.01	0.14 (–0.06 to 0.33)
Transition situations	12.7 (0.3)	11.7 (0.3)	1.0 (0.5 to 1.5)	<.001	0.25 (0.05 to 0.45)
Eating situations	7.2 (0.2)	6.6 (0.2)	0.5 (0.2 to 0.8)	.002	0.19 (–0.01 to 0.39)
Situations outside home	7.5 (0.2)	7.6 (0.2)	–0.1 (–0.4 to 0.2)	.67	–0.03 (–0.23 to 0.18)
Situations inside home	7.9 (0.2)	8.1 (0.2)	–0.2 (–0.5 to 0.2)	.42	–0.05 (–0.25 to 0.15)

[a]Includes both intervention completers and noncompleters.

[b]Model based on least squares means.

XSL•FO
RenderX

Table 5. Comparing baseline and follow-up changes in hyperactivity, conduct problems, affective reactivity, and daily situations between intervention completers and noncompleters.

	Completers (N=125)				Noncompleters (n=77)				Completers versus noncompleters		
	Baseline, mean[a] (SE)	Follow-up, mean[a] (SE)	Change[b]		Baseline, mean[a] (SE)	Follow-up, mean[a] (SE)	Change[b]				
			Mean[a] (95% CI)	*P* value			Mean[a] (95% CI)	*P* value	Mean[b,c] (95% CI)	*P* value	Cohen *d* (95% CI)
Behavioral outcome (n=198)											
Hyperactivity	4.1 (0.2)	3.8 (0.2)	0.3 (0.0 to 0.7)	.04	3.6 (0.3)	3.4 (0.3)	0.2 (–0.3 to 0.6)	.44	0.0 (–0.5 to 0.5)	.96	0.00 (–0.19 to 0.20)
Conduct problems	3.7 (0.2)	3.3 (0.2)	0.4 (0.1 to 0.7)	.006	3.4 (0.2)	3.0 (0.2)	0.4 (–0.0 to 0.7)	.077	0.0 (–0.5 to 0.5)	.98	0.00 (–0.19 to 0.20)
Affective reactivity (n=202)											
Total	4.4 (0.3)	3.7 (0.3)	0.6 (0.2 to 1.1)	.006	4.0 (0.3)	3.0 (0.2)	1.0 (0.6 to 1.4)	<.001	0.6 (0.0 to 1.1)	.03	0.21 (0.02 to 0.41)
Daily situations (n=190)											
Total score	35.6 (0.9)	34.0 (0.9)	1.5 (0.3 to 2.8)	.0 13	34.5 (1.0)	33.7 (1.0)	0.8 (–1.1 to 2.7)	.39	–0.2 (–2.2 to 1.9)	.8 8	–0.01 (–0.22 to 0.19
Transition situations	13.1 (0.4)	12.0 (0.4)	1.1 (0.5 to 1.7)	<.001	11.9 (0.4)	11.1 (0.4)	0.8 (–0.0 to 1.6)	.06	0.2 (– 0.7 to 1.1)	.6 3	0.05 (–0.15 to 0.25)
Eating situations	7.1 (0.3)	6.7 (0.3)	0.4 (0.0 to 0.8)	.03	7.1 (0.3)	6.5 (0.3)	0.7 (0.1 to 1.2)	.02	0.3 (–0.3 to 0.9)	.26	0.11 (–0.09 to 0.32)
Situations outside home	7.4 (0.2)	7.4 (0.2)	0.0 (–0.4 to 0.4)	1.00 0	7.6 (0.3)	7.8 (0.3)	–0.2 (–0.8 to 0.4)	.510	–0.2 (–0.8 to 0.4)	.54	–0.06 (–0.26 to 0.14)
Situations inside home	8.0 (0.3)	8.0 (0.3)	0.0 (–0.4 to 0.4)	.97	7.9 (0.3)	8.3 (0.4)	–0.5 (–1.2 to 0.3)	.21	–0.3 (–1.0 to 0.4)	.42	–0.08 (–0.28 to 0.12)

[a]Model based least squares means.

[b]Change from baseline to follow-up.

[c]Adjusted with corresponding baseline and child's sex.

Intervention Satisfaction

Parents reported high levels of satisfaction with the program (Table 6). Over 80% of the parents indicated that they would recommend the program to others and agreed that the program provided information about positive parenting skills as well as ways to notice the good in their child. Satisfaction was significantly higher among those who completed the program.

XSL•FO
RenderX

Table 6. Program satisfaction and acknowledged support in specific areas of parenting and comparison between completers and noncompleters.

Satisfaction-related factors[a]	Total (N=186), n (%)	Completers (n=123), n (%)	Noncompleters (n=63), n (%)	Completers versus non-completers (*P* value)
Satisfaction				
The program met my expectations				.001
Agree	142 (76.3)	104 (84.6)	38 (60.3)	
Neutral	27 (14.5)	13 (10.0)	16 (25.4)	
Disagree	17 (9.1)	11 (8.9)	9 (14.3)	
The program suited my needs				.005
Agree	141 (75.8)	102 (82.9)	39 (61.9)	
Neutral	21 (11.3)	11 (8.9)	10 (15.9)	
Disagree	24 (12.9)	10 (8.1)	14 (22.2)	
I could recommend the program to others				<.001
Agree	158 (84.9)	114 (92.7)	44 (69.8)	
Neutral	20 (10.8)	6 (4.9)	14 (22.2)	
Disagree	8 (4.3)	3 (2.4)	5 (7.9)	
I could join the program again				.10
Agree	147 (79.0)	102 (82.9)	45 (71.4)	
Neutral	21 (11.3)	13 (10.6)	8 (12.7)	
Disagree	18 (9.7)	8 (6.5)	10 (15.9)	
How satisfied you have been with the program				.003
Satisfied	128 (68.8)	94 (76.4)	34 (54.0)	
Neutral	37 (19.9)	21 (17.1)	13 (23.2)	
Dissatisfied	21 (11.3)	8 (6.5)	16 (25.4)	
Support in specific areas				
The program provided information about positive parenting skills				<.001
Agree	167 (89.8)	119 (96.7)	48 (76.2)	
Neutral	12 (6.5)	1 (0.8)	11 (17.5)	
Disagree	7 (3.8)	3 (2.4)	4 (6.3)	
The program offered ways to notice the good in my child				<.001
Agree	164 (88.2)	116 (94.3)	48 (76.2)	
Neutral	13 (7.0)	3 (2.4)	10 (15.9)	
Disagree	9 (4.8)	4 (3.3)	5 (7.9)	
The program gave me confidence in my ability to be a parent				.002
Agree	146 (78.5)	106 (86.2)	40 (63.5)	
Neutral	27 (14.5)	12 (9.8)	15 (23.8)	
Disagree	13 (7.0)	5 (4.1)	8 (12.7)	
The program offered ways to manage everyday situations				<.001
Agree	144 (77.4)	106 (86.2)	38 (60.3)	
Neutral	26 (14.0)	12 (9.2)	16 (25.4)	
Disagree	16 (8.6)	10 (8.1)	9 (14.3)	
The program offered ways to manage transitions				<.001
Agree	134 (72.0)	99 (80.5)	35 (55.6)	
Neutral	34 (18.3)	14 (11.4)	20 (31.7)	

XSL•FO
RenderX

Satisfaction-related factors[a]	Total (N=186), n (%)	Completers (n=123), n (%)	Noncompleters (n=63), n (%)	Completers versus non-completers (*P* value)
Disagree	18 (9.7)	10 (8.1)	8 (12.7)	
The program offered ways to be present with my child				.002
Agree	146 (78.5)	106 (86.2)	40 (63.5)	
Neutral	32 (17.2)	16 (12.3)	19 (30.2)	
Disagree	8 (4.3)	4 (3.3)	4 (6.3)	

[a]Disagree combines strongly disagree and disagree. Agree combines agree and strongly agree.

Discussion

Principal Findings

In this study, we assessed a universal unguided internet-based parenting intervention and whether it improves child behavior and daily activities. Hyperactivity, conduct problems, affective reactivity, and daily activities improved from baseline to follow-up, and parents reported a high level of satisfaction with the intervention and showed rather good adherence to it. Out of all the parents who started using the intervention, 36% (n=183) completed it.

The outcomes were also assessed by comparing those who completed the intervention and those who did not complete it, and we found that child irritability improved among the completers and the noncompleters. This finding is important; interventions targeting small children's irritability are central, as irritability in childhood predicts mental health problems, functional impairment, and outpatient treatment use in later life [29]. However, at the moment the number of studies on the topic is limited, especially studies on parent training for decreasing irritability [30]. Interestingly, noncompleters demonstrated greater improvement in affective reactivity compared with completers, and, at baseline, completers had more difficulties in affective reactivity. This may suggest that parents of children with greater initial difficulties were more motivated to engage with the program. It is also possible that this counterintuitive finding was because more severe difficulties are inherently harder to improve than milder ones. In this study, we also found decreases in child hyperactivity and conduct problems with small effect sizes.

Although effect sizes in universal programs are typically smaller than in targeted programs [31], our findings are encouraging and may reflect that universal internet-based interventions can improve child-specific mental health outcomes. In addition, internet-based parenting programs are effective for several parenting outcomes. A meta-analysis of parenting programs by Spencer et al [32] found no significant differences when comparing programs that included clinical support to programs that only contained internet-based components. Moreover, the parents reported high satisfaction with the interventions and felt that the interventions gave them confidence in their parenting capabilities, which is in line with previous studies in this field. Thus, considering the limited resources in society, internet-based parenting interventions may offer a wise way of supporting parenting.

Of the parents who started using the intervention, 36% completed it, forming an attrition rate of 64%. Previous studies on universal unguided parenting programs have reported similar attrition rates. For example, in the Australian internet-based self-directed parenting program ParentWorks, which included six sequenced modules, Dadds et al [13] found 68% attrition up until module 4 of the program. They defined full completion of the intervention as completing at least 5 core modules and the post-intervention survey. Among their 2967 parent participants, 8% were classified as full completers. In contrast, 17% of the participants in our study fully completed the intervention, resulting in a higher proportion of parents completing the entire program. One possible explanation for this discrepancy is that Dadds et al [13] only invited participants who had completed the core modules to fill out the postintervention survey, whereas in our study, the follow-up questionnaire was sent to all participants regardless of their completion status. ParentWorks was free of charge, and parents were recruited through a national media campaign and flyers distributed to child and family services. In another example, from the United States, which recruited parents from a well-child visit in primary care, Breitenstein et al [33] assessed a digitally delivered, self-administered parent training intervention, ezParent, and found that, out of the program's 6 modules, 33% of parents completed at least 4 modules. The completion rate is comparable to our study, in which 36% of parents who began the BePresent program completed at least 4 modules. This suggests that completion rates below 40% may be relatively typical for web-based modules, unguided universal parenting programs.

The main reasons for discontinuing the use of a universal parenting program include the parent being too busy or finding the program unhelpful [34]. In fact, a typical disadvantage of universal programs is their tendency to lose sufficient focus on the individualized needs of users, causing a modest matching of program goals with individual needs [35]. In our universal intervention, simple personalized and customized strategies were used, such as the possibility to choose individual strategies to complete the exercises. However, the main parts of the intervention were the same for everyone. Parenting programs that address families' actual needs facilitate program adherence [36], and personalized approaches could extend the program completion for those who currently do not benefit from such programs [37]. For example, in a recent pilot trial, Baumel et al [38] found that the program completion of an unguided digital parent training program aimed at treating child disruptive behaviors increased from 28% in a standard version (12/43 of

XSL•FO
RenderX

participating parents) to 69% (31/45 of participating parents) in an enhanced version, through the use of features such as intervention tailoring, timely reminders, automated monitoring, and feedback. BePresent had sequential modules, meaning that the modules were in a defined order and the order was the same for everyone.

Based on the findings of this study, some potential strategies to improve retention in universal internet-based parenting programs can be identified. First, motivation could potentially be increased and attrition rates decreased if programs allowed parents to choose the modules they felt would best benefit their situation and personalize the modules according to their needs. Second, tailored support, such as frequent and personalized reminders could motivate parents to use the program. These reminders could be based on parents' progress in the program or their specific needs. Third, it seems that the most common point for discontinuing the program was between modules 1 and 2, with a dropout rate of approximately 34%. Engagement should be especially supported here. Encouraging actions, especially at the beginning of the program might be highly beneficial. Universal parenting programs could benefit from evolving technology to provide more personalized approaches. For example, certain modules and different program lengths could be suggested based on parents' reports of their needs and the skills they already possess.

The current study and previous studies about universal parenting programs have shown short-term improvements in individual-level outcomes, but it is important to notice that universal interventions can also foster changes in public health at the population level, too. This means focusing on prevention at the population level, for example, by increasing knowledge or changing attitudes [39]. From a parent training perspective, universal programs can aim to prevent problems in parenting and in child behavior, and the results can be seen in the number of those who eventually seek treatment. However, when aiming for population-level changes, the reach of the intervention must be comprehensive enough. Universal interventions typically have low penetration rates, mainly because it is a family's own responsibility to seek out the intervention [35]. Our parent training program was offered when families attended a free of charge annual health check-up, something that all families in Finland are invited to participate in. Using this kind of setting, participation in the intervention was designed to be as easy as possible. Reach and easiness to participation should be carefully considered when implementing universal parenting programs. In addition, it would be important to also assess the population-level effects of universal parenting programs rather than focusing only on individual-level outcomes [39].

Limitations

The study has several limitations to consider. First, this was not a randomized controlled study; therefore, effectiveness and efficacy could not be assessed. Second, it was not possible for us to assess how many parents were offered the chance to participate in the program. The child health clinics were instructed to offer the program to all parents who fulfilled the inclusion criteria. Third, because this was an acceptability study, we were not aiming to reach the population level, so the reach of this study was limited. However, the program was studied in a wide geographical area, covering 67% of well-being services counties in Finland, making the study geographically highly representative. Fourth, the effect sizes seen in this study are low, reaching a Cohen *d* of 0.27, at the highest. The effect sizes in previous universal internet-based parenting interventions have been small to moderate [11]. Effect sizes of parent training trials based on clinical samples have typically been higher, including large effects [40]. One reason for the low effect size in our study was that the level of problems among the sample, taken from the general population, was less severe than that of clinical samples, indicating that there was less room for improvement. Fifth, the study did not ask participants their reasons for not completing the intervention; therefore, this could not be factored or taken into account in our analysis.

Conclusions

The universal internet-based parenting program BePresent may promote improvements in behavior and the daily lives of children. The completion rate was similar to that of other internet-based unguided universal parenting programs. In addition, parents were satisfied with the program and perceived that it supported their positive parenting skills. This type of program is low in cost, requires minimal resources, and can produce several benefits. Therefore, programs like this have the potentiality to be implemented widely in community settings to improve knowledge and positive parenting skills. However, there is a need to assess the impact of personalizing digital universal interventions to increase program adherence and to assess the efficacy of such programs in randomized controlled designs.

Acknowledgments

This project has received funding (ERC [European Research Council] Advanced; to AS) from the ERC under the European Union's Horizon 2020 research and innovation program (grant agreement 101020767). This research was supported by the INVEST Research Flagship Centre. The authors used OpenAI's ChatGPT for checking the spelling and grammar of the manuscript, after which the authors carefully checked the text and modified it. Artificial intelligence was not used to produce the content of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding authors on reasonable requests.

XSL•FO
RenderX

Conflicts of Interest

AB received grant funding from The Israel National Institute for Health Policy Research, National Institutes of Health, and the Israeli Scientific Foundation. He is providing consultancy services to ifeel online. AS is the CEO of Digifamilies Inc., which provides clinical consultation and digital interventions for specialized care. The remaining authors have no conflicts of interest to disclose.

Multimedia Appendix 1
TREND (Transparent Reporting of Evaluations with Nonrandomized Designs) Statement Checklist for complete and detailed reporting.
[PDF File (Adobe PDF File), 177 KB - jmir_v27i1e65391_app1.pdf]

References

1. Izett E, Rooney R, Prescott SL, De Palma M, McDevitt M. Prevention of mental health difficulties for children aged 0-3 years: a review. Front Psychol 2020;11:500361 [FREE Full text] [doi: 10.3389/fpsyg.2020.500361] [Medline: 34777074]
2. Jeong J, Franchett EE, Ramos de Oliveira CV, Rehmani K, Yousafzai AK. Parenting interventions to promote early child development in the first three years of life: a global systematic review and meta-analysis. PLoS Med 2021;18(5):e1003602 [FREE Full text] [doi: 10.1371/journal.pmed.1003602] [Medline: 33970913]
3. Scott S, Gardner F. Parenting programs. In: Thapar DS, Pine JF, Leckman S, Scott MJ, Snowling E, editors. Rutter's Child Adolescent Psychiatry. Hoboken: John Wiley & Sons; 2015.
4. Prevention of mental disorders : effective interventions and policy options : summary report / a report of the World Health Organization Dept. of Mental Health and Substance Abuse ; in collaboration with the Prevention Research Centre of the Universities of Nijmegen and Maastricht.World Health Organization. World Health Organization. 2004. URL: https://iris.who.int/handle/10665/43027 [accessed 2025-02-24]
5. Arango C, Díaz-Caneja CM, McGorry PD, Rapoport J, Sommer IE, Vorstman JA, et al. Preventive strategies for mental health. The Lancet Psychiatry 2018;5(7):591-604. [doi: 10.1016/s2215-0366(18)30057-9]
6. Singh V, Kumar A, Gupta S. Mental health prevention and promotion-a narrative review. Front Psychiatry 2022;13:898009 [FREE Full text] [doi: 10.3389/fpsyt.2022.898009] [Medline: 35958637]
7. Branco MSS, Altafim ERP, Linhares MBM. Universal intervention to strengthen parenting and prevent child maltreatment: updated systematic review. Trauma Violence Abuse 2022;23(5):1658-1676. [doi: 10.1177/15248380211013131] [Medline: 33973499]
8. de Oliveira RC, Altafim ERP, Gaspardo CM, Linhares MBM. Strengthening mother-child interactions among mothers and children with behavior problems at early development. Curr Psychol 2023:1-13 [FREE Full text] [doi: 10.1007/s12144-023-04786-9] [Medline: 37359573]
9. van Leuven L, Lalouni M, Forster M. "Will it work as well on Zoom?" a natural experiment during the Covid-19 pandemic of delivering parenting groups via video conferencing or in person. J Child Fam Stud 2023;32(1):67-80 [FREE Full text] [doi: 10.1007/s10826-022-02398-8] [Medline: 35965634]
10. Frost N, Dolan P. Theory, research and practice in child welfare: the current state of the art in social work. Child & Family Social Work 2021;26(3):498-506 [FREE Full text] [doi: 10.1111/cfs.12824]
11. Opie JE, Esler TB, Clancy EM, Wright B, Painter F, Vuong A, et al. Universal digital programs for promoting mental and relational health for parents of young children: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2024;27(1):23-52 [FREE Full text] [doi: 10.1007/s10567-023-00457-0] [Medline: 37917315]
12. Hall CM, Bierman KL. Technology-assisted interventions for parents of young children: emerging practices, current research, and future directions. Early Child Res Q 2015;33:21-32 [FREE Full text] [doi: 10.1016/j.ecresq.2015.05.003] [Medline: 27773964]
13. Dadds MR, Sicouri G, Piotrowska PJ, Collins DA, Hawes DJ, Moul C, et al. Keeping parents involved: predicting attrition in a self-directed, online program for childhood conduct problems. J Clin Child Adolesc Psychol 2019;48(6):881-893 [FREE Full text] [doi: 10.1080/15374416.2018.1485109] [Medline: 30067388]
14. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Lingley-Pottie P, et al. Internet-assisted parent training intervention for disruptive behavior in 4-year-old children: a randomized clinical trial. JAMA Psychiatry 2016;73(4):378-387. [doi: 10.1001/jamapsychiatry.2015.3411] [Medline: 26913614]
15. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Hinkka-Yli-Salomäki S, et al. Two-year follow-up of internet and telephone assisted parent training for disruptive behavior at age 4. J Am Acad Child Adolesc Psychiatry 2018;57(9):658-668.e1. [doi: 10.1016/j.jaac.2018.07.001] [Medline: 30196869]
16. Ristkari T, Kurki M, Suominen A, Gilbert S, Sinokki A, Kinnunen M, et al. Web-based parent training intervention with telephone coaching for disruptive behavior in 4-year-old children in real-world practice: implementation study. J Med Internet Res 2019;21(4):e11446 [FREE Full text] [doi: 10.2196/11446] [Medline: 30973337]

XSL•FO
RenderX

17. Sourander A, Ristkari T, Kurki M, Gilbert S, Hinkka-Yli-Salomäki S, Kinnunen M, et al. Effectiveness of an internet-based and telephone-assisted training for parents of 4-year-old children with disruptive behavior: implementation research. J Med Internet Res 2022;24(4):e27900 [FREE Full text] [doi: 10.2196/27900] [Medline: 35377332]
18. Sourander S, Sourander A, Hinkka-Yli-Salomäki S, Ristkari T, Kurki M. An internet-based parent training with telephone coaching on managing disruptive behavior in children at special family counseling centers during the COVID-19 pandemic: feasibility study. JMIR Pediatr Parent 2022;5(4):e40614 [FREE Full text] [doi: 10.2196/40614] [Medline: 36194895]
19. Finnish Institute for Healthwelfare. Health examinations at maternity and child healt clinics in 2022 and in school health care in school year 2022?2023. URL: https://thl.fi/tilastot-ja-data/tilastot-aiheittain/terveyspalvelut/terveystarkastukset-aitiys-ja-lastenneuvoloissa-seka-opiskeluhuollossa [accessed 2024-08-07]
20. Finlex. Finnish Health Care Act Terveydenhuoltolaki 1326/2010. 2010. URL: https://www.finlex.fi/fi/laki/smur/2010/20101326 [accessed 2025-02-25]
21. school and student health care services and preventive oral health care services [Valtioneuvoston asetus neuvolatoiminnasta, koulu- ja opiskeluterveydenhuollosta sekä lasten ja nuorten ehkäisevästä suun terveydenhuollosta. Decree of Government on maternity and child health clinics. URL: https://www.finlex.fi/fi/laki/smur/2011/20110338 [accessed 2025-02-05]
22. Goodman R. The strengths and difficulties questionnaire: a research note. J Child Psychol Psychiatry 1997;38(5):581-586. [doi: 10.1111/j.1469-7610.1997.tb01545.x] [Medline: 9255702]
23. Goodman R, Scott S. Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful? J Abnorm Child Psychol 1999;27(1):17-24. [doi: 10.1023/a:1022658222914] [Medline: 10197403]
24. Hajjar STEL. Statistical analysis: internal-consistency reliability and construct validity. 2018 IJQQRM 2018;6(1):46-57.
25. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 2025;16(3):297-334. [doi: 10.1007/BF02310555]
26. Stringaris A, Goodman R, Ferdinando S, Razdan V, Muhrer E, Leibenluft E, et al. The Affective Reactivity Index: a concise irritability scale for clinical and research settings. J Child Psychol Psychiatry 2012;53(11):1109-1017 [FREE Full text] [doi: 10.1111/j.1469-7610.2012.02561.x] [Medline: 22574736]
27. Wilson MK, Cornacchio D, Brotman MA, Comer JS. Measuring irritability in early childhood: a psychometric evaluation of the affective reactivity index in a clinical sample of 3- to 8-year-old children. Assessment 2022;29(7):1473-1481 [FREE Full text] [doi: 10.1177/10731911211020078] [Medline: 34060361]
28. Barkley RA, Murphy KR. Home situations questionnaire. In: Attention-Deficit Hyperactivity Disorder: A Clinical Workbook, 2nd Edition. New York City, New York: Guilford Press; 1998.
29. Dougherty LR, Smith VC, Bufferd SJ, Kessel E, Carlson GA, Klein DN. Preschool irritability predicts child psychopathology, functional impairment, and service use at age nine. J Child Psychol Psychiatry 2015;56(9):999-1007 [FREE Full text] [doi: 10.1111/jcpp.12403] [Medline: 26259142]
30. Fongaro E, Picot MC, Stringaris A, Belloc C, Verissimo AS, Franc N, et al. Parent training for the treatment of irritability in children and adolescents: a multisite randomized controlled, 3-parallel-group, evaluator-blinded, superiority trial. BMC Psychol 2022;10(1):273 [FREE Full text] [doi: 10.1186/s40359-022-00984-5] [Medline: 36414963]
31. Fisak B, Penna A, Mian ND, Lamoli L, Margaris A, Cruz SAMFD. The effectiveness of anxiety interventions for young children: a meta-analytic review. J Child Fam Stud 2023:1-12 [FREE Full text] [doi: 10.1007/s10826-023-02596-y] [Medline: 37362628]
32. Spencer CM, Topham GL, King EL. Do online parenting programs create change?: a meta-analysis. J Fam Psychol 2020;34(3):364-374. [doi: 10.1037/fam0000605] [Medline: 31697102]
33. Breitenstein SM, Fehrenbacher C, Holod AF, Schoeny ME. A randomized trial of digitally delivered, self-administered parent training in primary care: effects on parenting and child behavior. J Pediatr 2021;231:207-214.e4 [FREE Full text] [doi: 10.1016/j.jpeds.2020.12.016] [Medline: 33338496]
34. Fletcher R, Regan C, Dizon J, Leigh L. Understanding attrition in text-based health promotion for fathers: survival analysis. JMIR Form Res 2023;7:e44924 [FREE Full text] [doi: 10.2196/44924] [Medline: 37594788]
35. Dodge KA. Annual research review: universal and targeted strategies for assigning interventions to achieve population impact. J Child Psychol Psychiatry 2020;61(3):255-267. [doi: 10.1111/jcpp.13141] [Medline: 31643089]
36. Koerting J, Smith E, Knowles MM, Latter S, Elsey H, McCann DC, et al. Barriers to, and facilitators of, parenting programmes for childhood behaviour problems: a qualitative synthesis of studies of parents' and professionals' perceptions. Eur Child Adolesc Psychiatry 2013;22(11):653-670 [FREE Full text] [doi: 10.1007/s00787-013-0401-2] [Medline: 23564207]
37. McKay K, Kennedy E, Senior R, Scott S, Hill J, Doolan M, et al. Informing the personalisation of interventions for parents of children with conduct problems: a qualitative study. BMC Psychiatry 2020;20(1):513 [FREE Full text] [doi: 10.1186/s12888-020-02917-1] [Medline: 33081742]
38. Baumel A, Brandes O, Brendryen H, Muench F, Kane JM, Saar C. The impact of therapeutic persuasiveness on engagement and outcomes in unguided interventions: a randomized pilot trial of a digital parent training program for child behavior problems. Internet Interv 2023;34:100680 [FREE Full text] [doi: 10.1016/j.invent.2023.100680] [Medline: 37840647]
39. Greenberg MT, Abenavoli R. Universal interventions: fully exploring their impacts and potential to produce population-level impacts. Journal of Research on Educational Effectiveness 2016;10(1):40-67. [doi: 10.1080/19345747.2016.1246632]

XSL•FO
RenderX

40. Lundahl B, Risser HJ, Lovejoy MC. A meta-analysis of parent training: moderators and follow-up effects. Clin Psychol Rev 2006;26(1):86-104. [doi: 10.1016/j.cpr.2005.07.004] [Medline: 16280191]

Abbreviations

TREND: Transparent Reporting of Evaluations with Nonrandomized Designs

Edited by A Mavragani; submitted 29.08.24; peer-reviewed by M Ambrosio, K Ayers; comments to author 28.11.24; revised version received 22.01.25; accepted 23.01.25; published 13.03.25.

Please cite as:
Mishina K, Baumel A, Kinnunen M, Ristkari T, Heinonen E, Hinkka-Yli-Salomäki S, Sourander A
BePresent Universal Internet-Based Parenting Intervention: Single-Arm Pre-Post Intervention Study
J Med Internet Res 2025;27:e65391
URL: https://www.jmir.org/2025/1/e65391
doi:10.2196/65391
PMID:

©Kaisa Mishina, Amit Baumel, Malin Kinnunen, Terja Ristkari, Emmi Heinonen, Susanna Hinkka-Yli-Salomäki, Andre Sourander. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Profiling Generalized Anxiety Disorder on Social Networks: Content and Behavior Analysis

Linah Alhazzaa[1,2], MSc; Vasa Curcin[3], PhD

[1]Department of Informatics, King's College London, London, United Kingdom

[2]Department of Computer Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[3]Department of Population Health Sciences, King's College London, London, United Kingdom

Corresponding Author:
Linah Alhazzaa, MSc
Department of Informatics
King's College London
Strand
London, WC2R 2LS
United Kingdom
Phone: 44 2078365454
Email: linah.alhazzaa@kcl.ac.uk

Abstract

Background: Despite a dramatic increase in the number of people with generalized anxiety disorder (GAD), a substantial number still do not seek help from health professionals, resulting in reduced quality of life. With the growth in popularity of social media platforms, individuals have become more willing to express their emotions through these channels. Therefore, social media data have become valuable for identifying mental health status.

Objective: This study investigated the social media posts and behavioral patterns of people with GAD, focusing on language use, emotional expression, topics discussed, and engagement to identify digital markers of GAD, such as anxious patterns and behaviors. These insights could help reveal mental health indicators, aiding in digital intervention development.

Methods: Data were first collected from Twitter (subsequently rebranded as X) for the GAD and control groups. Several preprocessing steps were performed. Three measurements were defined based on Linguistic Inquiry and Word Count for linguistic analysis. GuidedLDA was also used to identify the themes present in the tweets. Additionally, users' behaviors were analyzed using Twitter metadata. Finally, we studied the correlation between the GuidedLDA-based themes and users' behaviors.

Results: The linguistic analysis indicated differences in cognitive style, personal needs, and emotional expressiveness between people with and without GAD. Regarding cognitive style, there were significant differences (P<.001) for all features, such as insight (Cohen d=1.13), causation (Cohen d=1.03), and discrepancy (Cohen d=1.16). Regarding personal needs, there were significant differences (P<.001) in most personal needs categories, such as curiosity (Cohen d=1.05) and communication (Cohen d=0.64). Regarding emotional expressiveness, there were significant differences (P<.001) for most features, including anxiety (Cohen d=0.62), anger (Cohen d=0.72), sadness (Cohen d=0.48), and swear words (Cohen d=2.61). Additionally, topic modeling identified 4 primary themes (ie, symptoms, relationships, life problems, and feelings). We found that all themes were significantly more prevalent for people with GAD than for those without GAD (P<.001), along with significant effect sizes (Cohen d>0.50; P<.001) for most themes. Moreover, studying users' behaviors, including hashtag participation, volume, interaction pattern, social engagement, and reactive behaviors, revealed some digital markers of GAD, with most behavior-based features, such as the hashtag (Cohen d=0.49) and retweet (Cohen d=0.69) ratios, being statistically significant (P<.001). Furthermore, correlations between the GuidedLDA-based themes and users' behaviors were also identified.

Conclusions: Our findings revealed several digital markers of GAD on social media. These findings are significant and could contribute to developing an assessment tool that clinicians could use for the initial diagnosis of GAD or the detection of an early signal of worsening in people with GAD via social media posts. This tool could provide ongoing support and personalized coping strategies. However, one limitation of using social media for mental health assessment is the lack of a demographic representativeness analysis.

(J Med Internet Res 2025;27:e53399) doi:10.2196/53399

XSL•FO
RenderX

KEYWORDS

generalized anxiety disorder; mental health; Twitter; social media analysis; natural language processing

Introduction

Background

Mental health disorders are increasingly common, affecting approximately 1 billion people worldwide [1]. Among these mental health disorders, anxiety disorders are prevalent, with 3.94% of the world's population affected in 2019 [2]. It is important to note that generalized anxiety disorder (GAD) is the most common anxiety disorder. According to the 2014 UK Adult Psychiatric Morbidity Survey, 5.9% of adults experienced GAD [3]. According to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) [4], GAD is characterized by extreme and persistent worry and anxiety about everyday activities lasting at least 6 months. Symptoms can be both mental and physical, including unrealistic anxiety, inability to relax, restlessness, headaches, rapid heart rate, and difficulty breathing [4].

Untreated GAD, in the long term, may lead to physical and mental health complications [5]. It also reduces the quality of life by affecting work productivity and physical and social functions [6]. In addition, GAD burdens health care systems and the economy. It increases health care costs due to higher use of services and decreases work productivity, leading to more days of disability [7]. Despite the availability of effective GAD treatments, only 43.2% of those who have it receive treatment [8]. Consequently, early detection of GAD could significantly reduce all these burdens and improve quality of life.

People use technology in nearly every aspect of their lives. Social media has become more popular, reaching 4.8 billion users worldwide in 2023 [9]. Individuals often share their mental health status on social media as a way of reducing isolation, seeking support, overcoming stigma, increasing awareness, coping with challenges, and sharing experiences without the feeling of being unfairly judged [10]. This generates valuable data on users' posts, profiles, and web-based activities that can provide insights into their mental health conditions. Therefore, exploiting data extracted from social networks is essential to revealing users' mental health statuses.

Consequently, this study presented an approach to identifying GAD markers via social media. By analyzing both linguistic and behavioral patterns on Twitter (subsequently rebranded as X), this approach constructed a comprehensive digital profile of individuals with GAD, distinguishing them from the control group. The methodology integrated analytical tools, such as Linguistic Inquiry and Word Count (LIWC; Pennebaker Conglomerates, Inc) for detailed linguistic analysis and GuidedLDA for thematic exploration, yielding a multidimensional understanding of language and topic patterns unique to GAD. Moreover, a thorough behavioral analysis examined hashtag use frequency, follower counts, interaction patterns, and other aspects, identifying GAD-associated behaviors. This integrated approach highlights social media's potential as a mental health assessment resource and support by identifying 47 statistically validated digital markers of GAD.

In summary, our primary contributions are (1) characterizing the digital markers of GAD by analyzing social media content and behaviors of those who have it, by identifying distinct patterns of language use, thematic topics, and behavioral traits that differentiated them from individuals in the control group; (2) conducting an extensive content-based analysis that leveraged LIWC for a detailed linguistic-based analysis and GuidedLDA for a comprehensive thematic analysis; (3) identifying significant behavioral differences based on user and tweet-level features, such as hashtag use frequency and interaction patterns with other users, as well as measuring the correlation between the themes derived from GuidedLDA and various metadata attributes for both the GAD and control groups; and (4) presenting our findings on 47 distinct GAD-associated digital markers through comprehensive statistical analysis.

Related Works

The widespread use of social media networks and the availability of massive data inspire researchers to investigate the characteristics of mental illnesses through user-generated content and users' activities on social media platforms. Defining these characteristics helps detect mental illnesses via social media, which could help psychiatrists assess a patient's diagnosis and recovery progress. Many studies have used natural language processing (NLP) and machine learning techniques for mental illness prediction via social networks. Several studies have focused mainly on depression prediction via social networks [11-24], and some have examined anxiety disorders [25-30], posttraumatic stress disorder [31-33], and eating disorders [15,34-37]. Therefore, identifying digital markers of mental health conditions could improve detection accuracy.

Textually generated content on social media platforms has been studied to identify indicators of mental health problems. Several studies in psycholinguistics have explored the link between language use and psychological health. Most of these works have used LIWC, a text analysis program that uses predefined dictionaries to capture various dimensions, such as social, emotional, and psychological states. Many studies have analyzed word use to extract psycholinguistic lexical features for identifying depression [11,12,18,25,33,38-40], posttraumatic stress disorder [31,33,40], bipolar disorder [25,33,40], eating disorders [25,34,35], seasonal affective disorder [25,33,40], anxiety disorders [25,29], obsessive-compulsive disorder [25], and schizophrenia [25,41].

A few works have studied the word use associated with psychological health in different languages to identify mental illnesses. The study by Hao et al [28] used a Chinese version of LIWC to investigate psychological characteristics from posts collected from Sina Weibo, a popular Chinese social media platform. In the study by Peng et al [42], they extracted emotional features for depression based on the Chinese Affective Lexicon Ontology, a Chinese emotional dictionary, by adding the depression class to other emotional classes in the emotional dictionary and excluding the surprise class because it is unrelated to depression.

XSL•FO
RenderX

Several studies have analyzed textual content using topic modeling techniques to reveal hidden topics in social media posts and define their relevance to mental health disorders. Latent Dirichlet allocation (LDA) is a widely used topic modeling technique in previous research, including studies on depression [11,18], anxiety disorders [29,30], anorexia [34], schizophrenia [41], and a set of mental health disorders [26].

Because of the significance of leveraging social interaction metadata from users with mental health conditions, some studies have been conducted to understand how individuals with a low mental health status interact on social media platforms. Therefore, studying users' behaviors could provide insights into their mental health characteristics. A survey [43] investigated the correlation between web behaviors and psychological health, concluding that some web activities, such as internet addiction, instant messaging, and web-based gaming, are highly related to mental health disorders.

In the context of social media, previous research has examined social media users' activities as an indicator of their mental health status. Vedula and Parthasarathy [19] studied the behaviors of people with depression on Twitter, whereas Peng et al [42] analyzed the depressive characteristics of Sina Weibo users based on posts and user profiles. In the study by de Choudhury et al [44], they investigated changes in new mothers' characteristics related to mood and behavior in the period following childbirth based on some measures of linguistic style, emotional state, and ego network (ie, the number of followers and followees, as well as engagement with the Twitter community). The work by Wang et al [35] characterized individuals with eating disorders via Twitter based on their social interactions, behaviors, and psycholinguistic properties.

It is worth noting that using social media data for mental health assessment offers valuable insights. However, it also presents ethical considerations. Privacy is the key ethical concern when using publicly available data [45,46]. Therefore, to maintain users' privacy, it is essential to apply anonymization techniques to the datasets collected from social media. This process involves removing all information that could reveal a person's identity.

Although previous research has conducted content analysis of users with anxiety disorders via social media, a limited number of works have focused mainly on GAD. Therefore, a thorough study of GAD characteristics on social media platforms is needed. Accordingly, this study provided an in-depth and detailed profiling of users with GAD via social media. We conducted a content analysis, including linguistic and thematic analysis, the web-based behavior of users with GAD, and the correlation between themes and users' behaviors. These analyses could contribute to a better understanding of and provide insights into GAD markers, helping extract significant features used in GAD detection via social media.

The rest of this paper is organized as follows. The methodology is outlined in the next section. The results of our work are then presented. The last section highlights the findings and discusses the limitations and possible future directions to extend our research.

Methods

Overview

To identify GAD characteristics in social media users, we performed several steps. The first step was collecting data using the Twitter application programming interface (API) [47] to obtain posts from both GAD and control groups. Second, a set of data-cleaning procedures was applied. Third, we investigated the dataset based on different aspects, including linguistic analysis, thematic analysis, and user behavior analysis.

LIWC was used for the linguistic-based analysis, whereas the thematic analysis was based on the GuidedLDA algorithm. The user- and tweet-level metadata were used to analyze user behaviors, and several features were extracted from tweets and user profiles. We then explored the results to shed light on GAD markers derived from the content and behavior analysis. Figure 1 illustrates the analytical workflow of GAD profiling via social media. The subsequent sections provide an in-depth explanation of each step.

XSL•FO
RenderX

Figure 1. The analytical workflow of generalized anxiety disorder (GAD) profiling via social media involved collecting data from Twitter for the GAD and control groups; performing a set of data preprocessing steps; and conducting a thorough analysis, including behavior-based and content-based analysis. API: application programming interface; DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; LIWC: Linguistic Inquiry and Word Count.

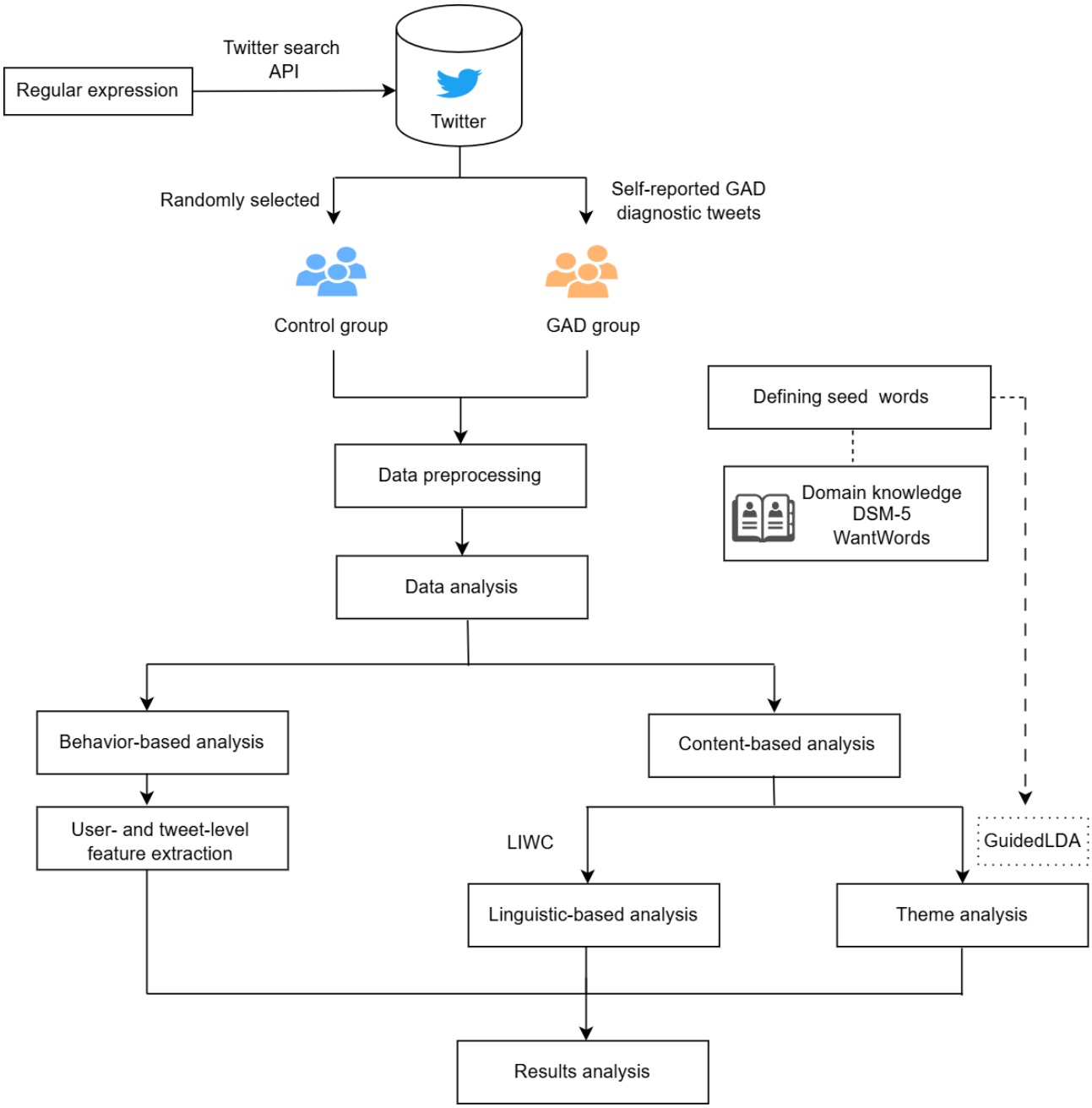

Data Collection

The data were collected from Twitter, a widely used social media platform. The reason for choosing Twitter was its popularity as a social media platform, with approximately 369 million active accounts every month worldwide in 2022 [48]. Another reason was that Twitter offers a significant amount of data for researchers on health topics because of the enormous number of users, its data accessibility, and the lack of difficulty retrieving and searching data [49].

The Twitter platform provides an API for pooling public posts, searching for relevant ones using regular expressions, and streaming real-time tweets. Through the Twitter API, we collected a dataset encompassing users with GAD and users comprising the control population. Identifying individuals with GAD was done by following the method proposed by Coppersmith et al [40]. This method uses a regular expression to search for tweets that include the diagnosis statement for a condition, such as "I was diagnosed with [condition name]," using the Twitter API. In our study, a group of users with GAD was selected using a regular expression to search tweets that contained the following statement—"I was diagnosed with Generalized Anxiety Disorder (GAD)"—to ensure that a user had publicly stated that they had been diagnosed with GAD. This resulted in 3445 tweets from 2721 unique potential users in the GAD group.

The Twitter API returned tweets comprising all words that matched the given expression, requiring a manual verification

XSL•FO
RenderX

procedure to eliminate tweets unrelated to the GAD diagnosis statement. This process ensured that the collected tweets specifically pertained to a self-reported diagnosis of GAD of users who posted them, so a manual review was conducted to ensure that the tweets did not mention quotations or stories of someone else with GAD. As a result, of the 2721 unique potential users, 1500 (55.13%) users with GAD were included in the dataset.

The control group consisted of randomly selected users representing a general population sample. This resulted in 1847 control group users being included in the dataset. In instances in which users were found to be present in both the GAD and control groups, they were excluded from the control group. The most recent tweets for users with GAD and control users were then retrieved considering the limitation of retrieving a maximum of 3200 tweets per user imposed by the Twitter API. Figure 2 illustrates the process of collecting the dataset.

As shown in Tables 1 and 2, the user and tweet metadata were also retrieved for each profile to study the users' behaviors and social engagement. It is worth noting that this study only collected data from users whose accounts were public, and all the collected data were deidentified to protect privacy.

Figure 2. The data collection process from Twitter for the generalized anxiety disorder (GAD) and control groups. It encompassed searching for a target group, verifying potential users, and retrieving their tweets. API: application programming interface.

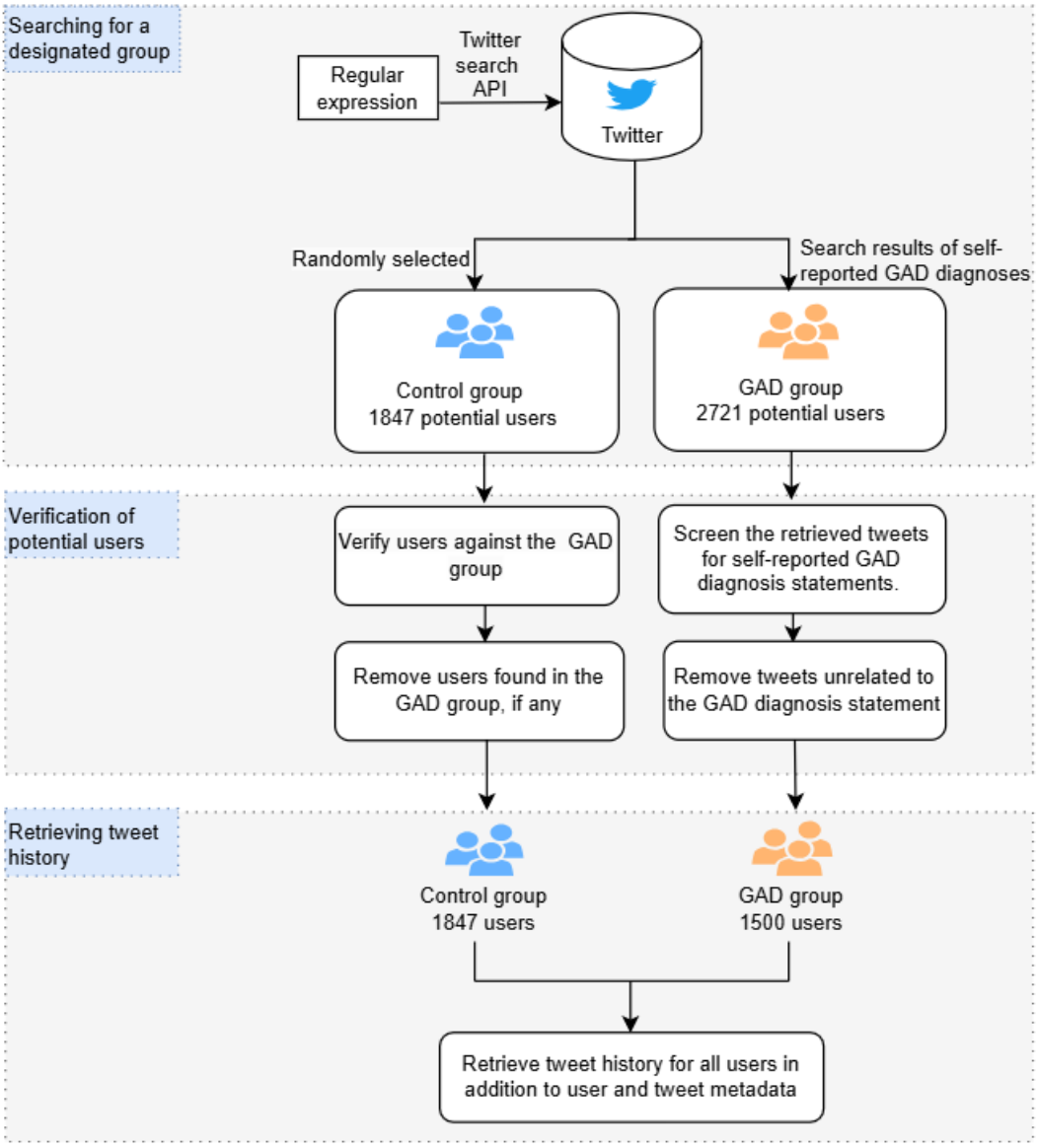

XSL•FO
RenderX

Table 1. A description of each type of user-level metadata [50] that was retrieved from Twitter for each user in both the generalized anxiety disorder and control groups.

Feature	Description
User ID	A unique ID for a particular user
Status count	It shows how many tweets were posted by a user.
Description	The user biography briefly describes the user's account.
Verified	It shows whether the user's account is verified.
Account creation date	It shows the date and time of the account's creation.
Follower count	The total amount of users who follow an account
Friend count	The number of users that are currently being followed by an account
Listed count	The number of public lists that a user has joined
Favorite count	The total number of tweets that have been marked as favorites

Table 2. A description of tweet-level metadata [51] that were retrieved from Twitter for each user in both the generalized anxiety disorder and control groups.

Feature	Description
Tweet ID	A unique ID for a tweet
Full tweet	The tweet text
Tweet created at	The date and time when a tweet was posted
Favorite count	The number of likes for a tweet
Retweet count	The number of retweets for a tweet
Reply to user ID	If a tweet is a response to another user's tweet, this attribute will show that user's unique identifier.

Data Preprocessing

Overview

Several preprocessing steps were undertaken to prepare the 2 collections of raw tweets for further analysis. Data transformation was applied first. We transformed the raw data (tweets) into a structured format in which the columns were the user ID, full tweet, and favorite count, among other data, as described in Tables 1 and 2. Therefore, each row included 1 tweet for a user ID. These tweets were then further processed as outlined in the following sections.

Filtering Out

As this study focused on analyzing posts in English, the dataset was filtered to exclude users with <80% of tweets in English. We then removed non–English-language tweets. Users with <100 tweets were excluded from the dataset. This resulted in a final dataset with a total of 7,313,769 tweets posted by 2779 users. Of these 2779 users, 1316 (47.36%) had been diagnosed with GAD and had posted a total of 46.04% (3,367,581/7,313,769) of the tweets, whereas 1463 (52.64%) of the users constituted the control group with a total of 53.96% (3,946,188/7,313,769) of the tweets. The average number of tweets was 2558.95 (SD 1315.98) for each individual with GAD, whereas it was 2697.33 (SD 918.30) for each user in the control group. Figure 3 illustrates the statistics of collected data for the GAD and control groups, showing the number of users and tweets for each group.

HTML tags and URLs were also eliminated. As part of the preprocessing step, tweets were converted into lower case to prevent case sensitivity in the subsequent data-cleaning steps. Furthermore, identifiable usernames or posts were eliminated to protect users' privacy.

The following data preprocessing steps were standard and have been applied in previous NLP-related studies [52,53].

XSL•FO
RenderX

Figure 3. The statistics for the collected data are presented for the generalized anxiety disorder (GAD) and control groups showing the number of users and tweets for both groups.

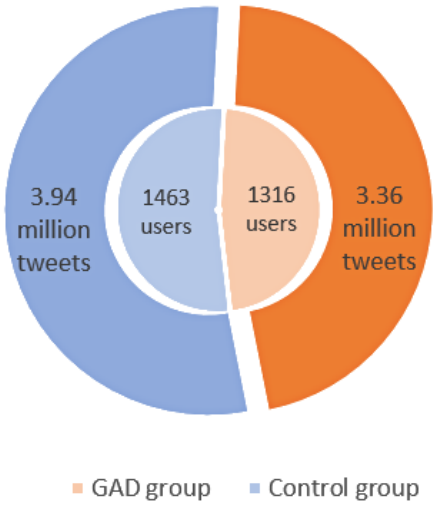

Tokenization

Tokenization is applied to split tweets into a list of tokens (ie, words) using the Natural Language Toolkit (NLTK; Team NLTK) tokenizer. The output of this process becomes the main input for punctuation and stop word removal, word normalization, and part-of-speech (POS) tagger steps.

Punctuation and Stop Word Removal

Stop words, such as articles and prepositions, are widely used in any natural language, but they are commonly ignored in text analysis because of their low meaning value and high frequency. Thus, we removed them from the data using the standard stop word list provided in the NLTK. Punctuation was also removed as it adds noise to the dataset. This step was conducted using regular expressions and the NLTK library.

POS Tagger

The POS tagger was applied to assign each token the appropriate POS, such as noun, verb, adverb, and adjective. This process enhances the quality of preprocessing. The POS must be specified for every token using the NLTK POS tagger to obtain the correct lemma.

Lemmatization

As words have different forms, it is essential to unify the word representation to analyze the text. Applying word normalization techniques eliminates the different forms of the same word. Lemmatization is a technique that reduces words to their base form (lemma) using a corpus, such as WordNet. We applied lemmatization to reduce each token to its root using the NLTK lemmatizer based on the WordNet dictionary (The Trustees of Princeton University).

LIWC for Text Analysis

Overview

Several studies have revealed a correlation between language use and an individual's mental health status. Therefore, we investigated the differences in word use patterns between users with GAD and control users to identify underlying potential indications of GAD. We used LIWC [54], a text analysis tool equipped with >100 built-in dictionaries to reveal individuals' psychological, social, and behavioral patterns from text, each including words, word stems, and emoticons [54]. These words could be classified into one or multiple categories or dictionaries.

The process of text analysis using LIWC involves several steps. These steps are reading the dataset, comparing every word with each dictionary entry, and calculating the percentage of words related to each built-in dictionary. A quantitative summary of all LIWC dimensions is then generated, and the user-level means for all dimensions are found. Finally, these results could be analyzed to capture underlying psychosocial indicators.

As mental health impacts our way of thinking, behaving, and feeling [1], we proposed 3 measurements based on LIWC dictionaries. These measurements, namely, cognitive style, personal needs, and emotional expressiveness, aimed to study the differences in language use between users with GAD and control users. We carefully defined these measures inspired by the method in the studies by Pavalanathan and de Choudhury [55] and Ernala et al [56]. This method is based on leveraging insights from an extensive review and synthesis of previous research in psychology and psycholinguistics [57-59]. The potential implications of studying linguistic style are significant as it could provide a deeper understanding of the linguistic markers of GAD and potentially lead to more effective interventions.

Cognitive Style

The cognitive style quantifies how individuals think, talk about their thinking, or retrieve information. It is based on the cognitive dimension of LIWC to measure related categories, including insight, causation, discrepancy, tentativeness, certitude, and differentiation. In addition, the memory category is used to help measure the retrieval or forgetfulness of information. The results of these analyses could reveal a decline in cognition associated with poor mental health status in an individual with GAD, including memory recall issues, difficulty reasoning, and decision-making inability.

XSL•FO
RenderX

Personal Needs

Identifying personal needs helps maintain mental well-being, which could be helpful to separate the GAD group from the control group as the former may not sufficiently satisfy various personal needs. Therefore, we investigated some personal needs, including essential needs and self-esteem, based on LIWC dictionaries. The essential needs involve being physically and mentally healthy and feeling safe, including seeking support and maintaining a healthy life, which was assessed using some LIWC categories (ie, *food*, *substances*, *death*, *want*, *need*, and *lack*).

Self-esteem needs involve belief in personal value and ability, which also impacts how to communicate with others. Accordingly, measuring self-esteem needs includes several categories (ie, *reward*, *curiosity*, *risk*, *interpersonal conflict*, and *communication*). The results of these categories could reflect mental well-being based on the definition of mental well-being according to the World Health Organization, which is a person's ability to recognize their skills, deal with everyday life stressors, effectively produce and accomplish, and participate in the community [60].

Emotional Expressiveness

Emotional expressiveness measurement assesses a person's tendency to express various emotions through language use. We used the *negative tone*, *anxiety*, *anger*, *sadness*, and *swear words* categories to measure emotional expressiveness, which could identify the patterns of expressing emotional sentiments that are common among people with GAD.

GuidedLDA

Topic modeling is a content analysis approach for uncovering hidden themes in textual data. One common topic modeling technique is LDA, an unsupervised machine learning mechanism. LDA helps extract topics with no advanced knowledge of themes. However, implementing LDA on a large dataset may generate meaningless topics [61]. To leverage previous research and the domain knowledge of GAD in topic modeling, we adopted the GuidedLDA technique for extracting the common topics discussed among people with GAD.

GuidedLDA is a semisupervised LDA algorithm based on setting predefined seed words to extract hidden themes of a particular direction and interest. Unlike the LDA topics initialized using randomly selected words, each seed word has a high probability of belonging to its particular topic in the initialization stage of GuidedLDA. We defined some possible seed words for each candidate topic related to GAD by relying on domain knowledge and previous research on GAD. We explain the configuration of GuidedLDA in the *Thematic Analysis* section.

Twitter Metadata

In addition to retrieving the user-level and tweet-level metadata shown in Tables 1 and 2, we extracted several features derived from Twitter metadata. These included tweet and behavior-based features, as shown in Table 3.

XSL•FO
RenderX

Table 3. A set of extracted features derived from Twitter metadata.

Feature	Description
FF[a] ratio	The ratio of followers to followees per user
Number of mentions	The average number of mentions per tweet
Number of tweets	The number of tweets per user across their history
Number of replies	The number of replies per user across their history
Reply-to-tweet ratio	The ratio of replies to the total number of tweets per user
Number of hashtags	The number of participating hashtags per user across their history
Hashtag ratio	The ratio of tweets including hashtags to the total number of tweets per user
Number of retweets	The number of retweets per user across their history
Retweet-to-tweet ratio	The ratio of retweets to the total number of tweets per user
Total number of mentions per user	The number of mentions per user across their history
Total number of mentions without replies per user	The number of mentions per user without considering mentions associated automatically with replies across their history
Mention ratio	The ratio of tweets with other people mentioned to the total number of tweets per user
Total tweets with favorites	The total number of tweets that received at least one favorite per user
Tweets with favorites ratio	The ratio of total tweets that received at least one favorite to the total tweets per user
Total tweets with retweets	Total tweets that received at least one retweet per user
Tweets with retweets ratio	The ratio of total tweets that received at least one retweet to the total tweets per user
Tweets per day	The average number of tweets per day per user
Replies per day	The average number of replies per day per user
Retweet per day	The average number of retweets per day per user
Mentions per day	The average number of tweets including mentions per day per user
Hashtag engagement per day	The average number of tweets including hashtags per day per user

[a]FF: follower-to-followee.

Data Analysis

We aimed to compare various features between individuals with GAD and control individuals. Therefore, the analysis involved several statistical methods, including the 1-tailed, 2-sample *t* test; *P* value for statistical significance; and Cohen *d* for effect size. First, descriptive statistics were calculated for each content-based and behavioral-based feature for both groups, including means and SDs to summarize the data.

Next, we conducted the 1-tailed ,2-sample *t* test to compare the means of the 2 groups. This test evaluated whether the mean differences between these groups were statistically significant, with a *P* value of ≤.05 indicating statistical significance. The Cohen *d* was also computed to measure the effect size to provide insights into the practice effect. The possible magnitude of these differences ranged from small to large. According to Cohen [62] and Sawilowsky [63], the effect sizes are interpreted as small if Cohen $d<0.20$, as medium if the Cohen *d* is between 0.20 and 0.50, and as large if the Cohen *d* is between 0.50 and 0.80. The *P* value and Cohen *d* provide insights into the results' statistical significance and practical importance.

Ethical Considerations

This study used publicly available data without direct user interaction or intervention, so an ethics review was unnecessary. All usernames, IDs, and tweets were de-identified to ensure privacy. In addition, data that could disclose personal identifiers were carefully omitted to maintain user privacy and anonymity.

Results

This section presents the results of exploring the data in terms of linguistic, thematic, and behavior analyses.

Linguistic-Based Analysis

Overview

We found differences in language use across different aspects, including cognitive style, personal needs, and emotional expressiveness, between the individuals in the GAD group and those in the control group. Table 4 provides the summary statistics of these differences based on the LIWC categories for people with GAD compared to those without GAD. Consequently, these differences in the LIWC categories demonstrate a correlation between language use and the mental health status of people with GAD.

XSL•FO
RenderX

Table 4. A comparison of linguistic-based measurements between the generalized anxiety disorder (GAD) and control groups based on the Linguistic Inquiry and Word Count categories[a].

Category	GAD group, mean (SD)	Control group, mean (SD)	*P* value	Cohen *d*
Cognitive style (percentage of total word count)				
Insight	1.59 (0.57)	0.90 (0.65)	<.001	1.13
Causation	1.028 (0.39)	0.60 (0.44)	<.001	1.03
Discrepancy	1.39 (0.47)	0.78 (0.58)	<.001	1.16
Tentativeness	1.40 (0.48)	0.74 (0.56)	<.001	1.27
Certitude	0.57 (0.29)	0.30 (0.31)	<.001	0.90
Differentiation	2.15 (0.71)	1.17 (0.88)	<.001	1.23
Memory	0.10 (0.07)	0.07 (0.09)	<.001	0.37
Personal needs (percentage of total word count)				
Need	0.35 (0.18)	0.20 (0.18)	<.001	0.83
Want	0.38 (0.25)	0.23 (0.21)	<.001	0.65
Lack	0.11 (0.20)	0.06 (0.07)	<.001	0.33
Reward	0.11 (0.22)	0.15 (0.46)	.003	0.11
Risk	0.19 (0.12)	0.11 (0.12)	<.001	0.67
Curiosity	5.79 (1.31)	3.82 (2.30)	<.001	1.05
Substances	0.05 (0.17)	0.03 (0.06)	<.001	0.16
Food	0.48 (0.33)	0.32 (0.44)	<.001	0.41
Death	0.19 (0.19)	0.11 (0.13)	<.001	0.49
Interpersonal conflict	0.27 (0.17)	0.18 (0.19)	<.001	0.50
Communication	3.50 (1.64)	4.80 (2.35)	<.001	0.64
Emotional expressiveness (percentage of total word count)				
Negative tone	4.39 (1.36)	4.24 (1.73)	.01	0.10
Anxiety	0.18 (0.26)	0.06 (0.08)	<.001	0.62
Anger	0.17 (0.12)	0.09 (0.10)	<.001	0.72
Sadness	0.22 (0.22)	0.13 (0.15)	<.001	0.48
Swear words	0.77 (0.74)	0.47 (0.72)	<.001	2.61

[a]*P* values were calculated using a 1-tailed, 2-sample *t* test, with $P \leq .05$ indicating statistical significance, and the effect size is reported using the Cohen *d*.

Differences in Cognitive Style

The GAD group had higher values for all categories used for assessing cognitive style than individuals in the control group, showing statistically significant differences between the 2 groups ($P<.001$) for insight (Cohen $d=1.13$), causation (Cohen $d=1.03$), discrepancy (Cohen $d=1.16$), tentativeness (Cohen $d=1.27$), certitude (Cohen $d=0.90$), differentiation (Cohen $d=1.23$), and memory (Cohen $d=0.37$), as shown in Table 4. Tweets posted by users with GAD included more certainty, discrepancy, and differentiation words, almost 2 times more than tweets posted by users in the control group. In addition, their tweets contained nearly twice as many tentative words as the control group. The results also indicated that the GAD group used nearly 1.5 times more memory words related to recollecting and forgetting, such as *forget* and *remember*, than the control group.

Differences in Personal Needs

Regarding essential needs, the results indicated that the tweets of the GAD group contained approximately 1.8 times more words from the *want*, *need*, and *lack* categories than the tweets of the control group. Moreover, the GAD group showed increased use of *substances*, *death*, and *food* words. The mean values of these categories were 0.05, 0.19, and 0.48, respectively, for the GAD group, which were higher than those for the control group.

Regarding self-esteem needs, the number of reward words in the control group was higher than that for the GAD group. Another result was a lower number of communication words in the GAD group than in the control group. In contrast, 1.5 times more words related to interpersonal conflict were observed in the GAD group compared to the control group.

XSL•FO
RenderX

Table 4 highlights the significant differences in most categories of personal needs, underscoring the depth of these differences. This was evident in the *need* (Cohen *d*=0.83), *want* (Cohen *d*=0.65), *lack* (Cohen *d*=0.33), *risk* (Cohen *d*=0.67), *curiosity* (Cohen *d*=1.05), *food* (Cohen *d*=0.41), *death* (Cohen *d*=0.49), *interpersonal conflict* (Cohen *d*=0.50), and *communication* (Cohen *d*=0.64) categories. Although the *reward* and *substances* categories were statistically significant (*P*=.003 and *P*<.001, respectively), the effect size was small (Cohen *d*=0.11 and 0.16, respectively).

Differences in Emotional Expressiveness

The GAD group expressed more negative tones. Notably, their tweets also showed extensive use of anxiety-related words, 3 times more than the control group. In addition, their tweets expressed anger and sadness more frequently. Moreover, the GAD group used a higher number of swear words.

We found statistical significance (*P*<.001) as well as a significant effect size for most features, including anxiety (Cohen *d*=0.62), anger (Cohen *d*=0.72), sadness (Cohen *d*=0.48), and swear words (Cohen *d*=2.61). However, the effect size of the negative tone was small (Cohen *d*=0.10).

Thematic Analysis

Overview

We present the results of the thematic analysis, which included 4 primary themes: symptoms, life problems, relationships, and feelings. The symptoms theme was derived from major symptoms according to the *DSM-5* for GAD. Other themes were derived from previous research [64] highlighting how GAD may affect an individual's perceived life satisfaction in different aspects such as relationships, health, and work. The seed words of the symptoms theme were defined based on GAD symptoms and the health-related problems of GAD according to the *DSM-5*. To generate seed words for other themes, we used a reverse dictionary named WantWords (The Natural Language Processing Group, Tsinghua University) [65] to obtain words that related semantically to the descriptions of these themes. Table 5 shows a sample of seed words for each theme.

Table 5. A sample of seed words used by GuidedLDA for each theme.

Theme	Seed words
Symptoms	"Worry," "concern," "anxiety," "twitch," "treatment," "restlessness," "concentration," "muscle," "mental," and "health"
Life problems	"Business," "money," "work," "school," "job," "salary," "promotion," "college," "office," and "experience"
Relationships	"Wife," "husband," "parents," "kids," "friends," "partners," "spouse," "fiancee," "fiance," and "father"
Feelings	"Feel," "cry," "respect," "release," "support," "emotion," "disclosure," "care," "sadness," "sense," and "sorry"

Symptoms Theme

The symptoms theme reflects the GAD symptoms and other health problems related to GAD among those who have it that were expressed in their tweets. Some tweets showed how individuals with GAD have extreme anxiety about various things, and others represented how individuals with GAD tend to expect the worst scenario for any issue. In addition, users with GAD often posted about experiencing difficulty concentrating and overthinking plans. Other tweets showed that people with GAD tend to experience tiredness and sleep disturbance because of insomnia.

The results also demonstrate that many people with GAD discuss their health status, therapy, and treatment with others. Some users also discussed how excessive worry caused other health problems, such as heart conditions and digestive issues.

Life Problems Theme

This theme denotes how individuals with GAD worry about various issues, including financial status, work, and school. Some users with GAD tended to share with others how GAD negatively impacted work because of difficulty concentrating on any task and accomplishing it on time, reducing work performance. Some users with GAD discussed the difficulty maintaining employment and the increased fear of financial strain. Other tweets showed that several people with GAD discussed the impact of GAD on their academic performance.

Relationships Theme

This theme highlights that people with GAD have impairments in relationships with partners, family members, and friends. Many users tended to discuss their problems in relationships with others. We found that some tweets showed that individuals with GAD had fewer relationships for several reasons, such as having trouble expressing their feelings, having poor self-esteem, and being irritable with others. In addition, some users discussed how GAD significantly affected their marriage, whereas some individuals reported that GAD caused the weakening of their relationships with their partners and decreased enjoyment of activities with them.

Feelings Theme

This theme indicates that several users with GAD tend to express and share their feelings and emotions with others on social media. We found that they disclosed their feelings to obtain support and help from professionals or other users or raise awareness of mental health issues.

Summary of Thematic Analysis Results

In Figure 4, the GuidedLDA-based thematic analysis shows that users in the GAD group had a higher frequency of these themes per tweet than those in the control group for every theme. The results indicate that the GAD group tended to discuss life problems and other themes 1.5 times and approximately 2 times more than the control group, respectively. Table 6 shows the statistically significant differences for all themes between users with GAD and control users. We found statistically significant

XSL•FO
RenderX

differences (P<.001) and significant effect sizes (Cohen d>0.50) for most themes. Examples of paraphrased tweets for each theme can be found in Multimedia Appendix 1.

Figure 4. The frequency of the GuidedLDA-based themes per tweet for each user profile in both groups. GAD: generalized anxiety disorder.

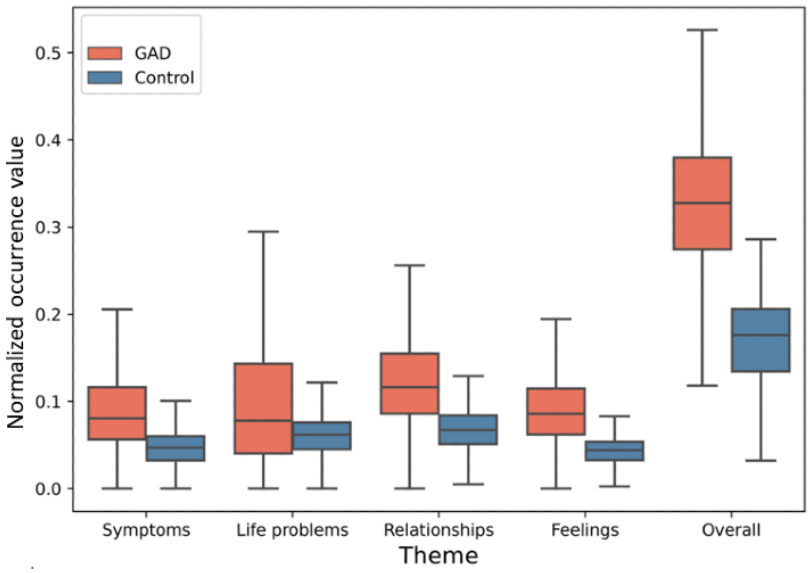

Table 6. A comparison of themes between the generalized anxiety disorder (GAD) and control groups[a].

Theme	GAD group, mean (SD)	Control group, mean (SD)	*P* value	Cohen *d*
Symptoms (percentage of total tweets per user profile)	10 (8)	5 (2)	<.001	0.86
Life problems (percentage of total tweets per user profile)	11 (10)	7 (7)	<.001	0.46
Relationships (percentage of total tweets per user profile)	13 (8)	7 (3)	<.001	0.99
Feelings (percentage of total tweets per user profile)	9 (5)	4 (2)	<.001	1.31

[a]*P* values were calculated using a 1-tailed, 2-sample *t* test, with $P \leq .05$ indicating statistical significance, and the effect size is reported using the Cohen *d*.

Behavioral Analysis

In this section, an analysis is presented to study the patterns of individuals with GAD concerning their behaviors and activities on Twitter. Moreover, we measured the correlation between the GuidedLDA-based themes and users' behaviors, including hashtag participation, volume, interaction patterns, social engagement, and reactive behavior.

Participation in Hashtags

Participating in hashtags shows the tendency of users to discuss various topics in public with others. The analysis showed that people with GAD participated in fewer hashtags than people without GAD. In addition, the results showed different preferences for the types of hashtag topics followed by people with GAD compared to people without GAD. People with GAD preferred to participate in hashtags related to their mental health status (eg, *#anxiety* and *#mentalhealth*), therapy, and treatments (eg, *#ketaminetherapy* and *#ketwell*). In addition, they liked to follow hashtags where they had the opportunity to discuss their mental health with others to seek support and reduce isolation (eg, *#MentalHealthMatters*).

Many people with GAD tended to participate slightly more in hashtags related to current events regarding health, politics, and other matters (eg, *#news* and *#COVID19*) than those without GAD. On the other hand, people without GAD followed hashtags related to daily life, ranging from family to entertainment, more than individuals with GAD (eg, *#NBATwitter* and *#movie*). We found that people with GAD tended to participate less using hashtags that those without GAD were usually more active with. The mean value of users' tweets per day containing a hashtag in the GAD group was 0.2 (SD 0.54; median 0.06, IQR 0-0.42), whereas the mean value of users' tweets per day containing a hashtag for those without GAD was significantly higher at 0.53 (SD 1.13; median 0.13, IQR 0-0.89; P<.001).

We also measured the ratio of all tweets that participated in a hashtag to the overall tweets. We found that 15% (591,928/3,946,188) of tweets from the control group included hashtags compared to 8.6% (289,611/3,367,581) of tweets from the GAD group. The median value for the former was 9% (IQR 6.75%-13.5%), whereas the median value for the latter was 4.7% (IQR 3.53%-7.06%).

However, those in the GAD group who participated in hashtags mainly focused on sharing their symptoms, as shown in Figure 5. We also identified a higher correlation (ρ) between all 4 identified themes overall and hashtag ratio than for those in the control group. This indicates that individuals without GAD participated in hashtags that were less correlated with these themes. The highest value for ρ was observed when individuals

XSL•FO
RenderX

with GAD shared their symptoms with others (0.34 compared to 0.05 for the control group).

Figure 5. The correlation between the GuidedLDA-based theme frequency and metadata for (A) the GAD group and (B) the control group in our dataset. GAD: generalized anxiety disorder.

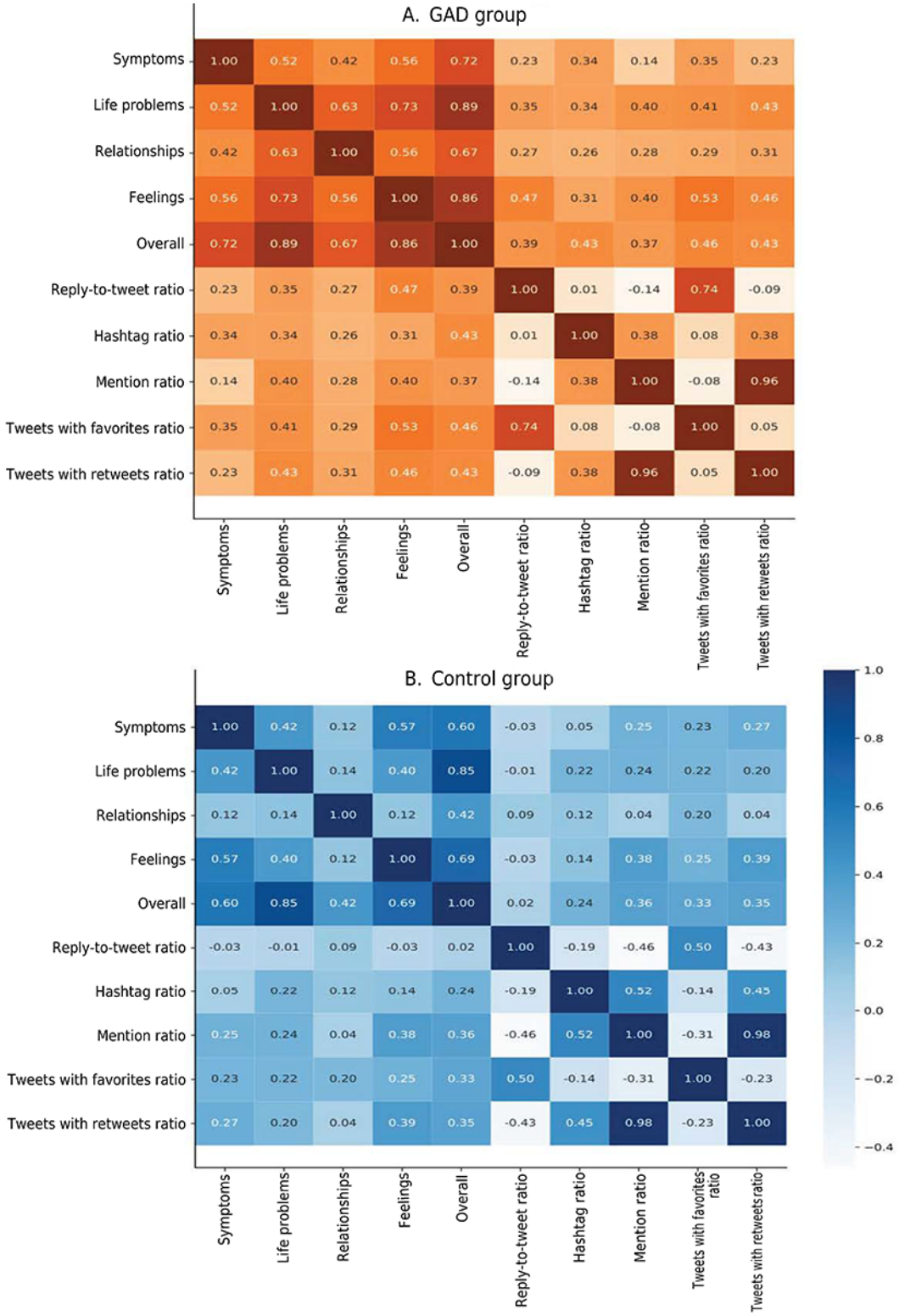

Metadata of Engagement Analysis

We explored the 2 groups through metadata analysis, such as how they interacted with each other in terms of replies, mentions, number of followers and followees, and volume of tweets.

Volume

We measured the volume of several aspects for each user from their profile history. Our scope spanned 12 aspects to be compared between the GAD and control groups, as shown in Figure 6. In general, we found that individuals in the control group had 46.93% and 30.5% more friends and followers,

XSL•FO
RenderX

respectively, and had a 31.53% higher listed count. In addition, we found that the control group was 52.33% more active in the participation in hashtags than the GAD group and tended to have 64.11% and 40.85% more retweets and mentions, respectively.

In addition, the control group tended to retweet popular tweets 42.47% more than the GAD group. However, we noticed a 36.01% lower volume of replies for the control group than for people with GAD. This might be due to the fact that the GAD group tends to be more reactive than proactive, which will be discussed later in the *Reactive Behavior* section.

Table 7 shows the statistical significance of the volume features. We found statistical significance and significant effect sizes for the number of replies (Cohen *d*=0.33), number of hashtags (Cohen *d*=0.51), number of retweets (Cohen *d*=0.62), total number of mentions per user (Cohen *d*=0.30), total number of mentions without replies per user (Cohen *d*=0.63), and total tweets with retweets (Cohen *d*=0.60). There were no statistically significant differences in the follower count and listed count features (Cohen *d*=0.06 and 0.07, respectively).

The results also showed a statically significant difference in friend count, favorite count, retweet count, and number of tweets. However, the effect size was small for these features (Cohen *d*=0.14, 0.20, 0.14, and 0.08, respectively).

Figure 6. A comparison of mean values across different volume metrics between the generalized anxiety disorder (GAD) and control groups.

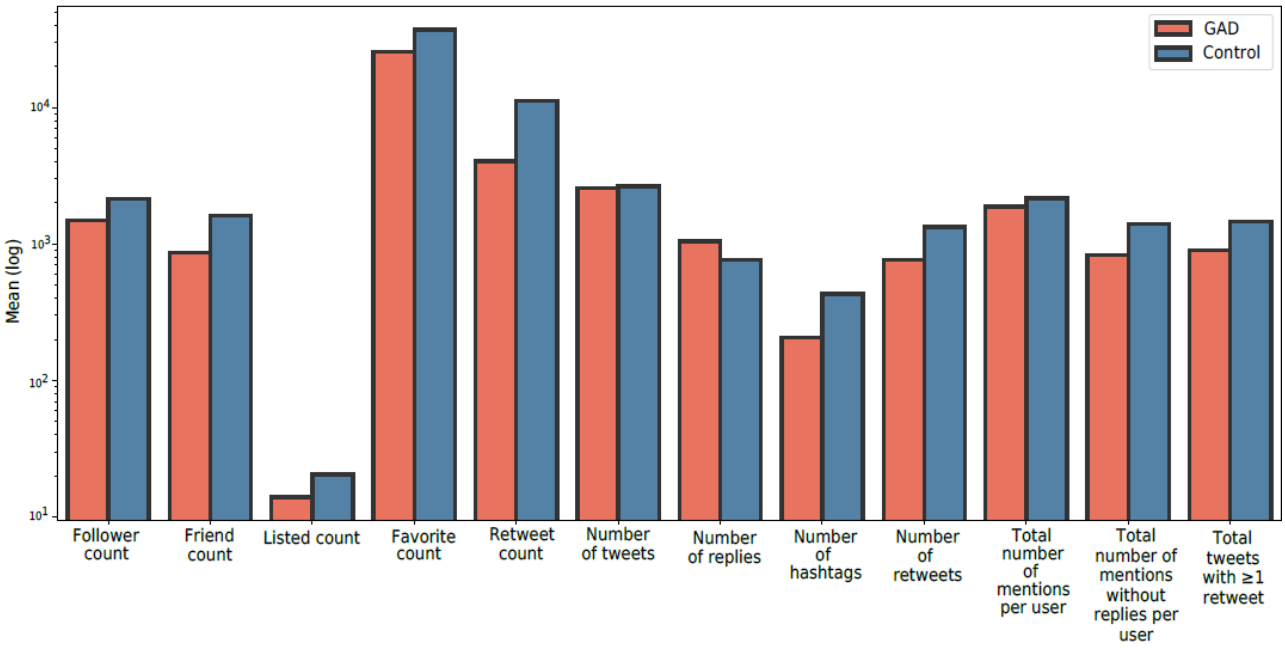

Table 7. Comparing the volume metrics between the generalized anxiety disorder (GAD) and control groups[a].

Feature	GAD group, mean (SD)	Control group, mean (SD)	*P* value	Cohen *d*
Follower count	1492.90 (7433.32)	2148.25 (12,479.92)	.09	0.06
Friend count	859.56 (2016.48)	1619.79 (7567.63)	<.001	0.14
Listed count	14.03 (65.89)	20.56 (111.17)	.06	0.07
Favorite count	25,480.23 (40,851.50)	37,215.05 (70,582.59)	<.001	0.20
Retweet count	4033.26 (25,873.79)	11,239.02 (70,861.73)	<.001	0.14
Number of tweets	2559.78 (1315.98)	2650.30 (918.30)	.04	0.08
Number of replies	1042.95 (873.12)	766.84 (792.72)	<.001	0.33
Number of hashtags	205.19 (305.56)	430.50 (539.18)	<.001	0.51
Number of retweets	767.65 (773.00)	1334.42 (1029.41)	<.001	0.62
Total number of mentions per user	1871.37 (1082.88)	2167.46 (909.44)	<.001	0.30
Total number of mentions without replies per user	828.42 (779.71)	1400.62 (1019.23)	<.001	0.63
Total tweets with retweets	902.59 (797.51)	1453.47 (1012.96)	<.001	0.60

[a]*P* values were calculated using a 1-tailed, 2-sample *t* test, with $P \leq .05$ indicating statistical significance, and the effect size is reported using the Cohen *d*.

XSL•FO
RenderX

Interaction Patterns

The mention and reply features on Twitter could reflect the interaction among Twitter users. Therefore, we analyzed the ratio of mentions and replies for people with and without GAD to study the extent of interaction in these 2 groups. We found that individuals with GAD had a lower mention ratio than those without GAD, indicating less interaction with others.

Another indication of a user interacting with the community is acting positively and communicating more with others. On Twitter, users tend to mention someone within their tweets to ask for participation in their opinion. Another example is to retweet others' ideas for either endorsement or to pass on an interesting tweet to their audience for further discussion. We found that control users tended to include more mentions per day than people with GAD. The mean number of mentions per day for the control group was 1.57 (SD 2.31; median 0.69, IQR 0.30-4.95). This is, in fact, more than double that of people with GAD (mean 0.78, SD 1.2; median 0.39, IQR 0.25-2.60).

For more confirmation of such behavior, we also found that 51% (2,012,555/3,946,188) of tweets from the control group had a mention, whereas 31.5% (1,060,788/3,367,581) of tweets included a mention in the GAD group. The mentions comprised those receiving a reply and those solely mentioned as both are technically considered *mentions* in the Twitter community. In Figure 5, users in the GAD group show a high correlation with mentions when talking about their life problems and relationships (approximately 0.4 and 0.28 compared to 0.24 and 0.04 for the control group, respectively).

However, the symptoms theme showed a lower correlation with mentions for the GAD group than for the control group (0.14 compared to 0.25). We found that the control group tended to talk directly to someone about health issues. In contrast, the GAD group associated health discussions with a hashtag to receive more attention and used direct mentions to seek support for life problems and relationships.

In addition, the retweets per day and their ratio to the overall tweets may indicate endorsement of others' tweets or the desire to pass them on to their audience for further discussion. We found a mean of 1.63 (SD 2.31; median 0.74, IQR 0.50-5.19) retweets per day for the control group and 0.86 (SD 1.26; median 0.44, IQR 0.40-2.87) retweets per day for the GAD group. Moreover, 48% (1,894,170/3,946,188) of the control group's tweets were retweets, whereas 28.91% (973,567/3,367,581) of the GAD group's tweets were retweets.

However, we also observe in Figure 5 that, in the GAD group, there is a high correlation between both the feelings and life problems themes and higher likes and retweets received from their audience when they tweet (ρ of 0.53 and 0.41 in the GAD group compared to 0.25 and 0.22 in the control group, respectively). This result might suggest that the GAD group interacts with others when they share the effects of GAD on their life, seeking support and a way to release and express their emotions.

Social Engagement

We used the number of followers and followees to measure the social engagement among users. The results showed that people with GAD had fewer followers and followees than those without GAD, as shown in Figure 6. This might reflect that the GAD group was less social than the control group and tended to limit their relationships.

However, we found that users with a large follower-to-followee ratio were associated with GAD, as shown in Figure 7. This result could be due to these users being more well known, making them more likely to receive criticism from individuals who are unfamiliar with them. In addition, people with GAD had fewer friends than the control group. To summarize these cases, we identified 2 categories within the GAD group. The first category was those with high pressure from a large number of followers, such as celebrities, whereas the second category was those who tended to live alone and have fewer friends.

The mean value of the follower-to-followee ratio in the GAD group was 5.34 (SD 93.82). On the other hand, the control group had a mean follower-to-followee ratio of 2.51 (SD 15.53). The large value for the SD is due to those who had a high number of followers and experienced pressure from their audience, leading to some GAD symptoms. Therefore, we measured the median for both groups, with 0.76 (IQR 0.25-157.91) and 0.88 (IQR 0.25-26.89) for the GAD and control group, respectively. Figure 7 shows the first category within the GAD group, whereas Figure 8 shows the second category.

XSL•FO
RenderX

Figure 7. A comparison of mean values across different features between the generalized anxiety disorder (GAD) and control groups. FF: follower-to-followee.

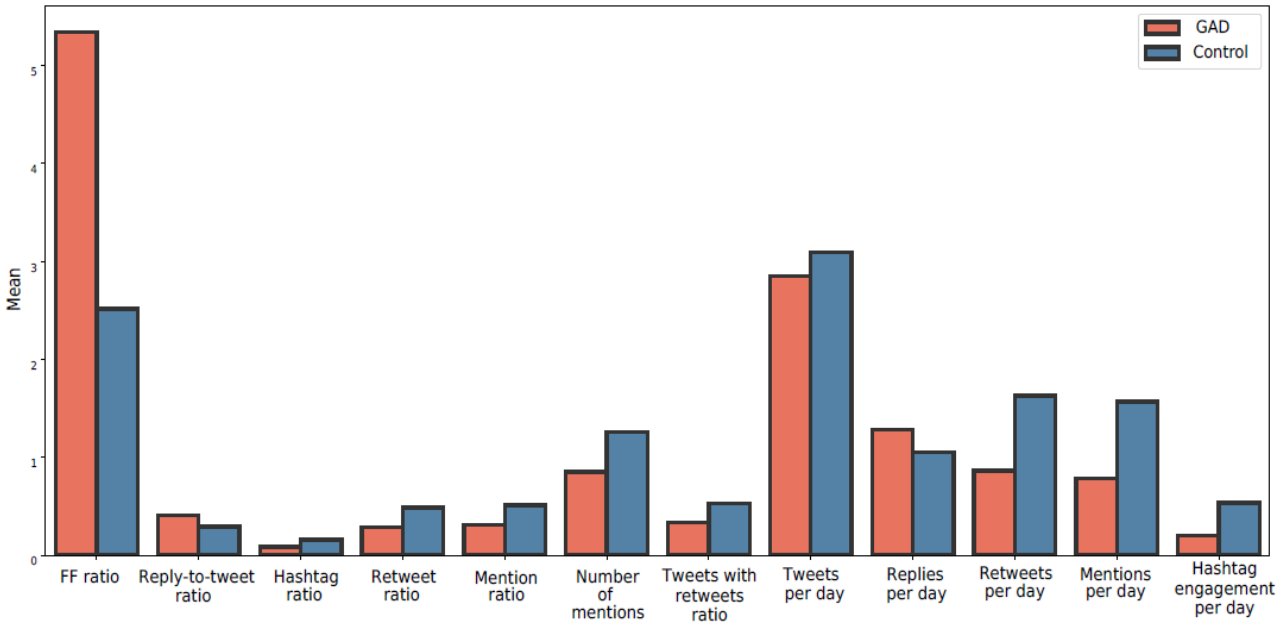

Figure 8. A comparison of median values across different features between the generalized anxiety disorder (GAD) and control groups. FF: follower-to-followee.

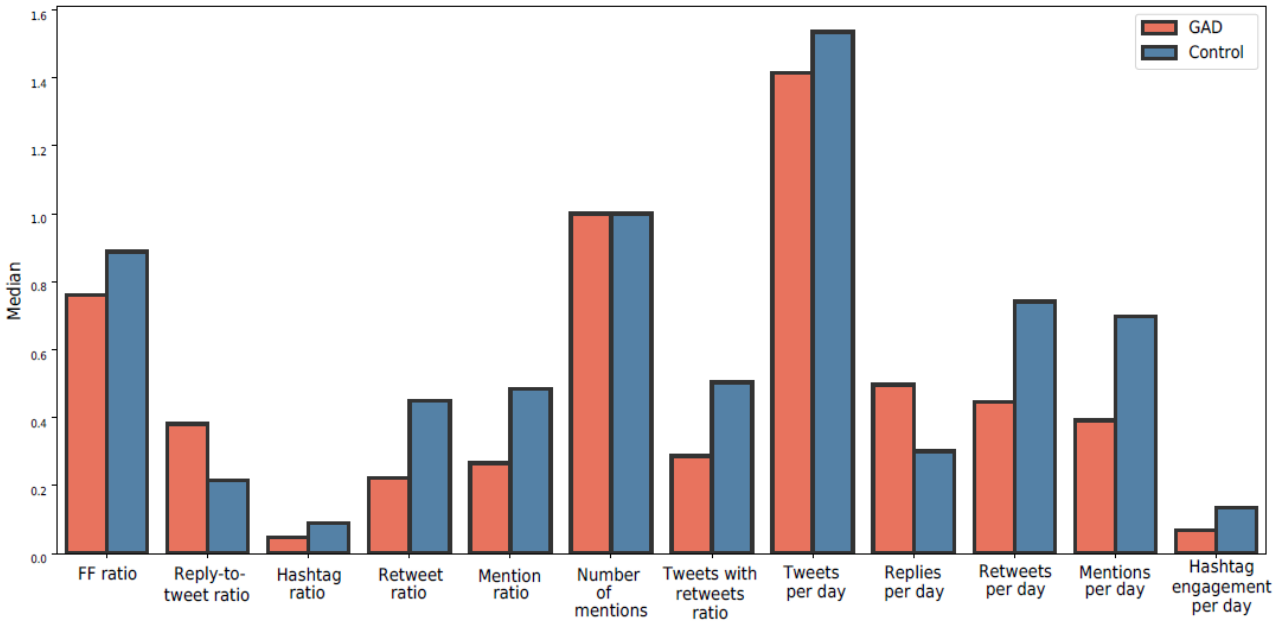

Reactive Behavior

We identified that people with GAD tended to reply more than write new tweets compared to those in the control group. Figure 7 shows that the mean value of users' replies per day for the GAD group was 1.28 (SD 2.12), whereas the mean value of users' replies per day in the control group was slightly lower at 1.05 (SD 2.05). The margin was even larger in terms of median values at 0.49 (IQR 0.35-4.40) and 0.30 (IQR 0.25-4.08) for the GAD and control groups, respectively. This may indicate a common aspect of those in the GAD group being more reactive to others rather than engaging proactively.

To clarify this, we also measured the ratio of replies to total tweets, which points to the same behavior. The mean value of this ratio for the GAD group was 0.41 (SD 0.24), and the median was 0.38 (IQR 0.30-0.82). In the control group, the mean value of the ratio of replies to total tweets was 0.29 (SD 0.26), and the median was 0.21 (IQR 0.18-0.69). We confirm the reactive behavior in Figure 5, where there is a higher correlation (ρ) for the GAD group between the *symptoms* theme, *life problems* theme, *relationships* theme, *feelings* theme, and the 4 identified themes overall and the reply-to-retweet ratio at 0.23, 0.35, 0.27, 0.47, and 0.39 compared to the control group at –0.03, –0.01, 0.09, –0.03, and 0.02, respectively. Table 8 shows the statistical significance for most of the behavioral features.

We found statistical significance and significant effect sizes for the reply-to-tweet ratio (Cohen d=0.47), hashtag ratio (Cohen d=0.49), retweet ratio (Cohen d=0.69), mention ratio (Cohen

XSL•FO
RenderX

d=0.68), tweets with retweets ratio (Cohen *d*=0.68), retweets per day (Cohen *d*=0.41), mentions per day (Cohen *d*=0.42), and hashtag engagement per day (Cohen *d*=0.38).

There were no statistically significant differences in the follower-to-followee ratio and tweets per day features (Cohen *d*=0.04 and 0.07, respectively). The results also show a statistically significant difference in mention count and replies per day. However, the effect size was small for these features (Cohen *d*=0.17 and 0.11, respectively).

Table 8. Comparing the behavior-based metrics between the generalized anxiety disorder (GAD) and control groups[a].

Feature	GAD group, mean (SD)	Control group, mean (SD)	*P* value	Cohen *d*
FF[b] ratio	5.34 (93.82)	2.52 (15.53)	.28	0.04
Reply-to-tweet ratio	0.41 (0.24)	0.29 (0.27)	<.001	0.47
Hashtag ratio	0.087 (0.11)	0.16 (0.18)	<.001	0.49
Retweet ratio	0.29 (0.24)	0.49 (0.33)	<.001	0.69
Mention ratio	0.32 (0.23)	0.51 (0.32)	<.001	0.68
Number of mentions	0.85 (0.84)	1.26 (3.27)	<.001	0.17
Tweets with retweets ratio	0.34 (0.23)	0.53 (0.32)	<.001	0.68
Tweets per day	2.85 (3.69)	3.09 (3.56)	.82	0.07
Replies per day	1.29 (2.12)	1.06 (2.05)	.004	0.11
Retweets per day	0.87 (1.27)	1.63 (2.32)	<.001	0.41
Mentions per day	0.79 (1.20)	1.57 (2.31)	<.001	0.42
Hashtag engagement per day	0.20 (0.54)	0.54 (1.14)	<.001	0.38

[a]*P* values were calculated using a 1-tailed 2-sample *t* test, with $P \leq .05$ indicating statistical significance, and the effect size is reported using the Cohen *d*.

[b]FF: follower-to-followee.

Discussion

Principal Findings

This study characterized the GAD digital markers using social media content and behavior analysis of individuals with GAD. The results show the importance of studying psychological states reflected in language. Investigating the word use patterns in people with GAD revealed differences in all the defined measurements (ie, cognitive style, personal needs, and emotional expressiveness) compared to individuals without GAD. First, all measurements of cognitive style were higher for those with GAD, including certainty, discrepancy, and tentativeness, which may reflect several GAD markers. One possible explanation for why users with GAD expressed more certainty is that they have difficulty dealing with uncertainty, which is aligned with the work by Miranda and Mennin [66]. Another possible reason is that people with GAD tend to have a negative outlook toward future events because of persistent anxiety [66]. The high number of tentative and discrepancy words in their tweets might be attributed to having difficulty putting worries aside or letting them go, having difficulty making decisions, and being indecisive, or it might also indicate inability to focus, consistent with the work by Koerner and Dugas [67].

In addition, individuals with GAD tended to use more memory words, which might reflect memory problems, including recalling and forgetfulness. This result might reveal some GAD symptoms (ie, difficulty concentrating and excessive worry) that may contribute to emerging memory issues, which is aligned with the work by Butters et al [68]. Second, we also found differences in personal needs regarding essential and self-esteem needs. All measurements of essential needs were higher for those with GAD. The increased use of word categories (ie, *want*, *need*, and *lack*) might indicate the desire to seek support and help or raise mental health awareness, which agrees with the work by Fine et al [69].

We also observed a higher number of references to food and death in the GAD group. This might indicate that GAD is more likely to be associated with unhealthy lifestyle behaviors, which is aligned with the work by Moselhy et al [70]. A possible reason for the increased use of food words in the GAD group might be the relationship between GAD and increased food intake, which agrees with the studies by Moselhy et al [70] and Fonseca et al [71]. Regarding self-esteem needs, people with GAD used fewer words related to rewards but more words related to curiosity and risk. This result might be explained by the fact that users who experience GAD have difficulty accomplishing tasks because of excessive anxiety, inability to focus, and overthinking, which is consistent with the work by Habib and Laidey [72].

Moreover, increased use of interpersonal conflict words and decreased use of communication words were observed in the GAD group. This might be due to being irritable and nervous, which are symptoms of GAD. Finally, the results also indicate differences in emotional expressiveness between the 2 groups. The increased negative tones, anger, anxiety, and sadness in tweets posted by individuals with GAD might be due to excessive anxiety, which is aligned with the work by Mennin

XSL•FO
RenderX

et al [73], which found difficulties in emotion regulation among people with GAD.

Our findings highlight the importance of thematic analysis in quantifying GAD signals. We found that users with GAD tended to post more about their symptoms, health status, life issues, friendship problems, and feelings. Interestingly, there were significant differences in all themes—symptoms, life problems, relationships, and feelings—between the GAD group and the control group. We note that relying on the *DSM-5* and previous work [64] related to the impact of GAD on individual life enhanced our analysis to generate themes more related to GAD, resulting in a detailed thematic analysis to distinguish GAD signals.

We found that the dominant themes in tweets posted by users with GAD tended to focus on life problems and relationships, followed by feelings and symptoms. This result may be explained by the fact that worsening symptoms and difficulty with emotion regulation contribute to difficulty handling life and relationship issues.

The other key finding was that GAD signals can be found in users' digital activities. The behavioral analysis focused on users' activities based on Twitter metadata regarding hashtag participation, volume, interaction patterns, social engagement, and reactive behaviors. We found that the GAD group used fewer hashtags than those without GAD, meaning that people with GAD might have fewer interactions with the public. These findings might indicate that people with GAD tend to be isolated by having fewer relationships and limiting their interaction with people, which is consistent with the work by Henning et al [64] and Revicki et al [74] that shows that GAD is associated with heightened social functioning impairment.

We also observed that people with GAD have different preferences regarding topics for participating in hashtags compared to users without GAD. We found that people with GAD preferred to participate more in hashtags related to mental health disorders and treatments. This might be due to their low mood or desire to seek support, raise mental health awareness, or overcome mental health stigma. The results also showed that users with GAD were more likely to follow hashtags about the latest news, which might reflect their low mood and contribute to heightened anxiety levels, which is aligned with the work by Cheun et al [75].

In addition, we found that users without GAD tended to have more followees, mentions, retweets, and hashtags. These findings might suggest that people with GAD are less active and tend to limit their interaction with others. This might reflect that people with GAD have impairments in relationships, which is in agreement with the work by Henning et al [64].

Moreover, this analysis measured the correlation between the GuidedLDA-based themes and users' behaviors. The results showed a high correlation between hashtag ratio and overall themes for the GAD group compared to the control group. For users with GAD, the results identified a high correlation between the mention ratio and the life problems and relationships themes, whereas the symptoms theme had a lower correlation with the mention ratio. This might indicate that the GAD group tends to discuss their health issues using related hashtags to obtain more attention and use mentions to seek direct help for life problems and relationship issues.

Our findings regarding language use patterns are consistent with the research presented in the studies by Coppersmith et al [25] and Shen and Rudzicz [29], which identify linguistic patterns for individuals with GAD. Regarding topic analysis, our study supports the conclusions of both the studies by Shen and Rudzicz [29] and Gkotsis et al [30] that individuals with GAD tend to focus on specific topics in their discussions. In our study, we comprehensively extended previous work by characterizing linguistic and thematic markers associated with GAD and providing a more detailed and in-depth examination of the textual content of individuals with GAD. Regarding the behavior analysis, previous studies [19,35,42,44] have covered the behaviors of individuals with a low mental health status on social media. Our exploration comprehensively analyzed how people with GAD interact and express themselves on the web. It highlights unique behavioral patterns that may contribute significantly to a deeper understanding of the disorder in the digital context. To the best of our knowledge, this is the first study that comprehensively explores the behaviors on social media of people with GAD, along with investigating the correlation between thematic content and users' behaviors on social media platforms. Therefore, our work aims to expand our understanding of the unique digital markers of GAD related to linguistic, thematic, and behavioral aspects.

It is worth noting that using publicly available social media data for mental health research raises ethical issues, including concerns about users' privacy and informed consent [45,46]. To address these ethical concerns, a comprehensive strategy to anonymize the collected datasets is essential to prevent the identification of users. This involves removing all identifiable information, such as usernames and real names, and paraphrasing tweets to ensure that there is no inferred information about those users [45,46].

Limitations

This study has a few limitations. First, we did not investigate the link between the mental health status of people with GAD and demographic characteristics, including age, gender, education, and employment status. These features cannot be retrieved from a user's profile using the Twitter API.

One limitation is that this study did not cover a user's medical and childhood history. Traumatic events in the early years of someone's life can negatively impact mental health throughout their life. Investigating a patient's medical history enhances the mental health diagnosis because it reveals their medical problems and the complications they went through, which may develop into psychological disorders. These factors contribute to increasing the likelihood of experiencing mental health disorders such as GAD.

Another limitation is that this study did not consider the severity of GAD on a scale from mild to severe. The severity of GAD may affect how it interferes with several aspects of daily living, so the degree to which GAD causes impairments may vary based on its severity. This study is also limited by the lack of

XSL•FO
RenderX

prospective data, which makes causal inference impossible; thus, only cross-sectional inferences can be drawn.

Regarding data collection, Twitter's API imposes limitations that might impact the comprehensiveness of collected datasets. According to the limit imposed by the Twitter API, only the 3200 most recent tweets can be retrieved. Another limitation in terms of the representativeness of the sample is that identifying potential members of a control group can pose a challenge as it might include individuals who experience GAD but choose not to disclose their diagnosis publicly. However, this study used a regular expression to search for self-reported diagnosis statements of a mental health disorder, which is a common technique used to identify individuals diagnosed with a mental health disorder on social media [24,40].

As future work, we intend to extend this study to conduct sentiment analysis of tweets, which could clarify any cues for GAD. In addition, a larger study could be conducted in which the data obtained from social media platforms could be combined with other information, such as medical history, gender, age, and employment status, to yield a more accurate picture of an individual's mental health and contribute to developing a robust GAD detection model via social media. Moreover, using more advanced NLP techniques beyond regular expressions to identify individuals with GAD could significantly enhance the ability to identify people with GAD via their social media posts. For instance, using machine learning models trained on a broader set of linguistic features allows for identifying markers of GAD effectively.

Conclusions

Our study provides several valuable insights into identifying the characteristics of GAD in social media networks. First, examining tweets from a linguistic perspective indicated that the language used by people with GAD reveals their mental health status. Second, thematic analysis of tweets highlighted which topics people with GAD tend to discuss more with others. Third, studying the behavior patterns showed how users with GAD interact with others, which provides more understanding of GAD.

In this study, we identified and characterized the digital markers associated with GAD by conducting a comprehensive analysis of social media activity from individuals diagnosed with GAD. Using content-based analysis techniques such as LIWC for linguistic analysis and GuidedLDA for thematic analysis, we thoroughly examined the language and themes in these individuals' social media posts.

Furthermore, we determined significant behavioral differences by analyzing user- and tweet-level features, such as hashtag participation and interaction patterns. We also investigated the correlation between the themes identified by GuidedLDA and various metadata for both the GAD and control groups within our dataset. Our findings include 47 distinct GAD digital markers, with results presented in terms of means, SDs, *P* values, and Cohen *d*, highlighting these markers' statistical significance and effect size.

In conclusion, understanding the behavior and tweet content of users diagnosed with GAD could help develop an assistant tool for diagnosis and recovery progress assessment to be used by mental health professionals. This assistant-diagnostic tool would integrate insights from patients' social media content and behaviors into personalized treatment plans. This approach may help clinicians develop timely updated personalized treatment plans by delivering continuous support and timely interventions, enhancing communication with health care providers, encouraging engagement in activities, and customizing coping techniques.

Acknowledgments

LA is fully supported by a scholarship from the Princess Nourah bint Abdulrahman University to study for a PhD. VC is supported by the Engineering and Physical Sciences Research Council–funded King's Health Partners Digital Health Hub (EP/X030628/1).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to adhering to Twitter's policies regarding the distribution of data collected using its application programming interface, but they are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of paraphrased tweets for each theme.
[DOCX File , 14 KB - jmir_v27i1e53399_app1.docx]

References

1. World mental health report: transforming mental health for all. World Health Organization. 2022 Jun 16. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2023-02-16]

XSL•FO
RenderX

2. Percentage of the global population that suffered from select mental health disorders in 2021. Statista. URL: https://www.statista.com/statistics/979852/prevalence-of-mental-health-disorders-globally/ [accessed 2023-02-18]
3. Baker C. Mental health statistics for England: prevalence, services and funding. House of Commons Library. 2020 Jan 23. URL: https://commonslibrary.parliament.uk/research-briefings/sn06988/ [accessed 2023-03-20]
4. Diagnostic and Statistical Manual of Mental Disorders DSM-5. Washington, DC: American Psychiatric Association; 2013.
5. Generalized anxiety disorder. Mayo Clinic. URL: https://tinyurl.com/y9j48m87 [accessed 2023-02-10]
6. Barrera TL, Norton PJ. Quality of life impairment in generalized anxiety disorder, social phobia, and panic disorder. J Anxiety Disord 2009 Dec;23(8):1086-1090 [FREE Full text] [doi: 10.1016/j.janxdis.2009.07.011] [Medline: 19640675]
7. DiBonaventura M, Toghanian S, Järbrink K, Locklear J. Economic and humanistic burden of illness in generalized anxiety disorder: an analysis of patient survey data in Europe. ClinicoEconomics Outcomes Res 2014 Apr;6:151-163. [doi: 10.2147/ceor.s55429]
8. Anxiety disorders - facts and statistics. Anxiety and Depression Association of America. URL: https://adaa.org/understanding-anxiety/facts-statistics/ [accessed 2023-03-06]
9. Dixon SJ. Number of social media users worldwide from 2017 to 2028. Statista. URL: https://tinyurl.com/24x27wph [accessed 2023-04-23]
10. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
11. Tadesse MM, Lin H, Xu B, Yang L. Detection of depression-related posts in Reddit social media forum. IEEE Access 2019;7:44883-44893. [doi: 10.1109/access.2019.2909180]
12. Islam MR, Kabir MA, Ahmed A, Kamal AR, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. Health Inf Sci Syst 2018 Dec 27;6(1):8 [FREE Full text] [doi: 10.1007/s13755-018-0046-0] [Medline: 30186594]
13. Sadeque F, Xu D, Bethard S. Measuring the latency of depression detection in social media. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018 Presented at: WSDM '18; February 5-9, 2018; Marina Del Rey, CA p. 495-503. [doi: 10.1145/3159652.3159725]
14. Briand A, Almeida H, Meurs MJ. Analysis of social media posts for early detection of mental health conditions. In: Bagheri E, Cheung J, editors. Advances in Artificial Intelligence. Cham, Switzerland: Springer; 2018:133-143.
15. Paul S, Jandhyala S, Basu T. Early detection of signs of anorexia and depression over social media using effective machine learning frameworks. Conference and Labs of the Evaluation Forum. 2018. URL: https://ceur-ws.org/Vol-2125/paper_182.pdf [accessed 2023-02-11]
16. Wu MY, Shen CY, Wang ET, Chen ALP. A deep architecture for depression detection using posting, behavior, and living environment data. J Intell Inf Syst 2018 Oct 30;54(2):225-244. [doi: 10.1007/s10844-018-0533-4]
17. Katchapakirin K, Wongpatikaseree K, Yomaboot P, Kaewpitakkun Y. Facebook social media for depression detection in the Thai community. In: Proceedings of the 15th International Joint Conference on Computer Science and Software Engineering. 2018 Presented at: JCSSE 2018; July 11-13, 2018; Nakhonpathom, Thailand p. 1-6. [doi: 10.1109/jcsse.2018.8457362]
18. Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, et al. Depression detection via harvesting social media: a multimodal dictionary learning solution. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017 Presented at: IJCAI 2017; August 19-25, 2017; Melbourne, Australia p. 3838-3844. [doi: 10.24963/ijcai.2017/536]
19. Vedula N, Parthasarathy S. Emotional and linguistic cues of depression from social media. In: Proceedings of the 2017 International Conference on Digital Health. 2017 Presented at: DH '17; July 2-5, 2017; London, UK p. 127-136. [doi: 10.1145/3079452.3079465]
20. Owen D, Antypas D, Hassoulas A, Pardiñas AF, Espinosa-Anke L, Collados JC. Enabling early health care intervention by detecting depression in users of web-based forums using language models: longitudinal analysis and evaluation. JMIR AI 2023 Mar 24;2:e41205 [FREE Full text] [doi: 10.2196/41205] [Medline: 37525646]
21. Cui B, Wang J, Lin H, Zhang Y, Yang L, Xu B. Emotion-based reinforcement attention network for depression detection on social media: algorithm development and validation. JMIR Med Inform 2022 Aug 09;10(8):e37818 [FREE Full text] [doi: 10.2196/37818] [Medline: 35943770]
22. Zhu J, Jin R, Kenne DR, Phan N, Ku WS. User dynamics and thematic exploration in r/depression during the COVID-19 pandemic: insights from overlapping r/SuicideWatch users. J Med Internet Res 2024 May 20;26:e53968 [FREE Full text] [doi: 10.2196/53968] [Medline: 38767953]
23. Kabir MK, Islam M, Kabir AN, Haque A, Rhaman MK. Detection of depression severity using Bengali social media posts on mental health: study using natural language processing techniques. JMIR Form Res 2022 Sep 28;6(9):e36118 [FREE Full text] [doi: 10.2196/36118] [Medline: 36169989]
24. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring depression trends on Twitter during the COVID-19 pandemic: observational study. JMIR Infodemiology 2021 Jul 18;1(1):e26769 [FREE Full text] [doi: 10.2196/26769] [Medline: 34458682]

XSL•FO
RenderX

25. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPsych@NAACL-HLT 2015; May 31-June 5, 2015; Denver, CO p. 1-10. [doi: 10.3115/v1/w15-1201]
26. Gaur M, Kursuncu U, Alambo A, Sheth A, Daniulaityte R, Thirunarayan K, et al. "Let me tell you about your mental health!": contextualized classification of Reddit posts to DSM-5 for web-based intervention. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018 Presented at: CIKM '18; October 22-26, 2018; Torino, Italy p. 753-762. [doi: 10.1145/3269206.3271732]
27. Thorstad R, Wolff P. Predicting future mental illness from social media: a big-data approach. Behav Res Methods 2019 Aug 29;51(4):1586-1600. [doi: 10.3758/s13428-019-01235-z] [Medline: 31037606]
28. Hao B, Li L, Li A, Zhu T. Predicting mental health status on social media: a preliminary study on microblog. In: Proceedings of the 5th International Conference on Cross-Cultural Design. 2013 Presented at: CCD 2013; July 21-26, 2013; Las Vegas, NV p. 101-110. [doi: 10.1007/978-3-642-39137-8_12]
29. Shen JH, Rudzicz F. Detecting anxiety through Reddit. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality. 2017 Presented at: CLPsych@ACL 2017; August 3, 2017; Vancouver, BC p. 58-65. [doi: 10.18653/v1/w17-3107]
30. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using informed deep learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
31. Coppersmith G, Harman C, Dredze M. Measuring post traumatic stress disorder in Twitter. Proc Int AAAI Conf Web Soc Media 2014 May 16;8(1):579-582. [doi: 10.1609/icwsm.v8i1.14574]
32. Ismail NH, Liu N, Du M, He Z, Hu X. A deep learning approach for identifying cancer survivors living with post-traumatic stress disorder on Twitter. BMC Med Inform Decis Mak 2020 Dec 14;20(Suppl 4):254 [FREE Full text] [doi: 10.1186/s12911-020-01272-1] [Medline: 33317508]
33. Chen X, Sykora M, Jackson T, Elayan S, Munir F. Tweeting your mental health: an exploration of different classifiers and features with emotional signals in identifying mental health conditions. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: HICSS 2018; January 3-6, 2018; Hilton Waikoloa Village, HI. [doi: 10.24251/hicss.2018.421]
34. Ramirez-Cifuentes D, Mayans M, Freire A. Early risk detection of anorexia on social media. In: Proceedings of the 5th International Conference on Internet Science. 2018 Presented at: INSCI 2018; October 24-26, 2018; St. Petersburg, Russia p. 3-14. [doi: 10.1007/978-3-030-01437-7_1]
35. Wang T, Brede M, Ianni A, Mentzakis E. Detecting and characterizing eating-disorder communities on social media. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. 2017 Presented at: WSDM '17; February 6-10, 2017; Cambridge, UK p. 91-100. [doi: 10.1145/3018661.3018706]
36. Benítez-Andrades JA, Alija-Pérez JM, Vidal ME, Pastor-Vargas R, García-Ordás MT. Traditional machine learning models and bidirectional encoder representations from transformer (BERT)-based automatic classification of tweets about eating disorders: algorithm development and validation study. JMIR Med Inform 2022 Mar 24;10(2):e34492 [FREE Full text] [doi: 10.2196/34492] [Medline: 35200156]
37. Yan H, Fitzsimmons-Craft EE, Goodman M, Krauss M, Das S, Cavazos-Rehg P. Automatic detection of eating disorder-related social media posts that could benefit from a mental health intervention. Int J Eat Disord 2019 Oct 05;52(10):1150-1156 [FREE Full text] [doi: 10.1002/eat.23148] [Medline: 31381168]
38. Islam MR, Kamal AR, Sultana N, Islam R, Moni MA, Ulhaq A. Detecting depression using k-nearest neighbors (KNN) classification technique. In: Proceedings of the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering. 2018 Presented at: IC4ME2 2018; February 8-9, 2018; Rajshahi, Bangladesh p. 1-4. [doi: 10.1109/ic4me2.2018.8465641]
39. Shah FM, Ahmed F, Saha Joy SK, Ahmed S, Sadek S, Shil R, et al. Early depression detection from social network using deep learning techniques. In: Proceedings of the IEEE Region 10 Symposium. 2020 Presented at: TENSYMP 2020; June 5-7, 2020; Dhaka, Bangladesh p. 823-826. [doi: 10.1109/tensymp50017.2020.9231008]
40. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPsych@ACL 2014; June 27, 2014; Baltimore, MD p. 51-60. [doi: 10.3115/v1/w14-3207]
41. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPsych@NAACL-HLT 2015; May 31-June 5, 2015; Denver, CO p. 11-20. [doi: 10.3115/v1/w15-1202]
42. Peng Z, Hu Q, Dang J. Multi-kernel SVM based depression recognition using social media data. Int J Mach Learn Cybern 2017 Jun 2;10(1):43-57. [doi: 10.1007/s13042-017-0697-1]
43. Zhang F, Zhu T, Li A, Li Y, Xu X. A survey of web behavior and mental health. In: Proceedings of the 6th International Conference on Pervasive Computing and Applications. 2011 Presented at: ICPCA 2011; October 26-28, 2011; Port Elizabeth, South Africa p. 189-195. [doi: 10.1109/icpca.2011.6106503]

XSL•FO
RenderX

44. de Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; April 27-May 2, 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]
45. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
46. Conway M. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014 Dec 22;16(12):e290 [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]
47. Twitter API documentation. X Platform. URL: https://developer.x.com/en/docs/x-api [accessed 2023-02-26]
48. Dixon SJ. Number of X (formerly Twitter) users worldwide from 2019 to 2024. Statista. URL: https://www.statista.com/statistics/303681/twitter-users-worldwide/ [accessed 2023-04-10]
49. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
50. User object. X Platform. URL: https://developer.x.com/en/docs/x-api/v1/data-dictionary/object-model/user [accessed 2023-02-26]
51. Tweet object. X Platform. URL: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet/ [accessed 2023-02-26]
52. Hoogendoorn M, Berger T, Schulz A, Stolz T, Szolovits P. Predicting social anxiety treatment outcome based on therapeutic email conversations. IEEE J Biomed Health Inform 2017 Sep;21(5):1449-1459 [FREE Full text] [doi: 10.1109/JBHI.2016.2601123] [Medline: 27542187]
53. Hoogendoorn M, Funk B. Machine Learning for the Quantified Self: On the Art of Learning from Sensory Data. Cham, Switzerland: Springer International Publishing; 2017.
54. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The development and psychometric properties of LIWC-22. University of Texas at Austin. 2022. URL: https://tinyurl.com/y2tp6zbf [accessed 2023-01-04]
55. Pavalanathan U, de Choudhury M. Identity management and mental health discourse in social media. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW '15 Companion; May 18-22, 2015; Florence, Italy p. 315-321. [doi: 10.1145/2740908.2743049]
56. Ernala SK, Rizvi AF, Birnbaum ML, Kane JM, de Choudhury M. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. Proc ACM Hum Comput Interact 2017 Dec 06;1(CSCW):1-27. [doi: 10.1145/3134678]
57. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. Use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
58. Chung CK, Pennebaker JW. The psychological functions of function words. In: Fiedler K, editor. Social Communication. New York, NY: Psychology Press; 2007:343-359.
59. de Choudhury M, De S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. Proc Int AAAI Conf Web Soc Media 2014 May 16;8(1):71-80. [doi: 10.1609/icwsm.v8i1.14526]
60. Promoting mental health: concepts, emerging evidence, practice: summary report. World Health Organization. 2004. URL: https://apps.who.int/iris/handle/10665/42940/ [accessed 2023-04-12]
61. Dahal B, Kumar SA, Li Z. Topic modeling and sentiment analysis of global climate change tweets. Soc Netw Anal Min 2019 Jun 10;9:24. [doi: 10.1007/s13278-019-0568-8]
62. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
63. Sawilowsky SS. New effect size rules of thumb. J Mod Appl Stat Methods 2009 Nov 01;8(2):597-599. [doi: 10.22237/jmasm/1257035100]
64. Henning ER, Turk CL, Mennin DS, Fresco DM, Heimberg RG. Impairment and quality of life in individuals with generalized anxiety disorder. Depress Anxiety 2007;24(5):342-349. [doi: 10.1002/da.20249] [Medline: 17091478]
65. Qi F, Zhang L, Yang Y, Liu Z, Sun M. WantWords: an Open-source Online Reverse Dictionary System. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020 Presented at: EMNLP 2020; November 16-20, 2020; Online p. 175-181. [doi: 10.18653/v1/2020.emnlp-demos.23]
66. Miranda R, Mennin DS. Depression, generalized anxiety disorder, and certainty in pessimistic predictions about the future. Cogn Ther Res 2007 Feb 21;31(1):71-82. [doi: 10.1007/s10608-006-9063-4]
67. Koerner N, Dugas MJ. A cognitive model of generalized anxiety disorder: the role of intolerance of uncertainty. In: Davey GC, Wells A, editors. Worry and Its Psychological Disorders: Theory, Assessment and Treatment. Hoboken, NJ: John Wiley & Sons; 2006:201-216.
68. Butters MA, Bhalla RK, Andreescu C, Wetherell JL, Mantella R, Begley AE, et al. Changes in neuropsychological functioning following treatment for late-life generalised anxiety disorder. Br J Psychiatry 2011 Sep 02;199(3):211-218 [FREE Full text] [doi: 10.1192/bjp.bp.110.090217] [Medline: 21727232]
69. Fine A, Nasiri K, Fotinos K, Anand L, Furtado M, Armata RS, et al. Examining predictors of help-seeking behaviours in patients with mood and anxiety symptoms. Psychiatry Res 2018 Jul;265:190-197. [doi: 10.1016/j.psychres.2018.04.040] [Medline: 29729499]

XSL•FO
RenderX

70. Moselhy HF, Ghubach R, El-Rufaie O, Zoubeidi T, Badrinath P, Sabri S, et al. The association of depression and anxiety with unhealthy lifestyle among United Arab Emirates adults. Epidemiol Psychiatr Sci 2011 Nov 03;21(2):213-219. [doi: 10.1017/s2045796011000709]
71. Fonseca NK, Molle RD, Costa MD, Gonçalves FG, Silva AC, Rodrigues Y, et al. Impulsivity influences food intake in women with generalized anxiety disorder. Braz J Psychiatry 2020 Aug;42(4):382-388 [FREE Full text] [doi: 10.1590/1516-4446-2019-0556] [Medline: 32074229]
72. Habib M, Laidey NM. Measuring the relationship between generalised anxiety disorder and job performance. Electron J Bus Manag 2021;6(2):65-82.
73. Mennin DS, Heimberg RG, Turk CL, Fresco DM. Preliminary evidence for an emotion dysregulation model of generalized anxiety disorder. Behav Res Ther 2005 Oct;43(10):1281-1310. [doi: 10.1016/j.brat.2004.08.008] [Medline: 16086981]
74. Revicki DA, Travers K, Wyrwich KW, Svedsäter H, Locklear J, Mattera MS, et al. Humanistic and economic burden of generalized anxiety disorder in North America and Europe. J Affect Disord 2012 Oct;140(2):103-112. [doi: 10.1016/j.jad.2011.11.014] [Medline: 22154706]
75. Cheun J, Dajani L, Thomas QB. Personalized detection of anxiety provoking news events using semantic network analysis. SMU Data Sci Rev 2019;2(3):5 [FREE Full text]

Abbreviations

API: application programming interface
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD: generalized anxiety disorder
LDA: latent Dirichlet allocation
LIWC: Linguistic Inquiry and Word Count
NLP: natural language processing
NLTK: Natural Language Toolkit
POS: part-of-speech

Edited by A Mavragani; submitted 05.10.23; peer-reviewed by H Zainal, G Padilla-Castillo; comments to author 10.04.24; revised version received 31.08.24; accepted 17.11.24; published 20.03.25.

Please cite as:
Alhazzaa L, Curcin V
Profiling Generalized Anxiety Disorder on Social Networks: Content and Behavior Analysis
J Med Internet Res 2025;27:e53399
URL: https://www.jmir.org/2025/1/e53399
doi:10.2196/53399
PMID:

©Linah Alhazzaa, Vasa Curcin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adolescent Cyberbullying and Cyber Victimization: Longitudinal Study Before and During COVID-19

Peter Johannes Schulz[1,2*], Prof Dr, PhD; Marc-Olivier Boldi[3*], PhD; Ann van Ackere[3*], Prof Dr, PhD

[1]Faculty of Communication, Culture and Society, Università della Svizzera Italiana, Lugano, Switzerland

[2]Department of Communication & Media, Ewha Womans University, Seoul, Republic of Korea

[3]Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland

[*]all authors contributed equally

Corresponding Author:
Peter Johannes Schulz, Prof Dr, PhD
Faculty of Communication, Culture and Society
Università della Svizzera Italiana
Via G. Buffi 13
Lugano, 6900
Switzerland
Phone: 41 58666724
Email: schulzp@usi.ch

Abstract

Background: Adolescent cyberbullying has been a persistent issue, exacerbated by the shift to remote learning and increased screen time during the COVID-19 pandemic. These changes have sparked concerns about potential increases in cyberbullying and its associated risks.

Objective: This study aims to explore how factors such as age, exposure to violent media, parental communication quality, internet access, sex, and sibling relationships influence cyberbullying behavior at school. Additionally, we examine how the COVID-19 pandemic may have altered these dynamics.

Methods: Leveraging a panel dataset, we examine the same group of adolescents both before and during the pandemic. The analysis focused on identifying relationships between the selected factors and cyberbullying perpetration and victimization, with an emphasis on the dynamics introduced by the COVID-19 pandemic.

Results: Perceived quality of parental communication was found to reduce the risk of both cyberbullying perpetration and victimization, with the former effect becoming more pronounced during the COVID-19 pandemic. Exposure to violent media increased both cyberbullying perpetration and victimization, but the effect on perpetration decreased during the COVID-19 pandemic. The well-established correlation between internet access and both cyberbullying perpetration and victimization remained unaffected by COVID-19. Surprisingly, adolescents with siblings were less likely to become victims or perpetrators of school-related cyberbullying, irrespective of the pandemic.

Conclusions: In hindsight, COVID-19, functioning as a kind of natural experiment, has provided us with a unique opportunity to examine the effects of a global event, forcing major behavioral changes on the persistent challenge of cyberbullying in middle schools.

(J Med Internet Res 2025;27:e70508) doi:10.2196/70508

KEYWORDS

cyberbullying and cyber victimization among adolescents; COVID-19; panel study; longitudinal data analysis; parental communication; exposure to violent media content

Introduction

Overview

The global issue of adolescents' aggressive behavior has long been acknowledged as a significant concern, predating the outbreak of COVID-19. This pervasive issue has garnered considerable attention in the literature, with studies highlighting its profound impact on the well-being of young individuals [1-6]. With the advent of digital technologies and the widespread use of social media platforms, adolescents are increasingly

XSL•FO
RenderX

exposed to online harassment, encompassing various forms of aggressive behaviors such as spreading rumors, sending threatening messages, or posting hurtful content. The consequences of cyberbullying extend beyond the digital realm, negatively affecting adolescents' mental health, self-esteem, and overall psychosocial development. Research consistently links cyberbullying to heightened levels of anxiety, depression, and decreased self-esteem [7], as well as emotional distress among victims [3]. Moreover, the persistent nature of online harassment can lead to feelings of social isolation and a decline in academic performance [8-10]. The long-term consequences may manifest in a reluctance to engage in social interactions, thereby hindering the development of healthy peer relationships [11].

With the onset of COVID-19 and the subsequent surge in remote learning and screen time during the pandemic, concerns have emerged regarding the potential escalation of cyberbullying incidents. Within just 1 year after March 11, 2020, when the World Health Organization declared COVID-19 a pandemic, the first studies were published comparing the prevalence of cyberbullying perpetration and victimization before and during COVID-19. In 2023, the first systematic review [12] and meta-analysis on this comparison were published [13]. As summarized by Vaillancourt et al [2] in a recent overview, the results are contradictory, with studies reporting an increased, decreased, or stable trend in cyberbullying. Most of the studies involve comparing 2 cross-sectional surveys, one conducted before and one during the pandemic, with only a few comparisons based on panel data.

The objectives of this study are 3-fold. First, we aim to investigate whether and, if so, in which direction the trend of cyberbullying evolves with age. The second objective delves into several factors associated with changes in this trend. Specifically, we aim to explore the role of factors such as exposure to violent media content, the quality of parental communication, access to the internet, as well as the influence of sex and sibling relationships. The third objective is to assess the extent to which changes in the trend can be attributed to COVID-19: our unique time-series data enable us to do this by comparing the same adolescents in periods before and during the COVID-19 pandemic. The data collection for this study began in September 2017 and concluded in the summer of 2021, spanning 6 waves. While the richness and longitudinal nature of such a comprehensive panel study are clear advantages, they also present challenges: the data, particularly from the early waves, may appear outdated, and organizing such a large dataset requires substantial effort, making it challenging to publish findings immediately after data collection is completed.

Literature Review and Formulation of Hypotheses

We review the literature on cyberbullying perpetration and victimization, with particular attention to those covering the COVID-19 period, to derive a set of hypotheses to achieve the objectives listed above.

Cyberbullying and COVID-19

To understand the impact of COVID-19 on the prevalence of aggressive behavior among adolescents, it is necessary to examine trends predating the pandemic. Cyberbullying perpetration is intrinsically related to cyberbullying victimization, which refers to the experience of being targeted and harmed by cyber aggression—those who engage in cyberbullying may themselves become victims of retaliatory behaviors [14,15].

Several studies on bullying patterns suggest that cyberbullying was already a concern before COVID-19 [16-18]. With the advent of digital technology and social media platforms, there has been a gradual increase in the incidence of cyberbullying. A study conducted in the United States spanning the period 2010-2017 reported a consistent growth in cyberbullying [19]. However, longitudinal studies tracking adolescents over periods longer than 6 to 12 months are scarce, limiting our understanding of the factors contributing to this upward trend [15,20]. Adolescents are believed to be particularly vulnerable to cyberbullying due to their frequent use of smartphones and social networking sites, which provide perpetrators with anonymity and easy access to their targets. Research conducted in the years preceding the pandemic underscored the detrimental effects on adolescents' mental health, academic performance, and social relationships. A systematic review of cyberbullying studies conducted worldwide from the beginning of 2015 to December 2019 suggests a continuous increase—prevalence rates of cyberbullying victimization ranged from 6% to 46.3% in 2015, increasing to 13.99% to 57.55% in 2019 [21].

The role of age in adolescent cyberbullying presents mixed evidence. While certain studies have argued that both perpetration and victimization occur equally across different age groups [22], others have found that cyberbullying is particularly prevalent in middle schools, with incidents increasing after the fifth grade and peaking in the eighth grade [5,23,24]. One explanation for the mixed evidence might be that most studies are either cross-sectional or based on repeated survey designs.

There is an ongoing controversy regarding whether cyberbullying perpetration and victimization went up, remained stable, or decreased during the lockdown caused by COVID-19. While several studies have found an increase [18,25-29], a UNICEF report from 2020 indicates a 17% decrease in cyberbullying during the lockdown. Other studies report not only an overall decrease in cyberbullying perpetration, but also a decrease in victimization, at least for those adolescents with moderate social media use [30].

Besides the question of whether a real change occurred in both periods, and if so, in which direction, another query arises: if such a change did occur, is it due directly to the pandemic or to a shift in the main factors associated with cyberbullying before and during the pandemic? To address this issue, we include COVID-19 as an indicator variable to enable testing for interaction effects.

Considering the lack of longitudinal studies, it is not surprising that Vallaincourt et al [2] conclude that due to the predominantly cross-sectional design of existing studies, it is not possible to make definitive statements about true change. Given the inconclusive evidence regarding the impact of COVID-19 and the role of age, we formulate the following hypothesis:

XSL•FO
RenderX

Cyberbullying perpetration and victimization initially increase with age and then stabilize. The observed relationship is unchanged before and after COVID-19. [H1]

Use of Information and Communication Technologies and Exposure to Violent Media Content

The use of information and communication technologies has long been recognized as a notable risk factor linked to cyberbullying perpetration and victimization [11,31]. Particularly for adolescents aged 12 to 17 years, social networking sites provide a platform where approximately 85% have reported witnessing negative interactions [32]. The amount of time adolescents spend using information and communication technologies represents another crucial factor [14,33], in particular the time of exposure to antisocial media content which, in turn, predicts that an adolescent is more likely to engage in cyberbullying over time [34]. A recent systematic review and meta-analysis comprising 46 studies found a remarkably strong increase in screen time of nearly 1.5 hours per day compared with pre–COVID-19 levels, constituting a surge of approximately 52%. Adolescents aged 12 to 18 years contributed significantly to this trend [35]. Applications like TikTok, Pinterest, Reddit, Facebook, Snapchat, Instagram, LinkedIn, and Twitter have seen an increase in active users, ranging from 8% to 38% [36]. Additionally, a study analyzing data from Twitter users in 16 countries and information gathered from Reddit in one country highlights a substantial growth in abusive content on both platforms [37]. Based on this, we formulate the next 2 hypotheses.

Increased access to the internet increases both the perpetration and victimization of cyberbullying. The observed relationship is unchanged before and after COVID-19. [H2]

Increased exposure to violent media content increases both perpetration and victimization of cyberbullying. The observed relationship is unchanged before and after COVID-19. [H3]

Parental Communication, Siblings, and Sex

Meta-analytic research has demonstrated that a positive family environment serves as a protective factor against bullying (and cyberbullying) perpetration and victimization [38-40]. Conversely, a perceived lack of quality in parental communication has been linked to instances of adolescent bullying and cyberbullying [38,41-43].

Some scholars posit that the shift to remote learning during COVID-19, combined with a stressful economic environment, has heightened the burden on families as they navigate the challenges of childcare, education, and professional challenges [13,44]. The resulting stress may negatively impact the quality of parental communication [45]. We thus formulate the following hypothesis:

Higher quality of parental communication decreases both the perpetration and victimization of cyberbullying. The observed relationship is unchanged before and after COVID-19. [H4]

The impact of siblings on cyberbullying has not really been researched. Existing studies focus on bullying between siblings, rarely on the impact of siblings on bullying by schoolmates. For instance, Borualogo and Casas [46], in their study involving young adolescents aged 10-12 years in Indonesia, concluded that physical bullying between siblings increased during COVID-19 compared with prepandemic levels, while incidents of school bullying notably decreased. Given this lack of prior studies, we formulate the following explorative hypothesis:

Having siblings is a risk factor for cyberbullying perpetration and victimization at school. The observed relationship is unchanged before and after COVID-19. [H5]

Concerning sex, certain studies [22,47] argue that this does not predict cyberbullying victimization, while others [1,30,48,49] conclude that boys report more perpetration and girls report more victimization. We formulate the following hypothesis:

There is a higher rate of cyberbullying perpetration and a lower rate of victimization among boys compared to girls; the difference persists with aging. The observed relationship is unchanged before and after COVID-19. [H6]

Methods

Sample Participants and Procedure

Data for this research are drawn from a 6-wave panel survey conducted among middle school students in the Canton of Ticino, Switzerland, between 2017 and 2021. Ticino is the southernmost canton of Switzerland and the only one with Italian as the official language. It is considered one of the most prosperous regions in Switzerland and Europe [50]. On a global scale, Switzerland falls within the average range for bullying incidents in compulsory schools, as reported by the Programme for International Students Assessment (Organisation for Economic Co-operation and Development; 2019). Exactly 1 month after the World Health Organization had declared COVID-19 a pandemic on March 11, 2020, all middle schools of the Canton of Ticino were closed, and students transitioned to online learning modalities. Two months later, in May 2020, all students in compulsory school years returned to offline schooling.

At the start of the study in September 2017, a random sample consisting of 66% of all first-year public middle school classes across the 5 areas of the canton was selected. Private schools with more than 2 first-year classes were also invited to participate, and one of the 3 schools that met this criterion accepted the invitation. From this school, one of the 2 first-year classes was randomly selected for inclusion. Thus, the research encompassed 101 classes randomly chosen from 1 private and 35 public middle schools.

Data were collected through self-administered, paper-and-pencil questionnaires completed in class under the supervision of a teacher. The data from the 6 waves were matched using a unique identifier linked to each student's name, which only collaborating school staff accessed for survey distribution purposes. To protect participants' privacy, teachers were

XSL•FO
RenderX

instructed to place all completed questionnaires in an envelope and seal it in front of the students. The remaining missing data were due to absences on the day of data collection or students moving out of the canton. Students who changed schools were retained in our sample as long as they moved to another school and class included in our sample. However, if they transferred to a school outside the selected sample, they were no longer included in the study.

For this research, data collected on 6 different occasions were analyzed: November and December 2017 (T1), May and June 2018 (T2), November and December 2018 (T3), May and June 2019 (T4), October and November 2020 (T5), and May and June 2021 (T6). The observation time thus ranged from the beginning of the first grade to the end of the fourth grade. The average age was 13.22 (SD 1.33) years. At time point T1, ages ranged from 10.71 to 13.94 years, and at T6, from 14.21 to 17.44 years.

The first 4 waves took place before the COVID-19 pandemic, as initially scheduled. The last 2 had to be delayed due to the closure of schools in May 2020. The final random sample included 2052 students who were invited to participate in T1. Of these 2052 students, a total of 2022 (98.54%) participated at T1, 1879 (91.57%) at T2, 1896 (92.4%) at T3, 1865 (90.89%) at T4, 1871 (91.18%) at T5, and 1869 (91.08%) at T6.

Ethical Considerations

The survey was conducted with the approval and collaboration of the Cantonal Department for School Affairs (Divisione della scuola e l'Ufficio dell'insegnamento) and involved a representative sample of students. In addition to authorization from the cantonal school authorities, parents were informed about the study and given the option to exclude their children (passive consent). Less than 4% of students (n=75) did not participate due to parental objections. The institutional research board of the University della Svizzera Italiana (USI) did not consider the study to require approval, provided that the school authorities in the Canton of Ticino and the parents of the adolescents had approved the study. Informed consent was obtained from all adolescents prior to their participation in the study (29-01-2019Ls). The consent process included a clear explanation of the study's purpose, procedures, potential risks, and benefits, as well as assurances of confidentiality and the voluntary nature of participation. The adolescents were informed that they could stop filling in the questionnaire at any time. The study ensures the privacy and confidentiality of participants by maintaining all collected data in an anonymous or deidentified format. No personal identifiers are linked to the data. All study data are stored on servers with access restricted to authorized research personnel only, ensuring high level of privacy and confidentiality for participants. Participants did not receive any compensation for their participation in the study.

Variable Selection

Table 1 provides an overview of all variables used in our analysis; it is followed by a description of each measure.

Table 1. Variable description.

Variable	Scale	Description
Dependent variables		
Cybperp	1 (never) to 4 (often)	Perception of cyberbullying perpetration
Cybvic	1 (low) to 4 (high)	Perception of cyberbullying victimization
Independent variables		
Medexp	1 (never) to 5 (very often)	Exposure to violent media content
Parqual	1 (disagree) to 5 (agree)	Perceived quality of communication with parents
Siblings	0 or 1	Indicator variable: only child or 1 or more siblings
Age	Years	Age of participants
COVID	0 or 1	Indicator variable: before or after COVID-19
Sex	0 or 1	Indicator variable: male or female
Device	0 or 1	Indicator variable: no access (0) or access (1) to a device connected to internet
Control variables		
Parents	0 or 1	Indicator variable: child lives with none or one (0) or both (1) parents
Econf	1 (not at all well) to 5 (very well)	Self-reported (by students) economic status of household

Perpetration of Cyberbullying

Perpetration of cyberbullying (abbreviated as "Cybperp") was assessed on all 6 occasions using 6 of the 11 items from the European Cyberbullying Intervention Project Questionnaire [51] to evaluate the perpetration of aggressive acts online. For more details regarding the multi-item measurement scale, refer to Bullo and Schulz [34]. Measurement invariance of cyber peer aggression was tested in a structural equation modeling framework using confirmatory factor analysis, which demonstrated that all 4 levels of measurement invariance (from configural factorial over weak and strong factorial to strict factorial invariance [52]) can be assumed for the data. Additional details are provided in Bullo and Schulz [34]. For each item,

XSL•FO
RenderX

participants were asked to indicate how often—in the previous 6 months—they had engaged in these behaviors. The response scale went from "Never" (1) to "Often" (4). McDonald ω was used to assess the reliability of the measure, which ranged between 0.70 and 0.88.

Cyberbullying Victimization

Cyberbullying victimization (abbreviated as "Cybvic") was measured using 6 items selected from the European Cyberbullying Intervention Project Questionnaire project. Participants reported the frequency with which they had been victims of various forms of cyber aggression, including direct verbal aggression (eg, being insulted via texts or web-based messages; being threatened), relational aggression (eg, being excluded in the virtual world; being a victim of badmouthing or gossiping), and media-related aggression (eg, unwanted sharing of embarrassing pictures or videos). The response scale ranged from "Never" (1) to "Often" (4). McDonald ω was used to assess the reliability of the measure, which ranged between 0.84 and 0.88.

Exposure to Violent Media Content

Exposure to violent media content (abbreviated as "Medexp") was assessed by calculating an average score from 6 items regarding how often adolescents have watched media (including television, smartphones, video games, or the internet) depicting people fighting, destroying others' property, stealing, shooting, taking drugs, or drinking alcohol (1=never to 5=very often). McDonald ω was used to assess the reliability of the measure, which ranged from 0.89 to 0.93.

Parental Communication Quality

Parental communication quality (abbreviated as "Parqual") was measured using a slightly modified version of a scale developed by Guilamo-Ramos et al [53], which includes dimensions of perceived parental expertise, trustworthiness, and accessibility. The scale consists of 4 items rated on a 5-point scale from "Totally disagree" (1) to "Totally agree" (5), with higher values indicating higher perceived quality in adolescents talking to their parents. More details are provided in Bullo and Schulz [41]. Again, McDonald ω was applied to assess the reliability of the measure, which ranged from 0.60 to 0.75.

For all 4 multi-item measurement scales (cyberbullying perpetration and victimization, exposure to violent media content, and perceived parental communication quality), the unidimensionality of the measures was confirmed using confirmatory factor analysis [54].

Siblings

Furthermore, adolescents were asked how many brothers and sisters lived with them in their household. We used a 0 or 1 indicator, where 0 indicates an only child, and 1 indicates at least 1 sibling.

Age

This variable represents the adolescents' age, expressed in years.

COVID

This 0 or 1 indicator variable captures for each wave whether it took place before (0) or after (1) the school closure due to the COVID-19 pandemic.

Sex

This is an indicator variable, where 0 indicates male and 1 female.

Access to Communication Devices

Access to communication devices (abbreviated as "Device") was assessed by asking participants whether they possess a cellphone with or without internet access, a tablet or computer with internet access, a PlayStation, and whether they have a television in their room. The measure of access was defined as a 0 or 1 indicator, a value of 1 indicating that the child has access to at least 1 device with internet access.

Parental Status

Parental status (abbreviated as "Parents") was assessed by asking adolescents whether they lived with both, one, or none of their parents. We used a 0 or 1 indicator, where 0 indicates that they live with at most one of their parents, and 1 indicates they live with both parents.

Perceived Economic Status of the Family

The perceived economic status of the family (abbreviated as "Econf") was assessed during the first data collection by asking students to indicate how financially well off their family is on a scale from "Not well at all" (1) to "Very well" (5).

Descriptive Statistics

Table 2 presents a demographic profile of the sample, together with means, SDs, and indices of skewness and kurtosis for key variables used in the formal models we tested.

XSL•FO
RenderX

Table 2. Descriptive statistics. Longitudinal assessment of cyberbullying perpetration, cyberbullying victimization, violent media exposure, perceived quality of parental communication, siblings, age, sex, parental status, internet access, and self-reported economic status of the household.

Statistics	T1	T2	T3	T4	T5	T6
Cyberbullying perpetration (Cybperp)						
Participants, n	1471	1440	1425	1409	1067	988
Mean (SD)	1.10 (0.21)	1.20 (0.36)	1.17 (0.34)	1.25 (0.43)	1.26 (0.39)	1.24 (0.44)
Skewness (SE)	0.21	0.36	0.34	0.43	0.39	0.44
Kurtosis (SE)	3.86	2.84	4.09	2.44	2.60	2.70
Cyberbullying victimization (Cybvic)						
Participants, n	1454	1433	1419	1402	1060	989
Mean (SD)	1.15 (0.31)	1.29 (0.45)	1.19 (0.38)	1.34 (0.53)	1.23 (0.39)	1.28 (0.46)
Skewness (SE)	0.31	0.45	0.38	0.53	0.39	0.46
Kurtosis (SE)	3.56	2.04	3.17	2.03	2.79	2.31
Violent media exposure (Medexp)						
Participants, n	1475	1440	1430	1413	1068	994
Mean (SD)	1.58 (0.82)	2.05 (1.05)	1.95 (1.06)	2.43 (1.16)	2.57 (1.16)	2.74 (1.17)
Skewness (SE)	0.82	1.05	1.06	1.16	1.16	1.17
Kurtosis (SE)	1.76	0.90	1.08	0.51	0.34	0.12
Perceived quality of parental communication (Parqual)						
Participants, n	1441	1445	1434	1414	1059	989
Mean (SD)	2.99 (0.89)	3.03 (0.82)	3.09 (0.81)	3.01 (0.85)	2.85 (0.83)	2.84 (0.83)
Skewness (SE)	0.89	0.82	0.81	0.85	0.83	0.83
Kurtosis (SE)	–0.50	–0.59	–0.71	–0.74	–0.42	–0.41
Siblings						
Participants, n	1480	1457	1448	1423	1070	997
Mean	0.82	0.81	0.80	0.79	0.79	0.79
Age (years)						
Participants, n	1497	1499	1498	1497	1506	1434
Mean (SD)	11.53 (0.40)	12.03 (0.40)	12.53 (0.40)	13.03 (0.40)	14.53 (0.40)	15.03 (0.40)
Sex						
Participants, n	1496	1499	1498	1496	1506	1001
Mean	0.49	0.49	0.49	0.49	0.49	0.50
Parental status (Parents)						
Participants, n	1480	1457	1449	1423	1070	996
Mean	0.79	0.77	0.77	0.75	0.74	0.74
Internet access (Device)						
Participants, n	1339	1215	1197	1231	888	933
Mean	0.80	0.84	0.90	0.93	0.98	0.99
Self-reported economic status of household (Econf)						
Participants, n	1435	1435	1435	1435	1435	1370
Mean (SD)	3.86 (0.70)	3.86 (0.70)	3.86 (0.70)	3.86 (0.70)	3.86 (0.70)	3.86 (0.69)

Statistical Methods

Two mixed-effect models were set up for the 2 dependent variables, namely “Cybperp,” and “Cybvic.” The independent variables were “Age,” “COVID,” “Sex,” “Siblings,” “Medexp,” “Parqual,” and “Device.” For the variable age, a polynomial of order 3 was fitted to test the variation of the effect. The 2-way

XSL•FO
RenderX

interactions between "COVID" and the other variables were included, as well as between "Age" and "Sex." The 3-way interaction between "COVID," "Age," and "Sex" was also included. A random intercept per participant was included in the model to account for the repeated measure framework. Finally, "Econf" and "Parents" were included as control variables. The models were fitted using the statistical software R (R Foundation for Statistical Computing) and its package *lme4* [55]. Multiple comparisons were performed using package *emmeans* [56] and Sidak-adjusted *P* values. Once fitted, a selection of variables was performed by removing the nonsignificant ones according to ANOVA *F* tests. These results are given in Tables S1 and S2 in Multimedia Appendix 1. Missing data cases were discarded. Cases were fixed based on the most complete model before the model selection. For the analysis, complete cases for the selected model's features were used.

Results

Overview

The victimization and perpetration model coefficients are given in Tables S3 and S4 in Multimedia Appendix 1. To test H1 and H6, which involve age, we set the comparisons at the mean age, minus and plus 1 SD, that is, at 11.8, 13.1, and 14.5 years. The results are presented below, per hypothesis, first for victimization and then for perpetration.

Hypothesis 1

Victimization

The hypothesis is validated for female individuals, but only partially for male individuals (Table S5 in Multimedia Appendix 1).

Indeed, for female individuals, the observed increase in victimization by "Age" is significant between 11.8 and 13.1 years (≈0.07, z=3.83, P<.001) and nonsignificant between 13.1 and 14.5 years (≈0.03, z=0.88, P=.62). For male individuals, this increase is significant in both cases (≈0.13, z=7.17, P<.001; ≈0.16, z=4.98, P<.001). As expected, there is no difference between before and after COVID-19, as no interaction involving "COVID" and "Age" is significant.

Perpetration

The hypothesis is validated for female and male individuals (Table S6 in Multimedia Appendix 1). Indeed, the observed increase in perpetration by "Age" is significant between 11.8 and 13.1 years (male: ≈0.04, z=3.53, P<.001; female: ≈0.04, z=3.04, P<.01) and nonsignificant between 13.1 and 14.5 years (male: ≈0.03, z=1.59, P=.21; female: ≈0.00, z=–0.08, P>.99). As expected, there is no difference between before and after COVID-19, as no interaction involving "COVID" and "Age" is significant.

Hypothesis 2

Victimization

The hypothesis is validated as the difference in victimization between with and without access to the internet is significant (≈0.09, z=5.03, P<.001; Table S7 in Multimedia Appendix 1). Without significant interaction between "COVID" and "Device," this effect remains unchanged, as expected.

Perpetration

The hypothesis is validated as the difference in perpetration between with and without access to the internet is significant (≈0.07, z=4.48, P<.001; Table S8 in Multimedia Appendix 1). Without significant interaction between "COVID" and "Device," this effect remains unchanged as expected.

Hypothesis 3

Victimization

The hypothesis is validated. Indeed, the effect of "Medexp" is significant and positive (≈0.06, t_{5791}=11.7, P<.001; Table S9 in Multimedia Appendix 1). Without significant interaction between "Medexp" and "COVID," this effect remains unchanged as expected.

Perpetration

The hypothesis is validated. The effect of "Medexp" is significant before and after COVID-19 (before: ≈0.09, z=16.88, P<.001; after: ≈0.07, z=10.32, P<.001; Table S10 in Multimedia Appendix 1). Although the effect remains significant and positive before and after COVID-19, the difference in the slope is mildly significant (≈0.02, z=2.06, P=.04; Table S11 in Multimedia Appendix 1). Thus, the effect of "Medexp" on cyberbullying perpetration decreased after COVID-19, contrary to our expectations.

Hypothesis 4

Victimization

The hypothesis is validated. The effect of "Parqual" is significant and negative (≈–0.06, t_{5789}=–8.9, P<.001; Table S9 in Multimedia Appendix 1). Without interaction between "COVID" and "Parqual," the relationship did not change after COVID-19.

Perpetration

The hypothesis is validated. Both before and after COVID-19, the effect of "Parqual" is negative and significant (before: ≈–0.02, z=–4.00, P<.001; after: ≈–0.05, z=–5.20, P<.001; Table S12 in Multimedia Appendix 1). In addition, contrary to our expectation, this deterring effect significantly increased after COVID-19 (≈0.03, z=2.34, P=.02; Table S13 in Multimedia Appendix 1).

Hypothesis 5

Victimization

The hypothesis is not validated. The effect of "Siblings" is significant and negative (≈–0.05, t_{2640}=–2.80, P<.01; Table S9 in Multimedia Appendix 1). Contrary to our expectations, the data indicate that having siblings is a protective factor: those with siblings are less likely to become victims of cyberbullying at school compared with those without siblings. As there is no significant interaction between "Siblings" and "COVID," this effect is the same before and after COVID-19.

XSL•FO
RenderX

Perpetration

The hypothesis is not validated. As for victimization, the effect of "Siblings" is significant and negative (≈–0.03, t_{2714}=–2.26, P=.02; Table S14 in Multimedia Appendix 1). Again, contrary to our expectations, the data indicate that having siblings is a deterring factor. Without significant interaction between "Siblings" and "COVID," that effect is the same before and after COVID-19.

Hypothesis 6

Victimization

The hypothesis is partially validated. The expected lower rate for male individuals is mildly significant at early age before COVID-19 (male – female; age 11.8 years; ≈–0.05, z=–2.78, P=.03; Table S15 in Multimedia Appendix 1) but not afterward (male – female; age 11.8 years; ≈–0.19, z=–2.63, P=.05). At any other age, before and after COVID-19, differences were not significant.

Perpetration

The hypothesis is not validated. There is no difference between male and female individuals at any age (Table S16 in Multimedia Appendix 1). In addition, without interaction between "COVID" and "Age" and between "COVID" and "Sex," this nonrelation remains the same before and after COVID-19.

Discussion

Principal Findings

The focus of our study is on the question of whether COVID-19 has changed the relationship between cyberbullying and several of its drivers, among adolescents. We were particularly interested in whether COVID-19 has made these relationships more pronounced, whether they have remained unchanged, or whether they have been mitigated. These questions are not primarily important because of the current demand for an analysis of COVID-19 and its impact on social life [57]. Rather, it is also expected that an unforeseen event like COVID-19 could lead to a better understanding of the factors associated with the increasing challenge of cyberbullying in schools [58]. Ultimately, this also involves the expectation of gaining a better understanding of the phenomenon of cyberbullying among adolescents.

Regarding our first hypothesis, namely the increase in cyberbullying victimization among adolescents with age, our data confirm this effect, known from previous studies, at least for the early middle school years, when adolescents are between 11 and 13 years old. This effect applies equally to girls and boys. However, from the age of 13 years onward, this seems to change, as the increase in cyberbullying victimization among girls is no longer significant, while cyberbullying perpetration among boys continues to grow.

When interpreting these data, it is important to note that the end of the increasing trend in cyberbullying victimization coincides with the onset of COVID-19. Therefore, it is difficult to determine to what extent the end of the increase among girls or the continuing upward trend among boys is attributable to age and how much is due to the onset of COVID-19. Previous studies on cyberbullying victimization have shown that this trend ends at a certain age among adolescents. It is also challenging to comment on the magnitude of this effect. In other words, it is unclear how strong the increase in cyberbullying victimization among boys would have been if COVID-19 had not occurred.

As to our second hypothesis, there is a well-documented correlation, supported by numerous studies [59], between adolescents owning devices that provide internet access and cyberbullying perpetration and victimization: when adolescents have access to a device that allows them to go online, the likelihood of them becoming a victim or perpetrator of cyberbullying rises. This correlation was observed in our study as well, both before and during the outbreak of COVID-19, with no significant change in the strength of this relationship during the pandemic.

It is important to consider that even without personal internet-enabled devices, adolescents can still fall victim to or engage in cyber aggression, for example, by using internet access on school computers. Moreover, by the end of middle school, only a small number of students in our study still lacked a device that enabled them to access the internet. Nevertheless, the lack of a personal internet-enabled device remains a protective factor, regardless of the age at which adolescents acquire such a device.

Regarding the relationship between exposure to violent media content and the perpetration or victimization of cyberbullying among adolescents, previous meta-analytical studies [14,20,24,60,61] provide strong evidence supporting this link. Children highly exposed to violent media content ("Medexp"=5) have on average a victimization score of 0.25 larger than nonexposed children ("Medexp"=1). This is a major increase, as the average "Cybvic" score ranges between 1.15 and 1.34 for the different data-collection waves. The impact of violent media exposure is even larger for the perpetrators: a score increase of 0.33 and 0.28 before and during COVID-19, respectively, when "Medexp" increases from 1 to 5, compared with average scores ranging between 1.10 and 1.26.

The debate often centers on how violent media content leads to such behaviors, whether through imitation or other mechanisms [20,61]. Therefore, it is not surprising that our study also confirms this association. Longitudinal studies do not provide insights into causal relationships between exposure to violent media content and cyberbullying perpetration or victimization. However, it is notable that the effect of violent media exposure on cyberbullying remains substantial and persistent over the 4 years of our study. Additionally, this effect appears to diminish during COVID-19. This reduction cannot be explained by the age-related decrease in cyberbullying tendencies discussed in hypothesis 1, but rather seems to be an interaction effect that lessens during the pandemic. In this sense, one might consider it a positive effect of COVID-19: the pandemic has reduced the impact of exposure to violent media content on aggressive behavior. The exact reasons for this reduction in the relationship between exposure to violent media and cyberbullying

XSL•FO
RenderX

victimization or perpetration, before and during COVID-19, remain unclear, but could be understood through Social Learning Theory: during COVID-19, changes in social contexts (eg, increased time spent at home and less face-to-face peer interaction) might have altered how adolescents process and enact behaviors learned from violent media. The reduced opportunities for social interaction and confrontation in physical spaces may have diminished the translation of media-inspired aggression into actual cyberbullying acts.

Hypotheses 2 and 3 are not unrelated, as access to internet and exposure to violent media are intrinsically linked. The finding that possessing an internet-enabled device consistently increases the likelihood of adolescents becoming victims or perpetrators of cyberbullying, with no significant change before and during COVID-19, can be explained by the general aggression model [24,62-64], which posits that exposure to aggressive stimuli (including violent media or environments) can increase aggressive thoughts, feelings, and behaviors. Access to the internet via a device provides a platform where aggressive behavior can be exhibited. The presence of a device enables easier access to both aggressive content and opportunities for cyberbullying, maintaining a consistent risk factor for both victimization and perpetration regardless of external circumstances like the pandemic.

Another possible explanation comes from Cohen and Felson's [65] routine activity theory [66], which suggests that the changes in routine activities, such as increased parental supervision, reduced peer interaction, and changes in web-based activities, may have disrupted the conditions that facilitate cyberbullying perpetration. The decrease in the effect of violent media on perpetration could be because adolescents were less motivated to act aggressively or had fewer opportunities to engage in cyberbullying due to increased monitoring and altered social dynamics.

Our hypothesis 4, namely the assumption that a higher quality relationship with parents serves as a protective factor for adolescents, making them less likely to fall victim to or become a perpetrator of cyber aggression, has largely been confirmed. Interestingly, the protective effect of positive parental communication in preventing adolescents from becoming victims of cyber aggression is comparable in magnitude to the harmful impact of exposure to violent media content. On average, an increase of 1 unit in exposure to violent media content can be offset by an equivalent increase in the perceived positive quality of the relationship between adolescents and their parents.

What is particularly noteworthy, however, is that a positive relationship with parents not only acts as a protective factor against becoming a victim of cyber aggression but also seems to have a (more limited) deterrent effect on adolescents' aggressive behavior toward others in cyberspace. This deterrent effect, although inferior to the impact of exposure to violent media, doubled during the COVID-19 crisis compared with the period before the pandemic.

We suggest that this finding can be explained through various theoretical perspectives. First, social control theory [67] posits that strong social bonds, such as those with family, serve as a deterrent to deviant behavior, including cyberbullying. During the pandemic, adolescents may have experienced increased stress, uncertainty, and isolation. In such a context, the role of positive communication with parents likely became more crucial in maintaining social control and discouraging deviant behavior. The intensification of family interactions due to lockdowns and restrictions may have strengthened these social bonds, leading adolescents to rely more on their parents for emotional support and guidance, thereby enhancing the protective effect against cyberbullying.

Another explanation comes from stress and coping theory, which posits that individuals cope with stress in various ways, including seeking support from social relationships [68]. Adolescents who could communicate positively with their parents may have experienced better emotional regulation and lower stress levels, reducing the likelihood of resorting to cyberbullying as a maladaptive coping strategy [69].

Regarding our fifth hypothesis, the results were contrary to our expectations: the data indicate that having siblings acts as a protective factor. Adolescents with siblings are less likely to become victims of cyberbullying at school than those without siblings. The magnitude of this protective effect is slightly lower than that observed in our previous hypotheses, suggesting that having siblings is, on average, equivalent to improving parental communication or decreasing exposure to violent media content by a bit less than 1 point on their respective scales. Similarly, our data show that having siblings also acts as a deterrent against becoming a cyberbully, although to a lesser extent, comparable to an increase of 1 point in the quality of parental communication, but much less than a decrease of 1 point in media exposure. We found no change in the effect of having siblings before and after the COVID-19 pandemic.

One possible explanation for these findings comes from the buffering hypothesis in social support theory [70,71], which posits that social relationships can buffer the impact of stressful events. Within this framework, siblings can offer emotional support, companionship, and a sense of belonging, which may protect against the isolation that can make a child more vulnerable to bullying [72].

Finally, based on findings from similar studies [73-75], we hypothesized that boys would exhibit a higher rate of cyberbullying perpetration, and a lower rate of victimization, compared with girls, with these differences persisting as they age. However, the lower rate of victimization among boys was only observed in the very early stage when they were around 11 years old. As they aged, the difference in victimization rates between boys and girls diminished and had already disappeared before the onset of COVID-19. Regarding the perpetration rate, no significant difference was found between male and female individuals.

Concluding Remarks

In summary, we can conclude that COVID-19 likely had a limited overall impact on both cyberbullying perpetration and victimization. Moreover, while certain factors were identified as either increasing or decreasing the likelihood of

XSL•FO
RenderX

cyberbullying, COVID-19 only partially influenced these relationships.

However, in certain cases, this impact is noteworthy. On the one hand, the perceived positive relationship with parents reduces both the risk of becoming a victim of cyberbullying and the likelihood of exhibiting aggressive behavior toward other adolescents, with the latter effect becoming more pronounced during COVID-19 compared with the period before. Notably, a positive relationship with parents served as a protective factor, reducing both the risk of cyberbullying victimization and the likelihood of engaging in cyberbullying. This protective effect was even stronger during the pandemic. These findings reaffirm previous research highlighting the role of family in mitigating cyberbullying risks [44] and emphasize its importance during times of crisis, such as COVID-19.

On the other hand, exposure to violent media was associated with increased rates of both cyberbullying perpetration and victimization, but interestingly, the effect of cyberbullying perpetration lessened during the pandemic. This suggests that while increased violent media consumption during an unexpected crisis may contribute to cyberbullying victimization, it does not necessarily translate into a rise in web-based aggressive behaviors. Additionally, an unexpected finding of our study revealed that adolescents with siblings were less likely to be involved as either victims or perpetrators of cyberbullying, independent of the pandemic's influence.

The question of whether COVID-19 impacted cyberbullying behavior has been widely discussed in numerous studies. A recent systematic review and meta-analysis [13] concluded that, based on pooled prevalence rates, the pandemic led to a significant reduction in overall cyberbullying perpetration. However, as the authors caution, the lack of longitudinal studies limits the ability to draw definitive conclusions about the underlying factors driving these changes. Only a few studies, such as those by Huang et al [13] and Vaillancourt et al [2], have compared cyberbullying behaviors in the same group of participants before and during COVID-19, and even then, these studies involved only 2 or 3 waves of data collection.

In contrast, our study, which spans 6 waves—4 before and 2 during the pandemic—offers a more comprehensive picture. By tracking the same participants over time, we were able to estimate both the baseline rates of cyberbullying perpetration and victimization and the changes in these behaviors before and during the pandemic. Furthermore, we examined key contributing factors that either increase the likelihood of cyberbullying or act as protective measures. A particular strength of our panel study is the consistent use of the same measures for both cyberbullying outcomes and predictors across all 6 waves, allowing for a more robust analysis of trends over time.

Limitations

Our study has several limitations due to the specific nature of our dataset. First, the sudden outbreak of the COVID-19 pandemic disrupted our planned data collection. While we were able to gather data from students at the start and end of each school year during the 4 pre–COVID-19 phases, this regularity was not maintained once the pandemic began. Data collection only resumed when students returned to school, resulting in an unbalanced distribution of waves between the prepandemic and postpandemic periods.

Second, our analysis is at a cohort level. We did not examine whether cyberbullying victims and perpetrators belong to specific groups, nor did we analyze the extent to which victims and perpetrators might be the same individuals, possibly at different points in time.

Finally, a strength of our dataset is that it is a panel study, following the same people over several years. The study includes 4 pre–COVID-19 waves (participants aged roughly 11-15 years) and 2 post–COVID-19 waves (participants aged roughly 14-17 years). The drawback is that we have few older adolescents before COVID-19 and few younger ones after COVID-19, which may be the cause of the lack of significance (large CIs) of certain age-related hypotheses.

In hindsight, COVID-19, functioning as a kind of natural experiment, has provided us with a unique opportunity to examine the effects of a global event, forcing major behavioral changes, on the persistent challenge of cyberbullying in middle schools.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 29 KB - jmir_v27i1e70508_app1.docx]

References

1. Vaillancourt T, Brittain H, Krygsman A, Farrell AH, Landon S, Pepler D. School bullying before and during COVID-19: results from a population-based randomized design. Aggress Behav 2021;47(5):557-569 [FREE Full text] [doi: 10.1002/ab.21986] [Medline: 34235773]

XSL•FO
RenderX

2. Vaillancourt T, Farrell AH, Brittain H, Krygsman A, Vitoroulis I, Pepler D. Bullying before and during the COVID-19 pandemic. Curr Opin Psychol 2023;53:101689 [FREE Full text] [doi: 10.1016/j.copsyc.2023.101689] [Medline: 37690185]
3. Gao L, Liu J, Yang J, Wang X. Longitudinal relationships among cybervictimization, peer pressure, and adolescents' depressive symptoms. J Affect Disord 2021;286:1-9 [FREE Full text] [doi: 10.1016/j.jad.2021.02.049] [Medline: 33647782]
4. Turliuc MN, Măirean C, Boca-Zamfir M. The relation between cyberbullying and depressive symptoms in adolescence. the moderating role of emotion regulation strategies. Comput Hum Behav 2020;109:106341 [FREE Full text] [doi: 10.1016/j.chb.2020.106341]
5. Hinduja S, Patchin JW. Cyberbullying: an exploratory analysis of factors related to offending and victimization. Deviant Behav 2008;29(2):129-156 [FREE Full text] [doi: 10.1080/01639620701457816]
6. Klomek AB, Sourander A, Elonheimo H. Bullying by peers in childhood and effects on psychopathology, suicidality, and criminality in adulthood. Lancet Psychiatry 2015;2(10):930-941 [FREE Full text] [doi: 10.1016/s2215-0366(15)00223-0]
7. Palermiti AL, Servidio R, Bartolo MG, Costabile A. Cyberbullying and self-esteem: an Italian study. Comput Hum Behav 2017;69:136-141 [FREE Full text] [doi: 10.1016/j.chb.2016.12.026]
8. Lee E, Schulz PJ, Lee HE. The impact of COVID-19 and exposure to violent media content on cyber violence victimization among adolescents in South Korea: national population-based study. J Med Internet Res 2024;26:e45563 [FREE Full text] [doi: 10.2196/45563] [Medline: 38517467]
9. Korchmaros JD, Mitchell KJ, Ybarra ML. Technology-based interpersonal victimization: predictors of patterns of victimization over time. J Interpers Violence 2014;29(7):1297-1317 [FREE Full text] [doi: 10.1177/0886260513506277] [Medline: 24269990]
10. Wright MF. Adolescents' cyber aggression perpetration and cyber victimization: the longitudinal associations with school functioning. Soc Psychol Educ 2015;18(4):653-666 [FREE Full text] [doi: 10.1007/s11218-015-9318-6]
11. Orben A, Tomova L, Blakemore SJ. The effects of social deprivation on adolescent development and mental health. Lancet Child Adolesc Health 2020;4(8):634-640 [FREE Full text] [doi: 10.1016/s2352-4642(20)30186-3]
12. Sorrentino A, Esposito A, Acunzo D, Santamato M, Aquino A. Onset risk factors for youth involvement in cyberbullying and cybervictimization: a longitudinal study. Front Psychol 2022;13:1090047 [FREE Full text] [doi: 10.3389/fpsyg.2022.1090047] [Medline: 36743647]
13. Huang N, Zhang S, Mu Y, Yu Y, Riem MME, Guo J. Does the COVID-19 pandemic increase or decrease the global cyberbullying behaviors? a systematic review and meta-analysis. Trauma Violence Abuse 2024;25(2):1018-1035 [FREE Full text] [doi: 10.1177/15248380231171185] [Medline: 37177992]
14. Chen L, Ho SS, Lwin MO. A meta-analysis of factors predicting cyberbullying perpetration and victimization: from the social cognitive and media effects approach. New Media Soc 2016;19(8):1194-1213 [FREE Full text] [doi: 10.1177/1461444816634037]
15. Camerini AL, Marciano L, Carrara A, Schulz PJ. Cyberbullying perpetration and victimization among children and adolescents: a systematic review of longitudinal studies. Telemat Inform 2020;49:101362 [FREE Full text] [doi: 10.1016/j.tele.2020.101362]
16. Jones LM, Mitchell KJ, Finkelhor D. Online harassment in context: trends from three youth internet safety surveys (2000, 2005, 2010). Psychol Violence 2013;3(1):53-69 [FREE Full text] [doi: 10.1037/a0030309]
17. Kessel Schneider S, O'Donnell L, Smith E. Trends in cyberbullying and school bullying victimization in a regional census of high school students, 2006-2012. J Sch Health 2015;85(9):611-620 [FREE Full text] [doi: 10.1111/josh.12290] [Medline: 26201758]
18. Trompeter N, Jackson E, Sheanoda V, Luo A, Allison K, Bussey K. Cyberbullying prevalence in Australian adolescents: time trends 2015-2020. J Sch Violence 2022;21(3):252-265 [FREE Full text] [doi: 10.1080/15388220.2022.2075881]
19. Kennedy RS. Bullying trends in the United States: a meta-regression. Trauma Violence Abuse 2021;22(4):914-927 [FREE Full text] [doi: 10.1177/1524838019888555] [Medline: 31789118]
20. Marciano L, Schulz PJ, Camerini AL. Cyberbullying perpetration and victimization in youth: a meta-analysis of longitudinal studies. J Comput Mediat Commun 2020;25(2):163-181 [FREE Full text] [doi: 10.1093/jcmc/zmz031]
21. Zhu C, Huang S, Evans R, Zhang W. Cyberbullying among adolescents and children: a comprehensive review of the global situation, risk factors, and preventive measures. Front Public Health 2021;9:634909 [FREE Full text] [doi: 10.3389/fpubh.2021.634909] [Medline: 33791270]
22. Tokunaga RS. Following you home from school: a critical review and synthesis of research on cyberbullying victimization. Comput Hum Behav 2010;26(3):277-287 [FREE Full text] [doi: 10.1016/j.chb.2009.11.014]
23. Kowalski RML, Susan, WA. Patricia. Cyberbullying: Bullying in the Digital Age. Hoboken: John Wiley & Sons; 2012.
24. Kowalski RM, Giumetti GW, Schroeder AN, Lattanner MR. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. Psychol Bull 2014;140(4):1073-1137 [FREE Full text] [doi: 10.1037/a0035618] [Medline: 24512111]
25. Barlett CP, Simmers MM, Roth B, Gentile D. Comparing cyberbullying prevalence and process before and during the COVID-19 pandemic. J Soc Psychol 2021;161(4):408-418 [FREE Full text] [doi: 10.1080/00224545.2021.1918619] [Medline: 34128768]

XSL•FO
RenderX

26. Patchin JW, Hinduja S. Cyberbullying among Asian American youth before and during the COVID-19 pandemic. J Sch Health 2023;93(1):82-87 [FREE Full text] [doi: 10.1111/josh.13249] [Medline: 36221854]
27. Morales-Arjona I, Pastor-Moreno G, Ruiz-Pérez I, Sordo L, Henares-Montiel J. Characterization of cyberbullying victimization and perpetration before and during the COVID-19 pandemic in spain. Cyberpsychol Behav Soc Netw 2022;25(11):733-743 [FREE Full text] [doi: 10.1089/cyber.2022.0041] [Medline: 36125383]
28. Choi YJ, Shin SY, Lee J. Change in factors affecting cyberbullying of Korean elementary school students during the COVID-19 pandemic. Int J Environ Res Public Health 2022;19(17):11046 [FREE Full text] [doi: 10.3390/ijerph191711046] [Medline: 36078772]
29. Forsberg JT, Thorvaldsen S. The severe impact of the COVID-19 pandemic on bullying victimization, mental health indicators and quality of life. Sci Rep 2022;12(1):22634 [FREE Full text] [doi: 10.1038/s41598-022-27274-9] [Medline: 36587112]
30. Wang C, Gao T, Cheng X, Li B. Social media use, cyber victimization, and adjustment during COVID-19 virtual learning: a short-term longitudinal study among Chinese middle school students. Sch Psychol 2024;39(2):176-186 [FREE Full text] [doi: 10.1037/spq0000582] [Medline: 37883009]
31. Odgers CL, Jensen MR. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. J Child Psychol Psychiatry 2020;61(3):336-348 [FREE Full text] [doi: 10.1111/jcpp.13190] [Medline: 31951670]
32. Kowalski RM, Limber SP, McCord A. A developmental approach to cyberbullying: prevalence and protective factors. Aggress Violent Behav 2019;45:20-32 [FREE Full text] [doi: 10.1016/j.avb.2018.02.009]
33. Shapka JD, Onditi HZ, Collie RJ, Lapidot-Lefler N. Cyberbullying and cybervictimization within a cross-cultural context: a study of canadian and tanzanian adolescents. Child Dev 2018;89(1):89-99 [FREE Full text] [doi: 10.1111/cdev.12829] [Medline: 28523643]
34. Bullo A, Schulz PJ. Do peer and parental norms influence media content-induced cyber aggression? Comput Hum Behav 2022;129:107136 [FREE Full text] [doi: 10.1016/j.chb.2021.107136]
35. Madigan S, Eirich R, Pador P, McArthur BA, Neville RD. Assessment of changes in child and adolescent screen time during the COVID-19 pandemic: a systematic review and meta-analysis. JAMA Pediatr 2022;176(12):1188-1198 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.4116] [Medline: 36342702]
36. Number of internet users worldwide from 2005 to 2024. Statista. URL: https://www.statista.com/statistics/273018/; [accessed 2024-10-28]
37. Babvey P, Capela F, Cappa C, Lipizzi C, Petrowski N, Ramirez-Marquez J. Using social media data for assessing children's exposure to violence during the COVID-19 pandemic. Child Abuse Negl 2021;116(Pt 2):104747 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104747] [Medline: 33358281]
38. Ding Y, Li D, Li X, Xiao J, Zhang H, Wang Y. Profiles of adolescent traditional and cyber bullying and victimization: the role of demographic, individual, family, school, and peer factors. Comput Hum Behav 2020;111:106439 [FREE Full text] [doi: 10.1016/j.chb.2020.106439]
39. Cook CR, Williams KR, Guerra NG, Kim TE, Sadek S. Predictors of bullying and victimization in childhood and adolescence: a meta-analytic investigation. Sch Psychol Q 2010;25(2):65-83 [FREE Full text] [doi: 10.1037/a0020149]
40. Zych I, Ortega-Ruiz R, Del Rey R. Systematic review of theoretical studies on bullying and cyberbullying: facts, knowledge, prevention, and intervention. Aggress Violent Behav 2015;23:1-21 [FREE Full text] [doi: 10.1016/j.avb.2015.10.001]
41. Bullo A, Schulz PJ. Parent-child communication, social norms, and the development of cyber aggression in early adolescence. J Youth Adolesc 2022;51(9):1774-1786 [FREE Full text] [doi: 10.1007/s10964-022-01625-1] [Medline: 35595923]
42. Hemphill SA, Tollit M, Kotevski A, Heerde JA. Predictors of traditional and cyber-bullying victimization: a longitudinal study of Australian secondary school students. J Interpers Violence 2015;30(15):2567-2590. [doi: 10.1177/0886260514553636] [Medline: 25315480]
43. Kretschmer T, Sentse M, Meeus W, Verhulst FC, Veenstra R, Oldehinkel AJ. Configurations of adolescents' peer experiences: associations with parent-child relationship quality and parental problem behavior. J Res Adolesc 2016;26(3):474-491 [FREE Full text] [doi: 10.1111/jora.12206] [Medline: 28581651]
44. Kang J, Kim J, Kim T, Song H, Han J. Experiencing stress during COVID-19: a computational analysis of stressors and emotional responses to stress. Cyberpsychol Behav Soc Netw 2022;25(9):561-570. [doi: 10.1089/cyber.2022.0052] [Medline: 36108282]
45. Stubbs-Richardson M, May DC. Social contagion in bullying: an examination of strains and types of bullying victimization in peer networks. Am J Crim Just 2020;46(5):748-769 [FREE Full text] [doi: 10.1007/s12103-020-09572-y]
46. Borualogo IS, Casas F. Sibling bullying, school bullying, and children's subjective well-being before and during the COVID-19 pandemic in Indonesia. Child Indic Res 2023;16(3):1203-1232 [FREE Full text] [doi: 10.1007/s12187-023-10013-5] [Medline: 36785618]
47. Wong RYM, Cheung CMK, Xiao B. Does gender matter in cyberbullying perpetration? an empirical investigation. Comput Hum Behav 2018;79:247-257 [FREE Full text] [doi: 10.1016/j.chb.2017.10.022]
48. Beckman L, Hagquist C, Hellström L. Discrepant gender patterns for cyberbullying and traditional bullying – an analysis of Swedish adolescent data. Comput Hum Behav 2013;29(5):1896-1903 [FREE Full text] [doi: 10.1016/j.chb.2013.03.010]

XSL•FO
RenderX

49. Cosma A, Bjereld Y, Elgar FJ, Richardson C, Bilz L, Craig W, et al. Gender differences in bullying reflect societal gender inequality: a multilevel study with adolescents in 46 countries. J Adolesc Health 2022;71(5):601-608 [FREE Full text] [doi: 10.1016/j.jadohealth.2022.05.015] [Medline: 35817675]
50. Baechler BJ. Parmi les régions les plus prospères d'Europe. Rapport sur le produit intérieur brut (PIB) romand [among the most prosperous regions of Europe. Report on the gross domestic product (GDP) of French-speaking Switzerland]. Banque cantonale du Valais [Cantonal Bank of Valais]. 2017. URL: https://www.bcvs.ch/fileadmin/user_upload/site/publications/pib-romand/Brochure_PIB_romand_2017.pdf [accessed 2025-02-15]
51. del Rey R, Elipe PE, Ortega-Ruiz R. Bullying and cyberbullying: overlapping and predictive value of the co-occurrence. Psicothema 2012;24(4):608-613. [Medline: 23079359]
52. Widaman KF, Reise SP. Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In: Bryant KJ, Windle M, West SG, editors. The Science of Prevention: Methodological Advances From Alcohol and Substance Abuse Research. Worcester: American Psychological Association; 1997:281-324.
53. Guilamo-Ramos V, Jaccard J, Dittus P. Parental Monitoring of Adolescents: Current Perspectives for Researchers and Practitioners. New York: Columbia University Press; 2010.
54. Jaccard J, Bo A. Prevention science and child/youth development: randomized explanatory trials for integrating theory, method, and analysis in program evaluation. J Soc Soc Work Res 2018 Dec 01;9(4):651-687 [FREE Full text] [doi: 10.1086/701388]
55. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
56. Lenth R, Singmann H, Love J, Buerkner P, Herve M. 2018 Emmeans: estimated marginal means, aka least-squares means. Scientific Research. 2023. URL: https://www.scirp.org/reference/referencespapers?referenceid=3437485 [accessed 2023-03-13]
57. de Figueiredo CS, Sandre PC, Portugal LCL, Mázala-de-Oliveira T, da Silva Chagas L, Raony, et al. COVID-19 pandemic impact on children and adolescents' mental health: biological, environmental, and social factors. Prog Neuropsychopharmacol Biol Psychiatry 2021;106:110171 [FREE Full text] [doi: 10.1016/j.pnpbp.2020.110171] [Medline: 33186638]
58. Subaramaniam K, Kolandaisamy ABJ, Indraah K. Cyberbullying challenges on society. J Posit Sch Psychol 2022;6(2):2174-2184 [FREE Full text]
59. Cheng C, Chan L, Chau CI. Individual differences in susceptibility to cybercrime victimization and its psychological aftermath. Comput Hum Behav 2020;108:106311 [FREE Full text] [doi: 10.1016/j.chb.2020.106311]
60. Barlett CP, Kowalewski DA, Kramer SS, Helmstetter KM. Testing the relationship between media violence exposure and cyberbullying perpetration. Psychol Pop Media Cult 2019;8(3):280-286 [FREE Full text] [doi: 10.1037/ppm0000179]
61. den Hamer AH, Konijn EA. Adolescents' media exposure may increase their cyberbullying behavior: a longitudinal study. J Adolesc Health 2015;56(2):203-208 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.09.016] [Medline: 25620303]
62. Bushman BJ, Anderson CA. Violent video games and hostile expectations: a test of the general aggression model. Pers Soc Psychol Bull 2002;28(12):1679-1686 [FREE Full text] [doi: 10.1177/014616702237649]
63. Ferguson CJ, Dyck D. Paradigm change in aggression research: the time has come to retire the general aggression model. Aggress Violent Behav 2012;17(3):220-228 [FREE Full text] [doi: 10.1016/j.avb.2012.02.007]
64. Allen J, Anderson CA, Bushman BJ. The general aggression model. Curr Opin Psychol 2018;19:75-80 [FREE Full text] [doi: 10.1016/j.copsyc.2017.03.034] [Medline: 29279227]
65. Cohen LE, Felson M. Social change and crime rate trends: a routine activity approach. Am Sociol Rev 1979;44(4):588. [doi: 10.2307/2094589]
66. Hutchings A, Hayes H. Routine activity theory and phishing victimisation: who gets caught in the 'Net'? Curr Issues Crim Justice 2018;20(3):433-452 [FREE Full text] [doi: 10.1080/10345329.2009.12035821]
67. Hirschi T. A control theory of delinquency. In: The Craft of Criminology. New York: Routledge; 2002:1-10.
68. Lazarus RS, Folkman S. Stress, appraisal, and coping. New York: Springer; 1984.
69. Elsaesser C, Russell B, Ohannessian CM, Patton D. Parenting in a digital age: a review of parents' role in preventing adolescent cyberbullying. Aggress Violent Behav 2017;35:62-72 [FREE Full text] [doi: 10.1016/j.avb.2017.06.004]
70. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985;98(2):310-357 [FREE Full text] [doi: 10.1037//0033-2909.98.2.310]
71. Cohen S, McKay G. Social support, stress and the buffering hypothesis: a theoretical analysis. In: In Handbook of psychology and health IV. Abingdon: Routledge; 2020.
72. Wright MF. The buffering effect of parental mediation in the relationship between adolescents? cyberbullying victimisation and adjustment difficulties. Child Abuse Review 2016;25(5):345-358 [FREE Full text] [doi: 10.1002/car.2448]
73. Li Q. Cyberbullying in schools. Sch Psychol Int 2006;27(2):157-170 [FREE Full text] [doi: 10.1177/0143034306064547]
74. Smith PK, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N. Cyberbullying: its nature and impact in secondary school pupils. J Child Psychol Psychiatry 2008;49(4):376-385 [FREE Full text] [doi: 10.1111/j.1469-7610.2007.01846.x] [Medline: 18363945]
75. Sun S, Fan X, Du J. Cyberbullying perpetration: a meta-analysis of gender differences. Trauma Violence Abuse 2023;11:61-81. [doi: 10.1177/15248380221082087] [Medline: 35603442]

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.12.24; peer-reviewed by S Kath, R Singh; comments to author 29.12.24; revised version received 09.01.25; accepted 10.01.25; published 25.03.25.

Please cite as:
Schulz PJ, Boldi MO, van Ackere A
Adolescent Cyberbullying and Cyber Victimization: Longitudinal Study Before and During COVID-19
J Med Internet Res 2025;27:e70508
URL: https://www.jmir.org/2025/1/e70508
doi:10.2196/70508
PMID:

©Peter Johannes Schulz, Marc-Olivier Boldi, Ann van Ackere. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Transdiagnostic Compulsivity Traits in Problematic Use of the Internet Among UK Residents: Cross-Sectional Network Analysis Study

Chang Liu[1], PhD; Samuel Chamberlain[2,3], PhD; Konstantinos Ioannidis[2,3], PhD; Jeggan Tiego[1], PhD; Jon Grant[4], JD, MD; Murat Yücel[1], PhD; Peter Hellyer[5], PhD; Christine Lochner[6], PhD; Adam Hampshire[7], PhD; Lucy Albertella[1], PhD

[1]School of Psychological Sciences, Monash University, Clayton, Australia

[2]Department of Psychiatry, University of Southampton, Southampton, United Kingdom

[3]Hampshire and Isle of Wight Healthcare NHS Foundation Trust, Southampton, United Kingdom

[4]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, United States

[5]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[6]Department of Psychiatry, Stellenbosch University, Stellenbosch, South Africa

[7]Department of Brain Sciences, Imperial College London, London, United Kingdom

Corresponding Author:
Chang Liu, PhD
School of Psychological Sciences
Monash University
Level 5, 18 Innovation Walk
Wellington Road
Clayton, 3800
Australia
Phone: 61 07 33620222
Email: chang.liu7@monash.edu

Abstract

Background: The societal and public health costs of problematic use of the internet (PUI) are increasingly recognized as a concern across all age groups, presenting a growing challenge for mental health research. International scientific initiatives have emphasized the need to explore the potential roles of personality features in PUI. Compulsivity is a key personality trait associated with PUI and has been recognized by experts as a critical factor that should be prioritized in PUI research. Given that compulsivity is a multidimensional construct and PUI encompasses diverse symptoms, different underlying mechanisms are likely involved. However, the specific relationships between compulsivity dimensions and PUI symptoms remain unclear, limiting our understanding of compulsivity's role in PUI.

Objective: This study aimed to clarify the unique relationships among different dimensions of compulsivity, namely, perfectionism, reward drive, cognitive rigidity, and symptoms of PUI using a symptom-based network approach.

Methods: A regularized partial-correlation network was fitted using a large-scale sample from the United Kingdom. Bridge centrality analysis was conducted to identify bridge nodes within the network. Node predictability analysis was performed to assess the self-determination and controllability of the nodes within the network.

Results: The sample comprised 122,345 individuals from the United Kingdom (51.4% female, age: mean 43.7, SD 16.5, range 9-86 years). The analysis identified several strong mechanistic relationships. The strongest positive intracluster edge was between reward drive and PUI4 (financial consequences due to internet use; weight=0.11). Meanwhile, the strongest negative intracluster edge was between perfectionism and PUI4 (financial consequences due to internet use; weight=0.04). Cognitive rigidity showed strong relationships with PUI2 (internet use for distress relief; weight=0.06) and PUI3 (internet use for loneliness or boredom; weight=0.07). Notably, reward drive (bridge expected influence=0.32) and cognitive rigidity (bridge expected influence=0.16) were identified as key bridge nodes, positively associated with PUI symptoms. Meanwhile, perfectionism exhibited a negative association with PUI symptoms (bridge expected influence=–0.05). The network's overall mean predictability was 0.37, with PUI6 (compulsion, predictability=0.55) showing the highest predictability.

XSL•FO
RenderX

Conclusions: The findings reveal distinct relationships between different dimensions of compulsivity and individual PUI symptoms, supporting the importance of choosing targeted interventions based on individual symptom profiles. In addition, the identified bridge nodes, reward drive, and cognitive rigidity may represent promising targets for PUI prevention and intervention and warrant further investigation.

(J Med Internet Res 2025;27:e66191) doi:10.2196/66191

KEYWORDS

compulsivity; problematic use of the internet; network analysis; perfectionism; reward drive; cognitive rigidity; transdiagnostic; PUI; mental health; intrapersonal factor; cognitive; internet use; network

Introduction

Problematic use of the internet (PUI) is characterized by excessive and maladaptive internet use that leads to significant functional impairment and distress [1]. PUI encompasses a variety of potentially harmful online behaviors such (but not limited to) as excessive gaming, viewing pornography, social networking, and streaming content consumption [2]. It is recognized as a global public health concern, receiving increasing attention from mental health researchers, clinicians and policy makers [1,2]. A recent meta-analysis showed that the rate of PUI is estimated to be 14.2% globally [3], though estimates vary depending on the operational definition of PUI. Mental and physical impairments associated with PUI are well-documented, including depression, anxiety, musculoskeletal problems, and digital eye strain [4-7]. Given the high rates and associated impairments of PUI, it is important to understand the underpinning mechanisms.

Clarifying the role of intrapersonal risk factors (eg, personality traits) in PUI has been identified as a research priority for understanding PUI [1,2]. Compulsivity is defined as the tendency toward repetitive actions that persist despite such actions being inconsistent with one's overall goal and associated negative consequences [8,9]. It is considered a hallmark feature of addictions and obsessive-compulsive disorder [10] and has been consistently linked to PUI [11-14]. For instance, recent evidence suggests that compulsivity may be more closely related to PUI than other well-established risk factors (ie, impulsivity traits [15]).

Compulsivity is oftentimes treated as a unitary construct when estimating its relationships with PUI. However, recent evidence suggests that compulsivity can be decomposed into various distinct but interrelated dimensions. Specifically, a large-scale network-based study [16] has shown that compulsivity may be composed of 3 dimensions: perfectionism (having high personal standards and striving to reach those standards), reward drive (approach tendencies toward immediate gratification and acting on urges) and cognitive rigidity (rigid and repetitive thinking patterns and behaviors). These distinct dimensions of compulsivity likely have different relationships with PUI. For instance, individuals high on reward drive may be more driven to use the internet for its rewarding features (eg, obtain social rewards, online gambling, and cybersex [17-19]). Alternatively, individuals high in cognitive rigidity may find it harder to redirect their attention away from internet-related cues and thoughts and may be prone to activities such as checking the internet due to health-related anxieties (ie, cyberchondria [20]). Finally, individuals high in perfectionism may be inclined to use the internet as a way to cope with the discrepancy between their ideal and actual life achievements [21]. Identifying how different compulsivity dimensions are differently linked with PUI may offer important insights into the mechanistic relationships between compulsivity and PUI.

Further, there is a lack of understanding of the specific symptom pathways between compulsivity dimensions and PUI. Using symptom sum scores as an index for PUI is a common practice when estimating the relationship between compulsivity and PUI. However, recent studies have demonstrated that PUI symptoms differ from each other in their relation to risk factors [15,22,23]. For instance, it has been found that compulsivity is more closely related to PUI symptoms characterized by negative consequences (eg, failure to fulfil role obligations [15]) than other PUI symptoms, such as staying online longer than intended, hiding online time from others or acting aggressively when being disrupted during online time. In addition, it has been observed that individuals with PUI symptoms characterized by continued or escalated usage report experiencing less fear of missing out [24]. Understanding these more specific relations may facilitate a more precise understanding of specific mechanisms underlying different PUI symptom profiles.

A network approach may be particularly useful to address the aforementioned issues and provide a more precise understanding of the relations between different dimensions of compulsivity and symptoms of PUI. The network approach to psychopathology posits that mental conditions can be conceptualized as a complex network, constituted by a set of nodes (eg, item-level characteristics and symptoms) and edges that connect them [25,26]. Accordingly, the associated analytic techniques focus on pinpointing specific and distinct relationships between nodes (eg, dimensions of compulsivity and symptoms of PUI) and visualizing these relationships in insightful ways [27]. Hence, the utilization of the network approach moves beyond examining how compulsivity traits may be associated with the PUI construct (indexed by the symptom sum score) to elucidate differential relationships between compulsivity dimensions and specific symptoms of PUI.

It is worth noting that the network approach also provides novel indices to identify influential nodes that may serve as targets of therapeutic intervention. The bridge expected influence calculates the sum of connectivity of a given node to all nodes of the other cluster (a subset of nodes in the network with denser connections among themselves, forming a subnetwork [28]). Nodes that exhibit high bridge centrality (bridge nodes) are

XSL•FO
RenderX

deemed to be integral to the co-occurrence of constructs, given their extensive interconnectivity with all components of the other constructs [25]. Theoretically, targeting such nodes has the potential to impede the transmission of activation from one construct to another [25,29]. Accordingly, it has been proposed that nodes with high bridge centrality may be ideal candidates for intervention. Given that PUI treatments (both psychological and pharmacological) are still in their infancy, such understanding is both timely and critical to push this field forward.

There is a paucity of research on the relationship between dimensions of compulsivity and PUI symptoms. The current study is the first attempt to fill the knowledge gap using a symptom-based network approach. By investigating the network structure and bridge centrality index, the current study aimed to (1) elucidate the distinct relationships between dimensions of compulsivity (ie, perfectionism, reward drive, and cognitive rigidity) and PUI symptoms, (2) quantify the extent to which each dimension of compulsivity is related to the PUI symptom cluster and identify bridge nodes, and (3) examine the controllability of each node within the network.

Methods

Participants and Procedure

The sample included 122,345 individuals (51.4% female) from the United Kingdom, with a mean age of 43.7 (SD 16.5, range 9-86) years, who participated in the online Great British Intelligence Test (GBIT). The GBIT is a collaborative citizen science project with BBC two Horizon Program, designed to investigate cognition and well-being in the general population. The study was an open, web-based survey using convenience sampling, accessible to all UK residents. It was promoted across BBC media channels, both online and offline, to recruit participants. Interested individuals could access the study through links provided in mainstream media and news articles, which directed them to the study's landing page on the Cognition website. The recruitment for the GBIT study commenced in December 2019, with the specific sample used in this paper recruited between May and June 2020. For the current study, we focused solely on individuals who provided complete responses for both measures. The compulsivity measure was introduced during a later phase of the GBIT. Among the 122,680 participants who took part after the compulsivity measure was included, only a small proportion of responses were missing, specifically, 335 for the compulsivity measure and 45 for PUI items. These incomplete responses were excluded from the data analysis.

Measures

Trait Compulsivity

The Cambridge-Chicago Compulsivity Trait Scale (CHI-T; [30]) was used to measure trait compulsivity. The CHI-T contains 15 items, which cover 3 dimensions of compulsivity (perfectionism, cognitive rigidity and reward-drive; [16]). The perfectionism subscale covers self-oriented perfectionism, which features having high personal standards and striving for perfection and "just right" (eg, "completion to a high standard" and "doing things just right"). The cognitive rigidity subscale covers the inability to change behaviors or mindset (eg, "repetitive thoughts" and "being stubborn or rigid"). The reward drive subscale covers approach tendencies toward immediate gratification and acting on urges (eg, "acting on urges" and "seeking immediate reward"). We have opted for the subscale score over the item score due to practical considerations. Currently, interventions for individual items of the CHI-T scale are undeveloped. In contrast, targeted therapies for specific subscales, like attentional bias modification therapy, already exist and may effectively reduce reward-drive tendencies. Therefore, using the subscale score may enhance the applicability of our findings in future intervention development.

PUI Symptoms

PUI symptoms were measured by 7 items based on expert consensus (Hampshire et al [31]). The items measure internet checking, internet use for distress relief, internet use for loneliness or boredom, financial consequences due to internet use, prioritizing internet use over role obligations, compulsion and failed attempts to cut down, in the past 4 weeks. Response options range from 0 (never) to 6 (more than hourly every day).

Data Analysis

Descriptive statistics were calculated for demographic variables, CHI-T subscales, and PUI items. Before conducting our formal analysis, we used the goldbricker function to check for significant topological overlap among nodes, which identifies pairs with highly similar connectivity patterns, indicative of duplicates [32]. Specifically, we looked for node pairs with a topological overlap exceeding 80%, which corresponds to less than 20% difference in their connectivity patterns (significant at $P<.01$). No topological overlap was identified (nor was it for the detailed item-level analysis in the supplement). We constructed a regularized partial correlation network, where the variables (ie, the 3 dimensions of compulsivity and individual PUI symptoms) are represented as nodes. The edges between nodes represent regularized partial correlations, controlling for all other variables within the network. The Graphical Least Absolute Shrinkage and Selection Operator algorithm was applied for regularization [33]. The regularization procedure forces weak or spurious partial correlations to become exactly zero, leading to a sparse network where only the most robust edges are retained [33]. Following recommendations from [33], we set the tuning parameter (γ) for regularization to 0.5, which balances the trade-off between model fit and sparsity of the network.

The Fruchterman-Reingold layout algorithm was used for network visualization [34]. Within each network, strongly related nodes were placed closer together, while weakly related nodes were placed further apart. The magnitudes of partial correlations (edge weights) are reflected as edge thickness, with thicker edges indicating stronger partial correlations between nodes. Blue edges represent positive partial correlations, and red edges represent negative partial correlations between nodes. The visualization procedures were conducted via the "qgraph" package in R (R Core Team) [34]. To enhance visual clarity, we adjusted the scaling of edge widths and color saturation using the cut argument. The improved visual representation of

XSL•FO
RenderX

the network is available (Figure S5 in Multimedia Appendix 1).

We used the "networktools" package [32] to quantify to what extent each compulsivity dimension was related to the PUI symptom cluster and identify influential bridge nodes. Number of 2 clusters (subnetworks) were predefined for the network, namely, the compulsivity cluster (consisting of 3 compulsivity dimensions) and the PUI community (consisting of PUI items). Bridge expected influence (ie, the sum of the edge weights connecting a given node to all nodes in the other cluster; [25] was computed for each node. The compulsivity dimension with a high positive bridge expected influence is positively related to the PUI symptom cluster. Conversely, the compulsivity dimension with a negative bridge expected influence is negatively related to the PUI symptom cluster.

We also examined the predictability of nodes, assessing how well a specific node can be predicted by all other nodes within the network [35,36]. Since all the variables included are continuous, we selected R^2 as the measure of predictability for the estimated network. A node with higher predictability suggests that a larger proportion of its variance can be explained by other nodes in the network [35]. Further, we also computed the mean predictability across nodes within the network. This allows us to determine whether, on average, the variance observed in the nodes of the network can be explained by other nodes within the network [35,36]. A high mean predictability indicates that a network is largely determined by internal nodes within itself [35,36].

To estimate the accuracy and stability of the network model, we followed the steps described in [37]. First, we bootstrapped 95% CIs (with 10,000 bootstrap samples) of edge weights to determine the accuracy of the estimated edges within each network. Second, we estimated the stability of edges and bridge centrality using the case-dropping bootstrap procedure (with 10,000 bootstrap samples) and calculated the correlation stability coefficient (CS-coefficient) for edge weight and bridge expected influence. An acceptable CS-coefficient should be above 0.25, and a value of 0.5 or higher indicates optimal stability. Finally, in alignment with the tutorial paper [37], we examined whether the weights for each edge and the bridge centrality estimates for each node differ significantly (α=.05) from one another. This was achieved using bootstrapped difference tests for edge weights and bridge centrality indices. The "bootnet" package was used to perform the accuracy and stability tests [38].

We acknowledge that the item-level findings may add further nuance to the current results. To offer a comprehensive overview, we have included the item-level network analysis in the for readers interested in detailed item-level analysis (Figure S6 in Multimedia Appendix 1).

Ethical Considerations

The study adhered to the ethical principles for medical research involving human subjects outlined in the Helsinki Declaration of 1975, as revised in 2008, and received approval from the institutional review board at Imperial College London (approval number 17IC4009). All participants were fully informed about the study details and provided their informed consent before participation. All data collected were anonymous. Participation was voluntary, and no financial incentives were offered.

Results

Table 1 presents demographic characteristics of the study sample. Descriptive statistics for CHI-T subscale scores and PUI symptom scores are provided (Table S1 in Multimedia Appendix 1). The clinical characteristics of the sample are presented (Table S2 in Multimedia Appendix 1).

Figure 1 presents the estimated compulsivity-PUI network and the bridge centrality plot (ie, raw bridge expected values of each node within the network). The network was densely connected, with 39 remaining edges (out of 45 possible edges). The strongest positive intracluster edge emerged between reward-drive and PUI4 (financial consequences due to internet use; edge weight=0.11). The strongest negative intracluster edge emerged between perfectionism and PUI4 (financial consequences due to internet use; edge weight=–0.04). Cognitive rigidity was strongly associated with PUI2 (internet use for distress relief; edge weight=0.06) and PUI3 (internet use for loneliness or boredom; edge weight=0.07). We found that reward-drive had the highest positive bridge expected influence, while perfectionism had the highest negative bridge expected influence.

The predictability of nodes was plotted as the colored area in the ring around each node. The nodes exhibiting the highest predictability were PUI6 (compulsion, predictability=0.55); PUI7 (failed attempts to cut down, predictability=0.49), and PUI3 (internet use for loneliness or boredom, predictability=0.44). The network's overall mean predictability is 0.37, which is comparable to the previously reported average level of predictability (mean predictability=0.32) in the literature [35]. The raw predictability scores can be found (Table S3 in Multimedia Appendix 1).

The narrow confidence intervals for edge weights (Figure S1) indicate that estimated edges within networks were accurate. The CS-coefficients for bridge expected influence were excellent in the estimated network (value=0.75), indicating optimal stability for the bridge centrality index. Results from the bootstrapped stability and accuracy tests are included (Figure S1 in Multimedia Appendix 1 and Figure S3 in Multimedia Appendix 1). The results of bootstrapped difference tests are included (Figure S2 in Multimedia Appendix 1 and Figure S4 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Descriptive statistics of demographic variables for UK residents who participated in the online Great British Intelligence Test and completed the problematic use of the internet and compulsivity measures between May and June 2020.

Variable	Values
Age (years), mean (SD)	43.7 (16.5)
Sex, n (%)	
Female	62286 (51.4)
Male	58728 (48)
Nonbinary/other	710 (0.6)
Education level, n (%)	
No schooling	290 (0.2)
Primary school	2592 (2.1)
High school or diploma	41469 (33.9)
Bachelor's or master's degree	68644 (54.1)
Doctoral degree	4621 (3.8)
Other	4729 (3.9)
Ethnicity, n (%)	
Mixed ethnicity	3367 (2.8)
North African	260 (0.2)
Rom, Sinti or Bedouin	100 (0.1)
Sub-Saharan African or Afro-American	551 (0.5)
West-central Asian	499 (0.4)
White	108497 (88.7)
Nondisclosure	9071 (7.4)

XSL•FO
RenderX

Figure 1. Compulsivity-PUI network and bridge centrality plot of UK residents who participated in the online GBIT and completed compulsivity and PUI measures between May and June 2020.

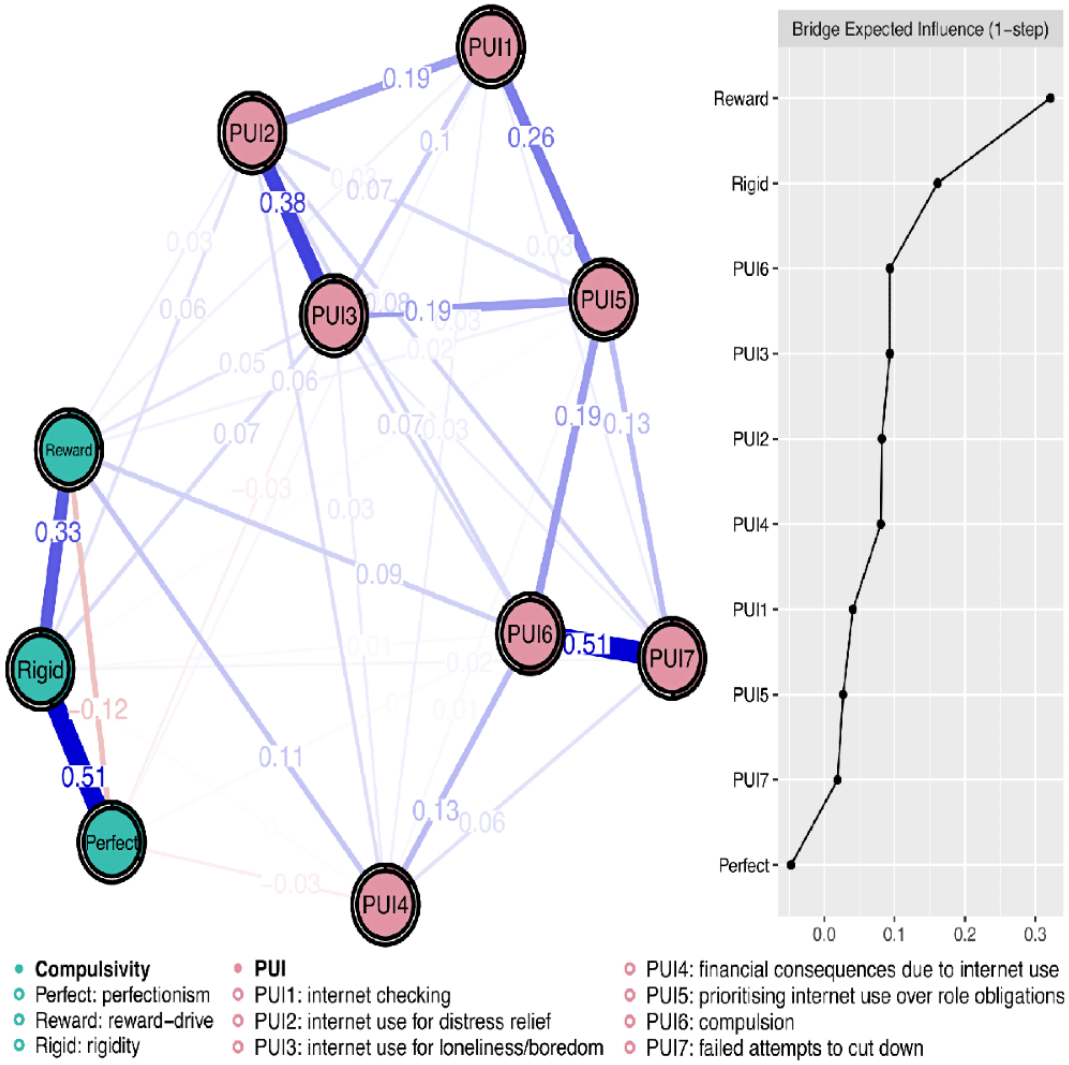

Discussion

Principal Findings

This study is the first and largest to date to explore the relationships between different dimensions of compulsivity and individual PUI symptoms through network analysis. Extensive intercluster connections were found, indicating that compulsivity and PUI were strongly interrelated. We identified several strong mechanistic pathways, notably the edge between cognitive rigidity and coping and the edge between reward drive and functional impairment (ie, negative financial consequences). Finally, reward drive and cognitive rigidity were identified as the bridge nodes.

When elucidating interrelationships between compulsivity dimensions and PUI symptoms, we found that cognitive rigidity was associated with coping-motivated internet use. The results are in line with previous research [39-41], which highlights the role of rigid coping tendencies as a core mechanism underlying the relationship between compulsivity and addictive behaviors. Specifically, individuals high on cognitive rigidity may have a narrowed coping repertoire when experiencing negative emotions that, in turn, limits engagement in alternative coping strategies [42]. Several features of internet use (eg, easy accessibility and providing immediate stress relief) make it easy for individuals to use it as a coping strategy. However, such a coping strategy may be counterproductive in the long run (eg, due to excessive time spent online and avoiding dealing with daily challenges). For individuals with high cognitive rigidity, the inability to generate and adopt alternative coping responses when they need these most (ie, under stress) may result in their persistent engagement in coping-motivated internet use, even if this coping strategy proves to be ineffective and impairing.

Another strong pathway emerged between reward drive and functional impairment (ie, negative financial consequences and sleep loss), which may be explained by the counterproductive attentional capture mechanism characteristic of compulsivity [11,42]. Specifically, individuals with high reward drive may exhibit stronger attraction toward reward-related cues, finding it difficult to resist approaching these cues even when doing so may be counterproductive. In these individuals, such approach tendencies could be triggered by reward-related stimuli presented on the internet (eg, applications and websites for recreational online gaming, gambling, and cybersex), leading to internet use despite such use not being in line with an individual's goals and resulting in negative consequences.

XSL•FO
RenderX

Interestingly, negative relationships were observed between perfectionism and PUI symptoms. The relationship between perfectionism and addictive behaviors tends to be inconsistent across studies, with some studies finding that perfectionism may actually protect individuals from addiction-related impairments [43] and problematic drinking [44]. This inconsistency may stem from variations in how perfectionism is measured. Perfectionism is considered as a multidimensional construct, encompassing types like self-oriented, other-oriented, and socially prescribed perfectionism [45]. In this study, the perfectionism dimension captures self-oriented perfectionism (ie, holding high standards and striving for perfection or "just right"). It has been posited that specific aspects of perfectionism (ie, self-oriented perfectionism) may be considered a form of hyperconscientiousness [44]. Such hyper-conscientiousness may protect individuals from particular addictive behaviors, which are considered impairing and not aligned with their goals (eg, being perfect in everything; [44]). Our results provide support for this hypothesis by showing that perfectionism was negatively associated with self-report items measuring functional impairment.

The bridge centrality analysis identified reward drive and cognitive rigidity as bridge nodes that positively relate to the PUI symptom cluster, whereas perfectionism negatively relates to it. The roles of reward drive and cognitive rigidity in PUI are in line with previous studies that have demonstrated attentional approach tendencies toward reward cues and rigidity of such attentional responses are linked to obsessive-compulsive and addictive behaviors, including PUI [11,41]. Meanwhile, by showing that perfectionism is negatively related to the PUI symptom cluster, our findings extend previous assumptions on the potential protective role of perfectionism (underpinned by conscientiousness) in problematic alcohol use [44] to suggest that such a protective effect may be transdiagnostic across substance use and behavioral addictions.

Current findings addressed several key research priorities identified by experts [1,2]. First, our findings help elucidate the relations between compulsivity dimensions and PUI symptoms, which addressed the need for clarifying the role of compulsivity in PUI [1,2]. The distinct connectivity between traits and symptoms may suggest that interventions should be personalized to address individual symptom profiles. For example, interventions might focus on reducing reward-driven tendencies rather than targeting perfectionism in individuals whose problematic usage is characterized by financial consequences, to achieve better intervention outcomes. Second, by demonstrating the role of reward drive and cognitive rigidity, our results may inform efforts to identify at-risk individuals at the earliest stages of PUI and provide early intervention [2]. Specifically, given that compulsivity traits can exist before the emergence of PUI symptoms, reward drive and cognitive rigidity traits should, in future work, be evaluated in terms of their potential value in identifying individuals who are most at risk of developing PUI. In turn, this may inform our screening processes.

Previous simulation research on the role of bridge nodes has demonstrated that "focusing treatment on the detected nodes is effective in preventing the contagion of comorbid disorders" ([25], P. 365). Empirically, if we can establish that the identified bridge nodes, namely reward drive and cognitive rigidity, exhibit temporal precedence, these findings could mark a significant step toward intervention target selection.

Finally, node predictability may shed several practical insights. For instance, using the internet to cope with loneliness or boredom was identified as one of the nodes with the highest predictability. The predictability allows an understanding of the clinical relevance of the identified mechanism. For instance, we found that a significant proportion of variance in using the internet to cope with loneliness or boredom may be explained by its neighboring nodes (reward-drive and cognitive rigidity). This may indicate that intervention on reward drive and cognitive rigidity may have a higher success potential in reducing coping-motivated internet use. Further, the mean predictability of the network suggests that although compulsivity dimensions may play significant roles in PUI, other factors may also be involved in driving PUI and should be considered when planning interventions.

Based on our findings, targeting reward drive and cognitive rigidity (the bridge nodes) via digital personality change interventions [46] may be expected to improve treatment outcomes. Specifically, digital personality change interventions incorporate evidence-based techniques that foster and sustain personality change (eg, tailored implementation intentions (if-then plans), psychoeducation, behavioral activation, self-reflection, resource activation, and individual progress feedback), alongside a chatbot that facilitates coaching dialogues to simulate a personal coaching process [47]. A large-scale randomized controlled trial (n=1523) has shown that a digital personality change intervention was associated with significant changes in big 5 personality traits, versus waiting list control, with effects persisting beyond the intervention [47]. The promising findings suggest that such interventions have promise in potentially improving compulsivity traits such as reward drive and cognitive rigidity. It should be noted that the aforementioned study did not directly measure compulsivity, but rather traits associated with it (eg, conscientiousness), as classic personality inventories typically do not assess compulsivity. Future studies may consider directly examining the extent to which the intervention can effectively reduce dimensions of compulsivity.

Considering the transdiagnostic characteristics of mechanisms discussed above (ie, counterproductive attentional capture, rigid coping styles, and conscientiousness), such interventions may potentially help reduce other compulsive behaviors (eg, problematic drinking, binge-eating, and obsessive-compulsive behaviors) that frequently co-occur with PUI [48]. These suggestions should be tested in future research studies.

Limitations

Several limitations should be noted when interpreting current findings. First, PUI symptoms in the current study were measured using items drawn from expert consensus, which may raise concerns regarding the psychometric properties of these items. This was the case for pragmatic reasons as the data were collected as part of a citizen science study [31] that needed to have relatively brief questions that were also acceptable to other

XSL•FO
RenderX

stakeholders. The advantage of this approach is that it enables a very large sample size to be collected. Second, the findings were analyzed and discussed with the theoretical assumption that personality traits act as predisposing factors and predict PUI (instead of the other way around; [49-52]. However, due to the cross-sectional nature of the study, we cannot disregard the possibility that PUI may contribute to compulsive traits. It will be important for future research to test the temporal precedence of compulsivity traits and PUI symptoms with longitudinal network analysis. Third, the majority of the sample self-identified as White or Caucasian and resided in a Western country (ie, the United Kingdom), potentially limiting the generalizability of these findings to individuals from other ethnic or cultural backgrounds. Fourth, the current findings were based on data from the general population, which may not be generalizable to individuals with severe levels of PUI. It is important for future research to replicate current findings among individuals with severe levels of PUI.

Future Directions

In theory, targeting bridge nodes could limit the transmission of activation between different psychological symptoms (or conditions), potentially preventing the onset of comorbid psychopathologies and thus making them valuable targets for intervention [25]. However, empirical evidence on the predictive validity of bridge nodes is inconsistent. For instance, findings from 6 randomized control trials suggest that bridge nodes may predict treatment outcomes in patients with depressive disorders [53]. In addition, studies indicate that bridge nodes can statistically predict the persistence of psychopathology, such as poorer psychosocial functioning and lower BMI in adolescents with anorexia nervosa [54], as well as the future onset of anxiety disorders [55]. Conversely, some evidence suggests that bridge nodes might only partially predict treatment outcomes (ie, predict weight restoration, but not full remission, in patients with anorexia nervosa [56]) or yield nonsignificant results in terms of relationships to outcomes [57]. Given these mixed results, future longitudinal and intervention research is essential to ascertain the predictive validity of bridge nodes before they can be confidently used as treatment targets.

Most research in the fields of psychology and psychiatry that uses network-based approaches has relied on cross-sectional, between-subject designs. However, there is an increasing shift toward using longitudinal designs, including panel data and intensive time-series data [58-60]. These methods enable researchers to differentiate within-subject and between-subject effects and to elucidate the temporal dynamics among psychological constructs. A recent study using graphical vector autoregressive network models to analyze panel data revealed that coping-motivated substance use could predict posttraumatic stress disorder symptoms over time [59]. In addition, when combined with intensive time-series data, graphical vector autoregressive network models can elucidate the interrelationships among psychological symptoms and the impact of risk factors on these symptoms at an individual level, offering insights into precision medicine (eg, by enabling personalized diagnostic profiling and treatment planning; [58,61]). To improve the clinical relevance of current findings, future studies should use longitudinal designs to further explore the network structure and bridge nodes identified in this study.

Conclusions

The current study addresses evidence gaps in the literature by investigating the relations between different dimensions of self-reported compulsivity and PUI symptoms. using network analysis, we identified distinct connections between dimensions of compulsivity and PUI symptoms. Specifically, we found that reward-drive and cognitive rigidity were differently associated with PUI symptom clusters, which may reflect distinct mechanisms (ie, counterproductive attentional capture and rigid coping response). Further, we identified “reward drive” and “cognitive rigidity” as bridge nodes within the network. These findings represent an initial exploration into the role of transdiagnostic compulsivity in developing future interventions for PUI. Examining the direction of the mechanistic pathways identified, whether compulsivity drives PUI or vice versa, through longitudinal studies represents a promising next step for applying network-informed treatment strategies in clinical practice for PUI.

Data Availability

Data sharing requests from researchers within academic institutions should be directed to Adam Hampshire.

Conflicts of Interest

LA receives a stipend for her work as Associate Editor at Neuroscience and Biobehavioral Reviews, and at Comprehensive Psychiatry. SRC receives a stipend for his work as Associate Editor at Neuroscience and Biobehavioral Reviews and at Comprehensive Psychiatry. JT is supported by a National Health and Medical Research Council Investigator Grant (NHMRC: 1197431). AH is founder and director of Future Cognition Ltd and cofounder and co-director of H2 Cognitive Designs, which develop custom cognitive assessment technology and provide online assessment technology as a service to third parties respectively, particularly in the research and healthcare sectors. The funders had no involvement in the conduct of the study, its interpretation, or the decision to submit it for publication. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Multimedia Appendix 1
The results of the Cambridge-Chicago Compulsivity Trait Scale–problematic use of the internet network.
[DOCX File , 890 KB - jmir_v27i1e66191_app1.docx]

XSL•FO
RenderX

References

1. Fineberg N, Menchón JM, Hall N, Dell'Osso B, Brand M, Potenza M, et al. Advances in problematic usage of the internet research - A narrative review by experts from the European network for problematic usage of the internet. Compr Psychiatry 2022 Oct;118:152346. [doi: 10.1016/j.comppsych.2022.152346] [Medline: 36029549]
2. Fineberg N, Demetrovics Z, Stein D, Ioannidis K, Potenza M, Grünblatt E, COST Action Network, et al. Manifesto for a European research network into Problematic Usage of the Internet. Eur Neuropsychopharmacol 2018 Nov;28(11):1232-1246 [FREE Full text] [doi: 10.1016/j.euroneuro.2018.08.004] [Medline: 30509450]
3. Meng SQ, Cheng JL, Li YY, Yang XQ, Zheng JW, Chang XW, et al. Global prevalence of digital addiction in general population: A systematic review and meta-analysis. Clin Psychol Rev 2022 Mar;92:102128. [doi: 10.1016/j.cpr.2022.102128] [Medline: 35150965]
4. Cerniglia L, Zoratto F, Cimino S, Laviola G, Ammaniti M, Adriani W. Internet Addiction in adolescence: Neurobiological, psychosocial and clinical issues. Neurosci Biobehav Rev 2017 May;76(Pt A):174-184. [doi: 10.1016/j.neubiorev.2016.12.024] [Medline: 28027952]
5. Mylona I, Deres ES, Dere GS, Tsinopoulos I, Glynatsis M. The impact of internet and videogaming addiction on adolescent vision: a review of the literature. Front Public Health 2020;8:63 [FREE Full text] [doi: 10.3389/fpubh.2020.00063] [Medline: 32195219]
6. Yang G, Cao J, Li Y, Cheng P, Liu B, Hao Z, et al. Association Between Internet Addiction and the Risk of Musculoskeletal Pain in Chinese College Freshmen - A Cross-Sectional Study. Front Psychol 2019;10:1959 [FREE Full text] [doi: 10.3389/fpsyg.2019.01959] [Medline: 31551859]
7. Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: A systematic review and meta-analysis. Psychiatry Res 2020 Feb;284:112686. [doi: 10.1016/j.psychres.2019.112686] [Medline: 31757638]
8. Luigjes J, Lorenzetti V, de Haan S, Youssef GJ, Murawski C, Sjoerds Z, et al. Defining Compulsive Behavior. Neuropsychol Rev 2019;29(1):4-13 [FREE Full text] [doi: 10.1007/s11065-019-09404-9] [Medline: 31016439]
9. Robbins TW, Gillan CM, Smith DG, de Wit S, Ersche KD. Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. Trends Cogn Sci 2012;16(1):81-91. [doi: 10.1016/j.tics.2011.11.009] [Medline: 22155014]
10. Figee M, Pattij T, Willuhn I, Luigjes J, van den Brink W, Goudriaan A, et al. Compulsivity in obsessive-compulsive disorder and addictions. Eur Neuropsychopharmacol 2016;26(5):856-868. [doi: 10.1016/j.euroneuro.2015.12.003] [Medline: 26774279]
11. Albertella L, Chamberlain SR, Le Pelley ME, Greenwood L, Lee RS, Den Ouden L, et al. Compulsivity is measurable across distinct psychiatric symptom domains and is associated with familial risk and reward-related attentional capture. CNS Spectr 2020;25(4):519-526 [FREE Full text] [doi: 10.1017/S1092852919001330] [Medline: 31645228]
12. Lochner C, Albertella L, Kidd M, Kilic Z, Ioannidis K, Grant JE, et al. The COVID-19 pandemic and problematic usage of the internet: findings from a diverse adult sample in South Africa. J Psychiatr Res 2022;153:229-235 [FREE Full text] [doi: 10.1016/j.jpsychires.2022.06.035] [Medline: 35841819]
13. Ioannidis K, Chamberlain SR, Treder MS, Kiraly F, Leppink EW, Redden SA, et al. Problematic internet use (PIU): associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. J Psychiatr Res 2016;83:94-102 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.08.010] [Medline: 27580487]
14. Raj K, Segrave R, Verdéjo-Garcia A, Yücel M. Cognitive inflexibility and repetitive habitual actions are associated with problematic use of the internet. Addict Behav 2023;139:107600. [doi: 10.1016/j.addbeh.2022.107600] [Medline: 36608592]
15. Liu C, Rotaru K, Ren L, Chamberlain SR, Christensen E, Brierley M, et al. Examining the unique relationships between problematic use of the internet and impulsive and compulsive tendencies: network approach. BJPsych Open 2024;10(3):e104 [FREE Full text] [doi: 10.1192/bjo.2024.59] [Medline: 38721785]
16. Liu C, Albertella L, Lochner C, Tiego J, Grant JE, Ioannidis K, et al. Conceptualising compulsivity through network analysis: a two-sample study. Compr Psychiatry 2023;127:152429 [FREE Full text] [doi: 10.1016/j.comppsych.2023.152429] [Medline: 37832377]
17. Rosenthal-von der Pütten AM, Hastall MR, Köcher S, Meske C, Heinrich T, Labrenz F, et al. “Likes” as social rewards: their role in online social comparison and decisions to like other People's selfies. Computers in Human Behavior 2019;92:76-86. [doi: 10.1016/j.chb.2018.10.017]
18. Konietzny J, Caruana A, Cassar ML. Fun and fair, and I don’t care: the role of enjoyment, fairness and subjective norms on online gambling intentions. Journal of Retailing and Consumer Services 2018;44:91-99. [doi: 10.1016/j.jretconser.2018.06.010]
19. Laier C, Pekal J, Brand M. Cybersex addiction in heterosexual female users of internet pornography can be explained by gratification hypothesis. Cyberpsychol Behav Soc Netw 2014;17(8):505-511. [doi: 10.1089/cyber.2013.0396] [Medline: 25080011]
20. Vismara M, Caricasole V, Starcevic V, Cinosi E, Dell'Osso B, Martinotti G, et al. Is cyberchondria a new transdiagnostic digital compulsive syndrome? A systematic review of the evidence. Compr Psychiatry 2020;99:152167 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152167] [Medline: 32146315]

XSL•FO
RenderX

21. Maurage P, Creupelandt C, Bollen Z, Pabst A, Fontesse S, Laniepce A, et al. Greater self-oriented and socially prescribed perfectionism in severe alcohol use disorder. Alcohol Clin Exp Res 2022;46(7):1340-1347. [doi: 10.1111/acer.14878] [Medline: 35913501]
22. Huang S, Lai X, Li Y, Luo Y, Wang Y. Understanding juveniles' problematic smartphone use and related influencing factors: a network perspective. J Behav Addict 2021;10(3):811-826 [FREE Full text] [doi: 10.1556/2006.2021.00048] [Medline: 34406975]
23. Liu C, Ren L, Rotaru K, Liu X, Li K, Yang W, et al. Bridging the links between big five personality traits and problematic smartphone use: a network analysis. J Behav Addict 2023;12(1):128-136 [FREE Full text] [doi: 10.1556/2006.2022.00093] [Medline: 36763335]
24. Li L, Niu Z, Griffiths MD, Mei S. Relationship between gaming disorder, self-compensation motivation, game flow, time spent gaming, and fear of missing out among a sample of Chinese university students: a network analysis. Front Psychiatry 2021;12:761519 [FREE Full text] [doi: 10.3389/fpsyt.2021.761519] [Medline: 34790137]
25. Jones PJ, Ma R, McNally RJ. Bridge centrality: a network approach to understanding comorbidity. Multivariate Behav Res 2021;56(2):353-367. [doi: 10.1080/00273171.2019.1614898] [Medline: 31179765]
26. Jones PJ, Robinaugh DR. An Answer to "So What?" implications of network theory for research and practice. Focus (Am Psychiatr Publ) 2021;19(2):204-210 [FREE Full text] [doi: 10.1176/appi.focus.20200050] [Medline: 34690584]
27. Bringmann LF, Eronen MI. Don't blame the model: reconsidering the network approach to psychopathology. Psychol Rev 2018;125(4):606-615. [doi: 10.1037/rev0000108] [Medline: 29952625]
28. Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. Defining and identifying communities in networks. Proc Natl Acad Sci U S A 2004;101(9):2658-2663 [FREE Full text] [doi: 10.1073/pnas.0400054101] [Medline: 14981240]
29. Robinaugh DJ, Millner AJ, McNally RJ. Identifying highly influential nodes in the complicated grief network. J Abnorm Psychol 2016;125(6):747-757 [FREE Full text] [doi: 10.1037/abn0000181] [Medline: 27505622]
30. Chamberlain SR, Grant JE. Initial validation of a transdiagnostic compulsivity questionnaire: the cambridge-Chicago compulsivity trait scale. CNS Spectr 2018;23(5):340-346 [FREE Full text] [doi: 10.1017/S1092852918000810] [Medline: 29730994]
31. Hampshire A, Hellyer PJ, Soreq E, Mehta MA, Ioannidis K, Trender W, et al. Associations between dimensions of behaviour, personality traits, and mental-health during the COVID-19 pandemic in the United Kingdom. Nat Commun 2021;12(1):4111 [FREE Full text] [doi: 10.1038/s41467-021-24365-5] [Medline: 34272365]
32. Jones P. Networktools: Tools for identifying important nodes in networks. R package version. 2018. URL: https://cran.r-project.org/web/packages/networktools/networktools.pdf [accessed 2025-02-18]
33. Epskamp S, Fried EI. A tutorial on regularized partial correlation networks. Psychol Methods 2018;23(4):617-634. [doi: 10.1037/met0000167] [Medline: 29595293]
34. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: network visualizations of relationships in psychometric data. J. Stat. Soft 2012;48(4):1-18. [doi: 10.18637/jss.v048.i04]
35. Haslbeck JMB, Fried EI. How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. Psychol Med 2017;47(16):2767-2776. [doi: 10.1017/S0033291717001258] [Medline: 28625186]
36. Haslbeck JMB, Waldorp LJ. How well do network models predict observations? On the importance of predictability in network models. Behav Res Methods 2018;50(2):853-861 [FREE Full text] [doi: 10.3758/s13428-017-0910-x] [Medline: 28718088]
37. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods 2018;50(1):195-212 [FREE Full text] [doi: 10.3758/s13428-017-0862-1] [Medline: 28342071]
38. Epskamp S, Fried E. Package 'bootnet'. R Package Version 1.3. 2020. URL: https://cran.r-project.org/web/packages/bootnet/bootnet.pdf [accessed 2025-02-18]
39. Liu C, Rotaru K, Lee RSC, Tiego J, Suo C, Yücel M, et al. Distress-driven impulsivity interacts with cognitive inflexibility to determine addiction-like eating. J Behav Addict 2021;10(3):534-539 [FREE Full text] [doi: 10.1556/2006.2021.00027] [Medline: 33909594]
40. Liu C, Rotaru K, Chamberlain SR, Yücel M, Grant JE, Lee RSC, et al. Distress-driven impulsivity interacts with trait compulsivity in association with problematic drinking: a two-sample study. Front Psychiatry 2022;13:938275 [FREE Full text] [doi: 10.3389/fpsyt.2022.938275] [Medline: 36203833]
41. Liu C, Rotaru K, Chamberlain SR, Ren L, Fontenelle LF, Lee RSC, et al. The moderating role of psychological flexibility on the association between distress-driven impulsivity and problematic internet use. Int J Environ Res Public Health 2022;19(15):9592 [FREE Full text] [doi: 10.3390/ijerph19159592] [Medline: 35954947]
42. Albertella L, Chamberlain S, Fontenelle L, Westbrook R. Linking addictive and obsessive-compulsive behaviours. In: Handbook of Substance Misuse and Addictions: From Biology to Public Health. Cham: Springer; 2022:1-18.
43. Simons JS, Christopher MS, McLaury AE. Personal strivings, binge drinking, and alcohol-related problems. Addict Behav 2004;29(4):773-779. [doi: 10.1016/j.addbeh.2004.02.006] [Medline: 15135560]
44. Flett GL, Goldstein A, Wall A, Hewitt PL, Wekerle C, Azzi N. Perfectionism and binge drinking in Canadian students making the transition to university. J Am Coll Health 2008;57(2):249-253. [doi: 10.3200/JACH.57.2.249-256] [Medline: 18809543]

XSL•FO
RenderX

45. Hewitt PL, Flett GL. Perfectionism in the self and social contexts: conceptualization, assessment, and association with psychopathology. J Pers Soc Psychol 1991;60(3):456-470. [doi: 10.1037//0022-3514.60.3.456] [Medline: 2027080]
46. Allemand M, Flückiger C. Personality change through digital-coaching interventions. Curr Dir Psychol Sci 2022;31(1):41-48. [doi: 10.1177/09637214211067782]
47. Stieger M, Flückiger C, Rüegger D, Kowatsch T, Roberts BW, Allemand M. Changing personality traits with the help of a digital personality change intervention. Proc Natl Acad Sci U S A 2021;118(8):e2017548118 [FREE Full text] [doi: 10.1073/pnas.2017548118] [Medline: 33558417]
48. Tiego J, Oostermeijer S, Prochazkova L, Parkes L, Dawson A, Youssef G, et al. Overlapping dimensional phenotypes of impulsivity and compulsivity explain co-occurrence of addictive and related behaviors. CNS Spectr 2019;24(4):426-440. [doi: 10.1017/S1092852918001244] [Medline: 30458896]
49. Billieux J. Problematic use of the mobile phone: a literature review and a pathways model. CPSR 2012;8(4):299-307. [doi: 10.2174/157340012803520522]
50. Billieux J, Maurage P, Lopez-Fernandez O, Kuss DJ, Griffiths MD. Can disordered mobile phone use be considered a behavioral addiction? An update on current evidence and a comprehensive model for future research. Curr Addict Rep 2015;2(2):156-162. [doi: 10.1007/s40429-015-0054-y]
51. Brand M, Young KS, Laier C, Wölfling K, Potenza MN. Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: an interaction of person-affect-cognition-execution (I-PACE) model. Neurosci Biobehav Rev 2016;71:252-266 [FREE Full text] [doi: 10.1016/j.neubiorev.2016.08.033] [Medline: 27590829]
52. Brand M, Wegmann E, Stark R, Müller A, Wölfling K, Robbins TW, et al. The interaction of person-affect-cognition-execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. Neurosci Biobehav Rev 2019;104:1-10 [FREE Full text] [doi: 10.1016/j.neubiorev.2019.06.032] [Medline: 31247240]
53. O'Driscoll C, Buckman JEJ, Fried EI, Saunders R, Cohen ZD, Ambler G, et al. The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: analysis of data from six randomised control trials. BMC Med 2021;19(1):109 [FREE Full text] [doi: 10.1186/s12916-021-01971-0] [Medline: 33952286]
54. Monteleone AM, Cascino G, Salerno L, Schmidt U, Micali N, Cardi V, et al. A network analysis in adolescent anorexia nervosa exploring the connection between both patient and carer reactions and outcome. Eur Eat Disord Rev 2023;31(1):65-75 [FREE Full text] [doi: 10.1002/erv.2933] [Medline: 35717595]
55. Rouquette A, Pingault JB, Fried EI, Orri M, Falissard B, Kossakowski JJ, et al. Emotional and behavioral symptom network structure in elementary school girls and association with anxiety disorders and depression in adolescence and early adulthood: a network analysis. JAMA Psychiatry 2018;75(11):1173-1181 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2119] [Medline: 30128480]
56. Hagan KE, Matheson BE, Datta N, L'Insalata AM, Onipede ZA, Gorrell S, et al. Understanding outcomes in family-based treatment for adolescent anorexia nervosa: a network approach. Psychol Med 2023;53(2):396-407 [FREE Full text] [doi: 10.1017/S0033291721001604] [Medline: 33952357]
57. Brown TA, Vanzhula IA, Reilly EE, Levinson CA, Berner LA, Krueger A, et al. Body mistrust bridges interoceptive awareness and eating disorder symptoms. J Abnorm Psychol 2020;129(5):445-456 [FREE Full text] [doi: 10.1037/abn0000516] [Medline: 32202809]
58. Ebrahimi OV, Borsboom D, Hoekstra RHA, Epskamp S, Ostinelli EG, Bastiaansen JA, et al. Towards precision in the diagnostic profiling of patients: leveraging symptom dynamics as a clinical characterisation dimension in the assessment of major depressive disorder. Br J Psychiatry 2024;224(5):157-163 [FREE Full text] [doi: 10.1192/bjp.2024.19] [Medline: 38584324]
59. Freichel R, Skjerdingstad N, Mansueto AC, Epskamp S, Hoffart A, Johnson SU, et al. Use of substances to cope predicts posttraumatic stress disorder symptom persistence: investigating patterns of interactions between symptoms and its maintaining mechanisms. Psychol Trauma 2025;17(1):216-224. [doi: 10.1037/tra0001624] [Medline: 38059942]
60. Peng J, Yuan S, Wei Z, Liu C, Li K, Wei X, et al. Temporal network of experience sampling methodology identifies sleep disturbance as a central symptom in generalized anxiety disorder. BMC Psychiatry 2024;24(1):241 [FREE Full text] [doi: 10.1186/s12888-024-05698-z] [Medline: 38553683]
61. Levinson CA, Hunt RA, Keshishian AC, Brown ML, Vanzhula I, Christian C, et al. Using individual networks to identify treatment targets for eating disorder treatment: a proof-of-concept study and initial data. J Eat Disord 2021;9(1):147 [FREE Full text] [doi: 10.1186/s40337-021-00504-7] [Medline: 34736538]

Abbreviations

CD-coefficient: correlation stability coefficient
CHI-T: Cambridge-Chicago Compulsivity Trait Scale
GBIT: Great British Intelligence Test
PUI: problematic use of the internet

XSL•FO
RenderX

Edited by A Mavragani; submitted 05.09.24; peer-reviewed by R Freichel, CP Delaquis; comments to author 29.11.24; revised version received 17.01.25; accepted 06.02.25; published 26.03.25.

Please cite as:
Liu C, Chamberlain S, Ioannidis K, Tiego J, Grant J, Yücel M, Hellyer P, Lochner C, Hampshire A, Albertella L
Transdiagnostic Compulsivity Traits in Problematic Use of the Internet Among UK Residents: Cross-Sectional Network Analysis Study
J Med Internet Res 2025;27:e66191
URL: https://www.jmir.org/2025/1/e66191
doi:10.2196/66191
PMID:40137076

©Chang Liu, Samuel Chamberlain, Konstantinos Ioannidis, Jeggan Tiego, Jon Grant, Murat Yücel, Peter Hellyer, Christine Lochner, Adam Hampshire, Lucy Albertella. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of e-Mental Health Tools for Suicide Prevention in Clinical Practice by Mental Health Professionals in NSW, Australia: Cross-Sectional Survey

Carol Hood[1,2,3], BPsyc(Hons); Sally Hunt[1,2], PhD; Alexandra P Metse[1,4], PhD; Rebecca K Hodder[2,5,6], PhD; Kim Colyvas[1], MStats; Rachel Sheather-Reid[3], PhD; David Duerden[3], MPH, MHA; Jenny Bowman[1,2], PhD

[1]School of Psychological Sciences, The University of Newcastle, Callaghan, Australia

[2]Hunter Medical Research Institute, Newcastle, Australia

[3]Central Coast Mental Health Service, Central Coast Local Health District, Gosford, Australia

[4]School of Health, University of the Sunshine Coast, Sippy Downs, Australia

[5]Hunter New England Population Health, Hunter New England Local Health District, Wallsend, Australia

[6]School of Medicine and Public Health, The University of Newcastle, Callaghan, Australia

Corresponding Author:
Carol Hood, BPsyc(Hons)
School of Psychological Sciences
The University of Newcastle
College of Engineering, Science and Environment
Callaghan, 2308
Australia
Phone: 61 2 4921 5958
Email: Carol.Hood@uon.edu.au

Abstract

Background: Suicide is a significant global health concern. In the context of increased demand for mental health services and workforce shortages, exacerbated by the COVID-19 pandemic, electronic mental health (eMH) tools represent a promising means of augmenting mental health care generally and for suicide prevention specifically. A significant research gap exists however with respect to the use and uptake of eMH tools, especially electronic mental health tools for suicide prevention (eMH-SP).

Objective: This study aimed to investigate the use of eMH tools by Australian mental health professionals, both in general and with respect to suicide prevention specifically, examining changes in use since COVID-19. Further, it explored factors associated with frequent use of eMH-SP, including sociodemographic and professional characteristics.

Methods: A web-based cross-sectional survey was conducted across 15 local health districts (LHDs) in New South Wales, Australia, from May 2022 to July 2023. The sample was drawn from over 10,000 mental health professionals working in government services statewide. The survey explored the use of electronic mental health tools for general mental health issues (eMH-gen) and eMH-SP, explored the changes in the use of both since COVID-19, and used multivariable logistic regression to identify factors associated with the current use of eMH-SP.

Results: Among 469 participants, increased use since COVID-19 was reported by over half (247/469, 52.7%) for eMH-gen, and by approximately one-third (141/386, 36.6%) for eMH-SP. The proportion reporting frequent use increased significantly from before to after COVID-19 for both eMH-gen (243/469, 51.8% to 283/469, 60.3%; $P<.001$) and eMH-SP (152/386, 39.4% to 170/385, 44.2%; $P=.01$). Since COVID-19, the most frequently used types of eMH tools for eMH-gen and eMH-SP, respectively, were information sites (231/469, 49.3% and 130/385, 33.8%), phone/online counseling (173/469, 36.9% and 130/385, 33.8%), and apps (145/469, 30.9% and 107/385, 27.8%). Professionals more likely to use eMH-SP frequently were females (odds ratio [OR] 3.32, 95% CI 1.88-5.87; $P<.001$) compared with males; peer workers (OR 2.17, 95% CI 1.0-4.71; $P<.001$) compared with nurses; those located in regional/rural LHDs (OR 1.65, 95% CI 1.04-2.61; $P=.03$) compared with metropolitan LHDs; and those practicing in emergency health care settings (OR 8.31, 95% CI 2.17-31.75; $P=.03$) compared with inpatient settings.

Conclusions: The study's findings highlight the increasing adoption of eMH tools and delivery of remote care by mental health professionals and provide valuable new insights into sociodemographic factors associated with the use of eMH for suicide

XSL•FO
RenderX

prevention specifically. Continued research on the role eMH is playing is essential for guiding policy, optimizing resources, and enhancing mental health care and suicide prevention efforts.

(J Med Internet Res 2025;27:e64746) doi:10.2196/64746

KEYWORDS

suicide prevention; digital mental health; mental health professionals; peer support; internet; mobile apps; clinical practice; cross-sectional survey; Australia; e-mental health tools

Introduction

Background

Suicide is a leading global cause of death. In 2019, over 700,000 deaths were attributed to suicide worldwide; approximately one in every 100 deaths (1.3%) [1]. Suicide has a profound impact that extends beyond individuals who die by suicide, with those bereaved by suicide at higher risk of suicide themselves [2], and extensive social and economic impacts across the broader community. Suicide prevention initiatives are therefore crucial for decreasing suicide mortality and its wider impacts. The World Health Organization considers suicide preventable if effective evidence-based interventions, treatment, and support are provided in a timely manner [3].

In many nations including Australia, the mental health sector faces challenges in meeting a growing demand for services coupled with existing workforce shortages [4,5]. Pressure on the mental health sector has been further exacerbated by the COVID-19 pandemic (hereafter referred to as "COVID"), where effects have included: increased rates of depression, anxiety, and suicidal behavior [4,6]; reduced opportunity for in-person service contacts; and compounding of a chronic shortage of mental health professionals [7]. Mental health services need accessible and effective solutions that can be scaled up to reach larger numbers of people during times of increased demand.

The use of electronic mental health (eMH) tools to provide remote access to mental health services has emerged as an important part of the solution to bridge this service gap. eMH tools consist of digital technologies, including mobile apps, information sites, online peer support, web-delivered programs, and phone/online counseling [8], which are delivered remotely via electronic communication channels (eg, internet, telephone), and may be clinician-guided or self-managed to provide information or therapy [9]. eMH tools may be designed for standalone use or for integration into routine care to complement in-person support [10]. Their advantages have been noted to include expanded access to mental health services for difficult-to-reach populations, such as clients in geographically remote communities [11], those experiencing stigma or reluctance to engage in-person, or those who prefer anonymity [12]; reduced costs of mental health support [13]; and enhanced continuity or intensity of support (such as when blended with face-to-face care) [14].

Increasing evidence supports the effectiveness of eMH tools in reducing symptoms of various mental health conditions and in preventing suicide. For example, systematic reviews investigating internet-delivered cognitive behavior therapy have found it to be as effective as traditional face-to-face cognitive behavior therapy for anxiety and depressive disorders [15]; and to significantly reduce suicidal ideation compared with controls [16]. Although maintaining end user engagement can be challenging [17], and related to factors such as age, sex, and geographic location [18], end users generally report eMH tools to be acceptable, and have a positive attitude about their use [15,19].

Research reporting the prevalence of use and uptake of eMH tools has focused largely on "consumers" (people with lived experience of mental ill-health) as end users, using a range of measurement approaches and indicating widely varying levels of use [20]. A systematic review into the use of self-help apps and web-based eMH tools for depression, anxiety, or mood enhancement for example (n=11 studies, mostly from Australia and North America), found up to 88% of users had engaged at least minimally with an eMH tool; and that some tools achieved engagement of over 40,000 registrations or downloads monthly [21].

Fewer studies have investigated the prevalence of use and uptake of eMH tools by clinicians and peer workers (hereafter collectively referred to as "professionals") in their work with consumers, with no systematic reviews synthesizing the findings of this limited body of literature being located to date. Two of the larger, more recent studies include a survey of 209 psychiatrists from 19 countries (including Australia) conducted prior to COVID, which indicated that 60% of psychiatrists did not recommend such tools to their patients [22]; and a repeated cross-sectional survey among a mix of psychologists, nurses, social workers, physicians and other professionals (n=1039) in the Netherlands, which found a significant increase in frequent use (monthly or weekly) of videoconferencing to deliver care since COVID [23].

While previous studies have explored the use of eMH tools for various mental health conditions, a crucial gap exists in understanding their use in suicide prevention. Prior to this research, we were able to identify only a recent, qualitative study of 15 Irish professionals describing the adoption of mobile apps for suicide prevention, which reported that while a majority (n=13) recommended apps to clients, only 3 professionals used the apps regularly in practice [24]. To evaluate the impact of eMH tools in the context of suicide prevention, it is vital to understand their use and uptake rates by professionals. This comprehensive study provides the first large-scale look at how mental health professionals use eMH tools for suicide prevention, offering new insights into current practices in this critical area of mental health care provision.

XSL•FO
RenderX

Objectives

This study aimed to first investigate the use of eMH tools by mental health professionals employed in government services in Australia via a web-based survey, including electronic mental health tools for general mental health issues (eMH-gen) and electronic mental health tools for suicide prevention specifically (eMH-SP), and explicitly examine changes in use since COVID. Second, the study explored factors that may influence the frequency of eMH-SP use in practice, including sociodemographic and professional characteristics (age, sex, professional role, rurality of service location, work setting, primary mode of service delivery [whether predominantly in-person, remote, or a blend of both]).

Methods

Design and Setting

A web-based cross-sectional survey of mental health professionals working in the 15 local health districts (LHDs) of the public health system in New South Wales (NSW), Australia, was undertaken. LHD mental health services provide state-wide care for both acute and chronic mental health needs as follows: 6 metropolitan LHDs service the broader Sydney area, with the largest servicing over 1 million residents, while the smallest of 9 regional and rural LHDs services just over 30,000 residents (across nearly 200,000 sq km).

Participants and Recruitment

Eligible participants were professionals (ie, nurses, occupational therapists, peer workers, psychologists, psychiatrists, and social workers) aged 18 years and older, currently providing mental health care within an NSW LHD service. Participants not conforming to these criteria were excluded from the study. Participants were invited to take part in the anonymous survey through emails sent from representatives within their organization. Participants were recruited between May 2022 and July 2023 for the survey, which was open for 8 weeks at each LHD. Three reminder emails were sent from most LHDs at fortnightly intervals to encourage participation (3 LHDs elected to send only 1 or 2 reminders).

Efforts to reduce bias included using neutral phrasing in survey questions, sending multiple reminders to encourage participation, and sampling across diverse demographics and work settings. No formal sample size and power calculation was carried out. Rather, an attempt was made to maximize the potential sample size by contacting all eligible professionals across the 15 NSW LHDs. The final sample size achieved was the result of the response rate to the survey.

Survey Variables and Data Measurement

Overview

The web-based survey was developed through an iterative process that included reviewing existing literature, key stakeholder consultation, and review by an expert panel of mental health professionals (including an Aboriginal scholar and mental health clinician). The survey was deployed on the Research Electronic Data Capture (REDCap) platform [25], where it underwent piloting, with items revised for clarity and appropriateness. Survey data collection comprised predictor variables such as sociodemographic and professional descriptors; an exposure variable of the COVID period (before vs since); and key variables related to the primary outcomes regarding frequency of use of eMH-gen and eMH-SP.

Sociodemographic and Professional Characteristics

Demographic data collected included age, sex, and Aboriginal or Torres Strait Islander origin. Professional characteristics included professional role (nurse; occupational therapist; peer worker; psychiatrist; psychologist; social worker; other), length of service as a mental health professional, suicide prevention training attendance within the previous 3 years (yes/no), LHD of employment (multiple selections permitted from Central Coast; Far West; Hunter New England; Illawarra Shoalhaven; Mid North Coast; Murrumbidgee; Nepean Blue Mountains; Northern NSW; Northern Sydney; South Eastern Sydney; South Western Sydney; Southern NSW; Sydney; Western NSW; Western Sydney), work setting (inpatient; emergency; community; other), and area of mental health practice (multiple selections permitted from: Aboriginal; adult; child and adolescent; consultation liaison; forensic; improvement or service development; neuropsychiatry or neuropsychology; older people; perinatal; special projects; other).

The primary mode of health care delivery was calculated based on participants' ratings of the proportion of care (0% to 100%) they delivered via in-person and remote (phone; video; online text) modes. Responses were sought with respect to delivery mode prior to COVID (ie, prior to the World Health Organization declaration of the pandemic on 11 March 2020 [26]) as well as currently (ie, at the time of survey completion).

Use of e-Mental Health Tools

The survey defined eMH tools as "digital resources and tools delivered via online or telephone technologies that support provision of mental health care", and categorized the types of eMH tools according to the e-Mental Health in Practice website [8], a comprehensive directory of evidence-based Australian, digital mental health resources. The categories of tool types include apps, information sites, online peer support, online self-directed programs, online coach-assisted programs, and phone/online counseling. The e-Mental Health in Practice directory groups eMH tools by topic, including general issues such as mental health conditions, relationship issues, or trauma (eMH-gen); and suicide prevention (eMH-SP).

Participants were asked about their use of eMH tools in terms of (1) change in the extent of eMH-gen and eMH-SP use overall since COVID, rated on a 5-point Likert scale (reduced considerably; reduced somewhat; no change; increased somewhat; increased considerably); (2) frequency of use, based on participants' most frequently used type of eMH tool, rated on a 5-point Likert frequency scale (never, less than monthly, monthly, weekly, daily; the responses were sought separately for eMH-gen and eMH-SP, and use was compared before and since COVID for both); and (3) types of eMH used, based on the frequency of use for each individual tool type rated on the same 5-point Likert scale (never to daily), with responses sought

XSL•FO
RenderX

separately for the types of eMH-gen and eMH-SP, and use compared before and since COVID for both.

Statistical Methods

Descriptive statistics were used to characterize the sample and explore the prevalence of the use of eMH, both eMH-gen, and eMH-SP. Continuous data were presented as means and SDs, and categorical data as frequencies and corresponding percentages. Condensed response categories were created based on age, length of service, LHD location, and delivery mode.

Binary categories were created for the frequency of eMH use and for types of eMH tools (rarely [never, less than monthly, and monthly], frequently [weekly and daily]). McNemar tests were used to assess the significance of differences before and since COVID. To assess differences in delivery modes before and since COVID, paired data were first assessed for normality using Shapiro-Wilk tests and skewness values. Based on these assessments, Wilcoxon signed rank tests (for non-normally distributed data) were carried out, with the α level set to .01 after applying the Bonferroni correction to adjust for multiple comparisons, reducing the risk of type I errors. To address the potential impact of participants dropping out early before completing questions on eMH-SP, such participants were compared with survey completers using chi-square analyses to examine for differences in sociodemographic characteristics (age, sex, and Aboriginal or Torres Strait Islander origin) and professional attributes (professional role, length of service, suicide prevention training, LHD location, and work setting).

Univariate chi-square tests were conducted to identify associations between categorical variables, in this case, factors suggested by the literature as theoretically likely to be associated with frequent eMH use. The following independent variables were considered: age, sex, professional role, LHD location, work setting, and delivery mode. Univariate analysis was carried out on a condensed 3-point Likert scale (never, occasionally [monthly or less than monthly], and frequently [weekly or daily]) using chi-square analysis to test the significance of the independent variables. Those that were significant were entered into a multiple-variable binary logistic regression model. To enable this, the outcome variable was dichotomized into binary categories (rarely [never or occasionally] and frequently) for the multivariable analysis. The independent variables considered in the logistic regression model were all the variables examined in the univariate analysis, as they were all significant. None of the variables were removed from the model to comprehensively assess the combined effects of these variables. The study was conducted and reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [27]. Data were analyzed using the IBM Statistical Package for the Social Sciences (version 29; IBM Corp).

Ethics Approval

Ethics approval was received from the Hunter New England Human Research Ethics Committee (approval 2021/ETH00613) and the University of Newcastle Human Research Ethics Committee (approval H-2022-0034). Local research governance authorization was obtained from each participating LHD. Participants received an email invitation with an information statement outlining the study's objectives, procedures, and potential risks. The information statement was also on the survey website, which participants reviewed before consenting to participate. Completing the web-based survey questions implied informed consent. Participation was voluntary, with the option to opt out at any time. Responses were anonymous, and the data were stored in encrypted electronic databases accessible only to the research team. No financial compensation was provided.

Results

Participants

Overview

Participants were drawn from all 15 LHDs across NSW, with over 10,000 mental health staff invited to participate (ranging from 33 to 1400 per LHD). Of the 713 respondents who accessed the survey, 677 consented, of whom 645 provided information allowing assessment of eligibility (with 7 then excluded). A final sample of 469 participants provided data on the use of eMH-gen; with 385 participants providing data that enabled inclusion in analysis for the use of eMH-SP (Figure 1). The sociodemographic and professional characteristics of the 469 participants whose data were analyzed in this study are reported in Table 1. No significant differences were found in sociodemographic or professional characteristics between participants who dropped out early and those who completed all survey questions.

Figure 1. Flowchart of the study population. eMH: electronic mental health; eMH-gen: electronic mental health tools for general mental health issues; eMH-SP: electronic mental health tools for suicide prevention.

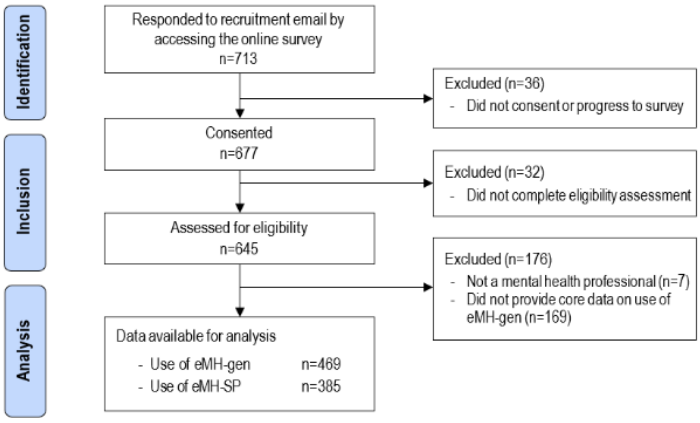

XSL•FO
RenderX

Table 1. Participant sociodemographic and professional characteristics (N=469).

Individual sociodemographic and professional characteristics	Values
Age (years), mean (SD)	46.2 (12.7)
Age (years), range	22-77
Age (years), n (%)	
18-34	103 (22)
35-54	225 (48)
55 and older	141 (30.1)
Sex , n (%)	
Female	339 (72.3)
Male	127 (27.1)
Prefer not to disclose	3 (0.6)
Aboriginal or Torres Strait Islander origin , n (%)	
Not Aboriginal or Torres Strait Islander	444 (94.7)
Aboriginal	13 (2.8)
Aboriginal or Torres Strait Islander	1 (0.2)
Do not know	3 (0.6)
Prefer not to disclose	8 (1.7)
Professional role , n (%)	
Nurse	216 (46.1)
Psychologist	55 (11.7)
Social worker	53 (11.3)
Peer worker	47 (10)
Occupational therapist	37 (7.9)
Psychiatrist	30 (6.4)
Other (eg, welfare officer, mental health practitioner)	31 (6.6)
Length of service (years), mean (SD)	14.3 (11.6)
Length of service (years), range	1-53
Length of service, n (%)	
0-5 years	133 (28.4)
6-10 years	92 (19.6)
11-20 years	131 (27.9)
21 years and over	113 (24.1)
Suicide prevention training within last 3 years, n (%)	294 (62.7)
Location of mental health service[a], n (%)	
Regional or rural LHD[b]	260 (55.4)
Metropolitan LHD	204 (43.5)
LHD not disclosed	5 (1.1)
Work setting, n (%)	
Community	277 (59.1)
Inpatient	139 (29.6)
Emergency	18 (3.8)
Other (eg, a combination of any or all the above)	35 (7.5)

XSL•FO
RenderX

Individual sociodemographic and professional characteristics	Values
Area of practice (multiple selections permitted), n (%)	
Adult	340 (72.5)
Child and adolescent	138 (29.4)
Older people	113 (24.1)
Aboriginal	99 (21.1)
Forensic	46 (9.8)
Perinatal	38 (8.1)
Consultation liaison	37 (7.9)
Special projects	33 (7)
Improvement or service development	24 (5.1)
Neuropsychiatry or neuropsychology	8 (1.7)
Other	31 (6.6)

[a]One participant worked at both a metropolitan and a regional/rural LHD (assigned to the metropolitan LHD), and several participants worked across multiple (2-3) metropolitan or regional/rural LHDs).

[b]LHD: local health district.

Primary Mode of Health Care Delivery

Initial analyses revealed significant deviations from normality in the data (Shapiro-Wilk tests: all $P<.001$; skewness values >±1, except one variable with skewness of –0.569), necessitating the use of Wilcoxon signed rank tests. The Wilcoxon signed rank tests revealed significant changes in the proportion of care delivered via the different modes (in-person and remote) from pre-COVID to since COVID. In-person care remained the primary mode of delivery but decreased significantly from 78.6% pre-COVID to 66.5% since COVID ($z=11.07$, $P<.001$, N=467). Conversely, all modes of remote care delivery increased, with significant rises in phone-based care from 13.2% to 17.0% ($z=6.96$, $P<.001$, N=463), video-based care from 2.7% to 10.3% ($z=10.98$, $P<.001$, N=458), and online text-based care from 2.3% to 3.7% ($z=2.71$, $P=.007$, N=462). While most participants used a combination of 2 to 3 delivery modes, the percentage using exclusively in-person delivery modes decreased significantly from 33.9% pre-COVID to 24.5% since COVID ($P<.001$).

Use of e-Mental Health Tools

Overview

This section presents findings on mental health professionals' use of eMH tools. The frequency of eMH use (including both detailed frequency levels and broader binary categories of frequently and rarely used), and the types of eMH tools they frequently use are presented in Table 2.

XSL•FO
RenderX

Table 2. Frequency of use of e-Mental Health (eMH) tools and types of eMH used frequently (daily or weekly), for electronic mental health tools for general mental health issues (eMH-gen) and electronic mental health tools for suicide prevention specifically (eMH-SP), measured both before and since the onset of COVID.

	Pre-COVID, n (%)	Since COVID, n (%)	Significance (McNemar test *P* value)
Frequency of use of eMH-gen			
Total, n	469	469	—[a]
Never	28 (6)	29 (6.2)	—
Less than monthly	80 (17.1)	50 (10.7)	—
Monthly	118 (25.2)	107 (22.8)	—
Weekly	164 (35)	177 (37.7)	—
Daily	79 (16.8)	106 (22.6)	—
Combined categories of use			
Frequently (weekly or daily)	243 (51.8)	283 (60.3)	<.001
Rarely (never or <monthly or monthly)	226 (48.2)	186 (39.7)	—
Frequency of use of eMH-SP			
Total, n	386	385	—
Never	53 (13.7)	45 (11.7)	—
Less than monthly	91 (23.6)	80 (20.8)	—
Monthly	90 (23.3)	90 (23.4)	—
Weekly	95 (24.6)	106 (27.5)	—
Daily	57 (14.8)	64 (16.6)	—
Combined categories of use			
Frequently (weekly or daily)	152 (39.4)	170 (44.2)	.01
Rarely (never or <monthly or monthly)	234 (60.6)	215 (55.8)	—
Types of eMH-gen used frequently (daily or weekly)			
Total, n	469	469	—
Information sites	180 (38.4)	231 (49.3)	<.001
Phone/online counseling	168 (35.8)	173 (36.9)	.67
Apps	120 (25.6)	145 (30.9)	.01
Online self-directed programs	65 (13.9)	90 (19.2)	<.001
Online peer support	47 (10)	68 (14.5)	.002
Online coach-assisted programs	43 (9.2)	60 (12.8)	.01
Other	37 (7.9)	50 (10.7)	.04
Types of eMH-SP used frequently (daily or weekly)			
Total, n	386	385	—
Information sites	114 (29.5)	130 (33.8)	.03
Phone/online counseling	113 (29.3)	130 (33.8)	.01
Apps	57 (14.8)	107 (27.8)	<.001
Online self-directed programs	42 (10.9)	62 (16.1)	.002
Online peer support	43 (11.1)	74 (19.2)	<.001
Online coach-assisted programs	26 (6.7)	43 (11.2)	<.001
Other	19 (4.9)	33 (8.6)	<.001

[a]Not applicable.

XSL•FO
RenderX

Changes in Use of eMH

Over half the participants (247/469, 52.7%) reported an increase in their use of eMH-gen since COVID; 43.3% (n=203) reported no change, and a small minority (19/469, 4%) reported a decrease in use. Regarding eMH-SP, over one-third of participants (141/385, 36.6%) reported an increase in their use, over half reported no change in their frequency of use (226/385, 58.7%), while less than 5% (n=18) reported a reduction in the frequency of eMH-SP use.

Frequency of Use of eMH

Among participants, 60.3% (n=283) reported currently using eMH-gen frequently, compared with 51.8% (n=243) pre-COVID, a significant increase ($P<.001$). Similarly, 44.2% (n=170) reported frequently using eMH-SP compared with 39.4% (n=152) pre-COVID, a significant increase ($P=.01$). Comparing frequent use of eMH-gen with eMH-SP since COVID, the difference was also significant ($P<.001$).

Types of eMH Tools Used

Of the 6 types of eMH tools assessed, the 3 eMH-gen currently used most frequently were information sites (231/469, 49.3%), phone/online counseling (173/469, 36.9%), and apps (145/469, 30.9%), with significant increases since COVID in both information sites ($P<.001$) and apps ($P=.01$). Among these 6 types of tools, the 3 eMH-SP currently used most frequently, all of which showed significant increases since COVID, were information sites (130/385, 33.8%; $P=.03$), phone/online counseling (130/385, 33.8%; $P=.01$), and apps (107/385, 27.8%; $P<.001$).

Sociodemographic and Professional Characteristics Associated With the Use of eMH-SP in Current Practice

Current frequency of use of eMH-SP was classified into one of 3 outcome categories (never, occasionally, and frequently), with possible associations explored with sociodemographic and professional characteristics (Table 3). Chi-square test results indicate significant univariate associations between all of the factors and the frequency of use of eMH-SP. Professionals of younger age, female sex, professional role of peer worker or social worker, working in regional or rural LHDs, in emergency health care settings, and using at least 1 remote delivery mode (as opposed to in-person mode only) were more likely to report current frequent (weekly or daily) use of eMH-SP.

Multivariable binary logistic regression was then used (categories: rarely; frequently) to further explore the relationship between these factors and frequent use of eMH-SP (ie, on a daily or weekly basis), as presented in Table 4. The overall model was found to be significant ($\chi^2_{14}=70.89$, n=378, and $P<.001$). Age ($P=.06$) and delivery mode ($P=.17$) were no longer significant, while significance was retained for all other factors. Frequent use of eMH-SP was 3.32 times more likely for females than males, 2.17 times more likely for peer workers than nurses, 1.65 times more likely for regional/rural LHDs than metropolitan LHDs, and 8.31 times more likely for emergency settings than inpatient settings.

XSL•FO
RenderX

Table 3. Sociodemographic and professional characteristics associated with the frequency of use of electronic mental health tools for suicide prevention (eMH-SP) in current practice (N=385, percentages by rows).

	Never, n (%)	Occasionally (monthly or less), n (%)	Frequently (weekly or daily), n (%)	Chi-square (*df*)	*P* value
Age (years)				13.65 (4)	.008
18-34 (n=79)	7 (8.9)	34 (43)	38 (48.1)		
35-54 (n=185)	15 (8.1)	78 (42.2)	92 (49.7)		
55 and older (n=121)	23 (19)	58 (47.9)	40 (33.1)		
Sex				18.76 (4)	<.001
Female (n=278)	28 (10.1)	109 (39.2)	141 (50.7)		
Male (n=104)	16 (15.4)	60 (57.7)	28 (26.9)		
Prefer not to disclose (n=3)	1 (33.3)	1 (33.3)	1 (33.3)		
Professional role				34.72 (12)	<.001
Peer worker (n=40)	0 (0)	14 (34.1)	26 (65.9)		
Social worker (n=40)	6 (15)	11 (27.5)	23 (57.5)		
Nurse (n=179)	24 (13.4)	81 (45.3)	74 (41.3)		
Psychologist (n=47)	2 (4.3)	27 (57.4)	18 (38.3)		
Psychiatrist (n=24)	4 (16.7)	14 (58.3)	6 (25)		
Occupational therapist (n=28)	7 (25.9)	15 (55.6)	6 (18.5)		
Other (n=27)	2 (7.4)	8 (29.6)	17 (63)		
LHD[a] location (n=381)				10.41 (2)	.005
Regional/rural LHD (n=213)	21 (9.9)	82 (38.5)	110 (51.6)		
Metropolitan LHD (n=168)	23 (13.7)	86 (51.2)	59 (35.1)		
Work setting				14.24 (6)	.03
Emergency (n=14)	1 (7.1)	3 (21.4)	10 (71.4)		
Community (n=226)	20 (8.8)	98 (43.4)	108 (47.8)		
Inpatient (n=113)	20 (17.7)	56 (49.6)	37 (32.7)		
Other (>1 setting, n=32)	4 (12.5)	13 (40.6)	15 (46.9)		
Delivery mode				18.21 (2)	<.001
In-person or remote (n=295)	24 (8.1)	129 (43.7)	142 (48.1)		
In-person only (n=90)	21 (23.3)	41 (45.6)	28 (31.1)		

[a]LHD: local health district.

XSL•FO
RenderX

Table 4. Binary logistic regression model results with regression coefficients (B), standard errors (SE), odds ratios (OR), and 95% CIs for factors associated with the likelihood of frequent use of electronic mental health tools for suicide prevention (eMH-SP) in current practice (N=378).

	Values, n (%)	B	SE	*P* value	Odds ratio (95% CI)	Likelihood ratio test	
						Chi-square (*df*)	*P* value
Age group (years)						5.570 (2)	.06
18-34	78 (20.6)	0.70	0.35	.04	2.01 (1.02-3.97)		
35-54	183 (48.4)	0.56	0.27	.04	1.75 (1.03-2.99)		
55 and older (reference)	117 (31)	0	—[a]	—	1.00[b]		
Sex						18.791 (1)	<.001
Female	276 (73)	1.20	0.29	<.001	3.32 (1.88-5.87)		
Male (reference)	102 (27)	0	—	—	1.00[b]		
Professional role						22.525 (6)	<.001
Peer worker	40 (10.6)	0.77	0.40	.05	2.17 (1.0-4.71)		
Social worker	39 (10.3)	0.16	0.40	.68	1.18 (0.54-2.56)		
Psychologist	47 (12.4)	–0.50	0.37	.18	0.61 (0.29-1.27)		
Psychiatrist	22 (5.8)	–0.88	0.58	.13	0.42 (0.13-1.30)		
Occupational therapist	27 (7.1)	–1.41	0.53	.01	0.24 (0.09-0.69)		
Other	26 (6.9)	0.70	0.47	.14	2.02 (0.80-5.08)		
Nurse (reference)	177 (46.8)	0	—	—	1.00[b]		
LHD[c] location						4.561 (1)	.03
Rural/regional	211 (55.8)	0.50	0.23	.03	1.65 (1.04-2.61)		
Metropolitan (reference)	167 (44.2)	0	—	—	1.00[b]		
Setting						9.294 (3)	.03
Emergency	14 (3.7)	2.12	0.68	.002	8.31 (2.17-31.75)		
Community	222 (58.7)	0.67	0.27	.01	1.95 (1.15-3.32)		
Other (eg, >1 setting)	31 (8.2)	0.65	0.46	.17	1.91 (0.77-4.73)		
Inpatient (reference)	111 (29.4)	0	—	—	1.00[b]		
Delivery mode						1.851 (1)	.17
In-person or remote	288 (76.2)	0.51	0.38	.18	1.66 (0.80-3.47)		
In-person only (reference)	90 (23.8)	0	—	—	1.00[b]		

[a]Not applicable.

[b]The parameter of B for each of the reference categories is set to 0 because it is redundant and there are no confidence intervals for the reference categories.

[c]LHD: local health district.

Discussion

Principal Findings

This study invited mental health professionals in government mental health services across NSW Australia to take part in a survey with respect to the use of eMH tools in their professional practice. It focused particularly on the use of eMH-SP, given the urgent need for strengthening suicide prevention efforts in the face of consistently high suicide rates. Understanding the uptake of eMH is especially important in the context of challenges in meeting increasing demand for mental health care, both generally and for suicide prevention; with demand heightened further by the COVID pandemic. Study findings reflected a shift toward the increasing incorporation of remote modes of care delivery (video, phone, online text) since COVID, with the proportion reporting delivering care solely “in-person” declining from one-third to one-quarter.

Over half (247/469, 52.7%) of participants reported an increase in their use of eMH-gen since COVID, while 43.3% (n=203) reported no change. Over one-third (141/386, 36.6%) similarly reported their use of eMH-SP had increased, with 58.7% (n=226) reporting their use remained unchanged. A majority (283/469, 60.3%) of professionals reported frequent use of eMH-gen (at least weekly); a proportion that had increased by approximately

XSL•FO
RenderX

10% since COVID. While a lesser proportion of participants reported frequent use of eMH-SP specifically, both before and since COVID, there was still a notable increase in its use. Nearly half (170/386, 44.2%) of the sample reported frequent use of eMH-SP since COVID, up from 39.4% (n=152) prior to the pandemic. Approximately half of the participants (231/469, 49.3%) reported information sites as the most frequently used type of eMH-gen, followed by phone/online counseling (173/469, 36.9%). The same types of tools were also most frequently used with respect to suicide prevention, with approximately one-third reporting information sites and phone/online counseling as their most frequently used tools. Since COVID, the use of phone/online counseling for suicide prevention has increased significantly, in contrast to its use for general mental health issues. This shift may have been driven by the pandemic, which limited in-person counseling and support for suicide prevention, instead prompting increased use of eMH-SP. The most marked increases in the frequency of use by types of eMH tools were observed for information sites in eMH-gen (from 180/469, 38.4% to 231/469, 49.3%), and for suicide prevention apps in eMH-SP (from 57/385, 14.8% to 107/385, 27.8%). This increase in overall use is consistent with the rapidly increasing availability and promotion of digital mental health resources in practice [28].

A small but consistent proportion of professionals (≈6% both pre-COVID and currently) reported never using eMH-gen, while roughly double this proportion (≈14% pre-COVID and 12% currently) reported never using eMH-SP. Over half of respondents reported no or very limited (monthly or less) use of eMH-SP currently, as compared with approximately one-third who reported no or very limited use of eMH-gen. To some extent, this lesser use for suicide prevention specifically might be related to a lack of knowledge about what eMH-SP are available and are indicated for use with clients experiencing suicidal ideation [29], as well as concerns about the safety and effectiveness of these tools in managing suicidal crises [30]. It is also however likely explained in part by some clinicians encountering clients at high risk of suicide relatively infrequently in their clinical practice. This is lent support by the finding that the use of eMH-SP did differ by setting; frequent use being highly reported by the small number of survey respondents who worked in emergency settings (less than 4% of participants), where presentations involving suicidal ideation and suicide attempts would be more common.

Frequent use of eMH-SP was also more likely for female mental health professionals as compared with males. While this finding aligns with the general trend of higher digital tool use among females [31], no corroborating evidence specific to mental health professionals was found in the existing literature. The finding that some professional groups, most notably peer workers and social workers, were more likely to report frequent use of eMH-SP than others, might suggest that daily routines and demands of the various professional roles influence levels of use of these tools, but further research is needed to elucidate this relationship. Notably, all peer workers used eMH-SP to some extent, with almost two-thirds using these tools frequently. It might be speculated that the finding of higher use in regional or rural LHDs may be due to participants in such locations being more likely to promote eMH adoption to compensate for limited access to mental health resources compared with those in metropolitan LHDs [10].

The delivery of mental health care and suicide prevention is continuing to evolve, with the findings of this study supporting previous research suggesting that remote delivery modes and the use of eMH tools are becoming more common [32]. This raises questions about potential barriers to eMH adoption, especially in suicide prevention, where clinicians may face added safety and ethical concerns about recommending appropriate eMH tools to consumers at heightened risk of suicide [33]. It also creates the need for mental health services to respond and provide appropriate leadership, infrastructure access, and training to allow mental health professionals to optimize the use of eMH tools, despite the challenges. The rapid growth in the number of available eMH-SP, coupled with the variable expertise involved in their development, has resulted in inconsistent quality across these tools [34,35]. The use of eMH is already quite significant and likely to continue rising, and brings with it the need for clear, best practice guidelines to direct mental health professionals toward the most effective and safest eMH-SP for clinical practice.

This study offers valuable insights into eMH use by diverse mental health professionals in NSW, both before and since the COVID pandemic. One significant strength is its comprehensive reach, encompassing all 15 LHDs across NSW and inviting over 10,000 mental health staff to participate. However, there are limitations to consider for the study. The results are based on a self-report survey, which may exhibit response bias. The research was conducted in a single Australian state, limiting the generalizability of the results to other contexts. However, NSW is the most populous state with the highest number of mental health professionals [36]. The survey had a modest participation rate (4.7%), which nevertheless equated to a substantial sample of 469 participants. The dropout rate was 17.9%, with 84 of the 469 eligible participants failing to complete the survey, although there were no significant differences in sociodemographic or professional characteristics between survey dropouts and completers. Self-selection bias may have been introduced, with those who opted to complete the survey likely to have a stronger interest in eMH and be more likely to use it.

Despite such limitations, this study makes a significant contribution to the field by systematically documenting real-life practices of mental health professionals' use of eMH tools since the COVID pandemic, with a focus on suicide prevention. Unlike previous research, this study provides a comprehensive, quantitative overview that not only assesses mental health professionals' adoption of eMH tools but also provides important insights about sociodemographic factors that may influence the uptake of these tools into practice. Building on previous qualitative insights, this study provides a more comprehensive, data-driven understanding of eMH tool use in suicide prevention. Using a quantitative approach, the study offers statistically robust evidence beyond the limited scope of existing literature.

XSL•FO
RenderX

Conclusion

It is vitally important to continue to explore avenues for increasing timely access to quality mental health support, particularly in the area of suicide prevention. In order to promote appropriate uptake of eMH-SP in practice, it is imperative to understand the perceived benefits, challenges, and barriers to use. Future research could explore the factors that clinicians report influence the adoption of eMH-SP in greater detail, possibly using quantitative or qualitative approaches with selected mental health professionals and perhaps different professional groups, to gain deeper insights into their experiences and motivations for using or not using eMH tools for suicide prevention. Implementation research is needed to identify strategies for mental health services to support clinicians in their optimal use of these tools, which may include, for instance, best practice guidelines. Research in these areas is essential for guiding policy and resource allocation, improving the quality of mental health care, and enhancing suicide prevention efforts in this critical area of public health.

Acknowledgments

The authors wish to express our sincere gratitude to the expert panel who reviewed the survey instrument and provided valuable feedback to strengthen its validity and relevance, and to the mental health professionals across New South Wales who participated in the online survey. The New South Wales Ministry of Health supported this research as part of the Towards Zero Suicides Initiative, which included the funding of a PhD scholarship. However, the funding source played no role in the writing of the manuscript or the decision to submit it for publication. Authors were not precluded from accessing data in the study, and they accept responsibility to submit for publication.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to ethical restrictions and privacy considerations (eg, data containing potentially identifying information), but are available from the corresponding author on reasonable request.

Authors' Contributions

All authors made substantial contributions to the conception and design, or acquisition of data, or analysis and interpretation of data, specifically. CH contributed to conception and design, methodology, data acquisition, analysis and interpretation, and project administration. JB, SH, APM, and RKH participated in supervision, conception and design, methodology, data analysis, and interpretation. KC contributed to methodology, formal data analysis, and interpretation. RSR and DD contributed to the conception, design, and methodology. CH drafted the original article, and all authors revised it critically for important intellectual content. All authors granted final approval of the version to be published. No generative artificial intelligence was used in the manuscript writing.

Conflicts of Interest

None declared.

References

1. Suicide worldwide in 2019: global health estimates. World Health Organization. Geneva; 2021. URL: https://www.who.int/publications/i/item/9789240026643 [accessed 2025-02-25]
2. Molina N, Viola M, Rogers M, Ouyang D, Gang J, Derry H, et al. Suicidal ideation in bereavement: a systematic review. Behav Sci (Basel) 2019;9(5):53 [FREE Full text] [doi: 10.3390/bs9050053] [Medline: 31091772]
3. Preventing suicide: a global imperative. World Health Organization. 2014. URL: https://www.who.int/publications/i/item/9789241564779 [accessed 2025-02-25]
4. Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief, 2 March 2022. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 [accessed 2025-02-25]
5. National mental health workforce strategy 2022-2032. Commonwealth of Australia (Department of Health and Aged Care). 2022. URL: https://www.health.gov.au/sites/default/files/2023-10/national-mental-health-workforce-strategy-2022-2032.pdf [accessed 2025-02-25]
6. Dubé JP, Smith MM, Sherry SB, Hewitt PL, Stewart SH. Suicide behaviors during the COVID-19 pandemic: a meta-analysis of 54 studies. Psychiatry Res 2021;301:113998 [FREE Full text] [doi: 10.1016/j.psychres.2021.113998] [Medline: 34022657]
7. Ellis LA, Meulenbroeks I, Churruca K, Pomare C, Hatem S, Harrison R, et al. The application of e-mental health in response to COVID-19: scoping review and bibliometric analysis. JMIR Ment Health 2021;8(12):e32948 [FREE Full text] [doi: 10.2196/32948] [Medline: 34666306]
8. Our digital mental health resource directory. eMHPrac. 2024. URL: https://www.emhprac.org.au/directory/ [accessed 2025-02-25]

XSL•FO
RenderX

9. Kreuze E, Jenkins C, Gregoski M, York J, Mueller M, Lamis DA, et al. Technology-enhanced suicide prevention interventions: a systematic review. J Telemed Telecare 2017;23(6):605-617. [doi: 10.1177/1357633X16657928] [Medline: 27377792]
10. Reynolds J, Griffiths KM, Cunningham JA, Bennett K, Bennett A. Clinical practice models for the use of e-mental health resources in primary health care by health professionals and peer workers: a conceptual framework. JMIR Ment Health 2015;2(1):e6 [FREE Full text] [doi: 10.2196/mental.4200] [Medline: 26543912]
11. Kavanagh BE, Corney KB, Beks H, Williams LJ, Quirk SE, Versace VL. A scoping review of the barriers and facilitators to accessing and utilising mental health services across regional, rural, and remote Australia. BMC Health Serv Res 2023;23(1):1060 [FREE Full text] [doi: 10.1186/s12913-023-10034-4] [Medline: 37794469]
12. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the james lind alliance priority setting partnership. Lancet Psychiatry 2018;5(10):845-854 [FREE Full text] [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
13. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GLD, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
14. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
15. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
16. Torok M, Han J, Baker S, Werner-Seidler A, Wong I, Larsen ME, et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2020;2(1):e25-e36 [FREE Full text] [doi: 10.1016/S2589-7500(19)30199-2] [Medline: 33328037]
17. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
18. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
19. Chan AHY, Honey MLL. User perceptions of mobile digital apps for mental health: acceptability and usability - an integrative review. J Psychiatr Ment Health Nurs 2022;29(1):147-168. [doi: 10.1111/jpm.12744] [Medline: 33604946]
20. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
21. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
22. Bauer R, Glenn T, Monteith S, Whybrow PC, Bauer M. Survey of psychiatrist use of digital technology in clinical practice. Int J Bipolar Disord 2020;8(1):29 [FREE Full text] [doi: 10.1186/s40345-020-00194-1] [Medline: 33009954]
23. Feijt M, de Kort Y, Westerink J, Bierbooms J, Bongers I, IJsselsteijn W. Integrating technology in mental healthcare practice: a repeated cross-sectional survey study on professionals' adoption of digital mental health before and during COVID-19. Front Psychiatry 2022;13:1040023 [FREE Full text] [doi: 10.3389/fpsyt.2022.1040023] [Medline: 36874171]
24. Melia R, Monahan L, Duggan J, Bogue J, O'Sullivan M, Young K, et al. Exploring the experiences of mental health professionals engaged in the adoption of mobile health technology in Irish mental health services. BMC Psychiatry 2021;21(1):412 [FREE Full text] [doi: 10.1186/s12888-021-03426-5] [Medline: 34412601]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. WHO director-general's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2025-02-25]
27. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
28. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]

XSL•FO
RenderX

29. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Understanding mental health professionals' perspectives and practices regarding the implementation of digital mental health: qualitative study. JMIR Form Res 2022;6(4):e32558 [FREE Full text] [doi: 10.2196/32558] [Medline: 35412459]
30. Hofstra E, van Nieuwenhuizen C, Bakker M, Özgül D, Elfeddali I, de Jong SJ, et al. Effectiveness of suicide prevention interventions: a systematic review and meta-analysis. Gen Hosp Psychiatry 2020;63:127-140. [doi: 10.1016/j.genhosppsych.2019.04.011] [Medline: 31078311]
31. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
32. Bouabida K, Lebouché B, Pomey M. Telehealth and COVID-19 pandemic: an overview of the telehealth use, advantages, challenges, and opportunities during COVID-19 pandemic. Healthcare (Basel) 2022;10(11):2293 [FREE Full text] [doi: 10.3390/healthcare10112293] [Medline: 36421617]
33. Rytterström P, Ovox SM, Wärdig R, Hultsjö S. Impact of suicide on health professionals in psychiatric care mental healthcare professionals' perceptions of suicide during ongoing psychiatric care and its impacts on their continued care work. Int J Ment Health Nurs 2020;29(5):982-991 [FREE Full text] [doi: 10.1111/inm.12738] [Medline: 32419316]
34. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
35. Torous J, Firth J, Huckvale K, Larsen ME, Cosco TD, Carney R, et al. The emerging imperative for a consensus approach toward the rating and clinical recommendation of mental health apps. J Nerv Ment Dis 2018;206(8):662-666. [doi: 10.1097/NMD.0000000000000864] [Medline: 30020203]
36. Mental health: workforce 2024. Australian Institute of Health and Welfare. 2024. URL: https://www.aihw.gov.au/mental-health/topic-areas/workforce [accessed 2024-11-20]

Abbreviations

eMH: electronic mental health
eMH-gen: electronic mental health tools for general mental health issues
eMH-SP: electronic mental health tools for suicide prevention
LHD: local health district
NSW: New South Wales
OR: odds ratio
REDCap: Research Electronic Data Capture
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by T de Azevedo Cardoso; submitted 29.07.24; peer-reviewed by A Teles, KHM O'Brien; comments to author 29.11.24; revised version received 20.12.24; accepted 14.02.25; published 26.03.25.

Please cite as:
Hood C, Hunt S, Metse AP, Hodder RK, Colyvas K, Sheather-Reid R, Duerden D, Bowman J
Use of e-Mental Health Tools for Suicide Prevention in Clinical Practice by Mental Health Professionals in NSW, Australia: Cross-Sectional Survey
J Med Internet Res 2025;27:e64746
URL: https://www.jmir.org/2025/1/e64746
doi:10.2196/64746
PMID:

©Carol Hood, Sally Hunt, Alexandra P Metse, Rebecca K Hodder, Kim Colyvas, Rachel Sheather-Reid, David Duerden, Jenny Bowman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

US State Public Health Agencies' Use of Twitter From 2012 to 2022: Observational Study

Samuel R Mendez[1], MS; Sebastian Munoz-Najar[2], PhD; Karen M Emmons[1], PhD; Kasisomayajula Viswanath[1,3], PhD

[1]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[2]Harvard Graduate School of Education, Cambridge, MA, United States

[3]Dana Farber Cancer Institute, Boston, MA, United States

Corresponding Author:
Samuel R Mendez, MS
Department of Social and Behavioral Sciences
Harvard T.H. Chan School of Public Health
677 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 617 432 1135
Email: smendez@g.harvard.edu

Abstract

Background: Twitter (subsequently rebranded as X) is acknowledged by US health agencies, including the US Centers for Disease Control and Prevention (CDC), as an important public health communication tool. However, there is a lack of data describing its use by state health agencies over time. This knowledge is important amid a changing social media landscape in the wake of the COVID-19 pandemic.

Objective: The study aimed to describe US state health agencies' use of Twitter from 2012 through 2022. Furthermore, we organized our data collection and analysis around the theoretical framework of the networked public to contribute to the broader literature on health communication beyond a single platform.

Methods: We used Twitter application programming interface data as indicators of state health agencies' engagement with the 4 key qualities of communication in a networked public: scalability, persistence, replicability, and searchability. To assess scalability, we calculated tweet volume and audience engagement metrics per tweet. To assess persistence, we calculated the portion of tweets that were manual retweets or included an account mention. To assess replicability, we calculated the portion of tweets that were retweets or quote tweets. To assess searchability, we calculated the portion of tweets using at least 1 hashtag.

Results: We observed a COVID-19 pandemic–era shift in state health agency engagement with scalability. The overall volume of tweets increased suddenly from less than 50,000 tweets in 2019 to over 94,000 in 2020, resulting in an average of 5.3 per day. Though mean tweets per day fell in 2021 and 2022, this COVID-19 pandemic–era low was still higher than the pre–COVID-19 pandemic peak. We also observed a more fragmented approach to searchability aligning with the start of the COVID-19 pandemic. More state-specific hashtags were among the top 10 during the COVID-19 pandemic, compared with more general hashtags related to disease outbreaks and natural disasters in years before. We did not observe such a clear COVID-19 pandemic–era shift in engagement with replicability. The portion of tweets mentioning a CDC account gradually rose and fell around a peak of 7.0% in 2018. Similarly, the rate of retweets of a CDC account rose and fell gradually around a peak of 5.4% in 2018. We did not observe a clear COVID-19 pandemic–era shift in persistence. The portion of tweets mentioning any account reached a maximum of 21% in 2013. It oscillated for much of the study period before dropping off in 2021 and reaching a minimum of 10% in 2022. Before 2018, the top 10 mentioned accounts included at least 2 non-CDC or corporate accounts. From 2018 onward, state agencies were much more prominent.

Conclusions: Overall, we observed a more fragmented approach to state health agency communication on Twitter during the pandemic, prioritizing volume over searchability, formally replicating existing messages, and leaving traces of interactions with other accounts.

(J Med Internet Res 2025;27:e59786) doi:10.2196/59786

XSL•FO
RenderX

KEYWORDS

social media; health communication; Twitter; tweet; public health; state government; government agencies; information technology; data science; communication tool; COVID-19 pandemic; data collection; theoretical framework; message; interaction

Introduction

Twitter as a Networked Public

This paper describes US state public health agencies' activity on Twitter (subsequently rebranded as X) from 2012 through 2022. We used the theoretical framework of a networked public, an interactive space that networked technologies make possible, as well as the collection of people that inhabit it [1].

According to this theoretical framework, a networked public's defining features are profiles, contact lists, and public communication tools. This framework outlines 4 emergent qualities of communication these features produce: scalability, persistence, replicability, and searchability. Twitter's features were always open to users with potentially conflicting motivations. Viewed generously, this aligns with notions of Twitter's democratization of health communication [2]. Viewed through a more critical lens, however, this illustrates how Twitter has been a site for both challenging and propagating inequities.

This tension was evident from Twitter's 2006 launch, with dominant narratives of its Silicon Valley origins obscuring its relationship to open-source innovations from activists [3]. The hashtag illustrates such long-term tensions that exist in networked publics, between corporate practices and grassroots discourse, as well as between social networks with opposing goals. Hashtag use began as an informal organizational practice among power users before Twitter formally incorporated it as a feature in 2007 [4]. While activists eventually adopted hashtags for social justice campaigns like #MeToo and #BlackLivesMatter, and corporations adopted hashtags for advertising purposes, hashtags were also tools of political marginalization. For example, the use of the hashtag #ChineseVirus was a marker of anti-Chinese sentiment in March 2020, associated with anti-Asian hashtags more broadly after the US president's use of the term "Chinese Virus" [5]. Hashtags can also be sites of conflict around public health topics, such as the coordinated flooding of the provaccine #DoctorsSpeakUp hashtag with antivaccine messaging in 2020 [6].

Twitter as a State Public Health Tool

Though the use of Twitter in public health has been critiqued as "tweeting to the choir" [7], this type of communication toward professional audiences could reflect a viable platform-specific communication strategy. For example, the US Centers for Disease Control and Prevention (CDC) has recommended that public health communicators and organizational leaders use Twitter to reach reporters with crisis communication [8]. This could help explain why US health agencies' COVID-19 tweets focused more on data than did their Facebook (Meta) posts [9].

Existing observational research reinforces this notion of Twitter as an information-focused public health communication tool. A study of state Medicaid program Twitter use between 2014 and 2019 found a similar informational focus, which garnered little audience engagement [10]. A study of Canadian public health agencies' Twitter use during the first half of 2020 found a similar informational focus, even as more action-oriented tweets received higher audience engagement [11]. A study of health agency use of Twitter across 7 countries during the spring of 2020 found evidence of this trend internationally, with announcements and reporting being the most common tweet theme [12]. At a local level in the US, there have been experiments in using Twitter as a community service tool, such as in the case of the Chicago Department of Public Health's foodborne illness response program [13].

Descriptive studies of US state public health agencies' Twitter use, however, are either not comprehensive or are not up to date. Twitter was found to be an emerging platform for state public health agencies, based on a study covering a 2-month period in 2011 [14]. An early content analysis of state public health agency tweets in 2012 found a focus on the transmission of personal health advice [15] and a much higher Twitter adoption rate than local health departments around the same time [16]. A more recent study analyzed all state health agency tweets from a 4-month period around the emergency use authorization of the first COVID-19 vaccine in the United States [17]. Findings indicated a much higher volume of tweets than earlier studies, as well as a lack of key terms related to vaccination, inequities, and racism, lagging behind on-the-ground efforts to address COVID-19 racial health inequities.

We present an instrumental case study [18] to provide a more complete picture of state public health agency Twitter use long-term. We chose this approach as it aligns with our research interest in producing descriptive knowledge with theoretical implications beyond the individual case of Twitter. Our goal was to provide transferable knowledge to help identify trends and relationships between theoretical constructs in health communication. This knowledge is important amid a changing social media landscape in the wake of the COVID-19 pandemic. We follow STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [19] in reporting this study. The completed STROBE statement checklist is available in Multimedia Appendix 1.

Methods

Overview

We conducted a descriptive analysis of state public health agency tweets from 2012 through 2022, structured as an instrumental case study organized around an issue question [19]: How did US state public health agencies engage with Twitter as a networked public?

Our data came from a download of public health agency tweets using the "academictwitteR" R package to access the Twitter application programming interface (API) through the "Twitter API v2 for Academic Research" on February 3, 2023 [18]. We

XSL•FO
RenderX

filtered the data to focus on tweets in English from state public health agencies. The language filter relied on labels available by Twitter's proprietary API data. We identified tweets from the state-level public health agency in each state by leveraging associations between usernames and author IDs recorded during the data download process [20]. These accounts were initially identified through manual searches of state public health agency websites, as well as searches on Google and Twitter where necessary. Refer to Multimedia Appendix 2 for a list of account usernames and the associated state public health agencies. The study period begins on January 1, 2012, which allowed us to build on a previous study taking a random sampling approach to analyze state public health agency tweets in the same year [15]. The study period ends on December 31, 2022, which was the most recent full calendar year at the time of data collection. We did not account for missing data, as the API download method includes an exhaustive record of all published tweets from all specified accounts during the study period. We conducted this analysis using R (version 4.2.3, R Studio version 2023.06.1+524). The scripts enabling this analysis are available in an Open Science Foundation repository [21].

We calculated indicator variables to describe health agencies' engagement with the key qualities of communication in a networked public: scalability, persistence, replicability, and searchability [1].

Scalability

"Scalability" describes social media content's hypothetical ability to reach an entire network. We assessed scalability by tweet volume and audience engagement. We calculated tweets per year, categorized into types: replies, retweets, or quote tweets (from their formal introduction in 2015) [22]. We further categorized replies into self-replies and other replies by comparing the author and in-reply-to user IDs. We used the regular expression "RT @" to identify "manual retweets," a custom metric describing text copies of messages. Here, we pooled manual retweets with formal retweets. Finally, we calculated the mean, median, and IQR of states' mean tweets per day each year, as well as the likes, replies, retweets, and quotes on non-retweets. "Non-retweets" include manual retweets, as they are unique posts rather than pointers to source tweets.

Persistence

"Persistence" describes the digital records of interactions on networked publics. We assessed persistence by manual retweets. We also calculated mentions (a hyperlinked account username) among non-retweets, as their records do not depend on other accounts' activity. We calculated the percentage of tweets that were manual retweets each year. We calculated the percentage of non-retweets with at least 1 mention each year.

Replicability

"Replicability" refers to social media content's infinite duplicability. We assessed replicability by retweets and quote tweets. We calculated the percentage of tweets that were formal retweets and quote tweets each year. We further calculated the percentage of tweets that were formal retweets of a CDC account, as well as the percentage of tweets including a CDC account mention. We used regular expressions to identify usernames starting with "CDC" or "NIOSH," which comprised the majority of usernames listed as official CDC accounts [23]. For other accounts, we relied on exact matches with entire usernames.

Searchability

"Searchability" refers to ways that social media content is findable, which we assessed by hashtag use. We calculated the percentage of tweets with at least 1 hashtag per year. We used regular expressions to extract hashtags and calculate the top 10 most common per year (after converting to lowercase). We further calculated the percentage of tweets per year using #flu or #hiv as hashtags with long-term national relevance.

Ethical Considerations

The study data consisted of publicly available information from government institutional authors. We did not seek an ethics review and have no individual participant protections in place, as they are not relevant to this study design.

Results

Overview

We identified 570,335 tweets from state public health agency accounts in English from 2012 through 2022. All 50 states were represented in this data set. However, all 50 states were only active on Twitter in 2018 and 2019, as states gradually adopted Twitter during the study period, and 1 state published its last tweet in 2019. We included 52 accounts in this data set to account for 1 state creating a replacement public health agency account during the study period, as well as another state communicating through 2 accounts under its public health agency. An exhaustive summary data table is available in Multimedia Appendix 3.

Scalability

In total, 41 states were active at any time in 2012, meaning they posted at least 1 tweet. This number increased each year, up to 50 in 2018. It decreased to 49 in 2020. The total number of tweets across all accounts increased from 25,276 in 2012 to 51,132 in 2018. The total rose to 94,205 in 2020 and decreased to 62,797 in 2022 (Figure 1). Overall, mean tweets per day rose from 1.7 in 2012 to 2.8 in 2018. It reached a maximum of 5.3 in 2020 and remained above the 2018 level. Figure 2 displays the distribution of mean tweets per day among active states each year. Delaware, North Carolina, Florida, Indiana, Pennsylvania, Massachusetts, and West Virginia each had the most tweets in a given year. Delaware, Florida, and West Virginia each had the most for multiple years in a row. Massachusetts was the most active in a single year, averaging >21 tweets per day in 2020.

XSL•FO
RenderX

Figure 1. Stacked bar chart displaying an exhaustive enumeration of tweets from US state public health agency accounts for each year from 2012 through 2022, divided into 5 categories: retweets, self-replies, other-replies, quote tweets, and standalone tweets.

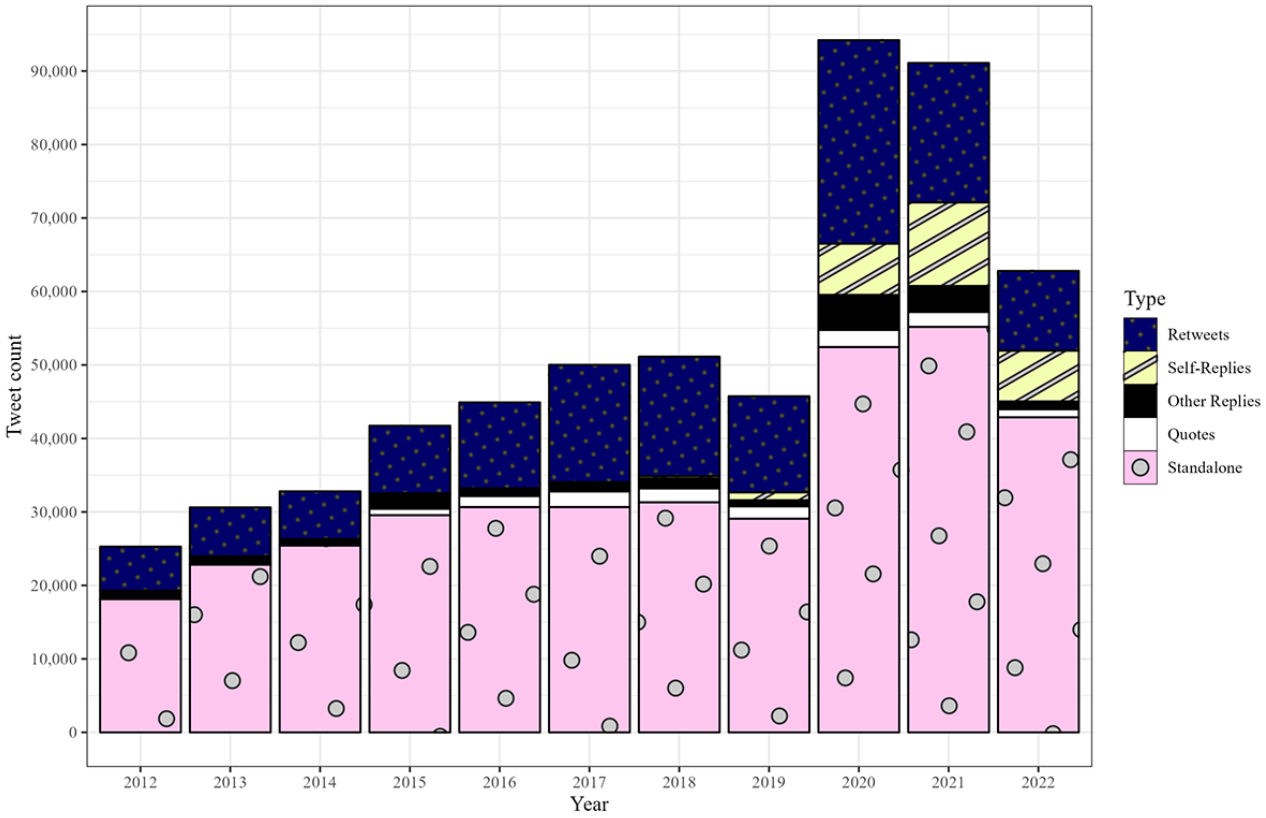

Figure 2. Among an exhaustive enumeration of tweets from US state public health agency accounts, violin plots displaying the distribution of the mean daily tweets per account, each year from 2012 through 2022. IQRs for each year are marked with horizontal lines. The mean of means for each year is marked with a dot.

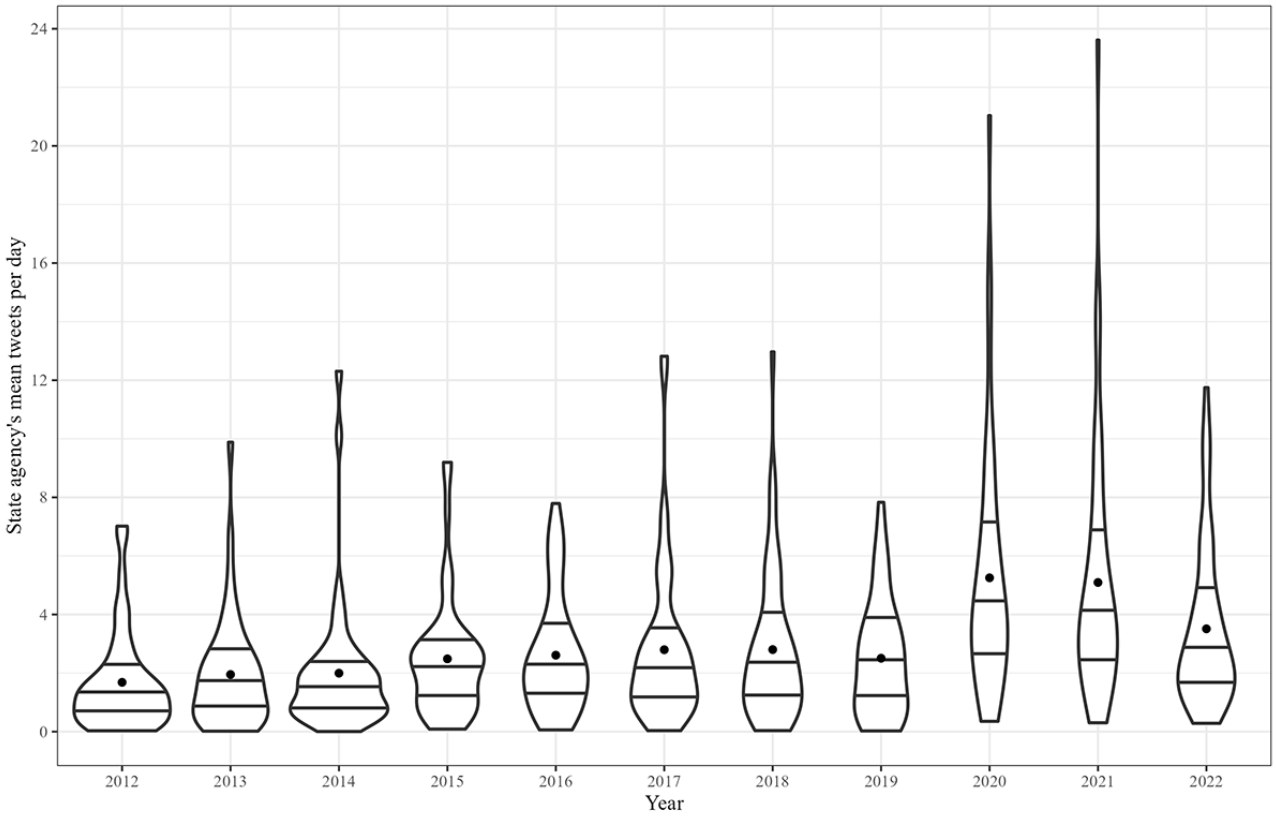

XSL•FO
RenderX

The percentage of non-retweets with 0 on all audience engagement metrics was 62% (12,965/21,023) in 2012, decreasing to 20% (6465/32,668) in 2019. It reached a minimum of 6% (4123/66,507) in 2020 and 12% (6306/51,950) in 2022.

Figure 3 presents each year's IQR for audience engagement on state agencies' non-retweets (ie, replies, quote tweets, and standalone tweets), summarized in the text below.

Figure 3. Among an exhaustive enumeration of non-retweets from US state public health agency accounts, grouped bar charts showing the IQR of audience engagement metrics, for each year from 2012 through 2022. The engagement metrics include quote tweets, replies, retweets, and likes. Note: medians are marked with a thick horizontal line. The bottom edge of each boxplot marks the 25th percentile. The top edge marks the 75th percentile.

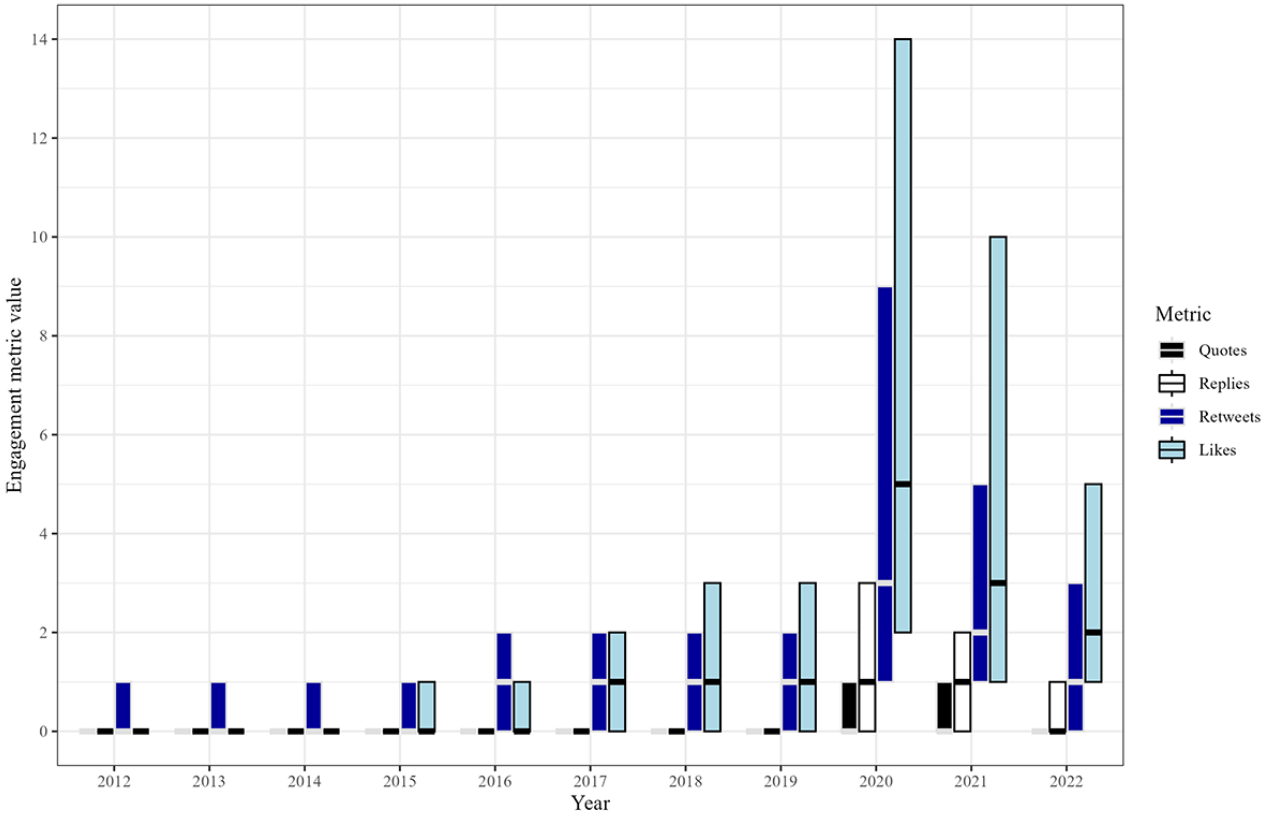

The mean number of likes was <0.6 before increasing to 1.2 in 2016 and then 2.7 in 2019. Its maximum was 17.6 in 2020. It decreased to 5.4 in 2022. The median was 0 before rising to 1 in 2017. Its maximum was 5 in 2020. It decreased to 2 in 2022. The maximum (n=69,798) was on a 2020 tweet about social distancing in Ohio. The second highest (n=12,022) was on a 2020 tweet about face coverings in California.

The mean number of replies was <0.3 before rising to a maximum of 3.5 in 2020. It decreased to 1.8 in 2022. The median was 0 every year except for 2020 and 2021, when it was 1. The maximum (n=2003) was on a 2022 tweet about the COVID-19 pandemic and breastfeeding or chestfeeding from Washington. The second highest (n=1954) was on a 2021 tweet about face coverings in Alabama.

The mean number of retweets was <1 until 2015, rising to 1.8 in 2018. Its maximum was 10.9, in 2020. It decreased to 2.9 in 2022. The median was 0 until rising to 1 in 2016. Its maximum was 3 in 2020. It decreased to 1 in 2022. The maximum (n=39,950) was on the previously mentioned Ohio tweet. The second highest (n=4844) was on a 2020 tweet about COVID-19 cases in Florida.

The mean number of quotes was <0.3 until rising to a maximum of 1.9 in 2020 and decreasing to 0.6 in 2022. The median was 0 for the entire study period. The maximum (n=7041) was on the previously mentioned Ohio tweet. The second highest (n=2552) was on a 2020 tweet about raw meat in Wisconsin.

Persistence

The percentage of non-retweets with at least 1 mention of an account other than the authoring state public health agency decreased from a maximum of 21% (5399/25,267) in 2013 before decreasing to 17% (4437/26,750) in 2014. It then went up and down within that range before dropping off to 13% (9661/72,124) in 2021 and a minimum of 10% (5402/51,950) in 2022.

Among non-retweets, the top-mentioned account each year was @CDCgov. Before 2018, the top 10 accounts typically included at least 2 accounts representing a company or a non-CDC federal government account. In 2018, 9 of the top 10 represented state government agencies or officials, with the top 10 maintaining a similar balance for several years. In 2022, the top 10 included 2 additional CDC accounts and 2 other federal agencies.

The overall percentage of manual retweets decreased from 7% (1729/25,276) in 2012 to 1% (427/32,800) in 2014, then to <0.1% (5/50,024) in 2017 and 0% (0/91,099) in 2021.

XSL•FO
RenderX

Replicability

The percentage of tweets that were formal retweets increased from 17% (4263/25,276) in 2012 to its maximum of 32% (15,955/50,024) in 2017. It decreased to 29% (27,698/94,205) in 2019 and increased slightly in 2020 before decreasing to 17% (10,847/62,797) in 2022.

The percentage of tweets that mentioned a CDC account was 4.2% (1056/25,276) in 2012, increasing to a maximum of 7% (3559/51,132) in 2018 and decreasing to a minimum of 3.5% (2218/62,797) in 2022. The percentage of tweets retweeting a CDC account increased from a minimum of 2.1% (521/25,276) in 2012 to a maximum of 5.4% (2756/51,132) in 2018. It decreased 2.2% (1363/62,797) in 2022.

The percentage of quote tweets increased from 2% (877/41,730) in 2015 to 4% (2129/50,024) in 2017 before decreasing each year to a minimum of 2% (1106/62,797) in 2022. Table 1 displays the retweet and quote tweet data by year.

Table 1. Among an exhaustive enumeration of all tweets from US state public health agency accounts, the portions that were either retweets or quote tweets, as well as the portion mentioning or retweeting a CDC[a] account.

Year	Retweets (%)	Quote tweets (%)	Tweets with ≥1 mention of a CDC account (%)	Retweets of a CDC account (%)
2012	16.9	—[b]	4.2	2.1
2013	17.5	—	4.7	2.4
2014	18.4	—	4.8	2.9
2015	21.8	2.1	4.4	3.2
2016	26.0	3.3	6.0	4.1
2017	31.9	4.3	6.1	4.8
2018	31.8	3.7	7.0	5.4
2019	28.9	3.7	5.4	3.6
2020	29.4	2.5	4.8	3.6
2021	20.1	2.2	3.7	2.6
2022	17.3	1.8	3.5	2.2

[a]CDC: US Centers for Disease Control and Prevention.

[b]Not applicable.

Searchability

The percentage of tweets with at least 1 hashtag increased from a minimum of 36% (9019/25,276) in 2012 to a maximum of 58% (24,017/41,730) in 2015. It decreased to 38% (23,565/62,797) in 2022.

The hashtags #health, #healthde, #ebola, #vibriovulnificus, #zika, #flu, and #covid19 were each the most common hashtags at 1 point during the study period. The #flu and #covid19 hashtags were each the most common for 3 consecutive years. The top 10 hashtags each year included at least 1 reference to natural disasters or disease outbreaks like #sandy or #ebola. From 2012 through 2016, top hashtags included at least 1 reference to a public health event, such as #cdcchat. In 9 years, the top 10 included #netde, a Delaware-specific hashtag.

Nine of the top 10 hashtags in 2020 and 2021 referred to the COVID-19 pandemic. This number decreased to 5 in 2022. In 2020, 5 of the top 10 hashtags were state-specific COVID-19 hashtags like #covid19ma. This decreased to 2 in 2021 and 2022.

The hashtag #flu was in the top 10 every year in the study period. The hashtag #hiv was in the top 10 from 2012 through 2017 and in 2019. Figure 4 presents the percentage of tweets, including #flu or #hiv, each year.

XSL•FO
RenderX

Figure 4. Among an exhaustive enumeration of tweets from US state public health agency accounts, grouped bar charts displaying the portion using the #flu or the #hiv hashtag, for each year from 2012 through 2022.

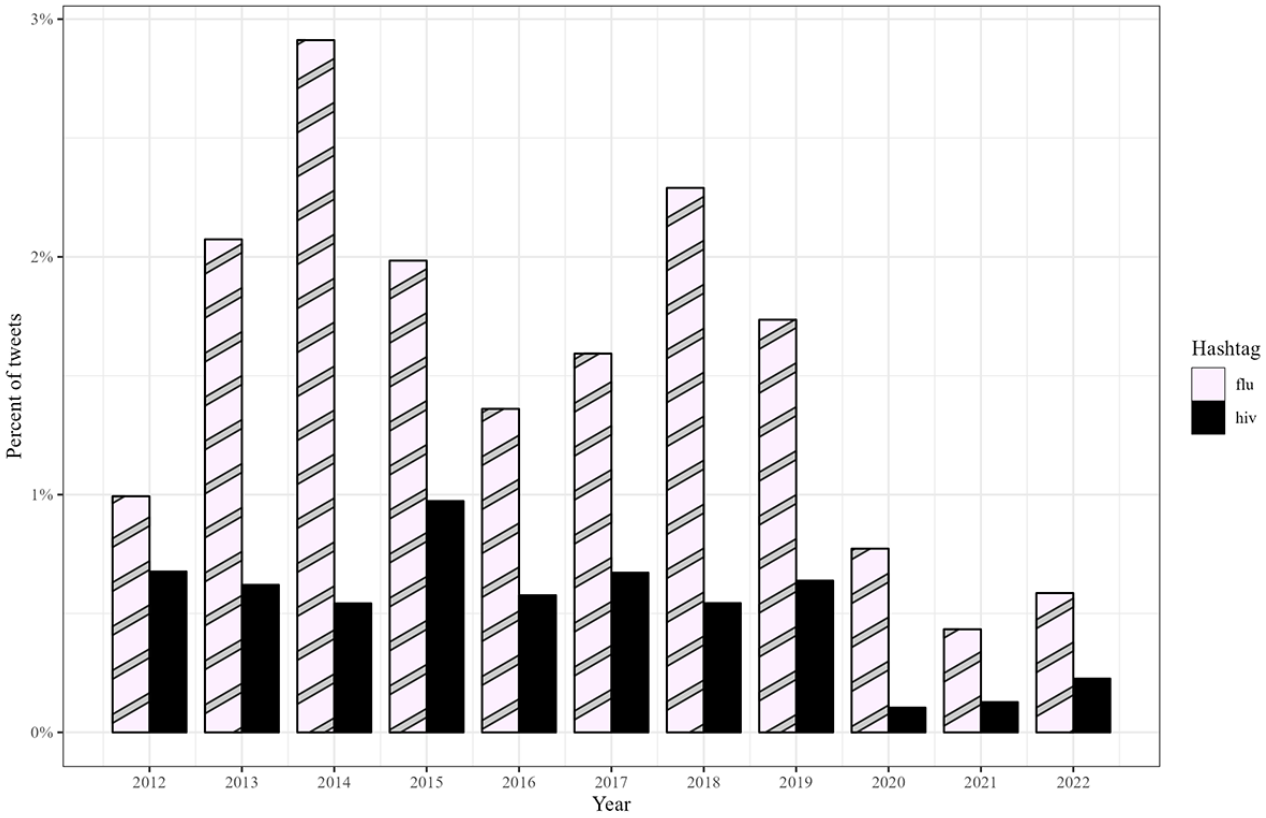

Discussion

Principal Findings

We analyzed state public health agencies' English-language tweets from 2012 through 2022 to provide a complete snapshot of Twitter use in the long term. We used Twitter API data as indicators for agencies' engagement with key qualities of communication in a networked public: scalability, persistence, replicability, and searchability. We found increasingly fragmented communication over time, with less emphasis on replicability and persistence. This involved a drop in public interactions with other accounts in the form of retweets and quote tweets, which coincided with a dramatic increase in overall tweet volume and audience engagement during the COVID-19 pandemic.

Indicators of scalability and searchability on Twitter display a drastic shift at the beginning of the pandemic. We observed a dramatic increase in Twitter outputs and a focus on state-specific searchability amid a decreased focus on searchability overall. Though some states were especially active, the entire interquartile range of states' mean Tweets per day shifted up throughout the pandemic. Audience engagement metrics followed a similar pattern, though still modest by marketing standards. Multiple COVID-19 hashtags were prominent from 2020 through 2022, with state-specific pandemic hashtags especially prominent in 2020. This contrasted with the prepandemic prominence of broader, more singular health emergency hashtags like #measles or #ebola.

Indicators of persistence and replicability did not display the same kind of COVID-19 pandemic–era shift. We observed more fragmented public health communication over time, relying less on replicating messages and traceable interactions with other accounts, on a time scale that did not align with the beginning of the pandemic. Mentions of state government officials became more prominent before the pandemic and remained so during the pandemic. The decrease in mentions of external accounts did not occur at the start of the pandemic. The rate of quote tweets decreased over time, starting before the pandemic. The rate of formal retweets followed a similar pattern but with a larger sudden decrease during the second year of the pandemic. The rate of mentions and retweets of CDC accounts decreased over time, starting before the pandemic.

Finally, an overarching finding in this study was between-state variation that did not fall neatly along the lines of geography, population, or a political binary of red and blue states. Florida, West Virginia, and Delaware were all highly active on Twitter for multiple years. Massachusetts, Wisconsin, and Ohio health agency communication styles included prominent use of state-specific pandemic hashtags. Tweets with record engagement were from Washington, Ohio, California, Florida, Wisconsin, and Alabama.

Limitations

Our findings should be interpreted in light of 4 limitations. First, we collected our data retrospectively, meaning Tweet deletion and changes in privacy settings potentially shaped our findings. By the nature of these underlying nonrandom mechanisms, it is not possible to estimate the magnitude of potential missing

XSL•FO
RenderX

data or how missingness may have changed over the study period. However, we argue this is not a critical issue for state health agencies, whose communications are part of the public record. This could, however, impact our findings around retweets and quote tweets of other accounts. We argue this is not likely a significant limitation, as the scale of "standalone tweets" we observed toward the beginning of our study period was similar to measures of "original tweets" from Thackeray et al [15]. A second limitation is our use of API indicators to filter tweets by language, relying on Twitter's proprietary processing. Third, we focused solely on text features, meaning we missed trends in multimedia content, for example, reuploaded graphics. Finally, we studied only 1 platform. While our numerical results are not generalizable to other platforms, the patterns we identified in terms of our theoretical framework can be directly compared with data from other networked publics.

Comparison With Previous Work

Our findings about the long-term prominence of self-replies and standalone tweets align with findings on the one-way nature of public health communication on Twitter from the previous decade [14,15]. We argue this could actually be a form of effective tailoring, for example, if known journalists draw information from Tweets for dissemination more often than they draw from press releases. However, there is a need for more research to make sense of our findings, such as exploring whether public health tweets during the pandemic have been helpful for these audiences.

Our findings differ from those of a study from Bradford et al [17], whose work covered a study period contained within our own. This appears to be due to their inclusion of only 1 Twitter account from Maine, in contrast with our inclusion of both Maine's Department of Health and Human Services (DHHS) and its Center for Disease Control and Prevention (MeCDC) accounts. We made this decision after noting the MeCDC serves as the primary public-facing communication arm of the state DHHS. This aligns with the finding from Bradford et al [17] that a Maine account was barely active during their study period. Our conclusion about the overall volume of tweets also differs from Bradford et al [17], due purely to a matter of interpretation. Bradford et al [17] interpreted their observations as signs of low Twitter activity despite most accounts tweeting multiple times per day. We interpreted that same volume of tweets as high activity due in part to the sharp contrast with prepandemic levels. This highlights the need for more research on state health agencies' social media strategies to help interpret our findings and those of other quantitative descriptions of public health communication on social media. Despite our differing conclusions, we echo the call of Bradford et al [17] for more research into the determinants of social media use at state health agencies. For example, we do not know to what degree our findings reflect purposeful decisions within state health agencies or the effects of changes in platform ownership, communication budgets, political factors, or platform design changes over time.

Our findings align with research on US state health departments' use of Facebook. Previous work has found the overall volume of Facebook posts by state health departments to vary [24]. Previous work has also identified different message frames between CDC and state and local health department Facebook posts [25], which could relate to our observed trends in decreasing engagement with CDC tweets and Twitter accounts in the latter half of our study period. Though Twitter and Facebook are prominent social media platforms, future research should quantitatively analyze state health departments' use of other platforms as well.

Conclusions

Our findings warrant further research to elucidate the factors driving the variation we saw between states and over time. This includes research into potential platform-specific communication strategies, health communication training, platform policy, and network norms. Such research should include qualitative content analyses of social media content to elucidate the actual use of social media in public health agencies, not just intended strategies. Further quantitative research should examine the impact of political factors on public health communication on social media. This includes state and federal political factors like budget changes and political divisions around the practice of public health. Further research is also warranted to better understand various audiences' interactions with health agencies in networked publics, as well as their perceptions of state-level health communication.

Our findings raise further questions about what state health agencies' communication should look like and how to support it. For example, it might be impractical for state agencies to target communities that have more contact with local health departments, even when there may be more extensive communication resources at the state level. There is not a robust national infrastructure to fill this gap, much less monitor and evaluate health communication on social media during a public health emergency. This gap is a critical contextual factor to consider when interpreting our findings around between-state variation.

Our study exemplifies the use of theoretical frameworks from media studies to understand trends in public health communication on social media. Though we examined metrics from one social media platform, our use of the networked public framework allows us to draw broader inferences about health communication. This increases the relevance of our findings despite differing features across time and across platforms. This is in contrast to the existing descriptive work of US state health agencies on Twitter. Integration of a theoretical framework should be a key part of future research on public health communication on social media, so as to extend its use beyond the limited scope of one platform's features.

Finally, our study highlights the vulnerability of public health communication in networked publics. A social media platform's use is subject to corporate decisions and community norms. If a platform shuts down or its communities deliberately hide health information, then public health investment in that platform may be suddenly irrelevant. For example, it is unclear whether the drop-off in tweets we observed in 2022 was the result of strategic decisions, insufficient resources, or platform changes. Even the ability to efficiently study public health communication largely depends on corporate policies around

XSL•FO
RenderX

data access. Our study highlights the importance of supporting health communication research. This includes regulation to ensure academic access to social media data, as well as building a robust communication infrastructure outside of private platforms.

Acknowledgments

This study was funded by the National Cancer Institute (grant 3U54CA156732-13S1). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the funder. Generative AI was not used in any portion of this study or in writing this manuscript.

Data Availability

The data analyzed during this study are available in a Harvard Dataverse repository [20]. The R code used to analyze the data and calculate descriptive statistics is available in an Open Science Framework repository [21].

Authors' Contributions

SRM contributed to conceptualization, methodology, software, formal analysis, data curation, writing (original draft), and visualization. SMNG managed conceptualization, methodology, supervision, and writing (review and editing). KME handled conceptualization, methodology, supervision, and writing (review and editing). KV contributed to conceptualization, methodology, supervision, and writing (review and editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.
[DOCX File , 32 KB - jmir_v27i1e59786_app1.docx]

Multimedia Appendix 2
Usernames of Twitter accounts included in this study.
[XLS File (Microsoft Excel File), 49 KB - jmir_v27i1e59786_app2.xls]

Multimedia Appendix 3
Exhaustive summary data table.
[XLS File (Microsoft Excel File), 54 KB - jmir_v27i1e59786_app3.xls]

References

1. Boyd D. Social network sites as networked publics: Affordances, dynamics, and implications. In: Papacharissi Z, editor. A Networked Self: Identity, Community, and Culture on Social Network Sites. New York, NY: Routledge; 2011:47-66.
2. Gibson CM. The democratization of medical research and education through social media: The potential and the peril. JAMA Cardiol 2017;2(1):9-11. [doi: 10.1001/jamacardio.2016.4933] [Medline: 27973665]
3. Costanza-Chock S. Design narratives: From TXTMob to Twitter. In: Design Justice: Community-Led Practices to Build the Worlds We Need. Cambridge, MA: The MIT Press; 2020:103-134.
4. Baym NK, Burgess J. Twitter: A Biography. New York, NY: New York University Press; 2020:59-80.
5. Hswen Y, Xu X, Hing A, Hawkins JB, Brownstein JS, Gee GC. Association of "#covid19" versus "#chinesevirus" with anti-Asian sentiments on Twitter: March 9-23, 2020. Am J Public Health 2021;111(5):956-964. [doi: 10.2105/AJPH.2021.306154] [Medline: 33734838]
6. Hoffman BL, Colditz JB, Shensa A, Wolynn R, Taneja SB, Felter EM, et al. #DoctorsSpeakUp: Lessons learned from a pro-vaccine Twitter event. Vaccine 2021;39(19):2684-2691 [FREE Full text] [doi: 10.1016/j.vaccine.2021.03.061] [Medline: 33863574]
7. Harris JK, Choucair B, Maier RC, Jolani N, Bernhardt JM. Are public health organizations tweeting to the choir? Understanding local health department Twitter followership. J Med Internet Res 2014;16(2):e31 [FREE Full text] [doi: 10.2196/jmir.2972] [Medline: 24571914]
8. Crisis and emergency risk communication: Working with the media. Centers for Disease Control and Prevention. 2014. URL: https://emergency.cdc.gov/cerc/ppt/CERC_Working_with_the_Media.pdf [accessed 2024-01-04]

XSL•FO
RenderX

9. DePaula N, Hagen L, Roytman S, Alnahass D. Platform effects on public health communication: A comparative and national study of message design and audience engagement across Twitter and Facebook. JMIR Infodemiology 2022;2(2):e40198 [FREE Full text] [doi: 10.2196/40198] [Medline: 36575712]
10. Zhu JM, Sarker A, Gollust S, Merchant R, Grande D. Characteristics of Twitter use by state Medicaid programs in the United States: Machine learning approach. J Med Internet Res 2020;22(8):e18401 [FREE Full text] [doi: 10.2196/18401] [Medline: 32804085]
11. Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N. Examining tweet content and engagement of Canadian public health agencies and decision makers during COVID-19: Mixed methods analysis. J Med Internet Res 2021;23(3):e24883 [FREE Full text] [doi: 10.2196/24883] [Medline: 33651705]
12. Li L, Aldosery A, Vitiugin F, Nathan N, Novillo-Ortiz D, Castillo C, et al. The response of governments and public health agencies to COVID-19 pandemics on social media: A multi-country analysis of Twitter discourse. Front Public Health 2021;9:716333 [FREE Full text] [doi: 10.3389/fpubh.2021.716333] [Medline: 34650948]
13. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J, Centers for Disease Control and Prevention. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep 2014;63(32):681-685 [FREE Full text] [Medline: 25121710]
14. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]
15. Thackeray R, Neiger BL, Burton SH, Thackeray CR. Analysis of the purpose of state health departments' tweets: Information sharing, engagement, and action. J Med Internet Res 2013;15(11):e255 [FREE Full text] [doi: 10.2196/jmir.3002] [Medline: 24217361]
16. Harris JK, Mueller NL, Snider D. Social media adoption in local health departments nationwide. Am J Public Health 2013;103(9):1700-1707. [doi: 10.2105/AJPH.2012.301166] [Medline: 23865660]
17. Bradford NJ, Amani B, Walker VP, Sharif MZ, Ford CL. Barely tweeting and rarely about racism: Assessing US state health department Twitter use during the COVID-19 vaccine rollout. Ethn Dis 2022;32(3):257-264 [FREE Full text] [doi: 10.18865/ed.32.3.257] [Medline: 35909637]
18. Stake RE. The art of case study research. In: Research Questions. Thousand Oaks, CA: SAGE Publications, Inc; 1995:3-4.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
20. Mendez SR. State public health agency, CDC, and FDA Tweets (2012 through 2022), V1. Harvard Dataverse. 2022. URL: https://doi.org/10.7910/DVN/VX4HK8 [accessed 2024-12-03]
21. Mendez S, Munoz-Najar S, Emmons KM, Viswanath K. Description of US state public health agencies, 2012 through 2022. Open Science Foundation. 2023. URL: https://osf.io/3hec6/ [accessed 2024-12-03]
22. Shu C. Twitter officially launches its “Retweet With Comment” feature. TechCrunch. 2015. URL: https://web.archive.org/web/20150408010147/https://techcrunch.com/2015/04/06/retweetception/ [accessed 2024-12-04]
23. CDC’s Social media channels. Centers for Disease Control and Prevention, Office of Communications. 2023. URL: http://web.archive.org/web/20231109023623/https://www.cdc.gov/digital-social-media-tools/Social-Media-Channels.html [accessed 2023-11-09]
24. Jha A, Lin L, Savoia E. The use of social media by state health departments in the US: Analyzing health communication through Facebook. J Community Health 2016;41(1):174-179. [doi: 10.1007/s10900-015-0083-4] [Medline: 26318742]
25. Meadows CZ, Meadows CW, Tang L. The CDC and state health department Facebook messages: An examination of frames and the extended parallel processing model. Communication Studies 2020;71(5):740-752. [doi: 10.1080/10510974.2020.1819839]

Abbreviations

API: application programming interface
CDC: US Centers for Disease Control and Prevention
DHHS: Department of Health and Human Services
MeCDC: Maine’s Center for Disease Control and Prevention
NIOSH: National Institute for Occupational Safety and Health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 22.04.24; peer-reviewed by D Boatman, N Hu; comments to author 17.09.24; revised version received 10.10.24; accepted 05.11.24; published 03.01.25.

Please cite as:
Mendez SR, Munoz-Najar S, Emmons KM, Viswanath K
US State Public Health Agencies' Use of Twitter From 2012 to 2022: Observational Study
J Med Internet Res 2025;27:e59786
URL: https://www.jmir.org/2025/1/e59786
doi:10.2196/59786
PMID:39752190

©Samuel R Mendez, Sebastian Munoz-Najar, Karen M Emmons, Kasisomayajula Viswanath. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Applying Natural Language Processing Techniques to Map Trends in Insomnia Treatment Terms on the r/Insomnia Subreddit: Infodemiology Study

Jack A Cummins[1]; Daniel J Gottlieb[2,3,4], MPH, MD; Tamar Sofer[2,3,5,6], PhD; Danielle A Wallace[3,7], MPH, PhD

[1]Manchester Essex Regional High School, Manchester, MA, United States

[2]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, United States

[3]Division of Sleep Medicine, Harvard Medical School, Boston, MA, United States

[4]VA Boston Healthcare System, Boston, MA, United States

[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[6]CardioVascular Institute (CVI), Beth Israel Deaconess Medical Center, Boston, MA, United States

[7]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Danielle A Wallace, MPH, PhD
Division of Sleep and Circadian Disorders
Department of Medicine
Brigham and Women's Hospital
221 Longwood Avenue
Boston, MA, 02115
United States
Phone: 1 617 732 5987
Email: dwallace5@bwh.harvard.edu

Abstract

Background: People share health-related experiences and treatments, such as for insomnia, in digital communities. Natural language processing tools can be leveraged to understand the terms used in digital spaces to discuss insomnia and insomnia treatments.

Objective: The aim of this study is to summarize and chart trends of insomnia treatment terms on a digital insomnia message board.

Methods: We performed a natural language processing analysis of the r/insomnia subreddit. Using Pushshift, we obtained all r/insomnia subreddit comments from 2008 to 2022. A bag of words model was used to identify the top 1000 most frequently used terms, which were manually reduced to 35 terms related to treatment and medication use. Regular expression analysis was used to identify and count comments containing specific words, followed by sentiment analysis to estimate the tonality (positive or negative) of comments. Data from 2013 to 2022 were visually examined for trends.

Results: There were 340,130 comments on r/insomnia from 2008, the beginning of the subreddit, to 2022. Of the 35 top treatment and medication terms that were identified, melatonin, cognitive behavioral therapy for insomnia (CBT-I), and Ambien were the most frequently used (n=15,005, n=13,461, and n=11,256 comments, respectively). When the frequency of individual terms was compared over time, terms related to CBT-I increased over time (doubling from approximately 2% in 2013-2014 to a peak of over 5% of comments in 2018); in contrast, terms related to nonprescription over-the-counter (OTC) sleep aids (such as Benadryl or melatonin) decreased over time. CBT-I–related terms also had the highest positive sentiment and showed a spike in frequency in 2017. Terms with the most positive sentiment included "hygiene" (median sentiment 0.47, IQR 0.31-0.88), "valerian" (median sentiment 0.47, IQR 0-0.85), and "CBT" (median sentiment 0.42, IQR 0.14-0.82).

Conclusions: The Reddit r/insomnia discussion board provides an alternative way to capture trends in both prescription and nonprescription sleep aids among people experiencing sleeplessness and using social media. This analysis suggests that language related to CBT-I (with a spike in 2017, perhaps following the 2016 recommendations by the American College of Physicians for CBT-I as a treatment for insomnia), benzodiazepines, trazodone, and antidepressant medication use has increased from 2013 to 2022. The findings also suggest that the use of OTC or other alternative therapies, such as melatonin and cannabis, among r/insomnia Reddit contributors is common and has also exhibited fluctuations over time. Future studies could consider incorporating

XSL•FO
RenderX

alternative data sources in addition to prescription medication to track trends in prescription and nonprescription sleep aid use. Additionally, future prospective studies of insomnia should consider collecting data on the use of OTC or other alternative therapies, such as cannabis. More broadly, digital communities such as r/insomnia may be useful in understanding how social and societal factors influence sleep health.

(J Med Internet Res 2025;27:e58902) doi:10.2196/58902

KEYWORDS

insomnia; natural language processing; NLP; social media; cognitive behavioral therapy; CBT; sleep initiation; sleep disorder; easly awakening; sleep aids; benzodiazepines; trazodone; antidepressants; melatonin; treatment

Introduction

Insomnia is a condition characterized by self-reported difficulty in falling asleep or staying asleep, with accompanying distress or impairment [1]. Insomnia symptoms can be chronic or acute [2]. Approximately 10%-30% of the population is estimated to experience either chronic or acute insomnia, with insomnia more common among women and older adults [3,4]. One prospective cohort study reported an insomnia incidence of 13.9%—of those, 37.5% developed chronic insomnia over a 5-year period [5].

People experiencing insomnia may report decreased quality of life and daily functioning [6,7], have higher injury risk [8], and may seek multiple treatment options to improve their sleep and address insomnia symptoms [9]. Insomnia treatment approaches have also changed over time. Hypnotics used to be the primary treatment option for insomnia until behavioral treatment options such as cognitive behavioral therapy for insomnia (CBT-I) were shown to be highly clinically effective [10]. Alternative herbal remedies for insomnia, such as herbal teas to induce somnolence, have also been popular throughout history [11]. Trends in social mention of insomnia treatment options may track with changes in approaches for treatment in the medical community, but they may also reflect additional community-based approaches due to limitations in access to treatment or cultural preferences. These community-based approaches, such as using herbal teas and supplements, may not be reflected in medical records or drug prescription rates.

The language used by individuals seeking and obtaining insomnia treatment may be heterogeneous and may differ from standard medical terminology for the treatment provided [10]. For example, one of the components of CBT-I consists of sleep hygiene, which addresses routines and environmental factors that influence sleep initiation and maintenance [12]; however, the term “hygiene” may connote a negative association, implying the person seeking help has “dirty” sleep habits [13]. Understanding the language that people use in digital insomnia-related communities and the sentiment attached to treatment terms may be useful for health practitioners as they work to understand how to discuss symptoms and treatment options with their patients, as well as for sleep researchers engaged in data collection. Mapping trends in language around insomnia may also provide insights helping to interpret the use of related terms over time and may identify temporal patterns in the use of medication and other therapies for insomnia.

Digital health-related message boards can be important sources of frank discussion and community and provide insight into beliefs, practices, and language. Reddit is one such digital community with medically related subreddits; prior analyses of these subreddits have used well-established, reliable, natural language processing (NLP) techniques such as bag of words (BOW), regular expressions (RE), and Valence Aware Dictionary and Sentiment Reasoner (VADER) to better understand patient or community experiences [14-16]. For example, Low et al [17] analyzed data from Reddit mental health communities using various NLP techniques to identify changes in various mental health–related subreddits during the COVID-19 pandemic. However, these techniques have not been previously applied to the r/insomnia subreddit. The r/insomnia subreddit may be a particularly rich source to examine insomnia using NLP techniques because sleeplessness is commonly self-treated outside of a medical setting. In a 2022 American Academy of Sleep Medicine (AASM) survey (n=2010 US adults), 64% of respondents reported using a substance to help them sleep, with 23% reporting the use of prescription medications and 41% reporting the use of an over-the-counter (OTC), herbal, or other substance; the use of a sleep aid was most prevalent among respondents aged 18-54 years [18]. Comments on r/insomnia, an anonymous forum, may be a resource for understanding self-treatment behavior and the use of prescription or nonprescription substances. We characterize the language used in a digital discussion board of insomnia using an NLP analysis of the r/insomnia subreddit to measure the discussion frequency of insomnia-related substances and to identify the sentiment (positive or negative) when substances are mentioned.

Methods

Overview

Our goals were to measure the discussion frequency of insomnia-related substances and to identify the sentiment. Thus, we applied a set of NLP methods that (1) extract common words from text, (2) identify substance-related treatment terms (relying on human recognition of some terms), and (3) assign sentiment to sentences containing such terms. First, the r/insomnia data were collected from the Pushshift dataset. Next, comments were converted to word counts using a BOW model. Using this BOW model, the most common nonstop words were identified, and the insomnia treatment–related terms were selected. The BOW approach was chosen because (1) it is a simple and efficient method and (2) summarizing and quantifying the popularity or frequency of treatment terms was a primary goal; in this way,

XSL•FO
RenderX

disadvantages of BOW, such as loss of word order, were not deemed to be problematic for our purposes. REs were then applied to quantify the use of these terms in the subreddit over time. The RE approach was chosen because (1) the percentage of comments containing a term is an effective way to gauge the popularity or interest in a term, and (2) it is able to find all comments that contain a pattern of letters rather than just individual words (like the BOW method). For example, the RE pattern "sleep" would also capture related words with the same root, such as "asleep," "sleep," "sleeping," and "sleeplessness." Finally, sentiment analysis was applied on the raw comments (not word counts; Figure S1 in Multimedia Appendix 1). Sentiment analysis was chosen because, in addition to measuring interest in these terms and treatments, the goal of this study was to gauge how social media users felt about the terms and treatments. In general, the average sentiment of comments containing the treatment term is expected to reflect the sentiment toward that treatment.

Data Source

All subreddit comments from the history of r/insomnia [19] in 2008 up to the end of 2022 were obtained from the top 20,000 subreddits Pushshift dataset, which is a publicly available dataset [20,21]. All comments from the r/insomnia subreddit from 2008 to 2022 were included in each data analysis step. Data analysis was performed with Python (version 3; Python Software Foundation), and statistical testing was performed with R (version 4.4.0; R Foundation for Statistical Computing). Analyses that were summarized by year used comments dated up to December 31, 2022. All methods of analysis detailed below refer to individual comments (responses, not including the original post). If a thread of conversation included multiple comment responses from different commenters or the same commenter, then each comment within the thread would be counted as a separate comment in the total. This study did not meet the criteria for human participant research as defined by Mass General Brigham Human Research Office policies and Health and Human Services regulations set forth in 45 CFR 46. The data were obtained from a publicly available anonymous digital forum.

Identifying Treatment-Related Words Using a BOW Model

A BOW model, which extracts text features without regard to the order, was created using the *gensim* package [22]. A BOW model counts the number of occurrences of each word in a comment (excluding "stop words," such as "the" and "a"). To create the BOW model, we first preprocessed all comments from the r/insomnia subreddit using the gensim simple preprocessor function, which removed symbols and numbers and converted the comments into lists of words. The natural language toolkit "wordNetLemmatizer" was used to convert words to their dictionary forms. For example, the word "dogs" is simplified to "dog," but the word "benzos" would not be shortened to "benzo" because it is an abbreviation and is not included in the dictionary. Stop words were removed from comments. Words used in fewer than 15 comments were removed. After the BOW model was created, the 1000 words with the highest comment frequency (number of comments that contain the word at least once) were retrieved. The list of 1000 words for the r/insomnia subreddit was reviewed manually by one of the authors (DAW) and reduced to words based on relevance to insomnia medication or insomnia-related treatment terms. The words are referred to henceforth as the "BOW list." To identify additional brand or generic drug names for the treatment terms that referred to a medication, we also performed an extensive search using the Epocrates drug reference platform in addition to Google and Wikipedia. Terms that referred to the same treatment, such as generic versus brand names of medications (eg, quetiapine and Seroquel), were combined in deriving term counts (Figure S1 in Multimedia Appendix 1).

Using REs to Identify Textual Patterns in the Data

In addition to BOW, we used RE to identify patterns in the text by identifying and counting comments that contained each of the selected words of interest from the BOW list within r/insomnia. The RE searches were set as case insensitive. We used Matplotlib [23] to visualize the percentage of r/insomnia comments containing each of the prespecified words each year. The percentage of comments containing a word each year was calculated as the total number of comments containing the specified word divided by the total number of comments in the subreddit extracted from that year, not accounting for authorship. To standardize the data and account for changes in total r/insomnia comment volume over time, all data are reported as the percentage of total annual comments that contain the specified term (ratio of comments containing the term relative to the total annual comment volume). The analysis focused on the years 2013 to 2022 because comment volume was low prior to 2013. For some related words, search terms were combined so that comments containing any of multiple words (eg, "CBT," "CBTI," or "hygiene") were counted together in the analysis.

Sentiment Analysis of Insomnia Treatments

Sentiment analysis was also applied to identify the overall emotional tone of the text containing the terms of interest. Sentiment analysis was performed using the VADER [20], a reliable and commonly used lexicon and rule-based tool specifically trained on and developed for use with social media data [22-25]; VADER has been applied in numerous prior health-related analyses of Reddit data [26-28]. VADER calculates the polarity of text on a normalized scale from –1 to 1, 1 being the most positive sentiment for a comment and –1 being the most negative sentiment for a comment. The resulting sentiment values from comments containing insomnia-related treatment terms were compared. To test whether sentiment by treatment term category significantly differed from 0, comments with treatment terms were subset to randomly select 1 comment per author (for independence; Figure 1), and sentiment tested to be different from 0 using 1-sample Wilcoxon signed rank tests (2-sided). Bonferroni correction was used to correct for multiple testing, and results (P<.003) were considered statistically significant.

XSL•FO
RenderX

Figure 1. Boxplot of sentiment of the treatment and medication terms identified in r/insomnia comments after randomly selecting 1 comment per author. The blue boxes depict from the first quartile to the third quartile, with the median marked by an orange line. The black lines extend to minimum and maximum values. Asterisks mark terms with sentiment significantly different from 0 in 1-sample Wilcoxon signed-rank test after Bonferroni correction for multiple testing (P<.003). CBD: cannabidiol; CBT: cognitive behavioral therapy; RE: regular expression.

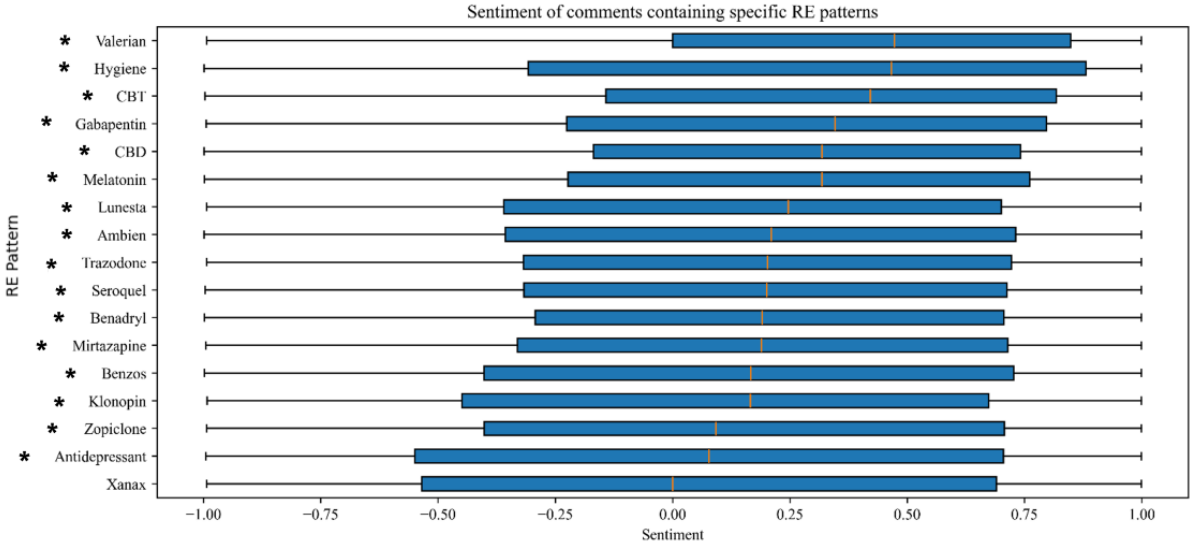

Ethical Considerations

Ethical concerns exist with the use of internet and social media data and should be considered prior to and throughout the data analysis process. Although our project received a research exemption, we believe it is important to mention some of the ethical issues around using Reddit data. The data we used in this analysis are publicly available. However, social media can also potentially contain identifying information if names or other identifiers are contained within the data. Reddit is considered an anonymous site where users create their own usernames or pseudonyms [29] compared to other social media sites, where people may use their real names; we also do not mention usernames or provide quotes from comments to protect the anonymity of commenters [30] and present aggregated results. Additionally, the r/insomnia subreddit has a sizeable community base compared to smaller subreddits, where commenters may know each other [30]. Given these considerations, we believe that our analysis aligns with current ethical guidelines around the use of internet data [31].

Results

There were 340,130 comments from the r/insomnia subreddit. Only data from the years 2013 to 2022 were included in the trend analyses due to sparse data (less than 2500 comments per year) through 2012.

Insomnia Subreddit Growth Over Time

The number of comments per year in the r/insomnia subreddit grew steadily from 306 comments in 2011 to 86,623 comments in 2022 (Table S1 and Figure S2 in Multimedia Appendix 1). The number of unique commenters also grew from 155 commenters in 2011 to 14,866 commenters in 2022 (Table S1 in Multimedia Appendix 1). The number of unique authors commenting on the subreddit increased every year except for 2021, when there was a 2.5% decrease in unique authors. The number of comments also increased every year. The subreddit grew the greatest in 2019, with 20,942 (79.5%) more comments and 4319 (88.5%) more unique commenters compared to 2018. The rate of increase in number of comments and commenters began to plateau from 2020 to 2022.

Patterns in Treatment and Medication Term Use

Unsurprisingly, the most common term in the r/insomnia subreddit was “sleep.” Of the top BOW 1000 terms, we selected 35 terms related to treatment and medication use (Table 1 and Tables S2 and S3 in Multimedia Appendix 1). Melatonin was the most common single treatment term, with 15,005 mentions, followed by terms related to CBT-I, with 13,461 mentions. When time trends were evaluated with RE, terms related to CBT-I spiked in 2017-2018, with a slight decrease and plateau from 2020 onward (Figure 2); the term “hygiene” alone, however, had less variable patterns (Figure S3 in Multimedia Appendix 1). The combined frequency of terms related to antidepressant medications (Table S3 in Multimedia Appendix 1) showed an increase in 2016-2017 and a gradual decline and plateau until a subsequent rise starting in 2020. Terms related to benzodiazepines overall showed a general increase in frequency over time (Figure 2), with a gradual increase from 2013 to 2016, a gradual decrease from 2016 to 2019, and a subsequent increase from 2020 to 2022.

XSL•FO
RenderX

Table 1. List of chosen terms related to treatment and medication use, in addition to the number of comments containing at least 1 occurrence of the specific term and the number of comments containing at least 1 occurrence of a combined set of similar terms (data from 2008 to 2022)[a].

Term	Comments, n
Melatonin	15,005
CBT[b,c]	9018
CBT-I[d]	1512
Hygiene	3846
Ambien	9971
Trazodone[e]	5753
Trazadone	3582
Benzos[f]	3987
Benzo	2289
Benzodiazepine	1158
Weed[g]	3726
Marijuana	1241
Cannabis	1207
CBD[h]	4545
THC[i]	1635
Seroquel	4968
Hygiene	3846
Benadryl	3715
Magnesium	3383
Xanax	3151
Lunesta	2997
Mirtazapine	2902
Antidepressant	2811
SSRI[j]	1469
Zopiclone	2537
Gabapentin[k]	1173
GABA[l]	1133
Valerian	2066
Antihistamine	1638
Klonopin	1514
Zolpidem	1389
Diphenhydramine	1341
Hydroxyzine	1165
Remeron	1047
Ativan	1032
Unisom	812
Quetiapine	940

[a]Combined number of comments are provided over groups of terms that are likely to refer to the same treatment. The number of comments for each unique term may not add up to the combined number of comments because some comments contain multiple terms in a group.

[b]The combined number of comments for the terms CBT, CBT-I, and hygiene together is 13,461.

XSL•FO
RenderX

[c]CBT: cognitive behavioral therapy.

[d]CBT-I: cognitive behavioral therapy for insomnia.

[e]The combined number of comments for the terms trazodone and trazadone together is 9277.

[f]The combined number of comments for the terms benzos, benzo, and benzodiazepine together is 6816.

[g]The combined number of comments for the terms weed, marijuana, cannabis, CBD, and THC together is 10,232.

[h]CBD: cannabidiol.

[i]THC: tetrahydrocannabinol.

[j]SSRI: selective serotonin reuptake inhibitor.

[k]The combined number of comments for the terms gabapentin and GABA together is 2230.

[l]GABA: γ-aminobutyric acid.

Figure 2. Trends in select treatment and medication terms (percent of comments each year containing specific terms) in the r/insomnia subreddit from 2013 to 2022. As indicated in titles, equivalent or related terms were combined; the numerical percent of comments counts all comments containing one or more of the listed terms. CBT: cognitive behavioral therapy; CBD: cannabidiol; SSRI: selective serotonin reuptake inhibitor; THC: tetrahydrocannabinol.

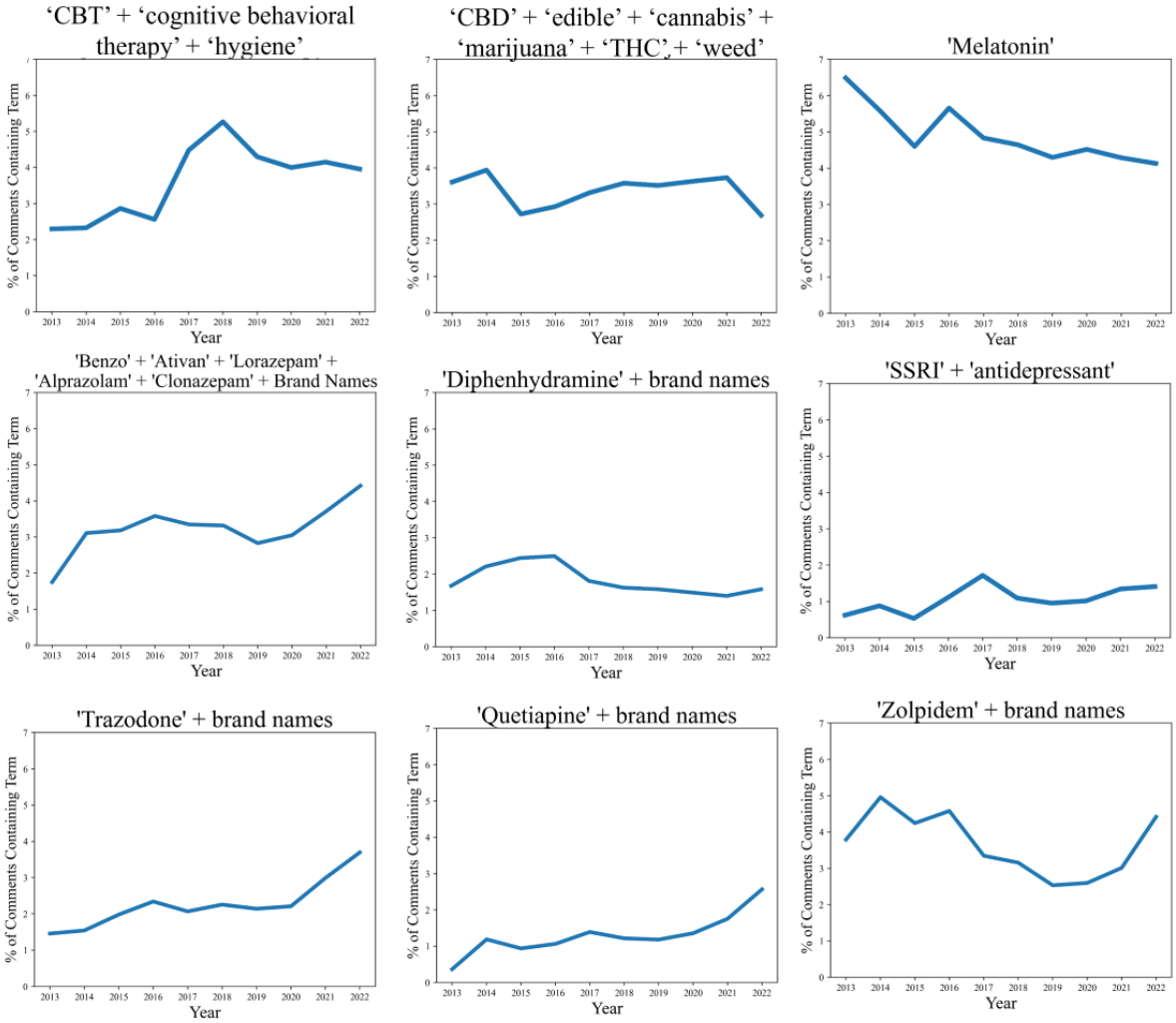

Results of the Treatment and Medication Term Sentiment Analysis

To better understand the emotionality associated with the discussion of treatment within the discussion board, we also conducted a sentiment analysis of the BOW medication and treatment terms. All medication and treatment terms were significantly different from 0 (P<.003), except for terms related to "Xanax." The terms with the most positive sentiment included "hygiene," "valerian," "CBT" (referencing CBT-I), "melatonin," and "CBD" (referencing cannabidiol; Figure 1). The terms with the lowest sentiment (although still positive) included "antidepressant," "zopiclone," "Seroquel," and "Klonopin."

Discussion

Principal Findings

In this analysis of a public internet discussion forum for insomnia, we identified patterns in the frequency of treatment and medication terms over time. Mention of CBT-I,

XSL•FO
RenderX

benzodiazepines, trazodone, and other antidepressants increased over time, while nonprescription terms showed varying fluctuations. Our results both align with prior studies and present new findings, particularly for treatments that are not well captured in medical records. Additionally, population-level patterns in sleep health and insomnia may also reflect larger societal changes, given the importance of sociocultural context in sleep behavior [24] as well as the popularity and availability of sleep treatments.

We applied 3 NLP methods. BOW was necessary to identify frequently used terms within the subreddit because it tokenizes each comment and counts words that are used in the greatest number of comments. After BOW was used to identify the most used terms and treatments, RE was a natural choice to determine the number of comments containing the term because RE can identify terms by patterns of letters. As a result, RE can identify not only words that exactly match but also all words that have the same root (including other tenses, plural, closely related words, and some misspelled versions of the word used by commenters). Finally, VADER, a widely used sentiment analysis tool that is specifically trained for social media data, was used to determine in quantifiable terms the degree to which comments containing a term demonstrate positive or negative sentiment.

Individual treatment terms related to melatonin and CBT-I were the most common across time. They were also associated with high positive sentiment. While melatonin frequency showed a slight decline from 2014 to 2022 in the RE analysis, the frequency of terms related to CBT-I spiked after 2016. This rapid increase in CBT-I may be linked to the guidelines published by the American College of Physicians in May 2016, which recommended CBT-I as a frontline treatment for insomnia [12]. However, starting in 2018, there has been a decrease and subsequent plateau in the frequency of CBT-I terms. There is a documented lack of providers trained in behavioral sleep medicine, particularly in CBT-I, where there are not enough CBT-I practitioners to meet patient demand [10,25]. While the reasons for the slight decrease in CBT-I frequency in 2018-2019 are unclear, the plateau from 2020 onward may be due to more limited ability to seek provider care or decreased availability of providers during the COVID-19–related disruptions in the health care system.

The overall increase in the mention of benzodiazepines may indicate a resurgence in benzodiazepine use. Benzodiazepines are a medication of concern because they can become habit-forming. Prior research examining sedative-hypnotic medication prescriptions in the US National Ambulatory Medical Care Survey reported an overall decrease in benzodiazepine prescriptions and an increase in nonbenzodiazepine receptor agonists from 1993 to 2010 [26]. A subsequent analysis of US data from the 1999-2014 National Health and Nutrition Examination Survey supported an increase in benzodiazepine use, which appeared to be driven by medium- and long-term use of the medications [27]. A more recent analysis of benzodiazepine prescriptions from 2018 to 2021 in the United States suggests an uptick during the 2020-2021 COVID-19 pandemic, particularly among women [28], aligning with our findings.

The frequency of terms related to trazodone, an antidepressant with sedative effects that is sometimes prescribed off-label for insomnia, increased steadily until 2016, followed by a plateau from 2017 to 2020, and then rose from 2021 to 2022. A prior analysis of US prescription data from 2011 to 2018 also supports a gradual increase in low-dose (<150 mg) trazodone prescriptions and a concomitant decrease in prescriptions for zolpidem (similar to our findings) [32]. The frequency of terms related to selective serotonin reuptake inhibitors and antidepressants in general differed, with an incline in 2016-2017, followed by a gradual decline and plateau and a slight uptick through 2021-2022. Interestingly, the drop-off in antidepressant term frequency coincides with the peak CBT-I term occurrence. These patterns align with and may be related to the updated clinical guidelines for insomnia treatments published in 2017 by the AASM. This report described an increase in physicians prescribing antidepressants with sedative properties, such as trazodone, as an alternative to benzodiazepines [33]. However, these guidelines also recommended against trazodone and other medications such as tiagabine, diphenhydramine (an antihistamine included in Benadryl), melatonin, and valerian for the treatment of sleep onset or sleep maintenance insomnia [33]. While these guidelines did suggest doxepin, another antidepressant, doxepin was not a common term in the r/insomnia message board. Despite these guidelines and off-label use, trazodone is one of the most commonly prescribed medications for the treatment of insomnia [34].

There were also trends indicating an overall decreased frequency of terms related to OTC medications (such as Benadryl) and marijuana or cannabis-related (non-cannabidiol) terms on r/insomnia. There were opposing trends between cannabidiol (CBD) and cannabis-related non-CBD terms, where CBD frequency had a sharp transient rise after 2016, and cannabis-related non-CBD terms gradually decreased from 2014 onward. This rise in CBD frequency may be due to the relaxation of CBD regulation, a growing interest in the use of cannabinoids for insomnia and sleep [35,36], and interest in the nonpsychoactive properties of CBD in comparison to other cannabinoids. By 2016, a majority of US states had legalized medical cannabis or CBD [37], and in 2018, hemp-derived CBD was removed as a Schedule I substance in accordance with the 2018 Agriculture Improvement Act [38], increasing its availability to US consumers; however, there was a rapid decline in mentions of CBD in 2022. Although the extent of cannabis use for insomnia treatment and directionality is unclear, prior work suggests a high prevalence of insomnia symptoms and sleep disturbances among people who use cannabis, with 97% increased odds of insomnia among participants reporting daily cannabis use [39]. As cannabis products become more widely available, these findings support the need for greater investigation of the prevalence, possible benefits, and possible drawbacks of cannabis use for insomnia symptoms.

The use of social data may provide an alternative method to capture treatment trends that may be underreported by conventional methods. The 2017 AASM report speculated that the true use of sleep medications for insomnia is higher than reported [33]. Likewise, a prior analysis of US 1999-2010 National Health and Nutrition Examination Survey data found

XSL•FO
RenderX

that of those participants who reported using a sleep aid medication, 58% did not provide information for prescription medication, suggesting widespread use of OTC or alternative treatments [40]. Our findings also suggest a high prevalence of the mention of nonprescription sleep aids on r/insomnia and suggest that digital discussion boards may provide an alternative means of investigating these patterns. Additionally, sleep is a physiological process that can be influenced by sociocultural factors, as it is nested within a socioecological context of the individual, social factors, and society [41-43]. Because of this, social and cultural phenomena, such as societal change and upheaval, can impact sleep and insomnia. For example, measures of self-reported sleep and mood worsened during the week of the US 2020 election compared to a baseline measure a few weeks prior in both US and non-US participants [44]. Our results also suggest changepoints in treatment terms around 2016 and 2020, which may be related to political election cycles or the COVID-19 pandemic. Future research could more explicitly investigate the relationships between sleep health and social and societal factors using r/insomnia and other digital communities.

In addition to CBT-I, r/insomnia comments reflected positive sentiment toward "natural" or herbal supplements and therapies, such as melatonin, valerian root, and cannabidiol. Counter to our expectations that the term would have a negative connotation and sentiment, the term "hygiene," referencing the sleep hygiene component of CBT-I, was the term with the highest positive sentiment. This may reflect that the commenters were not bothered by the term or that this CBT-I component was successful in improving insomnia symptoms. The median sentiment for comments containing each term was positive, which may suggest that the data from this subreddit may be more positive than the median sentiment from the dataset VADER was trained on.

Strengths and Limitations

Our analysis has multiple strengths and limitations, which should be considered in interpreting results. Because the r/insomnia subreddit is accessible to people all over the world, the term trends may reflect an international sample of posts. Yet, it is not representative of a population with defined characteristics, as individuals who participated in the subreddit are self-selected. Our analysis was at the level of comments in response to original posts, treating each comment as independent. Therefore, the analysis may be biased by the overrepresentation of large threads focusing on specific topics. While sentiment analysis attempts to identify the general emotionality of text, there are some limitations; it is not able to capture sarcasm, and because the text is considered as a whole, text with both positive and negative comments will be treated as a composite rather than as an individual component. While we sought to include alternative or brand names for medications in trend measurement, treatments with many options or medication names (such as that for selective serotonin reuptake inhibitors) may exhibit residual underrepresentation or bias.

Conclusions

The use of language related to CBT-I and medications such as benzodiazepines and trazodone and other antidepressants have fluctuated over time on the r/insomnia Reddit platform. Some of these trends, such as the rise in CBT-I in 2017, may reflect clinical treatment guidelines, while others align with nationwide prescription trends. Trends of treatment-related terms, such as the changepoints around 2016 and 2020, may also reflect larger societal events, such as the 2020 COVID-19 pandemic. The data also suggest that r/insomnia commenters mention treatment terms related to OTC and alternative therapies, such as melatonin and cannabis. The representation of OTC and alternative treatments that are not captured by prescription activity is an especially important aspect of this dataset, as the use of nonprescription sleep aids is difficult to track; such information may be useful for sleep health practitioners. The r/insomnia dataset or other alternative data sources may be a valuable addition to prescription records for future studies seeking to capture nonprescription sleep aid use. The prevalence of these terms in the data suggests that future studies of insomnia should consider collecting information on OTC treatment use, cannabis use, and other alternative therapies that people may seek when experiencing insomnia symptoms. Data from digital communities such as r/insomnia may also provide a deeper understanding of how social and societal factors shape sleep health globally.

Acknowledgments

This study was supported by funding from the National Institutes of Health (NIH-NHLBI T32HL007901 supported DAW, K99HL166700 to DAW, and R01HL161012 to TS). DAW declares grant support from the National Institutes of Health and the Sleep Research Society and a past Travel Award from the Sleep Research Society. DAW reports unpaid committee service for the Sleep Research Society. There are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability

The r/insomnia Reddit data analyzed in this study are publicly available in Academic Torrents [45-47].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables and figures of treatment terms and trends from the r/insomnia subreddit.

XSL•FO
RenderX

[DOCX File , 694 KB - jmir_v27i1e58902_app1.docx]

References

1. Roth T. Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med 2007;3(5 Suppl):S7-S10 [FREE Full text] [Medline: 17824495]
2. Buysse DJ. Insomnia. JAMA 2013;309(7):706-716 [FREE Full text] [doi: 10.1001/jama.2013.193] [Medline: 23423416]
3. Morin CM, Jarrin DC. Epidemiology of insomnia: prevalence, course, risk factors, and public health burden. Sleep Med Clin 2022;17(2):173-191. [doi: 10.1016/j.jsmc.2022.03.003] [Medline: 35659072]
4. Kravitz HM, Ganz PA, Bromberger J, Powell LH, Sutton-Tyrrell K, Meyer PM. Sleep difficulty in women at midlife: a community survey of sleep and the menopausal transition. Menopause 2003;10(1):19-28. [doi: 10.1097/00042192-200310010-00005] [Medline: 12544673]
5. Morin CM, Jarrin DC, Ivers H, Mérette C, LeBlanc M, Savard J. Incidence, persistence, and remission rates of insomnia over 5 years. JAMA Netw Open 2020;3(11):e2018782 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18782] [Medline: 33156345]
6. Ishak WW, Bagot K, Thomas S, Magakian N, Bedwani D, Larson D, et al. Quality of life in patients suffering from insomnia. Innov Clin Neurosci 2012;9(10):13-26 [FREE Full text] [Medline: 23198273]
7. Kessler RC, Berglund PA, Coulouvrat C, Hajak G, Roth T, Shahly V, et al. Insomnia and the performance of US workers: results from the America Insomnia Survey. Sleep 2011;34(9):1161-1171 [FREE Full text] [doi: 10.5665/SLEEP.1230] [Medline: 21886353]
8. Kessler RC, Berglund PA, Coulouvrat C, Fitzgerald T, Hajak G, Roth T, et al. Insomnia, comorbidity, and risk of injury among insured Americans: results from the America Insomnia Survey. Sleep 2012;35(6):825-834 [FREE Full text] [doi: 10.5665/sleep.1884] [Medline: 22654202]
9. Edinger JD, Arnedt JT, Bertisch SM, Carney CE, Harrington JJ, Lichstein KL, et al. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. J Clin Sleep Med 2021;17(2):263-298 [FREE Full text] [doi: 10.5664/jcsm.8988] [Medline: 33164741]
10. Muench A, Vargas I, Grandner MA, Ellis JG, Posner D, Bastien CH, et al. We know CBT-I works, now what? Fac Rev 2022;11:4 [FREE Full text] [doi: 10.12703/r/11-4] [Medline: 35156100]
11. Gyllenhaal C, Merritt SL, Peterson SD, Block KI, Gochenour T. Efficacy and safety of herbal stimulants and sedatives in sleep disorders. Sleep Med Rev 2000;4(3):229-251. [doi: 10.1053/smrv.1999.0093] [Medline: 12531167]
12. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016;165(2):125-133 [FREE Full text] [doi: 10.7326/M15-2175] [Medline: 27136449]
13. Robbins R, Depner C, Grandner M, Khosla S, Macedo D, Stewart N. The internet and social media: platforms that offer promise and peril for disseminating sleep health information and for sleep disorders awareness, evaluation, and treatment. 2022 Presented at: Sleep; June 4-8, 2022; Charlotte, NC, United States URL: https://www.sleepmeeting.org/wp-content/uploads/2022/03/SLEEP-2022-Preliminary-Program.pdf
14. Manchaiah V, Londero A, Deshpande AK, Revel M, Palacios G, Boyd RL, et al. Online discussions about tinnitus: what can we learn from natural language processing of Reddit posts? Am J Audiol 2022;31(3S):993-1002. [doi: 10.1044/2021_AJA-21-00158] [Medline: 35130042]
15. Cummins JA, Zhou G, Nambudiri VE. Natural language processing for large-scale analysis of eczema and psoriasis social media comments. JID Innov 2023;3(5):100210 [FREE Full text] [doi: 10.1016/j.xjidi.2023.100210] [Medline: 37564106]
16. Benson R, Hu M, Chen AT, Zhu SH, Conway M. Examining cannabis, tobacco, and vaping discourse on Reddit: an exploratory approach using natural language processing. Front Public Health 2021;9:738513 [FREE Full text] [doi: 10.3389/fpubh.2021.738513] [Medline: 35071153]
17. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: observational study. J Med Internet Res 2020;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
18. AASM Sleep Prioritization Survey: sleep aid use. American Academy of Sleep Medicine. 2022. URL: https://aasm.org/wp-content/uploads/2022/06/sleep-prioritization-survey-sleep-aids.pdf [accessed 2024-11-21]
19. R/Insomnia. Reddit. URL: https://www.reddit.com/r/insomnia/ [accessed 2023-08-14]
20. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit dataset. 2020 Presented at: International AAAI Conference on Web and Social Media; June 8-11, 2019; Atlanta, Georgia. [doi: 10.1609/icwsm.v14i1.7347]
21. Baumgartner J, Lazzarin E, Seiler A. Pushshift Reddit API. GitHub. URL: https://github.com/pushshift/api [accessed 2023-08-14]
22. Rehurek R, Sojka P. Gensim—Python framework for fast Vector Space Modelling. 2011 Presented at: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks; May 22, 2010; Valletta, Malta URL: https://pypi.org/project/gensim/
23. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng 2007;9(3):90-95. [doi: 10.1109/mcse.2007.55]

XSL•FO
RenderX

24. Colten H, Altevogt B. Improving awareness, diagnosis, and treatment of sleep disorders. In: Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington, DC: National Academies Press; 2006.
25. Thomas A, Grandner M, Nowakowski S, Nesom G, Corbitt C, Perlis ML. Where are the behavioral sleep medicine providers and where are they needed? A geographic assessment. Behav Sleep Med 2016;14(6):687-698 [FREE Full text] [doi: 10.1080/15402002.2016.1173551] [Medline: 27159249]
26. Kaufmann CN, Spira AP, Alexander GC, Rutkow L, Mojtabai R. Trends in prescribing of sedative-hypnotic medications in the USA: 1993-2010. Pharmacoepidemiol Drug Saf 2016;25(6):637-645 [FREE Full text] [doi: 10.1002/pds.3951] [Medline: 26711081]
27. Kaufmann CN, Spira AP, Depp CA, Mojtabai R. Long-term use of benzodiazepines and nonbenzodiazepine hypnotics, 1999-2014. Psychiatr Serv 2018;69(2):235-238 [FREE Full text] [doi: 10.1176/appi.ps.201700095] [Medline: 29089011]
28. Milani SA, Raji MA, Chen L, Kuo YF. Trends in the use of benzodiazepines, Z-hypnotics, and serotonergic drugs among US women and men before and during the COVID-19 pandemic. JAMA Netw Open 2021;4(10):e2131012 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.31012] [Medline: 34694388]
29. Tran T, Kavuluru R. Social media surveillance for perceived therapeutic effects of cannabidiol (CBD) products. Int J Drug Policy 2020 Mar;77:102688 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102688] [Medline: 32092666]
30. Gliniecka M. The ethics of publicly available data research: a situated ethics framework for Reddit. Social Media + Soc 2023 Aug 06;9(3). [doi: 10.1177/20563051231192021]
31. Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (Version 2.0). Association of Internet Researchers. 2012. URL: https://aoir.org/reports/ethics2.pdf [accessed 2024-12-02]
32. Wong J, Murray Horwitz M, Bertisch SM, Herzig SJ, Buysse DJ, Toh S. Trends in dispensing of zolpidem and low-dose trazodone among commercially insured adults in the United States, 2011-2018. JAMA 2020;324(21):2211-2213 [FREE Full text] [doi: 10.1001/jama.2020.19224] [Medline: 33258882]
33. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2017;13(2):307-349 [FREE Full text] [doi: 10.5664/jcsm.6470] [Medline: 27998379]
34. Pelayo R, Bertisch SM, Morin CM, Winkelman JW, Zee PC, Krystal AD. Should trazodone be first-line therapy for insomnia? A clinical suitability appraisal. J Clin Med 2023;12(8):2933 [FREE Full text] [doi: 10.3390/jcm12082933] [Medline: 37109268]
35. Whiting PF, Wolff RF, Deshpande S, Di Nisio M, Duffy S, Hernandez AV, et al. Cannabinoids for medical use: a systematic review and meta-analysis. JAMA 2015;313(24):2456-2473. [doi: 10.1001/jama.2015.6358] [Medline: 26103030]
36. Shannon S, Opila-Lehman J. Effectiveness of cannabidiol oil for pediatric anxiety and insomnia as part of posttraumatic stress disorder: a case report. Perm J 2016;20(4):16-005 [FREE Full text] [doi: 10.7812/TPP/16-005] [Medline: 27768570]
37. Pacula RL, Smart R. Medical marijuana and marijuana legalization. Annu Rev Clin Psychol 2017;13:397-419 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-045128] [Medline: 28482686]
38. H.R.2—Agriculture Improvement Act of 2018. 115th Congress (2017-2018). 2018. URL: https://www.congress.gov/bill/115th-congress/house-bill/2 [accessed 2024-11-14]
39. Coelho J, Montagni I, Micoulaud-Franchi JA, Plancoulaine S, Tzourio C. Study of the association between cannabis use and sleep disturbances in a large sample of university students. Psychiatry Res 2023;322:115096. [doi: 10.1016/j.psychres.2023.115096] [Medline: 36842936]
40. Bertisch SM, Herzig SJ, Winkelman JW, Buettner C. National use of prescription medications for insomnia: NHANES 1999-2010. Sleep 2014;37(2):343-349 [FREE Full text] [doi: 10.5665/sleep.3410] [Medline: 24497662]
41. Grandner MA. Social-ecological model of sleep health. In: Sleep and Health. San Diego, CA: Elsevier; 2019:45-53.
42. Grandner MA. Sleep, health, and society. Sleep Med Clin 2017;12(1):1-22 [FREE Full text] [doi: 10.1016/j.jsmc.2016.10.012] [Medline: 28159089]
43. Billings ME, Hale L, Johnson DA. Physical and social environment relationship with sleep health and disorders. Chest 2020;157(5):1304-1312 [FREE Full text] [doi: 10.1016/j.chest.2019.12.002] [Medline: 31870910]
44. Cunningham TJ, Fields EC, Denis D, Bottary R, Stickgold R, Kensinger EA. How the 2020 US presidential election impacted sleep and its relationship to public mood and alcohol consumption. Sleep Health 2022;8(6):571-579 [FREE Full text] [doi: 10.1016/j.sleh.2022.08.009] [Medline: 36280586]
45. r/insomnia subreddit comments. Academic Torrents. 2023 Feb 28. URL: https://academictorrents.com/details/c398a571976c78d346c325bd75c47b82edf6124e [accessed 2024-12-02]
46. Lo HZ, Cohen JP. Academic Torrents: scalable data distribution. 2016 Presented at: Neural Information Processing Systems Challenges in Machine Learning (CiML) Workshop; 2016; xx URL: http://arxiv.org/abs/1603.04395
47. Cohen JP, Lo HZ. Academic Torrents: a community-maintained distributed repository. 2014 Presented at: Annual Conference of the Extreme Science and Engineering Discovery Environment; July 13-18, 2014; Atlanta, GA, United States. [doi: 10.1145/2616498.2616528]

XSL•FO
RenderX

Abbreviations

AASM: American Academy of Sleep Medicine
BOW: bag of words
CBD: cannabidiol
CBT-I: cognitive behavioral therapy for insomnia
NLP: natural language processing
OTC: over-the-counter
RE: regular expression
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by A Mavragani; submitted 27.03.24; peer-reviewed by GMD Dore, C Thornton; comments to author 12.06.24; revised version received 08.08.24; accepted 16.10.24; published 09.01.25.

Please cite as:
Cummins JA, Gottlieb DJ, Sofer T, Wallace DA
Applying Natural Language Processing Techniques to Map Trends in Insomnia Treatment Terms on the r/Insomnia Subreddit: Infodemiology Study
J Med Internet Res 2025;27:e58902
URL: https://www.jmir.org/2025/1/e58902
doi:10.2196/58902
PMID:

©Jack A Cummins, Daniel J Gottlieb, Tamar Sofer, Danielle A Wallace. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predicting User Engagement in Health Misinformation Correction on Social Media Platforms in Taiwan: Content Analysis and Text Mining Study

Hsin-Yu Kuo[1*], MA; Su-Yen Chen[2*], PhD

[1]Department of Educational Psychology and Counseling, National Tsing Hua University, Hsinchu, Taiwan

[2]Institute of Learning Sciences and Technologies, National Tsing Hua University, Hsinchu, Taiwan

[*]all authors contributed equally

Corresponding Author:
Hsin-Yu Kuo, MA
Department of Educational Psychology and Counseling
National Tsing Hua University
101, Section 2, Kuang-Fu Road
Hsinchu, 300044
Taiwan
Phone: 886 3 571 5131 ext 34510
Email: kuohy@mx.nthu.edu.tw

Abstract

Background: Health misinformation undermines responses to health crises, with social media amplifying the issue. Although organizations work to correct misinformation, challenges persist due to reasons such as the difficulty of effectively sharing corrections and information being overwhelming. At the same time, social media offers valuable interactive data, enabling researchers to analyze user engagement with health misinformation corrections and refine content design strategies.

Objective: This study aimed to identify the attributes of correction posts and user engagement and investigate (1) the trend of user engagement with health misinformation correction during 3 years of the COVID-19 pandemic; (2) the relationship between post attributes and user engagement in sharing and reactions; and (3) the content generated by user comments serving as additional information attached to the post, affecting user engagement in sharing and reactions.

Methods: Data were collected from the Facebook pages of a fact-checking organization and a health agency from January 2020 to December 2022. A total of 1424 posts and 67,378 corresponding comments were analyzed. The posts were manually annotated by developing a research framework based on the fuzzy-trace theory, categorizing information into “gist” and “verbatim” representations. Three types of gist representations were examined: risk (risks associated with misinformation), awareness (awareness of misinformation), and value (value in health promotion). Furthermore, 3 types of verbatim representations were identified: numeric (numeric and statistical bases for correction), authority (authority from experts, scholars, or institutions), and facts (facts with varying levels of detail). The basic metrics of user engagement included shares, reactions, and comments as the primary dependent variables. Moreover, this study examined user comments and classified engagement as cognitive (knowledge-based, critical, and bias-based) or emotional (positive, negative, and neutral). Statistical analyses were performed to explore the impact of post attributes on user engagement.

Results: On the basis of the results of the regression analysis, risk (β=.07; *P*=.001), awareness (β=.09; *P*<.001), and facts (β=.14; *P*<.001) predicted higher shares; awareness (β=.07; *P*=.001) and facts (β=.24; *P*<.001) increased reactions; and awareness (β=.06; *P*=.005), numeric representations (β=.06; *P*=.02), and facts (β=.19; *P*<.001) increased comments. All 3 gist representations significantly predicted shares (risk: β=.08; *P*<.001, awareness: β=.08; *P*<.001, and value: β=.06; *P*<.001) and reactions (risk: β=.04; *P*=.007, awareness: β=.06; *P*<.001, and value: β=.05; *P*<.001) when considering comment content. In addition, comments with bias-based engagement (β=–.11; *P*=.001) negatively predicted shares. Generally, posts providing gist attributes, especially awareness of misinformation, were beneficial for user engagement in misinformation correction.

Conclusions: This study enriches the theoretical understanding of the relationship between post attributes and user engagement within web-based communication efforts to correct health misinformation. These findings provide a foundation for designing more effective content approaches to combat misinformation and strengthen public health communication.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e65631) doi:10.2196/65631

KEYWORDS

health misinformation; misinformation correction; fact-checking; content analysis; text mining; fuzzy-trace theory; social media; large language models; user engagement; health communication

Introduction

Background

Health misinformation negatively impacts individuals' and society's adaptive response to a health crisis [1,2]. As a communication tool, social media has worsened the problem by speeding up the spread of misinformation and strengthening people's existing beliefs [3-5]. Conversely, social media platforms can be effectively used for health communication to disseminate accurate information by correcting misinformation and engaging users by providing verified health advice [6-9]. Many organizations devote efforts to correcting misinformation and releasing them on social media. However, the efficacy of correction is sometimes limited for various reasons, such as people feeling uncomfortable with the knowledge gap when retracting misinformation [10]. Moreover, in an increasingly complex health information environment, people are flooded with knowledge they may not be able to use correctly [11]. In this context, the conditions and message designs that make misinformation correction an effective countermeasure have been examined in experimental studies [6,12,13]. As an interactive digital environment, social media also allows researchers to access data on user responses to health communication, obtaining results from different perspectives [7-9,14].

In this study, we investigate how to enhance the effectiveness of misinformation correction based on real-world data. We selected 2 representative Facebook fan pages in Taiwan (Facebook, Meta Platforms, Inc) that provide rich health misinformation corrections in post forms. Users engage with the content provided by these 2 organizations and other users' comments. In general, correction is a remedy that offers accurate information to counteract misinformation presented in media [10]. According to the observed data, the definition of health misinformation correction in this study is the content with the intent to resist health-related misinformation. This includes clarifications of health misinformation, fact-checking that covers the verification process, and content that addresses potential misinformation while not directly correcting a specific claim (examples are provided in Table S1 in Multimedia Appendix 1). The topics covered include diseases; health care; food safety; and issues extending from the COVID-19 pandemic, such as public welfare and information security. Then, we need a theoretical framework to identify the posts' attributes (called features or characters in some studies [7,9]). The fuzzy-trace theory (FTT) provides a framework that differentiates between 2 cognitive representations of a message, "gist" (top-down meaning) and "verbatim" (detailed information) [15], helping us focus on analyzing textual meaning in posts rather than merely considering surface attributes such as length and tone. The theory was also used to explore physician-patient, health, and scientific communication [7,16,17].

The responses social media users have when they see posts, particularly those explicitly displayed on social media, are referred to as user engagement in this study. These are important metrics for examining the impacts of post attributes in many studies [8,9,14]. This study explored user engagement in 2 dimensions. The first is based on the metrics Facebook provides (sharing, reacting, and commenting), which are the main focus of the predictions in this research. The second is to examine the content users create when they comment. We assume that users engage with different intentions in their comments, and this is reflected in the comment content, which is presented as additional information accessible to other users. Then, this study proposes a framework to classify the engagement categories in comments and examine if this content affects the other 2 metrics: sharing and reacting.

We then introduced prior work about health communication on social media, explained the research gap, and explored the perspective of FTT to identify post attributes and the relationship between post attributes and user engagement in this study.

Prior Work

In previous studies, the acceptance of misinformation correction across different contexts and content types has been examined through experiments. The variables often examined include simple versus detailed or factual elaboration [12,13,18]; correction sources [6,12,13,19]; and information content such as quality, intensity, or emphasis [18,20,21].

Another approach in the context of social media involves collecting data from these platforms, labeling them manually or automatically, and analyzing them. A study using FTT as a framework explored news reports about measles and vaccines and expressed the opinions of vaccination proponents or opponents as the gist. In this study, data were coded to predict the likelihood of news sharing on Facebook. The regression model included other variables, such as statistics and stories, but only the gist could significantly predict news sharing [14]. Health communication efforts increased significantly during the COVID-19 pandemic, allowing for analysis through publicly available data. A study of 1816 posts from hospitals in Taiwan during the first 4 months of the COVID-19 pandemic found that specific social media strategies significantly increased user engagement, indicating the potential to improve health literacy among health care organizations [7]. Such studies have also observed that users inevitably share misinformation through comments when discussing health issues on social media. A study analyzed 1534 comments from 2 Facebook groups discussing mental health symptoms between January 2019 and December 2021. It found that 26.1% of the comments contained medically inaccurate information, specifically regarding complementary and alternative medicine (60.3%) [22]. By observing long-term trends, it is also possible to track changes in the quality of health communication and user attitudes. An

XSL•FO
RenderX

analysis of 12,553 COVID-19 vaccine fact-checking posts and 122,362 comments from January 2020 to March 2022 showed increased analytical thinking and confidence in the posts, alongside decreased tentativeness among comments, showing an evolution in public sentiment over time [8]. Finally, a study analyzed 914 fact-checking news articles to understand how source transparency, contextual information, and vividness influence user engagement. The study found that these factors significantly affected user responses, especially likes and angry reactions, emphasizing the importance of content style and depth in capturing public attention [9]. These studies underline the critical role of social media in shaping public health narratives and emphasize the necessity of designing engaging yet accurate content during health crises.

Research Gap

Regarding the research gap, there is a lack of health communication studies that incorporate both data and theory, and finding a theory that fits the data is not easy. To explore the effect of misinformation correction, the primary theories considered include the inoculation theory [23], elaboration likelihood model [24], heuristic-systematic model [25], and FTT [15], all of which offer different interpretations from a cognitive perspective on how misinformation correction can be effectively achieved. The FTT emphasizes the importance of conveying the essence of a message, focusing on its core meaning rather than merely summarizing or simplifying it. It highlights the need to avoid overwhelming people with excessive quantitative details. [26]. Moreover, FTT has robust cognitive neuroscience foundations, making it suitable for scientific communication and compliant with nonpersuasive fact-checking methods that prioritize reference-based corrections. Therefore, this study adopted FTT for analysis. The method of conveying essential meaning is called gist communication [17].

In addition, we proposed a framework for classifying user-generated comment content, as previous studies have rarely examined this aspect. We have included them as independent variables in the prediction model to explore whether the FTT's gist communication remains effective even when different comments are present as additional post information.

The FTT and Health Misinformation Correction

Overview

The FTT explains how we process and remember information through 2 types of cognitive representation: "gist" and "verbatim." Gist captures the essential, top-down meaning, including causal relations and key concepts [27,28]. Verbatim captures details such as exact words and elements [15,16]. This theory was originally developed to study how individuals shape their memory and engage in reasoning and decision-making [15]. The gist is more than a mental shortcut; it is a distilled way of presenting objective facts to help individuals make health and medical decisions and navigate life [29].

According to the FTT, information containing the gist is likely to evoke motivationally relevant values that may be shared online, whether true or not [30]. This explains why people easily share misinformation that offers a clear core message and causal explanation [27]. This mechanism can be applied in fact-checking or correction as an intervention. FTT explains that providing a brief explanation has more lasting benefits than merely offering a false label [31], but offering decontextualized detailed information may be counterproductive [18]. The critical factor may be the inclusion of the gist. The gist of information triggers emotions and evokes core values [17]. Furthermore, when the content provides gist cues that connect with people's inner motivations and values, it is more likely to catch attention and be shared [30].

Health interventions should emphasize conveying the gist of information with an organized approach to highlight essential points [32]. The next question is to identify the gist of the health misinformation correction. Different gist representations exist in diverse contexts. In a study on physician-patient communication, the gist of the effectiveness and risks of medications was extracted from communication records to examine its relationship with patient satisfaction [16]. On the basis of the FTT and important concepts discussed in misinformation research, we proposed 3 gist representations: risks associated with misinformation, awareness of misinformation, and value in health promotion. In addition, we proposed 3 verbatim representations: numeric and statistical information as a reference for correction; authority from named experts, scholars, or official institutions used as references; and facts presented with varying levels of detail. These 6 representations, abbreviated as risk, awareness, value, numeric, authority, and facts, serve as attributes of misinformation correction posts.

Risk: Risks Associated With Misinformation

During health crises, it is important to communicate risks accurately and effectively [32,33], as it significantly correlates with adopting preventive health behaviors [34]. The risks associated with misinformation in this study included harm to personal and public health, social panic, and legal risks of spreading misinformation.

Awareness: Awareness of Misinformation

Public awareness plays a crucial role in improving resilience from misinformation, including awareness of exposure to misinformation and the ability to identify it [35]. Moreover, providing awareness and warnings beforehand usually has a beneficial impact [23]. In this study, awareness of misinformation involved practical information to remind and help individuals recognize and verify misinformation.

Value: Value in Health Promotion

Promoting health and well-being is a primary goal of health communication [36]. When correcting misinformation, it is recommended not only to mention what is false but also to provide alternative accurate information and encourage self-affirmation [10]. Conveying public trust and confidence is vital to combat anxiety-induced acceptance of misinformation [37]. Therefore, this study identified the third gist of health promotion as the core value toward health crises, including health literacy and a positive attitude.

XSL•FO
RenderX

Numeric: Numeric and Statistical Information as a Reference for Correction

Scientific communication often includes numbers and statistics as objective data. People's numeracy levels can affect their interpretations of risks, probabilities, and numerical results, leading to potential misunderstandings [38]. This study examined the use of numerical information to correct misinformation. Indeed, prior content analysis studies have considered statistical data as an attribute or characteristic of posts [9,14].

Authority: Authority From Named Experts, Scholars, or Official Institutions Used as References

Experts, scholars, and professional organizations are trusted sources of knowledge for authoritative communication, and health care professionals can correct health misinformation on social media [39,40]. Fact-checking organizations consult experts to improve the credibility of misinformation corrections. Thus, this study also investigated the effects of referencing authority.

Facts: Facts Presented With Varying Levels of Detail

People's attention span is generally limited on social media, and the effects of detailed factual information should be evaluated. The outcomes of detailed factual elaboration and simple rebuttal have been discussed in prior research [12,41]. In this study, fact representations refer to detailed and scientific elaborations with evidence for correcting misinformation. These can be categorized into three levels: (1) those that do not target specific misinformation; (2) simple corrections involving straightforward explanations; and (3) detailed corrections encompassing comprehensive information, typically with a complete structure, including background, reasons for errors, and alternative explanations.

User Engagement on Social Media

Social media offers benefits for health communication, such as improving the accessibility of health information and allowing users to generate and exchange information, but the quality and reliability of the shared information must be monitored [42]. The FTT suggests that information with gist cues supports reasoning and decision-making, and they are more likely to share when the cues connect their motivations or values. Facebook shares are commonly used as a metric, indicating users' affirmation and commitment to a post [43]. In addition to sharing, users can also engage through reactions (including likes and 6 emoticons) and comments on social media platforms. A study used the common-sense model of illness self-regulation to describe humans' cognitive and emotional systems to increase user engagement toward some attributes of illness representations on social media [44]; that is, shares, reactions, and comments were metrics driven by cognitive and emotional factors.

This study defined the basic metrics of user engagement as follows:

- Shares—user engagement in sharing, measured by share count
- Reactions—user engagement in reacting, measured by summing up clicks of likes and 6 other emoticons
- Comments—user engagement in commenting, measured by comment count

Metrics such as shares, reactions, and comments were commonly used in health communication studies to identify factors that increase user engagement [44-46]. However, user engagement is not always positive, especially in comments, which often contain errors, biases, and uncivil behavior [22,47,48]. The phenomenon of echo chambers on social media refers to selective exposure and confirmation bias, which limit diverse perspectives and foster polarized groups [3]. A study distinguishes between 2 belief systems among users: scientific and conspiracy [49]. Some studies have suggested that user-generated comments influence other users' perceptions of posts. For example, it was found that the perceived quality of journalistic content being commented on decreased due to a lack of reasoning and incivility in user comments [47]. When exposed to critical comments about a fake news article, people are less likely to share or comment positively on the article [50]. A study reported that positive comments influenced people's perceptions of a company's reputation in crisis [51]. However, well-informed individuals often remain quiet on certain issues, whereas misinformed individuals tend to be more vocal online [52].

This study categorized comments into cognitive and emotional engagement [44]. For cognitive engagement, we were concerned with the conspiracy belief group and its counterpart, the scientific group [49]. We observed the data potentially driven by conspiracy belief as bias-based engagement and identified 2 types of engagement driven by scientific belief: one focused on knowledge itself, called knowledge-based engagement, and the other on misinformation issues, called critical engagement. Furthermore, we observed user engagement driven by emotions, which can be divided into positive, negative, and neutral emotional engagement according to the positive-negative valence framework [53]. We classified comments into one of six categories, which are further defined as follows:

1. Knowledge-based engagement. This study observed users' knowledge-related behaviors on social media driven by scientific beliefs, which include providing additional information, clarifying facts, and asking questions.
2. Critical engagement. Developing critical thinking skills and pragmatic skepticism is crucial for citizens to combat misinformation [35]. Unlike knowledge-based engagement, which focuses on knowledge itself, critical engagement identified in this study focused on discussing misinformation by offering warnings or explanations.
3. Bias-based engagement. Bias-based engagement refers to users disseminating biased content driven by conspiracy beliefs, including misinformation, incorrect inferences, logical fallacies, or conspiracy theories, potentially undermining information reliability and impeding evidence-based discourse.
4. Positive emotional engagement. Positive emotional engagement is characterized by users expressing positive emotions, such as joy, happiness, and excitement.

XSL•FO
RenderX

5. Negative emotional engagement. Negative emotional engagement is characterized by users expressing negative emotions, such as anger, fear, sadness, or frustration.
6. Neutral emotional engagement. Neutral emotional engagement describes a user's subdued or indifferent emotional response to content or events.

Research Questions

User attitudes toward correction may change over time [8], and the misinformation contained in comments is also noteworthy [22]. Therefore, this study first investigated whether user engagement toward health misinformation correction has changed over the past 3 years. Second, according to the FTT, providing the gist as cues can effectively and directly convey messages to users [17]. Accordingly, we hypothesized that providing gist representations would help users grasp the main aspects of health misinformation issues. We also identified some important verbatim representations as post attributes to examine their relationship with user engagement. Finally, we considered whether the content of comments affects users' choice to share and react to a post. On the basis of the earlier discussion, the research questions (RQs) posed by this study are as follows:

- How has user engagement toward health misinformation correction changed during the COVID-19 pandemic? (RQ1)
- Can the gist and verbatim representations identified in this study as post attributes predict user engagement in shares, reactions, and comments? (RQ2)
- On the basis of RQ2, does the inclusion of user-generated comment content from engagement in commenting affect engagement in sharing and reactions? (RQ3)

Methods

Corpus and Sample Period

This study focused on health misinformation correction, drawing on 2 representative platforms, the Taiwan Fact-Checking Center (TFC) and the Ministry of Health and Welfare (MOHW), with rich data, including fact-checking and clarifying health-related event or policy posts and user feedback. Data were collected from the Facebook pages of the TFC, which was established in 2018 and has >160,000 followers on Facebook, and the MOHW, which has a significant following of approximately 1.52 million Facebook users, accounting for 6.45% of the total population of Taiwan. We collected 3769 posts from the TFC and 5103 posts from the MOHW from January 2020 to December 2022. The TFC targets various misinformation issues such as health, politics, economics, and celebrities. By contrast, the post content of MOHW focuses on public health, such as health policy announcements, and the correction of health misinformation accounts for a small portion. To ensure our focus was on health misinformation correction, we manually reviewed each post to identify those that involved both health and misinformation correction. For TFC posts, the reference keywords included health, pandemic, disease, food safety, folk remedies, vaccine, and health care. For MOHW posts, the reference keywords included misinformation, false information, clarification, and fake news. Because relying solely on keywords may either miss or capture non-target content, we primarily relied on manual review. After the filtering process, we found 34.89% (1315/3769) posts from the TFC source and 2.14% (109/5103) posts from the MOHW source. Corresponding to the posts from TFC and MOHW, 26,458 (including 913 self-replies by the TFC manager) and 40,920 comments (including 97 self-replies by the MOHW manager) were downloaded, respectively. Finally, the dataset consisted of 1424 posts and 67,378 comments from both platforms.

Research Process

Overview

This study included three phases, as shown in Figure 1: (1) obtaining research data and organizing them into a dataset, (2) processing post and comment data, and (3) data integration and statistical analyses. First, after defining the research participants, we used web scraping techniques to acquire the necessary posts and relevant metrics and obtain the content of comments. Subsequently, to process unstructured data into structured data for further analysis, manual annotation was applied to posts to obtain 1 indicator (readability) using computational methods, whereas a large amount of comment data was automatically annotated using GPT-3.5 (OpenAI). Finally, a statistical analysis was conducted. Detailed information regarding data annotation processing is provided in the following sections.

XSL•FO
RenderX

Figure 1. Research procedure and data processing.

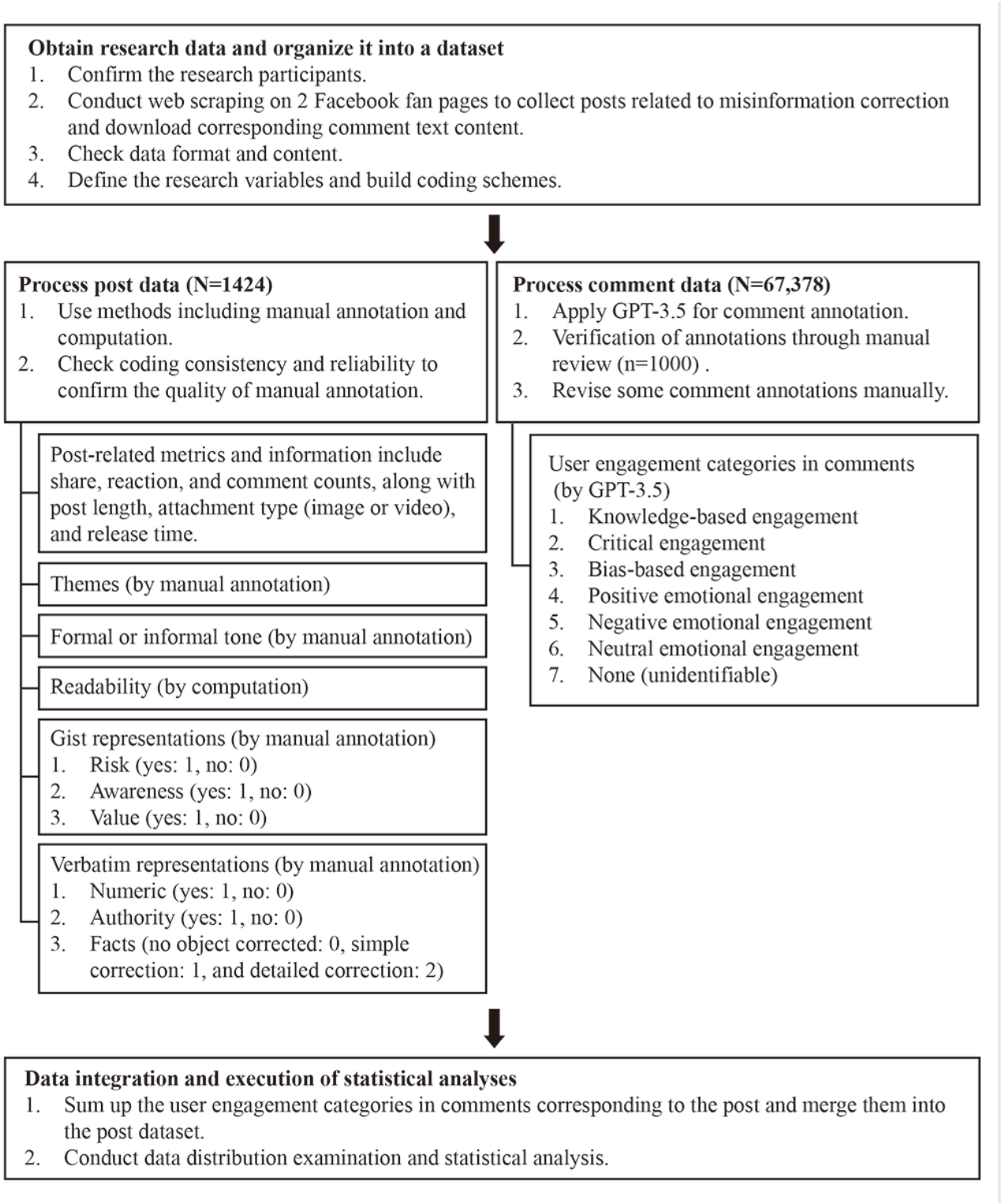

Annotation and Transformation Data for Posts

Regarding the posts, we categorized them into five main themes based on the emphasized content: (1) understanding and controlling diseases; (2) vaccine safety and policies; (3) health care, food safety, and home remedies; (4) livelihood, economy, and information security; and (5) misinformation awareness and identification. Livelihood, economy, and information security refers to daily life aspects impacted by the COVID-19 pandemic, such as subsidy applications. Two coders manually coded this component (Cohen κ coefficient=0.672). In addition, as control variables, this study considered the length of posts, their readability, whether they were from official platforms, whether they adopted a formal tone, and whether they used images.

Due to the lack of Chinese readability tools, this study created a difficult-word dictionary based on 2 resources: the “Chinese 8000-Word List” levels 3 to 5 [54] and the “National Academy for Educational Research Third Level Seven-Word List” level

XSL•FO
RenderX

7 [55]. In addition, the word frequency table of this dataset was examined, and 340 words, such as reagents and nucleic acid, were added to the dictionary. The difficult-word dictionary included 11,138 words. Then, we used the Dale-Chall readability formula to compute the scores. The formula of the Dale-Chall Readability Score is 0.1579 × [(number of difficult words / total number of words) × 100] + 0.0496 × (total number of words / total number of sentences) [56,57]. The original score ranges from 0 to 10, with higher scores indicating greater difficulty in reading. Furthermore, the additive inverse of this score was calculated for intuitive interpretation, with higher scores indicating easier readability.

Whether a post adopted a formal tone was determined by manual annotation (Fleiss κ=0.521). Because determining whether the tone is formal can be subjective, we gradually reached a consensus throughout the process. This included discussions on the use of common Chinese modal particles and slang. At least 2 annotators annotated each post, and after comparing the annotation results, we discussed and resolved any inconsistencies individually.

The last control variable was whether an image was used. The data we obtained through web scraping included information about attachments, such as images, links, or videos. Uploaded images were prioritized for display in Facebook posts. If no image was uploaded but the content included a link, a preview image was selected from the linked page for display. Similarly, a thumbnail image was shown when a video was uploaded in a post. When an image was used as an attachment, it was typically designed. We did not evaluate the content of the images. Of the 1424 posts, there were 1310 (91.99%) posts with uploaded images as attachments (ranging from 1 to 15 images), 109 (7.65%) posts that displayed an image from a link, 4 (0.28%) posts with an uploaded video, and only 1 (0.07%) post without any attachment. We assigned a dummy code of 1 to the 91.99% (1310/1424) posts with a designed image as an attachment, while the others were coded as 0.

This study focused on examining 6 representations: risk, awareness, value, numeric, authority, and facts, which serve as attributes of misinformation correction posts identified by the FTT and related literature. Please refer to Tables S2 and S3 in Multimedia Appendix 1 for the gist and verbatim coding schemes. This portion of the annotated data was converted into dummy variables. The first author and an annotator reviewed the data during the initial phase and conducted theoretical discussions to create the coding scheme. A 10.04% (143/1424) sample of the posts was manually annotated, and ambiguous concepts were clarified. Two annotators independently annotated another 9.97% (142/1424) of the sample. The interannotator agreement ranged from 87.32% to 94.37%, and Fleiss κ values ranged from 0.459 to 0.927. Each post was coded by at least 2 annotators, and any inconsistencies were discussed and confirmed to reach a consensus.

Annotation Data for Comments

Regarding the comments, we classified engagement in comments into 6 categories based on a literature review and empirical data observation. Three categories were related to cognition: knowledge-based, critical, and bias-based engagement. In addition, 3 categories were related to emotion: positive, negative, and neutral. Each comment was assigned 1 of these 6 constructs.

GPT-3.5 has been trained on sufficiently diverse multilingual data and can perform cross-lingual transfer learning by using knowledge from source languages to make inferences in target languages, even with limited target language data [58]. The comment content was mainly in Chinese, and the average length of a comment was 26.40 (SD 52.70) words. We assigned each comment to a category, although it is possible for a single comment to contain multiple categories in short texts. We prioritized using GPT-3.5 to determine cognitive engagement; if not, we assessed for emotional engagement. The comment was classified as None if it was indistinguishable and neither aspect could be determined from the content. Examples of the prompts and outputs are provided in Tables S4 and S5 in Multimedia Appendix 1.

GPT-3.5 has some issues with instability and hallucinations. Regarding annotation instability, we set the temperature parameter to 0.5. This parameter ranges from 0 to 1, with higher values increasing creativity. We found that setting the temperature too low could lead to conservative responses, resulting in frequent annotating as unidentifiable. In the study, approximately 100 comments were used for testing to improve the prompt and adjust the temperature parameter. We required GPT-3.5 to provide reasons during annotation to examine potential causes of annotating errors and incorporated more specific definitions into the prompt. Next, we performed automated annotation of all data. To ensure GPT-3.5 annotation reliability, we randomly selected 1000 comments for manual annotation. Cohen κ coefficient for manual annotations was 0.519 (moderate agreement). After discussion, a consensus on the annotation was reached. The consistency between manual and GPT-3.5 annotations was calculated, yielding a Cohen κ coefficient of 0.648 (substantial agreement). Then, to mitigate potential biases introduced by GPT-3.5, we systematically compared manual annotations with GPT-3.5 annotations and checked the comments for keywords that GPT-3.5 may not recognize. We revised 4233 comment annotations in this process. Finally, the data for the 6 categories of comments were summed and linked to their corresponding posts.

Data Analysis

We initially used descriptive statistics to provide an overall overview of the data to answer RQ1. The mean share count was 79.12 (SD 293.14; range 0-4259), with 57 posts without sharing. The mean reaction count was 914.35 (SD 2,582.93; range 13-34,403). The mean comment count, excluding self-replies by the platform manager was 46.61 (SD 212.82; range 0-5092), with 198 posts without commenting. The 1.5% (1010/67,378) comments made by the platform manager supplied full information links or expressed positive feedback to some comments. These comments were excluded because we aimed to understand engagement from regular users. After examining the data distribution with Q-Q plots, we found that share, reaction, and comment count were nonnormally distributed. In this situation, it may be appropriate to transform the dependent variable using a natural logarithm in a linear regression model

XSL•FO
RenderX

[9,59] or to use Poisson or negative binomial regression models [44,56]. We chose to use natural logarithms of the dependent variables as ln(shares+1), ln(reactions), and ln(comments+1). To improve the interpretability of the model, this study also transformed some independent variables using the natural logarithm, including ln(post length) and ln(engagement categories in comments+1). This study considered post length, readability, formal language use, and images as control variables based on prior literature [56,60]. For the regression models of 3 dependent variables, observations with studentized residuals that fall outside the mean of +3 or –3 were considered outliers and excluded. Then, regression analysis was performed to confirm the theory's applicability and answer the RQs.

Ethical Considerations

The study was approved by the National Tsing Hua University Research Ethics Committee (11308HT139). This study used publicly available data and processed and stored the data in compliance with the Personal Data Protection Act. As the data were publicly accessible, informed consent was not required.

Results

Overview

To answer RQ1 regarding user engagement trends with misinformation correction during the COVID-19 pandemic, we aggregated 3 years of data and analyzed various types of user engagement. For RQ2, we performed hierarchical multiple regression to examine the relationship between the identified attributes and user engagement in sharing and reactions. For RQ3, we considered the content in comments as additional information on the post, incorporating 6 categories of user engagement in commenting as independent variables in the models for RQ2.

Trends Related to Post Content and User Engagement During the COVID-19 Pandemic

First, this study organized the data to understand the data profile. We merged the MOHW and TFC datasets to obtain an overview of the situation. We categorized the data by themes based on a quarterly distribution, as shown in Figure 2, in which the distribution of misinformation correction themes aligned with the development trajectory of the COVID-19 pandemic. The peak of misinformation correction occurred in the second quarter of 2021, coinciding with Taiwan's level 3 alert and a vaccine shortage. The most common theme was health care, food safety, and home remedies, followed by understanding and controlling diseases and vaccine safety and policies.

Figure 2. Quarterly distribution of post counts by theme over 3 years.

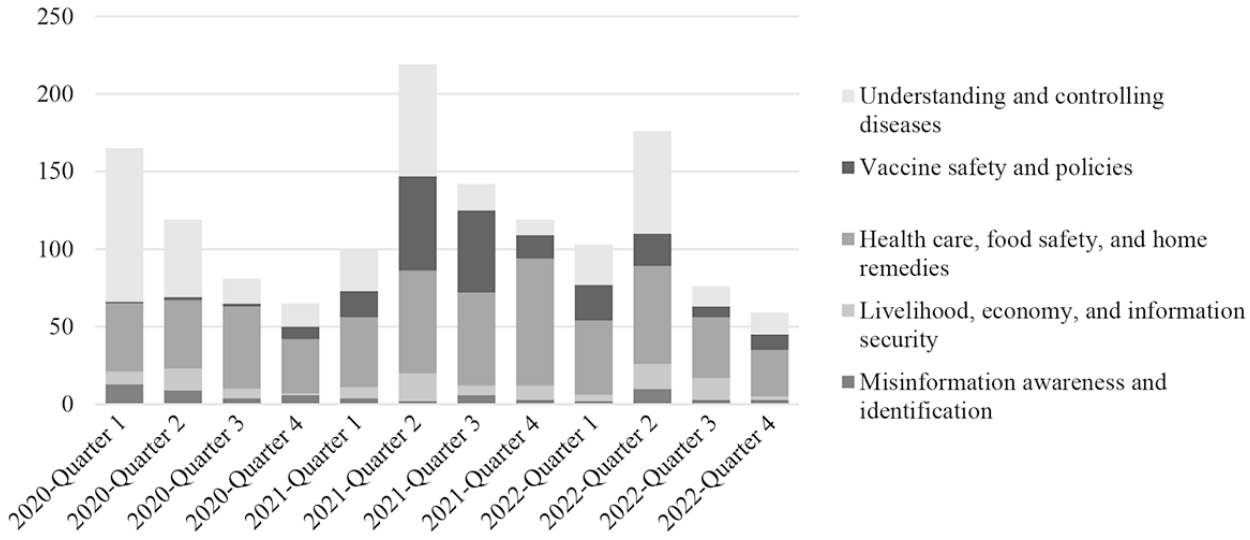

To answer RQ1, we conducted a 1-way ANOVA to analyze user engagement over 3 years, and the results are shown in Table 1. Through homogeneity tests, these counts were found to be heterogeneous, and Welch ANOVA and Games-Howell post hoc comparisons were performed. We found that sharing (Welch $F_{2,\ 805.887}$=31.082; P<.001) and reactions (Welch $F_{2,\ 802.838}$=40.620; P<.001) declined significantly over time, while comments (Welch $F_{2,\ 909.388}$=11.253; P<.001) remained high in the first 2 years, before dropping in the third year. Under the general trends, we further examined the 3-year change patterns of comments based on user engagement categories. The patterns of the knowledge-based engagement (Welch $F_{2,\ 931.301}$=19.470; P<.001) and 3 emotional engagements (positive emotional: Welch $F_{2,\ 741.713}$=22.636; P<.001; negative emotional: Welch $F_{2,\ 845.244}$=11.857; P<.001; neutral emotional: Welch $F_{2,\ 849.284}$=21.055; P<.001) in commenting were similar, remaining consistent in the first and second years but significantly declining in the third year. Critical engagement declined significantly over time (Welch $F_{2,\ 785.780}$=32.996; P<.001), while bias-based engagement peaked in the second year and then significantly declined in the third year (Welch $F_{2,\ 922.113}$=3.120; P=.045).

XSL•FO
RenderX

Table 1. Variation in user engagement over 3 years.

User engagement	2020 (n=430), mean (SD)	2021 (n=580), mean (SD)	2022 (n=414), mean (SD)	Tests of homogeneity of variances		Robust tests of equality of means		Mean difference between groups (Games-Howell post hoc tests)
				Levene statistic	*P* value	Welch statistic	*P* value	
Share	192.65 (496.70)	38.78 (104.67)	17.72 (70.42)	85.546	<.001	31.082	<.001	• 2020>2021, • 2020>2022, • 2021>2022
Reaction	1871.12 (3799.06)	664.91 (2001.51)	270.06 (960.27)	60.753	<.001	40.620	<.001	• 2020>2021, • 2020>2022, • 2021>2022
Comment	54.28 (131.03)	59.26 (302.90)	20.91 (92.98)	10.115	<.001	11.253	<.001	• 2020>2022, • 2021>2022
Knowledge-based engagement in commenting	7.04 (13.23)	5.42 (22.07)	2.16 (9.98)	11.194	<.001	19.470	<.001	• 2020>2022, • 2021>2022
Critical engagement in commenting	6.80 (13.83)	3.80 (13.32)	1.51 (4.39)	29.049	<.001	32.996	<.001	• 2020>2021, • 2020>2022, • 2021>2022
Bias-based engagement in commenting	11.44 (40.04)	25.65 (152.21)	9.00 (44.68)	13.045	<.001	3.120	.045	• 2021>2022
Positive emotional engagement in commenting	6.87 (20.27)	4.18 (17.30)	1.16 (4.67)	22.438	<.001	22.636	<.001	• 2020>2022, • 2021>2022
Negative emotional engagement in commenting	11.44 (33.40)	11.71 (63.18)	3.67 (17.16)	10.159	<.001	11.857	<.001	• 2020>2022, • 2021>2022
Neutral emotional engagement in commenting	5.00 (11.62)	3.22 (12.38)	1.27 (5.20)	16.948	<.001	21.055	<.001	• 2020>2022, • 2021>2022

We further explored the relative proportions among different categories of user engagement in commenting, as shown in Figure 3. In the first year of the COVID-19 pandemic, the engagement across categories was fairly balanced, with cognitive and emotional engagement each accounting for approximately half. The sum of knowledge-based and critical engagement comments exceeded those of bias-based engagement in 2020 in overall data, suggesting that the 2 beliefs could counterbalance each other. However, bias-based engagement became predominant by the second and third years on both platforms. In overall data, in 2020, bias-based engagement accounted for 23.55% (4920/20,893) of the comments; in 2021, it accounted for 47.51% (14,876/31,309); and in 2022, it accounted for 47.95% (3726/7770). Specifically, based on the MOHW data, in 2020, bias-based engagement was 32.97% (2776/8419), which rose to 52.39% (11,976/22,858) in 2021 and remained at 52.57% (2662/5064) in 2022. On the TFC data, bias-based engagement was 17.19% (2144/12,474) in 2020, 34.32% (2900/8451) in 2021, and 39.32% (1064/2706) in 2022. The discussions were dominated by bias-based engagement in the later stages of the COVID-19 pandemic. When comparing platforms, there was relatively more knowledge-based and critical engagement on the TFC platform, while the MOHW platform had relatively more positive emotional engagement in the first year.

XSL•FO
RenderX

Figure 3. Variation in the proportion of different types of user engagement in comments over 3 years. MOHW: Ministry of Health and Welfare; TFC: Taiwan Fact-Checking Center.

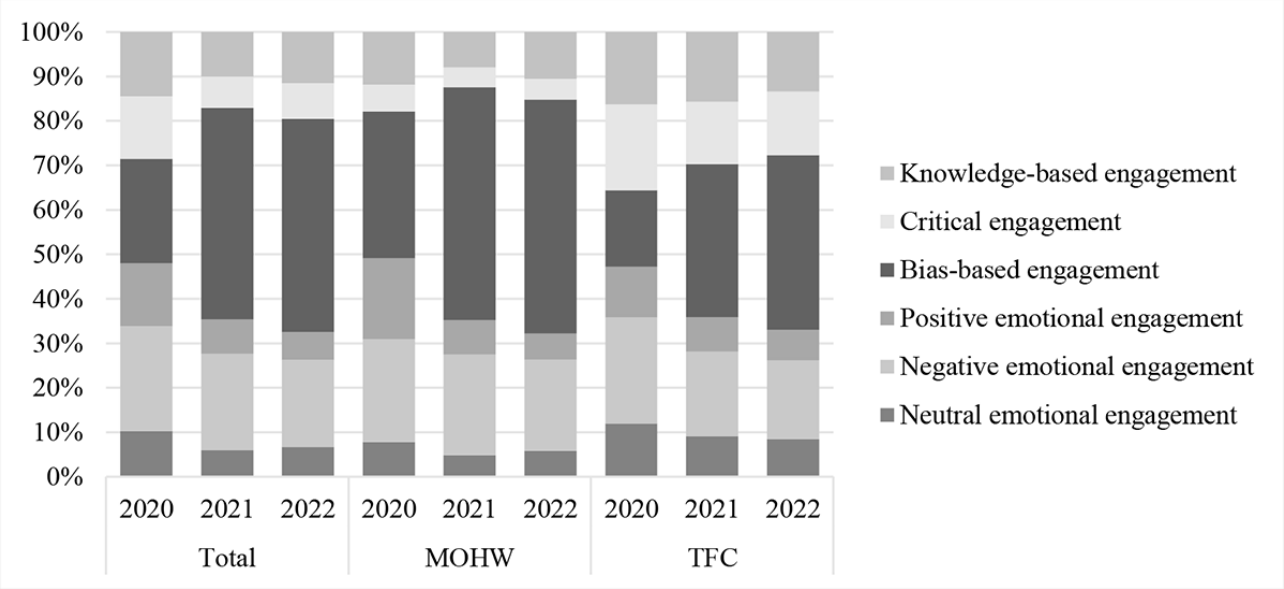

Objective metrics such as shares, reactions, and comments for user engagement showed that interest in health-related misinformation significantly decreased in the third year. However, regarding the number of deaths due to COVID-19, the pandemic was most severe in Taiwan in its third year, with 7 deaths in 2020, 843 in 2021, and 14,636 in 2022 [61].

Post Attributes and User Engagement

In the annotated dataset, of the 1424 posts, 261 (18.33%) contained risk, 868 (60.96%) contained awareness, 527 (37.01%) contained value, 163 (11.45%) contained numeric information, and 768 (53.93%) contained authority. In addition, of the 1424 posts, 202 (14.19%) did not target specific misinformation, 505 (35.46%) provided a simple correction, and 717 (50.35%) offered a detailed correction. Subsequently, concerning RQ2, we focused on the relationship between post attributes and 3 user engagement metrics: shares, reactions, and comments, as shown in Table 2. Model 1 controlled for background variables including post length (transformed using the natural logarithm), official status (converted to a dummy variable, where 1 represents the official platform [MOHW] and 0 represents the nonofficial platform [TFC]), readability, formal tone (where 1 indicates formal and 0 indicates informal), and image (where 1 indicates the attachment of at least 1 image and 0 indicates none). The dependent variables were also transformed using the natural logarithm. These background variables contributed significantly to the regression models for shares (R^2=0.334; $F_{5,1412}$=141.380; P<.001), reactions (R^2=0.407; $F_{5,1404}$=192.994; P<.001), and comments (R^2=0.328; $F_{5,1414}$=137.794; P<.001). All background variables were predictive of these engagement metrics, except for post length, which did not significantly predict comment count.

XSL•FO
RenderX

Table 2. Hierarchical multiple regression analysis predicting shares, reactions, and comments (models 1 and 2).

Predictors	Share[a]				Reaction[a]				Comment[a]			
	B[b] (SE[c])	β[d]	*P* value	VIF[e]	B (SE)	β	*P* value	VIF	B (SE)	β	*P* value	VIF
Model 1												
Constant	3.82 (0.36)	—[g]	<.001	—	6.19 (0.28)	—	<.001	—	1.76 (0.36)	—	<.001	—
Post length[a]	0.11 (0.05)	.07	.03	2.08	0.11 (0.04)	.08	.005	2.07	0.09 (0.05)	.06	.08	2.09
Official platform[f]	2.95 (0.14)	.49	<.001	1.15	2.80 (0.11)	.56	<.001	1.15	3.00 (0.14)	.51	<.001	1.15
Readability	0.39 (0.03)	.27	<.001	1.22	0.30 (0.03)	.25	<.001	1.21	0.16 (0.03)	.11	<.001	1.22
Formal tone[f]	0.53 (0.11)	.15	<.001	1.95	0.54 (0.08)	.19	<.001	1.94	0.77 (0.10)	.22	<.001	1.95
Image[f]	0.43 (0.13)	.07	.001	1.09	0.24 (0.10)	.05	.02	1.09	0.27 (0.13)	.05	.04	1.09
Model 2												
Constant	4.40 (0.43)	—	<.001	—	7.00 (0.34)	—	<.001	—	2.56 (0.43)	—	<.001	—
Post length[a]	–0.04 (0.06)	–.02	.56	3.21	–0.05 (0.05)	–.04	.29	3.20	–0.07 (0.06)	–.04	0.26	3.21
Official platform	2.88 (0.14)	.48	<.001	1.26	2.84 (0.11)	.57	<.001	1.26	3.11 (0.14)	.52	<.001	1.26
Readability	0.40 (0.03)	.27	<.001	1.26	0.31 (0.03)	.26	<.001	1.26	0.17 (0.03)	.12	<.001	1.27
Formal tone[f]	0.45 (0.11)	.13	<.001	2.08	0.42 (0.08)	.15	<.001	2.07	0.63 (0.11)	.18	<.001	2.09
Image[f]	0.21 (0.14)	.04	.13	1.25	–0.04 (0.11)	–.01	.73	1.26	0.00 (0.14)	.00	.99	1.25
Risk[f]	0.30 (0.09)	.07	.001	1.11	0.10 (0.07)	.03	.15	1.11	–0.07 (0.09)	–.02	.44	1.11
Awareness[f]	0.30 (0.07)	.09	<.001	1.14	0.20 (0.06)	.07	.001	1.14	0.21 (0.07)	.06	.005	1.14
Value[f]	0.14 (0.07)	.04	.051	1.09	0.08 (0.06)	.03	.17	1.08	–0.08 (0.07)	–.02	.28	1.09
Numeric[f]	0.15 (0.11)	.03	.18	1.10	0.13 (0.09)	.03	.13	1.10	0.27 (0.11)	.06	.02	1.10
Authority[f]	0.10 (0.10)	.03	.31	2.02	–0.02 (0.08)	–.01	.74	2.04	–0.02 (0.10)	–.01	.84	2.02
Facts	0.30 (0.08)	.14	<.001	2.71	0.44 (0.06)	.24	<.001	2.73	0.42 (0.08)	.19	<.001	2.72

[a]Transformed using the natural logarithm.

[b]Unstandardized regression coefficient.

[c]SE for unstandardized regression coefficient.

[d]Standardized regression coefficient.

[e]VIF: variance inflation factor.

[f]Converted into a dummy variable.

[g]Not applicable.

Model 2 incorporated gist and verbatim representations. The R^2 changes for the 3 models were significant. In the model focused on predicting shares (R^2=0.353; $F_{11,1406}$=69.808; P<.001; Δ R^2=0.020), risk (β=.07; P=.001), awareness (β=.09; P<.001), and facts (β=.14; P<.001) were significant predictors. When predicting reactions (R^2=0.435; $F_{11,1398}$=97.292; P<.001;

XSL•FO
RenderX

Δ R^2=0.028), awareness (β=.07; *P*=.001) and facts (β=.24; *P*<.001) were found to be significant factors. Similarly, when predicting comments (R^2=0.351; $F_{11,1408}$=69.158; *P*<.001; Δ R^2=0.023), awareness (β=.06; *P*=.005), numeric (β=.06; *P*=.02), and facts (β=.19; *P*<.001) played significant roles. This finding suggests that 2 of the gist representations, risk and awareness, can enhance post endorsement and sharing, but only awareness was predictive of reactions and comments. Moreover, facts positively influenced the 3 metrics in the verbatim dimension, indicating that detailed clarifications enhanced shares, reactions, and comments.

Post Attributes and User Engagement: Considering Comment Content

For RQ3, we incorporated comments generated by 6 categories of user engagement into model 3, as presented in Table 3. To ensure the validity of our findings, we used the variance inflation factor to diagnose multicollinearity, which was within acceptable limits (<10). In the model predicting shares (R^2=0.668; $F_{17,1400}$=165.904; *P*<.001; Δ R^2=0.315), the 3 gist representations (risk: β=.08; *P*<.001, awareness: β=.08; *P*<.001, and value: β=.06; *P*<.001) were significant predictors, while facts (β=.02; *P*=.36) no longer predicted it. Then, when considering user engagement in commenting, in the cognitive aspect, only bias-based engagement (β=–.11; *P*=001) negatively predicted shares, and knowledge-based and critical engagement positively predicted shares (knowledge-based: β=.28; *P*<.001 and critical: β=.15; *P*<.001). In the emotional aspect, positive and negative emotional engagement both predicted shares, while neutral did not predict it (positive: β=.29; *P*<.001, negative: β=.09; *P*=.02, and neutral: β=.05; *P*=.15). Similarly, in the model predicting reactions (R^2=0.772; $F_{17,1392}$=277.743; *P*<.001; Δ R^2=0.337), the 3 gist representations (risk: β=.04; *P*=.007, awareness: β=.06; *P*<.001, and value: β=.05; *P*<.001) significantly predicted the reaction count, and facts (β=.12; *P*<.001) also predicted it. All comment categories, except for bias-based engagement (β=–.04; *P*=.15), significantly predicted reactions (knowledge-based: β=.19; *P*<.001, critical: β=.14; *P*<.001, positive: β=.26; *P*<.001, negative: β=.15; *P*=<.001, and neutral: β=.08; *P*=.002).

XSL•FO
RenderX

Table 3. Hierarchical multiple regression analysis predicting share and reaction count (model 3).

Predictors	Share[a]				Reaction[a]			
	B[b] (SE[c])	β[d]	*P* value	VIF[e]	B (SE)	β	*P* value	VIF
Model 3								
Constant	2.34 (0.32)	—[f]	<.001		5.26 (0.22)	—	<.001	—
Post length[a]	0.08 (0.05)	.05	.07	3.27	0.05 (0.03)	.03	.13	3.27
Official platform[g]	0.72 (0.14)	.12	<.001	2.18	0.91 (0.09)	.18	<.001	2.18
Readability	0.24 (0.03)	.17	<.001	1.32	0.18 (0.02)	.15	<.001	1.31
Formal tone[g]	0.02 (0.08)	.01	.76	2.14	0.03 (0.05)	.01	.58	2.13
Image[g]	0.23 (0.10)	.04	.02	1.25	–0.02 (0.07)	.00	.79	1.26
Risk[g]	0.31 (0.07)	.08	<.001	1.11	0.12 (0.05)	.04	.007	1.11
Awareness[g]	0.25 (0.05)	.08	<.001	1.16	0.15 (0.04)	.06	<.001	1.15
Value[g]	0.19 (0.05)	.06	<.001	1.11	0.15 (0.04)	.05	<.001	1.10
Numeric[g]	0.01 (0.08)	.00	.92	1.11	0.00 (0.06)	.00	.93	1.11
Authority[g]	0.07 (0.07)	.02	.28	2.02	–0.04 (0.05)	–.01	.46	2.05
Facts	0.05 (0.06)	.02	.36	2.78	0.22 (0.04)	.12	<.001	2.80
Knowledge-based engagement[a]	0.42 (0.04)	.28	<.001	3.67	0.24 (0.03)	.19	<.001	3.68
Critical engagement[a]	0.23 (0.05)	.15	<.001	3.77	0.18 (0.03)	.14	<.001	3.78
Bias-based engagement[a]	–0.13 (0.04)	–.11	.001	4.29	–0.04 (0.03)	–.04	.15	4.26
Positive emotional engagement[a]	0.45 (0.05)	.29	<.001	4.36	0.34 (0.03)	.26	<.001	4.35
Negative emotional engagement[a]	0.12 (0.05)	.09	.02	6.93	0.17 (0.04)	.15	<.001	6.84
Neutral emotional engagement[a]	0.08 (0.05)	.05	.15	4.26	0.11 (0.04)	.08	.002	4.27

[a]Transformed using the natural logarithm.
[b]Unstandardized regression coefficient.
[c]SE for unstandardized regression coefficient.
[d]Standardized regression coefficient.
[e]VIF: variance inflation factor.
[f]Not applicable.
[g]Converted into a dummy variable.

Discussion

Principal Findings

First, in response to RQ1, all metrics unsurprisingly decreased, and it is noteworthy that the proportion of bias-based engagement in comments increased on both platforms. At the outbreak of the COVID-19 pandemic, despite the potential chaos in information, the comment categories remained relatively balanced and were not dominated by any category. The increase in bias-based engagement in the second year was likely due to the first COVID-19 community transmission in Taiwan and the rise of COVID-19 vaccine–related issues. The result aligns with findings from previous studies, which show that comments often contain inaccuracies and are predominantly made by uninformed users [22,52].

Then, we compared our findings with similar studies that explored the relationship between post attributes and user engagement. Regarding background variables, according to model 1, readability aligned with previous research findings [56], which explained why readability influences social media engagement through fluency. Posts using a formal tone also showed higher user engagement, consistent with the literature [60], in which formality was a heuristic for credibility and importance. However, we did not control for the topics or timing

XSL•FO
RenderX

associated with using a formal tone; more severe events may be more likely to use a formal tone. Posts containing images also positively impacted user engagement, a finding supported and explained in the literature by the concept of media type richness [59] or an approach-motivated perspective [44].

Drawing on the FTT and literature on misinformation, we identified 3 key gist representations and tested their effectiveness. The theory suggests that a message with a clear core and causal explanation is more likely to resonate with people, leading to behaviors such as sharing [30]. Moreover, this study suggests that gist representations provide cues that enhance the probability of users finding the respective information relatable or practical. Risk may involve terms such as danger or harm. Awareness often describes tips and tricks and is relatively neutral. Value sometimes refers to reassurance and tends to have more positive terms. However, only awareness is a significant predictor in each model. When considering the content of comments, all gist representations significantly predict reactions, indicating that risk and value retain their value within social media posts that include comments. In addition, value was the only gist of predicting comments negatively, although not significantly. Whether value can further reduce bias-based engagement remains to be examined in future studies. Overall, providing gist information increased shares, reactions, and comments, consistent with previous research showing that gist information enhances article sharing [14]. Awareness of misinformation is the most crucial factor in this study, as a previous study has emphasized awareness of misinformation as the foundational concept of citizen resilience to misinformation [35].

Regarding the verbatim dimension, we found that numeric information cannot significantly predict shares and reactions but can predict comments. Statistics also failed to predict sharing in another study based on the FTT [14]. According to discussions related to the FTT, individuals' numerical abilities influence the numeric factor [38], and users might engage in discussions based on different perceptions of numeric information. However, statistics as a factor of source transparency was one of the predictors of shares and comments, suggesting that numerical information in news articles may increase the audience's perceived accuracy [9]. A previous study indicated that participants engaged more with corrections even without statistical information due to the simplicity of the material, which made it easier to assess credibility. Then, they put less cognitive effort when the information included statistical evidence [20]. We offered a different perspective through FTT, suggesting that numeric information did not significantly affect user engagement in sharing and reacting, as these forms of engagement relied more on top-down meaning. However, numeric information increased commenting, which might rely on detailed information and higher cognitive engagement. Furthermore, this study identified authority as verbatim because the names of experts or institutes as references were detailed. Nevertheless, the attribute still conveyed a sense of credibility. The concept was similar to source transparency but with slight differences, and a previous study showed that using official reports increased shares and comments, and references to laws and news articles also increased comments [9]. In this study, this variable did not significantly predict any metrics. This could also be due to both platforms being trustworthy sources. Finally, facts significantly predicted all 3 metrics in model 2, indicating that posts with detailed and complete content generated higher user engagement, which is consistent with past experimental results [12,13,18]. Detailed corrections might contain alternative interpretations and context information for understanding and discussion, generating user engagement.

Finally, we considered that readers might refer to comment content to decide whether to share or react in model 3. An increase in knowledge-based and critical engagement comments, which reflect scientific beliefs, led to higher shares and reactions. Conversely, comments reflecting bias-based engagement negatively predicted shares, suggesting that these comments, much like uncivil comments [47], may diminish the quality of discussion threads and reduce share counts. Comments with positive and negative emotional engagement both predicted shares and reactions. Although comment categories could predict shares and reactions, all 3 gist representations still play an important role, indicating that gist communication can still be effective.

In summary, this study defined post attributes based on the FTT and considered a more diversified array of user engagement types to explore their relationship with post attributes. The main theoretical contribution of our study is that it presents a framework applied to misinformation issues validated using real social media data. It examined user engagement in commenting to better reflect the content users are most likely to encounter. The main practical contribution is the introduction of gist representations—risk, awareness, and value—that serve as simple but important cues when publishing correction messages, helping users grasp and address misinformation issues and encouraging content sharing.

Implications, Limitations, and Future Directions

FTT emphasizes the mutually supportive relationship between our cognition and social values and recommends using reasoned approaches to guide decision-making and emotional responses, rather than emotional appeals, to counter misinformation [17]. Building on this foundation, we identified 3 rational gist representations that connect with social values. Among these, awareness of misinformation proved to be a highly effective gist when comment content was not considered. In our data, this concept included elements such as increasing awareness of misinformation and fraud, encouraging verification of sources, promoting cautious sharing, and identifying characteristics and sources of misinformation. Mentioning the risks associated with misinformation can increase share counts. The data for this study referenced health risks, social panic, legal risks, and information security risks. Mentioning the value of health promotion is effective when considering comment content. The data for this study referenced health promotion and health literacy, a positive attitude toward the COVID-19 pandemic prevention, media or eHealth literacy, public trust, psychological well-being, and social confidence. These attributes serve as guidelines for designing content in misinformation correction and science communication.

XSL•FO
RenderX

This study had some limitations. First, this study examined data from 3 years during the COVID-19 pandemic, during which attention to health misinformation was heightened. There were also some health crises before and after the COVID-19 pandemic. Whether inferences based on the results of this study remain valid requires further examination. Second, we used GPT-3.5 to code comments efficiently, followed by a manual review. Cultural characteristics may have caused GPT-3.5 to misinterpret specific comments. For example, the colors blue and green sometimes represent political parties, and GPT-3.5 may not always accurately interpret this from the context, leading to some classification inaccuracies. Nevertheless, GPT-3.5 offers several advantages over previous machine learning methods, such as cross-language capabilities and emoji interpretation. Third, authority, as one of the independent variables in this study, has implications for credibility in some types of research. In the heuristic-systematic model, it is viewed as having an intuitive role, and in the elaboration likelihood model, it is considered a peripheral cue, suggesting the credibility of a message. We cannot confirm whether the dual roles cause construct confusion or whether the platform symbolizes knowledge authority. Authority did not significantly predict the outcomes in models 2 and 3; however, this does not imply that it is unimportant. Finally, we considered whether there was an attached image but did not examine its content or whether the images contained gist messages. Images attract user attention more quickly and should be considered in future research.

In addition to addressing the abovementioned limitations, future research could consider simplifying the 6 comment categories we used. For instance, knowledge-based and critical engagement could be combined, as well as positive and neutral emotional engagement. In addition, some comments directly responded to the content design of certain corrections. We observed that some posts were criticized for lacking focus, while others were praised for being practical. If these comments can be accurately extracted and systematically analyzed, it would also help assess whether this type of gist communication is well received. We found that posts containing numerical information, such as vaccine adverse reaction rates, received more comments, as users may interpret and engage more with rates and numbers. This area requires more systematic analysis and effective framing of numeric information in correction design to prevent misinterpretation on social media, enhancing the application of FTT. Finally, this study used regression analysis to explore the impact of 6 attributes on metrics; however, a single post often contains multiple attributes, and it may be possible to identify an optimal combination. Future studies could further explore this area.

Conclusions

This study enriches the theoretical understanding of the relationship between user engagement and the effectiveness of web-based communication strategies in conveying corrections for health misinformation. It provides evidence-based web content for delivering health misinformation corrections, and FTT explains user engagement with specific post attributes. The theoretical contribution strengthens the link between FTT and multiple engagement metrics beyond sharing and discusses post attributes in terms of gist and verbatim dimensions. In addition, by incorporating comment content, gist communication increases shares and reactions effectively. Finally, these findings provide a foundation for designing more effective content strategies to combat health misinformation, ultimately contributing to more resilient public health communication.

Acknowledgments

The authors are grateful for the funding support from Academia Sinica in Taiwan (AS-HLGC-111-03). The authors acknowledge the contributions of the 3 human annotators. This study used generative an artificial intelligence tool, GPT-3.5, for large language model–based annotation of comment data.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of health misinformation correction, coding schemes, prompts for GPT-3.5, and generated output examples.
[DOCX File , 33 KB - jmir_v27i1e65631_app1.docx]

References

1. Ecker UK, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, et al. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol 2022 Jan 12;1(1):13-29. [doi: 10.1038/S44159-021-00006-Y]
2. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 2020 Dec;4(12):1285-1293. [doi: 10.1038/s41562-020-00994-6] [Medline: 33122812]

XSL•FO
RenderX

3. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
4. Domenico GD, Sit J, Ishizaka A, Nunan D. Fake news, social media and marketing: a systematic review. J Bus Res 2021 Jan;124:329-341. [doi: 10.1016/j.jbusres.2020.11.037]
5. Yoon HY, You KH, Kwon JH, Kim JS, Rha SY, Chang YJ, et al. Understanding the social mechanism of cancer misinformation spread on YouTube and lessons learned: infodemiological study. J Med Internet Res 2022 Nov 14;24(11):e39571 [FREE Full text] [doi: 10.2196/39571] [Medline: 36374534]
6. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
7. Chu WM, Shieh GJ, Wu SL, Sheu WH. Use of Facebook by academic medical centers in Taiwan during the COVID-19 pandemic: observational study. J Med Internet Res 2020 Nov 20;22(11):e21501 [FREE Full text] [doi: 10.2196/21501] [Medline: 33119536]
8. Xue H, Gong X, Stevens H. COVID-19 vaccine fact-checking posts on Facebook: observational study. J Med Internet Res 2022 Jun 21;24(6):e38423 [FREE Full text] [doi: 10.2196/38423] [Medline: 35671409]
9. Kim HS, Suh YJ, Kim EM, Chong E, Hong H, Song B, et al. Fact-checking and audience engagement: a study of content analysis and audience behavioral data of fact-checking coverage from news media. Digital Journalism 2022 Feb 08;10(5):781-800. [doi: 10.1080/21670811.2021.2006073]
10. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 2012 Dec 17;13(3):106-131. [doi: 10.1177/1529100612451018] [Medline: 26173286]
11. Koh HK, Rudd RE. The arc of health literacy. JAMA 2015;314(12):1225-1226. [doi: 10.1001/jama.2015.9978] [Medline: 26247161]
12. van der Meer TG, Jin Y. Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source. Health Commun 2020 May;35(5):560-575. [doi: 10.1080/10410236.2019.1573295] [Medline: 30761917]
13. Zeng HK, Lo SY, Li SS. Credibility of misinformation source moderates the effectiveness of corrective messages on social media. Public Underst Sci 2024 Jul 31;33(5):587-603. [doi: 10.1177/09636625231215979] [Medline: 38160402]
14. Broniatowski DA, Hilyard KM, Dredze M. Effective vaccine communication during the disneyland measles outbreak. Vaccine 2016 Jun 14;34(28):3225-3228 [FREE Full text] [doi: 10.1016/j.vaccine.2016.04.044] [Medline: 27179915]
15. Reyna VF, Brainerd CJ. Fuzzy-trace theory: an interim synthesis. Learn Individ Differ 1995 Jan;7(1):1-75. [doi: 10.1016/1041-6080(95)90031-4]
16. Blalock SJ, DeVellis RF, Chewning B, Sleath BL, Reyna VF. Gist and verbatim communication concerning medication risks/benefits. Patient Educ Couns 2016 Jun;99(6):988-994 [FREE Full text] [doi: 10.1016/j.pec.2015.12.001] [Medline: 26786667]
17. Reyna VF. A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. Proc Natl Acad Sci U S A 2021 Apr 13;118(15):e1912441117 [FREE Full text] [doi: 10.1073/pnas.1912441117] [Medline: 32312815]
18. Martel C, Mosleh M, Rand DG. You're definitely wrong, maybe: correction style has minimal effect on corrections of misinformation online. Media Commun 2021 Feb 03;9(1):120-133. [doi: 10.17645/mac.v9i1.3519]
19. Wang Y. Debunking misinformation about genetically modified food safety on social media: can heuristic cues mitigate biased assimilation? Sci Commun 2021 Jun 18;43(4):460-485. [doi: 10.1177/10755470211022024]
20. Song Y, Wang S, Xu Q. Fighting misinformation on social media: effects of evidence type and presentation mode. Health Educ Res 2022 May 24;37(3):185-198 [FREE Full text] [doi: 10.1093/her/cyac011] [Medline: 35582892]
21. Sui Y, Zhang B. Determinants of the perceived credibility of rebuttals concerning health misinformation. Int J Environ Res Public Health 2021 Feb 02;18(3):1345 [FREE Full text] [doi: 10.3390/ijerph18031345] [Medline: 33540869]
22. Bizzotto N, Schulz PJ, de Bruijn GJ. The "Loci" of misinformation and its correction in peer- and expert-led online communities for mental health: content analysis. J Med Internet Res 2023 Sep 18;25:e44656 [FREE Full text] [doi: 10.2196/44656] [Medline: 37721800]
23. McGuire WJ. Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. J Abnorm Soc Psychol 1961 Sep;63(2):326-332. [doi: 10.1037/H0048344]
24. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: Berkowitz L, editor. Advances in Experimental Social Psychology. New York, NY: Academic Press; 1986:123-205.
25. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980 Nov;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
26. Edelson SM, Reyna VF. Who makes the decision, how, and why: a fuzzy-trace theory approach. Med Decis Making 2024 Aug 26;44(6):614-616. [doi: 10.1177/0272989X241263818] [Medline: 39056326]
27. Broniatowski DA, Reyna VF. To illuminate and motivate: a fuzzy-trace model of the spread of information online. Comput Math Organ Theory 2020 Dec;26:431-464 [FREE Full text] [doi: 10.1007/s10588-019-09297-2] [Medline: 33737859]

XSL•FO
RenderX

28. Reyna VF, Brainerd CJ. Dual processes in decision making and developmental neuroscience: a fuzzy-trace model. Dev Rev 2011 Sep;31(2-3):180-206 [FREE Full text] [doi: 10.1016/j.dr.2011.07.004] [Medline: 22096268]
29. Reyna VF, Edelson S, Hayes B, Garavito D. Supporting health and medical decision making: findings and insights from fuzzy-trace theory. Med Decis Making 2022 Aug 23;42(6):741-754 [FREE Full text] [doi: 10.1177/0272989X221105473] [Medline: 35735225]
30. Broniatowski DA, Hosseini P, Porter EV, Wood TJ. The role of mental representation in sharing misinformation online. J Exp Psychol Appl 2024 Jul 25. [doi: 10.1037/xap0000517] [Medline: 39052364]
31. Ecker UK, O'Reilly Z, Reid JS, Chang EP. The effectiveness of short-format refutational fact-checks. Br J Psychol 2020 Feb;111(1):36-54 [FREE Full text] [doi: 10.1111/bjop.12383] [Medline: 30825195]
32. Blalock SJ, Reyna VF. Using fuzzy-trace theory to understand and improve health judgments, decisions, and behaviors: a literature review. Health Psychol 2016 Aug;35(8):781-792 [FREE Full text] [doi: 10.1037/hea0000384] [Medline: 27505197]
33. Reyna VF. Risk perception and communication in vaccination decisions: a fuzzy-trace theory approach. Vaccine 2012 May 28;30(25):3790-3797 [FREE Full text] [doi: 10.1016/j.vaccine.2011.11.070] [Medline: 22133507]
34. Dryhurst S, Schneider CR, Kerr J, Freeman AL, Recchia G, van der Bles AM, et al. Risk perceptions of COVID-19 around the world. J Risk Res 2020 May 05;23(7-8):994-1006. [doi: 10.1080/13669877.2020.1758193]
35. Rodríguez-Pérez C, Canel MJ. Exploring European citizens' resilience to misinformation: media legitimacy and media trust as predictive variables. Media Commun 2023 Apr 28;11(2):30-41. [doi: 10.17645/mac.v11i2.6317]
36. Rimal RN, Lapinski MK. Why health communication is important in public health. Bull World Health Organ 2009 Apr;87(4):247 [FREE Full text] [doi: 10.2471/blt.08.056713] [Medline: 19551226]
37. Pan W, Liu D, Fang J. An examination of factors contributing to the acceptance of online health misinformation. Front Psychol 2021;12:630268 [FREE Full text] [doi: 10.3389/fpsyg.2021.630268] [Medline: 33732192]
38. Reyna VF, Brainerd CJ. Numeracy, gist, literal thinking and the value of nothing in decision making. Nat Rev Psychol 2023 May 18;2(7):1-19 [FREE Full text] [doi: 10.1038/s44159-023-00188-7] [Medline: 37361389]
39. Looi JC, Allison S, Bastiampillai T, Maguire PA. Clinical update on managing media exposure and misinformation during COVID-19: recommendations for governments and healthcare professionals. Australas Psychiatry 2021 Feb 08;29(1):22-25. [doi: 10.1177/1039856220963947] [Medline: 33031704]
40. Bautista JR, Zhang Y, Gwizdka J. Healthcare professionals' acts of correcting health misinformation on social media. Int J Med Inform 2021 Apr;148:104375. [doi: 10.1016/j.ijmedinf.2021.104375] [Medline: 33461008]
41. Chan MP, Jones CR, Hall Jamieson K, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychol Sci 2017 Nov;28(11):1531-1546 [FREE Full text] [doi: 10.1177/0956797617714579] [Medline: 28895452]
42. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
43. Kim C, Yang SU. Like, comment, and share on Facebook: how each behavior differs from the other. Public Relat Rev 2017 Jun;43(2):441-449. [doi: 10.1016/j.pubrev.2017.02.006]
44. Rus HM, Cameron LD. Health communication in social media: message features predicting user engagement on diabetes-related Facebook pages. Ann Behav Med 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
45. Jiang S, Tay J, Ngien A, Basnyat I. Social media health promotion and audience engagement: the roles of information dissemination, organization-audience interaction, and action confidence building. Health Commun 2024 Apr 25;39(1):4-14. [doi: 10.1080/10410236.2022.2150809] [Medline: 36426852]
46. Bhattacharya S, Srinivasan P, Polgreen P. Social media engagement analysis of U.S. Federal health agencies on Facebook. BMC Med Inform Decis Mak 2017 Apr 21;17(1):49 [FREE Full text] [doi: 10.1186/s12911-017-0447-z] [Medline: 28431582]
47. Weber P, Prochazka F, Schweiger W. Why user comments affect the perceived quality of journalistic content: the role of judgment processes. J Media Psychol 2019 Jan;31(1):24-34. [doi: 10.1027/1864-1105/A000217]
48. Mehmet M, Heffernan T, Algie J, Forouhandeh B. Harnessing social listening to explore consumer cognitive bias: implications for upstream social marketing. J Soc Mark 2021 Oct 02;11(4):575-596. [doi: 10.1108/jsocm-03-2021-0067]
49. Brugnoli E, Cinelli M, Quattrociocchi W, Scala A. Recursive patterns in online echo chambers. Sci Rep 2019 Dec 27;9(1):20118 [FREE Full text] [doi: 10.1038/s41598-019-56191-7] [Medline: 31882591]
50. Colliander J. "This is fake news": investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. Comput Human Behav 2019 Aug;97:202-215. [doi: 10.1016/j.chb.2019.03.032]
51. Hong S, Cameron GT. Will comments change your opinion? The persuasion effects of online comments and heuristic cues in crisis communication. J Conting Crisis Manag 2017 Nov 24;26(1):173-182. [doi: 10.1111/1468-5973.12215]
52. Bode L, Vraga EK. Correction experiences on social media during COVID-19. Soc Media Soc 2021 Apr 12;7(2):205630512110088. [doi: 10.1177/20563051211008829]
53. Russell JA, Barrett LF. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J Pers Soc Psychol 1999 May;76(5):805-819. [doi: 10.1037//0022-3514.76.5.805] [Medline: 10353204]

XSL•FO
RenderX

54. Chinese 8,000-word list. National Chinese Quiz Promotion Working Committee. URL: https://tocfl.edu.tw/tocfl/index.php/exam/download [accessed 2025-01-21]
55. Third level seven-word list. National Academy for Educational Research. URL: https://coct.naer.edu.tw/file/files/TBCL%E8%A9%9E%E8%AA%9E%E8%A1%A8_14420.pdf [accessed 2025-01-21]
56. Pancer E, Chandler V, Poole M, Noseworthy TJ. How readability shapes social media engagement. J Consum Psychol 2018 Nov 07;29(2):262-270. [doi: 10.1002/jcpy.1073]
57. Dale E, Chall JS. A formula for predicting readability: instructions. Educ Res Bull 1948;27(2):37-54 [FREE Full text]
58. Feng K, Huang L, Wang K, Wei W, Zhang R. Prompt-based learning framework for zero-shot cross-lingual text classification. Eng Appl Artif Intell 2024 Jul;133:108481. [doi: 10.1016/j.engappai.2024.108481]
59. Sabate F, Berbegal-Mirabent J, Cañabate A, Lebherz PR. Factors influencing popularity of branded content in Facebook fan pages. Eur Manag J 2014 Dec;32(6):1001-1011. [doi: 10.1016/j.emj.2014.05.001]
60. Linos E, Lasky-Fink J, Larkin C, Moore L, Kirkman E. The formality effect. Nat Hum Behav 2024 Feb 23;8(2):300-310. [doi: 10.1038/s41562-023-01761-z] [Medline: 37996499]
61. Statistics table of area, age, sex of death cases - COVID-19 (by month of death). Center for Diseases Control, Ministry of Health and Welfare. URL: https://data.cdc.gov.tw/en/dataset/death-date-statistics-cases-19cov-2 [accessed 2025-01-21]

Abbreviations

FTT: fuzzy-trace theory
MOHW: Ministry of Health and Welfare
RQ: research question
TFC: Taiwan Fact-Checking Center

Edited by A Mavragani; submitted 21.08.24; peer-reviewed by JE Ibarra-Esquer, SB Park, L Shang; comments to author 01.10.24; revised version received 05.11.24; accepted 27.11.24; published 23.01.25.

Please cite as:
Kuo HY, Chen SY
Predicting User Engagement in Health Misinformation Correction on Social Media Platforms in Taiwan: Content Analysis and Text Mining Study
J Med Internet Res 2025;27:e65631
URL: https://www.jmir.org/2025/1/e65631
doi:10.2196/65631
PMID:39847418

©Hsin-Yu Kuo, Su-Yen Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Forecasting the Incidence of Mumps Based on the Baidu Index and Environmental Data in Yunnan, China: Deep Learning Model Study

Xin Xiong[1*], BS; Linghui Xiang[1*], BS; Litao Chang[2], BS; Irene XY Wu[1,3], PhD; Shuzhen Deng[2], MS

[1]Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China

[2]Department of School Health, Yunnan Provincial Center for Disease Control and Prevention, Kunming, China

[3]Hunan Provincial Key Laboratory of Clinical Epidemiology, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Shuzhen Deng, MS
Department of School Health
Yunnan Provincial Center for Disease Control and Prevention
No.1177 Xianghe Street, Luolong Street, Chenggong District
Kunming, 650500
China
Phone: 86 15096624164
Email: 461447164@qq.com

Abstract

Background: Mumps is a viral respiratory disease characterized by facial swelling and transmitted through respiratory secretions. Despite the availability of an effective vaccine, mumps outbreaks have reemerged globally, including in China, where it remains a significant public health issue. In Yunnan province, China, the incidence of mumps has fluctuated markedly and is higher than that in mainland China, underscoring the need for improved outbreak prediction methods. Traditional surveillance methods, however, may not be sufficient for timely and accurate outbreak prediction.

Objective: Our study aims to leverage the Baidu search index, representing search volumes from China's most popular search engine, along with environmental data to develop a predictive model for mumps incidence in Yunnan province.

Methods: We analyzed mumps incidence in Yunnan Province from 2014 to 2023, and used time series data, including mumps incidence, Baidu search index, and environmental factors, from 2016 to 2023, to develop predictive models based on long short-term memory networks. Feature selection was conducted using Pearson correlation analysis, and lag correlations were explored through a distributed nonlinear lag model (DNLM). We constructed four models with different combinations of predictors: (1) model BE, combining the Baidu index and environmental factors data; (2) model IB, combining mumps incidence and Baidu index data; (3) model IE, combining mumps incidence and environmental factors; and (4) model IBE, integrating all 3 data sources.

Results: The incidence of mumps in Yunnan showed significant variability, peaking at 37.5 per 100,000 population in 2019. From 2014 to 2023, the proportion of female patients ranged from 41.3% in 2015 to 45.7% in 2020, consistently lower than that of male patients. After excluding variables with a Pearson correlation coefficient of <0.10 or *P* values of <.05, we included 3 Baidu index search term groups (disease name, symptoms, and treatment) and 6 environmental factors (maximum temperature, minimum temperature, sulfur dioxide, carbon monoxide, particulate matter with a diameter of 2.5 μm or less, and particulate matter with a diameter of 10 μm or less) for model development. DNLM analysis revealed that the relative risks consistently increased with rising Baidu index values, while nonlinear associations between temperature and mumps incidence were observed. Among the 4 models, model IBE exhibited the best performance, achieving the coefficient of determination of 0.72, with mean absolute error, mean absolute percentage error, and root-mean-square error values of 0.33, 15.9%, and 0.43, respectively, in the test set.

Conclusions: Our study developed model IBE to predict the incidence of mumps in Yunnan province, offering a potential tool for early detection of mumps outbreaks. The performance of model IBE underscores the potential of integrating search engine data and environmental factors to enhance mumps incidence forecasting. This approach offers a promising tool for improving public health surveillance and enabling rapid responses to mumps outbreaks.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e66072) doi:10.2196/66072

KEYWORDS

mumps; deep learning; baidu index; forecasting; incidence prediction; time series analysis; Yunnan; China

Introduction

Mumps is an acute respiratory infectious disease caused by the mumps virus, characterized by unilateral or bilateral facial swelling [1]. It is transmitted through contact with the respiratory secretions of infected individuals, with an incubation period typically ranging from 16 to 18 days [2]. Nonspecific symptoms such as fever, headache, and malaise could occur before the inflammation of the parotid and other salivary glands [1,2]. While mild symptoms are more common, severe complications such as aseptic meningitis, encephalitis, sensorineural deafness, and orchitis can also develop [3,4]. Out of 2 doses of the mumps vaccine have shown nearly 90% effectiveness in protecting against mumps [5], reducing the global incidence of mumps by 97% [6] Until 2019, a total of 122 countries have introduced mumps vaccination [7], and the number reached to 124 Member States by the end of 2023 [8]. However, in recent years, mumps outbreaks have reemerged in many countries, even those with high vaccine coverage, such as the United States [9], the United Kingdom [10], and Ireland [11], which may have many potential reasons, but the waning immunity may contribute majorly [12].

In China, mumps is classified as a Class C notifiable infectious disease. Despite being vaccine-preventable, mumps remained the most prevalent disease among the 44 notifiable infectious diseases from 2008 to 2017 in persons aged 6 to 22 years, with incidences of 83.207 per 100,000 population in 2008 and 69.771 per 100,000 population in 2017 [13]. Similar to the global condition, even after mumps was included in the expanded immunization program in mainland China in 2008, an abnormal increase in incidence was observed after the program [13,14]. Although no exact reason, waning immunity, changing environment, high population mobility, along with misinformation about the vaccination may play roles [7,15,16]. After the onset of the COVID-19 pandemic, the incidence of mumps in mainland China was 7.37 and 6.48 per 100,000 population in 2022 and 2023, ranking fourth and fifth among Class C notifiable infectious diseases, respectively [17,18]. These trends highlight the continued public health challenge posed by mumps in China.

Yunnan, a province located in southwestern China, has experienced marked variability in mumps incidence. The average annual percentage change in incidence was 13.97% from 2004 to 2018, ranking third, just slightly lower than that of Hunan (15.15%) and Hainan (15.10%) provinces [19]. Furthermore, the mumps incidence in 2022 and 2023 was 14.2 and 12.8 per 100,000 population, respectively, both of which were notably higher than the national averages. These sharp fluctuations in the incidence of mumps, along with relatively higher incidence presented challenges in predicting mumps outbreaks, particularly given the lag of traditional surveillance system, which rely on case reports from medical and health institutions.

Several studies have developed prediction models for the incidence of mumps in China [20,21]. However, due to regional variations of mumps incidence, these models might not be suitable for Yunnan province. In addition, these studies only used mumps incidence, environmental factors data or both as predictors, which may not be able to explain the nonseasonal fluctuations. Incorporating alternative predictors, such as search engine query data, may improve the timeliness and accuracy of outbreak predictions. Previous studies have identified search engine data, such as that from Twitter [rename as X], Google, and Baidu, along with environmental factors, as effective predictors of certain infectious diseases such as influenza and chicken pox, due to their lower cost and higher timeliness [22-24]. Baidu, the most popular search engine in China, provides a rich source of real-time data that can be leveraged for predictive purposes. Long short-term memory (LSTM) networks, introduced by Hochreiter and Schmidhuber [25] in 1997, are a specialized type of recurrent neural network designed to capture long-term dependencies in sequential data by mitigating the vanishing and exploding gradient problems [25]. These networks have shown exceptional performance in capturing temporal dependencies in larger horizons [26]. Therefore, our study aims to use multisource data to develop a prediction model of mumps incidence in Yunnan province based on LSTM.

Methods

Data Sources

Our study used a total of 3 time series data about Yunnan province, including the incidence of mumps, the Baidu index [27], and environmental factors.

We obtained the daily number of mumps cases with their age, sex, and careers, and the annual population data from the Yunnan Center for Disease Control and Prevention, which collected accurate case characteristics from the Chinese information system for infectious diseases control and prevention, for the period between January 1, 2014, and December 31, 2023. From the system, we collected a total of 20 career types in our study, reporting the number of patients of the top 6 most common ones while grouping the remaining careers under the category “other careers.” In addition, since more than 80% of patients were younger than 20 years, we organized the age distribution into 5-year groups for those younger than 20 years, while others were presented as a separate category. The Baidu index [27] is the weighted sum of the search frequencies of various search terms by scientifically analyzing and calculating based on users’ search volumes on Baidu, which is the most popular search engine in China. By extensive search, we initially collected daily Baidu index for 30 search terms in Yunnan from January 1, 2016, and December 31, 2023. After group discussion, we excluded 3 search terms due to irrelevance and another 3 because they were introduced after 2016, resulting

XSL•FO
RenderX

in incomplete data (Table S1 in Multimedia Appendix 1). Ultimately, we included 24 search terms in our study.

We collected information about environmental factors during January 1, 2016, and December 31, 2023, from China Meteorological Administration [28], including maximum temperature (°C), minimum temperature (°C), air quality index, particulate matter (PM) with an aerodynamic diameter less than or equal to 2.5 μm ($PM_{2.5}$; measured as μg/m^3), PM with an aerodynamic diameter less than or equal to 10 μm (PM_{10}; measured as μg/m^3), sulfur dioxide (SO_2; measured as μg/m^3), nitrogen dioxide (NO_2; measured as μg/m^3), carbon monoxide (CO; measured as mg/m^3), and ozone (O_3; measured as μg/m^3).

Statistical Analysis

Feature Selection and Descriptive Analysis

To predict the weekly incidence of mumps, we calculated the weekly values for each variable. The weekly mumps incidence and Baidu index [27] were obtained by summing their daily values, while the weekly environmental factors were calculated as the mean values for each week. We performed Pearson correlation tests to determine the correlation between each variable and the incidence of mumps. After excluding variables with a Pearson correlation coefficient (PCC) of <0.10 or a *P* value of <.05, we consequently included 16 search terms and 6 environmental factors (Table S2 in Multimedia Appendix 1). These search terms were classified into 3 groups based on their content: disease name, symptoms, and treatment. We then calculated the weekly sum of the Baidu index for these 3 groups as the predictors.

Our study used the Kolmogorov-Smirnov test to inspect the normality of continuous variables. Normally distributed continuous variables were described as mean and SD, while nonnormally distributed continuous variables are described as median and IQR.

Log Correlation Analysis

Previous studies have demonstrated that environmental factors and the Baidu index [27] had a lag impact on the incidence of infectious disease, which means exposure affects health for a period of days [22,29]. To capture these complex relationships, we used the distributed lag nonlinear model (DLNM). DLNM is a modeling framework that simultaneously represents nonlinear exposure-response relationships and lagged effects using cross-basis functions [30]. We used natural spline functions for both the exposure-response and lag-response components, allowing for flexible modeling of nonlinear associations. We selected DLNM over other nonlinear methods, such as generalized additive models, because DLNM is specifically tailored to handle both nonlinear and lagged effects, which is crucial for understanding the temporal dynamics of environmental influences on mumps incidence. Our study used R package DLNM (version 2.4.7) to make DLNM between the Baidu index [27] and environmental factors and mumps incidence [31].



Yt represents the outcome value (mumps incidence) at week t while E[Yt] is its expected value. α refers to the interception term. p refers to the maximum lag time, which was 3 weeks in our study considering the incubation and previous study [32]. βi the regression coefficient for the i-th lag time, representing the effect size of the exposure variable at lag i on the outcome variable. cb(Xt - i, argvar, arglag) is a cross-basis function used to capture the nonlinear relationship between the exposure variable Xt-i and the lag time. This function depends on 3 parameters: Xt-i represents the value of the exposure variable at time t-i, argvar were parameters defining the effect of the exposure variable, and arglag were parameters defining the lag effect. Relative risks (RR) were used to assess the effect of a specific exposure level at one lag time compared with the reference exposure value.

Model Construction and Performance Evaluation

Our study developed 4 models based on LSTM to predict mumps incidence 1 week in advance (Figure 1). First, to better capture the characteristics of the 3 time series data, we constructed 3 primary LSTM models for each data set. The primary model for mumps incidence included a dropout layer to prevent overfitting, followed by 2 LSTM layers and a dense layer. The Baidu index [27] and environmental factors models each consisted of 2 LSTM layers followed by a dense layer. Among the 3 models, we used data from the previous 2 weeks to predict the incidence of mumps one week in advance. Subsequently, we concatenated the outputs of these primary models to construct composite models. Model IB combined the mumps incidence and Baidu index outputs, model IE combined the mumps incidence and environmental factors outputs, model BE combined the Baidu index [27] and environmental factors outputs, and model IBE combined the outputs from all 3 primary models. Each composite model then passed the concatenated outputs through an additional dense layer to produce the final prediction. This approach ensured that primary models effectively integrated information from different data sources for accurate predictions. To simulate real-world scenarios where future data are unknown and must be predicted based on past observations, we divided our data sets into train, validation and test set based on time. The train set interval was from January 1, 2016, to December 31, 2021. Data from January 1, 2022, to December 31, 2022, and those from January 1, 2023, to December 31, 2023, were used as the validation and test sets, respectively. These models were constructed by keras3 package [ONEIROS] (version 1.0.0), tensorflow package [Google Brain Team] (version 2.16.0) in R (version 4.4.1; R Core Team).

XSL•FO
RenderX

Figure 1. Model constructions. LSTM: Long short-term memory. Model BE: model including Baidu index and environmental factors; Model IB: model including mumps incidence and Baidu index; Model IE: model including mumps incidence and environmental factors; Model IBE: model including mumps incidence and Baidu index and environmental factors.

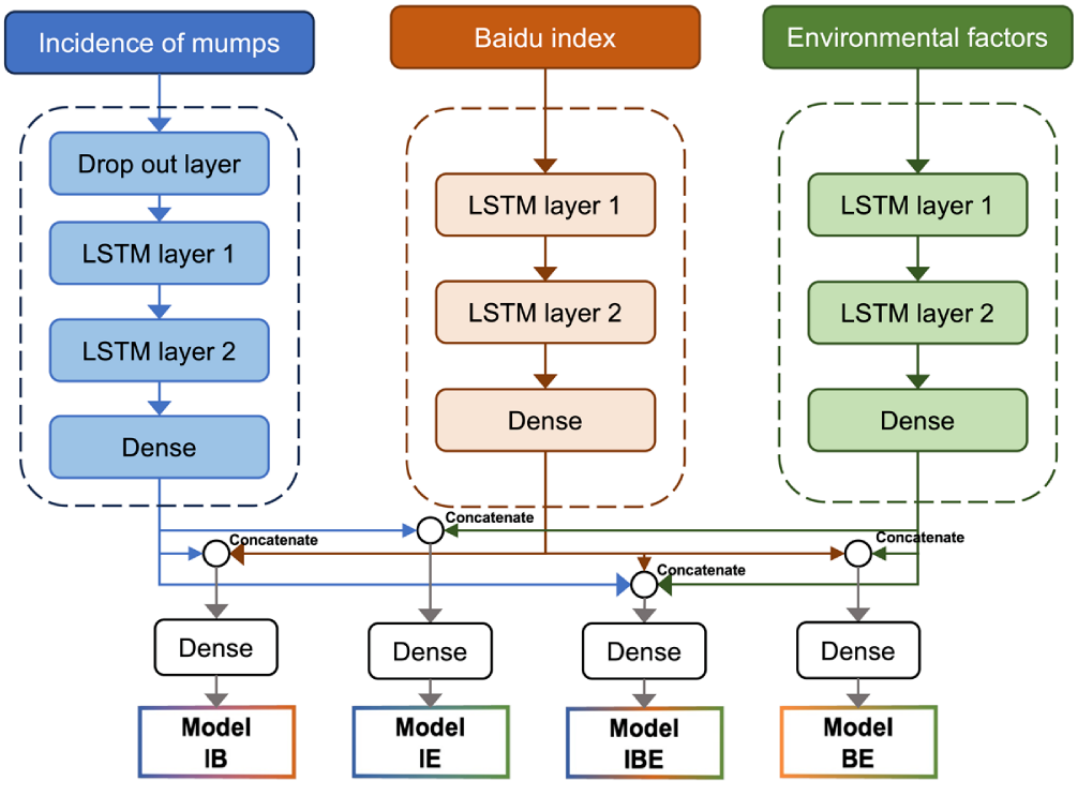

Out of 4 performance metrics were applied to evaluate these 4 models: coefficient of determination (R^2), mean absolute error (MAE), mean absolute percentage error (MAPE), and root-mean-square error (RMSE). R^2 refers to the proportion of the variance for a dependent variable that is explained by the model, with values less than 1, where larger values indicate better performance. MAE, MAPE, and RMSE measure the errors between predictions and actual values, with MAE and RMSE ranging greater than 0, and MAPE ranging from 0 to 100%, where lower values indicate better model accuracy. Compared with the MAE, MAPE represents the average of absolute percentage errors expressed as a percentage, making it more suitable for comparison across different data sets and models, while RMSE is more sensitive to outliers and larger errors.



In these equations, n was the number of observations, while y_i and y_i refer to the actual incidence of mumps in the i_{th} week and its prediction, respectively.

Ethical Considerations

Our research was approved by the Medical Ethics Committee, Xiangya School of Public Health, Central South University, Changsha, China (XYGW-2024-65). Informed consent was waived by our institutional review board because all data were fully anonymized before being accessed by the researchers, ensuring that no identifiable information about individuals was available. Therefore, compensation is not applicable in our study.

Results

The Temporal Distribution of Mumps Incidence and Predictors

In Yunnan Province, the incidence of mumps increased from 10.9 per 100,000 population in 2014, peaked at 37.5 per 100,000 population in 2019, and then steadily decreased to 12.8 per 100,000 population by 2023 (Figure 2). The outbreaks of mumps occurred from 2017 to 2020 in different cities in Yunnan. From 2014 to 2023, the proportion of female patients remained fairly stable, ranging from 41.3% in 2015 to 45.7% in 2020. (Table 1). The median age of patients showed some fluctuations over the years, starting at 7 years (IQR 4-12) in 2014, peaking at 11 years (IQR 7-14) in 2019, and then decreasing to 6 years (IQR 4-10) by 2023. The proportion of cases in children between 5 and 10 years old remained substantial, with the highest proportion recorded in 2023 (43.7%). Students consistently represented the largest affected group, comprising 44.3% to 67.5% of cases in 2014 and 2019, respectively.

XSL•FO
RenderX

Figure 2. Annual incidence of mumps (per 100,000 population) in Yunnan province from 2014 to 2023.

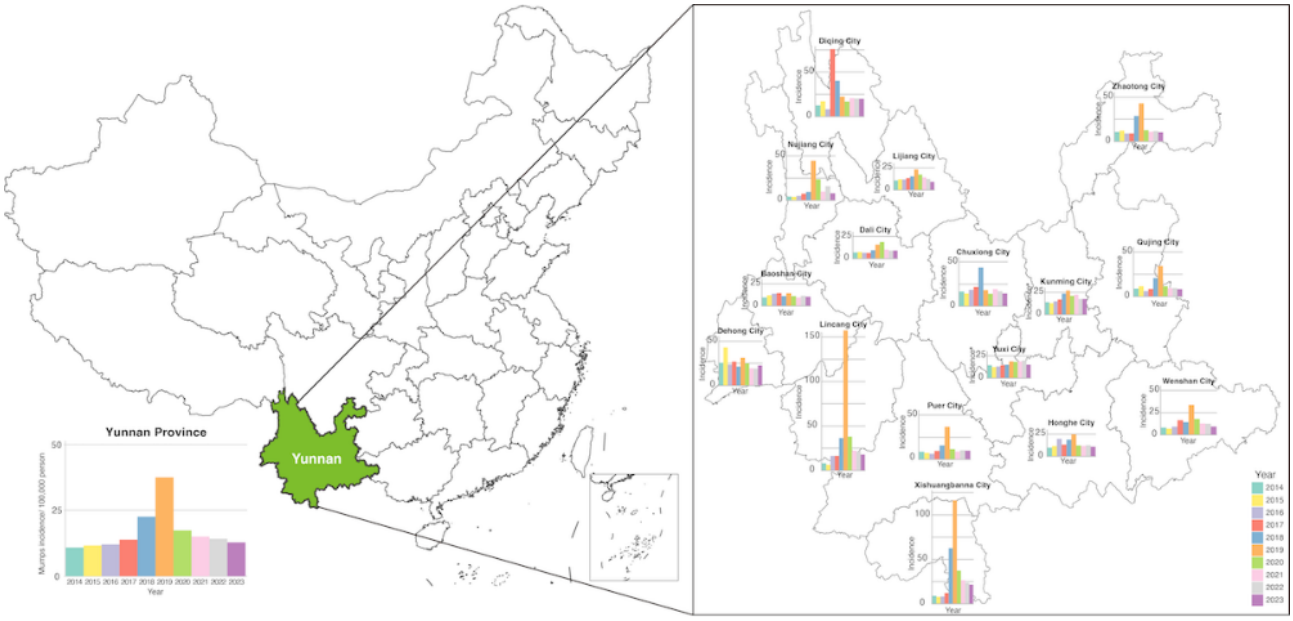

Table 1. Characteristics of reported mumps patients in Yunnan from 2014 to 2023.

	2014 (N=5343)	2015 (N=5611)	2016 (N=5850)	2017 (N=6843)	2018 (N=11,249)	2019 (N=18,703)	2020 (N=8592)	2021 (N=7296)	2022 (N=6788)	2023 (N=6085)
Female	2290 (42.9)	2320 (41.3)	2484 (42.5)	2927 (42.8)	4961 (44.1)	8214 (43.9)	3929 (45.7)	3250 (44.5)	2929 (43.1)	2599 (42.7)
Age (years), median (IQR)	7 (4-12)	8 (5-11)	8 (5-12)	9 (5-13)	10 (6-14)	11 (7-14)	8 (5-14)	6 (4-11)	6 (4-11)	6 (4-10)
Age groups (years), n (%)										
<5	1389 (26)	1387 (24.7)	1456 (24.9)	1329 (19.4)	1679 (14.9)	2266 (12.1)	2051 (23.9)	2183 (29.9)	1940 (28.6)	1700 (27.9)
5 to <10	2082 (39)	2224 (39.6)	2300 (39.3)	2484 (36.3)	3492 (31)	5236 (28.0)	2801 (32.6)	2828 (38.8)	2873 (42.3)	2661 (43.7)
10 to <15	1007 (18.8)	1105 (19.7)	1074 (18.4)	1807 (26.4)	3786 (33.7)	7019 (37.5)	1618 (18.8)	944 (12.9)	850 (12.5)	715 (11.8)
15 to <20	162 (3)	200 (3.6)	210 (3.6)	308 (4.5)	1098 (9.8)	2272 (12.1)	679 (7.9)	289 (4.0)	201 (3.0)	152 (2.5)
≥20	703 (13.2)	695 (12.4)	810 (13.8)	915 (13.4)	1194 (10.6)	1910 (10.2)	1443 (16.8)	1052 (14.4)	924 (13.6)	857 (14.1)
Careers, n (%)										
Student	2368 (44.3)	2608 (46.5)	2654 (45.4)	3627 (53)	7027 (62.5)	12618 (67.5)	3912 (45.5)	2794 (38.3)	2571 (37.9)	2269 (37.3)
Diaspora children	1246 (23.3)	1229 (21.9)	1073 (18.3)	1001 (14.6)	1169 (10.4)	1660 (8.9)	1259 (14.7)	1059 (14.5)	972 (14.3)	796 (13.1)
Nursery children	992 (18.6)	1041 (18.6)	1271 (21.7)	1267 (18.5)	1798 (16.0)	2302 (12.3)	1961 (22.8)	2398 (32.9)	2326 (34.3)	2163 (35.5)
Peasant	519 (9.7)	539 (9.6)	614 (10.5)	670 (9.8)	926 (8.2)	1546 (8.3)	1100 (12.8)	761 (10.4)	666 (9.8)	616 (10.1)
Unemployment	33 (0.6)	33 (0.6)	45 (0.8)	48 (0.7)	57 (0.5)	101 (0.5)	62 (0.7)	55 (0.8)	61 (0.9)	68 (1.1)
Teacher	17 (0.3)	20 (0.4)	24 (0.4)	29 (0.4)	56 (0.5)	100 (0.5)	62 (0.7)	32 (0.4)	30 (0.4)	19 (0.3)
Other careers	168 (3.1)	141 (2.5)	169 (2.9)	201 (2.9)	216 (1.9)	376 (2)	236 (2.7)	197 (2.7)	162 (2.4)	154 (2.5)

The temporal distribution from 2016 to 2023 revealed distinct trends in both the Baidu search index and environmental factors (Figure 3). The distribution of the 3 Baidu index mirrored the pattern observed in mumps incidence, with a peak in 2019 and followed by a decline. Most environmental factors except SO_2 displayed obvious seasonal cycles. Maximum and minimum

XSL•FO
RenderX

temperatures reached peaks in summer, while air pollutants, including CO, $PM_{2.5,}$ and PM_{10}, had higher concentrations in winter. In addition, SO_2 and CO showed a long-term decreasing trend over the 10-year period.

Figure 3. The temporal distribution of weekly Baidu index and environmental factors in Yunnan Province from 2016 to 2023 CO: carbon monoxide; SO2: sulfur dioxide; PM: particulate matter; PM2.5: particulate matter with a diameter of 2.5 μm or less; PM10: particulate matter with a diameter of 10 μm or less.

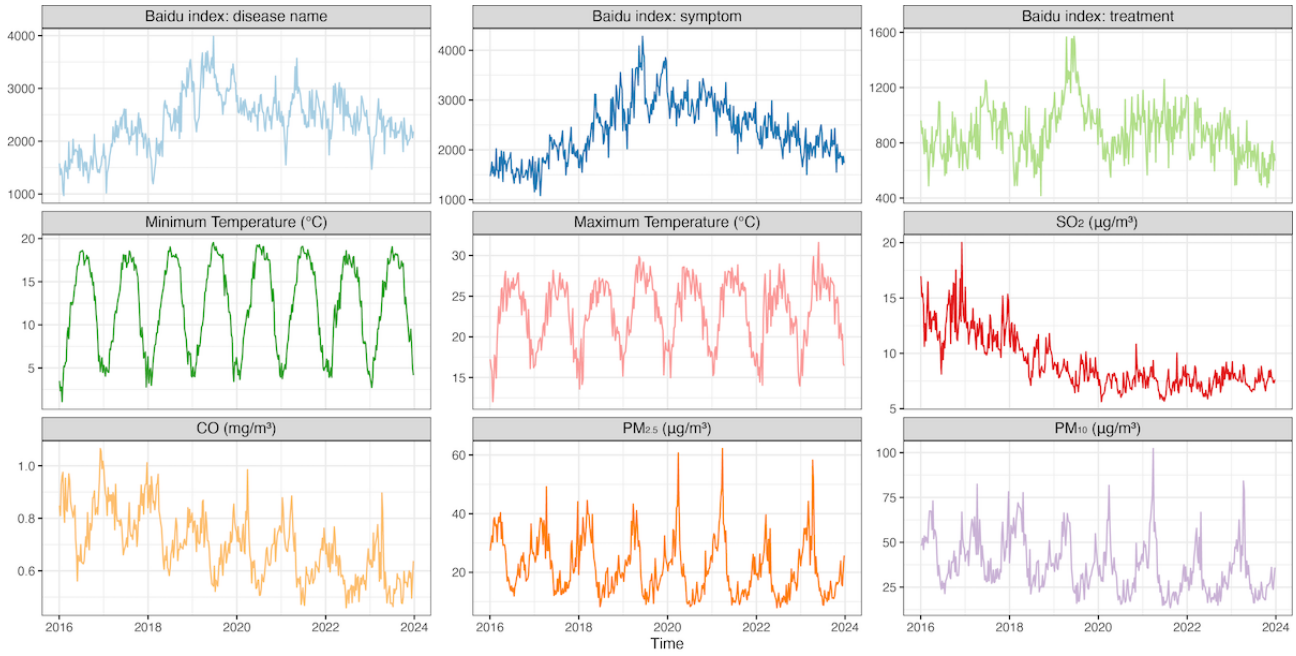

Correlation Between Predictors and Incidence of Mumps

The Pearson correlation analysis reveals that 3 Baidu search terms, disease name (r=0.67), symptom (r=0.66), and treatment (r=0.53), show strong positive correlations with mumps incidence in Yunnan Province from 2016 to 2023 (Table 2). Among environmental factors, both minimum (r=0.17) and maximum temperature (r=0.16) have weak positive correlations, while pollutants including SO_2 (r=–0.16), CO (r=–0.12), $PM_{2.5}$ (r=–0.20) and PM_{10} (r=–0.10) show weak negative correlations with mumps incidence.

XSL•FO
RenderX

Table 2. Distribution of predictors and Pearson correlation coefficient between weekly predictors and mumps incidence in Yunnan province from 2016 to 2023.

Predictors	Baidu search terms in Chinese	Median (IQR)	PCC[a]	*P* value
Baidu search terms				
Disease name	—	2394 (2053-2697)	0.67	<.001
Mumps	流行性腮腺炎	560 (457-655)	0.42	<.001
Parotitis	腮腺炎	1397 (1225-1564)	0.76	<.001
Zha sai (a colloquial name for mumps in Chinese)	痄腮	301 (174-418)	0.36	<.001
San xian yan (a term phonetically like “parotitis” in Chinese)	三线炎	114 (57-172)	0.39	<.001
Symptom	—	2360 (1928-2821)	0.66	<.001
Jaw pain	下巴疼	114 (57-174)	0.38	<.001
Swelling on one side of the face	半边脸肿	0 (0-57)	0.23	<.001
Pain at the base of the ear	耳根痛	0 (0-0)	0.16	<.001
Facial swelling	脸肿	440 (348-524)	0.39	<.001
Complications of parotitis	腮腺炎并发症	0 (0-57)	0.24	<.001
Early symptoms of parotitis	腮腺炎的早期症状	115 (57-294)	0.26	<.001
Symptoms of parotitis + parotitis symptoms	腮腺炎的症状+腮腺炎症状	1315 (876-1560)	0.33	<.001
Parotitis with orchitis	腮腺炎睾丸炎	0 (0-0)	0.13	.01
Is parotitis contagious?	腮腺炎传染吗	436 (172-508)	0.48	<.001
Treatment	—	860 (729-1000)	0.53	<.001
What medicine to take for parotitis	腮腺炎吃什么药	464 (441-506)	0.30	<.001
How to treat parotitis	腮腺炎怎么治疗	300 (233-405)	0.54	<.001
Treatment of parotitis	腮腺炎的治疗	57 (0-115)	0.13	.007
Environmental factors				
Maximum temperature, °C	—	24 (20.2-26.3)	0.16	<.001
Minimum temperature, °C	—	12.8 (7.6-17.2)	0.17	<.001
SO^b_2 (μg/m^3)	—	8.2 (7.4-10.6)	–0.20	<.001
CO^c (mg/m^3)	—	0.7 (0.6-0.8)	–0.12	.02
$PM^d_{2.5}$ (μg/m^3)	—	19.5 (13.8-27.4)	–0.20	<.001
PM_{10} (μg/m^3)		34.3 (25.5-46.0)	–0.10	.04

[a]PCC: Pearson correlation coefficient.

[b]SO: sulfur dioxide.

[c]CO: carbon monoxide.

[d]PM: particulate matter.

Considering the potential lag impact of these predictors on mumps incidence, we leveraged DLNM to explore their short-term impacts (Figure 4). The impact of various predictors on the incidence of mumps varies across different lag weeks. For 3 Baidu search terms, the RR consistently increases with the rising Baidu index. The analysis also revealed nonlinear associations between temperatures and incidence, with minimum temperature showing a U-shaped curve and maximum temperature exhibiting a complex pattern. In addition, elevated levels of SO_2, CO, $PM_{2.5}$, and PM_{10} generally corresponded with decreases in RR at a 1-week lag.

XSL•FO
RenderX

Figure 4. Relative risks for mumps incidence by predictors and lags using distributed lag nonlinear models in Yunnan province from 2016 to 2023, with references at median of each predictor.

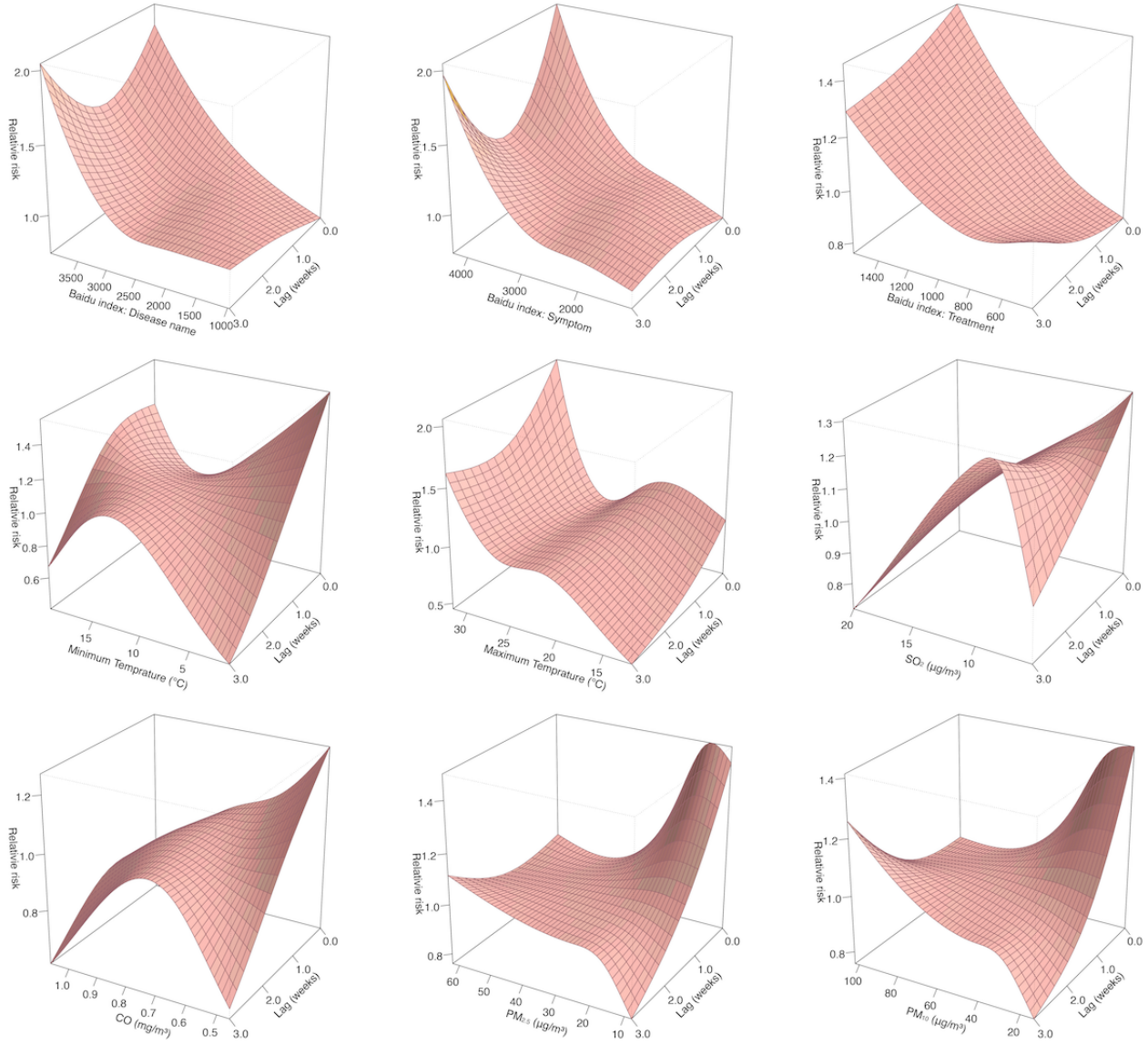

Model Performances and Comparisons

Table 3 shows the comparisons of the performance metrics of 4 models with different combinations of predictors. Model IBE, which includes all 3 predictors, consistently outperformed the other models in both the train and validation set, as well as the test set. Specifically, model IBE achieved the highest R^2 values (0.89 in the train and validation set, 0.72 in the test set) and the lowest error metrics, with an MAE of 0.51, MAPE of 15.2%, and RMSE of 0.74 in the train and validation set, and an MAE of 0.33, MAPE of 15.9%, and RMSE of 0.43 in the test set. In comparison, the IE model (R^2=0.54, MAE=0.45, MAPE=28.2%, and RMSE=0.55) and the IB model (R^2=0.49, MAE=0.45, MAPE=29.2%, and RMSE=0.58) showed similar levels of prediction accuracy in the test set, both outperforming model BE, which had the worst performance (R^2=0.33, MAE=0.54, MAPE=32.0%, and RMSE=0.66). The actual values and model predictions of the weekly incidence of mumps are presented in Figure 5.

XSL•FO
RenderX

Table 3. Performance metrics of different models.

	Model BE[a]	Model IE[b]	Model IB[c]	Model IBE[d]
Train and validation set				
R^2	0.80	0.84	0.84	0.89
MAE[e]	0.69	0.59	0.65	0.51
MAPE[f] (%)	21.2	17.1	21.6	15.2
RMSE[g]	0.97	0.87	0.88	0.74
Test set				
R^2	0.33	0.54	0.49	0.72
MAE	0.54	0.45	0.45	0.33
MAPE (%)	32.0	28.2	29.2	15.9
RMSE	0.66	0.55	0.58	0.43

[a]Model BE: the model included Baidu index and environmental factors.

[b]Model IB: the model included mumps incidence and Baidu index.

[c]Model IE: the model included mumps incidence and environmental factors.

[d]Model IBE: the model included mumps incidence, Baidu index, and environmental factors.

[e]MAE: mean absolute error.

[f]MAPE: mean absolute percentage error.

[g]RMSE: root-mean-square error.

Figure 5. Predictions of weekly mumps incidence 1 week in advance by different models in Yunnan province from 2016 to 2023. Model BE: a model including Baidu index and environmental factors; model IB: the model including mumps incidence and Baidu index; model IE: a model including mumps incidence and environmental factors; model IBE: the model including mumps incidence, Baidu index, and environmental factors.

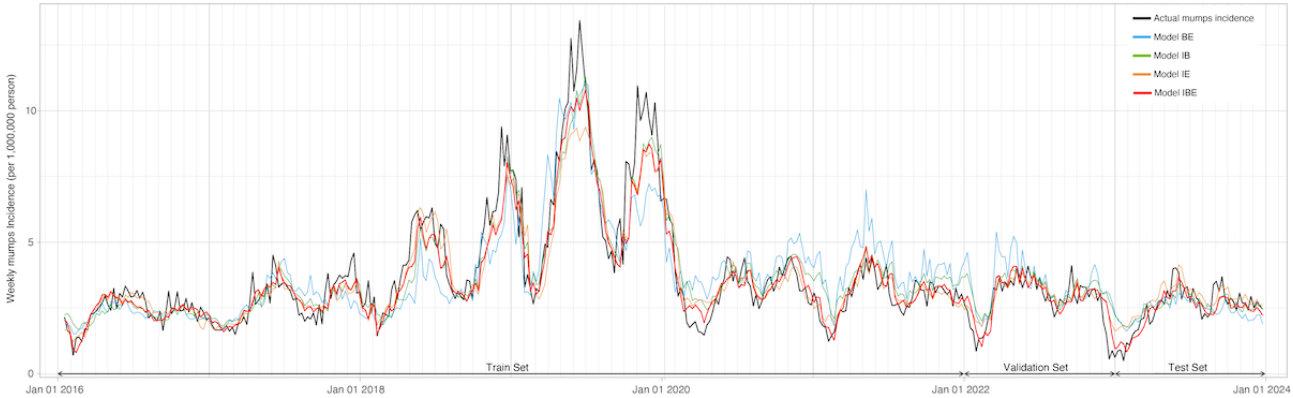

Discussion

Principal Findings

Our study presented the incidence of mumps from 2014 to 2023 in Yunnan, a representative province of irregular mumps activity, and found continuous mumps outbreaks in different cities from 2017 to 2020. Using multisource data, we constructed models for predicting the incidence of mumps one week in advance based on LSTM. Model IBE (ie, the model including the incidence of mumps, Baidu index, and environmental factors) showed the best performance among 4 models.

Comparison to Previous Work

The mean incidence of mumps in Yunnan from 2014 to 2023 was 16.9 per 100,000 persons, which is lower than that of the period between 2004 to 2012 (26.2/100,000) in Yunnan and comparable to the incidence in mainland China from 2013 to 2018 (16.9/100,000) [19]. The 10-year incidence pattern of mumps in Yunnan shows an increase followed by a decrease starting in 2020. The decline may be related to the COVID-19 pandemic, which extensively reduced opportunities for interpersonal contact, thereby blocking the spread of infectious diseases. However, we do not have a definite explanation for the peak in 2019. Given the rapidly increasing number of cases among students in 2019, factors such as vaccine coverage, waning immunity, and the close gathering of adolescents on campuses may have contributed to the outbreak [3,33].

Previous studies have found nonlinear associations between meteorological factors and the incidence of several infectious diseases including mumps [34,35], influenza [23,24,36], and chickenpox [37]. These findings suggest meteorological factors are common predictors for the field of infectious disease

XSL•FO
RenderX

prediction. Our study found 6 environmental factors with a PCC of >0.10 and presented their lag correlations with the incidence of mumps. The most relevant factors were $PM_{2.5}$ and SO_2, both showing weak negative correlations (PCC=–0.2) with mumps incidence in Yunnan. Notably, the PCC is obviously lower than that of other infectious disease. For example, a study in Singapore found a stronger correlation (PCC=0.32) between minimum temperature and the COVID-19 pandemic incidence [38]. This variability suggests that while environmental factors can be valuable for predicting some infectious diseases, they may not be sufficient for accurate prediction across all diseases, highlighting the need for additional predictors.

In recent years, using search engine data, such as Google Trends [39], and the Baidu index [24,40], to predict the incidence of infectious disease has become popular and demonstrated excellent predictive performance. The advantages of search engine data included timeliness, low cost, and sometimes stronger correlation with disease incidence than other data sources. Given the different disease outcomes, fairly comparing our study with other research using search engine data is challenging, as this is, to our knowledge, the first study to construct a mumps prediction model using search engine data. However, we noted that Yang et al [24] and Choi and Ahn [39], who predicted influenza incidence in mainland China and South Korea respectively, achieved similar or even better performance using Baidu index and Google Trends alone, compared with our study's combination of the Baidu index and environmental factors. It suggests that the contribution of search engine data to incidence prediction may vary depending on types of disease, regions, and data sources. In our study, the PCCs of the Baidu index of 3 search term groups were higher than those of environmental factors. This may be because the Baidu index, derived from individuals' active search activities, better reflected the incidence peak between 2018 to 2020. In contrast, the relatively stable environmental factors might contribute more to the seasonal trends in incidence but failed to capture the sharp fluctuations observed during outbreaks.

Our study deployed multisource data to construct and compare 4 models, with the model IBE demonstrating superior performance. Notably, when Baidu index data was incorporated, model IBE exhibited even greater predictive accuracy compared with the IE model. This demonstrates the capability of the Baidu index to predict mumps incidence and the potential of integrating multiple data sources to complement the surveillance systems in predicting mumps more accurately, as these data sets can complement one another and provide a more comprehensive understanding of the influencing factors and early signs of disease spread. Further research could build on the proposed framework in our study or explore the integration of additional data sources, such as social media activity and movement behaviors, to further enhance the predictive models.

Strengths and Limitations

To our knowledge, our study is the first to develop a prediction model for the incidence of mumps that combines the Baidu index and environmental factors as predictors, and it is also the first to create such a model specifically for Yunnan province, China. In addition, by leveraging data such as the Baidu index, our model offers a novel and practical approach to monitor mumps incidence 1 week in advance.

However, there are several limitations to our study. First, at the city level rather than the provincial level, we observed different outbreak times among cities from 2017 to 2020, which may indicate the spread trajectory of mumps and be important for its control and prevention. Unfortunately, the time-series forecast model in our study cannot leverage this geographical information. Future studies could explore integrating temporal and spatial data, which may be crucial for preventing widespread transmission. Second, some Baidu search terms were introduced after 2016 (eg, the search term "Symptoms and treatment of mump" was introduced on April 2, 2018), thereby decreasing the number of available predictors of the Baidu index, which may compromise the performance of Baidu index. Nonetheless, after extensive research, our study identified and included a sufficient number of predictors. Third, we did not have access to some meteorological factors such as sunshine duration and wind speed, which have been identified as effective predictors in other studies, potentially hindering better performance of our models. Future research should incorporate a broader range of meteorological variables to improve prediction accuracy further. Fourth, applying the model to practical public health interventions still faced several challenges. Future work should focus on developing a user-friendly tool that can automatically collect predictors and output results, making the model more accessible and actionable for public health authorities. Finally, due to economic and geographic variations in China [19], various epidemic patterns of infectious diseases, and the different nature of data sets from other search platforms, this model may not generalize well across different regions, diseases, or data sources without careful adaptation and validation. While the model in this study was applied to mumps, similar approaches could potentially be applied to other infectious diseases with seasonal or outbreak-driven incidence patterns. Diseases such as influenza, chickenpox, and hand, foot, and mouth disease, where public awareness, mobility, and environmental factors play a role in transmission, might benefit from similar predictive frameworks. Future studies should focus on tailoring models to specific diseases and data sets to evaluate their broader applicability.

Conclusions

In conclusion, our study showed the incidence of mumps in Yunnan province, China, from 2014 to 2023, highlighting a sharp peak around 2019. We used data from 2016 to 2023, including mumps incidence, the Baidu index, and environmental factors, to construct 4 prediction models based on LSTM networks. Among these models, the model IBE demonstrates the best performance. This indicates the feasibility of using search engine data and environmental factors to forecast the incidence of mumps, offering a valuable tool for public health surveillance and rapid outbreak assessment.

XSL•FO
RenderX

Acknowledgments

XX, LX, and IXW designed the study. XX extracted data. LX conducted data analysis. XX and LX wrote the draft. IX W, SD, and LC critically revised the manuscript.

Special thanks would be extended to IXW and SD for equqally invovling in the correspondence with collaborators and the review process. While not listed as a correspoding author, IXW made significant contributions throughout the process.

We would like to thank Baidu and the China Center for Disease Control and Prevention.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File , 25 KB - jmir_v27i1e66072_app1.docx]

References

1. Grennan D. Mumps. JAMA 2019;322(10):1022. [doi: 10.1001/jama.2019.10982] [Medline: 31503310]
2. Lam E, Rosen JB, Zucker JR. Mumps: an update on outbreaks, vaccine efficacy, and genomic diversity. Clin Microbiol Rev 2020;33(2):e00151-e00119 [FREE Full text] [doi: 10.1128/CMR.00151-19] [Medline: 32102901]
3. Principi N, Esposito S. Mumps outbreaks: a problem in need of solutions. J Infect 2018;76(6):503-506. [doi: 10.1016/j.jinf.2018.03.002] [Medline: 29678496]
4. Dayan GH, Quinlisk MP, Parker AA, Barskey AE, Harris ML, Schwartz JMH, et al. Recent resurgence of mumps in the United States. N Engl J Med 2008;358(15):1580-1589. [doi: 10.1056/NEJMoa0706589] [Medline: 18403766]
5. Di Pietrantonj C, Rivetti A, Marchione P, Debalini MG, Demicheli V. Vaccines for measles, mumps, rubella, and varicella in children. Cochrane Database Syst Rev 2021;11(11):CD004407 [FREE Full text] [doi: 10.1002/14651858.CD004407.pub5] [Medline: 34806766]
6. Galazka AM, Robertson SE, Kraigher A. Mumps and mumps vaccine: a global review. Bull World Health Organ 1999;77(1):3-14 [FREE Full text] [Medline: 10063655]
7. Kauffmann F, Heffernan C, Meurice F, Ota MOC, Vetter V, Casabona G. Measles, mumps, rubella prevention: how can we do better? Expert Rev Vaccines 2021;20(7):811-826 [FREE Full text] [doi: 10.1080/14760584.2021.1927722] [Medline: 34096442]
8. World Health Organization. Immunization coverage. URL: https://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage [accessed 2024-11-25]
9. Zamir CS, Schroeder H, Shoob H, Abramson N, Zentner G. Characteristics of a large mumps outbreak: clinical severity, complications and association with vaccination status of mumps outbreak cases. Hum Vaccin Immunother 2015;11(6):1413-1417 [FREE Full text] [doi: 10.1080/21645515.2015.1021522] [Medline: 25874726]
10. Aasheim ET, Inns T, Trindall A, Emmett L, Brown KE, Williams CJ, et al. Outbreak of mumps in a school setting, United Kingdom, 2013. Hum Vaccin Immunother 2014;10(8):2446-2449 [FREE Full text] [doi: 10.4161/hv.29484] [Medline: 25424953]
11. Ferenczi A, Gee S, Cotter S, Kelleher K, Mumps Outbreak Control Team. Ongoing mumps outbreak among adolescents and young adults, Ireland, August 2018 to January 2020. Euro Surveill 2020;25(4):2000047 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.4.2000047] [Medline: 32019666]
12. Marshall HS, Plotkin S. The changing epidemiology of mumps in a high vaccination era. Lancet Infect Dis 2019;19(2):118-119. [doi: 10.1016/S1473-3099(18)30541-3] [Medline: 30558996]
13. Dong Y, Wang L, Burgner DP, Miller JE, Song Y, Ren X, et al. Infectious diseases in children and adolescents in China: analysis of national surveillance data from 2008 to 2017. BMJ 2020;369:m1043 [FREE Full text] [doi: 10.1136/bmj.m1043] [Medline: 32241761]
14. Cui A, Zhu Z, Hu Y, Deng X, Sun Z, Zhang Y, et al. Mumps epidemiology and mumps virus genotypes circulating in mainland China during 2013-2015. PLoS One 2017;12(1):e0169561 [FREE Full text] [doi: 10.1371/journal.pone.0169561] [Medline: 28085897]
15. Paradise Vit A, Magid A. Differences in fear and negativity levels between formal and informal health-related websites: analysis of sentiments and emotions. J Med Internet Res 2024;26:e55151 [FREE Full text] [doi: 10.2196/55151] [Medline: 39120928]
16. Westphal DW, Eastwood A, Levy A, Davies J, Huppatz C, Gilles M, et al. A protracted mumps outbreak in Western Australia despite high vaccine coverage: a population-based surveillance study. Lancet Infect Dis 2019;19(2):177-184. [doi: 10.1016/S1473-3099(18)30498-5] [Medline: 30558994]

XSL•FO
RenderX

17. Department of Planning，Development，and Information Technology，National Health Commission of the People's Republic of China. Statistical bulletin of China's health development in 2023. Chin J Viral Dis 2024;14(05):416-424 [FREE Full text] [doi: 10.16505/j.2095-0136.2024.0083]
18. Department of Planning,Development and Information Technology,National Health Commission of the People's Republic of China. Statistical bulletin of China's health development in 2022. Chin J Viral Dis 2024;14(04):317-325 [FREE Full text] [doi: 10.16505/j.2095-0136.2024.0066]
19. Fu X, Ge M, Xu W, Yu M, Ju J, Zhong Y, et al. Epidemiological features and sociodemographic factors associated with mumps in mainland China from 2004 to 2018. J Med Virol 2022;94(10):4850-4859 [FREE Full text] [doi: 10.1002/jmv.27955] [Medline: 35739613]
20. Zhang H, Su K, Zhong X. Association between meteorological factors and mumps and models for prediction in Chongqing, China. Int J Environ Res Public Health 2022;19(11):6625 [FREE Full text] [doi: 10.3390/ijerph19116625] [Medline: 35682208]
21. Jia J, Liu M, Xue Z, Wang Z, Pan Y. Prediction of mumps incidence trend in China based on difference grey model and artificial neural network learning. Iran J Public Health 2021;50(7):1405-1415 [FREE Full text] [doi: 10.18502/ijph.v50i7.6630] [Medline: 34568179]
22. Wang Z, He J, Jin B, Zhang L, Han C, Wang M, et al. Using baidu index data to improve chickenpox surveillance in Yunnan, China: infodemiology study. J Med Internet Res 2023;25:e44186 [FREE Full text] [doi: 10.2196/44186] [Medline: 37191983]
23. Su K, Xu L, Li G, Ruan X, Li X, Deng P, et al. Forecasting influenza activity using self-adaptive AI model and multi-source data in Chongqing, China. EBioMedicine 2019;47:284-292 [FREE Full text] [doi: 10.1016/j.ebiom.2019.08.024] [Medline: 31477561]
24. Yang L, Zhang T, Han X, Yang J, Sun Y, Ma L, et al. Influenza epidemic trend surveillance and prediction based on search engine data: deep learning model study. J Med Internet Res 2023;25:e45085 [FREE Full text] [doi: 10.2196/45085] [Medline: 37847532]
25. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
26. Lara-Benítez P, Carranza-García M, Riquelme JC. An experimental review on deep learning architectures for time series forecasting. Int J Neural Syst 2021;31(3):2130001. [doi: 10.1142/S0129065721300011] [Medline: 33588711]
27. Baidu index. URL: https://index.baidu.com [accessed 2024-08-16]
28. Tianqihoubao. URL: http://www.tianqihoubao.com [accessed 2024-09-20]
29. Zhang T, Qin W, Nie T, Zhang D, Wu X. Effects of meteorological factors on the incidence of varicella in Lu'an, Eastern China, 2015-2020. Environ Sci Pollut Res Int 2023;30(4):10052-10062. [doi: 10.1007/s11356-022-22878-0] [Medline: 36066801]
30. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. Stat Med 2010;29(21):2224-2234 [FREE Full text] [doi: 10.1002/sim.3940] [Medline: 20812303]
31. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. J Stat Softw 2011;43(8):1-20 [FREE Full text] [Medline: 22003319]
32. Li R, Lin H, Liang Y, Zhang T, Luo C, Jiang Z, et al. The short-term association between meteorological factors and mumps in Jining, China. Sci Total Environ 2016;568:1069-1075. [doi: 10.1016/j.scitotenv.2016.06.158] [Medline: 27353959]
33. Vygen S, Fischer A, Meurice L, Mounchetrou Njoya I, Gregoris M, Ndiaye B, et al. Waning immunity against mumps in vaccinated young adults, France 2013. Euro Surveill 2016;21(10):30156 [FREE Full text] [doi: 10.2807/1560-7917.ES.2016.21.10.30156] [Medline: 26987576]
34. Yu G, Yang R, Yu D, Cai J, Tang J, Zhai W, et al. Impact of meteorological factors on mumps and potential effect modifiers: an analysis of 10 cities in Guangxi, Southern China. Environ Res 2018;166:577-587. [doi: 10.1016/j.envres.2018.06.047] [Medline: 29966878]
35. Hu W, Li Y, Han W, Xue L, Zhang W, Ma W, et al. Meteorological factors and the incidence of mumps in Fujian Province, China, 2005-2013: Non-linear effects. Sci Total Environ 2018;619-620:1286-1298 [FREE Full text] [doi: 10.1016/j.scitotenv.2017.11.108] [Medline: 29734606]
36. Athanasiou M, Fragkozidis G, Zarkogianni K, Nikita KS. Long short-term memory-based prediction of the spread of influenza-like illness leveraging surveillance, weather, and twitter data: model development and validation. J Med Internet Res 2023;25:e42519 [FREE Full text] [doi: 10.2196/42519] [Medline: 36745490]
37. Chen B, Sumi A, Wang L, Zhou W, Kobayashi N. Role of meteorological conditions in reported chickenpox cases in Wuhan and Hong Kong, China. BMC Infect Dis 2017;17(1):538 [FREE Full text] [doi: 10.1186/s12879-017-2640-1] [Medline: 28774264]
38. Pani SK, Lin N, RavindraBabu S. Association of COVID-19 pandemic with meteorological parameters over Singapore. Sci Total Environ 2020;740:140112 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.140112] [Medline: 32544735]
39. Choi SB, Ahn I. Forecasting seasonal influenza-like illness in South Korea after 2 and 30 weeks using Google Trends and influenza data from Argentina. PLoS One 2020;15(7):e0233855 [FREE Full text] [doi: 10.1371/journal.pone.0233855] [Medline: 32673312]

XSL•FO
RenderX

40. Chen S, Liu X, Wu Y, Xu G, Zhang X, Mei S, et al. The application of meteorological data and search index data in improving the prediction of HFMD: a study of two cities in Guangdong Province, China. Sci Total Environ 2019;652:1013-1021. [doi: 10.1016/j.scitotenv.2018.10.304] [Medline: 30380469]

Abbreviations

CO: carbon monoxide
DLNM: distributed nonlinear lag model
LSTM: long short-term memory
MAE: mean absolute error
MAPE: mean absolute percentage error
PCC: Pearson correlation coefficient
PM: particulate matter
PM2.5: particulate matter with a diameter of 2.5 μm or less
PM10: particulate matter with a diameter of 10 μm or less
RMSE: root-mean-square error
RR: relative risk
SO2: sulfur dioxide

Edited by A Schwartz; submitted 03.09.24; peer-reviewed by R Guo, T Zhang; comments to author 27.09.24; revised version received 04.10.24; accepted 04.12.24; published 06.02.25.

Please cite as:
Xiong X, Xiang L, Chang L, Wu IXY, Deng S
Forecasting the Incidence of Mumps Based on the Baidu Index and Environmental Data in Yunnan, China: Deep Learning Model Study
J Med Internet Res 2025;27:e66072
URL: https://www.jmir.org/2025/1/e66072
doi:10.2196/66072
PMID:

©Xin Xiong, Linghui Xiang, Litao Chang, Irene XY Wu, Shuzhen Deng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Spatiotemporal Characteristics and Influential Factors of Electronic Cigarette Web-Based Attention in Mainland China: Time Series Observational Study

Zhongmin Zhang[1], BMed; Hengyi Xu[1], BS; Jing Pan[2], BMed; Fujian Song[3], PhD; Ting Chen[1], PhD

[1]Healthy Hubei Development and Social Progress Research Center of the Key Research Base of Humanities and Social Sciences in Hubei Province, School of Public Health, Wuhan University of Science and Technology, Wuhan, China

[2]School of Xiangtao, Wuhan University of Science and Technology, Wuhan, China

[3]Norwich Medical School, University of East Anglia, Norwich, United Kingdom

Corresponding Author:
Ting Chen, PhD
Healthy Hubei Development and Social Progress Research Center of the Key Research Base of Humanities and Social Sciences in Hubei Province
School of Public Health
Wuhan University of Science and Technology
2 Huangjiahuxi Road
Hongshan District
Wuhan, 430065
China
Phone: 86 18120237582
Email: chent41@wust.edu.cn

Abstract

Background: The popularity of electronic cigarettes (e-cigarettes) has steadily increased, prompting a considerable number of individuals to search for relevant information on them. Previous e-cigarette infodemiology studies have focused on assessing the quality and reliability of website content and quantifying the impact of policies. In reality, most low-income countries and low- and middle-income countries have not yet conducted e-cigarette use surveillance. Data sourced from web-based search engines related to e-cigarettes have the potential to serve as cost-effective supplementary means to traditional monitoring approaches.

Objective: This study aimed to analyze the spatiotemporal distribution characteristics and associated sociodemographic factors of e-cigarette searches using trends from the Baidu search engine.

Methods: The query data related to e-cigarettes for 31 provinces in mainland China were retrieved from the Baidu index database from January 1, 2015, to December 31, 2022. Concentration ratio methods and spatial autocorrelation analysis were applied to analyze the temporal aggregation and spatial aggregation of the e-cigarette Baidu index, respectively. A variance inflation factor test was performed to avoid multicollinearity. A spatial panel econometric model was developed to assess the determinants of e-cigarette web-based attention.

Results: The daily average Baidu index for e-cigarettes increased from 53,234.873 in 2015 to 85,416.995 in 2021 and then declined to 52,174.906 in 2022. This index was concentrated in the southeastern coastal region, whereas the hot spot shifted to the northwestern region after adjusting for population size. Positive spatial autocorrelation existed in the per capita Baidu index of e-cigarettes from 2015 to 2022. The results of the local Moran's *I* showed that there were mainly low-low cluster areas of the per capita Baidu index, especially in the central region. Furthermore, the male-female ratio, the proportion of high school and above education, and the per capita gross regional domestic product were positively correlated with the per capita Baidu index for e-cigarettes. A higher urbanization rate was associated with a reduced per capita Baidu index.

Conclusions: With the increasing popularity of web-based searches for e-cigarettes, a targeted e-cigarette health education program for individuals in the northwest, males, rural populations, high school and above educated individuals, and high-income groups is warranted.

(J Med Internet Res 2025;27:e66446) doi:10.2196/66446

XSL•FO
RenderX

KEYWORDS

electronic cigarettes; Baidu index; web-based attention; spatiotemporal characteristics; China

Introduction

Electronic cigarettes (e-cigarettes), as emerging tobacco products, present a significantly appealing option to the public. In the European Union, 31.1% of current smokers, 10.8% of former smokers, and 2.3% of never smokers reported ever using e-cigarettes [1]. The world's first e-cigarette was invented in Guangzhou, China. To date, China remains the principal producer of e-cigarettes, responsible for 95% of global manufacturing [2]. In 2021, China's e-cigarette exports reached 101.5 billion RMB (US $15.91 billion; US $1=6.38 RMB), with a market size of 165.88 billion RMB (US $26 billion) [3]. Between 2015-2016 and 2018-2019, e-cigarette use among Chinese adults increased from 1.3% to 1.6% [4]. Similarly, the rate of e-cigarette use among middle school students rose from 1.2% in 2014 to 2.7% in 2019 [5].

The widespread marketing of e-cigarettes undoubtedly contributes significantly to this growth. In addition to conventional retail marketing tactics, e-cigarette manufacturers use web-based advertisement strategies to promote their merchandise [6]. On social media platforms, e-cigarettes are often portrayed as safer and healthier alternatives to traditional cigarettes [6-8]. Nonetheless, numerous studies have indicated that these products may elevate the risk of cardiovascular and respiratory diseases and adversely affect brain development and mental health [9,10]. Given the increasing prevalence of e-cigarette use and the persistent debate surrounding its health implications, an increasing number of individuals are turning to the web to access the latest information.

Web-based data offer a diverse array of perspectives and data points, with user query frequencies and concerns being well recorded. In contrast to data from traditional government communication channels, web-based data are typically anonymous, timely, and cost-effective [11,12]. Therefore, they were extensively used in disease incidence reporting, pandemic outbreak surveillance and forecasting, and public awareness analysis [13-15]. The seasonality and global nature of infectious diseases rendered them initial case studies for web-based surveillance. Tracking of influenza, HIV, and dengue outbreaks has been a continuous research hot spot [15,16]. Recently, the scope of web-based surveillance has expanded to encompass mental health, tobacco use, and noncommunicable diseases [12].

Scholarly inquiries into e-cigarette infodemiology have focused on trends in popularity, website content quality and reliability, policy impact assessments, and correlations with population use prevalence [8,17-22]. Ayers et al [17,18] used Google Trends to estimate the popularity of e-cigarettes in high-income countries such as the United States. Ghosh et al [20] suggested that web-based search data on e-cigarette–related topics in the United States and India correlate with their actual population-based prevalence. Another study conducted in the United Kingdom also confirmed this finding [19]. These studies demonstrated the great potential of using web-based search engine data to supplement and expand traditional channels for monitoring e-cigarette use. This is particularly important in low-income countries and low- and middle-income countries, where the majority have not started monitoring e-cigarette use and lack evidence to guide local policy and regulatory decisions [23,24].

To our knowledge, only 1 study has portrayed e-cigarette attention in China from 2011 to 2014 using search engine data [25]. It revealed regional differences in the Baidu search volume for e-cigarettes but did not delve into the long-term spatial and temporal evolution patterns, nor did it consider spatial effects when exploring the influencing factors. Spatiotemporal analysis can integrate multidimensional data, uncover spatiotemporal dynamic associations, and facilitate visual presentation. Previous research has used this method to explore web-based search characteristics for vaccines, migraine, and anxiety [26-28].

By December 2023, China's web user base had reached approximately 1.1 billion [29]. Baidu dominates the search engine market with a penetration rate of 90.9% [30]. Anchored in Baidu's vast search data, the Baidu index (BI) can quantify the search volume of specific terms in a certain geographical area within a given time range and convert it into a relative index value [31]. It largely reflects the search needs and awareness of Chinese web users and has been used for public health surveillance [26-28,31]. The purpose of this study was to use the BI to assess temporal and regional changes in web-based e-cigarette retrieval in mainland China and to identify factors associated with differences in public attention. The findings will contribute to enhancing and supplementing the traditional e-cigarette surveillance system in mainland China and provide a reference for the development and implementation of tobacco control policies and targeted health educational programs on e-cigarettes.

Methods

Data Sources

Web-Based Search Databases

The web-based search data related to e-cigarettes were derived from the official BI website. This platform allows researchers to capture search frequencies of different geographic regions using specific search terms over selected time frames. In this study, the corresponding Chinese characters for e-cigarettes and their synonyms were chosen as search terms. Considering the relatively low web-based search volume of e-cigarettes in China until 2015, the data sample period was set from January 1, 2015, to December 31, 2022. The geographical scope encompassed all 31 provinces in mainland China.

The daily average Baidu index (DBI) is the weighted sum of search frequency for search terms, calculated according to the daily search volume on Baidu. To further quantify the average number of searches per individual within each province, the per capita Baidu index (PBI) was defined as the ratio of the DBI to the year-end population.

XSL•FO
RenderX

Real-World Databases

To explore the influencing factors of web-based search volume, demographic and socioeconomic data were collected from the China Statistical Yearbook (2011-2022). The demographic variables included the male-female ratio, gross dependency ratio (the ratio of the nonworking-age individuals to the working-age population), and urbanization rate (the proportion of the urban population to the total population at the end of the year). The proportion of individuals with a high school education or above is the segment of the total population with at least a higher education, reflecting the educational level of the region. The per capita gross regional domestic product (GRDP), an indicator of economic development, was logarithmized in the analysis to eliminate the effects of heteroscedasticity and quantity stiffness. The descriptive statistical analysis of these variables is shown in Table S1 in Multimedia Appendix 1.

Statistical Analysis

Temporal Aggregation Analysis

The concentration ratio method was applied to analyze the temporal aggregation of the DBI for e-cigarettes [32]. The formula takes the following form:



(1)

where$R$ represents the degree of dispersion and $r_1$indicates the ratio of searches in month to the total searches for the year. M shows the degree of concentration, ranging from 0 to 1. A larger value of M indicates a stronger temporal aggregation and a value less than 0.3 suggests a relatively uniform time distribution.

Spatial Weight Matrix Definition

The spatial weight matrix is the main expression of spatial adjacency relationships. Its definition is a prerequisite for conducting spatial analysis. In our study, the widely used Rook adjacency method was adopted. This method assumes that if 2 provinces have a common boundary, they are considered adjacent with a weight of 1, otherwise, it is 0 [33]. Given that Hainan Province lacks a shared border with any neighboring provinces, we defined it as adjacent to the geographically closest Guangdong Province. The definition results of the spatial matrix Rook adjacency criterion are shown in Table S2 in Multimedia Appendix 1.

Spatial Autocorrelation Analysis

Spatial autocorrelation analysis was used to determine the spatial aggregation of the DBI for e-cigarettes [34]. It is generally categorized into global and local spatial analysis. We used the global Moran's I to analyze the spatial distribution pattern of the e-cigarette BI in each province of China, reflecting spatial dependence. The equation for the calculation is as follows:



(2)

Where w_{ij} represents the weight of the spatial relationship between observations i and j. The deviation of the attribute value of element i from its mean is denoted by . n is the total number of elements. The global Moran's I is between –1 and 1. I>0 suggests the presence of a positive spatial correlation, with values closer to 1 indicating a greater degree of spatial aggregation. Conversely, I<0 signifies a negative spatial correlation. When I equals 0, it is irrelevant.

Global spatial autocorrelation analysis fails to reveal the spatial relationships and evolutionary features within a region [35]. Therefore, the local Moran's I was applied to examine the correlation of provincial units and their neighboring provinces. It can identify the characteristics of aggregation and dissociation among districts. The local Moran's I is defined as follows:



(3)

The indicators within the formula carry similar meanings to those in the global Moran's I formula. The results of the local Moran's I analysis can be categorized into 4 significant types: high-high cluster, low-low cluster, high-low outlier, and low-high outlier. Specifically, a region with high values can be identified as a high-high cluster if its neighboring regions also have high values. If a regional unit with low values is surrounded by other regional units with low values, that unit may be classified as a low-value cluster. High-low outliers denote high-attention regions encircled by low-attention regions, whereas low-high outliers indicate low-attention regions surrounded by high-attention regions. The global and local Moran's I were computed using GeoDa 12.0 (Dr Luc Anselin's team). Regional distribution maps were drawn with ArcGIS 10.8 (Environmental Systems Research Institute).

Spatial Panel Econometric Model Description

Unlike traditional panel regression models, spatial panel econometric models consider the spatial correlation and heterogeneity among regions [36]. The classical spatial panel econometric analysis models include the spatial autoregressive regression model (SAR), spatial error model (SEM), and spatial Durbin model (SDM). These models differ in their assumptions regarding spatial interaction. SAR solely incorporates the spatial dependence of explanatory variables, SEM focuses exclusively on the spatial dependence of the random error term, and SDM integrates both effects. The basic formulas are defined as follows:



(4)

Where y_{it} is the dependent variable of province i at time t. x is the explanatory variable. α represents the intercept item. ρ is the spatial autocorrelation coefficient and w_{ij} is the spatial weight matrix. β is the vector of explanatory power, ε is the residual disturbance term, and λ is the spatial error coefficient. When λ=0 and β_2=0, the above model can directly degenerate into the SAR; when ρ=0 and β_2=0, the model can evolve into the SEM and when λ=0, the model transforms into the SDM.

XSL•FO
RenderX

Selection of a Suitable Spatial Panel Econometric Model

The Lagrange multiplier test was used to identify the presence of spatial effect [37]. The Wald test and likelihood ratio (LR) test were performed to evaluate whether SDM could be degraded into the SAR or SEM. The Hausman test was used to compare the suitability of fixed and random effects [38]. The LR time-space effect test was finally applied to determine the type of spatial effect. The selection process is shown in Figure S1 in Multimedia Appendix 1.

Table S3 in Multimedia Appendix 1 shows the results of the spatial panel model test. The Lagrange multiplier test suggested that the SDM was more appropriate than the SEM and the SAR. The Wald test and the LR test further confirmed that the SDM could not be degraded into the SAR and the SEM. The estimated values of the Hausman test rejected the null hypothesis, indicating that fixed effect was better than random effect in this study. In addition, the time-space double effect is more suitable for objective reality, as evidenced by the significant LR test results (LR-both/space=53.780; P<.001; LR-both/time=777.650; P<.001). Consequently, a time-space double fixed effects SDM was used to analyze the influencing factors of PBI for e-cigarettes. All tests were conducted using Stata/MP 16 (StataCorp). A 2-sided P value of <.1 was considered to indicate statistical significance.

Multicollinearity Test

Multicollinearity refers to the linear relationship between 2 or more independent variables. To address potential multicollinearity problems in this study, we conducted a variance inflation factor test. The results are shown in Table S4 in Multimedia Appendix 1. Notably, the variance inflation factor values for all variables did not exceed 10, indicating the absence of multicollinearity in our model.

Ethical Considerations

Data for this study were obtained from publicly available BI search data and statistical yearbooks. No human or animal subjects were involved. Data privacy and security were approved by the ethics committee of Wuhan University of Science and Technology (approval number 202 063).

Results

Temporal Characteristics

Table S5 in Multimedia Appendix 1 and Figure 1 show the temporal distribution characteristics of the DBI for e-cigarettes from January 2015 to December 2022. Over the 8-year period, the average DBI was 55,137.045. Generally, the DBI for e-cigarettes follows a fluctuating yet upward trajectory. The highest increase occurred in 2020-2021, with the DBI rising from 61,277.286 to 85,416.995. In 2022, the DBI decreased to 52,174.906. During the study period, the concentration ratios of DBI varied between 0.136 and 0.417. Specifically, the maximum values were recorded in 2019 (M=0.417) and 2020 (M=0.399). In both of these years, the peak DBI values appeared in October, November, and December.

Figure 1. Seasonal plot of the daily average Baidu index for electronic cigarettes (2015-2022).

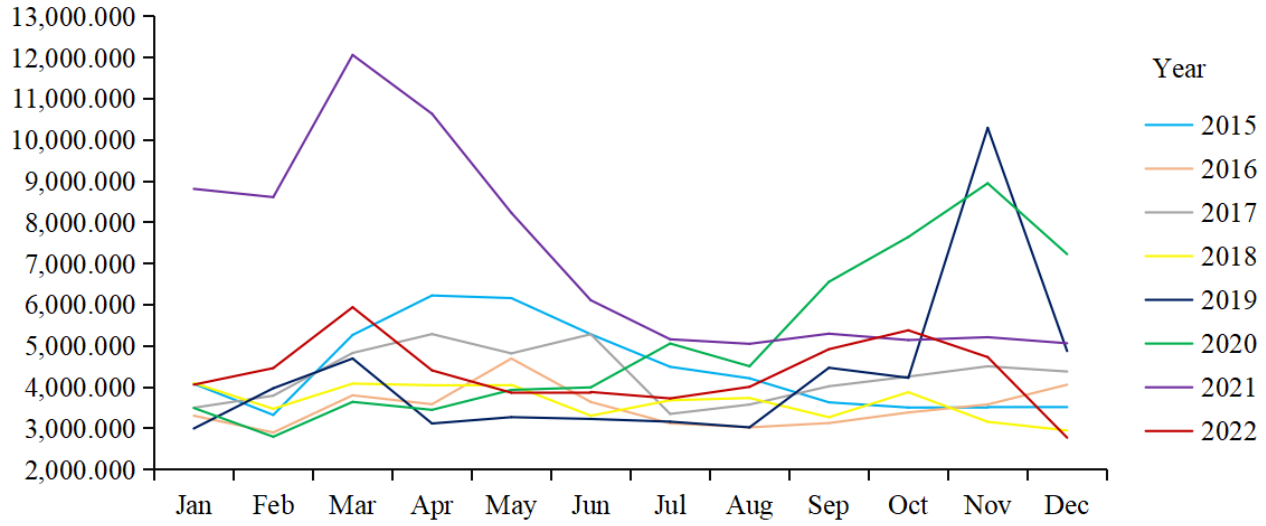

Spatial Characteristics

Figures 2 and 3 depict the DBI and the PBI of e-cigarettes in each Chinese province from 2015 to 2022, respectively. Provinces with higher DBI for e-cigarettes include Shandong, Jiangsu, Zhejiang, and Guangdong, all located on the southeast coast of China. These regions consistently rank at the top, with annual DBI exceeding 450. Qinghai, Tibet, and Ningxia maintained the lowest DBI levels. The geographical pattern of the PBI contrasts significantly with that of the DBI, as the former is concentrated in Tibet, Qinghai, and Ningxia. Notably, the PBI for e-cigarettes in Beijing far surpassed that in neighboring provinces.

Table 1 indicates the global Moran's I of the PBI for e-cigarettes from 2015 to 2022. Throughout the period, all years had a global Moran's I value greater than 0, ranging between 0.151 and 0.265. The z scores revealed that spatial clustering of the PBI for e-cigarettes was statistically significant. Despite fluctuations in the global Moran's I and z values, the clustering pattern and positive spatial autocorrelation remained stable. The local Moran's I findings for the PBI of e-cigarettes are detailed in

XSL•FO
RenderX

Table 2. Low-low clusters exist in the central region of China. Guizhou and Hubei consistently demonstrated significant low-low clusters over the 8-year period. Tianjin displayed a high-high cluster only in 2015. Furthermore, Xinjiang, located in the northwestern region, was identified as a low-high outlier area in 2016, 2017, and 2022.

Figure 2. The daily average Baidu index of electronic cigarettes by province in China (2015-2022).

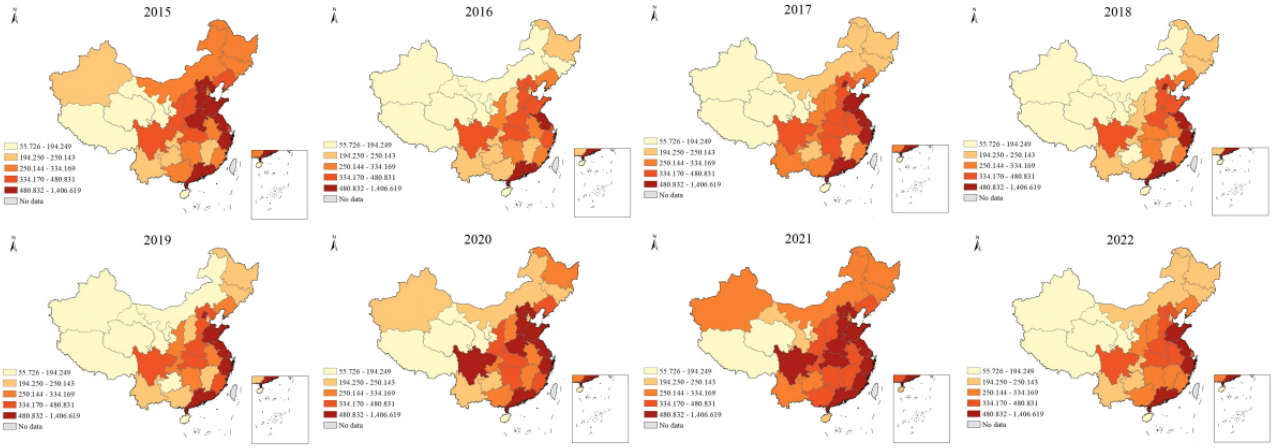

Figure 3. The per capita Baidu index of electronic cigarettes by province in China (2015-2022).

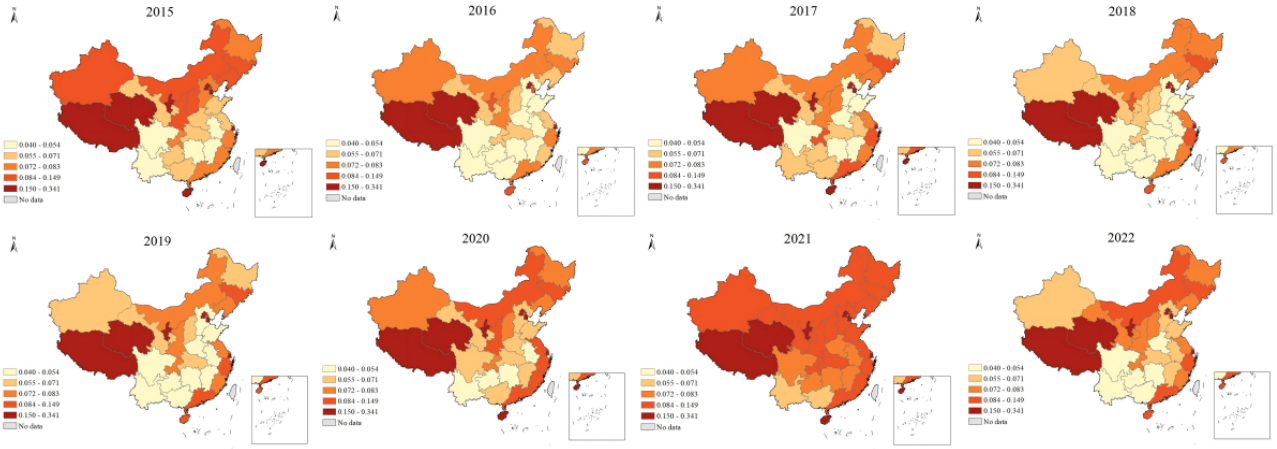

Table 1. The global Moran's I of the per capita Baidu index for electronic cigarettes (2015-2022)[a].

Year	Global Moran's *I*	*z* value	*P* value
2015	0.265	2.612	.009
2016	0.155	1.574	.07
2017	0.151	1.536	.08
2018	0.196	1.999	.04
2019	0.155	1.566	.07
2020	0.186	1.884	.04
2021	0.174	1.805	.05
2022	0.166	1.692	.06

[a]Significance levels were determined by *P* value tests of standardized *z* values.

XSL•FO
RenderX

Table 2. Local spatial aggregation of the per capita Baidu index for electronic cigarettes (2015-2022)[a].

Year	High-high cluster	High-low outlier	Low-high outlier	Low-low cluster
2015	Tianjin	—[b]	Hebei	Guizhou, Anhui, Chongqing, Jiangxi, Hubei, Hunan, and Guangxi
2016	—	—	Xinjiang	Guizhou, Anhui, Hubei, and Shandong
2017	—	—	Xinjiang	Guizhou, Chongqing, Hubei, Shandong, and Henan
2018	—	—	—	Guizhou, Chongqing, Hubei, Hunan, Henan, and Guangxi
2019	—	—	—	Guizhou, Chongqing, Hubei, Shandong, and Henan
2020	—	—	—	Guizhou, Chongqing, Hubei, Hunan, and Guangxi
2021	—	—	—	Guizhou, Chongqing, Hubei, Hunan, and Guangxi
2022	—	—	Xinjiang	Guizhou, Chongqing, Hubei, Hunan, Henan, and Guangxi

[a]Provinces with nonsignificant local Moran's *I* are not listed.

[b]Not available.

Influencing Factors

Table 3 illustrates the SDM results of the PBI for e-cigarettes. The male-female ratio, the proportion of high school and above education, and the GRDP were positively correlated with the PBI for e-cigarettes. A higher urbanization rate was associated with a reduced PBI. The PBI for e-cigarettes had significant positive spatial spillover effects (spatial ρ=0.467; *P*<.001). The proportion of high school and above education significantly impacted the PBI for e-cigarettes in adjacent proximity. To further analyze these spillover effects, the SDM was decomposed into direct and indirect effects using partial differential decomposition. In terms of the direct effect, there was a significant association between the PBI and male-female ratio, urbanization rate, and GRDP in the local region. The indirect effects reveal a significant negative association between the proportion of high school education and above and the PBI for e-cigarettes in neighboring areas. Moreover, both the urbanization rate and the proportion of high school and above education negatively influence the PBI of e-cigarettes across all areas.

XSL•FO
RenderX

Table 3. Spatial Durbin model results of the per capita Baidu index for electronic cigarettes.

	Coefficient	95% CI	*P* value
Main[a]			
Male-female ratio	0.057	0.008 to 0.106	.02
Gross dependency ratio	–0.002	–0.071 to 0.067	.96
Urbanization rate	–0.145	–0.282 to –0.009	.04
Proportion of education at high school and above	0.069	0.002 to 0.137	.04
Ln (per capita gross regional domestic product)[b]	0.017	0.001 to 0.032	.03
Wx[c]			
Male-female ratio	0.022	–0.077 to 0.122	.66
Gross dependency ratio	–0.042	–0.181 to 0.098	.56
Urbanization rate	–0.094	–0.381 to 0.193	.52
Proportion of education at high school and above	–0.203	–0.342 to –0.064	.004
Ln (per capita gross regional domestic product)	–0.009	–0.033 to 0.016	.49
Spatial ρ[d]	0.467	0.333 to 0.602	<.001
Direct effect[e]			
Male-female ratio	0.065	0.008 to 0.122	.03
Gross dependency ratio	–0.008	–0.08 to 0.064	.82
Urbanization rate	–0.161	–0.287 to –0.035	.01
Proportion of education at high school and above	0.046	–0.021 to 0.114	.18
Ln (per capita gross regional domestic product)	0.016	0.001 to 0.032	.03
Indirect effect[f]			
Male-female ratio	0.090	–0.094 to 0.274	.34
Gross dependency ratio	–0.070	–0.321 to 0.181	.59
Urbanization rate	–0.295	–0.756 to 0.167	.21
Proportion of education at high school and above	–0.288	–0.523 to –0.053	.02
Ln (per capita gross regional domestic product)	–0.001	–0.04 to 0.037	.94
Total effect[g]			
Male-female ratio	0.155	–0.067 to 0.376	.17
Gross dependency ratio	–0.078	–0.365 to 0.209	.59
Urbanization rate	–0.456	–0.938 to 0.026	.06
Proportion of education at high school and above	–0.242	–0.503 to 0.02	.07
Ln (per capita gross regional domestic product)	0.015	–0.029 to 0.059	.50

[a]Main denotes the impact coefficient of the variable on the local region.

[b]Ln (per capita gross regional domestic product) indicates the logarithm of per capita gross regional domestic product.

[c]Wx refers to the spatial spillover coefficient of the variable to adjacent regions.

[d]Spatial ρ represents spatial autocorrelation coefficients.

[e]Direct effect is the influence of the change in the local independent variable on the local dependent variable.

[f]Indirect effect is the influence of the change in the local independent variable on the adjacent region.

[g]The coefficient of total effects is the sum of direct and indirect effects.

XSL•FO
RenderX

Discussion

Principal Results and Comparison With Prior Work

In this study, we used the BI to explore the spatial and temporal distribution of web-based searches for e-cigarettes in mainland China from 2015 to 2022 and identified the complex associations between the e-cigarette searches and demographic characteristics, economic development level, and education attainment.

The upward trend in e-cigarette–related BI observed in our study aligns with evidence from high-income countries [17,18]. This may be attributed to the increasing prevalence of e-cigarette usage, improved health literacy, and increased web penetration [4,29,39]. Notably, a pronounced spike in search queries was detected in November 2019. A possible explanation for this temporary surge is "policy shock" [22,40]. At the beginning of that month, the Chinese government successively issued the "Notice on Further Protection of Minors from E-cigarettes" and the "Notice on Further Strengthening Tobacco Control Work Among Young People." These 2 notices explicitly prohibited manufacturers from advertising and retailers from selling e-cigarettes on the web to minors [41,42]. This prompted the public and the media to focus on web-based information about e-cigarettes. In addition, the COVID-19 pandemic might be a crucial factor for the rise in e-cigarette web-based information-seeking behavior in 2019 and 2020. The implementation of stay-at-home orders and temporary retail closures fostered an environment conducive to heightened web-based engagement [43]. During this period, 18.2%-19.8% of e-cigarette users shifted their purchasing habits from retail to web-based stores [44].

In the spatial dimension, significant regional differences exist in e-cigarette web-based attention across provinces in mainland China. The DBI for e-cigarettes is high in eastern coastal regions. Although national epidemiological data on e-cigarettes are lacking, this geographic distribution may reflect the socioeconomic and population conditions. These eastern coastal regions are characterized by dense populations and well-developed economies [45]. Residents here boasted better web-based access and public health awareness [27,31]. Among the provinces, Guangdong holds the highest DBI rankings for e-cigarettes [46]. Besides demographic and economic considerations, Guangdong was the pioneer manufacturing and production base for e-cigarettes in China and the world [47]. Regarding provinces with higher PBI, Tibet and Qinghai stand out. These 2 western provinces are distinguished by their high altitudes and unique climatic conditions In such environments, traditional tobacco combustion may be impeded. E-cigarettes show greater adaptability due to their closed system and ignition-free operation [48]. It is noteworthy that Guizhou and Hubei Province in the central region were identified as a significant low-low cluster over the 8-year period. These are both large traditional tobacco-consuming provinces, where smokers may have deep-rooted consumption habits of traditional cigarettes and are been less receptive to emerging e-cigarette products [49].

The Influencing Factors of E-Cigarette Web-Based Attention Are Diverse

First, males reported a greater frequency of e-cigarette web-based inquiries than females did. A study indicated that the proportion of males who had used and heard of e-cigarettes surpassed that of females in China [50]. Meanwhile, males are more likely to obtain e-cigarette information from online platforms and forums [51].

Second, we found a positive correlation between the ratio of high school and above education and the e-cigarette PBI. The previous study has also revealed that a greater percentage of individuals with a higher level of education had both used and been informed about e-cigarettes than did those with lower educational attainment [50]. An individual's education level is a critical determinant of public health information literacy [52]. With the accumulation of knowledge, the understanding of e-cigarette potential harm has gradually increased. Furthermore, highly educated groups tend to be more socially active and susceptible to peer pressure.

As expected, GRDP was significantly positively correlated with web-based attention. This finding was in line with the results on regional differences. The growth of GRDP is often accompanied by improved infrastructure, comprehensive services, and increased digitalization [26,27].

However, our study found a negative correlation between the urbanization rate and the PBI of e-cigarettes. On the one hand, in highly urbanized cities, such as Beijing and Shanghai, the prohibition of smoking in public places was effectively enforceable [53]. On the other hand, in regions with higher urbanization rates, consumers' awareness of e-cigarettes may have matured and their preferences have been largely established. New users' curiosity-driven search behavior was substantially reduced. Moreover, there are more e-cigarette retail shops in these areas. Consumers are more likely to have on-site experiences instead of conducting web-based searches.

Limitations

Although this study provided valuable evidence on the status and trends of e-cigarette popularity in China, there are still several limitations. First, web-based search query activities are inherently biased toward groups with higher socioeconomic levels and educational attainment. It fails to capture some important populations for e-cigarette use, especially adolescents. Second, the data available on the BI platform are limited to search terms that meet established user access thresholds. As a result, less common or infrequently used expressions may be excluded from trend analysis. Third, even though Baidu is the primary search engine among Chinese users, a minority still use alternative searches such as Google and Bing. Our study did not cover these populations. Finally, the BI data we used were collected at the provincial level. Further research should consider smaller geographic units to obtain more accurate evidence.

Conclusions

Public interest in e-cigarettes is increasing in China. There are notable differences in the PBI for e-cigarettes among provinces,

XSL•FO
RenderX

with northwest and economically prosperous coastal areas becoming hot spots. In addition, the male-female ratio, urbanization ratio, education attainment, and level of regional economic development were important factors contributing to the web-based attention to e-cigarettes. These findings underscore the necessity of devising targeted e-cigarette health education programs for individuals in the northwest, males, rural populations, high school and above educated individuals, and those with higher incomes. For less concerned groups, the focus should be on educating them about the accuracy of the information they access, thereby preventing the spread of misinformation. Our study not only contributes to supplementing and enhancing the e-cigarette use monitoring systems in China but also provides a valuable reference method for other low-income countries and low- and middle-income countries with similar patterns of web use.

Acknowledgments

This study was supported by “The 14th Five Year Plan” Hubei Provincial advantaged and characteristic disciplines (groups) project of Wuhan University of Science and Technology (grant number C0202).

Authors' Contributions

ZZ and TC conceptualized the study. ZZ and HX designed the methodology and conducted the data analysis. ZZ and JP were involved in data collection. ZZ drafted the original draft. TC and FS reviewed and edited the draft. All the authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and tables on the methodology and results in the manuscript.
[DOCX File , 29 KB - jmir_v27i1e66446_app1.docx]

References

1. Farsalinos KE, Poulas K, Voudris V, Le Houezec J. Electronic cigarette use in the European Union: analysis of a representative sample of 27 460 Europeans from 28 countries. Addiction 2016;111(11):2032-2040. [doi: 10.1111/add.13506] [Medline: 27338716]
2. Eriksen M, Mackay J, Schluger N, Islami F, Drop J. The Tobacco Atlas. 5th ed. Atlanta, GA: American Cancer Society; 2015.
3. iiMedia Report. Research on development of global e-cigarette industry from 2021 to 2022. 2022. URL: https://www.iimedia.cn/c400/83182.html [accessed 2024-12-13]
4. Zhao Z, Zhang M, Wu J, Xu X, Yin P, Huang Z, et al. E-cigarette use among adults in China: findings from repeated cross-sectional surveys in 2015-16 and 2018-19. Lancet Public Health 2020;5(12):e639-e649 [FREE Full text] [doi: 10.1016/S2468-2667(20)30145-6] [Medline: 33271077]
5. Chinese Center for Disease Control and Prevention. The results of the 2019 China Middle School Student Tobacco Survey were released. 2020. URL: https://www.chinacdc.cn/jkyj/yckz/jcpg/202005/t20200531_296450.html [accessed 2024-12-11]
6. Collins L, Glasser AM, Abudayyeh H, Pearson JL, Villanti AC. E-cigarette marketing and communication: how e-cigarette companies market e-cigarettes and the public engages with e-cigarette information. Nicotine Tob Res 2019;21(1):14-24 [FREE Full text] [doi: 10.1093/ntr/ntx284] [Medline: 29315420]
7. de Andrade M, Hastings G, Angus K. Promotion of electronic cigarettes: tobacco marketing reinvented? BMJ 2013;347:f7473. [doi: 10.1136/bmj.f7473] [Medline: 24361526]
8. Chen T, Gentry S, Qiu D, Deng Y, Notley C, Cheng G, et al. Online information on electronic cigarettes: comparative study of relevant websites from Baidu and Google search engines. J Med Internet Res 2020;22(1):e14725 [FREE Full text] [doi: 10.2196/14725] [Medline: 32012069]
9. Callahan-Lyon P. Electronic cigarettes: human health effects. Tob Control 2014;23 Suppl 2(Suppl 2):ii36-ii40 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051470] [Medline: 24732161]
10. Dinakar C, O’Connor GT. The health effects of electronic cigarettes. N Engl J Med 2016;375(14):1372-1381. [doi: 10.1056/nejmra1502466] [Medline: 27705269]
11. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
12. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): systematic review. J Med Internet Res 2020;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]

XSL•FO
RenderX

13. Wang Z, He J, Jin B, Zhang L, Han C, Wang M, et al. Using Baidu index data to improve chickenpox surveillance in Yunnan, China: infodemiology study. J Med Internet Res 2023;25:e44186 [FREE Full text] [doi: 10.2196/44186] [Medline: 37191983]
14. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google trends data: infodemiology study. J Med Internet Res 2020;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
15. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
16. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
17. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med 2011;40(4):448-453. [doi: 10.1016/j.amepre.2010.12.007] [Medline: 21406279]
18. Ayers JW, Althouse BM, Allem J, Leas EC, Dredze M, Williams RS. Revisiting the rise of electronic nicotine delivery systems using search query surveillance. Am J Prev Med 2016;50(6):e173-e181 [FREE Full text] [doi: 10.1016/j.amepre.2015.12.008] [Medline: 26876772]
19. Sindhoo Z, Sindhoo S, Ghosh A, Kaur S, Soyiri I, Ahmadi K. Online interest for electronic cigarettes using google trends in the UK: a correlation analysis. Subst Use Misuse 2023;58(14):1791-1797 [FREE Full text] [doi: 10.1080/10826084.2023.2247056] [Medline: 37671780]
20. Ghosh A, Kaur S, Roub F. Use and interest of electronic nicotine delivery systems (ENDS): assessing the validity of Google Trends. Am J Drug Alcohol Abuse 2021;47(4):414-421. [doi: 10.1080/00952990.2021.1944171] [Medline: 34187262]
21. Foxon F. Estimating e-cigarette use prevalence among US adolescents using vaping-related online search trends. Subst Use Misuse 2021;56(10):1559-1563. [doi: 10.1080/10826084.2021.1936054] [Medline: 34110977]
22. Boakye E, Dzaye O, Erhabor J, Osuji N, Obisesan O, Osei AD, et al. Impact of the food and drug administration enforcement policy on flavored e-cigarettes on the online popularity of disposable e-cigarettes: analyses of Google search query data. BMC Public Health 2022;22(1):1937 [FREE Full text] [doi: 10.1186/s12889-022-14367-3] [Medline: 36258175]
23. World Health Organization. WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. 2023. URL: https://www.who.int/publications/i/item/9789240077164 [accessed 2023-12-11]
24. Yoong SL, Hall A, Leonard A, McCrabb S, Wiggers J, Tursan d'Espaignet E, et al. Prevalence of electronic nicotine delivery systems and electronic non-nicotine delivery systems in children and adolescents: a systematic review and meta-analysis. Lancet Public Health 2021;6(9):e661-e673. [doi: 10.1016/s2468-2667(21)00106-7]
25. He G, Lin X, Ju G, Chen Y. Mapping public concerns of electronic cigarettes in China. J Psychoactive Drugs 2020;52(1):13-19. [doi: 10.1080/02791072.2019.1707334] [Medline: 31888426]
26. Hu F, Qiu L, Xia W, Liu C, Xi X, Zhao S, et al. Spatiotemporal evolution of online attention to vaccines since 2011: an empirical study in China. Front Public Health 2022;10:949482 [FREE Full text] [doi: 10.3389/fpubh.2022.949482] [Medline: 35958849]
27. Lin L, Zhu M, Qiu J, Li Q, Zheng J, Fu Y, et al. Spatiotemporal distribution of migraine in China: analyses based on Baidu index. BMC Public Health 2023;23(1):1958 [FREE Full text] [doi: 10.1186/s12889-023-16909-9] [Medline: 37817123]
28. Xie T, Huang Z, Tan Y, Tan T. Analysis of the situations and influencing factors of public anxiety in China: based on Baidu index data. Front Public Health 2024;12:1360119 [FREE Full text] [doi: 10.3389/fpubh.2024.1360119] [Medline: 38721539]
29. China Internet Network Information Center. The 53rd statistical report on the development of China's internet. 2024 Mar 22. URL: https://www.cnnic.net.cn/n4/2024/0322/c88-10964.html [accessed 2024-06-01]
30. China Internet Network Information Center. 2019 Research report on Chinese internet users' search engine usage. URL: https://www.cnnic.cn/n4/2022/0401/c122-1110.html [accessed 2024-06-01]
31. Wei S, Ma M, Wen X, Wu C, Zhu G, Zhou X. Online public attention toward premature ejaculation in mainland China: infodemiology study using the Baidu index. J Med Internet Res 2021;23(8):e30271 [FREE Full text] [doi: 10.2196/30271] [Medline: 34435970]
32. Lin JJ, Zhang TW, Li XY. [Research progress on spatiotemporal clustering of disease] [Article in Chinese]. Zhonghua Liu Xing Bing Xue Za Zhi 2020;41(7):1165-1170. [doi: 10.3760/cma.j.cn112338-20190806-00582] [Medline: 32741189]
33. Curts H. Adjacency table method of deriving minimal sums. IEEE Trans Comput 1977;C-26(11):1136-1141. [doi: 10.1109/tc.1977.1674760]
34. Getis A. Spatial autocorrelation. In: Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Berlin, Heidelberg: Springer; 2009.
35. Anselin L. Local indicators of spatial association—LISA. Geogr Anal 2010;27(2):93-115. [doi: 10.1111/j.1538-4632.1995.tb00338.x]
36. Elhorst JP. Spatial Econometrics: From Cross-Sectional Data to Spatial Panels. Berlin, Heidelberg: Springer; 2014.
37. Anselin L, Arribas-Bel D. Spatial fixed effects and spatial dependence in a single cross‐section. Pap Reg Sci 2013;92(1):3-18. [doi: 10.1111/j.1435-5957.2012.00480.x]

XSL•FO
RenderX

38. Bole V, Rebec P. Bootstrapping the Hausman test in panel data models. Commun Stat Simulation Comput 2013;42(3):650-670. [doi: 10.1080/03610918.2011.650261]
39. Regular press release. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/xcs/s7847/201908/2a77e09aed9d427ab6d4e472095ba020.shtml [accessed 2024-05-20]
40. Goniewicz ML, Kosmider L, Delijewski M, Knysak J, Ochota P, Sobczak A. The impact of the 2010 Polish smoke-free legislation on the popularity and sales of electronic cigarettes. Eur J Public Health 2014;24(3):471-473 [FREE Full text] [doi: 10.1093/eurpub/ckt214] [Medline: 24424581]
41. Notice on further protection of minors from e-cigarettes. Stata Tobacco Monopoly Administration. 2019. URL: https://www.gov.cn/xinwen/2019-11/01/content_5447612.htm [accessed 2024-05-20]
42. Notice on further strengthening tobacco control work. National Health Commission of the People's Republic of China. 2019. URL: https://www.gov.cn/xinwen/2019-11/08/content_5450004.htm [accessed 2024-05-20]
43. Mowbray H. Letter from China: covid-19 on the grapevine, on the internet, and in commerce. BMJ 2020;368:m643. [doi: 10.1136/bmj.m643] [Medline: 32075783]
44. Gaiha SM, Lempert LK, Halpern-Felsher B. Underage youth and young adult e-cigarette use and access before and during the coronavirus disease 2019 pandemic. JAMA Netw Open 2020;3(12):e2027572 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.27572] [Medline: 33270127]
45. Li J, Lu D, Xu C, Li Y, Chen M. Spatial heterogeneity and its changes of population on the two sides of Hu Line [Article in Chinese]. Acta Geographica Sinica 2017;72(1):10-160. [doi: 10.11821/dlxb201701012]
46. Huang J, Zheng R, Emery S. Assessing the impact of the national smoking ban in indoor public places in China: evidence from quit smoking related online searches. PLoS One 2013;8(6):e65577 [FREE Full text] [doi: 10.1371/journal.pone.0065577] [Medline: 23776504]
47. Henningfield JE, Zaatari GS. Electronic nicotine delivery systems: emerging science foundation for policy. Tob Control 2010;19(2):89-90. [doi: 10.1136/tc.2009.035279] [Medline: 20378582]
48. Trtchounian A, Williams M, Talbot P. Conventional and electronic cigarettes (e-cigarettes) have different smoking characteristics. Nicotine Tob Res 2010;12(9):905-912. [doi: 10.1093/ntr/ntq114] [Medline: 20644205]
49. Chinese Center for Disease Control and Prevention. Map of data on smoking prevalence among aged 15 and over by region, 2020. 2019. URL: https://www.chinacdc.cn/jkyj/yckz/jcpg/202203/t20220316_296452.html [accessed 2024-12-17]
50. Xiao L, Parascandola M, Wang C, Jiang Y. Perception and current use of e-cigarettes among youth in China. Nicotine Tob Res 2019;21(10):1401-1407 [FREE Full text] [doi: 10.1093/ntr/nty145] [Medline: 30053201]
51. Piñeiro B, Correa JB, Simmons VN, Harrell PT, Menzie NS, Unrod M, et al. Gender differences in use and expectancies of e-cigarettes: online survey results. Addict Behav 2016;52:91-97 [FREE Full text] [doi: 10.1016/j.addbeh.2015.09.006] [Medline: 26406973]
52. Xu C, Zhang X, Wang Y. Mapping of health literacy and social panic via web search data during the COVID-19 public health emergency: infodemiological study. J Med Internet Res 2020;22(7):e18831 [FREE Full text] [doi: 10.2196/18831] [Medline: 32540844]
53. Lin H, Chang C, Liu Z, Zheng Y. Subnational smoke-free laws in China. Tob Induc Dis 2019;17:78 [FREE Full text] [doi: 10.18332/tid/112665] [Medline: 31772556]

Abbreviations

BI: Baidu index
DBI: daily average Baidu index
e-cigarette: electronic cigarette
GRDP: per capita gross regional domestic product
HIC: high-income country
LR: likelihood ratio
PBI: per capita Baidu index
SAR: spatial autoregressive regression model
SDM: spatial Durbin model
SEM: spatial error model

XSL•FO
RenderX

Edited by A Mavragani; submitted 13.09.24; peer-reviewed by D Li, G Jin; comments to author 04.12.24; revised version received 23.12.24; accepted 29.12.24; published 10.02.25.

Please cite as:
Zhang Z, Xu H, Pan J, Song F, Chen T
Spatiotemporal Characteristics and Influential Factors of Electronic Cigarette Web-Based Attention in Mainland China: Time Series Observational Study
J Med Internet Res 2025;27:e66446
URL: https://www.jmir.org/2025/1/e66446
doi:10.2196/66446
PMID:39928402

©Zhongmin Zhang, Hengyi Xu, Jing Pan, Fujian Song, Ting Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analyzing Themes, Sentiments, and Coping Strategies Regarding Online News Coverage of Depression in Hong Kong: Mixed Methods Study

Sihui Chen[1], BA, MSc; Cindy Sing Bik Ngai[1], BA, MPhil, PhD; Cecilia Cheng[2], BSocSci, PhD; Yangna Hu[1], BA, MA, MEd

[1]Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hung Hom, Kowloon, China (Hong Kong)

[2]Social and Health Psychology Laboratory, Department of Psychology, University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Cindy Sing Bik Ngai, BA, MPhil, PhD
Department of Chinese and Bilingual Studies
Hong Kong Polytechnic University
AG502
Hung Hom, Kowloon
China (Hong Kong)
Phone: 852 27667465
Email: cindy.sb.ngai@polyu.edu.hk

Abstract

Background: Depression, a highly prevalent global mental disorder, has prompted significant research concerning its association with social media use and its impact during Hong Kong's social unrest and COVID-19 pandemic. However, other mainstream media, specifically online news, has been largely overlooked. Despite extensive research conducted in countries, such as the United States, Australia, and Canada, to investigate the latent subthemes, sentiments, and coping strategies portrayed in depression-related news, the landscape in Hong Kong remains unexplored.

Objective: This study aims to uncover the latent subthemes presented in the online news coverage of depression in Hong Kong, examine the sentiment conveyed in the news, and assess whether coping strategies have been provided in the news for individuals experiencing depression.

Methods: This study used natural language processing (NLP) techniques, namely the latent Dirichlet allocation topic modeling and the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis, to fulfill the first and second objectives. Coping strategies were rigorously assessed and manually labeled with designated categories by content analysis. The online news was collected from February 2019 to May 2024 from Hong Kong mainstream news websites to examine the latest portrayal of depression, particularly during and after the social unrest and the COVID-19 pandemic.

Results: In total, 2435 news articles were retained for data analysis after the news screening process. A total of 7 subthemes were identified based on the topic modeling results. *Societal system*, *law enforcement*, *global recession*, *lifestyle, leisure*, *health issues*, and *US politics* were the latent subthemes. Moreover, the overall news exhibited a slightly positive sentiment. The correlations between the sentiment scores and the latent subthemes indicated that the societal system, law enforcement, health issues, and US politics revealed negative tendencies, while the remainder leaned toward a positive sentiment. The coping strategies for depression were substantially lacking; however, the categories emphasizing *information on skills and resources* and *individual adjustment* to cope with depression emerged as the priority focus.

Conclusions: This pioneering study used a mixed methods approach where NLP was used to investigate latent subthemes and underlying sentiment in online news. Content analysis was also performed to examine available coping strategies. The findings of this research enhance our understanding of how depression is portrayed through online news in Hong Kong and the preferable coping strategies being used to mitigate depression. The potential impact on readers was discussed. Future research is encouraged to address the mentioned implications and limitations, with recommendations to apply advanced NLP techniques to a new mental health issue case or language.

(J Med Internet Res 2025;27:e66696) doi:10.2196/66696

XSL•FO
RenderX

KEYWORDS

online news coverage; depression; natural language processing; NLP; latent Dirichlet allocation; LDA; sentiment; coping strategies; content analysis

Introduction

Background

Depression is known as one of the world's most common mental disorders, with an estimated 5% of adults experiencing it around the globe [1]. The disorder involves a depressed mood, loss of interest in activities, feeling excessive guilt or tiredness, being disconnected from family and friends, and thoughts about dying or suicide that affect all aspects of life. According to the mental health review report [2] released by the Health Bureau in Hong Kong, approximately 1 in 7 Hong Kongers will experience a mental disorder at any given point in life.

Since the outbreak of civil unrest and the COVID-19 pandemic in Hong Kong, a surging number of studies have emerged to examine the intensity of local people's depression, particularly among adolescents, older adults, students, and health care workers [3-6]. In particular, these studies have been widely associated with social media use and its effect compared to other digital media use [7-11]. Although the online press is one of the most widely consumed digital media in Hong Kong, existing research on online news coverage of depression has primarily been conducted in the United States, Australia, Canada, and China [12-17]. Studies have rarely explored online news coverage of depression in Hong Kong, based on available research.

The consumption of news coverage on depression can have both beneficial and detrimental effects on its readers. According to the protection and motivation theory, which has been applied in investigating different health-related behaviors, the underpinned threat and coping appraisal shape individuals' belief of the potential dangers to their lives and the motivation to practice protective behaviors via communication [18,19]. Thereby, positively, news coverage can help to raise awareness, increase understanding of mental disorders, and may facilitate the intention of professional help-seeking behavior [18,20]. Negatively, news coverage can also facilitate misunderstanding of depression among the public, exacerbating the stigma regarding people with mental disorders and leading to undertreatment [12,16,21,22]. The impact of news consumption has been acknowledged, and in depression-related news, diverse subthemes have also been investigated.

According to research on US television news coverage, the primary latent subthemes associated with individual and family-related depression stories include entertainment and celebrity depression, personal experiences, crime, and public policy [23]. Most of the news reports were also found to be largely negative in tone and often occurred with the subthemes of violence and crime [16]. This was likely because negative news can disproportionally grab the readers' attention, while positive news is perceived as having minimal impact [24]. Such a trend of news coverage is thereby harmful to populations with depression as it may reinforce social stigma and decrease public support for individuals with mental health disorders [16].

Conversely, in other research, which analyzed the contents of Canadian print news, the primary theme was raising awareness with the subthemes of addressing policy issues or lack of resources, military mental health issues, workplace issues, stigma, and so on, followed by the theme of research advances encompassing subthemes of treatment discoveries, etiology or prevention, and personal stories of recovery [25]. These kinds of positive subthemes may bring a positive effect of news exposure, which leads to favorable outcomes on mental health, particularly in circumstances where access to social support is restricted [26]. It is a fact that depression is 1 of the 2 most frequent mental disorder diagnoses In Hong Kong and has continued to deteriorate in recent years [3,27]. Media coverage of depression may encourage positive outcomes such as enhancing social support for individuals with depression or otherwise lead to negative outcomes, such as strengthening the public stigma toward the population with depression and their self-stigma [15,16,22,26]. Investigating the underlying subthemes of online news could aid in comprehending the positive and negative impacts of news experienced by its readers and potential mental health outcomes on populations with depression. However, to the extent of available studies, limited research has investigated the subthemes associated with online news coverage of depression in Hong Kong.

Research Questions

This study aims to fill this gap by exploring the latent subthemes in the news, which constitutes our first research question (RQ): What are the latent subthemes in the online news coverage of depression in Hong Kong? (RQ1)

In contrast, the sentiment of the news, which can be categorized as positive, negative, or neutral, has a direct effect on the stigmatization or destigmatization of people with depression or mental disorders [28,29]. Compared to print media, online news coverage was found to have a more positive sentiment in Canada and Chile, which helps to raise awareness, increase rehabilitation and recovery, and reduce the stigmatization of populations with depression [29,30]. Conversely, another study in the United Kingdom found that half of the articles about mental health expressed a negative sentiment and were closely connected with subthemes of violence, danger, and criminality, which reinforced stigma toward individuals with depression [31]. Stigmatization of people with depression usually labels them as having bad character, being violent, and blaming their conditions as a personal weakness, consequently, adding more obstacles when they want to seek social support [13,21,25]. People who are experiencing depression may also internalize the stereotypes, leading to diminished self-esteem, reduced quality of life, reluctance to seek help, and an increased likelihood of engaging in suicidal behavior [22,25].

Therefore, the sentiment of online news can play a significant role in reproducing or challenging the stigma around depression [32]. Within the context of Hong Kong, the embedded cultural norms, particularly "face concern," make people with depression

XSL•FO
RenderX

avoid seeking help from mental health professionals to prevent shaming their own families [33]. A study that investigated the experiences of mental health service users comparatively in 2001 and 2017 found no positive attitudinal changes toward people with mental disorders in Hong Kong due to perceived stigmatization [34]. However, reporting of celebrity depression, particularly their stories of overcoming depression, has resulted in positive attitudinal changes in Hong Kong, motivating readers to learn about available treatments for depression [20]. News that expressed more negative sentiment is correlated with increased anxiety, depression, and psychological distress in its readers [35,36]. Given the potential impact of news, it is important to investigate its sentiment to enhance our comprehension of its influence on both populations with depression and the broader audience. As most journalists are unaware that their reporting may have varying impacts [37] and the sentiment of online news was also less studied in the Asian context, analyzing the sentiment could further provide practical suggestions for future news practices in Hong Kong based on solid research findings. Consequently, this study is also interested in exploring how different sentiments were reflected in different latent subthemes, which constitutes our second RQ: What sentiments are reflected in the online news coverage of depression in Hong Kong and are associated with the latent subthemes? (RQ2)

In addition, starting from early research, it is worth noticing that news coverage on mental disorders provided rare coping resources or information for populations with depression [12,15,25,37,38]. However, the provision of coping strategies (ie, information on skills and resources and emotional support) is important to reduce uncertainty about a certain disease and to facilitate the potential of help-seeking behavior [39]. Therefore, news coverage was suggested to leverage its impact and include more information and resources about depression treatments or prevention, as it may help to reduce the individual's uncertainty about depression, decrease public negative attitudes on depression, and increase public awareness and knowledge of this certain mental health disorder [13,18,40,41]. Therefore, this research also seeks to analyze whether online news coverage of depression provides coping strategies for supporting the population with depression in Hong Kong, which constitutes our third RQ: What kind of coping strategies were provided in the online news coverage of depression in Hong Kong? (RQ3)

Methods

Data Selection and Collection

This research aimed to collect depression-related news on Hong Kong mainstream news websites and analyze the collected news using the Natural Language Toolkit (NLTK) embedded in Python. The latent Dirichlet allocation (LDA) topic modeling was first conducted to address RQ1, followed by the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis to address RQ2. RQ3 involved manual coding of coping strategies by content analysis. The detailed methodology will be later elaborated in the Data Analysis subsection.

To conduct a comprehensive search and collection of depression-related news in Hong Kong, the *Hong Kong Free Press* (*HKFP*) and the *South China Morning Post* (*SCMP*) were carefully selected based on 3 criteria. First, news websites should be accessible to readers without membership needs to prevent potential news access privacy risks. Second, the news website could retrieve a certain amount of depression-related news (ie, >20) to guarantee adequate data for subsequent analysis. Third, the news websites are published in English. Given that Hong Kong is a multilingual society with diverse ethnicities, English is one of the most used languages and provides a broader reach among diverse reader groups in Hong Kong. Online news websites that do not meet these 3 criteria were excluded.

After selecting the websites, the news was searched by the keywords, "depression" and "depressed" separately in each news website to ensure all related news was included as per the search strategy used in previous research [21,41,42]. Only news articles published between February 2019 and May 2024 were collected during the search process. In February 2019, the introduction of the Fugitive Offenders and Mutual Legal Assistance in Criminal Matters Legislation (Amendment) Bill sparked protests and marches in Hong Kong due to divisive political viewpoints, leading to a significant rise in depression cases, which worsened in 2020 with the outbreak of COVID-19 [6,43,44]. Therefore, the data were aimed to be collected during and after this period to obtain a more updated picture of the media portrayal of depression in Hong Kong. News outside this time range was excluded, and data collection concluded on the author's final search day.

Eventually, 3648 news stories were collected. Specifically, when searching with the keyword "depression" in the *SCMP*, 2298 (62.99%) news articles were retrieved, and 1040 (28.51%) more news articles were collected by searching with the keyword "depressed." In the *HKFP*, each keyword retrieved 155 news articles, resulting in a total of 310 (8.5%) news articles being collected. Afterward, all collected news was manually exported to a Microsoft Excel file for screening and excluding duplicated news based on the titles. Meanwhile, nonnews articles, such as opinions and commentary, were excluded. Therefore, of 3338 *SCMP* news articles, 209 (6.26%) duplicated news articles and 846 (25.34%) nonnews articles were excluded, leaving 2283 (68.39%) news articles for subsequent analysis. Furthermore, 158 (51%) duplicated news articles were removed from 310 *HKFP* articles, resulting in 152 (49%) news articles. The dataset thus included 2435 news articles for data analysis.

Data Preprocessing

Before applying the natural language processing (NLP) techniques for each RQ, each news article was uploaded to Python and created as a news corpus. The corpus was first preprocessed and tokenized into words to ensure better clarity and results of analysis [45]. Tokenization in NLP means converting words or sentences in a text into understandable tokens that a program can work with. After tokenizing the corpus, lemmatization was performed to convert words to their root meaning, such as "works," "working," and "worked" to "work." Stop words were removed by using the NLTK library. Common stop words, such as, "the," "and," "in," and

XSL•FO
RenderX

punctuations were removed as they do not provide meaningful information for subsequent data analysis. Afterward, we observed the top 200 words from the corpus and created a stop word list to remove other uninformative words, such as "say," "year," and "people."

Data Analysis

To uncover the subthemes underneath the online news for RQ1, LDA was first performed on the combined dataset to investigate the "hidden" thematic topics. The distribution of these topics will be then examined in each dataset separately to allow a more nuanced understanding. Specifically, LDA is one of the unsupervised topic modeling techniques and is becoming increasingly popular in communication research due to the ability to quickly identify the thematic structure of large amounts of text and allow more efficient explorative and descriptive analyses [45]. LDA assumed that each document or news was represented as a mixture of topics, and each topic was characterized by a distribution of words [46]. The algorithms started by assigning topics randomly to the document and then computed the distributions of words within a topic, as well as the distribution of topics within a document, then continued to update and iterate until the most suitable match was found [47]. The *genism* package was used in the NLTK to perform LDA analysis.

Meanwhile, after performing LDA analysis, coherence is an essential metric that decides the best number of topics (k) for the designated corpus. It measures the quality of the topics generated by the model and how semantically interpretable the topics are. A coherence range of 0.3 to 0.8 is acceptable [48-50]. To gain a more comprehensive understanding of the intertopic relationships, multidimensional scaling (MDS) is also used to assess and visualize the topic distance between each other based on their word distributions [51]. Specifically, topics exhibiting similarities will be positioned closer to each other on the MDS map, while topics with fewer similarities will be placed farther apart. MDS helps to facilitate the decisions regarding whether certain topics should be merged or excluded based on their similarities or distances. In contrast, the LDAvis tool (set λ=1) assists with inspecting the top keywords or most associated terms under each topic based on their relevance [52]. Relevance refers to the probability of occurrences of a term under a topic and is ranked from top to bottom to aid topic interpretation.

Moreover, sentiment analysis is another NLP technique that enables the analysis of sentiment being expressed in online news to address RQ2 [53,54]. VADER is a lexicon-based sentiment analysis tool that is included in the NLTK and shows high accuracy in news sentiment analysis [55]. In sentiment analysis, each news article was treated as a document, and the polarities of individual words within the document were calculated and summed to categorize each news into positive, negative, or neutral stances by providing a compound sentiment score [28]. The compound sentiment score ranged from –1 to 1, where –1 indicated the most negative sentiment and +1 indicated the most positive sentiment [56].

For RQ3, 2 authors who are doctoral students in language and communication studies coded each piece of news sentence by sentence to see whether the news provided coping strategies for the population with depression. Although NLP techniques can efficiently process large datasets, they may lack the nuance and contextual understanding that human coders could interpret. Manual coding can offer a deeper investigation into the news and provide more robust research findings [39,57]. The coding scheme was adapted from 2 social media studies investigating depression and coping strategies in health crises [12,39]. Specifically, each news was manually read and labeled by the first and fourth author in 2 rounds of coding.

The first round involved coding for the absence or presence of coping strategies in the news. The second round involved coding in the news that presented coping strategies and further labeled them based on the following categories: (1) information on skills and resources, (2) emotional support, (3) individual adjustment, and (4) others. The first and second categories correspond to the framework designed by Ngai et al [39] to assess if the news articles include information on coping with depression and provide supportive resources as well as emotional support. The third category aligns with the coding done by Pan et al [12], which examines whether the news provides self-adjustment strategies to cope with depression without seeking external help and is found to be the most used coping strategy to mitigate depression on social media. The fourth category refers to news that includes other coping strategies that are not covered by the previous 3 categories.

To ensure consistency in the application of the coding, Cohen κ, a statistical measure, was applied to measure the intercoder reliability on 2435 news articles [57]. The κ value ranges from –1 to 1, suggesting that a value of 0.8 in each round is the minimum acceptable intercoder agreement [58]. The flowchart in Figure 1 outlines the overall process of data collection, preprocessing, and analysis.

XSL•FO
RenderX

Figure 1. Flowchart of data collection, preprocessing, and analysis. LDA: latent Dirichlet allocation; RQ: research question.

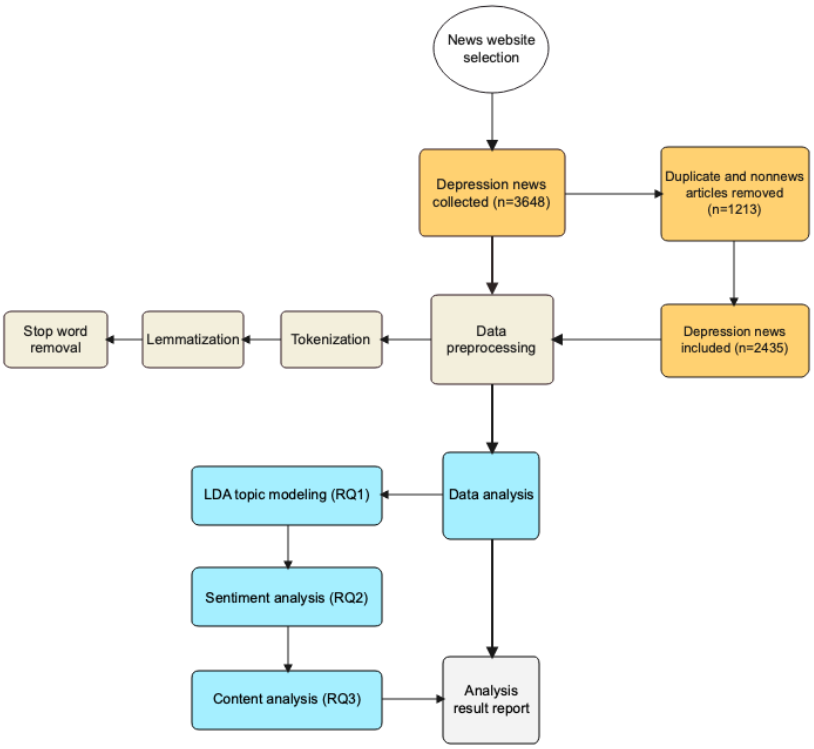

Ethical Considerations

This study was conducted in accordance with ethical standards and guidelines. Institutional ethical approval was not sought as the research did not involve interaction with human subjects [59]. The data used in this study were obtained from publicly available resources and were analyzed in a manner that ensured the privacy and confidentiality of any individuals indirectly involved.

Results

LDA Topic Modeling

After data preprocessing, the LDA topic modeling was first used for RQ1. We examined the coherence scores in different k from 1 to 15 and found the coherence score reached its peak at k=8, where the coherence value equaled 0.5834 (Figure 2). Meanwhile, we visualized the intertopic relationships between the 8 topics by MDS, as illustrated in Figure 3 [51]. The topic keywords under each topic were also presented by LDAvis on the same figure [52]. However, it is worth noticing that LDA-generated topics might not be readily interpretable. Previous research has suggested labeling the topics into meaningful frames to better interpret the LDA-generated topics [60,61]. Having an expert in a related research field to label the frame or review the top documents and top words under each LDA topic are both feasible approaches that have been applied in previous research [45]. Therefore, the third author, who is a public health expert, particularly in the field of social psychology, reviewed each topic, the most relevant news under each topic, and top keywords from LDAvis to label the topic frames, as shown in Table 1.

In addition, drawing from observation, topic 3 and topic 8 were merged into one frame as topic 8 only occupied a small share of the news according to MDS, and its underlying top keywords and news were more aligned with the content of topic 3. Subsequently, we examined the LDA topic labels within the *HKFP* dataset and revealed that a significant proportion of the news articles, specifically 74 (48.7%) of the 152 news articles, were allocated to topic 1. This was followed by 59 (38.8%) of the 152 news articles assigned to topic 2. In addition, 10 (6.6%) news articles were labeled under topic 5, while 9 (5.9%) articles were equally assigned to topics 7, 4, and 3. On the basis of the identified topics, we also examined the reporting time of the news under each topic to visualize the overall temporal trend (Figure 4).

XSL•FO
RenderX

Figure 2. Coherence scores of varied topic numbers.

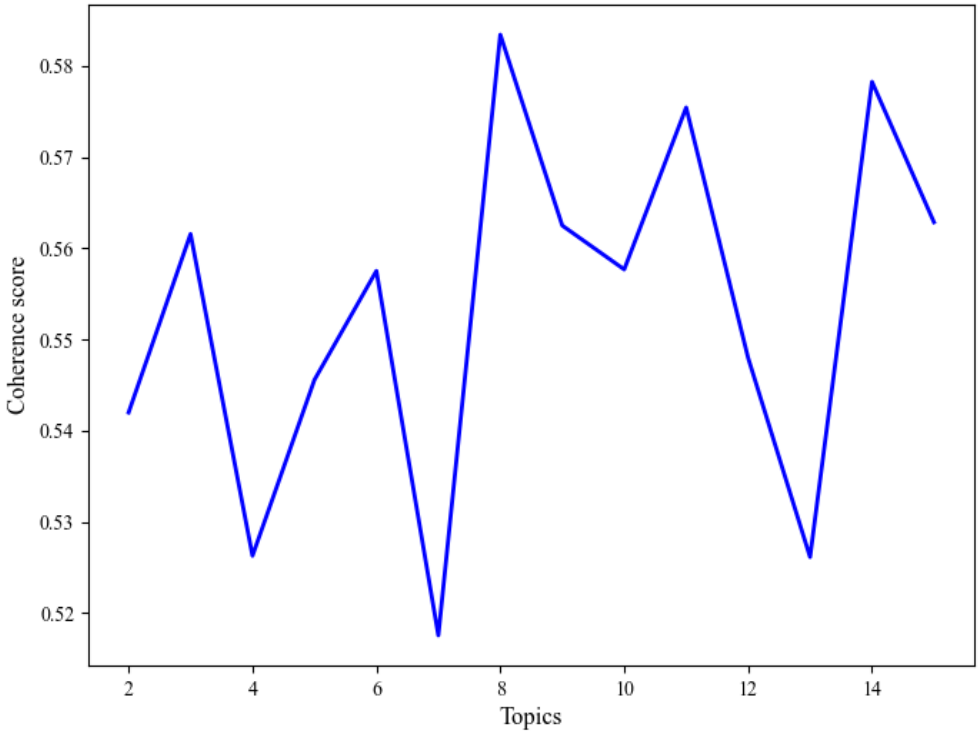

Figure 3. Intertopic distance via multidimensional scaling (MDS) and Top 30 words for topic 1 via LDAvis. PC1: Principal Component Analysis 1; PC2: Principal Component Analysis 2.

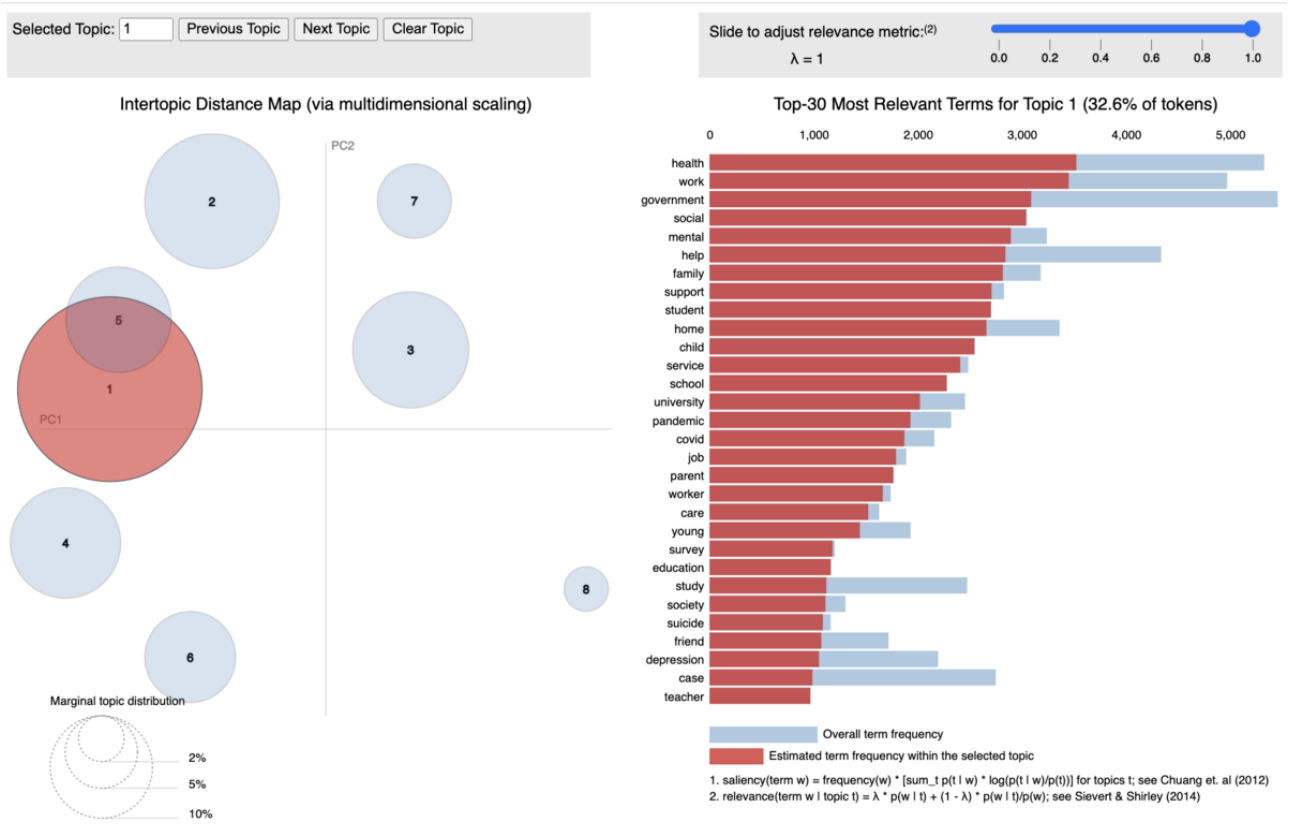

XSL•FO
RenderX

Table 1. Topic frames, top keywords, and illustrative quotes in Hong Kong depression news.

Topic frames or sub-themes	Topics	Top keywords	Illustrative quotes	News portion (N=2435), n (%)
Societal system	Topic 1	health, work, government, social, mental, help, family, support, student, and home	"The number of Hong Kong students with mental health problems shot past 1,400 in the last academic year...with experts and educators blaming the 2019 social unrest and the Covid-19 pandemic...Some feared that the actual number of young people with mental health issues could be much higher...included anxiety, depression, obsessive-compulsive disorder...and eating problems."	848 (34.83)
Law enforcement	Topic 2	police, law, protest, case, government, officer, china, protester, force, and security	"A real estate agent accused of slashing a man with a knife and assaulting another during an antigovernment protest on Friday has been remanded to a psychiatric facility, after a court heard he had been battling depression for a year."	458 (18.81)
Global recession	Topic 3	china, market, chinese, unitedstates, company, trade, business, Beijing, government, and bank	"After three years of the pandemic, the market should cope better than before, especially when valuation is so depressed,' he said. 'Nevertheless, investors should be prepared for market volatility ahead.'"	412 (16.92)
Global recession	Topic 8	economic, economic, growth, economist, lockdown, virus, recovery, spending, tax, and inflation	"New Zealand suffered its worst economic slump since the Great Depression in the second quarter as a strict nationwide lockdown to combat the coronavirus brought the country to a standstill."	3 (0.12)
Lifestyle	Topic 4	sleep, help, drink, eat, bed, body, work, stress, gas, and mind	"The genetic study...by the University of Colorado Boulder, the Broad Institute at MIT and the University of Harvard...found that waking just one hour earlier—but getting the same amount of sleep—reduces your risk of major depression by 23 percent."	196 (8.05)
Leisure	Topic 5	film, work, story, game, sport, woman, video, music, event, and book	"Osaka withdrew from last year's French Open after being fined and threatened with expulsion for refusing to do media duties, which she said had contributed to the depression she had struggled with for years."	273 (11.21)
Health issues	Topic 6	patient, health, treatment, study, symptom, doctor, diseases, cancer, drug, and hospital	"This discovery is expected to provide rapid and robust therapies, and most importantly, if the transcorneal electrical stimulation is effective in reducing the symptoms of depression and dementia, then this could be a major breakthrough for patients, families and caregivers."	169 (6.94)
US politics	Topic 7	election, trump, party, unitestates, candidate, democracy, president, political, vote, and gay	"Fetterman's dress style, or perhaps lack of style, became his signature on the campaign trail before entering the Senate this year. He also gained sympathy from many after he had to undergo treatment for clinical depression soon after taking office."	76 (3.12)

XSL•FO
RenderX

Figure 4. Trend of news articles by topic.

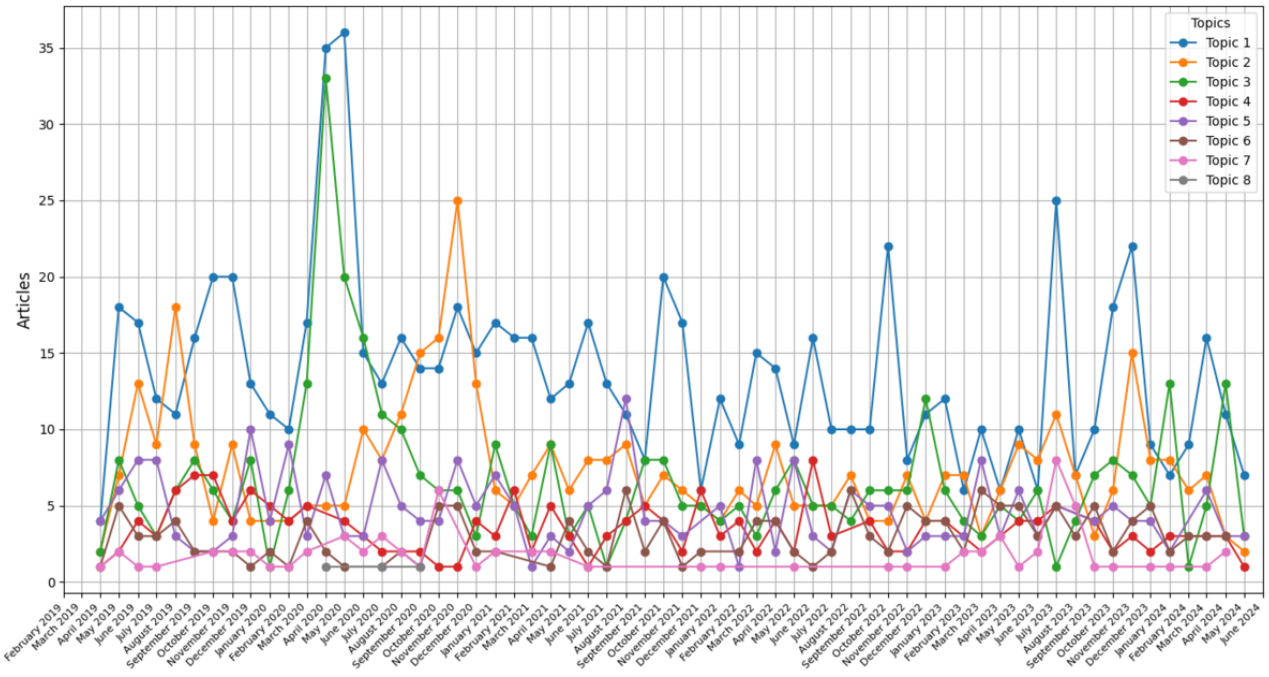

Sentiment Analysis

After uncovering the subthemes in the online news, our next objective was to ascertain the prevailing sentiment conveyed in the news. Sentiment scores were computed for each news using VADER. The findings indicated that the collective sentiment expressed in the news articles leaned toward the positive spectrum. Specifically, 50.51% (1230/2435) of the news articles reflected a positive stance, whereas 49.08% (1195/2435) of the news articles reflected a negative sentiment. Moreover, 10 (0.41%) pieces of news maintained reflected a neutral stance.

Consequently, we correlated the sentiment scores in different LDA topics to observe the sentiment variation in response to RQ2 (Figure 5). Notably, topics 1, 2, 6, 7, and 8 are more negatively inclined, while topics 3, 4, and 5 are positively oriented. Therefore, subthemes of the societal system, law enforcement, health issues, and US politics exhibited a predominantly negative sentiment. In contrast, subthemes of the global recession in a combination of topic 3 and topic 8 sentiment results as well as lifestyle and leisure exhibited a more positive sentiment.

Figure 5. Sentiment distribution by topic.

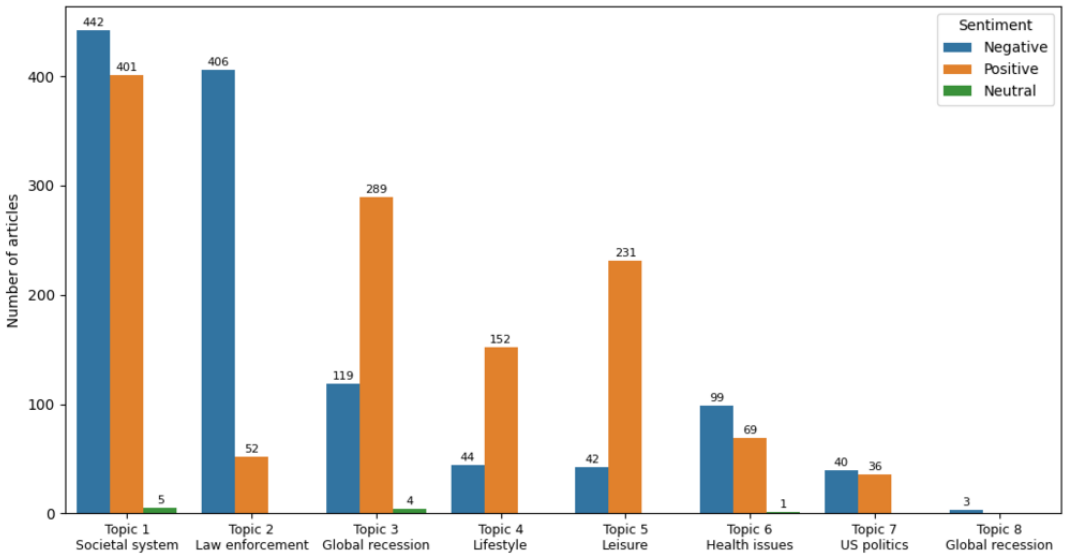

XSL•FO
RenderX

Coping Strategy Analysis

RQ3 focused on examining whether online news has provided depression-related coping strategies. It was found that >76% (1861/2435) of the news did not present coping strategies. After the two coders labeled all the news respectively, the interrater reliability in identifying and categorizing the coping strategies across 2 rounds was assessed using the Cohen κ statistic [58]. All disagreements between the coders were discussed. We used the SPSS (IBM Corp) crosstab function to calculate Cohen κ in both rounds of labeling. In the first round, the Cohen κ value obtained was 0.882, indicating a substantial level of agreement regarding the presence or absence of coping strategies in 2435 news articles. For the second round, we focused only on the news initially labeled as "2" in the first round, in which 501 (20.57%) news articles were consistently annotated by both coders, emphasizing particular coping strategies were presented. The κ value obtained for the second labeling round yielded a value of 0.849, suggesting a firm agreement between coders regarding the specific categories of coping strategies.

Among the news that presented coping strategies, most of the news articles (334/501, 66.7%) provided information on skills and resources to cope with depression. Under this category, local mental health organizations or charities launched diverse mental health programs and provided support hotlines to facilitate quick resource access. This was followed by individual adjustments, including making personal changes, such as adopting a healthier lifestyle through exercise, sleep, and diet. Emotional support was the third preferred coping strategy, including support from friends, family, partners, and pets to help individuals mitigate depression. The least mentioned category includes mental health advice suggested by experts, particularly focusing on parental guidance for the well-being of their children and coverage of past mental health initiatives.

Discussion

Principal Findings

This study investigated online news coverage of depression in Hong Kong, discovering prevalent subthemes (RQ1), overall sentiment (RQ2), and coping strategies (RQ3) in the news. In response to RQ1, the topic modeling results identified 8 LDA topics based on the coherence score, and 7 latent subthemes were eventually identified by using MDS, LDAvis, and expert insights. The primary subtheme is "societal system," accounting for a representation of 34.83% (848/2435) of overall news, followed by "law enforcement" (458/2435, 18.81%), "global recession" (415/2435, 17.04%), "lifestyle" (196/2435, 8.05%), "leisure" (273/2435, 11.21%), "health issues" (169/2435, 6.94%), and "US politics" (76/2435, 3.12%). The overall sentiment of the news was slightly more positive, which aligns with news reported in Canada and Chile [29,30,62]. Among the subthemes, "societal system," "law enforcement," "health issues," and "US politics" exhibited negative tendencies, while others tended to be more positive. The coping strategies were substantially inadequate, in line with previous research results [25,38,40]. However, this study has enhanced comprehension regarding preferred types of coping strategies provided in depression-related news, predominately featuring information on skills and resources and individual adjustment to cope with depression.

On the basis of the sentiment analysis, it becomes apparent that the primary subtheme, focusing on the societal system, exhibited a more negatively inclined sentiment. The subtheme was related to diverse systems in Hong Kong's society, including the health and social systems, such as family, education, and work. The amount of news by reporting time indicated a peak in reported news, notably during May 2020, April 2020, and July 2023. Therefore, we first examined the news content between April and May 2020 and found that COVID-19 lockdown measures greatly affected Hong Kong, contributing to a more depressed mood [6]. Specifically, the period signifies the second wave of the COVID-19 outbreak [63]. While the first wave in Hong Kong reported relatively few cases, the second wave saw a sharp increase in cases, leading to stricter measures to contain the virus, including school closures, work-from-home policies, and a ban on dine-in services [43]. School closures, as noted in the news, resulted in an increased rate of depression among children and heightened anxiety for parents who had to take care of both their children and the older adults at home. Work-family conflicts also elevated due to the disruption to normal work routine (ie, no office space and blurred working time), which was found to be depressing [8,64]. In July 2023, we continued to examine the news content and found that reported suicide cases surged. Notably, an influential star, Coco Lee, who experienced depression, died by suicide at home, raising more awareness for expanded mental health support services and enhanced accessibility, with support hotlines being frequently highlighted. This phenomenon is supported by previous research, which suggests that news of celebrity suicides is more likely to provide coping strategies, as readers, particularly fans, may become more proactive in seeking treatment for depression and increase their understanding of the condition [18,20].

The second leading subtheme focuses on law enforcement, mainly negative sentiment, and coincides with the subtheme in other news studies [23,42,65]. The amount of reported news peaked during November 2020 and August 2019. In the former month, individual crimes and court cases were reported, particularly centering on an event of a Hong Kong professor murdering his wife, which may have brought great social shock and attention. Notably, depression was often cited as an explanation for the perpetrator's illegal actions in the news, which may reinforce stigmatizing attitudes and portray people with depression as violent, dangerous, and more likely to commit crimes [13,16,23,31]. The stigmas associated with depression could have been strengthened by the news report, making the public more unlikely to associate with, employ, and work with people affected by depression [25]. In Hong Kong, a study on media reporting of homicides committed by patients with schizophrenia found an increase in negative attitudes and the belief that people with psychosis are dangerous [66]. In the latter month, the online news focused on antigovernment demonstrations regarding the extradition bill. A sequence of protests and marches transpired in opposition to the Hong Kong government's ratification of the bill, which would allow the transfer of criminal suspects to China, China Taiwan, and China Macau, raising local people's distress about the potential erosion

XSL•FO
RenderX

of autonomy [67]. This event marked the largest protest activities in Hong Kong's history, as evidenced by a survey conducted from June to August 2019, where >90% of protesters acknowledged that if the government persisted in advancing the bill, protest activities would likely escalate [5,67]. Meanwhile, the rate of depression among locals continued to rise, along with suicidal ideations, surpassing the levels recorded during the larger-scale protests of 2014 [68].

The third leading subtheme centers on the global recession resulting from the economic hardship induced by widespread COVID-19 across industries but with more contradicting positive sentiment. Distinctly, the news story's structure would initially depict economic challenges and then emphasize the initiative underway to facilitate recovery in response to the crisis. A significant change in overall sentiment occurred, where up to January 2021, the continued decrease in the quantity of positive news ceased. Subsequently, the aggregate number of positive news exceeded negative news almost every month, signalizing an overall shift from negative to positive sentiment surrounding the discussed economic circumstances. This may be attributed to the fact that people were forced to adapt to a "new normal" and start to restore health, economy, and societies after the pandemic had emerged for one year [69]. Strict social distancing measures were gradually relaxed worldwide, such as school reopening, shortened quarantine days, and global trade resumes [70]. Meanwhile, the Hong Kong government has also launched various initiatives and campaigns to stimulate the economic market through issuing digital consumption vouchers to eligible residents; supporting enterprises and the general public with tax cuts; and subsidizing older adults, working families, and child-raising families with more allowances [71].

The fourth and sixth subthemes share greater similarities, as observed on the MDS graph. The fourth subtheme, lifestyle, presented a more positive sentiment by offering different healthy lifestyle suggestions and ways to help its readers improve their well-being. This can be credited to the insights gained from the analysis of coping strategies. Individual adjustment emerged as the second most prominent coping strategy category, emphasizing ways to mitigate depression by considering altering their lifestyle to promote better health through adjustments, such as maintaining adequate sleep, doing regular exercise, and having a nutritious diet while quitting unhealthy behaviors, such as drinking, smoking, and staying up all night. In recent years, lifestyle-based mental health interventions were suggested to be implemented in clinical practices for adults with depression, with growing recommendations on practicing physical activity and exercise, relaxation techniques, work-directed interventions, sleep, and mindfulness-based therapies [72]. Likewise, the sixth subtheme, health issues, was more inclined toward negative sentiment and shared information about the symptoms and negative effects of mental illness as well as available treatments or therapies that are effective in tackling mental health conditions.

The final 2 subthemes were inclined toward positive and negative sentiment, respectively, focusing on leisure and US politics. The leisure subtheme shared depression-related dramas or narratives about celebrities who were and are experiencing depression. However, it is worth noticing that while reading the news, successful stories of how individuals combat depression are rare. Earlier research found that only 7% of depression-related news covered successful treatment or stories of individuals recovering from mental illness, and only 18% emphasized recovery and rehabilitation of those affected as part of their content [16,29]. The concluding subtheme centers on US politics, specifically emphasizing the rivalry and electoral process for the presidency as well as politicians from different parties who experienced depression.

Finally, based on the coping strategy analysis, we found that information on skills and resources was the leading coping strategy category to offer mental health support for populations with depression. Under this category, the suicide intention and prevention helplines were the most frequently cited information, as depression is a strong indicator of suicide [73]. In contrast, various mental health programs were introduced by local mental health organizations such as Mind HK, Kely Support Group, and the Mental Health Association of Hong Kong to offer counseling services and therapies for people with depression. Under the individual adjustment category, exercise is the most cited one, with leading recommendations on different types of running, yoga, and mindfulness activities to enhance self-efficacy in dealing with depression. Regular exercise can bring positive mood change and serve as an antidepressant for individuals with depression [74,75]. Regarding the emotional support category, friends, family, and loved ones are protective factors [76]. However, having a bad relationship may further exacerbate depressive symptoms [77]. In addition, companion pets played a critical role in reducing feelings of depression, particularly during the COVID-19 pandemic period with poor social contact [78]. Although there exists a variety of coping strategies, the overall quantity of strategies offered in the news is deemed rare. In addition, we have observed the inadequacy of policy interventions or support as well as financial assistance for individuals with depression, which implies the deficiency in the societal level of coping strategies in the news coverage. The news media should raise awareness by incorporating more varieties of coping strategies in depression-related news to leverage its positive impact, which might potentially foster positive mental outcomes among individuals with depression and facilitate early mental intervention [79,80].

Implications for Future Practices

The media portrayal of depression has a dual effect on its readers. In a positive way, the news can help to reduce the uncertainty experienced by an individual with depression, spread information and knowledge about depression, and increase the recognition and understanding of the public [39,81]. In contrast, news can also exacerbate the social stigma and self-stigma of individuals with depression, resulting in decreasing support resources or information from the public, which may deteriorate the well-being of individuals with depression, and may lead to less intentions of help-seeking [22,32,35]. The subthemes highlighted in the news were predominantly negative, suggesting that Hong Kong news agencies and journalists should expand the range of topics they cover when portraying depression. Preferably, incorporating successful treatment of depression and achievements of individuals with depression as potential subthemes can help to reduce the stigmatization of people with

XSL•FO
RenderX

depression and foster a better understanding, positive attitude, and accurate beliefs toward populations with depression among the public [14,24,30]. Meanwhile, the overall sentiment of the news was slightly more positive, with 50.51% (1230/2435) positive versus 49.07% (1195/2435) negative news, suggesting that more efforts are needed to report positive content, which could help reduce the stigmatization toward depression [29]. Containing more coping strategies in the news is also suggestive. Specifically, enhancing the inclusion of individual adjustment category in depression-related news was relatively more achievable and preferable as studies found that improving self-efficiency, a key component of the coping appraisal in the protection motivation theory, is more effective in forecasting health-related intention and behavior than eliciting fear or threat [12,18].

Implications for Future Research

On the basis of the methods and findings of this research, we would like to suggest several points for interested scholars to expand on the current research. First, we would highlight a potential future research direction focusing on coping strategies. In this study, we have analyzed the coping strategies in the online depression news; however, we do not know whether the public or individuals with depression will literally increase their knowledge and awareness of depression and facilitate health-seeking behaviors among populations with depression. Future research is recommended to explore the social media platforms of these news websites and examine the interactions between the news posts and users' engagement (ie, comments and likes) to see whether their interactions can promote attitudinal changes and potential links with behavioral change.

Second, from a methodological standpoint, sentiment analysis is a powerful tool, notably through techniques such as EmoBank, which analyzes emotions from both the reader's and writer's perspective and examines emotion dimensions, such as valence, arousal, and dominance [82,83]. This technique is constructive to explore the topic proposed earlier. Other more advanced techniques such as the multidimensional relation model (examining the relationships between valence, arousal, and dominance) and tree-based regional convolutional neural network–long short term memory (examining text sequential tendency, text regional information, and even nontext elements such as images in the news) are also encouraged for more in-depth research analysis [84,85]. In contrast, for acknowledgment, different score ranges of sentiment analysis are also available beyond –1 to 1. For example, the Semantic Orientation Calculator and AFINN lexicon-based resources provided a score range from –5 (most negative) to +5 (most positive) [86,87]. The Extended Affective Norms for English Words and Chinese EmoBank range from 1 to 9 (negative-neutral-positive), which may also provide a delicate interpretation and quantification of sentiment analysis results [83,88]. Meanwhile, it is also suggested to perform error analysis in sentiment analysis, as demonstrated in these 2 case studies, whether through different computational support or human inspection [89,90].

Third, we would also encourage the replication of this study design and its application to different mental health problems or language cases. However, it is important to note that the resources required for performing LDA and sentiment analysis may differ. For example, except for alphabetical languages such as English, nonalphabetic languages such as Chinese require other packages and resources such as Jieba and Chinese EmoBank due to differences in the language system [61,83].

Limitations

The study has several limitations that should be acknowledged and recommended for future research to address and advance. Due to the international nature of Hong Kong society, we primarily selected English-reported news as our focus. English news tends to reach more diverse audiences, including disadvantaged groups such as ethnic minorities and foreign domestic helpers who may not be proficient in Chinese, face more stigmas, and are more susceptible to mental health issues [91-93]. However, for future research, this study has presented the steps for investigating the news reports of a mental health issue. Future research can also consider incorporating Chinese as another mainstream language to provide a more comprehensive understanding of portraying depression in Hong Kong, particularly focusing on the local population.

Second, after news coverage collection, we recognized the practical flaw of the imbalance of the number of news items being collected from *SCMP* and *HKFP*. This imbalance may be primarily due to *SCMP* standing out as the most widely read paid newspaper in Hong Kong in terms of news and current affairs, according to the Hong Kong Media Survey report [94]. In contrast, the *SCMP* website places special emphasis on mental health and has dedicated a section called the "Wellbeing" page to report on mental health issues, including depression, which may contribute to the imbalance in coverage between these two news websites.

Third, while screening the news coverage, a further step can be implemented in the screening process, which is to exclude news stories that mention depression as only an incidental topic rather than as the main topic of the news coverage [42]. Future research could explore the possibility of excluding news articles that mention "depressed" or "depression" only once, whether through NLP techniques or human inspection, to ensure a more focused analysis targeting mental health–centered discussion.

Conclusions

In conclusion, this study is the first to explore latent subthemes, sentiment, and coping strategies in online depression-related news in Hong Kong based on available research. The findings offer a deeper understanding of the underlying subthemes, revealing a predominance of slightly more positive sentiments, inadequate coping strategies, and associated categories that warrant increased attention. For future news reporting, it is recommended that news organizations and journalists take into account the recommendations discussed in the Implications for Future Practices section. Future research is also recommended to explore the topics and methodologies discussed in the Implications for Future Research section and tackle the acknowledged limitations. We encourage the application of NLP techniques to other prevalent mental health disorders such as anxiety, schizophrenia, and bipolar disorder and in different

XSL•FO
RenderX

language contexts. In general, this study enhances our comprehension of how depression is portrayed in online news coverage in Hong Kong, particularly during and after periods of social unrest and the COVID-19 pandemic, by providing the latest findings and perspectives.

Acknowledgments

This research was funded by the Hong Kong Polytechnic University (P0046380).

Data Availability

The datasets generated and analyzed during this study are available from the first author on reasonable request.

Authors' Contributions

SC conceptualized the study and curated, collected, and analyzed the data. SC wrote the original draft of the manuscript and reviewed and edited it. CSBN supervised the quality of the manuscript, reviewed and edited it, and acquired funding. CC investigated the data and provided comments. YH analyzed the data.

Conflicts of Interest

None declared.

References

1. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-05-16]
2. Mental health review report. Food and Health Bureau. 2017 Mar. URL: https://www.healthbureau.gov.hk/download/press_and_publications/otherinfo/180500_mhr/e_mhr_full_report.pdf [accessed 2023-05-17]
3. Choi EP, Hui BP, Wan EY. Depression and anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740 [FREE Full text] [doi: 10.3390/ijerph17103740] [Medline: 32466251]
4. Lam SC, Arora T, Grey I, Suen LK, Huang EY, Li D, et al. Perceived risk and protection from infection and depressive symptoms among healthcare workers in Mainland China and Hong Kong during COVID-19. Front Psychiatry 2020 Jul 15;11:686 [FREE Full text] [doi: 10.3389/fpsyt.2020.00686] [Medline: 32765321]
5. Ni MY, Yao XI, Leung KS, Yau C, Leung CM, Lun P, et al. Depression and post-traumatic stress during major social unrest in Hong Kong: a 10-year prospective cohort study. Lancet 2020 Jan 25;395(10220):273-284. [doi: 10.1016/S0140-6736(19)33160-5] [Medline: 31928765]
6. Hou WK, Lee TM, Liang L, Li TW, Liu H, Ettman CK, et al. Civil unrest, COVID-19 stressors, anxiety, and depression in the acute phase of the pandemic: a population-based study in Hong Kong. Soc Psychiatry Psychiatr Epidemiol 2021 Aug;56(8):1499-1508 [FREE Full text] [doi: 10.1007/s00127-021-02037-5] [Medline: 33590313]
7. Tso IF, Park S. Alarming levels of psychiatric symptoms and the role of loneliness during the COVID-19 epidemic: a case study of Hong Kong. Psychiatry Res 2020 Nov;293:113423 [FREE Full text] [doi: 10.1016/j.psychres.2020.113423] [Medline: 32871487]
8. Wong FH, Liu T, Leung DK, Zhang AY, Au WS, Kwok WW, et al. Consuming information related to COVID-19 on social media among older adults and its association with anxiety, social trust in information, and COVID-safe behaviors: cross-sectional telephone survey. J Med Internet Res 2021 Feb 11;23(2):e26570 [FREE Full text] [doi: 10.2196/26570] [Medline: 33523825]
9. Yang C, Lai DW, Sun Y, Ma CY, Chau AK. Mobile application use and loneliness among older adults in the digital age: insights from a survey in Hong Kong during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Jun 23;19(13):7656 [FREE Full text] [doi: 10.3390/ijerph19137656] [Medline: 35805316]
10. Yu L, Du M. Social networking use, mental healtAh, and quality of life of Hong Kong adolescents during the COVID-19 pandemic. Front Public Health 2022 Oct 31;10:1040169 [FREE Full text] [doi: 10.3389/fpubh.2022.1040169] [Medline: 36388293]
11. Cheng C, Chen S, Chen S. Impact of national lockdown measures on the association between social media use and sleep disturbance during COVID-19: a meta-analysis of 21 nations. Cyberpsychol Behav Soc Netw 2024 Aug 27;27(8):527-538. [doi: 10.1089/cyber.2023.0571] [Medline: 38916117]
12. Pan J, Liu B, Kreps GL. A content analysis of depression-related discourses on Sina Weibo: attribution, efficacy, and information sources. BMC Public Health 2018 Jun 20;18(1):772 [FREE Full text] [doi: 10.1186/s12889-018-5701-5] [Medline: 29925363]
13. Wang W. Stigma and counter-stigma frames, cues, and exemplification: comparing news coverage of depression in the English- and Spanish-language media in the U.S. Health Commun 2019 Feb;34(2):172-179. [doi: 10.1080/10410236.2017.1399505] [Medline: 29131692]

XSL•FO
RenderX

14. Coleman CL, Corbitt JA. The social construction of depression in news stories: an exploratory study. Ecquid Novi Afr Journal Stud 2010 Dec 01;24(1):99-114. [doi: 10.1080/02560054.2003.9653263]
15. Francis C, Pirkis J, Blood RW, Dunt D, Burgess P, Morley B, et al. Portrayal of depression and other mental illnesses in Australian nonfiction media. J Community Psychol 2005 Mar 24;33(3):283-297. [doi: 10.1002/jcop.20050]
16. McGinty EE, Kennedy-Hendricks A, Choksy S, Barry CL. Trends in news media coverage of mental illness in the United States: 1995-2014. Health Aff (Millwood) 2016 Jun 01;35(6):1121-1129 [FREE Full text] [doi: 10.1377/hlthaff.2016.0011] [Medline: 27269031]
17. Penetration rate of digital media in Hong Kong as of 3rd quarter 2023, by type. Statista. 2024 Feb 23. URL: https://www.statista.com/statistics/1061608/hong-kong-digital-media-penetration-rate-by-type/ [accessed 2024-10-16]
18. Lee SY. Media coverage of celebrity suicide caused by depression and increase in the number of people who seek depression treatment. Psychiatry Res 2019 Jan;271:598-603. [doi: 10.1016/j.psychres.2018.12.055] [Medline: 30554108]
19. Milne S, Sheeran P, Orbell S. Prediction and intervention in health‐related behavior: a meta‐analytic review of protection motivation theory. J Appl Soc Psychol 2006 Jul 31;30(1):106-143 [FREE Full text] [doi: 10.1111/j.1559-1816.2000.tb02308.x]
20. Leung VS. The disclosure of celebrity major depressive disorder diagnoses in Hong Kong: its effects on public awareness and understanding toward the illness. Community Ment Health J 2019 Jan;55(1):120-128. [doi: 10.1007/s10597-018-0274-4] [Medline: 29679192]
21. Zhang Y, Jin Y. Who's responsible for depression? J Int Commun 2015 Jul 31;21(2):204-225. [doi: 10.1080/13216597.2015.1052532]
22. Goepfert NC, Conrad von Heydendorff S, Dreßing H, Bailer J. Effects of stigmatizing media coverage on stigma measures, self-esteem, and affectivity in persons with depression - an experimental controlled trial. BMC Psychiatry 2019 May 07;19(1):138 [FREE Full text] [doi: 10.1186/s12888-019-2123-6] [Medline: 31064374]
23. Myrick JG, Major LH, Jankowski SM. The sources and frames used to tell stories about depression and anxiety: a content analysis of 18 years of national television news coverage. Electron News 2014 Mar 03;8(1):49-63. [doi: 10.1177/1931243114523962]
24. Ng ZZ, Li G, Flynn S, Yow WQ. How COVID-19 news affect older adults' mental health-evidence of a positivity bias. Int J Environ Res Public Health 2023 Feb 23;20(5):3950 [FREE Full text] [doi: 10.3390/ijerph20053950] [Medline: 36900959]
25. Cummings LA, Konkle AT. The representation of depression in Canadian print news. Can J Community Ment Health 2016 Dec 01;35(3):97-112. [doi: 10.7870/cjcmh-2016-043]
26. Villasanta AP, Acosta AC, Tabo-Corpuz CE, Manaois JO. Exposure to COVID-19 news and its relation to stress, depression, and anxiety in the context of difficulty in accessing social support. J Ment Health 2023 Dec;32(6):1105-1110. [doi: 10.1080/09638237.2022.2069706] [Medline: 35583225]
27. Lam LC, Wong CS, Wang MJ, Chan WC, Chen EY, Ng RM, et al. Prevalence, psychosocial correlates and service utilization of depressive and anxiety disorders in Hong Kong: the Hong Kong Mental Morbidity Survey (HKMMS). Soc Psychiatry Psychiatr Epidemiol 2015 Sep;50(9):1379-1388. [doi: 10.1007/s00127-015-1014-5] [Medline: 25660760]
28. Samuels A, Mcgonical J. News sentiment analysis. arXiv Preprint posted online on June 5, 2020 [FREE Full text] [doi: 10.48550/arXiv.2007.02238]
29. Grandón P, Fernández Vega D, Sánchez Oñate AA, Vielma Aguilera AV, Villagrán Valenzuela L, Vidal Gutiérrez D, et al. Mental disorders in the media: a retrospective study of newspaper coverage in the Chilean Press. Int J Soc Psychiatry 2022 Nov 13;68(7):1351-1362. [doi: 10.1177/00207640211023086] [Medline: 34254543]
30. Whitley R, Wang J. Good news? A longitudinal analysis of newspaper portrayals of mental illness in Canada 2005 to 2015. Can J Psychiatry 2017 Apr;62(4):278-285 [FREE Full text] [doi: 10.1177/0706743716675856] [Medline: 27777273]
31. Chen M, Lawrie S. Newspaper depictions of mental and physical health. BJPsych Bull 2017 Dec 02;41(6):308-313 [FREE Full text] [doi: 10.1192/pb.bp.116.054775] [Medline: 29234506]
32. Scholz B, Crabb S, Wittert GA. "We've got to break down the shame": portrayals of men's depression. Qual Health Res 2014 Dec;24(12):1648-1657. [doi: 10.1177/1049732314549020] [Medline: 25212853]
33. Chen SX, Mak WW, Lam BC. Is it cultural context or cultural value? Unpackaging cultural influences on stigma toward mental illness and barrier to help-seeking. Soc Psychol Pers Sci 2020 Mar 12;11(7):1022-1031. [doi: 10.1177/1948550619897482]
34. Chung KF, Tse S, Lee CT, Wong MM, Chan WM. Experience of stigma among mental health service users in Hong Kong: are there changes between 2001 and 2017? Int J Soc Psychiatry 2019 Feb;65(1):64-72. [doi: 10.1177/0020764018815926] [Medline: 30488753]
35. Stainback K, Hearne BN, Trieu MM. COVID-19 and the 24/7 news cycle: does COVID-19 news exposure affect mental health? Socius 2020 Nov 02;6:1-15. [doi: 10.1177/2378023120969339]
36. Jones R, Mougouei D, Evans SL. Understanding the emotional response to COVID-19 information in news and social media: a mental health perspective. Hum Behav Emerg Technol 2021 Dec;3(5):832-842 [FREE Full text] [doi: 10.1002/hbe2.304] [Medline: 34901769]
37. Jamieson P, Jamieson KH, Romer D. The responsible reporting of suicide in print journalism. Am Behav Sci 2003 Aug;46(12):1643-1660. [doi: 10.1177/0002764203254620]

XSL•FO
RenderX

38. Wilkinson TJ, Lightfoot CJ, Palmer J, Smith AC. Navigating the COVID-19 infodemic in those living with kidney disease: access and trust in health information sources and the association with anxiety and depression. Curr Med Res Opin 2022 Jan;38(1):35-42 [FREE Full text] [doi: 10.1080/03007995.2021.1984221] [Medline: 34551667]
39. Ngai CS, Singh RG, Lu W, Yao L, Koon AC. Exploring the relationship between trust-building strategies and public engagement on social media during the COVID-19 outbreak. Health Commun 2023 Oct;38(10):2141-2157. [doi: 10.1080/10410236.2022.2055261] [Medline: 35473490]
40. Lawal O. Attitudes towards depression and suicide: an analysis of news coverage and online comments on Nigerian news sites. Georgia State University. 2018 Jul 08. URL: https://scholarworks.gsu.edu/communication_theses/120/ [accessed 2024-07-16]
41. Tobin G, Lyddy F. Media representation of depression in young people: a corpus-based analysis of Irish newspaper coverage. Ir J Psychol Med 2014 Mar;31(1):21-30. [doi: 10.1017/ipm.2013.64] [Medline: 30189473]
42. Wang W. Exemplification and stigmatization of the depressed: depression as the main topic versus an incidental topic in national US news coverage. Health Commun 2020 Jul;35(8):1033-1041. [doi: 10.1080/10410236.2019.1606874] [Medline: 31014117]
43. Kwok KO, Wei WI, Huang Y, Kam KM, Chan EY, Riley S, et al. Evolving epidemiological characteristics of COVID-19 in Hong Kong from January to August 2020: retrospective study. J Med Internet Res 2021 Apr 16;23(4):e26645 [FREE Full text] [doi: 10.2196/26645] [Medline: 33750740]
44. Shek DT. Protests in Hong Kong (2019-2020): a perspective based on quality of life and well-being. Appl Res Qual Life 2020;15(3):619-635 [FREE Full text] [doi: 10.1007/s11482-020-09825-2] [Medline: 32206136]
45. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, et al. Applying LDA topic modeling in communication research: toward a valid and reliable methodology. Commun Methods Meas 2018 Feb 16;12(2-3):93-118. [doi: 10.1080/19312458.2018.1430754]
46. Chen W, Rabhi FA, Liao W, Al-Qudah I. Leveraging state-of-the-art topic modeling for news impact analysis on financial markets: a comparative study. Electronics 2023 Jun 09;12:2605. [doi: 10.3390/electronics12122605]
47. Leung YT, Khalvati F. Exploring COVID-19-related stressors: topic modeling study. J Med Internet Res 2022 Jul 13;24(7):e37142 [FREE Full text] [doi: 10.2196/37142] [Medline: 35731966]
48. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text] [doi: 10.7551/mitpress/1120.003.0082]
49. Islam T. Yoga-veganism: correlation mining of Twitter health data. arXiv Preprint posted online on June 15, 2019 [FREE Full text] [doi: 10.48550/arXiv.1906.07668]
50. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Presented at: WSDM '15; February 2-6, 2015; Shanghai, China. [doi: 10.1145/2684822.2685324]
51. Parker MA, Valdez D, Rao VK, Eddens KS, Agley J. Results and methodological implications of the digital epidemiology of prescription drug references among Twitter users: latent Dirichlet allocation (LDA) analyses. J Med Internet Res 2023 Jul 28;25:e48405 [FREE Full text] [doi: 10.2196/48405] [Medline: 37505795]
52. Sievert C, Shirley K. LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. 2014 Presented at: ACL 2014; June 27, 2014; Baltimore, MD. [doi: 10.3115/v1/w14-3110]
53. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conf Web Soc Media 2014 May 16;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]
54. Yang M, Jiang B, Wang Y, Hao T, Liu Y. News text mining-based business sentiment analysis and its significance in economy. Front Psychol 2022 Jul 14;13:918447 [FREE Full text] [doi: 10.3389/fpsyg.2022.918447] [Medline: 35910983]
55. Nkongolo Wa Nkongolo M. News classification and categorization with smart function sentiment analysis. Int J Intell Syst 2023 Nov 13;2023(1):1-24. [doi: 10.1155/2023/1784394]
56. Bonta V, Kumaresh N, Janardhan N. A comprehensive study on lexicon based approaches for sentiment analysis. Asian J Comput Sci Technol 2019 Jan 22;8(S2):1-6. [doi: 10.51983/ajcst-2019.8.s2.2037]
57. Ngai CS, Singh RG, Yao L. Impact of COVID-19 vaccine misinformation on social media virality: content analysis of message themes and writing strategies. J Med Internet Res 2022 Jul 06;24(7):e37806 [FREE Full text] [doi: 10.2196/37806] [Medline: 35731969]
58. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
59. Institutional review board. The Hong Kong Polytechnic University. URL: https://www.polyu.edu.hk/irb/?sc_lang=en [accessed 2025-01-23]
60. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]

XSL•FO
RenderX

61. Kong D, Chen A, Zhang J, Xiang X, Lou WQ, Kwok T, et al. Public discourse and sentiment toward dementia on Chinese social media: machine learning analysis of Weibo posts. J Med Internet Res 2022 Sep 02;24(9):e39805 [FREE Full text] [doi: 10.2196/39805] [Medline: 36053565]
62. Goulden R, Corker E, Evans-Lacko S, Rose D, Thornicroft G, Henderson C. Newspaper coverage of mental illness in the UK, 1992-2008. BMC Public Health 2011 Oct 12;11:796 [FREE Full text] [doi: 10.1186/1471-2458-11-796] [Medline: 21992410]
63. Chan HY, Chen A, Ma W, Sze NN, Liu X. COVID-19, community response, public policy, and travel patterns: a tale of Hong Kong. Transp Policy 2021 Jun;106:173-184 [FREE Full text] [doi: 10.1016/j.tranpol.2021.04.002] [Medline: 33846671]
64. Li JC, Cheung CK, Sun IY, Cheung YK, Zhu S. Work-family conflicts, stress, and turnover intention among Hong Kong police officers amid the COVID-19 pandemic. Police Q 2022 Sep;25(3):281-309 [FREE Full text] [doi: 10.1177/10986111211034777] [Medline: 36065392]
65. Corrigan PW, Watson AC, Gracia G, Slopen N, Rasinski K, Hall LL. Newspaper stories as measures of structural stigma. Psychiatr Serv 2005 May;56(5):551-556. [doi: 10.1176/appi.ps.56.5.551] [Medline: 15872163]
66. Chan SK, Li OW, Hui CL, Chang WC, Lee EH, Chen EY. The effect of media reporting of a homicide committed by a patient with schizophrenia on the public stigma and knowledge of psychosis among the general population of Hong Kong. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):43-50. [doi: 10.1007/s00127-018-1610-2] [Medline: 30315332]
67. Lee FL, Yuen S, Tang G, Cheng EW. Hong Kong's summer of uprising: from anti-extradition to anti-authoritarian protests. China Rev 2019 Nov;19(4):1-32 [FREE Full text]
68. Hou WK, Hall BJ, Liang L, Li TW, Liu H, Galea S. Probable depression and suicidal ideation in Hong Kong amid massive civil unrest. Ann Epidemiol 2021 Feb;54:45-51. [doi: 10.1016/j.annepidem.2020.09.006] [Medline: 32949722]
69. Corpuz JC. Adapting to the culture of 'new normal': an emerging response to COVID-19. J Public Health (Oxf) 2021 Jun 07;43(2):e344-e345 [FREE Full text] [doi: 10.1093/pubmed/fdab057] [Medline: 33683346]
70. Mallah SI, Ghorab OK, Al-Salmi S, Abdellatif OS, Tharmaratnam T, Iskandar MA, et al. COVID-19: breaking down a global health crisis. Ann Clin Microbiol Antimicrob 2021 May 18;20(1):35 [FREE Full text] [doi: 10.1186/s12941-021-00438-7] [Medline: 34006330]
71. Budget speech. Level AA Conformance to Web Content Accessibility Guidelines 2.0. 2023. URL: https://www.budget.gov.hk/2023/eng/speech.html [accessed 2024-12-21]
72. Marx W, Manger SH, Blencowe M, Murray G, Ho FY, Lawn S, et al. Clinical guidelines for the use of lifestyle-based mental health care in major depressive disorder: World Federation of Societies for Biological Psychiatry (WFSBP) and Australasian Society of Lifestyle Medicine (ASLM) taskforce. World J Biol Psychiatry 2023 Jun;24(5):333-386 [FREE Full text] [doi: 10.1080/15622975.2022.2112074] [Medline: 36202135]
73. Zhang J, Li Z. The association between depression and suicide when hopelessness is controlled for. Compr Psychiatry 2013 Oct;54(7):790-796 [FREE Full text] [doi: 10.1016/j.comppsych.2013.03.004] [Medline: 23602028]
74. Harris AH, Cronkite R, Moos R. Physical activity, exercise coping, and depression in a 10-year cohort study of depressed patients. J Affect Disord 2006 Jul;93(1-3):79-85. [doi: 10.1016/j.jad.2006.02.013] [Medline: 16545873]
75. Pickett K, Yardley L, Kendrick T. Physical activity and depression: a multiple mediation analysis. Ment Health Phys Act 2012 Dec;5(2):125-134. [doi: 10.1016/j.mhpa.2012.10.001]
76. Roohafza HR, Afshar H, Keshteli AH, Mohammadi N, Feizi A, Taslimi M, et al. What's the role of perceived social support and coping styles in depression and anxiety? J Res Med Sci 2014 Oct;19(10):944-949 [FREE Full text] [Medline: 25538777]
77. Shao R, He P, Ling B, Tan L, Xu L, Hou Y, et al. Prevalence of depression and anxiety and correlations between depression, anxiety, family functioning, social support and coping styles among Chinese medical students. BMC Psychol 2020 Apr 22;8(1):38 [FREE Full text] [doi: 10.1186/s40359-020-00402-8] [Medline: 32321593]
78. Kogan LR, Currin-McCulloch J, Bussolari C, Packman W, Erdman P. The psychosocial influence of companion animals on positive and negative affect during the COVID-19 pandemic. Animals (Basel) 2021 Jul 13;11(7):2084 [FREE Full text] [doi: 10.3390/ani11072084] [Medline: 34359212]
79. Fu KW, Chan YY, Yip PS. Newspaper reporting of suicides in Hong Kong, Taiwan and Guangzhou: compliance with WHO media guidelines and epidemiological comparisons. J Epidemiol Community Health 2011 Oct;65(10):928-933. [doi: 10.1136/jech.2009.105650] [Medline: 20889589]
80. Magaard JL, Seeralan T, Schulz H, Brütt AL. Factors associated with help-seeking behaviour among individuals with major depression: a systematic review. PLoS One 2017 May 11;12(5):e0176730 [FREE Full text] [doi: 10.1371/journal.pone.0176730] [Medline: 28493904]
81. Bengs C, Johansson E, Danielsson U, Lehti A, Hammarström A. Gendered portraits of depression in Swedish newspapers. Qual Health Res 2008 Jul;18(7):962-973. [doi: 10.1177/1049732308319825] [Medline: 18552322]
82. Buechel S, Hahn U. Emobank: studying the impact of annotation perspective and representation format on dimensional emotion analysis. arXiv. Preprint posted online on May 4, 2022 2025 [FREE Full text] [doi: 10.48550/arXiv.2205.01996]
83. Lee LH, Li JH, Yu LC. Chinese EmoBank: building valence-arousal resources for dimensional sentiment analysis. ACM Trans Asian Low Resour Lang Inf Process 2022 Jan 19;21(4):1-18. [doi: 10.1145/3489141]
84. Xie H, Lin W, Lin S, Wang J, Yu LC. A multi-dimensional relation model for dimensional sentiment analysis. Inf Sci 2021 Nov;579:832-844. [doi: 10.1016/j.ins.2021.08.052]

XSL•FO
RenderX

85. Zhao J, Lin J, Liang S, Wang M. Sentimental prediction model of personality based on CNN-LSTM in a social media environment. J Intell Fuzzy Syst 2021 Feb 02;40(2):3097-3106. [doi: 10.3233/JIFS-189348]
86. Nielsen F. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. arXiv. Preprint posted online on March 15, 2011 2025 [FREE Full text] [doi: 10.48550/arXiv.1103.2903]
87. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput Linguist 2011 Jun 01;37(2):267-307. [doi: 10.1162/coli_a_00049]
88. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods 2013 Dec;45(4):1191-1207. [doi: 10.3758/s13428-012-0314-x] [Medline: 23404613]
89. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment analysis of COVID-19 tweets by deep learning classifiers-a study to show how popularity is affecting accuracy in social media. Appl Soft Comput 2020 Dec;97:106754 [FREE Full text] [doi: 10.1016/j.asoc.2020.106754] [Medline: 33013254]
90. Zimbra D, Abbasi A, Zeng D, Chen H. The state-of-the-art in Twitter sentiment analysis: a review and benchmark evaluation. ACM Trans Manage Inf Syst 2018 Aug 24;9(2):1-29. [doi: 10.1145/3185045]
91. Lui ID, Vandan N, Davies SE, Harman S, Morgan R, Smith J, et al. "We also deserve help during the pandemic": the effect of the COVID-19 pandemic on foreign domestic workers in Hong Kong. J Migr Health 2021;3:100037 [FREE Full text] [doi: 10.1016/j.jmh.2021.100037] [Medline: 33817682]
92. Yi Nam S, Yik Chun W, Tak Hing Michael W, Kit Wa Sherry C, Lai Ming Christy H, Ho Ming Edwin L, et al. Double stigma in mental health service use: experience from ethnic minorities in Hong Kong. Int J Soc Psychiatry 2023 Sep;69(6):1345-1353. [doi: 10.1177/00207640231161301] [Medline: 36938981]
93. Chen S, Niu M, Ngai CS. What is the next step of ICT development? The changes of ICT use in promoting elderly healthcare access: a systematic literature review. Heliyon 2024 Feb 07;10(3):e25197 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e25197] [Medline: 38371988]
94. Hong Kong media Survey 2024 results unveiled. Marketing Interactive. 2024 Oct 21. URL: https://www.marketing-interactive.com/hong-kong-media-survey-2024-results-unveiled [accessed 2025-01-21]

Abbreviations

HKFP: Hong Kong Free Press
LDA: latent Dirichlet allocation
MDS: multidimensional scaling
NLP: natural language processing
NLTK: Natural Language Toolkit
RQ: research question
SCMP: South China Morning Post
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by T de Azevedo Cardoso; submitted 20.09.24; peer-reviewed by Y Li, LC Yu; comments to author 29.11.24; revised version received 09.12.24; accepted 23.12.24; published 13.02.25.

Please cite as:
Chen S, Ngai CSB, Cheng C, Hu Y
Analyzing Themes, Sentiments, and Coping Strategies Regarding Online News Coverage of Depression in Hong Kong: Mixed Methods Study
J Med Internet Res 2025;27:e66696
URL: https://www.jmir.org/2025/1/e66696
doi:10.2196/66696
PMID:39946170

©Sihui Chen, Cindy Sing Bik Ngai, Cecilia Cheng, Yangna Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Teenager Substance Use on Reddit: Mixed Methods Computational Analysis of Frames and Emotions

Xinyu Zhang[1], MA; Jianfeng Zhu[2], PhD; Deric R Kenne[3], PhD; Ruoming Jin[2], Prof Dr

[1]College of Communication, Michigan State University, East Lansing, MI, United States

[2]Department of Computer Science, Kent State University, Kent, OH, United States

[3]College of Public Health, Kent State University, Kent, OH, United States

Corresponding Author:
Jianfeng Zhu, PhD
Department of Computer Science
Kent State University
241 Mathematical Sciences Building
Kent, OH, 44224
United States
Phone: 1 330 672 9980
Email: jzhu10@kent.edu

Abstract

Background: Adolescent substance use disorder is a pressing public health issue, with increasing prevalence as individuals age. Social media platforms like Reddit (Reddit Inc) serve as significant venues for teenagers to discuss and navigate substance use. Social media platforms, such as Reddit, serve as increasingly important spaces where teenagers discuss, share, and navigate their experiences with substance use, presenting unique opportunities and challenges for understanding and addressing this issue.

Objective: This study aims to explore how teenagers frame substance-use discussions on the r/teenagers subreddit, focusing on their personal interpretations, causal attributions, and the social and psychological contexts that shape these online support groups. By identifying these interpretive frames, we aimed to better understand the complex drivers of adolescent substance use behavior and their potential interventions.

Methods: Using natural language processing techniques, we analyzed 32,674 substance use–related posts from 2018 to 2022. A framing approach was used to identify and categorize prevalent themes, supplemented by emotional profiling using the EmoLLaMA-chat-13B model developed by Liu and colleagues.

Results: In total, 7 primary frames emerged: normalization, risk awareness, social integration, autonomy and rebellion, coping mechanisms, media influence, and stigmatization. These frames varied in prevalence and were associated with distinct emotional profiles, highlighting the complex interplay between substance use and adolescent experiences. We observed that, for example, the normalization frame was often associated with a mix of sadness and anxiety, while the coping frame exhibited elevated levels of anger, sadness, and anxiety. These distinctive emotional landscapes associated with each frame reveal unique insights into the mental state of adolescents navigating substance use.

Conclusions: The findings underscore the multifaceted nature of adolescent substance-use discussions on social media. Interventions must address underlying emotional and social factors as well as identity to effectively mitigate substance use disorder among adolescents. By understanding the frames teenagers use to interpret substance use, we can pave the way for more effective and personalized public health campaigns, and support services designed to resonate with adolescents' unique lived experiences.

(J Med Internet Res 2025;27:e59338) doi:10.2196/59338

KEYWORDS

teenager; substance use; Reddit; emotional analysis; bidirectional encoder representations from transformers; BERT; frame approach

XSL•FO
RenderX

Introduction

Overview

Adolescent substance use disorder (SUD), characterized by a problematic pattern of substance use causing clinically significant impairment or distress, remains a significant public health concern [1]. Data from the monitoring of the future survey in the United States underscore a clear age-related increase in substance use during adolescence. Cannabis use and nicotine vaping remain particularly prevalent, with 29% of 12th graders, 17.8% of 10th graders, and 8.3% of 8th graders reporting cannabis use in the past year [2]. This trend reflects the progressive escalation of substance use as adolescents age, with nicotine vaping and cannabis being the most commonly used substances [2]. These findings align with the 2019 National Survey on Drug Use and Health, which reported a 4.5% prevalence of SUD among adolescents aged 12-17 years, rising to 14.1% among young adults aged 18-25 years. This further illustrates the role of age in substance use patterns [3]. A global meta-analysis found that older teenagers (aged 15-18 years) exhibit higher rates of substance use, with cannabis and methamphetamine becoming more prevalent as age increases [4]. This progression mirrors US patterns, where substance use typically escalates during mid-to-late adolescence.

The high prevalence of SUD among adolescents is particularly concerning owing to the potential risks and negative impacts associated with substance use [5-8]. SUD poses immediate health risks and can precipitate mental health challenges, infectious diseases like HIV or AIDS, and other negative societal impacts [9]. Families of those struggling with SUD often face profound disruptions, including strained relationships, communication breakdowns, and financial difficulties, all of which could further elevate the risk of SUD transmission across generations [5]. Despite these detrimental impacts, the stigma surrounding drug addiction frequently deters individuals from seeking help through traditional channels such as individual or group counseling [10]. In contrast, the anonymous nature of social media platforms like Reddit enables users to engage in discussions about sensitive topics such as substance-use experiences and seek support for recovery without fear of judgment. These platforms offer insights into a wide array of experiences, underscoring their importance in addressing the challenges of SUD [6].

An increasing number of studies have explored discussions of substance use and mental health on social media platforms such as Reddit. For example, research highlights Reddit's role as a vital support network for opioid recovery, offering emotional and informational support [11,12]. Recent work using Reddit data reveals pandemic-related shifts in discussions about mental health, substance use, and lifestyle changes, highlighting correlations between the pandemic and shifts in health-related behaviors [13,14]. Other research focuses on specific emotional experiences related to substance use [15,16]. Collectively, this body of research illustrates the potential role of social media platforms in monitoring the evolving landscape of mental health and substance-use issues over time.

This research on discussions around substance use on Reddit has generally overlooked 2 critical areas. First, demographic factors, especially age differences, have largely been ignored. For instance, inhalant use is more prevalent among adolescents, whereas substances like nicotine, cannabis, alcohol, opioids, and benzodiazepines are more commonly used by individuals aged 21 to 30 years [17]. Existing studies of substance use discussions on Reddit have mainly focused on specific substances such as alcohol [12], marijuana [18], and opioids [19], disregarding the variations in substance use patterns and substance use among different age groups. Notably, there is a lack of focus on teenagers' discussions about substance use, despite its significance as a growing concern.

Second, there is a substantial gap in understanding the interpretations of substance use by teenagers on these platforms. While existing studies have shed light on aspects such as language use [14], emotional expression [15], prevalent topics [14,20], and topical correlation [13], they fall short in investigating the deeper discourse within these conversations. In addition, research has highlighted how media portrays substance use primarily as an issue of criminal justice [21], but this perspective overlooks the importance of individual framing [22,23]. Particularly in matters with adverse personal and social outcomes, understanding teenagers' personal interpretations and causal attributions is crucial for addressing the challenges associated with substance use.

To address these concerns, this study uses a framing approach to investigate teenagers' discussions about substance use on Reddit, focusing on personal interpretations, causal attributions, and the broader social and psychological contexts influencing these conversations. Reddit is a social media platform consisting of numerous communities called subreddits, each dedicated to specific topics. The subreddit r/teenagers is a community where individuals aged 13 to 19 years share experiences and discuss issues relevant to their age group [24]. Goffman's [25] concept of frames provides the theoretical foundation for our analysis. Frames are "schemata of interpretation" that enable individuals to "locate, perceive, identify, and label" occurrences within their life experiences and the world at large. By analyzing the frames used in r/teenagers discussions, we aim to reveal the interpretive lenses through which teenage users construct meaning around substance use. This approach allows us to uncover the narratives and attributions they use to make sense of the issue and how these framing processes relate to broader social and cultural contexts. Previous research has demonstrated the utility of framing theory in understanding health-related communications and behaviors. For example, studies have shown that framing influences individuals' perceptions and responses to health messages [26]. Applying framing analysis to social media discussions provides a unique opportunity to explore how teenagers frame substance uses in their own words, offering insights that may not be accessible through traditional survey methods.

To systematically analyze the large volume of textual data available on Reddit, we use natural language processing (NLP) techniques. Automated NLP methods enable the processing of extensive datasets, capturing a more comprehensive picture of the discourse as compared with manual qualitative analyses.

XSL•FO
RenderX

This approach enhances the reliability and generalizability of the findings by reducing potential biases introduced through selective sampling and human subjectivity [27]. The NLP techniques have been successfully used in previous studies to identify patterns and themes in social media data related to mental health and substance use [28], demonstrating their effectiveness in uncovering subtle linguistic and thematic patterns that might be overlooked in manual analysis. By leveraging these techniques, we aim to provide deeper insights into teenagers' perspectives on substance use, which can inform targeted interventions and support strategies. Ultimately, we hope that our findings will contribute to the literature on adolescent mental health and substance-use prevention, supporting efforts to address these critical public health concerns.

Literature Review

Teenagers' Substance Use

Adolescent substance use is influenced by a complex interplay of peer dynamics, family environment, romantic relationships, and external factors. Peer influence, crucial in this context, manifests through peer selection and socialization, where adolescents align with peers having similar substance use behaviors and attitudes [8]. Concurrently, family dynamics significantly impact substance use, with factors like parenting styles, parental substance use, and family conflicts playing key roles [29,30]. Adolescents from authoritarian or conflicted family backgrounds, or those with parents who abuse substances, demonstrate a higher propensity for substance use [29].

In the context of romantic relationships, substance use can serve as a response to relationship-induced stress and may also contribute to adolescent dating violence [31]. For example, substance use often initiates at the onset of relationships and may persist or escalate as a coping mechanism for relationship stress and breakups. Research by Rothman et al [31] indicates that adolescents view alcohol consumption as intensifying aggression in dating violence by increasing irritability and anger. Moreover, external factors such as the COVID-19 pandemic have influenced substance use trends. While lockdown-induced reductions in social interactions have generally led to decreased substance use among teenagers [32], Alambo et al [13] observed an uptick in online support group discussions related to substance use disorders during the pandemic, indicating heightened awareness or concern.

Despite the critical insights these factors provide into adolescent substance use, research focusing on teenagers' personal narratives within these contexts remains limited. A notable study in Australia identified four primary narratives among teenagers aged 13 to 15 years who were involved in alcohol and other drug programs: (1) viewing substance use as purposeful, (2) controlled, (3) a component of future competent use, and (4) not necessitating treatment [33]. However, a deeper understanding of how teenagers frame and make sense of substance use within their lived experiences and social contexts is still needed. This study aims to address this gap by using a framing approach [25] to analyze teenagers' narratives on substance use. Frames represent interpretive lenses through which individuals make sense of and assign meaning to issues. By examining the frames used in teenagers' discussions, we can gain insights into how they interpret and ascribe significance to substance use within their lived experiences and social environments.

Frame Approach

Frame research investigates the conceptual and interpretive mechanisms through which individuals comprehend the world. Rooted in interactionism and the School of Chicago, frame theory emphasizes the schemas that help individuals identify, perceive, and categorize experiences within their personal and broader environments [25]. These frames guide the interpretation of experiences and are shared by fellow actors, often taken for granted in daily interactions [25]. This study adopts an interpretive approach to frames, focusing on how individuals construct meaning and make sense of their lived experiences through cognitive schemas or "interpretive lenses."

Understanding the collective framing processes used by teenagers is crucial for several reasons. First, in online support communities such as Reddit's r/teenagers subreddit, users are not merely passive consumers but active creators of content. This user-generated content possesses the potential to influence and modify the interpretive frames of others, fostering a dynamic interplay between personal and collective framing processes [34]. As teenagers engage in discussions, share personal experiences, and collectively interpret substance use issues, collective action frames can organically emerge within the online support community.

Second, frames within social media discourse can significantly influence emotions, thereby shaping public understanding and engagement [35]. For instance, a study analyzing mental health frames on Twitter (X Corp) revealed 7 distinctive frames, including "Awareness," "Feelings and Problematization," "Classification," "Accessibility and Funding," "Stigma," "Service," and "Youth" [23]. These frames not only shape the conversation but also carry emotional weight that can vary in intensity, affecting the audience's likelihood to engage civically [36]. The relationship between emotion and action extends to other areas of public discourse, such as substance use. For instance, a teenager's Reddit post about recreational drug use can evoke a spectrum of emotions, ranging from concern to recognition of stress relief. Understanding these emotional responses is crucial for identifying discussions that may either hinder or promote positive societal change.

Several studies have individually examined the influence of specific frames surrounding substance use. One study explored how framing addiction as a brain disease within genetic and neuroscience research on nicotine addiction could undermine public health efforts, create unrealistic expectations regarding drug therapies, and divert attention from significant social factors [37]. Another study critiqued the framing of recreational substance use as potentially misleading and potentially detrimental [38]. While these studies highlighted issue-specific frames, they may not generalize across different contexts.

To date, there is a paucity of research investigating the frames and emotions associated with substance use discussions among teenagers. This gap is particularly concerning given the

XSL•FO
RenderX

heightened vulnerability of this demographic to substance use disorder and its potential long-term repercussions. Exploring the prevalent frames and emotions in teenagers' discourse surrounding substance use could yield valuable insights for developing tailored and effective prevention and intervention strategies. Consequently, the following research questions (RQs) are proposed: (1) RQ1: What frames do users in r/teenagers use to assign to substance use? (2) RQ2: Which of these frames on substance use are most prominent in these discussions? and (3) RQ3: Which emotions are commonly associated with individual frames in substance use discussions in r/teenagers?

Methods

Data Collection and Preparation

We analyzed originating posts from the r/teenagers subreddit, the largest online support forum for adolescents on Reddit, targeting users aged 13 to 19 years and boasting over 3 million subscribers. This subreddit offers a broad perspective on teenage life, making it a valuable resource for studying adolescent experiences, including substance use. Our analysis was confined to originating English-language posts, defined as the primary submissions made by users to the subreddit, rather than comments or replies to these submissions. This focus allowed us to prioritize the initial framing and self-expression by teenagers when discussing substance use.

To construct our dataset, we obtained 2,145,997 originating posts spanning from 2018 to 2022 through Reddit's submission archives [39]. Posts related to substance use were identified through a systematically curated keyword-based methodology involving 3 distinct stages to ensure comprehensiveness and linguistic inclusivity. First, an initial list of keywords was compiled from authoritative sources, including the 2022 National Survey on Drug Use and Health by Substance Abuse and Mental Health Services Administration, which documents substances commonly used by adolescents; peer-reviewed studies such as Rutherford et al [40]; and the Drug-Use Insights web application [41], a resource providing substance-related terminology observed in social media contexts. Second, this list was expanded to include synonyms, abbreviations, and slang terms frequently used by adolescents, derived from an analysis of online support forums, social media platforms, and adolescent behavioral studies. This step was critical to capturing the diverse linguistic expressions used by teenagers when discussing substance use. Third, the expanded list underwent systematic refinement to eliminate terms deemed ambiguous, redundant, or irrelevant; for example, broad and nonspecific terms such as "party" were excluded, while slang terms directly related to substance use, such as "zaza" (referring to high-grade cannabis), were retained for their contextual relevance. This iterative process resulted in a final list of 150 terms, categorized into substance-use areas (refer to Table 1), enabling the identification of 32,674 substance-use-related posts within our dataset. Table 1 represents the finalized coding schema used to identify substance use-related posts from Reddit's r/teenagers subreddit.

To prepare the extracted posts for analysis, we performed comprehensive text preprocessing to normalize the content and reduce noise. The steps included converting all text to lowercase; expanding contractions and replacing abbreviations using a custom dictionary tailored to adolescent slang (eg, "idk" to "I do not know"); removing HTML tags, numbers, punctuation, and special characters; normalizing elongated words and repeated characters (eg, "loooove" to "love"); removing stop words using natural language toolkit's English stop word list supplemented with context-specific terms; and applying lemmatization to reduce words to their base forms. Duplicate posts were removed based on the cleaned text, and posts that were excessively short (fewer than 30 characters) after cleaning were excluded to ensure data quality.

XSL•FO
RenderX

Table 1. Substance-use keyword schema.

Categories	Keywords
Alcohol use	alcohol, beer, wine, liquor, whiskey, vodka, gin, rum, tequila, drunk, intoxicated, booze, hammered, tipsy, binge drinking, wasted, brandy, cocktail
Tobacco and nicotine use	nicotine, vape, vaping, tobacco, cigarettes, cigs, cigars, e-cigarettes, e-cigs, JUUL, dip, smoking, snuff, chewing tobacco, JUULing, hookah, nicotine patches, nicotine gum
Cannabis use	marijuana, weed, pot, THC[a], cannabis, hash, hashish, ganja, skunk, dabbing, joints, blunts, reefer, Mary Jane, stoned, zaza (slang for high-grade cannabis), kush, 420, CBD[b], hemp
Opioids (Prescription)	oxycodone, OxyContin, hydrocodone, Vicodin, Percocet, codeine, morphine, fentanyl, Dilaudid, Tramadol, painkillers, percs (slang for Percocet), Lorcet, Norco
Benzodiazepines	Xanax, Valium, Ativan, diazepam, clonazepam, Klonopin, alprazolam, lorazepam, temazepam, Restoril, midazolam, Dormicum, oxazepam, Serax, chlordiazepoxide, Librium, estazolam, Prosom, nitrazepam, Mogadon, benzos (slang for benzodiazepines), xans (slang for Xanax), flurazepam, Dalmane, triazolam, Halcion
Stimulants (Prescription)	Adderall, Ritalin, Vyvanse, Dexedrine, amphetamines, study drugs, addy (slang for Adderall)
Opioids (Illicit)	heroin, smack (slang for heroin), black tar, china white, fentanyl, chasing the dragon (slang for inhaling vaporized heroin or fentanyl)
Cocaine	cocaine, yayo (slang for cocaine)
Amphetamine-type stimulants	methamphetamine, meth, crystal meth
MDMA[c]	MDMA, ecstasy, Molly, candy flipping (slang for combining MDMA and LSD[d])
Hallucinogens	LSD, shrooms, psilocybin, magic mushrooms, mescaline, DMT[e], peyote, ayahuasca, 2C-B[f], 2C-E[g], 2C-I[h], DOM[i], STP[j], DOB[k], NBOMe[l], 25I-NBOMe[m], 25C-NBOMe[n], 25B-NBOMe[o], Iboga, Salvia, Psilocin, Bufotenin, 5-MeO-DMT[p], Ergot, Ergine, Morning Glory, LSA[q], Fly Agaric, Mescalito
Dissociative drugs	ketamine, Special K, PCP[r], angel dust, DXM[s], dextromethorphan, robo (slang for DXM abuse), nitrous oxide, whippets, methoxetamine, MXE[t], Mexxy, eticyclidine, PCE[u], DXO[v], 2-oxo-PCE, O-PCE[w], ibogaine, muscimol, tiletamine
Synthetic cannabinoids	K2[x], fake weed, synthetic cannabinoids
Novel psychoactive substances	Flakka, research chemicals, RCs[y], designer drugs

[a]THC: tetrahydrocannabinol.
[b]CBD: cannabidiol.
[c]MDMA: methylenedioxymethamphetamine.
[d]LSD: lysergic acid diethylamide.
[e]DMT: dimethyltryptamine.
[f]2C-B: 2,5-dimethoxy-4-bromophenethylamine.
[g]2C-E: 2,5-dimethoxy-4-ethylphenethylamine.
[h]2C-I: 2,5-dimethoxy-4-iodophenethylamine.
[i]DOM: 2,5-dimethoxy-4-methylamphetamine.
[j]STP: serenity, tranquility, and peace.
[k]DOB: 2,5-dimethoxy-4-bromoamphetamine.
[l]NBOMe: N-methoxybenzyl derivatives of the 2C family.
[m]25I-NBOMe: 2C-I-NBOMe (2-(4-Iodo-2,5-dimethoxyphenyl)-N-(2-methoxybenzyl)ethanamine).
[n]25C-NBOMe: 2C-C-NBOMe (2-(4-Chloro-2,5-dimethoxyphenyl)-N-(2-methoxybenzyl)ethanamine).
[o]25B-NBOMe: 2C-B-NBOMe (2-(4-Bromo-2,5-dimethoxyphenyl)-N-(2-methoxybenzyl)ethanamine).
[p]5-MeO-DMT: 5-methoxy-N,N-dimethyltryptamine.
[q]LSA: lysergic acid amide (D-lysergamide)
[r]PCP: phencyclidine
[s]DXM: dextromethorphan
[t]MXE: methoxetamine
[u]PCE: eticyclidine (N-ethyl-1-phenylcyclohexan-1-amine)
[v]DXO: dextrorphan
[w]O-PCE: N-ethyl-1-(2-fluorophenyl)-2-(methylamino)propan-1-one.

XSL•FO
RenderX

[x]K2: synthetic cannabinoids.

[y]RCs: research chemicals.

Data Analysis

The labeling of Reddit posts was conducted through a systematic approach combining machine learning techniques with qualitative analysis. Initially, the BERT (Bidirectional Encoder Representations from Transformers)–based topic modeling algorithm identified topic clusters by capturing semantic relationships and patterns within the text. These clusters were provisionally labeled based on the most representative terms and phrases extracted by the algorithm.

To refine these labels, the author independently reviewed a representative sample of 30 posts from each cluster, conducting inductive coding to identify emergent themes, linguistic nuances, and contextual elements not fully captured by the algorithm. Following this independent analysis, collaborative discussions were held to reconcile differing interpretations, refine thematic categories, and develop operational definitions for each label. This iterative process ensured that the labels accurately represented the content and context of the posts, facilitating robust, contextually grounded labels tailored to the unique characteristics of the Reddit dataset.

Subsequently, the 34 identified topics were organized into 6 distinct frames, and the distribution of posts across these frames was calculated to assess their prevalence within the dataset. In the *Results* section, detailed descriptions and examples of these topics and frames are provided, illustrating their thematic relevance and enhancing the interpretability of the findings.

To analyze the sentiment characteristics of posts, we used the EmoLLaMA-chat-13B model, a state-of-the-art tool within the EmoLLMs series built on Meta's LLaMA2-chat-13B and fine-tuned using the affective analysis instruction dataset. This model notably outperforms previous state-of-the-art models on benchmarks like SemEval-2018, particularly in ordinal emotion intensity classification and emotion intensity regression [42]. In addition, it consistently outperforms advanced alternatives, including GPT-4 and ChatGPT, across various sentiment analysis tasks. It calculates emotion intensity scores for 7 core emotions: anger, sadness, anxiety, fear, worry, hope, and loneliness. These scores quantified the degree to which each emotion was expressed in the posts, allowing nuanced detection of emotional tones. For example, a post with scores of 0.33 for anger and 0.5 for sadness indicated moderate anger and substantial sadness, while other emotions remained minimal. This integrated approach facilitated a robust understanding of the emotional and thematic context of substance use discussions among adolescents. The average emotion scores for each frame were visualized using a heat map, as shown in Figure 1, further illustrating the emotional dynamics underlying these conversations. This heat map visualizes average intensity scores for 7 emotions across 7 substance-use frames identified in Reddit's r/teenagers posts (2018-2022). Color intensity indicates the degree of emotion expression, ranging from 0.05 (lighter blue) to 0.25 (darker blue) based on EmoLLaMA-chat-13B model profiling.

Figure 1. Frame-specific emotion intensity heat map.

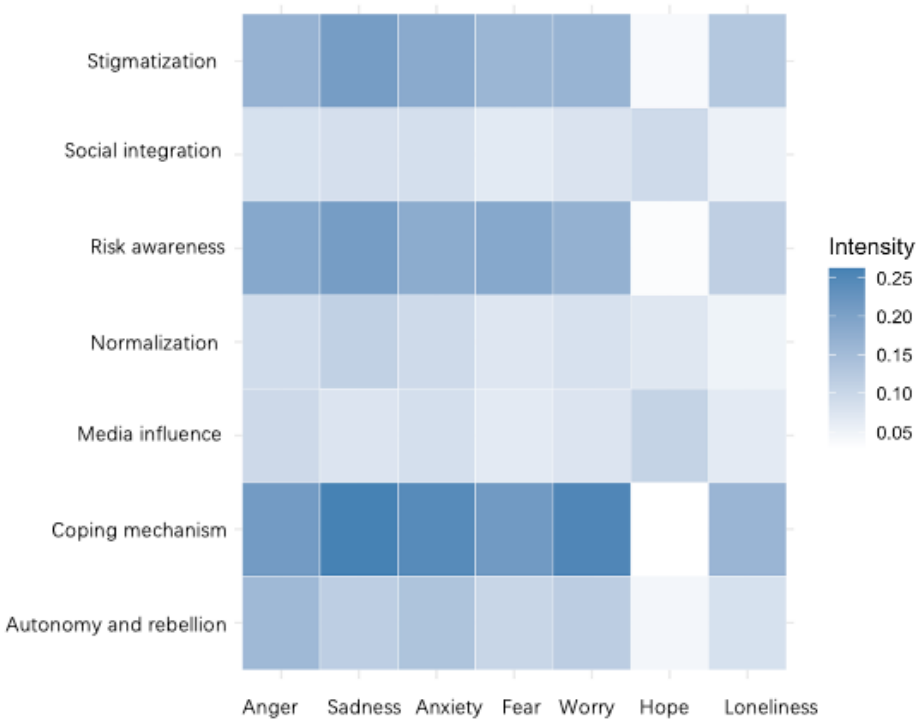

XSL•FO
RenderX

Ethical Considerations

The procedures conducted in this study adhered to ethical standards and were exempted from institutional review board review, as the research involved publicly available, anonymized data containing no identifiable information [28]. To further safeguard privacy, all posts included in the results were paraphrased to preserve their original meaning while ensuring that they could not be traced back to individual users.

Results

Overview

Table 2 displays the 3 topics identified by BERTopic. The "KeyBERT" column shows the top keywords associated with each topic, while the "Custom label" column was generated by our team through qualitative assessment. This process involved reviewing the top keywords and representative documents for each identified topic. Table 2 lists the 34 topics identified via BERTopic modeling from 32,674 Reddit posts on r/teenagers (2018-2022) related to substance use, along with their corresponding keywords ("KeyBERT") and custom labels derived through qualitative analysis.

In accordance with various frames from other studies and the framing attribute of substance use, our team collaboratively analyzed and interpreted the themes to identify 7 overarching frames related to substance use within the r/teenagers subreddit discussions. These frames were normalization (17014/32674, 38.14%), risk awareness (1379/32,674, 4.22%), social integration (9573/32,674, 29.3%), autonomy and rebellion (1146/32,674, 3.51%), coping mechanism (5762/32,674, 17.64%), digital media influence (1310/32,674, 4.01%), stigmatization (1042/32,674, 3.19%). The identification of these frames highlights the interconnectedness of various factors influencing adolescent substance use, including social, familial, psychological, and cultural influences (refer to Figure 1).

XSL•FO
RenderX

Table 2. BERTopic-generated topics on adolescent substance use.

ID	KeyBERT[a]	Custom label	Percentage (N=32,674), n (%)
0	dad, mom, family	Family dynamics reinforcing substance use	6495 (19.88)
1	friend, girl, like	Peer relationships encouraging substance use	5093 (15.59)
2	personal, event, back	Stress coping through substance use	4099 (12.55)
3	school, teacher, class	Stress from school normalizing substance use	2552 (7.81)
4	juul, vape, vaping	Vaping as part of everyday life	1892 (5.79)
5	social, accept, drug	Substances as tools for peer acceptance	1689 (5.17)
6	birthday, party, friend	Party culture and group bonding	1359 (4.16)
7	feel, crisis, suicide	Mental health struggles and self-medication	1157 (3.54)
8	alcohol, taste, drinking	Teenage drinking as a social routine	1111 (3.40)
9	media, alcohol, rubbing	Social media promoting substance glamor	833 (2.55)
10	age, alcohol, drinking	Challenging legal drinking and vaping ages	753 (2.25)
11	anti-vaping, ads, campaign	Anti-vaping campaigns and social stigma	647 (1.98)
12	chat, drunk, social media	Drunken interactions on social media	618 (1.89)
13	concern, survey, test	Drug testing and its social implications	484 (1.48)
14	christmas, party, nicotine	Celebratory use of nicotine	412 (1.26)
15	iphone, phone, apple	Platforms encouraging substance trends	376 (1.15)
16	drunk, love, tomorrow	Romanticizing drinking in social settings	363 (1.11)
17	friend, social, group	Social situations facilitating substance use	350 (1.07)
18	died, death, driver	Death and accidents from substance use	284 (0.87)
19	covid, pandemic, virus	COVID-19 and awareness of substance risks	268 (0.82)
20	environment, health, pollution	Environmental impacts of substance use	261 (0.80)
21	juul, buy, access	Accessing restricted substances	209 (0.64)
22	sleep, drunk, alcohol	Using alcohol to manage sleep deprivation	196 (0.60)
23	religion, Muslim, god	Religion’s moral judgment of substance use	176 (0.54)
24	soda, nicotine, celebration	Alcohol as a comfort	176 (0.54)
25	poem, music, writing	Personal expression reflecting emotional struggles	134 (0.41)
26	drunk, friend, trend	Online portrayals of substance use	101 (0.31)
27	consent, rape, alcohol	Alcohol, consent, and youthful rebellion	101 (0.31)
28	russia, america, geopolitics	Sociopolitical resistance and substance use	101 (0.31)
29	game, play, video	Social connections through gaming	101 (0.31)
30	injury, surgery, shoulder	Health consequences of substance use	82 (0.25)
31	adderall, medication, price	The role of medication in adolescence	82 (0.25)
32	irish, culture, nationality	Cultural identity and substance perception	75 (0.23)
33	gay, trap, chromosome	LGBTQ+[b] substance use and social stigma	62 (0.19)

[a]BERT: Bidirectional Encoder Representations from Transformers.

[b]LGBTQ+: lesbian, gay, bisexual, trans, and queer.

Frame 1: Normalization

The normalization frame, accounting for 38.14% (12462/32674) of the discussions, reveals a substantial perception among adolescents that substance use is an accepted and routine part of daily life. This frame intertwines several topics: family dynamics reinforcing substance use (6495/32,674, 19.88%), stress from school normalizing substance use (2552/32,674, 7.81%), vaping as part of everyday life (1892/32,674, 5.79%), teenage drinking as a social routine (1111/32,674, 3.40%), and celebratory use of nicotine (412/32,674, 1.26%). Collectively, these topics illustrate how various facets of adolescents’ environments contribute to the normalization of substance use.

XSL•FO
RenderX

Family behaviors and attitudes play a significant role, with substance use often perceived as a familial norm. Academic pressures lead to shared coping mechanisms like vaping and drinking, embedding these behaviors into daily routines. Social interactions further reinforce this normalization, as substance use becomes an expected component of communal activities and celebrations.

These shared understandings coalesce into collective action frames by fostering a collective identity where substance use is not only normalized but also encouraged. Adolescents internalize these behaviors as standard, leading to widespread participation and reinforcement within their social groups. One adolescent reflected, “In my family, everyone unwinds with a drink; so when I started drinking with friends, it just felt natural.” Another shared, “Vaping between classes is how we deal with stress—it’s just what we all do to get through the day.” Such narratives demonstrate how substance use is reinforced through family acceptance, peer practices, and shared coping strategies, solidifying its place in adolescent culture.

Emotional analysis within the normalization frame reveals moderate levels of sadness (0.092) and anxiety (0.112), indicating underlying emotional distress despite the outward acceptance of substance use. The topic “Stress from School Normalizing Substance Use” exhibits higher levels of sadness (0.14) and anxiety (0.18), suggesting that academic pressures significantly contribute to emotional turmoil and the adoption of substance use as a coping mechanism. Lower levels of positive emotions such as joy (0.072) imply that normalization does not necessarily enhance overall well-being. This suggests that while substance use is normalized, it may also be a response to deeper emotional challenges, further perpetuating its role in adolescents’ lives.

Frame 2: Risk Awareness

The risk awareness frame accounts for 4.22% (1379/32,674) of the discussions, indicating a growing consciousness among adolescents about the negative consequences of substance use. This frame interconnects topics such as “Drug Testing and Its Social Implications” (484/32,674, 1.48%), “Death and Accidents from Substance Use” (284/32,674, 0.87%), “Environmental Impacts of Substance Use” (261/32,674, 0.80%), “COVID-19 and Awareness of Substance Risks” (268/32,674, 0.82%), and “Health Consequences of Substance Use” (82/32,674, 0.25%). Collectively, these discussions reflect an increased awareness of health risks, legal implications, and broader societal impacts associated with substance use.

The association among these topics illustrates how adolescents are developing a comprehensive understanding of the multifaceted risks involved. They express concerns not only about personal health but also about legal repercussions and the broader impact on their communities and the environment. This shared awareness leads to a collective reassessment of substance use, as adolescents weigh the perceived benefits against potential negative outcomes. One teenager remarked, “After my school started random drug testing, I realized how serious the consequences could be—not just for me but for my family too.” Another shared, “When I heard about someone from our town dying in a drunk driving accident, it hit me hard. It made me think about how my actions could affect others.”

These shared understandings coalesce into a collective action frame centered on caution and responsibility. Adolescents are not only reevaluating their own behaviors but are also contributing to a broader dialogue about the risks associated with substance use. The heightened emotional responses—elevated levels of sadness (0.188), anxiety (0.208), fear (0.188), and worry (0.181)—underscore the depth of their concern. As one individual noted, “With COVID-19 affecting people’s lungs, continuing to vape just doesn't seem worth it anymore.” Another stated, “I’m starting to see how smoking and littering impact the environment—we have to think about the bigger picture.” These narratives demonstrate a collective shift towards greater risk awareness and a desire to make more informed, responsible choices.

Frame 3: Social Integration

Representing 29.30% (9573/32,674) of the discussions, the social integration frame underscores how adolescents use substance use to navigate social dynamics and achieve peer acceptance. This frame interweaves topics such as “Peer Relationships Encouraging Substance Use” (5093/32,674, 15.59%), “Substances as Tools for Peer Acceptance” (1689/32,674, 5.17%), “Party Culture and Group Bonding” (1359/32,674, 4.16%), “Social Situations Facilitating Substance Use” (618/32,674, 1.07%), “Drunken Interactions on Social Media” (363/32,674, 1.89%), “Romanticizing Drinking in Social Settings” (350/32,674, 1.11%), and “Social Connections Through Gaming” (101/32,674, 0.31%). Collectively, these topics illustrate the role of substance use as a means for group bonding and social interaction, revealing the complexities of adolescent social behavior.

The interconnections among these topics highlight a collective understanding among teenagers that substance use is integral to social belonging and acceptance. Adolescents often feel pressured to engage in substance use to align with peer norms and avoid isolation. One teenager shared, “If you’re not drinking at the party, it’s like you’re not really there. Joining in just makes you part of the group.” This sentiment reflects the drive to conform and the importance of participating in shared behaviors to maintain social bonds. Another adolescent noted, “Vaping with my friends between classes is just what we do; it’s our way of hanging out and decompressing together.” These narratives demonstrate how substance use becomes embedded in social rituals, reinforcing group cohesion and collective identity.

Moreover, the romanticization of drinking in social settings and the portrayal of “drunken interactions on social media” amplify this phenomenon, solidifying substance use’s role in adolescent culture. An individual observed, “Some of our best memories are made when we’re all a bit tipsy, laughing and just enjoying each other’s company.” Social media platforms often showcase these moments, further reinforcing the behavior. Despite this perceived enhancement of social connections, the emotional analysis reveals elevated levels of sadness (0.083), anxiety (0.085), and loneliness (0.054), suggesting that reliance on substances for social integration may not fully alleviate

XSL•FO
RenderX

underlying emotional distress. This collective action frame reflects a complex interplay between the desire for social connection and the potential emotional costs associated with substance use, highlighting the need for a deeper understanding of adolescent social dynamics.

Frame 4: Autonomy and Rebellion

Comprising 3.51% (1146/32,674) of the discussions, the autonomy and rebellion frame highlights how adolescents use substance consumption as a means of asserting independence and resisting authority. This frame intertwines topics such as "Challenging Legal Drinking and Vaping Ages" (735/32,674, 2.25%), "Accessing Restricted Substances" (209/32,674, 0.64%), "Sociopolitical Resistance and Substance Use" (101/32,674, 0.31%), and "Alcohol, Consent, and Youthful Rebellion" (101/32,674, 0.31%). Collectively, these topics reveal a shared desire among teenagers to assert control over their own decisions and to challenge societal restrictions, illustrating how substance use becomes a symbolic act of personal freedom and defiance.

The interconnections among these topics demonstrate a collective action frame where adolescents perceive substance use as a way to test boundaries and resist external control. One teenager asserted, "Why should the law tell me when I'm old enough to make my own choices? If I'm responsible enough to work and pay taxes, I should be able to decide if I want to drink." This sentiment reflects a common challenge to age-based regulations, emphasizing a desire for self-determination. Another shared, "They keep making it harder to get vapes, but that just makes us more creative. It's about proving that we won't be controlled." Such narratives underscore how obtaining and using restricted substances becomes a form of rebellion against perceived overreach by authorities.

Moreover, some adolescents connect their personal defiance to broader sociopolitical contexts. An individual remarked, "Using substances is our way of pushing back against a system that doesn't listen to us. It's not just about drinking or vaping; it's about showing that we have a voice." This perspective highlights how substance use can be intertwined with a larger critique of societal structures, reinforcing group identity through shared acts of resistance. Emotionally, this frame exhibits elevated levels of anger (0.156), sadness (0.114), and anxiety (0.136), reflecting the frustration and tension associated with challenging authority. Despite relatively low levels of hope (0.045) and loneliness (0.083), adolescents in this frame find solidarity in their collective pursuit of autonomy, united in their challenge to societal norms and expectations.

Frame 5: Coping Mechanism

Accounting for 17.64% (5762/32,674) of the discussions, the coping mechanism frame illustrates how adolescents use substances to manage emotional stress and personal challenges. This frame integrates topics such as "Stress Coping Through Substance Use" (4099/32,674, 12.55%), "Mental Health Struggles and Self-Medication" (1157/32,674, 3.54%), "Using Alcohol to Manage Sleep Deprivation" (196/32,674, 0.60%), "Alcohol as a Comfort" (176/32,674, 0.54%), and "Personal Expression Reflecting Emotional Struggles" (134/32,674, 0.41%). Collectively, these topics reflect the significant role that emotional distress plays in adolescents' decisions to engage in substance use, highlighting how substances become a means of coping with various stressors in their lives.

The interconnection among these topics underscores a shared understanding that substance use serves as a mechanism for self-medication and emotional regulation. Adolescents facing stress, mental health challenges, sleep deprivation, or emotional struggles turn to substances as temporary relief, seeking to alleviate negative emotions and escape from their problems. This collective behavior demonstrates how substance use becomes an ingrained coping strategy within this demographic, forming a collective action frame centered on managing personal hardships through substance use.

Teenagers' narratives vividly capture this phenomenon, illustrating how their shared experiences coalesce into this collective frame. One adolescent admitted, "When school gets overwhelming, vaping is the only thing that helps me relax." Another shared, "I can't sleep at night, so sometimes I drink a little to help me drift off." These statements reveal how substance use is perceived not merely as recreational but as a necessary tool for managing emotional and physical distress. Further emphasizing this point, an adolescent explained, "Dealing with anxiety is tough, and smoking weed is the only way I can feel normal." Another added, "Writing songs while I'm high helps me express feelings I can't put into words otherwise." These authentic accounts highlight the reliance on substances to navigate their emotional landscapes and the collective interpretation of substance use as essential for coping.

Emotionally, this frame exhibits elevated levels of sadness (0.212), anxiety (0.262), and anger (0.244), indicating significant emotional distress among adolescents using substances to cope. Notably, the topic "Mental Health Struggles and Self-Medication" shows particularly high levels of sadness (0.23), anxiety (0.28), and loneliness (0.18), reflecting severe emotional turmoil. One post expressed, "My depression gets so bad that drinking is the only thing that numbs the pain, even if just for a little while." Another shared, "I know relying on substances isn't healthy, but it's the only way I can get through the day." The elevated levels of fear (0.215) and worry (0.253) further underscore the depth of their struggles. The relatively low level of hope (0.029) suggests a lack of optimism about their situations, reinforcing the reliance on substances for temporary relief. These shared sentiments illustrate how adolescents, through collective experiences and interpretations, normalize substance use as a coping mechanism for emotional distress. This collective action frame emphasizes the urgent need for supportive interventions, as teenagers mobilize around shared understandings to navigate their challenges.

Frame 6: Media Influence

Accounting for 4.01% (1310/32,674) of the discussions, the media influence frame examines how social media and digital platforms shape adolescents' perceptions of substance use through media socialization. This frame integrates topics such as "Social Media Promoting Substance Glamor" (833/32,674, 2.55%), "Platforms Encouraging Substance Trends" (376/32,674, 1.15%), and "Online Portrayals of Substance Use"

XSL•FO
RenderX

(101/32,674, 0.31%). Collectively, these topics highlight the significant role of media in normalizing and glamorizing substance use among adolescents, thereby influencing their attitudes and behaviors.

The interconnections among these topics demonstrate the mechanisms by which media-driven norms operate. Adolescents are frequently exposed to portrayals of substance use as desirable and socially advantageous, fostering a perception of these behaviors as normative. This collective exposure leads to shared cognitive schemas and social comparisons that reinforce these impressions. The prevalence of substance-related content on social media platforms encourages adolescents to perceive substance use as an integral part of social identity and acceptance.

Teenagers' narratives vividly capture this phenomenon, illustrating how their shared experiences coalesce into this collective frame. One adolescent remarked, "Scrolling through Instagram, it seems like everyone's partying and having a great time with drinks in hand. It makes you feel like you're missing out if you're not doing the same." Another shared, "All the popular kids are posting videos with vapes and flashy setups. It's like you have to join in to be part of the trend." These authentic accounts highlight the influence of media representations in shaping adolescents' behaviors, as they adopt substance use to conform to perceived social expectations promoted by online support groups.

Emotionally, this frame is associated with elevated levels of sadness (0.096) and anxiety (0.086), suggesting that media portrayals contribute to negative effects among adolescents. The relatively high level of hope (0.108) indicates a complex emotional response, where optimism is intertwined with negative emotions. One user confessed, "I thought posting pics of me at parties would make me feel more connected, but I still feel alone when the likes stop coming in." Another reflected, "Seeing everyone else having fun online support group makes me wonder if they're really happy or just putting on a show like I am." These shared sentiments underscore how media influence fosters a collective understanding that substance use is necessary for social inclusion, while also highlighting the emotional complexities involved. Through these shared experiences, adolescents collectively reinforce the media influence frame, which shapes their perceptions and actions regarding substance use.

Frame 7: Stigmatization

The stigmatization frame, accounting for 3.19% (1042/32,674) of the discussions, centers on the moral and social judgments surrounding adolescent substance use, particularly within familial, religious, and cultural contexts. This frame includes topics such as "Anti-vaping Campaigns and Social Stigma" (647/32,674, 1.98%), "Religion's Moral Judgment of Substance Use" (176/32,674, 0.54%), "LGBTQ+ Substance Use and Social Stigma" (62/32,674, 0.19%), "Cultural Identity and Substance Perception" (75/32,674, 0.23%), and "The Role of Medication in Adolescence" (82/32,674, 0.25%). These discussions collectively highlight how adolescents face criticism and moral condemnation from societal norms and authority figures, shaping their behavior and self-perception.

The interconnectedness of these topics underscores the pervasive experience of moral judgment, leading to the stigmatization that fosters guilt, shame, or defiance among adolescents. Antivaping campaigns often reinforce social ostracization, as reflected in the sentiment of one adolescent: "I feel like everyone looks down on me just because I vape—like I'm some kind of delinquent." Such societal disapproval adversely impacts adolescents' self-image, reinforcing isolation. Religious and cultural expectations further intensify this stigmatization, as illustrated by another adolescent's experience: "In my community, drinking is considered a sin, but it's hard when all my friends do it. I constantly feel torn between my beliefs and wanting to fit in." These statements underscore the internal conflicts adolescents endure while negotiating personal desires and cultural values.

The stigmatization faced by lesbian, gay, bisexual, trans, and queer (LGBTQ+) adolescents is particularly acute, as they often experience dual layers of judgment concerning both their identity and substance use. One individual conveyed, "It's hard enough being judged for who I am, but when they find out I smoke to cope, the judgment just doubles." This compounded stigma exacerbates their sense of alienation. Similarly, the topic "The Role of Medication in Adolescence" reflects the stigmatization of prescribed medication use, even when these substances are essential for managing mental health. As one adolescent stated, "People think I'm weak because I take meds for my anxiety, but they don't understand how much it helps." These narratives demonstrate that stigmatization extends beyond illicit substances, affecting adolescents' willingness to seek help and manage their health needs effectively.

The emotional dynamics of the stigmatization frame are marked by high levels of sadness (0.168), anxiety (0.183), and anger (0.161), indicating significant emotional distress due to societal and familial judgment. The topic "Religion's Moral Judgment of Substance Use" exhibits particularly high levels of anger (0.256) and loneliness (0.145), suggesting that moral condemnation tied to religious beliefs exacerbates feelings of isolation and resentment. Moreover, the low levels of hope (0.039) within this frame indicate adolescents' difficulty in finding optimism amid pervasive stigmatizations. One adolescent reflected, "Sometimes it feels like no matter what I do, I'll never meet their expectations." These narratives illustrate a collective experience of stigmatization that impedes well-being, reinforcing a shared perception of being marginalized and judged.

Discussion

Principal Findings

The findings of this study provide insights into interpretive frames adolescents use to construct and negotiate their perspectives on substance use, revealing intricate connections to broader sociocultural, emotional, and developmental contexts. At a conceptual level, the research underscores how adolescents engage with substances used as a mechanism to address sociocultural and identity-related challenges, situating these behaviors within collective-action frames that reflect both individual agencies and shared cultural dynamics. This framing

XSL•FO
RenderX

highlights substance use not as an isolated behavior but as a socially embedded phenomenon with deep implications for adolescent development.

One key contribution is the identification of the normalization frame, which captures how substance use becomes embedded in adolescents' everyday lives through familial, social, and institutional influences. Unlike previous studies that narrowly focused on peer influence or environmental factors, this research situates normalization within a wider cultural framework. It demonstrates how academic stress, parental modeling, and peer interactions collectively foster a shared perception that substances are used as routine. Such findings broaden the scope of existing literature by emphasizing the systemic reinforcement of these behaviors across multiple social domains.

The emergence of the risk awareness frame provides a compelling counter-narrative to normalization, illuminating adolescents' critical reflections on substance-related risks despite their concurrent acceptance of such behaviors. This duality complicates simplistic dichotomies that frame adolescents as either deviant or compliant. Consistent with previous studies [43], this study shows heightened risk awareness, particularly in the wake of the pandemic, which intensified adolescents' health consciousness and self-reflective capacities. The findings also resonate with Marinelli et al [44], who documented how the pandemic prompted adolescents to re-evaluate habitual behaviors, fostering a unique environment for behavioral reassessment.

The social integration frame further underscores the role of substance use in mediating adolescent social dynamics, positioning substances as tools for navigating group identity and peer acceptance. However, this narrative is complicated by findings that reveal underlying emotional vulnerabilities, such as sadness and anxiety; these findings suggest that while substance use may facilitate social connection, it also masks significant emotional distress. This creates a paradoxical feedback loop in which substance use both addresses and exacerbates emotional challenges, underscoring the need for interventions that address these deeper vulnerabilities.

Adding another layer to this analysis, the autonomy and rebellion frame portrays substance use as a symbolic act of resistance against societal and institutional constraints. This dimension situates adolescent substance use within a broader context of sociopolitical critique, where behaviors function as both a developmental milestone of individuation and a performative challenge to authority. These findings align with theoretical perspectives on youth subcultures, which emphasize resistance as a core characteristic while raising critical questions about the efficacy of punitive or prohibitive approaches to regulating adolescent behaviors.

The coping mechanism frame provides perhaps the most poignant insights, revealing that emotional distress (manifested in elevated levels of sadness, anxiety, and anger) is a pervasive driver of substance use among adolescents. This frame underscores the urgent need for mental health interventions that address the root causes of emotional distress rather than merely its behavioral manifestations. By situating these coping behaviors within a collective framework, the study highlights how adolescents share and normalize their struggles, entrenching substance use as a widely accepted strategy for emotional regulation. The lack of hope observed in this frame reflects systemic failures to provide adolescents with viable alternatives for managing psychological challenges, further reinforcing the importance of holistic and proactive support systems.

The media influence frame introduces a contemporary perspective, illustrating how social media amplifies cultural narratives around substance use by portraying it as glamorous or socially advantageous. While such portrayals exert considerable influence, the study also reveals adolescents' growing skepticism and critique of these representations, reflecting an awareness of their constructed nature. This dual engagement suggests potential avenues for media literacy programs aimed at deconstructing harmful narratives and promoting more critical consumption of digital content.

Lastly, the stigmatization frame explores how moral and cultural judgments surrounding substance use profoundly affect adolescents' experiences, contributing to heightened feelings of sadness, anxiety, and anger. This frame is particularly salient for marginalized groups, such as LGBTQ+ adolescents, who face compounded stigmatization based on both identity and behavior. The findings align with critiques of stigma-based deterrents, advocating instead for inclusive approaches that address systemic inequalities and provide supportive alternatives to punitive or judgmental responses. Therefore, the focus needs to be shifted away from blame and judgment and towards inclusive approaches that address inequality and offer alternative supportive environments, as has also been discussed in light of the pandemic by Marinelli et al [44].

Moreover, the findings of this study are particularly pertinent in light of the ever-evolving landscape of new psychoactive substances (NPSs), which pose a significant and novel challenge to public health and adolescent mental health in particular. While our research focuses on how teenagers frame and interpret substance use in online support discussions, it is critical to acknowledge the broader context in which this occurs. As Rinaldi et al [45] emphasize, the complexities of these NPSs (their diverse chemical structures and their ability to evade traditional drug enforcement) create novel risks. These substances are not just a theoretical concern; they can be found on illicit markets and, increasingly, online support groups, shaping the very discourse we analyzed on Reddit. Moreover, consistent with Meherali et al [46], the isolation and social disruption brought about by the COVID-19 pandemic, which potentially exacerbated online support interactions, have created new opportunities for illicit drug markets, as noted by Marinelli et al [44] and Negro et al [47]. This highlights the importance of online support group spaces, where adolescents are engaging in substance use discussions, as potential avenues for both risk and intervention. This understanding calls for a need to not only analyze the frames used to discuss substance use but also to address the underlying context of rapidly evolving substance markets and information streams that influence these narratives.

Theoretically, this study integrates these diverse frames into a cohesive framework, challenging reductionist models that view substance use as purely pathological or recreational. It highlights

XSL•FO
RenderX

the interplay between individual agency and collective dynamics, offering a more comprehensive perspective on the social, emotional, and cultural dimensions of substance use. Practically, the findings advocate for targeted interventions tailored to the specific interpretive frames adolescents employ. For instance, strategies addressing normalization might focus on disrupting familial and peer modeling behaviors, while those targeting coping mechanisms should prioritize mental health resources and resilience-building initiatives.

By contextualizing substance use within broader sociocultural and emotional frameworks, this research deepens our theoretical understanding of adolescent behavior while offering actionable insights for fostering healthier environments. It calls for a shift in both policy and practice, emphasizing the importance of addressing systemic factors and respecting adolescents' lived experiences to promote meaningful and sustainable change.

Limitations and Futural Studies

While our study provides valuable insights into adolescent discourse on substance use, it is limited by the exclusive use of data from Reddit's r/teenagers subreddit, an English-speaking online support forum. This introduces inherent biases, restricting the generalizability of our findings to the global adolescent population, as the perspectives captured predominantly reflect Reddit's specific user demographic rather than a comprehensive view of teenage behaviors and attitudes worldwide. Moreover, our analysis does not fully contextualize substance use within the dynamics of the COVID-19 pandemic and the emergence of NPSs. Although we included keywords related to NPSs to capture emerging substances, we did not address the complex trafficking dynamics, pandemic-related changes, or broader mental health implications, which limits the depth of our findings concerning current societal influences on adolescent substance use.

Future research should address these limitations by incorporating more diverse data sources to enhance generalizability and by situating substance use within the context of the COVID-19 pandemic and NPS proliferation. Investigating discrepancies between online support discussions and offline attitudes can illuminate the influence of digital environments on real-world behaviors. In addition, examining the impact of significant societal events, such as public health crises and changes in drug legislation, on online support discourse could provide insights into the adaptability of adolescent attitudes and behaviors. By expanding the scope to include diverse data sources and contextual factors, future studies can deepen our understanding of adolescent substance use and inform more effective interventions. Such research would enhance the relevance and applicability of findings, ultimately contributing to solutions for the complex issues surrounding adolescent substance use.

Conclusion

This study reveals that adolescents' discussions about substance use on social media are predominantly framed around normalization, social integration, coping mechanisms, risk awareness, autonomy and rebellion, media influence, and stigmatization. The normalization and social integration frames indicate that substance use is perceived as a routine part of adolescent life, reinforced by family dynamics, peer relationships, and social media portrayals. Coping mechanisms highlight the use of substances to manage emotional distress, reflecting significant mental health challenges among youth. Although there is growing risk awareness, this consciousness does not consistently lead to behavioral change. The autonomy and rebellion frame illustrates substance use as a means for adolescents to assert independence and resist authority, while the stigmatization frame emphasizes the moral and social judgments contributing to feelings of isolation.

These findings underscore the multifaceted nature of adolescent substance use discussions and the necessity of addressing underlying emotional, social, and identity-related factors. Interventions should consider these frames to develop strategies that resonate with adolescents' experiences and perceptions. By fostering supportive environments and promoting open dialogues about mental health and identity, health communication efforts can more effectively mitigate substance use among adolescents.

Acknowledgments

The discussion was polished using ChatGPT [48] by OpenAI and subsequently reviewed and revised by the authors.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Miech R, Johnston L, Patrick M, O'Malley P, Bachman J, Schulenberg J. Monitoring the future national survey results on drug use, 1975-2022: secondary school students. Institute for Social Research. 2023. URL: https://eric.ed.gov/?id=ED604018 [accessed 2024-01-25]
2. Patrick M, Miech R, Johnston L, O?Malley P. Monitoring the future panel study annual report: national data on substance use among adults ages 19 to 65, 1976-2023. Institute for Social Research, University of Michigan. 2024. URL: https://monitoringthefuture.org/wp-content/uploads/2023/07/mtfpanel2023.pdf [accessed 2025-01-28]

XSL•FO
RenderX

3. Branca Vergano L, Monesi M, Vicenti G, Bizzoca D, Solarino G, Moretti B. Key substance use and mental health indicators in the United States: results from the 2022 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. 2020. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt42731/2022-nsduh-nnr.pdf [accessed 2025-01-28]
4. Armoon B, Griffiths MD, Mohammadi R. The global distribution and epidemiology of psychoactive substance use and injection drug use among street-involved children and youth: a meta-analysis. Subst Use Misuse 2023;58(6):746-764. [doi: 10.1080/10826084.2023.2181036] [Medline: 36924274]
5. Lander L, Howsare J, Byrne M. The impact of substance use disorders on families and children: from theory to practice. Soc Work Public Health 2013;28(3-4):194-205 [FREE Full text] [doi: 10.1080/19371918.2013.759005] [Medline: 23731414]
6. Chi Y, Chen HY. Investigating substance use via Reddit: systematic scoping review. J Med Internet Res 2023 Oct 25;25:e48905 [FREE Full text] [doi: 10.2196/48905] [Medline: 37878361]
7. Jordan A, Quainoo S, Nich C, Babuscio TA, Funaro MC, Carroll KM. Racial and ethnic differences in alcohol, cannabis, and illicit substance use treatment: a systematic review and narrative synthesis of studies done in the USA. Lancet Psychiatry 2022 Aug;9(8):660-675. [doi: 10.1016/S2215-0366(22)00160-2] [Medline: 35752192]
8. Henneberger AK, Mushonga DR, Preston AM. Peer influence and adolescent substance use: a systematic review of dynamic social network research. Adolescent Res Rev 2020 Jan 02;6(1):57-73. [doi: 10.1007/s40894-019-00130-0]
9. Volkow ND, Baler RD, Compton WM, Weiss SRB. Adverse health effects of marijuana use. N Engl J Med 2014 Jun 05;370(23):2219-2227 [FREE Full text] [doi: 10.1056/NEJMra1402309] [Medline: 24897085]
10. National Academies of Sciences E. Understanding stigma of mental and substance use disorders. In: Ending Discrimination Against People with Mental and Substance Use Disorders: The Evidence for Stigma Change. Washington, DC: National Academies Press; 2016.
11. Bunting AM, Frank D, Arshonsky J, Bragg MA, Friedman SR, Krawczyk N. Socially-supportive norms and mutual aid of people who use opioids: an analysis of Reddit during the initial COVID-19 pandemic. Drug Alcohol Depend 2021 May 01;222:108672 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108672] [Medline: 33757708]
12. Gauthier RP, Costello MJ, Wallace JR. "I Will Not Drink With You Today": a topic-guided thematic analysis of addiction recovery on reddit. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29 to May 5, 2022; New Orleans, LA p. 1-17.
13. Alambo A, Padhee S, Banerjee T, Thirunarayan K. COVID-19 and mental health/substance use disorders on Reddit: a longitudinal study. 2021 Presented at: Pattern Recognition. ICPR International Workshops and Challenges (ICPR 2021); January 10-15, 2021; Virtual event. [doi: 10.1007/978-3-030-68790-8_2]
14. Wanchoo K, Abrams M, Merchant RM, Ungar L, Guntuku SC. Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19. PLoS One 2023;18(2):e0280337 [FREE Full text] [doi: 10.1371/journal.pone.0280337] [Medline: 36735708]
15. Yang G, King SG, Lin HM, Goldstein RZ. Emotional expression on social media support forums for substance cessation: observational study of text-based Reddit posts. J Med Internet Res 2023 Jul 19;25:e45267 [FREE Full text] [doi: 10.2196/45267] [Medline: 37467010]
16. Jha D, Singh R. Analysis of associations between emotions and activities of drug users and their addiction recovery tendencies from social media posts using structural equation modeling. BMC Bioinformatics 2020 Dec 30;21(Suppl 18):554 [FREE Full text] [doi: 10.1186/s12859-020-03893-9] [Medline: 33375934]
17. Rather Y, Bashir W, Sheikh A, Amin M, Zahgeer Y. Socio-demographic and clinical profile of substance abusers attending a regional drug de-addiction centre in chronic conflict area: Kashmir, India. Malays J Med Sci': MJMS 2013;20(3):31-38 [FREE Full text]
18. Rhidenour KB, Blackburn K, Barrett AK, Taylor S. Mediating medical marijuana: exploring how veterans discuss their stigmatized substance use on Reddit. Health Commun 2022 Sep;37(10):1305-1315. [doi: 10.1080/10410236.2021.1886411] [Medline: 33602000]
19. Scurti P, Nunzi M, Leonardi C, Pierlorenzi C, Marenzi R, Lamartora V. The experience of buprenorphine implant in patients with opioid use disorder: a series of narrative interviews. Front Psychiatry 2023;14:1205285 [FREE Full text] [doi: 10.3389/fpsyt.2023.1205285] [Medline: 37720906]
20. Stevens RC, Brawner BM, Kranzler E, Giorgi S, Lazarus E, Abera M, et al. Exploring substance use tweets of youth in the United States: mixed methods study. JMIR Public Health Surveill 2020 Mar 26;6(1):e16191 [FREE Full text] [doi: 10.2196/16191] [Medline: 32213472]
21. McGinty EE, Kennedy-Hendricks A, Baller J, Niederdeppe J, Gollust S, Barry CL. Criminal activity or treatable health condition? News media framing of opioid analgesic abuse in the United States, 1998-2012. Psychiatr Serv 2016 Apr 01;67(4):405-411. [doi: 10.1176/appi.ps.201500065] [Medline: 26620290]
22. Benford RD, Snow DA. Framing processes and social movements: an overview and assessment. Annu. Rev. Sociol 2000 Aug;26(1):611-639. [doi: 10.1146/annurev.soc.26.1.611]
23. Pavlova A, Berkers P. "Mental Health" as defined by Twitter: frames, emotions, stigma. Health Commun 2022 May;37(5):637-647 [FREE Full text] [doi: 10.1080/10410236.2020.1862396] [Medline: 33356604]
24. r/teenagers. Reddit. URL: https://www.reddit.com/r/teenagers/ [accessed 2025-01-28]

XSL•FO
RenderX

25. Goffman E. Frame Analysis: An Essay on the Organization of Experience. Cambridge, MA: Harvard University Press; 1974.
26. Zhao X, Nan X. Influence of self-affirmation on responses to gain- versus loss-framed antismoking messages. Hum Commun Res 2010;36(4):493-511.
27. Mendelsohn J. How framing influences public opinion in Twitter and the real world: a case study of immigration using NLP. GitHub. 2019. URL: https://juliamendelsohn.github.io/GRFP_Research_Statement_2019.pdf [accessed 2025-01-28]
28. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying Reddit: a systematic overview of disciplines, approaches, methods, and ethics. Social Media + Society 2021 May 26;7(2):205630512110190. [doi: 10.1177/20563051211019004]
29. Shakya HB, Christakis NA, Fowler JH. Parental influence on substance use in adolescent social networks. Arch Pediatr Adolesc Med 2012 Dec;166(12):1132-1139 [FREE Full text] [doi: 10.1001/archpediatrics.2012.1372] [Medline: 23045157]
30. Smith JM, Estefan A. Families parenting adolescents with substance abuse--recovering the mother's voice: a narrative literature review. J Fam Nurs 2014 Nov;20(4):415-441. [doi: 10.1177/1074840714554397] [Medline: 25342471]
31. Rothman EF, Linden JA, Baughman AL, Kaczmarsky C, Thompson M. "The Alcohol Just Pissed Me Off": views about how alcohol and marijuana influence adolescent dating violence perpetration, results of a qualitative study. Youth & Society 2013;48(3):366-382. [doi: 10.1177/0044118x13491973]
32. Bourduge C, Teissedre F, Morel F, Flaudias V, Izaute M, Brousse G. Lockdown impact on stress, coping strategies, and substance use in teenagers. Front Psychiatry 2021 Jan 21;12:790704. [doi: 10.3389/fpsyt.2021.790704] [Medline: 35126203]
33. MacLean SJ, Bruun A, Mallett S. 'I've had a pretty tough life but that's not why I do this': narratives of autonomy and control among alcohol and drug service-engaged early teenagers. Journal of Youth Studies 2012 Jul 24;16(2):206-221. [doi: 10.1080/13676261.2012.704990]
34. Lewallen J. When image isn't everything: the effects of Instagram frames on social comparison. The Journal of Social Media in Society 2016 Sep 30;5(2):108-133 [FREE Full text]
35. Gross K. Framing persuasive appeals: episodic and thematic framing, emotional response, and policy opinion. Political Psychology 2008 Feb 29;29(2):169-192. [doi: 10.1111/j.1467-9221.2008.00622.x]
36. Major LH. Mental health news: how frames influence support for policy and civic engagement intentions. J Health Commun 2018;23(1):52-60. [doi: 10.1080/10810730.2017.1411994] [Medline: 29252120]
37. Dingel MJ, Karkazis K, Koenig BA. Framing nicotine addiction as a "Disease of the Brain": social and ethical consequences. Soc Sci Q 2012 Dec 01;92(5):1363-1388 [FREE Full text] [doi: 10.1111/j.1540-6237.2011.00822.x] [Medline: 23476081]
38. Sanchez ZM, Folgar MI, Matias JP, Pimentel MP, Burkhart G. Framing substance use as "recreational" is neither accurate nor helpful for prevention purposes. J Prev (2022) 2023 Dec;44(6):795-811. [doi: 10.1007/s10935-023-00745-z] [Medline: 37642907]
39. Reddit comments/submissions 2005-06 to 2022-12. Internet Archive. URL: https://archive.org/details/pushshift_reddit_200506_to_202212 [accessed 2024-02-22]
40. Rutherford BN, Lim CCW, Johnson B, Cheng B, Chung J, Huang S, et al. #TurntTrending: a systematic review of substance use portrayals on social media platforms. Addiction 2023 Feb;118(2):206-217 [FREE Full text] [doi: 10.1111/add.16020] [Medline: 36075258]
41. Prince Z, Jha D, Singh R. DUI: the drug use insights web server. Bioinformatics 2021 Dec 11;37(24):4895-4897. [doi: 10.1093/bioinformatics/btab461] [Medline: 34164647]
42. Liu Z, Yang K, Zhang T, Xie Q, Ananiadou S. EmoLLMs: a series of emotional large language models and annotation tools for comprehensive affective analysis. arXiv Preprint posted online on January 16, 2024 [FREE Full text] [doi: 10.1145/3637528.3671552]
43. Hawke LD, Barbic SP, Voineskos A, Szatmari P, Cleverley K, Hayes E, et al. Impacts of COVID-19 on youth mental health, substance use, and well-being: a rapid survey of clinical and community samples. Répercussions de la COVID-19 sur la santé mentale, l'utilisation de substances et le bien-être des adolescents : un sondage rapide d'échantillons cliniques et communautaires. Can J Psychiatry 2020 Oct;65(10):701-709 [FREE Full text] [doi: 10.1177/0706743720940562] [Medline: 32662303]
44. Marinelli S, Basile G, De Paola L, Napoletano G, Zaami S. Consequences of COVID-19 restrictions on adolescent mental health and drug abuse dynamics. Eur Rev Med Pharmacol Sci 2024 Jan;28(2):836-851 [FREE Full text] [doi: 10.26355/eurrev_202401_35085] [Medline: 38305627]
45. Rinaldi R, Bersani G, Marinelli E, Zaami S. The rise of new psychoactive substances and psychiatric implications: a wide-ranging, multifaceted challenge that needs far-reaching common legislative strategies. Hum Psychopharmacol 2020 May;35(3):e2727. [doi: 10.1002/hup.2727] [Medline: 32144953]
46. Meherali S, Punjani N, Louie-Poon S, Abdul Rahim K, Das JK, Salam RA, et al. Mental health of children and adolescents amidst COVID-19 and past pandemics: a rapid systematic review. Int J Environ Res Public Health 2021 Mar 26;18(7):3432 [FREE Full text] [doi: 10.3390/ijerph18073432] [Medline: 33810225]
47. Negro F, Di Trana A, Marinelli S. The effects of the COVID-19 pandemic on the use of the performance-enhancing drugs. Acta Biomed 2022 Jan 19;92(6):1-3 [FREE Full text] [doi: 10.23750/abm.v92i6.12377] [Medline: 35075058]
48. ChatGPT. OpenAI. URL: https://chatgpt.com/ [accessed 2025-02-18]

XSL•FO
RenderX

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
LGBTQ+: lesbian, gay, bisexual, trans, and queer
NLP: natural language processing
NPS: new psychoactive substance
RQ: research question
SUD: substance uses disorder

Edited by A Mavragani; submitted 09.04.24; peer-reviewed by F Lamy, S Zaami, E Eschliman; comments to author 11.09.24; revised version received 22.11.24; accepted 16.12.24; published 19.02.25.

Please cite as:
Zhang X, Zhu J, Kenne DR, Jin R
Teenager Substance Use on Reddit: Mixed Methods Computational Analysis of Frames and Emotions
J Med Internet Res 2025;27:e59338
URL: https://www.jmir.org/2025/1/e59338
doi:10.2196/59338
PMID:

©Xinyu Zhang, Jianfeng Zhu, Deric R Kenne, Ruoming Jin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Engagement With and Use of Health Information on Social Media Among US Latino Individuals: National Cross-Sectional Survey Study

Yonaira M Rivera[1], MPH, PhD; Kathryna Corpuz[2], MPH; Tahilin Sanchez Karver[3], MPH, PhD

[1]Department of Communication, School of Communication and Information, Rutgers University, New Brunswick, NJ, United States

[2]Office of Clinical and Health Affairs, Rutgers Biomedical & Health Sciences, New Brunswick, NJ, United States

[3]Department of Health, Behavior & Society, Johns Hopkins School of Public Health, Baltimore, MD, United States

Corresponding Author:
Yonaira M Rivera, MPH, PhD
Department of Communication
School of Communication and Information
Rutgers University
4 Huntington St
New Brunswick, NJ, 08901
United States
Phone: 1 8489327612
Email: yonaira.rivera@rutgers.edu

Abstract

Background: During the COVID-19 pandemic, US Latino individuals were more likely to report accessing coronavirus information on social media than other groups, despite copious amounts of health misinformation documented on these platforms. Among the existing literature on factors associated with engagement and use of health information, racial minority status has been associated with greater susceptibility to health misinformation. However, literature to date has not reported national trends on how Latino individuals engage with or use health information on social media compared to non-Latino White (NLW) individuals, nor whether perceptions of the amount of health misinformation on social media influence health information engagement and usage.

Objective: This study aimed to examine differences in engagement with and use of health information on social media among Latino and NLW individuals in the United States.

Methods: We examined a nationally representative cross-sectional sample of Latino (n=827) and NLW (n=2563) respondents of the 2022 Health Information National Trends Survey who used social media in 2022 to assess differences in engagement with and use of health information. Items related to the perceived quantity of health misinformation on social media, social media use frequency, health information engagement (sharing content; watching videos), and health information usage (health decision-making; discussions with health care providers) were selected to conduct weighted bivariate analyses and logistic regressions.

Results: Latino individuals perceive lower amounts of health misinformation on social media (28.9% perceived little to no misinformation vs 13.6% NLW individuals, $P<.001$). Latino audiences also reported higher health information engagement compared to NLW individuals (20% vs 10.2% shared information several times a month or more, $P<.001$; 42.4% vs 27.2% watched videos several times a month or more, $P<.001$), as well as higher information usage for health decision-making (22.8% vs 13.7%, $P=.003$). When controlling for ethnicity and other sociodemographic variables, perceiving lower amounts of health misinformation on social media was associated with higher odds of watching videos more frequently, making health decisions, and discussing health-related content with a health care provider ($P<.001$). Furthermore, Latino audiences were 1.85 times more likely to watch videos ($P<.001$), when controlling for the perceived amount of health misinformation and other sociodemographic variables. Finally, when compared to NLW individuals perceiving little to no health misinformation, Latino audiences perceiving little to no health misinformation were 2.91 times more likely to watch videos ($P<.001$).

Conclusions: The findings suggest that Latino individuals engage with visual health (mis)information at higher rates. Digital health literacy interventions should consider video formats and preferred social media platforms among Latino individuals. Further research is warranted to understand sociocultural factors important to Latino social media users when consuming health information, as these may impact the success of digital media literacy interventions that teach users how to navigate misinformation online.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e59387) doi:10.2196/59387

KEYWORDS

Latinos; health misinformation; engagement; utilization; social media; health information; United States; national; trends; survey; pandemic; non-Latino whites

Introduction

Background

Social media's ubiquitous presence in daily life has allowed for the democratization of health information and for individuals and communities to view, engage with, and share a variety of content with people in their networks. However, it has also contributed to the rise of health misinformation [1-4], which tends to evoke an emotional response and a reaction from its readers [5-7]. This drive to act can propel social media users to share viewed content with others without necessarily validating its credibility and integrate it into their own behaviors, including health-seeking behaviors. This can have a significant impact on their overall health outcomes [8], particularly in light of misinformation shared unintentionally to voice people's concerns, find answers to their questions, and understand those answers in the context of their own life experiences [9]. Understanding the ramifications of misinformation in health outcomes, thus, requires an interdisciplinary approach that intersects the fields of health communication, communication and technology, and public health education.

Social Media Usage and Health Information Consumption Among US Latino Individuals

Individuals identifying as Hispanic, Latino, Latina, or Latinx (henceforth Latino) are the largest and fastest growing minority group in the United States [10] and are also avid social media users [11]. In 2021, the Pew Research Center reported general social media usage rates among Latino adults in the United States, where 85% used YouTube, 72% used Facebook, 52% used Instagram, and 46% used WhatsApp [11]. YouTube (Google) has also been reported to be a top source for news content among Latino adults, particularly among those seeking content in Spanish: 48% of Latino households that predominantly speak Spanish rely on YouTube for news content compared to 36% among those that speak English [12]. Latino individuals also report a greater reliance on social media and messaging platforms than other groups, viewing platforms like WhatsApp, Instagram, and Facebook (Meta) as trustworthy tools to connect with families and friends in the United States and internationally [13]. This reliance on social media contributes to Latino individuals being at a higher risk for health misinformation exposure, as has been seen with topics like COVID-19, guns, and reproductive health matters including abortion [14]. For example, the National Association of Latino Elected and Appointed Officials' 2022 National Latino Voter tracking poll reported that 76% of Latino survey respondents were exposed to news that abortion is now illegal in the United States and that individuals can be imprisoned for seeking abortion services [15]. Furthermore, Latino individuals may engage with content in English, Spanish, or a combination of both [13], adding another challenging layer in efforts to understand how these audiences engage with and use health information found on these platforms. Previous literature has highlighted inequities in Spanish-speaking Latino individuals' ability to access and understand cancer information compared to English-speaking counterparts [16], yet studies exploring Spanish health content on social media are limited. This is particularly concerning given reports that platforms like Facebook have failed to flag misinformation in Spanish more so than misinformation in English, with one 2020 study reporting that an estimated 70% of Spanish-language misinformation on Facebook lacked warning labels [17].

Early during the COVID-19 pandemic, Latino individuals were reported to use social media to access COVID-19 information at higher rates than other groups in the United States [18]. However, given that social media platforms also contain user-generated content that is oftentimes not evidence-based, there is an increased likelihood for users to engage with and use health misinformation. For example, qualitative research exploring the Latino individuals' engagement with cancer information on Facebook pre-COVID found that participants mainly engaged with content from potentially unreliable sources predominantly shared by Facebook friends [19]. Some participants also reported using encountered cancer information to discuss with friends and family and to implement new or reinforce current health behaviors, despite this content not always being evidence-based or reliable [19]. Despite these findings, the literature to date has not reported national trends in how Latino individuals engage with health information on social media compared to non-Latino White (NLW) individuals, nor whether the information they have engaged with has triggered offline actions that can influence health outcomes. As such, our first research question (RQ) asks "Does engagement with and usage of health information encountered on social media differ between Latino and NLW social media users?"

Factors Associated With Health (Mis)Information Susceptibility and Their Role in Engagement and Usage

In an effort to better understand why some individuals may be more susceptible to health misinformation on social media, Nan and colleagues [20] conducted a systematic review of 61 studies to explore the psychological, demographic, and behavioral predictors of susceptibility. After presenting an integrative psychological model of susceptibility to health misinformation, authors found that psychological predictors included directionally motivated reasoning or actions driven by one's emotions; placing greater trust in information shared by family, friends, politicians, or journalists; conservative or political ideology; and religiosity. Greater social media use was generally observed to increase one's susceptibility to health misinformation, while demographic factors like age, education, sex at birth, and income showed mixed correlations to increased susceptibility to misinformation. The review also noted studies that suggest Latino individuals tend to be more susceptible

XSL•FO
RenderX

[21-23], although none of these explicitly focused on this ethnic group.

One area that has received less attention relates to whether an individual's perceptions of the amount of health misinformation that exists on social media has an impact on subsequent engagement and usage of that information. Chou et al [4] briefly bring attention to this issue in their 2020 commentary, "Where We Go From Here: Health Misinformation on Social Media," which calls for research identifying the characteristics that make certain populations more vulnerable to the impacts of health misinformation. More recently, Goldsmith et al [24] conducted a systematic review exploring the use of social media as a source of COVID-19 information among migrant and ethnic minority populations, which found that some populations may rely on health misinformation to shape their understanding and inform their actions. The authors specifically referenced a study by Cervantes et al [25] that had a sample of 60 US Latino adults who were hospitalized for COVID-19. Many individuals in the study discussed using social media for COVID-19 recommendations, despite some acknowledging that COVID-19 misinformation was circulating at the time [25]. Similarly, Rivera et al [26] reported cases of Latino interviewees not verifying cancer-related information from Facebook because "someone must have" previously done so. Combined, these qualitative findings raise the question of whether differences in the Latino individuals' perceptions about the amount of health misinformation on social media may impact engagement with and use of that content. It also begs the question of whether any such differences in perception impact engagement and usage outcomes differently than for NLW individuals. Therefore, we also ask the following:

RQ2. Do perceptions of the amount of health misinformation on social media differ between Latino and NLW social media users?

RQ3. Are perceptions of the amount of health misinformation on social media associated with engagement with and use of health information on social media upon controlling for social media usage and other sociodemographic variables?

RQ4. Does engagement with and use of encountered health information differ between Latino and NLW social media users who perceive the same amount of health misinformation on social media, upon controlling for social media usage and other sociodemographic variables?

Methods

Overview

This paper analyzes data from the 2022 Health Information National Trends Survey (referred to as HINTS 6). HINTS is a national cross-sectional survey conducted annually by the National Cancer Institute to assess access and use of health information by the general adult public in the United States. HINTS 6 covers topics related to looking for health information; finding information online; health care access and infrastructures; general health prevention; cancer screening and awareness; cancer beliefs; and cancer history [27]. HINTS 6 was conducted via a self-administered web or paper survey from March to November 2022. An equal-probability sample of addresses provided by Marketing Systems Group was selected from within four sampling strata (urban and rural areas with high vs low concentrations of minority adults) [28]. A total of 6252 individuals participated in the survey (response rate=28.1%). Full sample selection details are available in the HINTS 6 Methodology Report [28]. In order to explore engagement with and use of health information encountered on social media among individuals identified as Latino or Hispanic compared to NLW individuals, the analyses in this paper focus on participants who identified their race/ethnicity in the survey and reported using social media (n=3390). Acknowledging the heterogeneity that exists in Latino communities in the United States, it is important to delineate that HINTS 6 defines Latino or Hispanic participants as anyone who answers "yes" to the question, "Are you of Hispanic, Latino/a, or Spanish origin?" This includes anyone living in the United States who identifies as being part of any Latino subethnicity, regardless of racial background.

Measures

Dependent Variables

Engagement With Health-Related Information

Two items were selected to measure engagement with health-related information on social media: frequency of sharing general health–related information on social media and frequency of watching health-related videos on a social media site. These outcomes were assessed by survey items asking participants: "In the past 12 months, how often did you share general health-related information on social media (eg, a news article)?" and "In the past 12 months, how often did you watch a health-related video on a social media site (eg, YouTube)?" Response options for both items included 1=almost every day, 2=at least once a week, 3=a few times a month, 4=less than once a month, and 5=never. Exploratory data analyses reported skewed findings (mean scores for sharing information 4.48, SD 0.82, and mean scores for watching videos 3.93, SD 1.05). Therefore, final binary outcome variables were created by recoding each response option as 0=never to less than once a month and 1=a few times a month or more.

Use of Health-Related Information

Two items were selected to measure the use of encountered health-related information on social media: using information encountered on social media to make health decisions and discussing information encountered on social media with a health care provider. These outcomes were assessed by survey items asking participants how much they agreed or disagreed with the following statements: "I use information from social media to make decisions about my health" and "I use information from social media in discussions with my health care provider." Response options for both items included 1=strongly agree, 2=somewhat agree, 3=somewhat disagree, and 4=strongly disagree. Exploratory data analyses reported skewed findings (mean scores for making decisions 3.49, SD 0.76, and mean scores for discussing with health providers 3.38, SD 0.84). Therefore, final binary outcome variables were created by recoding each response option as 0=disagree and 1=agree.

XSL•FO
RenderX

Independent Variable: Perception of the Quantity of Misinformation on Social Media

The main factor associated with engagement with and usage of social media health information was the perception of the amount of health misinformation on social media. To explore differences in perception between Latino and NLW individuals, we first identified all participants who responded to the question "How much of the health information that you see on social do you think is false or misleading?". Response options included 1=a lot, 2=some, 3=a little, 4=none, and 5=I don't use social media; any participant who stated they did not use social media was dropped from the analysis. The remaining responses were recoded as 0=a little or none, 1=some, and 2=a lot.

To further assess whether differences in perceptions of the amount of health misinformation on social media between Latino participants and NLW participants were associated with engagement with and usage of social media health information, responses were divided by ethnicity (Latino vs NLW individuals). To do so, a final variable was created with the following six groups: 0=NLW individuals perceiving little to no health misinformation on social media, 1=NLW individuals perceiving some health misinformation on social media, 2=NLW individuals perceiving a lot of misinformation on social media, 3=Latino individuals perceiving little to no health misinformation on social media, 4=Latino individuals perceiving some health misinformation on social media, and 5=Latino individuals perceiving a lot of misinformation on social media.

Other covariates included categorical variables for age, education, sex at birth, and social media usage frequency (refer to Table 1 for category options). Social media use frequency was assessed by an item asking, "In the past 12 months, how often did you visit a social media site?" Responses were recorded as 0=never, 1=less than once a month, 2=a few times a month, 3=at least once a week, and 4=almost every day.

XSL•FO
RenderX

Table 1. Characteristics of study sample, Latino/a individuals and non-Latino/a White individuals, and weighted percentages (n=3390). All values presented, except for n, are weighted.

Variables	Latino/a, Unweighted (n=827), Weighted %	NLW[a], Unweighted (n=2563), Weighted %	*P* value[b]
Sex at birth, n (%)			.74
Male	287 (47.2)	1013 (48.5)	
Female	533 (52.8)	1545 (51.5)	
Age (year), n (%)			<.001
18-34	218 (40)	365 (24.2)	
35-49	249 (31.6)	533 (25.7)	
50-64	214 (20.3)	772 (30.2)	
65-74	111 (6.3)	572 (12.9)	
75+	29 (1.8)	309 (6.9)	
Education, n (%)			<.001
Less than high school	94 (13.9)	77 (4.3)	
High school	178 (27.8)	363 (19)	
Some college	259 (38.5)	681 (37.3)	
College graduate or more	294 (19.9)	1436 (39.4)	
Household income, n (%)			<.001
Less than US $20,000	174 (17.8)	239 (10.7)	
US $20,000 to less than US $35,000	138 (16.5)	267 (9.3)	
US $35,000 to less than US $50,000	117 (12.4)	278 (9.9)	
US $50,000 to less than US $75,000	131 (18.7)	432 (17)	
US $75,000 or more	237 (34.6)	1227 (53)	
Social media use frequency, n (%)			.04
Almost daily	567 (75.5)	1681 (67.8)	
At least once a week	100 (8.4)	358 (13.5)	
A few times a month	65 (8.3)	207 (8.3)	
Less than once a month	38 (3.7)	134 (4.2)	
Never	50 (4)	172 (6.3)	
Perceived amount of health misinformation on social media, n (%)			<.001
A little or none	223 (28.9)	369 (13.6)	
Some	352 (43.3)	1152 (46)	
A lot	252 (27.8)	1042 (40.5)	

[a]NLW: non-Latino White.

[b]Based on Pearson chi-square statistic.

Data Analysis

Data were analyzed using SPSS (version 28; IBM Corp). Any missing variable responses were dropped from the corresponding analyses. Sampling weights provided in the dataset were applied using Taylor Series for variance estimation [29]. Frequencies were calculated by comparing all independent variables by ethnic group (Latino: n=827; NLW: n=2563) using chi-square tests. Bivariate analyses were conducted comparing the effects of perceived amount of health misinformation on social media, ethnic group, age, sex at birth, and social media use frequency in the past 12 months on all four outcomes. This was followed by multivariate logistic regressions to assess whether the perceived amount of health misinformation on social media and ethnic groups was associated with each outcome. Only statistically significant covariates were included in logistic regressions. Adjusted odds ratios, confidence intervals, and their *P* values for the independent variables were reported, and the significance level of all analyses was set at α=.05.

Results

Participants in this sample were predominantly NLW individuals (75.6%, n=2563), female (61.5%, n=2078), 35-64 years old

XSL•FO
RenderX

(52.4%, n=1768), had a college degree or higher (51.2%, n=1730), and had incomes of US $50,000 or more (62.6%, n=2027). Weighted demographics are listed in Table 1. Compared to NLW participants, Latino participants were more likely to be younger (*P*<.001), less educated (*P*<.001), and have lower incomes (*P*<.001).

Perceptions of the Amount of Health Misinformation on Social Media Among Latino Audiences Compared to NLW Audiences

Table 1 also reports how participants use and perceive the social media environment. There were statistically significant differences in social media use, with the Latino individuals reporting more almost daily social media use than NLW individuals (75.5% vs 67.8%; *P*=.04). Latino individuals were also more likely to perceive there was little to no false or misleading health information on social media compared to NLW individuals (28.9% vs 13.6%; *P*<.001).

Differences in Engagement With and Use of Health Information Encountered on Social Media Between Latino and NLW Individuals

Overall, Latino individuals reported higher levels of engagement with health information on social media in the last 12 months compared to NLW individuals (Table 2). Compared to 10.2% of NLW individuals, 20% of Latino individuals shared general health information on social media several times a month or more (*P*<.001). They also watched health videos on social media several times a month or more at higher amounts than NLW individuals (42.4% vs 27.2%, *P*<.001). Upon controlling for the perceived amount of health misinformation, frequency of social media usage, age, sex at birth, and education (Table 3), Latino individuals were 1.85 times more likely to watch health videos several times a month or more than NLW individuals (CI 1.38-2.47). They were also 1.65 times more likely than NLW individuals to share general health information on social media several times a month or more; these findings were marginally significant (CI 0.97-2.80).

Table 2. Reported engagement with and usage of health information on social media by perceptions of the amount of health misinformation on social media, Latino/a versus non-Latino/a White (NLW) individuals, weighted percentages. All values presented, except for n, are weighted.

Outcome variables	Latino/a, Unweighted n (Weighted %)	NLW, Unweighted n (Weighted %)	*P* value[a]
Engagement			
Shared general health information on social media several times a month or more	137 (20)	248 (10.2)	<.001
Watched health-related videos on social media several times a month or more	330 (42.4)	658 (27.2)	<.001
Usage			
Used information to make a health decision	160 (22.8)	335 (13.7)	.003
Discussed information with health care provider	189 (22)	479 (18.4)	.12

[a]Based on Pearson chi-square statistic.

Latino individuals also reported higher levels of use of health information encountered on social media when compared to NLW individuals; specifically, 22.8% of Latino individuals reported using encountered health information for health decision-making, compared to 13.7% of NLW individuals (*P*=.003; Table 2). However, differences between Latino and NLW individuals were not statistically significant upon controlling for perceptions of the amount of health misinformation on social media, frequency of social media usage, age, sex at birth, and education (Table 4). No statistically significant differences in discussing health information encountered on social media with a health care provider were found.

XSL•FO
RenderX

Table 3. Weighted unadjusted and adjusted odds ratios of engagement with health information on social media by perceptions of the amount of health misinformation on social media, Latino/a individuals versus non-Latino/a White individuals. Weighted adjusted models include frequency of social media usage, age, sex at birth, and education.

	Shared		Watched	
	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Race or ethnicity				
NLW[b] (reference)				
Latino/a	2.21 (1.40-3.51)[c]	1.65 (0.97-2.80)[c]	1.97 (1.47-2.63)[b]	1.85 (1.38- 2.47)[b]
Perceived health misinformation				
A lot (reference)				
Some	1.57 (1- 2.48)[c]	1.04 (0.62- 1.77)	1.49 (1.15- 1.93)[c]	1.65 (1.26- 2.15)[c]
A little to none	2.14 (1.41- 3.24)[b]	1.43 (0.91- 2.24)	2.052 (1.47- 2.86)[c]	2.15 (1.53- 3.02)[c]

[a]OR: odds ratio.

[b]NLW: non-Latino White.

[c]*P*<.05.

Table 4. Weighted unadjusted and adjusted odds ratios of the use of health information on social media by perceptions of the amount of health misinformation on social media, Latino/a versus non-Latino/a White individuals. Weighted adjusted models include frequency of social media usage, age, sex at birth, and education.

	Decided		Discussed	
	Unadjusted OR[a] (95% CI)	Adjusted OR (95% CI)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Race and ethnicity				
NLW[b] (reference)				
Latino/a	1.87 (1.23- 2.82)[c]	1.41 (0.88- 2.25)	1.25 (0.94- 1.66)	1.10 (0.79- 1.54)
Perceived health misinformation				
A lot (reference)				
Some	3.26 (2.20- 4.83)[c]	4.15 (2.59- 6.65)[c]	1.96 (1.3- 2.93)[c]	2.34 (1.61- 3.40)[c]
A little to none	4.39 (2.59- 7.44)[c]	6.26 (3.95- 9.92)[c]	2.49 (1.67- 3.73)[c]	2.86 (1.97- 4.16)[c]

[a]OR: odds ratio.

[b]NLW: non-Latino White.

[c]*P*<.05.

Perceptions of Quantity of Health Misinformation on Social Media as a Factor Associated With Information Engagement and Use by Ethnicity

As seen in Tables 3 and 4, perceptions of the amount of health misinformation on social media were associated with engagement with and usage of encountered health information on social media in 3 of the 4 models, when controlling for ethnicity and other covariates. Specifically, individuals who perceived little to no health misinformation on social media were 2.15 times more likely to watch videos several times a month or more compared to individuals who believed there is a lot of health misinformation on social media (CI 1.53-3.02; Table 3). They were also 6.26 times more likely to use encountered information to make health decisions (CI 3.95-9.92) and 2.86 times more likely to discuss encountered health information with a health care provider (CI 1.97-4.16) than individuals who believed there is a lot of health misinformation on social media (Table 4).

Table 5 reports differences in engagement and usage between the Latino and NLW individuals by perceived amount of health misinformation to assess differences in these outcomes between ethnic groups who perceive the same amount of health misinformation on social media. Compared to NLW individuals perceiving little to no health misinformation on social media, Latino individuals perceiving little to no health misinformation were 2.91 times more likely to watch videos (CI 1.69-5.03), controlling for social media use frequency, age, sex at birth, and education. They also reported marginally significant higher odds of using encountered information to make health decisions when compared to NLW individuals perceiving little to no health misinformation (adjusted odds ratio 1.87, CI 0.96-3.63).

XSL•FO
RenderX

Table 5. Weighted adjusted odds ratios of perceived amount of health misinformation and ethnicity (Latino/a vs non-Latino/a White individuals) on engagement and usage of health information. Weighted adjusted models include frequency of social media usage, age, sex at birth, and education.

Perceived amount of health misinformation on social media	Engagement		Usage	
	Shared, Adjusted OR[a] (95% CI)	Watched, Adjusted OR (95% CI)	Decided, Adjusted OR (95% CI)	Discussed, Adjusted OR (95% CI)
Models 1-4[b]				
Latino audiences perceiving little to no health misinformation	1.67 (0.82-3.40)	2.91 (1.69- 5.03)[c]	1.87 (0.96- 3.63)	0.88 (0.46- 1.66)
Models 5-8[b]				
Latino audiences perceiving some health misinformation	1.51 (0.66- 3.46)	1.87 (1.22- 2.88)[c]	1.31 (0.66- 2.60)	1.12 (0.69- 1.82)
Models 9-12[b]				
Latino audiences perceiving a lot of health misinformation	1.84 (0.81- 4.17)	1.24 (0.72- 2.12)	0.92 (0.44- 1.91)	1.44 (0.78- 2.66)

[a]OR: odds ratio.

[b]Reference group for each is non-Latino White individuals with the same perceived amount of health misinformation on social media.

[c]$P<.05$.

When comparing Latino audiences to NLW audiences perceiving some health misinformation on social media, there were only statistically significant differences in the amount of health videos watched, with Latino audiences being 1.87 times more likely to watch videos on social media than NLW audiences (CI 1.87-2.88). No differences in engagement and usage were found when comparing Latino audiences to NLW audiences who perceived there is a lot of health misinformation on social media.

Discussion

Principal Findings

Multiple calls to action in the disciplines of public health and health communication have detailed the importance of honing in on subsets of communities that are most vulnerable to the impacts of health misinformation, so as to better target resources for intervention [4,30]. This work provides insights to the engagement and usage patterns of social media users who perceive little to no health misinformation exists on social media, despite evidence to the contrary [1-3]. Our findings suggest that perceptions of the amount of health misinformation may be an important factor of engagement and usage. More importantly, Latino populations appear to have lower perceptions of the amount of health information on social media compared to NLW populations. In addition, Latino individuals who actively seek health advice on social media have been reported to be more likely to hear or believe at least one false health claim [14]. Coupled with higher social media usage patterns [11,13], these trends may place members of the Latino population at a higher disadvantage when navigating the media landscape. It is important that future work focus on digital health literacy tailored to these subgroups.

Comparisons With Previous Work

A growing body of literature has specifically highlighted concerns about the amount of health misinformation the Latino population may be exposed to on social media through visual formats. Rivera and colleagues [19,26] found evidence of Latino individuals both engaging with cancer misinformation encountered on Facebook in predominantly visual formats (ie, videos and images) and using it to make health decisions, despite most of the content not being evidence-based. Similarly, COVID-19 vaccine narratives in video format were highlighted as being easily accessible and shareable among Latino audiences, particularly those available in Spanish and shared through closed network platforms like WhatsApp [31]. These studies are consistent with findings that the Latino population has higher odds of watching health-related videos multiple times a month. In fact, Latino audiences spend more time streaming video content than other groups in the United states, with YouTube being the most used streaming platform [32]. Given recent evidence of YouTube's poor quality of health-related content [33,34], future endeavors should target multilingual visual forms of health misinformation.

According to the Kaiser Family Foundation, almost half of Latino adults in the United States report using social media to seek health advice and information at least once a week [14]. In light of this, two additional findings merit further discussion despite their marginal significance. First, we found that Latino audiences were more likely to share general health information with others on social media than NLW audiences. Latino individuals have been reported higher trust in media shared and recommended by their friends and family than from other sources [31], signaling a potential point of intervention in efforts to minimize the dissemination of health misinformation and its downstream impacts. We also found that Latino audiences who perceived that little to no health misinformation exists on social media were more likely to make decisions with health-related content found on these platforms. This merits further exploration, as Rivera et al [19] also identified extreme cases where individuals reported potentially harmful behaviors (such as canceling a mammogram) in response to breast cancer misinformation videos shared via Facebook. Their qualitative work illuminated the role of interpersonal relationships in

XSL•FO
RenderX

enhancing (and sometimes superseding) engagement with culturally relevant cancer content that came from potentially unreliable online sources, highlighting the interrelated ways in which cultural values and message and source factors contribute to engagement with social media health information [19]. As a budding area of inquiry, future research should identify important sociocultural factors and critical thinking skills that may predict engagement with content on social media among individuals with low digital health literacy, so as to better design interventions that curtail its negative impacts.

Limitations

Several limitations to this work must be acknowledged. First, there is no way to ascertain how often surveyed individuals actually encountered health misinformation on social media. HINTS 6 items only ask about overall social media health information engagement and usage, which may include different rates of exposure among different populations [35]. Similarly, perceptions of misinformation were only assessed by one item and relied on personal interpretations of what misinformation is—a term that may be unfamiliar to some or misrecognized by others. This study also failed to control for regional differences, which may impact trends. Finally, given the cross-sectional nature of the dataset, we cannot make causal inferences related to the predictive nature of the amount of health misinformation individuals perceive exists on social media and how they engage with and use this information. It may be that engagement and usage impact perceptions of the amount of health misinformation on social media, as has been suggested by Stimpson and Ortega [36]. Future studies should design surveys to better assess causality.

Despite initial attempts to explore differences in engagement and usage by Latino subethnic groups, sample size limitations did not allow for these comparisons. HINTS 6 reports Latino population subethnicities as Mexican, Puerto Rican, Cuban, and Other. While a preliminary look at these subgroups suggests that Puerto Ricans living in the continental United States have engagement and usage rates similar to NLW individuals, a larger sample is needed to explore these factors. Similarly, the HINTS 6 dataset no longer includes items asking language preferences, making it difficult to explore the role of language as an antecedent to health information engagement and use of social media. While it does include an item for the selected survey language, bilingual individuals who prefer Spanish are missed. Finally, while HINTS 6 collects information about social media use generally, there are no items that identify the specific social media platforms used by respondents. Given Latino audiences' reliance on platforms that are less popular to other US populations (ie, WhatsApp) [37], future iterations of HINTS may consider including this measure in their survey.

Finally, it is important to note that, as has been delineated by scholars elsewhere [16,38], the HINTS dataset was not designed to examine differences within Latino subgroups, nor does it include questions that may provide important sociocultural factors at play within this heterogeneous community. These may include cultural values like "familismo," "personalismo," "respeto," and "simpatía"–values oftentimes associated with Latin American collectivist identities [39]. As researchers centering their work on the Latino community, we have discussed some of these complexities elsewhere [19,40] and believe future work should explore the impact these may have on social media engagement. Furthermore, there may be other mechanisms impacting engagement with online information such as trust [38], patient engagement [41], and media literacy [42] that should be explored in future studies. Nonetheless, the findings reported in this paper lay a foundation for this important line of inquiry.

Conclusions

Our findings present new considerations for social media interventions to adequately deliver evidence-based health information to Latino populations. It is evident that video formats may better reach Latino audiences, despite the poor validity of many health-related videos on platforms like YouTube [33]. This raises the importance of building the digital health literacy of individuals and populations to improve their perception of whether a specific video or other modality of health information is actually valid and coming from a trustworthy source that can be used for health decision-making. Interventions should also consider multilingual digital media literacy training tailored to the specific social media platforms used by Latino populations, teaching them how to navigate each platform's unique features and how to use these to better assess content that may be false. This is particularly important in light of recent changes in social media platforms that call for the cancelation of fact-checking services and replacing these with user-generated fact-checking labels [43,44].

Acknowledgments

YMR and KC contributed to the conceptualization. Methodology was designed by YMR and KC. Formal analysis was conducted by YMR and TSK; YMR and KC assisted with writing–original draft. YMR, KC, and TSK contributed to writing–review & editing.

Data Availability

The datasets analyzed during this study are publicly available at the National Cancer Institute's Health Information National Trends survey website.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
2. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Tsagkaris C, Parvanov ED, et al. Medical and health-related misinformation on social media: bibliometric study of the scientific literature. J Med Internet Res 2022;24(1):e28152 [FREE Full text] [doi: 10.2196/28152] [Medline: 34951864]
3. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
4. Sylvia Chou WY, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. Am J Public Health 2020;110(S3):S273-S275. [doi: 10.2105/AJPH.2020.305905] [Medline: 33001722]
5. Karanian JM, Rabb N, Wulff AN, Torrance MG, Thomas AK, Race E. Protecting memory from misinformation: warnings modulate cortical reinstatement during memory retrieval. Proc Natl Acad Sci U S A 2020;117(37):22771-22779 [FREE Full text] [doi: 10.1073/pnas.2008595117] [Medline: 32868423]
6. Acerbi A. Cognitive attraction and online misinformation. Palgrave Commun 2019;5(1). [doi: 10.1057/s41599-019-0224-y]
7. Freiling I, Krause NM, Scheufele DA, Brossard D. Believing and sharing misinformation, fact-checks, and accurate information on social media: the role of anxiety during COVID-19. New Media Soc 2023;25(1):141-162. [doi: 10.1177/14614448211011451] [Medline: 36620434]
8. Viswanath K, Ramanadhan S, Kontos EZ. Mass media. In: Macrosocial Determinants of Population Health. New York, NY: Springer; 2007:275-294.
9. Jost JT, van der Linden S, Panagopoulos C, Hardin CD. Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. Curr Opin Psychol 2018;23:77-83. [doi: 10.1016/j.copsyc.2018.01.003] [Medline: 29427900]
10. Passel JS, Lopez MH, Cohn D. U.S. Hispanic population continued its geographic spread in the 2010. Pew Research Center. 2022. URL: https://hispanicad.com/news/u-s-hispanic-population-continued-its-geographic-spread-in-the-2010s/ [accessed 2024-11-30]
11. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2024-11-30]
12. Sanchez GR, Bennett C. Why Spanish-language mis- and disinformation is huge issue in 2022. Brookings. URL: https://www.brookings.edu/articles/why-spanish-language-mis-and-disinformation-is-a-huge-issue-in-2022/ [accessed 2024-11-30]
13. Inclusion, information, and intersection. Nielsen. 2021. URL: https://www.nielsen.com/insights/2021/inclusion-information-and-intersection/ [accessed 2023-10-06]
14. Kearney A, Lopes L, Washington I, Valdes I, Yilma H, Hamel L. Addressing misinformation among Hispanic adults: snapshot from the KFF health misinformation tracking poll pilot. Kaiser Family Foundation. URL: https://www.kff.org/health-information-and-trust/poll-finding/addressing-misinformation-among-hispanic-adults-snapshot-from-the-kff-health-misinformation-tracking-poll-pilot/ [accessed 2024-01-18]
15. 2022 National Latino voter tracking poll: week five results. NALEO Educational Fund. 2022. URL: https://naleo.org/wp-content/uploads/2022/10/10_11_22_-_NEF_Release_-_Tracking_Poll_-_Week_5_FINAL.pdf [accessed 2024-12-01]
16. Clayman ML, Manganello JA, Viswanath K, Hesse BW, Arora NK. Providing health messages to Hispanics/Latinos: understanding the importance of language, trust in health information sources, and media use. J Health Commun 2010;15 Suppl 3(Suppl 3):252-263 [FREE Full text] [doi: 10.1080/10810730.2010.522697] [Medline: 21154097]
17. How Facebook can flatten the curve of the Coronavirus Infodemic. Avaaz. 2020. URL: https://secure.avaaz.org/campaign/en/facebook_coronavirus_misinformation/ [accessed 2024-12-01]
18. Cultural connectivity transformed. Nielsen. 2020. URL: https://www.nielsen.com/insights/2020/cultural-connectivity-transformed/ [accessed 2024-01-18]
19. Rivera YM, Moran MB, Thrul J, Joshu C, Smith KC. When engagement leads to action: understanding the impact of cancer (mis)information among Latino/a Facebook users. Health Commun 2022;37(9):1229-1241 [FREE Full text] [doi: 10.1080/10410236.2021.1950442] [Medline: 34256676]
20. Nan X, Wang Y, Thier K. Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. Soc Sci Med 2022;314:115398. [doi: 10.1016/j.socscimed.2022.115398] [Medline: 36327631]
21. Brian Houston J, Thorson E, Kim E, Mantrala MK. COVID-19 communication ecology: visualizing communication resource connections during a public health emergency using network analysis. Am Behav Sci 2021;65(7):893-913. [doi: 10.1177/0002764221992811]
22. Lunz Trujillo K, Motta M, Callaghan T, Sylvester S. Correcting misperceptions about the MMR vaccine: using psychological risk factors to inform targeted communication strategies. Political Research Quarterly 2020;74(2):464-478. [doi: 10.1177/1065912920907695]

XSL•FO
RenderX

23. Motta M, Callaghan T. The pervasiveness and policy consequences of medical folk wisdom in the U.S. Sci Rep 2020;10(1):10722 [FREE Full text] [doi: 10.1038/s41598-020-67744-6] [Medline: 32612260]
24. Goldsmith LP, Rowland-Pomp M, Hanson K, Deal A, Crawshaw AF, Hayward SE, et al. Use of social media platforms by migrant and ethnic minority populations during the COVID-19 pandemic: a systematic review. BMJ Open 2022;12(11):e061896 [FREE Full text] [doi: 10.1136/bmjopen-2022-061896] [Medline: 36396309]
25. Cervantes L, Martin M, Frank MG, Farfan JF, Kearns M, Rubio LA, et al. Experiences of latinx individuals hospitalized for COVID-19: a qualitative study. JAMA Netw Open 2021;4(3):e210684 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0684] [Medline: 33704475]
26. Rivera YM, Smith KC, Moran MB. "Why would someone intentionally lie?": assessing the credibility of cancer (mis)information on Facebook. AoIR Selected Papers of Internet Research. 2019. URL: https://spir.aoir.org/ojs/index.php/spir/article/view/11026 [accessed 2025-02-06]
27. Health Information National Trends Survey (HINTS) 6 Public Codebook. National Cancer Institute. URL: https://hints.cancer.gov/ [accessed 2023-05-16]
28. Health Information National Trends Survey 6 (HINTS 6): HINTS 6 Methodology Report. National Cancer Institute. 2006. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS_6_MethodologyReport.pdf [accessed 2023-05-16]
29. Overview of the HINTS 6 Survey (2022) and Data Analysis Recommendations. National Cancer Institute. URL: https://hints.cancer.gov/ [accessed 2023-05-16]
30. National Academies of Sciences National Academies of Sciences, Engineering, and Medicine. Addressing Health Misinformation with Health Literacy Strategies: Proceedings of a Workshop-in Brief. Washington, DC: National Academies Press; 2020.
31. Longoria J, Acosta D, Urbani S, Smith R. A limiting lens: how vaccine misinformation has influenced hispanic conversations online. MediaWell. URL: https://mediawell.ssrc.org/news-items/a-limiting-lens-how-vaccine-misinformation-has-influenced-hispanic-conversations-online-first-draft/ [accessed 2024-01-18]
32. Hispanic representation. Nielsen. URL: https://www.nielsen.com/about-us/business-inclusion-impact-belonging/hispanic/ [accessed 2024-01-18]
33. Haslam K, Doucette H, Hachey S, MacCallum T, Zwicker D, Smith-Brilliant M, et al. YouTube videos as health decision aids for the public: an integrative review. Can J Dent Hyg 2019;53(1):53-66 [FREE Full text] [Medline: 33240342]
34. Gruzd A, Abul-Fottouh D, Song MY, Saiphoo A. From facebook to youTube: the potential exposure to COVID-19 anti-vaccine videos on social media. Soc Media Soc 2023;9(1):20563051221150403 [FREE Full text] [doi: 10.1177/20563051221150403] [Medline: 36751363]
35. Borges do Nascimento IJ, Pizarro AB, Almeida JM, Azzopardi-Muscat N, Gonçalves MA, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022;100(9):544-561 [FREE Full text] [doi: 10.2471/BLT.21.287654] [Medline: 36062247]
36. Stimpson JP, Ortega AN. Social media users' perceptions about health mis- and disinformation on social media. Health Aff Sch 2023;1(4):qxad050 [FREE Full text] [doi: 10.1093/haschl/qxad050] [Medline: 38107206]
37. Scherer LD, McPhetres J, Pennycook G, Kempe A, Allen LA, Knoepke CE, et al. Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. Health Psychol 2021;40(4):274-284. [doi: 10.1037/hea0000978] [Medline: 33646806]
38. Camacho-Rivera M, Gonzalez CJ, Morency JA, Blake KD, Calixte R. Heterogeneity in trust of cancer information among hispanic adults in the United States: an analysis of the health information national trends survey. Cancer Epidemiol Biomarkers Prev 2020;29(7):1348-1356. [doi: 10.1158/1055-9965.EPI-19-1375] [Medline: 32611616]
39. Magaña D. Local voices on health care communication issues and insights on latino cultural constructs. Hispanic Journal of Behavioral Sciences 2020;42(3):300-323. [doi: 10.1177/0739986320927387]
40. Rivera YM. Reducing cancer health disparities among U.S. latinos: a freireian approach. Int J Hum Rights Healthc 2018;11(5):368-379 [FREE Full text] [doi: 10.1108/IJHRH-02-2018-0021] [Medline: 31032065]
41. Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
42. Roschke K, Koskan AM, Sivanandam S, Irby J. Partisan media, trust, and media literacy: regression analysis of predictors of COVID-19 knowledge. JMIR Form Res 2024;8:e53904 [FREE Full text] [doi: 10.2196/53904] [Medline: 39047283]
43. Isaac M, Schleifer T. Meta to end fact-checking program in shift ahead of Trump term. The New York Times. URL: https://www.nytimes.com/2025/01/07/technology/meta-fact-checking-facebook.html [accessed 2025-01-08]
44. Conger K. Elon musk wants people on X to police election posts. It's not working well. The New York Times. URL: https://www.nytimes.com/2024/07/25/technology/elon-musk-x-community-notes-election.html [accessed 2025-01-08]

Abbreviations

HINTS: Health Information National Trends Survey
NLW: non-Latino White
RQ: research question

XSL•FO
RenderX

Edited by C Argyropoulos; submitted 09.07.24; peer-reviewed by A Garg, V Gruest, D Chavez-Yenter; comments to author 08.11.24; revised version received 09.01.25; accepted 19.01.25; published 24.02.25.

Please cite as:
Rivera YM, Corpuz K, Karver TS
Engagement With and Use of Health Information on Social Media Among US Latino Individuals: National Cross-Sectional Survey Study
J Med Internet Res 2025;27:e59387
URL: https://www.jmir.org/2025/1/e59387
doi:10.2196/59387
PMID:

©Yonaira M Rivera, Kathryna Corpuz, Tahilin Sanchez Karver. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of Demographic and Clinical Subgroups in Google Trends Data: Infodemiology Case Study on Asthma Hospitalizations

Diana Portela[1,2], MD; Alberto Freitas[1,2], PhD; Elísio Costa[2,3], PhD; Mattia Giovannini[4,5], MD; Jean Bousquet[6,7,8], PhD; João Almeida Fonseca[1,2], MD, PhD; Bernardo Sousa-Pinto[1,2], PhD

[1]Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal

[2]Center for Health Technologies and Services Research, Health Research Network, University of Porto, Porto, Portugal

[3]Research Unit on Applied Molecular Biosciences, Faculty of Pharmacy, University of Porto, Porto, Portugal

[4]Allergy Unit, Department of Pediatrics, Meyer Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Florence, Italy

[5]Department of Health Sciences, University of Florence, Florence, Italy

[6]Institute of Allergology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[7]Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Allergology and Immunology, Berlin, Germany

[8]MASK-air, Montpellier, France

Corresponding Author:
Bernardo Sousa-Pinto, PhD
Department of Community Medicine, Information and Health Decision Sciences
Faculty of Medicine
University of Porto
R. Dr. Plácido da Costa
Porto, 4200-450
Portugal
Phone: 351 22 551 3622
Email: bernardosousapinto@protonmail.com

Abstract

Background: Google Trends (GT) data have shown promising results as a complementary tool to classical surveillance approaches. However, GT data are not necessarily provided by a representative sample of patients and may be skewed toward demographic and clinical groups that are more likely to use the internet to search for their health.

Objective: In this study, we aimed to assess whether GT-based models perform differently in distinct population subgroups. To assess that, we analyzed a case study on asthma hospitalizations.

Methods: We analyzed all hospitalizations with a main diagnosis of asthma occurring in 3 different countries (Portugal, Spain, and Brazil) for a period of approximately 5 years (January 1, 2012-December 17, 2016). Data on web-based searches on common cold for the same countries and time period were retrieved from GT. We estimated the correlation between GT data and the weekly occurrence of asthma hospitalizations (considering separate asthma admissions data according to patients' age, sex, ethnicity, and presence of comorbidities). In addition, we built autoregressive models to forecast the weekly number of asthma hospitalizations (for the different aforementioned subgroups) for a period of 1 year (June 2015-June 2016) based on admissions and GT data from the 3 previous years.

Results: Overall, correlation coefficients between GT on the pseudo-influenza syndrome topic and asthma hospitalizations ranged between 0.33 (in Portugal for admissions with at least one Charlson comorbidity group) and 0.86 (for admissions in women and in White people in Brazil). In the 3 assessed countries, forecasted hospitalizations for 2015-2016 correlated more strongly with observed admissions of older versus younger individuals (Portugal: Spearman ρ=0.70 vs ρ=0.56; Spain: ρ=0.88 vs ρ=0.76; Brazil: ρ=0.83 vs ρ=0.82). In Portugal and Spain, forecasted hospitalizations had a stronger correlation with admissions occurring for women than men (Portugal: ρ=0.75 vs ρ=0.52; Spain: ρ=0.83 vs ρ=0.51). In Brazil, stronger correlations were observed for admissions of White than of Black or Brown individuals (ρ=0.92 vs ρ=0.87). In Portugal, stronger correlations were observed for admissions of individuals without any comorbidity compared with admissions of individuals with comorbidities (ρ=0.68 vs ρ=0.66).

XSL•FO
RenderX

Conclusions: We observed that the models based on GT data may perform differently in demographic and clinical subgroups of participants, possibly reflecting differences in the composition of internet users' health-seeking behaviors.

(J Med Internet Res 2025;27:e51804) doi:10.2196/51804

KEYWORDS

infodemiology; asthma; administrative databases; multimorbidity; co-morbidity; respiratory; pulmonary; Google Trends; correlation; hospitalization; admissions; autoregressive; information seeking; searching; searches; forecasting

Introduction

The assessment of internet users' behavior can be a valuable source of information regarding their specific interests, preferences, and perceptions pertaining to diverse health topics. Such an assessment not only enables the identification and exploration of emerging trends in health-related interests but also facilitates an understanding of the factors influencing health information seeking, dissemination, and consumption in the digital age [1,2]. In this context, there are different methodological approaches that can be used, including the assessment of the relative volume of searches on specific health topics and keywords (i.e, assessing what internet users seek) [3-7] or the assessment of content available online, including social media posts by internet users [8-10]. These approaches are part of a recent field of studies termed "infodemiology," which is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [1,11].

Infodemiology studies have been conducted to accomplish different goals [7,12,13]. For instance, Google Trends (GT) data, which measure the relative volume of searches on a specific topic or term, have shown promising results as a complementary tool to classical surveillance methods [6], in forecasting influenza spread and hospitalizations [14-16], for modelling COVID-19 spread [17,18], and for forecasting asthma admissions [19]. However, GT data are not necessarily provided by a representative sample of individuals within a certain country or region [20] but can rather preferentially reflect demographic or clinical groups that are more likely to use the internet for health-related inquiries. For instance, the online behavior of younger, more educated, or technologically literate individuals may be overrepresented in GT data. Moreover, health-related search behaviors can be influenced by a host of factors, including the severity and type of health conditions, the availability and quality of health information online, and individual health literacy levels. Therefore, it is possible to hypothesize that GT data should not be seen as a "one-size-fits-all" tool for health research since we do not know the clinical and demographic composition of the individuals searching for a specific health term or topic. As such, it is probable that there may be relevant differences from what is observed in the general population, with relevant implications for the performance and interpretability of GT-based models.

Therefore, in this study, we aimed to assess whether GT-based models can have a different performance when considering different population subgroups (according to their clinical and demographic characteristics). To achieve that goal, we assessed a case study of asthma hospitalizations. Specifically, we (1) assessed the correlation between GT data for the *common cold* and the number of hospitalizations for asthma considering admissions of subgroups of patients (according to their age, sex, ethnicity, and presence of comorbidities) and (2) compared the performance of models predicting asthma hospitalizations based on GT for these specific participant segments (according to their age, sex, ethnicity, and presence of comorbidities).

Methods

Study Design

This study adhered to the methodological framework proposed by Mavragani and Ochoa [21]. In a previous study by our team, we had (1) established a correlation between GT data related to common cold–related search terms and asthma hospitalizations and (2) evaluated whether GT data on the *common cold*, combined with data on admissions, could help forecast asthma hospitalizations. In this study, we applied the same methodology (correlations and forecast models) and used the same GT data but specifically considered those hospital admissions occurring in patients of each sex, age group (18-64 years old versus ≥65 years old), ethnicity (White versus Black or Brown ["pardo"]), and the presence or absence of at least one Charlson comorbidity [22]. We assessed a period of approximately 5 years (2012-2016), assessing data from Portugal, Spain, and Brazil.

Data Sources and Variables

GT Data

GT is a tool that offers insight into the popularity of search terms by providing their relative search volume data on a scale of 0 to 100 (where 100 represents the peak interest at a specific time and location). It allows users to compare the popularity of different keywords, topics, or queries across regions and time periods. The data are indexed to show the proportion of searches for a specific term relative to all searches on Google at that specific time and location [6,12,23].

We obtained GT data on January 13, 2020, as already described by Sousa-Pinto et al [19]. In brief, we retrieved country-level GT on rhinovirus-related search terms for 5 years (between 2012 and 2016) in Portugal, Spain, and Brazil: 5 years is the maximum amount of time for which GT displays data on a weekly level. These countries and timeframe were chosen (1) to allow comparability with a previous study [19] and (2) due to the accessibility of nationwide data regarding the frequency of weekly hospitalizations presented by age and sex. No specific categories or subcategories of GT data were selected. We accessed GT data exclusively through its web interface, with a

XSL•FO
RenderX

single data extraction performed for each country included in the study.

For each country, we applied 2 different GT queries. The first query focused on the *pseudo-influenza syndrome topic*, which was subsequently renamed as the *common cold* topic (of note, "topics" encompass groups of search terms associated with a specific concept regardless of language [12,19,21,24]). The second query consisted of a combination of search terms related to the *common cold*, carefully selected through discussions with native speakers of each language:

- Portugal: constipação + resfriado
- Spain: resfriado + resfrío + catarro + constipado + refredat + constipate + arrefriado + hotzeri
- Brazil: resfriado

We did not include quotation marks for the search terms as each term represented a single word. Misspellings or nonaccentuated forms were also excluded from the search term combinations, as we identified identical relative search volumes observed whenever misspelt words were or were not included in search term combinations.

Asthma Hospitalization Data Sources

We analyzed hospitalization data from January 1, 2012, to December 17, 2016 (we excluded the last 2 weeks of 2016 due to unavailable information on discharges in Portugal and Brazil—as many patients admitted toward the end of 2016 were discharged in 2017). In the 3 countries under investigation, we examined all hospitalizations in which asthma was identified as the primary diagnosis. Specifically, we used the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), code 493.x or International Classification of Diseases, Tenth Revision (ICD-10-CM) [25], code J45.x to identify these cases. The hospitalization data were obtained from the following sources: (1) the Hospital Morbidity database, provided by the Portuguese Central Administration of the Healthcare System, for Portugal [26]; (2) the Hospital Morbidity Survey databases (*Encuesta de morbilidad hospitalaria, Instituto Nacional de Estadistica*) for Spain [27]; and (3) Departamento de Informática do Sistema Único de Saúde (DATASUS) data from the Single Health System (*Sistema Único de Saúde*) for Brazil [28].

For each country, separate analyses were performed by participants' sex and age group (we considered only the age groups 18-64 years and ≥65 years to ensure a sufficient weekly number of hospital admissions for each analysis). Based on data availability for Portugal [20], analyses were also separately performed for episodes with at least one comorbidity from the Charlson comorbidity group and without any comorbidity from the Charlson comorbidity group. Likewise, based on data availability for Brazil, analyses were also separately performed by participants' ethnicity (White versus Black or Brown). Information on ethnicity was not available for Portugal and Spain, and information on the presence of comorbidities was not available for Spain and Brazil.

Statistical Analysis

Data analysis encompassed 2 different steps: (1) assessing the correlations between GT data and asthma hospitalizations in each country after applying time series analysis methods and (2) building models forecasting asthma hospitalizations for a period of 1 year based on GT and hospitalization data from the previous 4 years. To evaluate the predictive capability of the models, we compared the forecasted asthma admissions with the observed hospitalization data. For both GT and the frequency of hospitalizations, we worked with data displayed on a weekly basis (as it allowed detection of short-term variations while mitigating the impact of large random fluctuations that can occur when data are examined on a daily basis).

First, we performed a cross-correlation analysis to examine the correlation between GT data and asthma hospitalizations (cross-correlation can be understood as a statistical method used to analyze the relationship [correlate] between 2 continuous variables that can be measured or sampled at different points in time) [29]. Given that (1) for GT data, a relevant secular trend was expected (reflecting an increase in Google searches over time) and (2) GT and hospitalization data are expressed on different scales (GT results are expressed as relative search values [ie, percentages in relation to the maximum observed value of the whole period], whereas hospitalizations are expressed as absolute values), we removed the secular trend component for both GT and hospitalization data then assessed the correlation between search volumes and asthma hospitalizations. We analyzed Spearman correlations between GT and hospitalization data registered in the same week, as well as cross-correlation coefficients for a lag of 1 and 2 weeks to determine if search volumes demonstrated a stronger correlation with asthma hospitalizations occurring afterward rather than those happening concurrently. Correlation coefficients were presented alongside 95% CIs, which were computed using bootstrap methods for Spearman correlation coefficients.

Second, we built seasonal autoregressive integrated moving average (ARIMA) models to forecast variations in asthma hospitalizations over a period of 1 year [19,30]. Seasonal ARIMA models are used to forecast time series data that exhibit repeating patterns over fixed intervals (in this case, yearly cycles). These models take into account both the nonseasonal patterns and the seasonal variations in the data to make accurate predictions for future time points [31]. For each analysis, seasonal ARIMA models parameters $(p, d, q)(P', D, Q)s$ were defined, where p denotes the order of autoregression, d denotes the degree of difference, q denotes the order of the moving average part, P' denotes the seasonal order of autoregression, D denotes the degree of difference following seasonal integration, Q denotes the seasonal moving average, and s denotes the length of the seasonal period. We chose d and D so that the 2012-2016 time series appeared stationary (ie, with constant variance and no extreme fluctuations or overall increasing or decreasing behavior), including by testing a measure of seasonal strength [32]; we chose s=52 weeks (since there are roughly 52 weeks in a year); we chose p and P' based on spikes in partial autocorrelation function plots; and we chose q and Q based on spikes in autocorrelation function plots. Identification of these parameters using autocorrelation and

XSL•FO
RenderX

partial autocorrelation plots allowed us to define candidate models. Final seasonal ARIMA models were then selected based on the results of the Ljung-Box test (which was applied to assess whether residuals look like white noise) and on the minimization of the Akaike information criteria and Bayesian information criteria (see Table S1 in Multimedia Appendix 1 for the parameters defined for each model). In this study, we used asthma hospitalization data alongside GT data to forecast future asthma hospitalizations. The data set was split into a training set and a testing set. Specifically, the training set comprised asthma hospitalizations and GT data collected from July 1, 2012, to June 20, 2015. We then used this trained model to forecast asthma hospitalizations for the testing set, which included hospitalizations between the weeks of June 21, 2015, and June 19, 2016. This split allowed for the evaluation of model performance on previously unseen data.

To evaluate the predictive performance of the models, several measures were used: (1) the Spearman correlation coefficients between the predicted variation in hospitalizations and the actual number of asthma hospitalizations (ie, without time series decomposition), (2) the average weekly difference between the numbers of predicted and observed hospitalizations, and (3) the number of weeks whose number of observed asthma hospitalizations fell outside the 95% CI for predicted admissions.

All analyses were performed using R software, version 4.3.0 (R Foundation for Statistical Computing) [33], using the *forecast* and *urca* packages.

Ethical Considerations

Ethics Approval

The data used in this study were provided by the Central Administration of the Health System (Administração Central do Sistema de Saúde [ACSS]) in accordance with their institutional data-sharing policies. These data consist of the Morbidity Hospital Database (Bases de Dados de Morbilidade Hospitalar), which includes anonymized and de-identified data. Per the ACSS's internal guidelines, data anonymization and de-identification are conducted before any access is granted to external researchers. As a result, specific ethical approval was not required, as the use of anonymized data aligns with both Portuguese data protection regulations and the institutional policy governing secondary data analysis. [34-36],

Privacy and Confidentiality

The ACSS guarantees that the provided data sets are fully anonymized, making it impossible to identify individual patients. In addition, strict data use agreements are in place, which ensure that external entities, such as the authors of this study, commit to (1) using the data exclusively for research within the scope of their project, ensuring secure and fair data processing; (2) requesting explicit authorization from ACSS for any other use beyond the agreed scope; (3) not sharing the data with third parties; (4) citing ACSS as the source of the data in any resulting publications; (5) providing ACSS with copies of all publications that use the data; and (6) taking full responsibility for any analysis or conclusions drawn from the provided data sets.

Compensation

No compensation was provided, as the study did not involve direct patient recruitment or interaction.

Additionally, any identification of specific hospitals or the disclosure of medical device pricing data requires explicit approval from the respective institutions. This confidentiality further strengthens the protection of sensitive information while allowing for the comprehensive analysis of anonymized data.

Results

Between 2012 and 2016, GT data for *pseudo-influenza syndrome* presented similar patterns across the 3 countries for which GT data were plotted, with peaks in the winter and valleys in the summer of the respective hemispheres (Figure 1). This pattern was also observed for asthma hospitalizations in each subgroup of patients in each country.

In the assessed countries and for each subgroup of admissions, correlations between GT on the *pseudo-influenza syndrome* topic (after removing the trend component) and asthma hospitalizations ranged between 0.33 (in Portugal for admissions with at least one Charlson comorbidity group) and 0.86 (for admissions of women and White individuals in Brazil; Table 1). Similar values were observed when analyzing the correlations between GT and terms for the *common cold*. In the 3 countries, stronger correlation coefficients were observed for admissions occurring for women. In Portugal and Spain, stronger correlations were found for admissions of younger individuals, while in Brazil, the inverse phenomenon occurred. In Brazil, no differences were observed between correlations for admissions of patients of White or Black/Brown ethnicity. In Portugal, stronger correlations were observed for admissions of patients without comorbidities than for those with at least one comorbidity.

In most cases, GT on the *pseudo-influenza syndrome* topic correlated more strongly with asthma hospitalizations occurring in the subsequent week than with those occurring in the same week (Table 1).

In the 3 assessed countries (Table 2; Figure 2), forecasted hospitalizations for 2015-2016 obtained through seasonal ARIMA models correlated more strongly with observed admissions of older adults versus younger individuals (Portugal: correlation coefficient [ρ]= 0.70 vs ρ= 0.56; Spain: ρ=0.88 vs ρ=0.76; Brazil: ρ=0.83 vs ρ=0.82). In Portugal and Spain, forecasted hospitalizations displayed a much stronger correlation with admissions occurring for women than for men (Portugal: ρ=0.75 vs ρ=0.52; Spain: ρ=0.83 vs ρ=0.51). Consistent results were observed when performing a sensitivity analysis by age group (Table S2 in Multimedia Appendix 1). In Brazil, stronger correlations were observed for admissions of White individuals than of Black or Brown individuals (ρ=0.92 vs ρ=0.87). In Portugal, stronger correlations were observed for admissions of individuals without any comorbidity compared with admissions of individuals with comorbidities (ρ=0.68 vs ρ=0.66). The numbers of weeks with observed hospitalizations outside the confidence interval for predicted values ranged between 1 and 7 for Portugal, 2 and 12 for Spain, and 0 and 1

XSL•FO
RenderX

for Brazil (according to subgroups of admissions in each country).

Figure 1. Google Trends (GT) data on pseudo-influenza syndrome and asthma hospitalizations (2012-2016) for (A.1) men in Portugal, (B.1) men in Spain, (C.1) men in Brazil, (A.2) women in Portugal, (B.2) women in Spain, (C.2) women in Brazil, (A.3) older adults in Portugal, (B.3) older adults in Spain, (C.3) older adults in Brazil, (A.4) younger adults in Portugal, (B.4) younger adults in Spain, (C.4) younger adults in Brazil, (C.5) White people in Brazil, (C.6) Black or Brown people in Brazil, (A.7) people with comorbidities in Portugal, and (A.8) people without comorbidities in Portugal. The trend component of the time series has been plotted after removing the seasonal effects and random error components.

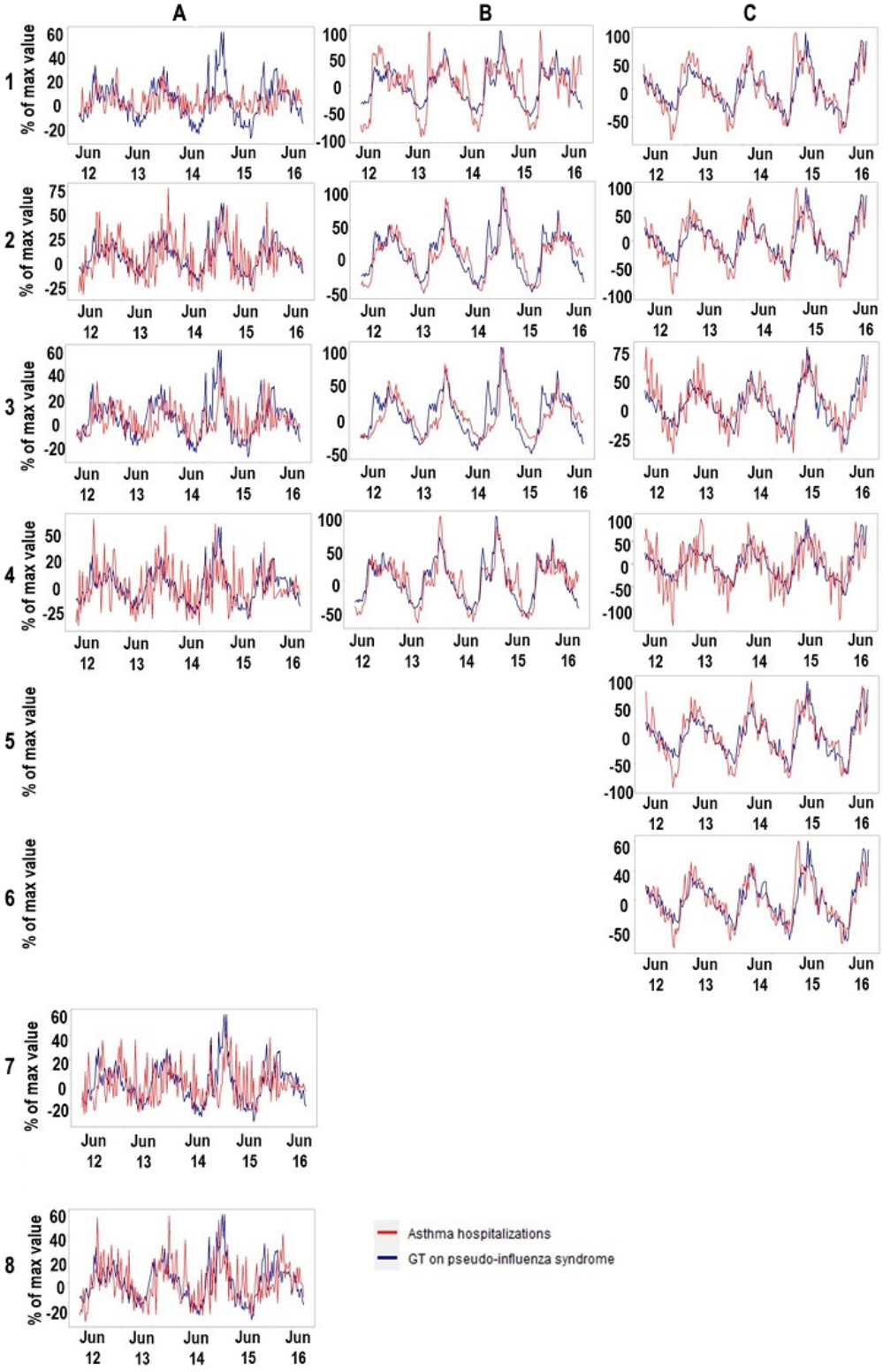

XSL•FO
RenderX

Table 1. Correlation and cross-correlation coefficients between common cold Google Trends data (ie, Google Trends data on the pseudo-influenza syndrome topic and on common cold search terms) and asthma hospitalizations (according to age, sex, ethnicity, and presence of comorbidities) for the period 2012-2016.

Categories	Results based on observed data, correlation coefficients (95% CI)	Results after removal of the trend component, cross-correlation coefficients (95% CI)		
		Week lag –0	Week lag –1	Week lag –2
Pseudo-influenza syndrome topic				
Portugal				
Sex				
Male	0.41 (0.30 to 0.52)	0.36 (–0.17 to 0.17)	0.37 (–0.17 to 0.17)	0.36 (–0.16 to 0.16)
Female	0.47 (0.36 to 0.58)	0.44 (–0.19 to 0.19)	0.51 (–0.20 to 0.20)	0.47 (–0.19 to 0.19)
Age group (years)				
>65	0.38 (0.25 to 0.50)	0.37 (–0.22 to 0.22)	0.47 (–0.22 to 0.22)	0.49 (–0.21 to 0.21)
18-64	0.43 (0.31 to 0.54)	0.41 (–0.15 to 0.15)	0.42 (–0.16 to 0.16)	0.34 (–0.16 to 0.16)
Comorbidities				
With comorbidities	0.33 (0.19 to 0.46)	0.32 (–0.18 to 0.18)	0.38 (–0.18 to 0.18)	0.38 (–0.18 to 0.18)
Without comorbidities	0.51 (0.40 to 0.61)	0.36 (–0.17 to 0.17)	0.37 (–0.16 to 0.16)	0.36 (–0.17 to 0.17)
Spain				
Sex				
Male	0.63 (0.53 to 0.72)	0.65 (–0.34 to 0.34)	0.61 (–0.35 to 0.35)	0.57 (–0.35 to 0.35)
Female	0.79 (0.72 to 0.84)	0.83 (–0.39 to 0.39)	0.86 (–0.40 to 0.40)	0.86 (–0.40 to 0.40)
Age group (years)				
>65	0.69 (0.60 to 0.76)	0.73 (–0.41 to 0.41)	0.80 (–0.40 to 0.40)	0.82 (–0.40 to 0.40)
18-64	0.81 (0.75 to 0.86)	0.83 (–0.34 to 0.34)	0.86 (–0.34 to 0.34)	0.83 (–0.34 to 0.34)
Brazil				
Sex				
Male	0.85 (0.80 to 0.88)	0.82 (–0.37 to 0.37)	0.75 (–0.37 to 0.37)	0.68 (–0.37 to 0.37)
Female	0.86 (0.82 to 0.89)	0.83 (–0.34 to 0.34)	0.78 (–0.33 to 0.33)	0.72 (–0.33 to 0.33)
Age group (years)				
>65	0.74 (0.67 to 0.78)	0.68 (–0.29 to 0.29)	0.70 (–0.29 to 0.29)	0.69 (–0.30 to 0.230)
18-64	0.70 (0.63 to 0.75)	0.65 (–0.27 to 0.27)	0.63 (–0.27 to 0.27)	0.59 (–0.27 to 0.27)
Ethnicity				
White	0.86 (0.82 to 0.89)	0.84 (–0.35 to 0.35)	0.79 (–0.35 to 0.35)	0.73 (–0.35 to 0.35)
Black or Brown	0.85 (0.80 to 0.88)	0.81 (–0.32 to 0.32)	0.75 (–0.32 to 0.32)	0.69 (–0.32 to 0.32)
Common cold search terms				
Portugal				
Sex				
Male	0.35 (0.23 to 0.48)	0.31 (–0.16 to 0.16)	0.30 (–0.16 to 0.16)	0.34 (–0.16 to 0.16)
Female	0.45 (0.33 to 0.56)	0.46 (–0.19 to 0.19)	0.46 (–0.18 to 0.18)	0.52 (–0.19 to 0.19)
Age group (years)				
>65	0.37 (0.24 to 0.50)	0.41 (–0.20 to 0.20)	0.45 (–0.20 to 0.20)	0.49 (–0.20 to 0.20)
18-64	0.42 (0.30 to 0.53)	0.41 (–0.15 to 0.15)	0.38 (–0.15 to 0.15)	0.42 (–0.16 to 0.16)
Comorbidities				
With comorbidities	0.33 (0.19 to 0.45)	0.34 (–0.16 to 0.16)	0.35 (–0.16 to 0.16)	0.42 (–0.15 to 0.15)

XSL•FO
RenderX

Categories	Results based on observed data, correlation coefficients (95% CI)	Results after removal of the trend component, cross-correlation coefficients (95% CI)		
		Week lag –0	Week lag –1	Week lag –2
Without comorbidities	0.48 (0.36 to 0.58)	0.48 (–0.21 to 0.21)	0.49 (–0.21 to 0.21)	0.51 (–0.20 to 0.20)
Spain				
Sex				
Male	0.63 (0.53 to 0.71)	0.64 (–0.34 to 0.34)	0.60 (–0.34 to 0.34)	0.57 (–0.34 to 0.34)
Female	0.78 (0.71 to 0.83)	0.83 (–0.39 to 0.39)	0.86 (–0.39 to 0.39)	0.85 (–0.39 to 0.39)
Age group (years)				
>65	0.69 (0.60 to 0.76)	0.73 (–0.40 to 0.40)	0.80 (–0.40 to 0.40)	0.81 (–0.40 to 0.40)
18-64	0.81 (0.74 to 0.86)	0.83 (–0.33 to 0.33)	0.85 (–0.36 to 0.36)	0.83 (–0.36 to 0.36)
Brazil				
Sex				
Male	0.84 (0.81 to 0.87)	0.81 (–0.34 to 0.34)	0.75 (–0.34 to 0.34)	0.66 (–0.34 to 0.34)
Female	0.86 (0.82 to 0.89)	0.82 (–0.32 to 0.32)	0.78 (–0.32 to 0.32)	0.70 (–0.31 to 0.31)
Age group (years)				
>65	0.74 (0.69 to 0.79)	0.69 (–0.29 to 0.2)	0.69 (–0.29 to 0.29)	0.70 (–0.29 to 0.29)
18-64	0.70 (0.63 to 0.76)	0.65 (–0.25 to 0.25)	0.65 (–0.25 to 0.25)	0.60 (–0.25 to 0.25)
Ethnicity				
White	0.86 (0.83 to 0.89)	0.84 (–0.33 to 0.33)	0.78 (–0.33 to 0.33)	0.71 (–0.34 to 0.34)
Black or Brown	0.84 (0.80 to 0.87)	0.80 (–0.31 to 0.31)	0.75 (–0.31 to 0.31)	0.68 (–0.31 to 0.31)

XSL•FO
RenderX

Table 2. Results of 1-year (June 2015 to June 2016) forecasts for the number of asthma hospitalizations (according to age, sex, ethnicity, and presence of comorbidities) based on autoregressive integrated moving average models including common cold–related Google Trends data and asthma hospitalizations from the previous 3 years.

Categories	Results for number of predicted and observed hospitalizations, correlation (95% CI)	Average difference in the absolute numbers of predicted and observed weekly hospitalizations, average	Weeks with observed hospitalizations outside the predicted 95% CIs, n (%)
Pseudo-influenza syndrome topic			
Portugal			
Sex			
Male	0.52 (0.26-0.71)	6.5	1 (1.9)
Female	0.75 (0.57-0.85)	26.3	5 (9.6)
Age group (years)			
>65	0.70 (0.51-0.83)	14.9	6 (11.5)
18-64	0.56 (0.38-0.70)	27.8	7 (13.5)
Comorbidities			
With comorbidities	0.66 (0.45-0.80)	21.9	1 (1.9)
Without comorbidities	0.68 (0.52-0.81)	9.5	1 (1.9)
Spain			
Sex			
Male	0.51 (0.22-0.72)	49.1	2 (3.9)
Female	0.83 (0.65-0.92)	111.2	12 (23.1)
Age group (years)			
>65	0.88 (0.74-0.95)	101.9	12 (23.1)
18-64	0.76 (0.59-0.88)	30.1	7 (13.5)
Sensitivity analyses by age group (years)			
>65	0.85 (0.73-0.91)	71.3	3 (5.7)
45-64	0.89 (0.78-0.94)	10.8	0 (0)
18-44	0.85 (0.72-0.92)	11.8	18 (1.8)
Brazil			
Sex			
Male	0.91 (0.83-0.94)	75.1	0 (0)
Female	0.89 (0.81-0.93)	71.8	0 (0)
Age group (years)			
>65	0.83 (0.71-0.90)	32.8	1 (1.9)
18-64	0.82 (0.69-0.89)	44.6	1 (1.9)
Sensitivity analyses by age group (years)			
>65	0.87 (0.78-0.92)	20.7	3 (5.7)
45-64	0.78 (0.63-0.88)	19.3	2 (3.8)
18-44	0.74 (0.57-0.85)	22.1	0 (0)
Ethnicity			
White	0.92 (0.84-0.95)	40.8	0 (0)
Black or Brown	0.87 (0.75-0.93)	71.1	0 (0)

XSL•FO
RenderX

Figure 2. Predicted number (estimated based on previous hospitalizations and Google Trends data) and observed number of asthma hospitalizations (June 2015 to June 2016) for (A) men in Portugal, (B) men in Spain, (C) men in Brazil, (D) women in Portugal, (E) women in Spain, (F) women in Brazil, (G) older adults in Portugal, (H) older adults in Spain, (I) older adults in Brazil, (J) younger adults in Portugal, (K) younger adults in Spain, (L) younger adults in Brazil, (M) White people in Brazil, (N) Black or Brown people in Brazil, (O) people with comorbidities in Portugal, and (P) people without comorbidities in Portugal.

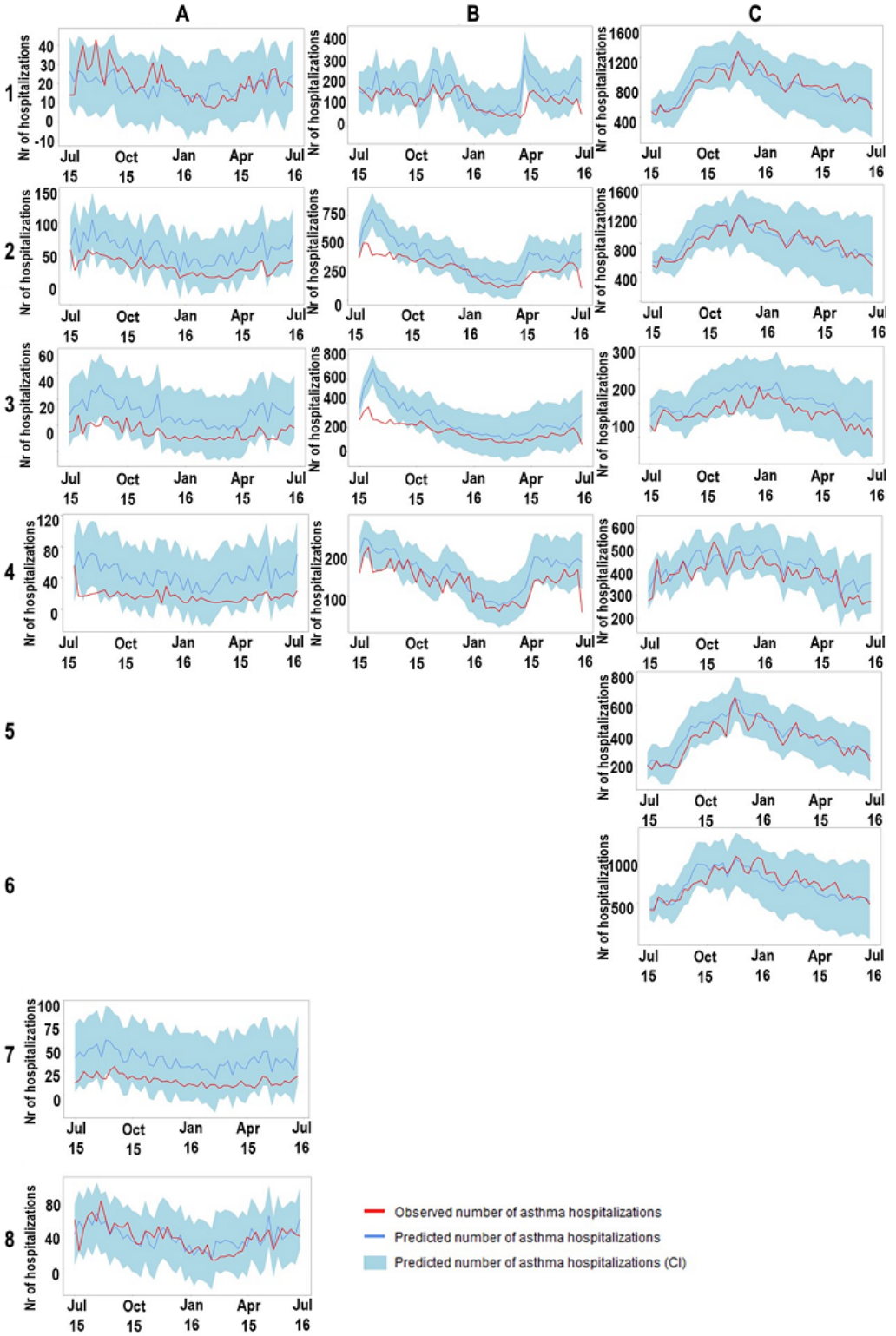

Discussion

Principal Findings

In this study, we assessed the correlations of GT data and the performance of GT-based models in different subgroups of patients (as defined by clinical and demographic characteristics: age, sex, ethnicity, and presence of comorbidities) using the case study of asthma. Overall, our results point out that GT-based models may not necessarily have the same performance in all subgroups of patients, highlighting that GT data may vary across different segments of users. In fact, we

XSL•FO
RenderX

observed stronger correlations between GT data and asthma hospitalization data or between forecasted and observed hospitalizations when assessing admissions of women or patients without comorbidities. Less consistent results were observed, in particular in Brazil, according to age group.

Overall, studies using GT data for surveillance purposes have obtained mixed results. On the one hand, GT has been shown to have an effective potential to monitor the spread of infectious diseases, track public interest in health-related topics, and identify emerging trends in public health [37,38]. However, there have also been instances where GT has shown inconsistencies or failed to provide accurate predictions, emphasizing the need to carefully interpret the data [23]. In part, these failures may be related to differences in the composition of internet users compared with that of patients with a particular disease. Although we cannot necessarily generalize the results observed in the use case of asthma hospitalizations to other conditions or countries, this paper is relevant from a methodological point of view, as it demonstrates, through a case study, how the association between GT and disease data is not always the same for all groups of individuals, pointing to the need to study these associations according to the characteristics of the patients.

This study is also relevant for asthma care, as this was the condition we particularly assessed. Regarding our findings of the performance of GT-based models in the distinct subgroups of asthma hospitalizations, we observed relevant gender-related differences. In fact, women have higher asthma prevalence, severity, and health care utilization than men [39,40]. The better correlations and model performance observed in female admissions may be related to the higher prevalence of asthma in this population. On the other hand, women often exhibit more proactive health information–seeking behaviors, with a particular emphasis on their own health and well-being as well as that of their families—which may partly explain the higher internet use by women than men [41]. In addition, in some cultures, women may have primary caregiving responsibilities for family members' health, including managing asthma [42]. This can also contribute to increasing interest and information-seeking behavior and enhance engagement with online platforms, possibly explaining the higher correlations and better performance of models in admissions of women.

Younger adults, especially those who are generally healthy, may exhibit different health-seeking behaviors than older adults or individuals with chronic illnesses [43]. Younger adults may tend to be more proactive in seeking health care information online and to be more likely to use search engines [44], in part given their historically higher access and rates of internet use [44] (which can be attributed to factors such as greater digital literacy, increased reliance on technology for information and communication, and higher rates of smartphone ownership [44,45]). However, that access has also been proliferating very quickly among older adults, who may possibly be more concerned about their health [46]. These changing patterns may partly explain the heterogeneity of our results obtained on age groups, with higher correlations found for younger adults in contrast with better performance of forecasting models in admissions of older adults. Such a pattern was also observed when performing separate analyses for age groups of 15 to 44-year-olds and 45 to 64-year-olds in Spain and Brazil. All things considered, our findings may offer insights into digital divides, hinting at disparities in internet access, digital literacy, and health information–seeking behaviors across demographic groups. The use of GT-based tools for complementing surveillance systems may have important implications in terms of health equity, considering the discrepancies in internet access across clinical and demographic subgroups.

Information on the presence of comorbidities or ethnicity of patients was only available for one country each (Portugal and Brazil, respectively). The presence of comorbidities has been associated with worse health outcomes for asthma admissions [47]. In addition, we observed a less pronounced seasonal pattern in Portuguese hospitalizations of individuals with comorbidities than in those without, potentially explaining the worse performance of GT-based models in forecasting those admissions. Small differences were observed regarding ethnicity, with a slightly better performance of models for Whites in Brazil, possibly reflecting different regional demographics, internet use patterns, or health care–seeking patterns.

Strengths and Limitations

Several limitations should be discussed. First, the differences observed in the performance of GT across different subgroups are not necessarily generalizable to other countries and conditions (eg, we cannot state that GT-based models always display better performance when considering data from female participants). Second, data availability on hospitalizations was limited to 3 countries and, regarding the presence of comorbidities or ethnicity, we only had that information for Portugal and for Brazil, respectively. In addition, the small frequency of weekly admissions precluded the comparison of the performance of the models' unspecific sets of Charlson comorbidity groups. Third, GT provides data on search term popularity and relative interest (ie, GT presents searches as a relative volume instead of as an absolute number of searches), which makes comparisons between queries difficult and reveals less information about the absolute search interest in each aspect being assessed. In addition, it does not provide detailed information about the context or intent behind the searches, thus making it prone to bias due to possible media coverage [48]. In the particular context of this study, we were not able to quantify the number of searches on the "common cold" resulting from individuals experiencing cold symptoms versus reflecting other intentions (eg, search for news on the common cold). This lack of specificity can make it challenging to establish a causal link between search behaviors and the studied outcomes. However, this is an inherent limitation of GT, and our goal was not so much to establish an association between searches on "common cold" and asthma hospitalizations but rather to assess how that association varies considering different subgroups. Finally, during the assessed period, there was an increase in the use of the internet. However, we applied time series analysis methods, removing the estimated trend components for both GT and hospitalization data.

This study also had several strengths. In particular, this study has an important novelty component—to the best of our

XSL•FO
RenderX

knowledge, this is the first time that the performance of GT-based models has been investigated across diverse demographic and clinical subgroups, with relevant potential implications for considering digital divides and health equity–related aspects in interpreting results of GT-based tools. In addition, we assessed 3 different countries (1 in Europe and 1 in South America) using nationwide data for a period of 5 years. We applied 2 different strategies to retrieve common cold–related GT data—GT data on the *pseudo-influenza syndrome* topic and search terms regarding the *common cold*—which obtained comparable results. We examined asthma as a case study since (1) asthma, in comparison with other diseases (such as COVID-19), is less subject to a high or variable media coverage, thus not particularly biased for GT data [30,48]; (2) the relationship between asthma admissions and GT data on the common cold has been already established [19]; and (3) the influence of patients' characteristics on asthma outcomes has been assessed [20]. Although this study relied on a case study on asthma admissions, there is potential application of this methodology to other diseases and segments of the population to better understand the context in which GT-based models can be better applied.

Conclusions

In this study, we observed better performance of models forecasting asthma hospitalizations in women, White individuals (Brazil), and patients without comorbidities (Portugal), suggesting that the models based on GT may perform differently in subgroups of participants, which may indicate variations in the patterns of health-related information seeking among different segments of internet users. Although GT data have increasingly been assessed as a potential complementary tool to more classical surveillance approaches, determining the best practices for using GT data and understanding its limitations requires exploring in which segments of users it performs better. Although this study assessed the use case of asthma in 3 countries and shows differences in different segments of the population, future studies should explore how GT-related models may differentially perform in accordance with other variables, such as sociodemographic variables (like age, gender, education, income, urban/rural context, underserved populations), as well as to test differences observed in other diseases, countries, and clinical data sources. This study contributes to advancing our understanding of the complexities inherent in the infodemiology field and hints at the need to consider population subgroups and health contexts for the applicability of GT-based surveillance systems.

Acknowledgments

The authors wish to thank the Portuguese Ministry of Health for providing access to the hospitalization data managed by the Portuguese Central Health System Administration (Administração Central do Sistema de Saúde).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Parameters used for autoregressive integrated moving average models and results of a 1-year (June 2015 to June 2016) forecasts for the number of asthma hospitalizations.
[DOCX File , 22 KB - jmir_v27i1e51804_app1.docx]

References

1. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
2. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
3. Nicholson KG, Kent J, Ireland DC. Respiratory viruses and exacerbations of asthma in adults. BMJ 1993 Oct 16;307(6910):982-986 [FREE Full text] [doi: 10.1136/bmj.307.6910.982] [Medline: 8241910]
4. Johnston SL, Pattemore PK, Sanderson G, Smith S, Lampe F, Josephs L, et al. Community study of role of viral infections in exacerbations of asthma in 9-11 year old children. BMJ 1995 May 13;310(6989):1225-1229 [FREE Full text] [doi: 10.1136/bmj.310.6989.1225] [Medline: 7767192]
5. Busse WW, Lemanske RF, Gern JE. Role of viral respiratory infections in asthma and asthma exacerbations. The Lancet 2010 Sep;376(9743):826-834. [doi: 10.1016/s0140-6736(10)61380-3]
6. Sousa-Pinto B, Antó JM, Sheikh A, de Lusignan S, Haahtela T, Fonseca J, et al. Comparison of epidemiologic surveillance and Google Trends data on asthma and allergic rhinitis in England. Allergy 2022 Feb;77(2):675-678 [FREE Full text] [doi: 10.1111/all.15139] [Medline: 34665881]

XSL•FO
RenderX

7. Mavragani A, Sampri A, Sypsa K, Tsagarakis KP. Integrating smart health in the US health care system: infodemiology study of asthma monitoring in the Google era. JMIR Public Health Surveill 2018 Mar 12;4(1):e24 [FREE Full text] [doi: 10.2196/publichealth.8726] [Medline: 29530839]
8. Gesualdo F, Stilo G, D'Ambrosio A, Carloni E, Pandolfi E, Velardi P, et al. Can Twitter be a source of information on allergy? Correlation of pollen counts with tweets reporting symptoms of allergic rhinoconjunctivitis and names of antihistamine drugs. PLoS One 2015;10(7):e0133706 [FREE Full text] [doi: 10.1371/journal.pone.0133706] [Medline: 26197474]
9. Joshi A, Sparks R, McHugh J, Karimi S, Paris C, MacIntyre CR. Harnessing tweets for early detection of an acute disease event. Epidemiology 2020 Jan;31(1):90-97 [FREE Full text] [doi: 10.1097/EDE.0000000000001133] [Medline: 31651659]
10. Wakamiya S, Morimoto O, Omichi K, Hara H, Kawase I, Koshiba R, et al. Exploring relationships between tweet numbers and over-the-counter drug sales for allergic rhinitis: retrospective analysis. JMIR Form Res 2022 Feb 02;6(2):e33941 [FREE Full text] [doi: 10.2196/33941] [Medline: 35107434]
11. Eysenbach G. Infodemiology: the epidemiology of (mis)information. The American Journal of Medicine 2002 Dec;113(9):763-765. [doi: 10.1016/S0002-9343(02)01473-0]
12. Bousquet J, Onorato G, Oliver G, Basagana X, Annesi-Maesano I, Arnavielhe S, et al. Google Trends and pollen concentrations in allergy and airway diseases in France. Allergy 2019 Oct;74(10):1910-1919 [FREE Full text] [doi: 10.1111/all.13804] [Medline: 30942904]
13. Osborne NJ, Alcock I, Wheeler BW, Hajat S, Sarran C, Clewlow Y, et al. Pollen exposure and hospitalization due to asthma exacerbations: daily time series in a European city. Int J Biometeorol 2017 Oct 12;61(10):1837-1848 [FREE Full text] [doi: 10.1007/s00484-017-1369-2] [Medline: 28500390]
14. Jabour AM, Varghese J, Damad AH, Ghailan KY, Mehmood AM. Examining the correlation of Google influenza trend with hospital data: retrospective study. J Multidiscip Healthc 2021;14:3073-3081 [FREE Full text] [doi: 10.2147/JMDH.S322185] [Medline: 34754195]
15. Kandula S, Pei S, Shaman J. Improved forecasts of influenza-associated hospitalization rates with Google Search Trends. J R Soc Interface 2019 Jun 28;16(155):20190080. [doi: 10.1098/rsif.2019.0080] [Medline: 31185818]
16. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
17. Saegner T, Austys D. Forecasting and surveillance of COVID-19 spread using Google Trends: literature review. Int J Environ Res Public Health 2022 Sep 29;19(19):12394 [FREE Full text] [doi: 10.3390/ijerph191912394] [Medline: 36231693]
18. Amusa LB, Twinomurinzi H, Okonkwo CW. Modeling COVID-19 incidence with Google Trends. Front Res Metr Anal 2022;7:1003972 [FREE Full text] [doi: 10.3389/frma.2022.1003972] [Medline: 36186843]
19. Sousa-Pinto B, Halonen JI, Antó A, Jormanainen V, Czarlewski W, Bedbrook A, et al. Prediction of asthma hospitalizations for the common cold using Google Trends: infodemiology study. J Med Internet Res 2021 Jul 06;23(7):e27044 [FREE Full text] [doi: 10.2196/27044] [Medline: 34255692]
20. Portela D, Pereira Rodrigues P, Freitas A, Costa E, Bousquet J, Fonseca JA, et al. Impact of multimorbidity patterns in hospital admissions: the case study of asthma. J Asthma 2023 Sep 21;60(9):1723-1733 [FREE Full text] [doi: 10.1080/02770903.2023.2185154] [Medline: 36848045]
21. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
22. Freitas A, Lema I, da Costa-Pereira A. Comorbidity Coding Trends in Hospital Administrative Databases. In: Rocha Á, Correia A, Adeli H, Reis L, Mendonça Teixeira M, editors. New Advances in Information Systems and Technologies. Advances in Intelligent Systems and Computing, vol 445. Cham, Switzerland: Springer; 2016:609-617.
23. Cebrián E, Domenech J. Is Google Trends a quality data source? Applied Economics Letters 2022 Jan 05;30(6):811-815 [FREE Full text] [doi: 10.1080/13504851.2021.2023088]
24. Bousquet J, Agache I, Anto J, Bergmann K, Bachert C, Annesi-Maesano I, et al. Google Trends terms reporting rhinitis and related topics differ in European countries. Allergy 2017 Aug;72(8):1261-1266 [FREE Full text] [doi: 10.1111/all.13137] [Medline: 28140507]
25. ICD-9-CM Official Guidelines for Coding and Reporting. Centers for Disease Control and Prevention. 2011. URL: https://www.cdc.gov/nchs/data/icd/icd9cm_guidelines_2011.pdf [accessed 2025-03-01]
26. Principle guidelines for the contracting of health services from a National Health Service (SNS) perspective. ACSS. 2017. URL: https://www.acss.min-saude.pt/wp-content/uploads/2017/04/Termos-de-Referencia-para-2017_vf.pdf [accessed 2023-08-01]
27. Encuesta de morbilidad hospitalaria. Año 2022. Instituto Nacional de Estadística. 2024 Mar 21. URL: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176778&menu=ultiDatos&idp=1254735573175 [accessed 2023-08-01]
28. Basic Health Indicators and Data - Brazil - 2002. DATASUS. 2002. URL: http://tabnet.datasus.gov.br/cgi/idb2002/apresent_eng.htm [accessed 2023-08-01]

XSL•FO
RenderX

29. Kumar S, Chong I. Correlation analysis to identify the effective data in machine learning: prediction of depressive disorder and emotion states. Int J Environ Res Public Health 2018 Dec 19;15(12):1 [FREE Full text] [doi: 10.3390/ijerph15122907] [Medline: 30572595]
30. Sousa-Pinto B, Heffler E, Antó A, Czarlewski W, Bedbrook A, Gemicioglu B, et al. Anomalous asthma and chronic obstructive pulmonary disease Google Trends patterns during the COVID-19 pandemic. Clin Transl Allergy 2020 Nov 02;10(1):47 [FREE Full text] [doi: 10.1186/s13601-020-00352-9] [Medline: 33292529]
31. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 2024. URL: http://otexts.org/fpp/ [accessed 2025-03-01]
32. Wang X, Smith K, Hyndman R. Characteristic-based clustering for time series data. Data Min Knowl Disc 2006 May 16;13(3):335-364. [doi: 10.1007/s10618-005-0039-x]
33. R Team. R Foundation for Statistical Computing. Vienna, Austria: R: A Language and Environment for Statistical Computing URL: https://www.R-project.org/ [accessed 2025-03-01]
34. Loureiro da Silva C, Rocha JV, Santana R. Economic and financial crisis based on Troika's intervention and potentially avoidable hospitalizations: an ecological study in Portugal. BMC Health Serv Res 2021 May 26;21(1):506 [FREE Full text] [doi: 10.1186/s12913-021-06475-4] [Medline: 34039326]
35. Rocha JA, Cardoso JC, Freitas A, Allison TG, Azevedo LF. Time-trends and predictors of interhospital transfers and 30-day rehospitalizations after acute coronary syndrome from 2000-2015. PLoS One 2021;16(7):e0255134 [FREE Full text] [doi: 10.1371/journal.pone.0255134] [Medline: 34293045]
36. Rocha A, Azevedo LF, Silva Cardoso JC, Allison TG, Freitas A. Internal deterministic record linkage using indirect identifiers for matching of same-patient hospital transfers and early readmissions after acute coronary syndrome in a nationwide hospital discharge database: a retrospective observational validation study. BMJ Open 2019 Dec 30;9(12):e033486 [FREE Full text] [doi: 10.1136/bmjopen-2019-033486] [Medline: 31892664]
37. Wang D, Guerra A, Wittke F, Lang JC, Bakker K, Lee AW, et al. Real-time monitoring of infectious disease outbreaks with a combination of Google Trends search results and the moving epidemic method: a respiratory syncytial virus case study. Trop Med Infect Dis 2023 Jan 19;8(2):75 [FREE Full text] [doi: 10.3390/tropicalmed8020075] [Medline: 36828491]
38. Arora VS, McKee M, Stuckler D. Google Trends: opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
39. Colombo D, Zagni E, Ferri F, Canonica GW, PROXIMA study centers. Gender differences in asthma perception and its impact on quality of life: a post hoc analysis of the PROXIMA (Patient Reported Outcomes and Xolair In the Management of Asthma) study. Allergy Asthma Clin Immunol 2019 Nov 06;15(1):65 [FREE Full text] [doi: 10.1186/s13223-019-0380-z] [Medline: 31708980]
40. Zein JG, Erzurum SC. Asthma is different in women. Curr Allergy Asthma Rep 2015 Jun 4;15(6):28 [FREE Full text] [doi: 10.1007/s11882-015-0528-y] [Medline: 26141573]
41. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
42. Sharma N, Chakrabarti S, Grover S. Gender differences in caregiving among family - caregivers of people with mental illnesses. World J Psychiatry 2016 Mar 22;6(1):7-17 [FREE Full text] [doi: 10.5498/wjp.v6.i1.7] [Medline: 27014594]
43. Ma X, Liu Y, Zhang P, Qi R, Meng F. Understanding online health information seeking behavior of older adults: a social cognitive perspective. Front Public Health 2023 Mar 3;11:1147789 [FREE Full text] [doi: 10.3389/fpubh.2023.1147789] [Medline: 36935731]
44. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021 Dec 16;9(12):1 [FREE Full text] [doi: 10.3390/healthcare9121740] [Medline: 34946466]
45. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: implications for public health practice. J Public Health Manag Pract 2017;23(4):380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
46. Sixsmith A, Horst BR, Simeonov D, Mihailidis A. Older people's use of digital technology during the COVID-19 pandemic. Bull Sci Technol Soc 2022 Jun 21;42(1-2):19-24 [FREE Full text] [doi: 10.1177/02704676221094731] [Medline: 38603230]
47. Gershon AS, Wang C, Guan J, To T. Burden of comorbidity in individuals with asthma. Thorax 2010 Jul;65(7):612-618. [doi: 10.1136/thx.2009.131078] [Medline: 20627918]
48. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google Trends data: infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]

Abbreviations

ACSS: Administração Central do Sistema de Saúde
ARIMA: autoregressive integrated moving average
DATASUS: Departamento de Informática do Sistema Único de Saúde

XSL•FO
RenderX

GT: Google Trends
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10-CM: International Classification of Diseases, Tenth Revision

Edited by T de Azevedo Cardoso; submitted 13.08.23; peer-reviewed by S Jankin, L Amusa, O Serban; comments to author 05.03.24; revised version received 15.04.24; accepted 04.10.24; published 10.03.25.

Please cite as:
Portela D, Freitas A, Costa E, Giovannini M, Bousquet J, Almeida Fonseca J, Sousa-Pinto B
Impact of Demographic and Clinical Subgroups in Google Trends Data: Infodemiology Case Study on Asthma Hospitalizations
J Med Internet Res 2025;27:e51804
URL: https://www.jmir.org/2025/1/e51804
doi:10.2196/51804
PMID:

©Diana Portela, Alberto Freitas, Elísio Costa, Mattia Giovannini, Jean Bousquet, João Almeida Fonseca, Bernardo Sousa-Pinto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Contextualizing Changes in e-Cigarette Use During the Early COVID-19 Pandemic and Accompanying Infodemic ("So Much Contradictory Evidence"): Qualitative Document Analysis of Reddit Forums

Shannon Lea Watkins[1], PhD; Katherine Snodgrass[2], BA, BS; Lexi Fahrion[1], MPH; Emily Shaw[1], MPH

[1]Department of Community and Behavioral Health, College of Public Health, University of Iowa, Iowa City, IA, United States

[2]College of Public Health, University of Iowa, Iowa City, IA, United States

Corresponding Author:
Shannon Lea Watkins, PhD
Department of Community and Behavioral Health
College of Public Health
University of Iowa
145 N Riverside Dr
Iowa City, IA, 52242
United States
Phone: 1 3194671489
Email: shannon-watkins@uiowa.edu

Abstract

Background: Understanding how social media platforms facilitate information exchange and influence behavior during health crises can enhance public health responses during times of uncertainty. While some risk factors for COVID-19 susceptibility and severity (eg, old age) were clear, whether e-cigarette use increased risk was not clear. People who used e-cigarettes had to navigate both the COVID-19 infodemic and a conflicting, politicized, and changing information environment about the interaction between COVID-19 and e-cigarette use.

Objective: This study aims to characterize and contextualize e-cigarette–related behavior changes during the early COVID-19 pandemic and illuminate the role that social media played in decision-making.

Methods: We conducted a qualitative analysis of COVID-19–related e-cigarette discussions on 3 Reddit forums about e-cigarettes. We collected 189 relevant discussion threads made in the first 6 months of the pandemic (collected from June 27, 2020, to July 3, 2020). Threads included 3155 total comments (mean 17 comments) from approximately 1200 unique Redditors. We developed and applied emergent codes related to e-cigarette perceptions and behaviors (eg, the role of nicotine in COVID-19 and do-it-yourself narratives) and web-based community interactions (eg, advice), identified thematic patterns across codes, and developed a model to synthesize the socioecological context of e-cigarette behaviors.

Results: e-Cigarette subreddits provided a platform for Redditors to discuss perceptions and experiences with e-cigarettes, make sense of information, and provide emotional support. Discussions reflected an array of e-cigarette–related behavioral responses, including increases and decreases in use intensity, changes in purchasing practices (eg, stockpiling), and changes in vaping practices (eg, reusing disposable pods). This study presented a theoretically and empirically informed model of how circumstances created by the pandemic (eg, changes in activity space and product shortages) compelled behavior changes. Redditors drew from their existing perceptions, intentions, and experiences with nicotine and tobacco products; their personal pandemic experiences; and their participation on Reddit to decide whether and how to change their e-cigarette behaviors during the early pandemic. Forums reflected uncertainty, stress, and debate about the rapidly evolving and complicated public health information. Consumption and discussion of media (eg, news articles and peer-reviewed publications) on Reddit informed e-cigarette perceptions and behaviors. Decisions were complicated by distrust of the media.

Conclusions: Variations in individual traits and environmental circumstances during the early COVID-19 pandemic provide context for why there was no unified direction of e-cigarette behavior change during this period. Information and discussion on Reddit also informed risk perceptions and decisions during the pandemic. Social media is an effective and important place to communicate public health information, particularly during crisis or disaster situations. Moving forward, transparent, accurate,

XSL•FO
RenderX

and specific message development should consider the stress, struggles, and stigma of people who use e-cigarettes and address the roles mistrust and misinformation play in decisions.

(J Med Internet Res 2025;27:e66010) doi:10.2196/66010

KEYWORDS

vaping; nicotine; tobacco; health communication; social media; new media

Introduction

Background

During the first year of the COVID-19 pandemic, risk perceptions, related behaviors, and public health guidance changed rapidly, as both the virus and our understanding of it evolved. The COVID-19 pandemic was accompanied by an "infodemic," a glut of accurate and inaccurate information about a disease outbreak, which can cause confusion and undermine public trust [1-3]. Conflicts and changes in public health messaging confused some members of the public and seemed untrustworthy to others [4,5].

In this era of high stress, evolving public health information, and politicization of science, the public used social media as a tool to inform their health-related decisions [6]. During crises and disasters, social media facilitates information dissemination and communication but can also facilitate the intentional and unintentional spread of information that is "inaccurate, incomplete, vague, or ambiguous," [7] that is, *misinformation* [8-10]. During the intense isolation of the early pandemic, social media also provided essential human connection. In 2020, nearly one-third of Americans reported viewing COVID-19–related content on social media daily, and almost two-thirds of social media users reported they were unlikely to fact-check information about COVID-19 they read on social media with a health professional [6]. Understanding how social media platforms facilitate information exchange and impact behavior during health crises is critical to improving infodemic management and future public health response [1]. Moreover, social media conversations provide a unique record of the human experience during the pandemic.

Globally, the pandemic influenced multiple addictive behaviors, including alcohol consumption [11,12], gambling [13], cannabis use [14,15], and tobacco and nicotine use [12,16-20]. Evidence suggests that nicotine e-cigarette use, purchasing, and risk perception were all impacted [21] and that changes were mixed (ie, toward e-cigarette use and away from it) for both adolescents and adults [22-25]. e-Cigarette use is influenced by personal beliefs and experiences and interpersonal, informational, organizational, community, and policy contexts (ie, socioecological context) [26]. Influences include risk perceptions, social norms [27], e-cigarette advertising [28], social media [29], and access to e-cigarettes [30]. By transforming multiple domains of an individual's life (eg, activity space and social connections), the pandemic likely influenced e-cigarette use through multiple pathways [31].

People who used e-cigarettes exchanged e-cigarette–related information on the web during the pandemic [4,32], leveraging already-established social media forums on which individuals discussed products, flavors, current regulations, and personal experiences [33]. While some risk factors for COVID-19 susceptibility and severity, such as old age, were clear, whether e-cigarette and tobacco use increased risk was not clear [34-36]. People who used e-cigarettes or tobacco products not only had to navigate the COVID-19 infodemic but also make sense of conflicting, politicized, and changing information about the interaction between COVID-19 and e-cigarette and other nicotine and tobacco product use [37]. The mixed effect of the pandemic on e-cigarette use, the compounding and interacting uncertainties about COVID-19 and e-cigarette use, and the existing e-cigarette communities on the web suggest e-cigarette use is an excellent context to investigate how social media captured lived experience during the pandemic and how it facilitates health-related information exchange during public health crises. The COVID-19 pandemic will influence health and well-being long after the threat of severe illness subsides. Documenting *how* the profound personal and social changes of the pandemic influenced health risk behaviors in the early months of the crisis can reveal facilitators and barriers to health promotion, help us contextualize and interpret behavioral trends as they unfold in the years following, and suggest future communication interventions during crises.

Objectives

We aimed to explicate the mechanisms of e-cigarette behavior change, including engagement in web-based forums, during the early months of the COVID-19 pandemic. We leveraged social media as a source of qualitative data and conceptualized social media use as a phenomenon of study, conducting a qualitative analysis of e-cigarette forums on the social media platform Reddit (Reddit, Inc). Understanding the discourse in web-based e-cigarette communities during the COVID-19 pandemic can illuminate the range of individual experiences of vaping during that time; show how health information, including misinformation, is consumed, analyzed, and implemented on the web; and illuminate how individuals make health decisions during crisis.

Methods

We thematically analyzed randomly selected Reddit posts (posted from January 31, 2020, to July 3, 2020) related to the COVID-19 pandemic from 3 e-cigarette subreddits.

Study Site

Reddit is a social media site that allows anonymous users (Redditors) to post content on discussion boards (subreddits) dedicated to specific topics. Redditors interact by creating an original post (OP) related to the subreddit, commenting on an OP, and "upvoting" (liking) or "downvoting" (disliking) OPs and comments. The term OP refers to both *original posts* and

XSL•FO
RenderX

the Redditor who created that OP. Because each subreddit is dedicated to a specific topic, it provides space for relevant discussions, questions, and advice. Each month, approximately 500 million visitors participate on Reddit, and the site hosts 73 million unique users daily [38]. Because of these attributes, Reddit has been effectively used to investigate e-cigarette perceptions and behaviors [39-42] and is uniquely situated to study web-based social interactions during the pandemic through unobtrusive observation. A large study of psychosocial language markers in health discussions on Reddit reported substantial differences in topic and sentiment before and during the pandemic [43]. This qualitative study included 3 e-cigarette subreddits over 6 months to capture a range of reactions to the pandemic.

Data Collection

In each subreddit, we searched for OPs using 8 terms: "COVID," "corona," "coronavirus," "pandemic," "quarantine," "virus," "lockdown," and "isolation." For inclusion, OPs had to be written in English and posted no earlier than January 31, 2020, when the United States declared the novel coronavirus a public health emergency. KS conducted searches from June 27, 2020, to July 3, 2020. We captured a range of reactions to COVID-19, as our data collection window included many other important dates, including the virus being named COVID-19 on February 11, Italy's lockdown beginning on February 23, the first community spread case in the United States on February 26, the first report of a US death on February 29, and the declaration of a pandemic on March 11. Figure 1 presents a timeline of COVID-19 pandemic events, data collection, and date range of OPs. Posts had to relate to both e-cigarette use and the COVID-19 pandemic. Because posts were made in e-cigarette subreddits, mentions of "vaping" were assumed to be related to nicotine vaping unless there was an explicit mention of cannabis vaping and no mention of nicotine vaping. To capture interactions between Redditors, eligible posts had to have at least 1 comment from another Redditor. We excluded OPs that mentioned cannabis vaping and made no reference to nicotine vaping or e-cigarettes.

Figure 1. A timeline of events related to the COVID-19 pandemic in early 2020. We collected Reddit data from June 27, 2020, to July 3, 2020. The final dataset included original posts made between February 19, 2020, and June 14, 2020. WHO: World Health Organization.

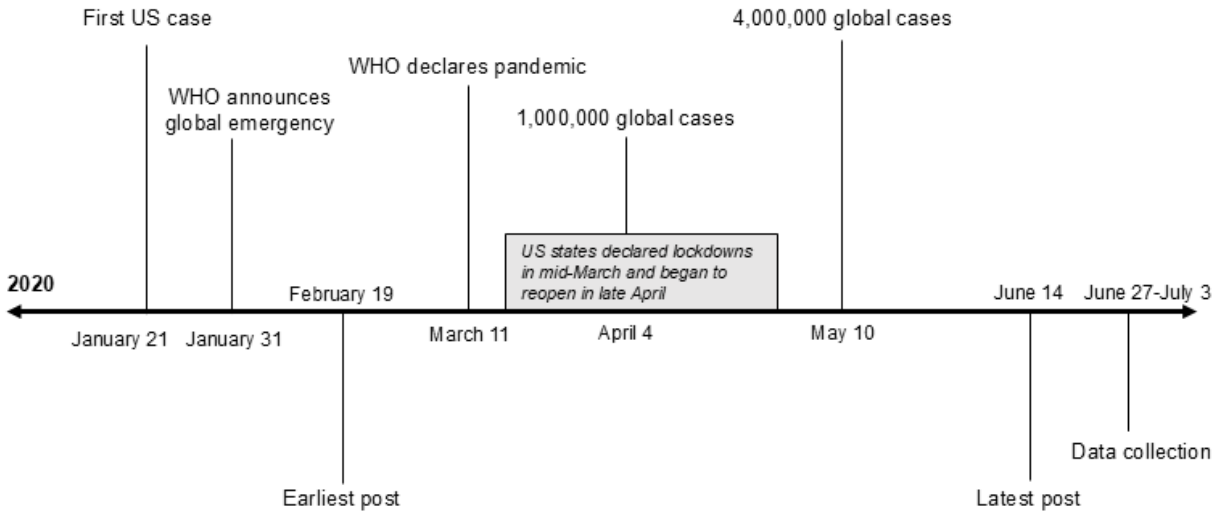

For each search, we screened OPs randomly (to decrease algorithmic bias where newer or more popular posts appeared first) until we collected 10 relevant posts and all comments (ie, threads) or reviewed all search results. KS saved 189 conversation threads (PDF) and recorded search metadata, including search terms, number of hits, number of ineligible posts removed before hitting 10 relevant posts, search date, and qualitative observations about the search (eg, June 27 entry for search term *coronavirus*, "I'm noticing individuals like to leave articles with little to no commentary and then ask for others' interpretations. In the comments, a robust discussion often occurs. It is interesting to see individuals using others' interpretations/community discussion."). When saved, OPs had an average of 17 comments (range 1-167; total: n=3155), an average of 25 "points" (assigned by Reddit), and were 81% upvoted on average. Approximately 1200 unique usernames engaged in conversation. If a Redditor deletes their account, their comment remains but their username does not. This occurred infrequently in our dataset. Still, we could not determine how many additional Redditors participated.

Data Analysis

We conducted a qualitative thematic analysis to (1) contextualize e-cigarette–related behavior changes during the early COVID-19 pandemic and (2) characterize web-based discussions about e-cigarettes during that time. After the first complete reading of the data, KS and ES led codebook development using an iterative process of reading, note-taking, and discussion to identify recurrent ideas and come to an agreement on codes and definitions. They developed an initial codebook of inductive codes, definitions, and examples; tested the codebook; and refined it to ensure comprehensiveness and clarity between codes. SLW provided suggestions and mentorship throughout the process. The final codebook included 7 codes related to e-cigarette perceptions and behaviors and 3 codes related to web-based community interactions. Finally, we used a 3-tiered quality code of "priority," "average," and "low priority" to

XSL•FO
RenderX

indicate richness (a qualitative determination based on the number, length, and detail of posts) of the thread's content.

KS completed coding and led the analysis. We applied codes to threads in their entirety to preserve the interrelated, unique interactions between Redditors within each threaded conversation, and codes were assigned based on relevant content in the OP, comments, or both. We assigned multiple codes to a single thread when relevant and used Microsoft Excel to track coding.

For each code, KS combined all relevant threads into a single document (PDF, then converted to a Microsoft Word document) and performed data reduction, a process that involved reading each post in its entirety, summarizing relevant content, saving illustrative quotes, and keeping analytic notes to document emerging ideas and interpretations. For each code, we used summaries, illustrative quotes, and analytic notes to identify patterns. *Priority* threads were analyzed first, and *average* threads were read if posts continued to add unique information to our analysis (ie, saturation was not yet met). KS and SLW used the collaborative web-based whiteboarding application Miro to organize cases and analytic ideas, returning to original texts to test emerging themes. During this analytic process, we used the socioecological model—which posits that individual behavior is influenced by their social, physical, and cultural environment—as an organizing framework to tease apart the external and internal factors Redditors mentioned. This iterative process produced the final model and thematic findings presented subsequently.

Ethical Considerations

This study was determined not to be human participants research by the University of Iowa Institutional Review Board (project title "202006140 e-cigarette conversations in online social media platforms"). All posts in our dataset were accessible on the web and might be considered publicly available data. Furthermore, Redditors select usernames, which can conceal their identity. However, they did not formally consent to the inclusion of their comments in research and might not have conceptualized their posts as truly "public." To avoid the reidentification of contributors, we used only short quotes in our examples, did not name the subreddits we collected data from, and did not identify Redditors by their usernames. We did not engage with Redditors for this study. Documents were stored on shared cloud storage through the University of Iowa and accessible only to team members.

Results

Overview

Reddit served as an information hub, debate platform, public journal, and a place of emotional support for people who used e-cigarettes early in the COVID-19 pandemic. Our analysis reflected the multifaceted experiences of people who use e-cigarettes, represented in a model of the environmental and individual influences on behavior change discussed by Redditors (Figure 2). Our findings indicate that circumstances created by the pandemic compelled behavior changes as they affected the context, pace, and environment in which individuals experienced life. Redditors drew from their existing perceptions, intentions, and experiences with e-cigarettes and cigarettes; their pandemic experience; and their participation on Reddit to decide whether and how to change their e-cigarette behaviors. Consumption of media and information (eg, peer-reviewed publications) and associated commentary shared on Reddit played a role in shaping behavior. Posts revealed how decisions were complicated by media distrust and by tobacco regulatory context.

Figure 2. A model of environmental and individual influences on 5 categories of behavior, developed through qualitative document analysis, visualizes the complexity of the pandemic's downstream impact. Changes in circumstances during the early COVID-19 pandemic compelled behavior changes (A) and changed perceptions and intentions (B), which in some instances also led to behavior change (C). Information and misinformation in the media and social media environment (D), especially information about the risks of COVID-19 for people who use e-cigarettes and tobacco, indirectly influenced behavior change through changes in perceptions of risk. Participation on Reddit itself exposed Redditors to information and advice of varying quality, which in turn might have impacted product perceptions and use practices. Findings are based on a qualitative analysis of 189 discussion threads from Reddit, posted between February 19, 2020, and June 14, 2020, and collected from June 27, 2020, to July 3, 2020.

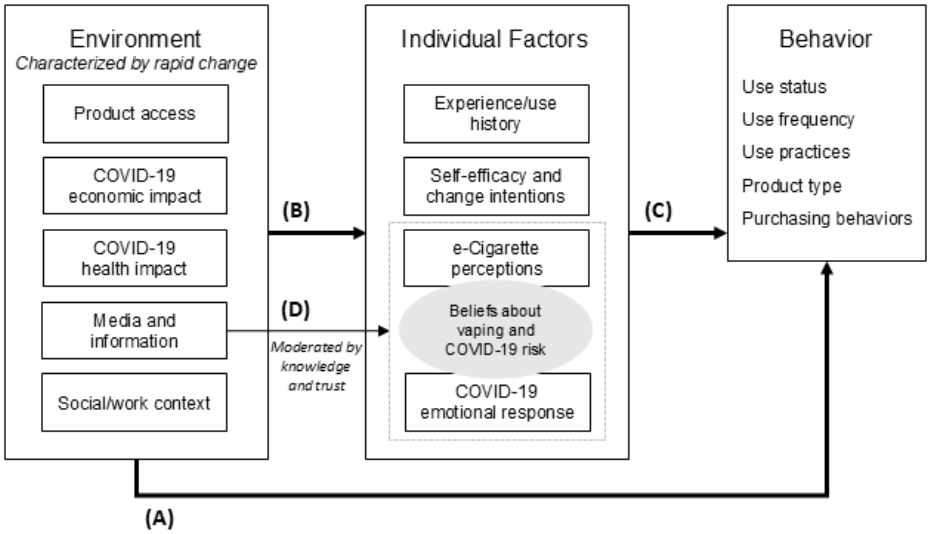

XSL•FO
RenderX

Reddit as a Source of Support, Information, Debate, and Influence

Rapidly evolving public health information about COVID-19 created uncertainty, stress, and confusion. The subreddits acted as community spaces for individuals seeking to express these emotions and receive advice and support. Subreddit posts during the pandemic included solicitation and sharing of e-cigarette knowledge. For example, one Redditor who could not afford to replace their broken device sought expertise from other Redditors, stating "All advice is appreciated." Suggestions included power cycling (ie, turning off and turning on again), ensuring the coil of the device was not damaged, and checking the charging pins. Other OPs asked subjective questions, such as advice on whether they should quit vaping because of the pandemic. One Redditor who reported having asthma and being a "pretty heavy vaper" explained they feared a more severe illness but a lack of "solid evidence" about vaping made them unsure about what they should do, so they asked others for their opinions.

Reddit also served as a space for discussion of information from other media, such as news articles, research publications, and blog posts related to the COVID-19 pandemic. The discourse about media seemed to change opinions for some Redditors, while for others it confirmed their existing beliefs. For example, when a commenter questioned a study's validity because it was published in the 1940s, the OP acknowledged the observation was a "good call." Another Redditor described propylene glycol (PG) and vegetable glycerin in detail, inaccurately claiming that vegetable glycerin had antimicrobial and antiviral properties, to which another Redditor expressed gratitude for the information.

Redditors shared anecdotes and stories, and other Redditors used these stories to form ideas about COVID-19. In response to an OP wondering whether they should quit vaping during the pandemic, one Redditor opined that there is "no evidence" of a relationship between COVID-19 and vaping because their whole family contracted COVID-19 except for them, and they were the only one in their family who vaped. The anecdote was used to support the OP's decision to continue vaping.

Comments providing emotional support or sharing experiences in solidarity reflected a sense of community. For example, one OP experiencing nicotine withdrawal asked whether their symptoms were normal, and a commenter affirmed their experience and assured them they would "wake up more and more happy" over time. Another Redditor received 6 recommendations after requesting e-cigarette device options for their mother who had lost her husband to COVID-19 and "desperately want[ed] to quit [cigarettes]." Posts reflected solidarity with people who were using e-cigarettes as a harm reduction strategy, encouraging people to avoid relapse (eg, "stay strong and don't do it") or advocating for vape shops to remain open to support the people who "rely on vape shops to make sure they don't return to cigarettes" and the many people who are "one cigarette away from giving up on vaping." Many Redditors disagreed, arguing their closure was important for COVID-19 prevention.

Behavior Changes and Their Contexts

Overview

Thematic analysis revealed how the early COVID-19 pandemic directly and indirectly motivated changes in e-cigarette use and e-cigarette practices. We found no consistent pattern of change; some Redditors reported increasing use, and others reported decreasing or pausing use (ie, at least temporary cessation). Synthesizing our thematic analysis, Figure 2 visualizes pandemic-related behavior changes as influenced by environmental factors (ie, physical, economic, and legal environment that constrained individual behavior) and personal characteristics (ie, individual perceptions and experiences with e-cigarettes), which could be modified during the pandemic by personal experiences and exposure to information and misinformation, including Reddit participation. Posts indicated that the impact of environmental context, consumed media, and personal experience depended on prepandemic risk perceptions and behaviors.

Direct Impacts of the COVID-19 Pandemic on e-Cigarette Behaviors

Redditors explicitly and implicitly referred to life changes during the COVID-19 pandemic, including working from home, job loss and financial insecurity, exposure to COVID-19, and reduced access to e-cigarettes. These changes in circumstances led to changes in behavior. For example, one Redditor reported having fewer hours of work and therefore more time to play video games. Their already-established practice of vaping while playing video games meant they increased their e-cigarette use when their gameplay increased. A Redditor who previously smoked cigarettes and lived in a state with a flavored product ban lost access to their usual flavored e-cigarette products when they stopped commuting to work across state lines. Because of this loss of access, they were considering reinitiating cigarette smoking. One Redditor who used e-cigarettes socially and could not spend time with friends because of the COVID-19 pandemic social distancing restrictions reported they "stop[ped] vaping for a while," although they implied that they might start again. In contrast, a few Redditors who worked in essential services or already worked from home cited their lack of job change as the reason they had not changed their vaping intensity.

COVID-19 itself prompted behavior change, largely temporary reduction or cessation, for a few Redditors who reported COVID-19 symptoms, such as shortness of breath, a cough, or fatigue that made vaping difficult or impossible. One Redditor with a "terrible cough" switched to nicotine patches, and another decreased use when COVID-19–like symptoms made vapor inhalation difficult. Some posts revealed potentially harmful practices, like one Redditor who diluted e-liquid with water and consumed it orally while they had a cough and suggested that other people who were sick try the same.

Several posts reflected e-cigarette changes in response to other effects of COVID-19, such as the loss of a loved one or high test positivity rates, including one account of increased vaping while coping with the loss of their grandfather to COVID-19.

The economic impact of the early pandemic left some Redditors with less discretionary income to spend on e-cigarettes. For

XSL•FO
RenderX

example, one Redditor shared that they had not worked in a week, ran out of e-liquid, and had < US $1 left in their bank account. Because they could no longer afford e-cigarette products, they "resorted" to their fiancé's cigarettes, sharing their disgust with the emojis " ."

Purchasing patterns adapted to changing circumstances; vape shop closures, curbside pickups, web-based purchases, limited supplies, and shipping delays altered purchasing behaviors. Some individuals began "stockpiling" e-cigarette devices and accessories because of real or anticipated shortages (Figure 3). In response to observed or anticipated reduction in access, some Redditors altered their use and personal practices, including reducing volume (eg, abstinence when unable to buy their favorite e-liquid) and changing products (eg, turning to pipe tobacco after running out of pods and e-liquid because they were "still addicted"). Redditors reported "DIY" activities in the face of shortages, for example mixing their own e-liquid from a bulk purchase.

Figure 3. e-Cigarette stockpile shared on Reddit.



Redditors referenced potentially risky adaptations, such as using other wires in place of standard product coils, attempting their own device repairs, or trying to clean disposable pods to reuse. Replies to these posts included encouragement, suggestions, and discouragement. One Redditor who realized they bought expired e-liquid still considered vaping it because it was their only option. Another shared that their local shop was selling 3-year expired juice with fake stickers to look fresh; some responses suggested it might be usable, but most responses strongly discouraged this behavior and some suggested reporting the retailer.

A Redditor shared an image of their stockpile after going to their local vape shop. The owner of the store was in the process of closing for the pandemic, but the Redditor shared they felt prepared (because of their stockpile), asserting "Bring on the lockdown."

Reddit discussions reflected an array of e-cigarette–related behavioral responses to the COVID-19 pandemic and infodemic, including increases and decreases in use intensity, changes in purchasing practices (eg, stockpiling), and changes in vaping practices (eg, reusing disposable pods). These findings are based on a qualitative analysis of 189 discussion threads from Reddit, posted between February 19, 2020, and June 14, 2020, and collected from June 27, 2020, to July 3, 2020.

People who used e-cigarettes were impacted by both the pandemic response and the existing e-cigarette regulatory scheme in their locale. A few Redditors made references to changing e-cigarette policy or stay-at-home orders, illustrating this impact. For example, some individuals had already stockpiled devices, e-liquid, or accessories in response to regulatory changes, and thus were not as severely impacted by the pandemic.

Impacts of Individual Perceptions and e-Cigarette or Cigarette History on e-Cigarette Behaviors

Two key beliefs anchored risk perception during the pandemic: (1) COVID-19 was a serious illness and (2) e-cigarette use increased COVID-19–related risk. The belief that e-cigarette use made COVID-19 more dangerous (either in disease susceptibility or severity) sufficiently increased the perceived risk of vaping to prompt behavior change for some Redditors. For example, one Redditor stopped vaping entirely because of an underlying belief that vaping was not "exactly 100% safe and healthy" and a desire for their body not to "suffer unnecessarily" if they contracted COVID-19. Others reduced their e-cigarette use "just in case" using e-cigarettes increased COVID-19 risk or severity. One Redditor who had COVID-19 shared that they felt their vaping did not make their symptoms worse but still took a 2-week hiatus when their symptoms were severe. Alternatively, Redditors who did not perceive both COVID-19 and e-cigarettes as risky were unlikely to alter their behavior. Some Redditors perceived COVID-19 as a nonserious or nonexistent condition. Others believed they had negligible exposure and were therefore unconcerned about contracting COVID-19. This lack of concern meant no change in behavior for one Redditor who asserted they would continue to vape

XSL•FO
RenderX

alone at home. While washing their hands. And not touching their face (punctuation in the original comment).

Perceived addiction, previous cigarette smoking, existing intentions to quit, and self-efficacy to change behavior help explain the extent to which people who used e-cigarettes changed their behavior in response to the pandemic. One Redditor's post reflected a tension between underlying intentions to quit and low self-efficacy, describing the pandemic as "the perfect reason" to fulfill their quit intentions but admitting that it has been hard due to the stress they had experienced. Another Redditor who read a report that vaping increased the severity of COVID-19 decided that it would be "way too difficult…[to quit] during a stressful time," even after the study made them believe quitting might be beneficial. Product shortages left one Redditor who had previously smoked with what he saw as 2 unappealing options: "go cold turkey" or "return to cigarettes."

e-Cigarette– and COVID-19–Related Beliefs and Their Development

The complex environment of COVID-19 and e-cigarette information during this time resulted in an array of interpretations of the health risks that COVID-19 posed to people who used e-cigarettes. Redditors offered multiple explanations for why e-cigarette use disproportionately increased the health risks for people who used e-cigarettes, including COVID-19 and vaping impacting the same bodily systems, the vape device itself acting as a transmission vector, or nicotine harmfully impacting the angiotensin converting enzyme 2 (ACE2; a protein) receptor system. For example, one Redditor expressed concern with vaping during the pandemic, stating nicotine may increase the number of ACE2 receptors, which are the "[receptors] this specific virus hijacks to get into cells." However, some Redditors suggested vaping may be beneficial for fighting or preventing COVID-19. One Redditor described how studies show inhaling PG can "prevent you from getting the flu" and argued PG has "incredible antimicrobial effects," reporting they would continue vaping. Another said "nicotine [and PG] kill bacteria," concluding that vaping was beneficial. Others made claims without proposing a mechanism, such as "vaping...help[s] against pneumonia."

Some posts reflected the salience of former smoking on e-cigarette– and COVID-19–related beliefs. For example, in response to an OP who asked whether concern about the risks of vaping and COVID-19 was causing people to stop vaping, several Redditors commented on how vaping "makes a huge difference" (compared to smoking cigarettes) on the severity or frequency of respiratory infections, implying that the relative wellbeing is evidence that vaping does not increase COVID-19 risks.

Reddit threads served as a forum to discuss information about the COVID-19 pandemic, and Redditors cited information from a variety of sources, including public health authorities and scientific organizations, traditional media sources, such as newspapers and televised broadcasts, and informal news sources, such as blog and social media posts. The thread mentioned earlier, which included the OP asking about the risk of vaping (submitted on March 18, 2020) and 120 comments, offers an interesting case of how Redditors pieced together the range of questions about vaping-related risks. The tone and quality of the debate varied across comments, which included personal anecdotes, scientific claims, and commentary on the quality of the discussion. Many posts included unsubstantiated claims. Other comments supported claims with a range of sources, including articles from PubMed, a publicly accessible database of scientific and medical research funded by the US federal government. Some comments asked for evidence to support others' claims, offered corrections, and identified logical fallacies (eg, noting the claim that vaping was better for respiratory health than smoking did not address the question of whether, in the absence of smoking, vaping was better than not). One post asked the subreddit to avoid circulating opinions and stay very close to high-quality sources of information to protect public safety, and a comment on another thread noted to the OP that the subreddit was the not the best place to ask about perceptions about vaping and covid risk, "seeing as everyone is going to have a biased opinion."

As we have documented, Redditors credited beliefs about whether and how COVID-19 and e-cigarette risk interacted for driving their behavior change. In some instances, it was apparent that the conversations in subreddits influenced the development of these beliefs. Some conversations suggested that the media individuals consumed had a profound impact on their perceptions, thoughts, and behaviors during this time. Perhaps most colorfully, one Redditor commented, "HOLY FUCK. There is so much contradictory evidence. I'm like a sheep. I read one comment and put my [vape] away. I read another and pick it up! I just need to stop using this shit."

We found evidence that individual factors moderated one's interpretation and understanding of Reddit and media content (Figure 2). Perceived contradictions between sources, a Redditor's level of trust in each source, assumptions, and understanding of the scientific process altered how Redditors weighed new information. For some Redditors unfamiliar with the research process, distrust emerged when recommendations for pandemic safety changed as additional research was published.

The underlying trust of an information source influenced a Redditor's assessment of risk regarding COVID-19, vaping, and their intersection. When an OP asked whether they should vape less during the pandemic, a commenter said no, stating "all the bs" in the media is grounded in "zero actual science." Another Redditor complained that officials "refer to all nicotine intake as 'smoking'" and emerging studies regarding ACE2 receptors and smoking show "how incompetent and ignorant" officials are in the United States. These generalizations across nicotine delivery methods and seemingly conflicting scientific evidence likely deepened the distrust of the media for this individual.

Discussion

Principal Findings

Our qualitative thematic analysis of e-cigarette–related conversations on Reddit during the early months of the COVID-19 pandemic revealed changes in amount, frequency,

XSL•FO
RenderX

product type, purchasing behaviors, and use practices from the widespread effects of the COVID-19 pandemic. By examining the "nuanced deliberations" [40] in social media communities of people who use e-cigarettes, our study illuminates how the variety in behavioral responses during the pandemic evolved in response to socioecological changes, individual nicotine and tobacco product use history, perceptions and beliefs, the information environment, and the interactions between these factors.

Our finding of mixed behavioral changes from the pandemic's impact on everyday life is consistent with other reports about e-cigarettes [24,44], and we hope our model can help piece together the growing work on the contextual and individual factors [45] to explain *why* behavior change was so varied. On the one hand, Reddit conversations suggested that reduced access to e-cigarette products, salient health messages and concerns [46,47] (although Klemperer et al [48] report mixed findings), and changes in circumstance encouraged cutting back and quitting. On the other hand, stress, boredom, unstable routines, and changes in activity space (eg, working from home [49]) encouraged more nicotine use [50] or deterred cessation. Other findings suggest that for individuals trying to quit, pandemic-related stress might have undermined coping skills and therefore suppressed cessation [51]. Studies on alcohol consumption [52] and cigarette smoking [12] suggest variations in the types and sources of stress may explain some of the divergence in e-cigarette use behavior. Our observation that some Redditors were concerned about product access and stockpiled vaping materials early in the pandemic is consistent with previous reports about e-cigarettes [53] and resembled stockpiling of essential goods (eg, meat and toilet paper). The similarity suggests that these individuals considered e-cigarettes an essential good.

Our research builds upon previous work that identifies social media as a key space for individuals to seek health-related advice, information, and community, demonstrating in the context of vaping during the pandemic how the community-building and knowledge-sharing functions of social media can affect people's health behavior. In another Reddit study, Sharma et al [40] found that people with mental illness perceived that web-based discussion about the shared interest in vaping facilitated social connectedness for them. We speculate that during the early months of the COVID-19 pandemic, when other socializing may have been scarce and access to substances was limited, Reddit and other social media platforms were particularly important for connectedness and support among individuals who use substances [54-56]. Still, using social media has also been associated with cigarette and e-cigarette use [29,57,58], and an intervention during the first year of the pandemic reported that a 30-minute reduction in daily social media use decreased smoking behavior [59].

Misinformation can easily circulate on the web because the credibility of information on social media cannot be certified, and social media algorithms create "echo chambers." The Online Misinformation Engagement Framework identifies four stages in which people engage with web-based misinformation as follows: (1) selecting information sources, (2) choosing to consume or ignore the information they are exposed to, (3) evaluating that information (ie, evaluation), and (4) engaging with the information (ie, reaction). In times of scientific and social uncertainty, when public health messaging is limited and competing with misinformation, people might turn to social media communities seeking advice, information, and strategies from people they perceive as trustworthy (ie, the pandemic might have impacted stage 1: source selection) [1]. Our findings suggest Reddit served as a trusted source of information about e-cigarettes for community members during the early pandemic. Previous work has documented social media conversations about the intersection of vaping and COVID-19, including health concerns and unsubstantiated claims (Twitter [60] and YouTube [32]). In the discussion we observed, some Redditors critically evaluated the claims (stage 3: evaluation) and engaged in peer-to-peer debunking (stage 4: reaction); however, we do not know the extent to which all viewers engaged in these behaviors and whether these actions impacted perceptions of those reading.

There remains no definitive information about the intersection of COVID-19 and vaping risk [36]. Some claims in our dataset do appear to be supported by scientific evidence [36]. In addition, it is likely that Redditors were sincere in their discussion, as evidence suggests that people rarely share misinformation intentionally. Still, most posts engaged with just one or a few dimensions of this complicated question, and it appears individuals made decisions based on incomplete information. Cognitive bias might also be at play. A UK study reported that adults who used e-cigarettes were more likely to assert that using e-cigarettes decreased or did not affect the risk of severe COVID-19 symptoms [34], although empirical evidence of the effects was uncertain. The conversations we analyzed likely reflected and potentially perpetuated this optimism bias. In addition, misinformation (including incomplete information) might spread more easily in vaping subreddits, which engage a relatively homogenous network than a more general subreddit with more diverse participants [61].

We posit that Reddit conversations impacted behavior both directly (eg, providing advice for vape repair) and indirectly by informing beliefs (including both accurate and inaccurate perceptions). Consistent with our finding that information on Reddit influenced decision-making for some Redditors, a study of college students found that exposure to and seeking COVID-19–related news was related to lower odds of pausing smoking or vaping [62]. We speculate that the sometimes heated discussion of information, including misinformation, in these subreddits might have increased some Redditors' already heightened pandemic-related stress and perhaps contributed to vaping escalation or avoided reduction in vaping intensity. Moreover, exposure to inaccurate information might have a lasting impact; for example, Redditors who vaped more believing that it would decrease their odds of contracting COVID-19 would inadvertently increase their e-cigarette–related health risks. An observational study of people who had tried to quit using e-cigarettes reported higher relapse among people who recalled at least one misinformation claim about COVID-19 and nicotine [63]. Other information might have also increased the risk. Redditors who followed suggestions on how to extend the life of their e-cigarette devices (eg, making their own coil) might have been exposed to more toxic chemicals. Studies on

XSL•FO
RenderX

the health effects of improper use of e-cigarette devices can inform our understanding of the extent to which the pandemic not only influenced use frequency but changed the risk profile of e-cigarette use.

Our study demonstrates the value of qualitative work in behavioral research, as it illuminates the nuance of how often, how much, or in what way people use vaping products. Our qualitative analysis elucidates the array of socioecological and personal considerations Redditors had at the time. This array of experiences might help explain conflicting findings between quantitative studies with different samples [17,64]. It also suggests that, because summary statistics obscure bidirectional impacts, cross-sectional analyses will likely underestimate the extent to which e-cigarette practices were impacted during the crisis. Longitudinal, quantitative surveillance will determine how widespread and long-lasting the pandemic's impacts were on vaping at a population level. Qualitative work can contextualize those impacts. In addition, while quantitizing social media data can reflect overall discussion topics [43], qualitative analysis can reveal the community dynamics, such as the role of personal anecdotes on risk perceptions, that play out in those discussions.

The lack of unified messaging about e-cigarettes reflected on Reddit is not a new problem and follows a decade of conflicting messages. The e-cigarette industry has advertised e-cigarettes as a cigarette cessation device. Scientific evidence is mixed on their effectiveness for cessation and relative health benefits. The outbreak of e-cigarette, or vaping, product use–associated lung injury resulted in a growing but unsubstantiated perception that e-cigarettes may be more dangerous than combustible cigarettes [65]. There is a need for consistent, easy-to-understand messaging to fight misinformation on the web about COVID-19 and the risks and practices associated with e-cigarettes. Applying the tools of infodemic management might improve vaping-related communication [3]. Although an experimental study reported that US adults perceived messages that linked smoking to COVID-19 as effective for decreasing vaping [66], our findings suggest that messages and evidence specific to e-cigarette use rather than tobacco products overall would be viewed as more trustworthy and acceptable. Leveraging many communication channels can be an effective way to reach target audiences with antivaping campaigns [67], and Reddit itself might be a potential avenue for health communication interventions [68].

Moreover, e-cigarette interventions might address underlying causes of substance use, misinformation, and distrust. Much of our data revealed the impact of stress, coping, mental health struggles, and feelings of stigmatization, including from public health officials. We expect cessation messaging that acknowledges the experiences and struggles of individuals who use e-cigarettes and the meaning that e-cigarettes have in their lives (eg, stress relief) might better resonate with this population. This and other work highlight the role of community dynamics of social media in information exchange; health communication might leverage these relationships to disseminate accurate health information. Structural- and individual-level misinformation interventions across stages of engagement (eg, content labels, introducing friction that encourages the viewer to pause before sharing information) will also likely improve the quality of decision-making [69] about vaping.

Limitations

Our analysis aimed to document e-cigarette behavior change and reveal mechanisms through which the early pandemic and related infodemic impacted behavior; our findings may be an incomplete list or overemphasize the experiences of more active Redditors. Moreover, many individuals may not have changed their e-cigarette use behavior at all and the behavior changes that we report may have been only temporary.

The unobtrusive methods that allowed us to observe Reddit conversations were limited in several ways. Because we did not communicate with Redditors, our ability to contextualize narratives in the specific national and local policy and social norms was restricted to the information Redditors disclosed, we are unable to determine how long changes lasted, and we cannot describe our sample's demographic characteristics or baseline use behaviors. Redditors are predominantly White and male; therefore, it is likely that our study overemphasized their experiences [70,71]. Pandemic-related stressors might have intersected with and compounded experiences of minority groups' stress [72,73] and resulted in increased vaping [74-76]; by contrast, resilience, optimism, and sense of community might have buffered some individuals with minoritized identities from that stress [77,78]. Work that investigates the digital experiences of minoritized groups during the pandemic is essential [79]. Our team is composed of female researchers in the United States; therefore, we might have missed cultural cues outside of our experience or expertise. We have not documented how perceptions and behaviors have evolved since June 2020 as new information emerged, policies changed, and more recent waves of the pandemic occurred. Although the specific perceptions about the relationship between vaping and COVID-19 might not be transferable today, we believe our findings speak to broader questions about the role of social media in information exchange during a crisis and the relationship between people who use e-cigarettes and public health.

Conclusions

This study elucidated how people make decisions about their health during a time of informational and social uncertainty. Information is gathered from a variety of sources, increasingly from social media platforms like Reddit, making public health research on these platforms critical. We were able to gain insight into the way people processed and discussed information and identify individual and environmental changes (eg, local laws and supply chain) that provide an explanation for why there was no unified direction of e-cigarette behavior change during the early pandemic. Social media is an effective and important place to communicate public health information, particularly during crisis or disaster situations, and it is crucial that e-cigarette information is accurate, transparent, accessible, nonstigmatizing, and specific to the target population.

XSL•FO
RenderX

Acknowledgments

The authors express gratitude to Minji Kim, Alyssa Bercasio, and Kamara Shaw for their help, support, and feedback on this work and to the Redditors for sharing their insights and experiences during this unprecedented time. This work was supported by the American Cancer Society (IRG 18-165-43), administered through the Holden Comprehensive Cancer Center at the University of Iowa, the University of Iowa Office of Undergraduate Research, and the University of Iowa College of Public Health. KS conducted this work as an undergraduate student at the University of Iowa. ES and LF completed this work as Master's of Public Health students at the University of Iowa.

Data Availability

The dataset generated and analyzed during this study is not publicly available; although Reddit posts are publicly available and anonymized by username, Redditors did not consent for their content to be shared publicly in a curated dataset. The dataset is available from the corresponding author on reasonable request for research purposes.

Authors' Contributions

SLW (principal investigator) contributed to conceptualization; methodology; investigation; validation; writing the original draft; writing, reviewing, and editing the manuscript; supervision; and funding acquisition. KS contributed to conceptualization; methodology; investigation; data curation; writing the original draft; writing, reviewing, and editing the manuscript; visualization; and funding acquisition. ES contributed to the methodology and writing, reviewing, and editing the manuscript. LF contributed to investigation; writing the original draft; and writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

References

1. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
2. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
3. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
4. Sauer MA, Truelove S, Gerste AK, Limaye RJ. A failure to communicate? How public messaging has strained the COVID-19 response in the United States. Health Secur 2021 Feb 01;19(1):65-74 [FREE Full text] [doi: 10.1089/hs.2020.0190] [Medline: 33606575]
5. Benham JL, Lang R, Kovacs Burns K, MacKean G, Léveillé T, McCormack B, et al. Attitudes, current behaviours and barriers to public health measures that reduce COVID-19 transmission: a qualitative study to inform public health messaging. PLoS One 2021;16(2):e0246941. [doi: 10.1371/journal.pone.0246941] [Medline: 33606782]
6. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021 Jun 11;23(6):e29802. [doi: 10.2196/29802] [Medline: 34043526]
7. Ruokolainen H, Widén G, Eskola EL. How and why does official information become misinformation? A typology of official misinformation. Libr Inf Sci Res 2023 Apr;45(2):101237. [doi: 10.1016/j.lisr.2023.101237]
8. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
9. Schultz F, Utz S, Göritz A. Is the medium the message? Perceptions of and reactions to crisis communication via Twitter, blogs and traditional media. Public Relat Rev 2011 Mar;37(1):20-27. [doi: 10.1016/j.pubrev.2010.12.001]
10. Papakyriakopoulos O, Medina Serrano JC, Hegelich S. The spread of COVID-19 conspiracy theories on social media and the effect of content moderation. Harv Kennedy Sch Misinformation Rev 2020 Aug 17:22. [doi: 10.37016/mr-2020-034]
11. Grossman ER, Benjamin-Neelon SE, Sonnenschein S. Alcohol consumption during the COVID-19 pandemic: a cross-sectional survey of US adults. Int J Environ Res Public Health 2020 Dec 09;17(24):1-10. [doi: 10.3390/ijerph17249189] [Medline: 33316978]
12. Koopmann A, Georgiadou E, Reinhard I, Müller A, Lemenager T, Kiefer F, et al. The effects of the lockdown during the COVID-19 pandemic on alcohol and tobacco consumption behavior in Germany. Eur Addict Res 2021 Apr 26;27(4):242-256. [doi: 10.1159/000515438] [Medline: 33902030]
13. Håkansson A. Changes in gambling behavior during the COVID-19 pandemic-a web survey study in Sweden. Int J Environ Res Public Health 2020 Jun 05;17(11):4013. [doi: 10.3390/ijerph17114013] [Medline: 32516880]

XSL•FO
RenderX

14. van Laar MW, Oomen PE, van Miltenburg CJA, Vercoulen E, Freeman TP, Hall WD. Cannabis and COVID-19: reasons for concern. Front Psychiatry 2020 Dec 21;11:601653. [doi: 10.3389/fpsyt.2020.601653] [Medline: 33408655]
15. Case KR, Clendennen SL, Shah J, Tsevat J, Harrell MB. Changes in marijuana and nicotine vaping perceptions and use behaviors among young adults since the COVID-19 pandemic: a qualitative study. Addict Behav Rep 2022 Jun;15:100408 [FREE Full text] [doi: 10.1016/j.abrep.2022.100408] [Medline: 35075434]
16. Yang H, Ma J. How the COVID-19 pandemic impacts tobacco addiction: changes in smoking behavior and associations with well-being. Addict Behav 2021 Aug;119:106917 [FREE Full text] [doi: 10.1016/j.addbeh.2021.106917] [Medline: 33862579]
17. Almeda N, Gómez-Gómez I. The impact of the COVID-19 pandemic on smoking consumption: a systematic review of longitudinal studies. Front Psychiatry 2022 Jul 12;13:941575. [doi: 10.3389/fpsyt.2022.941575] [Medline: 35903638]
18. Gallus S, Stival C, Carreras G, Gorini G, Amerio A, McKee M, et al. Use of electronic cigarettes and heated tobacco products during the COVID-19 pandemic. Sci Rep 2022 Jan 13;12(1):702 [FREE Full text] [doi: 10.1038/s41598-021-04438-7] [Medline: 35027590]
19. Layman HM, Thorisdottir IE, Halldorsdottir T, Sigfusdottir ID, Allegrante JP, Kristjansson AL. Substance use among youth during the COVID-19 pandemic: a systematic review. Curr Psychiatry Rep 2022 Jun 27;24(6):307-324 [FREE Full text] [doi: 10.1007/s11920-022-01338-z] [Medline: 35476186]
20. Wiley RC, Oliver AC, Snow MB, Bunn JY, Barrows AJ, Tidey JW, et al. The impact of the COVID-19 pandemic on smoking among vulnerable populations. Nicotine Tob Res 2023 Jan 05;25(2):282-290. [doi: 10.1093/ntr/ntac135] [Medline: 35605264]
21. Soule EK, Mayne S, Snipes W, Guy MC, Breland A, Fagan P. Impacts of COVID-19 on electronic cigarette purchasing, use and related behaviors. Int J Environ Res Public Health 2020 Sep 16;17(18):6762. [doi: 10.3390/ijerph17186762] [Medline: 32948084]
22. Kalkhoran SM, Levy DE, Rigotti NA. Smoking and e-cigarette use among U.S. adults during the COVID-19 pandemic. Am J Prev Med 2022 Mar;62(3):341-349. [doi: 10.1016/j.amepre.2021.08.018] [Medline: 34756629]
23. Denlinger-Apte R, Suerken CK, Ross JC, Reboussin BA, Spangler J, Wagoner KG, et al. Decreases in smoking and vaping during COVID-19 stay-at-home orders among a cohort of young adults in the United States. Prev Med 2022 Mar;156:106992. [doi: 10.1016/j.ypmed.2022.106992] [Medline: 35149114]
24. Barrington-Trimis JL, Klemperer EM, Majmundar A, Harrell M. The impact of the COVID-19 pandemic on nicotine use: what have we learned? Nicotine Tob Res 2023 Jan 05;25(2):175-176 [FREE Full text] [doi: 10.1093/ntr/ntac265] [Medline: 36478086]
25. Folayan MO, Alade OT, Sabbagh H, Oyapero A, Adeyemo YI, Popoola BO, et al. Factors associated with changes in e-cigarette use and tobacco smoking by adolescents and young people in Nigeria during the COVID-19 pandemic. Psychoactives 2023 Jan 12;2(1):23-36. [doi: 10.3390/psychoactives2010002]
26. Cheney MK, Gowin M, Clawson AH. Using the Ecological Model to understand influences on college student vaping. J Am Coll Health 2018 Oct 06;66(7):597-607. [doi: 10.1080/07448481.2018.1440578] [Medline: 29452051]
27. McDermott MS, East KA, Hitchman SC, McNeill A, Tountas Y, Demjén T, et al. Social norms for e-cigarettes and smoking: associations with initiation of e-cigarette use, intentions to quit smoking and quit attempts: findings from the EUREST-PLUS ITC Europe Surveys. Eur J Public Health 2020 Jul 01;30(Suppl_3):iii46-iii54. [doi: 10.1093/eurpub/ckaa014] [Medline: 32918818]
28. Giovenco DP, Casseus M, Duncan DT, Coups EJ, Lewis MJ, Delnevo CD. Association between electronic cigarette marketing near schools and e-cigarette use among youth. J Adolesc Health 2016 Dec;59(6):627-634. [doi: 10.1016/j.jadohealth.2016.08.007] [Medline: 27720358]
29. Vogel EA, Ramo DE, Rubinstein ML, Delucchi KL, Darrow SM, Costello C, et al. Effects of social media on adolescents' willingness and intention to use e-cigarettes: an experimental investigation. Nicotine Tob Res 2021 Mar 19;23(4):694-701 [FREE Full text] [doi: 10.1093/ntr/ntaa003] [Medline: 31912147]
30. D'Angelo H, Patel M, Rose SW. Convenience store access and e-cigarette advertising exposure is associated with future e-cigarette initiation among tobacco-naïve youth in the PATH study (2013-2016). J Adolesc Health 2021 Apr;68(4):794-800 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.08.030] [Medline: 33039271]
31. Papini S, López-Castro T, Swarbrick M, Paul LK, Stanley D, Bauer A, COVID-Dynamic Team, et al. Alcohol, cannabis, and nicotine use have distinct associations with COVID-19 pandemic-related experiences: an exploratory Bayesian network analysis across two timepoints. Drug Alcohol Depend 2023 Jul 01;248:109929 [FREE Full text] [doi: 10.1016/j.drugalcdep.2023.109929] [Medline: 37267744]
32. Lee J, Ouellette RR, Murthy D, Kong G. Content analysis of YouTube videos related to e-cigarettes and COVID-19. Subst Use Misuse 2024 Oct 05;59(1):143-149. [doi: 10.1080/10826084.2023.2262027] [Medline: 37798867]
33. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]

XSL•FO
RenderX

34. Herbec A, Brown J, Jackson SE, Kale D, Zatoński M, Garnett C, et al. Perceived risk factors for severe COVID-19 symptoms and their association with health behaviours: findings from the HEBECO study. Acta Psychol (Amst) 2022 Feb;222:103458 [FREE Full text] [doi: 10.1016/j.actpsy.2021.103458] [Medline: 34933210]
35. Phandthong R, Wong M, Song A, Martinez T, Talbot P. Does vaping increase the likelihood of SARS-CoV-2 infection? Paradoxically yes and no. Am J Physiol Lung Cell Mol Physiol 2024 Feb 01;326(2):L175-L189 [FREE Full text] [doi: 10.1152/ajplung.00300.2022] [Medline: 38147795]
36. Benowitz NL, Goniewicz ML, Halpern-Felsher B, Krishnan-Sarin S, Ling PM, O'Connor RJ, et al. Tobacco product use and the risks of SARS-CoV-2 infection and COVID-19: current understanding and recommendations for future research. Lancet Respir Med 2022 Sep;10(9):900-915 [FREE Full text] [doi: 10.1016/S2213-2600(22)00182-5] [Medline: 35985357]
37. Kamiński M, Muth A, Bogdański P. Smoking, vaping, and tobacco industry during COVID-19 pandemic: Twitter data analysis. Cyberpsychol Behav Soc Netw 2020 Dec 30;23(12):811-817. [doi: 10.1089/cyber.2020.0384] [Medline: 32757951]
38. United States Securities and Exchange Commission form S-1. Reddit Inc. URL: https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1q423.htm#i1b9a579e78a34dfa99f7f26daeec195b_88 [accessed 2024-04-29]
39. Struik L, Yang Y. e-Cigarette cessation: content analysis of a quit vaping community on Reddit. J Med Internet Res 2021 Oct 25;23(10):e28303 [FREE Full text] [doi: 10.2196/28303] [Medline: 34694229]
40. Sharma R, Wigginton B, Meurk C, Ford P, Gartner CE. Motivations and limitations associated with vaping among people with mental illness: a qualitative analysis of reddit discussions. Int J Environ Res Public Health 2016 Dec 22;14(1):e256 [FREE Full text] [doi: 10.3390/ijerph14010007] [Medline: 28025516]
41. Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage JUUL use patterns: content analysis of reddit messages. J Med Internet Res 2019 Sep 09;21(9):e13038 [FREE Full text] [doi: 10.2196/13038] [Medline: 31502542]
42. Wu D, Kasson E, Singh AK, Ren Y, Kaiser N, Huang M, et al. Topics and sentiment surrounding vaping on Twitter and Reddit during the 2019 e-cigarette and vaping use-associated lung injury outbreak: comparative study. J Med Internet Res 2022 Dec 13;24(12):e39460 [FREE Full text] [doi: 10.2196/39460] [Medline: 36512403]
43. Wanchoo K, Abrams M, Merchant RM, Ungar L, Guntuku SC. Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19. PLoS One 2023 Feb 3;18(2):e0280337 [FREE Full text] [doi: 10.1371/journal.pone.0280337] [Medline: 36735708]
44. Bennett M, Speer J, Taylor N, Alexander T. Changes in e-cigarette use among youth and young adults during the COVID-19 pandemic: insights into risk perceptions and reasons for changing use behavior. Nicotine Tob Res 2023 Jan 05;25(2):350-355 [FREE Full text] [doi: 10.1093/ntr/ntac136] [Medline: 35639822]
45. Yang JS, Pacheco C, Mackey TK. Factors affecting tobacco, nicotine, and cannabis product use among California young adults during the COVID-19 pandemic: a qualitative study. Addict Behav Rep 2022 Dec;16:100470 [FREE Full text] [doi: 10.1016/j.abrep.2022.100470] [Medline: 36405491]
46. Yang XY, Kelly BC, Pawson M, Vuolo M. Vaping in a time of pandemics: risk perception and motivations for electronic cigarette use. Nicotine Tob Res 2023 Jan 05;25(2):237-246 [FREE Full text] [doi: 10.1093/ntr/ntac050] [Medline: 35182425]
47. Lee E, Moore SP, Trapl E, Fryer CS, Gunzler D, Sterling KL. Changes in little cigar and cigarillo use during the COVID-19 pandemic: a cross-sectional analysis of a nationally representative sample of young adult users. Int J Environ Res Public Health 2022 Jul 22;19(15):8933 [FREE Full text] [doi: 10.3390/ijerph19158933] [Medline: 35897303]
48. Klemperer EM, West JC, Peasley-Miklus C, Villanti AC. Change in tobacco and electronic cigarette use and motivation to quit in response to COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1662-1663 [FREE Full text] [doi: 10.1093/ntr/ntaa072] [Medline: 32343816]
49. Koyama S, Tabuchi T, Okawa S, Kadobayashi T, Shirai H, Nakatani T, et al. Changes in smoking behavior since the declaration of the COVID-19 state of emergency in Japan: a cross-sectional study from the Osaka Health app. J Epidemiol 2021;31(6):378-386. [doi: 10.2188/jea.je20200533]
50. Popova L, Henderson K, Kute N, Singh-Looney M, Ashley DL, Reynolds RM, et al. "I'm bored and I'm stressed": a qualitative study of exclusive smokers, ENDS users, and transitioning smokers or ENDS users in the time of COVID-19. Nicotine Tob Res 2023 Jan 05;25(2):185-192 [FREE Full text] [doi: 10.1093/ntr/ntab199] [Medline: 34610133]
51. Schnitzer K, Jones S, Kelley JH, Tindle HA, Rigotti NA, Kruse GR. A qualitative study of the impact of COVID-19 on smoking behavior for participants in a post-hospitalization smoking cessation trial. Int J Environ Res Public Health 2021 May 19;18(10):5404 [FREE Full text] [doi: 10.3390/ijerph18105404] [Medline: 34069350]
52. Graupensperger S, Cadigan JM, Einberger C, Lee CM. Multifaceted COVID-19-related stressors and associations with indices of mental health, well-being, and substance use among young adults. Int J Ment Health Addict 2023 Aug 04;21(1):418-431 [FREE Full text] [doi: 10.1007/s11469-021-00604-0] [Medline: 34377106]
53. Berg CJ, Callanan R, Johnson TO, Schliecher NC, Sussman S, Wagener TL, et al. Vape shop and consumer activity during COVID-19 non-essential business closures in the USA. Tob Control 2021 Nov 19;30(e1):e41-e44 [FREE Full text] [doi: 10.1136/tobaccocontrol-2020-056171] [Medline: 33077506]
54. Arshonsky J, Krawczyk N, Bunting AM, Frank D, Friedman SR, Bragg MA. Informal coping strategies among people who use opioids during COVID-19: thematic analysis of Reddit forums. JMIR Form Res 2022 Mar 03;6(3):e32871 [FREE Full text] [doi: 10.2196/32871] [Medline: 35084345]

XSL•FO
RenderX

55. Bunting AM, Frank D, Arshonsky J, Bragg MA, Friedman SR, Krawczyk N. Socially-supportive norms and mutual aid of people who use opioids: an analysis of Reddit during the initial COVID-19 pandemic. Drug Alcohol Depend 2021 May 01;222:108672 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108672] [Medline: 33757708]
56. Alambo A, Padhee S, Banerjee T, Thirunarayan K. COVID-19 and mental health/substance use disorders on Reddit: a longitudinal study. In: Proceedings of the 2021 Conference on Pattern Recognition. ICPR International Workshops and Challenges. 2021 Presented at: ICPR '21; January 10-15, 2021; Virtual event p. 20-27 URL: https://link.springer.com/chapter/10.1007/978-3-030-68790-8_2 [doi: 10.1007/978-3-030-68790-8_2]
57. Albarracin D, Romer D, Jones C, Jamieson KH, Jamieson P. Misleading claims about tobacco products in YouTube videos: experimental effects of misinformation on unhealthy attitudes. J Med Internet Res 2018 Jun 29;20(6):e229 [FREE Full text] [doi: 10.2196/jmir.9959] [Medline: 29959113]
58. Clendennen SL, Loukas A, Vandewater EA, Perry CL, Wilkinson AV. Exposure and engagement with tobacco-related social media and associations with subsequent tobacco use among young adults: a longitudinal analysis. Drug Alcohol Depend 2020 Aug 01;213:108072 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108072] [Medline: 32563115]
59. Brailovskaia J, Swarlik VJ, Grethe GA, Schillack H, Margraf J. Experimental longitudinal evidence for causal role of social media use and physical activity in COVID-19 burden and mental health. Z Gesundh Wiss 2022 Sep 02;31(11):1-14 [FREE Full text] [doi: 10.1007/s10389-022-01751-x] [Medline: 36068852]
60. Majmundar A, Allem JP, Cruz TB, Unger JB. Public health concerns and unsubstantiated claims at the intersection of vaping and COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1667-1668 [FREE Full text] [doi: 10.1093/ntr/ntaa064] [Medline: 32285129]
61. Stein J, Keuschnigg M, van de Rijt A. Network segregation and the propagation of misinformation. Sci Rep 2023 Jan 17;13(1):917 [FREE Full text] [doi: 10.1038/s41598-022-26913-5] [Medline: 36650189]
62. Sokolovsky AW, Hertel AW, Micalizzi L, White HR, Hayes KL, Jackson KM. Preliminary impact of the COVID-19 pandemic on smoking and vaping in college students. Addict Behav 2021 Apr;115:106783 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106783] [Medline: 33360444]
63. Silver N, Kierstead E, Tran B, Sparrock L, Vallone D, Schillo B. Belief and recall of nicotine as therapeutic for COVID-19 may undermine e-Cigarette quitting behavior. Health Educ Behav 2022 Dec;49(6):929-933. [doi: 10.1177/10901981221109127] [Medline: 35848331]
64. Bakaloudi DR, Evripidou K, Siargkas A, Breda J, Chourdakis M. Impact of COVID-19 lockdown on smoking and vaping: systematic review and meta-analysis. Public Health 2023 May;218:160-172 [FREE Full text] [doi: 10.1016/j.puhe.2023.02.007] [Medline: 37043948]
65. Dave D, Dench D, Kenkel D, Mathios A, Wang H. News that takes your breath away: risk perceptions during an outbreak of vaping-related lung injuries. J Risk Uncertain 2020 Jun;60(3):281-307 [FREE Full text] [doi: 10.1007/s11166-020-09329-2] [Medline: 34504389]
66. Grummon AH, Hall MG, Mitchell CG, Pulido M, Mendel Sheldon J, Noar SM, et al. Reactions to messages about smoking, vaping and COVID-19: two national experiments. Tob Control 2022 May 13;31(3):402-410 [FREE Full text] [doi: 10.1136/tobaccocontrol-2020-055956] [Medline: 33188150]
67. Zeller M. Evolving "The Real Cost" campaign to address the rising epidemic of youth e-cigarette use. Am J Prev Med 2019 Feb;56(2 Suppl 1):S76-S78 [FREE Full text] [doi: 10.1016/j.amepre.2018.09.005] [Medline: 30661529]
68. Silberman WR, Record RA. We post it, U Reddit: exploring the potential of Reddit for health interventions targeting college populations. J Health Commun 2021 Jun 03;26(6):381-390. [doi: 10.1080/10810730.2021.1949648] [Medline: 34260329]
69. Geers M, Swire-Thompson B, Lorenz-Spreen P, Herzog SM, Kozyreva A, Hertwig R. The Online Misinformation Engagement Framework. Curr Opin Psychol 2024 Feb;55:101739 [FREE Full text] [doi: 10.1016/j.copsyc.2023.101739] [Medline: 38091666]
70. Ruths D, Pfeffer J. Social sciences. Social media for large studies of behavior. Science 2014 Nov 28;346(6213):1063-1064. [doi: 10.1126/science.346.6213.1063] [Medline: 25430759]
71. Amaya A, Bach R, Keusch F, Kreuter F. New data sources in social science research: things to know before working with Reddit data. Soc Sci Comput Rev 2019 Dec 18;39(5):943-960. [doi: 10.1177/0894439319893305]
72. Zhai Y, Du X. Disparities and intersectionality in social support networks: addressing social inequalities during the COVID-19 pandemic and beyond. Humanit Soc Sci Commun 2022 Apr 21;9(1):91. [doi: 10.1057/S41599-022-01163-Y]
73. O'Handley B, Courtice EL. Mental well-being, social support, and the unique experiences of transgender and nonbinary people during the COVID-19 pandemic. Can J Hum Sex 2022 Sep 01;31(2):163-175. [doi: 10.3138/cjhs.2022-0024]
74. Valera P, Owens M, Malarkey S, Acuna N. Exploring tobacco and e-Cigarette use among queer adults during the early days of the COVID-19 pandemic. Int J Environ Res Public Health 2021 Dec 08;18(24):12919 [FREE Full text] [doi: 10.3390/ijerph182412919] [Medline: 34948530]
75. Mamudu HM, Adzrago D, Dada O, Odame EA, Ahuja M, Awasthi M, et al. Examining disparities in current e-Cigarette use among U.S. adults before and after the WHO declaration of the COVID-19 pandemic in March 2020. Int J Environ Res Public Health 2023 Apr 26;20(9):5649 [FREE Full text] [doi: 10.3390/ijerph20095649] [Medline: 37174168]

XSL•FO
RenderX

76. Chen-Sankey JC, Broun A, Duarte DA, Ajith A, Jewett B, Smiley SL, et al. Exploring changes in cigar smoking patterns and motivations to quit cigars among black young adults in the time of COVID-19. Addict Behav Rep 2020 Dec;12:100317 [FREE Full text] [doi: 10.1016/j.abrep.2020.100317] [Medline: 33364325]
77. Abreu RL, Gonzalez KA, Arora S, Sostre JP, Lockett GM, Mosley DV. "Coming together after tragedy reaffirms the strong sense of community and pride we have:" LGBTQ people find strength in community and cultural values during the COVID-19 pandemic. Psychol Sex Orientat Gend Divers 2023 Mar;10(1):140-149. [doi: 10.1037/sgd0000516]
78. Singleton G, Johnson L, Singleton N, Li H. COVID-19-related anxiety: how do coping and optimism relate to substance use in African-American young adults? J Community Psychol 2023 Jul 20;51(6):2390-2407 [FREE Full text] [doi: 10.1002/jcop.22863] [Medline: 35441710]
79. Nguyen MH, Gruber J, Fuchs J, Marler W, Hunsaker A, Hargittai E. Changes in digital communication during the COVID-19 global pandemic: implications for digital inequality and future research. Soc Media Soc 2020 Jul 09;6(3):2056305120948255 [FREE Full text] [doi: 10.1177/2056305120948255] [Medline: 34192039]

Abbreviations

ACE2: angiotensin converting enzyme 2
OP: original post
PG: propylene glycol

Edited by A Mavragani; submitted 01.09.24; peer-reviewed by N Hu, Z Ehtesham; comments to author 26.09.24; revised version received 02.11.24; accepted 12.12.24; published 20.03.25.

Please cite as:
Watkins SL, Snodgrass K, Fahrion L, Shaw E
Contextualizing Changes in e-Cigarette Use During the Early COVID-19 Pandemic and Accompanying Infodemic ("So Much Contradictory Evidence"): Qualitative Document Analysis of Reddit Forums
J Med Internet Res 2025;27:e66010
URL: https://www.jmir.org/2025/1/e66010
doi:10.2196/66010
PMID:

©Shannon Lea Watkins, Katherine Snodgrass, Lexi Fahrion, Emily Shaw. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Optimizing Virtual Follow-Up Care: Realist Evaluation of Experiences and Perspectives of Patients With Breast and Prostate Cancer

Sarah Scruton[1], MSc; Geoff Wong[2], MD, PhD; Stephanie Babinski[1], MA, MPH; Lauren R Squires[1,3], MSc; Alejandro Berlin[4,5,6], MSc, MD; Julie Easley[7], PhD; Sharon McGee[8], PhD; Ken Noel[9], BSc; Danielle Rodin[4,5], MSc, MD; Jonathan Sussman[10], MSc, MD; Robin Urquhart[11], PhD; Jacqueline L Bender[1,3,12], PhD

[1]Cancer Rehabilitation and Survivorship, Department of Supportive Care, Princess Margaret Cancer Centre, Toronto, ON, Canada
[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[4]Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada
[5]Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, ON, Canada
[6]TECHNA Institute, University Health Network, Toronto, ON, Canada
[7]Department of Medical Education, Dr Everett Chalmers Regional Hospital, Horizon Health Network, Fredericton, NB, Canada
[8]Department of Medicine, Division of Medical Oncology, The Ottawa Hospital and the University of Ottawa, Ottawa, ON, Canada
[9]The Walnut Foundation, Brampton, ON, Canada
[10]Department of Oncology, McMaster University, Hamilton, ON, Canada
[11]Department of Community Health and Epidemiology, Dalhousie University, Halifax, ON, Canada
[12]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Jacqueline L Bender, PhD
Cancer Rehabilitation and Survivorship
Department of Supportive Care
Princess Margaret Cancer Centre
610 University Ave
Toronto, ON, MG5 2C1
Canada
Phone: 1 416 581 8606
Email: Jackie.Bender@uhn.ca

Abstract

Background: Virtual follow-up (VFU) has the potential to enhance cancer survivorship care. However, a greater understanding is needed of how VFU can be optimized.

Objective: This study aims to examine how, for whom, and in what contexts VFU works for cancer survivorship care.

Methods: We conducted a realist evaluation of VFU among patients with breast cancer and prostate cancer at an urban cancer center during the COVID-19 pandemic. Realist evaluations examine how underlying causal processes of an intervention (mechanisms) in specific circumstances (contexts) interact to produce results (outcomes). Semistructured interviews were conducted with a purposive sample of patients ≤5 years after diagnosis. Interviews were audio-recorded and analyzed using a realist logic of analysis.

Results: Participants (N=24; n=12, 50% with breast cancer and n=12, 50% with prostate cancer) had an average age of 59.6 (SD 10.7) years. Most participants (20/24, 83%) were satisfied with VFU and wanted VFU options to continue after the COVID-19 pandemic. However, VFU impacted patient perceptions of the quality of their care, particularly in terms of its effectiveness and patient centeredness. Whether VFU worked well for patients depended on patient factors (eg, needs, psychosocial well-being, and technological competence), care provider factors (eg, socioemotional behaviors and technological competence), and virtual care system factors (eg, modality, functionality, usability, virtual process of care, and communication workflows). Key mechanisms that interacted with contexts to produce positive outcomes (eg, satisfaction) were visual cues, effective and empathetic communication, and a trusting relationship with their provider.

XSL•FO
RenderX

Conclusions: Patients value VFU; however, VFU is not working as well as it could for patients. To optimize VFU, it is critical to consider contexts and mechanisms that impact patient perceptions of the patient centeredness and effectiveness of their care. Offering patients the choice of in-person, telephone, or video visits when possible, coupled with streamlined access to in-person care when required, is important. Prioritizing and addressing patient needs; enhancing physician virtual socioemotional behaviors and technology competency; and enhancing VFU functionality, usability, and processes of care and communication workflows will improve patient perceptions of the patient centeredness and effectiveness of virtual care.

(J Med Internet Res 2025;27:e65148) doi:10.2196/65148

KEYWORDS

cancer; follow-up; virtual; outcomes; realist evaluation; survivorship

Introduction

According to report "Canadian Cancer Statistics 2023," 1 in 2 Canadians will be diagnosed with cancer in their lifetime, and two-thirds of Canadians diagnosed with cancer today will live for ≥5 years after their diagnosis [1]. Among the most common cancers in men and women, the 5-year net survival rate is 87% for breast cancer (BC) and 95% for prostate cancer (PC) [1]. While these outcomes reflect advances in early detection and treatment, they have substantially increased the demand for follow-up care services. Optimal follow-up care consists of surveillance for cancer spread, recurrences, or secondary cancers; prevention and management of acute and persistent treatment side effects; and promotion of healthy behaviors to mitigate new and ongoing health concerns [2]. To be effective, this involves ongoing assessment and comprehensive care, including referral to specialized supportive care services to improve quality of life, reduce disability, and restore function [3,4].

In Canada, and internationally, follow-up cancer care is typically provided by oncologists in outpatient secondary or tertiary care settings. Evidence suggests that this model of in-person follow-up care with a specialist is not sustainable and is not working well for patients [5]. Oncology offices are often overcrowded with long wait times, short appointments, and high costs per visit [6]. Furthermore, the high burden of unmet posttreatment supportive care needs suggests that the existing model of follow-up care is not meeting patients' needs [7,8]. Alternatives such as follow-up with primary care providers are comparable to oncologist-led follow-up in terms of detection of recurrence, survival, and quality of life [9,10]. However, prior work [6,11,12] shows that many survivors of cancer prefer to be seen by an oncologist for follow-up due to concerns about the quality and continuity of care.

One method proposed to improve the experience and effectiveness of cancer follow-up care is virtual care, defined as remote interactions between patients and care providers using technology (ie, video calls, telephone, email, etc) [13]. Virtual follow-up (VFU) could alleviate pressure on clinicians and health care services and better meet patient needs by providing patients with access to timely, convenient, and tailored follow-up care [14]. Before the COVID-19 pandemic, virtual care was infrequently used in Canada, mainly to provide care to rural and remote populations [13], and the evidence of its effectiveness was limited [15]. During the COVID-19 pandemic, there was rapid and widespread adoption of virtual care in Canada and globally [16]. In a prior study, we demonstrated that most patients with cancer were satisfied with VFU during the COVID-19 pandemic, but some did not want VFU to continue after the COVID-19 pandemic [17]. In addition, while virtual visits made care more accessible for some, other studies have shown that virtual care has exacerbated access disparities by widening the digital divide for those who lack the knowledge or technological means to access care virtually [18,19].

To sustain VFU as an option for patients with cancer, we need a better understanding from patients on how, for whom, and in what contexts this model of care works. Patient perspectives on how to optimize VFU care can inform updates to telehealth clinical practice guidelines, such as those offered by the American Association of Clinical Oncology (ASCO) early in the COVID-19 pandemic [20], to ensure patient-centered virtual care. Realist evaluations are ideally suited to answer these questions. Realist evaluations are a proven, theory-driven method that enable a rigorous understanding of context and its influence on causal processes (ie, mechanisms) to explain how and for whom an innovation works best [21,22]. This method is increasingly used to assess complex innovations in health care, particularly those that involve multiple interconnected components and are dependent on individual responses and the wider context for their success [23,24]. Therefore, we conducted a realist evaluation [21] to examine how, for whom, and in what contexts VFU works for patients with BC and PC and conclude with recommendations for optimizing VFU for patients with cancer.

Methods

Ethical Considerations

The study was approved by the University Health Network Institutional Research Ethics Board (ID 21-5397). Informed consent was obtained from all study participants before data collection, and all data were deidentified before analysis. All study participants received a CAD $25 honorarium to compensate for their time. At the time of the study, US $1 was equal to CAD $1.30.

Study Design

We conducted a realist evaluation following the steps by Pawson and Tilley [21] and RAMESES [23] quality and reporting guidelines.

Step 1 involved developing an initial program theory, which is a kind of conceptual framework, to explain how VFU could be

XSL•FO
RenderX

delivered to work most effectively. This involved explaining how the underlying processes of VFU care (mechanisms) interact with specific circumstances (contexts) to produce results (outcomes). For this step, we used prior evidence to understand patient views [25] and the effectiveness of virtual versus in-person cancer care [15], theories from technology adoption [26-28] to understand factors that influence the adoption of VFU and theories from patient-provider communication [29] to determine the essential elements of effective clinical encounters, and the 6 domains of health care quality identified by the Agency for Healthcare Research and Quality [30] to understand the attributes of quality VFU. Using these information sources, we created a figure to illustrate how VFU can bring about the desired outcomes. We then held a 2-hour virtual workshop with project team members and stakeholders, consisting of clinicians, researchers, and patient partners (N=10) to obtain their views on the initial theory.

Step 2 involved collecting evidence on the contexts, mechanisms, and outcomes of VFU. For this, we (1) administered a web-based survey to patients at a cancer centre to obtain information on their sociodemographic characteristics, experience, and satisfaction with VFU [17]; (2) purposively recruited a subsample of survey respondents to participate in semistructured interviews to explore their views and experiences with VFU in greater depth; and (3) administered a short follow-up survey to collect information on interview participants' psychosocial well-being.

Steps 3 and 4 involved data integration and synthesis using a realist logic of analysis. The goal of this analysis was to determine whether data were functioning as context, mechanism, or outcome, and if so, within which context-mechanism-outcome configuration (CMOC) [23]. We then held a 2-hour virtual stakeholder workshop with a wider group of stakeholders (N=12) to obtain feedback on the relevance of our findings more broadly.

Setting and Participants

This study was conducted at the Princess Margaret (PM) Cancer Centre, a tertiary, university-affiliated, teaching hospital, which is a part of the University Health Network in Toronto, Canada. Before COVID-19, <1% of visits at PM were conducted virtually [31]. After the declaration of the COVID-19 pandemic, 68% of visits at PM were virtual [31]. The participants in this study comprised a subsample of patients with BC and PC who had completed the PM Virtual Care Evaluation Survey between May and July 2021 [17] and who agreed to participate in a follow-up interview. Interviews occurred between September and November 2021. Survey participants had to meet the following criteria: they must have received a cancer diagnosis, be aged >18 years, have participated in at least 1 virtual appointment (eg, by phone or video) in the last 12 months, and have a valid email address on file. We used the survey data to intentionally select interview participants who varied in age, ethnicity, geographic location, and satisfaction with VFU. In addition, to be eligible to participate in an interview, survey respondents must have agreed to be contacted about a follow-up interview, received a diagnosis of BC or PC within the last 5 years, and completed treatment for cancer or were receiving adjuvant hormone therapy. On the basis of previous research, a sample size of 15 per patient group (30 in total) was estimated to be sufficient [32,33]. However, data collection continued until theoretical saturation [34].

Data Collection

First, data on participants' sociodemographic characteristics and virtual care use and perceptions were extracted from the PM Virtual Care Evaluation Survey to describe the sample and inform each interview. Demographic data included age, whether they were born in Canada (yes or no), English as a first language (yes or no), race or ethnicity (10 response options), highest level of education (high school or less, college or technical school, university undergraduate, or postgraduate), and household income (CAD <$60,000, CAD $60,000-$99,000, or CAD ≥$100,000). Virtual care use and preferences were assessed by asking participants to indicate the type of virtual appointments they had received and would have liked to receive in the past 12 months (phone, video, or both). Patient satisfaction with virtual care was captured using a 5-point Likert response (very satisfied, satisfied, neutral, dissatisfied, or very dissatisfied). Patient desire for virtual care options after the COVID-19 pandemic was captured by asking participants to indicate their level of agreement with the statement "I would like to continue to have virtual options for some of my visits after the COVID-19 pandemic ends," using a 5-point Likert response option (strongly agree, agree, neutral, disagree, or strongly disagree).

Each participant then took part in one 45- to 60-minute semistructured interview based on the realist approach [21,22]. They were asked about their initial expectations and desired outcomes of VFU, how it worked or did not work for them, and how VFU could be made to work optimally. The interviews were conducted by a research coordinator with experience in qualitative interview methods (SB) who was trained in realist methodology by a realist expert on the team (GW). Interviews were audio-recorded and transcribed verbatim.

After the interview, participants completed a brief survey on their level of patient activation, anxiety, and depression. Patient activation was measured using the Patient Activation Measure [35], which scores a patient's knowledge, skills, and confidence for proactively managing their health and health care using a 13-item tool with 4-point Likert scales. The Patient Activation Measure scores (between 0 and 100) were converted into four levels of activation: (1) not believing (patient) activation is important (score of ≤47.0), (2) a lack of knowledge and confidence (score of 47.1-55.1), (3) beginning to take action (score of 55.2-67.0), and (4) taking action (score of ≥67.1). Anxiety was measured using the Generalized Anxiety Disorder 7-item [36], and depression was measured using the Patient Health Questionnaire-9 [37]. Anxiety and depression scores were converted into the following categories: minimal (score of 0-4), mild (score of 5-9), moderate (score of 10-14), moderately severe (depression score only [15-19]), and severe (score of 20-27 on Patient Health Questionnaire-9 or 15-21 on Generalized Anxiety Disorder 7-item). Fear of recurrence was measured using the Fear of Cancer Recurrence Inventory-Short Form [38], a 9-item severity scale that evaluates intrusive thoughts associated with fear of cancer recurrence using 4-point

XSL•FO
RenderX

Likert scales. A score of ≥13 on the Fear of Cancer Recurrence Inventory-Short Form denotes a clinically relevant fear of cancer recurrence.

Data Analysis

A realist logic of analysis was used to analyze interview transcripts [21], with the support of NVivo (version 14; Lumivero). The coding was both deductive (some codes were created in advance informed by the initial program theory) and inductive (some codes emerged from the data). The data were initially sorted into broad conceptual categories or "buckets" by 2 team members independently (SB and SS) through consultation with the lead author (JLB). The buckets were developed following data collection both deductively and inductively based on the initial program theory and what was found in the data. They were further analyzed by 1 team member (SS) to develop preliminary themes and CMOCs. As there were no meaningful differences across the 2 groups of BC and PC participants, the data were analyzed and reported jointly. The themes and CMOCs underwent an intensive refinement process with 3 research team members (JLB, GW, and SS) to ensure they were clear, concise, and representative of the interview data. Repetitive or redundant CMOCs were removed. The data were then used to refine the program theory. Descriptive statistics were used to summarize patient demographic characteristics, virtual care use, patient activation, and psychosocial well-being using R (version 4.1.2; R Foundation for Statistical Computing).

Results

Participant Characteristics

In total, 24 patients were interviewed; of these, 12 (50%) had been diagnosed and treated for BC and 12 (50%) had been diagnosed and treated for PC (Table 1). On average, the participants were aged 65.9 (SD 8.65) years, 100% (24/24) attended postsecondary school, and 71% (17/24) reported a household income of CAD ≥$100,000. Approximately half (11/24, 46%) of the sample identified as racialized, 50% (12/24) were born outside of Canada, and 20% (5/24) had a non-English first language. The most common treatment type was surgery (18/24, 75%), followed by radiation therapy (17/24, 71%) and chemotherapy (9/24, 37%). This was a highly activated sample, with most (18/24, 75%) at level 3 or 4 on the patient activation scale and most with relatively mild or minimal depression (18/24, 75%) or anxiety (20/24, 83%), and most participants (15/24, 63%) reported nonclinical levels of fear of recurrence.

XSL•FO
RenderX

Table 1. Participant characteristics (N=24).

Characteristic	Prostate cancer (n=12)	Breast cancer (n=12)	Total
Demographic characteristics			
Age (y), mean (SD)	65.92 (8.65)	53.25 (14.31)	59.58 (13.41)
Race or ethnicity, n (%)			
Arab or West Asian	0 (0)	1 (8)	1 (4)
Black	1 (8)	2 (17)	3 (13)
East Asian	0 (0)	1 (8)	1 (4)
Jewish	1 (8)	0 (0)	1 (4)
Latin American	0 (0)	2 (17)	2 (8)
Southeast Asian	0 (0)	2 (17)	2 (8)
South Asian	1 (8)	0 (0)	1 (4)
White	9 (75)	4 (33)	13 (54)
Born in Canada , n (%)			
Yes	5 (42)	7 (58)	12 (50)
English as a first language, n (%)			
Yes	10 (83)	9 (75)	19 (79)
Highest level of education received, n (%)			
High school or less	0 (0)	0 (0)	0 (0)
Postsecondary school	12 (100)	12 (100)	24 (100)
Household income (CAD $1.30=US $1), n (%)			
60,000-99,999	2 (17)	2 (17)	4 (17)
≥100,000	8 (67)	9 (75)	17 (71)
Prefer not to say	2 (17)	1 (8)	3 (12)
Clinical and psychosocial characteristics			
Treatments received, n (%)			
Drug or chemotherapy	0 (0)	9 (75)	9 (37)
Hormone therapy	1 (8)	5 (42)	6 (25)
Radiation therapy	6 (50)	11 (92)	17 (71)
Surgery	8 (67)	10 (83)	18 (75)
Other	1 (8)	0 (0)	1 (4)
Anxiety level , n (%)			
Minimal or mild	10 (83)	10 (83)	20 (83)
Moderate or severe	1 (8)	2 (17)	3 (13)
Depression level, n (%)			
Minimal or mild	11 (92)	7 (58)	18 (75)
Moderate, moderately severe, or severe	1 (8)	4 (33)	5 (21)
Level of patient activation , n (%)			
Not believing activation is important	2 (17)	1 (8)	3 (13)
A lack of knowledge or confidence	0 (0)	1 (8)	1 (4)
Beginning to take action	5 (42)	4 (33)	9 (37)
Taking action	4 (33)	5 (42)	9 (37)
Fear of recurrence, n (%)			
Clinical	4 (33)	3 (25)	7 (29)

XSL•FO
RenderX

Characteristic	Prostate cancer (n=12)	Breast cancer (n=12)	Total
Nonclinical	7 (58)	8 (67)	15 (63)
Missing	1 (8)	1 (8)	2 (8)

VFU Use and Preferences

All participants received a phone call appointment (Table 2). Most patients with BC also received a video appointment (20/24, 83%); none of the patients with PC received a video appointment. Overall, most participants (20/24, 83%) were satisfied or highly satisfied with their virtual care visits, and most participants (20/24, 83%) would want VFU options to continue after the COVID-19 pandemic. When asked which type of VFU appointment they would prefer in the future, most patients (15/24, 63%) indicated video appointments.

Table 2. Virtual follow-up (VFU) characteristics (N=24).

Characteristic	Prostate cancer (n=12)	Breast cancer (n=12)	All
Type of VFU appointment received, n (%)			
Only phone	12 (100)	2 (17)	14 (58)
Only video	0 (0)	0 (0)	0 (0)
Phone+video	0 (0)	10 (83)	10 (42)
Satisfaction with virtual care, n (%)			
Very satisfied	7 (58)	7 (58)	14 (58)
Satisfied	4 (33)	2 (17)	6 (25)
Dissatisfied	1 (8)	3 (25)	4 (17)
Would want the option for VFU after the COVID-19 pandemic, n (%)			
Strongly agree	6 (50)	7 (58)	13 (54)
Agree	4 (33)	3 (25)	7 (29)
Neutral	2 (17)	2 (17)	4 (17)
Disagree	0 (0)	0 (0)	0 (0)
Preferences for virtual care in the future, n (%)			
No preference	1 (8)	4 (33)	5 (21)
Phone	4 (33)	0 (0)	4 (17)
Video	7 (58)	8 (67)	15 (63)

Views and Experiences With VFU

Overview

We identified 8 overarching concepts that captured patients' views on how VFU could be made to work effectively. These 8 concepts consisted of 24 themes and 46 CMOCs (Multimedia Appendix 1). Due to the number of CMOCs that were developed, each CMOC is labeled with a code (eg, A1, A2) found in Multimedia Appendix 1 (for ease of reference). The findings related to each CMOC are labelled with a matching code throughout the Results section. The concepts, themes, and CMOCs were used to refine the program theory (Figure 1).

XSL•FO
RenderX

Figure 1. Refined program theory.

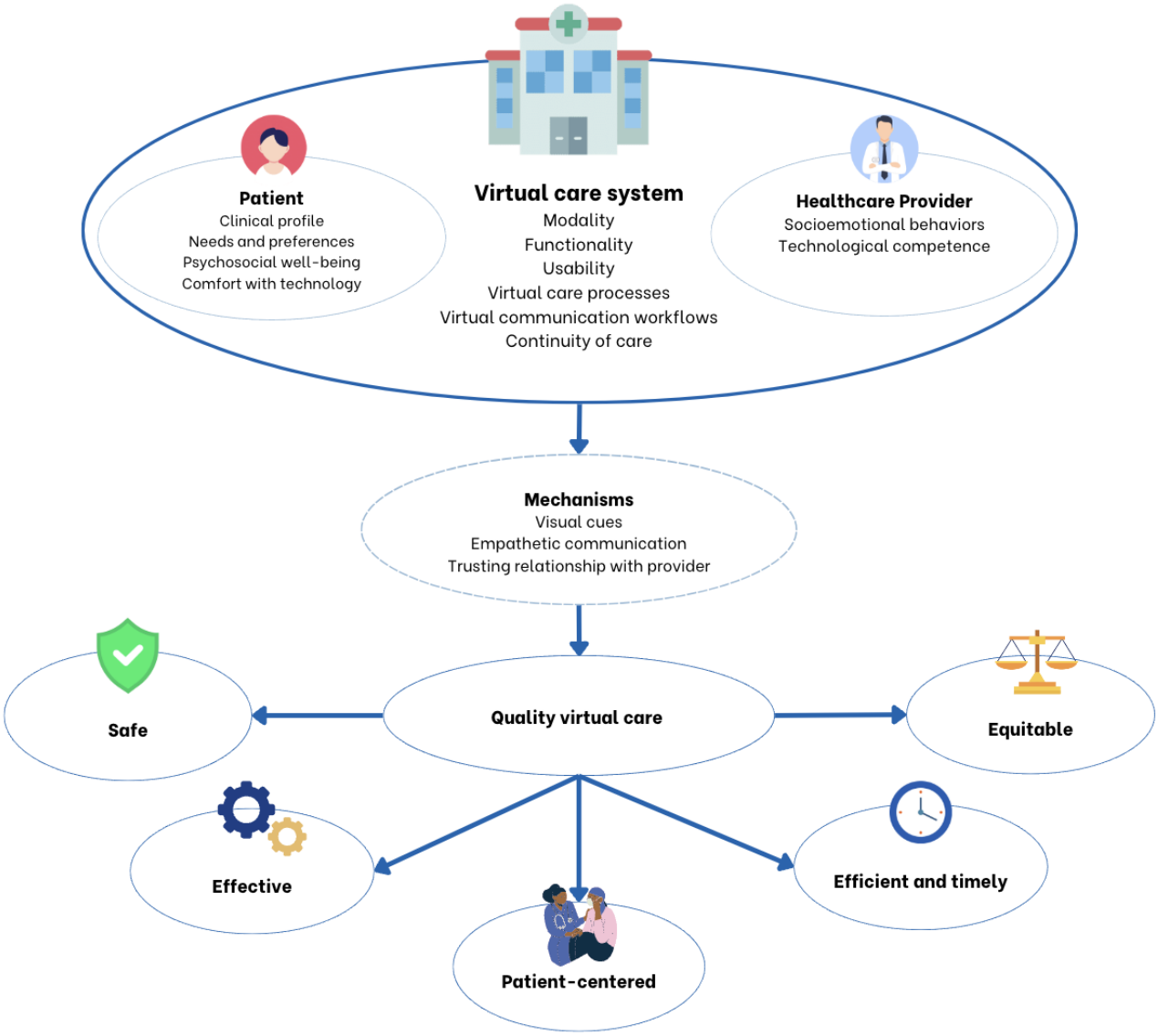

Effectiveness

Effectiveness of VFU appears to be influenced by patient-, physician-, and disease-related contextual factors. For example, VFU was considered ineffective if it replaced physical appointments all together, as physical examinations and speaking to a physician in person reduced patient anxiety about recurrence and management of ongoing side effects (A1). In-person appointments provided visual cues (ie, body language, facial expression, etc) that helped patients understand the meaning of information received from their provider and feel more confident in their care (A2). However, participants felt that this may be partly overcome through video appointments, as video allows patients and care providers to read each other's body language and facial expressions. This may be particularly important for patients experiencing high levels of anxiety about their cancer, as the lack of visual cues with phone calls causes VFU to be insufficiently reassuring (A3). Participants also felt that it was important for physicians to see the visual cues of the patient, as this would enable the physician to pick up on distress or concern that may not be verbally expressed by patients and act accordingly (A4). The final contextual factor described to impact effectiveness of VFU was the perceived loss of the appreciation of the seriousness of cancer. In-person appointments act as a reminder of the seriousness of the condition. Without in-person appointments, patients were concerned that they may become less vigilant in their self-care and self-management (eg, nonadherence to medication or forgetting self-examinations; A5).

Efficiency and Timeliness

If participants felt that the VFU they received was both efficient and timely, they were more likely to be satisfied with their care. These participants felt that VFU would likely improve the overall efficiency of follow-up care, as they believed hospitals could save time and money by scheduling more appointments per day (B1). Participants also believed that greater use of virtual appointments could improve the experience of in-person appointments (when required), as hospital waiting rooms would be less crowded and stressful (B2).

However, some participants expressed concerns that VFU could lead to delays in receiving care. Many described being given very long windows of time for their VFU appointments, which

XSL•FO
RenderX

were difficult and frustrating to plan around (B3). In some cases, physicians called outside of these time windows, which led to missed appointments or patients having to speak to someone other than their principal care provider (B4). These types of situations were anxiety inducing for patients, particularly if they had concerns to share with their provider as they feared that their condition would worsen while waiting for a rescheduled appointment (B5). Issues with timeliness may also occur if patients have difficulty booking timely appointments due to long wait times for callbacks or uncertainty about whom to contact (B3).

Patient Centeredness

The largest conceptual bucket by far was related to patient centeredness. Participants described ways in which VFU can be more or less patient-centered compared to in-person care, and how this impacted how well VFU worked for them. This included references to personal values, preferences, and expectations of VFU and how these measured up to their experiences. A patient's preferences and expectations were influenced not only by their unique life circumstances but also by health system characteristics, such as continuity of care. Participants were concerned that the number of care providers involved in their care may increase with VFU due to physicians being busier with more appointments. Participants explained that receiving care from multiple care providers made them feel less cared for because they had less time to form trusting relationships with each provider (C1). This was particularly anxiety inducing for those who were newly diagnosed, as they were particularly vulnerable and in need of a strong physician-patient relationship (C2). In addition, they explained that receiving care by multiple care providers led to poor continuity and quality of care due to a lack of communication between care providers. This was frustrating for participants, as they felt they had to continually repeat their history to their care providers, and they were concerned that their care would be negatively impacted by their physician not having a comprehensive understanding of their case (C3 and C4).

Difficulty forming relationships with care providers was also heightened when participants had never been given the opportunity to meet their provider face-to-face, as they felt this hindered their ability to open up emotionally (C5). Some participants felt that the care they received was more holistic and empathetic when provided face-to-face, and they valued the ability to visually see that they had the full attention of their care provider (C6). Therefore, participants appreciated when their care providers would make use of video calls for their VFU appointments. However, most participants also expressed that their VFU appointments felt rushed compared to in-person appointments, which made them feel less cared for as they were unable to take the time needed to process information and ask questions (C7). Many participants felt their VFU appointments only addressed their clinical or medical issues, as opposed to the personal issues that accompany a cancer diagnosis, which caused them to have unmet needs (C8). This may be because the nature of VFU tends to favor information provision rather than emotional support (C9).

When patients leave active treatment, some described feeling "lost in the system." This may be increased with VFU, as not entering hospitals means that it may be more difficult for patients to access support when needed (C10). Participants felt that VFU may be improved if it included straightforward processes to obtain information, support, and care when needed, which would prevent them from feeling abandoned. This could include the ability to communicate with their care provider via technology (ie, email) in between appointments (C11 and C12). In addition, participants expected their care providers to continue to provide referrals to supportive care programs, but this was not always the case with VFU (C13).

Finally, participants appreciated that VFU improved the convenience of care. Patients did not have to travel to the hospital for their appointment, which saved them time and money (C14). They also appreciated that they could attend their appointments from anywhere, including from the comfort of their own home, which avoided stressful and slow waiting rooms (C15). This also meant they could multitask while waiting for appointments, meaning VFU was less disruptive to their schedule (C15). VFU also allowed family members to be included in VFU appointments, which was particularly valuable when COVID-19 prevented family members from attending in-person appointments (C16). Many participants felt that these advantages of VFU outweighed the benefits of in-person care.

Equity

Follow-up care works best when it is equitable for every patient. Participants noted that VFU may limit access to care for those with certain sociodemographic characteristics such as a disability, older age, or low socioeconomic status. For example, participants expressed concern that someone with hearing loss may face difficulties speaking over the phone, and someone with vision loss may struggle with video appointments (D1). Some participants also considered that some individuals may not have access to fast internet or that those with a lower level of education may be unable to search for medical information on the internet if they are not given time to ask questions during VFU appointments (D2 and D3).

Safety

Participants noted that VFU could improve safety by decreasing their risk of being exposed to infectious diseases such as COVID-19 (E2). However, some participants felt that VFU may be less safe because it removed the ability to be physically examined by an oncologist, potentially resulting in poorer care. Turning elsewhere for care (such as the emergency room) caused participants to feel they were being inadequately cared for (E1).

Patient Characteristics

Whether VFU worked well depended on patient characteristics that impacted their suitability and preferences for virtual care. This included comfort with technology, as some participants described challenges that they and others may have with navigating and using VFU technology (F1). Participants who had experience working within a virtual environment had an easier time using VFU and therefore were more likely to be satisfied with it (F2).

XSL•FO
RenderX

Other patient factors that influenced whether VFU was suitable included clinical characteristics such as mental health, disease characteristics, or comorbidities. Those with complex issues (ie, fear of recurrence, severe side effects, anxiety, and recently completed treatment) reported a greater need for reassurance, acknowledgment, and help coping with their adversities, along with a higher reliance on physical examinations, making them a poorer candidate for VFU (F3). The opposite was true for those who experienced fewer side effects and whose appointments mainly addressed simple routine check-ups, making them ideal candidates for VFU (F4).

At the same time, VFU seemed to overcome access barriers experienced by those who are less likely to seek health care. Participants who described themselves as being reluctant to access the health care system (eg, due to a lack of time, downplaying their condition, not the type to seek care, etc) appreciated the convenience and efficiency of VFU, which they explained motivated them to access health care (F5). Finally, a patient's preference for VFU seemed to depend on their coping style. Those who described themselves as requiring considerable emotional support (ie, emotional coping) felt that VFU did not work as well for them because VFU tends to focus on information provision over emotional support (F6). In contrast, those who did not require as much emotional support (ie, problem-focused coping), appreciated the simple, quick information-provision appointments (F7).

Physician Characteristics

Many participants described the importance of physician competence and bedside manner regardless of whether the appointment was virtual or in person. If a physician did not appear to be comfortable with the VFU technology, participants judged the physician as less competent in providing care and were concerned that this could impact the quality of their care (G1). In addition, if a physician was less empathetic, patient, or comforting during VFU compared to in-person appointments, participants stated that they would not prefer this type of care as they would feel less cared for (G2). Most participants described feeling more rushed and less listened to during phone appointments, as if the appointments were only for the purpose of sharing information such as test results and not for the purpose of addressing emotional concerns (G3). Some participants felt that "webside manner" is something that could be improved with proper training (G4).

Virtual Care System Characteristics

The final factor found to impact whether VFU works well for patients was the characteristics of the virtual care system itself. The most discussed topic was VFU modality, as most patients were not offered the choice between having their VFU appointments via phone or video. Some participants were okay with this, as they felt the VFU modality that they were offered was effective for their needs. However, others were disappointed that they were not given a choice in the modality. They felt that being provided the option for video appointments would have increased the quality of their care as they could have benefitted from face-to-face communication (H1). That said, this was only true if the VFU technology worked well (functionality) and was easy to use (usability). Some participants were frustrated with the VFU technology because it was difficult for them to use and did not work well for them (H2). However, when VFU technology does work well, participants described being appreciative of the fact that they could easily access appointments and test results and communicate with their care providers. Participants felt that physicians should be responsible for offering fallback plans if the VFU technology did not work well, such as what to do when a call drops, so that they need not be fearful of the technology malfunctioning (H3). Some participants noted the lack of virtual care processes and communication channels (virtual communication workflows). This caused issues when they did not know how to contact support staff and schedule appointments, when they did not know if their appointment would occur outside of the scheduled time window, and when they did not know what to do if their VFU call or internet connection was dropped or the technology malfunctioned (H4).

Recommendations

The study findings point to several strategies that health care providers and health care systems can use to enhance the patient centeredness and perceived effectiveness of VFU appointments for patients. We have outlined these patient-derived recommendations in Textbox 1, which align with ASCO's guidelines for telehealth in oncology [38] and provide additional considerations to optimize VFU for patients.

Textbox 1. Patient-derived recommendations for improving virtual follow-up (VFU) care.

1. Offer patients the choice of an in-person or virtual visit
2. Offer patients the choice of virtual visit modality (eg, telephone or video)
3. Ensure there is an easy-to-use and efficient process for patients to schedule an in-person visit if needed
4. Clearly communicate how the appointment can be continued if the technology fails and provide technical support
5. Ensure that virtual visit booking and scheduling processes are flexible enough to cater to patient's changing needs and expectations and reduce wait times where possible
6. Make time to listen to patient concerns and provide emotional support during virtual visits
7. Demonstrate empathy and provide your full attention using active listening skills during virtual visits
8. Demonstrate competency with virtual care technology
9. Ensure patients are provided resources and referrals to survivorship care programs during or after a virtual visit
10. Improve the quality of VFU technology and work flows to create a seamless and reliable care experience for patients

XSL•FO
RenderX

Discussion

Principal Findings

To our knowledge, this is the first realist evaluation of VFU for cancer care. This study has helped to answer for whom and in which contexts VFU is most suitable and has identified ways in which VFU could be optimized to better meet the needs of patients with cancer receiving follow-up care. Overall, the patients with BC and PC in this study described themselves as being satisfied with VFU during the COVID-19 pandemic and wanted VFU options to continue after the COVID-19 pandemic. However, VFU impacted patient perceptions of the quality of their care, particularly in terms of its effectiveness and patient centeredness. Whether VFU worked well for patients depended on patient factors (eg, needs, psychosocial well-being, and technological competence), provider factors (eg, socioemotional behaviors and technological competence), and virtual care system factors (eg, modality, functionality, usability, and virtual processes of care and communication workflows). Key mechanisms that interacted with contexts to produce positive outcomes (eg, satisfaction and reassurance) were visual cues, clear empathetic communication, and a trusting relationship with their provider.

Is Virtual Care Suitable for All Cancer Follow-Up Appointments?

While follow-up appointments have been identified by patients with cancer and health care providers and recommended by clinical practice guidelines as suitable for virtual care delivery because of their relative brevity and simplicity [20,39,40], they can be anxiety provoking and complex depending on the circumstances. For example, moderate to severe “scanxiety” (the fear, stress, and anxiety in anticipation of surveillance tests in follow-up cancer care) is present in as many as 28% of survivors of cancer [41]. In a prior study, we reported that survivors of cancer who experienced distress (anxiety or depression) were less likely to be satisfied with VFU and less likely to want VFU in the future compared to those who did not report distress [17]. This study has revealed that patients experiencing distress may not be satisfied with VFU because the quality of emotional support provided virtually is less effective. Participants felt more comforted during in-person visits because they had more time to discuss their concerns and the type of emotional support received was more effective because the visual cues of the provider conveyed empathy and understanding. Interestingly, a systematic review comparing virtual to in-person cancer care on the psychosocial outcomes of patients with cancer found that virtual care from physicians or nurses was more beneficial for patients during active treatment than follow-up [42]. However, a systematic review comparing virtual versus in-person cancer care found virtual psychosocial counseling to be equally effective for follow-up care [15]. The ASCO telehealth guidelines recommend providing individualized orientation and instruction to patients and care providers on the specific technology that will be used for the virtual care interaction [20]. Our findings highlight a need for care provider training in empathetic virtual communication as well, which has been expressed by care providers themselves [43], accompanied with seamless referrals to psychosocial professionals for follow-up psychosocial issues.

For Whom and in Which Contexts VFU Is Most Suited

Importantly, this study has identified for whom and in which contexts VFU is most suited—a knowledge gap that has been identified by many [39,44,45]. In doing so, the study findings align with and considerably expand upon the recommendations of the European Society of Medical Oncology on which patients should be offered telehealth [46]. In its current form, this study has found that VFU is more suited for patients with cancer who are doing well—physically and emotionally—often termed “the well follow-up patient” [11]. From a clinical perspective, this includes patients with no evidence of disease; with limited or well managed treatment side effects including no-to-mild depression, anxiety, and fear of recurrence; who have problem-focused coping styles; and who require limited emotional support from professionals to deal with the effects of cancer and its treatment. From a sociodemographic perspective, this also includes patients who are comfortable speaking in English (for appointments that are in English); who do not have auditory, visual, cognitive, or physical impairments that limit their effective use of VFU technology; who have access to and experience with VFU technology; and who have the education and skills needed to proactively manage their health and health care. VFU is also well suited for patients with cancer with busy schedules and many competing demands, as they appreciate the convenience of virtual visits. It may even increase health care access for those who are typically reluctant to seek care due to the inconvenience of in-person visits.

Hence, it its current form, if VFU was the only option, it would likely increase disparities in access to care for some patients with cancer. Other studies have demonstrated this to be the case. Notably, a study of health care use during the COVID-19 pandemic in the United States found that Black patients had 0.6 times of the adjusted odds of accessing care through telemedicine compared to White patients [19]. However, subgroup analysis revealed that younger Black women were more likely to access care through telemedicine, particularly for urgent care issues. In contrast, a study of access disparities conducted at the same institution as this study identified no differences in virtual care use based on patient demographics [47]. This may be due to the higher socioeconomic status of their study sample, as only 7% identified as low income. Importantly, they found that regardless of visit type, patients who were structurally marginalized by ethnocultural, situational, and residential status, as well as gender, were less satisfied with their care whether it was virtual or in person. These findings highlight the need for proactive and continuous efforts to identify and intervene to address health disparities in access, experience, and outcomes.

How Health Care Providers Can Optimize Virtual Care for Patients

This study also identified strategies that care providers can use to enhance the patient centeredness and perceived effectiveness of VFU for patients. In fact, health care provider behavior was the key contextual factor that triggered the mechanisms (eg, assessment of visual cues, providing full attention, etc) that led

XSL•FO
RenderX

to positive outcomes (eg, patient responses such as trust, reassurance, feeling cared for, and satisfaction). VFU was considered ineffective for patients if care providers did not give patients the option of choosing in-person versus virtual appointments; were not competent using virtual care technology; did not give patients their full attention and time; were less empathetic, patient, and comforting during the virtual visit; or focused on collecting or providing information at the expense of emotional support. Collectively, except for technological competence, these behaviors reflect socioemotional behaviors critical for the effectiveness of the patient-physician encounter. Drawing on social interaction theory, Roberts and Aruguete [48] have shown that patients mostly recognize and react to physician socioemotional behaviors (eg, concern, affection, and attention displayed verbally and nonverbally) due to a lack of understanding of physician task behaviors (eg, explanation of etiology, symptoms, and treatment). In a series of experiments using videotapes of varied levels of physician task versus socioemotional behavior, they demonstrated that high levels of socioemotional behaviors increased patient self-disclosure, trust, satisfaction, and likelihood of recommending the physician. Participants in this study recognized that physician "webside manner" could be improved through training. Likewise, a study of key informants representing Canadian provincial and national health care organizations with expertise in virtual care delivery [43] identified a need to train care providers in how to use technology and integrate virtual care into practice, adapt their clinical skills (including examination skills) to virtual care, build rapport through good communication with patients, and understand when an in-person visit is necessary based on the nature of the appointment and patient contextual factors.

How Health Care Systems Can Optimize Virtual Care for Patients

System transformation is needed to deliver virtual care optimally and equitably to patients. First, health care systems need to enhance virtual care infrastructure and workflows for virtual care to realize its potential for greater efficiency. From a patient perspective, this includes ensuring that VFU technology is accessible, easy to use, and reliable. To increase accessibility, high technology (eg, video) and low technology (telephone and text) VFU solutions are required. Most Canadian studies of virtual care during the COVID-19 pandemic [17,30,49], including this study, reported greater use of telephone than video for virtual care. This was likely due to limited virtual care infrastructure, problems encountered using video, and limited video technology proficiency of patients and care providers. As virtual care technology innovates, low technology options should remain and be offered to patients to ensure equitable access, as broadband connectivity remains inconsistent across Canada [50]. However, the virtual care reimbursement structure needs to change to adequately reimburse care providers for telehealth services. Currently, Ontario's virtual care funding model limits telephone-based care by compensating physicians less for telephone visits compared to in-person visits or video visits [51]. Consequently, physicians might opt to limit their use of telephone visits, potentially leading to reduced access to care for those without high-speed internet access. Establishment of virtual care processes and workflows, along with patient education materials, are needed to seamlessly navigate patients through their virtual care appointment. From a patient perspective, this includes ensuring patients are given the option to choose their virtual visit modality, informed how to use VFU technology optimally, and what to do and who to contact if the technology fails or they require help. Support for technological issues needs to be provided by the system itself, to avoid the responsibility falling on health care providers. These strategies will enhance patient reactions and responses to virtual care, which will result in better outcomes by increasing patient confidence in their ability to use virtual care and patient confidence in the health system's ability to support them if the technology malfunctions or does not work as intended.

Limitations

This study has certain limitations. First, while theory-driven research can provide transferable context-specific understandings of phenomena (in the case of this study, VFU for cancer), inferences about transferability must be tested. In other words, while we may assume that a CMOC operates in a different setting due to the presence of similar mechanisms, this claim should be tested using locally collected primary data. Any testing is aided by the fact that CMOCs are a form of middle-range theory and are expressed in a way that permits empirical testing using primary data. Though our sample was diverse in terms of race, ethnicity, first language, etc, the participants were predominantly high-income, highly educated, and highly activated individuals. As such, our findings may not fully represent the perspective of less activated or less educated patients who may have lower health literacy, be less involved in their care, and face more challenges with virtual care. It is possible that these individuals would differ in their satisfaction with VFU and its suitability as a method of follow-up care. This highlights a need for future research to better understand how VFU could be optimized for populations who face systemic and structural barriers to the use of virtual care. Finally, it may be a limitation that a small subset of participants (6/24, 25%) did not experience an in-person appointment within 12 months of participation. Though it is likely they received an in-person appointment before this date, it is possible that they lacked experience with in-person cancer care with which to compare virtual cancer care. However, we found no obvious differences in the findings between participants who did and did not report in-person appointments during the 12-month period.

Conclusions

Patients value VFU as a part of their care; however, VFU is not working as well as it could for them. To optimize VFU, it is critical to consider the contextual factors and underlying mechanisms that influence patient perceptions of the patient centeredness and effectiveness of their care. Offering patients the choice of in-person, telephone, or video visits when possible, coupled with streamlined access to in-person care when required, is important. Prioritizing and addressing patient needs; enhancing physician online socioemotional behaviors and technology competency; and enhancing VFU functionality, usability, and processes of care and communication workflows will improve the patient centeredness of virtual care. By improving these contextual factors, VFU can be aligned with

XSL•FO
RenderX

patient needs and preferences, resulting in effective, efficient, safe, and equitable follow-up care.

Acknowledgments

The authors would like to gratefully acknowledge all participants who took the time to participate in this study as well as the patient partners who supported this study. This study was funded by the Canadian Institutes of Health Research (grant TT7 128272).

Authors' Contributions

JLB, GW, AB, JE, SM, KN, DR, JS, and RU contributed to the conceptualization of the project. JLB, SS, SB, LRS, and AB contributed to data curation. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to formal analysis. JLB acquired funding. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to study investigation. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to the methodology. JLB, SS, and SB were involved in project administration. JLB contributed to resources and software and supervised the study. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU contributed to data validation. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU prepared the original draft. JLB, SS, GW, SB, LRS, AB, JE, SM, KN, DR, JS, and RU reviewed and edited the manuscript.

Conflicts of Interest

JLB is the Associate Editor for *JMIR Cancer*. All other authors have no conflicts to declare.

Multimedia Appendix 1
Conceptual buckets, themes, and context-mechanism-outcome configurations.
[DOCX File , 23 KB - jmir_v27i1e65148_app1.docx]

References

1. Canadian cancer statistics 2023. Canadian Cancer Society. 2023. URL: https://cancer.ca/en/cancer-information/resources/publications/canadian-cancer-statistics-2023 [accessed 2024-03-19]
2. Hewitt M, Greenfield S, Stovall E. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: National Academies Press; 2005.
3. Silver JK, Baima J, Mayer RS. Impairment-driven cancer rehabilitation: an essential component of quality care and survivorship. CA Cancer J Clin 2013 Sep;63(5):295-317 [FREE Full text] [doi: 10.3322/caac.21186] [Medline: 23856764]
4. Cheville AL, Mustian K, Winters-Stone K, Zucker DS, Gamble GL, Alfano CM. Cancer rehabilitation: an overview of current need, delivery models, and levels of care. Phys Med Rehabil Clin N Am 2017 Feb;28(1):1-17. [doi: 10.1016/j.pmr.2016.08.001] [Medline: 27912990]
5. Urquhart R, Cordoba W, Bender J, Cuthbert C, Easley J, Howell D, et al. Risk stratification and cancer follow-up: towards more personalized post-treatment care in Canada. Curr Oncol 2022 May 03;29(5):3215-3223 [FREE Full text] [doi: 10.3390/curroncol29050261] [Medline: 35621651]
6. Mayer EL, Gropper AB, Neville BA, Partridge AH, Cameron DB, Winer EP, et al. Breast cancer survivors' perceptions of survivorship care options. J Clin Oncol 2012 Jan 10;30(2):158-163 [FREE Full text] [doi: 10.1200/JCO.2011.36.9264] [Medline: 22162585]
7. Paterson C, Robertson A, Smith A, Nabi G. Identifying the unmet supportive care needs of men living with and beyond prostate cancer: a systematic review. Eur J Oncol Nurs 2015 Aug;19(4):405-418. [doi: 10.1016/j.ejon.2014.12.007] [Medline: 25613370]
8. Shakeel S, Tung J, Rahal R, Finley C. Evaluation of factors associated with unmet needs in adult cancer survivors in Canada. JAMA Netw Open 2020 Mar 02;3(3):e200506 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0506] [Medline: 32142127]
9. Grunfeld E, Fitzpatrick R, Mant D, Yudkin P, Adewuyi-Dalton R, Stewart J, et al. Comparison of breast cancer patient satisfaction with follow-up in primary care versus specialist care: results from a randomized controlled trial. Br J Gen Pract 1999 Sep;49(446):705-710 [FREE Full text] [Medline: 10756611]
10. Grunfeld E, Levine MN, Julian JA, Coyle D, Szechtman B, Mirsky D, et al. Randomized trial of long-term follow-up for early-stage breast cancer: a comparison of family physician versus specialist care. J Clin Oncol 2006 Feb 20;24(6):848-855. [doi: 10.1200/JCO.2005.03.2235] [Medline: 16418496]
11. Kwan JY, Croke J, Panzarella T, Ubhi K, Fyles A, Koch A, et al. Personalizing post-treatment cancer care: a cross-sectional survey of the needs and preferences of well survivors of breast cancer. Curr Oncol 2019 Apr;26(2):e138-e146 [FREE Full text] [doi: 10.3747/co.26.4131] [Medline: 31043819]

XSL•FO
RenderX

12. Murchie P, Norwood PF, Pietrucin-Materek M, Porteous T, Hannaford PC, Ryan M. Determining cancer survivors' preferences to inform new models of follow-up care. Br J Cancer 2016 Dec 06;115(12):1495-1503 [FREE Full text] [doi: 10.1038/bjc.2016.352] [Medline: 27802453]
13. Virtual care in Canada: environmental scan from CPAC. Canadian Partnership Against Cancer. URL: https://www.partnershipagainstcancer.ca/topics/virtual-care-canada/ [accessed 2024-04-29]
14. Pham Q, Hearn J, Gao B, Brown I, Hamilton RJ, Berlin A, et al. Virtual care models for cancer survivorship. NPJ Digit Med 2020;3:113 [FREE Full text] [doi: 10.1038/s41746-020-00321-3] [Medline: 32923690]
15. Singh S, Fletcher GG, Yao X, Sussman J. Virtual care in patients with cancer: a systematic review. Curr Oncol 2021 Sep 08;28(5):3488-3506 [FREE Full text] [doi: 10.3390/curroncol28050301] [Medline: 34590602]
16. Husak L, Sovran V, Ytsma A, Comeau M. Impact of the COVID-19 pandemic on virtual care: a major shift for physicians and patients. Healthc Q 2022 Oct;25(3):11-13. [doi: 10.12927/hcq.2022.26948] [Medline: 36412522]
17. Bender JL, Scruton S, Wong G, Abdelmutti N, Berlin A, Easley J, et al. Virtual follow-up care among breast and prostate cancer patients during and beyond the COVID-19 pandemic: association with distress. Cancer Med 2024 Feb;13(4):e6948 [FREE Full text] [doi: 10.1002/cam4.6948] [Medline: 38466233]
18. Virtual care in Canada: progress and potential. Canadian Medical Association. URL: https://www.cma.ca/virtual-care-canada-progress-and-potential-report-virtual-care-task-force-0 [accessed 2024-04-29]
19. Chunara R, Zhao Y, Chen J, Lawrence K, Testa PA, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021 Jan 15;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]
20. Zon RT, Kennedy EB, Adelson K, Blau S, Dickson N, Gill D, et al. Telehealth in oncology: ASCO standards and practice recommendations. JCO Oncol Pract 2021 Sep;17(9):546-564. [doi: 10.1200/OP.21.00438] [Medline: 34319760]
21. Pawson R, Tilley N. Realistic Evaluation. Thousand Oaks, CA: Sage Publication; 1997.
22. Pawson R, Manzano-Santaella A. A realist diagnostic workshop. Evaluation 2012 Apr 22;18(2):176-191. [doi: 10.1177/1356389012440912]
23. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC Med 2016 Jun 24;14(1):96 [FREE Full text] [doi: 10.1186/s12916-016-0643-1] [Medline: 27342217]
24. Wong G, Westhorp G, Greenhalgh J, Manzano A, Jagosh J, Greenhalgh T. Quality and reporting standards, resources, training materials and information for realist evaluation: the RAMESES II project. Health Serv Deliv Res 2017;5(28):2-134 [FREE Full text] [doi: 10.3310/hsdr05280] [Medline: 29072890]
25. Traversa P. 36th annual CAPO conference: advocating for all: psychosocial oncology at the intersections of equity, diversity, and inclusion, 8-10 June 2021. Curr Oncol 2021 Jul 15;28(4):2579-2698 [FREE Full text] [doi: 10.3390/curroncol28040234] [Medline: 34287261]
26. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
27. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
28. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
29. Feldman-Stewart D, Brundage MD. A conceptual framework for patient-provider communication: a tool in the PRO research tool box. Qual Life Res 2009 Feb;18(1):109-114. [doi: 10.1007/s11136-008-9417-3] [Medline: 19043804]
30. Six domains of health care quality. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/talkingquality/measures/six-domains.html [accessed 2024-03-19]
31. Berlin A, Lovas M, Truong T, Melwani S, Liu J, Liu ZA, et al. Implementation and outcomes of virtual care across a tertiary cancer center during COVID-19. JAMA Oncol 2021 Apr 01;7(4):597-602 [FREE Full text] [doi: 10.1001/jamaoncol.2020.6982] [Medline: 33410867]
32. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and availability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
33. Ando H, Cousins R, Young C. Achieving saturation in thematic analysis: development and refinement of a codebook. Compr. Physiol 2014 Feb 20;3:03.CP.3.4. [doi: 10.2466/03.cp.3.4]
34. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA: Sage Publications; 1985.
35. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
36. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

XSL•FO
RenderX

38. Simard S, Savard J. Fear of cancer recurrence inventory: development and initial validation of a multidimensional measure of fear of cancer recurrence. Support Care Cancer 2009 Mar;17(3):241-251. [doi: 10.1007/s00520-008-0444-y] [Medline: 18414902]
39. Li CZ, Borycki EM, Kushniruk AW. Connecting the world of healthcare virtually: a scoping review on virtual care delivery. Healthcare (Basel) 2021 Oct 05;9(10):1325 [FREE Full text] [doi: 10.3390/healthcare9101325] [Medline: 34683005]
40. Abdelmutti N, Powis M, Macedo A, Liu Z, Bender JL, Papadakos J, et al. Virtual cancer care beyond the COVID-19 pandemic: patient and staff perspectives and recommendations. JCO Oncol Pract 2024 May;20(5):643-656. [doi: 10.1200/OP.23.00254] [Medline: 38266201]
41. Bui KT, Liang R, Kiely BE, Brown C, Dhillon HM, Blinman P. Scanxiety: a scoping review about scan-associated anxiety. BMJ Open 2021 May 26;11(5):e043215 [FREE Full text] [doi: 10.1136/bmjopen-2020-043215] [Medline: 34039571]
42. Caminiti C, Annunziata MA, Di Giulio P, Isa L, Mosconi P, Nanni MG, et al. Psychosocial impact of virtual cancer care through technology: a systematic review and meta-analysis of randomized controlled trials. Cancers (Basel) 2023 Mar 31;15(7):2090 [FREE Full text] [doi: 10.3390/cancers15072090] [Medline: 37046750]
43. Curran V, Hollett A, Peddle E. Training for virtual care: what do the experts think? Digit Health 2023;9:20552076231179028 [FREE Full text] [doi: 10.1177/20552076231179028] [Medline: 37274369]
44. Young K, Xiong T, Pfisterer KJ, Ng D, Jiao T, Lohani R, et al. A qualitative study on healthcare professional and patient perspectives on nurse-led virtual prostate cancer survivorship care. Commun Med (Lond) 2023 Nov 02;3(1):159 [FREE Full text] [doi: 10.1038/s43856-023-00387-6] [Medline: 37919491]
45. Turner K, Bobonis Babilonia M, Naso C, Nguyen O, Gonzalez BD, Oswald LB, et al. Health care providers' and professionals' experiences with telehealth oncology implementation during the COVID-19 pandemic: a qualitative study. J Med Internet Res 2022 Jan 19;24(1):e29635 [FREE Full text] [doi: 10.2196/29635] [Medline: 34907900]
46. Curigliano G, Banerjee S, Cervantes A, Garassino MC, Garrido P, Girard N, et al. Managing cancer patients during the COVID-19 pandemic: an ESMO multidisciplinary expert consensus. Ann Oncol 2020 Oct;31(10):1320-1335 [FREE Full text] [doi: 10.1016/j.annonc.2020.07.010] [Medline: 32745693]
47. Safavi AH, Lovas M, Liu ZA, Melwani S, Truong T, Devonish S, et al. Virtual care and electronic patient communication during COVID-19: cross-sectional study of inequities across a Canadian tertiary cancer center. J Med Internet Res 2022 Nov 04;24(11):e39728 [FREE Full text] [doi: 10.2196/39728] [Medline: 36331536]
48. Roberts CA, Aruguete MS. Task and socioemotional behaviors of physicians: a test of reciprocity and social interaction theories in analogue physician-patient encounters. Soc Sci Med 2000 Feb;50(3):309-315. [doi: 10.1016/s0277-9536(99)00245-2] [Medline: 10626757]
49. Watson L, Qi S, Delure A, Link C, Photitai E, Chmielewski L, et al. Virtual cancer care during the COVID-19 pandemic in Alberta: evidence from a mixed methods evaluation and key learnings. JCO Oncol Pract 2021 Sep;17(9):e1354-e1361 [FREE Full text] [doi: 10.1200/OP.21.00144] [Medline: 34351822]
50. Report 2: connectivity in rural and remote areas. Office of the Auditor General of Canada. URL: https://www.oag-bvg.gc.ca/internet/English/parl_oag_202303_02_e_44205.html [accessed 2023-07-19]
51. Virtual care 1: comprehensive and limited virtual care services. Ontario Ministry of Health. URL: http://www.ontario.ca/document/education-and-prevention-committee-billing-briefs/virtual-care-1-comprehensive-and-limited [accessed 2023-07-19]

Abbreviations

ASCO: American Association of Clinical Oncology
BC: breast cancer
CMOC: context-mechanism-outcome configuration
PC: prostate cancer
PM: Princess Margaret
VFU: virtual follow-up

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by K Le Du, J Li; comments to author 08.10.24; revised version received 28.10.24; accepted 07.11.24; published 03.01.25.

Please cite as:
Scruton S, Wong G, Babinski S, Squires LR, Berlin A, Easley J, McGee S, Noel K, Rodin D, Sussman J, Urquhart R, Bender JL Optimizing Virtual Follow-Up Care: Realist Evaluation of Experiences and Perspectives of Patients With Breast and Prostate Cancer J Med Internet Res 2025;27:e65148
URL: https://www.jmir.org/2025/1/e65148
doi:10.2196/65148
PMID:

XSL•FO
RenderX

©Sarah Scruton, Geoff Wong, Stephanie Babinski, Lauren R Squires, Alejandro Berlin, Julie Easley, Sharon McGee, Ken Noel, Danielle Rodin, Jonathan Sussman, Robin Urquhart, Jacqueline L Bender. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Stress Reduction in Perioperative Care: Feasibility Randomized Controlled Trial

Haridimos Kondylakis[1*], PhD; Irene Alice Chicchi Giglioli[2*], PhD; Dimitrios Katehakis[1*], MSc; Hatice Aldemir[2*], PhD; Paul Zikas[3*], PhD; George Papagiannakis[1*], PhD; Santiago Hors-Fraile[2*], PhD; Pedro L González-Sanz[2*], MD; Konstantinos Apostolakis[1*], PhD; Constantine Stephanidis[1*], Prof Dr; Francisco J Núñez-Benjumea[4*], PhD; Rosa M Baños-Rivera[5*], Prof Dr; Luis Fernandez-Luque[2*], PhD; Angelina Kouroubali[1*], PhD

[1]FORTH-ICS, Heraklion, Greece

[2]Adhera Therapeutics (United States), Wake Forest, CA, United States

[3]ORamaVR SA, Heraklion, Greece

[4]Innovation & Data Analysis Unit, Institute of Biomedicine of Seville, IBiS/Virgen Macarena University Hospital/CSIC/University of Seville, Seville, Spain

[5]Polibienestar Institute, University of Valencia, Valencia, Spain

[*]all authors contributed equally

Corresponding Author:
Haridimos Kondylakis, PhD
FORTH-ICS
N Plastira 100
Vassilika Vouton
Heraklion, GR70013
Greece
Phone: 30 2810391449
Email: kondylak@ics.forth.gr

Abstract

Background: Patients undergoing surgery often experience stress and anxiety, which can increase complications and hinder recovery. Effective management of these psychological factors is key to improving outcomes. Preoperative anxiety is inversely correlated with the amount of information patients receive, but accessible, personalized support remains limited, especially in preoperative settings. Face-to-face education is often impractical due to resource constraints. Digital health (DH) interventions offer a promising alternative, enhancing patient engagement and empowerment. However, most current tools focus on providing information, overlooking the importance of personalization and psychological support.

Objective: This study aimed to assess the viability of a DH intervention known as the Adhera CARINAE DH Program. This program is specifically designed to offer evidence-based and personalized stress- and anxiety-management techniques. It achieves this by using a comprehensive digital ecosystem that incorporates wearable devices, mobile apps, and virtual reality technologies. The intervention program also makes use of advanced data-driven techniques to deliver tailored patient education and lifestyle support.

Methods: A total of 74 patients scheduled for surgery across 4 hospitals in 3 European countries were enrolled in this study from September 2021 to March 2022. Surgeries included cardiopulmonary and coronary artery bypass surgeries, cardiac valve replacements, prostate or bladder cancer surgeries, hip and knee replacements, maxillofacial surgery, and scoliosis procedures. After assessment for eligibility, participants were randomized into 2 groups: the intervention group (n=23) received the Adhera CARINAE DH intervention in addition to standard care, while the control group (n=27) received standard care alone. Psychological metrics such as self-efficacy, self-management, and mental well-being were assessed before and after the intervention, alongside physiological markers of stress.

Results: The intervention group demonstrated significant improvements across several psychological outcomes. For example, Visual Analogue Scale Stress at the hospital improved at admission by 5% and at hospital discharge by 11.1% and Visual Analogue Scale Pain at admission improved by 31.2%. In addition, Hospital Anxiety and Depression Scale Anxiety after surgery improved by 15.6%, and Positive and Negative Affect Scale-Negative at hospital admission improved by 17.5%. Overall, patients in the intervention study spent 17.12% less days in the hospital. Besides these individual scores, the intervention group shows more

XSL•FO
RenderX

positive relationships among the psychological dimensions of self-efficacy, self-management, and mental well-being, suggesting that the CARINAE solution could have a positive effect and impact on the reduction of stress and negative emotions.

Conclusions: Our results provide an important first step toward a deeper understanding of optimizing DH solutions to support patients undergoing surgery and for potential applications in remote patient monitoring and communication.

Trial Registration: ClinicalTrials.gov NCT05184725; https://clinicaltrials.gov/study/NCT05184725

International Registered Report Identifier (IRRID): RR2-10.2196/38536

(J Med Internet Res 2025;27:e54049) doi:10.2196/54049

KEYWORDS

CARINAE; digital health; patient empowerment; stress and anxiety management; mHealth; virtual reality; feasibility; perioperative; randomized controlled trial; surgery; risk; patient empowerment; wearable devices; patient education; mobile app; psychological; self-efficacy; self-management; well-being; patient monitoring

Introduction

Patients undergoing surgical operations commonly experience symptoms of severe stress, anxiety, and fear due to the potentially threatening nature of surgeries [1,2]. Family caregivers, who play a crucial role in caring for these patients, also face emotional distress and physical challenges [3]. Surgeons and health care professionals use various stress coping strategies, as stressors can affect surgical performance and lead to complications [4]. Psychological support and patient education have proven effective in reducing stress and anxiety in surgical settings [5,6]. Providing information about the surgical procedure is essential for both patients and caregivers, as it helps decrease anxiety levels and surgical complications [7,8]. Patient empowerment, where patients actively participate in managing their diseases, leads to better self-care management and improved outcomes in terms of satisfaction, cost, health status, and function [9,10]. Addressing caregiver strain is also vital for pediatric patients [11] and those with special health care needs [12].

Being in unfamiliar surroundings and facing preoperative requirements can add to feelings of confusion and stress. It is common for individuals anticipating surgery to experience uncertainty, fear, hesitation, and anxiety. These emotions can affect patients' well-being and their ability to adhere to instructions throughout their perioperative journey. Research shows [13] that patient experience is closely tied to emotional health and surgical outcomes. Failure to address anxiety and stress adequately can lead to unnecessary discomfort, prolonged hospital stays, higher costs, and suboptimal clinical results.

Preventing presurgical anxiety can significantly contribute to positive outcomes in terms of health and well-being. Various strategies and techniques have been used to manage preoperative stress and anxiety, ranging from simple methods like listening to music and basic relaxation techniques to more involved approaches such as providing information, consulting with nurses, and using advanced interventions like patient education programs and modern information and communication technology tools [14,15].

For instance, allowing preoperative patients the opportunity to listen to music before surgery can prove to be an effective intervention in reducing anxiety and aiding patients in coping with what could potentially be a challenging or stressful procedure. Music has demonstrated its ability to alleviate anxiety among preoperative patients on numerous occasions [8]. However, it is noteworthy that only a small fraction of patients receive adequate stress relief support before undergoing surgery.

Digital health (DH) interventions have significantly supported enhancing health condition awareness and mental health management, using both nonimmersive systems, for example, mobile apps, and immersive systems, for example, virtual reality (VR) [16]. These interventions have shown promise in supporting patients in managing anxiety, stress, and pain [17-19]. Especially for VR, studies [20-23] have demonstrated benefits in various health care areas, including stress and pain reduction [23,24], medical practitioner training, patient counseling, cognitive rehabilitation, physical therapy, dentistry, mental health management, and surgery, as well as in managing pain and anxiety in pediatric patients [21,22,25]. VR's distractive properties make it an efficient tool for stress relief and pain management [26]. In addition, VR has been shown to reduce perceptions of anxiety in preoperative patients [27].

On the other hand, DH interventions generate vast amounts of data that can be used for personalization through artificial intelligence techniques [28-33]. These techniques capitalize on remote activity recognition and monitoring in order to provide recommendations, enabling the personalization of patient interventions based on their unique behavioral and health needs [34,35]. Perioperative stress varies greatly among patients, depending on the severity of their illness and the type of surgery required. Personalized interventions targeting stress management should take these individual differences into account, incorporating health behavior theories. Health Recommender Systems, like the I-Change model, guide the personalization of educational and behavioral interventions [36,37].

Independent of the progress in the domain, the feasibility of combining VR with mobile-based technologies as the primary means of delivering a DH intervention remains poorly understood [38]. This study aims to explore the feasibility of a DH intervention named CARINAE, which combines evidence-based perioperative stress management, anxiety, and pain relief techniques grounded in behavioral science. The intervention leverages mobile and VR technologies to help patients manage stress and anxiety during surgery, promoting healthy recovery, extensively reported already in another paper [39]. Through self-reported measurements and a control group,

XSL•FO
RenderX

the study assesses the impact of the CARINAE platform on perioperative stress, aiming to determine the effectiveness of the intervention.

Methods

Overview

In this section, we summarize the design and the protocol followed for the intervention. For more information on the details of the protocol and the technological solution, the interested reader is forwarded to the relevant paper [39].

Ethical Considerations

This study was approved by the Institutional Review Board of the four hospitals participating in the study (Hospital Universitario Reina Sofia, Spain; Instituto di Ricovero e Cura per Anziani, Italy; Sant Joan de Déu Hospital, Spain; and Fundació Parc Taulí, Spain) and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. Additionally, measures were taken to ensure the confidentiality and anonymity of patient data throughout the study. The study posed no significant risks to participants, and they were free to withdraw at any time without affecting their standard of care. Finally, no financial compensation was provided to the participants of the study.

Study Setting

The study, following ethical boards' approval, was conducted at the following European hospitals: (1) Hospital Universitario Reina Sofia (SAS; Spain)—Cardiothoracic Surgery Department; (2) Instituto di Ricovero e Cura per Anziani (INRCA; Italy)—Urology Department; (3) Sant Joan de Déu Hospital (SJD; Spain)—Orthopedics and Traumatology Department for Children; and (4) Fundació Parc Taulí (Parc Taulí; Spain)—Orthopedics and Traumatology Department for Adults. A fifth hospital had granted ethical approval to conduct the study; however, due to external reasons, no patients were recruited from this hospital, and so it did not participate.

The recruitment process started in September 2021 and finished in March 2022.

Clinical investigators prescreened the eligibility of the participants that they had available in their pool of participants that were proposed for one of the surgeries in the inclusion criteria that are reported in the sequel. Whenever a potentially eligible participant was identified, he or she was invited to the study, either by phone or during the routine consultation, whatever was more convenient. As soon as the patient showed interest in taking part in the study, she or he was referred to the research coordinator of the study who facilitated him or her by providing the patient information letter and the informed consent form and who solved any questions and concerns that the patient might have. Upon the signature of the informed consent, the patient was considered recruited for the trial.

Study Design

The CARINAE DH platform was tested in a multicentric trial conducted in 4 clinical settings across two European countries. The trial used a stratified randomized controlled design and aimed to address two main research questions.

RQ1: The first question seeks to determine the extent to which CARINAE impacts patients' stress, anxiety, and pain levels compared to those who receive the standard of care only. Additionally, the study investigated the effects on secondary outcomes, such as well-being and overall quality of life. As a side effect, the study investigated the impact of CARINAE on caregivers' stress and anxiety based on the groups they were assigned to.

RQ2: The second research question focuses on evaluating the overall usability of the CARINAE solution based on feedback and experiences from patients, caregivers, and health care professionals involved in the trial.

Eligibility Criteria

Inclusion Criteria

Participants aged 12 to 65 years who underwent various surgeries, including cardiopulmonary bypass surgery, coronary artery bypass surgery, cardiac valve replacement, prostate, kidney, or bladder cancer surgery, hip or knee replacement, maxillofacial surgery, orthognathic surgery, or scoliosis. Adult participants were required to have an Android smartphone and demonstrate basic digital literacy, while children's caregivers also needed to have an Android smartphone and basic digital literacy.

Exclusion Criteria

Participants who could not provide informed consent, communicate effectively in the native language, demonstrate basic digital literacy, exhibited symptoms of dementia, had allergies to dedicated wearable materials like steel and silicone, were pregnant, or were already enrolled in another clinical trial.

Interventions

Overview

The eligible participants were randomly allocated to either the experimental group or the control group using block randomization with a block size of 4. The randomization process was facilitated through the Sealed Envelope web-based tool [31]. The DH solution consisted of three distinct components: a mobile app, a VR component, and a clinical web application. Participants in the experimental group used the first and second components, while the third component was exclusively used by health care professionals for those patients. Both the control and experimental groups underwent the same visit schedule.

Control Group

The control group received standard care, which included four visits with the health care provider. During these visits, patients in the control group received instructions on diet and healthy lifestyle habits. In current health care settings, it is not common to provide patients with stress and anxiety relief support during the perioperative period. Assessments for the control group took place during the following visits: (1) the initial visit, where the health care provider communicated the need for surgery to the patient (2-4 weeks before the surgery); (2) hospital admission,

XSL•FO
RenderX

which occurred 1-3 days before the surgery; (3) hospital discharge, approximately 1 week after the surgery; and (4) remote follow-up 14 days after the surgery. After each visit, patients were administered several questionnaires capturing primary and secondary outcomes, as well as covariates.

Experimental (Intervention) Group

The participants in the experimental group received the digital solution CARINAE during the first visit, along with training on how to use the tool effectively. They were allowed to take CARINAE home and use it as frequently as desired. Following each of the 4 standard care visits, the experimental group completed the same questionnaires as the control group.

Sample Size Calculation

The study was a feasibility clinical trial, with the number of participants established before the beginning of the project by the clinics. According to the study protocol reported by Kondylakis et al [39], participants were randomized to achieve a balance between the 2 groups (intervention and control groups) according to the type of surgery and baseline characteristics (covariates). Considering a 1:1 random allocation, significance level of .05 (2-sided), and 80% power, 60 participants were needed to detect a 10% difference in stress, anxiety, and pain between the intervention and control groups. Sample size calculations are estimated using G*Power (version 3.1.9.2; University of Dusseldorf).

Outcome Variables

The primary outcome variables in this study were stress, anxiety, and pain, which were measured using paper-and-pencil questionnaires. The secondary outcome variables included overall quality of life, emotional status, mental well-being, self-efficacy perception, and patient activation during and after the hospital stay.

For assessing the primary outcome variables, the following questionnaires were used after each standard care visit and administered on paper:

- Patients' and caregivers' self-reported stress measured using a Visual Analogue Scale (VAS) at baseline, admission for surgery, hospital discharge, and 2 weeks after surgery [40].
- Patients' self-reported pain measured using a VAS at baseline, admission for surgery, hospital discharge, and 2 weeks after surgery [41].
- Patients' Hospital Anxiety and Depression (HADS) measured at admission for surgery, hospital discharge, and 2 weeks after surgery [42].

For assessing the secondary outcome variables, the following questionnaires were used after each standard care visit and administered on paper:

- Patients' health-related quality of life measured using the EQ-5D-3L questionnaire at baseline, admission for surgery, and clinical discharge [43].
- Patients' emotional status measured using the Positive and Negative affect Scale (PANAS) at baseline, admission for surgery, hospital discharge, and 2 weeks after surgery [44].
- Patients' and caregivers' mental well-being measured using the "Short Warwick Edinburgh Mental Well-Being Scale" (SWEMWBS) at baseline and 2 weeks after surgery [45].
- Patients' and caregivers' self-efficacy measured using the General Self-Efficacy in short form (GSE) questionnaire at baseline and 2 weeks after surgery [46].
- Patients' activation status measured using the Patient Activation Measure short form (PAM-13) at baseline, admission for surgery, and 2 weeks after surgery [47].

All these questionnaires have proven to be reliable and effective in measuring the respective variables.

Results

Overview

The diagram of the study is shown in Figure 1. From September 2021 to March 2022, a total of 74 patients were assessed for eligibility and 24 patients were excluded according to the following: (1) not meeting inclusion criteria (n=2); (2) declined to participate (n=8); (3) COVID-19 urgency (n=4); and (4) hospital acute admission (n=10). A total of 50 patients participated in the study, 23 patients were allocated to the CARINAE group (intervention group) and 27 participants were allocated to the control group. In total, 10 patients dropped out of the study (5 patients in the intervention group and 5 patients in the control group). In addition, 1 control group patient died during the surgery. A total of 39 patients finalized the trial and their data have been analyzed (21 patients in the control group and 18 patients in the intervention group; see the CONSORT [Consolidated Standards of Reporting Trials] flowchart in Figure 1 and CONSORT checklist in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of the trial.

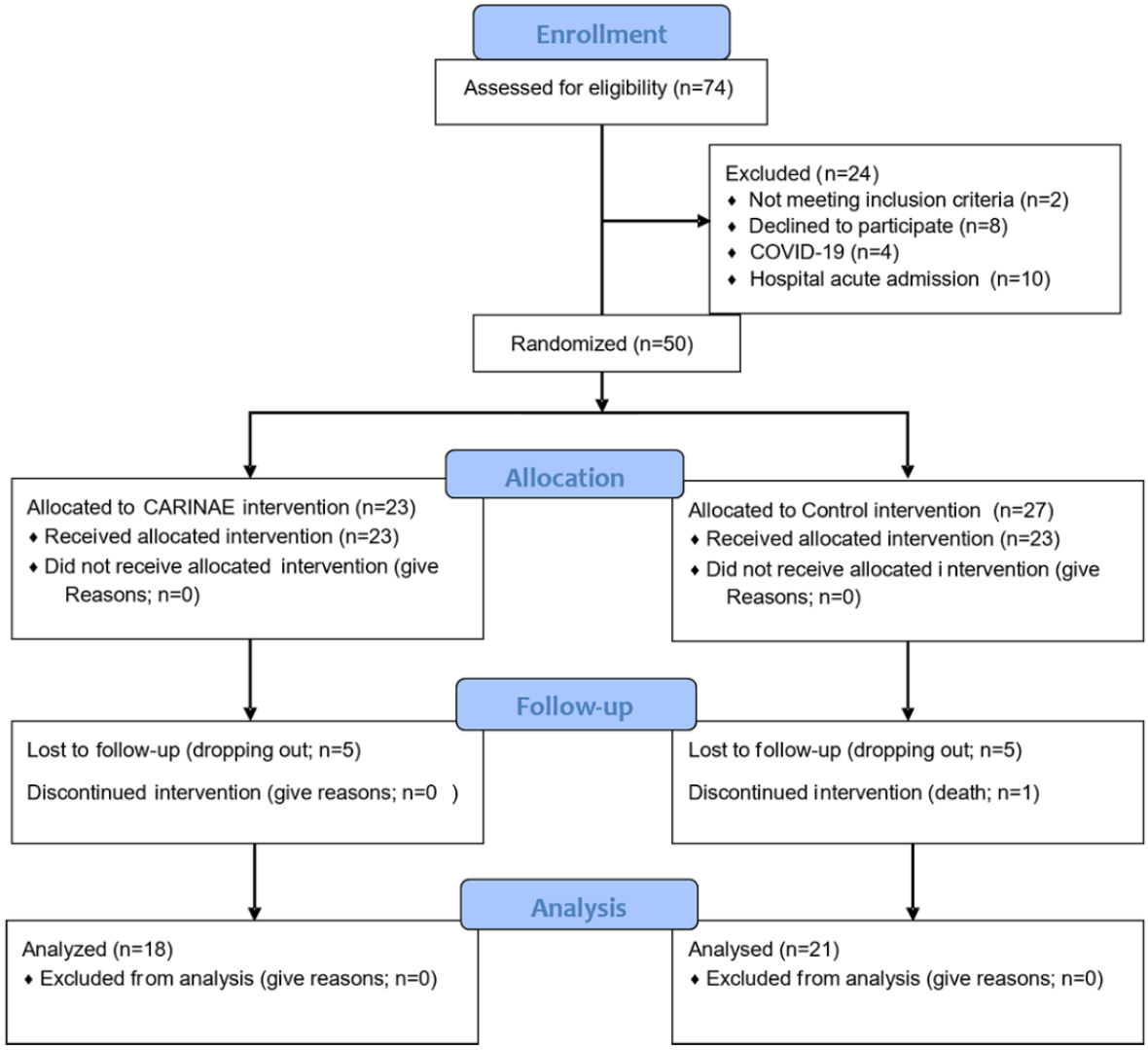

The intervention group consisted of 18 patients (10 male and 8 female patients; mean age 45.38, SD 26.2, range 12-91 years) and the control group consisted of 21 patients (11 male and 10 female patients; mean age 56.66, SD 27, range 12-91 years). Furthermore, intervention group surgeries included 1 patient undergone coronary bypass, 1 cardiac valve replacement, 4 scoliosis, 2 hip and 2 knee replacements, 3 prostate cancer, 2 bladder cancer surgeries, and 3 maxillofacial surgery types (see Multimedia Appendix 2 for sociodemographic characteristics).

The control group consisted of 3 patients undergone cardiac valve replacement, 2 scoliosis, 2 hip replacements, 4 knee replacements, 2 prostate cancer, 4 bladder cancer, and 4 maxillofacial surgery types (see Multimedia Appendix 2 for sociodemographic characteristics). The baseline table highlights a potential bias with regard to the age of the participants in both intervention and control groups (ca. 10 years), see the impact of that in the Discussion/Limitations section.

In addition, 22 caregivers participated in the study (11 in the control group and 11 in the intervention group). A total of 5 caregivers dropped out of the study (4 in the control group and 1 in the intervention group). A total of 17 caregivers finalized the trial and their data have been analyzed. This difference in drop out might be resulting from the lack of incentive to stay in the control group.

Finally, 12 health care professionals answered the three questionnaires (VAS, SWEMWBS, and Caregiver GSE), and their data were analyzed; 5 health care professionals belonged to INRCA—Italy, 4 to the SAS—Cordoba, Spain, 2 to the HSJD—Barcelona, Spain and 1 to the Parc Taulí, Spain.

Descriptive Statistics

Multimedia Appendix 2 presents descriptive statistics for the most important sociodemographic characteristics such as age, sex, digital expertise, educational level, type of medication, type of surgery, and previous comorbidities. Most of the participants had basic (n=13) or advanced digital skills (n=17) and 8 of them considered themselves experts. Regarding previous comorbidities, 11 of them had cancer in the past, 7 of them had cardiovascular disease, 1 had diabetes mellitus, 1 had renal disease, and 4 of them had mental diseases.

Clinical Outcome Parameters

Regarding the clinical variables, descriptive statistical metrics along with results of univariate analysis between intervention and control groups by *t* test method are presented in Multimedia Appendix 3.

XSL•FO
RenderX

Among all the variables in Multimedia Appendix 3, the VAS values for stress were lower for the intervention groups in all the overall cases (prehospitalization, hospitalization, and discharge) but they did not reach significance. There is only one that demonstrates statistical differences between control and intervention groups: HADS Depression Score during hospital admission in Parc Taulí. For the rest variables, the hypothesis of equal mean between intervention and control groups cannot be rejected due to the high *P* value outcome of the statistical tests. In addition, the mean VAS Stress, VAS Pain, HADS Depression, HADS Anxiety, PANAS Positive, and PANAS Negative score evolution throughout the 4 hospital visits follow-up period are illustrated in Figures 2-7.

Figure 2. Evolution of mean VAS Stress score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). VAS: Visual Analogue Scale.

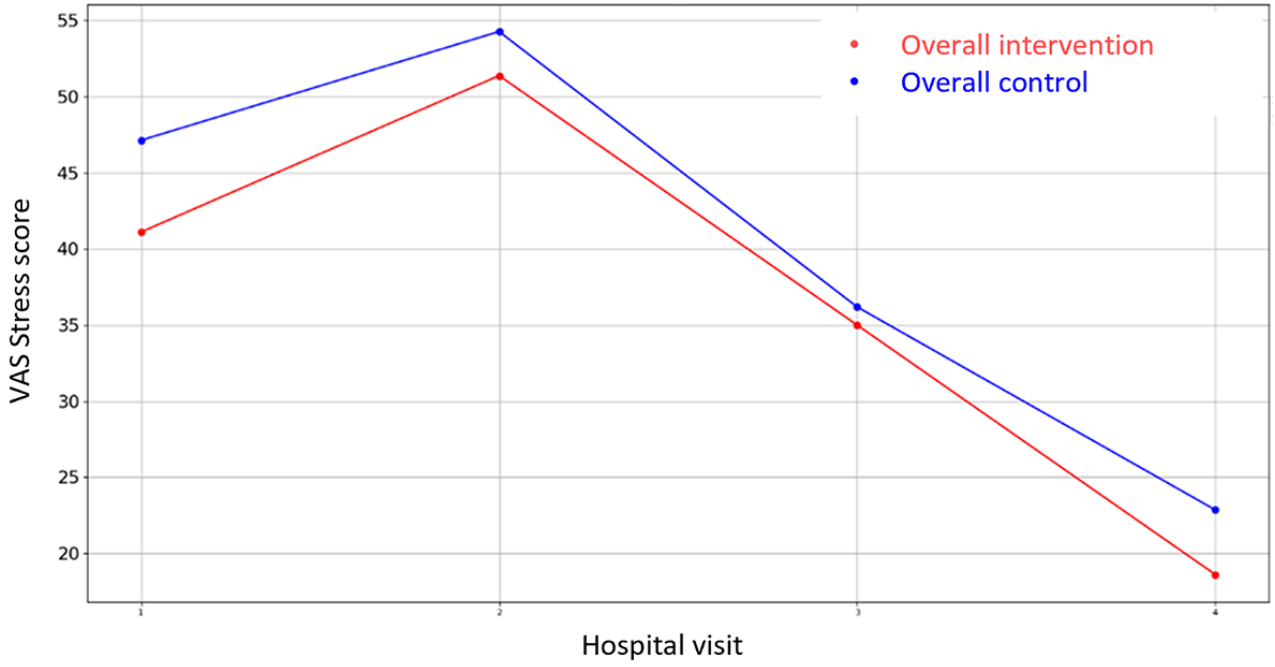

Figure 3. Evolution of mean VAS Pain score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). VAS: Visual Analogue Scale.

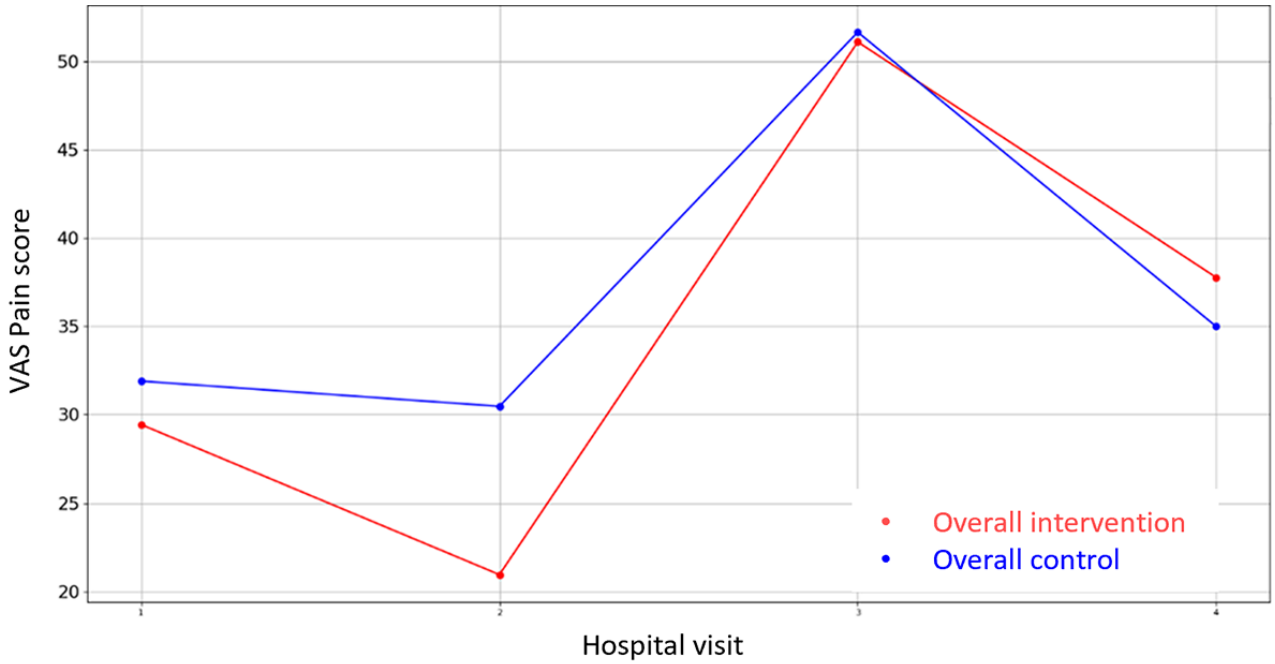

XSL•FO
RenderX

Figure 4. Evolution of mean HADS Depression score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). HADS: Hospital Anxiety and Depression Scale.

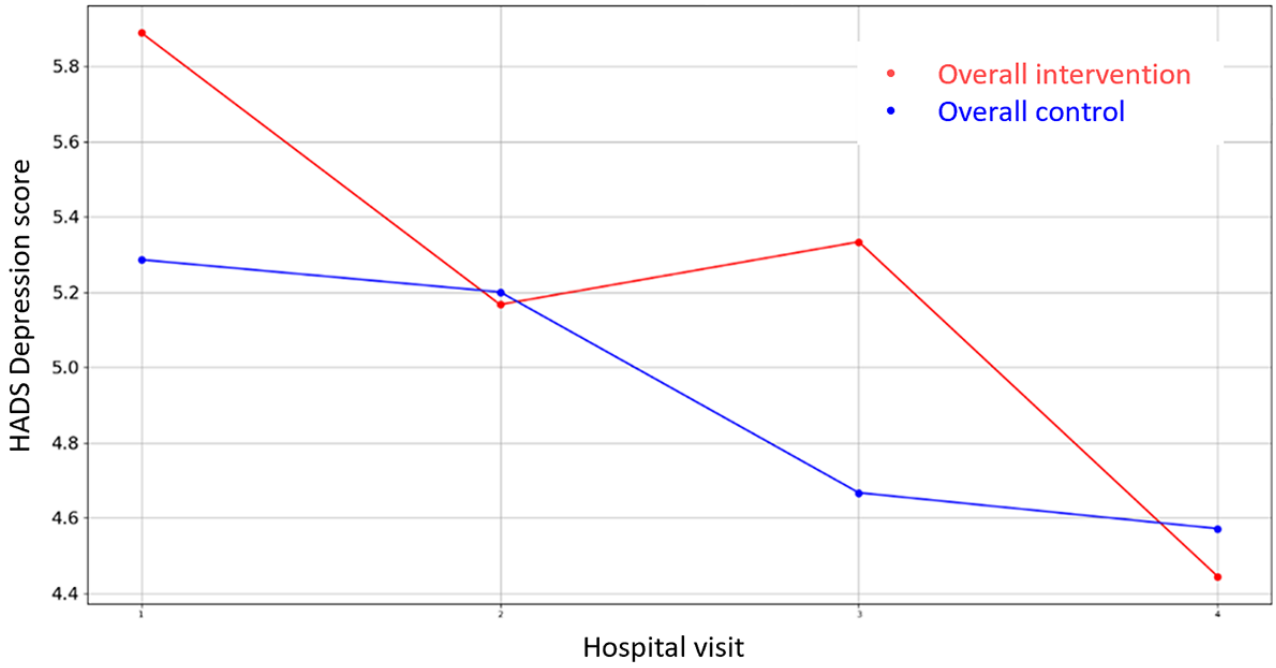

Figure 5. Evolution of mean HADS Anxiety score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). HADS: Hospital Anxiety and Depression Scale.

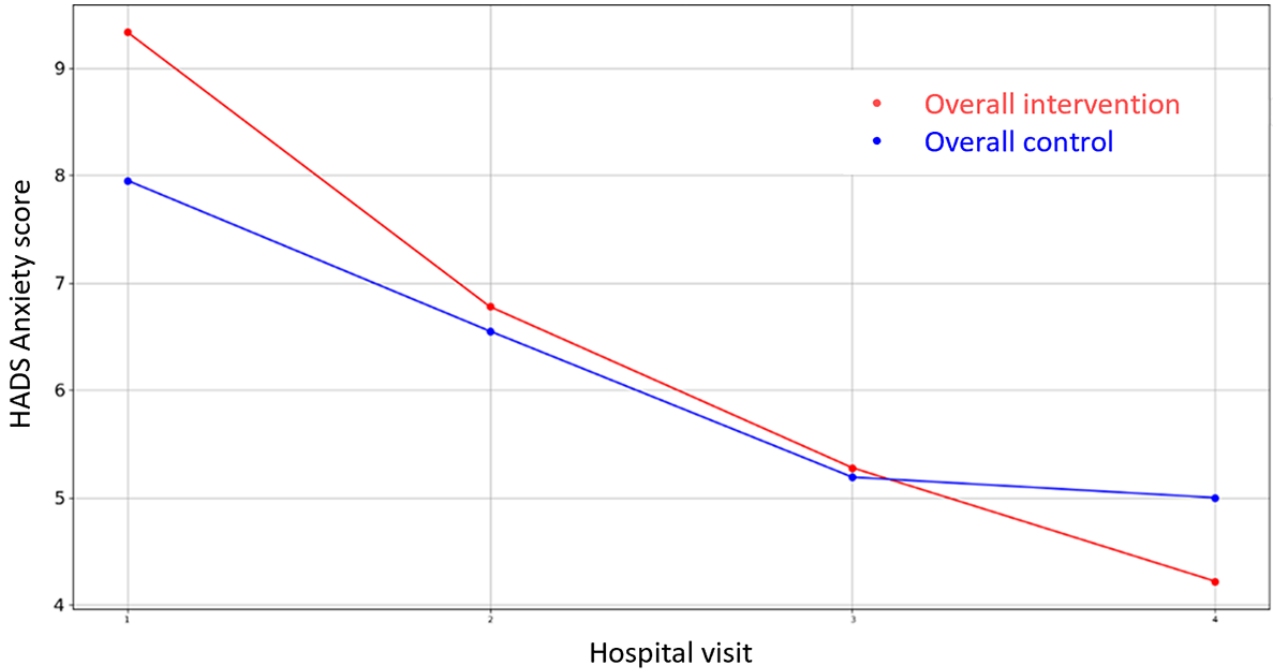

XSL•FO
RenderX

Figure 6. Evolution of mean PANAS Positive score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). PANAS: Positive and Negative Affect Scale.

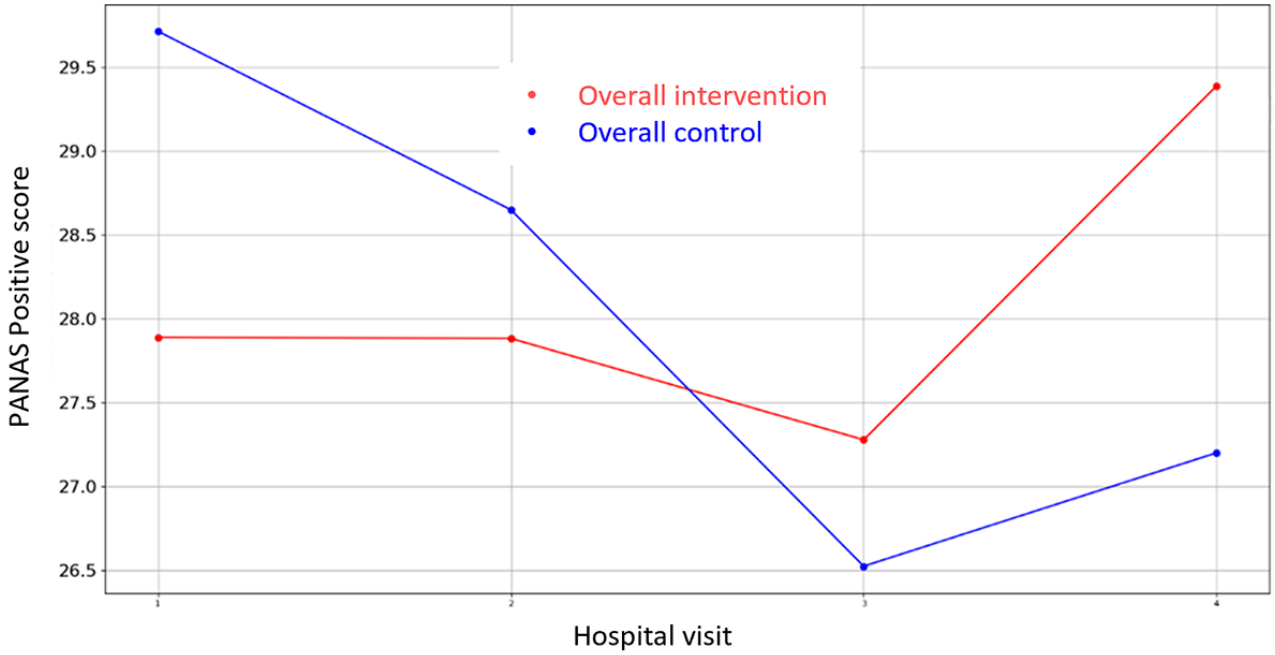

Figure 7. Evolution of mean PANAS (Positive and Negative Affect Scale) Negative score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery).

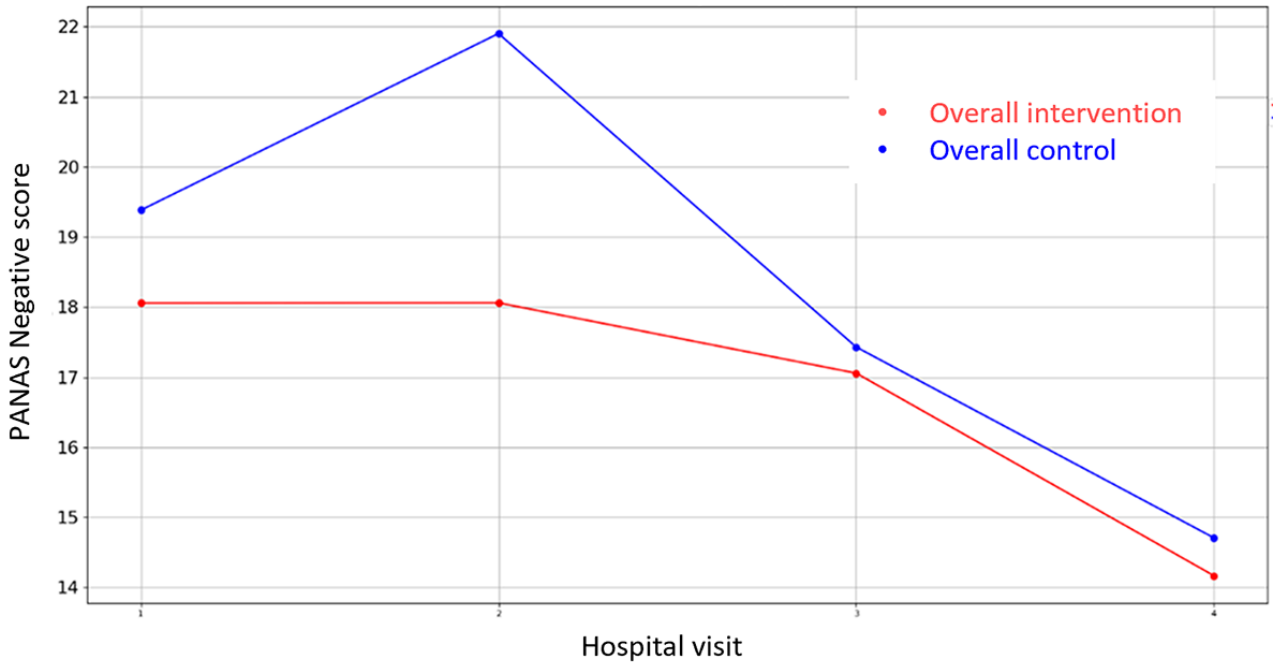

Likewise, regarding the nonclinical outcome parameters of Multimedia Appendix 4, the same statistical methodology has been designed and executed to identify the most statistically significant variables. In this case, there are three variables where the hypothesis of equal means can be rejected (P<.05), that is, PAM-13 Hospital Admission item 13, VAS Stress—hospital discharge, GSE, and the corresponding questionnaires' (PAM-13, GSE, SWEMWBS, QOLmean, VAS Stress Caregiver, SWEMWBS Caregiver, and GSE Caregiver) score evolution graphics are presented in Figures 8-16.

XSL•FO
RenderX

Figure 8. Evolution of mean PAM-13 score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). PAM-13: Patient Activation Measure 13.

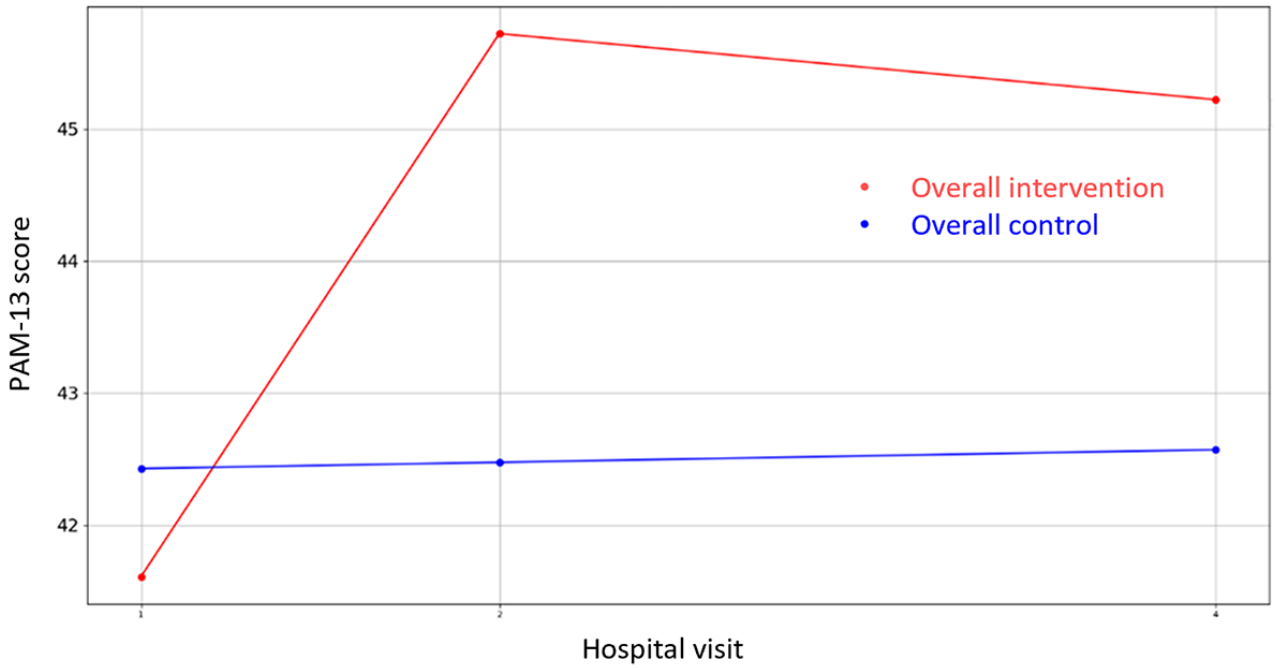

Figure 9. Evolution of mean GSE score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). GSE: General Self-Efficacy.

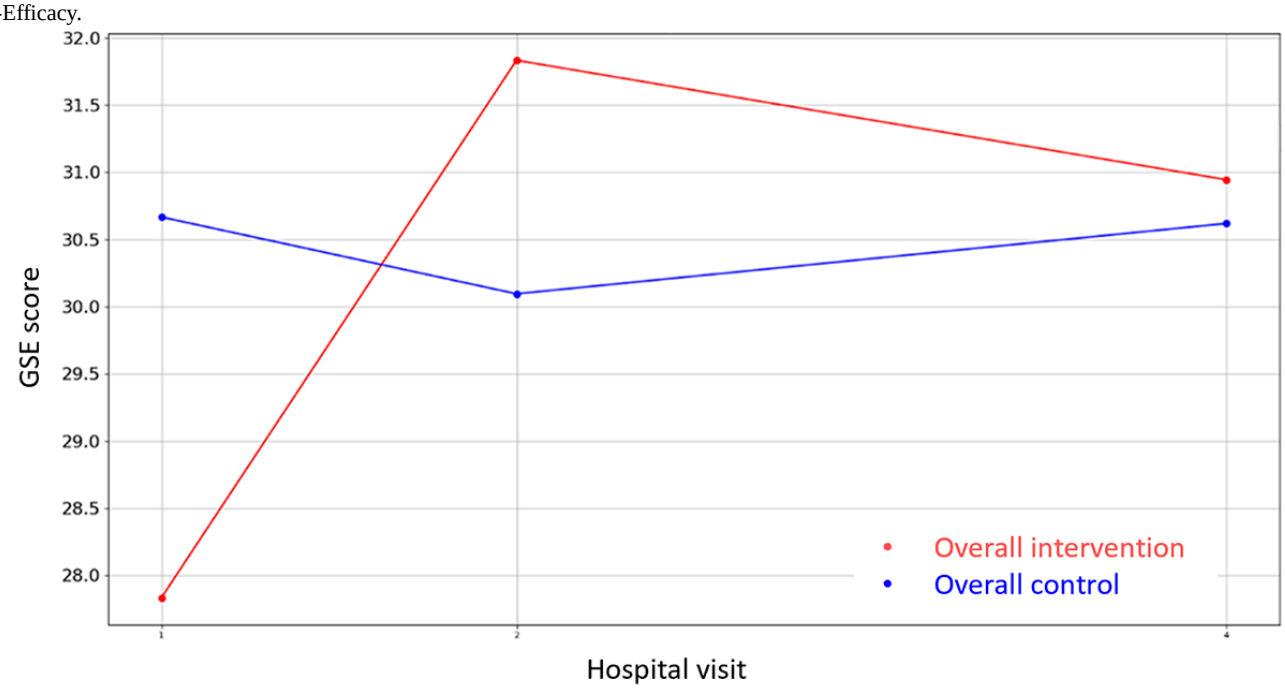

XSL•FO
RenderX

Figure 10. Evolution of mean SWEMWBS score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

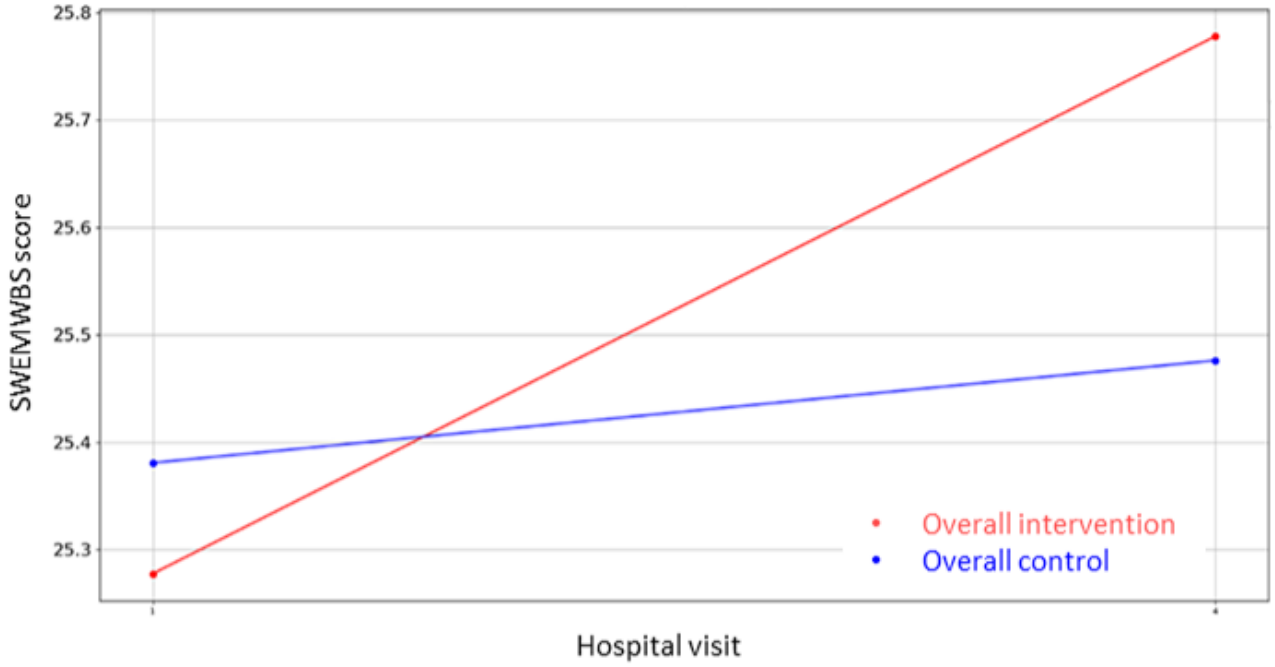

Figure 11. Evolution of mean SWEMWBS Metric score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

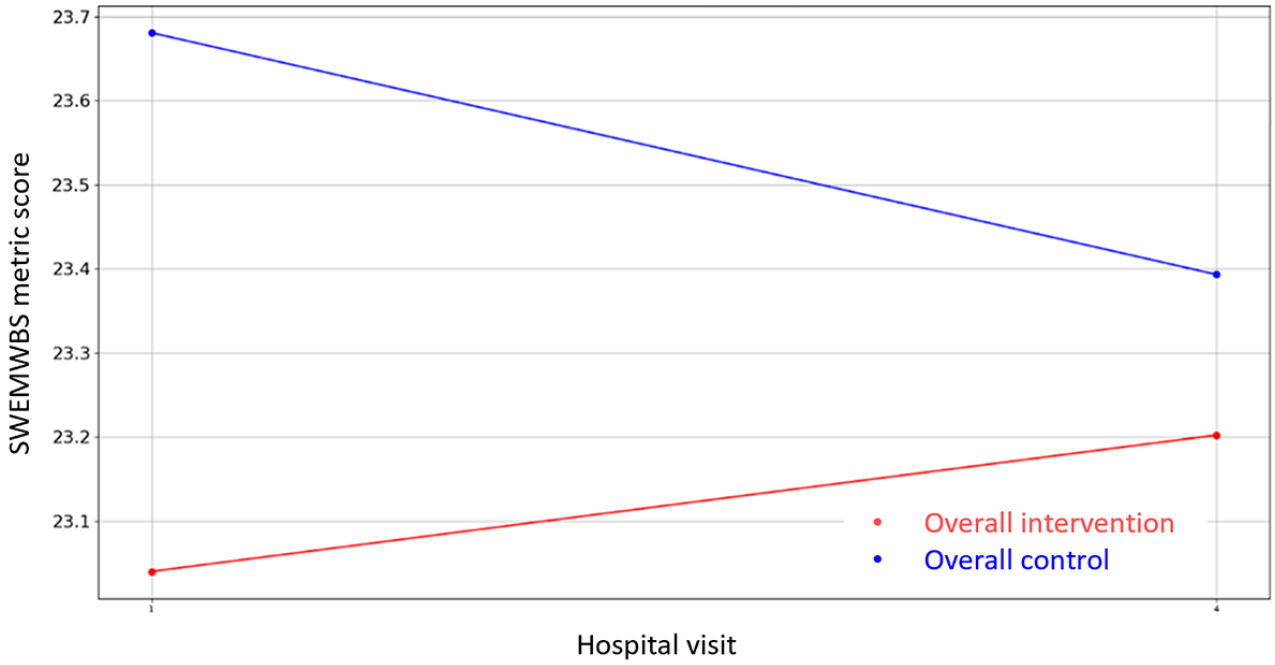

XSL•FO
RenderX

Figure 12. Evolution of mean QOL score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). QOL: Quality of Life.

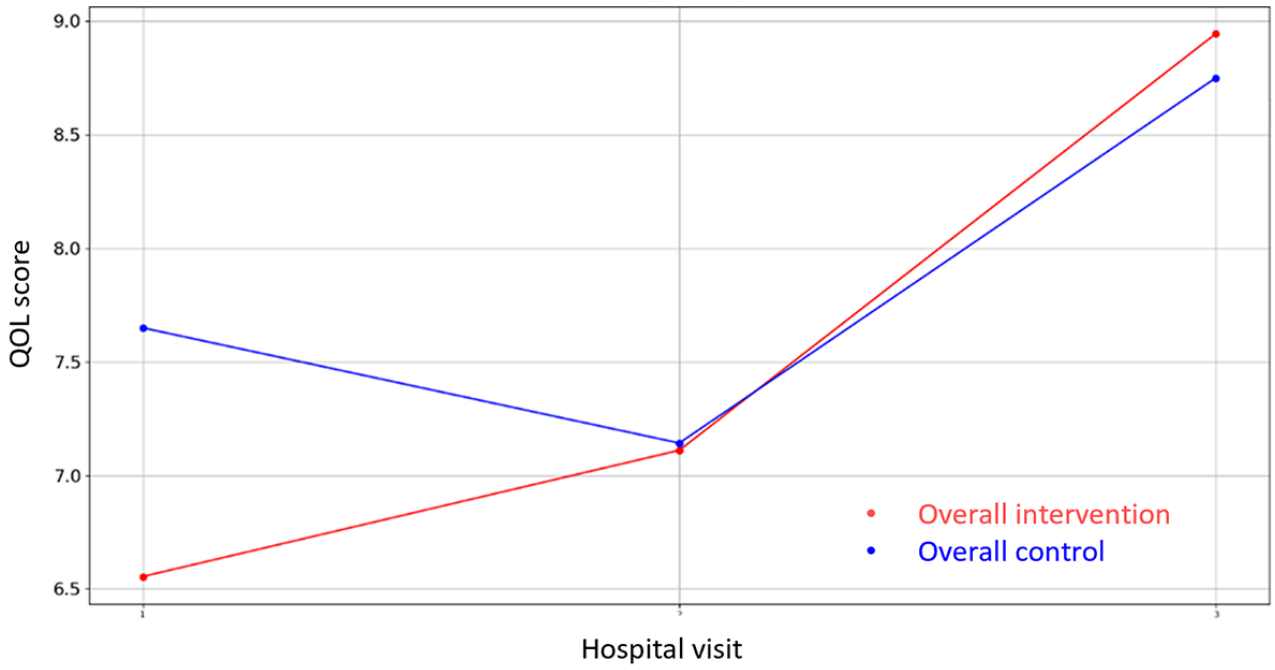

Figure 13. Evolution of mean QOL VAS score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). QOL: quality of life; VAS: Visual Analogue Scale.

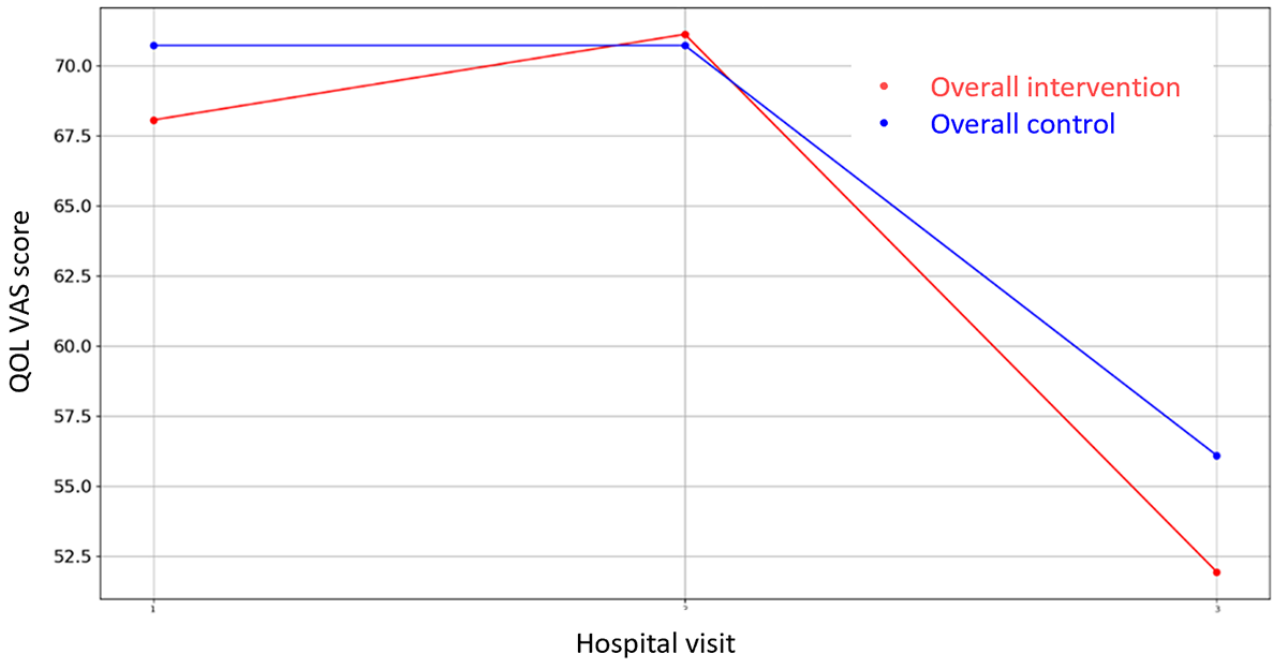

XSL•FO
RenderX

Figure 14. Evolution of Caregiver's VAS Stress score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). VAS: Visual Analogue Scale.

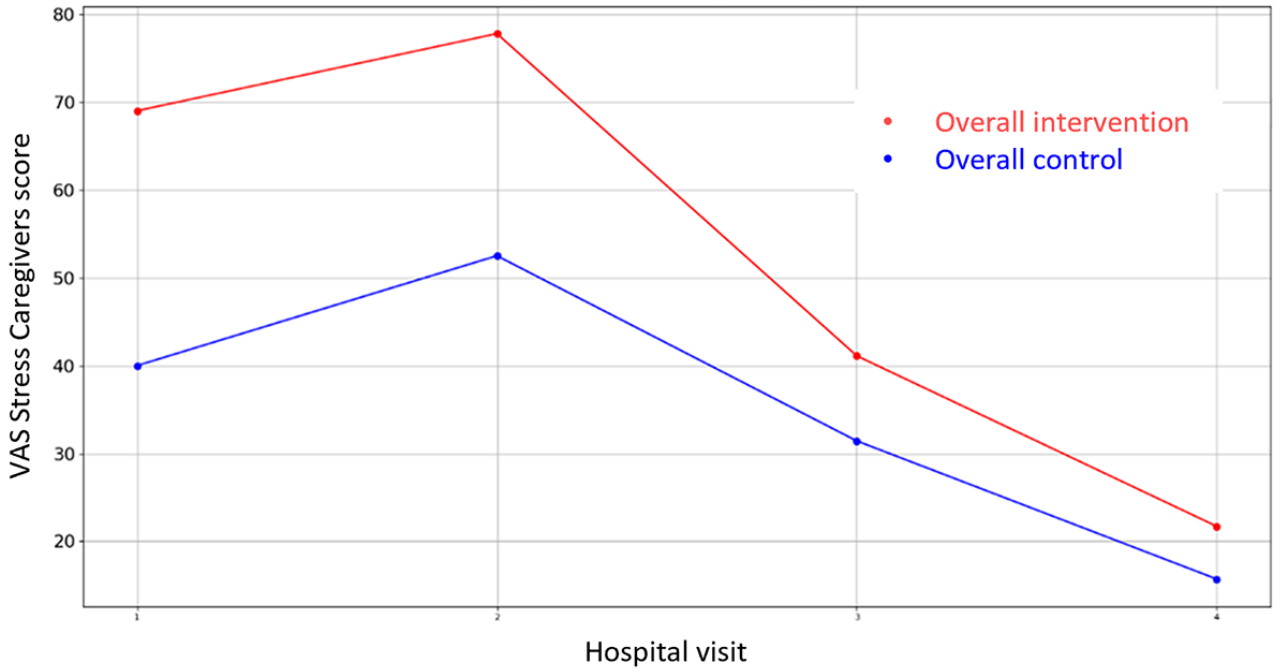

Figure 15. Evolution of Caregiver's SWEMWBS score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale.

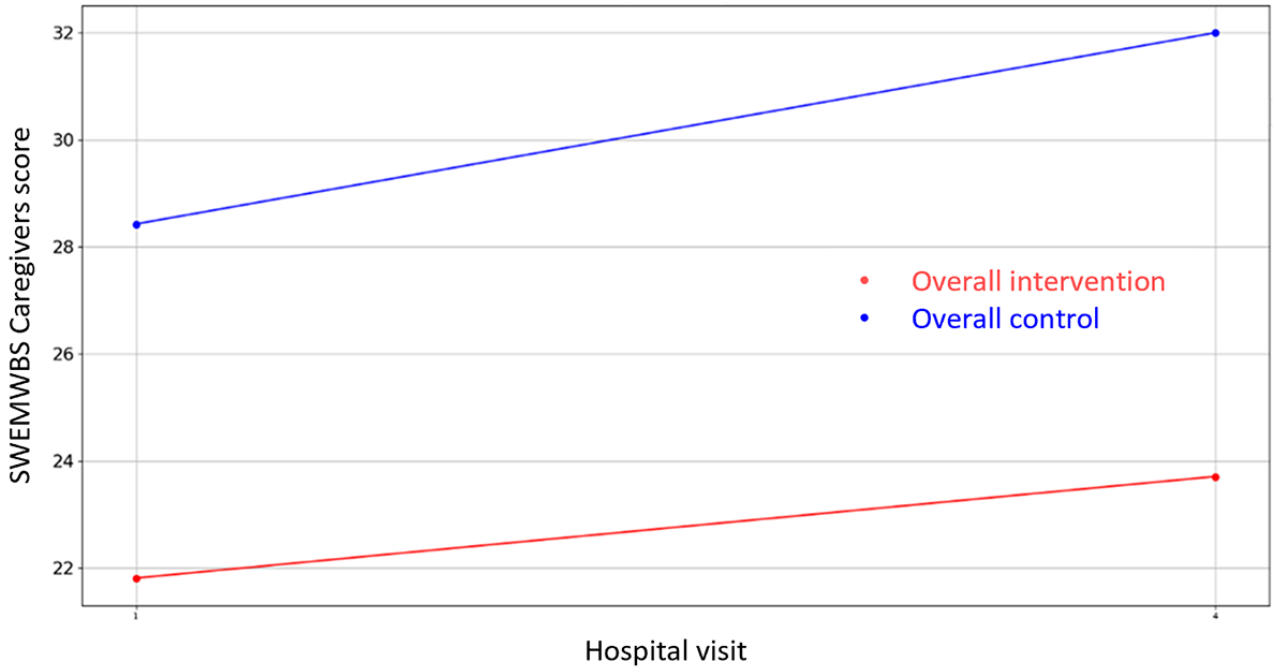

XSL•FO
RenderX

Figure 16. Evolution of Caregiver's GSE score during the 4 hospital visits (initial visit, hospital, hospital discharge, and 14 days postsurgery). GSE: General Self-Efficacy.

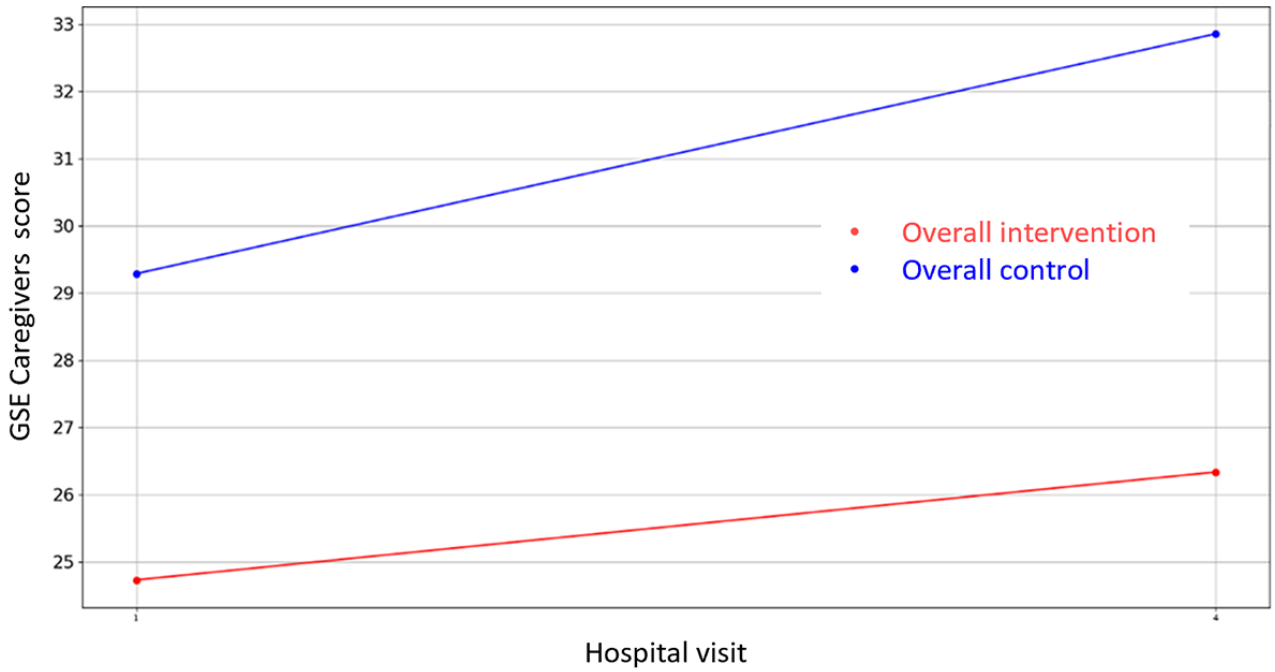

Surgery's Exploratory Analysis

ANOVA Analysis

This section aims to investigate any possible statistical effect of the type of surgery on the patients between the intervention and control groups. It uses the mixed-ANOVA statistical tests applied to each unique surgery type group. Thus, depending on the surgery, ANOVA results are as follows:

- Hip replacement surgery: $F_{29,29}$=1.38; P=.19; η_p^2=0.58
- Knee replacement surgery:$F_{29,116}$=259; P=.06; η_p^2=0.27
- Maxilofacial surgery: $F_{29,145}$=0.67; P=.89; η_p^2=0.11
- Cardiac valve replacement surgery: $F_{29,58}$=0.8, P=.73; η_p^2=0.28
- Bladder surgery: $F_{29,116}$=0.38; P=.99; η_p^2=0.08
- Prostate surgery: $F_{29,87}$=0.73; P=.82; η_p^2=0.19
- Scoliosis surgery: $F_{29,116}$=0.13; P=.007; η_p^2=0.32

The P value obtained from ANOVA analysis in scoliosis is significant (P<.05), and therefore, we conclude that there are significant differences among intervention and control patients undergone that type of surgery.

Tukey Approach

For further investigation of the scoliosis data, an essential pair-wise correlational analysis has been implemented to specify the exact pair of variables whose statistics are significantly different between the intervention and control groups. One of the most commonly used post hoc tests is Tukey's test, which allows us to make pairwise comparisons between the means of each group while controlling for the family wise error rate.

Thus, we would conclude that there is a statistically significant difference (P<.05) between the means of groups: VAS Stress (hospital discharge) – VAS Pain (postoperation day) VAS Stress (hospital discharge) – HADS Depression (enrollment) VAS Stress (hospital discharge) – HADS Depression (hospital admission) VAS Stress (hospital discharge) – HADS Depression (hospital discharge) VAS Stress (hospital discharge) – HADS Depression (postoperation day) VAS Stress (hospital discharge) – HADS Anxiety (enrollment) VAS Stress (hospital discharge) – HADS Anxiety (hospital admission) VAS Stress (hospital discharge) – HADS Anxiety (hospital discharge) VAS Pain (enrollment) – VAS Pain (postoperation day) VAS Pain (hospital discharge) – VAS Pain (postoperation day) VAS Pain (hospital discharge) – HADS Depression (enrollment) VAS Pain (hospital discharge) – HADS Depression (hospital admission) VAS Pain (hospital discharge) – HADS Depression (hospital discharge) VAS Pain (hospital discharge) – HADS Depression (postoperation day) VAS Pain (hospital discharge) – HADS Anxiety (enrollment) VAS Pain (hospital discharge) – HADS Anxiety (hospital admission) VAS Pain (hospital discharge) – HADS Anxiety (hospital discharge) HADS Depression (enrollment) – PAM (enrollment).

Linear Mixed Models

Apart from ANOVA tests, to identify any connections between the input data clustered by type of surgery, another statistical approach has been tested, the linear mixed modeling technique.

In this exploratory analysis VAS Stress, VAS Pain, HADS Depression, HADS Anxiety and PANAS questionnaires' scores have been submitted to a linear mixed models (LMM) analysis wherein "time," "group (intervention vs control)," their interaction (time × group), age, and sex have been included as fixed factors. On the other hand, "subjects" and "surgery" have been submitted as random factors. The most significant LMM model's results appear to be the ones with dependent variables the questionnaire VAS Stress and VAS Pain. The corresponding results are presented in Tables 1 and 2.

XSL•FO
RenderX

Table 1. Results of the linear mixed models for the VAS[a] Stress.

	Coefficient	SE	*P* value
Hospital admission score	8.59	5.7	.13
Hospital discharge score	–8.72	5.7	.13
Postoperation visit score	–23.46	5.7	<.001
Sex (female)	11.92	5.06	.02
Groups (intervention/control)	–8.99	4.616	.05
Age	–1.58	0.448	.001

[a]VAS: Visual Analogue Scale.

Table 2. Results of the linear mixed models for the VAS[a] Pain.

	Coefficient	SE	*P* value
Hospital admission score	–4.69	5.195	.36
Hospital discharge score	20.64	5.195	<.001
Postoperation visit score	5.513	5.195	.29
Sex (female)	0.85	4.6	.85
Groups (intervention/control)	–2.22	4.2	.59
Age	–0.28	0.349	.42

[a]VAS: Visual Analogue Scale.

Feature Importance Analysis

Feature importance refers to techniques that calculate a score for all the input features for a given model—the scores simply represent the "importance" of each feature. A higher score means that the specific feature will have a larger effect on the model that is being used to predict a certain variable. The following figures correspond to the computation of the most important factors linked to the principal questionnaires of this study (VAS Stress, VAS Pain, HADS Depression, HADS Anxiety, PANAS) using the XGBoost Regressor. The results are shown in the Multimedia Appendices 5-10.

Adherence and Engagement Analysis to CARINAE App

Adherence to VAS Pain and Stress, as well as to Wound Healing questionnaires into the CARINAE app have been analyzed to show the actions performed by the patients and the results show an overall adherence on VAS Pain of 47.23%, on VAS Stress of 68.41%, and Wound Healing of 2.5%. In the following table (Table 3), the adherence rates are also presented per hospital.

Further, the engagement table to the CARINAE app is presented in Table 4 for the various hospitals showing the total amount of interactions and patient sessions.

Table 3. Mean adherence to VAS[a] Pain, VAS Stress, and Wound Healing questionnaires into the CARINAE app.

Hospital	Days in hospital	Days in study	Completed VAS Pain	VAS Pain rate	Completed VAS Stress	VAS Stress rate	Wound Healing rate
INRCA[b]	3	33	23.33	73%	23	71.13%	3%
Parc Taulí	4	37	12.33	24.15%	15.6	30.50%	0%
Hospital Sant Joan de Déu	2.2	23.4	9.2	40.40%	20.2	79%	4.1%

[a]VAS: Visual Analogue Scale.

[b]INRCA: Instituto di Ricovero e Cura per Anziani.

XSL•FO
RenderX

Table 4. Engagement to CARINAE app in terms of patient and caregiver interactions, number of sessions, and session duration.

Hospital	Patients total interactions	Caregivers total interactions	Patients session duration (in seconds), mean (SD)	Caregivers session duration (in seconds), mean (SD)	Patients total number of sessions	Caregivers total number of sessions
INRCA[a]	526	N/A[b]	4.402 (132)	N/A	375	N/A
Parc Taulí	46	14	1.789 (54)	76 (22)	136	10
Hospital Sant Joan de Déu	264	43	3.456 (104)	2.143 (64)	190	167
SAS	34	N/A	87.683 (2630)	N/A	8	N/A

[a]INRCA: Instituto di Ricovero e Cura per Anziani.

[b]N/A: not applicable.

Discussion

Principal Findings

The main aim of this study was to test the feasibility and efficacy of a personalized stress and anxiety patient empowerment DH solution (CARINAE) on patients undergoing surgery compared to a control group. Furthermore, this study aimed to show the impact of the CARINAE solution on quality of life, emotional status, mental well-being, and self-efficacy, activation status on patients' knowledge, skills, and confidence for self-management. Besides, the CARINAE solution has been assessed by health care professionals and patients and caregivers allocated to the intervention group, in terms of engagement.

According to our results, at baseline, patients of both groups reported generally normal levels of pain and stress, with higher levels of stress. Patients undergoing coronary bypass and cardiac valve replacement surgeries indicated higher levels of stress before the surgery (VAS>80), followed by hip and knee replacement surgeries in both pain and stress (VAS>60) compared to the other patients' surgeries. Similar stress values can be observed by caregivers. Regarding depression, patients of both groups reported general normal levels but they indicated to experience anxiety (HADS score between 7 and 10). In detail, depression is primarily observed by older patients undergoing prostate and bladder cancer removal surgery, with values between the abnormal range (HADS depression abnormal range 11-21) and secondarily, with borderline values (HADS depression borderline range 8-10), by patients undergoing coronary bypass and cardiac valve replacement surgeries. In contrast, anxiety is more generalized among various patients, and the results showed that older patients undergoing an operation to remove prostate or bladder cancer report very high values, in the abnormal range (HADS anxiety >11), followed by the youngest undergoing scoliosis and maxillofacial surgeries (HADS anxiety=10), and hip and knee surgeries, as well as coronary bypass and cardiac valve replacement surgeries with borderline levels of anxiety (HADS anxiety borderline range 8-10). Regarding positive and negative affect scores, patients of both groups showed normal levels of positive affect (mean scores normal levels 33.3, SD 7.2) and higher values regarding negative affect (mean scores normal levels 17.4, SD 6.2). Regarding patients' knowledge, skills, and confidence for self-management, no patient of either group found it important to take action with respect to the situation and the self-efficacy perception resulted quite high in most patients and caregivers (GSE>25). Finally, with respect to mental well-being and quality of life, at baseline patients and caregivers of both groups present quite high rates of mental well-being and quality of life. With respect to the latter two measures, it can be emphasized that patients who are older and have undergone more delicate surgeries, such as coronary bypass and cardiac valve replacement surgeries showed lower levels than other surgeries and age. A potential bias is the difference in age between the control and intervention groups, the control group has on average 10 years more. In a small sample, this bias might have occurred inevitably, but it might be indicative of a selection bias. In future studies, we should consider the possibility of including a "control intervention" that also requires the use of some simple mobile-based technology (eg, watching educational videos on YouTube).

Looking at the relationship between the variables, overall patients show most strong associations between negative emotions, anxiety, and stress, especially in the preoperative phase (recruitment and hospital admission), which in some cases are also maintained in the postoperative phase (hospital discharge and postoperative follow-up visit), but nevertheless, the presence of a strong inverse relationship between mental well-being and anxiety at the postoperative follow-up visit suggests that as anxiety decreases, patients' mental well-being increases again.

Examining the 2 groups separately, the control group shows greater relationships between the more negative psychological dimensions, such as anxiety, stress, negative emotions, and depression, whereas the intervention group shows more positive relationships between the psychological dimensions of self-efficacy, self-management, and mental well-being, suggesting that CARINAE solution could have a positive effect and impact on the reduction of stress and negative emotions. In detail, the control group showed levels of stress, depression, anxiety, and negative emotions that remained constant throughout the entire perioperative process, and only at the end of the process could changes in the perception of greater mental and physical well-being be observed. There were positive trends in the intervention group, although not significant from a statistical point of view. The intervention group shows a greater perception of mental well-being, self-efficacy, and self-management during the entire perioperative process and not only at the end of the process. Considering the clinical

XSL•FO
RenderX

outcome parameters on pain, stress, anxiety, and depression, a significant difference between the two groups has been found in the depression subscale regarding the hip and knee surgeries, still showing values within the normal range. Although no other statistically significant differences were found, it can be seen how the values with respect to pain remain low at the time of hospitalization and how they increase after the operation and then decrease again at the follow-up visit. Pain is known to be related to stress levels. The highest postoperative pain values are perceived more in those patients who underwent a more invasive operation (eg, coronary bypass and cardiac valve replacement surgeries) than in those whose operation was less invasive (prostate and bladder cancer). As expected, as far as stress was concerned, no significant differences were found, but the values increased at hospital admission, compared to baseline, and these values decreased at discharge and more so at the follow-up visit. Looking at the individual operations, the cardiac valve replacement surgeries and hip and knee replacement surgeries on admission to the hospital are the ones that cause the most stress for patients. Finally, the values of anxiety and depression remained within the normal ranges, and with respect to the baseline, it can be observed that in patients undergoing prostate or bladder cancer removal surgery who had high levels, the levels of both parameters steadily decreased at various stages of the perioperative process. We should notice that the project and the intervention aim at reducing stress levels, and not clinical anxiety disorder or major depression. This is quite important since a narrower focus on major mental health disorders might reduce the applicability of the intervention to most of the population who undergoes surgery. We should be aware that the exploratory analyses might, due to multiple testing, be subject to type 1 error, and cautious interpretation of the analysis is warranted.

Regarding the nonclinical outcome parameters, no significant differences have been found. However, the positive effect values remain stable and in the normal range during the perioperative process, increasing more after the hospital discharge, while negative values decrease after the operation and at the follow-up visit. Regarding patients' knowledge, skills, and confidence for self-management, no patient of either group found it important to take action with respect to the situation, and the self-efficacy perception resulted higher in patients and caregivers after the surgeries than the baseline (GSE>30). Finally, with respect to mental well-being and quality of life, in the follow-up visit, patients and caregivers of both groups presented higher rates of mental well-being and lower rates of quality of life than baseline, suggesting a faster recovery process of mental well-being in terms of recovery of feeling relaxed, optimistic, thinking clearly, and dealing with problems well and a slower recovery of quality of life in terms of recovery of mobility, autonomy, and usual activities. These results are not conclusive, and a bigger study might be needed to identify the level of effectiveness of the intervention. However, major aspects deserve further attention prior to a larger study. The low awareness about the importance of self-management might be an indicator of low health literacy levels. It might be wise to consider strategies that as part of the intervention address awareness of lifestyle factors for recovery even before the intervention itself.

The founded differences between the two groups may be linked to the young age of the participants and consequently to the related hormonal and neurodevelopment changes. Indeed, these changes are currently conceptualized in terms of imbalance between systems supporting reactivity and regulation, specifically nonlinear changes in reactivity networks and linear changes in regulatory networks [48].

In addition, as identified by LMMs, stress and pain showed other statistical differences in accordance with other parameters, showing that experiences of stress seem to differ according to age and sex. The level of stress rises with aging and is higher in female patients than in male patients. In addition, both stress and pain also differed depending on the phase of the perioperative process in which patients find themselves. In detail, pain presented more effect and impact at hospital discharge, showing the highest levels of the perioperative process, and stress at the end of the entire process (at follow-up visit), showing the lowest levels of the perioperative process. Further analyzing the importance of the various features using XGBoost revealed that stress, anxiety, pain, and negative affect measures are interrelated as expected and that physical activity is also important.

Finally, according to the adherence to the psychometric questionnaires in the CARINAE app, we observed that younger and older participants were more involved in the self-evaluation regarding stress with 79% and 71%, respectively. On pain, older participants were involved more than younger with 73% of responses by the older and 40% by the younger. These data are also reflected in the use of the CARINAE app, where the greatest interactions were found among older and younger patients. These data suggest, on the one hand, a greater provision of time by older people, and on the other hand, extensive use of technology by younger people and overall high involvement in the use of CARINAE for both ages.

Comparison to Literature

Although no significant differences have been found between groups, advances in DH interventions are playing great support in enhancing awareness about health conditions and for the management of mental health by relying on both nonimmersive (eg, such as mobile apps) and immersive systems (eg, VR) [16]. These include support for patients in the management of anxiety, stress [37], and pain [18]. On one hand, mobile apps for perioperative processes are becoming a hot topic, allowing to provide psychoeducational contents, mental well-being activities to reduce pain and stress, up-to-date information, tracking personal health data, reminding and engaging patients, and communicating in a cost-effective way. On the other, VR has been used in multiple health care applications including reducing stress and pain, training medical practitioners, patient counseling, cognitive rehabilitation, physical therapy in medicine, and for diagnostic and treatment needs in dentistry, mental health management, and surgery [19]. A multiuser immersive VR system was developed and used during presurgical discussions in a prospective patient cohort undergoing cerebrovascular surgery [20]. An immersive VR intervention adopted in pediatric patients to manage pain and anxiety provided a new, easy, and cost-effective intervention

XSL•FO
RenderX

that can be applied to other painful and stressful medical procedures [49]. Pain is a highly distressing symptom for patients in all clinical settings and stress and anxiety levels influence it. VR applications have proven to be efficient in stress relief and pain management, mainly due to their distractive properties [22].

In the literature, there is a lack of understanding of the feasibility of combining VR with mobile-based technologies as the main channel of the provision of a DH intervention [36]. In this study, what is explored is the feasibility of a DH intervention that leverages the latest mobile and VR technologies within the use case of helping patients manage stress and anxiety during surgery while promoting healthy recovery.

The findings of this project, reveal the importance of addressing this type of intervention as a service design approach with a strong focus on implementation aspects. A recent guideline by the World Health Organization in DH reinforces the importance of addressing training and supporting the environment as key success aspects in DH implementations [50]. In addition, infrastructure was a key element (eg, connectivity issues). The workforce is not only crucial from a service delivery point of view, but also a key channel to facilitate patient DH literacy which in this very complex setup with multiple devices and features is especially important.

Strengths and Limitations

Several strategies and techniques have been proposed to manage preoperative stress and anxiety that can be effective in supporting patients to cope with a wide range of stressful health situations. In the current health care settings, however, it is not very common to provide patients with stress and anxiety relief support prior to a surgical procedure. Usually, VR-enhanced solutions focus only on providing informative content, neglecting the importance of patient empowerment with a more robust educational curriculum. This study has been among the first to evaluate the potential effectiveness of a comprehensive technology in reducing perioperative stress and anxiety. CARINAE provided a unique combination of endpoints and the integration of knowledge deriving from several domains, including stress or anxiety management, patient empowerment, communication with medical professionals, adaptation to illness, self-regulation and self-management, and adaptation to medical procedures. The study has been successful in terms of identifying the key impacts of such type of intervention and provided enough exploratory insights to redesign the intervention and establish a new randomized controlled trial design to provide more conclusive data on efficacy. At the same time, this study allowed us to identify implementation issues that need to be addressed prior to larger studies and implementations. For example, larger trials with more participants are advisable for future work, which will facilitate the application of ANCOVA analysis.

This solution allowed participants to receive constant feedback to improve their appraisal and coping skills in an entertaining and motivating manner. It focuses on patient empowerment through active participation in the process and is dynamically adapted according to operation type, patient preferences, needs, and medical history all the way through the preclinical phase, admission, and discharge, in a continuous and personalized way. At the same time, it facilitated effective interactions between patients and health care professionals, through user-friendly and intelligent communication. It used the spaced learning methodology to help patients understand and learn the diverse aspects of their surgical process, from presurgery requirements to recovery steps with stress management all along the process, and provides multichannel anxiety and stress relief personalized content.

CARINAE solution, combining mobile health and VR technologies with a web app, provided positive preliminary results in reducing perioperative stress and creating effective collaborations between physicians or surgeons and their patients whilst supporting them in improving their knowledge in related domains. CARINAE has shown the potential to improve physical and emotional reactions to a stressor, such as surgical operations, to increase the levels of calmness to promote a sense of well-being and to empower patients in preoperative conditions. Information provided through the platform advances and enhances health literacy and digital competence and increases the participation of the patient in the decision-making process. Integration with third-party applications has facilitated the exchange of important information between patients and physicians, as well as between personal applications and clinical health systems.

While the findings of this study are interesting and valuable, it has some limitations. First, the small sample size, as well as the high variability of the ages and surgeries might limit the generalizability of the results. Given the pragmatic nature of this feasibility study, its small sample size was anticipated, further compounded by emergent challenges amid the COVID-19 crisis. The study did not aim to have a conclusive answer in terms of effectiveness, but to gather insight on whether it can be done as suggested in related literature [51]. As such, this study provides insights to support the design of a future randomized controlled trial. While a prospective observational design was considered, its susceptibility to confounding due to temporal changes prompted the inclusion of a control group, albeit with a heightened risk of limited significance. Moving forward, future research should prioritize more robust, adequately powered clinical studies to offer comprehensive insights.

Furthermore, pain and anxiety management guidelines might differ from one clinical setting to another, thus providing limited evidence based on raw cross-comparisons between different health care providers. Finally, the different durations of each surgery and each process limited the engagement data to the DH solution. Third, no data about use have been gathered from the VR app, limiting the quantitative performance rates and engagement to the VR component.

Recommendations for Future Research, Development, and Clinical Practice

An area of future improvement that both this feasibility study and other DH solutions demonstrate the need for is long-term patient engagement. Indeed, according to the literature, as mobile health apps grow more prevalent, many have low attrition rates, reducing the relevance of data collection for the

XSL•FO
RenderX

study [51]. Potential hurdles to long-term involvement include the patients' functional condition and the workload associated with performing in-app tasks. Our findings are aligned with those studies. This is especially true for procedures such as cancer and cardiac surgeries, where the target population is older and potentially more functionally constrained than younger patient populations. To address the issue of long-term engagement, potential solutions include developing even more simplified versions of the solution that are easier to use for those unfamiliar with technology, using SMS text messaging–only modes of communication, and using shorter surveys that take less time to complete. For younger patients, on the contrary, more videos and images instead of text could engage more. Additionally, it may empower patients further if they are shown the outcomes of their data and are able to visualize and quantify in real-time, how their condition has improved. Previous studies have noted that patients provided positive feedback when presented with their study results and often noted reluctance to participate because they had not received personalized feedback from their care teams in previous app-based studies [51]. In this study, differences across sites might be due to differences in populations but also on how the solutions were promoted or introduced in each site. When integrating new technology into any health care setting, provider input may significantly streamline the integration of applications into already complex clinical workflows. Continually enhancing the experience and usability, from both the patient and provider perspectives, will allow the realization of the full potential of DH solutions for patients undergoing interventional procedures.

Conclusions

We evaluated the feasibility of using CARINAE, a comprehensive DH solution, for patients undergoing surgery. This study examined the user experience and engagement with the solution, as well as the ability to collect patient-reported outcomes. Our results provide an important first step toward a deeper understanding of optimizing DH solutions to support patients undergoing surgery and for potential applications in remote patient monitoring and communication. This study highlighted how DH may be integrated into major procedures to improve patient education, emotional management, and engagement, and serve as a versatile clinical and research tool. Future studies will aim to test CARINAE in a larger cohort to establish its potential impact on health care resource utilization.

Acknowledgments

This study was part of the STARS-EU-PCP project that has received funding from the European Union's Horizon 2020 Research and Innovation Program (grant: 727585).

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

IACG, HA, SH-F, PLG-S, FJN-B, and LF-L developed the patient view of the digital solution and coordinated clinical studies across all hospitals. In addition, they performed the initial data analysis. HK, DK, and AK developed the clinician view of the digital solution and contributed to the analysis of the results and their interpretation. RMB-R focused on the analysis of the results from the psychological perspective. GP, PZ, KA, and CS focused on the development of the virtual reality technology of the solution and contributed to the analysis of the results. All authors contributed to the writing of the manuscript, and reviewed, edited, and approved the final manuscript for submission.

Conflicts of Interest

IACG, HA, SH-F, and PLG-S are employees of Adhera Therapeutics, which owns the Adhera CARINAE DH Program. LF-L and SH-F were co-founders of Salumedia Labs, which was acquired by Adhera Therapeutics. All other authors declare no conflicts of interest.

Multimedia Appendix 1
The completed CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 1291 KB - jmir_v27i1e54049_app1.pdf]

Multimedia Appendix 2
Baseline characteristics of the control and intervention arms, both overall and per hospital.
[DOCX File , 43 KB - jmir_v27i1e54049_app2.docx]

Multimedia Appendix 3
Clinical outcome parameters of the control and intervention arms, both overall and per hospital.
[DOCX File , 42 KB - jmir_v27i1e54049_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Non-clinical outcome parameters of the control and intervention arms, both overall and per hospital.
[DOCX File , 63 KB - jmir_v27i1e54049_app4.docx]

Multimedia Appendix 5
HADS (Hospital Anxiety and Depression) Depression Feature Importance Analysis by XGBoost.
[PNG File , 21 KB - jmir_v27i1e54049_app5.png]

Multimedia Appendix 6
Panas Positive Feature Importance Analysis by XGBoost.
[PNG File , 19 KB - jmir_v27i1e54049_app6.png]

Multimedia Appendix 7
PANAS Negative Feature Importance Analysis by XGBoost.
[PNG File , 50 KB - jmir_v27i1e54049_app7.png]

Multimedia Appendix 8
VAS Pain Feature Importance Analysis by XGBoost.
[PNG File , 21 KB - jmir_v27i1e54049_app8.png]

Multimedia Appendix 9
HADS Anxiety Feature Importance Analysis by XGBoost.
[PNG File , 20 KB - jmir_v27i1e54049_app9.png]

Multimedia Appendix 10
VAS Stress Feature Importance Analysis by XGBoost.
[PNG File , 53 KB - jmir_v27i1e54049_app10.png]

References

1. Kumar A, Das S, Chauhan S, Kiran U, Satapathy S. Perioperative anxiety and stress in children undergoing congenital cardiac surgery and their parents: effect of brief intervention—a randomized control trial. J Cardiothorac Vasc Anesth 2019 May;33(5):1244-1250. [doi: 10.1053/j.jvca.2018.08.187] [Medline: 30243867]
2. Mathew JP, Machovec KA. A new paradigm of anxiolysis. J Cardiothorac Vasc Anesth 2019 May;33(5):1251-1252. [doi: 10.1053/j.jvca.2018.09.022] [Medline: 30529182]
3. Anton NE, Montero PN, Howley LD, Brown C, Stefanidis D. What stress coping strategies are surgeons relying upon during surgery? Am J Surg 2015;210(5):846-451. [doi: 10.1016/j.amjsurg.2015.04.002] [Medline: 26105802]
4. Sherman DW. A review of the complex role of family caregivers as health team members and second-order patients. Healthcare (Basel) 2019;7(2):63 [FREE Full text] [doi: 10.3390/healthcare7020063] [Medline: 31022837]
5. Kondylakis H, Bucur A, Dong F, Renzi C, Manfrinati A, Graf N. iManageCancer: developing a platform for empowering patients and strengthening self-management in cancer diseases. IEEE; 2017 Presented at: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); June 22-24, 2017; Thessaloniki, Greece p. 755-760. [doi: 10.1109/CBMS.2017.62]
6. Cheung KL, Durusu D, Sui X, de Vries H. How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. Digital Health 2019;5:2055207618824727 [FREE Full text] [doi: 10.1177/2055207618824727] [Medline: 30800414]
7. Nigussie S, Belachew T, Wolancho W. Predictors of preoperative anxiety among surgical patients in Jimma University specialized teaching hospital, South Western Ethiopia. BMC Surg 2014;14:67 [FREE Full text] [doi: 10.1186/1471-2482-14-67] [Medline: 25189274]
8. Thompson M, Moe K, Lewis CP. The effects of music on diminishing anxiety among preoperative patients. J Radiol Nurs 2014;33(4):199-202. [doi: 10.1016/j.jradnu.2014.10.005]
9. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
10. Katehakis DG, Kondylakis H, Koumakis L, Kouroubali A, Marias K. Integrated care solutions for the citizen: personal health record functional models to support interoperability. EJBI 2017;13(1):51-58. [doi: 10.24105/ejbi.2017.13.1.8]
11. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015;21(1):37-44 [FREE Full text] [doi: 10.1177/1357633X14562734] [Medline: 25475220]

XSL•FO
RenderX

12. Pilapil M, Coletti DJ, Rabey C, DeLaet D. Caring for the caregiver: supporting families of youth with special health care needs. Curr Probl Pediatr Adolesc Health Care 2017;47(8):190-199. [doi: 10.1016/j.cppeds.2017.07.003] [Medline: 28803827]
13. Kiecolt-Glaser J, Page G, Marucha P, MacCallum R, Glaser R. Psychological influences on surgical recovery: perspectives from psychoneuroimmunology. Am Psychol 1998 Nov;53(11):1209-1218. [doi: 10.1037//0003-066x.53.11.1209] [Medline: 9830373]
14. Gürsoy A, Candaş B, Güner Ş, Yılmaz S. Preoperative stress: an operating room nurse intervention assessment. J Perianesth Nurs 2016;31(6):495-503. [doi: 10.1016/j.jopan.2015.08.011] [Medline: 27931701]
15. Ibrahim MS, Khan MA, Nizam I, Haddad FS. Peri-operative interventions producing better functional outcomes and enhanced recovery following total hip and knee arthroplasty: an evidence-based review. BMC Med 2013;11:37 [FREE Full text] [doi: 10.1186/1741-7015-11-37] [Medline: 23406499]
16. East ML, Havard BC. Mental health mobile apps: from infusion to diffusion in the mental health social system. JMIR Ment Health 2015;2(1):e10 [FREE Full text] [doi: 10.2196/mental.3954] [Medline: 26543907]
17. Kondylakis H, Kouroubali A, Koumakis L, Rivero-Rodriguez A, Hors-Fraile S, Katehakis D. Designing a novel technical infrastructure for chronic pain self-management. Stud Health Technol Inform 2018;249:203-207. [doi: 10.3233/978-1-61499-868-6-203] [Medline: 29866983]
18. Kondylakis H, Hors-Fraile S, Koumakis L, Kouroubali A, Notas G, Rivero-Rodriguez A, et al. An innovative, information and communication technology supported approach, towards effective chronic pain management. In: Alternative Pain Management: Solutions for Avoiding Prescription Drug Overuse. Hershey, Pennsylvania, USA: IGI Global; 2020:21.
19. Pillai AS, Mathew PS. Impact of virtual reality in healthcare: a review. Virtual Augment Real Ment Health Treat 2019:17-31. [doi: 10.4018/978-1-5225-7168-1.ch002]
20. Sugiyama T, Clapp T, Nelson J, Eitel C, Motegi H, Nakayama N, et al. Immersive 3-dimensional virtual reality modeling for case-specific presurgical discussions in cerebrovascular neurosurgery. Oper Neurosurg (Hagerstown) 2021;20(3):289-299 [FREE Full text] [doi: 10.1093/ons/opaa335] [Medline: 33294936]
21. Wong CL, Lui MMW, Choi KC. Effects of immersive virtual reality intervention on pain and anxiety among pediatric patients undergoing venipuncture: a study protocol for a randomized controlled trial. Trials 2019;20(1):369 [FREE Full text] [doi: 10.1186/s13063-019-3443-z] [Medline: 31221208]
22. Scates D, Dickinson JI, Sullivan K, Cline H, Balaraman R. Using nature-inspired virtual reality as a distraction to reduce stress and pain among cancer patients. Environ Behav 2020;52(8):895-918. [doi: 10.1177/0013916520916259]
23. Prabhu VG, Stanley L, Morgan R. A biofeedback enhanced adaptive virtual reality environment for managing surgical pain and anxiety. Int J Semant Comput 2020;14(3):375-393. [doi: 10.1142/s1793351x20400152]
24. Prabhu VG, Stanley L, Morgan R, Shirley B. Designing and developing a nature-based virtual reality with heart rate variability biofeedback for surgical anxiety and pain management: evidence from total knee arthroplasty patients. Aging Ment Health 2024;28(5):738-753. [doi: 10.1080/13607863.2023.2270442] [Medline: 37850735]
25. Ryu JH, Park JW, Nahm FS, Jeon YT, Oh AY, Lee HJ, et al. The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: a prospective, randomized, and controlled trial. J Clin Med 2018;7(9):284 [FREE Full text] [doi: 10.3390/jcm7090284] [Medline: 30227602]
26. Nayarisseri A, Khandelwal R, Tanwar P, Madhavi M, Sharma D, Thakur G, et al. Curr Drug Targets 2021;22(6):631-655. [doi: 10.2174/1389450122999210104205732] [Medline: 33397265]
27. Hendricks TM, Gutierrez CN, Stulak JM, Dearani JA, Miller JD. The use of virtual reality to reduce preoperative anxiety in first-time sternotomy patients: a randomized controlled pilot trial. Mayo Clin Proc 2020;95(6):1148-1157. [doi: 10.1016/j.mayocp.2020.02.032] [Medline: 32498771]
28. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y. Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: a systematic review. J Biomed Inform 2018;87:138-153 [FREE Full text] [doi: 10.1016/j.jbi.2018.09.002] [Medline: 30267895]
29. Li L, Novillo-Ortiz D, Azzopardi-Muscat N, Kostkova P. Digital data sources and their impact on people's health: a systematic review of systematic reviews. Front Public Health 2021;9:645260 [FREE Full text] [doi: 10.3389/fpubh.2021.645260] [Medline: 34026711]
30. Ho D, Quake SR, McCabe ER, Chng WJ, Chow EK, Ding X, et al. Enabling technologies for personalized and precision medicine. Trends Biotechnol 2020;38(5):497-518 [FREE Full text] [doi: 10.1016/j.tibtech.2019.12.021] [Medline: 31980301]
31. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision medicine, AI, and the future of personalized health care. Clin Transl Sci 2021;14(1):86-93 [FREE Full text] [doi: 10.1111/cts.12884] [Medline: 32961010]
32. Bertoncello C, Colucci M, Baldovin T, Buja A, Baldo V. How does it work? Factors involved in telemedicine home-interventions effectiveness: a review of reviews. PLoS One 2018;13(11):e0207332 [FREE Full text] [doi: 10.1371/journal.pone.0207332] [Medline: 30440004]
33. De Croon R, Van Houdt L, Htun NN, Štiglic G, Abeele VV, Verbert K. Health recommender systems: systematic review. J Med Internet Res 2021;23(6):e18035 [FREE Full text] [doi: 10.2196/18035] [Medline: 34185014]
34. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. PBSIJ 2017;2(2):555585. [doi: 10.19080/pbsij.2017.02.555585]

XSL•FO
RenderX

35. Hors-Fraile S, Malwade S, Luna-Perejon F, Amaya C, Civit A, Schneider F, et al. Opening the black box: explaining the process of basing a health recommender system on the I-change behavioral change model. IEEE Access 2019;7:176525-176540. [doi: 10.1109/access.2019.2957696]
36. Repetto C, Gaggioli A, Pallavicini F, Cipresso P, Raspelli S, Riva G. Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial. Pers Ubiquit Comput 2011;17(2):253-260. [doi: 10.1007/s00779-011-0467-0]
37. Kouroubali A, Kondylakis H, Karadimas E, Kavlentakis G, Simos A, Baños RM, et al. Digital health tools for preoperative stress reduction in integrated care. Eur J Biomed Inform 2019;16(2):7-13 [FREE Full text]
38. Kouroubali A, Kondylakis H, Koumakis L, Papagiannakis G, Zikas P, Katehakis DG. iSupport: building a resilience support tool for improving the health condition of the patient during the care path. Stud Health Technol Inform 2019;261:253-258. [Medline: 31156125]
39. Kondylakis H, Giglioli IAC, Katehakis DG, Aldemir H, Zikas P, Papagiannakis G, et al. A digital health intervention for stress and anxiety relief in perioperative care: protocol for a feasibility randomized controlled trial. JMIR Res Protoc 2022;11(11):e38536 [FREE Full text] [doi: 10.2196/38536] [Medline: 36445734]
40. Mitchell AM, Crane PA, Kim Y. Perceived stress in survivors of suicide: psychometric properties of the perceived stress scale. Res Nurs Health 2008;31(6):576-585. [doi: 10.1002/nur.20284] [Medline: 18449942]
41. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: visual analog scale for pain (VAS Pain), numeric rating scale for pain (NRS Pain), mcGill pain questionnaire (MPQ), short-form mcGill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
42. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
43. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
44. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
45. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the warwick-edinburgh mental well-being scale (WEMWBS): a rasch analysis using data from the Scottish health education population survey. Health Qual Life Outcomes 2009;7:15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
46. Schwarzer R. General perceived self-efficacy in 14 cultures. URL: https://userpage.fu-berlin.de/~health/world14.htm [accessed 2024-04-01]
47. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
48. Flannery JE, Giuliani NR, Flournoy JC, Pfeifer JH. Neurodevelopmental changes across adolescence in viewing and labeling dynamic peer emotions. Dev Cogn Neurosci 2017;25:113-127 [FREE Full text] [doi: 10.1016/j.dcn.2017.02.003] [Medline: 28262423]
49. Wong CL, Lui MMW, Choi KC. Effects of immersive virtual reality intervention on pain and anxiety among pediatric patients undergoing venipuncture: a study protocol for a randomized controlled trial. Trials 2019;20(1):369 [FREE Full text] [doi: 10.1186/s13063-019-3443-z] [Medline: 31221208]
50. WHO Guideline. Recommendations on digital interventions for health system strengthening. 2023. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-11-20]
51. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DH: digital health
GSE: General Self-Efficacy
HADS: Hospital Anxiety and Depression
INRCA: Instituto di Ricovero e Cura per Anziani
LMM: linear mixed model
PAM-13: Patient Activation Measure 13
PANAS: Positive and Negative Affect Scale
SAS: Hospital Universitario Reina Sofia
SJD: Sant Joan de Déu Hospital

XSL•FO
RenderX

SWEMWBS: Short Warwick Edinburgh Mental Well-Being Scale
VAS: Visual Analogue Scale
VR: virtual reality

Edited by A Mavragani; submitted 27.10.23; peer-reviewed by V Girishan Prabhu, A Dimitrios; comments to author 12.03.24; revised version received 02.05.24; accepted 14.10.24; published 07.01.25.

Please cite as:
Kondylakis H, Giglioli IAC, Katehakis D, Aldemir H, Zikas P, Papagiannakis G, Hors-Fraile S, González-Sanz PL, Apostolakis K, Stephanidis C, Núñez-Benjumea FJ, Baños-Rivera RM, Fernandez-Luque L, Kouroubali A
Stress Reduction in Perioperative Care: Feasibility Randomized Controlled Trial
J Med Internet Res 2025;27:e54049
URL: https://www.jmir.org/2025/1/e54049
doi:10.2196/54049
PMID:

©Haridimos Kondylakis, Irene Alice Chicchi Giglioli, Dimitrios Katehakis, Hatice Aldemir, Paul Zikas, George Papagiannakis, Santiago Hors-Fraile, Pedro L González-Sanz, Konstantinos Apostolakis, Constantine Stephanidis, Francisco J Núñez-Benjumea, Rosa M Baños-Rivera, Luis Fernandez-Luque, Angelina Kouroubali. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Long-Term Uptake of Home Spirometry in Regular Cystic Fibrosis Care: Retrospective Multicenter Observational Study

Pia Bertram[1*], MD; Martinus C Oppelaar[1*], MD; Michiel AGE Bannier[2], MD, PhD; Monique HE Reijers[3], MD, PhD; Hester van der Vaart[4], MD, PhD; Renske van der Meer[5], MD; Josje Altenburg[6], MD, PhD; Lennart Conemans[7,8], MD; Bart L Rottier[9,10], MD, PhD; Marianne Nuijsink[11], MD, PhD; Lara S van den Wijngaart[1], MD, PhD; Peter JFM Merkus[1], MD, PhD; Jolt Roukema[1], MD, PhD

[1]Department of Pediatric Pulmonology, Amalia Children's Hospital, Radboud University Medical Center, Nijmegen, Netherlands

[2]Department of Pediatric Pulmonology, MosaKids Children's Hospital, Maastricht University Medical Centre+, Maastricht, Netherlands

[3]Radboud University Medical Center, Department of Pulmonary Diseases, Nijmegen, Netherlands

[4]Department of Pulmonary Diseases, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[5]Department of Pulmonology, Haga Teaching Hospital, The Hague, Netherlands

[6]Department of Respiratory Medicine, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands

[7]Department of Respiratory Medicine, Maastricht University Medical Centre+, Maastricht, Netherlands

[8]Division of Respiratory & Age-related Health, Department of Respiratory Medicine, NUTRIM Institute of Nutrition and Translational Research in Metabolism, Maastricht, Netherlands

[9]Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University Medical Center Groningen, Groningen, Netherlands

[10]Groningen Research Institute for Asthma and COPD, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[11]Juliana Children's Hospital, Haga Teaching Hospital, The Hague, Netherlands

[*]these authors contributed equally

Corresponding Author:
Martinus C Oppelaar, MD
Department of Pediatric Pulmonology
Amalia Children's Hospital
Radboud University Medical Center
Geert Grooteplein Zuid 32
Nijmegen, 6526 GA
Netherlands
Phone: 31 243614430
Fax: 31 243616428
Email: marc.oppelaar@radboudumc.nl

Abstract

Background: Home spirometers have been widely implemented in the treatment of people with cystic fibrosis (CF). Frequent spirometry measurements at home could lead to earlier detection of exacerbations. However, previous research indicates that the long-term use of home spirometry is not well maintained by people with CF.

Objective: We aimed to gain insight into the long-term uptake of home spirometry in regular multicenter CF care.

Methods: Home spirometers combined with a remote monitoring platform were introduced in the treatment of people with CF in 5 Dutch CF centers starting in April 2020. Usage data from April 2020 to December 2022 were analyzed retrospectively. Survival analyses were conducted to assess use consistency over time, and *t* tests were used to evaluate the impact of increased pulmonary symptoms on home spirometry frequency. The effect of the initiation of a new treatment, Elexacaftor/Tezacaftor/Ivacaftor, on use frequency over time was assessed in a subgroup of participants with repeated measures ANOVA.

Results: During the observation period, a total of 604 people with CF were enrolled in the remote monitoring platform and 9930 home spirometry measurements were performed. After the initiation of home spirometry use, the number of users declined rapidly. One year after the initiation, 232 (54.2%) people with CF stopped using home spirometry. During the observation period, 67 (11.1%) users performed more than 20 measurements. Furthermore, the number of consistent home spirometry users decreased

XSL•FO
RenderX

over time. After 600 days, only 1% of users had measured their lung function consistently every 31 days. Use frequency slightly increased during periods with increased pulmonary symptoms (ΔMean=0.45, $t_{497.278}$=–4,197; P<.001) and showed an initial rise followed by a decrease after starting treatment with Elexacaftor/Tezacaftor/Ivacaftor (ΔMean=0.45, $t_{497.278}$=–4,197; P<.001).

Conclusions: Consistent uptake of home spirometry in people with CF is low but increases around periods of changing symptoms. A clear strategy for the organization of remote care seemed to improve the long-term uptake of home spirometry. Nevertheless, home spirometry and its intensity are not a goal on their own but should be used as a tool to reach individual goals within local contexts.

(J Med Internet Res 2025;27:e60689) doi:10.2196/60689

KEYWORDS

telemonitoring; eHealth; spirometry; adherence; pulmonary medicine; home spirometers; cystic fibrosis; autosomal disease; treatment; remote monitoring; survival analyses; frequency; digital health; telehealth

Introduction

Cystic fibrosis (CF) is a recessive autosomal disease caused by dysfunction of the CF transmembrane conductance regulator (CFTR) channel. Dysfunction of the CFTR channel leads to increased pulmonary mucus viscosity that causes chronic inflammation and predisposes to recurrent infections [1]. Early detection and treatment of pulmonary decline is essential to prevent mortality and morbidity in people with CF. The forced expiratory volume in 1 second (FEV1) measured by spirometry is the most reliable clinical outcome for this end [2,3].

Portable home spirometers that measure FEV1 have been available for multiple years. However, they have recently become more accessible thanks to advances in digital patient environments, the widespread availability of smartphones to operate them with, and a pivot toward home monitoring driven by the COVID-19 pandemic [4,5]. In theory, home spirometers allow people with CF to identify lung function decline sooner which could help them seek timely medical intervention and thus reduce mortality, morbidity, and health care costs [6].

Early detection of pulmonary decline with home spirometry would require regular self-monitoring of people with CF over extended periods, also when their condition is stable. In recent work, we found there is little incentive for people with CF to frequently self-monitor during periods of well-being, or when there is no reduction in therapy burden elsewhere [7]. This is in line with findings from previous prospective studies that reported declines in monitoring adherence to study protocols over time, especially in adults [6,8-11].

Understanding how people with CF incorporate home spirometry into their treatment routine is crucial for evaluating its added value. There is a specific need for more insights into the long-term uptake of home spirometry in regular CF care. This study aimed to examine the uptake of home spirometry in regular CF care in 5 Dutch CF centers over 2.5 years. From previous work, we hypothesized that the uptake of home spirometry is not well-maintained over time, that home spirometry uptake increases during periods with increased pulmonary symptoms, and that home spirometry uptake declines after the initiation of treatment with a new drug that specifically targets the CFTR channel (Elexacaftor/Tezacaftor/Ivacaftor) [7].

Methods

Study Design

This was a retrospective, multicenter, observational study conducted in 5 Dutch CF centers (Radboud University Medical Center Nijmegen, University Medical Center Groningen, Maastricht University Medical Center, Amsterdam University Medical Center, Haga Hospital, The Hague).

Remote Monitoring Platform

Home spirometry testing combined with a remote monitoring platform (RMP) was implemented on April 1, 2020, to prevent unnecessary outpatient visits during the COVID-19 pandemic in 5 Dutch CF centers. Details of the RMP have been published elsewhere [7]. In short, the RMP was used in addition to regular care in order to reduce physical hospital visits when possible. Considering the turbulent period of the COVID-19 pandemic, no formal implementation period or protocols were used. Instead, the implementation of the RMP and its use varied between departments based on their local settings, capacities, and needs. Therefore, RMP implementation across the 5 centers and enrollment of individual users occurred gradually during the COVID-19 period [7]. As a general rule, people with CF were asked to measure their lung function monthly, but health care professionals and users had the freedom to deviate from this based on individual contexts.

Home spirometry was performed with the Spirobank Smart (Medical Internet Research). Home spirometers were donated by the Dutch Cystic Fibrosis Foundation. In addition, 4 of the 5 CF centers used a digital symptom questionnaire based on the Modified Fuchs criteria to track symptoms indicative of a pulmonary exacerbation. How often this questionnaire was used varied between centers and people with CF. If 2 or more symptoms were reported, automated feedback provided the advice to measure the lung function and contact the CF team as needed [12].

Study Population

This study included all people with CF who measured their lung function at home with the RMP. Home spirometry and symptom survey data of all users of the RMP were extracted anonymously. As this was a retrospective observational trial with anonymous data, no formal eligibility criteria were used. In practice, eligibility for enrollment on the RMP is decided by the treating

XSL•FO
RenderX

health care professionals and patients themselves based on whether the user is able and willing to measure their lung function at home. In general, children had to be older than 6 years old to be considered for enrolment on the RMP. Individual start dates of Elexacaftor/Tezacaftor/Ivacaftor were available for a subgroup of people with CF who participated in a previous study and who gave informed consent for additional data collection [7].

Primary and Secondary Outcome Measures

Primary outcomes included (1) the median number of days between each consecutive home spirometry measurement day during the observation period (time-to-next lung function [TTN]) and (2) the proportion of users who continued self-monitoring with consistent frequencies (ie, monthly, 2-monthly, or quarterly) over time. Secondary outcomes included (1) the difference in the average intra-person frequency of home spirometry measurements between periods of increased pulmonary symptoms (ie, a symptom survey score ≤5) and few pulmonary symptoms (ie, a symptom survey score ≥6) and (2) the difference in the average intra-person frequency of home spirometry following initiation of Elexacaftor/Tezacaftor/Ivacaftor.

Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics (version 27). User characteristics were described for all users and user groups based on home spirometry uptake categories (ie, 0, 1-20, and >20 unique measurement days). User characteristics of the group without measurements and the group with >20 unique measurement days were compared with each other using the Pearson chi-square test for proportions. To compare uptake categories with CF departments the Monte Carlo method was used to estimate the Fisher exact test. As data of users were gathered anonymously, only age, sex, and treating CF centers could be described. Treating CF centers were anonymized.

Consistent uptake of home spirometry was assessed using the Kaplan-Meier estimator. A total of 2 different definitions of consistent uptake were used in 2 separate analyses. First, we examined the number of users who self-monitored every 31, 62, or 93 days without breaking this pattern (ie, TTN remained below 31, 62, or 93 days). A break in this pattern was considered a loss of survival. Second, we allowed 1 break in the pattern after every 2 measurements following the pattern of 31 or 62 days. The break could be no longer than twice the number of the pattern (ie, 62 or 124 days, respectively). Users who had never measured their lung function or only once were excluded from the analyses. Separate subgroup analyses were performed for age categories at the first measurement, CF centers, and the adult and pediatric CF departments of the CF centers.

An independent *t* test was used to compare the mean number of home spirometry measurement days within 28 days of all symptom surveys with a score ≤5 (ie, a period of increased symptoms) with those taken within 28 days of all symptom surveys with a score ≥6 (ie, a period of few symptoms). In a sensitivity analysis, the intraindividual change in home spirometry use related to the number of symptoms was examined. A dependent *t* test was used to compare the number of home spirometry measurement days during the first period of increased symptoms of an individual after the initial 100 days of use with the closest period of few symptoms.

Changes in use frequency after the introduction of Elexacaftor/Tezacaftor/Ivacaftor were examined in a subgroup of people with CF. Data on Elexacaftor/Tezacaftor/Ivacaftor use was only available for a subgroup of 81 users who participated in a previous study [7]. Change in frequency of use was assessed with repeated measures ANOVA between multiple intervals around the initiation of Elexacaftor/Tezacaftor/Ivacaftor: the 30 days before initiation of Elexacaftor/Tezacaftor/Ivacaftor; and 0-30 days, 30-60 days, 60-90 days, and 90-120 days after initiation of Elexacaftor/Tezacaftor/Ivacaftor. Users were only included if they started Elexacaftor/Tezacaftor/Ivacaftor at least 120 days before the end of the study period.

Ethical Considerations

Local ethical committees waived formal approval considering the negligible burden of participation and absence of imposed risks (file number local ethical committee Arnhem-Nijmegen region: 2021-13214). A subgroup of people with CF provided electronic informed consent for the use of additional pseudonymized data to answer the research questions of this study.

Results

Patient Characteristics

A total of 604 people with CF were included in the analysis. Demographics are presented in Table 1. The median time people with CF were enrolled on the RMP during the observation period was 816 (IQR 716-856, range 37-998) days. Throughout the 998-day observation period, a total of 9930 home spirometry measurements were performed of which 5104 were unique day measurements. Overall, 428/604 (70.1%) people with CF performed at least 1 home spirometry measurement, 365/604 (60.4%) had 2 or more unique home spirometry measurements days, and this number decreased further with more unique measurement days (Table 1). One year after the initiation of home spirometry, 232 (54.2%) people with CF stopped using home spirometry. For users with 2 or more unique home spirometry measurement days (N=365), the median participant TTN was 25 (IQR 12-76, range 1-541) days for the total period. The median number of days between the first and last unique home spirometry day was 334.5 (IQR 59.25-632.75, range 0-966).

XSL•FO
RenderX

Table 1. User characteristics of total population and subgroups based on the number of unique measurement days (none, 1-20, and >20).

	Total	No measurements	1-20 measurements	>20 measurements
	604 (100)	176 (29.1)	361 (59.8)	67 (11.1)
Sex, n (%)				
Male	307 (50.8)	97 (55.1)	180 (49.9)	30 (44.8)
Missing	34 (5.6)	9 (5.1)	21 (5.8)	4 (6)
Age category[a,b], n (%)				
<12 years	85 (14.1)	27 (15.3)	39 (10.8)	19 (28.4)
12-18 years	75 (12.4)	24 (13.6)	37 (10.2)	14 (20.9)
>18 years	432 (71.5)	123 (69.9)	277 (76.7)	18 (47.8)
Missing	12 (2.0)	2 (1.1)	8 (2.2)	2 (3.0)
CF[c] department[d,e,f], n (%)				
1	31 (5.1)	4 (2.3)	20 (5.5)	7 (10.4)
2	49 (8.1)	7 (4.0)	37 (10.2)	5 (7.5)
3	16 (2.6)	0 (0)	6 (1.7)	10 (14.9)
4	46 (7.6)	11 (6.3)	23 (6.4)	12 (17.9)
5	34 (5.6)	24 (13.6)	10 (2.8)	0 (0)
6	136 (22.5)	59 (33.5)	75 (20.8)	2 (3.0)
7	51 (8.4)	7 (4.0)	29 (8.0)	15 (22.4)
8	76 (12.6)	11 (6.3)	60 (16.6)	5 (7.5)
9	28 (4.6)	16 (9.1)	11 (3.0)	1 (1.5)
10	125 (20.7)	35 (19.9)	82 (22.7)	8 (11.9)
Missing	12 (2.0)	2 (1.1)	8 (2.2)	2 (3.0)
TTN[g], median (IQR)	25 (12-76)	—[h]	34 (14-91)	13 (7-17)
Time in days between first and last measurement, median (IQR)	334.5 (59.25-632.75)	—	237 (21-524.5)	779 (633-832)

[a]Significant difference between the subgroup without measurements and the subgroup with >20 measurements based on the Pearson chi-square test of proportions.

[b]Significant difference between the subgroup with 1-20 measurements and the subgroup with >20 measurements based on the Pearson chi-square test of proportions.

[c]CF: cystic fibrosis.

[d]Significant difference between the subgroup without measurement and the subgroup with 1-20 measurements based on the Monte Carlo estimated Fisher exact test.

[e]Significant difference between the subgroup without measurements and the subgroup with >20 based on the Monte Carlo estimated Fisher exact test.

[f]Significant difference between the subgroup with 1-20 measurements and the subgroup with >20 measurements based on the Monte Carlo estimated Fisher exact test.

[g]TTN: time-to-next lung function measurement.

[h]Not applicable.

Long-Term Use

At 600 days, only 1% of users had measured their lung function at home consistently every 31 days. The decline was largest during the first 150 days across all intervals. For the more lenient criteria, the decline was less steep but followed a similar curve. Figure 1 shows Kaplan-Meier curves for the different uptake patterns. On average, consistent use frequency patterns were broken after 41 days (95% CI 31-51 days) for the 31-day interval, after 105 days (95% CI 86-123 days) for the 62-day interval, and after 172 days (95% CI 95-118 days) for the 92-day interval. After 150 days, 93% of users had broken the 31-day pattern, 79% had broken the 62-day pattern, and 68% had broken the 93-day pattern (Figure 1).

XSL•FO
RenderX

Figure 1. Kaplan-Meier survival curve of the consistency of home spirometry use (n=365). Upper panel: Consistent users with every FEV1 measurement performed with a maximum time-to-next lung function of 31, 62, or 93 days. Lower panel: consistent users with a more lenient approach, where one FEV1 measurement can be performed with a longer time-to-next lung function if two subsequent measurements are taken with a maximum time-to-next lung function of 31, or 62 days. People with cystic fibrosis who are consistent users at the end of the observation period are censored with dots. Yellow: 93-day interval; green: 62-day interval; blue: 31-day interval.

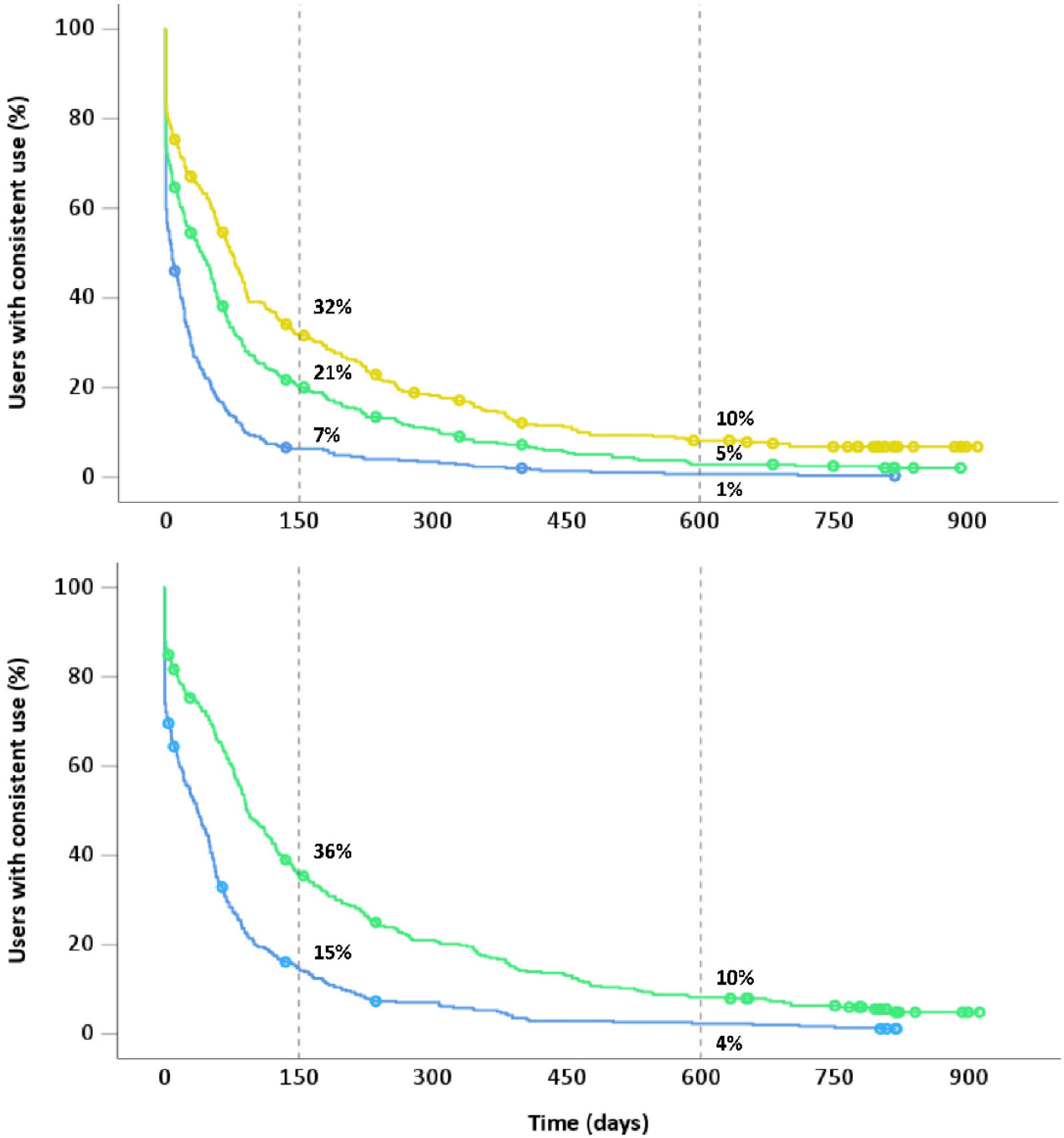

Multimedia Appendix 1 shows Kaplan-Meier curves for the strict use frequency patterns with subgroups based on age categories (6-12 years, 12 -18 years, ≥18 years; Figure S1 in Multimedia Appendix 1), sex (Figure S2 in Multimedia Appendix 1), treating CF center (Figure S3 in Multimedia Appendix 1), and age categories within treating CF center (<18 years within treating CF center vs ≥18 years within treating CF center; Figure S4 in Multimedia Appendix 1). There was no difference between long-term use consistency and sex. Children and teenagers generally seemed to use their home spirometer consistently longer than adults. Average consistent use frequency patterns were maintained longer in one CF center than in the others with a mean survival of the strict 31-day pattern of 125 days (95% CI 70-179 days). Both the pulmonary and pediatric pulmonary departments of this CF center had better maintained consistent use rates than all other departments (Figure S3 in Multimedia Appendix 1).

XSL•FO
RenderX

Impact of Increased Pulmonary Symptoms on Home Spirometry Use Frequency

A total of 381 periods with increased symptoms and 1340 periods with few symptoms were compared. There was a small, but significant difference between the mean number of home spirometry tests performed during periods of increased symptoms (mean=1.33) and those with few symptoms (mean=0.88; ΔMean=0.45, $t_{497.278}$=–4,197; *P*<.001). In the sensitivity analysis, periods of increased and few symptoms could be compared for 76 users. Dependent *t* test showed no significant difference between the number of home spirometry tests performed between these periods (ΔMean=0.37, t_{75}=–1.843; *P*=.07). Of the users in the sensitivity analysis, 14 (18.4%) measured only during periods of increased symptoms, 5 (6.6%) only during periods with few symptoms, and 44 (57.9%) measured their lung function in both periods. A total of 13 (17.1%) of all users did not measure their lung function in both periods. In an exploratory analysis, the FEV1 during the first period with increased symptoms after 100 days of use was compared with the nearest period with few symptoms for 44 users with a Wilcoxon signed rank test. The FEV1 was significantly lower during the period with increased symptoms (median=1.75, IQR 1.41-2.59) compared with the period with low symptoms (median=1.97, IQR 1.52-2.89; ΔMedian=0.22; *Z*=–3.46; *P*<.001).

Influence of Elexacaftor/Tezacaftor/Ivacaftor Treatment Initiation on Use Frequency

A total of 63 out of 81 (78%) people with CF for whom additional data were available started Elexacaftor/Tezacaftor/Ivacaftor treatment during the observation period. More users measured their lung function at home within 30 days after initiation of Elexacaftor/Tezacaftor/Ivacaftor (N=42/63, 67%) compared with the other time intervals (Figure 2). The mean number of home spirometry measurements for the consecutive time intervals were 0.76, 2.11, 0.97, 0.81, and 0.86, respectively. Repeated measures ANOVA with Greenhouse-Geyser correction showed a significant effect of time in relation to Elexacaftor/Tezacaftor/Ivacaftor initiation on the number of home spirometry measurements ($F_{2.36,146.26}$=13.73, *P*<.001). Tests of within-subjects contrasts showed a significant difference between the month before initiation and 0-30 days after initiation ($F_{1.62}$=19, *P*<.001) and between 0-30 days and 30-60 days after initiation ($F_{1.62}$=15.37, *P*<.001) but not between 30-60 days and 60-90 days, or 60-90 days and 90-120 days after initiation of Elexacaftor/Tezacaftor/Ivacaftor treatment.

Figure 2. Mean number and 95% CI values of home spirometry measurements per user in the month before and months after initiation of Elexacaftor/Tezacaftor/Ivacaftor treatment. N indicates the number of users who measured their lung function at least once during the intervals.

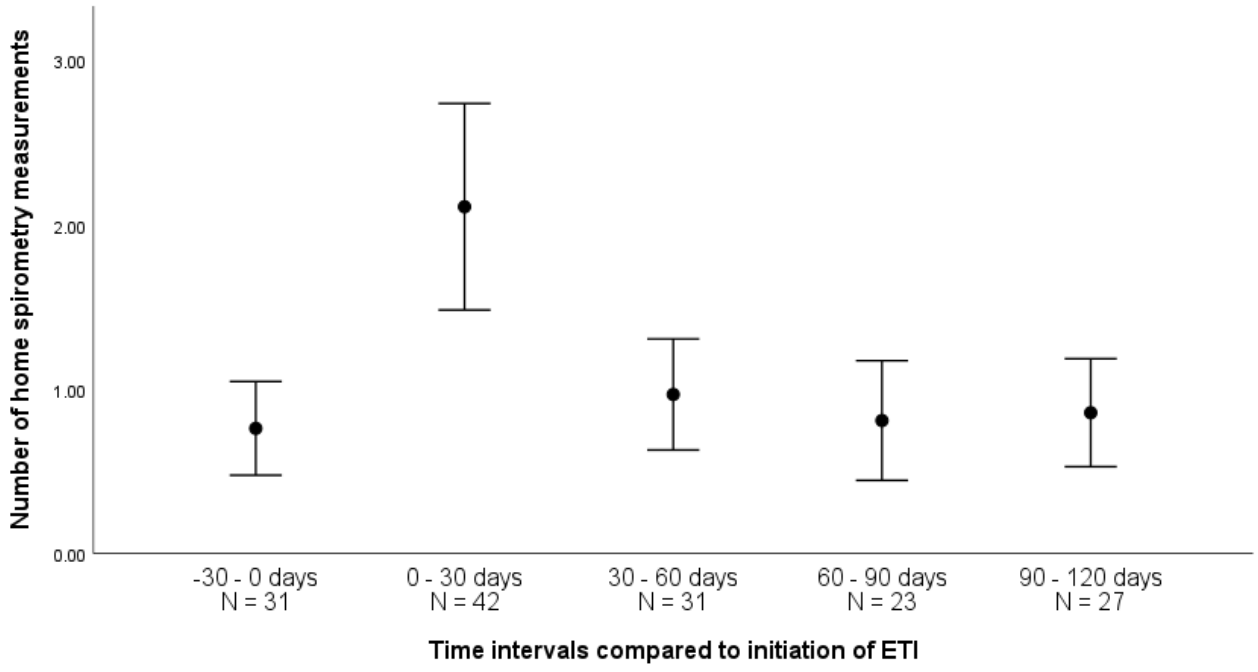

Discussion

Principal Findings

This study aimed to examine the use of home spirometry in regular CF care for 2.5 years in 5 Dutch CF centers. Our results showed a fast decrease in both use frequency and the number of users of home spirometry over time. Only a small group of people with CF continued to use home spirometry consistently during the study period. We observed a small but significant increase in the absolute number of home spirometry measurements during periods of increased pulmonary symptoms on a group level as well as shortly after initiation of Elexacaftor/Tezacaftor/Ivacaftor treatment. Moreover, during increased symptoms, up to a fifth of users measured their lung

XSL•FO
RenderX

function at home whereas they did not during periods of few pulmonary symptoms.

Comparison With Literature

Low adherence to home spirometry protocols has been identified as a recurring problem in many prospective trials [8-11,13]. Our results show that the frequency of home spirometry monitoring might be even lower in routine care. This discrepancy could partly be attributed to differences in study designs. First, selection bias is inherent to trials as they select highly skilled and motivated patients or patients with few comorbidities for whom preventing a decline in pulmonary function is the highest priority. Second, this was a retrospective study in the time frame of the COVID-19 pandemic and during the introduction of Elexacaftor/Tezacaftor/Ivacaftor treatment in the Netherlands. These enormous changes in CF care and the lack of a home spirometry protocol might have limited the uptake of home spirometry in our study.

In previous work, we identified multiple incentives (eg, changes in the physical condition and positive psychosocial effects) and disincentives (eg, technical difficulties and lack of perceived need) for regular home monitoring of lung function [7]. Regular use of home spirometry requires patients to allocate time and effort to perform the measurements and this increases the already high treatment burden of people with CF [14-16]. When there are no clear benefits from home spirometry or reductions in overall treatment burden, people with CF might have little motivation to keep measuring their lung function at home regularly [7]. This is emphasized by our findings that an increase in pulmonary symptoms and initiation of Elexacaftor/Tezacaftor/Ivacaftor temporarily increased both the number of users and the number of home spirometry tests. Increases in symptoms and new modulator therapies provide clear incentives and a perceived necessity for people with CF to measure their lung function, but during stable periods these incentives may be less evident.

A previous study showed prolonged uptake in home spirometry in children compared with adults, but the authors also reported that increased supervision of health care professionals in pediatric care might have attributed to this difference [11]. In our study, younger people with CF had just slightly better uptakes of home spirometry than adults. However, we found large differences between CF centers, with much more consistent users in the adult and pediatric departments of one CF center. Interestingly, in this CF center, CF nurses cared for both the adult and pediatric populations. These CF nurses had created a clear strategy for the organization of remote care with policies for the initiation, follow-up, and administration of remote monitoring [7]. This strategy was likely used for both the adult and pediatric populations, whereas similar strategies were not necessarily present in other centers due to the turbulent period of the COVID-19 pandemic [7]. We already know that CF nurses play an important role in the successful implementation of home monitoring interventions, and these findings suggest that differences in uptake might not be due to the age of patients but rather due to the differences in the organization of care and policies regarding home spirometry across CF centers and departments [7]. This strengthens the conclusion that clear agreements within departments and with patients as well as clear policies for the initiation and follow-up of home spirometry may enhance its sustainability.

Strength and Limitations

To our knowledge, this is the first long-term study into the uptake of home spirometry in a real-world clinical setting. Moreover, our study benefited from the large number of participants, the multicenter setting, and the long follow-up time that increases generalizability to other settings. As stated earlier, the time of initiation of home spirometry was turbulent due to the COVID-19 pandemic and the introduction of Elexacaftor/Tezacaftor/Ivacaftor treatment. This may have affected the implementation of home spirometry and therefore use frequency. Moreover, CF departments were provided with spirometers and access to the RMP but were allowed to implement home spirometry according to their preferences. The uptake of home spirometry differed between CF departments and this may limit our overall findings. But these differences between centers also allowed us to hypothesize about the underlying causes of varying uptake such as the availability of motivated CF nurses and clear policies. Finally, because this study used anonymized data, we were unable to study the frequency of home spirometry within individual contexts such as treatment, health care consumption, socioeconomic background, and other important determinants for digital health use. In addition, due to the use of anonymized data, data on Elexacaftor/Tezacaftor/Ivacaftor treatment use was only available for a small number of users.

Future Directions

The role of home spirometry in the field of CF and other pulmonary diseases is becoming increasingly evident. However, to ensure that home spirometry does not become an inefficient and underused tool, it is imperative that we regard it as a means to an end instead of a goal on its own [17]. Goals should be tailored to individual needs and contexts, and the intensity of home spirometry monitoring should be adjusted accordingly. Centers should create clear policies for the organization of remote care that address how to set, initiate, follow up, and evaluate these goals on an individual basis to ensure successful implementation.

We previously identified four “Sensible Strategies for Remote Monitoring” as a set of predefined goals that can be adopted in CF care [7]. In our study, CF centers mostly seemed to adopt symptom-driven strategies where home spirometry was used to quantify symptoms when needed rather than at regular intervals. However, the positive attitude from our questionnaire results implies that it is not just the frequency of use that creates the benefits [7]. Nevertheless, some people with CF might benefit from regular home monitoring within patient-driven strategies that address individual needs, but a widespread application seems limited.

There is an increasing interest in prediction-driven strategies that use regular home spirometry to predict a decline in pulmonary function or other clinical outcomes [9,10,18]. However, these strategies only work when people with CF monitor themselves frequently over extended periods of time,

XSL•FO
RenderX

also during periods of stability. Our results show these strategies are likely not feasible for the long term or the majority of users. To mitigate this, studies will have to devise solutions to solve the problem of low adherence many of these studies face. They will have to devise solutions and address the lack of perceived necessity for people with CF to measure their lung function during periods of stability. This could include limiting their studies to people with CF who deteriorate often or reducing treatment burdens elsewhere to ensure that home spirometry is not just an additional burden (eg, care-driven strategies that replace regular care with remote monitoring) [15,19].

Hence, there should be a stronger emphasis on whether home spirometry can support individual patients and what intensity of home monitoring best fits individual goals. The frequency of home spirometry on its own is meaningless when not coupled with clear strategies. This approach will not only prevent the introduction of new "monitoring-adherence discussions" in our clinics, but it will also ensure that remote monitoring remains value-driven.

Conclusion

Consistent uptake of home spirometry in regular long-term multicenter CF care is low but appears to increase around periods of changing symptoms. A clear strategy for the organization of remote care seems to stimulate long-term uptake of home spirometry. Nevertheless, frequent and regular home spirometry monitoring might not be feasible for all patients and all the time. Instead, home spirometry and its intensity are not a goal but should be used as a tool to reach individual patient goals within local contexts.

Acknowledgments

We would like to thank the Dutch CF Foundation for donating the portable spirometers. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Anonymized data from this study will be made available upon reasonable request. No generative AI was used at any stage of this study.

Data Availability

Anonymized data from this study will be made available upon reasonable request.

Authors' Contributions

All authors were involved in the conceptualization of this study and data collection. PB, MO, JR, and PM were responsible for the study design. PB and MO managed the project administration, quantitative analyses, visualization of results, and writing of the manuscript. Analyses were supervised by PM and JR. All authors were involved in reviewing the manuscript. PB and MO contributed equally to this manuscript and shared the first authorship.

Conflicts of Interest

MO, LvdW, JR, and PM are involved in the development of the software used in this study. The software is the property of the Radboud University Medical Center and hence does not generate any profits for the authors.

Multimedia Appendix 1
Kaplan-Meier Plots for the consistent use of home spirometry using strict criteria.
[DOCX File , 389 KB - jmir_v27i1e60689_app1.docx]

References

1. Ratjen F, Bell SC, Rowe SM, Goss CH, Quittner AL, Bush A. Cystic fibrosis. Nat Rev Dis Primers 2015 May 14;1(1):15010 [FREE Full text] [doi: 10.1038/nrdp.2015.10] [Medline: 27189798]
2. Bell S, Mall M, Gutierrez H, Macek M, Madge S, Davies J, et al. The future of cystic fibrosis care: a global perspective. The Lancet Respiratory Medicine 2020 Jan;8(1):65-124. [doi: 10.1016/S2213-2600(19)30337-6] [Medline: 31669225]
3. Castellani C, Duff AJA, Bell SC, Heijerman HGM, Munck A, Ratjen F, et al. ECFS best practice guidelines: the 2018 revision. J Cyst Fibros 2018;17(2):153-178 [FREE Full text] [doi: 10.1016/j.jcf.2018.02.006] [Medline: 29506920]
4. Dowd C, Van Citters AD, Dieni O, Willis A, Powell L, Sabadosa KA. Design and methods for understanding the state of cystic fibrosis care amid the COVID-19 pandemic. J Cyst Fibros 2021;20 Suppl 3:3-8 [FREE Full text] [doi: 10.1016/j.jcf.2021.08.028] [Medline: 34930539]
5. Richardson CH, Orr NJ, Ollosson SL, Irving SJ, Balfour-Lynn IM, Carr SB. Initiating home spirometry for children during the COVID-19 pandemic - a practical guide. Paediatr Respir Rev 2022 Jun;42:43-48 [FREE Full text] [doi: 10.1016/j.prrv.2021.02.001] [Medline: 33773928]
6. Calthorpe RJ, Smith S, Gathercole K, Smyth AR. Using digital technology for home monitoring, adherence and self-management in cystic fibrosis: a state-of-the-art review. Thorax 2020 Jan;75(1):72-77 [FREE Full text] [doi: 10.1136/thoraxjnl-2019-213233] [Medline: 31594802]

XSL•FO
RenderX

7. Oppelaar MC, Emond Y, Bannier MAGE, Reijers MHE, van der Vaart H, van der Meer R, et al. Potential, pitfalls, and future directions for remote monitoring of chronic respiratory diseases: multicenter mixed methods study in routine cystic fibrosis care. J Med Internet Res 2024;26:e54942 [FREE Full text] [doi: 10.2196/54942] [Medline: 39106098]
8. Shakkottai A, Nasr SZ. The use of home spirometry in pediatric cystic fibrosis patients: results of a feasibility study. Glob Pediatr Health 2017;4:2333794X17690315 [FREE Full text] [doi: 10.1177/2333794X17690315] [Medline: 28229102]
9. van Horck M, Winkens B, Wesseling G, van Vliet D, van de Kant K, Vaassen S, et al. Early detection of pulmonary exacerbations in children with cystic fibrosis by electronic home monitoring of symptoms and lung function. Sci Rep 2017 Sep 27;7(1):12350 [FREE Full text] [doi: 10.1038/s41598-017-10945-3] [Medline: 28955051]
10. Lechtzin N, Mayer-Hamblett N, West NE, Allgood S, Wilhelm E, Khan U, eICE Study Team. Home monitoring of patients with cystic fibrosis to identify and treat acute pulmonary exacerbations. eICE Study results. Am J Respir Crit Care Med 2017 Nov 01;196(9):1144-1151 [FREE Full text] [doi: 10.1164/rccm.201610-2172OC] [Medline: 28608719]
11. Beaufils F, Enaud R, Gallode F, Boucher G, Macey J, Berger P, et al. Adherence, reliability, and variability of home spirometry telemonitoring in cystic fibrosis. Front Pediatr 2023;11:1111088 [FREE Full text] [doi: 10.3389/fped.2023.1111088] [Medline: 36911035]
12. Bilton D, Canny G, Conway S, Dumcius S, Hjelte L, Proesmans M, et al. Pulmonary exacerbation: towards a definition for use in clinical trials. report from the EuroCareCF working group on outcome parameters in clinical trials. Journal of Cystic Fibrosis 2011;10:S79-S81. [doi: 10.1016/s1569-1993(11)60012-x]
13. Davis J, Ryan M, Marchetti P, Dahlberg SE, Greenberg J, Bacon C, et al. Real-world feasibility of short-term, unsupervised home spirometry in CF. Pediatr Pulmonol 2022 Dec;57(12):3129-3135. [doi: 10.1002/ppul.26147] [Medline: 36124390]
14. Calthorpe RJ, Smith SJ, Rowbotham NJ, Leighton PA, Davies G, Daniels T, et al. What effective ways of motivation, support and technologies help people with cystic fibrosis improve and sustain adherence to treatment? BMJ Open Respir Res 2020;7(1):e000601 [FREE Full text] [doi: 10.1136/bmjresp-2020-000601] [Medline: 32816834]
15. Edmondson C, Westrupp N, Seddon P, Olden C, Wallis C, Dawson C, et al. The feasibility of home monitoring of young people with cystic fibrosis: Results from CLIMB-CF. J Cyst Fibros 2022;21(1):70-77 [FREE Full text] [doi: 10.1016/j.jcf.2021.09.018] [Medline: 34635459]
16. Kruizinga MD, van der Heide N, Moll A, Zhuparris A, Yavuz Y, Kam MLD, et al. Towards remote monitoring in pediatric care and clinical trials-Tolerability, repeatability and reference values of candidate digital endpoints derived from physical activity, heart rate and sleep in healthy children. PLoS One 2021;16(1):e0244877 [FREE Full text] [doi: 10.1371/journal.pone.0244877] [Medline: 33411722]
17. Oppelaar MC, van den Wijngaart LS, Merkus PJFM, Croonen EA, Hugen CAC, Brouwer ML, et al. It is not just the FEV1 that matters, but the personal goals we reach along the way: qualitative, multicenter, prospective, observational study. J Med Internet Res 2021;23(10):e29218 [FREE Full text] [doi: 10.2196/29218] [Medline: 34668868]
18. Yanaz M, Yilmaz Yegit CY, Gulieva A, Kalyoncu M, Selcuk M, Uzunoglu B, et al. Electronic home monitoring of children with cystic fibrosis to detect and treat acute pulmonary exacerbations and its effect on 1-year FEV. J Cyst Fibros 2024;23(2):329-333. [doi: 10.1016/j.jcf.2023.09.007] [Medline: 37748990]
19. Solomon GM, Bailey J, Lawlor J, Scalia P, Sawicki GS, Dowd C, et al. Patient and family experience of telehealth care delivery as part of the CF chronic care model early in the COVID-19 pandemic. J Cyst Fibros 2021;20(3):41-46 [FREE Full text] [doi: 10.1016/j.jcf.2021.09.005] [Medline: 34930542]

Abbreviations

CF: cystic fibrosis
CFTR: CF transmembrane conductance regulator
FEV1: forced expiratory volume in one second
RMP: remote monitoring platform
TTN: time-to-next lung function measurement

Edited by T de Azevedo Cardoso; submitted 19.05.24; peer-reviewed by D Drummond, C Addy, L Naehrlich; comments to author 14.08.24; revised version received 07.09.24; accepted 21.10.24; published 09.01.25.

Please cite as:
Bertram P, Oppelaar MC, Bannier MAGE, Reijers MHE, van der Vaart H, van der Meer R, Altenburg J, Conemans L, Rottier BL, Nuijsink M, van den Wijngaart LS, Merkus PJFM, Roukema J
The Long-Term Uptake of Home Spirometry in Regular Cystic Fibrosis Care: Retrospective Multicenter Observational Study
J Med Internet Res 2025;27:e60689
URL: https://www.jmir.org/2025/1/e60689
doi:10.2196/60689
PMID:

XSL•FO
RenderX

©Pia Bertram, Martinus C Oppelaar, Michiel AGE Bannier, Monique HE Reijers, Hester van der Vaart, Renske van der Meer, Josje Altenburg, Lennart Conemans, Bart L Rottier, Marianne Nuijsink, Lara S van den Wijngaart, Peter JFM Merkus, Jolt Roukema. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Early Intervention for Children With Developmental Disabilities and Their Families via Telehealth: Systematic Review

Yoomi Shin[1], BSN, RN; Eun Ju Park[2], MSN, RN; Anna Lee[2,3,4], PhD, RN

[1]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

[2]College of Nursing, Yonsei University, Seoul, Republic of Korea

[3]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[4]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Anna Lee, PhD, RN
Mo-Im Kim Nursing Research Institute
College of Nursing
Yonsei University
50-1, Yonsei-Ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3284
Email: annalee@yuhs.ac

Abstract

Background: Early intervention during the first 3 years of life is crucial for children with developmental disabilities to optimize developmental outcomes. However, access to such services is often limited by geographical distance and resource constraints. Telehealth can be part of a solution for overcoming these barriers, enabling the delivery of early intervention services. However, a comprehensive understanding of the efficacy and implementation of telehealth in early interventions remains elusive, particularly for children aged 0-3 years.

Objective: This systematic review aims to synthesize existing research on the effectiveness and implementation of telehealth interventions in infants and toddlers (aged 0–3 years) who are at risk of or diagnosed with developmental disabilities. The primary objective of the study is to evaluate the ways that telehealth compares to conventional in-person interventions in improving developmental outcomes for children and supporting family well-being.

Methods: A systematic search was conducted of 4 electronic databases (PubMed, Embase, CINAHL, and Web of Science), focusing on studies published between 2010 and 2024. The inclusion criteria were studies involving telehealth interventions for children aged 0-3 years who were at high risk or had developmental disabilities, which involved active interactions between the providers and the families. Study quality was assessed using the mixed methods appraisal tool, and a narrative synthesis was used to analyze the data.

Results: Eighteen studies met the inclusion criteria: 12 single-case designs, 4 randomized controlled trials, and 2 nonequivalent control group designs. All studies involved caregiver-child dyads, with child ages ranging from 5 to 37 months and having or at risk of autistic spectrum disorder (n=10, 56%), cerebral palsy (n=4, 22%), and other conditions (n=4, 22%). Synchronous videoconferencing was the primary modality for caregiver training and coaching (n=17, 94%) while 1 intervention used an Internet of Things system. Outcomes were identified in child communication (n=9, 50%), physical (n=6, 33%), social or emotional (n=6, 33%), and adaptive behavior (n=4, 22%), as well as caregiver implementation (n=12, 66%). Telehealth demonstrated comparable or superior effectiveness to traditional in-person methods in 2 studies. However, the focus on specific conditions and limited research on cognitive development were notable gaps.

Conclusions: Telehealth can be a viable alternative to traditional in-person early interventions for young children who have developmental disabilities and their families. It enhances accessibility and interactions between families and providers at a distance while promoting family-centered care. Challenges exist, including those of technological literacy, and the lack of research on cognitive outcomes must be addressed. Future work should explore more comprehensive interventions, including multidisciplinary approaches and expanded family outcomes, to solidify the role that telehealth plays in early intervention.

Trial Registration: PROSPERO CRD42024551286; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=551286

XSL•FO
RenderX

(J Med Internet Res 2025;27:e66442) doi:10.2196/66442

KEYWORDS

developmental disabilities; developmental delay; early intervention; telehealth; digital intervention; autistic spectrum disorder; cerebral palsy; family-centered care; multidisciplinary care; systematic review

Introduction

Child mortality rates have decreased dramatically in recent decades, but the prevalence of developmental disabilities in children is gradually increasing [1,2]. Developmental disabilities are a group of conditions due to an impairment in physical, learning, language, or behavior areas, including but not limited to autistic spectrum disorder (ASD), down syndrome (DS), attention-deficit/hyperactivity disorder, cerebral palsy (CP), and learning disabilities [3]. These generally manifest during the developmental period, and their effects typically persist throughout life [3].

Historically speaking, developmental disabilities have been predominantly recognized for their negative impacts on the physical and psychosocial health of affected individuals [4]. Indeed, developmentally disabled children and their families can expect to encounter an endless series of daily challenges that can undermine the overall functioning of the entire family [5,6]. However, positive experiences have also been reported by parents raising children who have developmental disabilities, such as their feelings of appreciation, improved tolerance and sensitivity toward others, opportunities for learning, and deeper familial connections [7,8]. Effective advocacy, including promoting family-centered interventions and socioeconomic support, can reinforce well-being and resilience in such families [9,10].

Early intervention is an umbrella term encompassing a range of services and support that are provided for developmentally vulnerable infants, toddlers, and their families to enable them to achieve their optimal competencies and quality of life [11,12]. Robust neurological evidence supports the strong recommendation to provide early intervention in the first 3 years of life, which is a period of high brain plasticity [13-19]. This addresses 5 critical domains of development—the cognitive, physical, adaptive behavioral, social or emotional, and communication domains—to mitigate developmental challenges and maximize children's developmental potential [13]. In addition, in that multidisciplinary and family-centered care (FCC) are critical principles for early intervention [14-18], the positive effects of early intervention extend to family outcomes including parents' competency and appraisal, family relationships, and family quality of life [19-21].

In spite of substantial research that has demonstrated the positive effects of early interventions on children who have developmental disabilities and their families [4,22], several factors continue to impede the accessibility of such services, including complex referral and provision systems, long wait times, and geographical distances [23,24]. The COVID-19 pandemic has only exacerbated the challenges faced by these families, reducing their access to in-person services [25]. Rural areas and underdeveloped societies tend to be more vulnerable to these issues, aggravating existing health disparities [26,27]. In response to these challenges, telehealth has emerged as a solution that can help address the limitations of traditional early intervention methods, such as accessibility issues and resource shortages.

For several decades, telehealth has been extending its sphere of influence and making remarkable advancements [28]. Recently, the COVID-19 pandemic has accelerated its growth, going beyond a proper alternative to becoming an essential method of global health care [29]. While the terminology of telehealth is often used interchangeably with other terms, including telemedicine, telepractice, and teleintervention, telehealth is the broadest term that encompasses all types of remote health care services delivered via technology, as suggested in multiple definitions of telehealth [30-32]. World Health Organization defines it as any method of delivering "health care services, where patients and providers are separated by distance," so long as they use "information and communication technologies for the exchange of information" [32]. Medicaid suggests a narrower definition of telehealth as permitting "2-way, real-time interactive communication between the patient and the physician or practitioner at the distant site" via audio or video components [30].

Previous research has investigated the effectiveness of telehealth across health conditions and health care settings [33,34]. Telehealth is a promising means of overcoming distance, contributing to the prevention of infections, and narrowing health disparities between urban and rural areas, as well as between developed and underdeveloped societies [28].

Young children who have developmental disabilities are among the populations that benefit from telehealth. Previous research has found that telehealth is as effective as or better than traditional methods of delivering rehabilitation, parent-mediated intervention, and parent coaching to children who have developmental disabilities [35-37]. A few reviews have demonstrated the success of early interventions via telehealth. Still, these include a wider range of ages, up to 7 years [38,39], despite the heterogeneity in children's developmental stages. Further, these reviews tend to limit their focus to specific diagnoses (eg, ASD, CP) or particular types of intervention (eg, mobile apps, websites) [40,41]. To understand the implementation and effectiveness of early intervention via telehealth precisely, it is indispensable to focus on children aged 0-3 years and not to limit types of developmental disability or delay and telehealth modalities.

This systematic review compiled and synthesized the literature on early intervention via telehealth for children who have a high risk of developmental disabilities and their families, focusing on evaluating the implementation process and the effectiveness of these interventions. The main research question, structured in Patient, Interested Intervention, Comparison, Outcome (PICO)

XSL•FO
RenderX

format, was "Is early intervention via telehealth (I) effective in improving outcomes (O) for infants and toddlers aged 0-3 years with or at high risk for developmental disabilities and their families (P), relative to conventional in-person intervention (C)?"

Methods

Methodology

The review was conducted and reported using systematic review methodologies from the Joanna Briggs Institute [42].

Selection Criteria

The inclusion and exclusion criteria are described in Textbox 1. In this study, early intervention via telehealth was defined as using information and communication technologies to enable active interactions between children or families and service providers, either synchronously or asynchronously. This definition reflects the nature of early intervention, which involves active engagement and collaboration of families and providers [13]. Thus, studies adopting telehealth modalities only for recording, monitoring, self-management, assessment, or diagnosis were excluded, as they do not involve the interactive process integral to early intervention practices.

Textbox 1. Inclusion and exclusion criteria for study selection.

Inclusion criteria

- Population: children aged 0-3 years with or at high risk of developmental disability
- Intervention: telehealth interventions involving interactions with a health care provider
- Outcomes: results including child or family outcomes
- Study design: empirical studies
- Year: January 2010 to April 2024
- Language: available in English

Exclusion criteria

- Population: children older than 3 years of age
- Intervention: adoption of telehealth modalities only for recording, monitoring, self-management, assessment, or diagnosis without interaction or feedback from health care providers
- Outcomes: results not reporting child outcomes
- Study design: nonempirical studies (eg, editorials, study protocols, or review studies)

Search Strategy

A comprehensive literature search of 4 bibliographic databases (PubMed, Embase, CINAHL, and Web of Science) was conducted to identify studies using free keywords and the index vocabulary tailored for each database. The search strategy adopted was developed in collaboration with a librarian expert in the field. The search terms were a combination of three categories: (1) developmental disabilities, (2) telehealth, and (3) early intervention (Multimedia Appendix 1).

Study Selection

For the study selection, all search references were imported into Covidence (Veritas Health Innovation), a management tool used to review studies. After the exclusion of duplicates, 2 reviewers independently screened the title and abstract, followed by a full-text review. Disagreements were resolved by the corresponding author. For data extraction, a template was developed using the software program Microsoft Excel to collate specific information from the included studies.

Quality Assessment

Quality assessment was conducted using the mixed methods appraisal tool [43] to ensure the methodological quality of the studies included (Multimedia Appendix 2 [44-61]). This tool includes 5 distinct checklists for empirical studies across categories. Each criterion could be assigned to "yes," "no," or "can't tell." The total quality score for each article ranged from 0% to 100%, where higher scores indicated higher methodological quality. Two researchers independently appraised each study, and conflicts or criteria that were answered as "can't tell" were discussed with the corresponding author. The threshold for inclusion was 60% or more.

Data Extraction and Synthesis

Due to the heterogeneity of the populations, interventions, and outcome measures across studies, a narrative synthesis was conducted following the guidance of Popay et al [62]. Two researchers investigated the selected studies thoroughly to develop a standardized tabular form for the extraction and synthesis of the data. The synthesis included (1) the general characteristics of the studies that were included (authors, study location, study design, target functions, and technological modalities); (2) their population characteristics (diagnosis or condition, sample size, age, and sex); (3) the intervention characteristics (program title or theoretical base, provider or profession, description, control, and duration), and (4) the outcome characteristics (variables, measurements, and main findings). The frequency of the technological modalities and target functions across the included studies were described in independent matrix tables. Minor disagreements that arose during the process were resolved by consensus between the

XSL•FO
RenderX

investigators. The corresponding author was brought in when consensus could not be reached.

Results

Overview

An electronic search yielded 2468 studies, of which 18 were ultimately included in the review. The screening process is detailed in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram, shown in Figure 1.

Figure 1. Flow diagram of study selection following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline.

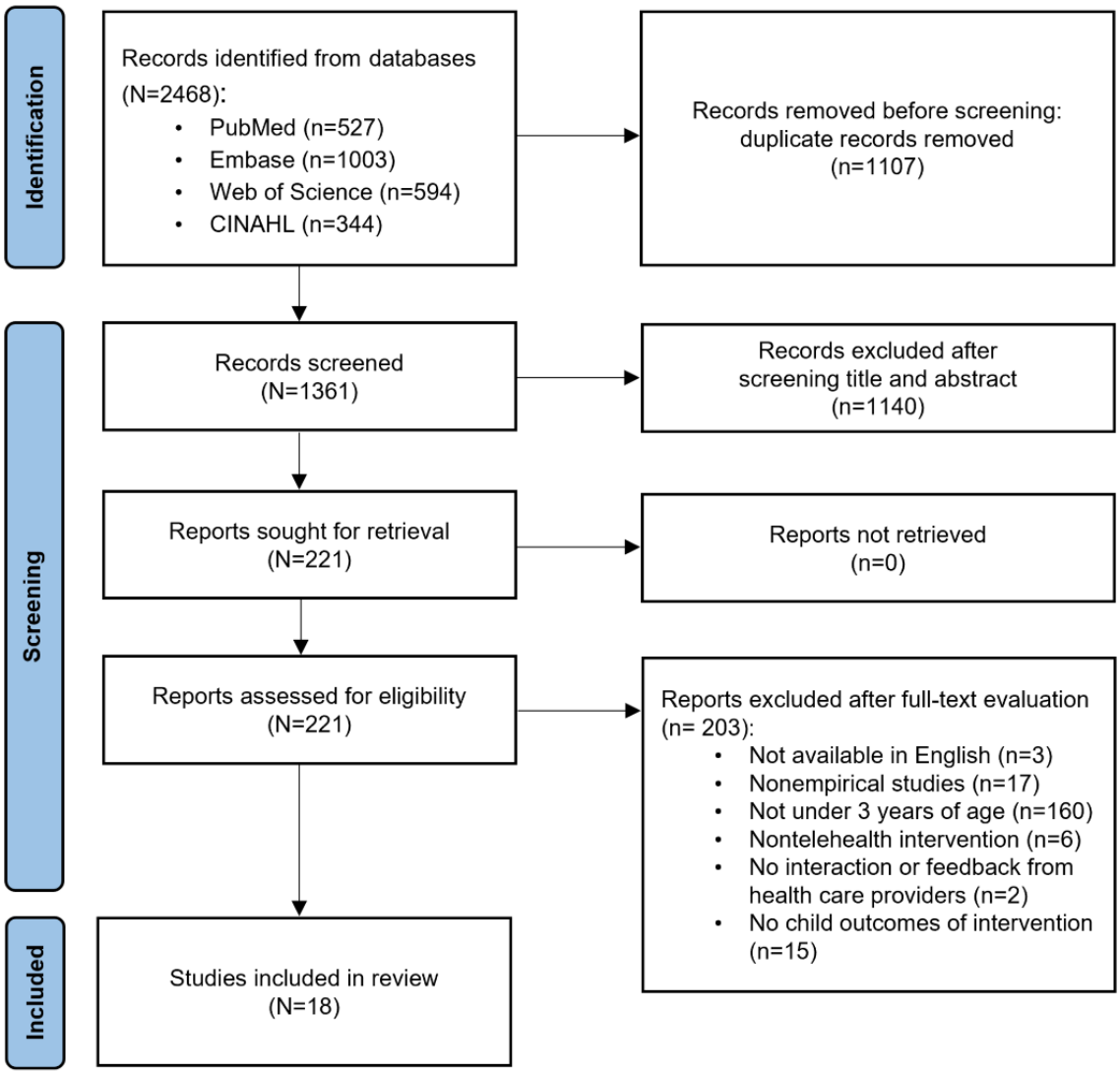

Quality Assessment

Of the 18 selected studies, 4 studies were evaluated using the checklist for quantitative randomized controlled trials, all scoring 100% (Multimedia Appendix 2). A checklist for quantitative non-randomized research was used to assess 13 studies. Of these, 7 studies scored 100%, 5 studies scored 80%, and 1 study scored 60%, due to an unclear indication of participant eligibility (n=3, 50%), an incomplete exhibition of outcome data (n=2, 33%), or an inadequate consideration of potential confounders (n=2, 33%). One study was categorized as a mixed methods study and scored 100% on the quality assessment.

General Characteristics of the Reviewed Studies

Of the 18 studies included, 13 were published between 2020 and 2024, following the onset of the COVID-19 pandemic (Multimedia Appendix 3 [44-61]). The studies were primarily conducted in the United States (n=10, 56%), with others from Canada (n=2, 11%), Brazil (n=3, 17%), Italy or Denmark (n=1, 6%), Iran (n=1, 6%), and Sweden (n=1, 6%). The designs of the studies were varied, including single-case designs (n=11, 61%), randomized controlled trials (n=4, 22%), and nonequivalent control group designs (n=3, 17%). The sizes of the samples ranged from single cases to 82 participants, including 8 studies featuring 6 or fewer participants.

XSL•FO
RenderX

Participant Characteristics

Most studies focused on children at risk for or diagnosed with ASD (n=10, 56%; Multimedia Appendix 3 [44-61]). Other studies included children at risk for CP (n=4, 22%), born preterm (n=1, 6%), with DS or developmental delays (n=1, 6%), with DS and visual impairment (n=1, 6%), or with a hearing impairment (n=1, 6%). Among the 12 that specified the sex of the child participant, 147 were boys (62.3%), and 89 were girls (37.7%). The ages of the children ranged from 5 to 37 months.

All studies, with the exception of one, which included a child and both of her parents [44], recruited participants as caregiver-child dyads. In all, 14 studies specified the caregiver's sex; 170 (88.1%) were female, and 23 (11.9%) were male. Although none of the studies purposefully recruited mother-child dyads, seven ended up having 100% female caregivers, all of whom were the child's mothers, except for 1 grandmother [45]. Eight studies reported the ages of the caregivers, which ranged from 20 to 48 years.

Intervention Characteristics

Technological Modalities

Videoconferencing using devices such as smartphones, tablets, or personal computers was predominantly used (n=17, 94%) to deliver interventions (Table 1). The only exception was Sgandurra et al [46] who used an Internet of Things (IoT) baby gym, which enabled remote, highly personalized rehabilitation and monitoring. Websites or multimedia modules (n=2, 11%) and data sharing platforms or clouds (n=8, 44%) were used to facilitate data sharing between providers and families. Two studies used an IoT system or a wearable device to monitor children's functioning [46,47].

Table 1. Frequencies of telehealth technological modalities used in the reviewed studies.

Author (year)	Videoconferencing	Website or module	Data sharing platform or cloud	IoT[a] system or device
Akemoğlu et al (2022) [56]	✓	✓	✓	
Azzano et al (2023) [44]	✓	✓		
Bailey et al (2024) [48]	✓		✓	
Brian et al (2022) [45]	✓			
Daczewitz et al (2020) [49]	✓			
de Almeida Rodrigues et al (2023) [61]	✓			
Kunze et al (2021) [50]	✓		✓	
Lee et al (2023) [60]	✓		✓	
Lima et al (2023) [55]	✓			
Meadan et al (2016) [51]	✓		✓	
Pietruszewski et al (2020) [47]	✓			✓
Sadeghi et al (2022) [59]	✓			
Schlichting et al (2022) [52]	✓			
Sgandurra et al (2017) [46]				✓
Svensson et al (2024) [53]	✓			
Vismara et al (2013) [54]	✓		✓	
Vismara et al (2018) [58]	✓		✓	
Vismara et al (2012) [57]	✓		✓	

[a]IoT: Internet of Things.

Intervention Contents

The intervention characteristics are given in Multimedia Appendix 4 [44-61]. Among the 17 studies that used videoconferencing, all but 1 study [45], which included both individual and group sessions, conducted one-on-one sessions to provide parent training, coaching, or both. In the sessions, the providers delivered knowledge of strategy, answered questions, monitored caregiver-child interactions, and provided real-time feedback.

Nine out of 17 studies had children participate in live sessions to observe caregivers' strategy implementation, monitor their progress, and provide coaching at the moment [44,47-54]. The remaining 8 studies involved only researchers and caregivers, focusing on the delivery of strategies so that caregivers could apply them to their children later. Four studies prescribed dosage for caregivers' self-practice outside of the videoconferencing sessions [54,55,57, 60], where 3 required caregivers to log their progress, whether online or offline. In addition, 7 studies individually tailored the goals and plans of the programs as a result of discussions conducted between providers and caregivers, which increased the extent of the individualization of the interventions offered [49,51,54,55,57,58,61].

XSL•FO
RenderX

In addition to videoconferencing, in 5 studies, the caregivers [53,54,56-58] received training in multimedia modules provided through web platforms or DVDs, primarily containing video and written instructions. The caregivers were instructed to use the modules for self-teaching in a predetermined sequence or at their own pace. The IoT baby gym system that was used by Sgandurra et al [46] provided daily goal-directed activities based on the child's condition monitored through video cameras, wearable sensors, and other sensorized components, such as mats and toys.

The interventions for all 18 studies were provided by researchers who had relevant experience with the target population. Nine studies indicated the professions or professional qualifications of service providers, which included special educators (n=3, 17%), physiotherapists/occupational therapists (n=3), behavior analysts (n=2, 11%), and psychologists (n=1, 6%). The average intervention was approximately 10.2 weeks long, ranging from 4 to 29 weeks, except for 1 study, which only stated the number of sessions (42-57 sessions) [48]. The average frequency was approximately 1.6 sessions per week, with the 4 aforementioned studies including additional self-practice components [52,53,55,60]. Each session continued for an average of approximately 54.3 minutes, with a range of 15 to 90 minutes.

Effectiveness of Interventions

All of the included studies suggested the overall effectiveness of the interventions on child outcomes, with some degree of variability (Multimedia Appendices 3 and 4 [44-61]). In each of the 18 studies, the interventions targeted one or more domains of child development (Table 2). Among the 5 early-intervention domains, the development of communication was most frequently addressed (n=9, 50%), as measured through variables such as communicative behaviors (eg, initiation, response, and vocal imitation) and the number of words used. Studies targeting social/emotional development (n=6, 33%) assessed social behaviors, including joint attention, social turn-taking, and behavior imitation. General or specific motor skills were measured for physical development (n=6, 33%). The development of adaptive behavior (n=4, 22%) was evaluated with flexible and inflexible behavior, repetitive and restricted behaviors, instrumental turn-taking, and daily activities. It should be noted that cognitive development was not specifically targeted as part of any of the included studies.

Table 2. Frequencies of child functions targeted in the reviewed studies based on early intervention domains [13].

Author (year)	Target function domain				
	Cognitive (n=0)	Physical (n=6)	Adaptive behavior (n=4)	Social or emotional (n=6)	Communication (n=9)
Akemoğlu et al (2022) [56]					✓
Azzano et al (2023) [44]				✓	✓
Bailey et al (2024) [48]					✓
Brian et al (2022) [45]					✓
Daczewitz et al (2020) [49]					✓
de Almeida Rodrigues et al (2023) [61]		✓	✓		
Kunze et al (2021) [50]			✓		
Lee et al (2023) [60]			✓	✓	
Lima et al (2023) [55]		✓		✓	
Meadan et al (2016) [51]					✓
Pietruszewski et al (2020) [47]		✓			
Sadeghi et al (2022) [59]			✓		
Schlichting et al (2022) [52]		✓			
Sgandurra et al (2017) [46]		✓			
Svensson et al (2024) [53]		✓			
Vismara et al (2012) [57]				✓	✓
Vismara et al (2013) [54]				✓	✓
Vismara et al (2018) [58]				✓	✓

More particularly, among the child outcomes, the communication skills of children with or at risk for ASD were the most frequently reported (n=7, 39%), followed by the motor function of infants having or at biological risk for CP (n=4, 22%). The adaptive behavior of children with or at risk for ASD has been assessed in 3 studies [50,59,60]. Six studies included control groups, three of which provided in-person or traditional care to the control groups. Two studies reported nonsignificant differences between the invention (telehealth) and the control (in-person or traditional care) groups.

The outcomes for caregivers and families were reported in 12 studies, among which 10 demonstrated increased or appropriate levels of caregivers' implementation fidelity for learned

XSL•FO
RenderX

strategies. Implementation-related outcomes, such as acceptability, feasibility, satisfaction, and cost efficiency were also measured. In addition, interventions were reported to improve clinical outcomes for caregivers and families, including parenting stress (n=2, 11%), caregiver engagement (n=1, 6%), self-efficacy (n=1, 6%), and home environment (n=1, 6%), although these did not consistently reach statistical significance.

Discussion

Principal Findings

This review synthesized the existing literature on early interventions through telehealth for young children who have developmental risks and examined their implementation and effectiveness. A distinguishing feature of this review is that the child participants were restricted to the ages of 0-3 years, strictly following the extensively recommended criteria for early intervention. Thereby, this study enabled a concrete inspection of current practices in telehealth interventions that were precisely aimed at young children during their first 3 years and their impact.

Effectiveness

This review identified the overall effectiveness of early interventions via telehealth for both children and their families. In particular, children showed improvement in their targeted abilities and behaviors, in general, corresponding to the caregivers' enhanced or sufficient fidelity for implementing learned strategies. The effectiveness was comparable to those of in-person approaches. This finding is consistent with previous studies that have addressed the effectiveness of telehealth interventions for children and youths with developmental disabilities [35]. McCarthy et al [63] and Bharat et al [40] suggested a potential for telehealth or mobile apps in delivering early intervention for children aged under 6 with ASD or hearing difficulties. For older children and adolescents, multiple reviews have indicated that telehealth interventions for ASD [41,64,65] and CP [66] are at least as effective as traditional methods. Telehealth interventions demonstrate sustained effectiveness throughout childhood in children who have developmental challenges.

Scope of Child Conditions and Target Outcomes

However, early interventions through telehealth included in the studies were less comprehensive, with limited populations or target functions relative to traditional in-person interventions. Most interventions targeted either communication development in children who have ASD during their second or third years of life or motor development in infants with CP; none of the included studies targeted or examined child outcomes in relation to the cognitive domain. Cognitive ability serves as the foundation for early childhood development, and it affects overall health and socioeconomic status during the later stages of life [67,68]. This underscores the importance of early intervention for enhancing cognitive development [69,70], highlighting further attention to this omission. The alienation of cognitive outcomes is also observed in telehealth interventions for children who have developmental disabilities and are aged more than 3 years [41,65,71]. This suggests that adopting telehealth in comprehensive interventions including the cognitive domain for children having developmental disabilities or delays could accompany difficulties. Some conditions require specialized skills that are difficult to deliver virtually [72], and certain assessments require long periods or physical materials that may not be feasible using telehealth [73,74].

Despite these challenges, a comprehensive approach is uncompromisable for early intervention for children with developmental disabilities [75,76]. As a result of the interrelated nature of all aspects of development, such that the growth or deficiency in 1 area influences that of others, thereby either promoting or hindering overall development [13]. Thus, the study of early intervention for young children with developmental challenges has consistently striven to encompass cognitive, physical, adaptive behavior, social or emotional, and communication domains, and it has even expanded to include sensory integration, nutrition, and sleep [75,77-80]. To establish telehealth as a viable method for the delivery of early intervention, future research is encouraged to cover a more comprehensive range of developmental challenges.

Telehealth Modalities

The most widely used form of early intervention through telehealth was synchronous caregiver training or coaching through videoconferencing, engaging caregivers as the main implementers. The caregivers generally reported satisfaction with telehealth care. Real-time videoconferencing sessions between providers and caregivers produced positive feedback, in particular when it was combined with multimedia modules.

Telehealth interventions commonly meet certain barriers, including technological issues, lack of digital literacy, and negative perceptions of telehealth [81,82]. In pediatric care, telehealth assigns extra responsibility to caregivers for delivering interventions and maintaining an optimal environment for remote communication, which some find burdensome [83,84]. The caregivers in the included studies reported occasional internet issues and discomfort in moderating the home setting; however, these did not critically reduce their satisfaction with telehealth modalities.

Consistent reports of decent social validity could be attributed to the participants' ages and locations. Youths, internet access, and technological competency facilitate the use of telehealth, while old age, technological disconnection, and insufficient understanding of telehealth pose barriers [81,85]. Studies of early-intervention studies typically involve young individuals who have high digital literacy [86] and are often conducted in environments having stable internet connections, which likely contributes to positive feedback on telehealth. Future research should account for the digital divide according to age and region, as it prevents the generalizability of the results.

Family-Centered Care

The concept of early intervention is grounded in the principle of FCC, as has been established in key definitions and purposes emphasizing enhancements to parental capacity and supporting parent-mediated interventions [11,13]. The integration of telehealth into early intervention strengthens home-based family

XSL•FO
RenderX

participation and caregiver competence [87-89], promoting active family engagement within a natural environment, and leading to favorable outcomes for both child and family [39].

While the included studies incorporated FCC, there is potential for further improvement in its application in early intervention, in particular through telehealth [90]. Dyad-recruiting and caregiver coaching/training, the most common strategies used by the included studies, are reported to be effective in child development, caregiver implementation, and child-caregiver interaction [13,22,91-93]. However, a notable sex imbalance was seen in caregivers who were participating in the studies, with mothers predominantly participating. This inclination toward the mother aligns with prior FCC research, which attributes it to traditional sex roles, lack of tailored services, and limited awareness of both families and providers [22,94-97]. Nonetheless, fathers' engagement is linked to the improved cognitive, emotional, behavioral, and psychopathological development of young children [98-100], and involving siblings as playmates or models is associated with enhanced peer relationships and social skills [101,102]. Future research should explore strategies to increase father and sibling involvement in FCC interventions.

In addition, most included studies had family outcomes that were limited to caregivers' implementation in the absence of clinical outcomes, underscoring the need for further exploration of caregiver and family outcomes to develop and evaluate FCC-based early interventions through telehealth. Research indicates that parents' physical and mental health directly affects their ability to provide care and support for their children, which, in turn, impacts children's developmental progress and overall well-being [103,104]. By contrast, the severity of children's developmental disabilities can also affect parental health, producing an interplay within families [105-107]. Therefore, more comprehensive research is required to explore the whole family's outcomes when evaluating the effects of telehealth early interventions.

Multidisciplinary Approach

A multidisciplinary approach should also be considered, as each intervention in the included studies was conducted by individuals from a few professions, including physiotherapists, behavior analysts, and special educators. While the focused expertise of the providers could allow for intensive interventions in particular areas, collaboration among diverse professions fosters a comprehensive understanding of conditions, integrative problem-solving, and effective risk management, ultimately enhancing the quality and outcomes of the intervention [108-110]. A multidisciplinary team, including education, medical care, rehabilitation, community service, and family, can reflect the holistic nature of children who have developmental disabilities and their families [111,112]. In particular, thanks to early intervention, a multidisciplinary approach can be posited as an essential principle for the promotion of individualization and family engagement and achieving optimal development [113].

Limitations

To the best of our knowledge, this is the first systematic review to synthesize the use of telehealth in early intervention for young children aged 0-3 years and their families, involving developmental challenges. However, this study has some limitations.

First, a large portion of the included studies used single-case designs, which pose challenges in generalizing findings to broader populations. This limitation hindered our analysis to compare the overall effectiveness of telehealth interventions to traditional methods or evaluate variability in effectiveness across different groups based on factors such as participants' age, sex, and developmental disabilities of the children.

Consequently, we were unable to calculate effect sizes for a statistical synthesis, limiting our ability to draw definitive conclusions about the effectiveness of early intervention delivered through telehealth.

While our search terms were meticulously developed in collaboration with an institutional librarian specializing in medical research, the scope of our search could be further refined. The exclusion of gray literature may have omitted informal data or professional experiences beyond traditional research settings. Additionally, limiting the review to published studies introduces a potential risk of publication bias, as studies with positive results are more likely to be published. Future research incorporating a wider range of integrative works, including gray literature, could enhance the generalizability and provide a more comprehensive understanding of telehealth's impact on early intervention.

Finally, the findings predominantly come from developed, primarily Western countries. This limitation may be due to restricting the search to English-language studies, as well as the concentration of related research in these regions. In many low- and middle-income countries, early intervention remains a lesser-known concept [114], and the necessary infrastructure for telehealth delivery is often lacking [115]. These considerations may explain the geographic biases of this review; nonetheless, the findings serve as a foundation for guiding future research and policies to expand access to early intervention via telehealth in underserved communities.

Conclusions

This review highlights the potential that telehealth has for early intervention for young children having developmental challenges and their families. By synthesizing the existing literature, we observed that telehealth, mainly in the form of synchronous caregiver training or coaching, can effectively enhance child and family outcomes across various areas. Despite its limitations, the adoption of telehealth technologies in early intervention is a promising direction for researchers, clinicians, and educators to deliver holistic, FCC for young children with developmental disabilities at a distance. This integrative approach will ultimately benefit children who have developmental challenges and their families in remote or underserved areas by improving developmental outcomes and the quality of life within families. Finally, we call for additional high-quality research that expands the scope of telehealth early

XSL•FO
RenderX

interventions and ensures a comprehensive, multidisciplinary approach for children with developmental disabilities and their families, which can generate robust and generalizable evidence.

Acknowledgments

This study received no external funding, and the authors have no financial interest in the interventions discussed in the review.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search terms.
[DOCX File , 14 KB - jmir_v27i1e66442_app1.docx]

Multimedia Appendix 2
The quality assessment based on the MMAT (Mixed Methods Appraisal Tool) tool.
[DOCX File , 37 KB - jmir_v27i1e66442_app2.docx]

Multimedia Appendix 3
Summary of main characteristics of the included studies' locations, designs, participants, primary technological modalities, and main findings.
[DOCX File , 36 KB - jmir_v27i1e66442_app3.docx]

Multimedia Appendix 4
Characteristics of early interventions via telehealth and measurements for children aged 0-3 years in the reviewed studies.
[DOCX File , 43 KB - jmir_v27i1e66442_app4.docx]

Multimedia Appendix 5
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 135 KB - jmir_v27i1e66442_app5.pdf]

References

1. Li Q, Li Y, Zheng J, Yan X, Huang J, Xu Y, et al. Prevalence and trends of developmental disabilities among US children and adolescents aged 3 to 17 years, 2018-2021. Sci Rep 2023;13(1):17254 [FREE Full text] [doi: 10.1038/s41598-023-44472-1] [Medline: 37828147]
2. Olusanya BO, Davis AC, Wertlieb D, Boo N, Nair M, Halpern R, et al. Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Glob Health 2018;6(10):e1100-e1121 [FREE Full text] [doi: 10.1016/s2214-109x(18)30309-7]
3. Holm VA. Developmental disabilities: delivery of medical care for children and adults. JAMA 1989;262(20):2935-2936. [doi: 10.1001/jama.1989.03430200183054]
4. Smythe T, Zuurmond M, Tann C, Gladstone M, Kuper H. Early intervention for children with developmental disabilities in low and middle-income countries—the case for action. Int Health 2021;13(3):222-231 [FREE Full text] [doi: 10.1093/inthealth/ihaa044] [Medline: 32780826]
5. Leeman J, Crandell JL, Lee A, Bai J, Sandelowski M, Knafl K. Family functioning and the well-being of children with chronic conditions: a meta-analysis. Res Nurs Health 2016;39(4):229-243. [doi: 10.1002/nur.21725] [Medline: 27128982]
6. Masefield SC, Prady SL, Sheldon TA, Small N, Jarvis S, Pickett KE. The caregiver health effects of caring for young children with developmental disabilities: a meta-analysis. Matern Child Health J 2020;24(5):561-574 [FREE Full text] [doi: 10.1007/s10995-020-02896-5] [Medline: 32048172]
7. Sutherland D, Griffin J, Flynn S, Thompson PA, Hastings RP. Positive family connections for families of children with a developmental disability: Cluster randomized controlled trial. J Fam Psychol 2024;38(4):559-570. [doi: 10.1037/fam0001211] [Medline: 38497720]

XSL•FO
RenderX

8. Beighton C, Wills J. Are parents identifying positive aspects to parenting their child with an intellectual disability or are they just coping? A qualitative exploration. J Intellect Disabil 2017;21(4):325-345 [FREE Full text] [doi: 10.1177/1744629516656073] [Medline: 27352854]
9. Emmanuel CJ, Knafl KA, Hodges EA, Docherty SL, O'Shea TM, Santos HP. Family members' experience of well-being as racial/ethnic minorities raising a child with a neurodevelopmental disorder: a qualitative meta-synthesis. Res Nurs Health 2022;45(3):314-326. [doi: 10.1002/nur.22217] [Medline: 35141915]
10. Rios K, Aleman-Tovar J, Burke M. Exploring patterns of advocacy and well-being of parents of children with intellectual and developmental disabilities. Intellect Dev Disabil 2021;59(6):459-471. [doi: 10.1352/1934-9556-59.6.459] [Medline: 34814169]
11. Dunst CJ. Early intervention for infants and toddlers with developmental disabilities. In: Odom S, Horner R, Snell M, Blacher J, editors. Handbook on Developmental Disabilities. New York, NY: Guilford Press; 2007.
12. American Academy of Pediatrics Council on Children With Disabilities, Duby JC. Role of the medical home in family-centered early intervention services. Pediatrics 2007;120(5):1153-1158. [doi: 10.1542/peds.2007-2638] [Medline: 17974754]
13. McIntyre LL, Kunze M, Barton H. Early intervention for children with intellectual and developmental disabilities. In: Glidden LM, Abbeduto L, McIntyre LL, Tassé MJ, editors. APA Handbook of Intellectual and Developmental Disabilities: Clinical and Educational Implications: Prevention, Intervention, and Treatment, Vol. 2. Washington, DC: American Psychological Association; 2021.
14. Mota LAT, Silva MZ, Dos Santos M, Pfeifer LI. The processes and outcomes related to 'family-centred care' in neuromotor and functional rehabilitation contexts for children with cerebral palsy: a scoping review. Child Care Health Dev 2024;50(3):e13271. [doi: 10.1111/cch.13271] [Medline: 38738842]
15. Shen J, Zhao S, Horn T, Benkart R, Busch T, Vrabec A, et al. Family matters: a systematic review and meta-analysis on the efficacy of family-oriented interventions for children with acquired brain injuries. Clin Psychol Rev 2023;99:102218 [FREE Full text] [doi: 10.1016/j.cpr.2022.102218] [Medline: 36401964]
16. Fazel M, Townsend A, Stewart H, Pao M, Paz I, Walker J, et al. Integrated care to address child and adolescent health in the 21st century: a clinical review. JCPP Adv 2021;1(4):e12045 [FREE Full text] [doi: 10.1002/jcv2.12045] [Medline: 37431408]
17. Marshall KH, Pincus HA, Tesson S, Lingam R, Woolfenden SR, Kasparian NA. Integrated psychological care in pediatric hospital settings for children with complex chronic illness and their families: a systematic review. Psychol Health 2024;39(4):452-478. [doi: 10.1080/08870446.2022.2072843] [Medline: 35635028]
18. Leach KF, Stack NJ, Jones S. Optimizing the multidisciplinary team to enhance care coordination across the continuum for children with medical complexity. Curr Probl Pediatr Adolesc Health Care 2021;51(12):101128. [doi: 10.1016/j.cppeds.2021.101128]
19. Bailey DB, Hebbeler K, Spiker D, Scarborough A, Mallik S, Nelson L. Thirty-six-month outcomes for families of children who have disabilities and participated in early intervention. Pediatrics 2005;116(6):1346-1352. [doi: 10.1542/peds.2004-1239] [Medline: 16322157]
20. McCarthy E, Guerin S. Family-centred care in early intervention: a systematic review of the processes and outcomes of family-centred care and impacting factors. Child Care Health Dev 2022;48(1):1-32. [doi: 10.1111/cch.12901] [Medline: 34324725]
21. Chiu S, Lin I, Chou Y, Chien L. Family quality of life among Taiwanese children with developmental delay before and after early intervention. J Intellect Disabil Res 2020;64(8):589-601. [doi: 10.1111/jir.12754] [Medline: 32567788]
22. Ward R, Reynolds JE, Pieterse B, Elliott C, Boyd R, Miller L. Utilisation of coaching practices in early interventions in children at risk of developmental disability/delay: a systematic review. Disabil Rehabil 2020;42(20):2846-2867. [doi: 10.1080/09638288.2019.1581846] [Medline: 31021669]
23. Sapiets SJ, Totsika V, Hastings RP. Factors influencing access to early intervention for families of children with developmental disabilities: a narrative review. J Appl Res Intellect Disabil 2021;34(3):695-711 [FREE Full text] [doi: 10.1111/jar.12852] [Medline: 33354863]
24. Jimenez ME, Fiks AG, Shah LR, Gerdes M, Ni AY, Pati S, et al. Factors associated with early intervention referral and evaluation: a mixed methods analysis. Acad Pediatr 2014;14(3):315-323. [doi: 10.1016/j.acap.2014.01.007] [Medline: 24767785]
25. Shorey S, Lau L, Tan J, Ng ED, Aishworiya R. Families with children with neurodevelopmental disorders during COVID-19: a scoping review. J Pediatr Psychol 2021;46(5):514-525 [FREE Full text] [doi: 10.1093/jpepsy/jsab029] [Medline: 33758930]
26. McAuliffe T, Vaz S, Falkmer T, Cordier R. A comparison of families of children with autism spectrum disorders in family daily routines, service usage, and stress levels by regionality. Dev Neurorehabil 2017;20(8):483-490. [doi: 10.1080/17518423.2016.1236844] [Medline: 27739909]
27. Smythe T, Scherer N, Nanyunja C, Tann CJ, Olusanya BO. Strategies for addressing the needs of children with or at risk of developmental disabilities in early childhood by 2030: a systematic umbrella review. BMC Med 2024;22(1):51 [FREE Full text] [doi: 10.1186/s12916-024-03265-7] [Medline: 38302917]

XSL•FO
RenderX

28. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state Of telehealth evidence: a rapid review. Health Aff 2018;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
29. Latifi R, Doarn CR. Perspective on COVID-19: finally, telemedicine at center stage. Telemed J E Health 2020;26(9):1106-1109. [doi: 10.1089/tmj.2020.0132] [Medline: 32408804]
30. Telehealth. Medicaid. URL: https://www.medicaid.gov/medicaid/benefits/telehealth/index.html [accessed 2024-12-11]
31. ATA A. Telehealth: defining 21st century care;. URL: https://www.americantelemed.org/resource/why-telemedicine/ [accessed 2024-12-11]
32. WHO Global Observatory for eHealth. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth. Global Observatory for eHealth Series, 2. Geneva: World Health Organization; 2010. URL: https://iris.who.int/handle/10665/44497
33. Garfan S, Alamoodi A, Zaidan B, Al-Zobbi M, Hamid RA, Alwan JK, et al. Telehealth utilization during the Covid-19 pandemic: a systematic review. Comput Biol Med 2021;138:104878 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
34. Khoshrounejad F, Hamednia M, Mehrjerd A, Pichaghsaz S, Jamalirad H, Sargolzaei M, et al. Telehealth-based services during the COVID-19 pandemic: a systematic review of features and challenges. Front Public Health 2021;9:711762 [FREE Full text] [doi: 10.3389/fpubh.2021.711762] [Medline: 34350154]
35. Ogourtsova T, Boychuck Z, O'Donnell M, Ahmed S, Osman G, Majnemer A. Telerehabilitation for children and youth with developmental disabilities and their families: a systematic review. Phys Occup Ther Pediatr 2023;43(2):129-175. [doi: 10.1080/01942638.2022.2106468] [Medline: 36042567]
36. Yosep I, Prayogo SA, Kohar K, Andrew H, Mardhiyah A, Amirah S, et al. Managing autism spectrum disorder in the face of pandemic using internet-based parent-mediated interventions: a systematic review of randomized controlled trials. Children 2022;9(10):1483 [FREE Full text] [doi: 10.3390/children9101483] [Medline: 36291419]
37. Wu X, Deng H, Jian S, Chen H, Li Q, Gong R, et al. Global trends and hotspots in the digital therapeutics of autism spectrum disorders: a bibliometric analysis from 2002 to 2022. Front Psychiatry 2023;14:1126404 [FREE Full text] [doi: 10.3389/fpsyt.2023.1126404] [Medline: 37255688]
38. Pacia C, Holloway J, Gunning C, Lee H. A systematic review of family-mediated social communication interventions for young children with autism. Rev J Autism Dev Disord 2022;9(2):208-234 [FREE Full text] [doi: 10.1007/s40489-021-00249-8] [Medline: 33821200]
39. Retamal-Walter F, Waite M, Scarinci N. Exploring engagement in telepractice early intervention for young children with developmental disability and their families: a qualitative systematic review. Disabil Rehabil Assist Technol 2023;18(8):1508-1521. [doi: 10.1080/17483107.2022.2048098] [Medline: 35287526]
40. Bharat R, Uzaina YT, Yadav T, Niranjan S, Kurade P. mHealth apps delivering early intervention to support parents of children with autism spectrum disorder: a scoping review. Indian Pediatr 2023;60(3):224-230. [doi: 10.1007/s13312-023-2840-1]
41. de Nocker YL, Toolan CK. Using telehealth to provide interventions for children with ASD: a systematic review. Rev J Autism Dev Disord 2023;10(1):82-112 [FREE Full text] [doi: 10.1007/s40489-021-00278-3] [Medline: 34306962]
42. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews. Int J Evid Based Healthc 2015;13(3):132-140. [doi: 10.1097/xeb.0000000000000055]
43. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018;34(4):285-291. [doi: 10.3233/EFI-180221]
44. Azzano A, Vause T, Ward R, Feldman MA. Telehealth parent training for a young child at risk for autism spectrum disorder. Behav Interv 2022;38(1):140-158. [doi: 10.1002/bin.1917]
45. Brian J, Solish A, Dowds E, Roth I, Bernardi K, Perry K, et al. "Going Mobile"—increasing the reach of parent-mediated intervention for toddlers with ASD via group-based and virtual delivery. J Autism Dev Disord 2022;52(12):5207-5220 [FREE Full text] [doi: 10.1007/s10803-022-05554-7] [Medline: 35608785]
46. Sgandurra G, Lorentzen J, Inguaggiato E, Bartalena L, Beani E, Cecchi F, et al. A randomized clinical trial in preterm infants on the effects of a home-based early intervention with the 'CareToy System'. PLoS One 2017;12(3):e0173521 [FREE Full text] [doi: 10.1371/journal.pone.0173521] [Medline: 28328946]
47. Pietruszewski L, Burkhardt S, Yoder PJ, Heathcock J, Lewandowski DJ, Maitre NL. Protocol and feasibility-randomized trial of telehealth delivery for a multicomponent upper extremity intervention in infants with asymmetric cerebral palsy. Child Neurol Open 2020;7:2329048X20946214. [doi: 10.1177/2329048x20946214]
48. Bailey KM, Rodgers ME, Quinn ED, Thompson S, Nietfeld J, Kaiser AP. Just-in-time: a caregiver-mediated intervention for toddlers with autism. Am J Speech Lang Pathol 2024;33(3):1209-1225. [doi: 10.1044/2023_ajslp-23-00279]
49. Daczewitz M, Meadan-Kaplansky H, Borders C. PiCs: Telepractice coaching for a parent of a child who is hard-of-hearing. Deafness Educ Int 2019;22(2):113-138. [doi: 10.1080/14643154.2019.1587235]
50. Kunze MG, Machalicek W, Wei Q, St Joseph S. Coaching via telehealth: caregiver-mediated interventions for young children on the waitlist for an autism diagnosis using single-case design. J Clin Med 2021;10(8):1654 [FREE Full text] [doi: 10.3390/jcm10081654] [Medline: 33924440]

XSL•FO
RenderX

51. Meadan H, Snodgrass MR, Meyer LE, Fisher KW, Chung MY, Halle JW. Internet-based parent-implemented intervention for young children with autism. J Early Interv 2016;38(1):3-23. [doi: 10.1177/1053815116630327]
52. Schlichting T, Martins da Silva K, Silva Moreira R, Marques de Moraes MV, Cicuto Ferreira Rocha NA, Boyd RN, et al. Telehealth program for infants at risk of cerebral palsy during the Covid-19 pandemic: a pre-post feasibility experimental study. Phys Occup Ther Pediatr 2022;42(5):490-509. [doi: 10.1080/01942638.2022.2057209] [Medline: 35341469]
53. Svensson K, Sundelin H, Eliasson A. Outcomes of a parent-delivered baby-mCIMT model for infants at high risk of unilateral cerebral palsy using remote coaching in telerehabilitation. Children 2024;11(1):101 [FREE Full text] [doi: 10.3390/children11010101] [Medline: 38255414]
54. Vismara LA, McCormick C, Young GS, Nadhan A, Monlux K. Preliminary findings of a telehealth approach to parent training in autism. J Autism Dev Disord 2013;43(12):2953-2969. [doi: 10.1007/s10803-013-1841-8] [Medline: 23677382]
55. Lima CRG, Abreu RWF, Verdério BN, Brugnaro BH, Santos MMD, Dos Santos AN, et al. Early intervention involving specific task-environment-participation (STEP) protocol for infants at risk: a feasibility study. Phys Occup Ther Pediatr 2023;43(3):303-320. [doi: 10.1080/01942638.2022.2142084] [Medline: 36329671]
56. Akemoğlu Y, Laroue D, Kudesey C, Stahlman M. A module-based telepractice intervention for parents of children with developmental disabilities. J Autism Dev Disord 2022;52(12):5177-5190. [doi: 10.1007/s10803-022-05549-4] [Medline: 35391629]
57. Vismara LA, Young GS, Rogers SJ. Telehealth for expanding the reach of early autism training to parents. Autism Res Treat 2012;2012:121878 [FREE Full text] [doi: 10.1155/2012/121878] [Medline: 23227334]
58. Vismara LA, McCormick CEB, Wagner AL, Monlux K, Nadhan A, Young GS. Telehealth parent training in the early start Denver model: results from a randomized controlled study. Focus Autism Other Dev Disabl 2018;33(2):67-79. [doi: 10.1177/1088357616651064]
59. Sadeghi S, Pouretemad HR, Shalani B. Internet-based versus face-to-face intervention training for parents of young children with excessive screen-time and autism spectrum disorder-like symptoms: a comparative study. Int J Dev Disabil 2022;68(5):744-755 [FREE Full text] [doi: 10.1080/20473869.2021.1895699] [Medline: 36210895]
60. Lee K, Godina F, Pike D. A social turn-taking, parent mediated learning intervention for a young child with autism: findings of a pilot telehealth study. Early Child Educ J 2023;52(4):1-13 [FREE Full text] [doi: 10.1007/s10643-023-01467-x] [Medline: 37360607]
61. de Almeida Rodrigues L, de Melo A, de Medeiros N, Camargos AR. Family-professional collaborative intervention via telehealth with an infant with Down syndrome and visual impairment: a case report. Physiother Theory Pract 2024;40(10):2458-2467. [doi: 10.1080/09593985.2023.2244067] [Medline: 37540216]
62. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. CiteSeerX. 2006. URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ed8b23836338f6fdea0cc55e161b0fc5805f9e27 [accessed 2006-04-01]
63. McCarthy M, Leigh G, Arthur-Kelly M. Telepractice delivery of family-centred early intervention for children who are deaf or hard of hearing: a scoping review. J Telemed Telecare 2018;25(4):249-260. [doi: 10.1177/1357633x18755883]
64. Ellison KS, Guidry J, Picou P, Adenuga P, Davis TE. Telehealth and autism prior to and in the age of COVID-19: a systematic and critical review of the last decade. Clin Child Fam Psychol Rev 2021;24(3):599-630 [FREE Full text] [doi: 10.1007/s10567-021-00358-0] [Medline: 34114135]
65. Parsons D, Cordier R, Vaz S, Lee HC. Parent-mediated intervention training delivered remotely for children with autism spectrum disorder living outside of urban areas: systematic review. J Med Internet Res 2017;19(8):e198 [FREE Full text] [doi: 10.2196/jmir.6651] [Medline: 28807892]
66. Pagaki-Skaliora M, Morrow E, Theologis T. Telehealth and remote interventions for children with cerebral palsy: scoping review. JMIR Rehabil Assist Technol 2022;9(4):e36842 [FREE Full text] [doi: 10.2196/36842] [Medline: 36041012]
67. Bowe AK, Staines A, Murray DM. Below average cognitive ability—an under researched risk factor for emotional-behavioural difficulties in childhood. Int J Environ Res Public Health 2021;18(24):12923 [FREE Full text] [doi: 10.3390/ijerph182412923] [Medline: 34948532]
68. Elshani H, Dervishi E, Ibrahimi S, Nika A, Maloku Kuqi M. The impact of cognitive impairment in children with intellectual disabilities. J Int Cooperation Dev 2020;3(2):25. [doi: 10.36941/jicd-2020-0013]
69. Koshy B, Srinivasan M, Srinivasaraghavan R, Roshan R, Mohan VR, Ramanujam K, et al. Structured early childhood education exposure and childhood cognition—evidence from an Indian birth cohort. Sci Rep 2024;14(1):12951. [doi: 10.1038/s41598-024-63861-8] [Medline: 38839990]
70. Hirve R, Adams C, Kelly CB, McAullay D, Hurt L, Edmond KM, et al. Effect of early childhood development interventions delivered by healthcare providers to improve cognitive outcomes in children at 0-36 months: a systematic review and meta-analysis. Arch Dis Child 2023;108(4):247-257. [doi: 10.1136/archdischild-2022-324506] [Medline: 36732037]
71. Alonazi A. Effectiveness and acceptability of telerehabilitation in physical therapy during COVID-19 in children: findings of a systematic review. Children (Basel) 2021;8(12):1101 [FREE Full text] [doi: 10.3390/children8121101] [Medline: 34943295]

XSL•FO
RenderX

72. Farmani E, Fekar Gharamaleki F, Nazari MA. Challenges and opportunities of tele-speech therapy: before and during the COVID-19 pandemic. J Public Health Res 2024;13(1):22799036231222115 [FREE Full text] [doi: 10.1177/22799036231222115] [Medline: 38333616]
73. Buchi AB, Langlois DM, Northway R. Use of telehealth in pediatrics. Prim Care 2022;49(4):585-596. [doi: 10.1016/j.pop.2022.04.005] [Medline: 36357063]
74. Wells MJ, Dukarm P, Mills A. Telehealth in rehabilitation psychology and neuropsychology. Phys Med Rehabil Clin N Am 2021;32(2):405-418 [FREE Full text] [doi: 10.1016/j.pmr.2020.12.009] [Medline: 33814065]
75. Morgan C, Fetters L, Adde L, Badawi N, Bancale A, Boyd RN, et al. Early intervention for children aged 0 to 2 years with or at high risk of cerebral palsy: international clinical practice guideline based on systematic reviews. JAMA Pediatr 2021;175(8):846-858 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.0878] [Medline: 33999106]
76. Guralnick MJ. Applying the developmental systems approach to inclusive community-based early intervention programs: process and practice. Infants Young Child 2020;33(3):173-183 [FREE Full text] [doi: 10.1097/IYC.0000000000000167] [Medline: 32612322]
77. Blanche EI, Chang MC, Gutiérrez J, Gunter JS. Effectiveness of a sensory-enriched early intervention group program for children with developmental disabilities. Am J Occup Ther 2016;70(5):7005220010p1-7005220010p8. [doi: 10.5014/ajot.2016.018481] [Medline: 27548865]
78. Spittle A, Orton J, Anderson PJ, Boyd R, Doyle LW. Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. Cochrane Database Syst Rev 2015;2015(11):CD005495 [FREE Full text] [doi: 10.1002/14651858.CD005495.pub4] [Medline: 26597166]
79. Zwicker JG, Lee EJ. Early intervention for children with/at risk of developmental coordination disorder: a scoping review. Dev Med Child Neurol 2021;63(6):659-667 [FREE Full text] [doi: 10.1111/dmcn.14804] [Medline: 33426644]
80. Vaivada T, Gaffey MF, Bhutta ZA. Promoting early child development with interventions in health and nutrition: a systematic review. Pediatrics 2017;140(2):e20164308. [doi: 10.1542/peds.2016-4308] [Medline: 28771408]
81. Steindal SA, Nes AAG, Godskesen TE, Holmen H, Winger A, Österlind J, et al. Advantages and challenges of using telehealth for home-based palliative care: systematic mixed studies review. J Med Internet Res 2023;25:e43684 [FREE Full text] [doi: 10.2196/43684] [Medline: 36912876]
82. Bravo A, Ibañez LV, Scott S, Dick C, Carpentier P, Stone WL. Telehealth delivery in part C early intervention: provider and caregiver perspectives. J Autism Dev Disord 2023;53(12):4545-4559 [FREE Full text] [doi: 10.1007/s10803-022-05734-5] [Medline: 36153443]
83. Bourke A, O'Connor U, Boyle B, Kennedy J, Lynch H. Bringing therapy home: exploring parents' experiences of telehealth for children with developmental coordination disorder. Br J Occup Ther 2023;86(10):714-722. [doi: 10.1177/03080226231181018]
84. Cheung WC, Aleman-Tovar J, Johnston AN, Little LM, Burke MM. A qualitative study exploring parental perceptions of telehealth in early intervention. J Dev Phys Disabil 2023;35(3):353-373 [FREE Full text] [doi: 10.1007/s10882-022-09853-w] [Medline: 35789590]
85. Soltany KA, Segovia Molina R, Pappo C, Thomson S, Pring K, Cox S, et al. Interest and satisfaction of telemedicine use among ambulatory neurology patients in Western North Carolina during the COVID-19 pandemic. Telemed J E Health 2024;30(4):e1071-e1080. [doi: 10.1089/tmj.2023.0039] [Medline: 37883644]
86. Phenicie R, Acosta Wright R, Holzberg J. Patient satisfaction with telehealth during COVID-19: experience in a rural county on the United States-Mexico border. Telemed J E Health 2021;27(8):859-865. [doi: 10.1089/tmj.2021.0111] [Medline: 34096790]
87. Hinton V, Akemoğlu Y, Tomeny K, McWilliam RA. Supporting families from a distance: implementing routines-based home visits via telepractice. Early Child Educ J 2023:1-8 [FREE Full text] [doi: 10.1007/s10643-023-01464-0] [Medline: 37360610]
88. Abid MH, Abid M, Shahid R, Al Nofeye J, Ratnani I. Patient and family engagement during challenging times: What works and what does not? Cureus 2021;13(5):e14814 [FREE Full text] [doi: 10.7759/cureus.14814] [Medline: 34094768]
89. Barnett ER, Knight-Zhang EM, Haskell HG, Merriweather KA, Gaspar HA, Parton RR, et al. Family engagement in services during COVID-19: a mixed-methods study of caregiver and staff perspectives. J Pediatr Health Care 2023;37(2):142-152 [FREE Full text] [doi: 10.1016/j.pedhc.2022.10.002] [Medline: 36372630]
90. Traynor NM, Hughes M, Wood A, Onslow G, Edwards K, Clements J, et al. Supporting families with complex early parenting needs through a virtual residential parenting service: an investigation of outcomes, facilitators and barriers. J Clin Nurs 2024;33(3):1122-1133. [doi: 10.1111/jocn.16927] [Medline: 37962242]
91. Huber E, Ferjan Ramírez N, Corrigan NM, Kuhl PK. Parent coaching from 6 to 18 months improves child language outcomes through 30 months of age. Dev Sci 2023;26(6):e13391. [doi: 10.1111/desc.13391] [Medline: 36999222]
92. Ellis PK, Scherer NJ, Lien KM. Enhanced milieu teaching with phonological emphasis: a pilot, telepractice study for parent training in early intervention. Int J Speech Lang Pathol 2024;26(2):179-193. [doi: 10.1080/17549507.2023.2199176] [Medline: 37265060]

XSL•FO
RenderX

93. Brian J, Drmic I, Roncadin C, Dowds E, Shaver C, Smith IM, et al. Effectiveness of a parent-mediated intervention for toddlers with autism spectrum disorder: evidence from a large community implementation. Autism 2022;26(7):1882-1897. [doi: 10.1177/13623613211068934] [Medline: 35037520]
94. Garcia A, Pérez‐Figueroa R, Cozart D, Cook V, Damron J, Clark S, et al. "They don't trust me": Service providers reflections of father‐centric treatment and engagement. Fam Relations 2024;73(4):2544-2563. [doi: 10.1111/fare.13038]
95. Gur A, Hindi TN. Exploring fathers' perspectives on family-centered services for families of children with disabilities. Res Dev Disabil 2022;124:104199. [doi: 10.1016/j.ridd.2022.104199] [Medline: 35176723]
96. Diniz E, Brandão T, Veríssimo M. Father involvement during early childhood: a systematic review of qualitative studies. Fam Relations 2023;72(5):2710-2730. [doi: 10.1111/fare.12858]
97. Acar S, Chen CI, Xie H. Parental involvement in developmental disabilities across three cultures: a systematic review. Res Dev Disabil 2021;110:103861. [doi: 10.1016/j.ridd.2021.103861] [Medline: 33482560]
98. Sarkadi A, Kristiansson R, Oberklaid F, Bremberg S. Fathers' involvement and children's developmental outcomes: a systematic review of longitudinal studies. Acta Paediatr 2008;97(2):153-158. [doi: 10.1111/j.1651-2227.2007.00572.x] [Medline: 18052995]
99. Yoon S, Kim M, Yang J, Lee JY, Latelle A, Wang J, et al. Patterns of father involvement and child development among families with low income. Children 2021;8(12):1164 [FREE Full text] [doi: 10.3390/children8121164] [Medline: 34943359]
100. Jansen E, Marceau K, Sellers R, Chen T, Garfield CF, Leve LD, et al. The role of fathers in child development from preconception to postnatal influences: opportunities for the National Institutes of Health Environmental Influences on Child Health Outcomes (ECHO) program. Dev Psychobiol 2024;66(2):e22451. [doi: 10.1002/dev.22451] [Medline: 38388196]
101. Beffel JH, Gerde HK, Nuttall AK. Siblings and interventions: how siblings influence development and why practitioners should consider including them in interventions. Early Childhood Educ J 2021;50(8):1439-1448. [doi: 10.1007/s10643-021-01273-3]
102. Lynam A, Smith MM. Sibling involvement in interventions for children with a disability: a systematic review. Disabil Rehabil 2022;44(17):4579-4589. [doi: 10.1080/09638288.2021.1913247] [Medline: 33878995]
103. Vilaseca R, Rivero M, Bersabé RM, Cantero M, Navarro-Pardo E, Valls-Vidal C, et al. Demographic and parental factors associated with developmental outcomes in children with intellectual disabilities. Front Psychol 2019;10:872 [FREE Full text] [doi: 10.3389/fpsyg.2019.00872] [Medline: 31068864]
104. García-Grau P, Martínez-Rico G, González-García RJ, Escorcia-Mora CT, Cañadas-Pérez M. Caregiver burden and family quality of life in early intervention: the role of mothers and family confidence. Eur J Investig Health Psychol Educ 2024;14(5):1325-1337 [FREE Full text] [doi: 10.3390/ejihpe14050087] [Medline: 38785585]
105. Brekke I, Alecu A. The health of mothers caring for a child with a disability: a longitudinal study. BMC Womens Health 2023;23(1):639 [FREE Full text] [doi: 10.1186/s12905-023-02798-y] [Medline: 38037017]
106. Lee J, Gao M, Lee C. Gendered racial disparities in health of parents with children with developmental disabilities. Front Psychol 2022;13:926655 [FREE Full text] [doi: 10.3389/fpsyg.2022.926655] [Medline: 36118507]
107. Marquis SM, McGrail K, Hayes MV. Mental health outcomes among parents of a child who has a developmental disability: comparing different types of developmental disability. Disabil Health J 2020;13(2):100874. [doi: 10.1016/j.dhjo.2019.100874] [Medline: 31796336]
108. Douglas N, Mays N, Al-Haboubi M, Manacorda T, Thana L, Wistow G, et al. Observations of community-based multidisciplinary team meetings in health and social care for older people with long term conditions in England. BMC Health Serv Res 2022;22(1):758 [FREE Full text] [doi: 10.1186/s12913-022-07971-x] [Medline: 35676685]
109. Heinzmann J, Simonson A, Kenyon D. A transdisciplinary approach is essential to community-based research with American Indian populations. Am Indian Alsk Native Ment Health Res 2019;26(2):15-41 [FREE Full text] [doi: 10.5820/aian.2602.2019.15] [Medline: 31550377]
110. Durand F, Fleury MJ. A multilevel study of patient-centered care perceptions in mental health teams. BMC Health Serv Res 2021;21(1):44 [FREE Full text] [doi: 10.1186/s12913-020-06054-z] [Medline: 33413371]
111. Frye RE. A personalized multidisciplinary approach to evaluating and treating autism spectrum disorder. J Pers Med 2022;12(3):464 [FREE Full text] [doi: 10.3390/jpm12030464] [Medline: 35330464]
112. Leung C, Lai C, Lau D, Leung S, Pin TW. Effectiveness of a multidisciplinary parent training program for children with developmental disabilities: a single-blind randomized waitlist controlled trial. J Child Health Care 2020;24(4):560-576. [doi: 10.1177/1367493519880447] [Medline: 31601123]
113. Position statement on early intervention for children with developmental disabilities database on the internet. The Royal Australasian College of Physicians. 2013. URL: https://www.racp.edu.au/docs/default-source/advocacy-library/early-intervention-for-children-with-developmental-disabilities.pdf [accessed 2024-09-12]
114. Tiwari BB, Kulkarni A, Zhang H, Khan MM, Zhang DS. Utilization of telehealth services in low- and middle-income countries amid the COVID-19 pandemic: a narrative summary. Glob Health Action 2023;16(1):2179163 [FREE Full text] [doi: 10.1080/16549716.2023.2179163] [Medline: 36803538]
115. Combi C, Pozzani G, Pozzi G. Telemedicine for developing countries: a survey and some design. Appl Clin Inform 2017;7(4):1025-1050. [doi: 10.4338/aci-2016-06-r-0089]

XSL•FO
RenderX

Abbreviations

ASD: autistic spectrum disorder
CP: cerebral palsy
DS: down syndrome
FCC: family-centered care
IoT: Internet of Things
PICO: Patient, Interested Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 13.09.24; peer-reviewed by L Zhou, B Fadojutimi, N Raju; comments to author 13.11.24; revised version received 27.11.24; accepted 28.11.24; published 17.01.25.

Please cite as:
Shin Y, Park EJ, Lee A
Early Intervention for Children With Developmental Disabilities and Their Families via Telehealth: Systematic Review
J Med Internet Res 2025;27:e66442
URL: https://www.jmir.org/2025/1/e66442
doi:10.2196/66442
PMID:

©Yoomi Shin, Eun Ju Park, Anna Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A New Smartphone-Based Cognitive Screening Battery for Multiple Sclerosis (icognition): Validation Study

Stijn Denissen[1,2*], PT, PhD; Delphine Van Laethem[1,3*], MSc, MD; Johan Baijot[1], PhD; Lars Costers[1,2], PhD; Annabel Descamps[2], MSc; Ann Van Remoortel[1,4], BSc, RN; Annick Van Merhaegen-Wieleman[1,5], RN, MSc; Marie D'hooghe[4,6], MD, PhD; Miguel D'Haeseleer[4,5,6], MD, PhD; Dirk Smeets[1,2], PhD; Diana M Sima[1,2], PhD; Jeroen Van Schependom[1,7], PhD; Guy Nagels[1,5,8], MD, PhD

[1]AIMS Lab, Center for Neurosciences, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels, Belgium

[2]icometrix, Leuven, Belgium

[3]Department of Physical and Rehabilitation Medicine, Universitair Ziekenhuis Brussel, Brussels, Belgium

[4]Neurology Department, National Multiple Sclerosis Center, Melsbroek, Belgium

[5]Neurology Department, Universitair Ziekenhuis Brussel, Brussels, Belgium

[6]Center for Neurosciences, Vrije Universiteit Brussel, Brussels, Belgium

[7]Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium

[8]St Edmund Hall, University of Oxford, Oxford, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Stijn Denissen, PT, PhD
AIMS Lab
Center for Neurosciences
Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel
Laarbeeklaan 103
Brussels, 1090
Belgium
Phone: 32 2 477 41 02
Email: stijn.denissen@vub.be

Abstract

Background: Cognitive deterioration is common in multiple sclerosis (MS) and requires regular follow-up. Currently, cognitive status is measured in clinical practice using paper-and-pencil tests, which are both time-consuming and costly. Remote monitoring of cognitive status could offer a solution because previous studies on telemedicine tools have proved its feasibility and acceptance among people with MS. However, existing smartphone-based apps include designs that are prone to motor interference and focus primarily on information processing speed, although memory is also commonly affected.

Objective: This study aims to validate a smartphone-based cognitive screening battery, icognition, to detect deterioration in both memory and information processing speed.

Methods: The icognition screening battery consists of 3 tests: the Symbol Test for information processing speed, the Dot Test for visuospatial short-term memory and learning, and the visual Backward Digit Span (vBDS) for working memory. These tests are based on validated paper-and-pencil tests: the Symbol Digit Modalities Test, the 10/36 Spatial Recall Test, and the auditory Backward Digit Span, respectively. To establish the validity of icognition, 101 people with MS and 82 healthy participants completed all tests. Of the 82 healthy participants, 20 (24%) repeated testing 2 to 3 weeks later. For each icognition test, validity was established by the correlation with its paper-and-pencil equivalent (concurrent validity), the correlation and intraclass correlation coefficient (ICC) between baseline and follow-up testing (test-retest reliability), the difference between people with MS and healthy participants, and the correlation with other clinical parameters such as the Expanded Disability Status Scale.

Results: All icognition tests correlated well with their paper-and-pencil equivalents (Symbol Test: r=0.67; P<.001; Dot Test: r=0.31; P=.002; vBDS: r=0.69; P<.001), negatively correlated with the Expanded Disability Status Scale (Symbol Test: ρ=–0.34; P<.001; Dot Test: ρ=−0.32; P=.003; vBDS: ρ=−0.21; P=.04), and showed moderate test-retest reliability (Symbol Test: ICC=0.74; r=0.85; P<.001; Dot Test: ICC=0.71; r=0.74; P<.001; vBDS: ICC=0.72; r=0.83; P<.001). Test performance was comparable between people with MS and healthy participants for all cognitive tests, both in icognition (Symbol Test: U=4431; P=.42; Dot

XSL•FO
RenderX

Test: *U*=3516; *P*=.32; vBDS: *U*=3708; *P*=.27) and the gold standard paper-and-pencil tests (Symbol Digit Modalities Test: *U*=4060.5, *P*=.82; 10/36 Spatial Recall Test: *U*=3934; *P*=.74; auditory Backward Digit Span: *U*=3824.5, *P*=.37).

Conclusions: icognition is a valid tool to remotely screen cognitive performance in people with MS. It is planned to be included in a digital health platform that includes volumetric brain analysis and patient-reported outcome measures. Future research should establish the usability and psychometric properties of icognition in a remote setting.

(J Med Internet Res 2025;27:e53503) doi:10.2196/53503

KEYWORDS

multiple sclerosis; telemedicine; cognition; memory; information processing speed; mobile phone

Introduction

Background

Medicine is increasingly digitalizing, and there are compelling reasons to stimulate this trend. Clinicians can more easily access and share electronic health records, and storing data in a digital format facilitates visualization and organization in research databases, yielding new insights into pathology and disease management. Moreover, electronic health records drive artificial intelligence research [1], while artificial intelligence, in turn, further stimulates storing records digitally [2], closing a positive feedback loop. Far from being a mere "nice to have," digital medicine was crucial during the COVID-19 pandemic [3], enabling telemedicine services to be provided when social distancing was essential.

Telemedicine provides practical solutions for people with multiple sclerosis (MS), a chronic disease characterized by inflammation and degeneration of the central nervous system [4]. Telemedicine tools are well accepted by patients, and their feasibility and cost-effectiveness have been established previously [5]. Moreover, patients tend to objectively benefit from the use of these tools, which can, for example, aid in fatigue management [6] and improve cognitive function [7]. The latter is important because nearly half of the people with MS have cognitive impairment [8], which has significant repercussions on daily life activities, societal participation, employment, and susceptibility to psychiatric disorders [9].

Telemedicine could also facilitate cognitive monitoring of people with MS. First, it would allow increasing temporal resolution of cognitive trajectories because cognitive tests are usually performed during routine follow-up sessions that are months apart. This is problematic because cognitive decline can be sudden, unexpected, and severe (eg, in case of a relapse) [10,11]. Remote testing could help detect minimal changes early, enabling timely intervention. Second, it would unburden clinicians because current practice relies on paper-and-pencil tests administered under the supervision of a trained examiner.

Smartphones especially provide a window of opportunity, with an estimated 6.7 billion subscriptions worldwide (69% of the population) [12]. However, current smartphone-based cognitive assessments focus primarily on information processing speed (IPS) [13]; yet, besides slowed IPS, the hallmark cognitive problem in MS is impaired memory [14], for which the first smartphone test was only recently introduced by Podda et al [15]. Memory assessments emerged earlier on tablet devices [16] but are less suitable for consistent follow-up because tablet devices are used far less frequently than smartphones. Furthermore, smartphone tests might be prone to motor interference. They are predominantly digital versions of the Symbol Digit Modalities Test (SDMT) [17], which is a popular test in clinical practice to measure IPS and has excellent psychometric properties [18]. Digitalizing the SDMT allows randomizing its key, which could reduce practice effect as reported in the study by Pereira et al [19]. However, creating an exact digital replica of the SDMT requires patients to choose from 9 small buttons on the screen, which could cause motor interference because fine motor skills are commonly affected in MS [20].

Objectives

To tackle these limitations, in this study, we aim to validate a new smartphone-based cognitive screening battery called "icognition." It is a quick, smartphone-based screening tool for remote follow-up of the 2 most commonly impaired cognitive domains in MS: IPS and memory [14]. It is intended to be part of the recently established icompanion app, a digital diary for people with MS [21]. Regular remote screening could enable faster confirmation of cognitive deterioration by a neuropsychologist and prompt intervention by the patient's neurologist.

Methods

Study Design

This is an observational case-control study designed to validate the icognition smartphone app.

Participants

Study participants were recruited between June 17, 2021, and January 3, 2023. People with MS were enrolled from the outpatient clinics of the neurology department at Universitair Ziekenhuis Brussel (secondary care) and the National MS Center at Melsbroek (tertiary care). They were recruited by the local study nurse during their follow-up visit. Healthy control participants were recruited via leaflets and the social networks of the researchers involved. Inclusion criteria for people with MS were a confirmed diagnosis of MS according to the McDonald criteria [22]. People with MS were excluded if they had been hospitalized for reasons other than rehabilitation or if they had experienced a relapse within the past month. Both people with MS and healthy control participants were excluded if they had any other neurological or psychiatric disorder or learning disorder. A total of 101 people with MS and 82 healthy control participants (matched on age, sex, and education level)

XSL•FO
RenderX

met the inclusion and exclusion criteria for this study. All participants were either Dutch-speaking or bilingual, including Dutch, and were aged ≥18 years.

Ethical Considerations

This study was approved by the medical ethics committee of Universitair Ziekenhuis Brussel (BUN 143201940335) and the National MS Center at Melsbroek. All participants signed informed consent (in Dutch) before inclusion. Data were pseudonymized and stored in the protected OneDrive cloud service (Microsoft Corporation) of Vrije Universiteit Brussel. When publicly shared via GitHub, as explained in the Data Availability section, the data will be anonymized. This manuscript presents the primary analysis of these data. Participants did not receive compensation for taking part in the study.

icognition Screening Battery

The icognition cognitive screening battery consists of 3 tests (Figure 1).

Figure 1. Screenshots of the icognition tests (although the instructions were in Dutch during testing, they are presented here in English). (A) The Symbol Test for information processing speed (score=number of correct responses [does the symbol combination occur in the key?] in 90 s). (B) The Dot Test for visuospatial short-term learning and memory (score=total number of correctly indicated dots across 10 trials [maximum score=30]). (C) Visual Backward Digit Span (vBDS) for working memory (score=sum of span lengths of correct spans [correct inversion of shown span]).

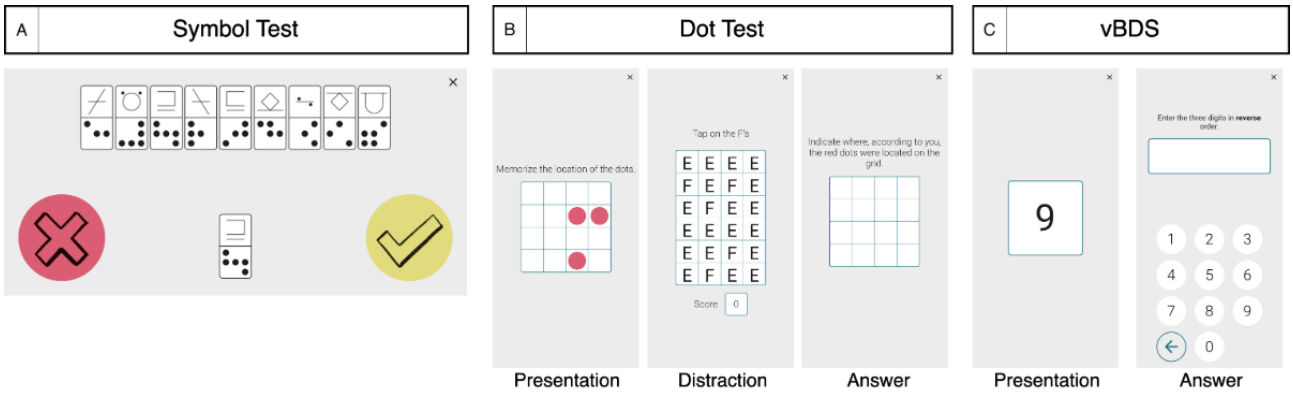

The Symbol Test is based on the computerized Digit Symbol Substitution Test presented in the study by Rypma et al [23]. In the Symbol Test, a combination of symbols is presented to the participants, one at a time. A key, consisting of 9 pairs of symbols, is displayed at the top and is shuffled for each trial. For each trial, the participant needs to indicate whether the presented combination appears in the key. The total score is the number of correct answers provided in 90 seconds. This test is designed to assess IPS.

The Dot Test consists of 3 phases. In the first phase, a participant is presented a 4×4 grid in which 3 dots are shown for 3 seconds. Next, as a distractor task, the participant is shown a 4×6 grid of "E" and "F" shapes and must identify as many "F" shapes as possible in 4 seconds. In the last phase, the participant must indicate in an empty 4×4 grid where the 3 dots of the first phase were located. The Dot Test is inspired by the Dot Memory Test presented in the study by Sliwinski et al [24], with all grids reduced in size compared to their version (5×5 grids). We also implemented a criterion for the distractor task, requiring the participant to identify at least 3 "F" shapes. If this criterion was not met, the trial was restarted. The 3 dots could not be aligned on 1 line or form an L-shape within a 2×2 block of cells. The total score is the number of correctly indicated dots across 10 trials. This test is designed to assess visuospatial short-term memory and learning.

In the visual Backward Digit Span (vBDS), a series of digits is presented on the screen one by one, each for 1 second, as described in the study by Hilbert et al [25]. The participant must then list the digits in reverse order. Spans were randomly generated with digits between 0 and 9, with the following constraints: a digit can appear only once in the span, and a chain of ≥3 digits cannot have a fixed increment or decrement of 1 or 2 digits, in accordance with Woods et al [26]. Scoring is based on the sum of all correct span lengths; for example, if a participant correctly recalls 2 spans of length 3 and 1 span of length 4, the total score is 10. This test is designed to assess working memory.

All tests were performed on a Samsung Galaxy A10 smartphone (6.2-inch screen size) and were supervised by a test examiner. Each test was directly preceded by a practice phase to familiarize the participants with the respective test. This phase consisted of 5 trials for the Symbol Test, 3 for the Dot Test, and 4 (2 of length 3 and 2 of length 4) for the vBDS. The Symbol Test is performed with the smartphone in landscape orientation, whereas for the Dot Test and vBDS, the smartphone must be in portrait orientation. In the design of icognition, careful consideration was given to the potential biasing influence of fine motor impairment in MS [20]. Motor interference was minimized by using 2 large buttons for the Symbol Test (in contrast to digital SDMT variants where participants must choose from 9 smaller buttons [27]) and not placing any restrictions on the response time in the other icognition tests.

Validation Procedure

The procedure to validate icognition is based on the study by Benedict et al [28] and involves assessing 4 criteria, as outlined in the following subsections.

Concurrent Validity

We assessed how well each icognition test correlates with its paper-and-pencil equivalent. For the Symbol Test, the equivalent

XSL•FO
RenderX

was the SDMT [17]. In the SDMT, a sheet is presented to the participant with a key of 9 symbol-digit pairs at the top and a list of symbols without corresponding digits. In 90 seconds, the participant must convert as many symbols to digits as possible from the list, reading them out loud to the examiner, using the key. The Dot Test is based on the 10/36 Spatial Recall Test (SPART) [29]. In the SPART, the participant is shown a 6×6 grid with a pattern of 10 dots for 10 seconds. Subsequently, the grid is removed, and the participant is asked to replicate the pattern using 10 checkers. This process is repeated 3 times. The final score is the total number of correctly placed checkers across all trials. Finally, for the vBDS, a modified version of the Wechsler Adult Intelligence Scale, Fourth Edition, auditory Backward Digit Span [30] was used. In the auditory Backward Digit Span, digit spans are read out loud to the participant, who is asked to repeat them in reverse order. The original test consists of 2 trials for each span length, starting with a span of 2 digits and increasing by 1 digit each time the participant correctly completes at least 1 of the 2 trials. As discussed in Woods et al [31], in the original Wechsler Adult Intelligence Scale, Fourth Edition, design, participants with the same score can have a different number of correct spans. To mitigate this, we used a fixed number of spans, ranging from 3 to 7 digits in length (Table S1 in Multimedia Appendix 1). The complete list was always administered for each participant. The scoring metric is the same as described earlier for the icognition Backward Digit Span.

Test-Retest Reliability

Benedict et al [28] mention that test-retest reliability should be assessed on a "small sample" of either people with MS or healthy controls [28]. We aimed to retest the healthy controls 2 to 3 weeks after baseline testing. We used the intraclass correlation coefficient (ICC) to assess agreement between baseline and retesting, which is explained in more detail in the Statistical Analyses subsection.

Comparison of Performance

For each test, we compared the performance of people with MS to that of age-, sex-, and education level–matched healthy controls using the Mann-Whitney *U* test. The analysis was then repeated after correcting test performance for age, sex, and education level (Figure S1 in Multimedia Appendix 1). Details of the correction methodology are presented in the Statistical Analyses subsection.

Assessment of Correlations

Finally, we assessed the correlation of each icognition test with the Expanded Disability Status Scale (EDSS) [32], disease duration, the Beck Depression Inventory (BDI) [33], the Fatigue Scale for Motor and Cognitive Functions (FSMC) [34], education level, and age.

Data Curation

Data were entered independently by 2 researchers (SD and DVL). Conflicts in data entry were resolved through mutual discussion.

Statistical Analyses

We used an α level of .05 for all analyses. Participants with missing data on a certain test were only excluded for that specific test. We used the Spearman correlation for nonlinear and categorical variables (EDSS, BDI, FSMC, and education level), whereas the Pearson correlation was used otherwise. Mann-Whitney *U* tests were used for between-group distribution comparisons.

To assess test-retest reliability, following the approach of van Oirschot et al [35] and the guidelines by Koo and Li [36], we used the following ICC type: "two-way mixed effects, absolute agreement, single rater/measurement" (ICC[A,1]). In this study, the smartphone app served as the sole "rater," which is why we used a 2-way mixed effects model, appropriate when the "selected raters are the only raters of interest" [36]. For the same reason, the "type" of ICC was "single rater/measurement," while for "definition," we used "absolute agreement," reflecting the extent to which 1 rater's score (in our case, baseline testing) equals the other rater's score (in our case, retesting). The decision process is illustrated in the decision flowchart in the study by Koo and Li [36]. To interpret the magnitude of the test-retest reliability, we used the guidelines by Koo and Li [36]: poor (<0.50), moderate (0.50-0.75), good (0.75-0.90), and excellent (>0.90).

The magnitude of correlations was interpreted using the classification by Portney and Watkins [37]: small (<0.25), fair (0.25-0.50), moderate to good (0.50-0.75), and excellent (>0.75).

Finally, for each cognitive test, we fitted a regression equation with age, sex, and education level as independent variables and test performance as dependent variable on the healthy control data. This allows calculation of the expected score of a participant, given their age, sex, and education level, and subsequently comparison with the actual score of the participant, resulting in a *z* score. The technical details of the procedure, including the required values for performing *z* normalization on other data, are presented in Multimedia Appendix 1, while Table S2 in Multimedia Appendix 1 presents the values necessary to perform the procedure. Test performance was considered impaired if the *z* score was ≤−1.5, based on the study by Benedict et al [38].

Results

Overview

Participant characteristics are presented in Table 1.

XSL•FO
RenderX

Table 1. Participant characteristics (n=183)[a].

	People with MS[b] (n=101)	Healthy controls (n=82)	*P* value
Demographics			
Age (y), mean (SD)	45.4 (10.0)	46.8 (14.7)	.59[c]
Sex, n (%)			.74[d]
Female	63 (62.4)	54 (65.9)	
Male	38 (37.6)	28 (34.1)	
Education, median (IQR)	15 (12-17)	15 (13-17)	.74[c]
MS specific			
Disease duration (y), mean (SD)	11.7 (7.4)	—[e]	—
MS type, n (%)		—	—
RRMS[f]	86 (85.1)		
SPMS[g]	8 (7.9)		
PPMS[h]	7 (6.9)		
EDSS[i] , median (IQR)	3 (2-4)	—	—
Paper-and-pencil tests			
SDMT[j] , mean (SD)	58.5 (10.0)	58.3 (9.9)	.82[c]
Those considered impaired, n (%)	7 (6.9)	6 (7.3)	.99[d]
10/36 SPART[k] , mean (SD)	20.6 (4.3)	20.1 (4.6)	.74[c]
Those considered impaired, n (%)	8 (7.9)	4 (4.9)	.60[d]
Auditory BDS[l] , mean (SD)	48.4 (18.7)	46.0 (17.6)	.37[c]
Those considered impaired, n (%)	9 (8.9)	6 (7.3)	.90[d]
BDI[m] , mean (SD)	9.3 (5.7)	5.8 (5.1)	<.001[c]
FSMC[n] , mean (SD)	58.7 (16.4)	39.9 (12.1)	<.001[c]
icognition tests			
Symbol Test, mean (SD)	24.8 (6.3)	25.4 (6.4)	.42[c]
Number impaired, n (%)	13 (12.9)	7 (8.5)	.49[d]
Dot Test, mean (SD)	21.8 (5.1)	21.2 (4.9)	.32[c]
Number impaired, n (%)	9 (8.9)	7 (8.5)	.99[d]
Visual BDS, mean (SD)	46.9 (16.9)	43.8 (16.7)	.27[c]
Number impaired, n (%)	8 (7.9)	9 (11)	.65[d]

[a]The following variables had missing values: Expanded Disability Status Scale (n=7), 10/36 Spatial Recall Test (people with multiple sclerosis: n=1; healthy controls: n=1), Beck Depression Inventory (people with multiple sclerosis: n=2; healthy controls: n=1), Fatigue Scale for Motor and Cognitive Functions (people with multiple sclerosis: n=2; healthy controls: n=7), Dot Test (people with multiple sclerosis: n=3; healthy controls: n=1), and visual Backward Digit Span (people with multiple sclerosis: n=1). An additional 3 participants were excluded from the Dot Test analyses because they had been tested with an earlier version of icognition where a larger grid size was used. We reduced the grid size after these 3 participants were tested because we deemed this test to be too difficult. However, these 3 participants were included when the mean (SD) and number of participants considered impaired was calculated.

[b]MS: multiple sclerosis.

[c]Mann-Whitney *U* test.

[d]Chi-square test.

[e]Not applicable.

[f]RRMS: relapsing-remitting multiple sclerosis.

XSL•FO
RenderX

[g]SPMS: secondary progressive multiple sclerosis.
[h]PPMS: primary progressive multiple sclerosis.
[i]EDSS: Expanded Disability Status Scale
[j]SDMT: Symbol Digit Modalities Test.
[k]SPART: 10/36 Spatial Recall Test.
[l]BDS: Backward Digit Span
[m]BDI: Beck Depression Inventory.
[n]FSMC: Fatigue Scale for Motor and Cognitive Functions.

Concurrent Validity

Figure 2 shows the scatterplot of each icognition test with its paper-and-pencil equivalent. The Symbol Test showed a significant moderate to good correlation with SDMT performance (healthy controls: r=0.68; P<.001; people with MS: r=0.67; P<.001). There was also a significant fair correlation between the Dot Test and the SPART (healthy controls: r=0.30; P=.007; people with MS: r=0.31; P=.002) and a moderate to good correlation between the vBDS and its auditory equivalent (healthy controls: r=0.69; P<.001; people with MS: r=0.69; P<.001).

Figure 2. Concurrent validity. Scatterplots comparing the scores of each icognition test (x-axis) with those of its corresponding paper-and-pencil equivalent (y-axis). Correlations for both people with multiple sclerosis (MS) and healthy controls were (A) moderate to good, (B) fair, and (C) moderate to good. aBDS: auditory backward digit span; SDMT: Symbol Digit Modalities Test; SPART: 10/36 Spatial Recall Test; vBDS: visual Backward Digit Span.

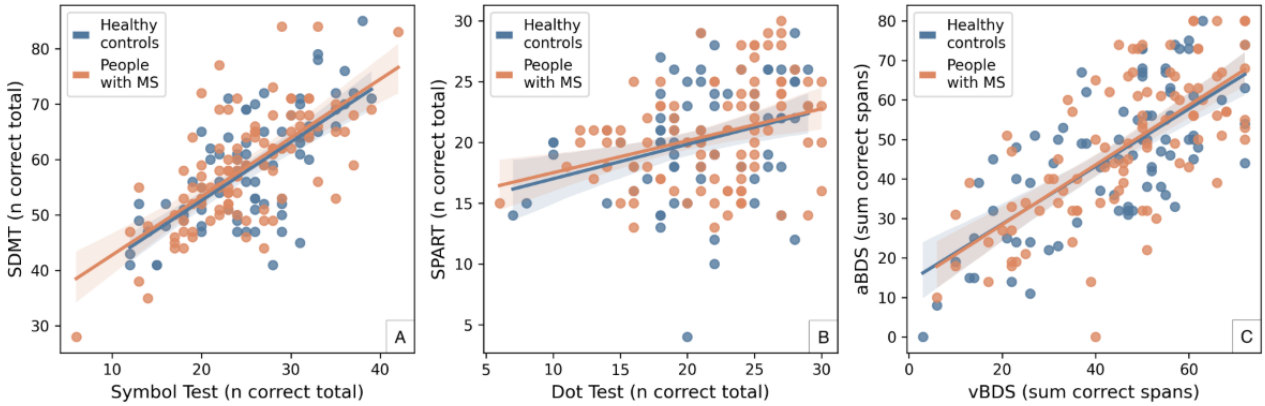

Test-Retest Reliability

In total, 20 healthy controls were retested with an average intertest interval of 18 (SD 3, range 14-23) days after initial testing to establish test-retest reliability. For the Dot Test, there was 1 missing value at baseline testing (n=19). Test-retest reliability (Figure 3) was moderate for the Symbol Test (ICC=0.74, r=0.85; P<.001), moderate for the Dot Test (ICC=0.71, r=0.74; P<.001), and moderate for the vBDS (ICC=0.72, r=0.83; P<.001).

Figure 3. Test-retest reliability. Scatterplots comparing the scores of each icognition test at baseline (x-axis) with the retest scores an average of 18 (SD 3) days later (y-axis). All tests demonstrated moderate test-retest reliability. vBDS: visual Backward Digit Span.

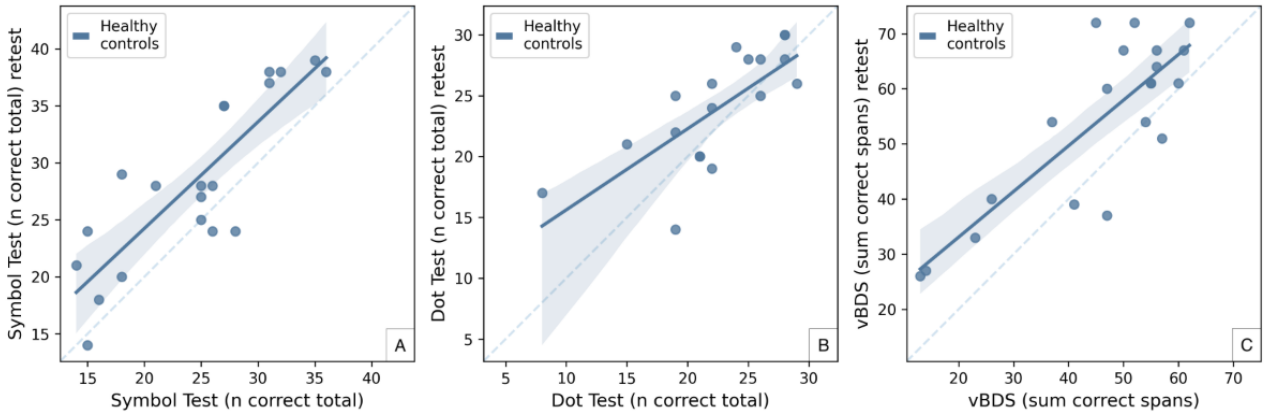

XSL•FO
RenderX

Difference Between People With MS and Healthy Controls

For all icognition tests, there was no significant difference in performance between healthy controls and people with MS (Symbol Test: U=4431; P=.42; Dot Test: U=3516; P=.32; vBDS: U=3708; P=.27; Figure 4). The results were similar for all paper-and-pencil tests (SDMT: U=4060.5; P=.82; SPART: U=3934; P=.74; auditory Backward Digit Span: U=3824.5; P=.37; Figure S2 in Multimedia Appendix 1).

Figure 4. Comparison of the performance of healthy controls and people with multiple sclerosis (MS) on the icognition tests. Test performance was not significantly different between people with MS and healthy controls for all icognition tests. vBDS: visual Backward Digit Span.

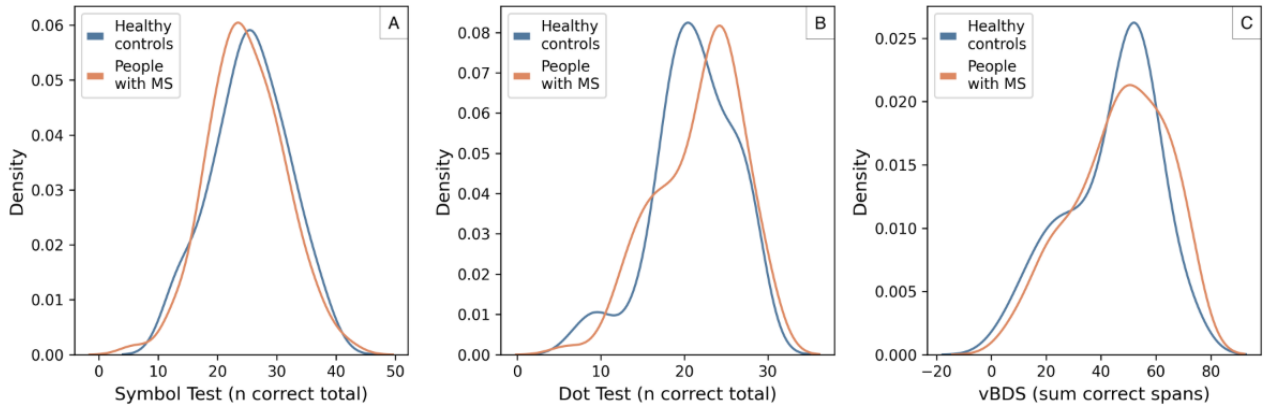

Correlations With Clinical Parameters

A correlation matrix between the icognition tests and different clinical tests is presented in Table 2. In general, higher scores on the icognition tests were associated with less physical disability (measured using the EDSS), younger age, and a higher education level. Furthermore, all tests except the Symbol Test showed a correlation with disease duration, whereas fatigue (measured using the FSMC) was negatively associated with Dot Test performance. No correlation was found between any icognition test and depression (measured using the BDI). Correlation magnitudes were small or fair, with the exception of the moderate to good correlation between the Symbol Test and age ($|r|$=0.52).

Table 2. Correlation matrix of icognition tests with clinical variables. The Spearman correlation was used for the Beck Depression Inventory (BDI), Fatigue Scale for Motor and Cognitive Functions (FSMC), Expanded Disability Status Scale (EDSS), and education level, while the Pearson correlation was used for age and disease duration.

	Symbol Test		Dot Test		vBDS[a]	
	Correlation	*P* value	Correlation	*P* value	Correlation	*P* value
Age	−0.52	<.001	−0.29	.004	−0.27	.008
Education level	0.20	.04	0.37	<.001	0.22	.03
BDI	−0.03	.78	−0.12	.27	0.03	.80
FSMC	−0.10	.31	−0.26	.01	−0.03	.79
EDSS	−0.34	<.001	−0.32	.003	−0.21	.04
Disease duration	−0.17	.08	−0.28	.006	−0.31	.002

[a]vBDS: visual Backward Digit Span.

Discussion

Overview

In this paper, we present the results of the validation process of a smartphone-based screening battery for cognitive problems in people with MS. All tests correlated with their paper-and-pencil equivalents (concurrent validity), although correlation between the Dot Test and the SPART was only fair, most likely due to the fact that the SPART—the gold standard—lacks a distractor phase. The distractor phase was initially added to avoid ceiling effects and increase sensitivity in detecting cognitive decline. However, other test differences between the Dot Test and the SPART, such as grid size and the number of trials, could also cause the correlation to be weaker. Furthermore, all tests correlated with various clinical variables (age, education level, and EDSS) and showed moderate test-retest reliability. These findings indicate the suitability of the battery for routine remote screening of cognitive problems, a condition that many people with MS are likely to develop over time [8].

State of the Art

We performed a systematic search (Multimedia Appendix 1) to identify smartphone-based cognitive tests for MS, which are summarized in Table 3. In brief, these apps began to emerge in 2020 and predominantly use a typical SDMT design. In this format, the mobile phone is held in landscape orientation, and

XSL•FO
RenderX

a key of 9 symbol-digit pairs is displayed at the top. For 90 seconds, 1 symbol at a time is shown in the center of the screen, and the participant selects as quickly as possible the matching digit from 9 digit buttons located at the bottom of the screen. These tests are primarily designed to measure IPS, although they are also used to measure cognitive fatigability [39,40]. Other tests designed to measure specific cognitive domains include the smartphone version of the Trail Making Test Part B [41] for executive function and the Auditory Test of Processing Speed for IPS [42]. Furthermore, smartphone keystroke analysis [43,44] and various cognitive training games [45] were not designed for a specific cognitive domain, but their relationships with multiple cognitive domains was assessed. Recently, the first smartphone-based cognitive screening battery (DIGICOG-MS [15]) was introduced, marking an important next step in smartphone-based cognitive assessment.

XSL•FO
RenderX

Table 3. Summary of the state of the art in smartphone-based cognitive tests for people with multiple sclerosis (MS). The "Target domain" column indicates the specific cognitive domain assessed by each cognitive test.

Name	Summary	Target domain
MCT[a] [46]	The MCT, part of MSCopilot, is a digital version of the SDMT[b]. In the MCT, a key displaying 9 symbol-digit pairs is presented at the top of the screen, with an exit button located in the top right corner. One symbol at a time is presented below the key. No details pertaining to the length of the test or scoring method are provided in the paper.	Information processing speed
sSDMT[c] [35]	The sSDMT, part of MS Sherpa, is a digital version of the SDMT. In the sSDMT, a symbol-digit key (with 9 pairs) is presented at the top of the screen. A single symbol is then presented in the center of the screen, and participants must quickly tap the correct matching digit from 9 digit buttons presented at the bottom. The total score is the number of correct digits selected in 90 s. A timer is displayed, and the key is randomized at each trial.	Information processing speed
Voice-controlled DSST[d] [47]	The voice-controlled DSST is part of the elevateMS app. Although limited details are provided, the app seems to be a variant of the SDMT, in which responses are collected via microphone.	Information processing speed and working memory
sSDMT [27]	The sSDMT, part of the Neurological Functional Test Suite, resembles the sSDMT (described previously). However, 3 differences are noteworthy. First, no timer is provided. Second, the test duration is 75 s; a 90-s performance score is calculated by multiplying the result by 90/75. Finally, the app also supports vocalized responses using a microphone and voice recognition to test participants with severe motor impairments.	Information processing speed
Mobile Trail Making Test Part B [41]	The app displays 13 circles, 6 containing a letter and 7 containing a number. The goal of the test is to connect the circles as quickly as possible in order, alternating between numbers and letters (1-A-2-B and so on, up to 7). The smartphone test was introduced by Ross et al [48] and assessed in people with MS in the study by Chen et al [41].	Executive function
Neurokeys [43]	Neurokeys is an alternative keyboard that records typing events (press and release) and subsequently calculates a range of features such as press duration and time interval between 2 presses.	No specific cognitive domain
BiAffect [44]	Similar to Neurokeys, the BiAffect app features a custom keyboard that temporarily replaces the smartphone's default keyboard and is used to collect keystrokes and subsequently calculate various features.	No specific cognitive domain
cFAST[e] [40]	The cFAST is similar to the sSDMT but includes an exit button and a progress bar to indicate the remaining response time. Unlike the 90-s duration of the sSDMT, the cFAST test lasts 5 minutes because the goal is to assess cognitive fatigability rather than information processing speed.	Cognitive fatigability
eSDMT[f] [49]	The eSDMT, part of Floodlight, is an electronic version of the SDMT. Although few details are provided, the app seems to use a similar design to the sSDMT [39,50]. However, no timer or randomization of the key is mentioned.	Information processing speed
Cognitive games in the dreaMS app [45]	Several cognitive games were included in the dreaMS app featured in the study by Pless et al [45].	No specific cognitive domain
Auditory Test of Processing Speed [42]	Participants are presented auditory digits between 1 and 99 across 3 trials (20 digits each), with each trial involving different questions for the participants to answer (ie, trial 1: is the digit greater than 50? Trial 2: is the digit greater than 50 or odd? Trial 3: is the digit greater than 50 and odd?). The test examiner records the participants' responses, and response time is calibrated before testing.	Information processing speed
MS Care Connect [51]	The MS Care Connect app is freely available for remote health assessments, including cognitive tests, in people with MS. Although limited information is available about the cognitive tests, a scoping review by Michaud et al [52] mentions it contains an SDMT-like test. According to a video on the app's website, the test is administered in portrait orientation, with the key (9 symbol-digit combinations) at the top of the screen, a symbol presentation field below it, and a 3×3 answer keypad at the bottom [53].	Information processing speed
Konectom [54]	The test is designed to measure cognitive processing speed, with the smartphone held in portrait orientation. A symbol-digit key (with 9 pairs) is displayed at the top of the screen and remains fixed for the 90-s test duration. Symbols are presented in the center of the screen, and a 3×3 answer keyboard (digits 1-9) is located at the bottom. To correct for visuomotor interference, a separate test is included in which participants are asked to tap, in the same 3×3 answer keyboard and for 20 s, as many numbers as possible that are consecutively displayed in the center of the screen.	Information processing speed
DIGICOG-MS [15]	This is a smartphone-based cognitive battery with tests for memory, semantic fluency, and information processing speed. All tests seem to be designed for use in portrait orientation. A symbol-digit key (with 9 pairs) is presented at the top of the screen and remains fixed for the 90-s test duration. Four symbols are shown at a time in the center of the screen, and a 3×3 answer keyboard (digits 1-9) is located at the bottom.	Visuospatial memory, verbal memory, semantic fluency, and information processing speed

[a]MCT: Mobile Cognition Test.

[b]SDMT: Symbol Digit Modalities Test.

XSL•FO
RenderX

[c]sSDMT: smartphone Symbol Digit Modalities Test.

[d]DSST: Digit Symbol Substitution Test.

[e]cFAST: Cognitive Fatigability Assessment Test.

[f]eSDMT: electronic Symbol Digit Modalities Test.

The icognition screening battery differs from the state of the art in 2 ways. First, fine motor and visual problems are frequently reported in people with MS [55,56]. These symptoms are likely to interfere with time-based digital processing speed tests and could be affected by design considerations, as seen in the common digital SDMT design. This design includes multiple answer buttons (typically 9), and if configured in portrait orientation may result in a small symbol-digit key, especially on smaller devices. Although the problem is recognized by Woelfle et al [50] and the Scaramozza et al [54], the latter implementing a separate test to quantify motor interference, we opted to minimize motor and visual interference through design. Specifically, we used a simplistic design in landscape orientation, featuring 2 large answer buttons (Figure 1). Second, rather than a single test, icognition features a battery of tests designed to capture a broader cognitive profile because cognitive problems in people with MS are not limited to slowed IPS [14]. To the best of our knowledge, only Podda et al [15] have recently published a smartphone-based cognitive battery for people with MS. Moreover, icognition will be included in a digital care platform assessing symptoms, imaging biomarkers, and patient-reported outcomes [21] to provide a holistic picture of a patient's well-being.

A Cognitively Preserved Sample of People With MS

People with MS scored equally well on all tests compared to healthy controls. However, as can be observed in Table 1 and Figure S2 in Multimedia Appendix 1, this was also the case for the paper-and-pencil tests. Indeed, we seem to have included a sample of people with MS with relatively preserved cognition. We tested this in a post hoc analysis using the ANOVA method described by Anders [57], comparing the SDMT performance scores of our sample of people with MS (mean 58.5, SD 10.0; n=101; Table 1) with those reported in previous studies. López-Góngora et al [58] report an average SDMT performance score of 54.3 (SD 13.4; N=237), while Sousa et al [59] mention an average SDMT performance score of 53.51 (SD 11.76; n=115). We found that the people with MS included in this study performed significantly better (comparison with López-Góngora et al [58]: $F_{1,336}$=8.01; P=.005; comparison with Sousa et al [59]: $F_{1,214}$=11.1; P=.001).

The underlying reason for having selected a cognitively preserved sample is most likely the sampling bias resulting from the limited inclusion and exclusion criteria and the recruitment of patients during outpatient consultation in secondary and tertiary MS care settings. In these settings, participants are tested cognitively at least annually, which might enhance their familiarity with typical cognitive tests for MS, with a carryover effect to derivative tests, such as those in icognition. Moreover, people with MS who were able and willing to participate might (1) have better cognitive abilities, (2) be better at handling a smartphone, and (3) be familiar with scientific studies. The reason for being less strict on inclusion was to be able to offer the screening battery to any patient with MS who is regularly followed up in an outpatient setting. Future studies should confirm the validity of icognition in people with MS who are cognitively impaired. Moreover, combined with the publicly available data underlying this study, the sensitivity of icognition in detecting cognitive impairment could be established.

To investigate the impact of cognitive impairment on our results, we performed a post hoc analysis excluding participants who were considered impaired on at least 1 paper-and-pencil test (Figures S3-S5 and Table S3 in Multimedia Appendix 1). Compared to the original analysis, the most notable differences were observed in the correlations with clinical parameters, where the correlation between the Symbol Test and disease duration became significant (P=.04), while significance dropped for the relationship between the vBDS and education level (P=.06), the Dot Test and FSMC (P=.18), and the vBDS and EDSS (P=.20). Results for concurrent validity, test-retest reliability, and the difference between people with MS and healthy controls were comparable.

The Benefits of Regular Digital Follow-Up

Proper and regular follow-up of cognitive function is important to capture fluctuations in cognitive state, such as those caused by a disease exacerbation or a relapse. Although Giedraitiene et al [60] show that a short cognitive screening tool such as the Brief International Cognitive Assessment for Multiple Sclerosis [18] can detect these cognitive fluctuations [60], the cognitive dip is likely missed in reality because cognitive assessments are usually performed only once or twice a year during clinical follow-up visits. The icognition screening battery allows for more frequent cognitive assessments. Furthermore, testing can be performed wherever and whenever a patient feels ready for it, reducing biasing effects such as fatigue [61]. Moreover, digitalization facilitates data collection and allows the extraction of more information from a cognitive test. As icognition associates every response with a 13-digit time stamp (with millisecond precision), cognitive fatigue or other performance metrics such as the "maximum gap time between correct responses" [39] can be tracked with high temporal resolution. For the Symbol Test, analogous examples are provided in the studies by Ganzetti et al [39] and Barrios et al [40].

Implementing icognition in a health care platform that includes fine motor assessment [62], symptom logging, and patient-reported outcome measures [21] enables patients to take an active role in their disease management; they can provide the individual information that complements the professional knowledge of the treating physician [63]. Moreover, collecting confounding variables, such as fatigue [21], at cognitive assessment improves the interpretation of cognitive scores.

Limitations and Future Work

This study has some limitations. This validation study used a cross-sectional design, aside from the assessment of test-retest reliability. Therefore, we were unable to map the cognitive

XSL•FO
RenderX

evolution of people with MS over time. These and other validated digital cognitive tests could considerably facilitate longitudinal testing in future studies, analogous to the 1-year follow-up study by Lam et al [64].

Popular screening batteries such as the Brief International Cognitive Assessment for Multiple Sclerosis also include a measure of verbal memory, which is indeed commonly impaired in MS [14]. However, we chose not to implement a digital test of verbal memory in icognition because such tests usually rely heavily on language, which complicates international use with the need for language-specific validation. Moreover, this introduces new technological difficulties, such as variations in microphone and speaker quality or the accuracy of speech recognition software. Although the aim of this study was to create a quick, home-based cognitive screening assessment, we acknowledge the importance of more in-depth multidomain cognitive assessment, including the assessment of verbal memory.

In the initial version of the Dot Test, the final 5 trials included remembering the position of the 3 dots on a 5×5 grid, but, after testing 3 participants with MS, we decided to consistently use a 4×4 grid, given the difficulty of the task, and excluded these participants from the Dot Test analyses. However, including these 3 participants in a post hoc analysis yielded similar results.

Furthermore, we tested consistently on the same Android smartphone (Samsung Galaxy A10) to avoid a biasing influence of model-specific factors, such as screen size, weight, and responsiveness, because these factors might impact performance and user experience. Although we expect our results to be applicable to other smartphone models, this limits the study findings to this smartphone model. However, as icognition has been designed using the Flutter framework (Google LLC), it can also be deployed on operating systems other than Android. An external validation study is planned to show the robustness of the app on different devices in a home setting.

To test concurrent validity, we used the SPART as the ground truth test for visuospatial memory because this closely resembled the Dot Memory Test presented in the study by Sliwinski et al [24]. Although recent evidence suggests that the SPART is among the most sensitive memory assessments in MS [65], Strober et al [66] found the Brief Visuospatial Memory Test–Revised to be more sensitive than the SPART. One might argue that the sole criterion of the ground truth test is that it measures visuospatial memory, regardless of its similarity to the digital test. In this context, using the Brief Visuospatial Memory Test–Revised would have been justified. Although the concurrent validity of the Dot Test was only fair, it demonstrated an acceptable test-retest reliability and correlated with several clinical variables, underscoring its value for clinical practice. Unlike the Symbol Test, which can be seen as a digital alternative to its paper-and-pencil reference test (SDMT), more research is necessary to establish the ground truth reference for the Dot Test. Furthermore, future research could subject icognition to a real-world (referred to later in this subsection), longitudinal study to establish its sensitivity to change at the individual level.

Test-retest reliability was assessed only in the healthy controls as per the recommendations of Benedict et al [28], who stated that test-retest reliability “can be investigated in a small sample of MS and/or healthy volunteers over 1-3 weeks.” Although we deem this to be a good indication for test-retest reliability in people with MS, substantiated by similar ICC reliability scores in the smartphone SDMT study by van Oirschot et al [35], test-retest reliability is not guaranteed to be similar for the cognitive tests in icognition. As we strove to perform test-retest reliability in similar circumstances (eg, location), assessing it for people with MS was difficult because they were tested during a 1-time visit to the outpatient MS clinics for consultation. Furthermore, we acknowledge that, although we followed the recommendations of Benedict et al [28], the sample sizes for testing test-retest reliability were small (Symbol Test and vBDS: n=20; Dot Test: n=19).

By recruiting a cognitively preserved sample of people with MS, this study was unfortunately unable to provide a realistic assessment of the sensitivity of the app to discriminate between (1) people with MS and healthy controls and (2) people with MS who are cognitively impaired and people with MS who are cognitively preserved. As the goal of icognition is not the diagnosis of MS but to screen for people with MS who are cognitively impaired, we especially recognize the latter as a limitation of this study. Indeed, of the 101 participants with MS, only 7 (6.9%), 8 (7.9%), and 9 (8.9%) presented impairment on the Symbol Test, Dot Test, and vBDS, respectively, resulting in a substantial class imbalance with respect to preserved test performance. However, because we make our data publicly available, this analysis can be conducted by future studies evaluating icognition on a sample of people with MS who are cognitively impaired.

This study validated the icognition app in controlled laboratory conditions under the supervision of a test examiner. However, we acknowledge that future studies should confirm psychometric properties such as test-retest reliability in remote, unsupervised settings. Furthermore, as recently demonstrated by Podda et al [15], the usability of cognitive apps should be assessed to assure a smooth transition to a real-life setting. A usability study is planned for icognition as well.

Conclusions

This study established the reliability and validity of a newly developed smartphone app, icognition, for remotely screening cognitive impairment in terms of IPS and memory in people with MS. This allows for regular screening of cognitive performance to more quickly detect and respond to potential deterioration.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Steve De Backer for his efforts in the development of the icognition app, as well as Eva Keytsman and Florine Wöhler for their support in recruiting and testing participants for this study. SD was funded by a Baekeland grant from Flanders Innovation and Entrepreneurship (HBC.2019.2579), a grant from Vrije Universiteit Brussel (SRP85), and a junior postdoctoral grant from Fonds Wetenschappelijk Onderzoek–Flanders (12A6U25N). Both DVL (1SD5322N) and GN (1805620N) are supported by a personal grant from Fonds Wetenschappelijk Onderzoek–Flanders. This study is partly funded by the CLAIMS (Clinical Impact Through AI-Assisted MS Care) project, supported by the Innovative Health Initiative joint undertaking (101112153). The joint undertaking is supported by the European Union's Horizon Europe research and innovation program as well as COCIR (European Coordination Committee of the Radiological, Electromedical, and Healthcare Information Technology Industry), the European Federation of Pharmaceutical Industries and Associations, EuropaBio, MedTech Europe, Vaccines Europe, AB Science SA, and icometrix NV.

Data Availability

The anonymized source data on which this paper relies, as well as the code used for statistical analysis and the creation of figures, are available in the GitHub repository of our laboratory [67].

Conflicts of Interest

This work was part of the industrial PhD trajectory of SD in collaboration with icometrix. LC, AD, DS, and DMS are employees of icometrix. GN and DS are shareholders of icometrix. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Supplementary text, tables, and figures.
[DOCX File , 791 KB - jmir_v27i1e53503_app1.docx]

References

1. Lee S, Kim HS. Prospect of artificial intelligence based on electronic medical record. J Lipid Atheroscler 2021 Sep;10(3):282-290 [FREE Full text] [doi: 10.12997/jla.2021.10.3.282] [Medline: 34621699]
2. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. Sci Rep 2022 Oct 26;12(1):17981 [FREE Full text] [doi: 10.1038/s41598-022-22514-4] [Medline: 36289266]
3. Getachew E, Adebeta T, Muzazu SG, Charlie L, Said B, Tesfahunei HA, et al. Digital health in the era of COVID-19: reshaping the next generation of healthcare. Front Public Health 2023;11:942703 [FREE Full text] [doi: 10.3389/fpubh.2023.942703] [Medline: 36875401]
4. Jakimovski D, Bittner S, Zivadinov R, Morrow SA, Benedict RH, Zipp F, et al. Multiple sclerosis. Lancet 2024 Jan 13;403(10422):183-202. [doi: 10.1016/S0140-6736(23)01473-3] [Medline: 37949093]
5. Yeroushalmi S, Maloni H, Costello K, Wallin MT. Telemedicine and multiple sclerosis: a comprehensive literature review. J Telemed Telecare 2020;26(7-8):400-413. [doi: 10.1177/1357633X19840097] [Medline: 31042118]
6. Khan F, Amatya B, Kesselring J, Galea M. Telerehabilitation for persons with multiple sclerosis. Cochrane Database Syst Rev 2015 Apr 09;2015(4):CD010508 [FREE Full text] [doi: 10.1002/14651858.CD010508.pub2] [Medline: 25854331]
7. Charvet LE, Yang J, Shaw MT, Sherman K, Haider L, Xu J, et al. Cognitive function in multiple sclerosis improves with telerehabilitation: results from a randomized controlled trial. PLoS One 2017;12(5):e0177177 [FREE Full text] [doi: 10.1371/journal.pone.0177177] [Medline: 28493924]
8. Ruano L, Portaccio E, Goretti B, Niccolai C, Severo M, Patti F, et al. Age and disability drive cognitive impairment in multiple sclerosis across disease subtypes. Mult Scler 2017 Aug 13;23(9):1258-1267. [doi: 10.1177/1352458516674367] [Medline: 27738090]
9. DeLuca J, Chiaravalloti ND, Sandroff BM. Treatment and management of cognitive dysfunction in patients with multiple sclerosis. Nat Rev Neurol 2020 Jun 05;16(6):319-332. [doi: 10.1038/s41582-020-0355-1] [Medline: 32372033]
10. Pozzilli V, Cruciani A, Capone F, Motolese F, Rossi M, Pilato F, et al. A true isolated cognitive relapse in multiple sclerosis. Neurol Sci 2023 Jan;44(1):339-342. [doi: 10.1007/s10072-022-06441-w] [Medline: 36207651]
11. Benedict RH, Morrow S, Rodgers J, Hojnacki D, Bucello MA, Zivadinov R, et al. Characterizing cognitive function during relapse in multiple sclerosis. Mult Scler 2014 Nov;20(13):1745-1752. [doi: 10.1177/1352458514533229] [Medline: 24842959]
12. Global smartphone penetration rate as share of population from 2016 to 2023. Statista. URL: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/ [accessed 2024-05-23]
13. Foong YC, Bridge F, Merlo D, Gresle M, Zhu C, Buzzard K, et al. Smartphone monitoring of cognition in people with multiple sclerosis: a systematic review. Mult Scler Relat Disord 2023 May;73:104674 [FREE Full text] [doi: 10.1016/j.msard.2023.104674] [Medline: 37001409]

XSL•FO
RenderX

14. Macías Islas M, Ciampi E. Assessment and impact of cognitive impairment in multiple sclerosis: an overview. Biomedicines 2019 Mar 19;7(1):22 [FREE Full text] [doi: 10.3390/biomedicines7010022] [Medline: 30893874]
15. Podda J, Tacchino A, Ponzio M, Di Antonio F, Susini A, Pedullà L, et al. Mobile health app (DIGICOG-MS) for self-assessment of cognitive impairment in people with multiple sclerosis: instrument validation and usability study. JMIR Form Res 2024 Jun 20;8:e56074 [FREE Full text] [doi: 10.2196/56074] [Medline: 38900535]
16. Beier M, Alschuler K, Amtmann D, Hughes A, Madathil R, Ehde D. iCAMS: assessing the reliability of a brief international cognitive assessment for multiple sclerosis (BICAMS) tablet application. Int J MS Care 2020;22(2):67-74 [FREE Full text] [doi: 10.7224/1537-2073.2018-108] [Medline: 32410901]
17. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, et al. Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr;23(5):721-733 [FREE Full text] [doi: 10.1177/1352458517690821] [Medline: 28206827]
18. Langdon DW, Amato MP, Boringa J, Brochet B, Foley F, Fredrikson S, et al. Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). Mult Scler 2012 Jun;18(6):891-898 [FREE Full text] [doi: 10.1177/1352458511431076] [Medline: 22190573]
19. Pereira DR, Costa P, Cerqueira JJ. Repeated assessment and practice effects of the written symbol digit modalities test using a short inter-test interval. Arch Clin Neuropsychol 2015 Aug;30(5):424-434. [doi: 10.1093/arclin/acv028] [Medline: 25994156]
20. Lamers I, Kerkhofs L, Raats J, Kos D, van Wijmeersch B, Feys P. Perceived and actual arm performance in multiple sclerosis: relationship with clinical tests according to hand dominance. Mult Scler 2013 Sep;19(10):1341-1348. [doi: 10.1177/1352458513475832] [Medline: 23407701]
21. van Hecke W, Costers L, Descamps A, Ribbens A, Nagels G, Smeets D, et al. A novel digital care management platform to monitor clinical and subclinical disease activity in multiple sclerosis. Brain Sci 2021 Sep 03;11(9):1171 [FREE Full text] [doi: 10.3390/brainsci11091171] [Medline: 34573193]
22. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 2018 Feb;17(2):162-173. [doi: 10.1016/S1474-4422(17)30470-2] [Medline: 29275977]
23. Rypma B, Berger JS, Prabhakaran V, Bly BM, Kimberg DY, Biswal BB, et al. Neural correlates of cognitive efficiency. Neuroimage 2006 Nov 15;33(3):969-979. [doi: 10.1016/j.neuroimage.2006.05.065] [Medline: 17010646]
24. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
25. Hilbert S, Nakagawa TT, Puci P, Zech A, Bühner M. The digit span backwards task: verbal and visual cognitive strategies in working memory assessment. Eur J Psychol Assess 2015;31(3):174-180. [doi: 10.1027/1015-5759/A000223]
26. Woods DL, Herron TJ, Yund EW, Hink RF, Kishiyama MM, Reed B. Computerized analysis of error patterns in digit span recall. J Clin Exp Neuropsychol 2011 Aug;33(7):721-734. [doi: 10.1080/13803395.2010.550602] [Medline: 21957866]
27. Pham L, Harris T, Varosanec M, Morgan V, Kosa P, Bielekova B. Smartphone-based symbol-digit modalities test reliably captures brain damage in multiple sclerosis. NPJ Digit Med 2021 Feb 24;4(1):36 [FREE Full text] [doi: 10.1038/s41746-021-00401-y] [Medline: 33627777]
28. Benedict RH, Amato MP, Boringa J, Brochet B, Foley F, Fredrikson S, et al. Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. BMC Neurol 2012 Jul 16;12(1):55 [FREE Full text] [doi: 10.1186/1471-2377-12-55] [Medline: 22799620]
29. Gerstenecker A, Martin R, Marson DC, Bashir K, Triebel KL. Introducing demographic corrections for the 10/36 spatial recall test. Int J Geriatr Psychiatry 2016 Apr;31(4):406-411 [FREE Full text] [doi: 10.1002/gps.4346] [Medline: 26270773]
30. Raiford SE, Coalson DL, Saklofske DH, Weiss LG. Practical issues in WAIS-IV administration and scoring. In: Weiss LG, Saklofske DH, Coalson DL, Raiford SE, editors. WAIS-IV Clinical Use and Interpretation: A Volume in Practical Resources for the Mental Health Professional. New York, NY: Academic Press; 2010:25-59.
31. Woods DL, Kishiyamaa MM, Lund EW, Herron TJ, Edwards B, Poliva O, et al. Improving digit span assessment of short-term verbal memory. J Clin Exp Neuropsychol 2011 Jan;33(1):101-111 [FREE Full text] [doi: 10.1080/13803395.2010.493149] [Medline: 20680884]
32. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
33. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
34. Penner IK, Raselli C, Stöcklin M, Opwis K, Kappos L, Calabrese P. The Fatigue Scale for Motor and Cognitive functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue. Mult Scler 2009 Dec;15(12):1509-1517. [doi: 10.1177/1352458509348519] [Medline: 19995840]
35. van Oirschot P, Heerings M, Wendrich K, den Teuling B, Martens MB, Jongen PJ. Symbol digit modalities test variant in a smartphone app for persons with multiple sclerosis: validation study. JMIR Mhealth Uhealth 2020 Oct 05;8(10):e18160 [FREE Full text] [doi: 10.2196/18160] [Medline: 33016886]

XSL•FO
RenderX

36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
37. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. London, UK: Pearson/Prentice Hall; 2009.
38. Benedict RH, Cookfair D, Gavett R, Gunther M, Munschauer F, Garg N, et al. Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). J Int Neuropsychol Soc 2006 Jul 27;12(4):549-558. [doi: 10.1017/s1355617706060723] [Medline: 16981607]
39. Ganzetti M, Graves JS, Holm SP, Dondelinger F, Midaglia L, Gaetano L, et al. Neural correlates of digital measures shown by structural MRI: a post-hoc analysis of a smartphone-based remote assessment feasibility study in multiple sclerosis. J Neurol 2023 Mar 05;270(3):1624-1636 [FREE Full text] [doi: 10.1007/s00415-022-11494-0] [Medline: 36469103]
40. Barrios L, Amon R, Oldrati P, Hilty M, Holz C, Lutterotti A. Cognitive fatigability assessment test (cFAST): development of a new instrument to assess cognitive fatigability and pilot study on its association to perceived fatigue in multiple sclerosis. Digit Health 2022;8:20552076221117740 [FREE Full text] [doi: 10.1177/20552076221117740] [Medline: 36046638]
41. Chen MH, Cherian C, Elenjickal K, Rafizadeh CM, Ross MK, Leow A, et al. Real-time associations among MS symptoms and cognitive dysfunction using ecological momentary assessment. Front Med (Lausanne) 2022;9:1049686 [FREE Full text] [doi: 10.3389/fmed.2022.1049686] [Medline: 36714150]
42. Weinstock ZL, Jaworski M, Dwyer MG, Jakimovski D, Burnham A, Wicks TR, et al. Auditory test of processing speed: preliminary validation of a smartphone-based test of mental speed. Mult Scler 2023 Nov;29(13):1646-1658. [doi: 10.1177/13524585231199311] [Medline: 37842763]
43. Lam KH, Twose J, Lissenberg-Witte B, Licitra G, Meijer K, Uitdehaag B, et al. The use of smartphone keystroke dynamics to passively monitor upper limb and cognitive function in multiple sclerosis: longitudinal analysis. J Med Internet Res 2022 Nov 07;24(11):e37614 [FREE Full text] [doi: 10.2196/37614] [Medline: 36342763]
44. Chen MH, Leow A, Ross MK, DeLuca J, Chiaravalloti N, Costa SL, et al. Associations between smartphone keystroke dynamics and cognition in MS. Digit Health 2022 Dec 05;8:20552076221143234 [FREE Full text] [doi: 10.1177/20552076221143234] [Medline: 36506490]
45. Pless S, Woelfle T, Naegelin Y, Lorscheider J, Wiencierz A, Reyes Ó, et al. Assessment of cognitive performance in multiple sclerosis using smartphone-based training games: a feasibility study. J Neurol 2023 Jul;270(7):3451-3463 [FREE Full text] [doi: 10.1007/s00415-023-11671-9] [Medline: 36952010]
46. Maillart E, Labauge P, Cohen M, Maarouf A, Vukusic S, Donzé C, et al. MSCopilot, a new multiple sclerosis self-assessment digital solution: results of a comparative study versus standard tests. Eur J Neurol 2020 Mar;27(3):429-436. [doi: 10.1111/ene.14091] [Medline: 31538396]
47. Pratap A, Grant D, Vegesna A, Tummalacherla M, Cohan S, Deshpande C, et al. Evaluating the utility of smartphone-based sensor assessments in persons with multiple sclerosis in the real-world using an app (elevateMS): observational, prospective pilot digital health study. JMIR Mhealth Uhealth 2020 Oct 27;8(10):e22108 [FREE Full text] [doi: 10.2196/22108] [Medline: 33107827]
48. Ross MK, Demos AP, Zulueta J, Piscitello A, Langenecker SA, McInnis M, et al. Naturalistic smartphone keyboard typing reflects processing speed and executive function. Brain Behav 2021 Nov;11(11):e2363 [FREE Full text] [doi: 10.1002/brb3.2363] [Medline: 34612605]
49. Montalban X, Graves J, Midaglia L, Mulero P, Julian L, Baker M, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. Mult Scler 2022 Apr;28(4):654-664 [FREE Full text] [doi: 10.1177/13524585211028561] [Medline: 34259588]
50. Woelfle T, Pless S, Wiencierz A, Kappos L, Naegelin Y, Lorscheider J. Practice effects of mobile tests of cognition, dexterity, and mobility on patients with multiple sclerosis: data analysis of a smartphone-based observational study. J Med Internet Res 2021 Nov 18;23(11):e30394 [FREE Full text] [doi: 10.2196/30394] [Medline: 34792480]
51. Metzger R, Garrett K, Christensen A, Foley J. Digital performance measures show sensitivity to demographic and disease characteristics in a multiple sclerosis cohort utilizing the MS care connect mobile app in a real-world setting (P18-4.004). Neurology 2022 May 03;98(18_supplement):11. [doi: 10.1212/wnl.98.18_supplement.2053]
52. Michaud JB, Penny C, Cull O, Hervet E, Chamard-Witkowski L. Remote testing apps for multiple sclerosis patients: scoping review of published articles and systematic search and review of public smartphone apps. JMIR Neurotech 2023 Feb 6;2:e37944. [doi: 10.2196/37944]
53. MS Care Connect. URL: https://www.mscareconnect.com/ [accessed 2024-05-29]
54. Scaramozza M, Chiesa PA, Zajac L, Sun Z, Tang M, Juraver A, et al. Konectom™ cognitive processing speed test enables reliable remote, unsupervised cognitive assessment in people with multiple sclerosis: exploring the use of substitution time as a novel digital outcome measure. Mult Scler 2024 Aug;30(9):1193-1204. [doi: 10.1177/13524585241259650] [Medline: 38912764]
55. van der Feen FE, de Haan GA, van der Lijn I, Huizinga F, Meilof JF, Heersema DJ, et al. Recognizing visual complaints in people with multiple sclerosis: prevalence, nature and associations with key characteristics of MS. Mult Scler Relat Disord 2022 Jan;57:103429 [FREE Full text] [doi: 10.1016/j.msard.2021.103429] [Medline: 34863058]

XSL•FO
RenderX

56. Ghandi Dezfuli M, Akbarfahimi M, Nabavi SM, Hassani Mehraban A, Jafarzadehpur E. Can hand dexterity predict the disability status of patients with multiple sclerosis? Med J Islam Repub Iran 2015;29:255 [FREE Full text] [Medline: 26793646]
57. Anders K. Resolution of students t-tests, ANOVA and analysis of variance components from intermediary data. Biochem Med (Zagreb) 2017 Jun 15;27(2):253-258 [FREE Full text] [doi: 10.11613/BM.2017.026] [Medline: 28740445]
58. López-Góngora M, Querol L, Escartín A. A one-year follow-up study of the Symbol Digit Modalities Test (SDMT) and the Paced Auditory Serial Addition Test (PASAT) in relapsing-remitting multiple sclerosis: an appraisal of comparative longitudinal sensitivity. BMC Neurol 2015 Mar 22;15(1):40 [FREE Full text] [doi: 10.1186/s12883-015-0296-2] [Medline: 25886168]
59. Sousa C, Rigueiro-Neves M, Passos AM, Ferreira A, Sá MJ, Group for Validation of the BRBN-T in the Portuguese MS Population. Assessment of cognitive functions in patients with multiple sclerosis applying the normative values of the Rao's brief repeatable battery in the Portuguese population. BMC Neurol 2021 Apr 21;21(1):170 [FREE Full text] [doi: 10.1186/s12883-021-02193-w] [Medline: 33882847]
60. Giedraitiene N, Kaubrys G, Kizlaitiene R. Cognition during and after multiple sclerosis relapse as assessed with the brief international cognitive assessment for multiple sclerosis. Sci Rep 2018 May 25;8(1):8169 [FREE Full text] [doi: 10.1038/s41598-018-26449-7] [Medline: 29802384]
61. Andreasen AK, Iversen P, Marstrand L, Siersma V, Siebner HR, Sellebjerg F. Structural and cognitive correlates of fatigue in progressive multiple sclerosis. Neurol Res 2019 Feb;41(2):168-176. [doi: 10.1080/01616412.2018.1547813] [Medline: 30513278]
62. van Laethem D, Denissen S, Costers L, Descamps A, Baijot J, van Remoortel A, et al. The finger dexterity test: validation study of a smartphone-based manual dexterity assessment. Mult Scler 2024 Jan;30(1):121-130. [doi: 10.1177/13524585231216007] [Medline: 38140857]
63. Holman H, Lorig K. Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. BMJ 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]
64. Lam KH, Bucur IG, van Oirschot PV, de Graaf FD, Weda H, Strijbis E, et al. Towards individualized monitoring of cognition in multiple sclerosis in the digital era: a one-year cohort study. Mult Scler Relat Disord 2022 Apr;60:103692 [FREE Full text] [doi: 10.1016/j.msard.2022.103692] [Medline: 35219240]
65. Raimo S, Giorgini R, Gaita M, Costanzo A, Spitaleri D, Palermo L, et al. Sensitivity of conventional cognitive tests in multiple sclerosis: application of item response theory. Mult Scler Relat Disord 2023 Jan;69:104440. [doi: 10.1016/j.msard.2022.104440] [Medline: 36495845]
66. Strober L, Englert J, Munschauer F, Weinstock-Guttman B, Rao S, Benedict R. Sensitivity of conventional memory tests in multiple sclerosis: comparing the Rao Brief Repeatable Neuropsychological Battery and the Minimal Assessment of Cognitive Function in MS. Mult Scler 2009 Sep;15(9):1077-1084. [doi: 10.1177/1352458509106615] [Medline: 19556311]
67. AIMS-VUB / smartphone_tests. GitHub. URL: https://github.com/AIMS-VUB/smartphone_tests [accessed 2024-12-30]

Abbreviations

BDI: Beck Depression Inventory
EDSS: Expanded Disability Status Scale
FSMC: Fatigue Scale for Motor and Cognitive Functions
ICC: intraclass correlation coefficient
IPS: information processing speed
MS: multiple sclerosis
SDMT: Symbol Digit Modalities Test
SPART: 10/36 Spatial Recall Test
vBDS: visual backward digit span

Edited by A Mavragani; submitted 10.10.23; peer-reviewed by M Ghozali, O Mirmosayyeb, FB Santinelli; comments to author 06.04.24; revised version received 30.05.24; accepted 07.11.24; published 20.01.25.

Please cite as:
Denissen S, Van Laethem D, Baijot J, Costers L, Descamps A, Van Remoortel A, Van Merhaegen-Wieleman A, D'hooghe M, D'Haeseleer M, Smeets D, Sima DM, Van Schependom J, Nagels G
A New Smartphone-Based Cognitive Screening Battery for Multiple Sclerosis (icognition): Validation Study
J Med Internet Res 2025;27:e53503
URL: https://www.jmir.org/2025/1/e53503
doi:10.2196/53503
PMID:

XSL•FO
RenderX

©Stijn Denissen, Delphine Van Laethem, Johan Baijot, Lars Costers, Annabel Descamps, Ann Van Remoortel, Annick Van Merhaegen-Wieleman, Marie D'hooghe, Miguel D'Haeseleer, Dirk Smeets, Diana M Sima, Jeroen Van Schependom, Guy Nagels. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Video Teletherapy in Treating Obsessive-Compulsive Disorder in Children and Adolescents With Exposure and Response Prevention: Retrospective Longitudinal Observational Study

Jamie D Feusner[1,2,3,4], MD; Nicholas R Farrell[1], PhD; Mia Nunez[1], PhD; Nicholas Lume[1], BS; Catherine W MacDonald[1], BA; Patrick B McGrath[1], PhD; Larry Trusky[1], BS; Stephen Smith[1], BA; Andreas Rhode[1], PhD

[1]NOCD, Inc, Chicago, IL, United States

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[3]Centre for Addiction and Mental Health, Toronto, ON, Canada

[4]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Jamie D Feusner, MD
NOCD, Inc
225 N Michigan Ave Suite 1430
Chicago, IL, 60601
United States
Email: jamie@nocdhelp.com

Abstract

Background: An effective primary treatment for obsessive-compulsive disorder (OCD) in children and adolescents as well as adults is exposure and response prevention (ERP), a form of intervention in the context of cognitive-behavioral therapy. Despite strong evidence supporting the efficacy and effectiveness of ERP from studies in research and real-world settings, its clinical use remains limited. This underuse is often attributed to access barriers such as the scarcity of properly trained therapists, geographical constraints, and costs. Some of these barriers may be addressed with virtual behavioral health, providing ERP for OCD through video teletherapy and supplemented by app-based therapeutic tools and messaging support between sessions. Studies of teletherapy ERP in adults with OCD have shown benefits in research and real-world settings in both small and large samples. However, studies of teletherapy ERP in children and adolescents thus far have been in small samples and limited to research rather than real-world settings.

Objective: This study reports on the real-world effectiveness of teletherapy ERP for OCD in the largest sample (N=2173) of child and adolescent patients to date.

Methods: Children and adolescents with OCD were treated with live, face-to-face video teletherapy sessions, with parent or caregiver involvement, using ERP. Assessments were conducted at baseline, after 7-11 weeks, and after 13-17 weeks. Additionally, longitudinal assessments of OCD symptoms were performed at weeks 18-30, 31-42, and 43-54. We analyzed longitudinal outcomes of OCD symptoms, depression, anxiety, and stress using linear mixed models.

Results: Treatment resulted in a median 38.46% (IQR 12.50%-64.00%) decrease in OCD symptoms at 13-17 weeks, and 53.4% of youth met full response criteria at this point. Improvements were observed in all categories of starting symptom severity: mild (median 40.3%, IQR 8.5%-79.8%), moderate (median 38.4%, IQR 13.3%-63.6%), and severe (median 34.1%, IQR 6.6%-58.5%). In addition, there were significant reductions in the severity of depression, anxiety, and stress symptoms. The median amount of therapist involvement was 13 (IQR 10.0-16.0) appointments and 11.5 (IQR 9.0-15.0) hours. Further, symptom improvements were maintained or improved upon in the longitudinal assessment periods of weeks 18-30, 31-42, and 43-54.

Conclusions: These results show that remote ERP treatment, assisted by technology, can effectively improve both core OCD and related depression, anxiety, and stress symptoms in children and adolescents with OCD in a real-world setting. Notable outcomes were achieved in a relatively small amount of therapist time, demonstrating its efficiency. Demonstrating the usefulness of a delivery format that overcomes several traditional barriers to treatment, these findings have implications for widespread dissemination of accessible, evidence-based care for children and adolescents with OCD.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e66715) doi:10.2196/66715

KEYWORDS

digital behavioral health; youth; cognitive-behavioral therapy; exposure and response prevention; CBT; ERP; OCD; psychiatry; clinical trial; psychology; video therapy; teletherapy; e-therapy; e-counseling; cyber-counseling; adolescents; adolescence; obsessive-compulsive disorder; retrospective study; longitudinal study; observational study; ERP therapy

Introduction

Obsessive-compulsive disorder (OCD) is a common and often disabling mental health condition that affects 0.25%-4% of children and adolescents [1-3]. Without treatment, OCD can persist into adulthood. Further, it can significantly interfere with a young person's development, education, and relationships. Fortunately, OCD in young people can be effectively managed through psychotherapy, medication, or a combination of the two. Exposure and response prevention (ERP), a form of intervention in the context of cognitive-behavioral therapy (CBT), is particularly effective for treating OCD in children and adolescents. It has been extensively tested in clinical trials [4-6] and is recommended as the first-line treatment for OCD [4,7-10].

However, accessing ERP can be challenging due to a shortage of therapists trained in this specialized technique, along with the costs and geographical limitations of attending in-person therapy sessions [11-13]. Therapists with adequate training in ERP for children and adolescents are even more scarce [14,15]. Remotely delivered ERP, for example, therapy that delivers ERP via the phone or video, can potentially overcome many of these obstacles by increasing access to specialized providers, reducing travel time and costs for families, and enabling therapists to work with patients in their natural environments that trigger OCD symptoms.

Meta-analyses of remotely delivered CBT with ERP for OCD have shown benefits in studies for adults and children or adolescents [16-18]. The largest meta-analysis (22 studies; N=1796) that included studies of both adults and children or adolescents (9 studies) found that remote ERP resulted in superior reduction of OCD symptoms compared with control conditions (Hedges g=0.94, 95% CI 0.60-1.27; P<.001), and there was no significant difference in efficacy compared with in-person CBT or ERP (g=–0.104, 95% CI –0.391 to 0.184; P=.479) [17]. The majority of these studies have included mostly self-referred participants, those with any severity of OCD, concurrent psychiatric medications, and comorbidities such as depression and anxiety disorders. Further, in several countries, remote ERP, delivered both with and without therapist support, has been implemented in routine care and found to be safe and efficacious, with medium to large effect sizes [19-21]. Remote video telemedicine treatment, in general, has shown noninferiority to traditional in-person treatment for OCD, anxiety, and depression in a head-to-head study in adults [22]. Non-inferiority of remote CBT delivered by telephone compared with face-to-face CBT has also specifically been demonstrated in adolescents with OCD [23]. Remote treatments have the added benefit of allowing therapy to take place in the patient's everyday environment. This can be especially helpful for younger patients, as it enables therapists to work with them in the settings that trigger their OCD symptoms, such as at home or school.

The convenience of teletherapy, coupled with the widespread ownership of smartphones, makes this form of treatment a promising option for families seeking help for OCD. NOCD is a virtual behavioral health provider that specializes in the delivery of evidence-based treatments, including ERP, for obsessive-compulsive and related disorders. NOCD has created a virtual therapy program that delivers ERP through video teletherapy. The program aims to maximize therapeutic impact while minimizing the time required from therapists by eliminating travel time and allowing for more efficient, remote delivery sessions. The program uses treatment design elements from an open clinical trial in adults that demonstrated the effectiveness of combining the NOCD app with face-to-face therapy sessions in reducing OCD symptoms significantly and efficiently while achieving good patient satisfaction [24]. To extend the reach of its treatment and enhance its effectiveness, NOCD provides one-on-one video sessions with a therapist, along with additional support through messaging with their therapist, a web-based OCD community, and peer support. We previously reported clinical outcomes from 3552 adults with OCD treated with ERP [25]. In this retrospective observational study, the median improvement was a 45% reduction in OCD symptoms. Further, 62.9% met the criteria for a full response.

The treatment for children and adolescents at NOCD follows a similar structure as for adults, although it includes involvement of parents and other caregivers. This comprehensive model aims to improve access to effective OCD treatment for children and adolescents, thereby addressing a significant gap in mental health services for young people. The goal of this retrospective observational longitudinal study was to determine clinical outcomes in a large naturalistic sample of children and adolescents with a primary diagnosis of OCD who received ERP treatment via video teletherapy.

Methods

Research Design

This was a retrospective, observational longitudinal analysis of clinical data from children and adolescents with a primary OCD diagnosis who received ERP treatment at NOCD from July 23, 2020 (when NOCD began offering child and adolescent treatment), to May 3, 2024, when we froze the data for analysis.

Population and Sample

From the total pool of child and adolescent patients who received treatment at NOCD between July 2020 and May 2024, our analytical sample included 2173 patients who met the following criteria: (1) primary diagnosis of OCD according to *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders*

XSL•FO
RenderX

[5th Edition]) criteria and Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders (DIAMOND) assessment, (2) completion of baseline Dimensional Obsessive-Compulsive Scale (DOCS) assessment, and (3) completion of at least one DOCS assessment during weeks 13-17 of treatment.

As part of the general clinical enrollment at NOCD, parents or legal guardians initially reached out to the intake team as self-referrals or received a referral from their insurance or medical provider. Therapists trained by NOCD in OCD assessment and treatment conducted the initial diagnostic evaluations. These evaluations occurred over the first 2 sessions and included a thorough clinical review covering the biopsychosocial aspects of the individual's history and a semi-structured diagnostic interview using the DIAMOND [26]. Those diagnosed with OCD as their primary concern, according to *DSM-5* criteria [27] and the DIAMOND, received treatment. The majority of candidates with an “extreme” rating on the DIAMOND clinician-rated severity scale were directed to more intensive treatment options, such as intensive outpatient programs, partial hospitalization programs, or residential treatment programs. Exceptions to referring patients with an “extreme” DIAMOND rating to more intensive treatment were made on a case-by-case basis, based on clinician judgment. Factors considered included, for example, the patient's level of motivation to engage in outpatient ERP treatment, the presence of comorbidities that might interfere with treatment, as well as the availability and involvement of supportive family members or caregivers who could assist with the treatment process. Referrals for other significant psychiatric or substance use issues, if they were deemed to potentially interfere with ERP treatment, were also made as needed (eg, to child and adolescent psychiatrists or other specialty providers). NOCD generally provides services to individuals aged 5 years and older, although exceptions were made for some 4-year-old patients.

Treatment Approach

The treatment plan included weekly or twice-weekly 60-minute ERP sessions via video for the first 3 weeks for most, followed by, typically, 10-14 weeks of weekly 60-minute sessions to support the continuation of ERP exercises. During this phase, some transitioned to 30-minute check-in sessions based on their clinical progress. The decision to transition patients to 30-minute check-in sessions was based on clinical judgment considering factors such as symptom reduction, demonstrated mastery of ERP principles, and consistent engagement with between-session exposures. Therapists, while aiming to adhere to this structured approach, could adjust the number or frequency of sessions to meet clinical needs and accommodate patients’ and their parents’ or caregivers’ schedules. Family therapy sessions were scheduled when the therapist deemed them necessary. Treatment duration, number of visits completed, and total therapist contact hours were obtained for each patient. Additionally, patients and their parents or caregivers could engage in asynchronous in-app messaging with their therapists for guidance on homework assignments or for support. App usage (number of app opens) and messaging activity between therapists and patients were tracked to measure treatment engagement. NOCD also offered continuously available support through online monitored community groups. Patients and their parents or caregivers also had access to the NOCD app for tools to set up ERP exercises with their therapist, record the results of their ERP exercises, and access and interact with the online NOCD community. The NOCD app’s ERP tools include ways to set up exposures with instructions, images, scripts, audio recordings, and external links; means to record obsessions and track distress reduction associated with exposure exercises (with graphical depictions); and response prevention tips (see Multimedia Appendix 1 for example screenshots of the NOCD app's key features). Parents or caregivers and caregivers of youth with OCD were able to access weekly support groups facilitated by a trained clinician to receive ongoing support and guidance in reducing accommodation of OCD in the home. Lastly, parents or caregivers received personalized psychoeducation on OCD symptoms in youth and on the maintaining role of familial accommodation and were given guidance in developing a plan to progressively reduce and eventually eliminate any accommodation behaviors.

Technology Platform

All sessions used a secure, US HIPAA (Health Insurance Portability and Accountability Act)-compliant (and compliant with other countries’ health information privacy regulations) version of Zoom, accessible via personal computing or mobile devices, with live technical support available during business hours to address connectivity issues.

Therapist Qualifications and Training

Study therapists held master’s or doctoral degrees and were licensed in the states of patients to whom they provided treatment or were associate clinicians who were supervised by licensed therapists. Therapists underwent comprehensive ERP training from NOCD. This included multiple days of focused instruction on OCD and ERP techniques. This was followed by evaluations, including written tests and mock diagnostics, education, and ERP treatment sessions. Therapists were required to pass these written and practical evaluations before starting to treat patients. Additionally, clinicians received further training and guided practice opportunities in the following areas: (1) providing psychoeducation to youth using age-appropriate language and examples, (2) tailoring ERP for OCD in youth, and (3) addressing familial accommodation. Ongoing consultation was provided, including weekly group sessions and periodic case reviews.

Data Collection and Assessment Instruments

Patients completed self-report assessments to avoid therapist bias. These included the DOCS, which was the primary OCD outcome measure, and the Depression, Anxiety, and Stress Scale (DASS-21) to measure commonly occurring comorbid depressive, anxiety, and stress symptoms. Links to these assessments were sent to patients or parents or caregivers via the NOCD app every 3 weeks. Assessments were completed collaboratively between the patient and their parent or caregiver. Therapists also completed the DIAMOND severity scale, a clinician-rated measure of OCD severity.

XSL•FO
RenderX

DOCS

The DOCS [28] is a 20-item self-report measure of OCD symptom severity across 4 domains: contamination, responsibility for harm or mistakes, unacceptable thoughts, and incompleteness or symmetry. Respondents rate the severity of their symptoms on a scale from 0 (no symptoms) to 4 (extreme symptoms) across 5 items: time occupied, avoidance behavior, associated distress, functional interference, and difficulty disregarding obsessions or refraining from compulsions. The 4 subscale scores (range 0-20) can be summed to produce a total DOCS score (range 0-80). A DOCS total score of 18 or above optimally distinguishes someone with OCD from someone without a psychiatric diagnosis [29]. The DOCS has shown good psychometric properties, including strong convergent validity with the Yale-Brown Obsessive Compulsive Scale (r=0.54) and the Obsessive-Compulsive Inventory-Revised (r=0.69), and is sensitive to the effects of treatment. Further, it shows strong correlations with the Obsessive-Compulsive Inventory-Children's Version Revised (OCI-CV-R) [30] at baseline (r=0.77), at 3 weeks (r=0.80), and at 9 weeks (r=0.80) of treatment [31]. For the purposes of this analysis, a full response was defined as a ≥35% reduction in DOCS scores and a partial response as a 25%-35% reduction, aligned with recommended criteria for treatment response and remission in OCD [29,31].

DIAMOND

The DIAMOND severity scale [26] is a 2-item clinician-rated assessment of the overall severity of an individual's emotional distress and functional impairment related to OCD symptoms. The clinician makes separate ratings of an individual's emotional distress and functional impairment on a scale ranging from 1 (normal) to 7 (extreme), and the higher of the 2 ratings is taken as the total severity score. On the DIAMOND clinician-rated severity scale, a score of 7 indicates an "extreme" level of OCD symptoms and functional impairment.

DASS-21

The DASS-21 [32] is a 21-item self-report measure of symptoms of depression, anxiety, and stress. It has been widely used in previous research and has consistently shown good psychometric qualities.

Statistical Analysis

Data, pseudoanonymized before analysis, were examined using a linear mixed model approach, with time points as fixed factors and patients as random factors. This statistical method allowed us to include all available data points while accounting for the repeated measures structure of the data. The time points for rating scale scores included ratings at baseline, the most recent rating obtained between weeks 7-11, and the most recent rating obtained between weeks 13-17. The primary measurement of interest was DOCS score changes from baseline to week 13-17. Since there was some degree of flexibility in the treatment, not everyone had rating scales done at precisely the same session or week of their treatment. Thus, these windows allowed us to measure symptom improvement at approximately a midpoint in treatment (weeks 7-11) and at the end of the active treatment period (weeks 13-17). We only included patients who had data from at least both of these time points. Of the patients who completed both baseline and week 13-17 assessments, a subset also completed assessments during weeks 7-11.

The primary (DOCS scores) and secondary (DASS-21 depression, anxiety, and stress scores) outcome measures were analyzed using linear mixed models, with time points as fixed factors and patients as random factors and with statistical significance set at an α of 0.05. We calculated effect sizes using Hedges g. Descriptive statistics, including treatment duration and mean and median symptom improvements, were calculated for those who had a baseline and at least one subsequent rating at 13-17 weeks. The primary analysis included data from both children and adolescents together, with subsequent post hoc analyses examining the effect of age group (children: ages 4-12 years, and adolescents: ages 13-17 years). The model included time (assessment_bin), age group, and their interaction as fixed effects. Similarly, we analyzed the effect of racial and ethnic category with a model including time (assessment_bin), racial and ethnic category, and their interaction as fixed effects. We also analyzed the effect of therapy over time on individual OCD domains, with the respective DOCS subscore as the outcome variable. The model included time (assessment_bin), OCD subtype (subscore_name), and their interaction as fixed effects. A random intercept for each user accounted for individual variability. Analyses were conducted in R.

Ethical Considerations

The analyses in this study did not require research ethics board review, as this does not meet the criteria for human subject research as defined by federal regulations for human subject protections, 45 CFR 46.102(e); this is a secondary analysis of de-identified data from clinical records, obtained and analyzed retrospectively, and was not the result of a research intervention or interaction. The University of California, Los Angeles (UCLA) institutional review board (IRB) office confirmed that this study did not meet the criteria for human subject research (PRE#20-008583) and thus did not require approval. Compliance with data protection laws was ensured through NOCD's privacy policy, which all patients agreed to, outlining data use and protection measures.

The UCLA IRB determined that analyses conducted in this study did not meet criteria for human subject research (PRE#20-008583) as defined by federal regulations for human subject protections, 45 CFR 46.102(e). This determination was made because this is a secondary analysis of de-identified data from clinical records, obtained and analyzed retrospectively, and was not the result of a research intervention or interaction. Regarding informed consent, since this was a secondary analysis of existing clinical data, additional research consent was not required beyond the original consent provided for receiving clinical care. All patients agreed to NOCD's privacy policy and terms of service during their clinical care, which included provisions for the use of deidentified data for research purposes. Patients had the ability to opt out of data sharing. Data privacy and confidentiality were maintained through comprehensive de-identification procedures. All direct identifiers were removed from the dataset before analysis. The de-identified data was stored and analyzed on secure, HIPAA-compliant servers with

XSL•FO
RenderX

restricted access limited to authorized clinical and research personnel. No direct participant compensation was provided for this retrospective analysis, as it did not involve any additional participation beyond standard clinical care.

Results

Table 1 summarizes the demographic and psychometric characteristics of the study participants.

XSL•FO
RenderX

Table 1. Demographics and psychometrics, N=2173[a].

Characteristics	Values
Age, mean (SD)	13.44 (2.77)
Gender, n (%)	
Male	1036 (47.68)
Female	763 (35.11)
Not reported or prefer not to say	374 (17.21)
Race and ethnicity, n (%)	
White	949 (43.67)
Not recorded	721 (33.18)
More than one race	177 (8.15)
Hispanic or Latino	126 (5.80)
Asian, Asian American, or Pacific Islander	95 (4.37)
Undisclosed	42 (1.93)
Other	40 (1.84)
Black or African American	22 (1.01)
American Indian or Alaska Native	1 (0.05)
Insurance pay, n (%)	1102 (50.71)
Cash pay, n (%)	1071 (49.29)
Medication, n (%)	
Currently taking	1058 (48.69)
Not currently taking	1115 (51.31)
DOCS[b] (baseline), mean (SD)	28.16 (13.50)
DASS-21[c] (baseline), mean (SD)	
Anxiety	10.96 (8.58)
Depression	11.71 (10.17)
Stress	17.58 (9.02)
Psychiatric medication, n (%)[d]	
SSRI[e]	1270 (58.44)
ADHD[f] medication[g]	232 (10.68)
Antianxiety or sedative-hypnotic[h]	124 (5.70)
Antipsychotic[i]	109 (5.02)
Other antidepressant[j]	59 (2.72)
SNRI[k]	49 (2.25)
Anticonvulsant or mood stabilizer[l]	33 (1.52)
TCA[m]	30 (1.38)
Comorbid diagnoses, n (%)[d]	
Anxiety disorders	569 (26.18)
Obsessive-compulsive and related disorders	448 (20.62)
Mood disorders	339 (15.60)
ADHD	228 (10.49)
Tic or Tourette’s disorder	202 (9.30)

XSL•FO
RenderX

Characteristics	Values
Autism spectrum disorder	64 (2.95)
Trauma and stress-related disorders	42 (1.93)
Substance use disorders	2 (0.09)
Other	492 (22.64)
At least one comorbidity	1200 (55.22)
Single comorbidity	704 (32.40)
Multiple comorbidities	496 (22.83)

[a]This sample includes individuals with DOCS assessments available at least at baseline and at 13-17 weeks.

[b]DOCS: Dimensional Obsessive-Compulsive Scale.

[c]DASS-21: Depression Anxiety and Stress Scale-21.

[d]As some individuals were on more than one psychiatric medication and had more than one comorbidity, the total across categories is greater than 100%.

[e]SSRI: selective serotonin reuptake inhibitor.

[f]ADHD: attention-deficit and hyperactivity disorder.

[g]ADHD medications included stimulants (eg, methylphenidate, amphetamine derivatives), non-stimulants (eg, atomoxetine), and alpha-2 adrenergic agonists (eg, clonidine and guanfacine).

[h]Anti-anxiety or sedative-hypnotics included non-benzodiazepine anti-anxiety agents (eg, hydroxyzine, and buspirone), benzodiazepines (eg, clonazepam and lorazepam), and beta blockers (eg, propranolol).

[i]Antipsychotics included both typical and atypical antipsychotics (eg, aripiprazole, quetiapine, risperidone, and ziprasidone).

[j]Other antidepressants included norepinephrine-dopamine reuptake inhibitors (eg, bupropion), serotonin antagonist and reuptake inhibitors (eg, trazodone), and tetracyclic antidepressants (eg, mirtazapine).

[k]SNRI: serotonin–norepinephrine reuptake inhibitor.

[l]Anticonvulsant or mood stabilizers included anticonvulsants (eg, lamotrigine, gabapentin, and oxcarbazepine) and non-anticonvulsant mood stabilizers (eg, lithium).

[m]TCA: tricyclic antidepressant (eg, clomipramine).

Sample

The total sample size of individuals who had at least a baseline and weeks 13-17 DOCS completed was 2173. The number of individuals who had a baseline, weeks 7-11, and weeks 13-17 DOCS completed was 1797. As shown in Figure 1A, the age distribution of participants at treatment initiation included both children and adolescents. The sample represented diverse geographic regions across the United States (Figure 1B).

Figure 1. Age and US geographic distributions for the total sample. (A) Participants' age at the beginning of treatment. Blue indicates children, and red indicates adolescents. (B) Percentages of patients by US geographic location of residence.

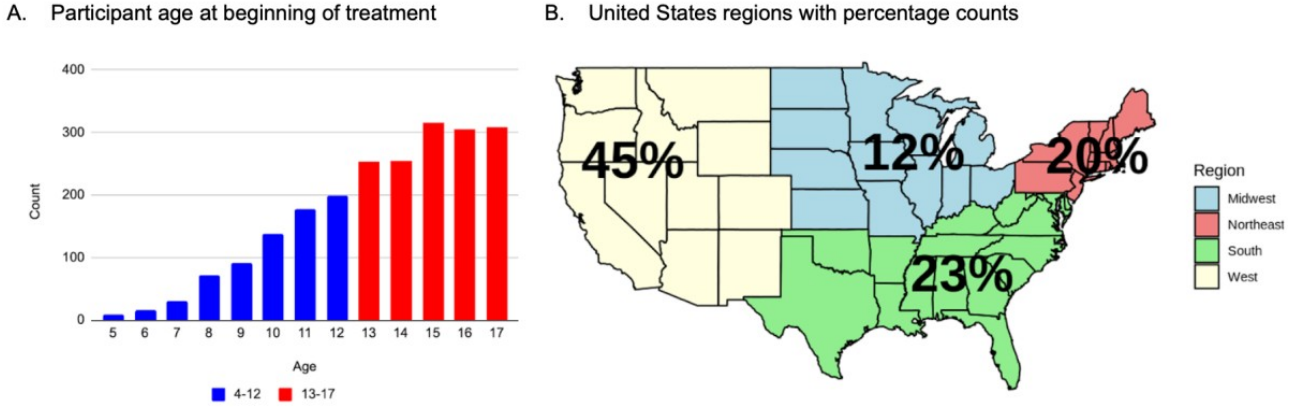

App Use and Messaging

In terms of app usage, 2129 (97.98%) opened the app more than 10 times. In total, 2127 (97.88%) had at least one chat message with their therapist, while 1835 (84.45%) had 10 or more chat messages. The mean number of chat messages for those who had at least one message was 53.53 (SD 87).

Treatment Duration

The mean treatment duration was 15.12 (SD 7.96, median 15, IQR 1-.86-17.00, mode 16) weeks, the mean number of visits was 13.70 (SD 5.83, median 13, IQR 10.00-16.00, mode 13), and the mean number of therapist hours was 12.60 (SD 5.58, median 11.50, IQR 9.00-15.00, mode 10.50).

XSL•FO
RenderX

OCD Symptom Results

NOCD treatment resulted in a significant decrease in patient-rated OCD symptoms over time (DOCS scores; effect of time: $F_{1,6927.7}$=2785, P<.001; $t_{6927.69}$=–52.77, P<.001; initial to endpoint Hedges g=–0.65: "medium" effect size, 95% CI –0.59 to –0.70). From baseline to week 7-11, DOCS scores decreased from a mean of 28.16 (SD 13.50) to a mean of 20.58 (SD 13.00), representing a mean –7.7 (95% CI –7.1 to –8.2) point decrease (27.3%) among participants who completed the forms at the midpoint. By week 13-17, DOCS scores improved to a mean of 17.94 (SD 12.9), representing a mean –10.22 (95% CI –9.7 to –10.7) point decrease (38.4%; see Figure 2 and Table 2). On the individual patient level, median DOCS score improvement was 38.46% (IQR 12.5%-64%). Further, 54.0% had a ≥35% reduction in OCD symptoms and were categorized as full "responders" [33]. A total of 64.6% achieved either partial (25%-35% reduction) or full response.

An additional analysis using the assessment time period as a categorical fixed effect (rather than ordinal, as was done in the main analysis) confirmed the overall effect of duration on scores ($F_{2,6545.5}$=1610.4, P<.001). Specifically, there was a significant decrease in scores between baseline and week 7-11 of 7.58 points (SE .179, t_{6805}=44.707, P<.001). There was an additional significant decrease in scores between weeks 7-11 and weeks 13-17 of 1.94 points (SE .224, t_{6310}=8.674, P<.001). Thus, longer treatment duration is associated with continued significant improvements in OCD symptoms.

Figure 2. Changes in obsessive-compulsive disorder symptoms as assessed by the Dimensional Obsessive-Compulsive Scale (DOCS) with treatment. Median and interquartile ranges are indicated in the box-and-whisker plots. P<.001 for the effect of assessment period.

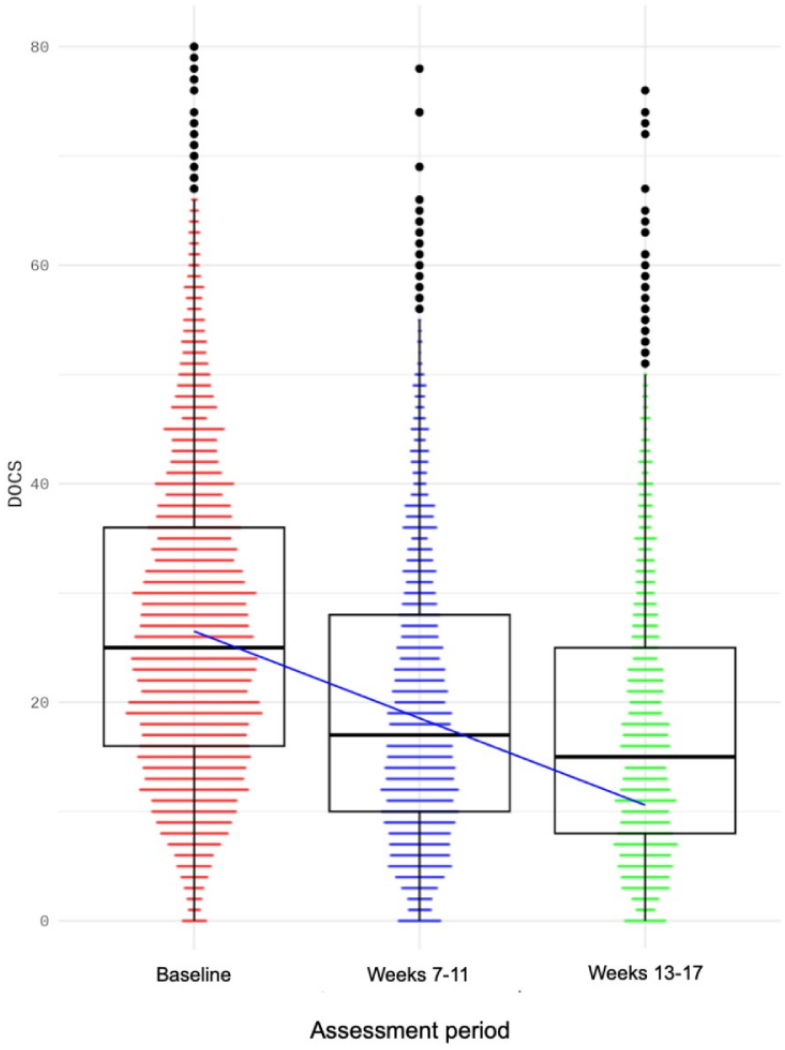

XSL•FO
RenderX

Table 2. Clinical assessments by treatment time point.

Outcome scale and assessment time point	Valid, n	Missing, n			Mean	SD
		Total missing	Did not complete form	Left treatment		
DOCS[a]						
Baseline	2173	0	0	0	28.16	13.50
Weeks 7-11	1797	376[b]	376	0	20.58	13.00
Weeks 13-17	2173	0	0	0	17.94	12.90
Weeks 18-30	1779	394	67	327	15.26	12.01
Weeks 31-42	1257	916	67	849	14.04	11.71
Weeks 43-54	859	1314	0	1314	13.51	11.87
DASS-21[c] depression						
Baseline	2173	0	—[d]	—	11.71	10.17
Weeks 7-11	1527	646[b]	—	—	9.00	9.38
Weeks 13-17	1766	407	—	—	8.04	9.19
DASS-21 anxiety						
Baseline	2173	0	—	—	10.96	8.58
Weeks 7-11	1527	646[b]	—	—	7.94	7.63
Weeks 13-17	1766	407	—	—	6.93	7.18
DASS-21 stress						
Baseline	2173	0	—	—	17.58	9.02
Weeks 7-11	1527	646[b]	—	—	13.65	8.79
Weeks 13-17	1766	407	—	—	12.60	8.73

[a]DOCS: Dimensional Obsessive-Compulsive Scale.

[b]Missing data at weeks 7-11 for DOCS and DASS-21 scales were due to the patient not completing the rating scale(s). Missing data for DOCS at weeks 18-30, 31-42, or 43-54 were due to either the patient not completing the rating scale or the patient having left treatment.

[c]DASS-21: Depression Anxiety and Stress Scale-21.

[d]Not applicable.

Post Hoc Analysis of Treatment Outcomes by Age Group and Racial and Ethnic Categories

Adolescents had higher baseline scores compared to children (β=4.52, $t_{6045.16}$=11.09, P<.001). However, the interaction between time and age group was not significant (β=–0.21, $t_{5127.48}$=–0.80, P=.42), indicating that the rate of improvement over time was not significantly different between the 2 age groups. An additional analysis of responses in children (ages 4-12 years) and adolescents (ages 13-17 years), separately, revealed similar results: in children, the effect of time on DOCS scores: β=–5.10, $t_{2132.88}$=–27.06, P<.001; in adolescents, the effect of time on DOCS scores (β=–5.39, $t_{4071.89}$=–40.99, P<.001).

An analysis testing the effects of racial and ethnic categories did not reveal any significant effects of these categories on the overall outcomes (all P values≥.10).

OCD Subtype Results

Treatment resulted in a significant decrease in patient-rated OCD symptoms (DOCS subscores) over time. With the "contamination" subtype taken as the reference group, the reported main effect of time was significant ($F_{1,52,742}$=2213.70, $t_{48,510}$=–16.31, P<.001). A significant main effect of OCD subtype ($F_{3,45,585}$=564.98, P<.001) indicated that baseline symptom severity differed across subtypes. Patients with the "thoughts" subtype had the highest baseline scores, followed by "responsibility," "symmetry," and "contamination."

The interaction between time and subtype was significant ($F_{3,45,585}$=68.85, P<.001), indicating that the rate of symptom improvement varied by subtype. The "thoughts" subtype showed the steepest decline in symptoms over time, changing by –1.85 points per assessment time point window ($t_{48,510}$=–13.81, P<.001), followed by "responsibility" (–1.49 points per assessment time point window, $t_{48,510}$=–8.77, P<.001) and "symmetry" (–1.20 points per assessment time point window, $t_{48,510}$=–4.80, P<.001). The "contamination" subtype, serving

XSL•FO
RenderX

as the reference group, experienced the slowest decline (–0.86 points per assessment time point window, $t_{48,510}$=–16.31, P<.001). These results suggest that NOCD treatment is effective across all OCD subtypes, with different magnitudes of improvement over time.

Depression, Anxiety, and Stress Results

Treatment resulted in significant improvements on the DASS-21 depression ($t_{6415.03}$=–30.55, P<.001; initial to endpoint Hedges g=–.37, 95% CI –0.32 to –0.42), DASS-21 anxiety ($t_{6571.91}$=–34.05, P<.001; initial to endpoint Hedges g=–.43, 95% CI –0.38 to –0.49), DASS-21 stress ($F_{7123.66}$=–36.66, P<.001; initial to endpoint Hedges g=–.52, 95% CI –0.47 to –0.57; Figure 3 and Table 2).

Figure 3. Changes in (A) depression, (B) anxiety, and (C) stress symptoms with treatment, as assessed by the Depression Anxiety and Stress Scale (DASS-21). Median and interquartile ranges are indicated in the box-and-whisker plots. P<.001 for the effect of assessment period.

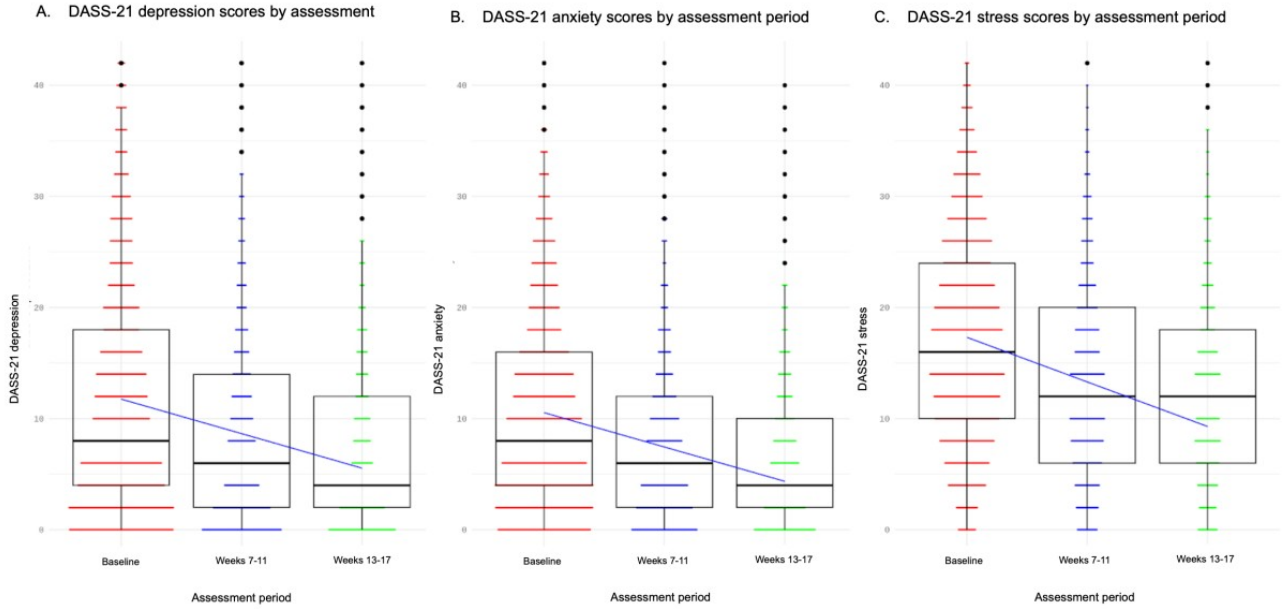

Post Hoc Analysis of Outcomes Stratified by Starting Clinician-Rated Severity Level

To determine how treatment response differed by different initial severity levels of OCD, we used the DIAMOND clinician-rated severity scale at the initial assessment to stratify patients into 3 groups of severity ratings: “Mild” (severity score of 2 or 3), “Moderate” (severity score of 4 or 5), or “Severe” (severity score of 6 or 7) (n=4 had missing DIAMOND data). For DOCS scores, on the individual patient level, the mild group (n=264) had a median 40.31% (IQR 8.51%-79.80%) reduction, the moderate group (n=1761) a median 38.36% (IQR 13.33%-63.64%) reduction, and the severe group (n=144) a median 34.07% (IQR 6.56%-58.52%) reduction. Response rates from the DOCS were 55.3% for mild, 54.23% for moderate, and 49.30% for severe. A chi-square test showed that these response rates did not differ significantly between severity groups (χ^2_2=1.49, P=.47), suggesting that treatment was similarly effective regardless of initial severity level.

Longitudinal Follow-Up

We conducted an analysis of longitudinal follow-up of OCD symptom scores after the active treatment period (weeks 13-17). At the follow-up windows of weeks 18-30, 31-42, and 43-54, most patients maintained their gains or made further improvements (effect of time: $t_{7970.47}$=–61.99, P<.001; Figure 4 and Table 2). From baseline to week 18-30 (n=1779), DOCS scores decreased from a mean of 28.16 (SD 13.50) to a mean of 15.26 (SD 12.01), representing a mean –12.9 (95% CI –12.11 to –13.7) point decrease (45.81%; Hedges g=–1, 95% CI –0.94 to –1.07). By weeks 31-42 (n=1257), DOCS scores improved to a mean of 14.04 (SD 11.71), representing a mean –14.12 (95% CI –13.27 to –14.99) point decrease (50.14%; Hedges g=–1.09, 95% CI –1.02 to –1.17). By weeks 43-54 (n=859), DOCS scores improved to a mean of 13.51 (SD 11.87), representing a mean –14.65 (95% CI –13.68 to –15.63) point decrease (52.02%) from baseline (Hedges g=–1.12, 95% CI –1.04 to –1.2).

XSL•FO
RenderX

Figure 4. Longitudinal follow-up of obsessive-compulsive disorder symptoms as assessed by the Dimensional Obsessive-Compulsive Scale (DOCS). Median and interquartile ranges are indicated in the box-and-whisker plots. *P*<.001 for the effect of assessment period.

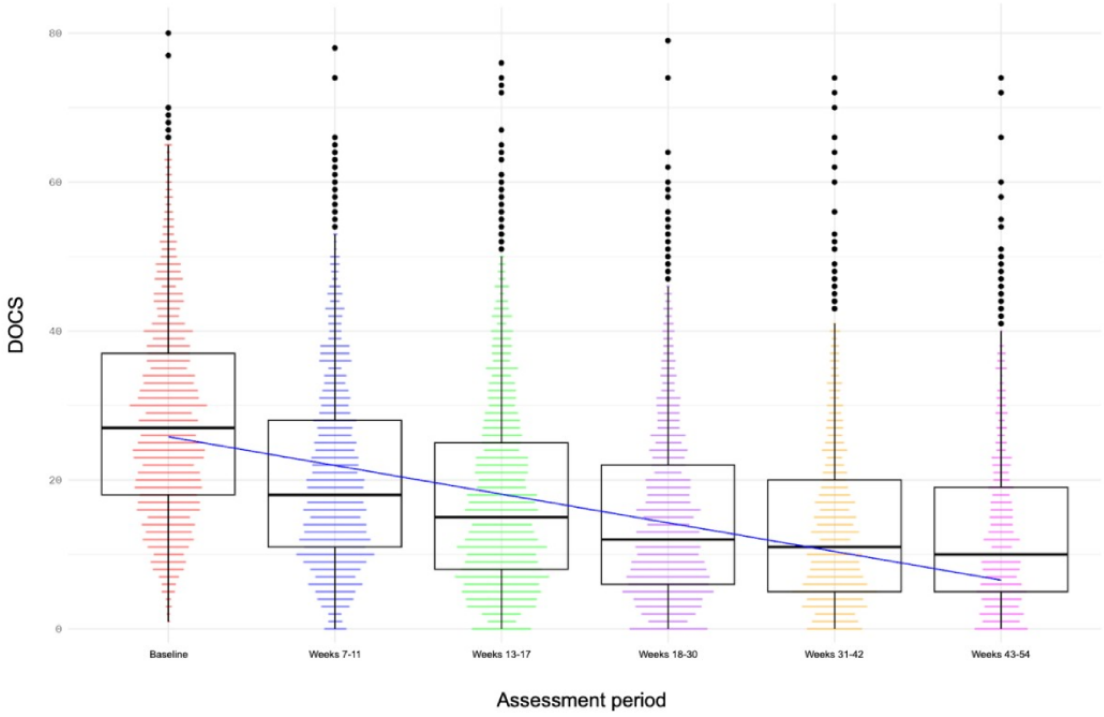

Discussion

Principal Findings

The aim of this retrospective observational longitudinal study was to assess the clinical outcomes of a large, naturalistic sample of children and adolescents with a primary diagnosis of OCD who received ERP treatment using video teletherapy. Children and adolescents exhibited significant reductions in their symptoms, demonstrating the effectiveness of this approach for younger populations. Specifically, symptom reduction was notable with a median 38.5% reduction, 54.0% of participants achieving a full response, and 64.6% showing either a partial of full response. The treatment also led to improvements in commonly co-occurring symptoms such as depression, anxiety, and stress symptoms from baseline to end point (weeks 13-17).

There were similar outcomes for children as for adolescents, and outcomes did not significantly differ by racial and ethnic categories. Moreover, the results were largely maintained, or improved upon, for those in long-term follow-up. These outcomes underscore the potential of targeted OCD treatment to alleviate a range of severe and distressing symptoms that stem from their primary OCD diathesis. This is particularly important given that an estimated 50%-87.5% of OCD cases onset before the age of 21 [34,35], and OCD is a chronic condition if left untreated. Further, individuals often endure symptoms for an average of 11 years before receiving treatment [36].

The results highlight the substantial impact and efficiency of this treatment model for OCD and associated symptoms, offering both time and cost savings. The rapid timeframe of these improvements, achieved in a median of 13 sessions and 11.5 hours, represents a significant reduction in both therapist time and treatment duration compared to what has been observed in treatment-as-usual outpatient CBT (37.0, SD 45.0 sessions) [37]. This efficiency has implications for considerable cost reductions for families and insurance providers.

Comparison to Prior Work

Meta-analyses of remotely delivered CBT with ERP for OCD have consistently shown benefits across studies of both adults, adolescents, and children. The earliest meta-analysis [16] (N=420) found that technology-delivered CBT demonstrated a moderate effect size (d=0.82) for OCD symptom reduction compared to control conditions (although this analysis primarily focused on adults, without specifying results for children or adolescents). A subsequent meta-analysis [18] (N=823) similarly reported a large effect size (g=1.17) for remote treatment of OCD symptoms compared to control, including both children and adults in their analysis but without segregating effect sizes by age group. The most recent and largest meta-analysis [17] (N=1796) found that remotely delivered CBT resulted in significant improvement in OCD symptoms compared to control conditions (g=0.936) and showed no significant difference in efficacy compared to face-to-face CBT (g=–0.104). Notably, this analysis included 9 studies specifically examining children and adolescents, representing 36.8% of the total sample. A pilot study published after these meta-analyses, focusing specifically on video-based CBT for pediatric OCD [38], demonstrated promising results, with a 90% response rate in a small sample (N=29) of children and adolescents. While encouraging, this high response rate may reflect the carefully controlled conditions typical of pilot investigations. Additionally, Weidle's study used clinician-rated outcomes rather than patient-reported metrics.

Importantly, to date, no other studies have examined the real-world effectiveness of video teletherapy for treating OCD in children and adolescents. The current retrospective

XSL•FO
RenderX

observational analysis represents the largest naturalistic examination of this treatment approach, with over 2000 patients receiving care within a community-based online clinical practice. Comparison of these results with results from controlled studies is hindered by multiple factors, including differences in sample size, self-selection (in controlled studies), treatment protocols, and outcome measures used, which may have contributed to differences in response rates. These data were from patient-reported OCD symptom scale (DOCS) rather than clinician-rated assessments, are retrospective or observational rather than a prospective study, and have a longitudinal design to evaluate the durability of treatment gains. Despite these methodological distinctions, the findings from our observational analysis, together with findings from previous research studies, suggest video-based delivery of exposure-based therapy is both effective and efficacious for treating OCD in younger populations. Our large-scale, real-world investigation provides promising evidence for the viability of this innovative service delivery model.

The treatment methodology for children and adolescents, as for adults, was influenced by a previously developed and evaluated approach to provide evidence-based ERP treatment efficiently in terms of therapist time [24]. The symptom reduction achieved in this study aligns with that found in earlier research, although direct comparisons are somewhat limited due to differences in setting (real-world clinical versus controlled research environments) and outcome measures used (patient-rated DOCS versus clinician-rated Yale-Brown Obsessive-Compulsive Scale) [36]. Additionally, the use of face-to-face teletherapy distinguishes this treatment from in-person methods previously studied.

In terms of comparison with previously-reported adult outcomes [25], the results in this child and adolescent cohort show a slightly lower magnitude of symptom reduction (median 7.7% lower). This may be due to several factors. Children and adolescents may not as readily comprehend the rationale behind ERP, which might result in some resistance in completing all required homework. In general, the idea of intentionally experiencing distress in the interest of overcoming symptoms may be highly counterintuitive, especially for children. The concepts of habituation with repeated exposures and interruption of compulsions as a way to break the cycle of obsessions and compulsions are abstract concepts that may exceed the cognitive developmental capacities of some. These potential barriers are partially mitigated, however, by parental psychoeducation and their involvement in the treatment.

Further, in this cohort, symptom improvements were relatively consistent across mild, moderate, and severe cases. This indicates the treatment model's broad applicability and effectiveness across different severity levels of OCD, even in those with severe OCD, who achieved a median of 34.07% symptom reduction. This finding emphasizes the treatment's capacity to address the needs of a diverse group of young patients in a time-efficient manner.

The treatment also resulted in decreases in depression, anxiety, and stress symptoms from baseline to endpoint (weeks 13-17). This was unsurprising as, for many with OCD, these symptoms are secondary to their core OCD symptoms; thus, as the OCD symptoms improve, improvements in the secondary symptoms often follow. This can occur to a certain extent even without specific treatments directed at them.

An innovative aspect of the NOCD model was the inclusion of additional patient support mechanisms, such as between-session SMS messaging with therapists, which was used by about 98% of members and in those who used it at an average of about 54 times. Also available was 24-hour access to NOCD's online support community. It is possible that this might facilitate a sense of belonging and understanding among participants and their parents or caregivers and help normalize their experiences by connecting them with peers facing similar challenges. This peer support, especially from individuals who had successfully completed the NOCD treatment, likely encouraged ongoing engagement and adherence to the therapy process, which is critical given the inherently challenging nature of ERP.

The use of technology, including video teletherapy and integrated communication tools, was pivotal in engaging and effectively treating a wide demographic of young patients across various locations. These technological solutions allowed for the execution of in-session, in vivo exercises tailored to the individual's symptoms and environments, enhancing the relevance and impact of the therapy. Previous research supports the efficacy of remote therapy [17], and the significant symptom improvement rates observed in this study further validate the effectiveness of virtual ERP, comparable to traditional in-person therapy.

Strengths and Limitations

There are several limitations to this analysis that should be considered. First, the use of the DOCS as the primary outcome measure is a limitation, as it was not designed specifically for use in children and adolescents. Nevertheless, we found it to be correlated with a validated self-report child-specific measure (the OCI-CV-R) [30]. However, because we did not use a rating scale that is widely used in children (such as the Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS) [39], direct comparison of results to prior studies that used other scales to assess OCD symptom severity and treatment response in pediatric populations is hindered. In future work, it would be beneficial to incorporate additional parent- or caregiver-completed assessments, such as the CY-BOCS Parent Version [40]. Second, there was a high percentage of missing data for the DOCS for weeks 13-17 and in the longitudinal follow-up periods and for the DASS-21 for weeks 7-11 and 13-17. Using linear mixed models for analyses allowed us to include all available data points (without having to discard all of a patient's data if, for example, a midpoint assessment was missing). However, the extent of missing data limits our ability to estimate responses for OCD symptoms at midpoint (weeks 13-17), in the follow-up period, and secondary outcomes of depression, anxiety, and stress. Third, this was an observational study design, which inherently precludes making causal inferences about the specific effects of the video teletherapy intervention since there was not a control arm. Relatedly, there were no standardized treatment fidelity checks, typically found in controlled trials, although NOCD therapists received the same

XSL•FO
RenderX

standardized training and were audited for adherence. Finally, while we collected satisfaction ratings after sessions, the response rate was low, limiting our ability to comprehensively assess families' satisfaction with the treatment approach. Despite these limitations, the flexibility allowed within the treatment model provides real-world applicability and potential for adaptation to the highly varied individual patient and parent needs. Additionally, the large sample size and longitudinal design offer substantial evidence for the effectiveness of this innovative service delivery approach in a naturalistic clinical setting.

Conclusions

Overall, ERP delivered via technology-assisted video teletherapy results in clinically significant improvements in OCD, depression, anxiety, and stress symptoms for children and adolescents with OCD. This is achieved in approximately 65% fewer sessions than treatment-as-usual outpatient CBT with ERP. This could translate to lower costs than such lower-intensity approaches and may also translate to reduced costs compared with higher-intensity programs for those with severe symptoms. Further, it is effective for moderate and severe OCD, which for some may prevent the need for higher levels of care such as intensive outpatient, partial hospitalization, residential, or inpatient treatment. Because OCD in this population intimately involves the family, healthier children and adolescents could translate to less stress and better health in family members. In sum, this treatment modality offers a scalable, effective option for accessing evidence-based care, potentially reducing the burden of OCD on young individuals and their families while also presenting an opportunity for significant cost savings compared to traditional treatment methods.

Acknowledgments

The authors would like to thank the NOCD therapists and member advocates for their help in facilitating treatment and care experiences. We would also like to thank Clare Beatty, MA, for her assistance with manuscript preparation. There was no external funding for the analysis.

Data Availability

The datasets analyzed during this study are proprietary business assets of NOCD Inc, and are not publicly available. The data may be available from the corresponding author on reasonable request with appropriate data use agreements in place.

Authors' Contributions

JDF and SS were involved in conceptualization; NL and AR were involved in data curation; formal analysis was done by AR; JDF and AR were involved in investigation and methodology development; JDF and LT contributed to project administration; LT and SS contributed to resources; supervision was done by JDF, NRF, MN, PBM, LT, and SS; visualization was done by AR; JDF, NRF, MN, CWM, and AR were involved in writing—original draft; and JDF, NRF, MN, NL, CWM, PBM, LT, SS, AR were involved in writing—review & editing.

Conflicts of Interest

JDF, NRF, MN, NL, CWM, PBM, LT, SS, and AR report personal fees from NOCD Inc.

Multimedia Appendix 1
Supplementary Figures. Screenshots demonstrating key features of the NOCD app interface used during treatment, including the home screen, fear hierarchy building tools, exposure exercise support features, and crisis management resources.
[DOCX File , 4599 KB - jmir_v27i1e66715_app1.docx]

References

1. Flament MF, Whitaker A, Rapoport JL, Davies M, Berg CZ, Kalikow K, et al. Obsessive compulsive disorder in adolescence: an epidemiological study. J Am Acad Child Adolesc Psychiatry 1988;27(6):764-771. [doi: 10.1097/00004583-198811000-00018] [Medline: 3264280]
2. Heyman I, Fombonne E, Simmons H, Ford T, Meltzer H, Goodman R. Prevalence of obsessive-compulsive disorder in the British nationwide survey of child mental health. Br J Psychiatry 2001;179:324-329. [doi: 10.1192/bjp.179.4.324] [Medline: 11581112]
3. Douglass HM, Moffitt TE, Dar R, McGee R, Silva P. Obsessive-compulsive disorder in a birth cohort of 18-year-olds: prevalence and predictors. J Am Acad Child Adolesc Psychiatry 1995;34(11):1424-1431. [doi: 10.1097/00004583-199511000-00008] [Medline: 8543509]
4. Fisher PL, Wells A. How effective are cognitive and behavioral treatments for obsessive-compulsive disorder? A clinical significance analysis. Behav Res Ther 2005;43(12):1543-1558. [doi: 10.1016/j.brat.2004.11.007] [Medline: 16239151]
5. Reid JE, Laws KR, Drummond L, Vismara M, Grancini B, Mpavaenda D, et al. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of

XSL•FO
RenderX

randomised controlled trials. Compr Psychiatry 2021;106:152223 [FREE Full text] [doi: 10.1016/j.comppsych.2021.152223] [Medline: 33618297]
6. Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol 2009;77(4):595-606 [FREE Full text] [doi: 10.1037/a0016032] [Medline: 19634954]
7. Abramowitz JS. Effectiveness of psychological and pharmacological treatments for obsessive-compulsive disorder: a quantitative review. J Consult Clin Psychol 1997;65(1):44-52. [doi: 10.1037//0022-006x.65.1.44] [Medline: 9103733]
8. Abramowitz JS. The psychological treatment of obsessive-compulsive disorder. Can J Psychiatry 2006;51(7):407-416. [doi: 10.1177/070674370605100702] [Medline: 16838822]
9. Koran LM, Hanna GL, Hollander E, Nestadt G, Simpson HB, American Psychiatric Association. Practice guideline for the treatment of patients with obsessive-compulsive disorder. Am J Psychiatry 2007 Jul;164(7 Suppl):5-53. [Medline: 17849776]
10. Overview | Obsessive-compulsive disorder and body dysmorphic disorder: treatment | Guidance | NICE. 2005. URL: https://www.nice.org.uk/guidance/cg31 [accessed 2024-04-19]
11. Mancebo MC, Eisen JL, Sibrava NJ, Dyck IR, Rasmussen SA. Patient utilization of cognitive-behavioral therapy for OCD. Behav Ther 2011 Sep;42(3):399-412 [FREE Full text] [doi: 10.1016/j.beth.2010.10.002] [Medline: 21658523]
12. O'Neill J, Feusner JD. Cognitive-behavioral therapy for obsessive-compulsive disorder: access to treatment, prediction of long-term outcome with neuroimaging. Psychol Res Behav Manag 2015;8:211-223 [FREE Full text] [doi: 10.2147/PRBM.S75106] [Medline: 26229514]
13. Marques L, LeBlanc NJ, Weingarden HM, Timpano KR, Jenike M, Wilhelm S. Barriers to treatment and service utilization in an internet sample of individuals with obsessive-compulsive symptoms. Depress Anxiety 2010;27(5):470-475. [doi: 10.1002/da.20694] [Medline: 20455248]
14. Reid AM, Guzick AG, Fernandez AG, Deacon B, McNamara JP, Geffken GR, et al. Exposure therapy for youth with anxiety: utilization rates and predictors of implementation in a sample of practicing clinicians from across the United States. J Anxiety Disord 2018;58:8-17. [doi: 10.1016/j.janxdis.2018.06.002] [Medline: 29929139]
15. Whiteside SPH, Deacon BJ, Benito K, Stewart E. Factors associated with practitioners' use of exposure therapy for childhood anxiety disorders. J Anxiety Disord 2016;40:29-36 [FREE Full text] [doi: 10.1016/j.janxdis.2016.04.001] [Medline: 27085463]
16. Dèttore D, Pozza A, Andersson G. Efficacy of technology-delivered cognitive behavioural therapy for OCD versus control conditions, and in comparison with therapist-administered CBT: meta-analysis of randomized controlled trials. Cogn Behav Ther 2015;44(3):190-211. [doi: 10.1080/16506073.2015.1005660] [Medline: 25705787]
17. Salazar de Pablo G, Pascual-Sánchez A, Panchal U, Clark B, Krebs G. Efficacy of remotely-delivered cognitive behavioural therapy for obsessive-compulsive disorder: an updated meta-analysis of randomised controlled trials. J Affect Disord 2023 Feb 01;322:289-299 [FREE Full text] [doi: 10.1016/j.jad.2022.11.007] [Medline: 36395988]
18. Wootton BM. Remote cognitive-behavior therapy for obsessive-compulsive symptoms: a meta-analysis. Clin Psychol Rev 2016 Feb;43:103-113. [doi: 10.1016/j.cpr.2015.10.001] [Medline: 26494179]
19. Luu J, Millard M, Newby J, Haskelberg H, Hobbs MJ, Mahoney AE. Internet-based cognitive behavioural therapy for treating symptoms of obsessive compulsive disorder in routine care. J Obsessive Compuls Relat Disord 2020 Jul;26(6):100561. [doi: 10.1016/j.jocrd.2020.100561]
20. Wootton BM, Karin E, Dear BF, Staples L, Nielssen O, Kayrouz R, et al. Internet-delivered cognitive-behaviour therapy (ICBT) for obsessive-compulsive disorder when delivered as routine clinical care: a phase IV clinical trial. J Anxiety Disord 2021 Aug;82:102444 [FREE Full text] [doi: 10.1016/j.janxdis.2021.102444] [Medline: 34273871]
21. Flygare O, Lundström L, Andersson E, Mataix-Cols D, Rück C. Implementing therapist-guided internet-delivered cognitive behaviour therapy for obsessive-compulsive disorder in the UK's IAPT programme: a pilot trial. Br J Clin Psychol 2022 Nov;61(4):895-910 [FREE Full text] [doi: 10.1111/bjc.12365] [Medline: 35315523]
22. Kishimoto T, Kinoshita S, Kitazawa M, Hishimoto A, Asami T, Suda A, J‐PROTECT collaborators. Live two-way video versus face-to-face treatment for depression, anxiety, and obsessive-compulsive disorder: a 24-week randomized controlled trial. Psychiatry Clin Neurosci 2024;78(4):220-228. [doi: 10.1111/pcn.13618] [Medline: 38102849]
23. Turner CM, Mataix-Cols D, Lovell K, Krebs G, Lang K, Byford S, et al. Telephone cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled non-inferiority trial. J Am Acad Child Adolesc Psychiatry 2014 Dec;53(12):1298-1307.e2 [FREE Full text] [doi: 10.1016/j.jaac.2014.09.012] [Medline: 25457928]
24. Gershkovich M, Middleton R, Hezel DM, Grimaldi S, Renna M, Basaraba C, et al. Integrating exposure and response prevention with a mobile app to treat obsessive-compulsive disorder: feasibility, acceptability, and preliminary effects. Behav Ther 2021 Mar;52(2):394-405. [doi: 10.1016/j.beth.2020.05.001] [Medline: 33622508]
25. Feusner JD, Farrell NR, Kreyling J, McGrath PB, Rhode A, Faneuff T, et al. Online video teletherapy treatment of obsessive-compulsive disorder using exposure and response prevention: clinical outcomes from a retrospective longitudinal observational study. J Med Internet Res 2022 May 19;24(5):e36431 [FREE Full text] [doi: 10.2196/36431] [Medline: 35587365]

XSL•FO
RenderX

26. Tolin DF, Gilliam C, Wootton BM, Bowe W, Bragdon LB, Davis E, et al. Psychometric properties of a structured diagnostic interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders. Assessment 2018 Jan;25(1):3-13. [doi: 10.1177/1073191116638410] [Medline: 26988404]
27. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5th ed. Washington, DC: American Psychiatric Association; 2013.
28. Abramowitz JS, Deacon BJ, Olatunji BO, Wheaton MG, Berman NC, Losardo D, et al. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the Dimensional Obsessive-Compulsive Scale. Psychol Assess 2010 Mar;22(1):180-198. [doi: 10.1037/a0018260] [Medline: 20230164]
29. Storch EA, De Nadai AS, Conceição do Rosário M, Shavitt RG, Torres AR, Ferrão YA, et al. Defining clinical severity in adults with obsessive-compulsive disorder. Compr Psychiatry 2015 Nov;63:30-35 [FREE Full text] [doi: 10.1016/j.comppsych.2015.08.007] [Medline: 26555489]
30. Abramovitch A, Abramowitz JS, McKay D, Cham H, Anderson KS, Farrell L, et al. The OCI-CV-R: a revision of the Obsessive-Compulsive Inventory - Child Version. J Anxiety Disord 2022 Mar;86:102532 [FREE Full text] [doi: 10.1016/j.janxdis.2022.102532] [Medline: 35091252]
31. Farrell NR, MacDonald CW, Nuñez M, Rhode A, Lume N, McGrath PB, et al. A psychometric examination of the dimensional obsessive compulsive scale in a treatment-seeking youth sample. medRxiv 2024. [doi: 10.1101/2024.09.06.24313128] [Medline: 272435099]
32. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
33. Mataix-Cols D, Fernández de la Cruz L, Nordsletten AE, Lenhard F, Isomura K, Simpson HB. Towards an international expert consensus for defining treatment response, remission, recovery and relapse in obsessive-compulsive disorder. World Psychiatry 2016 Feb;15(1):80-81 [FREE Full text] [doi: 10.1002/wps.20299] [Medline: 26833615]
34. Delorme R, Golmard J, Chabane N, Millet B, Krebs M, Mouren-Simeoni MC, et al. Admixture analysis of age at onset in obsessive-compulsive disorder. Psychol Med 2005 Feb;35(2):237-243. [doi: 10.1017/s0033291704003253] [Medline: 15841681]
35. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry 2010;15(1):53-63 [FREE Full text] [doi: 10.1038/mp.2008.94] [Medline: 18725912]
36. Pinto A, Mancebo MC, Eisen JL, Pagano ME, Rasmussen SA. The Brown Longitudinal Obsessive Compulsive Study: clinical features and symptoms of the sample at intake. J Clin Psychiatry 2006 May;67(5):703-711 [FREE Full text] [doi: 10.4088/jcp.v67n0503] [Medline: 16841619]
37. Mancebo MC, Eisen JL, Pinto A, Greenberg BD, Dyck IR, Rasmussen SA. The brown longitudinal obsessive compulsive study: treatments received and patient impressions of improvement. J Clin Psychiatry 2006 Nov;67(11):1713-1720. [doi: 10.4088/jcp.v67n1107] [Medline: 17196050]
38. Weidle B, Babiano-Espinosa L, Skokauskas N, Wolters LH, Henriksen M, Arntzen J, et al. Online CBT versus standard CBT for pediatric obsessive-compulsive disorder. Child Psychiatry Hum Dev 2024 Aug 17. [doi: 10.1007/s10578-024-01745-8] [Medline: 39152274]
39. Storch EA, McGuire JF, Wu MS, Hamblin R, McIngvale E, Cepeda SL, et al. Development and psychometric evaluation of the children's Yale-Brown obsessive-compulsive scale second edition. J Am Acad Child Adolesc Psychiatry 2019;58(1):92-98 [FREE Full text] [doi: 10.1016/j.jaac.2018.05.029] [Medline: 30577944]
40. Storch EA, Murphy TK, Adkins JW, Lewin AB, Geffken GR, Johns NB, et al. The children's Yale-Brown obsessive-compulsive scale: psychometric properties of child- and parent-report formats. J Anxiety Disord 2006;20(8):1055-1070. [doi: 10.1016/j.janxdis.2006.01.006] [Medline: 16503111]

Abbreviations

CBT: cognitive-behavioral therapy
CY-BOCS: Children’s Yale-Brown Obsessive Compulsive Scale
DASS-21: Depression, Anxiety, and Stress Scales
DIAMOND: Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders
DOCS: Dimensional Obsessive-Compulsive Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (5th Edition)
ERP: exposure and response prevention
HIPAA: US Health Insurance Portability and Accountability Act
IRB: institutional review board
OCD: obsessive-compulsive disorder
OCI-CV-R: Obsessive-Compulsive Inventory-Children’s Version Revised
UCLA: University of California, Los Angeles

XSL•FO
RenderX

Edited by A Coristine; submitted 21.09.24; peer-reviewed by K Hollmann, S Manurung; comments to author 23.10.24; revised version received 25.11.24; accepted 27.12.24; published 27.01.25.

Please cite as:
Feusner JD, Farrell NR, Nunez M, Lume N, MacDonald CW, McGrath PB, Trusky L, Smith S, Rhode A
Effectiveness of Video Teletherapy in Treating Obsessive-Compulsive Disorder in Children and Adolescents With Exposure and Response Prevention: Retrospective Longitudinal Observational Study
J Med Internet Res 2025;27:e66715
URL: https://www.jmir.org/2025/1/e66715
doi:10.2196/66715
PMID:

©Jamie D Feusner, Nicholas R Farrell, Mia Nunez, Nicholas Lume, Catherine W MacDonald, Patrick B McGrath, Larry Trusky, Stephen Smith, Andreas Rhode. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reconstructing Risk Dimensions in Telemedicine: Investigating Technology Adoption and Barriers During the COVID-19 Pandemic in Taiwan

Tzu-Chi Wu[1,2,3], MD, PhD; Chien-Ta Ho[1], DBA

[1]Graduate Institute of Technology Management, National Chung-Hsing University, Taichung, Taiwan

[2]Department of Emergency Medicine, Show Chwan Memorial Hospital, Changhua, Taiwan

[3]College of Health Sciences, Central Taiwan University of Science and Technology, Taichung, Taiwan

Corresponding Author:
Tzu-Chi Wu, MD, PhD
Department of Emergency Medicine
Show Chwan Memorial Hospital
542, Sec 1, Zhongshan Rd.
Changhua, 500009
Taiwan
Phone: 886 958352193
Email: j10062008@hotmail.com

Abstract

Background: The COVID-19 pandemic has shifted health care toward virtual and online models, impacting both users and providers. Numerous user concerns and perceived risks related to telemedicine are continually evolving and adjusting in response to the pandemic. In many countries, there has been a substantial increase in the use of virtual health care visits, which offers a unique opportunity for researchers to explore these user concerns.

Objective: This study aimed to first reconstruct the risk dimensions associated with telemedicine, then identify the risk factors affecting users' adoption, and finally propose effective solutions to mitigate these concerns. By integrating the newly constructed perceived risk with the technology acceptance model (TAM), we scrutinized various dimensions of perceived risk and their influence on users' perceptions of ease of use, perceived usefulness, and use intention (UI).

Methods: Our target population consists of adults aged ≥18 years who have used or may use telemedicine services, recruited through an anonymous, voluntary, open, web-based survey. We collected responses and used part of them to reconstruct risk dimensions using exploratory factor analysis. Subsequently, we analyzed the intricate relationship between perceived risk, the TAM, and the acceptance of telemedicine using structural equation modeling with another part of the responses.

Results: A total of 1600 valid responses were collected. Eight distinct risk dimensions were reconstructed, revealing a substantial negative impact of performance risk on UI. The psychological and social risk was the strongest barrier to the ease of using telemedicine. Time risk, provider risk, and privacy risk were not statistically significant to the TAM. The resulting model elucidates a remarkable 66% variance in UI for telemedicine services.

Conclusions: This study substantially advances the field of telemedicine research by reconstructing and redefining 8 risk dimensions and confirming the statistical significance of 5 perceived risks on the adoption of telemedicine services. These insights are poised to facilitate the promotion and enhancement of telemedicine services in the health care sector.

(J Med Internet Res 2025;27:e53306) doi:10.2196/53306

KEYWORDS

telemedicine; perceived risk; technology acceptance model; risk; performance risk

XSL•FO
RenderX

Introduction

Background

During the COVID-19 crisis, telemedicine gained global prominence and became more accessible and applicable than ever before [1]. Social distancing measures necessitated the use of telemedicine as a substitute for in-person health care interactions in many countries. Health care users and providers, who had previously been reluctant to adopt telemedicine, were compelled to embrace it [2]. A substantial disruption such as a pandemic can act as a catalyst for the acceptance of telemedicine services, potentially leading to a shift in users' perceptions and attitudes [3]. Consequently, the current expansion of telehealth has provided an opportunity to explore and reevaluate the concepts of technology acceptance and risk.

Telemedicine is an umbrella term defined as the use of technology to deliver health care and health information to individuals who are located at a distance from health care providers [4]. While the concept of telemedicine is not new, its wide-scale adoption began in the 1990s. However, telehealth is currently operating within a more information-rich environment. Due to the rapid development of telecommunication technologies and the high cost of devices, its adoption continues to face various challenges, particularly in certain regions where the full benefits of telemedicine services have not yet been realized [5,6].

According to Chen et al [7], the impact of technology on the development of telemedicine in Taiwan began with the National Information Infrastructure project, which used various network bandwidths for telemedicine starting in 1995. Before the legalization of telemedicine through an amendment to the Physician Act in 2018, Taiwan's medical laws restricted its practice to remote areas, mountainous regions, and outlying islands [8]. Following this amendment, the Taiwan Ministry of Health and Welfare expanded the nationalized health insurance program to include coverage for telemedicine outpatient consultations across different specialties in remote areas. Despite the technological advancements, use of telemedicine remains limited in Taiwan. Reports indicate a disappointingly low rate of use, with only approximately 5.83% of individuals quarantined at home in Hsinchu City using telemedicine consultations from February to May 2020 [8]. With the onset of the COVID-19 pandemic, the government further promoted telemedicine services, resulting in a larger pool of virtual visits. This expansion provides researchers with a valuable opportunity to better understand user concerns. During the COVID-19 pandemic, telemedicine in Taiwan was primarily used for treating confirmed COVID-19 cases and chronic diseases care. Due to quarantine and social distancing requirements, patients used mobile phones, tablets, or computers to engage in telemedicine consultations with health care professionals from the comfort of their homes.

Telemedicine service is not only a technological innovation but also a sociocultural phenomenon [5]. It transforms the traditional physician-patient relationship, altering both its complexity and essence [5]. A systematic review of telemedicine emphasizes the need for more patient-oriented studies, recognizing it as a complex collaborative process [9]. In addition, Bashshur et al [5] highlight the lack of research on the social, cultural, and psychological dimensions of telemedicine.

This Study

This study aims to approach the impact of risk from psychological and social perspectives, comprehending its influence on telemedicine, and proposing strategies for development, implementation, and acceptance [5].

Perceived risk is a substantial barrier to the adoption of telemedicine services, encompassing social and psychological aspects [10]. Bakshi and Tandon [11] introduced and validated a theory-driven framework that uncovers the facets of perceived risk as barriers to telemedicine acceptance in their study involving physicians in north India. Harst et al [12] argued that user-centered design and theory-guided approaches are crucial in telemedicine development. The impact of risk on the adoption of a technology is undeniable. However, with the increasing stream of literature, many different dimensions of risk have been proposed. Some studies adopt a limited number of risk dimensions, while others introduce risk dimensions from other fields; some risks share similar concepts or have a high degree of overlap. First, there is a shortage of effective categorization and definition among the various risks. Second, there remains a lack of research that comprehensively assesses the multiple dimensions of risk in telemedicine. Therefore, this study addresses substantial gaps in the telemedicine literature related to perceived risk facets and provides solutions. We aim to reconstruct the effective dimensions of perceived risk and explore what primarily influences users' acceptance of telemedicine services. We seek to evaluate the integration of the technology acceptance model (TAM) into the perceived risk dimensions, providing a solid theoretical foundation for telemedicine acceptance during a crisis.

Therefore, we posed the following research questions: (1) What are the evidence-based effective risk dimensions for telemedicine services adoption? and (2) Which risks effectively influence users' intention to use telemedicine during the COVID-19 pandemic?

To address these questions, we will conduct a comprehensive literature review to identify potential risk types in telemedicine from existing literature, providing a strong theoretical foundation. We will then empirically reconstruct these risk dimensions through exploratory factor analysis (EFA) [13]. By applying confirmatory factor analysis (CFA), we will validate and redefine the risk factors identified in the literature. Subsequently, we will integrate these redefined risk dimensions with the TAM and use CFA to explore the relationship between the 2, aiming to provide robust and credible results to strengthen the theoretical framework.

Theoretical Background

TAM Overview

The TAM was designed as a model of users' acceptance of information technologies and systems in 1986, and 3 years later, it was used to understand the attitude of computer users [14,15]. The final version of the TAM, introduced by Venkatesh and

XSL•FO
RenderX

Davis [16] in 1996, eliminated the construct of attitude and found that both perceived ease of use (PEOU) and perceived usefulness (PU) positively influence use intention (UI).

The TAM model included constructs of PU, PEOU, and UI. PU was defined as the degree of users' belief that using a particular system or service would enhance their performance [17]. PEOU means the degree of their belief that using a particular system or service would be free from effort [17]. TAM theorizes that PU and PEOU are antecedents to the adoption of technology determined by external factors [17]. In prior studies on telemedicine, PEOU positively affects PU, and both PU and PEOU positively affect UI [17].

The TAM has been used to explain various forms of computer technology and is also one of the most applied theories in the field of telemedicine [12]. Many studies considered that the TAM is an appropriate model that can be adapted to the study of technology acceptance and use of telemedicine [18,19]. One systematic review by Harst et al [12] indicates that the adaption of telemedicine service can be evaluated by using the theories of TAM and acceptance is predicted by PU, social influences, and attitude. In our study, we used the TAM to evaluate the adoption of telemedicine and hypothesized that it would be influenced by perceived risk.

Perceived Risk

Overview

The perceived risk was developed from the psychological theory proposed by Bauer [20] in 1960 to explore consumer behavior. The perceived risk is defined as a combination of uncertainty plus seriousness of the outcome involved [20]. Featherman and Pavlou [21] also adopted the same definition. Perceived risk refers to an individual's subjective expectation of experiencing a loss in pursuit of the desired outcome [21,22]. Perceived risk consists of individuals' subjective feelings and the self-evaluation of the size of the potential loss [23]. Thereafter, researchers increasingly applied perceived risk to consumer behavior and many dimensions of perceived risk were identified [21,24,25].

Cunningham [23] identified performance and psychosocial elements as perceived risk's 2 major categories and typified perceived risk into 6 dimensions, such as performance, safety, financial loss, time, and social and psychological loss [23]. Then, the importance of privacy concerns has been gradually recognized and the rich literature on consumer risk perception appeared [21]. Provider risk, physical risk, and technology risk have been presented to better understand the adoption of e-services [21]. The fear that has stemmed from the COVID-19 crisis has recently caused psychological changes in terms of consumer behavior. Information on the objective risk of the pandemic is typically scarce and is featured as unpredictable and uncertain. It remains unclear how people perceive infection risks and whether these perceptions influence their decision to adopt telemedicine [26,27].

Perceived risk has been a prominent subject of discussion within the realms of telemedicine and health care in general. In Table 1, we present a comprehensive review of previous studies that have incorporated perceived risk or its related dimensions as variables in telemedicine research, using structural equation modeling (SEM) as the primary analytical approach. The literature on perceived risk suggests different types of risk, which can be flexibly applied according to the characteristics of the product or service [28]. Among these studies, only the work of Bakshi and Tandon [11] attempts to understand the barriers to telemedicine adoption by exploring different facets of perceived risk. It is worth noting that not all risk dimensions have been included in SEM models or assessed for patients in prior studies. However, a substantial number of studies investigating the relationship between telemedicine and perceived risk did not adopt SEM as their chosen research methodology.

In the subsequent sections, we will define each risk construct and provide ample justification for selecting these constructs through the existing literature, supported by theoretical backing. Furthermore, we will reconstruct the risks using EFA and obtain empirical evidence with CFA. This approach will be used to reconstruct the effective risk dimensions and validate the reliability and influence of these risk constructs on UI.

Ten dimensions of risk were selected based on our literature review. We will discuss their relationship with telemedicine in the subsequent sections.

XSL•FO
RenderX

Table 1. Previous literature about perceived risk and telemedicine using structural equation modeling (SEM).

Study, year	Participants	PR[a]	FNR[b]	TMR[c]	PFR[d]	PLR[e]	SCR[f]	PRR[g]	PSR[h]	TNR[i]	PVR[j]	COR[k]
Rho et al [29], 2015	Users									✓	✓	
Kondrateva et al [30], 2020	Students	✓						✓				
Kamal et al [10], 2020	Users	✓						✓				
Baudier et al [31], 2021	Users	✓										
Esmaeilzadeh and Mirzaei [32], 2021	Users					✓		✓				
Kaphzan et al [33], 2022	Psychiatrists	✓										
Rad et al [34], 2022	Providers										✓	
Bakshi and Tandon [11], 2022	Physicians		✓	✓			✓	✓		✓		
Xiao et al [35], 2023	Users		✓	✓				✓				
Liu et al [36], 2024	Users	✓										
Bahari et al [37], 2024	Users			✓		✓		✓				
Singh et al [38], 2024	Older users		✓			✓				✓		

[a]PR: perceived risk.
[b]FNR: financial risk.
[c]TMR: time risk.
[d]PFR: performance risk.
[e]PLR: psychological risk.
[f]SCR: social risk.
[g]PRR: privacy risk.
[h]PSR: physical risk.
[i]TNR: technology risk.
[j]PVR: provider risk.
[k]COR: COVID-19 infection risk.

Financial Risk

The financial risk refers to potential monetary loss owing to transaction errors or subsequent maintenance costs of the productor service [11]. Financial risk in telemedicine pertains to insurance, billing, payment system, and equipment costs. Historically, insurance coverage and reimbursement for telemedicine services has been a major barrier to adoption [39]. Before the COVID-19 pandemic, most commercial health insurance plans restricted the types of services covered, and many excluded telehealth services. Users were worried that insurance payment regulations did not allow the coverage of telemedicine, which would cause extra loss of money [39]. Second, the lack of awareness about appropriate service charges and the way of payment is another cause of risk. Because of payment restrictions and complex and diversified billing before the pandemic, users' fear about the perceived risk of monetary loss was caused by transaction errors, though some countries had legalized payment parity for telehealth and more user-friendly payment systems have been developed as of now. Third, the purchasing and maintenance cost of the equipment for telemedicine and as well as the internet is relatively high, particularly in developing countries [39]. Topacan et al [40] opined that users and designers of health services should be concerned about the cost of telemedicine services, which should be even less than the cost of visiting a physician [40]. These factors lead to the following concerns: (1) using telemedicine (video consultations) might involve additional costs (financial risk 1); (2) uncertainty about the fees for telemedicine worries me (financial risk 2); (3) complex online payment systems

XSL•FO
RenderX

during telemedicine could lead to financial losses (financial risk 3); and (4) uncertainty regarding insurance coverage during telemedicine might result in financial setbacks (financial risk 4).

Performance Risk

A performance risk is the potential that a product or service will not deliver as much value and desired benefits as required [25]. Inconsistent service quality and failure to meet expectations contribute to perceived risk and consumer dissatisfaction [41]. Generally, the content of the service should be related to the physicians' professionalism and users' requirements. Deviations from one's expectations of care may result in reduced satisfaction [42]. Topacan et al [40] stated that users attach importance to obtaining accurate, relevant, and quality information in a minimum response time. A cross-sectional survey of 699 German mayors identified misdiagnosis and differences in disease applications as the primary perceived risks of telemedicine [43,44]. Users are concerned that not only would the same patient-physician connection not be established but they would not receive the level of complex care they may require because diseases could not be diagnosed through telemedicine services compared with face-to-face consultations [45,46]. These reviews and observations lead to the following survey questions: (1) telemedicine services might not be able to address my concerns (performance risk 1); (2) telemedicine might not meet my expected standards (performance risk 2); and (3) the effectiveness of telemedicine might not compare to in-person consultations (performance risk 3).

Technology Risk

Technology risk can be defined as a technological failure that disrupts business and increases the users' perceived uncertainty about telemedicine technology. Bakshi and Tandon [11] revealed that technology risk had a negative effect on a group of physicians' intention to adopt telemedicine. Consumers are reported to have the same concern. The most frequently cited barriers among patients on dialysis who use home telehealth services are malfunctioning equipment or trouble with internet connections [47]. In addition, a cross-sectional survey in Australia revealed that factors including ineffective communication, the lack of scope for physical examination, and difficulty obtaining prescriptions or examination results were reasons for perceived risk. Technological limitation is the reason why user's telehealth experiences were poorer than traditional face-to-face consultation experiences [46]. Hence, the survey includes the following questions: (1) I worry that the equipment for telemedicine might not function properly (technology risk 1); (2) I am concerned that physicians might not gather my information thoroughly during telemedicine consultations (technology risk 2); (3) I fear the absence of physical examinations (such as touch, percussion, or auscultation) during telemedicine services (technology risk 3); (4) I am worried about the effectiveness of communication with the physician during telemedicine consultations (technology risk 4); and (5) I am concerned about difficulties in obtaining medications or learning about test results through telemedicine services (technology risk 5).

Psychological Risk

Perceived psychological risk refers to users' perception of any possible psychological frustration or anxiety resulting from the use of telemedicine services, which is relatively complicated and with which users are unfamiliar compared to long-term face-to-face consultation [48]. Thus, users may operate the service system unsuccessfully, which can cause anxiety and psychological pressure, called technology anxiety. The term is derived from studies on computer anxiety in which users hold anxious and negative attitudes toward computers [49]. Previous studies among physicians in Ethiopia have shown that computer systems can bring about anxiety; users tend to have an emotional response and become anxious when they are requested to use the service [50]. The following questions are based on these observations: (1) using telemedicine services might make me anxious (perceived psychological risk 1); (2) using telemedicine services might make me feel uncomfortable (perceived psychological risk 2); (3) using telemedicine services might create stress for me, such as worrying about technical errors (perceived psychological risk 3); and (4) using telemedicine services might negatively impact my self-image and self-esteem (perceived psychological risk 4).

Social Risk

Featherman and Pavlou [21] defined social risk as the "potential loss of status in user's social group as a result of adopting the service, looking foolish, or untrendy" [21]. Social risk and influence may come from subjective norms, social factors, and specific interpersonal agreements [40]. People are influenced by others' opinions and behaviors. Individuals have favorable or unfavorable perceptions of telemedicine that in turn affects their perspectives; alternatively, not adopting telemedicine may also have positive or negative connotations [51]. Harst et al [12] demonstrated that not only the target users but the user's social environment should be considered when planning telemedicine service. This is because of the importance of users' social environment and because individuals' social group can support their use of telemedicine. One worldwide study showed that peer experience is one of the major factors behind the overall adoption of telemedicine, as it seems more attractive to them after seeing their family and friends successfully using it. The following questions are based on these observations: (1) using telemedicine service might lower my self-esteem (social risk 1); and (2) using telemedicine service might lead my family and friends to have negative opinions about me (social risk 2).

Physical Risk

Physical risk means the risk of potential threat to users' safety, physical health, or well-being. Physical risk in telemedicine services is related to the safety and health of individuals within high-risk environments [25,52]. During the COVID-19 pandemic, telemedicine could help mildly ill patients to receive the medical care they need at home while minimizing their exposure to the coronavirus [53]. Therefore, the survey contains the following questions: (1) using telemedicine service might pose risks to my physical health (physical risk 1); (2) using telemedicine service might harm my well-being (physical risk 2); and (3) using telemedicine service might lead to illness or infectious diseases, such as pneumonia (physical risk 3).

XSL•FO
RenderX

Privacy Risk

Privacy risk refers to the possibility of the loss of private data owing to fraud or hacking, which compromises the information security of online users [51]. Privacy concerns are a major barrier to online technology adoption [54]. Issues in telemedicine included substantial concerns about privacy and information security [55]. The users' critical concern about privacy included the maximum protection of personal electronic records and clinical data. Transmission and storage of sensitive information must be transparent [56]. This guarantee must be provided by those who supply the device; they must ensure that the devices are safe and noninvasive so that the users do not feel they were spied on [55]. Studies in the United States suggest that people tend to trust hospitals, which substantially moderate the negative privacy effects [54]. Therefore, the survey contains the following questions: (1) I feel that my privacy might not be secure when using telemedicine service (privacy risk 1); (2) I worry that if health care facilities do not have proper storage and backup, my personal data might be lost during the use of telemedicine service (privacy risk 2); (3) I am concerned that my personal information or medical condition might be leaked during telemedicine service (privacy risk 3); and (4) I fear that my personal data might be stolen or misused during telemedicine service (privacy risk 4).

Provider Risk

Provider risk refers to the degree to which users perceive uncertainty about telemedicine providers, including companies, hospitals, organizations, primary care providers, and physicians [29]. A nationwide survey in the United States emphasizes the importance of providers, stating that respondents become less willing to use telemedicine as they become further detached from their own providers [57]. On the basis of another nationwide survey involving 4345 adults in the United States, most participants expressed hesitation when it came to using telemedicine for consultations with health care providers not affiliated with their current organization. Only a mere 19% indicated a willingness to use it for a provider from an entirely different organization [58]. Therefore, the survey contains the following questions: (1) I worry that the providers of telemedicine service might not be knowledgeable (provider risk 1); (2) I am concerned whether professional medical personnel provide this telemedicine service (provider risk 2); and (3) I fear that the providers (physicians or hospitals) might not have enough capability during telemedicine service (provider risk 3).

Time Risk

Time risk is defined as the loss of time and inconvenience caused by delays in finding appropriate services and in learning to use particular technology [11,21]. Most users do not prefer using medical services that demand more time [40]. During the process, consumers may waste time learning how to use telemedicine services and are forced to switch to face-to-face consultation if it does not meet their expectations. Moreover, in developing economies, inadequate provision of information and communication technology infrastructure and poor internet connectivity are still formidable barriers leading to the wastage of time [57]. In contrast, a worldwide study showed that one of the major factors behind the overall adoption of telemedicine is that it saves time, as telemedicine can reduce physical travel and queue times at the hospital [59]. In addition, in some countries, telemedicine visits usually take less time than in-person visits, which are substantially associated with higher levels of satisfaction [60]. So, the survey contains the following questions: (1) accepting telemedicine service might require additional time (time risk 1); (2) slow internet speeds could potentially waste my time (time risk 2); and (3) If telemedicine service do not meet expectations, it might lead to even more time wastage (time risk 3).

Perceived COVID-19 Infection Risk

There are 2 dimensions in perceived COVID-19 infection risk: the cognitive and affective dimension and the dimension related to fear and general concerns. The cognitive or affective dimension is the qualitative dimension in the perception of COVID-19, such as its severity, controllability, and personal impact. However, the dimension related to general concerns focuses on personal feelings of fear and insecurity. Perceived risk related to COVID-19 does not come under the perceived risk of telemedicine; however, a crisis situation such as the COVID-19 pandemic influences the perceptions of telemedicine [59]. Following these theories, we can expect that people with higher risk perceptions are more likely to adopt telemedicine services. Therefore, the survey includes the following questions: (1) I feel my health is at risk during the COVID-19 pandemic (perceived COVID-19 infection risk 1); (2) I believe it is challenging to control the COVID-19 outbreak (perceived COVID-19 infection risk 2); (3) I worry about severe harm to my body if I contract the virus (perceived COVID-19 infection risk 3); (4) I think the current COVID-19 situation is more serious than before (perceived COVID-19 infection risk 4); and (5) I fear I might get infected by the coronavirus (perceived COVID-19 infection risk 5).

Methods

Instrument Development

The target population in our study were users and potential users of telemedicine services in Taiwan. The questionnaire for the survey was adopted from previous telemedicine studies or studies on fields related to health care. Both constructs suggested by the TAM model and perceived risk dimensions were included in the questionnaire [21]. To improve its quality and clarity, the questionnaire's preliminary version was tested to a pilot group of 60 citizens through the web-based survey and 3 individuals holding master's degree suggested modifications in the wording of items for conciseness. Finally, some questions were deleted owing to their low α reliability. The measurement scales were assessed on a 5-point Likert scale—from 1 for "strongly disagree" to 5 for "strongly agree." Then, it was determined whether the value of Cronbach α in all constructs was >0.65, which meant that the internal consistency and reliability of summated rating scales were acceptable [61].

Ethical Considerations

We conducted a noninterventional questionnaire-based study, using social science methods to gain a deeper understanding of

XSL•FO
RenderX

participants' perceptions. According to the Taiawn Ministry of Health and Welfare's Human Research Act, Article 5, research conducted in public settings involving nonspecific individuals, which is noninterventional, anonymous, and where the collected information cannot be used to identify specific individuals, is exempt from review [62]. At the initial stage of questionnaire distribution, participants were informed of the study's purpose and made aware that their participation was voluntary and anonymous, with no personal or private data being collected. All procedures were conducted in accordance with the ethical guidelines of the Declaration of Helsinki.

Data Collection

The study uses SurveyCake, a widely used and professional questionnaire distribution platform in Taiwan. Our target population includes individuals who have previously used telemedicine services or those who may potentially use them in the future. Recruitment was restricted to adults aged ≥18 years, as they possess medical decision-making autonomy. We primarily used public social media platforms, such as Facebook (Meta Platforms, Inc), to reach a broad audience. To ensure the representativeness of our sample, we extended our outreach to relevant online communities across various geographical regions, aiming to achieve balance in terms of age and urban-rural distribution. In addition, for populations less familiar with the internet, we used local community networks to raise awareness about recruitment. By engaging participants through a diverse range of online and offline platforms, we aimed to ensure that the sample accurately reflected Taiwan's current population, particularly in terms of sex, age, education level, and regional distribution.

Data Management

After the questionnaire collection is complete, we conducted preliminary processing of the results, including the deletion of incomplete responses. In addition, some questionnaires deemed to have low reliability were removed, including those with reverse order questions and extreme outlier values (such as very short response times and selecting the same option for all questions). Next, the collected questionnaires were randomly divided into 2 groups: group A and group B, using simple random sampling in SPSS (version 20; IBM Corp). Subsequently, normality tests were conducted to determine whether the datasets approximated a normal distribution by testing the normality of the data using the Kolmogorov-Smirnov residual method with the appropriate criteria.

The exported data were managed by SPSS and the relationship between variables was evaluated with SPSS Amos (version 26). SEM is the preferred technique among many standard methods in software and enables the evaluation of complex, multiple latent constructs and relationships [63]. Moreover, because the study analyzes the cause and different facets of perceived risk and its effect on technology acceptance, SEM has been preferred over regression. SEM comprises 2 parts, EFA and CFA. The questionnaires from group A, used for EFA, were used to ensure that the variables adequately represented the domains of the factors and to exclude variables from unrelated domains, thereby reconstructing several evidence-based risk dimensions. Meanwhile, the questionnaires from group B were used for CFA to validate the hypothesized model and its paths [64].

EFA, Construct Extraction, and Definition

First, the questionnaires from group A were used to justify the application of EFA by accessing the Bartlett test of sphericity with Kaiser-Meyer-Olkin, which was conducted with Varimax rotation and the results of EFA [65,66]. Next, we redesigned the facets of perceived risk based on eigenvalues exceeding 1.000, as determined by CFA. The EFA aims to extract domains with eigenvalues greater than 1.000, and to name the newly generated dimensions based on a combination of data analysis results (such as factor loadings) and theoretical background. In this step, we reconstructed latent variables of perceived risk, and informed the hypothesis according to the abovementioned literature review.

Bias, Reliability, and Validity

To address common method bias (CMB), we conducted a Harman 1-factor test, which involved performing an EFA on all study variables. The results indicated that the first factor accounted for >50% of the variance among the variables [67]. To mitigate concerns of multicollinearity, we used the variance inflation factor to evaluate the potential bias generated by the measurement design and determine the degree of multicollinearity among the multiple regression variables, with a suggested threshold of <10 [68].

To assess the validity of the model, 2 types of validities were used: convergent and discriminant validities. The convergent validity reflects the extent to which 2 measures capture a common construct [69]. Convergent validity is estimated by using the factor loading of all indicators and the average variance extracted (AVE). If the value of the factor loading of all items is >0.50, it is considered substantial. In addition, if the AVE is >0.5 and composite reliability is not <0.7, it indicates adequate convergence [70]. Moreover, the discriminant validity reflects the degree to which items are discriminated against amid constructs or individual conceptions are assessed for the measurement model. Two criteria are used for the estimation of discriminant validity. The first one is that the AVE itself is greater than correlations in a given construct's column or row [71]. The second one is that the square root of AVE of the corresponding construct exceeds any correlation with another construct, which means the discriminant validity is considered to be established [72].

Model Fit

To ensure the fit between the data and the measurement model or to assess the goodness of fit, 4 different measures were used, including chi-square—the ratio of the chi-square statistic to the respective *df* (χ^2/df), the goodness of fit index (GFI), root mean square error of approximation, and the comparative fit index. The chi-square is an absolute fit index, with a low chi-square value relative to the df indicating a better model fit. However, the chi-square fit statistic is affected by large samples and a ratio is <5 for the χ^2/df is preferred for acceptable model fit evaluation [73-75].

XSL•FO
RenderX

Hypothesis Testing

After the measurement model was confirmed, we used the structural model to explain the relationship among the variables and discover the connection between constructs followed by hypothesis testing. Then, bootstrapping resampling technique was used to examine model paths in the structural model—the *P* values <.05 were considered statistically significant [76].

Results

Overview

The web-based survey of our study was open to the public in Taiwan from June 13 to 30, 2022. A total of 1638 responses to the questionnaire were received, of which 38 were deleted owing to incompletion or low reliability caused by short use of answer time. Therefore, 1600 valid responses were collected for the analysis with a valid response rate of 98.2%. Using simple random sampling, all the questionnaires were divided into 2 groups: group A (600 questionnaires, 37.5%) and group B (1000 questionnaires, 62.5%). Table 2 reports the characteristics of all respondents in more detail, including reasonable representations of sex, age, and educational qualifications. Normality was assessed using the 1-sample Kolmogorov-Smirnov test, with significance values for group A and group B recorded as 0.200 and 0.550, respectively. Because both values exceed 0.05, we conclude that the data are normally distributed.

Table 2. Sample demographic of the respondents.

Characteristics	Group A (n=600)	Group B (n=1000)	Total (N=1600)
Sex, n (%)			
Male	301 (50.1)	474 (47.4)	775 (48.4)
Female	299 (49.9)	526 (52.6)	825 (51.6)
Age (y), n (%)			
18-29	213 (35.5)	318 (31.8)	531 (33.2)
30-39	116 (19.3)	169 (16.9)	285 (17.8)
40-49	98 (16.4)	214 (21.4)	312 (19.5)
>50	173 (28.8)	299 (29.9)	472 (29.5)
Education level, n (%)			
High school	71 (11.8)	129 (12.9)	200 (12.5)
Bachelor	357 (59.5)	607 (60.7)	964 (60.3)
Master	155 (25.8)	241 (24.1)	396 (24.7)
Doctoral	17 (2.8)	23 (2.3)	40 (2.5)
Experience of telemedicine, n (%)			
None	510 (85)	839 (83.9)	1349 (84.3)
1-3 times	85 (14.1)	147 (14.7)	232 (14.5)
>3 times	5 (0.9)	14 (1.4)	19 (1.2)
History of chronic disease, n (%)			
None	477 (79.5)	801 (80.1)	1278 (79.8)
Yes	123 (20.5)	199 (19.9)	322 (20.2)

Thus, group A were used to justify the application of EFA by accessing the Bartlett test of sphericity with Kaiser-Meyer-Olkin value was 0.925 and the Bartlett test of sphericity was significant (*P*<.001) [65,66]. The test was conducted with Varimax rotation and the results of EFA indicated that 8 different factors, with eigenvalues being superior to 1.000, accounted for 69.9% of the cumulative variance. Technology risk 1, technology risk 2, technology risk 4, and perceived COVID-19 infection risk 3 were deleted owing to lower factor loading, wrong concept of contrast, or the crossing of 2 different contrasts during the EFA process.

According to the results of EFA, 2 risk facets, namely "technology risk" and "social risk," were removed from the analysis. In addition, the facet originally labeled as "perceived psychological risk" was renamed "psychological and social risks" because the 2 items related to social risk were classified under the same category as perceived psychological risk by the EFA. As a result, a total of 8 facets related to perceived risk were identified (Figure 1) and details including factor loading are presented in Table 3. To sum up, 8 risk dimensions were reconstructed, including financial risk, performance risk, psychological and social risk, physical risk, privacy risk, provider risk, time risk, and perceived COVID-19 infection risk.

XSL•FO
RenderX

Figure 1. Eight risk dimensions reconstructed through exploratory factor analysis (EFA). FNR: financial risk; PCOR: perceived COVID-19 infection risk; PFR: performance risk; PLR: psychological risk; PLSR: psychological and social risks; PSR: physical risk; PRR: privacy risk; PVR: provider risk; SCR: social risk; TMR: time risk; TNR: technology risk.

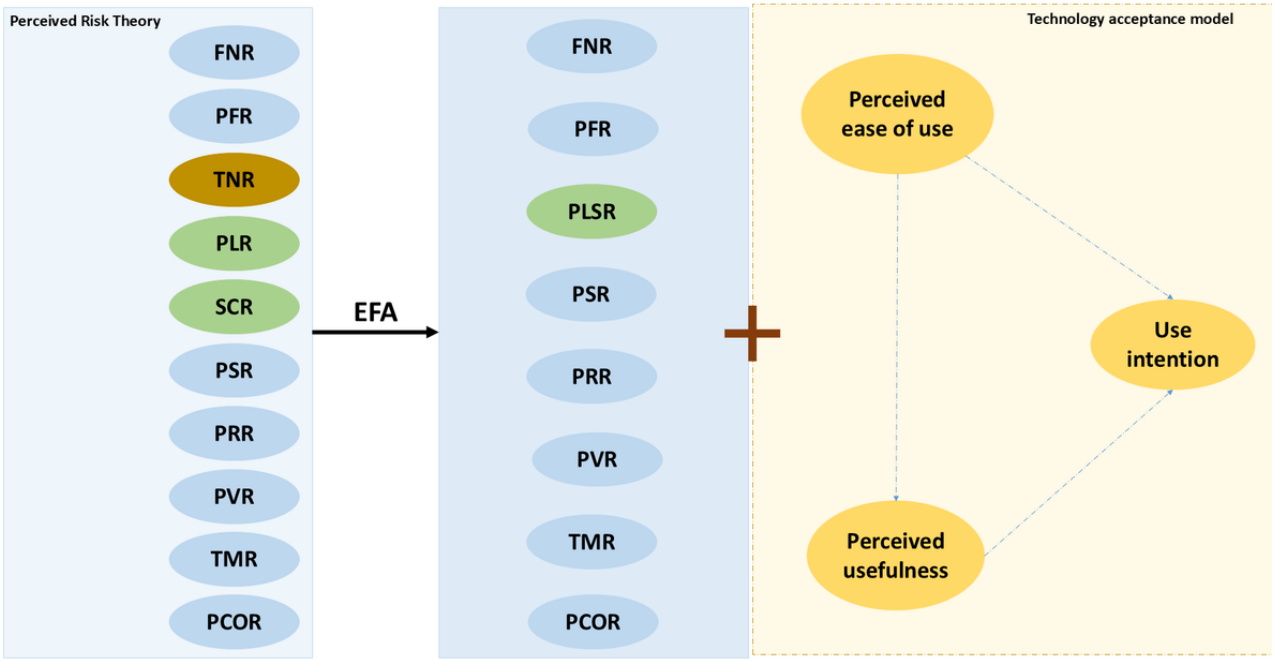

XSL•FO
RenderX

Table 3. Varimax with Kaiser normalization.

Factor and item	Factor loading of components							
	1	2	3	4	5	6	7	8
FNR[a]								
FNR1	0.702	__[b]	—	—	—	—	—	—
FNR2	0.737	—	—	—	—	—	—	—
FNR3	0.704	—	—	—	—	—	—	—
FNR4	0.649	—	—	—	—	—	—	—
PFR[c]								
PFR1	—	0.828	—	—	—	—	—	—
PFR2	—	0.822	—	—	—	—	—	—
PFR3	—	0.767	—	—	—	—	—	—
TNR3	—	0.610	—	—	—	—	—	—
PLSR[d]								
PLR1	—	—	0.741	—	—	—	—	—
PLR2	—	—	0.823	—	—	—	—	—
PLR3	—	—	0.689	—	—	—	—	—
PLR4	—	—	0.676	—	—	—	—	—
SCR1	—	—	0.611	—	—	—	—	—
SCR2	—	—	0.596	—	—	—	—	—
PSR[e]								
PSR1	—	—	—	0.800	—	—	—	—
PSR2	—	—	—	0.823	—	—	—	—
PSR3	—	—	—	0.714	—	—	—	—
PRR[f]								
PRR1	—	—	—	—	0.660	—	—	—
PRR2	—	—	—	—	0.775	—	—	—
PRR3	—	—	—	—	0.865	—	—	—
PRR4	—	—	—	—	0.833	—	—	—
PVR[g]								
PVR1	—	—	—	—	—	0.639	—	—
PVR2	—	—	—	—	—	0.691	—	—
PVR3	—	—	—	—	—	0.583	—	—
TNR5	—	—	—	—	—	0.589	—	—
TMR[h]								
TMR1	—	—	—	—	—	—	0.743	—
TMR2	—	—	—	—	—	—	0.777	—
TMR3	—	—	—	—	—	—	0.587	—
PCOR[i]								
PCOR1	—	—	—	—	—	—	—	0.721
PCOR2	—	—	—	—	—	—	—	0.813
PCOR4	—	—	—	—	—	—	—	0.756

XSL•FO
RenderX

Factor and item	Factor loading of components							
	1	2	3	4	5	6	7	8
PCOR5	—	—	—	—	—	—	—	0.714

[a]FNR: financial risk.
[b]Not applicable.
[c]PFR: performance risk.
[d]PLSR: psychological and social risks.
[e]PSR: physical risk.
[f]PRR: privacy risk.
[g]PVR: provider risk.
[h]TMR: time risk.
[i]PCOR: perceived COVID-19 infection risk.

Definition of Factors of Risk

The definitions of the 8 facets of perceived risk and constructs of TAM are provided in Table 4. The 8 risk dimensions included financial risk, performance risk, psychological and social risk, physical risk, privacy risk, provider risk, time risk, and perceived COVID-19 infection risk.

Table 4. The definition of the constructs.

Constructs	Operational definition	Reference
Financial risk	Potential for financial loss due to online payments, limited insurance coverage, and maintenance costs for products like mobile phones and the internet.	[39,77]
Performance risk	The result does not match the desired benefits.	[21,48,78]
Psychological and social risks	The risk that lowers the user's self-image, leads to a loss of self-esteem and embarrassment, and generates pressure and anxiety resulting from the use of a service.	[79]
Physical risk	The risk to the user's safety and physical health in using telemedicine.	[78]
Privacy risk	Potential loss of control over personal health information without permission.	[77,80]
Provider risk	The risk of the user's perceived uncertainty about the telemedicine provider.	[29,81]
Time risk	The risk of the user's perception of wastage of time in learning how to use the service.	[21,77]
Perceived COVID-19 infection risk	The fear and general concerns about people's perception of COVID-19 infection.	[82]
Perceived ease of use	The degree of users' belief that using telemedicine service would be free from effort.	[83]
Perceived usefulness	The degree of users' belief that using telemedicine service would enhance their performance.	[83]
Use intention	Users' desire to use telemedicine services in the future.	[10,77]

Hypothesis Formation

According to the aforementioned literature review, 18 hypotheses were derived (Textbox 1).

XSL•FO
RenderX

Textbox 1. Derived hypotheses.

- Hypothesis 1a: financial risk negatively influences the perceived ease of use (PEOU) of telemedicine services.
- Hypothesis 1b: financial risk negatively influences the usage intention (UI) in telemedicine services.
- Hypothesis 2a: performance risk negatively influences the PEOU in telemedicine services.
- Hypothesis 2b: performance risk negatively influences the perceived usefulness (PU) of telemedicine services.
- Hypothesis 2c: performance risk negatively influences the UI in telemedicine services.
- Hypothesis 3a: psychological and social risks negatively influences the PEOU of telemedicine services.
- Hypothesis 3b: psychological and social risk negatively influences PU in telemedicine services.
- Hypothesis 3c: psychological and social risk negatively influences UI in telemedicine services.
- Hypothesis 4a: physical risk negatively influences the PU of telemedicine services.
- Hypothesis 4b: physical risk negatively influences the UI in telemedicine services.
- Hypothesis 5: privacy risk negatively influences the UI in telemedicine services.
- Hypothesis 6: provider risk negatively influences the UI in telemedicine services.
- Hypothesis 7: time risk negatively influences the UI in telemedicine services.
- Hypothesis 8a: perceived COVID-19 infection risk positively influences the PU of telemedicine services.
- Hypothesis 8b: perceived COVID-19 infection risk positively influences the UI in telemedicine services.
- Hypothesis 9: PEOU positively influences the PU in telemedicine services.
- Hypothesis 10: PEOU positively influences the UI in telemedicine services.
- Hypothesis 11: PU positively influences the UI in telemedicine services.

Research Model

The research model (Figure 2) consists of 8 redefined dimensions of risk combined with the TAM, resulting in a total of 18 hypotheses.

Figure 2. Research model and hypothesis. FNR: financial risk; PFR: performance risk; PLSR: psychological and social risks; PSR: physical risk; PRR: privacy risk; TMR: time risk.

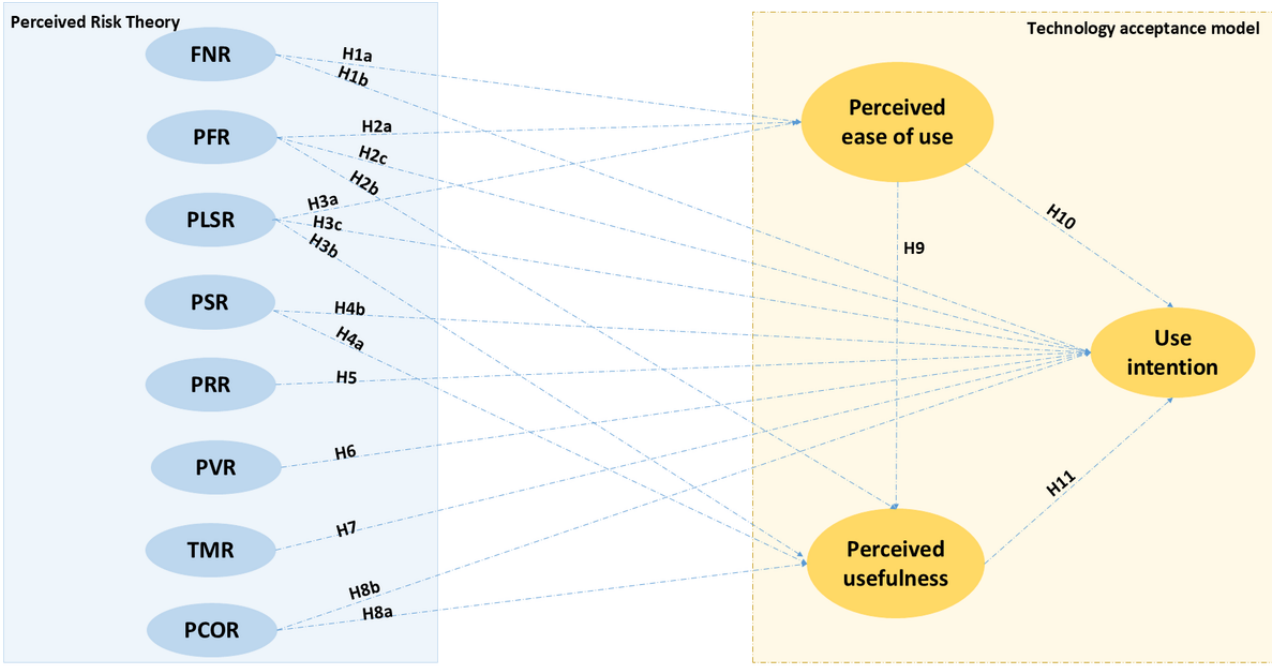

XSL•FO
RenderX

Reliability, Validity, and Bias

In CFA for the reliability and validity test (Table 5), financial risk 1 and perceived COVID-19 infection risk 5 were deleted owing to lower standard loading. PU1 was deleted from the subsequent analysis according to the result of modification indices because of a higher correlation with UI and failure of the test of discriminant validity (the square root of the AVE for PU 0.715 is less than PU of the correlations with UI 0.766). After that, the tests indicate adequate discriminant validity. Two criteria regarding discriminant validity showed that not only all AVEs were greater than the correlations in that construct's rows and columns but also the square root of AVE (diagonal) was greater than off-diagonal elements in rows and columns. The two tests mentioned earlier support discriminant validity (Table 6).

In addition, the result of the evaluation showed that all Cronbach α values were >0.7 and the internal consistency of items was considered. Moreover, all item loads on their own factors were >0.5 and the internal consistency reliability for each construct was >0.6. Most constructs' AVE values met the minimum requirement because the constructs exceeded the criterion of 0.50, though the construct—"perceived COVID-19 infection risk"—was 0.394 and "financial risk" was 0.483. However, we could accept AVE<0.5 if the composite reliability was >0.6 because the convergent validity of the construct was still adequate. Therefore, these tests indicate adequate convergent validity [72].

Financial risk, performance risk, psychological and social risk, physical risk, privacy risk, provider risk, time risk, and perceived COVID-19 infection risk had mean scores of 3.33 (SD 0.854), 3.58 (SD 0.783), 2.43 (SD 0.769), 2.49 (SD 0.897), 3.39 (SD 1.00), 3.32 (SD 0.878), 3.37 (SD 0.831), and 3.81 (SD 0.751), respectively. The mean scores of predictors of TAM such as PU, PEOU, and UI, were 4.00 (0.761), 3.81 (SD 0.804), and 3.87 (SD 0.768), respectively.

The values of variance inflation factor for all the latent variables in the model were between 1.107 and 2.241, all of which were lower than the suggested threshold, indicating low concerns about multicollinearity. For addressing CMB, the Harman 1-factor test during the EFA on all study variables showed the first factor accounted for 31% of the variance among the variables, which was lower than the suggested threshold, indicating no CMB issues.

XSL•FO
RenderX

Table 5. Reliability, validity test, and means among latent constructs.

Items	Standard loading	Values, mean (SD)	AVE[a]	CR[b]	Cronbach α
FNR[c]		3.33 (0.854)	0.483	0.780	0.807
FNR2	0.727				
FNR3	0.776				
FNR4	0.791				
PFR[d]		3.58 (0.783)	0.599	0.853	0.845
PFR1	0.842				
PFR2	0.880				
PFR3	0.765				
TNR3	0.571				
PLSR[e]		2.43 (0.769)	0.613	0.905	0.902
PLR1	0.781				
PLR2	0.825				
PLR3	0.733				
PLR4	0.807				
SCR1	0.799				
SCR2	0.749				
PSR[f]		2.49 (0.897)	0.632	0.835	0.819
PSR1	0.827				
PSR2	0.906				
PSR3	0.626				
PRR[g]		3.39 (1.00)	0.702	0.903	0.897
PRR1	0.685				
PRR2	0.811				
PRR3	0.930				
PRR4	0.903				
PVR[h]		3.32 (0.878)	0.585	0.847	0.840
PVR1	0.755				
PVR2	0.823				
PVR3	0.849				
TNR5	0.609				
TMR[i]		3.37 (0.831)	0.505	0.749	0.745
TMR1	0.543				
TMR2	0.745				
TMR3	0.816				
PCOR[j]		3.81 (0.804)	0.394	0.716	0.705
PCOV1	0.569				
PCOV2	0.764				
PCOV4	0.673				
PU[k]		4.00 (0.761)	0.655	0.792	0.791
PU2	0.799				

XSL•FO
RenderX

Items	Standard loading	Values, mean (SD)	AVE[a]	CR[b]	Cronbach α
PU3	0.821				
PEOU[l]		3.81 (0.751)	0.703	0.904	0.904
PEOU1	0.841				
PEOU2	0.812				
PEOU3	0.843				
PEOU4	0.857				
UI[m]		3.87 (0.768)	0.656	0.884	0.882
UI1	0.849				
UI2	0.811				
UI3	0.747				
UI4	0.828				

[a]AVE: average variance extracted.
[b]CR: critical ratio.
[c]FNR: financial risk.
[d]PFR: performance risk.
[e]PLSR: psychological and social risks.
[f]PSR: physical risk.
[g]PRR: privacy risk.
[h]PVR: provider risk.
[i]TMR: time risk.
[j]PCOR: perceived COVID-19 infection risk.
[k]PU: perceived usefulness.
[l]PEOU: perceived ease of use.
[m]UI: use intention.

XSL•FO
RenderX

Table 6. Correlation matrix of variables.

	1	2	3	4	5	6	7	8	9	10	11
FNR[a]	0.695	—[b]	—	—	—	—	—	—	—	—	—
PFR[c]	0.473	0.774	—	—	—	—	—	—	—	—	—
PLSR[d]	0.468	0.443	0.783	—	—	—	—	—	—	—	—
PSR[e]	0.388	0.387	0.658	0.795	—	—	—	—	—	—	—
PRR[f]	0.571	0.548	0.471	0.433	0.838	—	—	—	—	—	—
PVR[g]	0.672	0.584	0.520	0.450	0.773	0.775	—	—	—	—	—
TMR[h]	0.643	0.687	0.350	0.294	0.531	0.579	0.711	—	—	—	—
PCOR[i]	0.213	0.185	0.108	0.093	0.214	0.204	0.235	0.627	—	—	—
PU[j]	–0.184	–0.342	–0.365	–0.294	–0.220	–0.234	–0.260	0.086	0.810	—	—
PEOU[k]	–0.335	–0.306	–0.483	–0.323	–0.279	–0.325	–0.249	0.029	0.607	0.838	—
UI[l]	–0.205	–0.460	–0.356	–0.341	–0.256	–0.280	–0.311	0.046	0.775	0.625	0.810

[a]FNR: financial risk.
[b]Not applicable.
[c]PFR: performance risk.
[d]PLSR: psychological and social risks.
[e]PSR: physical risk.
[f]PRR: privacy risk.
[g]PVR: provider risk.
[h]TMR: time risk.
[i]PCOR: perceived COVID-19 infection risk.
[j]PU: perceived usefulness.
[k]PEOU: perceived ease of use.
[l]UI: use intention.

Model Fit

In the following procedure, the measurement model (Figure 3) assessment was done, and a summary of the model fit measurements is presented in Table 7, in which most results surpass the acceptable levels but the value of GFI is 0.840 which is below 0.9. As Doll et al [84] suggest, GFI in the 0.8-0.9 range presents a reasonable fit, particularly for more variance [84]. Overall, the measurement model indicates a satisfactory fit.

After the measurement model was confirmed, we used the structural model (Figure 4) to discover how closely the theory is supported by empirical data and whether the theory is empirically confirmed [85]. The result of the structural model and the summary of the model fit measurement are given in Table 8. The structural model indicates a satisfactory fit. We also tested for second-order factors in multifaceted risk perception. However, the model fit indices were inferior to those of the first-order model. Therefore, we decided to retain the original structural model.

XSL•FO
RenderX

Figure 3. Confirmatory factor analysis (CFA) for the measurement model. FNR: financial risk; PCOR: perceived COVID-19 infection risk; PEOU: perceived ease of use; PFR: performance risk; PLSR: psychological and social risks; PRR: privacy risk; PU: perceived usefulness; PVR: provider risk; TMR: time risk; UI: use intention.

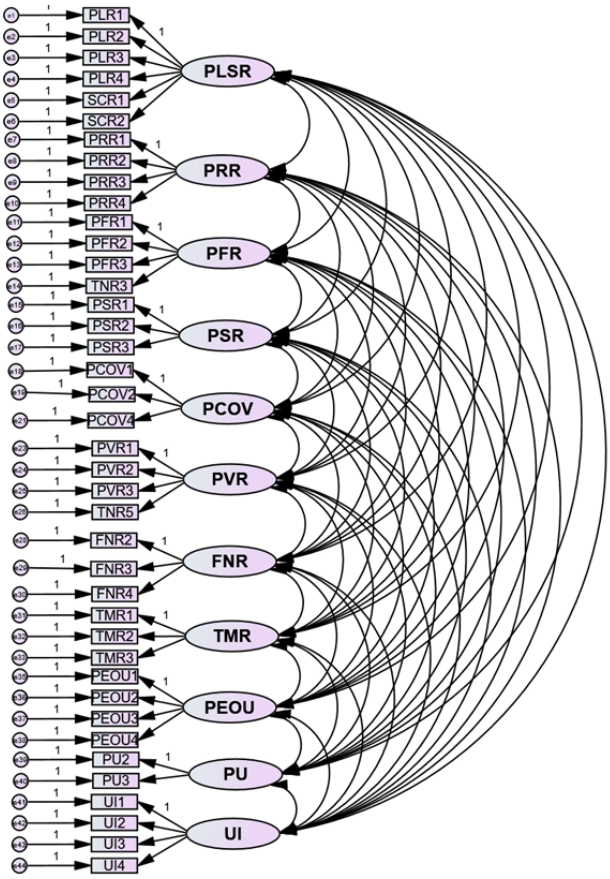

Table 7. Fit measures in the structural model.

Fit measure	Measurement model	Recommended values
χ^2 (*df*)	3116 (685)	—[a]
χ^2/df	4.55	<5
Goodness of fit index	0.840	>0.9
Comparative fit index	0.902	>0.9
Root mean square error of approximation	0.060	<0.08

[a]Not available.

XSL•FO
RenderX

Figure 4. The structural model. FNR: financial risk; PCOR: perceived COVID-19 infection risk; PEOU: perceived ease of use; PFR: performance risk; PLSR: psychological and social risks; PRR: privacy risk; PU: perceived usefulness; PVR: provider risk; TMR: time risk; UI: use intention.

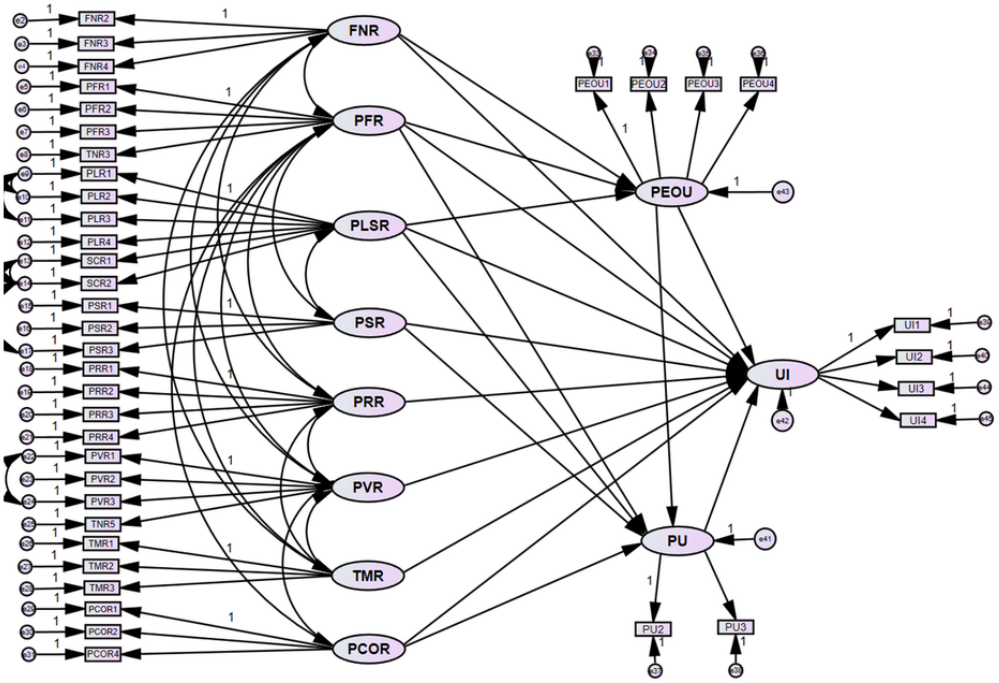

Table 8. Results of hypotheses testing.

Fit measure	Structural model	Recommended values
χ^2 *(df)*	2908 (703)	—[a]
χ^2 */df*	4.136	<5
Goodness of fit index	0.869	>0.9
Comparative fit index	0.911	>0.9
Root mean square error of approximation	0.056	<0.08

[a]Not available.

Result of Hypotheses

The hypothesis testing showed that 5 out of 8 redefined risks significantly affect the TAM. The result of the hypotheses testing is shown in Table 9. Three kinds of perceived risk—financial risk, performance risk and psychological and social risk—had significant negative effects on PEOU. The psychological and social risk showed a stronger effect (β=−0.377; *P*<.001) than the financial risk (β=−0.116; *P*=.003) and performance risk (β=−0.091; *P*=.006). PU was affected by 2 types of perceived risk. The performance risk (β=−0.133; *P*<.001) had a significant negative effect, but the perceived COVID-19 infection risk (β=0.120; *P*=.003) had a significant positive effect on PU. Furthermore, performance risk was the only kind of risk that had a significant direct negative impact on all 3 variables: PEOU, PU, and UI. However, time risk, provider risk, and privacy risk did not significantly impact the 3 contrasts of TAM. PU was significantly influenced by PEOU (β=0.511; *P*<.001). Moreover, both PU and PEOU were found to significantly affect UI (*P*<.001). In terms of goodness of fit indicators, the research model explained 66% of the variance in UI of telemedicine service as shown in Figure 5. To further assess the significance of the direct and indirect effects of predictor variables on the UI of telemedicine service, a decomposition of the effects analysis was conducted in Table 10. Overall, among the 8 types of risks, performance risk had the most significant total effect on UI (β=−0.362).

XSL•FO
RenderX

Table 9. Results of hypotheses testing.

Hypothesis	Path coefficient (SE)	CR[a]	*P* value	Result
Hypothesis 1a: PEOU[b]<-FNR[c]	−0.116 (0.039)	−2.996	.003	Accepted
Hypothesis 1b: UI[d]<-FNR	0.105 (0.048)	2.183	.03	Accepted
Hypothesis 2a: PEOU<-PFR[e]	−0.091 (0.033)	−2.754	.006	Accepted
Hypothesis 2b: PU[f]<-PFR	−0.133 (0.029)	−4.542	<.001	Accepted
Hypothesis 2c: UI<-PFR	−0.220 (0.041)	−5.419	<.001	Accepted
Hypothesis 3a: PEOU<-PLSR[g]	−0.377 (0.038)	−9.799	<.001	Accepted
Hypothesis 3b: PU<-PLSR	−0.007 (0.048)	−0.140	.89	Rejected
Hypothesis 3c: UI<-PLSR	0.109 (0.044)	2.465	.01	Accepted
Hypothesis 4a: PU<-PSR[h]	−0.050 (0.037)	−1.359	.17	Rejected
Hypothesis 4b: UI <-PSR	−0.101 (0.034)	−2.979	.003	Accepted
Hypothesis 5: UI<-PRR[i]	0.031 (0.043)	0.717	.47	Rejected
Hypothesis 6: UI <- PVR[j]	0.000 (0.040)	0.001	<.99	Rejected
Hypothesis 7: UI <- TMR[k]	−0.031 (0.067)	−0.467	.64	Rejected
Hypothesis 8a: PU <-PCOR[l]	0.120 (0.040)	2.992	.003	Accepted
Hypothesis 8b: UI <-PCOR	0.007 (0.037)	0.190	.85	Rejected
Hypothesis 9: PU <-PEOU	0.511 (0.038)	13.319	<.001	Accepted
Hypothesis 10: UI <-PEOU	0.287 (0.042)	6.896	<.001	Accepted
Hypothesis 11: UI <-PU	0.649 (0.049)	13.281	<.001	Accepted

[a]CR: critical ratio.
[b]PEOU: perceived ease of use.
[c]FNR: financial risk.
[d]UI: use intention.
[e]PFR: performance risk.
[f]PU: perceived usefulness.
[g]PLSR: psychological and social risks.
[h]PSR: physical risk.
[i]PRR: privacy risk.
[j]PVR: provider risk.
[k]TMR: time risk.
[l]PCOR: perceived COVID-19 infection risk.

XSL•FO
RenderX

Figure 5. Results of hypotheses testing. FNR: financial risk; PFR: performance risk; PLSR: psychological and social risks; PSR: physical risk; PRR: privacy risk; PVR: provider risk. *P<.05, ***P<.001.

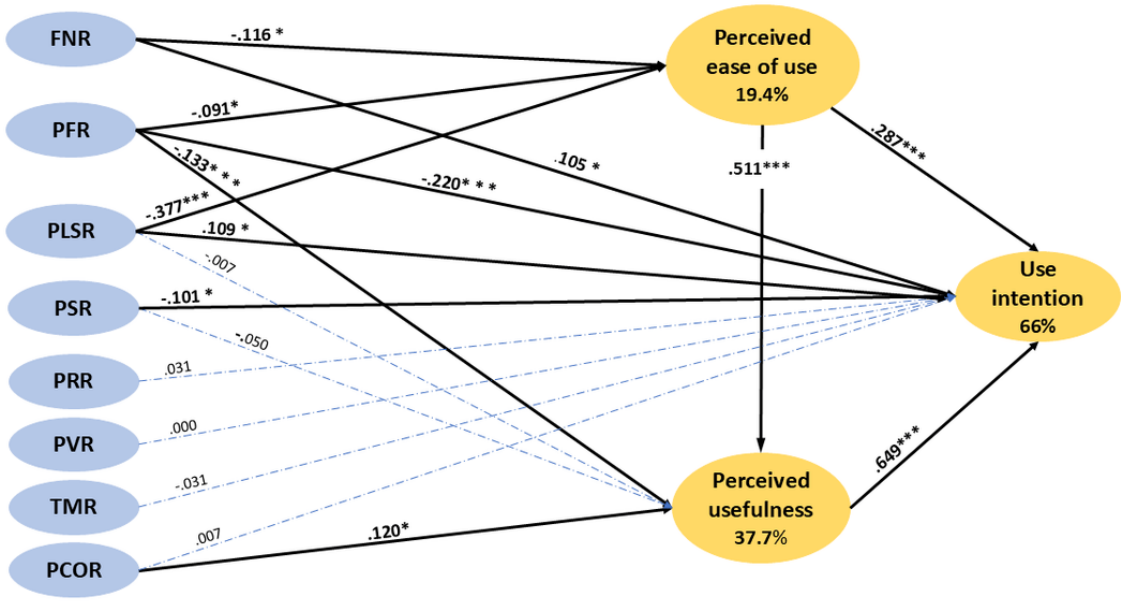

Table 10. Direct effects, indirect effects, and total effects.

	PFR[a]	PLSR[b]	PSR[c]	PCOR[d]	FNR[e]	PRR[f]	PVR[g]	TMR[h]	PEOU[i]	PU[j]
PEOU										
Direct effect	−0.091	−0.377	__[k]	—	−0.116	—	—	—	—	—
Indirect effect	—	—	—	—	—	—	—	—	—	—
Total effect	−0.091	−0.377	—	—	−0.116	—	—	—	—	—
PU										
Direct effect	−0.133	−0.007	−0.050	0.120	—	—	—	—	0.511	—
Indirect effect	−0.046	−0.192	−0.048	—	−0.059	—	—	—	—	—
Total effect	−0.179	−0.199	−0.098	0.120	−0.059	—	—	—	0.511	—
UI										
Direct effect	−0.220	0.109	−0.101	0.007	0.105	0.031	—	−0.031	0.287	0.649
Indirect effect	−0.142	−0.237	−0.033	0.078	−0.072	—	—	—	0.331	—
Total effect	−0.362	−0.128	−0.134	0.085	0.033	0.031	—	−0.031	0.618	0.649

[a]PFR: performance risk.
[b]PLSR: psychological and social risks.
[c]PSR: physical risk.
[d]PCOR: perceived COVID-19 infection risk.
[e]FNR: financial risk.
[f]PRR: privacy risk.
[g]PVR: provider risk.
[h]TMR: time risk.
[i]PEOU: perceived ease of use.
[j]PU: perceived usefulness.
[k]Not available.

XSL•FO
RenderX

Discussion

Principal Findings

The study developed a theoretical model based on TAM to determine the barriers to UI in telemedicine services. By reconstructing the risk dimensions, 8 facets of perceived risk were extracted through EFA, and 5 facets of perceived risk were considered to substantially affect the TAM of telemedicine services during the COVID-19 pandemic in Taiwan. Three risk dimensions have been adjusted according to the results of EFA. Perceived psychological risk and social risk have been categorized into psychological and social risk, mainly due to their high degree of overlap. Social risk emphasizes the potential for image damage due to lack of recognition by others, while perceived psychological risk highlights the potential for inconsistency with self-identity. Social perceptions may also influence one's self-image, as well as attitudes toward and anxiety about telemedicine. Technology risk, which is not traditionally categorized under risk perception, was included in previous studies primarily from a technological perspective to assess the impact of risks. This inclusion leads to some overlap with the original risk dimensions. For instance, survey questions asked users whether they were concerned about the poor performance of telemedicine due to various technological limitations (such as network, equipment, and physical examination). Although these are technological limitations, they fundamentally address concerns about performance. This was the purpose of EFA—to reconstruct 8 effective types of risk.

The psychological and social risk seems to be the strongest barrier to perceived risk affecting the ease of using telemedicine. Our study supports the effect of technology anxiety and pressure, which have been shown to affect users' resistance to using new technology [10]. Technology anxiety and uncertainty can lead to resistance to change owing to unexpected errors in technology [49]. Therefore, when users are faced with a new telemedicine service, they may be reluctant to switch from a face-to-face service to an alternative one. In addition, our results agree with the past finding that psychological and social risk prevents not only providers but also users from adopting technology [11]. Users care about social pressure from their friends, family members, and work groups with regard to telemedicine service.

According to the results of the total effect, performance risk is the most substantial risk factor affecting UI when Taiwanese individuals consider the adoption of telemedicine. In addition, performance risk was the only risk factor that had a negative effect both on UI directly and on PEOU and PU at the same time. This indicates that PEOU and PU are the mediators that partially explain the association between performance risk and UI. Concerns about the performance, technology, and quality of service have been expressed by stakeholders. For both health care providers and users, performance always plays a vital role. A large number of medical studies on the effectiveness of telemedicine have compared it with traditional in-person visits with regard to different diseases, such as anxiety disorders [2] and chronic kidney diseases [86]. Variations of quality and effectiveness in telemedicine and in-person care may lead to performance risk. Another factor leading to performance risk is distrust and concern regarding the technology. As shown in the previous research, the result provides more empirically grounded evidence for the importance of technology concerns perhaps caused by inadequate communication effects between physicians and patients and by disruptions in communication owing to an unstable broadband connection, which may alter the accuracy of instructed physical examination results and even cause medical error. Therefore, the "digital divide" means that telemedicine service cannot completely substitute for face-to-face visits. Indeed, a substantial concern shared by both physicians and users is the inability to conduct an in-person physical examination [87]. In the study, the item technology risk 3, "I fear the absence of physical examinations during telemedicine services," was classified as a factor related to performance risk during the EFA. This indicates that individuals not only have concerns about the limitations of the technology itself but also about the impact it may have and the resulting performance outcomes. As Nguyen et al [42] mentioned in the review, promoting realistic expectations before the telemedicine visit takes place is essential to effectively improve patient satisfaction [42]. Providers should try to increase technology acceptance and a sense of trust by assuring the users of various technological solutions to reduce the risk or by showing evidence for the similarity between technological use and in-person examination components through virtual assessments [88].

We found that the perceived COVID-19 infection risk has a positive effect on PU with P value of <.001. This is because telemedicine was globally used by force of circumstance and because of social distancing as a substitute for face-to-face interactions in enhancing clinical care, health promotion, and disease prevention. The substantial negative effect of financial risk on the PEOU of telemedicine is notable. In this context, financial risk refers to payment models, personal and governance insurance coverage, and online payment systems. One possible reason for the impact of financial risk could be the general public's understanding of Taiwan's nationalized health insurance program, which may not be fully aware of the extent and scope of insurance coverage for telemedicine services. In addition, users may have concerns about online transactions due to their perception of financial loss risks. Studies on online payment highlight the importance of website security signs and quality information, as these factors can influence customers' perception of financial risk [89]. These results may underscore the importance of trust in online trading platforms and payment systems.

Three facets of perceived risk—time risk, provider risk, and privacy risk—do not appear to be statistically significant to TAM in telemedicine service. In the past, internet privacy concerns were one major barrier to the adoption of digital technologies or online services [90]. However, the absence of a substantial effect of privacy risk on UI in telemedicine in our hypotheses agrees with the findings of the study by Kamal et al [10], in which perceived privacy risk was found to be an insignificant determinant of UI to use telemedicine services. Moreover, as shown by studies among 1059 US residents, just like environmental risks similar to the COVID-19 pandemic, privacy concerns do not negatively impact UI. Another

XSL•FO
RenderX

cross-sectional study in Germany also revealed that data privacy and security concerns were seen as less important in case of an emergency condition [44]. One possible explanation by Kato-Lin and Thelen [54] is that people may be involved in a different decision process when they are facing immediate and severe risks. Another explanation may be the result of a tradeoff during the crisis between the privacy concern and the perceived risk of COVID-19 infection.

In general, telemedicine services were expected to reduce the need for transportation, save travel time, and provide cost-saving for the public [87,91]. However, it is worth noting that the health care system in Taiwan is characterized by good accessibility, comprehensive population coverage, and short waiting times [92]. Compared to the study by Weibel et al found that patients using telemedicine saved an average of 438 driving miles per visit, the study also highlighted the potential risks related to time and distance [42]. In Taiwan, seeking medical services is extremely convenient, with shorter time and distance required for medical visits. These advantages make it less likely for the public to experience time risk.

Regarding provider risk, previous studies found that most patients prefer to use telemedicine with their physician with whom they have an established relationship [58]. One possible explanation for the insignificant hypothesis is that most users in Taiwan primarily used telemedicine services for COVID-19 infection diagnosis and treatment during that period. This is different from the context of general chronic disease management, where the importance of the service provider is comparatively reduced. Therefore, it is the result of tradeoff among provider risk, government policy, and perceived of COVID-19 infection. Consistent with previous research, our study found that PEOU positively influences PU, and both PU and PEOU positively impact UI. However, PU exerts a stronger influence on UI [17].

As >80% of respondents had no prior telemedicine experience, the findings are particularly relevant to nonadopters and may encourage them to consider using telemedicine. Risk perception is a dynamic and continually adjusting process. Nonadopters often perceive increased uncertainty and risks due to their lack of experience with technology, driven by unfamiliarity, fear of the unknown, and privacy concerns [93]. Therefore, these perceived risks have a greater potential for mitigation through awareness campaigns and other adjustments.

Implications and Limitations

In the promotion of telemedicine, addressing psychological and social risk is crucial in influencing user behavior. To alleviate technology anxiety and discomfort, health care providers should prioritize the establishment of emotional bonds and genuine care, particularly when catering to older adult users engaging in mobile health services [94]. Demonstrating empathy and inclusiveness toward users, particularly those experiencing anxiety and discomfort, is essential. Swiftly resolving issues and offering efficient technical support, particularly for new services, can effectively reduce user concerns and mitigate risks associated with social origins. Encouraging users to share positive experiences contributes to diminishing social risk and generating excitement among potential users.

Regarding the main barrier to adoption, performance risk, which may stem from the public's unfamiliarity and lack of knowledge about telemedicine, as well as the unknowns and misunderstandings surrounding its performance. According to consistent evidence from systematic reviews, the quality of care in telemedicine is comparable to, or even superior to, face-to-face consultations [2]. Governments and health care providers play a critical role in informing individuals about this evidence to alleviate concerns of potential users. Close monitoring of telemedicine service quality and providing additional education tailored to underserved users can substantially enhance adoption rates.

There are some limitations to this study. First, the web-based survey using the platform SurveyCake has a methodological limitation. The sample may produce a bias toward individuals more likely to be on the web and have internet access. Second, we focused on users in Taiwan、 and our result for perceived risk may not be representative of other countries. Furthermore, at the time of the study, approximately 80% of the population in Taiwan still did not have experience with telemedicine services, which may limit the applicability of the study's findings. Third, this survey would benefit from longitudinal data, which represents only a single point in time and does not assess trends or tradeoffs in patients' perceived risk over time. In future studies, an examination of user's perceived risk at other time points will help understand the change of tradeoffs among the risks. Second, to understand the effect of experience or age difference on users' perceived risk, future studies can recruit subgroups and compare the findings with those of our study.

Conclusions

This study aimed to enhance previous research efforts by reconstructing the facets of perceived risk at a more granular level and further refining the nomological fit of the TAM with the perceived risk dimensions. Eight effective risk dimensions were reconstructed based on a solid review of the background literature through empirical techniques. Upon analyzing the relationships between perceived risk and the TAM, we determined that performance risk was the most substantial risk factor and substantially influenced telemedicine adoption both directly and indirectly. In addition, performance risk, financial risk, and psychological and social risk substantially impacted PEOU. These findings deepen our understanding of user perceived risk and its impact on telemedicine adoption, highlighting the need for strategies to address users' uncertainties and concerns.

On the basis of the identified factors, including performance risk and psychological and social risk, telemedicine service providers, planners, and policy makers can design better strategies for the successful implementation and adoption of the services in countries that have not yet benefited from telemedicine. To decrease public anxiety and pressure, public awareness about the adoption process should be increased, and the willingness to embrace telemedicine should be instilled in the public as a convenient way to obtain high-quality health care services. Health care providers and governments must also improve the quality of technology and widely disseminate

XSL•FO
RenderX

evidence related to the effectiveness and safety of telemedicine, which will decrease performance risk and fundamentally promote the public's intention to use telemedicine services.

Authors' Contributions

TCW drafted the manuscript. TCW and CTH revised it critically for important intellectual content. All authors have read and agreed to the published version of the manuscript. This research received no external funding.

Conflicts of Interest

None declared.

References

1. Gillman-Wells CC, Sankar TK, Vadodaria S. COVID-19 reducing the risks: telemedicine is the new norm for surgical consultations and communications. Aesthetic Plast Surg 2021 Feb;45(1):343-348 [FREE Full text] [doi: 10.1007/s00266-020-01907-8] [Medline: 32885319]
2. Krzyzaniak N, Greenwood H, Scott AM, Peiris R, Cardona M, Clark J, et al. The effectiveness of telehealth versus face-to face interventions for anxiety disorders: a systematic review and meta-analysis. J Telemed Telecare 2024 Feb;30(2):250-261. [doi: 10.1177/1357633X211053738] [Medline: 34860613]
3. Heinonen K, Strandvik T. Reframing service innovation: COVID-19 as a catalyst for imposed service innovation. J Serv Manag 2020 Sep 15;32(1):101-112. [doi: 10.1108/josm-05-2020-0161]
4. Roine R, Ohinmaa A, Hailey D. Assessing telemedicine: a systematic review of the literature. CMAJ 2001 Sep 18;165(6):765-771 [FREE Full text] [Medline: 11584564]
5. Bashshur RL, Reardon TG, Shannon GW. Telemedicine: a new health care delivery system. Annu Rev Public Health 2000;21:613-637. [doi: 10.1146/annurev.publhealth.21.1.613] [Medline: 10884967]
6. Wilson LS, Maeder AJ. Recent directions in telemedicine: review of trends in research and practice. Healthc Inform Res 2015 Oct;21(4):213-222 [FREE Full text] [doi: 10.4258/hir.2015.21.4.213] [Medline: 26618026]
7. Chen HS, Guo FR, Lee RG, Lin CC, Chen JH, Chen CY, et al. Recent advances in telemedicine. J Formos Med Assoc 1999 Nov;98(11):767-772. [Medline: 10705694]
8. Chu CM, Wang TH, Lee HC, Lin CS, Feng CC. Emergency physicians' role in telemedicine care during the coronavirus disease pandemic: experiences from Taiwan. Emerg Med Australas 2021 Jun;33(3):569-571 [FREE Full text] [doi: 10.1111/1742-6723.13731] [Medline: 33474829]
9. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
10. Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technol Soc 2020 Feb;60:101212 [FREE Full text] [doi: 10.1016/j.techsoc.2019.101212]
11. Bakshi S, Tandon U. Understanding barriers of telemedicine adoption: a study in North India. Syst Res Behav Sci 2021 Jan 04;39:128-142. [doi: 10.1002/sres.2774]
12. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
13. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 1999;4(3):272-299. [doi: 10.1037//1082-989x.4.3.272]
14. Davis FD. A technology acceptance model for empirically testing new end-user information systems : theory and results. Massachusetts Institute of Technology, Sloan School of Management. 1986. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2024-12-25]
15. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
16. Venkatesh V, Davis FD. A model of the antecedents of perceived ease of use: development and test. Decis Sci 1996 Sep;27(3):451-481. [doi: 10.1111/j.1540-5915.1996.tb00860.x]
17. Su YY, Huang ST, Wu YH, Chen CM. Factors affecting patients' acceptance of and satisfaction with cloud-based telehealth for chronic disease management: a case study in the workplace. Appl Clin Inform 2020 Mar;11(2):286-294 [FREE Full text] [doi: 10.1055/s-0040-1708838] [Medline: 32294772]
18. Rho MJ, Choi IY, Lee J. Predictive factors of telemedicine service acceptance and behavioral intention of physicians. Int J Med Inform 2014 Aug;83(8):559-571. [doi: 10.1016/j.ijmedinf.2014.05.005] [Medline: 24961820]
19. Rouidi M, Elouadi A, Hamdoune A. Acceptance and use of telemedicine technology by health professionals: development of a conceptual model. Digit Health 2022;8:20552076221081693 [FREE Full text] [doi: 10.1177/20552076221081693] [Medline: 35223077]
20. Bauer RA. Consumer behavior as risk taking. In: Proceedings of the 43rd Conference of the American Marketing Association. 1960 Presented at: AMA 1960; June 15-17, 1960; Chicago, IL.

XSL•FO
RenderX

21. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum Comput Stud 2003 Oct;59(4):451-474 [FREE Full text] [doi: 10.1016/S1071-5819(03)00111-3]
22. Wang CC. Factors influencing the adoption of online group-buying in virtual community. Multim Tools Appl 2017 Feb 7;76:11751-11768. [doi: 10.1007/s11042-017-4470-y]
23. Cunningham SM. The major dimensions of perceived risk. In: Cox DF, editor. Risk Taking and Information Handling in Consumer Behavior. Cambridge, MA: Harvard University Press; 1967:82-102.
24. Featherman M, Fuller M. Applying TAM to e-services adoption: the moderating role of perceived risk. In: Proceedings of the 36th Hawaii International Conference on System Sciences. 2003 Presented at: HICSS 2003; January 6-9, 2003; Big Island, HI.
25. Luo X, Li H, Zhang J, Shim JP. Examining multi-dimensional trust and multi-faceted risk in initial acceptance of emerging technologies: an empirical study of mobile banking services. Decis Support Syst 2010 May;49(2):222-234 [FREE Full text] [doi: 10.1016/j.dss.2010.02.008]
26. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):300-302. [doi: 10.1016/s2215-0366(20)30073-0]
27. Omar NA, Nazri MA, Ali MH, Alam SS. The panic buying behavior of consumers during the COVID-19 pandemic: examining the influences of uncertainty, perceptions of severity, perceptions of scarcity, and anxiety. J Retail Consum Serv 2021 Sep;62:102600 [FREE Full text] [doi: 10.1016/j.jretconser.2021.102600]
28. Kaplan LB, Szybillo GJ, Jacoby J. Components of perceived risk in product purchase: a cross-validation. J Appl Psychol 1974 Jun;59(3):287-291. [doi: 10.1037/h0036657]
29. Rho MJ, Yoon KH, Kim HS, Choi IY. Users' perception on telemedicine service: a comparative study of public healthcare and private healthcare. Multim Tools Appl 2014 Mar 25;74:2483-2497. [doi: 10.1007/s11042-014-1966-6]
30. Kondrateva G, Baudier P, Ammi C. The digital natives' paradox: adoption of telemedicine cabin. International Marketing Trends Conference. 2024. URL: https://archives.marketing-trends-congress.com/2020/pages/PDF/079.pdf [accessed 2024-07-14]
31. Baudier P, Kondrateva G, Ammi C, Chang V, Schiavone F. Patients' perceptions of teleconsultation during COVID-19: a cross-national study. Technol Forecast Soc Change 2021 Feb;163:120510 [FREE Full text] [doi: 10.1016/j.techfore.2020.120510] [Medline: 33318716]
32. Esmaeilzadeh P, Mirzaei T. Do hospitals need to extend telehealth services? An experimental study of different telehealth modalities during the COVID-19 pandemic. Methods Inf Med 2021 Sep;60(3-04):71-83. [doi: 10.1055/s-0041-1735947] [Medline: 34598297]
33. Kaphzan H, Sarfati Noiman M, Negev M. The attitudes and perceptions of Israeli psychiatrists toward telepsychiatry and their behavioral intention to use telepsychiatry. Front Psychiatry 2022 Mar 21;13:829965 [FREE Full text] [doi: 10.3389/fpsyt.2022.829965] [Medline: 35386519]
34. Rad PM, Chiang D, Connolly J, Mahmood A. Distance healthcare in Taiwan. J Supply Chain Oper Manag 2022 Oct;20(1):75-88.
35. Xiao F, Zhu Y, Wei X. Study on the mechanism of influencing patients' willingness to use online consultation based on SOR theory. Creat Educ 2023 Apr;14(4):844-852. [doi: 10.4236/ce.2023.144055]
36. Liu C, Wang J, Chen R, Zhou W. Exploring the influence of Chinese online patient trust on telemedicine behavior: insights into perceived risk and behavior intention. Front Public Health 2024 Aug 23;12:1415889 [FREE Full text] [doi: 10.3389/fpubh.2024.1415889] [Medline: 39247232]
37. Bahari G, Mutambik I, Almuqrin A, Alharbi ZH. Trust: how it affects the use of telemedicine in improving access to assistive technology to enhance healthcare services. Risk Manag Healthc Policy 2024 Jul 24;17:1859-1873 [FREE Full text] [doi: 10.2147/RMHP.S469324] [Medline: 39072188]
38. Singh G, Bhatt S, Jhamb D. Impact of privacy, technology readiness, and perceived crowding on adoption of telemedicine services. Int J Electron Mark Retail 2024 Jul 02;15(4):455-478. [doi: 10.1504/ijemr.2024.139380]
39. Bajowala SS, Milosch J, Bansal C. Telemedicine pays: billing and coding update. Curr Allergy Asthma Rep 2020 Jul 27;20(10):60 [FREE Full text] [doi: 10.1007/s11882-020-00956-y] [Medline: 32715353]
40. Topacan U, Basoglu N, Daim T. Health information service adoption: case of telemedicine. In: Proceedings of the 42nd Hawaii International Conference on System Sciences. 2009 Presented at: HICSS 2009; January 5-8, 2009; Waikoloa, HI. [doi: 10.1109/hicss.2009.237]
41. Crosby LA, Evans KR, Cowles D. Relationship quality in services selling: an interpersonal influence perspective. J Mark 1990 Jul;54(3):68-81. [doi: 10.1177/002224299005400306]
42. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72. [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
43. Sloan M, Lever E, Harwood R, Gordon C, Wincup C, Blane M, et al. Telemedicine in rheumatology: a mixed methods study exploring acceptability, preferences and experiences among patients and clinicians. Rheumatology (Oxford) 2022 May 30;61(6):2262-2274 [FREE Full text] [doi: 10.1093/rheumatology/keab796] [Medline: 34698822]

XSL•FO
RenderX

44. Weißenfeld MM, Goetz K, Steinhäuser J. Facilitators and barriers for the implementation of telemedicine from a local government point of view - a cross-sectional survey in Germany. BMC Health Serv Res 2021 Sep 06;21(1):919 [FREE Full text] [doi: 10.1186/s12913-021-06929-9] [Medline: 34488753]
45. Freiman S, Schwabe MT, Barrack RL, Nunley RM, Clohisy JC, Lawrie CM. Telemedicine for patients undergoing arthroplasty : access, ability, and preference. Bone Joint J 2021 Jul;103-B(7 Supple B):98-102. [doi: 10.1302/0301-620X.103B7.BJJ-2020-2420.R1] [Medline: 34192903]
46. Isautier JM, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, et al. People's experiences and satisfaction with telehealth during the COVID-19 pandemic in Australia: cross-sectional survey study. J Med Internet Res 2020 Dec 10;22(12):e24531 [FREE Full text] [doi: 10.2196/24531] [Medline: 33156806]
47. Minatodani DE, Chao PJ, Berman SJ. Home telehealth: facilitators, barriers, and impact of nurse support among high-risk dialysis patients. Telemed J E Health 2013 Aug;19(8):573-578. [doi: 10.1089/tmj.2012.0201] [Medline: 23742629]
48. Yang Y, Liu Y, Li H, Yu B. Understanding perceived risks in mobile payment acceptance. Ind Manag Data Syst 2015;115(2):253-269. [doi: 10.1108/IMDS-08-2014-0243]
49. Tsai TH, Lin WY, Chang YS, Chang PC, Lee MY. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020 Jan 13;15(1):e0227270. [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
50. Xue Y, Liang H, Mbarika V, Hauser R, Schwager P, Kassa Getahun M. Investigating the resistance to telemedicine in Ethiopia. Int J Med Inform 2015 Aug;84(8):537-547 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.04.005] [Medline: 25991059]
51. Lee MC. Factors influencing the adoption of internet banking: an integration of TAM and TPB with perceived risk and perceived benefit. Electron Commer Res Appl 2009;8(3):130-141. [doi: 10.1016/j.elerap.2008.11.006]
52. Pi SM, Sangruang J. The perceived risks of online shopping in Taiwan. Soc Behav Pers Int J 2011 Mar 01;39(2):275-286. [doi: 10.2224/sbp.2011.39.2.275]
53. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
54. Kato-Lin YC, Thelen ST. Privacy concerns and continued use intention of telemedicine during COVID-19. Telemed J E Health 2022 Oct 01;28(10):1440-1448. [doi: 10.1089/tmj.2021.0603] [Medline: 35235434]
55. Nittari G, Khuman R, Baldoni S, Pallotta G, Battineni G, Sirignano A, et al. Telemedicine practice: review of the current ethical and legal challenges. Telemed J E Health 2020 Dec;26(12):1427-1437 [FREE Full text] [doi: 10.1089/tmj.2019.0158] [Medline: 32049608]
56. Jalali MS, Landman A, Gordon WJ. Telemedicine, privacy, and information security in the age of COVID-19. J Am Med Inform Assoc 2021 Mar 01;28(3):671-672 [FREE Full text] [doi: 10.1093/jamia/ocaa310] [Medline: 33325533]
57. Bali S, Gupta A, Khan A, Pakhare A. Evaluation of telemedicine centres in Madhya Pradesh, Central India. J Telemed Telecare 2015 Jul 08;22(3):183-188. [doi: 10.1177/1357633x15593450]
58. Welch BM, Harvey J, O'Connell NS, McElligott JT. Patient preferences for direct-to-consumer telemedicine services: a nationwide survey. BMC Health Serv Res 2017 Nov 28;17(1):784 [FREE Full text] [doi: 10.1186/s12913-017-2744-8] [Medline: 29183372]
59. Benis A, Banker M, Pinkasovich D, Kirin M, Yoshai BE, Benchoam-Ravid R, et al. Reasons for utilizing telemedicine during and after the COVID-19 pandemic: an internet-based international study. J Clin Med 2021 Nov 25;10(23):5519 [FREE Full text] [doi: 10.3390/jcm10235519] [Medline: 34884221]
60. Vosburg RW, Robinson KA. Telemedicine in primary care during the COVID-19 pandemic: provider and patient satisfaction examined. Telemed J E Health 2022 Feb;28(2):167-175. [doi: 10.1089/tmj.2021.0174] [Medline: 33999740]
61. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
62. Human research related business. Ministry of Health and Welfare. URL: https://statics.teams.cdn.office.net/evergreen-assets/safelinks/1/atp-safelinks.html [accessed 2024-12-25]
63. Lowry PB, Gaskin J. Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: when to choose it and how to use it. IEEE Trans Profess Commun 2014 Jun;57(2):123-146. [doi: 10.1109/tpc.2014.2312452]
64. Watkins MW. Exploratory factor analysis: a guide to best practice. J Black Psychol 2018 Apr 27;44(3):219-246. [doi: 10.1177/0095798418771807]
65. Bartlett MS. A note on the multiplying factors for various $\chi 2$ approximations. J R Stat Soc Series B Methodol 1954;16(2):296-298. [doi: 10.1111/j.2517-6161.1954.tb00174.x]
66. Kaiser HF. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36. [doi: 10.1007/bf02291575]
67. Podsakoff PM, Organ DW. Self-reports in organizational research: problems and prospects. J Manage 1986 Dec 01;12(4):531-544. [doi: 10.1177/014920638601200408]
68. Kock N. Common method bias in PLS-SEM: a full collinearity assessment approach. Int J e-Collaboration 2015 Oct;11(4):1-10. [doi: 10.4018/ijec.2015100101]

XSL•FO
RenderX

69. Carlson KD, Herdman AO. Understanding the impact of convergent validity on research results. Organ Res Methods 2010 Dec 30;15(1):17-32. [doi: 10.1177/1094428110392383]
70. Cheung GW, Wang C. Current approaches for assessing convergent and discriminant validity with SEM: issues and solutions. In: Proceedings of the Academy of Management Annual Meeting. 2017 Presented at: AMP 2017; August 4-9, 2017; Atlanta, GA. [doi: 10.5465/ambpp.2017.12706abstract]
71. Chin WW. The partial least squares approach for structural equation modeling. In: Marcoulides GA, editor. Modern Methods for Business Research. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 1998:295-336.
72. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 1981 Aug 01;18(3):382-388. [doi: 10.1177/002224378101800313]
73. Alavi M, Visentin DC, Thapa DK, Hunt GE, Watson R, Cleary M. Chi-square for model fit in confirmatory factor analysis. J Adv Nurs 2020 Sep 07;76(9):2209-2211. [doi: 10.1111/jan.14399] [Medline: 32323338]
74. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980 Nov;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
75. Snijders TA, Bosker RE. A Beginner's Guide to Structural Equation Modeling. London, UK: Routledge; 2022.
76. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2014 Dec 08;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
77. Bakshi S, Tandon U, Mittal. A. Drivers and barriers of telemedicine in india: seeking a new paradigm. J Comput Theory Nanosci 2019 Oct 01;16(10):4367-4373. [doi: 10.1166/jctn.2019.8527]
78. Jacoby J, Kaplan LB. The components of perceived risk. In: Proceedings of the Third Annual Convention of the Association for Consumer Research. 1972 Presented at: ACR 1972; November 2-5, 1972; Chicago, IL.
79. Crespo Á, del Bosque IR, de los Salmones Sánchez MM. The influence of perceived risk on internet shopping behavior: a multidimensional perspective. J Risk Res 2009 Mar;12(2):259-277. [doi: 10.1080/13669870802497744]
80. Luo L, Zheng Z, Luo J, Jia Y, Zhang Q, Wu C, et al. Spatial agglomeration of manufacturing in the Wuhan metropolitan area: an analysis of sectoral patterns and determinants. Sustainability 2020 Sep 28;12(19):8005. [doi: 10.3390/su12198005]
81. Lim N. Consumers' perceived risk: sources versus consequences. Electron Commerce Res Appl 2003 Sep;2(3):216-228 [FREE Full text] [doi: 10.1016/S1567-4223(03)00025-5]
82. Botzen WJ, Duijndam SJ, Robinson PJ, van Beukering P. Behavioral biases and heuristics in perceptions of COVID-19 risks and prevention decisions. Risk Anal 2022 Dec 29;42(12):2671-2690. [doi: 10.1111/risa.13882] [Medline: 35092967]
83. Aggelidis VP, Chatzoglou PD. Using a modified technology acceptance model in hospitals. Int J Med Inform 2009 Feb;78(2):115-126 [FREE Full text] [doi: 10.1016/j.ijmedinf.2008.06.006] [Medline: 18675583]
84. Doll WJ, Xia W, Torkzadeh G. A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Q 1994 Dec;18(4):453-461. [doi: 10.2307/249524]
85. Hair JFJ, Ringle CM, Sarstedt M. Partial least squares structural equation modeling: rigorous applications, better results and higher acceptance. Long Range Plan 2013 Feb;46(1-2):1-12. [doi: 10.1016/j.lrp.2013.01.001]
86. Androga LA, Amundson RH, Hickson LJ, Thorsteinsdottir B, Garovic VD, Manohar S, et al. Telehealth versus face-to-face visits: a comprehensive outpatient perspective-based cohort study of patients with kidney disease. PLoS One 2022 Mar 11;17(3):e0265073 [FREE Full text] [doi: 10.1371/journal.pone.0265073] [Medline: 35275958]
87. Saljoughian M. The benefits and limitations of telehealth. US Pharm 2021;46(8):5-8.
88. Lu AD, Veet CA, Aljundi O, Whitaker E, Smith WB2, Smith JE. A systematic review of physical examination components adapted for telemedicine. Telemed J E Health 2022 Dec 01;28(12):1764-1785. [doi: 10.1089/tmj.2021.0602] [Medline: 35363573]
89. Haddad GE, Aïmeur E, Hage H. Understanding trust, privacy and financial fears in online payment. In: Proceedings of the 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering. 2018 Presented at: TrustCom/BigDataSE 2018; August 1-3, 2018; New York, NY. [doi: 10.1109/trustcom/bigdatase.2018.00015]
90. Bélanger F, Crossler RE. Privacy in the digital age: a review of information privacy research in information systems. MIS Q 2011 Dec;35(4):1017. [doi: 10.2307/41409971]
91. Bali S. Barriers to development of telemedicine in developing countries. In: Heston TF, editor. Telehealth. London, UK: IntechOpen; 2019.
92. Wu TY, Majeed A, Kuo KN. An overview of the healthcare system in Taiwan. London J Prim Care (Abingdon) 2010 Dec;3(2):115-119 [FREE Full text] [doi: 10.1080/17571472.2010.11493315] [Medline: 25949636]
93. Cocosila M, Turel O. Adoption and non-adoption motivational risk beliefs in the use of mobile services for health promotion. Internet Res 2019 Aug 05;29(4):846-869. [doi: 10.1108/intr-04-2018-0174]
94. Al Meslamani AZ, Aldulaymi R, El Sharu H, Alwarawrah Z, Ibrahim OM, Al Mazrouei N. The patterns and determinants of telemedicine use during the COVID-19 crisis: a nationwide study. J Am Pharm Assoc (2003) 2022;62(6):1778-1785 [FREE Full text] [doi: 10.1016/j.japh.2022.05.020] [Medline: 35710898]

XSL•FO
RenderX

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CMB: common method bias
EFA: exploratory factor analysis
GFI: goodness of fit index
PEOU: perceived ease of use
PU: perceived usefulness
SEM: structural equation modeling
TAM: technology acceptance model
UI: use intention

Edited by A Coristine; submitted 03.10.23; peer-reviewed by A Gangadhara Rao, Z Yu; comments to author 06.04.24; revised version received 25.07.24; accepted 25.11.24; published 03.02.25.

Please cite as:
Wu TC, Ho CT
Reconstructing Risk Dimensions in Telemedicine: Investigating Technology Adoption and Barriers During the COVID-19 Pandemic in Taiwan
J Med Internet Res 2025;27:e53306
URL: https://www.jmir.org/2025/1/e53306
doi:10.2196/53306
PMID:

©Tzu-Chi Wu, Chien-Ta Ho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Long-Term Monitoring of Individuals With Chronic Obstructive Pulmonary Disease Using Digital Health Technology: Qualitative Study

Shih-Ying Chien[1,2,3], PhD; Han-Chung Hu[4,5], MD; Hsiu-Ying Cho[5], MBA

[1]Department of Industrial Design, Chang Gung University, Taoyuan, Taiwan

[2]Department of Public Health & Medical Humanities, School of Medicine, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

[3]Department of Physical Medicine and Rehabilitation, Linkou Chang Gung Memorial Hospital (CGMH), Taoyuan, Taiwan

[4]Department of Thoracic Medicine, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

[5]Department of Respiratory Therapy, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan

Corresponding Author:
Shih-Ying Chien, PhD
Department of Industrial Design
Chang Gung University
No 259, Wen-Hwa 1st Road, Kweishan
Taoyuan, 333
Taiwan
Phone: 886 32118800 ext 5669
Email: shihying@mail.cgu.edu.tw

Abstract

Background: Digital health adoption in clinical practice has been widespread, yet there remains further potential for optimizing care specifically for chronic obstructive pulmonary disease (COPD). This study therefore conducted qualitative research involving 35 health care professionals from a range of hospitals in Taiwan.

Objective: This study aims to investigate barriers and facilitators related to the implementation of digital health technology (DHT) in the long-term monitoring of individuals with COPD based on clinical experiences in Taiwan. The perspectives of Taiwanese health care professionals provided valuable insights into the challenges and opportunities associated with using DHT for the management and enhancement of respiratory rehabilitation and long-term monitoring of patients with COPD.

Methods: Several key themes related to the development of DHT were identified. Barriers encompassed concerns pertaining to digital safety, insurance coverage, constraints related to medical resources, and the presence of a digital divide. Facilitators included the potential for cost reduction, personalized prescriptions, and instilling motivation in users.

Results: To enhance the acceptance and use of DHT, embracing a user-centered approach that prioritizes the distinct needs of all parties involved is recommended. Moreover, optimizing and leveraging the effective use of DHT in managing the health of individuals with COPD promises to deliver care characterized by greater precision and efficiency.

Conclusions: Overall, the benefits of using DHT for the long-term care of patients with COPD outweigh the disadvantages. After the COVID-19 pandemic, there has been an increased emphasis in Taiwan on the effectiveness of DHT in managing chronic diseases. Relevant studies including this paper have suggested that web-based exercise management systems could benefit patients with COPD in rehabilitation and tracking. Our findings provide meaningful directions for future research endeavors and practical implementation. By addressing identified barriers and capitalizing on facilitators, advancements can be made in the development and use of DHT, especially in overcoming challenges such as information security and operational methods. The implementation of the recommended strategies will likely lead to improved COPD care outcomes.

(J Med Internet Res 2025;27:e63660) doi:10.2196/63660

KEYWORDS

digital health; chronic obstructive pulmonary disease; digital health technology; barriers and facilitators; Taiwan

XSL•FO
RenderX

Introduction

Chronic obstructive pulmonary disease (COPD) is a major public health problem and is the third-leading cause of morbidity and mortality by disease worldwide [1,2]. Indeed, over 328 million people were diagnosed with this disease in 2019 [2,3]. Caused by smoking and air pollution [4,5], COPD is typically associated with reduced physical, emotional, and social functioning, resulting in a decline in overall quality of life [4,5]. Health care professionals have reported that physical activity and effective self-management are strongly correlated with improved lung function [6,7], reduced exacerbations, and significant improvements in the quality of life for hospitalized and discharged patients [7,8]. However, decreased exercise tolerance is linked with dyspnea, which is the primary symptom experienced by individuals with COPD [9-11]. Consequently, this symptom often leads to poor adherence to self-care management and exercise in patients with COPD [12]. Research has shown that lower levels of physical activity are associated not only with increased hospitalization but also with an elevated risk of premature death [13,14]. Thus, respiratory rehabilitation plays a vital role in improving lung function and aiding recovery in patients with lung damage or injury [15].

To enhance adherence to exercise and facilitate health management in patients with COPD, the implementation of digital health (DH) has been suggested as an innovative approach [16,17]. DH models offer opportunities to optimize COPD care by leveraging accumulated big data to assess individual exercise capacity and provide personalized exercise prescriptions, thereby enhancing treatment efficacy [18]. DH primarily involves the application of information and communications technologies, leading to the emergence of concepts such as mHealth or eHealth [19,20]. Digital health technology (DHT) refers to technologies enabling the remote monitoring of physiological markers such as oxygen saturation and the self-monitoring of physical activity using wearable activity trackers. It also encompasses the use of web-based apps to support disease management, facilitate behavioral change, provide medication reminders, and predict personalized exercise prescriptions.

Compared to alternative therapies, digital pulmonary rehabilitation is a cost-effective and beneficial intervention that encompasses various components, including diagnosis, training, and evaluation [21-23]. This comprehensive approach effectively delays the deterioration of lung function and slows the progression of the disease [24,25]. Moreover, digital pulmonary rehabilitation empowers patients to have real-time control over their condition [26,27]. Furthermore, it benefits health care professionals by enabling the collection of longitudinal health data from patients, thereby facilitating informed decision-making and the provision of personalized care [28,29].

However, systematic reviews have found no significant evidence demonstrating the effective impact of digital pulmonary rehabilitation on COPD management [30,31]. This could be attributed to the advanced age or limited motivation for exercise in the majority of patients with COPD, who often have impaired lung function. Additionally, their limited familiarity with digital spaces, devices, and knowledge further contributes to this issue [32,33]. Notably, prior exploration of the needs and preferences of health care professionals and patients as well as careful consideration of intervention design are essential factors that have rarely been addressed in the implementation of digital pulmonary rehabilitation for patients with COPD [34,35]. Failure to consider the needs of health care professionals and patients from a user-centric perspective may lead to user experience issues, such as lack of clinical use, disruption of patients' workflow, and negative engagement experiences with DHT [36,37]. Consequently, users may become unwilling to continue using and adhering to treatment regimens [38-40].

Prior to conducting the primary survey in this study, we conducted informal interviews with health care professionals within the hospital to gather their perspectives on using DHT for managing remote pulmonary rehabilitation. Although the sample size was small (n=10), the majority of health care professionals emphasized that remote rehabilitation is an important development for the future. This was particularly evident during the COVID-19 outbreak when many hospitals faced closures, preventing patients from continuing their rehabilitation programs and hindering their overall recovery progress. Most Health care professionals believed that implementing DH models for respiratory rehabilitation was necessary. However, both health care professionals and patients had limited knowledge about DH. Moreover, despite the advancements in medical technology in Taiwan, the application of DH in the care of patients with COPD remains subject to various barriers, such as information security measures, disparities between urban and rural areas, and the acceptance of digital tools among professional caregivers.

This study explores a broader range of barriers to using DHT in respiratory rehabilitation from the perspective of health care professionals. We also assessed the feasibility of implementing DHT in the management of COPD. Given Taiwan's relatively rich medical resources and its urban-rural disparities, the findings from this research serve as a valuable reference for similar implementations worldwide. This study not only contributes to optimizing clinical rehabilitation designs of DH from a user perspective but also significantly enhances the benefits for patients throughout the treatment process.

Methods

Study Design

A qualitative descriptive methodology was used to investigate the barriers encountered when using DH for managing pulmonary rehabilitation among patients with COPD. A snowball sampling technique was used to recruit 35 diverse health care professionals in this study. The participants included 12 respiratory physiotherapists, 8 respiratory specialist nurses, 5 rehabilitation physicians, 2 thoracic surgeons, 5 physical therapists, and 3 sports medicine specialists. The interviewees were representatives in the field of cardiopulmonary rehabilitation in Taiwan, with an average age of 52 (SD 8) years and a minimum of 6.3 (SD 7.4) years of relevant work experience.

XSL•FO
RenderX

Ethical Considerations

To investigate the key elements of pulmonary rehabilitation for patients across different regions, this study was conducted in the Department of Physical Medicine and Rehabilitation across 6 affiliated hospitals situated in different regions of Taiwan. The research protocol received approval from the Research Ethics Board of Chang Gung Hospital (202200070B0). All participants possessed experience in providing care for patients with COPD. Before the commencement of interviews, participants were provided with physical copies of consent documents. All individuals who participated in the study provided informed written consent. Among these consents, 1 copy was retained by the researcher, while another was retained by the participant. Subsequently, semistructured one-on-one telephone interviews were conducted, each lasting approximately 45 to 60 minutes. The interview data were processed with unique identifiers to ensure participants' rights. All participants voluntarily participated in the study without receiving any compensation, ensuring the fairness of the experiment.

Procedure

The study was conducted in June 2022 and completed in May 2023. All interviews were conducted by the first author (SYC), who is an experienced qualitative researcher. The interviews were carried out in a one-to-one semistructured format. An interview topic guide in Textbox 1, outlined in Tables 1 and 2, was used as a framework for the interviews. The initial version of the topic guide was developed based on relevant literature and consultations with respiratory therapists and rehabilitation physicians. The interviews were designed to gather information through semistructured and open-ended questions, allowing for in-depth discussions on the interview content.

Textbox 1. Interview topic guide (including clinical, home care, and digital health aspects).

Clinical perspective

- Please outline the basic requirements for conducting pulmonary exercise training for patients with chronic obstructive pulmonary disease (COPD).
- Please describe the information and physiological parameters that need to be collected for conducting pulmonary exercise training.
- Please describe the current limitations and challenges in clinical respiratory care.

Self-care management

- Please describe health management and follow-up methods for inpatients and discharged patients.
- Please discuss the key factors influencing patients' self-management abilities.
- Please describe the difficulties of health management for patients after discharge.

Digital health

- Please describe your views on digital health (DH).
- Please describe the benefits and challenges that DH may bring to patients with COPD.
- Please describe your experiences of implementing DH in clinical practice. What are your views on incorporating DH into respiratory rehabilitation?
- Please describe the factors that need to be considered when integrating DH into the workflow of respiratory rehabilitation care.
- Please describe the potential risks of using DH technology in clinical settings.
- Please describe the indicators that clinicians wish to track for patients after discharge.

Table 1. Participant demographics and characteristics (N=35).

	Values
Sex (male:female), n:n	18:17
Age (years), range (mean, SD)	37-67 (52, 8)
Years of experience in pulmonary rehabilitation, range (mean, SD)	5-36 (6.3, 7.4)

XSL•FO
RenderX

Table 2. Health care professionals in pulmonary rehabilitation.

Type of health care professionals	Participants, n (%)
Respiratory physiotherapists	12 (34)
Respiratory specialist nurses	8 (23)
Rehabilitation physicians	5 (14)
Thoracic surgeons	2 (6)
Physical therapists	5 (14)
Sports medicine specialists	3 (9)

Data Analysis

Participants were informed that all study data would be treated anonymously. Each participant voluntarily consented to join the study, with no compensation offered during the research period. The entire interview process was audio-recorded, and unique identifiers were assigned to each participant to ensure confidentiality. The transcripts were generated based on the interview topic guide, and NVivo software (version 12.2; QSR International) was used to conduct a thematic analysis (TA) of the transcripts.

This study follows the reflexive TA method developed by Braun and Clarke [41,42], using a coding-first approach to identify themes that help understand and construct the clinical experiences of health care providers. This approach aims to authentically capture the participants' perspectives on events and uncover the meanings behind surface-level expressions. The TA consists of 6 meticulous steps, ensuring consistency, logic, and comprehensiveness in the research process. It has been demonstrated to possess confirmability, reliability, and credibility [41,42]. Moreover, this descriptive method distinctly articulates the researcher's interpretation of the data, rather than simply reiterating the participants' statements.

The analysis of the interviews followed an inductive and cyclical process. This involved a thorough reading and rereading of each transcript to gain familiarity with the data [43]. The data were then coded with unique labels, and themes and subthemes were generated based on the coded data. To ensure analytical rigor, the first author independently coded the data. Then, 2 experienced researchers were invited to review and validate the coding process.

In reference to the study by Hennink and Kaiser [44] on sample sizes for saturation in qualitative research, this research also considers the work of Rowlands et al [45], which proposes calculating thematic saturation using a lognormal distribution with a selected confidence level. In this study, the data obtained from 35 participants is expected to hold significant reference value, aligning with these methodological guidelines [44,45].

Results

Overview

After nearly a year of duration, this study completed interviews with 35 participants. Based on the interview records, an analysis was conducted, resulting in the generation of themes and subthemes for this study. The analysis has identified several themes: digital safety and reliability, insurance coverage and safety concerns, medical resource constraints, and the digital divide, which are presented as barriers. Additionally, the themes of the advantages of using DHT, reduced health care expenditure, personalized prescribing, and motivation for use are presented as facilitators.

Barriers

Digital Safety and Reliability

The majority of the interviewees (n=21, 60%) expressed concerns about the use of DHT in health care services, particularly in relation to information security and patient privacy. As the primary objective of health care is to ensure patient safety, any potential breach of privacy resulting from seeking medical treatment is considered harmful to patients. Therefore, information security in the health care field is a highly sensitive issue for patients. The promotion of DHT in the health care industry is critically dependent on network security and data protection.

Furthermore, more than half of the interviewees (n=19, 54%) had concerns about the accuracy of the generated data. Inaccurate data have the potential to lead to erroneous medical decisions and in severe cases to even jeopardize the health and safety of patients.

> *Patient information is highly sensitive, and any data breaches or information errors in a hospital can result in significant losses. While I acknowledge the many benefits that DHT can bring, I believe that ensuring information security and data accuracy should be a minimum requirement.* [Thoracic surgeon 26]

> *Technology has been advancing rapidly, especially after the pandemic. Currently, almost every industry relies on the internet to handle various aspects of daily life. However, the healthcare field is different. There are two main concerns. First, there is a fear of data breaches and the potential leakage of sensitive information. Second, there is a worry that if the machines make mistakes, it could jeopardize the safety and well-being of patients.* [Respiratory physiotherapist 3]

> *Regardless of everything, patient safety always comes first for healthcare professionals. Digital health is just starting to thrive, and it seems to bring much convenience. However, I believe there is still room for improvement. For example, how do you prove that it is accurate?* [Thoracic surgeon 27]

XSL•FO
RenderX

The interviewees highlighted the issue of low patient adherence to medical instructions, particularly among patients with COPD, which encompasses various complex aspects such as medication adherence, exercise control, and smoking cessation. Most interviewees emphasized that patients with COPD tend to dislike exercise due to the physical burden it imposes (n=26, 74%), as they believe it may cause discomfort and even pose a threat to their livelihood. Consequently, after being discharged from the hospital, patients often struggle to comply with basic exercise instructions, resulting in a decline in pulmonary function. Even with the use of optimal rehabilitation assistive devices, there remains an insufficient improvement in the pulmonary function of patients. Moreover, the interviewees (n=19, 54%) expressed concerns regarding the long-term observation and accumulation of training hours required to assess patients' recovery progress in pulmonary function training. Consequently, when using DHT, there may be concerns about the reliability of generated medical recommendations if the accumulated data are incomplete or if the duration of data collection is insufficient to provide meaningful results.

> *Patients can be categorized into inpatients and outpatients. Generally, inpatients tend to be more compliant and follow the doctor's treatment instructions. Managing the health of outpatients can be challenging because many patients become difficult to reach once they return home, and it is uncertain whether they are taking medication and engaging in physical activity as prescribed.* [Respiratory physiotherapist 1]

> *Our patients are not very keen on physical activity because it often triggers breathlessness and discomfort. This makes it challenging for them to engage in exercise. Many patients who cooperate well during their hospital stay exhibit different behaviors once they return home. Therefore, for patients with chronic conditions such as COPD, it is crucial that we have digital health technology to assist us in monitoring their condition closely.* [Respiratory physiotherapist 9]

> *After being discharged, patients often resume smoking, engage in unhealthy eating habits, and neglect physical activity due to a lack of supervision. As a result, their lung function deteriorates rapidly, leading to a cycle of readmissions until their unfortunate demise.* [Respiratory physiotherapist 6]

> *After discharge, patients often lack motivation and health awareness, which lead to poor compliance with medical instructions. Some may occasionally follow them, while others resist any form of supervision. Both types of patients are not suitable candidates for digital health technology-based healthcare prescription advice because insufficient information would be collected, which may result in inaccurate prescriptions and potential harm to their health.* [Respiratory specialist nurse 17]

Insurance Coverage and Safety Concerns

Prior to the outbreak of the COVID-19 pandemic, global insurance coverage for telemedicine was subject to stringent limitations. However, as the severity of the pandemic increased, regions including Taiwan gradually relaxed regulations surrounding telehealth. Nevertheless, participants universally emphasized that DHT continues to face multiple constraints and considerations. Specifically, not all medical conditions are suitable for DHT adoption, and a lack of sufficient information impedes health care professionals from making accurate medical decisions. Moreover, vulnerable patients, particularly elderly patients, often face potential risks of falls and emergencies.

> *In the past, telemedicine was not very successful because insurance coverage was limited. Even though there are more flexible rules now due to the pandemic, digital health still has its challenges. Some diseases cannot be diagnosed through video calls alone, and there are also concerns about risks for weaker patients, such as the possibility of accidents or falls.* [Rehabilitation physician 22]

> *I believe digital health technology can play a significant role in managing health information, while the diagnosis and treatment depend on the situation. Of course, conditions vary in different regions, so I think digital health technology should consider factors such as disease type, severity, and the availability of alternative options in diagnosis and treatment.* [Thoracic surgeon 26]

Medical Resource Constraints

The interviewees discussed the challenges of implementing DHT in different hospitals, which stem from variations in health care resources. For urban medical centers, the promotion of DHT is relatively easier compared to regional hospitals located in remote rural or island areas. This is due to the limitations in health care resources, knowledge gaps, and a shortage of health care professionals in rural areas, leading to an increasing disparity in health care between urban and rural regions. However, it is precisely because of the scarcity of health care resources in remote rural and island areas that the introduction of DHT is needed in order to bridge the gap in health care access between urban and rural areas.

> *In my opinion, not every region can implement digital health technology because conditions vary. For example, in rural areas, many places lack internet signal, making it impossible for activities such as video consultations or data transmission to take place.* [Sports medicine specialist 34]

> *Urban and rural areas have different living conditions. In rural areas, apart from limited healthcare resources, there is also a shortage of staff. Additionally, there are disparities in education and professional expertise across different regions. As a result, the acceptance and trust in digital health technology vary among healthcare professionals and patients in different areas.* [Physical therapist 31]

XSL•FO
RenderX

> *Digital health technology can be a valuable tool for rural or island areas, but before considering its use in healthcare services, it is important to assess whether digital health technology is suitable for the specific region. This involves examining what technologies can address the local challenges, improve specific issues, and enhance overall benefits.* [Respiratory specialist nurse 14]

Digital Divide

As mentioned earlier, both health care professionals and patients may face digital disparities. These disparities are more prevalent in rural or island areas due to factors such as economic conditions, educational levels, living environments, and comorbidities, including cognitive impairments, which can potentially impact the opportunities and abilities of individuals to use digital products such as computers or the internet. Consequently, there exists a significant gap in the health care capacity for medical examinations, monitoring, and care between urban and rural areas.

> *Typically, COPD patients do not just have a single condition. They may also experience physical, cognitive, or visual impairments. This means they might face difficulties with tasks such as connecting to or operating computers, or expressing themselves.* [Rehabilitation physician 21]

> *"Based on my past experience, although we use smartphones daily for information, older doctors like us may find it challenging to practice and become proficient with new IT products, let alone the patients.* [Thoracic surgeon 27]

Facilitators

Advantages of Using DHT

Interviewees emphasized the need for an adaptation period when adopting DHT (n=26, 74%), for both health care professionals and patients. They unanimously acknowledged the significant clinical benefits of using DHT to assist patients with COPD in respiratory rehabilitation. Moreover, they recognized the advantages of DHT in facilitating the recovery of emergency inpatients and monitoring discharged patients. For the majority of patients who are unable to remain hospitalized for an extended period due to insurance regulations, continuous monitoring and remote care treatment at home are essential. DHT is regarded as the optimal clinical tool choice in such cases.

> *I believe that every disease requires going through four stages: testing, treatment, monitoring, and health management. It is just a matter of the extent to which each stage is needed. COPD patients, in particular, require them because their condition does not improve, so having long-term tracking and health management tools is crucial to assist us in our clinical work.* [Rehabilitation physician 22]

> *Due to insurance regulations, most hospitalized patients are discharged within a certain number of days. If there are still relevant technologies or tools available for continuous remote monitoring or care of COPD patients after discharge, it would be a significant advancement.* [Respiratory specialist nurse 14]

Reduced Health Care Expenditure

Most interviewees noted that patients with COPD often face limitations in adhering to medical instructions (n=23, 66%), for example, taking medications and following exercise prescriptions, due to factors such as home environment, declining physical function, and transportation difficulties after discharge. The application of DHT is seen as an effective solution to mitigate these limitations.

> *Some patients face challenges in contacting or returning for follow-up care after discharge due to their residential or transportation constraints. If there is technology available to assist us in regularly monitoring discharged patients, their willingness to cooperate would likely be higher.* [Respiratory specialist nurse 20]

Personalized Prescriptions

Most participants highlighted the potential of DHT in enabling personalized prescriptions through the accumulation of data (n=24, 69%). They emphasized that the use of big data and machine learning can facilitate data processing, enabling the identification, inference, and prediction of individuals' health statuses and required rehabilitation interventions based on changes in their health behaviors.

> *To be honest, I have high expectations for big data because humans cannot remember everything, but machines can. Moreover, the larger the data are, the more accurate their judgments become. It is a great clinical tool.* [Thoracic surgeon 26]

> *I believe we should understand that artificial intelligence does not provide a definitive answer, but it is certainly the best supporting evidence for making judgments.* [Sports medicine specialist 35]

Increased Motivation to Participate in Rehabilitation Programs

The motivation of patients to use DHT will be a key determinant of the success of its implementation. Our participants discussed how issues related to hardware and software devices, technology, and operation are likely to affect user motivation. While they generally recognized the convenience and clinical benefits of DHT, many pointed out the difficulties encountered by older health care professionals when learning new technology. Thus, the interface design of DHT must be optimized for users, both patient users and health care professionals. This will increase motivation to participate in remote rehabilitation programs.

> *Learning and adapting to new tools is not easy for us or for patients, so interface design is also a crucial factor.* [Thoracic surgeon 27]

> *I believe motivation plays a significant role in many things. If digital healthcare can help alleviate our*

XSL•FO
RenderX

clinical workload, of course, we are all willing to use it, as are the patients. If they know it is for their own good, I believe they will be more than willing to give it a try. [Rehabilitation physician 23]

Discussion

Overview

This study provides insights into the perspectives of health care professionals on the use of DHT in the care of patients with COPD and explores potential issues faced by both the clinical and patient sides. The findings highlight the urgent need to promote DHT at the clinical level, particularly in the postpandemic era, to reduce interpersonal contact, interaction, and health care costs for hospitals and patients.

Findings

Previous research has often neglected the user-centered perspectives and diverse needs of stakeholders when investigating DHTs. Additionally, the demands and feasibility of DHT vary across regions due to population demographics and contextual differences. Addressing these challenges requires tailored investigations of stakeholder needs, encompassing health care providers and users, considering specific regional characteristics. Collaborative or participatory design approaches can facilitate the development of innovative DHT solutions that are adaptable to different regional contexts, including urban, rural, and island areas, with the ultimate goal of maximizing the benefits of DHTs in health care.

In summary, this study reveals that due to significant individual differences among patients with COPD, a singular treatment approach cannot be universally applicable, and a uniform health care service model cannot be implemented across regions. Respondents preferred personalized prescriptions that consider crucial factors such as patients' health conditions, geographic locations, and educational levels, thus facilitating the provision of diverse DH services.

Furthermore, the aging population and lifestyle changes in the postpandemic era have underscored the importance of pulmonary rehabilitation as a key indicator for DHT development. The findings indicate that both health care professionals and patients have unmet digital literacy needs, emphasizing the importance of considering usability when implementing DHT and constructing pulmonary rehabilitation systems. This study contributes valuable insights into the demands and preferences of health care professionals regarding DHT in the care of patients with COPD. The user-centric design approach not only benefits health care professionals in clinical care for inpatients but also expands the applications of DHT in postdischarge monitoring, observation, and ongoing health care services for patients with COPD.

Despite health care professionals having concerns about the health information obtained from patients using DHT, they still recognize numerous advantages of DHT in clinical practice. For instance, DHT can streamline repetitive clinical tasks and proactively collect patient data through recommended system and device designs, thereby reducing the chances of human errors in documentation and medical decision-making. Simultaneously, simplified hardware and software designs can facilitate the collection and management of patient exercise levels and fundamental physiological data, particularly for discharged patients. Despite the mentioned information security issues, DHT faces additional risks and limitations in clinical applications.

Health care professionals acknowledge that insurance coverage and personal safety have hindered the successful implementation of DHT for many years [19]. Prior to the pandemic, many doctors faced challenges in understanding and navigating the insurance system, resulting in unsuccessful insurance claims. Furthermore, older patients may struggle with using internet-connected devices, making remote consultations difficult. Health care professionals also consider the advanced age of patients with COPD, raising concerns about potential falls during simple rehabilitation exercises at home. Additionally, various constraints in manpower, equipment, and infrastructure within the clinical setting hinder the widespread adoption of DHT. For instance, patients with severe respiratory disease rely on large respiratory devices and specialized personnel for treatment.

Past studies have identified resource constraints as the primary hindrance to the widespread adoption of DH, mainly due to the inability to generate reliable data. However, this investigation reveals that the lack of consideration for local geographical conditions and population structures has impeded the proliferation of DH. This results in a disconnect between health care providers and patients, either due to communication barriers or disparities in digital literacy, thereby reducing acceptance. It is important to acknowledge that different regions, including urban, rural, and island areas, exhibit varying living environments, necessitating the use of different DH tools to ensure the generation of precise, reliable, and trustworthy data. Moreover, given the significant individual differences among patients with COPD, a one-size-fits-all approach is deemed inappropriate. Consequently, emphasizing the significance of personalized prescriptions, this study highlights the potential of big data in achieving tailored and accurate treatments. Finally, contrary to previous research attributing patients' inadequate self-management to a lack of motivation, this study proposes that their inability to effectively engage in health management may stem from practical constraints such as the unavailability of necessary equipment or the complexity of operations. In summary, by adopting a human-centered approach and considering the needs and circumstances of various stakeholders, the development of health care services can be optimized, ultimately enhancing patient adherence to medical instructions.

Future Directions

The survey findings of this study indicate that the majority of health care professionals perceive the potential of DHT to facilitate the care of patients with COPD. However, both health care professionals and patients require a period of time to adapt to and become familiar with the use of digital tools. The novelty of this study lies in proposing that each disease undergoes 4 stages: detection, treatment, monitoring, and health management. However, the distribution of needs across these stages varies for different diseases. For instance, in the context of COPD, a

XSL•FO
RenderX

condition that cannot be clinically cured, patients require long-term monitoring of their condition and self-health management. Therefore, DHT should be designed with individualized approaches for different diseases. Participants widely believed that DHT tools not only assist patients in understanding their own condition but also alleviate the burden of clinical work.

Previous studies have frequently cited inadequate compliance with exercise programs due to a lack of motivation and home environment factors. However, this study posits that the appropriate use of DHT can enhance patient motivation for exercise and reduce health care expenditures. By leveraging DHT, it becomes possible to synchronously execute consultations, treatments, monitoring, and overall health management for patients with COPD.

While the reliability of big data may still be questioned in many studies, the accumulation and learning capabilities of big data over the years will serve as a reference for future health care professionals when making medical decisions, given the inherent limitations of human memory and abilities. The information provided by DHT may not be the sole definitive answer, but it is certainly the best evidence supported by a scientific basis. The personalized prescriptions inferred by DHT are accumulated from individual health behavior changes, enabling the identification, inference, and prediction of individual needs for rehabilitative interventions. This will lead to more accurate health care decisions and provide an additional level of patient safety.

Finally, the issue of low patient adherence to health behaviors is often discussed in many studies. However, this problem also applies to health care professionals. Learning new things is inherently challenging for everyone, especially for older individuals. Therefore, for both patients and health care professionals, the motivation to use and accept DHT is crucial. If the use of DHT can help health care professionals reduce their clinical workload, make more accurate judgments, and lower the rate of medical decision errors, it is expected to increase the chances of clinical acceptance of DHT interventions. On the other hand, if DHT can assist patients in effectively managing and controlling their own health [20] while making the treatment process more interesting and reducing discomfort, it is believed that the majority of patients would be willing to try it. Of course, it is crucial that the devices, systems, or equipment designed for DHT adhere to design principles that are simple and intuitive and minimize errors. With this in mind, popularizing DHT should not be a challenging task.

Study Limitations

Given the global attention to patients with COPD and the increasing demand for care and management of chronic lung diseases, along with the reliance on DHT during the pandemic, particularly in the health care sector, it has been confirmed that DHT can effectively perform the 4 major clinical tasks of diagnosis, treatment, monitoring, and health management even in special circumstances. Despite the rigorous investigation conducted over the course of one year, this study still has some limitations. First, the study used a snowball sampling method, which may introduce information bias and deviation due to the potential similarity of perspectives. Second, the study overlooked the age and work experience differences among health care professionals involved in the care of patients with COPD, resulting in a broad and unfocused information scope. Finally, the research solely relied on interview-based data collection without conducting a questionnaire survey, potentially leading to unverified differences between the interview findings and questionnaire responses.

Based on the above, this study represents a novel postpandemic reevaluation of the clinical applicability of DHT for managing patients with COPD. It aims to explore the potential of using DHT in enhancing health management for patients with COPD while considering the associated obstacles and potential benefits from a clinical perspective. According to the findings of the survey, the majority of health care professionals express concerns regarding the security and reliability of data generated by DHTs. These concerns can be attributed to two primary reasons: (1) apprehension regarding potential privacy breaches for patients; and (2) the lack of a validation platform for DHT, which raises doubts about its accuracy. Many respondents also mentioned incomplete data and potential risks to patient safety due to information security issues. Despite having high expectations for DHT, there is still perceived room for improvement. Existing research suggests poor adherence among patients with chronic diseases, especially patients with COPD who may exhibit reluctance to regular physical activity due to compromised lung function and struggle to maintain complete smoking cessation due to self-discipline challenges. Moreover, they face difficulties related to the absence of effective tracking and management systems. These factors hinder patients' ability to consistently comply with data collection, thereby impacting the accuracy and reliability of health decision-making.

Overall, using DHT for the long-term monitoring of individuals with COPD offers more benefits than drawbacks. In the context of global digital transformation, medical digitalization is an inevitable trend. Therefore, limitations in equipment and urban-rural disparities will improve with advancements in information security and changes in insurance systems. Digitalization will significantly reduce overall health care costs, enhance precision, and transform personal health behaviors and motivations, ultimately achieving a patient-centered vision of health and well-being.

Acknowledgments

The authors wish to express their gratitude to the individuals who dedicated their time, invaluable experiences, and insightful perspectives to enrich the depth of this study. Furthermore, the authors' sincere appreciation is extended to the entire team of health care professionals at Chang Gung Hospital, each of whom contributed to the support of this research initiative. Notably, special recognition is reserved for the Respiratory Therapy Team, whose commitment and collaboration were integral in facilitating

XSL•FO
RenderX

the completion of this research. Their dedication and assistance in logistical arrangements have been pivotal to this study. This study was supported by the National Science and Technology Council, Taiwan, Republic of China (grant 111-2222-E-182-0002-MY2).

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to participant protection rights but are available from the corresponding author upon reasonable request.

Authors' Contributions

SYC contributed to the study design, data analysis, and drafting of the manuscript. HCH and HYC contributed to the study design.

Conflicts of Interest

None declared.

References

1. López-Campos JL, Tan W, Soriano JB. Global burden of COPD. Respirology 2016;21(1):14-23 [FREE Full text] [doi: 10.1111/resp.12660] [Medline: 26494423]
2. Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D, et al. An official American Thoracic Society public policy statement: novel risk factors and the global burden of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2010;182(5):693-718. [doi: 10.1164/rccm.200811-1757ST] [Medline: 20802169]
3. Kim EJ, Yoon SJ, Kim YE, Go DS, Jung Y. Effects of aging and smoking duration on cigarette smoke-induced COPD severity. J Korean Med Sci 2019;34:e90 [FREE Full text] [doi: 10.3346/jkms.2019.34.e90] [Medline: 30923493]
4. Raherison C, Girodet PO. Epidemiology of COPD. Eur Respir Rev 2009;18(114):213-221 [FREE Full text] [doi: 10.1183/09059180.00003609] [Medline: 20956146]
5. Lebowitz KR, Suh S, Diaz PT, Emery CF. Effects of humor and laughter on psychological functioning, quality of life, health status, and pulmonary functioning among patients with chronic obstructive pulmonary disease: a preliminary investigation. Heart Lung 2011;40(4):310-319. [doi: 10.1016/j.hrtlng.2010.07.010] [Medline: 21724041]
6. Dineen-Griffin S, Garcia-Cardenas V, Williams K, Benrimoj SI. Helping patients help themselves: a systematic review of self-management support strategies in primary health care practice. PLoS One 2019;14(8):e0220116 [FREE Full text] [doi: 10.1371/journal.pone.0220116] [Medline: 31369582]
7. Jonsdottir H. Self-management programmes for people living with chronic obstructive pulmonary disease: a call for a reconceptualisation. J Clin Nurs 2013;22(5-6):621-637. [doi: 10.1111/jocn.12100] [Medline: 23398312]
8. Wang Q, Bourbeau J. Outcomes and health-related quality of life following hospitalization for an acute exacerbation of COPD. Respirology 2005;10(3):334-340. [doi: 10.1111/j.1440-1843.2005.00718.x] [Medline: 15955146]
9. O'Donnell DE. Hyperinflation, dyspnea, and exercise intolerance in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2006;3(2):180-184. [doi: 10.1513/pats.200508-093DO] [Medline: 16565429]
10. O'Donnell D, Flüge T, Gerken F, Hamilton A, Webb K, Aguilaniu B, et al. Effects of tiotropium on lung hyperinflation, dyspnoea and exercise tolerance in COPD. Eur Respir J 2004;23(6):832-840 [FREE Full text] [doi: 10.1183/09031936.04.00116004] [Medline: 15218994]
11. Ries AL, Kaplan RM, Limberg TM, Prewitt LM. Effects of pulmonary rehabilitation on physiologic and psychosocial outcomes in patients with chronic obstructive pulmonary disease. Ann Intern Med 1995;122(11):823-832. [doi: 10.7326/0003-4819-122-11-199506010-00003] [Medline: 7741366]
12. Riegel B, Carlson B. Facilitators and barriers to heart failure self-care. Patient Educ Couns 2002;46(4):287-295. [doi: 10.1016/s0738-3991(01)00165-3] [Medline: 11932128]
13. Albarrati AM, Gale NS, Munnery MM, Cockcroft JR, Shale DJ. Daily physical activity and related risk factors in COPD. BMC Pulm Med 2020;20(1):60 [FREE Full text] [doi: 10.1186/s12890-020-1097-y] [Medline: 32138714]
14. Pitta F, Troosters T, Probst VS, Spruit MA, Decramer M, Gosselink R. Physical activity and hospitalization for exacerbation of COPD. Chest 2006;129(3):536-544. [doi: 10.1378/chest.129.3.536] [Medline: 16537849]
15. Hodgkin JE, Celli BR, Connors GL. Pulmonary Rehabilitation: Guidelines to Success. St. Louis, MO: Mosby; 2023.
16. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. Nat Rev Rheumatol 2020;16(9):525-535. [doi: 10.1038/s41584-020-0461-x] [Medline: 32709998]
17. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. A qualitative study of chronic obstructive pulmonary disease patient perceptions of the barriers and facilitators to adopting digital health technology. Digit Health 2019;5:2055207619871729 [FREE Full text] [doi: 10.1177/2055207619871729] [Medline: 31489206]
18. Gibson AL, Wagner DR, Heyward VH. Advanced Fitness Assessment and Exercise Prescription. Champaign, IL: Human kinetics; 2024.

XSL•FO
RenderX

19. Zeadally S, Isaac JT, Baig Z. Security attacks and solutions in electronic health (E-health) systems. J Med Syst 2016;40(12):263. [doi: 10.1007/s10916-016-0597-z] [Medline: 27730389]
20. Njoroge M, Zurovac D, Ogara EAA, Chuma J, Kirigia D. Assessing the feasibility of eHealth and mHealth: a systematic review and analysis of initiatives implemented in Kenya. BMC Res Notes 2017;10(1):90 [FREE Full text] [doi: 10.1186/s13104-017-2416-0] [Medline: 28183341]
21. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
22. Darmstadt GL, Bhutta ZA, Cousens S, Adam T, Walker N, de Bernis L, Lancet Neonatal Survival Steering Team. Evidence-based, cost-effective interventions: How many newborn babies can we save? Lancet 2005;365(9463):977-988. [doi: 10.1016/S0140-6736(05)71088-6] [Medline: 15767001]
23. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
24. Hill CW, Jones GR, Schilling MA. Strategic Management: Theory & Cases: An Integrated Approach. Boston, Massachusetts: Cengage Learning; 2014.
25. Black TR. Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics. Thousand Oaks, CA: Sage Publications; 2012.
26. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015;206(4):263-265 [FREE Full text] [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
27. Tu J, Torrente‐Rodríguez RM, Wang M, Gao W. The era of digital health: a review of portable and wearable affinity biosensors. Adv Funct Mater 2019;30(29):1906713 [FREE Full text] [doi: 10.1002/adfm.201906713]
28. Holzner B, Giesinger JM, Pinggera J, Zugal S, Schöpf F, Oberguggenberger AS, et al. The computer-based health evaluation software (CHES): a software for electronic patient-reported outcome monitoring. BMC Med Inform Decis Mak 2012;12:126 [FREE Full text] [doi: 10.1186/1472-6947-12-126] [Medline: 23140270]
29. Downing GJ, Boyle SN, Brinner KM, Osheroff JA. Information management to enable personalized medicine: stakeholder roles in building clinical decision support. BMC Med Inform Decis Mak 2009;9:44 [FREE Full text] [doi: 10.1186/1472-6947-9-44] [Medline: 19814826]
30. Gordon CS, Waller JW, Cook RM, Cavalera SL, Lim WT, Osadnik CR. Effect of pulmonary rehabilitation on symptoms of anxiety and depression in COPD: a systematic review and meta-analysis. Chest 2019;156(1):80-91. [doi: 10.1016/j.chest.2019.04.009] [Medline: 31034818]
31. Jones AW, Taylor A, Gowler H, O'Kelly N, Ghosh S, Bridle C. Systematic review of interventions to improve patient uptake and completion of pulmonary rehabilitation in COPD. ERJ Open Res 2017;3(1) [FREE Full text] [doi: 10.1183/23120541.00089-2016] [Medline: 28154821]
32. Lall P, Rees R, Law GCY, Dunleavy G, Cotič Ž, Car J. Influences on the implementation of mobile learning for medical and nursing education: qualitative systematic review by the digital health education collaboration. J Med Internet Res 2019;21(2):e12895 [FREE Full text] [doi: 10.2196/12895] [Medline: 30816847]
33. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
34. Watson JS, Adab P, Jordan RE, Enocson A, Greenfield S. Referral of patients with chronic obstructive pulmonary disease to pulmonary rehabilitation: a qualitative study of barriers and enablers for primary healthcare practitioners. Br J Gen Pract 2020;70(693):e274-e284 [FREE Full text] [doi: 10.3399/bjgp20X708101] [Medline: 31988083]
35. Slevin, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. Exploring the barriers and facilitators for the use of digital health technologies for the management of COPD: a qualitative study of clinician perceptions. QJM 2020;113(3):163-172. [doi: 10.1093/qjmed/hcz241] [Medline: 31545374]
36. Carayon P, Karsh B, Gurses AP, Holden RJ, Hoonakker P, Hundt AS, et al. Macroergonomics in healthcare quality and patient safety. Rev Hum Factors Ergon 2013;8(1):4-54 [FREE Full text] [doi: 10.1177/1557234X13492976] [Medline: 24729777]
37. Rebhan M. Towards a systems approach for chronic diseases, based on health state modeling. F1000Res 2017;6:309 [FREE Full text] [doi: 10.12688/f1000research.11085.1] [Medline: 28529704]
38. Efficace F, Patriarca A, Luppi M, Potenza L, Caocci G, Tafuri A, et al. Physicians' perceptions of clinical utility of a digital health tool for electronic patient-reported outcome monitoring in real-life hematology practice. Evidence from the GIMEMA-ALLIANCE platform. Front Oncol 2022;12:826040 [FREE Full text] [doi: 10.3389/fonc.2022.826040] [Medline: 35372079]
39. Petersen JA, Rasmussen LS, Rydahl-Hansen S. Barriers and facilitating factors related to use of early warning score among acute care nurses: a qualitative study. BMC Emerg Med 2017;17(1):36 [FREE Full text] [doi: 10.1186/s12873-017-0147-0] [Medline: 29191159]

XSL•FO
RenderX

40. Hoge CW. Interventions for war-related posttraumatic stress disorder: meeting veterans where they are. JAMA 2011;306(5):549-551. [doi: 10.1001/jama.2011.1096] [Medline: 21813436]
41. Cooper H, Coutanche MN, McMullen LM, Panter AT, Rindskopf D, Sher KJ. APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2012.
42. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. Los Angeles, CA: SAGE Publications Ltd; 2016.
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
45. Rowlands T, Waddell N, McKenna B. Are we there yet? A technique to determine theoretical saturation. J Comput Inf Syst 2015;56(1):40-47. [doi: 10.1080/08874417.2015.11645799]

Abbreviations

COPD: chronic obstructive pulmonary disease
DH: digital health
DHT: digital health technology
TA: thematic analysis

Edited by A Coristine; submitted 26.06.24; peer-reviewed by YW Shau, KM Liao; comments to author 19.08.24; revised version received 28.08.24; accepted 11.12.24; published 05.02.25.

Please cite as:
Chien SY, Hu HC, Cho HY
Long-Term Monitoring of Individuals With Chronic Obstructive Pulmonary Disease Using Digital Health Technology: Qualitative Study
J Med Internet Res 2025;27:e63660
URL: https://www.jmir.org/2025/1/e63660
doi:10.2196/63660
PMID:

©Shih-Ying Chien, Han-Chung Hu, Hsiu-Ying Cho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mental Health Changes in US Transgender Adults Beginning Hormone Therapy Via Telehealth: Longitudinal Cohort Study

Jae Downing Corman[1], PhD; Jaclyn M W Hughto[2], MPH, PhD; Theresa I Shireman[2], PhD; Kellan Baker[3], PhD; Kate Steinle[1], NP; Michelle Forcier[1,2], MD

[1]FOLX Health, Boston, MA, United States

[2]Brown University, Providence, RI, United States

[3]Whitman-Walker Health, Washington, DC, United States

Corresponding Author:
Jae Downing Corman, PhD
FOLX Health
1 Boston Pl
Ste 2600
Boston, MA, 02108
United States
Phone: 1 617 792 6855
Email: jae@folxhealth.com

Abstract

Background: Gender-affirming hormone therapy (GAHT) has shown potential for improving mental health outcomes among transgender and gender-diverse adults. How clinical outcomes change among adults receiving GAHT via telehealth across the United States is not well known.

Objective: This study evaluated the relationship between initiating GAHT via a telehealth clinic and changes in depression, anxiety, and suicide ideation over a 3-month period.

Methods: This cohort study evaluated the relationship between initiating GAHT via a telehealth clinic and changes in mental health over a 3-month period. Data were collected at baseline and 3 months later among adults who had their first GAHT visit between August and November 2023. The study included adults aged 18 years and older initiating GAHT for the first time, with a final sample of 342 adults across 43 states (192 initiated estrogen and 150 initiated testosterone therapy). The primary outcomes were depression symptoms using the Patient Health Questionnaire-9 (PHQ-9), anxiety symptoms using the General Anxiety Disorder-7 (GAD-7), and suicide ideation in the past 2 weeks.

Results: Before GAHT initiation, 40% (136/342) of participants reported depression (PHQ-9 ≥10), 36% (120/342) reported anxiety (GAD-7 ≥8), and 25% (91/342) reported suicidal ideation. By follow-up, significant reductions were observed in PHQ-9 (−2.4, 95% CI −3.0 to −1.8) and GAD-7 scores (−1.5, 95% CI −2.0 to −1.0). Among those with elevated symptoms, 40% (48/120) to 42% (56/133) achieved a clinically meaningful response (≥50% reduction in baseline scores), and 27% (36/133) to 28% (33/120) achieved remission (PHQ-9 or GAD-7 score <5). Of those with suicide ideation at baseline, 60% (50/83) had none at follow-up.

Conclusions: This study highlights the important relationship between telehealth-delivered GAHT and mental health, emphasizing the importance of accessible and timely care.

(J Med Internet Res 2025;27:e64017) doi:10.2196/64017

KEYWORDS

transgender; LGBT persons; telehealth; depression; anxiety; suicide; mental health; adult; virtual care; longitudinal cohort study; gender-affirming hormone therapy; United States; observational study; adolescent; mobile health

Introduction

Transgender and gender-diverse (TGD) individuals have poorer mental health outcomes than their cisgender peers, including depression, anxiety, and suicide ideation [1-3]. TGD individuals consistently face significant barriers to accessing gender-affirming medical and surgical treatments, leading many to delay or altogether forgo necessary care [3]. These challenges include difficulty in locating providers who are affirming and knowledgeable, financial barriers to paying for care, and

XSL•FO
RenderX

frequent discrimination in health care settings [4-6]. Despite legislation against gender identity discrimination [7], an unprecedented number of states enacted banned medical care for transgender people between 2021 and 2024, further reducing access to care [8]. These barriers to care can intensify psychological distress and create a syndemic environment that significantly increases the risk of poor mental health for TGD people [6].

Gender-affirming hormone therapy (GAHT), which involves medications such as testosterone, estrogens, and antiandrogens [9], has been associated with improvements in depression, anxiety, suicidality, and overall quality of life for transgender people [4-6]. A 2022 randomized controlled trial (RCT) conducted in Australia found a significant reduction in depression and suicidality within 3 months for patients beginning testosterone therapy compared to those for whom access to care was delayed [10]. However, the study may not represent individuals using estrogen or those outside of Australia, where care barriers differ from those in the United States. For example, TGD individuals in the United States face varied access challenges [10-12] particularly in states [10] where recent legislation led to legal penalties for providers of gender-affirming care [13], clinic closures, and care disruptions. These laws profoundly impact TGD individuals' mental health [14,15]. Studies that include a nationwide population of TGD adults can help ensure that the findings are applicable across the United States.

Further, the rapid expansion of telehealth and acceptance of telehealth-delivered GAHT [7] after COVID-19 has shifted the way this care is provided. Telehealth is a critical modality for primary care [8], mental [16], and behavioral health [15], yet its efficacy in delivering GAHT remains less explored [11,17]. TGD adults report frequent use of telemedicine [10], and express high levels of satisfaction with telehealth-delivered GAHT [12]. Therefore, understanding changes in health outcomes among those initiating GAHT via telehealth is essential to ensure that this increasingly widespread model of care provides effective, equitable, and accessible treatment.

Factors that are associated with improved mental health outcomes for those initiating GAHT are not well understood. Previous research found the mental health of TGD individuals varies based on insurance coverage [13], legal affirmation of gender [14,18], gender-affirming surgery access [19], geography [20], gender identity [21], race and ethnicity [9], employment status, and mental health treatment [22]. Few studies have explored the association between this comprehensive set of factors and changes in mental health outcomes.

We conducted an observational study of adults in 43 states initiating GAHT for the first time at a telehealth clinic to assess changes in depression, anxiety, and suicide ideation over a 3-month period. In addition, we examined whether these changes varied based on demographic factors, insurance coverage, and mental health treatment.

Methods

Overview

Ethical Considerations

This study received an exemption from the Western Institutional Review Board-Copernicus Group (WCG) institutional review board. The study follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Study Procedures (Participant Selection and Study Framework)

We conducted a single-arm, observational study at FOLX Health, a telehealth clinic licensed across all 50 US states and the District of Columbia, with adults aged 18 years and older who initiated testosterone or estrogen therapy between August 3, 2023, and November 6, 2023. Participants were required to complete an intake questionnaire before their initial visit and a follow-up assessment 3-4 months later. All had consented to care and were first-time GAHT recipients, and the study did not influence their treatment outcomes.

All participants self-scheduled their initial GAHT visit and completed a comprehensive intake questionnaire before the visit, and a follow-up questionnaire three months later to assess mental health and care satisfaction. Alerts were sent if urgent mental health issues were reported. Initial Zoom visits lasted 30 minutes, with payment options including insurance or cash, and prescriptions could be sent to local pharmacies or delivered directly to homes. Treatment options, such as injectable testosterone cypionate or enanthate, transdermal testosterone gel, estradiol tablets, injectable estradiol cypionate or valerate, and transdermal estrogen patches, were chosen based on standard care practices and patient goals, under clinician guidance [23]. Psychotherapy was not a requirement for initiation of GAHT [23].

Measures

Psychological Outcomes

The primary outcomes of this study were the resolution of suicide ideation, and Healthcare Effectiveness Data and Information Set measures for response and remission from depression and anxiety [24], assessed using the Patient Health Questionnaire-9 (PHQ-9) [25] for depression and the General Anxiety Disorder-7 (GAD-7) for anxiety [24,26]. Elevated symptoms were defined as PHQ-9 ≥10 and GAD-7 ≥8 [27,28]. Response, our primary outcome, was defined as a 50% reduction in these scores [29,30], accommodating patients with various baseline severities [29]. Remission, chosen as a secondary outcome for its correlation with improved daily function and long-term prognosis, was defined as scores below 5 on both tools. Completion of the PHQ-9 and GAD-7 was mandatory at baseline, ensuring that all respondents provided data for these instruments.

Suicidal ideation was evaluated with item 9 of the PHQ-9, "Thoughts that you would be better off dead or of hurting yourself in some way." Consistent with existing research and

XSL•FO
RenderX

the relevance of this measure for gender-diverse populations [2,31], any response other than "not at all" indicated possible suicidal ideation. Resolution was defined as a "not at all" response at the follow-up (refer to Multimedia Appendix 1).

Covariates Plausibly Associated With Changes in Outcomes

We included demographic factors (age, sex assigned at birth, race or ethnicity, and legal change of gender marker on identity document), access factors (insurance coverage, urbanicity of living environment, educational level, and unemployment status), and health care factors (history of a gender-affirming surgery (GAS); use of mental health treatment, including medication for anxiety or depression; use of talk therapy in the year previous [32] to initiating hormones; and use of talk therapy during the study period).

Statistical Analysis

First, we present baseline descriptive statistics, categorizing continuous variables as mean (SD) and categorical variables as frequency (percentage). These statistics are provided separately for individuals seeking testosterone and estrogen therapy.

Second, we calculated the mean changes in PHQ-9 and GAD-7 scores for all participants, along with their 95% CIs. Next, we analyzed the proportion of individuals achieving response and remission from depression and anxiety, and resolution of suicide ideation.

Third, for each group —those with baseline elevated symptoms of depression (PHQ-9 ≥10) anxiety (GAD-7 ≥8), and suicide ideation— we estimated 5 separate models for each outcome: resolution of suicide ideation, and response and remission of anxiety and depression. We used generalized estimating equations with a binomial family and a logit link function to estimate adjusted odds ratios for outcomes while accounting for individual-level clustering. These models estimated the association between outcomes and any mental health treatment, health insurance, education, urbanicity of residence, age, and sex assigned at birth. All models included time to follow-up (in weeks), and because remission does not take into account baseline symptom severity, we also included baseline PHQ-9 and GAD-7 scores in those models. We excluded covariates with a prevalence of <5% and state of residence from the final model to prevent overfitting and ensure the reliability of our estimates.

Participants with missing data on PHQ-9 or GAD-7 scores at baseline or follow-up were excluded from the analysis. For all other variables, missing data were categorized separately and included in the analysis to maintain the integrity and size of the sample.

We conducted a sensitivity analysis in which we reran all models after excluding individuals who had experienced mental health treatment, as well as those who had undergone GAS or legally changed their gender. This approach aimed to help mitigate the effects of unobserved confounders like symptom persistence and perceived need for care, which could influence mental health independently.

Finally, to assess potential differences between participants who completed the follow-up survey and those who did not, we analyzed the association of completing the follow-up survey with baseline PHQ-9 and GAD-7 scores, controlling for time and covariates described previously.

Statistical analyses were conducted using Python (version 3.11.7; Python Software Foundation) in March 2024.

Results

Overview

A total of 669 adults initiated GAHT via telehealth during the dates specified in our study period, with 98% (658/669) completing an initial intake survey; and 52% (342/669) participating in a follow-up survey. A detailed comparison between participants who completed the follow-up and those who did not is available in Multimedia Appendix 2. There were no baseline mental health differences between participants who did and did not complete the follow-up survey.

Our final sample included 342 adults ages 18 years to 67 years (mean 26.5 years [SD 8.5 years]). Among 342 adults, 192 (56.1%) were prescribed estrogen therapy and 150 (43.9%) were prescribed testosterone therapy.

Prescriptions were issued primarily by nurse practitioners (291/342, 85%), and the remainder were primary care physicians (51/342, 15%). Most patients (298/342, 87%) completed their first appointment within 7 days of registration. Prescription fulfillment was split, with 50% (171/342) receiving medications through direct mail and 50% (171/342) through local pharmacies. Estradiol tablets were the predominant form of estrogen therapy, while injectable testosterone cypionate was the most common testosterone therapy (Multimedia Appendix 3).

Overall, 71% (243/342) identified their race as white only, and 12% (42/342) identified their ethnicity as Latinx (Table 1). Geographic distribution showed a concentration in the southern United States (170/342, 50%), and socioeconomic data revealed that 59% (201/342) were uninsured and 6% (18/342) unemployed, with educational attainment distributed as follows: 36% (122/342) with high school education or less, 37% (124/342) with some college education, and 27% (93/342) holding at least a college degree.

XSL•FO
RenderX

Table 1. Baseline characteristics of adults initiating gender-affirming hormone therapy through telehealth.

	All (n=342)	Estrogen (n=192)	Testosterone (n=150)	*P* value[a]
Age group, n (%)				
18-21	122 (35.67)	56 (29.17)	66 (44)	.04
22-26	84 (24.56)	51 (26.56)	33 (22)	
27-31	68 (19.88)	41 (21.35)	27 (18)	
≥32	68 (19.88)	44 (22.92)	24 (16)	
Region[b], n (%)				
Midwest	43 (12.57)	29 (15.1)	14 (9.33)	.08
North	54 (15.79)	36 (18.75)	18 (12)	
South	170 (49.71)	90 (46.88)	80 (53.33)	
West	75 (21.93)	37 (19.27)	38 (25.33)	
Urbanicity, n (%)				
City	75 (21.99)	36 (18.75)	39 (26.17)	.35
Suburb	112 (32.84)	68 (35.42)	44 (29.53)	
Small town	123 (36.07)	69 (35.94)	54 (36.24)	
Rural	31 (9.09)	19 (9.9)	12 (8.05)	
Missing	1 (0.3)	0 (0)	1 (0.02)	
White alone, n (%)				
Yes	243 (71.47)	142 (74.74)	101 (67.33)	.17
No	97 (28.53)	48 (25.26)	49 (32.67)	
Missing	2 (0.58)	2 (0.1)	0 (0)	
Latinx, n (%)				
Yes	42 (12.28)	18 (9.38)	24 (16)	.09
No	300 (87.72)	174 (90.62)	126 (84)	
Changed legal gender, n (%)				
Yes	16 (4.68)	5 (2.6)	11 (7.33)	.07
No	326 (95.32)	187 (97.4)	139 (92.67)	
Gender identity, n (%)				
Woman or trans woman	121 (35.48)	118 (61.78)	3 (2)	<.001
Man or trans man	108 (31.67)	11 (5.76)	97 (64.67)	
Nonbinary	99 (29.03)	55 (28.8)	44 (29.33)	
Other	13 (3.81)	7 (3.66)	6 (4)	
Missing	1 (0.29)	1 (0.52)	0	
Education category, n (%)				
High school or less	122 (35.99)	65 (34.21)	57 (38.26)	.69
Some college	124 (36.58)	70 (36.84)	54 (36.24)	
College graduate or higher	93 (27.43)	55 (28.95)	38 (25.5)	
Unemployment status, n (%)				
Yes	18 (5.57)	5 (2.76)	13 (9.15)	.03
No	305 (94.43)	176 (97.24)	129 (90.85)	
Missing	19 (5.53)	11 (5.72)	8 (5.33)	
Insurance type, n (%)				

XSL•FO
RenderX

	All (n=342)	Estrogen (n=192)	Testosterone (n=150)	*P* value[a]
In-network	101 (29.53)	50 (26.04)	51 (34)	.13
Out-of-network	40 (11.7)	20 (10.42)	20 (13.33)	
Uninsured	201 (58.77)	122 (63.54)	79 (52.67)	
Pharmacy type, n (%)				
Local pharmacy	170 (49.71)	95 (49.48)	75 (50)	1.00
Shipped to home	172 (50.29)	97 (50.52)	75 (50)	
Had a mental health visit in past year, n (%)				
Yes	132 (38.6)	63 (32.81)	69 (46)	.02
No	210 (61.4)	129 (67.19)	81 (54)	
Number of mental health visits in previous year, mean (SD)	6.01 (12.51)	4.62 (10.64)	7.78 (14.39)	.03
Suicide ideation at baseline, n (%)				
Yes	85 (24.85)	52 (27.08)	33 (22)	.34
No	257 (75.15)	140 (72.92)	117 (78)	
Baseline PHQ-9[c] category, n (%)				
None or minimal: 0-4	128 (37.43)	72 (37.5)	56 (37.33)	.38
Mild: 5-9	78 (22.81)	44 (22.92)	34 (22.67)	
Moderate: 10-14	76 (22.22)	37 (19.27)	39 (26)	
[d]Moderately Severe: 15-19	42 (12.28)	26 (13.54)	16 (10.67)	
[d]Severe: 20-27	18 (5.26)	13 (6.77)	5 (3.33)	
Baseline GAD-7[e]category, n (%)				
None/Minimal: 0-4	127 (37.13)	73 (38.02)	54 (36)	.91
Mild: 5-7	91 (26.61)	51 (26.56)	40 (26.67)	
[f]Mild: 8-9	40 (11.7)	20 (10.42)	20 (13.33)	
[f]Moderately Severe: 10-14	51 (14.91)	28 (14.58)	23 (15.33)	
[f]Severe: 15-21	33 (9.65)	20 (10.42)	13 (8.67)	
Prescribed a medication for anxiety or depression, n (%)				
Yes	82 (23.98)	36 (18.75)	46 (30.67)	.02
No	260 (76.02)	156 (81.25)	104 (69.33)	
Had gender-affirming surgery, n (%)				
Yes	14 (4.09)		14 (9.33)	<.001
No	328 (95.91)	192 (100)	136 (90.67)	

[a]*P* values represent the statistical significance of differences between adults prescribed testosterone and estrogen, calculated using appropriate tests for each variable type (eg, chi-square test for categorical variables and *t* test for continuous variables).

[b]Patients resided in 43 states.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]Probable major depression.

[e]GAD-7: General Anxiety Disorder-7

[f]Probably anxiety disorder.

Only 5% (16/342) of participants had legally changed their gender on their license or identification card, and 4% (14/342) had undergone GAS. Meanwhile, 39% (132/342) had consulted a mental health professional in the past year, with an average of six visits, and 24% (82/342) were prescribed medication for anxiety or depression.

Individuals seeking testosterone therapy were younger, used more mental health care services, had higher rates of

XSL•FO
RenderX

prescriptions for anxiety or depression, and were more likely to have undergone one or more GAS compared to those seeking estrogen therapy.

Baseline and Follow-Up Depression and Anxiety Scores

There were 333 individuals who reported depression and anxiety scores at both baseline and follow-up. The mean baseline PHQ-9 score was 7.6 (Figure 1). At follow-up, there was a statistically significant decrease in PHQ-9 scores (mean difference, −2.4 points; 95% CI −3.0 to −1.8 points).

Figure 1. Baseline and follow-up depression (PHQ-9) and anxiety (GAD-7) scores among adults who initiated gender-affirming hormone therapy. GAD-7: General Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9.

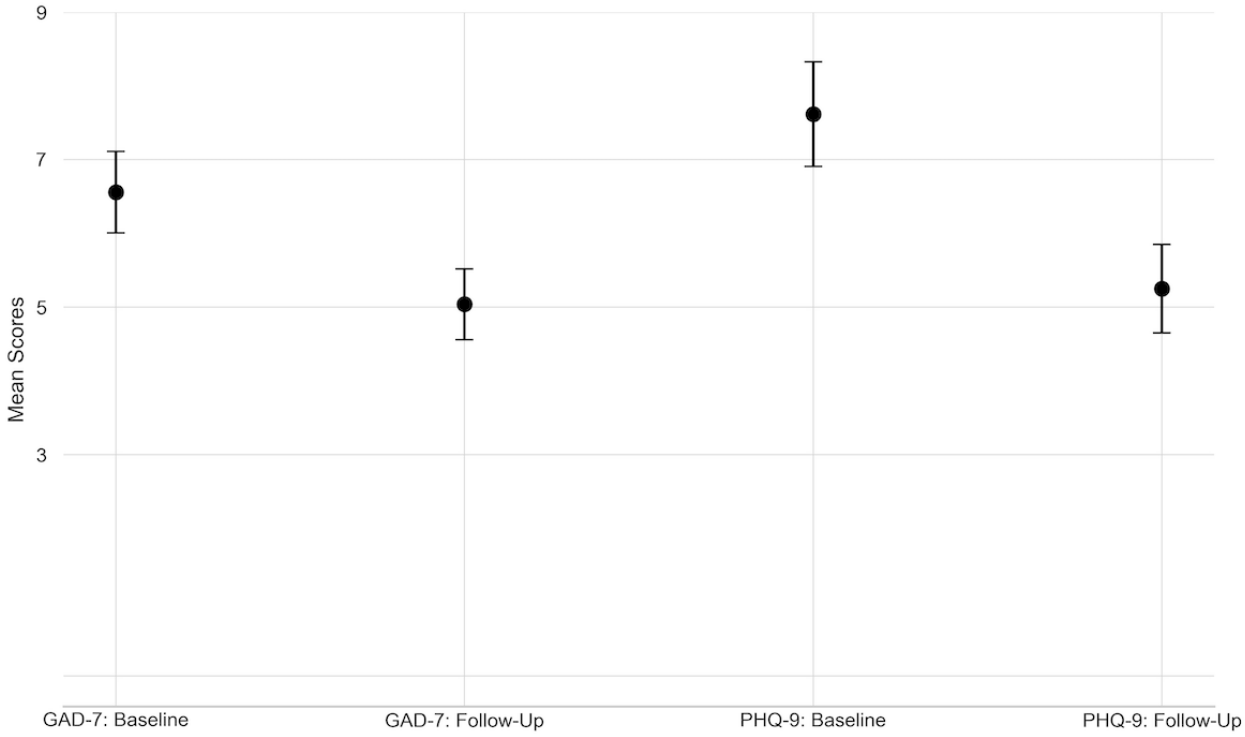

Mean and 95% CIs are represented for 333 adults who completed baseline and follow-up PHQ-9 and GAD-7 assessments.

In total, 40% (136/342) had elevated symptoms of depression (PHQ-9 ≥10) at baseline. This included 76 individuals who scored moderate (10 to 14), 42 individuals who scored moderately severe (15 to 19), and 18 individuals who scored severe (20 to 27). For those without elevated symptoms, the baseline PHQ-9 score was none/minimal (0 to 4) in 128 individuals (128/342, 37%), and mild (5 to 9) in 78 individuals (78/342, 23%).

The mean baseline GAD-7 score was 6.6. At follow-up, there was a statistically significant decrease in GAD-7 scores (mean difference −1.5 points, 95% CI −2.0 to −1.0 points).

In total, 36% (120/342) had elevated symptoms of anxiety (GAD-7 ≥8). This included 40 individuals who scored mild (8-9), 51 individuals who scored moderately severe (10 to 14), and 33 individuals who scored severe (15 to 21). For those without elevated symptoms, the baseline GAD-7 score was none or minimal (0 to 4) in 127 individuals (127/342, 37%), and mild (5 to 7) in 91 individuals (91/342, 27%).

Response and Remission from Depression and Anxiety

Among the 133 individuals with PHQ-9 ≥10 at baseline, 42% (n=56) achieved response (50% reduction in score), and 27% (n=36) achieved remission (PHQ-9 <5) at follow-up (Figure 2).

XSL•FO
RenderX

Figure 2. The proportion of adults who initiated GAHT (gender-affirming hormone therapy) achieving remission and response from elevated symptoms of depression and anxiety, and resolution of suicide ideation at follow-up.

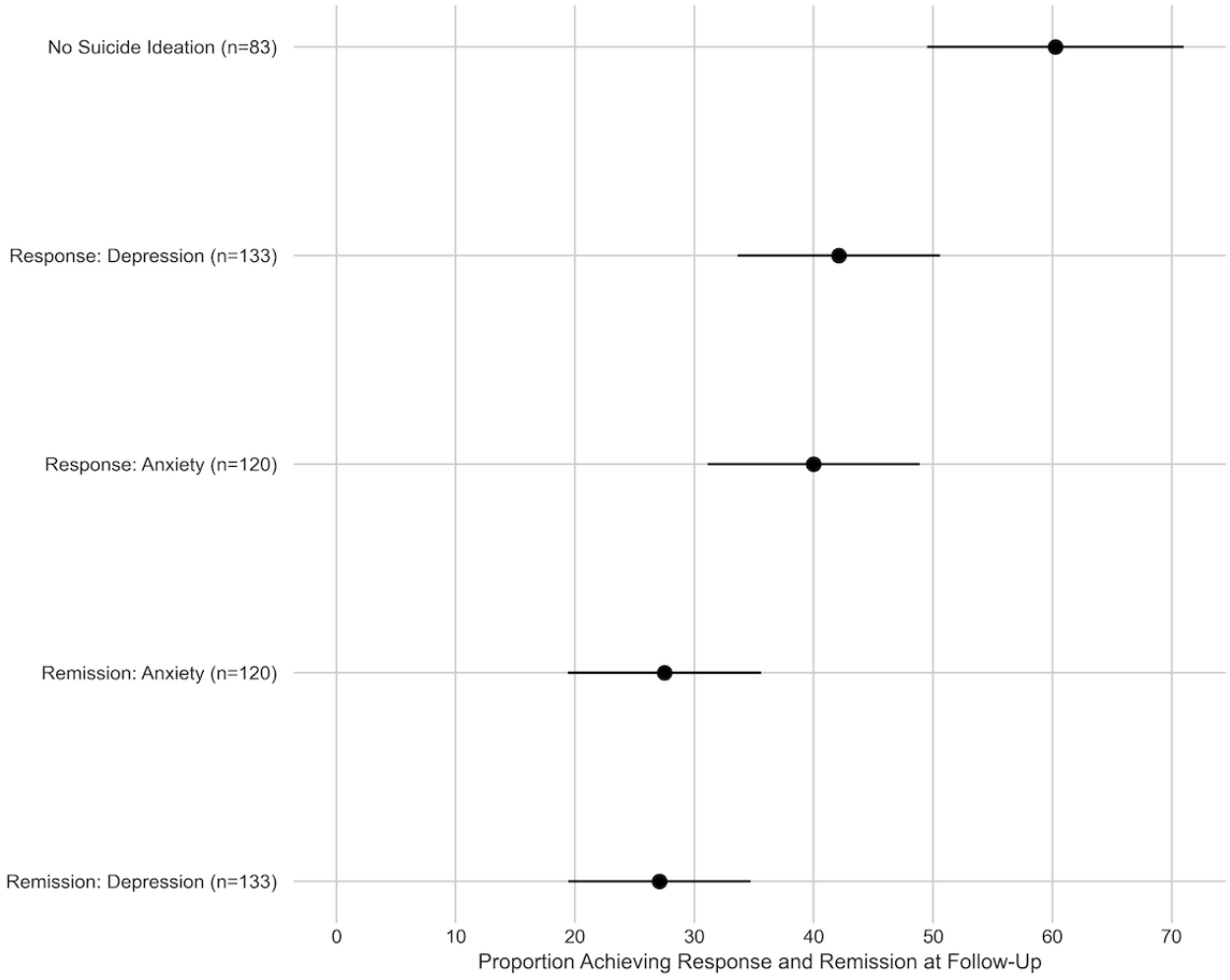

Among the 120 individuals with GAD-7 ≥8 at baseline, 40% (n=48) achieved response, and 28% (n=33) achieved remission at follow-up.

Resolution of Suicide Ideation

At baseline, 83 individuals (83/342, 25%) reported suicide ideation. Among these individuals, 60% (50/83) did not report suicide ideation at follow-up.

Adjusted Odds of Response and Remission

Use of mental health treatment before or during the period was associated with a 0.43 (95% CI 0.19-0.93) lower odds of depression response and 0.43 (95% CI 0.19-0.98) lower odds of anxiety response. Those who were female sex assigned at birth (FSAB) had a 0.38 (95% CI 0.18-0.82) lower odds of depression response. No statistically significant association was found between depression or anxiety symptom response and age, race or ethnicity, insurance coverage, educational level, or use of mental health treatment (Table 2).

No associations were observed between mental health treatment or other baseline factors and remission of depression, anxiety, and resolution of suicide ideation (Table 3).

XSL•FO
RenderX

Table 2. Association between demographic and treatment factors and mental health response among adults initiating gender-affirming hormone therapy with elevated symptoms of depression or anxiety.

	Response from depression (n= 133)[a]		Response from anxiety (n= 120)[b]	
	Adjusted odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Any mental health treatment[c]	0.43 (0.19-0.93)	.03	0.43 (0.19-0.98)	.04
Age (in years)	1.06 (0.99-1.13)	.10	1.03 (0.97-1.11)	.34
Sex assigned at birth				
Male sex assigned at birth	1 (Reference)		1 (Reference)	
Female sex assigned at birth	0.38 (0.18-0.82)	.01	0.77 (0.34-1.73)	.52
Non-Latinx White alone	1.66 (0.73-3.75)	.22	0.77 (0.32-1.83)	.55
Insurance status				
Uninsured	1 (Reference)		1 (Reference)	
Insured	0.64 (0.29-1.43)	.27	0.81 (0.37-1.79)	.61
Urbanicity of residence				
City or suburb	1 (Reference)		1 (Reference)	
Rural or small town	1.22 (0.56-2.63)	.62	1.36 (0.62-3.02)	.44
Education				
Some college or higher	1 (Reference)		1 (Reference)	
High school or less	0.83 (0.38-1.82)	.63	0.86 (0.36-2.02)	.73

[a]All models use Generalized Estimating Equations with a binomial family and a logit link function, with clustering at the individual level and controlling for time to follow-up. We excluded employment status, change of legal gender, or history of a gender-affirming surgery because of small cell sizes. This model includes those with a Patient Health Questionnaire-9 score ≥10 at baseline. The response is defined as a ≥50% improvement in the Patient Health Questionnaire-9 score from baseline.

[b]Model includes those with a General Anxiety Disorder-7 score ≥8 at baseline. The response is defined as a ≥50% improvement in the General Anxiety Disorder-7 score.

[c]Any mental health treatment was defined as using an antidepressant or talk therapy a year before or during the study period.

XSL•FO
RenderX

Table 3. Association between demographic and treatment factors and mental health remission among adults initiating gender-affirming hormone therapy.

	Remission from depression[a] (n=133)		Remission from anxiety[b] (n=120)		Remission from suicide ideation[c] (n=83)	
	Adjusted odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Any mental health treatment[d]	0.42 (0.17-1.03)	.06	0.49 (0.19-1.26)	.14	2.46 (0.87-6.99)	.09
Age (in years)	1.06 (0.99-1.14)	.11	1.01 (0.94-1.08)	.84	1.01 (0.91-1.11)	.86
Sex assigned at birth						
Male	1 (Reference)		1 (Reference)		1 (Reference)	
Female	0.52 (0.21-1.25)	.14	0.58 (0.25-1.38)	.22	0.43 (0.14-1.30)	.14
Non-Latinx White alone	2.01 (0.82-4.92)	.13	0.75 (0.28-1.98)	.56	0.75 (0.27-2.09)	.58
Insurance status						
Uninsured	1 (Reference)		1 (Reference)		1 (Reference)	
Insured	0.55 (0.21-1.44)	.22	0.76 (0.31-1.88)	.55	0.46 (0.17-1.26)	.13
Urbanicity of residence						
City or suburb	1 (Reference)		1 (Reference)		1 (Reference)	
Rural or small town	1.53 (0.64-3.62)	.34	1.61 (0.68-3.81)	.28	1.17 (0.43-3.21)	.76
Education						
Some college or higher	1 (Reference)		1 (Reference)		1 (Reference)	
High school or less	1.25 (0.54-2.93)	.61	0.53 (0.20-1.43)	.21	1.40 (0.48-4.09)	.53
GAD-7[e] or PHQ-9[f] baseline	0.93 (0.84-1.03)	.14	0.91 (0.80-1.04)	.16		

[a]All models use Generalized Estimating Equations with a binomial family and a logit link function, with clustering at the individual level and controlling for time to follow-up. We excluded employment status, change of legal gender, or history of a gender-affirming surgery because of small cell sizes. This model includes those with PHQ-9 ≥10 at baseline. Remission is defined as a PHQ-9 <5 at follow-up

[b]Model includes those with GAD-7 ≥8 at baseline. Remission is defined as a GAD-7 score <4 at follow-up.

[c]Model includes those with suicide ideation at baseline. Remission includes those who do not have suicide ideation at follow-up.

[d]Any mental health treatment was defined as using an antidepressant or using talk therapy a year before or during the study period.

[e]GAD-7: General Anxiety Disorder-7.

[f]PHQ-9: Patient Health Questionnaire-9.

Sensitivity Analyses

After excluding those without mental health treatment, GAS, or a legal gender change, no statistically significant association was observed between FSAB and odds of depression and response or remission (Multimedia Appendices 4 and 5).

Discussion

Principal Findings

Among adults newly initiating testosterone or estrogen therapy via telehealth services with elevated depression and anxiety symptoms, 40%-42% achieved a clinically meaningful response. Of those with suicide ideation, 60% saw resolution within 3 months. These results are consistent with improvements reported in previous studies [4,5,33]; however, most of these studies were conducted outside the United States, inclusive of youth, and before the COVID-19 pandemic, thus telehealth modalities were not specifically explored. Moreover, very few studies have used similar outcome measures. One such example is an observational study of 315 youth aged 13 to 20 from Seattle. This study reported no change in moderate to severe depressive symptoms at 3 or 6 months, with improvements only evident after a year on GAHT [34].

Other US-based studies, including smaller cohorts from New York, California, Boston, and Illinois, also documented mental health improvements in youth initiating GAHT using other measures [6,35]. While comparisons between observational studies and RCTs, may raise concerns due to differences in study design, it is noteworthy that 61% of Australian adults receiving testosterone saw a reduction of five or more points on the PHQ-9 score, and 52% experienced the resolution of suicide ideation within three months [34]. Our study fills a critical gap by providing contemporary insights on hormone therapy via telehealth, demonstrating significant mental health improvements in a setting largely unexplored in previous research.

Although GAHT is not explicitly prescribed for anxiety or depression, its impact on mental health can be profound, and in this study, it compares favorably with treatments that target mental health conditions directly. For instance, antidepressants

XSL•FO
RenderX

typically show an improvement in symptoms for less than 20% of adults with moderate to severe depression [36,37]. Other types of treatments show similar effects, with 13-25% of adults achieving remission from depression in traditional and telehealth-enabled collaborative care models [38,39]. While GAHT is not a replacement for mental health condition–specific treatments, this research strongly suggests that it is an important factor in mental health wellness for TGD adults.

Our finding that individuals who used mental health treatment had lower odds of achieving a depression response, though not remission, suggests that those with a history of mental health treatment may have more persistent symptoms of depression. In contrast, individuals without prior treatment may have more emergent symptoms. Furthermore, engagement in mental health treatment is not randomly assigned; rather, it is heavily influenced by both perceived needs and available resources [36]. Notably, less than 40% of this population had health insurance, which likely limited their access to consistent and comprehensive mental health treatment. Our findings imply that improvements in mental health observed in individuals using hormone therapy occurred independently of their use of other mental health treatments.

Prior studies have reported that adults with multiple marginalized identities are at risk for worse mental health as a result of compounded discrimination [21]. While our findings showed no significant differences in mental health remission based on race, ethnicity, or education, we observed that FSAB had lower odds of depression response than male sex assigned at birth adults. However, this association disappeared when excluding individuals already receiving mental health treatment. The higher rates of mental health treatment among FSAB individuals suggest variations in underlying symptomatology, which our main models could not fully adjust for. Future research should further elucidate how concurrent GAHT and mental health treatment influence mental health symptoms.

Limitations

A primary limitation is the lack of a control group of participants who did not receive GAHT, which would be both infeasible and unethical. This makes our observational study unable to definitively link improvements in mental health to GAHT alone due to potential unmeasured factors. In addition, with only a 3-month follow-up, the long-term effects of GAHT remain unclear. Our reliance on patient-reported outcomes may introduce bias, although baseline mental health was comparable between those lost to follow-up and those retained. Furthermore, the quick scheduling of initial visits within a week, contrasting with the months- to years-long delays at traditional gender specialty clinics, may also influence the generalizability of our results.

The generalizability of our findings may be limited to similar care models, as our cohort was drawn from a specialized telehealth-only clinic—one of a few in the United States providing GAHT. Although telehealth has increasingly been recognized as a way to broaden geographic access to such care [7,40], it remains out of reach for some populations. Compared to other nationally representative samples of transgender and gender-diverse adults, our participants were generally younger, more educated, predominantly white, more rural, and less insured [1,41] Despite these differences, the prevalence of mental health symptoms before GAHT mirror those of other clinics providing in-person care, suggesting that our results may be comparable with other health care settings [2,27].

Conclusions

In this observational study, adults accessing GAHT at a telehealth clinic experienced significant improvements in depression, anxiety, and suicide ideation after 3 months. Although GAHT is not prescribed for mental health, it may be an important part of mental health wellness for TGD adults. Future studies with longer follow-up periods are essential to better understand these findings.

Acknowledgments

FOLX Health paid the salaries of Jae Downing Corman, Kate Steinle, and Michelle Forcier. No funding was provided for the other co-authors. Jae Corman conceptualized the analysis, drafted the paper, and conducted the analysis. All other co-authors contributed to revisions of all drafts.

Conflicts of Interest

JDC, KS, and MF are employed by FOLX Health.

Multimedia Appendix 1
Contains detailed description of measures used in the study.
[PDF File (Adobe PDF File), 44 KB - jmir_v27i1e64017_app1.pdf]

Multimedia Appendix 2
Differences in baseline characteristics of participants and nonparticipants in the follow-up assessment.
[PDF File (Adobe PDF File), 51 KB - jmir_v27i1e64017_app2.pdf]

Multimedia Appendix 3
Proportion of study participants by type of initial prescription of estrogen or testosterone.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 26 KB - jmir_v27i1e64017_app3.pdf]

Multimedia Appendix 4
Sensitivity analysis: association between demographic factors and mental health response among those without mental health treatment, gender-affirming surgery, or legal gender change.
[PDF File (Adobe PDF File), 43 KB - jmir_v27i1e64017_app4.pdf]

Multimedia Appendix 5
Sensitivity analysis: association between demographic and mental health remission among those without mental health treatment, gender-affirming surgery, or legal gender change.
[PDF File (Adobe PDF File), 49 KB - jmir_v27i1e64017_app5.pdf]

References

1. Downing JM, Przedworski JM. Health of transgender adults in the U.S., 2014-2016. Am J Prev Med 2018;55(3):336-344 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.045] [Medline: 30031640]
2. Kirakosian N, Stanton AM, McKetchnie SM, King D, Dolotina B, O'Cleirigh C, et al. Suicidal ideation disparities among transgender and gender diverse compared to cisgender community health patients. J Gen Intern Med 2023;38(6):1357-1365 [FREE Full text] [doi: 10.1007/s11606-022-07996-2] [Medline: 36650322]
3. Stanton AM, Batchelder AW, Kirakosian N, Scholl J, King D, Grasso C, et al. Differences in mental health symptom severity and care engagement among transgender and gender diverse individuals: findings from a large community health center. PLoS One 2021;16(1):e0245872 [FREE Full text] [doi: 10.1371/journal.pone.0245872] [Medline: 33493207]
4. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. Transgend Health 2016;1(1):21-31 [FREE Full text] [doi: 10.1089/trgh.2015.0008] [Medline: 27595141]
5. Baker KE, Wilson LM, Sharma R, Dukhanin V, McArthur K, Robinson KA. Hormone therapy, mental health, and quality of life among transgender people: a systematic review. J Endocr Soc 2021;5(4):bvab011 [FREE Full text] [doi: 10.1210/jendso/bvab011] [Medline: 33644622]
6. Chen D, Berona J, Chan YM, Ehrensaft D, Garofalo R, Hidalgo MA, et al. Psychosocial functioning in transgender youth after 2 years of hormones. N Engl J Med 2023;388(3):240-250 [FREE Full text] [doi: 10.1056/NEJMoa2206297] [Medline: 36652355]
7. Dowshen N, Lett E. Telehealth for gender-affirming care: challenges and opportunities. Transgend Health 2022;7(2):111-112 [FREE Full text] [doi: 10.1089/trgh.2021.0206] [Medline: 36644512]
8. Reed ME, Huang J, Graetz I, Lee C, Muelly E, Kennedy C, et al. Patient characteristics associated with choosing a telemedicine visit vs office visit with the same primary care clinicians. JAMA Netw Open 2020;3(6):e205873 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5873] [Medline: 32585018]
9. Vance SR, Boyer CB, Glidden DV, Sevelius J. Mental health and psychosocial risk and protective factors among Black and Latinx transgender youth compared with peers. JAMA Netw Open 2021;4(3):e213256 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3256] [Medline: 33769506]
10. Data from: household pulse survey, 2021-2022. Telemedicine use. Statistics NCfH. 2021. URL: https://www.cdc.gov/nchs/covid19/pulse/telemedicine-use.htm [accessed 2022-10-05]
11. Lock L, Anderson B, Hill BJ. Transgender care and the COVID-19 pandemic: exploring the initiation and continuation of transgender care in-person and through telehealth. Transgend Health 2022;7(2):165-169. [doi: 10.1089/trgh.2020.0161] [Medline: 35582360]
12. Apple DE, Lett E, Wood S, Freeman Baber K, Chuo J, Schwartz LA, et al. Acceptability of telehealth for gender-affirming care in transgender and gender diverse youth and their caregivers. Transgend Health 2022;7(2):159-164 [FREE Full text] [doi: 10.1089/trgh.2020.0166] [Medline: 35586576]
13. Carter SP, Cowan T, Snow A, Cerel J, Tucker R. Health insurance and mental health care utilization among adults who identify as transgender and gender diverse. Psychiatr Serv 2020;71(2):151-157. [doi: 10.1176/appi.ps.201900289] [Medline: 31658897]
14. King WM, Gamarel KE. A scoping review examining social and legal gender affirmation and health among transgender populations. Transgend Health 2021;6(1):5-22 [FREE Full text] [doi: 10.1089/trgh.2020.0025] [Medline: 33644318]
15. Allen LD. Navigating the path to effective, equitable, and evidence-based telehealth for opioid use disorder treatment. JAMA Netw Open 2023;6(10):e2336885 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36885] [Medline: 37851449]
16. Cantor JH, McBain RK, Ho P, Bravata DM, Whaley C. Telehealth and in-person mental health service utilization and spending, 2019 to 2022. JAMA Health Forum 2023;4(8):e232645 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.2645] [Medline: 37624614]

XSL•FO
RenderX

17. Grasso C, Campbell J, Yunkun E, Todisco D, Thompson J, Gonzalez A, et al. Gender-affirming care without walls: utilization of telehealth services by transgender and gender diverse people at a federally qualified health center. Transgend Health 2022;7(2):135-143 [FREE Full text] [doi: 10.1089/trgh.2020.0155] [Medline: 36644515]
18. Yee K, Lind BK, Downing J. Change in gender on record and transgender adults' mental or behavioral health. Am J Prev Med 2022;62(5):696-704 [FREE Full text] [doi: 10.1016/j.amepre.2021.10.016] [Medline: 34920916]
19. Almazan AN, Keuroghlian AS. Association between gender-affirming surgeries and mental health outcomes. JAMA Surg 2021;156(7):611-618 [FREE Full text] [doi: 10.1001/jamasurg.2021.0952] [Medline: 33909023]
20. White Hughto JM, Murchison GR, Clark K, Pachankis JE, Reisner SL. Geographic and individual differences in healthcare access for U.S. transgender adults: a multilevel analysis. LGBT Health 2016;3(6):424-433 [FREE Full text] [doi: 10.1089/lgbt.2016.0044] [Medline: 27636030]
21. Robertson L, Akré ER, Gonzales G. Mental health disparities at the intersections of gender identity, race, and ethnicity. LGBT Health 2021;8(8):526-535. [doi: 10.1089/lgbt.2020.0429] [Medline: 34591707]
22. Bränström R, Pachankis JE. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: a total population study. Am J Psychiatry 2020;177(8):727-734. [doi: 10.1176/appi.ajp.2019.19010080] [Medline: 31581798]
23. Coleman E, Radix AE, Bouman WP, Brown GR, de Vries ALC, Deutsch MB, et al. Standards of care for the health of transgender and gender diverse people, version 8. Int J Transgend Health 2022;23(Suppl 1):S1-S259 [FREE Full text] [doi: 10.1080/26895269.2022.2100644] [Medline: 36238954]
24. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
25. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
27. US Preventive Services Task Force, Barry MJ, Nicholson WK, Silverstein M, Coker TR, Davidson KW, et al. Screening for anxiety disorders in adults: US Preventive Services Task Force recommendation statement. JAMA 2023;329(24):2163-2170. [doi: 10.1001/jama.2023.9301] [Medline: 37338866]
28. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020;77(9):906-914 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
29. Coley RY, Boggs JM, Beck A, Hartzler AL, Simon GE. Defining success in measurement-based care for depression: a comparison of common metrics. Psychiatr Serv 2020;71(4):312-318. [doi: 10.1176/appi.ps.201900295] [Medline: 31847739]
30. Segal ZV, Dimidjian S, Beck A, Boggs JM, Vanderkruik R, Metcalf CA, et al. Outcomes of online mindfulness-based cognitive therapy for patients with residual depressive symptoms: a randomized clinical trial. JAMA Psychiatry 2020;77(6):563-573 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.4693] [Medline: 31995132]
31. Xu L, Chang R, Wang H, Xu C, Yu X, Chen H, et al. Validation of the Patient Health Questionnaire-9 for suicide screening in transgender women. Transgend Health 2023;8(5):450-456. [doi: 10.1089/trgh.2021.0075] [Medline: 37810941]
32. National health interview survey, 2021. Public-use data file and documentation. Statistics NCfH. 2022. URL: https://healthdata.gov/dataset/National-Health-Interview-Survey/q7bn-d69a/about_data [accessed 2024-08-07]
33. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. Transgend Health 2016;1(1):21-31 [FREE Full text] [doi: 10.1089/trgh.2015.0008] [Medline: 27595141]
34. Nolan BJ, Zwickl S, Locke P, Zajac JD, Cheung AS. Early access to testosterone therapy in transgender and gender-diverse adults seeking masculinization: a randomized clinical trial. JAMA Netw Open 2023;6(9):e2331919 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.31919] [Medline: 37676662]
35. Achille C, Taggart T, Eaton N, Osipoff J, Tafuri K, Lane A, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. Int J Pediatr Endocrinol 2020;2020:8 [FREE Full text] [doi: 10.1186/s13633-020-00078-2] [Medline: 32368216]
36. Depression: learn more – how effective are antidepressants? InformedHealth.org. 2020. URL: https://www.ncbi.nlm.nih.gov/books/NBK361016/ [accessed 2024-04-15]
37. Almohammed OA, Alsalem AA, Almangour AA, Alotaibi LH, Al Yami MS, Lai L. Antidepressants and health-related quality of life (HRQoL) for patients with depression: analysis of the medical expenditure panel survey from the United States. PLoS One 2022;17(4):e0265928 [FREE Full text] [doi: 10.1371/journal.pone.0265928] [Medline: 35442954]
38. Katzelnick DJ, Duffy FF, Chung H, Regier DA, Rae DS, Trivedi MH. Depression outcomes in psychiatric clinical practice: using a self-rated measure of depression severity. Psychiatr Serv 2011;62(8):929-935. [doi: 10.1176/ps.62.8.pss6208_0929] [Medline: 21807833]
39. Das S, Wang J, Chen SY, Chen CE. Telemental health collaborative care medication management: implementation and outcomes. Telemed J E Health 2022;28(7):1035-1043 [FREE Full text] [doi: 10.1089/tmj.2021.0401] [Medline: 34939839]

XSL•FO
RenderX

40. Radix AE, Bond K, Carneiro PB, Restar A. Transgender individuals and digital health. Curr HIV/AIDS Rep 2022;19(6):592-599 [FREE Full text] [doi: 10.1007/s11904-022-00629-7] [Medline: 36136217]
41. Feldman JL, Luhur WE, Herman JL, Poteat T, Meyer IH. Health and health care access in the US transgender population health (TransPop) survey. Andrology 2021;9(6):1707-1718 [FREE Full text] [doi: 10.1111/andr.13052] [Medline: 34080788]

Abbreviations

FSAB: female sex assigned at birth
GAD-7: General Anxiety Disorder-7
GAHT: gender-affirming hormone therapy
GAS: gender-affirming surgery
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
STROBE: Reporting of Observational Studies in Epidemiology
TGD: transgender and gender diverse
WCG: WIRB-Copernicus Group

Edited by T McCall; submitted 05.07.24; peer-reviewed by S Mehta, L Aljerf, D Valvi; comments to author 01.10.24; revised version received 08.11.24; accepted 13.01.25; published 14.02.25.

Please cite as:
Corman JD, Hughto JMW, Shireman TI, Baker K, Steinle K, Forcier M
Mental Health Changes in US Transgender Adults Beginning Hormone Therapy Via Telehealth: Longitudinal Cohort Study
J Med Internet Res 2025;27:e64017
URL: https://www.jmir.org/2025/1/e64017
doi:10.2196/64017
PMID:

©Jae Downing Corman, Jaclyn M W Hughto, Theresa I Shireman, Kellan Baker, Kate Steinle, Michelle Forcier. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usage Trends and Data Sharing Practices of Healthcare Wearable Devices Among US Adults: Cross-Sectional Study

Ranganathan Chandrasekaran[1], PhD; Muhammed Sadiq T[2], MSc; Evangelos Moustakas[3], PhD

[1]Department of Information & Decision Sciences / Department of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL, United States

[2]Department of Management Studies, Indian Institute of Technology Madras, Chennai, India

[3]Department of Communication and Media, Faculty of Communication, Arts and Sciences, Canadian University of Dubai, Dubai, United Arab Emirates

Corresponding Author:
Ranganathan Chandrasekaran, PhD
Department of Information & Decision Sciences / Department of Biomedical and Health Information Sciences
University of Illinois at Chicago
2428 Univ Hall
601 S Morgan St
Chicago, IL, 60607
United States
Phone: 1 3129962676
Email: ranga@uic.edu

Abstract

Background: Health care wearable devices can transform health care delivery by enabling real-time, continuous monitoring that facilitates early disease detection, personalized treatments, and improved patient engagement. The COVID-19 pandemic has heightened awareness of the importance of health technology, accelerating interest in wearables as tools for monitoring health and managing chronic conditions. As we navigate the postpandemic era, understanding the adoption and data-sharing behaviors associated with wearable devices has become increasingly critical. Despite their potential, challenges and low adoption rates persist, with significant gaps in understanding the impact of sociodemographic factors, health conditions, and digital literacy on the use and data-sharing behaviors of these devices.

Objective: This study aimed to explore the usage and data-sharing practices (willingness to share wearable data and actual data-sharing behavior) of wearable devices among US adults specifically during the later phases of the COVID-19 pandemic.

Methods: Using cross-sectional data from the National Cancer Institute's Health Information National Trends Survey 6, conducted from March to November 2022, this study uses responses from 5591 US adults to examine wearable use, willingness to share wearable data with providers, family, and friends, and the wearable data-sharing behavior.

Results: The results indicate an increase in wearable device adoption to 36.36% (2033/5591) in 2022, up from 28%-30% in 2019. We also find a significant discrepancy between the willingness to share data, with 78.4% (1584/2020) of users open to sharing with health care providers, and the actual sharing behavior, where only 26.5% (535/ 2020) have done so. Higher odds of using wearables were associated with female gender (odds ratio [OR] 1.49, 95% CI 1.17-1.90, P<.01) and higher income levels (OR 2.65, 95% CI 1.42-4.93, P<.01 for incomes between US $50,000 and US $75,000, and OR 3.2, 95% CI 1.71-5.97, P<.01 for incomes above US $75,000). However, the likelihood of usage and data sharing declines significantly with age. Compared with African American respondents, Hispanic respondents were more willing to share wearable data with providers (OR 1.92, 95% CI 1.02-3.62, P<.05), though the odds of their actual sharing of wearable data with providers was relatively less (OR 0.44, 95% CI 0.20-0.97, P<.05). Frequency of provider visits (OR 1.23, 95% CI 1.08-1.39, P<.01), and total medical conditions (OR 1.35, 95% CI 1.05-1.73, P<.01) were significant predictors of data-sharing behavior. The study also identified weight, frequency of provider visits, technological self-efficacy and frequent physical activity as predictors for higher wearable use.

Conclusions: Insights from this study are crucial for health care providers and policy makers aiming to leverage wearable technology to enhance health outcomes. Addressing the disparities and barriers identified can lead to more effective integration of these technologies in health care systems, thereby maximizing the potential of digital health tools to improve public health outcomes.

(J Med Internet Res 2025;27:e63879) doi:10.2196/63879

XSL•FO
RenderX

KEYWORDS

healthcare wearable devices; data-sharing behavior; willingness to share wearable data; activity trackers; wearable use; post-pandemic; wearables; healthcare delivery; disease detection; patient engagement; digital literacy; adults; United States; survey; cross-sectional survey; data sharing

Introduction

Health care wearable devices, encompassing a diverse array of technologies ranging from smartwatches, activity trackers, to biosensors, are revolutionizing the landscape of health care delivery [1]. These devices, which include skin-based wearables, biofluidic-based wearables, and drug delivery systems [2], are typically worn on or close to the body, gather physiological data, and provide actionable insights into an individual's health and well-being. By continuously monitoring vital signs, activity levels, and other biometric parameters, health care wearables offer opportunities for early disease detection, personalized treatment, and remote patient monitoring [3-5]. The potential of these devices to transform health care delivery is underscored by their ability to empower individuals to take proactive measures toward their health [6-8] and to facilitate health care professionals in delivering more personalized and timely interventions [9-11]. Furthermore, the integration of health data generated from wearable devices with mobile apps enhances their use, allowing for real-time tracking, monitoring, and analysis of health metrics [12]. With advancements in sensor technology, miniaturization, and data analytics, the market for health care wearables has witnessed exponential growth in recent years. The global market for health care wearables was valued at US $33.85 billion in 2023 and is projected to reach US $250 billion by 2030 [13], driven by increasing consumer demand for continuous health monitoring and the growing adoption of telehealth services.

Wearable health care devices have been found to be effective in gathering reliable health information for various conditions including cardiovascular diseases [14-17], chronic respiratory conditions like asthma, and chronic obstructive pulmonary disease [18-20], Parkinson disease [21-23], and mental health conditions such as stress, anxiety, depression, and behavioral changes [24-27]. Despite such promising benefits, adoption rates of wearable health care devices have varied across different demographics and health conditions [28-30]. While some individuals enthusiastically embrace these technologies for health monitoring purposes, others remain skeptical or face barriers such as cost, usability issues, and concerns regarding data privacy and security [31-33]. For instance, studies have indicated that older adults [34,35], individuals with lower income [36,37], and those with limited digital literacy [38-40] may experience greater difficulties in accessing and using wearable technology. In addition, questions have also been raised about the accuracy and reliability of data captured by wearable devices [14,41,42]. Therefore, to fully realize the potential benefits offered by wearable health care devices, effective sharing of wearable health data from patients to health care providers is critical.

Many studies have examined individuals' behavioral intentions to use wearable devices [43-46] and their willingness to share health data generated by wearables with health care providers [47,48]. Other studies have documented the willingness of individuals to share the data generated from their wearable health devices for research purposes [49-51]. However, despite documented intentions, there remains a notable gap in understanding the actual, actionable health data-sharing behaviors of users of health care wearable devices. This study aims to address this gap by investigating both self-reported willingness and actual behavior in sharing health data generated by wearable devices. We present findings based on a national survey conducted among US adults to ascertain the current landscape of health care wearable usage. We build upon and extend our earlier work [30] that examined wearable usage in the prepandemic period (2019) by using a recent dataset from 2022 to provide insights into the current use of wearable devices, with a particular focus on wearable data-sharing practices. In addition, we explore potential disparities in wearable use across socioeconomic demographics, health conditions, levels of technological self-efficacy, and engagement in physical activities. Furthermore, we examine whether these factors are associated with variations in individuals' willingness to share health data with health care providers and family or friends. Importantly, we assess the actual data-sharing behavior to identify predictors associated with the willingness-action gap and compare our findings with similar studies to highlight both similarities and differences (Figure 1).

This study leverages the most recent, nationally representative dataset of US adults, gathered in the later phases of COVID-19 pandemic (2022), a period marked by heightened health awareness and evolving health behaviors [52]. As we transition into the postpandemic era, understanding the use and data-sharing behaviors associated with wearable devices has become increasingly critical. The data collected during this period could offer valuable insights into shifts in health behaviors and highlights the role of wearable technology in managing and improving health engagement.

This study aims to answer the following questions:

1. How has wearable device usage among US adults changed since 2019, and what factors are associated with wearable use in the postpandemic era?
2. What are the key sociodemographic, health, and technological factors that influence US adults' willingness to share health data from wearable devices with health care providers, and with their family and friends?
3. How do actual data-sharing behaviors with health care providers compare with individuals' self-reported willingness to share wearable health data, and what are the key predictors of the actual data-sharing behavior?

XSL•FO
RenderX

Figure 1. Research model.

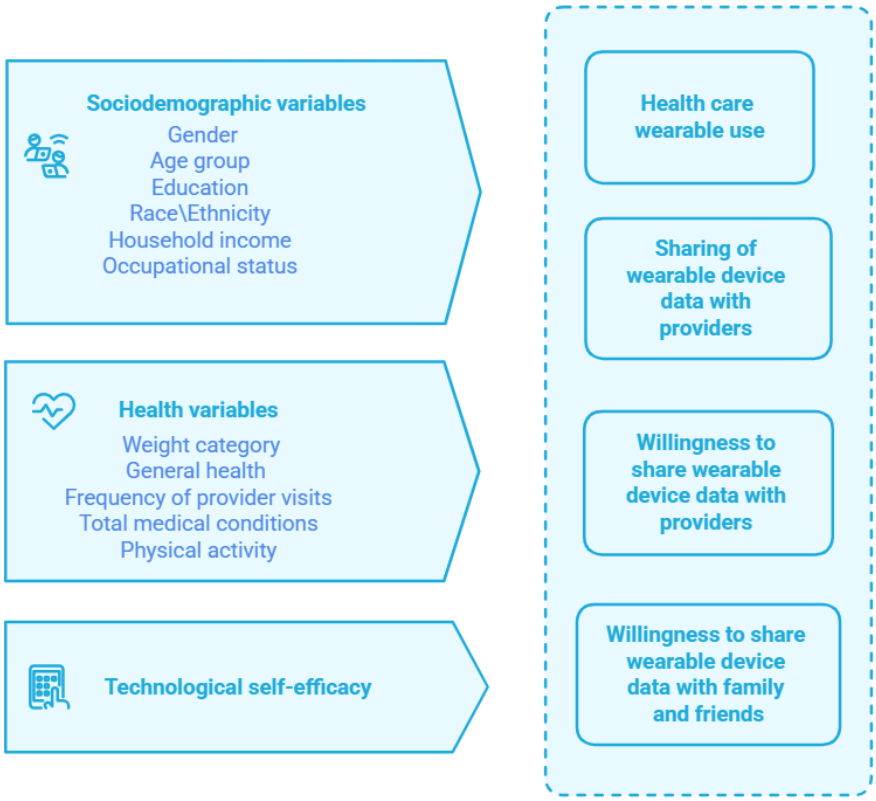

Methods

Sample and Data Collection

Data for this study were sourced from the sixth cycle of the Health Information National Trends Survey (HINTS 6), conducted by the National Cancer Institute from March through November 2022. HINTS 6 is a nationally representative, probability-based cross-sectional survey. In this cycle, adult, civilian, noninstitutionalized individuals were invited to participate through self-administered questionnaires, with respondents completing the survey either on paper or online. The survey used a nationally representative sampling strategy, targeting areas with high concentrations of minorities and those with lower minority concentrations, in addition to stratification by rural and urban geographic areas.

The survey was designed to collect data in 2 modes, giving respondents the option to complete the questionnaire through paper or through the web, with certain incentives offered for web participation. The sampling method continued to use a 2-stage design, initially selecting a stratified sample of addresses, followed by the selection of 1 adult per household from these addresses. This methodology ensures a broad representation of the US population, though the reliance on self-reported data may introduce certain biases in reporting and recall. To include respondents in the analytic sample, we implemented a 2-step filtering process: (1) respondents must have reported having a smart device, such as a smartphone or tablet, or multiple smart devices, and (2) they must have provided a response to the question regarding their use of a wearable device for tracking health or activity in the past 12 months. Specifically, only those smart device owners who answered the question, “In the last 12 months, have you used an electronic wearable device to monitor or track health or activity? For example, a Fitbit, Apple Watch, or Garmin Vivofit,” were included in the dataset for analysis. This approach resulted in 5591 completed survey responses, ensuring consistency with our research goals pertaining to current usage patterns of wearable health care devices and also into the data-sharing behaviors with health care providers and personal networks.

Variables and Measures

Outcome Variables

The primary focus of this study was on the use of wearable health care devices, assessed through a binary (yes or no) question asking respondents if they had used such an electronic wearable device to monitor or track their health or activity in the past 12 months. In addition, we also explored the responses about the frequency of wearable health care device use in the past month (everyday, almost every day, 1-2 times a week, <1 time a week, never used). In addition, the study explored respondents’ willingness to share data from their wearables with health care providers and with family and friends, also assessed as yes or no responses. To understand actual data-sharing behavior, we examined a survey question about whether respondents had shared health information from a wearable device with a health professional in the last 12 months.

Independent Variables

The study incorporated data from the survey questionnaire to capture respondents’ self-reported characteristics, which

XSL•FO
RenderX

included sociodemographic factors (such as age, gender, race/ethnicity, education, occupational status, and annual household income), health-related aspects (general health status, the total number of medical conditions, frequency of health care provider visits, and weight category), and technology-related factors (technology self-efficacy). General health was evaluated on a Likert scale from 1 (poor) to 5 (excellent). To quantify the total number of medical conditions, we counted 5 self-reported conditions such as diabetes or high blood sugar, hypertension, heart conditions (eg, heart attack, angina, or congestive heart failure), chronic lung diseases (such as asthma, emphysema, or chronic bronchitis), and mental health issues like depression or anxiety disorders. The values for total medical conditions ranged from 0 to 5. The frequency of health care provider visits was measured on a scale based on the number of visits in the past year, with values ranging from 0 (none) to 6 (10 times or more). Weight categories were determined based on BMI values: a BMI under 18.5 indicates underweight, and values between 18.5 and 25 indicates a normal healthy status, 25 to 29.9 is considered overweight, and a BMI of 30 or above is categorized as obese.

Previous research has demonstrated that technological self-efficacy, or comfort with technology, significantly influences the use of wearables [35,53,54]. Building on these findings, we measured respondents' technological self-efficacy based on their engagement in 4 internet-based health activities over the past year: searching for health information, electronic communication with health providers, accessing medical test results online, and scheduling appointments online. Each activity was scored as yes (1) or no (0), with total scores ranging from 0 to 4, indicating increasing comfort with technology.

We also explored the relationship between physical activity exercise routines and wearable device usage. Previous research suggests that individuals who are regularly motivated to exercise and engage in physical activities are more inclined to adopt and use wearable devices [55,56]. To assess the physical activity levels of respondents, we examined a question asking how many days per week they engage in any physical activity or exercise of at least moderate intensity (0-7 days a week). Examples provided included brisk walking, bicycling at a regular pace, and swimming at a regular pace.

Statistical Analysis

All the statistical analyses were performed using STATA software (version 18; StataCorp). We began with a descriptive analysis of our data sample. To explore the associations between the use of wearable health care devices and various sociodemographic predictors, we generated crosstab tables and evaluated them using the Wald chi-square test. This approach was repeated to analyze the willingness to share wearable health data with providers, family, and friends, as well as the actual sharing of such data with health care providers.

Given that our primary variables of interest were all binary, and our predictor variables included both categorical and continuous types, logistic regression analysis was used (Figure 1). We ran 4 survey logistic regression models (using svy: logistic command in STATA) with the following outcome variables: (1) use of wearable health care device (0/1), (2) willingness to share health care wearable data with providers (0/1), (3) willingness to share health care wearable data with family and friends (0/1), and (4) actual sharing of wearable data with providers (0/1). The predictor variables were the sociodemographic factors (gender, age group, education level, race/ethnicity, annual household income, and occupation status), health related variables (weight category based on BMI, self-reported general health, frequency of provider visits, total medical conditions, and extent of physical activity) and technology self-efficacy. We calculated adjusted odds ratios (ORs) and CIs for these predictors. To account for the complex survey design, we applied weights with jackknife replications in our statistical analyses, using *svy:* commands in STATA. Since the svy: logistic model in STATA does not assess multicollinearity, we initially ran standard logistic regressions and examined variance inflation factor values with a cutoff value of 4, before performing svy: logistic regression assessments.

Ethical Considerations

HINTS data are publicly available and collected as deidentified information from participants, specifically designed for research purposes [57]. The HINTS 6 survey was classified as "exempt research" under 45 CFR 46.104 and approved by the Westat institutional review board (IRB) on May 10, 2021 (Project #6632.03.51), with an amendment approved on November 24, 2021 (Amendment ID #3597). It also received a "Not Human Subjects Research" determination from the National Institutes of Health (NIH) Office of IRB Operations on August 16, 2021 (iRIS reference #562715). Therefore, no additional approvals were required from authors' institutions. Further details on the survey instrument, and the methodology and data handling for HINTS 6 can be explored through the National Cancer Institute's official resources [58,59].

Results

Overview

Table 1 describes the characteristics of survey respondents. A total of 5591 respondents had indicated having a smart device that is needed for storing and sharing data from a health care wearable device, in addition to responding to the question about use of a wearable device. A total of 3161 out of 5591 (60.45%) were female, 3405 out of 5591 (61.55%) were aged 50 years or older, 2600 out of 5591 (49.67%) had a college degree or higher, and 2479 out of 5591 (44.34%) were employed full-time. Among the respondents, 2925 out of 5591 (57.97%) identified their race\ethnicity as White, 785 out of 5591 (15.56%) as Black, 893 out of 5591 (17.7%) as Hispanic, and 271 out of 5591 (5.37%) as Asian. Furthermore, 2100 out of 5591 (42.24%) indicated their annual household incomes to be US $75,000 or more.

XSL•FO
RenderX

Table 1. Socioeconomic demographic characteristics of survey respondents.

Characteristics	Survey respondents, n (%)
Gender (n=5229)	
Male	2068 (39.55)
Female	3161 (60.45)
Age group (years; n=5532)	
18-34	927 (16.76)
35-49	1200 (21.69)
50-64	1639 (29.63)
65-74	1168 (21.11)
75+	598 (10.81)
Education (n=5235)	
Less than high school	268 (5.12)
High school	849 (16.22)
Some college	1518 (29)
College graduate or more	2600 (49.67)
Race\Ethnicity (n=5046)	
Non-Hispanic White	2925 (57.97)
Non-Hispanic African American	785 (15.56)
Hispanic	893 (17.7)
Non-Hispanic Asian	271 (5.37)
Others	172 (3.41)
Household income (US $; n=4972)	
< 20,000	719 (14.46)
20,000 to <35,000	613 (12.33)
35,000 to <50,000	666 (13.4)
50,000 to <75,000	874 (17.58)
> 75,000	2100 (42.24)
Occupational status (n=5591)	
Employed full-time	2479 (44.34)
Employed part-time	366 (6.55)
Not employed or retired	2768 (49.11)

Health Care Wearable Use and Frequency of Use

Out of 5591 respondents, 2033 (36.36%) reported using a health care wearable device in the past 12 months (Table 2). In the exploratory analysis, we observed statistically significant differences in wearable use (Table 3) across age groups (χ^2_4=261.7, *P*<.01), gender (χ^2_1=25.7, *P*<.01), education (χ^2_3=170.6, *P*<.01), occupation status (χ^2_2=143.1, *P*<.01), race\ethnicity (χ^2_4=13.7, *P*<.01) and income levels (χ^2_4=290.1, *P*<.01; Figure 2). When asked about health care wearable use in the past month, 880 out of 5591 (43.48%) reported daily use, 557 out of 5591 (27.52%) used them almost every day, 217 out of 5591 (10.72%) used them 1-2 times per week, and 156 out of 5591 (7.71%) used them less than once a week. Notably, 214/5591 (10.57%) indicated they had not used their wearables at all in the past month.

XSL•FO
RenderX

Table 2. Usage, willingness, and actual sharing of wearable health data.

Variable	Response	
	Yes, n (%)	No, n (%)
Health care wearable device use (n=5591)	2033 (36.4)	3558 (63.6)
Health information sharing from wearable device (n=2020)	535 (26.5)	1485 (73.5)
Willingness to share wearable health data with provider (n=2020)	1584 (78.4)	436 (21.6)
Willingness to share wearable health data with family and friends (n=2014)	1266 (62.9)	748 (37.1)

Table 3. Variations in wearable use based on demographics.

Respondent demographics	Use of wearable health care device in the past 12 months (n=5591)			
	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* values
Gender			25.7 (1)	<.01
Male	654 (12.51)	1414 (27.04)		
Female	1217 (23.27)	1944 (37.18)		
Age group (years)			267.7 (4)	<.01
18-34	463 (8.37)	463 (8.37)		
35-49	579 (10.47)	621 (11.23)		
50-64	549 (9.93)	1090 (19.71)		
65-74	298 (5.39)	870 (15.73)		
75+	130 (2.35)	468 (8.46)		
Education			170.6 (3)	<.01
Less than high school	47 (0.9)	221 (4.22)		
High school	193 (3.78)	651 (12.44)		
Some college	494 (9.44)	1024 (19.56)		
College graduate or more	1131 (21.6)	1469 (28.06)		
Race\ethnicity			13.7 (4)	<.01
Non-Hispanic White	1065 (21.11)	1860 (36.86)		
Non-Hispanic African American	254 (5.03)	531 (10.52)		
Hispanic	331 (6.56)	562 (11.14)		
Non-Hispanic Asian	120 (2.38)	151 (2.99)		
Others	57 (1.13)	115 (2.28)		
Household income (US $)			290.1 (4)	<.01
< 20k	128 (2.57)	591 (11.89)		
20k to <35k	152 (3.06)	461 (9.27)		
35k to <50k	197 (3.96)	469 (9.43)		
50k to <75k	312 (6.28)	562 (11.30)		
> 75k	1019 (20.49)	1081 (21.74)		
Occupation status			143.2 (2)	<.01
Employed full time	1113 (19.91)	1366 (24.43)		
Employed part time	123 (2.2)	243 (4.35)		
Not employed or retired	797 (14.26)	1949 (34.86)		

XSL•FO
RenderX

Figure 2. Health care wearable use across different demographic segments.

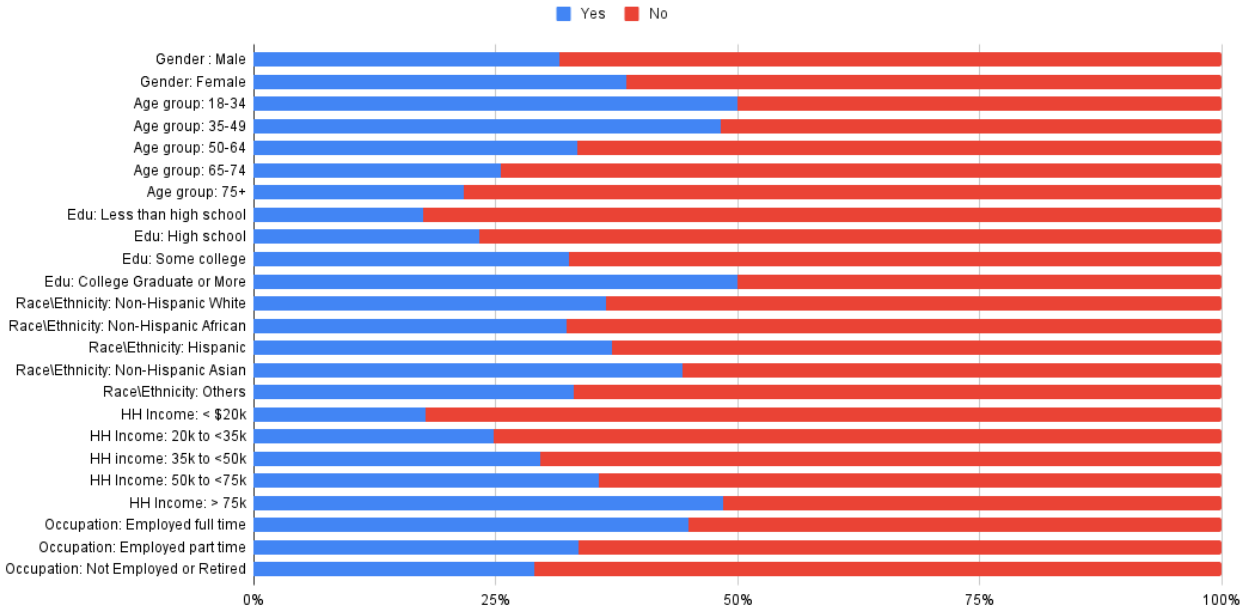

Wearable Data Sharing: Willingness and Sharing Behavior

Out of 2020 wearable users, 1584 (78.4%) expressed willingness to share data from health care wearables with their providers, and 1266 out of 2014 (62.95%) users indicated willingness to share health data with family or friends. Despite this willingness, only 535 out of 2020 (26.5%) users actually shared data from wearable devices with providers within the past 12 months (Table 2). This highlights a significant gap between the intention to share health data and actual sharing behavior. Table 4 reveals demographic disparities in health data sharing behaviors based on exploratory chi-square tests: gender (χ^2_1=4.08, P=.04), age groups (χ^2_4=50.45, P<.01), race/ethnicity (χ^2_4=12.79, P<.01), and occupational status (χ^2_2=18.56, P<.01). Differences in the willingness to share wearable data with health providers were significant only by race/ethnicity (χ^2_4=10.35, P<.03). Similarly, the willingness to share data with family and friends showed significant variation across different age groups (χ^2_4=20.44, P<.01) and occupational status (χ^2_2=5.64, P<.05).

XSL•FO
RenderX

Table 4. Variations in data sharing and willingness to share wearable data based on demographics.

Respondent demographics	Willingness to share wearable data with health care provider (n=2023)				Willingness to share wearable data with family or friends (n=2014)				Actual sharing of health data with health professional in the past 12 months (n=2020)			
	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value
Gender			1.3 (1)	<.25			0.2 (1)	<.68			4.1 (1)	<.05
Male	520 (27.96)	129 (6.94)			403 (21.73)	244 (13.15)			190 (10.22)	463 (24.89)		
Female	943 (50.7)	268 (14.41)			764 (41.19)	444 (23.94)			299 (16.08)	908 (48.82)		
Age group (years)			4.9 (4)	<.29			20.4 (4)	<.01			50.4 (4)	<.01
18-34	355 (17.69)	105 (5.23)			317 (15.83)	143 (7.14)			92 (4.59)	367 (18.3)		
35-49	445 (22.17)	132 (6.58)			384 (19.18)	193 (9.64)			122 (6.08)	454 (22.63)		
50-64	426 (21.23)	119 (5.93)			311 (15.53)	233 (11.64)			156 (7.78)	392 (19.54)		
65-74	244 (12.16)	52 (2.59)			173 (8.64)	121 (6.04)			103 (5.13)	194 (9.67)		
75+	106 (5.28)	23 (1.15)			77 (3.85)	50 (2.5)			56 (2.79)	70 (3.49)		
Education			4.8 (3)	<.19			1.9 (3)	<.58			4.6 (3)	<.21
Less than high school	34 (1.83)	13 (0.7)			25 (1.35)	20 (1.08)			9 (0.48)	37 (1.99)		
High school	145 (7.8)	50 (2.69)			122 (6.58)	72 (3.88)			43 (2.31)	152 (8.18)		
Some college	382 (20.55)	110 (5.92)			300 (16.18)	190 (10.25)			142 (7.64)	350 (18.83)		
College Graduate or More	900 (48.41)	225 (12.1)			717 (38.67)	408 (22.01)			295 (15.87)	831 (44.7)		
Race\ethnicity			10.3 (4)	<.03			6.9 (4)	<.14			12.8 (4)	<.01
Non-Hispanic White	855 (47.06)	206 (11.34)			682 (37.6)	378 (20.84)			279 (15.35)	781 (42.98)		
Non-Hispanic African American	193 (10.62)	58 (3.19)			140 (7.72)	110 (6.06)			81 (4.46)	171 (9.41)		
Hispanic	259 (14.25)	71 (3.91)			205 (11.3)	124 (6.84)			69 (3.8)	260 (14.31)		
Non-Hispanic Asian	81 (4.46)	37 (2.04)			73 (4.02)	45 (2.48)			26 (1.43)	93 (5.12)		
Others	47 (2.59)	10 (0.55)			39 (2.15)	18 (0.99)			20 (1.1)	37 (2.04)		
Household income (US $)			0.9 (4)	<.91			6.9 (4)	<.13			6.9 (4)	<.14
<20,000	98 (5.45)	30 (1.67)			67 (3.74)	59 (3.29)			42 (2.34)	82 (4.56)		
20,000 to <35,000	118 (6.57)	34 (1.89)			93 (5.19)	59 (3.29)			42 (2.34)	110 (6.12)		
35,000 to <50,000	154 (8.57)	41 (2.28)			129 (7.19)	65 (3.63)			57 (3.17)	139 (7.73)		
50,000 to <75,000	249 (13.86)	61 (3.39)			200 (11.15)	109 (6.08)			73 (4.06)	236 (13.13)		

XSL•FO
RenderX

Respondent demographics	Willingness to share wearable data with health care provider (n=2023)				Willingness to share wearable data with family or friends (n=2014)				Actual sharing of health data with health professional in the past 12 months (n=2020)			
	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value	Yes, n (%)	No, n (%)	Chi-square (*df*)	*P* value
> 75,000	799 (44.46)	213 (11.85)			643 (35.86)	369 (20.58)			252 (14.02)	765 (42.55)		
Occupation status			0.4 (2)	<.80			5.6 (2)	<.05			18.6 (2)	<.1
Employed full time	862 (42.67)	245 (12.13)			712 (35.35)	394 (19.56)			258 (12.77)	852 (42.18)		
Employed part time	97 (4.8)	26 (1.29)			84 (4.17)	39 (1.94)			37 (1.34)	96 (4.75)		
Not employed or retired	625 (30.94)	165 (8.17)			470 (23.34)	315 (15.64)			250 (12.38)	537 (26.58)		

Predictors of Health Care Wearable Use

Table 5 presents results from a logistic regression analysis examining predictors of health care wearable use. Women, compared with men, are more likely to use health care wearables (OR 1.49, 95% CI 1.17-1.9, *P*<.01). The likelihood of using health care wearables declines with age; compared with the 18-34 years age group, the 50-64 years (OR 0.49, 95% CI 0.36-0.67, *P*<.01), 65-74 years (OR 0.37, 95% CI 0.26-0.54, *P*<.01), and over 75 years (OR 0.35, 95% CI 0.20-0.61, *P*<.01) age groups showed significantly lower odds. Higher annual household incomes are also associated with increased likelihood of wearable use, notably in the US $50,000 to US $75,000 range (OR 2.65, 95% CI 1.42-4.93, *P*<.01) and above US $75,000 (OR 3.21, 95% CI 1.71-5.97, *P*<.01). Weight categories, as assessed by BMI significantly influenced wearable use. Overweight individuals (BMI 25-29.9) were more likely to use wearables compared with those with normal weight (BMI 18.5-24.9; OR 1.36, 95% CI 0.98-1.89, *P*<.05). Conversely, underweight individuals (BMI<18.5) were less likely to use wearables (OR 0.31, 95% CI 0.09-1.04, *P*<.05). Obese individuals (BMI≥30) also showed higher odds of use (OR 1.33, 95% CI 0.93-1.90, *P*<.10), though this was only significant at the 10% level.

Among the health-related variables, the frequency of provider visits was significantly associated with an increase in wearable use (OR 1.08, 95% CI 1.02-1.15, *P*<.01), whereas neither self-rated general health nor the number of medical conditions were significant predictors. In addition, our analysis indicated that individuals who exercised more frequently were more likely to use wearables (OR 1.15, 95% CI 1.11-1.20, *P*<.01).

As reported in Table 5, our findings indicated a strong association between technological self-efficacy and wearable use (OR 1.29, 95% CI 1.15-1.44, *P*<.01).

XSL•FO
RenderX

Table 5. Factors associated with health care wearable use: results of logistic regression.

Predictors and subcategory	Odds ratio (95% CI)	*P* value
Gender (reference category: male)		
Female	1.49 (1.17-1.90)	.01
Age group (years; reference category: 18-34 years)		
35-49	1.09 (0.79-1.51)	.57
50-64	0.49 (0.36-0.67)	<.001
65-74	0.37 (0.26-0.54)	<.001
75+	0.35 (0.2-0.61)	<.001
Education (reference category: college graduate or higher)		
Less than high school	0.78 (0.38-1.62)	.50
High school	0.76 (0.54-1.07)	.11
Some college	0.89 (0.66-1.19)	.43
Race/ethnicity (reference category: non-Hispanic African American)		
Non-Hispanic White	1.19 (0.87-1.65)	.25
Hispanic	1.38 (0.99-1.91)	.06
Non-Hispanic Asian	1.39 (0.67-2.85)	.36
Others	1.15 (0.59-2.22)	.67
Household income (US $; reference category: less than US $20,000)		
20,000 to <35,000	1.61 (0.82-3.13)	.15
35,000 to <50,000	1.58 (0.84-2.96)	.15
50,000 to <75,000	2.65 (1.42-4.93)	.01
75,000 or more	3.21 (1.71-5.97)	.01
Occupational status (reference category: not employed or retired)		
Employed full-time	1.02 (0.79-1.31)	.85
Employed part-time	0.89 (0.51-1.54)	.68
Weight category (reference category: normal weight)		
Underweight	0.31 (0.09-0.84)	.05
Overweight	1.36 (1.08-1.89)	.05
Obesity	1.33 (0.93-1.9)	.11
Health		
General health	1.02 (0.87-1.21)	.77
Frequency of provider visits	1.08 (1.02-1.15)	.01
Total medical conditions	0.96 (0.83-1.11)	.58
Physical activity (times exercise)	1.15 (1.11-1.2)	.01
Technology self-efficacy		
Electronic usage	1.29 (1.15-1.44)	.01

Predictors of Health Data Sharing and Willingness to Share Wearable Data

A series of 3 logistic regressions were used to explore the relationships between various predictors and both the behavior of sharing health data and the willingness to share data from wearable health devices with providers, family, and friends. These findings are detailed in Table 6.

While women were more likely to use health care wearable devices compared with men, they were less likely to share the data from these devices with providers (OR 0.66, 95% CI 0.44-0.99, $P<.05$). No significant associations were found between gender and the willingness to share wearable data with either providers or family and friends.

XSL•FO
RenderX

Table 6. Key predictors of wearable data sharing behavior and willingness to share wearable data: logistic regression results.

Predictors and subcategory	Wearable health data sharing with provider		Willingness to share wearable data with provider		Willingness to share wearable data with family and friends	
	Odds ratio (95% CI)	*P* values	Odds ratio (95% CI)	*P* values	Odds ratio (95% CI)	*P* values
Gender (Ref: male)						
Female	0.66 (0.44-0.99)	.04	0.80 (0.54-1.2)	.27	1.13 (0.83-1.54)	.42
Age group (years; ref: 18-34 years)						
35-49	1.08 (0.61-1.92)	.80	1.14 (0.66-1.99)	.63	0.68 (0.4-1.18)	.17
50-64	1.47 (0.82-2.64)	.19	0.94 (0.48-1.86)	.85	0.39 (0.22-0.71)	.01
65-74	1.64 (0.74-3.63)	.22	1.26 (0.48-3.28)	.63	0.35 (0.21-0.59)	.01
75+	4.14 (1.1-14.6)	.03	2.52 (0.74-8.64)	.14	0.46 (0.23-0.93)	.03
Education (ref: college graduate or higher)						
Less than High School	0.95 (0.26-3.45)	.94	1.03 (0.33-3.29)	.95	0.72 (0.1-4.98)	.73
High School	0.78 (0.36-1.71)	.53	0.66 (0.3-1.46)	.3	1.11 (0.6-2.06)	.73
Some College	1.36 (0.91-2.02)	.13	1.15 (0.64-2.06)	.64	1.14 (0.8-1.65)	.46
Race/ethnicity (ref: Non-Hispanic African American)						
Non-Hispanic White	0.53 (0.25-1.14)	.10	1.74 (1.08-3.09)	.05	1.81 (1.04-3.16)	.04
Hispanic	0.44 (0.2-0.97)	.05	1.92 (1.02-3.62)	.04	1.34 (0.74-2.42)	.32
Non-Hispanic Asian	0.41 (0.13-1.23)	.11	0.87 (0.39-1.92)	.72	1.02 (0.45-2.31)	.95
Others	0.66 (0.21-2.03)	.46	2.57 (0.6-11.09)	.20	2 (0.8-5.02)	.13
Household income (US $; ref: Less than US $20,000)						
20,000 to <35,000	0.34 (0.08-1.35)	.12	0.96 (0.29-3.12)	.94	0.98 (0.39-2.46)	.95
35,000 to <50,000	0.45 (0.12-1.65)	.22	0.55 (0.2-1.53)	.25	1.33 (0.51-3.42)	.55
50,000 to <75,000	0.28 (0.08-0.94)	.04	0.93 (0.32-2.69)	.89	1.1 (0.41-2.91)	.84
75,000 or more	0.31 (0.09-0.99)	.05	0.60 (0.22-1.67)	.32	1.02 (0.45-2.27)	.96
Occupational status						
Employed Full time	0.87 (0.49-1.55)	.63	0.83 (0.47-1.45)	.51	0.97 (0.63-1.51)	.90
Employed Part time	0.88 (0.32-2.41)	.80	1.38 (0.53-3.61)	.50	1.78 (0.87-3.61)	.11
Weight category						
Underweight	4.29 (0.66-27.56)	.12	1.92 (0.51-7.30)	.33	0.82 (0.23-2.94)	.76
Overweight	1.34 (0.85-2.11)	.20	1.34 (0.79-2.26)	.28	0.97 (0.61-1.55)	.90
Obesity	1.42 (0.89-2.27)	.14	1.40 (0.85-2.3)	.19	0.96 (0.61-1.52)	.86
Health						
General Health	1.16 (0.84-1.60)	.35	1.10 (0.85-1.42)	.46	1.21 (0.97-1.52)	.09
Frequency of Provider Visits	1.23 (1.08-1.39)	.01	1.29 (1.08-1.43)	.04	1.10 (0.98-1.23)	.15
Total Medical Conditions	1.35 (1.05-1.73)	.01	1.06 (0.87-1.29)	.53	1.06 (0.84-1.32)	.54
Physical Activity (Times Exercise)	1.09 (0.98-1.21)	.10	0.99 (0.91-1.07)	.81	1.08 (1.07-1.17)	.05
Technology self-efficacy						
Electronic Usage	1.26 (0.92-1.71)	.14	1.13 (0.94-1.36)	.17	0.94 (0.8-1.11)	.46

Older adults aged 75 years and above showed higher odds of sharing wearable device data with providers (OR 4.14, 95% CI 1.18-14.6, *P*<.01) compared with the younger (18-34 years) age group. Although no significant differences were noted among age groups in their willingness to share wearable data with providers, a negative association was observed in their willingness to share with family and friends. Specifically, individuals aged 50-64 years (OR 0.39, 95% CI 0.22-0.71, *P*<.01), 65-74 years (OR 0.35, 95% CI 0.23-0.93, *P*<.01), and older than 75 years (OR 0.46, 95% CI 0.23-0.93, *P*<.05) were less likely to share health data from wearable devices with family or friends.

XSL•FO
RenderX

Compared with non-Hispanic Black respondents, Hispanic respondents (OR 1.92, 95% CI 1.02-3.62, *P*<.05) and non-Hispanic White respondents (OR 1.74, 95% CI 1.08-3.09, *P*<.05) exhibited higher odds of willingness to share health data from wearables with providers. Furthermore, non-Hispanic White respondents showed a greater tendency to share wearable data with family and friends (OR 1.81, 95% CI 1.04-3.16, *P*<.05). However, in comparison with non-Hispanic Black respondents, actual sharing behavior was lower for Hispanic respondents (OR 0.44, 95% CI 0.20-0.97, *P*<.05). Altogether, these results indicate racial disparities in the actual sharing of health data despite a higher willingness.

No significant differences were observed in the willingness to share health data with providers, as well as family and friends, based on annual household income, yet the likelihood of sharing health data with providers decreased with higher income levels. Those earning between US $50,000 and US $75,000 (OR 0.28, 95% CI 0.08-0.94, *P*<.05), and those with incomes above US $75,000 (OR 0.31, 95% CI 0.09-0.99, *P*<.05) showed lesser odds of sharing.

Frequency of provider visits emerged as a significant predictor for both willingness to share and actual sharing behavior with providers. Those who visited their providers more frequently were more willing to share data from wearable devices (OR 1.29, 95% CI 1.08-1.43, *P*<.05) and more actively engaged in sharing this data (OR 1.23, 95% CI 1.08-1.39, *P*<.01). Individuals with more medical conditions also showed higher odds of sharing data with providers (OR 1.35, 95% CI 1.05-1.73, *P*<.01). A weak association was observed between self-reported general health and the willingness to share data with family and friends (OR 1.21, 95% CI 0.97-1.52, *P*<0.1). In addition, individuals engaging in more exercise and physical activities were found more likely to share wearable health care data with family and friends (OR 1.08, 95% CI 1-1.17, *P*<.05).

Discussion

Principal Findings

Wearable health care devices such as smartphones and activity trackers offer continuous and real-time monitoring of an individual's health metrics. These tools are becoming increasingly popular as they not only empower users to take proactive control over their health but also provide health care providers with invaluable data for better patient care. Despite their potential, significant gaps remain in understanding the predictors that influence wearable device use, and data-sharing behaviors associated with these devices. This study aimed to address these gaps by identifying key factors that influence both the use of wearable devices, the willingness to share the data they collect and actual data-sharing behavior. We observed an increase in wearable device usage from 28%-30% in 2019 to 36.36% in 2022, reflecting a broader trend of heightened health awareness, particularly following the COVID-19 pandemic. In addition, despite a high willingness to share wearable data with health care providers and family or friends, there is a significant gap between willingness and actual sharing behaviors, with only a quarter of users actively sharing data with providers.

Our results highlight a significant uptick in the adoption and use of health care wearable devices, from 28%-30% in 2019 [30,60] and 21% according to a 2020 Pew Research Study [61], to 36.36% in 2022. Our findings from HINTS 6 dataset, collected during a period of accelerating vaccinations and evolving public health measures, reflect how the COVID-19 pandemic has increased the adoption of wearable devices and shifted consumer attitudes toward self-health monitoring. The rise in wearable usage in 2022 signifies a growing health consciousness and commitment to self-health monitoring, suggesting that wearable technology will be crucial in fostering ongoing health engagement and preventive care in the postpandemic landscape [62,63].

Our analysis found that 78.4% of US adults were willing to share data from wearable devices with providers, and 62.95% were willing to share with family and friends. However, only 26.5% actually shared data with providers. This willingness-action gap may stem from factors such as privacy concerns and technological challenges. Notably, there is a decrease in willingness compared with previous years; studies from 2019 reported that 81.85% of respondents were willing to share wearable data with providers, and 69.51% with family or friends [47]. Other studies have reported even higher levels of willingness [30,48], particularly among subgroups such as cardiovascular patients and older individuals [35,64]. This decline in willingness suggests that while interest in wearable technology remains high, barriers to data sharing may have increased, hindering the translation of willingness into actual sharing behavior with providers.

Our analysis also reveals that wearable use is not uniform across demographic segments. Notably, women are more likely than men to use these devices. This trend is partly attributed to women's greater health consciousness and proactive engagement in health monitoring and management activities [65,66]. Previous research has also found that women with one or more chronic conditions are more likely to use wearable devices [67]. However, we also find that women are less likely to share data from wearable devices with health providers. This reluctance, relatively more in women, may stem from privacy concerns, fear of data misuse, and lack of trust in how their health information will be handled [68,69]. It is essential to develop targeted education and intervention programs for women that emphasize benefits of data sharing for personalized health care and improved medical outcomes. In addition, ensuring robust data security measures and transparent communication about how data will be used can help build trust and encourage more individuals to share their health information from wearable devices.

We also find that wearable usage decreases with age, with individuals over 75 showing the least likelihood of using wearables. This finding is consistent with previous studies [35]. This decline could be attributable to varying levels of technological comfort or perceived use and usability of the devices among the older adults [34,70]. Furthermore, older adults could find the devices to be too complex to use and integrate into their daily lives [71]. However, we also found older adults to be less willing to share wearable device data with family and friends but were actually more engaged in

XSL•FO
RenderX

sharing this data with their providers. This behavior suggests that older adults may prioritize sharing health information with professionals for medical oversight and care management, highlighting a trust in professional health care settings over personal networks.

Racial and ethnic disparities also manifest in wearable device use, with Hispanic individuals showing a greater propensity to share data from these devices compared with non-Hispanic Black individuals. This could reflect cultural differences in health management practices or disparities in access to technology. Our findings are consistent with previous studies that have documented increased interest in wearable devices among Hispanic communities, especially after the COVID-19 pandemic [72]. Challenges in obtaining health resources and general distrust in health care infrastructure and systems have also motivated Hispanic individuals to look at wearable devices to better monitor their health care. Hispanic individuals have also been engaging more with electronic tools to acquire health information using online resources and for communicating with providers [73]. While non-Hispanic White and Hispanic individuals showed a higher willingness to share data with providers, their actual sharing rates did not align with these intentions, indicating possible systemic or personal barriers that prevent translating willingness into action.

Our findings reveal that economic factors significantly influence wearable device usage, with individuals having higher annual household incomes being more likely to use these technologies, likely due to greater access to and affordability of emerging health technologies. Interestingly, despite their higher use rates, these individuals are less inclined to share wearable data with health care providers, possibly due to heightened privacy concerns [74,75] or a greater sense of autonomy in managing their health data independently.

Our analysis extends into health-related variables and conditions influencing wearable adoption. Frequent visits to health care providers and the presence of multiple health conditions are associated with higher likelihood of wearable usage and sharing of wearable data with providers, suggesting that individuals more engaged with their health are more receptive to using technology to manage their health. Our findings complement previous studies that have documented wearable use among patients with cardiovascular conditions [76], and diabetes [77,78].

Furthermore, our findings suggest that individuals who regularly engage in physical activities are not only more likely to use wearables but are also more willing to share the data generated, especially with family and friends potentially to monitor their fitness progress or health status more effectively. Many users share their physical activities with a group of friends using fitness and health apps, which helps to boost motivation and receive positive reinforcement [79,80]. These social features and gamification in wearable health apps allow users to join challenges, share achievements, and provide mutual support, significantly enhancing adherence to fitness goals and creating a sense of community [81].

Implications

The findings from this study on wearable health care devices suggest several implications for health care providers, policy makers, regulators, and industry stakeholders. Wearable devices can provide health data relevant to users' specific health needs and concerns, yet access to these benefits is not equally distributed among all users. Those with greater digital literacy and socioeconomic resources tend to have better access to the health monitoring and predictive capabilities of wearable technologies.

Health care providers can enhance wearable data sharing by addressing patient concerns, particularly among women and older adults, through targeted education and communication that underscores the benefits of sharing health wearable data. There is also a need for simplified training programs to increase wearable device use among older adults. Policy makers and regulators should focus on improving accessibility and affordability for lower-income individuals, ensuring robust data protection standards, and promoting interoperability between wearable devices and health systems to build trust and use. For racial and ethnic minorities, culturally tailored programs that address disparities in technology access and use are crucial. In addition, health insurers and technology companies could offer incentives for wearable data sharing and develop features that enhance user engagement through social connectivity and gamification, fostering a community-focused approach to health management. Collectively, these efforts can lead to better integration of wearable technologies in health care, providing valuable insights for preventive health measures and personalized care, thereby maximizing the potential of digital health tools in improving public health outcomes.

Limitations

This study has several limitations. Primarily, it relies on self-reported information by survey respondents, which can be subject to recall bias and social desirability bias. Respondents may not accurately remember past behaviors or may portray themselves in a more favorable light. For some variables, which may introduce inherent biases. In addition, the cross-sectional design of the survey restricts the ability to establish causality between variables. Since wearable usage may change over time, this type of study does not capture the dynamic changes in patterns of wearable use over time. We were also limited by the data already gathered by HINTS. Further research could explore reasons for nonuse or nonsharing of wearable data, and could examine additional questions pertaining to privacy, trust and design issues.

Conclusion

This study highlights critical insights into the use and data-sharing behaviors associated with wearable health care technology. We have detailed the latest trends in the adoption and use of wearable devices, outlining the frequency and demographics of users. In addition, we explored how US adults engage in sharing data from wearable devices, including their willingness to share this information with health care providers and with family and friends. Our study noted major gaps between willingness to share health care wearable data, and the

XSL•FO
RenderX

actual sharing behavior with providers. By identifying the sociodemographic, health and technological factors that influence both the use of wearables and the willingness to share data, our findings underscore the necessity for tailored interventions. Clinicians can enhance patient engagement by addressing the specific barriers identified, particularly among demographics with lower sharing rates. Policy makers should prioritize initiatives that promote accessibility and digital literacy, ensuring equitable benefits from wearable technologies. These targeted approaches can help bridge the gap between users' willingness to share health data and their actual sharing behaviors, ultimately advancing the integration of wearable devices in personalized health care.

Looking forward, this study lays the groundwork for future research aimed at enhancing the design of digital health interventions and improving the integration of wearable technology in clinical practices. By continuing to investigate the evolving dynamics of user engagement and the barriers to data sharing, we can better leverage wearable technology to advance personalized health care and public health outcomes.

Data Availability

The Health Information National Trends Survey 6 (HINTS 6) data is publicly available and seeks to provide a comprehensive view of health communication trends among US adults. Conducted by the National Cancer Institute, this data collection focuses on how people access and use health-related information and behaviors. The complete datasets, including methodological reports and questionnaires, are accessible on the HINTS website [82].

Authors' Contributions

MST contributed to data cleaning and data analysis. EV managed study conceptualization, study design and writing. RC handled study design, conceptualization, data analysis, and writing.

Conflicts of Interest

None declared.

References

1. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
2. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. npj Flex Electron 2021;5(1):1-14. [doi: 10.1038/s41528-021-00107-x]
3. Guk K, Han G, Lim J, Jeong K, Kang T, Lim E, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel) 2019;9(6):813 [FREE Full text] [doi: 10.3390/nano9060813] [Medline: 31146479]
4. Mattison G, Canfell O, Forrester D, Dobbins C, Smith D, Töyräs J, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022;24(7):e36690 [FREE Full text] [doi: 10.2196/36690] [Medline: 35776492]
5. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillian B, Bower P, et al. Interventions using wearable physical activity trackers among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open 2021;4(7):e2116382 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.16382] [Medline: 34283229]
6. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
7. Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of wearable trackers on physical activity in healthy adults: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020;8(7):e15576 [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]
8. Jin D, Halvari H, Maehle N, Olafsen AH. Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. Behav. Inf. Technol 2020;41(2):242-261. [doi: 10.1080/0144929x.2020.1801840]
9. Volpato L, Del Río Carral M, Senn N, Santiago Delefosse M. General practitioners' perceptions of the use of wearable electronic health monitoring devices: qualitative analysis of risks and benefits. JMIR Mhealth Uhealth 2021;9(8):e23896 [FREE Full text] [doi: 10.2196/23896] [Medline: 34383684]
10. Linton SC, De Boer C, Tian Y, Alayleh A, Bouchard ME, Figueroa A, et al. Effect of consumer-grade wearable device data on clinician decision making during post-discharge telephone calls after pediatric surgery. J Pediatr Surg 2022;57(9):137-142. [doi: 10.1016/j.jpedsurg.2021.09.040] [Medline: 34732297]
11. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021;4(1):45 [FREE Full text] [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]

XSL•FO
RenderX

12. Moorthy P, Weinert L, Schüttler C, Svensson L, Sedlmayr B, Müller J, et al. Attributes, methods, and frameworks used to evaluate wearables and their companion mHealth apps: scoping review. JMIR Mhealth Uhealth 2024;12:e52179 [FREE Full text] [doi: 10.2196/52179] [Medline: 38578671]
13. Grand View Research. URL: https://www.grandviewresearch.com/industry-analysis/wearable-medical-devices-market [accessed 2025-02-04]
14. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
15. Sarhaddi F, Kazemi K, Azimi I, Cao R, Niela-Vilén H, Axelin A, et al. A comprehensive accuracy assessment of Samsung smartwatch heart rate and heart rate variability. PLoS One 2022;17(12):e0268361 [FREE Full text] [doi: 10.1371/journal.pone.0268361] [Medline: 36480505]
16. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021;18(8):581-599. [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
17. Strik M, Ploux S, Ramirez FD, Abu-Alrub S, Jaîs P, Haïssaguerre M, et al. Smartwatch-based detection of cardiac arrhythmias: beyond the differentiation between sinus rhythm and atrial fibrillation. Heart Rhythm 2021;18(9):1524-1532. [doi: 10.1016/j.hrthm.2021.06.1176] [Medline: 34147700]
18. Pipek LZ, Nascimento RFV, Acencio MMP, Teixeira LR. Comparison of SpO2 and heart rate values on apple watch and conventional commercial oximeters devices in patients with lung disease. Sci Rep 2021;11(1):18901 [FREE Full text] [doi: 10.1038/s41598-021-98453-3] [Medline: 34556765]
19. Wu R, Calligan M, Son T, Rakhra H, de Lara E, Mariakakis A, et al. Impressions and perceptions of a smartphone and smartwatch self-management tool for patients with COPD: a qualitative study. COPD 2024;21(1):2277158 [FREE Full text] [doi: 10.1080/15412555.2023.2277158] [Medline: 38348964]
20. Coutu F, Iorio OC, Ross BA. Remote patient monitoring strategies and wearable technology in chronic obstructive pulmonary disease. Front Med (Lausanne) 2023;10:1236598 [FREE Full text] [doi: 10.3389/fmed.2023.1236598] [Medline: 37663662]
21. Powers R, Etezadi-Amoli M, Arnold EM, Kianian S, Mance I, Gibiansky M, et al. Smartwatch inertial sensors continuously monitor real-world motor fluctuations in Parkinson's disease. Sci Transl Med 2021;13(579):eabd7865. [doi: 10.1126/scitranslmed.abd7865] [Medline: 33536284]
22. Adams JL, Kangarloo T, Tracey B, O'Donnell P, Volfson D, Latzman RD, Parkinson Study Group Watch-PD Study InvestigatorsCollaborators. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. NPJ Parkinsons Dis 2023;9(1):64 [FREE Full text] [doi: 10.1038/s41531-023-00497-x] [Medline: 37069193]
23. Fay-Karmon T, Galor N, Heimler B, Zilka A, Bartsch RP, Plotnik M, et al. Home-based monitoring of persons with advanced Parkinson's disease using smartwatch-smartphone technology. Sci Rep 2024;14(1):9 [FREE Full text] [doi: 10.1038/s41598-023-48209-y] [Medline: 38167434]
24. Lui GY, Loughnane D, Polley C, Jayarathna T, Breen PP. The apple watch for monitoring mental health-related physiological symptoms: literature review. JMIR Ment Health 2022;9(9):e37354 [FREE Full text] [doi: 10.2196/37354] [Medline: 36069848]
25. Lui GY, Loughnane D, Polley C, Jayarathna T, Breen PP. The apple watch for monitoring mental health-related physiological symptoms: literature review. JMIR Ment Health 2022;9(9):e37354 [FREE Full text] [doi: 10.2196/37354] [Medline: 36069848]
26. Sultana M, Al-Jefri M, Lee J. Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study. JMIR Mhealth Uhealth 2020;8(9):e17818 [FREE Full text] [doi: 10.2196/17818] [Medline: 32990638]
27. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using smartphones and wearable devices to monitor behavioral changes during COVID-19. J Med Internet Res 2020;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
28. Friel CP, Garber CE. Who uses wearable activity trackers and why? A comparison of former and current users in the United States. Am J Health Promot 2020;34(7):762-769. [doi: 10.1177/0890117120919366] [Medline: 32329355]
29. Venn RA, Khurshid S, Grayson M, Ashburner JM, Al-Alusi MA, Chang Y, et al. Characteristics and attitudes of wearable device users and nonusers in a large health care system. J Am Heart Assoc 2024;13(1):e032126 [FREE Full text] [doi: 10.1161/JAHA.123.032126] [Medline: 38156452]
30. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by us adults: insights from a national survey. J Med Internet Res 2020;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
31. Huarng KH, Yu TH, Lee CF. Adoption model of healthcare wearable devices. Technol. Forecast. Soc. Change 2022;174:121286. [doi: 10.1016/j.techfore.2021.121286]
32. Ferguson C, Hickman LD, Turkmani S, Breen P, Gargiulo G, Inglis SC. "Wearables only work on patients that wear them": barriers and facilitators to the adoption of wearable cardiac monitoring technologies. Cardiovasc Digit Health J 2021;2(2):137-147 [FREE Full text] [doi: 10.1016/j.cvdhj.2021.02.001] [Medline: 35265900]

XSL•FO
RenderX

33. Holko M, Litwin TR, Munoz F, Theisz KI, Salgin L, Jenks NP, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. NPJ Digit Med 2022;5(1):53 [FREE Full text] [doi: 10.1038/s41746-022-00593-x] [Medline: 35469045]
34. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
35. Chandrasekaran R, Katthula V, Moustakas E. Use of wearable healthcare devices by older adults and their willingness to share health data with providers. Health Informatics J 2021;27(4):14604582211058073 [FREE Full text] [doi: 10.1177/14604582211058073] [Medline: 34802315]
36. Kim EH, Jenness JL, Miller AB, Halabi R, de Zambotti M, Bagot KS, et al. Association of demographic and socioeconomic indicators with the use of wearable devices among children. JAMA Netw Open 2023;6(3):e235681 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.5681] [Medline: 36995714]
37. Bao H, Lee EWJ. Examining the antecedents and health outcomes of health apps and wearables use: an integration of the technology acceptance model and communication inequality. Behav. Inf. Technol 2023;43(4):695-716. [doi: 10.1080/0144929x.2023.2183062]
38. Verma B, Tandon U. Modelling barriers to wearable technologies in Indian context: validating the moderating role of technology literacy. Glob. Knowl. Mem. Commun 2022;73(6/7):984-1004 [FREE Full text] [doi: 10.1108/GKMC-08-2022-0209]
39. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]
40. Chandrasekaran R, Sharma P, Moustakas E. Exploring disparities in healthcare wearable use among cardiovascular patients: findings from a national survey. Rev Cardiovasc Med 2023;24(11):307 [FREE Full text] [doi: 10.31083/j.rcm2411307] [Medline: 39076432]
41. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022;24(1):e30791. [doi: 10.2196/30791] [Medline: 35060915]
42. Chow H, Yang C. Accuracy of optical heart rate sensing technology in wearable fitness trackers for young and older adults: validation and comparison study. JMIR Mhealth Uhealth 2020;8(4):e14707 [FREE Full text] [doi: 10.2196/14707] [Medline: 32343255]
43. Adebesin F, Mwalugha R. The mediating role of organizational reputation and trust in the intention to use wearable health devices: cross-country study. JMIR Mhealth Uhealth 2020;8(6):e16721 [FREE Full text] [doi: 10.2196/16721] [Medline: 32348260]
44. Chang C. Exploring the usage intentions of wearable medical devices: a demonstration study. Interact J Med Res 2020;9(3):e19776 [FREE Full text] [doi: 10.2196/19776] [Medline: 32945778]
45. Wang H, Tao D, Yu N, Qu X. Understanding consumer acceptance of healthcare wearable devices: an integrated model of UTAUT and TTF. Int J Med Inform 2020;139:104156. [doi: 10.1016/j.ijmedinf.2020.104156] [Medline: 32387819]
46. Singh N, Misra R, Singh S, Rana NP, Khorana S. Assessing the factors that influence the adoption of healthcare wearables by the older population using an extended PMT model. Technol Soc 2022;71:102126. [doi: 10.1016/j.techsoc.2022.102126]
47. Rising CJ, Gaysynsky A, Blake KD, Jensen RE, Oh A. Willingness to share data from wearable health and activity trackers: analysis of the 2019 Health information national trends survey data. JMIR Mhealth Uhealth 2021;9(12):e29190 [FREE Full text] [doi: 10.2196/29190] [Medline: 34898448]
48. Hyde ET, Omura JD, Fulton JE, Weldy A, Carlson SA. Physical activity surveillance using wearable activity monitors: are US adults willing to share their data? Am J Health Promot 2020;34(6):672-676. [doi: 10.1177/0890117119900587] [Medline: 31957481]
49. Shandhi MMH, Singh K, Janson N, Ashar P, Singh G, Lu B, et al. Assessment of ownership of smart devices and the acceptability of digital health data sharing. NPJ Digit Med 2024;7(1):44 [FREE Full text] [doi: 10.1038/s41746-024-01030-x] [Medline: 38388660]
50. Taylor CO, Flaks-Manov N, Ramesh S, Choe EK. Willingness to share wearable device data for research among mechanical turk workers: web-based survey study. J Med Internet Res 2021;23(10):e19789 [FREE Full text] [doi: 10.2196/19789] [Medline: 34673528]
51. Hirst Y, Stoffel ST, Brewer HR, Timotijevic L, Raats MM, Flanagan JM. Understanding public attitudes and willingness to share commercial data for health research: survey study in the United Kingdom. JMIR Public Health Surveill 2023;9:e40814 [FREE Full text] [doi: 10.2196/40814] [Medline: 36951929]
52. Gramlich J. Two years into the pandemic, Americans inch closer to a new normal. Pew Research Center. URL: https://www.pewresearch.org/2022/03/03/two-years-into-the-pandemic-americans-inch-closer-to-a-new-normal/ [accessed 2025-02-04]
53. Lee SY, Lee K. Factors that influence an individual's intention to adopt a wearable healthcare device: The case of a wearable fitness tracker. Technol Forecast Soc Change 2018;129:154-163 [FREE Full text] [doi: 10.1016/j.techfore.2018.01.002]
54. Abouzahra M, Ghasemaghaei M. The antecedents and results of seniors' use of activity tracking wearable devices. Health Policy Technol 2020;9(2):213-217 [FREE Full text] [doi: 10.1016/j.hlpt.2019.11.002]

XSL•FO
RenderX

55. Nuss K, Li K. Motivation for physical activity and physcial activity engagement in current and former wearable fitness tracker users: a mixed-methods examination. Comput. Hum. Behav 2021;121:106798. [doi: 10.1016/j.chb.2021.106798]
56. Deranek K, Hewitt B, Gudi A, McLeod A. The impact of exercise motives on adolescents' sustained use of wearable technology. Behav. Inf. Technol 2020;40(7):691-705. [doi: 10.1080/0144929x.2020.1720295]
57. Finney Rutten LJ, Blake KD, Skolnick VG, Davis T, Moser RP, Hesse BW. Data resource profile: The National Cancer Institute's Health Information National Trends Survey (HINTS). Int J Epidemiol 2020;49(1):17-17j [FREE Full text] [doi: 10.1093/ije/dyz083] [Medline: 31038687]
58. Methodology reports. National Cancer Institute at the National Institutes of Health. URL: https://hints.cancer.gov/data/methodology-reports.aspx [accessed 2025-02-04]
59. Survey instruments. National Cancer Institute at the National Institutes of Health. URL: https://hints.cancer.gov/data/survey-instruments.aspx [accessed 2025-02-04]
60. Xie Z, Jo A, Hong Y. Electronic wearable device and physical activity among US adults: an analysis of 2019 HINTS data. Int J Med Inform 2020;144:104297. [doi: 10.1016/j.ijmedinf.2020.104297] [Medline: 33059242]
61. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020. URL: https://www.pewresearch.org/short-reads/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2025-02-04]
62. Wunsch K, Kienberger K, Niessner C. Changes in physical activity patterns due to the Covid-19 pandemic: a systematic review and meta-analysis. Int J Environ Res Public Health 2022;19(4):2250 [FREE Full text] [doi: 10.3390/ijerph19042250] [Medline: 35206434]
63. Tong HL, Maher C, Parker K, Pham TD, Neves AL, Riordan B, et al. The use of mobile apps and fitness trackers to promote healthy behaviors during COVID-19: a cross-sectional survey. PLOS Digit Health 2022;1(8):e0000087 [FREE Full text] [doi: 10.1371/journal.pdig.0000087] [Medline: 36812578]
64. Seifert A, Vandelanotte C. The use of wearables and health apps and the willingness to share self-collected data among older adults. Aging Health Res 2021;1(3):100032. [doi: 10.1016/j.ahr.2021.100032]
65. Lupton D, Maslen S. How women use digital technologies for health: qualitative interview and focus group study. J Med Internet Res 2019;21(1):e11481 [FREE Full text] [doi: 10.2196/11481] [Medline: 30681963]
66. Del Busso L, Brottveit G, Torp Løkkeberg S, Gluppe G. Women's embodied experiences of using wearable digital self-tracking health technology: a review of the qualitative research literature. Health Care Women Int 2022;43(12):1355-1379 [FREE Full text] [doi: 10.1080/07399332.2021.1884682] [Medline: 33900152]
67. Ajayi KV, Wachira E, Onyeaka HK, Montour T, Olowolaju S, Garney W. The use of digital health tools for health promotion among women with and without chronic diseases: insights from the 2017-2020 health information national trends survey. JMIR Mhealth Uhealth 2022;10(8):e39520 [FREE Full text] [doi: 10.2196/39520] [Medline: 35984680]
68. Lin X, Wang X. Examining gender differences in people's information-sharing decisions on social networking sites. Int. J. Inf. Manag 2020;50:45-56. [doi: 10.1016/j.ijinfomgt.2019.05.004]
69. Tech companies have a trust gap to overcome-especially with women. Deloitte Insights. 2023. URL: https://www2.deloitte.com/us/en/insights/industry/technology/bridge-data-privacy-concerns-in-women-with-technology.html [accessed 2025-02-04]
70. Jeng MY, Pai FY, Yeh TM. Antecedents for older adults' intention to use smart health wearable devices-technology anxiety as a moderator. Behav Sci (Basel) 2022;12(4):114 [FREE Full text] [doi: 10.3390/bs12040114] [Medline: 35447686]
71. Farivar S, Abouzahra M, Ghasemaghaei M. Wearable device adoption among older adults: a mixed-methods study. Int J Inf Manage 2020;55:102209 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102209] [Medline: 32834339]
72. Cruz S, Lu C, Ulloa M, Redding A, Hester J, Jacobs M. Perceptions of wearable health tools post the COVID-19 emergency in low-income latin communities: qualitative study. JMIR Mhealth Uhealth 2024;12:e50826 [FREE Full text] [doi: 10.2196/50826] [Medline: 38717816]
73. Rajamani G, Kurina L, Rosas LG. Investigating health information technology usage by sociodemographic subpopulations to increase community engagement in healthcare: an analysis of the health information national trends survey. AMIA Annu Symp Proc 2021;2021:1029-1038 [FREE Full text] [Medline: 35308912]
74. Kang H, Jung EH. The smart wearables-privacy paradox: acluster analysis of smartwatch users. Behav. Inf. Technol 2020;40(16):1755-1768 [FREE Full text] [doi: 10.1080/0144929X.2020.1778787]
75. Singh S. The moderating role of privacy concerns on intention to use smart wearable technologies: an integrated model combining UTAUT2 theoretical framework and privacy dimensions. J. Glob. Mark 2022;36(2):93-111 [FREE Full text] [doi: 10.1080/08911762.2022.2141167]
76. Dhingra LS, Aminorroaya A, Oikonomou EK, Nargesi AA, Wilson FP, Krumholz HM, et al. Use of wearable devices in individuals with or at risk for cardiovascular disease in the US, 2019 to 2020. JAMA Netw Open 2023;6(6):e2316634 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.16634] [Medline: 37285157]
77. Walle AD, Jemere AT, Tilahun B, Endehabtu BF, Wubante SM, Melaku MS, et al. Intention to use wearable health devices and its predictors among diabetes mellitus patients in Amhara region referral hospitals, Ethiopia: using modified UTAUT-2 model. Inform. Med. Unlocked 2023;36:101157. [doi: 10.1016/j.imu.2022.101157]

XSL•FO
RenderX

78. Ahmed A, Aziz S, Abd-Alrazaq A, Farooq F, Sheikh J. Overview of artificial intelligence-driven wearable devices for diabetes: scoping review. J Med Internet Res 2022;24(8):e36010 [FREE Full text] [doi: 10.2196/36010] [Medline: 35943772]
79. Tsai J, Hung S, Lin G. Continued usage of smart wearable devices (SWDs): cross-level analysis of gamification and network externality. Electron Mark 2022;32(3):1661-1676 [FREE Full text] [doi: 10.1007/s12525-022-00575-7] [Medline: 35965737]
80. Windasari NA, Lin F. Why do people continue using fitness wearables? The effect of interactivity and gamification. Sage Open 2021;11(4):215824402110566. [doi: 10.1177/21582440211056606]
81. Hydari MZ, Adjerid I, Striegel AD. Health wearables, gamification, and healthful activity. Manage Sci 2023;69(7):3920-3938 [FREE Full text] [doi: 10.1287/mnsc.2022.4581] [Medline: 37547027]
82. HINTS data. National Cancer Institute at the National Institutes of Health. URL: https://hints.cancer.gov/data/Default.aspx [accessed 2025-01-31]

Abbreviations

HINTS: Health Information National Trends Survey
IRB: institutional review board
NIH: National Institutes of Health
OR: odds ratio

Edited by T de Azevedo Cardoso; submitted 02.07.24; peer-reviewed by K Ajayi, Y Chen, D Valvi, Q Shi; comments to author 27.09.24; revised version received 28.10.24; accepted 13.01.25; published 21.02.25.

Please cite as:
Chandrasekaran R, Sadiq T M, Moustakas E
Usage Trends and Data Sharing Practices of Healthcare Wearable Devices Among US Adults: Cross-Sectional Study
J Med Internet Res 2025;27:e63879
URL: https://www.jmir.org/2025/1/e63879
doi:10.2196/63879
PMID:39982763

©Ranganathan Chandrasekaran, Muhammed Sadiq T, Evangelos Moustakas. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Correlation Between the Online Visiting Time and Frequency Increase in Telemedicine Services Offered by Health Care Providers Before, During, and After the COVID-19 Pandemic in China: Cross-Sectional Study

Weiyi Wang[1*], PhD; Yuntian Chu[1,2*], PhD; Fangfang Cui[1,2], MM; Xiaobing Shi[1], MSc; Xu Zhang[1], MSc; Dongxu Sun[1], MSc; Jinming Shi[1], MSc; Jie Zhao[1,2,3], PhD

[1]National Engineering Laboratory for Internet Medical Systems and Applications, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[2]Institute of Intelligent Medicine, Henan Academy of Innovations in Medical Science, Zhengzhou, China

[3]Shanghai Artificial Intelligence Laboratory, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Jie Zhao, PhD
National Engineering Laboratory for Internet Medical Systems and Applications
The First Affiliated Hospital of Zhengzhou University
3 Jianshe East Rd
Erqi District
Zhengzhou, 450001
China
Phone: 86 15093316769
Email: zhaojie@zzu.edu.cn

Abstract

Background: China has changed its COVID-19 prevention and control status since 2023. However, what role telemedicine will play post–COVID-19 is still uncertain.

Objective: We aimed to determine the frequency change in health care providers offering telemedicine services before, during, and after COVID-19, as well as the correlation between the frequency change and telemedicine visit time.

Methods: The Telemedicine Informationization Professional Committee of China (TIPC) carried out a nationwide questionnaire survey. We adopted data from part of the questionnaires that answered questions regarding the frequency of offering telemedicine services before, during, and after the COVID-19 explosion. Chi-square tests were applied to compare general differences in the between-group telemedicine frequency. Regression models were performed to analyze correlations between the frequency change and the time spent in online versus in-person visits.

Results: Questionnaires from 428 providers were included. As reported, 39 (9.11%) providers often and 159 (37.15%) always offered telemedicine services before COVID-19 exploded. The component ratio increased to 12.38% (n=53) of providers often and 45.79% (n=196) always offering telemedicine during COVID-19 explosion and 12.62% (n=54) often and 50% (n=214) always offering telemedicine after pandemic control was relaxed. The increase in frequency shown as a difference between the before and during groups (χ2=17.21, *P*.002) and between the before and after groups (χ2=30.17, *P*<.001) was significant, while it was insignificant between the during and after groups (χ2=2.89, *P*.57). Senior professional titles (odds ratio [OR] 4.38, 95% CI 1.72-11.6) and longer (OR 3.87, 95% CI 1.95-7.89) and shorter (OR 2.04, 95% CI 1.11-3.87) online visits were correlated with the increase in frequency during versus before COVID-19. In addition, senior professional titles (OR 3.47, 95% CI 1.46-8.49), longer (OR 3.14, 95% CI 1.64-6.11) and shorter (OR=2.27, 95% CI 1.31-4.07) online visits, and using third-party telemedicine platforms (OR 0.51, 95% CI 0.29-0.86) were correlated with the increase in frequency after versus before COVID-19. No factor was significantly correlated with the frequency change after versus during COVID-19. In stratified analysis, longer online visits were correlated with both during versus before (OR 3.84, 95% CI 1.73-8.83) and after versus before (OR 3.40, 95% CI 1.61-7.34) groups for providers using hospital-run platforms, while shorter online visits were correlated with both during versus before (OR 8.16, 95% CI 1.39-68.3) and after versus before (OR 5.70, 95% CI 1.22-33.6) groups for providers using third-party platforms.

XSL•FO
RenderX

Conclusions: The frequency of telemedicine has increased since the COVID-19 pandemic exploded and is correlated with the time spent in online versus in-person visits. The correlation is different for providers using hospital-run and third party platforms. On a hospital-run platform, providers with longer online visits have a higher frequency of offering telemedicine, while on a third-party platform, providers with shorter online visits are more likely to offer telemedicine.

(J Med Internet Res 2025;27:e65092) doi:10.2196/65092

KEYWORDS

telemedicine; post–COVID-19; provider's perspective; length of online visit; COVID-19; pandemic; China; prevention; questionnaire; survey; healthcare provider

Introduction

The COVID-19 explosion has made significant changes to health care systems worldwide [1,2]. Telemedicine services, also referred to as virtual care in some studies, are defined in this paper as services that use digital technologies to provide health care at a distance and to connect doctors with each other. These services, mainly including online medical visits, virtual consultations, and remote monitoring and diagnosis, developed rapidly during the COVID-19 pandemic period [3-6]. As proved by former studies, telemedicine or virtual care has many advantages, such as decreased exposure risk to COVID-19 and other infectious diseases, ensuring the sustainability of medical services during quarantine; increased accessibility of medical services; lower costs for patients on both the money and time spent on the way to clinic; lower costs for medical institutes on patient flow management; relatively increased patient compliance, especially in chronic disease management; and early detection of subclinical changes [7-13]. Based on these advantages, telemedicine has been accepted, and both patients and providers are satisfied; it was even promoted by some policies as the pandemic proceeded [1,14-16].

Telemedicine in China developed early since the first remote consultation case in the mid-1980s, and the "Opinions on Promoting Telemedicine Services in Medical Institutions" policy issued in 2014 by the Chinese government accelerated its development [17.] In 2017, telemedicine platforms had been established in 22 of 34 provincial administrative regions in China that covered 13,000 medical institutions, which indicated that a telemedicine service system had been formed already [17,18]. When the COVID-19 pandemic exploded, this telemedicine system was rapidly put into practice. Virtual consultations were performed, live-stream platforms were used to offer health education to patients and the general public, virtual care bills were covered by the national health insurance, and medication-delivering systems were formed as the pandemic proceeded [2,19]. At some places, experts were gathered to offer preliminary screening for community residents with symptoms [20]. In Henan Province, the local telemedicine platform connected 147 COVID-19–designated hospitals, which offered prevention, diagnosis, and treatment of COVID-19 using modules such as real-time monitors, intractable case transfer, and remote guidance [21]. Implementation of telemedicine systems was on an unprecedented scale and pace.

In January 2023, the management of COVID-19 in China was changed from class A to class B, which means no longer isolating patients with COVID-19, no longer identifying close contacts, no longer designating high- and low-risk areas, and adjusting the testing strategy to voluntary testing. However, the pressure on the health care system rapidly increased in a short period after prevention measures were relaxed, which has led to different requirements for the application of telemedicine. Post–COVID-19, it is still uncertain how telemedicine would integrate in the health care system for a long time [22].

As a result, in this study, we aimed to compare the frequency of offering telemedicine services before the COVID-19, during the pandemic explosion, and after China reopened post–COVID-19 and to find out what kind of providers would like to offer more telemedicine services, even after pandemic prevention was relaxed.

Methods

Study Setting and Data Collection

The data for this study were taken from the 2023 Chinese National Survey of Telemedicine Development in Hospitals, conducted from September to October 2023, which used online questionnaires. The Telemedicine Informationization Professional Committee of China (TIPC) and the National Telemedicine Center of China (NTCC) were in charge of data collection for the survey.

A multistage sampling method was used to recruit respondents. Initially, 11 provinces were strategically selected from across eastern, central, and western regions for the survey, considering prior collaborative experiences. Subsequently, a TIPC member was designated to manage the survey operations in each province. They were tasked with overseeing hospital coordination and selection. Following receipt of consent from the chosen hospitals, a dedicated staff member was appointed in each facility by the TIPC members to handle the recruitment of participants, all of whom had undergone training by 3 experienced researchers from the NTCC. Ultimately, electronic questionnaires, accompanied by quick response (QR) codes and detailed instructions, were distributed to health care providers engaged in telemedicine services by the staff member in each hospital. Participation was voluntary, allowing participants to withdraw anytime. Data anonymity and confidentiality were guaranteed, and no monetary compensation was offered. Researchers from the NTCC periodically reviewed the data collected during the survey, offered prompt feedback, and carried out centralized data cleansing to ensure data integrity and effectiveness.

At the end of the survey, 996 questionnaires were received from all TIPC members. After excluding questionnaires with repeated

XSL•FO
RenderX

information, incomplete responses, inconsistency, and other problems (eg, some providers reported having worked in telemedicine longer than their overall working years), there were 831 (83.43%) valid questionnaires left. Data used in this study were from questionnaires filled by health care providers who had ever practiced telemedicine and has answered the question "How frequently did you offer telemedicine services before/during/after the COVID-19 pandemic explosion?" As a result, only 428 (51.5%) questionnaires were included.

The questionnaires included the following demographic items:

- Age, sex, professional title, education, and working experience, especially in offering telemedicine services recorded by year
- Hospital information, such as province, level of the hospital, and which department the provider was from
- The provider's feelings about validity and reliability (scoring from 1 to 10)
- Whether the provider engaged in telemedicine through a hospital self-operated platform or a platform run by a third party, which mainly refers to companies such as Chunyu Doctor
- The provider's feelings about whether the time spent in offering telemedicine services is longer or shorter or not different compared to in-person visits and other offline medical services
- The provider's feelings about how easy it is to acquire patients' tests and examine results during the telemedicine service procedure
- The provider's feelings about the frequency of offering telemedicine services before the COVID-19 pandemic outbreak, during the pandemic explosion and pandemic prevention, and after pandemic prevention control was relaxed

Professional titles included junior, intermediate, and senior. Education recorded doctor, master's, and bachelor's degrees, as well as education below the bachelor level. The province was then classified as the eastern, central, or western region of China. Hospital levels were categorized as tertiary or nontertiary. Departments of providers were sorted as clinic and nonclinic departments. How easy it was to acquire test results online and the frequency of telemedicine at 3 time points were scored on a 5-point Likert scale as "always," "often," "occasionally," "rarely," and "never."

Ethical Considerations

This study was conducted using an online survey method, collecting data that were limited to demographic information, such as age and gender, with no collection of sensitive or individually identifiable biological information. All procedures were approved by the Medical Ethics Committee (2020-KY-0379-002) of the First Affiliated Hospital of Zhengzhou University. Informed consent was obtained from all individual participants involved in the study. For secondary analyses of research data, we confirmed that the original informed consent or Institutional Review Board (IRB) approval included provisions allowing for secondary analysis without additional consent from the participants.

In this study, we prioritized the privacy and confidentiality of participants. The online survey was designed to collect only nonsensitive demographic information, such as age and gender, without any personal identifiers that could compromise the privacy of the respondents. All responses were collected anonymously, ensuring that no individual could be identified using the data.

Participants were not offered any form of compensation for their participation. No images of individual participants or users were included in the manuscript or multimedia appendices that would allow for identification.

Outcomes

The general frequency of offering telemedicine services was calculated at each time point, and the difference in the frequency between the 3 groups was considered the primary outcome of this study. As the frequency was an ordered categorical variable, it was reassigned a score of 1-5 at each time point. A change in the frequency of offering telemedicine services as the pandemic progressed was calculated, via subtraction, as the difference in the frequencies between the 3 time points for each participant, and the change was categorized as an increase or a decrease, depending on whether the result of the subtraction was >0 or <0. Thus, binary variables showing whether the frequency increased from before the pandemic outbreak to during its explosion, increased from during the explosion to after controls were relaxed, and from before the outbreak to after controls were relaxed were calculated as secondary outcomes of this study.

Statistical Analysis

Age was presented as the mean (SD). Other continuous variables were presented as the median (IQR) as their distribution was not symmetric. Categorical variables were described as counts and percentages. Chi-square tests were performed to compare differences in the frequency of offering telemedicine services between all groups. Single-variate and multiple regression analyses were used to explore correlations between the increase in frequency and other questionnaire responses. Stratified analysis was also performed in using the multiple regression model in different groups. A 2-sided *P* value of <05 was considered statistically significant in all analyses. Statistical analysis was performed using R version 4.3.1 (R Foundation for Statistical Computing). Figures demonstrating intergroup distributions and advantages and disadvantages were created in Microsoft Excel.

Results

Participant Details

A total of 428 health care providers (n=199, 46.5%, males and n=229, 53.5%, females) filled in the questionnaires and were included in this study. The general characteristics of participants and their responses to the questionnaires are summarized in Table 1. The participants' mean age was 38.6 (SD 8.51) years. Over half were bachelors (n=257, 60.05%), with a median of 12 (IQR 7-20) years of working experience and 4 (IQR 2-6) years working in telemedicine. The majority were from tertiary hospitals (n=229, 53.5%) and located in central (n=218, 50.93%)

XSL•FO
RenderX

and western (n=151, 35.28%) parts of China. Most providers were engaged in telemedicine through platforms operated by their own institutes (n=317, 74.07%), while the rest were engaged through third-party platforms usually operated by internet companies, such as Chunyu Doctor. Participants reported that test and examination results could be generally acquired (n=125, 29.21%, always; n=191, 44.63%, often) during the telemedicine procedure. Most participants believed that the reliability and validity of telemedicine are high, with a median score of 9 (IQR 8-10) and 9 (IQR 7.75-10), respectively, on a scale of 0-10. Regarding the complexity of the diagnosis procedure, 200 (46.73%) participants indicated online virtual visits are quicker than in-person outpatient visits, 99 (23.13%) participants reported the length of virtual medical service is even longer than an offline visit, and the rest reported that the time spent in the 2 methods is not different.

XSL•FO
RenderX

Table 1. General characteristics and responses of participants (N=428).

Characteristics	Value
Age (years), mean (SD)	38.6 (8.51)
Sex, n (%)	
Female	229 (53.50)
Professional title, n (%)	
Junior	123 (28.74)
Intermediate	146 (34.11)
Senior	159 (37.15)
Education, n (%)	
Doctor	23 (5.37)
Master’s degree	88 (20.56)
Bachelor’s degree	257 (60.05)
Below bachelor’s degree	60 (14.02)
Working years, median (IQR)	12 (7-20)
Telemedicine working years, median (IQR)	4 (2-6)
Department, n (%)	
Clinic	370 (86.45)
Nonclinic	58 (13.55)
Region, n (%)	
Eastern	59 (13.79)
Central	218 (50.93)
Western	151 (35.28)
Type of hospital, n (%)	
Tertiary	229 (53.50)
Nontertiary	199 (46.50)
Telemedicine platform, n (%)	
Self-operated	317 (74.07)
Third party	111 (25.93)
Time spent in visits, n (%)	
Shorter online visits	200 (46.73)
Not different	129 (30.14)
Longer online visits	99 (23.13)
Acquire test results online, n (%)	
Never	6 (1.40)
Rarely	28 (6.54)
Occasionally	78 (18.22)
Often	191 (44.63)
Always	125 (29.21)
Validity of telemedicine, median (IQR)	9 (8-10)
Reliability of telemedicine, median (IQR)	9 (7.75-10)

XSL•FO
RenderX

Outcomes

According to the providers' responses, the frequency of providing telemedicine services, including online visits, generally increased as the pandemic proceeded through the 3 stages: before, during, and after the COVID-19 explosion. As shown in Table 2, approximately 198 (46.26%) providers always or often offered telemedicine before the pandemic outbreak, 249 (58.17%) during the pandemic explosion, and 268 (62.62%) after the pandemic explosion. The differences between these groups, calculated using the chi-squared test, indicated that generally the frequency of providing telemedicine services increased since the pandemic exploded and stayed at similar levels after the explosion (Table 3).

Table 2. Frequency of offering telemedicine before, during, and after the COVID-19 pandemic.

Grade	Before the pandemic outbreak, n (%)	During the pandemic explosion, n (%)	After prevention measures were relaxed, n (%)
Never	29 (6.78)	11 (2.57)	9 (2.10)
Rarely	76 (17.76)	60 (14.02)	46 (10.75)
Occasionally	125 (29.21)	108 (25.23)	105 (24.53)
Often	159 (37.15)	196 (45.79)	214 (50.00)
Always	39 (9.11)	53 (12.38)	54 (12.62)

Table 3. Chi-square test results of differences between the frequencies of offering telemedicine.

Comparison between groups	χ^2 (*df*)	*P* value
Before vs during the pandemic	17.21 (4)	.002[a]
During vs after the pandemic	2.89 (4)	.57
Before vs after the pandemic	30.17 (4)	<.001[a]

[a]*P*<.05.

Regression models were performed in single and multiple variate analyses, and whether offering telemedicine services increased during the pandemic explosion versus before the pandemic outbreak, after the pandemic explosion versus during the explosion, and after the pandemic explosion versus before the pandemic outbreak was considered a dependent variable. As demonstrated in Table 4, age, working experience, telemedicine working experience, and providers' report of the validity of telemedicine were positively correlated or negatively correlated with an increase in the frequency of offering telemedicine services as the pandemic explosion proceeded, with significant correlation but a coeffective value close to 1. The providers' professional titles were significantly correlated with the frequency increase in the single-variate model: they had an increased frequency compared to those with junior titles. Participants from western regions compared to central regions and from nontertiary hospitals compared to tertiary hospitals were less likely to increase telemedicine services as the pandemic exploded. Compared to participants who felt the time spent between online and in-person visits was not different, those who felt online services take longer (during vs before: odds ratio [OR] 4.27, 95% CI 2.30-8.15; after vs before: OR 3.22, 95% CI 1.79-5.92) or shorter (during vs before: OR 2.09, 95% CI 1.19-3.80; after vs before: OR 2.15, 95% CI 1.28-3.72) increased telemedicine services.

XSL•FO
RenderX

Table 4. Single-variate analysis of telemedicine service frequency increase during versus before the COVID-19 pandemic and after versus before the pandemic.

Characteristics	During vs before the pandemic		After vs before the pandemic	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Age	1.04 (1.01-1.07)	.004[b]	1.01 (0.98-1.03)	.60
Sex				
Male	0.91 (0.59-1.40)	.70	0.82 (0.54-1.24)	.40
Professional title				
Intermediate	1.97 (1.05-3.79)	.04	1.17 (0.68-2.04)	.60
Senior	3.99 (2.22-7.47)	<.001[b]	1.98 (1.19-3.36)	.01[b]
Education				
Bachelor's degree	2.04 (0.99-4.63)	.07	1.38 (0.73-2.74)	.30
Master's degree	2.93 (1.31-7.09)	.01[b]	1.97 (0.96-4.22)	.07
Doctor	2.48 (0.78-7.76)	.12	2.11 (0.74-5.93)	.20
Working years	1.04 (1.01-1.06)	.003[b]	1.01 (0.99-1.03)	.50
Telemedicine working years	0.99 (0.94-1.04)	.70	0.95 (0.89-0.99)	.04
Department				
Nonclinic	1.28 (0.69-2.32)	.40	1.57 (0.88-2.77)	.12
Region				
Eastern	1.69 (0.92-3.05)	.09	1.18 (0.64-2.13)	.60
Western	0.54 (0.32-0.88)	.02[b]	0.64 (0.40-1.02)	.07
Type of hospital				
Nontertiary	0.45 (0.29-0.70)	<.001[b]	0.63 (0.41-0.95)	.03[b]
Telemedicine platform				
Third party	0.74 (0.44-1.22)	.30	0.61 (0.36-0.99)	.05
Time spent in visits				
Longer online visits	4.27 (2.30-8.15)	<.001[b]	3.22 (1.79-5.92)	<.001[b]
Shorter online visits	2.09 (1.19-3.80)	.01[b]	2.15 (1.28-3.72)	.005[b]
Acquire test result online				
Often	1.02 (0.61-1.70)	.90	0.93 (0.57-1.52)	.80
Occasionally	1.03 (0.54-1.94)	.90	1.28 (0.70-2.33)	.40
Rarely	1.12 (0.43-2.70)	.80	1.08 (0.43-2.56)	.90
Never	0.56 (0.03-3.63)	.60	1.14 (0.15-6.13)	.90
Validity of telemedicine	0.85 (0.75-0.97)	.02[b]	0.92 (0.81-1.05)	.20
Reliability of telemedicine	0.82 (0.72-0.94)	.70	0.89 (0.79-1.01)	.08

[a]OR: odds ratio.

[b]*P*<.05.

In the multivariate analysis, age, professional title, education, years working in telemedicine, region, type of hospital, platform provider engaged in telemedicine, time spent in online versus in-person service, and validity of telemedicine were included as independent variables in regression models. According to variance inflation factor test results shown in Table S1 in Multimedia Appendix 1, working years (during vs before: 10.85; after vs during: 9.99; after vs before: 10.86) and reliability of telemedicine (during vs before: 6.9; after vs during: 6.72; after vs before: 6.85) were not adjusted in the multivariate regression model. As shown in Table 5, providers with senior professional titles compared to junior titles, who were engaged in telemedicine through third party–operated platforms compared to hospital-run platforms, and who felt online visits are shorter

XSL•FO
RenderX

or longer than in-person visits had significant coeffective values. In Tables S2 and S3 in Multimedia Appendix 2, single and multiple regression models involving frequency increases after versus during the pandemic explosion are reported, and no significant correlations were found.

Table 5. Multivariate analysis of telemedicine service frequency increase during versus before the COVID-19 pandemic and after versus before the pandemic.

Characteristics	During vs before the pandemic		After vs before the pandemic	
	OR[a] (95% CI)[a]	*P* value	OR (95% CI)	*P* value
Age	0.99 (0.95-1.03)	.70	0.97 (0.93-1.01)	.20
Professional title				
Intermediate	1.79 (0.84-3.95)	.14	1.38 (0.70-2.75)	.40
Senior	4.38 (1.72-11.6)	.002[b]	3.47 (1.46-8.49)	.005[b]
Education				
Bachelor's degree	0.71 (0.29-1.85)	.50	0.74 (0.34-1.68)	.50
Master's degree	0.69 (0.24-2.09)	.50	0.84 (0.32-2.22)	.70
Doctor	0.39 (0.09-1.59)	.20	0.83 (0.23-2.96)	.80
Telemedicine working years	0.97 (0.92-1.03)	.40	0.93 (0.87-1.0)	.04[b]
Region				
Eastern	1.82 (0.94-3.51)	.07	1.19 (0.61-2.26)	.60
Western	0.6 (0.32-1.10)	.10	0.62 (0.35-1.08)	.10
Type of hospital				
Nontertiary	0.79 (0.44-1.41)	.40	1.07 (0.62-1.86)	.80
Telemedicine platform				
Third party	0.57 (0.32-1.00)	.05	0.51 (0.29-0.86)	.02[b]
Time spent in visits				
Longer online visits	3.87 (1.95-7.89)	<.001[b]	3.14 (1.64-6.11)	<.001[b]
Shorter online visits	2.04 (1.11-3.87)	.02[b]	2.27 (1.31-4.07)	.004[b]
Validity of telemedicine	0.96 (0.83-1.12)	.60	1 (0.86-1.16)	.90

[a]OR: odds ratio.

[b]*P*<.05.

Stratified analysis was performed. The distribution of the online service length between 2 types of platforms (Table S4 in Multimedia Appendix 3) showed that over half of the providers on self-operated platforms (160/317, 50.47%) felt online visits are shorter, while on third party–operated platforms, only 40 (36.04%) of 111 providers felt online visits are shorter. The difference in distribution between the 2 types of platforms was significant. Thus, the correlation of whether online service takes longer and the frequency increase in telemedicine might be different between the 2 types of platforms. As a result, providers were grouped by whether they chose to offer telemedicine on platforms operated by their own hospitals or on third party platforms.

As demonstrated in Table 6, multiple regression analysis was applied, and the variables age, professional title, education, years working in telemedicine, region of hospital, type of hospital, and validity of telemedicine reported by providers were adjusted as covariables in the regression model in each group. In the self-operated platform group, compared to providers who felt the time spent in online and in-person visits is not different, those who felt online visits are longer were significantly more likely to increase offering telemedicine services when COVID-19 exploded (OR 3.84, 95% CI 1.73-8.83) and when pandemic prevention controls were relaxed (OR 3.40, 95% CI 1.61-7.34) compared to before the pandemic. The differences in the telemedicine offer frequency increase for providers who felt online service time is shorter than in-person visits were not significant. In the third party–operated platform group, the increase in telemedicine offer frequency was significant among providers who felt online visits are shorter than in-person visits both when the pandemic exploded (OR 8.16, 95% CI 1.39-68.3) and when prevention measures were relaxed (OR 5.70, 95% CI 1.22-33.6) compared to before the pandemic, while the differences for providers who felt online visits are longer were not significant.

XSL•FO
RenderX

Table 6. Stratified analysis of the relationship between visit length and frequency increase[a].

Platform type, visit length, and frequency change duration	OR[b] (95% CI)	*P* value
Self-operated: longer online visits		
During vs before the pandemic	3.84 (1.73-8.83)	.001[c]
After vs before the pandemic	3.40 (1.61-7.34)	.001[c]
Self-operated: shorter online visits		
During vs before the pandemic	1.61 (0.81-3.29)	.20
After vs before the pandemic	1.78 (0.96-3.41)	.07
Third party: longer online visits		
During vs before the pandemic	6.51 (1.05-54.7)	.06
After vs before the pandemic	1.87 (0.32-11.6)	.50
Third party: : shorter online visits		
During vs before the pandemic	8.16 (1.39-68.3)	.03[c]
After vs before the pandemic	5.70 (1.22-33.6)	.04[c]

[a]Adjusted for age, professional title, education, years working in telemedicine, region, type of hospital, and validity of telemedicine.

[b]OR: odds ratio.

[c]*P*<.05.

Discussion

Principal Findings

Based on a national survey of telemedicine providers, this study explored the change in the frequency of offering telemedicine services at the time before COVID-19 exploded, during the pandemic explosion, and after prevention measures were relaxed; analyzed the correlation between other factors and the frequency change; and located health care providers who would like to offer more telemedicine services in China even after the pandemic is over.

In general, our study found that the frequency of health care providers offering telemedicine increased both during the pandemic and after prevention measures were relaxed compared to before the pandemic. Specifically, we found that providers' feelings about the online service length compared to in-person visits were correlated with the telemedicine offer frequency increase as the pandemic proceeded. Among providers offering telemedicine through platforms run by their own hospitals, those who felt online visits take longer were more likely to increase offering telemedicine. Among providers engaged through third-party platforms, those who felt online visits are shorter were more likely to increase offering telemedicine.

Comparison With Prior Work

Studies worldwide have already proved that telemedicine technologies were helpful tools during the COVID-19 explosion and were generally welcomed after the pandemic was over [1,3,9,23,24]. Considering the condition changed in China, it was significant that the telemedicine offer frequency was still high after pandemic controls were relaxed [22].

In our study, the difference in the telemedicine frequency change was not significant between different types of hospitals but was significant between the 2 types of platforms. As Chinese telemedicine-promoting policies started early, there were various ways of offering telemedicine services at the time the pandemic exploded [17]. In May 2019, there were 158 telemedicine platforms known as "internet hospitals" operated by local public hospitals around the nation and video-based platforms, such as the Huawei telepresence conference system [18]. Private companies, such as Chunyu Doctor (founded in 2011) and Ping An Health (founded in 2015), established and began operating their teleconsulting and diagnosis platforms even earlier. Some physicians said they would still contact patients through phone calls, text messages, or instant messaging software, such as WeChat, but a few responded that they would offer telemedicine services or only consultations through these ways. As a result, common tools, such as WhatsApp, telephone, or email, were not included as a way of offering telemedicine in this study. The difference in frequency between the 2 kinds of platforms was not the same as the platforms or tools used in former studies [25,26].

The 2 types of platforms in this study included providers engaging in telemedicine publicly or privately. As all our participants were from public hospitals, the self-operated platforms were similar to those run by public hospitals. However, the third-party platforms were mainly run by private companies. Former studies on cognition and telemedicine services between public and private providers are not consistent. Galle et al [27] and Cordioli et al [28] suggested that providers in the private sector or private clinics have a higher frequency of offering telemedicine services. Scheffer et al [29] claimed physicians working in both public and private sectors are the most frequent users. Elhadi et al [30] indicated there are no differences in attitude, cognition, knowledge, or skills in providers between work in public hospitals, private hospitals, or both. Mazouri-Karker et al [31] reported that differences

XSL•FO
RenderX

between public and private providers are only their preference in tools. In our results, public providers reported that a higher frequency increase after pandemic controls were relaxed might be because public hospitals, especially tertiary public hospitals in China, usually have a better reputation, whereas private third-party telemedicine platforms or physicians who work a "part-time job" on these platforms can hardly acquire patient trust.

We also found that some providers feel that online visits take longer, others believe that online visits are shorter, while still others feel that they take the same time as in-person visits. This is consistent with Ly's [32] results that some physicians report telemedicine increases their efficiency and lowers the workload, while other physicians report telemedicine takes more time and increases their workload because of having to learn new techniques and because communication on-screen is difficult. Although some researchers believe that one of the advantages of adopting telemedicine during COVID-19 was that it could save time and increase efficiency, it actually saved patients' traffic-related costs on both time and money and increased efficiency by reducing unnecessary visits through pretriage online [24,30,33]; however, we are still uncertain about whether adopting telemedicine could have saved the providers' time. Naamani et al [34] suggested that telemedicine could save providers' time because of pretriage online. Shaarani et al [26] proved that the average time spent in in-person visits reported by former studies is longer than 15 minutes, while providers in their survey reported a shorter average online visit time. Silver et al [35] found out that providers who spent an average visiting time longer than 14 minutes would prefer offering visits in person instead of virtually. Ramsey et al [36] indicated that providers would have to spend more time on documentation, which made the procedure more complicated and online visits took longer.

Providers in our study who reported that online visits are shorter than in-person visits have a good reason to increase the frequency of offering telemedicine services as these have higher efficiency. Providers who reported that online visits are longer than in-person visits, compared to those who believed there is no difference between the 2 service types, would also like to increase the telemedicine usage frequency. This might be because telemedicine has other advantages as well, such as saving patients' costs, even though it would increase the time spent by providers. As Ramsey et al [36] suggested, the other benefits for clinicians, patients, and quality of care could motivate health care professionals to offer more telemedicine services. In addition, in China, the average outpatient visiting time is much shorter, especially in tertiary hospitals [37]. This indicates that even though online visits take longer, they might still be shorter than average in-person visits in other countries. Providers who feel online visits take more time are not against offering telemedicine according to Silver et al [35], as also shown by our results of stratified analysis. Providers offering telemedicine services through their own hospitals' platforms might be more responsible and attach importance to their reputation. However, providers using third-party platforms might pay more attention to efficiency. As a result, on self-operated platforms, the increase in the telemedicine offer frequency by providers who felt online visits take longer was significant, while on third-party platforms, the increase in the telemedicine offer frequency by providers who felt online visits are shorter was significant.

Limitations

Our study showed that the length of online visits is correlated with the frequency change in offering telemedicine by providers, and the frequency increased for both longer and shorter online visits. However, our questionnaire was not designed to acquire the average number of minutes the providers spent in online and in-person visits. Further studies could focus on this to obtain a more detailed and convincing result. The results from this study were all from a cross-sectional survey, though it was nationwide. Thus, the frequencies of offering telemedicine at the 3 time points and the frequency change data were all obtained from providers' self-reports, which may have led to self-report and selection bias. It would be more accurate after regional online prescription–examining platforms in China are established and we can acquire data on them.

Moreover, some questions in this study were rather subjective. As the questionnaire used a 5-point Likert scale, questions regarding the frequency of offering telemedicine was set with 5 levels from "never" to "always," without specifying the exact time or duration for each level. "What platform do you mainly use to offer telemedicine services?" was set as a single-choice question, which was less suitable for more complicated situations, such as providers using both types of platforms.

Conclusion

The frequency of telemedicine services offered by health care providers has increased since the COVID-19 pandemic exploded and was still high after pandemic prevention controls were relaxed in China. Differences in providers' professional titles, types of platforms, and length of online visits compared to in-person visits are correlated with the increase in frequency. In using hospital-run platforms, providers with longer online visits are more likely to offer telemedicine services, while on third-party platforms, providers with shorter online visits would like to increase telemedicine services.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (grant 2022ZD0160704), the Science and Technology Research Program in Henan Province (grant 242102311021), and the Joint Co-construction Program of Medical Science and Technology of Henan Province (grants LHGJ20220433 and LHGJ20230178). The funders of the study played no role in the design, data collection, or interpretation of the work. The views expressed in the paper are those of the

XSL•FO
RenderX

authors and do not necessarily reflect the position of the funding bodies. No artificial intelligence was used in any part of the paper.

Data Availability

The datasets generated and analyzed during this study are not publicly available because they are questionnaire data obtained from a nationwide survey initiated by the Telemedicine Informationization Professional Committee of China, but they are available from the corresponding author upon reasonable request.

Authors' Contributions

JZ, WW, and YC made substantial contributions to the design of this work. WW and YC designed, led, and initiated the survey; conducted data analysis; and drafted the initial paper. FC and XS played an important role in data preprocessing. XZ and DS were involved in the discussion of the paper structure and results. JS reviewed and revised the paper. All authors contributed to writing and revising the manuscript and have approved the final version of the manuscript, as submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variance inflation factor test.
[DOCX File , 15 KB - jmir_v27i1e65092_app1.docx]

Multimedia Appendix 2
Single and multiple regression model for frequency increase between after and during groups.
[DOCX File , 21 KB - jmir_v27i1e65092_app2.docx]

Multimedia Appendix 3
Group difference in online visit lengths between platforms.
[DOCX File , 13 KB - jmir_v27i1e65092_app3.docx]

References

1. Bokolo Jr. A. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
2. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
3. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. Ir J Med Sci 2021 Feb;190(1):1-10 [FREE Full text] [doi: 10.1007/s11845-020-02299-z] [Medline: 32642981]
4. Silva RSD, Schmtiz CAA, Harzheim E, Molina-Bastos CG, Oliveira EBD, Roman R, et al. The role of telehealth in the covid-19 pandemic: a Brazilian experience. Cien Saude Colet 2021 Jun;26(6):2149-2157 [FREE Full text] [doi: 10.1590/1413-81232021266.39662020] [Medline: 34231727]
5. Combi C, Pozzani G, Pozzi G. Telemedicine for developing countries. a survey and some design issues. Appl Clin Inform 2016 Nov 02;7(4):1025-1050 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0089] [Medline: 27803948]
6. Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. Lancet Digit Health 2020 Jun 06;2(6):e282-e285 [FREE Full text] [doi: 10.1016/S2589-7500(20)30104-7] [Medline: 32382724]
7. Maroongroge S, De B, Woodhouse KD, Bassett RL, Lee P, Bloom ES, et al. Physician perspectives on telemedicine in radiation oncology. Adv Radiat Oncol 2023;8(2):101005 [FREE Full text] [doi: 10.1016/j.adro.2022.101005] [Medline: 36311822]
8. Liu J, Liu S, Zheng T, Bi Y. Physicians' perspectives of telemedicine during the COVID-19 pandemic in China: qualitative survey study. JMIR Med Inform 2021 Jun 01;9(6):e26463 [FREE Full text] [doi: 10.2196/26463] [Medline: 33945493]
9. Cerda IH, Therond A, Moreau S, Studer K, Donjow AR, Crowther JE, et al. Telehealth and virtual reality technologies in chronic pain management: a narrative review. Curr Pain Headache Rep 2024 Mar;28(3):83-94. [doi: 10.1007/s11916-023-01205-3] [Medline: 38175490]
10. Bellicini MG, D'Altilia FP, Gussago C, Adamo M, Lombardi CM, Tomasoni D, et al. Telemedicine for the treatment of heart failure: new opportunities after COVID-19. J Cardiovasc Med (Hagerstown) 2023 Oct 01;24(10):700-707. [doi: 10.2459/JCM.0000000000001514] [Medline: 37409660]
11. Sumner L, Schmidt H, Minden S, Falkenberg N, Sun L, McBurney R, et al. Use of telemedicine among people with multiple sclerosis before and during the COVID-19 pandemic. Telemed J E Health 2023 Aug;29(8):1152-1163. [doi: 10.1089/tmj.2022.0284] [Medline: 36493358]

XSL•FO
RenderX

12. Saleem MK, Sattar K, Ejaz KF, Rehman MU, Saleem H, Khursheed S, et al. Use of telemedicine to tackle health problems in South Asia during the COVID-19 era and beyond: a systematic review. Ann Med Surg (Lond) 2024 Feb;86(2):1012-1020 [FREE Full text] [doi: 10.1097/MS9.0000000000001649] [Medline: 38333256]
13. Mahajan V, Singh T, Azad C. Using telemedicine during the COVID-19 pandemic. Indian Pediatr 2020 Jul 15;57(7):652-657 [FREE Full text] [Medline: 32412914]
14. Aldekhyyel RN, Alshuaibi F, Alsaaid O, Bin Moammar F, Alanazy T, Namshah A, et al. Exploring behavioral intention to use telemedicine services post COVID-19: a cross sectional study in Saudi Arabia. Front Public Health 2024;12:1385713 [FREE Full text] [doi: 10.3389/fpubh.2024.1385713] [Medline: 38689764]
15. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
16. Kim J, Cai ZR, Chen ML, Onyeka S, Ko JM, Linos E. Telehealth utilization and associations in the United States during the third year of the COVID-19 pandemic: population-based survey study in 2022. JMIR Public Health Surveill 2024 Apr 26;10:e51279 [FREE Full text] [doi: 10.2196/51279] [Medline: 38669075]
17. Zhang W, He D, Wang G, Zhu C, Evans R. Analyzing national telemedicine policies in China from the perspective of policy instrument (1997-2020). Int J Med Inform 2022 Oct;166:104854. [doi: 10.1016/j.ijmedinf.2022.104854] [Medline: 35981479]
18. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e18426 [FREE Full text] [doi: 10.2196/18426] [Medline: 33095175]
19. Chen R, Jiang Q. Evolution of telemedicine in China during COVID-19 pandemic: from 2020 to 2022. J Public Health Policy 2022 Sep;43(3):469-472 [FREE Full text] [doi: 10.1057/s41271-022-00353-x] [Medline: 35676328]
20. Song X, Liu X, Wang C. The role of telemedicine during the COVID-19 epidemic in China-experience from Shandong Province. Crit Care 2020 Apr 28;24(1):178 [FREE Full text] [doi: 10.1186/s13054-020-02884-9] [Medline: 32345359]
21. Gao J, Ma Q, Sun D, Yang Y, Ren M, Wang L, et al. Telemedicine in the battle with 2019 novel coronavirus disease (COVID-19) in Henan Province, China: a narrative study. Telemed J E Health 2023 Aug;29(8):1211-1223. [doi: 10.1089/tmj.2022.0247] [Medline: 36602780]
22. Ge J. The COVID-19 pandemic in China: from dynamic zero-COVID to current policy. Herz 2023 Jun;48(3):226-228 [FREE Full text] [doi: 10.1007/s00059-023-05183-5] [Medline: 37294456]
23. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: sustaining telehealth beyond COVID-19. J Telemed Telecare 2022 May 26;28(4):301-308. [doi: 10.1177/1357633X20960638] [Medline: 32985380]
24. Gareev I, Gallyametdinov A, Beylerli O, Valitov E, Alyshov A, Pavlov V, et al. The opportunities and challenges of telemedicine during COVID-19 pandemic. Front Biosci (Elite Ed) 2021 Dec 20;13(2):291-298 [FREE Full text] [doi: 10.52586/E885] [Medline: 34937315]
25. Nies S, Patel S, Shafer M, Longman L, Sharif I, Pina P. Understanding physicians' preferences for telemedicine during the COVID-19 pandemic: cross-sectional study. JMIR Form Res 2021 Aug 13;5(8):e26565 [FREE Full text] [doi: 10.2196/26565] [Medline: 34227993]
26. Shaarani I, Ghanem A, Jounblat M, Jounblat H, Mansour R, Taleb R. Utilization of telemedicine by the Lebanese physicians during time of pandemic. Telemed J E Health 2022 Feb;28(2):219-226. [doi: 10.1089/tmj.2021.0090] [Medline: 33999731]
27. Galle A, Semaan A, Huysmans E, Audet C, Asefa A, Delvaux T, et al. A double-edged sword-telemedicine for maternal care during COVID-19: findings from a global mixed-methods study of healthcare providers. BMJ Glob Health 2021 Feb;6(2):e004575 [FREE Full text] [doi: 10.1136/bmjgh-2020-004575] [Medline: 33632772]
28. Cordioli E, Giavina-Bianchi M, Pedrotti CHS, Podgaec S. Brazilian Medical Survey on Telemedicine since the onset of COVID-19. Einstein (Sao Paulo) 2023;21:eAE0428 [FREE Full text] [doi: 10.31744/einstein_journal/2023AE0428] [Medline: 37729352]
29. Scheffer M, Cassenote A, de Britto E Alves MTSS, Russo G. The multiple uses of telemedicine during the pandemic: the evidence from a cross-sectional survey of medical doctors in Brazil. Global Health 2022 Sep 19;18(1):81 [FREE Full text] [doi: 10.1186/s12992-022-00875-9] [Medline: 36123696]
30. Elhadi M, Elhadi A, Bouhuwaish A, Bin Alshiteewi F, Elmabrouk A, Alsuyihili A, et al. Telemedicine awareness, knowledge, attitude, and skills of health care workers in a low-resource country during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2021 Feb 25;23(2):e20812 [FREE Full text] [doi: 10.2196/20812] [Medline: 33600350]
31. Mazouri-Karker S, Lüchinger R, Braillard O, Bajwa N, Achab S, Hudelson P, et al. Perceptions of and preferences for telemedicine use since the early stages of the COVID-19 pandemic: cross-sectional survey of patients and physicians. JMIR Hum Factors 2023 Nov 07;10:e50740 [FREE Full text] [doi: 10.2196/50740] [Medline: 37934574]
32. Ly BA, Labonté R, Bourgeault IL, Niang MN. The individual and contextual determinants of the use of telemedicine: a descriptive study of the perceptions of Senegal's physicians and telemedicine projects managers. PLoS One 2017;12(7):e0181070 [FREE Full text] [doi: 10.1371/journal.pone.0181070] [Medline: 28732028]

XSL•FO
RenderX

33. Grossman Z, Chodick G, Reingold SM, Chapnick G, Ashkenazi S. The future of telemedicine visits after COVID-19: perceptions of primary care pediatricians. Isr J Health Policy Res 2020 Oct 20;9(1):53 [FREE Full text] [doi: 10.1186/s13584-020-00414-0] [Medline: 33081834]
34. El Naamani K, Abbas R, Mukhtar S, El Fadel O, Sathe A, Kazan AS, et al. Telemedicine during and post-COVID 19: the insights of neurosurgery patients and physicians. J Clin Neurosci 2022 May;99:204-211 [FREE Full text] [doi: 10.1016/j.jocn.2022.03.006] [Medline: 35286972]
35. Silver SL, Lewis MN, Ledford CJW. A stepwise transition to telemedicine in response to COVID-19. J Am Board Fam Med 2021 Feb;34(Suppl):S152-S161 [FREE Full text] [doi: 10.3122/jabfm.2021.S1.200358] [Medline: 33622831]
36. Ramsey A, Mustafa SS, Portnoy JM. Patient and clinician attitudes toward telemedicine for allergy and immunology. J Allergy Clin Immunol Pract 2022 Oct;10(10):2493-2499 [FREE Full text] [doi: 10.1016/j.jaip.2022.05.008] [Medline: 35618214]
37. Ren W, Sun L, Tarimo CS, Li Q, Wu J. The situation and influencing factors of outpatient satisfaction in large hospitals: evidence from Henan Province, China. BMC Health Serv Res 2021 May 25;21(1):500 [FREE Full text] [doi: 10.1186/s12913-021-06520-2] [Medline: 34034724]

Abbreviations

NTCC: National Telemedicine Center of China
OR: odds ratio
TIPC: Telemedicine Informationization Professional Committee of China

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by W Zhu, D Dinh; comments to author 11.12.24; revised version received 07.01.25; accepted 23.01.25; published 26.02.25.

Please cite as:
Wang W, Chu Y, Cui F, Shi X, Zhang X, Sun D, Shi J, Zhao J
Correlation Between the Online Visiting Time and Frequency Increase in Telemedicine Services Offered by Health Care Providers Before, During, and After the COVID-19 Pandemic in China: Cross-Sectional Study
J Med Internet Res 2025;27:e65092
URL: https://www.jmir.org/2025/1/e65092
doi:10.2196/65092
PMID:

©Weiyi Wang, Yuntian Chu, Fangfang Cui, Xiaobing Shi, Xu Zhang, Dongxu Sun, Jinming Shi, Jie Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Telemedicine Use During the COVID-19 Pandemic in 8 Countries From the International Sexual Health and Reproductive Health Consortium: Web-Based Cross-Sectional Survey Study

Rayner Kay Jin Tan[1,2], PhD; Devon Hensel[3,4], PhD; Olena Ivanova[5,6], PhD; Raquel Gomez Bravo[7], PhD; Adesola Olumide[8], PhD; Emmanuel Adebayo[8], PhD; Amanda Cleeve[9,10], PhD; Amanda Gesselman[11], PhD; Sonam Jyoti Shah[12,13], MPH, JD; Helen Adesoba[8], PhD; Gifty Marley[2,14], PhD; Weiming Tang[2,13,14], PhD

[1]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
[2]University of North Carolina Project-China, Guangzhou, China
[3]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, United States
[4]Department of Sociology, Indiana University School of Medicine, Indianapolis, IN, United States
[5]Institute of Infectious Diseases and Tropical Medicine, LMU University Hospital, LMU Munich, Munich, Germany
[6]German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany
[7]Department of Behavioural and Cognitive Sciences, University of Luxembourg—Campus Belval, Belval, Luxembourg
[8]Institute of Child Health, College of Medicine, University of Ibadan and University College Hospital, Ibadan, Nigeria
[9]Department of Global Health, Karolinska Institutet, Stockholm, Sweden
[10]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden
[11]Kinsey Institute, Indiana University, Bloomington, IN, United States
[12]Institute for Global Health and Infectious Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[13]School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[14]Dermatology Hospital of Southern Medical University, Guangzhou, China

Corresponding Author:
Rayner Kay Jin Tan, PhD
Saw Swee Hock School of Public Health
National University of Singapore and National University Health System
12 Science Drive 2, #10-01 MD1
Singapore, 117549
Singapore
Phone: 65 91878576
Email: rayner.tan@nus.edu.sg

Abstract

Background: Telemedicine is an important way to fill in the access gap to in-person health care services during challenging times like pandemics.

Objective: This study aimed to investigate the role that telemedicine played during the COVID-19 pandemic by multicountry comparison of the use of telemedicine prior to and during the pandemic.

Methods: This study analyzes data from the second wave of the International Sexual Health and Reproductive Health study. This included data collected between April 2021 and July 2022 in 8 countries, including Armenia (n=296), Egypt (n=889), Germany (n=138), Moldova (n=311), Nigeria (n=205), Portugal (n=951), Singapore (n=13), and Spain (n=54). This study covered sociodemographics, sexual and reproductive health (SRH), and telemedicine use. Descriptive statistics and multilevel modeling were used to assess the factors influencing the use of telemedicine.

Results: Overall, 2857 participants were recruited. Approximately 57.6% (n=1646) of participants had never used telemedicine prior to COVID-19 measures, while 45.9% (n=1311) of participants required health care but reported not using telemedicine services following the introduction of COVID-19 measures. In high-income countries, the most common mode reported was audio-based telemedicine services, with 283 (71.8%) and 417 (73.5%) participants doing so before and during COVID-19, respectively. This was followed by text-based telemedicine services, with 152 (38.6%) and 173 (30.5%) participants doing so before and during COVID-19, respectively. In low- to middle-income countries, many participants also reported using audio-based

XSL•FO
RenderX

telemedicine services, with 288 (35.3%) and 237 (40.8%) participants doing so before and during COVID-19, respectively. This was followed by chat-based telemedicine services, with 265 (32.4%) and 217 (37.3%) participants doing so before and during COVID-19, respectively. Multilevel modeling revealed that those who were older (adjusted odds ratio [aOR] 0.99, 95% CI 0.99-1.00) and were in countries with a higher gross domestic product per capita (aOR 0.99, 95% CI 0.98-1.00) were less likely to have ever used telemedicine. Participants who were of male sex assigned at birth (aOR 0.79, 95% CI 0.65-0.96) were less likely to use telemedicine during the pandemic. Participants who perceived that they were worse off financially were more likely to have switched to telemedicine during COVID-19 (aOR 1.39, 95% CI 1.02-1.89) and were more likely to report having a poor or fair experience of telemedicine services (aOR 1.75, 95% CI 1.34-2.29). When sexual orientation was included in the model, nonheterosexual individuals were more likely to ever use telemedicine prior to COVID-19 (aOR 1.35, 95% CI 1.08-1.69), more likely to have used telemedicine during COVID-19 (aOR 1.58, 95% CI 1.24-2.02), and more likely to have switched to telemedicine during COVID-19 (aOR 1.55, 95% CI 1.09-2.21).

Conclusions: Telemedicine played a key role in addressing health care needs during the COVID-19 pandemic. Age, sex, economic status, and sexual orientation influenced its use.

(J Med Internet Res 2025;27:e60369) doi:10.2196/60369

KEYWORDS

COVID-19; telemedicine; sexual and reproductive health; pandemic; web-based survey; sexual health; reproductive health; communication technology; medical education; contraception; abortion; health care delivery; care; chronic condition

Introduction

The World Health Organization defines telemedicine as "healing at a distance," enabling the use of information and communication technology tools to improve the quality of medical services while overcoming the obstacles posed by travel [1]. It has also been described as "a branch of e-health that uses communication networks for the delivery of healthcare services and medical education from one geographic location to another" [2]. Telemedicine plays an important role in enhancing or substituting access to health care services, particularly during challenging circumstances like a pandemic or in hard-to-reach areas [3]. Moreover, telemedicine technologies have the potential to address unmet sexual and reproductive health (SRH) needs and improve SRH services. This is especially important as the World Health Organization estimates that everyone of reproductive age would not have access to at least 1 essential reproductive health intervention over the course of their lives [4].

The role of telemedicine in health care has grown rapidly and concomitantly with the expansion of communication technology. The significance of telemedicine was especially underscored during COVID-19 infection control measures and lockdowns. During this time, telemedicine provided the general public with a convenient and secure method to seek guidance from health care professionals regarding COVID-19 symptoms, preventive and treatment measures, psychological concerns, and various other matters [5]. It is estimated that about 116 million people had access to digital doctor consultations worldwide in 2024 compared to 57 million in 2019 [6]. During the pandemic, telemedicine was also recommended to be used for contraception counseling, shared decision-making, and managing potential side effects [7]. It was also used as a means of accessing abortion-related services during the pandemic [8]. Country-specific studies demonstrate that telemedicine was successfully implemented to provide SRH services during the pandemic, and it was a convenient and comfortable approach, especially for young people [9,10].

Telemedicine helped fill several health care service gaps that emerged during the pandemic, especially in the context of SRH [11]. First, telemedicine allowed patients to consult with health care providers remotely, reducing the need for in-person visits to health care facilities [12]. This was particularly important during the pandemic when minimizing physical contact was essential to curb the spread of the virus. Second, telemedicine platforms were used for COVID-19 screening and triage [5]. Patients could use digital visits to discuss symptoms with health care professionals, who could then determine the appropriate course of action, whether it be self-isolation, testing, or in-person care. Third, many routine and nonemergency health care services were disrupted due to lockdowns and restrictions [13,14]. Telemedicine provided a way for patients to continue receiving medical care for chronic conditions, follow-up appointments, medication management, and health services without needing to visit a health care facility. Fourth, the pandemic had a significant impact on mental health, and telemedicine became a valuable tool for delivering mental health services [15]. Digital counseling and therapy sessions allowed individuals to access support from the safety of their homes.

In the past 2021-2022 period, many countries have relaxed COVID-19 restrictions. This provided a unique opportunity to understand how social and behavioral factors related to decreasing COVID-19 restrictions may have influenced sexual behaviors in diverse settings, including the use of telemedicine. The second International Sexual Health and Reproductive Health (I-SHARE)-2 study examined the 2021-2022 period and focused on similar sexual or reproductive health data from I-SHARE-1, over and above the addition of telemedicine-related variables. The aim of this study is to describe the use of telemedicine before and during the COVID-19 pandemic in 8 countries. The data for this study are a part of the cross-sectional multicountry project called "International Sexual Health and Reproductive Health during COVID-19" [16]. The I-SHARE project brought together SRH researchers from different countries to conduct a web-based survey using a standardized instrument. The main aim of this multinational study was to gain insights into SRH

XSL•FO
RenderX

before and during the initial phase of the COVID-19 pandemic in each country.

The I-SHARE, multicountry, web-based, cross-sectional study was developed to assess SRH during and after COVID-19 restrictions. The initial multicountry study examined the 2020-2021 period and compared SRH in the pre–COVID-19 and COVID-19 periods [17,18]. In this study, we found that the pandemic hindered access to SRH care. However, telemedicine-related variables were not measured in the initial study. This study aimed to understand the prevalence of general telemedicine use and correlates of its use during the COVID-19 period as well as prior to COVID-19.

Methods

Study Design and Participants

This study analyzes data from the I-SHARE study. The consortium was developed to better understand SRH during the COVID-19 pandemic. The initial round of data collection (named the "I-SHARE survey") covered the period from July 20, 2020, to February 15, 2021. This I-SHARE-2 survey included data collected between April 2021 and July 2022 in 8 countries, including Armenia (n=296), Egypt (n=889), Germany (n=138), Moldova (n=311), Nigeria (n=205), Portugal (n=951), Singapore (n=13), and Spain (n=54).

Each country included details of SRH resources at the end of the survey. The in-country team translated the survey into local languages, field-tested the survey instrument, and submitted the protocol to a local institutional review committee for review. Field testing included giving the survey to at least 10 individuals who provided feedback about translation and sensitive topics. Some countries organized a second round of field testing using the web-based version of the survey questionnaire. All countries recruited participants on the web and through the help of community organizations. More details of the methods of the initial I-SHARE survey are described in the survey protocol [19]. The survey took approximately 15-20 minutes to complete. Open Data Kit software (version 1.16; Get ODK Inc) was used to collect data from participants on a cell phone, laptop, or other electronic device.

Eligibility criteria for participation included being at least 18 years of age, residing in the country where the survey was conducted, and being able to provide web-based informed consent. Researchers from each country were invited to join working groups focused on analyzing multicountry data. Survey data were included if they had received institutional review board approval and field-tested the survey. The survey questionnaire included domains on sociodemographic characteristics, sexual relationships, compliance with COVID-19 restrictions (eg, social distancing), web-based harassment through digital media, intimate partner violence, and HIV or sexually transmitted disease testing. The complete survey instrument is included in Table S1 in Multimedia Appendix 1.

Measures

Telemedicine use was measured through a series of questions regarding its (1) use prior to the introduction of COVID-19 measures (never, rarely, sometimes, often, and always), (2) use following the introduction of COVID-19 measures (yes, no, and did not need to access health care), (3) the various types of telemedicine modalities used (audio, video, text, chat, or other forms), as well as (4) participants' satisfaction with telemedicine services during COVID-19 (excellent, good, fair, and poor). While this was a survey that focused on SRH, our questions on telemedicine use were not limited to just SRH services, and we were not able to measure the use of telemedicine in the context of specific types of services. We constructed a new variable for participants who never used telemedicine prior to the introduction of COVID-19 measures but needed health care and switched to the use of telemedicine following the introduction of COVID-19 measures (switched to telemedicine).

Individual-level variables included sociodemographic characteristics such as sex assigned at birth, age (in years), sexual orientation, place of residence, educational attainment, employment status, partner living arrangement, perceived changes to one's economic situation as a result of COVID-19, and having children at home. Country-level variables included gross domestic product (GDP) per capita.

Data Analysis

Sociodemographic characteristics as well as frequencies and proportions around the use of telemedicine were summarized using descriptive statistics. Subsequently, we proceeded to conduct multilevel modeling (using Stata, version 16; StataCorp LLC), which included country-level variables. These multilevel models were constructed to assess the odds of ever using telemedicine, of using telemedicine during COVID-19 if health care services were needed, of switching to telemedicine during COVID-19, as well as having fair or poor experiences of telemedicine during COVID-19. We omitted sexual orientation from the main models, as participants in several countries largely omitted responding to this question but provided supplemental analyses including sexual orientation as a variable. Statistical significance was set at $P<.05$.

Ethical Considerations

Each country's ethical review committee approved the respective national web-based study prior to launch. We obtained approval from Ghent University (approval BC-07988) and the University of North Carolina at Chapel Hill (approval 295989) for multicountry analyses based on deidentified data, which allowed for secondary analyses without additional consent. A single data-sharing agreement covered multicountry analyses. In-country leads made final decisions about data sharing and data management practices. Survey participants were required to provide informed consent prior to the survey and were allowed to stop at any point and leave any item blank. We did not collect any personal identifiers and therefore all responses were anonymous. In-country leads also made decisions on the forms of remuneration provided to participants that were appropriate for the respective countries and approved by the local ethics board.

Results

A total of 2857 participants were recruited in the 8 countries of our sample. Most participants identified as cisgender female

XSL•FO
RenderX

(n=1757, 61.5%), were heterosexual or straight (n=1313, 70.8%), reported living in the city (n=1392, 49.7%), had completed college (n=1421, 51%), were formally employed (n=1663, 58.2%), were not partnered or not living with their partners (n=1714, 60%), perceived that the economic situation of their household had remained the same (n=1400, 50.6%), and did not have children (n=1538, 60.1%). The average age of participants was 37.4 (SD 16.0) years, and the average GDP per capita at the country level was US $13,900 (SD US $13,600). Table 1 demonstrates the sociodemographic attributes of the study sample.

XSL•FO
RenderX

Table 1. Sociodemographic attributes of International Sexual Health and Reproductive Health-2 study participants in 8 countries, April 2021 to July 2022 (N=2857).

Demographic variables	Values
Sex, n (%)	
Female	1757 (61.5)
Male	1088 (38.1)
Other	12 (0.4)
Total	2857 (100)
Age (years), mean (SD)	37.4 (16)
Sexual orientation, n (%)	
Heterosexual or straight	1313 (70.8)
Minority sexual orientation	542 (29.2)
Total	1855 (100)
Area, n (%)	
City	1392 (49.7)
Suburb of city	312 (11.1)
Town	640 (22.8)
Remote or rural area	420 (15)
Other	39 (1.4)
Total	2806 (100)
Schooling, n (%)	
Secondary school and below	626 (22.4)
Some college	609 (21.8)
Completed college	1421 (51)
Other	134 (4.8)
Total	2793 (100)
Employment status, n (%)	
Formally employed	1663 (58.2)
Not formally employed	1194 (41.8)
Total	2857 (100)
Relationship (participants can select multiple options; denominator is n=2857), n (%)	
Not partnered or not living with a partner (single or partnered but not living or separated)	1714 (60)
Partnered, living with partner	1143 (40)
Total	2857 (100)
Economic situation, n (%)	
Economic situation of household became worse	1155 (41.8)
Economic situation of household remained the same	1400 (50.6)
Economic situation of household became better	211 (7.6)
Total	2769 (100)
Has children, n (%)	
No	1538 (60.1)
Yes	1023 (39.9)
Total	2546 (100)
Gross domestic product per capita (US $), mean (SD)	13,900 (13,600)

XSL•FO
RenderX

Demographic variables	Values
Country income group, n (%)	
High-income country (Germany, Portugal, Singapore, and Spain)	1156 (40.5)
Low- to middle-income country (Armenia, Egypt, Moldova, and Nigeria)	1701 (59.5)
Total	2857 (100)

We have found that approximately 57.6% (n=1646) of participants had never used telemedicine before the introduction of COVID-19 measures, while 45.9% (n=1311) of participants required health care but reported not using telemedicine services following the introduction of COVID-19 measures. A total of 370 (26.3%) participants who never used telemedicine services prior to the introduction of COVID-19 measures, and required health care services during the pandemic, switched to using telemedicine during the pandemic. A total of 482 participants (33.3%) reported a fair or poor experience with using telemedicine services during COVID-19. Table 2 summarizes the frequencies and proportions for variables regarding telemedicine use before and following the implementation of COVID-19 measures as well as for participants who switched or did not switch to telemedicine use during COVID-19.

Table 2. Use of telemedicine among International Sexual Health and Reproductive Health-2 study participants in 8 countries, April 2021 to July 2022 (N=2857).

Telemedicine variables	Values, n (%)
Telemedicine use before the introduction of COVID-19 measures	
Never	1646 (57.6)
Rarely	519 (18.2)
Sometimes	496 (17.4)
Often	144 (5)
Always	52 (1.8)
Total	2857 (100)
Telemedicine use following the introduction of COVID-19 measures	
Yes	1148 (40.2)
No	1311 (45.9)
Did not need access to health care	398 (13.9)
Total	2459 (100)
Switched from not using telemedicine before to using telemedicine during COVID-19	
Yes	370 (26.3)
No	1036 (73.7)
Total	1406 (100)
Satisfaction with telemedicine services during COVID-19	
Excellent	316 (21.8)
Good	650 (44.9)
Fair	403 (27.8)
Poor	79 (5.5)
Total	1448 (100)

Table 3 provides a summary of the proportions of participants by country who reported ever using telemedicine services before COVID-19 as well as telemedicine use following the introduction of COVID-19 measures (excluding participants who did not require health care services during this time). In high-income countries, a majority of participants reported using audio-based telemedicine services, with 283 (71.8%) and 417 (73.5%) participants doing so before and during COVID-19, respectively. This was followed by text-based telemedicine services, with 152 (38.6%) and 173 (30.5%) participants doing so before and during COVID-19, respectively. In low- to middle-income countries, many participants also reported using audio-based telemedicine services, with 288 (35.3%) and 237 (40.8%) participants doing so before and during COVID-19, respectively. This was followed by chat-based telemedicine services, with 265 (32.4%) and 217 (37.3%) participants doing so before and during COVID-19, respectively, as well as text-based telemedicine services, with 265 (32.4%) and 217

XSL•FO
RenderX

(37.3%) participants doing so before and during COVID-19, respectively. Table 4 summarizes the frequencies and proportions of participants who reported using either audio, video, text, chat, or other forms of telemedicine modalities during COVID-19. Country-level data may be found in Figures S1 and S2 in Multimedia Appendix 1.

Table 3. Telemedicine use before and during COVID-19 measures by country among International Sexual Health and Reproductive Health-2 study participants in 8 countries, April 2021 to July 2022 (N=2857).

Country and time frame	Used telemedicine, n (%)	Did not use telemedicine, n (%)
Armenia		
Before COVID-19 (n=296)	168 (56.8)	128 (43.2)
During COVID-19 (n=251)	133 (53)	118 (47)
Egypt		
Before COVID-19 (n=889)	431 (48.5)	458 (51.5)
During COVID-19 (n=723)	261 (36.1)	462 (63.9)
Germany		
Before COVID-19 (n=138)	48 (34.8)	90 (65.2)
During COVID-19 (n=117)	40 (34.2)	77 (65.8)
Moldova		
Before COVID-19 (n=311)	136 (43.7)	175 (56.3)
During COVID-19 (n=280)	127 (45.4)	153 (54.6)
Nigeria		
Before COVID-19 (n=205)	82 (40)	123 (60)
During COVID-19 (n=173)	60 (34.7)	113 (65.3)
Portugal		
Before COVID-19 (n=951)	325 (34.2)	626 (65.8)
During COVID-19 (n=853)	485 (56.9)	368 (43.1)
Singapore		
Before COVID-19 (n=13)	3 (23.1)	10 (76.9)
During COVID-19 (n=12)	5 (41.7)	7 (58.3)
Spain		
Before COVID-19 (n=54)	18 (33.3)	36 (66.7)
During COVID-19 (n=50)	37 (74)	13 (26)

Table 4. Telemedicine modalities used before and during COVID-19 measures by country income groupings among International Sexual Health and Reproductive Health-2 study participants in 8 countries, April 2021 to July 2022 (N=2857).

Types of telemedicine modalities used	High-income countries[a]		Low- to middle-income countries[b]		All countries[c]	
	Before COVID-19, n (%)	During COVID-19, n (%)	Before COVID-19, n (%)	During COVID-19, n (%)	Before COVID-19, n (%)	During COVID-19, n (%)
Audio	283 (71.8)	417 (73.5)	288 (35.3)	237 (40.8)	571 (47.2)	654 (57)
Video	36 (9.1)	116 (20.5)	113 (13.8)	81 (13.9)	149 (12.3)	197 (17.2)
Text	152 (38.6)	173 (30.5)	259 (31.7)	211 (36.3)	411 (33.9)	384 (33.4)
Chat	44 (11.2)	50 (8.8)	265 (32.4)	217 (37.3)	309 (25.5)	267 (23.3)
Other	68 (17.3)	84 (14.8)	58 (7.1)	36 (6.2)	126 (10.4)	120 (10.5)

[a]A total of 394 participants used telemedicine before COVID-19, and 567 participants used telemedicine during COVID-19. Countries included Germany, Portugal, Singapore, and Spain.

[b]A total of 817 participants used telemedicine before COVID-19, and 581 participants used telemedicine during COVID-19. Countries included Armenia, Egypt, Moldova, and Nigeria.

[c]A total of 1211 participants used telemedicine before COVID-19, and 1148 participants used telemedicine during COVID-19.

XSL•FO
RenderX

In Table 5, we have summarized the binomial logistic multilevel model assessing the odds of ever using telemedicine, of using telemedicine during COVID-19 if health care services were needed, of switching to telemedicine during COVID-19, as well as having fair or poor experiences of telemedicine during COVID-19. First, for individuals who ever used telemedicine prior to COVID-19, those who were older (adjusted odds ratio [aOR] 0.99, 95% CI 0.99-1.00) and were in countries with a higher GDP per capita (aOR 0.99, 95% CI 0.98-1.00) were less likely to have ever done so. Second, for telemedicine use during COVID-19, those who were of male sex assigned at birth (aOR 0.79, 95% CI 0.65-0.96) were less likely to use telemedicine.

XSL•FO
RenderX

Table 5. Correlates for telemedicine variables among International Sexual Health and Reproductive Health-2 study participants in 8 countries, April 2021 to July 2022 (N=2857).

	Telemedicine variables			
	Telemedicine use before COVID-19 (n=2485), aOR[a] (95% CI)	Telemedicine use during COVID-19 (n=2147), aOR (95% CI)	Switch to telemedicine during COVID-19 (n=1259), aOR (95% CI)	Poor experience with telemedicine during COVID-19 (n=1271), aOR (95% CI)
Variables for main multivariable analysis				
Sex assigned at birth (reference=female)				
Male	1.16 (1.16-1.39)	*0.79* [b] *(0.65-0.96)* [c]	0.77 (0.56-1.06)	0.92 (0.69-1.23)
Other	8.22 (8.23-70.03)	4.91 (0.53-45.27)	__[d]	1.69 (0.31-9.11)
Age (years)	*0.99* [b] *(0.99-1.00)*	0.99 (0.98-1.01)	1 (0.99-1.02)	1 (0.99-1.02)
Place of residence (reference=city)				
Suburb	1.12 (1.12-1.47)	1.19 (0.89-1.6)	1.1 (0.72-1.69)	1.28 (0.87-1.87)
Town	1.06 (1.06-1.32)	8.42 (0.66-1.07)	0.97 (0.66-1.42)	0.95 (0.68-1.33)
Remote or rural area	0.85 (0.85-1.1)	0.8 (0.61-1.06)	0.75 (0.47-1.18)	1.09 (0.74-1.61)
Other area	0.69 (0.69-1.45)	0.59 (0.26-1.36)	0.68 (0.18-2.6)	2.11 (0.73-6.05)
Educational attainment (reference=secondary school and below)				
Some college	1.03 (1.03-1.34)	1.05 (0.79-1.39)	1.37 (0.87-2.15)	0.92 (0.62-1.35)
Completed college	1.1 (1.1-1.38)	1.09 (0.86-1.4)	1.23 (0.81-1.85)	0.83 (0.6-1.16)
Other	1.44 (1.44-2.19)	1.11 (0.71-1.75)	1.07 (0.51-2.26)	0.7 (0.37-1.3)
Employment status (reference=not formally employed)				
Formally employed	1.16 (1.16-1.4)	0.99 (0.81-1.22)	1.07 (0.78-1.46)	1.02 (0.78-1.34)
Economic situation (reference=stayed the same)				
Got worse	0.87 (0.87-1.04)	1.02 (0.84-1.24)	*1.39* [b] *(1.02-1.89)*	*1.75* [e] *(1.34-2.29)*
Got better	1.26 (1.26-1.72)	1.06 (0.75-1.5)	1.18 (0.67-2.11)	1.29 (0.8-2.06)
Cohabitation status (reference=not living with a partner)				
Living with partner	1.08 (1.08-1.31)	1.13 (0.91-1.39)	0.82 (0.59-1.14)	0.88 (0.66-1.17)
Children at home				
Has children at home	1.13 (1.13-1.45)	1.08 (0.83-1.41)	1.05 (0.69-1.61)	1.01 (0.71-1.42)
Gross domestic product per capita (US $, thousands)	*0.99* [f] *(0.98-1.00)*	1.01 (0.99-1.03)	1.02 (0.98-1.05)	1.00 (0.98-1.01)
Supplementary multivariable analysis with sexual orientation variable[g,h]				
Sexual orientation (reference=heterosexual)				
Nonheterosexual	*1.35* [f] *(1.08-1.69)*	*1.58* [e] *(1.24-2.02)*	*1.55* [b] *(1.09-2.21)*	0.91 (0.66-1.26)

[a]aOR: adjusted odds ratio.

[b]*P*<.05.

[c]Values in italics format indicate findings where *P*<.05 (statistical significance).

[d]Not available.

[e]*P*<.001.

[f]*P*<.01.

[g]Please note that sexual orientation was not measured systematically in several countries and therefore was not included as a variable in the main analysis. However, it was an important factor; therefore, we have chosen to display the role of sexual orientation in telemedicine use, adjusted for all of the above variables in a separate model.

[h]Telemedicine use before COVID-19: n=1657, telemedicine use during COVID-19: n=1467, switched to telemedicine during COVID-19: n=881, and poor experience with telemedicine during COVID-19: n=870.

Third, for those who never used telemedicine prior to COVID-19 but switched to using telemedicine during COVID-19, participants who perceived that they were worse off financially were more likely to have switched to telemedicine during

XSL•FO
RenderX

COVID-19 (aOR 1.39, 95% CI 1.02-1.89). For those who reported a fair or poor experience of telemedicine, participants who perceived that they were worse off financially were more likely to report doing so (aOR 1.75, 95% CI 1.34-2.29). When sexual orientation was included in the model, we noted that nonheterosexual individuals were more likely to ever use telemedicine prior to COVID-19 (aOR 1.35, 95% CI 1.08-1.69), more likely to have used telemedicine during COVID-19 (aOR 1.58, 95% CI 1.24-2.02), and more likely to have switched to telemedicine during COVID-19 (aOR 1.55, 95% CI 1.09-2.21).

Discussion

Principal Findings

This study explored the use of telemedicine during COVID-19 through an analysis of the I-SHARE-2 study that took place amid the lifting of COVID-19 restrictions. Slightly over half the participants we surveyed had never used telemedicine prior to the COVID-19 pandemic. However, this proportion increased during the pandemic, with about a quarter of those who never used telemedicine before switching to the use of telemedicine. About a third of participants reported a fair or poor telemedicine experience during this time period. This study extends the literature through an analysis of multiple countries and provides insight into the use of different telemedicine modalities and the factors associated with them during the COVID-19 pandemic. In this section, we discuss these findings considering the available evidence and suggest implications of our data for policy and future research.

Comparison to Prior Work

Our study found that about less than half of the participants had ever used telemedicine services prior to the COVID-19 pandemic. This is lower than estimates published by Ipsos in 2018, which estimated that globally, only 10% of individuals ever reported using telemedicine. Nevertheless, such estimates vary widely across contexts, with Saudi Arabia (31%), India (27%), China (24%), Malaysia (15%), and the United States (15%) being the highest adopters in 2018 [20].

Furthermore, we found that those who were older were less likely to have used telemedicine prior to the pandemic. The finding on the association between older age and lower use of telemedicine has been supported in past studies around the world due to barriers that older patients face in terms of accessing these technological platforms [21-23]. It is worth noting that we did not find an association between age and switching to the use of telemedicine during the pandemic. The findings on this association are mixed. Some contexts find no difference in age and use rates for telemedicine [23]. However, this contrasts with other studies, which found that older adults were less likely to take up telemedicine services compared to younger people during the pandemic [24-26]. Nevertheless, the negative correlation between increasing age and low telemedicine use is not universal and may differ across country settings, disease areas, and types of telemedicine platforms (eg, video, audio, or text-based) [27]. Further nuanced research on different older adult populations, as well as modes of using telehealth, is warranted.

Our study also indicated that individuals who identified as sexual minorities were more likely to have ever used telemedicine services. They were also more likely to have used telemedicine services during the pandemic. These results are unsurprising, given that traditional facility-based health services have often been associated with stigma and suboptimal care for sexual minorities [19,28] and that digital health interventions targeted at sexual minorities have been expanding across most settings [29]. There has also been evidence to show that services for sexual minorities had successfully pivoted to provide telehealth services during the pandemic, thus offering targeted interventions for this population that might have increased their access to telemedicine services.

The adoption of telemedicine during the pandemic, among individuals who previously never used telemedicine, was also influenced by economic circumstances. Our results indicated that those who perceived themselves as financially worse off during the pandemic were more likely to switch into digital care but at the same time also report fair or poor telemedicine experiences. This substantial proportion of individuals reporting fair or poor experiences with telemedicine use is noteworthy, given that a systematic review on telemedicine use in low- to middle-income countries found high patient satisfaction with such services [30]. Another systematic review found high satisfaction across multiple medical disciplines [31]. Ultimately, our data may reflect how some individuals who may have competing financial demands may not have been able to leverage technology effectively to enjoy telemedicine services compared to others who were more financially stable or better off during the pandemic.

Audio-based telemedicine services were the most used platform across both high- and low-income countries in our sample. We noted that more participants in high-income countries switched to the use of video-based telemedicine services (before: n=36, 9.1% and during: n=116, 20.5%) compared to those in low-income countries (before: n=113, 13.8% and after: n=81, 13.9%). We also note that chat-based services were the second most popular platform for the use of telemedicine services in low-income countries in our sample. These findings may reflect disparities in technological and telecommunication infrastructures between high- and low-income settings, which may shape access to telemedicine. Studies have found that chat-based platforms may be preferred over video-based platforms in low- to middle-income countries due to the reduced need for data bandwidth and stability of communications [30].

We are mindful of several limitations of this study. This study relies on self-reported data, which may be subject to recall bias and social desirability bias. Participants might not accurately remember their use of telemedicine or may report what they perceive as socially acceptable answers. There is also the possibility of nonresponse bias of those individuals who did not participate in the survey and might have different experiences or views about telemedicine compared to those who participated. Moreover, the survey was conducted on the web, which might result in a selection bias where respondents who are more technology-savvy or have better access to digital devices or internet services were recruited. Such participants may already have better experiences of telemedicine given their technological

XSL•FO
RenderX

savvy. The cross-sectional design also constitutes a limitation to establish causality. Additionally, the study methodology is basically quantitative, which may not capture the nuanced experiences, perceptions, and challenges faced by users of telemedicine. As only 8 countries have been included, the findings may not be generalizable to all populations and may not represent the global diversity in health care systems, technological infrastructure, and cultural attitudes toward telemedicine. Moreover, the study groups various forms of telemedicine (audio, video, text, and chat) into a single category. The experiences and effectiveness of these different modalities can vary greatly, and lumping them together might oversimplify the analysis.

Conclusions

Most of the participants had never used telemedicine before the pandemic, and a significant proportion did not use it either following the introduction of COVID-19 measurements. Only a quarter of the participants who never used it before switched into it during the pandemic. However, about one-third of the participants reported fair or poor experiences using it during COVID-19. Telemedicine has emerged as a critical tool for expanding access to health care services, particularly in the field of SRH. The COVID-19 pandemic has accelerated its adoption, highlighting the importance of telemedicine during crises, emergencies, or natural disasters. Understanding all the factors that influence the implementation of digital consultations, barriers and facilitators could help not only health care professionals but also policy makers and stakeholders to tailor strategies that improve its accessibility and effectiveness in delivering proper care. Finally, the adoption and implementation of telemedicine may be cost-saving but might also incur costs that vary greatly across different countries and contexts due to diverse health system configurations and digital and telecommunication infrastructures [32]. Future work should therefore consider how telemedicine services can be used to deliver care to people in a cost-effective manner, especially across high-income and low- to middle-income settings.

Acknowledgments

The authors acknowledge the research teams in Armenia, Egypt, Germany, Moldova, Nigeria, Portugal, Singapore, and Spain who contributed data to International Sexual Health and Reproductive Health-2 as well as the International Sexual Health and Reproductive Health-1 team for the initial work on setting up the consortium.

Conflicts of Interest

DH is a paid research consultant with For Goodness Sake, LLC. All other authors declared no conflicts of interest.

Multimedia Appendix 1

Supplementary data containing the survey instrument as well as figures summarizing telemedicine modalities used before and during COVID-19 measures by country among International Sexual Health and Reproductive Health-2 study participants in 8 countries, April 2021 to July 2022.

[DOCX File , 33 KB - jmir_v27i1e60369_app1.docx]

References

1. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth. World Health Organization. 2010. URL: https://iris.who.int/handle/10665/44497 [accessed 2025-02-03]
2. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
3. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
4. Universal access to sexual and reproductive health. World Health Organization. 2022. URL: https://www.who.int/news/item/19-07-2022-universal-access-to-sexual-and-reproductive-health [accessed 2024-11-26]
5. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
6. Telemedicine—statistics and facts. Statista. 2024. URL: https://www.statista.com/topics/12106/telemedicine/#topicOverview [accessed 2024-11-26]
7. Nanda K, Lebetkin E, Steiner MJ, Yacobson I, Dorflinger LJ. Contraception in the era of COVID-19. Glob Health Sci Pract 2020;8(2):166-168. [doi: 10.9745/ghsp-d-20-00119]
8. Qaderi K, Khodavirdilou R, Kalhor M, Behbahani BM, Keshavarz M, Bashtian MH, et al. Abortion services during the COVID-19 pandemic: a systematic review. Reprod Health 2023;20(1):61 [FREE Full text] [doi: 10.1186/s12978-023-01582-3] [Medline: 37055839]
9. Bittleston H, Goller JL, Temple-Smith M, Hocking JS, Coombe J. Telehealth for sexual and reproductive health issues: a qualitative study of experiences of accessing care during COVID-19. Sex Health 2022;19(5):473-478. [doi: 10.1071/sh22098]

XSL•FO
RenderX

10. Cheng Y, Boerma C, Peck L, Botfield JR, Estoesta J, McGeechan K. Telehealth sexual and reproductive health care during the COVID-19 pandemic. Med J Aust 2021;215(8):371-372 [FREE Full text] [doi: 10.5694/mja2.51219] [Medline: 34374090]
11. VanBenschoten H, Kuganantham H, Larsson EC, Endler M, Thorson A, Gemzell-Danielsson K, et al. Impact of the COVID-19 pandemic on access to and utilisation of services for sexual and reproductive health: a scoping review. BMJ Glob Health 2022;7(10):e009594 [FREE Full text] [doi: 10.1136/bmjgh-2022-009594] [Medline: 36202429]
12. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
13. Elson EC, Oermann C, Duehlmeyer S, Bledsoe S. Use of telemedicine to provide clinical pharmacy services during the SARS-CoV-2 pandemic. Am J Health Syst Pharm 2020;77(13):1005-1006 [FREE Full text] [doi: 10.1093/ajhp/zxaa112] [Medline: 32369111]
14. Perrin PB, Pierce BS, Elliott TR. COVID-19 and telemedicine: a revolution in healthcare delivery is at hand. Health Sci Rep 2020;3(2):e166 [FREE Full text] [doi: 10.1002/hsr2.166] [Medline: 32500101]
15. Arafat MY, Zaman S, Hawlader MDH. Telemedicine improves mental health in COVID-19 pandemic. J Glob Health 2021;11:03004 [FREE Full text] [doi: 10.7189/jogh.11.03004] [Medline: 34326984]
16. Michielsen K, Larrson EC, Kågesten A, Erausquin JT, Griffin S, van de Velde S, et al. International Sexual Health and Reproductive Health (I-SHARE) survey during COVID-19: study protocol for online national surveys and global comparative analyses. Sex Transm Infect 2021;97(2):88-92. [doi: 10.1136/sextrans-2020-054664] [Medline: 33082232]
17. Erausquin J, Tan RKJ, Uhlich M, Francis JM, Kumar N, Campbell L, et al. The International Sexual Health and Reproductive Health Survey (I-SHARE-1): a multi-country analysis of adults from 30 countries prior to and during the initial COVID-19 wave. Clin Infect Dis 2022;75(1):e991-e999 [FREE Full text] [doi: 10.1101/2021.09.18.21263630] [Medline: 34704103]
18. Michielsen K, Campbell L, de la Torre FF. The impact of COVID-19 on sexual and reproductive health in Eastern Europe and Central Asia. UNFPA. 2022 Sep 26. URL: https://eeca.unfpa.org/en/publications/impact-covid-19-sexual-and-reproductive-health-eastern-europe-and-central-asia [accessed 2025-02-21]
19. Tan RKJ, Kaur N, Kumar PA, Tay E, Leong A, Chen MI, et al. Clinics as spaces of costly disclosure: HIV/STI testing and anticipated stigma among gay, bisexual and queer men. Cult Health Sex 2020;22(3):307-320. [doi: 10.1080/13691058.2019.1596313] [Medline: 30975036]
20. Global views on healthcare in 2018. Ipsos. 2018. URL: https://www.ipsos.com/en-sg/global-views-healthcare [accessed 2024-11-26]
21. Pasquinelli MM, Patel D, Nguyen R, Fathi J, Khan M, Fernandez K, et al. Age-based disparities in telehealth use in an urban, underserved population in cancer and pulmonary clinics: a need for policy change. J Am Assoc Nurse Pract 2022;34(5):731-737. [doi: 10.1097/JXX.0000000000000708] [Medline: 35353071]
22. Braswell M, Wally MK, Kempton LB, Seymour RB, Hsu JR, Karunakar M, et al. Age and socioeconomic status affect access to telemedicine at an urban level 1 trauma center. OTA Int 2021;4(4):e155 [FREE Full text] [doi: 10.1097/OI9.0000000000000155] [Medline: 34765905]
23. Hayrapetian L, Zepp M, Rao S, Hennessey M, Houle M, Atienza M, et al. Expanding telehealth options during the COVID pandemic eliminated racial and age disparities in electronic communication by inflammatory bowel disease patients. J Natl Med Assoc 2021;113(4):474-477. [doi: 10.1016/j.jnma.2021.03.005] [Medline: 33941370]
24. Miyawaki A, Tabuchi T, Ong MK, Tsugawa Y. Age and social disparities in the use of telemedicine during the COVID-19 pandemic in Japan: cross-sectional study. J Med Internet Res 2021;23(7):e27982 [FREE Full text] [doi: 10.2196/27982] [Medline: 34259641]
25. Choi NG, DiNitto DM, Marti CN, Choi BY. Telehealth use among older adults during COVID-19: associations with sociodemographic and health characteristics, technology device ownership, and technology learning. J Appl Gerontol 2022;41(3):600-609 [FREE Full text] [doi: 10.1177/07334648211047347] [Medline: 34608821]
26. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
27. Friedman EE, Devlin SA, Gilson SF, Ridgway JP. Age and racial disparities in telehealth use among people with HIV during the COVID-19 pandemic. AIDS Behav 2022;26(8):2686-2691 [FREE Full text] [doi: 10.1007/s10461-022-03607-7] [Medline: 35133528]
28. O'Hara CA, Foon XL, Ng JC, Wong CS, Wang FY, Tan CY, et al. Lesbian, gay, bisexual, transgender, queer and intersex (LGBTQI+) healthcare in Singapore: perspectives of non-governmental organisations and clinical year medical students. Med Educ Online 2023;28(1):2172744 [FREE Full text] [doi: 10.1080/10872981.2023.2172744] [Medline: 36744296]
29. Waad A. Caring for Our Community: telehealth interventions as a promising practice for addressing population health disparities of LGBTQ+ communities in health care settings. Dela J Public Health 2019;5(3):12-15 [FREE Full text] [doi: 10.32481/djph.2019.06.005] [Medline: 34467036]
30. Tiwari BB, Kulkarni A, Zhang H, Khan MM, Zhang DS. Utilization of telehealth services in low- and middle-income countries amid the COVID-19 pandemic: a narrative summary. Glob Health Action 2023;16(1):2179163 [FREE Full text] [doi: 10.1080/16549716.2023.2179163] [Medline: 36803538]

XSL•FO
RenderX

31. Pogorzelska K, Chlabicz S. Patient satisfaction with telemedicine during the COVID-19 pandemic—a systematic review. Int J Environ Res Public Health 2022;19(10):6113 [FREE Full text] [doi: 10.3390/ijerph19106113] [Medline: 35627650]
32. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]

Abbreviations

aOR: adjusted odds ratio
GDP: gross domestic product
I-SHARE: International Sexual Health and Reproductive Health
SRH: sexual and reproductive health

Edited by A Mavragani; submitted 09.05.24; peer-reviewed by A Abdelhamed, C Lapadula; comments to author 06.11.24; revised version received 28.11.24; accepted 08.12.24; published 04.03.25.

Please cite as:
Tan RKJ, Hensel D, Ivanova O, Bravo RG, Olumide A, Adebayo E, Cleeve A, Gesselman A, Shah SJ, Adesoba H, Marley G, Tang W Telemedicine Use During the COVID-19 Pandemic in 8 Countries From the International Sexual Health and Reproductive Health Consortium: Web-Based Cross-Sectional Survey Study
J Med Internet Res 2025;27:e60369
URL: https://www.jmir.org/2025/1/e60369
doi:10.2196/60369
PMID:40053813

©Rayner Kay Jin Tan, Devon Hensel, Olena Ivanova, Raquel Gomez Bravo, Adesola Olumide, Emmanuel Adebayo, Amanda Cleeve, Amanda Gesselman, Sonam Jyoti Shah, Helen Adesoba, Gifty Marley, Weiming Tang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Use of Remote Assessment Tools to Substitute Routine Outpatient Care: Scoping Review

Jayne Barclay[1*], BAPSC, MICT; Clair Sullivan[2,3*], PhD; Michael Beckmann[1,4*], PhD; Graeme Mattison[5,6], MBChB; Rebecca Runciman[4], BSc; Elizabeth Martin[1,7*], PhD

[1]School of Medicine, University of Queensland, Brisbane, Australia

[2]Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[3]Clinical Informatics Director (Research), Metro North HHS, Queensland Health, Brisbane, Australia

[4]Mater Misericordiae Ltd, Brisbane, Australia

[5]QLD Digital Health Centre, University of Queensland, Brisbane, Australia

[6]Department of Respiratory Medicine, The Mater Hospital Brisbane, Brisbane, Australia

[7]Wesley Research Institute, Brisbane, Australia

[*]these authors contributed equally

Corresponding Author:
Jayne Barclay, BAPSC, MICT
School of Medicine
University of Queensland
St Lucia
Brisbane, 4072
Australia
Phone: 61 425302145
Email: j.barclay@uq.net.au

Abstract

Background: The increasing global demand for health care, driven by demographic shifts, the rise of personalized medicine, and technological innovations necessitate novel approaches to health care delivery. Digital remote assessment tools have emerged as a promising solution, enabling hybrid care models that combine traditional and remote patient management. These tools support the quadruple aim of health care by enhancing the monitoring and evaluation of patient-reported data, thereby improving patient care, boosting operational efficiency, reducing costs, and improving the experience of patients and clinicians. This review seeks to understand how remote assessment tools are used for routine consultation substitution in adult tertiary care centers.

Objective: This scoping review aims to evaluate the implementation and health outcomes of digital remote assessment tools used for routine consultation substitutions in adult tertiary care centers. The objectives include assessing the extent of use, types, and effectiveness of these tools in substituting conventional outpatient care.

Methods: A comprehensive scoping review was conducted, adhering to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. The review focused on studies that used internet-dependent remote assessment technologies for patient data transfer in tertiary care settings. A detailed search strategy was used across multiple databases, with studies selected based on predefined inclusion and exclusion criteria. Data extraction and analysis were performed by independent reviewers, with a focus on the functionalities of the tools and their alignment with the Quadruple Aim of Healthcare.

Results: The review included 12 studies, highlighting a growing interest in remote assessment technologies across diverse clinical settings. The interventions varied in length, from 4 weeks to 12 months, and demonstrated a range of functionalities, including symptom monitoring and postsurgical follow-ups. The use of these tools was associated with improved clinical outcomes, such as timely intervention for clinical deterioration and enhanced clinical protocol adherence. Additionally, a small number of studies identified potential cost savings in terms of reduced unplanned health care contacts and optimized clinical resource use. Patient and clinician experiences were generally positive, with high adherence to remote assessments and an appreciation for the personalized and timely care facilitated by these technologies. Barriers included high initial setup costs for digital technologies, leading to an inflated cost per patient in small sample studies.

XSL•FO
RenderX

Conclusions: Digital remote assessment tools offer significant potential to enhance health care delivery by improving health outcomes, reducing costs, and enriching patient and clinician experiences. Their flexibility and adaptability make them suitable for various clinical contexts, supporting the personalization of care and operational efficiency. While digital remote assessment tools offer significant potential, careful consideration of implementation strategies, equity, cost, and clinician and patient experiences is crucial for successful clinical integration.

(J Med Internet Res 2025;27:e65938) doi:10.2196/65938

KEYWORDS

remote assessment; remote patient monitoring; telemedicine; mHealth; applications; patient-reported outcome measures; self-reported; health care cost reduction; hybrid care models

Introduction

Background

Globally, the demand for health care services is set to rise due to an increasing population, an aging demographic, the rise of personalized medicine, and the introduction of innovative technologies. Escalating costs, digital disruption, and consumer-centered care are likely to impact the effectiveness and viability of health care systems to meet this demand [1]. Novel approaches to health care are needed.

Well-established digital health interventions, such as telehealth have only marginally enhanced the capacity of public health systems and health care services [2]. Efforts to explore new strategies for enhancing health system sustainability persist, including digital interventions that use data to triage patients to optimal care pathways [3]. Remote management of diagnosed chronic diseases using biometric and wearable technologies is also becoming more common [3]. Such interventions streamline patient screening and triage, center the process around the patient, and in some instances, obviate the need for physical tertiary facility visits [4]. Provider-delivered remote monitoring tools and digital patient-reported outcome measurement tools are becoming prevalent. These approaches enable the introduction of new care models that may alleviate operational health system pressures and the realization of personalized care models [5].

Digital remote assessment tools are an emerging area with the potential to revolutionize traditional patient care approaches. These technologies enable hybrid care delivery models consisting of asynchronous patient-reported follow-up outside of conventional care settings and enable health providers to clinically supervise remotely located patients [5]. Such tools serve to capture, monitor, evaluate, and guide care decisions undertaken by health care professionals using the data reported by the patient, including pain, function, and symptoms.

Remote assessment tools cover a broad range of methods and technologies that can deliver interactive, patient-provider communications remotely through the digital collection of clinical assessment data. The implementation of care substitution using technology could yield benefits for both patients and providers by reducing unplanned health care contact post surgery, decreasing travel and costs, and optimizing workflows and human resources [6,7]. There is also some evidence that substituting postoperative follow-up visits with clinical assessment data collected from smartphone apps is acceptable to patients [6,7]. The broad use of "care substitution" encompasses various technologies that facilitate patient-provider interactions through the digital collection of clinical assessment data, thereby offering an alternative to traditional face-to-face or videoconferencing consultation.

The Quadruple Aim of Healthcare [8] provides a comprehensive framework for evaluating new interventions by focusing on 4 key dimensions: improving population health, enhancing patient and clinician experiences, and reducing health care costs, offering a holistic measure of an intervention's impact and effectiveness. This general approach is particularly valuable in assessing diverse interventions, as it captures a wide range of outcomes and benefits, ensuring a thorough evaluation despite the variability in study designs and methodologies. The integration of innovative technologies to complement or substitute existing clinical pathways warrants further exploration, particularly the use of asynchronous remote consultations to replace routine tertiary facility care. There are no published systematic or broad scoping reviews that assess the outcomes of using only remote assessment tools to substitute routine outpatient care.

This review seeks to understand, through the available literature: How are adult patient interventions using remote assessment tools used, practically, and effectively, for routine consultation substitution in tertiary care centers?

Study Objectives

The objectives of this scoping review are to (1) investigate how many studies have used remote assessment tools as the primary data collection or assessment method for routine care substitution; (2) explore what are the types, frequencies, and durations of interventions using digital assessment tools; and (3) map the published outcomes of interventions using digital assessment tools, including how they are evaluated and measured against the quadruple aim framework. This research can inform health care systems and providers regarding their appropriate use.

Methods

Overview

A scoping review was conducted, following PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) methods [9], because of the broad research questions and the emerging field of digital remote assessments. The PRISMA-ScR checklist,

XSL•FO
RenderX

provided in Multimedia Appendix 1, was used to ensure comprehensive reporting and methodological rigor.

Study Selection and Search Strategy

The intervention of interest for this scoping review was remote assessment tools using digital internet technologies incorporating patient-reported data in a tertiary health setting. The full inclusion and exclusion criterion developed is shown in Textbox 1.

The initial literature search was conducted in April 2023. An updated search was completed in March 2024. The search terms used to describe the intervention were "survey," "telemedicine," "mobile applications," and "tertiary" with synonyms built out from these initial terms. Studies published in the last 10 years in English and indexed in the following databases were searched: PubMed, CINAHL, Embase, Web of Science, and Cochrane. The PubMed search strategy is included in Multimedia Appendix 2.

Textbox 1. Search strategy inclusion and exclusion criteria.

Inclusion criteria

- Study type: Peer-reviewed, full-length studies.
- Language: English
- Study design: Primary intervention studies with or without a control group, including randomized controlled trials, nonrandomized controlled studies, controlled before-after studies, observational cohort studies, and preliminary and pilot studies.
- Population: Study participants aged 18 years or older, who are under management of a hospital or similar tertiary facility.
- Intervention: Remote assessments using internet-dependent technologies to transfer patient-reported data from the patient's home to a hospital institution or tertiary health provider as the primary intervention.
- Comparison: Health facility or hospital where remote patient assessments were compared to standard or routine care, specifically face-to-face or in-person clinical consultations
- Outcome: Hospitalization rate, mortality, quality of life, health service efficiency, economic impact, clinical workflows, patient experience, clinician experience, and access to health services.

Exclusion criteria

- Study type: Systematic reviews, scoping reviews, study protocols, conference abstracts, case studies, or editorials or commentary.
- Language: Languages other than English.
- Study design: Studies without primary data, descriptive studies only, theoretical frameworks, conference abstracts, or posters.
- Population: Pediatric patients and patients managed outside of a tertiary or hospital team setting.
- Intervention: Exclude video consultations and technologies that directly transfer data from biometric or wearable devices; exclude the use of photographic data; and exclude studies that use educational information or content.
- Comparison: Studies assessing rehabilitation, exercise, mobility, or clinical education use and adoption compared to in-person consultations.
- Outcome: Educational outcomes and training effectiveness

All identified citations were uploaded into EndNote (version 20.6/2023; Clarivate Analytics) and initial duplicates were removed. The remaining studies were then imported into Covidence, a web-based literature review platform, where the identification of studies was undertaken by 3 independent reviewers. All 3 reviewers screened the titles and abstracts for eligibility, with any combination of 2 reviewers per study. A study was included if both reviewers agreed that it met the inclusion criteria. Conflicts in study selection were resolved during team meetings. Full-text review of studies was randomly distributed across the team of 3 reviewers, with any combination of 2 reviewers requiring agreement for the study to be included. Extraction of data from the final full-text studies was shared across all reviewers, with all reviewers systematically and independently documenting data as per a previously agreed data extraction format. The data extraction fields were then compared, discussed, and finalized with all reviewers present.

To help understand the functions of the tools used, the data was used to create a functionality score based on the Incidental Medical Services (IMS) Institute for Healthcare Informatics functionality score description [10]. This functionality score consists of 7 functionality criteria and 4 functional subcategories. The complete structure of the IMS Institute for Healthcare Informatics functionality scoring can be found in Table 1. Each intervention was scored using 1 to indicate function was present, and 0 function was not present.

To facilitate consistency across the studies, the results were mapped across the 4 dimensions of the Quadruple Aim of Healthcare to assist in the comparison of benefits and outcomes. The quadrants used were (1) improve population health, (2) enhance patient experience, (3) enhance clinician experience, and (4) reduce health system costs. In the context of this review, factors contributing to the reduction of health system costs under investigation included cost of care or service (overall operational cost to provide a service), workforce use, and reduction in avoidable or unnecessary hospital admissions.

XSL•FO
RenderX

Table 1. IMS[a] Institute for Healthcare Informatics functionality criteria assessment [10].

Functionality criteria	Description
Inform	Provides information in a variety of formats (text, photo, and video)
Instruct	Provides instructions to the user
Record	Capture user entered data
Collect data	Able to enter and store health data
Share data	Able to transmit health data
Evaluate data	Able to evaluate the entered health data by patient and provider, provider and administrator, or patient and caregiver
Intervene	Able to send alerts based on the data collected or propose behavioral intervention or changes
Display	Graphically display user entered data or output user–entered data
Guide	Provide guidance based on user entered information, and may further offer a diagnosis, or recommend a consultation with a physician or a course of treatment
Remind or alert	Provides reminders to the user
Communicate	Provide communication with health care provider or provide links to social network

[a]IMS: Incidental Medical Services.

Results

Overview

The literature searches yielded 5758 unique citations, of which 12 studies were included in the final review. Figure 1 demonstrates the combined search results, with the first search completed in April 2023, and an updated search completed in March 2024.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for combined searches.

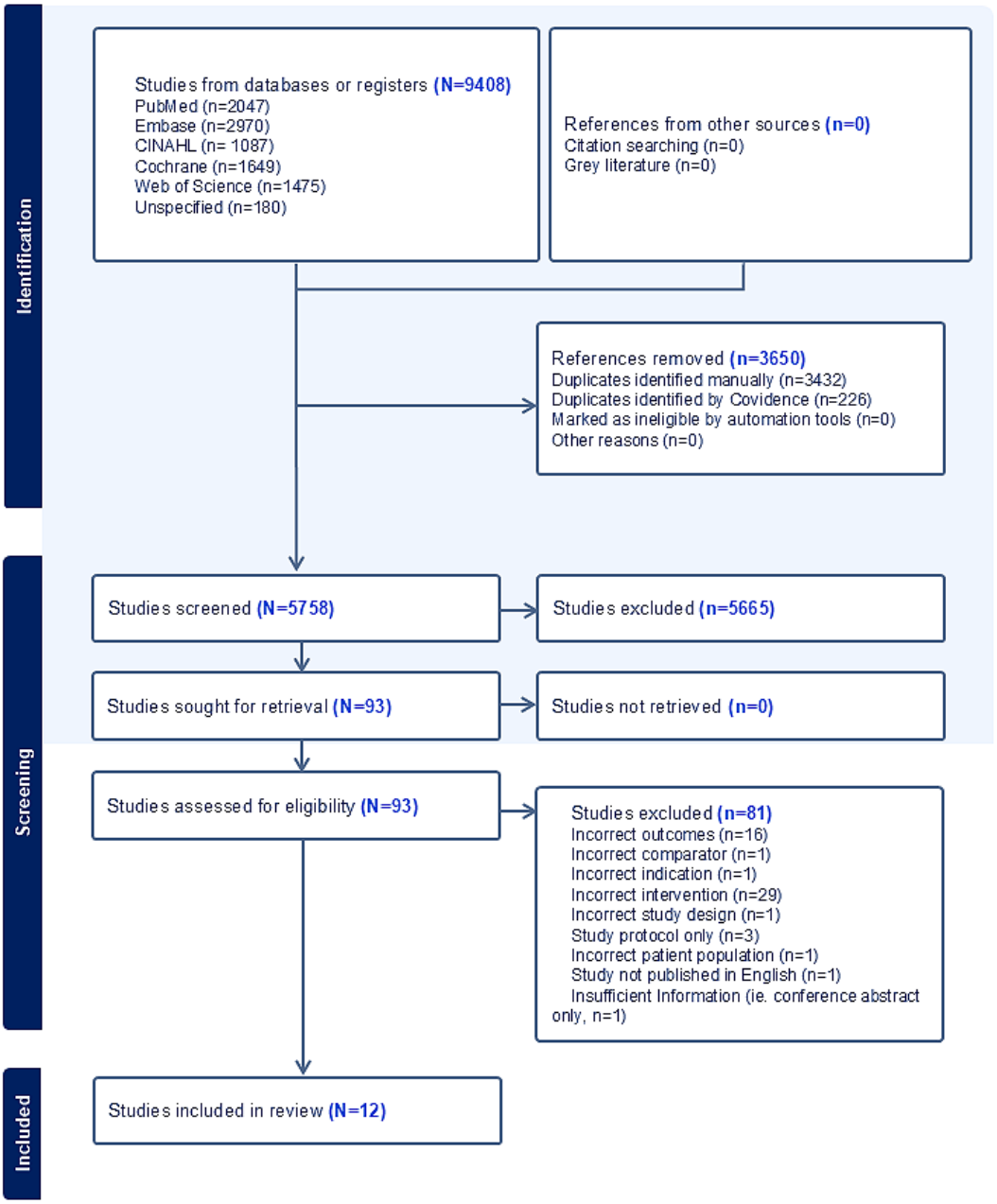

Summary of Included Studies

The 12 included studies span a publication period from June 2020 to January 2024, illustrating a growing contemporary interest in the domain of digital remote assessments. Geographically, the research landscape is diverse with studies originating from regions including the United Kingdom [11-13], Europe [13-17], North America [18,19], Asia [20], and Africa [21]. As for clinical focus, the studies encompass a wide range of patient populations, including those undergoing cancer treatment, individuals with cardiac or rheumatoid conditions, and postsurgery patients. Five of the studies included specific participant criteria that included participants who communicated in a specific or native language [12,15-17,19].

Intervention descriptions are varied, from real-time digital monitoring systems to smartphone apps. While some studies use the term “remote symptom monitoring” explicitly, others refer to “digital health tool,” “mHealth,” or “connected monitoring.” It should be noted that there were a considerable number of conference extracts that were excluded due to insufficient information published. Characteristics of included studies are presented in Multimedia Appendix 3 [11-22].

Types of Remote Assessment Interventions

Observations from the results on interventions involving remote assessment tools in various clinical scenarios reveal certain care models and distinctions. The length of interventions varied across studies, ranging from a period of 4 weeks [14] to a more extended duration of up to 12 months [20], with others such as Sanyal et al [18] remaining unspecified.

Data Sampling

Some studies adopted a weekly sampling approach [13,14,22], while others used more intermittent sampling, evident in Liu et al's method of sampling [20] at 90, 180, and 365 days. Richards

XSL•FO
RenderX

et al [12] introduced a mixed approach, beginning with thrice weekly visits in the initial week and then weekly assessments to the eighth week. The assessments' purpose ranged from symptom monitoring, and feedback collection to postsurgical outcomes and patient experience, with the data collection formats and tools varying accordingly. This included data collected through specific clinical symptom questionnaires designed by local clinical teams or disease-specific assessments, pain scores, patient experience Likert measures, distress thermometers, and disease-specific patient-reported outcome measures in various combinations.

Technology

Interventions leveraged digital platforms, either on computers or smartphone devices. Several studies used digital, web-based questionnaire portals [11,19], with other studies [18,22] using smartphone-native apps, some of which had additional functionalities like real-time alerts for clinical teams based on patient input. Pienaar et al [21] use a "2-way-texting" platform that requires only native SMS text responses from patients. Functional variation of the technology used is described in Figure 2, mapped against the IMS Institute for Healthcare Informatics mHealth functionality criteria [10]. Only 1 study [18] incorporated a dual approach using both SMS and email to communicate with patients. The majority of interventions [11-19,22] used functionalities to trigger specific clinical team alerts or interventions when customized thresholds in symptom reports were reached. Only the functions of recording, collecting, and sharing data were consistent across all interventions reviewed.

Figure 2. Study intervention assessed by digital health function criteria [11-22].

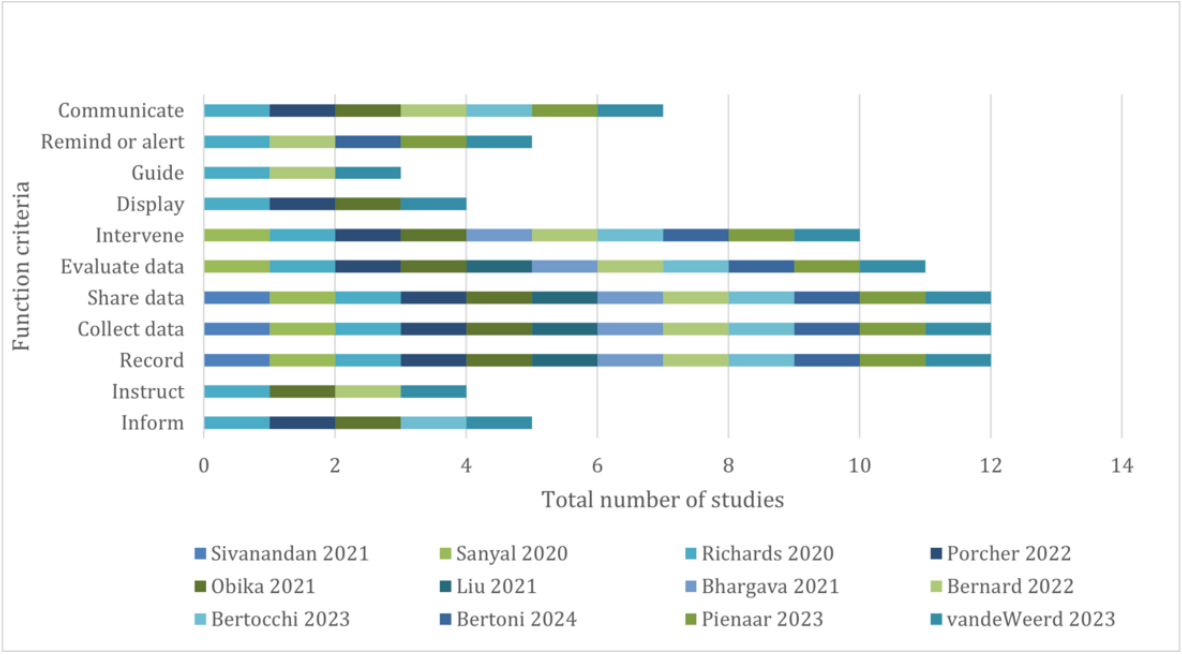

Evaluation of Outcomes Mapped to the Quadruple Aim

With an extensive range of variation across the types of interventions reviewed, there was not a single or consistent evaluation method to measure the outcomes and benefits of the intervention. The themes of the included studies are summarized in Multimedia Appendix 4 [11-22].

Population Health Outcomes

Overall, the use of remote assessment tools to directly manage patient health outcomes indicated positive results, particularly in the management of clinical deterioration, timely clinical intervention, and clinical protocol adherence. Accurate identification of clinical deterioration was achieved in most studies [12-14,19,22], whereby remote clinical review was initiated after interpretation or alert of the data, and patient treatment schedules or pathways were adjusted to prevent harm or further deterioration of health. This included 1 study that reported no patients on the intervention attending an emergency department for symptom management during the study period [19] and a reduction of the 30-day readmission rate for patients with high protocol adherence in a separate study [15]. In another study, quality of life outcomes (quality-adjusted life years) improved by 0.07, on a scale of 0 to 1, for those patients who underwent remote digital assessment as compared to traditional in-person monitoring and consultations [13]. The number of potential adverse events was also tracked over the intervention period in 1 study, where potential adverse events were low on day 1 post surgery and peaking on day 3 of 13-day monitoring [21].

One study explored the consistency of patients' data provided over a telephone interview with a clinician versus patient self-reporting via instant messaging. The results showed up to 10% variance on some symptom assessment data points, which may impact accurate management of health outcomes [20].

Health System Costs

Cost of Care or Service

Two studies [13,19] attempted to understand service costs using cost utility analysis, assessing the overall cost saving of providing care through remote assessment interventions

XSL•FO
RenderX

compared to usual care. While both studies reported a cost saving delivered by the intervention compared to routine care, there was commentary on an inflated cost per patient due to the small sample sizes undertaken, n=13 [19] and n=89 [13], and the baseline costs of establishing the digital technologies for such interventions. The majority of patients in 1 (60%) study indicated that the digital intervention saved them time and money from attending in-person clinic visits [21].

Workforce Use

The impact on the use of clinical resources through the substitution of remote assessment was reported across several studies [12-14,19,20], including the assumed suggestion of potential cost savings for the health service through triaged use of multidisciplinary teams, asynchronous communication, and reduction of routine consultation methods. The ability to effectively use different roles where appropriate was highly dependent on predefined patient pathways and parameters that could be automated through the digital tools and workflows used [12-14,19,22].

Reduction in Avoidable or Unnecessary Hospital Admissions

There was no clear reporting on the reduction in avoidable or unnecessary hospital admissions. The number of hospital readmissions for the intervention group of postsurgical patients was reported, however, no comparator to readmissions for a usual care approach was provided to establish if this was an improvement or otherwise [12]. One study reported a correlation of reduction in readmission rate for patients with high adherence to the digital protocol (P=.004) [15]. A reduction in hospitalization rate over a 6-month period was reported in patients with chronic disease when remotely monitored (rate of hospitalization=1.3) compared to the conventional monitoring group (rate of hospitalization=2.3) [13].

Patient Experience

Patient-experience measures were commonly captured through the use and adherence of the protocol, as well as qualitative surveying of the patient's service experience using Likert scoring.

Overall, there was positive patient acceptance of services using digital remote assessments [13-16,19,21,22], with adherence levels to the assessment reporting ranging between 62%-93% across the full study periods of the relevant interventions.

One study noted that patients with high function and low symptom burden were more likely to withdraw, implying the tool was more beneficial for patients experiencing significant symptoms [19]. It was further noted that there seemed to be other influencing factors to patient use such as physical health, in being able to complete self-reporting assessments [19].

Further thematic analysis of 1 intervention [12] highlighted that most patients valued the advice provided by the technology-enabled system. Key themes included reassurance and the relevance of advice which felt tailored to their specific needs. Patient perception of the usefulness of the interventions varied, with 1 study [17] reporting that 63% of patients thought the intervention was useful to support the management of health and access to services (score of 4.42/7), while another scored above 92% for patients feeling comfortable and safe using the digital service [21].

Clinician Experience

The 2 study interventions where clinical experience measures were reported both strongly indicate positive outcomes [12,19]. Measures indicated a consensus among clinical staff that using digital assessments improved clinical confidence in the delivery of care. Commentary in these studies indicated clinicians' belief that no clinical symptoms were overlooked, and the system was perceived to provide additional safety in highlighting areas of patient health that might not otherwise be reported or escalated by the patient [12,19].

One study highlighted that the clinicians found the ability to access the reported data through the hospitals. Electronic medical record systems useful, however, were contingent upon their accessibility to a computer within the hospital [12].

Additionally, 75% of clinicians from 1 study [19] perceived that the quality of life for their patients had improved owing to the remote assessment system in place.

Discussion

Summary of Evidence

This scoping review seeks to explore how remote assessment tools are used, practically and effectively, for routine consultation substitution in tertiary care centers. The review identifies the risks and benefits of these tools to guide health care systems in adopting technology-assisted models of care. Specifically, it investigates studies using remote assessment tools as the primary method for routine care substitution, the types and durations of these interventions, and how published outcomes align with the Quadruple Aim framework.

The review identified 12 studies, indicating that while digital remote assessment tools may be broadly implemented across health care facilities, limited research has been published on their specific applications for routine care substitution. Despite this small body of literature, these tools appear to be effective in capturing and assessing clinical deterioration, enabling proactive management and personalized treatment pathways. They may influence care pathway stratification and multidisciplinary team use and have the potential to improve health service efficiency.

The geographic distribution of studies suggests a concentrated focus in regions such as the United Kingdom and Europe, where conditions appear favorable for this type of research. However, many studies excluded participants due to language barriers, raising concerns about equity and inclusivity. While digital translation tools are widely available in other industries, health care settings face unique challenges in integrating such solutions effectively [23].

Notably, all studies were published from 2020 onward, highlighting the role of the COVID-19 pandemic as a catalyst for the adoption and evaluation of digital remote assessment tools. This trend underscores the growing interest in leveraging digital health technologies to address emerging pressures on

XSL•FO
RenderX

health care systems. However, the limited number of studies underscores the need for further research to establish robust evidence for the risks, benefits, and broader applicability of these interventions.

Principal Findings

When analyzed using the Quadruple Aim framework, the findings demonstrate an emphasis on measuring population health outcomes and patient experience. However, the neglect of health system costs and clinician experience in many studies reveals a significant gap in literature. To create a sustainable health care system capable of meeting both current and future challenges, balanced attention to all 4 dimensions of the Quadruple Aim is essential [8].

Only 2 studies [13,19] examined health system costs, underscoring the limited focus on this critical area. This neglect may stem from the challenges associated with costing health care delivery comprehensively, particularly in complex, innovative care models like digital health interventions. Traditional costing methods often fail to capture the variability and complexity of care pathways, making it difficult for policymakers to evaluate economic implications accurately [24]. Without robust cost analyses, the ability to optimize resource allocation, reduce unnecessary expenditures, and understand long-term sustainability remains limited [25].

Similarly, only 2 studies [12,19] addressed clinician experience, a surprising finding given the pivotal role clinicians play in the success of digital health technologies. While the focus on patient-centered care is appropriate, neglecting clinician perspectives risks compromising the balance necessary to achieve the Quadruple Aim [26]. Clinicians must be confident that digital remote assessment tools enhance, rather than burden, their workflows and align with their commitment to patient care. Failure to address these concerns could lead to implementation challenges or resistance, undermining the benefits of digital interventions.

Limitations

The study has several limitations. The narrow selection criteria focused exclusively on interventions using remote assessment data as the primary data source, excluding studies that integrated additional data inputs such as biomedical devices, photographic data, or web-based education. As a result, relevant studies incorporating more comprehensive data sources were not included.

Additionally, a substantial number of studies (n=27) were excluded due to insufficient information to answer the research questions, many of which were conference abstracts. The lack of randomized controlled trials or comparator studies further limits the robustness of the findings, potentially introducing bias and inflating the perceived benefits of these interventions. This reflects a broader challenge in digital health research, where the rapid pace of technological development renders traditional randomized controlled trials less feasible. Innovative methods, such as observational studies, may provide more pragmatic and timely approaches for evaluating digital health interventions [27].

Future Research

Future research should prioritize direct assessments of health system costs associated with remote assessment interventions, incorporating time-driven activity-based costing to provide more accurate and comprehensive insights [25]. Understanding funding policies and potential revenue streams from public and private health funders will also be essential to support scalability and sustainability [27-29].

Additionally, research should explore clinician experiences with reduced patient contact due to digital interventions and identify systemic strategies to mitigate burnout and promote health care workers' well-being. This area remains underexplored despite its importance for successful implementation.

Broader Implications

The integration of digital remote assessment tools into health systems holds substantial promise for improving health care delivery. However, their success hinges on addressing critical challenges, including equity, cost analysis, and clinician engagement. Tailoring interventions to specific clinical contexts, integrating them seamlessly with existing workflows, and incorporating patient and clinician feedback are essential steps toward maximizing their impact.

To ensure long-term sustainability, future research and implementation efforts must adopt a holistic approach that balances the dimensions of the Quadruple Aim. This includes developing robust cost analyses, fostering clinician buy-in, and addressing barriers to accessibility. Ultimately, by addressing these factors, digital health interventions can contribute to a more equitable, efficient, and patient-centered health care system capable of withstanding evolving challenges.

Conclusions

This review highlights the broad-reaching potential and adaptability of digital remote assessment interventions. Despite variability in duration and methodologies, these technologies consistently demonstrate accurate data recording, collection, and dissemination. More mature interventions that guide clinical responses offer additional benefits.

Digital health solutions can advance the Quadruple Aim of Healthcare by enhancing population health outcomes, optimizing resource allocation, and improving patient and clinician experiences. Future research should establish standardized measures of health system effectiveness, consider clinician adoption, and investigate the long-term sustainability and impact on access to care.

While digital remote assessment tools offer significant potential, careful consideration of implementation strategies, equity, cost, and clinician and patient experiences is crucial for successful clinical integration.

XSL•FO
RenderX

Acknowledgments

JB received a higher education scholarship fund through the University of Queensland to support their research.

Data Availability

The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist was included in the manuscript submission and can be consulted in Multimedia Appendix 1. The datasets generated and analyzed during this study are available from the corresponding author (JB) on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for scoping review study design.
[PDF File (Adobe PDF File), 101 KB - jmir_v27i1e65938_app1.pdf]

Multimedia Appendix 2
PubMed search strategy used to undertake scoping review.
[DOCX File , 14 KB - jmir_v27i1e65938_app2.docx]

Multimedia Appendix 3
Summary of study characteristics.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v27i1e65938_app3.xlsx]

Multimedia Appendix 4
Summary of study themes mapped across the Quadruple Aims of Healthcare.
[XLSX File (Microsoft Excel File), 30 KB - jmir_v27i1e65938_app4.xlsx]

References

1. Clark HE, Colombo F, Candeia V. Global Future Council on Health and Healthcare 2018-2019: A vision for the future: transforming health systems. World Economic Forum. 2019. URL: https://www3.weforum.org/docs/WEF_GFC_reflection_paper.pdf [accessed 2025-02-06]
2. Queenland Audit Office. Improving access to specialist outpatient services: Report 8; 2021-22. 2021. URL: https://documents.parliament.qld.gov.au/tp/2021/5721T2069.pdf [accessed 2021-12-06]
3. Williams GJ, Al-Baraikan AB, Rademakers FE, Ciravegna F, van de Vosse FN, Lawrie A, et al. Wearable technology and the cardiovascular system: the future of patient assessment. Lancet Digit Health 2023;5(7):e467-e476 [FREE Full text] [doi: 10.1016/S2589-7500(23)00087-0] [Medline: 37391266]
4. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
5. Manteghinejad A, Javanmard SH. Challenges and opportunities of digital health in a post-COVID19 world. J Res Med Sci 2021;26(11):11 [FREE Full text] [doi: 10.4103/jrms.jrms_1255_20]
6. Belarmino A, Walsh R, Alshak M, Patel N, Wu R, Hu JC. Feasibility of a mobile health application to monitor recovery and patient-reported outcomes after robot-assisted radical prostatectomy. Eur Urol Oncol 2019;2(4):425-428 [FREE Full text] [doi: 10.1016/j.euo.2018.08.016] [Medline: 31277778]
7. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A mobile health application to track patients after gastrointestinal surgery: results from a pilot study. J Gastrointest Surg 2017;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]
8. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
9. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
10. Aitken M, Gauntlett C. Patient apps for improved healthcare: From novelty to mainstream. IMS Institute for Healthcare Informatics. 2013. URL: https://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2025-02-08]
11. Sivanandan MA, Sharma C, Bullard P, Christian J. Digital patient-reported outcome measures for monitoring of patients on cancer treatment: cross-sectional questionnaire study. JMIR Form Res 2021;5(8):e18502 [FREE Full text] [doi: 10.2196/18502] [Medline: 34398785]

XSL•FO
RenderX

12. Richards HS, Blazeby JM, Portal A, Harding R, Reed T, Lander T, et al. A real-time electronic symptom monitoring system for patients after discharge following surgery: a pilot study in cancer-related surgery. BMC Cancer 2020;20(1):543 [FREE Full text] [doi: 10.1186/s12885-020-07027-5] [Medline: 32522163]
13. Bernard L, Valsecchi V, Mura T, Aouinti S, Padern G, Ferreira R, et al. Management of patients with rheumatoid arthritis by telemedicine: connected monitoring. a randomized controlled trial. Joint Bone Spine 2022;89(5):105368 [FREE Full text] [doi: 10.1016/j.jbspin.2022.105368] [Medline: 35248737]
14. Porcher L, Perron V, Blanc J, Kaderbhai CG, Tharin Z, Schmitt A, et al. Smartphone-based application and nurses' interventions for symptoms monitoring in patients treated with oral anticancer agents: a 1-year follow-up in a tertiary cancer center. J Oncol Pharm Pract 2022:10781552221117731 [FREE Full text] [doi: 10.1177/10781552221117731] [Medline: 35938191]
15. Bertocchi E, Barugola G, Masini G, Guerriero M, Menestrina N, Gentile I, et al. iColon, a patient-focused mobile application for perioperative care in colorectal surgery: results from 444 patients. J Telemed Telecare 2023:1357633X231203064 [FREE Full text] [doi: 10.1177/1357633X231203064] [Medline: 37820368]
16. Bertoni S, M M K Magema JP, Van Geluwe B, Abbes Orabi N, Bislenghi G, D'Hoore A, et al. Digital postoperative follow-up after colorectal resection: a multi-center preliminary qualitative study on a patient reporting and monitoring application. Updates Surg 2024;76(1):139-146. [doi: 10.1007/s13304-023-01671-9] [Medline: 37943493]
17. van de Weerd C, Ebbers T, Smilde DEM, van Tol-Geerdink JJ, Takes RP, van den Broek GB, et al. Evaluation of a remote monitoring app in head and neck cancer follow-up care. Cancer Med 2023;12(14):15552-15566. [doi: 10.1002/cam4.6202] [Medline: 37293944]
18. Sanyal A, Heun JM, Sweeney, JD, J, Janssen C. Mobile-health tool to improve care of patients with hematological malignancies. Blood 2020;136:35-36 [FREE Full text] [doi: 10.1182/blood-2020-143405]
19. Bhargava R, Keating B, Isenberg SR, Subramaniam S, Wegier P, Chasen M. RELIEF: a digital health tool for the remote self-reporting of symptoms in patients with cancer to address palliative care needs and minimize emergency department visits. Curr Oncol 2021;28(6):4273-4280 [FREE Full text] [doi: 10.3390/curroncol28060363] [Medline: 34898539]
20. Liu F, Guo P, Su X, Cui M, Jiang J, Wang S, et al. A novel remote follow-Up tool based on an instant messaging/social media app for the management of patients with low anterior resection syndrome: pilot prospective self-control study. JMIR Mhealth Uhealth 2021;9(3):e22647 [FREE Full text] [doi: 10.2196/22647] [Medline: 33739295]
21. Pienaar J, Day S, Setswe G, Wasunna B, Ncube V, Ndebele F, et al. 'I understood the texting process well'. Participant perspectives on usability and acceptability of SMS-based telehealth follow-up after voluntary medical male circumcision in south africa. Digit Health 2023;9:20552076231194924 [FREE Full text] [doi: 10.1177/20552076231194924] [Medline: 37654716]
22. Obika BD, Dolezova N, Ponzo S, Valentine S, Shah S, Gledhill J, et al. Implementation of a mHealth solution to remotely monitor patients on a cardiac surgical waiting list: service evaluation. JAMIA Open 2021;4(3):ooab053 [FREE Full text] [doi: 10.1093/jamiaopen/ooab053] [Medline: 34355134]
23. Panayiotou A, Gardner A, Williams S, Zucchi E, Mascitti-Meuter M, Goh AM, et al. Language translation apps in health care settings: expert opinion. JMIR Mhealth Uhealth 2019;7(4):e11316 [FREE Full text] [doi: 10.2196/11316] [Medline: 30964446]
24. Guinness LGS, Ghosh S, Mehndiratta A, Shah HA. Role of healthcare cost accounting in pricing and reimbursement in low-income and middle-income countries: a scoping review. BMJ Open 2022;12(9):e065019 [FREE Full text] [doi: 10.1136/bmjopen-2022-065019] [Medline: 36171039]
25. Kaplan RS, Porter ME. How to solve the cost crisis in health care. Harv Bus Rev 2011;89(9):46-64. [Medline: 21939127]
26. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
27. Hayes CJ, Dawson L, McCoy H, Hernandez M, Andersen J, Ali MM, et al. Utilization of remote patient monitoring within the United States health care system: a scoping review. Telemed J E Health 2023;29(3):384-394 [FREE Full text] [doi: 10.1089/tmj.2022.0111] [Medline: 35819861]
28. Dvir R, Goldsmith C, Seavey I, Vedlitz A, Hammett J, Bonet S, et al. The policy environment of remote patient monitoring: evaluating stakeholders' views. Healthc Technol Manag 2023:249-275.
29. Rakers MM, van Os HJ, Recourt K, Mosis G, Chavannes NH, Struijs JN. Perceived barriers and facilitators of structural reimbursement for remote patient monitoring, an exploratory qualitative study. Health Policy Technol 2023;12(1):100718 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100718]

Abbreviations

IMS: Incidental Medical Services
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

XSL•FO
RenderX

Edited by A Mavragani; submitted 30.08.24; peer-reviewed by LMV Schougaard, M Wilkes; comments to author 29.11.24; revised version received 30.12.24; accepted 28.01.25; published 04.03.25.

Please cite as:
Barclay J, Sullivan C, Beckmann M, Mattison G, Runciman R, Martin E
Use of Remote Assessment Tools to Substitute Routine Outpatient Care: Scoping Review
J Med Internet Res 2025;27:e65938
URL: https://www.jmir.org/2025/1/e65938
doi:10.2196/65938
PMID:40053751

©Jayne Barclay, Clair Sullivan, Michael Beckmann, Graeme Mattison, Rebecca Runciman, Elizabeth Martin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Telehealth Versus In-Person Informed Consent: Randomized Study of Comprehension and Decision-Making

Saif Khairat[1,2,3,4], MPH, PhD; Paige Ottmar[1], MHA, MSIS; Prabal Chourasia[1], MD, PSM; Jihad Obeid[5], MD

[1]Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]Cecil G Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[4]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, United States

[5]Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Saif Khairat, MPH, PhD
Carolina Health Informatics Program
University of North Carolina at Chapel Hill
Carrington Hall
Chapel Hill, NC
United States
Phone: 1 9198435413
Email: saif@unc.edu

Abstract

Background: Obtaining informed consent (IC) is vital for ethically and effectively recruiting participants in research projects. However, traditional in-person IC approaches encounter notable obstacles, such as geographic barriers, transportation expenses, and literacy challenges, which can lead to delays in enrollment and increased costs. Telehealth, especially teleconsent, offers a potential way to overcome these obstacles by facilitating the IC process in a digital setting. Nonetheless, there are concerns about whether teleconsent can achieve levels of understanding and involvement that are equivalent to those of in-person IC meetings.

Objective: This study aims to evaluate comprehension and decision-making in participants undergoing teleconsent versus traditional in-person IC. We used validated assessments to determine whether teleconsent is a viable alternative that maintains participants' understanding and decision-making abilities.

Methods: A randomized comparative study design was used, recruiting potential participants for a parent study assessing patient experiences with patient portals. Participants were randomly assigned to 2 groups: teleconsent and in-person consent. The teleconsent group used Doxy.me software, allowing real-time interaction between researchers and participants while reviewing and electronically signing the IC documents. Recruitment involved using an institutional web-based platform to identify interested individuals, who were then contacted to assess eligibility and gather demographic information. The Decision-Making Control Instrument (DMCI) survey was used to assess the perceived voluntariness, trust, and decision self-efficacy. The Quality of Informed Consent (QuIC) was used to measure the comprehension level of the consent form. The validated Short Assessment of Health Literacy-English tool was used to measure participants' health literacy levels.

Results: A total of 64 participants were enrolled in the study, with 32 in the teleconsent group and 32 in the in-person group. Of 64 participants, 32 (50%) were in the teleconsent group, 54 (84.4%) were females, 44 (68.7%) were aged 18-34 years, 50 (78.1%) were White, and 31 (48.4%) had a bachelor degree. The mean SAHL-E scores were different between the teleconsent and in-person groups (16.72, SD 1.88 vs 17.38, SD 0.95; *P*=.03). No significant differences were found between the average scores at baseline and follow-up for QuIC part A (*P*=.29), QuIC part B (*P*=.25), and DMCI (*P*=.38) within the teleconsent and in-person groups. Additionally, there were no significant differences in QuIC or DMCI between subgroups based on age, sex, and ethnicity.

Conclusions: This study assessed the effectiveness of IC processes through telehealth compared to traditional in-person visits. Findings indicate that telehealth offers similar participant understanding and engagement while overcoming geographic and accessibility barriers. As health care adopts digital solutions, these results highlight telehealth's potential to improve recruitment and retention in clinical research, suggesting that policy makers should integrate telehealth practices into regulations for better access and health outcomes.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63473) doi:10.2196/63473

KEYWORDS

telehealth; informed; consent; comprehension; decision-making; cross-sectional study; cross-section; telemedicine; eHealth; health care services; mHealth; effectiveness; informed consent; statistical analysis; feasibility

Introduction

Obtaining informed consent (IC) supports the successful recruitment of participants and the overall success of research studies [1]. However, obtaining IC faces challenges related to the cost and time to travel to and from the study site, geographic barriers in rural or remote areas, and literacy barriers [2]. Lack of obtaining IC leads to delayed enrollment, low retention, and engagement rates, and increased costs for additional resources [3].

Traditional methods of obtaining IC usually involve face-to-face interactions in which health care providers explain procedures, risks, benefits, and alternatives to patients, often using written consent forms [4,5]. These methods facilitate direct communication and allow patients to ask questions, which helps build understanding and trust [5,6]. However, several challenges are associated with this process. Patients often have varying levels of health literacy, and time constraints can prevent thorough discussions [7,8]. Additionally, the use of complex medical jargon can confuse patients [9,10].

Research has shown that many patients are unaware of their rights and may not receive enough information to make fully informed decisions [11,12]. Challenges in decision-making and comprehension of in-person IC are multifaceted. Patients often have varying levels of health literacy, which can significantly impact their understanding of medical information and their ability to make informed decisions [13]. Time constraints during consultations can further hinder thorough discussions, leaving patients with insufficient information [7]. Additionally, the use of complex medical jargon can confuse patients, making it difficult for them to grasp the details of their treatment options [14]. Research has shown that many patients are unaware of their rights and may not receive enough information to make fully informed decisions [11]. Furthermore, the responsibility for obtaining consent can sometimes be unclear, leading to inconsistencies in the process [15,16]. To address these challenges, it is essential to improve communication strategies and ensure that consent forms are clear and comprehensive.

The use of telehealth, namely teleconsent, has the potential to address these challenges by conducting the IC process digitally [17]. Teleconsent involves live video calls during which the researcher and participant can see and interact with each other in real-time [18,19]. There are concerns about the ability of teleconsent to deliver similar levels of comprehension and engagement to in-person IC visits. This noninferiority study aimed to examine the comprehension and decision-making levels for teleconsent versus in-person IC visits using validated comprehension assessment survey instruments in conjunction with health literacy evaluation.

Methods

Study Design

In this randomized comparative study, we obtained the IC of potential participants for a parent study that assesses patients' experience using patient portals. Participants were recruited through an institutional web-based recruitment platform or word of mouth. Participants were randomly assigned to in-person or teleconsent. The teleconsent group used Doxy.me software, which enables the researcher to share the IC document on-screen, collaboratively complete it with the participant and sign it electronically. Individuals in the in-person group met with the study research assistant in a private office.

Recruitment Process

We listed the study on an institutional recruitment platform such that when an individual expressed interest in the study, the research assistant received an automated notification email containing the potential participant's contact information.

The research assistant had initiated contact via email acknowledging the individual's interest and requesting demographic and eligibility information to ensure participant diversity. The research assistant provided a link to a Qualtrics survey and asked qualified individuals to digitally sign a consent form. The research assistant emphasized confidentiality and mentioned an upcoming interview, which could be in person or via teleconsent. Participants would also receive a gift card for their involvement in both the initial interview and a follow-up session.

Participants designated for teleconsent received an email from the research assistant confirming their eligibility for remote participation via computer (microphone and camera required; smartphones and tablets not allowed). They were asked to complete a scheduling form for a 45-minute session, which included the consent process, a semistructured interview on MyChart usability, and follow-up questionnaires. The research assistant also mentioned a gift card incentive for both initial and follow-up sessions.

In cases where participants were scheduled for an in-person session, the research assistant issued a similarly structured email. The message confirmed their eligibility and presented them with the option to choose between 2 locations: a public library or a private office at the University. Participants were encouraged to fill out a scheduling form with their preferred time slots. The research assistant outlined the session's duration and activities, mirroring the information provided for teleconsent, along with the mention of the gift card reward for study completion.

XSL•FO
RenderX

Consent Process

The consent form was 6 pages long, and it outlined the purpose, procedures, risks, and participant rights related to the study. The consent form was designed to inform potential participants about the study, ensuring they can make an informed decision about their involvement. It invited them to ask questions and clarified their rights as research subjects.

This consent form explained the nature of the study to evaluate participants' perceptions of the MyChart patient portal's usability and differences in consent understanding between teleconsent and traditional methods. The participants were adults who were users of MyChart, a prominent web-based patient portal. Participation in this study was voluntary. The form explained the chronological order of events in this study and that it will begin with eligibility questions and the collection of demographic information. Participants will then be randomly assigned to either teleconsent or in-person consent. Following this, they will take part in an interview with a research assistant regarding their experiences with MyChart, which will be recorded for transcription purposes. Additionally, participants will complete questionnaires on consent comprehension and health literacy.

The form also stated that the study will last no more than 1 hour and will be conducted at a designated location either in person or via videoconference. The research assistant also read the section about the potential risks, including confidentiality breaches, which will be minimized by assigning codes to data and ensuring anonymity during analysis. Finally, the form covers that the study is protected by a Certificate of Confidentiality, which prevents the use of identifying information in legal scenarios without consent, with some exceptions such as mandatory reporting laws. To verify the identity of the individuals in the teleconsent group, we asked that they enable their cameras during the entirety of the session. In addition, when signing the consent form, we used a feature in Doxy.me that enables taking a screenshot along with a timestamp to accompany the individual's wet signature obtained live during the session. This feature ensures that the individual who signed the consent form is the intended individual, a necessary step in teleconsent [20].

After the session, participants completed the Decision-Making Control Instrument (DMCI) and Quality of Informed Consent (QuIC) surveys after the consent session (ie, baseline) and 30 days after the session (ie, follow-up). Data collection occurred between April 30, 2019, and June 29, 2021. The institution review board approval was obtained.

We used the DMCI, a 15-item validated instrument, to assess participants' perceived voluntariness, trust, and decision self-efficacy in the consent form [21]. The DMCI assesses participants' feelings of influence over decisions and their confidence in the fairness and transparency of the process. The maximum possible score is 30 and higher scores indicate greater perceived autonomy and trust in decision-making.

We used the QuIC to measure comprehension of the parent study consent form [22]. The QuIC consists of 2 parts: part A (14 items) assesses objective knowledge of study details, testing participants' factual understanding of the consent process; and part B (6 items) measures perceived understanding, evaluating how confident participants feel about comprehending the consent materials. Together, the QuIC thoroughly evaluates both actual and perceived IC quality. The maximum possible score is 80 and higher scores reflect better comprehension and confidence, ensuring that consent is truly informed.

We used the validated Short Assessment of Health Literacy-English (SAHL-E) tool to measure participants' health literacy levels. The SAHL-E is an 18-item tool that assesses functional health literacy by testing participants' ability to understand and apply health-related terms. It uses a word association format to evaluate vocabulary and conceptual knowledge. The maximum possible score is 18, and higher scores indicate better health literacy, which correlates with improved capacity to make informed health decisions (Multimedia Appendix 1).

Ethical Considerations

The University of North Carolina at Chapel Hill institutional review board deemed this study exempt (17-2870). Participants consented to participate in the study and to be assigned to either the control or intervention groups. Participant responses were kept anonymous, and participants were only referred to by subject ID. All data were kept separate from names and housed behind a secure, Health Insurance Portability and Accountability Act (HIPAA)–compliant server. Participants received a US $15 gift card upon completing the baseline surveys and another US $15 gift card upon completing the follow-up surveys.

Data Analysis

We compared the demographic distribution of participants regarding sex, age, race, education, and income between teleconsent and in-person groups. We used a 2-sample, 2-tailed *t* test assuming unequal variances to compare the SAHL-E scores between the teleconsent and in-person groups to compare literacy levels. We compared the scores for the DMCI and QuIC surveys for both the teleconsent and in-person groups between baseline and follow-up using 2-sample *t* tests. We used a 2-sample, 2-tailed *t* test assuming unequal variances to compare the DMCI and QuIC score for subgroups on sex (male and female), age (18-34 years, 35-50 years, 51-64 years, and 65 years and older), and race (Asian, Black or African American, White, and mixed race).

We aimed to enroll 64 participants to provide 80% power to detect modest differences (effect sizes approximately 0.50) between groups with respect to the survey instrument scores. For example, we expect to be able to detect a 3.5-unit difference on the DMCI composite score (which has been reported to have a mean of 46.1, SD 7.0) and a 4.7-unit difference on the QuIC (which has been reported to have a mean of 77.8, SD 9.4).

Results

Overview

Of 64 participants, 32 (50%) were in the teleconsent group, 54 (84%) were females, 44 (68%) were aged 18-34 years, 50 (78%) were White, and 31 (48%) had a bachelor degree. The mean

XSL•FO
RenderX

SAHL-E scores were different between the teleconsent and in-person groups (16.72, SD 1.88 vs 17.38, SD 0.95; *P*=.03; Table 1).

At baseline, the teleconsent group had a higher mean QuIC score than the in-person group for part A (7.38, SD 1.27 vs 6.84, SD 2.46; *P*=.29) and part B (53.88, SD 1.80 vs 53.19, SD 2.79; *P*=.25). Both groups had similar mean DMCI scores (19.44, SD 1.12 vs 19.84, SD 2.31; *P*=.38; Figure 1).

Table 1. Demographic characteristics and health literacy scores of participants in teleconsent and in-person consent groups.

Category	All	Teleconsent (n=32)	In-person (n=32)	*P* value
Sex, n (%)				.49
Male	10 (15)	4 (12)	6 (18)	
Female	54 (84)	28 (87)	26 (81)	
Age (years), n (%)				.22
18-34	44 (68)	23 (71)	21 (65)	
35-50	15 (23)	7 (21)	8 (25)	
51-64	4 (6)	2 (6)	2 (6)	
65 and older	1 (1)	0 (0)	1 (3)	
Race, n (%)				.40
American Indian	2 (3.1)	1 (3)	1 (3)	
Asian	6 (9)	3 (9)	3 (9)	
Black or African American	5 (7)	2 (6)	3 (9)	
White	50 (78)	25 (78)	25 (78)	
White Asian	1 (1)	1 (3)	0 (0)	
Education, n (%)				.08
Bachelor degree in college (4 years)	31 (48)	14 (43)	17 (53)	
Master degree	13 (20)	6 (18)	7 (21)	
Associate degree	5 (7)	3 (9)	2 (6)	
Professional or doctoral degree	4 (6)	0 (0)	4 (12)	
Some college but no degree	10 (15)	8 (25)	2 (6)	
High school graduate	1 (1)	1 (3)	0 (0)	
Income (US $), n (%)				<.001
Less than 10,000	4 (6)	4 (12)	0 (0)	
10,000-49,999	23 (35)	17 (53)	6 (18)	
50,000-99,000	21 (32)	8 (25)	13 (40)	
100,000-149,999	12 (18)	3 (9)	9 (28)	
150,000 and more	4 (6)	0 (0)	4 (12)	
SAHL-E[a]	17.05	16.71	17.37	.03

[a]SAHL-E: Short Assessment of Health Literacy-English.

XSL•FO
RenderX

Figure 1. Average scores for (A,B) QuIC: (A) part A (maximum possible score=24) and (B) Part B (maximum possible score=56); and (C) DMCI (maximum possible score=30) at baseline and follow-up for the teleconsent and in-person groups. DMCI: Decision-Making Control Instrument; QuIC: Quality of Informed Consent.

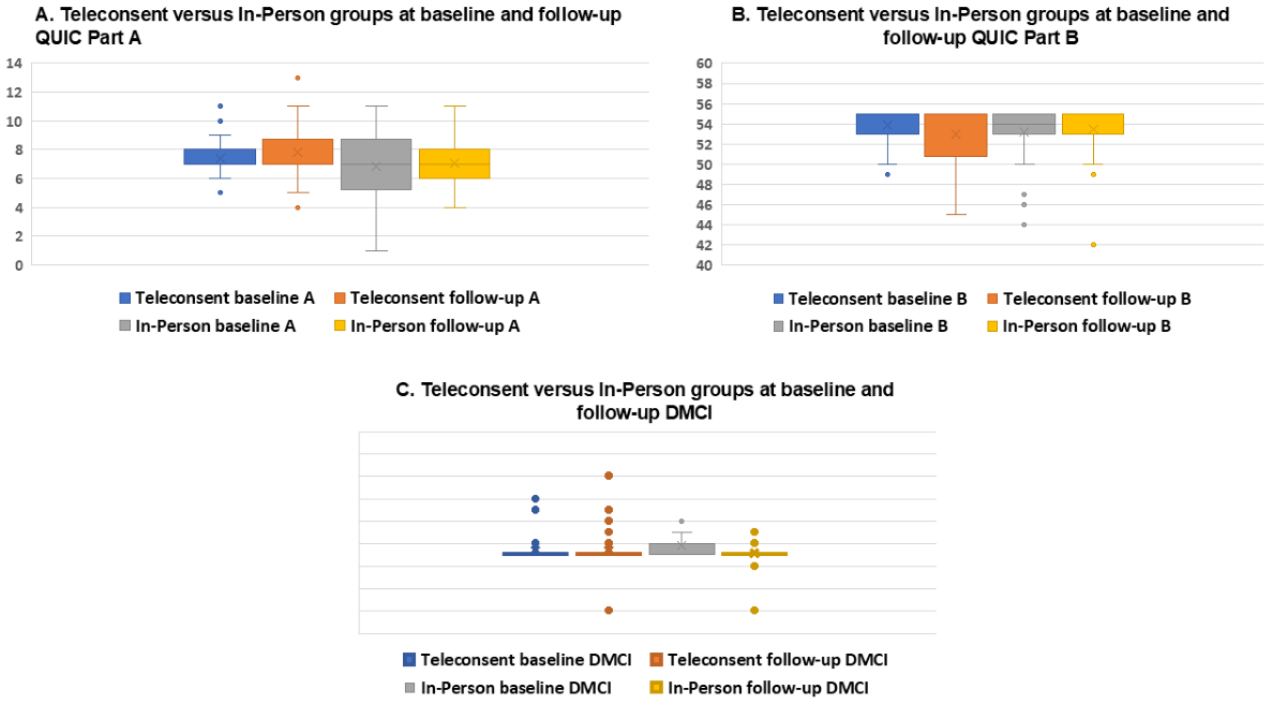

At follow-up, the teleconsent group had a higher mean score than the in-person for the QuIC part A (7.84, SD 1.99 vs 7.03, SD 1.61; *P*=.08) and the DMCI (19.41, SD 1.77 vs 19.09, SD 1.13; *P*=.41). In-person had a higher mean score for the QuIC part B (52.94, SD 3.03 vs 53.50, SD 2.68; *P*=.44).

No significant differences were found between the average scores at baseline and follow-up for QuIC part A, QuIC part B, and DMCI within the teleconsent and in-person groups (Multimedia Appendix 2).

Within the teleconsent group, scores for QuIC part A increased from a baseline average of 7.38 (SD 1.27) to 7.84 (SD 1.99), while QuIC part B showed a slight decrease from 53.88 (SD 1.80) to 52.94 (SD 3.03). The mean DMCI scores remained relatively stable, with a minor decrease from 19.44 (SD 1.12) to 19.41 (SD 1.77). However, none of these changes were statistically significant.

Within the in-person group, the mean QuIC part A scores increased slightly from 6.84 (SD 2.46) to 7.03 (SD 1.61), and QuIC part B scores rose marginally from 53.19 (SD 2.79) to 53.50 (SD 2.68). The mean DMCI scores saw a notable drop from 19.84 (SD 2.31) to 19.09 (SD 1.13). Similar to the teleconsent group, none of these differences reached statistical significance.

DMIC Subgroup Analysis

The results of the teleconsent and in-person consent assessments, measured by the DMCI scores, showed no significant differences, Table 2. However, males exhibited higher baseline scores in the teleconsent group (mean 20.25, SD 1.64) compared to females (mean 19.32, SD 0.97), although both groups showed similar follow-up scores. In-person assessments also indicated male scores were slightly higher at baseline (mean 19.50, SD 0.5) compared to females (mean 19.92, SD 2.54), but follow-up scores highlighted a more significant disparity between the groups.

XSL•FO
RenderX

Table 2. Comparison of DMCI[a] scores by sex, age, and race for teleconsent and in-person consent groups at baseline and follow-up (maximum possible sore=30).

	Teleconsent			In-person		
	Baseline DMCI	Follow-up DMCI	*P* value	Baseline DMCI	Follow-up DMCI	*P* value
Sex, mean (SD)						
Male	20.25 (1.64)	20.50 (1.12)	.83	19.50 (0.5)	19.33 (0.94)	.73
Female	19.32 (0.97)	19.25 (1.79)	.85	19.92 (2.54)	19.04 (1.16)	.12
Age (years), mean (SD)						
18-34	19.52 (1.28)	19.17 (0.64)	.26	20.28 (2.75)	19.09 (1.38)	.09
35-50	19.14 (0.35)	20.14 (3.48)	.51	19 (0)	19.12 (0.33)	.35
51-64	19.5 (0.5)	19.5 (0.5)	≥.99	19 (0)	19 (0)	N/A[b]
65 and older	0 (0)	0 (0)	N/A	19 (0)	19 (0)	N/A
Race, mean (SD)						
Asian	20.33 (1.89)	20 (1.41)	.85	24.33 (5.44)	19.66 (0.94)	.35
Black or African American	19 (0)	19.4 (0.8)	.37	19 (0)	19.5 (0.5)	.50
White	22 (2)	21 (2)	.75	19.5 (0.5)	19 (0)	.50
Mixed race	19.18 (0.39)	19.18 (1.85)	≥.99	19.4 (0.75)	19 (1.2)	.17

[a]DMCI: Decision-Making Control Instrument.

[b]N/A: not applicable.

Age-related trends showed that younger participants (aged 18-34 years) had DMCI scores of 19.52 (SD 1.28) at baseline and 19.17 (SD 0.64) at follow-up via teleconsent, which were not significantly different (*P*=.26). Comparatively, older age groups showed varied results, with the 35- to 50-year-olds' scores being 19.14 (SD 0.35) and 20.14 (SD 3.48) for teleconsent, yielding a *P* value of .51. The 51- to 64-year age group had stable scores of 19.5 (SD 0.5) across both assessments, while there were no participants aged 65 years and older.

Racial comparisons show that Asian participants recorded the highest baseline score in the teleconsent group (mean 20.33, SD 1.89), while White participants presented more stable averages across both modes of consent. Asian participants had mean (SD) DMCI scores of 20.33 (1.89) and 20.00 (1.41) at baseline and follow-up via teleconsent (*P*=.85). Notably, the performance of Black or African American and mixed-race individuals illustrated lesser variability in scores.

QuIC Subgroup Analysis

There were no significant differences in comprehension scores between teleconsent and in-person methods across the sex categories, with males reporting a baseline mean (SD) score of 61.25 (1.92) for teleconsent and 60 for in-person consent, while females scored 61.25 (2.18) and 60.04, respectively (Table 3). However, the *P* values for both sexes were .85 and .36, indicating no statistically significant differences between baseline and follow-up.

XSL•FO
RenderX

Table 3. Comparison of QuIC[a] scores by sex, age, and race for teleconsent and in-person consent groups at baseline and follow-up (maximum possible sore=80).

	Teleconsent			In-Person		
	Baseline QuIC	Follow-up QuIC	*P* value	Baseline QuIC	Follow-up QuIC	*P* value
Sex, mean (SD)						
Male	61.25 (1.92)	60.75 (4.09)	.85	60 (3.74)	59.33 (5.12)	.82
Female	61.25 (2.18)	60.79 (4.21)	.61	60.04 (3.17)	60.81 (2.70)	.36
Age (years), mean (SD)						
18-34	61.30 (1.78)	61.78 (3.89)	.60	60.24 (3.28)	60.43 (3.92)	.87
35-50	61.14 (3.23)	58.86 (4.23)	.21	59.88 (2.76)	60.88 (1.76)	.43
51-64	61 (1)	58 (2)	.41	62 (0)	61.5 (0.5)	.50
65 and older	0 (0)	0 (0)	N/A[b]	53 (0)	58 (0)	N/A
Race, mean (SD)						
Asian	61.66 (1.25)	60 (4.08)	.64	60 (3.56)	63.33 (0.47)	.32
Black or African American	62.2 (0.98)	59.4 (4.08)	.25	62.5 (0.5)	61.5 (0.5)	.29
White	61.05 (2.33)	61.27 (3.60)	.81	59.92 (3.31)	60.28 (3.45)	.71
Mixed race	60.5 (2.5)	60 (8)	.96	59 (3)	58.5 (3.5)	.92

[a]QuIC: Quality of Informed Consent.

[b]N/A: not applicable.

The 18-34 age group had a teleconsent baseline mean (SD) score of 61.30 (1.78) compared to 60.24 (3.28) for in-person, with a *P* value of .87 indicating no significant difference. The 35-50 age group's scores exhibited a teleconsent baseline of 61.14 (3.23) and an in-person score of 59.88 (2.76), with a *P* value of .43, again highlighting a lack of significant differences. The older age groups (51-64 years) demonstrated teleconsent baseline scores of 61 (SD 1) and 58 (SD 2) respectively for in-person, yielding a *P* value of .41.

Regarding race, the Black or African American group exhibited the highest baseline teleconsent score of 62.20 (SD 0.98), while the lowest was seen in the mixed-race category, scoring 60.50 (SD 2.50). Mixed race individuals exhibited a teleconsent score of 60.5 (SD 2.5) and 59 (SD 3) in-person, yielding a *P* value of .96, while White participants had baseline scores of 61.05 (SD 2.33) for teleconsent and 59.92 (SD 3.31) for in-person, with a *P* value of .81. The *P* values across race categories ranged from .25 to .96, indicating a lack of statistical significance in the differences observed.

Discussion

Principal Findings

Our study provides important insights into how participants understand and make decisions during teleconsent sessions compared to traditional in-person IC visits. Notably, our results suggest that telehealth can match or even improve participants' understanding of the consent process in certain situations. This study provides evidence that telehealth can effectively obtain IC for research studies in health. This challenges the common belief that in-person interactions are always better at facilitating informed consent. It also emphasizes the need to reevaluate consent methods in the evolving digital health landscape.

Findings from this study provide evidence that using telehealth to obtain IC is a viable method that allows participants to obtain similar levels of comprehension of the study consent form and make voluntary decisions as the in-person method. In some cases, we found that participants retained information better in the teleconsent group 30 days after the IC session compared to the in-person group. This supports previous findings suggesting that obtaining IC digitally can streamline the IC process compared to in-person methods [23].

The subgroup analysis suggests no significant differences in DMCI or QuIC scores across sexes, age groups, or races when comparing teleconsent and in-person methods. We found no significant differences in teleconsent participants' comprehension levels of perceived voluntariness, trust, and decision self-efficacy from baseline to 30-day follow-up. The lack of significant differences provides evidence that using teleconsent among various populations can yield similar results to in-person IC efforts.

Our findings, aligned with prior studies, suggest the efficacy of telehealth in increasing accessibility and improving participant engagement in clinical research. For example, previous research has demonstrated that telehealth significantly reduces geographic barriers to seeking care [24]. Our findings build on this foundation by specifically targeting the comprehension aspect of the consent process, thereby filling an important gap in current scholarships.

However, teleconsent may also exacerbate health inequities among individuals and communities that lack access to digital tools or lack digital literacy [25,26]. Therefore, it is essential to carefully examine the use of teleconsent by conducting a needs assessment or feasibility study for the target audience to understand if teleconsent is an appropriate IC method.

XSL•FO
RenderX

The implications of our study findings are profound for future clinical research. As health care systems continue to adapt and incorporate telehealth, there is a clear opportunity to standardize teleconsent in many studies. Teleconsent has the potential to streamline participant enrollment and also foster a more inclusive environment where individuals who might otherwise face barriers to participation, such as those in remote areas or those with mobility challenges, can take part in important research without compromising their understanding of the process. Ultimately, this evolution in practice could improve participant understanding and engagement, advancing the goals of ethical, effective, and translational clinical research.

This study had limitations. It is a single-site study, and the diversity in some demographics was limited due to the difficulty of recruiting participants during a public health emergency. Also, we did not collect the duration of the consent session for each group, which would have added more evidence regarding differences in efficiencies between in-person and teleconsent methods. Future research should explore the long-term effects of teleconsent on participant engagement and retention, particularly among different populations. Studying how different technologies, such as smartphone apps, compare to traditional methods could provide valuable insights for optimizing telehealth strategies.

The findings of this study emphasize that study participants using teleconsent and in-person were able to comprehend and make voluntary decisions related to IC alike, that is, comprehension and voluntariness are not negatively impacted by this type of remote IC modality.

Conclusions

This study aimed to assess the comprehension and decision-making effectiveness of IC processes conducted via telehealth compared to traditional in-person visits. Key findings revealed that telehealth maintains comparable levels of participant understanding and engagement, overcoming significant barriers such as geographic limitations and accessibility challenges. As health care increasingly adopts digital solutions, these findings highlight the potential of telehealth to enhance participant recruitment and retention in clinical research. This is particularly relevant for policy makers, who should integrate telehealth practices into regulatory frameworks to promote equitable access and improve health outcomes in diverse populations.

Acknowledgments

This study was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (R21TR002088 and RC2TR004380). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, nor does it mention of department or agency names imply endorsement by the US government. All authors had access to the data and a role in writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey Instrucments.
[PDF File (Adobe PDF File), 199 KB - jmir_v27i1e63473_app1.pdf]

Multimedia Appendix 2
Comparison of baseline and follow-up scores for teleconsent and in-person surveys.
[DOCX File , 15 KB - jmir_v27i1e63473_app2.docx]

References

1. Grant SC. Informed consent—we can and should do better. JAMA Netw Open 2021;4(4):e2110848 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.10848] [Medline: 33909058]
2. Moser DJ, Arndt S, Kanz JE, Benjamin ML, Bayless JD, Reese RL, et al. Coercion and informed consent in research involving prisoners. Compr Psychiatry 2004;45(1):1-9. [doi: 10.1016/j.comppsych.2003.09.009] [Medline: 14671730]
3. Berrios C, James CA, Raraigh K, Bollinger J, Murray B, Tichnell C, et al. Enrolling genomics research participants through a clinical setting: the impact of existing clinical relationships on informed consent and expectations for return of research results. J Genet Couns 2018;27(1):263-273 [FREE Full text] [doi: 10.1007/s10897-017-0143-2] [Medline: 28932961]
4. Moeini S, Shahriari M, Shamali M. Ethical challenges of obtaining informed consent from surgical patients. Nurs Ethics 2020;27(2):527-536. [doi: 10.1177/0969733019857781] [Medline: 31296111]
5. Hall DE, Prochazka AV, Fink AS. Informed consent for clinical treatment. CMAJ 2012;184(5):533-540 [FREE Full text] [doi: 10.1503/cmaj.112120] [Medline: 22392947]
6. Grady C. Enduring and emerging challenges of informed consent. N Engl J Med 2015;372(22):2172 [FREE Full text] [doi: 10.1056/NEJMc1503813] [Medline: 26017840]

XSL•FO
RenderX

7. McNaughton CD, Cawthon C, Kripalani S, Liu D, Storrow AB, Roumie CL. Health literacy and mortality: a cohort study of patients hospitalized for acute heart failure. J Am Heart Assoc 2015;4(5):e001799 [FREE Full text] [doi: 10.1161/JAHA.115.001799] [Medline: 25926328]
8. Altin SV, Stock S. The impact of health literacy, patient-centered communication and shared decision-making on patients' satisfaction with care received in German primary care practices. BMC Health Serv Res 2016;16(1):450 [FREE Full text] [doi: 10.1186/s12913-016-1693-y] [Medline: 27577702]
9. Schenker Y, Meisel A. Informed consent in clinical care: practical considerations in the effort to achieve ethical goals. JAMA 2011;305(11):1130-1131. [doi: 10.1001/jama.2011.333] [Medline: 21406651]
10. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
11. Tam NT, Huy NT, Thoa LTB, Long NP, Trang NTH, Hirayama K, et al. Participants' understanding of informed consent in clinical trials over three decades: systematic review and meta-analysis. Bull World Health Organ 2015;93(3):186-98H [FREE Full text] [doi: 10.2471/BLT.14.141390] [Medline: 25883410]
12. Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. BMC Med Ethics 2013;14(1):28 [FREE Full text] [doi: 10.1186/1472-6939-14-28] [Medline: 23879694]
13. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
14. Schillinger D, Bindman A, Wang F, Stewart A, Piette J. Functional health literacy and the quality of physician-patient communication among diabetes patients. Patient Educ Couns 2004;52(3):315-323. [doi: 10.1016/S0738-3991(03)00107-1] [Medline: 14998602]
15. Kass NE, Taylor HA, Ali J, Hallez K, Chaisson L. A pilot study of simple interventions to improve informed consent in clinical research: feasibility, approach, and results. Clin Trials 2015;12(1):54-66 [FREE Full text] [doi: 10.1177/1740774514560831] [Medline: 25475879]
16. Manta CJ, Ortiz J, Moulton BW, Sonnad SS. From the patient perspective, consent forms fall short of providing information to guide decision making. J Patient Saf 2021;17(3):e149-e154 [FREE Full text] [doi: 10.1097/PTS.0000000000000310] [Medline: 27490160]
17. Khairat S, Tirtanadi K, Ottmar P, Sleath B, Obeid J. Evaluating the perceptions of teleconsent in urban and rural communities. Eur J Biomed Inform 2019;15(2) [FREE Full text] [Medline: 32802171]
18. Welch BM, Marshall E, Qanungo S, Aziz A, Laken M, Lenert L, et al. Teleconsent: a novel approach to obtain informed consent for research. Contemp Clin Trials Commun 2016;3:74-79 [FREE Full text] [doi: 10.1016/j.conctc.2016.03.002] [Medline: 27822565]
19. Sawant S, Madathil KC, Molloseau M, Obeid J. Overcoming recruitment hurdles in clinical trials: an investigation of remote consenting modalities and their impact on workload, workflow, and usability. Appl Ergon 2024;114:104135. [doi: 10.1016/j.apergo.2023.104135] [Medline: 37713927]
20. Newlin T, McCall T, Ottmar P, Welch B, Khairat S. Assessing the satisfaction of citizens using teleconsent in clinical research. Stud Health Technol Inform 2018;247:685-689. [Medline: 29678048]
21. Miller VA, Ittenbach RF, Harris D, Reynolds WW, Beauchamp TL, Luce MF, et al. The decision making control instrument to assess voluntary consent. Med Decis Making 2011;31(5):730-741 [FREE Full text] [doi: 10.1177/0272989X11398666] [Medline: 21402793]
22. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent: a new measure of understanding among research subjects. J Natl Cancer Inst 2001;93(2):139-147. [doi: 10.1093/jnci/93.2.139] [Medline: 11208884]
23. Chimonas S, Lipitz-Snyderman A, Matsoukas K, Kuperman G. Electronic consent in clinical care: an international scoping review. BMJ Health Care Inform 2023;30(1):e100726 [FREE Full text] [doi: 10.1136/bmjhci-2022-100726] [Medline: 37423643]
24. Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, et al. Advancing health equity and access using telemedicine: a geospatial assessment. J Am Med Inform Assoc 2019;26(8-9):796-805 [FREE Full text] [doi: 10.1093/jamia/ocz108] [Medline: 31340022]
25. Nouri S, Khoong EC, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catal Innov Care Deliv 2020;1 [FREE Full text] [doi: 10.1056/CAT.20.0123]
26. Geracitano J, Barron L, McSwain D, Khairat S. How is digital health suitability measured for communities? A systematic review. Digit Health 2024;10:20552076241288316 [FREE Full text] [doi: 10.1177/20552076241288316] [Medline: 39398896]

Abbreviations

DMCI: Decision-Making Control Instrument
HIPAA: Health Insurance Portability and Accountability Act

XSL•FO
RenderX

IC: informed consent
QuIC: Quality of Informed Consent
SAHL-E: Short Assessment of Health Literacy-English

Edited by A Coristine; submitted 21.06.24; peer-reviewed by E Cohen, T Church; comments to author 15.10.24; revised version received 04.11.24; accepted 16.01.25; published 05.03.25.

Please cite as:
Khairat S, Ottmar P, Chourasia P, Obeid J
Effectiveness of Telehealth Versus In-Person Informed Consent: Randomized Study of Comprehension and Decision-Making
J Med Internet Res 2025;27:e63473
URL: https://www.jmir.org/2025/1/e63473
doi:10.2196/63473
PMID:40053795

©Saif Khairat, Paige Ottmar, Prabal Chourasia, Jihad Obeid. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact on Patient Outcomes of Continuous Vital Sign Monitoring on Medical Wards: Propensity-Matched Analysis

Bradley Rowland[1], MD; Amit Saha[2,3,4], MS, PhD; Vida Motamedi[5], MD; Richa Bundy[1], MS; Scott Winsor[6], MD; Daniel McNavish[6], MD; William Lippert[1], MD, MPH; Ashish K Khanna[2,3,4], MD, MS

[1]Department of Internal Medicine, Section of Hospital Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[2]Department of Anesthesiology, Wake Forest University School of Medicine, Winston Salem, NC, United States

[3]Department of Anesthesiology, Perioperative Outcomes and Informatics Collaborative, Winston-Salem, NC, United States

[4]Department of Anesthesiology, Outcomes Research Consortium, Cleveland, OH, United States

[5]Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, United States

[6]Corewell Health Frederik Meijer Heart & Vascular Institute, Michigan State University, Grand Rapids, MI, United States

Corresponding Author:
Bradley Rowland, MD
Department of Internal Medicine
Section of Hospital Medicine
Wake Forest University School of Medicine
1 Medical Center Boulevard
Winston-Salem, NC, 27157
United States
Phone: 1 336 716 9218
Email: barowlan@wakehealth.edu

Abstract

Background: Continuous and wireless vital sign (VS) monitoring on hospital wards is superior to intermittent VS monitoring at detecting VS abnormalities; however, the impact on clinical outcomes remains to be confirmed. A recent propensity-matched study of primary surgical patients found decreased odds of intensive care unit (ICU) admission and mortality in patients receiving continuous monitoring. Primary surgical patients are inherently different from their medical counterparts who typically have high morbidity, including frailty. Continuous monitoring research has been limited in primary medical patients.

Objective: This study aims to evaluate the clinical outcomes of primary medical patients who received either continuous or, as a contemporaneous control, intermittent vital monitoring as the standard of care using propensity matching.

Methods: Propensity-matched analysis of a population-based sample of 7971 patients admitted to the medical wards between January 2018 and December 2019 at a single, tertiary United States medical center. The continuous monitoring device measures oxygen saturation, heart rate, respiratory rate, continuous noninvasive blood pressure, and either 3-lead or 5-lead electrocardiogram. Patients received either 12 hours or more of continuous and wireless VS monitoring (n=1450) or intermittent VS monitoring (n=6521). The primary outcome was the odds of a composite of in-hospital mortality or ICU transfer during hospitalization. Secondary outcomes were the odds of individual components of the primary outcome, as well as heart failure (HF), myocardial infarction (MI), acute kidney injury (AKI), and rapid response team (RRT) activations.

Results: Those who received intermittent VS monitoring had greater odds of a composite of in-hospital mortality or ICU admission (odds ratio [OR] 2.79, 95% CI 1.89-4.25; *P*<.001) compared with those who had continuous and wireless VS monitoring. The odds of HF (OR 1.03, 95% CI 0.83-1.28; *P*=.77), MI (OR 1.58, 95% CI 0.77-3.47; *P*=.23), AKI (OR 0.74, 95% CI 0.62-1.02; *P*=.06), and RRT activation (OR 0.94, 95% CI 0.75-1.19; *P*=.62) were similar in both groups.

Conclusions: In this propensity-matched study, medical ward patients who received standard of care intermittent VS monitoring were at nearly 3 times greater odds of transfer to the ICU or death compared with those who received continuous VS monitoring. Our study was primarily limited by the inability to match patients on admission diagnosis due to limitations in electronic health record data. Other limitations included the number of and reasons for false alarms, which can be challenging with continuous monitoring strategies. Given the limitations of this work, these observations need to be confirmed with prospective interventional trials.

(J Med Internet Res 2025;27:e66347) doi:10.2196/66347

XSL•FO
RenderX

KEYWORDS

clinical; continuous; monitoring; outcomes; medical ward; wireless; wireless monitoring; vital sign; ward; patient outcome; hospital ward; clinical outcome; contemporaneous control; contemporaneous; teenager; young adult; adult; monitoring device; wireless device; wearable; patient monitoring

Introduction

Vital sign (VS) monitoring is routine practice on the general medical wards and is typically performed as an intermittent "spot check," at 4 to 8-hour intervals [1]. However, this intermittent VS monitoring may fail to detect VS abnormalities [2-5], and perturbations in VS may be sustained for long periods between spot check intervals [6,7]. The failure to detect abnormal VS or their trends has significant implications for clinical care [8]. For example, transient hypotension is associated with worse outcomes [9-14]. Furthermore, the failure to act on or recognize deterioration was identified as the most common reason for harm in a study of 2010 hospitalized patient deaths in the United Kingdom between 2010 and 2012 [15].

Aggregate VS scoring systems such as the early warning score (EWS) have been developed to aid in the identification and risk stratification of patients at risk of or experiencing clinical deterioration. Once identified, clinical response teams, often termed "rapid response teams" (RRTs), may intervene to provide stabilization or transfer to a higher level of care. This identification and response have been described as the afferent and efferent arms, respectively, of rapid response [16]. The afferent response is contingent upon appropriate and timely recognition of clinical deterioration, which is accomplished primarily through the detection of abnormal VS. The use of intermittent VS monitoring may fail to detect clinical deterioration, thereby precluding or delaying this process.

The advent of wearable and wireless devices capable of continuous VS monitoring integrated into the clinical workflow offers a solution to the failure-to-detect paradigm associated with intermittent VS monitoring; however, clinical outcome data are needed. Evidence regarding outcomes with continuous VS monitoring relies almost exclusively on before-and-after comparison studies, and the research has also focused primarily on surgical patients [17]. Two systematic reviews of continuous VS monitoring concluded that existing study quality is low-to-moderate and does not show benefit of continuous compared with intermittent VS monitoring [18,19]. However, these reviews were limited by the heterogeneity of devices included in the studies and narrow scope of the patient populations. That is, the analyses included over 15 different device types making cross comparison and generalization challenging. Additionally, the analyses included almost entirely surgical populations, with individual studies including niche surgical areas such as gastrointestinal cancer, again limiting the generalizability of the analyses.

The granular detection capabilities of continuous and wireless VS monitoring mitigate the variability that is inevitable in today's complex health care systems, allowing capture and intervention of clinical deterioration to improve quality of care.

At Atrium Health Wake Forest Baptist Medical Center, continuous VS monitoring using wearable and portable monitoring device technology on surgical and medical wards has been in place since 2015 and is coupled with VS threshold–based alarms and interventions based on best practices. An earlier report from our institution demonstrated a significant decrease in RRT calls after implementation [20,21]. We have also reported substantial changes in heart rate (HR) and blood pressure (BP) detected by these monitoring devices that would have gone undetected with traditional intermittent VS monitoring. Furthermore, we recently reported a propensity-matched analysis of 36,000 surgical ward patients, where 12,345 intermittently monitored patients were compared with 7955 continuously monitored patients, and a nearly 2 times increased odds of intensive care unit (ICU) admission or in-hospital mortality was seen in the intermittently monitored group [22]. The inherent differences between a primary surgical population, where nearly all continuous postprocedure monitoring research has been conducted, and a primary medical population is paramount to the significance of this work. Primary medical admissions are inherently different from surgical ones. Medical patients often have multiple acute problems related to single or multiple intersected disease processes. Often, patients admitted to a medical service are deemed too sick or high risk to undergo surgery.

Therefore, we sought to evaluate and compare clinical outcomes in a large propensity-matched medical ward population admitted to a single institution who received either continuous or, as a contemporaneous control, intermittent VS monitoring as the standard of care every 4 hours to 8 hours. We hypothesized that earlier detection of VS aberrations through continuous VS monitoring would reduce ICU transfer and in-hospital mortality. Secondary outcomes included heart failure (HF), myocardial infarction (MI), acute kidney injury (AKI), and RRT activations.

Methods

Population and Setting

We first identified adult patients (aged ≥18 years) admitted to a tertiary care center (Atrium Health Wake Forest Baptist Medical Center main campus) over 2 years (between January 1, 2018, and December 31, 2019) under a primary medical service.

Primary medical services included hospitalist, general internal medicine, and family medicine. Provider staffing varied by service but included physician attendings, advanced practice providers, and medical residents. Primary medical services represented the typical breadth of patients admitted to hospital medicine services. Admission diagnoses were not included in propensity matching. Patients having any surgery during admission were excluded from analysis. Race and ethnicity were self-described by patients on admission then retrospectively collected with other demographics from the electronic health record (EHR).

XSL•FO
RenderX

Ethical Considerations

Regarding human subject research ethics review, our study was a secondary analysis and did not actively involve nor enroll human subjects. Our study and a priori defined statistical analysis plan were approved by the Wake Forest University Health Sciences Institutional Review Board (IRB; approval number IRB00083520). Informed consent from human subjects was not obtained for this secondary analysis. Primary consent for data collection was allowed by the IRB without additional consent. Privacy and confidentiality protection were maintained using de-identified data for analysis and using only data needed for the analysis. Data were stored in a secure location per IRB requirements.

Monitoring Protocols

Patients admitted to the medical wards either received (1) VS monitoring with a continuous monitoring device, ViSi Mobile, (Sotera Wireless Inc), or (2) standard intermittent VS monitoring (every 4-8 hours). Briefly, the ViSi Mobile system is a portable wrist-mounted system that received clearance from the Food and Drug Administration in 2012. These ViSi devices were purchased and implemented by the Wake Forest Baptist Hospital system in 2015. Today, ViSi monitors are on several different medical and surgical units; however, these devices are not available on all units. If units do not have ViSi monitors or patients decline to wear the monitor, then intermittent monitoring is used.

Hospital units with ViSi devices are staffed by the same number of personnel as those not doing this monitoring. Nurse-to-patient ratios on our floors range from 1:3 to 1:6. For this analysis, patients were not randomized into continuous or intermittent VS monitoring. The type of monitoring was dependent upon the floor a patient was admitted to, which was dependent on bed availability and not patient clinical status. Based on clinical experience, there would not have been a significant crossover between groups; that is, once a patient arrived on the floor and was allocated to either continuous or intermittent VS monitoring, they would then most likely continue that strategy until discharge—unless, for example, a patient was transferred to a different unit and received a different type of monitoring strategy there.

The ViSi device records oxygen saturation (SpO_2), HR, respiratory rate (RR), continuous noninvasive BP, and either 3-lead or 5-lead electrocardiogram. Continuous BP measurements are estimated from pulse arrival time, specifically the time that elapses between the R wave being detected and the arrival of the resulting pulse at the SpO_2 finger sensor. According to the manufacturer, the estimated maximal mean error for BP measurements is 5 mm Hg with a standard deviation ≤8 mm Hg [23].

ViSi monitors were calibrated daily with the oscillometric brachial cuff method and connected to the hospital's wireless network, and measurements were distributed to the central nurse's station for that unit and continuously displayed. The ViSi monitoring system measures VS nearly continuously, and these data are transmitted to the central nurse's station for each respective unit; however, these data are not transmitted nor recorded directly in the EHR. That is, nurses or nursing assistants must still see the patient and collect VS from the device or the monitoring station remotely and then manually document (chart) these vitals in the EHR. No matter what, the VS assessment includes direct patient visualization to measure and record patient alertness. The VS data not recorded in the EHR are temporarily stored on the ViSi manufacturer's cloud for a set period of time and are then destroyed. Unfortunately, we were *only* able to acquire and evaluate the EHR data.

VS abnormalities exceeding established thresholds for the local hospital system generated alerts at the floor's central nursing station and to the nurses' hospital-supplied phones (ASCOM). Alarms not addressed by the primary nurses were escalated to other floor nurses, then to the unit manager. Time to escalation depended on the VS and the relative degree of abnormality. For example, SpO_2 abnormality alerts, if not acknowledged, were quickly escalated over minutes.

EWS at our institution are based on the patient's weighted VS of HR, RR, SpO_2, temperature, BP, rate and percentage of inspired oxygen, and level of awareness determined by the patient's nurse [24]. EWS are calculated by the EHR every 4 hours based on data input from ViSi monitors, which can be collected from the central display or at the bedside, or those collected by intermittent monitoring. Patients, regardless of monitoring strategy, must be observed at bedside to determine level of awareness for EWS calculation. Total scores range from 0 to 20. Ad hoc EWS may be generated by the clinical team in response to changes in a patient's clinical status through either monitoring strategy. For example, if abnormal VS are detected by ViSi monitoring and an alarm is generated at the central nurse's station, the clinical team may assess the patient and generate an updated EWS. An RRT activation is triggered if the EWS is ≥8.

Outcomes

Our primary outcome was the risk of a composite of in-hospital mortality or ICU admission during the hospital encounter. Secondary outcomes were the individual components of the primary outcome (in-hospital mortality and ICU admission), as well as HF, MI, AKI, and RRT activations throughout the hospitalization. Exploratory outcomes included hospital length of stay (LOS), ICU LOS, and median EWS values for the entire LOS. Diagnoses of HF, MI, and AKI were identified using International Classification of Diseases, Tenth Revision billing codes available in the EHR.

Statistical Analysis

Statistical analysis was performed in R v3.6.1 (R Foundation for Statistical Computing) using RStudio environment v1.1.456. The participants were first stratified into continuous VS monitoring and contemporaneous controls as intermittent VS monitoring groups. Sequential random nearest neighbor matching with a propensity score ratio of 1:3 and with a caliper of 0.25 [25,26] was used to balance the potential factors affecting the outcome including selection bias. To create propensity score–matched pairs, we performed random 1-to-3 matching using the MatchIt package in R, in which the control variables in our study were race, gender, ethnicity, age,

XSL•FO
RenderX

quarter-year of admission, BMI, Charlson Comorbidity Index (CCI), patient's primary insurance type, and source of admission [27]. The variable selection for propensity score matching was based on feedback from clinical expertise. To adjust for the effect of time and variability of the continuous wireless monitoring device use, the quarter-year of admission was used as a covariate in the model. Standardized differences were used as indicators of intergroup balance. If the standardized differences were <10%, the covariates between the 2 groups were considered well-balanced.

For summary statistics, categorical variables were described as numbers and percentages, and non-normally distributed continuous variables were described as median and IQRs.

Primary, secondary, and exploratory outcomes were compared between groups and subgroups. Multivariable logistic regression analysis was performed to detect differences in outcomes between groups. Initially, restricted cubic splines were included with age and CCI to relax the linearity assumptions of the multivariable logistic regression model. As the nonlinear components of age and CCI were not statistically significant, the splines were removed from the variables. The multivariable logistic regression adjusted model was controlled for race, gender, ethnicity, age, BMI, patient's primary insurance type, source of admission, and CCI. The *P* value for significance level within the primary and secondary outcome analyses was set at .004 for each group of analyses after Bonferroni correction for multiple comparisons.

To test the influence of unobserved confounding, we used an E-value analysis to quantify the potential implications of unmeasured confounders. E-value is defined as the minimum strength of an association on the risk ratio scale that an unmeasured confounder would need to have with both the exposure and the outcome, conditional on the measured covariates, to fully explain away a specific exposure-outcome association.

We further performed multivariate logistic regression analysis for the composite outcome on the whole study cohort (n=7971) to assess the effect of continuous monitoring sensitivity after controlling for race, gender, ethnicity, age, BMI, and CCI as a sensitivity analysis.

Results

We identified 7971 patients who met inclusion criteria (Figure 1). Over the study period, 1450 patients received 12 hours or more of continuous VS monitoring, compared with 6521 contemporaneous controls who received intermittent VS monitoring (Table 1).

Figure 1. CONSORT (Consolidated Standards Of Reporting Trials) diagram of the retrospective propensity-matched analysis of continuous vital sign monitoring (CVSM) versus intermittent vital sign monitoring (IM) in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.

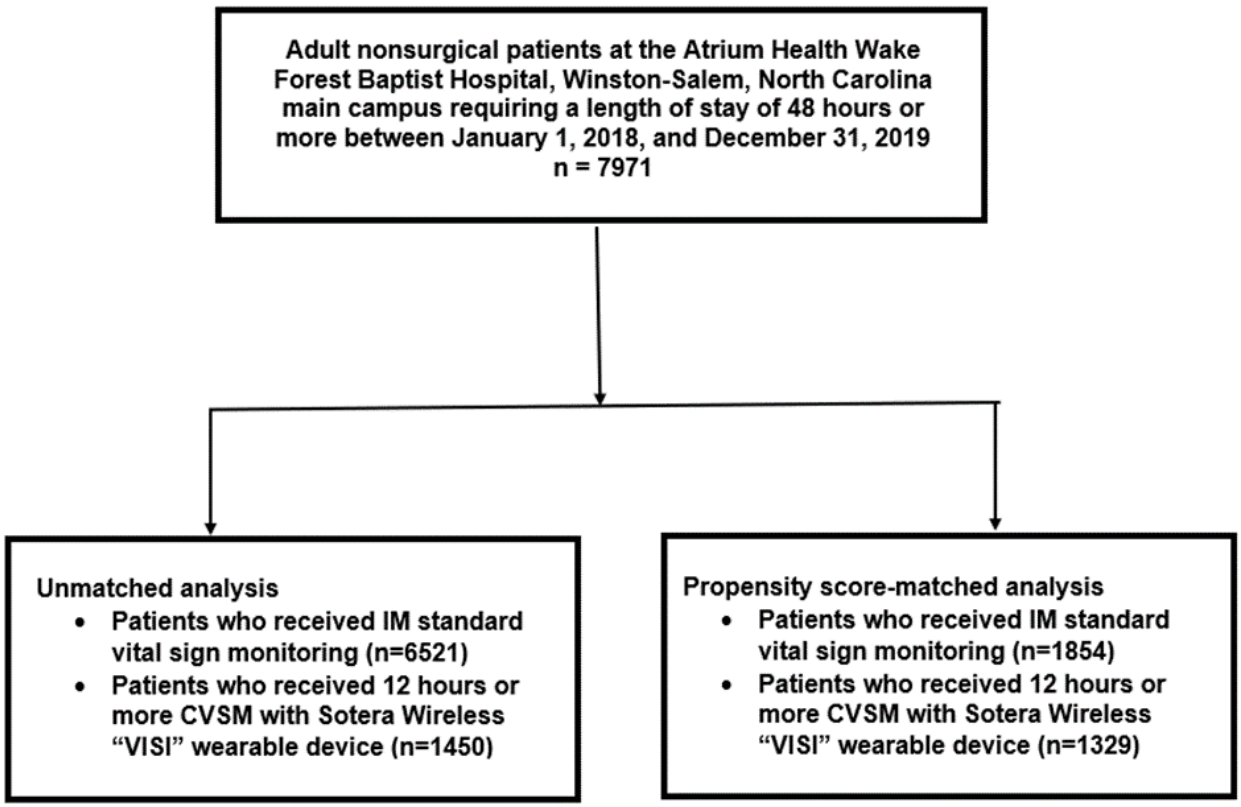

XSL•FO
RenderX

Table 1. Baseline patient demographics and characteristics of the retrospective propensity-matched analysis of continuous versus intermittent vital sign monitoring of primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.

Characteristics	Unmatched cohort analysis			Propensity score–matched cohort analysis		
	Continuous vital sign monitoring (n=1450)	Intermittent vital sign monitoring (n=6521)	SMD[a]	Continuous vital sign monitoring (n=1329)	Intermittent vital sign monitoring (n=1854)	SMD
Quartile (Q)-year, n (%)			1.936			0.107
Q1-2018	411 (28.3)	417 (6.4)		374 (28.1)	417 (22.5)	
Q2-2018	420 (29.0)	452 (6.9)		372 (28)	452 (24.4)	
Q3-2018	467 (32.2)	528 (8.1)		431 (32.4)	528 (28.5)	
Q4-2018	30 (2.1)	926 (14.2)		30 (2.3)	87 (4.7)	
Q1-2019	10 (0.7)	1324 (20.3)		10 (0.8)	26 (1.4)	
Q2-2019	10 (0.7)	1211 (18.6)		10 (0.8)	44 (2.4)	
Q3-2019	49 (3.4)	1082 (16.6)		49 (3.7)	119 (6.4)	
Q4-2019	53 (3.7)	581 (8.9)		53 (4)	181 (9.8)	
Age (years), median (IQR)	61.00 (48.00-72.00)	62.00 (48.00-73.00)	0.012	61.00 (48.00-72.00)	61.00 (46.00-74.00)	0.018
Sex, n (%)			0.082			0.038
Male	678 (46.8)	3316 (50.9)		643 (48.4)	932 (50.3)	
Female	772 (53.2)	3205 (49.1)		686 (51.6)	922 (49.7)	
Ethnicity (Hispanic/Latino), n (%)	77 (5.3)	339 (5.2)	0.005	69 (5.2)	96 (5.2)	0.001
Racial identity, n (%)			0.184			0.067
White	887 (61.2)	4520 (69.3)		861 (64.8)	1256 (67.7)	
Black or African American	473 (32.6)	1594 (24.4)		386 (29)	484 (26.1)	
Other	90 (6.2)	407 (6.2)		82 (6.2)	114 (6.1)	
CCI[b], median (IQR)	3.00 (1.00-5.00)	3.00 (1.00-5.00)	0.055	3.00 (1.00-5.00)	3.50 (1.00-5.00)	0.079
Hypertension, n (%)	605 (41.7)	2750 (42.2)	0.009	527 (39.7)	699 (37.7)	0.040
BMI, median (IQR)	27.77 (23.17-32.88)	27.91 (22.99-32.70)	0.001	27.76 (23.19-32.45)	27.80 (22.81-32.73)	0.020
Insurance class, n (%)			0.052			0.034
Commercial	235 (16.2)	941 (14.4)		212 (16)	273 (14.7)	
Government	1078 (74.3)	4921 (75.5)		995 (74.9)	1410 (76.1)	
Other	137 (9.4)	659 (10.1)		122 (9.2)	171 (9.2)	
Source of admission, n (%)			0.039			0.015
ED[c]	1199 (82.7)	5487 (84.1)		1124 (83.2)	1551 (83.7)	
Non-ED	251 (17.3)	1034 (15.9)		227 (16.8)	301 (16.3)	
Hospital service, n (%)			0.146			0.057
Hospitalist	822 (56.7)	4011 (61.5)		798 (60)	1119 (60.4)	
General medicine	463 (31.9)	2028 (31.1)		433 (32.6)	573 (30.9)	
Family medicine	165 (11.4)	482 (7.4)		98 (7.4)	162 (8.7)	

[a]SMD: standardized mean difference.

[b]CCI: Charlson Comorbidity Index.

[c]ED: emergency department.

XSL•FO
RenderX

For propensity matching, 1329 (1329/1450, 91.7%) patients receiving continuous wireless VS monitoring were matched with 1854 (1854/6521, 28.4%) contemporaneous patients receiving intermittent VS monitoring. When propensity matching was not possible, they were excluded from propensity score–matched analysis. Multimedia Appendix 1 shows the patient demographics and clinical characteristics of patients not matched.

Before propensity matching, 365 (365/6521, 5.6%) of intermittent VS monitoring patients had in-hospital mortality or ICU admission compared with 29 (29/1450, 2%) in the continuous wireless VS monitoring group (Multimedia Appendix 2).

After propensity matching, patients who had intermittent VS monitoring had greater odds of the primary composite outcome, compared with patients who were on continuous VS monitoring (odds ratio 2.79, 95% CI 1.89-4.25; *P*<.001; Table 2, Figure 2).

Table 2. Primary and secondary outcomes in the unmatched and matched models from the retrospective propensity-matched analysis of continuous versus intermittent vital sign monitoring in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.

Outcomes	Unmatched model		Propensity-matched model		E value
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	
Composite of in-hospital mortality or ICU[a] admission	2.79 (1.96-4.15)	<.001	2.79 (1.89-4.25)	<.001	2.73
In-hospital mortality	2.11 (0.99-4.47)	.08	1.91 (0.79-4.31)	.18	2.11
ICU admission	3.02 (2.06-4.63)	<.001	3.07 (2.03-4.82)	<.001	2.91
Heart failure	0.96 (0.81-1.14)	.63	1.03 (0.83-1.28)	.77	1.14
Myocardial infarction	1.22 (0.69-2.38)	.52	1.58 (0.77-3.47)	.23	1.83
Acute kidney injury	1.01 (0.73-1.03)	.58	0.74 (0.62-1.02)	.06	1.60
RRT[b] activation	1.05 (0.84-1.22)	.92	0.94 (0.75-1.19)	.62	1.21

[a]ICU: intensive care unit.

[b]RRT: rapid response team.

Figure 2. Impact of continuous wireless monitoring on the primary composite outcome of intensive care unit (ICU) admission and mortality from the retrospective propensity-matched analysis of continuous versus intermittent vital sign monitoring in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.

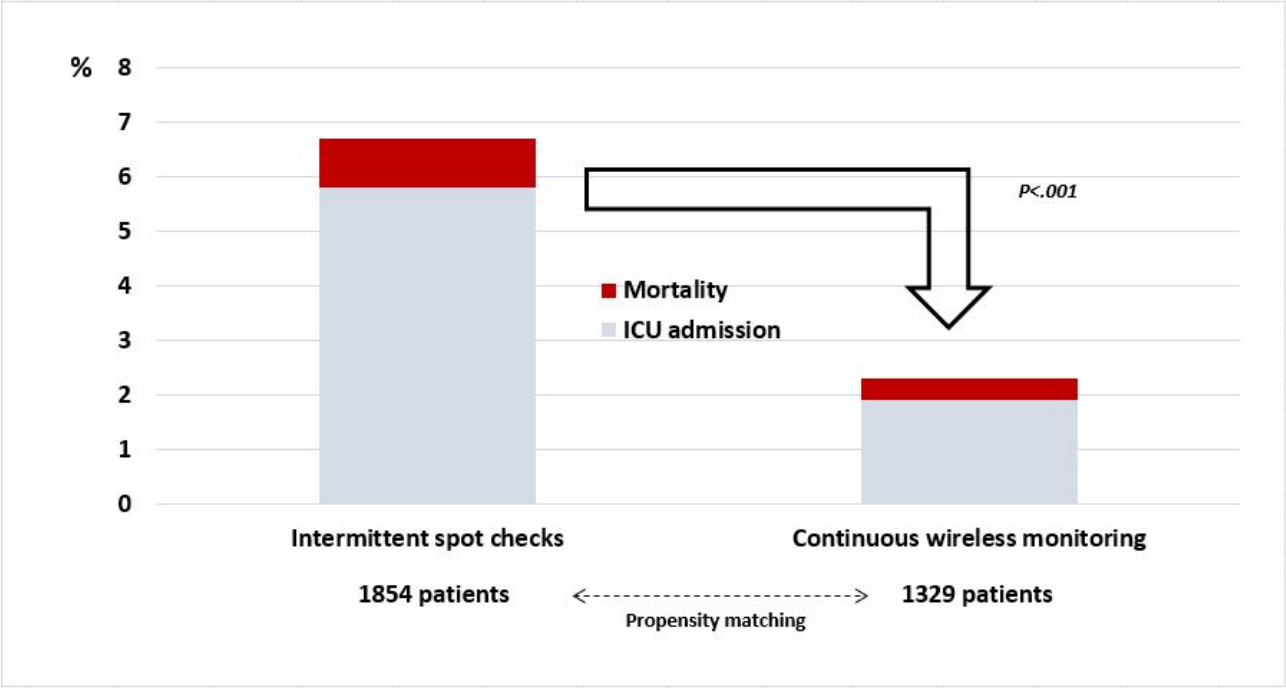

The primary composite outcome in the matched cohort was primarily driven by ICU admission (in-hospital mortality: 5/1329, 0.4% vs 17/1854, 0.9%; ICU admission: 25/1329, 1.9% vs 106/1854, 5.7%). Secondary outcome odds ratios (in-hospital mortality, ICU admission, HF, MI, AKI, and RRT activation) are shown in Table 2 and Figure 3. Intermittently VS-monitored patients had significantly increased odds of in-hospital mortality and ICU admission compared with continuous wireless VS

XSL•FO
RenderX

monitoring. Hospital LOS, ICU LOS, and median EWS during the LOS are shown in Table 3.

Figure 3. Forest plot of odds ratios (95% CIs) of primary and secondary outcomes for patients receiving intermittent monitoring vs continuous wireless vital sign monitoring from the retrospective propensity-matched analysis in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019. ICU: intensive care unit; RRT: rapid response team.

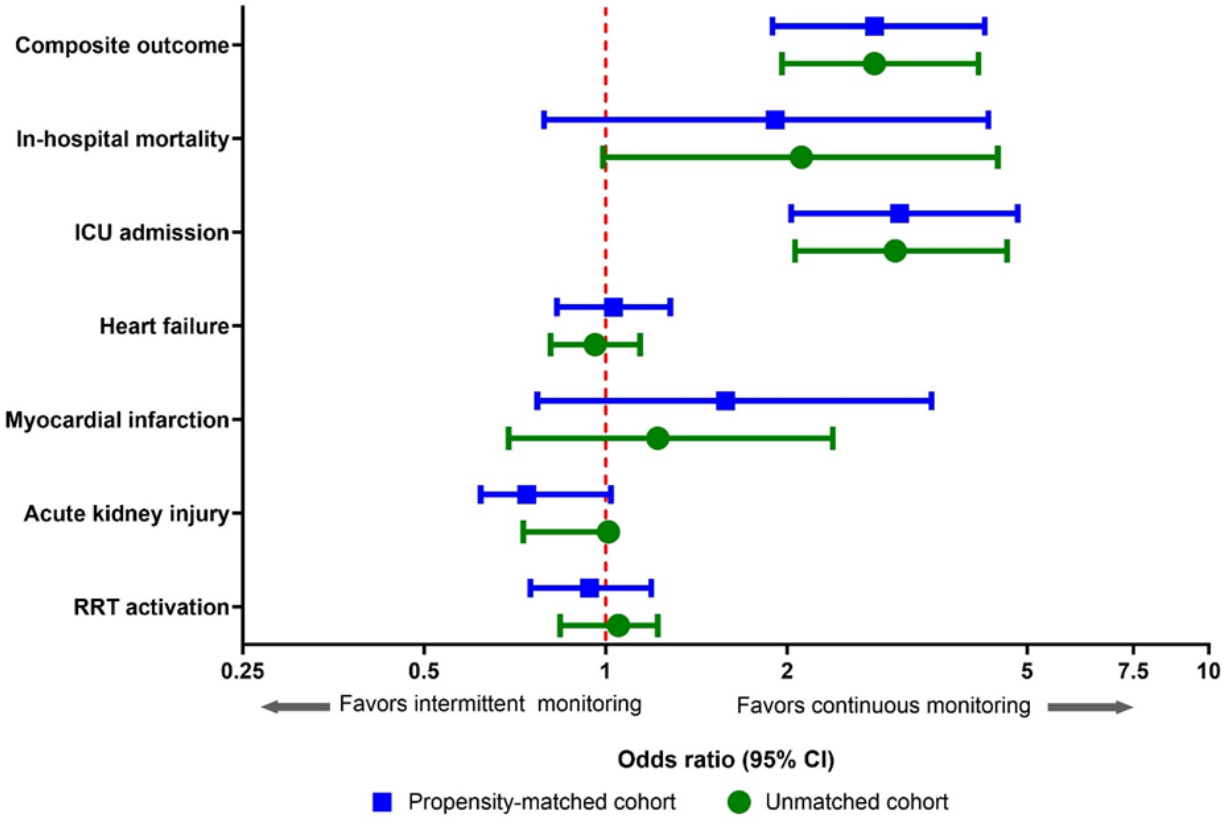

Table 3. Table 3. Exploratory outcomes for propensity-matched cohorts from the retrospective propensity-matched analysis of continuous versus intermittent vital sign monitoring in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.

Outcomes	Continuous vital sign monitoring (n=1329)	Intermittent vital sign monitoring (n=1854)	*P* value
Hospital LOS[a] (days), median (IQR)	3.12 (2.07-5.25)	3.21 (2.09-5.30)	.80
ICU[b] LOS (days), median (IQR)	3.32 (2.01-4.75)	2.45 (1.73-4.00)	.10
EWS[c] during LOS, median (minimum-maximum)	2.00 (1.00-9.00)	1.00 (0.00-9.00)	<.001

[a]LOS: length of stay.

[b]ICU: intensive care unit.

[c]EWS: early warning score.

Multimedia Appendix 3 shows the covariate balance plot before and after propensity score matching between the continuous monitoring cohort and intermittent monitoring cohort. A caliper width of 0.25 was used for the propensity match.

Discussion

Principal Findings

In this retrospective propensity-matched analysis, the standard of care intermittent VS monitoring of adult medical ward patients was associated with nearly 3-fold greater odds of ICU admission or in-hospital mortality as compared with continuous VS monitoring. Our work is novel in that this is, to the best of our knowledge, the first propensity-matched study of continuous VS monitoring in the medical ward. Primary medical admissions are inherently different from surgical ones. Medical patients often have multiple acute problems related to single or multiple intersected disease processes.

Comparison With Prior Work

Our findings corroborate several before-and-after studies mainly conducted with surgical cohorts and are additionally supported by a recently published propensity-matched study by our group in a primary surgical population [28-30]. We hypothesized that, through earlier detection with continuous VS monitoring, patients were able to receive clinical attention sooner, thereby precluding severe clinical deterioration requiring ICU admission.

XSL•FO
RenderX

This hypothesis is supported by similar RRT activation rates but greater odds of ICU admission of patients receiving intermittent VS monitoring.

We report a significant difference in favor of continuous monitoring for the primary composite outcome; however, this composite is only from ICU admissions. When compared with a recently published national database study of unplanned ICU admissions of patients who received standard of care intermittent monitoring, we had a greater percentage of patients in our intermittent monitoring arm transfer to the ICU (2.5% vs 5.8%) [31]. Baseline characteristics, including median CCI, were similar; however, the national study included patients with COVID-19 (2021), while our study examined a prepandemic cohort (2018-2019). Moreover, hospitals in the national database varied significantly from ours in terms of size, teaching status, and regional location, which could have influenced outcomes.

Limitations

Our study was limited primarily by its retrospective approach. First, although propensity matching allowed us to match patients on baseline covariates, we were unable to match based on admission diagnosis or ongoing changes in a patient's hospital course. We were unable to match on admission diagnosis due to limitations in our EHR data (eg, admission diagnosis missing, inaccurate, or incorrect). Changes over the hospital course occurred after receiving the intervention (continuous versus intermittent monitoring); therefore, the changes were not used to match patients. The inability to match admission diagnosis is a major limitation, but we feel this is offset by our hospital's triage system by bed availability, which is agnostic to monitoring strategy. The use of continuous wireless VS monitoring or intermittent VS monitoring was not based on patient condition, but it is possible that, over time, patients admitted to continuous VS monitoring units may have represented a subset of clinically different patients from their intermittent VS monitoring counterparts. That is, patients receiving intermittent VS monitoring may have initially been sicker, or more continuous wireless VS monitoring use may have been deployed over a time course to a healthier group. To account for these differences, cohorts were matched on a combination of baseline covariates present on admission along with every quarter-year of admission to adjust for time variance of continuous wireless VS monitoring use throughout our large hospital system. The difference between the number of patients receiving continuous VS monitoring in 2018 compared with 2019 is related to an issue that occurred with the continuous VS monitoring server. Our approach with propensity matching was set at a caliper of 0.25, and the drop-off for continuously monitored patients does demonstrate that patient groups may be different at baseline within the defined caliper. Most of the imbalance was due to the time of year of admission and CCI, which was matched adequately. An increase in the caliper from 0.25 to 0.35 did not show any significant change in the matched cohort number. Second, we relied on reported rates of conditions from discharge summaries and administrative codes, which have their inherent flaws.

Third, our large administrative data set does not have granular information that would allow a deeper understanding of the mechanisms of favorable outcomes with continuous VS monitoring. This includes the number of and reasons for alarms and false alarms as well as the VS captured by the continuous devices. Similarly, we had limited information on EWS values at the time of RRT activation but did include median, minimum, and maximum EWS and absolute numbers of RRT activations for the entire LOS for both groups. A smaller pilot from the same institution and devices showed a rate of 2.3 alarms per patient per day and a significant decrease in RRT calls, from 189 to 158 per 1000 discharges after implementation using a before-and-after study design [21].

Future Directions

Interesting future outcome research could include whether continuous VS monitoring was associated with changes in medical management, including the timing of antibiotics, discontinuation of supplemental oxygen, and antihypertensive prescriptions. Future work should consider the limitations of this study in order to strengthen the approach, including admission diagnosis matching, granular VS capture, and comparisons between monitoring strategies.

Conclusion

The ability to continuously monitor patient VS outside of the critical care unit or emergency department using a wearable and wireless device represents a significant technological advancement. However, the quality of evidence for the benefit of continuous VS monitoring is low to moderate, and most studies have been performed in the postoperative setting with before-and-after comparisons. This is the first propensity-matched study using a large data set with contemporaneous controls and a focus on patients with primary medical problems. The inherent differences between a primary surgical population, where nearly all continuous postprocedure monitoring research has been conducted, and a primary medical population is paramount to the significance of this work.

In conclusion, patients receiving intermittent VS monitoring had nearly a factor-of-3 increase in the odds of ICU transfer and in-hospital mortality compared with those receiving continuous VS monitoring. Prospective interventional studies are necessary to confirm the impact of continuous VS monitoring on morbidity and mortality.

Acknowledgments

AK is funded with a Clinical and Translational Science Institute National Institutes of Health National Center for Advancing Translational Sciences (KL2 TR001421) award for a trial on continuous postoperative hemodynamic and saturation monitoring. The Department of Anesthesiology at Wake Forest University School of Medicine is supported by Edwards Lifesciences under a master clinical trials agreement and receives grant funding from Masimo and Medtronic.

XSL•FO
RenderX

Authors' Contributions

BR, SW, and DM undertook the primary conceptualization of the study. RB curated the data. AS performed the formal analysis, created the visualizations, and conducted the methodology. BR, VM, and AK wrote the original manuscript draft. BR and AK reviewed and edited the manuscript. WL and AK supervised the study.

Conflicts of Interest

AK consults for Medtronic, Edwards Life Sciences, Philips Research North America, Baxter, GE Healthcare, Potrero Medical, Retia Medical, and Caretaker Medical.

Multimedia Appendix 1
Patient demographics and clinical characteristics of the patients who had no match versus propensity-matched patients for each monitoring strategy from the retrospective propensity matched analysis of continuous versus intermittent vital sign monitoring in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.
[DOCX File , 14 KB - jmir_v27i1e66347_app1.docx]

Multimedia Appendix 2
Absolute values of outcomes measures for the study cohort from the retrospective propensity-matched analysis of continuous versus intermittent vital sign monitoring in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019.
[DOCX File , 13 KB - jmir_v27i1e66347_app2.docx]

Multimedia Appendix 3
Covariate balance plot before and after propensity score matching between the continuous wireless vital sign monitoring cohort and the control monitoring cohort from the retrospective propensity-matched analysis of continuous versus intermittent vital sign monitoring in primary medical patients at an academic tertiary medical center from January 1, 2018, to December 31, 2019. Caliper width of 0.25 was used for the propensity match.
[PNG File , 169 KB - jmir_v27i1e66347_app3.png]

References

1. Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. Crit Care 2019 May 30;23(1):194 [FREE Full text] [doi: 10.1186/s13054-019-2485-7] [Medline: 31146792]
2. Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: a systematic review. PLoS One 2019 Jan 15;14(1):e0210875 [FREE Full text] [doi: 10.1371/journal.pone.0210875] [Medline: 30645637]
3. Eddahchouri Y, Koeneman M, Plokker M, Brouwer E, van de Belt TH, van Goor H, et al. Low compliance to a vital sign safety protocol on general hospital wards: a retrospective cohort study. Int J Nurs Stud 2021 Mar;115:103849 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103849] [Medline: 33517078]
4. Saab R, Wu BP, Rivas E, Chiu A, Lozovoskiy S, Ma C, et al. Failure to detect ward hypoxaemia and hypotension: contributions of insufficient assessment frequency and patient arousal during nursing assessments. Br J Anaesth 2021 Nov;127(5):760-768 [FREE Full text] [doi: 10.1016/j.bja.2021.06.014] [Medline: 34301400]
5. Michard F, Sessler D. Ward monitoring 3.0. Br J Anaesth 2018 Nov;121(5):999-1001 [FREE Full text] [doi: 10.1016/j.bja.2018.07.032] [Medline: 30336872]
6. Sun Z, Sessler D, Dalton J, Devereaux PJ, Shahinyan A, Naylor AJ, et al. Postoperative hypoxemia Is common and persistent: a prospective blinded observational study. Anesth Analg 2015 Sep;121(3):709-715 [FREE Full text] [doi: 10.1213/ANE.0000000000000836] [Medline: 26287299]
7. Turan A, Chang C, Cohen B, Saasouh W, Essber H, Yang D, et al. Incidence, severity, and detection of blood pressure perturbations after abdominal surgery: a prospective blinded observational study. Anesthesiology 2019 Apr;130(4):550-559. [doi: 10.1097/ALN.0000000000002626] [Medline: 30875354]
8. Khanna A, Bergese S, Jungquist C, Morimatsu H, Uezono S, Lee S, PRediction of Opioid-induced respiratory Depression In patients monitored by capnoGraphY (PRODIGY) Group Collaborators. Prediction of opioid-induced respiratory depression on inpatient wards using continuous capnography and oximetry: an international prospective, observational trial. Anesth Analg 2020 Oct;131(4):1012-1024 [FREE Full text] [doi: 10.1213/ANE.0000000000004788] [Medline: 32925318]
9. Franklin C, Mathew J. Developing strategies to prevent inhospital cardiac arrest. Critical Care Medicine 1994;22(2):244-247. [doi: 10.1097/00003246-199402000-00014]
10. Jones AE, Yiannibas V, Johnson C, Kline JA. Emergency department hypotension predicts sudden unexpected in-hospital mortality. Chest 2006 Oct;130(4):941-946. [doi: 10.1016/S0012-3692(15)51124-0]

XSL•FO
RenderX

11. Kenzaka T, Okayama M, Kuroki S, Fukui M, Yahata S, Hayashi H, et al. Importance of vital signs to the early diagnosis and severity of sepsis: association between vital signs and sequential organ failure assessment score in patients with sepsis. Intern Med 2012;51(8):871-876 [FREE Full text] [doi: 10.2169/internalmedicine.51.6951] [Medline: 22504241]
12. Van Leuvan CH, Mitchell I. Missed opportunities? An observational study of vital sign measurements. Critical Care and Resuscitation 2008 Jun;10(2):111-115. [doi: 10.1016/s1441-2772(23)01322-4]
13. Levin N, Horton D, Sanford M, Horne B, Saseendran M, Graves K, et al. Failure of vital sign normalization is more strongly associated than single measures with mortality and outcomes. The American Journal of Emergency Medicine 2020 Dec;38(12):2516-2523. [doi: 10.1016/J.AJEM.2019.12.024]
14. Liem V, Hoeks S, Mol K, Potters JW, Grüne F, Stolker RJ, et al. Postoperative hypotension after noncardiac surgery and the association with myocardial injury. Anesthesiology 2020 Sep;133(3):510-522. [doi: 10.1097/ALN.0000000000003368] [Medline: 32487822]
15. Donaldson LJ, Panesar SS, Darzi A. Patient-safety-related hospital deaths in England: thematic analysis of incidents reported to a national database, 2010-2012. PLoS Med 2014 Jun 24;11(6):e1001667 [FREE Full text] [doi: 10.1371/journal.pmed.1001667] [Medline: 24959751]
16. Taenzer AH, Spence BC. The afferent limb of rapid response systems: continuous monitoring on general care units. Crit Care Clin 2018 Apr;34(2):189-198. [doi: 10.1016/j.ccc.2017.12.001] [Medline: 29482899]
17. Michard F, Kalkman C. Rethinking patient surveillance on hospital wards. Anesthesiology 2021 Sep 01;135(3):531-540. [doi: 10.1097/ALN.0000000000003843] [Medline: 34237129]
18. Areia C, Biggs C, Santos M, Thurley N, Gerry S, Tarassenko L, et al. The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis. Crit Care 2021 Sep 28;25(1):351 [FREE Full text] [doi: 10.1186/s13054-021-03766-4] [Medline: 34583742]
19. Cardona-Morrell M, Prgomet M, Turner RM, Nicholson M, Hillman K. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. Int J Clin Pract 2016 Oct 31;70(10):806-824. [doi: 10.1111/ijcp.12846] [Medline: 27582503]
20. Weller RS, Foard KL, Harwood TN. Evaluation of a wireless, portable, wearable multi-parameter vital signs monitor in hospitalized neurological and neurosurgical patients. J Clin Monit Comput 2018 Oct 6;32(5):945-951. [doi: 10.1007/s10877-017-0085-0] [Medline: 29214598]
21. Khanna AK, O'Connell NS, Ahuja S, Saha AK, Harris L, Cusson BD, et al. Incidence, severity and detection of blood pressure and heart rate perturbations in postoperative ward patients after noncardiac surgery. J Clin Anesth 2023 Oct;89:111159. [doi: 10.1016/j.jclinane.2023.111159] [Medline: 37295123]
22. Rowland B, Motamedi V, Michard F, Saha AK, Khanna AK. Corrigendum to 'Impact of continuous and wireless monitoring of vital signs on clinical outcomes: a propensity-matched observational study of surgical ward patients' (Br J Anaesth 2024; 132: 519-27). Br J Anaesth 2024 Oct;133(4):922-923. [doi: 10.1016/j.bja.2024.07.005] [Medline: 39043523]
23. ViSi Mobile System Technical Specifications. Medaval. 2015 Jul. URL: https://medaval.ie/docs/specs/Sotera-VisiMobile-Specs.pdf [accessed 2025-03-04]
24. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
25. Austin PC. Optimal caliper widths for propensity‐score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical Statistics 2011 Mar 29;10(2):150-161. [doi: 10.1002/PST.433]
26. Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. Pharmacoepidemiol Drug Saf 2012 May 03;21 Suppl 2(S2):69-80. [doi: 10.1002/pds.3263] [Medline: 22552982]
27. Ho D, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. Journal of Statistical Software 2011;42(8):1-28 [FREE Full text] [doi: 10.18637/jss.v042.i08]
28. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. Am J Med 2014 Mar;127(3):226-232. [doi: 10.1016/j.amjmed.2013.12.004] [Medline: 24342543]
29. Eddahchouri Y, Peelen RV, Koeneman M, Touw HR, van Goor H, Bredie SJ. Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study. Br J Anaesth 2022 May;128(5):857-863 [FREE Full text] [doi: 10.1016/j.bja.2022.01.036] [Medline: 35282866]
30. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
31. Khanna A, Moucharite MA, Benefield P, Kaw R. Patient characteristics and clinical and economic outcomes associated with unplanned medical and surgical intensive care unit admissions: a retrospective analysis. CEOR 2023 Sep;Volume 15:703-719. [doi: 10.2147/ceor.s424759]

Abbreviations

AKI: acute kidney injury
BP: blood pressure
CCI: Charlson Comorbidity Index

XSL•FO
RenderX

EHR: electronic health record
EWS: early warning score
HF: heart failure
HR: heart rate
ICU: intensive care unit
IRB: institutional review board
LOS: length of stay
MI: myocardial infarction
RR: respiratory rate
RRT: rapid response team
SpO2: oxygen saturation
VS: vital sign

Edited by A Coristine; submitted 10.09.24; peer-reviewed by M Breteler, C Roman; comments to author 29.11.24; revised version received 13.12.24; accepted 31.01.25; published 11.03.25.

Please cite as:
Rowland B, Saha A, Motamedi V, Bundy R, Winsor S, McNavish D, Lippert W, Khanna AK
Impact on Patient Outcomes of Continuous Vital Sign Monitoring on Medical Wards: Propensity-Matched Analysis
J Med Internet Res 2025;27:e66347
URL: https://www.jmir.org/2025/1/e66347
doi:10.2196/66347
PMID:

©Bradley Rowland, Amit Saha, Vida Motamedi, Richa Bundy, Scott Winsor, Daniel McNavish, William Lippert, Ashish K Khanna. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prevalence of Antibiotic Prescribing for Acute Respiratory Tract Infection in Telehealth Versus Face-to-Face Consultations: Cross-Sectional Analysis of General Practice Registrars' Clinical Practice

Yu Gao[1,2], FRACGP; Parker Magin[1,2,3], PhD; Amanda Tapley[1,2], MMedStat; Elizabeth Holliday[2], PhD; Jason Dizon[4], MMedStat; Katie Fisher[1,2], FRACGP; Mieke van Driel[5], PhD; Joshua S Davis[2], PhD; Andrew Davey[1,2], FRACGP; Anna Ralston[1,2], GradCertClinEpid; Alison Fielding[1,2], PhD; Dominica Moad[1,2], MPH; Katie Mulquiney[1,2], MClinEpid; Lisa Clarke[6], FRACGP; Alexandria Turner[1,2,5], PhD

[1]GP Training Research, Royal Australian College of General Practitioners, Mayfield West, Australia
[2]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia
[3]School of Population Health, University of New South Wales, Kensington, Australia
[4]Clinical Research Design IT and Statistical Support Unit, Hunter Medical Research Institute, New Lambton Heights, Australia
[5]General Practice Clinical Unit, Faculty of Medicine, The University of Queensland, Herston, Australia
[6]GP Training Medical Education, Royal Australian College of General Practitioners, Hobart, Australia

Corresponding Author:
Alexandria Turner, PhD
General Practice Clinical Unit
Faculty of Medicine
The University of Queensland
Level 8, Health Sciences Building, Royal Brisbane & Women's Hospital
Herston, 4029
Australia
Phone: 61 3 9698 0655
Email: alex.turner@racgp.org.au

Abstract

Background: Antimicrobial resistance is a global threat. Australia has high antibiotic prescribing rates with the majority of antibiotics prescribed by general practitioners (GPs) for self-limiting acute respiratory tract infection (ARTIs). Australian GP trainees' (registrars') prescribing for ARTIs may have been affected by the introduction of remunerated telehealth consultations in 2020. Understanding of the impact of telehealth on antibiotic stewardship may inform registrar educational programs.

Objective: This study aimed to compare the prevalence of antibiotic prescribing by GP registrars in telehealth versus face-to-face (F2F) consultations for common cold (upper respiratory tract infection [URTI]), bronchitis, sore throat, acute otitis media, and sinusitis.

Methods: A cross-sectional analysis of data from the Registrar Clinical Encounters in Training (ReCEnT) study, a multicenter inception cohort study of registrars' in-consultation clinical and educational experiences. Analysis used univariable and multivariable logistic regression using 2020-2023 ReCEnT data. The outcome variable was "antibiotic prescribed" for new presentations of URTI, acute sore throat, acute bronchitis, acute sinusitis, and acute otitis media. The study factor was consultation type (telehealth or F2F).

Results: A total of 2392 registrars participated (response rate=93.4%). The proportions of diagnoses that were managed via telehealth were 25% (5283/21384) overall, 19% (641/3327) for acute sore throat, 29% (3733/12773) for URTI, 21% (364/1772), for acute bronchitis, 4.1% (72/1758) for acute otitis media, and 27% (473/1754) for acute sinusitis. Antibiotics were prescribed for 51% (1685/3327) of sore throat diagnoses, 6.9% (880/12773) of URTI diagnoses, 64% (1140/1772) of bronchitis diagnoses, 61% (1067/1754) of sinusitis diagnoses, and 73% (1278/1758) of otitis media diagnoses. On multivariable analysis, antibiotics were less often prescribed in telehealth than F2F consultations for sore throat (adjusted odds ratio [OR] 0.69, 95% CI 0.55-0.86; *P*=.001), URTI (adjusted OR 0.64, 95% CI 0.51-0.81; *P*<.001), and otitis media (adjusted OR 0.47, 95% CI 0.26-0.84; *P*=.01).

XSL•FO
RenderX

There were no significant differences for acute bronchitis (adjusted OR 1.07, 95% CI 0.79-1.45; *P*=.66) or acute sinusitis (adjusted OR 1, 95% CI 0.76-1.32; *P*=.99).

Conclusions: GP registrars are less likely to prescribe antibiotics for sore throat, URTI, and otitis media when seeing patients by telehealth versus F2F. Understanding the reason for this difference is essential to help guide educational efforts aimed at decreasing antibiotic prescribing by GPs for conditions such as ARTIs where they are of little to no benefit. There was no evidence in this study that telehealth consultations were associated with greater registrar antibiotic prescribing for ARTIs. Therefore, there is no deleterious effect on antibiotic stewardship.

(J Med Internet Res 2025;27:e60831) doi:10.2196/60831

KEYWORDS

antimicrobial resistance; antibiotics stewardship; telehealth; general practice; registrars; acute respiratory tract infection; antibiotics; prescription; respiratory tract infection; RTIs; Australia; consultations; teleconsultation; teleconsult; bronchitis; sore throat; acute otitis; sinusitis; in-consultation; upper respiratory tract infection

Introduction

Antimicrobial resistance is a major threat to global health [1]. Australia's antibiotic prescribing rates are much higher than countries of similar socioeconomic status. In 2021, the rate of Australian community antimicrobial use was 17.5 defined daily doses (DDD) per 1000 inhabitants per day, compared with England (15.9 DDD), Canada (11.4 DDD), and the Netherlands (7.9 DDD) [2]. The majority of antibiotics in Australia are prescribed in general practice (family practice). Inappropriate prescribing of antibiotics in outpatient settings, especially in primary care or general practice, is a major driver of antimicrobial resistance [3]. Therefore, antibiotic stewardship in general practice is vital.

In 2022, overall 36.6% of Australians were prescribed at least one antibiotic in the community [2]. The most common reason for prescription of antibiotics in general practice is nonpneumonia, self-limiting acute respiratory tract infections (ARTIs) [4]. ARTIs (common cold [upper respiratory tract infection {URTI}], acute bronchitis, acute sore throat, acute otitis media, and acute sinusitis) are not generally recommended to be treated with antibiotics, irrespective of symptom severity (in the case of URTI and bronchitis) [2,5-7]. Established Australian general practitioners (GPs) have been found to prescribe antibiotics for ARTIs at a higher rate than European GPs [2,8].

Vocational trainees in general practice are a group of particular interest regarding antibiotic stewardship. Trainees are developing and establishing patterns of practice that may be stable over time [9-11], including antibiotic prescribing [12]. In Australia, GP trainees ("GP registrars") comprised 16% of the GP workforce by headcount in 2022-2023 [13]. They have lower antibiotic prescribing rates for ARTIs compared with established Australian GPs [14]. In 2019, established GPs prescribed antibiotics in 36% of acute URTI and 82% of bronchitis cases, whereas registrars prescribed antibiotics in 12% of acute URTI and 72% of bronchitis cases. A similar trend exists for Australian GPs and registrars' prescribing for sore throat, otitis media, and sinusitis [15,16].

Furthermore, in longitudinal analyses between 2010 and 2019, registrars' prescribing for URTI, acute bronchitis, sore throat, sinusitis, and otitis media reduced significantly [16,17]. URTI prescribing decreased from 24% of presentations in 2010 to 12% in 2019, acute bronchitis or bronchiolitis prescribing from 84% to 72%, sore throat from 76% to 60%, otitis media from 88% to 77%, and sinusitis from 84% to 66% [16,17]. While these temporal trends are mirrored in both established GP and registrar populations [15-17] and despite lower prescribing rates than established GPs, registrars' antibiotic prescribing for most ARTIs still exceeds international benchmarks [5,8].

Due to community lockdown and implementation of strict infection prevention during the COVID-19 pandemic, remuneration for Australian GP telehealth consultations was introduced (via the Medicare Benefits Schedule of medical services subsidized by the Australian Government [18]). In the second quarter of 2020, overall 36% of GP consultations in Australia were conducted via telehealth and it is now an established mode of consultation in Australian general practice [19].

Antibiotic prescribing may be affected by consultation modality, that is, remote telehealth consultations as compared with face-to-face (F2F) consultations. It is plausible that clinical uncertainty related to reduced diagnostic confidence consequent upon limited ability to examine patients might drive inappropriate antibiotic prescribing [20,21]. This has implications for antibiotic stewardship. A systemic review and meta-analysis reported that antibiotics were more highly prescribed in telehealth consultations for otitis media and pharyngitis. However, no difference was found in prescribing between modalities for sinusitis, upper respiratory infections, or urinary tract infections [22]. In another systematic review and meta-analysis of 13 studies, antibiotic prescribing for respiratory, urinary, or skin and soft tissue infections in telehealth compared with F2F consultations varied depending on infection type [23]. When the 10 observational studies were pooled, there was significantly less prescribing for sinusitis in telehealth consultations and significantly more prescribing for otitis media in telehealth consultations. Pharyngitis, conjunctivitis, and urinary tract infections had nonsignificantly higher antibiotic prescribing rates in telehealth consultations. There was also no significant difference in prescribing for bronchitis between the 2 modalities [23]. However, the methodological quality of these studies indicates that more research is needed into the effect of consultation modality on antibiotic prescribing rates.

XSL•FO
RenderX

A further systematic review found that studies (published before February 2020) of remote prescribing (telephone, video, internet-based, and text-based consultations) of antibiotics for respiratory tract disease, compared with F2F consultations, had similarly conflicting results [24].

A qualitative study of Australian GP registrars and their clinical supervisors found concerns regarding impaired diagnostic capacity for ARTIs during telehealth consultations and suggested that this could lead to organizing further in-person assessment (arrangement of a F2F follow-up appointment to confirm the diagnosis) or to overprescribing of antibiotics [21]. Overall, there remains a large evidence gap relating to the effects of consultation modality on antibiotic prescribing and the potential reasons for this effect.

The context of, and impetus for, the introduction remunerated telehealth in Australia was the COVID-19 pandemic. An interrupted time-series analysis in England found the overall volume of prescribing of antibiotics primarily used for ARTIs increased between national lockdown in March 2020 and May 2022 [25]. Some Australian studies have found that there was a substantial decrease in dispensing of antimicrobials in Australia in 2020 [2], especially for antibiotics commonly prescribed for ARTIs [26]. A retrospective study also found that the proportion of respiratory tract infections (not further defined) for which antibiotics were prescribed via telehealth was lower than that of F2F consultations at an early stage of the pandemic. However, this percentage progressively increased over time, and by the end of 2021 was almost equivalent to that of F2F consultations [27].

Therefore, there is still some uncertainty regarding whether consultation modality affects antibiotic prescribing rates in primary care settings. GP registrars are a physician demographic of singular interest in this aspect of antibiotic stewardship. They are in a formative stage of their careers when practice patterns are being formed, and once established, GPs may continue to prescribe antibiotics [11,12]. A particular cause for concern related to telehealth during this formative practice period is that clinical uncertainty may drive registrars' inappropriate antibiotic prescribing [20,21]. Given constraints on important diagnostic components (physical examination) during telehealth, it is plausible that widespread uptake of telehealth may be problematic for registrars' antibiotic stewardship.

To the best of our knowledge, no studies have been conducted on the prevalence of GP registrars' or trainees' antibiotic prescribing for respiratory tract infections through telehealth. Understanding antibiotic prescribing patterns in telehealth versus F2F consultations will help inform GP registrar education and training in antimicrobial stewardship.

This study aimed to address this evidence gap by estimating the prevalence of antibiotic prescribing by GP registrars for URTI, bronchitis, sore throat, acute otitis media, and sinusitis during telehealth as compared with F2F consultations. We hypothesized that registrars would prescribe antibiotics in a greater proportion of telehealth consultations for URTI, bronchitis, sore throat, acute otitis media, and sinusitis than prescribed in F2F consultations.

Methods

Study Design

This was a cross-sectional analysis nested within the Registrar Clinical Encounters in Training (ReCEnT) study.

ReCEnT Setting and Participants

ReCEnT is a multicenter inception cohort study of Australian GP registrars' in-consultation clinical and educational experience. The detailed methodology is published elsewhere [28]. ReCEnT is a routine part of the educational program of registrars in participating training regions. Registrars may also elect to provide informed consent for the data to be used for research purposes. During the period reported here, all registrars training in New South Wales, Tasmania, and eastern Victoria (approximately 43% of Australian GP registrars) [29] were eligible for participation.

ReCEnT Data Collection

In ReCEnT, each registrar electronically records details of 60 consecutive consultations, once in each of their first 3 general practice training terms [28]. At the beginning of each collection period, registrars provide demographics and practice data via a questionnaire. In each of the 60 individual consecutive consultations, patient demographics, diagnosis formulations, and clinical and educational actions are recorded. Data are recorded via a dedicated internet-based portal.

For the analyses reported here, data from 7 six-monthly data collection rounds between 2020 and 2023 (following the onset of COVID-19 pandemic with consequent expanded telehealth remuneration for GPs) were used.

Outcome and Study Factor

The outcome factor was "antibiotics prescribed" (yes or no). This was determined using the Anatomical Therapeutic Chemical Classification J01 [30]. The study factor (independent variable of interest) was "consultation modality" (F2F or telehealth). "Telehealth" included both telephone and videoconference consultations.

Independent Variables

Analyses were adjusted for a large range of independent covariates. These included registrar, patient, practice, and consultation variables. The registrar variables were age, gender, training term, full-time, or part-time training status, if qualified as doctor in Australia, and if the registrar had worked at the practice previously.

Patient variables were age group specific to the ARTI problem type (refer to Table S1 in Multimedia Appendix 1), gender, Aboriginal and/or Torres Strait Islander status, and whether the patient was new to the practice or to the registrar.

Practice variables were training region, rurality determined by the Australian Standard Geographical Classification [31], practice size (practices with <6 full-time equivalent doctors are considered "small practices"), socioeconomic status of the practice location (Socioeconomic Index for Areas-Index of Relative Social Disadvantage) [32], and billing policy (whether

XSL•FO
RenderX

the practice routinely bulk bills all patients, ie, accepts the government-provided rebate as payment in full).

Consultation content variables were consultation duration, number of diagnoses managed, and whether registrars sought assistance or information during the consultation (no information source used, supervisor called, and another source used).

Consultation action variables included whether pathology was ordered, whether imaging was ordered, whether referrals were made, whether follow-ups were arranged, and whether learning goals were generated.

Analysis Level

Analyses were conducted at the level of individual diagnosis (rather than at the registrar or consultation level). The diagnoses were restricted to initial presentations of several self-limited and nonpneumonia ARTIs: acute bronchitis, URTI, acute otitis media, acute sore throat, and acute sinusitis. Refer to Table S1 in Multimedia Appendix 1 for the International Classification of Primary Care 2 plus codes used for these diagnoses [33].

Statistical Methods

Participant Characteristics

Overall characteristics of the registrars who participated in this study were calculated at either the registrar level or registrar-round level (each individual registrar contributes up to 3 rounds of data) and reported as n (%) or mean (SD).

Descriptive Analyses

Total diagnoses of the conditions of interest and the proportions of these diagnoses managed by telehealth were calculated. The proportion of telehealth consultations that were undertaken via video compared with phone was calculated.

The proportion of antibiotics prescribed for all new ARTI diagnoses, and individually for new URTIs, acute bronchitis, acute sore throat, acute otitis media, and acute sinusitis, was calculated for telehealth consultations and for F2F consultations. The proportion of types of antibiotics by ARTI diagnoses and consultation modality was also calculated.

Main Analyses

To estimate the association between consultation modality and antibiotic prescription, univariable and multivariable logistic regression models were estimated with the outcome “antibiotic prescribed” for new cases of each of the following: URTI, acute bronchitis, acute sore throat, acute otitis media, and acute sinusitis. Models were estimated within the generalized estimating equations framework to account for repeated measures within registrars. An exchangeable working correlation structure was assumed. An augmented backward selection process was followed to select explanatory variables for inclusion in the final multivariable model for each diagnosis. Variables in the model with *P* values of >.20 were tested for removal. A variable was removed if the resulting model did not have substantively different effect sizes than the previous model (more than 10% different from its value in the previous model). Model fit was assessed using the Hosmer-Lemeshow goodness of fit test. The logistic model assumption of linearity in the log-odds for continuous variables was also checked.

Regressions modeled the log-odds that antibiotics were prescribed for a given presentation. Effects are expressed as odds ratios (ORs) with 95% CIs. Significance was declared at the conventional .05 level, with the magnitude and precision of effect estimates also used to interpret results. Analyses were programmed using STATA (version 16.0; StataCorp) and SAS (version 9.4; SAS Institute).

Post Hoc Analysis

In post hoc analyses, the proportion of registrars’ arrangement of follow-up appointments (with themselves or another GP at the practice) following the index consultation for each ARTI, separately for telehealth and F2F index consultations, was calculated. Differences for telehealth compared with F2F index consultations were tested with chi-square tests.

Missing data were handled using complete case analysis. No imputation was performed to address the missing data (Tables S2-S6 in Multimedia Appendix 2).

Ethical Considerations

Ethics approval was by the University of Newcastle Human Research Ethics Committee (Reference H-2009-0323) and the RACGP National Research and Evaluation Ethics Committee (NREEC-23-0000000161). Registrars provided written informed consent for data routinely collected as part of the educational program to also be used for research purposes. All data were deidentified. No compensation was provided to participants as ReCEnT is a routine part of training.

Results

Participant Characteristics

A total of 2392 registrars (response rate: 93.4%) contributed data from 301,403 consultations, entailing in 425,059 diagnoses. Table 1 provides details of the characteristics of participating registrars.

XSL•FO
RenderX

Table 1. Registrar and practice demographics of participants.

Variables and class	Values
Registrar characteristics (n=2392), n (%)	
Registrar gender	
Woman	1373 (57.9)
Has Australian medical degree, n (%)	
Yes	1880 (79.6)
Year of graduation, mean (SD)	2014.5 (4.9)
Pathway registrar enrolled in (general or rural), n (%)	
General	1557 (66)
Has postgraduate qualifications, n (%)	
Yes	642 (27.2)
College for which seeking fellowship, n (%)	
RACGP[a]	2880 (95.3)
ACRRM[b]	70 (2.9)
Both	10 (0.4)
Registrar-round/practice characteristics (n=5060)	
Registrar age (years), mean (SD)	32.8 (5.9)
Registrar works part-time (rather than full-time), n (%)	
Yes	1169 (24.1)
Registrar training term, n (%)	
Term 1	1839 (36.3)
Term 2	1546 (30.6)
Term 3	1675 (33.1)
Practice routinely bulk bills all patients, n (%)	
Yes	1653 (34.1)
Registrar had worked at practice previously, n (%)	
Yes	1284 (26.5)
Size of practice, n (%)	
Small (≤5 FTE[c] GPs[d])	2004 (41.5)
Rurality of practice, n (%)	
Major city	2873 (56.8)
Inner regional	1876 (37.1)
Outer regional, remote, and very remote	308 (6.1)
SEIFA-IRSD[e] decile of practice, mean (SD)	5.4 (2.8)
SEIFA-IRSD decile of practice, n (%)	
Decile 1	469 (9)
Decile 2	505 (10)
Decile 3	532 (11)
Decile 4	633 (13)
Decile 5	621 (12)
Decile 6	477 (9)
Decile 7	538 (11)

XSL•FO
RenderX

Variables and class	Values
Decile 8	382 (8)
Decile 9	424 (8)
Decile 10	469 (9)

[a]RACGP: Royal Australian College of General Practitioners.

[b]ACRRM: Australian College of Rural and Remote Medicine.

[c]FTE: full-time equivalent.

[d]GP: general practitioner.

[e]SEIFA-IRSD: Socioeconomic Indexes for Areas-Index of Relative Social Disadvantage [32].

Descriptive Findings

Overall, there were 21,384 new diagnoses of our 5 nominated conditions available for analysis. These included 3327, 12,773, 1772, 1754, and 1758 new diagnoses of sore throat, URTI, bronchitis, sinusitis, and otitis media representing 0.7%, 3%, 0.4%, 0.4%, and 0.4%, of all diagnoses, respectively.

The proportions of diagnoses managed via telehealth were as follows: ARTI, 25% (5283/21384); acute sore throat, 19% (641/3327); URTI, 29% (3733/12773); acute bronchitis, 21% (364/1772); acute otitis media; 4.1% (72/1758) and acute sinusitis; and 27% (473/1754). Of the total ARTI consultations managed via telehealth (n=5283), most were managed over the phone (video consult rate: 3.8% [201/5283]). Patient and consultation factors are given in Table 2 for telehealth versus F2F consultations.

For all ARTIs, overall, antibiotics were prescribed in 20% (1035/5283) of telehealth diagnoses compared with 31% (5015/16101) of F2F diagnoses. Refer to Table S7 in Multimedia Appendix 3 for the most commonly prescribed antibiotics for each condition. Refer to Tables S8-S12 in Multimedia Appendix 4 for the characteristics associated with prescribing antibiotics for each ARTI condition.

Table 2. Patient and consultation factors according to consultation mode for all ARTI[a] diagnoses.

Variables and class	Telehealth	Face-to-face
Patient characteristics, n (%)		
Patient gender		
Men	2034 (39)	7197 (45)
Women	3241 (61)	8897 (55)
Patient age (year), mean (SD)	32 (20)	25 (22)
Patient status, n (%)		
Existing patient	1423 (27)	4057 (26)
New to registrar	3748 (71)	10277 (64)
New to practice	112 (2)	1766 (11)
Consultation characteristics, mean (SD)		
Consultation duration (min)	12 (6)	16 (7)
Number of problems managed	1.2 (0.5)	1.3 (0.6)

[a]ARTI: acute respiratory tract infection.

Main Analyses

Examining each of the 5 individual ARTI conditions: the proportions for which antibiotics were prescribed, and the unadjusted and adjusted ORs (from the univariable and multivariable logistic models) for prescribing antibiotics in telehealth, as compared with F2F consultations, are given in Table 3 (refer to Tables S13-S17 in Multimedia Appendix 5 for the full models). The univariable analyses showed reduced odds of prescribing antibiotics via telehealth (compared with F2F) for sore throat (OR 0.62, 95% CI 0.52-0.74; *P*<.001), and URTI (OR 0.62, 95% CI 0.52-0.75; *P*<.001) but greater odds of prescribing for acute bronchitis (OR 1.40, 95% CI 1.11-1.76; *P*=.005), whilst in multivariable analyses, there were statistically significant differences for sore throat (adjusted OR 0.69, 95% CI 0.55-0.86; *P*=.001), URTI (adjusted OR 0.64, 95% CI 0.51-0.81; *P*<.001), and otitis media (adjusted OR 0.47, 95% CI 0.26-0.84; *P*=.01).

XSL•FO
RenderX

Table 3. Univariable and multivariable analyses: antibiotic prescribing for new diagnoses of acute respiratory tract infections during telehealth compared with face-to-face consultations.

Antibiotic prescribed (Yes/No)	All diagnoses (telehealth and face-to-face), n (%)	Diagnoses managed via telehealth, n (%)	Diagnoses managed via face-to-face, n (%)	Uni-variable regression (outcome 'antibiotics prescribed'): odds ratio for 'by telehealth' OR[a] (95% CI)	*P* value	Adjusted regression: (outcome 'antibiotics prescribed'): odds ratio for 'by telehealth' OR (95% CI)	*P* value
Sore throat				0.62 (0.52-0.74)	<.001	0.69 (0.55-0.86)	.001
Yes	1685 (51)	261 (41)	1424 (53)				
No	1642 (49)	380 (59)	1262 (47)				
URTI[b]				0.62 (0.52-0.75)	<.001	0.64 (0.51-0.81)	<.001
Yes	880 (6.9)	178 (4.8)	702 (7.8)				
No	11,893 (93)	3555 (95)	8338 (92)				
Sinusitis				1.10 (0.89-1.38)	.38	1.00 (0.76- 1.32)	.99
Yes	1067 (61)	294 (62)	773 (60)				
No	687 (39)	179 (38)	508 (40)				
Bronchitis				1.40 (1.11-1.76)	.005	1.07 (0.79-1.45)	.66
Yes	1140 (64)	255 (70)	885 (63)				
No	632 (36)	109 (30)	523 (37)				
Otitis media				0.69 (0.42-1.14)	.15	0.47 (0.26-0.84)	.01
Yes	1278 (73)	47 (65)	1231 (73)				
No	480 (27)	25 (35)	455 (27)				

[a]OR: odds ratio.

[b]URTI: upper respiratory tract infection.

Post Hoc Analysis

The percentages of registrars' new ARTI diagnoses with follow-up GP appointments organized are presented in Table 4. There were no statistically significant changes in follow-up, but there was some evidence for more follow-ups following F2F for new otitis media problems (*P*=.06) with clinically significant effect size n/N (11.5%).

XSL•FO
RenderX

Table 4. Percentage of follow-up appointments arranged following consultations for new acute respiratory tract infections.

Follow-up arranged (Yes/No)	Telehealth	Face-to-face	*P* value
Sore throat, n (%)			.40
Yes	259 (40)	1037 (39)	
No	382 (60)	1649 (61)	
URTI[a], n (%)			.23
Yes	1087 (29)	2729 (30)	
No	2646 (71)	6311 (70)	
Sinusitis, n (%)			.84
Yes	152 (32)	418 (33)	
No	321 (68)	863 (67)	
Bronchitis, n (%)			.74
Yes	172 (47)	679 (48)	
No	192 (53)	729 (52)	
Otitis media, n (%)			.06
Yes	28 (39)	849 (50)	
No	44 (61)	837 (50)	

[a]URTI: upper respiratory tract infection.

Discussion

Principal Findings

On multivariable analysis, antibiotics were (statistically and clinically) significantly more likely to be prescribed in F2F consultations than in telehealth consultations for sore throat, URTI, and otitis media problems. For sinusitis and bronchitis, there were no significant differences in prescribing. There were no significant differences in rates of follow-up consultations organized for ARTIs between the initial consultation modalities, though there was some evidence (*P*=.06) for follow-up appointments to be made more frequently following F2F appointments when the diagnosis was acute otitis media.

Comparison to Previous Work

A context for our findings (which are specific to antibiotic prescribing for ARTIs) is that previous studies have reported significantly higher rates of prescribing of medicines in general in F2F consultations compared with telehealth consultations, both by Australian GP registrars [34] and in the wider Australian general practice context [35]. In line with this trend, a recent international publication which compared antibiotic prescribing rates for infections (including but not limited to ARTIs) reported higher prescribing rates in F2F consultations [36].

Our results similarly suggest that antibiotics were more likely to be prescribed in F2F consultations for URTI, otitis media, and sore throat. However, these findings differ from results from 2 systematic reviews and meta-analyses [22,23]. Bakhit et al [23] reported greater antibiotic prescribing for otitis media in telehealth consultations, but no difference between consultation modalities in antibiotic prescribing for pharyngitis. Suzuki et al [22] similarly reported that antibiotics were more frequently prescribed via telehealth for otitis media and pharyngitis. This may be partially due to differences in disease classifications. For example, our classification of sore throat includes, but is not limited to, pharyngitis, whereas both of these systematic reviews examine pharyngitis alone, or group all ARTIs together [22,23]. Furthermore, our study was restricted to a GP registrar population, compared with a wider GP population in these systematic reviews.

Regarding sinusitis and bronchitis, our results are consistent with previous studies that reported no significant difference in antibiotic prescribing rates between consultation modalities for sinusitis [22] or bronchitis [23]. In contrast, some studies have reported that fewer antibiotics were prescribed in telehealth vs F2F consultations for sinusitis [23,37,38]. Previous studies reported that the overall antibiotic prescription rates were lower in telehealth consultations for respiratory tract infections [27], especially in early stages of the COVID-19 pandemic (between 2020 2021), both in Australia [26] and internationally [39,40]. Therefore, it is possible that some of these discrepancies may relate to changes in antibiotic prescribing by modality over time, given that our data are from 2020 to 2023.

One possible explanation for our finding of fewer antibiotic prescribing in telehealth consultations is that registrars organized prompt review appointments (eg, to be able to adequately examine the patient) and deferred antibiotic prescription to the F2F review. In a previous qualitative study, some Australian GPs indicated they were not likely to prescribe antibiotics for acute infections without physically examining the patient (in a follow-up consultation) [21]. However, in our post hoc univariate analyses, we found no significant evidence to support this contention. Similarly, a study looking at the management of sinusitis reported no difference in the follow-up rate between e-visits and F2F consultations [38].

XSL•FO
RenderX

Managing patient expectations may be easier in a telehealth consultation compared with a F2F consultation. A recent qualitative study on the management of acute infections via telehealth reported that some GPs found it easier to deny antibiotics during a telehealth consultation, which could be justified by the lack of physical examination [21]. Conversely, other GPs found themselves prescribing more antibiotics in telehealth consultations as a way to mitigate diagnostic uncertainty [21]. Therefore, while it may be easier to manage patient demand via telehealth, the evidence remains mixed.

Strengths and Limitations

The ReCEnT study has a large sample size, a singularly high response rate for studies of GPs [41], and collects are large amount of registrar, patient and practice variables, which allows for fine-grained adjustment for potential confounding in relationships of factors (eg, consultation mode) with registrars' prescribing. A particular strength is the tight linkage of prescribed medicine with the diagnosis for which it was prescribed.

A further strength is that, in contrast to a number of other studies of this topic, we examined individual clinical presentations (sore throat, acute bronchitis, acute otitis media, URTI, and acute sinusitis) rather than a heterogeneous combination of conditions, such as "acute respiratory infections" [27].

The generalizability of our registrar population results to the wider, established GP population is potentially limited as vocational training is a singular period in GPs' professional lives. However, within the apprenticeship-like model of GP training in Australia, registrar prescribing behavior (including antibiotic prescribing) is strongly influenced by, and reflects, the prescribing of their supervisors and senior GP colleagues [20,42]. This is reflected in the temporal pattern of Australian registrars' antibiotic prescribing mirroring that of established GPs, though at slightly lower rates [15-17]. The generalizability of our findings to vocational training settings in countries beyond Australia is strong in countries with similar apprenticeship-like training structures (eg, the United Kingdom, New Zealand, Ireland, and several European countries). However, generalizability to countries with "residency" training programs, such as in North America [43], is less certain.

This study has several limitations. First, patients' past medical history and the severity of the index ARTI, which could affect patients and practices electing to schedule telehealth or F2F consultations, were not documented. However, antibiotics are not recommended by authoritative Australian evidence-based guidelines for URTI and acute bronchitis [6], irrespective of severity. For conditions such as acute otitis media, sore throat, and acute sinusitis, where antibiotics are indicated in some circumstances and symptom severity may influence both decision-making and the decisions to consult by F2F or telehealth, symptom severity is an unmeasured potential confounder in our analyses. However, it should be noted that symptom severity is not reliably recorded in primary care databases and is infrequently adjusted for in analyses of antibiotic prescribing [8,23,27,36,44].

Second, ReCEnT does not collect data on whether an antibiotic prescription was filled by the patient. However, this limitation is not of significant importance as the aim of the study was to establish registrars' prescribing behavior rather than patients' adherence.

Third, the cross-sectional study design does not allow for implications of causality. However, this study design was appropriate for the aim of estimating the prevalence of antibiotic prescribing by GP registrars for ARTIs during telehealth compared with F2F consultations.

Future Directions

It is possible that there is more clinical uncertainty in telehealth consultations due to the lack of physical examination, communication barriers, and potential miscommunication from poor internet connection or slow speeds [45]. However, concerns that diagnostic uncertainty (especially due to limitations to physical examination) may drive inappropriate antibiotic prescribing for ARTIs and compromise antimicrobial stewardship have not been supported by our findings, which may, in fact, suggest the contrary. Noting the heterogeneity in the literature on this topic, further research is indicated to establish if our findings are particular to registrars, during the COVID-19 and immediately after the COVID-19 period, or if there is unmeasured confounding in our findings.

Our findings also demonstrate the importance of considering individual respiratory tract infections rather than relying solely on analyses of grouped heterogeneous respiratory infections.

Registrars' antibiotic prescribing for sore throat, URTI, and otitis media diagnoses was significantly higher in F2F consultation than in telehealth consultations. However, there were no significant differences in antibiotic prescribing for bronchitis or sinusitis between the 2 consultation modalities. Therefore, we have no evidence from our work to discourage registrars from conducting telehealth consultations for ARTIs on the basis of concerns regarding antibiotic stewardship. Further research is needed to explore the reasons why registrars prescribe antibiotics more highly in F2F consultations than telehealth consultations for sore throat, URTI, and otitis media consultations.

Acknowledgments

We acknowledge the contribution of the GP registrars and the practices of the participating Regional Training Organizations, GP Synergy, Eastern Victoria GP Training, and General Practice Training Tasmania, and of the previous participating Regional Training Providers. In 2023, GP Synergy, Eastern Victoria GP Training, and General Practice Training Tasmania merged to deliver GP Training as part of the Royal Australian College of General Practitioners, funded by the Australian Department of Health. Alexandria Turner is funded through a postdoctoral fellowship from the National Health and Medical Research Council's (NHMRC) Centre for Research Excellence in Minimising Antibiotic Resistance in the Community (CRE-MARC).

XSL•FO
RenderX

Authors' Contributions

Conceptualization: ReCEnT project, PM, MvD, AD; this study, PM, A Turner, YG, and KF.

Data curation: A Tapley, AD, AF, KM, and DM.

Formal analysis: A Tapley, EH, JD, and AR.

Funding acquisition: PM.

Investigation: YG, PM, A Tapley, EH, JD, MvD, JSD, AD, AR, and A Turner.

Methodology: ReCEnT project, PM, MvD, AD; this study, YG, PM, A Tapley, EH, AD, AR, and A Turner.

Project administration: PM, AF, and DM.

Software: A Tapley, EH, and JD.

Supervision: PM, EH, and A Turner.

Visualization: YG, A Tapley, JD, and A Turner.

Writing—original draft: YG, PM, A Tapley, JD, and A Turner.

Writing—review & editing: KF, EH, MvD, JSD, AD, AR, AF, KM, DM, and LC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ICPC2+ code inclusions for each problem/diagnoses.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v27i1e60831_app1.xlsx]

Multimedia Appendix 2
Missing data for antibiotic prescribing for each problem/diagnoses.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v27i1e60831_app2.xlsx]

Multimedia Appendix 3
Most commonly prescribed antibiotics for acute respiratory tract infections by consultation mode.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v27i1e60831_app3.xlsx]

Multimedia Appendix 4
Characteristics associated with prescribing antibiotics for each problem/diagnoses.
[XLSX File (Microsoft Excel File), 27 KB - jmir_v27i1e60831_app4.xlsx]

Multimedia Appendix 5
Univariable and multivariable associations with prescribing antibiotics for each problem/diagnoses.
[XLSX File (Microsoft Excel File), 28 KB - jmir_v27i1e60831_app5.xlsx]

References

1. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 2022;399(10325):629-655 [FREE Full text] [doi: 10.1016/S0140-6736(21)02724-0] [Medline: 35065702]
2. AURA 2023: fifth Australian report on antimicrobial use and resistance in human health. Australian Commission on Safety and Quality in Health Care. 2023. URL: https://tinyurl.com/4r2mc9ce [accessed 2023-11-16]
3. Magin PJ, Morgan S, Tapley A, Henderson KM, Holliday EG, Ball J, et al. Changes in early-career family physicians' antibiotic prescribing for upper respiratory tract infection and acute bronchitis: a multicentre longitudinal study. Fam Pract 2016;33(4):360-367. [doi: 10.1093/fampra/cmw025] [Medline: 27095798]
4. Magin P, Tapley A, Morgan S, Davis JS, McElduff P, Yardley L, et al. Reducing early career general practitioners' antibiotic prescribing for respiratory tract infections: a pragmatic prospective non-randomised controlled trial. Fam Pract 2018;35(1):53-60. [doi: 10.1093/fampra/cmx070] [Medline: 28985369]
5. McCullough AR, Pollack AJ, Hansen MP, Glasziou PP, Looke DF, Britt HC, et al. Antibiotics for acute respiratory infections in general practice: comparison of prescribing rates with guideline recommendations. Med J Aust 2017;207(2):65-69. [doi: 10.5694/mja16.01042] [Medline: 28701117]

XSL•FO
RenderX

6. Electronic therapeutic guidelines complete (eTG complete). Therapeutic Guidelines of Australia. Melbourne, VIC: Therapeutic Guidelines Ltd; 2019. URL: https://tgldcdp.tg.org.au/etgcomplete [accessed 2025-02-27]
7. Sur DKC, Plesa ML. Antibiotic use in acute upper respiratory tract infections. Am Fam Physician 2022;106(6):628-636 [FREE Full text] [Medline: 36521460]
8. Smith DRM, Dolk FCK, Pouwels KB, Christie M, Robotham JV, Smieszek T. Defining the appropriateness and inappropriateness of antibiotic prescribing in primary care. J Antimicrob Chemother 2018;73(suppl_2):ii11-ii18 [FREE Full text] [doi: 10.1093/jac/dkx503] [Medline: 29490061]
9. Shashar S, Ellen M, Codish S, Davidson E, Novack V. Medical practice variation among primary care physicians: 1 decade, 14 health services, and 3,238,498 patient-years. Ann Fam Med 2021;19(1):30-37 [FREE Full text] [doi: 10.1370/afm.2627] [Medline: 33431388]
10. Taylor RJ, Bond CM. Change in the established prescribing habits of general practitioners: an analysis of initial prescriptions in general practice. Br J Gen Pract 1991;41(347):244-248 [FREE Full text] [Medline: 1817480]
11. Davis P, Gribben B, Scott A, Lay-Yee R. Do physician practice styles persist over time? Continuities in patterns of clinical decision-making among general practitioners. J Health Serv Res Policy 2000;5(4):200-207. [doi: 10.1177/135581960000500403] [Medline: 11184955]
12. Björnsdóttir I, Kristinsson KG, Hansen EH. Diagnosing infections: a qualitative view on prescription decisions in general practice over time. Pharm World Sci 2010;32(6):805-814. [doi: 10.1007/s11096-010-9441-6] [Medline: 20931359]
13. General practice workforce providing primary care services in Australia. Australian Government, Department of Health and Aged Care. URL: https://hwd.health.gov.au/resources/data/gp-primarycare.html [accessed 2024-10-28]
14. Davey A, Tapley A, Mulquiney KJ, van Driel M, Fielding A, Holliday E, et al. Immediate and delayed antibiotic prescribing strategies used by Australian early-career GPs: a cross-sectional analysis. Br J Gen Pract 2021;71(713):e895-e903 [FREE Full text] [doi: 10.3399/BJGP.2021.0026] [Medline: 34097641]
15. AURA 2021: fourth Australian report on antimicrobial use and resistance in human health. Australian Commission on Safety and Quality in Health Care. 2021. URL: https://www.safetyandquality.gov.au/publications-and-resources/resource-library/aura-2021-fourth-australian-report-antimicrobial-use-and-resistance-human-health [accessed 2025-02-04]
16. Turner A, van Driel ML, Mitchell BL, Davis JS, Fielding A, Davey A, et al. Temporal patterns of antibiotic prescribing for sore throat, otitis media, and sinusitis: a longitudinal study of general practitioner registrars. Fam Pract 2024;41(3):283-291 [FREE Full text] [doi: 10.1093/fampra/cmad055] [Medline: 37226282]
17. Baillie EJ, Merlo G, Magin P, Tapley A, Mulquiney KJ, Davis JS, et al. Antibiotic prescribing for upper respiratory tract infections and acute bronchitis: a longitudinal analysis of general practitioner trainees. Fam Pract 2022;39(6):1063-1069 [FREE Full text] [doi: 10.1093/fampra/cmac052] [Medline: 35640041]
18. Medicare benefits schedule online. Australian Government, Department of Health and Aged Care. 2024. URL: https://www.mbsonline.gov.au/ [accessed 2024-04-16]
19. Telehealth and coronavirus: medicare benefits schedule (MBS) activity in Australia. The University of Queensland. URL: https://coh.centre.uq.edu.au/telehealth-and-coronavirus-medicare-benefits-schedule-mbs-activity-australia [accessed 2024-04-16]
20. Dallas A, van Driel M, van de Mortel T, Magin P. Antibiotic prescribing for the future: exploring the attitudes of trainees in general practice. Br J Gen Pract 2014;64(626):e561-e567 [FREE Full text] [doi: 10.3399/bjgp14X681373] [Medline: 25179070]
21. Baillie EJ, Merlo G, Biezen R, Boaitey KP, Magin PJ, van Driel ML, et al. Diagnosis and management of acute infections during telehealth consultations in Australian general practice: a qualitative study. BJGP Open 2024;8(1):BJGPO.2023.0142 [FREE Full text] [doi: 10.3399/BJGPO.2023.0142] [Medline: 38092442]
22. Suzuki H, Marra AR, Hasegawa S, Livorsi DJ, Goto M, Perencevich EN, et al. Outpatient antibiotic prescribing for common infections via telemedicine versus face-to-face visits: systematic literature review and meta-analysis. Antimicrob Steward Healthc Epidemiol 2021;1(1):e24 [FREE Full text] [doi: 10.1017/ash.2021.179] [Medline: 36168456]
23. Bakhit M, Baillie E, Krzyzaniak N, van Driel M, Clark J, Glasziou P, et al. Antibiotic prescribing for acute infections in synchronous telehealth consultations: a systematic review and meta-analysis. BJGP Open 2021;5(6):BJGPO.2021.0106 [FREE Full text] [doi: 10.3399/BJGPO.2021.0106] [Medline: 34497096]
24. Han SM, Greenfield G, Majeed A, Hayhoe B. Impact of remote consultations on antibiotic prescribing in primary health care: systematic review. J Med Internet Res 2020;22(11):e23482 [FREE Full text] [doi: 10.2196/23482] [Medline: 33031045]
25. Yang Z, Bou-Antoun S, Gerver S, Cowling TE, Freeman R. Sustained increases in antibiotic prescriptions per primary care consultation for upper respiratory tract infections in England during the COVID-19 pandemic. JAC Antimicrob Resist 2023;5(1):dlad012 [FREE Full text] [doi: 10.1093/jacamr/dlad012] [Medline: 36789176]
26. Gillies MB, Burgner DP, Ivancic L, Nassar N, Miller JE, Sullivan SG, et al. Changes in antibiotic prescribing following COVID-19 restrictions: lessons for post-pandemic antibiotic stewardship. Br J Clin Pharmacol 2022;88(3):1143-1151 [FREE Full text] [doi: 10.1111/bcp.15000] [Medline: 34405427]
27. Imai C, Amin J, Prgomet M, Pearce C, Georgiou A. An increase in antibiotic prescribing for respiratory tract infections through telehealth consultations: retrospective study in Australian general practice. J Med Internet Res 2022;24(10):e40876 [FREE Full text] [doi: 10.2196/40876] [Medline: 36256826]

XSL•FO
RenderX

28. Davey A, Tapley A, van Driel M, Holliday E, Fielding A, Ball J, et al. The registrar clinical encounters in training (ReCEnT) cohort study: updated protocol. BMC Prim Care 2022;23(1):328 [FREE Full text] [doi: 10.1186/s12875-022-01920-7] [Medline: 36527002]
29. Taylor R, Clarke L, Radloff A. Australian General Practice Training Program: National report on the 2020 National Registrar Survey. 2021. URL: https://www.health.gov.au/resources/publications/agpt-program-national-report-on-the-2020-national-registrar-survey?language=en [accessed 2025-02-04]
30. Anatomical therapeutic chemical (ATC) classification. World Health Organization. Oslo; 2025. URL: https://www.who.int/tools/atc-ddd-toolkit/atc-classification [accessed 2025-02-04]
31. Australian standard geographical classification (ASGC). Australian Bureau of Statistics. 2011. URL: https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/32FBEDE1EA4C5800CA25791F000F2E1C/$File/att98dqt.pdf [accessed 2024-05-03]
32. Socio-economic indexes for areas. Autralian Bureau of Statistics. 2016. URL: https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa [accessed 2024-05-03]
33. Britt H. A new coding tool for computerised clinical systems in primary care--ICPC plus. Aust Fam Physician 1997;26 Suppl 2:S79-S82. [Medline: 9254947]
34. Fisher K, Tapley A, Ralston A, Davey A, Fielding A, van Driel M, et al. General practice trainees' telehealth use during the COVID-19 pandemic: a cross-sectional study. Fam Pract 2023;40(5-6):638-647 [FREE Full text] [doi: 10.1093/fampra/cmad022] [Medline: 36882013]
35. Wabe N, Thomas J, Sezgin G, Sheikh MK, Gault E, Georgiou A. Medication prescribing in face-to-face versus telehealth consultations during the COVID-19 pandemic in Australian general practice: a retrospective observational study. BJGP Open 2022;6(1):BJGPO.2021.0132 [FREE Full text] [doi: 10.3399/BJGPO.2021.0132] [Medline: 34819296]
36. Wallman A, Svärdsudd K, Bobits K, Wallman T. Antibiotic prescribing by digital health care providers as compared to traditional primary health care providers: cohort study using register data. J Med Internet Res 2024;26:e55228 [FREE Full text] [doi: 10.2196/55228] [Medline: 38924783]
37. Johnson KM, Dumkow LE, Burns KW, Yee MA, Egwuatu NE. Comparison of diagnosis and prescribing practices between virtual visits and office visits for adults diagnosed with sinusitis within a irimary care network. Open Forum Infect Dis 2019;6(9):ofz393 [FREE Full text] [doi: 10.1093/ofid/ofz393] [Medline: 31660415]
38. Penza KS, Murray MA, Myers JF, Furst JW, Pecina JL. Management of acute sinusitis via e-visit. Telemed J E Health 2021;27(5):532-536. [doi: 10.1089/tmj.2020.0047] [Medline: 32522103]
39. Bara W, Brun-Buisson C, Coignard B, Watier L. Outpatient antibiotic prescriptions in France: patients and providers characteristics and impact of the COVID-19 pandemic. Antibiotics (Basel) 2022;11(5):643 [FREE Full text] [doi: 10.3390/antibiotics11050643] [Medline: 35625287]
40. Barbieri E, Liberati C, Cantarutti A, Di Chiara C, Lupattelli A, Sharland M, et al. Antibiotic prescription patterns in the paediatric primary care setting before and after the COVID-19 pandemic in Italy: an analysis using the aWaRe metrics. Antibiotics (Basel) 2022;11(4):457 [FREE Full text] [doi: 10.3390/antibiotics11040457] [Medline: 35453209]
41. Bonevski B, Magin P, Horton G, Foster M, Girgis A. Response rates in GP surveys - trialling two recruitment strategies. Aust Fam Physician 2011;40(6):427-430. [Medline: 21655493]
42. Deckx L, Anthierens S, Magin PJ, Morgan S, McArthur L, Yardley L, et al. Focus on early-career GPs: qualitative evaluation of a multi-faceted educational intervention to improve antibiotic prescribing. Fam Pract 2018;35(1):99-104. [doi: 10.1093/fampra/cmx074] [Medline: 28985300]
43. Roberts RG, Hunt VR, Kulie TI, Schmidt W, Schirmer JM, Villanueva T, et al. Family medicine training--the international experience. Med J Aust 2011;194(11):S84-S87. [doi: 10.5694/j.1326-5377.2011.tb03135.x] [Medline: 21644860]
44. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. Br J Clin Pharmacol 2010;69(1):4-14 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03537.x] [Medline: 20078607]
45. Fisher K, Davey AR, Magin P. Telehealth for Australian general practice: the present and the future. Aust J Gen Pract 2022;51(8):626-629 [FREE Full text] [doi: 10.31128/AJGP-11-21-6229] [Medline: 35908757]

Abbreviations

ARTI: acute respiratory tract infection
DDD: defined daily doses
F2F: face-to-face
GP: general practitioner
OR: odds ratio
ReCEnT: Registrar Clinical Encounters in Training
URTI: upper respiratory tract infection

XSL•FO
RenderX

Edited by A Schwartz; submitted 22.05.24; peer-reviewed by J Bleidorn, T Wallman, C Cortie; comments to author 11.09.24; revised version received 04.11.24; accepted 22.01.25; published 13.03.25.

Please cite as:
Gao Y, Magin P, Tapley A, Holliday E, Dizon J, Fisher K, van Driel M, Davis JS, Davey A, Ralston A, Fielding A, Moad D, Mulquiney K, Clarke L, Turner A
Prevalence of Antibiotic Prescribing for Acute Respiratory Tract Infection in Telehealth Versus Face-to-Face Consultations: Cross-Sectional Analysis of General Practice Registrars' Clinical Practice
J Med Internet Res 2025;27:e60831
URL: https://www.jmir.org/2025/1/e60831
doi:10.2196/60831
PMID:

©Yu Gao, Parker Magin, Amanda Tapley, Elizabeth Holliday, Jason Dizon, Katie Fisher, Mieke van Driel, Joshua S Davis, Andrew Davey, Anna Ralston, Alison Fielding, Dominica Moad, Katie Mulquiney, Lisa Clarke, Alexandria Turner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Emotion Forecasting: A Transformer-Based Approach

Leire Paz-Arbaizar[1], MSc; Jorge Lopez-Castroman[2,3,4], MD, PhD; Antonio Artés-Rodríguez[1,3,5,6], PhD; Pablo M Olmos[1,5], PhD; David Ramírez[1,5], PhD

[1]Signal Theory and Communications Department, Universidad Carlos III de Madrid, Leganés, Spain

[2]Department of Psychiatry, Public Health, Radiology, Nursing and Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

[3]Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain

[4]Department of Adult Psychiatry, Nimes University Hospital, Nimes, France

[5]Instituto de Investigacion Sanitaria Gregorio Marañón, Madrid, Spain

[6]Evidence-Based Behavior (eB2), Leganés, Spain

Corresponding Author:
Leire Paz-Arbaizar, MSc
Signal Theory and Communications Department
Universidad Carlos III de Madrid
Av. de la Universidad, 30
Leganés, 28911
Spain
Phone: 34 91 624 9157
Email: lpaz@pa.uc3m.es

Abstract

Background: Monitoring the emotional states of patients with psychiatric problems has always been challenging due to the noncontinuous nature of clinical assessments, the effect of the health care environment, and the inherent subjectivity of evaluation instruments. However, mental states in psychiatric disorders exhibit substantial variability over time, making real-time monitoring crucial for preventing risky situations and ensuring appropriate treatment.

Objective: This study aimed to leverage new technologies and deep learning techniques to enable more objective, real-time monitoring of patients. This was achieved by passively monitoring variables such as step count, patient location, and sleep patterns using mobile devices. We aimed to predict patient self-reports and detect sudden variations in their emotional valence, identifying situations that may require clinical intervention.

Methods: Data for this project were collected using the Evidence-Based Behavior (eB2) app, which records both passive and self-reported variables daily. Passive data refer to behavioral information gathered via the eB2 app through sensors embedded in mobile devices and wearables. These data were obtained from studies conducted in collaboration with hospitals and clinics that used eB2. We used hidden Markov models (HMMs) to address missing data and transformer deep neural networks for time-series forecasting. Finally, classification algorithms were applied to predict several variables, including emotional state and responses to the Patient Health Questionnaire-9.

Results: Through real-time patient monitoring, we demonstrated the ability to accurately predict patients' emotional states and anticipate changes over time. Specifically, our approach achieved high accuracy (0.93) and a receiver operating characteristic (ROC) area under the curve (AUC) of 0.98 for emotional valence classification. For predicting emotional state changes 1 day in advance, we obtained an ROC AUC of 0.87. Furthermore, we demonstrated the feasibility of forecasting responses to the Patient Health Questionnaire-9, with particularly strong performance for certain questions. For example, in question 9, related to suicidal ideation, our model achieved an accuracy of 0.9 and an ROC AUC of 0.77 for predicting the next day's response. Moreover, we illustrated the enhanced stability of multivariate time-series forecasting when HMM preprocessing was combined with a transformer model, as opposed to other time-series forecasting methods, such as recurrent neural networks or long short-term memory cells.

Conclusions: The stability of multivariate time-series forecasting improved when HMM preprocessing was combined with a transformer model, as opposed to other time-series forecasting methods (eg, recurrent neural network and long short-term memory), leveraging the attention mechanisms to capture longer time dependencies and gain interpretability. We showed the potential to assess the emotional state of a patient and the scores of psychiatric questionnaires from passive variables in advance. This allows real-time monitoring of patients and hence better risk detection and treatment adjustment.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63962) doi:10.2196/63962

KEYWORDS

affect; emotional valence; machine learning; mental disorder; monitoring; mood; passive data; Patient Health Questionnaire-9; PHQ-9; psychological distress; time-series forecasting

Introduction

Background

The presence of a specific mood status is a necessary criterion for many psychiatric diagnoses, as outlined in the *Diagnostic and Statistical Manual of Mental Disorders* [1], and self-perceived mood is a fundamental component of assessing mental states in psychiatry [2,3]. Therefore, precise monitoring and following of mood conditions play a vital role in mental health care. For instance, both positive and negative mood states, as well as their fluctuations, have demonstrated their predictive value for significant outcomes, such as compulsive overeating episodes in bulimia nervosa, adherence to treatment in bipolar disorder, and opioid use disorders [4,5]. In recent years, technological advancements have facilitated the real-time tracking of individuals' self-reported mood status. One notable advancement is the use of smartphone-delivered ecological momentary assessment (EMA), which allows for the analysis of an individual's experiences, behavior, and emotions as they unfold in their natural environments [6]. However, the effectiveness of this method of mood state evaluation largely depends on the individual's level of self-awareness and ability to interact with the EMA platform. In many cases, psychiatric disorders can cause behavioral changes that decrease the likelihood of individuals engaging with an EMA tool, resulting in missing data. Consequently, a crucial research priority is the development of objective behavioral biomarkers for mood states that can be passively sensed without requiring active involvement from individuals.

By harnessing the power of patients' mobile phones and wearable devices, it has become feasible to gather continuous sensor data in a noninvasive manner, providing valuable insights into their daily activity patterns [7]. These models hold the potential to forecast mental health crises and detect abnormal behavioral patterns, facilitating early intervention [8].

In this study, we examined daily summaries of passively collected behavioral data, treating them as multivariate time series. Passively collected data refer to behavioral information recorded through sensors embedded in devices such as wearables and mobile phones, where patient interaction is not required for data capture. This passive data collection is supplemented by actively provided inputs from patients, such as self-reported emotions or responses to questionnaires tailored to the study context. The dataset used in this research was obtained using the Evidence-Based Behavior (eB2) app, a platform developed by our research group and implemented across multiple studies. These studies encompass a variety of patient cohorts, providing extensive behavioral data complemented by responses to psychological, quality-of-life, and nutrition-related questionnaires, depending on the specific objectives of each study.

For time-series analysis, we used transformer models, which are particularly well suited for capturing long-range dependencies within sequential data [9]. Leveraging the advantages offered by transformer models for time-series analysis, we aimed to uncover underlying patterns in the data collected over time. The attention mechanism in transformers is a fundamental feature that enables the model to prioritize relevant temporal dependencies within the input sequence. By assigning varying weights to different time samples in the sequence, transformers can effectively capture and integrate the most pertinent information for accurate forecasting [10,11]. Furthermore, transformer models exhibit the ability to handle sequences of variable lengths. This flexibility renders transformers scalable to datasets containing many time points, accommodating diverse time-series lengths without sacrificing performance [12]. These attributes position transformers as a promising option for time-series forecasting, as they facilitate the modeling of complex temporal patterns and enhance forecasting accuracy.

Related Work

Overview

This section covers different aspects related to depression diagnosis and tracking, including research in patient emotional state monitoring and the reliability of the Patient Health Questionnaire-9 (PHQ-9) for diagnosis. Moreover, it explores how researchers are using passive observations to obtain standardized and objective insights into patients' self-reported states. Following this, we introduce works that apply machine learning and deep learning methods to analyze such data for enhanced diagnosis accuracy and longitudinal patient tracking. Finally, we provide some background on transformer-type attention models for time-series forecasting.

Depression Diagnosis and Tracking

As numerous studies have shown [13-15], depression remains the worldwide leading cause of disability. However, conventional diagnosis and tracking approaches primarily rely on self-reported depressive symptoms in clinical settings—methodologies that were established over half a century ago. These methods typically entail survey completion or face-to-face interviews, offering limited accuracy, ecological validity, and reliability while imposing significant costs for monitoring and scalability [16]. Moreover, the subjective nature of patient and clinician evaluations, combined with the fluctuating nature of mental health conditions over time, emphasizes the need for ongoing, longitudinal assessments to accurately capture these nuances effectively. In terms of assessing depression, the PHQ-9 has shown to be reliable for the criteria-based diagnosis of this disorder, alongside giving a valid measure of depression severity. These characteristics, plus its brevity, make the PHQ-9 a useful clinical and research tool [17].

XSL•FO
RenderX

A set of studies focused on the relationship between mood variability and some psychiatric disorders. In this work, we used mood in the sense of the subjective variation of the patient's emotional state, as expressed by the patient [18]. Research indicates that mood variability including hypomania, cyclothymia, and hyperthymia has been described in 40% to 50% of patients with depression and that such variability could also characterize anxiety disorders [19,20]. Over the past 2 decades, there has been a surge in research linking various patterns of short-term emotional change to adaptive or abnormal psychological functioning, often with conflicting results [21]. Psychiatric decompensations are characterized by specific patterns of emotional fluctuations across time and provide insight into what constitutes optimal and suboptimal emotional functioning.

Advances in EMA and Passive Monitoring

To prevent bias and capture changes in behavior over time and across different contexts, several researchers proposed a shift from relying solely on global retrospective self-reports collected during research or clinic visits in clinical psychology assessment [22]. In EMA, data are repeatedly collected on an individual's current behaviors and experiences as they happen in their everyday environments [6,23]. EMA helps reduce memory bias, provides more accurate insights into daily life, and allows for the study of small-scale influences on behavior in real-world settings. Technologies used for EMA range from traditional written diaries and phone calls to electronic diaries and physiological sensors [22]. In this case, an EMA-style monitoring is conducted to rely solely on passively collected variables through mobile phone sensors and wearables, excluding all the patient-reported data.

Several studies that reviewed the passive follow-up of patients with different conditions, including bipolar disorder, schizophrenia, and depression, highlighted the potential of passive biomarkers for the monitoring of different types of disorders. Particularly, variables such as accelerometry, location, audio, and use data showed a high general performance [24]. Other studies explored the detection of daily-life behavioral markers through mobile phone GPS and use sensors. They showed that features extracted from these sensors provided markers strongly related to depressive symptom severity [21]. In this line, one of the variables that has shown a correlation with the individual's mental state is daily activity [25,26]. A study focused on detecting emotional state instabilities through passive data found that 3 weeks of continuous, passive recordings were enough to reliably predict mood changes, obtaining average and median errors of mood instability scores within a margin of 5% [27].

Machine Learning and Deep Learning Approaches for Behavioral Data Analysis

Due to the potential demonstrated by these variables in predicting emotional states and their variability in patients, previous works have focused on the application of machine learning and deep learning algorithms to analyze these data [28]. A study by Ghandeharioun et al [13] applied machine learning methods to data on sleep behavior, motion, phone-based communication, location changes, and phone use patterns to impute missing clinical scores from self-reported measures and predict depression severity from these continuous sensor measurements. Similarly, other studies evaluated the performance of random forest and support vector machine classifiers for binary classification of the PHQ-9 score, resulting in 60.1% and 59.1% accuracy, respectively, demonstrating a proof of concept for the detection of depression from passive features [29]. Following this approach, diagnostic meta-analyses have demonstrated the effectiveness of the PHQ-9 for depression screening using mobile devices through various machine learning techniques [30].

Recent studies [30,31] have emphasized the importance of continuous follow-up and the analysis of temporal behavioral patterns in the longitudinal monitoring of patients. Aiming to address the lack of clarity on the temporal scale, specificity, and person-specific nature of the associations between smartphone data and affective symptoms, a study was conducted on smartphone-based passive sensing to identify within- and between-person digital markers of depression and anxiety symptoms over time [31]. Here, hierarchical linear regression models and temporal windows were used to understand the timescale at which sensed features relate to mental health symptoms and explore the predictions in the distal, medial, and proximal times. In line with this, other studies used multilevel modeling to examine the relationships between daily mood and mood variability with symptoms of depression, generalized anxiety, and social anxiety to confirm the empirical evidence linking EMA of mood variability with psychiatric disorders [32]. The findings showed both common and specific emotional dynamics that defined the severity of affective symptoms.

Our research aligned with previous studies by aiming to predict self-reported emotional states and their fluctuations as well as the PHQ-9 scores [33]. However, our methodology differs from conventional approaches by integrating into patients' natural daily routines rather than relying on an experimental setup. Besides, our sample included patients with a variety of disorders who were expected, but not compelled, to actively report data via the app. This setup introduced challenges, notably the substantial presence of missing data, which we addressed using a hidden Markov model (HMM), as documented in previous literature [34]. We build upon this paper, which predicted emotional valence from passive variables, using HMM to handle missing values and classification methods. With the proposed model, we enhanced the emotional valence prediction achieved in this paper and obtained a more reliable prediction as the time horizon expanded. In addition, we included the prediction of the scores of the PHQ-9 questions.

To leverage the continuous acquisition of passive variables and work with these temporal data sequences, this study adopted a transformer-based approach. Transformer models, based on attention mechanisms, offer several advantages for time-series forecasting. The transformer architecture, as introduced by Vaswani et al [10], excels at capturing long-range dependencies, making it well suited for time-series data where distant historical information can be crucial. Attention mechanisms within transformers enable contextual understanding, allowing the model to weigh the significance of past elements, thereby improving forecasting accuracy [10,12].

XSL•FO
RenderX

In addition to the previously mentioned studies, several publications delved into the application of transformers to multivariate time-series data [12,35,36], showcasing the flexibility and adaptability of the transformer architecture in capturing complex temporal relationships. Initially developed for natural language processing tasks, transformers have demonstrated a seamless transition to time-series forecasting due to their inherent capability to model temporal data. This shift underscores the versatility and robustness of transformer-based models in addressing diverse sequential data tasks beyond language processing. Furthermore, pretrained transformer models can be fine-tuned for specific forecasting tasks, leveraging insights from diverse datasets, as demonstrated in various transfer learning applications. Finally, attention weights also contribute to their utility in analyzing and forecasting time-series data in a more interpretable way [37].

To the best of our knowledge, no previous studies have used transformer models specifically for emotion recognition relying solely on behavioral (passive) data from mobile devices. This represents a novel direction in leveraging behavioral data for emotion classification and change detection, particularly in real-world, noninvasive contexts.

In other domains, emotion recognition has been explored using transformer models applied to various data modalities. For text-based emotion recognition, transformer models such as Bidirectional Encoder Representations from Transformers (BERT) and GPT proved to be effective in capturing emotional nuances in textual data. For instance, Xie et al. [38] used GPT (developed by OpenAI) to encode dialogue features, while Li et al. [39] combined Robustly Optimized BERT Pretraining Approach (RoBERTa, developed by Facebook AI) embeddings with wav2vec (version 2.0, developed by Facebook AI) for cross-modal emotion recognition.

For audio-based emotion recognition, speech data have been modeled using architectures such as wav2vec (version 2.0). Chen and Rudnicky [40] fine-tuned wav2vec (version 2.0) for detailed speech emotion analysis, while Sun et al [41] combined it with BERT for multimodal integration.

Transformers have also been used for visual-based emotion recognition, where facial expressions and gestures were analyzed. Praveen et al [42] used the multihead attention mechanism to fuse visual and audio features, effectively capturing both spatial and temporal dynamics.

Multimodal emotion recognition, which combines text, audio, and visual data, has been a common approach. Xie et al [38] proposed a Crossmodality Fusion Transformer, and Zhao et al [43] introduced MEmoBERT for cross-modal emotion classification tasks.

In the field of physiological emotion recognition, transformer-based models such as modality-agnostic transformer based self-supervised learning [44] and conformer [45] have demonstrated high accuracy in analyzing electroencephalogram and electrocardiogram signals for emotion classification.

In contrast to these approaches, our work focused exclusively on behavioral data collected passively through smartphones and wearable devices, without relying on more invasive techniques such as video, voice, or physiological signals such as electroencephalogram. This distinction offered significant advantages. By aligning with natural patient behavior and environments, our method reduced intrusiveness, ensured scalability, and facilitated seamless integration into daily life, making it particularly suitable for real-world applications.

Objectives

The general goal was to acquire objective indicators of patients' conditions and fluctuations in their emotional well-being from passive biomarkers. For this, we focused on predicting the emotional valence of patients as well as the PHQ-9 score. This approach aimed to address the challenge of subjectivity and the absence of continuous monitoring in psychiatry, ultimately aiding in the identification of potentially risky situations. This would facilitate timely intervention and treatment adaptation, thus improving the quality of life of the patients and their environment. We strived to achieve this prediction several days in advance of the actual event. Moreover, the incorporation of attention mechanisms served the additional purpose of advancing our understanding of behavioral patterns.

Methods

Recruitment: Patient Inclusion and Exclusion Criteria and Indications

Participants were eligible for inclusion in the study if they were aged ≥18 years and were diagnosed as clinical outpatients with mental disorders by specialists or if they were attending therapy groups. Among these groups, there were 3 main categories: high suicidal risk, eating disorder, and common mental disorder (CMD). CMD encompasses a group of distress states manifesting with anxiety, depression, and unexplained somatic symptoms that are typically encountered in community and primary care settings [46]. Furthermore, patients from other studies, including cohorts of patients affected by cancer, HIV, obstructive sleep apnea, and cardiac conditions, as well as control patients, were included.

Participants were required to own a smartphone running on Android or iOS operating systems, which they connected to a Wi-Fi network at least once per week. Only participants who provided written informed consent for the eB2 study were included.

Patients received instructions from the clinicians at the beginning of follow-up. For most cohorts, only information regarding the app's functionality was provided. However, 2 groups received specific guidance: patients with eating disorders were required to compulsorily complete meal entries, while patients with CMD were encouraged to regularly log their emotions and periodically complete the PHQ-9. There was no obligation or a set number of required responses for these tasks.

Passive data sources included both mobile and wearable devices. In some studies involving patients with HIV, cancer, or obstructive sleep apnea, patients were provided with wearables. For the remaining studies, if patients had their own wearable, information was extracted from the most reliable data source.

XSL•FO
RenderX

Data

This study was conducted on a sample of 4403 patients from 8 distinct cohorts, each characterized by a different pathology or condition. Out of the 4358 patients, the patient cohorts included individuals with CMD (n=1785, 40.54%), eating disorder (n=1477, 33.54%), high suicidal risk (n=413, 9.38%), cancer (n=84, 1.91%), obstructive sleep apnea (n=48, 1.09%), HIV (n=24, 0.54%), cardiology-related conditions (n=20, 0.45%), as well as control participants (n=507, 11.51%).

The data for this study were derived from secondary analyses of multiple clinical studies conducted in collaboration with various hospitals. All studies were designed to monitor patients with distinct health conditions using the eB2 platform. Despite differences in data sources, standardized protocols for data collection were applied across all studies to ensure consistency in data quality. All users were Spanish or French. Among them, 57.48% (2531/4403) were female, 40.93% (1802/4403) were male, and sex information was missing for the remaining 1.59% (70/4403) of the users. All age groups were well represented, with a mean age of 46 (SD 14.8; range 18-77) years at the start of the measurement period.

Patient monitoring was conducted using the eB2 MindCare app [47,48]. The eB2 app operates by harnessing information from diverse sources within the patient's ecosystem. Using phone sensors, Google Fit (Google LLC), and wearable devices, it acquires data at varying intervals, facilitating a nuanced understanding of the patient's daily activities. In parallel with this passive monitoring, patients had the option to input subjective experiences, sleep patterns, and emotional states throughout the day. Emotional states were cataloged within the app, offering 20 options in a spectrum from anger to delight [49,50].

Daily summaries of these data were the primary focus, although alternative granularities were also considered (such as hours and minutes). Therefore, when predicting sequences or emotions at the next temporal moment, this temporal interval was in days. The data collection period for this study spanned 8 years, from 2016 to 2023, and the average duration of passive activity sequences for patients was 224 days, with an SD of 200 days.

The data used as inputs for training the models were daily summaries of the following passive variables: step count, covered distance, sleep hours, app use, time at home, number of visited places, and practiced sport (a binary variable indicating whether the patient practiced sports during the day). The targets we aimed to predict from these passive data were patients' emotional valence and PHQ-9 scores. The ground truth values for these targets were derived from self-reported emotions and questionnaire answers that patients completed through the app. A more detailed description of the different cohorts is included in Multimedia Appendix 1.

Preprocessing

To ensure the variables were sampled at the same frequency, daily summaries were created for each variable. These daily summaries were derived by aggregating the data according to the specific variable.

Sleep hours, distance, steps, and app use were calculated using passive data collection methods from various sources. Sleep hours were determined based on a prioritization hierarchy: user-entered data were the most prioritized, followed by wearable devices; mobile data; and finally, a sleep estimation model. Daily distance was calculated using GPS signals from mobile devices, collected every 5 minutes. Step counts were gathered at intervals of 1 to 5 minutes, depending on the manufacturers of the smartwatch devices. To calculate daily totals, step data from each slot were merged by priority (wearables over mobile). App use was recorded every 5 minutes, capturing which apps were used and the duration of their use. For these 4 variables, once the data were collected and selected from the priority sources, the data were summed to obtain the daily summary of each variable.

To identify home and work locations, the DenStream algorithm was used. Location data were collected over a 15-day period to form clusters. Once the clusters were established, incoming location data were tagged in real time as either home, work, or other. The cluster definitions were updated every 30 days following the initial 15-day period. On the basis of this clustering process, the time a user spent at home was calculated. Data were collected every 5 minutes and tagged upon entry to indicate the cluster corresponding to home. The total time at home was measured in seconds and aggregated for daily summaries.

The variable practiced sport was a Boolean value indicating whether a user engaged in physical activity lasting at least 15 minutes during the day. This variable was updated whenever new physical activity data were received. Physical activity detection was based on three primary sources: (1) activities automatically labeled by devices (eg, mobile phones and wearables), (2) manually logged activities by users through the provider's app, and (3) activities logged directly into the eB2 MindCare platform.

To address the variations between sensors and data formats, which resulted in anomalies and noise in the information, a preprocessing stage was implemented. This included removing negative values, thresholding the time-related variables to 24 hours, the time step count to 30,000 per day, and the distance to 500 km. Finally, data were standardized over all the patients' sequences (mean 0, SD 1 for input features). Standardization was performed to ensure all variables had a uniform scale, which improved the efficiency and performance of machine learning algorithms. This was particularly important for models sensitive to differences in feature scales, such as neural networks or distance-based methods. Figure 1 shows the sequence of data preprocessing.

XSL•FO
RenderX

Figure 1. The data preprocessing pipeline: data acquisition, obtention of the daily summaries, and standardization of behavioral data. The sequence shown in the daily summaries displays the temporal sequence of passive data for 5 different patients. The intensity of the lines indicates the amount of nonmissing behavioral data the patient has for that day.

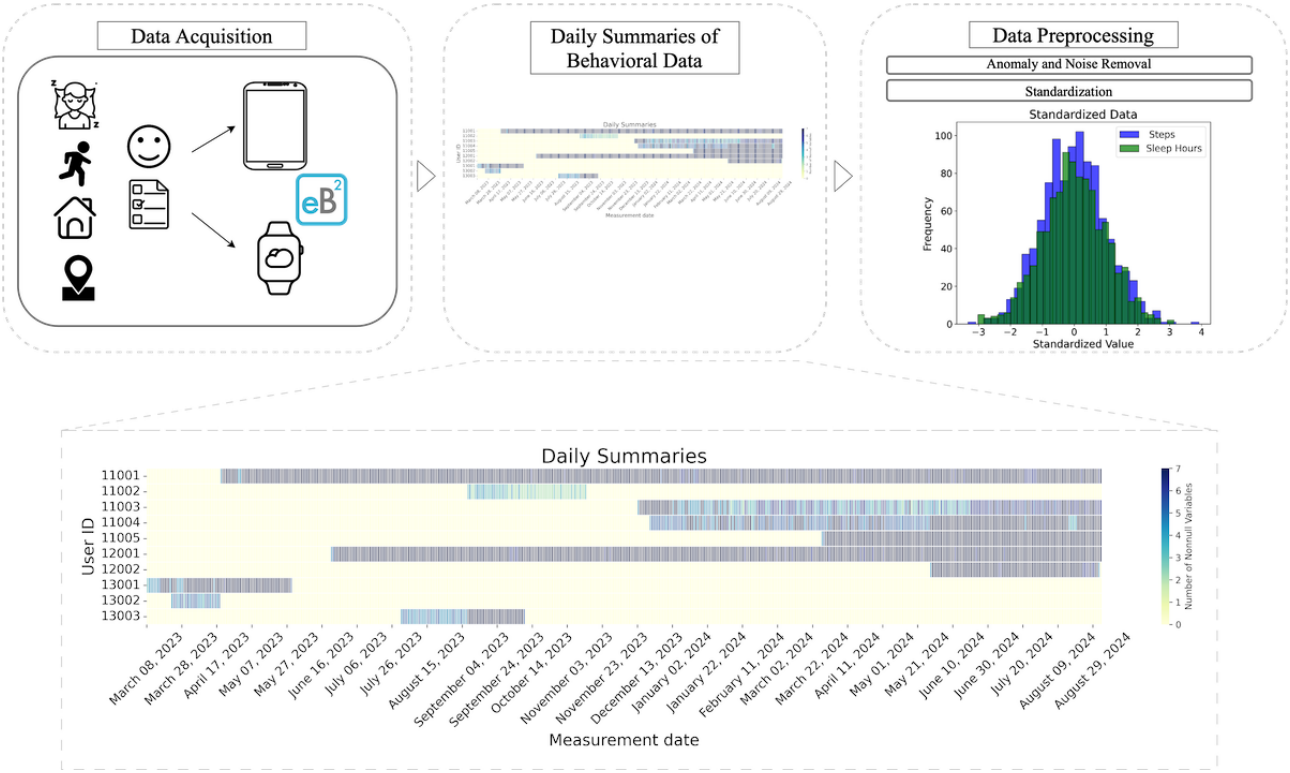

Regarding passive data, the mean percentage of missing data was approximately 60.23% (594,699/986,909), with step count (519,168/986,909, 52.61% of total missing) having the fewest missing values and time at home being the least complete (685,614/986,909, 69.47% total missing). Table 1 shows the percentage of missing values per passive variable grouped by year.

Regarding targets, Table 1 shows the percentage of missing emotion ratio and PHQ-9 answers in comparison with the passive data. It can be observed that the variables reflecting the patients' mood (self-reported emotions and PHQ-9 answers), which they must enter actively, had the lowest percentage of registered data. This is why we aimed to predict these 2 variables using passive data, as they were more complete and allowed for continuous monitoring of the patient's condition without relying on the patients to input the data.

In this study, we handled emotions following the classification scheme by Russell [51], which characterizes emotions in a 2D space. The circumplex model by Russell [52] proposed that emotions could be understood along 2 independent and bipolar dimensions: valence (pleasantness or unpleasantness) and activation (high arousal or low arousal). Independence implied that valence and activation were uncorrelated, while bipolarity suggested that opposite emotions lied at opposite poles of each dimension. For example, happiness and sadness represented opposite ends of the pleasantness spectrum, whereas emotions such as tension and sleepiness were at opposite extremes of the activation dimension.

Using this framework, emotions reported by patients were assigned a level of valence as positive (pleasant), negative (unpleasant), or neutral. For instance, positive valence emotions included happiness, enthusiasm, or satisfaction, while negative valence emotions included sadness, anger, or frustration. Neutral valence emotions fell in the center of the pleasantness spectrum, representing a lack of strong affective polarity [53]. Related studies [54] similarly relied on the model proposed by Russell [52] to categorize emotions, either in 1D using valence or in 2D with both valence and arousal. From this scheme, we focused on emotional valence, which was the first prediction target.

The daily valence was determined by the difference between the counts of positive and negative emotions and can take values between 0 and 2 (negative, neutral, and positive valence).

In addition, the PHQ-9, comprising the 9-item depression module extracted from the full Patient Health Questionnaire, was examined as another target variable. According to this questionnaire, diagnosis of major depression was established if "more than half the days" over the past 2 weeks exhibited the presence of ≥5 depressive symptom criteria, with one of these symptoms being either depressed mood or anhedonia [17]. The responses to the questionnaire were derived from the cohort of patients with CMD, comprising a total of 597 completed surveys. In the context of this investigation, each of the 9 questionnaire items was treated independently. This approach was adopted due to the diverse nature of the questions, which collectively encompassed various facets of the patients' daily experiences. The simultaneous prediction of all scores posed a considerable methodological challenge in this initial endeavor.

XSL•FO
RenderX

Table 1. Percentage of missing daily data for each passive variable[a] by year.

Year	Total number of records per year, n	Steps, n (%)	Distance, n (%)	Sleep, n (%)	App use, n (%)	Time at home, n (%)	Location clusters, n (%)	Emotions ratio, n (%)	PHQ-9[b], n (%)
2016	9548	1243 (13.01)	9548 (100.00)	9511 (99.612)	9548 (100)	9548 (100)	9548 (100)	—[c]	—
2017	19,009	2122 (11.16)	19,009 (100)	18,938 (99.62)	19,009 (100)	19,009 (100)	19,009 (100)	—	—
2018	104,755	63,291 (60.42)	41,528 (39.64)	74,188 (70.82)	94,501 (90.21)	51,521 (49.18)	46,886 (44.76)	—	—
2019	183,919	92,552 (50.32)	72,906 (39.64)	137,889 (74.97)	144,283 (78.45)	93,272 (50.71)	88,998 (48.39)	—	—
2020	139,296	82,317 (59.10)	75,816 (54.43)	10884 (78.14)	105,514 (75.75)	91,684 (65.82)	89,711 (64.4)	—	—
2021	165,154	99,234 (60.09)	121,355 (73.48)	93,075 (56.36)	86,382 (52.3)	128,809 (77.99)	128,276 (77.67)	—	—
2022	239,773	119,631 (49.89)	189,051 (78.85)	128,221 (53.48)	139,454 (58.16)	192,852 (80.43)	191,660 (79.93)	—	—
2023	125,455	58,778 (46.85)	96,824 (77.18)	65,042 (51.84)	69,014 (55.01)	98,919 (78.85)	98,449 (78.47)	—	—
Total	986,909	519,168 (52.60)	626,037 (63.43)	635,705 (64.41)	667,705 (67.66)	685,614 (69.47)	672,537 (68.15)	950,833 (96.34)	967,171 (98.00)

[a]Passive variables include steps, distance, sleep, app use, time at home, location clusters, emotions ratio, and PHQ-9.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]Not available.

Originally, this score comprised 4 classes. However, due to the insufficient number of answers for each type and the frequency-based nature of the responses (ranging from “not at all” to “nearly every day”), the intermediate classes sometimes merged with the 2 extreme groups. Consequently, we decided to classify the answers into 2 broader classes: 0 (representing “not at all” and “several days” as low frequency) and 1 (representing “more than half of the days” and “nearly every day” as high frequency).

Models

Overview

Our global model, illustrated in Figure 2, consisted of 3 main submodels that were trained separately. Hence, we considered that the global model was trained in three stages: (1) training of the HMM to deal with missing fields, (2) the autoregressive transformer to pretrain the model to capture the temporal structure of the data, and (3) the classification model itself.

Initially, the raw data underwent the preprocessing and quality control mentioned previously. To prevent data leakage between the different training phases and biases, the original dataset was divided into 3 subsets for training and validation purposes, with 30% (296,072/986,909), 40% (394,765/986,909), and 30% (296,072/986,909) of the data allocated to the HMM, transformer, and classification layers, respectively, as shown in Figure 2. The partitions were made by assigning equitable percentages of each cohort to each subset to mitigate biases, assigning unique different users to each of them. The following sections describe the 3 main stages in detail.

XSL•FO
RenderX

Figure 2. A graphical abstract of the proposed scheme illustrating the underlying architecture, comprising 3 primary blocks. The first block involves the use of hidden Markov models (HMMs) to address missing data and extract posterior state probabilities. The second block uses a transformer model equipped with an attention mechanism to facilitate pattern recognition and time-series forecasting. Finally, the scheme incorporates a classification layer responsible for patient-reported variables, namely, emotional valence or Patient Health Questionnaire-9 score. eB2: Evidence-Based Behavior; FFN: feed-forward network; t : initial time of the sequence; t _enc: length of the sequence of observed data used as input for the encoder; the encoder is fed with observed data from t to t _enc, in this case, 30 days; t _dec: length of the sequence of observed data used as input for the decoder; the decoder is fed with data from t _enc to t _dec, in this case, for training purposes, the observed sequence consisted of 20 days; tN: forecasting horizon in days, it does not have a fixed value, for example, in this study, N is either 1 or 3, depending on the prediction horizon.

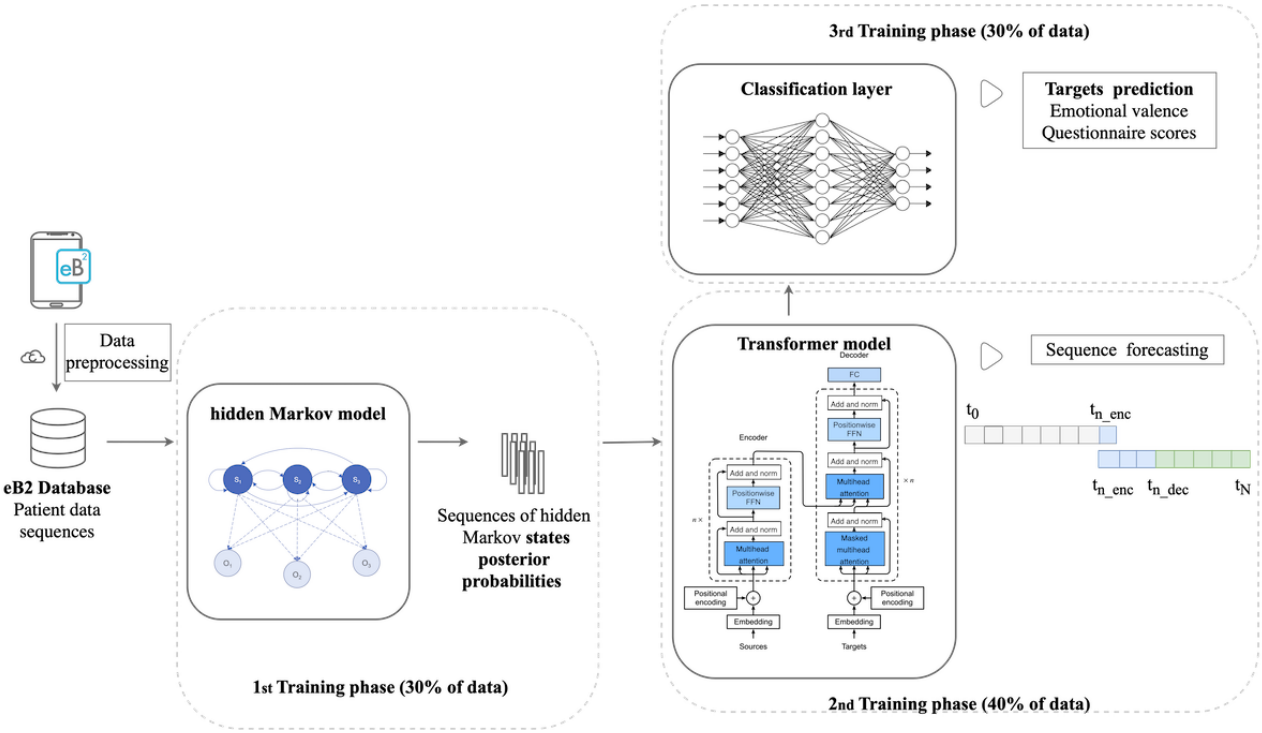

Probabilistic Generative Model for Dealing With Missing Data: HMMs

Our objective was to evaluate the efficacy of transformer models in enhancing time-series forecasting. However, transformer models do not inherently accommodate missing data, presenting a challenge for our analysis. To address this limitation, we used HMMs. In addition, we leveraged the latent space representation provided by HMMs, using the posterior probabilities of the hidden states for training the transformer model.

HMMs, commonly used in time-series analysis, represent a temporal variant of Markov models [55,56]. These generative models are characterized by a collection of observable variables, and a notable advantage lies in their ability to manage missing data without necessitating prior imputation, achieved through marginalization. In the HMM, a sequence of observable variables O is generated by a corresponding sequence of internal hidden states S. However, these hidden states are not observed directly. Instead, transitions between hidden states follow the assumption of a first-order Markov chain. This transition process is defined by a start probability vector π and a transition probability matrix A. In addition, each observable emission is associated with a probability distribution, conditioned on the current hidden state. These emission probabilities are specified by parameter B. Together, these parameters $\lambda=(\pi, A, B)$ fully define the HMM.

The dataset used in this study exhibited heterogeneity, encompassing both categorical variables such as practiced sports as well as continuous variables presumed to take real values (Table 1). For this purpose, we used the heterogenous-HMMs implementation from the HMMs in Python (PyHMM) library, which facilitates the use of various distributions to manage the emission probabilities of each feature type, as depicted in Figure 3 [34,55,57].

In this initial phase, the HMM was trained using 30% (296,072/986,909) of the available data. Once trained, the model was used to infer the state posterior probabilities. At each time step, the set of passive data for a given day and patient was represented by a hidden state posterior probability vector, with a total of 7 hidden states in this instance. Previous studies [34] tested different hidden state configurations, and 7 hidden components were found to effectively capture the underlying patterns in the data. The optimal number of hidden states was determined using the Bayesian Information Criterion and Akaike Information Criterion on a randomly selected subset of sequences with varying levels of missingness [58]. This number of hidden states also led to the best results when a classifier was applied to predict emotions.

XSL•FO
RenderX

Figure 3. Architecture of the heterogeneous hidden Markov model. The model is described by their hidden states sequence (s0:Tn), continuous observations sequence (y0:Tn), and discrete observations sequence (l0:Tn). In our case, (l0:Tn) corresponded to the sequence of discrete observation: practiced sport and (y0:Tn) was the sequence of continuous observations: steps, location distances, sleep time, app use, time at home, and location clusters count.

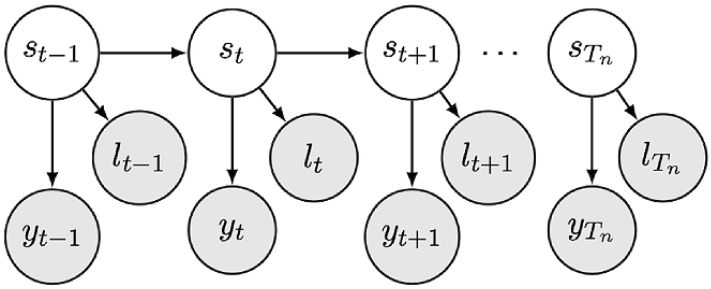

The hidden states' posterior probabilities refer to the probabilities of the hidden states given the observed data. These probabilities are calculated using the observations and the model parameters. The sequence of hidden states probability vectors served as the embedding of the daily information for each patient, enabling the training of subsequent models.

Sequence Forecasting With the Transformer Model

Due to the limited data on emotions and PHQ-9 responses, to obtain a more informative representation of the time series, we first performed a phase of self-supervised training, following a forecasting approach. To achieve this, we used a transformer model for time-series forecasting, leveraging its strengths in handling sequential data. Transformers have shown promising results in capturing long-range dependencies and can provide interpretability regarding the most relevant parts of the sequence for each specific task. Our transformer followed the basic encoder-decoder structure, as depicted in Figure 4 [10], incorporating elements from the encoder of the Informer model [12].

The encoder within the transformer architecture comprised a series of 3 layers. Each layer incorporated 2 distinct sublayers: a multihead self-attention mechanism and a position-wise fully connected feedforward network. Following each sublayer was a residual connection and layer normalization. Likewise, the decoder consisted of 3 layers. In addition to the 2 sublayers present in each encoder layer, the decoder incorporated a third sublayer responsible for performing cross-attention over the output generated by the encoder stack. Residual connections and layer normalization were similarly applied around each sublayer to the encoder. To prevent the decoder from attending to subsequent positions, modifications were made to the self-attention sublayer within the decoder stack, implementing causal (masked) self-attention.

Figure 4. (A) Simplified encoder-decoder architecture of the transformer model. (B) Example of model training and inference. During training, the model learns the parameters to forecast the future state, using the real data as a reference. During inference, the transformer's decoder is responsible for forecasting the future state value in an autoregressive manner. The blue line shows the true states, and the orange line shows those predicted by the model. During inference, the possible future states have associated probabilities, which are illustrated in the graph with different-colored margins around the orange line. Xenc: input sequence for the encoder; Xdec: input sequence for the decoder.

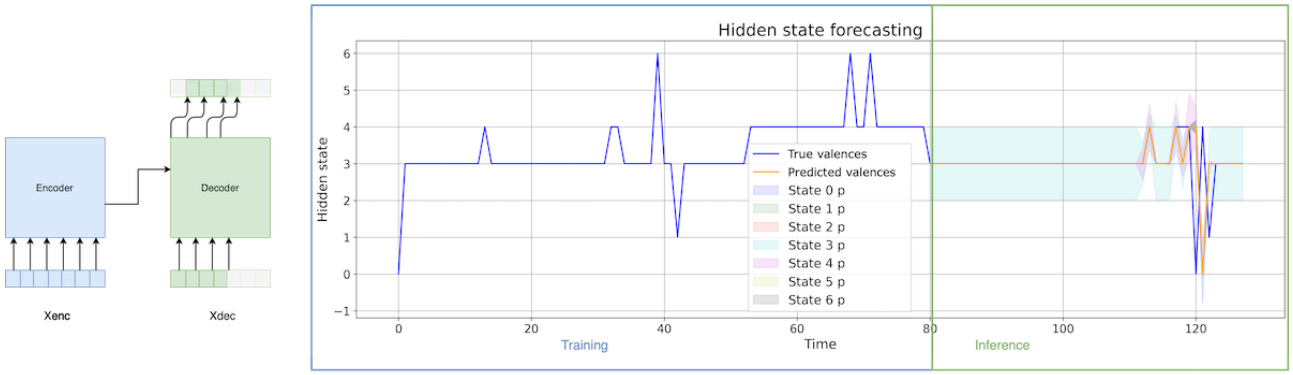

Through the encoder, we obtained contextual information, and with the decoder, we performed the prediction of future observations, considering both the contextual information and the current observations. To incorporate information about the relative or absolute position of tokens in the time series, we combined a positional encoding with the input embeddings at the beginning of both the encoder and decoder stacks [10].

The transformer model was trained from scratch in a forecasting paradigm, wherein the transformer's output was compared to the actual output shifted 1 day ahead. This approach enabled the model to predict future time points with precision. As our dataset consisted of real-world data, it was uncommon for the patients to have extensive sequences of data. Because of this limitation, for model training, we experimented with different sequence lengths, aiming to strike a balance between model performance and sequence duration. Through experimentation with various sequence lengths for both the encoder and decoder, we determined that a sufficiently large sequence (between 25

XSL•FO
RenderX

and 30 days) was required to adequately capture relationships and facilitate accurate forecasting.

For training, we used diverse training schedules, integrating early stopping and dropout techniques to mitigate data overfitting. The loss functions we tested included both mean squared error and mean absolute error, computed between the predicted and actual future sequences. Optimization was carried out using the Adam optimizer across 80 epochs.

A hyperparameter grid was defined, including values for embedding dimensions (32, 64, 128), which represent the dimension of the embedding vector for each data point; number of attention heads (4, 8, 16), indicating the number of heads in each attention layer; number of encoder and decoder layers (2, 3, 4, 6), referring to the attention blocks that include the attention, add and normalization, feedforward, and add and normalization layers; feedforward dimensions (128, 256, 512), which correspond to the dimension of the feedforward layers; dropout rate (0.3, 0.5); and learning rate (0.01, 0.001, 0.0005).

The best-performing model, which we presented in our results, was trained using sequences spanning 50 days of passive data, with 30 days allocated for the encoder and 20 days for the decoder. During inference, using the preceding 30 days of collected data was sufficient for forecasting, as the model could then predict further into the future autoregressively. The optimal model configuration was achieved with an embedding dimension of 32, 4 attention heads, 3 layers, a feedforward dimension of 128, a dropout rate of 0.3, and a learning rate of 0.001. The results obtained for the different combinations of hyperparameters, as well as a more detailed explanation of the used architecture, are included in Multimedia Appendix 2 [10,12,34,57].

Emotional State and PHQ-9 Score Classification

For this section, we used the output of the transformer model to train the subsequent classification layers, aimed at predicting specific targets. The objective was to predict emotional valence and the PHQ-9 scores for the following day.

For the prediction of emotional states, the HMM was first applied to the data sequences to obtain the posterior probabilities of the hidden states of the Markov model, which represented the embeddings of the passive data sequences. Subsequently, forecasting for the following days was conducted using the transformer, which was trained in the previous phase. These outputs were then used to train a model for emotional valence classification.

Several classification models were experimented with, including multilayer perceptron, ensemble models, support vector classifier, and extreme gradient boosting (XGB). Among these, random forest classifiers, as part of the ensemble models, and XGB demonstrated superior performance. This constituted the third training phase, which we evaluated through the F_1-score, accuracy, precision-recall curves, and receiver operating characteristic (ROC) curves, along with their respective area under the curves (AUCs). For each input day, these models provided a probability distribution of the predicted emotional valence for the following day, which in this case was a probability distribution among 3 possible outcomes (0, 1, and 2). In this regard, we analyzed both the probability distributions of the valence predicted by the model and the concrete valence estimated for the following day (the one with the highest probability).

To predict the binary PHQ-9 score for each question, we compared several classifiers by training them with temporal windows of 15, 7, and 3 days using the transformer output (the contextual embeddings), the emotions predicted by the classifier, or a combination of the hidden states' posterior probabilities and the predicted emotions.

Ethical Considerations

Each clinical study received approval from the relevant institutional review board in compliance with ethical standards and the Declaration of Helsinki. Institutional review board approval numbers are indicated in brackets and correspond to the center where approval was obtained for each project and country. Patients at high risk of suicide were identified through collaborations with the Jiménez Díaz Foundation (FJD, EC005-21), Montpellier University Hospital (CPP Ouest IV 20/18_2), and Clínica Nuestra Señora de la Paz. Patients with CMDs were recruited from FJD (PIC148-22), while those with eating disorders were monitored at specialized mental health centers, including Adalmed and Ita mental health clinics. The study also includes patients with cancer monitored in partnership with Gregorio Marañón Hospital (EB2COLON2023), Centro Nacional de Investigaciones Oncológicas, and Fuenlabrada Hospital; patients with HIV/AIDS from Gregorio Marañón (MICRO.HGUGM.2022-002); patients with heart problems from Clínico San Carlos Hospital (19/239-O_P); and patients with obstructive sleep apnea monitored at FJD (PIC163-22). Informed consent was obtained from every participant at the time of inclusion, ensuring adherence to ethical guidelines and participant rights.

Results

Time-Series Forecasting

In this section, we first compare the time-series forecasting capabilities of the transformer and the HMM, without considering specific target predictions, solely as a comparison of their ability to capture the temporal structure and prediction performance of both models.

For the forecasting of the sequences of the posterior probabilities of the hidden states, which represent the information of the passive observations, our results demonstrated enhanced stability over time when comparing the transformer model to the HMM. Specifically, when both models performed autoregressive forecasting, predictions made by the transformer model exhibited greater similarity to the true values as we project further into the future beyond the last observed day.

Figure 5 presents the accuracy of state forecasting, measured as the rate of correctly predicted states for 0 to 7 days into the future. While both models performed similarly for immediate next-day predictions, the transformer's predictions decayed more gradually than those of the HMM as the forecasting horizon extended. The chosen configuration for the transformer

XSL•FO
RenderX

model involved training with sequences of 50 days—30 days for the encoder and 20 days for the decoder—during the initial training phase, encompassing all potential patient sequences. Training with shorter sequences yielded inferior forecasting performance.

Figure 5. Comparison of models in state match forecasting. HMM: hidden Markov model.

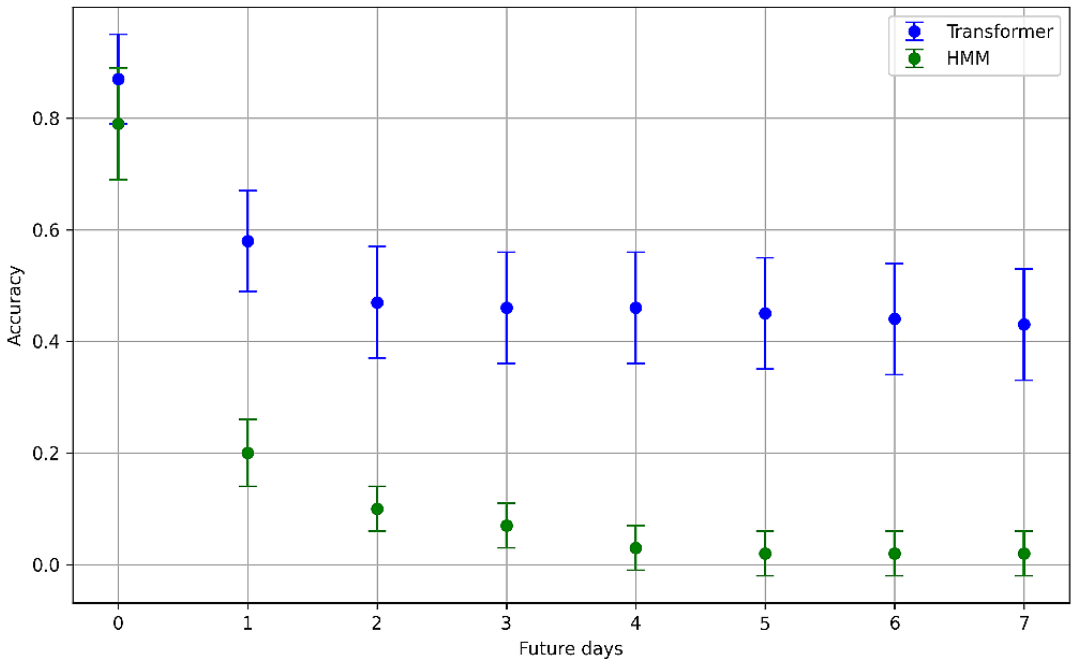

Emotional Valence Forecasting

Regarding emotional valence forecasting, our results focused on predicting emotional state 1 day ahead. The most effective approach involved using the XGB algorithm with a 7-day window of the transformer model decoder as input for the classification layers.

Alternative approaches were explored for the classification. One included incorporating model-embedded vectors of the hidden states, not the final output of the decoder, to potentially enhance informativeness; however, this yielded comparable results to only using the transformer's decoder output. Another strategy involved retraining the entire model (including the transformer and classifier) or some of the layers via fine-tuning, as opposed to solely training the classifier while keeping the transformer parameters frozen to improve task performance. However, the results were similar, and given the time and resource requirements for complete retraining, we opted to train the classification layers for the specific task.

Results across different models used for classification exhibited minor variations, potentially attributed to the embedded vectors effectively capturing necessary information for forecasting, irrespective of the classification model's complexity.

Table 2 shows the results for valence classification. Notably, the XGB model performed best with a ROC AUC of 0.982 and an accuracy of 0.93. Of particular interest was the discrimination of class number 2 (representing a neutral state), which is challenging to classify due to its scarcity. Despite a slight bias toward negative valence, models achieved robust discrimination for negative and positive states and notably good results for the neutral state, accurately predicting events 1 day in advance.

XSL•FO
RenderX

Table 2. Metrics for emotion classification models[a].

Metric	MLP[b]	RF[c]	XGB[d]	XGB retrained	XGB embedding
Precision	0.86	0.87	0.89	*0.89*	0.84
Recall	0.82	*0.86*	*0.86*	0.85	0.8
F_1-score	0.84	*0.87*	*0.87*	0.86	0.8
Accuracy	0.91	0.93	0.93	*0.94*	0.9
AUC[e]	0.97	0.98	*0.98*	0.98	0.92
PR[f]	0.89	0.92	*0.93*	0.9	0.88

[a]For each metric, the results obtained with the best model are italicized.

[b]MLP: multilayer perceptron.

[c]RF: random forest classifier.

[d]XGB: extreme gradient boosting.

[e]AUC: area under the curve.

[f]PR: precision-recall.

Figure 6 illustrates individual patients' emotional series, where the blue line represents the actual emotion for the current day, the orange line represents the forecasted emotion for the following day, and the circles represent the probabilities predicted by the classification model for each emotional valence for the next day. The position and size of the circles correspond to their probability values, with higher probabilities resulting in larger circles. Purple indicates valence 0 (negative), green indicates valence 1 (neutral), and red indicates valence 2 (positive).

Figure 6. Sequences of real emotions (blue) and those predicted by the model (orange) for the following day. The colored circles denote the probability distribution of the emotional state. Purple indicates state 0 (negative valence), green indicates state 1 (neutral), and red indicates state 2 (positive valence). This figure pertains to a patient (ID: 218) selected from a set of patients with high variability of emotion in their temporal sequence.

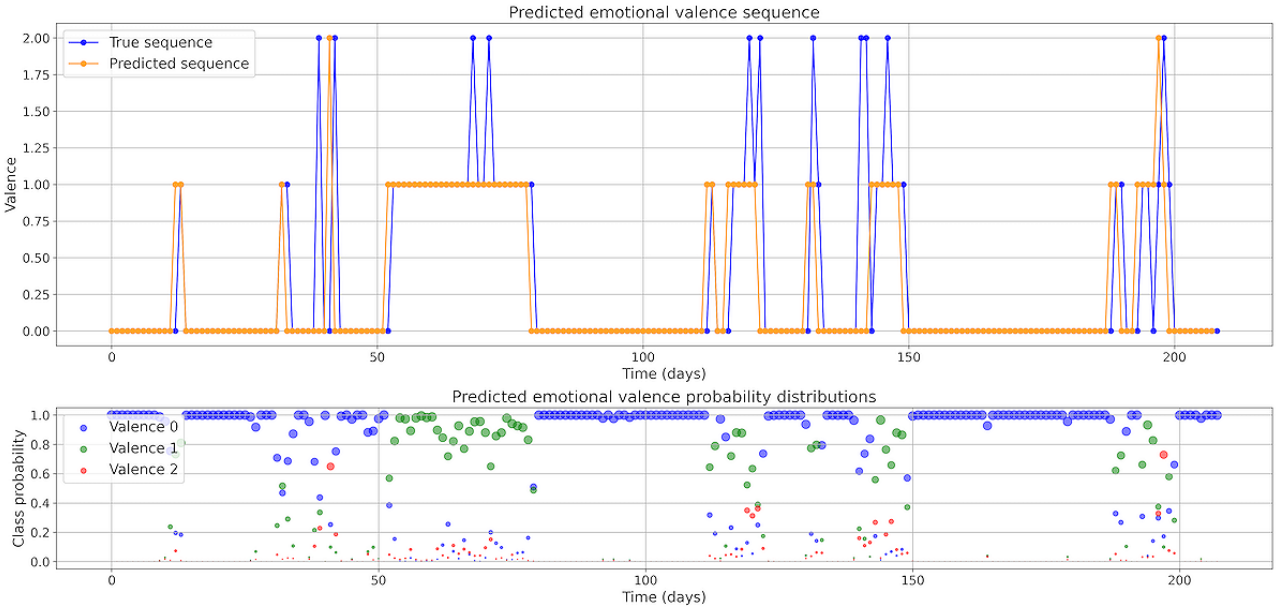

Regarding emotional valence change detection, analyzing the emotional valence distributions revealed a pattern. During stable periods, the distributions were skewed, meaning that most of the probability mass was concentrated around a single valence outcome. However, near a change, the distributions become closer to a uniform distribution. That is, the probability mass was distributed more evenly among the possible outcomes.

To detect the change, we considered calculating the entropy and the Jensen-Shannon (JS) divergence within a temporal window of the sequence. We calculated the entropy of the valences' probability distributions (equation 1) for a 3-day window, including the previous 3 days to each event. The JS divergence (equation 2) was computed between the present-day distribution and an average distribution over a temporal window to determine which approach could better determine the shift in emotional state: whether an increment in disorder within the window or a comparison between the actual and previous distribution.

$$H(x) = -\Sigma\, p_i \log_2 p_i \quad \textbf{(1)}$$

$$D_{JS}(P||Q) = \tfrac{1}{2}\Sigma\,(p_i \log_2 p_i/q_i + q_i \log_2 q_i/p_i) \quad \textbf{(2)}$$

XSL•FO
RenderX

In Figure 7, we display the real changes in emotional valence alongside the JS divergence, and in Figure 8, we display real changes and change detection through entropy. An increase in entropy within a 3-day temporal window, as seen in this case, is associated with a shift in emotional state. Similarly, divergence remains low, close to 0, during stable periods, with most peaks corresponding to real shifts in emotion.

Figure 9 shows the ROC AUC for emotional change detection with entropy and JS divergence. The results of change detection are shown for patients with mental disorders, for patients who have not been diagnosed with mental disorders, and for all study participants to compare detection across different scenarios. For the global case, lower thresholds for both metrics yielded better results, with JS divergence outperforming entropy by 0.13 ROC AUC for change detection.

Figure 7. The Jensen-Shannon (JS) divergence calculated in a temporal window (blue curve) and the true emotion changes (gray vertical lines) for 2 patient sequences.

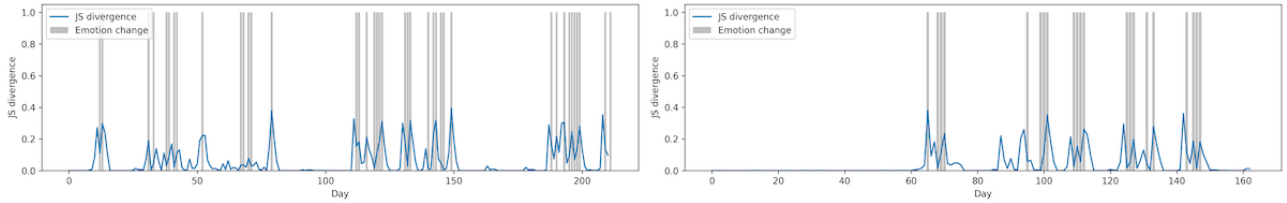

Figure 8. Entropy calculated in a temporal window (blue curve) and the true emotion changes (gray vertical lines) for 2 patient sequences.

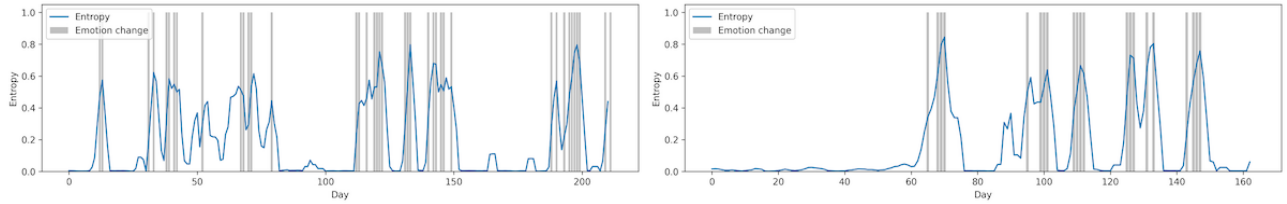

Figure 9. The receiver operating characteristic (ROC) curves for change detection with Jensen-Shannon (JS) divergence and entropy. The curves are obtained for change detection in patients with mental disorders, in patients without any diagnosed mental disorders, and for the entire study cohort (global). AUC: area under the curve; Mental dis.: patients with mental disorders; no mental dis.: patients without any diagnosed mental disorders.

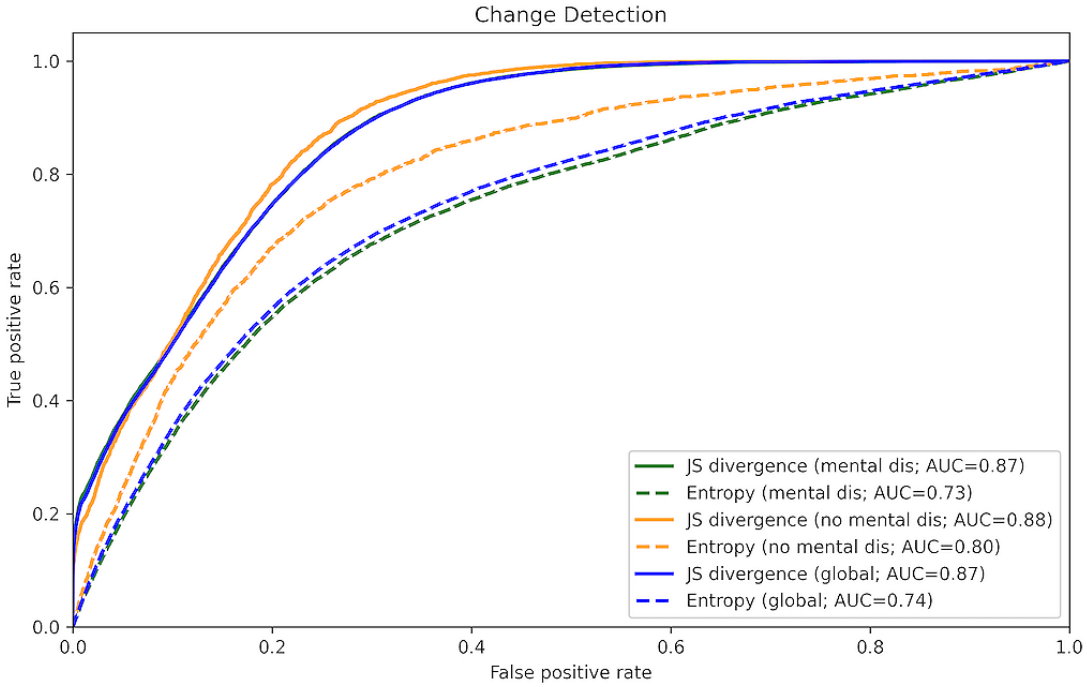

PHQ-9 Score Forecasting

Table 3 presents the accuracy and AUC results for the questionnaire score forecasting, delineated by the classification into 2 classes over a temporal window of 7 days using the decoder transformer outputs. The predictions were obtained using both a random forest classifier and an XGB classifier, both of which demonstrated similar results. Among the results presented in Table 3, the difference lies in the input provided to the classifier: either solely the transformer output (hidden states), the transformer output combined with the corresponding emotional state forecasting, or solely the emotional state. Notably, the results across all 3 configurations were similar,

XSL•FO
RenderX

with slight improvements when using either the hidden states alone or solely the emotion distribution sequence as input.

When comparing different questions, certain inquiries demonstrated higher predictability than others. Notably, question 9 consistently yielded the most accurate predictions across all configurations. This question pertained to whether patients have experienced thoughts of being better off dead or hurting themselves. In addition, question 8, which addresses difficulties in movement or speech that others may notice, exhibited relatively accurate predictions, along with question 5, which was related to appetite, and question 7, which was related to concentration problems.

Table 3. Patient Health Questionnaire-9 (PHQ-9) score forecasting results for questions 1 to 9[a].

PHQ-9 question	Transformer embeddings[b]		Transformer embeddings and emotion[c]		Emotion sequence (3 days)[d]		Emotion sequence (15 days)[d]	
	Accuracy	ROC[e] AUC[f]	Accuracy	ROC AUC	Accuracy	ROC AUC	Accuracy	ROC AUC
1	0.57	0.62	0.57	0.60	*0.59* [g]	*0.63*	0.57	0.61
2	0.6	0.60	0.57	0.59	*0.63*	*0.64*	0.63	0.63
3	0.55	0.51	0.55	0.50	0.55	0.55	*0.58*	*0.62*
4	*0.62*	*0.61*	0.53	0.49	0.53	0.54	0.57	0.57
5	0.54	0.57	0.55	0.60	*0.64*	*0.69*	0.53	0.60
6	0.46	0.46	0.47	0.51	0.6	0.51	0.61	0.65
7	0.52	0.54	0.55	0.54	0.61	0.62	*0.67*	*0.64*
8	0.64	0.52	0.61	0.53	0.68	0.50	*0.72*	*0.56*
9	*0.91*	*0.77*	0.91	0.74	0.91	0.60	0.91	0.74

[a]Results correspond to the classifier performance with different input sets. All the inputs correspond to sequences of different data types (as specified in the columns) before the day with a PHQ-9 record.

[b]Classifier inputs: sequence of embeddings obtained from the transformer.

[c]Classifier inputs: sequence of embeddings obtained from the transformer along with emotion valence prediction obtained from the emotion classifier.

[d]Classifier inputs: the inputs consist exclusively of the emotion predictions obtained from the emotion classifier. Therefore, these columns correspond to sequences of valence predictions of 3 and 15 days before the day with a PHQ-9 entry that is being predicted.

[e]ROC: receiver operating characteristic.

[f]AUC: area under the curve.

[g]For each question, the results obtained with the best model are italicized.

Discussion

Principal Findings

Regarding time-series forecasting, our findings indicated that the transformer model surpassed the HMM in predicting future time steps, resulting in more stable predictions. This suggests that attention mechanisms within the transformer model are effective in capturing longer temporal dependencies, leading to improved prediction stability. Such capabilities are particularly beneficial for assessing a patient's state several days in advance, providing valuable insights into their potential behavior and enabling the early detection of high-risk situations.

The findings concerning emotional changes were generally positive, encompassing both global emotion detection results and change detection 1 day in advance. The variance in outcomes across different machine learning classification algorithms was minimal, highlighting the robustness of the variables and their latent representation. This indicated that accurately predicting a patient's emotional state could be achieved solely through passive variables, with the representation of these variables in the posterior probabilities of the hidden states (obtained with generative models) proving to be informative.

When examining individual patient sequences, it became apparent that in many instances, we could correctly anticipate mood shifts in advance. As for change detection using measures such as entropy or divergence, there appeared to be a correlation between the disorder in emotional valence probability distribution within a specific temporal window and the subsequent emotional state change. Furthermore, global findings indicated that divergence yielded superior results for change detection, suggesting that identifying relative disorder between the current prediction distribution and the past-window distribution holds greater significance than overall disorder within the window, although both contribute to change detection. Similar results were obtained when comparing change detection in patients without any mental disorders and patients with mental disorders. This indicated that the detection was accurate in both cases, with slightly better results for the patients without mental disorders. This can be expected, as the literature suggests that these patients exhibit less fluctuation in their emotions compared to patients with mental disorders.

XSL•FO
RenderX

Regarding the PHQ-9 answers, our model aimed to predict responses with considerable accuracy across both high- and low-frequency classes, allowing for a comprehensive overview of the PHQ-9 outcomes. Upon analyzing differences between answers, those better predicted were typically the ones where patients found it easier to detect their emotions regarding the topic and the frequency of their thoughts. For instance, question 9, corresponding to "Thoughts that you would be better off dead, or of hurting yourself," exhibited the highest classification AUC and accuracy. Conversely, answers such as "Trouble falling or staying asleep, or sleeping too much" and "Feeling bad about yourself or that you are a failure or have let yourself or your family down" yielded poorer results.

Having an approximate score on this questionnaire proved to be a valuable tool, facilitating the monitoring of changes in depressive symptoms over time and guiding treatment decisions. Consequently, it can be used for screening purposes to identify individuals who may require further evaluation for depression.

Limitations

There was a high percentage of missing data in the passive variables, with the minimum percentage of missing data being 52.6% (519,168/986,909) for the number of steps and the maximum percentage being 69.47% (685,614/986,909) for time at home. In addition, there was a very high rate of missing data in the active variables we aimed to predict, such as emotional valence, which had 96.34% (950,833/986,909) missing data.

The recording of emotions was slightly imbalanced, with a higher rate of negative emotions recorded (negative: 16,536/36,076, 45.83%; positive: 10,972/36,076, 30.41%; and neutral: 8568/36,076, 23.74%). Consequently, there was greater sensitivity for detecting negative valence compared to neutral and positive states.

We had limited responses (n=549) for the PHQ-9, and they were sporadic for most of the patients. The mean interval between 2 responses was 25.27 days, with a mode of 14.5 days. On average, each patient responded 4.65 times. Thus, predicting the global score based on passive data was challenging.

Conclusions

Our study yielded several key findings. First, the use of passive variables led to favorable outcomes in both emotion valence detection and change analysis. This underscores the potential of leveraging passive data sources for monitoring and understanding emotional states. Moreover, using temporal methods enabled accurate prediction of emotional states up to a day in advance, with stable results for subsequent days. This temporal stability in prediction highlights the robustness of our approach and suggests its potential applicability in real-world settings where timely intervention is crucial. Further exploration may elucidate the potential for extending prediction horizons beyond a single day, thereby enhancing the utility of our method in long-term monitoring scenarios. Furthermore, our model demonstrates promising performance in predicting PHQ-9 scores, providing an approximate understanding derived solely from passive data. This highlights the utility of this approach as a screening tool for identifying individuals at higher risk of having a crisis or whose depressive symptoms are changing, thereby enabling timely intervention or adjustment in treatment.

Future work should focus on the interpretability of the data. This includes exploring the impact of each variable to assess the patient's condition as well as the effects of removing or adding new variables (such as heart rate and oxygen saturation). In addition, delving deeper into the understanding and extraction of information from the time series could help identify behavioral patterns that determine patient progression as well as pinpoint specific moments that are most relevant for changes in a patient's emotional state, ultimately aiding in adapting the treatment.

While there is room for improvement, particularly in refining the predictive accuracy and expanding the scope of our analysis, our findings represent a significant step forward in the development of in situ support and unobtrusive monitoring strategies for mental health disorders.

Acknowledgments

LP-A was supported by the Spanish Instituto de Salud Carlos III (PMP22/00032) and by the Spanish Ministry of Science, Innovation, and Universities through a doctoral FPU grant (Formación de Profesorado Universitario; FPU23/03723). AA-R was partially supported by the Spanish Instituto de Salud Carlos III (PMP22/00032); Ministerio de Ciencia, Innovación y Universidades (CPP2022- 009537); and Ministerio de Ciencia e Innovación jointly with the European Commission (ERDF, European Regional Development Fund; PID2021-123182OB-I00 and PID2021-125159NB-I00). DR was partially supported by the Office of Naval Research Global (N62909-23-1-2002), the Comunidad de Madrid (IRIS: Hacia un asistente personal para la mejora del bienestar emocional. Ayudas para la realización de Doctorados Industriales en la Comunidad de Madrid 2023; IND2023/TIC-27508), the Spanish Ministry of Economic Affairs and Digital Transformation, and the European Union-NextGenerationEU through the UNICO 5G I+D SORUS project. PMO was supported by the Comunidad de Madrid (IND2022/TIC-23550), ELLIS Unit Madrid (European Laboratory for Learning and Intelligent Systems), and 2024 Leonardo Grant for Scientific Research and Cultural Creation from the BBVA Foundation (Banco Bilbao Vizcaya Argentaria). Both DR and PMO were supported by MICIU/AEI/10.13039/501100011033/FEDER, UE (PID2021-123182OB-I00; EPiCENTER). This work was also supported by the Comunidad de Madrid through the IDEA-CM project (TEC-2024/COM-89).

Conflicts of Interest

AA-R is the founder of Evidence-Based Behavior. All other authors declare no other conflicts of interest.

XSL•FO
RenderX

Multimedia Appendix 1
Cohort overview and patient profiles.
[DOCX File , 15 KB - jmir_v27i1e63962_app1.docx]

Multimedia Appendix 2
Model description and implementation details.
[DOCX File , 35 KB - jmir_v27i1e63962_app2.docx]

References

1. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
2. Panlilio LV, Stull SW, Kowalczyk WJ, Phillips KA, Schroeder JR, Bertz JW, et al. Stress, craving and mood as predictors of early dropout from opioid agonist therapy. Drug Alcohol Depend 2019 Sep 01;202:200-208. [doi: 10.1016/j.drugalcdep.2019.05.026] [Medline: 31357121]
3. Ortiz A, Grof P. Electronic monitoring of self-reported mood: the return of the subjective? Int J Bipolar Disord 2016 Dec;4(1):28 [FREE Full text] [doi: 10.1186/s40345-016-0069-x] [Medline: 27900735]
4. Anestis MD, Selby EA, Crosby RD, Wonderlich SA, Engel SG, Joiner TE. A comparison of retrospective self-report versus ecological momentary assessment measures of affective lability in the examination of its relationship with bulimic symptomatology. Behav Res Ther 2010 Jul;48(7):607-613 [FREE Full text] [doi: 10.1016/j.brat.2010.03.012] [Medline: 20392437]
5. Gershon A, Kaufmann CN, Torous J, Depp C, Ketter TA. Electronic Ecological Momentary Assessment (EMA) in youth with bipolar disorder: demographic and clinical predictors of electronic EMA adherence. J Psychiatr Res 2019 Sep;116:14-18. [doi: 10.1016/j.jpsychires.2019.05.026] [Medline: 31176107]
6. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
7. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. In: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services. 2013 Presented at: MobiSys '13; June 25-28, 2013; Taipei, Taiwan. [doi: 10.1145/2462456.2464449]
8. Garriga R, Mas J, Abraha S, Nolan J, Harrison O, Tadros G, et al. Machine learning model to predict mental health crises from electronic health records. Nat Med 2022 Jun;28(6):1240-1248 [FREE Full text] [doi: 10.1038/s41591-022-01811-5] [Medline: 35577964]
9. Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? Proc AAAI Conf Artif Intell 2023 Jun 26;37(9):11121-11128 [FREE Full text] [doi: 10.1609/aaai.v37i9.26317]
10. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, CA. [doi: 10.5555/3295222.3295349]
11. Catania F, Spitale M, Garzotto F. Conversational agents in therapeutic interventions for neurodevelopmental disorders: a survey. ACM Comput Surv 2023 Feb 02;55(10):1-34. [doi: 10.1145/3564269]
12. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. Proc AAAI Conf Artif Intell 2021 May 18;35(12):11106-11115. [doi: 10.1609/aaai.v35i12.17325]
13. Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In: Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction. 2017 Presented at: ACII 2017; October 23-26, 2017; San Antonio, TX. [doi: 10.1109/acii.2017.8273620]
14. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2024-05-07]
15. Angermann CE, Ertl G. Depression, anxiety, and cognitive impairment: comorbid mental health disorders in heart failure. Curr Heart Fail Rep 2018 Dec 6;15(6):398-410. [doi: 10.1007/s11897-018-0414-8] [Medline: 30402659]
16. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3 [FREE Full text] [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
18. LeDoux JE, Hofmann SG. The subjective experience of emotion: a fearful view. Curr Opin Behav Sci 2018 Feb;19:67-72. [doi: 10.1016/j.cobeha.2017.09.011] [Medline: 39355512]
19. Bowen R, Clark M, Baetz M. Mood swings in patients with anxiety disorders compared with normal controls. J Affect Disord 2004 Mar;78(3):185-192. [doi: 10.1016/s0165-0327(02)00304-x]

XSL•FO
RenderX

20. Bowen R, Baetz M, Hawkes J, Bowen A. Mood variability in anxiety disorders. J Affect Disord 2006 Apr;91(2-3):165-170. [doi: 10.1016/j.jad.2005.12.050] [Medline: 16458367]
21. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175. [doi: 10.2196/jmir.4273] [Medline: 26180009]
22. Ram N, Brinberg M, Pincus AL, Conroy DE. The questionable ecological validity of ecological momentary assessment: considerations for design and analysis. Res Hum Dev 2017 Aug 10;14(3):253-270 [FREE Full text] [doi: 10.1080/15427609.2017.1340052] [Medline: 30613195]
23. Coppersmith DD, Kleiman EM, Millner AJ, Wang SB, Arizmendi C, Bentley KH, et al. Heterogeneity in suicide risk: evidence from personalized dynamic models. Behav Res Ther 2024 Sep;180:104574. [doi: 10.1016/j.brat.2024.104574] [Medline: 38838615]
24. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
25. Dillon CB, McMahon E, O'Regan G, Perry IJ. Associations between physical behaviour patterns and levels of depressive symptoms, anxiety and well-being in middle-aged adults: a cross-sectional study using isotemporal substitution models. BMJ Open 2018 Jan 21;8(1):e018978 [FREE Full text] [doi: 10.1136/bmjopen-2017-018978] [Medline: 29358436]
26. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
27. Morshed MB, Saha K, Li R, D'Mello SK, De Choudhury M, Abowd GD, et al. Prediction of mood instability with passive sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-21 [FREE Full text] [doi: 10.1145/3351233]
28. Kołakowska A, Szwoch W, Szwoch M. A review of emotion recognition methods based on data acquired via smartphone sensors. Sensors (Basel) 2020 Nov 08;20(21):6367 [FREE Full text] [doi: 10.3390/s20216367] [Medline: 33171646]
29. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
30. Kim S, Lee K. Screening for depression in mobile devices using Patient Health Questionnaire-9 (PHQ-9) data: a diagnostic meta-analysis via machine learning methods. Neuropsychiatr Dis Treat 2021 Nov;Volume 17:3415-3430. [doi: 10.2147/ndt.s339412]
31. Stamatis CA, Meyerhoff J, Meng Y, Lin ZC, Cho YM, Liu T, et al. Differential temporal utility of passively sensed smartphone features for depression and anxiety symptom prediction: a longitudinal cohort study. Npj Ment Health Res 2024 Jan 04;3(1):1 [FREE Full text] [doi: 10.1038/s44184-023-00041-y] [Medline: 38609548]
32. Lakhtakia T, Smith SR, Mohr DC, Stamatis CA. Longitudinal associations of daily affective dynamics with depression, generalized anxiety, and social anxiety symptoms. J Affect Disord 2024 May 01;352:437-444. [doi: 10.1016/j.jad.2024.01.250] [Medline: 38286236]
33. Baryshnikov I, Aledavood T, Rosenström T, Heikkilä R, Darst R, Riihimäki K, et al. Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: a prospective ecological momentary assessment study on 80 psychiatric outpatients. J Affect Disord 2023 Mar 01;324:170-174 [FREE Full text] [doi: 10.1016/j.jad.2022.12.127] [Medline: 36586594]
34. Sükei E, Norbury A, Perez-Rodriguez MM, Olmos PM, Artés A. Predicting emotional states using behavioral markers derived from passively sensed data: data-driven machine learning approach. JMIR Mhealth Uhealth 2021 Mar 22;9(3):e24465 [FREE Full text] [doi: 10.2196/24465] [Medline: 33749612]
35. Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. A transformer-based framework for multivariate time series representation learning. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021 Presented at: KDD '21; August 14-18, 2021; Virtual Event, Singapore. [doi: 10.1145/3447548.3467401]
36. Tong J, Xie L, Yang W, Zhang K, Zhao J. Enhancing time series forecasting: a hierarchical transformer with probabilistic decomposition representation. Inf Sci 2023 Nov;647:119410. [doi: 10.1016/j.ins.2023.119410]
37. Chefer H, Gur S, Wolf L. Transformer interpretability beyond attention visualization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021 Presented at: CVPR 2021; June 20-25, 2021; Nashville, TN. [doi: 10.1109/cvpr46437.2021.00084]
38. Xie B, Sidulova M, Park CH. Robust multimodal emotion recognition from conversation with transformer-based crossmodality fusion. Sensors (Basel) 2021 Jul 19;21(14):4913 [FREE Full text] [doi: 10.3390/s21144913] [Medline: 34300651]
39. Li F, Luo J, Xia W. WavFusion: towards wav2vec 2.0 multimodal speech emotion recognition. In: Proceedings of the 31st International Conference on Multimedia Modeling. 2025 Presented at: MMM 2025; January 8-10, 2025; Nara, Japan. [doi: 10.1007/978-981-96-2071-5_24]
40. Chen LW, Rudnicky A. Exploring Wav2vec 2.0 fine tuning for improved speech emotion recognition. In: Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. 2023 Presented at: ICASSP 2023; June 04-10, 2023; Rhodes Island, Greece. [doi: 10.1109/icassp49357.2023.10095036]

XSL•FO
RenderX

41. Sun D, He Y, Han J. Using auxiliary tasks in multimodal fusion of Wav2vec 2.0 and Bert for multimodal emotion recognition. In: Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. 2023 Presented at: ICASSP 2023; June 04-10, 2023; Rhodes Island, Greece. [doi: 10.1109/icassp49357.2023.10096586]
42. Praveen RG, Carneiro de Melo W, Ullah N, Aslam H, Zeeshan O, Denorme T, et al. A joint cross-attention model for audio-visual fusion in dimensional emotion recognition. arXiv Preprint posted online on March 28, 2022 [FREE Full text] [doi: 10.1109/cvprw56347.2022.00278]
43. Zhao J, Li R, Jin Q, Wang X, Li H. Memobert: pre-training model with prompt-based learning for multimodal emotion recognition. In: Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. 2022 Presented at: ICASSP 2022; May 23-27, 2022; Singapore, Singapore. [doi: 10.1109/icassp43922.2022.9746910]
44. Goetz T, Arora P, Erick FX, Holzer N, Sawant S. Self-supervised representation learning using multimodal transformer for emotion recognition. In: Proceedings of the 8th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence. 2023 Presented at: iWOAR '23; September 21-22, 2023; Lübeck, Germany. [doi: 10.1145/3615834.3615837]
45. Gulati A, Qin J, Chiu CC, Parmar N, Zhang Y, Yu J, et al. Conformer: convolution-augmented transformer for speech recognition. arXiv Preprint posted online on May 16, 2020 [FREE Full text] [doi: 10.21437/interspeech.2020-3015]
46. Risal A. Common mental disorders. Kathmandu Univ Med J (KUMJ) 2011 Jun 13;9(35):213-217. [doi: 10.3126/kumj.v9i3.6308] [Medline: 22946143]
47. Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Carmona Camacho R, Baca-García E, et al. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the evidence-based behavior (eB2) study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e197 [FREE Full text] [doi: 10.2196/mhealth.9472] [Medline: 30530465]
48. Carretero P, Campana-Montes JJ, Artes-Rodriguez A. Ecological momentary assessment for monitoring risk of suicide behavior. Curr Top Behav Neurosci 2020;46:229-245. [doi: 10.1007/7854_2020_170] [Medline: 32797403]
49. Evidence-based behavior homepage. Evidence-Based Behavior. URL: https://eb2.tech/?lang=en [accessed 2024-02-22]
50. Bonilla-Escribano P, Ramirez D, Sedano-Capdevila A, Campana-Montes JJ, Baca-Garcia E, Courtet P, et al. Assessment of e-social activity in psychiatric patients. IEEE J Biomed Health Inform 2019 Nov;23(6):2247-2256. [doi: 10.1109/JBHI.2019.2918687] [Medline: 31135374]
51. Russell JA. Core affect and the psychological construction of emotion. Psychol Rev 2003;110(1):145-172. [doi: 10.1037/0033-295X.110.1.145]
52. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980 Dec;39(6):1161-1178. [doi: 10.1037/h0077714]
53. Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 2005;17(3):715-734 [FREE Full text] [doi: 10.1017/S0954579405050340] [Medline: 16262989]
54. Mitsios M, Vamvoukakis G, Maniati G, Ellinas N, Dimitriou G, Markopoulos K, et al. Improved text emotion prediction using combined valence and arousal ordinal classification. arXiv Preprint posted online on April 2, 2024 [FREE Full text] [doi: 10.18653/v1/2024.naacl-short.72]
55. Bishop CM. Pattern Recognition and Machine Learning. Cham, Switzerland: Springer; 2006.
56. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989;77(2):257-286. [doi: 10.1109/5.18626]
57. Moreno-Pino F, Sükei E, Olmos PM, Artés-Rodríguez A. PyHHMM: a python library for heterogeneous hidden Markov models. arXiv Preprint posted online on January 12, 2022 [FREE Full text] [doi: 10.1201/9781003226581-6]
58. Dridi N, Hadzagic M. Akaike and Bayesian information criteria for hidden Markov models. IEEE Signal Process Lett 2019 Feb;26(2):302-306. [doi: 10.1109/lsp.2018.2886933]

Abbreviations

AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
CMD: common mental disorder
eB2: Evidence-Based Behavior
EMA: ecological momentary assessment
HMM: hidden Markov model
JS: Jensen-Shannon
PHQ-9: Patient Health Questionnaire-9
PyHMM: hidden Markov models in Python
RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach
ROC: receiver operating characteristic
XGB: extreme gradient boosting

XSL•FO
RenderX

Edited by F Xie; submitted 05.07.24; peer-reviewed by R Wu, S Kakuba, Y Wang, J Devaraj, B Ozek, F Yu; comments to author 01.11.24; revised version received 25.12.24; accepted 16.01.25; published 18.03.25.

Please cite as:
Paz-Arbaizar L, Lopez-Castroman J, Artés-Rodríguez A, Olmos PM, Ramírez D
Emotion Forecasting: A Transformer-Based Approach
J Med Internet Res 2025;27:e63962
URL: https://www.jmir.org/2025/1/e63962
doi:10.2196/63962
PMID:

©Leire Paz-Arbaizar, Jorge Lopez-Castroman, Antonio Artés-Rodríguez, Pablo M Olmos, David Ramírez. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Application of Internet Hospitals in the Disease Management of Patients With Ulcerative Colitis: Retrospective Study

Tianzhi Yu[1*], PHD; Wanyu Li[2*], BS; Yingchun Liu[3*], MS; Chunjie Jin[1], MD; Zimin Wang[2], BS; Hailong Cao[2], MD, PHD

[1]Internet Hospital, Tianjin Medical University General Hospital, Tian Jin, China

[2]Department of Gastroenterology, National Key Clinical Specialty, Tianjin Medical University General Hospital, Tianjin, China

[3]School of Management, Tianjin University of Traditional Chinese Medicine, Tian Jin, China

[*]these authors contributed equally

Corresponding Author:
Hailong Cao, MD, PHD
Department of Gastroenterology
National Key Clinical Specialty
Tianjin Medical University General Hospital
154 Anshan Road in Heping District
Tianjin, 300052
China
Phone: 86 +86 022 6036155
Email: caohailong@tmu.edu.cn

Abstract

Background: Ulcerative colitis (UC) is a chronic disease characterized by frequent relapses, requiring long-term management and consuming substantial medical and social resources. Effective management of UC remains challenging due to the need for sustainable remission strategies, continuity of care, and access to medical services. Intelligent diagnosis refers to the use of artificial intelligence–driven algorithms to analyze patient-reported symptoms, generate diagnostic probabilities, and provide treatment recommendations through interactive tools. This approach could potentially function as a method for UC management.

Objective: This study aimed to analyze the diagnosis and treatment data of UC from both physical hospitals and internet hospitals, highlighting the potential benefits of the intelligent diagnosis and treatment service model offered by internet hospitals.

Methods: We collected data on the visits of patients with UC to the Department of Gastroenterology at Tianjin Medical University General Hospital. A total of 852 patients with UC were included between July 1, 2020, and June 31, 2023. Statistical methods, including chi-square tests for categorical variables, *t* tests for continuous variables, and rank-sum tests for visit numbers, were used to evaluate the medical preferences and expenses of patients with UC.

Results: We found that internet hospitals and physical hospitals presented different medical service models due to the different distribution of medical needs and patient groups. Patients who chose internet hospitals focused on disease consultation and prescription medication (3295/3528, 93.40%). Patients' medical preferences gradually shifted to web-based services provided by internet hospitals. Over time, 58.57% (270/461) of patients chose either web-based services or a combination of web-based and offline services for UC diagnosis and treatment. The number of visits in the combination of web-based and offline service modes was the highest (mean 13.83, SD 11.07), and younger patients were inclined to visit internet hospitals (49.66%>34.71%). In addition, compared with physical hospitals, there was no difference in testing fees and examination fees for patients with UC in internet hospitals, but medicine fees were lower.

Conclusions: The intelligent diagnosis and treatment model provided by internet hospitals demonstrates the potential benefits in managing UC, including feasibility, accessibility, convenience, and economics.

(J Med Internet Res 2025;27:e60019) doi:10.2196/60019

KEYWORDS

inflammatory bowel disease; ulcerative colitis; intelligent diagnosis and treatment service; internet hospital; chronic disease management

XSL•FO
RenderX

Introduction

Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) affecting the rectum and colon, characterized by a relapsing and remitting course. As of 2023, UC affects an estimated 5 million individuals globally, with a rapidly increasing incidence [1]. This growing prevalence imposes a substantial economic burden on both health care systems and patients.

Uncontrolled UC can lead to severe complications, including malnutrition [2], intestinal infections [3,4], colon cancer [5], anemia, perforation, toxic megacolon [6], and sepsis. These complications not only decrease the quality of life but also increase the risk of mortality. Effective management strategies are essential to reduce these risks.

Traditionally, patients with UC have required frequent clinic visits and hospitalizations due to the complexity of the disease. To overcome these challenges, several telehealth models have been developed to monitor disease activity and facilitate management, reducing the number of outpatient visits and hospital admissions while improving medication adherence [7,8]. Recent advancements in artificial intelligence (AI), particularly large language models such as ChatGPT, show promise in supporting chronic disease management [9,10]. Studies have highlighted their potential for interpreting medical information and generating responses to common patient queries, thereby enhancing patient education and engagement. Despite these advancements, limitations remain, including short follow-up periods, restricted generalizability to broader UC populations, lack of common-sense judgment, and delays in information updates. In addition, existing telehealth models often rely heavily on pharmacists for medication management [11-13], but greater clinician involvement is essential to ensure comprehensive care for patients with UC.

The internet hospital, as a digital expansion of traditional medical services, uses advanced information and communication technologies to offer consultation, prescription, and follow-up services [14]. Internet hospitals established by physical hospitals can connect web-based and offline medical services, ensuring the continuity of care and adequately meeting the medical needs of patients. The COVID-19 pandemic significantly accelerated their development, as government initiatives promoted the use of internet-based health care platforms to enhance access and continuity of care [15-17]. By early 2023, over 3000 internet hospitals were operating across China, demonstrating the growing importance of this model in managing chronic diseases such as UC. Since internet hospitals are not constrained by geographical limitations, patients can conveniently download the client software for internet hospital services on their mobile phones and schedule appointments for various medical services. Consequently, internet hospitals enhance the accessibility and convenience of health care delivery, particularly in addressing the challenges faced by chronic patients who frequently visit hospitals and experience prolonged waiting times for medical treatment.

Based on the current situation and the development trends in health care, we propose 4 hypotheses regarding internet hospitals in the context of UC management. First, internet hospitals can comprehensively meet the medical service needs of patients with UC by leveraging intelligent tools for diagnosis, treatment, and follow-up. Second, more patients with UC will opt for internet hospital services in the future, drawn by the convenience of digital health care. Third, the accessibility of internet hospitals will lead to an increase in the frequency of visits by patients with UC. Finally, internet hospital services are more economical, reducing costs such as transportation and medicine fees.

Methods

Data Sources

The data for this study were obtained from Tianjin Medical University General Hospital, a Grade-A tertiary general hospital located in Tianjin. In response to the COVID-19 epidemic and the need for effective prevention and control measures, Tianjin Medical University General Hospital established an internet hospital, which officially commenced operations in March 2020. The intelligent diagnosis system in the internet hospital uses AI-based tools, such as the Left-Hand Doctor software, as an auxiliary diagnosis and treatment tool., This system uses deep learning algorithms and big data processing to evaluate symptoms and generate diagnostic probabilities. It also integrates interactive dialogue systems to guide patients through their treatment journey [18,19]. The flowchart of the internet hospital's web-based service is shown in Figure 1.

XSL•FO
RenderX

Figure 1. The flowchart of the internet hospital's web-based service process. AI: artificial intelligence.

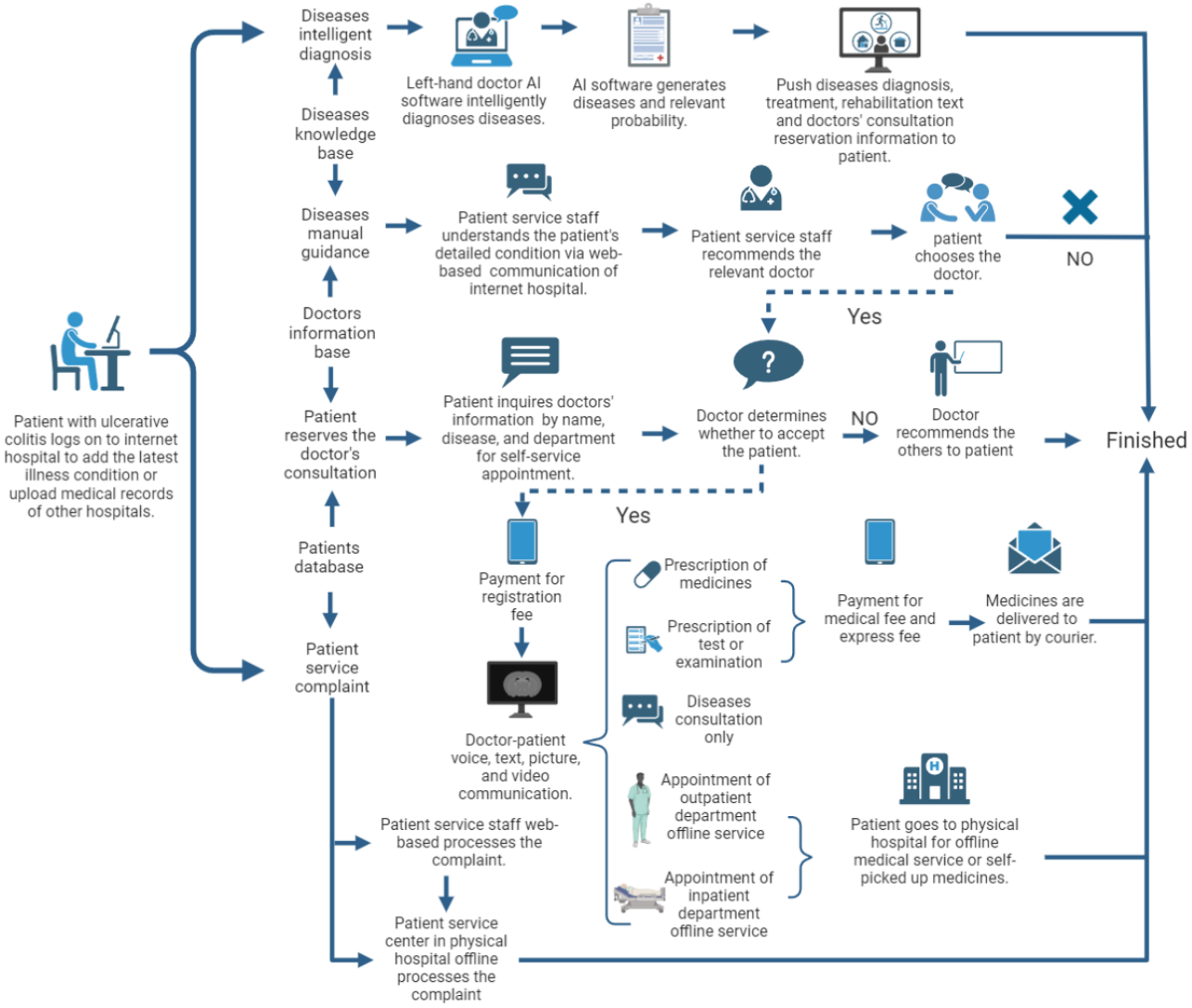

Data Collection and Inclusion Criteria

We retrieved 15,248 medical records using keywords related to IBD and UC, including *ICD* (*International Classification of Diseases*) codes K52.915 and K51.900, from the database of Tianjin Medical University General Hospital. After excluding 21 incomplete records, 1999 cases with no more than 4 consultations were excluded, as UC is a chronic disease requiring regular follow-up. We believe that if patients visit fewer than 4 times in 3 years, their data may be incomplete, and new patients may not have developed medical habits. In addition, infrequent visits do not accurately reflect preferences for offline or web-based services. We also excluded 126 cases involving only COVID-19 nucleic acid testing and 3285 Crohn disease records. Ultimately, 852 patients were included, with a total of 9817 UC-related visits over 3 years. Of these, 445 (52.2%) were male patients, 407 (47.8%) were female patients, and 649 (76.2%) patients had basic health insurance. This data collection flowchart is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Data collection flowchart. ICD: International Classification of Diseases.

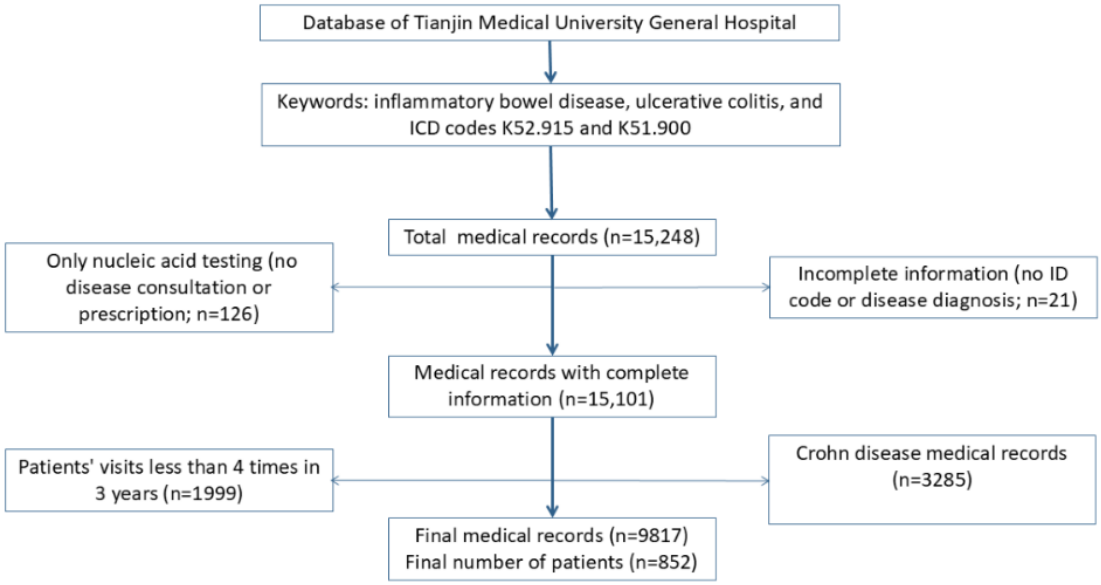

Study Design

In this study, we extracted the diagnosis and treatment data of patients with UC from the Department of Gastroenterology between July 1, 2020, and June 31, 2023. The dataset includes data from in-person consultations at the physical hospital outpatient department and web-based consultations through the internet hospital over the same period. The data also includes information about patients with UC, such as their gender, age, disease diagnosis, treatment time, treatment method, medical expenses, and medical insurance. Our research team tracked the associated patients' visits to UC over 3 years, using their identification card numbers as unique identifiers. Starting in July 2020, the data were divided into 6-month periods to discover the changing trends in the medical services sought by patients with UC, as the promotion and application of internet hospitals continued to grow.

Measures

Medical preferences refer to the types of medical modalities that patients with UC select based on their individual needs, experiences, and the advantages of each modality. From an empirical perspective, we analyze the number of patients using different medical service modes over time to illustrate trends.

The economy of internet hospitals specifically pertains to the comparison of web-based medical fees with those charged in offline settings, including testing fees, examination fees, and medicine fees. Furthermore, patients who opt for internet hospital services can save time and transportation costs for themselves and their companions.

Statistical Analysis

We used chi-square tests to analyze differences in medical preferences and service demands between patients with UC using physical or internet hospitals. We used *t* tests to compare medical expenses across the 2 service modes, while rank-sum tests were used to evaluate differences in the number of patient visits among different medical modes. Frequency and percentages were used to characterize noncontinuous variables, including gender, age group, patient needs, and the number of patients. In contrast, mean and SD values were used to describe continuous variables such as the number of visits and medical fees. All data analysis was performed using IBM SPSS Statistics (version 25).

Ethical Considerations

This study was reviewed and approved by the Ethics Committee of the Tianjin Medical University General Hospital (approval number IRB2024-YX-066-01). The study involved analyzing patient data from the hospital database and did not involve human experimentation or compensation. The data used for this research were extracted from the hospital's information system with previous approval from the Tianjin Medical University General Hospital. To ensure patient privacy and confidentiality, identifying information such as names, addresses, and contact details were removed during data extraction. As a result, the datasets used in the study were anonymized, and no personally identifiable information was accessible to the research team. All patient-related datasets met ethical and legal requirements for data protection. The study adhered to the principles of informed consent, and participants were fully informed of the study's purpose and procedures. In addition, the research team complied with all relevant local, national, and international laws and regulations regarding the protection of personal information, privacy, and human rights. The Ethics Committee of the Tianjin Medical University General Hospital granted approval for this study following a thorough review of the research protocol.

Results

Characteristics of Web-Based and Offline Service Demand

The web-based service mode of the internet hospital can provide equal diagnosis and treatment as a physical hospital. However, web-based consultations (n=3528) were fewer than offline consultations (n=6289), with a significant difference in service demands (P<.001, Table 1). The primary web-based service was prescribing medication (2071/3528, 58.70%), followed by

XSL•FO
RenderX

disease consultations (1224/3528, 34.70%). These findings indicated that web-based services played an important role in the diagnosis and treatment needs of patients with UC, meeting their basic needs, and facilitating follow-up and disease management for chronic diseases. These findings highlight the role of web-based platforms in addressing the basic needs of patients with UC and supporting follow-up care and disease management.

Table 1. Patients' service needs and numbers of visits across different age cohorts over 3 years (N=9817; χ23=879.9, *P*<.001).

Patients' demands for diagnosis and treatment services	Web-based service, n/N (%)	Offline service, n/N (%)
Prescribe medicine	2071/3528 (58.70)	4634/6289 (73.69)
Prescribe testing and examination	125/3528 (3.54)	553/6289 (8.79)
Prescribe drugs, testing, and examination	108/3528 (3.06)	419/6289 (6.66)
Disease consultation only	1224/3528 (34.70)	683/6289 (10.86)

Preference of Different Ages for Medical Treatment Modes

Patients with UC in the age groups of younger than 30 years and between 30 and 40 years demonstrated a greater preference for web-based consultations, with rates of 49.66% and 34.71%, respectively. In contrast, patients aged 50 years and older predominantly opted for in-person consultations at physical hospitals, as indicated by their preference rates of 43.09% compared with 28.03% in Table 2.

Table 2. Patients' visit numbers for different age cohorts across different diagnosis and service modes over 3 years (N=9817; χ24=278.7, *P*<.001).

Patients' visit numbers of different age cohorts	Web-based service, n/N (%)	Offline service, n/N (%)
Age≤30 years	621/3528 (17.60)	712/6289 (11.32)
30<Age≤40 years	1131/3528 (32.06)	1471/6289 (23.39)
40<Age≤50 years	787/3528 (22.31)	1396/6289 (22.20)
50<Age≤60 years	493/3528 (13.97)	1156/6289 (18.38)
Age>60 years	496/3528 (14.06)	1554/6289 (24.71)

Changing Trend of Diagnosis and Treatment Service Modes

During period 1, only 3.52% of patients opted for completely web-based services, while the majority (n=318, 79.90%) preferred entirely offline services, as given in Table 3. By period 6, there was a significant increase in the percentage of patients choosing either completely web-based services (n=146, 31.67%) or a combination of web-based and offline services (n=124, 26.90%), together comprising a total of 58.57%. In contrast, the proportion of patients selecting exclusively offline services decreased from 79.90% in Period 1 to just 41.43% by period 6. The trend of patients selecting various diagnosis and treatment service modes over the 6 periods is given in Figure 3. This trend reflects the growing acceptance and use of web-based diagnosis and treatment options among patients with UC.

Table 3. Number (percentage, %) of patients choosing different diagnosis and treatment service modes across 6 periods (χ210=368.9, *P*<.001).

Periods	Offline service, n (%)	Web-based and offline service combination, n (%)	Web-based service, n (%)
Period 1	318 (79.90)	66 (16.58)	14 (3.52)
Period 2	306 (71.50)	83 (19.39)	39 (9.11)
Period 3	322 (66.53)	91 (18.80)	71 (14.67)
Period 4	262 (37.92)	291 (42.11)	138 (19.97)
Period 5	229 (45.35)	145 (28.71)	131 (25.94)
Period 6	191 (41.43)	124 (26.90)	146 (31.67)

XSL•FO
RenderX

Figure 3. Percentage of patients choosing different diagnosis and treatment service modes across 6 periods.

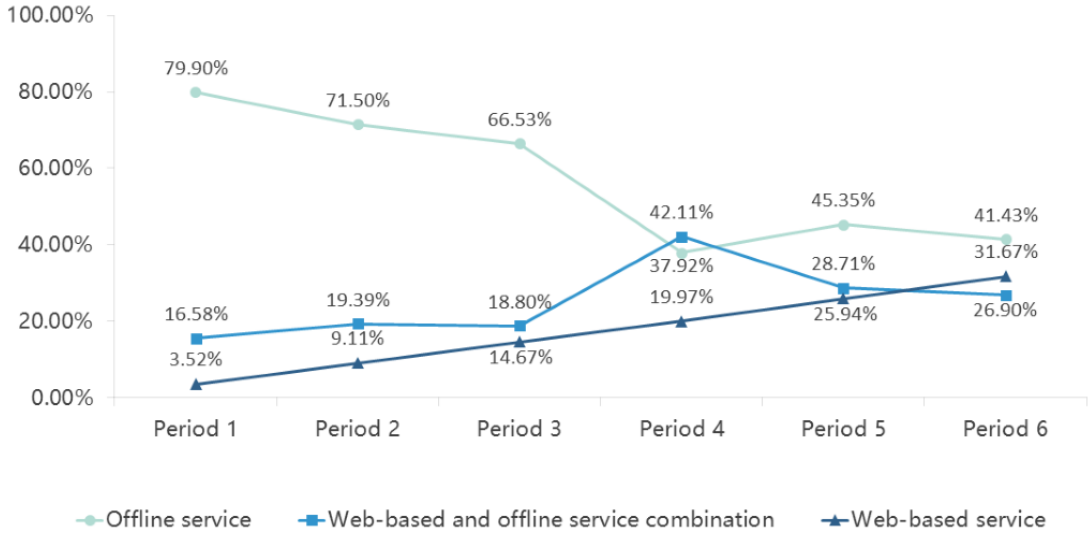

Preference of Medical Treatment Modes

The number of patients (n=498, 58.45%) who opted for a combination of web-based and offline services, as well as completely web-based consultations, exceeded that of those choosing physical hospitals (354/852, 41.55%) in Table 4. However, patients who chose the combination of web-based and offline care had the highest number of visits (mean 13.83, SD 11.07).

Table 4. Number of patient visits for consultations with different medical service modes over 3 years (N=852; Kruskal-Wallis H=51.39, *P*<.001).

Service modes	Number of patients, n (%)	Number of visits, mean (SD)
Offline service	354 (41.55)	9.33 (6.78)
Web-based and offline service combination	424 (49.77)	13.83 (11.07)
Web-based service	74 (8.68)	8.80 (6.81)

Difference in the Fee of Web-Based and Offline Treatment

The testing fees and examination fees for patients with UC in different age groups showed no statistical difference. However, there was a significant disparity in medicine fees for web-based and offline services among patients with UC of all age groups, with web-based drug fees being lower than offline ones, as given in Table 5.

Table 5. Web-based and offline service expenses (yuan) for different age cohorts (N=9817).

Service expenses	Age≤30 years	30<Age≤40 years	40<Age≤50 years	50<Age≤60 years	Age>60 years
Testing fee					
Web-based service, mean (SD)	289.19 (24.79)	293.49 (23.74)	288.75 (25.32)	325.23 (41.22)	264.90 (46.69)
Offline service, mean (SD)	322.26 (29.99)	325.18 (23.76)	307.42 (37.13)	328.38 (25.38)	266.10 (19.10)
T-value	0.92	0.94	0.43	0.06	0.06
P value	.36	.35	.67	.95	.98
Examination fee					
Web-based service, mean (SD)	300.00 (13.63)	393.66 (38.22)	381.03 (40.41)	364.44 (19.17)	413.44 (42.49)
Offline service, mean (SD)	342.14 (11.73)	391.58 (13.30)	404.55 (15.87)	385.99 (16.20)	384.38 (13.48)
T-value	1.81	0.06	0.60	0.57	0.62
P value	.72	.95	.55	.57	.53
Medicine fee					
Web-based service, mean (SD)	731.65 (24.41)	807.03 (14.65)	747.66 (19.40)	700.99 (23.34)	823.04 (23.95)
Offline service, mean (SD)	922.16 (27.14)	996.65 (17.25)	879.55 (15.79)	912.57 (17.65)	914.70 (14.74)
T-value	5.22	8.38	5.27	7.23	3.26
P value	<.001	<.001	<.001	<.001	<.001

XSL•FO
RenderX

Discussion

Principal Findings

The findings indicate that internet hospitals can effectively address the needs of patients with UC. An increasing number of patients with UC, particularly younger individuals, are opting for internet hospitals for their treatment and follow-up care. This trend not only facilitates better management of UC but also contributes to a reduction in health care costs associated with the disease.

The Tianjin Medical University internet hospital platform features the Left Hand Doctor AI software, which provides relevant information on disease diagnosis, treatment, rehabilitation, and available specialists. Consequently, the AI-driven model enhances patients' ability to choose the appropriate clinical departments and health care providers. In contrast to certain internet hospitals that lack support from physical facilities, medications prescribed by this platform can be obtained either through self-pickup at a physical location or via express delivery. In addition, examination and testing procedures can be conducted at a physical hospital. Patients unable to receive a diagnosis through web-based consultations may be referred to a physical hospital's outpatient department for further treatment. Those needing inpatient care can schedule hospitalization appointments directly via the internet hospital platform. In addition, medical disputes can be mediated at the reception service center of the physical hospital. This indicates that the internet hospital effectively facilitates continuity between web-based and offline medical services.

The COVID-19 pandemic has significantly accelerated the adoption of internet hospitals, with video consultations increasingly replacing many in-person visits [20]. Our findings demonstrate that web-based platforms, such as internet hospitals, play a pivotal role in addressing the long-term management needs of patients with UC. The majority of patients using web-based services sought medication prescriptions (58.70%) and disease consultations (34.70%), indicating the practicality of these platforms for chronic disease management and effectively reducing the burden on offline medical resources. Furthermore, our discoveries reveal a significant increase over time in the number of patients choosing fully web-based services or a combination of web-based and offline services. This trend suggests that internet hospitals will become increasingly accessible in the future. Together, these findings highlight the value of integrating web-based and offline services to ensure comprehensive care, especially in resource-constrained settings.

Age-based preferences indicated that younger patients (younger than 40 years) were more likely to use web-based consultations (49.66%), while older patients (50 years and older) predominantly chose offline services (43.09%). This suggests that emerging digital health methods, such as internet hospitals, are more easily accepted by younger people. The flexibility and convenience of web-based platforms appeal to younger demographics, whereas older patients may favor the familiarity and perceived reliability of in-person care. These trends highlight the need for a hybrid model to accommodate diverse patient preferences. Furthermore, web-based services significantly reduced medicine costs compared with offline services (P<.001), with consultation fees also being lower under the unified internet hospital policy in Tianjin. These cost advantages make web-based platforms a more economical option for the long-term management of UC, alleviating the financial burden on patients.

Our findings align with previous studies that demonstrated the growing role of telemedicine in managing chronic diseases. For example, de Jong et al [21] showed that telemedicine systems, such as myIBDcoach, significantly reduced hospital admissions and improved medication adherence in patients with IBD [21]. Similarly, 2 other studies highlighted the effectiveness of telehealth models in improving patient outcomes [22,23]. While these studies focused on specific telehealth tools, our research provides a broader analysis of internet hospital services, demonstrating their capacity to integrate web-based and offline care while maintaining cost-effectiveness and accessibility. However, unlike earlier studies that relied on short follow-up periods [24], our analysis spans 3 years, offering a more comprehensive view of the long-term impact of internet hospitals. A recent study by Al-Sheikh et al [25] emphasized the efficacy of telehealth models in reducing health care costs in IBD management [25]. However, it was unable to eliminate the cost variance resulting from the disparity in hospital registration techniques, as the research participants were from different hospitals. The participants of our study are all from the General Hospital of Tianjin Medical University, thereby effectively avoiding the interference of hospital technical factors.

Strengths and Limitations

Our study has several strengths. First, this retrospective study provides insights into the use of internet and physical hospitals for UC management over 3 years, with a representative sample from a Grade-A tertiary hospital. Second, it is the first to investigate how internet hospitals can facilitate standardized care for patients with UC outside traditional settings. However, this study has limitations. The internet hospital system lacks data on the quality of care and patient satisfaction, which limits the evaluation of service outcomes. A service quality model, a satisfaction degree questionnaire, and even a comprehensive satisfaction degree index measured by a 5-point Likert scale are proposed to be designed based on the conditions of the internet hospitals and then integrated into the hospital information system. Therefore, quality and satisfaction data can be automatically obtained from the hospital database. Furthermore, long-term studies are needed to assess the sustained effects of internet hospitals on patient outcomes, as well as the accuracy and precision of AI diagnosis tools such as the Left-Hand Doctor mentioned in the paper.

Conclusion

Internet hospitals represent an innovative approach to UC management, offering a series of web-based intelligent diagnosis and treatment services. Through seamless integration of web-based and offline care, they ensure continuity of care. Compared with the traditional medical model, this approach highlights the potential benefits in terms of feasibility, accessibility, convenience, and economy. This intelligent mode

XSL•FO
RenderX

can also be applied to the medical service and disease management of other chronic diseases.

Acknowledgments

The authors sincerely thank the Information Center of Tianjin Medical University General Hospital for retrieving medical records. This study was supported by grants from the Tianjin Municipality Education Commission Major Research Project of Social Sciences (2022JWZD21).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TY contributed to the conceptualization, formal analysis, and funding acquisition. WL contributed to the data curation and validation. YL contibuted to the methodology and writing-orginal draft. CJ and ZW contibuted to the writing-review & editing. HC contributed to the supervision and project administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The full list of keywords and search strategy of the database.
[DOCX File , 15 KB - jmir_v27i1e60019_app1.docx]

References

1. Le Berre C, Honap S, Peyrin-Biroulet L. Ulcerative colitis. Lancet 2023;402(10401):571-584. [doi: 10.1016/S0140-6736(23)00966-2] [Medline: 37573077]
2. Balestrieri P, Ribolsi M, Guarino MPL, Emerenziani S, Altomare A, Cicala M. Nutritional aspects in inflammatory bowel diseases. Nutrients 2020;12(2):372 [FREE Full text] [doi: 10.3390/nu12020372] [Medline: 32023881]
3. Barber GE, Hendler S, Okafor P, Limsui D, Limketkai BN. Rising incidence of intestinal infections in inflammatory bowel disease: a nationwide analysis. Inflamm Bowel Dis 2018;24(8):1849-1856. [doi: 10.1093/ibd/izy086] [Medline: 29722832]
4. Spartz EJ, DeDecker LC, Fansiwala KM, Noorian S, Roney AR, Hakimian S, et al. Recent trends and risk factors associated with Clostridioides difficile infections in hospitalized patients with inflammatory bowel disease. Aliment Pharmacol Ther 2024;59(1):89-99. [doi: 10.1111/apt.17777] [Medline: 37873878]
5. Lakatos PL, Lakatos L. Risk for colorectal cancer in ulcerative colitis: changes, causes and management strategies. World J Gastroenterol 2008;14(25):3937-3947 [FREE Full text] [doi: 10.3748/wjg.14.3937] [Medline: 18609676]
6. Autenrieth DM, Baumgart DC. Toxic megacolon. Inflamm Bowel Dis 2012;18(3):584-591. [doi: 10.1002/ibd.21847] [Medline: 22009735]
7. Kang J, Su J, Wang J, Ren H, Yin A, Li J, et al. An education-based telehealth group improves the management and clinical outcomes of patients with inflammatory bowel disease in China (THEIM study). United European Gastroenterol J 2023;11(10):960-972 [FREE Full text] [doi: 10.1002/ueg2.12470] [Medline: 37921707]
8. Cross RK, Cheevers N, Rustgi A, Langenberg P, Finkelstein J. Randomized, controlled trial of home telemanagement in patients with ulcerative colitis (UC HAT). Inflamm Bowel Dis 2012;18(6):1018-1025 [FREE Full text] [doi: 10.1002/ibd.21795] [Medline: 21688350]
9. Gravina AG, Pellegrino R, Palladino G, Imperio G, Ventura A, Federico A. Charting new AI education in gastroenterology: cross-sectional evaluation of ChatGPT and perplexity AI in medical residency exam. Dig Liver Dis 2024;56(8):1304-1311. [doi: 10.1016/j.dld.2024.02.019] [Medline: 38503659]
10. Gravina AG, Pellegrino R, Cipullo M, Palladino G, Imperio G, Ventura A, et al. May ChatGPT be a tool producing medical information for common inflammatory bowel disease patients' questions? An evidence-controlled analysis. World J Gastroenterol 2024;30(1):17-33 [FREE Full text] [doi: 10.3748/wjg.v30.i1.17] [Medline: 38293321]
11. Li H, Zheng S, Li D, Jiang D, Liu F, Guo W, et al. The establishment and practice of pharmacy care service based on internet social media: telemedicine in response to the COVID-19 pandemic. Front Pharmacol 2021;12:707442 [FREE Full text] [doi: 10.3389/fphar.2021.707442] [Medline: 34658854]
12. Jiang Y, Lu R, Ou M, Zhou Q, Du Z, Zhu H. Application of "Internet +" pharmaceutical consultation services in psychiatric hospital during the epidemic. Asian J Psychiatr 2023;82:103532 [FREE Full text] [doi: 10.1016/j.ajp.2023.103532] [Medline: 36842265]

XSL•FO
RenderX

13. Bu F, Sun H, Li L, Tang F, Zhang X, Yan J, et al. Artificial intelligence-based internet hospital pharmacy services in China: perspective based on a case study. Front Pharmacol 2022;13:1027808 [FREE Full text] [doi: 10.3389/fphar.2022.1027808] [Medline: 36438784]
14. Zhi L, Yin P, Ren J, Wei G, Zhou J, Wu J, et al. Running an internet hospital in China: perspective based on a case study. J Med Internet Res 2021;23(9):e18307 [FREE Full text] [doi: 10.2196/18307] [Medline: 34342267]
15. Liu L, Shi L. Chinese patients' intention to use different types of internet hospitals: cross-sectional study on virtual visits. J Med Internet Res 2021;23(8):e25978 [FREE Full text] [doi: 10.2196/25978] [Medline: 34397388]
16. Sun S, Yu K, Xie Z, Pan X. China empowers internet hospital to fight against COVID-19. J Infect 2020;81(1):e67-e68 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.061] [Medline: 32251688]
17. Venkatesh KP, Raza MM, Kvedar J. Has increased telehealth access during COVID-19 led to over-utilization of primary care? NPJ Digit Med 2022;5(1):178 [FREE Full text] [doi: 10.1038/s41746-022-00740-4] [Medline: 36513808]
18. Zhang C. Intelligent application: a breakthrough medicine for internet healthcare. China Hospital CEO 2020(16):84-87.
19. Liu W. The third party in the clinic is triggering a revolution. China Hospital CEO 2020(19):88-89.
20. van den Bosch SC, van Dalen D, Meinders M, van Goor H, Bergé S, Stommel M, et al. Outpatient video visits during the COVID-19 pandemic: cross-sectional survey study of patients' experiences and characteristics. J Med Internet Res 2024;26:e49058 [FREE Full text] [doi: 10.2196/49058] [Medline: 38536236]
21. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. Lancet 2017;390(10098):959-968. [doi: 10.1016/S0140-6736(17)31327-2] [Medline: 28716313]
22. Cohen-Mekelburg S, Valicevic A, Lin LA, Saini SD, Kim HM, Adams MA. Inflammatory bowel disease hospitalizations are similar for patients receiving televisit-delivered outpatient care and those receiving traditional in-person care. Am J Gastroenterol 2024;119(8):1555-1562. [doi: 10.14309/ajg.0000000000002703] [Medline: 38314800]
23. Pang L, Liu H, Liu Z, Tan J, Zhou LY, Qiu Y, et al. Role of telemedicine in inflammatory bowel disease: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(3):e28978 [FREE Full text] [doi: 10.2196/28978] [Medline: 35323120]
24. Aguas M, Del Hoyo J, Vicente R, Barreiro-de Acosta M, Melcarne L, Hernandez-Camba A, et al. Telemonitoring of active inflammatory bowel disease using the app TECCU: short-term results of a multicenter trial of GETECCU. J Med Internet Res 2024;26:e60966 [FREE Full text] [doi: 10.2196/60966] [Medline: 39189160]
25. Al-Sheikh M, Ankersen DV, Olsen J, Spanggaard M, Peters-Lehm CT, Naimi RM, et al. The costs of home monitoring by telemedicine vs standard care for inflammatory bowel diseases-a danish register-based, 5-year follow-up study. J Crohns Colitis 2025;19(2):jjae120. [doi: 10.1093/ecco-jcc/jjae120] [Medline: 39110131]

Abbreviations

AI: artificial intelligence
IBD: inflammatory bowel disease
ICD: International Classification of Diseases
UC: ulcerative colitis

Edited by A Schwartz; submitted 02.05.24; peer-reviewed by A Ren, Z Lawal, R Pellegrino, J Soldera; comments to author 17.11.24; revised version received 20.12.24; accepted 21.02.25; published 18.03.25.

Please cite as:
Yu T, Li W, Liu Y, Jin C, Wang Z, Cao H
Application of Internet Hospitals in the Disease Management of Patients With Ulcerative Colitis: Retrospective Study
J Med Internet Res 2025;27:e60019
URL: https://www.jmir.org/2025/1/e60019
doi:10.2196/60019
PMID:

©Tianzhi Yu, Wanyu Li, Yingchun Liu, Chunjie Jin, Zimin Wang, Hailong Cao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Associated With the Availability of Virtual Consultations in Primary Care Across 20 Countries: Cross-Sectional Study

Gabriele Kerr[1,2], MSc; Geva Greenfield[1,2], PhD; Edmond Li[1,3], MBChB; Thomas Beaney[1], MBBS; Benedict W J Hayhoe[1,2], MBBS; Josip Car[1,4], PhD; Ana Clavería[5,6], PhD; Claire Collins[7,8], PhD; Gustavo Gusso[9], PhD; Robert D Hoffman[10], MD; Geronimo Jimenez[11], PhD; Tuomas H Koskela[12,13], MD; Liliana Laranjo[14], PhD; Heidrun Lingner[15], MD; Ensieh Memarian[16], PhD; Katarzyna Nessler[17], PhD; Davorina Petek[18], PhD; Rosy Tsopra[19,20], MD, PhD; Azeem Majeed[1,2], MBBS; Ana Luisa Neves[1,2], PhD

[1]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom
[2]NIHR Applied Research Collaboration Northwest London, London, United Kingdom
[3]Institute of Global Health Innovation, Department of Surgery and Cancer, Imperial College London, London, United Kingdom
[4]School of Life Course and Population Sciences, King's College London, London, United Kingdom
[5]Primary Care Research Unit, Servizo Galego de Saúde, Vigo, Spain
[6]I-Saúde Group, Galicia Sur Health Research Institute, Vigo, Spain
[7]Irish College of General Practitioners, Dublin, Ireland
[8]Department of Public Health and Primary Care, Ghent University, Ghent, Belgium
[9]Department of Internal Medicine, Universidade de São Paulo, São Paulo, Brazil
[10]Department of Family Medicine, Medical Faculty, Tel Aviv University, Tel Aviv, Israel
[11]Department of Public Health and Primary Care, Leiden University, Leiden, The Netherlands
[12]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[13]The Wellbeing Services County of Pirkanmaa, Tampere, Finland
[14]Westmead Applied Research Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[15]Center for Public Health and Healthcare, Department of Medical Psychology Unit 5430, Hannover Medical School, Hannover, Germany
[16]Internal Medicine Research Group, Department of Clinical Sciences, Faculty of Medicine, Lund University, Malmö, Sweden
[17]Department of Family Medicine, Jagiellonian University Medical College, Krakow, Poland
[18]Department of Family Medicine, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
[19]Centre de Recherche des Cordeliers, Université Paris Cité & Sorbonne Université, Paris, France
[20]Assistance Publique-Hôpitaux de Paris, Department of Medical Informatics, Hôpital Européen Georges-Pompidou et Hôpital Necker-Enfants Malades, Paris, France

Corresponding Author:
Gabriele Kerr, MSc
Department of Primary Care and Public Health
Imperial College London
White City Campus, 90 Wood Lane
London, W12 0BZ
United Kingdom
Phone: 44 20 7594 9826
Email: gk916@ic.ac.uk

Abstract

Background: Virtual consultations represent a notable change in health care delivery following the COVID-19 pandemic. Understanding the dynamics of virtual consultations is critical in assessing health care system resilience and adaptability in times of crisis.

Objective: This study aimed to describe the availability and hours of use of telephone, video, and human chat consultations before and during the COVID-19 pandemic period, and identify factors associated with their availability.

XSL•FO
RenderX

Methods: Primary care physicians (PCPs) from 20 upper-middle– and high-income countries completed a cross-sectional web-based survey in 2020. Factors associated with availability were investigated using chi-square tests and effect size (ES) estimates calculated using Cramer V.

Results: A total of 1370 PCPs were included in this study (85.4% of the total sample of 1605). Telephone consultations were the most frequently available type of virtual consultations before and during the pandemic (73.1% and 90.4%, respectively). Significant increases in availability and use were observed during the pandemic for all the types of virtual consultations. The largest absolute increase in availability was observed for video consultations (39.5%), followed by telephone (17.3%) and chat (8.6%; all $P<.001$). The largest increase in use was observed for telephone consultations (+11 hours per week, $P<.001$). Digital maturity of the practice was weakly associated with availability of video consultations both before (ES 0.2) and during (ES 0.2) the pandemic ($P<.001$ for both), and with chat consultations before the pandemic only (ES 0.1, $P=.001$). Greater availability of video and chat consultations was found in PCPs who had completed digital health training, both before and during the pandemic ($P<.001$ for all). There was significant country-level variation in the use and availabilities of the technologies between both time periods. The association between country and the availability of telephone consultations changed from strong (ES 0.5, $P<.001$) to weak (ES 0.2, $P=.03$), while the relationship between country and video consultations changed from moderate (ES 0.3, $P<.001$) to strong (ES 0.5, $P<.001$).

Conclusions: Our study demonstrates the transformative impact of the COVID-19 pandemic on the availability of virtual consultations globally, and how practice-level factors, predominantly digital maturity, digital health training, and country, were associated with the availability of virtual consultations. Further exploration of drivers of availability, particularly at the national level, is needed to ensure sustained and effective implementation of virtual consultations.

International Registered Report Identifier (IRRID): RR2-10.2196/30099

(J Med Internet Res 2025;27:e65147) doi:10.2196/65147

KEYWORDS

digital health; primary care; telemedicine; virtual consultation; healthcare delivery; online consultation; primary care physician; upper-middle income; upper-middle income countries; high-income countries; online survey; chi-squared test; remote healthcare; video consultation; chat consultation; telephone consultations; digital technology; virtual care; teleconsultation; telehealth; remote consultation

Introduction

The emergence of virtual consultations, defined as remote health care interactions facilitated by digital technologies, is a significant evolution in health care delivery. Telephone, video, and chat consultations may be more accessible than in-person appointments, as they offer rapid real-time communications with providers without a need to travel [1,2]. Despite these potential benefits, before the COVID-19 pandemic, these technologies, particularly telephone consultations, were steadily gaining traction, but had not reached widespread integration into most mainstream primary health care systems [3,4].

During the COVID-19 pandemic, virtual care became vital to the safe and efficient continuation of primary care delivery, when minimizing in-person encounters was essential to protect both health care staff and patients from the risk of infection [5,6]. Many health systems adopted some form of “virtual first” approach to primary health care provision [4]. The initial virtual encounter aimed to manage patients’ needs without in-person contact wherever possible while reserving “higher risk” face-to-face visits for those at greatest need, and where physical examination was deemed to be essential.

Throughout the pandemic, primary care physicians (PCPs) faced barriers in adopting and implementing virtual consultations, with potential consequences impacting the quality of care delivered to patients [2]. The ability of PCPs to effectively transition to virtual service delivery depends on multiple factors, including organizational and policy incentives, digital health infrastructure capacity and investment, cultural norms and attitudes, and the digital health literacy and skills of PCPs and patient populations [3,4,7-9]. These factors would act as barriers or drivers to differing degrees depending on the specific consultation technology, with likely fewer infrastructural or skill barriers for telephone consultations compared to video or chat consultations [10]. These factors may have resulted in variation in adoption and use of different virtual consulting technologies between PCPs and providers from different settings [2,4].

With growing demand for rapid and convenient access to primary care, alongside financial constraints requiring efficiency gains, virtual care appeared as an attractive solution to enhance patient accessibility [2]. Consequently, virtual consultations continue as a core component of health care delivery in many upper-middle– and high-income countries beyond the pandemic [4].

Examining the landscape of virtual health care technologies before and during the pandemic can help us better understand the magnitude of the transition to these new models of care. This transition highlights the investments made into the digital health capacity of primary care systems, carrying significant long-term implications for how care is delivered. However, how the availability and uptake of virtual consultations varied across PCPs from different settings, including different countries, is uncertain [11].

The aim of this study was to analyze access and use of virtual consultations before and during the pandemic, and factors

XSL•FO
RenderX

associated with availability of virtual consulting technologies between PCPs from different settings. Specific aims include to analyze the availability and hours of use of telephone, video and chat consultations before and during the COVID-19 pandemic, and to identify factors associated with their availability.

Methods

Study Design

This study used data from a cross-sectional web-based questionnaire completed by PCPs of 20 upper-middle- and high-income countries (Australia, Brazil, Canada, Chile, Colombia, Croatia, Finland, France, Germany, Ireland, Israel, Italy, Poland, Portugal, Spain, Slovenia, Sweden, Türkiye, the United Kingdom, and the United States). The questionnaire was designed and administered by the inSIGHT Research Group, a consortium of academic primary care researchers from the 20 countries previously listed. The study adheres to the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) guidelines for reporting observational studies [12].

Data Collection

Participants were eligible if they were practicing PCPs in 1 of the 20 countries listed above, between March and September 2020. The study was conducted between June and September 2020. National leads in each country invited PCPs through their formal organizations or personal networks through email or social media (ie, Facebook and Twitter [subsequently rebranded X]). The questionnaire was hosted on Qualtrics (Silver Lake) and was available in English, French, German, Italian, Spanish, and Portuguese. A complete description of the study protocol, including the full questionnaire and power analyses, has been previously published [13]. Sections of the questionnaire relevant to this study are included in pages 2-5 in Multimedia Appendix 1.

Study Variables

Participants were asked to answer whether chat (ie, using a text-based messaging system), telephone or video consultations were available in their practice before or during the COVID-19 pandemic (from March 11, 2020). Respondents ticked a box for each period considered (ie, before or during the COVID-19 pandemic) to indicate a technology was available in that period.

Respondents were subsequently asked how many hours they spent per week on each type of consultation in each time period. Before analysis, hours per week spent on each of the 3 virtual consultation technologies were cleaned to remove answers of ≥100 hours per week. A response of >0 hours spent on a technology was considered evidence for the technology being available. This study includes PCPs who responded to at least 1 question on the availability or hours of use of virtual consultation technologies.

Predictor variables included country, urbanicity (rural, mixed, and urban), and practice digital maturity. Practice digital maturity was assessed using the digital maturity framework developed by Flott et al [14], which considers the 6 dimensions of usage, resources, and abilities (organizational and individual), interoperability, general evaluation methodology, and impact. PCPs could agree or disagree with 6 statements about their practice's digital maturity, corresponding with the 6 dimensions. A digital maturity score was calculated for each PCP by granting 1 point for each statement with which the PCP indicated agreement, giving a possible range of 0 to 6 where a score of 6 indicates high digital maturity. PCPs were also asked whether they have completed training on digital technologies before or during the COVID-19 pandemic.

Statistical Analysis

The total number of hours spent per week before and during the pandemic on virtual consultation technologies was calculated for PCPs who reported availability of at least one of the technologies in the period. For PCPs who reported the technology as available in both time periods, the number of hours spent by PCPs on each technology before and during the COVID-19 pandemic were compared using paired Wilcoxon signed rank tests, and the relationship between the predictors and change in hours of use of each technology was investigated using univariable linear regression models.

McNemar tests were conducted to compare availability of each technology before and during the COVID-19 pandemic period. Absolute differences in the percentage of PCPs with each technology available in each time period were described. Plots were created to visualize changes in technology availability and hours of use by country of PCP employment.

Cramer V was calculated to estimate the effect size (ES) of practice factors upon the variation in the availability of digital technologies before and during the COVID-19 pandemic. Cramer V estimates of ES to describe the strength of association between the predictors and outcomes were categorized as weak (0-0.29), moderate (0.3-0.49), or strong (≥0.5). The change in percentage of PCPs with each technology available was visualized by country. *P* values for statistical tests were adjusted for multiple comparisons using the Holm-Bonferroni method [15]. All analyses were performed in R (version 4.3; R Foundation for Statistical Computing) [16], and a significance level of .05 was used throughout.

Ethical Considerations

Ethical approval was granted by the UK Imperial College Research Ethics Committee (reference 20IC5956), which oversees health-related research with human participants. Survey participants gave their written informed consent to participate in the study. Data collected were anonymous and no compensation was given for participation in the survey.

Results

Overview

This study includes responses from 1370 PCPs who responded to at least 1 question on the availability and/or hours of use of virtual consultation technologies, representing 85.4% of the total sample of 1605 (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the 1370 surveyed primary care physicians.

Characteristic	Primary care physicians, n (%)
Age category (years)	
>30	79 (5.8)
30-39	434 (31.7)
40-49	356 (26)
50-59	289 (21.1)
60-69	192 (14)
70+	16 (1.2)
Prefer not to answer	4 (0.3)
Missing	0 (0)
Gender	
Female	827 (60.4)
Male	535 (39.1)
Other	1 (0.1)
Prefer not to answer	7 (0.5)
Missing	0 (0)
Urbanicity	
Mixed	307 (22.4)
Rural	211 (15.4)
Urban	852 (62.2)
Missing	0 (0)
Years of PCP[a] Experience (years)	
<5	265 (19.3)
5-10	295 (21.5)
10-15	210 (15.3)
15-20	156 (11.4)
>20	444 (32.4)
Missing	0 (0)
Digital maturity score	
0	108 (7.9)
1	112 (8.2)
2	130 (9.5)
3	249 (18.2)
4	268 (19.6)
5	234 (17.1)
6	269 (19.6)
Country of PCP employment	
Australia	69 (5)
Brazil	48 (3.5)
Canada	47 (3.4)
Chile	52 (3.8)
Colombia	60 (4.4)
Croatia	55 (4)

XSL•FO
RenderX

Characteristic	Primary care physicians, n (%)
Finland	43 (3.1)
France	56 (4.1)
Germany	46 (3.4)
Ireland	237 (17.3)
Israel	65 (4.7)
Italy	89 (6.5)
Poland	49 (3.6)
Portugal	77 (5.6)
Slovenia	66 (4.8)
Spain	85 (6.2)
Sweden	67 (4.9)
Turkey	51 (3.7)
United Kingdom	55 (4)
United States	53 (3.9)
Missing	0 (0)

[a]PCP: primary care physician.

The majority (827/1370, 60.4%) of the respondents were female and 57.6% (790/1370) were aged between 30-49 years. Almost a third (444/1370, 32.4%) of the respondents had clinical experience of more than 20 years. PCPs spent a median of 36 (IQR 28-40) hours on clinical work per week. The highest proportion of the respondents (852/1370, 62.2%) worked in practices based in urban areas. The median digital maturity score of their practices as reported by PCPs was 4 (IQR 2-5). Training on digital-first technologies was undertaken by (312/1370, 22.8%) PCPs before the pandemic and by (375/1370, 27.4%) PCPs during the pandemic period. A breakdown of PCP characteristics by country is available in Tables S1 and S2 in Multimedia Appendix 1.

Use of Digital Technologies

PCPs reported spending a median of 3 hours per week using these tools (IQR 1-5), increasing to 15 (IQR 8-25) during the pandemic period (P<.001). Hours spent per week on specific technologies are shown in Table 2.

The average number of hours per week spent on each type of virtual consultation increased during the pandemic (Table 2). The greatest change was observed for time spent on telephone consultations (+11.0 hours/week, P<.001), on which 91.8% of PCPs reported spending increased time.

Country of PCP employment was associated with changes in hours spent per week on telephone (R^2=0.2, P<.001) and chat consultations (R^2=0.1, P=.001), but not with changes in hours spent on video consultations (R^2=0.1, P=.73). The increase in hours spent per week on telephone consultations was largely driven by PCPs from Poland, Spain, Canada, Chile, and Portugal, who spent more than 15 additional hours per week on telephone consultations during the COVID-19 pandemic compared with before (Figure 1).

No association was found between changes in hours of use of any of the 3 virtual consultation technologies and any of practice digital maturity score, training, or urbanicity (Figure S1 in Multimedia Appendix 1).

Table 2. Average change in hours spent on virtual consultations by primary care physicians before and during the COVID-19 pandemic period, amongst primary care physicians who had the technology available in both time periods. The COVID-19 pandemic period was defined as the period from March 11, 2020 onwards to the date of survey completion by the primary care physicians (between June and September 2020).

Technology	Denominator	Mean hours spent per week before the pandemic	Mean hours spent per week during the pandemic	Mean difference in hours, mean (SD)[a]	P values[b]
Telephone consultations	883	3.8	14.2	+11 (10.7)	<.0001
Video consultations	127	1.3	4.3	+4.5 (7.3)	<.0001
Chat consultations (ie, using a messaging system)	365	2.4	5.3	+3.4 (6.7)	<.0001

[a]The mean difference describes the mean of the change in hours spent by each primary care physician on the technology.

[b]P values correspond to 2-sample Wilcoxon tests.

XSL•FO
RenderX

Figure 1. Mean change in hours spent per week on virtual consultation technologies by primary care physicians during versus before the COVID-19 pandemic period, by country of primary care physician employment. The COVID-19 pandemic period was defined as the period from March 11, 2020, onwards to the date of survey completion by the primary care physician (between June and September 2020). Grey cells indicate where no data was available for change in hours of use.

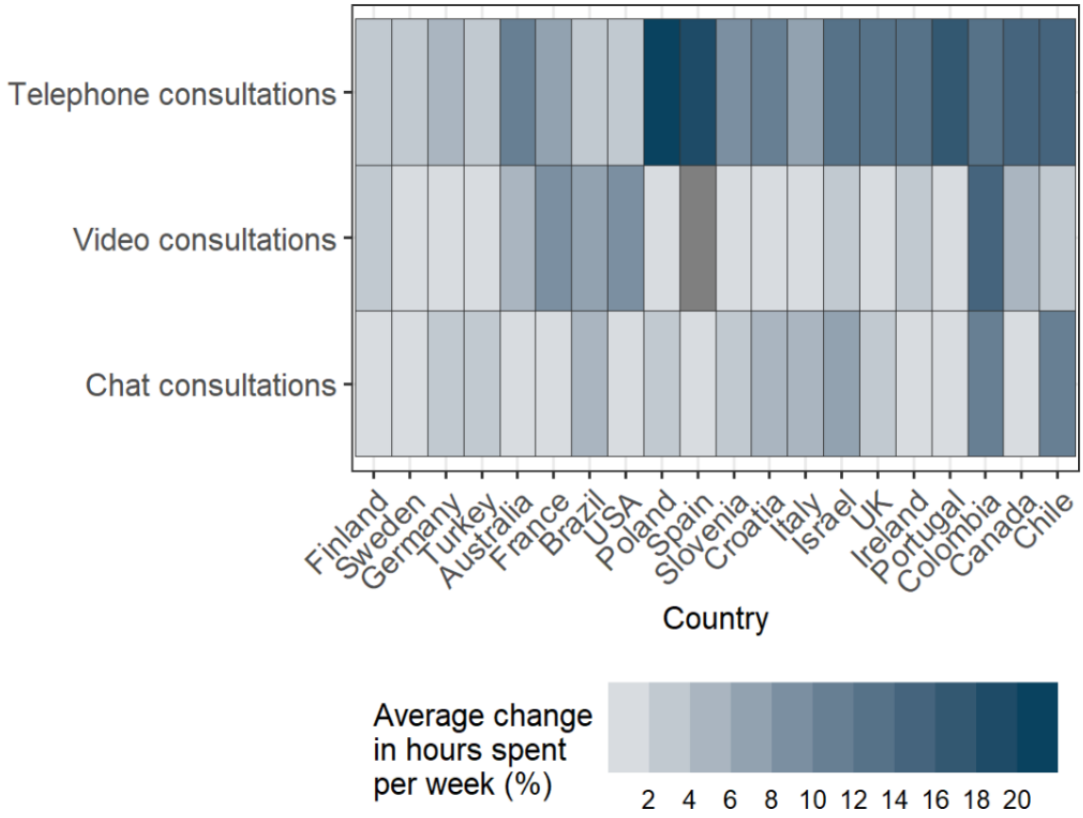

Availability of Virtual Consultations Before and During the COVID-19 Pandemic

Before the pandemic, telephone consultations were the most frequently available virtual consultation technology (73.1%, 1002), followed by chat consultations (33.7%, 462) and video consultations (12.8%, 176). During the pandemic, telephone consultations remained the most frequently available solution (90.3%, 1238), followed by video consultations (52.3%, 717) and chat consultations (42.3%, 580).

Statistically significant increases were observed for all types of virtual consultation during the pandemic. During the pandemic, telephone consultations remained the most commonly available technology (90.4%, 1238; Figure 2). The largest absolute increase in availability was observed for video consultations (+39.5%, $P<.001$; Figure 2; Table S3 in Multimedia Appendix 1).

Of the 365 PCPs who lacked access to telephone consultations before the COVID-19 pandemic, 87.7% gained availability to this type of virtual consultation during the crisis. Correspondingly, 49.3% (n=584), and 21.8% (n=196) of PCPs who previously did not have access to video consultations, or chat consultations, respectively, reported gaining access to these technologies during the COVID-19 pandemic period.

XSL•FO
RenderX

Figure 2. Percentages of primary care physicians who reported virtual consultation technologies as available to them before versus during the COVID-19 pandemic period. The COVID-19 pandemic period was defined as the period from March 11, 2020, to the date of survey completion by the PCP (between June and September 2020). ***: adjusted *P* values <.001.

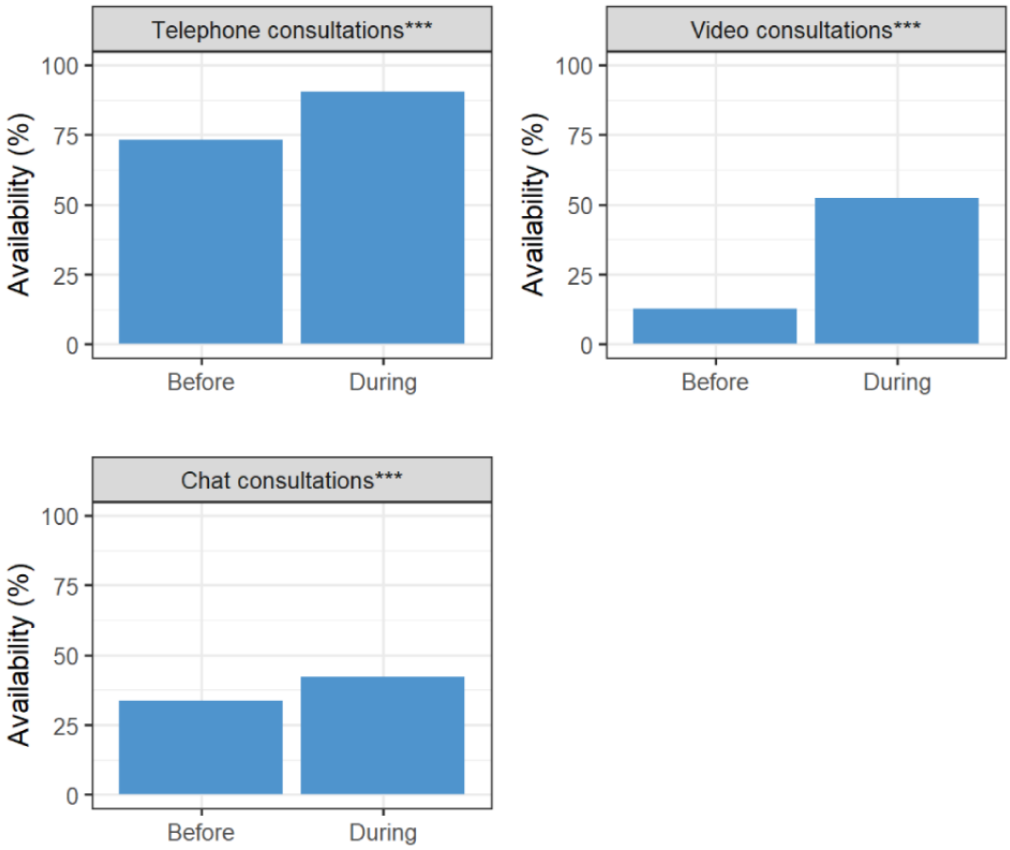

Factors Associated With Availability

Before the COVID-19 pandemic, practice urbanicity was weakly associated with availability of telephone consultations (ES 0.1, *P*<.001). Digital health training was weakly associated with availability of video (*P*<.001) and chat consultations (*P*=.04). Digital maturity score was weakly associated with increased availability of video consultations (*P*<.001) and chat consultations (*P*=.001). There was moderate to strong association between the country and availability of each of the technologies (Figure 2; ES range: 0.3-0.5, *P*<.001 for all).

During the pandemic, significant associations remained between availability and country, digital health training, and digital maturity score (Figure 2). Country persisted as significantly associated with the availability of chat consultations (*P*<.001), video consultations (*P*<.001), and telephone consultations (*P*=.03). Digital maturity score remained only weakly associated with the availability of video consultations (*P*<.001), but not with telephone or chat consultations. Digital health training was weakly associated with the availability of video (as observed before the pandemic), but also with chat consultations (*P*<.001 for both). Practice digital maturity score was no longer associated with chat consultations during the pandemic.

The strength of univariable associations between the availability of the technologies and the predictors differed before and during the COVID-19 pandemic. The strength of the associations between the country and telephone consultations decreased from strong to weak between the 2 time periods. In contrast, the strength of association between video consultations and country increased from moderate to strong.

A detailed overview of the nature of such associations is provided below.

Country Variations

Availability of chat consultations varied greatly by country for both time periods, ranging from 9% to 78.7% for before the COVID-19 pandemic, and 6.5%-75.4% during the COVID-19 pandemic. Most countries showed only small changes in availability of chat consultations from before to during the COVID-19 pandemic (Figure 3). The largest changes in availability were observed for PCPs from Chile (+38.5%), Colombia (+33.3%), Brazil (+33.3%), and the United Kingdom (+23.6%). Changes in availability of <10% were observed for PCPs from 13 of the 20 countries surveyed.

There was less variation in availability of telephone consultations between countries during than before the COVID-19 pandemic period (Figure 4). Before the pandemic, availability of telephone consultations across countries ranged from 25% to 100%, while during the pandemic, availability ranged from 78.4% to 100%. There were distinct differences in change in telephone consultation availability by country. Average availability of telephone consultations decreased amongst PCPs from countries, which reported >90% availability of telephone consultations before the COVID-19 pandemic while increases in availability were observed for all other countries.

XSL•FO
RenderX

Availability of video consultations was low across all countries before the COVID-19 pandemic period, with less than 35% of PCPs from each country having them available (range 1.1%-33.4%). Availability of video consultations increased on average for PCPs from all countries, to differing degrees by country (Figure 4). The largest increases were observed for PCPs from the United Kingdom (+81.8%), followed by PCPs from France (+71.4%), Colombia (+60.0%), and Ireland (+57.8%).

Figure 3. Absolute difference in percentage of primary care physicians from each country reporting the technology as available to them before versus during the COVID-19 pandemic. The COVID-19 pandemic period was defined as the period from March 11, 2020, to the date of survey completion by the primary care physician (between June and September 2020). PCP: primary care physician.

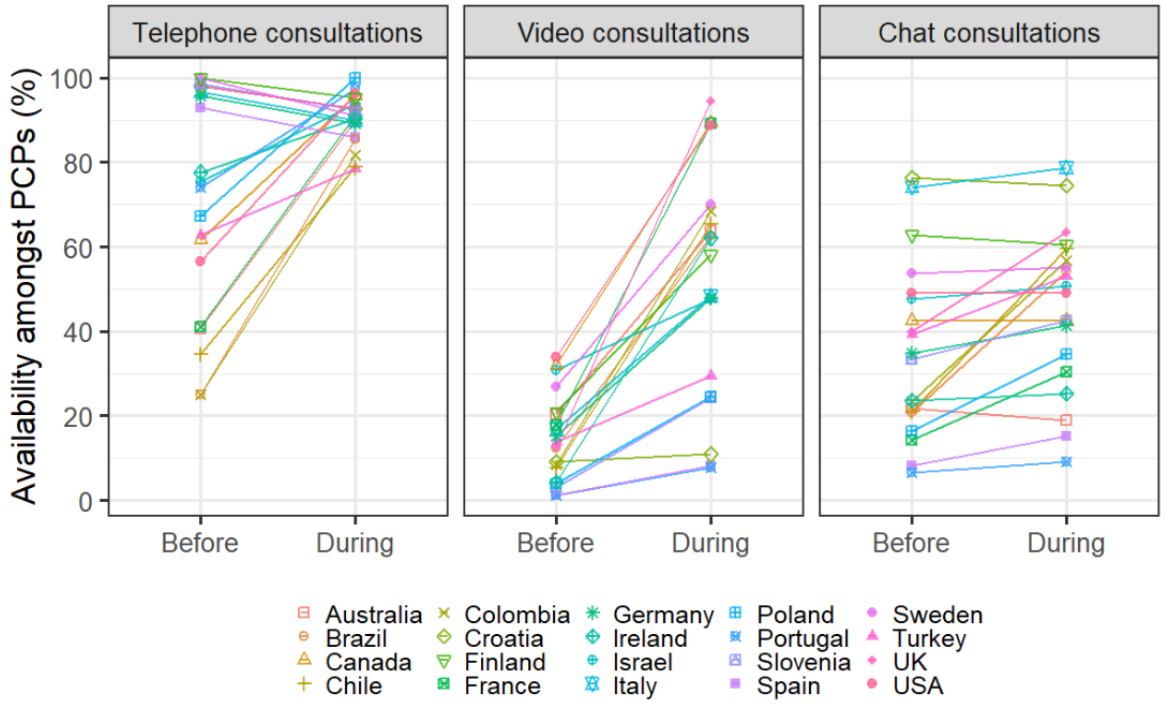

Figure 4. Effect size of primary care physician and practice factors on the reported availability of digital consultation technologies before and during the COVID-19 pandemic. The COVID-19 pandemic period was defined as the period from March 11, 2020, to the date of survey completion by the primary care physician (between June and September 2020). Effect sizes correspond to Cramer V measures of association; larger effect sizes indicate a stronger relationship between the predictor and availability. Estimates for nonsignificant relationships are not shown.

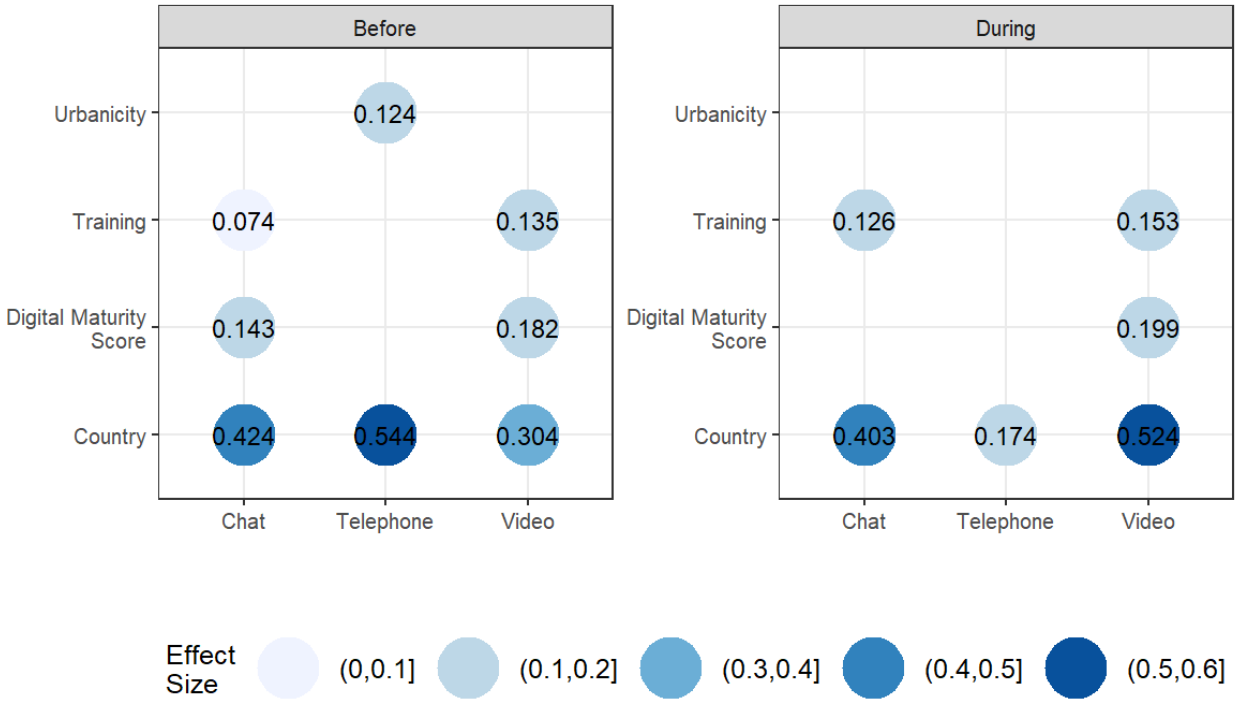

XSL•FO
RenderX

Urbanicity

Before the pandemic, the percentage of PCPs with telephone consultations available was highest amongst those from rural practices (83.4%), compared with mixed (77.5%) or urban settings (69%). This difference by practice urbanicity did not persist into the pandemic period. Availabilities of video and chat consultations were similar across PCPs from urban, mixed, and rural settings in both time periods.

Digital Health Training

Higher availability of video consultations was reported amongst PCPs who had completed, versus never completed, training in digital-first technologies, both before (18% vs 9%, $P<.001$) and during (61.1% vs 45.8%, $P<.001$) the pandemic. PCPs who had completed training reported greater availability of chat consultations before (37.8% vs 30.7%, $P=.04$) and during (49.5% vs 37%, $P<.001$) the pandemic period.

Digital Maturity

Availability of video consultations before and during the pandemic was greater amongst PCPs from more digitally mature practices. Availability of chat consultations before the pandemic was highest in PCPs from practices with a digital maturity score of 6 (42.7%), followed by 4 (39.6%) and 1 (35.7%). After adjustment for multiple testing, there was no association detected between digital maturity and availability of chat consultations during the pandemic.

Discussion

Principal Findings

Telephone consultations were the most frequently available type of virtual consultations both before and during the pandemic (73.1% and 90.4%, respectively). Significant increases in availability during the pandemic were observed for all the types of virtual consultations, alongside significant increases in hours spent on every type of virtual consultation. The largest increase in availability was observed for video consultations, whereas, although a minority of countries did display large increases in availability of chat consultations (Figure 3), a smaller change emerged in availability of chat consultations overall.

Regarding the factors associated with availability, PCPs from rural practices reported greater availability of telephone consultations before the COVID-19 pandemic but this association did not persist in the pandemic. Practice digital maturity was significantly (although weakly) associated with the availability of video consultations both before and during the pandemic, and with chat consultations before the pandemic only. Digital health training was weakly associated with the availability of both video and chat consultations, both before and during the pandemic.

There was significant country-level variation in the hours of use and availabilities of the technologies between both time periods (Figures 1 and 3). The association between country and the availability of telephone consultations changed from strong to weak, while the relationship between country and video consultations changed from moderate to strong. There was similarly strong country-level variation in availability of chat consultations in both periods.

Comparison With Previous Work

Telephone consultations were the most frequently available and used virtual consultation modality, increasing during the pandemic compared with before. Their higher use and uptake were likely driven by their lower resource requirements and maintenance costs compared with video and chat consultations [7,10]. In addition, telephones are readily available to most patients and telephone consultations were already widely used in primary care in many places (Figure 3) [4], reducing the need for additional infrastructure or training. Supporting this, telephone consultation availability was independent of practicing digital maturity level or training in digital-first technologies, unlike video or chat technologies (Figure 2).

Before the pandemic, rural PCPs reported greater availability of telephone consultations compared with PCPs from urban or mixed settings. This is unsurprising, given the benefits of virtual consultations where geographic isolation can limit health care accessibility [17]. However, during the pandemic, availability of telephone consultations became similarly high amongst PCPs from rural, mixed, and urban settings, likely attributable to the need for social distancing and consequent adoption of telephone consultations in urban areas. Future research should address whether these changes persisted in the postpandemic period.

Smaller increases were apparent in the availability and use of chat consultations during the pandemic, compared with video. This may reflect specific implementation barriers for this type of virtual consultation, alongside their perception as an adjunct to, rather than as a replacement for, other consultation methods [18]. Previous UK research found that most online consultations required in-person or telephone follow-up [19,20]. There are safety considerations with chat consultations, including the challenges of identifying patient cues solely from written communication [21]. The proportionately greater increase in availability of video is likely explained by the ability to see the patient, which contributes substantially to the confidence of professionals in making a clinical assessment [22].

Adherence to data privacy regulations poses a particular challenge for implementing chat consultations, potentially discouraging uptake. Fulfilment of the legal obligation to record and store patient information can be difficult for chat consultations, necessitating PCPs to keep separate clinical records [23-25]. Most commonly used commercial messaging systems (eg, Telegram, iMessage, and WhatsApp) do not comply with health data privacy and security regulations [23,25]. Despite potential nonadherence to ethical or data privacy guidelines, commercial instant-messaging providers are widely used for clinical purposes by patients and health care staff [24,26-28].

Among the factors examined, the country had the strongest association with availability of virtual consultation technologies and was the only significant predictor of change in hours of use. The varied ability to transition to virtual service delivery between countries is likely attributable to various governance and infrastructural factors. Some countries have national

XSL•FO
RenderX

long-term digitization goals for primary care, including strategies for virtual consultation adoption [29-31]. Coupled with guidelines on their effective and safe use [32,33], these would have facilitated greater adoption by PCPs. Countries also varied in their organizational and IT readiness to incorporate new consultation technologies into existing operations [9]. In the case of video and chat, regional variation in the availability of suitable platforms, internet coverage, and smart devices may have affected the feasibility of these consultations, contributing to a digital divide [10,31,34]. Implementation of video consultations in some countries was impeded by the need to update national health data regulations [3] and reimbursement policies [4,7,17,33]. It would be a valuable area of future work to map systems-level characteristics and explore associations between them and availability and use of virtual consultations.

Country-level variation in telephone consultation availability reduced during the pandemic compared with before, while the variation for video consultations increased (Figures 2 and 4). This indicates that the COVID-19 pandemic amplified discrepancies in barriers and facilitators of video consultation implementation between countries. Before the COVID-19 pandemic, video consultations were in the earlier stages of adoption in many countries, whereas telephone consultations were already widely available (Figure 3) and easier to scale up, for reasons previously stated [4]. Many countries continue to lack strategies for interoperability or digital education [29], likely contributing to the low adoption of video consultations post the pandemic [9]. Future research should investigate whether there have been sustained country-level differences in video consultation implementation beyond the pandemic.

Strengths and Limitations

A primary strength of our study is the large number of PCPs surveyed from 20 countries, which included a mix of urban and rural settings, during a critical transition period for primary care service delivery. However, the findings must be interpreted in light of some accompanying weaknesses. The generalizability of the study's findings may be limited by the reduced representativeness introduced by use of convenience sampling. Convenience sampling may introduce some self-selection bias for PCPs who hold stronger views about the research topic and are more vocal in sharing their experiences. Use of an anonymous web-based survey, disseminated by email and social media, prevented the identification of whether multiple respondents were employed at the same organization. In addition, due to the cross-sectional nature of the survey, we cannot establish directionality in the associations detected and causality should not be inferred. Nonetheless, these limitations are inherent to most survey-based studies and should not detract from the value of our findings.

In addition, the survey was not available in all the languages spoken by the countries surveyed which possibly excluded some PCPs from participating or affected their interpretation of questions. The study did not consider the type or size of health care organizations. Smaller practices may have incurred greater difficulties in transitioning to virtual service delivery models, particularly those in lower income areas, due to facing higher operating costs [7,10]. Another limitation is that PCPs were surveyed exclusively from upper-middle and high-income countries, restricting the generalizability of the findings to health care systems of similar economic contexts.

Finally, there have likely been significant changes in the implementation and perceptions of digital health technologies since the survey administration in 2020. An examination of availability and use of these virtual consultation technologies in more recent years, including in postpandemic contexts, would be a valuable area of future work. Nevertheless, these findings reflect a critical period for understanding the adaptability of health care systems in times of crisis.

Implications for Policy and Practice

Understanding the variations in the availability of virtual consultation technologies within and between countries is essential to ensure that their continued use does not impose additional barriers [35]. While the pandemic reduced country-level discrepancies in the availability of telephone consultations, a widening gap emerged with the availability of video consultations. Further investigation is needed to determine if these disparities reflect variations in patient, clinician, or health care organization preferences, or if they stem from digital capacity limitations.

To fully harness the potential of digital health innovations, health care providers must possess a robust understanding of their capabilities, limitations, and ethical implications. However, despite the finding of a positive relationship between training in digital first technologies and availability of chat and video consultations, less than a third (27.4%) of PCPs had completed such training. There is therefore a need for comprehensive digital health training for physicians, ensuring that they are equipped with the digital health literacy essential for delivering optimal patient care in the modern health care landscape [11].

As video consultations experienced the most significant rise, it is crucial to establish through further research whether this pattern persists in current practice, and whether this consultation modality offers substantial advantages beyond simply reducing in-person interactions. It is possible that video consultations served primarily as a tool for clinical risk mitigation during the pandemic; as restrictions on in-person appointments have stopped, the high use of video consultations may have declined [9]. This underlines the need for further studies to understand the postpandemic landscape.

Conclusions

This study highlights the significant role the COVID-19 pandemic played in driving the global adoption of virtual consultations in primary care. The increased use of virtual consultation technologies during the COVID-19 pandemic underscores the flexibility of primary care systems to adapt rapidly to the constraints imposed by the pandemic. This shift enabled continued service delivery while minimizing exposure risks for both patients and health care staff.

This research identified practice-level factors, particularly the country of practice and practice digital maturity, and digital health training, as key factors associated with the availability of these technologies. Although the COVID-19 pandemic

XSL•FO
RenderX

motivated increased usage of virtual consultations overall, it also revealed widened discrepancies between countries in their ability to implement video consultations. Systems-level research is necessary to identify the country-level facilitators and barriers toward implementation of video consultations, to ensure their continued use.

Acknowledgments

ALN, G Greenfield, GK, BH, and AM are supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Northwest London. ALN is also funded by the NIHR Patient Safety Research Collaboration, with infrastructure support from Imperial NIHR Biomedical Research Centre. The views expressed in this publication are those of the authors and not necessarily those of NIHR or the Department of Health and Social Care. Generative artificial intelligence was not used in any aspect of the manuscript preparation.

Data Availability

The datasets generated and/or analyzed in this study are available upon reasonable request from the corresponding author.

Authors' Contributions

ALN and the inSIGHT group (comprising JC, AC, CC, G Gusso, RDH, GJ, THK, LL, HL, EM, KN, DP, and RT) contributed to the conception and design of the study. GK contributed to data preparation and analysis. GK, ALN, TB, G Greenfield, and BWJH contributed to interpretation of results and drafting the manuscript. AC, CC, DP, EL, G Gusso, GJ, HL, JC, KN, LL, RDH, RT, AM, ALN, and THK revised the manuscript. All authors provided critical review and approved the final version of the manuscript.

Conflicts of Interest

BWJH also works for eConsult Health, provider of a platform for online consultations for primary, secondary and urgent or emergency care in the National Health Service.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 88 KB - jmir_v27i1e65147_app1.docx]

References

1. Car J, Koh GCH, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
2. Neves AL, van Dael J, O'Brien N, Flott K, Ghafur S, Darzi A, et al. Use and impact of virtual primary care on quality and safety: the public's perspectives during the COVID-19 pandemic. J Telemed Telecare 2024;30(2):393-401 [FREE Full text] [doi: 10.1177/1357633X211066235] [Medline: 34935535]
3. Wherton J, Greenhalgh T, Hughes G, Shaw SE. The role of information infrastructures in scaling up video consultations during COVID-19: mixed methods case study into opportunity, disruption, and exposure. J Med Internet Res 2022;24(11):e42431 [FREE Full text] [doi: 10.2196/42431] [Medline: 36282978]
4. The COVID-19 pandemic and the future of telemedicine. OECD. 2023. URL: https://doi.org/10.1787/ac8b0a27-en [accessed 2025-02-14]
5. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
6. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. BMJ Health Care Inform 2020;27(1):e100166 [FREE Full text] [doi: 10.1136/bmjhci-2020-100166] [Medline: 32565418]
7. Weigel G, Ramaswamy A, Sobel L, Salganicoff A, Cubanski J, Freed M. Opportunities and Barriers for Telemedicine in the U.S. During the COVID-19 Emergency and Beyond. KFF. 2020. URL: https://www.kff.org/womens-health-policy/issue-brief/opportunities-and-barriers-for-telemedicine-in-the-u-s-during-the-covid-19-emergency-and-beyond/ [accessed 2024-05-24]
8. Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health 2023;5:1264780 [FREE Full text] [doi: 10.3389/fdgth.2023.1264780] [Medline: 38046643]
9. Assing Hvidt E, Atherton H, Keuper J, Kristiansen E, Lüchau EC, Lønnebakke Norberg B, et al. Low adoption of video consultations in post-COVID-19 general practice in Northern Europe: barriers to use and potential action points. J Med Internet Res 2023;25:e47173 [FREE Full text] [doi: 10.2196/47173] [Medline: 37213196]

XSL•FO
RenderX

10. Chang JE, Lai AY, Gupta A, Nguyen AM, Berry CA, Shelley DR. Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-COVID era. Milbank Q 2021;99(2):340-368 [FREE Full text] [doi: 10.1111/1468-0009.12509] [Medline: 34075622]
11. Walley D, McCombe G, Broughan J, O'Shea C, Crowley D, Quinlan D, et al. Use of telemedicine in general practice in Europe since the COVID-19 pandemic: a scoping review of patient and practitioner perspectives. PLOS Digit Health 2024;3(2):e0000427 [FREE Full text] [doi: 10.1371/journal.pdig.0000427] [Medline: 38354211]
12. von Elm E, Altman M, Egger SJ, Pocock PC, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Epidemiology 2007 Nov;18(6):800-804. [doi: 10.1097/EDE.0b013e3181577654] [Medline: 18049194]
13. Neves AL, Li E, Serafini A, Jimenez G, Lingner H, Koskela TH, et al. Evaluating the impact of COVID-19 on the adoption of virtual care in general practice in 20 countries (inSIGHT): protocol and rationale study. JMIR Res Protoc 2021;10(8):e30099 [FREE Full text] [doi: 10.2196/30099] [Medline: 34292867]
14. Flott K, Callahan R, Darzi A, Mayer E. A patient-centered framework for evaluating digital maturity of health services: a systematic review. J Med Internet Res 2016;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
15. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979;6:65-70 [FREE Full text]
16. Core Team R. The R project for statistical computing. R foundation. 2023. URL: https://www.R-project.org [accessed 2025-02-14]
17. Lin CC, Dievler A, Robbins C, Sripipatana A, Quinn M, Nair S. Telehealth in health centers: key adoption factors, barriers, and opportunities. Health Aff (Millwood) 2018;37(12):1967-1974. [doi: 10.1377/hlthaff.2018.05125] [Medline: 30633683]
18. Eldh AC, Sverker A, Bendtsen P, Nilsson E. Health care professionals' experience of a digital tool for patient exchange, anamnesis, and triage in primary care: qualitative study. JMIR Hum Factors 2020;7(4):e21698 [FREE Full text] [doi: 10.2196/21698] [Medline: 33315014]
19. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018;68(666):e1-e8 [FREE Full text] [doi: 10.3399/bjgp17X693509] [Medline: 29109115]
20. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
21. Bakhai M, Atherton H. How to conduct written online consultations with patients in primary care. BMJ 2021;372:n264. [doi: 10.1136/bmj.n264] [Medline: 33627324]
22. Wanderås MR, Abildsnes E, Thygesen E, Martinez SG. Video consultation in general practice: a scoping review on use, experiences, and clinical decisions. BMC Health Serv Res 2023;23(1):316 [FREE Full text] [doi: 10.1186/s12913-023-09309-7] [Medline: 36997997]
23. Morris C, Scott RE, Mars M. WhatsApp in clinical practice-the challenges of record keeping and storage. A scoping review. Int J Environ Res Public Health 2021;18(24):13426 [FREE Full text] [doi: 10.3390/ijerph182413426] [Medline: 34949033]
24. Using mobile messaging. NHS England. 2022. URL: https://transform.england.nhs.uk/information-governance/guidance/use-mobile-messaging-software-health-and-care-settings/ [accessed 2025-02-16]
25. Information governance considerations for staff on the use of instant messaging software in the NHS Scotland. The Scottish Government. 2020. URL: https://www.informationgovernance.scot.nhs.uk/blog/2020/03/17/information-governance-considerations-for-staff-on-the-use-of-instant-messaging-software-in-n/ [accessed 2025-02-16]
26. Marin-Gomez FX, Garcia Cuyas F, Reig-Bolano R, Mendioroz J, Roura-Poch P, Pico-Nicolau M, et al. Social networking app use among primary health care professionals: web-based cross-sectional survey. JMIR Mhealth Uhealth 2018;6(12):e11147 [FREE Full text] [doi: 10.2196/11147] [Medline: 30578175]
27. Martins TGDS. Use of smartphone-based instant messaging services in medical practice: a cross-sectional study. Sao Paulo Med J 2020;138(3):269-270 [FREE Full text] [doi: 10.1590/1516-3180.2020.0228.15052020] [Medline: 32578746]
28. Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency. U.S. Department of Health and Human Services (HHS). 2020. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/emergency-preparedness/notification-enforcement-discretion-telehealth/index.html [accessed 2025-02-14]
29. Exploring the digital health landscape in the WHO European region: digital health country profiles. World Health Organization. 2024. URL: https://www.who.int/europe/publications/i/item/9789289060998 [accessed 2025-02-14]
30. LeRouge CM, Gupta M, Corpart G, Arrieta A. Health system approaches are needed to expand telemedicine use across nine latin American nations. Health Aff (Millwood) 2019;38(2):212-221. [doi: 10.1377/hlthaff.2018.05274] [Medline: 30715970]
31. Dos Santos ADF, Mata-Machado ATGD, Melo MDCBD, Fonseca Sobrinho D, Araújo LL, Silva É, et al. Implementation of telehealth resources in primary care in Brazil and its association with quality of care. Telemed J E Health 2019;25(10):996-1004. [doi: 10.1089/tmj.2018.0166] [Medline: 30592699]

XSL•FO
RenderX

32. Telemedicina durante la epidemia de COVID-19 en Chile: guías de buenas prácticas y recomendaciones. Organización Panamericana de la Salud. 2020. URL: https://www.paho.org/es/documentos/telemedicina-durante-epidemia-covid-19-chile-guia-buenas-practicas-recomendaciones-abril [accessed 2025-02-14]
33. Gunasekeran DV, Tham YC, Ting DSW, Tan GSW, Wong TY. Digital health during COVID-19: lessons from operationalising new models of care in ophthalmology. Lancet Digit Health 2021;3(2):e124-e134 [FREE Full text] [doi: 10.1016/S2589-7500(20)30287-9] [Medline: 33509383]
34. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
35. Norberg BL, Austad B, Kristiansen E, Zanaboni P, Getz LO. The impact and wider implications of remote consultations for general practice in Norway: qualitative study among norwegian contract general practitioners. JMIR Form Res 2024;8:e63068 [FREE Full text] [doi: 10.2196/63068] [Medline: 39688890]

Abbreviations

ES: effect size
PCP: primary care physician
STROBE: STrengthening the Reporting of OBservational studies in Epidemiology

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by K Thompson, T Vagg, J Wherton; comments to author 07.01.25; revised version received 24.01.25; accepted 25.01.25; published 19.03.25.

Please cite as:
Kerr G, Greenfield G, Li E, Beaney T, Hayhoe BWJ, Car J, Clavería A, Collins C, Gusso G, Hoffman RD, Jimenez G, Koskela TH, Laranjo L, Lingner H, Memarian E, Nessler K, Petek D, Tsopra R, Majeed A, Neves AL
Factors Associated With the Availability of Virtual Consultations in Primary Care Across 20 Countries: Cross-Sectional Study
J Med Internet Res 2025;27:e65147
URL: https://www.jmir.org/2025/1/e65147
doi:10.2196/65147
PMID:40105882

©Gabriele Kerr, Geva Greenfield, Edmond Li, Thomas Beaney, Benedict W J Hayhoe, Josip Car, Ana Clavería, Claire Collins, Gustavo Gusso, Robert D Hoffman, Geronimo Jimenez, Tuomas H Koskela, Liliana Laranjo, Heidrun Lingner, Ensieh Memarian, Katarzyna Nessler, Davorina Petek, Rosy Tsopra, Azeem Majeed, Ana Luisa Neves. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of a Telemedicine Model on Patients With Heart Failure With Reduced Ejection Fraction in a Resource-Limited Setting in Vietnam: Cohort Study

Hoai Thi Thu Nguyen[1,2], MD, PhD; Hieu Ba Tran[1,2], MD, MSc; Phuong Minh Tran[2], MD, MSc; Hung Manh Pham[3], MD, PhD; Co Xuan Dao[4,5], MD, PhD; Thanh Ngoc Le[5], MD, PhD; Loi Doan Do[2,3], MD, PhD; Ha Quoc Nguyen[6], PhD; Thom Thi Vu[5], PhD; James Kirkpatrick[7†], MD; Christopher Reid[8,9], PhD; Dung Viet Nguyen[1,2], MD, MSc

[1]Department of Cardiology, Faculty of Medicine, University of Medicine and Pharmacy, Vietnam National University, Hanoi, Hanoi, Vietnam
[2]Vietnam National Heart Institute, Bach Mai Hospital, Hanoi, Vietnam
[3]Department of Cardiology, Hanoi Medical University, Hanoi, Vietnam
[4]Bach Mai Hospital, Hanoi, Vietnam
[5]University of Medicine and Pharmacy, Vietnam National University, Hanoi, Hanoi, Vietnam
[6]Department of Science and Technology, Hanoi People's Committee, Hanoi, Vietnam
[7]Division of Cardiology, Department of Medicine, University of Washington Medical Center, Seattle, WA, United States
[8]School of Population Health, Curtin University, Perth, Australia
[9]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[†]deceased

Corresponding Author:
Dung Viet Nguyen, MD, MSc
Department of Cardiology
Faculty of Medicine
University of Medicine and Pharmacy, Vietnam National University, Hanoi
Y1 Building
2B Pham Van Dong Street
Hanoi, 10000
Vietnam
Phone: 84 966756865
Fax: 84 243 7450 146
Email: vietdung.ump@vnu.edu.vn

Abstract

Background: Heart failure (HF) is a complex, life-threatening condition marked by high morbidity, mortality, reduced functional capacity, poor quality of life, and substantial health care costs. HF with reduced ejection fraction (HFrEF) represents the subgroup of HF with the highest risks of mortality and hospitalization, necessitating the prioritization of care and management models to optimize treatment outcomes in these patients. Currently, data on the effectiveness of telemedicine models in resource-limited settings, such as low- and middle-income countries, are scarce.

Objective: This study aimed to evaluate the impact of telemedicine on improving prognosis in patients with HFrEF in Vietnam.

Methods: In this prospective cohort study, we recruited patients who received either remote monitoring and management (telemedicine) or standard monitoring and management (usual care) in the outpatient department of the Vietnam National Heart Institute, Bach Mai Hospital, Hanoi, Vietnam. Eligible patients were ≥18 years old, had a diagnosis of HFrEF defined as left ventricular ejection fraction (LVEF) ≤40%, had a history of HF hospitalization within the past 12 months, and presented with clinical symptoms classified as New York Heart Association (NYHA) II or III. The primary composite outcome was defined as the time to the first unplanned HF hospitalization or all-cause mortality. The follow-up period for all outcomes extended to 12 months.

Results: In total, 426 patients (298/426, 70% male; 128/426, 30% female) with a mean age of 61.3 (SD 14.6) years and a mean LVEF of 32.1% (SD 6.0%) were included in our study. Of these patients, 121 received telemedicine care, while 305 received usual care. The primary outcome occurred in 23 (23/121, 19%) patients in the telemedicine group and 82 (82/305, 26.9%) patients

XSL•FO
RenderX

in the usual care group during the follow-up period, indicating a significant reduction in risk (adjusted hazard ratio [aHR] 0.57, 95% CI 0.35-0.94; *P*=.03). However, this effect was primarily driven by a significant reduction in unplanned HF hospital admissions (aHR 0.57, 95% CI 0.33-0.98; *P*=.04) rather than in all-cause mortality (aHR 0.77, 95% CI 0.36-1.63; *P*=.49).

Conclusions: This study demonstrates that a simplified telemedicine model, even in resource-limited settings such as Vietnam, can effectively facilitate the remote monitoring and management of patients with HFrEF, resulting in significant reductions in HF-related hospitalizations and all-cause mortality.

Trial Registration: National Agency for Science and Technology Information (NASATI), Vietnam CT07/01-2022-3; https://nsti.vista.gov.vn/projects/dth/xay-dung-mo-hinh-theo-doi-va-tu-van-suc-khoe-tim-mach-tu-xa-tai-thanh-pho-ha-noi-109276.html

(J Med Internet Res 2025;27:e67228) doi:10.2196/67228

KEYWORDS

heart failure; HFrEF; telemedicine; telecare; remote monitoring; remote management; heart failure hospitalization; all-cause mortality; Vietnam; telehealth; heart; cardiology; cohort; remote; monitoring; resource-limited; cost-effective; low-cost; cardiovascular disease; hospitalization; mortality

Introduction

The universal definition of heart failure (HF) describes it as a clinical syndrome characterized by symptoms or signs caused by structural or functional cardiac abnormalities, corroborated by elevated natriuretic peptide levels or objective evidence of pulmonary or systemic congestion [1,2]. HF has long been regarded as a global pandemic, with an estimated 64 million individuals worldwide with this condition, and this number continues to rise [3]. The economic burden of HF on global health care systems is substantial and continuously increasing due to the growing incidence of HF [3]. In the United States alone, the total cost of HF is projected to reach approximately US $70 billion by 2030 [4].

Despite significant advances in HF treatment over the past decades, mortality rates remain high, with 1-year mortality rates ranging from 14% to 30% and 5-year rates ranging from 50% to 75% [3,5]. Additionally, patients with HF frequently require hospitalization due to acute exacerbations of the disease, making HF the leading cause of hospitalization for individuals older than 65 years in Europe [2,5-7]. The frequency of hospitalizations is strongly associated with disease progression and increases the mortality risk in this patient population [2,5,7,8]. HF is classified based on left ventricular ejection fraction (LVEF) into 3 categories: HF with reduced LVEF (HFrEF), HF with mildly reduced LVEF, and HF with preserved LVEF, with LVEF values of ≤40%, 41% to 49%, and ≥50%, respectively [2,8]. HFrEF accounts for approximately one-half of all patients with HF and has the highest mortality and hospitalization rates among the 3 groups [2,5]. Furthermore, guided by abundant evidence synthesized from landmark trials, the management of HFrEF based on guideline-directed medical therapy recommended by leading cardiovascular societies promises significant benefits for this patient group [2,6]. In contrast, most clinical trials have failed to identify truly effective therapies for patients with HF with mildly reduced LVEF and HF with preserved LVEF, resulting in relatively limited and low-grade evidence in the treatment recommendations for these HF subtypes. Given this context, in resource-limited settings such as Vietnam, it is essential to prioritize the development of strategies, such as telemedicine models, to optimize the care and management of high-risk patients, particularly those with HFrEF, who are more likely to derive substantial benefits from proper monitoring and care. This approach represents a more resource-efficient strategy than focusing on the other 2 HF subtypes.

Over the past decade, remote monitoring and telemedicine solutions have emerged as promising strategies to enhance HF care and management by enabling closer patient monitoring and facilitating timely, personalized clinical decision-making by health care providers [7,9,10]. Numerous clinical trials have demonstrated the benefits of telemedicine for patients with HF, particularly for reducing the risk of HF-related hospitalizations [9-12]. However, most of these studies have been conducted in high-income countries using advanced patient monitoring technologies in telemedicine models. Conversely, data on the benefits of telemedicine in low- and middle-income countries (LMIC) like Vietnam, where HF is rapidly becoming a significant health care burden, remain limited [13].

Against this backdrop, we conducted this study to evaluate the effectiveness of a simplified telemedicine model designed to accommodate the limited resources available in LMIC at improving the prognosis of patients with HFrEF in Vietnam.

Methods

Study Design, Setting, and Participants

This prospective cohort study was conducted at the Vietnam National Heart Institute (VNHI), Bach Mai Hospital, Hanoi, Vietnam. Between January 2023 and October 2023, the study recruited 2 cohorts of patients managed and monitored in the outpatient department at VNHI. One cohort consisted of patients who received remote monitoring and management (telemedicine model), while the second cohort received standard monitoring and management (usual care model). This study was registered with the National Agency for Science and Technology Information, Vietnam (registration number: CT07/01-2022-3). As the first study to evaluate the effectiveness of a telemedicine model for patients with HF in Vietnam, the findings and experiences obtained during the implementation of this telemedicine program will provide a foundation for refining the model and expanding its application to other health care facilities across the country. Bach Mai Hospital was selected as the sole study site because it is the only center classified as the highest

XSL•FO
RenderX

level in northern Vietnam's hierarchical health care system. Geographically, its location in Hanoi, the capital city and central hub of northern Vietnam, facilitates access to patients from diverse regions across the north. Furthermore, with a team of highly experienced physicians and nurses specializing in HF management and medical education, the implementation of the telemedicine model at Bach Mai Hospital was expected to not only ensure the highest quality of care for patients enrolled in this study but also enhance the potential for successful knowledge and technique transfer to lower-level health care facilities in the future.

Participants were eligible if they (1) were aged 18 years or older, (2) were diagnosed with HF according to European Society of Cardiology criteria with an LVEF ≤40% [2], (3) had been admitted to the hospital for acute HF within 12 months prior to starting the study, and (4) had a clinical presentation of NYHA classification of II or III. Patients were excluded from the study if they had severe mental or physical disorders (eg, major depression), were undergoing hemodialysis, or had been admitted to the hospital for any reason within the previous week. Additionally, the exclusion criteria included patients with a left ventricular assist device or those who had experienced a pulmonary embolism, acute stroke, cardiogenic shock, cardiac surgery, coronary revascularization, or cardiac resynchronization therapy implantation within 1 month prior to the study. Patients scheduled for coronary revascularization, transcatheter aortic valve implantation, transcatheter mitral valve repair, or cardiac resynchronization therapy implantation within the next 3 months were also excluded. Furthermore, patients with severe pulmonary diseases, including chronic obstructive pulmonary disease (stage C/D), uncontrolled asthma, or pulmonary hypertension (World Health Organization [WHO] class III-IV), as well as pregnant women, were not eligible for inclusion. In our study, the necessary costs (for patient management and provision of measurement devices) for patients in the telemedicine group were significantly higher than the costs for patients in the usual care group (where care and management continue as usual, with coverage by insurance or out-of-pocket payment by patients). To optimize the research resources, we chose a ratio of 1:2.5 rather than a balanced patient ratio of 1:1 between the 2 groups. This approach helped reduce the number of patients required in the telemedicine group, thus lowering the total costs while still ensuring sufficient statistical power. Using a sample size calculation for a time-to-event end point with a maximum follow-up period of 12 months and assuming median survival times of 17 months for the usual care group and 29 months for the treatment group, with a significance level of .05 and patient ratio of 1:2.5 between the telemedicine and usual care groups, the minimum required sample size was calculated to be 120 patients for the telemedicine group and 298 patients for the usual care group to achieve 80% power [9,14]. To account for potential withdrawals, an additional 3.5% was added, resulting in a final estimated sample size of 433 patients.

Study Procedures

The remote monitoring and management program for outpatients, referred to as the telemedicine model, was approved and implemented in the outpatient department at VNHI, Bach Mai Hospital, beginning in January 2023. Due to the unavailability of advanced specialized monitoring devices in Vietnam, our center developed a telemedicine model using simple monitoring devices that patients and caregivers can easily operate at home, such as automatic electronic blood pressure monitors and weighing scales. Throughout the study period, all potential patients seen and managed in the outpatient department were assessed by their primary treating physicians and invited to participate in the telemedicine program.

Our recruitment process consisted of 2 phases. In Phase 1, all patients meeting the inclusion criteria and consenting to participate in the study were assigned to the telemedicine group. Patients who declined participation in the telemedicine group were assigned to the usual care group. This process continued until the target number of patients for the telemedicine group (n=123) was reached. In Phase 2, all patients meeting the inclusion criteria and consenting to participate in the study were allocated to the usual care group. This phase continued until the target number of patients for the usual care group (n=310) was achieved.

Upon enrollment in the telemedicine program, patients were required to report their health status to the telemedicine center daily. The information provided included blood pressure, heart rate, and body weight measurements taken at home using validated devices (either by the patients themselves or their caregivers), as well as any new or unusual symptoms such as dyspnea, fatigue, chest pain, palpitations, swelling, cough, or fever. This reporting followed the Self-Check Plan for HF management recommended by the American Heart Association [6]. Reporting of information was encouraged to be conducted via a specialized app developed for our telemedicine program, accessible on smartphones, either by the patients themselves or their caregivers. Symptoms were entered as a checklist, while measurement results were input into designated fields, and any specific signs were reported as free-text entries. The app was designed to validate the completeness of the entered information and display warnings if essential data were missing. Additionally, the app issued daily reminders at 8 PM or a time selected by the patient, alerting them to submit their health information to the telemedicine center if they had not already done so. Information collected through this method was automatically stored in the patient's profile within the telemedicine system database. For patients without access to smartphones or those experiencing difficulties using the app, alternative reporting methods were available. Patients or caregivers could report information via SMS text messaging (including photos of measurement results displayed on devices) or through direct video or phone calls to telemedicine center nurses. In such cases, the nurses manually entered the reported information into the system database. At 8 PM daily, or at the patient's chosen time, if no health updates had been received, the system sent alerts to the telemedicine center staff, prompting them to proactively contact the patient to collect the necessary information. Patient data were securely stored in the telemedicine center's online database, accessible at any time by cardiologists on our clinical team.

Based on the daily information provided by patients and the health history recorded in the database, nurses at the

XSL•FO
RenderX

telemedicine center assessed whether a more detailed evaluation by a specialist was necessary. The following signs or symptoms (either newly developed or pre-existing but worsening during this episode) were considered indicative of an unstable condition with a high risk of acute decompensated HF, thereby necessitating a review of the patient's health status by a cardiologist: frequent cough, shortness of breath at rest, inability to lie flat due to dyspnea, confusion, dizziness, loss of appetite, greater difficulty sleeping, sadness or depression, increased swelling in the lower extremities, sudden weight gain of more than 1 kg within 24 hours or 2 kg in a week, systolic blood pressure (SBP) <90 mm Hg, diastolic blood pressure (DBP) <60 mm Hg, heart rate >100 bpm, blood pressure >180/110 mm Hg, or SBP or DBP increase or decrease >20 mm Hg compared with the average SBP or DBP of the previous day. In certain cases, an additional real-time interview with patients was arranged to collect further information. After gathering all necessary information from the patient's online medical records and direct interviews, physicians were granted autonomy to make decisions based on their clinical judgment and the European Society of Cardiology guidelines for the diagnosis and management of acute HF [2]. We did not propose a rigid protocol for physicians within this telemedicine model, as flexibility was deemed essential given the variability in patients' clinical conditions, geographic differences, and the availability of health care services in their respective areas. Physicians' decisions may include advising the patient to visit the emergency department, seek hospitalization at a cardiology center, or consult a local clinic for further evaluation (eg, electrocardiography, chest X-ray, echocardiography, laboratory tests). Other options may involve adjusting medication (limited to oral drugs), recommending lifestyle changes or home exercises, or determining that no further action is required. In more complex cases, a multidisciplinary team at the telemedicine center would discuss the situation to reach a final decision. In addition to nurse-initiated evaluations, each patient's health records were reviewed biweekly by telemedicine center physicians based on the online database.

The usual care group comprised patients who were monitored and managed through the standard process in the outpatient department at the VNHI. Throughout the study, these patients remained under the supervision of cardiologists and other physicians in the outpatient department. Treatment adjustments or prescriptions were based solely on the independent decisions of their primary treating doctors and were not influenced by the patient's participation in the study. Follow-up visits were typically scheduled every 3 months but could vary depending on the patient's condition, as assessed by the primary treating doctors.

Baseline data (including demographics; anthropometrics; clinical history; signs and symptoms; medication lists; and data from electrocardiograms, echocardiograms, and lab tests) were collected at the time of patient enrollment. BMI was classified according to the WHO diagnostic criteria for Asian adults: A BMI of 18.5-22.9 kg/m² was considered as normal weight, <18.5 kg/m² as underweight, 23-24.9 kg/m² as overweight, and ≥25 kg/m² as obese [15]. We did not use medication adherence of patients as an outcome of interest due to concerns that missing data from patients who died or were lost to follow-up could introduce significant bias.

Study Outcomes

The primary outcome of the study was defined as a composite end point consisting of the first unplanned hospitalization due to HF or all-cause mortality, evaluated over a follow-up period of up to 12 months after patient enrollment. Secondary outcomes included (1) all-cause mortality and (2) the first unplanned hospitalization specifically due to HF.

Statistical Analysis

Baseline characteristics were summarized as mean (SD) or median (IQR), as appropriate, for continuous variables and as numbers (percentages) for categorical variables. The between-group differences in continuous variables were tested using either the independent Student *t* test or the Wilcoxon rank-sum test, depending on their distribution. Categorical variables were compared using Pearson chi-square tests or Fisher exact tests, as appropriate.

All survival analyses of outcomes were conducted on a time-to-first-event basis. Cumulative incidence curves for all-cause mortality were generated using the Kaplan-Meier method to visually represent survival probabilities over time. Differences between the survival curves of the 2 groups were evaluated using the log-rank test to determine statistical significance. To assess the impact of telemedicine on the primary and secondary outcomes, multivariable Cox-proportional hazards regression models were applied, estimating hazard ratios (HRs) and the corresponding 95% CIs while adjusting for potential confounders. The confounding variables incorporated into the multivariable analysis were determined by a panel of HF experts from the Vietnam Heart Association. These variables were selected based on well-documented predictors of mortality and HF hospitalization in patients with HFrEF, as outlined in clinical guidelines and supported by previous studies [1,2,6,8,13,16]. Specifically, the variables included baseline demographic characteristics (age, sex), BMI classification, comorbidities, LVEF, the use of guideline-directed medical therapy for HFrEF, and N-terminal pro-B-type natriuretic peptide (NT-proBNP) levels. In addition, these variables, as well as the analyses of the primary composite end point and secondary end points (mortality and HF hospitalization), were prespecified before the commencement of the study. Furthermore, we conducted a sensitivity analysis, incorporating additional variables (cardiac comorbidities, relevant medications, socioeconomic status, and health literacy–related factors) into the multivariable Cox regression models to validate the robustness of the results from the primary analyses. Event rates were reported as the number of events per 100 patient-years of follow-up, accounting for any censored data due to loss to follow-up or the end of the study period.

All statistical analyses were performed using Stata SE version 18.0. All comparisons were 2-tailed, and *P* values <.05 were considered statistically significant.

XSL•FO
RenderX

Ethical Considerations

The study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the Institutional Review Board of Bach Mai Hospital, Hanoi, Vietnam (number: 1973/BM-HĐĐĐ). All participants provided written informed consent prior to enrollment, after receiving a comprehensive explanation of the study's purpose, procedures, and potential risks and benefits. To protect participant privacy, all personal identifiers were removed from the data set.

Results

Participants

A total of 426 patients with HFrEF (298/426, 70% male; mean age 61.3, SD 14.6 years), including 121 patients in the telemedicine group (81/121, 66.9% male; mean age 60.6, SD 15.4 years) and 305 patients in the usual care group (217/305, 71.2% male; mean age 61.6, SD 14.4 years), were involved in this study (Multimedia Appendix 1). Of the 273 eligible patients who initially declined participation in the study, 187 declined because they did not plan to continue HF management at Bach Mai Hospital, 41 declined due to concerns about the safety of their health during participation, 29 declined due to concerns about the additional time or costs associated with participation, and 16 declined due to concerns regarding the security of their privacy and personal information. At the end of Phase 1 of the recruitment process, a total of 123 patients were recruited into the telemedicine group, while 15 patients were assigned to the usual care group. A survey identified 2 primary reasons for patients agreeing to participate in the study but refusing to join the telemedicine group: (1) concerns about the treatment efficacy of the telemedicine model, as it was a new approach not previously implemented in Vietnam (13/15, 87%), and (2) concerns about potential additional costs associated with participation in the telemedicine model (2/15, 13%). In Phase 2, an additional 295 patients were recruited into the usual care group to reach the target number of 310. Notably, among the 15 participants who opted out of the telemedicine intervention, 11 (73%) resided within a 5-km radius of our hospital, whereas only 42 (34%) of the 123 participants included in the telemedicine group lived within this proximity. This proximity, which facilitates regular follow-up visits at the hospital, may have contributed to the participants' reluctance to choose telemedicine, a recently implemented health care model in Vietnam. The majority of patients were married (401/426, 94.1%) and resided in rural areas (259/426, 60.8%). Most reported a monthly income of US $300 to US $450 (275/426, 64.6%), slightly above the national average of US $300. Notably, all patients had health insurance (426/426, 100%), largely attributable to Vietnam's universal health care coverage policy. In Table 1, we present the detailed baseline characteristics of the participants in the 2 groups. The mean LVEF of the overall population was 32.1% (SD 6.0%) and comparable between the telemedicine group (32.1%, SD 6.0%) and the usual care group (32.1%, SD 6.0%; P=.56).

The most common comorbidities among the study participants were hypertension (128/426, 30.1%), coronary artery disease (109/426, 25.6%), chronic kidney disease (93/426, 21.8%), and diabetes (91/426, 21.4%). The mean BMI of the study population was 22.1 (SD 2.9) kg/m^2, with an obesity rate of 15.3% (65/426), based on the WHO classification for Asians, and an underweight rate of 10.6% (45/426). At the start of the study, 52.4% (223/426) of patients were being treated with an angiotensin-neprilysin inhibitor, 22.3% (95/426) with an angiotensin-converting enzyme inhibitor or angiotensin receptor blocker, 80.8% (344/426) with a sodium-glucose cotransporter-2 inhibitor, 70% (209/426) with beta-blockers, 68.5% (292/426) with a mineralocorticoid receptor antagonist (only spironolactone), 61.7% (263/426) with loop diuretics, and 7% (30/426) with thiazide-like diuretics (no patients were prescribed thiazide diuretics). Analysis results indicated no statistically significant differences in demographic characteristics, BMI, echocardiographic findings, laboratory measurements, comorbidities, socioeconomic status, health literacy, or therapeutic regimens between the telemedicine and usual care groups.

During the follow-up period, data from our telemedicine database indicated that, among the 121 patients in the telemedicine group, 75 patients (62%) reported a complete set of daily measurements on at least 90% of the total days enrolled in the telemedicine program. Additionally, 30 patients (30/121, 25.6%) provided complete measurements on 70% to 90% of the days, 13 patients (13/121, 9.9%) on 50% to 70% of the days, and only 3 patients (3/121, 2.5%) on fewer than 50% of the days.

XSL•FO
RenderX

Table 1. Baseline characteristics of study participants.

Characteristic	Total sample (n=426)	Usual care (n=305)	Telemedicine (n=121)	*P* value
Age (years), mean (SD)	61.3 (14.6)	61.6 (14.4)	60.6 (15.4)	.50
Gender (male), n (%)	298 (70)	217 (71.2)	81 (66.9)	.39
BMI (kg/m^2), mean (SD)	22.1 (2.9)	22.1 (2.9)	22.2 (2.8)	.74
BMI classifications, n (%)				.88
Underweight	45 (10.6)	32 (10.5)	13 (10.7)	
Normal weight	232 (54.5)	165 (54.1)	67 (55.4)	
Overweight	84 (19.7)	63 (20.7)	21 (17.4)	
Obesity	65 (15.3)	45 (14.8)	20 (16.5)	
Locality				.73
Rural	259 (60.8)	187 (61.3)	72 (59.5)	
Urban	167 (39.2)	118 (38.7)	49 (40.5)	
Marital status				.60
Never married	10 (2.4)	6 (2)	4 (3.3)	
Married	401 (94.1)	287 (94.1)	114 (94.2)	
Widowed/divorced	15 (3.5)	12 (3.9)	3 (2.5)	
Education				.99
School	299 (70.2)	214 (70.2)	85 (70.3)	
College/university	127 (29.8)	91 (29.8)	36 (29.8)	
Monthly income (US $), n (%)				.80
High (>450)	151 (35.5)	107 (35.1)	44 (36.4)	
Medium (300-450)	275 (64.6)	198 (64.9)	77 (63.6)	
Monthly treatment affordability (US $), n (%)				.80
High (>125)	172 (40.4)	122 (40)	50 (41.3)	
Medium (75-125)	254 (59.6)	183 (60)	71 (58.7)	
Health insurance (yes), n (%)	426 (100)	305 (100)	121 (100)	>.99
Comorbidities , n (%)				
Hypertension	128 (30.1)	85 (27.9)	43 (35.5)	.12
Diabetes	91 (21.4)	64 (21.0)	27 (22.3)	.76
Coronary artery disease	109 (25.6)	81 (26.6)	28 (23.1)	.47
Chronic kidney disease	93 (21.8)	69 (22.6)	24 (19.8)	.53
Previous stroke	25 (5.9)	20 (6.6)	5 (4.1)	.34
Atrial fibrillation	54 (12.7)	37 (12.1)	17 (14.1)	.59
Documented VT[a]	42 (9.9)	30 (9.8)	12 (9.9)	.98
Diagnosed VHD[b]	188 (44.1)	134 (43.9)	54 (44.6)	.90
Diagnosed CM[c]	162 (38.0)	119 (39.0)	43 (35.5)	.51
Pharmacotherapy , n (%)				
ARNI[d]	223 (52.4)	163 (53.4)	60 (50)	.47
ACEIs[e]/ARBs[f]	95 (22.3)	70 (23)	25 (20.7)	.61
SGLT2[g] inhibitors	344 (80.8)	247 (81)	97 (80.2)	.85
Beta-blockers	298 (70)	217 (71.2)	81 (66.9)	.39

XSL•FO
RenderX

Characteristic	Total sample (n=426)	Usual care (n=305)	Telemedicine (n=121)	*P* value
MRAs[h]	292 (68.5)	210 (68.9)	82 (67.8)	.83
Loop diuretics	263 (61.7)	184 (60.3)	79 (65.3)	.34
Thiazide-like diuretics	30 (7)	20 (6.6)	10 (8.3)	.54
Ivabradine	90 (21.1)	67 (22)	23 (19)	.50
Echocardiography findings, mean (SD)				
LVEF[i] (%)	32.1 (6.0)	32.0 (6.0)	32.4 (6.2)	.56
PAP[j] (mm Hg)	38.0 (11.3)	37.8 (10.5)	38.4 (12.9)	.64
TAPSE[k] (mm)	19.2 (2.7)	19.2 (2.7)	19.2 (2.9)	.96
FAC[l] (%)	38.9 (4.8)	38.7 (4.8)	39.3 (4.8)	.30
LAVi[m] (mL/m^2)	54.0 (26.2)	55.2 (26.9)	50.8 (24.1)	.12
Laboratory measurements , median (IQR)				
NT-proBNP[n] (pmol/L)	1064 (807-1436)	1081 (818-1458)	1024 (749-1405)	.41
Serum potassium (mmol/L)	4.2 (4.0-4.5)	4.2 (4.0-4.5)	4.1 (3.9-4.4)	.10
Serum sodium (mmol/L)	138 (136-139)	138 (136-139)	138 (136-139)	.61
Haemoglobin (g/L)	141 (129-153)	141 (128-154)	140 (131-151)	.94
eGFR[o] (mL/min/1.73m^2)	53.2 (37.4-68.7)	51.9 (36.8-67.2)	56.7 (40.5-72.8)	.13

[a]VT: ventricular tachycardia.
[b]VHD: valvular heart disease.
[c]CM: cardiomyopathy.
[d]ARNI: angiotensin-neprilysin inhibitor.
[e]ACEIs: angiotensin-converting enzyme inhibitors.
[f]ARBs: angiotensin receptor blockers.
[g]SGLT2: sodium-glucose cotransporter-2.
[h]MRAs: mineralocorticoid receptor antagonists.
[i]LVEF: left ventricular ejection fraction.
[j]PAP: pulmonary artery pressure.
[k]TAPSE: tricuspid annular plane systolic excursion.
[l]FAC: fractional area change.
[m]LAVi: left atrial volume index.
[n]NT-proBNP: N-terminal pro-B-type natriuretic peptide.
[o]eGFR: estimated glomerular filtration rate.

Primary Outcome

During the 12-month follow-up period, the primary end point occurred in 23 of 121 patients (19%) in the telemedicine group and in 82 of 305 patients (26.9%) in the usual care group. Specifically, there were 10 deaths (10/121, 8.3%) and 19 first unplanned HF hospitalizations (19/121, 15.7%) in the telemedicine group, compared with 29 deaths (29/305, 9.5%) and 71 first unplanned HF hospitalizations (71/305, 23.3%) in the usual care group. This indicates a significant reduction in the risk of the composite end point (first unplanned HF hospitalization or all-cause mortality) over the 12-month follow-up in the telemedicine group (adjusted HR 0.57, 95% CI 0.35-0.94; *P*=.03; Figure 1 and Table 2).

XSL•FO
RenderX

Figure 1. Kaplan-Meier cumulative event curve for the primary composite outcome.

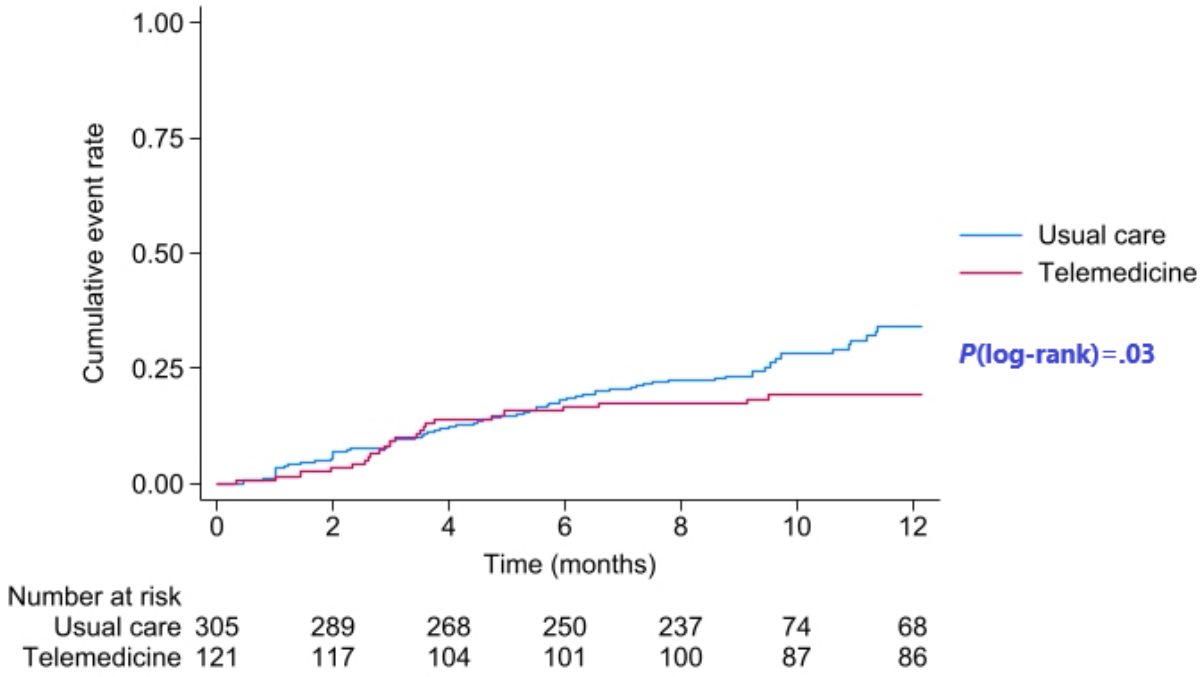

XSL•FO
RenderX

Table 2. Multivariable Cox proportional hazards regression assessing the effect of telemedicine on outcomes.

Variables	Primary outcome		All-cause mortality		Heart failure hospitalization	
	aHR[a] (95% CI)	*P* value	aHR (95% CI)	*P* value	aHR (95% CI)	*P* value
Care model						
Telemedicine	0.57 (0.35-0.94)	.03	0.77 (0.36-1.63)	.49	0.57 (0.33-0.98)	.04
Usual care	1.0 (reference)	—[b]	1.0 (reference)	—	1.0 (reference)	—
Age (years)	1.01 (1.00-1.03)	.11	1.02 (1.00-1.05)	.10	1.02 (1.00-1.04)	.06
Sex						
Male	1.26 (0.79-2.02)	.34	0.99 (0.49-2.02)	.98	1.37 (0.81-2.30)	.24
Female	1.0 (reference)	—	1.0 (reference)	—	1.0 (reference)	—
LVEF[c] (%)	0.92 (0.89-0.96)	<.001	0.98 (0.92-1.04)	.45	0.91 (0.88-0.95)	<.001
BMI classification						
Underweight	1.10 (0.58-2.08)	.77	1.19 (0.47-3.03)	.72	1.06 (0.52-2.14)	.87
Normal weight	1.0 (reference)	—	1.0 (reference)	—	1.0 (reference)	—
Overweight	0.82 (0.46-1.45)	.49	0.77 (0.28-2.10)	.61	0.93 (0.51-1.70)	.81
Obese	1.36 (0.76-2.41)	.30	1.19 (0.43-3.30)	.73	1.44 (0.76-2.72)	.26
Hypertension	0.65 (0.39-1.07)	.09	0.40 (0.16-1.00)	.051	0.76 (0.45-1.29)	.31
Diabetes	1.87 (1.17-2.98)	.009	1.04 (0.45-2.39)	.93	1.90 (1.16-3.14)	.01
CAD[d]	1.15 (0.73-1.84)	.54	0.93 (0.42-2.08)	.87	1.44 (0.89-2.34)	.14
CKD[e]	1.64 (0.99-2.72)	.056	1.78 (0.79-3.98)	.16	1.90 (1.11-3.25)	.02
Prior stroke	1.96 (0.97-3.93)	.06	3.38 (1.16-9.86)	.03	1.67 (0.77-3.63)	.19
Atrial fibrillation	1.08 (0.59-1.97)	.80	1.87 (0.80-4.35)	.15	1.10 (0.57-2.10)	.78
NT-proBNP[f] (per 200 ng/mL)	1.00 (1.00-1.01)	.37	1.00 (0.99-1.01)	.99	1.00 (1.00-1.01)	.29
Medication use						
ARNI[g] /ACEI[h] /ARB[i]						
Not used	1.0 (reference)	—	1.0 (reference)	—	1.0 (reference)	—
ACEIs/ARBs	1.82 (0.91-3.63)	.09	1.35 (0.42-4.37)	.62	1.55 (0.74-3.23)	.24
ARNI	1.83 (0.87-3.83)	.22	1.84 (0.55-6.13)	.32	1.73 (0.79-3.77)	.17
SGLT2[j] inhibitors	0.97 (0.56-1.69)	.92	0.58 (0.24-1.41)	.23	1.09 (0.59-2.03)	.77
Beta-blockers	0.59 (0.38-0.92)	.02	0.46 (0.23-0.93)	.03	0.59 (0.36-0.95)	.03
MRAs[k]	0.78 (0.44-1.40)	.41	1.27 (0.47-3.45)	.63	0.82 (0.44-1.53)	.54
Loop diuretics	1.03 (0.60-1.75)	.92	1.25 (0.51-3.08)	.63	0.92 (0.53-1.62)	.78

[a]aHR: adjusted hazard ratio.

[b]Not applicable.

[c]LVEF: left ventricular ejection fraction.

[d]CAD: coronary artery disease.

[e]CKD: chronic kidney disease.

[f]NT-proBNP: N-terminal pro b-type natriuretic peptide.

[g]ARNI: angiotensin–neprilysin inhibitor.

[h]ACEI: angiotensin-converting enzyme inhibitor.

[i]ARB: angiotensin receptor blocker.

[j]SGLT2: Sodium-glucose cotransporter-2.

[k]MRAs: mineralocorticoid receptor antagonist.

XSL•FO
RenderX

Secondary Outcomes

Regarding the secondary end points, after adjusting for covariables, telemedicine continued to demonstrate a significant reduction in the risk of HF hospitalization (adjusted HR 0.57, 95% CI 0.33-0.98; *P*=.04; Figure 2 and Table 2). Conversely, although there was a trend suggesting a slight reduction in mortality risk in the telemedicine group, the difference was not statistically significant (adjusted HR 0.77, 95% CI 0.36-1.63; *P*=.49; Figure 2 and Table 2).

Figure 2. Kaplan-Meier cumulative event curve for the secondary outcomes: (A) first unplanned heart failure hospitalization, (B) all-cause mortality.

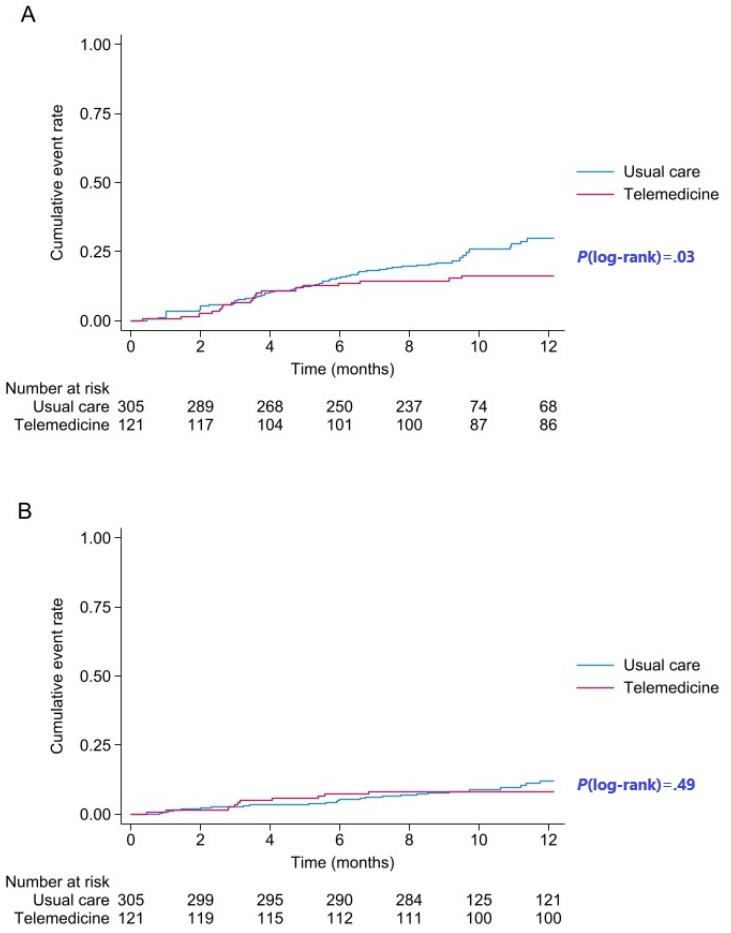

Sensitivity Analysis

The sensitivity analysis, incorporating additional variables (documented ventricular tachycardia, diagnosed valvular heart disease, diagnosed cardiomyopathy, thiazide-like use, ivabradine use, locality, marital status, education, monthly income, monthly treatment affordability) into the multivariable Cox regression models (Multimedia Appendix 2) for the outcomes of interest, demonstrated that the efficacy of telemedicine remained consistent with the primary analysis (Table 2), suggesting the robustness of the results.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to evaluate the efficacy of telemedicine for reducing mortality and HF hospitalization among patients with HFrEF in Vietnam. In this prospective cohort study, we demonstrated that a resource-limited telemedicine model, which does not require the support of advanced specialized devices, could still provide significant prognostic benefits for HFrEF patients. Specifically, patients monitored and managed through telemedicine had a 43% lower risk of experiencing the primary outcome compared with those receiving usual care (adjusted HR 0.57, 95% CI 0.35-0.94; *P*=.03). However, this composite primary outcome was primarily driven by a significant reduction in unplanned HF hospital admissions (adjusted HR 0.57, 95% CI 0.33-0.98; *P*=.04) rather than in all-cause mortality (adjusted HR 0.77, 95% CI 0.36-1.63; *P*=.49).

Comparison With Prior Work

Patients with HFrEF exhibit the highest rates of hospitalization and mortality among patients with HF [5,17,18]. Furthermore, there is a well-documented pattern of gradually increasing cardiac filling pressures, even during the “stable” phase following hospital discharge [19,20]. This progressive rise in congestion ultimately leads to the recurrence of symptoms and, consequently, the need for rehospitalization [16,21]. As a result, this population requires close monitoring and management to prevent and address exacerbations while simultaneously promoting patient self-management and empowerment. Remote monitoring and telemedicine have emerged as critical components in the management of HF, offering significant

XSL•FO
RenderX

improvements in patient outcomes while optimizing health care resource allocation [7,22]. A primary benefit is the increased accessibility to care, facilitating real-time monitoring of physiological parameters and symptoms from patients' homes [7,22]. This approach minimizes the necessity for frequent in-person clinic visits and promotes patient engagement in self-management, thereby reducing the logistical burden of travel [7]. Furthermore, evidence suggests that remote monitoring allows for the early detection of HF deterioration, enabling timely therapeutic adjustments and decreasing the risk of hospitalizations [23,24]. A meta-analysis of 29 randomized controlled trials (RCTs) published in 2019 demonstrated that telemedicine significantly improves quality of life and reduces the risk of HF-related hospitalization in patients with HF [25]. One notable RCT is the TIM-HF2 study conducted in Germany, which included 1571 patients with NYHA class II or III HF with an LVEF ≤45%, with a maximum follow-up duration of 393 days [9]. The telemedicine model used in this trial was a structured remote patient management intervention in which patients were monitored daily via a telemonitoring system that included a 3-channel electrocardiogram device, a device to measure blood pressure, a device to measure SpO_2, and weighing scales. This telemonitoring system also facilitated the transmission of data from the patient's home to the telemedical center, where specialized software classified patient risk based on preprogrammed algorithms. The subsequent actions taken by the telemedical center's physicians were informed by the patient's risk level and health data within the system. The results of the TIM-HF2 trial revealed that the remote patient management group had a significantly lower all-cause mortality rate compared with the usual care group (HR 0.70, 95% CI 0.50-0.96; *P*=.03), although there was no significant difference observed in cardiovascular mortality (HR 0.67, 95% CI 0.45-1.01; *P*=.056) [9]. More recently, the multicenter AMULET trial conducted in Poland with 605 patients with HF and an LVEF ≤49% was published in 2021 [10]. Patients in the intervention group (telecare) were monitored and managed under the AMULET model, which included regular consultations with nurses at ambulatory care points. Information collected after each visit, including HF signs and symptoms and measurements of vital parameters, showed that, after 12 months of follow-up, patients in the telecare group experienced a 31% lower rate of composite events (unplanned HF hospitalizations or cardiovascular deaths) compared with those receiving standard care (HR 0.69, 95% CI 0.48-0.99; *P*=.04) [10]. Notably, the implementation of AMULET telecare primarily reduced the risk of first hospitalizations due to HF (HR 0.64, 95% CI 0.41-0.99; *P*=.04) but did not significantly impact the risks of cardiovascular mortality (HR 1.03, 95% CI 0.54-1.98; *P*=.93) or all-cause mortality (HR 0.99, 95% CI 0.59-1.67; *P*=.98) [10].

Although direct comparisons between our study and others using different telemedicine models may be limited due to variations in research methodologies, study populations, and telemedicine processes (eg, monitoring technologies, interventions, and communication methods), the comparable benefits observed in our research underscore the viability of remote health care models adapted to resource-limited settings, such as Vietnam and other LMIC [10,22]. Despite the absence of advanced monitoring devices, our telemedicine model demonstrated good efficacy, likely due to 3 main factors. First, the model ensures that physicians can regularly monitor patients' weight, cardiovascular symptoms, and vital signs, which are the most critical information needed for early assessment and timely advice [7]. This helps to improve the prevention of congestion and reduce the risk of HF-related hospitalizations. Second, our protocol allows for direct communication between physicians and patients when necessary. This ensures the accuracy and clarity of information exchanged and enhances patient trust and adherence to treatment [7]. Finally, our study population consisted of high-risk HF patients (mean age >60 years, mean LVEF approximately 30%, with a history of HF hospitalization in the past 12 months), which may explain the significant benefit observed from the telemedicine model [7,9,23].

In high-income countries, health care systems are typically well-developed and uniformly distributed, with both urban and rural areas having well-resourced health care centers capable of providing high-quality care for a wide range of diseases. In contrast, Vietnam and most LMIC operate under a hierarchical health care model due to limited resources. In this model, only a few medical centers and hospitals in major cities receive prioritized investments, allowing them to employ highly skilled medical staff and use advanced technologies to provide high-quality care and optimize outcomes for complex medical conditions. Health care centers in rural and remote areas often receive minimal resources and lack specialists due to lower income and limited career development opportunities in these regions [26]. Consequently, these facilities can only provide basic health care services and lack the capacity to manage complex conditions like HF effectively. The implementation of telemedicine in Vietnam and other developing countries holds significant potential to overcome these limitations [26,27]. For patients in remote or underserved areas or those who face challenges in accessing specialized centers, telemedicine provides an opportunity for regular management, monitoring, and consultation with experienced physicians. This approach could improve health care quality, reduce costs and travel time, and enhance convenience for patients [22]. However, widespread adoption of telemedicine for HF management in resource-limited settings faces challenges due to the high costs of advanced monitoring devices commonly used in developed countries. Given the resource constraints in Vietnam and other developing countries, we propose a cost-effective, scalable telemedicine model that uses phone-based monitoring and affordable devices to measure essential parameters for HF management, such as blood pressure, heart rate, and body weight. The findings of this study demonstrate that this low-cost telemedicine model can provide significant benefits for patients with HFrEF in Vietnam compared with usual care. This approach may represent a viable strategy for enhancing HF care in resource-limited settings.

Strengths and Limitations

Although this study represents the first evaluation of telemedicine's efficacy in managing patients with HF in Vietnam using a simplified model designed for resource-limited settings, several limitations warrant consideration. First, as this

XSL•FO
RenderX

was not an RCT, residual confounding variables may persist despite multivariable adjustments. This limitation could introduce bias, emphasizing the need for a well-designed trial to validate our findings in the future. Second, the single-center design limits the generalizability of the findings. The results may not be applicable to other health care settings or populations, particularly in regions or countries with differing health care infrastructures. Additionally, the cohort's sociodemographic and clinical characteristics likely reflect the specific health care context in Vietnam, including unique patterns of comorbidities, health care access, and cultural factors. This may reduce the external validity of the findings when applied to different settings.

Conclusions

Our study demonstrates that a simplified telemedicine model, without the inclusion of advanced monitoring devices in resource-limited settings such as Vietnam, can still effectively monitor and manage patients with HFrEF remotely, yielding significant benefits for reducing the risk of composite outcomes. However, this effect is primarily attributed to the reduction in unplanned HF hospitalizations rather than mortality. Future studies with more robust designs conducted across diverse regions and populations are necessary to validate these findings.

Acknowledgments

This research was funded by the Department of Science and Technology of Hanoi People's Committee (grant ID: 1438/TB-SKHCN).

Data Availability

The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HTTN, CXD, TNL, HQN, HBT, LDD, HMP, TTV, and DVN conceptualized the study. HTTN, DVN, HBT, and PMT curated the data and performed the investigation. HTTN and DVN performed the formal analysis, created the visualizations, and wrote the original manuscript draft. HTTN, CXD, and HQN acquired the funding. HTTN, DVN, JK, and CR performed the methodology and reviewed and edited the manuscript. HTTN, CXD, and DVN performed project administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study Flow Chart.
[PNG File , 674 KB - jmir_v27i1e67228_app1.png]

Multimedia Appendix 2
Sensitivity Analysis.
[PDF File (Adobe PDF File), 126 KB - jmir_v27i1e67228_app2.pdf]

References

1. Bozkurt B, Coats AJ, Tsutsui H, Abdelhamid M, Adamopoulos S, Albert N, et al. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. J Card Fail 2021 Mar 01:1. [doi: 10.1016/j.cardfail.2021.01.022] [Medline: 33663906]
2. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, ESC Scientific Document Group. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021 Sep 21;42(36):3599-3726 [FREE Full text] [doi: 10.1093/eurheartj/ehab368] [Medline: 34447992]
3. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
4. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2021 update: a report from the American Heart Association. Circulation 2021 Feb 23;143(8):e254-e743 [FREE Full text] [doi: 10.1161/CIR.0000000000000950] [Medline: 33501848]
5. Shahim B, Kapelios CJ, Savarese G, Lund LH. Global public health burden of heart failure: an updated review. Card Fail Rev 2023;9:e11 [FREE Full text] [doi: 10.15420/cfr.2023.05] [Medline: 37547123]
6. Writing Committee Members, ACC/AHA Joint Committee Members. 2022 AHA/ACC/HFSA guideline for the management of heart failure. J Card Fail 2022 May;28(5):e1-e167. [doi: 10.1016/j.cardfail.2022.02.010] [Medline: 35378257]

XSL•FO
RenderX

7. Stevenson LW, Ross HJ, Rathman LD, Boehmer JP. Remote monitoring for heart failure management at home. J Am Coll Cardiol 2023 Jun 13;81(23):2272-2291 [FREE Full text] [doi: 10.1016/j.jacc.2023.04.010] [Medline: 37286258]
8. Nguyen DV, Le TN, Truong BQ, Nguyen HTT. Efficacy and safety of angiotensin receptor-neprilysin inhibition in heart failure patients with end-stage kidney disease on maintenance dialysis: a systematic review and meta-analysis. Eur J Heart Fail 2025 Jan;27(1):72-84. [doi: 10.1002/ejhf.3454] [Medline: 39252524]
9. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
10. Krzesiński P, Jankowska EA, Siebert J, Galas A, Piotrowicz K, Stańczyk A, et al. Effects of an outpatient intervention comprising nurse-led non-invasive assessments, telemedicine support and remote cardiologists' decisions in patients with heart failure (AMULET study): a randomised controlled trial. Eur J Heart Fail 2022 Mar;24(3):565-577 [FREE Full text] [doi: 10.1002/ejhf.2358] [Medline: 34617373]
11. Bekfani T, Fudim M, Cleland JG, Jorbenadze A, von Haehling S, Lorber A, et al. A current and future outlook on upcoming technologies in remote monitoring of patients with heart failure. Eur J Heart Fail 2021 Jan 18;23(1):175-185 [FREE Full text] [doi: 10.1002/ejhf.2033] [Medline: 33111389]
12. Dierckx R, Inglis SC, Clark RA, Prieto-Merino D, Cleland JG. Telemedicine in heart failure: new insights from the Cochrane meta-analyses. Eur J Heart Fail 2017 Mar 01;19(3):304-306 [FREE Full text] [doi: 10.1002/ejhf.759] [Medline: 28251777]
13. Nguyen HTT, Ha TTT, Tran HB, Nguyen DV, Pham HM, Tran PM, et al. Relationship between BMI and prognosis of chronic heart failure outpatients in Vietnam: a single-center study. Front Nutr 2023;10:1251601 [FREE Full text] [doi: 10.3389/fnut.2023.1251601] [Medline: 38099185]
14. Chow SC, Shao J, Wang H, Lokhnygina Y. Sample Size Calculations in Clinical Research. New York, NY: Taylor & Francis; 2017.
15. Regional Office for the Western Pacific. The Asia-Pacific perspective: redefining obesity and its treatment. World Health Organization. 2000. URL: https://iris.who.int/handle/10665/206936 [accessed 2025-03-14]
16. Nohria A, Tsang SW, Fang JC, Lewis EF, Jarcho JA, Mudge GH, et al. Clinical assessment identifies hemodynamic profiles that predict outcomes in patients admitted with heart failure. J Am Coll Cardiol 2003 May 21;41(10):1797-1804 [FREE Full text] [doi: 10.1016/s0735-1097(03)00309-7] [Medline: 12767667]
17. Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. Heart 2022 Aug 11;108(17):1351-1360 [FREE Full text] [doi: 10.1136/heartjnl-2021-320131] [Medline: 35042750]
18. Bozkurt B, Ahmad T, Alexander KM, Baker WL, Bosak K, Breathett K, Writing Committee Members. Heart failure epidemiology and outcomes statistics: a report of the Heart Failure Society of America. J Card Fail 2023 Oct;29(10):1412-1451 [FREE Full text] [doi: 10.1016/j.cardfail.2023.07.006] [Medline: 37797885]
19. Zile MR, Bennett TD, St. John Sutton M, Cho YK, Adamson PB, Aaron MF, et al. Transition from chronic compensated to acute decompensated heart failure. Circulation 2008 Sep 30;118(14):1433-1441. [doi: 10.1161/circulationaha.108.783910]
20. Adamson PB. Pathophysiology of the transition from chronic compensated and acute decompensated heart failure: new insights from continuous monitoring devices. Curr Heart Fail Rep 2009 Dec;6(4):287-292. [doi: 10.1007/s11897-009-0039-z] [Medline: 19948098]
21. Punnoose LR, Gopal DM, Stevenson LW. Time to update our profiles. Eur J Heart Fail 2019 Nov;21(11):1366-1369 [FREE Full text] [doi: 10.1002/ejhf.1549] [Medline: 31321859]
22. Piskulic D, McDermott S, Seal L, Vallaire S, Norris CM. Virtual visits in cardiovascular disease: a rapid review of the evidence. Eur J Cardiovasc Nurs 2021 Nov 28;20(8):816-826 [FREE Full text] [doi: 10.1093/eurjcn/zvab084] [Medline: 34632501]
23. Ploux S, Strik M, Ramirez FD, Buliard S, Chauvel R, Dos Santos P, et al. Remote management of worsening heart failure to avoid hospitalization in a real-world setting. ESC Heart Fail 2023 Dec;10(6):3637-3645 [FREE Full text] [doi: 10.1002/ehf2.14553] [Medline: 37797957]
24. Kuan PX, Chan WK, Fern Ying DK, Rahman MAA, Peariasamy KM, Lai NM, et al. Efficacy of telemedicine for the management of cardiovascular disease: a systematic review and meta-analysis. The Lancet Digital Health 2022 Sep;4(9):e676-e691. [doi: 10.1016/s2589-7500(22)00124-8]
25. Takeda A, Martin N, Taylor RS, Taylor SJ. Disease management interventions for heart failure. Cochrane Database Syst Rev 2019 Jan 08;1(1):CD002752 [FREE Full text] [doi: 10.1002/14651858.CD002752.pub4] [Medline: 30620776]
26. Eberly LA, Tennison A, Mays D, Hsu C, Yang C, Benally E, et al. Telephone-based guideline-directed medical therapy optimization in Navajo nation: the Hózhó randomized clinical trial. JAMA Intern Med 2024 Jun 01;184(6):681-690 [FREE Full text] [doi: 10.1001/jamainternmed.2024.1523] [Medline: 38583185]
27. Takahashi EA, Schwamm LH, Adeoye OM, Alabi O, Jahangir E, Misra S, American Heart Association Council on Cardiovascular RadiologyIntervention, Council on Hypertension, Council on the Kidney in Cardiovascular Disease,Stroke Council. An overview of telehealth in the management of cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2022 Dec 20;146(25):e558-e568. [doi: 10.1161/CIR.0000000000001107] [Medline: 36373541]

XSL•FO
RenderX

Abbreviations

DBP: diastolic blood pressure
HF: heart failure
HFrEF: heart failure with reduced ejection fraction
HR: hazard ratio
LMIC: low- and middle-income countries
LVEF: left ventricular ejection fraction
NT-proBNP: N-terminal pro-B-type natriuretic peptide
NYHA: New York Heart Association
RCT: randomized controlled trial
SBP: systolic blood pressure
VNHI: Vietnam National Heart Institute
WHO: World Health Organization

Edited by A Coristine; submitted 18.10.24; peer-reviewed by T Kaas; comments to author 23.12.24; revised version received 12.01.25; accepted 19.02.25; published 19.03.25.

Please cite as:
Nguyen HTT, Tran HB, Tran PM, Pham HM, Dao CX, Le TN, Do LD, Nguyen HQ, Vu TT, Kirkpatrick J, Reid C, Nguyen DV
Effect of a Telemedicine Model on Patients With Heart Failure With Reduced Ejection Fraction in a Resource-Limited Setting in Vietnam: Cohort Study
J Med Internet Res 2025;27:e67228
URL: https://www.jmir.org/2025/1/e67228
doi:10.2196/67228
PMID:

©Hoai Thi Thu Nguyen, Hieu Ba Tran, Phuong Minh Tran, Hung Manh Pham, Co Xuan Dao, Thanh Ngoc Le, Loi Doan Do, Ha Quoc Nguyen, Thom Thi Vu, James Kirkpatrick, Christopher Reid, Dung Viet Nguyen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Differences in Telemedicine Use for Patients With Diabetes in an Academic Versus Safety Net Health System: Retrospective Cohort Study

Jonathan J Shih[1*], BS; Magdalene Kuznia[2*], MS-HCI, RN; Sarah Nouri[3], MPH, MD; Elizabeth B Sherwin[4], MPH; Kathryn E Kemper[4,5], MPH; Anna D Rubinsky[6], PhD; Courtney R Lyles[7,8], PhD; Elaine C Khoong[5,9], MS, MD

[1]School of Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[3]Division of Palliative Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[4]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States

[5]UCSF Action Research Center for Health Equity, University of California, San Francisco, San Francisco, CA, United States

[6]Academic Research Services, Information Technology, University of California, San Francisco, San Francisco, CA, United States

[7]Center for Healthcare Policy and Research, University of California, Davis, Davis, CA, United States

[8]Department of Public Heath Science, University of California, Davis, Davis, CA, United States

[9]Department of Medicine, Division of General Internal Medicine at Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Elaine C Khoong, MS, MD
Department of Medicine
Division of General Internal Medicine at Zuckerberg San Francisco General Hospital
University of California, San Francisco
2540 23rd Street, Room 4708
San Francisco, CA, 94110
United States
Phone: 1 415 502 6300
Fax: 1 415 443 4584
Email: Elaine.Khoong@ucsf.edu

Abstract

Background: The COVID-19 public health emergency catalyzed widespread adoption of both video- and audio-only telemedicine visits. This proliferation highlighted inequities in use by age, race and ethnicity, and preferred language. Few studies have investigated how differences in health system telemedicine implementation affected these inequities.

Objective: This study aims to describe patients who used telemedicine during the public health emergency and identify predictors of telemedicine use across 2 health systems with different telemedicine implementations.

Methods: This retrospective cohort study included adults with diabetes receiving primary care between July 2020 and March 2021 at 2 independent health systems in San Francisco, California. Participant sociodemographic characteristics, health information, and telemedicine utilization were acquired from electronic health records. The primary outcome was visit type (any audio or video telemedicine vs in-person only) during the study period. We used multivariable logistic regression to assess the association between visit type and key predictors associated with digital exclusion (age, race and ethnicity, preferred language, and neighborhood socioeconomic status), adjusting for baseline health. We included an interaction term to evaluate health system impact on each predictor and then stratified by health system (academic, which prioritized video-enabled visits, vs safety net, which prioritized audio-only visits).

Results: Among 10,201 patients, we found higher odds of telemedicine use in the safety net system compared with the academic system (adjusted odds ratio [aOR] 2.94, 95% CI 2.48-3.48). Patients with younger age (18-34 years: aOR 2.55, 95% CI 1.63-3.97; 35-49 years: aOR 1.39, 95% CI 1.12-1.73 vs 75+ years) and Chinese-language preference (aOR 2.04, 95% CI 1.66-2.5 vs English) had higher odds of having a telemedicine visit. Non-Hispanic Asian (aOR 0.67, 95% CI 0.56-0.79), non-Hispanic Black (aOR

XSL•FO
RenderX

0.83, 95% CI 0.68-1), and Hispanic or Latine (aOR 0.76, 95% CI 0.61-0.95) patients had lower odds of having a telemedicine visit than non-Hispanic White patients. We found significant interactions between health system and age, race and ethnicity, and preferred language (P<.05). After stratifying by health system, several differences persisted in the academic system: non-Hispanic Asian (aOR 0.57, 95% CI 0.46-0.70) and Latine (aOR 0.67, 95% CI 0.50-0.91) patients had lower odds of a telemedicine visit, and younger age groups had higher odds (18-34 years: aOR 3.97, 95% CI 1.99-7.93; 35-49 years: aOR 1.86, 95% CI 1.36-2.56). In the safety net system, Chinese-speaking patients had higher odds of having a telemedicine visit (aOR 2.52, 95% CI 1.85-3.42).

Conclusions: We found disparities in telemedicine utilization by age, race and ethnicity, and preferred language, primarily in the health system that used more video visits. While telemedicine expanded rapidly recently, certain populations remain at risk for digital exclusion. These findings suggest that system-level factors influence telemedicine adoption and implementation decisions impact accessibility for populations at risk for digital exclusion.

(J Med Internet Res 2025;27:e64635) doi:10.2196/64635

KEYWORDS

telehealth; telemedicine; diabetes; delivery of health care; primary health care; health care utilization; health care disparities; chronic disease; COVID-19; safety net providers; public health emergency; older adults

Introduction

Prior to the COVID-19 public health emergency (PHE), the adoption of telemedicine (defined here as real-time virtual visits between patients and clinicians) in the United States was limited due to both policy and implementation challenges [1,2]. The PHE served as a catalyst for widespread adoption: telemedicine policies expanded coverage to include reimbursement parity for both video- and audio-only visits, with notably high adoption of audio-only visits in safety net settings [3-5]. The proliferation of telemedicine, however, also introduced important concerns about equity [6]. Persistent inequities in utilization and access have been documented, with studies indicating that older adults, racial and ethnic minority groups, individuals with limited English proficiency, and those with lower income are less likely to adopt telemedicine, particularly video-enabled services [7-12]. These patterns of underutilization highlight the broader issues of digital exclusion, driven by challenges such as limited digital literacy and lack of access to technology and web-based service [6,12].

However, an understudied aspect in these studies is how health system factors and their implementation of telemedicine impact these disparities. Audio visits, provided more frequently by community health centers and safety net systems, are predominantly used by low-income, older, or racially and ethnically minoritized patients [3]. It remains unclear whether these patterns in telemedicine utilization are a result of patient characteristics or reflective of health systems being underresourced. Understanding the drivers of these patterns of telemedicine utilization is crucial, as it informs the development of interventions that address disparities and enhance equitable distribution of telehealth services.

To address this gap in the literature, we evaluated telemedicine use in 2 distinct health systems in San Francisco, California: University of California, San Francisco (UCSF), an academic medical system, and San Francisco Health Network (SFHN), a safety net system. Telemedicine implementation during the PHE differed across the systems. UCSF started implementing video visits and training prior to the PHE and encouraged video as the predominant modality during the PHE, resulting in 50% of all visits (including in-person and telemedicine) being video visits [13]. In contrast, SFHN encountered delays in video implementation and did not prioritize video visits, resulting in telemedicine visits being primarily delivered as audio-only encounters (>95%) [14]. In this study, we aimed to (1) describe characteristics of patients with and without a telemedicine visit during the PHE in each health system, and (2) evaluate associations between key sociodemographic factors—age, race and ethnicity, preferred language, and neighborhood socioeconomic status (nSES)—and telemedicine use, including health system interaction effects. We sought to identify whether disparities in telemedicine use were driven by patient characteristics and by variations in the implementation of telemedicine across systems.

Methods

Study Setting and Data Source

This retrospective cohort study included patients with diabetes receiving primary care between July 1, 2020, and March 31, 2021, at 1 of 2 independent health care systems in San Francisco: UCSF and SFHN. UCSF is an academic tertiary care medical care center with 4 primary care practices. SFHN is a safety net health system for San Francisco County with 14 primary care practices.

Study Sample

The study cohort was identified using electronic health records (EHRs) and included adult patients (aged 18 years and older) with diabetes mellitus who had at least 1 primary care or endocrinology encounter between July 1, 2020, and March 31, 2021 (hereafter referred to as the study period). Participants were required to have a diagnosis of diabetes prior to the beginning of the study period. In addition, participants must have been empaneled in primary care as of April 1, 2019 (approximately 1 year before start of the PHE), defined as having assigned a primary care provider and at least 1 primary care encounter within the 3 years prior to April 1, 2019. Details on this cohort are provided in a prior publication [14].

XSL•FO
RenderX

Ethical Considerations

Data were deidentified for analysis and reporting. This study was approved by the UCSF institutional review board (#20-31253) with waiver of informed consent.

Variables

Participants' sociodemographic characteristics, health information (eg, comorbidities, vitals, and laboratory results), and telemedicine utilization were derived from EHRs. The primary outcome was defined as whether participants had any telemedicine visit (audio- or video-only visit vs in-person visit only) with their diabetes care team (primary care or endocrinology) during the study period. Key predictor variables were as follows: age (18-34, 35-49, 50-64, 65-74, and ≥75 years); race and ethnicity (non-Hispanic Asian, non-Hispanic Black, Hispanic or Latine, non-Hispanic White, and non-Hispanic other race); preferred language (English, Spanish, Chinese, and other); and nSES quintile at the census tract level based on geocoded residential address [14]. nSES is a validated composite SES measure that incorporates 5-year census tract level data from the 2013-2017 American Community Survey on income, education, poverty, employment, occupation, housing, and rent values [15]. nSES quintiles are based on the distribution of nSES across US Census tracts from Bay Area counties, with the first quintile (Q1) reflecting lowest nSES and the fifth quintile (Q5) reflecting highest nSES.

Covariates included average baseline hemoglobin A_{1c} control, baseline blood pressure (BP) control (≤120/80, 121/81-140/90, and >140/90 mm Hg), patient sex (male or female), insurance type (Medicare, Medicaid, commercial, Healthy Workers [county administered], or uninsured), patient portal enrollment (yes or no, as a proxy for digital access), and Charlson Comorbidity Index (CCI) score category (0-2 or 3+). Hemoglobin A_{1c} control, BP control, and CCI were used to adjust for differences in utilization based on medical need.

Statistical Analysis

We described patient-level differences in telemedicine use associated with each predictor, overall and by health care system, using chi-square tests for categorical predictors and 2-tailed *t* tests for continuous predictors. Our regression analyses included a multivariable logistic regression model using the combined health system data to determine the adjusted effects of key predictors on telemedicine utilization and a second model using an interaction term to estimate the effect of health system on each of our key predictors (age, race and ethnicity, preferred language, and nSES). We calculated predicted margins for the significant key predictors. Next, we stratified models by health system. Each logistic regression model controlled for our listed covariates. Missing data for hemoglobin A_{1c} values and BP measurements were imputed using multiple imputation by chained equations [16]. Results were presented as adjusted odds ratios (aORs) with 95% CIs. Two-tailed significance was determined with α value of .05. All analyses were performed using Stata 17 [17].

Results

Included Participants

The final study cohort included a total of 10,201 adult patients with diabetes across the 2 health systems (Table 1 and Multimedia Appendix 1). As shown in Multimedia Appendix 1, there was different racial and ethnic, linguistic, and nSES distribution in the 2 health systems. The academic health system had mostly non-Hispanic White and non-Hispanic Asian patients (929/3623, 26%, and 1430/3623, 39%, respectively), while the safety net system had more non-Hispanic Asian and Hispanic or Latine patients (2302/6578, 35%, and 2244/6578, 34%, respectively). Most of the patients in the academic health system had an English-language preference (2935/3623, 81%); however, less than half of the patients at the safety net system preferred the English language (2847/6578, 43%). Furthermore, the safety net system mostly took care of patients living in the lowest nSES quintile (2551/6578, 39%), whereas the academic health system had about 20% (Q1: 640/3623, 17.7%; Q2: 682/3623, 188.%; Q3: 716/3623, 198.%; Q4: 883/3623, 24.4%; and Q5: 702/3623, 19.4%) of patients in each nSES quintile.

Of the 10,201 patients in the cohort, 81% (n=8305) had at least 1 telemedicine visit during the study period. Patients in the safety net system were more likely to use telemedicine than patients in the academic health system (5655/6578, 86% vs 2650/3623, 73%; *P*<.001). Among the combined health systems, there were significant differences in telemedicine use by age (*P*<.001), language preference (*P*<.001), and nSES quintile (*P*<.001). Within the academic health system, there were only significant differences in telemedicine use by age group, with younger patients showing higher adoption (*P*<.001), and by racial or ethnic groups, with lower adoption by non-Hispanic Asian patients (*P*<.001). The safety net system showed significant differences in telemedicine use by racial and ethnic groups, with highest adoption by non-Hispanic Asian patients and lowest by non-Hispanic Black patients (*P*<.001), and by language preferences, with highest adoption by Chinese speakers (*P*<.001).

XSL•FO
RenderX

Table 1. Patient characteristics by health system among patients with any visits.

Patient characteristics	Combined health systems (N=10,201)			Academic system (n=3623)			Safety net system (n=6578)		
	In-person visits only	Telemedicine visit	*P* value	In-person visits only	Telemedicine visit	*P* value	In-person visits only	Telemedicine visit	*P* value
Patients, n (%)	1896 (19)	8305 (81)		973 (27)	2650 (73)		923 (14)	5655 (86)	
Age group (years), n (%)			<.001			<.001			.16
18-34	26 (10)	223 (90)		10 (10)	90 (90)		16 (11)	133 (89)	
35-49	228 (16)	1198 (84)		80 (19)	346 (81)		148 (15)	852 (85)	
50-64	731 (18)	3375 (82)		285 (26)	797 (74)		446 (15)	2578 (85)	
65-74	552 (20)	2192 (80)		320 (31)	712 (69)		232 (14)	1480 (86)	
75+	359 (21)	1317 (79)		278 (28)	705 (72)		81 (12)	612 (88)	
Race and ethnicity, n (%)			.47			<.001			<.001
Non-Hispanic White	306 (18)	1351 (82)		201 (22)	728 (78)		105 (14)	623 (86)	
Non-Hispanic Asian	706 (19)	3026 (81)		459 (32)	971 (68)		247 (11)	2055 (89)	
Non-Hispanic Black or African American	285 (19)	1192 (81)		119 (24)	371 (76)		166 (17)	821 (83)	
Hispanic or Latine	469 (18)	2211 (83)		112 (26)	324 (74)		357 (16)	1887 (84)	
Other or unknown	130 (20)	525 (80)		82 (24)	256 (76)		48 (15)	269 (85)	
Language, n (%)			<.001			.17			<.001
English	1223 (21)	4541 (79)		768 (26)	2149 (74)		455 (16)	2392 (84)	
Spanish	315 (16)	1630 (84)		28 (24)	91 (76)		287 (16)	1539 (84)	
Chinese	176 (11)	1371 (89)		94 (32)	203 (68)		82 (7)	1168 (93)	
Other or unknown	182 (19)	763 (81)		83 (29)	207 (71)		99 (15)	556 (85)	
nSES[a] quintiles, n (%)			<.001			.26			.05
1 (lowest)	537 (17)	2654 (83)		156 (24)	484 (76)		381 (15)	2170 (85)	
2	421 (17)	2005 (83)		173 (25)	509 (75)		248 (14)	1496 (86)	
3	324 (19)	1413 (81)		206 (29)	510 (71)		118 (12)	903 (88)	
4	364 (21)	1396 (79)		251 (28)	632 (72)		113 (13)	764 (87)	
5 (highest)	250 (23)	837 (77)		187 (27)	515 (73)		63 (16)	322 (84)	
Baseline hemoglobin A_{1c} (%), mean (SD)	7.5 (1.7)	7.6 (1.7)	.02	7.2 (1.3)	7.4 (1.6)	<.001	7.9 (2.1)	7.7 (1.8)	.02
Blood pressure (mm Hg), n (%)			.17			.91			.59
≤120/80	402 (18)	1856 (82)		175 (26)	488 (74)		227 (14)	1368 (86)	
≤140/90	992 (18)	4471 (82)		526 (27)	1451 (73)		466 (13)	3020 (87)	
>140/90	413 (20)	1670 (80)		242 (27)	645 (73)		171 (14)	1025 (86)	
Sex, n (%)			<.001			.02			<.001
Female	916 (17)	4,494 (83)		502 (25)	1480 (75)		414 (12)	3014 (88)	
Male	980 (20)	3811 (80)		471 (29)	1170 (71)		509 (16)	2641 (84)	
Insurance, n (%)			<.001		.04			<.001	
Private	448 (27)	1184 (73)		443 (28)	1124 (72)		5 (8)	60 (92)	
Public	529 (17)	2512 (83)		102 (22)	354 (78)		427 (17)	2158 (83)	
Medicare	721 (19)	3083 (81)		428 (27)	1172 (73)		293 (13)	1911 (87)	
Uninsured	111 (15)	650 (85)		N/A[b]	N/A		111 (15)	650 (85)	
Healthy Workers[c]	87 (9)	876 (91)		N/A	N/A		87 (9)	876 (91)	

XSL•FO
RenderX

Patient characteristics	Combined health systems (N=10,201)			Academic system (n=3623)			Safety net system (n=6578)		
	In-person visits only	Telemedicine visit	*P* value	In-person visits only	Telemedicine visit	*P* value	In-person visits only	Telemedicine visit	*P* value
Patient portal enrollment, n (%)			<.001			<.001			.007
Activated	746 (22)	2655 (78)		685 (24)	2126 (76)		61 (10)	529 (90)	
Charlson Comorbidity Index, n (%)			.18			<.001			.06
0-2	1531 (19)	6591 (81)		785 (29)	1879 (71)		746 (14)	4712 (86)	
3+	365 (18)	1714 (82)		188 (20)	771 (80)		177 (16)	943 (84)	

[a]nSES: neighborhood socioeconomic status.

[b]N/A: not applicable.

[c]Healthy Workers is an insurance plan provided to temporary exempt employees of the City and County of San Francisco, including providers of in-home care.

Adjusted Analyses

Combined Health System Data

After adjusting for key predictors (age, race and ethnicity, language, and nSES) and covariates (baseline hemoglobin A_{1c}, baseline BP, sex, insurance type, patient portal enrollment status, and CCI), safety net patients had higher odds of having a telemedicine visit compared with academic health system patients (aOR 2.94, 95% CI 2.48-3.48; Table 2). Patients aged 18-34 and 35-49 years also had higher odds of having a telemedicine visit than patients older than 75 years (aOR 2.55, 95% CI 1.63-3.97; aOR 1.39, 95% CI 1.12-1.73). Non-Hispanic Asian, non-Hispanic Black or African American, and Hispanic or Latine patients had lower odds of using telemedicine than non-Hispanic White patients (aOR 0.67, 95% CI 0.56-0.79; aOR 0.83, 95% CI 0.68-1.00; aOR 0.76, 95% CI 0.61-0.95, respectively), whereas patients with Chinese as their preferred language had higher odds of using telemedicine than those with an English-language preference (aOR 2.04, 95% CI 1.66-2.50). No significant differences were found among nSES. Results are displayed in Table 2. There was a significant interaction between health care system and each individual-level key predictor ($P<.05$; Multimedia Appendix 2). Health system significantly modified the effect of telemedicine use across age, racial and ethnic, and language preference groups, which is shown in Figure 1A-C, respectively. Estimated predicted probabilities for significant key predictors after applying the health system interaction term are provided in Multimedia Appendix 3. The interaction between health system and nSES was not statistically significant.

XSL•FO
RenderX

Table 2. Adjusted odds ratios of predictors of telemedicine use during the public health emergency.

Predictors	Combined health systems (N=10,201)		Academic system (N=3623)		Safety net system (N=6578)	
	aOR[a]	95% CI	aOR	95% CI	aOR	95% CI
Health system (reference: UCSF[b])						
SFHN[c]	2.94[d]	2.48-3.48	N/A[e]	N/A	N/A	N/A
Age group (years; reference: 75+)						
18-34	2.55[d]	1.63-3.97	3.97[d]	1.99-7.93	1.42	0.78-2.60
35-49	1.39[f]	1.12-1.73	1.86[d]	1.36-2.56	0.98	0.70-1.37
50-64	1.09	0.92-1.30	1.15	0.91-1.45	0.87	0.65-1.16
65-74	0.95	0.81-1.12	0.91	0.74-1.11	0.84	0.64-1.11
Race and ethnicity (reference: White)						
Asian	0.67[d]	0.56-0.79	0.57[d]	0.46-0.70	0.88	0.67-1.17
Black or African American	0.83[g]	0.68-1.00	0.81	0.61-1.07	0.90	0.68-1.19
Hispanic or Latine	0.76[f]	0.61-0.95	0.67[g]	0.50-0.91	0.86	0.62-1.18
Other or unknown	0.89	0.70-1.13	0.86	0.63-1.16	0.94	0.64-1.38
Language (reference: English)						
Spanish	1.04	0.83-1.30	1.49	0.90-2.48	1.02	0.78-1.35
Chinese	2.04[d]	1.66-2.50	1.16	0.86-1.56	2.52[d]	1.85-3.42
Other or unknown	1.11	0.91-1.35	1.09	0.81-1.47	0.99	0.76-1.30
nSES[h] quintiles (reference: 5)						
1	1.08	0.90-1.30	1.16	0.89-1.52	1.10	0.81-1.48
2	1.14	0.95-1.38	1.11	0.86-1.43	1.19	0.88-1.62
3	1.11	0.92-1.35	0.95	0.74-1.21	1.36	0.97-1.91
4	1.05	0.87-1.27	0.95	0.75-1.19	1.23	0.88-1.73
Baseline hemoglobin A_{1c}	1.03	0.99-1.07	1.11[d]	1.05-1.18	0.99	0.95-1.04
Baseline BP[i] (mm Hg; reference: ≤120/80)						
≤140/90	1.08	0.95-1.23	1.04	0.85-1.28	1.10	0.93-1.31
>140/90	1.04	0.89-1.23	0.99	0.78-1.26	1.09	0.88-1.36
Sex (reference: male)						
Female	1.36[d]	1.22-1.51	1.36[d]	1.16-1.59	1.34[d]	1.15-1.55
Insurance type (reference: Medicare)						
Private	0.85[g]	0.72-1.00	0.82[g]	0.68-0.99	2.02	0.80-5.11
Public	0.84[g]	0.72-0.98	1.03	0.78-1.37	0.83	0.69-1.00
Uninsured	0.95	0.73-1.23	N/A	N/A	1.03	0.77-1.37
Healthy Workers[j]	1.35[g]	1.04-1.76	N/A	N/A	1.18	0.89-1.57
Patient portal status (reference: inactivated)						
Activated	1.71[d]	1.47-2.00	1.86[d]	1.55-2.23	1.46[f]	1.10-1.94
Charlson Comorbidity Index (reference: 0-2)						
3+	1.33[d]	1.17-1.53	1.90[d]	1.57-2.30	0.93	0.77-1.12

[a]aOR: adjusted odds ratio.

XSL•FO
RenderX

[b]UCSF: University of California, San Francisco.

[c]SFHN: San Francisco Health Network.

[d]*P*<.001.

[e]N/A: not applicable.

[f]*P*<.01.

[g]*P*<.05.

[h]nSES: neighborhood socioeconomic status.

[i]BP: blood pressure.

[j]Healthy Workers is an insurance plan provided to temporary exempt employees of the City and County of San Francisco, including providers of in-home care.

Figure 1. Predicted probabilities of telemedicine visits with 95% CIs for interactions between health system and (A) age, (B) race and ethnicity, and (C) language preference.

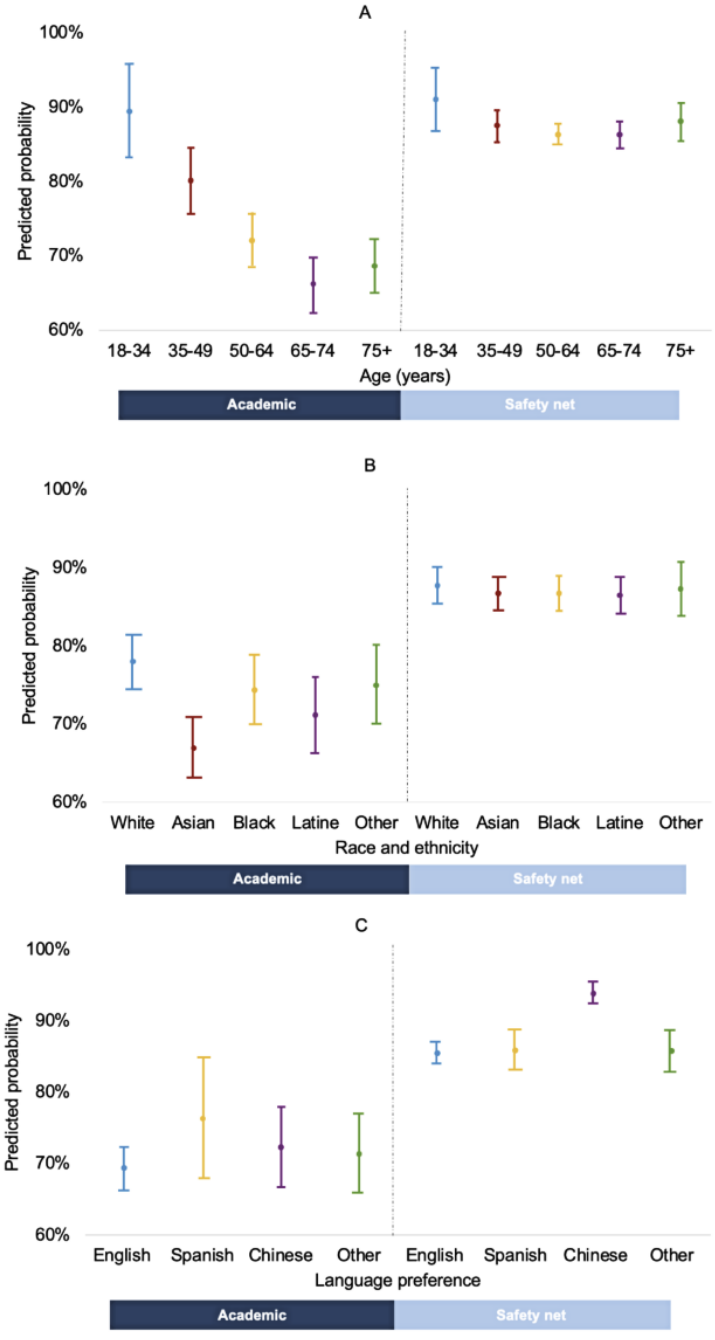

XSL•FO
RenderX

Stratified Data by Health System

Given the significant interactions by health system, we analyzed each system independently. At the academic health system, participants in the younger age groups (aged 18-34 years and 35-49 years) had higher odds of having a telemedicine visit than those aged 75 years and older (aOR 3.97, 95% CI 1.99-7.93; and aOR 1.86, 95% CI 1.36-2.56, respectively; Table 2). In the academic health system, non-Hispanic Asian (aOR 0.57, 95% CI 0.46-0.70) and Hispanic or Latine (aOR 0.67, 95% CI 0.50-0.91) participants were less likely to have a telemedicine visit than non-Hispanic White participants. There were no significant differences in telemedicine utilization based on language preference or nSES in the academic health system.

On the other hand, in the safety net system, patients with a Chinese language preference had higher odds of having a telemedicine visit (aOR 2.52, 95% CI 1.85-3.42; Table 2). In the safety net system, there were no significant differences in telemedicine utilization based on age, race and ethnicity, or nSES. These findings were consistent with the interaction analysis (Figure 1).

Discussion

Principal Results

Our study explored differences in telemedicine uptake among a cohort of primary care patients with diabetes across 2 large, unaffiliated health systems in San Francisco, California, during the COVID-19 PHE. Our final adjusted results highlight that health system is a significant predictor of telemedicine use and that disparities in telemedicine utilization may arise from differing implementation strategies by health systems, such as prioritization of audio-only versus video visits. Potential inequities could arise if audio-only visits are not widely implemented or perceived as inferior, despite payor policies that are meant to increase their implementation. Health systems may need to tailor and evaluate telemedicine implementation strategies to address disparities in utilization and ensure equitable access to services, especially for populations at greatest risk for digital exclusion.

Comparison With Prior Work

Our findings using the combined health system data are consistent with prior studies and indicated that groups traditionally vulnerable to digital exclusion (older and minoritized racial and ethnic groups) have lower odds of digital health uptake. These disparities in the context of telemedicine utilization have been well documented since the early stages of the PHE and have also been noted in the post-PHE era [6,11,18-22]. Novel to other studies, we found that these disparities differed across health systems with different telehealth implementation, even after adjustment for patient characteristics. When stratifying our models by health system, more disparities were found within the academic health system, where older, Asian, and Hispanic or Latine patients had lower odds of having a telemedicine visit than younger and non-Hispanic White patients, respectively; these disparities were not significant in the safety net system. In contrast, within the safety net system, Chinese-speaking patients had higher odds of having a telemedicine visit than other language preference groups.

Offering Audio-Only Visits May Increase Telemedicine Accessibility

The safety net system evaluated in our study serves a diverse group of patients who are at an increased risk for digital exclusion [14]. Despite this challenge, its adoption of a telemedicine infrastructure that primarily uses audio-only visits may have mitigated potential disparities in telemedicine access. While in-person visits remain the gold standard for primary care, robust audio-only options can augment the accessibility of care to patients reliant on telemedicine [23]. Studies have suggested that patients are more satisfied with telehealth modalities overall compared with in-person visits in terms of access and care provided, with no preference between video and audio-only visits [24,25]. Audio-only visits may be more accessible than video visits due to increased ease of use and lower technological requirements such as high-speed web access, smartphones, and laptops. Excluding or limiting these audio-only options may disproportionately impair access for minoritized racial and ethnicity groups, older adult patients, and non–English-speaking patients.

Our findings may have implications for patients in rural areas. Like urban safety net systems, patient served by rural health care systems are often at risk for digital exclusion, including having limited broadband access, lower rates of device ownership, and lower rates of digital literacy, suggesting that audio-only telemedicine may be necessary to prevent exacerbating health access disparities in these regions [26].

Considerations for Policy Makers and Health Care System Leaders to Ensure Telemedicine Equity

Our findings raise important considerations for future telemedicine policy, especially when considering the disparities observed within the academic health system. These disparities may partly be a consequence of the delayed and temporary approval for reimbursement of audio-only visits at a level similar to in-person visits by the Centers for Medicare & Medicaid Services [4]. The safety net system primarily served uninsured patients or those with public insurance (Medi-Cal) and Healthy Workers, where reimbursement rates were similar across audio-only and in-person visits. In contrast, the academic health system serves patients with private insurance or Medicare, both of which reimbursed video visits at lower rates than in-person or audio-only visits and thus had different incentives for telemedicine implementation. Inconsistent reimbursement across different telemedicine modalities can disproportionately impact populations at risk of digital exclusion and exacerbate existing health inequities. While video-enabled telemedicine care has been perceived in the United States to be more equivalent to in-person care and, therefore, reimbursed at a similar level for a longer period, audio-only telemedicine care has been perceived as providing lower quality and inadequate care, and therefore is reimbursed often at a much lower rate. This has resulted in limited motivation from many health care systems to provide audio-only encounters, which are more accessible to the populations that have experienced barriers to accessing

XSL•FO
RenderX

traditional face-to-face health care and have limited digital literacy or access to digital tools [27].

Policy makers should assess the impact of reimbursement policies across different telemedicine modalities and payors to ensure equitable adoption and appropriate utilization among all patient groups, especially patients who may forgo care otherwise without accessible telemedicine modalities. While most payor are concerned that increased accessibility may lead to higher utilization and costs, payors should also recognize the value of this opportunity to improve health care access for historically excluded populations.

Policies for encouraging delivery of more accessible telemedicine services should be implemented in conjunction with policies that address the barriers to accessing video-based telemedicine care. This may include policies that increase affordability of web access and devices as well as targeted investments in digital literacy programs [28]. Community-based initiatives, such as mobile technology training program and device distribution efforts, may reduce barriers to telehealth adoption [29]. These programs can be scaled to address the needs of populations with low digital literacy or limited access to digital tools [29].

Health systems leaders need to evaluate both the technologic needs and the health care requirements of their populations when developing telemedicine strategies, ensuring accessibility and equity while adapting to evolving policies and regulations. Incorporating multiple telemedicine options into care models can improve accessibility for older adults, patients with language barriers, and those in areas with limited broadband connectivity.

Both policy makers and health care leaders must evaluate the impact of policies on health equity over time. Consistent, systematic data collection and reporting on telemedicine utilization and its impact on clinical outcomes by sociodemographic characteristic such as race and ethnicity, language preference, and insurance coverage can help identify persistent gaps and guide future policy. When policies and implementation methods are aligned with real-world patient needs, particularly for populations at high risk of digital exclusion, telemedicine can serve as a tool to effectively reduce health disparities.

Limitations

Our sample included data from 2 urban health systems through March 2021, limiting generalizability to other systems and to later periods of the PHE. Future research evaluating the impact of telemedicine implementation in rural settings and examining trends in later post-PHE periods could further validate and broaden the applicability of these findings. However, more than 80% of the US population resides in urban areas, so our findings still have great relevance to many individuals [30]. In addition, differences in patient populations and institutional practices may be contributed to our findings. The safety net system may have had fewer disparities in telehealth use due to greater experience serving patients with limited digital access and stronger support for digital health literacy. However, our findings align with prior studies suggesting that telephone-based telehealth may improve accessibility and adoption among patient facing barriers to video-based services. We did not have access to the EHRs from outside systems, so we could not account for external encounters, although we limited our cohort to patients actively receiving primary care at UCSF or SFHN to mitigate this. Baseline measures of diabetes and BP control relied on single time points and may not perfectly reflect control through the study period. Incorporating longitudinal data in future analyses may help provide a more nuanced understanding of how disease control may be associated with telemedicine use.

Conclusions

This retrospective study examined a cohort of 10,201 primary care patients with diabetes during the PHE and found that inequities in telemedicine utilization varied across health systems with differing telemedicine implementation. Notably, disparities related to patient age and racial and ethnic identity underscore potential barriers to health care access. Compared with the safety net system with widespread adoption of audio-only visits, patients in the academic health system were less likely to have any telemedicine visit. Furthermore, lower telemedicine utilization was observed among older patients and minoritized racial and ethnic groups in the academic health system—disparities that were not present in the safety net system, which serves a more diverse population. Ultimately, our findings suggest that rapid, widespread telemedicine implementation prioritizing video telemedicine may exacerbate disparities in telemedicine utilization. Providing both audio-only and video appointments may enhance accessibility to care for populations historically at risk for digital exclusion.

Acknowledgments

This study is supported by the University of California, San Francisco (UCSF) Center for Healthcare Value; the UCSF Population Health Data initiative; and grants from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK P30DK092924); the National Heart, Lung, and Blood Institute (NHLBI K23HL157750); the National Institute on Aging (NIA K76AG088348); and the National Center for Advancing Translational Sciences (UCSF-CTSI grant UL1TR001872). MK is supported by UCSF School of Nursing Doctoral Student Funding. KEK and ADR are supported by the UCSF Population Health Data Initiative through residual class settlement funds in the matter of April Krueger v Wyeth, Inc, case no. 03-cv-2496 (US District Court, SD of California).

Funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Its contents are

XSL•FO
RenderX

solely the responsibility of the authors and do not necessarily represent the official views of UCSF or the National Institutes of Health.

Data Availability

The data collected during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SN, CRL, and ECK made substantial contributions to the conception or design of the work. All made substantial contribution to the analysis and interpretation of data for the work. JJS and MK drafted the manuscript. All critically revised the work, provided final approval of the submitted version, and agreed to be accountable for the accuracy and integrity of all parts of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient characteristics by health system.
[DOCX File , 19 KB - jmir_v27i1e64635_app1.docx]

Multimedia Appendix 2
Predictors of telemedicine use during the public health emergency by patient characteristics and health system interaction (N=10,201).
[DOCX File , 18 KB - jmir_v27i1e64635_app2.docx]

Multimedia Appendix 3
Estimated predicted probabilities for significant key predictors of telemedicine use after applying the health system interaction term.
[DOCX File , 16 KB - jmir_v27i1e64635_app3.docx]

References

1. Lin CC, Dievler A, Robbins C, Sripipatana A, Quinn M, Nair S. Telehealth in health centers: key adoption factors, barriers, and opportunities. Health Aff (Millwood) 2018;37(12):1967-1974. [doi: 10.1377/hlthaff.2018.05125] [Medline: 30633683]
2. Adler-Milstein J, Kvedar J, Bates DW. Telehealth among US hospitals: several factors, including state reimbursement and licensure policies, influence adoption. Health Aff (Millwood) 2014;33(2):207-215. [doi: 10.1377/hlthaff.2013.1054] [Medline: 24493762]
3. Uscher-Pines L, McCullough CM, Sousa JL, Lee SD, Ober AJ, Camacho D, et al. Changes in in-person, audio-only, and video visits in California's federally qualified health centers, 2019-2022. JAMA 2023;329(14):1219-1221 [FREE Full text] [doi: 10.1001/jama.2023.1307] [Medline: 37039799]
4. Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020. Wikipedia. URL: https://en.wikipedia.org/wiki/Coronavirus_Preparedness_and_Response_Supplemental_Appropriations_Act,_2020 [accessed 2025-02-14]
5. Medicare and Medicaid programs, basic health program, and exchanges; additional policy and regulatory revisions in response to the COVID-19 public health emergency and delay of certain reporting requirements for the skilled nursing facility quality reporting program. Federal Register. 2020. URL: https://tinyurl.com/3c8j4nre [accessed 2023-10-16]
6. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the COVID-19 pandemic. NEJM Catalyst 2020 May 4 [FREE Full text]
7. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
8. Eberly LA, Khatana SAM, Nathan AS, Snider C, Julien HM, Deleener ME, et al. Telemedicine outpatient cardiovascular care during the COVID-19 pandemic: bridging or opening the digital divide? Circulation 2020;142(5):510-512 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.048185] [Medline: 32510987]
9. Madjid K, Lee E, Couture S, Gonzales A, Grigorescu V, Smith S, et al. National Survey Trends in Telehealth Use in 2021: Disparities in Utilization and Audio vs. Video Services. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, Office of Health Policy; Feb 01, 2022:15.
10. Hsueh L, Huang J, Millman AK, Gopalan A, Parikh RK, Teran S, et al. Disparities in use of video telemedicine among patients with limited english proficiency during the COVID-19 pandemic. JAMA Netw Open 2021;4(11):e2133129 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.33129] [Medline: 34735015]

XSL•FO
RenderX

11. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]
12. Chang E, Penfold RB, Berkman ND. Patient characteristics and telemedicine use in the US, 2022. JAMA Netw Open 2024;7(3):e243354 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.3354] [Medline: 38517438]
13. Lyles C, Fields J, Sharma A, Aulakh V, Sarkar U. Launching a toolkit for safety-net clinics implementing telemedicine during the COVID-19 pandemic. Commonwealth Fund 2020.
14. Nouri S, Lyles CR, Sherwin EB, Kuznia M, Rubinsky AD, Kemper KE, et al. Visit and between-visit interaction frequency before and after COVID-19 telehealth implementation. JAMA Netw Open 2023;6(9):e2333944 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.33944] [Medline: 37713198]
15. Gomez SL, Glaser SL, McClure LA, Shema SJ, Kealey M, Keegan THM, et al. The California neighborhoods data system: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. Cancer Causes Control 2011;22(4):631-647 [FREE Full text] [doi: 10.1007/s10552-011-9736-5] [Medline: 21318584]
16. van Buuren S, Groothuis-Oudshoorn CG. MICE: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
17. StataCorp. Stata Statistical Software: Release 17. College Station, TX: StatCorp LLC; 2021.
18. Zeng B, Rivadeneira NA, Wen A, Sarkar U, Khoong EC. The impact of the COVID-19 pandemic on internet use and the use of digital health tools: secondary analysis of the 2020 health information national trends survey. J Med Internet Res 2022;24(9):e35828 [FREE Full text] [doi: 10.2196/35828] [Medline: 36041005]
19. Khoong EC, Butler BA, Mesina O, Su G, DeFries TB, Nijagal M, et al. Patient interest in and barriers to telemedicine video visits in a multilingual urban safety-net system. J Am Med Inform Assoc 2021;28(2):349-353 [FREE Full text] [doi: 10.1093/jamia/ocaa234] [Medline: 33164063]
20. Chunara R, Zhao Y, Chen J, Lawrence K, Testa P, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]
21. Ruberto RA, Schweppe EA, Ahmed R, Swindell HW, Cordero CA, Lanham NS, et al. Disparities in telemedicine utilization during covid-19 pandemic: analysis of demographic data from a large academic orthopaedic practice. JB JS Open Access 2022;7(2):e21.00116 [FREE Full text] [doi: 10.2106/JBJS.OA.21.00116] [Medline: 35425872]
22. Chang JE, Lai AY, Gupta A, Nguyen AM, Berry CA, Shelley DR. Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-COVID era. Milbank Q 2021;99(2):340-368 [FREE Full text] [doi: 10.1111/1468-0009.12509] [Medline: 34075622]
23. Baughman DJ, Jabbarpour Y, Westfall JM, Jetty A, Zain A, Baughman K, et al. Comparison of quality performance measures for patients receiving in-person vs telemedicine primary care in a large integrated health system. JAMA Netw Open 2022;5(9):e2233267 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.33267] [Medline: 36156147]
24. Allen AZ, Zhu D, Shin C, Glassman DT, Abraham N, Watts KL. Patient satisfaction with telephone versus video-televisits: a cross-sectional survey of an urban, multiethnic population. Urology 2021;156:110-116. [doi: 10.1016/j.urology.2021.05.096] [Medline: 34333039]
25. Chen K, Lodaria K, Jackson HB. Patient satisfaction with telehealth versus in-person visits during COVID-19 at a large, public healthcare system. J Eval Clin Pract 2022;28(6):986-990 [FREE Full text] [doi: 10.1111/jep.13770] [Medline: 36148479]
26. Cuadros DF, Moreno CM, Miller FD, Omori R, MacKinnon NJ. Assessing access to digital services in health care–underserved communities in the United States: a cross-sectional study. Mayo Clin Proc Digital Health 2023;1(3):217-225. [doi: 10.1016/j.mcpdig.2023.04.004]
27. Fisher R, Sarkar U, Adler-Milstein J. Audio-only telemedicine in primary care: embraced in the NHS, second rate in the US. Health Affairs Forefront 2022.
28. Sheon A, Khoong E. Pillars, policies, and plausible pathways linking digital inclusion and health equity. Health Affairs 2024.
29. Lyles CR, Nguyen OK, Khoong EC, Aguilera A, Sarkar U. Multilevel determinants of digital health equity: a literature synthesis to advance the field. Annu Rev Public Health 2023;44:383-405 [FREE Full text] [doi: 10.1146/annurev-publhealth-071521-023913] [Medline: 36525960]
30. United States Census Bureau. Urban areas facts. 2021. URL: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/ua-facts.html [accessed 2024-12-21]

Abbreviations

aOR: adjusted odds ratio
BP: blood pressure
CCI: Charlson Comorbidity Index
EHR: electronic health record

XSL•FO
RenderX

nSES: neighborhood socioeconomic status
PHE: public health emergency
SFHN: San Francisco Health Network
UCSF: University of California, San Francisco

Edited by A Coristine; submitted 23.07.24; peer-reviewed by D Valvi, Z Ehtesham, ND Berkman; comments to author 07.12.24; revised version received 31.01.25; accepted 06.02.25; published 24.03.25.

Please cite as:
Shih JJ, Kuznia M, Nouri S, Sherwin EB, Kemper KE, Rubinsky AD, Lyles CR, Khoong EC
Differences in Telemedicine Use for Patients With Diabetes in an Academic Versus Safety Net Health System: Retrospective Cohort Study
J Med Internet Res 2025;27:e64635
URL: https://www.jmir.org/2025/1/e64635
doi:10.2196/64635
PMID:40126552

©Jonathan J Shih, Magdalene Kuznia, Sarah Nouri, Elizabeth B Sherwin, Kathryn E Kemper, Anna D Rubinsky, Courtney R Lyles, Elaine C Khoong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

The Effect of Nurse Navigators in Digital Remote Monitoring in Cancer Care: Case Study Using Structural Equation Modeling

Etienne Minvielle[1,2], MD, PhD; Joel Perez-Torrents[2], PhD; Israa Salma[1], PhD; Philippe Aegerter[3], MD, PhD; Marie Ferrua[4], MSc; Charles Ferté[5], MD, PhD; Henri Leleu[6], MD, PhD; Delphine Mathivon[1], RN, MSc; Claude Sicotte[7], PhD; Mario Di Palma[1], MD; Florian Scotté[1], MD

[1]Interdisciplinary Department for the Organization of Patient Pathways, Gustave Roussy, Villejuif, France

[2]i3-CRG, École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

[3]Université de Versailles Saint-Quentin-en-Yvelines, Versailles, France

[4]Centre Hospitalier d'Argenteuil, Argenteuil, France

[5]Resilience Care, Paris, France

[6]Public Health Expertise, Paris, France

[7]Healthcare Management Department, Université de Montreal, Montreal, QC, Canada

Corresponding Author:
Etienne Minvielle, MD, PhD
Interdisciplinary Department for the Organization of Patient Pathways
Gustave Roussy
114 rue Edouard Vaillant
Villejuif, 94800
France
Phone: 33 0662102533
Email: Etienne.MINVIELLE@gustaveroussy.fr

Abstract

Background: The purpose of digital remote monitoring (DRM) is improving cancer care management. However, its effectiveness largely depends on the role of nurse navigators (NNs) within these systems to process data and lead action.

Objective: This study aims to fill gaps in our understanding of the role of NNs within a specific system, drawing on the Cancérologie parcours région Ile-de-France (CAPRI) DRM program applied to oncology patients.

Methods: The CAPRI DRM, targeting patients taking oral anticancer agents, combines digital interfaces with NN interventions. A phase 3 randomized controlled trial involving 559 patients assessed its safety and efficacy, with the primary end point being the relative dose intensity. This report focuses on patients in the CAPRI arm, evaluating the impact of NN interventions on outcomes such as toxicity, hospitalization, and emergency visits. Data on patient characteristics, NN interventions, and patient satisfaction surveys were analyzed using structural equation modeling.

Results: The study included 187 patients. Patient characteristics were significantly correlated with outcomes. Across all the models we used, the quality of NN interventions was consistently associated with higher patient satisfaction, with correlation coefficients ranging from 0.332 (95% CI 0.154-0.510; $P<.001$) to 0.366 (95% CI 0.182-0.550; $P<.001$). The number of grade ≥3 toxicity events correlated positively with NN referrals to oncologists. Hospitalization length was positively related to NN referral (coefficient 0.102, 95% CI 0.051-0.153; $P<.001$) and inversely to NN advice (coefficient –0.045, 95% CI –0.096 to 0.006; $P=.08$). Emergency visits showed a negative correlation with NN actions (coefficient –0.478, 95% CI –0.923 to 0.033; $P=.04$) and a positive correlation with NN calls and referrals (coefficient 0.516, 95% CI 0.069-0.963; $P=.02$).

Conclusions: This study shows the central role of NNs in making DRM effective. Despite the study's limitations, these results support the design of DRM as a hybrid model of automated digital tools and human support. Future research should explore the applicability of such a DRM model in various clinical settings to clarify the optimal association between automated systems and NN expertise.

Trial Registration: ClinicalTrials.gov NCT02828462; https://www.clinicaltrials.gov/study/NCT02828462

(J Med Internet Res 2025;27:e66275) doi:10.2196/66275

XSL•FO
RenderX

KEYWORDS

digital remote monitoring; nurse navigators; patient care; oncology; toxicity; patient satisfaction; hospitalization

Introduction

Background

The tremendous uptake of digital care (mobile apps and web portals) has paved the way for improving the use of remote patient monitoring (ie, the digitally remote collection of patient-generated data that is automatically sent to health care providers), and it has created opportunities to further integrate digital remote monitoring (DRM) into the delivery of various health services. Several clinical trials implementing DRM tools showed that their use may improve the quality of care compared to standard care, reduce costs, supplement (or replace) in-hospital care, and improve management of clinical events in different chronic conditions [1,2], including cancer [3-7].

However, implementing supportive digital remote monitoring interventions remains challenging, with concerns about potential overuse and exacerbation of health inequalities due to the digital divide [8-11]. Successful DRM implementation relies on several factors, combining patient-reported outcomes, carefully designed technology, and specialist health care professionals' expertise [7-12].

What Role Is Played by Nurse Navigators?

Considering the importance of professional expertise, the actions of nurse navigators (NNs) may become crucial. Their expertise lies in clinical care and patient navigation, meaning that NNs can enhance the usability and safety of DRM technologies [12]. They can be a critical link between patients and the health care system, ensuring digital tools are effectively integrated into patient care. NNs can interpret data from digital tools, provide timely interventions, and guide patients through their treatment journey [12-14].

Despite the postulated importance of NNs in DRM, more research is needed to explore the extent to which they contribute to the effectiveness of these systems, especially in oncology. This study aims to fill this gap by examining the impact of NNs in the Cancérologie parcours région Ile-de-France (CAPRI) DRM program, considering factors such as toxicity management and care use.

Methods

An Example of DRM: CAPRI

CAPRI is an application of DRM in oncology. It specifically targets patients receiving oral anticancer agents, focusing on improving adherence, managing toxicities, and enhancing overall patient care. The core of CAPRI lies in its dual components: a digital interface (including a smartphone app and a web portal) and the active involvement of NNs in patient monitoring and intervention [14]. NNs involved in the CAPRI study received in-house training during two 2-hour sessions on how to use the application.

The safety and efficacy of CAPRI were assessed in a phase 3 randomized controlled trial (RCT) comparing the CAPRI DRM intervention against the standard of care. The trial included 559 randomized patients followed for 6 months. The study's primary end point was the relative dose intensity (RDI), defined as the ratio of the dose delivered over time to the prescribed dose intensity. Secondary end points included treatment adherence, toxicity management, patient-reported outcomes, and overall quality of life. The trial's design and results have been published elsewhere [15]. The main results showed that the CAPRI group (n=272) showed a significant improvement in RDI (93.4% vs 89.4%; *P*=.04), an enhanced patient experience (Patient Assessment of Chronic Illness Care score 2.94 vs 2.67; *P*=.01), reduced hospitalization length (2.82 vs 4.44 days; *P*=.02), and decreased grade ≥3 toxicities (27.6% vs 36.9%; *P*=.02).

This study is an ancillary analysis of the 272 patients included in the CAPRI intervention arm. Its objective is to detail the mechanisms of the NN interventions' effects on patient outcomes in the intervention arm.

Data Collection

Three distinct types of data collected during the RCT were used: patient characteristics and outcomes, including demographics and outcomes that have been previously published [15]; NN interventions, extracted from the NN dashboard used during the trial (Figure 1); and patient perceptions and satisfaction, based on a dedicated patient questionnaire administered at the end of the RCT to assess patients' use, perceived utility, and satisfaction with the program. Data were available for 69% of patients (n=187).

XSL•FO
RenderX

Figure 1. Screenshots of the patient smartphone app (A) and nurse navigators' dashboard (B).

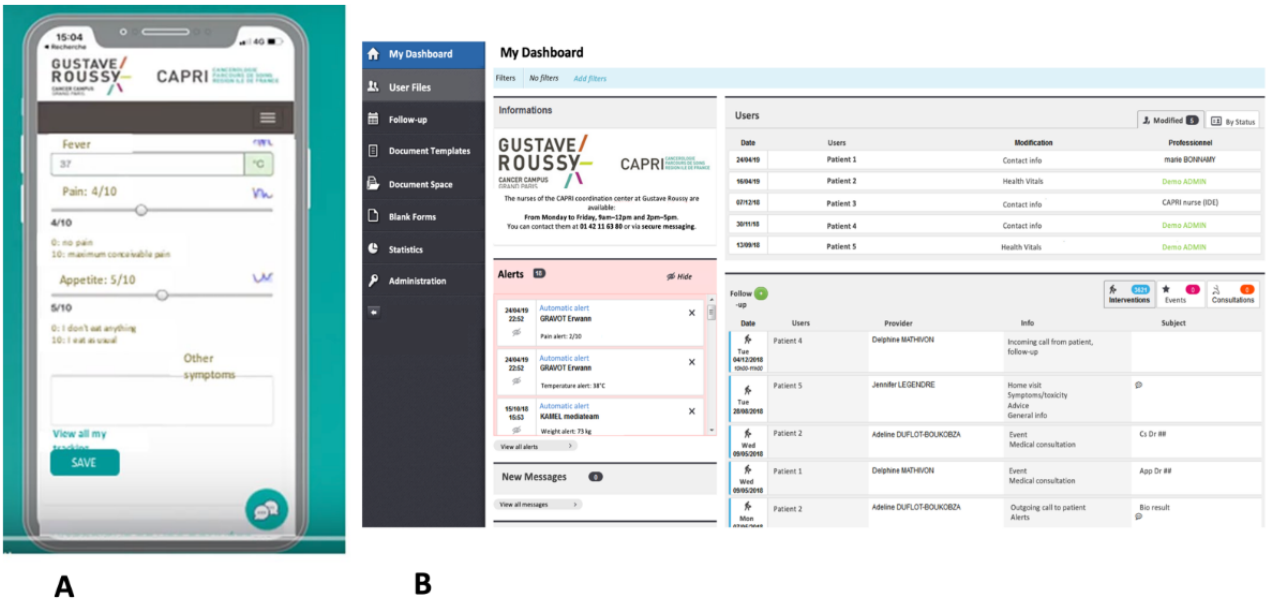

Data Analysis

The role of NN interventions in outcomes, including toxicity, hospitalization length, and emergency service visits, was explored using a structural equation modeling approach. As shown in Figure 2, the framework used draws inspiration from the DeLone and McLean [16] model, which explains the success of information systems like DRM based on the roles of system quality, user satisfaction, and use.

Figure 2. Structural modeling framework used to explore the effects of nurse navigator (NN) interventions on outcomes.

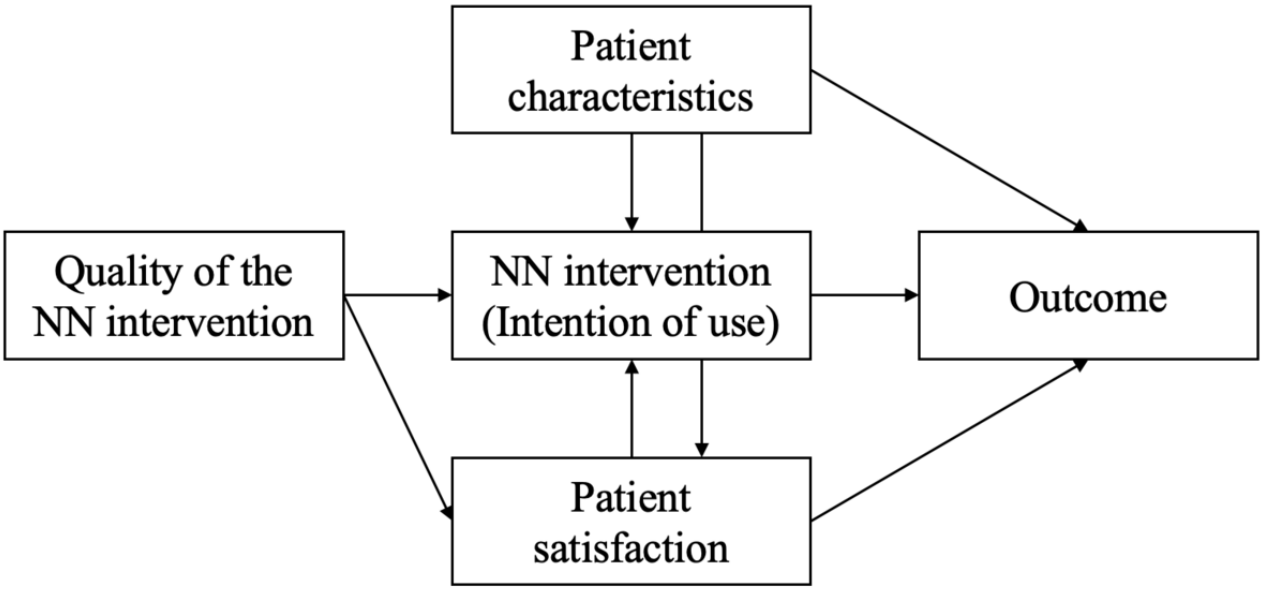

The structure used considers that differences in outcomes are mediated through the NN interventions after adjusting to individual patient's characteristics. The quality of the NN interventions and patients' satisfaction were assessed using the patient questionnaire. Through factor analysis, it was verified that the associated questionnaire items corresponded to shared underlying dimensions, thereby allowing the reduction of these items to the unified latent variables used in the structural model.

CAPRI use was based on the number and frequency of NN interventions throughout the RCT, separated into (1) the number of inbound or outbound calls, (2) actions taken (appointment management, electronic health record management, information about administrative issues), (3) referrals (external health care providers, internal oncologists), (4) advice to patients, and (5) coordination actions. Patient characteristics included sociodemographic data, fitness (World Health Organization score) and metastatic cancer status. Before introduction into the model, variables were standardized to account for different measurement scales, while missing data were imputed using a complete information maximum likelihood approach.

Final models were selected using an iterative process based on the model's fit, which was evaluated using several global

XSL•FO
RenderX

indicators, including the standardized root mean square residual, adjusted goodness of fit, Tucker-Lewis index, comparative fit index, and root mean squared error of approximation. Our sample size (n=187) refers to general rules adopted for structural equation models with a small numbers of variables (4 to 7, depending on the model) [17]. A complete model and methodological description are available in Multimedia Appendix 1.

Ethical Considerations

This study was a secondary analysis based on the results of the CAPRI study, which was conducted according to applicable laws and regulations and the Declaration of Helsinki and was approved by an ethics committee (CPP Paris-Ile-de-France IV [2016/20SC] and US Department of Health and Human Services (00003835). The study was registered on ClinicalTrials.gov (NCT02828462). All patients included in the study provided informed consent to participate in the study. The informed consent form clarified that the collected data were part of the CAPRI study and would be used for primary and secondary analyses. To protect participants' privacy, all data were anonymized. No compensation was offered to participants.

Results

Patient Characteristics

Overall, 187 patients from the intervention arm of the CAPRI trial were included. The characteristics of the respondents were similar to those of the complete set of patients, except for a longer follow-up period, as shown in Table 1.

Table 1. Patient characteristics.

Characteristics	Participants in this study (n=187)	Participants in the CAPRI[a] study intervention arm (n=272)	*P* value
Age (years), mean (SD)	59.8 (14.2)	59.1 (14.0)	.59
Female participants, n (%)	103 (55.8)	156 (57.3)	.63
Performance status[b], n (%)			.24
WHO[c] 0	92 (49.2)	121 (44.5)	
WHO 1	82 (43.9)	120 (44.1)	
Metastatic cancer, n (%)	147 (78.6)	222 (81.6)	.42
Follow-up duration (weeks), mean (SD)	148.7 (48.3)	133.0 (54.2)	.02

[a]CAPRI: Cancérologie parcours région Ile-de-France.

[b]Performance status reflects cancer patients' functional ability, ranging from normal activity (0) to completely bedridden (4).

[c]World Health Organization.

As expected, patients' characteristics, including age, performance status, and metastatic stage, were significantly associated with the selected outcomes (Multimedia Appendix 1, Figures S3 to S5).

The Role of NN Interventions

Structural model results, which do not show patients' characteristics, are presented in Figure 3.

Across all the models we used, the quality of NN interventions was consistently associated with higher patient satisfaction, with correlation coefficients ranging from 0.332 (95% CI 0.154-0.510; SE 0.091; z=3.64; P<.001) to 0.366 (95% CI 0.182-0.550; SE 0.094; z=3.90; P<.001). Likewise, patient satisfaction was consistently and positively associated with NN interventions, with similarly strong correlation coefficients ranging from 0.207 (SE 0.088, 95% CI 0.035-0.380; z=2.341; P=.02) to 0.283 (SE 0.079, 95% CI 0.128-0.438; z=3.598; P<.001).

The number of grade ≥ 3 toxicity events was positively associated with the number of referrals to oncologists and coordination actions. Hospitalization length was positively correlated with the number of referrals to oncologists (coefficient 0.102, 95% CI 0.051-0.153; SE 0.026; z=3.90; P<.001) and negatively associated with the number of times advice was given (coefficient –0.045, 95% CI –0.096 to 0.006; SE 0.026; z=–1.73; P=.08). Emergency visits were negatively related to the number of actions taken (coefficient –0.478, 95% CI –0.923 to –0.033; SE 0.227; z=–2.10; P=.04) and positively associated with the number of calls and the number of referrals (coefficient 0.516, 95% CI 0.069-0.963; SE 0.228; z=2.26; P=.02).

XSL•FO
RenderX

Figure 3. Simplified results (correlation coefficients) of the final structural models (excluding patient characteristics) for each outcome. NN: nurse navigator.

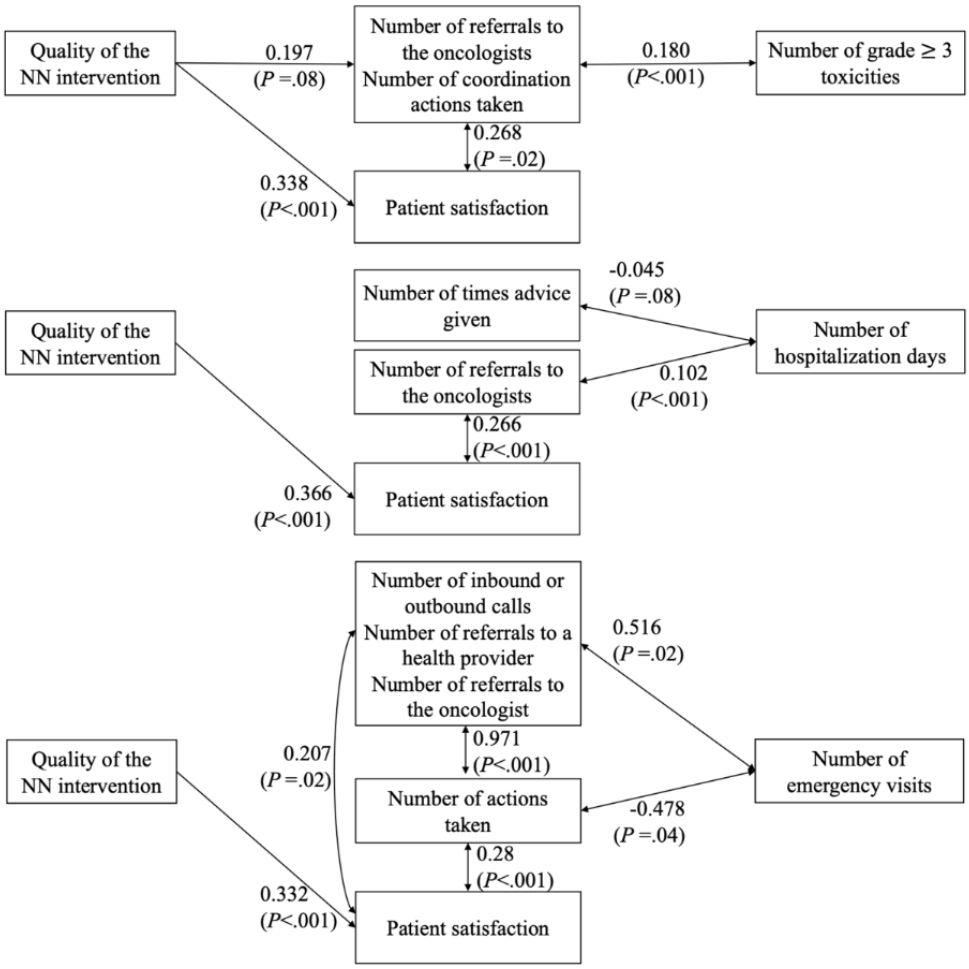

Discussion

This study corroborates past findings that nurse-led interventions in a DRM program for cancer patients can impact health outcomes [9-11]. The actions taken by NNs within CAPRI were associated with reduced emergency service use and hospitalization length. This suggests that NNs are critical in providing appropriate advice and avoiding unnecessary emergency service use and hospital interventions. The positive associations between emergency service use, hospitalization length, and grade ≥ 3 toxicity events and NN interventions, such as coordination or referrals, likely reflect a prioritization in monitoring and treating severe cases. This also highlights the versatility of NNs in managing most patient interaction and orientation without direct oncologist referrals, confirming their potential in delegating certain oncologist tasks [18].

These findings emphasize the importance of considering DRM as a hybrid model that combines digital tools with human support, especially in complex cases. Additionally, patients' preferences for different platforms (phone or web based) or reluctance to use digital platforms highlights the need for flexibility and human availability in DRM implementation [19].

Despite its contributions, this study had limitations, including its single-center nature, which might affect the external validity of the findings. This hybrid model's application to other cancer conditions and its generalizability requires further investigation.

In conclusion, the study illuminates the role of skilled nurses in DRM implementation, suggesting the need for future research to explore the division of tasks between automated alert systems and human expertise, particularly in managing complex cases and cancer pathways. The potential managerial implications underline the importance of clearly defining the role of NNs and considering their activity in payment models for digital remote monitoring systems [20].

Acknowledgments

The authors thank the Gustave Roussy clinical and research teams, as well as the patients and their families. The study was funded by the French National Research Agency, the Agence Régionale de Santé Ile-de-France, Philanthropia Lombard Odier Foundation, Novartis, and AstraZeneca. The funding bodies were not involved in the study design.

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: EM, JP-T, PA, MF, CF, HL, DM, CS, MDP, FS
Formal analysis: EM, JP-T, HL
Investigation: EM, JP-T, HL
Methodology: EM, JP-T, HL
Validation: EM, JP-T, HL
Writing—original draft: EM, JP-T, HL, IS
Writing—review and editing: EM, JP-T, IS, PA, MF, CF, HL, DM, CS, MDP, FS

Conflicts of Interest

CF is a full-time employee of Resilience Care. FS declares conflicts related to BMS, Pfizer, Sanofi, Roche, MSD, Leo Pharma, AMGEN, Pierre Fabre Oncologie, Gilead, Pharmanovia, Daiichi Sankyo, Thermo Fisher, Viatris, and GSK. After this study, Resilience Care derived a commercial solution from Cancérologie parcours région Ile-de-France (CAPRI). However, Resilience Care was not involved in the study design or analysis.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 418 KB - jmir_v27i1e66275_app1.docx]

Multimedia Appendix 2
Checklist of iCHECK-DH guidelines.
[PDF File (Adobe PDF File), 178 KB - jmir_v27i1e66275_app2.pdf]

References

1. Using remote patient monitoring technologies for better cardiovascular disease outcomes. American Heart Association. 2019. URL: https://www.heart.org/-/media/files/about-us/policy-research/policy-positions/clinical-care/remote-patient-monitoring-guidance-2019.pdf?la=en&hash=A98793D5A043AB9940424B8FB91D2E8D5A5B6BEB [accessed 2024-10-24]
2. McLean S, Protti D, Sheikh A. Telehealthcare for long term conditions. BMJ 2011 Feb 03;342:d120. [doi: 10.1136/bmj.d120] [Medline: 21292710]
3. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
4. Garg S, Williams NL, Ip A, Dicker AP. Clinical integration of digital solutions in health care: an overview of the current landscape of digital technologies in cancer care. JCO Clin Cancer Inform 2018 Dec;2:1-9. [doi: 10.1200/CCI.17.00159] [Medline: 30652580]
5. Maguire R, McCann L, Kotronoulas G, Kearney N, Ream E, Armes J, et al. Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART). BMJ 2021 Jul 21;374:n1647 [FREE Full text] [doi: 10.1136/bmj.n1647] [Medline: 34289996]
6. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 01;109(9):djx029. [doi: 10.1093/jnci/djx029] [Medline: 28423407]
7. McGregor BA, Vidal GA, Shah SA, Mitchell JD, Hendifar AE. Remote oncology care: review of current technology and future directions. Cureus 2020 Aug 31;12(8):e10156 [FREE Full text] [doi: 10.7759/cureus.10156] [Medline: 33014652]
8. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2025-03-10]
9. Mecklai K, Smith N, Stern AD, Kramer DB. Remote patient monitoring - overdue or overused? N Engl J Med 2021 Apr 15;384(15):1384-1386. [doi: 10.1056/NEJMp2033275] [Medline: 33853209]
10. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]
11. Franzoi MA, Ferreira AR, Lemaire A, Rodriguez J, Grosjean J, Ribeiro JM, et al. Implementation of a remote symptom monitoring pathway in oncology care: analysis of real-world experience across 33 cancer centres in France and Belgium. Lancet Reg Health Eur 2024 Sep;44:101005 [FREE Full text] [doi: 10.1016/j.lanepe.2024.101005] [Medline: 39444707]

XSL•FO
RenderX

12. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018 May 23;38:122-134 [FREE Full text] [doi: 10.1200/EDBK_200383] [Medline: 30231381]
13. Manderson B, McMurray J, Piraino E, Stolee P. Navigation roles support chronically ill older adults through healthcare transitions: a systematic review of the literature. Health Soc Care Community 2012 Mar;20(2):113-127. [doi: 10.1111/j.1365-2524.2011.01032.x] [Medline: 21995806]
14. Charalambous A, Wells M, Campbell P, Torrens C, Östlund U, Oldenmenger W, et al. A scoping review of trials of interventions led or delivered by cancer nurses. Int J Nurs Stud 2018 Oct;86:36-43. [doi: 10.1016/j.ijnurstu.2018.05.014] [Medline: 29960894]
15. Mir O, Ferrua M, Fourcade A, Mathivon D, Duflot-Boukobza A, Dumont S, et al. Digital remote monitoring plus usual care versus usual care in patients treated with oral anticancer agents: the randomized phase 3 CAPRI trial. Nat Med 2022 Jun;28(6):1224-1231. [doi: 10.1038/s41591-022-01788-1] [Medline: 35469070]
16. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
17. Kaplan D. Structural Equation Modeling: Foundations and Extensions. New York, NY: SAGE Publications; 2008.
18. Ferrua M, Mathivon D, Duflot-Boukobza A, Abbas M, Charles C, Barrais A, et al. Nurse navigators' telemonitoring for cancer patients with COVID-19: a French case study. Support Care Cancer 2021 Aug;29(8):4485-4492 [FREE Full text] [doi: 10.1007/s00520-020-05968-y] [Medline: 33462724]
19. Mooney KH, Beck SL, Wong B, Dunson W, Wujcik D, Whisenant M, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. Cancer Med 2017 Mar;6(3):537-546 [FREE Full text] [doi: 10.1002/cam4.1002] [Medline: 28135050]
20. Agarwal SD, Barnett ML, Souza J, Landon BE. Adoption of Medicare's transitional care management and chronic care management codes in primary care. JAMA 2018 Dec 25;320(24):2596-2597. [doi: 10.1001/jama.2018.16116] [Medline: 30575868]

Abbreviations

CAPRI: Cancérologie parcours région Ile-de-France
DRM: digital remote monitoring
iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations
NN: nurse navigator
RCT: randomized controlled trial
RDI: relative dose intensity

Edited by N Cahill; submitted 11.09.24; peer-reviewed by J Ordieres-Meré, K Le Du; comments to author 30.10.24; revised version received 14.12.24; accepted 31.12.24; published 28.03.25.

Please cite as:
Minvielle E, Perez-Torrents J, Salma I, Aegerter P, Ferrua M, Ferté C, Leleu H, Mathivon D, Sicotte C, Di Palma M, Scotté F
The Effect of Nurse Navigators in Digital Remote Monitoring in Cancer Care: Case Study Using Structural Equation Modeling
J Med Internet Res 2025;27:e66275
URL: https://www.jmir.org/2025/1/e66275
doi:10.2196/66275
PMID:

©Etienne Minvielle, Joel Perez-Torrents, Israa Salma, Philippe Aegerter, Marie Ferrua, Charles Ferté, Henri Leleu, Delphine Mathivon, Claude Sicotte, Mario Di Palma, Florian Scotté. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Topic Modeling of Nursing Issues in the Media During 4 Emerging Infectious Disease Epidemics in South Korea: Descriptive Analysis

Jungok Kim[1*], PhD; Eun Kyoung Yun[1*], PhD

College of Nursing Science, Kyung Hee University, Seoul, Republic of Korea

[*]all authors contributed equally

Corresponding Author:
Eun Kyoung Yun, PhD
College of Nursing Science
Kyung Hee University
26, Kyungheedae-ro
Dongdaemun-gu
Seoul, 02447
Republic of Korea
Phone: 82 2 961 2348
Fax: 82 2 961 9398
Email: ekyun@khu.ac.kr

Abstract

Background: Emerging infectious disease disasters receive extensive media coverage and public attention. Nurse burnout and attrition peak during health crises such as pandemics. However, there is limited research on nursing issues related to repeated emerging infectious disease crises over time.

Objective: The purpose of this study was to analyze and draw implications from changes in key nursing issues reported by the news media during the outbreaks of severe acute respiratory syndrome (SARS; 2003), influenza A (2009), Middle East respiratory syndrome (MERS; 2015), and COVID-19 (2020) in Korea using topic modeling.

Methods: A total of 51,489 news articles were extracted by searching for the keywords “nursing” or “nurse” in the title or body of articles published from April 2003 to May 2021 (during new infectious disease outbreaks) in the open integrated database. The selected news articles were preprocessed then analyzed for text and structure using a 3-step keyword analysis method, latent Dirichlet allocation topic modeling, and keyword network analysis.

Results: Among the 51,489 news articles collected with the search terms “nursing” and “nurse,” 17,285 (33.6%) were selected based on the eligibility criteria and used in the final analysis. Using topic modeling, we derived 5 topics each for SARS, influenza A, and MERS and 6 topics for COVID-19. The themes commonly identified through topic modeling and keyword network analysis across the 4 epidemics were “response to emerging infectious diseases in Korea,” “demand for nurses,” “vulnerability in the work environment,” and “roles and responsibilities of nurses.” Although the topic names were the same, the meanings implied by the comprehensive keywords for each epidemic varied depending on the epidemic and the times.

Conclusions: Analysis of the identified themes and associated keyword network revealed that issues related to nurse shortages, working conditions, and poor treatment were not unique to the COVID-19 pandemic but rather recurring themes from previous epidemics. Our findings can be used to inform strategies to improve the professional roles, work environment, and treatment of nurses during health crises. Suggestions for future nursing-related policy impact and change research are also provided.

(J Med Internet Res 2025;27:e60446) doi:10.2196/60446

KEYWORDS

topic modeling; news articles; nursing issues; text analysis; emerging infectious disease

Introduction

An emerging infectious disease (EID) is a newly recognized or previously known disease characterized by new virulence or spread in previously unaffected areas [1]. Since the 2000s, South Korea has had 4 EIDs: severe acute respiratory syndrome (SARS) in 2003, novel swine-origin influenza A (H1N1) in 2009, Middle East respiratory syndrome (MERS) in 2015, and COVID-19 in 2020. Among them, influenza A and COVID-19

XSL•FO
RenderX

were declared as pandemics by the World Health Organization (WHO). Antiviral treatments were provided free of charge in Korea starting in August 2009 during the SARS outbreak, after which the number of infections and deaths were no longer regularly recorded. Before then, 270 deaths and 2417 infections were recorded, and at the time of manuscript submission, SARS had resulted in an additional 0 deaths and 4 infections according to official reports. Although the exact number of infections has not been counted since August 2009, more than 700,000 cases had been prescribed antiviral drugs from August 2009 to December 2009. During the MERS period, there were 39 deaths and 186 infections. Regarding COVID-19, 34,572,552 confirmed cases and 35,605 deaths were reported from January 20, 2020, to August 30, 2023, the period involving full monitoring [2].

EID catastrophes receive extensive media coverage and public attention. Indeed, public interest in nurses peaks during health crises, such as infectious disease epidemics [3]. EID outbreaks increase public recognition of and attention to the dedication of nurses working on the front lines of the disaster response [4]. Meanwhile, nurses experience physical and psychological burnout as well as professional ethical conflicts owing to the risk of infection, heavy workloads, patient violence, and the stigma of being a carrier of infection [5-8]. Rising rates of nurse burnout and resignation during inadequate infectious disease crisis responses have become global concerns [9,10]. The loss of skilled nurses may lead to a decline in the quality of nursing care and organizational nursing capacity, which may pose a risk to national health crisis responses and public health.

Although infectious disease outbreaks can be traumatic for individuals and communities, they can also lay the groundwork for health system transitions after the outbreak ends, through institutional and policy improvements that identify and address health system weaknesses revealed during the response. With respect to issues in the nursing workforce, scholars have suggested that societal consensus for improvements in working conditions and institutional arrangements for health resources and protection are needed to minimize nurse burnout and attrition in special situations, such as infectious disease outbreaks [6,8,11,12]. At the national level, discussions about appropriate investments and compensation systems for the training and deployment of infectious disease professionals and collaboration with the medical community are needed, whereas at the societal level, increasing public awareness of health workers and building public consensus through media coverage have been emphasized as obligations [7,13,14]. Media's reports of issues in the field of infectious diseases play an important role in forming society's perceptions of an infectious disease response, nursing, and nurses. Problems that are not resolved over a long period of time, occur repeatedly, and increase public interest become social issues, which become part of a public agenda that the general public recognizes as appropriate for the government to solve. This public agenda is then developed and set as a specific policy agenda.

Previous studies analyzing media coverage of infectious disease outbreaks focused on public perceptions of nursing portrayals and issues [3,4,15,16] and were mostly limited to specific epidemic periods. However, these studies did not track how nursing issues were portrayed during recurrent outbreaks since the 2000s, the unique characteristics of each period, and how characteristics have changed. In the context of COVID-19, which was declared as an endemic by the WHO, policies that address the underlying issues and promote a better future urgently need to be developed. Prior studies using content analysis techniques to analyze data collected through standardized questionnaires or media inevitably had limited data to reflect public perceptions and trends. To compensate for this, it is necessary to grasp the meaning inherent in big data such as online news and social media, and topic modeling and keyword network analysis methods, which have been widely used recently to analyze unstructured text in big data, overcome the limitations of manual analysis by humans [17-19].

Therefore, this study aimed to analyze the changes in major nursing issues reported by the news media when 4 new infectious diseases occurred in the 2000s using topic modeling and keyword network analysis methods and derive the implications. Specifically, we aimed to identify the issues that should be considered to effectively respond to new infectious disease crises that may occur in the future and ensure public safety. Our findings could provide the basis for the development of desirable infectious disease response strategies.

Methods

Data Extraction

We collected data from the Korea Press Foundation's article information website [20], an open-access database containing news articles from 54 media organizations. We collected data published between April 2003 and May 2021 corresponding to the periods of EID epidemics. Specifically, we analyzed news articles published during the SARS, influenza A, MERS, and COVID-19 outbreaks.

A total of 51,489 news articles were extracted by searching for "nursing" or "nurse" as the keyword in the title or text of the article during the outbreak of each new infectious disease. Although the terms "nursing" and "nurse" appeared in the text of these articles, irrelevant content, such as promotional material, was excluded using the selection and removal criteria detailed in Figure 1. These criteria were informed by advice from experts in the field of nursing research and clinical experience. A total of 17,285 news articles were selected using these criteria and used for the final analysis. Figure 1 details the data extraction process used in this study.

XSL•FO
RenderX

Figure 1. Study process. MERS: Middle Eastern respiratory syndrome; SARS: severe acute respiratory syndrome.

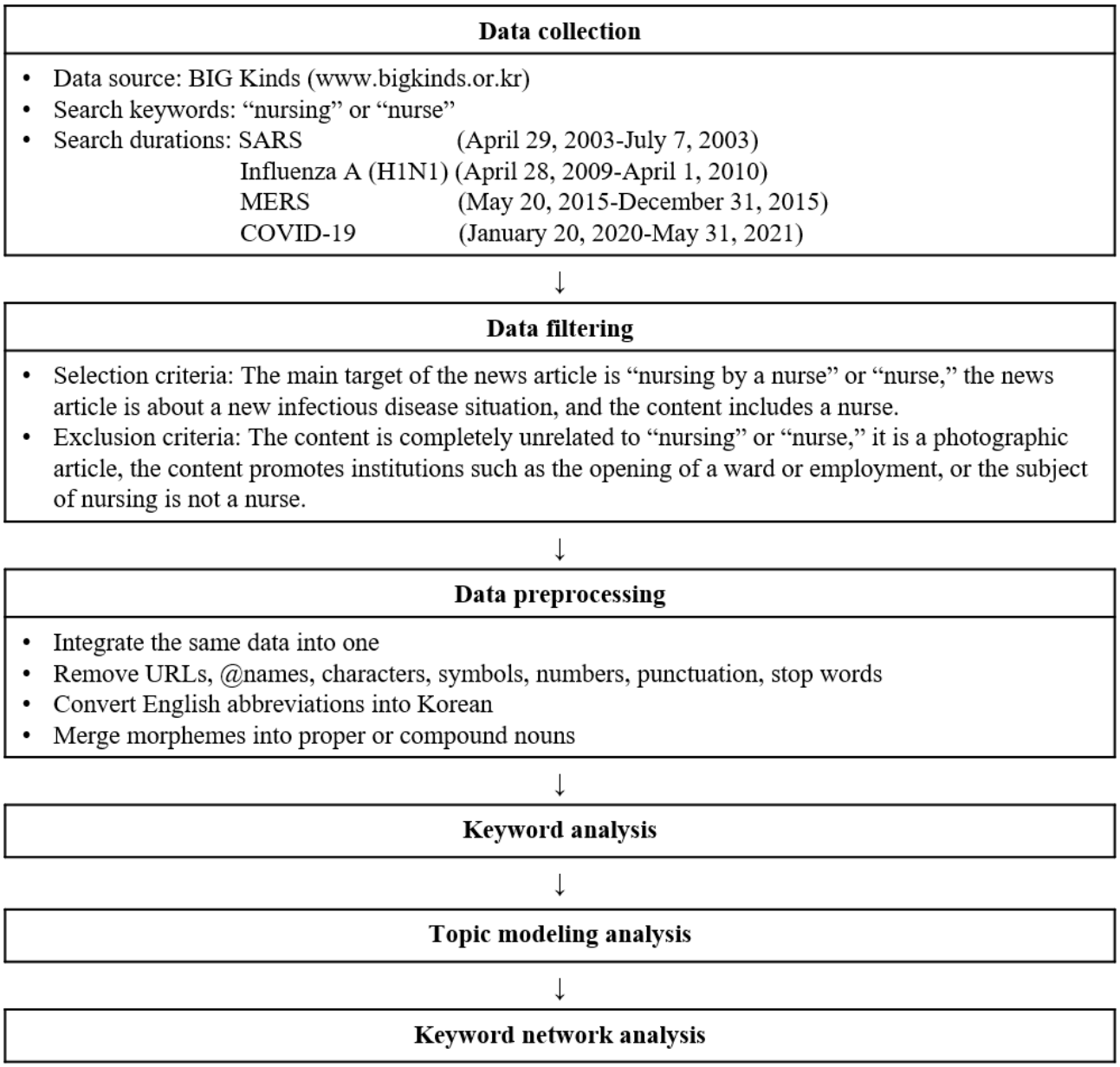

Data Preprocessing

We preprocessed the metadata of the selected news articles using the *KoNLP* and *tm* packages for natural language processing in R version 4.1.0. Specifically, we removed terms, including symbols, special characters, names of private individuals, and media company names. English terms, such as "SARS" and "OECD," were occasionally removed during terminology preprocessing; these terms were converted into their Korean equivalents to maintain consistency. During morpheme analysis, complex nouns like "living treatment center" were segmented into "living," "treatment," and "center," causing loss of meaning. To address this, these terms were reconstituted as proper or compound nouns based on N-gram analysis and the context of the original article. Abbreviated Korean terms were replaced with their official names to ensure terminological consistency.

Keyword Analysis

We used the *tm* package in R version 4.1.0 to calculate word frequencies and determine the importance of specific terms using the term frequency-inverse document frequency metric value. Statistical weights were calculated to assess the significance of specific terms within the corpus for keyword analysis of the news articles.

Topic Modeling

Topic modeling classifies and integrates topics in extensive document sets using probabilistic statistical methods to identify latent high-level concepts, extract the topics implicit in the documents, and derive common high-level concepts latent in the topics [21]. We used the latent Dirichlet allocation (LDA) algorithm of the *topicmodels* package of R version 4.1.0 to analyze the topics embedded in the documents. The coding of the analysis was based on the work of Grün and Hornik [16] and open-source code on the GitHub site.

XSL•FO
RenderX

We configured the LDA algorithm by setting the document extraction parameter (α) to 50/k, and the previous parameter delta (or eta) of the word distribution constituting the topic was set to 0.1. Gibbs sampling was used to estimate the topic distribution for the document, with 1000 repetitions used for Gibbs sampling [16,22].

We also conducted a quantitative analysis to select the appropriate number of topics. We calculated perplexity values ranging from 2 to 20 using the *topic models* and *LDAvis* packages of the R program. The topic modeling analysis was repeated by applying different numbers of topics based on the perplexity values. For the analysis, we repeated the adjustment of the λ value of the LDA visualization and reviewed the original text of the news articles to extract a list of the 30 most representative keywords for each topic.

Keyword Network Analysis

Keyword network analysis visualizes keyword relationships by calculating co-occurrence frequencies and metrics such as connection centrality, proximity centrality, and mediation centrality [23]. We performed keyword network analysis using the *Network*, *sna*, and *ggplot2* packages in the R program to improve the specific content and structure of the topics and the words constituting the topics derived from thematic modeling. Considering the visibility of the network around thematic keywords, the correlation coefficient was established, and the relationship between the simultaneous appearance of words was analyzed.

The top 15 keywords were selected as the main keywords for each topic through quantitative analyses including topic modeling analysis, keyword network analysis, and verification by nursing experts, and the topic names reflected the same higher-level concept to examine the trends of the topics and keywords that make up the topic by period of the new infectious disease based on the main keywords. In this study, 15 keywords for the new infectious disease topic and topic are presented.

Ethical Considerations

All news article data analyzed in this study were publicly available and not subject to copyright restrictions. The Institutional Review Board of Kyung Hee University approved this study (IRB number KHSIRB-21-334(EA)).

Results

News and Keyword Counts

After preprocessing the data from a total of 17,285 news articles, we extracted 1974 words for the SARS epidemic, 9520 words for influenza A, 15,639 words for MERS, and 24,888 words for COVID-19, resulting in a total of 52,021 words. The numbers of news articles and keyword types analyzed increased progressively across successive infectious disease outbreaks, and the average daily number of news articles was the highest during the MERS period.

Top Topics and Keywords During Epidemics

Table 1 presents the keywords and topic names identified for each epidemic period. We selected 5 topics each for SARS (perplexity=1599.241, λ=0.7), influenza A (perplexity=977.101, λ = 0.8), and MERS (perplexity=1717.432, λ=0.7) and 6 topics for COVID-19 (perplexity=2515.911, λ=0.6). We sorted the topics in the order of the largest percentage of word clusters that make up the topic and identified them as themes.

XSL•FO
RenderX

Table 1. Themes and keywords of the topics.

Themes and topics	Top keywords contributing to the topic model	Token, %
Theme 1: Response to emerging infectious diseases in Korea		
SARS[a] (Topic 3)	SARS, patient, hospital bed, infection, isolation, disinfection, suspected patient, outbreak, designation, health personnel, presumed patient, National Institute of health, constitute, dedication, countermeasure	18.2
Influenza A (H1N1; Topic 2)	Influenza A virus(H1N1), infection, hub-hospital, spread, outbreak, prevention, medical staff, treatment, mask, test, suspected case, confirmed case, countermeasure, the dead, Tamiflu	21.7
MERS[b] (Topic 3)	MERS, confirmed case, infection, addition, medical staff, hospital, emergency room, treatment, exposed, isolation, Central MERS Control Countermeasure Headquarters, closed, S-hospital[c], K-hospital[c], P-hospital[c]	21.5
COVID-19 (Topic 2)	COVID-19, medical staff, mask, protective clothing, world, treatment, China, pandemic, virus, outbreak, hospital room, situation, the dead, hospitalization, pneumonia	20.8
Theme 2: Demand for nurses		
SARS (Topic 1)	The aged, nursing care, society, nurse, facility, family, role, profession, psychiatric, system, push ahead, policy, dementia, the disabled, welfare	31.2
Influenza A (H1N1; Topic 1)	Nurse, nursing college, education, service, push ahead, nursing department, the aged, establishment, system, manpower, expansion, employment, quota, the disabled, insufficiency	26.3
MERS (Topic 1)	Expansion, comprehensive nursing service, push ahead, countermeasure, hospital, provide, nursing staff, system, infection disease, caregiving, nurse's aide, measure, reinforce, improvement, nationwide	23.6
COVID-19 (Topic 3)	Securing, support, treatment, medical worker, insufficiency, deploy, dispatch, nurse, critical patient, Daegu, living treatment center, response, allowance, recruitment, dedicated hospital	17.7
Theme 3: Vulnerability in the working environment		
SARS (Topic 2)	Son and daughter, woman, child, risk, exercise, person, workplace, night duty, trust, health, nurse, research, thought, parent, prevention	19.7
Influenza A (H1N1; Topic 4)	Doctor, nurse, patient, medical, hospital bed, OECD[d], level, medical service, facility, medical expenses, Korea, region, insufficiency, intensive care unit, increase	18.7
MERS (Topic 5)	MERS, isolation, patient, suspicion, contact, action, duty, hospital, Korea Centers for Disease Control and Prevention, nurse, infection, test result, infected person, addition, emergency room	16
COVID-19 (Topic 4)	Confirmed case, infection, nurse, hospital, group infection, duty, contact, convalescent hospital, screening center, rest, emergency room, working environment, site, manpower, protective clothing	17.5
Theme 4: Roles and responsibilities of nursing professionals		
SARS (Topic 5)	Nurse, hospital, suspicion, blood, doctor, use, uncover, sexual harassment, nurse's aide, investigation, contravention, drunk driving, illegality, measurement, injection	15.4
Influenza A (H1N1; Topic 3)	Hospital, nurse, doctor, treatment, family, death, incident, contravene, criminal investigation, grandmother, death with dignity, obstetrics and gynecology, adjudge, false, infringement	21.1
MERS (Topic 2)	Hospital, medical staff, nurse, doctor, treatment, family, thought, person, heart, intensive care unit, child, oneself, protective clothing, gratitude, cheer	21.8
COVID-19 (Topic 1)	Doctor, government, Korean Medical Association, expansion, policy, push ahead, national assembly, strike, region, improvement, medical resident, Korean Nursing Association, the public, medical school quota, found	23.9
Theme 5: Government response and public opinion		
MERS (Topic 4)	MERS, situation, virus, the public, outbreak, symptom, response, fear, transmission, government, information, preventive measures, open, anxiety, the Middle East	17

XSL•FO
RenderX

Themes and topics	Top keywords contributing to the topic model	Token, %
COVID-19 (Topic 5)	Nurse, president, the public, gratitude, heart, commitment, encouragement, message, social network service, cheer, consolation, controversy, criticism, expression, labor	13
Theme 6: Vaccination for emerging infectious diseases		
Influenza A (H1N1; Topic 5)	Vaccination, student, influenza A virus(H1N1), school, vaccination, public health center, start, doctor, nurse, constitute, school nurse, plan, securing, school parent, insufficiency	12.2
COVID-19 (Topic 6)	Vaccine, inoculation, AstraZeneca, injection, Korea Disease Control and Prevention Agency, start, public health center, side effect, Pfizer, effect, nurse, nursing facility, targets, worker, launch	7.1
Theme 7: Response to emerging infectious diseases in other countries		
SARS (Topic 4)	SARS, China, the dead, Taiwan, spread, outbreak, WHO[e], Hong Kong, Singapore, infected person, death, Canada, Vietnam, transmission	15.5

[a]SARS: severe acute respiratory syndrome.

[b]MERS: Middle East respiratory syndrome.

[c]Abbreviations for specific hospital names.

[d]OECD: Organisation for Economic Co-operation and Development.

[e]WHO: World Health Organization.

Trends in Nursing-Related Topics in Media Coverage

According to the topic modeling analysis, we classified a total of 7 topics from all news articles, and the topics commonly derived from the 4 infectious disease periods were "response to emerging infectious diseases in Korea," "demand for nurses," "vulnerability in the working environment," and "roles and responsibilities of nursing professionals." Although the overarching concepts remained consistent, the keywords comprising each topic varied across the duration of infectious diseases. Notably, the keywords that comprised the topics for each epidemic period differed. We observed that the topic "vaccination for emerging infectious diseases" appeared only in the influenza A and COVID-19 pandemics, and "government response and public opinion" appeared only in the MERS and COVID-19 periods. Moreover, the topic "response to emerging infectious diseases in other countries" appeared only during SARS (Table 2). Table 1 shows the topics and word clusters identified by our model.

Table 2. Trends in nursing-related topics in media coverage, reported as tokens for each outbreak.

Trends	SARS[a], %	Influenza A (H1N1), %	MERS[b], %	COVID-19, %
Response to emerging infectious diseases in Korea	18.2	21.7	21.5	20.8
Demand for nurses	31.2	26.3	23.6	17.7
Vulnerability in the working environment	19.7	18.7	16	17.5
Roles and responsibilities of nursing professionals	15.4	21.1	21.8	23.9
Government response and public opinion	—[c]	—	17.1	13
Vaccination for emerging infectious diseases	—	12.2	—	7.1
Response to emerging infectious diseases in other countries	15.5	—	—	—

[a]SARS: severe acute respiratory syndrome.

[b]MERS: Middle East respiratory syndrome.

[c]Not applicable.

Theme 1. Response to Emerging Infectious Diseases in Korea

During the 4 periods, the "response to emerging infectious diseases in Korea" topic accounted for approximately 20%, ranking third during the SARS and MERS outbreaks and second during the influenza A and COVID-19 outbreaks. The keyword patterns were similar between SARS and MERS and between influenza A and COVID-19, reflecting trends observed in the keyword frequency analysis.

Keywords for each period include the name of the respective virus and terms describing the outbreak situation, such as "infection," "suspected patient," "outbreak," and "presumed patient" during the SARS period; "infection," "spread," "outbreak," "confirmed patient," and "deaths" during the influenza A period; "confirmed case," "infection," "additional," "quarantine," and "closure" during the MERS period; and

XSL•FO
RenderX

"China," "pandemic," "outbreak," "situation," "world," and "the dead" during the COVID-19 period, indicating the evolving scale and severity of infectious disease outbreaks in Korea over time.

In addition, keywords related to infectious disease response and preparedness such as "isolation," "disinfection," "designation," "National Institute of Health," "constitute," and "dedication" during SARS changed to "hub-hospital," "prevention," "treatment," "mask," "countermeasure," "test," and "Tamiflu" during the influenza A and MERS periods and keywords related to preparation such as specific hospital names and "hospital," "emergency room," "Central MERS Control Countermeasure Headquarters," "treatment," "isolation," and "closed" during the COVID-19 period. Keywords related to preparation, such as "mask," "protective clothing," "treatment," and "hospitalization" diminished in number, with response-focused keywords emerging in their place.

Theme 2. Demand for Nurses

The topic "demand for nurses" ranked first and represented the largest proportion during the SARS to MERS periods; during COVID-19, the topics of "roles and responsibilities of nursing professionals" and "response to emerging infectious diseases in Korea" represented higher proportions, with "demand for nurses" dropping to third place. When examining the keyword flow of topics by period, the keywords "system," "policy," "countermeasure," and "response" were associated with "the aged," "society," "facility," "psychiatric," "dementia," "the disabled," and "welfare" during the SARS period; "service," "the aged," and "the disabled" during the influenza A period; "comprehensive nursing service," "hospital," and "infection disease" during the MERS period; and "Daegu," "living treatment center," and "dedicated hospital" during the COVID-19 period, showing that the areas in which nurses were needed and the reasons for a nursing demand changed as a result of health and welfare systems and policies for infectious disease responses.

In addition, "nursing care," "family," "profession," and "role" during the SARS period; "manpower" during the influenza A period; "nursing staff," "caregiving," and "nurse's aide" during the MERS period; and "treatment," "medical worker," and "critical patient" during the COVID-19 period represent keywords related to the perception of the role of nurses. The keywords "nursing college," "nursing department," "expansion," and "quota" during the influenza A period; "nurse's aide" during the MERS period; and "support," "deploy," "dispatch," "allowance," and "recruitment" during the COVID-19 period represented changes in ways to meet the nursing demand by period.

Theme 3. Vulnerability in the Working Environment

The topic "vulnerability of the working environment" had the second highest weight during the SARS period, but its weight decreased during the influenza A and MERS periods. Its weight increased to 4th place during the COVID-19 period. During the SARS period, keywords such as "son and daughter," "woman," "parent," "workplace," "thought," "risk," "exercise," "night duty," "health," and "prevention" were topics about women's conflicts between work and childcare and health risks related to work, while during the influenza A period, keywords such as "patient," "hospital bed," "OECD," "level," "facility," "insufficiency," "intensive care unit," and "region" showed a quantitative shortage of nurses and medical facilities. During the MERS period, news reports focused on topics such as "contact," "infection," "addition," "duty," and "emergency room" regarding nurse infections, while during the COVID-19 period, news reports focused on topics such as "group infection," "duty," "contact," "working environment," "site," and "rest."

Theme 4. Roles and Responsibilities of Nursing Professionals

The topic "roles and responsibilities of nursing professionals," which was the lowest during the SARS period, showed an upward trend from the influenza A period to the COVID-19 period, and during the COVID-19 period, it had the largest change among the 6 derived topics.

The keywords of the "roles and responsibilities of nursing professionals" topic changed mainly based on specific incidents reflecting the situation at the time. During the SARS period, specific incidents related to keywords such as "blood," "drunk driving," "hospital," and "sexual harassment" included a case in which a nurse switched the blood of a drunk driving colleague with her own blood and submitted it, a case in which a doctor sexually harassed a nurse, and an illegal medical practice by a nurse or nursing assistant. During the influenza A period, keywords related to the incidents and accidents at the time, such as a case in which a newborn was switched due to a nurse's mistake, included "adjudge" and "obstetrics and gynecology" and related to bioethics, such as euthanasia, included "grandmother" and "death with dignity." During the MERS period, keywords such as "family," "thought," "heart," "gratitude," and "cheer" showed support for doctors and nurses who were infected during the MERS response process, which occurred mainly in specific hospitals. During the COVID-19 period, keywords such as "medical school quota," "expansion," "Korean Medical Association," and "Korean Nurses Association" mainly focused on issues such as doctors' strikes and nurses who stayed by patients' sides during the infectious disease outbreak.

Theme 5-7. Topics Derived Only During a Particular Infectious Disease Period

The fifth topic "government response and public opinion" was derived during the MERS and COVID-19 periods, and it had the second lowest proportion in both periods. During the MERS period, from the keywords "virus," "the public," "outbreak," "anxiety," "information," "fear," and "open," we could infer that the public experienced anxiety due to a lack of information on new infectious diseases from the government's disclosure of information. During the COVID-19 period, keywords such as "nurse," "president," "gratitude," "encouragement," "controversy," and "criticism" emphasized the controversy over the president's thank you message to nurses during the doctors' strike and the public's support for the struggles of medical staff, such as the "Thanks to Challenge," as social issues.

XSL•FO
RenderX

"Vaccination for emerging infectious diseases" was derived only during the influenza A and COVID-19 periods, and this topic had the lowest proportion of all topics in both periods. During the influenza A period, the keywords "student," "school," and "school parent" indicate that children and adolescents were the main targets of vaccination, while during the COVID-19 period, the keywords "nursing facility," "targets," and "worker" show that the targets of vaccination changed to older adults and health care workers. "AstraZeneca," "side effects," "Pfizer," and "effect," which appeared only during the COVID-19 period, indicate the situation of and public interest in the development of COVID-19 vaccines, which were not available at the beginning of the outbreak, and vaccination.

The main keywords of the "response to emerging infectious diseases in other countries" topic, which was derived only during the SARS period, consisted of the names of countries that experienced significant damage from new infectious diseases, such as "China," "Taiwan," and "Hong Kong," and keywords about the situation of the outbreak, such as "death," "outbreak," "addition," and "transmission."

Discussion

Principal Findings

This study was based on the basic premise that, in public health crises such as new infectious diseases, media reports related to the field, including nursing and nurses, become more active, reflecting high social interest in major issues. This study attempted to explore the problems that were highlighted during 4 new infectious diseases that occurred in Korea and considerations regarding social awareness and improvement to policies in relation to major nursing issues. Therefore, this study compared and analyzed the main agenda and potential meaning of news articles on nursing and nurses for each new infectious disease period using the big data analysis techniques of topic modeling and keyword network analysis. In addition, we discussed the understanding of current nursing phenomena and the trends and implications of public opinion surrounding nursing based on the results.

As a result of topic modeling analysis, 7 topics were classified from all news articles. The commonly derived topics from the 4 infectious disease periods included "demand for nurses," "response to emerging infectious diseases in Korea," "vulnerability in the working environment," and "roles and responsibilities of nursing professionals." Topics with the same higher concept had changes in the associated keywords according to the infectious disease period. The following sections discuss the potential meaning in each topic by referring to the visualization results of the related keyword network map, focusing on the topics commonly derived during the new infectious disease periods.

Response to Emerging Infectious Diseases in Korea

First, the keywords for the topics during the SARS, MERS, influenza A, and COVID-19 periods showed similar patterns when analyzing the relevance of the topic by period. Depending on the duration of the infectious disease, the keywords "outbreak," "spread," "addition," and "pandemic" were used to express the seriousness of the domestic infectious disease situation. The potential meaning of the words related to the keywords "National Institute of Health," "dedication," and "designation" during SARS and "hub-hospital" during influenza A was the government's practical policy proposal to prepare for new infectious diseases. These policies included the designation of regional hub hospitals and securing state-designated inpatient treatment beds that were proposed during SARS and influenza A and were implemented and partially improved after the end of those infectious disease periods. However, the number of public medical resources, such as treatment beds and health professionals, that could be practically used was significantly insufficient for the number of infected patients during the MERS and COVID-19 periods [24].

In addition, keywords such as "medical staff," "infection," "closed," "hospital room," and "hospitalization" during MERS and COVID-19 highlighted vulnerabilities in emergency room environments, including inadequate infection control measures and gaps in infection control in nursing hospitals and nursing facilities where older adult patients, who are vulnerable to infection, are mainly hospitalized. The emergency room is where infected patients primarily access the hospital. Problems including the lack of facilities such as separate isolation zones and negative pressure isolation rooms for the treatment of infected patients revealed during the MERS period, the lack of education on infection control, and the response of emergency medical workers increase the risk of medical staff becoming infected and are viciously linked to patient infection and the lack of response personnel [25]. In addition, in emergency rooms, there is a greater need for systematic infection control by professional personnel such as nurses because older adult patients in nursing hospitals and facilities are more likely to contract infectious diseases due to low immunity and group living and the risk of death from infection is high [26,27]. This is a matter to be considered first in terms of policy for infection control in environments where patients and medical personnel are at a high risk of infection. After MERS, the enforcement regulations outlined in the Emergency Medical Service Act were revised to improve emergency room facilities and manpower standards [28], and through the revision of the enforcement regulations of the Medical Act during the COVID-19 period, the standards and fees for mandatory infection control rooms for infection control were expanded [29]. However, the effectiveness of the policy is still limited because it does not resolve gaps in infection control outside the scope of the policy, such as the lack of infection control experts and nursing hospitals and facilities with fewer than 100 beds. Although policies such as the revision of the Emergency Medical Service Act aimed to address these gaps, their effectiveness remains limited due to insufficient implementation and resource allocation. Similar challenges in emergency room preparedness were noted in countries such as Italy and the United States during the COVID-19 pandemic, underscoring the global nature of these vulnerabilities.

Demand for Nurses

In the "demand for nurses" topic, the need to increase the number of nurses has been steadily highlighted over the past

XSL•FO
RenderX

20 years, and the keyword "insufficiency," which appeared only during the influenza A and COVID-19 periods shows that the shortage of nurses during the COVID-19 period was also a problem that was raised during the influenza A period. In the topic analysis results, the focus on nursing demand changed from health and welfare services to responding to infectious diseases, and in the change of keywords over time, the area of public interest also changed according to the reason for the demand in nurses and the field of activity. The areas where nurses' activities are performed are diverse and include medical institutions, communities, schools, and businesses; nurses are also recognized as particularly indispensable for the national welfare service, especially in a society where the number of older adult and single-person households is increasing [30,31]. Among the government's systems related to the demand for nurses, the comprehensive nursing service (currently, the integrated nursing care service), which was introduced in 2013 to reduce the burden of care and expenses for people in a nuclear family and an aging society, accelerated the expansion of the service, as domestic nursing and door-to-door cultural issues were raised in terms of hospital infection control during the MERS period in 2015. Nurses with specialized knowledge and skills were recognized as a major element of the service, which proceeded to actively use nurses [32,33]. Many previous studies on integrated nursing care services have shown positive effects of these services for improving patient safety, reducing hospital infections, and increasing patient satisfaction with medical services [32,34,35]. Accordingly, there seems to be no difference in public opinion on the need for a change in the demographic and social structure, the response to infectious disasters at the health care level, a quantitative increase in nurse demand and quantity, and policies to increase the number nurses.

In terms of meeting the demand for nursing during periods of new infectious diseases, the focus was on recruiting nurses and distributing manpower to prevent the infection of the vulnerable, such as older adults, and school quarantine during the SARS and influenza A periods. Changes for the issue of nursing supply were mainly made to comprehensive nursing services as a preventive measure against hospital infections. During the COVID-19 period, there was a shortage of nursing personnel who could be immediately placed in response to infectious diseases in medical institutions in nonmetropolitan areas such as Daegu, life treatment centers, and hospitals dedicated to infectious diseases. The government filled the demand for nurses with temporary measures through the dispatch of nursing personnel recruited through the input of nursing officers and allowances. However, the urgent recruitment of the personnel needed in the field of infectious diseases was ineffective due to insufficient nursing experience, education, and training [36,37]. It is believed that the gap in the nursing field continues because there is a real need for nurses with the experience and capacity to care for infectious disease patients and intensive care patients, while the government's policies for and public opinion about increasing the nursing supply focused purely on the number of nurses rather than the quality and capacity of nursing.

The fact that the policy to increase the number of nurses recognized nursing as a job that not only resolves the imbalance of local medical personnel or meets medical needs but also resolves employment difficulties and guarantees employment after graduation is similar to the ideas expressed in the original news article [38] titled "Graduation is Employment, Nursing and Explosion of Popularity" and research showing that high school students who are about to pursue careers perceive nursing as an economically stable and good job compared with other occupations [39]. In addition, in research analyzing the meaning of work for new nurses in 2021, more than 50% of the nurses had family and relatives who recommended they pursue nursing, or they applied to the nursing department because they were able to easily get a job after graduation [40].

Since 2008, the year just before the influenza A outbreak, Korea implemented a policy to establish a nursing department and increase the number of admissions. As a result of this policy, during the swine flu period, there were 414,000 licensed nurses in 2019, and 20,000 new nurses entered the workforce every year [30]. However, there are only 215,000 nurses working in medical and health institutions, which represents about one-half of the licensed nurses [41]. According to data from a survey on the status of hospital nursing staff placements every year since 2010, the turnover of new nurses increased from 30.5% in 2011 to 47.7% in 2020, and the number one reason for resignation within a year after the announcement was "work maladjustment" [42,43]. The burden of work on new nurses was correlated with the reality shock caused by the conflict between the hospital organizational culture and nursing professionals [44]. Nurses are aware of the risk of infection at infectious disease sites, but they have a sense of calling and must perform patient care at the forefront [45,46]. In patient care, the experience of a conflict between infection risk situations and professional beliefs increased nurses' intentions to leave more than the general situation of care [10,47]. Based on these preceding studies, there is likely a difference between the actual clinical field and the public's perception of nursing in terms of the purpose of employment identified in the results of this study. Although economic feasibility cannot be excluded from policies to resolve employment difficulties, it is necessary to consider nursing professionals in policy and media in order to prevent the departure of new nurses, increase the number of experienced nurses, and change the public's perception of nurses.

Vulnerability in the Working Environment

In the "demand for nurses" topic during the influenza A and COVID-19 outbreaks, a critical issue was the shortage of skilled nurses—particularly in nonmetropolitan areas—in addition to skills to handle special equipment such as ventilators and extracorporeal membrane oxygen supply required for intensive care nursing as well as clinical experience [36]. Given the strong link between nursing demand, supply, and working conditions, the topics "demand for nurses" and "vulnerability in the working environment" are closely interconnected [37] and can be interpreted together. In this study, hot topics included conflicts between women's work-family balance during the SARS period, health risks due to night work, domestic medical facilities and manpower at a lower level than in OECD countries during the influenza A period, and the shortage of nurses in nonmetropolitan areas. The changes in keywords such as high secondary infections of nurses related to new infections, emergency rooms, and nursing hospitals; work weighting; and

XSL•FO
RenderX

society's expectations for sacrifice and commitment should be considered as major variables to resolve the shortage of nurses.

According to a survey by the Hospital Nurses Association, the average resignation rate of domestic nurses over the past decade has been over 13%. When the reasons for resignation in 2010 and 2020 were compared, the second highest reason of "marriage, childbirth, and childcare" in 2010 fell significantly from 14.8% of responses in 2010 to 6.3% of responses in 2020. "Work maladjustment," which was ranked third (13.1%) in 2010, was ranked first (17.1%) in 2020, and "transition to other jobs," which was ranked second (5.9%) in 2010, increased to 12.2% in 2020 [42,43]. This is likely the result of government policy and efforts to improve the conflict that comes from the balance between work and family for female nurses, which appeared in the topic during SARS. However, high-risk working environments for infection and the lack of experienced nurses during MERS and COVID-19, during which there was a focus on the quantitative increase of nurses following the influenza period; the policy to increase the number of nurses; and the poor treatment suggest that more active interventions and improvement are needed at the organizational level.

Inadequate facilities in which to rest and the social stigma of being perceived as infection carriers—both personally and for family members—accelerate burnout and increase turnover intentions among nurses, and secondary infections among nurses create a vicious cycle of labor shortages and increased patient infections [48,49]. The government expanded the admission quota for nursing colleges starting in 2008 in an effort to increase the absolute total number of nurses, but it did not have a significant impact on resolving the shortage of nurses in local small and medium-sized hospitals [50]. In addition, in March 2018, the Ministry of Health and Welfare announced policies aimed at improving nurses' working conditions, addressing human rights violations such as workplace bullying and sexual harassment, expanding nursing personnel, strengthening professionalism, improving the quality of nursing services, and creating a policy foundation for nursing personnel [51]. In this regard, the system for calculating inpatient nursing management fees was revised to base payments on patient-to-nurse ratios according to the actual situation in local hospitals in April 2018. The system was improved so that additional profits from hospitalization fees generated by changes based on the number of patients were used to improve the treatment of nurses. In addition, the night-only nurse management fee and the night-care fee for general wards established in October 2019 were applied to general hospitals and hospitals outside of Seoul so that additional allowances for night work could be paid. However, since both systems are merely recommendations, there were practical limitations for nurses to feel that their treatment was improved [52]. In February 2019, a Task Force Team was promoted to oversee nursing policies such as managing the supply and demand of nursing personnel and improving the working environment. In May 2021, a nursing policy department was established in the Ministry of Health and Welfare with social consensus on the need for a dedicated department during the COVID-19 period [42]. As such, the government has continuously implemented policies to address the shortage of nurses, and evaluation of the effectiveness of some policies is expected to take some time. This is an opportune time to reconsider whether fragmented and temporary approaches to nursing challenges have exacerbated systemic issues, such as the revision of the medical law that tried to reorganize nursing assistants from the "demand for nurses" topic during the MERS period or the "vulnerability in the working environment" topic, which is related to a high risk of infectious diseases and the need for high-intensity work. In addition, active monitoring is needed to determine how closely a series of policies has approached the practical problems of the nursing field, and feedback mechanisms during a changing infectious disease situation are needed. Although these policies mark progress, their voluntary nature limits their effectiveness in achieving widespread improvements.

Roles and Responsibilities of Nursing Professionals

The keyword changes in the topic of "roles and responsibilities of nursing professionals" mainly included incidents and accidents related to the morality of nurses, such as drunk driving and illegal medical practices during the SARS and influenza A periods. Given the focus on keywords related to infectious diseases during the MERS and COVID-19 periods, the conflict over work ethics in the nursing of infected patients versus the nurse's own safety and social expectations was considered a major agenda.

Since the image of nurses formed in the media has been considered to influence potential future nurses' career choices and intentions and some stories about nurses can negatively influence not only nursing supply but also potential improvements in the working environment and treatment of nurses, many previous studies have proposed monitoring of the images of nursing expressed in the media and internal management of the nursing community [3,53].

The analysis of correlations between keywords during the MERS period highlighted social expectations for nurses in special situations such as public health disasters as well as conflicts between nurses' protective instincts and work ethics and their families and colleagues. In this regard, the original news articles reviewed in this study described the critical condition of medical personnel who had secondary MERS infections and experiences with social isolation, such as nurses' children being ostracized by their neighbors due to the stigma of working with MERS-infected patients [54]. The ambivalence of social expectations for nurses responding to infectious diseases is reflected by the description of medical personnel as heroes through the "Thanks to the Challenge" campaign and the praise for volunteer nurses in Daegu and Gyeongbuk. The precedent of the government's legal response to the collective resignation of Taiwanese medical staff due to the infection and death of medical personnel is another example and was captured in the topic "response to emerging infectious diseases in other countries" during SARS. The ambivalence was also present in an overseas example [3] that described a nurse volunteering at an infectious disease site during the Ebola outbreak as a virus eradication hero versus an infection spreader and villain. Since social stigma and isolation during an infectious disease outbreak add to the mental and social stress of nurses and influence nursing and turnover intentions, the following are needed:

XSL•FO
RenderX

accurate information delivery by the media, a stable medical system, and an organizational protection and support system [48,49].

Meanwhile, in the context of the COVID-19 period, the government attempted to implement a policy to expand the number of medical schools to respond to infectious diseases, and the related medical strike led to the aggravation of nurses' work [55]. Previous studies have highlighted the 2-sided phenomenon of social expectations, which consists of systemic problems such as poor hospital infection control, inadequate government response, and increased public fear through negative media reports rather than the attitude of medical personnel, including nurses, when responding to infectious diseases [6,48]. This is similar to the relationships between the topics of "government response and public opinion" and "roles and responsibilities of nursing professionals," which include indiscriminate disclosure of information about infectious diseases in the media during the MERS period and the controversy over the president's thank you message to nurses during a doctor strike during the COVID-19 period. This suggests that active interest and the adoption of a leading role by the government are needed to avoid delays in policy and social consensus as a result of insufficient government responses and issue-oriented media reports.

Limitations

In this study, limitations were related to the search used to identify issues in the nursing field as well as the data collection period, which was restricted to the periods around outbreaks of new infectious diseases and limited the ability to compare the situation before and after an outbreak. Therefore, future research should adopt a multifaceted approach that incorporates time series analysis, dynamic topic modeling, and expanded data collection periods to capture broader trends. In addition, as this study started during the COVID-19 outbreak and we were unable to predict when the outbreak would end, we collected data only until May 31, 2021, due to limited time available for data analysis. Therefore, any changes in the situation after that data were not reflected in the research results, and we recommend collecting and analyzing data after that date.

Conclusion

This study analyzed the relevance of nursing-related agendas in news articles during 4 major EID outbreaks. Our analysis revealed that nurse shortages, poor working conditions, and inadequate treatment of nurses were recurring issues across the epidemics and were not unique to the COVID-19 pandemic. Although society and nursing professionals agree on the essential role of nurses in the health sector, significant differences persist in perceptions of their roles and responsibilities. The results suggest that retaining experienced nurses to strengthen crisis response capabilities should be closely linked to the working environment. Addressing nurse shortages requires not only increasing their numbers but also implementing substantial policy support to ensure safe working environments, better treatment conditions, and heightened societal recognition of their roles. Medical organizations, professional associations, governments, and the media should develop policies that incorporate nurses' perspectives. These efforts can foster positive social consensus on nursing and support a stable health care system during health crises. Our findings are expected to be used as a basis for establishing strategies to improve the professional role, environment, and treatment of nurses during health crisis situations. Finally, these findings emphasize the need for holistic reforms to strengthen nursing's role in mitigating future health crises. We recommend further research on the effects and changes of nursing-related policies.

Acknowledgments

This research was supported by the Government-wide R&D Fund for Infectious Disease Research (GFID), Republic of Korea (grant number RS-2022-KH125928 [HG22C0051]).

Authors' Contributions

EKY and JO collaboratively developed the study design and co-authored the initial manuscript draft. EKY supervised the study and validated the findings, while JO collected and preprocessed the data and conducted the analysis. Both authors interpreted the results and contributed to the writing and revision of the final manuscript.

Conflicts of Interest

None declared.

References

1. McArthur DB. Emerging infectious diseases. Nurs Clin North Am 2019 Jun;54(2):297-311 [FREE Full text] [doi: 10.1016/j.cnur.2019.02.006] [Medline: 31027668]
2. Infectious Disease Information. Korea Disease Control and Prevention Agency. URL: https://dportal.kdca.go.kr/pot/cv/trend/dmstc/selectMntrgSttus.do [accessed 2024-11-16]
3. McGillis Hall L, Kashin J. Public understanding of the role of nurses During Ebola. J Nurs Scholarsh 2016 Jan;48(1):91-97. [doi: 10.1111/jnu.12182] [Medline: 26642005]
4. Nayna Schwerdtle P, Connell CJ, Lee S, Plummer V, Russo PL, Endacott R, et al. Nurse expertise: a critical resource in the COVID-19 pandemic response. Ann Glob Health 2020 May 11;86(1):49 [FREE Full text] [doi: 10.5334/aogh.2898] [Medline: 32435602]

XSL•FO
RenderX

5. Chitty KK. Professional Nursing: Concepts and Challenges. Philadelphia, PA: Saunders; 2005.
6. Choi EK. Ethical responses to public health emergencies: the 2015 MERS outbreak in South Korea. Korean J Med Ethics 2016 Sep 30;19(3):358-374 [FREE Full text] [doi: 10.35301/ksme.2016.19.3.358]
7. Donelan K, Buerhaus P, DesRoches C, Dittus R, Dutwin D. Public perceptions of nursing careers: the influence of the media and nursing shortages. Nurs Econ 2008;26(3):143-50, 165. [Medline: 18616051]
8. Ehrenstein BP, Hanses F, Salzberger B. Influenza pandemic and professional duty: family or patients first? A survey of hospital employees. BMC Public Health 2006 Dec 28;6:311 [FREE Full text] [doi: 10.1186/1471-2458-6-311] [Medline: 17192198]
9. The COVID-19 Effect: World's nurses facing mass trauma, an immediate danger to the profession and future of our health systems. International Council of Nurses. 2021 Jan 13. URL: https://www.icn.ch/news/covid-19-effect-worlds-nurses-facing-mass-trauma-immediate-danger-profession-and-future-our [accessed 2024-05-09]
10. Labrague LJ, de Los Santos JAA. Fear of COVID-19, psychological distress, work satisfaction and turnover intention among frontline nurses. J Nurs Manag 2021 Apr;29(3):395-403 [FREE Full text] [doi: 10.1111/jonm.13168] [Medline: 32985046]
11. Ruderman C, Tracy CS, Bensimon CM, Bernstein M, Hawryluck L, Shaul RZ, et al. On pandemics and the duty to care: whose duty? who cares? BMC Med Ethics 2006 Apr 20;7:E5 [FREE Full text] [doi: 10.1186/1472-6939-7-5] [Medline: 16626488]
12. Sokol DK. Virulent epidemics and scope of healthcare workers' duty of care. Emerg Infect Dis 2006 Aug;12(8):1238-1241 [FREE Full text] [doi: 10.3201/eid1208.060360] [Medline: 16965703]
13. Brody H, Avery EN. Medicine's duty to treat pandemic illness: solidarity and vulnerability. Hastings Cent Rep 2009;39(1):40-48. [doi: 10.1353/hcr.0.0104] [Medline: 19213194]
14. Martin SD. Nurses' ability and willingness to work during pandemic flu. J Nurs Manag 2011 Jan;19(1):98-108. [doi: 10.1111/j.1365-2834.2010.01190.x] [Medline: 21223410]
15. Gollust SE, Fowler EF, Niederdeppe J. Television news coverage of public health issues and implications for public health policy and practice. Annu Rev Public Health 2019 Apr 01;40:167-185. [doi: 10.1146/annurev-publhealth-040218-044017] [Medline: 30633711]
16. Grün B, Hornik K. topicmodels: an R package for fitting topic models. Journal of Statistical Software 2011;40(13):1-30 [FREE Full text] [doi: 10.18637/jss.v040.i13]
17. Kang J, Kim S, Roh S. [A topic modeling analysis for online news article comments on nurses' workplace bullying]. J Korean Acad Nurs 2019 Dec;49(6):736-747. [doi: 10.4040/jkan.2019.49.6.736] [Medline: 31932568]
18. Kim BK, Kim MK. Keywords network analysis of temple stay using big-data. Tourism Studies 2015;39(5):27-40 [FREE Full text] [doi: 10.17086/JTS.2015.39.5.27.40]
19. Yoon YM, Kim SK, Kim HK, Kim EJ, Jeong Y. Comparison of topics related to nurse on the internet portals and social media before and during the COVID-19 era using topic modeling. Journal of Muscle and Joint Health 2020;27(3):255-267 [FREE Full text] [doi: 10.5953/JMJH.2020.27.3.255]
20. Bigkinds. URL: https://www.bigkinds.or.kr/ [accessed 2024-05-09]
21. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text]
22. Steyvers M, Griffiths T. Probabilistic topic models. In: Landauer TK, McNamara DS, Dennis S, Kintsch W, editors. Handbook of Latent Semantic Analysis. New York, NY: Psychology Press; 2007:427-448.
23. Dotsika F, Watkins A. Identifying potentially disruptive trends by means of keyword network analysis. Technological Forecasting and Social Change 2017 Jun;119:114-127. [doi: 10.1016/j.techfore.2017.03.020]
24. Yoon KJ. Infectious diseases and public health care as seen through the response to COVID-19. Health and Welfare Issues & Focus 2020;377:1-11 [FREE Full text]
25. Yoon S. Deriving the effective improvement elements of the emergency room in large general hospitals. Journal of Critical Care Nursing 2015;8(2):66-79 [FREE Full text]
26. Son C, Yoon M, Kim S, Jo Y. Current Status and Challenges of COVID-19 Infections in Seoul City's Nursing Homes. The Seoul Institute. 2021 May 04. URL: https://www.si.re.kr/node/64880 [accessed 2024-12-27]
27. Choi MJ, Joung YS, Lee HY, Jeong HS. Analysis of needs for infection control education in long-term care hospitals for the epidemic of COVID-19: Borich priority formula and the locus for focus model. Health and Social Welfare Review 2021;41(1):8-21 [FREE Full text] [doi: 10.15709/HSWR.2021.41.1.8]
28. Kim Y. Healthcare reform after MERS outbreak: progress to date and next steps. J Korean Med Assoc 2016;59(9):668. [doi: 10.5124/jkma.2016.59.9.668]
29. The 27th Health Insurance Policy Review Committee in 2021. Korean Ministry of Health and Welfare. 2021 Dec 23. URL: https://www.mohw.go.kr/board.es?mid=a20401000000&bid=0032&tag=&act=view&list_no=369057 [accessed 2024-12-28]
30. Kim J, Kwon H, Park E, Choi H, Lee S. Nursing workforce supply and demand: review and prospect in South Korea. Health Economics and Policy Research 2021;27(4):27-49 [FREE Full text]
31. Beggs-Yeager C, Sharts-Hopko N, McDermott-Levy R. The role of nurses in surveillance to enhance global health security: A delphi study. Nurs Outlook 2021;69(6):1021-1029. [doi: 10.1016/j.outlook.2021.05.011] [Medline: 34332765]

XSL•FO
RenderX

32. Kim J, Kim S, Park E, Jeong S, Lee E. Policy issues and new direction for comprehensive nursing service in the National Health Insurance. J Korean Acad Nurs Adm 2017;23(3):312. [doi: 10.11111/jkana.2017.23.3.312]
33. Jun B. Middle East respiratory syndrome outbreak and infectious disease control in Korea. J Korean Med Assoc 2015;58(7):590. [doi: 10.5124/jkma.2015.58.7.590]
34. Park JH, Lee MH. Effects of nursing and care-giving integrated service on nursing work performance, nurses' job satisfaction and patient safety. Journal of Home Health Care Nursing 2017;24(1):14-22 [FREE Full text] [doi: 10.22705/jkashcn.2017.24.1.014]
35. Son J, Park BK, Lee CH, Ahn KH, Kim JN, Park MH, et al. Incidence rate and characteristics of falls in patients using comprehensive nursing care service. Health Policy and Management 2019;29(2):172-183. [doi: 10.4332/KJHPA.2019.29.2.172]
36. Park S. Corona emergency, but only 4 government-supported nurses. Kangwon News. 2020 Nov 27. URL: https://www.kwnews.co.kr/page/view/2020112600000000114 [accessed 2024-05-09]
37. Shin KR. COVID-19 counter measures and action plan for improvement of nursing treatment. HIRA Research 2021;1(1):103-107 [FREE Full text] [doi: 10.52937/hira.21.1.1.103]
38. Choi C. Graduation leads to employment. Nursing department popularity explodes. Mail News. URL: https://www.imaeil.com/page/view/2009092909391865942 [accessed 2024-05-09]
39. Yeun EJ, Han SY, Lee IJ. Subjectivity on nursing images among high school students: an application of Q-methodology. Korean Society for Subjectivity Studies 2010;21:179-194 [FREE Full text]
40. Kim M, Seo H, Lee J. Relationships between meaning of work, organizational commitment, professional self-image, and retention intention among new nurses. Journal of Korean Clinical Nursing Research 2022;28(1):54-65. [doi: 10.22650/JKCNR.2022.28.1.54]
41. Nursing Statistics. Korean Nursing Association. 2022. URL: https://www.koreanurse.or.kr/resources/statistics.php [accessed 2024-05-09]
42. Analysis of trends in hospital nursing staff deployment status and working conditions survey from 2010 to 2019. Reference of Hospital Nurses Association. URL: https://khna.or.kr/home/pds/utilities.php?bo_table=board1&wr_id=8105 [accessed 2024-05-09]
43. 2021 Hospital Nursing Staffing Status Survey. Reference of Hospital Nurses Association. URL: https://khna.or.kr/home/pds/utilities.php?bo_table=board1&wr_id=8103&page=1 [accessed 2024-05-09]
44. Kim WS. New nurses on reality shock and organizational commitment convergence study. Journal of Digital Convergence 2020;18(11):377-386. [doi: 10.14400/JDC.2020.18.11.377]
45. Article 5 of the Korean Nurses' Ethical Guidelines (Provision of Ethical Nursing). Korean Nursing Association. URL: http://www.koreanurse.or.kr/about_KNA/ethics.php [accessed 2024-11-30]
46. Hsin DH, Macer DR. Heroes of SARS: professional roles and ethics of health care workers. J Infect 2004 Oct;49(3):210-215 [FREE Full text] [doi: 10.1016/j.jinf.2004.06.005] [Medline: 15337337]
47. Bruyneel A, Dello S, Dauvergne JE, Kohnen D, Sermeus W. Prevalence and risk factors for burnout, missed nursing care, and intention-to-leave the job among intensive care unit and general ward nurses: A cross-sectional study across six European countries in the COVID-19 era. Intensive Crit Care Nurs 2025 Feb;86:103885. [doi: 10.1016/j.iccn.2024.103885] [Medline: 39522308]
48. Kim JY. Nurses' experience of Middle East Respiratory Syndrome patients care. Journal of the Korea Academia-Industrial cooperation Society 2017;18(10):185-196 [FREE Full text] [doi: 10.5762/KAIS.2017.18.10.185]
49. Jin D, Lee G. Experiences of nurses at a general hospital in Seoul which is temporarily closed due to COVID-19. J Korean Acad Soc Nurs Educ 2020 Nov;26(4):412-422. [doi: 10.5977/jkasne.2020.26.4.412]
50. Bae H, Kim J. Impact of increasing an enrollment quota in nursing schools on nurse staffing level in local hospitals. Korean Data Anal Soc 2017;19(6):3301-3316 [FREE Full text] [doi: 10.37727/jkdas.2017.19.6.3301]
51. Ministry of Health and Welfare announces “Measures to Improve Nurses’ Working Environment and Treatment”. Korean Ministry of Health and Welfare. 2018 Mar 20. URL: https://www.mohw.go.kr/board.es?mid=a10503010100&bid=0027&act=view&list_no=344262&tag=&nPage=1 [accessed 2024-11-30]
52. Cho S, Hong KJ, Lee J. Estimation of revenue growth generated by changing nurse staffing grades and fees and introducing night shift nursing fees. J Korean Acad Nurs Adm 2021;27(3):191. [doi: 10.11111/jkana.2021.27.3.191]
53. Hall LM, Angus J, Peter E, O'Brien-Pallas L, Wynn F, Donner G. Media portrayal of nurses' perspectives and concerns in the SARS crisis in Toronto. J Nurs Scholarsh 2003 Apr 23;35(3):211-216 [FREE Full text] [doi: 10.1111/j.1547-5069.2003.00211.x] [Medline: 14562487]
54. MERS Fight Medical Staff Bullying Growing...Even Children Suffer. Kim IY. 2015 Jun 16. URL: https://www.yna.co.kr/view/AKR20150616167200063?input=1195m [accessed 2024-05-09]
55. Lee J. Nurses collapse trying to fill doctors' absence. The Hankyoreh Newspaper. 2020 Aug 25. URL: https://www.hani.co.kr/arti/society/society_general/959334.html [accessed 2024-05-09]

XSL•FO
RenderX

Abbreviations

EID: emerging infectious disease
LDA: latent Dirichlet allocation
MERS: Middle East respiratory syndrome
OECD: Organisation for Economic Co-operation and Development
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by A Mavragani; submitted 11.05.24; peer-reviewed by H Lee, PH Liao, CY Han, S Pesälä; comments to author 05.11.24; revised version received 07.12.24; accepted 21.12.24; published 06.01.25.

Please cite as:
Kim J, Yun EK
Topic Modeling of Nursing Issues in the Media During 4 Emerging Infectious Disease Epidemics in South Korea: Descriptive Analysis
J Med Internet Res 2025;27:e60446
URL: https://www.jmir.org/2025/1/e60446
doi:10.2196/60446
PMID:

©Jungok Kim, Eun Kyoung Yun. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Willingness of Doctors to Adopt Artificial Intelligence–Driven Clinical Decision Support Systems at Different Hospitals in China: Fuzzy Set Qualitative Comparative Analysis of Survey Data

Zhongguang Yu[1,2*], PhD; Ning Hu[3*], MSc; Qiuyi Zhao[4], MSc; Xiang Hu[5], PhD; Cunbo Jia[6], PhD; Chunyu Zhang[7], PhD; Bing Liu[6], MSc; Yanping Li[1], PhD

[1]Economics and Management School, Wuhan University, Wuhan, China

[2]Respiratory Centre, China-Japan Friendship Hospital, Beijing, China

[3]School of Management, Beijing University of Chinese Medicine, Beijing, China

[4]The Fourth Affiliated Hospital of Soochow University (Suzhou Dushu Lake Hospital), Suzhou, China

[5]Business School, Hubei University, Wuhan, China

[6]Hospital Office, China-Japan Friendship Hospital, Beijing, China

[7]Department of Human Resources, China-Japan Friendship Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yanping Li, PhD
Economics and Management School
Wuhan University
299 Bayi Road, Wuchang District
Wuhan, 430072
China
Phone: 86 68753084
Fax: 86 68754150
Email: ypli@whu.edu.cn

Abstract

Background: Artificial intelligence–driven clinical decision support systems (AI-CDSSs) are pivotal tools for doctors to improve diagnostic and treatment processes, as well as improve the efficiency and quality of health care services. However, not all doctors trust artificial intelligence (AI) technology, and many remain skeptical and unwilling to adopt these systems.

Objective: This study aimed to explore in depth the factors influencing doctors' willingness to adopt AI-CDSSs and assess the causal relationships among these factors to gain a better understanding for promoting the clinical application and widespread implementation of these systems.

Methods: Based on the unified theory of acceptance and use of technology (UTAUT) and the technology-organization-environment (TOE) framework, we have proposed and designed a framework for doctors' willingness to adopt AI-CDSSs. We conducted a nationwide questionnaire survey in China and performed fuzzy set qualitative comparative analysis to explore the willingness of doctors to adopt AI-CDSSs in different types of medical institutions and assess the factors influencing their willingness.

Results: The survey was administered to doctors working in tertiary hospitals and primary/secondary hospitals across China. We received 450 valid responses out of 578 questionnaires distributed, indicating a robust response rate of 77.9%. Our analysis of the influencing factors and adoption pathways revealed that doctors in tertiary hospitals exhibited 6 distinct pathways for AI-CDSS adoption, which were centered on technology-driven pathways, individual-driven pathways, and technology-individual dual-driven pathways. Doctors in primary/secondary hospitals demonstrated 3 adoption pathways, which were centered on technology-individual and organization-individual dual-driven pathways. There were commonalities in the factors influencing adoption across different medical institutions, such as the positive perception of AI technology's utility and individual readiness to try new technologies. There were also variations in the influence of facilitating conditions among doctors at different medical institutions, especially primary/secondary hospitals.

XSL•FO
RenderX

Conclusions: From the perspective of the 6 pathways for doctors at tertiary hospitals and the 3 pathways for doctors at primary/secondary hospitals, performance expectancy and personal innovativeness were 2 indispensable and core conditions in the pathways to achieving favorable willingness to adopt AI-CDSSs.

(J Med Internet Res 2025;27:e62768) doi:10.2196/62768

KEYWORDS

artificial intelligence; clinical decision support systems; willingness; technology adoption; fuzzy set qualitative comparative analysis; fsQCA; pathways

Introduction

Background

Artificial intelligence–driven clinical decision support systems (AI-CDSSs), aimed at enhancing medical decision-making, have been in use since the 1980s [1]. These systems leverage clinical knowledge, patient information, and health data to improve health care services by allowing intelligent detection, management, and improvement of patient health conditions [2,3]. According to statistics, the total number of outpatient visits to medical and health institutions in China reached 5.5 billion in 2023, with doctors at public hospitals averaging 7.1 outpatient visits and 2.3 hospital visits per day and with the burden on primary health care institutional doctors being even greater [4]. According to "World Health Statistics 2023: Monitoring Health for the SDGs, sustainable development goals" released by the World Health Organization (WHO), China has a low doctor density and uneven distribution of health care resources [5]. To relieve the pressure on doctors and improve the efficiency of medical services, China has actively explored the application of AI-CDSSs, with the National Health Commission of China issuing a national policy on the promotion of clinical decision support systems (CDSSs) in 2023.

With the evolution of artificial intelligence (AI) technology, current AI-CDSSs can generate specific assessments and recommendations through logical reasoning, analyze patient data from electronic medical records, and offer decision support to health care providers [6]. AI-CDSSs are widely used in clinical diagnoses, drug treatments, preventive measures, and patient management, and are recognized for their roles in enhancing health care decision-making and doctor performance [7]. For instance, scholars, such as Hanson et al [8], have conducted research from the perspective of patient management, evaluating the application of AI-CDSSs in the nursing field, and have confirmed that certain systems are able to improve the accuracy and comprehensiveness of nursing treatment. Alsharqi et al [9], among others, have evaluated the application effects of AI-CDSSs in the automatic image selection field of echocardiography. They found that AI-CDSSs can effectively identify, distinguish, and explain images through machine learning models. Islam et al [10] studied how AI-CDSSs could help patients continuously observe different parameters for controlling insulin levels, automatically analyzing the personal data of diabetic patients. The application of AI-CDSSs in the field of anesthesiology is also extensive. According to existing research, AI-CDSSs can improve the preoperative use of antibiotics and beta-blockers, reduce the use of inhaled anesthetics, and assist in completing anesthesia records and billing work [11-13].

AI-CDSSs are primarily targeted at physicians, whose acceptance is pivotal for their implementation. Sambasivan et al [14] extended the unified theory of acceptance and use of technology (UTAUT) model and explored the factors influencing doctors' willingness to adopt new CDSSs in a developing country context, considering user involvement in decision-making, perceived threat to autonomy, effort expectancy, and performance expectancy. Their study involved doctors from 12 hospitals representing 10 different specialty areas. However, physicians' attitudes toward these systems varied, with some exhibiting favorable inclinations and others exhibiting reluctance, primarily attributed to apprehensions about technological readiness, data confidentiality, absence of personalized interaction, and skepticism toward emerging technologies.

Thus, although some doctors have expressed positive attitudes toward adopting AI-CDSSs, others still harbor reservations, citing concerns over technology immaturity, data privacy and security, absence of human touch, and a distrust of new technologies. For instance, Wagner et al [15] found that only 54.9% of family doctors were willing to use AI for medical diagnosis, while O'Leary et al [16] reported that 82% of health care professionals saw the usefulness of AI in diagnosing rare diseases. In their 2020 study, Park et al [17] discovered that over 75% of American radiology students believed AI would play a crucial role in future medicine, while almost half of them expressed decreased enthusiasm for radiology due to AI adoption. Scheetz et al [18] found that 71% of the health care professionals surveyed in Australia and New Zealand believed AI would improve medicine, while 85.7% felt it would impact health care manpower. In a South Korean study, only 5.9% of doctors were very familiar with AI [19]. Similarly, in a UK study, medical students indicated they were not adequately prepared for AI. Scholars have also expressed concerns about humans being replaced by AI in health care [20].

Thus, despite the potential benefits, skepticism persists among scholars and doctors regarding the integration of AI in health care, citing concerns about technology replacing human roles. For instance, Poon and Sung [21] highlighted doctors' skepticism toward AI technology in clinical practice that impeded the progress of AI applications owing to a lack of trust.

Investigating doctors' willingness to adopt AI-CDSSs and understanding the factors influencing their acceptance can have a significant impact on the comprehensive integration of AI-CDSSs into clinical applications. Sambasivan et al [14]

XSL•FO
RenderX

employed structural equation modeling to explore doctors' willingness to use AI-CDSSs in developing countries, revealing that concerns about potential threats to professional autonomy could dampen doctors' willingness to embrace these systems. Conversely, active involvement in the planning, design, and implementation of AI-CDSSs was associated with increased acceptance and readiness to use these technologies. Similarly, Laka et al [22] utilized logistic regression analysis to explore the adoption willingness of AI-CDSSs, finding that doctors in local primary care facilities, as opposed to those in larger hospitals, identified factors, such as time constraints, perceived threats to professional autonomy, and considerations of patient preferences, as significant barriers to adopting AI-CDSSs.

Current research has focused mainly on examining the impact of individual factors or certain factors affecting the allocation of medical resources. Thus, the literature lacks in-depth analysis of the interactive mechanisms and synergistic effects of factors influencing doctors' intentions to adopt AI-CDSSs. There is also a lack of comprehensive analysis that combines the multiple factors influencing doctors' intentions to adopt AI-CDSSs for assessing their causal relationships. To explore the driving mechanisms of multiple conditional linkages on doctors' adoption willingness, our study focuses on doctors at different medical institutions in China, not only providing important assistance for the application of AI-CDSSs in hospitals at all levels in China, but also offering experiential references for other developing countries in the use and adoption of AI-CDSSs. Using fuzzy set qualitative comparative analysis (fsQCA), we clarify the synergistic effects of multiple factors influencing doctors' intentions to adopt AI-CDSSs, thereby providing theoretical support for promoting their application in the medical field.

Framework for AI-CDSS Adoption Willingness

The integration of the UTAUT was proposed in 2003 by Venkatesh et al [23] to explain the relevant factors influencing an individual's willingness to accept or use new technology. The UTAUT consists of 4 key factors: performance expectancy, effort expectancy, social influence, and facilitating conditions, which influence behavior through willingness [24]. The UTAUT also considers the moderating effects of sex, age, experience, and voluntary use. The theory was established in the context of the organizational implementation of new technology, with the influencing factors having clear utilitarian characteristics. With the emergence of AI technology, an increasing number of scholars are using the UTAUT to study individual AI technology adoption issues. The technology-organization-environment (TOE) framework proposed by Tornatzky et al [25] suggests that technological, organizational, and external environmental factors also have a certain impact on an organization's adoption and implementation of new technology.

There is an interactive relationship between medical institutions and doctors. An organization's attitude toward innovation will affect employees' acceptance of new technology [26]. In addition, the adoption willingness and behavior of AI technology in an organization may be affected by the adoption willingness of employees [27]. Therefore, based on the UTAUT model and TOE framework, this study constructed a multi-layer dynamic impact model for AI technology adoption among physicians.

Technical Factors Influencing the Adoption of AI-CDSSs by Doctors

Performance expectancy is one of the key constructs in the UTAUT model used to explain and predict individual technology acceptance behavior [28]. In a hospital setting, it can capture the extent that doctors believe that using new technology will help improve their job performance [29]. Previous studies have shown that performance expectancy is crucial for doctors' adoption and acceptance of AI-CDSSs [30], similar to the perceived usefulness in the technology acceptance model (TAM). Compared with other technology application scenarios, doctors place more emphasis on the impact of technology on their job performance when adopting new technology. Currently, AI technology has the ability to assist in eliminating redundant work steps, providing decision support, and improving job performance [31,32]. However, there are also issues, such as communication barriers between doctors and AI technology, which can impact work efficiency. The effectiveness of AI technology in the workplace is yet to be widely validated [33]. Therefore, performance expectancy still plays an important role in doctors' willingness to adopt AI technology.

Perceived risk refers to the degree of insecurity that doctors perceive when they are using technology to execute tasks and exchange data [34]. AI technology requires big data to achieve powerful learning, which means that AI technology may involve the input of data from various parties, such as individuals and vendors. When there is a risk of information leakage, doctors using AI technology may face legal, moral, and ethical issues. This can have a significant impact on their willingness to adopt AI technology.

Organizational Factors Influencing the Adoption of AI-CDSSs by Doctors

Social influence refers to the influence doctors feel from their social environment regarding a specific behavior. It is a key factor in the UTAUT model that affects an individual's willingness to adopt new technology [30]. In an organizational context, doctors are frequently influenced by colleagues and leaders, and they enhance their sense of belonging by conforming to these groups. When faced with emerging technologies, such as AI, there may not be enough information for an informed decision, making doctors more susceptible to peer influence. However, leaders also influence doctors' willingness to adopt AI technology [35] and have the power to authorize subordinates, determine job promotions, provide rewards, and administer punishments. Thus, doctors align with leaders to receive recognition.

The UTAUT construct "facilitating conditions" refers to the extent that doctors perceive that the necessary infrastructure and resources in the organization support their use of new technology [30]. Thus, this would also influence a doctor's willingness to adopt AI technology. The promotion of new technology requires organizations to provide various resources such as knowledge, funds, and technology. Simpler and more

XSL•FO
RenderX

convenient external support conditions are associated with a greater likelihood of doctors adopting AI technology. Research has shown that facilitating conditions positively influence an individual's willingness to adopt AI technology [36,37].

Individual Factors Influencing the Adoption of AI-CDSSs by Doctors

Technology anxiety refers to an individual's emotional anxiety or fear of the performance of new technology. For example, when individuals believe that the technology may threaten their sense of self, they may experience technology anxiety, which reduces their willingness to adopt it [28]. The existing and potential capabilities of AI technology to replace human abilities are constantly increasing, causing individuals to experience stronger feelings of anxiety compared with other technologies. Therefore, our study included technology anxiety as a factor in our framework.

Personal innovativeness reflects an individual's willingness to try something new. Innovation diffusion theory suggests that owing to differences in innovation capabilities, individuals' willingness and behaviors vary in this respect. Some scholars have proven that in consumer scenarios, personal innovativeness positively influences individuals' willingness to adopt self-service technologies [38]. As AI technology is a revolutionary innovation, personal innovativeness is needed to drive doctors toward a greater willingness to adopt it. Therefore, we assume that personal innovativeness significantly impacts doctors' willingness to adopt AI technology.

Study Model

Our study applies the UTAUT model for the basis of our research framework to analyze the factors influencing doctors' adoption of AI-CDSSs, adding other factors as well. Our model considers 3 perspectives: technology, organizations, and individuals. The technical factors included are performance expectancy and perceived risk, the organizational factors are social influence and facilitating conditions, and the individual factors are technology anxiety and personal innovativeness. We have incorporated performance expectancy, social influence, and facilitating conditions from the UTAUT model, and based on the research by Chen et al [37], personal innovativeness and perceived risk have been introduced into the model and confirmed, along with key factors in the UTAUT model that significantly influence doctors' acceptance of AI technology. Huang et al [39] demonstrated how technology anxiety reflected an individual's emotional anxiety or fear regarding the performance of AI technology. When individuals perceived the technology as threatening their sense of self, they experienced technology anxiety, which reduced their willingness to adopt the technology. Therefore, we have included technology anxiety in the model. Considering that AI technology often does not require users to learn how to operate it, as it possesses anthropomorphic characteristics that are different from nonintelligent technologies [8], we did not include effort expectancy. The factors ultimately in our analysis framework are performance expectancy, perceived risk, facilitating conditions, social influence, technology anxiety, and personal innovativeness.

Thus, we assume that doctors' willingness to adopt AI technology will be influenced by these various factors. Although some studies have investigated the individual effects of these elements on doctors' AI adoption willingness, providing a foundation for our understanding of the factors influencing the willingness to adopt AI technology, research struggles to answer how these factors interact to influence the willingness to adopt AI technology under multiple situational conditions. Additionally, research has not identified the deep-rooted causal relationships that are affecting doctors' willingness to adopt AI technology. To fill this gap in the literature, based on the UTAUT model and incorporating the TOE framework, we have explored the complex causal mechanisms of how environmental, technological, and individual factors influence doctors' willingness to adopt AI technology from a configurational perspective, proposing a theoretical model (Figure 1).

XSL•FO
RenderX

Figure 1. Analysis framework of the factors influencing Chinese doctors' willingness to adopt artificial intelligence–driven clinical decision support systems.

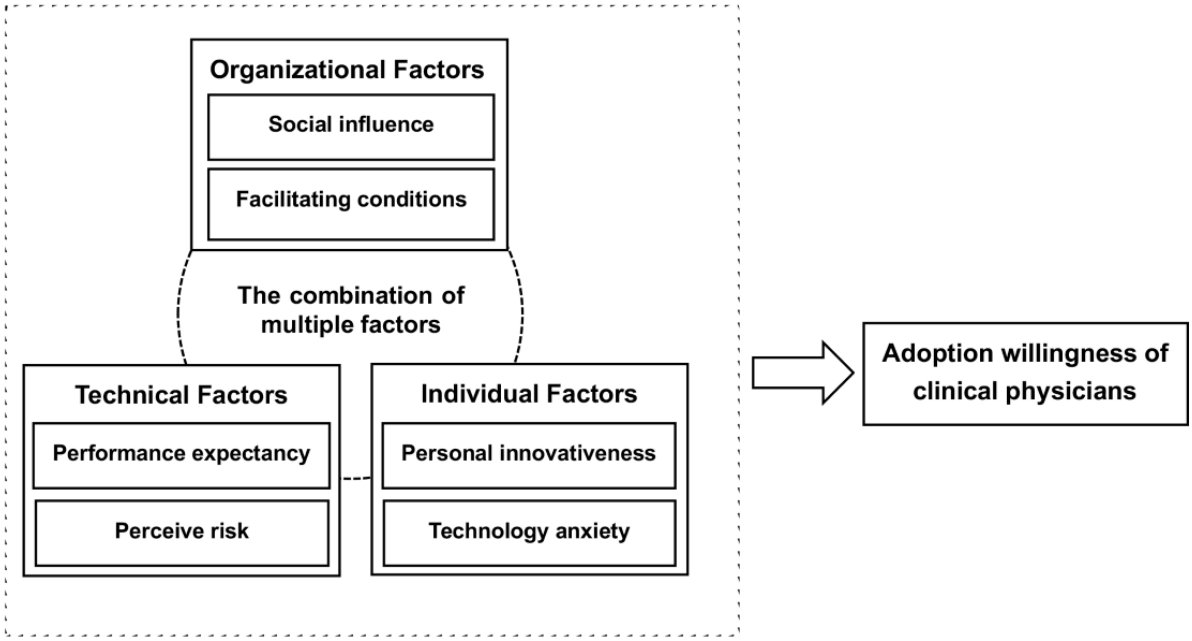

Methods

Method Selection

We used fsQCA to explore the complex causal mechanisms influencing clinical doctors' willingness to adopt AI-CDSSs, primarily for the following reasons. First, using this method uncovers the nonlinear relationships between various influencing factors and the doctors' willingness to adopt AI-CDSSs. The fsQCA method also explores combinations of influencing factors instead of individual factors [40]. Second, as our research question was "Which factors can lead clinical doctors to have a higher willingness to adopt AI-CDSSs?", the use of this method can reveal multiple equivalent paths that influence doctors' willingness to adopt AI-CDSSs. Third, compared with other qualitative comparative analysis methods, fsQCA is more suitable for handling continuous variables. Fourth, fsQCA can be applied to different sample sizes ranging from very small (<50 cases) to very large (thousands of cases) [41].

Data Collection

An online questionnaire was developed and shared with doctors in China who met specific criteria related to their work experience. We distributed 578 questionnaires through the Wenjuanxing platform [42]. After eliminating any invalid submissions, we had 450 valid responses, resulting in a questionnaire response rate of 77.9%. Detailed demographic information on the participants is presented in Table 1.

According to "Hospital Classification Management Measures" issued by the National Health Department, hospitals in China are classified into 3 levels. Tertiary hospitals provide medical and health services across regions, provinces, and cities, and nationwide. Secondary hospitals provide comprehensive medical and health services to multiple communities and undertake teaching and research tasks in regional hospitals. Primary hospitals are grassroot hospitals and health centers that provide preventive, medical, health, and rehabilitation services to their local communities [43]. In this study, we divided hospitals into only 2 categories for simplicity: tertiary hospitals and primary/secondary hospitals. Among the respondents, there were 332 responses from clinical doctors working in tertiary medical institutions and 118 responses from clinical doctors in primary/secondary medical institutions, as mentioned above.

XSL•FO
RenderX

Table 1. Basic respondent demographic statistics.

Category	Tertiary hospitals (N=332), n (%)	Primary/secondary hospitals (N=118), n (%)
Sex		
Male	171 (51.5)	53 (44.9)
Female	161 (48.5)	65 (55.1)
Age (years)		
Under 25	27 (8.1)	12 (10.2)
25 to 34	109 (32.8)	31 (26.3)
35 to 44	142 (42.8)	35 (29.7)
45 to 54	51 (15.4)	34 (28.8)
Above 54	3 (0.9)	6 (5.1)
Education		
Bachelor's degree	89 (26.8)	85 (72.0)
Master's degree	176 (53.0)	13 (11.0)
Doctoral degree	66 (19.9)	1 (0.9)
Others	1 (0.3)	19 (16.1)
Major title		
Resident physician	100 (30.1)	46 (39.0)
Attending physician	113 (34.0)	39 (33.1)
Associate chief physician	81 (24.4)	19 (16.1)
Chief physician	38 (11.5)	14 (11.9)
Duration of employment (years)		
1 or less	28 (8.4)	15 (12.7)
2 to 5	86 (25.9)	25 (21.2)
6 to 10	71 (21.4)	17 (14.4)
11 to 15	66 (19.9)	23 (19.5)
16 to 20	41 (12.4)	5 (4.2)
21 to 25	19 (5.7)	11 (9.3)
25 or more	21 (6.3)	22 (18.6)

Variable Measurement and Calibration

To ensure the reliability and validity of our scales, we based them on mature scales developed by scholars in the field of technology adoption and appropriately adjusted them for our research question to create measurement items for our main variables [23]. The main variables were performance expectancy, perceived risk, facilitating conditions, social influence, technology anxiety, personal innovativeness, and adoption willingness. Participant responses were on a 5-point Likert scale assessing the extent that they agreed with the content described in the items. We calculated the average score of the corresponding items within each scale to measure the variable. We use SPSS (version 25.0) to analyze the reliability and validity of the scales. All variables had a Cronbach α greater than .8, composite reliability greater than 0.8, Kaiser-Meyer-Olkin (KMO) values greater than 0.7, and average variance extracted values greater than 0.5, indicating that the scales had good reliability and validity. To ensure the overall reliability of our results, we removed items with factor loadings less than 0.6, and the conditions were still met after these items were deleted (Table 2).

To meet the Boolean logic requirements for qualitative comparative analysis, variables need to be transformed into sets and cases need to be assigned to the sets before conducting fsQCA, a process known as data calibration. In this process, we need to establish calibration points for "full membership," "crossing point," and "full nonmembership." We adopted a scholar's calibration method for the Likert scale questionnaire data, coding "completely agrees (5)" as "full membership," "neutral (3)" as "crossing point," and "completely disagrees (1)" as "full nonmembership." By setting these 3 thresholds, we converted the original data into fuzzy scores ranging from 0 to 1, using the calibrate (x, n1, n2, n3) function in the fsQCA software.

XSL•FO
RenderX

Table 2. Reliability and validity analysis.

Variable and item	Factor loading	Cronbach α	KMO[a]	CR[b]	AVE[c]
Performance expectancy		.831	0.797	0.916	0.611
1	0.913				
2	0.925				
3	0.884				
4	0.543				
Perceived risk		.879	0.848	0.934	0.568
1	0.567				
2	0.674				
3	0.885				
4	0.875				
5	0.838				
Facilitating conditions		.857	0.821	0.929	0.640
1	0.775				
2	0.893				
3	0.768				
4	0.825				
Social influence		.883	0.811	0.940	0.864
1	0.842				
2	0.913				
3	0.810				
4	0.686				
Technology anxiety		.921	0.900	0.890	0.572
1	0.839				
2	0.833				
3	0.859				
4	0.620				
5	0.723				
Personal innovativeness		.865	0.831	0.896	0.650
1	0.786				
2	0.824				
3	0.912				
4	0.880				
Adoption willingness		.887	0.747	0.931	0.768
1	0.961				
2	0.953				
3	0.946				

[a]KMO: Kaiser-Meyer-Olkin.

[b]CR: composite reliability.

[c]AVE: average variance extracted.

Ethical Considerations

This study was approved by the Clinical Research Ethics Committee of China-Japan Friendship Hospital (number: 2024-KY-254). All participants provided informed consent before the investigation began. Furthermore, information on

XSL•FO
RenderX

the research participants was kept confidential, and personal private information was not disclosed.

Results

Necessary Condition

According to the fsQCA method, before conducting the configuration analysis, the first step is to perform a necessity analysis on the individual condition variables, with the results reflected through consistency and coverage. Consistency represents the degree that the condition variables are a subset of the outcome variables. Identifying a necessary condition generally requires a consistency score higher than 0.9. Coverage represents the extent that the condition variables explain the outcome. This is only meaningful for conditions that pass the consistency test, with no acceptable threshold. The results of the necessity test for the individual conditions are shown in Table 3.

In Table 3, we can see that the consistency of personal innovativeness in influencing doctors' willingness to adopt AI-CDSSs at primary/secondary hospitals was higher than 0.9 and coverage was as high as 0.933. This indicates that personal innovativeness is a necessary condition influencing doctors' willingness to adopt AI-CDSSs in these hospitals. The consistency of the other variables was less than 0.9, indicating that they are not sufficient to constitute the necessary conditions affecting doctors' willingness to adopt AI-CDSSs, with these variables having relatively weak independent explanatory powers. Thus, we need to further analyze the combining effects of these condition variables and their impacts on our outcome variable.

Table 3. Necessary condition analysis.

Condition variable	Tertiary hospitals		Primary/secondary hospitals	
	Consistency	Coverage	Consistency	Coverage
Performance expectancy	0.882	0.917	0.872	0.832
~[a]Performance expectancy	0.376	0.852	0.423	0.784
Perceived risk	0.471	0.952	0.570	0.890
~Perceived risk	0.775	0.853	0.722	0.763
Facilitating conditions	0.653	0.955	0.757	0.924
~Facilitating conditions	0.611	0.849	0.567	0.739
Social influence	0.787	0.944	0.854	0.934
~Social influence	0.487	0.855	0.479	0.711
Technology anxiety	0.707	0.939	0.661	0.908
~Technology anxiety	0.555	0.854	0.658	0.766
Personal innovativeness	0.861[b]	0.949	0.930[b]	0.933
~Personal innovativeness	0.414	0.834	0.414	0.702

[a]"~" indicates the negation of the condition.

[b]Consistency exceeds 0.8.

Adequacy Analysis of Configuration

As mentioned, we conducted fsQCA separately for tertiary and primary/secondary hospitals. According to the principles of fsQCA, we included 6 condition variables. We retained 85% of the case numbers to set the frequency threshold, with case number thresholds of 5 and 2 for tertiary hospitals and primary/secondary hospitals, respectively. The consistency threshold for each configuration was higher than 0.8, and the proportional reduction in inconsistency threshold was greater than 0.75. A configuration with consistency below the threshold was assigned a value of 0 (Tables 4 and 5).

After standard analysis of the improved truth table, we found 3 types of solutions: complex, intermediate, and parsimonious. Among them, we obtained the intermediate solution through a counterfactual analysis, assuming that the emergence of personal innovativeness may increase doctors' willingness to adopt AI-CDSSs, whereas the other individual conditions may contribute to doctors' willingness to adopt AI-CDSSs. We identified the core conditions for each configuration by comparing the nested relationships between the intermediate and parsimonious solutions. The conditions appearing in both the parsimonious and intermediate solutions were considered core conditions for that configuration, whereas those appearing only in the intermediate solution were considered marginal conditions (Table 6).

In Table 6, we can see that there are 3 pathways leading to positive doctors' willingness to adopt AI-CDSSs in tertiary hospitals, which have been presented below.

XSL•FO
RenderX

Table 4. Truth table for doctors at tertiary hospitals.

Conditional variable						Number	Outcome	Raw consistency	PRI[a] consistency	SYM[b] consistency
A	B	C	D	E	F		Y			
1	1	1	1	1	1	32	1	0.997789	0.994404	0.994404
1	1	0	1	1	1	7	1	0.997023	0.989236	0.989237
1	0	1	1	1	1	38	1	0.995619	0.988805	0.988805
1	1	1	1	0	1	5	1	0.996540	0.984514	0.987172
1	0	0	1	1	1	22	1	0.994604	0.984282	0.984282
1	0	1	1	0	1	44	1	0.992934	0.981120	0.981120
1	0	0	0	1	1	19	1	0.992382	0.976033	0.978558
1	0	0	1	0	1	10	1	0.990879	0.962754	0.963750
1	0	0	0	1	0	8	1	0.988872	0.942290	0.942291
0	0	1	1	0	1	5	1	0.988860	0.929057	0.932779
1	0	0	0	0	1	13	1	0.982544	0.921498	0.931327
1	0	0	1	0	0	8	1	0.984218	0.907492	0.907493
0	0	0	0	1	1	8	1	0.981872	0.904679	0.909656
0	0	0	1	0	1	6	1	0.985961	0.902044	0.902046
1	0	1	1	0	0	10	1	0.971584	0.850015	0.866264
0	0	0	0	0	1	10	1	0.973338	0.826399	0.828025
1	0	0	0	0	0	11	1	0.964521	0.796661	0.797960
0	0	0	0	0	0	24	0	0.895842	0.463338	0.467589

[a]PRI: proportional reduction in inconsistency.

[b]SYM: symmetric.

XSL•FO
RenderX

Table 5. Truth table for doctors at primary/secondary hospitals.

Conditional variable						Number	Outcome	Raw consistency	PRI[a] consistency	SYM[b] consistency
A	B	C	D	E	F		Y			
1	0	1	1	1	1	7	1	0.995976	0.986577	0.986577
1	1	1	1	1	1	15	1	0.991665	0.977517	0.977518
1	0	0	1	1	1	2	1	0.992201	0.963961	0.963961
1	1	0	1	1	1	3	1	0.988830	0.953054	0.953054
1	1	0	1	0	1	3	1	0.992045	0.952049	0.952050
1	0	1	1	0	1	17	1	0.982767	0.947533	0.953350
0	0	1	1	0	1	2	1	0.989427	0.927492	0.927492
1	1	1	1	0	1	4	1	0.984086	0.922251	0.948591
1	0	0	1	0	1	3	1	0.986501	0.920241	0.920242
1	0	1	0	0	1	3	1	0.975536	0.830054	0.830053
1	0	0	0	0	1	5	1	0.969439	0.797216	0.797216
0	0	0	1	0	1	2	0	0.968835	0.732559	0.732558
0	0	0	0	0	1	6	0	0.946550	0.654310	0.654310
1	0	1	1	0	0	4	0	0.965158	0.631557	0.631558
0	0	0	0	1	0	2	0	0.912641	0.349602	0.349602
0	0	0	0	0	0	16	0	0.842963	0.315754	0.327492
1	1	0	0	0	0	3	0	0.937445	0.308522	0.308522
0	1	0	0	1	0	2	0	0.937238	0.290340	0.348758
1	0	0	0	0	0	10	0	0.861165	0.251138	0.251138

[a]PRI: proportional reduction in inconsistency.

[b]SYM: symmetric.

Table 6. Adoption willingness of doctors in different medical institutions.

Conditions	Tertiary hospitals[a]						Primary/secondary hospitals[b]		
	Pathway S1		Pathway S2		Pathway S3		Pathway N1	Pathway N2	
	S1a	S1b	S2a	S2b	S3a	S3b		N2a	N2b
Performance expectancy	PC[c]	PC	—[d]	—	PC	PC	PC	PC	—
Perceived risk	AM[e]	AM	AM	AM	—	—	—	AM	AM
Facilitating conditions	AM	—	AM	—	PM[f]	—	—	—	PM
Social influence	AM	PM	AM	PM	PM	PM	PM		PM
Technology anxiety	—	AM	—	AM	—	PM	—	AM	AM
Personal innovativeness	—	—	PC	PC	PC	PC	PC	PC	PC
Consistency	0.963	0.970	0.964	0.985	0.992	0.994	0.972	0.961	0.983
Raw coverage	0.416	0.451	0.413	0.444	0.591	0.549	0.769	0.514	0.479
Unique coverage	0.018	0.015	0.017	0.011	0.022	0.040	0.279	0.023	0.017

[a]The solution consistency value was 0.804 and solution coverage value was 0.953.

[b]The solution consistency value was 0.810 and solution coverage value was 0.961.

[c]PC: presence of a core causal condition.

[d]The condition can or cannot exist in the configuration.

[e]AM: absence of a marginal causal condition.

[f]PM: presence of a marginal causal condition.

XSL•FO
RenderX

Configuration for Doctors at Tertiary Hospitals

Technology Driven

The core condition for both pathways S1a and S1b was performance expectancy, which played a dominant role in the pathways. The findings indicate that these doctors believe that AI-CDSSs are helpful in clinical work, can improve work efficiency, and can enhance work quality. Pathway S1a indicated that under high performance expectancy, even without perceived technical risks for AI-CDSSs and without organizational factors, such as convenience of AI use and social influence, doctors have favorable adoption willingness for AI-CDSSs. Pathway S1b indicated that under high performance expectancy, without technology anxiety and perceived risk but with social influence, such as influence from surrounding groups, doctors still have favorable adoption willingness for AI-CDSSs.

Individual Driven

The core condition for pathways S2a and S2b was personal innovativeness, which played a primary role in these pathways. The findings indicate that these doctors are willing to try new AI technology, as they typically favor innovativeness that enables continuous learning of new medical technologies and treatment methods. Pathway S2a showed that when doctors reflect strong personal innovativeness, without perceived risks, convenience factors, and social influence, they have positive adoption willingness for AI-CDSSs. Pathway S2b indicated that when doctors have personal innovativeness without perceived risks and technology anxiety for AI-CDSSs, but with a certain degree of social influence, they still tend to have positive adoption willingness for AI-CDSSs.

Technology-Individual Dual Driven

The core conditions for pathways S3a and S3b were performance expectancy and personal innovativeness, indicating that doctors with both high performance expectancy and high personal innovativeness develop strong adoption willingness. Pathway S3a demonstrated that doctors with high performance expectancy and high personal innovativeness need support from certain convenience factors as well as a certain degree of social influence to engender strong adoption willingness for AI-CDSSs. Pathway S3b showed that when doctors have high performance expectancy, high personal innovativeness, and some degree of technology anxiety and social influence, they still tend to have high adoption willingness for AI-CDSSs.

Configuration for Doctors at Primary and Secondary Hospitals

Technology-Individual Dual Driven

The core conditions for pathways N1 and N2a were performance expectancy and personal innovativeness, indicating that doctors with both high performance expectancy and high personal innovativeness have strong adoption willingness for AI-CDSSs. Pathway N1 showed that doctors with high performance expectancy and high personal innovativeness have strong adoption willingness when influenced by leaders, colleagues, and other people regarding the use of AI. Pathway N2a demonstrated that doctors with high performance expectancy, high personal innovativeness, and no perceived risks or technology anxiety for AI technology have strong adoption willingness.

Organization-Individual Dual Driven

The core conditions for pathway N2b were convenience factors and personal innovativeness, indicating that doctors who receive support from their workplace and from technology and have high personal innovativeness develop strong adoption willingness. In other words, with high convenience factors, personal innovativeness, and low social influence, even without perceived risks and technology anxiety, doctors have strong adoption willingness for AI-CDSSs.

Comparative Analysis

Comparing doctors at tertiary hospitals with those at primary/secondary hospitals, we can point out the similarities and differences. The pathways reflect what drives the doctors at these different medical institutions to stronger adoption willingness for AI-CDSSs.

Similarities

AI technology and personal factors play dominant roles in influencing the adoption of AI-CDSSs by doctors at all the hospitals analyzed. There are not only single-dimensional factors that affect doctors' adoption willingness, but also combinations of factors, as in the technology-individual dual model, that have an impact on adoption. At the technological level, doctors believed that the application of AI-CDSSs in clinical diagnosis and treatment processes can provide efficient diagnostic support and improve the quality of clinical services. Although issues, such as overdiagnosis, may exist with AI-CDSSs, doctors considered the overall technology of AI-CDSSs to be safe and reliable. At the individual level, doctors demonstrated strong acceptance and openness to AI technology, showing no anxiety regarding the emergence of new technologies. In particular, they appeared willing to try new AI technologies.

Differences

Convenience factors had a greater impact on the adoption willingness of doctors at primary/secondary hospitals than on the adoption willingness of doctors at tertiary hospitals. According to pathways S1a and S2a, a lack of convenience factors did not affect the strong adoption willingness of doctors at tertiary hospitals. This finding indicates that even when the marginal condition of the convenience factor is missing, these doctors still have a positive adoption willingness for AI-CDSSs. In contrast, looking at pathway N2b, convenience factors are a necessary condition for doctors at primary/secondary hospitals, and these doctors will only favor adoption of AI-CDSSs when convenience factors are present. Thus, convenience factors are the objective material factors influencing the willingness of these doctors to adopt AI-CDSSs. As tertiary hospitals are generally regional hospitals that have comprehensive hospital facilities and advanced information systems, such convenience factors are present; thus, they do not significantly influence the adoption of AI-CDSSs among doctors at these facilities. Unlike tertiary hospitals, primary/secondary hospitals are often county hospitals or primary health care institutions, with some located in remote rural areas and having fewer hardware and software resources to support their doctors. Therefore, the impact of

XSL•FO
RenderX

convenience factors on the adoption willingness of doctors at these medical institutions is greater. The survey results imply that only when there is sufficient external support for doctors at these primary and secondary institutions will they actively adopt AI-CDSSs.

Robustness Test

To test the robustness of our results, we adjusted the consistency threshold from 0.8 to 0.85 and 0.72 [44]. There were no substantial changes observed in the configuration of the pathways or parameters. The results indicated that the adjusted structure remained consistent with the original structure and the pathways were the same as those before the adjustment. Therefore, the results remained robust.

Discussion

Principal Findings

We constructed a theoretical framework based on the UTAUT and TOE framework using configurational thinking and fsQCA to configure 6 conditional elements. We explored the multiple concurrent factors and causal complex mechanisms that influence the willingness to adopt AI-CDSSs among clinical doctors at different medical institutions in China from technological, organizational, and individual perspectives. Our results provide a theoretical basis for the further integration of AI-CDSSs into clinical applications. The following are our key results.

We found that the paths driving high AI-CDSS adoption willingness among clinical doctors in tertiary hospitals fell into 6 categories, which were summarized in 3 configurations: technology driven, individual driven, and technology-individual dual driven. The paths driving high AI-CDSS adoption willingness among clinical doctors at primary/secondary medical institutions fell into 3 categories, which were summarized in 2 configurations: technology-individual and organization-individual dual driven.

Comparing tertiary hospitals with primary/secondary medical institutions, we observed some commonalities and some differences in the paths driving doctors to a positive willingness to adopt AI-CDSSs. In terms of commonalities, AI technology and individual factors play dominant roles in doctors' adoption willingness. The doctors indicated their beliefs that AI-CDSSs can provide efficient diagnostic support and improve the quality of medical services. Moreover, they indicated that they are willing to try new technologies. In terms of differences, convenience factors had a greater impact on doctors at primary/secondary medical institutions. These doctors would actively adopt AI-CDSSs only with sufficient external support.

By studying doctors from different levels of medical institutions in China and their adoption paths of AI-CDSSs, we found that resource availability may be an important factor influencing the adoption willingness of doctors at different medical institutions. For example, tertiary hospitals have greater access to resources, such as funding and technical support, compared with primary/secondary hospitals. Additionally, differences in organizational culture and management styles and values may impact doctors' attitudes toward AI-CDSSs. The characteristics of patient populations served by different levels of medical institutions may also influence doctors' perceptions of AI-CDSSs, as patient needs and complexities can vary across hospital settings.

Finally, our research results can serve as a reference for other developing countries for promoting the application of AI-CDSSs or AI technologies in clinical treatment, contributing to enhancing the medical service capabilities of medical institutions.

Theoretical Contributions

Our study contributes to the literature in several ways. First, we built a comprehensive analytical framework for doctors' willingness to adopt AI technology. This is based on the UTAUT and TOE framework combined with specific characteristics and application scenarios for AI technology, with a focus on technical, organizational, and individual factors. Previous studies have focused mainly on the causal relationships between individual variables and the willingness to adopt AI technology. However, our study introduces multiple factors that can influence doctors' willingness to adopt AI technology, namely, performance expectancy, perceived risks, convenience factors, social influence, technology anxiety, and personal innovativeness. Building on the UTAUT model, we incorporated additional factors, specifically personal innovativeness, technology anxiety, and perceived risks, into the analytical framework. As such, we were able to enrich the theoretical research regarding the factors influencing doctors' willingness to adopt AI technology.

Second, we explored the synergistic effects influencing doctors' willingness to adopt AI-CDSSs from a configuration perspective, expanding the application of the UTAUT model in explaining causal complexity. Although the UTAUT model is widely used to explain individual adoption of new technologies within organizations, existing studies on AI adoption have largely overlooked the complexity of causal relationships. Owing to limitations in research methods, existing technology adoption models have been unable to test and explain the impact of multiple conditions on doctors' willingness to adopt AI-CDSSs. In our study, we empirically investigated the synergistic effects of 6 specific factors related to technology, organization, and individual aspects on doctors' willingness to adopt AI-CDSSs from a configuration perspective. By addressing the aforementioned issues, we expand the application of the UTAUT model in explaining causal complexity.

Finally, we used the fsQCA method to analyze the configurations of doctors' willingness to adopt AI-CDSSs at different medical institutions. Our results indicate that across different medical institutions, performance expectancy and personal innovativeness are the 2 important conditions for doctors to engender strong adoption willingness for AI-CDSSs, whereas perceived risks hinder adoption. Social influence can either promote or hinder doctors' willingness to adopt AI-CDSSs, and convenience factors have a greater impact on doctors' adoption willingness at primary/secondary medical institutions. In summary, our research extends the literature on doctors' willingness to adopt AI-CDSSs and provides theoretical support for future practical applications.

XSL•FO
RenderX

Practical Implications

From the perspective of the 6 configurations among the doctors at tertiary hospitals and the 3 configurations among the doctors at primary/secondary medical institutions, performance expectancy and personal innovativeness were the 2 indispensable and core conditions in the pathways to achieving strong willingness to adopt AI-CDSSs. Thus, AI product providers and health care managers should look closely at these factors when designing and implementing such systems. The organizational factor of facilitating conditions for doctors at primary/secondary medical institutions also appeared to be a necessary condition influencing the adoption willingness of doctors at these institutions. As such, we recommend the following measures for the AI-CDSS process.

Importance of Performance Expectancy in Adoption

For AI product providers, improvements in the quality and applicability of AI products can be achieved by focusing on the designing of AI-CDSSs that meet the specific needs of different clinical doctors. This means involving these doctors in the design and development process of AI-CDSSs to ensure that these systems address practical needs. In addition, to ensure that AI-CDSSs comply with data privacy and security standards, security measures should be included to increase doctors' trust in the systems. The systems should also be evaluated regularly for effectiveness and impact, and should be continuously improved based on feedback to ensure they are meeting user needs and expectations.

Importance of Personal Innovativeness

Health care institution managers should provide the appropriate training and support for doctors before introducing AI-CDSSs. This can be done through various activities, such as videos and practical exercises, among others. Such training can help doctors become comfortable with using AI-CDSSs and increase their effectiveness, boosting doctors' confidence in the systems and their willingness to adopt the systems. In addition, doctors should be actively encouraged to be innovative, cultivate innovation awareness, and improve their acceptance of new technologies.

Importance of Addressing Technology Anxiety

Individual doctors need to actively participate in training to understand the basic functions and uses of AI-CDSSs. Some doctors may be skeptical of AI-CDSSs, fearing that these systems will replace their work or reduce work quality. Appropriate training can help change this negative mindset, encouraging doctors to recognize that these systems are meant to assist and enhance the efficiency and accuracy of clinical work, and not to replace doctors.

Limitations

This study has some limitations that can serve as starting points for future research. First, our research scope was limited to medical institutions in mainland China, thereby lacking comparisons with doctors in foreign medical institutions regarding the adoption paths of AI-CDSSs. Second, differences between medical institutions at different levels, such as funding, staffing, and equipment configuration, may influence doctors' perceptions, acceptance, and experience with AI-CDSSs, leading to variations in questionnaire responses. In addition, we used fsQCA to explore the driving mechanism underlying doctors' adoption willingness based on the interactive matching of multiple conditions, which does not verify the impact of individual variables or a few factors on adoption willingness. To address the limitations of the fsQCA method, future research could consider including the use of structural equation modeling (SEM) to design a more complex model structure by considering the relationships between latent variables and the correlations between multiple observed indicators, thus providing a more comprehensive data analysis and explanation. Finally, the sample size of the questionnaire in primary/secondary hospitals was less than that in tertiary hospitals, but the hospitals included in this study were representative. Because AI has more value in areas with weak medical resources, our future research will focus on the willingness of primary hospitals to use AI. We will refine research plans, expand the sample size of the survey, and improve the compliance of the hospitals included.

Conclusion

Our study built a comprehensive analytical framework for doctors' willingness to adopt AI-CDSSs, considering specific characteristics and application scenarios for AI technology, with a focus on technical, organizational, and individual factors. We used fsQCA to explore doctors' willingness to adopt AI-CDSSs in different types of medical institutions in China along with the factors influencing their willingness. From the perspectives of the 6 pathways of the doctors at tertiary hospitals and the 3 pathways of the doctors at primary/secondary hospitals, performance expectancy and personal innovativeness were 2 indispensable and core conditions in the pathways to achieving favorable willingness to adopt AI-CDSSs. The comparative analysis revealed both similarities and differences between doctors at tertiary hospitals and those at primary/secondary hospitals in terms of their adoption of AI-CDSSs. While technical and individual factors were found to be influential in driving adoption willingness across all hospitals, the impact of facilitating conditions differed between the different levels of medical institutions. Facilitating conditions were identified as a significant driver for adoption among doctors at primary/secondary hospitals, underscoring the importance of external support and resources in facilitating the adoption of AI-CDSSs in these settings.

In conclusion, the results of our research provide valuable insights into the factors influencing doctors' willingness to adopt AI-CDSSs in different health care settings. By addressing performance expectancy, personal innovativeness, and organizational support, health care organizations can promote a more favorable environment for the implementation and utilization of AI technologies, ultimately enhancing clinical decision-making systems and improving patient care outcomes. Continued research and implementation of these strategies can further advance the integration of AI-CDSSs in health care and pave the way for the widespread application of AI technology in clinical practice.

By exploring the positive practice and policy promotion of AI-CDSS application in China, our research can provide a

XSL•FO
RenderX

positive reference for the governments of developing countries with similar conditions and uneven distribution of medical resources. At the technical level, the effectiveness and safety of AI technology should be ensured, so that AI can meet the needs of clinical practice. At the organizational level, medical institutions should organize technical training, so that doctors can understand and learn the combination of AI technology and clinical practice, and build a new model of human-machine collaboration. For areas with relatively weak medical resources, the government and medical institutions should increase the infrastructure construction required for AI applications and provide adequate technical support to help doctors solve technical problems.

Acknowledgments

We would like to thank the interviewees who participated in the study. This work was supported by the National Natural Science Fund for Young Scholars of China (72104255) and Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2021-I2M-1-046).

Data Availability

The raw data supporting the study results are available from the authors upon reasonable request.

Authors' Contributions

ZY, NH, and YL played significant roles in study design, recruitment, data coding, and paper writing. QZ was responsible for investigation, data collection, and editing. XH and CJ were responsible for methodology and writing (review and editing). CZ and BL were responsible for methodology, writing (review and editing), and supervision.

Conflicts of Interest

None declared.

References

1. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
2. Osheroff J, Teich J, Levick D, Saldana L, Velasco F, Sittig D, et al. Improving Outcomes with Clinical Decision Support: An Implementer's Guide, Second Edition. New York, NY: HIMSS Publishing; 2012.
3. Shankar P, Anderson N. Advances in sharing multi-sourced health data on decision support science 2016-2017. Yearb Med Inform 2018 Aug 29;27(1):16-24 [FREE Full text] [doi: 10.1055/s-0038-1641215] [Medline: 30157504]
4. China's health industry development Statistical Bulletin for 2023. National Health Commission of the People's Republic of China. 2023. URL: http://www.nhc.gov.cn/guihuaxxs/s3585u/202408/6c037610b3a54f6c8535c515844fae96.shtml [accessed 2024-11-04]
5. World health statistics 2023: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2023:119.
6. Lu Y, Melnick ER, Krumholz HM. Clinical decision support in cardiovascular medicine. BMJ 2022 May 25;377:e059818. [doi: 10.1136/bmj-2020-059818] [Medline: 35613721]
7. Ash J, Sittig D, Guappone A, Dykstra R, Richardson J, Wright A, et al. Recommended practices for computerized clinical decision support and knowledge management in community settings: a qualitative study. BMC Med Inform Decis Mak 2012;12:6 [FREE Full text] [doi: 10.1186/1472-6947-12-6] [Medline: 22333210]
8. Hanson L, Zimmerman S, Song M, Lin F, Rosemond C, Carey T, et al. Effect of the goals of care intervention for advanced dementia: a randomized clinical trial. JAMA Intern Med 2017;177(1):24-31 [FREE Full text] [doi: 10.1001/jamainternmed.2016.7031] [Medline: 27893884]
9. Alsharqi M, Upton R, Mumith A, Leeson P. Artificial intelligence: a new clinical support tool for stress echocardiography. Expert Rev Med Devices 2018;15(8):513-515. [doi: 10.1080/17434440.2018.1497482] [Medline: 29992841]
10. Islam R, Sultana A, Tuhin MN, Saikat MSH, Islam MR. Clinical decision support system for diabetic patients by predicting type 2 diabetes using machine learning algorithms. J Healthc Eng 2023 May 30:6992441 [FREE Full text] [doi: 10.1155/2023/6992441] [Medline: 37287539]
11. Nair BG, Peterson GN, Newman S, Wu W, Kolios-Morris V, Schwid HA. Improving documentation of a beta-blocker quality measure through an anesthesia information management system and real-time notification of documentation errors. Jt Comm J Qual Patient Saf 2012;38(6):283-288. [doi: 10.1016/s1553-7250(12)38036-7] [Medline: 22737780]
12. Nair BG, Peterson GN, Neradilek MB, Newman SF, Huang EY, Schwid HA. Reducing wastage of inhalation anesthetics using real-time decision support to notify of excessive fresh gas flow. Anesthesiology 2013 Apr;118(4):874-884 [FREE Full text] [doi: 10.1097/ALN.0b013e3182829de0] [Medline: 23442753]

XSL•FO
RenderX

13. Freundlich RE, Barnet CS, Mathis MR, Shanks AM, Tremper KK, Kheterpal S. A randomized trial of automated electronic alerts demonstrating improved reimbursable anesthesia time documentation. J Clin Anesth 2013 Mar;25(2):110-114. [doi: 10.1016/j.jclinane.2012.06.020] [Medline: 23333782]
14. Sambasivan M, Esmaeilzadeh P, Kumar N, Nezakati H. Intention to adopt clinical decision support systems in a developing country: effect of physician's perceived professional autonomy, involvement and belief: a cross-sectional study. BMC Med Inform Decis Mak 2012 Dec 05;12:142. [doi: 10.1186/1472-6947-12-142] [Medline: 23216866]
15. Wagner G, Raymond L, Paré G. Understanding Prospective Physicians' Intention to Use Artificial Intelligence in Their Future Medical Practice: Configurational Analysis. JMIR Med Educ 2023 Mar 22;9:e45631. [doi: 10.2196/45631] [Medline: 36947121]
16. O'Leary P, Carroll N, Richardson I. The Practitioner's Perspective on Clinical Pathway Support Systems. 2015 Presented at: 2014 IEEE International Conference on Healthcare Informatics; 15-17 September 2014; Verona, Italy. [doi: 10.1109/ICHI.2014.33]
17. Park CJ, Yi PH, Siegel EL. Medical student perspectives on the impact of artificial intelligence on the practice of medicine. Curr Probl Diagn Radiol 2021;50(5):614-619. [doi: 10.1067/j.cpradiol.2020.06.011] [Medline: 32680632]
18. Scheetz J, Rothschild P, McGuinness M, Hadoux X, Soyer HP, Janda M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 2021 Mar 04;11(1):5193. [doi: 10.1038/s41598-021-84698-5] [Medline: 33664367]
19. Oh S, Kim JH, Choi SW, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 2019;21(3):e12422. [doi: 10.2196/12422] [Medline: 30907742]
20. Sit C, Srinivasan R, Amlani A, Muthuswamy K, Azam A, Monzon L, et al. Attitudes and perceptions of UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights into Imaging 2020;11:14. [doi: 10.1186/s13244-019-0830-7] [Medline: 32025951]
21. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. J Gastroenterol Hepatol 2021;36(3):581-584. [doi: 10.1111/jgh.15384] [Medline: 33709609]
22. Laka M, Milazzo A, Merlin T. Factors that impact the adoption of clinical decision support systems (CDSS) for antibiotic management. Int J Environ Res Public Health 2021 Feb 16;18(4):1901. [doi: 10.3390/ijerph18041901] [Medline: 33669353]
23. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
24. Bendary N, Al-Sahouly I. Exploring the extension of unified theory of acceptance and use of technology, UTAUT2, factors effect on perceived usefulness and ease of use on mobile commerce in Egypt. Journal of Business & Retail Management Research 2018 Jan;12(02):60-71. [doi: 10.24052/jbrmr/v12is02/eteoutoaauotufeopuaeouomcie]
25. Tornatzky LG, Klein KJ. Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. IEEE Transactions on Engineering Management 1982 Feb;EM-29(1):28-45. [doi: 10.1109/tem.1982.6447463]
26. Li YW. Enterprise IT/IS adopt decision behavior model analysis. Modern Management Science 2006(2):29-42 [FREE Full text]
27. Edwards JS, Duan Y, Robins PC. An analysis of expert systems for business decision making at different levels and in different roles. European Journal of Information Systems 2000;9:36-46. [doi: 10.1057/palgrave.ejis.3000344]
28. Menon D, Shilpa K. "Chatting with ChatGPT": Analyzing the factors influencing users' intention to Use the Open AI's ChatGPT using the UTAUT model. Heliyon 2023 Oct 18;9(11):e20962 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e20962] [Medline: 37928033]
29. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157-178. [doi: 10.2307/41410412]
30. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello CP, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111. [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
31. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons 2018 Nov;62(1):15-25. [doi: 10.1016/j.bushor.2018.08.004]
32. Xu P, Xu XY. The logic and analytical framework of enterprise management change in the era of artificial intelligence. Management World 2020;36:122-129+238 [FREE Full text]
33. Qiu Y, He Q. Research progress on the impact of artificial intelligence on employment and theoretical analysis framework in the Chinese context. Human Resources Development of China 2020(2):90-103. [doi: 10.16471/j.cnki.11-2822/c.2020.2.007]
34. Pillai R, Sivathanu B. Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. Benchmarking An International Journal 2020 Aug 14;27(9):2599-2629. [doi: 10.1108/bij-04-2020-0186]
35. Chatterjee S, Bhattacharjee KK. Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling. Educ Inf Technol 2020 Mar 31;25(5):3443-3463. [doi: 10.1007/s10639-020-10159-7]
36. Sohn K, Kwon O. Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. Telematics and Informatics 2020 Apr;47:101324. [doi: 10.1016/j.tele.2019.101324]
37. Chen LSL, Wu KIF. Antecedents of intention to use CUSS system: moderating effects of self-efficacy. Serv Bus 2014;8(4):615-634. [doi: 10.1007/s11628-013-0210-1]

XSL•FO
RenderX

38. Chen J, Li R, Gan M, Fu ZY, Yuan FT. Public acceptance of driverless buses in china: an empirical analysis based on an extended UTAUT model. Discrete Dynamics in Nature and Society 2020;2020:1-13. [doi: 10.1155/2020/4318182]
39. Tsai TH, Lin WY, Chang YS, Chang PC, Lee MY. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020 Jan 13;15(1):e0227270. [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
40. Vis B, Dul J. Analyzing Relationships of Necessity Not Just in Kind But Also in Degree: Complementing fsQCA With NCA. Sociol Methods Res 2018 Nov;47(4):872-899. [doi: 10.1177/0049124115626179] [Medline: 30443090]
41. Pappas IO, Woodside AG. Fuzzy-set qualitative comparative analysis (fsQCA): guidelines for research practice in information systems and marketing. International Journal of Information Management 2021 Jun;58(3):102310. [doi: 10.1016/j.ijinfomgt.2021.102310]
42. Wenjuanxing platform. URL: https://www.wjx.cn/ [accessed 2024-12-16]
43. Du C. An empirical analysis of hierarchical hospital management system in China. Social Science Front 2018(04):84-94.
44. White L, Lockett A, Currie G, Hayton J. Hybrid Context, Management Practices and Organizational Performance: A Configurational Approach. Journal of Management Studies 2020;58(3):718-748. [doi: 10.1111/joms.12609]

Abbreviations

AI: artificial intelligence
AI-CDSS: artificial intelligence–driven clinical decision support system
CDSS: clinical decision support system
fsQCA: fuzzy set qualitative comparative analysis
TOE: technology-organization-environment
UTAUT: unified theory of acceptance and use of technology

Edited by G Tsafnat; submitted 31.05.24; peer-reviewed by J Walsh, M Zahmatkeshan, M Popovic; comments to author 12.10.24; revised version received 04.11.24; accepted 25.11.24; published 07.01.25.

Please cite as:
Yu Z, Hu N, Zhao Q, Hu X, Jia C, Zhang C, Liu B, Li Y
The Willingness of Doctors to Adopt Artificial Intelligence–Driven Clinical Decision Support Systems at Different Hospitals in China: Fuzzy Set Qualitative Comparative Analysis of Survey Data
J Med Internet Res 2025;27:e62768
URL: https://www.jmir.org/2025/1/e62768
doi:10.2196/62768
PMID:

©Zhongguang Yu, Ning Hu, Qiuyi Zhao, Xiang Hu, Cunbo Jia, Chunyu Zhang, Bing Liu, Yanping Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health and Experiences During the COVID-19 Pandemic Among Children and Young People: Analysis of Free-Text Responses From the Children and Young People With Long COVID Study

Natalia K Rojas[1], BSc; Sam Martin[1], PhD; Mario Cortina-Borja[2], PhD; Roz Shafran[2], PhD; Lana Fox-Smith[2], BSc; Terence Stephenson[2], DM; Brian C F Ching[3], BSc, MSc; Anaïs d'Oelsnitz[2], BA; Tom Norris[1], PhD; Yue Xu[4], MSci; Kelsey McOwat[5], MSc; Emma Dalrymple[2], MSc; Isobel Heyman[2,6], PhD; Tamsin Ford[6], PhD; Trudie Chalder[7], PhD; Ruth Simmons[5], PhD; CLoCk Consortium[8]; Snehal M Pinto Pereira[1], PhD

[1]Division of Surgery & Interventional Science, Faculty of Medical Sciences, University College London, London, United Kingdom

[2]UCL Great Ormond Street Institute of Child Health, London, United Kingdom

[3]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[4]Division of Psychology and Language Sciences, Faculty of Brain Science, University College London, London, United Kingdom

[5]Immunisation Department, UK Health Security Agency, London, United Kingdom

[6]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

[7]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[8]See Authors' Contributions

Corresponding Author:
Natalia K Rojas, BSc
Division of Surgery & Interventional Science
Faculty of Medical Sciences
University College London
43-45 Foley St, W1W 7TY
London
United Kingdom
Phone: 44 (0) 20 7679 200
Email: n.rojas@ucl.ac.uk

Abstract

Background: The literature is equivocal as to whether the predicted negative mental health impact of the COVID-19 pandemic came to fruition. Some quantitative studies report increased emotional problems and depression; others report improved mental health and well-being. Qualitative explorations reveal heterogeneity, with themes ranging from feelings of loss to growth and development.

Objective: This study aims to analyze free-text responses from children and young people participating in the Children and Young People With Long COVID study to get a clearer understanding of how young people were feeling during the pandemic.

Methods: A total of 8224 free-text responses from children and young people were analyzed using InfraNodus, an artificial intelligence–powered text network analysis tool, to determine the most prevalent topics. A random subsample of 411 (5%) of the 8224 responses underwent a manual sentiment analysis; this was reweighted to represent the general population of children and young people in England.

Results: Experiences fell into 6 main overlapping topical clusters: *school*, *examination stress*, *mental health*, *emotional impact of the pandemic*, *social and family support*, and *physical health (including COVID-19 symptoms)*. Sentiment analysis showed that statements were largely negative (314/411, 76.4%), with a small proportion being positive (57/411, 13.9%). Those reporting negative sentiment were mostly female (227/314, 72.3%), while those reporting positive sentiment were mostly older (170/314, 54.1%). There were significant observed associations between sentiment and COVID-19 status as well as sex (P=.001 and P<.001, respectively) such that the majority of the responses, regardless of COVID-19 status or sex, were negative; for example, 84.1% (227/270) of the responses from female individuals and 61.7% (87/141) of those from male individuals were negative. There were no observed associations between sentiment and all other examined demographics. The results were broadly similar when

XSL•FO
RenderX

reweighted to the general population of children and young people in England: 78.52% (negative), 13.23% (positive), and 8.24% (neutral).

Conclusions: We used InfraNodus to analyze free-text responses from a large sample of children and young people. The majority of responses (314/411, 76.4%) were negative, and many of the children and young people reported experiencing distress across a range of domains related to school, social situations, and mental health. Our findings add to the literature, highlighting the importance of specific considerations for children and young people when responding to national emergencies.

(J Med Internet Res 2025;27:e63634) doi:10.2196/63634

KEYWORDS

children and young people; text mining; free-text responses; experiences; COVID-19; long COVID; InfraNodus; sentiment analysis; discourse analysis; AI; artificial intelligence

Introduction

Background

The emergence and rapid spread of SARS-CoV-2 thoroughly disrupted day-to-day life globally, whether through the direct effects of infection or the implementation of preventive control measures [1]. Adolescence is a sensitive period when friendships and social interactions are important for mental health, brain development, and self-concept construction [2]. Thus, the COVID-19 pandemic was a fertile ground for the development and exacerbation of concerns related to children and young people and their mental health, with experts anticipating that it would lead to a higher prevalence of mental health conditions and increased demand on services [3].

In England, COVID-19 preventive measures comprised school closures (with a transition to web-based learning), lockdowns, and restrictions on leisure activities. These restrictions encompassed limits on time spent outdoors, as well as regulations on whom individuals could interact with and how they did so (ie, social distancing). The implementation of such measures was feared to be detrimental to the mental health of children and young people [4]; for example, the effects of school closures and social distancing, including their resultant loneliness, can be long-lasting, with evidence from previous pandemics suggesting that associations with poorer mental health can persist for nearly a decade [5]. Indeed, pandemics can exert a broad negative impact on children and young people, causing stress and feelings of helplessness, as well as behavioral, sleep, and eating problems, with higher rates of depression and anxiety likely during and after a pandemic [6].

Regarding the COVID-19 pandemic specifically, the literature is equivocal as to whether the predicted deterioration in mental health came to fruition. Some quantitative studies report more children and young people reaching thresholds for mental health conditions [7] and increased emotional problems and depression [4,8]. However, other studies report improved mental health and well-being, including sleep hygiene [9] and fewer externalizing and internalizing problems [10-12]. Therefore, it is unsurprising that a systematic review of 51 studies revealed heterogeneity regarding the impact of COVID-19 on the mental health of children and young people [12]. Similarly, qualitative explorations of the experiences of children and young people during the pandemic revealed heterogeneity, with prominent themes ranging from feelings of loss to growth and development [13]; for example, while most children and young people adapted well to school closures and web-based learning, experiences ranged from a lack of motivation to complete work to improved productivity [14]. Equally, while some children and young people from Black and ethnic minority backgrounds reported a negative mental health impact, others reported improved well-being and better coping strategies after lockdown [15]. More broadly, many children and young people aged 13 to 24 years across the United Kingdom reported worse mental health, school-related concerns, and socializing concerns, while a minority viewed pandemic experiences and the resultant changes as helpful (eg, strengthened familial relationships) [16]. Reasons for the observed heterogeneity are multifactorial and could include a research focus on specific subpopulations of children and young people (eg, those with special educational needs [17], specific ethnic groups [15], or groups considered vulnerable and high risk [7,11]) rather than the general population of children and young people; alternatively, experiences of children and young people could vary by pandemic stage [18]. Indeed, a lack of generalizability was a limitation highlighted in previous studies with small samples (eg, N=168 [4] and N=37 [16]) recruited from specific regions [4,8-11,14,15,19]. Similarly, qualitative research draws on the experiences and beliefs of small, selective samples and can rarely be considered to provide generalizable conclusions (although generalizability is not the primary aim of such research) [20].

Objectives

To overcome this limitation, we analyzed 8224 self-reported free-text responses from children and young people. Specifically, we used free-text responses from the Children and Young People With Long COVID (CLoCk) study, which recruited >30,000 children and young people aged 11 to 17 years when they underwent polymerase chain reaction (PCR) testing for SARS-CoV-2 between September 2020 and March 2021 [21]. At recruitment, children and young people were asked “to tell us about [their] health or how the pandemic or lockdown affected [them].” We aimed to analyze these responses to determine the most important topics and the main “sentiment” to understand the impact of the pandemic. To ensure that our findings were representative of the broader population of children and young people in England, we applied survey weights [22] to the sentiment analysis.

XSL•FO
RenderX

Methods

Study Design and Setting

In this study, we focus on the free-text responses of children and young people (aged 11 to 17 y at index PCR test) collected at study enrollment, which was 3, 6, or 12 months after their index PCR test (ie, April 2021-December 2021). During this period, most pandemic-related restrictions were gradually lifted (April-July 2021), until the beginning of December 2021 when measures were reimplemented to slow the spread of the SARS-CoV-2 Omicron variant [23].

The UK Health Security Agency's Second Generation Surveillance System, which is a dataset containing results of all SARS-CoV-2 PCR tests conducted by hospital and public health laboratories as part of national mandatory testing, was used to identify 219,175 children and young people (n=91,014, 41.53% tested positive, while n=128,161, 58.47% tested negative) who underwent PCR testing between September 2020 and March 2021 [21,24]. Children and young people who tested positive were matched, at CLoCk study invitation, to those who tested negative based on age, sex, region of residence, and month of test. Index of Multiple Deprivation (a proxy for socioeconomic status, grouped into quintiles) was determined via the lower super output area (ie, small local area level–based geographic hierarchy) in which the children and young people resided [25]. Consenting children and young people completed web-based questionnaires 3, 6, 12, and 24 months after their index PCR test, although the sweeps of data collection varied by month of test [21,24]; for example, children and young people who tested in January to March 2021 were able to provide data 3, 6, 12, and 24 months after the test compared to those who tested in October to December 2020 who were only able to do this at 6, 12, and 24 months. The questionnaire included demographics, elements of the International Severe Acute Respiratory and Emerging Infection Consortium pediatric COVID-19 questionnaire, and several validated scales (eg, the Strengths and Difficulties Questionnaire [26]). The questionnaire ended with a final, optional, free-text response question prompting children and young people to "Please use this space if there is anything else you would like to tell us about your health or how the pandemic or lockdown have affected you." This was the final question before survey submission. This study focused solely on the free-text responses provided at study enrollment.

Participants and Sample Size

In total, 31,012 children and young people enrolled into the CLoCk study [21] (ie, 31,012 children and young people submitted their enrollment questionnaire). Of these 31,012 respondents, 10,580 (34.12%) completed the free-text response question (Figure 1). The web-based questionnaire platform truncated free-text responses at 621 characters; to ensure that the experiences of the children and young people were not misrepresented, we excluded from analysis all responses exceeding 619 characters (370/10,580, 3.5%). Thus, 10,210 (96.5%) of the 10,580 responses were subject to data cleaning and relevancy assessment as described in the next subsection.

XSL•FO
RenderX

Figure 1. Participant selection flowchart. CLoCk: Children and Young People With Long COVID.

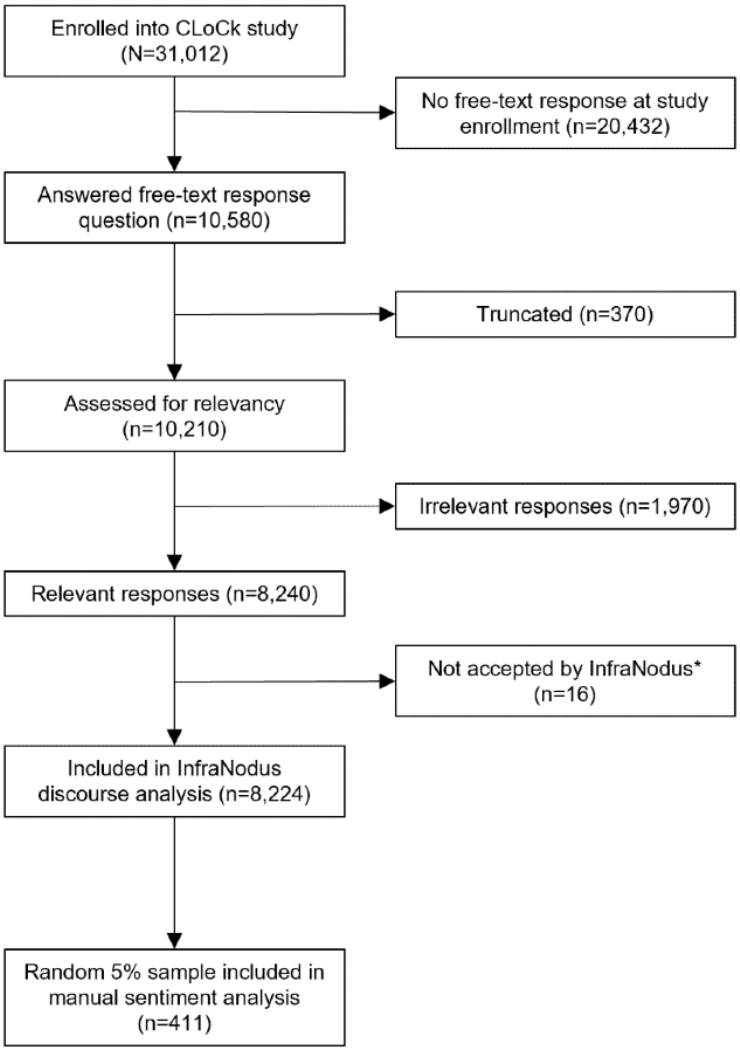

Data Cleaning and Relevancy Assessment

Free-text responses were fully anonymized—any potentially identifiable information (eg, child's or young person's name, email addresses, and hospital names) was redacted; for example, an anonymized response was as follows:

> *[T]he only thing that affected me is the move from [city name] to [city name] but regarding covid my mental health is ok although during the lockdown it was often low.*

All 10,210 responses were assessed for relevancy (refer to Textbox 1; of note, statements regarding COVID-19 vaccination, its immediate symptoms, timing with regard to SARS-CoV-2 infection or both were excluded because we are interested in the effects of SARS-CoV-2 itself and the associated COVID-19 preventive measures, eg, lockdowns, rather than vaccination side effects; moreover, the exclusion criteria are not mutually exclusive: a response may have been excluded for multiple reasons). One author (NKR) screened all responses. Any uncertainties were discussed with another author (SMPP) to reduce subjectivity and potential bias. Of the 10,210 responses assessed for relevancy, 1528 (15%) were discussed. Certain words or phrases were standardized across responses for consistency; for example, all variations of "COVID-19" (eg, "covid-19," "COVID 19," and "covid19") were revised to read "COVID." The final stage of data cleaning involved the checking and correction of spelling mistakes by 3 authors (NKR, LFS, and AD), with any uncertainties on how to rectify a spelling mistake (33/8240, 0.4%) discussed to achieve consensus.

XSL•FO
RenderX

Textbox 1. Relevancy assessment criteria and selected examples of free-text responses (included and excluded) provided by children and young people and not truncated at study enrollment (n=10,210).

Inclusion criteria (included: n=8240)

- Statements about the respondents' health related to SARS-CoV-2 infection, their health during the COVID-19 pandemic more generally or both.
- Statements about the respondents' experience of the pandemic, lockdown or both.
- Examples
  - "During the pandemic, I did lose quite a few friends and my mental health got worse"
  - "I enjoyed online learning as it gave me the opportunity to self teach something I found I am good at and I feel it is a useful skill for development. I was fitter during lockdown as I had more time to exercise and many good traits I picked up in lockdown I have kept"
  - "The first lockdown triggered phobias and anxiety. This worsened after grandad was in hospital with COVID"

Exclusion criteria (excluded: n=1970)

- Null responses (eg, "No," "N/A," or equivalent)
- Statements unrelated to the respondents' health or experiences of the pandemic or lockdown
- Nonsensical responses
- Statements revealing diagnoses before the pandemic (eg, autism)
- Statements discussing medication, health, or experiences before the pandemic with no comparison to medication, health, or experiences during the pandemic
- Statements about COVID-19 vaccination: its immediate symptoms and timing with regard to SARS-CoV-2 infection
- Responses related to whether the respondents tested positive for SARS-CoV-2 and its immediate impact (eg, on cough or smell within the first few weeks after infection)
- Statements on health guidance (eg, opinions on vaccination strategy) with no reference to their impact on the respondents personally
- Current colds, coughs, influenza, hay fever, seasonal allergies, and so on
- Correcting or clarifying previous questionnaire entries and expressing dissatisfaction with survey
- Examples
  - "TEST"
  - "are we going to go into another lockdown?"
  - "icciycih"
  - "I called [footballer] for my birthday, it was a small party. Suddenly, he arrived at the restaurant with a box of chocolates, all ashamed. He apologised for the gift and said he didn't know what to get because he had never been invited to a birthday party before"
  - "Have chronic fatigue syndrome (ME) since before COVID 19"
  - "I am allergic to aspirin"
  - "I have had glandular fever so most of these problems are and can be linked to it"
  - "I have had two Pfizer vaccinations, with an achy arm for a day or two as my only symptoms"
  - "China made covid Please tell boris that I want to go to him next party"
  - "Boris Johnson is rude"
  - "Help the NHS!! No more pandemic and vaccine effort until they have sufficient support in my opinion"

Quantitative Text Analysis

InfraNodus, an artificial intelligence (AI)–powered tool that processes and visualizes text-based data [27], was used. The tool follows these steps [28]:

1. Text normalization: words are converted into their lemmas (ie, the dictionary form) to reduce redundancy (eg, "families" becomes "family"), and syntax elements (eg, commas) are removed.
2. Stop word removal: words without significant semantic meaning (eg, "is" and "are") are filtered out. InfraNodus generates a default stop word list based on the uploaded text, which can be customized (see InfraNodus Discourse Analysis section).
3. Network graph visualization: words are displayed as a network graph that highlights the most influential words, with the option to detect topical clusters (ie, groups of words that tend to co-occur). The network graph is interactive,

XSL•FO
RenderX

allowing users to read through the text by relevant concepts (vs reading in a linear fashion).

InfraNodus Discourse Analysis

The terms "health," "COVID," "pandemic," and "lockdown" were part of the question wording. Therefore, before analysis, these words were included as stop words because they could potentially conceal other important topics. InfraNodus was then used to determine the most influential words and measure patterns occurring within the relevant 8224 free-text responses examining the experiences and health during the pandemic of children and young people. This was achieved using "betweenness centrality" scores [28]. This is a numerical measure of a word's influence over information flow and its ability to link different parts of the text network graph [28]. Words with the highest betweenness centrality scores were represented as larger nodes on the graph. Subsequently, topic modeling was undertaken. To achieve this, InfraNodus applied a community detection algorithm that identified groups of words more densely connected to each other than to the rest of the network, yielding an output of topical clusters [28].

Sentiment Analysis

Sentiment analysis measured the positive, negative, or neutral feelings reported by children and young people in the free-text responses. Initially, we set out to use InfraNodus for the sentiment analysis of all eligible responses; however, there were major discrepancies between researcher-assigned and InfraNodus-assigned sentiment (refer to Table S1 in Multimedia Appendix 1 for examples). Thus, to assess the main "sentiment" of children and young people during the COVID-19 pandemic, sentiment analysis was undertaken manually using the sentiment framework outlined in Textbox 2. A random sample was chosen for the sentiment analysis for pragmatic reasons, given the time available. This sample size represented 5% (411/8224) of the available data. The random selection aimed to mitigate potential biases inherent in other selection approaches, such as sequential sampling, which may have resulted in selection of responses predominantly from a specific time point. NKR and a second rater (LFS) met to discuss in depth the sentiment framework presented in Textbox 2, going through examples of positive, negative, and neutral statements. Subsequently, NKR's rater reliability was assessed by LFS who assigned a sentiment to a random subsample of 206 (50.1%) of the 411 responses. Agreement between the raters was 94.7% (195/206). Chi-square tests were used to examine associations between free-text sentiments and demographics.

XSL•FO
RenderX

Textbox 2. Sentiment framework for classification of free-text responses provided by children and young people at study enrollment.

Positive sentiment

- Largely communicating relatively high or high levels of well-being during the pandemic
- Largely communicating satisfaction with health guidelines and COVID-19 preventive measures (eg, school closures and lockdowns) and associated outcomes
- Largely communicating good health during the pandemic
- Examples
  - "I think the pandemic and lockdown has motivated me to be healthier and I am in overall better mental and physical health than before the pandemic"
  - "I liked being at home and doing schoolwork at home. But I missed seeing people"
  - "I loved the lie ins"

Neutral sentiment

- No positive or negative expressions
- General, factual statements
- Examples
  - "Made me be more independent- online learning and having to study the curriculum for my GCSE exams in 2021"
  - "Gained some weight"
  - "I live in an SGO I have not been able to see my mum as she lives in a hostel for over a year"

Negative sentiment

- Largely communicating poor well-being during the pandemic, including concerns about family members' health
- Largely communicating upset, frustration, or concern with health guidelines or COVID-19 preventive measures (eg, school closures and lockdowns) and associated outcomes
- Largely communicating poor health
- Examples
  - "I have not been able to see my friends and feel lonely. It scares me and makes me anxious, I have trouble concentrating"
  - "Found it very hard to stay on top of schoolwork so it has affected my mental health as I am quite stressed"
  - "I feel that my education has suffered because of lockdowns and within lockdowns I have gained problems with eating."

Reweighting: Sentiment Analysis

Sentiment analysis reweighting was conducted using Stata 17 (StataCorp LLC) [29]. Multimedia Appendix 1 provides detailed information on survey weight development and application, which is briefly summarized herein. Survey weights were developed to account for (1) not all children and young people invited to participate enrolling into the study, (2) not all children and young people enrolling providing a free text response, (3) not all free text respondents having done so relevantly (including within the character limit), and (4) not all those responding relevantly being randomly sampled for manual sentiment analysis. These 4 developed survey weights were combined to obtain a fifth survey weight that was used to reweight the analytic sample (n=411) to the target population (ie, all those invited to enroll; N=219,145). As the target population may not be representative of children and young people in England [22], we recalibrated our sentiment analysis to the general population of children and young people in England using census data from 2021 [30].

Ethical Considerations

Research ethics approval was granted by the Yorkshire and The Humber—South Yorkshire Research Ethics Committee (21/YH/0060; Integrated Research Application System project ID: 293495). Web-based informed consent was obtained from parents or carers of children and young people aged ≤15 years, with these children and young people also providing web-based assent. Those aged 16 to 17 years provided web-based informed consent, but their parents or carers did not, as per UK Health Research Authority–recommended processes [31]. Data privacy and confidentiality were maintained by redacting any potentially identifiable information in the free text (eg, child's or young person's name, email addresses, and hospital names) before analysis to achieve full anonymization. Those who completed and submitted questionnaires were compensated with a £10 (US $12.23) voucher for each completed questionnaire.

XSL•FO
RenderX

Results

Participant Demographics

Compared to those invited to participate in the CLoCk study, those included in the InfraNodus and sentiment analyses were broadly similar, with, for example, participants more likely to be female individuals (Table 1). There were some demographic differences in those excluded due to truncation and irrelevancy; for example, while more female individuals (279/370, 75.4%) and older children and young people (253/370, 68.4%) were excluded due to truncation, younger children and young people (1088/1970, 55.2%) were more likely excluded due to irrelevancy (Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Demographics of children and young people who (1) were invited, (2) enrolled, (3) answered the free-text question at enrollment, (4) were included in the InfraNodus analysis, and (5) were included in the manual sentiment analysis.

	Invited (N=219,175)	Enrolled (n=31,012)	Answered free-text question (n=10,580)	Included in InfraNodus analysis (n=8224)	Included in sentiment analysis (n=411)
COVID-19 status at study invitation					
SARS-CoV-2 positive	91,014 (41.53)	13,690 (44.14)	4743 (44.83)	3804 (46.25)	180 (43.8)
SARS-CoV-2 negative	128,161 (58.47)	17,322 (55.86)	5837 (55.17)	4420 (53.75)	231 (56.2)
Sex					
Male	103,939 (47.42)	11,961 (38.57)	3920 (37.05)	2905 (35.32)	141 (34.31)
Female	115,236 (52.58)	19,051 (61.43)	6660 (62.95)	5219 (64.68)	270 (65.69)
Age (y) at study invitation					
11-14	112,057 (51.13)	14,857 (47.91)	5184 (49)	3973 (48.31)	189 (45.99)
15-17	107,118 (48.87)	16,155 (52.09)	5396 (51)	4251 (51.69)	222 (54.01)
Ethnicity					
Asian or Asian British	—[a]	4553 (14.68)	1581 (14.94)	1143 (13.9)	60 (14.6)
Black British, African British, or Caribbean British	—	933 (3.01)	343 (3.24)	264 (3.21)	18 (4.38)
White	—	23,198 (74.8)	7838 (74.08)	6167 (74.99)	308 (74.94)
Mixed	—	1615 (5.21)	564 (5.33)	464 (5.64)	21 (5.11)
Other	—	524 (1.69)	187 (1.77)	135 (1.64)	1 (0.24)
Prefer not to say	—	189 (0.61)	67 (0.63)	51 (0.62)	3 (0.73)
Region of residence (England)					
East Midlands	14,109 (6.44)	2210 (7.13)	774 (7.32)	601 (7.31)	30 (7.3)
East of England	38,901 (17.75)	6047 (19.5)	1967 (18.59)	1543 (18.76)	63 (15.33)
London	46,300 (21.12)	6157 (19.85)	2107 (19.91)	1565 (19.03)	72 (17.52)
North East England	8613 (3.93)	1198 (3.86)	403 (3.81)	329 (4)	18 (4.38)
North West England	31,289 (14.28)	3606 (11.63)	1229 (11.62)	966 (11.75)	49 (11.92)
South East England	31,567 (14.40)	4917 (15.86)	1729 (16.34)	1347 (16.38)	80 (19.46)
South West England	8139 (3.71)	1514 (4.88)	499 (4.72)	396 (4.82)	25 (6.08)
West Midlands	22,681 (10.35)	3032 (9.78)	1046 (9.89)	822 (10)	37 (9)
Yorkshire and the Humber	17,576 (8.02)	2331 (7.52)	826 (7.81)	655 (7.96)	37 (9)
IMD[b] quintile					
1 (most deprived)	54,079 (24.67)	5345 (17.24)	1813 (17.14)	1397 (16.99)	74 (18)
2	44,757 (20.42)	5548 (17.89)	1943 (18.36)	1484 (18.04)	80 (19.46)
3	39,876 (18.19)	5792 (18.68)	1985 (18.76)	1563 (19.01)	80 (19.46)
4	39,996 (18.25)	6656 (21.46)	2220 (20.98)	1764 (21.45)	91 (22.14)
5 (least deprived)	40,467 (18.46)	7671 (24.74)	2619 (24.75)	2016 (24.51)	86 (20.92)

[a]Ethnicity data were not available for children and young people at study invitation; this was self-reported at study enrollment.

[b]IMD: Index of Multiple Deprivation.

InfraNodus Discourse Analysis

The InfraNodus discourse analysis revealed 6 main topical clusters related to *school*, *examination stress*, *mental health*, *emotional impact of the pandemic*, *COVID-19 symptoms*, and *social and family support*. Each cluster is described in detail in the following subsections (of note, the majority of statements fell into multiple clusters).

School

Children and young people reported struggling academically, with participants stating that the pandemic "affected my school

XSL•FO
RenderX

learning a lot" (PID 8126), and that absence due to restrictions caused much stress, upset, and worry due to finding it "difficult to catch up with work I have missed" (PID 7995). Concentration was commonly reported as an issue, with a respondent reporting having found "it harder to concentrate in class, harder to complete homework without being distracted and more stressful to complete tests" (PID 6851). However, there were some children and young people who enjoyed web-based classes during COVID-19 restrictions, even going as far as to say, "I liked online school" (PID 8161) and "I prefer home schooling" (PID 8196). Nevertheless, others struggled, commenting as follows: "I...hated having to do my work on a laptop" (PID 8076) and "I cannot concentrate as well as in the classroom" (PID 8051). Moreover, social skills seem to have been affected by school closures, with a participant stating that they "find it difficult to communicate with people and [I] get uncomfortable when people approach [me]" (PID 8146), while another shared as follows:

> *I feel it negatively affected my social skills slightly, making it harder for me to talk to acquaintances.* [PID 8046]

Examination Stress

In this cluster, children and young people reported high levels of stress due to COVID-19 restrictions and their impact on examinations. One young person reported as follows:

> *[S]chool is way more stressful, there's so many changes to how exams are going or if we don't have any etc. such as what subjects are being dropped or topics we don't need etc.* [PID 7531]

Another person berated the government, noting that "the way the government dealt with my GCSEs was very stressful" (PID 6101). A common complaint was the government's handling of examinations, with this varying depending on pandemic stage; for example, some had qualms with the sudden news of an exam:

> *[D]on't tell people they are not going to have exams and then give them exams with less than a month's warning. It's honestly a joke.* [PID 5386]

Others struggled with the consequences of GCSEs being cancelled:

> *As my GCSEs were obviously cancelled I have been sitting up to 5 exams every day for weeks.* [PID 5996]

Some expressed frustration at "all that study [being] wasted" (PID 4891) and even attributed the physical symptoms felt (eg, tiredness and weakness) to the "uncertainty around my exams" (PID 5996).

Mental Health

Children and young people became more aware of their mental health during the pandemic, with a respondent stating, "I think I have some mental illnesses" (PID 8221) and another sharing that "the lockdowns have severely affected my mental health" (PID 8176). Many children and young people reported having ≥1 mental health conditions (eg, oppositional defiant disorder, obsessive-compulsive disorder, eating disorders, and depression) and commented on the impact that COVID-19 preventive measures had on them:

> *What affected me the most was having to be in and out of school when there was a case of coronavirus as I find it hard to go to school and this made my anxiety disorder worse as I had to restart going back after two weeks of isolation.* [PID 8096]

Nevertheless, some children and young people reported that their "mental health got better" (PID 5629), while one respondent commented as follows:

> *[T]he pandemic helped me to focus on my mental health and forced me to better look after myself so, in a way, it was positive for me.* [PID 7716]

Emotional Impact of the Pandemic

Anxiety was a common complaint:

> *I have felt very anxious about COVID. It has made me not want to leave my house or parents in case they die.* [PID 51]
>
> *[T]he pandemic makes me more worried in general about the small things in everyday life.* [PID 31]

Similarly, a respondent expressed being "a bit more worried about the future and also what may happen as a direct impact of the pandemic" (PID 346). Many children and young people reported experiencing frustration, feeling unhappy, and being more prone to worrying during the pandemic. Moreover, loneliness was also a common complaint, with a respondent reporting as follows:

> *[T]he lockdown and the pandemic have made me feel a lot more isolated than I previously did.* [PID 291]

Physical Health (Including COVID-19 Symptoms)

Within this topical cluster, the children and young people mainly reported the long-lasting impact of testing positive for COVID-19 infection, with changes to smell and taste being prominent:

> *The fact that food tastes and smells bad in some cases (smells bad in many cases) has impacted how much I eat and my diet, for instance I can't eat eggs anymore as they taste rotten. My diet has changed as a result.* [PID 46]
>
> *I have not been able to taste very well for about the past 3 months ever since I got COVID. This includes foods tasting very weird, not eating food I used to like, not being able to smell clear things other people can.* [PID 196]

The impact of changes to smell and taste seemed to be long-lasting; for example, a respondent stated they got their "taste partially back 11 months after I was positive for coronavirus" (PID 2451). Some children and young people made a reference to the term "long COVID":

> *I have long COVID...where I have distorted/lost my smell and taste.* [PID 1621]

Sleep disturbances were also commonly reported:

XSL•FO
RenderX

> *"[M]y sleeping pattern started to go downhill. I was awake all night and sleeping through the day.* [PID 861]

However, there were some who did not report such difficulties, stating they "only had minor symptoms which went away within a week" (PID 1851) or that they "don't feel I have had any long term issues as a result of having COVID" (PID 2786).

Social and Family Support

There was marked heterogeneity in terms of the impact of COVID-19 on social and family support. Some children and young people reported feeling distanced:

> *During COVID I didn't have as much contact with friends or family so I became a bit distant.* [PID 736]

> *It affected me as I couldn't hang out with my friends and I feel like I've lost a connection between them.* [PID 2251]

Others shared the opposite experience:

> *COVID has been devastating for everyone. It's a very difficult and challenging time. However, it has helped me become closer with my family and friends.* [PID 756]

Some made use of web-based platforms to engage with their friends (eg, game consoles and Microsoft Teams), which helped mitigate the impact of the COVID-19 preventive measures. A respondent reported having "been able to spend a lot of time with my friends virtually using my games console so haven't felt bored or lonely" (PID 1016). Though children and young people acknowledged the impact of the lockdown in relation to loss of support:

> *[M]y support system of friends were suddenly cut off and couldn't be there physically for me anymore.* [PID 1026]

However, some also described how it expanded social circles:

> *[O]ver the course of the pandemic I will say that I have made more friends than I would have otherwise because people would add me and my friends into group chats. Some of my closest friends have come from this.* [PID 1026]

Familial problems were reported by some children and young people, with a respondent stating that they "argued with my parents because I spend all my time with them" (PID 3676). Another stated as follows:

> *I think the pandemic has affected me badly. Being confined in a space with my mum and sister and not having a personal space to debrief has been difficult. I have realised a lot about myself and my family and they're not good things but I am grateful that I know this information otherwise I would have gone living my life like I was wrong all the time.* [PID 3271]

Sentiment Analysis: Original and Reweighted

Manual sentiment analysis of the 411 randomly sampled free-text responses yielded 314 (76.4%) negative, 57 (13.87%) positive, and 40 (9.73%) neutral statements. There were observed associations between the free-text sentiment and COVID-19 status as well as sex (P=.001 and P<.001, respectively; Table S3 in Multimedia Appendix 1) such that most of the responses, regardless of COVID-19 status or sex, were negative; for example, 84.1% (227/270) of the responses from female individuals and 61.7% (87/141) of those from male individuals were negative (Table S3 in Multimedia Appendix 1). There were no observed associations between the free-text sentiment and all other examined demographics (ie, age, ethnicity, region of residence, and Index of Multiple Deprivation quintile). When reweighted to the general population of children and young people in England, sentiment proportions remained largely consistent: 78.52% (negative), 13.23% (positive), and 8.24% (neutral). Most of those with negative sentiment were female (227/314, 72.3%), while those with positive sentiment were mostly older (35/57, 61%; Table 2).

XSL•FO
RenderX

Table 2. Unweighted and reweighted demographics of the randomly sampled children and young people stratified by the sentiment of the free-text response provided at study enrollment (n=411).

	Positive sentiment (n=57)		Neutral sentiment (n=40)		Negative sentiment (n=314)	
	Observed (%)	Reweighted (%)	Observed (%)	Reweighted (%)	Observed (%)	Reweighted (%)
COVID-19 status at index PCR[a] test						
SARS-CoV-2 positive	13.33	13.52	3.89	3.71	82.78	82.77
SARS-CoV-2 negative	14.29	12.96	14.29	12.58	71.43	74.46
Sex						
Male	21.28	20.5	17.02	13.84	61.7	65.66
Female	10	9.13	5.93	5.08	84.07	85.79
Age (y) at study invitation						
11-14	11.64	11.32	12.17	10.04	76.19	78.63
15-17	15.77	14.86	7.66	6.7	76.58	78.43
Ethnicity						
Asian or Asian British	18.33	15.57	10	11.38	71.67	73.05
Black British, African British, or Caribbean British	16.67	13.06	0	0	83.33	86.94
White	12.01	12.13	10.71	8.85	77.27	79.01
Mixed	28.57	24.83	4.76	2.31	66.67	72.86
Other	0	0	0	0	100	100
Prefer not to say	0	0	0	0	100	100
Region of residence (England)						
East Midlands	6.67	6.49	13.33	11.85	80	81.66
East of England	11.11	14.02	11.11	11.54	77.78	74.44
London	19.44	20.29	12.5	11.58	68.06	68.13
North East England	5.56	4.71	16.67	13.17	77.78	82.12
North West England	4.08	5.82	6.12	4.78	89.8	89.4
South East England	22.5	17.97	10	8.46	67.5	73.57
South West England	16	16.81	4	2.87	80	80.32
West Midlands	16.22	15.91	2.7	2.72	81.08	81.37
Yorkshire and the Humber	8.11	8.11	10.81	9.87	81.08	82.01
IMD[b] quintile						
1 (most deprived)	14.86	12.66	8.11	6.33	77.03	81.01
2	17.5	16.35	11.25	10.35	71.25	73.3
3	6.25	4.94	10	6.44	83.75	88.63
4	10.99	14.08	6.59	7.24	82.42	78.68
5 (least deprived)	19.77	19.1	12.79	11.37	67.44	69.53

[a]PCR: polymerase chain reaction.

[b]IMD: Index of Multiple Deprivation.

Discussion

Principal Findings

Using an innovative AI method and self-reported data from 8224 children and young people during the COVID-19 pandemic, we found that the experiences of the children and young people fell into 6 main topical clusters related to *school*, *examination stress*, *mental health*, *emotional impact of the pandemic*, *physical health (including COVID-19 symptoms)*, and *social and family support*. Importantly, there was substantial overlap between the topical clusters, such that, for example, *examination stress* and *school* could be collapsed into 1 cluster. Similarly, *mental health* and *emotional impact of the pandemic*

XSL•FO
RenderX

could be classified together under *well-being*. Manual sentiment analysis showed that the majority of statements made by children and young people were largely negative (314/411, 76.4%), with only a small proportion (57/411, 13.9%) classified as positive. When reweighted to the general population of children and young people in England, the overriding sentiment prevalences were largely unchanged.

Our findings are in line with previous literature using different methods that saw COVID-19–related school closure as a contributing factor to the increased stress, frustration, anxiety, and loneliness experienced by children and young people [32]. Repeated cross-sectional comparisons (before and after the pandemic) of children and young people referred to specialist Child and Adolescent Mental Health Services found an increase in emotional problems and depression as well as the number of children and young people reaching thresholds for mental health conditions in those recruited at the time of schools reopening [7]. The authors suggest that this may be due to stresses involved in adjusting to the return to school. Qualitative descriptions paint a similar picture, with a minority benefiting from school closures (eg, because of reduced stress due to the lack of examinations), while the majority wanted to return for the social aspect and reported concentration difficulties, fears of falling behind, and frustration at their seemingly fruitless efforts [33]. Likewise, the examination by Scott et al [34] of the diary entries of children and young people over a 6-week period between July and October 2020 revealed that the pandemic had a negative impact on the mental health of children and young people, with many reporting anxiety, loss, stress, and loneliness, which is in line with the mental health and emotional impact of the pandemic clusters found in this study. Similar to our findings, while concerns surrounding school-related matters were prevalent (eg, long-term consequences of school closures and cancelled examinations), a minority of children and young people viewed school closures positively [34]. Sleep disturbances and loss of taste and smell commonly reported within the topical cluster of COVID-19 symptoms were in line with previous findings, with sleep hygiene negatively affected [11]. In terms of social and family support, the loneliness reported by some children and young people within this cluster is in line with quantitative data from the CLoCk study [23,35]. More broadly, previous literature indicates that 27% and 8% of children and young people reported “slightly more” and “much more” loneliness, respectively, during the first UK national lockdown [19].

The heterogeneity highlighted in this study brings together seemingly opposing findings in the literature; for example, some previous studies reported improved familial and peer relationships [9], whereas others reported tensions at home [36]. An explanation for previous discrepant findings and our current heterogeneous results may be explained by noting that variation in health and well-being by pandemic stage is possible; for example, in the Co-SPACE study, trends in emotional, conduct, and hyperactivity problems followed changes in COVID-19 preventive measures, that is, most problems increased after lockdown announcements and decreased as restrictions eased and schools reopened [18]. Notably, studies reporting improved mental health were typically conducted during the first UK national lockdown (March 2020-June 2020), which was novel because it was children’s and young people’s first experience of a lockdown and may have been qualitatively different from subsequent lockdowns [9]. Using data that largely come from children and young people themselves allows for their voices to be heard both in terms of the content of their written expressions and also the sentiment of the words used. Although previous studies have indicated the heterogeneity of experiences [37], this is the first study that used sentiment analysis to quantify the heterogeneity of experiences and health of children and young people during the pandemic. We found that more than three-quarters of the written expressions of children and young people were negative (314/411, 76.4%), and most of these negative statements were written by female individuals (227/314, 72.3%). By contrast, <15% (57/411) of the statements were positive, and these were mostly written by older children and young people (35/57, 61%). Understanding the basis for the heterogeneity of experiences and health during the pandemic is important because it helps identify factors that may contribute to resilience in the face of adversity as in the case of the pandemic. Notwithstanding these findings, it has also been highlighted that psychological distress during the COVID-19 pandemic varied across countries [38], with, for example, a quarter of the children and young people surveyed in Malaysia reporting depressive symptoms during lockdown and amid rising cases and fatalities [39], while 37% of children and young people aged 18 to 24 years in Southeast and South Asia reported positive attitudes toward the pandemic [40]. Thus, our findings might be restricted to children and young people in England.

Our study’s large sample size is a strength, taking into account the responses of >8000 children and young people regarding their experiences during the pandemic. Nonetheless, study limitations are acknowledged; for example, InfraNodus topical clusters were not entirely distinct: *examination stress* and *school* were similar, as were *mental health* and *emotional impact of the pandemic*. While the AI tool we used considered these as discrete clusters, we acknowledge that children and young people are complex beings, and the thoughts and emotions expressed in their written responses could straddle multiple clusters. While manual examination of the responses did not yield substantive differences between some clusters, the AI software did not allow for the identified clusters to be collapsed. Software limitations were further highlighted in the sentiment analysis. We intended to use the AI software to run the sentiment analysis on all 8224 free-text responses. However, our analysis highlighted several notable discrepancies between InfraNodus-assigned and human-assigned sentiment, especially in complex, nuanced statements. Examples of these discrepancies are provided in Table S1 in Multimedia Appendix 1. This issue reflects a limitation in the AI tool’s capacity to capture the full emotional range in the responses, particularly in contexts involving children’s expressions, which may convey implicit distress or complex emotional states not adequately interpreted by keyword-based algorithms. InfraNodus was not specifically trained on data from children and young people, which likely contributed to its misinterpretations. Sentiment analysis tools developed for general audiences often lack the contextual sensitivity needed for younger populations, whose expressions may differ significantly in language complexity,

XSL•FO
RenderX

structure, and sentiment cues. This highlights the need for AI sentiment tools to be trained specifically on data from children and young people to improve their interpretive accuracy. Furthermore, while we considered applying a natural language processing approach to extend the themes generated from our sentiment analysis and to predict themes across the full sample, this was not feasible within our current resources. Developing a contextual sentiment analysis framework for this purpose, which would allow for predictive insights and the potential identification of overlooked themes, would have required substantial computational support and additional funding. As such, this remains a limitation of our study. We recommend that future research consider integrating natural language processing methods such as topic modeling to analyze the remaining 95% (7813/8224) of the free-text responses, thereby enhancing theme coverage and providing a more comprehensive understanding of the thematic patterns across all data. In addition, reweighting the discourse analysis would have helped ensure representativeness, although this was not possible with the software used. These limitations highlight the importance of not relying exclusively on AI-generated interpretations of text and the need for specialist, substantive knowledge of the topics and data examined.

A small percentage of responses (370/10,580, 3.5%) were truncated and excluded from analysis. However, as the prevalence of truncation was low, it is unlikely that this exclusion introduced bias into our analysis. More generally, although the children and young people invited to take part in the CLoCk study were a nationally representative sample, they are a specific population (children and young people aged 11-17 y who underwent PCR testing). Therefore, children and young people who are invisible to the public health system may be excluded. In addition, there is further potential for selection bias because only 34.12% (10,580/31,012) of those completing the questionnaire answered the free-text response question. However, there seem to be no major demographic differences between those enrolling and those who answered the free-text question. To mitigate this bias, future research might consider making the free-text response question compulsory. The underrepresentation of positive sentiments may reflect a tendency for those more negatively impacted by the pandemic to respond to both the questionnaire and provide a relevant free-text response. To mitigate these selection issues, the sentiment analysis was reweighted to the general population of children and young people in England using census data from 2021 [30]. Although the questionnaire was designed to be completed by children and young people themselves, there was no way to check that responses were written by the children and young people as opposed to a parent or carer. Moreover, parental influence may have affected responses among those children and young people who completed it with parental assistance. Future studies would benefit from distinguishing between the input of children and young people and parental input by, for example, including a question asking whether the questionnaire was completed solely by the child or young person, solely by a parent or carer, or primarily by the child or young person with assistance from a parent or carer.

As in any manual qualitative data analysis, unconscious rater bias had the potential to affect the results. Thus, measures were taken to minimize its impact: for the manual sentiment analysis, a second rater assessed the sentiment of half the responses (206/411, 50.1%), yielding excellent agreement (195/206, 94.7%) between the raters. We acknowledge that the relevancy assessment of all 10,580 free-text responses by a second rater would have been beneficial; however, this was not possible due to limited capacity. Nevertheless, subjectivity in determining the relevance of statements during the relevancy assessment was mitigated by discussing all uncertainties with another researcher and by adopting a collaborative approach to developing the relevancy criteria. Finally, we acknowledge that temporal variation in health and experiences is possible and remains a literature gap that needs addressing. However, this was beyond the scope of this study. Ongoing work is examining how the health and experiences of children and young people evolved throughout the pandemic to provide a more comprehensive understanding of how the pandemic impacted their physical, mental, and emotional well-being, as well as their access to health care, education, and social support systems.

Conclusions

In conclusion, the study used InfraNodus to analyze >8000 free-text responses from children and young people during the COVID-19 pandemic. What emerged was that the majority of the text responses (314/411, 76.4%) indicated negative experiences, and many of the respondents reported experiencing distress across a range of domains related to school, social situations, and mental health. In terms of practical implications for policy makers, educators, and health care providers, this analysis highlights the importance of specific considerations for children and young people when responding to national emergencies; for example, the insights gained show that consistency in relation to education guidance, mitigations for social isolation and loneliness, and increased support on return to education would have reduced much stress; therefore, the application of these measures during and after national emergencies should be considered.

Acknowledgments

This work is independent research jointly funded by the National Institute for Health and Care Research (NIHR) and UK Research and Innovation (UKRI; Children and Young People With Long COVID study; COV-LT-0022). All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. SMPP was supported by a UK Medical Research Council Career Development Award (MR/P020372/1) and a Senior Nonclinical Fellowship (MR/Y009398/1). The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, UKRI, or the Department of Health and Social Care. The Department of Health and Social Care, as the NIHR and UKRI awarded grant COVLT0022, but was not involved in

XSL•FO
RenderX

study design, data collection, analysis, interpretation, or manuscript writing. Michael Lattimore, UK Health Security Agency, is acknowledged as project officer for the Children and Young People With Long COVID study.

Data Availability

Children and Young People With Long COVID study data are publicly available in the UK Data Service repository [41].

Authors' Contributions

NKR conducted the analysis for the study, accessed and verified the data, and drafted the manuscript. SM assisted with the analysis, including its design, and contributed to the drafting of the manuscript. MC-B assisted with the analysis, including its design, and reviewed the manuscript. RS contributed to the design of the Children and Young People With Long COVID (CLoCk) study, submitted the ethics and research and development applications, and reviewed the manuscript. LF-S contributed to data cleaning and analysis and reviewed the manuscript. BCFC reviewed the manuscript. AdO contributed to data cleaning and reviewed the manuscript. TN assisted with the analysis, including its design, and reviewed the manuscript. YX assisted with the literature review and reviewed the manuscript. KM adapted the questionnaire for the online Snap Surveys platform and reviewed the manuscript. ED, IH, TF, and TC contributed to the design of the CLoCk study and reviewed the manuscript. RS accessed and verified the data and designed the participant sampling and dataflow for the CLoCk study. TS conceived the idea for the CLoCk study, submitted the successful grant application, and reviewed the manuscript. SMPP conceived the idea for this study, assisted with the statistical analyses, accessed and verified the data, and drafted the manuscript. All members of the CLoCk Consortium made contributions to the conception or design of the work, were involved in drafting both the funding application and the manuscript, approved the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Members of the CLoCk Consortium: Dougal Hargreaves, Shamez Ladhani, Marta Buszewicz, Esther Crawley, Bianca De Stavola, Shruti Garg, Anthony Harnden, Michael Levin, Vanessa Poustie, Kishan Sharma (sadly deceased), Olivia Swann, Roz Shafran, Terence Stephenson, Isobel Heyman, Tamsin Ford, Trudie Chalder, Emma Dalrymple, Terry Segal, Malcolm Semple, and Elizabeth Whittaker.

Conflicts of Interest

TS is chair of the UK Health Research Authority and therefore recused himself from the research ethics application. TC is part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, King's College London. She has grants from Guy’s St Thomas Charity Grants, NIHR, and UKRI for post–COVID-19 syndromes. TC is a member of the National Institute for Health and Care Excellence committee for post–COVID-19 condition. She has received travel expenses, accommodation costs, and honorarium for several lectures in Europe and has received travel expenses and accommodation costs for attending the American Thoracic Society Conference. She was on the Expert Advisory Panel for COVID-19 Rapid Guidelines. She has written self-help books on chronic fatigue and has conducted workshops on chronic fatigue and postinfectious syndromes. TF’s research group receives payment for research consultancy to Place2Be, a third-sector organization providing mental health training and interventions in UK schools. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Supplementary text and tables.
[DOCX File , 277 KB - jmir_v27i1e63634_app1.docx]

References

1. Ng CS, Ng SS. Impact of the COVID-19 pandemic on children's mental health: a systematic review. Front Psychiatry 2022 Oct 18;13:975936 [FREE Full text] [doi: 10.3389/fpsyt.2022.975936] [Medline: 36329921]
2. Orben A, Tomova L, Blakemore SJ. The effects of social deprivation on adolescent development and mental health. Lancet Child Adolesc Health 2020 Aug;4(8):634-640 [FREE Full text] [doi: 10.1016/S2352-4642(20)30186-3] [Medline: 32540024]
3. Torjesen I. COVID-19: mental health services must be boosted to deal with "tsunami" of cases after lockdown. BMJ 2020 May 15;369:m1994. [doi: 10.1136/bmj.m1994] [Medline: 32417756]
4. Bignardi G, Dalmaijer ES, Anwyl-Irvine AL, Smith TA, Siugzdaite R, Uh S, et al. Longitudinal increases in childhood depression symptoms during the COVID-19 lockdown. Arch Dis Child 2021 Jul 19;106(8):791-797 [FREE Full text] [doi: 10.1136/archdischild-2020-320372] [Medline: 33298552]
5. Loades ME, Chatburn E, Higson-Sweeney N, Reynolds S, Shafran R, Brigden A, et al. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. J Am Acad Child Adolesc Psychiatry 2020 Nov;59(11):1218-39.e3 [FREE Full text] [doi: 10.1016/j.jaac.2020.05.009] [Medline: 32504808]

XSL•FO
RenderX

6. Meherali S, Punjani N, Louie-Poon S, Abdul Rahim K, Das JK, Salam RA, et al. Mental health of children and adolescents amidst COVID-19 and past pandemics: a rapid systematic review. Int J Environ Res Public Health 2021 Mar 26;18(7):3432 [FREE Full text] [doi: 10.3390/ijerph18073432] [Medline: 33810225]
7. Sayal K, Partlett C, Bhardwaj A, Dubicka B, Marshall T, Gledhill J, et al. Mental health in clinically referred children and young people before and during the COVID-19 pandemic. Eur Child Adolesc Psychiatry 2023 Dec 17;32(12):2657-2666 [FREE Full text] [doi: 10.1007/s00787-022-02115-2] [Medline: 36526804]
8. Wright N, Hill J, Sharp H, Pickles A. Interplay between long-term vulnerability and new risk: young adolescent and maternal mental health immediately before and during the COVID-19 pandemic. JCPP Adv 2021 Jan 18;1(1):e12008 [FREE Full text] [doi: 10.1111/jcv2.12008] [Medline: 34485987]
9. Soneson E, Puntis S, Chapman N, Mansfield KL, Jones PB, Fazel M. Happier during lockdown: a descriptive analysis of self-reported wellbeing in 17,000 UK school students during Covid-19 lockdown. Eur Child Adolesc Psychiatry 2023 Jun 17;32(6):1131-1146 [FREE Full text] [doi: 10.1007/s00787-021-01934-z] [Medline: 35174418]
10. Penner F, Hernandez Ortiz J, Sharp C. Change in youth mental health during the COVID-19 pandemic in a majority Hispanic/Latinx US sample. J Am Acad Child Adolesc Psychiatry 2021 Apr;60(4):513-523. [doi: 10.1016/j.jaac.2020.12.027] [Medline: 33359408]
11. Larsen L, Helland MS, Holt T. The impact of school closure and social isolation on children in vulnerable families during COVID-19: a focus on children's reactions. Eur Child Adolesc Psychiatry 2022 Aug 26;31(8):1-11 [FREE Full text] [doi: 10.1007/s00787-021-01758-x] [Medline: 33770275]
12. Newlove-Delgado T, Russell AE, Mathews F, Cross L, Bryant E, Gudka R, et al. Annual research review: the impact of COVID-19 on psychopathology in children and young people worldwide: systematic review of studies with pre- and within-pandemic data. J Child Psychol Psychiatry 2023 Apr 24;64(4):611-640 [FREE Full text] [doi: 10.1111/jcpp.13716] [Medline: 36421049]
13. Demkowicz O, Ashworth E, O'Neill A, Hanley T, Pert K. Will my young adult years be spent socially distancing?: a qualitative exploration of adolescents' experiences during the COVID-19 UK lockdown. J Adolesc Res 2022;39(6):1476-1511. [doi: 10.21203/rs.3.rs-451696/v1]
14. Widnall E, Adams EA, Plackett R, Winstone L, Haworth CM, Mars B, et al. Adolescent experiences of the COVID-19 pandemic and school closures and implications for mental health, peer relationships and learning: a qualitative study in South-West England. Int J Environ Res Public Health 2022 Jun 10;19(12):7163 [FREE Full text] [doi: 10.3390/ijerph19127163] [Medline: 35742408]
15. Lenoir R, Wong KK. Impact of the COVID-19 pandemic on young people from black and mixed ethnic groups' mental health in West London: a qualitative study. BMJ Open 2023 May 05;13(5):e071903 [FREE Full text] [doi: 10.1136/bmjopen-2023-071903] [Medline: 37147089]
16. McKinlay AR, May T, Dawes J, Fancourt D, Burton A. 'You're just there, alone in your room with your thoughts': a qualitative study about the psychosocial impact of the COVID-19 pandemic among young people living in the UK. BMJ Open 2022 Feb 09;12(2):e053676 [FREE Full text] [doi: 10.1136/bmjopen-2021-053676] [Medline: 35140155]
17. Paterson J, McCarthy M, Triantafyllopoulou P. The impact of the coronavirus pandemic on the lives of children and young people who have special educational needs and/or disabilities in the UK: a scoping review. J Res Spec Educ Needs 2023 Jul 12;24(1):12-24. [doi: 10.1111/1471-3802.12608]
18. Guzman Holst C, Bowes L, Waite P, Skripkauskaite S, Shum A, Pearcey S, et al. Examining Children and adolescent mental health trajectories during the COVID-19 pandemic: findings from a year of the Co-SPACE study. JCPP Adv 2023 Jun 18;3(2):e12153 [FREE Full text] [doi: 10.1002/jcv2.12153] [Medline: 37753152]
19. Geulayov G, Mansfield K, Jindra C, Hawton K, Fazel M. Loneliness and self-harm in adolescents during the first national COVID-19 lockdown: results from a survey of 10,000 secondary school pupils in England. Curr Psychol 2022 Sep 15;43(15):1-12 [FREE Full text] [doi: 10.1007/s12144-022-03651-5] [Medline: 36124048]
20. Waite P, Pearcey S, Shum A, Raw JA, Patalay P, Creswell C. How did the mental health symptoms of children and adolescents change over early lockdown during the COVID-19 pandemic in the UK? JCPP Adv 2021 Apr 28;1(1):e12009 [FREE Full text] [doi: 10.1111/jcv2.12009] [Medline: 34485988]
21. Nugawela M, Pinto Pereira SM, Rojas NK, McOwat K, Simmons R, Dalrymple E, CLoCk Consortium. Data resource profile: the children and young people with long COVID (CLoCk) study. Int J Epidemiol 2024 Feb 01;53(1):1 [FREE Full text] [doi: 10.1093/ije/dyad158] [Medline: 37990948]
22. Rojas NK, De Stavola BL, Norris T, Cortina-Borja M, Nugawela MD, Hargreaves D, CLoCk Consortium, et al. Developing survey weights to ensure representativeness in a national, matched cohort study: results from the children and young people with Long Covid (CLoCk) study. BMC Med Res Methodol 2024 Jun 20;24(1):134 [FREE Full text] [doi: 10.1186/s12874-024-02219-0] [Medline: 38902672]
23. Schneider V, Norris T, Nugawela M, Dalrymple E, Hargreaves D, Käll A, CLoCk Consortium members. Loneliness trajectories, associated factors and subsequent health in children and young people during the COVID-19 pandemic: a national matched cohort study. Psychol Res Behav Manag 2023;16:4461-4477 [FREE Full text] [doi: 10.2147/PRBM.S421165] [Medline: 37936971]

XSL•FO
RenderX

24. Stephenson T, Shafran R, De Stavola B, Rojas N, Aiano F, Amin-Chowdhury Z, et al. Long COVID and the mental and physical health of children and young people: national matched cohort study protocol (the CLoCk study). BMJ Open 2021 Aug 26;11(8):e052838. [doi: 10.1136/bmjopen-2021-052838]
25. English indices of deprivation 2015. Ministry of Housing Communities & Local Government. URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 [accessed 2024-04-29]
26. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry 2001 Nov;40(11):1337-1345. [doi: 10.1097/00004583-200111000-00015] [Medline: 11699809]
27. InfraNodus. Nodus Labs. URL: https://infranodus.com/ [accessed 2024-04-29]
28. Paranyushkin D. InfraNodus: generating insight using text network analysis. In: Proceedings of the 2019 Conference on the World Wide Web. 2019 Presented at: WWW '19; May 13-17, 2019; San Francisco, CA p. 3584-3589 URL: https://dl.acm.org/doi/10.1145/3308558.3314123 [doi: 10.1145/3308558.3314123]
29. Stata statistical software. StataCorp LLC. URL: https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/ [accessed 2024-04-29]
30. Census 2021 results. Office for National Statistics. URL: https://census.gov.uk/census-2021-results [accessed 2024-04-29]
31. Research involving children. Health Research Authority. URL: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/research-involving-children/ [accessed 2024-04-29]
32. Chaabane S, Doraiswamy S, Chaabna K, Mamtani R, Cheema S. The impact of COVID-19 school closure on child and adolescent health: a rapid systematic review. Children (Basel) 2021 May 19;8(5):415 [FREE Full text] [doi: 10.3390/children8050415] [Medline: 34069468]
33. Fisher H, Lambert H, Hickman M, Yardley L, Audrey S. Experiences of the coronavirus disease-19 (COVID-19) pandemic from the perspectives of young people: rapid qualitative study. Public Health Pract (Oxf) 2021 Nov;2:100162 [FREE Full text] [doi: 10.1016/j.puhip.2021.100162] [Medline: 34254057]
34. Scott S, McGowan VJ, Visram S. 'I'm gonna tell you about how Mrs Rona has affected me'. exploring young people's experiences of the COVID-19 pandemic in north-east England: a qualitative diary-based study. Int J Environ Res Public Health 2021 Apr 06;18(7):3837 [FREE Full text] [doi: 10.3390/ijerph18073837] [Medline: 33917557]
35. McOwat K, Pinto Pereira SM, Nugawela MD, Ladhani SN, Newlands F, Stephenson T, Consortium, et al. The CLoCk study: a retrospective exploration of loneliness in children and young people during the COVID-19 pandemic, in England. PLoS One 2023 Nov 21;18(11):e0294165 [FREE Full text] [doi: 10.1371/journal.pone.0294165] [Medline: 37988366]
36. Fleming P, O'Hara J. Impact on family life during COVID19 pandemic: Barnardos. J Leis Res 2020;35:163-189 [FREE Full text]
37. Cross L, Carey E, Benham-Clarke S, Hartley A, Mathews F, Burn AM, et al. Navigating changes: reflecting on children and young people's experiences of public health and social measures during the COVID-19 pandemic- a purposive, qualitative follow-up from a national probability sample. Emot Behav Diffic 2024 Jul 11;29(1-2):18-35. [doi: 10.1080/13632752.2024.2360778]
38. Marzo RR, Ismail Z, Nu Htay MN, Bahari R, Ismail R, Villanueva 3rd EQ, et al. Psychological distress during pandemic COVID-19 among adult general population: result across 13 countries. Clin Epidemiol Glob Health 2021 Apr;10:100708 [FREE Full text] [doi: 10.1016/j.cegh.2021.100708] [Medline: 33619459]
39. Marzo RR, Vinay V, Bahari R, Chauhan S, Ming DA, Nelson Fernandez SF, et al. Depression and anxiety in Malaysian population during third wave of the COVID-19 pandemic. Clin Epidemiol Glob Health 2021 Oct;12:100868 [FREE Full text] [doi: 10.1016/j.cegh.2021.100868] [Medline: 34549098]
40. Rahman MM, Marzo RR, Chowdhury S, Qalati SA, Hasan MN, Paul GK, et al. Knowledge, attitude and practices toward coronavirus disease (COVID- 19) in southeast and south Asia: a mixed study design approach. Front Public Health 2022 Jun 21;10:875727 [FREE Full text] [doi: 10.3389/fpubh.2022.875727] [Medline: 35801233]
41. Long COVID in children and young people (the CLoCk study): a national cohort study, 2020-2022. UK Data Service. URL: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=9203 [accessed 2024-04-29]

Abbreviations

AI: artificial intelligence
CLoCk: Children and Young People With Long COVID
PCR: polymerase chain reaction

XSL•FO
RenderX

Edited by A Mavragani; submitted 26.06.24; peer-reviewed by RR Marzo, K Wong; comments to author 22.10.24; revised version received 22.11.24; accepted 05.12.24; published 28.01.25.

Please cite as:
Rojas NK, Martin S, Cortina-Borja M, Shafran R, Fox-Smith L, Stephenson T, Ching BCF, d'Oelsnitz A, Norris T, Xu Y, McOwat K, Dalrymple E, Heyman I, Ford T, Chalder T, Simmons R, CLoCk Consortium, Pinto Pereira SM
Health and Experiences During the COVID-19 Pandemic Among Children and Young People: Analysis of Free-Text Responses From the Children and Young People With Long COVID Study
J Med Internet Res 2025;27:e63634
URL: https://www.jmir.org/2025/1/e63634
doi:10.2196/63634
PMID:

©Natalia K Rojas, Sam Martin, Mario Cortina-Borja, Roz Shafran, Lana Fox-Smith, Terence Stephenson, Brian C F Ching, Anaïs d'Oelsnitz, Tom Norris, Yue Xu, Kelsey McOwat, Emma Dalrymple, Isobel Heyman, Tamsin Ford, Trudie Chalder, Ruth Simmons, CLoCk Consortium, Snehal M Pinto Pereira. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of Mobile Phone Usage on Sleep Quality Among Medical Students Across Latin America: Multicenter Cross-Sectional Study

Juan S Izquierdo-Condoy[1], MD, MSc; Clara Paz[2], PhD; H A Nati-Castillo[3], MD; Ricardo Gollini-Mihalopoulos[4], MD; Telmo Raul Aveiro-Róbalo[5], MD, MSc; Jhino Renson Valeriano Paucar[6], MD; Sandra Erika Laura Mamami[7], MD; Juan Felipe Caicedo[8], MD; Valentina Loaiza-Guevara[9], MD; Diana Camila Mejía[8], MD; Camila Salazar-Santoliva[1], MD; Melissa Villavicencio-Gomezjurado[1], MD; Cougar Hall[10], PhD; Esteban Ortiz-Prado[1], MD, MPH, MSc, PhD

[1]One Health Research Group, Universidad de las Américas, Quito, Ecuador

[2]Grupo de Investigación Bienestar, Salud y Sociedad, Escuela de Psicología y Educación, Universidad de Las Américas, Quito, Ecuador

[3]Interinstitutional Internal Medicine Group (GIMI 1), Department of Internal Medicine, Universidad Libre, Cali, Colombia

[4]Facultad de Medicina, Universidad de Panamá, Ciudad de Panamá, Panama

[5]Facultad de Medicina, Universidad del Pacifico, Asunción, Paraguay

[6]Facultad de Medicina Humana, Universidad Nacional del Altiplano, Puno, Peru

[7]Facultad de Ciencias de la Salud, Universidad Privada Franz Tamayo, La Paz, Bolivia

[8]Facultad de Ciencias de la Salud, Universidad del Quindío, Armenia, Colombia

[9]Facultad de Medicina, Institución Universitaria Visión de las Américas, Pereira, Colombia

[10]Public Health Department, Bringham Young University, Provo, UT, United States

Corresponding Author:
Esteban Ortiz-Prado, MD, MPH, MSc, PhD
One Health Research Group
Universidad de las Américas
Calle de los Colimes
Quito, 170137
Ecuador
Phone: 593 0995760693
Email: e.ortizprado@gmail.com

Abstract

Background: The ubiquitous use of mobile phones among medical students has been linked to potential health consequences, including poor sleep quality.

Objective: This study investigates the prevalence of mobile phone addiction and its association with sleep quality among medical students across 6 Latin American countries.

Methods: A descriptive, cross-sectional, multicenter study was conducted between December 2023 and March 2024 using a self-administered online survey. The survey incorporated the Mobile Phone Addiction Scale and the Pittsburgh Sleep Quality Index to evaluate mobile phone addiction and sleep quality among 1677 medical students. A multiple regression model was applied to analyze the relationship between mobile phone addiction and poor sleep quality, adjusting for sex, age, and educational level to ensure robust results.

Results: Mobile phone addiction was identified in 32.5% (545/1677) of participants, with significant differences across countries. The overall mean Pittsburgh Sleep Quality Index score was 7.26, indicating poor sleep quality. Multiple regression analysis revealed a strong association between mobile phone addiction and poor sleep, controlled for demographic variables (β=1.4, 95% CI 1.05-1.74).

Conclusions: This study underscores a significant prevalence of mobile phone addiction among medical students and its detrimental association with sleep quality in Latin America. The findings advocate for the need to address mobile phone usage to mitigate its negative implications on student health and academic performance. Strategies to enhance digital literacy and promote healthier usage habits could benefit medical education and student well-being.

(J Med Internet Res 2025;27:e60630) doi:10.2196/60630

XSL•FO
RenderX

KEYWORDS

mobile phone; addiction behavior; sleep quality; medical students; Latin America

Introduction

By 2024, approximately 17.8 billion mobile phones are projected to be in use globally, according to Statista [1]. Initially designed for voice communication, mobile phones have evolved into essential tools for text messaging, complex communication, and daily activities [2]. The advent of smartphones with advanced computational capabilities has further integrated these devices into modern life [3]. While they offer substantial benefits, such as enhanced access to educational and informational resources, excessive use also poses significant risks [4]. Acute or chronic overuse of mobile phones has been associated with physical issues like headaches, carpal tunnel syndrome, tendinitis, and accidents caused by distraction [5]. In addition, excessive use can result in less overt but impactful effects, including reduced motivation, impaired memory and concentration, sleep disturbances, and learning difficulties [6].

One of the most well-documented negative consequences of mobile phone usage is its disruption of sleep patterns and quality. Prolonged screen exposure and late-night cellphone use adversely affect both the quantity and quality of sleep [7,8]. Poor sleep has been linked to a range of health problems, including cardiovascular diseases, mental health disorders, neurodegenerative conditions, musculoskeletal issues, and a reduced overall quality of life [9,10]. University students, particularly those belonging to Generation Z (Gen Z), also known as iGen or postmillennials, face compounded risks. Born between 1996 and 2015, Gen Z is the first generation to grow up with the internet and portable digital technology. These “digital natives” excel in mobile phone use and social media interaction but have less experience with traditional communication methods, such as print media. For medical students, the demands of their academic schedules, combined with excessive mobile phone usage, pose risks to mental health, sleep quality, and academic performance [11].

In recent years, mobile phones have become indispensable tools in the academic training of medical students [12]. Originally designed for personal communication, these devices have evolved into essential academic resources, facilitating learning and collaboration. However, this transition has also been associated with dependency and excessive use, raising concerns about potential addiction [13]. Constant internet connectivity provided by smartphones fosters engagement not only with educational resources but also with addictive behaviors and content [3]. These include compulsive information seeking, excessive use of online games and shopping platforms, and problematic online interactions such as cybersexuality and cybercontacts [14]. For medical students, in particular, problematic smartphone use, nomophobia (a fear of not having mobile phone connectivity), and general dependency not only negatively impacts academic performance but also poses significant health risks [15-18].

Recent studies in the United States and Turkey have identified a correlation between increased mobile phone use and poor sleep quality among university and medical students [19,20]. A meta-analysis further highlighted that sleep quality is closely linked to academic performance in medical students, emphasizing the need for interventions to address this issue [14]. Despite this growing body of evidence, research on the impact of mobile phone addiction on sleep quality among medical students in Latin America remains limited.

Medical students in Latin America face unique challenges, including demanding schedules, irregular sleep patterns, and high stress levels, which may make them particularly vulnerable to mobile phone addiction. In addition, the reliance on mobile phones for communication and education in resource-limited health care education systems could exacerbate these issues.

This study aims to investigate the prevalence of mobile phone addiction and its association with sleep quality among medical students in 6 Latin American countries. By addressing this gap, we hope to provide insights that can inform targeted interventions to improve the well-being and academic success of medical students in the region.

Methods

Study Design

A descriptive, cross-sectional, multicenter study was conducted from December 2023 to March 2024.

Setting and Participants

This study used a self-administered online survey targeting medical students enrolled at universities across 6 Latin American countries: Bolivia, Colombia, Ecuador, Panama, Paraguay, and Peru. Participants included undergraduate students from the first to the final years (sixth or seventh) of their medical degree. Participants were selected using a nonprobability convenience sampling method through the online survey platform SurveyMonkey. Participation was voluntary.

Survey Development and Measures

The research team designed a 49-item anonymous online survey to collect data on demographic characteristics, mobile phone usage, and sleep quality among the Latin American medical student population. An initial version of the questionnaire was reviewed by a public health expert to ensure its relevance, accuracy, and to preempt potential errors. It was then pilot tested with 20 medical students from Ecuador to fine tune comprehension and design; data from this phase were excluded from the final analysis. Revisions after pilot test produced the final version of the survey, which was drafted in Spanish. The report on the online survey used in this research, prepared following the CHERRIES (Checklist for Reporting Results of Internet E-Surveys), is provided in Table S1 in Multimedia Appendix 1.

Questionnaire and Variables

The final version of the questionnaire was structured into 3 sections to meet the research objectives.

XSL•FO
RenderX

Demographic Data

The first section collected demographic information such as sex, age, country of residence, year of study, and type of university, categorized by funding source (public or private). It also included an additional year category (7th year) to account for students in countries such as Colombia and Panama, where medical training can extend up to 13 semesters.

Mobile Phone Addiction

The second section used the Mobile Phone Addiction Scale, designed and validated by Basu et al [21], which adheres to *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) criteria for substance dependence syndrome. It comprised 20 questions rated on a 6-point Likert scale (from 1-completely disagree to 6-completely agree). The questionnaire evaluated 6 addiction components: intense desire, poor control, abstinence, tolerance, diminished interest in alternative pleasures, and harmful use. Addiction presence was defined if half or more of the questions in each component were affirmed; global addiction was noted when three or more components were present.

Sleep Quality

The third section used the standardized Pittsburgh Sleep Quality Index (PSQI) [22,23], designed to evaluate sleep quality through 19 questions covering 7 components: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. Scores ranged from 0 (no difficulty) to 3 (very severe difficulty), with a total score from 0 to 21. A score of 5 or less indicated normal rest, while scores greater than 5 indicated poor rest, categorizing the overall sleep quality level.

Data Collection and Management

The online platform SurveyMonkey was used to facilitate data collection. Participants accessed the survey through a unique link shared across social media channels, including Facebook and WhatsApp (Meta). The preamble of the survey outlined the study's aims, assured confidentiality, and sought informed consent. Participants were required to agree to the informed consent and accept the participation agreement to proceed, reinforcing the survey's anonymity. To ensure high data integrity, all responses were meticulously reviewed for potential errors or inconsistencies, such as implausible age ranges or respondents selecting all available answers. This scrutiny led to the exclusion of questionnaires from an original sample of 1798 survey responses before filtering to a final total of 1677 responses which were considered valid and included in the study (Figure 1).

Figure 1. Flowchart of the Strengthening the Reporting of Observational Studies in Epidemiology–based sample selection process for this study.

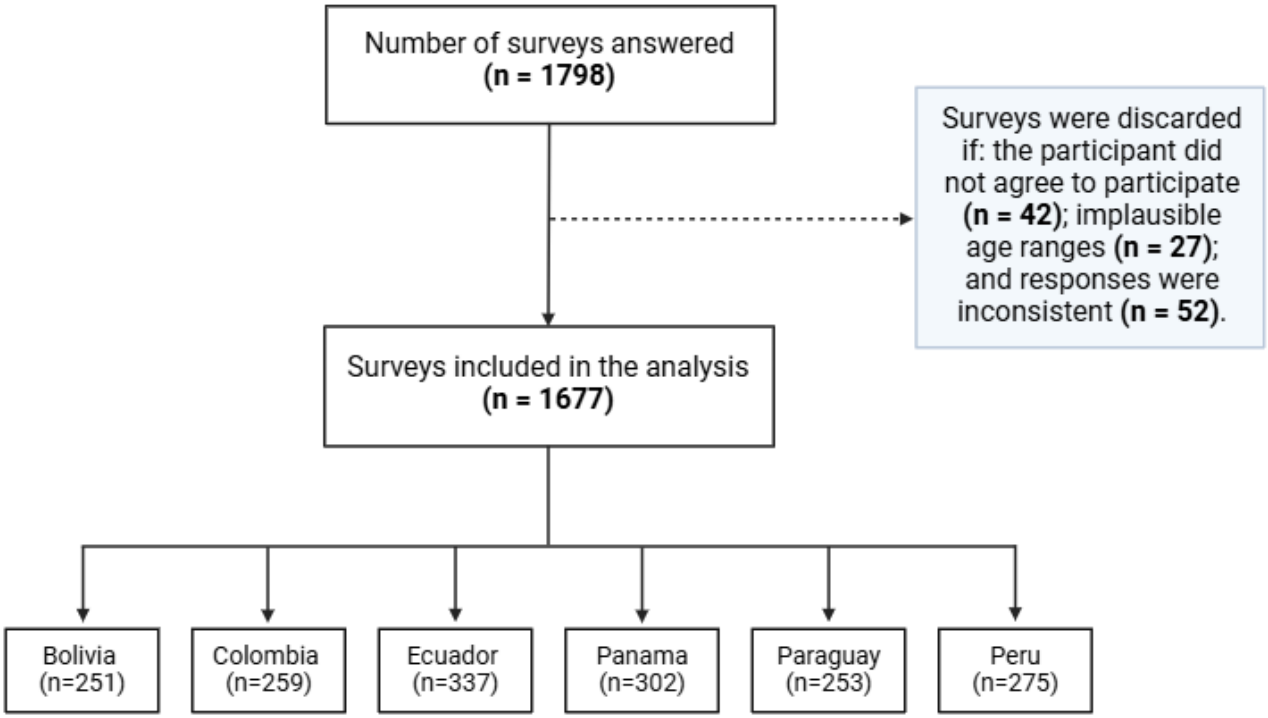

Bias

To mitigate potential bias during data collection and management, several strategies were used. For example, IP addresses were used to eliminate duplicate submissions by configuring the SurveyMonkey platform to limit 1 response per IP address. To ensure anonymity the questionnaire was carefully curated to avoid collecting any identifiable information, including IP addresses, beyond what was necessary for filtering duplicates. In the analysis phase, additional steps were taken to minimize bias. One of the researchers meticulously reviewed the results, and any discrepancies identified were collaboratively resolved. This rigorous approach ensured that only valid and genuine responses were included in the final analysis, substantially enhancing the reliability and credibility of the study's findings. The use of a nonprobability convenience

XSL•FO
RenderX

sampling method was justified in this study of medical students across 6 Latin American countries due to logistical, resource, and time constraints. This method enabled the inclusion of participants from multiple universities, making the study feasible and cost-effective while capturing valuable insights into issues such as mobile phone addiction and sleep quality. Although convenience sampling introduces potential biases, including selection bias and limited generalizability, it was the most practical option given the difficulty of accessing a more representative sample across diverse regions. Despite these limitations, the approach allowed for timely data collection and provided exploratory insights into the specific challenges faced by medical students.

Statistical Analysis

The current study used descriptive statistical tests to analyze the responses for each categorical variable, including the calculation of frequencies and percentages. For quantitative variables, measures of central tendency (mean) and dispersion (SD) were evaluated.

To evaluate the relationships between the variables studied, inferential statistical tests were used. Chi-square tests were conducted to assess the influence of participant characteristics on the presence of mobile phone addiction. *t* tests and one-way ANOVA were used to examine the influence of characteristics on sleep quality as determined by mean PSQI scores.

Ordinary least squares (OLS) regression was conducted as part of a multiple regression model adjusted for sex, age, and educational level was used to estimate the association between mobile phone addiction and poor sleep quality, expressed by β and 95% CI. In all statistical analyses, a *P* value of <.05 was considered statistically significant. All the analysis was conducted using R (R Foundation for Statistical Computing).

Ethical Considerations

This study was conducted in strict adherence to the ethical standards set forth in the Declaration of Helsinki. It also followed the ethical protocols approved by the Ethics Committee of Universidad de Las Américas, under code 2023-EXC-004. The research team was trained to ensure that all aspects of the study upheld the principles of participant anonymity and voluntary participation. No personally identifiable or sensitive information was solicited or included in the research data.

Results

Demographic Characteristics

A total of 1677 medical students from 6 Latin American countries participated in this study. Ecuador and Panama had the highest representations at 20.1% and 18.0%, respectively. The majority of participants were female (1083/1677, 64.4%), and over half (973/1677, 58%) were in their initial years of undergraduate study (1st to 3rd year). The majority (967/1677, 57.7%) were enrolled in private universities (Table 1).

Table 1. Demographic characteristics of medical students from 6 Latin American countries.

Characteristics	Bolivia (n=251	Colombia (n=259)	Ecuador (n=337)	Panama (n=302)	Paraguay (n=253)	Peru (n=275)	Overall (n=1677)
Sex, n (%)							
Male	99 (39.4)	103 (39.8)	118 (35)	113 (37.4)	73 (28.9)	88 (32)	594 (35.4)
Female	152 (60.6)	156 (60.2)	219 (65)	189 (62.6)	180 (71.1)	187 (68)	1083 (64.6)
Age, mean (SD)							
	20.2 (3.14)	21.6 (3.46)	21.7 (2.18)	22.1 (2.27)	23 (4.58)	22.3 (3.76)	21.8 (3.36)
Year of study, n (%)							
First	111 (44.2)	31 (12)	12 (3.6)	22 (7.3)	67 (26.5)	35 (12.7)	278 (16.6)
Second	104 (41.4)	51 (19.7)	43 (12.8)	52 (17.2)	33 (13.0)	64 (23.3)	347 (20.7)
Third	28 (11.2)	53 (20.5)	124 (36.8)	60 (19.9)	35 (13.8)	48 (17.5)	348 (20.8)
Fourth	4 (1.6)	34 (13.1)	124 (36.8)	20 (6.6)	58 (22.9)	38 (13.8)	278 (16.6)
Fifth	2 (0.8)	65 (25.1)	14 (4.2)	41 (13.6)	32 (12.6)	33 (12)	187 (11.2)
Sixth	2 (0.8)	18 (6.9)	8 (2.4)	102 (33.8)	20 (7.9)	19 (6.9)	169 (10.1)
Seventh	0 (0)	7 (2.7)	12 (3.6)	5 (1.7)	8 (3.2)	38 (13.8)	70 (4.2)
Level of education, n (%)							
Initial (first-third year)	243 (96.8)	135 (52.1)	179 (53.1)	134 (44.4)	135 (53.4)	147 (53.5)	973 (58)
Advanced (fourth-seventh year)	8 (3.2)	124 (47.9)	158 (46.9)	168 (55.6)	118 (46.6)	128 (46.5)	704 (42)
Type of university, n (%)							
Private university	248 (98.8)	89 (34.4)	256 (76%)	41 (13.6%)	229 (90.5)	104 (37.8)	967 (57.7)
Public university	3 (1.2)	170 (65.6)	81 (24%)	261 (86.4%)	24 (9.5)	171 (62.2)	710 (42.3)

XSL•FO
RenderX

Smartphone Usage

Smartphones were used by nearly the entire study population (1651/1677, 98.4%), with 94.9% (1591/1677) reporting using mobile phones for academic purposes. However, computers or laptops were preferred for studies (797/1677, 47.5%), while mobile phones were the least preferred option for studying (108/1677, 6.4%; Figure 2).

Figure 2. Characteristics of cell phone use, academic preparation, and distribution of cell phone addiction and sleep quality by country of residence among Latin American medical students.

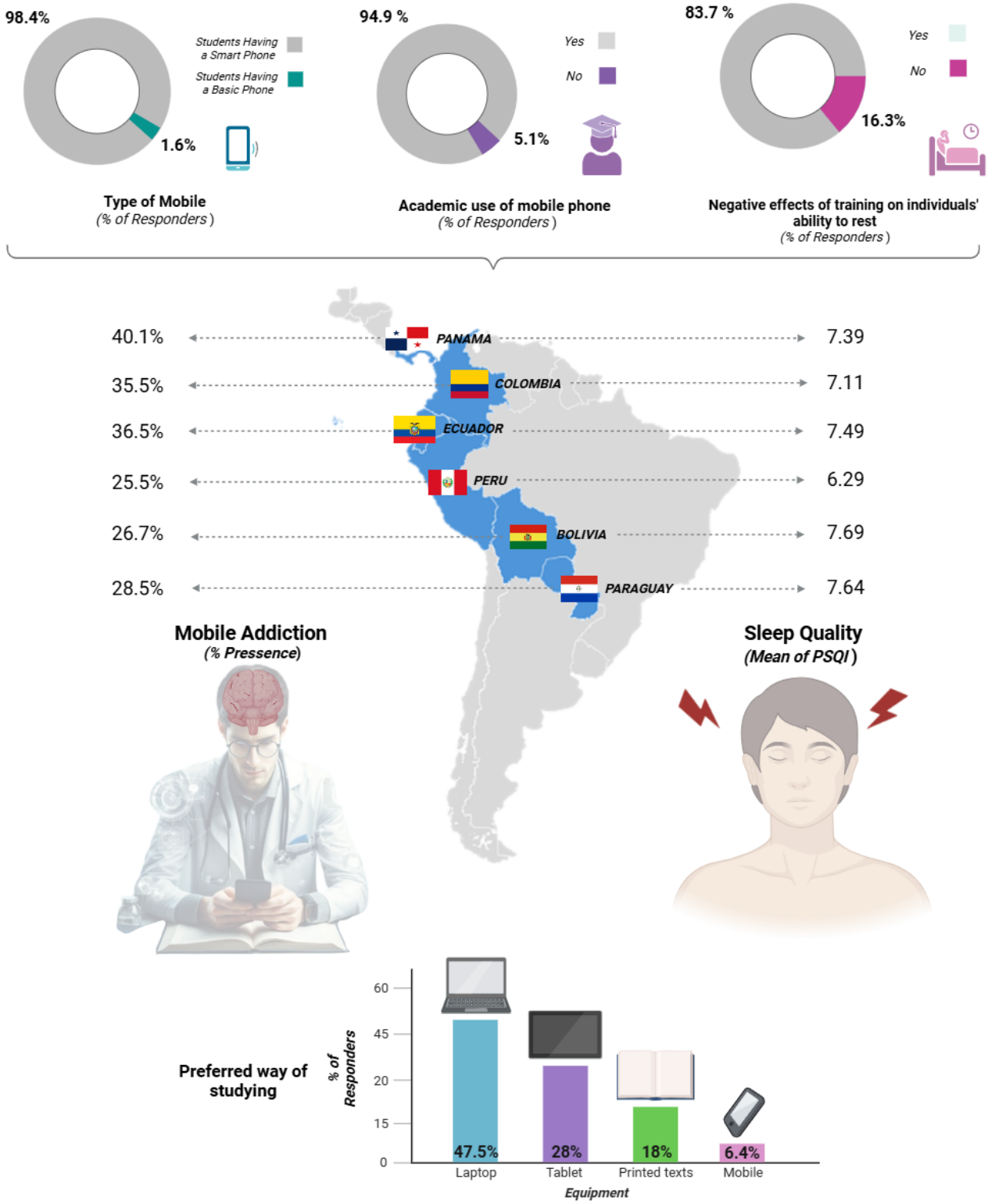

Mobile Phone Addiction

According to the Mobile Phone Addiction Scale, 32.5% (545/1677) of participants exhibited mobile phone addiction. The domain most affected was tolerance to use (974/1677, 58.1%), followed by an intense desire to use the phone (708/1677, 42.2%). A decrease in pleasure was the least

XSL•FO
RenderX

common symptom, noted in only 9.8% (165/1677) of the sample (Table S2 in Multimedia Appendix 1).

According to the distribution among the participating countries, the highest prevalence of mobile phone addiction was observed in medical students from Panama (121/302, 40.1%), while the lowest was in Peru (70/253, 25.5%; *P*<.001; Figure 2). Furthermore, the presence of mobile phone addiction was associated with certain participant characteristics, being significantly higher (38/86, 44.2%) among medical students who stated they did not use their phones for academic purposes (*P*=.02), as well as among those who reported that their undergraduate preparation negatively affected their rest (448/1251, 35.8%; *P*<.001). In addition, medical students who preferred printed texts had the lowest percentage of mobile phone addiction (81/302, 22.8%) when compared with those participants using screen devices (mobile phone, computer, and tablet; *P*=.002; Table 2).

Table 2. Relationship between the characteristics of the participants with mobile phone addiction and sleep quality.

Characteristics	Mobile addiction			PSQI[a]	
	Absence, n (%)	Presence, n (%)	*P* value	Mean (SD)	*P* value
Sex			.92		.02
Male (n=594)	400 (67.3)	194 (32.7)		7.04 (3.34)	
Female (n=1083)	732 (67.6)	351 (32.4)		7.39 (3.5)	
Level of education			.85		.22
Initial (n=973)	655 (67.3)	318 (32.7)		7.37 (3.41)	
Advanced (n=704)	477 (67.8)	227 (32.2)		7.11 (3.5)	
Type of university			.58		.001
Private university (n=967)	658 (68)	309 (32)		7.57 (3.48)	
Public university (n=710)	474 (66.8)	236 (33.2)		6.85 (3.37)	
Academic mobile phone use			.02		.72
No (n=86)	48 (55.8)	38 (44.2)		7.40 (3.5)	
Yes (n=1591)	1084 (68.1)	507 (31.9)		7.26 (3.45)	
Preferred way of studying			.002		.02
Mobile phone (n=108)	61 (56.5)	47 (43.5)		8.21 (3.71)	
Computer/laptop (n=797)	551 (69.1)	246 (30.9)		7.13 (3.41)	
Tablet/iPad (n=470)	299 (63.6)	171 (36.4)		7.22 (3.47)	
Prefer printed texts (n=302)	221 (73.2)	81 (22.8)		7.34 (3.4)	
Negative effects of your academic training on your rest			<.001		<.001
No (n=426)	329 (77.2)	97 (22.8)		6.08 (3.63)	
Yes (1251)	803 (64.2)	448 (35.8)		7.49 (3.37)	

[a]PSQI: Pittsburgh Sleep Quality Index.

Sleep Quality

As measured by the PSQI, a general mean of poor sleep quality of 7.26 (SD 3.45) was found among medical students from the 6 study countries. The components with the highest scores (indicating the worst sleep quality) were subjective sleep quality (1.33, SD 0.77) and daytime dysfunction (1.31, SD 0.84), while the lowest scores were found in the use of sleep medication (0.27, SD 0.70) and habitual sleep efficiency (0.89, SD 1.12; Table 3).

XSL•FO
RenderX

Table 3. Characteristics of the components of sleep quality in medical students in Latin America.

PSQI[a] values	Bolivia (n=251), mean (SD)	Colombia (n=259), mean (SD)	Ecuador (n=337), mean (SD)	Panama (n=302), mean (SD)	Paraguay (n=253), mean (SD)	Peru (n=275), mean (SD)	Overall (n=1677), mean (SD)
PSQI	7.69 (3.61)	7.11 (3.43)	7.47 (3.31)	7.39 (3.49)	7.64 (3.52)	6.29 (3.22)	7.26 (3.45)
Sleep quality components							
Subjective sleep quality	1.40 (0.87)	1.31 (0.75)	1.31 (0.69)	1.42 (0.74)	1.38 (0.8)	1.13 (0.75)	1.33 (0.77)
Sleep latency	1.15 (0.91)	1.20 (0.97)	1.11 (0.86)	1.08 (0.97)	1.39 (0.97)	0.996 (0.88)	1.15 (0.93)
Sleep duration	1.33 (1.13)	1.09 (1.05)	1.23 (1.04)	1.23 (1.09)	1.01 (1.07)	0.924 (1.01)	1.14 (1.07)
Habitual sleep efficiency	0.95 (1.16)	0.76 (1.05)	0.88 (1.09)	0.96 (1.16)	0.97 (1.15)	0.79 (1.07)	0.89 (1.12)
Sleep disturbance	1.31 (0.66)	1.18 (0.54)	1.26 (0.57)	1.04 (0.5)	1.21 (0.49)	1.16 (0.55)	1.19 (0.56)
Use of sleeping medication	0.25 (0.6)	0.29 (0.73)	0.27 (0.72)	0.25 (0.68)	0.38 (0.86)	0.17 (0.55)	0.27 (0.7)
Daytime dysfunction	1.29 (0.9)	1.28 (0.78)	1.41 (0.84)	1.43 (0.86)	1.28 (0.79)	1.11 (0.81)	1.31 (0.84)

[a]PSQI: Pittsburgh Sleep Quality Index.

Differences in sleep quality by country showed the highest scores for Bolivian (7.69, SD 3.61) and Paraguayan (7.64, SD 3.52) students, while Peru had the lowest score (6.29, SD 3.22; *P*<.001). Complementary post hoc analyses revealed that Peruvian medical students had significantly lower PSQI scores (better sleep quality) compared with their counterparts from Bolivia (*P*<.001), Ecuador (*P*<.001), Panama (*P*=.001), and Paraguay (*P*<.001; Figure 2).

Furthermore, characteristics associated with higher PSQI scores (worse sleep quality) were found in female medical students (7.39, SD 3.5; *P*=.02), those from private universities (7.57, SD 3.48; *P*=.001), those who preferentially use mobile phones for studying (8.21, 3.71; *P*=.02), and those who consider that academic preparation affects their sleep quality (7.49, SD 3.37; *P*<.001; Table 2).

Effect of Mobile Phone Addiction on Sleep Quality

The negative influence of mobile phone addiction on sleep quality was statistically significant across all seven components of the PSQI, showing higher mean scores (worse quality) in the group of medical students with mobile phone addiction, including in the dimensions of habitual sleep efficiency and use of sleeping medication (Table 4).

Table 4. Effect of mobile phone addiction on the components of sleep quality of medical students in Latin America.

Components	Mobile addiction		*P* value
	Absence (n=1132), mean (SD)	Presence (n=545), mean (SD)	
Subjective sleep quality	1.27 (0.76)	1.44 (0.76)	<.001
Sleep latency	1.08 (0.9)	1.28 (0.97)	<.001
Sleep duration	1.06 (1.05)	1.29 (1.09)	<.001
Habitual sleep efficiency	0.83 (1.09)	1 (1.17)	.003
Sleep disturbance	1.15 (0.54)	1.29 (0.59)	<.001
Use of sleeping medication	0.23 (0.67)	0.33 (0.77)	.01
Daytime dysfunction	1.19 (0.81)	1.56 (0.84)	<.001

A multiple regression was testes to demonstrated that mobile phone addiction is a factor associated with poor sleep quality as measured by mean PSQI scores even after adjusting for sex, age, and level of studies. The model largely met the assumptions of OLS regression, including linearity, independence, and homoscedasticity. However, the normality test of the residuals did not fully satisfy the assumption of normality. Despite this, we chose to retain the OLS model for several reasons. First, OLS is robust to deviations from normality, particularly when the sample size is large, as was the case in this study. Second, the central limit theorem suggested that the distribution of the residuals would approximate normality with a sufficiently large sample, even if the data were not perfectly normally distributed. Finally, the main goal of this model was to estimate the relationships between variables, and OLS provided unbiased, consistent, and efficient estimates under the assumption of homoscedasticity and linearity, which the model satisfied. Therefore, despite the minor concern regarding normality, the OLS model remained appropriate for the analysis. The model demonstrated the relationship between mobile phone usage and sleep quality among medical students in the study (β=1.4, 95% CI 1.05-1.74; Table 5).

XSL•FO
RenderX

Table 5. Regression analysis of mobile phone addiction impact on sleep disturbance (by Pittsburgh Sleep Quality Index mean scores). Analyses of sleep disturbances predicted by cell phone were controlled for gender, age and educational level (R2=0.06).

Variables	Beta values (95% CI)
Mobile phone addiction	1.40 (1.05 to 1.74)[a]
Sex (female)	0.40 (0.06 to 0.74)[b]
Age	0.32 (0.13 to 0.51)[a]
Level of education (advanced)	–0.46 (–0.85 to –0.06)[b]
Country	
Colombia	–0.62 (–1.23 to –0.02)[b]
Ecuador	–0.31 (–0.88 to –0.26)
Panama	–0.42 (–1.02 to –0.17)
Paraguay	–0.17 (–0.79 to –0.44)
Peru	–1.41 (–2 to –0.81)[a]

[a]*P*<.001.

[b]*P*<.05.

Discussion

Principal Findings

By the third decade of the 21st century, access to and use of mobile phones has become nearly universal, a reality that includes medical students in Latin America [24]. This research aimed to explore mobile phone addiction among medical students across 6 Latin American countries, as well as to evaluate their sleep quality and the potential impact of excessive and unhealthy mobile phone use on sleep.

This multicenter study included 1677 undergraduate medical students, revealing a notable predominance of female participants. This finding aligns with recent research indicating that a majority of medical students in Latin America are female [25,26]. Nearly 100% of participants reported using a smartphone, a rate comparable with those reported among medical students in other regions [27]. In addition, 94.9% of participants used their devices for academic purposes, a figure higher than in other regions but consistent with global trends favoring the academic use of mobile phones among undergraduate medical students [12,28,29].

Although smartphones are primarily associated with barrier-free communication through social networks, they also possess numerous tools that are increasingly integrated into academic training, particularly among medical students. However, prolonged and uncontrolled use of these devices can have significant health effects on users [30-32]. Medical students are uniquely affected by mobile phone addiction and poor sleep quality due to the combination of demanding academic schedules, high stress levels, and the pervasive nature of digital technology. The rigorous nature of medical training often necessitates extensive study hours, leading students to rely heavily on smartphones for quick access to educational materials. This reliance, coupled with the need for social interaction and recreation through the same devices, often results in prolonged and uncontrolled usage.

Despite being regarded as a tool for academic enhancement, approximately one-third (32.5%) of the study population exhibited mobile phone addiction, with rates ranging from 25.5% among Peruvian students to 40.1% among Panamanian students. Comparing addiction rates globally is challenging due to variations in the questionnaires used to assess mobile phone use, addiction, or overuse, such as the Smartphone Addiction Scale (SAS-SV) or the Cell Phone Overuse Scale (COS) [33]. For example, Kamal et al [34] reported a 48% addiction rate among Pakistani medical and dental students using a similar instrument, while other studies found addiction rates ranging from 10.7% among Iranian medical students [35] to 34.4% among first and second-year medical students in Srinagar, India [36]. The questionnaire used in the present study [21] found that addiction was most commonly characterized by tolerance (using the phone more than expected and considering the usage excessive) and a strong desire to use the phone, particularly for immediate interaction on social networks or phone use even when others are nearby, traits typical of addictive behaviors [37]. Despite the heterogeneity in assessment tools, the body of research underscores the pervasive nature of mobile phone addiction among medical students worldwide, including in Latin America.

Interestingly, this study found no significant demographic variables related to overuse or addiction to mobile phones, such as gender [38,39]. However, there was a concerning relationship between students who did not use mobile phones for academic purposes and higher rates of addiction, potentially due to a preference for leisure activities such as social networking, gaming, or streaming platforms [40]. In addition, students who reported that their studies negatively affected their rest exhibited higher rates of addiction, suggesting an indirect relationship between mobile phone addiction and poor sleep. This notion is supported by findings showing that students who preferred printed texts for studying had lower addiction rates.

Sleep quality studies have become more standardized with the widespread acceptance of the PSQI. The average PSQI score

XSL•FO
RenderX

among students in this study (7.26, SD 3.45) indicated poor sleep quality, considerably higher than scores reported in similar studies, including military medical students (PSQI score=5.78, SD 2.26) [41], Indian medical students (PSQI score=4.43, SD 2.62) [39], and Iranian medical students (PSQI score=5.38, SD 2.31). Systematic reviews report meta-mean PSQI scores between 5.95 and 6.1 [33,42]. Poor sleep quality among medical students is well-documented and has been attributed to heavy study loads, irregular schedules, and a preference for study over sleep [43,44]. In this study, students from private universities and women exhibited worse sleep quality, a trend previously observed in other research [41].

There was a significant and consistent association between mobile phone addiction and poor sleep quality, with addicted students scoring higher across all PSQI dimensions (subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction) compared with those without addiction (P<.05). Regression analyses revealed a global association between mobile phone addiction and diminished sleep quality, even after adjusting for sex, age, and educational level (β=1.4; 95% CI 1.05-1.74; P<.001). This association has been widely evidenced in previous studies of medical students across different countries and academic levels [33,35,38,39,41,45].

The findings of this study offer several plausible explanations for the observed association between mobile phone addiction and poor sleep quality. Smartphones facilitate a wide range of activities, including gaming, streaming, and social media use, all of which compete for users' limited time and attention, often displacing sleep. Social media platforms, in particular, are deliberately designed to capture and retain user engagement through features such as "infinite scroll," notifications, direct messaging (DMs), and interaction metrics like likes, follows, and shares [46,47]. These features create an attention-driven ecosystem where users are incentivized to remain active, with their engagement monetized for advertising purposes [46].

In addition, the blue light emitted by smartphones has been widely recognized as a disruptor of circadian rhythms [48-50]. Artificial light exposure, particularly at night, delays the release of melatonin, the hormone responsible for regulating sleep-wake cycles, exacerbating sleep disturbances. Among medical students, this effect is particularly concerning due to their already irregular schedules and high stress levels [51-53]. The combination of attention-capturing smartphone apps and blue light exposure likely contributes to prioritizing phone use oversleep, providing a compelling explanation for the current study's findings.

This study highlights a critical issue, emphasizing the dual impact of mobile phones on the present and future of academic medical training [54]. While these devices provide significant benefits, including portable and digital access to textbooks, scientific articles, and practical tools such as video content, infographics, and podcasts [24,55-57], heir use also presents risks. These resources can enhance knowledge acquisition, even in resource-limited settings like Latin America. However, decision-makers, including university institutions and educators, must be cognizant of the potential negative consequences of mobile phone use in academic environments. These include cross-device addiction, indirect effects on sleep quality, and other physical and mental health challenges.

The broader implications of these findings extend beyond sleep disturbances. Poor sleep quality and mobile phone addiction can significantly impair cognitive function, emotional regulation, and physical health, which are vital for medical students' academic success and future professional practice. Sleep deprivation and distractions caused by addiction may hinder students' ability to learn effectively, retain critical information, and perform competently in clinical settings, ultimately impacting patient care.

Addressing these challenges requires urgent intervention. Institutions should integrate sleep education into medical curricula, emphasizing its importance for cognitive performance and overall well-being. Behavioral strategies, such as implementing "digital detox" periods or encouraging the use of apps that monitor and limit screen time, can help students manage their smartphone use. Furthermore, counseling and support services tailored to stress management and time optimization can address the underlying causes of overuse and improve sleep hygiene.

Limitations

This study has several limitations that should be acknowledged. First, the survey was conducted online, which may have introduced a sampling bias, favoring medical students with easier access to technology or greater familiarity with online platforms. In addition, the reliance on self-reported data poses the risk of social desirability bias, where participants may exaggerate their academic use of smartphones or provide responses they perceive as socially acceptable.

The subjective sleep assessments used in this study, specifically the PSQI, are practical, cost-effective, and widely used in population-based research due to their ease of administration and scalability. However, subjective measures are inherently limited by recall bias and participants' ability to accurately evaluate their sleep habits. In contrast, objective sleep assessment methods, such as actigraphy and polysomnography, provide direct and measurable data on sleep parameters and are considered more precise. These objective tools often reveal discrepancies compared with subjective assessments, as personal perceptions may lead to overestimation or underestimation of sleep duration or quality. Nevertheless, objective methods may fail to account for psychological factors influencing perceived sleep quality, which subjective tools can capture.

In this study, the reliance solely on subjective measures, such as the PSQI, represents a limitation. This approach may not fully reflect the complexity of sleep patterns or detect potential clinical sleep disturbances. Future research should consider integrating objective assessments to complement and validate self-reported sleep data, providing a more comprehensive understanding of sleep behaviors. Furthermore, the cross-sectional design of this study precludes the establishment of causality between mobile phone addiction and poor sleep quality. It remains unclear whether academic mobile phone use contributes to addiction and reduced sleep quality, or if

XSL•FO
RenderX

academically successful students inherently use smartphones more frequently and experience shorter sleep durations.

Finally, while the study included participants from 6 Latin American countries, the findings may not be universally generalizable across the entire region due to cultural, economic, and educational differences. Future studies should aim to include more diverse populations to enhance the external validity of the results.

Conclusions

In conclusion, this study reveals a substantial prevalence of mobile phone addiction among medical students from 6 Latin American countries, affecting approximately one-third of the sample. In addition, the overall sleep quality of student participants was found to be poor, as indicated by a mean PSQI score higher than previously reported. Findings suggest that using mobile phones for studying can indirectly contribute to poorer sleep quality by diverting attention to other activities during study sessions. This underscores the dual nature of mobile technology, which can serve both educational and recreational purposes, each with differing impacts on the lives and health of students. Importantly, we identified a significant association between mobile phone addiction and poor sleep quality. Overall, this study confirms the widespread academic integration of mobile phones among medical students in Latin America and highlights the significant need for strategies to address the use of these devices in academia and their potential impact on health and academic success.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 18 KB - jmir_v27i1e60630_app1.docx]

References

1. Forecast number of mobile devices worldwide from 2020 to 2025 (in billions)*.: Statista; 2024. URL: https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/ [accessed 2024-04-28]
2. Morgan C. Draw Your Phone: The Cellphone as an Intimate, Everyday Artefact. Newland: Teaching and Learning the Archaeology of the Contemporary Era Bloomsbury Publishing; 2023:101.
3. De-Sola Gutiérrez J, Rodríguez de Fonseca F, Rubio G. Cell-phone addiction: a review. Front Psychiatry 2016;7:175 [FREE Full text] [doi: 10.3389/fpsyt.2016.00175] [Medline: 27822187]
4. Naeem Z. Health risks associated with mobile phones use. Int J Health Sci (Qassim) 2014;8(4):V-VI [FREE Full text] [Medline: 25780365]
5. Zirek E, Mustafaoglu R, Yasaci Z, Griffiths MD. A systematic review of musculoskeletal complaints, symptoms, and pathologies related to mobile phone usage. Musculoskelet Sci Pract 2020;49:102196. [doi: 10.1016/j.msksp.2020.102196] [Medline: 32861360]
6. Westerman R, Hocking B. Diseases of modern living: neurological changes associated with mobile phones and radiofrequency radiation in humans. Neurosci Lett 2004;361(1-3):13-16. [doi: 10.1016/j.neulet.2003.12.028] [Medline: 15135881]
7. Exelmans L, Van den Bulck J. Bedtime mobile phone use and sleep in adults. Soc Sci Med 2016;148:93-101. [doi: 10.1016/j.socscimed.2015.11.037] [Medline: 26688552]
8. White AG, Buboltz W, Igou F. Mobile phone use and sleep quality and length in college students. International Journal of Humanities and Social Science 2011;1(18):51-58.
9. Ai S, Ye S, Li G, Leng Y, Stone KL, Zhang M, et al. Association of disrupted delta wave activity during sleep with long-term cardiovascular disease and mortality. J Am Coll Cardiol 2024;83(17):1671-1684. [doi: 10.1016/j.jacc.2024.02.040] [Medline: 38573282]
10. Addo PNO, Mundagowa PT, Zhao L, Kanyangarara M, Brown MJ, Liu J. Associations between sleep duration, sleep disturbance and cardiovascular disease biomarkers among adults in the United States. BMC Public Health 2024;24(1):947 [FREE Full text] [doi: 10.1186/s12889-024-18381-5] [Medline: 38566084]
11. Almutairi H, Alsubaiei A, Abduljawad S, Alshatti A, Fekih-Romdhane F, Husni M, et al. Prevalence of burnout in medical students: a systematic review and meta-analysis. Int J Soc Psychiatry 2022;68(6):1157-1170. [doi: 10.1177/00207640221106691] [Medline: 35775726]
12. Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, et al. Perceptions of digital health education among european medical students: mixed methods survey. J Med Internet Res 2020;22(8):e19827 [FREE Full text] [doi: 10.2196/19827] [Medline: 32667899]
13. Li S, Feng N, Cui L. Network analysis of social anxiety and problematic mobile phone use in Chinese adolescents: a longitudinal study. Addict Behav 2024;155:108026. [doi: 10.1016/j.addbeh.2024.108026] [Medline: 38593598]
14. Seoane HA, Moschetto L, Orliacq F, Orliacq J, Serrano E, Cazenave MI, et al. Sleep disruption in medicine students and its relationship with impaired academic performance: a systematic review and meta-analysis. Sleep Med Rev 2020;53:101333. [doi: 10.1016/j.smrv.2020.101333] [Medline: 32485517]

XSL•FO
RenderX

15. Rozgonjuk D, Saal K, Täht K. Problematic smartphone use, deep and surface approaches to learning, and social media use in lectures. Int J Environ Res Public Health 2018;15(1):92 [FREE Full text] [doi: 10.3390/ijerph15010092] [Medline: 29316697]
16. Thapa K, Lama S, Pokharel R, Sigdel R, Rimal SP. Mobile phone dependence among undergraduate students of a medical college of eastern Nepal: a descriptive cross-sectional study. JNMA J Nepal Med Assoc 2020;58(224):234-239 [FREE Full text] [doi: 10.31729/jnma.4787] [Medline: 32417860]
17. Mengi A, Singh A, Gupta V. An institution-based study to assess the prevalence of Nomophobia and its related impact among medical students in Southern Haryana, India. J Family Med Prim Care 2020;9(5):2303-2308 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_58_20] [Medline: 32754492]
18. Copaja-Corzo C, Aragón-Ayala CJ, Taype-Rondan A, Nomotest-Group N. Nomophobia and its associated factors in peruvian medical students. Int J Environ Res Public Health 2022;19(9):5006 [FREE Full text] [doi: 10.3390/ijerph19095006] [Medline: 35564400]
19. Kaya F, Bostanci Daştan N, Durar E. Smart phone usage, sleep quality and depression in university students. Int J Soc Psychiatry 2021;67(5):407-414. [doi: 10.1177/0020764020960207] [Medline: 32969293]
20. Herrell C, Foster S. Can't stop won't stop: problematic phone use, sleep quality, and mental health in U.S. graduate students. J Am Coll Health 2024;28:1-7. [doi: 10.1080/07448481.2024.2334068] [Medline: 38546702]
21. Basu S, Garg S, Singh MM, Kohli C. Addiction-like behavior associated with mobile phone usage among medical students in Delhi. Indian J Psychol Med 2018;40(5):446-451 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_59_18] [Medline: 30275620]
22. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
23. Hita-Contreras F, Martínez-López E, Latorre-Román PA, Garrido F, Santos MA, Martínez-Amat A. Reliability and validity of the Spanish version of the pittsburgh sleep quality index (PSQI) in patients with fibromyalgia. Rheumatol Int 2014;34(7):929-936. [doi: 10.1007/s00296-014-2960-z] [Medline: 24509897]
24. Izquierdo-Condoy JS, Arias-Intriago M, Nati-Castillo HA, Gollini-Mihalopoulos R, Cardozo-Espínola CD, Loaiza-Guevara V, et al. Exploring smartphone use and its applicability in academic training of medical students in Latin America: a multicenter cross-sectional study. BMC Med Educ 2024;24(1):1401. [doi: 10.1186/s12909-024-06334-w] [Medline: 39616324]
25. Izquierdo-Condoy JS, Simbaña-Rivera K, Nati-Castillo HA, Cassa Macedo A, Cardozo Espínola CD, Vidal Barazorda GM, et al. How much do Latin American medical students know about radiology? Latin-American multicenter cross-sectional study. Med Educ Online 2023;28(1):2173044 [FREE Full text] [doi: 10.1080/10872981.2023.2173044] [Medline: 36722612]
26. Izquierdo-Condoy JS, Montiel-Alfonso MA, Nati-Castillo HA, Saucedo R, Jaramillo-Aguilar DS, Nanjari-Barrientos C, et al. Knowledge, perceptions, and practices on risks and disasters among medical students. A multicenter cross-sectional study in 9 Latin American and Caribbean countries. Adv Med Educ Pract 2023;14:225-235 [FREE Full text] [doi: 10.2147/AMEP.S394331] [Medline: 36941904]
27. Baticulon RE, Sy JJ, Alberto NRI, Baron MBC, Mabulay REC, Rizada LGT, et al. Barriers to online learning in the time of COVID-19: A national survey of medical students in the Philippines. Med Sci Educ 2021;31(2):615-626 [FREE Full text] [doi: 10.1007/s40670-021-01231-z] [Medline: 33649712]
28. Sharma N, Advani U, Sharma L, Jain M, Sharma K, Dixit A. Pattern of mobile phone usage among medical students. Int J Acad Med 2019;5(2):118. [doi: 10.4103/ijam.ijam_61_18]
29. Singh K, Sarkar S, Gaur U, Gupta S, Adams OP, Sa B, et al. Smartphones and educational apps use among medical students of a smart university campus. Front Commun 2021;6. [doi: 10.3389/fcomm.2021.649102]
30. Chase TJG, Julius A, Chandan JS, Powell E, Hall CS, Phillips BL, et al. Mobile learning in medicine: an evaluation of attitudes and behaviours of medical students. BMC Med Educ 2018;18(1):152 [FREE Full text] [doi: 10.1186/s12909-018-1264-5] [Medline: 29945579]
31. Montagni I, Donisi V, Tedeschi F, Parizot I, Motrico E, Horgan A. Internet use for mental health information and support among European university students: The e-MentH project. Digit Health 2016;2:2055207616653845 [FREE Full text] [doi: 10.1177/2055207616653845] [Medline: 29942558]
32. Ratan ZA, Parrish A, Zaman SB, Alotaibi MS, Hosseinzadeh H. Smartphone addiction and associated health outcomes in adult populations: a systematic review. Int J Environ Res Public Health 2021;18(22):12257 [FREE Full text] [doi: 10.3390/ijerph182212257] [Medline: 34832011]
33. Leow MQH, Chiang J, Chua TJX, Wang S, Tan NC. The relationship between smartphone addiction and sleep among medical students: a systematic review and meta-analysis. PLoS One 2023;18(9):e0290724 [FREE Full text] [doi: 10.1371/journal.pone.0290724] [Medline: 37713408]
34. Kamal S, Kamal S, Mubeen SM, Shah AM, Samar SS, Zehra R, et al. Smartphone addiction and its associated behaviors among medical and dental students in Pakistan: a cross-sectional survey. J Educ Health Promot 2022;11:220 [FREE Full text] [doi: 10.4103/jehp.jehp_494_21] [Medline: 36177415]

XSL•FO
RenderX

35. Mohammadbeigi A, Absari R, Valizadeh F, Saadati M, Sharifimoghadam S, Ahmadi A, et al. Sleep quality in medical students; the impact of over-use of mobile cell-phone and social networks. J Res Health Sci 2016;16(1):46-50 [FREE Full text] [Medline: 27061997]
36. Nowreen N, Ahad F. Effect of smartphone usage on quality of sleep in medical students. Natl J Physiol Pharm Pharmacol 2018;8(10):1366. [doi: 10.5455/njppp.2018.8.0620009062018]
37. Shoukat S. Cell phone addiction and psychological and physiological health in adolescents. EXCLI J 2019;18:47-50 [FREE Full text] [Medline: 30956638]
38. Nikolic A, Bukurov B, Kocic I, Vukovic M, Ladjevic N, Vrhovac M, et al. Smartphone addiction, sleep quality, depression, anxiety, and stress among medical students. Front Public Health 2023;11:1252371 [FREE Full text] [doi: 10.3389/fpubh.2023.1252371] [Medline: 37744504]
39. Divya YL, Dampetla S, Manjulatha K, Manikanta, Srikanth. Smart phone use and its effect on quality of sleep in medical students. Int J Acad Med Pharm 2023;5(6):213-217. [doi: 10.47009/jamp.2023.5.6.43]
40. Yu S, Sussman S. Does smartphone addiction fall on a continuum of addictive behaviors? Int J Environ Res Public Health 2020;17(2):422 [FREE Full text] [doi: 10.3390/ijerph17020422] [Medline: 31936316]
41. Xu JZ. Correlation study on mobile phone addiction and sleep quality of military medical students. Academic Journal of Second Military Medical University 2019;12:1389-1392.
42. Rao WW, Li W, Qi H, Hong L, Chen C, Li C, et al. Sleep quality in medical students: a comprehensive meta-analysis of observational studies. Sleep Breath 2020;24(3):1151-1165. [doi: 10.1007/s11325-020-02020-5] [Medline: 32072469]
43. Preišegolavičiūtė E, Leskauskas D, Adomaitienė V. Associations of quality of sleep with lifestyle factors and profile of studies among Lithuanian students. Medicina (Kaunas) 2010;46(7):482-489 [FREE Full text] [Medline: 20966622]
44. Almojali AI, Almalki SA, Alothman AS, Masuadi EM, Alaqeel MK. The prevalence and association of stress with sleep quality among medical students. J Epidemiol Glob Health 2017;7(3):169-174 [FREE Full text] [doi: 10.1016/j.jegh.2017.04.005] [Medline: 28756825]
45. Boonluksiri P. Effect of smartphone overuse on sleep problems in medical students. The Asia Pacific Scholar 2018;3(2):25-28. [doi: 10.29060/taps.2018-3-2/oa1039]
46. Bhargava HK. Infinite Scroll: Addiction by Design in Information Platforms.: School of Management University of California Davis; 2023. URL: https://www.teis-workshop.org/papers/2023/Digital_Addiction_and_Advertising.pdf [accessed 2025-01-07]
47. Rixen JO, Meinhardt LM, Glöckler M, Ziegenbein ML, Schlothauer A, Colley M, et al. The loop and reasons to break it: investigating infinite scrolling behaviour in social media applications and reasons to stop. Proc ACM Hum-Comput Interact 2023;7(MHCI):228. [doi: 10.1145/3604275]
48. Oh JH, Yoo H, Park HK, Do YR. Analysis of circadian properties and healthy levels of blue light from smartphones at night. Sci Rep 2015;5:11325 [FREE Full text] [doi: 10.1038/srep11325] [Medline: 26085126]
49. Wahl S, Engelhardt M, Schaupp P, Lappe C, Ivanov IV. The inner clock-Blue light sets the human rhythm. J Biophotonics 2019;12(12):e201900102 [FREE Full text] [doi: 10.1002/jbio.201900102] [Medline: 31433569]
50. West KE, Jablonski MR, Warfield B, Cecil KS, James M, Ayers MA, et al. Blue light from light-emitting diodes elicits a dose-dependent suppression of melatonin in humans. J Appl Physiol (1985) 2011;110(3):619-626 [FREE Full text] [doi: 10.1152/japplphysiol.01413.2009] [Medline: 21164152]
51. Belsare V, Tekade M, Munghate SC, Belsare H. Effect of sleep and smart phone on serum melatonin level in first year medical students. Int J Adv Med 2020;7(7):1121. [doi: 10.18203/2349-3933.ijam20202587]
52. Randjelović P, Stojiljković N, Radulović N, Ilić I, Stojanović N, Ilić S. The association of smartphone usage with subjective sleep quality and daytime sleepiness among medical students. Biological Rhythm Research 2018;50(6):857-865. [doi: 10.1080/09291016.2018.1499374]
53. Shrivastava A, Saxena Y. Effect of mobile usage on serum melatonin levels among medical students. Indian J Physiol Pharmacol 2014;58(4):395-399. [Medline: 26215007]
54. Gavali MY, Khismatrao DS, Gavali YV, Patil KB. Smartphone, the new learning aid amongst medical students. J Clin Diagn Res 2017;11(5):JC05-JC08 [FREE Full text] [doi: 10.7860/JCDR/2017/20948.9826] [Medline: 28658804]
55. Spicer JO, Coleman CG. Creating effective infographics and visual abstracts to disseminate research and facilitate medical education on social media. Clin Infect Dis 2022;74(Suppl_3):e14-e22. [doi: 10.1093/cid/ciac058] [Medline: 35568482]
56. Krumm IR, Miles MC, Clay A, Carlos Ii WG, Adamson R. Making effective educational videos for clinical teaching. Chest 2022;161(3):764-772 [FREE Full text] [doi: 10.1016/j.chest.2021.09.015] [Medline: 34587482]
57. Franz A, Oberst S, Peters H, Berger R, Behrend R. How do medical students learn conceptual knowledge? High-, moderate- and low-utility learning techniques and perceived learning difficulties. BMC Med Educ 2022;22(1):250 [FREE Full text] [doi: 10.1186/s12909-022-03283-0] [Medline: 35387637]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COS: Cell Phone Overuse Scale

XSL•FO
RenderX

DM: direct message
Gen Z: generation Z
ICD-10: International Statistical Classification of Diseases, Tenth Revision
OLS: ordinary least squares
PSQI: Pittsburgh Sleep Quality Index
SAS-SV: Smartphone Addiction Scale-Short Version

Edited by S He; submitted 16.05.24; peer-reviewed by S Ghazi, J Bartwal; comments to author 14.10.24; revised version received 13.12.24; accepted 14.12.24; published 10.02.25.

Please cite as:
Izquierdo-Condoy JS, Paz C, Nati-Castillo HA, Gollini-Mihalopoulos R, Aveiro-Róbalo TR, Valeriano Paucar JR, Laura Mamami SE, Caicedo JF, Loaiza-Guevara V, Mejía DC, Salazar-Santoliva C, Villavicencio-Gomezjurado M, Hall C, Ortiz-Prado E
Impact of Mobile Phone Usage on Sleep Quality Among Medical Students Across Latin America: Multicenter Cross-Sectional Study
J Med Internet Res 2025;27:e60630
URL: https://www.jmir.org/2025/1/e60630
doi:10.2196/60630
PMID:

©Juan S Izquierdo-Condoy, Clara Paz, H A Nati-Castillo, Ricardo Gollini-Mihalopoulos, Telmo Raul Aveiro-Róbalo, Jhino Renson Valeriano Paucar, Sandra Erika Laura Mamami, Juan Felipe Caicedo, Valentina Loaiza-Guevara, Diana Camila Mejía, Camila Salazar-Santoliva, Melissa Villavicencio-Gomezjurado, Cougar Hall, Esteban Ortiz-Prado. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A New Computer-Based Cognitive Measure for Early Detection of Dementia Risk (Japan Cognitive Function Test): Validation Study

Hiroyuki Shimada[1], MSc, PhD; Takehiko Doi[1], MSc, PhD; Kota Tsutsumimoto[1], MSc, PhD; Keitaro Makino[1], MSc, PhD; Kenji Harada[1], MSc; Kouki Tomida[1], MSc, PhD; Masanori Morikawa[1], MSc, PhD; Hyuma Makizako[1,2], MSc, PhD

[1]Department of Preventive Gerontology, Centre for Gerontology and Social Science, National Center for Geriatrics and Gerontology, Ōbu, Japan
[2]Department of Physical Therapy, School of Health Sciences, Faculty of Medicine, Kagoshima University, Kagoshima, Japan

Corresponding Author:
Hiroyuki Shimada, MSc, PhD
Department of Preventive Gerontology, Centre for Gerontology and Social Science
National Center for Geriatrics and Gerontology
7-430, Morioka-cho
Ōbu, 474-8511
Japan
Phone: 81 0562 46 2311
Email: shimada@ncgg.go.jp

Abstract

Background: The emergence of disease-modifying treatment options for Alzheimer disease is creating a paradigm shift in strategies to identify patients with mild symptoms in primary care settings. Systematic reviews on digital cognitive tests reported that most showed diagnostic performance comparable with that of paper-and-pencil tests for mild cognitive impairment and dementia. However, most studies have small sample sizes, with fewer than 100 individuals, and are based on case-control or cross-sectional designs.

Objective: This study aimed to examine the predictive validity of the Japanese Cognitive Function Test (J-Cog), a new computerized cognitive battery test, for dementia development.

Methods: We randomly assigned 2520 older adults (average age 72.7, SD 6.7 years) to derivation and validation groups to determine and validate cutoff points for the onset of dementia. The Mini-Mental State Examination (MMSE) was used for comparison purposes. The J-Cog consists of 12 tasks that assess orientation, designation, attention and calculation, mental rotation, verbal fluency, sentence completion, working memory, logical reasoning, attention, common knowledge, word memory recall, and episodic memory recall. The onset of dementia was monitored for 60 months. In the derivation group, receiver operating characteristic curves were plotted to determine the MMSE and J-Cog cutoff points that best discriminated between the groups with and without dementia. In the validation group, Cox proportional regression models were developed to predict the associations of the group classified using the cutoff points of the J-Cog or MMSE with dementia incidence. Harrell C-statistic was estimated to summarize how well a predicted risk score described an observed sequence of events. The Akaike information criterion was calculated for relative goodness of fit, where lower absolute values indicate a better model fit.

Results: Significant hazard ratios (HRs) for dementia incidence were found using the MMSE cutoff between 23 and 24 point (HR 1.93, 95% CI 1.13-3.27) and the J-Cog cutoff between 43 and 44 points (HR 2.42, 95% CI 1.50-3.93). In the total validation group, the C-statistic was above 0.8 for all cutoff points. Akaike information criterion with MMSE cutoff between 23 and 24 points as a reference showed a poor fit for MMSE cutoff between 28 and 29 points, and a good fit for the J-Cog cutoff between 43 and 44 points.

Conclusions: The J-Cog has higher accuracy in predicting the development of dementia than the MMSE and has advantages for use in the community as a test of cognitive function, which can be administered by nonprofessionals.

(J Med Internet Res 2025;27:e59015) doi:10.2196/59015

XSL•FO
RenderX

KEYWORDS

cognition; neurocognitive test; dementia; Alzheimer disease; aged; MMSE; cognitive impairment; Mini-Mental State Examination; monitoring; eHealth

Introduction

Dementia is a syndrome associated with neurodegenerative diseases, characterized by a general decline in cognitive abilities, which affects daily activities. It ultimately has a significant impact on patients and their careers and social relationships. The most common symptoms of dementia include problems with memory, thinking, behavior, emotions, language, and reduced motivation [1]. The population of Japan is aging more rapidly than that of any other country. As of 2020, the number of Japanese people older than 65 years of age has reached a record high of 36 million, or 29% of the total population of 125 million. The number of people with dementia in Japan is projected to reach 7 million by 2025. To address this issue, the Basic Act for Dementia to Promote an Inclusive Society was approved on June 14, 2023, and took effect on January 1, 2024. This law requires national and local governments to ensure the social participation of people with dementia, develop a consultation system, and promote research for the development of a symbiotic society so that people with dementia can live with peace of mind and protect their human rights [2].

The Basic Act for Dementia to Promote an Inclusive Society includes measures to promote early detection, diagnosis, and response to dementia regarding prevention strategies. Early identification through screening would also allow patients and their families to receive care at an earlier stage in the disease process, potentially facilitating discussions regarding decision-making (eg, health care, financial, and legal), while the patient still has decision-making capacity [3]. Several progressive local governments have already initiated screening programs for the early detection of dementia risk. However, the cognitive function tests used as screening methods are Mini-Mental State Examination (MMSE) or similar conventional paper-and-pencil tests. The delay in digitizing cognitive tests might limit the number of older adults who are eligible for primary screening. A meta-analysis reported a high rate of undetected dementia, particularly in populations with low socioeconomic status [4]. This is likely due to the lack of widespread implementation of appropriate cognitive screening tests. Furthermore, the meta-analysis results suggested that studies that used the *Diagnostic and Statistical Manual of Mental Disorders, Third andFourth Editions* criteria to identify dementia had a higher rate of underdetection than those that used the MMSE [4].

Lecanemab (Leqembi), a treatment agent for early Alzheimer disease (AD), can significantly alter the disease course for patients in the early stages [5]. Among participants with early symptomatic AD and amyloid and tau pathology, donanemab significantly slowed clinical progression in those with low or medium tau and in the combined low or medium and high tau pathology population [6]. The World Health Organization guidelines on risk reduction for cognitive decline and dementia provide evidence-based recommendations regarding lifestyle behaviors and interventions to delay or prevent cognitive decline and dementia [7]. These preventive paradigms create the need to identify patients with mild symptoms, typically in primary care settings, and refer them to AD specialists for formal diagnosis and to determine treatment or prevention eligibility.

Primary screening of cognitive function in the community is required to facilitate early hospital visits for early diagnosis of AD. The involvement of psychologists and other professionals in administering specialized neuropsychological assessments in the community is limited, and a screening test for cognitive function that can be administered by nonprofessionals would improve primary screening. To address these issues, cognitive testing as a digital biomarker that is reliable and self-administered and can detect cognitive dysfunction at an early stage is required. A review and recommendation on routine screening for cognitive impairment in adults aged 65 years and older by the US Preventive Services Task Force asks that clinicians remain alert to early signs of cognitive impairment for individual evaluation [3].

A systematic review on digital cognitive tests reported that most of these tests showed diagnostic performance comparable with that of paper-and-pencil tests for mild cognitive impairment (MCI) and dementia [8]. Furthermore, digital biomarkers related to memory and executive function might be more sensitive than those related to other cognitive domains [9]. These findings suggest that digital cognitive battery tests are promising sensitive clinical tools for detecting MCI and early dementia. However, most studies have small sample sizes, with fewer than 100 individuals, and are based on case-control or cross-sectional designs [9]. Further studies may require the use of digital cognitive biomarkers in long-term follow-up to monitor the conversion from normal cognition to dementia.

We developed a digital cognitive test, the Japanese Cognitive Function Test (J-Cog), which assesses orientation, verbal ability, attention, visuospatial performance, reasoning, and memory, to identify older adults with cognitive impairment in primary settings. The purpose of this study was to evaluate the effectiveness of the J-Cog in predicting the development of dementia and compare the accuracy of the J-Cog with that of the MMSE [10], one of the most widely used tests of cognitive function. If the detective validity of the J-Cog is confirmed, it can be widely used in the community as a cognitive screening tool.

Methods

Participants

In total, 4167 community-dwelling older adults who were recruited from Takahama City, Japan, for the National Center for Geriatric and Gerontology-Study of Geriatric Syndromes (NCGG-SGS) [11], which is a Japanese national cohort study, were enrolled in this study. The inclusion criteria were residence in Takahama and age ≥60 years at the time of the examination (September 2015 to February 2017). The exclusion criteria were

XSL•FO
RenderX

(1) availing the Japanese certified public long-term care insurance system, which requires support or care (n=76); (2) not availing the certified public long-term care insurance system but having the inability to perform basic activities of daily living (such as eating, dressing, bathing, climbing stairs) and maintain functional mobility, personal hygiene, grooming, and toilet hygiene (n=9); (3) history of dementia (n=7) and the following conditions or symptoms suspected of causing dementia: Parkinson disease (n=16), stroke (n=219), or depression (n=97), and a score of <21 on the MMSE [12] (n=67); and (4) missing data for cognitive assessments and other measurements (n=1156). Of the initial 4167 participants, 1647 were excluded, and data from 2520 older adults were analyzed (1017 men and 1503 women). The mean age was 72.7 (SD 6.7; range 60-96) years. Analyses with dementia incidence as an outcome were conducted on 2292 participants who were available for follow-up.

Measurements of Cognitive Function

Overview

Cognitive tests included the J-Cog and MMSE, administered once on the same day by well-trained assessors. The MMSE is commonly used to assess cognitive problems and is often a component of dementia diagnosis (range 0-30, higher scores indicating better functioning) [10].

Japanese Cognitive Function Test

Overview

The J-Cog was developed by neuropsychological researchers at the National Center for Geriatrics and Gerontology as a cognitive test for measuring the domains of cognitive impairment that are generally recognized in MCI. The initial version comprised 15 tasks, which were evaluated in terms of their simplicity, importance, and feasibility. Based on these assessments, the final version, comprising 12 tasks, was developed.

The J-Cog was presented on an iPad (Apple Inc) with a 9.7-inch touch display. The task instructions and questions were presented on the display with a letter size of at least 1.0 cm^2. This battery consists of 12 tasks—orientation, designation, attention and calculation, mental rotation, verbal fluency, sentence completion, working memory, logical reasoning, attention, common knowledge, word memory recall, and episodic memory recall [13]—that assess cognitive functions that tend to decline with age. The participants were given 20-30 minutes to complete the battery (described further in this study). Higher J-Cog scores indicate better cognitive ability (range 0-72). An operator supported each participant in setting up the iPad, understanding the task protocols, and recording their data. The participants only needed to touch the display to complete the tasks using a digital pen. Because the J-Cog is a mixture of examinations in which the display of the screen is made invisible to the participant and examinations in which the display of the screen is shown to the participant, the examinations were conducted by well-trained operators.

Task 1: Orientation

The participants indicated the month, day, and weekday of the test date. The participants also indicated the prefecture and city of their current location. All 5 questions in each task were shown, and we calculated the total number of correct answers (0-5 points).

Task 2: Designation

The participants looked at pictures of eyeglasses, keys, wallets, watches, and cell phones and responded with the name of the object. One point was awarded for each item (0-5).

After naming the objects, they had 15 seconds to memorize the 5 objects they named. The participants then chose 1 of 5 items and answered where it would be stored, that is, under the television, on the dresser, on the shoebox, under the table, or in the dresser. These encoding tasks were scored during memory recall.

Task 3: Attention and Calculation

The participants responded to additions and subtractions of up to 2 digits by rote calculations. There were 6 questions, each presented only once (0-6 points). If there was obvious noise, the questions were presented again.

Task 4: Mental Rotation

The participants saw a completed figure with 3×3 and 4×4 squares showing kana (Japanese alphabet) and completed an incomplete figure with rotated kana missing (0-5 points). First, the participants were asked to answer example questions. If the answer to the example question was incorrect, the correct answer was taught, and the test was conducted after the participants understood the rules. The time limit was 20 seconds for the 3×3 task and 30 seconds for the 4×4 task.

Task 5: Verbal Fluency

The participants were asked to say as many words as possible to fit the category (eg, for vegetables, answer carrots, onions, and lettuce). They had 15 seconds to respond to as many categories as possible (0-20 points).

Task 6: Sentence Completion

The participants responded to the words that applied to the blanks in the sentence from 5 choices (0-2 points). They read a sentence with a missing word and selected the word that fits the blank based on the context of the sentence; 2 types of sentences were presented, each containing 5 choices.

The participants were asked to read the question silently and provide the correct answer. If they did not know the answer, they were asked to select “don’t know.” After each question, the participants were asked to press the next button to complete their full response.

Task 7: Working Memory

Participants were shown images of blue dots. After a while, some of the dots changed to white, then returned to blue. They were then asked to indicate which dots had changed to white. In addition, 2 questions were asked: one with 3×3 dots, and the other with 4×4 dots (0-7 points).

XSL•FO
RenderX

Task 8: Logical Reasoning

The participants responded to a logical reasoning text task and 4 fill-in-the-blank questions (0-4 points). Participants were asked to read the instructional text and question silently and select the correct response according to the instructions.

Task 9: Attention

The participants were shown 2 similar pictures and had 30 seconds to find the 1 shape present in only 1 of the pictures (0-1 points).

Task 10: Common Knowledge

The participants looked at pictures of national flags and were asked to choose the corresponding country from 8 options. There were 8 questions (0-8 points).

Task 11: Word Memory Recall

The participants responded to the 5 objects memorized in task 2 (0-5 points).

Task 12: Episodic Memory Recall

The participants responded with the object and location of the object that they stored away in task 2. If the participant were unable to answer, they listened to the options and selected the correct answer (0-4 points).

Measurement of Incidence of Dementia

Participants who did not have dementia at baseline and were diagnosed with dementia during the 60-month, follow-up period were considered to have new-onset dementia. Data were collected from the Japanese National Health Insurance System and the Japanese long-term care insurance system. In Japan, all adults aged ≥65 years have public health insurance, which includes one of the following: health insurance for employed individuals (Employees' Health Insurance), national health insurance for unemployed and self-employed individuals aged <75 years (Japanese National Health Insurance), or health care for individuals aged ≥75 years (Later-Stage Medical Care). In this study, participants were followed up monthly for new-onset dementia (AD, vascular dementia, frontotemporal dementia, or other dementia subtypes), as recorded by the Japanese National Health Insurance System and Later-Stage Medical Care. Participants were considered to have dementia based on a diagnosis made by a physician according to the *International Classification of Diseases, Tenth Revision*. We identified cases of dementia from the NCGG-SGS data according to the corresponding *International Classification of Diseases, Tenth Revision* codes (Multimedia Appendix 1). The diagnosis of dementia from the UK Biobank inpatient data, which have criteria similar to those of the NCGG-SGS (Multimedia Appendix 1), was previously validated and showed a positive predictive value of 87.3% for any dementia compared with the clinical expert adjudication of full-text medical records [14].

The long-term care insurance system is a mandatory form of social insurance that supports the daily lives of older adults with disabilities [15,16]. In Japan, all individuals aged ≥65 years are eligible for institutional or community-based services depending on their disability level, and those aged ≥40 years with disabilities related to aging, such as cerebrovascular disease, are also eligible to access benefits. The disability certification process consists of two parts: (1) the degree of disability based on a questionnaire developed by the Ministry of Health, Labor, and Welfare of Japan, and (2) a physician's written opinion prepared by the attending physician [17]. We identified dementia based on the level of independence in the daily living of older adults with dementia in the disability certification process. The level of independence was classified into 5 ranks, from mild to severe (I, II, III, IV, and M); individuals with moderate dementia-related behavioral disturbances and cognitive impairment with slight dependence were classified as rank II. We defined dementia as symptoms of rank II or higher, as derived from a primary care physician's written opinion. These criteria were the same as those used in previous studies in Japan [18-21].

Analysis of Potential Confounding Factors

Dementia results from a combination of factors that occur over time, including demographics and chronic conditions. All multivariate-adjusted models reported included the following covariates unless otherwise noted: age at enrollment, sex, educational level, drinking, smoking, presence of chronic conditions, BMI, and regular exercise. The self-reported chronic conditions included in the model were hypertension, diabetes, cancer, and osteoarthritis [22,23].

Ethical Considerations

This study was approved by the ethics committee of the Institutional Review Board of the National Center for Geriatrics and Gerontology (approval 770). All the participants were informed of the study and provided their written consent before their participation in this study, agreeing to the anonymous use of their data for publication. Data were deidentified before analysis and stored securely in a password-protected file on a password-protected computer. No compensation was provided to the study participants.

Statistical Analysis

First, those who participated at baseline and could follow the development of dementia during the follow-up period were randomly divided into the derivation and validation groups (1:1). Data homogeneity analysis regarding the control variables and J-Cog scores was tested using independent sample 2-tailed *t* tests or chi-square tests for contingency tests, depending on the variable structure. Incident dementia diagnoses per 1000 person-years were calculated for all participants and for participants classified by the cutoff point. Chi-square tests were used to compare the groups for incident dementia.

In the derivation group, receiver operating characteristic curves were plotted to determine the MMSE and J-Cog cutoff points that best discriminated between groups with and without dementia (the cutoff points are denoted with the 2 points on either side of the cutoff separated by a "/"). The area under the receiver operating characteristic curve was calculated, and the cutoff points for maximizing the sensitivity and specificity of each test were determined using the Youden index [24]. Based on these cutoff points, participants were categorized as having either high or low cognitive ability. We compared the accuracy of predicting the development of dementia by classifying

XSL•FO
RenderX

participants using the MMSE cutoff point 23/24, which is used to determine the risk of dementia [10].

In the validation group, 3 Cox proportional regression models were developed to predict the associations of the group (lower or higher cognitive ability) classified using the J-Cog or MMSE with dementia incidence. Specifically, the first and second models used MMSE score cutoffs of 23/24 and 28/29, respectively, and the third model used a J-Cog score cutoff of 43/44. All the models were adjusted for the potential confounding factors mentioned above. The adjusted hazard ratio (HR) and 95% CIs were obtained from the respective models. Harrell C-statistic was estimated to summarize how well a predicted risk score described an observed sequence of events. Generally, a C-statistic value greater than 0.7 indicates a good match [25,26]. The Akaike information criterion (AIC) was calculated for relative goodness of fit, where lower absolute values indicate a better model fit. Differences in Akaike information criterion (ΔAIC) of 3 or more were considered meaningful [27]. Harrell C-statistic and AIC were calculated using the *dynpred* (version 0.1.2) and *survival* (version 3.5.7) R package [28,29]. As sensitivity analyses, Cox proportional regression models were used, stratified by age group (65-74 years vs 75 years or more) and educational history (less than 9 years vs 10 years or more). These are variables that have a robust relationship with cognitive decline and dementia [30,31], and we felt that the stratified analyses would show the benefits of population-specific cognitive testing.

All statistical tests were considered significant at $P<.05$, and statistical analyses were performed using IBM SPSS Statistics (version 24.0) and R version 4.3.2 (R Core Team).

Results

Of the 2520 participants, 181 (7.2%) were diagnosed with dementia, 29 (1.2%) moved away from Takahama City, 138 (5.5%) died during the follow-up period, and 61 (2.4%) terminated the study for other reasons (Figure 1). Incident dementia diagnosis per 1000 person-years (95% CI) for total participants was 15.7 (13.6-18.1).

We examined whether the possible confounders of dementia incidence and cognitive tests differed at baseline between the derivation and validation groups (Table 1). No significant differences were observed in any of the measurements between the 2 groups (all $P>.05$). Table 2 presents the possible confounding factors for dementia incidence and cognitive test results for participants grouped according to the presence or absence of dementia. In the derivation group, significant differences were found for age ($P<.001$), educational level ($P<.001$), drinking ($P=.02$), diabetes ($P=.002$), cancer ($P=.04$), MMSE ($P<.001$), all subscores in the J-Cog (all $P<.05$), and total J-Cog score ($P<.001$) between participants with and without dementia (Table 2). In the validation group, significant differences were found for age ($P<.001$), sex ($P=.02$), education level ($P<.001$), hypertension ($P=.05$), osteoarthritis ($P<.001$), MMSE, and all subscores in the J-Cog (all $P<.05$) except for orientation and designation, and total J-Cog score ($P<.001$) between the participants with and without dementia incidence (Table 2).

The Youden index was used to determine the cutoff points for future dementia incidence during a 60-month period, which were 28/29 and 43/44 points on the MMSE and J-Cog, respectively. Incident dementia diagnoses per 1000 person-years with 95% CI for the low and high cognitive ability groups were 21.0 (17.9-24.6) and 7.1 (5.1-10.0) according to the MMSE 28/29 cutoff, 35.0 (29.5-41.4) and 6.0 (4.5-8.0) according to the J-Cog 43/44 cutoff, and 42.5 (31.1-58.0) and 13.2 (11.2-15.5) according to the MMSE 23/24 cutoff. The low cognitive ability groups, which were divided according to the 3 cutoff points, included significantly more participants with incident dementia than those with high cognitive ability ($P<.001$).

The predictive performances of all 3 risk cutoff points for dementia are presented in Table 3. In the total validation group participants, significant HRs for dementia incidence were found using the MMSE 23/24 and J-Cog 43/44 cutoffs (MMSE 23/24: HR 1.93, 95% CI 1.13-3.27; J-Cog 434/44: HR 2.42, 95% CI 1.50-3.93; Figure 2). The C-statistic was above 0.7 for all cutoff points. The ΔAIC with the MMSE 23/24 cutoff as a reference indicated that the MMSE 28/29 cutoff had poor discrimination, and the J-Cog 43/44 cutoff had good discrimination.

A subgroup analysis of participants aged 65-74 years showed no significant HRs (HR 1.31, 95% CI 0.53-3.22) at the MMSE 28/29 cutoff and a poor fit (ΔAIC=5.6). Participants aged 75 years and older showed a good fit at the J-Cog 43/44 cutoff (ΔAIC=–23.2), with significant HRs (HR 2.71, 95% CI 1.84-4.00) and good discrimination (C-statistics=0.714). In participants with less than 9 years of education, significant HRs for dementia incidence were found at the MMSE 23/24 and J-Cog 43/44 cutoff points (MMSE 23/24: HR 1.83, 95% CI 1.18-2.83; and J-Cog 43/44: HR 2.26, 95% CI 1.38-3.69). The ΔAIC using the MMSE 23/24 cutoff as a reference showed a poor fit at the MMSE 28/29 cutoff (ΔAIC=4.9) and a good fit at the J-Cog 43/44 cutoff (ΔAIC=–5.3). In participants with more than 10 years of education, significant HRs for dementia incidence were found at the MMSE 28/29 and J-Cog 43/44 cutoff points (MMSE 28/29: HR 1.84, 95% CI 1.06-3.22; and J-Cog 43/44: HR 3.53, 95% CI 2.10-5.93). The ΔAIC using the MMSE 23/24 cutoff as a reference showed good fit at the MMSE 28/29 (ΔAIC=–3.5) and J-Cog 43/44 cutoff points (ΔAIC=–22.7).

XSL•FO
RenderX

Figure 1. Flow chart of participant enrollment. MMSE: Mini-Mental State Examination.

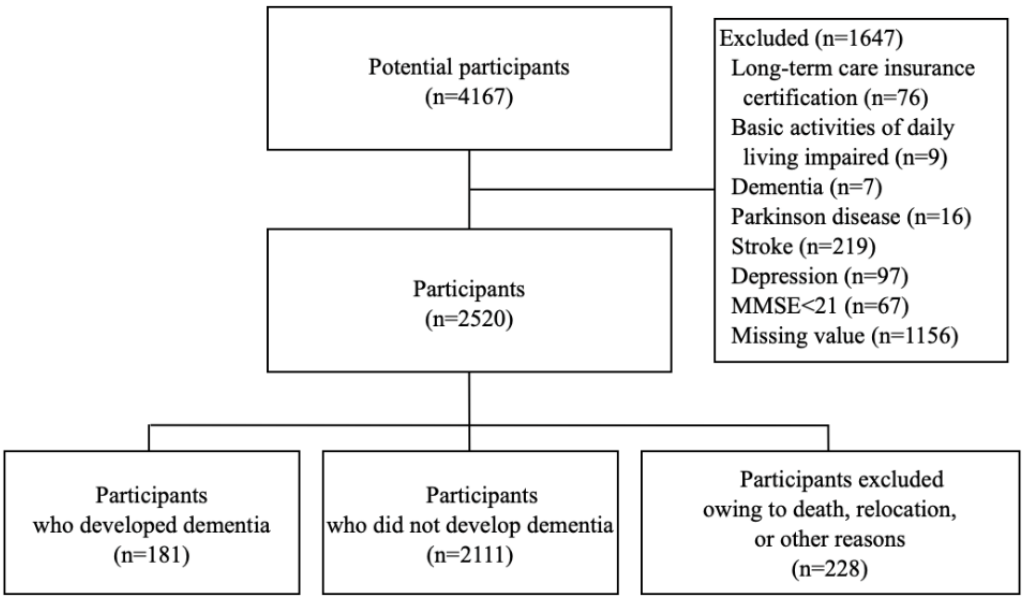

Table 1. Baseline characteristics of the participants.

Characteristic	Derivation group (n=1260)	Validation group (n=1260)	*P* value
Age, years (mean, SD)	72.7 (6.7)	72.7 (6.7)	.78
Sex, female (n, %)	756 (60)	747 (59.3)	.72
Education, years (mean, SD)	11.0 (2.3)	11.1 (2.4)	.44
Drinking, no (n, %)	863 (68.5)	865 (68.7)	.93
Smoking, yes (n, %)	121 (9.6)	111 (8.8)	.49
Hypertension, yes (n, %)	611 (48.5)	613 (48.7)	.94
Diabetes, yes (n, %)	182 (14.4)	171 (13.6)	.53
Cancer, yes (n, %)	152 (12.1)	161 (12.8)	.59
Osteoarthritis, yes (n, %)	235 (18.7)	236 (18.7)	.96
BMI (mean, SD)	23.4 (3.2)	23.5 (3.3)	.59
Regular exercise, no (n, %)	299 (23.7)	312 (24.8)	.55
Mini-mental state examination, point (mean, SD)	27.1 (2.4)	27.2 (2.4)	.40
J-Cog[a] , point (mean, SD)			
Orientation	5.0 (0.3)	4.9 (0.3)	.35
Designation	5.0 (0.2)	5.0 (0.2)	.27
Attention and calculation	4.2 (1.6)	4.2 (1.6)	.68
Mental rotation	3.8 (1.5)	3.8 (1.5)	.79
Verbal fluency	2.7 (1.0)	2.7 (1.0)	.84
Sentence completion	1.6 (0.5)	1.6 (0.5)	.38
Working memory	6.7 (0.7)	6.7 (0.6)	.12
Logical reasoning	3.5 (0.7)	3.5 (0.7)	.89
Attention	0.7 (0.5)	0.7 (0.5)	.28
Common knowledge	5.0 (1.7)	5.1 (1.8)	.20
Word and episodic memory recall	6.9 (1.9)	7.0 (1.9)	.63
Total	45.0 (6.0)	45.2 (6.0)	.35

[a]J-Cog: Japanese Cognitive Function Test.

XSL•FO
RenderX

Table 2. Baseline characteristics of study participants stratified by development of dementia.

Characteristic	Derivation group (n=1143)			Validation group (n=1149)		
	Participants who did not develop dementia (n=1051)	Participants who developed dementia (n=92)	*P* value	Participants who did not develop dementia (n=1060)	Participants who developed dementia (n=89)	*P* value
Age (years), mean (SD)	71.8 (6.3)	80.3 (4.7)	<.001	72.0 (6.3)	79.7 (5.4)	<.001
Sex (female), n (%)	643 (61.2)	54 (59)	.64	630 (59.4)	64 (72)	.02
Education (years), mean (SD)	11.1 (2.3)	10.2 (2.4)	<.001	11.2 (2.4)	10.2 (2.1)	<.001
Drinking (no), n (%)	706 (67.2)	73 (79)	.02	733 (69.2)	70 (79)	.06
Smoking (yes), n (%)	97 (9.2)	8 (9)	.87	974 (91.9)	82 (92)	.93
Hypertension (yes), n (%)	497 (47.3)	49 (53)	.27	503 (47.5)	52 (58)	.05
Diabetes (yes), n (%)	137 (13)	23 (25)	.002	141 (13.3)	11 (12)	.80
Cancer (yes), n (%)	119 (11.3)	17 (19)	.04	133 (12.5)	9 (10)	.503
Osteoarthritis (yes), n (%)	186 (17.7)	23 (25)	.08	187 (17.6)	31 (35)	<.001
BMI (kg/m ^{2}), mean (SD)	23.4 (3.2)	23 (3.3)	.18	23.6 (3.3)	23 (3.5)	.12
Regular exercise (no), n (%)	245 (23.3)	25 (27)	.40	261 (24.6)	22 (25)	.98
Mini-Mental State Examination (point), mean (SD)	27.3 (2.3)	25.8 (2.5)	<.001	27.3 (2.3)	26.0 (2.7)	<.001
J-Cog[a] (point), mean (SD)						
Orientation	5.0 (0.2)	4.8 (0.5)	.005	5.0 (0.3)	4.9 (0.4)	.15
Designation	5.0 (0.1)	4.9 (0.3)	.04	5.0 (0.2)	4.9 (0.3)	.12
Attention and calculation	4.3 (1.6)	3.6 (1.7)	<.001	4.3 (1.5)	3.6 (1.6)	<.001
Mental rotation	3.9 (1.4)	2.8 (1.8)	<.001	3.9 (1.5)	3.0 (1.7)	<.001
Verbal fluency	2.8 (1.0)	2.1 (0.9)	<.001	2.8 (0.9)	2.2 (0.9)	<.001
Sentence completion	1.6 (0.5)	1.4 (0.6)	<.001	1.7 (0.5)	1.4 (0.6)	<.001
Working memory	6.7 (0.6)	6.3 (0.9)	<.001	6.8 (0.6)	6.5 (0.6)	.001
Logical reasoning	3.5 (0.7)	3.1 (1.0)	<.001	3.5 (0.7)	3.3 (0.8)	.007
Attention	0.7 (0.5)	0.5 (0.5)	.003	0.7 (0.5)	0.5 (0.5)	<.001
Common knowledge	5.1 (1.7)	4.1 (1.7)	<.001	5.2 (1.7)	4.2 (1.7)	<.001
Word and episodic memory recall	7.1 (1.7)	5.1 (2.2)	<.001	7.1 (1.8)	5.4 (2.2)	<.001
Total	45.7 (5.7)	39.0 (6.1)	<.001	45.8 (5.8)	39.8 (6.1)	<.001

[a]J-Cog: Japanese Cognitive Function Test.

XSL•FO
RenderX

Table 3. Hazard ratio for development of dementia in the low and high cognitive ability groups.

Group and cutoff points	HR[a] (95% CI)	*P* value	C-statistic	AIC[b]	ΔAIC[c]
Validation group					
MMSE[d] 23/24	1.93 (1.13-3.27)	.02	0.82	1147.6	Reference
MMSE 28/29	1.34 (0.80-2.24)	.26	0.82	1151.7	4.1[e]
J-Cog[f] 43/44	2.42 (1.50-3.93)	<.001	0.83	1139.2	–8.4[g]
Participants aged 65-74 years					
MMSE 23/24	5.30 (1.61-17.38)	.006	0.83	306.1	Reference
MMSE 28/29	1.31 (0.53-3.22)	.56	0.83	311.7	5.6[e]
J-Cog 43/44	2.86 (1.13-7.22)	.03	0.84	307.2	1.1
Participants aged 75 years and older					
MMSE 23/24	1.64 (1.11-2.43)	.01	0.69	2110.9	Reference
MMSE 28/29	1.69 (1.10-2.62)	.02	0.69	2110.5	–0.4
J-Cog 43/44	2.71 (1.84-4.00)	<.001	0.71	2087.7	–23.2[g]
Participants with less than 9 years of education					
MMSE 23/24	1.83 (1.18-2.83)	.007	0.81	1348.5	Reference
MMSE 28/29	1.43 (0.83-2.46)	.20	0.81	1353.4	4.9[e]
J-Cog 43/44	2.26 (1.38-3.69)	.001	0.81	1343.1	–5.3[g]
Participants with 10 or more years of education					
MMSE 23/24	1.50 (0.72-3.11)	.28	0.84	964	Reference
MMSE 28/29	1.84 (1.06-3.22)	.03	0.85	960.5	–3.5[g]
J-Cog 43/44	3.53 (2.10-5.93)	<.001	0.87	941.3	–22.7[g]

[a]HR: hazard ratio.

[b]AIC: Akaike information criterion.

[c]ΔAIC: differences in Akaike information criterion.

[d]MMSE: Mini-Mental State Examination.

[e]Poor fit compared with Mini-Mental State Examination cutoff point 23/24 (ie, differences in Akaike information criterion greater than 3).

[f]J-Cog: Japanese Cognitive Function Test.

[g]Good fit compared with Mini-Mental State Examination cutoff point 23/24 (ie, differences in Akaike information criterion greater than –3).

XSL•FO
RenderX

Figure 2. Cox survival estimates for dementia incidence according to cognitive ability: (A) hazard estimates of dementia incidence with MMSE cutoff point 23/24; (B) hazard estimates of dementia incidence with MMSE cutoff point 28/29; and (C) hazard estimates of dementia incidence with J-Cog cutoff point 43/44. The cutoff points are denoted by the 2 points on either side of the cutoff separated by a "/". J-Cog: Japanese Cognitive Function Test; MMSE: Mini-Mental State Examination.

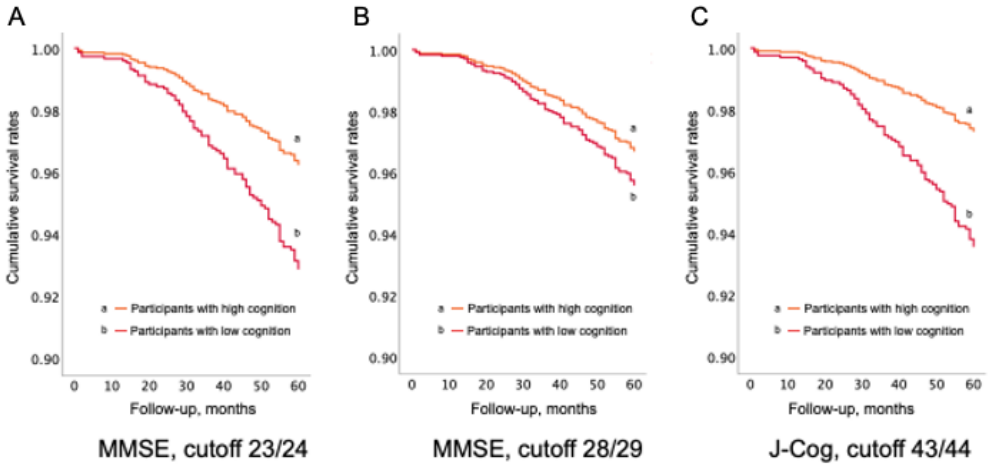

Discussion

Principal Findings

In this study, we developed and validated the J-Cog to establish a digital cognitive marker with a low economic and human burden to serve a large number of older adults. The J-Cog is a promising free tool for use in community centers and primary care clinics so that at-risk older adults can receive timely cognitive screening in the community. This study showed that the J-Cog has good criteria and predictive validity for identifying older adults who have developed dementia. Automated administration and scoring algorithms can reduce the workload of professionals and health care staff.

In this prospective study, dementia risk was significantly associated with the MMSE 23/24 and J-Cog 43/44 cutoff points, and a good fit at the J-Cog 43/44 cutoff point. The advantages of the MMSE include rapid administration and high acceptance as a diagnostic tool among health professionals and researchers [32]. Traditionally, a 23/24 cutoff has been used to select patients with suspected cognitive impairment or dementia [33]. However, sociocultural variables, such as age and education, among others, may affect individual scores [34,35]. A review to determine the accuracy of MMSE for the early detection of dementia reported no evidence supporting the substantial role of the MMSE as a stand-alone, single-administration test in the identification of participants who could develop dementia [36].

The optimal cutoff point for the development of dementia in this study was a high MMSE score of 28/29. The mean MMSE score was 27 (SD 2), and the median score was 28 (IQR 25-29). Only 211 (8.4%) of 2520 participants had MMSE scores below 23, which is traditionally considered a high-risk factor for dementia [10]. The reason for the higher cutoff point in this study may be the high distribution of MMSE scores.

The MMSE cutoff point of 23/24 showed better accuracy than 28/29 for participants in the total validation group, those aged 65-74 years, and those with less than 9 years of education. These results suggest that the MMSE 23/24 cutoff is a robust indicator of the development of dementia. However, the MMSE 28/29 cutoff point of showed a better fit than that of 23/24 for participants with more than 10 years of education. This result is in line with that of a previous study that reported the need for a higher MMSE cutoff for diagnosing dementia for patients with high educational levels [37].

The J-Cog showed a better fit for detecting the onset of dementia than the MMSE 23/24 cutoff points. The findings were similar in analyses in which participants were stratified by age and educational level, indicating the superiority of the J-Cog over MMSE in predicting incident dementia. In addition to predictive validity, the J-Cog has certain advantages over pen-and-paper cognitive function tests, such as the fact that (1) it does not require a professional to administer the test, (2) it can be administered free of charge, and (3) scoring and judgment are automated. These benefits are important for covering a large number of participants in a community setting for cognitive testing.

Comparison With Previous Work

The emergence of disease-modifying treatment options for AD has led to a paradigm shift in treatment strategies. Amyloid-targeted therapies are ideally used in the early stages of the disease, at the stage of MCI, and no later than mild dementia, because they slow disease progression but cannot reverse decline [38]. This preventive paradigm creates the need to identify patients with mild symptoms, typically in primary care settings, and many digital biomarkers have been developed. Systematic reviews on digital cognitive tests reported that most showed diagnostic performance comparable with that of paper-and-pencil tests for MCI and dementia. However, most studies have small sample sizes of fewer than 100 individuals and are based on case-control or cross-sectional designs [8,9]. In this study, a large population was followed for 60 months after implementation of the J-Cog to determine the predictive validity of the development of dementia in community-dwelling older adults. A recent systematic review reported the evaluation of digital cognitive testing applications, with a particular focus on the older people [39]. The review included an analysis of 20 articles. The number of cognitive tasks included in the accepted digital cognitive tests ranged from 1 to 10, with only 2 (10%)

XSL•FO
RenderX

tests including more than 8 tasks. The cognitive domains included executive functions, memory, visuospatial, attention, language, orientation, processing speed, learning, praxis or gnosis, abstractions, and calculations [39]. Many of these domains were included in the J-Cog. Previous studies have not found any validation of predictive validity with dementia onset as an outcome. This study provides the first evidence that digital cognitive tests are a tool that can predict the future onset of dementia.

Limitations

Although the J-Cog has some advantages over pen-and-paper cognitive function tests, there are some limitations to interpreting the results of this study. First, the participants were not randomly recruited, which may have resulted in an underrepresentation of cognitive impairment, because the participants were sufficiently healthy to undergo health examinations. Second, data on dementia subtypes were not collected (eg, AD, vascular dementia, dementia with Lewy bodies, and frontotemporal dementia). Thus, the data do not allow inferences regarding the correlation between the J-Cog and dementia pathology. Third, we were unable to contact an informant (eg, a family member) to verify the medical records, lifestyle information, or asymptomatic aberrant behavior. In addition, this study only identified dementia using specific *International Classification of Diseases, Tenth Revision* codes and information from the long-term care insurance system, so it was not possible to identify the onset of dementia in people who had dementia due to other factors or who had no record. These may have led to an underestimation of early-onset dementia in particular. Finally, information about the participants' medical conditions and comorbidities was collected through self-reports, and because we did not have access to medical records, we were unable to confirm these reports.

Conclusions

In this study, the J-Cog, a computerized cognitive battery test, was shown to be valid for detecting dementia development. The J-Cog is a cognitive function test that can be widely used in the community, and its social implementation should be promoted in the future so that AD treatment can be implemented in a timely and appropriate manner. A future extension of this study would be to improve the usability of the J-Cog so that it could be administered remotely and to remove the need for an operator to improve efficiency compared with paper-and-pencil tasks.

Acknowledgments

The study was supported by Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (A), 26242059); Ministry of Health, Labour and Welfare (Research project on health and welfare promotion for the elderly); the National Center for Geriatrics and Gerontology (Research Funding for Longevity Sciences 27-22); Japan Agency for Medical Research and Development (15dk0107003h0003, 15dk0207004h0203); Research Institute of Science and Technology for Society (RISTEX); and Japan Science and Technology Agency (JST) (Japan and Solution-Driven Co-creative R&D Program for SDGs [SOLVE for SDGs]: Preventing Social Isolation & Loneliness and Creating Diversified Social Networks: 22-221037254). The funders had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The authors acknowledge the substantial contributions of the dedicated and skilled staff at the National Center for Geriatric and Gerontology-Study of Geriatric Syndromes (NCGG-SGS). The authors are also grateful to the NCGG-SGS who volunteered to participate in this study.

Authors' Contributions

HS had full access to all data in the study and took responsibility for the integrity of the data and accuracy of the data analysis. Concept and design were contributed by HS and HM. Acquisition, analysis, and interpretation of data were contributed by TD, K Tsutsumimoto, KM, KH, K Tomida, and MM. Drafting of the manuscript was handled by HS. Critical revision of the manuscript for important intellectual content was contributed by all authors. Statistical analysis was contributed by HS and MM. Funding was handled by HS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ICD-10 codes used to identify dementia cases in the UK Biobank and National Center for Geriatric and Gerontology-Study of Geriatric Syndromes. ICD-10: International Statistical Classification of Diseases, Tenth Revision.
[DOCX File , 21 KB - jmir_v27i1e59015_app1.docx]

References

1. World Health Organization. Global Action Plan on the Public Health Response to Dementia 2017 - 2025. Geneva, Switzerland: World Health Organization; 2017:2025-2092.
2. Ministry of Justice. The Basic Act on Dementia to Promote an Inclusive Society (outline). Japanese Law Translation Database System. URL: https://www.japaneselawtranslation.go.jp/outline/92/905R510.pdf [accessed 2025-01-01]

XSL•FO
RenderX

3. US Preventive Services Task Force, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al. Screening for cognitive impairment in older adults: US preventive services task force recommendation statement. JAMA 2020;323(8):757-763. [doi: 10.1001/jama.2020.0435] [Medline: 32096858]
4. Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open 2017;7(2):e011146 [FREE Full text] [doi: 10.1136/bmjopen-2016-011146] [Medline: 28159845]
5. van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in early Alzheimer's disease. N Engl J Med 2023;388(1):9-21. [doi: 10.1056/NEJMoa2212948] [Medline: 36449413]
6. Sims JR, Zimmer JA, Evans CD, Lu M, Ardayfio P, Sparks J, et al. Donanemab in early symptomatic Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. JAMA 2023;330(6):512-527 [FREE Full text] [doi: 10.1001/jama.2023.13239] [Medline: 37459141]
7. WHO. Risk Reduction of Cognitive Decline and Dementia: WHO Guidelines. Geneva, Switzerland: World Health Organization; 2019.
8. Chan JYC, Yau STY, Kwok TCY, Tsoi KKF. Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: a systematic review. Ageing Res Rev 2021;72:101506. [doi: 10.1016/j.arr.2021.101506] [Medline: 34744026]
9. Ding Z, Lee TL, Chan AS. Digital cognitive biomarker for mild cognitive impairments and dementia: a systematic review. J Clin Med 2022;11(14):4191 [FREE Full text] [doi: 10.3390/jcm11144191] [Medline: 35887956]
10. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
11. Shimada H, Makizako H, Lee S, Doi T, Lee S, Tsutsumimoto K, et al. Impact of cognitive frailty on daily activities in older persons. J Nutr Health Aging 2016;20(7):729-735. [doi: 10.1007/s12603-016-0685-2] [Medline: 27499306]
12. Shimada H, Tsutsumimoto K, Lee S, Doi T, Makizako H, Lee S, et al. Driving continuity in cognitively impaired older drivers. Geriatr Gerontol Int 2016;16(4):508-514. [doi: 10.1111/ggi.12504] [Medline: 25953032]
13. Lezak MD. Neuropsychological Assessment. New York, NY: Oxford University Press; 2004.
14. Wilkinson T, Schnier C, Bush K, Rannikmäe K, Henshall DE, Lerpiniere C, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. Eur J Epidemiol 2019;34(6):557-565 [FREE Full text] [doi: 10.1007/s10654-019-00499-1] [Medline: 30806901]
15. Ikegami N. Public long-term care insurance in Japan. JAMA 1997;278(16):1310-1314. [Medline: 9343449]
16. Tsutsui T, Muramatsu N. Care-needs certification in the long-term care insurance system of Japan. J Am Geriatr Soc 2005;53(3):522-527. [doi: 10.1111/j.1532-5415.2005.53175.x] [Medline: 15743300]
17. Moriyama Y, Tamiya N, Kamimura A, Sandoval F, Luptak M. Doctors' opinion papers in long-term care need certification in Japan: comparison between clinic and advanced treatment hospital settings. Public Policy and Administration Research 2014(4):31-37.
18. Ikeda A, Yamagishi K, Tanigawa T, Cui R, Yao M, Noda H, et al. Cigarette smoking and risk of disabling dementia in a Japanese rural community: a nested case-control study. Cerebrovasc Dis 2008;25(4):324-331 [FREE Full text] [doi: 10.1159/000118377] [Medline: 18303251]
19. Yamamoto T, Kondo K, Hirai H, Nakade M, Aida J, Hirata Y. Association between self-reported dental health status and onset of dementia: a 4-year prospective cohort study of older Japanese adults from the Aichi Gerontological Evaluation Study (AGES) project. Psychosom Med 2012;74(3):241-248. [doi: 10.1097/PSY.0b013e318246dffb] [Medline: 22408130]
20. Yamagishi K, Ikeda A, Chei CL, Noda H, Umesawa M, Cui R, et al. Serum α-linolenic and other ω-3 fatty acids, and risk of disabling dementia: community-based nested case-control study. Clin Nutr 2017;36(3):793-797. [doi: 10.1016/j.clnu.2016.05.011] [Medline: 27265182]
21. Ihira H, Sawada N, Inoue M, Yasuda N, Yamagishi K, Charvat H, et al. Association between physical activity and risk of disabling dementia in Japan. JAMA Netw Open 2022;5(3):e224590 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.4590] [Medline: 35348711]
22. Peters R, Booth A, Rockwood K, Peters J, D'Este C, Anstey KJ. Combining modifiable risk factors and risk of dementia: a systematic review and meta-analysis. BMJ Open 2019;9(1):e022846 [FREE Full text] [doi: 10.1136/bmjopen-2018-022846] [Medline: 30782689]
23. Anstey KJ, Ee N, Eramudugolla R, Jagger C, Peters R. A systematic review of meta-analyses that evaluate risk factors for dementia to evaluate the quantity, quality, and global representativeness of evidence. J Alzheimers Dis 2019;70(s1):S165-S186 [FREE Full text] [doi: 10.3233/JAD-190181] [Medline: 31306123]
24. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163(7):670-675 [FREE Full text] [doi: 10.1093/aje/kwj063] [Medline: 16410346]
25. Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. J Biomed Inform 2020;108:103496 [FREE Full text] [doi: 10.1016/j.jbi.2020.103496] [Medline: 32652236]

XSL•FO
RenderX

26. Sun W, Jiang YZ, Liu YR, Ma D, Shao ZM. Nomograms to estimate long-term overall survival and breast cancer-specific survival of patients with luminal breast cancer. Oncotarget 2016;7(15):20496-20506 [FREE Full text] [doi: 10.18632/oncotarget.7975] [Medline: 26967253]
27. Fayosse A, Nguyen DP, Dugravot A, Dumurgier J, Tabak AG, Kivimäki M, et al. Risk prediction models for dementia: role of age and cardiometabolic risk factors. BMC Med 2020;18(1):107 [FREE Full text] [doi: 10.1186/s12916-020-01578-x] [Medline: 32423410]
28. Putter H. dynpred: companion package to "Dynamic Prediction in Clinical Survival Analysis". R Project. 2011. URL: https://cran.r-project.org/web/packages/dynpred/index.html [accessed 2025-01-01]
29. Therneau TM, Lumley T, Elizabeth A, Cynthia C. survival: survival analysis. R Project. 2011. URL: https://cran.r-project.org/web/packages/survival/index.html [accessed 2025-01-01]
30. Alzheimer's Association. 2020 Alzheimer's disease facts and figures. Alzheimers Dement 2020 Mar 1 [FREE Full text] [doi: 10.1002/alz.12068] [Medline: 32157811]
31. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Lee S, Hotta R, et al. Cognitive functioning and walking speed in older adults as predictors of limitations in self-reported instrumental activity of daily living: prospective findings from the obu study of health promotion for the elderly. Int J Environ Res Public Health 2015;12(3):3002-3013 [FREE Full text] [doi: 10.3390/ijerph120303002] [Medline: 25768239]
32. Nieuwenhuis-Mark RE. The death knoll for the MMSE: has it outlived its purpose? J Geriatr Psychiatry Neurol 2010;23(3):151-157. [doi: 10.1177/0891988710363714] [Medline: 20231732]
33. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. J Am Geriatr Soc 1992;40(9):922-935. [doi: 10.1111/j.1532-5415.1992.tb01992.x] [Medline: 1512391]
34. Brayne C, Calloway P. The association of education and socioeconomic status with the Mini Mental State Examination and the clinical diagnosis of dementia in elderly people. Age Ageing 1990;19(2):91-96. [doi: 10.1093/ageing/19.2.91] [Medline: 2337015]
35. Bleecker ML, Bolla-Wilson K, Kawas C, Agnew J. Age-specific norms for the Mini-Mental State Exam. Neurology 1988;38(10):1565-1568. [doi: 10.1212/wnl.38.10.1565] [Medline: 3419600]
36. Arevalo-Rodriguez I, Smailagic N, Roqué-Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). Cochrane Database Syst Rev 2021;7(7):CD010783 [FREE Full text] [doi: 10.1002/14651858.CD010783.pub3] [Medline: 34313331]
37. Kochhann R, Varela JS, Lisboa CSDM, Chaves MLF. The Mini Mental State Examination: review of cutoff points adjusted for schooling in a large Southern Brazilian sample. Dement Neuropsychol 2010;4(1):35-41 [FREE Full text] [doi: 10.1590/S1980-57642010DN40100006] [Medline: 29213658]
38. Cummings J, Rabinovici GD, Atri A, Aisen P, Apostolova LG, Hendrix S, et al. Aducanumab: appropriate use recommendations update. J Prev Alzheimers Dis 2022;9(2):221-230 [FREE Full text] [doi: 10.14283/jpad.2022.34] [Medline: 35542993]
39. Cubillos C, Rienzo A. Digital cognitive assessment tests for older adults: systematic literature review. JMIR Ment Health 2023;10:e47487 [FREE Full text] [doi: 10.2196/47487] [Medline: 38064247]

Abbreviations

AD: Alzheimer disease
AIC: Akaike information criterion
ΔAIC: differences in Akaike information criterion
HR: hazard ratio
J-Cog: Japanese Cognitive Function Test
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
NCGG-SGS: National Center for Geriatric and Gerontology-Study of Geriatric Syndromes

Edited by A Schwartz; submitted 30.03.24; peer-reviewed by X Li, J Robertson; comments to author 13.06.24; revised version received 25.09.24; accepted 28.11.24; published 14.02.25.

Please cite as:
Shimada H, Doi T, Tsutsumimoto K, Makino K, Harada K, Tomida K, Morikawa M, Makizako H
A New Computer-Based Cognitive Measure for Early Detection of Dementia Risk (Japan Cognitive Function Test): Validation Study
J Med Internet Res 2025;27:e59015
URL: https://www.jmir.org/2025/1/e59015
doi:10.2196/59015
PMID:

XSL•FO
RenderX

©Hiroyuki Shimada, Takehiko Doi, Kota Tsutsumimoto, Keitaro Makino, Kenji Harada, Kouki Tomida, Masanori Morikawa, Hyuma Makizako. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Association of Screen Content With Early Development Among Preschoolers in Shanghai: 7-Day Monitoring Study With Auto Intelligent Technology

Hao Chen[1*], PhD; Yi Sun[1*], MPH; Sha Luo[2*], MD; Yingyan Ma[3,4*], MD, PhD; Chenshu Li[1], MPH; Yingcheng Xiao[1], MPH; Yimeng Zhang[1], MPH; Senlin Lin[3], MSc; Yingnan Jia[1], PhD

[1]Preventive Medicine and Health Education Department, School of Public Health, Fudan University, Shanghai, China

[2]Xuhui Maternity and Child Healthcare Center, Shanghai, China

[3]Shanghai Eye Hospital, Shanghai Eye Diseases Prevention & Treatment Center, Shanghai, China

[4]Department of Ophthalmology, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Yingnan Jia, PhD
Preventive Medicine and Health Education Department
School of Public Health
Fudan University
No 138 Yixueyuan Road
Shanghai, 200032
China
Phone: 86 13764665540
Email: jyn@fudan.edu.cn

Abstract

Background: It is unclear how exposure to different types of screen content is associated with early development among preschool children.

Objective: This study aims to precisely evaluate the screen exposure time across different content types and to explore the associations with the Ages and Stages Questionnaire, Third Edition (ASQ-3) score and 5 capacity domains in children aged 34.5-66 months.

Methods: This monitoring study used intelligent technology to collect data on the 7-day screen time and the time spent viewing each content type. The participants were 2 groups of Shanghai kindergarten kids. The data were collected between March 2023 and July 2023. Screen exposure data (total daily time and time for each type of content) were collected from children aged between 34.5 and 66 months. A self-designed questionnaire and the Healthy Screen Viewing for Children intelligent technology app were used to assess screen exposure to all media and tablets. The ASQ-3 was used to assess early development in children aged 34.5-66 months.

Results: In the 535-child sample, the results of linear regression analysis indicated that both screen time of more than 60 minutes and exposure to smartphones and tablets were negatively associated with ASQ-3 score. Among 365 participants with data collected by the Healthy Screen Viewing for Children app, median regression showed that the median total ASQ-3 score was negatively associated with screen time for noneducational content (β=–.055; 95% CI –0.148 to –0.006; *P*=.03), screen time for both educational and noneducational content (β=–.042; 95% CI –0.081 to –0.007; *P*=.001), and fast-paced content (β=–.034; 95% CI –0.062 to –0.011; *P*=.049). The median gross motor score was negatively associated with screen time for parental guidance-13–rated content (β=–.015; 95% CI –0.022 to 0.009; *P*=.03), educational and noneducational content (β=–.018, 95% CI –0.038 to –0.001; *P*=.02), static content (β=–.022; 95% CI –0.050 to 0.007; *P*=.02). This study also revealed that the median fine motor score was negatively associated with screen time for guidance–rated content (β=–.032, 95% CI –0.057 to –0.003; *P*=.006), parental guidance (PG) rated content (β=–.020; 95% CI –0.036 to –0.007; *P*=.004), noneducational content (β=–.026; 95% CI –0.067 to –0.003; *P*=.01), both educational and noneducational content (β=–.020; 95% CI –0.034 to –0.001; *P*<.001), fast-paced content (β=–.022; 95% CI –0.033 to –0.014; *P*<.001), static content (β=–.034; 95% CI –0.050 to 0.018; *P*<.001), animated content (β=–.038; 95% CI –0.069 to –0.001; *P*=.004), and screen use during the daytime (β=–.026; 95% CI –0.043 to 0.005; *P*=.005).

XSL•FO
RenderX

Conclusions: The results indicated that the time spent viewing noneducational, static, fast-paced, and animated content was negatively associated with early development among preschool children. Limiting screen time in relevant aspects is recommended.

(J Med Internet Res 2025;27:e65343) doi:10.2196/65343

KEYWORDS

types of screen content; screen time; intelligent technology; early development; preschool

Introduction

Since iPads and similar touch-enabled media technological devices are becoming increasingly popular worldwide, the increasingly younger trend of screen exposure may threaten the early development of toddlers and preschool children [1]. Although the World Health Organization [2] and pediatric societies worldwide (eg, Australian [3] and Canadian [4] 24-Hour Movement Guidelines) have recommended no more than 1-2 hours per day for children aged 2 to 5 years, the global prevalence of children who meet the screen time guidelines (1 hour/day) is only 35.6% (95% CI 30.6%-40.9%) [5]. Excessive screen time is associated with child language skills [6]; development [7]; mental health symptoms, including irritability, inattention, and hyperactivity [8]; obesity [9,10]; and physical inactivity [11,12].

The most active and most vulnerable period for brain and neural development is from the ages of 2 to 5 years [13] because children are able to generalize an action learned based on 2D screen images to a real 3D situation beginning at 2 years of age [14]. In addition to excessive screen time, there is still controversy about whether and how exposure to screen content during early life could influence children's physical, psychological, and social development. Previous studies have shown that early exposure to violent entertainment programs is associated with aggressive behavior in childhood [13], and watching entertainment programs has adverse effects on cognitive development [15]; moreover, a cohort study showed that viewing entertainment programs was associated with greater expressive word skills [16]. Several intervention experiments have shown that watching cartoons with a fast screen-switching rate reduces executive function in children, but exposure to content with a slow screen-switching rate has no significant effect on executive function in children [17,18]. Several studies have suggested that playing games can improve children's cognitive abilities, such as attention, visual processing, and spatial imagination, in the short term [19], but the long-term learning and educational benefits are not obvious [20].

In general, there is an urgent need for greater awareness about the complexities of the impact of exposure to screen content on children's development, especially regarding touch-enabled media technology [21]. The most recent umbrella review of screen time studies, also calls for more objective measurements [22]. However, previous research seems simple and unable to match the pace of technological advances because the majority of screen exposure measurements are based on parent-reported screen viewing time and content, which might be susceptible to recall bias [1,23]. Conversely, the World Health Organization recommends that toddlers view high-quality content but fails to specify what defines such quality or offers representative examples [2]. Therefore, the primary aim of this study was to collect real-time screen exposure data through intelligent monitoring technology and then examine the association between screen viewing content and time and children's early development. These results may be beneficial for defining high-quality evidence for direct future recommendations.

Methods

Study Design and Population

This monitoring study was conducted in 2 districts of Shanghai, Xuhui District (urban), and Pudong New Area (suburban), from March 2023 to July 2023. A random sampling method was used to select a kindergarten from each of the 2 districts. For each kindergarten, a cluster sampling method was used to select 5 classes, in which the average number of children in each class was 30, from the junior, middle, and senior grades. A total of 535 preschool children were included in this study, and they were required to use the Huawei iPad for 1 week. During the recruitment process, we repeatedly emphasized that children should use tablets according to their daily habits at home and that parents should neither indulge in children's use which may lead to an overestimation of screen exposure duration nor overly regulate it which may result in an underestimation of screen exposure duration. All the children were provided with intelligent monitoring technology and completed a developmental screening questionnaire. The minimum sample size was calculated to be 462 by using the following formula:  , where the reported prevalence of abnormal development among nationwide Chinese preschool children was approximately 10.8% to 26.2% [24]. In this study, the prevalence (p) was 10.8%, the type I error (α) was .05, the $z_{1-/2}$ was 1.96, the precision (d) was 0.04, and the design effect (deff) was 2 [25]. The inclusion criteria for participant enrollment were as follows: (1) children aged between 2 and 6 years; (2) children whose parents or other caregivers allowed them to use screen devices, such as televisions, smartphones, iPads, and computers; and (3) children whose parents or other caregivers understood the questionnaire.

Measures

Screen Exposure Monitoring Technology

The "Healthy Screen Viewing for Children (HSVC)" mobile app, developed by the Department of Preresearch of Huawei Corporations and installed on the Huawei iPad, was used to automatically collect data on screen time and screen content while running in the background. The screen time was calculated by obtaining the start time point and end time point as well as the sum of the duration of use for every app. Correspondingly, app name labels, namely, screen content information, were also

XSL•FO
RenderX

collected simultaneously (the specific data collection and storage procedures are shown in Figure S1 in Multimedia Appendix 1). Considering that video applications (eg, YouTube and Netflix) cannot reflect the viewing program according to the application name, the HSVC application takes screenshots every 60 seconds to obtain images of the screen as an important source of data on-screen content. All the abovementioned screen exposure data were uploaded in real-time to the cloud server of Huawei Corp and can be downloaded from the Huawei Research Platform. Based on the data of 268 preschool children from the accuracy survey, the accuracy between the screen-time category from the HSVC records and the 24-hour diary according to the Bayesian classification of the testing set and training set were 0.96 and 0.93, respectively; moreover, the κ value of representativeness was 0.61.

Classification of Screen Content

The 7-day monitoring study ultimately collected 138.98 thousand screen records and 41,812 screenshot images. Based on previous studies and the Delphi method, screen content type was classified into the following six regulations.

- App function direction: entertainment videos, video games, tools, educational videos, and short videos.
- Suggested audience: General audience, parental guidance suggested (PG), parents strongly cautioned (PG-13), and restricted according to the film rating system formulated by the Motion Picture Association of America [26].
- Educational direction: Educational and noneducational content.
- Screen pace direction: The pace of content was defined as 2 episodes viewed for the number of times a complete change occurred (eg, from the living room to the stadium) according to a previous standard [18], and we categorized pace into fast-paced, medium-paced, and slow-paced content (switched more than 6 times per minute, 2 to 6 times per minute, and less than 2 times per minute, respectively).
- Screen interaction direction: Static, touch-enabled, and interactive for children and caregivers.
- Realistic or animated direction: The content corresponded to the program or game, which mainly included realistic content, animated content, and both realistic and animated content.

The specific classification procedures and standards are shown in Table S1 in Multimedia Appendix 1.

Developmental Screening

The Chinese version of the Ages and Stages Questionnaire, Third Edition (ASQ-3), which is a widely used, parent-reported screening tool, was used for developmental screening. The ASQ-3 comprises the following 5 domains and includes 30 items: communication (6 items), gross motor (6 items), fine motor (6 items), problem-solving (6 items), and personal-social (6 items) [27]. Each item is scored on a 3-point scale ranging from 10=yes to 5=sometimes to 0=not yet. The total ASQ-3 score is obtained by summing the corresponding item scores, with a higher ASQ-3 score indicating a preferable development level [28,29]. After validation via standardized testing of developmental (Bayley Scales of Infant Development) and intellectual (Stanford-Binet Intelligence Test–4th Edition) skills, the ASQ-3 was found to have moderate to high sensitivity (0.70-0.90) and specificity (0.76-0.91), as well as strong concurrent validity (r=0.85) and internal consistency (α=.51-.87) [30,31]. The ASQ-3 was released web-based through the Wenjuanxing platform. It included the filling instructions for ASQ-3. Parents were required to conduct evaluations on children in a quiet environment and then fill in the questionnaires themselves. We also provided contact information so that parents could give feedback in a timely manner and solve problems when encountering difficulties in filling out the questionnaires.

Covariates

Based on the results of univariate analysis of demographic characteristics, lifestyles, Ages and Stages Questionnaire (ASQ) scores, and each dimension (Table S2 in Multimedia Appendix 1), we respectively selected the indicators that had statistically significant associations with ASQ and its 5 dimensions as covariates and incorporated them into the regression models of ASQ and each dimension respectively, in order to evaluate the associations between screen exposure content and ASQ more accurately. In addition, we investigated children's screen exposure in kindergartens. The results of 70 questionnaires showed that the screen exposure in schools was background screen exposure (such as playing videos and music in the classroom during recess gymnastics time while children were doing sports outdoors). The median duration of background screen exposure was P_{50}=30 (20-30) minutes, and children rarely actively obtained information from background exposure. Therefore, this study did not include background screen exposure as a covariate in the analysis.

Caregiver-Reported Questionnaire

Based on the literature, the covariates in this study included child age, child gender, caregiver, maternal educational attainment, paternal educational attainment, monthly income, single child status, and health-related behaviors, including day sleep time, night sleep time, indoor physical activity level, outdoor physical activity level and screen control by caregivers.

A screen exposure questionnaire was used to measure screen time on all media to compensate for the inability of the HSVC to collect data from television, computers, and other nontablet media. The screen times for televisions, personal computers, smartphones, and tablets were derived from graded responses (0 minute, 1-29 minutes, 30-59 minutes, 60-120 minutes, or >120 minutes) to the question "How long does your child spend using TVs/PCs/cell phones/tablets on an average day?" The age at first screen exposure was derived from frequency responses (under the age of 1 year, 1-2 years, 2-3 years, or older than 3 years). The types of on-screen content were derived from ranked responses (eg, mainly educational or mainly entertainment or both educational and entertainment) to the question "What types of on-screen content is your child exposed to?" The response methods used for the type of on-screen content, namely, age restriction, animated direction, fast-paced, and touch-enabled, were similar to those used for educational direction. Caregivers' interactions with children when viewing programs and playing games are related to early development; thus, the use of co-viewing programs and co-playing games was derived from

XSL•FO
RenderX

frequency responses, which included seldom, sometimes, and often.

Statistical Analysis

Participant characteristics are presented as frequencies with percentages and as the means with SDs. First, we used descriptive analysis to show the demographic characteristics of the participants in the whole sample with questionnaire data and in the subsample with HSVC device records. Second, a 2-tailed Student *t* test and ANOVA were performed to explore the distribution of the total ASQ-3 score for the whole sample and the scores on the 5 domains according to demographic characteristics, lifestyle, screen time category of the 4 media types, screen use styles, and type of screen exposure content, as shown in Tables S2 and S3 in Multimedia Appendix 1. Then, 6 multiple linear regression analyses were performed to examine the associations between the total ASQ score and the scores on the 5 domains (dependent variable), screen time category of the 4 media types, screen use styles, and type of screen exposure content (independent variables) adjusted for age, gender, parental educational attainment, and monthly income (control variable; Table S4 in Multimedia Appendix 1). Furthermore, Spearman correlation analysis was performed to examine the univariate association between the total ASQ score and the scores on the 5 domains and between screen viewing time and exposure to different types of content according to the HSVC application, as shown in Table S5 in Multimedia Appendix 1. Finally, median regressions were used to examine the nonlinear correlation trend between the median score on the ASQ and median time of exposure to content according to the HSVC sample application, as shown in Table S6 in Multimedia Appendix 1. The estimates of the independent variables for the total ASQ score and the domain scores were summarized using regression coefficients (β) and 95% CIs. Statistical analyses were performed using R software (version 4.3.2; R Foundation for Statistical Computing).

Ethical Considerations

Before the participants underwent screen exposure monitoring and questionnaire surveys, we informed their parents (legal guardians) of the whole process of participation in the surveys as well as the benefits and risks involved, and the legal guardians signed the informed consent. This study conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Ethics Committee for Medical Research at the School of Public Health, Fudan University (IRB#2023-11-1088). When extracting data from the Huawei Research Platform and the web-based questionnaire library for analysis, all the data were deidentified, and an anonymous study ID was used as an identifier for each participant. In this study, we provided a gift of a paintbrush set, worth approximately US $15, to each child participating in the research.

Results

Participant Characteristics

As shown in Table 1, among a total of 535 children and 365 preschool children with tablet use, 298 (55.7%) and 205 (56.2%) were boys, and more than half of the participants' caregivers were their mothers (n=297, 55.5% and n=205, 55.6%, respectively). The majority of participants' parents had a bachelor degree (mothers: n=290, 54.2% and n=196, 53.7%; fathers: n=286, 53.5% and n=194, 53.2%). Most of the participants' monthly incomes were over US$ 1750 (n=309, 57.8% and n=202, 55.4%). Nearly half of the participants were only children (n=272, 50.8% and n=185, 50.7%). Only 50 (9.3%) and 33 (9.1%) of participants had insufficient sleep during the day, while 412 (77%) and 280 (76.7%) of participants had insufficient sleep during the night (less than 10 hours). Less than half of the participants reported insufficient indoor physical activity (209, 39.1% and 138, 37.8%) or insufficient outdoor physical activity (183, 34.2% and 120, 32.9%).

The daily median screen time on all media was 98.6 (60-65.5) minutes, and the screen time on tablets was 41.1 (0-107.2) minutes. The mean total score and score for each domain of the ASQ for the total sample and tablet sample were 269.3 (SD 33.33) and 269.89 (SD 30.75); 55.05 (SD 7.47) and 54.86 (SD 5.51) for communication); 51.54 (SD 10.39) and 51.61 (SD 10.12) for gross motor; 50.38 (SD 11.26) and 50.45 (SD 10.86) for fine motor; 56.61 (SD 6.92) and 56.82 (SD 6.01) for problem-solving; and 55.73 (SD 7.11) and 56.14 (SD 6.45) for personal-social, respectively. The participant characteristics and total ASQ and domain scores did not significantly differ between the 2 sample populations (P>.05).

XSL•FO
RenderX

Table 1. Demographic characteristics.

Variables	Total sample	Tablet use sample	*P* value
Age, n (%)			.80
Approximately 34 months 16 days to 50 months 30 days	54 (10.1)	32 (8.8)	
Approximately 51 months 0 days to 56 months 30 days	105 (19.6)	74 (20.3)	
Approximately 57 months 0 days to 66 months 0 days	376 (70.3)	259 (70.9)	
Gender, n (%)			.89
Girl	237 (44.3)	160 (43.8)	
Boy	298 (55.7)	205 (56.2)	
Caregiver, n (%)			.95
Mother only	297 (55.5)	203 (55.6)	
Father only	35 (6.5)	22 (6.0)	
Grand parents	203 (38.0)	140 (38.4)	
Mother's education, n (%)			.98
≤Secondary school	93 (17.4)	65 (17.8)	
Senior high school	152 (28.4)	104 (28.5)	
≥Bachelor	290 (54.2)	196 (53.7)	
Father's education, n (%)			.89
≤Secondary school	69 (12.9)	51 (13.9)	
Senior high school	180 (33.6)	120 (32.9)	
≥Bachelor	286 (53.5)	194 (53.2)	
Monthly income (US $)[a], n (%)			.73
<1000	145 (27.1)	102 (27.9)	
Approximately 1000 to 1750	81 (15.1)	61 (16.7)	
>1750	309 (57.8)	202 (55.4)	
Single child, n (%)			.96
One	272 (50.8)	185 (50.7)	
Two or more	263 (49.2)	180 (49.3)	
Day sleep time, n (%)			.85
Normal	249 (46.6)	164 (44.9)	
Insufficient	50 (9.3)	33 (9.1)	
Excessive	236 (44.1)	168 (46.0)	
Night sleep time, n (%)			.92
Sufficient	123 (23.0)	85 (23.3)	
Insufficient	412 (77.0)	280 (76.7)	
Indoor physical activity, n (%)			.70
Sufficient	326 (60.9)	227 (62.2)	
Insufficient	209 (39.1)	138 (37.8)	
Outdoor physical activity, n (%)			.68
Sufficient	352 (65.8)	245 (67.1)	
Insufficient	183 (34.2)	120 (32.9)	
Screen guard, n (%)			.31
Strictly	159 (29.7)	97 (26.6)	
Occasionally	376 (70.3)	268 (73.4)	

XSL•FO
RenderX

Variables	Total sample	Tablet use sample	*P* value
Daily total screen time, n (%)			.69
<1 hour	103 (19.3)	62 (17.0)	
1~2 hours	200 (37.4)	139 (38.1)	
>2 hours	232 (43.3)	164 (44.9)	
Score of ASQ[b], Mean (SD)			
Total score of ASQ	269.3 (33.33)	269.89 (30.75)	.79
Score of Communication	55.05 (7.47)	54.86 (5.51)	.71
Score of gross motor	51.54 (10.39)	51.61 (10.12)	.92
Score of fine motor	50.38 (11.26)	50.45 (10.86)	.93
Score of problem-solving	56.61 (6.92)	56.82 (6.01)	.62
Score of personal-social	55.73 (7.11)	56.14 (6.45)	.37

[a]Monthly income: The original options are <7500 RMB, 7500-12500 RMB, >12500 RMB. To convert the amounts from RMB to US dollars, multiply them by 0.137.

[b]ASQ: Ages and Stages Questionnaire.

Associations Between Screen Exposure Type and ASQ-3 Scores

The univariate associations between demographic characteristics and lifestyle variables and ASQ-3 scores are shown in Table S2 in Multimedia Appendix 1. Age, gender, maternal and paternal educational attainment, and monthly income were associated with the total ASQ score and domain scores (*P*<.05). The results of the univariate association between screen exposure types and ASQ-3 scores are shown in Table S3 in Multimedia Appendix 1. Smartphone screen time, iPad screen time, education content category, animated or realistic content, co-viewing programs, and co-playing games were associated with the total ASQ score and the domain scores (*P*<.05).

Furthermore, as shown in Table S4 in Multimedia Appendix 1, the results of the fully adjusted linear regression model suggested that older age, compared with an age of approximately 34 months 16 days to 50 months 30 days, was more strongly associated with the personal-social score (β=2.29; 95% CI 0.02-4.56; *P*=.048 and β=2.79; 95% CI 0.75-4.83; *P*=.008). Male gender was negatively associated with the total ASQ score (β=–8.92; 95% CI –14.41 to –3.43; *P*=.002), fine motor score (β=–4.16, 95% CI –6.04 to –2.28; *P*<.001), and personal-social score (β=–2.52; 95% CI –3.71 to –1.33; *P*<.001). Comparing with no exposure to smartphones, the total ASQ-3 score, gross motor scores were negatively associated with the use of smartphones for less than 30 minutes (β=–8.62; 95% CI –15.58 to –1.67; *P*=.01 and β=–3.16, 95% CI –5.4 to –0.92; *P*=.01), from 30-60 minutes (β=–13.28; 95% CI –21.98 to –4.58, *P*=.003 and β=–4.52; 95% CI –7.32 to –1.71; *P*=.002) and for more than 60 minutes (β=–12.89; 95% CI –24.75 to –1.02; *P*=.03 and β=–4.71; 95% CI –8.54 to –0.88; *P*=.02). The duration of iPad screen time exceeding 1 hour (β=2.45; 95% CI 0.39-4.51; *P*=.02) had a more negative association with the fine motor score than did never using iPads. Engagement in co-playing games with caregivers was positively associated with the total ASQ-3 score (β=11.64; 95% CI 2.66-20.62; *P*=.01), communication (β=2.19; 95% CI 0.1-4.27; *P*=.04), and fine motor scores (β=4.76; 95% CI 1.68-7.84; *P*=.003), respectively.

Associations Between Screen Content and ASQ-3 Score

The Spearman correlation coefficients between the screen content and ASQ-3 scores based on the HSVC data are shown in Table S5 in Multimedia Appendix 1. The results showed that education content, pace, screen interaction, animated and realistic content, date, and screen time duration were associated with the total ASQ-3 score, fine motor score, problem-solving score, and personal-social score (*P*<.05).

Furthermore, the results of the median regression between screen content and the ASQ-3 score after adjusting for age, gender, monthly income, co-viewing programs, and co-playing games are shown in Table S6 in Multimedia Appendix 1. This study showed that the median total score on the ASQ-3 was negatively associated with screen time for noneducational content (β=–.055; 95% CI –0.148 to –0.006; *P*=.03), educational and noneducational content (β=–.042, 95% CI –0.081 to –0.007; *P*=.001), and fast-paced content (β=–.034; 95% CI –0.062 to –0.011; *P*=.049). The median gross motor score was negatively associated with screen time for PG-13 rated content (β=–.015; 95%CI –0.022 to 0.009; *P*=.03), educational and noneducational content (β=–.018, 95% CI –0.038 to –0.001; *P*=.02), static content (β=–.022, 95% CI –0.050 to 0.007; *P*=.02). This study also found that the median fine motor score was negatively associated with the time spent viewing G rated content (β=–.032; 95% CI –0.057 to –0.003; *P*=.006), PG rated content (β=–.020; 95% CI –0.036 to –0.007; *P*=.004), noneducational content (β=–.026; 95% CI –0.067 to –0.003; *P*=.01), both educational and noneducational content (β=–.020; 95% CI –0.034 to –0.0010; *P*<.001), fast-paced content (β=–.022; 95% CI –0.033 to –0.014; *P*<.001), static content (β=–.034; 95% CI –0.050 to 0.018; *P*<.001), animated content (β=–.038; 95% CI –0.069 to –0.001; *P*=.004), and screen use during the daytime (β=–.026; 95% CI –0.043 to 0.005; *P*=.005). This study suggested that screen time during the daytime and the time spent viewing touch-based content were negatively associated with the median

XSL•FO
RenderX

problem-solving (β=–.017; 95% CI –0.050 to –0.007; *P*=.03) and personal-social scores (β=–.016; 95% CI –0.033 to 0.003; *P*=.006), respectively.

Discussion

Principal Findings

Based on 7 days of monitoring data collected by the HSVC application from 2 representative samples of kindergarten children aged 34.5 to 66 months, we analyzed the association of different screen content types and the ASQ-3 score. We found that greater exposure to noneducational programs, fast-paced content, static, animated content, and screens during the daytime was associated with a lower median gross motor score and median fine motor score after controlling for covariates; moreover, greater exposure to noneducational programs and fast speed programs was also associated with lower total ASQ-3 scores.

The results showed that the mean (SD) total screen time and tablet screen time was 111.6 (140.9) minutes and 31.1 (42.5) minutes per day for children aged 34 months 16 days to 50 months 30 days, 172.7 (286.5) minutes and 41.2 (45.5) minutes per day for children aged 51 months 0 days to 56 months 30 days, and 159.3 (191.5) minutes and 42.4 (52.3) minutes per day for children aged 57 months 0 days to 66 months 0 days; these findings are similar to previous studies [32-34]. Whether educational programs are beneficial for early development in toddlers and preschool children is still controversial. This study also did not observe a positive or negative association; however, a significant finding was revealed that exposure to noneducational content was negatively associated with early development (total score on the ASQ-3) or fine motor in preschool children included in this monitoring study. Although there is little direct evidence of a relationship between the ASQ score and its domain scores, several longitudinal studies have shown that exposure to noneducational content before 3 years of age is negatively related to executive functioning and cognitive development at older ages, which are fundamental to the development of fine motor [35-37]. Another longitudinal study confirmed that screen use time was negatively correlated with fine motor skills 1 year after birth among children aged 3-6 years by cross-lagged path analysis [38]. The earlier cross-lagged conclusion may be explained by the results of this study, in which screen time in relation to categories containing both educational and noneducational content was still negatively associated with fine motor skills because it is inevitable to avoid noneducational content for preschoolers.

This study also revealed that exposure to fast-paced content (more than 6 times per minute) was negatively associated with the total ASQ score and fine motor score, while these relationships were not found for low-paced (less than 2 times per minute) or medium-paced (2 to 6 times per minute) content. Furthermore, the daily tablet screen time via tablet for fast-paced content (76.4 minutes) was far greater than that for medium- (8 minutes) and slow-paced content (9.1 minutes); namely, the predominant pace of the screen content for children aged 2-6 years was fast, which is consistent with the findings of previous studies showing that preschool children often view fast-paced content, with rapid screenshots and angle changes [18]. Several controlled experimental studies have shown that watching fast-paced television significantly impairs executive function immediately in children aged 4-6 years [17,18], and these test programs could appropriately reflect fine motor, problem-solving, communication, and personal-social skills.

Touch-enabled media technological devices (tablets and smartphones) play an important role not only because of their growing popularity among children but also because of their influence on early development; the use of tablets and smartphones was associated with decreased scores on the ASQ and its domains in this study, while other media did not show this relationship. Although exposure to touch-enabled content was not associated with the ASQ score, exposure to static content was negatively associated with gross motor and fine motor scores. There is some experimental evidence that toddlers can more readily learn from touchscreen devices than from televisions [39], possibly because the use of tablets is beneficial for motor development in preschoolers [40].

Animated content is widespread in children's media, and even realistic programs incorporate unrealistic elements. This study revealed that exposure to fantastic and both fantastic and realistic content was negatively associated with the fine motor score; these results are consistent with previous experimental studies [17]. In the working memory system, the content of screen stimuli is encoded, processed, and stored, while animated content is difficult to process because there is no stored schematic to use for interpreting animated content, especially for children's immature brains [17,41]. Engagement in such activities could lead to a decrease in local neurotransmitters in the lateral prefrontal cortex, subsequently resulting in a reduced availability of these neurotransmitters [42].

This study also revealed that general audience and PG content were associated with lower fine motor scores, while PG-13 content was associated with lower gross motor scores. A plausible explanation regarding this result may be that regardless of what the suggested audience of the screen content is, screen use can replace the time spent on physical activity [12,43]. A previous study revealed that the viewing of PG-13 and restricted content was associated with higher hyperactivity and aggression scores and a lower social skills rating. However, this association was not found in early childhood [44].

Limitations and Strengths

To our knowledge, this study was the first to examine preschool children's tablet screen time for different content types via 7-day intelligent monitoring records. By analyzing a sample of representative kindergarteners, we analyzed the associations between screen time and exposure to content types and early development. Several limitations of this study, however, should be mentioned.

First, because of the 7-day monitoring design, we could not calculate the causal relationship between screen exposure and early development. Longitudinal prospective studies and randomized clinical trials are needed to determine causal and cause-and-effect relationships regarding genetic and environmental factors. Second, this monitoring study could

XSL•FO
RenderX

accurately measure tablet screen exposure, while other media, such as televisions and smartphones, were measured via questionnaires on the frequency of different types of screen use; thus, the HSVC application should be appropriately used in other devices, such as smartphones or game machines. Third, although the participants were from developed provinces in China, we recruited kindergarteners from urban and suburban areas, which might influence the generalizability of the results to lower-income districts because all areas in Shanghai have relatively high incomes compared with the national average income level. Fourth, in the screen time assessment collected by questionnaires, a limitation was that no distinction was made between weekends and weekdays in the screen time question asked to families and the real-time screen time calculation, which may result in the existence of deviations in the estimation of the screen time. Fifth, although we informed parents to let children use tablets according to their usual habits at home, this process might still increase or change children's screen exposure behaviors to some extent. In future studies, we will explore ways to make parents accept the installation of intelligent monitoring technology on their own tablets and smartphones so as to restore children's screen exposure behaviors to the greatest extent possible. Finally, although this survey investigated the age intervals of children according to the age versions of ASQ-3, this study failed to investigate the specific dates of birth of children. Therefore, it may affect the rigor of statistical analysis to some extent.

Conclusions

This monitoring study showed that screen time associated with exposure to different content types was significantly associated with early development among preschool children, especially for noneducational, static, fast-paced, and animated content, which were associated not only with the total ASQ score but with gross motor and fine motor scores. The highest priority should be placed on limiting the aggregated screen time of noneducational, static, fast-paced, and animated content programs or games, which provide insight into defining high-quality screen content in guidelines.

Acknowledgments

The authors are grateful to Huawei for developing intelligent software for children's screen exposure. The authors also thank the excellent work of all kindergarten teachers from the Pudong New District and Xuhui District and all study participants. This work was supported by the National Natural Science Foundation of China (grant 82304260); China Postdoctoral Science Foundation (grant 2023M730626); and Shanghai Municipal Health Commission (grant GWV-10.2-YQ23). The 6th Round Three-Year Initiative Plan for Strengthening Public Health System Construction in Shanghai (project GWVI-6).

Data Availability

The data that support the findings of this study are available on request from the corresponding author upon reasonable request.

Authors' Contributions

YJ had full access to all of the data in the study, took responsibility for the integrity of the data and the accuracy of the data analysis, participated in concept and design, critical review of the manuscript for important intellectual content, obtained funding, provided administrative, technical, or material support, and supervised the study. S Lin participated in concept and design, acquisition, analysis, or interpretation of data, and critical review of the manuscript. HC participated in the concept and design, acquisition, analysis, or interpretation of data, drafting of the manuscript, critical review of the manuscript, statistical analysis, obtained funding, and supervised the study. YS participated in concept and design, acquisition, analysis, or interpretation of data, drafting of the manuscript, critical review of the manuscript, and statistical analysis. S Luo participated in concept and design, acquisition, analysis, or interpretation of data, critical review of the manuscript, provided administrative, technical, or material support, and supervised the study. YM had full access to all data, took data-related responsibilities, participated in concept and design, acquisition, analyses, or interpretation of data, critical review of the manuscript, provided administrative, technical, or material support, and supervised the study. CL participated in the acquisition, analysis, or interpretation of data, and critical review of the manuscript. YX participated in the acquisition, analysis, or interpretation of data, statistical analysis, and critical review of the manuscript. YZ participated in the acquisition, analysis, or interpretation of data, and critical review of the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables and figures.
[DOCX File , 559 KB - jmir_v27i1e65343_app1.docx]

References

XSL•FO
RenderX

1. Christakis DA. Interactive media use at younger than the age of 2 years: Time to rethink the American Academy of Pediatrics guideline? JAMA Pediatr 2014;168(5):399-400. [doi: 10.1001/jamapediatrics.2013.5081] [Medline: 24615347]
2. Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550536 [accessed 2025-02-07]
3. 24-hour movement guidelines – birth to 5 years – brochure. Australian Government Department of Health and Aged Care. 2016. URL: https://www.health.gov.au/resources/publications/24-hour-movement-guidelines-birth-to-5-years-brochure?language=en [accessed 2025-02-07]
4. Tremblay MS, Chaput J, Adamo KB, Aubert S, Barnes JD, Choquette L, et al. Canadian 24-hour movement guidelines for the early years (0-4 years): an integration of physical activity, sedentary behaviour, and sleep. BMC Public Health 2017;17:874 [FREE Full text] [doi: 10.1186/s12889-017-4859-6] [Medline: 29219102]
5. McArthur BA, Volkova V, Tomopoulos S, Madigan S. Global prevalence of meeting screen time guidelines among children 5 years and younger: a systematic review and meta-analysis. JAMA Pediatr 2022;176(4):373-383 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.6386] [Medline: 35157028]
6. Madigan S, McArthur BA, Anhorn C, Eirich R, Christakis DA. Associations between screen use and child language skills: a systematic review and meta-analysis. JAMA Pediatr 2020;174(7):665-675 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.0327] [Medline: 32202633]
7. Madigan S, Browne D, Racine N, Mori C, Tough S. Association between screen time and children's performance on a developmental screening test. JAMA Pediatr 2019;173(3):244-250 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5056] [Medline: 30688984]
8. Li X, Vanderloo LM, Keown-Stoneman CDG, Cost KT, Charach A, Maguire JL, et al. Screen use and mental health symptoms in Canadian children and youth during the COVID-19 pandemic. JAMA Netw Open 2021;4(12):e2140875 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.40875] [Medline: 34962557]
9. Fakhouri THI, Hughes JP, Brody DJ, Kit BK, Ogden CL. Physical activity and screen-time viewing among elementary school-aged children in the United States from 2009 to 2010. JAMA Pediatr 2013;167(3):223-229. [doi: 10.1001/2013.jamapediatrics.122] [Medline: 23303439]
10. Oude Groeniger J, de Koster W, van der Waal J. Time-varying effects of screen media exposure in the relationship between socioeconomic background and childhood obesity. Epidemiology 2020;31(4):578-586 [FREE Full text] [doi: 10.1097/EDE.0000000000001210] [Medline: 32483068]
11. Janssen X, Martin A, Hughes AR, Hill CM, Kotronoulas G, Hesketh KR. Associations of screen time, sedentary time and physical activity with sleep in under 5s: a systematic review and meta-analysis. Sleep Med Rev 2020;49:101226 [FREE Full text] [doi: 10.1016/j.smrv.2019.101226] [Medline: 31778942]
12. Pedersen J, Rasmussen MGB, Sørensen SO, Mortensen SR, Olesen LG, Brønd JC, et al. Effects of limiting recreational screen media use on physical activity and sleep in families with children: a cluster randomized clinical trial. JAMA Pediatr 2022;176(8):741-749 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.1519] [Medline: 35604678]
13. Eirich R, McArthur BA, Anhorn C, McGuinness C, Christakis DA, Madigan S. Association of screen time with internalizing and externalizing behavior problems in children 12 years or younger: a systematic review and meta-analysis. JAMA Psychiatry 2022;79(5):393-405 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0155] [Medline: 35293954]
14. Guellai B, Somogyi E, Esseily R, Chopin A. Effects of screen exposure on young children's cognitive development: a review. Front Psychol 2022;13:923370 [FREE Full text] [doi: 10.3389/fpsyg.2022.923370] [Medline: 36059724]
15. Levine LE, Waite BM, Bowman LL, Kachinsky K. Mobile media use by infants and toddlers. Comput Hum Behav 2019;94:92-99. [doi: 10.1016/j.chb.2018.12.045]
16. Linebarger DL, Walker D. Infants' and toddlers' television viewing and language outcomes. Am Behav Sci 2005;48(5):624-645. [doi: 10.1177/0002764204271505]
17. Lillard AS, Drell MB, Richey EM, Boguszewski K, Smith ED. Further examination of the immediate impact of television on children's executive function. Dev Psychol 2015;51(6):792-805. [doi: 10.1037/a0039097] [Medline: 25822897]
18. Lillard AS, Peterson J. The immediate impact of different types of television on young children's executive function. Pediatrics 2011;128(4):644-649 [FREE Full text] [doi: 10.1542/peds.2010-1919] [Medline: 21911349]
19. Thorell LB, Lindqvist S, Bergman Nutley S, Bohlin G, Klingberg T. Training and transfer effects of executive functions in preschool children. Dev Sci 2009;12(1):106-113. [doi: 10.1111/j.1467-7687.2008.00745.x] [Medline: 19120418]
20. Subrahmanyam K, Renukarya B. Digital games and learning: identifying pathways of influence. Educ Psychol 2016;50(4):335-348. [doi: 10.1080/00461520.2015.1122532]
21. Chen W, Adler JL. Assessment of screen exposure in young children, 1997 to 2014. JAMA Pediatr 2019;173(4):391-393 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5546] [Medline: 30776061]
22. Sanders T, Noetel M, Parker P, Del Pozo Cruz B, Biddle S, Ronto R, et al. An umbrella review of the benefits and risks associated with youths' interactions with electronic screens. Nat Hum Behav 2024;8(1):82-99. [doi: 10.1038/s41562-023-01712-8] [Medline: 37957284]
23. Chen B, Bernard JY, Padmapriya N, Ning Y, Cai S, Lança C, et al. Associations between early-life screen viewing and 24 hour movement behaviours: findings from a longitudinal birth cohort study. Lancet Child Adolesc Health 2020;4(3):201-209. [doi: 10.1016/S2352-4642(19)30424-9] [Medline: 32004497]

XSL•FO
RenderX

24. Huang XN, Zhang Y, Feng WW, Wang HS, Cao B, Zhang B, et al. [Reliability and validity of warning signs checklist for screening psychological, behavioral and developmental problems of children]. Zhonghua Er Ke Za Zhi 2017;55(6):445-450. [doi: 10.3760/cma.j.issn.0578-1310.2017.06.010] [Medline: 28592013]
25. Chen H, Li X, Gao J, Liu X, Mao Y, Wang R, et al. Health belief model perspective on the control of COVID-19 vaccine hesitancy and the promotion of vaccination in China: web-based cross-sectional study. J Med Internet Res 2021;23(9):e29329 [FREE Full text] [doi: 10.2196/29329] [Medline: 34280115]
26. Film ratings. Motion Picture Association. URL: https://www.motionpictures.org/film-ratings/ [accessed 2025-02-07]
27. Bian XY, Yao GY, Squires J, Wei M, Chen CI, Fang BH. [Studies of the norm and psychometric properties of ages and stages questionnaires in shanghai children]. Zhonghua Er Ke Za Zhi 2010;48(7):492-496. [Medline: 21055084]
28. Shuffrey L, Firestein M, Kyle M. Association of birth during the COVID-19 pandemic with neurodevelopmental status at 6 months in infants with and without in utero exposure to maternal SARS-CoV-2 infection. JAMA Pediatrics 2022;176(6). [doi: 10.3410/f.741406907.793592121]
29. Squires JTE, Bricker D, Potter L. (ASQ-3) Ages & Stages Questionnaires. Washington, DC: Brookes Publishing; 2009.
30. Squires J, Bricker D, Potter L. Revision of a parent-completed development screening tool: ages and stages questionnaires. J Pediatr Psychol 1997;22(3):313-328. [doi: 10.1093/jpepsy/22.3.313] [Medline: 9212550]
31. Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. J Dev Behav Pediatr 2011;32(7):499-511. [doi: 10.1097/DBP.0b013e31822552e9] [Medline: 21760526]
32. Wang H, Zhao J, Yu Z, Pan H, Wu S, Zhu Q, et al. Types of on-screen content and mental health in kindergarten children. JAMA Pediatr 2024;178(2):125-132. [doi: 10.1001/jamapediatrics.2023.5220] [Medline: 38048076]
33. Neville RD, McArthur BA, Eirich R, Lakes KD, Madigan S. Bidirectional associations between screen time and children's externalizing and internalizing behaviors. J Child Psychol Psychiatry 2021;62(12):1475-1484. [doi: 10.1111/jcpp.13425] [Medline: 33945153]
34. Goncalves WSF, Byrne R, de Lira PIC, Viana MT, Trost SG. Adherence to 24-hour movement guidelines among rural Brazalian preschool children: associations with parenting practices. Int J Behav Nutr Phys Act 2022;19(1):133 [FREE Full text] [doi: 10.1186/s12966-022-01369-y] [Medline: 36271449]
35. Christakis DA, Zimmerman FJ, DiGiuseppe DL, McCarty CA. Early television exposure and subsequent attentional problems in children. Pediatrics 2004;113(4):708-713. [doi: 10.1542/peds.113.4.708] [Medline: 15060216]
36. Landhuis CE, Poulton R, Welch D, Hancox RJ. Does childhood television viewing lead to attention problems in adolescence? Results from a prospective longitudinal study. Pediatrics 2007;120(3):532-537. [doi: 10.1542/peds.2007-0978] [Medline: 17766526]
37. Zimmerman FJ, Christakis DA. Children's television viewing and cognitive outcomes: a longitudinal analysis of national data. Arch Pediatr Adolesc Med 2005;159(7):619-625 [FREE Full text] [doi: 10.1001/archpedi.159.7.619] [Medline: 15996993]
38. Martzog P, Suggate SP. Screen media are associated with fine motor skill development in preschool children. Early Child Res Q 2022;60:363-373. [doi: 10.1016/j.ecresq.2022.03.010]
39. Kirkorian HL, Choi K, Pempek TA. Toddlers' word learning from contingent and noncontingent video on touch screens. Child Dev 2016;87(2):405-413. [doi: 10.1111/cdev.12508] [Medline: 27018327]
40. Souto PHS, Santos JN, Leite HR, Hadders-Algra M, Guedes SC, Nobre JNP, et al. Tablet use in young children is associated with advanced fine motor skills. J Mot Behav 2020;52(2):196-203. [doi: 10.1080/00222895.2019.1602505] [Medline: 31007146]
41. Lillard AS, Li H, Boguszewski K. Television and children's executive function. Adv Child Dev Behav 2015;48:219-248. [doi: 10.1016/bs.acdb.2014.11.006] [Medline: 25735946]
42. Brzezicka A, Kamiński J, Wróbel A. Local resource depletion hypothesis as a mechanism for action selection in the brain. Behav Brain Sci 2013;36(6):682-683. [doi: 10.1017/S0140525X13000940] [Medline: 24304779]
43. Yuan R, Zhang J, Song P, Qin L. The relationship between screen time and gross motor movement: a cross-sectional study of pre-school aged left-behind children in China. PLoS One 2024;19(4):e0296862 [FREE Full text] [doi: 10.1371/journal.pone.0296862] [Medline: 38578800]
44. Conners-Burrow NA, McKelvey LM, Fussell JJ. Social outcomes associated with media viewing habits of low-income preschool children. Early Educ Dev 2011;22(2):256-273. [doi: 10.1080/10409289.2011.550844]

Abbreviations

ASQ: Ages and Stages Questionnaire
ASQ-3: Ages and Stages Questionnaire, Third Edition
HSVC: Healthy Screen Viewing for Children
PG: parental guidance

XSL•FO
RenderX

Edited by A Coristine; submitted 13.08.24; peer-reviewed by ME Nergiz, A Beynon; comments to author 25.11.24; revised version received 23.12.24; accepted 31.01.25; published 05.03.25.

Please cite as:
Chen H, Sun Y, Luo S, Ma Y, Li C, Xiao Y, Zhang Y, Lin S, Jia Y
Association of Screen Content With Early Development Among Preschoolers in Shanghai: 7-Day Monitoring Study With Auto Intelligent Technology
J Med Internet Res 2025;27:e65343
URL: https://www.jmir.org/2025/1/e65343
doi:10.2196/65343
PMID:40053802

©Hao Chen, Yi Sun, Sha Luo, Yingyan Ma, Chenshu Li, Yingcheng Xiao, Yimeng Zhang, Senlin Lin, Yingnan Jia. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Monitoring Sleep Quality Through Low α-Band Activity in the Prefrontal Cortex Using a Portable Electroencephalogram Device: Longitudinal Study

Chuanliang Han[1*], PhD; Zhizhen Zhang[2*], MPhil; Yuchen Lin[3*], PhD; Shaojia Huang[4], MBA; Jidong Mao[4], BSc; Weiwen Xiang[4], BSc; Fang Wang[4], BEd; Yuping Liang[4], BSc; Wufang Chen[4], BSc; Xixi Zhao[5,6], PhD

[1]School of Biomedical Sciences and Gerald Choa Neuroscience Institute, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Department of Mathematics and Statistics, University of Massachusetts at Amherst, Amherst, MA, United States

[3]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[4]Shenzhen Shuimu AI Technology Co., Ltd, Shenzhen, China

[5]National Clinical Research Center for Mental Disorders & National Center for Mental Disorders, Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China

[6]Advanced Innovation Center for Human Brain Protection and Laboratory for Clinical Medicine,, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xixi Zhao, PhD
National Clinical Research Center for Mental Disorders & National Center for Mental Disorders, Beijing Key Laboratory of Mental Disorders
Beijing Anding Hospital
Capital Medical University
No. 5 Ankang Hutong
Xicheng District
Beijing, 100088
China
Phone: 86 15501193896
Email: zhaoxixi@ccmu.edu.cn

Abstract

Background: The pursuit of sleep quality has become an important aspect of people's global quest for overall health. However, the objective neurobiological features corresponding to subjective perceptions of sleep quality remain poorly understood. Although previous studies have investigated the relationship between electroencephalogram (EEG) and sleep, the lack of longitudinal follow-up studies raises doubts about the reproducibility of their findings.

Objective: Currently, there is a gap in research regarding the stable associations between EEG data and sleep quality assessed through multiple data collection sessions, which could help identify potential neurobiological targets related to sleep quality.

Methods: In this study, we used a portable EEG device to collect resting-state prefrontal cortex EEG data over a 3-month follow-up period from 42 participants (27 in the first month, 25 in the second month, and 40 in the third month). Each month, participants' sleep quality was assessed using the Pittsburgh Sleep Quality Index (PSQI) to estimate their recent sleep quality.

Results: We found that there is a significant and consistent positive correlation between low α band activity in the prefrontal cortex and PSQI scores (r=0.45, P<.001). More importantly, this correlation remained consistent across all 3-month follow-up recordings (P<.05), regardless of whether we considered the same cohort or expanded the sample size. Furthermore, we discovered that the periodic component of the low α band primarily contributed to this significant association with PSQI.

Conclusions: These findings represent the first identification of a stable and reliable neurobiological target related to sleep quality through multiple follow-up sessions. Our results provide a solid foundation for future applications of portable EEG devices in monitoring sleep quality and screening for sleep disorders in a broad population.

(J Med Internet Res 2025;27:e67188) doi:10.2196/67188

XSL•FO
RenderX

KEYWORDS

EEG; electroencephalogram; alpha oscillation; prefrontal cortex; sleep; portable device

Introduction

Sleep is a fundamental behavior closely related to daily life, and sleep quality plays a crucial role in human health [1-4]. Poor sleep quality often leads to various types of physical ailments [5-7]. Previous studies have conducted systematic research across different species (flies [8], rodents [9-14], cats [15,16], nonhuman primates [17], and humans [18-21]) to characterize the neural mechanisms across multiple stages of sleep. However, there are still many challenges in translating findings from animal experiments to humans, and sleep remains a complex concept. Our understanding of the relationship between sleep quality and the brain is still incomplete.

Electroencephalogram (EEG) is widely used in human studies to investigate sleep-related issues [22-24]. Different types of neural oscillations typically represent different stages of sleep [10] and are applied in various research areas. However, sleep quality is a subjective experience, and we hypothesize that this subjective perception stems from the brain's (eg, prefrontal cortex) daily neural responses. If this hypothesis holds, then under resting-state conditions, we should be able to identify certain neurobiological characteristics in the brain that are significantly associated with subjective sleep quality. Neural oscillations [25,26] are prominent biological features in EEG [27-31], such as α oscillations (8-13 Hz). Previous studies have explored the association between various types of neural oscillations and sleep quality [32-37]. However, much of the current research progress relies on single-session data collection, and there is a lack of follow-up studies. Hence, the stability of these associations remains unclear.

Therefore, in this study, to test the above hypothesis, we used a portable EEG device to collect resting-state EEG data (both open-eye and closed-eye conditions) from participants while also assessing their sleep quality (Pittsburgh Sleep Quality Index [PSQI]). The study involved a 3-month follow-up, with a total of 42 participants. In the first month, 27 people participated, followed by 25 people in the second month, and 40 people in the third month. A total of 22 participants completed all 3 sessions. We performed spectral analysis and correlation analysis on the data and further explored the periodic and aperiodic components of the spectra.

Methods

Participants

A total of 42 right-handed healthy individuals (19 females and 23 males; age: 34.9, SD 8.0 years) were recruited for this study, who were mainly from nearby communities through posters posted on the publicity walls. The sample size was calculated by G*Power software (Heinrich-Heine-Universität Düsseldorf), the sample size is calculated with 28 (given effect size of 0.65, α of 0.05, and β of 0.05 in the difference between means [matched pairs in eyes-open and eyes-closed state]). Hence, on average, we recruited around 30 participants each month. In the first month, 27 people participated, followed by 25 people in the second month, and 40 people in the third month. A total of 22 participants participated in all 3 months, 24 people participated in months 1 and 2, 25 people participated in months 1 and 3, and 23 people participated in months 2 and 3.

Ethical Considerations

Informed consent was obtained from all participants. All procedures were performed according to the National Institutes of Health Guidelines, and the research protocol was approved by the Shenzhen Institute of Advanced Science and Technology, Chinese Academy of Science (SIAT-IRB-240915-H0913). The study data are anonymous. The participant compensation is calculated based on the duration of the experiment (US $14.1 per hour) each month. The timer starts when the participant begins preparation for the experiment and stops when the experiment ends (after EEG data collection and completion of the PSQI questionnaire). The total duration of each experiment is typically 15-20 minutes.

Neurophysiological Recording

Each month, resting-state EEG in both open-eye and closed-eye states (counterbalanced order) was recorded (3 minutes each) using a 2-electrode (FP1 and FP2) setup (Brain Pro, Shenzhen Shuimu AI Technology Co Ltd, China). During the recording, the reference electrode was set at the earlobe, and real-time filtering was implemented with a bandpass filter with cutoff frequencies of 0.5 and 30 Hz. EEG data are wirelessly transmitted to a computer via Bluetooth. Each participant collected data in the morning, usually 2-3 hours after waking up. The sampling rate was 512 Hz and the scalp impedance was maintained below 10 kΩ for the electrode. Eye-blink artifacts were removed with unsupervised machine-learning algorithms [38].

Estimation of Sleep Quality

The PSQI [39] was used to evaluate the participants' sleep quality after monthly EEG recordings. The evaluation included several indicators: sleep quality, time to fall asleep, sleep duration, sleep disorders, daytime dysfunction, and total sleep score. Each factor is scored in 4 levels based on a score of 0-3. The cumulative score for each factor was summed to yield the PSQI total score (0-21 points), with a higher total score indicating poorer sleep quality.

Power Spectrum Analysis

The power spectrum of the EEG response was estimated using the multitaper method [40,41] (time-bandwidth product, 3; tapers, 5; Chronux toolbox, which was implemented using custom software written in MATLAB. Similar methods have been used in several studies in the field of biomedical science (eg, biological science [42-45], medical science [28,30], and public health [46-50]). Relative power can be defined as Equation 1 (Figure 1C).

XSL•FO
RenderX

Figure 1. Three-month follow-up electroencephalogram recording using portable device. (A) Averaged power spectrum in FP1 during months 1, 2, and 3 under both open-eye and closed-eye states. (B) Pairwise comparison of relative α power in FP1 between open-eye and closed-eye states in months 1, 2, and 3, with *** indicating *P*<.001. (C) Comparison of relative α power in FP1 between months 1, 2, and 3 in open-eye and closed-eye states, respectively. (D) Averaged power spectrum in FP2 during months 1, 2, and 3 under both open-eye and closed-eye states. (E) Pairwise comparison of relative α power in FP2 between open-eye and closed-eye states in months 1, 2, and 3, with *** indicating *P*<.001. (F) Comparison of relative α power in FP2 between months 1, 2, and 3 in open-eye and closed-eye states, respectively.

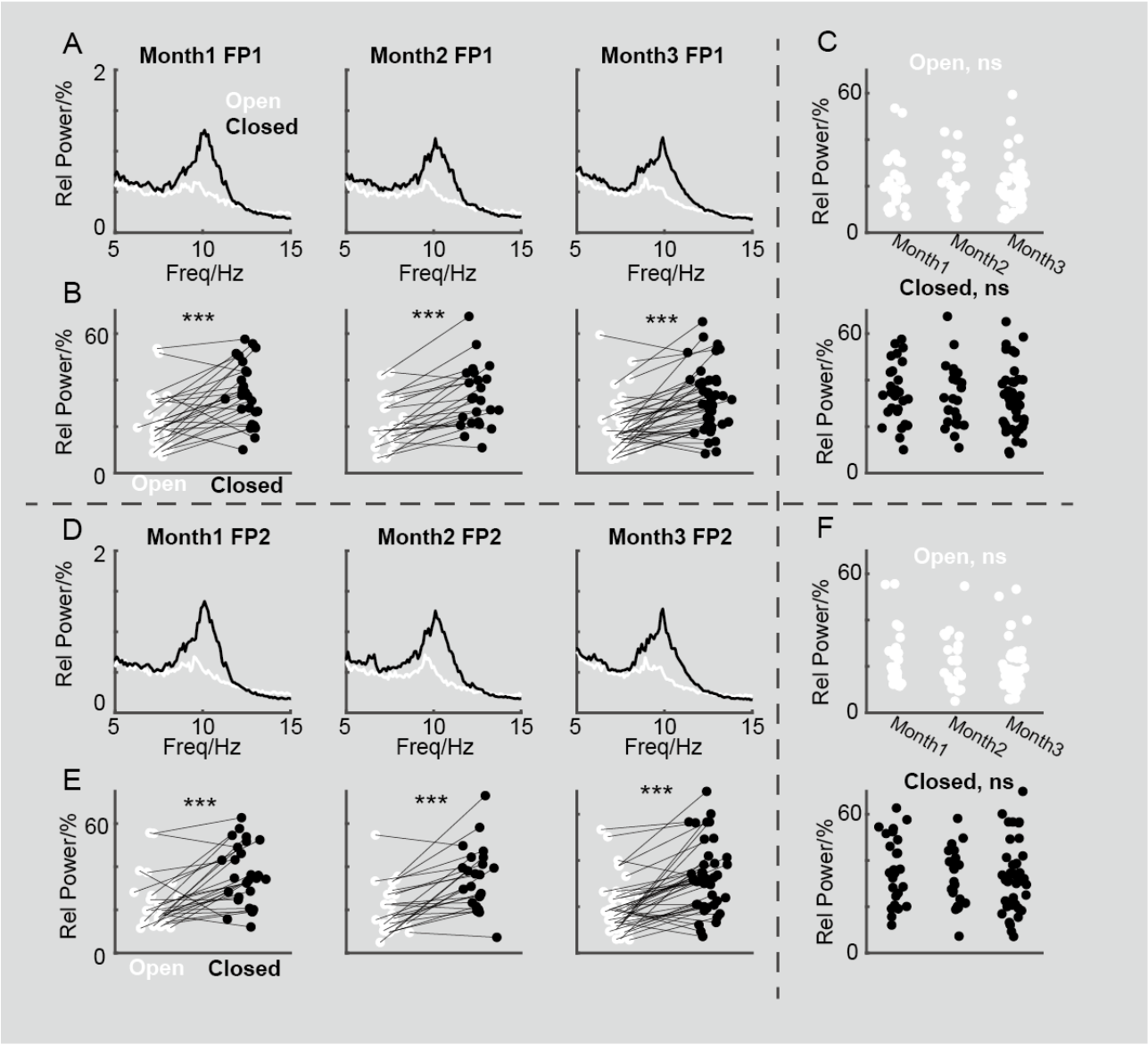

(1)

Descriptive Model for Dissecting Aperiodic and Periodic Activity

The power spectrum was considered as the sum of 2 components: aperiodic and periodic components. Aperiodic activity in the power spectrum refers to the frequency components of brain activity that do not follow a regular periodic or oscillatory pattern. It does not have a clear, repeating waveform or regular frequency. This activity is typically more broadband and can be described by a continuous distribution of power across a range of frequencies. Periodic activity in the power spectrum refers to frequency components of brain signals that exhibit a regular, repetitive oscillatory pattern over time. These oscillations have specific, identifiable peaks in the power spectrum at distinct frequencies. The aperiodic component of the power spectrum was extracted using a 1/f-like function (Equation 2), where a, b, c, and d represent the model parameters to be estimated. The periodic component was defined as the residual obtained by subtracting the aperiodic component from the raw power spectrum. This method has been used in the characterization of γ-band and α-band activities [51-53].

Statistical Analysis

In Figures 1B and 1E, a pairwise *t* test was examined to test the differences in relative α power between open-eye and closed-eye states in 3 months respectively. Similarly, in Figure 2C, a pairwise *t* test was used to test the differences in relative α

XSL•FO
RenderX

power associated with aperiodic activity between these states over the same time period. In Figure 2D, a pairwise *t* test was used to evaluate the differences in relative α power related to periodic activity between open-eye and closed-eye states in 3 months respectively. In Figures 1C and 1F, a *t* test with Bonferroni correction was used to compare the differences in relative α power among 3 months for both open-eye and closed-eye states. Additionally, in Figures 3A and 3B and Figures 4A and 4B, Pearson correlation analysis was conducted to measure the relationship between relative power at each frequency and the PSQI. Figures 3C-E and 4C-E display Pearson correlations between PSQI and relative power in low α (LA, 7-8.5 Hz), medium α (MA, 9-10.5 Hz), and high α (HA, 11-13 Hz) bands, respectively. Figures 5 and 6 illustrate the Pearson correlation between PSQI and relative power in aperiodic and periodic activity across the LA, MA, and HA bands, respectively.

Figure 2. Dissection of aperiodic and periodic activity. (A) Averaged raw power spectrum showing aperiodic activities in both open-eye and closed-eye states over a 3-month period. (B) Averaged periodic activities in both open-eye and closed-eye states over a 3-month period. (C) Pairwise comparisons of relative α power in aperiodic activity between open-eye and closed-eye states in months 1, 2, and 3. (D) Pairwise comparisons of relative α power in periodic activity between open-eye and closed-eye states in months 1, 2, and 3, with *** indicating *P*<.001.

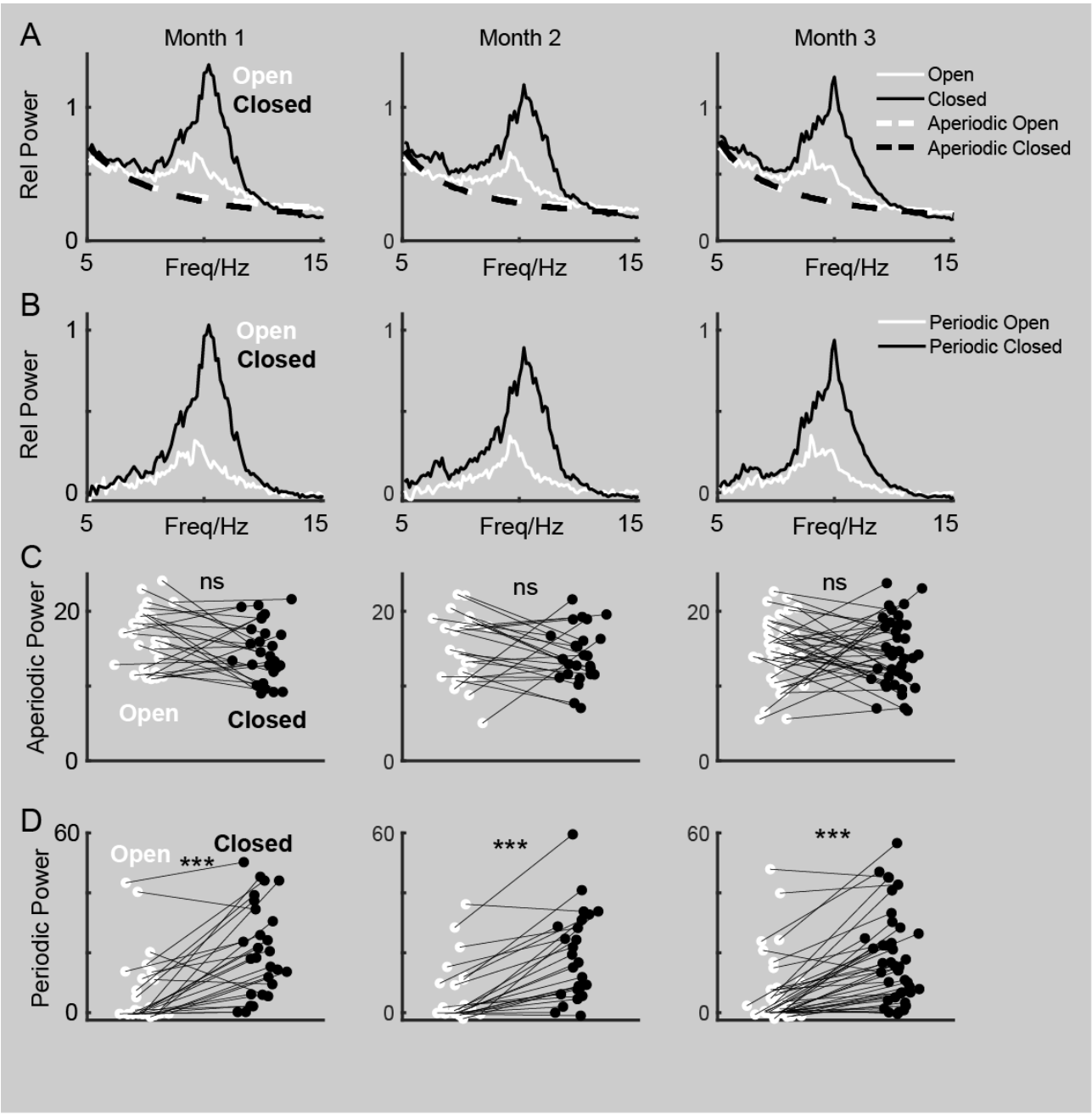

XSL•FO
RenderX

Figure 3. Relationship between PSQI and strength of sub-α oscillations for all participants. (A) Correlation between PSQI and relative power in difference frequencies under open-eye and closed-eye states over 3 months for all participants. The dot indicates significance at *P*<.05. (B) Averaged 3-month correlation between PSQI and relative power in difference frequencies under open-eye and closed-eye states. The dot indicates significance at *P*<.05. (C) Scatter plot showing the relationship between low α power in both open-eye and closed-eye states and PSQI for all participants in months 1, 2, and 3. (D) Scatter plot showing the relationship between medium α power in both open-eye and closed-eye states and PSQI for all participants in months 1, 2, and 3. (E) Scatter plot showing the relationship between high α power in both open-eye and closed-eye states and PSQI for all participants in months 1, 2, and 3. PSQI: Pittsburgh Sleep Quality Index.

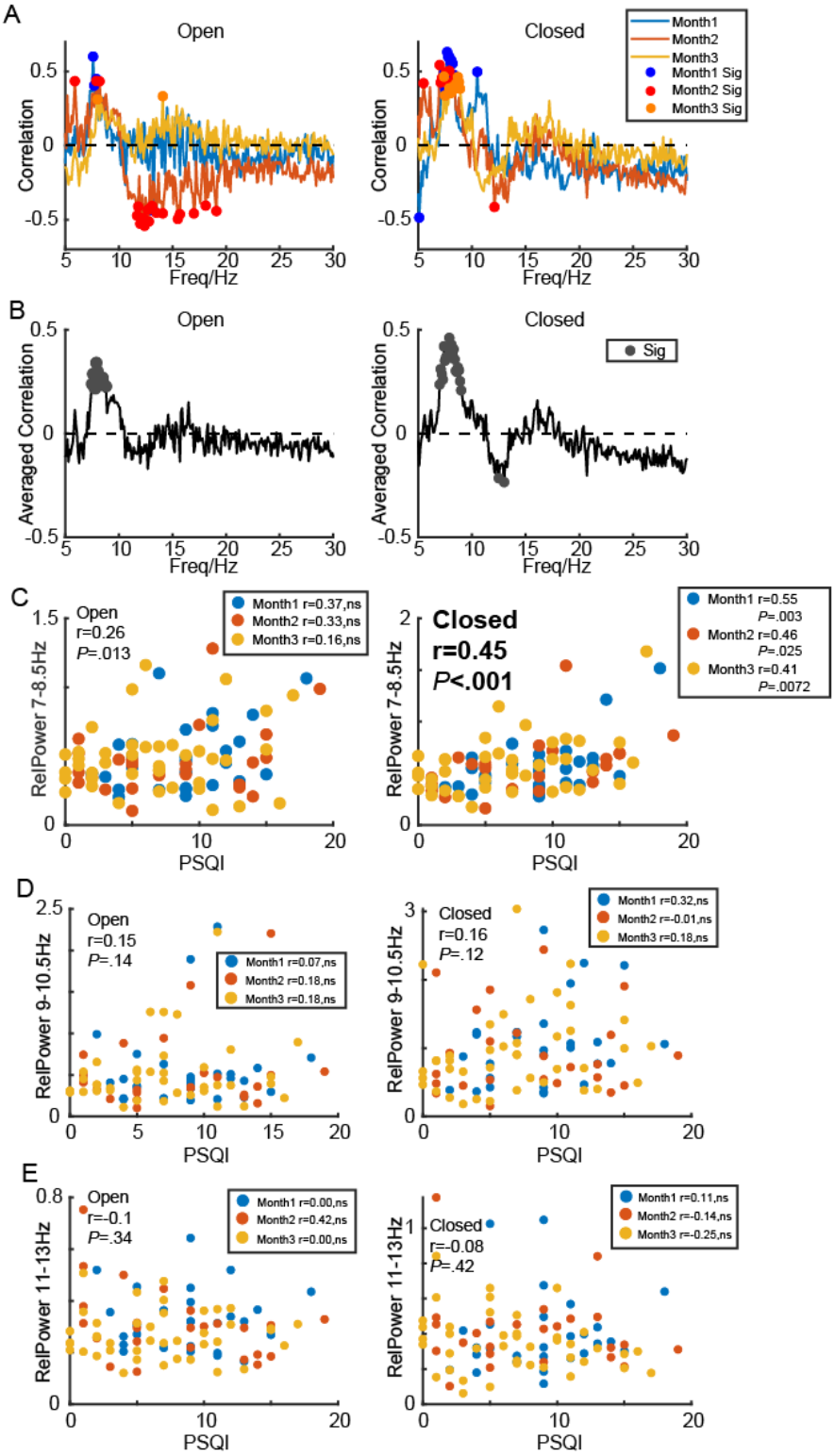

XSL•FO
RenderX

Figure 4. Relationship between PSQI and strength of sub-α oscillations for participants who participated in the all 3-time experiment. (A) Correlation between PSQI and relative power in difference frequencies in open-eye and closed-eye states in 3 months for participants who participated in the all 3-time experiment. The dot indicates significance at *P*<.05. (B) Averaged 3-month correlation between PSQI and relative power in difference frequencies in open-eye and closed-eye states. The dot indicates significance at *P*<.05. (C) Scatter plot showing the relationship between low α power in both open-eye and closed-eye states and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. (D) Scatter plot showing the relationship between medium α power in both open-eye and closed-eye states and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. (E) Scatter plot showing the relationship between high α power in both open-eye and closed-eye states and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. PSQI: Pittsburgh Sleep Quality Index.

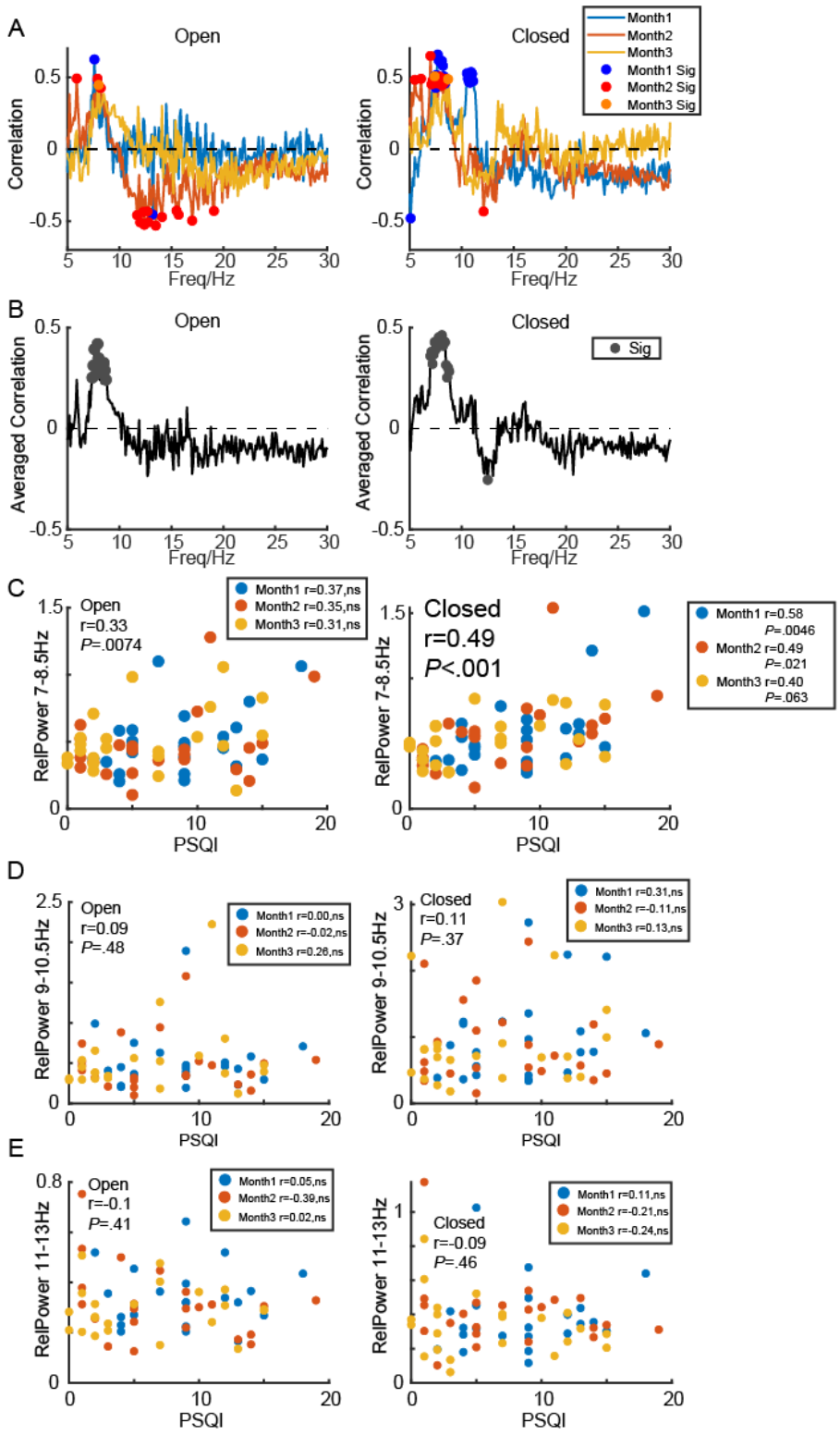

XSL•FO
RenderX

Figure 5. Relationship between PSQI and strength of sub-α power in aperiodic and periodic activity for all participants. (A) Scatter plot showing the relationship between low α power of aperiodic and periodic activity in the open-eye state and PSQI for all participants in months 1, 2, and 3. (B) Scatter plot showing the relationship between low α power of aperiodic and periodic activity in the closed-eye state and PSQI for all participants in months 1, 2, and 3. (C) Scatter plot showing the relationship between medium α power of aperiodic and periodic activity in open-eye and closed-eye states and PSQI for all participants in months 1, 2, and 3. (D) Scatter plot showing the relationship between high α power of aperiodic and periodic activity in open-eye and closed-eye states and PSQI for all participants in months 1, 2, and 3. PSQI: Pittsburgh Sleep Quality Index.

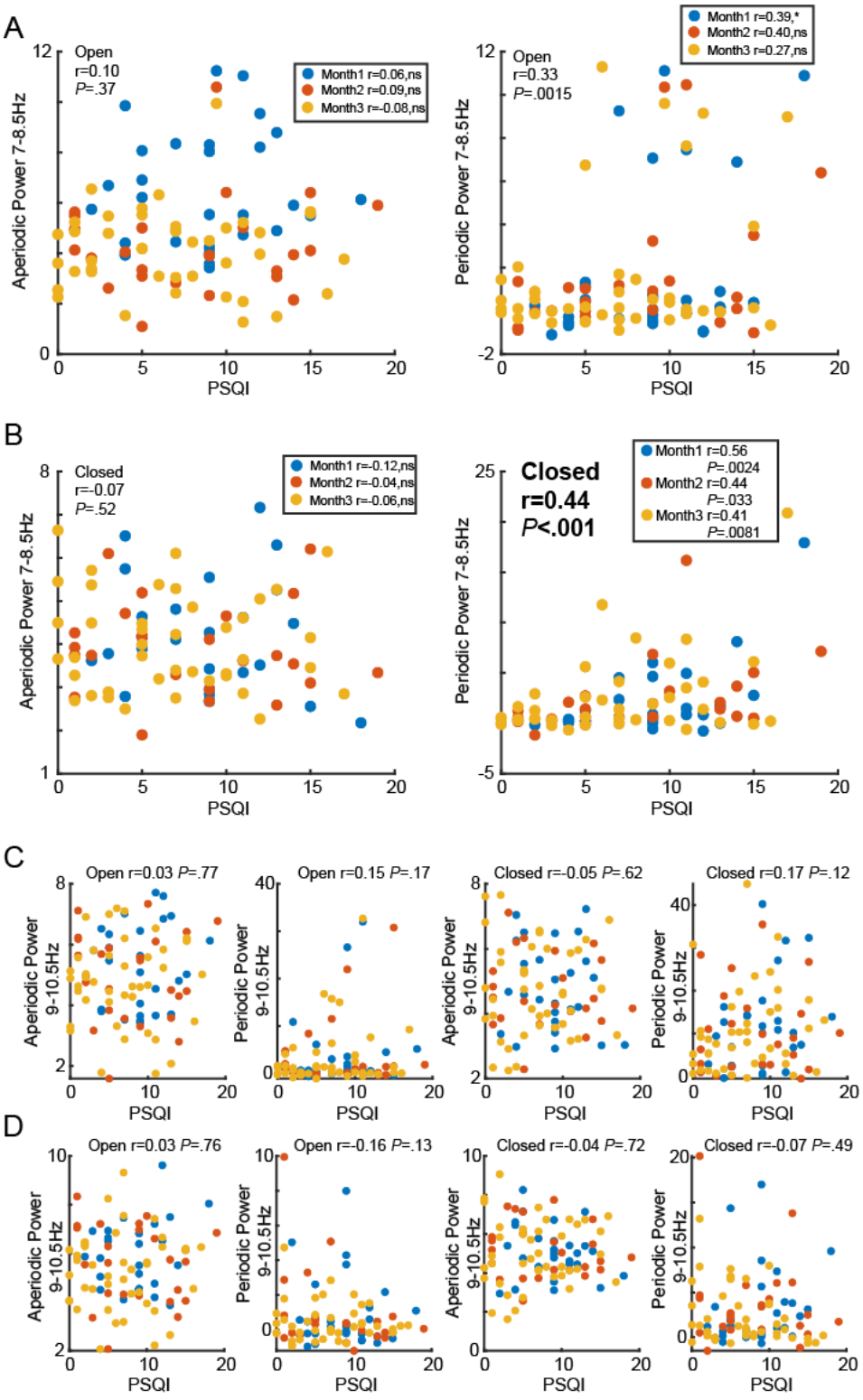

XSL•FO
RenderX

Figure 6. Relationship between PSQI and strength of sub-α power in aperiodic and periodic activity for participants who participated in the all 3-time experiment. (A) Scatter plot showing the relationship between low α power of aperiodic and periodic activity in the open-eye state and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. (B) Scatter plot showing the relationship between low α power of aperiodic and periodic activity in closed-eye states and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. (C) Scatter plot showing the relationship between medium α power of aperiodic and periodic activity in open-eye and closed-eye states and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. (D) Scatter plot showing the relationship between high α power of aperiodic and periodic activity in open-eye and closed-eye states and PSQI for participants who participated in all 3-time experiments in months 1, 2, and 3. PSQI: Pittsburgh Sleep Quality Index.

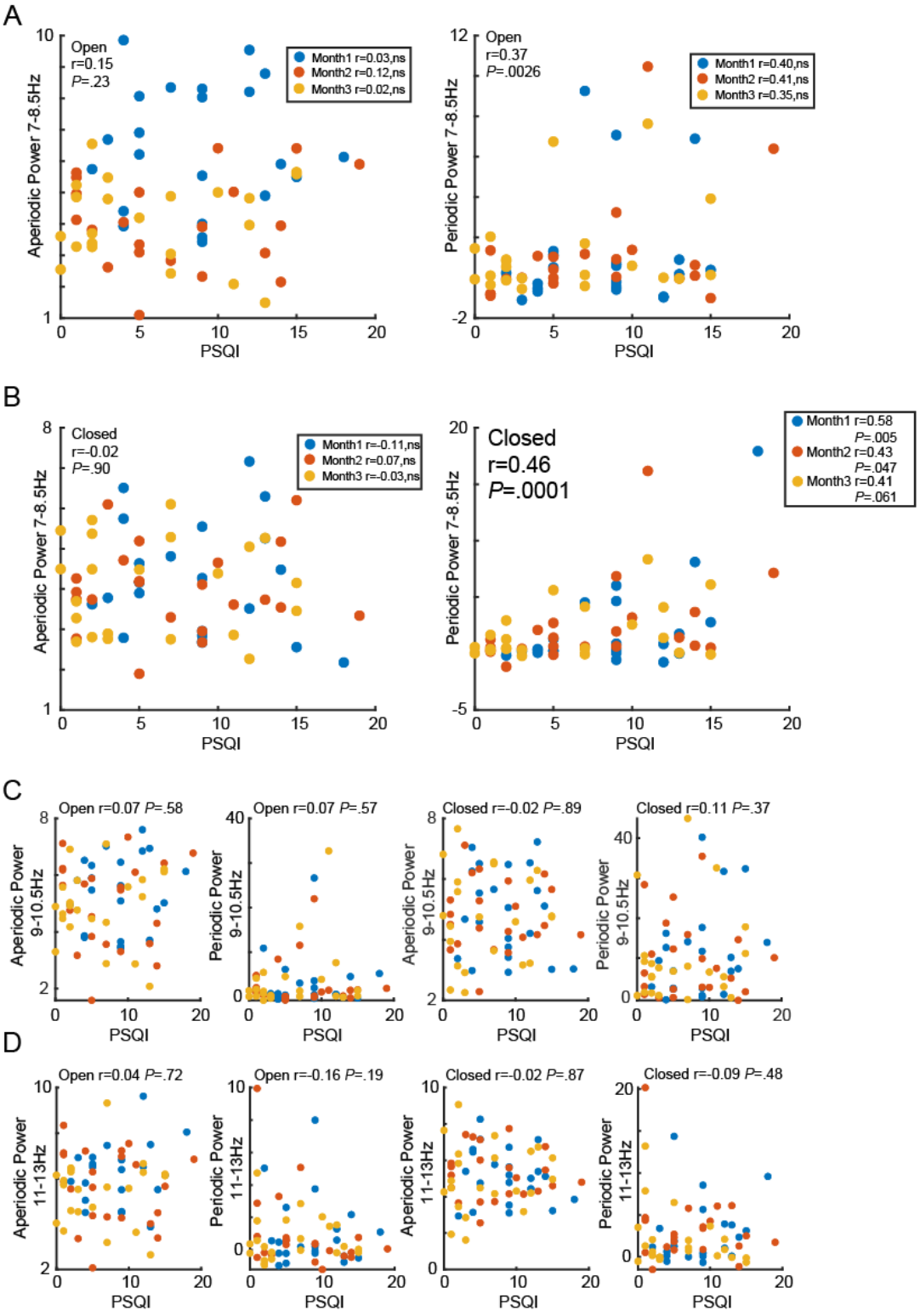

XSL•FO
RenderX

Results

Overview

In this study, we aimed to explore the relationship between different frequency components of portable EEG and sleep quality, and whether a stable neurobiological feature strongly associated with sleep quality can be found through portable EEG. First, we collected a batch of portable EEG data from participants in the first month and identified some features as potential targets that are significantly correlated with sleep quality. However, it cannot be determined whether this result can be repeated and stable. Therefore, we collected 2 additional batches of data in the second and third months, with 22 participants from the 2 follow-up periods also participating in the first data collection. Through this, we verified that the significant relationship between the target feature and sleep quality can be repeated and stable. During these 3 months of data collection, other participants also participated in the data collection.

Stable Neural Recording Using Portable EEG Device in 3 Months

Our data collection lasted for 3 months, with EEG data collected once each month and the PSQI scale used to assess sleep quality. Before conducting the correlation analysis, we needed to verify the signal quality of the portable EEG device, which had 2 recording electrodes in the prefrontal region (FP1 and FP2). We performed frequency analysis on the EEG data collected each month (in both open-eye and closed-eye conditions), and after averaging at the group level, we observed an obvious peak around 10 Hz (FP1, Figure 1A; FP2, Figure 1D), representing α oscillations. The spectral shapes were similar, and the α oscillation intensity in the closed-eye condition was significantly stronger than in the open-eye condition across all 3 months (P<.001; FP1, Figure 1B; FP2, Figure 1E), which is consistent with previous research [54-57]. The relative power in the α band (8-13 Hz) showed no significant differences (P>.05) between open-eye and closed-eye conditions over the 3 months (FP1, Figure 1C; FP2, Figure 1F). Therefore, in subsequent analyses, we will average the spectra of the 2 prefrontal electrodes to represent the neural oscillatory activity in the prefrontal region.

Low α-Band Activity Significant Correlated With Sleep Quality

After ensuring the stability of the portable EEG signal recordings, we first performed a correlation analysis between relative power across different frequency bands (5-30 Hz) and PSQI (Figure 3A, with the first month represented in blue, the second in orange, and the third in yellow). Although significant correlations appeared in multiple frequency bands each month (as indicated by the solid-colored dots), one band consistently showed a significant positive correlation across all 3 months, specifically the LA band (around 7-8.5 Hz), both in open-eye (Figure 3A left) and closed-eye (Figure 3A right) conditions. When we averaged the correlation coefficients across the 3 months, this phenomenon became even more pronounced (Figure 3B).

Based on previous studies [54,58,59], the traditional α band (8-13 Hz) may contain several suboscillations, such as LA, MA, and HA. In the subsequent analysis, we fixed 3 frequency bands to correspond to LA (Figure 3C), MA (Figure 3D), and HA (Figure 3E) to explore the relationship between the relative energy of these α subbands and sleep quality. We found a significant positive correlation between the relative power in the LA band and PSQI, a result that did not appear in the MA and HA bands. This significance was stronger in the closed-eye condition (r=0.45, P<.001) than in the open-eye condition (r=0.26, P=.01).

Since the number of participants varied each month, Figure 4 presents a generalized group-level result across a broad population. However, 22 participants in our data set completed all 3 experiments, allowing us to examine whether this LA biomarker remains consistent within the same group. The results from this consistent group (Figure 4) were also aligned with those of the broader population. Our findings suggest that the increase in the relative power within the LA band is significantly correlated with a decline in sleep quality, a result that remained consistent across multiple follow-ups in a broad population.

Periodic Activity in LA Band Contributes to the Relationship Between

Furthermore, to more precisely characterize the different components in the spectrum, we used a classic descriptive mathematical model to decompose each participant's spectral data for each recording into periodic activity (Figure 2A) and aperiodic activity (Figure 2B). After decomposition, we found no significant differences in aperiodic activity between open-eye and closed-eye conditions (Figure 2C); the differences were primarily observed in the periodic activity (Figure 2D). These comparisons remained consistent across the 3 months of recordings (Figure 2A-D).

The next question we addressed was whether the significant correlation between sleep quality and LA intensity shown in the previous results was mainly contributed by the periodic or aperiodic components (Figure 5). After the decomposition, we plotted scatterplots to show the correlations between PSQI and both periodic and aperiodic activities in the LA, MA, and HA bands. We found that, in the LA band, periodic activity was significantly correlated with PSQI (P<.01) in both open-eye and closed-eye conditions, while aperiodic activity was not. In the MA and HA bands, neither periodic nor aperiodic activities were significantly correlated with PSQI (P>.05).

Additionally, the results from the consistent group of participants (Figure 6) were aligned with those from the broader population (Figure 5). Our findings suggest that the increase in periodic activity in the LA band is significantly associated with a decline in sleep quality, a result that remained consistent across multiple follow-ups in a broad population.

Discussion

Principal Results

Rapid screening for cognitive functions, such as sleep quality, requires identifying stable and reliable neurobiological targets.

XSL•FO
RenderX

Additionally, for convenience, portable EEG devices are essential for quick data collection. In this study, we explored the relationship between the neurobiological characteristics of portable EEG and sleep quality through multiple follow-ups for the first time. We found that, at the group level, the relative power in the LA band of the prefrontal region was significantly positively correlated with recent sleep quality. More importantly, this correlation remained consistent after several months of continuous follow-up, regardless of whether we considered the same population or an expanded cohort. Furthermore, through detailed decomposition of the spectrum, we discovered that the periodic components of the LA band primarily contributed to this significant correlation. Our research lays a solid foundation for using portable EEG for rapid screening of sleep-related issues, bringing this technology into households everywhere in the future.

Comparison With Prior Work

Sleep is a classic issue in the field of neuroscience [60-64]. During sleep, the EEG signals show specific patterns in different stages. In the beginning, high-frequency (β, γ, and α), low-voltage waves will gradually switch to higher voltage, slower waves (θ and δ), which are the feature patterns of nonrapid eye movement (NREM) sleep [18,20,65]. During NREM sleep, stages 1, 2, and 3 are classified depending on the EEG frequency and specific pattern. Spindles or bursting firing of neurons were found shortly after falling asleep. During rapid eye movement sleep, the EEG patterns were similar to those during awake [18]. In rodent studies, compared with humans, rats exhibit a more fragmented sleep pattern, with most research primarily focusing on rapid eye movement and NREM sleep stages [9,24]. During deep sleep, δ band activity often predominates, while during light sleep, θ oscillations are more prevalent. α activity typically appears in the resting state with eyes closed before sleep or during drowsiness with eyes open [66,67], which suggests that α oscillations themselves are a prelude to the sleep phase. If there is an abnormal α oscillation in the brain, it is likely to affect sleep [68,69], particularly during the process of falling asleep. However, it remains unclear how different sub-α oscillations are related to sleep quality, and this is the basis of our research. Therefore, our study does not focus on changes in neural oscillations during sleep but instead examines whether resting-state EEG has oscillatory characteristics that are significantly and consistently associated with subjective sleep quality. One possible intervention for improving sleep is yoga or meditation training, which has been shown in previous studies to significantly enhance sleep quality [70,71] and is also closely related to α oscillations [72]. Our research further refines this relationship by finding that, among various sub-α oscillations, LA band activity is significantly negatively correlated with sleep quality at the group level, while no such correlation was found with medium or HA band activity. This result could also be applied to the field of neuromodulation in the future. Improving sleep quality is a concern for many people, and previous studies have found that techniques like TACS can enhance sleep quality [65,73-75]. However, a fundamental question is how to select the appropriate frequency band for stimulation. Our research will provide important insights in this regard.

Mechanisms of Suboscillations in α Band Activity

Our study focused primarily on the α frequency band, as it is widely recognized in EEG research as a prominent biomarker with strong underlying neurobiological mechanisms [76-79]. However, our analysis also included other frequency bands, such as θ and β (Figures 3A and 4A). These bands, however, did not show stable and consistent significant correlations with sleep quality. Although the 8-13 Hz α oscillation band is narrow, it includes various suboscillations. Some studies suggest that these α suboscillations may have different origins [58,59] and are associated with different cognitive functions [80,81], but the finer neural mechanisms underlying these suboscillations remain unclear. From a dynamical systems perspective [82], the generation of different α suboscillations could be related to the differing time constants of excitatory and inhibitory neurons within the system, directly leading to frequency changes. This implies that different α suboscillations may have hierarchical relationships. It may also explain the discrepancies in previous studies regarding the mechanisms of α oscillations. For instance, some believe they originate subcortically (such as the pulvinar thalamus [83-86]), while others argue for a cortical origin (such as layer 5 of the cortex [87,88]). Our study used electrodes to record from the prefrontal cortex, but due to the relatively low spatial resolution of EEG, we cannot guarantee that the signal originates solely from the prefrontal cortex. It could also come from other brain regions or nuclei. Future studies will need to design more experiments to verify the relationship between different α suboscillations and various brain regions.

Potential Applications on Large-Scale Screening of Sleep Disorders in the Population

Currently, compared with other methods for capturing brain neural responses (functional magnetic resonance imaging, functional near-infrared spectroscopy, magnetoencephalography, and various invasive recording methods), functional magnetic resonance imaging and magnetoencephalography are expensive and have strict use conditions, while functional near-infrared spectroscopy lacks good spatial and temporal resolution. The most likely method to be widely applied in human communities is portable EEG. Although it does not have good spatial resolution, its temporal resolution and ability to reliably capture electrical signals make it particularly effective. This is especially true for the precise detection of α oscillations, which can be correlated with cellular-level or local circuit level of α oscillation mechanisms observed in animal experiments [89,90]. Just as we hoped for a rapid screening method to detect COVID-19 infections during the pandemic [48,91], today, we also desire a rapid screening tool for detecting cognitive function abnormalities in large populations. The results from current portable EEG research happen to meet this need, and our study provides strong support for this approach. In the future, we will design more experiments related to cognitive functions to fully leverage the advantages of portable EEG.

Limitations

One limitation of this study is that it only explored the relationship between the prefrontal region and sleep quality. However, other brain regions may also be associated with sleep quality. Due to the limited number of electrodes in the portable

XSL•FO
RenderX

EEG system, we were unable to address this question in this study. Therefore, in future experiments, we plan to develop portable EEG systems with more electrodes to further investigate this issue. Another issue is that there are various types of artifacts in EEG data. Since we are using a portable EEG device, efforts have been made to make it as portable as possible, which resulted in a reduction in the number of electrodes used. The device only has 2 electrodes, and therefore, the EEG data from this device is not suitable for traditional artifact removal methods, such as principal component analysis or independent component analysis [92,93]. We acknowledge that there may be potential muscle artifacts in the FP1 and FP2 channels, but these artifacts primarily affect the β or γ frequency bands. Our study, however, specifically avoids these controversial frequency bands and focuses solely on the α band, which is safe and uncontroversial. One more issue is that the source of EEG signals has always been a topic of debate [94,95]. While we observed this phenomenon in the prefrontal cortex, we cannot currently determine whether it originates solely from the prefrontal region without the involvement of other brain areas, this question could potentially be addressed by combining intracranial EEG with surface EEG for source localization. Last, this study does not argue that abbreviated analyses should replace more detailed methods universally, but rather that they offer a viable alternative for certain types of studies, especially those that prioritize efficiency. A deeper, more comprehensive analysis of EEG data can yield valuable insights as well, the trade-offs study between abbreviated and more comprehensive EEG analysis methods should be further compared in the future.

Conclusions

These findings mark the first identification of a stable and reliable neurobiological target—relative power in the LA band during the closed-eye state—associated with sleep quality across multiple follow-up sessions. Our results lay a robust foundation for future applications of portable EEG devices in monitoring sleep quality and screening for sleep disorders across diverse populations.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (82201701) and the Beijing Municipal Hospital Research and Development Project (PX2021068), and we gratefully acknowledge the financial support and technical resources provided by Shenzhen Shuimu AI Technology Co Ltd.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the policy of the institution but are available from the corresponding author on reasonable request.

Authors' Contributions

CH, XZ, and Y Lin conceived and designed the study. WX, JM, SH, Y Lia, WC, and FW contributed to data collection. CH and ZZ contributed to the literature search, data analysis, and interpretation of results. All authors contributed to writing the paper.

Conflicts of Interest

SH, JM, WX, FW, Y Lia, and WC are employees of Shenzhen Shuimu AI Technology Co Ltd. The authors declare that no other competing interests exist. The other authors declare no competing financial conflicts of interest to disclose.

References

1. Cimenser A, Hempel E, Travers T, Strozewski N, Martin K, Malchano Z, et al. Sensory-evoked 40-Hz gamma oscillation improves sleep and daily living activities in Alzheimer's disease patients. Front Syst Neurosci 2021;15:746859 [FREE Full text] [doi: 10.3389/fnsys.2021.746859] [Medline: 34630050]
2. Chan D, Suk H, Jackson B, Milman N, Stark D, Klerman E, et al. 40 Hz sensory stimulation induces gamma entrainment and affects brain structure, sleep and cognition in patients with Alzheimer's dementia.. medRxiv Preprint posted online on May 17, 2021. [doi: 10.1101/2021.03.01.21252717]
3. Feinsilver SH. Normal and abnormal sleep in the elderly. Clin Geriatr Med 2021;37(3):377-386. [doi: 10.1016/j.cger.2021.04.001] [Medline: 34210444]
4. Charest J, Grandner MA. Sleep and athletic performance: impacts on physical performance, mental performance, injury risk and recovery, and mental health. Sleep Med Clin 2020;15(1):41-57 [FREE Full text] [doi: 10.1016/j.jsmc.2019.11.005] [Medline: 32005349]
5. Tsuno N, Besset A, Ritchie K. Sleep and depression. J Clin Psychiatry 2005;66(10):1254-1269. [doi: 10.4088/jcp.v66n1008] [Medline: 16259539]
6. Palagini L, Hertenstein E, Riemann D, Nissen C. Sleep, insomnia and mental health. J Sleep Res 2022;31(4):e13628. [doi: 10.1111/jsr.13628] [Medline: 35506356]
7. Grandner MA. Sleep, health, and society. Sleep Med Clin 2017;12(1):1-22 [FREE Full text] [doi: 10.1016/j.jsmc.2016.10.012] [Medline: 28159089]

XSL•FO
RenderX

8. Yap MHW, Grabowska MJ, Rohrscheib C, Jeans R, Troup M, Paulk AC, et al. Oscillatory brain activity in spontaneous and induced sleep stages in flies. Nat Commun 2017;8(1):1815 [FREE Full text] [doi: 10.1038/s41467-017-02024-y] [Medline: 29180766]
9. Vyazovskiy VV, Olcese U, Hanlon EC, Nir Y, Cirelli C, Tononi G. Local sleep in awake rats. Nature 2011;472(7344):443-447 [FREE Full text] [doi: 10.1038/nature10009] [Medline: 21525926]
10. Montgomery SM, Sirota A, Buzsáki G. Theta and gamma coordination of hippocampal networks during waking and rapid eye movement sleep. J Neurosci 2008;28(26):6731-6741 [FREE Full text] [doi: 10.1523/JNEUROSCI.1227-08.2008] [Medline: 18579747]
11. Watson BO, Levenstein D, Greene JP, Gelinas JN, Buzsáki G. Network homeostasis and state dynamics of neocortical sleep. Neuron 2016;90(4):839-852 [FREE Full text] [doi: 10.1016/j.neuron.2016.03.036] [Medline: 27133462]
12. Li SB, Damonte VM, Chen C, Wang GX, Kebschull JM, Yamaguchi H, et al. Hyperexcitable arousal circuits drive sleep instability during aging. Science 2022;375(6583):eabh3021 [FREE Full text] [doi: 10.1126/science.abh3021] [Medline: 35201886]
13. Van Dort CJ, Zachs DP, Kenny JD, Zheng S, Goldblum RR, Gelwan NA, et al. Optogenetic activation of cholinergic neurons in the PPT or LDT induces REM sleep. Proc Natl Acad Sci U S A 2015;112(2):584-589 [FREE Full text] [doi: 10.1073/pnas.1423136112] [Medline: 25548191]
14. van Betteray JN, Vossen JM, Coenen AM. Behavioural characteristics of sleep in rats under different light/dark conditions. Physiol Behav 1991;50(1):79-82. [doi: 10.1016/0031-9384(91)90501-e] [Medline: 1946735]
15. Chauvette S, Crochet S, Volgushev M, Timofeev I. Properties of slow oscillation during slow-wave sleep and anesthesia in cats. J Neurosci 2011;31(42):14998-15008 [FREE Full text] [doi: 10.1523/JNEUROSCI.2339-11.2011] [Medline: 22016533]
16. Destexhe A, Contreras D, Steriade M. Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. J Neurosci 1999;19(11):4595-4608 [FREE Full text] [doi: 10.1523/JNEUROSCI.19-11-04595.1999] [Medline: 10341257]
17. Xu W, de Carvalho F, Jackson A. Sequential neural activity in primary motor cortex during sleep. J Neurosci 2019;39(19):3698-3712 [FREE Full text] [doi: 10.1523/JNEUROSCI.1408-18.2019] [Medline: 30842250]
18. Andrillon T, Nir Y, Cirelli C, Tononi G, Fried I. Single-neuron activity and eye movements during human REM sleep and awake vision. Nat Commun 2015;6:7884 [FREE Full text] [doi: 10.1038/ncomms8884] [Medline: 26262924]
19. Manshanden I, De Munck JC, Simon NR, Lopes da Silva FH. Source localization of MEG sleep spindles and the relation to sources of alpha band rhythms. Clin Neurophysiol 2002;113(12):1937-1947. [doi: 10.1016/s1388-2457(02)00304-8] [Medline: 12464331]
20. Nir Y, Staba RJ, Andrillon T, Vyazovskiy VV, Cirelli C, Fried I, et al. Regional slow waves and spindles in human sleep. Neuron 2011;70(1):153-169 [FREE Full text] [doi: 10.1016/j.neuron.2011.02.043] [Medline: 21482364]
21. Lane JM, Qian J, Mignot E, Redline S, Scheer FAJL, Saxena R. Genetics of circadian rhythms and sleep in human health and disease. Nat Rev Genet 2023;24(1):4-20 [FREE Full text] [doi: 10.1038/s41576-022-00519-z] [Medline: 36028773]
22. Linkowski P, Kerkhofs M, Hauspie R, Susanne C, Mendlewicz J. EEG sleep patterns in man: a twin study. Electroencephalogr Clin Neurophysiol 1989;73(4):279-284. [doi: 10.1016/0013-4694(89)90106-5] [Medline: 2477214]
23. Armitage R, Landis C, Hoffmann R, Lentz M, Watson N, Goldberg J, et al. Power spectral analysis of sleep EEG in twins discordant for chronic fatigue syndrome. J Psychosom Res 2009;66(1):51-57 [FREE Full text] [doi: 10.1016/j.jpsychores.2008.08.004] [Medline: 19073294]
24. Campos-Beltrán D, Marshall L. Changes in sleep EEG with aging in humans and rodents. Pflugers Arch 2021;473(5):841-851 [FREE Full text] [doi: 10.1007/s00424-021-02545-y] [Medline: 33791849]
25. Buzsáki G. Rhythms of the Brain. Oxford, England: Oxford University Press; 2009.
26. Buzsáki G, Draguhn A. Neuronal oscillations in cortical networks. Science 2004;304(5679):1926-1929. [doi: 10.1126/science.1099745] [Medline: 15218136]
27. Wang J, Zhao X, Bi Y, Jiang S, Sun Y, Lang J, et al. Executive function elevated by long term high-intensity physical activity and the regulation role of beta-band activity in human frontal region. Cogn Neurodyn 2023;17(6):1463-1472. [doi: 10.1007/s11571-022-09905-z] [Medline: 37974584]
28. Han C, Guo M, Ke X, Zeng L, Li M, Haihambo N, et al. Oscillatory biomarkers of autism: evidence from the innate visual fear evoking paradigm. Cogn Neurodyn 2023;17(2):459-466 [FREE Full text] [doi: 10.1007/s11571-022-09839-6] [Medline: 37007195]
29. Han C, Zhao X, Li M, Haihambo N, Teng J, Li S, et al. Enhancement of the neural response during 40 Hz auditory entrainment in closed-eye state in human prefrontal region. Cogn Neurodyn 2023;17(2):399-410 [FREE Full text] [doi: 10.1007/s11571-022-09834-x] [Medline: 37007205]
30. Cao Y, Han C, Peng X, Su Z, Liu G, Xie Y, et al. Correlation between resting theta power and cognitive performance in patients with schizophrenia. Front Hum Neurosci 2022;16:853994 [FREE Full text] [doi: 10.3389/fnhum.2022.853994] [Medline: 35529780]

XSL•FO
RenderX

31. Wang B, Li M, Haihambo N, Qiu Z, Sun M, Guo M, et al. Characterizing major depressive disorder (MDD) using alpha-band activity in resting-state electroencephalogram (EEG) combined with MATRICS consensus cognitive battery (MCCB). J Affect Disord 2024;355:254-264 [FREE Full text] [doi: 10.1016/j.jad.2024.03.145] [Medline: 38561155]
32. Mentink LJ, Thomas J, Melis RJF, Olde Rikkert MGM, Overeem S, Claassen JAHR. Home-EEG assessment of possible compensatory mechanisms for sleep disruption in highly irregular shift workers - the ANCHOR study. PLoS One 2020;15(12):e0237622 [FREE Full text] [doi: 10.1371/journal.pone.0237622] [Medline: 33382689]
33. Wang H, Hou Y, Zhan S, Li N, Liu J, Song P, et al. EEG biofeedback decreases theta and beta power while increasing alpha power in insomniacs: an open-label study. Brain Sci 2023;13(11):1542 [FREE Full text] [doi: 10.3390/brainsci13111542] [Medline: 38002502]
34. Lin S, Du Y, Xia Y, Xiao L, Wang G. Resting-state EEG as a potential indicator to predict sleep quality in depressive patients. Int J Psychophysiol 2023;191:1-8. [doi: 10.1016/j.ijpsycho.2023.06.007] [Medline: 37348762]
35. Aseem A, Chaudhry N, Hussain ME. Cranial electrostimulation improves slow wave sleep in collegiate population: a polysomnographic study. Sleep Sci 2022;15(1):88-94 [FREE Full text] [doi: 10.5935/1984-0063.20220029] [Medline: 35662960]
36. Penalba-Sánchez L, Silva G, Crook-Rumsey M, Sumich A, Rodrigues PM, Oliveira-Silva P, et al. Classification of sleep quality and aging as a function of brain complexity: a multiband non-linear EEG analysis. Sensors (Basel) 2024;24(9):2811 [FREE Full text] [doi: 10.3390/s24092811] [Medline: 38732917]
37. Gudberg C, Stevelink R, Douaud G, Wulff K, Lazari A, Fleming MK, et al. Individual differences in slow wave sleep architecture relate to variation in white matter microstructure across adulthood. Front Aging Neurosci 2022;14:745014 [FREE Full text] [doi: 10.3389/fnagi.2022.745014] [Medline: 36092806]
38. Maddirala AK, Veluvolu KC. Eye-blink artifact removal from single channel EEG with k-means and SSA. Sci Rep 2021;11(1):11043 [FREE Full text] [doi: 10.1038/s41598-021-90437-7] [Medline: 34040062]
39. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
40. Bell B, Percival DB, Walden AT. Calculating Thomson's spectral multitapers by inverse iteration. J Comput Graph Stat 1993;2(1):119-130. [doi: 10.1080/10618600.1993.10474602]
41. Percival DB, Walden AT. Spectral Analysis for Physical Applications. Cambridge, England: Cambridge University Press; 1993.
42. Han C, Wang B, Yang G, Dai W, Yang Y, Wang T, et al. Neural mechanism of orientation selectivity for distinct gamma oscillations in cat V1. J Vis 2020;20(11):1116 [FREE Full text] [doi: 10.1167/jov.20.11.1116]
43. Han C. The oscillating mystery: the effects of forty‐hertz entrainment in treating Alzheimer's disease. Brain-X 2023;1(2):e14. [doi: 10.1002/brx2.14]
44. Han C, Shapley R, Xing D. Gamma rhythms in the visual cortex: functions and mechanisms. Cogn Neurodyn 2021;16(4):745-756 [FREE Full text]
45. Wang B, Han C, Wang T, Dai W, Li Y, Yang Y, et al. Superimposed gratings induce diverse response patterns of gamma oscillations in primary visual cortex. Sci Rep 2021;11(1):4941 [FREE Full text] [doi: 10.1038/s41598-021-83923-5] [Medline: 33654121]
46. Cao Y, Li M, Haihambo N, Zhu Y, Zeng Y, Jin J, et al. Oscillatory properties of class C notifiable infectious diseases in China from 2009 to 2021. Front Public Health 2022;10:903025 [FREE Full text] [doi: 10.3389/fpubh.2022.903025] [Medline: 36033737]
47. Han C, Li M, Haihambo N, Cao Y, Zhao X. Enlightenment on oscillatory properties of 23 class B notifiable infectious diseases in the mainland of China from 2004 to 2020. PLoS One 2021;16(6):e0252803 [FREE Full text] [doi: 10.1371/journal.pone.0252803] [Medline: 34106977]
48. Han C, Li M, Haihambo N, Babuna P, Liu Q, Zhao X, et al. Mechanisms of recurrent outbreak of COVID-19: a model-based study. Nonlinear Dyn 2021;106(2):1169-1185 [FREE Full text] [doi: 10.1007/s11071-021-06371-w] [Medline: 33758464]
49. Zhao X, Li M, Haihambo N, Jin J, Zeng Y, Qiu J, et al. Changes in temporal properties of notifiable infectious disease epidemics in China during the COVID-19 pandemic: population-based surveillance study. JMIR Public Health Surveill 2022;8(6):e35343 [FREE Full text] [doi: 10.2196/35343] [Medline: 35649394]
50. Zhao X, Li M, Haihambo N, Wang X, Wang B, Sun M, et al. Periodic characteristics of hepatitis virus infections from 2013 to 2020 and their association with meteorological factors in Guangdong, China: surveillance study. JMIR Public Health Surveill 2023;9:e45199 [FREE Full text] [doi: 10.2196/45199] [Medline: 37318858]
51. Han C, Wang T, Yang Y, Wu Y, Li Y, Dai W, et al. Multiple gamma rhythms carry distinct spatial frequency information in primary visual cortex. PLoS Biol 2021;19(12):e3001466 [FREE Full text] [doi: 10.1371/journal.pbio.3001466] [Medline: 34932558]
52. Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, et al. Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci 2020;23(12):1655-1665 [FREE Full text] [doi: 10.1038/s41593-020-00744-x] [Medline: 33230329]

XSL•FO
RenderX

53. Han C, Wang T, Wu Y, Li Y, Yang Y, Li L, et al. The generation and modulation of distinct gamma oscillations with local, horizontal, and feedback connections in the primary visual cortex: a model study on large-scale networks. Neural Plast 2021;2021:8874516 [FREE Full text] [doi: 10.1155/2021/8874516] [Medline: 33531893]
54. Lin Y, Huang S, Mao J, Li M, Haihambo N, Wang F, et al. The neural oscillatory mechanism underlying human brain fingerprint recognition using a portable EEG acquisition device. Neuroimage 2024;294:120637 [FREE Full text] [doi: 10.1016/j.neuroimage.2024.120637] [Medline: 38714216]
55. Barry RJ, Clarke AR, Johnstone SJ, Magee CA, Rushby JA. EEG differences between eyes-closed and eyes-open resting conditions. Clin Neurophysiol 2007;118(12):2765-2773. [doi: 10.1016/j.clinph.2007.07.028] [Medline: 17911042]
56. Barry RJ, Clarke AR, Johnstone SJ, Brown CR. EEG differences in children between eyes-closed and eyes-open resting conditions. Clin Neurophysiol 2009;120(10):1806-1811. [doi: 10.1016/j.clinph.2009.08.006] [Medline: 19748828]
57. Barry RJ, De Blasio FM. EEG differences between eyes-closed and eyes-open resting remain in healthy ageing. Biol Psychol 2017;129:293-304. [doi: 10.1016/j.biopsycho.2017.09.010] [Medline: 28943465]
58. Barzegaran E, Vildavski VY, Knyazeva MG. Fine structure of posterior alpha rhythm in human EEG: frequency components, their cortical sources, and temporal behavior. Sci Rep 2017;7(1):8249 [FREE Full text] [doi: 10.1038/s41598-017-08421-z] [Medline: 28811538]
59. Han C, Wang T, Wu Y, Li H, Wang E, Zhao X, et al. Compensatory mechanism of attention-deficit/hyperactivity disorder recovery in resting state alpha rhythms. Front Comput Neurosci 2022;16:883065 [FREE Full text] [doi: 10.3389/fncom.2022.883065] [Medline: 36157841]
60. Sulaman BA, Wang S, Tyan J, Eban-Rothschild A. Neuro-orchestration of sleep and wakefulness. Nat Neurosci 2023;26(2):196-212. [doi: 10.1038/s41593-022-01236-w] [Medline: 36581730]
61. Konkoly K, Appel K, Chabani E, Mironov AY, Mangiaruga A, Gott J, et al. Real-time dialogue between experimenters and dreamers during rem sleep. SSRN J 2020:1-11. [doi: 10.2139/ssrn.3606772]
62. Cirelli C, Tononi G. Cortical development, electroencephalogram rhythms, and the sleep/wake cycle. Biol Psychiatry 2015;77(12):1071-1078 [FREE Full text] [doi: 10.1016/j.biopsych.2014.12.017] [Medline: 25680672]
63. Ketz N, Jones AP, Bryant NB, Clark VP, Pilly PK. Closed-loop slow-wave tACS improves sleep-dependent long-term memory generalization by modulating endogenous oscillations. J Neurosci 2018;38(33):7314-7326 [FREE Full text] [doi: 10.1523/JNEUROSCI.0273-18.2018] [Medline: 30037830]
64. Adamantidis AR, Gutierrez Herrera C, Gent TC. Oscillating circuitries in the sleeping brain. Nat Rev Neurosci 2019;20(12):746-762. [doi: 10.1038/s41583-019-0223-4] [Medline: 31616106]
65. Ayanampudi V, Kumar V, Krishnan A, Walker MP, Ivry RB, Knight RT, et al. Personalized transcranial alternating current stimulation improves sleep quality: initial findings. Front Hum Neurosci 2022;16:1066453 [FREE Full text] [doi: 10.3389/fnhum.2022.1066453] [Medline: 36704097]
66. Bollimunta A, Chen Y, Schroeder CE, Ding M. Neuronal mechanisms of cortical alpha oscillations in awake-behaving macaques. J Neurosci 2008;28(40):9976-9988 [FREE Full text] [doi: 10.1523/JNEUROSCI.2699-08.2008] [Medline: 18829955]
67. Bollimunta A, Mo J, Schroeder CE, Ding M. Neuronal mechanisms and attentional modulation of corticothalamic α oscillations. J Neurosci 2011;31(13):4935-4943 [FREE Full text] [doi: 10.1523/JNEUROSCI.5580-10.2011] [Medline: 21451032]
68. Wu J, Zhou Q, Li J, Chen Y, Shao S, Xiao Y. Decreased resting-state alpha-band activation and functional connectivity after sleep deprivation. Sci Rep 2021;11(1):484 [FREE Full text] [doi: 10.1038/s41598-020-79816-8] [Medline: 33436726]
69. Zhao W, Van Someren EJW, Li C, Chen X, Gui W, Tian Y, et al. EEG spectral analysis in insomnia disorder: a systematic review and meta-analysis. Sleep Med Rev 2021;59:101457. [doi: 10.1016/j.smrv.2021.101457] [Medline: 33607464]
70. Büssing A, Michalsen A, Khalsa SBS, Telles S, Sherman KJ. Effects of yoga on mental and physical health: a short summary of reviews. Evid Based Complement Alternat Med 2012;2012:165410 [FREE Full text] [doi: 10.1155/2012/165410] [Medline: 23008738]
71. Cramer H, Lauche R, Langhorst J, Dobos G. Yoga for rheumatic diseases: a systematic review. Rheumatology (Oxford) 2013;52(11):2025-2030. [doi: 10.1093/rheumatology/ket264] [Medline: 23934220]
72. Chandra S, Sharma G, Mittal AP, Jha D. Effect of Sudarshan Kriya (meditation) on gamma, alpha, and theta rhythm during working memory task. Int J Yoga 2016;9(1):72-76 [FREE Full text] [doi: 10.4103/0973-6131.171715] [Medline: 26865775]
73. Robinson CSH, Bryant NB, Maxwell JW, Jones AP, Robert B, Lamphere M, et al. The benefits of closed-loop transcranial alternating current stimulation on subjective sleep quality. Brain Sci 2018;8(12):204 [FREE Full text] [doi: 10.3390/brainsci8120204] [Medline: 30469495]
74. Motamedi GK, Jeliazkov PG, Oyegbile-Chidi TO, Song SS, Sharma VD, Mete M, et al. Transcranial alternating current stimulation (tACS) as a treatment for insomnia. Can J Neurol Sci 2023;50(3):446-449 [FREE Full text] [doi: 10.1017/cjn.2022.33] [Medline: 35321778]
75. Lu H, Li J, Yang NS, Lam LCW, Ma SL, Wing YK, et al. Using gamma-band transcranial alternating current stimulation (tACS) to improve sleep quality and cognition in patients with mild neurocognitive disorders due to Alzheimer's disease: a study protocol for a randomized controlled trial. PLoS One 2023;18(8):e0289591 [FREE Full text] [doi: 10.1371/journal.pone.0289591] [Medline: 37540692]

XSL•FO
RenderX

76. Zikereya T, Lin Y, Zhang Z, Taguas I, Shi K, Han C. Different oscillatory mechanisms of dementia-related diseases with cognitive impairment in closed-eye state. Neuroimage 2024;304:120945 [FREE Full text] [doi: 10.1016/j.neuroimage.2024.120945] [Medline: 39586346]
77. Zhang Y, Zhang Z, Du F, Song J, Huang S, Mao J, et al. Shared oscillatory mechanisms of alpha-band activity in prefrontal regions in eyes open and closed state using a portable EEG acquisition device. Sci Rep 2024;14(1):26719 [FREE Full text] [doi: 10.1038/s41598-024-78173-0] [Medline: 39496816]
78. Bazanova OM, Vernon D. Interpreting EEG alpha activity. Neurosci Biobehav Rev 2014;44:94-110. [doi: 10.1016/j.neubiorev.2013.05.007] [Medline: 23701947]
79. Hindriks R, van Putten MJ, Deco G. Intra-cortical propagation of EEG alpha oscillations. Neuroimage 2014;103:444-453. [doi: 10.1016/j.neuroimage.2014.08.027] [Medline: 25168275]
80. Jia J, Fang F, Luo H. Selective spatial attention involves two alpha-band components associated with distinct spatiotemporal and functional characteristics. Neuroimage 2019;199:228-236. [doi: 10.1016/j.neuroimage.2019.05.079] [Medline: 31154048]
81. Ben-Simon E, Podlipsky I, Arieli A, Zhdanov A, Hendler T. Never resting brain: simultaneous representation of two alpha related processes in humans. PLoS One 2008;3(12):e3984 [FREE Full text] [doi: 10.1371/journal.pone.0003984] [Medline: 19096714]
82. Mejias JF, Murray JD, Kennedy H, Wang X. Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. Sci Adv 2016;2(11):e1601335 [FREE Full text] [doi: 10.1126/sciadv.1601335] [Medline: 28138530]
83. Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S. The pulvinar regulates information transmission between cortical areas based on attention demands. Science 2012;337(6095):753-756 [FREE Full text] [doi: 10.1126/science.1223082] [Medline: 22879517]
84. Kastner S, Fiebelkorn IC, Eradath MK. Dynamic pulvino-cortical interactions in the primate attention network. Curr Opin Neurobiol 2020;65:10-19 [FREE Full text] [doi: 10.1016/j.conb.2020.08.002] [Medline: 32942125]
85. Vijayan S, Kopell NJ. Thalamic model of awake alpha oscillations and implications for stimulus processing. Proc Natl Acad Sci U S A 2012;109(45):18553-18558 [FREE Full text] [doi: 10.1073/pnas.1215385109] [Medline: 23054840]
86. Hughes SW, Crunelli V. Thalamic mechanisms of EEG alpha rhythms and their pathological implications. Neuroscientist 2005;11(4):357-372. [doi: 10.1177/1073858405277450] [Medline: 16061522]
87. van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M, Poort J, van der Togt C, et al. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci U S A 2014;111(40):14332-14341 [FREE Full text] [doi: 10.1073/pnas.1402773111] [Medline: 25205811]
88. Silva LR, Amitai Y, Connors BW. Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. Science 1991;251(4992):432-435. [doi: 10.1126/science.1824881] [Medline: 1824881]
89. Hughes SW, Lörincz M, Cope DW, Blethyn KL, Kékesi KA, Parri HR, et al. Synchronized oscillations at alpha and theta frequencies in the lateral geniculate nucleus. Neuron 2004;42(2):253-268 [FREE Full text] [doi: 10.1016/s0896-6273(04)00191-6] [Medline: 15091341]
90. Lopes Da Silva F, Storm Van Leeuwen W. The cortical source of the alpha rhythm. Neurosci Lett 1977;6(2-3):237-241. [doi: 10.1016/0304-3940(77)90024-6] [Medline: 19605058]
91. Han C, Liu Y, Tang J, Zhu Y, Jaeger C, Yang S. Lessons from the mainland of China's epidemic experience in the first phase about the growth rules of infected and recovered cases of COVID-19 worldwide. Int J Disaster Risk Sci 2020;11(4):497-507. [doi: 10.1007/s13753-020-00294-7]
92. Turnip A, Junaidi E. Removal artifacts from EEG signal using independent component analysis and principal component analysis. 2014 Presented at: 2nd International Conference on Technology, Informatics, Management, Engineering & Environment; 2014 August 19-21; Bandung, Indonesia.
93. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9-21 [FREE Full text] [doi: 10.1016/j.jneumeth.2003.10.009] [Medline: 15102499]
94. Caso F, Cursi M, Magnani G, Fanelli G, Falautano M, Comi G, et al. Quantitative EEG and LORETA: valuable tools in discerning FTD from AD? Neurobiol Aging 2012;33(10):2343-2356. [doi: 10.1016/j.neurobiolaging.2011.12.011] [Medline: 22244088]
95. Nishida K, Yoshimura M, Isotani T, Yoshida T, Kitaura Y, Saito A, et al. Differences in quantitative EEG between frontotemporal dementia and Alzheimer's disease as revealed by LORETA. Clin Neurophysiol 2011;122(9):1718-1725. [doi: 10.1016/j.clinph.2011.02.011] [Medline: 21396882]

Abbreviations

EEG: electroencephalogram
HA: high α
LA: low α
MA: medium α

XSL•FO
RenderX

NREM: nonrapid eye movement
PSQI: Pittsburgh Sleep Quality Index

Edited by A Coristine; submitted 04.10.24; peer-reviewed by K Shi, C Kerson; comments to author 29.11.24; revised version received 20.12.24; accepted 04.02.25; published 10.03.25.

Please cite as:
Han C, Zhang Z, Lin Y, Huang S, Mao J, Xiang W, Wang F, Liang Y, Chen W, Zhao X
Monitoring Sleep Quality Through Low α-Band Activity in the Prefrontal Cortex Using a Portable Electroencephalogram Device: Longitudinal Study
J Med Internet Res 2025;27:e67188
URL: https://www.jmir.org/2025/1/e67188
doi:10.2196/67188
PMID:

©Chuanliang Han, Zhizhen Zhang, Yuchen Lin, Shaojia Huang, Jidong Mao, Weiwen Xiang, Fang Wang, Yuping Liang, Wufang Chen, Xixi Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigating the Magnitude and Persistence of COVID-19–Related Impacts on Affect and GPS-Derived Daily Mobility Patterns in Adolescence and Emerging Adulthood: Insights From a Smartphone-Based Intensive Longitudinal Study of Colorado-Based Youths From June 2016 to April 2022

Jordan D Alexander[1], MA; Kelly A Duffy[1], PhD; Samantha M Freis[2,3], PhD; Sy-Miin Chow[4], PhD; Naomi P Friedman[2,3], PhD; Scott I Vrieze[1], PhD

[1]Department of Psychology, University of Minnesota, Minneapolis, MN, United States

[2]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, United States

[3]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, United States

[4]College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States

Corresponding Author:
Jordan D Alexander, MA
Department of Psychology
University of Minnesota
75 East River Parkway
Minneapolis, MN, 55455
United States
Phone: 1 6126252818
Fax: 1 6126262079
Email: alexa877@umn.edu

Abstract

Background: The onset of the COVID-19 pandemic in early 2020 introduced unprecedented disruptions impacting the emotional well-being and daily routines of US youths. However, the patterns and persistence of these impacts over the pandemic's multiyear course remain less well understood.

Objective: This study examined longitudinal changes in affect and daily mobility patterns observed in adolescence and young adulthood from June 2016 to April 2022. The study aimed to quantify changes in youths' mood and daily routines following the pandemic's onset and in response to local COVID-19 case rates as well as the persistence of these effects over the pandemic's multiyear course.

Methods: Colorado-based adolescent and young adult twins (N=887; n=479, 54% female; $mean_{age}$ 19.2, SD_{age} 1.5 years on January 01, 2020) participating in the CoTwins study between June 2016 and April 2022 were followed via a smartphone app, which recorded persistent GPS location data and, beginning in February 2019, administered an abbreviated Positive and Negative Affect Schedule every 2 weeks. Nonlinear trajectories in affect and daily mobility over time and in response to local COVID-19 counts were modeled via generalized additive mixed models, while the magnitude and persistence of pandemic-related changes were quantified via linear mixed effects regressions.

Results: Between January and April 2020, participants experienced a 28.6% decline in daily locations visited (from 3.5 to 2.5; SD 0.9) and a 60% reduction in daily travel distance (from 20.0 to 8.0 km; SD 9.4). Mean positive affect similarly declined by 0.3 SD (from 3.0 to 2.79; SD 0.6), while, correspondingly, mean negative affect increased by 0.3 SD (from 1.85 to 2.10; SD 0.6). Though mobility levels partially recovered beginning in the summer of 2020, daily locations visited remained slightly below 2019 levels through the study's conclusion in April 2022 (standardized β=–0.10; P<.001). Average positive affect similarly remained slightly below (standardized β=–0.20; P<.001) and negative affect slightly above (standardized β=0.14; P=.04) 2019 levels through April 2022. Weekly county-level COVID-19 transmission rates were negatively associated with mobility and positive affect and positively with negative affect, though these effects were greatly weakened later in the pandemic (eg, early 2022) or when transmission rates were high (eg, >200 new cases per 100,000 people per week).

XSL•FO
RenderX

Conclusions: Findings demonstrate large initial declines in daily mobility, a moderate decline in positive affect, and a moderate increase in negative affect following the pandemic's onset in 2020. Though most effects attenuated over time, affect and mobility levels had not recovered to prepandemic levels by April 2022. Findings support theories of hedonic adaptation and resiliency while also identifying lingering emotional and behavioral consequences. The study highlights both youth's resiliency in adapting to major stressors while also underscoring the need for continued support for youth mental health and psychosocial functioning in the pandemic's aftermath.

(J Med Internet Res 2025;27:e64965) doi:10.2196/64965

KEYWORDS

adolescence; emerging adulthood; intensive longitudinal assessment; COVID-19; affect; GPS; mobility patterns; smartphone data; respiratory; infectious; pulmonary; pandemic; adolescents; teens; teenagers; mobility; apps; smartphones; intensive longitudinal panel studies; emotional well-being; well-being; daily routines; affect survey

Introduction

Individuals throughout the world faced widespread disruptions and uncertainty during the COVID-19 pandemic. From January 2020 to March 2023, the United States confirmed more than 100,000,000 cases and 1,100,000 deaths from COVID-19 [1]. Beginning in March 2020, many states closed schools and nonessential businesses and encouraged residents to minimize trips away from home [2]. Youths (ie, late adolescents and emerging adults) in particular faced unique stressors during the pandemic, most notably disruptions to developmental milestones, such as transitioning into college or the workforce, peer group formation, and parental separation [3-5], often with significant impacts on their mental health, including increases in rates of self-reported anxiety and depression symptoms [6-8].

Consistent with these mental health impacts, modest declines in positive emotional experiences and increases in negative emotional experiences following the pandemic's onset are widely reported, though stronger impacts have been reported for adolescents and younger adults, especially in response to stressful pandemic-related experiences [9-14]. While such findings have been widely replicated, studies of the pandemic's psychological impacts are frequently limited by measurement and study design concerns. Most lack pre–COVID-19 data, and report perceived changes in emotional well-being, which may be especially prone to recall biases in the context of traumatic events like the pandemic [15,16] rather than actual changes in emotional experiences before and after the pandemic's onset. Furthermore, such studies were generally collected at only a few time points, often during the first few months of 2020, limiting their ability to observe changes in well-being over the multiyear course of the pandemic.

Psychological theories of adaptation to stressful experiences offer different predictions on the persistence of pandemic-related distress. Hedonic adaptation theory [17] predicts that, following stressful life events, emotional well-being may initially be substantially impacted but will recover to a pre-existing "set point" over the subsequent months or weeks [18,19]. This theory predicts that pandemic-related impacts on emotional experiences gradually attenuated over subsequent weeks and months as individuals adjusted to the event. Contrastingly, theories regarding the psychological impact of traumatic experiences assert that sufficiently distressing events may have lasting emotional consequences, suggesting many individuals may have faced lasting distress in the pandemic's aftermath [20,21].

One response to the challenge of validly capturing pandemic-related impacts via retrospective self-report was the use of behavioral "proxy" measures, like smartphone-based GPS location data, for which prospective data were available and which did not use self-report [22]. Such data can capture features of daily routine mobility patterns, including their frequency, distance, duration, and regularity [23-26]. This research has identified widespread changes in mobility patterns following the pandemic's onset and the implementation of lockdown policies, including fewer locations visited and shorter travel distance, and frequently a gradual recovery toward prepandemic mobility patterns following the first few months of the pandemic [27-33]. Several studies found that mobility changes were moderated by factors like local pandemic severity and sociodemographic characteristics like age, race or ethnicity, and income levels [28,29,32,34-36].

Pandemic-related mobility disruptions are widely studied among adults, though little research has investigated effects on youth mobility. Youths are less risk averse than adults [37,38], experienced lower risk of severe illness from COVID-19 [39], and were likely differentially impacted by pandemic-related policies (eg, school closures). Hence, youths and adults likely experienced a different set of salient restrictions and motivations impacting their daily routines.

GPS location data are useful for identifying changes in daily routines, though they cannot capture the pandemic's psychological impacts. Such impacts are predominantly studied via self-report instruments. However, as previously discussed, the rapid onset of the pandemic precluded prospective data collection for many studies, leading much of the literature on COVID-19–related psychological impacts especially prone to recall biases. Ongoing longitudinal studies with prospectively measured psychological surveys, especially those conducted over multiple years before and during the pandemic, are key sources of such prospective data. Such studies are well suited to improve and expand upon the COVID-19 literature, allowing for the measurement of the pandemic's impact on psychological outcomes with greater robustness to self-reporting biases and greater information on the persistence of pandemic-related effects.

XSL•FO
RenderX

This study investigated the magnitude and persistence of changes in youth affect and daily mobility patterns following the COVID-19 pandemic, using GPS mobility data and biweekly affect surveys collected for up to 70 months between June 2016 and April 2022 from a sample of Colorado-based adolescent and young adult twins. These data are unique in that they pair multiple years of continuously collected intensive longitudinal behavioral surveys and persistent GPS location data, with data collection beginning years before and concluding years after the COVID-19 pandemic's onset, allowing for fine-grained analyses of changes in mood and daily routine both prior to and during the pandemic's multiyear course. Investigating initial pandemic-related impacts on youth affect and mobility and the persistence of these changes over subsequent years, this study aimed to inform ongoing efforts to both understand and mitigate the lasting consequences of pandemic-related experiences on youth mental health and psychosocial functioning.

We had 2 primary aims. First, to investigate the magnitude of COVID-19–related disruptions to youth daily routines, measured via GPS-based mobility measures, and emotional experiences, measured by biweekly surveys measuring positive and negative affect, and the persistence of these effects over the pandemic's multiyear course. Second, we estimated the effect of local pandemic severity (measured by the weekly incidence of county-level COVID-19 cases) on daily routine and affect as well as whether case count effects differed over the course of the pandemic.

Consistent with similar work in adults [32], we predicted sharp declines in measures of daily mobility in March 2020, following the implementation of COVID-19 mitigation policies in many US states, and a gradual return to prepandemic mobility patterns over the following months, as COVID-19–related restrictions eased, vaccines became available, and attitudes toward social gatherings grew more permissive. Similarly, we predicted moderate to large declines in positive affect and increases in negative affect in early 2020, with a return to baseline over the following months due to hedonic adaptation and increasing opportunities for enjoyment and social engagement with the loosening of pandemic-related restrictions. Finally, we predicted reduced mobility and poorer emotional well-being in areas with greater COVID-19 transmission rates and that this relationship would be strongest during the initial months of the pandemic, reflecting greater fear and uncertainty about the pandemic's impacts.

Methods

Participants

Potential CoTwins participants were identified from birth records maintained by the Colorado Department of Health and Human Services as twins between the ages of 14 and 17 years upon recruitment. Recruitment procedures included both English and Spanish language digital and physical advertising as well as phone-based recruitment by study research assistants. Participants in this study were 887 twins who, at the time of their initial intake visit, were between the ages of 14 and 17 years, enrolled in a Colorado high school, and owned their own Android or iOS smartphone device (see Table 1 for participant demographics).

Parents and children both provided informed consent and assent prior to participation, and the study was approved by institutional review boards at both the University of Minnesota and the University of Colorado, Boulder. Intake visits for the first 670 participants were conducted between April 2015 and October 2016 on the University of Colorado, Boulder campus. These participants were initially recruited to participate for 1 year with the opportunity to continue for an additional year. Participating youths and their parents provided demographic information and baseline youth or parent-report data during an initial in-person assessment visit, which included cognitive testing, interviews, and both youth and parent-report questionnaires [40]. Youth participants then completed routine surveys and provided GPS data via a smartphone app for the duration of their participation in the study. A second wave of recruitment occurred between October 2018 and July 2021, with participants agreeing to participate for an additional 3 years. Of the original 670 twins, 76.1% (n=510) participated in this wave, and 217 new participants were recruited. Due to COVID-19 university-mandated lockdowns, no intake visits occurred between February 2020 and February 2021, though new participants were still recruited at this time, completing both baseline and follow-up surveys remotely via the CoTwins smartphone app (Vrieze lab). Data collection was completed in April 2022.

XSL•FO
RenderX

Table 1. CoTwins sample demographics (N=887)[a].

	Values
Sex, n (%)	
Female	478 (53.9)
Male	409 (46.1)
Race, n (%)	
American Indian or Alaska Native	12 (1.4)
Asian	4 (0.5)
Black or African American	12 (1.4)
Native Hawaiian or other Pacific Islander	2 (0.2)
White	715 (80.6)
More than 1 race	90 (10.2)
Declined to provide	56 (6.3)
Ethnicity, n (%)	
Hispanic or Latino	138 (15.6)
Not Hispanic or Latino	749 (84.4)
Annual family income (US $), n (%)	
Less than $30,000	30 (3.4)
$31,000-$60,000	98 (11.0)
$61,000-$100,000	162 (18.3)
$100,000-$150,000	208 (23.4)
Greater than $150,000	225 (25.4)
Declined to provide	164 (18.5)
Highest attained parent education, n (%)	
Less than high school	4 (0.1)
High school diploma or graduate equivalency degree	42 (4.7)
Some college or associate degree	230 (25.9)
Bachelor degree	310 (34.9)
Master degree or higher	249 (28.1)
Declined to provide	52 (5.9)
Age (years), mean (SD)	
Age on January 1, 2020	19.18 (1.54)

[a]Participants were 887 adolescent and young adult twins initially recruited as high school students in the state of Colorado who provided affect surveys and persistent GPS location data over multiple years between June 2016 and April 2022. Participant demographic information was collected via parent report during an initial in-person intake assessment.

Procedure

The CoTwins study's primary aim was to investigate the development of substance use and executive functions in adolescence and emerging adulthood. The study used a twin design to allow genetic and environmental influences on these behaviors to be quantified. Upon enrolling in the study, CoTwins participants completed an in-person intake visit on the University of Colorado, Boulder campus that included baseline assessments and the installation of the CoTwins smartphone app. This app was then used to administer regular self-report assessments and collect time-stamped GPS location data for the duration of the study. Study staff regularly reviewed questionnaire completion rates and GPS data collection and contacted twins to assist with technical issues as needed. Additional information on the CoTwins study procedure is available in both Alexander et al [23] and Freis et al [40].

Measures

Positive and Negative Affect

Positive and negative affect surveys were deployed to participants' smartphones once every 2 weeks via an abbreviated form of the Positive and Negative Affect Schedule (PANAS) [41,42]. The abbreviated PANAS consisted of 5 items assessing

XSL•FO
RenderX

negative affect and 5 items assessing positive affect. Example items include “indicate to what extent you have felt afraid over the past few days” (negative affect) and “indicate to what extent you have felt inspired over the past few days” (positive affect). Each PANAS item was answered on a 5-point scale from 1=very slightly to 5=extremely. The average past week positive and negative affect were both computed by taking the mean score of all the positive or negative items.

Daily Mobility Measures

Participant location was collected via the CoTwins smartphone app, installed on each twin’s personal phone. To limit battery drain, the “significant change” location application programming interface was used on iOS devices, such that a participant location was recorded each time the device registered a “significant” change in location (eg, a movement greater than roughly 100-200 m). On Android devices, the app was designed to record the participant’s location once every 5 minutes.

After visual inspection and data cleaning steps, which included removing duplicated, incomplete, inaccurately measured, or highly improbable locations (such as those implying participant movement speeds greater than 600 km per hour), locations were aggregated into points of interest, also called “staypoints,” which were locations where participants were estimated to have spent at least 30 minutes within a 200 m radius [43]. This was done to help standardize the number and meaning of GPS locations observed each day between Android and iOS participants or between participants living in urban or rural areas. All locations recorded outside the United States were removed from analyses. This allowed for harmonization with COVID-19 case count data, ensured consistent data quality and availability, and reduced possible confounding effects due to heterogenous policies, cultural environments, and atypical travel experiences. From an initial sample of 42.0 million unique GPS locations, data cleaning and aggregation procedures yielded a dataset of 2.1 million staypoints from 598,966 participant days.

Two daily mobility measures were computed from these staypoints: daily locations visited, defined as the number of staypoints recorded by a participant on a given day, and daily travel distance, the straight-line distance (in kilometers) between a day’s consecutively recorded staypoints. Travel days, in which the daily travel distance exceeded 500 km, were excluded from analyses to reduce the effect of outlying values. Additional information on the computation and measurement properties of these measures is available in Alexander et al [23].

COVID-19 Case Count

The Johns Hopkins Coronavirus Research Center provided daily data on the number of new COVID-19 cases recorded in each US county between January 20, 2020, and March 10, 2023 [1]. These were mapped to participants’ modal county each day (defined as the county where a participant recorded the greatest number of staypoints) to measure the degree of COVID-19 transmission in a participant’s environs. To help account for reporting artifacts (eg, counties reporting weekend cases on the following Monday), daily county-level case count was aggregated by week and standardized by the county’s population as weekly county-level case count per 100,000 people.

Date

A date term was used to model changes in affect and mobility over time both before and during the pandemic. This was included in statistical models as the number of days since January 20, 2020, the date of the first US COVID-19 case. Data collected prior to pandemic onset were assigned negative date values. Several follow-up analyses included a categorical effect of date to compare affect and mobility levels between January and April (corresponding to the first 4 months of the COVID-19 pandemic) in 2019-2022. Categorical date effects were restricted to data from January to April each year. This was done to account for seasonality effects, to isolate the especially large effect of the first months of the pandemic in early 2020, and to account for the lack of available data after April 18, 2022. This categorical date variable had 4 levels: indicating whether an observation was recorded from January 20 to April 30, 2019, from January 20 to April 30, 2020, from January 20 to April 30, 2021, or from January 20 to April 18, 2022 (the final day with available affect or mobility data in the study).

Model Covariates

Participant sex was coded as male (0) or female (1). Participant age on January 20, 2020 (included as a constant to reduce collinearity with date effects) was recorded in years. Weekday or weekend status was coded as weekday (0) or weekend (1). The operating system was recorded as either Apple iOS (0) or Android OS (1).

Missing Data

To help account for any systematic missingness in either the affect or mobility measures, a common challenge in intensive longitudinal research [44], we included the proportion of missing days of location data and the proportion of missed PANAS responses as model covariates. The proportion of missing days of location data was defined as the proportion of days between a participant’s first and last recorded location with no recorded staypoints. The proportion of missed PANAS surveys was computed as 1 minus the number of recorded affect surveys divided by a “theoretical maximum” number of affect surveys a person could have recorded, defined as the number of surveys a participant would have completed if they had completed 1 affect survey every 2 weeks between the date of their first and last survey response (rounded upward to the nearest integer to prevent fractional values).

Ethical Considerations

All study protocols were reviewed and approved by institutional review boards at both the University of Minnesota (STUDY00000748) and the University of Colorado, Boulder (14-0433), to ensure compliance with ethical standards for the treatment of research participants. Participants’ parents provided written consent, and participants provided verbal assent during the study recruitment and prior to their initial baseline visit. Data were securely stored on both local encrypted devices at the Universities of Minnesota and Colorado, Boulder, and on the remote encrypted data storage platform, Box. Data identifiers, including participants’ names, birthdates, and home addresses, were stored separately from their behavioral and GPS data, which were assigned anonymized individual and family

XSL•FO
RenderX

ID codes (to retain twin pair relationships in the data). Raw GPS data, which potentially allow for the inference of participant addresses and other significant locations, were deemed inherently identifying and were thus subject to additional privacy protections, including a separate participant ID, which could only be reconciled with participants' behavioral data using a securely stored ID mapping available only to the study investigators. Only location metadata, which were not inherently identifying, including the number of daily locations visited and daily travel distance, were considered in this investigation or paired with participants' behavioral data. Participants were compensated US $100 for their first baseline visit (plus reimbursement for any travel costs), up to US $150 (for completion of all surveys and complete GPS location data) per year of participation, and an additional US $100 after offboarding from the study.

Analyses

All models reported in the Results section are presented in Table 2 along with their corresponding interpretation. A variety of statistical approaches were considered to model changes in affect and mobility patterns following the onset of the COVID-19 pandemic and in response to local COVID-19 cases, including linear mixed effects models, structural equation–based latent growth curve models, and generalized additive mixed models ("GAMMs"). Mobility and affect data were correlated at the individual and family level, were collected frequently at unevenly spaced intervals, and were expected to exhibit highly nonlinear trajectories both over time and in response to local COVID-19 cases. We therefore determined that GAMMs were best suited to accommodate these correlated, time unstructured data and expected nonlinear relationships [45,46].

GAMMs were fit using the *gamm4* package in R (R Foundation for Statistical Computing) [46]. Models are presented in Table 2. Linear fixed effects included sex, age, smartphone operating system, and the proportion of missing location and survey data. Random intercepts of individuals nested within families and nonlinear "smooth" terms (fit via penalized regression splines) of either date or case count were also included. Models of daily mobility further included a covariate, indicating whether the observation was recorded on a weekday or a weekend. Fixed effects covariates of both month and season were considered to correct for possible seasonality in affect and mobility patterns, though these were ultimately excluded due to a lack of sufficient prepandemic affect data to infer typical seasonal trends (particularly for affect data, which were first available in February 2019), as well as collinearity between season, COVID-19 onset, and COVID-19 case counts. Smooth effects for date were fit using thin plate regression splines with k=35 basis functions, while smooth effects for case counts were fit with k=10 basis functions. Selection of the number of basis functions was supported via the k-basis dimension test to prevent model underidentification [46,47]. To assess whether mobility and affect were significantly increasing or decreasing at a given date or case count level, we computed first derivatives and 99% CIs of the smooth date and case count terms at 10,000 equally spaced points on the smooth's curve using the R package *gratia* [48,49].

Such GAMMs are useful for visualizing complex mean trajectories, but they consequently do not provide readily interpretable fixed effects coefficients for date or case count effects. Hence, complementary to these generalized additive models, we further fit linear mixed effects models, via the R package *lme4* [50], which included all fixed effects covariates included in the GAMMs, nested random intercepts for individuals nested within families, and both fixed and random effects of the year (see the Measures section) or local COVID-19 transmission rates. Similarly, to assess whether affect or mobility differed in their responsiveness to the local case count over time, we also fit mixed effects models with these same covariates along with a linear case count effect, year effect (where year indicated whether an observation was recorded in 2020, 2021, or 2022), and a case count×year interaction effect (see Table 2 for formal and text-based descriptions of each model). In an attempt to control for seasonal effects, only data collected from January through April in 2019-2022 (coinciding with the largest pandemic-related impacts in 2020) were included in these linear mixed effects models.

We considered several different strategies to reduce the impact of missing surveys and location data. The proportion of missing affect surveys and missing days of GPS location data for each participant were recorded as model covariates. Relationships between missingness, participant demographics, and affect and mobility measures were explored to identify possible attrition effects on outcomes. Additional strategies were considered, such as running analyses in a latent growth curve modeling framework with full information maximum likelihood estimation [51] or multiple imputation of missing data. However, given that predictor variables were only rarely missing, measurement occasions were irregularly spaced, and affect and mobility trajectories were highly nonlinear, we concluded that GAMMs remained better suited to modeling these relationships [45,46]. Similarly, multiple imputation was ultimately not used due to computational constraints and to the large number of missing days of data, implying a large fraction of missing information impacting the accuracy of imputation results [52]. Finally, additional biometric "ACE" decomposition models [53] were devised to investigate the extent to which COVID-19 experiences and corresponding affect and mobility changes were subject to genetic influences, though these models were ultimately not used after initial efforts demonstrated inadequate statistical power (eg, heritability estimates for local COVID-19 case counts that ranged from 0% to 100%) when estimating monthly and quarterly biometric variance parameters (identifying changes in genetic and environmental contributions over time, a significant difference between 2 or more variances, was further underpowered and typically requires thousands of twin pairs, rather than the hundreds available in CoTwins).

XSL•FO
RenderX

Table 2. Description of models, terms included, and related research aims[a].

Research aim	Model	Terms
Quantify mean changes in affect or mobility over time or across COVID-19 transmission levels.	$Y_{ij}=f_1(u_{1ij})+X_{1ij}\beta_1+Z_{1ij}b_1+\varepsilon_{ij}$ (1)	• Y_{ij}: Vector of participant *i* in family *j*'s affect or mobility • $f_1(u_{1ij})$: Smoothing function representing the nonlinear effect of date or local case counts on affect or mobility • $X_{1ij}\beta_1$: Linear effect β_1 of covariates X_{1ij} • $Z_{1ij}b_1$: Random effects (intercept) b_1 for individual *j* nested in family *j* • ε_{ij}: Random error term
Obtain parameter estimates for affect and mobility changes by year and across COVID-19 transmission levels.	$Y_{ij}=X_{2ij}\beta_2+Z_{2ij}b_2+\varepsilon_{ij}$ (2)	• Y_{ij}: Vector of participant *i* in family *j*'s affect or mobility • $X_{2ij}\beta_2$: Linear effect β_2 of covariates X_{2ij}, including an effect of year or local case count • $Z_{2ij}b_2$: Nested random effects b_2 of individual i within family *j*, including a random slope of year or local case count • ε_{ij}: Random error term
Test whether the effect of local COVID-19 case counts on affect or mobility is moderated by year.	$Y_{ij}=X_{3ij}\beta_3+Z_{3ij}b_3+\varepsilon_{ij}$ (3)	• Y_{ij}: Vector of participant *i* in family *j*'s affect or mobility • $X_{3ij}\beta_3$: Linear effect β_3 of covariates X_{3ij}, including an effect of year, local case count, and a year by local case count interaction term • $Z_{3ij}b_3$: Nested random effects b_3 of individual *i* within family *j*, including random slopes of year and case count • ε_{ij}: Random error term

[a]Description of each specific research aim, the corresponding statistical model, and definitions for each term within each model. Models were run using data provided by 887 adolescent and young adult twins initially recruited as high school students in the state of Colorado, who provided biweekly affect surveys and persistent GPS location data over multiple years via smartphone between June 2016 and April 2022.

Results

Descriptive Statistics

Descriptive statistics for affect and mobility measures are provided in Table 3. To assess the extent of reliability of the remote affect surveys over the multiyear duration of the study, Cronbach αs were computed at each assessment for participants' first 70 remote affect surveys. These αs ranged from 0.67 to 0.91 for positive affect (mean α=0.78) and from 0.75 to 0.89 (mean α=0.83) for negative affect, indicating that affect surveys were adequately reliable for the duration of the study. Intraclass correlation coefficients (single random raters) for the affect surveys over time were 0.52 for average positive affect and 0.55 for average negative affect, indicating that individuals' self-reported positive and negative affect varied between assessments but was moderately highly correlated with affect measurements at other time points. Mobility measure intraclass correlation coefficients were 0.14 for daily locations visited and 0.04 for daily travel distance, suggesting highly variable mobility patterns over the multiyear duration of the study. A greater percentage of both location data and affect surveys were missing for younger participants (β=–0.40 to –1.12; all *P*<.001) and Android users (β=0.97-44.05; all *P*<.001). Missing location data were significantly more prevalent in female participants (β=0.22; *P*<.001), while missing affect surveys were more common among participants (β=2.88; *P*<.001).

XSL•FO
RenderX

Table 3. Descriptive statistics for affect and mobility measures[a].

Variable	Grand mean[b]		Within-subject mean[c]		N_{Total}[d]	$N_{Participant}$[e]		% Missing[f]	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Daily locations visited	3.5 (2.1)	3.0 (2.0-5.0)	3.4 (0.9)	3.4 (2.8-3.9)	592,834	689.3 (498.4)	575.0 (271.5-1090.5)	46.1 (26.4)	40.0 (23.7-62.0)
Daily travel distance (km)	20.3 (42.0)	6.4 (0.6-21.2)	18.1 (9.4)	16.9 (11.5-23.4)	592,834	689.3 (498.4)	575.0 (271.5-1090.5)	46.1 (26.4)	40.0 (23.7-62.0)
Positive affect (0-5)	2.9 (0.7)	2.8 (2.4-3.4)	2.9 (0.6)	2.9 (2.5-3.2)	15,501	23.3 (18.0)	19.0 (8-36)	47.1 (24.7)	50.0 (25.0-66.8)
Negative affect (0-5)	2.0 (0.8)	1.8 (1.4-2.6)	2.1 (0.6)	2.0 (1.6-2.5)	15,500	23.3 (18.0)	19.0 (8-36)	47.1 (24.7)	50.0 (25.0-66.8)

[a]Descriptive statistics characterize data provided by 887 adolescent and young adult twins initially recruited as high school students in the state of Colorado, who provided biweekly affect surveys and persistent GPS location data via smartphone over multiple years between June 2016 and April 2022.

[b]The average value for a measure across all participants and days.

[c]The mean of participant-mean values for a measure.

[d]The total number of recorded observations across all participants for a measure.

[e]The average number of observations recorded per participant.

[f]The mean participant's portion of missing days of location data or completed affect surveys (see the Missing Data section for additional information on missingness).

Change Over Time and Pandemic Onset Effects on Affect and Mobility

Table S1 in Multimedia Appendix 1 reports the results of GAMMs of daily locations visited, daily travel distance, positive affect, and negative affect conditioned on a smooth date term, fixed effects covariates, and random intercepts of individuals nested in families. Significant fixed effects of covariates are reported in Table S1 in Multimedia Appendix 1 and included male sex (positive affect and negative affect), age on January 20, 2020 (locations visited and travel distance), weekend day (locations visited and travel distance), Android OS (locations visited, positive affect, and negative affect), percent missing location data (locations visited and travel distance), and percent missing affect surveys (locations visited and travel distance). Effective degrees of freedom tests of nonlinearity [47] indicated that date smooths were significantly nonlinear for all outcome variables ($F_{31.2,33.8}$=29.0-400.7; all P<.001), while k-basis dimension tests [47] suggested that basis dimensions were sufficient to prevent underfitting (k=0.98-1.01; P=.14-.81).

Smooth date effects (adjusted for fixed effects covariates), characterizing the change in mean affect and mobility levels over time, are presented in Figure 1. Prior to the onset of the pandemic in January 2020, mobility measures were largely stable over time, though with seasonal fluctuations of higher mobility in the summer months and lower mobility during the winter. During the first several months of the COVID-19 pandemic, from January to May 2020, both daily locations visited and daily travel distance fell sharply to their lowest levels observed during the study, falling by 29%, from 3.5 to 2.5 locations per day, and by 60%, from 20 to 8 km per day, respectively. Both mobility measures sharply rebounded during the summer of 2020, with daily locations visited increasing by 68% and daily travel distance increasing by 200%. They then plummeted between September and December, with mean daily locations visited falling from a high of 4.2 locations per day to a low of 2.8 locations per day and mean daily travel distance falling from a high of 24 km per day to a low of 12 km per day, coincident with a large outbreak of COVID-19 cases. Beginning in December 2020 and January 2021, roughly coincident with when COVID-19 vaccines were first made broadly available to adults older than 16 years of age, daily mobility measures steadily increased for much of 2021 and 2022, though with significant decreases of ~0.6 locations per day and ~8 km per day between September 2021 and January 2022, possibly reflecting the return of seasonal declines in mobility during the winter observed prior to the pandemic's onset. By mid-April 2022, participants visited an average of 3.3 locations per day with a daily travel distance of 28 km, both of which were similar to levels observed in the spring of 2019.

During 2019, positive affect significantly increased during the summer months and significantly decreased during the fall and winter months (all P<.01). In contrast, negative affect significantly decreased during the summer of 2019, significantly increased in the late summer and fall, and significantly declined in the late fall and winter (all P<.01). Average positive affect saw modest changes following the pandemic's onset in early 2020, initially increasing from a mean of 2.85 in January to 3.0 in February 2020 before declining to 2.79 by April 2020 (a decline of 0.3 SDs), its lowest observed level during the study period. Contrastingly, mean negative affect initially increased from 1.85 to 2.10 between January and March 2020 (an increase of 0.3 SDs) before declining slightly to 1.91 in April and May 2020, concurrent with the implementation of COVID-19 lockdown policies. Over the subsequent 2 years, positive affect remained consistently below prepandemic levels, though with significant fluctuations: declining significantly during 2 large COVID-19 outbreaks in the fall and winter of 2020 and 2021

XSL•FO
RenderX

and increasing significantly in late 2020 or early 2021 (coincident with the rollout of the COVID-19 vaccine) and in early 2022. In contrast, mean negative affect levels continued to modestly fluctuate over the subsequent months and years of the pandemic, though generally remaining slightly above levels observed prior to the pandemic's onset.

The results of mixed effects models quantifying differences in average affect or mobility from January to April in 2019-2022 are presented in Table S2 in Multimedia Appendix 1 as well as in Figure 2. Fixed effects covariate estimates were largely consistent with those obtained via GAMMs (Table S1 in Multimedia Appendix 1). Daily locations visited between January and April 2020, 2021, and 2022 were all significantly below mean levels observed between January and April 2019 (standardized β=–0.24 to –0.10; all P<.001). Similarly, mean daily distance traveled remained below January-April 2019 levels in January-April 2020 and 2021 (standardized β=–0.13 to 0.07; all P<.001), though they were not significantly different from 2019 levels in January-April 2022 (standardized β=–0.03; P=.29). Mean positive affect remained significantly below early 2019 levels in January-April 2020, 2021, and 2022 (standardized β=–0.20 to –0.16; all P<.001). Relative to early 2019 levels, the mean negative affect was not significantly different in January-April 2020 or 2021 (standardized β=0.00-0.09; P=.99-.07) but was significantly higher in January-April 2022 (standardized β=0.14; P=.04).

Figure 1. Pandemic onset effects on affect and mobility measures over time. Affect and mobility measures regressed onto a smooth date term (fit via penalized spline regression). Participants were 887 adolescent and young adult twins initially recruited as high school students in the state of Colorado who provided biweekly affect surveys and persistent GPS location data over multiple years via smartphone between June 2016 and April 2022. Colored bands indicated 95% CIs. Blue segments of the curve indicate areas where the derivative of the curve is significantly greater than 0 (P<.01), while red segments indicate areas where the derivative is significantly less than 0 (P<.01). For reference, the number of daily national COVID-19 cases is presented in dark gray behind each curve. WHO: World Health Organization.

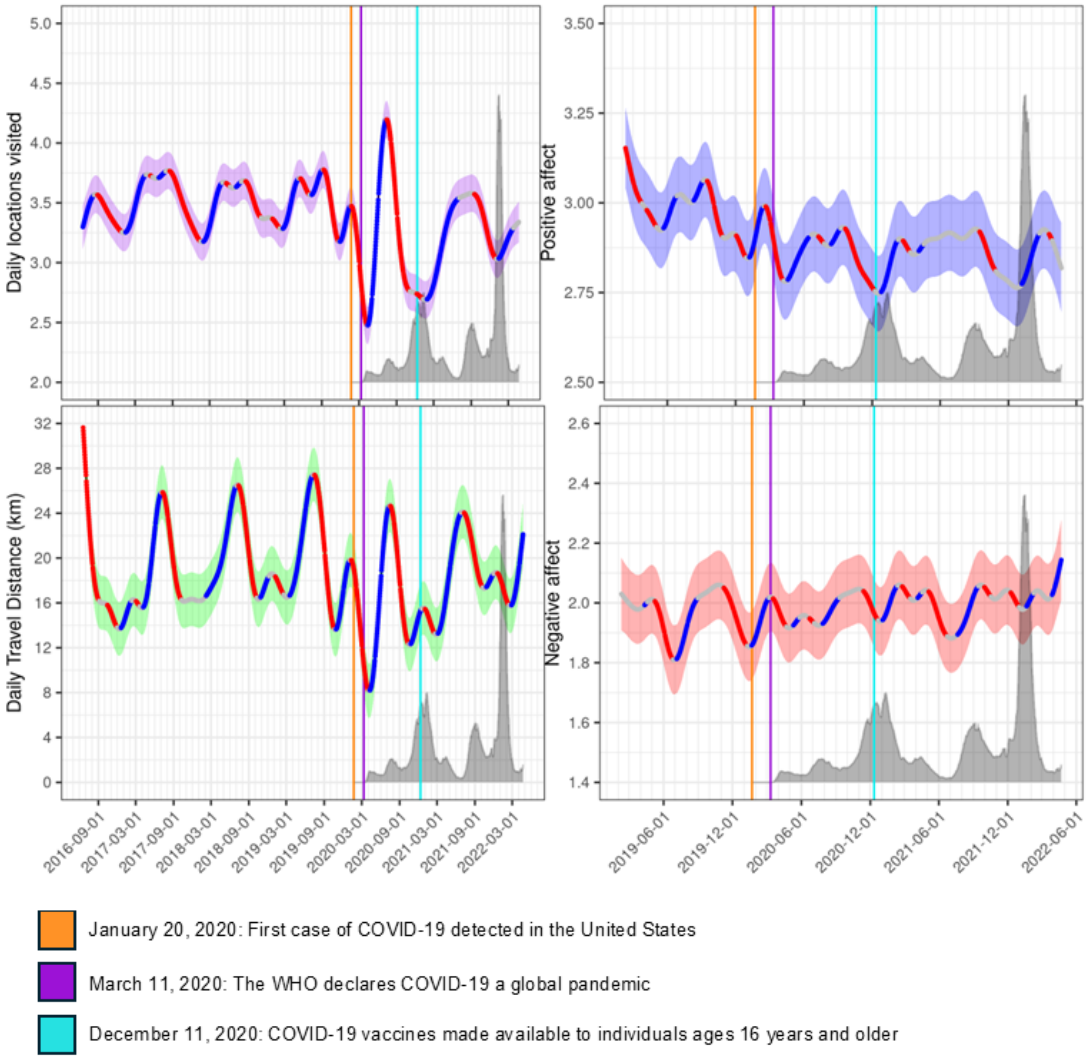

XSL•FO
RenderX

Figure 2. Differences in average affect and mobility: January-April 2019-2022. Predicted affect and mobility values from January 20 to May 1, 2019-2022, according to linear mixed effects models. Estimates are derived from 887 adolescent and young adult twins initially recruited as high school students in the state of Colorado who provided affect surveys and persistent GPS location data via smartphone over multiple years between June 2016 and April 2022. Affect values are represented in SD units for interpretability. Red dashed lines denote the estimated level of each outcome in 2019. Error bars represent 95% CIs.

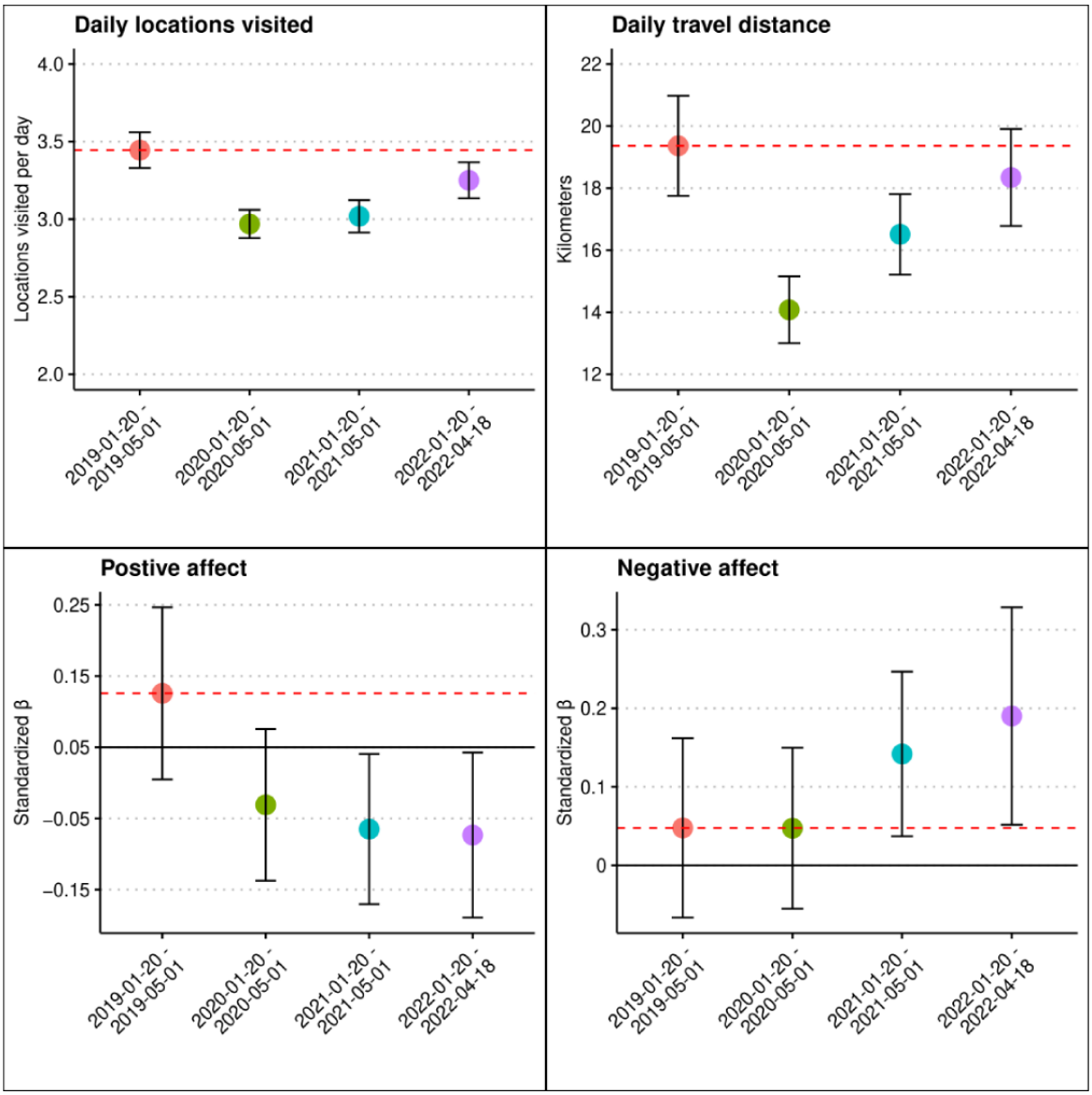

Local Case Count Effects on Affect and Mobility

Results of GAMMs of affect and mobility measures regressed on a smooth effect of local past-week COVID-19 cases per 100,000, fixed effects covariates, and random intercepts of individuals nested in families are presented in Table S3 in Multimedia Appendix 1. Fixed effects covariates showed largely the same relationships to affect and mobility measures as in the smooth date GAMMs (Table S1 in Multimedia Appendix 1). Case count smooths were significantly nonlinear for all outcome variables ($F_{5.77,8.55}$=23.04-338.9; *P*s<.001). *k*-Basis tests were significant for negative affect (*k*-index=0.93; *P*<.001) and daily travel distance (*k*-index=0.97; *P*=.03), suggesting possible underfitting, though these models' effective degrees of freedom, 6.54 and 5.77, respectively, were deemed sufficiently different from *k*, 9.00, that it was not necessary to refit these models with additional basis dimensions [47].

The effects of local case count smooths, corrected for fixed effects covariates, on affect and mobility measures are presented in Figure 3. Increased county-level COVID-19 transmission was associated with small or moderate declines in positive affect and both measures of mobility, though increases above several hundred past-week county-level cases per 100,000 were generally not significantly related to either positive affect or mobility, with the exception of a modest but significant increase in daily locations visited at very high levels of transmission above 750 cases per participant. Correspondingly, negative affect exhibited a small but statistically significant increase of 0.06 SDs, as local COVID-19 incidence increased from 0 to 190 past-week local cases per 100,000, though further increases in local COVID-19 transmission generally exhibited nonsignificant relationships with negative affect.

Mixed effects models testing for interactions between local case count and the year in which the observation was reported are

XSL•FO
RenderX

presented in Table S4 in Multimedia Appendix 1 and visualized in Figure 4. Covariate effects were largely consistent with those observed in previous models. An increase of 100 additional weekly cases per 100,000 was associated with significantly greater reductions in daily locations visited, daily travel distance, and positive affect during 2020 than during 2021 or 2022 (standardized β=0.096-1.527; all P<.001) and with significantly attenuated reductions in negative affect during 2022 than during 2020 (standardized β=0.184; P<.001).

Figure 3. Local case count effects on affect and mobility measures. Affect and mobility measures regressed onto smooth past-week county-level COVID-19 cases per 100,000 people. Trajectories were estimated from biweekly affect surveys and persistent GPS location data provided by 887 adolescent and young adult twins initially recruited as high school students in the state of Colorado, who provided data via smartphone over multiple years between June 2016 and April 2022. Colored bands indicated 95% CIs. Blue segments of the curve indicate areas where the derivative of the curve is significantly greater than 0 (P<.01), while red segments indicate areas where the derivative is significantly less than 0 (P<.01).

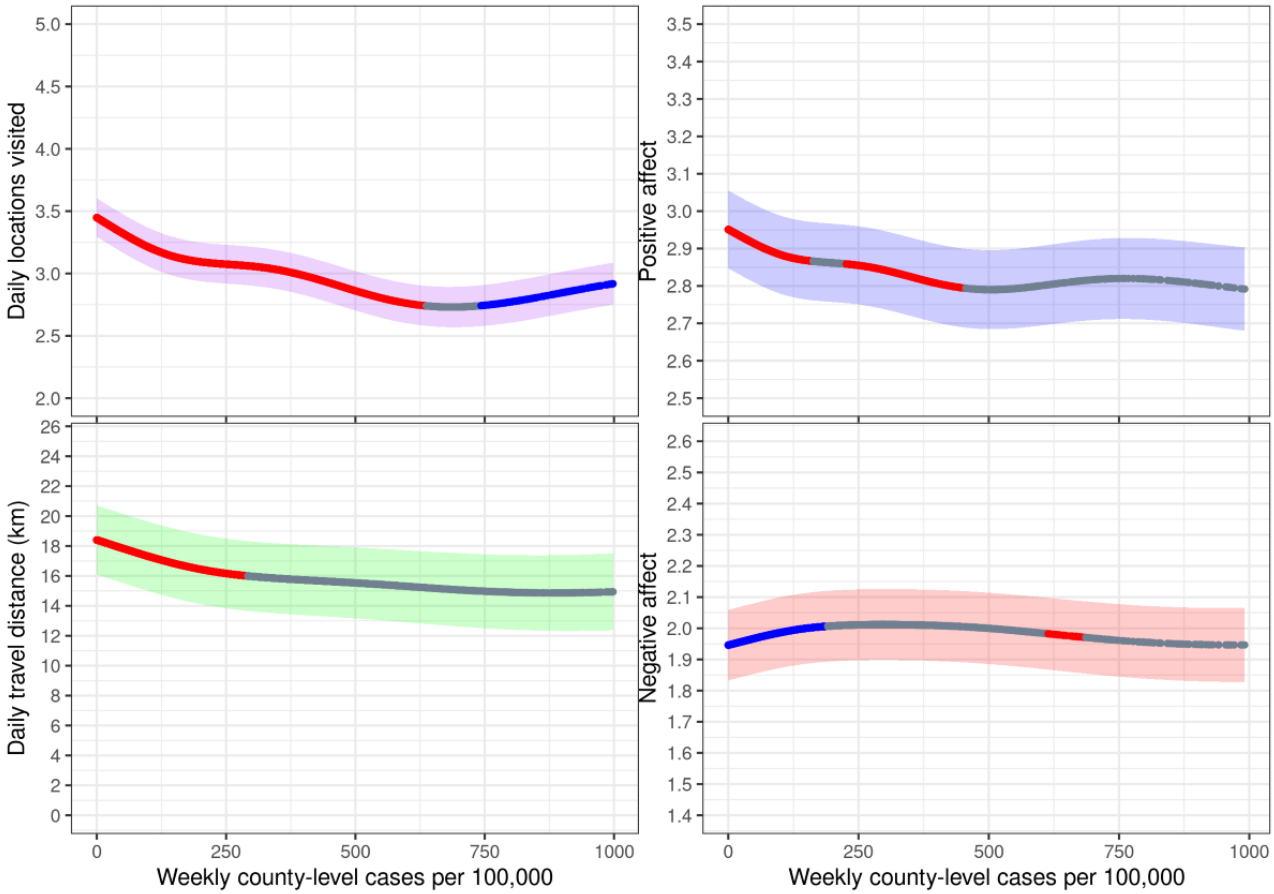

XSL•FO
RenderX

Figure 4. Effects of local case count on affect and mobility measures moderated by date. The linear effect of COVID-19 case counts on affect and mobility between January 20 and May 01 in 2020 (the initial phase of the COVID-19 pandemic), 2021, and 2022 obtained via linear mixed effects models. Estimates are derived from linear mixed effects models of biweekly affect surveys and persistent GPS location data collected from 887 adolescent and young adult twins, who provided data over multiple years via smartphone between June 2016 and April 2022. Shaded regions indicate 95% CIs.

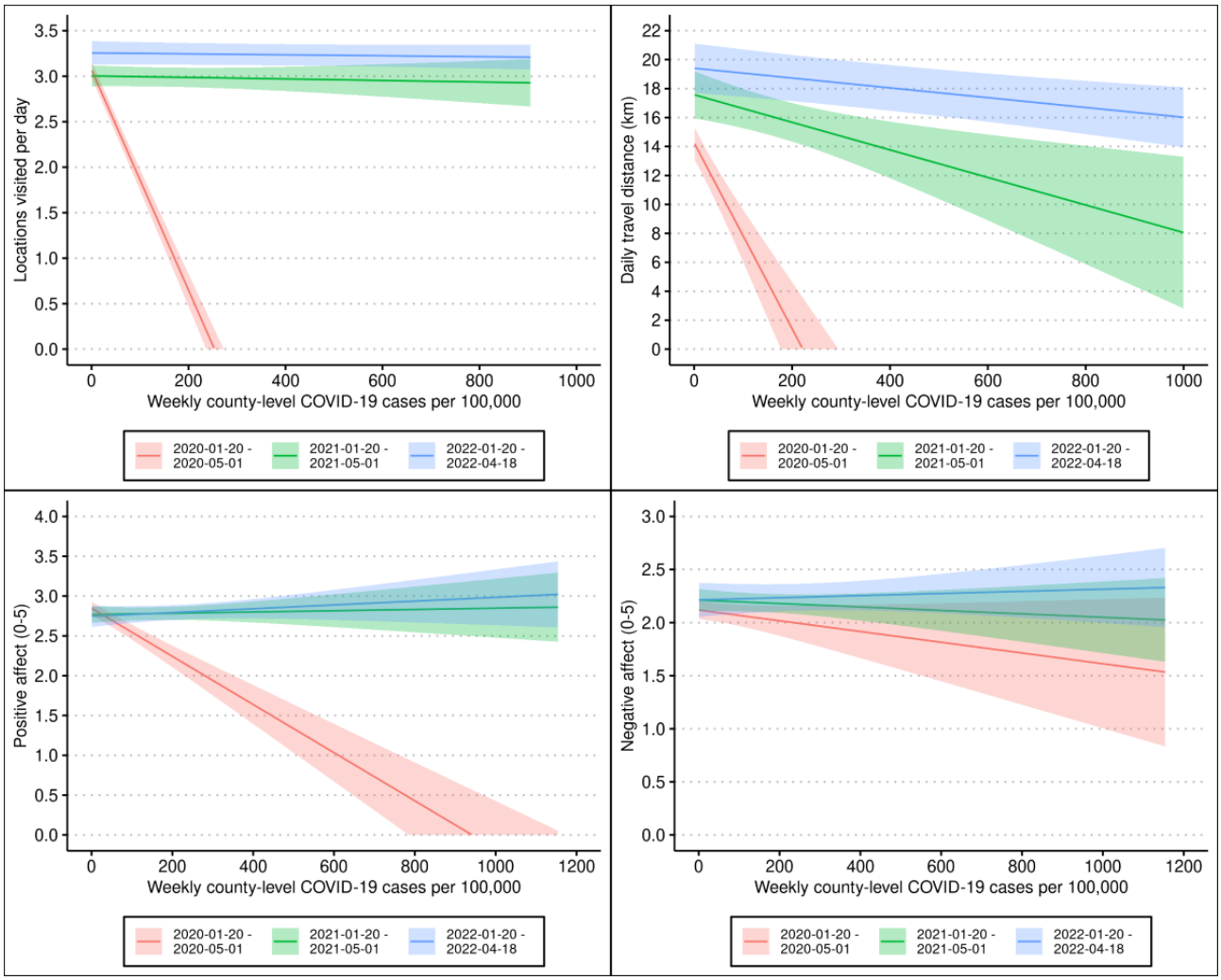

Discussion

Principal Findings and Comparison to Prior Work

This study investigated changes in affect and mobility patterns in American youths throughout the COVID-19 pandemic. As a nearly 6-year intensive longitudinal design, this study represents a lengthy and longitudinally rich investigation into the pandemic's psychological impact on youths.

Consistent with expectations, during the first pandemic months in 2020, we observed large decreases in participants' average locations visited per day and daily travel distance, a moderate decrease in mean positive affect, and a modest increase in mean negative affect. These changes are consistent with prior research conducted in adults [10,14,29,32,34,54], suggesting that youths exhibited behavioral and emotional changes similar to adults. However, we found that mean changes in affect during the first months of the pandemic were quite small; the average positive and negative affect between January 20 and May 1, 2020, differed by less than 0.2 SDs from average levels over the same period in 2019. These youths were, on average, resilient in the face of the pandemic's dramatic uncertainties.

Other studies have noted small or null changes in young adults' positive and negative affect following the pandemic's onset across varying national contexts, including in samples of college students in the northeastern United States [12], Germany [55], and China [56] and in a national web-based survey of British youths [57]. Taken together, both our results and the broader literature suggest that, across a wide range of cultural contexts, policy environments, and COVID-19–related experiences, average impacts on youth affect were likely small in magnitude even at the pandemic's onset. However, these small average impacts do not negate the possibility that youths exposed to particularly stressful life events during the pandemic remained vulnerable, as is suggested by several papers that find youths exhibited greater emotional lability in response to pandemic-related stressors relative to middle-aged and older adults [54,58,59].

Contrary to expectation, daily locations visited and positive affect remained significantly below and negative affect significantly above early 2019 levels through at least April 2022. This incomplete return to mean prepandemic levels more than 2 years after the pandemic's onset is more consistent with trauma-based theories on the psychology of disasters [20,21],

XSL•FO
RenderX

which predict lasting emotional impacts following traumatic experiences like natural disasters. Prior research has noted postpandemic increases in youth anxiety and depression symptoms [6-8] as well as broader secular trends of increasing youth mental health problems both before and during the pandemic [60,61]. Persistent postpandemic affective changes may be a further contributor to increasing youth mental health problems. That said, effect sizes are minimal if significant, and the lasting scar of the pandemic is of uncertain significance in the lives of these youths, at least on average.

These persistent differences—years after the pandemic onset—may reflect psychological scarring or may represent the ongoing spread of COVID-19 in April 2022. Yet, we found that the relationship between affect and case count greatly attenuated with time, showing almost no relationship with affect by 2022. Another explanation is that these differences represent developmental effects associated with participants transitioning from adolescence to early adulthood. However, existing research indicates that the transition from adolescence to emerging adulthood is associated with an increase in positive affect [5] and daily mobility [23] and a decrease in negative affect [5,62], opposite of our findings. Identifying the causes of these persistent affect and mobility differences 2 years after the pandemic's onset is likely beyond the scope of any single study; yet, speculatively, a combination of sociocultural changes, ongoing pandemic-related stressors, and lasting psychological impacts in the aftermath of the pandemic represents reasonable initial hypotheses for contributing factors.

Our hypothesis that local case count would be associated with reduced mobility and positive affect and increased negative affect was partially supported. These effects were strong earlier in the pandemic and decayed later that year and in the months surrounding the availability of vaccines. The reasons for this change could be myriad, including vaccine availability, the loosening of COVID-19 restrictions, a better understanding of the limited effects of infection on youths, or more simply changing attitudes toward the pandemic.

Limitations

Some limitations of this study are of note. Though we find clear changes in affect and mobility upon the pandemic's onset and extending over multiple years because of the complexity of the relationships between COVID-19–related policies, COVID-19 spread, and participant responses (as well as a lack of comprehensive data on COVID-19–related policy implementation), it was not possible in this study to disambiguate the specific features of the pandemic driving these changes, which likely included both individuals' reactions to the spread of COVID-19 and to state and local policies enacted to reduce the pandemic's severity (eg, school and business closures). Participants were adolescents or young adults and tended to be wealthy, educated, and White, albeit 15.6% (n=138) of Hispanic ethnicity, largely consistent with the ethnic diversity of Colorado. Generalization to other socioeconomic or ethnic contexts may not be straightforward. Indeed, generalizability may be particularly challenging for studies of pandemic-related behavior, as experiences of the pandemic varied considerably across geographic and sociodemographic contexts [31,34,35]. Twin participants are necessarily more similar to one another in their behaviors, personalities, and life circumstances than unrelated individuals. Hence, the effective sample size of the study is likely smaller than suggested by the study (N=887). While twin samples can be used to facilitate biometric analyses, such as the extent to which behavioral phenotypes and their relationships are subject to genetic influences, this sample lacked adequate statistical power to undertake these analyses in this study, at least as they pertained to the study's objectives to examine pandemic-related changes.

Missingness remains a particularly important challenge to the collection and interpretation of intensive longitudinal data, especially over multiple years. As expected in lengthy intensive longitudinal studies, both affect surveys and mobility data were subject to substantial missing data [63,64]. Efforts to characterize and limit this impact included covarying for the proportion of missing data in all models and reporting and investigating relationships between attrition, demographics, and model outcomes and covariates. We identified several significant relationships between attrition and demographics, affect, and mobility, suggesting that, though we partially account for missingness via a fixed effects covariate, this strategy may not have fully accounted for missingness effects, especially if missingness was influenced by unanticipated and unmeasured confounds [65].

Conclusions

In summary, we found substantial reductions in daily mobility and modest changes in positive and negative affect following the onset of the COVID-19 pandemic in early 2020, with reductions in daily locations visited and positive affect and increases in negative affect relative to early 2019 levels persisting through at least mid-2022. We further found that increases in local COVID-19 case counts were associated with reduced mobility and positive affect as well as increased negative affect, though with diminishing effects above several hundred cases per 100,000 and weaker effects in 2021 or 2022 relative to 2020. Lasting changes in affect and mobility do not appear to be well explained by the ongoing spread of COVID-19 during mid-2022 and instead may reflect numerous other factors, possibly including lasting pandemic-related harms, sociocultural changes, or disruptions to normative trajectories in social, emotional, professional, or educational development.

Acknowledgments

This research was supported by research grants from the National Institute on Drug Abuse (U01 DA 046413, R01 DA 054087, and R01 DA 042755), the National Center for Advancing Translational Sciences (UL1TR002014-06), and the National Institute of Diabetes, Digestive & Kidney Diseases (R01DK134863).

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not currently publicly available, though deidentified data (excluding GPS coordinates) are currently being uploaded to the University of Michigan's Inter-University Consortium for Political and Social Research web-based data repository. Data are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mixed effects model summaries.
[DOCX File , 526 KB - jmir_v27i1e64965_app1.docx]

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
2. Bergquist S, Otten T, Sarich N. COVID-19 pandemic in the United States. Health Policy Technol 2020;9(4):623-638 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.007] [Medline: 32874854]
3. Breaux R, Cash AR, Lewis J, Garcia KM, Dvorsky MR, Becker SP. Impacts of COVID-19 quarantine and isolation on adolescent social functioning. Curr Opin Psychol 2023;52:101613 [FREE Full text] [doi: 10.1016/j.copsyc.2023.101613] [Medline: 37364468]
4. de Figueiredo CS, Sandre PC, Portugal LCL, Mázala-de-Oliveira T, da Silva Chagas L, Raony Í, et al. COVID-19 pandemic impact on children and adolescents' mental health: biological, environmental, and social factors. Prog Neuropsychopharmacol Biol Psychiatry 2021;106:110171 [FREE Full text] [doi: 10.1016/j.pnpbp.2020.110171] [Medline: 33186638]
5. Furlong A. In: Furlong A, editor. Routledge Handbook of Youth and Young Adulthood. London: Taylor & Francis Group; 2016:1-9.
6. Viner R, Russell S, Saulle R, Croker H, Stansfield C, Packer J, et al. School closures during social lockdown and mental health, health behaviors, and well-being among children and adolescents during the first COVID-19 wave: a systematic review. JAMA Pediatr 2022;176(4):400-409. [doi: 10.1001/jamapediatrics.2021.5840] [Medline: 35040870]
7. Ludwig-Walz H, Dannheim I, Pfadenhauer LM, Fegert JM, Bujard M. Increase of depression among children and adolescents after the onset of the COVID-19 pandemic in Europe: a systematic review and meta-analysis. Child Adolesc Psychiatry Ment Health 2022;16(1):109 [FREE Full text] [doi: 10.1186/s13034-022-00546-y] [Medline: 36587221]
8. Ludwig-Walz H, Dannheim I, Pfadenhauer LM, Fegert JM, Bujard M. Anxiety increased among children and adolescents during pandemic-related school closures in Europe: a systematic review and meta-analysis. Child Adolesc Psychiatry Ment Health 2023;17(1):74 [FREE Full text] [doi: 10.1186/s13034-023-00612-z] [Medline: 37344892]
9. Bera L, Souchon M, Ladsous A, Colin V, Lopez-Castroman J. Emotional and behavioral impact of the COVID-19 epidemic in adolescents. Curr Psychiatry Rep 2022;24(1):37-46 [FREE Full text] [doi: 10.1007/s11920-022-01313-8] [Medline: 35102536]
10. Blasco-Belled A, Tejada-Gallardo C, Fatsini-Prats M, Alsinet C. Mental health among the general population and healthcare workers during the COVID-19 pandemic: a meta-analysis of well-being and psychological distress prevalence. Curr Psychol 2022;43(9):1-12 [FREE Full text] [doi: 10.1007/s12144-022-02913-6] [Medline: 35250245]
11. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
12. Deng W, Gadassi Polack R, Creighton M, Kober H, Joormann J. Predicting negative and positive affect during COVID-19: a daily diary study in youths. J Res Adolesc 2021;31(3):500-516 [FREE Full text] [doi: 10.1111/jora.12646] [Medline: 34448307]
13. Klaiber P, Wen JH, DeLongis A, Sin NL. The ups and downs of daily life during COVID-19: age differences in affect, stress, and positive events. J Gerontol B Psychol Sci Soc Sci 2021;76(2):e30-e37 [FREE Full text] [doi: 10.1093/geronb/gbaa096] [Medline: 32674138]
14. Lades LK, Laffan K, Daly M, Delaney L. Daily emotional well-being during the COVID-19 pandemic. Br J Health Psychol 2020;25(4):902-911 [FREE Full text] [doi: 10.1111/bjhp.12450] [Medline: 32573074]
15. Paulus DL, Vazire S. The self-report method. In: Handbook of Research Methods in Personality Psychology. New York, NY: Guilford Press; 2007.
16. Russell DW, Russell CA. The stress bias in mental health reporting: death anxiety biases mental health self-assessments amongst deployed soldiers. Psychol Serv 2021;18(2):237-248. [doi: 10.1037/ser0000391] [Medline: 31512906]
17. Frederick S, Loewenstein G. Hedonic adaptation. In: Well-Being: The Foundations of Hedonic Psychology. New York, NY: Russell Sage Foundation; 1999.

XSL•FO
RenderX

18. Diener E, Lucas RE, Scollon CN. Beyond the hedonic treadmill: revising the adaptation theory of well-being. In: Diener E, editor. The Science of Well-Being: The Collected Works of Ed Diener. Social Indicators Research Series. Dordrecht: Springer Netherlands; 2009.
19. Lyubomirsky S. Hedonic Adaptation to Positive and Negative Experiences. Oxford: Oxford University Press; 2010.
20. Goldmann E, Galea S. Mental health consequences of disasters. Annu Rev Public Health 2014;35(1):169-183. [doi: 10.1146/annurev-publhealth-032013-182435]
21. Norris FH, Wind, LH. The experience of disaster. In: Mental Health and Disasters. Cambridge: Cambridge University Press; 2009.
22. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
23. Alexander JD, Zhou Y, Freis SM, Friedman NP, Vrieze SI. Individual differences in adolescent and young adult daily mobility patterns and their relationships to big five personality traits: a behavioral genetic analysis. J Res Pers 2022;100:104277. [doi: 10.1016/j.jrp.2022.104277] [Medline: 35991708]
24. Andrade T, Cancela B, Gama J. Mining human mobility data to discover locations and habits. 2019 Presented at: Machine Learning and Knowledge Discovery in Databases; September 16-20, 2019; Würzburg, Germany p. 390-401 URL: https://link.springer.com/chapter/10.1007/978-3-030-43887-6_32#citeas [doi: 10.1007/978-3-030-43887-6_32]
25. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, et al. Human mobility: models and applications. Phys Rep 2018;734:1-74. [doi: 10.1016/j.physrep.2018.01.001]
26. González MC, Hidalgo CA, Barabási AL. Understanding individual human mobility patterns. Nature 2008;453(7196):779-782. [doi: 10.1038/nature06958] [Medline: 18528393]
27. Dahlberg M, Per-Anders Edin PA, Grönqvist E, Lyhagen J, Östh J, Siretskiy A, et al. Effects of the COVID-19 pandemic on population mobility under mild policies: causal evidence from Sweden. ArXiv Preprint posted online on April 20, 2020 [FREE Full text]
28. Galeazzi A, Cinelli M, Bonaccorsi G, Pierri F, Schmidt AL, Scala A, et al. Human mobility in response to COVID-19 in France, Italy and UK. Sci Rep 2021;11(1):13141 [FREE Full text] [doi: 10.1038/s41598-021-92399-2] [Medline: 34162933]
29. Hu S, Xiong C, Yang M, Younes H, Luo W, Zhang L. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. Transp Res Part C Emerg Technol 2021;124:102955 [FREE Full text] [doi: 10.1016/j.trc.2020.102955] [Medline: 33456212]
30. Klein B, LaRock T, McCabe S, Torres L, Privitera F, Lake B, et al. Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak. Network Science Institute, Northeastern University. 2021. URL: https://www.networkscienceinstitute.org/publications/assessing-changes-in-commuting-and-individual-mobility-in-major-metropolitan-areas-in-the-united-states-during-the-covid-19-outbreak [accessed 2023-09-13]
31. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. Lancet Digit Health 2020;2(12):e638-e649 [FREE Full text] [doi: 10.1016/S2589-7500(20)30243-0] [Medline: 33163951]
32. Santana C, Botta F, Barbosa H, Privitera F, Menezes R, Di Clemente R. COVID-19 is linked to changes in the time-space dimension of human mobility. Nat Hum Behav 2023;7(10):1729-1739. [doi: 10.1038/s41562-023-01660-3] [Medline: 37500782]
33. Wellenius GA, Vispute S, Espinosa V, Fabrikant A, Tsai TC, Hennessy J, et al. Impacts of social distancing policies on mobility and COVID-19 case growth in the US. Nat Commun 2021;12(1):3118 [FREE Full text] [doi: 10.1038/s41467-021-23404-5] [Medline: 34035295]
34. Bonaccorsi G, Pierri F, Cinelli M, Flori A, Galeazzi A, Porcelli F, et al. Economic and social consequences of human mobility restrictions under COVID-19. Proc Natl Acad Sci USA 2020;117(27):15530-15535 [FREE Full text] [doi: 10.1073/pnas.2007658117] [Medline: 32554604]
35. Showalter E, Vigil-Hayes M, Zegura E, Sutton, R, Belding E. Tribal mobility and COVID-19: an urban-rural analysis in new mexico. 2021 Presented at: Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications; February 24, 2021; Virtual, United Kingdom. [doi: 10.1145/3446382.3448654]
36. Weill JA, Stigler M, Deschenes O, Springborn MR. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. Proc Natl Acad Sci USA 2020;117(33):19658-19660 [FREE Full text] [doi: 10.1073/pnas.2009412117] [Medline: 32727905]
37. Crone EA, van Duijvenvoorde ACK. Multiple pathways of risk taking in adolescence. Dev Rev 2021;62:100996. [doi: 10.1016/j.dr.2021.100996]
38. Leather NC. Risk-taking behaviour in adolescence: a literature review. J Child Health Care 2009;13(3):295-304. [doi: 10.1177/1367493509337443] [Medline: 19713410]
39. Starke KR, Reissig D, Petereit-Haack G, Schmauder S, Nienhaus A, Seidler A. The isolated effect of age on the risk of COVID-19 severe outcomes: a systematic review with meta-analysis. BMJ Glob Health 2021;6(12):e006434 [FREE Full text] [doi: 10.1136/bmjgh-2021-006434] [Medline: 34916273]

XSL•FO
RenderX

40. Freis SM, Alexander JD, Anderson JE, Corley RP, De La Vega AI, Gustavson DE, et al. Associations between executive functions assessed in different contexts in a genetically informative sample. J Exp Psychol Gen 2024;153(1):70-85. [doi: 10.1037/xge0001471] [Medline: 37668562]
41. Crawford JR, Henry JD. The Positive and Negative Affect Schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 2004;43(Pt 3):245-265. [doi: 10.1348/0144665031752934] [Medline: 15333231]
42. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
43. Zheng Y, Zhang L, Ma Z, Xie X, Ma WY. Recommending friends and locations based on individual location history. ACM Trans Web 2011;5(1):1-44. [doi: 10.1145/1921591.1921596]
44. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45. [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
45. McNeish D, Matta T. Differentiating between mixed-effects and latent-curve approaches to growth modeling. Behav Res Methods 2018;50(4):1398-1414. [doi: 10.3758/s13428-017-0976-5] [Medline: 29067672]
46. Wood S, Scheipl F. gamm4. R. 2020. URL: https://cran.r-project.org/web/packages/gamm4/gamm4.pdf [accessed 2021-08-03]
47. Wood SN. Generalized Additive Models: An Introduction with R. 2nd Edition. Boca Raton, FL: CRC Press/Taylor & Francis Group; 2017.
48. Simpson G. Confidence intervals for derivatives of splines in GAMs. From the Bottom of the Heap. URL: https://fromthebottomoftheheap.net/2014/06/16/simultaneous-confidence-intervals-for-derivatives/ [accessed 2023-11-17]
49. Simpson GL, Singmann H. gratia: Graceful 'ggplot'-Based Graphics and Other Functions for GAMs Fitted Using 'mgcv'. 2023. URL: https://cran.r-project.org/web/packages/gratia/index.html [accessed 2023-08-01]
50. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
51. Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Struct Equ Model Multidiscip J 2001;8(3):430-457. [doi: 10.1207/s15328007sem0803_5]
52. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol 2019;110:63-73 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.02.016] [Medline: 30878639]
53. Beam CR, Kim AJ, Polderman TJC. Chapter 13—Twin studies of complex traits and diseases. In: Tarnoki A, Tarnoki D, Harris J, Segal N, editors. Twin Research for Everyone. Cambridge: Academic Press; 2022.
54. Carstensen LL, Shavit YZ, Barnes JT. Age advantages in emotional experience persist even under threat from the COVID-19 pandemic. Psychol Sci 2020;31(11):1374-1385. [doi: 10.1177/0956797620967261] [Medline: 33104409]
55. Krautter K, Friese M, Hart A, Reis D. No party no joy?—Changes in university students' extraversion, neuroticism, and subjective well-being during two COVID-19 lockdowns. Appl Psychol Health Well Being 2022;14(4):1314-1332. [doi: 10.1111/aphw.12336] [Medline: 34994098]
56. Li HY, Cao H, Leung DYP, Mak YW. The psychological impacts of a COVID-19 outbreak on college students in China: a longitudinal study. Int J Environ Res Public Health 2020;17(11):3933. [doi: 10.3390/ijerph17113933] [Medline: 32498267]
57. Foa RS, Fabian M, Gilbert S. Subjective well-being during the 2020-21 global coronavirus pandemic: evidence from high frequency time series data. PLoS One 2022;17(2):e0263570 [FREE Full text] [doi: 10.1371/journal.pone.0263570] [Medline: 35171919]
58. Reitsema AM, Jeronimus BF, Bos EH, PsyCorona Collaboration, de Jonge P, Leander P. Age differences in hedonic adaptation to societal restrictions? Positive and negative affect trajectories during the first wave of the COVID-19 pandemic in 33 nations. Emotion 2023;23(5):1440-1457. [doi: 10.1037/emo0001149] [Medline: 36107649]
59. Birditt KS, Turkelson A, Fingerman KL, Polenick CA, Oya A. Age differences in stress, life changes, and social ties during the COVID-19 pandemic: implications for psychological well-being. Gerontologist 2021;61(2):205-216 [FREE Full text] [doi: 10.1093/geront/gnaa204] [Medline: 33346806]
60. Armitage JM, Collishaw S, Sellers R. Explaining long-term trends in adolescent emotional problems: what we know from population-based studies. Discov Soc Sci Health 2024;4(1):4-14. [doi: 10.1007/s44155-024-00076-2]
61. Blanchflower DG, Bryson A, Lepinteur A, Piper A. Further evidence on the global decline in the mental health of the young. National Bureau of Economic Research Working Paper Series. 2024. URL: https://www.nber.org/system/files/working_papers/w32500/w32500.pdf [accessed 2024-09-12]
62. Bailen NH, Green LM, Thompson RJ. Understanding emotion in adolescents: a review of emotional frequency, intensity, instability, and clarity. Emot Rev 2018;11(1):63-73. [doi: 10.1177/1754073918768878]
63. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York, NY: Guilford Press; 2013.
64. Lydon-Staley DM, Bassett DS. The promise and challenges of intensive longitudinal designs for imbalance models of adolescent substance use. Front Psychol 2018;9:1576. [doi: 10.3389/fpsyg.2018.01576]
65. Dong Y, Peng CJ. Principled missing data methods for researchers. SpringerPlus 2013;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]

XSL•FO
RenderX

Abbreviations

GAMM: generalized additive mixed model
PANAS: Positive and Negative Affect Schedule

Edited by A Mavragani; submitted 01.08.24; peer-reviewed by I Nehring, M Bujard; comments to author 18.12.24; revised version received 26.01.25; accepted 05.02.25; published 17.03.25.

Please cite as:
Alexander JD, Duffy KA, Freis SM, Chow SM, Friedman NP, Vrieze SI
Investigating the Magnitude and Persistence of COVID-19–Related Impacts on Affect and GPS-Derived Daily Mobility Patterns in Adolescence and Emerging Adulthood: Insights From a Smartphone-Based Intensive Longitudinal Study of Colorado-Based Youths From June 2016 to April 2022
J Med Internet Res 2025;27:e64965
URL: https://www.jmir.org/2025/1/e64965
doi:10.2196/64965
PMID:

©Jordan D Alexander, Kelly A Duffy, Samantha M Freis, Sy-Miin Chow, Naomi P Friedman, Scott I Vrieze. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of the COVID-19 Pandemic on Oldest-Old Social Capital and Health and the Role of Digital Inequalities: Longitudinal Cohort Study

Luca Guido Valla[1], BSc, MSc, PhD; Michele Rossi[2], BSc; Alessandra Gaia[3,4], BSc, MSc, PhD; Antonio Guaita[2], MD; Elena Rolandi[2,5], BSc, MSc

[1]Department of Statistical Sciences, University of Padua, Padua, Italy

[2]Golgi Cenci Foundation, Abbiategrasso, Italy

[3]Centre for Longitudinal Studies, UCL Social Research Institute, University College London, London, United Kingdom

[4]Department of Sociology and Social Research, University of Milano-Bicocca, Milan, Italy

[5]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy

Corresponding Author:
Alessandra Gaia, BSc, MSc, PhD
Centre for Longitudinal Studies
UCL Social Research Institute
University College London
Room 101
55-59 Gordon Square
London, WC1H 0AL
United Kingdom
Phone: 44 02076792000
Email: a.gaia@ucl.ac.uk

Abstract

Background: During the COVID-19 pandemic, information and communication technology (ICT) became crucial for staying connected with loved ones and accessing health services. In this scenario, disparities in ICT use may have exacerbated other forms of inequality, especially among older adults who were less familiar with technology and more vulnerable to severe COVID-19 health consequences.

Objective: This study investigated changes in ICT use, psychological and physical health, and social capital before and after the pandemic among the oldest old population (aged 80 years or older after the pandemic) and explored how internet use influenced these changes.

Methods: We leveraged data from the InveCe.Ab study, a population-based longitudinal cohort of people born between 1935 and 1939 and living in Abbiategrasso, a municipality on the outskirts of Milan, Italy. Participants underwent multidimensional assessment at baseline (2010) and after 2, 4, 8, and 12 years. We restricted our analysis to cohort members who participated in the last wave (ie, 2022) and who did not have a diagnosis of dementia (n=391). We used linear mixed models to assess the impact of COVID-19 and time on changes in social capital, physical and psychological health, and ICT use in a discontinuity regression design while controlling for age, sex, education, and income satisfaction. Then, we assessed the influence of internet use and its interaction with COVID-19 on these changes.

Results: COVID-19 had a significant impact on social relationships (β=–4.35, 95% CI 6.38 to –2.32; *P*<.001), cultural activities (β=–.55, 95% CI –0.75 to –0.35; *P*<.001), cognitive functioning (β=–1.00, 95% CI –1.28 to –0.72; *P*<.001), depressive symptoms (β=.42, 95% CI 0.10-0.74; *P*=.009), physical health (β=.07, 95% CI 0.04-0.10; *P*<.001), and ICT use (β=–.11, 95% CI –0.18 to –0.03; *P*=.008). Internet use predicts reduced depressive symptoms (β=–.56, 95% CI –1.07 to –0.06; *P*=.03) over time. The interaction between internet use and COVID-19 was significant for cultural activities (β=–.73, 95% CI –1.22 to –0.24; *P*=.003) and cognitive functioning (β=1.36, 95% CI 0.67-2.05; *P*<.001).

Conclusions: The pandemic had adverse effects on older adults' health and social capital. Contrary to expectations, even ICT use dropped significantly after the pandemic. Internet users maintained higher psychological health regardless of time and

XSL•FO
RenderX

COVID-19 status. However, COVID-19 was associated with a steeper decline in cognitive functioning among internet nonusers. Policy makers may develop initiatives to encourage ICT adoption among older adults or strengthen their digital skills.

Trial Registration: ClinicalTrials.gov NCT01345110; https://clinicaltrials.gov/study/NCT01345110

(J Med Internet Res 2025;27:e62824) doi:10.2196/62824

KEYWORDS

older adults; information and communication technology; ICT; ICT use; COVID-19; social capital; health; mental health; digital divide

Introduction

Information and Communication Technology Use Among Older Adults in Italy

In the last few decades, the use of information and communication technology (ICT) has become pervasive and consistently more frequent at all ages [1,2]. The digitalization of public services has made it necessary for citizens to develop a certain degree of confidence in using digital devices [3]. In addition to population aging and contextual factors, this situation has contributed to an increase in the number of older adults who use ICT [4].

Indeed, the most recent Eurostat data shows that in 2023, the prevalence of internet use among the Italian population in the age group of 75 years or older has reached 25.7%, up from 13.1% in 2019. However, this strong increase in internet use among the oldest old segments of the population does not even remotely bridge the grey digital divide, that is, differences in technology adoption among the younger and oldest segments of the population. Indeed, internet use among individuals of all ages in Italy is 60 percentage points higher, reaching 87.7% in 2023 [5].

A growing number of studies have focused on different aspects of older individuals' use of ICT, including social media [6], smartphones [7], and computers [8]. Concurrently, the concept of the "digital divide," that is, the difference between those who have access to and use (effectively) digital media and those who do not [9], has become highly relevant in the literature on ICT use among older adults [10,11].

One key area of investigation is the relationship between digital inequalities, access to health services, and social capital formation and maintenance, which is often fueled by digital interactions. Given recent research [12] that has advocated for the adoption of a multidimensional perspective in the study of inequalities, this study focuses on the interrelation between digital inequalities and inequalities in health and social capital in the context of the COVID-19 pandemic using longitudinal data on the oldest old segments of the population (older than 80 years) living in the area of Abbiategrasso (Italy).

This study's context is particularly interesting in the context of research on older adults' habits and health, as well as the impact of COVID-19 on these dimensions. Indeed, Italy has one of the oldest populations in the world. Abbiategrasso is in Lombardy, a region that was among the first to be affected by the COVID-19 pandemic in the West, with dramatic consequences on the mortality rate, health, and social life of older adults.

Social Capital and ICT Use Among Older Adults

One of the main motives behind ICT use is the desire to stay connected with others. As a matter of fact, several existing studies on the issue have focused on the effects of ICT use on social capital, as Warburton et al [13] highlighted in their review. In this respect, one key challenge in this literature is the lack of consensus on the operationalization of social capital [14]. A pioneering work by Putnam et al [15] defined social capital as a set of networks, rules, and trust that yields mutual benefits. An alternative interpretation is the one by Bourdieu [16] who viewed social capital as the sum of potential resources that people can use when they are members of a group. Alternatively, Coleman [17] interpreted social capital as the construct that facilitates collective action within a social structure and distinguishes social capital within and outside the family.

Notwithstanding the aforementioned different interpretations of social capital, its relationship with ICT use is bidirectional. On the one hand, the possibility of staying connected with others is one of the main motives for the use of ICT among the general population and in old age. On the other hand, building and maintaining social relationships, especially intergenerational relationships, is crucial to helping older adults improve their digital skills. However, the vast literature on the relationship between ICT use and social capital in old age (for a review, see [13]) has numerous limitations due to a lack of consensus in the operationalization of social capital, lack of identification of mechanisms that drive the formation and maintenance of social capital in old age, and limited focus on the oldest old people (those aged older than 80 years).

An additional level of difficulty is related to the role of ICT in the formation and maintenance of social capital among older adults. With the increasingly widespread use of digital media, the very concept of social capital may change [18] to include forms of socialization other than physical meetings and in-person participation in groups. Accordingly, research on this topic has considered the role of ICT use in these new forms of social capital [19]. The influence of ICT in stimulating face-to-face interactions (and other forms of social capital in its "traditional" sense, as presented above) has been widely investigated [20,21]. In addition, research has explored the effects of ICT use on dimensions of social capital, such as social participation, social exclusion, and loneliness [22]. Nonetheless, additional research is needed not only to validate and replicate prior research results but also to reflect on how, in the postpandemic context, continuously evolving digital technologies and new forms of digital interactions may foster social capital.

XSL•FO
RenderX

Health and ICT Use Among Older Adults

The relationship between ICT use and health in older individuals seems complex. This is due not only to the multifaceted nature of health but also to the diverse ways ICT can be used in the context of health care, ranging from telehealth to the e-booking of medical appointments to web-based searches for health-related information.

Research on cognitive function shows that nondaily users of digital tools have overall lower cognitive performance than daily users [23], highlighting the beneficial effects of the use of digital tools on memory processes [24]. With respect to psychological health, research shows that people who use ICT experience fewer depressive symptoms and greater well-being [25-28].

The effects of ICT use on physical health in older adults have been studied less extensively. The literature mostly focuses on rehabilitation protocols using digital tools such as mobile apps. In this respect, the use of digital devices has proven to be effective in improving older people's autonomy in daily activities [29] and recovery from falls [30], which are likely to occur in old age [31].

Since multimorbidity is a relatively frequent condition in old age [32], studying the impact of ICT use on the management of related symptoms is a particularly interesting endeavor. Research has shown that the use of digital technologies by older people with two or more long-term health conditions can help them face day-to-day issues related to symptom management and self-reported perceptions of health conditions [33]. Overall, it seems that the use of digital technologies has positive effects on managing symptoms of coexisting pathologies and their perceived severity in old age.

The Impact of the COVID-19 Pandemic on Older Adults' Social Capital, Health, and ICT Use

The COVID-19 pandemic had a remarkable impact on all the dimensions of interest in this research: social capital, health, and ICT use. First, social distancing measures limited older people's ability to meet with family members and friends and hindered the possibility of participating in face-to-face volunteer work, associations, and activities within the local community [34].

Several activities have been moved to the digital sphere. Older people seem to have shifted traditional face-to-face relationships to nonphysical contact through video calls, social messaging, and social media [35]. In this respect, the pandemic seems to have accelerated the digitalization of society, and in fact, the importance of access to and the ability to proficiently use digital technologies in day-to-day activities from socializing to accessing health services. Indeed, recent research has documented a large amount of ICT adoption among the older people population during the pandemic. It seems that the shift from in-person to web-based activities during the most acute phase of the pandemic encouraged older adults to start or continue using digital technologies [36].

With respect to health, the COVID-19 pandemic clearly had important direct and indirect effects on the physical and mental health of older people. First, as widely documented, the severity of symptoms and the COVID-19–related mortality rate increase with age [37]. However, the pandemic may also have had a strong indirect impact on the health and well-being of older people [38,39]. In particular, the congestion of several nations' health services due to the unprecedented demand for health care resulted in delays in diagnostic and therapeutic medical visits for health conditions other than those strictly related to COVID-19. This circumstance may have led to medical conditions that required attention to be overlooked.

Other studies have examined the role of ICT use in recovery from the SARS-CoV-2 infection. On this matter, existing protocols, which are regularly under evaluation [40], show that the use of digital devices to recover from COVID-19 is favored by patients [41] and allows more personalized treatment plans [42]. While limited evidence exists to date, it is reasonable to anticipate a growing body of research in this area in the future.

Research Questions and Hypotheses

The evidence discussed above shows a multifaceted state of the art. While the number of studies investigating this topic has recently increased, existing research still has numerous potential shortcomings that this study aims to address.

First, limited research has been conducted on the effects of ICT use on health and social capital in adults aged 80 years and older, a rapidly growing portion of the population with unique health and social care needs. However, this topic has rarely been studied by researchers. We believe that this methodological choice might lead to an incomplete understanding of the issue. This research aims to fill this gap by studying the effects of ICT use with data from a sample of people around and older than 80 years.

Second, most existing studies were cross-sectional [43,44]. This type of research design is suitable for studying the associations between different dimensions, but it is not optimal for drawing conclusions on the direction of these associations. One limitation that affects the literature on the relationship between ICT use and health is a lack of assessment of the direction of the association between these two dimensions. While some of the aforementioned studies seem to suggest that ICT use drives better health outcomes, good health could be seen as a prerequisite (and a driver) for the acquisition of digital skills and digital technology adoption. Accordingly, longitudinal studies provide a favorable environment to establish the order of events (whether ICT adoption is associated with subsequent health outcomes), aiding in the assessment of causation [45].

Furthermore, analyzing prospective longitudinal cohort data is appropriate for capturing changes over time, which is crucial for studying a rapidly growing and dynamic phenomenon such as ICT use in older individuals.

Third, many studies on this topic have been conducted using qualitative research designs [46,47]. The advantages of these designs include a deeper investigation of subjective experiences and the possibility of investigating complex concepts without being bound to a predefined set of possible responses. However, qualitative studies lack standardized measurement, which is

XSL•FO
RenderX

crucial for tracking evolution over time and assessing the magnitude of effects.

Furthermore, most existing studies have focused on only some of the dimensions considered in this study, namely, the effects of ICT use on cognitive functioning [48] and social relationships [43]. However, to our knowledge, no study to date has explored the evolution of ICT use, cognitive functions, and social relationships concurrently over time while adopting a multidimensional and multidisciplinary approach.

Considering the above, this research's objective is 2-fold. First, this study assesses the effects of the COVID-19 pandemic on changes in the ICT use, social capital, and physical and psychological health of the oldest old population. Second, this study aims to investigate whether ICT use plays a role in mitigating the impact of the COVID-19 pandemic on the aforementioned dimensions.

Specifically, two research questions (RQs) guided the investigation presented in this work.

RQ1: Is there any change in ICT use, psychological and physical health, and social capital before and after the pandemic among the oldest old population?

RQ2: Do higher levels of ICT use have an impact on psychological and physical health and social capital over time in the context of the COVID-19 pandemic?

In light of current knowledge on this topic, we hypothesized that during the COVID-19 pandemic, health and social capital deteriorated while ICT increased. We also hypothesized that prepandemic ICT use is positively associated with social capital and health as it facilitates interpersonal communication, promotes social gatherings, and facilitates recovery from health issues.

Methods

Study Context

The data used in this research were collected as part of the study InveCe.Ab [49], a population-based cohort study conducted in Abbiategrasso, a municipality on the outskirts of the Metropolitan City of Milan (Italy). This study aimed to analyze the effects of aging on cerebral functions and the biological, social, clinical, and neuropsychological factors related to the onset of dementia. The data were collected at baseline (2010) and at follow-ups (2012, 2014, 2018, and 2022). The target population included older adults born between 1935 and 1939 who resided in Abbiategrasso on the prevalence day (November 1, 2009). Data in each wave were collected through a multidimensional assessment that included a social and lifestyle questionnaire, a medical visit, a neuropsychological evaluation, and the collection of anthropometric measures and blood samples.

Participants

The baseline sample of the InveCe.Ab study included 1321 people. Figure 1 shows the flowchart of study participation throughout the observation period (12 years, from 2010 to 2022). In this work, we analyzed data from a subsample composed of people who participated in the 2022 wave (after the COVID-19 outbreak), who were not previously diagnosed with any form of dementia, and who completed the full version of the questionnaire on social and lifestyle variables in the postpandemic wave in 2022 (n=391). We excluded sample members diagnosed with dementia from the analysis because we assumed that cognitive impairment could lead to measurement errors in self-reported health and social relationships, as well as endogeneity. Moreover, we excluded data collected through a shortened version of the social and lifestyle questionnaire since it did not include some of the variables considered in this study. We compared the main sociodemographic features at baseline between the study sample (n=391) and the InveCe.Ab participants who were not included, excluding 156 cases with incident dementia (from 2010 to 2018) and 8 with incomplete baseline assessment (n=766). They were comparable for sex (study sample=57% female, n=223; unselected=53%, n=404 female; P=.17), while the study sample was slightly younger (study sample mean 72.0, SD 1.3; unselected mean 72.2, SD 1.3; P=.02) and highly educated (study sample mean 7.4, SD 3.3; unselected mean 6.6, SD 3.3; P<.001).

XSL•FO
RenderX

Figure 1. Flowchart of the InveCe.Ab study and participants selection for the present investigation. At baseline, we enrolled 1,321 individuals from the target population (recruitment rate: 80.35%). At each follow-up, all alive InveCe.Ab participants still residing in Abbiategrasso were recontacted. Reasons for non-participation at each wave are detailed on the right side of the flow chart, while, on the left side, we reported the number of participants enrolled among those who did not participate in the previous follow-up.

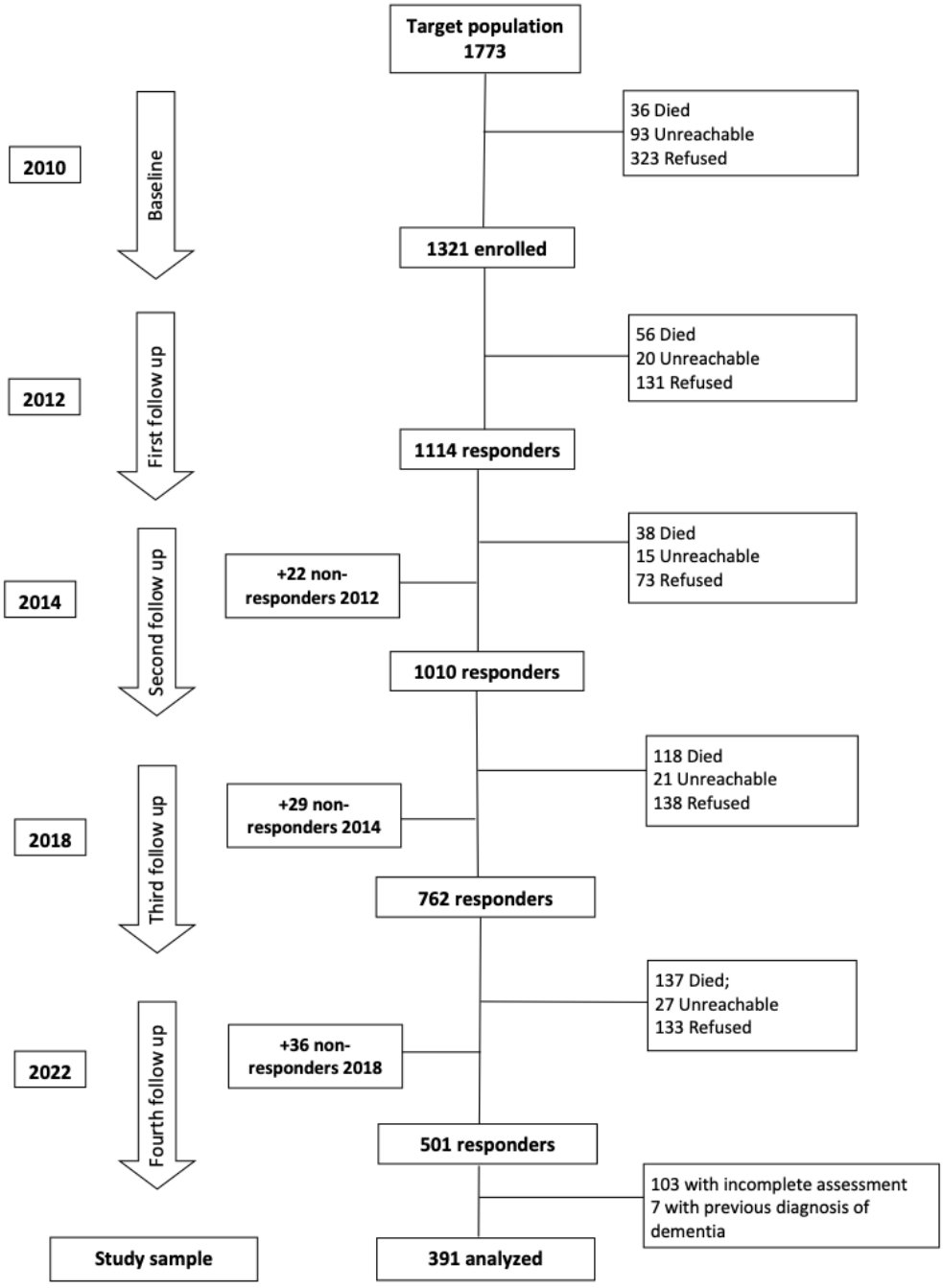

Measures

In the following, we describe the measures used in the analysis.

ICT Use

ICT use was measured with the following three questions: "Do you use a mobile phone?" "Do you use a computer?" and "Do you use the internet?" The answer options were "yes," "no," and "I don't know." A derived variable was computed as the sum of the 3 questions.

Social Capital

To capture the multidimensional nature of social capital as much as possible, we computed two variables, namely, "Social

XSL•FO
RenderX

relationships" and "Cultural activities." The former was the product of the monthly frequency of in-person contact with family, friends, and colleagues, ranging from 0 (there are none or not living together) to 2 (we meet frequently [four or more times a month]), and self-reported satisfaction with these contacts, ranging from 0 (not at all satisfied) to 4 (very satisfied).

The "Cultural activities" variable was computed as the sum of cultural activities (in the last year) that entailed contact with other people. These included going to the cinema or the theatre, attending courses, volunteering, participating in associations or groups, and playing cards. Responses ranged from 0 (no involvement) to 2 (frequent involvement).

Physical Health

This dimension was quantified with the Cumulative Illness Rating Scale (CIRS) [50], a 14-category measure of clinical and functional acuteness. The first 13 elements of the CIRS scale were averaged to form the CIRS severity index.

Psychological Health

The assessment of psychological health comprised 2 dimensions: cognitive functioning and mood. The former was explored with the administration of the Mini-Mental State Examination, which is a screening test used to detect cognitive impairment [51]. The latter was assessed with the Geriatric Depression Scale, a self-reported measure of depressive symptoms in old age [52].

Time

The passing of time was measured with a variable that could take five values (2010, 2012, 2014, 2018, and 2022) corresponding to the years the InveCe.Ab waves took place.

COVID-19

The presence or absence of the COVID-19 pandemic in a specific wave was operationalized with a dichotomous variable that took a value of 1 in 2022 and 0 in all other years.

Sociodemographic Variables

Age at baseline, sex, education, and the perception of the adequacy of one's income were considered relevant confounding factors due to their established association with ICT adoption and use.

Statistical Analysis

Power

To determine the minimum sample size required for the analyses, we conducted a power analysis with G*Power (version 3.1.9.7) [53]. Under the linear multiple regression-random model group, we set the effect size at $\rho 2=0.05$, the significance criterion at $\alpha=0.05$, and the power at 0.80. Our sample size was sufficiently large to conduct the analyses, as it exceeded 309, which is the sample size required for conducting this study, based on the aforementioned calculations.

Data Analysis

Inferential analysis was conducted using R [54] and RStudio [55].

To investigate the effects of the pandemic on changes in the dimensions of interest (RQ1), we ran a series of linear mixed models (LMMs) with social capital, physical and psychological health, and ICT use as the outcomes and time and COVID-19 as the predictors in a discontinuity regression design [56]. Age, sex, education, and the perception of the adequacy of one's income were included as controls. Since we leveraged longitudinal data with repeated measures over the waves, we computed these models with random intercepts for the participants. To further assess the impact of the COVID-19 pandemic in a discontinuity regression framework, we compared the predicted values (PVs) of the postpandemic wave (2022) based on the prespecified models considering the first four waves (2010-2018) and compared them with the observed values (OVs) measured in 2022 with 2-tailed paired-sample *t* tests.

To address RQ2, we used the same approach but included: internet use and the interactions between internet use and time, and between internet use and COVID-19, as predictors, and time, COVID-19, age, sex, education, and the perception of the adequacy of one's income as covariates. The COVID-19 variable was conceived as a dichotomous variable that took a value of 1 in 2022 and 0 in all other years, while time was calculated as years from baseline. We adopted internet use as a predictor (rather than the abovementioned ICT use-derived variable) because web-based activities are likely to have a greater impact on the variables of interest than other forms of ICT use. Thus, this model included a time-varying predictor (internet use) and time-varying covariates.

Ethical Considerations

The study procedures were in accordance with the principles outlined in the Declaration of Helsinki of 1964 and the subsequent amendments. The InveCe.Ab study protocol was approved by the Ethics Committee of the University of Pavia on October 6, 2009 (Committee report 3/2009).

Results

Sample Features

Descriptive statistics are reported in Table 1. On average, the participants were 72 (SD 1.3) years old when the InveCe.Ab study started (2010) and had an average education level of 7.4 (SD 3.3) years. The majority of the sample was composed of female participants (n=223, 57%). Most people perceived their income as adequate or just enough. In the prepandemic wave conducted in 2018, approximately 7% (n=30) of the participants reported that they did not use any ICT devices, while approximately 22% (n=87) were internet users.

XSL•FO
RenderX

Table 1. Descriptive statistics of the sample at baseline (2010).

Variables	Study sample (N=391)
Sociodemographic characteristics	
Age at baseline (years), mean (SD)	72.0 (1.3)
Sex (female), n (%)	223 (57)
Education, mean (SD)	7.4 (3.3)
Perceived adequacy of one's income, n (%)	
Not adequate	30 (7)
Just enough	171 (43.7)
Adequate	190 (48.6)
Variables of interest, mean (SD)	
Social relationships	51.5 (16.8)
Cultural activities	6.6 (1.9)
Physical health (CIRS[a] severity index)	1.5 (0.3)
Cognitive functioning (MMSE[b])	27.8 (2.3)
Depressive symptoms (GDS[c])	2.3 (2.6)
Internet use, n (%)	51 (13)
Computer use, n (%)	72 (18)
Mobile phone use, n (%)	350 (89.5)

[a]CIRS: Cumulative Illness Rating Scale.

[b]MMSE: Mini-Mental State Examination.

[c]GDS: Geriatric Depression Scale.

Over the years, our sample has shown a steady increase in the percentage of internet users, rising from 13% (n=51) in 2010 to 16% (n=64) in 2012, 17% (n=67) in 2014, and eventually reaching 27% (n=106) by 2022. Meanwhile, the proportion of sample members who do not use ICTs has remained relatively stable, fluctuating slightly around 9% throughout the period, with specific figures of 8% (n=34) in 2010, 9% (n=38) in both 2012 and 2014 (n=36), and 10% (n=42) in 2022.

The Impact of the COVID-19 Pandemic

Table 2 shows the results of the LMM models used to evaluate the effect of COVID-19 on the longitudinal changes in the measures of interest.

COVID-19 appears to have reduced social relationships (β=–4.35; conditional R^2=0.58), cultural activities (β=–.55; conditional R^2=0.67), cognitive functioning (β=–1.00; conditional R^2=0.57), and ICT use (β=–.11; conditional R^2=0.77), and increased depressive symptoms (β=.42; conditional R^2=0.58) and multimorbidity (β=.07; conditional R^2=0.57). These effects were independent of time, sex, age, education, and perception of the adequacy of one's income.

These results were further supported by 2-tailed paired-sample *t* tests that tested the differences between the mean PVs on the basis of waves 1 through 4 and the mean OVs in wave 5. Such *t* tests showed that the participants' social relationships (mean difference=–4.37; *P*<.001; PV=48.66; OV=44.39), cultural activities (mean difference=–0.56; *P*<.001; PV=6.36; OV=5.80), cognitive functioning (mean difference=–1.01; *P*<.001; PV=27.49; OV=26.51), and ICT use (mean difference=–0.10; *P*<.001; PV=1.42; OV=1.32) after the spread of the COVID-19 pandemic were significantly lower than expected. Concurrently, multimorbidity (mean difference=0.07; *P*<.001; PV=1.62; OV=1.68) and depression (mean difference=0.43; *P*<.001; PV=2.97; OV=3.42) were significantly greater than expected.

XSL•FO
RenderX

Table 2. Linear mixed models estimating the effect of COVID-19 on social relationships, cultural activities, cognitive functioning, depressive symptoms, physical health, and ICT[a] use.

	Estimates (SE)	95% CI[b]	*P* value	ICC[c]
Social relationships	–4.35 (1.04)	–6.38 to –2.32	<.001[d]	0.53
Cultural activities	–0.56 (0.10)	–0.76 to –0.37	<.001[d]	0.61
Cognitive functioning	–1.00 (0.14)	–1.28 to –0.72	<.001[d]	0.45
Depressive symptoms	0.42 (0.16)	0.10 to 0.74	.009[d]	0.50
Physical Health	0.07 (0.02)	0.04 to 0.10	<.001[d]	0.54
ICT use	–0.11 (0.04)	–0.18 to –0.03	.008[d]	0.69

[a]ICT: information and communication technology.

[b]CIs were computed using the Wald method. The control variables are time, sex, age, education, and income adequacy.

[c]ICC: IntraClass Correlation coefficient.

[d]*P* values <.05.

The Role of Internet Use

Table 3 shows the results of the LMM models assessing the effect of internet use on the longitudinal changes in the measures of interest. Figure 2 shows the changes in the measures of interest across the assessment wave stratified by internet use.

We found that internet use had a significant effect on depressive symptoms (β=–.56; conditional R^2=0.58) while only a marginally significant effect (*P*<.10) was found for social relationships (β=2.91; conditional R^2=0.58) and cultural activities (β=0.28; conditional R^2=0.68). The interaction term between internet use and time was not significant, indicating that the effects of internet use on the measures of interest were stable across waves. However, the interaction between internet use and COVID-19 was significant for cultural activities (β=–0.73; conditional R^2=0.68) and cognitive functioning (β=1.36; conditional R^2=0.58), thus indicating that the effect of internet use on the measures of interest was modified by the pandemic, resulting in a steeper decline for internet users in cultural activities and a steeper decline in cognitive functioning for internet nonusers. A trend toward the significance of the interaction term was further found for physical health (β=–.07; conditional R^2=0.58). These effects were independent of time, COVID-19 status, sex, age, education, and the perception of the adequacy of one's income.

Table 3. Linear mixed models estimating the effect of internet use on social relationships, cultural activities, cognitive functioning, depressive symptoms, and physical health during the COVID-19 pandemic.

	Social relationships	Cultural activities	Cognitive functioning	Depressive symptoms	Physical health
Internet use					
Estimates (SE)	2.91 (1.67)	0.28 (0.17)	0.16 (0.22)	–0.56 (0.26)	0.03 (0.03)
95% CI[a]	–0.37 to 6.18	–0.05 to –0.60	–0.42 to 0.45	–1.07 to –0.06	–0.02 to 0.09
P value	.08	.09	.94	.03[b]	.19
Interaction between internet use and time					
Estimates (SE)	0.10 (0.26)	–0.00 (0.03)	–0.00 (0.04)	–0.01 (0.04)	–0.01 (0.00)
95% CI	–0.41 to 0.62	–0.05 to 0.05	–0.07 to 0.07	–0.09 to 0.07	–0.01 to 0.00
P value	.69	.95	.99	.74	.18
Interaction between internet use and COVID-19					
Estimates (SE)	–0.22 (2.58)	–0.73 (0.25)	1.36 (0.35)	0.30 (0.40)	–0.07 (0.04)
95% CI	–5.27 to 4.83	–1.22 to –0.24	0.67 to 2.05	–0.49 to 1.09	–0.15 to 0.01
P value	.93	.003[b]	<.001[b]	.46	.08
ICC[c]	0.53	0.62	0.45	0.50	0.55

[a]CIs were computed using the Wald method. The control variables are time, sex, age, education, and income adequacy.

[b]*P* values <0.05.

[c]ICC: intraclass correlation coefficients.

XSL•FO
RenderX

Figure 2. Visual representation of the changes in the measures of interest according to internet used status based on the linear mixed models performed (research question 2). The blue line represents internet users, and the red line represents nonusers. CIRS: Cumulative Illness Rating Scale; GDS: Geriatric Depression Scale; MMSE: Mini-Mental State Examination.

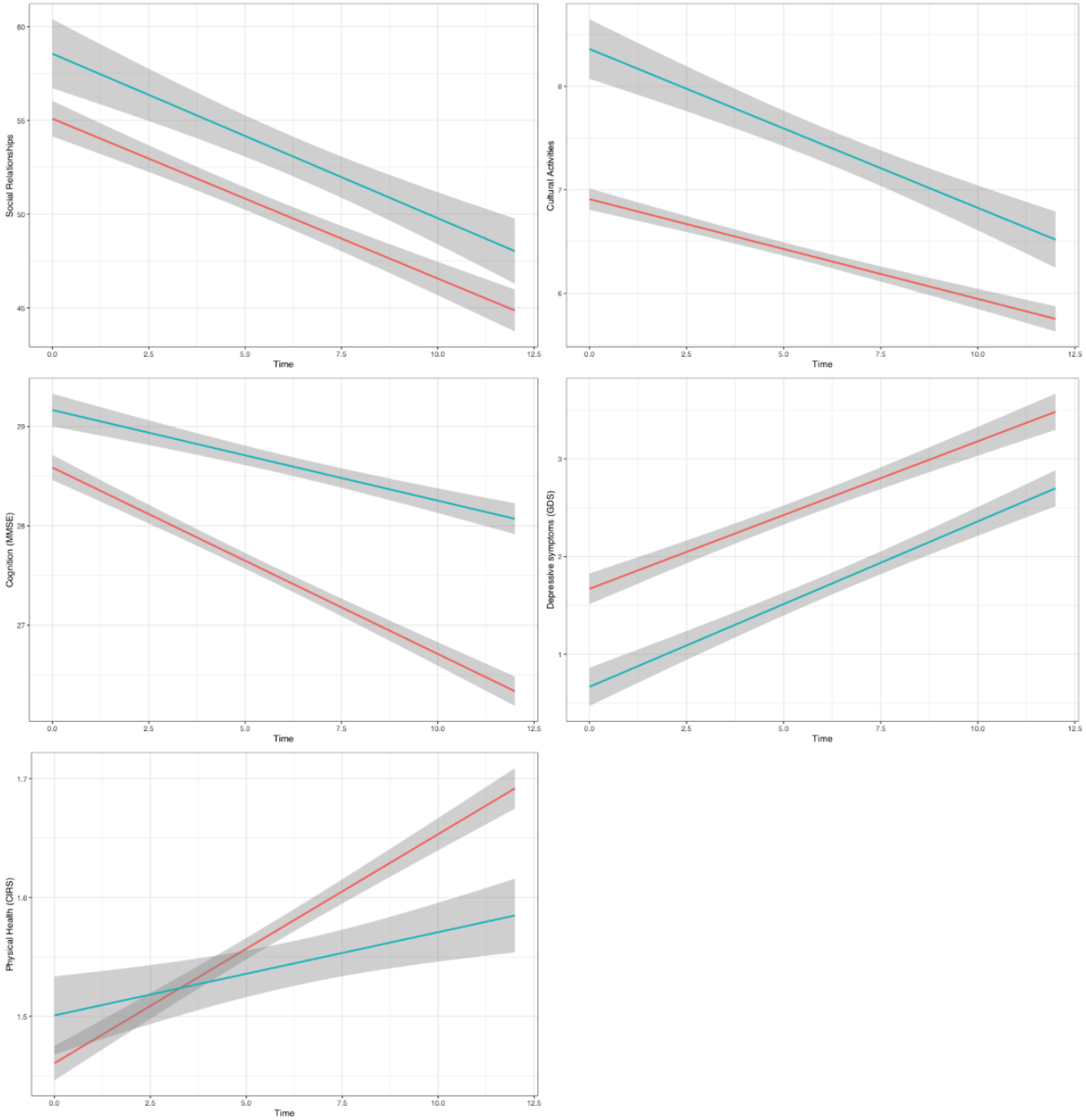

Discussion

Principal Results

This study explored the impact of the COVID-19 pandemic on social capital, health, and ICT use among older adults (RQ1) and investigated the effects of internet use on older adults' social capital and health throughout time and in the context of the COVID-19 pandemic (RQ2).

With regard to RQ1, our results largely confirmed our hypotheses. In particular, the COVID-19 pandemic predicted reduced social capital in terms of social relationships and cultural activities that entail contact with other people. In addition, the results indicated that the pandemic was predictive of worsened physical condition. Consistently, we found that the pandemic predicted lower levels of psychological health, which was assessed through the Mini-Mental State Examination and Geriatric Depression Scale to capture different aspects of this dimension, namely, cognitive functioning and depression. With respect to ICT use, we expected that the COVID-19 pandemic would predict heightened ICT use. This hypothesis has roots in the body of knowledge showing that the pandemic yielded higher levels of use of digital technologies among older adults to stay connected with family and friends as a result of restrictions on in-person contact imposed by authorities. However, this hypothesis was not confirmed by the study results. In contrast, we found that the pandemic predicted a lower level of ICT use.

XSL•FO
RenderX

With respect to RQ2, the results partially confirmed the hypotheses. More specifically, internet use was found to be positively associated with older adults' psychological health. Our results confirmed this hypothesis. Internet use was inversely related to depressive symptoms over time. Additionally, we found that COVID-19 modified the relationship between internet use and cognitive functioning, with nonusers showing a steeper decline in cognitive functioning during the pandemic.

We also expected internet use to be predictive of higher social capital levels. However, internet use did not predict the number of cultural activities that involve contact with other people, such as attending courses and participating in associations. This may be because training and associateship still follow channels other than digital channels. Interestingly, the relationship between internet use and cultural activities was modified by the pandemic and showed a greater reduction in cultural activities related to internet use. During the COVID-19 pandemic, cultural activities that entailed contact with other people were significantly limited and not recommended, especially for older adults. We, therefore, speculate that the impact was greater for those who were habitually more involved in these kinds of activities. Although the participants' sociodemographic characteristics were accounted for in our models, we cannot rule out the possibility that internet users might have a sociodemographic profile that entails greater involvement in cultural activities that imply contact with other people. This is confirmed by internet users' higher levels in this dimension at the baseline, as Figure 2 suggests.

Moreover, internet use is not significantly correlated (at standard significance levels, ie, $P<.05$) with the level of social relationships, which is the other variable considered in this study for the assessment of social capital. This lack of association could be attributed to the fact that our measure of social relationships measures the frequency of in-person interactions with family, friends, and colleagues. Instead, internet use may have primarily facilitated increased communication through digital channels rather than face-to-face contact.

It should be noted, however, that the cut-off for statistical significance may be relaxed in our analysis in light of the relatively small sample size. As such, we consider marginally statistically significant the association between internet use and social relationship ($P=.08$) and cultural activities dimensions ($P=.09$).

Finally, internet use did not affect physical health. A possible explanation for this result is that the oldest-old segments of the population might use the internet mainly for communicating and may lack the skills required for digital access to health care services.

Comparisons With Previous Work

With regard to RQ1, our results align with the findings of previous studies on the negative effects of the COVID-19 pandemic on social activities [57,58]. In relation to physical health, our results are in line with those of other studies showing that older adults are more vulnerable than other age groups to COVID-19 symptoms [59]. This outcome is consistent with existing evidence regarding older people's health susceptibility to unfavorable consequences of viral infections [60]. Regarding the adverse impact of the pandemic on psychological health, our results reflect existing knowledge on the topic that highlights the adverse impact of the pandemic on older adults' health, such as lower cognitive functioning [61] and higher levels of depression [62]. With respect to the negative effect on ICT use, this finding might be explained by the return to in-person contact after the pandemic (in 2022, when our postpandemic data were collected) and the resulting lower levels of digital technology use by older adults. Another aspect should be considered when interpreting these results: given the unexpected and extraordinary nature of the pandemic, older people may have faced these challenging times with the skills and tools that they had already mastered, which did not necessarily include ICT devices. Indeed, in our sample, only 13% (n=51) of older adults used the internet at baseline, and 22% (n=87) used it in the last wave before the pandemic (ie, 2018). Rather than increasing ICT take-up, this age group may have turned to alternative sources of support. Such interpretation may be confirmed by the percentage of people who reported no use of ICT at all (around 10%, n=42) versus the percentage of ICT nonusers in 2018 (around 7%, n=30). Additionally, a possible decline in participants' physical and cognitive conditions may have led to a lower level of ICT use.

Moving to RQ2, the lack of statistically significant predictive effect of internet use on greater social capital is consistent with the groundbreaking study conducted by Kraut et al [63], showing that greater internet use in older age was associated with lower levels of communication with relatives in the household and reduction in the social circle size. This result was partially corroborated by the study by Zhao [64], which showed that people who use the internet for nonsocial purposes do not differ from internet nonusers when it comes to network size.

In relation to the effects of internet use on psychological health, our results are consistent with existing studies that show an overall positive effect of ICT use on older adults' psychological health [65]. The effects of internet use on cognitive functioning may have different interpretations. Several studies have shown that the use of technology fosters cognitive function since it requires several cognitive skills. For instance, Choi and Hart [66] reported that greater ICT use is associated with better performance in tasks related to episodic memory and executive functions. Similarly, Yeung et al [67] demonstrated that higher levels of ICT use are associated with increased working memory and visuospatial ability. On the other hand, we cannot exclude reverse causality, with increased cognitive decline due to COVID-19 explaining a reduction in internet use.

The lack of effects of internet use on physical health is consistent with prior research that highlighted the sense of helplessness and insecurity of older adults in accessing digitally led health care [68]. Such concerns are greater in rural areas, where older adults show even lower levels of digital literacy and technology acceptance associated with health care [69]; hence, these findings may also apply to the suburban context of Abbiategrasso. Further research should explore different types of internet use among the oldest old people.

XSL•FO
RenderX

In summary, our results integrated the current understanding of the topic and added new layers of knowledge. The longitudinal research design allowed us to explore the medium- to long-term effects of the dimensions considered herein, namely, older adults' ICT use, social capital, and health. In addition, we studied the impact of the COVID-19 pandemic on these dimensions. Overall, the pandemic seemed to have a widespread negative impact on all of these aspects. Nonetheless, internet use seemed to predict better psychological health levels over time. Thus, internet use is a protective factor as people age, independent of sociodemographic characteristics and the effect of COVID-19. We also found that COVID-19 modulated the relationships between internet use and several cultural activities and cognitive functioning, although in opposite directions.

This work has several strengths. First, to our knowledge, this is the first work that analyses ICT use, social capital, and health in the oldest old population (aged older than 80 years) and is based on probability-based longitudinal data and multidimensional assessments collected over a 12-year period. In particular, medical and neuropsychological evaluations were combined with a self-reported appraisal of social activities and lifestyles to provide detailed information on each dimension.

Second, participants from a limited geographical area were included in this study, which limited the potential biases that might arise when the sample comprises individuals from different communities. Moreover, the fact that the data collection waves were conducted before and after the most acute phase of the COVID-19 pandemic allowed us to compare the impact of this event on the dimensions of interest.

Finally, by adopting a longitudinal design, we could observe ICT use at baseline and outcomes of interest at later time points. This allowed us to overcome one key methodological limitation of the literature, which has adopted cross-sectional designs and failed to identify the direction of associations between the use of technologies, health, and social capital.

Limitations

This research has several limitations. First, since this was an observational study and did not entail the experimental manipulation of the variables of interest, the exploration of the direction of causal inferences has a certain degree of ambiguity. Additionally, ICT was measured as a dichotomous variable that assessed only use or nonuse. A more refined approach would capture nuances in the frequency of ICT use. Furthermore, the longitudinal data were limited to one wave after the most acute phase of the pandemic. Future research may replicate this analysis should additional survey waves become available. Finally, the generalizability of our results to populations with different sociodemographic characteristics is limited. Future studies may address this issue by conducting similar research in different contexts and among populations with more heterogeneous characteristics.

Conclusions

This study assessed the impact of the pandemic on older adults' health and social capital and explored the role of inequalities in ICT access by leveraging a population-based sample of older adults followed by a multidimensional assessment over a 12-year period.

We believe our results will be of interest not only to the scientific community but also to policy makers by assisting them in assessing the importance of fostering ICT use in old age and drafting policy measures accordingly. First, the pandemic significantly affected multiple dimensions of older adults' lives. Measures aimed at aiding older people are particularly necessary in areas most significantly affected by the pandemic. As internet use seems to have beneficial effects on psychological health in older adults, policy makers may develop initiatives that encourage the adoption of digital devices among older adults or strengthen their digital skills. Such initiatives may include but are not limited to, intergenerational educational programs through which young adults act as "cyber tutors" for older people [70,71] and courses to improve digital literacy in old age [72].

This work seeks to enrich the current body of knowledge on the effects of ICT use on older adults' social capital and health. This study provides direct evidence on the longitudinal trajectory of these relevant dimensions in a population-based sample of older adults residing in a specific context during the COVID-19 pandemic. Future research may address the limitations of this study and further develop a thorough understanding of these rapidly changing topics.

Acknowledgments

We are grateful to Fondazione Serpero which supported data acquisition for the InveCe.Ab fifth wave. We thank Prof Valentina Pacetti for her insightful contributions to the operationalization of social capital. This work has been supported by Fondazione Cariplo (grant 2022-1686).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

AGaia conceived the study and acquired funding for the present work. LGV had a key role in data management, performed the analysis, and drafted the manuscript. ER contributed to the study design, and coordinated and supervised research activities. MR contributed to the study design and planned and supervised data analysis. AGuaita provided resources for data acquisition and

XSL•FO
RenderX

data management. All the authors contributed to the interpretation of the results and reviewed the manuscript for important intellectual content.

Conflicts of Interest

None declared.

References

1. Borgonovi F, Pokropek M. The evolution of the association between ICT use and reading achievement in 28 countries. Comput Educ Open 2021 Dec;2:100047. [doi: 10.1016/j.caeo.2021.100047]
2. Monachesi P. Sustainable development and ICT use among elderly: a comparison between the Netherlands and Italy. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population. Design for the Elderly and Technology Acceptance. HCII 2019. Lecture Notes in Computer Science. Cham, Switzerland: Springer; 2019:450-462.
3. Nevado-Peña D, López-Ruiz VR, Alfaro-Navarro JL. Improving quality of life perception with ICT use and technological capacity in Europe. Technol Forecasting Social Change 2019 Nov;148:119734. [doi: 10.1016/j.techfore.2019.119734]
4. Li W, Ornstein KA, Li Y, Liu B. Barriers to learning a new technology to go online among older adults during the COVID-19 pandemic. J Am Geriatr Soc 2021 Nov;69(11):3051-3057 [FREE Full text] [doi: 10.1111/jgs.17433] [Medline: 34409589]
5. Eurostat. Individuals—internet use. 2024. URL: https://ec.europa.eu/eurostat/databrowser/view/ISOC_CI_IFP_IU/default/table?lang=en [accessed 2024-11-24]
6. Nam SJ. Mediating effect of social support on the relationship between older adults' use of social media and their quality-of-life. Curr Psychol 2019;40(9):4590-4598. [doi: 10.1007/s12144-019-00399-3]
7. Busch PA, Hausvik GI, Ropstad OK, Pettersen D. Smartphone usage among older adults. Comput Hum Behav 2021;121:106783. [doi: 10.1016/j.chb.2021.106783]
8. Zhang S, Boot WR, Charness N. Does computer use improve older adults' cognitive functioning? Evidence from the personal reminder information and social management trial. Gerontologist 2022;62(7):1063-1070. [doi: 10.1093/geront/gnab188]
9. Van Dijk DJ. The Digital Divide. Hoboken: John Wiley & Sons; 2020.
10. Chee SY. Age-related digital disparities, functional limitations, and social isolation: unraveling the grey digital divide between baby boomers and the silent generation in senior living facilities. Aging Mental Health 2024;28(4):621-632. [doi: 10.1080/13607863.2023.2233454] [Medline: 37424361]
11. Zhao Y, Zhang T, Dasgupta RK, Xia R. Narrowing the age‐based digital divide: developing digital capability through social activities. Inf Syst J 2022 Jul 14;33(2):268-298. [doi: 10.1111/isj.12400]
12. Anand P, Chiappero-Martinetti E, Corneo G, Mcknight A, Moro E, O'Brien D, et al. In: Torreiro MD, editor. Multidimensional Perspectives on Inequality: Conceptual and Empirical Challenges. Luxembourg: Publications Office of the European Union; 2020.
13. Warburton J, Cowan S, Bathgate T. Building social capital among rural, older Australians through information and communication technologies: a review article. Australas J Ageing 2013;32(1):8-14. [doi: 10.1111/j.1741-6612.2012.00634.x] [Medline: 23521728]
14. Fukuyama F. Social capital and development: the coming agenda. SAIS Rev 2002;22(1):23-37. [doi: 10.1353/sais.2002.0009]
15. Putnam RD, Leonardi R, Nanetti RY. Making Democracy Work: Civic Traditions in Modern Italy. Princeton, New Jersey: Princeton University Press; 1993.
16. Bourdieu P. The forms of capital. In: Richardson JG, editor. Handbook of Theory and Research for the Sociology of Education. Westport, CT: Greenwood; 1986:241-258.
17. Coleman JS. Social capital in the creation of human capital. Am J Sociol 1988;94:S95-S120. [doi: 10.1086/228943]
18. Ragnedda M, Ruiu ML. Social capital and the three levels of digital divide. In: Ragnedda M, Muschert GW, editors. Theorizing Digital Divides. London, UK: Routledge; 2017:21-34.
19. Van Leeuwen C, Jacobs A, Mariën I. Catching the digital train on time: older adults, continuity, and digital inclusion. Soc Inclusion 2023;11(3):239-250. [doi: 10.17645/si.v11i3.6723]
20. Lee OE, Kim DH, Beum KA. Factors affecting information and communication technology use and eHealth literacy among older adults in the US and South Korea. Educ Gerontol 2020;46(9):575-586. [doi: 10.1080/03601277.2020.1790162]
21. Zhou S, Ogihara A, Nishimura S, Jin Q. Analyzing the changes of health condition and social capital of elderly people using wearable devices. Health Inf Sci Syst 2018;6(1):4 [FREE Full text] [doi: 10.1007/s13755-018-0044-2] [Medline: 29692887]
22. Srivastava SK, Panigrahi PK. Social participation among the elderly: moderated mediation model of information and communication technology (ICT). Commun Assoc Inf Syst 2019;44:698-717. [doi: 10.17705/1cais.04433]
23. Wu YH, Lewis M, Rigaud AS. Cognitive function and digital device use in older adults attending a memory clinic. Gerontol Geriatr Med 2019;5:1-7 [FREE Full text] [doi: 10.1177/2333721419844886] [Medline: 31080848]
24. Chan MY, Haber S, Drew LM, Park DC. Training older adults to use tablet computers: does it enhance cognitive function? Gerontologist 2016;56(3):475-484 [FREE Full text] [doi: 10.1093/geront/gnu057] [Medline: 24928557]
25. Francis J, Kadylak T, Cotten SR, Rikard RV. When it comes to depression, ICT use matters: a longitudinal analysis of the effect of ICT use and mattering on depression among older adults. In: Stephanidis C, editor. HCI International 2016—Posters'

XSL•FO
RenderX

Extended Abstracts. HCI 2016. Communications in Computer and Information Science. Cham, Switzerland: Springer; 2016:301-306.
26. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol, Behav, Soc Networking 2015;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
27. Nedeljko M, Bogataj D, Kaučič BM. The use of ICT in older adults strengthens their social network and reduces social isolation: literature review and research agenda. IFAC-PapersOnLine 2021;54(13):645-650. [doi: 10.1016/j.ifacol.2021.10.524]
28. Sims T, Reed AE, Carr DC. Information and communication technology use is related to higher well-being among the oldest-old. J Gerontol Ser B 2017;72(5):761-770. [doi: 10.1093/geronb/gbw130] [Medline: 27702839]
29. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018;6(2):e34 [FREE Full text] [doi: 10.2196/mhealth.8345] [Medline: 29391342]
30. Hawley-Hague H, Tacconi C, Mellone S, Martinez E, Ford C, Chiari L, et al. Smartphone apps to support falls rehabilitation exercise: app development and usability and acceptability study. JMIR Mhealth Uhealth 2020;8(9):e15460 [FREE Full text] [doi: 10.2196/15460] [Medline: 32985992]
31. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 2006;35(Suppl 2):ii37-ii41. [doi: 10.1093/ageing/afl084] [Medline: 16926202]
32. Salive ME. Multimorbidity in older adults. Epidemiol Rev 2013;35(1):75-83. [doi: 10.1093/epirev/mxs009] [Medline: 23372025]
33. Mateo-Abad M, Fullaondo A, Merino M, Gris S, Marchet F, Avolio F, et al. Impact assessment of an innovative integrated care model for older complex patients with multimorbidity: the CareWell project. Int J Integr Care 2020;20(2):8 [FREE Full text] [doi: 10.5334/ijic.4711] [Medline: 32477037]
34. Shimokihara S, Maruta M, Hidaka Y, Akasaki Y, Tokuda K, Han G, et al. Relationship of decrease in frequency of socialization to daily life, social life, and physical function in community-dwelling adults aged 60 and over after the COVID-19 pandemic. Int J Environ Res Public Health 2021;18(5):2573 [FREE Full text] [doi: 10.3390/ijerph18052573] [Medline: 33806599]
35. Arpino B, Pasqualini M, Bordone V. Physically distant but socially close? Changes in non-physical intergenerational contacts at the onset of the COVID-19 pandemic among older people in France, Italy and Spain. Eur J Ageing 2021;18(2):185-194 [FREE Full text] [doi: 10.1007/s10433-021-00621-x] [Medline: 33935612]
36. Freedman V, Hu M, Kasper JD. Changes in older adults' social contact during the COVID-19 pandemic. J Gerontol Ser B 2022;77(7):e160-e166 [FREE Full text] [doi: 10.1093/geronb/gbab166] [Medline: 34529083]
37. Ferroni E, Rossi PG, Alegiani SS, Trifirò G, Pitter G, Leoni O, et al. Survival of hospitalized COVID-19 patients in Northern Italy: a population-based cohort study by the ITA-COVID-19 network. Clin Epidemiol 2020;12:1337-1346. [doi: 10.2147/clep.s271763]
38. De Pue S, Gillebert C, Dierckx E, Vanderhasselt MA, De Raedt R, Van den Bussche E. The impact of the COVID-19 pandemic on wellbeing and cognitive functioning of older adults. Sci Rep 2021;11(1):4636. [doi: 10.1038/s41598-021-84127-7] [Medline: 33633303]
39. Zaninotto P, Iob E, Demakakos P, Steptoe A. Immediate and longer-term changes in the mental health and well-being of older adults in England during the COVID-19 pandemic. JAMA Psychiatry 2022;79(2):151-159 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3749] [Medline: 34935862]
40. Hyde L, Simpson AJ, Nettleton M, Shepherdson J, Killingback C, Marshall P, et al. Tele-rehabilitation for patients who have been hospitalised with COVID-19: a mixed-methods feasibility trial protocol. Phys Ther Rev 2022;27(3):230-238. [doi: 10.1080/10833196.2022.2028963]
41. Killingback C, Thompson M, Nettleton M, Hyde L, Marshall P, Shepherdson J, et al. Telerehabilitation for patients who have been hospitalised with COVID-19: a qualitative study. Disability Rehabil 2024;46(1):150-158. [doi: 10.1080/09638288.2022.2159075] [Medline: 36629074]
42. Komilova DR, Khusainboyev JD, Rustamov TR. Home-based rehabilitation for COVID-19 patients: a vital step towards recovery. 2023 Presented at: International Conference on Scientific Research in Natural and Social Sciences; September 5, 2023; Toronto, Canada p. 60-64 URL: https://econferenceseries.com/index.php/srnss/article/view/2622
43. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
44. Talmage CA, Knopf RC, Wu T, Winkel D, Mirchandani P, Candan KS. Decreasing loneliness and social disconnectedness among community-dwelling older adults: the potential of information and communication technologies and ride-hailing services. Act Adapt Aging 2020;45(2):89-117. [doi: 10.1080/01924788.2020.1724584]
45. Closer. Strengths of longitudinal data. Learning Hub. 2024. URL: https://learning.closer.ac.uk/learning-modules/introduction/what-can-longitudinal-studies-show-us/strengths-of-longitudinal-studies/ [accessed 2024-10-24]

XSL•FO
RenderX

46. Berg T, Winterton R, Petersen M, Warburton J. 'Although we're isolated, we're not really isolated': the value of information and communication technology for older people in rural Australia. Australas J Ageing 2017;36(4):313-317. [doi: 10.1111/ajag.12449] [Medline: 28639293]
47. Quan-Haase A, Mo GY, Wellman B. Connected seniors: how older adults in East York exchange social support online and offline. Inf Commun Soc 2017;20(7):967-983. [doi: 10.1080/1369118x.2017.1305428]
48. Kim YK, Han SH. Internet use and cognitive functioning in later life: focus on asymmetric effects and contextual factors. Gerontologist 2022;62(3):425-435 [FREE Full text] [doi: 10.1093/geront/gnab149] [Medline: 34614179]
49. Guaita A, Colombo M, Vaccaro R, Fossi S, Vitali SF, Forloni G, et al. Brain aging and dementia during the transition from late adulthood to old age: design and methodology of the "Invece.Ab" population-based study. BMC Geriatr 2013;13(1):98. [doi: 10.1186/1471-2318-13-98] [Medline: 24063518]
50. Parmelee PA, Thuras PD, Katz IR, Lawton MP. Validation of the cumulative illness rating scale in a geriatric residential population. J Am Geriatr Soc 1995;43(2):130-137. [doi: 10.1111/j.1532-5415.1995.tb06377.x] [Medline: 7836636]
51. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
52. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
53. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
54. R Core Team RA. Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
55. Posit team. RStudio: Integrated Development Environment for R. Boston, MA: Posit Software PBC; 2023.
56. Bor J, Moscoe E, Mutevedzi P, Newell ML, Bärnighausen T. Regression discontinuity designs in epidemiology. Epidemiology 2014;25(5):729-737. [doi: 10.1097/ede.0000000000000138]
57. Choi EY, Farina MP, Wu Q, Ailshire J. COVID-19 social distancing measures and loneliness among older adults. J Gerontol Ser B 2022;77(7):e167-e178. [doi: 10.1093/geronb/gbab009] [Medline: 33428753]
58. Krendl AC, Perry BL. The impact of sheltering in place during the COVID-19 pandemic on older adults' social and mental well-being. J Gerontol Ser B 2021;76(2):e53-e58 [FREE Full text] [doi: 10.1093/geronb/gbaa110] [Medline: 32778899]
59. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
60. Thompson WW, Shay DK, Weintraub E, Brammer L, Cox N, Anderson LJ, et al. Mortality associated with influenza and respiratory syncytial virus in the United States. JAMA 2003;289(2):179-186. [doi: 10.1001/jama.289.2.179] [Medline: 12517228]
61. Sardella A, Chiara E, Alibrandi A, Bellone F, Catalano A, Lenzo V, et al. Changes in cognitive and functional status and in quality of life of older outpatients during the COVID-19 pandemic. Gerontology 2022;68(11):1285-1290 [FREE Full text] [doi: 10.1159/000525041] [Medline: 35671739]
62. Briggs R, McDowell CP, De Looze C, Kenny RA, Ward M. Depressive symptoms among older adults pre- and post-COVID-19 pandemic. J Am Med Direct Assoc 2021;22(11):2251-2257 [FREE Full text] [doi: 10.1016/j.jamda.2021.09.003] [Medline: 34597531]
63. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukophadhyay T, Scherlis W. Internet paradox: a social technology that reduces social involvement and psychological well-being? Am Psychol 1998;53(9):1017-1031. [doi: 10.1037/0003-066X.53.9.1017]
64. Zhao S. Do internet users have more social ties? A call for differentiated analyses of internet use. J Comput-Mediated Commun 2006;11(3):844-862. [doi: 10.1111/j.1083-6101.2006.00038.x]
65. Fang Y, Chau AKC, Wong A, Fung HH, Woo J. Information and communicative technology use enhances psychological well-being of older adults: the roles of age, social connectedness, and frailty status. Aging Mental Health 2018;22(11):1516-1524. [doi: 10.1080/13607863.2017.1358354] [Medline: 28777010]
66. Choi S, Hart PS. The influence of different efficacy constructs on energy conservation intentions and climate change policy support. J Environ Psychol 2021;75:101618. [doi: 10.1016/j.jenvp.2021.101618]
67. Yeung DY, Chow LC, Ho AKK, Chung EKH. The effect of information and communications technology use on the well-being of older Hong Kong Chinese adults. Educ Gerontol 2022;49(8):710-725. [doi: 10.1080/03601277.2022.2160131]
68. Raja M, Uhrenfeldt L, Galvin KT, Kymre IG. Older adults' sense of dignity in digitally led healthcare. Nurs Ethics 2022;29(6):1518-1529 [FREE Full text] [doi: 10.1177/09697330221095140] [Medline: 35724431]
69. Park HK, Chung J, Ha J. Acceptance of technology related to healthcare among older Korean adults in rural areas: a mixed-method study. Technol Soc 2023;72:102182. [doi: 10.1016/j.techsoc.2022.102182]
70. Casanova G, Zaccaria D, Rolandi E, Guaita A. The effect of information and communication technology and social networking site use on older people's well-being in relation to loneliness: review of experimental studies. J Med Internet Res 2021;23(3):e23588 [FREE Full text] [doi: 10.2196/23588] [Medline: 33439127]

XSL•FO
RenderX

71. Rolandi E, Sala E, Colombo M, Vaccaro R, Guaita A. Designing an innovative intergenerational educational program to bridge the digital divide: the cyber school for grandparents initiative. In: Gao Q, Zhou J, editors. Human Aspects of IT for the Aged Population. Technology in Everyday Living. HCII 2022. Lecture Notes in Computer Science. Cham, Switzerland: Springer; 2022:398-412.
72. Martin A. Digital literacy for the third ageustaining identity in an uncertain world. ELearning Pap 2009;12:1-12 [FREE Full text]

Abbreviations

CIRS: Cumulative Illness Rating Scale
ICT: information and communications technology
LMM: linear mixed model
OV: observed value
PV: predicted value
RQ: research question

Edited by A Schwartz; submitted 07.06.24; peer-reviewed by Y Si, S Shimokihara; comments to author 25.07.24; revised version received 15.08.24; accepted 15.11.24; published 09.01.25.

Please cite as:
Valla LG, Rossi M, Gaia A, Guaita A, Rolandi E
The Impact of the COVID-19 Pandemic on Oldest-Old Social Capital and Health and the Role of Digital Inequalities: Longitudinal Cohort Study
J Med Internet Res 2025;27:e62824
URL: https://www.jmir.org/2025/1/e62824
doi:10.2196/62824
PMID:

©Luca Guido Valla, Michele Rossi, Alessandra Gaia, Antonio Guaita, Elena Rolandi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Association Between the Digital Divide and Health Inequalities Among Older Adults in China: Nationally Representative Cross-Sectional Survey

Mengqiu Wu[1*], BS; Yongxi Xue[2], BS; Chengyu Ma[2*], PhD

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Genetics, Peking University Cancer Hospital & Institute, Beijing, China

[2]School of Public Health, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Chengyu Ma, PhD
School of Public Health
Capital Medical University
10 Xitoutiao, Youanmenwai
Fengtai District
Beijing, 100069
China
Phone: 86 10 83911576
Email: machengyu@ccmu.edu.cn

Abstract

Background: Health inequalities among older adults become increasingly pronounced as aging progresses. In the digital era, some researchers argue that access to and use of digital technologies may contribute to or exacerbate these existing health inequalities. Conversely, other researchers believe that digital technologies can help mitigate these disparities.

Objective: This study aimed to investigate the relationship between the digital divide and health inequality among older adults and to offer recommendations for promoting health equity.

Methods: Data were obtained from the 2018 and 2020 waves of the China Health and Retirement Longitudinal Study. Physical, mental, and subjective health were assessed using the Activities of Daily Living (ADL) scale, the Instrumental Activities of Daily Living scale, the Mini-Mental State Examination scale, and a 5-point self-rated health scale, respectively. The chi-square and rank sum tests were used to explore whether internet use and access were associated with health inequality status. After controlling for confounders, multiple linear regression models were used to further determine this association. Sensitivity analysis was conducted using propensity score matching, and heterogeneity was analyzed for different influencing factors.

Results: The 2018 analysis highlighted widening health disparities among older adults due to internet access and use, with statistically significant increases in inequalities in self-rated health (3.9%), ADL score (5.8%), and cognition (7.5%). Similarly, internet use widened gaps in self-rated health (7.5%) and cognition (7.6%). Conversely, the 2020 analysis demonstrated that internet access improved health disparities among older adults, reducing gaps in self-rated health (3.8%), ADL score (2.1%), instrumental ADL score (3.5%), and cognition (7.5%), with significant results, except for ADL. Internet use also narrowed disparities, with significant effects on self-rated health (4.8%) and cognition (12.8%). The robustness of the results was confirmed through propensity score–matching paired tests. In addition, the study found heterogeneity in the effects of internet access and use on health inequalities among older adults, depending on sex, age, education, and region.

Conclusions: The impact of internet access and use on health inequalities among older adults showed different trends in 2018 and 2020. These findings underscore the importance of addressing the challenges and barriers to internet use among older adults, particularly during the early stages of digital adoption. It is recommended to promote equitable access to the health benefits of the internet through policy interventions, social support, and technological advancements.

(J Med Internet Res 2025;27:e62645) doi:10.2196/62645

KEYWORDS

older adults; digital divide; internet use; internet access; health inequalities

XSL•FO
RenderX

Introduction

Background

According to data from China's National Bureau of Statistics, by the end of 2023, the population aged ≥60 years in China had surpassed 290 million. This aging trend exceeds the world average and continues to rise. In response, China has introduced a national strategy to address population aging during the 14th Five-Year Plan (2021-2025) period. Health inequalities among older adults are widespread across various age groups and social structures [1]. These disparities are influenced by factors such as individual characteristics, the distribution of medical resources, regional economic development, and public policies [1]. They include differences based on socio-economic status [2,3], sex [4], education [5], and area of residence [6].

Meanwhile, the rise of information and communication technology has introduced digital health as a substantially factor in improving health outcomes. According to findings published by the China Internet Network Information Center [7] in 2020, only 24% of the older adult population in China has adopted internet use. Older adults face substantial challenges in accessing digital resources compared to other demographics, leading to a new form of social structural inequality known as the "digital divide" [8]. This divide is more pronounced among older adults due to limited internet access, the complexity of operating digital devices, and slower acceptance of new technologies [9].

Previous research presents differing views on the relationship between health inequity and the digital divide. One perspective holds that digital technology can enhance health, particularly in remote areas, through telemedicine, optimizing resource distribution, and promoting health equity [8]. Conversely, another perspective suggests that the digital divide may widen health disparities between older and middle-aged adults as well as among different groups of older adults. Digital health technologies, such as electronic medical records, telemedicine, disease surveillance, diagnostics, and drug discovery, are well established [10]. However, barriers to online health care modalities among older adults exacerbate health disparities, affecting equity in health care use [11]. For instance, some researchers have highlighted that inequality in access to and use of digital health is one of the key factors contributing to increased mortality rates among older adults during the COVID-19 pandemic [12].

Most previous studies have focused on the impact of internet use on health, yet it remains unclear whether such use has widened or narrowed health inequalities among different older adult populations. Given the varied effects of internet use across different demographics and the scarcity of research on its relationship with health inequalities, this paper uses cross-sectional data from China to construct a health inequality index. The aim is to explore the correlation between the digital divide and health inequality among older adults, thereby addressing gaps in existing research and providing strategies for improving health equity for older adults. This paper contributes by exploring how internet technology can be leveraged to improve the health status and equity of older adults, ultimately providing effective solutions to enhance social equity and reduce the digital divide.

Status of Health Inequalities Among Older Adults

Older adults bear a greater health burden as a group considered vulnerable, and health inequalities are more pronounced due to their living conditions and health-related behaviors. Health encompasses more than just the physical integrity of an individual [13]; it represents a comprehensive state of well-being requiring evaluation through various metrics. Health inequality indicators can be measured through subjective and objective measures. Subjective indicators are generally assessed by calculating the relative deprivation index using self-reported health status [14], while objective indicators encompass both physical and psychological levels. Physical health is commonly evaluated through one's ability to perform activities of daily living (ADL) [15] and instrumental ADL (iADL) [16]. Mental health in older adults, specifically depression and cognitive abilities, can be quantified using validated tools such as the Center for Epidemiologic Studies Depression Scale, the Symptom Checklist-90, the Montreal Cognitive Assessment, and the Mini-Mental State Examination (MMSE).

Various factors influence health inequalities among older adults, including residence area (urban or rural) [17,18], socioeconomic status [19], educational attainment [20], race [21], and early life experiences [22]. For instance, men from lower-income groups tend to have higher rates of smoking, leading to a higher prevalence of smoking-related illnesses, thereby contributing to health inequalities among different groups of older adults [23].

Dimensions of the Digital Divide

The digital divide refers to disparities in accessing, using, and benefiting from digital technologies among various groups. Within the framework of the 3-stage digital divide, this concept can be divided into 3 dimensions: the digital access divide, digital use divide, and digital outcome divide. The accessibility divide arises from limited internet connectivity; lack of computing devices, software, and accessories; and reduced interest in using the internet [24]. Internet access is typically assessed through household devices and network conditions [25,26]. With an increase in internet penetration, the access divide has narrowed, shifting focus toward diversified differences in use patterns and application depth, known as the "internet use divide." The use divide is related to technical expertise and effective use of digital technologies [27]. Scholars [28,29] have attempted to quantify these differences through use patterns and digital skills to measure individual use variations. The outcome divide dimension refers to differences in offline benefits derived from internet use [26,30]. The weak information acquisition and mastery abilities of older adults are proposed as key reasons for the emergence of the digital divide [31]. Thus, this study focuses on the impact of access and use divides on health inequalities.

Association Between Health Inequalities and Internet Use

The impact of the digital divide on health inequalities remains contested. Some scholars argue that the spread of the internet

XSL•FO
RenderX

has bridged the access divide to some extent, changing health information dissemination patterns and facilitating balanced information flow [32]. This can help reduce health inequalities among older adults, who are often disadvantaged in accessing information [33]. Services derived from digital technology, such as remote consultations, have the potential to alleviate imbalances in health care resource distribution, improving disparities between urban and rural health outcomes [34,35]. However, other studies suggest that while telemedicine can enhance health care access, it may not significantly impact health inequalities [36].

Conversely, it has been argued that the digital divide can exacerbate the inverse care law, potentially excluding those who would benefit most from digitally delivered health care [37]. Research has indicated that older adults may experience negative health effects due to the digital divide hindering their access to mobile health services [38]. A study in the Netherlands found that while older adults benefit from internet use, they do so to a lesser extent than younger people, which further exacerbates health disparities between age groups [39]. Research from Finland suggests that online health care services may intensify existing health disparities due to inequalities in access, skill, and use [40]. In addition, factors such as age [41], sex [42], socioeconomic status [43], and education level [44] influence digital health literacy [45] and telemedicine use, which in turn affect older adults' ability to understand and use digital health information and services. However, most previous studies exploring the relationship between the digital divide and health inequalities are review articles, with relatively few cross-sectional studies. Moreover, these studies often have limitations such as small sample sizes, unbalanced sex distributions, and higher educational attainment, which may introduce biases

Objectives

On the basis of the 2018 and 2020 cross-sectional data from the China Health and Retirement Longitudinal Study (CHARLS), this study addresses the limitations of previous research by examining the impact of internet use on health inequality among older adults at a micro level. It focuses on the digital access divide and use divide to investigate their influence on older adults' health equity, including self-rated health, ADL, iADL [17], and MMSE scores. The theoretical framework is shown in Figure 1. This study is guided by the following research questions: (1) Does access to and use of the internet exacerbate or bridge health disparities among older adults? and (2) Does the impact of internet access and use on health inequality among older adults vary across different demographics?

Figure 1. Theoretical framework ("hukou" is a Chinese term for "household registration"). ADL: activity of daily living; iADL: instrumental activity of daily living.

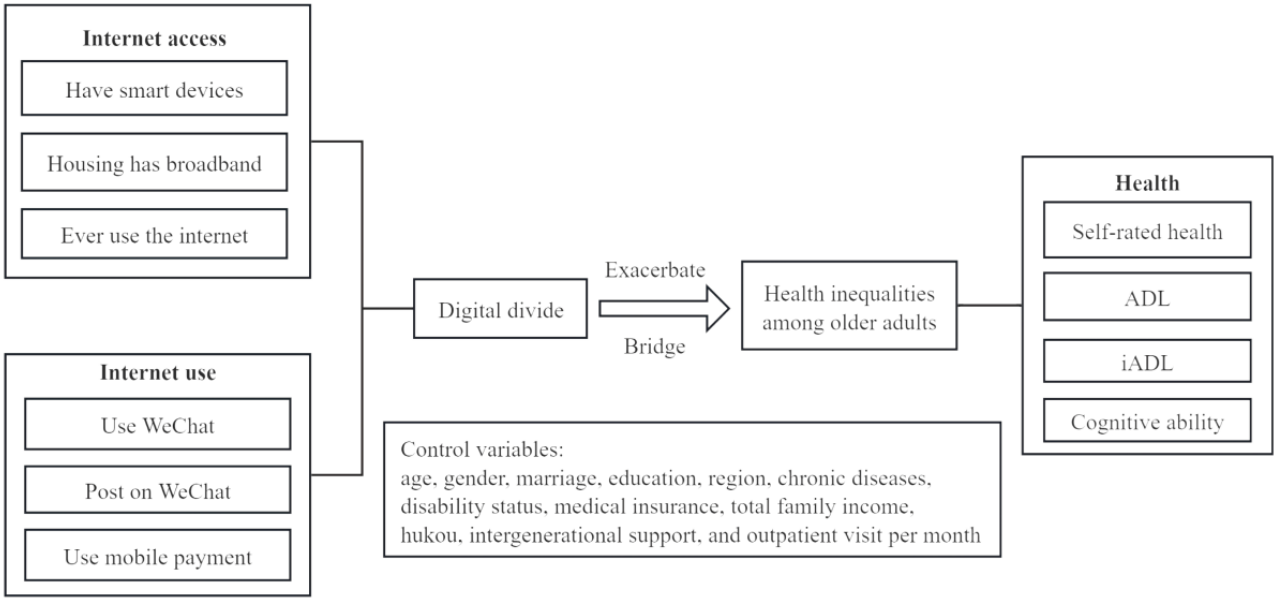

Methods

Data Sources

This study conducted a secondary analysis using data from CHARLS, a nationally representative longitudinal survey conducted in China by Peking University. CHARLS aims to understand the conditions of adults aged ≥45 years, along with their spouses. CHARLS currently possesses 5 waves of data spanning from 2011 to 2020. For this study, data from 2018 and 2020 were used. According to the World Health Organization, older adults are defined as those aged ≥60 years. However, considering that individuals aged 50 to 60 years will soon transition into older adulthood, it is important to study the impact of digital technology on their intentions and behavior; therefore, the age range for this study was ≥50 years. CHARLS 2020 included 19,367 participants; after excluding 2074 (10.71%) participants aged <50 years and 5120 (26.44%) with missing key variable data, 12,173 (62.85%) participants were retained. CHARLS 2018 included 19,827 participants; after excluding 2149 (10.84%) participants aged <50 years and 10,034 (50.61%) with missing key variable data, 7644 (38.55%) participants were retained.

XSL•FO
RenderX

Ethical Considerations

Prior the survey, all participants provided written informed consent, and the survey protocols received approval from the Peking University Ethics Review Board (IRB00001052-11015; Figure 2) [46]. The survey was also anonymous, and answers were protected by privacy law.

Figure 2. Flowchart of sample selection. CHARLS: China Health and Retirement Longitudinal Study.

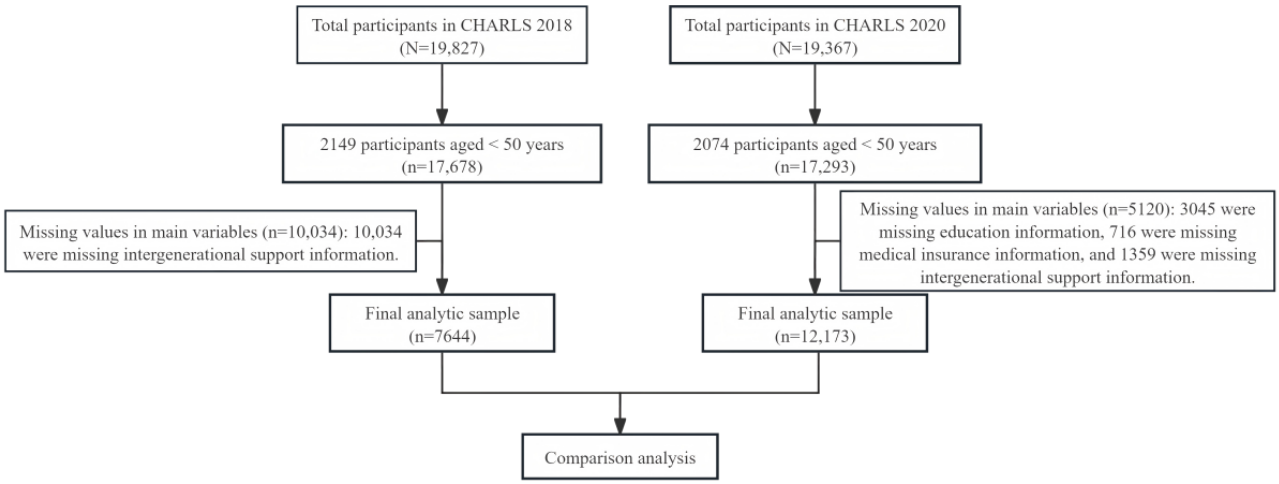

Variables and Measurements

Dependent Variable

This study used subjective and objective health indicators. Subjective indicators were measured through self-rated health, while objective indicators were divided into physical and psychological dimensions. Physical health was assessed based on the ability to perform ADL and iADL [17], and mental health was evaluated using cognitive function measured by the MMSE. Self-rated health consisted of five levels: (1) very unhealthy, (2) moderately unhealthy, (3) average, (4) moderately healthy, and (5) very healthy.

The ADL scale comprises 6 items: dressing, bathing or showering, self-feeding, getting into or out of bed, using the toilet (including getting up and down), and controlling urination and defecation (including catheter or pouch use). Responses were categorized into 4 levels: no difficulty, difficulty but able to perform, difficulty needing assistance, and unable to perform. Participants indicating *no difficulty* were scored as 1, while those indicating any level of difficulty were scored from 2 to 4 [47,48]. The total score ranges from 6 to 24 points.

The iADL scale includes 6 items: cooking, housework, shopping, managing money, making phone calls, and taking medication. Responses were assessed on a 4-point scale: 1=no difficulty, 2=difficulty but manageable, 3=difficulty needing assistance, and 4=inability to perform. Scores range from 6 to 24 points, with 6 indicating intact daily living skills and a higher score indicating greater impairment [47,48].

Cognitive function was evaluated using the MMSE [49], a concise assessment of adult cognitive status. The MMSE is used for screening cognitive impairment, gauging its severity, monitoring changes over time, and assessing treatment response. It consists of 30 items covering 7 cognitive domains: temporal orientation, spatial orientation, registration, attention and computation, recall, language, and praxis. Scores range from 0 to 30, with higher scores indicating better cognitive function.

According to the relative deprivation theory, older adults with poorer health status experience a greater sense of relative deprivation due to accumulating health disadvantages, leading to increased health inequality. We use the Kakwani relative deprivation index, which combines the results of self-rated health status, ADLs, iADLs, and cognitive function, to measure the degree of health inequality. The Kakwani index satisfies the problems of dimensionlessness, regularity, and transfer invariance. In addition, the Kakwani index can overcome the shortcomings of the Gini coefficient, which does not satisfy the additivity and decomposability. Therefore, this paper uses the Kakwani relative deprivation index to calculate the degree of health inequality of older adults. The measurement method is as follows: *Y* represents the sample size of older adults group of *n*, and arranged in order of the degree of health from smallest to the largest, the health distribution function for this group of older adults is as follows:



In equation 1, RD (y_j, y_i) denotes the relative deprivation index of health of the *j*-th older adult to the *i*-th older adult, and the relative deprivation index of health of the *i*-th older adult *n* is obtained by summing *j* and dividing it by the mean of the health level of all older adults in the group; hence, the degree of health inequality is as follows:



In equation 2, μ_Y represents the mean health level of all older adults in Y,  represents the mean health level of samples in Y with a health level of more than y_i, and  is the number of samples in Y with a health level of more than y_i as a proportion of the total sample.

XSL•FO
RenderX

Independent Variables

The digital divide is generally considered to have 3 dimensions: access divide, use divide, and outcome divide. This study focused on the first 2 dimensions. Three indicators were used to measure internet access level: the number of smart devices owned, broadband access at home, and the number of digital devices owned. The indicators were converted into a binary variable, where values 1 to 2 were assigned as 1. If any of the 3 indicators for internet access were 1, the individual was considered to have internet access.

To measure internet use level, 3 indicators were chosen: use of WeChat perform for socializing, posting on WeChat Moments, and using mobile payments [27]. WeChat, owned by the Chinese tech giant Tencent, is comparable to other social networking sites (SNSs) in western countries, such as Facebook and Twitter. The results of each question were denoted as 0 or 1, with higher values representing greater internet use capacity (ranging from 0 to 3).

Control Variables

Control variables were organized based on demographic characteristics: sex [50] (male=1 and female=0), age [51] (continuous variable), marital status [52] (married=1 and unmarried=0, where “unmarried” includes separation, divorce, widowhood, and never married), education [53] (no formal education=0, elementary school=1, and junior high school and above=2), chronic diseases [54] (yes=1 and no=0), disability status [55] (yes=1 and no=0), medical insurance [56] (urban employee medical insurance=1, urban and rural resident medical insurance=2, and other medical insurance=3), total family income [57] (log-transformed after bilateral trimming), *hukou* [58] (household registration; urban=0 and rural=1), intergenerational support [59] (determined by frequency of interactions with children; yes=1 and no=0), and outpatient visits [4] per month (range 0-30). Region of residence [60] (east, central, west, and northwest) was represented by 3 dummy variables.

Statistical Methods

Overview

Statistical analyses were conducted using Stata (version 15.0; StataCorp). Statistical significance was defined as 2-tailed *P* values <.05. Continuous data were presented as means and SDs, while categorical data were expressed as frequencies and percentages. The relationship between digital access (use) and health inequalities was examined using univariate analysis. Categorical variables were assessed using the chi-square test, continuous variables following a normal distribution were analyzed using the 2-tailed *t* test, and continuous variables not conforming to a normal distribution were evaluated using the Wilcoxon rank sum test.

Multiple Linear Regression

Multiple linear regression models were used to test the association between multiple dimensions of the digital divide and health inequalities. Subgroup analyses were conducted based on sex, age, *hukou* (household registration), educational level, and region. In addition to fully adjusted models, moderated effect analyses were conducted using interaction terms of internet access and use with other covariates to determine if the association between the digital divide and health inequality was moderated by these factors.

Propensity Score Matching Method

Sensitivity analyses were performed using propensity score matching (PSM). To analyze the impact of internet access and use on health inequality among older adults, the sample was divided into an experimental group (those who accessed or used the internet) and a control group (those who did not access or use the internet). Internet access (binary categorical variable) and internet use (converted into a binary variable using the median) were core variables for matching, and the nearest neighbor matching (1:1) principle was used to calculate the propensity score.

To further analyze the effects and mechanisms of internet access and use on health inequalities among older adults, the following model was constructed:



Where Y*i* denotes the degree of health inequality among older adults, D_i denotes the access or use of the Internet among older adults, X_i denotes other influences on the degree of health inequality among older adults, and ε is a random perturbation term. On the basis of the PSM model, internet access and internet use is set as a binary variable, and $D_i = \{0,1\}$ denotes whether the *i*-th older adult has access or use of the internet or not, that is, $D_i=1$ is access or use of the internet, and $D_i=0$ is no access or no use of the internet. There are 2 different scenarios of health inequality among older adults, one is the level of health inequality among older adults when accessing or using the internet (y_{1i}) and the other is the level of health inequality among older adults when not accessing or not using the internet (y_{0i}). The average treatment effect for the impact of the internet on health inequalities was calculated as:



In equation 5, average treatment effect denotes the difference between the level of health inequality among older adults with access and use of the internet $E(y_{1i}|Di=1)$ and the level of health inequality among older adults without access and use of the internet $E(y_{0i}|Di=1)$. Due to the unavailability of $E(y_{0i}|Di=1)$, $E(y_{0i}|Di=0)$ for the control group was used in the model as a proxy for the level of health inequality among older adults who use the internet.

Results

Characteristics of Study Participants

Table 1 presents the descriptive statistics of the sample, comprising 7644 observations from 2018 and 12,173 observations from 2020. In 2018, in total, 48.59% (3714/7644) of the participants were male, while 51.41% (3930/7644) were female. The average age of older adults was 65.11 (SD 10.02) years. In total, 73.31% (5604/7644) of the participants were

XSL•FO
RenderX

married at the time of the survey. Given that the average age of participants was approximately 65 years and education levels in China were generally low in the 1960s, this study classified participants with an education level of junior high school or above as “better educated.” The descriptive results indicate that only 30.65% (2343/7644) of the participants met this criterion. Most participants were rural residents (6449/7644, 84.37%). More than half of the respondents (4095/7644, 53.57%) had chronic diseases, and 14.21% (1086/7644) of the respondents had disabilities. Furthermore, 62.17% (4752/7644) of the respondents reported strong intergenerational support, indicating a tendency to contact and communicate with their children. Geographically, the distribution of older adults was fairly even across the eastern, central, and western regions, with the western region having the highest percentage (2642/7644, 34.56%) and the northeastern region having the fewest (491/7644, 6.43%).

XSL•FO
RenderX

Table 1. Characteristics of the study participants.

Variable	Total (N=19,817)	2018 (n=7644)	2020 (n=12,173)
Broadband access, n (%)			
No	9845 (49.68)	4646 (60.78)	5199 (42.71)
Yes	9972 (50.32)	2998 (39.22)	6974 (57.29)
Number of smart devices owned, n (%)			
0	14,693 (74.14)	6841 (89.50)	7852 (64.50)
1	4570 (23.06)	651 (8.52)	3919 (32.20)
≥2	554 (2.80)	152 (1.98)	402 (3.30)
Mobile payment use, n (%)			
No	2230 (11.25)	387 (5.06)	1843 (15.14)
Yes	2894 (14.61)	416 (5.44)	2478 (20.36)
Unknown	14,693 (74.14)	6841 (89.50)	7852 (64.50)
WeChat use, n (%)			
No	512 (2.59)	54 (0.70)	458 (3.76)
Yes	4612 (23.27)	749 (9.80)	3863 (31.74)
Unknown	14,693 (74.14)	6841 (89.50)	7852 (64.50)
WeChat Moments use, n (%)			
No	1870 (9.43)	199 (2.60)	1671 (13.72)
Yes	2742 (13.84)	550 (7.20)	2192 (18.01)
Unknown	15,205 (76.73)	6895 (90.20)	8310 (68.27)
Age (y), mean (SD)	64.41 (9.54)	65.11 (10.02)	63.97 (9.20)
Sex , n (%)			
Female	10,201 (51.48)	3930 (51.41)	6271 (51.52)
Male	9616 (48.52)	3714 (48.59)	5902 (48.48)
Marriage, n (%)			
Married	15,465 (78.04)	5604 (73.31)	9861 (81.01)
Unmarried	4352 (21.96)	2040 (26.69)	2312 (18.99)
Education, n (%)			
No formal education	4810 (24.27)	1917 (25.08)	2893 (23.77)
Elementary school	8610 (43.45)	3384 (44.27)	5226 (42.93)
Junior high school and above	6397 (32.28)	2343 (30.65)	4054 (33.30)
Current address, n (%)			
Rural	15,905 (80.26)	6449 (84.37)	9456 (77.68)
Urban	3912 (19.74)	1195 (15.63)	2717 (22.32)
Medical insurance, n (%)			
Urban employee medical insurance	2414 (12.33)	759 (10.25)	1655 (13.60)
Urban and rural resident medical insurance	11,058 (56.49)	849 (11.47)	10,209 (83.87)
Other medical insurance	6104 (31.18)	5795 (78.28)	309 (2.53)
Chronic disease, n (%)			
No	11,361 (57.33)	4095 (53.57)	7266 (59.69)
Yes	8456 (42.67)	3549 (46.43)	4907 (40.31)
Disability, n (%)			
No	16,943 (85.50)	6558 (85.79)	10,385 (85.31)

XSL•FO
RenderX

Variable	Total (N=19,817)	2018 (n=7644)	2020 (n=12,173)
Yes	2874 (14.50)	1086 (14.21)	1788 (14.69)
Intergenerational support, n (%)			
Weak	9775 (49.33)	2892 (37.83)	6883 (56.54)
Strong	10.042 (50.67)	4752 (62.17)	5290 (43.46)
Region, n (%)			
Eastern region	5859 (29.57)	2089 (27.33)	3770 (30.96)
Central region	6105 (30.80)	2422 (31.68)	3683 (30.26)
Western region	6649 (33.55)	2642 (34.56)	4007 (32.92)
Northeast region	1204 (6.08)	491 (6.43)	713 (5.86)
Log(income), mean (SD)	0.72 (0.34)	0.71 (0.35)	0.72 (0.33)
Outpatient visit per month, mean (SD)	0.45 (1.43)	0.39 (1.37)	0.49 (1.47)
Health, n (%)			
Very good	1718 (8.67)	500 (6.54)	1218 (10.01)
Good	3072 (15.50)	1737 (22.72)	1335 (11.00)
Fair	8992 (45.38)	3431 (44.89)	5561 (45.68)
Poor	3107 (15.68)	874 (11.43)	2233 (18.34)
Very poor	1572 (7.93)	769 (10.06)	803 (6.60)
Unknown	1356 (6.84)	333 (4.36)	1023 (8.37)
ADL[a], mean (SD)	6.43 (3.44)	5.93 (4.79)	6.75 (2.15)
iADL, mean (SD)	7.88 (4.14)	8.82 (5.15)	7.29 (3.22)
Cognition, mean (SD)	20.76 (8.3)	22.90 (7.97)	19.53 (8.04)

[a]ADL: Activities Of Daily Living scale.

[b]iADL: Instrumental Activities of Daily Living scale.

In 2020, in total, 48.48% (5902/12,173) of the participants were male and 51.52% (6271/12,173) were female. The mean age of older adults was 63.97 (SD 9.20) years. More than 80% (9861/12,173) of the participants were married. For educational attainment, only 33.3% (4054/12,173) of the participants had a junior high school education or higher. Most older adults were rural residents (9456/12,173, 77.68%), and the primary health insurance was urban and rural residents' health insurance (10,209/12,173, 83.87%). More than half of the participants (7266/12,173, 59.69%) had chronic diseases, while 14.69% (1788/12,173) had disabilities. In addition, 43.46% (5290/12,173) of the participants reported strong intergenerational support. The regional distribution of older adults in 2020 was similar to that in 2018, with the Western region having the highest proportion (4007/12,173, 32.92%) and the Northeast having the lowest (713/12,173, 5.86%).

Regarding internet access and use among older adults, only 39.22% (2998/7644) of the households had broadband in 2018, whereas by 2020, more than half of the households (6974/12,173, 57.29%) had broadband. Smart device ownership also increased significantly, with only 8.52% (651/7644) of the older adults owning a smartphone in 2018 compared to 32.2% (3919/12,173) in 2020.

The health status of older adults was also analyzed descriptively. For subjective health evaluations, the largest proportion of participants in both 2018 and 2020 rated their health as "fair" (3431/7644, 44.89% and 5561/12,173, 45.68%, respectively). However, the proportion of participants rating their health as "very good" increased in 2020 compared to 2018, accounting for 10.01% (1218/12,173). In terms of objective health indicators, older adults had better ADL scores in 2018 compared to 2020, while iADL scores were worse in 2018 compared to 2020. Finally, older adults had higher cognitive ability scores in 2018, with an average score of 22.90 (SD 7.97), compared to 2020.

Results of Simple Regression

This section investigated the effects of internet access and use on health inequalities among older adults, using data from the CHARLS collected in 2018 and 2020. In CHARLS 2018, the findings indicated that internet access and use generally widened health inequalities among older adults. Specifically, internet access was found to have amplified disparities in subjective self-rated health, ADL, and cognitive ability, with effect sizes of 3.9%, 5.8%, and 7.5%, respectively, all of which were statistically significant. In contrast, internet access showed a significant narrowing effect on inequalities related to iADL, with an effect size of 5.1%. Similarly, internet use was associated with an increase in inequalities in subjective self-rated health and cognitive ability, with effect sizes of 7.5% and 7.6%, respectively, and these results were statistically significant.

XSL•FO
RenderX

However, internet use had a mitigating effect on iADL and ADL inequalities, with effect values of 5% and 0.8%, although these effects were not statistically significant (Tables 2 and 3).

Table 2. Regression results for health based on internet access level in 2018 (N=7644)[a,b].

Attribute and level	Self-rated health[c]		ADL[d]		iADL[e]		Cognition[f]	
	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value
Internet access level	0.039 (0.005)	.001	0.058 (0.008)	<.001	–0.051 (0.003)	<.001	0.075 (0.008)	<0.001
Male	0.088 (0.005)	<.001	0.194 (0.008)	<.001	0.095 (0.003)	<.001	–0.086 (0.008)	<.001
Age	–0.075 (0.0003)	<.001	–0.222 (0.0004)	<.001	–0.311 (0.0001)	<.001	–0.504 (0.001)	<.001
Rural	–0.033 (0.006)	.006	–0.007 (0.012)	.57	0.039 (0.003)	<.001	0.037 (0.010)	<.001
Secondary schools	–0.052 (0.006)	<.001	0.013 (0.009)	.32	0.142 (0.004)	<.001	–0.212 (0.010)	<.001
Junior high school and above	0.012 (0.007)	.49	0.095 (0.012)	<.001	0.130 (0.004)	<.001	–0.178 (0.012)	<.001
Married	–0.002 (0.006)	.87	–0.005 (0.009)	.63	0.047 (0.003)	<.001	–0.155 (0.009)	<.001
Urban and rural resident medical insurance	0.006 (0.007)	.60	0.002 (0.012)	.85	0.017 (0.004)	.10	–0.021 (0.011)	.03
Chronic	–0.174 (0.004)	<.001	–0.126 (0.008)	<.001	–0.035 (0.002)	.001	–0.014 (0.007)	.16
Disability	–0.096 (0.005)	<.001	–0.086 (0.010)	<.001	–0.074 (0.004)	<.001	0.029 (0.011)	.009
Outpatient visit	0.085 (0.002)	<.001	–0.071 (0.003)	<0.001	–0.039 (0.001)	.001	0.005 (0.003)	.643
Intergenerational support	–0.031 (0.005)	.006	0.020 (0.008)	.06	0.072 (0.003)	<.001	–0.051 (0.008)	<.001
Log(income)	0.015 (0.006)	.19	0.021 (0.011)	.05	0.052 (0.003)	<.001	–0.013 (0.010)	.20
Central region	–0.047 (0.006)	.001	–0.041 (0.010)	.002	0.086 (0.003)	<.001	–0.009 (0.009)	.45
Western region	–0.055 (0.006)	<.001	–0.040 (0.010)	.002	0.104 (0.003)	<.001	–0.011 (0.009)	39
Northeast region	0.006 (0.010)	.64	–0.039 (0.017)	.001	–0.020 (0.005)	.07	–0.004 (0.015)	.69

[a]R^2 for the above models: 0.0901, 0.1779, 0.2138, and 0.2482.

[b]Perform 2-tailed winsorization and logarithmic transformation on income.

[c]Self-rated health (reference: reporting negative self-reported health).

[d]ADL: activity of daily living. Reporting difficulty with ADL (reference: no).

[e]iADL: instrumental activity of daily living. Reporting difficulty with iADL (reference: no).

[f]Cognition (reference: reporting negative cognition).

XSL•FO
RenderX

Table 3. Regression results for health based on internet use level in 2018 (N=1187)[a].

Attribute and level	Self-rated health		ADL[b]		iADL[c]		Cognition	
	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value
Internet use level	0.075 (0.008)	.03	0.076 (0.016)	.03	–0.050 (0.004)	.16	–0.008 (0.010)	.77
Male	0.086 (0.015)	02	0.206 (0.028)	<.001	0.064 (0.006)	.06	–0.029 (0.015)	.20
Age	–0.116 (0.001)	.002	–0.194 (0.002)	<.001	–0.202 (0.001)	<.001	–0.772 (0.002)	<.001
Rural	–0.073 (0.016)	.045	0.010 (0.030)	.78	0.199 (0.007)	<.001	–0.012 (0.017)	.62
Secondary schools	0.023 (0.053)	.84	0.025 (0.091)	.80	0.248 (0.025)	.04	–0.214 (0.058)	.007
Junior high school and above	0.037 (0.053)	.74	0.162 (0.091)	.11	0.273 (0.025)	.03	–0.210* (0.058)	.01
Married	0.052 (0.021)	.11	–0.043 (0.042)	.20	0.006 (0.009)	.86	–0.050 (0.027)	.06
Urban and rural resident medical insurance	–0.023 (0.028)	.53	0.026 (0.047)	.41	0.013 (0.011)	.70	–0.001 (0.024)	.97
Chronic	–0.123 (0.015)	.001	–0.094 (0.029)	.008	–0.017 (0.006)	.60	–0.005 (0.015)	.82
Disability	–0.085 (0.024)	.007	–0.110 (0.043)	<.001	0.013 (0.011)	.72	0.012 (0.026)	.60
Outpatient visit	–0.114 (0.006)	.001	–0.080 (0.010)	.01	–0.081 (0.003)	.03	–0.040 (0.007)	.15
Intergenerational support	0.027 (0.019)	.42	0.005 (0.036)	.89	0.027 (0.007)	.36	0.026 (0.019)	.24
Log(income)	0.006 (0.024)	.88	0.047 (0.041)	.18	0.079 (0.009)	.02	–0.038 (0.022)	.11
Central region	0.031 (0.019)	.47	–0.059 (0.035)	.14	0.251 (0.007)	<.001	0.001 (0.018)	.97
Western region	–0.019 (0.020)	.65	–0.083 (0.037)	.04	0.217 (0.008)	<.001	–0.002 (0.021)	.96
Northeast region	0.016 (0.026)	.68	–0.040 (0.047)	.28	0.207 (0.011)	.48	0.028 (0.026)	.28

[a]R^2 for the above models: 0.0928, 0.1640, 0.1952, and 0.6065.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Interestingly, the findings for CHARLS 2020 revealed a contrasting trend. In 2020, internet access was associated with a reduction in health inequalities in subjective self-rated health, ADL, iADL, and cognitive ability, with effect values of 3.8%, 2.1%, 3.5%, and 7.5%, respectively, all of which were statistically significant except for ADL. Likewise, internet use exhibited a narrowing effect on health inequalities, with respective effect sizes for subjective self-rated health, ADL, iADL, and cognition being 4.8%, 0.3%, 1.7%, and 12.8%. Except for ADL and iADL, these results were statistically significant. Notably, internet access and use demonstrated the largest impact on cognitive inequalities, highlighting cognition as a critical area affected by the digital divide (Tables 4 and 5).

XSL•FO
RenderX

Table 4. Regression results for health based on internet access level in 2020 (N=12,173)[a].

Attribute and level	Self-rated health		ADL[b]		iADL[c]		Cognition	
	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value
Internet access level	–0.038 (0.004)	<.001	–0.021 (0.001)	.053	–0.035 (0.001)	<.001	–0.075 (0.005)	<.001
Male	–0.054 (0.003)	<.001	0.059 (0.001)	<.001	0.058 (0.001)	<.001	–0.005 (0.004)	.56
Age	–0.009 (0.0002)	.38	–0.082 (0.0001)	.001	–0.129 (0.0001)	<.001	0.155 (0.0003)	<.001
Rural	0.022 (0.004)	.02	0.010 (0.001)	.28	0.011 (0.001)	.24	0.031 (0.005)	<.001
Secondary schools	0.007 (0.005)	.59	–0.005 (0.001)	.65	0.042 (0.002)	<.001	–0.372 (0.007)	<.001
Junior high school and above	–0.026 (0.005)	.06	0.022 (0.001)	.08	0.077 (0.002)	<.001	–0.483 (0.007)	<.001
Married	0.016 (0.005)	.11	–0.022 (0.001)	.04	–0.022 (0.002)	.02	–0.057 (0.006)	<.001
Urban and rural resident medical insurance	0.038 (0.004)	<.001	0.053 (0.001)	<.001	0.042 (0.002)	<.001	0.041 (0.005)	<.001
Chronic	0.156 (0.003)	<.001	–0.044 (0.001)	<.001	–0.047 (0.001)	<.001	0.012 (0.004)	.13
Disability	0.271 (0.006)	<.001	–0.227 (0.001)	<.001	–0.337 (0.002)	<.001	0.092 (0.008)	<.001
Outpatient visit	0.155 (0.002)	<.001	–0.043 (0.0003)	<.001	–0.037 (0.0004)	<.001	–0.019 (0.001)	.02
Intergenerational support	–0.011 (0.003)	.24	–0.009 (0.001)	.31	0.005 (0.001)	.57	–0.024 (0.004)	.003
Log(income)	0.009 (0.005)	.32	0.036 (0.001)	<.001	0.050 (0.002)	<.001	–0.094 (0.007)	<.001
Central region	0.005 (0.004)	.59	0.102 (0.001)	<.001	0.110*** (0.001)	<.001	0.030 (0.005)	.002
Western region	0.013 (0.004)	.21	0.070 (0.001)	<.001	0.066 (0.001)	<.001	0.008 (0.005)	.43
Northeast region	0.024 (0.007)	.005	0.076 (0.002)	<.001	0.093 (0.002)	<.001	–0.003 (0.008)	.67

[a]R^2 for the above models: 0.1639, 0.0972, 0.1973, and 0.3058.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

XSL•FO
RenderX

Table 5. Regression results for health based on internet use level (N=4321) in 2020[a].

Attribute and level	Self-rated health		ADL[b]		iADL[c]		Cognition	
	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value	β coefficient (SE)	*P* value
Internet use level	–0.048 (0.003)	.003	–0.003 (0.001)	.85	–0.017 (0.001)	.27	–0.128 (0.003)	<.001
Male	–0.051 (0.005)	.001	0.081 (0.001)	<.001	0.068 (0.002)	<.001	0.025 (0.005)	.10
Age	0.010 (0.0004)	.53	–0.106 (0.0001)	<.001	–0.100 (0.0001)	<.001	0.044* (0.0004)	.01
Rural	0.015 (0.005)	.37	0.031 (0.001)	.06	0.029 (0.002)	.08	–0.011 (0.005)	.497
Secondary schools	–0.004 (0.012)	.90	–0.017 (0.002)	.58	0.038 (0.004)	.27	–0.373 (0.015)	<.001
Junior high school and above	–0.030 (0.012)	.42	0.014 (0.002)	.66	0.095 (0.004)	.008	–0.531 (0.015)	<.001
Married	–0.023 (0.008)	.15	–0.009 (0.002)	.55	–0.029 (0.003)	.06	–0.050 (0.009)	.005
Urban and rural resident medical insurance	0.051 (0.006)	.001	0.105 (0.001)	<.001	0.095 (0.002)	<.001	0.038 (0.006)	.02
Chronic	0.164 (0.005)	<.001	–0.029 (0.001)	.049	–0.033 (0.002)	.03	0.047 (0.005)	.002
Disability	0.247 (0.015)	<.001	–0.138 (0.003)	<.001	–0.225 (0.004)	<.001	0.058 (0.013)	.003
Outpatient visit	0.176 (0.002)	<.001	–0.038 (0.001)	.04	–0.053 (0.001)	.008	–0.029 (0.001)	.005
Intergenerational support	–0.023 (0.005)	.11	0.019 (0.001)	.20	0.007 (0.002)	.65	–0.022 (0.005)	.13
Log(income)	–0.018 (0.007)	.23	0.047 (0.002)	.001	0.056 (0.002)	<.001	–0.051 (0.008)	.003
Central region	0.018 (0.006)	.28	0.176 (0.001)	<.001	0.168 (0.002)	<.001	0.046 (0.006)	.008
Western region	0.039 (0.006)	.02	0.116 (0.001)	<.001	0.127 (0.002)	<.001	0.001 (0.006)	.95
Northeast region	0.023 (0.008)	.10	0.143 (0.002)	<.001	0.172 (0.003)	<.001	0.015 (0.009)	.37

[a]R^2 for the above models: 0.1752, 0.0941, 0.1336, and 0.1397.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Sensitivity Analysis: PSM

To assess the robustness of the results, PSM was performed to estimate the net effect of internet access and use on health inequalities among older adults. Before applying PSM, a balance test was conducted to ensure comparability between the treatment and control groups after matching, apart from the independent variables. The kernel density trend graphs for the 2 subsamples indicated improved convergence after matching across different submodels, with satisfactory results for sample support and equilibrium tests (Multimedia Appendix 1). In addition, standardized mean differences were used to measure the balance of covariates before and after matching. Table S1 in Multimedia Appendix 1 shows that the standardized mean differences for all key variables were reduced to <20% after matching, indicating adequate balance and supporting the robustness of the procedure.

The results of the PSM analysis were consistent with the regression analysis. In 2018, internet access significantly widened health inequalities, particularly in subjective self-rated health, ADL, and cognition, with effect sizes of 2%, 6.5%, and 2.2%, respectively. Internet use also widened disparities in subjective health and ADL, with effect sizes of 5.7% and 8.4%, respectively. However, in 2020, internet access and use showed a trend toward reducing health inequalities, with the greatest effect seen in narrowing cognitive disparities, at 4.6% and 5.2%, respectively. Overall, the consistency between the regression and PSM results affirms the robustness of the model (Tables 6 and 7).

XSL•FO
RenderX

Table 6. Sensitivity analysis using propensity score matching for the year 2018.

Attribute and levels	Internet access level				Internet use level			
	Treatment group	Control group	ATT[a] (*P* value)	SE	Treatment group	Control group	ATT (*P* value)	SE
Self-rated health	0.192	0.173	0.020 (.008)	0.008	0.241	0.184	0.057 (.001)	0.024
ADL[b]	0.454	0.389	0.065 (<.001)	0.013	0.556	0.472	0.084 (.03)	0.043
iADL[c]	0.243	0.248	–0.005 (.26)	0.004	0.260	0.261	–0.001 (.55)	0.010
Cognition	0.432	0.410	0.022 (.01)	0.013	0.495	0.507	–0.012 (.13)	0.038

[a]ATT: average treatment effect.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Table 7. Sensitivity analysis using propensity score matching for the year 2020.

Attribute and levels	Internet access level				Internet use level			
	Treatment group	Control group	ATT[a] (*P* value)	SE	Treatment group	Control group	ATT (*P* value)	SE
Self-rated health	0.166	0.185	–0.020 (<.001)	0.008	0.143	0.167	–0.025 (<.001)	0.009
ADL[b]	0.099	0.101	–0.00 (.68)	0.002	0.102	0.103	–0.001 (.47)	0.002
iADL[c]	0.159	0.162	–0.003 (.74)	0.003	0.167	0.173	–0.005 (.04)	0.003
Cognition	0.173	0.219	–0.046 (<.001)	0.011	0.092	0.144	–0.052 (<.001)	0.010

[a]ATT: average treatment effect.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Heterogeneity Analysis

To further explore the effect of internet access and use on health inequalities, a heterogeneity analysis was conducted using the 2020 CHARLS data, focusing on sex, age, education level, and urban and rural status. The subgroup analysis demonstrated that internet access had a stronger impact on reducing health inequalities among rural residents compared to urban residents. For rural populations, greater internet access significantly contributed to reducing health disparities (Table 8). Sex differences showed that internet use had a larger impact on reducing inequalities among males, with an effect value of 9.5% (*P*<.001). Regarding age, younger adults aged <60 years experienced greater benefits from internet access in reducing inequalities in self-assessed health and cognitive ability, with effect values of 5.6% (*P*<.01) and 14.8% (*P*<.001), respectively. Regarding education level, both internet access and use had a significant impact on reducing health inequalities among those with higher education, potentially due to their greater information literacy and ability to access health services online. Complete results are shown in Tables 8-15.

XSL•FO
RenderX

Table 8. Results for urban and rural subgroups at the internet access level[a].

Attribute and levels	Urban subgroups (n=2717)				Rural subgroups (n=9456)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet access level	–0.042 (.06)	–0.016 (.48)	–0.036 (.10)	–0.055 (.01)	–0.036 (.001)	–0.022 (.07)	–0.036 (.001)	–0.076 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.191	0.118	0.218	0.240	0.156	0.093	0.196	0.302

[a]The numbers in the table represent the standardized β coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Table 9. Results for urban and rural subgroups at the internet use level[a].

Attribute and levels	Urban subgroups (n=1524)				Rural subgroups (n=2797)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet use level	–0.030 (.29)	–0.043 (.11)	–0.019 (.48)	–0.175 (<.001)	–0.056 (.004)	0.015(.46)	–0.020 (.29)	–0.109 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.199	0.125	0.150	0.103	0.165	0.078	0.134	0.159

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Table 10. Results for sex subgroups at the internet access level[a].

Attribute and levels	Female (n=6271)				Male (n=5902)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet access level	–0.038 (.006)	–0.031 (.04)	–0.046 (.001)	–0.073 (<.001)	–0.036 (.01)	–0.011 (.47)	–0.023 (.09)	–0.079 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.155	0.094	0.189	0.345	0.166	0.097	0.199	0.222

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

XSL•FO
RenderX

Table 11. Results for sex subgroups at the internet use level[a].

Attribute and levels	Female (n=2019)				Male (n=2302)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet use level	0.002 (.94)	–0.026 (.29)	–0.030 (.22)	–0.135 (<.001)	–0.095 (<.001)	0.012 (.59)	–0.008 (.70)	–0.124 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.181	0.102	0.125	0.183	0.168	0.082	0.137	0.097

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Table 12. Results for age subgroups at the internet access level.

Attribute and levels	<60 years (n=4337)				≥60 years (n=7836)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet access level	–0.056 (.001)	–0.007 (.70)	0.003 (.88)	–0.148 (<.001)	–0.027 (.03)	–0.026 (.04)	–0.043 (<.001)	–0.050 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.178	0.043	0.098	0.201	0.158	0.111	0.213	0.295

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Table 13. Results for age subgroups at the internet use level[a].

Attribute and levels	<60 years (n=2649)				≥60 years (n=1672)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet use level	–0.047 (.03)	–0.019 (.34)	–0.049 (.01)	–0.115 (<.001)	–0.051 (.04)	0.025 (.32)	0.026 (.30)	–0.139 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.181	0.067	0.114	0.125	0.174	0.129	0.156	0.156

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

XSL•FO
RenderX

Table 14. Results for education subgroups at the internet access level[a].

Attribute and levels	No formal education (n=2893)				Elementary school (n=5226)				Junior high school and above (n=4054)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL, β coefficient (*P* value)	iADL, β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet access level	–0.004 (.83)	–0.040 (.04)	–0.049 (.006)	–0.074 (<.001)	–0.039 (.007)	0.002 (.92)	–0.005 (.73)	–0.101 (<.001)	–0.060 (.001)	–0.026 (.19)	–0.047 (.01)	–0.048 (.02)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.132	0.111	0.222	0.137	0.161	0.089	0.170	0.102	0.182	0.094	0.160	0.049

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Table 15. Results for education subgroups at the internet use level[a].

Attribute and levels	No formal education (n=279)				Elementary school (n=1517)				Junior high school and above (n=2525)			
	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)	Self-rated health, β coefficient (*P* value)	ADL[b], β coefficient (*P* value)	iADL[c], β coefficient (*P* value)	Cognition, β coefficient (*P* value)
Internet use level	0.045 (.40)	0.048 (.45)	0.012 (.83)	–0.166 (.008)	–0.022 (.38)	0.004 (.86)	–0.018 (0.47)	–0.148 (<.001)	–0.079 (<.001)	–0.012 (=.59)	–0.010 (=.64)	–0.110 (<.001)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.264	0.156	0.181	0.081	0.176	0.093	0.138	0.052	0.149	0.090	0.127	0.036

[a]The numbers in the table represent the standardized regression coefficients for each submodel.

[b]ADL: activity of daily living.

[c]iADL: instrumental activity of daily living.

Discussion

Principal Findings

This study investigates the impact of internet access and use on health disparities among older adults in China, using multiple linear regression analysis and validating the findings with PSM methods. The results indicate that, in 2018, internet access and use tended to widen health inequalities among older adults, whereas in 2020, the effects were found to mitigate these inequalities. Factors such as urban or rural residence, sex, age, and education level also influenced the extent of these effects.

First, dual inequalities were observed in both the digital divide and health disparities among older adults. Significant improvements in internet access and use were noted between 2018 and 2020. In 2020, broadband coverage among older adults' households reached 57.29% (6974/12,173), the proportion of older adults owning smart devices was 35.5% (4321/12,173), and 3.3% (402/12,173) owned multiple digital devices. In contrast, these figures were 39.22% (2998/7644), 10.5% (803/7644), and 1.98% (152/7644), respectively, in 2018. Regarding health indicators, the proportion of older adults rating their health as "very good," along with ADL, iADL, and cognitive ability scores, all changed between 2018 and 2020, showing mixed trends. There was a slight improvement in self-rated health and ADL scores, whereas iADL and cognitive ability scores declined.

Second, results showed that internet access and use had differing impacts on health inequalities among older adults in 2018 and 2020. Our analysis of the 2018 data revealed that internet access and use generally exacerbated health inequalities for older adults. Specifically, internet access worsened disparities in subjective self-rated health, ADL, and cognition. However, it improved health inequalities in iADL. Through the internet, people could more conveniently handle daily affairs, reducing their reliance on physical services or facilities to a certain extent

XSL•FO
RenderX

and contributing to the enhancement of iADL levels. Particularly during the COVID-19 pandemic, services such as online shopping and telemedicine enabled individuals to complete many daily activities at home, minimizing outdoor activities and the risk of infection. This trend may have, to some degree, ameliorated health inequalities related to iADLs. Furthermore, iADLs may be more sensitive to changes in older adults' functional abilities, and even at lower levels of internet access, it could sensitively detect improvements in health equity. Furthermore, while the use of the internet appeared to have mitigated the disparity in iADL performance and cognitive abilities, it significantly intensified the imbalance between ADL performance and subjective health assessments. These findings remained robust after conducting a PSM test. This is consistent with previous research that used 2018 data from the Norwegian Diabetes Association to demonstrate that the digital divide can maintain and create health inequalities [61], which is basically consistent with our research results.

It was interesting to note that the results from 2020 showed that internet access and use actually contributed to improving health inequalities among older adults. Specifically, internet access had a mitigating effect on disparities in self-rated health, ADL, iADL, and cognitive inequality. In addition, internet use also narrowed inequalities in subjective self-rated health, ADL, iADL, and cognition. These findings remained robust following a PSM test. A 2020 study focusing on Chinese older adults found that digital participation played a mediating role in reducing depressive risks by indirectly improving mental health through the promotion of healthy lifestyles [62]. During the use of the internet, interaction and communication could stimulate thinking and broaden horizons, potentially mitigating differences in cognitive abilities to a certain extent. Meanwhile, given the severe aging issue in China, with many older individuals living alone, their children and friends could provide more timely care through mobile phones, which could reduce feelings of loneliness and benefit mental health among older adults.

During the COVID-19 pandemic in 2020, social distancing and travel restrictions prompted older adults to rely more heavily on the internet for socializing, shopping, and telemedical consultations, with the scale of China's mobile health care users reaching 661 million. This trend led to an increase in internet use. Furthermore, the COVID-19 pandemic itself could impact health inequalities among older adults. During this period, older adults faced greater psychological pressure and loneliness, adverse to cognitive function, while psychological panic might also introduce bias in self-assessments of health status. The inability to undergo physical examinations in person due to lockdowns and other restrictions negatively impacted their physical health. Older adults with lower ADL scores might encounter greater challenges in maintaining their daily living activities. However, internet use also provided conveniences for older adults, offering easier access to services and resources necessary for daily life, such as online shopping and remote household services. Older adults could better manage their daily lives using online calendars, reminder functions, and similar tools. In addition, they could access health information and participate in online health courses to support their health management.

Third, the difference in results between 2018 and 2020 can be attributed to the different stages of internet access, which had different effects on health equity. During the initial stages, as observed in 2018, the rate of internet access was low, potentially leading to widened health disparities among older adults due to their internet access and use. This was primarily because there was a learning curve associated with adopting new technologies, which was particularly challenging for older adults to master.

As internet technology became more widespread and society began to pay more attention to older adults' internet use, as seen in 2020, the gap in health inequalities started to diminish. The internet has been evolving with simpler and more intuitive designs, making it easier for older adults to navigate. Society has also been offering more internet education and training resources to help older adults improve their digital skills. For instance, in 2019, the Chinese government issued relevant policies to promote healthy aging. Furthermore, the range of health information and services available on the web has been expanding, including telemedicine and online health management. These services have become more accessible to older adults, regardless of their socioeconomic status [63]. These factors have all been working together to reduce the disparity in internet use among older adults, ultimately leading to a more equitable health landscape for this demographic.

Fourth, this study examines the heterogeneity of internet access and use on health inequalities in older adults concerning urban and rural areas, sex, age, and education level. In the older adult population, the influence of internet access on mitigating health inequalities is more pronounced among rural residents, although it also has a significant effect on urban dwellers. This could be explained by the fact that older adults in rural areas, through internet connectivity, gain access to health resources that were previously unavailable to them, whereas older adults in urban areas, who have had such access before, may experience less significant changes. However, older adults in urban areas tend to benefit more from acquiring and using the internet, as it facilitates improved health care information and services as well as living conditions. Conversely, older adults in rural areas may encounter greater life pressures and health challenges, coupled with a scarcity of resources and skills necessary for effective internet use. Consequently, these 2 groups exhibit varying degrees of impact on health outcomes. In terms of urban and rural areas, internet access has a more substantial impact on reducing health inequality among rural residents, who have limited access to health information. Accessing health information through the internet is crucial for improving the health of older adults in rural areas. Regarding sex, the effects of internet access on health inequalities are similar for men and women, whereas the effects of internet use are more prominent in men, possibly due to their greater aptitude for learning about electronics compared to women. In terms of age, the impact of internet access on reducing self-rated health inequalities and cognitive ability is more significant among younger adults aged <60 years, while the effect of internet use on both is essentially the same. Older adults approaching retirement tend to have stronger cognitive abilities and find it easier to learn how to use digital devices and online skills, resulting in a more significant

XSL•FO
RenderX

impact. Regarding education level, internet access has a greater impact on self-rated health among the highly educated population. Subjective self-assessment of health is an individual's personal evaluation, which is more influenced by the health knowledge they acquire. The application of science and technology promotes the popularization of health knowledge [64]. Highly educated individuals generally pay more attention to their health conditions and possess better information literacy [65,66]. Consequently, their subjective self-evaluation of health becomes more distinct from that of the unconnected population.

Finally, although the initial impact of internet access and use on health inequalities in older adults tends to increase before decreasing, this observation does not imply that we can neglect inequalities during the early stages. Instead, we should focus more on addressing the challenges and barriers that older adults face when initially using the internet. Through policy guidance, social support, and ongoing technological innovation, we can promote more equitable access for older adults to the health and well-being benefits that the internet provides. On the basis of this understanding, the following policy recommendations are proposed:

- Expand internet coverage to enable older adults to benefit more from digital resources.
- Promote active aging by enhancing older adults' information literacy, including training on digital skills through community programs and volunteerism.
- Strengthen internet coverage in rural areas to improve access to health information.
- Encourage collaboration between different sectors to address multiple social inequalities, such as interdepartmental cooperation to provide financial and technical support for the care of older adults.

There are several limitations to this paper. First, this study only used cross-sectional data from 2018 and 2020 to analyze the relationship between digital access and use and health outcomes among older adults, which limits our ability to establish causality and introduces the potential for reverse causality bias. Further longitudinal analysis is needed to assess the long-term effects and clarify the temporal sequence of these associations. Second, while this research attributes the difference in the impact of the internet on health inequality between 2018 and 2020 to varying levels of internet access, the specific impact mechanism and regulatory effects require additional verification in subsequent steps. Finally, although we conducted extensive stratified analyses across regions, education levels, urban and rural status, sex, and age, and included a wide range of control variables, we did not use survey weights. This choice was made to focus on associations rather than population-level inference, but it may affect the representativeness of our findings.

Conclusions

This study examined the relationship between internet access and use on health inequalities among older adults, finding that their impact initially widened and then narrowed these inequalities. As internet technology became more widespread and advanced, increased social attention and support for older adults' digital engagement contributed to this reduction. The internet is a powerful tool for improving health equity, but older adults often face barriers such as age and educational level, which make access to digital health benefits more challenging. Particularly, older adults in rural areas and with lower education levels have less access to information, making internet access even more crucial for improving health outcomes in these populations. Governments and society can promote equal, safe, and effective access to the health benefits of the internet for older adults through policy, social support, and technological innovation.

Acknowledgments

This study was funded by the General Project of Ministry of Education Foundation on Humanities and Social Sciences, China (22YJAZH082), and Youth Program of the National Natural Science Foundation of China (72304012).

Data Availability

The datasets generated and analyzed during this study are available in the China Health and Retirement Longitudinal Study data repository [67].

Authors' Contributions

CM and MW contributed to the study's conception and design. MW handled data acquisition, organization, and analysis. YX and MW drafted the initial manuscript, which underwent revisions by all authors. CM oversaw the study's overall management. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Common support and balance test.
[DOCX File , 335 KB - jmir_v27i1e62645_app1.docx]

XSL•FO
RenderX

References

1. Kaikeaw S, Punpuing S, Chamchan C, Prasartkul P. Socioeconomic inequalities in health outcomes among Thai older population in the era of universal health coverage: trends and decomposition analysis. Int J Equity Health 2023 Aug 02;22(1):144 [FREE Full text] [doi: 10.1186/s12939-023-01952-0] [Medline: 37533003]
2. Shang XT, Wei ZH. Socio-economic inequalities in health among older adults in China. Public Health 2023 Jan;214:146-152. [doi: 10.1016/j.puhe.2022.11.013] [Medline: 36549024]
3. Steele J, Shen J, Tsakos G, Fuller E, Morris S, Watt R, et al. The interplay between socioeconomic inequalities and clinical oral health. J Dent Res 2015 Jan;94(1):19-26. [doi: 10.1177/0022034514553978] [Medline: 25344336]
4. Wu L, Liu Q, Fu R, Ma J. Unmet healthcare needs, health outcomes, and health inequalities among older people in China. Front Public Health 2023 Jun 15;11:1082517 [FREE Full text] [doi: 10.3389/fpubh.2023.1082517] [Medline: 37397766]
5. Rohrbacher M, Hasselhorn HM. The contribution of work and health-related lifestyle to educational inequalities in physical health among older workers in Germany. A causal mediation analysis with data from the lidA cohort study. PLoS One 2023 Aug 9;18(8):e0285319 [FREE Full text] [doi: 10.1371/journal.pone.0285319] [Medline: 37556415]
6. Jiang CY, Han K, Yang F, Yin SY, Zhang L, Liang BY, et al. Global, regional, and national prevalence of hearing loss from 1990 to 2019: a trend and health inequality analyses based on the Global Burden of Disease Study 2019. Ageing Res Rev 2023 Dec;92:102124 [FREE Full text] [doi: 10.1016/j.arr.2023.102124] [Medline: 37972859]
7. The 51st statistical report on internet development in China. China Internet Network Information Center. 2020 Apr 27. URL: https://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm [accessed 2024-04-29]
8. Understanding the digital divide. Organisation for Economic Co-operation and Development. 2001. URL: https://www.oecd.org/en/publications/understanding-the-digital-divide_236405667766.html
9. Song Y, Qian C, Pickard S. Age-related digital divide during the COVID-19 pandemic in China. Int J Environ Res Public Health 2021 Oct 27;18(21):11285 [FREE Full text] [doi: 10.3390/ijerph182111285] [Medline: 34769801]
10. Thilak S, Brown P, Whitehouse T, Gautam N, Lawrence E, Ahmed Z, et al. Diagnosis and management of subarachnoid haemorrhage. Nat Commun 2024 Feb 29;15(1):1850 [FREE Full text] [doi: 10.1038/s41467-024-46015-2] [Medline: 38424037]
11. Gulumbe BH, Yusuf ZM, Hashim AM. Harnessing artificial intelligence in the post-COVID-19 era: a global health imperative. Trop Doct 2023 Oct 20;53(4):414-415 [FREE Full text] [doi: 10.1177/00494755231181155] [Medline: 37340738]
12. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
13. Cheshmehzangi A, Zou T, Su Z. The digital divide impacts on mental health during the COVID-19 pandemic. Brain Behav Immun 2022 Mar;101:211-213 [FREE Full text] [doi: 10.1016/j.bbi.2022.01.009] [Medline: 35041937]
14. Sharpe M. Time to transcend "physical" and "mental" health. BMJ 2020 Feb 24;368:m609. [doi: 10.1136/bmj.m609] [Medline: 32094143]
15. Silva-Peñaherrera M, Lopez-Ruiz M, Merino-Salazar P, Gómez-García AR, Benavides FG. Health inequity in workers of Latin America and the Caribbean. Int J Equity Health 2020 Jul 01;19(1):109 [FREE Full text] [doi: 10.1186/s12939-020-01228-x] [Medline: 32611402]
16. Solé-Auró A, Alcañiz M. Educational attainment, gender and health inequalities among older adults in Catalonia (Spain). Int J Equity Health 2016 Aug 04;15(1):126 [FREE Full text] [doi: 10.1186/s12939-016-0414-9] [Medline: 27491677]
17. Kino S, Aida J, Kondo K, Kawachi I. Do disasters exacerbate socioeconomic inequalities in health among older people? Int J Disaster Risk Reduct 2023 Nov;98:104071. [doi: 10.1016/j.ijdrr.2023.104071] [Medline: 37982017]
18. Deaton A. Health, inequality, and economic development. J Econ Lit 2003 Mar;41(1):113-158. [doi: 10.1257/.41.1.113]
19. Nie W, Hu M, Ye X. Internet use and rural-urban mental health inequalities: evidence from China. Front Public Health 2023 Jan 26;11:1107146 [FREE Full text] [doi: 10.3389/fpubh.2023.1107146] [Medline: 36778581]
20. Mu A, Deng Z, Wu X, Zhou L. Does digital technology reduce health disparity? Investigating difference of depression stemming from socioeconomic status among Chinese older adults. BMC Geriatr 2021 Apr 21;21(1):264 [FREE Full text] [doi: 10.1186/s12877-021-02175-0] [Medline: 33882865]
21. Kowal S, Ng CD, Schuldt R, Sheinson D, Jinnett K, Basu A. Estimating the US baseline distribution of health inequalities across race, ethnicity, and geography for equity-informative cost-effectiveness analysis. Value Health 2023 Oct;26(10):1485-1493 [FREE Full text] [doi: 10.1016/j.jval.2023.06.015] [Medline: 37414278]
22. Neelsen S, Stratmann T. Effects of prenatal and early life malnutrition: evidence from the Greek famine. J Health Econ 2011 May;30(3):479-488. [doi: 10.1016/j.jhealeco.2011.03.001] [Medline: 21546107]
23. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco product use among adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1736-1742 [FREE Full text] [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]
24. van Deursen AJ, van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media Soc 2019 Feb 07;21(2):354-375 [FREE Full text] [doi: 10.1177/1461444818797082] [Medline: 30886536]
25. Cruz-Jesus F, Vicente M, Bacao F, Oliveira T. The education-related digital divide: an analysis for the EU-28. Comput Human Behav 2016 Mar;56:72-82. [doi: 10.1016/j.chb.2015.11.027]

XSL•FO
RenderX

26. Guo C, Wan B. The digital divide in online learning in China during the COVID-19 pandemic. Technol Soc 2022 Nov;71:102122 [FREE Full text] [doi: 10.1016/j.techsoc.2022.102122] [Medline: 36118102]
27. Hargittai E. Second-level digital divide: differences in people's online skills. First Monday 2002;7(4):10. [doi: 10.5210/fm.v7i4.942]
28. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
29. van Deursen AJ, Helsper EJ, Eynon R. Development and validation of the Internet Skills Scale (ISS). Inf Commun Soc 2015 Aug 25;19(6):804-823. [doi: 10.1080/1369118X.2015.1078834]
30. Deursen AJ, van Dijk JA. Measuring internet skills. Int J Hum Comput Interact 2010 Sep 17;26(10):891-916. [doi: 10.1080/10447318.2010.496338]
31. Wilson KR, Wallin JS, Reiser C. Social stratification and the digital divide. Soc Sci Comput Rev 2003 May 01;21(2):133-143. [doi: 10.1177/0894439303021002001]
32. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
33. Rains SA. Health at high speed: broadband internet access, health communication, and the digital divide. Commun Res 2008 Apr 14;35(3):283-297. [doi: 10.1177/0093650208315958]
34. Daniels B, McGinnis C, Topaz LS, Greenwald P, Turchioe MR, Creber RM, et al. Bridging the digital health divide-patient experiences with mobile integrated health and facilitated telehealth by community-level indicators of health disparity. J Am Med Inform Assoc 2024 Apr 03;31(4):875-883. [doi: 10.1093/jamia/ocae007] [Medline: 38269583]
35. Irfan B, Yaqoob A. The digital lifeline: telemedicine and artificial intelligence synergy as a catalyst for healthcare equity in Pakistan. Cureus 2024 Feb;16(2):e54017 [FREE Full text] [doi: 10.7759/cureus.54017] [Medline: 38476785]
36. Katzow MW, Steinway C, Zuzarte A, Chen J, Fishbein J, Jan S. Sociodemographic disparities in ambulatory pediatric telemedicine utilization during COVID-19. Telemed J E Health 2024 Jan 09;30(1):57-66. [doi: 10.1089/tmj.2023.0005] [Medline: 37579076]
37. Hart JT. The inverse care law. Lancet 1971 Feb;297(7696):405-412. [doi: 10.1016/S0140-6736(71)92410-X]
38. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Jan 29;28(6):995-1019 [FREE Full text] [doi: 10.1111/isj.12179]
39. van Deursen AJ, Helsper EJ. The third-level digital divide: who benefits most from being online? In: Robinson L, Cotten SR, Schulz J, Hale TM, Williams A, editors. Communication and Information Technologies Annual, Volume 10. London, UK: Emerald Publishing Limited; 2015.
40. Heponiemi T, Jormanainen V, Leemann L, Manderbacka K, Aalto AM, Hyppönen H. Digital divide in perceived benefits of online health care and social welfare services: national cross-sectional survey study. J Med Internet Res 2020 Jul 07;22(7):e17616 [FREE Full text] [doi: 10.2196/17616] [Medline: 32673218]
41. Jella TK, Cwalina TB, Sachdev RR, Otteson T, Fowler N. Sociodemographic disparities in the use of health information technology by a national sample of head and neck cancer patients. Am J Otolaryngol 2022 Mar;43(2):103308. [doi: 10.1016/j.amjoto.2021.103308] [Medline: 34999347]
42. Rahman S, Amit S, Kafy AA. Gender disparity in telehealth usage in Bangladesh during COVID-19. SSM Ment Health 2022 Dec;2:100054 [FREE Full text] [doi: 10.1016/j.ssmmh.2021.100054] [Medline: 35036972]
43. Wegermann K, Wilder JM, Parish A, Niedzwiecki D, Gellad ZF, Muir AJ, et al. Racial and socioeconomic disparities in utilization of telehealth in patients with liver disease during COVID-19. Dig Dis Sci 2022 Jan;67(1):93-99 [FREE Full text] [doi: 10.1007/s10620-021-06842-5] [Medline: 33507442]
44. Pendl D, Maitz KM, Gasteiger-Klicpera B. Examining the relationship between health literacy and individual and sociodemographic factors in secondary school students. Z Gesundh Wiss 2023 Feb 02;32(3):1-12 [FREE Full text] [doi: 10.1007/s10389-023-01836-1] [Medline: 36747506]
45. Cheng C, Gearon E, Hawkins M, McPhee C, Hanna L, Batterham R, et al. Digital health literacy as a predictor of awareness, engagement, and use of a national web-based personal health record: population-based survey study. J Med Internet Res 2022 Sep 16;24(9):e35772 [FREE Full text] [doi: 10.2196/35772] [Medline: 36112404]
46. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb 12;43(1):61-68 [FREE Full text] [doi: 10.1093/ije/dys203] [Medline: 23243115]
47. Ruan Z, Qian ZM, Guo Y, Zhou J, Yang Y, Acharya BK, et al. Ambient fine particulate matter and ozone higher than certain thresholds associated with myopia in the elderly aged 50 years and above. Environ Res 2019 Oct;177:108581. [doi: 10.1016/j.envres.2019.108581] [Medline: 31323395]
48. Hong YA, Zhou Z, Fang Y, Shi L. The digital divide and health disparities in China: evidence from a national survey and policy implications. J Med Internet Res 2017 Sep 11;19(9):e317 [FREE Full text] [doi: 10.2196/jmir.7786] [Medline: 28893724]
49. Evandrou M, Falkingham J, Feng Z, Vlachantoni A. Individual and province inequalities in health among older people in China: evidence and policy implications. Health Place 2014 Nov;30:134-144 [FREE Full text] [doi: 10.1016/j.healthplace.2014.08.009] [Medline: 25262491]

XSL•FO
RenderX

50. The Lancet Public Health. Addressing health inequalities in gender diverse people. Lancet Public Health 2024 Feb;9(2):e68 [FREE Full text] [doi: 10.1016/S2468-2667(24)00006-9] [Medline: 38307680]
51. Esteve-Matalí L, Llorens-Serrano C, Alonso J, Vilagut G, Moncada S, Navarro-Giné A. Mental health inequalities in times of crisis: evolution between 2005 and 2021 among the Spanish salaried population. J Epidemiol Community Health 2022 Nov 07;77(1):38-43. [doi: 10.1136/jech-2022-219523] [Medline: 36344271]
52. Thoits PA. Stress and health: major findings and policy implications. J Health Soc Behav 2010;51 Suppl:S41-S53. [doi: 10.1177/0022146510383499] [Medline: 20943582]
53. Hübelová D, Caha J, Janošíková L, Kozumplíková A. A holistic model of health inequalities for health policy and state administration: a case study in the regions of the Czech Republic. Int J Equity Health 2023 Sep 05;22(1):183 [FREE Full text] [doi: 10.1186/s12939-023-01996-2] [Medline: 37670373]
54. Wang YJ, Chen XP, Chen WJ, Zhang ZL, Zhou YP, Jia Z. Ethnicity and health inequalities: an empirical study based on the 2010 China Survey of Social Change (CSSC) in Western China. BMC Public Health 2020 May 07;20(1):637 [FREE Full text] [doi: 10.1186/s12889-020-08579-8] [Medline: 32380963]
55. Falkingham JC, Chepngeno-Langat G, Kyobutungi C, Ezeh A, Evandrou M. Does socioeconomic inequality in health persist among older people living in resource-poor urban slums? J Urban Health 2011 Jun 23;88 Suppl 2(Suppl 2):S381-S400 [FREE Full text] [doi: 10.1007/s11524-011-9559-4] [Medline: 21431465]
56. Xian W, Xu X, Li J, Sun J, Fu H, Wu S, et al. Health care inequality under different medical insurance schemes in a socioeconomically underdeveloped region of China: a propensity score matching analysis. BMC Public Health 2019 Oct 25;19(1):1373 [FREE Full text] [doi: 10.1186/s12889-019-7761-6] [Medline: 31653250]
57. Oña A, Athanasios K, Tederko P, Escorpizo R, Arora M, Sturm C, et al. Unmet healthcare needs and health inequalities in people with spinal cord injury: a direct regression inequality decomposition. Int J Equity Health 2023 Mar 30;22(1):56 [FREE Full text] [doi: 10.1186/s12939-023-01848-z] [Medline: 36998015]
58. Yuan L, Yu B, Gao L, Du M, Lv Y, Liu X, et al. Decomposition analysis of health inequalities between the urban and rural oldest-old populations in China: evidence from a national survey. SSM Popul Health 2023 Mar;21:101325 [FREE Full text] [doi: 10.1016/j.ssmph.2022.101325] [Medline: 36618546]
59. Pei Y, Cong Z, Wu B. Education, adult children's education, and depressive symptoms among older adults in rural China. Soc Sci Med 2020 May;253:112966. [doi: 10.1016/j.socscimed.2020.112966] [Medline: 32247217]
60. Jiang W, Zhao Y, Wu X, Du Y, Zhou W. Health inequalities of global protein-energy malnutrition from 1990 to 2019 and forecast prevalence for 2044: data from the Global Burden of Disease Study 2019. Public Health 2023 Dec;225:102-109 [FREE Full text] [doi: 10.1016/j.puhe.2023.10.003] [Medline: 37924634]
61. Carotenuto A, Rea R, Traini E, Ricci G, Fasanaro AM, Amenta F. Cognitive assessment of patients with Alzheimer's disease by telemedicine: pilot study. JMIR Ment Health 2018 May 11;5(2):e31 [FREE Full text] [doi: 10.2196/mental.8097] [Medline: 29752254]
62. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen, Wangberg SC, et al. Inequalities in the use of eHealth between socioeconomic groups among patients with type 1 and type 2 diabetes: cross-sectional study. J Med Internet Res 2019 May 29;21(5):e13615 [FREE Full text] [doi: 10.2196/13615] [Medline: 31144669]
63. Tang H, Yuan D, Wang M, Yang H, Guo C. [Sequential mediating role of digital participation and health lifestyle in the relationship between socioeconomic status and depression of older adults]. Beijing Da Xue Xue Bao Yi Xue Ban 2024 Apr 18;56(2):230-238 [FREE Full text] [doi: 10.19723/j.issn.1671-167X.2024.02.005] [Medline: 38595238]
64. Żarnowski A, Jankowski M, Gujski M. Use of mobile apps and wearables to monitor diet, weight, and physical activity: a cross-sectional survey of adults in Poland. Med Sci Monit 2022 Sep 09;28:e937948 [FREE Full text] [doi: 10.12659/MSM.937948] [Medline: 36081328]
65. Dunn P, Hazzard E. Technology approaches to digital health literacy. Int J Cardiol 2019 Oct 15;293:294-296. [doi: 10.1016/j.ijcard.2019.06.039] [Medline: 31350037]
66. Chung MH, Chen LK, Peng LN, Chi MJ. Development and validation of the health literacy assessment tool for older people in Taiwan: potential impacts of cultural differences. Arch Gerontol Geriatr 2015 Sep;61(2):289-295. [doi: 10.1016/j.archger.2015.06.015] [Medline: 26153552]
67. Home page. The China Health and Retirement Longitudinal Study (CHARLS). URL: https://charls.pku.edu.cn/en/ [accessed 2024-04-29]

Abbreviations

ADL: activity of daily living
CHARLS: China Health and Retirement Longitudinal Study
iADL: instrumental activity of daily living
MMSE: Mini-Mental State Examination
PSM: propensity score matching

XSL•FO
RenderX

Edited by A Mavragani; submitted 29.05.24; peer-reviewed by M Chen, DJ Simon, KT Pritchard; comments to author 27.09.24; revised version received 17.10.24; accepted 17.11.24; published 15.01.25.

Please cite as:
Wu M, Xue Y, Ma C
The Association Between the Digital Divide and Health Inequalities Among Older Adults in China: Nationally Representative Cross-Sectional Survey
J Med Internet Res 2025;27:e62645
URL: https://www.jmir.org/2025/1/e62645
doi:10.2196/62645
PMID:

©Mengqiu Wu, Yongxi Xue, Chengyu Ma. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Latinx and White Adolescents' Preferences for Latinx-Targeted Celebrity and Noncelebrity Food Advertisements: Experimental Survey Study

Marie A Bragg[1,2*], PhD; Samina Lutfeali[3*], MPH; Daniela Godoy Gabler[1,2*], MPH; Diego A Quintana Licona[1,4*], BS; Jennifer L Harris[5], MBA, PhD

[1]Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States

[2]Food Environment and Policy Research Coalition, NYU Grossman School of Medicine, New York, NY, United States

[3]Marketing Department, Stanford Graduate School of Business, Stanford University, Stanford, CA, United States

[4]Mailman School of Public Health, Columbia University, New York, United States

[5]Rudd Center for Food Policy & Health, University of Connecticut, Hartford, CT, United States

[*]these authors contributed equally

Corresponding Author:
Marie A Bragg, PhD
Department of Population Health
NYU Grossman School of Medicine
180 Madison Avenue
New York, NY, 10016
United States
Phone: 1 6465012717
Email: marie.bragg@nyulangone.org

Abstract

Background: Exposure to food advertisements is a major driver of childhood obesity, and food companies disproportionately target Latinx youth with their least healthy products. This study assessed the effects of food and beverage advertisements featuring Latinx celebrities versus Latinx noncelebrities on Latinx and White adolescents.

Objective: This web-based within-subjects study aims to assess the effects of food and beverage advertisements featuring Latinx celebrities versus Latinx noncelebrities on Latinx and White adolescents' preferences for the advertisements and featured products.

Methods: Participants (N=903) were selected from a volunteer sample of adolescents, aged 13-17 years, who self-identified as Latinx or White, had daily internet access, and could read and write in English. They participated in a web-based Qualtrics study where each participant viewed 8 advertisements for novel foods and beverages, including 4 advertisements that featured Latinx celebrities and the same 4 advertisements that featured Latinx noncelebrities (matched on all other attributes), in addition to 2 neutral advertisements (featuring bland, nontargeted products and did not feature people). Primary outcomes were participants' ratings of 4 advertisements for food and beverage brands featuring a Latinx celebrity and the same 4 advertisements featuring a Latinx noncelebrity. Multilevel linear regression models compared the effects of celebrities and differences between Latinx and White participants on attitudes (advertisement likeability; positive affect; and brand perceptions) and behavioral intentions (consumption; social media engagement—"liking;" following; commenting; tagging a friend).

Results: Latinx (n=436; 48.3%) and White (n=467; 51.7%) participants rated advertisements featuring Latinx celebrities more positively than advertisements featuring noncelebrities on attitude measures except negative affect (*Ps*≤.002), whereas only negative affect differed between Latinx and White participants. Two of the 5 behavioral intention measures differed by celebrity advertisement status (*P*=.02*; P*<.001). Additionally, the interaction between celebrity and participant ethnicity was significant for 4 behavioral intentions; Latinx, but not White, participants reported higher willingness to consume the product (*P*<.001), follow brands (*P*<.001), and tag friends (*P*<.001). While White and Latinx adolescents both reported higher likelihoods of "liking" advertisements on social media endorsed by Latinx celebrities versus noncelebrities, the effect was significantly larger among Latinx adolescents (*P*<.01).

Conclusions: This study demonstrates the power of Latinx celebrities in appealing to both Latinx and White adolescents but may be particularly persuasive in shaping behavioral intentions among Latinx adolescents. These findings suggest an urgent need

XSL•FO
RenderX

to reduce celebrity endorsements in ethnically targeted advertisements that promote unhealthy food products to communities disproportionately affected by obesity and diabetes. The food industry limits food advertising to children ages 12 years and younger, but industry self-regulatory efforts and policies should expand to include adolescents and address disproportionate marketing of unhealthy food to Latinx youth and celebrity endorsements of unhealthy products.

(J Med Internet Res 2025;27:e53188) doi:10.2196/53188

KEYWORDS

Latinx; Hispanic; adolescents; marketing; celebrities; Spanish; advertisements; products; brands; food; unhealthy; beverages; diet; nutrition; consumers; intention; purchasing; attitudes; perceptions; preferences; youth

Introduction

Developing obesity during adolescence increases future risk of diabetes, cardiovascular disease, and diet-related cancers during adulthood [1-7]. Youth of color experience higher rates of obesity relative to White youth [8-10]. More than 37% of Latinx adolescents are overweight or obese relative to 26.5% of White adolescents [1,11], indicating an urgent need to reduce rates of obesity among Latinx youth.

Food marketing is a major driver of adolescent and childhood obesity [12-16]. Food companies spend more than US $1.8 billion on youth-directed marketing [17,18], which primarily promotes products that are energy-dense and nutrient-poor [19-22]. Such exposure is concerning given the large body of evidence showing that viewing food advertising increases children's preferences for promoted foods and brands, caloric intake, purchases, and requests for advertised products [2,16,23-26].

Adolescents and youth of color represent key consumer groups to food marketers not only because of their spending power [27] and trendsetting abilities [28,29] but also because of their unique responsiveness to advertising [19]. Compared to children and adults, adolescents are more impulsive and have lower inhibitory control [30]. They use brands to help elevate their feelings of self-worth [30], fit in with their desired peer group [31], and distinguish themselves from their caregivers [32]. Food and beverage brands aim to appeal to Latinx youth through targeted or multiethnic marketing, which refers to identifying a community that shares some common needs or characteristics that an organization—in this case, marketers—uses to appeal to that group [33]. While multiethnic advertising aims to reach consumers of diverse ethnicities, monoethnic advertising reaches members of a single ethnic group [33]. A study by Shao et al [34] found that socially distinctive consumers (eg, Latinx minorities) are more likely to prefer same-race monoethnic advertisements because this exposure promotes a sense of shared identity and belonging [34]. Janssen et al [35] explored advertising on Instagram and found that celebrity influencers with a high number of followers have more "likes" and more followers with positive attitudes toward the products they endorse; this results in greater purchase intention [35]. While targeted marketing is not inherently problematic, monoethnic advertising coupled with celebrity or influencer sponsorships can potentially target Latinx adolescents and Generation Z teens with health-harming products that may exacerbate health disparities.

Adolescents are exposed to food marketing on social media [36-38], but few studies have examined how adolescents respond to such marketing. One cross-sectional study showed that among a nationally representative sample of adolescents, more than 70% of them self-reported that they "liked," shared, or followed food and beverage brand accounts [39]. The same study also found that Black and less acculturated Latinx adolescents were more likely to self-report that they engaged with brand accounts than White adolescents [39]. Recent experimental studies have revealed that adolescents rate social media advertisements highly when those advertisements have high numbers of "likes" [36]. One randomized trial examining the impact of influencer marketing—the promotion of products by popular web-based celebrities—showed that influencer marketing of unhealthy foods on social media significantly increased children's consumption of unhealthy snacks [40,41].

Exposure to unhealthy food advertisements that feature celebrity influencers is concerning as it can capture adolescents' attention through its appealing features. Viewing these advertisements can increase positive affect and feelings of connection to the celebrity or influencer because they are often seen as trustworthy and are already well-liked by the adolescent [42-44]. These positive feelings toward the celebrity or influencer then elicit positive feelings (greater positive affect, liking, and desire for the product) toward the unhealthy food and beverage they are promoting [45]. Previous studies have assessed participants' responses to advertisements using 3 attitudinal ratings, including how much they like the advertisement, brand, and product, which allows researchers to compare such ratings across experimental conditions [46]. Such attitudinal ratings are measured through self-report questions that ask how much the participant likes the advertisement, how much they like the brand, and how much they like the product featured in the advertisement. Consequently, the likelihood of the adolescent engaging with the advertisement posts (eg, through "likes," comments, and shares) increases, along with the likelihood that they will subsequently purchase or consume unhealthy food [42,47-51]. Purchase intentions are measured via a self-report question and defined as a participant's willingness to purchase the product in the near future. Previous studies have used purchase intention questions as an outcome when objective measurements of purchasing behavior are not possible [52-57]. Previous research has measured and experimentally compared this engagement by participants with brands and advertisements on social media. Further, Social Norms Theory proposes that adolescents in particular, may be especially vulnerable to many of the racially targeted marketing practices that companies use on social media [58,59] because adolescents are highly sensitive

XSL•FO
RenderX

to their peers [60,61], popularity cues, and conformity pressures [62,63].

Research has shown that these feelings of connection may be even stronger when the race of the person featured in the advertisement matches the race of the adolescent [42,64]. A number of theoretical frameworks in social psychology may inform an understanding of this effect. According to the social identity theory, individuals derive a significant portion of their self-concept from their group memberships, leading to the formation of social identities [65]. Emphasizing the importance of social categorization, social comparison, and group identification in shaping attitudes and behaviors, social identity theory suggests that Latinx adolescents' ethnic identity may play a crucial role in shaping their self-concept. Racially targeted marketing that features Latinx celebrities can tap into this social identity, thereby fostering a sense of belonging and positively influencing attitudes toward the advertised product. Relatedly, Distinctiveness Theory may help explain the heightened appeal and influence of food advertisements featuring Latinx celebrities. Distinctiveness Theory explains that individuals define themselves based on rare traits in their environment [42,66]. Ethnicity in particular is an attribute related to distinctiveness and has been shown to hold greater significance in self-definition than other social categories [66]. Because Latinx adolescents are part of a distinctive minority group, distinctiveness theory suggests that they may be acutely attuned to and influenced by ethnically targeted marketing.

Despite the pervasiveness of targeted marketing, no experimental studies have examined the effects of Latinx-targeted food marketing on Latinx and White adolescents. This study assessed the effects of food and beverage advertisements featuring a Latinx celebrity on the same advertisements featuring a Latinx noncelebrity (ie, an individual who is not famous), including differential effects on Latinx and White adolescents. Examining the influence of celebrity-endorsed food advertisements on the preferences of Latinx and White youth will enhance the field's theoretical understanding of how ethnically targeted marketing uniquely affects Latinx youth, which can help inform practical implications on how public policies can address unhealthy food marketing that targets youth of color. By harnessing a rigorous 2×2 mixed factorial design, this study is the first to objectively compare celebrities' and noncelebrities' food endorsements on Latinx and White youth. Our research questions, therefore, are as follows:

- To what extent does exposure to Latinx celebrities' food advertisements increase adolescents' preferences for the advertisement and the featured product, compared to Latinx-noncelebrities' food advertisements?
- To what extent does exposure to Latinx people in food advertisements increase preferences for the advertisements among Latinx adolescents compared to White adolescents?

We predicted that (1) both Latinx and White adolescents would report higher preferences, including more positive attitudes (advertisement likeability; positive affect; and brand perceptions) and behavioral intentions (product consumption and social media engagement) when advertisements featured Latinx celebrities versus Latinx noncelebrities; (2) Latinx adolescents would report higher preferences for all advertisements compared to White adolescents; and (3) Latinx adolescents would show a stronger positive response to Latinx celebrity advertisements than White adolescents.

Methods

Study Design

This web-based study used a rigorous 2×2 mixed-factorial experimental design with 2 independent variables: the ethnicity of the participant (Latinx or White) and whether the person in the advertisement was a celebrity or noncelebrity. Latinx and White adolescent participants (aged 13-17 years) each viewed 8 advertisements for novel foods and beverages, including 4 advertisements that featured Latinx celebrities and the same 4 advertisements that featured Latinx noncelebrities (matched on all other attributes), in addition to 2 neutral advertisements. Novel food advertisements feature brands that participants are likely to be unfamiliar with, either created by researchers or brands from other countries. Using novel brands minimizes the likelihood that participants' ratings will be driven by previous brand preferences and experiences. For example, using a well-known brand such as Coca Cola would drown out our ability to examine the effects of Latinx celebrities' endorsements due to preexisting attitudes and preferences by consumers. After viewing each advertisement, participants provided attitude measures (for someone like me, advertisement likeability, positive and negative affect, and brand perceptions) and behavioral intentions (consumption, "liking," following, commenting, and tagging a friend), before viewing the next advertisement. These outcome measures have been used in other food marketing research studies as primary and secondary outcomes [39,41,52].

Participant Recruitment

In January 2019, we recruited participants through Dynata, a survey firm that maintains panels of thousands of adolescents and adults in the United States who are interested in completing surveys in exchange for incentives such as cash, prize drawings, or donations to charity. Dynata's recruitment involves a rigorous process that uses multiple steps to recruit and screen potential participants. First, randomly selected participants from members of Dynata's panels are combined with a pool of potential participants who are joining Dynata for the first time after responding to web-based recruitment materials. Those recruitment materials include banner advertisements on social media platforms and other commonly visited websites where participants might notice and respond to the flyers. This combined group of potential participants then receives an invitation to "take a survey," but are not provided with additional details about surveys that might introduce selection bias (eg, "take a survey about food" might attract people with an interest in food). Second, to be eligible for a study, potential participants complete quality control questions and are randomly matched with surveys for which they likely qualify. Finally, among the randomly selected participants who expressed an interest in and have been randomly matched with a survey, a subset is

XSL•FO
RenderX

randomized to complete a specific survey and then receives a link to the survey.

Pretest (N=50) and study participants (N=903) were adolescents, aged 13-17 years, who self-identified as Latinx, or White, had daily internet access, and could read and write in English. Of the 1509 eligible participants who responded to the survey, 992 participants (65.7%) completed all questions. An additional 89 participants did not answer 8 of 10 attention-check questions correctly and were excluded from the analyses. These attention-check questions asked, "What does this advertisement say?" with 3 multiple-choice answers. The final sample included 903 adolescents (436 (48.3%) Latinx adolescents and 467 (51.7%) White adolescents).

Ethical Considerations

The Institutional Review Board for the New York University School of Medicine approved the study (protocol #115-0087), in accordance with all applicable regulations. Parents voluntarily provided consent and adolescent participants provided assent after reading general information about the study, explaining the nature and possible consequences of the study, and the length of the survey. Study data were deidentified to protect participants' privacy and confidentiality. Participants were compensated by our recruitment firm, Dynata, with incentives such as cash, prize drawings, or donations to charity.

Materials

To develop the 8 food and beverage advertisements, we first created pairs of advertisements for 8 novel foods and beverages using Photoshop (Adobe). Each pair included 1 advertisement that featured a Latinx celebrity and a matched advertisement with a Latinx noncelebrity. Because brand familiarity for well-known products (eg, Coca Cola) heavily influences advertisement ratings and the difficulty of changing preexisting attitudes about known brands [67], we selected brands that were only available in countries outside the United States. To increase the likelihood that the advertisements were perceived as Latinx-targeted and that participants would recognize the celebrities, the advertisements included a slogan with at least 1 Spanish word or the name of the celebrity in the slogan or as a signature.

The 16 advertisements (8 pairs) for pretesting included 4 male and 4 female celebrity or noncelebrity pairs and 4 foods and 4 beverages. We pretested these advertisements among 20 Latinx and White adolescents to match the advertisement pairs on the attractiveness of the person in the advertisement, perceived target audience, and celebrity recognition.

Pretest participants first rated the attractiveness of the person in the advertisement, indicated the perceived race or ethnicity of the person, and identified whether the person in the advertisement was a celebrity. They also indicated how much they thought the advertisement was intended for people like them and for people in their age group.

We excluded 1 pair of advertisements because only 10% of participants (n=2) perceived the celebrity to be Latinx and another pair because the majority did not recognize the celebrity. We excluded 2 additional pairs because they scored the lowest on both celebrity recognition and attractiveness measures. The final set of 4 advertisement pairs (Figure 1) included 3 advertisement pairs that featured Latina women alongside a sugar-sweetened beverage (n=2 pairs) or unhealthy food product (n=1 pair) and one that featured a Latinx male alongside a sugar-sweetened beverage. Similar to previous research, we blurred the faces and names of the celebrities in the advertisements in Figure 1. The participants saw unblurred images [35,68].

We also used Photoshop to create 2 additional advertisements to serve as neutral nontargeted stimuli. These 2 advertisements featured bland, unsavory products (ie, water and saltine crackers) and did not show people (Figure 2).

They were used to compare advertisement ratings by Latinx and White adolescents to enable us to control for overall differences in advertisement ratings in our final models.

XSL•FO
RenderX

Figure 1. Final set of 4 pairs of advertising stimuli embedded into web-based survey. Each pair included (A) an advertisement featuring a Latinx noncelebrity and (B) an advertisement featuring a Latinx celebrity. Faces and names of the celebrities are blurred in the figure, but participants saw unblurred images.

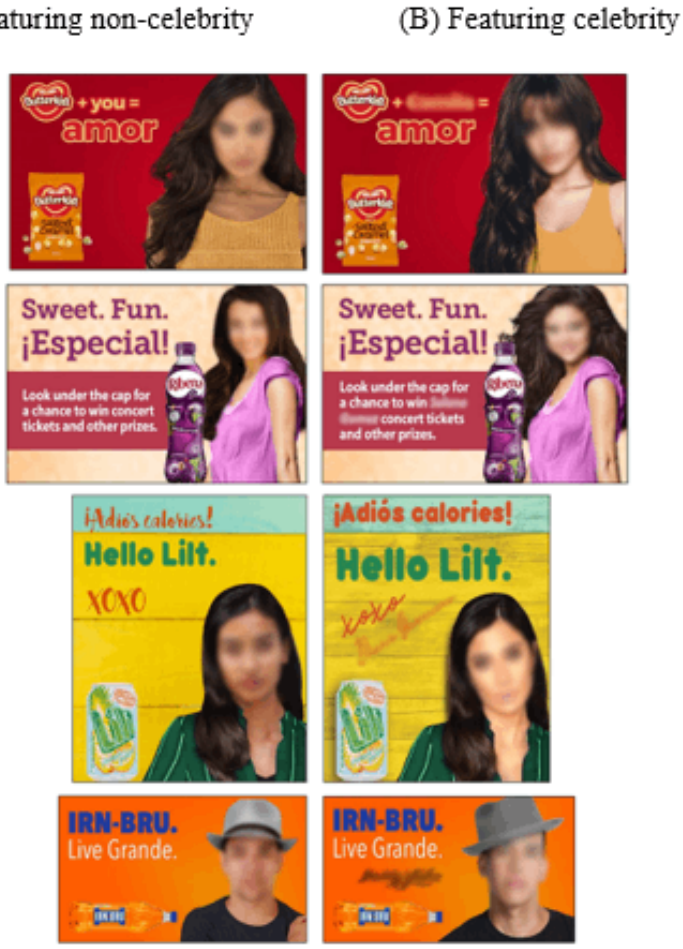

Figure 2. Final two 2 neutral, nontargeted advertising stimuli (control) embedded into web-based survey.



XSL•FO
RenderX

Survey Procedures

The web-based survey was designed using Silver Lake and CPP Investment's Qualtrics software, and the functionality of the survey was tested by the investigators before launching the study. Participants were required to complete the survey on a computer (instead of a mobile device) to ensure the visibility of advertisements. Participating on a mobile device might have made it more difficult to view the features of the advertisements in the survey due to the smaller screen size compared to desktops. Median survey completion time was 23 minutes. The study setting was anywhere a participant could access a computer to complete the experimental survey.

After providing assent and answering eligibility questions, participants were told the following cover story to conceal the true purpose of the study: "We are working with MTV to select commercials that their viewers might like and are interested in your attitudes and opinions on products featured in these advertisements. Participation in this study will involve viewing several different food and beverage advertisements and answering questions about them." All participants viewed and rated each of the 10 advertisements in a specific order as follows: they first viewed the neutral advertisement for water; then 2 celebrity advertisements and 2 noncelebrity advertisements for different brands presented in random order; then viewed the neutral advertisement for crackers; and finally, the remaining celebrity and noncelebrity advertisements presented in random order. This order ensured that participants did not see advertisements for 2 brands in the same group of 4 advertisements.

Primary Outcome Measures

The primary outcome measures and dependent variables are listed in Table 1.

Table 1. Primary outcome measures for cognitive responses to advertisements, advertisement efficacy, and behavioral intentions after viewing advertisements.

Outcomes	Survey questions
Advertisement rating questions: Cognitive responses to advertisements and advertisement efficacy	
Likeability[a]	"How much did you like this ad?" (0="strongly disliked it" to 100="liked it a lot") [16,36]
Positive affect	"Did the ad for [Coke] make you feel: cheerful; good; pleased; stimulated; soothed?" (0="not at all" to 100="very much so". Scores were averaged into a composite positive affect score [69].
Negative affect	"Did the ad for [Coke] make you feel: insulted; irritated; repulsed?" (0="not at all" to 100="very much so". Scores were averaged into a composite negative affect score [69].
Taste preference	"How much do you think you would like the taste of this product" (0="Not at all" to 100="A lot") [53].
Advertisement efficacy[b]	"I like the brand in this ad," "I react favorably to the brand in this ad," "The brand in this ad is appealing," "The brand in this ad is good," and "The brand in this ad is pleasant," (0="disagree" to 100="agree") [70]. Brand liking statements were averaged for a combined brand perception rating.
Behavioral intention questions	
Likelihood to consume the advertised brand	"How likely would you be to consume this product if it was offered to you?" (0="very unlikely" to 100="very likely") [71]
Likelihood of following or liking the brand on social media[a]	"What is the likelihood you would Follow or Like this brand on Facebook, Instagram, Twitter, Snapchat, or Tumblr?" (0="very unlikely" to 100="very likely") [16,36]
Willingness to engage with the brand or advertisement on social media[a]	How likely would you be to engage with this brand on social media?: "I would comment on the brand's post" (0="very unlikely" to 100="very likely") ; "I would "Like" this brand's post" (0="very unlikely" to 100="very likely") ; "I would tag a friend" (0="very unlikely" to 100="very likely") [16,36]
Likelihood they were the intended audience for the post	"I feel the advertisement was intended for people like me" (0="disagree" to 100="agree"); "I do not believe the company was targeting consumers like me" (0="disagree" to 100="agree"); "The advertiser made this advertisement for people like me." (0="disagree" to 100="agree"). The inverse of question 2 was calculated and the scores of the 3 questions were averaged into a composite score [72].

[a]This item has not been validated but has been used before in previous research [16,36].

[b]This item is based on a published commonly used marketing scale study but was adapted [70].

Participants provided their attitudes about advertisements and brands immediately after viewing each advertisement, before viewing the next advertisement, using a sliding scale (0-100) for all questions (Table 1). We used 0-100 response scales to minimize bias and lack of reliability versus Likert scales [73], and, among some participant populations (eg, under-resourced communities), Likert scales contradict other response formats [74]. Other studies have shown that 0-100 scales offer stronger psychometric properties than Likert scales [75].

The question, "How much did you like this ad?" (0="strongly disliked it" to 100="liked it a lot") assessed advertisement likeability. A commonly used and validated marketing survey was used to evaluate affective advertisement responses [69]. Participants responded to the following 8 prompts: "Did the ad for [Coke] make you feel…cheerful; good; pleased; stimulated (positive affect); soothed; insulted; irritated; repulsed (negative affect)?" Affective advertisement responses included positive (cheerful, good, pleased, stimulated, and soothed) and negative

XSL•FO
RenderX

(insulted, irritated, and repulsed) attributes, rated as 0="not at all" to 100="very much so." Positive and negative attributes were averaged into separate composite positive and negative affect scores. Taste preference was measured with 1 question: "How much do you think you would like the taste of this product" (0="Not at all" to 100="A lot") [69]. Advertisement efficacy was measured with an adapted version of a commonly used marketing scale that asked participants the following 5 questions: "I like the brand in this ad," "I react favorably to the brand in this ad," "The brand in this ad is appealing," "The brand in this ad is good," and "The brand in this ad is pleasant," and rated from 0="disagree" to 100="agree" [71]. Brand liking statements were averaged for a combined brand perception rating [70].

Behavioral intention measures included the likelihood to consume the advertised brand [71] and willingness to engage with the brand or advertisement on social media, including the following questions: "What is the likelihood you would Follow or Like this brand on Facebook, Instagram, Twitter, Snapchat, or Tumblr?"; How likely would you be to engage with this brand on social media? ("I would comment on the brand's post" [0="very unlikely" to 100="very likely"]; "I would "Like" this brand's post" [0="very unlikely" to 100="very likely"]; "I would tag a friend" [0="very unlikely" to 100="very likely"]). The willingness to engage measures have not been validated but have been used before in previous research [16,36]. All responses were assessed on a sliding scale from 0="very unlikely" to 100="very likely" (Figure 3).

Figure 3. Latinx (n=436) and White (n=467) adolescents' ratings of likelihood of (A) consuming the product if it was offered to them; (B) following or liking the brand on Facebook, Instagram, Twitter, Snapchat, or Tumblr; (C) "Liking" the post on social media; and (D) tagging a friend on social media. Error bars indicate SDs.

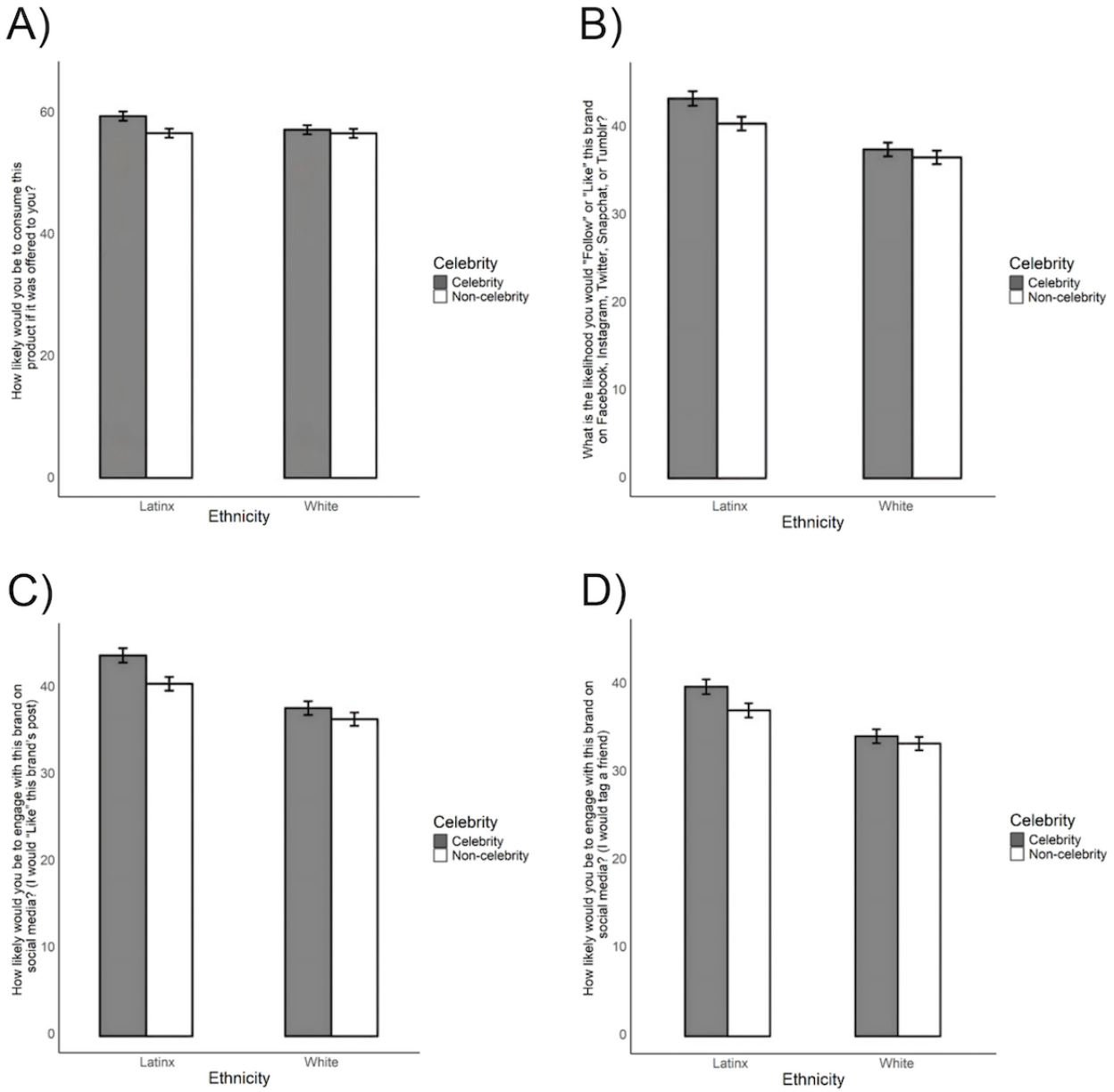

XSL•FO
RenderX

Participants rated the likelihood they were the intended audience for the post across all 8 advertisements that featured people, using a scale developed by Aaker et al [72]. The scale asked three questions: (1) "I feel the advertisement was intended for people like me," (2) "I do not believe the company was targeting consumers like me," and 3) "The advertiser made this advertisement for people like me." We assessed these responses using a sliding scale from 0="disagree" to 100="agree." We reverse-coded question 2 and then averaged the scores.

Other Measures

Participants provided demographic characteristics, including age, race, ethnicity, grade level, parental education, and height and weight. They then indicated how often they checked social media if they had access to a smartphone, and which social media platforms they used. Participants who identified as Latinx also completed the Multigroup Ethnic Identity Measure-Revised, a 6-item measure that assesses exploration of and commitment to ethnic identity on a 5-point Likert scale: 1=strongly disagree to 5=strongly agree. The overall composite Multigroup Ethnic Identity Measure-Revised score is calculated by averaging the scores for each of the 6 items. The scale has been validated among adolescents to measure ethnic identity (among Latinx adolescents, α=.930) [76-78].

Statistical Analysis

We used 2-tailed *t* tests (continuous variables) and chi-square tests (categorical variables) to compare the demographic characteristics of Latinx and White participants. We also compared Latinx and White participants' ratings of the 2 neutral advertisements on all outcomes to determine whether the 2 groups systematically rated advertisements differently on any of the measures.

We used multilevel linear regression to account for multiple ratings by each participant. Dependent variables included all attitude and behavioral measures. Independent variables included binary terms for participant ethnicity and presence of a celebrity in the advertisement, an interaction term between participant ethnicity and celebrity, random effect of participant, and categorical brand indicator. There were no differences in Latinx and White adolescents' ratings of the neutral advertisements that did not feature people, so we did not control for that variable in the model. Across all the outcomes except targetedness, Latinx participants did not rate the neutral advertisements any differently than the White participants. They differed on targetedness—Latinx participants said they thought the neutral advertisements were more targeted to them. Due to differences in age by participant ethnicity (*P*=.04), we included a categorical indicator of age as a covariate in all models. All analyses were conducted in R (R Foundation for Statistical Computing) in February 2020.

Results

Overview

More than half of participants (n=540, 59.8%) were aged 13-15 years and 40.2% (n=363) were aged 16-17 years (Table 2). Approximately one-half identified as Latinx (n=436, 48.3%). There were no significant differences in parental education (*P*=.21) or participant sex (*P*=.13) between Latinx and White participants. The majority of participants (n=870, 96.3%) reported having one or more social media accounts, and 41.1% reported using a smartphone "almost constantly" to access social media.

XSL•FO
RenderX

Table 2. Sociodemographic and behavioral characteristics of the study sample (N=903) and self-reported social media use survey responses.

	Total sample (N=903)	Latinx adolescents (n=436)	White adolescents (n=467)	*P* values for chi-squared tests[a]
Sociodemographic characteristics				
Age[b] (years), n (%)				.04
13-15	540 (59.8)	276 (63.3)	264 (56.5)	
16-17	363 (40.2)	160 (36.7)	203 (43.5)	
Highest parent educational level[c], n (%)				.21
High school, GED[d], or lower	191 (21.2)	81 (18.6)	110 (23.6)	
Vocational school or some college	327 (36.2)	167 (38.3)	160 (34.3)	
Bachelor degree	239 (26.5)	111 (25.5)	128 (27.4)	
Master degree or advanced graduate work	138 (15.3)	71 (16.3)	67 (14.4)	
Not sure or unknown	8 (0.9)	6(1.4)	2 (0.4)	
Sex, n (%)				.13
Male	148 (16.4)	79 (18.1)	69 (14.8)	
Female	753 (83.4)	355 (81.4)	398 (85.2)	
Nonbinary or other	2 (0.2)	2 (0.5)	0 (0.0)	
MEIM-R[e] (α=.930), n (%)	N/A[f]	60.23 (24.3)	N/A	
Less than or equal to median (60.8), n (%)	N/A	219 (50.2)	N/A	
Greater than median (60.8)), n (%)	N/A	217 (49.8)	N/A	
Behavioral characteristics				
How often do you use your smartphone to access social media accounts, such as Facebook, Instagram, and Snapchat, each day?[c], n (%)				.07
Almost constantly	371 (41.1)	193 (44.3)	178 (38.1)	
Several times a day	354 (39.2)	169 (38.8)	185 (39.6)	
Once a day or less frequently	178 (19.7)	74 (16.97)	104 (22.3)	
Social media access on smartphones[c], n (%)				.12
I have my own smartphone	833 (92.3)	409 (93.8)	424 (90.8)	
I do not have my own smartphone	70 (7.8)	27 (6.2)	43 (9.2)	
Which of the following social media accounts do you have?, n (%)				
Facebook[b]	644 (71.3)	296 (67.9)	348 (74.5)	
Instagram[b]	650 (71.98)	334 (76.6)	316 (67.7)	
Snapchat[b]	601 (66.6)	325 (74.5)	276 (59.1)	
Twitter	308 (34.1)	147 (33.7)	161 (34.5)	
Tumblr	66 (7.3)	37 (8.5)	29 (6.2)	
Other	14 (1.6)	6 (1.4)	8 (1.7)	
How many active social media accounts do you have?[b], n (%)				.10
0	33 (3.1)	11 (2.5)	22 (4.7)	
1	177 (19.8)	75 (17.2)	102 (21.8)	
2	228 (24.8)	110 (25.2)	118 (25.3)	
3	254 (28.0)	136 (31.2)	118 (25.3)	
4	171 (18.9)	80 (18.4)	91 (19.5)	
5	36 (3.99)	22 (5.1)	14 (3)	

XSL•FO
RenderX

	Total sample (N=903)	Latinx adolescents (n=436)	White adolescents (n=467)	*P* values for chi-squared tests[a]
6	4 (0.4)	2 (0.4)	2 (0.4)	
Average number of social media accounts, mean (SD)[b]	2.53 (1.2)	2.63 (1.2)	2.44 (1.3)	.02
Correctly identified celebrity status, (N=7224, 8 advertisements per person)[b], n (%)	5006 (69.3)	2572 (73.4)	2434 (65.2)	<.001

[a]*P* values for chi-square tests comparing the demographic data between Latinx and White adolescents.

[b]*P*<.05 comparing distributions between Latinx and White adolescents.

[c].05<*P*<.25 comparing distributions between Latinx and White adolescents.

[d]GED: General Educational Development.

[e]MEIM-R: Multigroup Ethnic Identity Measure-Revised.

[f]N/A: not applicable.

Attitudes About the Advertisement and the Brand

On average, participants reported liking the 8 advertisements that featured people (mean 53.58, SD 28.70; Table 3).

Positive affective responses were rated neutrally (mean 49.44, SD 28.85), while participants reported overall low negative affective responses to advertisements (mean 18.77, SD 25.79). Participants rated their perception of the brand on average as 51.63 (SD 29.25) across all 8 advertisements that featured people.

For all attitude measures except negative affect, all participants rated advertisements significantly more positively for advertisements with celebrities than with noncelebrities; there were no main effects of participant ethnicity nor interactions between celebrity and participant ethnicity. However, Latinx participants reported significantly more negative affect toward all advertisements (*P*=.01).

For brand perception, we found a main effect of celebrity but did not find an interaction, indicating that Latinx and White adolescents rated the brands with celebrities significantly higher than the brands without celebrities.

XSL•FO
RenderX

Table 3. Latinx (n=436) and White (n=467) adolescents' ratings of attitudes toward food advertisements featuring Latinx celebrities and Latinx noncelebrities, and behavioral intentions postadvertisement exposure[a,b].

Outcome	Latinx participants (n=436)		White participants (n=467)		Main effect of celebrity in the advertisement	Main effect of participant ethnicity	Interaction (participant ethnicity and celebrity in advertisement)
	Celebrity, mean (SD)	Noncelebrity, mean (SD)	Celebrity, mean (SD)	Noncelebrity, mean (SD)	β (*P* value)	β (*P* value)	β (*P* value)
Intended audience (targeting)	58.71 (0.49)	56.51 (0.47)	55.51 (0.51)	55.48 (0.50)	.03 (.95)	1.12 (.33)	2.17 (<.001)
Attitudes							
How much did you like or dislike the ad?	55.53 (0.69)	52.72 (0.67)	53.85 (0.68)	52.30 (0.66)	2.15 (<.001)	1.22 (.44)	—[c]
Positive affective response	51.52 (0.68)	48.93 (0.67)	49.45 (0.69)	47.96 (0.67)	2.02 (<.001)	1.66 (.32)	—
Negative affective response	20.69 (0.66)	20.69 (0.64)	17.05 (0.57)	16.58 (0.55)	0.24 (.46)	3.77 (.01)	—
How much do you think you would like the taste of this product?	57.15 (0.71)	55.65 (0.71)	55.49 (0.72)	54.73 (0.71)	1.12 (.01)	1.42 (.38)	—
Brand perception	53.55 (0.69)	50.94 (0.68)	51.54 (0.70)	50.59 (0.69)	0.96 (.07)	0.52 (.76)	1.66 (.03)
Behavioral intentions							
How likely would you be to consume this product if it was offered to you?	58.88 (0.73)	56.12 (0.73)	56.67 (0.75)	56.09 (0.73)	0.59 (.36)	0.13 (.94)	2.17 (.02)
What is the likelihood you would "Follow" or "Like" this brand on Facebook, Instagram, Twitter, Snapchat, or Tumblr?	43.23 (0.82)	40.38 (0.78)	37.44 (0.78)	36.53 (0.76)	0.92 (.08)	4.01 (.049)	1.94 (.01)
How likely would you be to engage with this brand on social media?							
I would comment on the brand's post	40.13 (0.81)	37.85 (0.79)	35.74 (0.78)	34.88 (0.76)	1.54 (.01)	3.79 (.06)	—
I would "Like" this brand's post	43.91 (0.82)	40.65 (0.79)	37.84 (0.80)	36.56 (0.77)	1.28 (.02)	4.27 (.04)	1.98 (.01)
I would tag a friend	39.28 (0.83)	36.63 (0.80)	33.71 (0.79)	32.90 (0.76)	0.81 (.12)	3.92 (.06)	1.85 (.01)

[a]All models included age as a covariate.

[b]All response items were on a scale ranging from 0 to 100.

[c]Not available.

Behavioral Intentions

Participants reported that they would likely consume the product (mean 56.92, SD 31.22), and there was no main effect of celebrity or participant ethnicity. However, we found a significant interaction between participant ethnicity and celebrity (Table 3). Only Latinx participants indicated significantly higher consumption intention when the advertisement featured a celebrity compared to advertisements featuring a noncelebrity (*P*<.001).

On measures of social media engagement, participants rated their likelihood of following the brand on average 39.31 (SD 33.51) across all 8 advertisements. They indicated a similar likelihood to comment on the social media post (mean 37.09; SD 33.34), "like" the post (mean 39.65, SD 33.81), and "tag" a friend in the post (mean 35.55, SD 33.88).

The main effect of celebrity was significant for two of the four social media engagement measures: (1) participants were more likely to comment on or "Like" posts that featured celebrities in their advertisements compared to posts that featured noncelebrities (Table 3); (2) Latinx participants were more likely to "Follow" or "Like" brands or posts than White participants. There was a main effect of celebrity on participant's likelihood of commenting on the post such that participants reported a higher likelihood of commenting on posts that featured

XSL•FO
RenderX

celebrities (mean 38.1, SE 1.04 compared to those without (mean 36.5, SE 1.04; β=1.54; t_{6320}=4.25; *P*<.001). We also found the main effects of celebrity (β=1.28; t_{6319}=2.38; *P*=.02) and participant ethnicity (β=4.27; t_{968}=2.09; *P*=.04) for willingness to "like" the brand. The main effects were qualified by an interaction (β=1.98; t_{6319}=2.57; *P*=.01). While White participants reported higher willingness to "like" advertisements with celebrities compared to noncelebrities (mean 36.7, SE 1.42 vs mean 38.0, SE 1.42; *z*=2.38; *P*=.02); the effect of celebrity was significantly greater among Latinx participants (mean 41.0, SE 1.49 vs mean 44.3, SE 1.49; *z*=5.87; *P*<.001).

For willingness to "tag" a brand, we found no main effects of celebrity (t_{6319}=1.55; *P*=.12) or participant ethnicity (t_{963}=1.91; *P*=.06). However, we did find a significant interaction (t_{6319}=2.46; *P*=.01). While White participants were equally likely to "tag" an advertisement with a celebrity (mean 33.9, SE 1.43) compared with an advertisement without a celebrity (mean 33.1, SE 1.43; *z*=1.55; *P*=.12), Latinx participants were more likely to "tag" an advertisement with a celebrity (mean 39.7, SE 1.50) compared to an advertisement without a celebrity (mean 37.0, SE 1.50; *z*=4.91; *P*<.001).

Finally, for willingness to follow a brand on social media, we found no main effect of celebrity (t_{6319}=1.76; *P*=.08), but did find evidence of a main effect of participant ethnicity (β=4.01; t_{964}=1.98; *P*=.05), as well as a significant interaction (β=1.94; t_{6319}=2.59; *P*=.01). White adolescents reported the same willingness to follow a brand on social media regardless of whether the advertisement featured a celebrity (mean 37.6, SE 1.42) or not (mean 36.7, SE 1.42; *P*=.08). However, Latinx participants reported higher likelihood of following the brand on social media when the advertisement featured a celebrity (mean 43.6, SE 1.48 vs mean 40.7, SE 1.48; *z*=5.30; *P*<.001).

Discussion

Principal Findings

Our hypothesis for research question 1, is that both Latinx and White adolescents would report higher preferences, including more positive attitudes (advertisement likeability; positive affect; and brand perceptions) and behavioral intentions (product consumption; social media engagement), when advertisements featured Latinx celebrities versus Latinx noncelebrities was mostly supported. Specifically, both Latinx and White participants in our sample reported significantly higher positive attitudes and brand perceptions for brands endorsed by Latinx celebrities compared to Latinx noncelebrities. However, we found significant main effects of celebrity advertisements for just 2 behavioral intentions: likelihood to comment and like brand posts. The findings that celebrities increased the appeal of the advertisements and products reinforces previous research that showed celebrity-endorsed products are more appealing to adolescents [41] and preadolescents [79] compared to products endorsed by noncelebrities.

Our hypothesis for research question 2, that Latinx participants would report more positive attitudes and behavioral intentions for both Latinx celebrity and Latinx noncelebrity advertisements compared to White participants was supported for just 2 outcomes: likelihood to follow brands on social media and like a brand post. We found no significant main effects of participant ethnicity on positive attitudes, brand perceptions, or intent to consume. However, we did find the hypothesized interaction between celebrity and participant ethnicity for most behavioral intentions. Relative to White adolescents, Latinx adolescents reported a higher willingness to "like" and follow brands that featured celebrities compared to brands that featured noncelebrities. In addition, Latinx adolescents reported higher consumption intentions and likelihood to tag a post when advertisements featured celebrities compared to noncelebrities.

These results reinforce other studies on the effects of celebrity endorsements and targeted food marketing. Our findings that both Latinx and White adolescents reported a higher likelihood to comment, "like," or tag a friend in the post on social media when the advertisements featured celebrities versus noncelebrities is similar to previous research showing that celebrity endorsements increase willingness to engage with brands [80]. Like other studies demonstrating that celebrity endorsements made adolescents more likely to choose the endorsed product [79], we identified a main effect of celebrity on adolescents' willingness to consume the product—but only for Latinx adolescents.

Latinx celebrities may disproportionately influence Latinx adolescents' behavioral intentions, including intent to consume and follow brands on social media, compared to White adolescents and noncelebrity advertisements. One potential explanation for the disproportionate influence of Latinx celebrities on Latinx youth is that Latinx communities are underrepresented in popular media, and positive portrayals in the media may heighten Latinx adolescents' responses to targeted advertisements. Future studies should examine the extent to which exposure to Latinx-targeted advertisements increases Latinx adolescents' actual food purchases and caloric consumption.

This study makes several contributions to the literature on the effects of food advertising and exposure to racially targeted food advertising. First, this is the first experimental study to examine the intersection of food advertising and celebrity endorsements among Latinx youth. The results indicate a direct causal effect of celebrities on willingness to consume the product and follow or "like" the brand on social media for both Latinx and White youth, even though—after controlling for other possible explanations—Latinx youth were more likely to believe that the advertisements with Latinx celebrities targeted them. These findings demonstrate the power of celebrity endorsements on adolescents' preferences. Second, this study demonstrates that the intersection between celebrity endorsements and Latinx ethnicity is complex and warrants further research to disentangle these effects. Celebrities influenced adolescents' attitudes about the advertisements, but it appears that Latinx celebrities may have a disproportionate influence on Latinx adolescents' behavioral intentions.

Limitations

This study has several limitations. First, although we only included brands from outside the United States to avoid brand

XSL•FO
RenderX

familiarity potentially influencing outcomes, our pretest did not assess brand familiarity among participants. Some adolescents may have been familiar with the international brands in our advertisements, thereby influencing their ratings. An additional limitation is that we only included Latinx celebrities and Latinx noncelebrities. The lack of White people in advertisements limits our ability to discern if our outcomes were shaped by ethnic congruity or some other factor. Another limitation is that in using Dynata for participant recruitment, our sample was not nationally representative. Participants in Dynata's sample are more likely to use the internet than the general US population, but since our research studies social media, this sample may better reflect our population of interest. Another limitation of opt-in panels is the risk of response bias (eg, providing socially desirable answers). However, we believe this risk is greatly reduced because the survey responses are anonymous, and research involving Dynata shows that adolescents express a variety of preferences (ie, they do not give socially desirable answers) [52-57]. Another limitation is that within-subjects study designs introduce possible "order effects" bias. Because our participants saw each advertisement (eg, the popcorn advertisement featuring a celebrity) and later saw its counterpart (eg, popcorn advertisement featuring a noncelebrity), order effects bias would suggest that the ratings of the second advertisement in a given pair could be affected by their reaction to the first advertisement in that pair. We aimed to reduce such bias, however, by (1) separating the presentation order of advertisements so that participants never saw two similar advertisements in a row; and (2) randomizing the order (eg, for each advertisement pair, some participants saw the celebrity advertisement first, whereas others saw the noncelebrity advertisement first). Although we would have preferred to avoid any possibility of order effects by using randomization via a between-subjects design, within-subjects is valuable when large sample sizes are not feasible—when designing the study, we knew we would not be able to recruit a large enough sample of Latinx adolescents to warrant a between-subjects design with multiple study conditions. Within-subjects' designs are commonly used in experimental research studies, including studies focused on food-related behaviors [81-83]. Finally, we did not assess participants' familiarity with the people featured in the advertisements, nor did we ask how much they liked the person in the advertisement. Both of those ratings should have been controlled in our analyses because they could have affected participants' ratings. Our study has major strengths, though, including the tightly controlled manipulation of images and the pretesting of those matched celebrities and noncelebrities on perceived attractiveness and target audience. Research on the effects of monoethnic and celebrity advertising effectiveness on the Generation Z cohort should also be explored.

Conclusions

Our data advance the theoretical understanding of targeting marketing by showing that viewing ethnicity-congruent food advertisements is particularly powerful for Latinx youth, which is concerning given Latinx communities experience high rates of diet-related diseases. These findings can help support policies that address unhealthy food marketing that targets communities of color, similar to how tobacco research on targeted marketing helped shed light on tobacco companies' targeted marketing practices [84-87]. Furthermore, in response to concerns about targeted marketing, Facebook now prohibits any advertisement from targeting users based on race [88]. New York State's proposed Predatory Marketing Prevention Act would aim to prevent advertisements from misleading consumers. These findings can be communicated to policymakers to consider including racial/ethnic-targeted marketing in such bills [89]. These findings suggest an urgent need to reduce celebrity endorsements in ethnically targeted advertisements that promote unhealthy food products to communities disproportionately affected by obesity and diabetes. Celebrities should consider choosing endorsements that support the health of their young fans, and powerful celebrities may be able to encourage companies to allow them to endorse healthier products. The food industry limits food advertising to children aged 12 years and younger, but our findings indicate that industry self-regulatory efforts should expand to include adolescents and address disproportionate advertising of unhealthy food to Latinx youth.

Acknowledgments

The authors would like to thank the following New York University Food Environment and Policy Research Coalition research assistants and staff who have no conflicts of interest to report: Alysa Miller, Yrvane Pageot, Margaret Eby, Tenay Greene, Joshua Arshonsky, Chelsea Mangold, Shirley Valerio, Robert Suss, Sana Husain, Natasha Pandit, Carolyn Fan, Anne Dumadag, Andrea Rodriguez Barrio, Ruchi Desai, Ingrid Wells, Jessica Osterman, and Jemar Bather. Generative artificial intelligence was not used in any portion of the manuscript writing. Data collection and analyses were supported by the National Institutes of Health (NIH) Early Independence Award (DP5OD021373-01; principal investigator: MAB) from the NIH Office of the Director. The NIH had no role in the design and conduct of the study.

Data Availability

The data generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Al-Hamad D, Raman V. Metabolic syndrome in children and adolescents. Transl Pediatr 2017;6(4):397-407 [FREE Full text] [doi: 10.21037/tp.2017.10.02] [Medline: 29184820]
2. Cook S, Weitzman M, Auinger P, Nguyen M, Dietz WH. Prevalence of a metabolic syndrome phenotype in adolescents: findings from the third National Health and Nutrition Examination Survey, 1988-1994. Arch Pediatr Adolesc Med 2003;157(8):821-827. [doi: 10.1001/archpedi.157.8.821] [Medline: 12912790]
3. Kansra AR, Lakkunarajah S, Jay MS. Childhood and adolescent obesity: a review. Front Pediatr 2021;8:581461 [FREE Full text] [doi: 10.3389/fped.2020.581461] [Medline: 33511092]
4. Lister NB, Baur LA, Felix JF, Hill AJ, Marcus C, Reinehr T, et al. Child and adolescent obesity. Nat Rev Dis Primers 2023;9(1):24. [doi: 10.1038/s41572-023-00435-4] [Medline: 37202378]
5. Todd AS, Street SJ, Ziviani J, Byrne NM, Hills AP. Overweight and obese adolescent girls: the importance of promoting sensible eating and activity behaviors from the start of the adolescent period. Int J Environ Res Public Health 2015;12(2):2306-2329 [FREE Full text] [doi: 10.3390/ijerph120202306] [Medline: 25690003]
6. Alberga AS, Sigal RJ, Goldfield G, Prud'homme D, Kenny GP. Overweight and obese teenagers: Why is adolescence a critical period? Pediatr Obes 2012;7(4):261-273. [doi: 10.1111/j.2047-6310.2011.00046.x] [Medline: 22461384]
7. Major JM, Cross AJ, Watters JL, Hollenbeck AR, Graubard BI, Sinha R. Patterns of meat intake and risk of prostate cancer among African-Americans in a large prospective study. Cancer Causes Control 2011;22(12):1691-1698. [doi: 10.1007/s10552-011-9845-1] [Medline: 21971816]
8. Cuevas AG, Krobath DM, Rhodes-Bratton B, Xu S, Omolade JJ, Perry AR, et al. Association of racial discrimination with adiposity in children and adolescents. JAMA Netw Open 2023;6(7):e2322839 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.22839] [Medline: 37432683]
9. Johnson VR, Acholonu NO, Dolan AC, Krishnan A, Wang EHC, Stanford FC. Racial disparities in obesity treatment among children and adolescents. Curr Obes Rep 2021;10(3):342-350 [FREE Full text] [doi: 10.1007/s13679-021-00442-0] [Medline: 33988825]
10. Mahmood N, Sanchez-Vaznaugh EV, Matsuzaki M, Sánchez BN. Racial/ethnic disparities in childhood obesity: the role of school segregation. Obesity 2022;30(5):1116-1125 [FREE Full text] [doi: 10.1002/oby.23416] [Medline: 35470976]
11. Alemán JO, Almandoz JP, Frias JP, Galindo RJ. Obesity among Latinx people in the United States: a review. Obesity 2023;31(2):329-337 [FREE Full text] [doi: 10.1002/oby.23638] [Medline: 36695058]
12. Boyland E, Muc M, Kelly B, Halford JCG, Vohra J, Rosenberg G, et al. Indirect associations between commercial television exposure and child body mass index. J Nutr Educ Behav 2021;53(1):20-27. [doi: 10.1016/j.jneb.2020.10.016] [Medline: 33423753]
13. Cairns G, Angus K, Hastings G, Caraher M. Systematic reviews of the evidence on the nature, extent and effects of food marketing to children. A retrospective summary. Appetite 2013;62:209-215 [FREE Full text] [doi: 10.1016/j.appet.2012.04.017] [Medline: 22561190]
14. Lobstein T, Jackson-Leach R, Moodie ML, Hall KD, Gortmaker SL, Swinburn BA, et al. Child and adolescent obesity: part of a bigger picture. Lancet 2015;385(9986):2510-2520 [FREE Full text] [doi: 10.1016/S0140-6736(14)61746-3] [Medline: 25703114]
15. Harris JL, Taillie LS. More than a nuisance: implications of food marketing for public health efforts to curb childhood obesity. Annu Rev Public Health 2024;45(1):213-233 [FREE Full text] [doi: 10.1146/annurev-publhealth-090419-102616] [Medline: 38109516]
16. Bragg M, Lutfeali S, Greene T, Osterman J, Dalton M. How food marketing on Instagram shapes adolescents' food preferences: online randomized trial. J Med Internet Res 2021;23(10):e28689 [FREE Full text] [doi: 10.2196/28689] [Medline: 34677136]
17. Bankole E, Harris N, Rutherford S, Wiseman N. A systematic review of the adolescent-directed marketing strategies of transnational fast food companies in low- and middle-income countries. Obes Sci Pract 2023;9(6):670-680 [FREE Full text] [doi: 10.1002/osp4.676] [Medline: 38090687]
18. Food marketing to kids. Center for Science in the Public Interest. URL: https://www.cspinet.org/advocacy/nutrition/food-marketing-kids [accessed 2024-03-15]
19. Vassallo AJ, Kelly B, Zhang L, Wang Z, Young S, Freeman B. Junk food marketing on Instagram: content analysis. JMIR Public Health Surveill 2018;4(2):e54 [FREE Full text] [doi: 10.2196/publichealth.9594] [Medline: 29871854]
20. Barker AB, Parkin M, Sinha S, Wilson E, Murray RL. A content analysis of 'junk food' content in children's TV programmes: a comparison of UK broadcast TV and video-on-demand services. J Public Health 2022;44(4):e506-e513 [FREE Full text] [doi: 10.1093/pubmed/fdac067] [Medline: 35731995]
21. Edwards CG, Pollack CC, Pritschet SJ, Haushalter K, Long JW, Masterson TD. Prevalence and comparisons of alcohol, candy, energy drink, snack, soda, and restaurant brand and product marketing on Twitch, Facebook gaming and YouTube gaming. Public Health Nutr 2022;25(1):1-12 [FREE Full text] [doi: 10.1017/S1368980021004420] [Medline: 34693900]
22. Review of food marketing to children and adolescents—follow-up report. Federal Trade Commission. 2012. URL: https://www.ftc.gov/reports/review-food-marketing-children-adolescents-follow-report [accessed 2024-03-15]

XSL•FO
RenderX

23. Backholer K, Gupta A, Zorbas C, Bennett R, Huse O, Chung A, et al. Differential exposure to, and potential impact of, unhealthy advertising to children by socio-economic and ethnic groups: a systematic review of the evidence. Obes Rev 2021;22(3):e13144. [doi: 10.1111/obr.13144] [Medline: 33073488]
24. Boyland E, Maden M, Coates AE, Masterson TD, Alblas MC, Bruce AS, et al. Food and non-alcoholic beverage marketing in children and adults: a systematic review and activation likelihood estimation meta-analysis of functional magnetic resonance imaging studies. Obes Rev 2024;25(1):e13643. [doi: 10.1111/obr.13643] [Medline: 37766661]
25. Boyland E, McGale L, Maden M, Hounsome J, Boland A, Angus K, et al. Association of food and nonalcoholic beverage marketing with children and adolescents' eating behaviors and health: a systematic review and meta-analysis. JAMA Pediatr 2022;176(7):e221037 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.1037] [Medline: 35499839]
26. Harris JL, Yokum S, Fleming-Milici F. Hooked on junk: emerging evidence on how food marketing affects adolescents' diets and long-term health. Curr Addict Rep 2021;8(1):19-27. [doi: 10.1007/s40429-020-00346-4]
27. Heller L. Move over Millennials, Generation Z is in charge. Forbes. 2015. URL: https://www.forbes.com/sites/lauraheller/2015/08/14/move-over-millennials-generation-z-is-in-charge/ [accessed 2024-03-15]
28. Zmuda N. How Coke is targeting Black consumers. Ad Age. 2009. URL: https://adage.com/article/the-big-tent/marketing-coke-targeting-african-american-consumers/137716 [accessed 2009-07-01]
29. Brooks R, Christidis R, Carah N, Kelly B, Martino F, Backholer K. Turning users into 'unofficial brand ambassadors': marketing of unhealthy food and non-alcoholic beverages on TikTok. BMJ Glob Health 2022;7(6):e009112 [FREE Full text] [doi: 10.1136/bmjgh-2022-009112] [Medline: 35879104]
30. Pechmann C, Levine L, Loughlin S, Leslie F. Impulsive and self-conscious: adolescents' vulnerability to advertising and promotion. J Public Policy Mark 2005;24(2):202-221. [doi: 10.1509/jppm.2005.24.2.202]
31. Gorrese A, Ruggieri R. Peer attachment: a meta-analytic review of gender and age differences and associations with parent attachment. J Youth Adolesc 2012;41(5):650-672. [doi: 10.1007/s10964-012-9759-6] [Medline: 22476726]
32. Armsden GC, Greenberg MT. The inventory of parent and peer attachment: individual differences and their relationship to psychological well-being in adolescence. J Youth Adolesc 1987;16(5):427-454. [doi: 10.1007/BF02202939] [Medline: 24277469]
33. Kumanyika S, Grier S. Targeting interventions for ethnic minority and low-income populations. Future Child 2006;16(1):187-207. [doi: 10.1353/foc.2006.0005] [Medline: 16532664]
34. Shao W, Zhang Y, Cheng A, Quach S, Thaichon P. Ethnicity in advertising and millennials: the role of social identity and social distinctiveness. Int J Advert 2023;42(8):1377-1418. [doi: 10.1080/02650487.2023.2172663]
35. Janssen L, Schouten AP, Croes EAJ. Influencer advertising on Instagram: product-influencer fit and number of followers affect advertising outcomes and influencer evaluations via credibility and identification. Int J Advert 2022;41(1):101-127. [doi: 10.1080/02650487.2021.1994205]
36. Lutfeali S, Ward T, Greene T, Arshonsky J, Seixas A, Dalton M, et al. Understanding the extent of adolescents' willingness to engage with food and beverage companies' Instagram accounts: experimental survey study. JMIR Public Health Surveill 2020;6(4):e20336 [FREE Full text] [doi: 10.2196/20336] [Medline: 33107836]
37. Ares G, Antúnez L, de León C, Alcaire F, Vidal L, Natero V, et al. 'Even if you don't pay attention to it, you know it's there': a qualitative exploration of adolescents' experiences with digital food marketing. Appetite 2022;176:106128. [doi: 10.1016/j.appet.2022.106128] [Medline: 35718311]
38. Arya V, Sambyal R, Sharma A, Dwivedi YK. J Consum Behav 2023;23(2):556-585. [doi: 10.1002/cb.2214]
39. Fleming-Milici F, Harris JL. Adolescents' engagement with unhealthy food and beverage brands on social media. Appetite 2020;146:104501. [doi: 10.1016/j.appet.2019.104501] [Medline: 31669579]
40. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What people "Like": analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. J Med Internet Res 2018;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
41. Coates AE, Hardman CA, Halford JCG, Christiansen P, Boyland EJ. Social media influencer marketing and children's food intake: a randomized trial. Pediatrics 2019;143(4):e20182554. [doi: 10.1542/peds.2018-2554] [Medline: 30833297]
42. Forbes-Bell S, Bardey AC, Fagan P. Testing the effect of consumer-model racial congruency on consumer behavior. Int J Market Res 2019;62(5):599-614. [doi: 10.1177/1470785319865892]
43. Sherman LE, Payton AA, Hernandez LM, Greenfield PM, Dapretto M. The power of the like in adolescence: effects of peer influence on neural and behavioral responses to social media. Psychol Sci 2016;27(7):1027-1035 [FREE Full text] [doi: 10.1177/0956797616645673] [Medline: 27247125]
44. Scissors L, Burke M, Wengrovitz S. What's in a Like?: Attitudes and behaviors around receiving Likes on Facebook. 2016 Presented at: CSCW '16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; February 27-March 2, 2016; San Francisco, CA p. 1501-1510. [doi: 10.1145/2818048.2820066]
45. Keller KL. Conceptualizing, measuring, and managing customer-based brand equity. J Mark 1993;57(1):1-22. [doi: 10.1177/002224299305700101]
46. Harris JL, Sacco SJ, Fleming-Milici F. TV exposure, attitudes about targeted food ads and brands, and unhealthy consumption by adolescents: modeling a hierarchical relationship. Appetite 2022;169:105804. [doi: 10.1016/j.appet.2021.105804] [Medline: 34780811]

XSL•FO
RenderX

47. Fazio RH, Powell MC, Williams CJ. The role of attitude accessibility in the attitude-to-behavior process. J Consum Res 1989;16(3):280-288. [doi: 10.1086/209214]
48. Muntinga DG, Moorman M, Smit EG. Introducing COBRAs. Int J Advertising 2015;30(1):13-46. [doi: 10.2501/ija-30-1-013-046]
49. Bargh JA, Chen M, Burrows L. Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. J Pers Soc Psychol 1996;71(2):230-244. [doi: 10.1037//0022-3514.71.2.230] [Medline: 8765481]
50. Bongers P, Jansen A, Havermans R, Roefs A, Nederkoorn C. Happy eating: the underestimated role of overeating in a positive mood. Appetite 2013;67:74-80. [doi: 10.1016/j.appet.2013.03.017] [Medline: 23583314]
51. Evers C, Adriaanse M, de Ridder DTD, de Witt Huberts JC. Good mood food. Positive emotion as a neglected trigger for food intake. Appetite 2013;68:1-7. [doi: 10.1016/j.appet.2013.04.007] [Medline: 23602962]
52. Bragg MA, Miller AN, Kalkstein DA, Elbel B, Roberto CA. Evaluating the influence of racially targeted food and beverage advertisements on Black and White adolescents' perceptions and preferences. Appetite 2019;140:41-49 [FREE Full text] [doi: 10.1016/j.appet.2019.05.001] [Medline: 31055011]
53. Roberto CA, Bragg MA, Schwartz MB, Seamans MJ, Musicus A, Novak N, et al. Facts up front versus traffic light food labels: a randomized controlled trial. Am J Prev Med 2012;43(2):134-141. [doi: 10.1016/j.amepre.2012.04.022] [Medline: 22813677]
54. Gorski Findling MT, Werth PM, Musicus AA, Bragg MA, Graham DJ, Elbel B, et al. Comparing five front-of-pack nutrition labels' influence on consumers' perceptions and purchase intentions. Prev Med 2018;106:114-121 [FREE Full text] [doi: 10.1016/j.ypmed.2017.10.022] [Medline: 29066375]
55. Moran AJ, Roberto CA. Health warning labels correct parents' misperceptions about sugary drink options. Am J Prev Med 2018;55(2):e19-e27 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.018] [Medline: 29903567]
56. Musicus AA, Moran AJ, Lawman HG, Roberto CA. Online randomized controlled trials of restaurant sodium warning labels. Am J Prev Med 2019;57(6):e181-e193. [doi: 10.1016/j.amepre.2019.06.024] [Medline: 31753271]
57. VanEpps EM, Roberto CA. The influence of sugar-sweetened beverage warnings: a randomized trial of adolescents' choices and beliefs. Am J Prev Med 2016;51(5):664-672 [FREE Full text] [doi: 10.1016/j.amepre.2016.07.010] [Medline: 27617366]
58. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychol Sci 2007;18(5):429-434. [doi: 10.1111/j.1467-9280.2007.01917.x] [Medline: 17576283]
59. Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. In: Zanna MP, editor. Advances in Experimental Social Psychology. Amsterdam, Netherlands: Elsevier; 1991:201-234.
60. Casey BJ, Getz S, Galvan A. The adolescent brain. Dev Rev 2008;28(1):62-77 [FREE Full text] [doi: 10.1016/j.dr.2007.08.003] [Medline: 18688292]
61. Lakon CM, Hipp JR, Wang C, Butts CT, Jose R. Simulating dynamic network models and adolescent smoking: the impact of varying peer influence and peer selection. Am J Public Health 2015;105(12):2438-2448. [doi: 10.2105/AJPH.2015.302789] [Medline: 26469641]
62. Cho H, Li W, Shen L, Cannon J. Mechanisms of social media effects on attitudes toward E-cigarette use: motivations, mediators, and moderators in a national survey of adolescents. J Med Internet Res 2019;21(6):e14303 [FREE Full text] [doi: 10.2196/14303] [Medline: 31250830]
63. Anderson L, McCabe DB. A coconstructed world: adolescent self-socialization on the internet. J Public Policy Mark 2012;31(2):240-253. [doi: 10.1509/jppm.08.043]
64. Downs JS, Bruine de Bruin W, Fischhoff B, Murray PJ. Behavioral decision research intervention reduces risky sexual behavior. Curr HIV Res 2015;13(5):439-446 [FREE Full text] [doi: 10.2174/1570162x13666150511145328] [Medline: 26149165]
65. Tajfel H, Turner J. Social identity theory and self-categorization theory: a historical review. In: Hornsey MJ, editor. Social and Personality Psychology Compass. Oxford, United Kingdom: Blackwell Publishing; 2008.
66. Grier SA, Brumbaugh AM. Consumer distinctiveness and advertising persuasion. In: Diversity in Advertising. London, United Kingdom: Psychology Press; 2004:217-237.
67. Harris JL, Haraghey KS, Lodolce M, Semenza NL. Teaching children about good health? Halo effects in child-directed advertisements for unhealthy food. Pediatr Obes 2018;13(4):256-264. [doi: 10.1111/ijpo.12257] [Medline: 29076259]
68. Schouten AP, Janssen L, Verspaget M. Celebrity vs. influencer endorsements in advertising: the role of identification, credibility, and product-endorser fit. Int J Advertising 2019;39(2):258-281. [doi: 10.1080/02650487.2019.1634898]
69. Olney TJ, Holbrook MB, Batra R. Consumer responses to advertising: the effects of ad content, emotions, and attitude toward the ad on viewing time. J Consum Res 1991;17(4):440-453. [doi: 10.1086/208569]
70. Delgado-Ballester E, Alemán JLM. Brand trust in the context of consumer loyalty. Eur J Mark 2001;35((11/12)):1238-1258. [doi: 10.1108/eum0000000006475]
71. Spears N, Singh SN. Measuring attitude toward the brand and purchase intentions. J Curr Issues Res Advertising 2004;26(2):53-66. [doi: 10.1080/10641734.2004.10505164]
72. Aaker JL, Brumbaugh AM, Grier SA. Nontarget markets and viewer distinctiveness: the impact of target marketing on advertising attitudes. J Consum Psychol 2008;9(3):127-140. [doi: 10.1207/s15327663jcp0903_1]

XSL•FO
RenderX

73. Dolnicar S. 5/7-point “Likert scales” aren't always the best option. Ann Tourism Res 2021;91(4):103297. [doi: 10.1016/j.annals.2021.103297]
74. Ogden J, Lo J. How meaningful are data from Likert scales? An evaluation of how ratings are made and the role of the response shift in the socially disadvantaged. J Health Psychol 2012;17(3):350-361. [doi: 10.1177/1359105311417192] [Medline: 21824896]
75. Kan A. Effect of scale response format on psychometric properties in teaching self-efficacy. Eurasian J Educ Res 2009;8(34):215-228 [FREE Full text]
76. Norris AE, Ford K, Bova CA. Psychometrics of a brief acculturation scale for Hispanics in a probability sample of urban Hispanic adolescents and young adults. Hisp J Behav Sci 1996;18(1):29-38. [doi: 10.1177/07399863960181004]
77. Phinney JS, Ong AD. Conceptualization and measurement of ethnic identity: current status and future directions. J Couns Psychol 2007;54(3):271-281. [doi: 10.1037/0022-0167.54.3.271]
78. Burrow-Sanchez JJ. Measuring ethnic identity in Latino adolescents with substance use disorders. Subst Use Misuse 2014;49(8):982-986. [doi: 10.3109/10826084.2013.794839] [Medline: 24779497]
79. Dixon H, Scully M, Niven P, Kelly B, Chapman K, Donovan R, et al. Effects of nutrient content claims, sports celebrity endorsements and premium offers on pre-adolescent children's food preferences: experimental research. Pediatr Obes 2014;9(2):e47-e57. [doi: 10.1111/j.2047-6310.2013.00169.x] [Medline: 23630014]
80. Duthie E, Veríssimo D, Keane A, Knight AT. The effectiveness of celebrities in conservation marketing. PLoS One 2017;12(7):e0180027 [FREE Full text] [doi: 10.1371/journal.pone.0180027] [Medline: 28686595]
81. Goffe L, Wrieden W, Penn L, Hillier-Brown F, Lake AA, Araujo-Soares V, et al. Reducing the salt added to takeaway food: within-subjects comparison of salt delivered by five and 17 holed salt shakers in controlled conditions. PLoS One 2016;11(9):e0163093 [FREE Full text] [doi: 10.1371/journal.pone.0163093] [Medline: 27668747]
82. Liu Y, Roefs A, Nederkoorn C. Food palatability directs our eyes across contexts. Front Psychol 2021;12:664893 [FREE Full text] [doi: 10.3389/fpsyg.2021.664893] [Medline: 34135823]
83. Passia N, Chaar MS, Krummel A, Nagy A, Freitag-Wolf S, Ali S, et al. Influence of the number of implants in the edentulous mandible on chewing efficacy and oral health-related quality of life—a within-subject design study. Clin Oral Implants Res 2022;33(10):1030-1037. [doi: 10.1111/clr.13984] [Medline: 35869607]
84. Harris JL. Targeted food marketing to Black and Hispanic consumers: the tobacco playbook. Am J Public Health 2020;110(3):271-272. [doi: 10.2105/AJPH.2019.305518] [Medline: 32023097]
85. Cruz TB, Rose SW, Lienemann BA, Byron MJ, Meissner HI, Baezconde-Garbanati L, et al. Pro-tobacco marketing and anti-tobacco campaigns aimed at vulnerable populations: a review of the literature. Tob Induc Dis 2019;17:68 [FREE Full text] [doi: 10.18332/tid/111397] [Medline: 31582956]
86. Moore DJ, Williams JD, Qualls WJ. Target marketing of tobacco and alcohol-related products to ethnic minority groups in the United States. Ethn Dis 1996;6(1-2):83-98. [Medline: 8882838]
87. Kelly B, Backholer K, Boyland E, Kent MP, Bragg MA, Karupaiah T, et al. Contemporary approaches for monitoring food marketing to children to progress policy actions. Curr Nutr Rep 2023;12(1):14-25 [FREE Full text] [doi: 10.1007/s13668-023-00450-7] [Medline: 36746878]
88. Bond S. Facebook scraps ad targeting based on politics, race and other 'sensitive' topics. National Public Radio. 2021. URL: https://www.npr.org/2021/11/09/1054021911/facebook-scraps-ad-targeting-politics-race-sensitive-topics [accessed 2024-03-15]
89. Senate Bill S7487C: 2021-2022 legislative session. The New York State Senate. URL: https://www.nysenate.gov/legislation/bills/2021/S7487 [accessed 2024-03-15]

Edited by A Mavragani; submitted 29.09.23; peer-reviewed by S Liu, S Bidmon, D Dinh; comments to author 23.02.24; revised version received 28.03.24; accepted 28.10.24; published 31.01.25.

Please cite as:
Bragg MA, Lutfeali S, Gabler DG, Quintana Licona DA, Harris JL
Latinx and White Adolescents' Preferences for Latinx-Targeted Celebrity and Noncelebrity Food Advertisements: Experimental Survey Study
J Med Internet Res 2025;27:e53188
URL: https://www.jmir.org/2025/1/e53188
doi:10.2196/53188
PMID:

©Marie A Bragg, Samina Lutfeali, Daniela Godoy Gabler, Diego A Quintana Licona, Jennifer L Harris. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted

XSL•FO
RenderX

use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Colorectal Cancer Racial Equity Post Volume, Content, and Exposure: Observational Study Using Twitter Data

Chau Tong[1,2], PhD; Drew Margolin[3], PhD; Jeff Niederdeppe[3,4], PhD; Rumi Chunara[5,6], PhD; Jiawei Liu[7,8], PhD; Lea Jih-Vieira[9], ME; Andy J King[10,11], PhD

[1]School of Journalism, University of Missouri, Columbia, MO, United States
[2]Institute for Data Science and Informatics, University of Missouri, Columbia, MO, United States
[3]Department of Communication, Cornell University, Ithaca, NY, United States
[4]Jeb E. Brooks School of Public Policy, Cornell University, Ithaca, NY, United States
[5]Department of Biostatistics, New York University, New York, NY, United States
[6]Department of Computer Science & Engineering, New York University, New York, NY, United States
[7]STEM Translational Communication Center, College of Journalism and Communications, University of Florida, Gainesville, FL, United States
[8]Department of Advertising, College of Journalism and Communications, University of Florida, Gainesville, FL, United States
[9]School of Engineering and Applied Science, University of Virginia, Charlottesville, VA, United States
[10]Cancer Control & Population Sciences, Huntsman Cancer Institute, Salt Lake, UT, United States
[11]Department of Communication, University of Utah, Salt Lake, UT, United States

Corresponding Author:
Chau Tong, PhD
School of Journalism
University of Missouri
140B Walter Williams
Columbia, MO, 65203
United States
Phone: 1 573 882 7875
Email: ctong@missouri.edu

Abstract

Background: Racial inequity in health outcomes, particularly in colorectal cancer (CRC), remains one of the most pressing issues in cancer communication and public health. Social media platforms like Twitter (now X) provide opportunities to disseminate health equity information widely, yet little is known about the availability, content, and reach of racial health equity information related to CRC on these platforms. Addressing this gap is essential to leveraging social media for equitable health communication.

Objective: This study aims to analyze the volume, content, and exposure of CRC racial health equity tweets from identified CRC equity disseminator accounts on Twitter. These accounts were defined as those actively sharing information related to racial equity in CRC outcomes. By examining the behavior and impact of these disseminators, this study provides insights into how health equity content is shared and received on social media.

Methods: We identified accounts that posted CRC-related content on Twitter between 2019 and 2021. Accounts were classified as CRC equity disseminators (n=798) if they followed at least 2 CRC racial equity organization accounts. We analyzed the volume and content of racial equity–related CRC tweets (n=1134) from these accounts and categorized them by account type (experts vs nonexperts). Additionally, we evaluated exposure by analyzing follower reach (n=6,266,269) and the role of broker accounts—accounts serving as unique sources of CRC racial equity information to their followers.

Results: Among 19,559 tweets posted by 798 CRC equity disseminators, only 5.8% (n=1134) mentioned racially and ethnically minoritized groups. Most of these tweets (641/1134, 57%) addressed disparities in outcomes, while fewer emphasized actionable content, such as symptoms (11/1134, 1%) or screening procedures (159/1134, 14%). Expert accounts (n=479; 716 tweets) were more likely to post CRC equity tweets compared with nonexpert accounts (n=319; 418 tweets). Broker accounts (n=500), or those with a substantial portion of followers relying on them for equity-related information, demonstrated the highest capacity for exposing followers to CRC equity content, thereby extending the reach of these critical messages to underserved communities.

Conclusions: This study emphasizes the critical roles played by expert and broker accounts in disseminating CRC racial equity information on social media. Despite the limited volume of equity-focused content, broker accounts were crucial in reaching

XSL•FO
RenderX

otherwise unexposed audiences. Public health practitioners should focus on encouraging equity disseminators to share more actionable information, such as symptoms and screening benefits, and implement measures to amplify the reach of such content on social media. Strengthening these efforts could help bridge disparities in cancer outcomes among racially minoritized groups.

(J Med Internet Res 2025;27:e63864) doi:10.2196/63864

KEYWORDS

racial equity information; information exposure; health disparities; colorectal cancer; cancer communication; Twitter; X

Introduction

Colorectal cancer (CRC) is among the most common and deadly cancers in the United States [1]. Although there have been advancements in CRC prevention and treatment, racial disparities in CRC morbidity and mortality persist, with Black Americans facing higher CRC risks [2]. Thus, promoting CRC awareness and screening behaviors among Black Americans for CRC prevention and early detection to reduce racial disparities has become an urgent task for CRC-related health communication and education [3-9]. Improving the quality of information about CRC prevention, detection, and advocacy is an important step in providing education and resources for improved decision-making to contribute to efforts to reduce CRC disparities among Black Americans [10-14].

Web-based sources—such as social media and related platforms—are an increasing source of health information for the general public, particularly for Black Americans. Black Americans are more likely than White Americans to rely on a wide variety of sources to acquire CRC screening-specific information—including web-based sources [15-20]. Research on CRC information has generally focused on the extent to which information presented is misleading, distracting, or inaccurate. For example, researchers found that medical professionals classified almost 40% of CRC YouTube content as "not useful," and such content was viewed and engaged with more frequently than higher-quality content [17]. TikTok content varied in its utility and accuracy, but researchers commonly identified inaccurate and misleading content [18]. On the other hand, other research on Twitter (now X) indicated that tweets about CRC were mostly accurate [19], while other studies noted that CRC received less Twitter attention than other (eg, breast and prostate) cancer sites [20].

For cancer topics where there are disparities by race, ethnicity, or other identities, misleading information is not the only way that members of the public can be underserved. Specifically, if specific information relevant to a particular community does not reach this audience [14,21-23], the information can be "misleading" in another sense. For example, since the rate of mortality from CRC among Black Americans is higher compared with other racial groups, misperceptions prevail as to the recommended screening age for this group, even more so after the official guideline changes from the recommended screening age of 50 to 45 years [16]. Thus, if Black Americans are only exposed to messages that describe overall population averages or recommendations, they may be inadvertently "misled" in the sense that they are not given access to important information to inform their decisions about prevention and screening [14,22].

This study focuses on an aspect of this problem of specialized/targeted information reaching populations that experience health inequities. Specifically, we examine the extent to which messages sent about and related to inequities regarding CRC by those motivated to address the issue are likely to reach historically marginalized populations or whether they tend to remain within the professional circles of the health experts themselves. The basic problem is that while in theory, social media can connect anyone to anyone, in practice, people tend to connect with similar others [24], resulting in clusters of tight-knit groups with overlapping, shared relationships [25]. These structures, sometimes referred to as "echo chambers," can create a particular problem for the dissemination of so-called "expert" information that, by definition, comes from specialized sources like medical journals and scholarly publications. For example, if a particular group is underrepresented in the scientific community, and the latest expert findings regarding the structural barriers to treatment and elevated CRC mortality faced by this underrepresented group are shared only among a cluster of experts, this important information may not reach and inform prevention, screening, and treatment decisions for people with historically minoritized identities.

Our study attempts to understand this distinction—between what is said about racial equity and CRC by accounts that are concerned with it and what is shared with populations facing these health outcome inequities—by analyzing both the volume and content of racial equity–related tweets from CRC equity disseminator accounts—an original concept that we introduced, defined as Twitter accounts that disseminate information related to racial equity concerning CRC outcomes.

Some individuals break out of communicating only within smaller, isolated clusters. Social network theory refers to these individuals as brokers [26,27]. Brokers are connected to different communities and are thus a key source of information transfer and diffusion [25,26]. They thus have an outsized influence on the information exposure for individuals and groups who are not connected to experts or other specialized communities. In essence, those who are disconnected do not hear everything that is said, but they hear what the brokers say. For this reason, this study also addresses a gap in the literature on CRC social media content and exposure by examining the influence of broker accounts—a special subset of CRC equity disseminator accounts that is more likely to reach unique audiences in disseminating racial equity content.

By addressing both aspects of content and dissemination, this study contributes to a more nuanced understanding of the public communication environment surrounding CRC racial equity, which will help inform future health communication and education efforts aimed at reducing CRC disparities.

XSL•FO
RenderX

Methods

Identifying CRC Equity Organization Accounts

We first identified a list of organizations whose primary focus is health equity or the health of racially and ethnically minoritized populations and the promotion of preventive health behaviors (including CRC screening; Figure 1). We used the US Department of Health and Human Services' National Minority Organizations list [28] to identify these health agencies (such as the National Institute on Minority Health and Health Disparities) and supplemented the list with organizations specifically dedicated to CRC prevention (such as the Association of Black Gastroenterologists and Hepatologists and the Colorectal Cancer Alliance). We then looked for the corresponding public Twitter accounts of these organizations and narrowed our pool from 18 to 7 accounts that had at least 1500 followers. The average number of followers was 9204, with the National Institute on Minority Health and Health Disparities being the most followed account (27,451 followers) and the Center for Black Health being the least followed (1594 followers). We retrieved the list of followers of these 7 organization accounts, resulting in a total of 59,669 unique followers (average number of organizations followed=1.08, SD 0.32). The majority of these users (n=55,725, 93.4%) only followed 1 of the 7 health equity organizations. A much smaller proportion, 3944 (6.6%) users followed at least 2 of the 7 organizations and were considered as having a potential interest in the topic of racial equity regarding CRC.

Figure 1. Workflow for mapping CRC equity networks and roles. CRC: colorectal cancer.

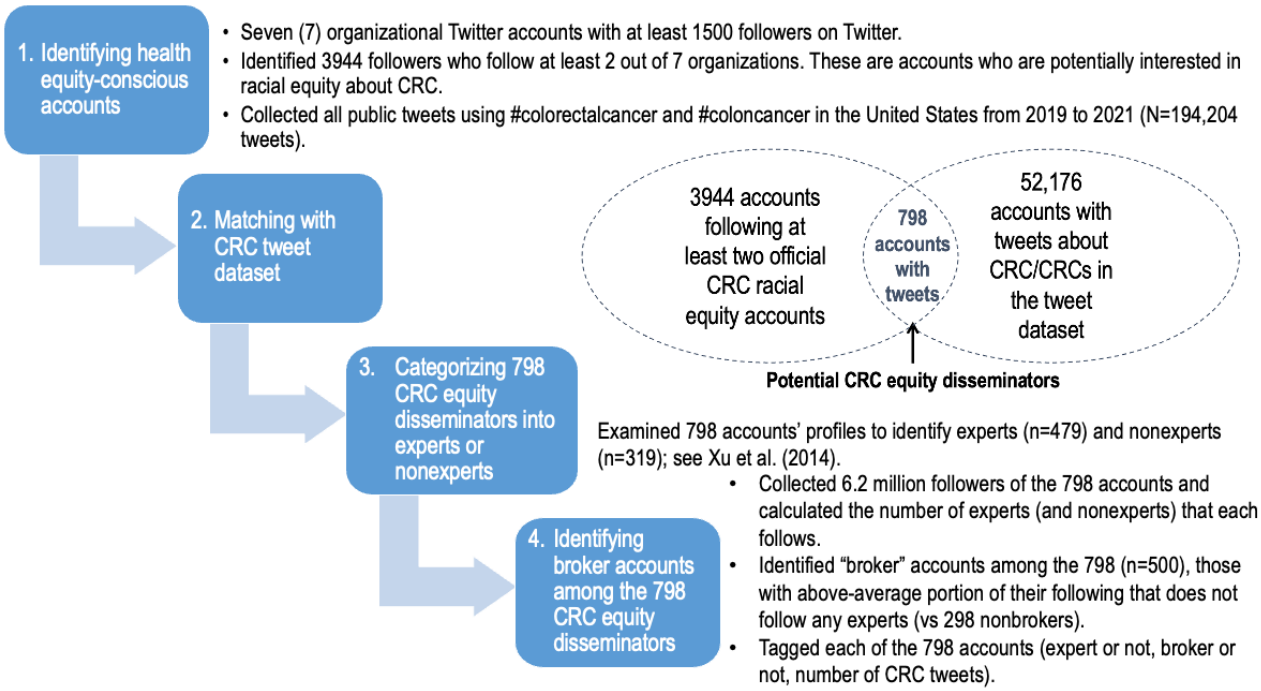

Identifying CRC Equity Disseminators

We collected public tweets about CRC from 2019 to 2021 using the keywords "coloncancer," "colorectalcancer," "colon cancer," and "colorectal cancer" and restricted the region to the United States (N=194,204 tweets) due to possible country-related differences about CRC advocacy and screening recommendations (eg, free screening services in some countries). Of the 3944 users identified as having a potential interest in racial equity about CRC, there were 798 accounts that produced tweets containing any of the terms above (n=25,093 CRC-related tweets; 19,559 of them are unique). We considered these 798 accounts as potential CRC equity disseminators; those who could play a crucial role in disseminating information about CRC racial equity to people who may not follow equity organizations. We note that only 0.7% of equity disseminators' followers also followed any CRC equity organization account, suggesting that these users' exposure to CRC racial equity content is less likely to come from equity organizations but more likely to come from equity disseminators themselves.

Identifying Expert and Nonexpert Accounts Among the CRC Equity Disseminators

To identify the identities of the 798 CRC equity disseminators, we categorized Twitter users based on their self-provided Twitter profiles. Following the inductive categorization framework introduced by Xu et al [29] and their typology for health knowledge sharing in a Twitter-based community of practice, in which users were grouped by their degree of health care involvement (ie, expert, those with experience, those interested in health topics, and the general public) and their health care roles (ie, providers, advocacy, engaged or average consumers, media, government, and non–health care–related organizations), we classified each of our 798 as either "expert" (including professional health service providers such as health care practitioners, health scientists, organizations, research laboratories, and medical centers) or "non-expert" (including advocacy groups, organizations, or individuals; average or engaged consumers; media; and government agencies; see Xu

XSL•FO
RenderX

et al [29] for details on each category and definition). This resulted in 479 experts versus 319 nonexperts.

Identifying Broker and Nonbroker Accounts Among the CRC Equity Disseminators

In addition to tagging each of the 798 accounts as either an expert or nonexpert account, we also determined whether each was a broker or nonbroker. To arrive at this categorization, we collected all the followers of the 798 accounts (N=6,266,269 unique followers) and calculated the number of expert and nonexpert accounts each of the followers followed. An account was classified as a broker account if more than half of its followers did not follow any experts (other than this one if the account in focus itself was an expert account). This resulted in 500 brokers and 298 nonbrokers.

Intersection of Expertise and Brokerage: The Role of Expert Brokers

Now that each of the 798 CRC equity disseminator accounts is tagged as either an expert or not and a broker or not, an account was considered an expert-broker account if it fell into both categories. Conceptually, an account is an expert-broker if it is an expert account, and more than half of its followers do not follow any other experts. Based on this definition, among the 479 experts, 254 (53%) were expert-brokers. These expert-brokers are especially important because they can uniquely connect people to content produced by experts. For example, a doctor who is followed by accounts that do not follow other doctors is more likely to be a crucial unique source of information for their followers who would not get important information (eg, about screening recommendations) elsewhere. Because our study is about access to medical information pertinent to minoritized populations, we are interested in the role of these expert-broker accounts as opposed to expert-nonbrokers (ie, whose followers tend to follow other experts as well, in this case, a doctor whose followers are mostly individuals who also follow many other doctors).

Coding of CRC Equity Tweets

We are interested in the specific CRC content about racial equity sent by the 798 CRC equity disseminators. We identified these tweets using a combination of (computational) dictionary tagging and manual (human) coding. We first tagged CRC-related tweets using a custom dictionary consisting of terms and phrases signaling historically minoritized racial groups (see full list of these terms and phrases in Table 1). Several specific terms were adapted from previous research such as [30,31].

Next, we analyzed the major themes in these CRC racial equity tweets and inductively identified 13 nonmutually exclusive categories for manual coding (Table 2). Two coders went through the codebook training process and reliably coded 10% of random tweets into 13 content subcategories (Krippendorf α ranged from 0.85 to 0.94). The subcategories are not mutually exclusive, so that a given tweet may fall into multiple categories. In Table 2, examples of tweets, in line with best practices of social media research [32], we have removed the names associated with certain hashtags and mentions to ensure the tweets are anonymized and cannot be traced back to their original authors.

Table 1. Keywords for retrieving potential racial equity content from colorectal cancer tweets.

Categories	Keywords
People of color	“people of color,” “community of color,” “communities of color,” “colored,” “poc,” “woc,” “woman of color,” “minority,” “minorities,” “racial disparity,” “racial disparities,” and “racism”
Black	“black*,” “african american*,” “af american,” “brown,” “blk,” “blm,” “brother,” “brotha,” “m4bl,” “sister,” and “sista”
Hispanic/Latino/a/x	“hispanic*,” “latino*,” “latinx*,” and “latina*”
Indigenous	“indigenous,” “american indian*,” and “native american*”
Asian	“asian*” and “asian american*”

XSL•FO
RenderX

Table 2. Categories, definitions, examples of colorectal cancer (CRC) racial equity tweets, and intercoder agreement.

Category	Definition	Examples of tweets	Krippendorf α
Outcome disparity	Mention of information about CRC or CRCs specific to more than 1 race or ethnicity; comparison/contrast between groups	Murphy et al show black/white disparity in rise in incidence driven by rectal cancer. Any data on Hispanics and Asians? Research presented at #[national research conference] shed some light on disparities in outcomes for African American and Caucasian #colorectalcancer patients. Listen to the story from @[media outlet]:	0.92
Call to action	Suggesting people talk to their doctors about screening; encouraging people to get screened as recommended by their doctors; reminding people to stay up to date on screening, etc.	Black men & women manifest at an earlier age with #colorectalcancer & should start screening at 40-45yo. If you have not had colonoscopy, please talk to your PCP. @FightCRC @CCAlliance #disparities Let's change the stats. Help us prevent #cancer by encouraging brothers (ages 45-75) in Utah to visit #[national research conference] #CuttingCRC #MinorityHealth #BlackHistoryMonth	0.94
Risk factors	Mention of information of risks that are associated with CRC, such as family history, physical inactivity, diet, alcohol use, lynch syndrome, being overweight, being older, etc.	What are some preventive measures can #Latinos and all people take to lower their #colorectalcancer risk? #coloncancerawarenessmonth Get a colonoscopy if you smoke or if you are a Black male or if there is cancer in your family.	0.89
Symptoms	Mention of known indicators of CRC, such as blood in stool, changes in bowel habits, bloating, cramps, gas, pain, weight loss for no reason, etc.	On world cancer day 2021, I'm urging everyone to get screened for colorectal cancer. Esp. Black and brown folks. If you are 45 or older or are experiencing symptoms (abdominal pain, thin stool, blood in stool, fatigue, unexplained weight loss), talk to your doctor. A colonoscopy could save your life.	0.88
Raising awareness	Mention or stress the need of more widespread and open communication about CRC	Just like Black people need to have the talk with their sons about sex and being careful of cops, they need to be talking about colorectal cancer and how to prevent it. It is not just an old person's disease.	0.85
Advocacy	Advocating or endorsing screening in general, or specific CRC screening options	African Americans have the highest incidence of #colorectalcancer and highest mortality rate of any racial or ethnic group, according to the @AmericanCancer Society. Get screened! If not for you, do it for your loved ones! #BlackHistoryMonth #coloncancer #cancer #colonoscopy Promote colonoscopy screening among low-income Latinos at average risk of #ColorectalCancer. Here's new research using randomized clinical trial.	0.9
Celebrities' stories	Individual stories concerning public figures who are patients, advocates, or survivors of CRC	In GQ, Author Ibram X. Kendi writes about his #diagnosis of #coloncancer at 35, the early #warningsigns of #CRC and his experience going thru #treatment. #Blackmen are 40% more likely to die of #colorectalcancer than other races. Kendi is now #cancerfree. Back on @[media outlet] to provide commentary on the stigma of colon cancer and the benefit of the @CCAlliance Buddy system. The segment focused on @[user], and her story of being diagnosed w/ colon cancer at 50.	0.94
Personal stories	Individual stories concerning regular people, patients, advocates related to CRC and CRC screening	An artist, an advocate, he gave my black boys a superhero with whom to identify. He reminds us #ColonCancer screening isn't elegant but necessary and should start 5yrs earlier in Af Am patients #womenshealth #parenting #healthdisparities. [User], a staunch patient advocate and #colorectalcancer survivor, is out here raising much-needed awareness about a #cancer highly preventable in most, and all too common, esp in the Black community.	0.93
Mortality	Mention of the severity of CRC, mortality rates	March is Colorectal Cancer Awareness Month! Colon cancer is the second most common cancer among Indigenous people, and the second leading cause of cancer death. #GetBehindCRCScreening to help us end colon cancer in Indian Country! Unfortunately, colorectal cancer is becoming a killer of young Black men, which we can conquer only by talking openly about symptoms, the value of colonoscopy screening, and sharing our experiences.	0.9

XSL•FO
RenderX

Category	Definition	Examples of tweets	Krippendorf α
Incidence of CRC	Mention of the prevalence of CRC and statistics of its impact	#ColorectalCancer is more common in men than women and among those of African American descent. The rate of new cases of colorectal cancer was 38.2 per 100,000 men and women per year based on 2013-2017 cases, age adjusted. Learn more from our page on #CRC. [User] discusses the concerning trends of #ColorectalCancer rates in younger patients and the higher incidence and lower survival rates being seen in Black people. #[local cancer center] #local cancer center].	0.89
Screening prevalence	Mention of the general prevalence of CRC screening or statistics about screening rates	Fewer than half of Native Americans over 50 are up-to-date with #ColorectalCancer screening. Learn how CDC is working with @[government agency] and @[government agency] to help. Due to COVID-19, colonoscopy screening for colon cancer among minorities declined nearly 90%. We are going #BlueForCRC to raise awareness and encourage preventative screenings.	0.89
Screening information	Providing details about screening options, or screening logistics	NEW ACG Clinical Guidelines on CRC Screening! info: -screen @ 45 (avg-risk adults) -recs if family Hx -qual indicators (ADR & withdrawal) -recs on aspirin for risk -1 versus 2-step tests -interventions to increase screening, esp. among African Americans “Mailed stool blood tests; followup phone calls are examples of effective strategies that increased #colorectalcancer screening among African Americans” @FightCRC	0.87
Benefits of screening	Using reasons to appeal to why screening is necessary, worth the effort, and beneficial	Getting routine screenings is imperative, regardless of family history. It’s worth it to have peace of mind that you don’t have colorectal cancer. #blackcommunity Colorectal cancer screening can detect cancer early—when it is most curable. Learn more about your screening options! #Latinxin-Medicine	0.86

Volume of Exposure to CRC Equity Tweets

We then examined the extent to which the followers of the 798 accounts were exposed to CRC equity tweets. Different from the analysis of content, which deals with unique tweets or unique equity tweets only (n=1134), the analysis of exposure deals with nonunique equity tweets (n=1333) to adjust for the cumulative effect of exposure to both unique tweets and retweets. As the exact mechanisms behind Twitter’s algorithms to display specific content to a user at a given time are unknown, we relied on Twitter’s public documentation [33], assuming that by subscribing to an account, a user could be exposed to all posts and updates from that account. Based on this logic, for each of the 6,266,269 unique followers of the 798 CRC potential disseminator accounts, we calculated the number of the accounts they followed, as well as the number of (nonunique) general CRC tweets and CRC equity tweets they could have been exposed to (which is equal to the number of the general CRC tweets and CRC equity tweets posted by the accounts they followed). Finally, we estimated the volume of exposure to CRC equity tweets among these followers overall and by account types (ie, whether the followed account is an expert, or a broker).

Ethical Considerations

All analyses are based on publicly available data. To respect privacy, we do not disclose usernames and identifiable information in our paper and only report aggregate results.

Results

Volume and Content of CRC Racial Equity Tweets

We begin by analyzing the content produced by the 798 potential equity disseminator accounts. Of the 19,559 (unique) CRC tweets these accounts produced, only 5.8% (n=1134) mentioned a historically minoritized racial group and thus were deemed likely to be specifically about CRC racial equity, disparity, or racial identity–specific impact/information (although not all of them necessarily discussed racial information in ways that signaled fairness and justice).

Table 3 provides a summary of the tweet categories’ respective volumes and proportions. Notably, outcome disparity (ie, specifying disparities in CRC outcomes between racial and ethnic groups) was the most common type of content, appearing in 56.53% (641/1134) of unique CRC racial equity tweets. Other common content categories were a call to action (ie, encouraging CRC detection/prevention; (505/1134, 44.53%), CRC risks or risk factors (351/1134, 30.95%), and raising awareness of CRC

XSL•FO
RenderX

(ie, emphasizing the need to communicate more about CRC; 337/1134, 29.72%). In contrast, details about CRC screening logistics or options (159/1134, 14.02%), benefits (98/1134, 8.64%), prevalence of screening (66/1134, 5.82%), and CRC symptoms (11/1134, 0.97%) were less common, though still present.

We next investigated which types of accounts tended to post different kinds of messages. Experts were more likely than nonexperts to send tweets about CRC racial equity, (z=14.88, P<.001). Specifically, of the 8210 unique CRC tweets produced by experts, 8.7% (n=716) of them mentioned at least 1 racially minoritized group. Nonexpert accounts, by contrast, only mentioned a racially minoritized group in 3.7% (418/11,349) of unique tweets they produced. There were also differences between expert and nonexpert accounts in their proportions of tweets for specific content subcategories. However, the only statistically significant difference is that experts focused more on celebrities' stories than nonexperts (P=.04; Table 3).

Table 3. Volume and proportion of colorectal cancer (CRC) racial equity tweets by content and account type. Proportions of specific content subcategories are compared between expert versus nonexpert accounts, brokers versus nonbrokers, and expert-brokers versus expert-nonbrokers.

Content type	Number of unique racial equity tweets from CRC equity disseminators (n=1134 tweets), n (%)	Number of unique racial equity tweets from experts (n=716 tweets), n (%)	Number of unique racial equity tweets from nonexperts (n=418 tweets), n (%)	Number of unique racial equity tweets from brokers (n=518 tweets), n (%)	Number of unique racial equity tweets from nonbrokers (n=616 tweets), n (%)	Number of unique racial equity tweets from expert-nonbrokers (n=556 tweets), n (%)	Number of unique racial equity tweets from expert-brokers (n=160 tweets), n (%)
Outcome disparity	641 (56.53)	412 (57.54)	229 (54.78)	282 (54.44)	360 (58.44)	322 (57.91)	90 (56.25)
Call to action	505 (44.53)	312 (44.83)	184 (44.02)	229 (44.21)	276 (44.80)	252 (45.32)	69 (43.13)
Risk factors	351 (30.95)	227 (31.70)	124 (29.67	168 (32.43)	183 (29.70	160 (28.78)[a]	67 (41.88)[a]
Raising awareness	337 (29.72)	201 (28.07)	136 (32.54)	162 (31.27)	175 (28.41)	154 (27.70)	47 (29.38)
Advocacy	263 (23.19)	171 (22.91)	99 (23.68)	124 (23.94)	139 (22.56)	129 (23.20)	35 (21.88)
Celebrities' stories	245 (21.60)	164 (23.88)[a]	74 (17.70)[a]	110 (21.24)	135 (21.92)	123 (22.12)[a]	48 (30.00)[a]
Mortality or death	194 (17.11)	130 (18.16)	64 (15.31)	83 (16.02)	111 (18.02)	97 (17.45)	33 (20.63)
Incidence of CRC	160 (14.11)	116 (13.97)	60 (14.35)	74 (14.29)	86 (13.96)	72 (12.95)	28 (17.50)
Screening information	159 (14.02)	100 (16.20)	43 (10.29)	70 (13.51)	89 (14.44)	80 (14.39)[a]	36 (22.50)[a]
Personal stories	125 (11.02)	75 (10.47)	50 (11.96)	69 (13.32)[a]	56 (9.09)[a]	54 (9.71)	21 (13.13)
Benefits of screening	98 (8.64)	61 (8.52)	37 (8.85)	52 (10.04)	46 (7.47)	42 (7.55)	19 (11.88)
Screening prevalence	66 (5.82)	39 (5.45)	27 (6.46)	27 (5.21)	39 (6.33)	34 (6.12)	5 (3.13)
Symptoms	11 (0.97)	9 (1.26)	2 (0.48)	4 (0.77)	7 (1.14)	7 (1.26)	2 (1.25)

[a]Values different at P<.05 level.

We also compared brokers to nonbrokers. Recall the definition of brokers as a subset of CRC equity disseminator accounts where more than half of their followers did not follow any experts (or any experts other than itself, if the account in focus was an expert account). Overall, brokers were less likely to talk about equity (518 tweets) than nonbrokers (616 tweets).

Examining the volume of equity tweets by expert brokers versus expert nonbrokers, we found that expert brokers produced fewer equity tweets (160 tweets) compared with expert nonbrokers (556 tweets). An examination at the account level revealed that only 30% (77/254) of experts who are brokers tweeted at least once about equity, compared with 46% (103/225) of expert-nonbrokers who did so. This pattern indicates that those who talk about equity more frequently are those with less reach beyond the community and who would otherwise not learn about this topic.

Exposure to CRC Racial Equity Tweets

We now turn to the question of how different types of accounts impact the kind of information that their followers are exposed to. In other words, which of the 798 potential CRC equity disseminators do the "best" job of exposing their followers to equity content. Since both tweets and retweets (not just unique tweets) contribute to exposure, the outcome of interest is the estimated nonunique equity tweets to which the followers of the 798 accounts (n=6,266,269) can potentially be exposed (mean 0.99, SD 3.89). Linear regression results showed that the number of accounts followed was positively associated with the number of racial equity tweets potentially seen (β=0.06, SE 0.000; P<.001). However, since only 309 (38.7%) out of the 798 potential disseminators posted any equity-oriented content, the estimated exposure for more than half (3,490,864/6,266,269, 55.7%) of all followers would be zero CRC racial equity tweets.

XSL•FO
RenderX

To determine which potential disseminator account has the most unique reach into the communities who would otherwise not learn about the topic, we evaluated the portion (%) of each account's followers for whom they are the only source of racial equity content. In other words, for what portion of their followers does each account category tend to be the only source? The results showed that brokers were substantially more likely to serve that role. For brokers, the average proportion of followers for whom they were the only source of equity tweets is 30%, with the median being 27% (IQR 17%-39%). By contrast, for nonbrokers, the average percentage is 11%, with the median being 10% (IQR 7.4%-14.5%). A linear regression was run using the percentage of followers for whom the source is the unique source as the dependent variable. Brokers, compared with nonbrokers, had a significantly higher proportion of their followers subscribing to them as unique sources of equity content (β=18.9, *P*<.001). These figures do not differ substantially between expert versus nonexperts (β=–0.5, *P*=.90), and including expert as an interaction term does not produce a difference (β=1.1, *P*=.80). The results indicate that broker accounts (regardless of expert or nonexpert status) who tweeted about equity had a unique potential, since they were the only source of this information for a much larger portion of their followers compared with nonbrokers who did so.

Discussion

Principal Findings

This study examined CRC content in the context of Twitter, focusing on racial equity–related discussions from CRC equity disseminator accounts on the platform. We analyzed the volume of CRC racial equity tweets overall and by specific content categories (eg, outcome disparity, screening, and narratives) and account types (experts vs nonexperts, brokers vs nonbrokers), and we identified the kinds of accounts that were most likely to uniquely expose their followers to racial equity related information. Findings provide insights into promoting cancer equity through health communication focusing on the role of brokers on social media.

Regarding the volume of CRC racial equity tweets, race and ethnicity was infrequently discussed on Twitter even among potential disseminators (accounting for only 5.8% of the unique tweets). This finding is consistent with previous research on health news stories from newspapers and local TV programs, which also found scant mentions of race or ethnicity [8,9]. Without featuring race and ethnicity–specific health information, people may not pay attention to CRC-related racial inequities. A lack of content tailored to racially and ethnically minoritized groups is concerning in light of the fact that tailored communication could better motivate individual and collective action [14].

While race and ethnicity is not a frequent feature overall in tweets from potential equity disseminators, outcome disparity is the most common content type when race and ethnicity is depicted. Emphasizing the differences in CRC risks and outcomes between racial groups may increase people's awareness of the disparity gap. However, research also found that when racial health disparity information was presented alone, it might have an adverse impact on the group with higher disease risks (the disadvantaged group), reducing their intention to engage in health behaviors [11-13]. This concern is intensified as the least frequent types of content in racial equity tweets are about strategies that could address CRC outcomes, including screening procedures, prevalence, and benefits. Featuring comparatively high disease risks without also highlighting detailed information about what can be done, how to do it, and the effectiveness of the solutions (eg, CRC screening) could trigger message resistance and behavioral inhibition [12]. Screening prevalence among Black Americans emphasizing progress toward addressing inequities may better motivate screening behavior compared with disparity information in and of itself [13]. The benefits of screening content could also convey the effectiveness of screening, which may serve as behavioral evaluation and response efficacy information both of which are important behavioral antecedents [34,35]. In addition, past research has found that Black Americans were more likely to come across screening-specific information via the web compared with White Americans, which was positively associated with screening behavior [11]. Thus, CRC racial equity tweets may benefit from including more screening-specific information.

However, the overall impact of these tweets may be less than one might assume. Our analysis showed that in line with our concern about the potential for CRC equity messages to reach new audiences, the accounts that talked the most about equity—experts—were less likely to reach unique audiences who would otherwise not learn about this topic. Consistent with network theory, brokers who tweeted about equity were more likely to serve that role since they were the only source of such information for a major portion of their followers. In essence, while experts were more vocal, the more insular nature of their following and networks may make them less capable as disseminators. This suggests that experts or others concerned about the dissemination of equity content might consider allying with accounts that have a greater or unique reach, identifying messages that these far-reaching broker accounts are likely to retweet, rather than simply producing more messages themselves.

Limitations

The study has several limitations. We identified potential disseminator accounts based on the 7 major CRC racial equity organization accounts with the most followers on Twitter (1500 or more). However, they represent only a part of the many equity-oriented organizations in the United States, and future research may include other related equity organizations with fewer followers on Twitter. Similarly, we only focused on the tweets from these accounts even though other Twitter accounts that are were not in this sample may also post CRC racial equity content. Thus, a user may be exposed to racial equity content, despite not following any of our potential disseminator accounts. Moreover, by focusing on only account followers, our measure of exposure to racial equity tweets might be a more conservative one. Future research could consider other measures of diffusion beyond follower-following relationship, such as retweet networks, for more complex exposure estimation. This paper also did not examine the actual audiences (eg, demographics of

XSL•FO
RenderX

different followers) and the impact of racial equity information diffusion on the audiences of such tweets. If members of historically minoritized racial groups do not see the tweets, this can undoubtedly mitigate some of the impact of such communication. Exploring this important research question will be crucial in gaining a more comprehensive understanding of how sharing such messages can lead to desirable outcomes. Last but not least, the official change of Twitter to X in July 2023, despite being irrelevant to the period of this study, means that future research might not be able to study this topic or other health topics on Twitter in the same way due to issues such as limited data access, shifts in user demographics, and changes in technical affordances. With the increasing prominence of platforms featuring long- and short-form video formats such as YouTube or TikTok [36], more research is needed to understand the availability and exposure to racial equity information on these platforms.

Conclusions

To conclude, the findings from this study highlight the importance of social media accounts that are in a position to diffuse racial equity information produced by equity organizations to otherwise disconnected audiences. Public health officials should encourage these accounts to post more information that focuses on CRC racial equity–related content and tailor their information to the needs of specific racial groups. In doing so, they should emphasize CRC symptoms and details about the screening guidelines/procedures, prevalence, and benefits in their racial equity posts to center and amplify information about CRC detection and prevention.

Acknowledgments

This work was supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under awards R37CA259156 and P30CA040214. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Colorectal cancer statistics. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/cancer/colorectal/statistics/index.htm [accessed 2022-01-05]
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin 2022;72(1):7-33 [FREE Full text] [doi: 10.3322/caac.21708] [Medline: 35020204]
3. Agunwamba AA, Zhu X, Sauver JS, Thompson G, Helmueller L, Finney Rutten LJF. Barriers and facilitators of colorectal cancer screening using the 5As framework: a systematic review of US studies. Prev Med Rep 2023;35:102353 [FREE Full text] [doi: 10.1016/j.pmedr.2023.102353] [Medline: 37576848]
4. Liu J, King AJ, Margolin D, Niederdeppe J. Information seeking and scanning about colorectal cancer screening among black and white Americans, ages 45-74: comparing information sources and screening behaviors. J Health Commun 2020;25(5):402-411. [doi: 10.1080/10810730.2020.1776424] [Medline: 32529955]
5. Cram P, Fendrick AM, Inadomi J, Cowen ME, Carpenter D, Vijan S. The impact of a celebrity promotional campaign on the use of colon cancer screening: the katie couric effect. Arch Intern Med 2003;163(13):1601-1605. [doi: 10.1001/archinte.163.13.1601] [Medline: 12860585]
6. Vos SC, Sutton J, Gibson CB, Butts CT. Celebrity cancer on Twitter: mapping a novel opportunity for cancer prevention. Cancer Control 2019;26(1):1073274819825826 [FREE Full text] [doi: 10.1177/1073274819825826] [Medline: 30816059]
7. Francis DB, Zelaya CM. Cancer fatalism and cancer information seeking among black women: examining the impact of aretha franklin's death on cancer communication outcomes. J Cancer Educ 2021;36(4):763-768. [doi: 10.1007/s13187-020-01701-9] [Medline: 32020521]
8. Kim AE, Kumanyika S, Shive D, Igweatu U, Kim S. Coverage and framing of racial and ethnic health disparities in US newspapers, 1996–2005. Am J Public Health 2010;100(S1):S224-S231. [doi: 10.2105/ajph.2009.171678]
9. Xu Y, Farkouh EK, Dunetz CA, Varanasi SL, Mathews S, Gollust SE, et al. Local TV news coverage of racial disparities in COVID-19 during the first wave of the pandemic, March-June 2020. Race Soc Probl 2023;15(2):201-213 [FREE Full text] [doi: 10.1007/s12552-022-09372-5] [Medline: 35855105]
10. Liu J, Niederdeppe J. Effects of communicating health disparities using social comparison framing: a comprehensive review. Soc Sci Med 2024;348:116808 [FREE Full text] [doi: 10.1016/j.socscimed.2024.116808] [Medline: 38537451]
11. Liu J, Niederdeppe J. Effects of race-specific prevalence and racial disparity information about obesity and diabetes. Health Commun 2024;39(7):1269-1278. [doi: 10.1080/10410236.2023.2210384] [Medline: 37157159]

XSL•FO
RenderX

12. Liu J, Niederdeppe J. Effects of communicating health disparities using social comparison framing: a comprehensive review. Soc Sci Med 2024;348:116808 [FREE Full text] [doi: 10.1016/j.socscimed.2024.116808] [Medline: 38537451]
13. Nicholson RA, Kreuter MW, Lapka C, Wellborn R, Clark EM, Sanders-Thompson V, et al. Unintended effects of emphasizing disparities in cancer communication to African-Americans. Cancer Epidemiol Biomarkers Prev 2008;17(11):2946-2953 [FREE Full text] [doi: 10.1158/1055-9965.EPI-08-0101] [Medline: 18990735]
14. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]
15. Pellino G, Simillis C, Qiu S, Rasheed S, Mills S, Warren O, et al. Social media and colorectal cancer: a systematic review of available resources. PLoS One 2017;12(8):e0183031 [FREE Full text] [doi: 10.1371/journal.pone.0183031] [Medline: 28832603]
16. Liu J, Niederdeppe J, Tong C, Margolin D, Chunara R, Smith T, et al. Associations between news coverage, social media discussions, and search trends about celebrity deaths, screening, and other colorectal cancer-related events. Prev Med 2024;185:108022. [doi: 10.1016/j.ypmed.2024.108022] [Medline: 38823651]
17. Brar J, Ferdous M, Abedin T, Turin TC. Online information for colorectal cancer screening: a content analysis of YouTube videos. J Cancer Educ 2021;36(4):826-831. [doi: 10.1007/s13187-020-01710-8] [Medline: 32072485]
18. Selvan B, Tran M, OConnell CB, Wilder J. What are TikTok users saying about colorectal cancer? an examination of content, quality, and emerging themes. Am J Gastroenterol 2023;118(10S):S192. [doi: 10.14309/01.ajg.0000950664.49054.6c]
19. Park S, Oh H, Park G, Suh B, Bae WK, Kim JW, et al. The source and credibility of colorectal cancer information on Twitter. Medicine (Baltimore) 2016;95(7):e2775 [FREE Full text] [doi: 10.1097/MD.0000000000002775] [Medline: 26886625]
20. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging social media to promote public health knowledge: example of cancer awareness via Twitter. JMIR Public Health Surveill 2016;2(1):e17. [doi: 10.2196/publichealth.5205] [Medline: 27227152]
21. Kreps GL. Disseminating relevant health information to underserved audiences: implications of the digital divide pilot projects. J Med Libr Assoc 2005;93(4 Suppl):S68-S73 [FREE Full text] [Medline: 16239960]
22. van der Heide I, Uiters E, Jantine Schuit A, Rademakers J, Fransen M. Health literacy and informed decision making regarding colorectal cancer screening: a systematic review. Eur J Public Health 2015;25(4):575-582. [doi: 10.1093/eurpub/ckv005] [Medline: 25733553]
23. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
24. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Sociol 2001;27(1):415-444. [doi: 10.1146/annurev.soc.27.1.415]
25. Monge PR, Contractor NS. Theories of Communication Networks. Oxford, United Kingdom: Oxford University Press; 2003.
26. Burt RS. Brokerage and Closure: An Introduction to Social Capital. Oxford, United Kingdom: Oxford University Press; 2005.
27. Long JC, Cunningham FC, Braithwaite J. Bridges, brokers and boundary spanners in collaborative networks: a systematic review. BMC Health Serv Res 2013;13:158 [FREE Full text] [doi: 10.1186/1472-6963-13-158] [Medline: 23631517]
28. Minority health resource guide on national minority organizations. US Department of Health and Human Services. URL: https://minorityhealth.hhs.gov/resource-guide [accessed 2022-01-01]
29. Xu WW, Chiu IH, Chen Y, Mukherjee T. Twitter hashtags for health: applying network and content analyses to understand the health knowledge sharing in a Twitter-based community of practice. Qual Quant 2014;49(4):1361-1380. [doi: 10.1007/s11135-014-0051-6]
30. Strayhorn SM, Carter A, Harmon BE, Hébert JR. An examination of culturally relevant health messages in African-American churches. J Relig Health 2023;62(4):2547-2562. [doi: 10.1007/s10943-022-01638-x] [Medline: 35994186]
31. Kvasny L, Igwe CF. An African American weblog community's reading of AIDS in black America. J Comput Mediat Commun 2008;13(3):592. [doi: 10.1111/j.1083-6101.2008.00411.x]
32. Bruckman A, Luther K, Fiesler C. When should we use real names in published accounts of internet research? In: Hargittai E, Sandvig C, editors. Digital Research Confidential. Cambridge, MA: MIT Press; 2015.
33. Following FAQs. X Help Center. URL: https://help.twitter.com/en/using-twitter/following-faqs [accessed 2022-03-05]
34. Champion VL, Skinner CS. The health belief model. In: Health Behavior and Health Education: Theory, Research, and Practice. 4th ed. San Francisco, CA: Jossey-Bass; 2008.
35. Popova L. The extended parallel process model: illuminating the gaps in research. Health Educ Behav 2012;39(4):455-473. [doi: 10.1177/1090198111418108] [Medline: 22002250]
36. Tong C, Margolin D, Chunara R, Niederdeppe J, Taylor T, Dunbar N, et al. Search term identification methods for computational health communication: word embedding and network approach for health content on YouTube. JMIR Med Inform 2022;10(8):e37862 [FREE Full text] [doi: 10.2196/37862] [Medline: 36040760]

XSL•FO
RenderX

Abbreviations

CRC: colorectal cancer

Edited by A Mavragani; submitted 08.07.24; peer-reviewed by JGZ Rodriguez, S Mai; comments to author 04.11.24; revised version received 18.11.24; accepted 18.12.24; published 03.02.25.

Please cite as:
Tong C, Margolin D, Niederdeppe J, Chunara R, Liu J, Jih-Vieira L, King AJ
Colorectal Cancer Racial Equity Post Volume, Content, and Exposure: Observational Study Using Twitter Data
J Med Internet Res 2025;27:e63864
URL: https://www.jmir.org/2025/1/e63864
doi:10.2196/63864
PMID:

©Chau Tong, Drew Margolin, Jeff Niederdeppe, Rumi Chunara, Jiawei Liu, Lea Jih-Vieira, Andy J King. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perspectives of Hispanic and Latinx Community Members on AI-Enabled mHealth Tools: Qualitative Focus Group Study

Stephanie A Kraft[1], JD; Shaan Chopra[2], MS; Miriana C Duran[3], MPH, MD; Janet A Rojina[3], MPH; Abril Beretta[4], BS; Katherine I López[3], BA; Russell Javan[3], MPH; Benjamin S Wilfond[4,5], MD; Margaret Rosenfeld[5,6], MPH, MD; James Fogarty[2], PhD; Linda K Ko[3], PhD

[1]Department of Bioethics and Decision Sciences, Geisinger College of Health Sciences, Danville, PA, United States

[2]Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, United States

[3]Department of Health Systems and Population Health, University of Washington School of Public Health, Seattle, WA, United States

[4]Treuman Katz Center for Pediatric Bioethics and Palliative Care, Seattle Children's Research Institute, Seattle, WA, United States

[5]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, United States

[6]Center for Respiratory Biology and Therapeutics, Seattle Children's Research Institute, Seattle, WA, United States

Corresponding Author:
Stephanie A Kraft, JD
Department of Bioethics and Decision Sciences
Geisinger College of Health Sciences
100 N. Academy Ave.
Danville, PA, 17822
United States
Phone: 1 5702140506
Email: skraft1@geisinger.edu

Abstract

Background: Mobile health (mHealth) tools have the potential to reduce the burden of chronic conditions that disproportionately affect Hispanic and Latinx communities; however, digital divides in the access to and use of health technology suggest that mHealth has the potential to exacerbate, rather than reduce, these disparities.

Objective: A key step toward developing health technology that is accessible and usable is to understand community member perspectives and needs so that technology is culturally relevant and appropriately contextualized. In this study, we aimed to examine the perspectives of Hispanic and Latinx community members in Washington State about mHealth.

Methods: We recruited English- and Spanish-speaking Hispanic or Latinx adults to participate in web-based focus groups through existing community-based networks across rural and urban regions of Washington State. Focus groups included a presentation of narrative slideshow materials developed by the research team depicting mHealth use case examples of asthma in children and fall risk in older adults. Focus group questions asked participants to respond to the case examples and to further explore mHealth use preferences, benefits, barriers, and concerns. Focus group recordings were professionally transcribed, and Spanish transcripts were translated into English. We developed a qualitative codebook using deductive and inductive methods and then coded deidentified transcripts using the constant comparison method. The analysis team proposed themes based on review of coded data, which were validated through member checking with a community advisory board serving Latino individuals in the region and finalized through discussion with the entire research team.

Results: Between May and September 2023, we conducted 8 focus groups in English or Spanish with 48 participants. Focus groups were stratified by language and region and included the following: 3 (n=18, 38% participants) Spanish urban groups, 2 (n=14, 29% participants) Spanish rural groups, 1 (n=6, 13% participants) English urban group, and 2 (n=10, 21% participants) English rural groups. We identified the following seven themes: (1) mHealth is seen as beneficial for promoting health and peace of mind; (2) some are unaware of, unfamiliar with, or uncomfortable with technology and may benefit from individualized support; (3) financial barriers limit access to mHealth; (4) practical considerations create barriers to using mHealth in daily life; (5) mHealth raises concern for overreliance on technology; (6) automated mHealth features are perceived as valuable but fallible, requiring human input to ensure accuracy; and (7) data sharing is seen as valuable for limited uses but raises privacy concerns. These themes illustrate key barriers to the benefits of mHealth that communities may face, provide insights into the role of mHealth within families, and examine the appropriate balance of data sharing and privacy protections.

XSL•FO
RenderX

Conclusions: These findings offer important insights that can help advance the development of mHealth that responds to community values and priorities.

(J Med Internet Res 2025;27:e59817) doi:10.2196/59817

KEYWORDS

wearable electronic devices; qualitative research; mobile health; mHealth; digital health; privacy; data sharing; artificial intelligence; AI; community; chronic conditions; chronic disease

Introduction

Background

Mobile health (mHealth) tools, broadly defined to include patient-facing, health-related mobile or wireless devices, such as sensors, trackers, and wearables [1,2], offer the potential to improve health care access and encourage health-promoting behaviors through personal health monitoring. In particular, mHealth powered by artificial intelligence to make actionable, personalized predictions is a growing field that aims to advance broad-reaching, patient-driven preventive care [3]. For example, recent work has demonstrated the efficacy of an artificial intelligence–aided home stethoscope in detecting asthma exacerbations in children, which can facilitate time-sensitive decisions about management [4]. Others are using mHealth monitoring devices, such as smart inhalers and smartwatches, to collect personal and environmental data to facilitate the development of predictive models for asthma attacks [5,6]. Mobile data collection and device availability beyond clinical settings may be particularly valuable for overcoming geographic, financial, and other structural barriers to health care that disproportionately affect marginalized and minoritized groups, including Hispanic and Latinx communities [7].

mHealth tools have the potential to reduce disease burden and have been shown to be especially well suited for managing chronic health conditions, such as asthma, and particularly pediatric asthma [8-11], by facilitating continuous monitoring [12]. In the United States, pediatric asthma disproportionately affects Hispanic and Latinx children. An analysis of health record data from community health centers across 18 states showed Spanish-preferring Latinx children have a higher likelihood of clinic visits for asthma exacerbations [13], and overall, Hispanic children have a 40% higher death rate from asthma compared to non-Hispanic White children [14]. Existing mHealth tools for pediatric asthma monitor a wide range of factors, including air quality, lung function, physical activity, sleep, and cough [15]. Hispanic individuals in the United States own smartphones at similar rates to other racial and ethnic groups and are more likely than White individuals to use a smartphone as their primary means of internet access [16]. However, at least 1 recent cross-sectional study has shown that Hispanic and Latinx families are significantly less likely to have reliable, high-speed internet as compared to White families [17], illustrating the persistence of the digital divide that imposes disproportionate barriers to accessing digital technologies such as mHealth. For example, an evaluation of remote digital studies found that minoritized groups, including Hispanic and Latinx individuals, were overrepresented in clusters with lower app use, suggesting a need for improved engagement approaches [18]. The digital divide may be even greater among those living in rural areas [16]. These patterns suggest that digital technologies may replicate and exacerbate existing disparities if devices are not implemented in a manner that is culturally appropriate and responsive to individuals' social determinants and lived realities [19-22]. Cultural appropriateness is essential to ensure a tool is acceptable and effective for its target population and can work toward maximizing meaningful benefits and minimizing important risks [23,24]. Likewise, sensitivity to how social and economic structures differentially shape populations' access to and use of health technology is necessary to support fairness when implementing devices into real-world settings [24]. A 2018 systematic literature review found that SMS text message–based mHealth tools were consistently studied among marginalized populations; however, there were gaps in studies of other mHealth strategies as well as their real-world implementation [25].

This Study

As a step toward informing health technology that is responsive to individuals' sociocultural contexts, we conducted an in-depth focus group study to provide formative research on the values and lived experiences of Hispanic and Latinx community members in Washington State as they relate to mHealth. We used focus group methods, incorporating relevant mHealth case examples, to allow for group discussion and deliberation [26]. Our specific aim was to describe the perspectives of Hispanic and Latinx community members in rural and urban regions of Washington State about mHealth.

Methods

Participants

We recruited participants through community-based networks in urban western Washington and rural central Washington in several ways. First, we contacted people who had signed up to be contacted about research opportunities on the Community Voices registry, a longstanding partnership with Hispanic communities in central Washington led by a study team member (LKK). Second, we contacted previous participants in other studies who had agreed to be contacted for future research. Third, we shared recruitment flyers with networks serving local Hispanic and Latinx communities.

Individuals were eligible to participate if they were aged ≥18 years, spoke either English or Spanish, and identified as Hispanic, Latino, Latina, or Latinx. All study materials were offered in both English and Spanish, and bilingual study team members were available for all participant interactions [27]. Participants completed a brief demographic questionnaire.

XSL•FO
RenderX

Focus Groups

We developed a focus group guide (Multimedia Appendix 1) to explore key topics related to health technology use, including potential patient-level barriers and concerns as well as relevant mHealth design decisions, drawing on our multidisciplinary team's expertise in these topics. Our approach was also informed, in part, by the lived informatics model, which delineates a process through which people decide to use, experience using, and may ultimately abandon technology in real-world settings [28].

To ensure all participants had a shared understanding to foster robust deliberation [29], we created multimedia informational aids to present during the focus groups. We collaborated with Booster Shot Media, a private health-focused communication company comprised of a health communication expert and a pediatrician who is also a cartoonist with over a decade of experience developing multimedia tools for use in patient-centered research, to develop a series of illustrated narrative slideshows depicting Latinx families using mHealth devices. Our team first identified several clinical scenarios that would reflect realistic near-future uses of mHealth and would also be relevant to focus group participants' lived experiences. These scenarios were based on our collective knowledge of the communities where we planned to recruit and of mHealth development. Through iterative team discussion, we identified pediatric asthma as a paradigmatic example of a condition that disparately impacts Hispanic and Latinx populations in the United States, which is common in the region where we conducted our study and has several distinct opportunities for mHealth intervention. We also identified a second example case of fall risk in older adults as similarly amenable to mHealth intervention and relevant to this population. While we used these 2 example cases to provide tangible, relevant starting points for discussion, conversations were not limited to these clinical cases; instead, our discussion questions built on and expanded from these cases.

We developed the narratives of our slideshows by drawing on team members' personal and professional experiences with Latinx families as well as professional experiences with mHealth and pediatric asthma, and we worked with Booster Shot Media to finalize and illustrate the narratives. Booster Shot Media reviewed photos of Latinx individuals living in the regions where we recruited participants to create culturally relevant illustrations, which were refined through iterative review and feedback from team members with direct experience with the communities where we recruited. We also sought feedback on the slideshows from bilingual and bicultural colleagues outside of our study team. The final versions of the slideshows first presented concrete examples of how mHealth could be used to predict asthma exacerbations and manage asthma in children, how different devices might work, and how data might be shared. They described these examples through a family-based narrative showing parents and children working together to manage pediatric asthma. The second example showed an older adult, the family's "abuelita" (grandmother), using a wearable device to predict fall risk, which was a briefer narrative intended to explore potential differences in perspectives based on the individual using the device and the condition being tracked. These examples described how the depicted mHealth tools could make personalized predictions and improve those predictions based on collected data, but we did not explicitly use the term "artificial intelligence" to avoid raising potentially preconceived notions about artificial intelligence as depicted in the broader media. Example slideshow images are shown in Figure 1. The full slide decks and text are included in Multimedia Appendix 2 (English version) and Multimedia Appendix 3 (Spanish version).

Figure 1. Example narrative slideshow of families using mobile health to control pediatric asthma and monitor fall risk in an older adult. The slideshows were developed in collaboration with Booster Shot Media.

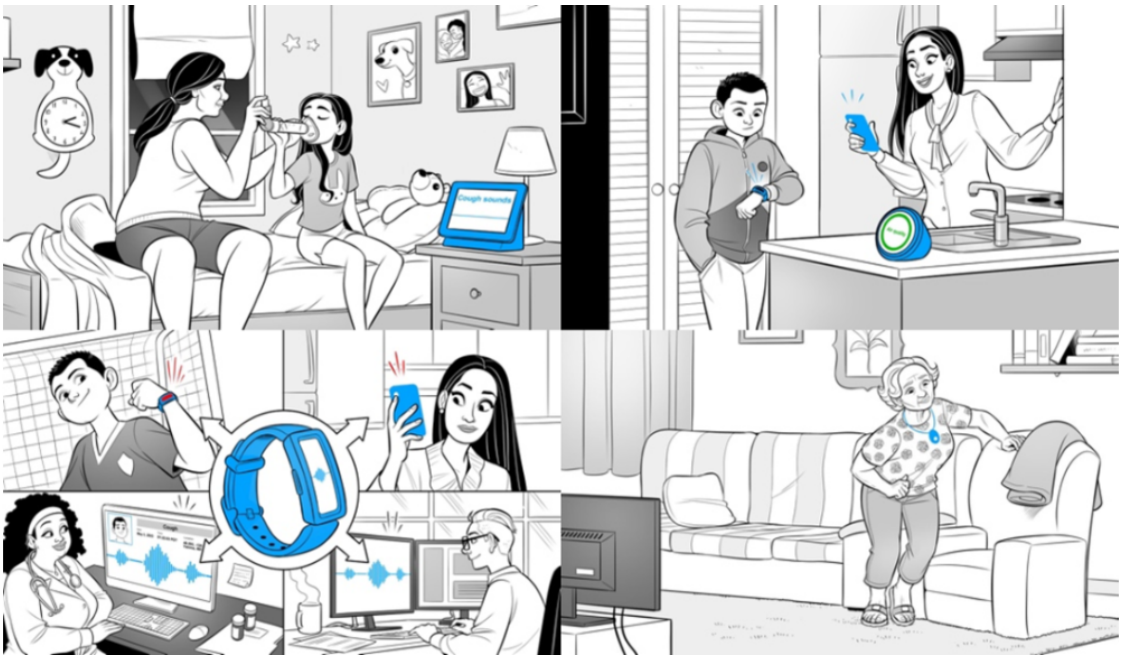

XSL•FO
RenderX

Focus groups were held on Zoom (Zoom Video Communications) and moderated by at least 2 members of the research team, with at least 1 additional team member present for note-taking and technology support. Native speakers in the research team facilitated the Spanish focus groups. Audio recordings were professionally transcribed verbatim, and Spanish focus groups were professionally translated into English. Transcripts were reviewed by the research team for accuracy of transcription and translation, deidentified, and uploaded to the cloud-based platform Dedoose (SocioCultural Research Consultants, LLC) [30] for qualitative analysis management.

We developed an initial codebook based on a deductive review of the focus group moderator guide topic areas. Two team members (MCD and SC) reviewed 2 focus group transcripts, revised the initial codebook using inductively derived codes from the transcripts, and coded those transcripts using the constant comparison method [31]. They continued this iterative process with additional transcripts until the codebook needed no further revision. A third coder (AB) was then trained. All transcripts were coded by the coding team and reviewed for consensus.

Coded excerpts were extracted for review and identification of themes [32,33]. We divided excerpts for review and summarization following the codebook structure, as our initial review suggested that themes aligned with the previously identified codes. This step allowed us to identify key themes and condense the data to facilitate reporting. Two team members (MCD and SC) summarized a portion of the excerpts and proposed themes, and to ensure credibility, a third member (SAK) independently summarized and proposed themes from all excerpts. Proposed themes were compared and discussed among the analysis team, which included a community-based researcher, a computer scientist, and a bioethicist. Our coding team additionally included a bioethics research assistant, and 2 of 3 coders were bilingual in English and Spanish. A summary of themes was shared with a community advisory board serving Latino individuals in central Washington for member checking, which validated and identified important nuances in our findings [34]. We then revised and finalized our themes through discussion with the entire research team.

Ethical Considerations

This study was approved as exempt by the institutional review boards at the University of Washington (STUDY00016183) and Seattle Children's Hospital (STUDY00004107). All participants reviewed an information sheet with a study team member in their preferred language and provided verbal consent. All transcripts were deidentified before analysis, and only deidentified quotes were reported. Participants received a US $50 gift card following the focus group.

Results

Participants

We conducted 8 focus groups: 5 in Spanish, 3 in English, and 4 each with rural and urban participants. These focus groups included 48 participants overall, with an average of 6 (SD 2) participants per focus group and a range of 3 to 9. Participants reported a mean age of 45 (SD 14) years, and most (43/48, 90%) were women. Most (38/48, 79%) indicated they were born outside the United States, including nearly all (31/32, 97%) participants in the Spanish focus groups and half (5/10, 50%) of participants in the rural English focus groups. Most (29/48, 60%) participants reported having either no health insurance or state-funded health insurance, and a majority (31/48, 65%) reported no formal post–high school education. Detailed participant characteristics, including self-reported experience with mHealth and asthma, are shown in Table 1.

XSL•FO
RenderX

Table 1. Participant characteristics (N=48).

	Spanish, urban (n=18)	Spanish, rural (n=14)	English, urban (n=6)	English, rural (n=10)	Total (N=48)
Age (y), mean (SD)	54 (13)	42 (11)	33 (9)	39 (11)	45 (14)
Gender identity, n (%)					
Women	16 (89)	13 (93)	5 (83)	9 (90)	43 (90)
Men	2 (11)	1 (7)	1 (17)	1 (10)	5 (10)
Race or ethnicity (check all that apply), n (%)					
Native American or Alaska Native	1 (6)	0 (0)	0 (0)	0 (0)	1 (2)
Hispanic, Latino, or of Spanish origin	17[a] (94)	14 (100)	6 (100)	10 (100)	47[a] (98)
White	1[a] (6)	0 (0)	0 (0)	0 (0)	1[a] (2)
Birthplace, n (%)					
Outside the United States[b]	17 (94)	14 (100)	2 (33)	5 (50)	38 (79)
Educational attainment, n (%)					
Elementary school or lower	3 (17)	6 (43)	0 (0)	0 (0)	9 (19)
Some high school	3 (17)	1 (7)	0 (0)	1 (10)	5 (10)
High school graduate or GED[c]	9 (50)	5 (36)	1 (17)	2 (20)	17 (35)
Technical school diploma	1 (6)	0 (0)	0 (0)	0 (0)	1 (2)
Some college	0 (0)	0 (0)	0 (0)	2 (20)	2 (4)
College graduate	2 (11)	2 (14)	1 (17)	2 (20)	7 (15)
Graduate school degree	0 (0)	0 (0)	4 (67)	3 (30)	7 (15)
Health insurance (check all that apply), n (%)					
Employer-sponsored	3 (17)	4 (29)	3 (50)	6 (60)	16 (33)
Medicare	3[a] (17)	0 (0)	1 (17)	0 (0)	4[a] (8)
Medicaid, Washington Apple Care, or coupons	4[a] (22)	3 (21)	0 (0)	3 (30)	10[a] (21)
None	8 (44)	6 (43)	1 (17)	1 (10)	16 (33)
Other	1 (6)	0 (0)	1 (17)	0 (0)	2 (4)
Prefer not to answer	0 (0)	1 (7)	0 (0)	0 (0)	1 (2)
Annual household income (US $), n (%)					
<15,000	1 (6)	1 (7)	0 (0)	0 (0)	2 (4)
15,000 to <30,000	2 (11)	3 (21)	0 (0)	2 (20)	7 (15)
30,000 to <45,000	1 (6)	2 (14)	0 (0)	2 (20)	5 (10)
45,000 to <60,000	3 (17)	0 (0)	0 (0)	1 (10)	4 (8)
60,000 to <75,000	1 (6)	1 (7)	4 (67)	1 (10)	7 (15)
≥75,000	0 (0)	2 (14)	2 (33)	4 (40)	8 (17)
Do not know or prefer not to answer	10 (56)	5 (36)	0 (0)	0 (0)	15 (31)
Have you or someone you know been diagnosed with asthma? (check all that apply), n (%)					
Yes, I have been diagnosed with asthma	3 (17)	2 (14)	0 (0)	2[a] (20)	7[a] (15)
Yes, I have a family member or close friend with asthma	6 (33)	3 (21)	6[a] (100)	8[a] (80)	23[a] (48)
Yes, I know someone else with asthma	3 (17)	2 (14)	2[a] (33)	0 (0)	7 (15)
No	6 (33)	7 (50)	0 (0)	1 (10)	15 (31)
How comfortable are you with using mobile devices, for example, smartphones? n (%)					
Very comfortable	9 (50)	6 (43)	5 (83)	6 (60)	26 (54)

XSL•FO
RenderX

	Spanish, urban (n=18)	Spanish, rural (n=14)	English, urban (n=6)	English, rural (n=10)	Total (N=48)
Somewhat comfortable	5 (28)	6 (43)	1 (17)	3 (30)	15 (31)
Not very comfortable	2 (11)	1 (7)	0 (0)	1 (10)	4 (8)
Not at all comfortable	2 (11)	1 (7)	0 (0)	0 (0)	3 (6)
Have you ever used any technology to track your personal health or wellness, for example, apps or devices related to diet, sleep, exercise, or steps? n (%)					
Yes, I currently use technology to track my health or wellness	7 (39)	5 (36)	4 (67)	6 (60)	22 (46)
Yes, I have used technology to track my health or wellness in the past but do not use it currently	1 (6)	3 (21)	1 (17)	2 (20)	7 (15)
No, I have never used technology to track my health or wellness	10 (56)	6 (43)	1 (17)	2 (20)	19 (40)

[a]At least 1 participant selected this category and at least 1 additional category on a question with a "Check all that apply" instruction.

[b]Participants (N=48) indicated the following countries of origin: Mexico (n=35, 73%), Argentina (n=2, 4%), and Colombia (n=1, 2%).

[c]GED: General Educational Development test.

Focus Group Themes

Overview

We identified seven overarching themes from our focus groups: (1) mHealth is seen as beneficial for promoting health and peace of mind; (2) some are unaware of, unfamiliar with, or uncomfortable with technology and may benefit from individualized support; (3) financial barriers limit access to mHealth; (4) practical considerations create significant barriers to using mHealth in daily life; (5) mHealth raises concern for overreliance on technology; (6) automated mHealth features are perceived as valuable but fallible, requiring human input to ensure accuracy; and (7) data sharing is seen as valuable for limited uses but raises privacy concerns.

We examined the data for differences in themes based on rural or urban location and English and Spanish language but did not identify distinct themes in any subgroups. Illustrative quotes for each theme are provided throughout.

Theme 1: mHealth Is Seen as Beneficial for Promoting Health and Peace of Mind

Most participants described mHealth as beneficial, discussing how this technology could help track chronic conditions and prevent serious health events through earlier intervention. For example, 1 participant noted the potential value of tracking their health conditions:

> *I would think it would be fantastic because I suffer from an autoimmune disease, I have no balance, so I have suffered several falls, and that would be a help for me, a help to be alert all the time. I would love to have that device.* [Urban; Spanish-speaking participant; translated; FG US3]

Another described how mHealth would have helped their family navigate interactions with their physician to improve their child's care:

> *I insisted a lot with the doctors. Because they all sent [my child with asthma] home, that his cough was normal, that his cough was temporary...I wish someone had given us that monitor, because we would go to the doctor and they would say: "And how is his cough? And how often is he coughing?" And I'd be like, "We've already had so many months without being able to sleep, like the least I pay attention to is the cough, what the cough sounds like." Yes, I would have liked to have had a device so it would have been easier for the doctor to know what his cough was like.* [Urban; Spanish-speaking participant; translated; FG US3]

One participant described the potential impact of earlier detection on their rural setting:

> *Here in [rural area], we only have one hospital, the emergency room is packed sometimes. We don't have enough beds on the floors. I feel it would really benefit in preventing people from going really bad to the hospital and flooding the hospital with people. Instead, they can use their devices and go to their doctor, take a little better care of their health or at least keep track of—especially with chronic issues. They can monitor if they're starting to notice some changes, then they can act on it earlier than later.* [Rural; English-speaking participant; FG RE2]

Many participants also discussed that these tools could promote peace of mind and feelings of safety. They linked the presented examples to their own experiences caring for children with asthma or supporting older family members in living independently, highlighting how technology could allow them to feel comfortable without needing to continuously monitor their family members. Several participants reflected on their fears as parents of children with asthma:

> *I used to take [my children] to my room so I can be watching them because it was scary. And especially at night, because during the day, they were fine. But at night, they usually had those asthma attacks when they were little. And we ended up in the hospital...As they get older, I mean, they tell you or they get out of*

XSL•FO
RenderX

> *their bedroom or whatever to tell you. But when they're little, they don't. So you have to be watching them.* [Rural; English-speaking participant; FG RE2]
>
> *I also think it's a very good idea because someone in my family has asthma, so, maybe that would have saved me a lot of sleepless nights.* [Urban; Spanish-speaking participant; translated; FG US2]

Others spoke about the value of mHealth monitoring for older people living independently:

> *This scenario made me think of my grandma who lives by herself, and she's very much a lot older. And she has a lot of health issues, right? So I think if her pressure is low or something and she falls, that could be—I think [an mHealth device] would be really helpful.* [Urban; English-speaking participant; FG UE1]
>
> *I have a neighbor who lives alone and is already old, and that makes me have all the access of her movements. Yes I would like that device to monitor...Because nobody is aware of her, she is left alone, and I think it would be very helpful.* [Rural; Spanish-speaking participant; translated; FG RS1]

In addition, some discussed the role of mHealth in allowing people to continue their regular activities, sometimes referencing our slideshow example of the boy using mHealth to manage his asthma while continuing to play soccer. A participant highlighted the importance of allowing a child to play sports without the added burden of fear about asthma:

> *I think it is good because the child would not abandon the games that he likes. And if that [mobile device to track asthma] did not exist, a mother would say, "You better not play." And if it did exist, she would say, "Yes, play. I am going keep an eye on you." Then, the child will no longer say, "If I do this, this will happen to me."* [Rural; Spanish-speaking participant; translated; FG RS2]

Theme 2: Some Are Unaware of, Unfamiliar With, or Uncomfortable With Technology and May Benefit From Individualized Support

Many participants noted that they had used mHealth apps for tracking things like physical activity or diet, but overall, most expressed a general unfamiliarity with mHealth technologies used to track and make predictions about risk for specific conditions. Several mentioned that no health professional had ever made them aware of such devices:

> *Well, I think [it would be good] for them to let us know that they have this kind of device, because I never heard that they have this device. They never told me.* [Rural; English-speaking participant; FG RE2]
>
> *[I would have questions about] the name of the devices, which company can provide them, if there is any support to be able to have those devices, if they are going to start providing them in the clinics, where one can get more information, or for example, in the clinic, who can one address specifically to say, "Hey, do you know if the device is already available".... In other words, how can one go about obtaining these aids.* [Rural; Spanish-speaking participant; FG RS2]

Participants further noted that some people might feel uncomfortable with using technology. They cited examples of older family members either avoiding technology in general or specifically avoiding wearable health devices. Some used these examples to highlight how the value of a device is contingent on an individual's comfort. For example, a participant provided an example of how their family supported their mother in getting comfortable with technology despite hesitancy about wearables and challenges with mobile devices:

> *So I know my mom wouldn't wear one, and we always worry about her also. She'll lose her phone and stuff, and we'll be trying to call her, and she won't find her phone, and so my best thing was to get her [a voice-activated device]. So I showed her how to use that, and she's pretty okay with it. So it's nothing she has to wear, but whenever she needs to make a phone call and she can't find her phone, that's what she uses.* [Rural; English-speaking participant; FG RE1]

Some participants expressed that most people could get used to new technology over time. One highlighted how quickly this could happen for children:

> *They just hop onto everything, whatever, and technology is not very difficult for them to get used to and learn it, and they pretty much get used to stuff real quick that has to do with—and especially if it has to do with their health is like you need to do it.* [Rural; English-speaking participant; FG RE1]

Others emphasized the need for practice over time:

> *Even if they want to, if it is not being used, they have the training, the direction—whatever they want to call it—to be able to use it and know what is going to happen, to know about the program or the equipment so they can help. If we don't know how to use something, it doesn't help us at all, on the contrary, it will give us more problems.* [Urban; Spanish-speaking participant; FG UE3]
>
> *I don't think it will be hard to learn. It would just take some time. To get used to anything, you get used to a new phone, you get used to anything you get new. So you get used to it. So it takes practice.* [Rural; English-speaking participant; FG RE2]

Some participants also identified a need for individualized support, such as developer-created tutorials or one-on-one support, to facilitate comfort with mHealth. One participant discussed the value of personal support:

> *I think, especially for Latino families, I think we need more person-to-person support. Like, "Take this, read it. If you have any questions, find me, call me." Because yes, sometimes we bring things into the house and we don't understand it and we don't know who to ask or how to use it. So, having that reinforcement, having that person who can guide you, help you and*

XSL•FO
RenderX

get you out of doubt when using new things. [Urban; Spanish-speaking participant; FG UE3]

Theme 3: Financial Barriers Limit Access to mHealth

The cost of devices was a concern for many participants, who raised questions about whether devices would be covered by insurance or what the cost would be for those without insurance coverage. Several called out the likely or actual cost of these devices as a barrier:

It'll probably come with a hefty price tag. [Urban; English-speaking participant; FG UE1]

I haven't used [mHealth], as the price is a bit high. [Rural; Spanish-speaking participant; translated; FG RS2]

I had one that was given to me as a gift, but it was one of the first ones that came out. Then, after that, well, I didn't—the truth is, financially I never put money to buy a new one for me, because they were more sophisticated, more expensive. [Urban; Spanish-speaking participant; translated; FG US2]

Some participants specifically noted the potential expense if multiple family members needed their own devices:

I would imagine for those who have more than one person in the family, it will cost a lot of money. [Rural; English-speaking participant; FG RE2]

If there are three, four, five children in the family, it is not enough to be buying—we are talking about the economic [impact], to be buying devices of all kinds and that [have] everything the child needs. [Rural; Spanish-speaking participant; translated; FG RS1]

Several participants highlighted the importance of insurance coverage for affordability, including how inconsistent insurance coverage could impact people's use of mHealth:

[MHealth devices] are expensive and all this. But if you have insurance and if it's within the insurance budget, they're not expensive. [Urban; Spanish-speaking participant; translated; FG US2]

The cost, that would be a problem that many people face, such as the costs of the products that provide us with these types of services. Like, for example, in my case I use a sensor, if it were not because I have an insurance, I couldn't afford it. [Urban; Spanish-speaking participant; translated; FG US1]

I think it also depends on cost or insurance, because a lot of us don't have access. An example. Right now, I no longer have access to insurance since I stopped working, because I have a temporary job, so in two months I run out of insurance. So, that's when the benefit comes or we no longer have benefit, for the drugs, the appointment, everything that comes with it is more expensive. So certain people who have or do not have health insurance, that's where I think the conflict really comes. But I think a lot of people would be willing to use it if they are cheap. [Rural; Spanish-speaking participant; translated; FG RS1]

One participant expressly noted that these resource limitations may disproportionately affect Latino individuals:

And of course, always the Latinos, right [laughs], we don't have the resources of those who speak English...Well, yes, the whites, they have them, we don't have them. [Urban; Spanish-speaking participant; translated; FG US2]

Theme 4: Practical Considerations Create Barriers to Using mHealth in Daily Life

Participants also discussed potential technical barriers, including inconsistent internet access, limited phone storage capacity, limited battery life, and technology crashing or simply not working. For example, a rural participant commented on the challenges of relying on inconsistent internet and cell service in their area:

How many times does our internet go away? How many times do the poles get knocked down or whatever, or there's lost connections? Something that we can have, like maybe going back to those dial tones. I don't know. Plugging them into the wall. I don't know, but trying to find something that's not going to crash on us, especially if your child is having one of these asthmatic attacks. [Rural; English-speaking participant; FG RE2]

Other participants also commented on inconsistencies in internet access:

Not everybody is going to have access to it. As they say now, "The Internet," but there are still places and there are still people who do not have access to it. [Urban; Spanish-speaking participant; translated; FG US3]

The only problem would be if, for example, how the device works; if it is with electricity, and if one day the power goes out; or if it is by means of wifi, even if the power goes out, it stops working. [Urban; Spanish-speaking participant; translated; FG US2]

Participants also described the breadth of challenges that people experience with their mobile devices:

Every device can fail at some point, or maybe it is low on battery or I don't know what it will have or—there is always a problem, but there is always—nothing is 100% safe. [Urban; Spanish-speaking participant; translated; FG US1]

The charge doesn't last long. So maybe having something that doesn't die so quickly. Again, mine could be because it's an older version or maybe it's time. Maybe it's a hint of having to go get a new one. I don't know. But the fact that it dies really, really fast because it's calculating everything. The more apps you have on it, the more battery it takes. [Rural; English-speaking participant; FG RE2]

It adds very quickly to your memory, it fills up very quickly. Well, make it less. Because can you imagine if we have a disease app? People who sometimes have so many diseases are going to fill up the phone and

XSL•FO
RenderX

> *sometimes the memory is not enough for them. Let it be less, please.* [Rural, Spanish-speaking participant; translated; FG RS2]

Participants further noted that the benefits of mHealth would be limited if not embedded within an environment that facilitated those benefits. For example, a child using a smartwatch to manage their asthma would need their school to allow them to use the device and coordinate with the family:

> *I wanted to ask you, to see if this device can be taken to school and the teacher can also have the information about the child, to be able to help him, because, just as the other family members can also help, they will have the information, the teacher can also help you.* [Urban; Spanish-speaking participant; translated; FG US1]

Similarly, another participant noted that not all work settings facilitate the use of mHealth devices:

> *Sometimes at work you can't even see the clock, much less the telephone. So, I think it would be a little difficult.* [Urban; Spanish-speaking participant; translated; FG US2]

Theme 5: mHealth Raises Concern for Overreliance on Technology

Some participants raised concerns about mHealth leading to overdependence on technology. For example, one participant described mHealth as a part of a broader societal move toward technology dependence:

> *Yes, we are advancing and we are advanced, but they are the things that leave us like a bad taste in our mouths: we are not up to date in the moment, in the second. Because I saw that suddenly, "Oh, three more hours on my phone. Leave me, I reinstall it, because it is not up to date."* [Rural; Spanish-speaking participant; translated; FG RS1]

Another participant discussed how technology could undermine one's self-reliance, especially for a child:

> *One of the challenges that I noticed even in the cartoon characters, that they showed Arturo's discouragement and feeling like he's trapped in all this technology causing him to not be the person that he wants to be because he wants to be a good soccer player, for sure. So that could interfere with his psyche and his confidence in himself.* [Rural; English-speaking participant; FG RE1]

Others spoke about the potential impact on family relationships. Some raised concerns about family members relying on devices and consequently neglecting to engage directly with their family members about their health:

> *I think there is a lot of elderly very, very alone, and this will make it worse, like, "Oh, okay. It will let me know. The little thing will send me a message if she falls." And I won't be able to like it. I'm kind of more old fashioned. I think maybe I can get more involved or ask a friend or ask more family to take care or be a company [] instead of just put a collar on it and just forget about it.* [Urban; English-speaking participant; FG UE1]
>
> *I have a teenager, and sometimes it's hard to make a conversation with them, right? So I think it will be more challenged to connect with them if you're just looking at the phone instead of asking him, "How you feel during the day?" or, "How was the challenge? How you feel when you run and jump?" And if I'm just worried about the app or what the phone said or the wrist coughing track, for me, it would be like I don't care how—it's going to be like stopping the communication. I would be more worried about the monitor and not asking my child. And maybe they will feel like, "Oh, she knows now that I've been coughing because she's been monitoring," instead of telling me, "How you feel with this and that?" And maybe the monitor is tracking just the coughing, but it didn't tell me what activities they were doing, and I think we'll be much better to stay with one-on-one communication with our kids.* [Urban; English-speaking participant; FG UE1]

Theme 6: Automated mHealth Features are Perceived as Valuable but Fallible, Requiring Human Input to Ensure Accuracy

Limitations of technology also came up in terms of data collection. Although automatic, continuous data collection was often seen as beneficial and easy, some participants noted that manual input would be more reliable in some cases, most notably in tracking mood and mental health. One participant suggested that manual input for a period would improve subsequent accuracy:

> *You wouldn't want or you wouldn't be able to just be with the child 24/7 all the time, which is why automatic kind of makes sense in this really busy world. But I guess in something like this where it's life or death and it's your child, you're willing to drop everything to be with them until this gets settled. So whether it's the day, the week, the weeks to then get it honed down to the accuracy and the level of punctuality that you need, it would make sense to do it manually as well. And then work automatically.* [Rural; English-speaking participant; FG RE1]

Another highlighted an example from their own experience of needing a human to step in when health technology fails:

> *I think we are going to realize if something is wrong with the device or with our child. I tell you this because my grandson uses a feeding pump and when the pump is failing, we can tell. So, then we call the technician and we have the possibility to feed him manually.* [Urban; Spanish-speaking participant; translated; FG US3]

Others discussed the importance of checking on data from the device to verify the information it provided, viewing the role of the device as providing them with a high level of information and then allowing them to make decisions. For example, one participant described the need to balance the ability to take

XSL•FO
RenderX

action when warranted with avoiding panic due to frequent alarms:

> *Imagine if [the device] is alarming every hour, then that can cause not only that the person gets scared or the child gets scared, but the others stay in panic, because they don't know if it is right, if it is correct. So it would be like, maybe, give them time to be able to verify all the symptoms and that yes, instead of sending and sending alerts.* [Urban; Spanish-speaking participant; translated; FG US3]

Another expressed skepticism that the information would provide clinical value:

> *For me, when my kids get sick or something, I never been asked by the pediatrician how many times it was coughing, never had to track that. It's just in general, how the kids are eating or playing or how they look, their appearance. Sometimes they look very tired, or you can tell when they couldn't breathe very well...So I think the monitor will be hard to track that, and I won't trust completely like, "Oh, you coughed 100 times. Let's go to the hospital."* [Urban; English-speaking participant; FG UE1]

Despite these concerns, participants generally indicated that they would prefer to receive false positive alerts rather than miss true positives:

> *I think, for me, it would be more stressful if I didn't detect a real asthma attack. Since it's about saving the child's life, what does it matter if it's a false alarm? Well, yes, I'm going to be a little stressed, but I'd rather have that than it not detect an asthma attack. I know that computers can't replace humans and they're not perfect either, but yes, I'd rather have the device be more sensitive and have a false alarm once in a while than not detect, not be sensitive enough to detect all asthma attacks.* [Urban; Spanish-speaking participant; translated; FG US2]

> *I would rather have a million false alarms than the one where they direly needed help and that one wasn't alerted.* [Rural; English-speaking participant; FG RE1]

Theme 7: Data Sharing Is Seen as Valuable for Limited Uses but Raises Privacy Concerns

Overall, participants endorsed the value of data sharing for uses such as providing emergency support or sharing clinical information with key family members or clinical team members. Participants also supported the use of data for improving the technology and making better predictions, at least on the level of the individual device. For example, one person described the importance of collecting location data as well as health data, both for emergency services and to improve accurate predictions:

> *I think it's very good, because it is giving the place where he is at. And let's say he is feeling very, very bad and can't even talk, then all that helps. And if someone is watching him, it would help, if he is in a place where no one else is around or close by, to find him, to alert the doctor, to know that the child is going to need help or that he is in difficulty...And maybe they can add more information, like "where you are, you're prone to have really bad air at this time of day." All of that could help them, the kids, make the best decisions for themselves.* [Urban; Spanish-speaking participant; translated; FG US3]

However, another participant described data sharing as a "double-edged sword" that would increase privacy concerns:

> *I know that it is to improve the service or the product, but it is also like a double-edged sword, because your information goes into more hands so, if it is no longer completely clear to you how your information is handled, when it is already in the hands of other people. So, in that aspect, I would like to have a little more privacy.* [Urban; Spanish-speaking participant; translated; FG US1]

Other participants also identified privacy concerns, highlighting the possibility of data breaches or data misuse:

> *So if somebody who isn't a professional hacks into his watch, that would be invasion of his privacy. If his friends, people at school, hack into his watch, he would probably be embarrassed. And again, it's just taking away his privacy, which the hacking is—could be an issue.* [Rural; English-speaking participant; FG RE1]

> *I would be afraid that, let's say, the medicine companies might say, "In this area there are many children with asthma or more attacks" and they might want to take advantage of that situation in some way. I don't know. And also, since all the information is virtual, and even if one is careful and cautious, there are many places where one's private information has been seen by many people who have nothing to do with it.* [Urban; Spanish-speaking participant; translated; FG US3]

These concerns were heightened for information they deemed sensitive, such as location data and information about sensitive or stigmatized health conditions. One participant identified mental health as an area that might raise particular concerns:

> *I think the disadvantage of using these technologies for more private things, as a mental health issue, is that there are times that they do like hacking, that they get into your phone and grab information that, obviously, you don't want other people to know or that is very sensitive. That's why I sometimes get nervous when I'm putting medical stuff on my phone.* [Rural; Spanish-speaking participant; translated; FG RS1]

Participants also suggested ways to find a balance between the value of data sharing in limited cases and the risks of privacy breaches. For example, one person proposed limits on the types of data to be shared:

> *Sometimes, everyone wants to keep [their information] private, and even more so when it comes*

XSL•FO
RenderX

> *to children, so maybe you are supposed to be sending the information to trusted people, but, for example, as a mother I would feel safer just having my daughter's location. So, or her father, as [another participant] mentioned. That is, the people as closer. But from then on, like, for example, the activity of coughing or anything else, obviously it does have to be shared, but for me, the location [should not be shared]*. [Urban; Spanish-speaking participant; translated; FG US1]

Another suggested that those accessing data enter into explicit agreements about data use:

> *The only ones that have access to the location would be his doctor or his doctor's office or the emergency room, the ambulance. I think that that would be—I know as a parent myself, I would be scared about people knowing where my kids' locations is at, right?...But giving access just to— even signing something saying that they're the only ones that are going to have access to their location in case of an emergency. Something written, not verbal. I would get something documented, written, that says, "I am only going to have access to their location in case of an emergency.*" [Rural; English-speaking participant; FG RE2]

Discussion

Principal Findings

Our focus group findings report important perspectives on mHealth from Hispanic and Latinx individuals across rural and urban regions of Washington State. Although the themes we report share similarities with perspectives of other communities, they highlight the importance of tailoring mHealth tools to the context in which they will be used, specifically among a community with high rates of underinsurance and limited financial resources, including many individuals in rural areas and many whose primary language is Spanish. Our results illustrate key barriers to the benefits of mHealth in these communities, provide insights into the role of mHealth within families, and examine the appropriate balance of data sharing and privacy protections.

Barriers to the Benefits of mHealth

Participants perceived mHealth as beneficial overall; however, our findings highlight several critical barriers to accessing those benefits. First, there was little awareness of the potential for mHealth to be used to manage chronic conditions. Participants with experience with asthma or other conditions expressed particular interest in having a health professional tell them about these types of tools. Second, many participants expressed that they or their family members might be uncomfortable using mHealth technologies, which could lead to them either never initiating or stopping the use of a device. Suggestions to provide time for practice, tutorials, or even personalized support arose in multiple focus groups as strategies to improve comfort and familiarity. Third, participants reported resource constraints and the lack of insurance coverage as significant barriers to accessing mHealth. Finally, participants identified both features of the device (eg, technical usability) and broader contextual features (eg, integration with school systems) as limiting the usability of mHealth tools if not adequately addressed. These findings, particularly related to resources and provision of support, are mirrored in qualitative work with other populations, including residents of rural areas [35] and English- and Spanish-speaking patients in a safety net setting (ie, care providers that serve patients regardless of their insurance status or ability to pay) [36]. Another pilot study that included interviews with 10 Latina women similarly identified limited access and underutilization of mHealth as limiting the technology's social value to the broader community [37]. The imbalance of high perceived benefits with significant barriers illustrates the need to approach mHealth research and development from a justice-oriented perspective, as Herington et al [24] argue in the context of digital health research that works on both decreasing risks and, importantly, developing specific, community-informed mechanisms to equitably share benefits.

The Role of mHealth in the Family

Participants also discussed the complex role of mHealth within family dynamics. Many identified relational benefits of using mHealth devices to make personalized predictions about risk when used in a family setting, providing caregivers of children and older adults with peace of mind. At the same time, participants acknowledged that their family members, particularly older adults, may view such monitoring as interfering with their autonomy. Some also worried about the risk of undermining interpersonal relationships from relying too heavily on technology. These dynamics reflect the value of considering a "family informatics" approach to mHealth design, which explicitly recognizes the interrelatedness of family members and their health [38]. Family connectedness and commitment are core cultural values that may shape how Latinx individuals approach their use of mHealth devices [23,39]. Thus, family informatics may be particularly relevant for ensuring devices are culturally appropriate. In addition, to navigate concerns for overreliance on technology, developers should consider hybrid solutions that integrate the benefits of automated technology alongside human input [40].

Data Sharing and Privacy Concerns

A key issue in the use of mHealth is how data are shared, including whether and how artificial intelligence incorporates user data [41]. In discussions of data sharing, participants identified specific benefits in the use of data sharing to facilitate emergency responses and improve personalized prediction accuracy. Although disclosing individual data was generally seen as acceptable for these beneficial uses, this was balanced with privacy considerations, particularly around location data and sensitive health information. Privacy concerns and the risks of breaches of privacy may be heightened among individuals from marginalized groups, who differentially experience harmful impacts from discriminatory social structures [24,42]. Primary language is one relevant dimension of marginalization; a recent survey of parents of children with asthma found that Spanish-speaking parents were more likely than their English-speaking counterparts to report privacy concerns about digital health tools [17].

XSL•FO
RenderX

Participants suggested solutions to limit data access to only those necessary to produce the benefits identified as valuable. Discussion about data sharing for purposes of improving predictions more generally was limited but similarly focused on benefits and risks. A key way to accomplish this appropriate balance will be to ensure that the benefits of artificial intelligence–enabled devices do, in fact, reach the communities that are taking on privacy risks [41,43]. Future work should continue to examine the nuances of community members' perspectives on this benefit-risk balance in the context of predictive technologies.

Limitations

Our study represents the views of community members drawn from a specific region of the United States; people with other life experiences, cultural, ethnic, or national backgrounds [44], and values may have different perspectives. Of note, most (43/48, 90%) participants were women, and most (35/48, 73%) reported that Mexico was their country of origin. Nevertheless, by situating our findings within the broader literature on community perspectives about mHealth and other health technology, we aim to provide nuance from the perspective of this underrepresented group.

Further, although we explored topics related to the use of mHealth by children and older adults, children were not eligible to participate in this study, and participants were in midadulthood on average. Additional perspectives would be important to include in guidance on devices targeted at additional age groups. Future work is also needed to explore mHealth as applied to specific conditions, such as mental health conditions that may raise additional privacy considerations.

Finally, while our case examples illustrated artificial intelligence–enabled mHealth tools and focus group discussions built on these examples, we were not able to fully explore the complexities of attitudes about artificial intelligence within this study. Future work should continue to seek ways to engage with community members about these topics and their implications.

Conclusions

This focus group study with Hispanic and Latinx community members illustrates the importance of developing mHealth with input from underrepresented communities. The in-depth perspectives of these participants highlight the necessity of developing and implementing tools with attention to barriers, family context, and privacy concerns to avoid exacerbating disparities in the use and value of health technology. Future work should incorporate these perspectives into guidance to support mHealth developers in creating tools that meet community needs and advance equity.

Acknowledgments

The authors would like to thank the team at the Institute of Translational Health Sciences for supporting this supplement, Booster Shot Media for collaborating with us on the development of our narrative slide presentations, and study participants for contributing their time and perspectives. This project was funded by the National Institutes of Health, through National Center for Advancing Translational Sciences grant 3UL1TR002319-06S1, with additional support from National Library of Medicine grant R01LM012810.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable written request.

Authors' Contributions

SAK designed the overall project; developed data collection instruments and procedures; acquired, analyzed, and interpreted data; and drafted the manuscript. SC contributed to data collection instruments and procedures; acquired, analyzed, and interpreted data; and reviewed the manuscript. MCD contributed to data collection instruments and procedures; acquired, analyzed, and interpreted data; and reviewed the manuscript. JAR contributed to data collection instruments and procedures, acquired and interpreted data, and reviewed the manuscript. AB contributed to data collection instruments and procedures; acquired, analyzed, and interpreted data; and reviewed the manuscript. KIL contributed to data collection instruments and procedures, acquired and interpreted data, and reviewed the manuscript. RJ contributed to data collection instruments and procedures, acquired data, and reviewed the manuscript. BSW designed the overall project, developed data collection instruments and procedures, interpreted data, and reviewed the manuscript. MR designed the overall project, developed data collection instruments and procedures, interpreted data, and reviewed the manuscript. JF designed the overall project, developed data collection instruments and procedures, interpreted data, and reviewed the manuscript. LKK designed the overall project; developed data collection instruments and procedures; acquired, analyzed, and interpreted data; and reviewed the manuscript. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Focus group guide.
[DOCX File , 27 KB - jmir_v27i1e59817_app1.docx]

Multimedia Appendix 2
Narrative slideshow (English version).
[PDF File (Adobe PDF File), 1329 KB - jmir_v27i1e59817_app2.pdf]

Multimedia Appendix 3
Narrative slideshow (Spanish version).
[PDF File (Adobe PDF File), 1338 KB - jmir_v27i1e59817_app3.pdf]

References

1. WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://iris.who.int/handle/10665/44607 [accessed 2024-04-03]
2. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980 [FREE Full text] [doi: 10.2196/37980] [Medline: 36315221]
3. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR Mhealth Uhealth 2022 Jun 09;10(6):e35053 [FREE Full text] [doi: 10.2196/35053] [Medline: 35679107]
4. Emeryk A, Derom E, Janeczek K, Kuźnar-Kamińska B, Zelent A, Łukaszyk M, et al. Home monitoring of asthma exacerbations in children and adults with use of an AI-aided stethoscope. Ann Fam Med 2023;21(6):517-525 [FREE Full text] [doi: 10.1370/afm.3039] [Medline: 38012028]
5. Tsang KC, Pinnock H, Wilson AM, Salvi D, Shah SA. Predicting asthma attacks using connected mobile devices and machine learning: the AAMOS-00 observational study protocol. BMJ Open 2022 Oct 03;12(10):e064166 [FREE Full text] [doi: 10.1136/bmjopen-2022-064166] [Medline: 36192103]
6. Tsang KC, Pinnock H, Wilson AM, Salvi D, Shah SA. Home monitoring with connected mobile devices for asthma attack prediction with machine learning. Sci Data 2023 Jun 08;10(1):370 [FREE Full text] [doi: 10.1038/s41597-023-02241-9] [Medline: 37291158]
7. Rodriquez EJ, Pérez-Stable EJ. The time is now for eHealth research with Latinos. Am J Public Health 2017 Nov;107(11):1705-1707. [doi: 10.2105/AJPH.2017.304055] [Medline: 29019786]
8. Meng G, Jan Ali M, Tse SM. Caregivers' perceptions, needs, and data sharing concerns in mHealth research on pediatric asthma: cross-sectional survey study. JMIR Pediatr Parent 2023 Dec 19;6:e49521 [FREE Full text] [doi: 10.2196/49521] [Medline: 38127911]
9. Dauletbaev N, Oftring ZS, Akik W, Michaelis-Braun L, Korel J, Lands LC, et al. A scoping review of mHealth monitoring of pediatric bronchial asthma before and during COVID-19 pandemic. Paediatr Respir Rev 2022 Sep;43:67-77. [doi: 10.1016/j.prrv.2022.01.002] [Medline: 35131174]
10. Fedele DA, Thomas JG, McConville A, McQuaid EL, Voorhees S, Janicke DM, et al. Using mobile health to improve asthma self-management in early adolescence: a pilot randomized controlled trial. J Adolesc Health 2021 Dec;69(6):1032-1040. [doi: 10.1016/j.jadohealth.2021.06.011] [Medline: 34274211]
11. Iio M, Miyaji Y, Yamamoto-Hanada K, Narita M, Nagata M, Ohya Y. Beneficial features of a mHealth asthma app for children and caregivers: qualitative study. JMIR Mhealth Uhealth 2020 Aug 24;8(8):e18506 [FREE Full text] [doi: 10.2196/18506] [Medline: 32831181]
12. Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
13. Kaufmann J, Marino M, Lucas J, Bailey SR, Giebultowicz S, Puro J, et al. Racial and ethnic disparities in acute care use for pediatric asthma. Ann Fam Med 2022;20(2):116-122 [FREE Full text] [doi: 10.1370/afm.2771] [Medline: 35346926]
14. Asthma and Hispanic Americans. U.S. Department of Health and Human Services Office of Minority Health. URL: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=60 [accessed 2023-08-03]
15. van der Kamp MR, Hengeveld VS, Brusse-Keizer MG, Thio BJ, Tabak M. eHealth technologies for monitoring pediatric asthma at home: scoping review. J Med Internet Res 2023 Jul 21;25:e45896 [FREE Full text] [doi: 10.2196/45896] [Medline: 37477966]
16. Atske S, Perrin A. Home broadband adoption, computer ownership vary by race, ethnicity in the U.S. Pew Research Center. 2021 Jul 16. URL: https://www.pewresearch.org/short-reads/2021/07/16/home-broadband-adoption-computer-ownership-vary-by-race-ethnicity-in-the-u-s/ [accessed 2023-08-03]
17. Kan K, Morales L, Shah A, Simmons E, Barrera L, Massey L, et al. Digital technology characteristics and literacy among families with children with asthma: cross-sectional study. JMIR Pediatr Parent 2023 Nov 29;6:e48822 [FREE Full text] [doi: 10.2196/48822] [Medline: 38031428]

XSL•FO
RenderX

18. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
19. Vogels EA. Some digital divides persist between rural, urban and suburban America. Pew Research Center. 2021 Aug 19. URL: https://www.pewresearch.org/short-reads/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/ [accessed 2023-08-03]
20. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453 [FREE Full text] [doi: 10.1126/science.aax2342] [Medline: 31649194]
21. Seyyed-Kalantari L, Zhang H, McDermott MB, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med 2021 Dec;27(12):2176-2182 [FREE Full text] [doi: 10.1038/s41591-021-01595-0] [Medline: 34893776]
22. Park J, Arunachalam R, Silenzio V, Singh VK. Fairness in mobile phone-based mental health assessment algorithms: exploratory study. JMIR Form Res 2022 Jun 14;6(6):e34366 [FREE Full text] [doi: 10.2196/34366] [Medline: 35699997]
23. Parker LJ, Marx KA, Nkimbeng M, Johnson E, Koeuth S, Gaugler JE, et al. It's more than language: cultural adaptation of a proven dementia care intervention for Hispanic/Latino caregivers. Gerontologist 2023 Mar 21;63(3):558-567 [FREE Full text] [doi: 10.1093/geront/gnac120] [Medline: 35951488]
24. Herington J, Connelly K, Illes J. Ethical imperatives for working with diverse populations in digital research. J Med Internet Res 2023 Sep 18;25:e47884 [FREE Full text] [doi: 10.2196/47884] [Medline: 37721792]
25. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
26. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2014.
27. Halcomb EJ, Gholizadeh L, DiGiacomo M, Phillips J, Davidson PM. Literature review: considerations in undertaking focus group research with culturally and linguistically diverse groups. J Clin Nurs 2007 Jun;16(6):1000-1011. [doi: 10.1111/j.1365-2702.2006.01760.x] [Medline: 17518876]
28. Epstein DA, Ping A, Fogarty J, Munson SA. A lived informatics model of personal informatics. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2804250]
29. Rothwell E, Anderson R, Botkin JR. Deliberative discussion focus groups. Qual Health Res 2016 May;26(6):734-740 [FREE Full text] [doi: 10.1177/1049732315591150] [Medline: 26078330]
30. Dedoose homepage. Dedoose. URL: https://www.dedoose.com/ [accessed 2024-04-03]
31. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Thousand Oaks, CA: SAGE Publications Ltd; 2006.
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
34. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep;25(9):1212-1222. [doi: 10.1177/1049732315588501] [Medline: 26184336]
35. Schoenberg N, Dunfee M, Yeager H, Rutledge M, Pfammatter A, Spring B. Rural residents' perspectives on an mHealth or personalized health coaching intervention: qualitative study with focus groups and key informant interviews. JMIR Form Res 2021 Feb 26;5(2):e18853 [FREE Full text] [doi: 10.2196/18853] [Medline: 33635278]
36. Purnell L, Sierra M, Lisker S, Lim MS, Bailey E, Sarkar U, et al. Acceptability and usability of a wearable device for sleep health among English- and Spanish-speaking patients in a safety net clinic: qualitative analysis. JMIR Form Res 2023 Jun 05;7:e43067 [FREE Full text] [doi: 10.2196/43067] [Medline: 37098152]
37. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
38. Pina LR, Sien SW, Ward T, Yip JC, Munson SA, Fogarty J, et al. From personal informatics to family informatics: understanding family practices around health monitoring. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR. [doi: 10.1145/2998181.2998362]
39. Garcia CN, Duran MC, Ramirez M. Refining cultural adaptations of a behavioral intervention for Latino caregivers of people living with dementia: qualitative interview study in Washington state. JMIR Aging 2024 Jan 11;7:e53671 [FREE Full text] [doi: 10.2196/53671] [Medline: 38206663]
40. Choe EK, Abdullah S, Rabbi M, Thomaz E, Epstein DA, Cordeiro F, et al. Semi-automated tracking: a balanced approach for self-monitoring applications. Pervasive Computing. 2017. URL: https://courses.cs.washington.edu/courses/cse440/20wi/assets/readings/PersonalInformatics-Choe2017.pdf [accessed 2024-04-03]

XSL•FO
RenderX

41. Hussein R, Griffin AC, Pichon A, Oldenburg J. A guiding framework for creating a comprehensive strategy for mHealth data sharing, privacy, and governance in low- and middle-income countries (LMICs). J Am Med Inform Assoc 2023 Mar 16;30(4):787-794 [FREE Full text] [doi: 10.1093/jamia/ocac198] [Medline: 36259962]
42. Gangadharan SP. The downside of digital inclusion: expectations and experiences of privacy and surveillance among marginal internet users. New Media Soc 2015 Nov 09;19(4):597-615 [FREE Full text] [doi: 10.1177/1461444815614053]
43. Kraft SA. Centering patients' voices in artificial intelligence based telemedicine. Am J Public Health 2023 May;113(5):470-471. [doi: 10.2105/AJPH.2023.307270] [Medline: 36926958]
44. Gonzalez M, Sanders-Jackson A, Wright T. Web-based health information technology: access among Latinos varies by subgroup affiliation. J Med Internet Res 2019 Apr 16;21(4):e10389 [FREE Full text] [doi: 10.2196/10389] [Medline: 30990462]

Abbreviations

mHealth: mobile health

Edited by A Schwartz; submitted 03.05.24; peer-reviewed by S Oesterle, L Palazzo; comments to author 15.08.24; revised version received 26.09.24; accepted 24.12.24; published 06.02.25.

Please cite as:
Kraft SA, Chopra S, Duran MC, Rojina JA, Beretta A, López KI, Javan R, Wilfond BS, Rosenfeld M, Fogarty J, Ko LK
Perspectives of Hispanic and Latinx Community Members on AI-Enabled mHealth Tools: Qualitative Focus Group Study
J Med Internet Res 2025;27:e59817
URL: https://www.jmir.org/2025/1/e59817
doi:10.2196/59817
PMID:

©Stephanie A Kraft, Shaan Chopra, Miriana C Duran, Janet A Rojina, Abril Beretta, Katherine I López, Russell Javan, Benjamin S Wilfond, Margaret Rosenfeld, James Fogarty, Linda K Ko. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding the Association Between Home Broadband Connection and Well-Being Among Middle-Aged and Older Adults in China: Nationally Representative Panel Data Study

Lu Yang[1,2,3], PhD; Chris Lynch[2,3,4], PhD; John Tayu Lee[2,5], PhD; Brian Oldenburg[2,3,4], PhD; Tilahun Haregu[2,3,6], PhD

[1]School of Sociology and Population Studies, Nanjing University of Posts and Telecommunications, Nanjing, China

[2]School of Psychology and Public Health, La Trobe University, Melbourne, Australia

[3]Non-Communicable Diseases and Implementation Science, Baker Heart and Diabetes Institute, Melbourne, Australia

[4]National Health and Medical Research Council Centers of Research Excellence in Digital Technology to Transform Chronic Disease Outcomes, Melbourne, Australia

[5]College of Public Health, National Taiwan University, Taipei, China

[6]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Lu Yang, PhD
School of Sociology and Population Studies
Nanjing University of Posts and Telecommunications
9 Wenyuan Road
Nanjing, 210049
China
Phone: 86 18094233761
Email: Lindsay.yanglu@outlook.com

Abstract

Background: Access to digital technology is among the major social determinants of health, and digital divide impacts health inequality. Yet, the impact of digital connectivity on the well-being and psychosocial outcomes in adults has not been fully studied.

Objective: The aim of this study was to investigate the association of home broadband connection with health and well-being of middle-aged adults and adults older than 45 years in China.

Methods: A panel data study design of the national sample of China Health and Retirement Longitudinal Study (CHARLS) was conducted in 2015, 2018, and 2020. This study included 16,185 participants older than 45 years. The associations between digital connectivity (home broadband connection), loneliness, social participation, and life satisfaction were assessed using mixed effects logistic regression models, adjusting for socioeconomic factors, behavioral factors, and locality. Broadband internet connectivity, feelings of loneliness, social participation, and satisfaction with life were measured using the self-reported CHARLS questionnaire.

Results: We observed a substantial increase in digital connectivity from 29.5% in 2015 to 59.8% in 2020. Broadband internet connection at home was positively correlated with social participation (adjusted odds ratio [AOR] 1.34, 95% CI 1.28-1.41) and life satisfaction (AOR 1.30, 95% CI 1.20-1.40), after adjusting for confounding factors, while the absence of broadband internet connection was associated with increased loneliness (AOR 0.81, 95% CI 0.77-0.86). These associations were consistent across age, gender, socioeconomic groups, and geographic areas.

Conclusions: This study highlights the potential additional health benefits of digital connectivity beyond the known advantages. Our results suggest the importance of expanding broadband access to enhance social inclusion and life satisfaction. Further research is needed to understand the broader implications and digital determinants of health associated with digital connectivity.

(J Med Internet Res 2025;27:e59023) doi:10.2196/59023

KEYWORDS

digital divide; health inequity; China; longitudinal study; broadband; internet connection; internet; well-being; psychosocial; middle age; older adult; inequality; digital connectivity; logistic regression; questionnaire; survey; panel data approach

XSL•FO
RenderX

Introduction

Loneliness and social isolation are clearly recognized as significant threats to public health [1]. Although social isolation is defined as an objective lack of interactions with others or the wider community, loneliness is a subjective feeling of the absence of a social network or a companion [2]. These 2 concepts do not necessarily coexist: a person may be socially isolated but not lonely or socially connected but still feel lonely. However, both social isolation and loneliness lead to poorer health outcomes and higher health care expenditure [3,4]. Notably, the COVID-19 pandemic led to a significant increase in social isolation and loneliness among older adults in some countries; for example, in Canada, there was a 37%-67% increase in loneliness during April-December 2020 compared to 2011-2015 [5,6]. Throughout the pandemic, governments, organizations, and technology companies made efforts to help prevent and address social isolation and loneliness [1]. Broadband connection at home provides a faster way to connect older adults to reduce feelings of loneliness and isolation via online discussion forums, online peer support platforms, social media, and telehealth [5,7,8]. The availability of home broadband means faster, reliable delivery of internet-based services and facilitates access to a spectrum of resources that encompass education, social activities, and health care services, all of which help to foster a broader and deeper level of digital engagement [9,10].

In our ever-connected society, home broadband connection has become essential for daily living. Globally, there has been a rapid increase in the number of households equipped with broadband internet service in recent years [11,12]. According to the Organization for Economic Co-operation and Development broadband portal data, the broadband subscriptions in 32 countries grew to 2.25 billion in December 2022, up from 1.99 billion at the end of 2019—an increase of 260 million in 3 years [13]. As of 2023, China is the leading market for the number of internet users in the world, followed by India and the United States [14]. According to the latest data from the China Internet Network Information Center, 2023 witnessed 1.11 billion internet broadband access ports actively connected, with a reported household broadband penetration rate exceeding 90% in 2020 [15]. However, there still remains a substantial gap between the most developed regions in China and the rest of the country, contributing to a widening digital divide [16]. Findings from 5 longitudinal cohort studies across 23 countries, including China, underscored a significant association between digital exclusion and difficulties in performing daily activities [17]. Furthermore, the China Household Finance Survey data suggest that while the digital divide index has gradually reduced from 58.36 in 2017 to 49.89 in 2019, its effect has led to a significant reduction in household consumption. This reduction is more prominent within the older, less educated, unhealthy, and low-income groups [16].

The rapid acceleration of the aging population has created a major challenge for health care services and industrial transformation process in China [18,19]. To understand more about these challenges, the China Health and Retirement Longitudinal Study (CHARLS) was designed to be comparable with the US Health and Retirement Study and related aging surveys around the world (eg, the English Longitudinal Survey of Aging and the Survey of Health, Aging, and Retirement in Europe), while being sensitive to the specific conditions of China [20]. The national baseline survey was conducted in 2011-2012, with continued waves in 2013, 2015, 2018, and most recently in 2020. Historically, CHARLS datasets have been used to provide individual social, economic, behavioral, and health data to understand and serve the needs of older adults, which have greatly enriched the international landscape of aging studies [21]. The introduction of the national-level Smart Eldercare action plan in 2017 [22] led to an increase in internet-related health research among older adults, such as internet use and its association with physical, mental, and subjective health through to health disparities, social activities, incidence of chronic diseases, and disease management [23-27]. The most recent study on internet access from CHARLS, using data from 2011-2018, identified that internet access improves older adults' self-reported health, mental health, and activities of daily living [28]. Another study used CHARLS data from 2011 and 2015 and reported that social isolation, rather than loneliness, was significantly associated with functional disability over 4 years among women (but not men) [29]. However, we are not aware of any recently published research on home broadband connection and its impact on the well-being and psychosocial outcomes among older people. This gap is critical for 3 reasons. First, broadband availability is associated with greater frequency of internet use and more engagement with internet activities [10]. Second, older adults showed higher preferences for home care [30]. Third, the recent advances in the internet of things technologies [31] have made home care solutions viable and appealing for older people who wish to age at home.

Digital technologies are now being recognized as important determinants of health [32]. In 2021, the Geneva Charter for Well-Being called for action to address the digital determinants of health [33]. It recognized that avoiding digital exclusion is the key to achieve universal health coverage, making full use of the digital transformation and ensuring access and meaningful participation. Alongside this, the World Health Organization highlighted the role of broadband and the internet as an example of the digital determinants of health [34]. Moreover, the US Federal Communications Commission reported that not having access to broadband internet had impact on physical or mental disability, as telemedicine requires high-speed internet access [35]. China has positioned the broadband network as a national strategic infrastructure for economic and social development since 2013 [36]. Notably, despite over 90% penetration now, the broadband penetration rate was only around 60% in 2020 when the CHARLS survey was conducted, which presents an opportunity to examine the potential impact of this technology in different social contexts. Yet, to date, no study has examined the relationship of home broadband connection with life satisfaction, loneliness, and social participation in China. This paper aims to fill this gap by being the first to analyze population-based nationally representative data from 2015 to 2020, to explore whether home broadband connection is associated with well-being and psychosocial characteristics such

XSL•FO
RenderX

as life satisfaction, loneliness, and social participation among older adults in China.

Methods

Study Design and Participants

This study complies with PRICSSA (Preferred Reporting Items for Complex Sample Survey Analysis) guidelines. Data were obtained from CHARLS in 2015, 2018, and 2020, which collects information on individuals aged 45 years or older in China and their household characteristics. Thirty province-level administrative units were selected by stratified random sampling with probabilities proportional to the size to be representative of the general population aged 45 years and older [37]. There were 17,708 respondents in the national baseline survey in 2011 who were followed up using a face-to-face computer-assisted personal interview every 2 years. A detailed description of CHARLS is published elsewhere [38]. For this study, we included data from participants who were included for all 3 waves in 2015, 2018, and 2020, after China introduced "Broadband China" Strategy and Implementation Plan in 2013. After data cleaning, the sample comprised 16,185 individuals.

Ethics Approval

The original CHARLS was approved by the ethics review committee of Peking University (IRB00001052-11015), and all participants signed informed consent at the time of participation. For this study, we examined the home broadband connection status of all adults aged 45 years and older in the deidentified datasets.

Measures

The key predictor is home broadband connection. Respondents who answered affirmatively to the question "Does your residence have broadband internet connection?" were defined as having a home broadband connection.

Demographic Characteristics

Demographic variables included area of living (rural/urban), gender, age, marital status (married and partnered, unmarried, and others), education level (primary school and below, secondary school, and college and above), work status (yes or no), socioeconomic groups, and health insurance (Urban Employee Basic Medical Insurance, Urban Resident Basic Medical Insurance, New Rural Cooperative Medical Scheme, other insurance, and without insurance). Specifically, we used annual per capita consumption spending as a proxy for socioeconomic status. We defined 4 socioeconomic groups on the basis of quartiles of per capita household consumption expenditure [39].

Dependent Variables

Life satisfaction was measured in the questionnaire "Please think about your life-as-a-whole. How satisfied are you with it? Are you completely satisfied, very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied?" with 5 Likert scale responses ranging from "1=completely satisfied," "2=very satisfied," "3=somewhat satisfied," "4=not very satisfied," and "5=not at all satisfied." To enable the use of a chi-square test, scales 1, 2, 3 were recoded as "satisfied" and scales 4-5 were recoded as "not satisfied."

Loneliness

Loneliness was measured with a direct single question in the questionnaire "I felt lonely." The 4-point response scale ranged from "1=rarely or none of the time (<1 day)," "2=some or a little of the time (1-2 days)," "3=occasionally or a moderate amount of the time (3-4 days)," to "4=most or all of the time (5-7 days)." To ensure use of a chi-square test, scale 1 was recoded as "no loneliness" and scales 2-3 were recoded as "yes loneliness."

Social Participation

Social participation status was measured from the question "Have you done any of these activities in the last month?" For social participation, which coded as 1=Yes, it means that the participants checked the following social activities: interacted with friends; played Mah-Jong (a tile-based game developed and widely played in China), played cards, or went to a community club; went to a sporting event or participated in a social group; provided help to family, friends, or neighbors who do not live together; went to a sport, social, or other kind of club; took part in a community-related organization; participated in voluntary or charity work; and attended an educational or training course. For those participants who did not check any of the above social activities, this was coded as 0=No.

Covariates

We included demographic characteristics as covariates. Some health-related variables were also included, such as self-report health status, duration of sleep, and number of noncommunicable diseases. Self-report health status was measured by: "How is your health, overall?" The response options lay on a 4-point Likert scale: very good=0, good=1, not so good=2, and not good at all=3. We recoded the scale to good=1, fair=2, and poor=3, so that higher scores reflected poorer self-rated health condition. Fifteen self-reported chronic diseases, each prompted by "Has a doctor or nurse ever told you that you have…," were asked to the respondents. The number of chronic diseases was considered in this study, which comprises arthritis, heart diseases, diabetes, stroke, chronic lung disease, kidney disease, stomach or other digestive diseases, asthma, depression, hypertension, dyslipidemia, Alzheimer disease, cancer, dementia, and liver illness.

Statistical Analyses

We used a panel data approach of mixed effects logistic regression to examine the associations between home broadband connection and life satisfaction, loneliness, and social participation. Frequencies, percentages, mean, and standard deviations were used to report descriptive data. The chi-square test was used to compare the differences in the demographic characteristics, health behavior, health status, and access to broadband internet connection with the outcome measures (life satisfaction, social participation, and loneliness). To explore the differential effect in population groups, we conducted subgroup analyses, stratified by age group, gender, area of residence, and socioeconomic status. To examine the possible heterogeneity in the relationships between home broadband

XSL•FO
RenderX

connection and psychosocial outcomes, we conducted a series of stratified analyses by age, gender, rural-urban, and socioeconomic groups. All tests were conducted using STATA software, and statistical significance was considered at $P<.05$.

Results

Descriptive Analysis

There were 16,185 respondents aged 45 years and older who completed the CHARLS survey in all 2015, 2018, and 2020 waves, which were included in the analysis. There were 2735 participants who did not have home broadband connection in 2015 and 2018 but had home broadband connection in 2020. Most of the participants were in the 45-60 years age group (6914/15,958, 43.3%) and 60-70 years age group (5534/15,958, 34.7%); 7526 (46.8%) participants were males, 8562 (53.2%) were females, and 12,830 (79.4%) were married or partnered. A total of 8958 (66.8%) participants had primary education or below, and 13,362 (74.8%) were residing in rural areas; 15,713 (94.1%) had at least one kind of health insurance, with 10,442 (67.5%) respondents enrolled in the New Rural Cooperative Medical Scheme.

Multimedia Appendix 1 shows the prevalence of social participation, life satisfaction, and loneliness from 2015 to 2020 by the status of home broadband connection. The proportion of people with home broadband connection increased substantially over time. In 2015, the overall prevalence of broadband connection was 29.5% (4741/16,070), while the prevalence increased to 45.7% (7382/16,143) in 2018 and 59.8% (9659/16,161) in 2020. The mean age of the respondents was 61.6 years in 2020. Table 1 presents the findings of the association of home broadband connection with socioeconomic and demographic characteristics, health behavior characteristics, health status, social participation, life satisfaction, and loneliness. For those who had home broadband connection, they were more likely to be within the age group of 45-60 years, married or in partnership, having higher education level, living in rural areas, having New Rural Cooperative Medical Scheme, having higher socioeconomic status (quartile 4), not drinking alcohol, having more sleep, self-rating their health status as good or fair, having participated in social activities, not feeling lonely, and satisfied with life. Interestingly, in both 2015 and 2018, for those who had home broadband connection, they were more likely to have no chronic diseases, while in 2020, those who had home broadband connection were more likely to have more than 3 types of chronic diseases. In 2015 and 2018, for those who had home broadband connection, they were more likely to be nonsmokers, while in 2020, this relationship was not significant.

XSL•FO
RenderX

Table 1. Association between home broadband connection and demographic and socioeconomic characteristics in 2015, 2018, and 2020.

	2015 (n=19,977)		2018 (n=19,151)		2020 (n=19,093)	
	Values	*P* value	Values	*P* value	Values	*P* value
Age group (years), n (%)		<.001		<.001		<.001
45-60	3544 (58.5)		5111 (57.5)		6348 (54.5)	
60-70	1679 (27.7)		2566 (28.8)		3584 (30.8)	
70-80	600 (9.9)		936 (10.5)		1315 (11.3)	
80 and older	236 (3.9)		283 (3.2)		392 (3.4)	
Gender, n (%)		.12		.44		.24
Male	2940 (48.7)		4298 (47.8)		5584 (47.3)	
Female	3099 (51.3)		4687 (52.2)		6235 (52.7)	
Marital status, n (%)		<.001		<.001		<.001
Unmarried and others	836 (13.6)		1581 (17.6)		2585 (21.9)	
Married and live with spouse	5334 (86.4)		7404 (82.4)		9234 (79.1)	
Education, n (%)		<.001		<.001		<.001
Less than elementary school	1921 (45.9)		4672 (52)		6601 (55.8)	
Middle school and higher education	2265 (54.1)		4313 (48)		5218 (44.2)	
Residence status, n (%)		<.001		<.001		<.001
Urban	2915 (47.3)		3580 (39.8)		5215 (44.2)	
Rural	3242 (52.7)		5405 (60.2)		6596 (55.8)	
Health insurance, n (%)		<.001		<.001		<.001
None	455 (8.6)		179 (2)		432 (3.7)	
Urban Employee Basic Medical Insurance	1012 (19.2)		1713 (19.1)		2191 (18.5)	
Urban Resident Basic Medical Insurance	403 (7.7)		1424 (15.9)		1700 (14.4)	
New Rural Cooperative Medical Scheme	2806 (53.3)		4852 (54)		7194 (60.9)	
Others	589 (11.2)		810 (9)		302 (2.5)	
Socioeconomic group, n (%)		<.001		<.001		<.001
Quartile 1	516 (10.5)		1116 (12.6)		1626 (13.8)	
Quartile 2	968 (19.7)		1453 (16.4)		2225 (18.8)	
Quartile 3	1358 (27.6)		2618 (29.5)		4049 (34.2)	
Quartile 4	2078 (42.2)		3683 (41.5)		3919 (33.2)	
Work status, n (%)		.001		.09		<.001
No	2215 (35.8)		3332 (37.1)		4171 (35.3)	
Yes	3973 (64.2)		5655 (62.9)		7648 (64.7)	
Drinking alcohol, n (%)		<.001		<.001		<.001
No	3731 (60.5)		5597 (62.3)		7257 (61.5)	
Yes	2432 (39.5)		3382 (37.7)		4551 (38.5)	
Smoking, n (%)		<.001		.003		.75
No	3648 (58.9)		5284 (58.8)		6877 (58.2)	
Yes	2540 (41.1)		3703 (41.2)		4942 (41.8)	
Sleep time (h), mean (SD)	6.54 (1.63)	<.001	6.33 (1.78)	<.001	6.16 (1.73)	<.001

XSL•FO
RenderX

	2015 (n=19,977)		2018 (n=19,151)		2020 (n=19,093)	
	Values	*P* value	Values	*P* value	Values	*P* value
Number of noncommunicable diseases, n (%)		<.001		<.001		<.001
0	3150 (50.9)		5117 (56.9)		2264 (19.2)	
1	1371 (22.2)		2426 (27)		2624 (22.2)	
2	856 (13.8)		909 (10.1)		2391 (20.2)	
≥3	811 (13.1)		535 (6)		4540 (38.4)	
Self-assessed health status, n (%)		<.001		<.001		<.001
Good	1825 (31.6)		2464 (29.5)		2859 (26.5)	
Fair	3026 (52.4)		4262 (51.1)		5636 (52.3)	
Poor	924 (16)		1615 (19.4)		2288 (21.2)	
Social participation, n (%)		<.001		<.001		<.001
No	1911 (34.5)		3297 (39.6)		5561 (47.1)	
Yes	3632 (65.5)		5034 (60.4)		6247 (52.9)	
Life satisfaction, n (%)		<.001		<.001		<.001
No	339 (5.9)		745 (9)		1032 (9.6)	
Yes	5397 (94.1)		7546 (91)		9748 (90.4)	
Loneliness, n (%)		<.001		<.001		<.001
No	4725 (82.5)		6071 (74.3)		7959 (74.5)	
Yes	1005 (17.5)		2098 (25.7)		2720 (25.5)	

Home Broadband Connection and Social Participation

Table 1 shows that the prevalence of social participation among home broadband users was 65.5% (3632/5543) as compared to 50.1% (6743/13,466) among nonusers—a difference of 15.4% in 2015. Similarly, there was a difference of 17.3% and 12.5% between home broadband users and nonusers in 2018 and 2020, respectively. Overall, we noticed that social participation rates decreased in both groups from 2015 to 2020. After adjusting for covariates (Table 2), broadband users were 34% more likely to participate in social activities as compared to nonusers (odds ratio 1.34, 95% CI 1.28-1.41). Stratified analyses (Figures 1-4) showed that home broadband connection and social participation do not have significant relationships among those older than 80 years but still have a significant relationship across gender, socioeconomic groups, and areas of residence.

XSL•FO
RenderX

Table 2. Mixed effects logistic regression results of effects of home broadband connection on social participation, life satisfaction, and loneliness.

Variables	Social participation		Life satisfaction		Loneliness	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Broadband connection						
No	1.00	Reference	1.00	Reference	1.00	Reference
Yes	1.34 (1.28-1.41)	<.001	1.30 (1.20-1.40)	<.001	0.81 (0.77-0.86)	<.001
Age group (years)						
45-60	1.00	Reference	1.00	Reference	1.00	Reference
60-70	0.83 (0.79-0.87)	<.001	1.51 (1.39-1.64)	<.001	0.95 (0.89-1.00)	.053
70-80	0.77 (0.72-0.82)	<.001	1.89 (1.70-2.11)	<.001	0.96 (0.89-1.03)	.27
80 and older	0.73 (0.65-0.82)	<.001	2.74 (2.21-3.41)	<.001	0.76 (0.80-1.00)	.04
Gender						
Male	1.00	Reference	1.00	Reference	1.00	Reference
Female	1.36 (1.27-1.46)	<.001	0.76 (0.67-0.85)	<.001	1.29 (1.20-1.39)	<.001
Marital status						
Unmarried and others	1.00	Reference	1.00	Reference	1.00	Reference
Married and live with spouse	0.89 (0.85-0.94)	<.001	1.57 (1.44-1.70)	<.001	0.42 (0.40-0.44)	<.001
Education						
Less than elementary school	1.00	Reference	1.00	Reference	1.00	Reference
Middle school and higher education	1.41 (1.34-1.48)	<.001	1.14 (1.05-1.25)	.002	0.94 (0.89-1.00)	.04
Residence status						
Urban	1.00	Reference	1.00	Reference	1.00	Reference
Rural	1.05 (1.00-1.11)	.08	0.97 (0.88-1.06)	.49	1.25 (1.17-1.33)	<.001
Health insurance						
None	1.00	Reference	1.00	Reference	1.00	Reference
Urban Employee Basic Medical Insurance	1.92 (1.69-2.17)	<.001	2.51 (2.04-3.08)	<.001	0.73 (0.63-0.84)	<.001
Urban Resident Basic Medical Insurance	1.34 (1.20-1.51)	<.001	2.03 (1.71-2.42)	<.001	0.86 (0.75-0.98)	.02
New Rural Cooperative Medical Scheme	1.14 (1.03-1.26)	.01	1.66 (1.44-1.91)	<.001	0.90 (0.80-1.00)	.06
Others	1.81 (1.57-2.09)	<.001	1.95 (1.55-2.45)	<.001	0.78 (0.66-0.91)	.002
Socioeconomic group						
Quartile 1	1.00	Reference	1.00	Reference	1.00	Reference
Quartile 2	1.09 (1.03-1.16)	.005	1.17 (1.06-1.28)	.002	0.85 (0.80-0.91)	<.001
Quartile 3	1.20 (1.13-1.28)	<.001	1.20 (1.09-1.32)	<.001	0.86 (0.80-0.92)	<.001
Quartile 4	1.30 (1.22-1.39)	<.001	1.13 (1.01-1.26)	.03	0.77 (0.72-0.83)	<.001
Work status						
No	1.00	Reference	1.00	Reference	1.00	Reference
Yes	1.02 (0.97-1.08)	.41	1.00 (0.92-1.09)	.97	1.08 (1.02-1.15)	.006
Drinking alcohol						
No	1.00	Reference	1.00	Reference	1.00	Reference
Yes	1.48 (1.38-1.52)	<.001	1.04 (0.96-1.13)	.34	1.01 (0.94-1.05)	.74
Smoking						

XSL•FO
RenderX

Variables	Social participation		Life satisfaction		Loneliness	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
No	1.00	Reference	1.00	Reference	1.00	Reference
Yes	1.17 (1.09-1.27)	<.001	0.85 (0.76-0.94)	.003	1.08 (0.99-1.18)	.03
Sleep time	1.00 (1.00-1.01)	.66	1.11 (1.09-1.13)	<.001	0.91 (0.90-0.92)	<.001
Number of noncommunicable diseases						
0	1.00	Reference	1.00	Reference	1.00	Reference
1	1.11 (1.05-1.17)	<.001	1.03 (0.93-1.13)	.59	1.09 (1.02-1.16)	.01
2	1.17 (1.09-1.25)	<.001	0.96 (0.85-1.07)	.44	1.23 (1.14-1.32)	<.001
≥3	1.31 (1.22-1.40)	<.001	0.83 (0.74-0.92)	.001	1.52 (1.41-1.64)	<.001
Self-assessed health status						
Good	1.00	Reference	1.00	Reference	1.00	Reference
Fair	0.91 (0.86-0.96)	<.001	0.47 (0.42-0.54)	<.001	1.61 (1.51-1.72)	<.001
Poor	0.76 (0.71-0.81)	<.001	0.19 (0.16-0.21)	<.001	2.59 (2.40-2.79)	<.001
Random effects parameter						
Year[a]	0.026 (0.005-0.131)	N/A[b]	0.044 (0.008-0.229)	N/A	0.027 (0.005-0.137)	N/A

[a]The level used are years of survey.

[b]N/A: not applicable.

Figure 1. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and social participation by area of residence.

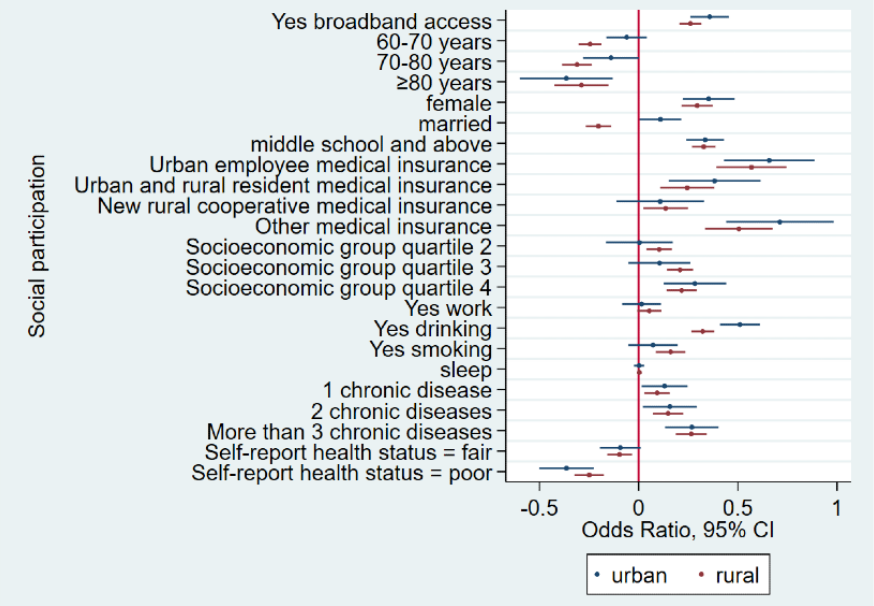

XSL•FO
RenderX

Figure 2. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and social participation by gender.

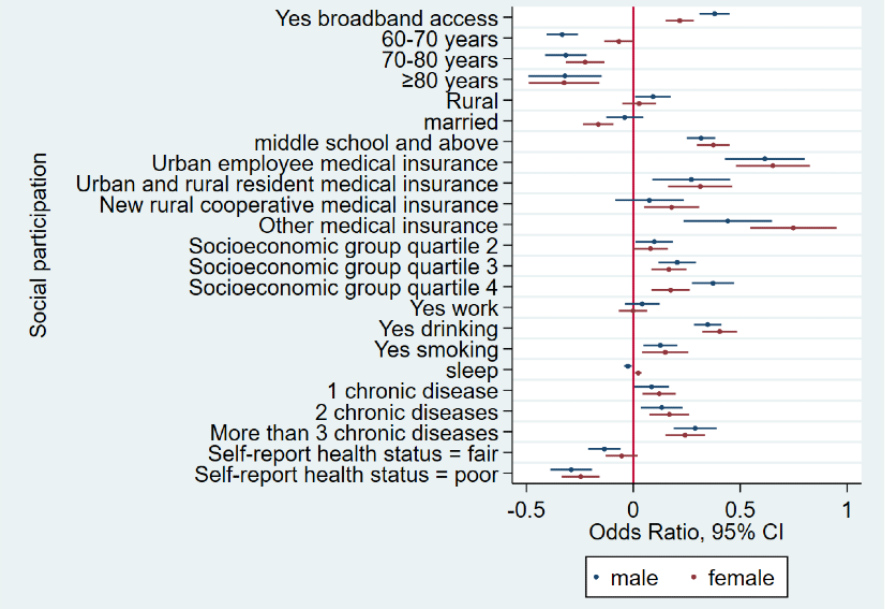

Figure 3. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and social participation by age groups (45-60 years/60-70 years/70-80 years/80 years and older).

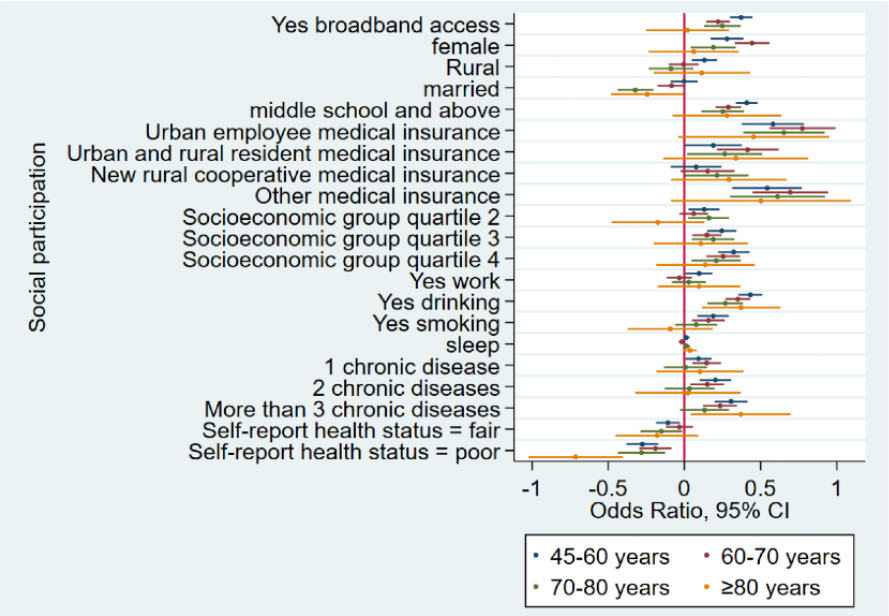

XSL•FO
RenderX

Figure 4. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and social participation by socioeconomic status from the lowest to the highest (quartile 1/quartile 2/quartile 3/quartile 4).

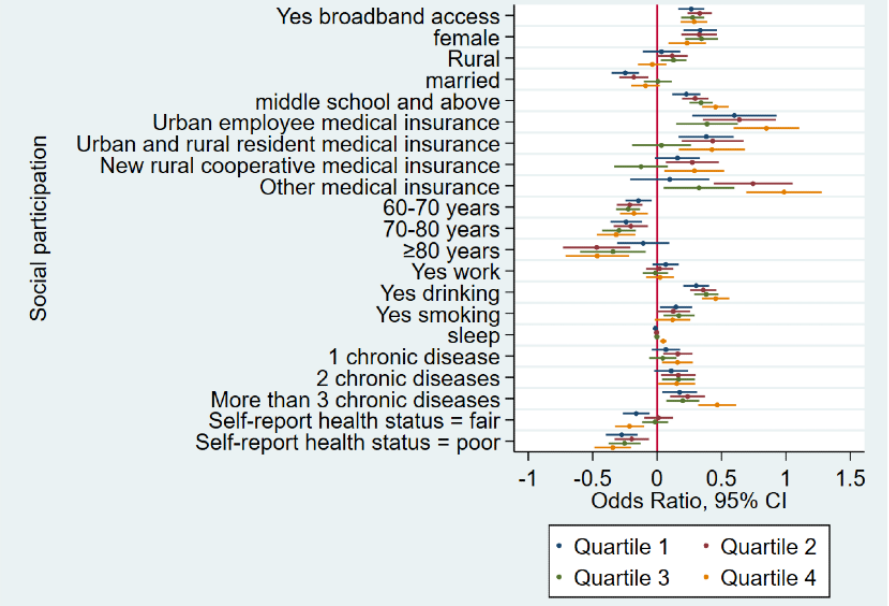

Home Broadband Connection at Home and Life Satisfaction

Table 1 shows that the prevalence of life satisfaction among broadband users was 94.1% (5397/5736) as compared to 90.4% (12,126/13,413) among nonusers—a difference of 3.7% in 2015. Similarly, there was a difference of 4.9% and 4% in 2018 and 2020, respectively. Overall, life satisfaction rates decreased in the broadband user group. After adjusting for covariates, Table 2 shows that broadband users were 1.30 times more likely to feel life satisfaction when compared to nonusers (adjusted odds ratio 1.30, 95% CI 1.20-1.40). Stratified analyses (Figures 5-8) showed that home broadband connection and life satisfaction still have a significant relationship across various ages, genders, socioeconomic groups, and geographic areas.

Figure 5. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and life satisfaction by area of residence.

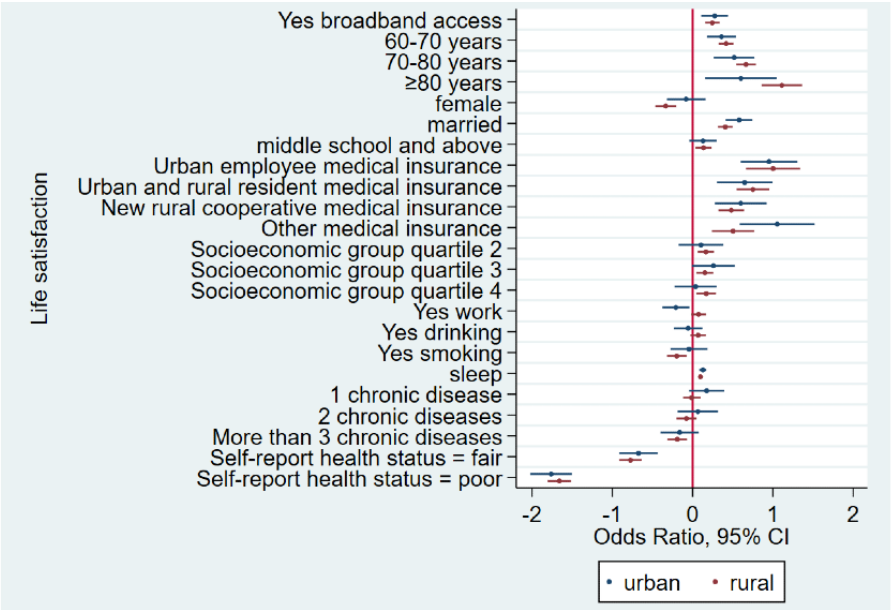

XSL•FO
RenderX

Figure 6. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and life satisfaction by gender.

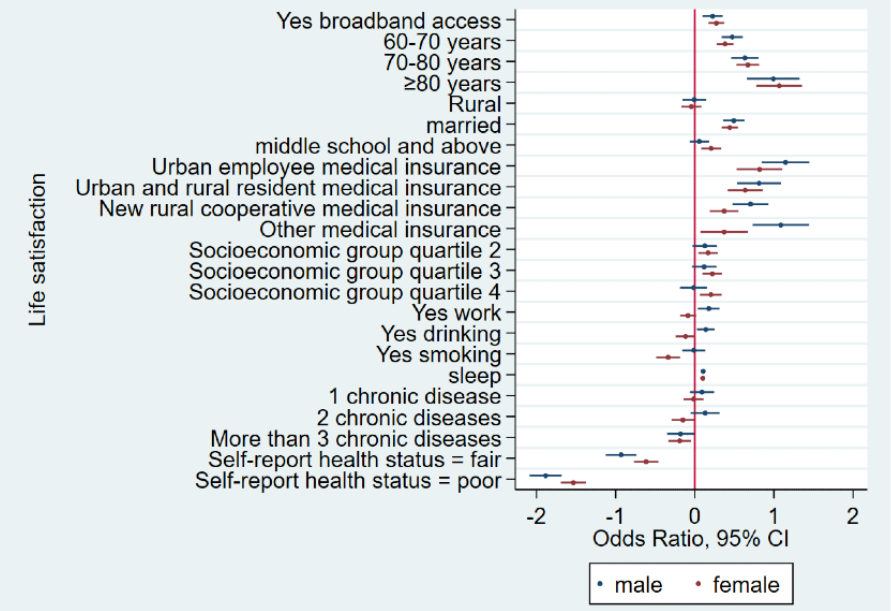

Figure 7. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and life satisfaction by age groups (45-60 years/60-70 years/70-80 years/80 years and older).

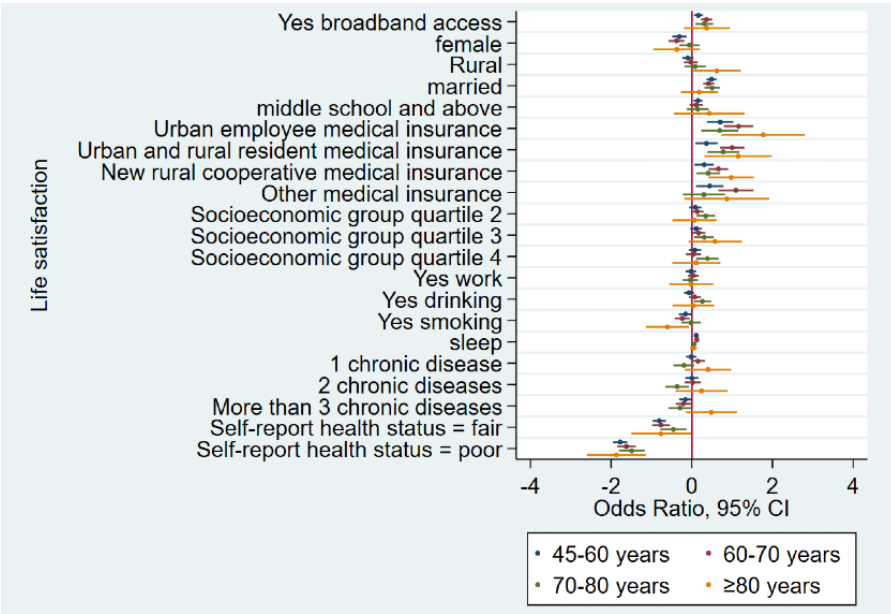

XSL•FO
RenderX

Figure 8. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and life satisfaction by socioeconomic status from the lowest to the highest (quartile 1/quartile 2/quartile 3/quartile 4).

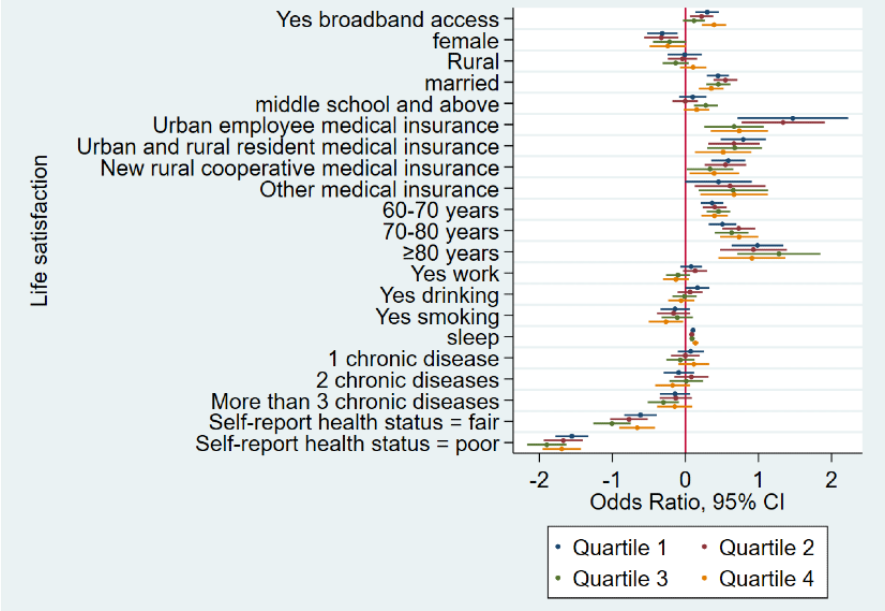

Home Broadband Connection at Home and Loneliness

Table 1 shows that the prevalence of loneliness among broadband users was 17.5% (1005/5730) as compared to 28.5% (3814/13376) among nonusers—a difference of 11% in 2015. Similarly, there was a difference of 9.6% and 9.8% in 2018 and 2020, respectively. Overall, loneliness rates decreased in both groups in 2015 and 2018 but remained similar in 2020. After adjusting for covariates, Table 2 shows that broadband users were 19% less likely to feel loneliness when compared to nonusers (adjusted odds ratio 0.81, 95% CI 0.77-0.86). Stratified analyses (Figures 9-12) showed that the relationship between home broadband connection and loneliness was consistent across various ages, genders, socioeconomic groups, and geographic areas.

Figure 9. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and loneliness by area of residence.

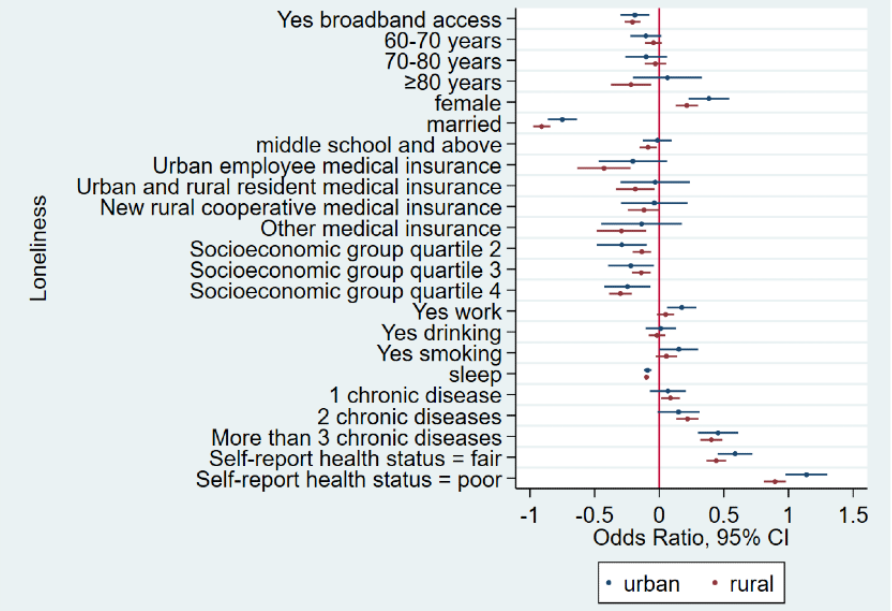

XSL•FO
RenderX

Figure 10. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and loneliness by gender.

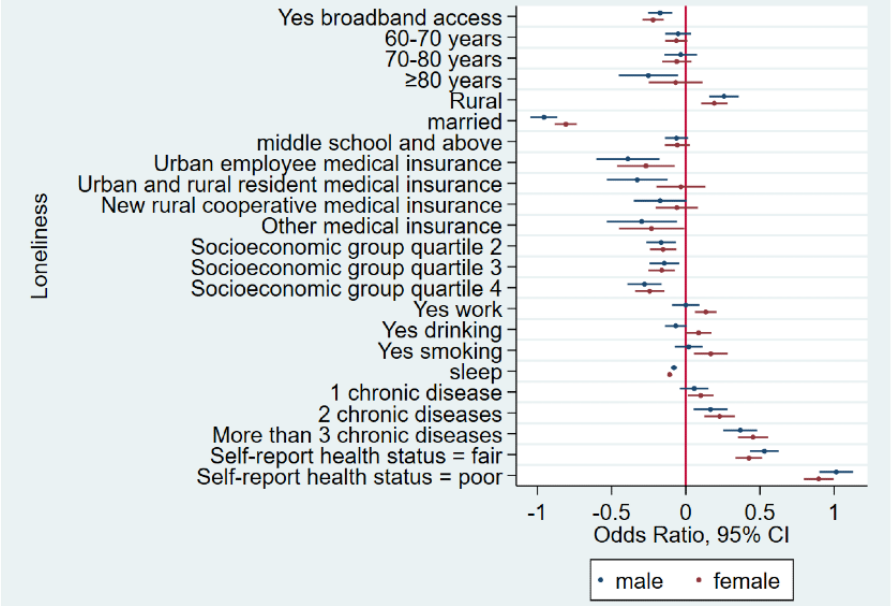

Figure 11. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and loneliness by age groups (45-60 years/60-70 years/70-80 years/80 years and older).

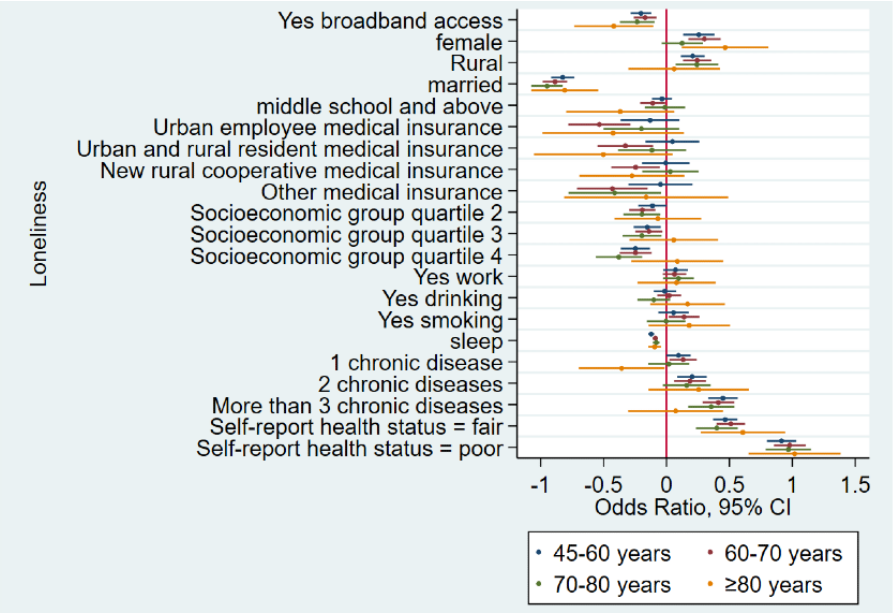

XSL•FO
RenderX

Figure 12. Coefficient plots showing multivariate logistic regression analysis of the association between home broadband connection and loneliness by socioeconomic status from the lowest to the highest (quartile 1/quartile 2/quartile 3/quartile 4).

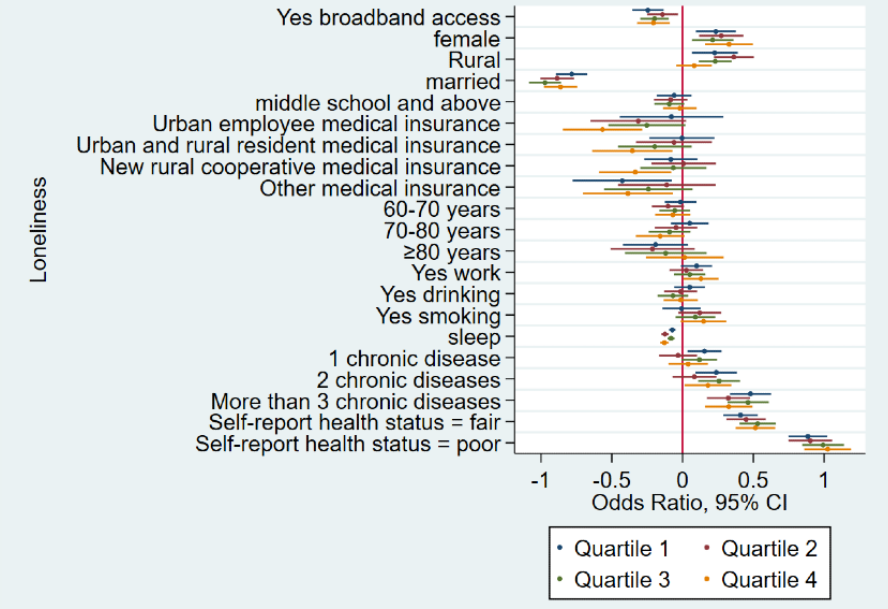

Discussion

Principal Findings

Different from previous analyses of the CHARLS dataset on the association between broadband or internet connection and health-related characteristics, this study examines the broader potential contribution of home broadband connection related to social participation, life satisfaction, and loneliness based on a national longitudinal sample from China. There are 3 key findings from this research: (1) home broadband connection has a positive impact on life satisfaction, (2) social activities increased with broadband connection at home, and (3) feelings of loneliness were decreased with home broadband connection. This research contributes to the existing literature in this field of digital determinants of health in 2 ways. First, our study utilizes the latest CHARLS datasets from 2015 to 2020, whereas other similar works focused on the earlier datasets or only used 1 dataset as the cross-sectional study [24,28]. Second, this study is the first to not only explore the changes of home broadband connection from 2015 to 2020 but also show the impact of home broadband connection on psychosocial characteristics among people older than 45 years in China.

Consistent with previous findings, home broadband connection has a positive impact on life satisfaction [40-42]. The study using the 2019 Chinese Social Survey data of 2105 adults older than 60 years showed that internet use was positively associated with older adults' life satisfaction [42]. Moreover, findings from a large-scale survey of Israel from 2003 to 2012 revealed that internet adoption increases life satisfaction [43]. A UK study using waves 6-8 of the English Longitudinal Study of Aging found that infrequent internet use (monthly or less vs daily) was predictive of deteriorating life satisfaction [44]. Since the COVID-19 pandemic, there has been an increase in the literature focusing on internet addiction or problematic internet use, associated with negative life satisfaction, especially among younger adults [45,46]. Further, studies from the United States reported that time spent browsing the web was negatively related to life satisfaction based on cross-sectional data collected from adult residents in 2004 and 2005 [47]. This discrepancy may be the result of differences in age groups and frequency of internet use. Given the emerging evidence on the importance of home broadband connection impacting life satisfaction, policy makers could consider using broadband internet connectivity as a platform to improve psychosocial outcomes.

The second finding is that, interestingly, the likelihood of having social activities increased with home broadband connection. This result is in accordance with a previous study, suggesting that internet use promotes social activities compared with those who do not use the internet [48]. In addition, there are studies showing that frequent internet use was associated with reductions in the likelihood of visiting family or friends [48,49]. Older adults and people with disabilities and chronic diseases are more vulnerable to social isolation than the rest of the population; it can be challenging for them to maintain regular social participation, especially for those living alone [50-52]. Increasing research demonstrates that digital technologies can be impactful tools for generating and supporting social participation [53,54]. Access to information and communication technologies such as the use of internet-based platforms and telehealth services has been found to increase the quality of life of those populations by saving travel time and expenditure [55-57]. However, such research is underdeveloped for those vulnerable populations; although online social participation expands social experiences and connections, face-to-face experiences were more valued [58]. There is a need to understand the broad range of technologies that are being applied to address social participation and research-integrated (online-offline) social participation to develop a connected life for those populations.

The third finding of this study suggests that the likelihood of feeling lonely was decreased by home broadband connection, which is in line with that reported in other studies in Ireland

XSL•FO
RenderX

[10], Germany [59], and the United States [60]. An Irish study that examined patterns of internet use and psychosocial outcomes for over 3500 people older than 50 years found that high-speed broadband availability was associated with higher reported levels of home internet access, greater frequency of internet use, and more engagement with internet activities and that loneliness scores were significantly lower among those who used the internet daily compared with those who never used the internet [10]. The reason can be explained by that those with access to broadband internet showed higher digital literacy [10], which plays an important role in bridging the digital divide, facilitating digital engagement and allowing for a connected society [61]. In contrast, there are studies indicating that internet addiction or problematic internet use could induce loneliness [62,63]. There are growing research evidences of the biological, psychological, social, and economic impacts of loneliness, which has drawn knowledge, service providers, and policy makers together, in order to raise awareness and reduce the negative effects of loneliness [64,65]. Use of the internet has been shown to enrich the lives of isolated older adults by lowering perceived life stress and increasing perceptions of self-efficacy [66]. Although there is evidence that broadband use among older people has been expanded due to COVID-19 experience [10], more evidence is required for exploring the role of home broadband connection beyond COVID-19. Interventions to reduce loneliness and its health consequences may need to consider the home broadband connection. Such findings may help inform governmental policies on areas such as active aging, infrastructure provision, and digital skills development.

Policy Implications From This Research

As broadband internet connectivity is one of the social determinants of health [67], which affects health care delivery, health outcomes, health education, public health prevention and promotion, advocating for digital equity is vital to people, community, and government. According to our research, having a home broadband connection is associated with higher satisfaction of life and social participation in China. As such, it is important to continue providing broadband connectivity and improve digital infrastructure, especially for those older adults feeling lonely and needing social connection. Since 2015, China has created an “Internet Plus” initiative aiming at connecting a growing economy to the power of connected services, from manufacturing to agriculture, which includes health management services using the internet to exchange health data to achieve quicker, safer, and more integrative care. This national initiative is not only aligned with the sustainable development goals, focusing on areas such as decent work and economic growth, but it also contributes to the enhancement of public health and the reduction of health disparities. Although “Internet plus” should progress to support novel applications of digital health such as artificial intelligence, internet of things, virtual reality, augmented reality, and big data analytics [68], it requires strong governance and political efforts. Moreover, as internet promotes active aging, China has developed National Action Plans for the Development of Smart Health and Eldercare Industry, which integrates digital technology with the older adult care service industry, addressing the needs and health status of older adults [22]. Aligned with World Health Organization’s Universal Health Coverage (all people have access to the full range of quality health services they need, when, and where they need them), this evidence-based research on well-being and psychosocial outcomes can inform policy makers consider reducing health inequality and improving social capital and cohesion, while expanding the home broadband connection.

Strengths and Limitations

Our study’s strengths lie in its use of data from a period marked by the expansion of broadband internet connection, allowing for meaningful comparisons between individuals with and without such connectivity. The meaning of home broadband connection is one of the limitations, as it does not include the speed of the internet or whether it is mobile broadband or fixed broadband, which requires more granular detail to determine what people are doing with this internet access. The variable we used (home broadband connection) may not capture all types of internet connections. Hence, our findings may not apply to access for all types of internet connections (including broadband connection outside home). Utilizing a panel data study design with a national sample, this research ensures robust findings. Moreover, the study effectively adjusts for socioeconomic factors, behavioral factors, and locality, enhancing the control over potential confounding variables. However, as an observational study, causal conclusions are limited, and potential omitted variable bias and unobserved factors may impact associations. Evidence from randomized controlled trials or quasi-experience study designs is needed to provide more robust evidence on the link between broadband connectivity on health and well-being.

Conclusion

In conclusion, this paper contributes some important insights to the field of digital determinants of health by highlighting the potential additional benefits of broadband internet connectivity on well-being and psychosocial outcomes, beyond the known advantages. Since the internet may be beneficial for decreasing loneliness and increasing social participation and life satisfaction, this study underscores the need for further research to fully understand the intricate relationship between broadband internet connection, interactions with the health system and society, and their implications for health and well-being.

Acknowledgments

We thank all the participants of the China Health and Retirement Longitudinal Study as well as the recruitment staff. LY is funded by the National Social Science Fund of China Program (21BRK019). JTL is funded by the Yushan Fellow Program by the Ministry of Education, Republic of China (MOE-112-YSFMN-0003-002-P1).

XSL•FO
RenderX

Authors' Contributions

LY was responsible for conceptualization, literature search, formal analysis, methodology, project administration, validation, and writing the original draft. JTL was responsible for methodology and writing the original draft. CL was responsible for literature search and writing the original draft. TH was responsible for methodology, formal analysis, and validation. BO was responsible for supervision. All authors were responsible for reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prevalence of social participation, life satisfaction, and loneliness among middle-aged adults and adults older than 45 years in China from 2015 to 2020 by their status of home broadband connection.
[PNG File , 112 KB - jmir_v27i1e59023_app1.png]

References

1. Murthy VH. COVID-19 pandemic underscores the need to address social isolation and loneliness. Public Health Rep 2021;136(6):653-655 [FREE Full text] [doi: 10.1177/00333549211045425] [Medline: 34586920]
2. Leigh-Hunt N, Bagguley D, Bash K, Turner V, Turnbull S, Valtorta N, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
3. Mihalopoulos C, Le LK, Chatterton ML, Bucholc J, Holt-Lunstad J, Lim MH, et al. The economic costs of loneliness: a review of cost-of-illness and economic evaluation studies. Soc Psychiatry Psychiatr Epidemiol 2020 Jul;55(7):823-836. [doi: 10.1007/s00127-019-01733-7] [Medline: 31119308]
4. National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Health and Medicine Division, Board on Behavioral, Cognitive, and Sensory Sciences, Board on Health Sciences Policy, Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington DC: National Academies Press; 2020.
5. MacLeod S, Tkatch R, Kraemer S, Fellows A, McGinn M, Schaeffer J, et al. COVID-19 era social isolation among older adults. Geriatrics (Basel) 2021 May 18;6(2):52 [FREE Full text] [doi: 10.3390/geriatrics6020052] [Medline: 34069953]
6. Briere J, Wang SH, Khanam UA, Lawson J, Goodridge D. Quality of life and well-being during the COVID-19 pandemic: associations with loneliness and social isolation in a cross-sectional, online survey of 2,207 community-dwelling older Canadians. BMC Geriatr 2023 Sep 30;23(1):615 [FREE Full text] [doi: 10.1186/s12877-023-04350-x] [Medline: 37777717]
7. Thangavel G, Memedi M, Hedström K. Customized information and communication technology for reducing social isolation and loneliness among older adults: scoping review. JMIR Ment Health 2022 Mar 07;9(3):e34221 [FREE Full text] [doi: 10.2196/34221] [Medline: 35254273]
8. Zubatsky M, Berg-Weger M, Morley J. Using telehealth groups to combat loneliness in older adults through COVID‐19. J Am Geriatr Soc 2020 Aug;68(8):1678-1679 [FREE Full text] [doi: 10.1111/jgs.16553] [Medline: 32392617]
9. Zapletal A, Wells T, Russell E, Skinner MW. On the triple exclusion of older adults during COVID-19: Technology, digital literacy and social isolation. Soc Sci Humanit Open 2023;8(1):100511 [FREE Full text] [doi: 10.1016/j.ssaho.2023.100511] [Medline: 37021073]
10. Mohan G, Lyons S. High-speed broadband availability, Internet activity among older people, quality of life and loneliness. New Media & Society 2022 Jun 04;26(5):2889-2913. [doi: 10.1177/14614448221095218]
11. Ronquillo C, Currie L. The digital divide: Trends in global mobile and broadband Internet access from 2000-2010. NI 2012 (2012) 2012;2012:346 [FREE Full text] [Medline: 24199118]
12. The state of broadband 2023 digital connectivity: A transformative opportunity. ITU/UNESCO Broadband Commission for Sustainable Development. 2023. URL: https://www.broadbandcommission.org/publication/state-of-broadband-2023/ [accessed 2025-01-17]
13. OECD. Growth of mobile data usage continues - but at a slower pace. Satellite Markets & Research. 2023. URL: https://testdesign.satellitemarkets.com/market-trends/growth-mobile-data-usage-continues-slower-pace [accessed 2025-01-17]
14. Statista. URL: https://tinyurl.com/scnzuj77 [accessed 2025-01-17]
15. The 52nd statistical report on China's internet development. China Internet Network Information Center. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202311/P020231121355042476714.pdf [accessed 2025-01-17]
16. Wang J, Yin Z, Jiang J. The effect of the digital divide on household consumption in China. International Review of Financial Analysis 2023 May;87:102593. [doi: 10.1016/j.irfa.2023.102593]
17. Lu X, Yao Y, Jin Y. Digital exclusion and functional dependence in older people: Findings from five longitudinal cohort studies. EClinicalMedicine 2022 Dec;54:101708 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101708] [Medline: 36353265]
18. Zeng Y, Hesketh T. The effects of China's universal two-child policy. The Lancet 2016 Oct;388(10054):1930-1938. [doi: 10.1016/s0140-6736(16)31405-2]

XSL•FO
RenderX

19. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. The Lancet 2019 Sep;394(10204):1192-1204. [doi: 10.1016/s0140-6736(19)32136-1]
20. Zhao Y, Strauss J, Yang G, Giles J, Hu P, Hu Y. China health and retirement longitudinal study: 2011-2012 national baseline users' guide. CHARLS. URL: https://charls.pku.edu.cn/en/doc/User2011.pdf [accessed 2025-01-17]
21. Chen X, Wang Y, Strauss J, Zhao Y. China health and retirement longitudinal study (CHARLS). In: Encyclopedia of Gerontology and Population Aging. United States: Springer International Publishing; Dec 2021:948-956.
22. Chen H, Hagedorn A, An N. The development of smart eldercare in China. Lancet Reg Health West Pac 2023 Jun;35:100547 [FREE Full text] [doi: 10.1016/j.lanwpc.2022.100547] [Medline: 37424692]
23. Wen W, Zhang Y, Shi W, Li J. Association between internet use and physical health, mental health, and subjective health in middle-aged and older adults: nationally representative cross-sectional survey in China. J Med Internet Res 2023 Mar 21;25:e40956 [FREE Full text] [doi: 10.2196/40956] [Medline: 36943368]
24. Liu J, Peng J, Chen M, Zhang T. Mediating and moderating effects of internet use on urban-rural disparities in health among older adults: nationally representative cross-sectional survey in China. J Med Internet Res 2023 Sep 28;25:e45343. [doi: 10.2196/45343]
25. Du X, Liao J, Ye Q, Wu H. Multidimensional internet use, social participation, and depression among middle-aged and elderly Chinese individuals: nationwide cross-sectional study. J Med Internet Res 2023 Aug 30;25:e44514 [FREE Full text] [doi: 10.2196/44514] [Medline: 37647119]
26. Li P, Zhang C, Gao S, Zhang Y, Liang X, Wang C, et al. Association between daily internet use and incidence of chronic diseases among older adults: prospective cohort study. J Med Internet Res 2023 Jul 17;25:e46298 [FREE Full text] [doi: 10.2196/46298] [Medline: 37459155]
27. Jin Y, Jing M, Zhang L, Song S, Ma X. Internet access and hypertension management among the elderly population: a nationally representative cross-sectional survey in China. J Med Internet Res 2019 Jan 31;21(1):e11280. [doi: 10.2196/11280]
28. Ding X, Yuan L, Zhou Y. Internet access and older adults' health: Evidence from China. China Economic Review 2023 Dec;82:102047. [doi: 10.1016/j.chieco.2023.102047]
29. Guo L, An L, Luo F, Yu B. Social isolation, loneliness and functional disability in Chinese older women and men: a longitudinal study. Age Ageing 2021 Jun 28;50(4):1222-1228. [doi: 10.1093/ageing/afaa271] [Medline: 33352582]
30. Leng A, Liu J, Maitland E, Li S, Nicholas S, Ma B, et al. Older adults preferences for long-term caregivers in China: a discrete choice experiment. Public Health 2024 Jun;231:158-165 [FREE Full text] [doi: 10.1016/j.puhe.2024.03.022] [Medline: 38692091]
31. Philip NY, Rodrigues JJPC, Wang H, Fong SJ, Chen J. Internet of Things for in-home health monitoring systems: Current advances, challenges and future directions. IEEE J Select Areas Commun 2021 Feb;39(2):300-310. [doi: 10.1109/jsac.2020.3042421]
32. Kickbusch I, Holly L. Addressing the digital determinants of health: health promotion must lead the charge. Health Promot Int 2023 Jun 01;38(3):1-3. [doi: 10.1093/heapro/daad059] [Medline: 37264549]
33. The Geneva Charter for well-being. World Health Organization. URL: https://www.who.int/publications/m/item/the-geneva-charter-for-well-being [accessed 2025-01-17]
34. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://tinyurl.com/3zachapy [accessed 2025-01-17]
35. Chidambaram S, Jain B, Jain U, Mwavu R, Baru R, Thomas B, et al. An introduction to digital determinants of health. PLOS Digit Health 2024 Jan;3(1):e0000346 [FREE Full text] [doi: 10.1371/journal.pdig.0000346] [Medline: 38175828]
36. Hong J, Shi F, Zheng Y. Does network infrastructure construction reduce energy intensity? Based on the “Broadband China” strategy. Technological Forecasting and Social Change 2023 May;190:122437. [doi: 10.1016/j.techfore.2023.122437]
37. Liu N, Cadilhac D, Kilkenny M, Liang Y. Changes in the prevalence of chronic disability in China: evidence from the China Health and Retirement Longitudinal Study. Public Health 2020 Aug;185:102-109. [doi: 10.1016/j.puhe.2020.03.032] [Medline: 32603874]
38. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb;43(1):61-68 [FREE Full text] [doi: 10.1093/ije/dys203] [Medline: 23243115]
39. Zhao Y, Atun R, Oldenburg B, McPake B, Tang S, Mercer SW, et al. Physical multimorbidity, health service use, and catastrophic health expenditure by socioeconomic groups in China: an analysis of population-based panel data. The Lancet Global Health 2020 Jun;8(6):e840-e849. [doi: 10.1016/s2214-109x(20)30127-3]
40. Deutrom J, Katos V, Ali R. Loneliness, life satisfaction, problematic internet use and security behaviours: re-examining the relationships when working from home during COVID-19. Behaviour & Information Technology 2021 Sep 11;41(14):3161-3175. [doi: 10.1080/0144929x.2021.1973107]
41. Zhang J, Cheng M, Yu N. Internet use and lower life satisfaction: The mediating effect of environmental quality perception. Ecological Economics 2020 Oct;176:106725. [doi: 10.1016/j.ecolecon.2020.106725]
42. Cheng X, Ge T, Cosco TD. Internet use and life satisfaction among Chinese older adults: the mediating effects of social interaction. Curr Psychol 2023 Feb 11:1-8 [FREE Full text] [doi: 10.1007/s12144-023-04303-y] [Medline: 36819750]
43. Lissitsa S, Chachashvili-Bolotin S. Life satisfaction in the internet age – Changes in the past decade. Computers in Human Behavior 2016 Jan;54:197-206. [doi: 10.1016/j.chb.2015.08.001]

XSL•FO
RenderX

44. Lam SSM, Jivraj S, Scholes S. Exploring the relationship between internet use and mental health among older adults in England: longitudinal observational study. J Med Internet Res 2020 Jul 28;22(7):e15683 [FREE Full text] [doi: 10.2196/15683] [Medline: 32718913]
45. Castellacci F, Schwabe H. Internet, unmet aspirations and the U-shape of life. PLoS One 2020;15(6):e0233099 [FREE Full text] [doi: 10.1371/journal.pone.0233099] [Medline: 32502145]
46. Evli M, Şimşek N. The effect of COVID-19 uncertainty on internet addiction, happiness and life satisfaction in adolescents. Arch Psychiatr Nurs 2022 Dec;41:20-26 [FREE Full text] [doi: 10.1016/j.apnu.2022.07.008] [Medline: 36428049]
47. Stepanikova I, Nie NH, He X. Time on the Internet at home, loneliness, and life satisfaction: Evidence from panel time-diary data. Computers in Human Behavior 2010 May;26(3):329-338. [doi: 10.1016/j.chb.2009.11.002]
48. Sun K, Zhou J. Understanding the impacts of Internet use on senior Citizens' social participation in China: Evidence from longitudinal panel data. Telematics and Informatics 2021 Jun;59:101566. [doi: 10.1016/j.tele.2021.101566]
49. Irina S, Robert K, Lee R. The internet and social participation: contrasting cross-sectional and longitudinal analyses. Journal of Computer-Mediated Communication 2004 Nov 1:1. [doi: 10.1111/j.1083-6101.2004.tb00226.x]
50. Zaine I, Frohlich DM, Rodrigues KRDH, Cunha BCR, Orlando AF, Scalco LF, et al. Promoting social connection and deepening relations among older adults: design and qualitative evaluation of media parcels. J Med Internet Res 2019 Oct 03;21(10):e14112 [FREE Full text] [doi: 10.2196/14112] [Medline: 31584001]
51. Griffith LE, Raina P, Levasseur M, Sohel N, Payette H, Tuokko H, et al. Functional disability and social participation restriction associated with chronic conditions in middle-aged and older adults. J Epidemiol Community Health 2016 Oct 17;71(4):381-389. [doi: 10.1136/jech-2016-207982]
52. Jespersen LN, Michelsen SI, Tjørnhøj-Thomsen T, Svensson MK, Holstein BE, Due P. Living with a disability: a qualitative study of associations between social relations, social participation and quality of life. Disabil Rehabil 2019 Jun 22;41(11):1275-1286. [doi: 10.1080/09638288.2018.1424949] [Medline: 29357697]
53. Heins P, Boots LMM, Koh WQ, Neven A, Verhey FRJ, de Vugt ME. The effects of technological interventions on social participation of community-dwelling older adults with and without dementia: A systematic review. J Clin Med 2021 May 25;10(11):2308 [FREE Full text] [doi: 10.3390/jcm10112308] [Medline: 34070660]
54. Benoit-Dubé L, Jean EK, Aguilar MA, Zuniga A, Bier N, Couture M, et al. What facilitates the acceptance of technology to promote social participation in later life? A systematic review. Disabil Rehabil Assist Technol 2023 Apr;18(3):274-284. [doi: 10.1080/17483107.2020.1844320] [Medline: 33156714]
55. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: A randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016 Oct 15;122(20):3166-3174 [FREE Full text] [doi: 10.1002/cncr.30172] [Medline: 27332968]
56. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: systematic review. JMIR Med Inform 2020 Aug 12;8(8):e20359. [doi: 10.2196/20359]
57. Yeroushalmi S, Maloni H, Costello K, Wallin MT. Telemedicine and multiple sclerosis: A comprehensive literature review. J Telemed Telecare 2019 May 01;26(7-8):400-413. [doi: 10.1177/1357633x19840097]
58. Kennedy J, Lynch H. A shift from offline to online: Adolescence, the internet and social participation. Journal of Occupational Science 2016 Jan 13;23(2):156-167. [doi: 10.1080/14427591.2015.1117523]
59. Rennoch G, Schlomann A, Zank S. The relationship between internet use for social purposes, loneliness, and depressive symptoms among the oldest old. Res Aging 2023 Jan 06;45(9-10):630-642. [doi: 10.1177/01640275221150017] [Medline: 36606365]
60. Schmidt D, Power SA. Offline world: The internet as social infrastructure among the unconnected in quasi-rural Illinois. Integr. psych. behav 2020 Aug 21;55(2):371-385. [doi: 10.1007/s12124-020-09574-9]
61. Lee H, Park N, Hwang Y. A new dimension of the digital divide: Exploring the relationship between broadband connection, smartphone use and communication competence. Telematics and Informatics 2015 Feb;32(1):45-56. [doi: 10.1016/j.tele.2014.02.001]
62. Costa RM, Patrão I, Machado M. Problematic internet use and feelings of loneliness. Int J Psychiatry Clin Pract 2019 Jun;23(2):160-162. [doi: 10.1080/13651501.2018.1539180] [Medline: 30570343]
63. Dalton F, Cassidy T. Problematic internet usage, personality, loneliness, and psychological well-being in emerging adulthood. Anal Soc Iss & Public Policy 2020 Dec 14;21(1):509-519. [doi: 10.1111/asap.12224]
64. Reinhardt G, Vidovic D, Hammerton C. Understanding loneliness: a systematic review of the impact of social prescribing initiatives on loneliness. Perspect Public Health 2021 Jul;141(4):204-213 [FREE Full text] [doi: 10.1177/1757913920967040] [Medline: 34159848]
65. Gerst-Emerson K, Jayawardhana J. Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. Am J Public Health 2015 May;105(5):1013-1019. [doi: 10.2105/ajph.2014.302427]
66. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39. [doi: 10.2196/jmir.2306]

XSL•FO
RenderX

67. Benda N, Veinot T, Sieck C, Ancker J. Broadband internet access is a social determinant of health. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/ajph.2020.305784]
68. Mbunge E, Muchemwa B, Jiyane S, Batani J. Sensors and healthcare 5.0: transformative shift in virtual care through emerging digital health technologies. Global Health Journal 2021 Dec;5(4):169-177. [doi: 10.1016/j.glohj.2021.11.008]

Abbreviations

AOR: adjusted odds ratio
CHARLS: China Health and Retirement Longitudinal Study
PRICSSA: Preferred Reporting Items for Complex Sample Survey Analysis

Edited by A Schwartz; submitted 07.04.24; peer-reviewed by Y Si, XF Zhong; comments to author 03.09.24; revised version received 29.10.24; accepted 24.12.24; published 10.02.25.

Please cite as:
Yang L, Lynch C, Lee JT, Oldenburg B, Haregu T
Understanding the Association Between Home Broadband Connection and Well-Being Among Middle-Aged and Older Adults in China: Nationally Representative Panel Data Study
J Med Internet Res 2025;27:e59023
URL: https://www.jmir.org/2025/1/e59023
doi:10.2196/59023
PMID:

©Lu Yang, Chris Lynch, John Tayu Lee, Brian Oldenburg, Tilahun Haregu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Competence of Arabic-Speaking Immigrant and Refugee Older Adults Enacting Agency and Navigating Barriers: Qualitative Descriptive Study

Jordana Salma[1], MN, PhD; Alesia Au[1*], BSCN, MN; Ghada Sayadi[1*], MA; Manal Kleib[1*], MSN, MBA, PhD

Faculty of Nursing, University of Alberta, Edmonton, AB, Canada

[*]these authors contributed equally

Corresponding Author:
Alesia Au, BSCN, MN
Faculty of Nursing
University of Alberta
Edmonton Clinic Health Academy (ECHA), 11405 87 Avenue NW
Edmonton, AB, T6G 1C9
Canada
Phone: 1 7804929469
Email: alesia1@ualberta.ca

Abstract

Background: Canada's immigrant and refugee older adult population is projected to grow substantially, making equitable access to information and communications technologies (ICTs) vital for enhancing quality of life in older age. Strengthening the digital competence of immigrant and refugee older adults can improve their social connectedness and access to local information.

Objective: This study explored the digital competence of Arabic-speaking immigrant and refugee older adults, focusing on how they engage with ICTs to meet their information and communication needs and the strategies they use to navigate digital barriers.

Methods: A qualitative descriptive methodology within a social constructivist paradigm was adopted, incorporating triangulated data collection and iterative co-design cycles. The qualitative approach facilitated an in-depth exploration of participants' experiences, skills, and emotions and the contextual factors influencing their digital competence. Data were collected through storytelling approaches, qualitative interviews, and focus group discussions, which were effective in capturing the experiential aspects of aging and technology use. Co-design cycles informed 6 digital learning sessions tailored to participants' immediate learning needs, fostering motivation and engagement and allowing for observation of ICT use. Digital competence was mapped across the learning domains of the Digital Competence Framework for Citizens 2.2.

Results: This study engaged 31 Arabic-speaking immigrant and refugee older adults residing in Canada. Most participants had limited formal education (19/31, 61%), lived with family (22/31, 70%), and reported a low income (21/31, 68%). All participants (31/31, 100%) used smartphones as their primary ICT device, whereas few (3/31, 10%) had access to a computer. In total, 3 themes were identified from the analysis, grounded in Digital Competence Framework for Citizens 2.2 competencies on information and data literacy, communication and collaboration, and safety and problem-solving. The themes focused on agency, which is enhanced or constrained using ICTs, impacting older adults' desire and ability to use these technologies to independently meet their daily needs.

Conclusions: Immigrant and refugee older adults require support to navigate digital barriers and gain digital competence. Smartphones serve as a critical tool for enhancing digital agency, which can lead to greater social connectedness and improved access to local resources in older age. The findings will inform the design of future digital competence programs for older migrants, emphasizing community partnership and reciprocal learning.

(J Med Internet Res 2025;27:e60547) doi:10.2196/60547

KEYWORDS

digital competence; Arab; older adult; co-design; migrant; information and communications technology; ICT

XSL•FO
RenderX

Introduction

Background

Information and communications technologies (ICTs) are defined as the spectrum of digital and social media used for communication and access to information [1]. Internet access rates have more than doubled for those aged ≥65 years in Canada [2]. Access to ICTs has been declared a fundamental human right, with strong evidence that the digital divide—lack of or limited access to the benefits of technology—impacts quality of life in older age [3]. ICTs provide several benefits for older adults, including social connectivity, opportunities for continuous learning, access to health-related information, and entertainment [4-8]. The benefits that older adults obtain from technology depend on the types of technologies adopted and their digital competence [9-12].

Digital competence is defined as "the confident, critical and responsible use of, and engagement with, digital technologies for learning, at work, and for participation in society. It includes information and data literacy, communication and collaboration, media literacy, digital content creation (including programming), safety (including digital well-being and competences related to cybersecurity), intellectual property related questions, problem solving and critical thinking" [13]. Digital competence is a newer concept in comparison to digital literacy, with both terms being used interchangeably [14] or viewed as different with some shared dimensions related to technological knowledge, skills, and attitudes [15,16]. Digital competence is a useful concept for older adults as it incorporates attitudes and motivations that are essential predictors of ICT use in older age [17-19]. Enhancing the benefits of using ICTs requires not only improving technological skills and knowledge but also understanding the attitudinal and socioemotional aspects of technology use [20]. Older adults are willing to adopt technology that is deemed beneficial to them and that has positive implications for their daily life [6,21,22]. Older adults use ICTs to maintain social connections, strengthen family ties, and support activities of daily living and leisure [22,23]. These digital tools can enhance older adults' independence, autonomy, and self-confidence [12,22]; however, the digital divide persists due to a range of factors, such as living with disabilities, lack of social support, lower educational level, and low income [12,24]. During the COVID-19 pandemic, physical distancing and the health risks to older adults necessitated creative digital solutions, such as ICT-mediated social and recreation programs, which reinforced the potential benefits of ICTs [25,26].

Improving digital access and competence is one strategy to support healthy aging in migrant populations. Canada continues to welcome high numbers of immigrants and refugees, with an increasing percentage from Arabic-speaking countries [27]. Younger cohorts of immigrants demonstrate high levels of proficiency in using technologies to adapt within postmigration contexts [28-30]. The extent to which this proficiency applies to immigrant and refugee older adults has been less explored despite well-documented ICT adoption to facilitate social engagement and sustain family ties [31,32]. Older Arabic-speaking migrants experience loneliness, social isolation, and barriers to access and use of aging-supportive services [33,34]. Arab migrants traditionally belong to extended family models and value close interpersonal relationships within local and transnational kin networks [35,36], with the adoption of ICTs influenced by social identities and roles [37,38]. In addition, gender, income, educational level, and technology infrastructure in their countries of origin shape the use of technologies and motivations to develop particular knowledge and skills [24,39,40].

To support digital competence and create avenues for digital learning, it is critical to understand the individual and social influences on ICT use among older migrants [41]. The lack of theorization linking aging and technology use is well cited in the literature [42], with little integration of immigrant and refugee older adult perspectives in gerontechnology [37]. This study aimed to explore the "messiness of practice" [43] that sheds light on the ways in which immigrant and refugee older adults interact with technologies in their daily lives to solve problems, their exercise of agency, and their digital competence.

Objectives

This study aimed to explore the digital competence of immigrant and refugee older adults with a specific focus on (1) the ways in which immigrant and refugee older adults engage with ICTs to meet their needs for accessing information and communication and (2) the strategies used to problem solve and navigate barriers to using ICTs.

Conceptual Framework

The Digital Competence Framework for Citizens (DigComp) 2.2 encompasses 5 key areas defining the scope of digital competence: information and data literacy, communication and collaboration, digital content creation, safety, and problem-solving [44]. The initial 3 domains focus on competencies associated with particular tasks and applications. The fourth and fifth domains, "safety" and "problem-solving," are relevant across all digital activities. Each competence area incorporates skills, knowledge, and attitudes that are essential for digitally literate citizens, with proficiencies that range from basic to advanced and that can be used to design competence assessment tools, support education, and tailor training [45]. This framework was selected for its comprehensiveness and its attention to attitudes across the range of competence areas that were deemed relevant to older adults in a growing technology-dependent world [5,46,47].

Methods

Research Design

A qualitative descriptive methodology within a social constructivist paradigm was used and included triangulated data collection methods and iterative co-design cycles with immigrant and refugee older adults to explore ICT use, digital competence, and related learning needs. A qualitative approach was selected to allow for in-depth experiential explorations of digital competencies. The goal was to conduct a holistic investigation into the ways in which older adults engaged with technologies, including their skills, knowledges, emotions, and influencing contextual factors. Triangulation of data sources over a period

XSL•FO
RenderX

is required to obtain descriptively rich insights on technology use in older age [48]. In addition, older adults respond best to storytelling approaches for data collection that are flexible and engaging, which makes qualitative interviews and focus group (FG) discussions most effective in expanding understandings of aging [49]. Co-design cycles informed a sequence of 6 learning sessions to address immediate learning needs identified over the course of data collection as older adults' motivation and interests are key drivers of digital engagement [12,20]. Reciprocity via supporting their immediate digital learning needs helped meet ethical obligations to the community and build sustainable partnerships [50].

Ethical Considerations

This study received ethics approval from the University of Alberta Research Ethics Board (ethics number Pro00103587). All participants provided informed consent before data collection and were informed of their right to withdraw at any time during the study. Identifying information was removed from interview and FG transcripts and from observation notes and then shared with the research team for analysis and knowledge dissemination. Participants were provided with an honorarium of CAD $50 (US $35.20) for participation in the data collection sessions.

Sample and Recruitment

Study participants were community-dwelling individuals who (1) were aged ≥55 years, (2) self-identified as first-generation Arabic-speaking migrants, and (3) reported using ICTs in their daily lives. We selected the age of ≥55 years to define "older adult" because some refugees in the Arab community identified as older adults at <65 years [51,52]. Despite the World Health Organization using the age of ≥60 years to define older adults, the definition acknowledges variations in life expectancy and cultural perceptions of aging worldwide [53]. Cohorts of Arab-Canadian migrants differ in levels of education, reasons for migration, religious traditions, and experiences of macro and micro historical and political events [54]. Most Arab older adults in this study were Syrian, Lebanese, Jordanian, or Palestinian; ascribed to the Muslim faith; and had varying levels of education, income levels, and English-language fluency. A social service organization that catered to newcomers and a mosque that included a congregation of Arabic-speaking older adults supported the project by hosting the data collection sessions and aiding recruitment. Flyers in English and Arabic were distributed by the 2 partner organizations, and staff helped explain the purpose of the study to potential participants. These organizations also provided space for the group-based digital learning sessions. Purposive sampling began with service providers at these organizations connecting with immigrant and refugee older adults via phone calls and in person to introduce the study. Immigrant and refugee older adults with interest in participating were then contacted by the research team for screening and formal consent to participate. Data saturation was determined by repetitive patterns of experiences, perceptions, and needs identified in the data over the course of the study.

Data Collection

Overview

Data collection included 2 phases. Phase 1 involved individual (n=8) and group (n=2) semistructured interviews to explore experiences of using ICTs. Phase 2 included digital learning sessions where observations and FGs facilitated a detailed documentation of immigrant and refugee older adults' digital competencies and learning needs. Triangulation of data sources over a period allows for access to descriptively rich insights on technology use in older age [22,38]. All individual interviews and FGs were audiotaped and transcribed verbatim by a professional transcriptionist trained in the Arabic language.

Semistructured Interviews

Individual interviews were conducted in participants' homes from April 2022 to June 2022 by an Arabic-speaking interviewer and lasted, on average, 1 to 1.5 hours. For the first phase, an interview guide using the DigComp 2.2 guided the types of questions asked (Figure 1). Learning session topics were derived from the priority areas identified in the phase-1 interviews and included learning about the essentials of smartphone hardware and software, social media communication tools such as Facebook and WhatsApp, and using email and online access to information.

XSL•FO
RenderX

Figure 1. Phase 1 interview questions from the Digital Competence Framework for Citizens 2.2.

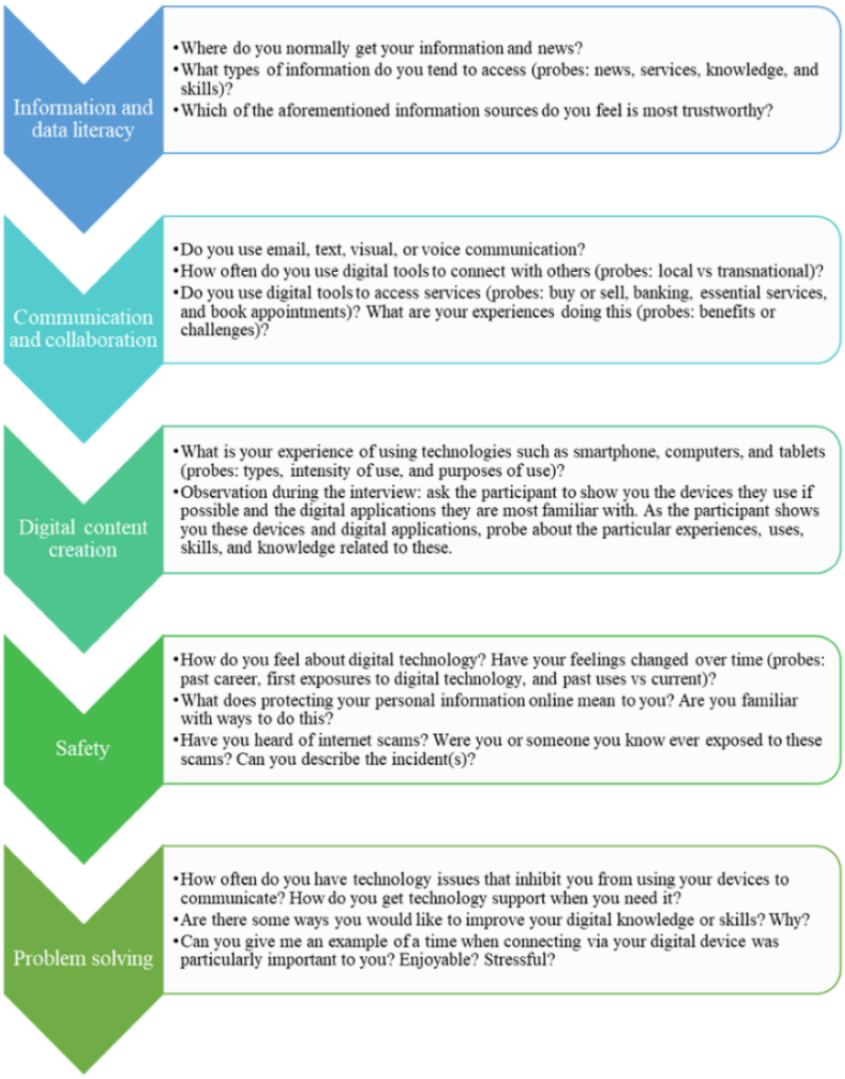

Co-Design Sessions

In phase 2, we used a participatory framework [12,20] to co-design ICT learning sessions between September 2022 and November 2022. The session instructor was a computer engineering university student, and he was supported by an Arabic-speaking research assistant who translated for participants who only spoke Arabic. The participatory co-design framework facilitated collaborative approaches that actively involved older adults in the design and decision-making processes of the learning sessions to ensure that the sessions were engaging, built on the lived experiences of participants, and met their self-identified learning needs [12,20] (Au et al [55] reported in detail on the co-design process). The project lead, who mentored the session instructor and research assistant, had >10 years of experience in participatory research methods and provided training on co-design processes during weekly team meetings. In each session, participants started with a 30- to 45-minute FG discussion related to the main learning topic of the day to identify the ways in which they engaged with that particular ICT in their daily lives; the challenges they faced; and their attitudes, emotions, and learning needs (Figure 2). Phase 2 FGs were used to deepen insights from phase 1, triangulate data for validation, tailor learning sessions to participants' needs, and engage them in identifying their learning requirements and experiences with digital technologies. Feedback on each learning session helped the instructor refine subsequent session delivery and content. The purpose of this study was not to develop or test a complete digital learning program; rather, the purpose was to create a supportive environment where immigrant and refugee older adults had some direct benefit from participating in the research study while the research team gained in-depth understanding of their digital practices. Participants for these sessions were divided into 3 groups, with separate men's (3/21, 14%) and women's (10/21, 48%) groups in the faith-based organization and a mixed-gender group (8/21, 38%) in the social service organization. Observations (54 hours of observational data) that occurred during these sessions entailed a researcher documenting group learning processes and the digital competencies of participants as they navigated their devices and worked together through lesson material.

XSL•FO
RenderX

Figure 2. Phase 2 focus group guiding questions.

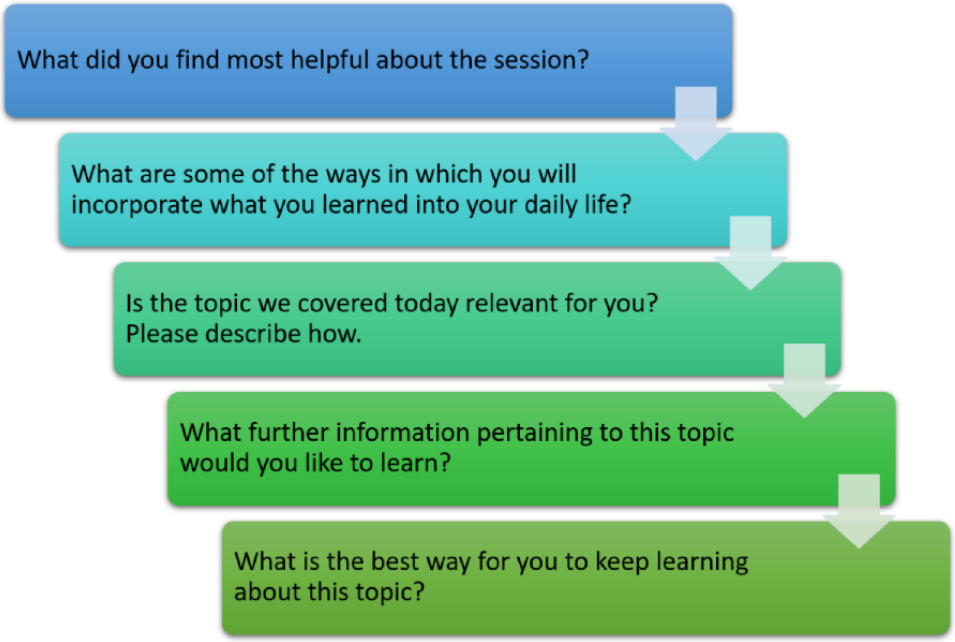

Data Analysis

Arabic conversations were initially analyzed by bilingual team members, with key quotes and codes then translated into English for team discussion and use in knowledge dissemination. A reflexive thematic data analysis approach was used to identify, analyze, and report patterns within the data [56] facilitated by the NVivo qualitative data management software (QSR International). The 6-phase approach included familiarizing ourselves with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and writing a report [57]. Initial codes were organized around the DigComp 2.2 competencies, followed by linking and comparing codes within and across competencies to arrive at overall patterns of understanding about participants' experiences of engaging with and learning about ICTs (Figure 3).

To ensure trustworthiness in data analysis, we used peer debriefing, data triangulation, and reflexivity [58]. Peer debriefing provides external validation, which was done through weekly team conversations to discuss the ways in which observation, FG, and interview data reflected particular competencies. Data were analyzed iteratively, meaning that the data were coded systematically once collected and then recoded as codes became more refined over time with deeper understanding. Team consensus was reached with a topical coding tree based on the DigComp 2.2 competencies. One team member (JS) led the refinement of coding to develop the final themes for this paper, and all coauthors provided critical feedback at multiple points during this process. The authors have expertise in qualitative methods, aging, and migration. In total, 2 authors are bilingual Arabic-speaking academics, with the remaining authors being second-generation or first-generation immigrants to Canada and academic researchers.

XSL•FO
RenderX

Figure 3. Digital competencies and thematic analysis process. DigComp: Digital Competence Framework for Citizens; ICT: information and communications technology.

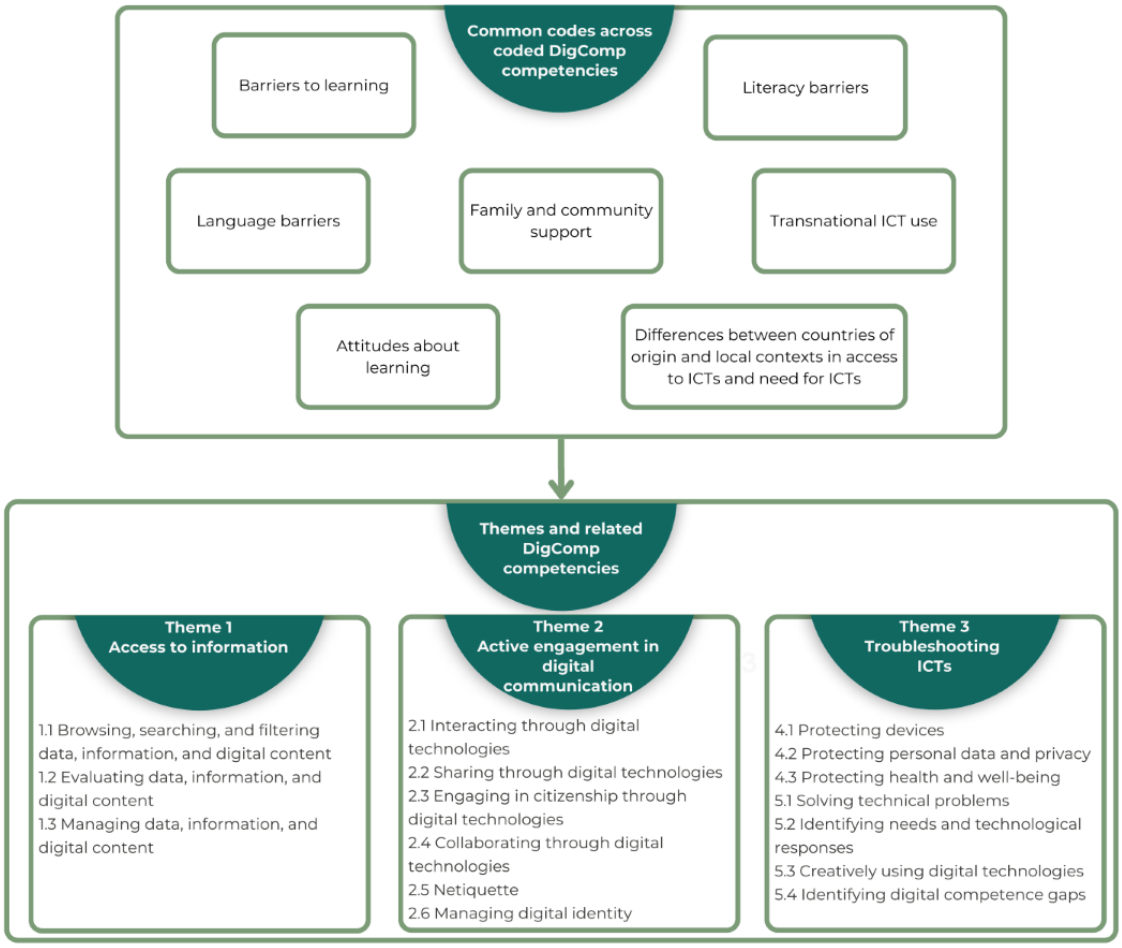

Results

Overview

A total of 31 immigrant and refugee older adults participated in this study, with an average age of 64.1 (SD 7.7) years. Most participants were women (20/31, 65%), identified as Arab (29/31, 93%), and were immigrants (19/31, 61%) or refugees (10/31, 32%). The highest level of education was primary or secondary school for most participants (19/31, 61%). In total, 2 non–Arabic-speaking participants were included because they were spouses of Arabic-speaking participants attending the program. More than half of the participants lived with their spouse, their children, or both (22/31, 71%); reported a family income of <CAD $20,000 (US $14,078.30; 21/31, 68%), which is categorized as low income in Canada; and received some form of government financial support (21/31,68%; Table 1).

All study participants (31/31, 100%) used smartphones as their primary ICTs, with only a few (3/31, 10%) reporting having access to and knowing how to use computers. Smartphones were cost-effective, accessible, and familiar to immigrant and refugee older adults, and hence, the focus of the learning sessions throughout this study was on exploring ICT use through these devices. The themes highlight the ways in which agency is enhanced or constrained using ICTs and draw attention to older adults' desire to use these technologies independently to meet their daily needs. The 3 themes were developed from across four dimensions of the DigComp 2.2: (1) information and data literacy, (2) communication and collaboration, (3) safety, and (4) problem-solving. The dimension of "digital content creation" was not inductively discussed by participants, and the scope was beyond the capacity of this study to address (Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Demographic characteristics of the participants (N=31).

Characteristics	Overall	Phase 1 (n=10)	Phase 2	
			Organization 1 (n=9)	Organization 2 (n=12)
Age (y), mean (SD)	64.1 (7.7)	61.2 (3.1)	64.2 (6.0)	66.3 (9.5)
Gender, n (%)				
Women	20 (64)	5 (50)	5 (56)	10 (83)
Men	11 (36)	5 (50)	4 (44)	2 (17)
Country of birth, n (%)				
Lebanon	5 (16)	0 (0)	1 (11)	4 (33)
Syria	16 (52)	5 (50)	6 (67)	5 (42)
Palestine	3 (10)	1 (10)	0 (0)	2 (17)
Sudan	3 (10)	3 (30)	0 (0)	0 (0)
Egypt	2 (6)	0 (0)	1 (11)	1 (8)
Afghanistan	1 (3)	0 (0)	1 (11)	0 (0)
Iraq	1 (3)	1 (10)	0 (0)	0 (0)
Dwelling, n (%)				
Spouse	10 (32)	1 (10)	6 (67)	3 (25)
Children	2 (6)	1 (10)	1 (11)	0 (0)
Spouse and children	10 (32)	5 (50)	0 (0)	5 (42)
Alone	6 (19)	1 (10)	1 (11)	4 (33)
Other	3 (10)	2 (20)	1 (11)	0 (0)
Family income, n (%)				
<CAD $20,000 (US $14,078.30)	21 (68)	7 (70)	7 (78)	7 (58)
CAD $20,000 to CAD $40,000 (US $14,078.30 to US $28,156.60)	7 (23)	1 (10)	2 (22)	4 (33)
CAD $40,000 to CAD $60,000 (US $28,156.60 to US $42,235)	2 (6)	1 (10)	0 (0)	1 (8)
Unknown	1 (3)	1 (10)	0 (0)	0 (0)
Time lived in Canada (y), n (%)				
<5	12 (39)	5 (50)	2 (22)	5 (42)
5-10	8 (26)	3 (30)	4 (44)	1 (8)
>10	11 (36)	2 (20)	3 (33)	6 (50)
English or French fluency, n (%)				
Excellent	3 (10)	2 (20)	1 (11)	0 (0)
Good	13 (42)	3 (30)	4 (44)	6 (50)
Minimal	14 (45)	5 (50)	4 (44)	5 (42)
None	1 (3)	0 (0)	0 (0)	1 (8)
Highest educational level, n (%)				
Primary school	9 (29)	1 (10)	2 (22)	6 (50)
Secondary school	10 (32)	2 (20)	4 (44)	4 (33)
Postsecondary school	11 (36)	7 (70)	2 (22)	2 (17)
Other	1 (3)	0 (0)	1 (11)	0 (0)

XSL•FO
RenderX

Theme 1: "Here You Need to Go on the Website and See for Yourself" (Access to Information)

All participants described using their smartphones to access information via social media platforms, with WhatsApp and Facebook being the most used, followed by YouTube. Search engines such as Google were rarely described as information sources by participants. All participants were skeptical about the credibility and authenticity of the information on social media. Participants most often used family and friends for information checks to confirm or disconfirm what they viewed on these platforms:

> *Yes, I subscribed with (news media channel) two or three times but I've seen fake news on there because we hear things and they broadcast different and fake things. We ask our friends and family in Iraq if the stuff they talked about has really happened there and they say "no," so all their news is fake.* [P5; older man; interview]

Immigrant and refugee older adults were savvy in evaluating information, especially with regard to world events. The information accessed related to Canada and their countries of origin as they followed ongoing political and social events. Accessing trustworthy information was heavily reliant on connections within transnational social networks that were made possible via the use of ICTs to communicate at a distance via free social media apps that allowed for texting, voice messages, and phone calls to gather information in real time:

> *Nowadays, to get the right information, sometimes I have to check more than one source and I compare the information...I exchange information with my friends and see if it's accurate. See what they've heard, so they tell me just to try to get the most accurate information...some [friends] are here, and some are back home. It depends. Some in Egypt, some in a different country.* [P12; older man; interview]

Beyond social and political news, participants used ICTs to search for information related to their hobbies and religious practices and to resources for immigration and health information:

> *For entertainment on YouTube I go and see how to bake a cake, I tried banana bread, amazing. I make things if I like the recipes.* [P5; older woman; FG]

Participants often received links to videos and images on social media that allowed them to explore a range of topics of interest, and many reported high daily uses for entertainment and information access:

> *Sometimes you will find it [health information] on YouTube and sometimes on Facebook. I type the name of the doctor on YouTube and I listen to what he says about health and about medicinal herbs...if I think the information is correct I follow it and if I do not think it is correct I do not follow it.* [P1; older woman; FG]

Refugee newcomers struggled as they quickly realized that they were expected to access online information needed for settlement. These participants were socially isolated during the COVID-19 pandemic and struggled to use email to access information critical to their daily lives, such as opening newsletters, paying utility bills, reading and responding to government emails, and accessing online banking:

> *No, so they don't know anything about accessing health information or banking information online, or any essential services, my understanding.* [Translator for P1 and P2; older couple; group interview]

While newcomers are connected to a settlement worker in Canada for their first year, this support was not always consistent or available in a timely manner or did not continue for a sufficient length of time:

> *Even here when you go to a government office to do anything, for example, they give you a website and ask you to see for yourself...in Syria, they would do that for you, if you needed anything, government employees would communicate clear instructions and you would get it done. Here you need to go on the website and see for yourself and this is hard for seniors, other than the language problem, they don't even know how to open or go on a website. For example, even though I am able to rely on myself, I still find it difficult because of the language. Even when I go on certain websites, their website design is sometimes a little different in comparison to what I was used to.* [P12; older man; interview]

All participants spoke Arabic, Kurdish, or Farsi as their native language, displaying diverse levels of proficiency in both written and spoken English. Language posed a challenge to browsing, searching, and filtering digital content in English. Consuming information via social media and specifically audiovisuals was easier for participants. Accessing information via ICTs was challenging not only due to lower digital competence and language barriers but also due to skepticism about the trustworthiness of the information and the applicability of this information to their unique situations. Furthermore, newcomers struggled with engaging with ICTs after migration due to the difference between technology in Canada and in their home country as they had shifted into a digital world later in life driven by the necessities of migration.

Finding trustworthy sources of information can be difficult for isolated immigrant and refugee older adults who do not have strong local social networks. Language barriers prevented accessing local trusted sources of information, such as health care practitioners and service providers:

> *There are issues with accessing banking or health information because often we opted to ask people around us locally, but have found that the information that we gathered could be inaccurate and the wrong information can spread quickly.* [P6 and P7; older couple; group interview]

Finding these trusted individuals was especially challenging for newcomers who might not know how to find a physician who speaks their language or might be unaware of older adult service providers in their local region. Arabic social media groups and

XSL•FO
RenderX

social media influencers who shared information after migration about settlement in Arabic were especially helpful:

> *In relation to the YouTube English lessons, someone who used to be an engineer in Syria, he started making YouTube videos in Edmonton that is English and Syrian translation. This is the content that they watch and he get the video link through WhatsApp through his cousin. They shared that thanks to the YouTube algorithm, they can watch the rest of his videos on his channel.* [P7; older man; interview]

Immigrant and refugee older adults who had English-language fluency or strong social connections within local Arabic-speaking communities were better connected to needed information and did not report these challenges. These participants reported more nuanced challenges such as storing and retrieving information:

> *I don't know how to [retrieve and save information that I have accessed]...I don't know where to save it.* [P2; older man; FG]

Others struggled with understanding search algorithms and the ways in which data were filtered by social media platforms, which caused frustration and inhibited access to preferred information:

> *P3 has a question about how to cancel certain posts that she sees on Facebook. She is confused as to why they show up when she has never pressed "like" on these posts.* [Older woman; observation]

In summary, immigrant and refugee older adults described a high level of skepticism about the information they found on social media, but often this was the most accessible source of information for those who were isolated with low digital competence and language barriers. Immigrant and refugee older adults were more comfortable navigating social media platforms in Arabic, and the availability of voice and video options supported those with low reading literacy. This meant that immigrant and refugee older adults were subjected to social media trends and algorithms that might bias what information they saw. This was countered by a reliance on trusted social contacts locally and transnationally to verify information from these digital sources. This coping strategy was ineffective in local contexts, where newcomers with few social contacts lacked access to reliable and needed information about their health and aging-supportive resources after migration.

Theme 2: "We Are Mute" (Active Engagement in Digital Communication)

A variety of platforms were used to meet participants' needs to communicate with family, friends, and others in their social networks. The most common social media platforms used were WhatsApp and Facebook. For participants who were illiterate or had low literacy, the voice message and video features of these platforms allowed them to engage with others. Some individuals used Zoom for group lessons and meetings with family. Email was the least used communication tool. Having left Syria before electricity and the internet were widely available in their village, one participant explained how the need and opportunity to use social media to connect with family began with their migration journey to Turkey—"We didn't know Facebook in Syria not even WhatsApp...when we went to Turkey we started a little to learn" (P1; older man; FG)—whereas the use of email followed to communicate with immigration officials and service providers in Canada. Responding to emails can be particularly important to access government social and health services, but a combination of language barriers and digital competence struggles prevented some newcomer refugees from using this communication tool:

> *P1 after seeing this slide has mentioned that she gets email messages and she doesn't understand, she doesn't open the message and it is left unread. Doesn't know how to respond to voicemail.* [Older woman; observation]
>
> *Participant knows how to open Arabic emails and uses it to track for school or government, doesn't send emails himself due to English barrier, and he communicates through social media more.* [Older man; observation]

Agency to create a digital identity and manage communication with others was important for participants. Some participants only received communication and were not able to reciprocate or initiate communication, which was a source of frustration:

> *I have been here 46 years, I went to school, I went to high school, I went to university, I did my education. At the time, I did all those but no technology was at the time, and now I have been retired for 10 years. Like on Facebook, if somebody sends me a message, I go back and answer them but for me to open my telephone and call them or send them a message, I do not know how.* [P8; older woman; interview]

Similar comments were made by women in the study, who reported wanting to learn to use email to expand beyond the usual familiar tools:

> *To enter into society...meaning if my friend sends me an email, I want to learn to respond to her...we read the message but we don't know how to respond.* [P6; older woman; FG]

Other participants reported communicating by proxy, where family members navigated their devices for them, responding to emails for communication with service providers and initiating contact with family and friends:

> *Sometimes there are things like the skype and internet-related things like video calls when I talk with my kids, my wife helps me with that. I am not very good with technology because we're a bit old school.* [P10; older man; interview]

While this support was appreciated by some, most reported wanting independence to alleviate the burden on their children. There were frequent comments about frustration with not keeping up with the times, being behind their grandkids in understanding ICTs, and not being able to maximize the use of their devices:

> *...but how much help from your kids! You know they are always busy, especially for mama, ok later later,*

XSL•FO
RenderX

> *tomorrow, tomorrow, and then tomorrow they never showed up for me.* [P8; older woman; FG]

For most, the lack of agency was revealed in observations during learning sessions despite their assertion initially that ICTs were accessible and had a high degree of integration into their daily lives. Immigrant and refugee older adults had difficulty creating digital content, such as initiating picture sharing, starting conversations, and deleting posts. Most participants struggled with these skills and required assistance to learn how to complete them during the learning sessions:

> *P1 knows how to share existing posts, is unable to save her own picture and share it.* [Observation]

This resulted in engagement being one-sided, eliminating the participants' autonomy to connect on their own terms to support preferences and needs. Passive use of ICTs, where participants observed but did not comment or create digital content, was common:

> *I access Facebook daily. I do not participate much but I benefit from world news, political news, religious news.* [P8; older woman; FG]

> *A lot of the participants don't post, rather they are looking at posts.* [Observation]

For those who had basic competence levels to communicate by posting, sending voice or text-based messages, and responding to social media posts, the more technical features of these applications were cited as challenging, such as deleting posts, adding friends, hiding posts, and creating private groups. Learning to block others, create private groups for discussion, and add or delete contacts is essential for agency in digital spaces:

> *P3 mentioned that she only blocks people on her phone, not on Facebook. She thinks it is really important to learn how to do this. Going over the Facebook slide, P4 wants to learn more about it...She knows how to add people and to search for content. She doesn't know how to block people on Facebook.* [Observation]

Transnational communication was prevalent within this population, and many participants connected with family, friends, and acquaintances overseas, with the lack of digital skills translating to some evidence of gendered power inequities, where women reported not being able to filter information or restrict access by families within transnational spaces:

> *P5 has a large friend request list containing her in-law side of the family. Although P5 knows them, she didn't want to add them as her friends on Facebook. She wants to keep her privacy, and noted that her husband has a large family (largest family in town), and she doesn't want them on her Facebook. But G. noticed that she also doesn't really post. H. wonders if once we teach them how to manage privacy settings on posts, maybe she may want to post more and may feel more comfortable to do so.* [Observation]

Communication via ICTs was an ongoing feature of participants' lives, and for some who were physically more isolated, it was the main source of daily communication. While, on the surface, it appeared that immigrant and refugee older adults had a high degree of competence due to their intense reliance on ICTs, it became apparent that agency was constrained due to limited control over how, when, and what was communicated. Creating digital content, communicating with preferred social contacts, and using communication to access local sources of support were major challenges.

Theme 3: "Don't Give Me a Fish, but Teach Me How to Catch It" (Troubleshooting ICTs)

Safety online and troubleshooting to address issues such as storage, function changes, and device accessibility were challenges for immigrant and refugee older adults. Safety related to sharing information online, password management, and privacy in online communication. Participants relied on practical approaches to store their passwords, such as digital word documents, notebooks, or family members. Learning sessions were often interrupted to troubleshoot password and security functions of their devices:

> *Their bewilderment and lack of awareness says a lot—they're learning that their children have set up a lot of the functions on their phones for them...so the participants never needed to deal with the set-up process and account/password management.* [Observation]

One participant described using the Notes app on their phone to store their passwords but was fearful as to what would happen to their digital identity if they misplaced their phone:

> *She has issues with remembering her passwords. She had written down her passwords in her phone but when she lost her phone, she couldn't recover her passwords.* [P6; older woman; FG]

Participants had a difficult time creating strong passwords and tended to use the same password across applications. They reported trouble remembering passwords and were required to use the password reset feature, which posed additional complexities. One participant felt overwhelmed with the number of passwords that are required to use ICTs and the need to synthesize complex passwords:

> *Yeah, I can't deal with, you know, like if I want to search about something on a Google, I pray, I go, or if they ask me to put my password, probably I have 200 passwords and I don't know which one. You know, like I'm not capable to action and I write down my password. I write down my email, whatever, and the next day I forgot something or some letter.* [P10; older man; interview]

Generally, the participants were well equipped to deal with security issues. There were instances in which participants could recognize scam calls and phishing emails. They were hesitant about making new friends online, joining Facebook groups that they did not recognize, and sharing photographs online. This hesitance was outlined by one participant:

> *You don't trust like on the Internet to find somebody new, or you don't know nothing, and you don't know*

XSL•FO
RenderX

> *if he's a con artist or thieves or, you know, eat your life. I don't like to meet those people that I don't know them. But if I knew a neighbour, you know, "Good morning, good afternoon." She visits me, I visit her. Yeah, I would like to communicate and be a friend and help her and she help me, accepted. But somebody like introduced, met on a—or I met on the Internet, and I don't believe those to be follow.* [Older woman; FG]

Instances of scams were not described by participants except for one who talked about learning to avoid accepting phone calls from unknown numbers. While most participants reported not responding to calls from individuals whom they did not know, this participant described in detail learning to be safer using ICTs after falling victim to a scam:

> *In the beginning, I have many problems with this. Sometimes I answer and I give my, you know, my information. Yeah, because, when I came here, I thought that it's trustworthy...this is a country of laws and never imagined this type of stuff would happen here to be honest. I had a few issues and I even had to cancel my credit card and so I learned, so now, whatever I don't understand, I don't engage in...Sometimes I don't even answer and I miss appointments... Sometimes they are calls from Canada that are important for me and I miss them.* [P12; older woman; interview]

Privacy when using digital devices was of concern for most participants, who tried to navigate this by not sharing or posting on social media and not adding people who were not part of their social network:

> *Daily I get friend requests...people I don't know so I ignore them completely.* [P3; older woman; FG]

Privacy for women related to religious and cultural expectations of modesty and to perceptions of avoiding envy or gossip. Simply not posting was easier versus restricting privacy settings so only certain people could view the posts. According to one participant, "I do not post my pictures on Facebook" (P4; older woman; FG). Another participant agreed and added that "because this is something private, it is not necessary for the whole world to know about it" (P3; older woman; FG). A group of women discussed an incident in which a video of their social gathering, involving clapping and singing, was shared without the consent of one of the participants on social media. Due to religious views about modesty, the women requested that this video be removed. The conversation centered on the importance of privacy and consent when posting content on public forums:

> *It was a misunderstanding...we let our guard down with our friends...she was afraid and didn't want this to spread.* [P4; older woman; FG 3]

However, often, immigrant and refugee older adults were unaware of how to navigate privacy settings on different devices and platforms:

> *Activity on how to make account more private: Participants needed to be verbally and individually guided to find the settings menu for privacy. All participants changed their settings to be more private for friends' requests, even when facilitators described it in a neutral way.* [Observation]

Troubleshooting for accessibility of devices was an issue in that not everyone knew the possibilities with Google Translate and other apps. Many participants were only literate in their first language, Arabic, and some were illiterate in Arabic, so they used their phones to watch and listen but not to read:

> *The message comes to us via phone and we understand a little of it but we then take it to someone who reads English, a friend...I send it to him via phone and he tell me what it is about.* [Older man; FG]

The use of text-to-voice apps and translation functions was not described by most participants outside of receiving and sending voice messages using social media platforms. Most participants who had family living with them or nearby used them for support with troubleshooting their devices. Immigrant and refugee older adults all preferred to learn how to troubleshoot on their own without relying on family and wanted community supports to help expand their digital competence. One participant shared that "They don't have time" (P6; older woman; FG). Another participant agreed and shared that family members lacked patience as "They take the device from you, do this and that, okay, how did you do that? How will you learn it if you are not doing it? Don't give me a fish, but teach me how to catch it" (P1; older woman; FG).

Some immigrant and refugee older adults did not have family or friends nearby or lived alone:

> *I don't have somebody live with me, so when I stuck, I get so upset so that's why I never try again.* [P4; older woman; interview]

These individuals reported frustration and a desire to access support to enhance their digital competence, citing this as a main reason for joining the digital learning sessions. Participants rarely reported accessing services or programs that help with digital competence, especially those who were newcomers to Canada and felt unfamiliar with avenues for this type of support. None reported using the library or other mainstream technology support programs for older adults. Local ethnocultural and faith-based organizations and acquaintances were cited as helping at times:

> *...there is no one to help us, we do not have children here, they are all in Syria and Jordan and Lebanon...We take it to the mosque or to someone we know.* [Older men and women; FG]

Overall, immigrant and refugee older adults experienced significant challenges navigating the technical features of their devices, which, in turn, restricted the scope and range of digital engagement. Assistance from family, when available, was often a source of support to use ICTs. Immigrant and refugee older adults preferred to expand their knowledge and skills to gain further independence in navigating their devices. Participants did not report enrolling in any formal digital learning programs before taking part in this research project.

XSL•FO
RenderX

Discussion

Principal Findings

This study explored the ways in which immigrant and refugee older adults use ICTs in their daily lives, their sense of agency, and their digital competence. This study addressed a significant gap by focusing on digital competence among Arabic-speaking older adults, a population largely overlooked in existing research. While the DigComp 2.2 framework has seen limited application with older adults, its detailed competency structure offers valuable insights into learning needs and serves as a robust tool for future research and interventions. The findings highlight how personal digital competence influences older migrants' engagement with ICTs for local and transnational purposes, providing actionable insights for designing interventions or exploring specific competencies. In addition, digital agency emerged as a critical yet understudied construct, particularly within the context of aging and migration, warranting further exploration.

The sociotechnical dimensions of ICT are described as the broader social practices and structures within which immigrant and refugee older adults are embedded and which shape their ICT use and choices [59,60]. Participants primarily relied on social media platforms to access information, entertainment, and social connections. Similarly to other migrant populations, Arabic-speaking older adults use a range of technologies to maintain transnational social networks, cultivate belonging to places of origin, practice valued social roles such as grandparenting, and navigate barriers to local integration [38,61,62]. Observing older adults use ICTs during the learning sessions was a rich data source, highlighting their creativity and agency despite constraints imposed by low digital competence.

Agency relates to using ICTs in ways that meets one's preferences and needs [5]. Agency has been a core concept in understanding the interplay of internal and external factors that shape technology adoption and the impacts of technology use [60]. Participants in this study struggled to meet their needs via ICTs and, simultaneously, exercised agency to overcome digital barriers. Digital use was on a spectrum from low to advanced agency in managing digital tools to meet needs and preferences during and after settlement in Canada. Constrained agency was most evident when lower digital competence was coupled with minimal environmental and social support to engage with ICTs. This was demonstrated when participants could not search for needed information on online platforms, were unable to control who they communicated with via blocking a contact or creating a private friend group, and stopped using particular apps or phone features because they did not know how to troubleshoot their devices. While immigrant and refugee older adults lacked individual digital competence (microlevel barrier), they were also disadvantaged by technosocial systems that were not designed with and for older migrants (system-level barrier).

Digital competence in searching for and filtering information was limited, especially when seeking information related to settlement or aging-supportive resources. This study showcases multiple exemplars of unsupportive technosocial systems for older migrants. Information about services for older adults is increasingly digitized, which makes it inaccessible to those with poorer digital and other literacies. A mismatch was evident between what services were available and what immigrant and refugee older adults had access to, which is mirrored in other studies on service access barriers [37]. Arabic translations of social and health service information are sparse in Canada despite the Arabic-speaking demographic increasing steadily over time. While immigrant and refugee older adults preferred information shared through trusted social networks, either in person or ICT mediated, newcomers had limited access to this avenue after migration. This was especially evident for refugee newcomers who had arrived during the pandemic, which mirrors what has been reported elsewhere [63,64]. While Arabic social media platforms are readily available, participants felt that the information was not always trustworthy or tailored to their local needs. Technologies, even when not adopted by older adults, influence their lives by changing the social world around them [59]. Study participants were deprived of local knowledge due to limited access to government websites and difficulty searching for online information in English, whereas the Arabic social media platforms that they did use were not specifically oriented to older adults living in Canada. The example of an Arabic-speaking social media influencer who provided local knowledge of the city showcases the potential for social media as a place for effective knowledge dissemination to older migrants.

Despite constrained agency, immigrant and refugee older adults were not always passive ICT users but rather adapted to, overcame, or disengaged from these technologies [59]. Participants disengaged from some ICT-mediated activities such as sharing content on social media because of privacy concerns and their lack of competence in controlling privacy settings or creating private groups. Participants used creative ways to store information, such as sending the information to a peer via WhatsApp so that it was stored in the conversation thread. "Hybrid practices," where technology is used to adapt ways of doing things [5], were also evident, such as accessing information via social media and then cross-checking accuracy with a trusted social network. Stereotypes of older adults influence the technologies that become available and adapted for them and also shape older adults' self-perceptions of their digital capabilities [12,24,65]. It is critical to challenge stereotypes of older migrants as nonusers and passive users of ICTs to ensure that these technologies continue to adapt while considering older adults as a key user demographic.

Family is a central source of support for older adults' adoption and navigation of new technologies [21,41], which mirrors the findings of this study. However, participants wanted to improve their digital competence to navigate ICTs independently and were engaged in the learning sessions. Challenges with troubleshooting their devices were a cause of frustration, and some participants reported an overreliance on family to facilitate technology use, which was not always effective in improving their digital knowledge and skills. Other studies have shown the importance of social support from family and friends in facilitating older adults' access to and use of technologies [12,24], and this holds true for immigrant and refugee older adults. Family members of some immigrant and refugee older

XSL•FO
RenderX

adults may also have limited digital competence, or immigrant and refugee older adults might be socially isolated without family support. Offering digital learning opportunities to increase their competence is a way to enhance digital agency [24]. Older adults might use technologies for necessity or leisure, so developing safe and engaging learning environments that meet their preferred needs is critical [12,21,22].

This study expands on existing literature on older migrants and ICT use [38,62,66] by exploring the digital competence of Arabic-speaking immigrant and refugee older adults who creatively circumvent challenges as they navigate the digital divide [24]. Migration shapes motives and needs for ICT adoption, with particular challenges arising for older migrants. Older migrants desire to stay connected with those left behind in their countries of origin, and they need to access local sources of support in their new country. ICTs will continue to play a vital role in communication patterns and information sharing for this population. Effective digital learning programs must be designed to overcome the intersecting barriers of age-related disabilities, cultural differences, language barriers, and low literacies. Canada has digital competence programs that target underserved older adults, such as the Digital Literacy Exchange Program [67] and the SUCCESS Digital Literacy Exchange Program for newcomers [68]. Participants in this study did not report accessing digital competence programs outside of the research study. While agism is often associated with the digital divide [24,59,65], digital exclusion based on language and language minority status must also be addressed when considering the opportunities available to older adults. Digital learning programs that integrate elements of English-as-an-additional-language learning, focus on problem-solving and creative uses of technologies, and tailor learning outcomes to the pragmatic needs of older migrants for transnational communication and local integration are likely to be the most successful. Finally, a discussion of ICT design is beyond the scope of this paper, but digital technologies that are acceptable to older adults and meet their needs, including for older adults with language and literacy barriers, are much needed as this demographic continues to grow worldwide.

Strengths and Limitations

This study allowed for an in-depth exploration over a short period of Arabic-speaking immigrant and refugee older adults' digital competence. A longitudinal design over a longer period could add further insights into evolving digital competencies in migrant older adults. Most study participants (20/31, 65%) were women, and hence, we were unable to explore gender differences. Some exceptions in this study related to women who reported constrained communication choices within transnational families and the importance of privacy in social media interactions. Other studies have shown how gender shapes the use of and access to technologies [40,69-71], and this is an area for further exploration. Most participants were healthy and younger older adults. Digital technology use and health are linked [12,72]; hence, further exploration of technology use in frail older migrants is required to meet their needs for social connectivity and information access. With the central role of social support and family in ICT adoption, future studies that include the perspectives of family members and caregivers might be critical in designing digital interventions for this population. Finally, the focus on Arabic-speaking older adults means that other populations of migrants might have differing experiences that warrant investigation.

Conclusions

Canada's immigrant and refugee older adult population will continue to grow, and addressing equitable access to ICTs is imperative to enhance quality of life in older age. Strengthening agency to use ICTs has the potential to markedly improve their social connectedness and access to local information. Digital competence education tailored to the needs of older migrants can empower active participation in the digital world and meet settlement needs after migration. These programs would need to focus on the competencies of communication and information access with attention to language preferences, transnational ICT-mediated practices, different types of literacies (eg, basic and informational), and creative troubleshooting strategies. Further research is needed to inform policy on innovative strategies for disseminating local information and resources to older migrants, such as leveraging social media platforms as well as developing ICTs that meet the needs of older adults more generally.

Acknowledgments

The authors would like to express their sincere gratitude and appreciation to the Edmonton community organizations, Al-Rashid Mosque and the Islamic Family and Social Services Association, who contributed generously to the activities of this project. The authors are indebted to the participants of this study, who shared their time and knowledge. They also want to thank Hesham Shammy, who was the instructor for the sessions. This work was supported by a Social Sciences and Humanities Research Council of Canada Insight Development Grant (RES0049609).

Data Availability

The datasets generated or analyzed during this study are not publicly available, due the nature of qualitative data where migrant research participants might be easily identifiable, but are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

JS and MK conceptualized this study and wrote the grant proposal. AA and GS led data collection in the field. All team members contributed to data analysis. JS led the writing of the paper, with feedback and final approval of the manuscript provided by all members of the team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topical codes—digital competence of participants.
[DOCX File , 21 KB - jmir_v27i1e60547_app1.docx]

References

1. Collin S, Karsenti T, Calonne O. Migrants' use of technologies: an overview of research objects in the field. J Technol Human Usability 2015;10(3-4):15-29. [doi: 10.18848/2381-9227/cgp/v10i3-4/56424]
2. Davidson J, Schimmele C. Evolving internet use among Canadian seniors. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2019015-eng.htm [accessed 2024-04-29]
3. Olsson T, Samuelsson U, Viscovi D. At risk of exclusion? Degrees of ICT access and literacy among senior citizens. Inf Commun Soc 2017 Jul 20;22(1):55-72. [doi: 10.1080/1369118x.2017.1355007]
4. Chen X, Frennert S, Östlund B. The use of information and communication technology among older immigrants in need of home care: a systematic literature review. Ageing Int 2021 Jul 13;47(2):238-264. [doi: 10.1007/S12126-021-09417-X]
5. Quan-Haase A, Martin K, Schreurs K. Interviews with digital seniors: ICT use in the context of everyday life. Inf Commun Soc 2016 Feb 17;19(5):691-707. [doi: 10.1080/1369118x.2016.1140217]
6. Peek ST, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016 Jun 5;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
7. Sen K, Prybutok G, Prybutok V. The use of digital technology for social wellbeing reduces social isolation in older adults: a systematic review. SSM Popul Health 2022 Mar;17:101020 [FREE Full text] [doi: 10.1016/j.ssmph.2021.101020] [Medline: 35024424]
8. Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021 Oct 12;11(10):e053440 [FREE Full text] [doi: 10.1136/bmjopen-2021-053440] [Medline: 34642200]
9. Gleason K, Suen JJ. Going beyond affordability for digital equity: closing the "Digital Divide" through outreach and training programs for older adults. J Am Geriatr Soc 2022 Jan 22;70(1):75-77 [FREE Full text] [doi: 10.1111/jgs.17511] [Medline: 34648663]
10. Köttl H, Allen LD, Mannheim I, Ayalon L. Associations between everyday ICT usage and (self-)ageism: a systematic literature review. Gerontologist 2023 Aug 24;63(7):1172-1187. [doi: 10.1093/geront/gnac075] [Medline: 35670751]
11. Tomczyk Ł, Mróz A, Potyrała K, Wnęk-Gozdek J. Digital inclusion from the perspective of teachers of older adults - expectations, experiences, challenges and supporting measures. Gerontol Geriatr Educ 2022 Sep 24;43(1):132-147. [doi: 10.1080/02701960.2020.1824913] [Medline: 32967582]
12. Kebede AS, Ozolins LL, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 07;24(12):e40192 [FREE Full text] [doi: 10.2196/40192] [Medline: 36477006]
13. Key competences for lifelong learning. European Commission, Directorate-General for Education, Youth, Sport and Culture. URL: https://data.europa.eu/doi/10.2766/569540 [accessed 2024-04-29]
14. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021 Feb 03;23(2):e26145 [FREE Full text] [doi: 10.2196/26145] [Medline: 33533727]
15. Mbandje DC, Loureiro MJ, Lucas M. Digital competence and information literacy: clarifying concepts based on a literature review. Educ Media Int 2024 Mar 05;60(3-4):306-316. [doi: 10.1080/09523987.2023.2324584]
16. Spante M, Hashemi SS, Lundin M, Algers A. Digital competence and digital literacy in higher education research: systematic review of concept use. Cogent Educ 2018 Oct 23;5(1):1519143. [doi: 10.1080/2331186x.2018.1519143]
17. Skov A. What is digital competence. Center for Digital Dannelse. URL: https://digital-competence.eu/front/what-is-digital-competence/ [accessed 2024-04-29]
18. Szabo A, Allen J, Stephens C, Alpass F. Longitudinal analysis of the relationship between purposes of internet use and well-being among older adults. Gerontologist 2019 Jan 09;59(1):58-68. [doi: 10.1093/geront/gny036] [Medline: 29688332]
19. Tyler M, Simic V, De George-Walker L. Older adult Internet super-users: counsel from experience. Act Adapt Aging 2018 Jan 26;42(4):328-339. [doi: 10.1080/01924788.2018.1428472]
20. Schirmer M, Dalko K, Stoevesandt D, Paulicke D, Jahn P. Educational concepts of digital competence development for older adults-a scoping review. Int J Environ Res Public Health 2023 Jun 30;20(13):6269 [FREE Full text] [doi: 10.3390/ijerph20136269] [Medline: 37444116]

XSL•FO
RenderX

21. Wynia Baluk K, Detlor B, La Rose T, Alfaro-Laganse C. Exploring the digital literacy needs and training preferences of older adults living in affordable housing. J Technol Hum Serv 2023 Jul 27;41(3):203-229. [doi: 10.1080/15228835.2023.2239310]
22. Pihlainen K, Ehlers A, Rohner R, Cerna K, Kärnä E, Hess M, et al. Older adults' reasons to participate in digital skills learning: an interdisciplinary, multiple case study from Austria, Finland, and Germany. Stud Ed Adults 2022 Oct 20;55(1):101-119. [doi: 10.1080/02660830.2022.2133268]
23. Arthanat S, Vroman KG, Lysack C, Grizzetti J. Multi-stakeholder perspectives on information communication technology training for older adults: implications for teaching and learning. Disabil Rehabil Assist Technol 2019 Jul 25;14(5):453-461 [FREE Full text] [doi: 10.1080/17483107.2018.1493752] [Medline: 30044659]
24. Fang ML, Canham SL, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring privilege in the digital divide: implications for theory, policy, and practice. Gerontologist 2019 Jan 09;59(1):e1-15. [doi: 10.1093/geront/gny037] [Medline: 29750241]
25. Garcia KR, Rodrigues L, Pereira L, Busse G, Irbe M, Almada M, et al. Improving the digital skills of older adults in a COVID-19 pandemic environment. Educ Gerontol 2021 Mar 30;47(5):196-206. [doi: 10.1080/03601277.2021.1905216]
26. Weil J, Kamber T, Glazebrook A, Giorgi M, Ziegler K. Digital inclusion of older adults during COVID-19: lessons from a case study of older adults technology services (OATS). J Gerontol Soc Work 2021 Sep 22;64(6):643-655. [doi: 10.1080/01634372.2021.1919274] [Medline: 33882782]
27. Canada's population estimates: record-high population growth in 2022. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/230322/dq230322f-eng.htm [accessed 2024-04-29]
28. Alencar A. Refugee integration and social media: a local and experiential perspective. Inf Commun Soc 2017 Jun 21;21(11):1588-1603. [doi: 10.1080/1369118x.2017.1340500]
29. Diminescu D. The connected migrant: an epistemological manifesto. Soc Sci Inf 2008 Dec 01;47(4):565-579. [doi: 10.1177/0539018408096447]
30. Fortunati L, Pertierra R, Vincent J. Migration, Diaspora and Information Technology in Global Societies. New York, NY: Routledge; 2012.
31. Lin CC, Li NJ, Lee E. Exploring the intersections of immigrant seniors' digital literacies and social connectedness: a Canadian study. Front Educ 2023 Sep 14;8:e251. [doi: 10.3389/feduc.2023.1250240]
32. Baldassar L, Wilding R. Migration, aging, and digital kinning: the role of distant care support networks in experiences of aging well. Gerontologist 2020 Feb 24;60(2):313-321. [doi: 10.1093/geront/gnz156] [Medline: 31812983]
33. Au A, Hilario C, Meherali S, Salma J. Unveiling social belonging: exploring the narratives of immigrant Muslim older women. Health Soc Care Community 2024 Jul 30;2024(1):5598247. [doi: 10.1155/2024/5598247]
34. Salma J, Salami B. "Growing old is not for the weak of heart": social isolation and loneliness in Muslim immigrant older adults in Canada. Health Soc Care Community 2020 Mar 21;28(2):615-623. [doi: 10.1111/hsc.12894] [Medline: 31750986]
35. Arab immigration to Canada hits record high. Canadian Arab Institute. URL: https://static1.squarespace.com/static/5e09162ecf041b5662cf6fc4/t/5e5c6fd9b728573d1bad10b7/1583116249563/Arab+Immigration+to+Canada+Hits+Record+High.pdf [accessed 2024-04-29]
36. Hennebry J, Momani B. Targeted transnationals: the state, the media, and Arab Canadians. UBC Press. URL: https://www.ubcpress.ca/targeted-transnationals [accessed 2024-05-06]
37. Georgeou N, Schismenos S, Wali N, Mackay K, Moraitakis E. A scoping review of aging experiences among culturally and linguistically diverse people in Australia: toward better aging policy and cultural well-being for migrant and refugee adults. Gerontologist 2023 Jan 24;63(1):182-199 [FREE Full text] [doi: 10.1093/geront/gnab191] [Medline: 34969076]
38. Millard A, Baldassar L, Wilding R. The significance of digital citizenship in the well-being of older migrants. Public Health 2018 May;158:144-148. [doi: 10.1016/j.puhe.2018.03.005] [Medline: 29625734]
39. Haight M, Quan-Haase A, Corbett BA. Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. Inf Commun Soc 2014 Mar 07;17(4):503-519. [doi: 10.1080/1369118x.2014.891633]
40. Kim J, Lee HY, Christensen MC, Merighi JR. Technology access and use, and their associations with social engagement among older adults: do women and men differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]
41. Schreurs K, Quan-Haase A, Martin K. Problematizing the digital literacy paradox in the context of older adults' ICT use: aging, media discourse, and self-determination. Can J Commun 2017 May 23;42(2):359-377. [doi: 10.22230/cjc.2017v42n2a3130]
42. Schulz R, Wahl HW, Matthews JT, De Vito Dabbs A, Beach SR, Czaja SJ. Advancing the aging and technology agenda in gerontology. Gerontologist 2015 Oct 27;55(5):724-734 [FREE Full text] [doi: 10.1093/geront/gnu071] [Medline: 25165042]
43. Kania-Lundholm M, Manchester H. Ageing with digital technologies: from theory to agency and practice. Int J Ageing Later Life 2022 Apr 21;15(2):9-21. [doi: 10.3384/ijal.1652-8670.4309]
44. Vuorikari R, Kluzer S, Punie Y. DigComp 2.2: the digital competence framework for citizens - with new examples of knowledge, skills and attitudes. European Commission. 2022. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC128415 [accessed 2024-04-29]

XSL•FO
RenderX

45. Punie Y, Ferrari A, Brečko BN. DIGCOMP : a framework for developing and understanding digital competence in Europe. Joint Research Centre, Institute for Prospective Technological Studies, European Commission. 2013. URL: https://data.europa.eu/doi/10.2788/52966 [accessed 2024-04-29]
46. Seifert A, Schelling HR. Seniors online: attitudes toward the internet and coping with everyday life. J Appl Gerontol 2018 Jan 20;37(1):99-109. [doi: 10.1177/0733464816669805] [Medline: 27655237]
47. Marston HR, Musselwhite CB. Improving older people's lives through digital technology and practices. Gerontol Geriatr Med 2021 Sep 11;7:23337214211036255 [FREE Full text] [doi: 10.1177/23337214211036255] [Medline: 34527764]
48. Connelly K, ur Rehman Laghari K, Mokhtari M, Falk TH. Approaches to understanding the impact of technologies for aging in place: a mini-review. Gerontology 2014 Jan 22;60(3):282-288. [doi: 10.1159/000355644] [Medline: 24457288]
49. Phoenix C. Why qualitative research is needed in gerontology and how we can do it better. J Gerontol B Psychol Sci Soc Sci 2018 Sep 20;73(7):e1-85. [doi: 10.1093/geronb/gby079] [Medline: 29947779]
50. Minkler M, Wallerstein N. Community–Based Participatory Research for Health: From Process to Outcomes. Hoboken, NJ: John Wiley & Sons; 2008.
51. Hachem S, Ali S, Al-Omari S, Abi Chahine M, Fahme SA, Mehio Sibai A. "Older people tend to be invisible": a qualitative study exploring the needs and inclusion of older Syrian refugees in the context of compounding crises in host country, Lebanon. Confl Health 2022 Nov 19;16(1):61 [FREE Full text] [doi: 10.1186/s13031-022-00496-4] [Medline: 36403023]
52. Hanley J, Hassan M, Al Mhamied AA, Guruge S, Hajjar O, Hynie M, et al. Resettlement of older Syrian refugees in Canada: key individual factors of social inclusion. Refuge 2024 Jun 26;40(1):1-23. [doi: 10.25071/1920-7336.41223]
53. Ageing and health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-04-29]
54. Asal H. Identifying as Arab in Canada: A Century of Immigration History. Halifax, NS: Fernwood Publishing; 2020.
55. Au A, Siddiqi H, Sayadi G, Zhao T, Kleib M, Tong H, et al. Digital learning preferences of Arabic-speaking older immigrants in Canada: a qualitative case study. Educ Gerontol 2024 Jun 23;50(11):1006-1028. [doi: 10.1080/03601277.2024.2370114]
56. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
57. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014 Oct 15;9(1):26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
58. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA: Sage Publications; 1985.
59. Peine A, Neven L. The co-constitution of ageing and technology – a model and agenda. Ageing Soc 2020 Jun 03;41(12):2845-2866. [doi: 10.1017/s0144686x20000641]
60. Neves BB, Waycott J, Malta S. Old and afraid of new communication technologies? Reconceptualising and contesting the 'age-based digital divide'. J Sociol 2018 Apr 04;54(2):236-248. [doi: 10.1177/1440783318766119]
61. Khvorostianov N. "Thanks to the internet, we remain a family": ICT domestication by elderly immigrants and their families in Israel. J Fam Commun 2016 Aug 02;16(4):355-368. [doi: 10.1080/15267431.2016.1211131]
62. Nguyen HT, Baldassar L, Wilding R. Lifecourse transitions: how ICTS support older migrants' adaptation to transnational lives. Soc Incl 2022 Sep 26;10(4):181-193. [doi: 10.17645/si.v10i4.5735]
63. Edmonds J, Flahault A. Refugees in Canada during the first wave of the COVID-19 pandemic. Int J Environ Res Public Health 2021 Jan 22;18(3):947 [FREE Full text] [doi: 10.3390/ijerph18030947] [Medline: 33499071]
64. Leung D, Lee C, Wang AH, Guruge S. Immigrants' and refugees' experiences of access to health and social services during the COVID-19 pandemic in Toronto, Canada. J Health Serv Res Policy 2023 Jan 15;28(1):34-41 [FREE Full text] [doi: 10.1177/13558196221109148] [Medline: 35971256]
65. Köttl H, Gallistl V, Rohner R, Ayalon L. "But at the age of 85? Forget it!": internalized ageism, a barrier to technology use. J Aging Stud 2021 Dec;59:100971 [FREE Full text] [doi: 10.1016/j.jaging.2021.100971] [Medline: 34794716]
66. Wilding R, Gamage S, Worrell S, Baldassar L. Practices of 'digital homing' and gendered reproduction among older Sinhalese and Karen migrants in Australia. J Immigr Refug Stud 2022 Apr 12;20(2):220-232. [doi: 10.1080/15562948.2022.2046895]
67. Digital literacy exchange program. Government of Canada. URL: https://ised-isde.canada.ca/site/digital-literacy-exchange-program/en/digital-literacy-exchange-program [accessed 2024-04-29]
68. Digital literacy exchange program: immigrant seniors go digital. SUCCESS. URL: https://successbc.ca/seniors-go-digital/services/immigrant-seniors-go-digital/ [accessed 2024-11-25]
69. Adnan M, Özbek Ç. Digital competences of older women in Turkey: gender and ageing as double danger. Educ Gerontol 2023 May 09;49(12):1082-1099. [doi: 10.1080/03601277.2023.2209452]
70. Suntai Z, Beltran SJ. The intersectional impact of race/ethnicity and sex on access to technology among older adults. Gerontologist 2023 Aug 24;63(7):1162-1171. [doi: 10.1093/geront/gnac178] [Medline: 36477498]
71. Yang J, Du P. Gender, capital endowment and digital exclusion of older people in China. Ageing Soc 2020 May 11;41(11):2502-2526. [doi: 10.1017/S0144686x20000434]
72. Kouvonen A, Kemppainen T, Taipale S, Olakivi A, Wrede S, Kemppainen L. Health and self-perceived barriers to internet use among older migrants: a population-based study. BMC Public Health 2022 Mar 23;22(1):574 [FREE Full text] [doi: 10.1186/s12889-022-12874-x] [Medline: 35321678]

XSL•FO
RenderX

Abbreviations

DigComp: Digital Competence Framework for Citizens
FG: focus group
ICT: information and communications technology

Edited by T de Azevedo Cardoso; submitted 14.05.24; peer-reviewed by B Detlor, X Du; comments to author 27.09.24; revised version received 28.11.24; accepted 07.02.25; published 25.03.25.

Please cite as:
Salma J, Au A, Sayadi G, Kleib M
Digital Competence of Arabic-Speaking Immigrant and Refugee Older Adults Enacting Agency and Navigating Barriers: Qualitative Descriptive Study
J Med Internet Res 2025;27:e60547
URL: https://www.jmir.org/2025/1/e60547
doi:10.2196/60547
PMID:

©Jordana Salma, Alesia Au, Ghada Sayadi, Manal Kleib. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patterns of Internet Use in People Diagnosed With Severe Mental Illness: Qualitative Interview Study

Ruth Wadman[1*], PhD; Lauren Walker[2*], PhD; Olivia Taylor[1], MSc; Paul Heron[1], MRES; Elizabeth Newbronner[3], PhD; Panagiotis Spanakis[4], PhD; Suzanne Crosland[1], PGdip; Emily Jane Peckham[5], PhD

[1]Department of Health Sciences, University of York, York, United Kingdom

[2]School of Medical and Health Sciences, Bangor University, Wrexham, United Kingdom

[3]Department of Sociology, University of York, York, United Kingdom

[4]Department of Psychology, University of Crete, Rethymno, Greece

[5]School of Medical and Health Sciences, Bangor University, Bangor, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Lauren Walker, PhD
School of Medical and Health Sciences
Bangor University
Wrexham Technology Park,
Croesnewydd Road
Wrexham, LL13 7YP
United Kingdom
Phone: 44 7934877022
Email: lrw24ltt@bangor.ac.uk

Abstract

Background: People with severe mental illness (SMI) face profound health inequalities, which may be exacerbated by increased rates of digital exclusion, especially as health services move to online provision. The activities that people carry out online can affect how they feel about the internet and may determine whether a person has a positive or negative experience when using the internet. This, in turn, could affect their mental health. To support people with SMI in using digital technology and the internet safely, it is important to understand the internet and digital technology use of those with SMI and their perceived positive or negative impact on their mental health.

Objective: This study aimed to explore the internet and digital technology use of those with SMI, with particular focus on any association between greater use of the internet and poorer self-reported mental health.

Methods: We carried out a qualitative interview study with 16 people with SMI. The sample was drawn from a wider investigation of the impact of the pandemic and its restrictions on the health and well-being of 367 people with SMI. We purposively sampled from the wider study based on age, gender, frequency of internet use, and self-reported mental health. The data were analyzed by 2 researchers using framework analysis.

Results: Participant experiences fell into 3 broad categories: those who had a positive or neutral internet-based experience, those who had negative or difficult experiences, and low users or those with poor digital literacy. Those who had positive or neutral experiences could be broken down into 2 subcategories: first, those with positive or neutral experiences of the internet who were similar in terms of the activities participated in, feelings reported, and their concerns about the internet, and second, conscious users who were mindful of their interaction with the internet world. Participants with difficult experiences fell into 2 categories: those with worries and fears related to using the internet and those who had difficulty limiting their internet use.

Conclusions: People with SMI, similarly the general population, are expected to conduct more of their activities of daily living online in the postpandemic world. This research shows that most internet users with SMI have positive or neutral experiences. However, our typology reveals subgroups of the population with SMI for whom there is a relationship between internet use and difficult feelings. These subgroups can be identified by asking questions about online activities; time spent online; feelings, difficulties, or issues experienced; and use of gambling, dating, adult content, and conspiracy theory websites. Our findings point to further work in collaboration with people with lived experience to modify and test this typology.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e55072) doi:10.2196/55072

KEYWORDS

severe mental illness; internet use; qualitative; typology; protective strategies; digital divide

Introduction

Background

In the general UK population, changes in internet use and its effects on mental health and well-being occurred during the COVID-19 pandemic. For example, consumption of COVID-19–related news on social media was associated with increased depression and anxiety, but during the pandemic, approximately 50% of UK internet users disclosed going online to support their mental and physical health and to feel less lonely [1,2]. Understanding the experiences of use of the internet in people with mental illness, especially those with severe mental illness (SMI), is important as they may experience unique challenges. People with SMI already experience significant health inequalities compared with people without SMI, including greater levels of physical illness and reduced life expectancy, with an estimated 20-year mortality gap [3,4]. People with SMI are at risk of digital exclusion, which, in the context of increased digital provision of health and other services (eg, video consultations, websites, and apps), could compound these health inequalities.

Research carried out before the COVID-19 pandemic presented a mixed picture of digital engagement and exclusion among people with SMI. Levels of digital exclusion decreased between 2011 (30% digitally excluded) and 2016 (18% digitally excluded) in a longitudinal study of people with psychosis [5]. However, high levels of digital exclusion were observed in people requiring community mental health rehabilitation support (such as residential placements and supported living), where only 14% used the internet [6]. Previous research suggests that most people with SMI have access to a digital device, such as a computer or smartphone. However, low rates of smartphone ownership, particularly in those with psychosis-spectrum disorders, have been reported [7-9].

A qualitative study of digital exclusion in 20 mental health service users (half of whom had psychosis) explored the factors that drive digital exclusion, highlighting a perceived lack of knowledge, being unable to access the necessary technology and services owing to personal circumstances, and the barriers presented by mental health difficulties [10]. Specific facilitators for overcoming digital exclusion included intrinsic motivation and a personalized learning format that reflects the individual's unique needs and preferences [10]. A qualitative study investigating how service users feel about digital health interventions for people experiencing SMI found that participants reported feelings of empowerment and being liberated because of an increase in choice, tools, and control [11]. The aforementioned study also identified barriers to accessing digital health interventions, with the assumption that these would replace face-to-face care, excluding those with poor access or abilities, and would be problematic for data protection [11].

Previously, people with SMI have reported that they would be less likely to use digital technology when unwell in comparison to when they are feeling well [7]. Research focused on internet use in general and internet use related to mental health for those with psychosis (eg, digital mental health tools, internet-based communities, and discussion boards) has found that general internet use is higher for younger people and those with low symptom severity and fewer cognitive difficulties [12]. Interestingly, use of the internet for mental health information was not associated with symptom severity or cognitive difficulties, indicating that there were no differences in interest in accessing online mental health resources between those with high symptom severity and cognitive difficulties and those with low symptom severity and cognitive difficulties [12].

During the COVID-19 pandemic, we explored the use of the internet, digital skills, and mental health in people with SMI. We found that most (61%) people with SMI were limited or nonusers of the internet despite most (approximately 80%) having access to a digital device and the internet at home [13]. We found that using the internet "a lot" (vs "just a bit" or "not at all") was associated with younger age, a diagnosis of bipolar disorder compared to a psychosis-spectrum disorder, and a report of a decline in mental health [14]. This finding that high internet use was associated with poorer self-reported mental health was somewhat counterintuitive as we expected greater digital engagement to contribute to greater support and inclusivity. However, we appreciated that it is not only how engaged someone is with the internet but what people do when they are online that affects this. This could determine whether a person has a positive or negative experience using the internet, which could affect their mental health (eg, being online to access support or being online to access potentially scary and stressful information).

In an era in which the use of the internet is becoming increasingly important to access public services, health care, and leisure, it is important to explore the association between greater use of the internet and poorer self-reported mental health to support people with SMI in using digital technology and the internet safely. To do this, we carried out a qualitative study to explore the internet and digital technology use of those with SMI, focusing on time spent online, types of activities, and their perceived positive or negative impact on mental health.

Objectives

Therefore, the aim of this study was to explore the internet and digital technology use of those with SMI, with a particular focus on any association between greater use of the internet and poorer self-reported mental health.

XSL•FO
RenderX

Methods

Participants

Between June 2021 and October 2021, qualitative interviews were undertaken via telephone to explore internet use and abilities for adults who have SMI.

A purposive sample was drawn from the wider Optimising Wellbeing During Self-Isolation (OWLS) study [13]. The OWLS study was a series of 5 surveys tracking the health and well-being of people with SMI during the COVID-19 pandemic. In the initial OWLS study, we asked people whether they would consent to taking part in a qualitative study. From those who consented, we purposively sampled based on age, gender, frequency of internet use, and self-reported mental health. As this study aimed to explore the drivers of the relationship between high internet use and poorer self-reported mental health, the sample was chosen to include individuals with a high use of the internet. In this survey study, we purposively sampled from a group known to use the internet more. This means that the sample reported in this paper had a higher-than-average level of digital skills.

The study team approached 28 participants who had expressed an interest in taking part in an interview study, and 16 (57%) gave informed consent and completed the interviews. After interviewing 16 participants, no new data emerged, and therefore, data saturation was achieved. The final sample included 56% (9/16) female and 44% (7/16) male participants aged between 24 and 68 years; 56% (9/16) of the participants had a schizophrenia-spectrum disorder diagnosis, 25% (4/16) had a diagnosis of bipolar disorder, and 19% (3/16) were diagnosed with another SMI. In total 19% (3/16) of the participants reported outstanding knowledge with regard to use of the internet, 38% (6/16) reported good knowledge, 25% (4/16) reported fair knowledge, and 19% (3/16) reported poor knowledge. Most participants (15/16, 94%) were White British. Most (13/16, 81%) were not employed at the time of the interview. The participants varied in the extent to which they used digital devices and the internet, from limited use to frequent use.

Interviews

Semistructured interviews were conducted following a topic guide (Multimedia Appendix 1). This was devised from previous findings of the OWLS survey to elaborate on information gathered regarding internet and digital device use in the previous 12 months and the importance of digital skills. The aim of the interviews was to explore how much people were using the internet, what they were using it for, and how using the internet made them feel. This was to further understand the relationship we had previously noted between high internet use and poorer self-reported mental health. Before conducting the interviews, the topic guide was piloted with a member of the OWLS lived-experience advisory group. The interviews were conducted between June 2021 and October 2021 by RW, PH, and LW. The researchers conducting the interviews had either a PhD or Master of Science and were a mix of ages and genders; one of the interviewers has lived experience of SMI. The interviews typically lasted between 30 and 40 minutes.

The topic guide began with introductory questions to gather general information regarding the frequency, duration and main purpose of internet use (eg, “Roughly how many hours a day do you spend doing things online” and “Could you tell me about the different types of activities you do online?”). This was followed by questions to gain an insight into the thoughts and feelings of participants when they used the internet (eg, “Do you think the online activities you do affect your mental health in any way?”). We did not specifically include questions about the use of the internet to receive formal or informal mental health services. This did not preclude participants talking about support via the internet, if they felt it was relevant to their experiences. The interview recordings were transcribed verbatim.

Analysis

An analysis was conducted with the aim of *exploring the relationship between higher internet use and poorer self-reported mental health.* Following a preliminary analysis by OT, the transcripts were analyzed by LW and EP using framework analysis. This method was chosen as it allows for the identification, description, and interpretation of key patterns in datasets where there is some comparison between groups or concepts to answer a research question. Framework analysis was originally developed for policy research in the 1980s by Richie and Spencer [15] but since then has been used across disciplines (eg, health research and psychology), usually with but not limited to datasets in which there is some comparison between groups or concepts [16,17]:

> *Framework analysis and applied qualitative research can be a perfect match, in large part because framework analysis was developed for the explicit purpose of analysing qualitative data in applied policy research.*

The five stages of framework analysis were followed: (1) data familiarization; (2) framework identification—deductive themes were derived from the research question, interview schedule, and literature and confirmed from data familiarization [18], and these themes were used to create a framework; (3) indexing—researchers (LW and EJP) coded the transcripts to the framework; (4) charting—the data were abstracted and tabulated; and (5) mapping and interpretation—this led to the creation of a typology.

Findings from a 2019 study that looked at internet use in people with mental illness in Poland created clusters of user types (leisure-seeking omnivores, gamers, and passive-selective users) [18]. To create these clusters, they used an adapted version of the Classification of Internet Usage Activities (Table 1) by van Deursen and van Dijk [19]. We used the adaptation by Ng et al [18] as a basis for our classification of what people were doing online, with the further addition of gambling, dating, and adult content websites (from our topic guide), and then divided these activities into consuming (passive internet use), including searching for information and news and shopping, or interaction (active internet use), including social media, study (attending education online), and some types of gaming. This division was made as, from immersion in the transcripts, we speculated that

XSL•FO
RenderX

there might be a different experience between those interacting with others online and those with more passive internet use.

Table 1. Original classification of activities that the internet may be used for by van Deursen and van Dijk [19] on which Ng et al [18] based their clusters.

Factor	Items
Personal development	• Finding online courses and training • Following online courses • Find vacancies and applying for jobs • Independent learning
Leisure	• Downloading music and video • Hobbies • Free surfing
Commercial transaction	• Using sites such as eBay • Acquiring product information • Shopping or ordering products
Social interaction	• Using social networking sites • Chatting • Sharing photos and videos
Information	• Using search systems • Searching for information
News	• News services • Newspapers and online magazines
Gaming	• Playing games on the internet

Typology (Mapping and Interpretation)

One of the researchers (LW) created a chart to answer questions about activities participated in online, feelings about the online experience, time spent online each day, and any particular issues or difficulties. A typology was developed from the charting process that categorized participants as having positive or neutral experiences, negative or difficult experiences, and low digital literacy or minimal use. The low digital literacy or minimal use category was classified separately as any feelings or experiences seemed to be in response to apprehension about the unknown or lack of digital skills. As these participants were online for <5 minutes per day, their experiences were not categorized as a response to what they were doing online.

Ethics Approval

Ethics approval was granted by the Health Research Authority North West – Liverpool Central Research Ethics Committee (reference 20/NW/0276), and all participants gave informed consent before taking part in an interview. Only the researcher who conducted the interview had access to the original recordings, which were securely stored and protected. To ensure anonymity, all participants were assigned unique codes, and personal identifiers were removed during data transcription to safeguard the privacy of the participants. Participants were offered a £20 (US $25.46) voucher as a token of thanks for their time.

Results

Overview

Our results originally showed 6 different types of internet users (Table 2). Table 3 shows a simplified version of the typology. Participants with positive and neutral experiences were merged into 1 category as there were few differences between the 2 groups in terms of activities participated in and feelings reported (no particular strong feelings were described), and the 2 groups were similar in terms of their concerns (mainly regarding scam attempts). For the purposes of answering the research question, separating the neutral and positive categories did not provide any additional insights.

XSL•FO
RenderX

Table 2. Charting of the activities that people conducted online in terms of frequency of activity and whether carrying out the activity was a positive, negative, or neutral experience.

	Negative or difficult experiences		Positive or neutral experiences			Low use—low digital literacy
	Worries and fears	Difficulties limiting use	Positive	Neutral	Conscious users	
Time daily	• Low; average of 45 min	• High; average of 9 h, 30 min	• Medium; average of 3 h	• Low but up from before the pandemic; average of 1 h, 10 min	• Medium; average of 4.5 h	• Very low; average of <5 min
Activities	• Participated in leisure; commercial—retail; information; and gambling, dating, adult content, or conspiracy theory websites • Did not participate in social, study, career, news, and gaming	• Participated in leisure; commercial—retail; social; career; information; news; gaming; and gambling, dating, adult content, or conspiracy theory websites • Did not participate in study	• Participated in leisure, commercial—retail, study, and information • Did not participate in: career; gaming; and gambling, dating, adult content, or conspiracy theory websites • Mixed responses: social	• Participated in leisure, commercial—retail, commercial—services, social—video calls and email only, study, and information • Did not participate in career; gaming; and gambling, dating, adult content, or conspiracy theory websites	• Participated in leisure, commercial—retail, commercial—services, social, study, information, and news • Did not participate in gaming; gambling; and dating, adult content, or conspiracy theory websites	• Participated in leisure, commercial—services, and social—video calls and email only • Did not participate in study; career; information; news; gaming; and gambling, dating, adult content, or conspiracy theory websites
Feelings	• Strong negative (fear based)	• Negative (boredom and exhaustion based)	• Positive	• Neutral	• Less about feeling and more about choices; both mentioned spiritual belief	• Neutral mixed with some distrust and apprehension
Difficulties or issues	• Worries and fears about being tracked	• Did not necessarily see their high internet use as a problem	• Hard to use, scam attempts, or reading bad news	• Concerns about online banking and scams	• Choosing and being mindful	• Lack of understanding or distrust

Table 3. Characteristics of the 6 different types of internet users.

	Type of internet-based experience					
	Negative or difficult experiences			Positive or neutral experiences		Low digital literacy or low use
	Worries and fears	Difficulty limiting use	Polarized	Positive or neutral	Conscious users	
Time spent online	Low	High	Moderate	Moderate	Moderate	Very low
Activities done online	Consuming only	Interacting, consuming, and gaming	Interacting and consuming	Interacting and consuming	Interacting and consuming	Interacting and consuming
Do these activities include dating, adult content, or conspiracy sites?	Yes	Yes	Yes	No	No	No
Feelings about the experience with the internet	Strong negative (fear based)	Negative (boredom and exhaustion based)	Mixed	Positive or neutral	Choices over feelings	Neutral feelings about actual use (fear of the unknown)

XSL•FO
RenderX

The following sections use quotes from the transcripts to illustrate the main findings for each of the original 6 groups and to link the typology to the original data.

Negative or Difficult Experiences

Participants who had negative or difficult experiences of the internet could be divided into 2 types: those who had worries or fears and those who had difficulties limiting their internet use.

Worries and Fears

This was a group of participants with worries and fears about being online who described strong negative (fear based) emotions in relation to their internet use; did a limited number of activities online, which involved mainly consuming content and very little interacting; had used gambling, dating, adult content, or conspiracy theory websites (specifically, conspiracy theory websites were mentioned); and spent an average of 45 minutes per day online. The participant quotes from this group illustrate the participant experiences typical of this group in terms of time spent online; activities participated in; feelings, issues, or difficulties; and experiences with gambling, dating, adult content, or conspiracy theory websites (Textbox 1).

Textbox 1. Participant quotes from the worries and fears group.

Time

- "So, I mean I've gone through different periods of using the internet. Sometimes we use if quite intensely over times like now I'm just using it for, like I don't know, I'm looking at computer programming at the minute. So I just usually look for bits of information about computer programming that I need and things like that and [unclear]. About half an hour a day or summat like that I'd say." [Participant 274]
- "My general use no more than, definitely no more than an hour." [Participant 167]

Activities

- "Amazon, I've used Amazon to get a few pairs of tracksuit bottoms and things like that...Yeah, I actually do quite a bit of shopping on-line to be fair. I get quite a lot of books on the internet." [Participant 274]
- "Yeah just food shop and sometimes a little bit of, you know, search words or shops and that on Google search and that but pretty much just shopping and that's it. Not very often like or I might check my bank on-line, you know, things like that or I pay a bill or summat." [Participant 167]
- "YouTube is a big one. I don't know much about Netflix. My brother bought me an Android tablet and put Netflix on it but I can't remember what's happened with that though." [Participant 274]

Feelings

- "I don't know the best way to describe how it makes me feel. It doesn't make me feel good I know that much." [Participant 167]
- "Yeah, well it (the internet) has and it hasn't (had a positive impact on my mental health) because I've tried to kill me self and finished up in hospital." [Participant 274]

Issues or difficulties

- "I recently went on Facebook for probably the second time I've been on Facebook. What I don't like about Facebook is that you set up an account and then you can't delete it, so in a way it's a bit...there's no-one to look after you on Facebook. I mean you can report stuff that you think is not there to be put up on-line, you can report things and stuff like that. There's no-one for you to contact if you want to delete your account. You can't get rid of it and stuff like that. Plus, I think to be fair it's a bit spooky because I don't know if police can pull you for certain things if you've been doing something and you've mentioned it all on Facebook." [Participant 274]

Gambling, dating, adult content, or conspiracy theory websites

- "Yeah, I used to do a bit of [unclear] years ago, you know, not very often but, you know, here and there like but that's not summat I've done in over 4 years." [Participant 167]
- "Well I've never really been a gambler if I'm being honest but other things, you know." [Participant 167]

Difficulties Limiting Internet Use

This was a group of participants who had difficulties limiting their internet use who described negative feelings regarding boredom and exhaustion in relation to their internet use; did many different things online (this was the only group that reported gaming); used gambling, dating, adult content, or conspiracy theory websites (specifically, gambling and dating sites); and spent an average of 9 and a half hours per day online. The participant quotes from this group illustrate the participant experiences typical of this group in terms of time spent online; activities participated in; feelings, issues, or difficulties; and experiences with gambling, dating, adult content, or conspiracy theory websites (Textbox 2).

XSL•FO
RenderX

Textbox 2. Participant quotes from the difficulties limiting internet use group.

Time

- "Most of them [hours] when I'm not asleep. A lot of hours, most of the day I use the internet...I work from home so I'm on a laptop from about 8 till 4 or 9 to 5. I game a lot in the evening, use the night on-line services. TV is on a lot. Always on my phone, so constantly." [Participant 216]

Activities

- "I do tend to look at quite a lot of stuff and buy quite a lot of stuff which I probably don't need. I think the tablets make me do that as well [The medication that I take]." [Participant 215]
- "My screen time on my phone is quite high, so I'll scroll through Twitter throughout the day, that's where I get most of my news. If something big has happened that's where I'll usually see it first. Tik Tok is really bad for a rabbit hole so I just keep scrolling because it's addictive. Facebook chatting with my friends. Instagram. Yeah all of my banking is on my phone. Pretty much everything I do is on my phone, so I use contactless payments, I don't even use cards, like physical cards. Let me just look at the apps I have. Uber, so I use that like all the time. I haven't got that much public transport. I have an electric toothbrush that has an app. My toothbrush is connected to the internet." [Participant 216]

Feelings

- "Well I think it's due to my isolation and the tablets that I take, I'm sort of bored, so I'll try and buy something, electrical goods or something or something else that I don't need, normally anyway. Yeah, that's quite a big part of my internet usage. I'm trying to cut down on that. But yeah, it's quite interesting. It's quite engaging." [Participant 215]
- "Well yeah it means that I'm glued to a screen all day and there's no like face to face meetings to communicate with people, it's really different and it seems to be more exhausting and I Googled why, it's because you can't interpret people's body language or like the only communication you've really got is speech and that's quite exhausting trying to work out things about other people." [Participant 216]

Difficulties or Issues

- "It makes me want to buy stuff more. I would say to sort of fill any other void in yourself, I'll buy something I'll treat myself to something. It's all the time. I've realised that I do it quite a lot." [Participant 215]
- "I find it hard to go outside now because I'm so used to being in and talking to people online." [Participant 216]
- "So like I'll scroll through my phone in bed a lot, it's harder to switch off and because obviously blue light from screens that I'm looking at all the time, it's harder to go to sleep and even...because we don't have an aerial in our house. We just use like YouTube and stuff, so literally everything we do is, it's not even like going through into the TV, it's just the internet all the time. If the internet goes down then my life is ruined." [Participant 216]

Gambling, dating, adult content, or conspiracy theory websites

- "I get addicted to things quite easily. So, if I think something is unhealthy, like mentally, I probably [unclear]. I've gambled a little bit." [Participant 216]
- "Not for a while anyway, not since I was younger, so yeah. Yeah, I don't really use any of that stuff to be honest. I've tried internet dating once or twice but that never works. That's about it really." [Participant 215]

Positive or Neutral Experiences

Overview

Those who had positive or neutral experiences were categorized into 3 groups: those with positive experiences of being online, those describing neutral experiences, and conscious users. The positive and neutral experience groups were merged into 1 category as the 2 groups did not appear to be significantly different from each other.

Those with positive and neutral experiences were a group of participants who described positive or neutral feelings in relation to their internet use; did a mixture of things online involving both interacting and consuming; did not use gambling, dating, adult content, or conspiracy theory websites; and spent 1 to 3 hours per day online. The participant quotes from this group illustrate the participant experiences typical of this group in terms of time spent online; activities participated in; feelings, issues, or difficulties; and experiences with gambling, dating, adult content, or conspiracy theory websites (Textbox 3).

XSL•FO
RenderX

Textbox 3. Participant quotes from the group with positive and neutral experiences.

Positive

Time:

- "I suppose then these days when I get up and look at my phone, I will go through WhatsApp and I will look at emails and, you know, sometimes there's the odd message from somebody that doesn't use WhatsApp. But I'd say generally [unclear 28.20] spend an hour or so messing about on it. To me it's a lot, remember people of my age didn't get used to computers at school and all that, you know, we are the older generation." [Participant 533]
- "Not much at all. Two hours." [Participant 182]

Activities:

- "Like I say I've got my bank account and mobile banking on my phone. I use all that which is handy as hell. I use it all the time. I'm forever doing that. I transfer money." [Participant 348]
- "Googling stuff about me own mental health...I like to read up on that...It was giving us more information about what was actually going on in me head." [Participant 183]
- "It's mainly social media, Twitter, Instagram, Facebook, catching up with family and friends really and messaging." [Participant 182]

Feelings:

- "I wouldn't say it's [internet activities] have any negative effects on us, no." [Participant 183]
- "Yeah I would say it's helped us [having the internet]. Just having that extra thing there. Just like keep your mind ticking over, you know, help." [Participant 183]
- "Oh yeah very pleasurable especially with the jokes and knowing that some people enjoyed some jokes and it's a laugh and then they send you a little smiley face back. I think that's all very positive especially in the pandemic. It has for me become the normal way to contact people." [Participant 533]
- "I suppose I wanted to feel part of the world if you know what I mean, part of like by going on twitter and Facebook, it felt like I was participating." [Participant 182]
- "Music tends to chill me out and calm me down, so I just listen to that over again. I mean YouTube, all your songs are there for you and it's free too." [Participant 183]

Difficulties or issues:

- "I wish they were easier to use sometimes, some of these websites...They don't always have instructions as you're going along, you have to find out what to do yourself or they say press there and it comes up with something else. Some of them are just quite difficult to use." [Participant 182]
- "I've given up watching the news years and years ago because being bipolar you just don't want to listen to the news because it's all bad news. So it's just that really. You get your bad news through the internet now, not necessarily listening to the news on TV or reading newspapers but that's just the way things are, isn't it." [Participant 533]
- "I've had a few cranks that have come over the internet and tried to get money out of me but, you know, I've just got rid of them, so that's happened. And I've had a few people try to extort money out of me over the internet which I didn't fall for. I was just laughing at them. I knew what they were doing, so I just played along with it and then when they realised that I wasn't going to pay out any money they soon got rid of me." [Participant 348]

Gambling, dating, adult content, or conspiracy theory websites:

- "So, I wouldn't say that...the negative content, you know, it is an issue, I just don't bother with it. I don't read it. I read hardly anything. I don't get in contact with stuff that stresses me apart from the adverts." [Participant 533]
- "Well what I did was I got this bloody, it was [name] next door and it was one of these dating apps, you know these dating app things. I just wanted rid of it. But I got it and I thought this is absolutely useless. I don't know why I went on and it took me ages to get rid of it. I didn't like it. I just want to meet people naturally rather than going on-line and doing it." [Participant 348]

Neutral

Time:

- "Not a lot to tell you the truth. Oh, I'd say about 20minutes, if that." [Participant 314]
- "Oh roughly? Probably about an hour." [Participant 332]

Activities:

- "So, we have a club in our village, the supper club which meets once a month and obviously we weren't able to meet, so they did run some Zoom sessions on that. Yeah that was a good social thing." [Participant 317]

XSL•FO
RenderX

- "As I say my husband tends to use it more than I do. I do try learn off him, you know, what he's doing and stuff and that's it really." [Participant 314]
- "Well I've been doing shopping. What else? Looking up holidays and things like that...Oh I've been on Zoom talking to people [family and friends]...I was playing scrabble at one time and doing jigsaws on-line as well." [Participant 332]
- "I don't do the social media stuff. I don't get involved in it at all." [Participant 317]

Feelings:

- "...yes, it [online scrabble and jigsaws] helps me sort of unwind a bit." [Participant 332]
- "I'm still a bit apprehensive about it [online banking] because I'm frightened that we're going to get conned [laugh]." [Participant 314]
- "Yeah I would probably say that, yeah it [use of the internet] probably wasn't a bad thing when we were in lockdown." [Participant 317]

Difficulties or issues:

- "I'm frightened, you know, that something will happen and we'll lose money. But I mean, touch wood, since we've been doing it we seem to be doing it right and we've had no problems with it." [Participant 314]

Gambling, dating, adult content, or conspiracy theory websites:

- "No, I don't gamble. I don't go on the adult sites or anything like that. In fact, I wouldn't know how to get on them I don't think, so yeah...No I don't do anything like that." [Participant 314]

Conscious Users

This was a group of participants who used the internet consciously who talked more about choices over online activity than feelings; used the internet for many things (with the exception of gaming); may have come across gambling, dating, adult content, or conspiracy theory websites and chose not to use them; and spent an average of 4 and a half hours per day online. The participant quotes from this group illustrate the participant experiences typical of this group in terms of time spent online; activities participated in; feelings, issues, or difficulties; and experiences with gambling, dating, adult content, or conspiracy theory websites (Textbox 4).

XSL•FO
RenderX

Textbox 4. Participant quotes from the group of conscious users.

Time

- "So probably about 3 to 4 hours at work and maybe 2 hours at home." [Participant 178]
- "I don't know. It fluctuates. It would be, you know, going on...I don't actually have a laptop. I either go on the phone or we have an iPad." [Participant 352]

Activities

- "So, you know, even during that time I'll be mostly using systems, you know, using Chrome, using different systems for work. Maybe have some music while I'm working." [Participant 352]
- "Yeah, I mean Facebook, Amazon. Yeah, you know, I've just acquired a kindle, so yeah. Facebook, Outlook, WhatsApp, Amazon, I've got Prime which is a blessing! You know maybe reading or listening to something like a Jewish Torah lecture or something like this. Lesser organise Instagram. I've not done this in a while but I was actually using Zoom so in addition to these like, you know, communal Zoom calls that I mentioned like courses that I do, I was using Zoom for some remote one to one Alexander technique lessons." [Participant 352]
- "Yeah, so at work I do a lot of academic research, reviewing papers and stuff and ordering, stuff like that. At home it tends to be kind of messaging, social media based kind of stuff." [Participant 178]
- "I use banking like once a week to make sure everything looks fine and sometimes we order online but only kind of a necessity really." [Participant 178]

Feelings

- "Yeah I'm definitely not negative. Just probably being mindful, you know, for example, Amazon not to spend too much. To get unnecessary items." [Participant 352]
- "I think maybe also realising how much time I was spending on it because I'm now a lot more conscious about how much I am spending on the internet and a lot of the time not being very productive and I've realised I would rather be doing other things. I think it was helpful in different ways at times but there was definitely a point where it was more harm than good because a lot of the time scrolling social media was just like doom and gloom and it was really hard not to focus on it." [Participant 178]

Difficulties or issues

- "...keep in touch with people and that became quite difficult and I had to make some self-preservation decisions. I don't need to be up-to-date with how my friends are right now so maybe stepping away from that." [Participant 178]
- "Okay. I mean yeah, I mean yeah what to say. I mean pornography. I mean I think, you know, our society, you know, pornography...look before you even get to hard core stuff that's on-line, you know, on the various sites but I mean just the objectification pictures and kind of stuff. I mean you don't even have to go on sites on the internet to see that, that's on mainstream tv shows now, you know, or, you know. I mean just not even looking for it, I mean this is what people post really. I mean people are posting all kinds of stuff on Instagram, Facebook. Yeah even before you get to like all the kinds of specific like hard core stuff you can get into on the internet. I mean there's plenty more you can get on, it's just out there. I mean Netflix or whatever, you know." [Participant 352]

Gambling, dating, adult content, or conspiracy theory websites

- "...very easily accessible, objectification images there when not looking for them." [Participant 352]
- "I don't really use any of them. I don't gamble or anything like that because I would lose all my money, so yeah. I don't think I have much to contribute to that one." [Participant 178]

Low Digital Literacy or Low Use

This was a group of participants with low digital literacy and minimal use who described neutral feelings about their actual internet use and expressed distrust with regard to potential use. This group described using the internet for leisure, services, and emails or video calls only and mentioned others in the household doing internet-based tasks on their behalf; did not use gambling, dating, adult content, or conspiracy theory websites; and typically spent 0 to 5 minutes a day online. The participant quotes from this group illustrate the participant experiences typical of this group in terms of time spent online; activities participated in; feelings, issues, or difficulties; and experiences with gambling, dating, adult content, or conspiracy theory websites (Textbox 5).

XSL•FO
RenderX

Textbox 5. Participant quotes from the group with low digital literacy or low use.

Time

- "Oh, very little. I might get on it once a month." [Participant 143]
- "Not hours but minutes [per day]." [Participant 159]

Activities

- "I only do sort of like go on YouTube and I don't really do anything else." [Participant 159]
- "I do emails. Hmm what else? Finding things out and that's about it." [Participant 143]

Feelings

- "I'm still a bit apprehensive about it because I am frightened that we're going to get conned...I'm frightened, you know, that something will happen and we'll lose money." [Participant 143]
- "I've never bought anything on-line because I basically don't trust with my card details on the internet. I'd rather go to a shop." [Participant 159]
- "I'm very aware of problems with the computer, so I'm always a bit careful what I'm doing and when. I don't particularly trust them. I've always been like that." [Participant 143]

Difficulties or issues

- "Well if you're trying to find something out, you can't get to the right page or you don't seem to have the information to hand and I find it takes me ages to do anything." [Participant 143]
- "...tried Facebook but couldn't work out how to use it and tried to read the instructions for Facebook." [Participant 159]

Gambling, dating, adult content, or conspiracy theory websites

- "I wouldn't know how to do any of that at all or do that myself, no." [Participant 159]

Discussion

Principal Findings

The patterns of internet use and feelings in people with SMI described in the typology are of interest because, as more of an online presence is necessary for many life tasks and situations, including access to health services, it is important to understand the barriers to inclusion that specific populations may face [20]. The typology of internet users with SMI demonstrates a breadth of experiences and challenges when using the internet. When thinking about addressing the digital divide in this population, a generic approach to digital upskilling (eg, a self-navigated online course) is unlikely to be sufficient.

In response to the findings of the OWLS survey and qualitative work that highlighted a lack of digital skills as the driver of the digital divide in the population with SMI, there may be moves to support people with SMI in increasing their digital skills [13,14]. The experiences of specific subsections of this potentially vulnerable population need to be understood and considered to provide tailored support that understands population-specific barriers and use patterns.

The typology developed in this work revealed patterns that deserve further investigation and offers some guidance at this stage for potential subsections of the population with SMI for whom being online could be problematic.

The *worries and fears* group seemed to try to limit their internet use to mainly consumption activities. Their descriptions of their experiences and feelings suggest that interacting online could be problematic for their mental health. The group who struggled to limit their internet use did a wide range of activities online (including consuming and interacting activities), said that they might seriously struggle without the internet, and talked about shopping in response to isolation and boredom.

Interestingly, both groups of participants who described negative or difficult experiences described using or having previously used gambling, dating, adult content, or conspiracy theory websites, whereas those with positive or neutral experiences or low digital literacy did not report internet use of this type. The exception to this was the conscious use group, in which a participant described this kind of content being difficult to avoid but choosing not to engage with it.

Conscious use is of interest in that these participants described a specific type of relationship to their internet use where they were aware of the possible downsides and negative impacts of the web and made active choices to mitigate these effects. Perhaps there is some learning from the experiences of the conscious user group that could be applied to the group with negative or difficult experiences. This could include the option of conscious limitation of internet use to consumption activities (rather than interaction) for people whose experiences most closely align with the *worries and fears* group; the option of mindfulness about time spent online and types of activities for the group who find it difficult to limit their use; and potential co-designed education on ways to handle internet-based content such as gambling, dating, adult content, and conspiracy theory websites. Perhaps any attempts to increase presence online for this population could include a mindful internet use session created in collaboration with people with SMI.

XSL•FO
RenderX

These postulations about the potential application of the insights from this interview study suggest a very strong case for collaborative work with people with lived experience of SMI. This typology could be shared for validation or otherwise from a lived-experience perspective, and such input could be used to modify and refine the typology. Following this, any advice on internet use for subsections of the population with SMI should be created and verified from a lived-experience perspective (co-design) and eventually tested in a larger, probably quantitative study.

From the standpoint of the results of this work so far, services that support people with SMI should be mindful and understanding of their concerns and barriers faced in relation to using the internet and doing online activities. People with SMI should be offered a one-to-one conversation before embarking on digital upskilling to identify any concerns that may require tailored support.

Any digital upskilling endeavors with people with SMI should devote time to informing them how to safely use the internet and also reassuring them that it is possible to keep themselves safe online.

Limitations and Strengths

Although this research was carried out during the COVID-19 pandemic, we believe that many of the difficulties described by the participants, such as fears about internet scams, are applicable to the postpandemic world [21,22]. The *worries and fears* participants described little or no change in their internet use as compared to before the pandemic in terms of time spent online. Changes to services and activities demand a greater web presence, particularly access to health services [20]. Corresponding author LW has direct lived experience of SMI and brings this perspective to the work. LW can only bring their own perspective to this work, and their background in coproduced research leads them to believe that further collaborative and coproduced work is required in this area.

Comparison With Prior Work

A 2015 systematic review showed that people with psychosis seem to spend more time in chat rooms or playing online games than control groups. The use of other internet-based tools such as Facebook or communication through email was lower or the same as that of controls. Online social networking was used by people with psychosis for establishing new relationships, maintaining relationships or reconnecting with people, and receiving online peer support [23]. Our findings suggest that a subgroup of people with SMI may use the internet in these ways and have trouble limiting their use.

Previous research conducted internationally [24] reveals a similar level of internet use in inpatients with schizophrenia to that of the general population and a difference in terms of higher use of the internet as a source of health information than in the general population. Patterns and levels of use may vary depending on global location and inpatient or community settings.

As far as the authors are aware, there has been little research specifically looking at the link between greater internet use and poorer self-reported mental health in people with SMI. This work builds on a previous body of work rather than acting as a comparator with previous findings.

This typology builds on our series of previous publications [13,14,25] investigating internet use, digital health literacy, and the digital divide in people with SMI. We know from previous quantitative research that people with SMI are more likely to be limited users or nonusers of the internet [13] and, as a population, are lacking digital skills [14]. For some of the participants in this study, the internet was seen as useful for practical tasks such as internet shopping and banking. It was important for staying connected and accessing health information. Participants did not always speak directly about internet use affecting their mental health. Worries, fears, and difficulty limiting internet use, as well as feelings of distrust and apprehension about using the internet, were notable, and a subsection of participants appeared to be having negative experiences in relation to going online.

In a previous study, people with SMI were found to increase their use of the internet over the course of the pandemic, but approximately half rated their knowledge of the internet as fair or worse than fair [14]. In addition, this group was found to have moderate levels of digital health literacy. This underpins the need for digital skill training for people with SMI who would like to access this, with a focus on digital health literacy (eg, the ability to find and understand health-related information online and apply this knowledge to make health care decisions and self-manage their conditions). A partnership approach with a clinician for seeking mental health information has been found to be useful for people with psychosis, and this should also be considered [26].

Conclusions

This was a small sample, and our findings require further investigation. They are not generalizable. However, as far as we know, there is little research looking at patterns of internet use and feelings in this population, and the authors believe that the findings and typology can provide the beginnings of guidance for tailored support for people with SMI who wish to increase their digital skills or presence on the internet.

Our research shows that most internet users with SMI have positive or neutral experiences. However, our typology reveals subgroups of the population with SMI for whom there is a relationship between internet use and difficult feelings. These subgroups can be identified by asking questions about online activities; time spent online; feelings, difficulties, or issues experienced; and use of gambling, dating, adult content, and conspiracy theory websites. This exploration is particularly pertinent in light of findings highlighting the need for training and support among people with SMI to increase digital skills, facilitate digital engagement, and reduce digital exclusion [25]. Our conclusions point to further work in collaboration with people with lived experience to modify and test this typology.

XSL•FO
RenderX

Acknowledgments

The authors extend their gratitude to the participants for sharing their experiences to make this study possible. This report is independent research supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Yorkshire and Humber. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. This study was supported by the Medical Research Council (grant reference MR/V028529) and links with the Closing the Gap cohort, which was partly funded by the Wellcome Trust (reference 204829) through the Centre for Future Health at the University of York, UK Research and Innovation (reference ES/S004459/1), and the NIHR Applied Research Collaboration Yorkshire and Humberside. Generative artificial intelligence was not used in any portion of manuscript writing.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic guide.
[DOCX File , 21 KB - jmir_v27i1e55072_app1.docx]

References

1. Neill RD, Blair C, Best P, McGlinchey E, Armour C. Media consumption and mental health during COVID-19 lockdown: a UK cross-sectional study across England, Wales, Scotland and Northern Ireland. Z Gesundh Wiss 2023;31(3):435-443 [FREE Full text] [doi: 10.1007/s10389-021-01506-0] [Medline: 33777650]
2. UK consumer digital index 2021. Lloyds Bank. URL: https://www.lloydsbank.com/assets/media/pdfs/banking_with_us/whats-happening/210513-lloyds-consumer-digital-index-2021-report.pdf [accessed 2024-04-29]
3. Hayes JF, Marston L, Walters K, King MB, Osborn DP. Mortality gap for people with bipolar disorder and schizophrenia: UK-based cohort study 2000-2014. Br J Psychiatry 2017 Sep;211(3):175-181 [FREE Full text] [doi: 10.1192/bjp.bp.117.202606] [Medline: 28684403]
4. Closing the gap: priorities for essential change in mental health. Social Care, Local Government and Care Partnership Directorate. 2014. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/281250/Closing_the_gap_V2_-_17_Feb_201pdf [accessed 2024-04-29]
5. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016 Nov 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
6. Tobitt S, Percival R. Switched on or switched off? A survey of mobile, computer and Internet use in a community mental health rehabilitation sample. J Ment Health 2019 Feb;28(1):4-10. [doi: 10.1080/09638237.2017.1340623] [Medline: 28675329]
7. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 04;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]
8. Hidalgo-Mazzei D, Nikolova VL, Kitchen S, Young AH. Internet-connected devices ownership, use and interests in bipolar disorder: from desktop to mobile mental health. Digit Psychiatry 2019 May 28;2(1):1-7. [doi: 10.1080/2575517X.2019.1616476]
9. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
10. Greer B, Robotham D, Simblett S, Curtis H, Griffiths H, Wykes T. Digital exclusion among mental health service users: qualitative investigation. J Med Internet Res 2019 Jan 09;21(1):e11696 [FREE Full text] [doi: 10.2196/11696] [Medline: 30626564]
11. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
12. Villagonzalo KA, Arnold C, Farhall J, Rossell SL, Foley F, Thomas N. Predictors of overall and mental health-related internet use in adults with psychosis. Psychiatry Res 2019 Aug;278:12-18. [doi: 10.1016/j.psychres.2019.05.034] [Medline: 31132571]

XSL•FO
RenderX

13. Spanakis P, Heron P, Walker L, Crosland S, Wadman R, Newbronner E, et al. Use of the internet and digital devices among people with severe mental ill health during the COVID-19 pandemic restrictions. Front Psychiatry 2021;12:732735 [FREE Full text] [doi: 10.3389/fpsyt.2021.732735] [Medline: 34630183]
14. Spanakis P, Wadman R, Walker L, Heron P, Mathers A, Baker J, et al. Measuring the digital divide among people with severe mental ill health using the essential digital skills framework. Perspect Public Health 2024 Jan;144(1):21-30 [FREE Full text] [doi: 10.1177/17579139221106399] [Medline: 35929589]
15. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. Analyzing Qualitative Data. New York, NY: Routledge; 1994:173-194.
16. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
17. Goldsmith LJ. Using framework analysis in applied qualitative research. Qual Rep 2021 Jun 20:2061-2076 [FREE Full text] [doi: 10.46743/2160-3715/2021.5011]
18. NG YL, Chan K, Balwicki Ł, Huxley P, Chiu MY. The digital divide, social inclusion and health among persons with mental illness in Poland. Int J Commun 2019;13:1652-1672 [FREE Full text]
19. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media Soc 2013 Jun 07;16(3):507-526 [FREE Full text] [doi: 10.1177/1461444813487959]
20. A plan for digital health and social care: policy paper. NHS England. URL: https://www.gov.uk/government/publications/a-plan-for-digital-health-and-social-care/a-plan-for-digital-health-and-social-care [accessed 2023-10-06]
21. Guedes I, Martins M, Cardoso CS. Exploring the determinants of victimization and fear of online identity theft: an empirical study. Secur J 2022 Jul 21;36(3):472-497. [doi: 10.1057/S41284-022-00350-5]
22. Monteith S, Bauer M, Alda M, Geddes J, Whybrow PC, Glenn T. Increasing cybercrime since the pandemic: concerns for psychiatry. Curr Psychiatry Rep 2021 Mar 03;23(4):18 [FREE Full text] [doi: 10.1007/s11920-021-01228-w] [Medline: 33660091]
23. Highton-Williamson E, Priebe S, Giacco D. Online social networking in people with psychosis: a systematic review. Int J Soc Psychiatry 2015 Feb;61(1):92-101. [doi: 10.1177/0020764014556392] [Medline: 25381145]
24. Žaja N, Vukojević J, Žarko T, Marelić M, Vidović D, Vukušić Rukavina T. Internet use among patients with schizophrenia and depression. Int J Environ Res Public Health 2022 May 07;19(9):5695 [FREE Full text] [doi: 10.3390/ijerph19095695] [Medline: 35565091]
25. Spanakis P, Lorimer B, Newbronner E, Wadman R, Crosland S, Gilbody S, et al. Digital health literacy and digital engagement for people with severe mental ill health across the course of the COVID-19 pandemic in England. BMC Med Inform Decis Mak 2023 Sep 26;23(1):193 [FREE Full text] [doi: 10.1186/s12911-023-02299-w] [Medline: 37752460]
26. Aref-Adib G, O'Hanlon P, Fullarton K, Morant N, Sommerlad A, Johnson S, et al. A qualitative study of online mental health information seeking behaviour by those with psychosis. BMC Psychiatry 2016 Jul 11;16:232 [FREE Full text] [doi: 10.1186/s12888-016-0952-0] [Medline: 27400874]

Abbreviations

OWLS: Optimising Wellbeing During Self-Isolation
SMI: severe mental illness

Edited by A Mavragani; submitted 01.12.23; peer-reviewed by A Rotondi, T Vukušić Rukavina, S Rajput; comments to author 13.02.24; revised version received 08.04.24; accepted 24.01.25; published 28.03.25.

Please cite as:
Wadman R, Walker L, Taylor O, Heron P, Newbronner E, Spanakis P, Crosland S, Peckham EJ
Patterns of Internet Use in People Diagnosed With Severe Mental Illness: Qualitative Interview Study
J Med Internet Res 2025;27:e55072
URL: https://www.jmir.org/2025/1/e55072
doi:10.2196/55072
PMID:

©Ruth Wadman, Lauren Walker, Olivia Taylor, Paul Heron, Elizabeth Newbronner, Panagiotis Spanakis, Suzanne Crosland, Emily Jane Peckham. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Noninvasive Oral Hyperspectral Imaging–Driven Digital Diagnosis of Heart Failure With Preserved Ejection Fraction: Model Development and Validation Study

Xiaomeng Yang[1,2,3*], MD; Zeyan Li[1,2,3*], MD; Lei Lei[4*], PhD; Xiaoyu Shi[1,2,3*], MD; Dingming Zhang[5], PhD; Fei Zhou[6,7], MD, PhD; Wenjing Li[6,7], MD; Tianyou Xu[1,2,3], MD; Xinyu Liu[1,2,3], MD; Songyun Wang[1,2,3,8], MD, PhD; Quan Yuan[9,10], PhD; Jian Yang[6,7], MD, PhD; Xinyu Wang[8,11], PhD; Yanfei Zhong[4,8], PhD; Lilei Yu[1,2,3,8], MD, PhD

[1]Cardiovascular Hospital, Renmin Hospital of Wuhan University, Wuhan, China

[2]Hubei Key Laboratory of Autonomic Nervous System Modulation, Wuhan University, Wuhan, China

[3]Cardiac Autonomic Nervous System Research Center, Wuhan University, Wuhan, China

[4]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

[5]College of Geomatics, Xi'an University of Science and Technology, Xi'an, China

[6]Department of Cardiology, The First College of Clinical Medical Science, Yichang Central People's Hospital, Yichang, China

[7]Hubei Key Laboratory of Ischemic Cardiovascular Disease, China Three Gorges University, Yichang, China

[8]Medical Remote Sensing Information Cross-Institute, Wuhan University, Wuhan, China

[9]College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, Wuhan University, Wuhan, China

[10]lnstitute of Molecular Medicine, Renmin Hospital of Wuhan University, Wuhan, China

[11]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Lilei Yu, MD, PhD
Cardiovascular Hospital
Renmin Hospital of Wuhan University
No. 238 Jiefang Road
Wuhan, 430060
China
Phone: 86 02788041911
Email: lileiyu@whu.edu.cn

Abstract

Background: Oral microenvironmental disorders are associated with an increased risk of heart failure with preserved ejection fraction (HFpEF). Hyperspectral imaging (HSI) technology enables the detection of substances that are visually indistinguishable to the human eye, providing a noninvasive approach with extensive applications in medical diagnostics.

Objective: The objective of this study is to develop and validate a digital, noninvasive oral diagnostic model for patients with HFpEF using HSI combined with various machine learning algorithms.

Methods: Between April 2023 and August 2023, a total of 140 patients were recruited from Renmin Hospital of Wuhan University to serve as the training and internal testing groups for this study. Subsequently, from August 2024 to September 2024, an additional 35 patients were enrolled from Three Gorges University and Yichang Central People's Hospital to constitute the external testing group. After preprocessing to ensure image quality, spectral and textural features were extracted from the images. We extracted 25 spectral bands from each patient image and obtained 8 corresponding texture features to evaluate the performance of 28 machine learning algorithms for their ability to distinguish control participants from participants with HFpEF. The model demonstrating the optimal performance in both internal and external testing groups was selected to construct the HFpEF diagnostic model. Hyperspectral bands significant for identifying participants with HFpEF were identified for further interpretative analysis. The Shapley Additive Explanations (SHAP) model was used to provide analytical insights into feature importance.

Results: Participants were divided into a training group (n=105), internal testing group (n=35), and external testing group (n=35), with consistent baseline characteristics across groups. Among the 28 algorithms tested, the random forest algorithm demonstrated superior performance with an area under the receiver operating characteristic curve (AUC) of 0.884 and an accuracy of 82.9%

XSL•FO
RenderX

in the internal testing group, as well as an AUC of 0.812 and an accuracy of 85.7% in the external testing group. For model interpretation, we used the top 25 features identified by the random forest algorithm. The SHAP analysis revealed discernible distinctions between control participants and participants with HFpEF, thereby validating the diagnostic model's capacity to accurately identify participants with HFpEF.

Conclusions: This noninvasive and efficient model facilitates the identification of individuals with HFpEF, thereby promoting early detection, diagnosis, and treatment. Our research presents a clinically advanced diagnostic framework for HFpEF, validated using independent data sets and demonstrating significant potential to enhance patient care.

Trial Registration: China Clinical Trial Registry ChiCTR2300078855; https://www.chictr.org.cn/showproj.html?proj=207133

(J Med Internet Res 2025;27:e67256) doi:10.2196/67256

KEYWORDS

heart failure with preserved ejection fraction; HFpEF; hyperspectral imaging; HSI; diagnostic model; digital health; Shapley Additive Explanations; SHAP; machine learning; artificial intelligence; AI; cardiovascular disease; predictive modeling; oral health

Introduction

About one-half of patients with chronic heart failure have heart failure with preserved ejection fraction (HFpEF), which has received wide attention in recent years and poses a serious threat to the management of patient health [1,2]. The pathogenesis of HFpEF is complex, and the pathological mechanism is still unclear, so a swift, noninvasive diagnostic strategy is still lacking [3]. Early diagnosis of HFpEF is often difficult because the left ventricular ejection fraction of patients with HFpEF is within the normal range and the symptoms of heart failure are often nonspecific. The diagnosis of HFpEF depends on clinical symptoms, laboratory tests, echocardiography, and an invasive hemodynamics test [4]. However, descriptions of clinical symptoms are subjective, and different clinical symptoms cannot be measured using a unified standard. Echocardiography and hemodynamics tests need to be performed by experienced clinicians in hospitals, thus limiting early identification and timely diagnosis of HFpEF, which introduces obstacles to the health management of individuals with HFpEF [5].

The health of the oral environment has an important impact on cardiovascular health, and oral microenvironmental disorders have been associated with an increased risk of HFpEF [6]. Poor oral health may cause an inflammatory response that is strongly associated with heart failure [7,8]. The tongue, as an important part of the oral cavity, plays an important role in oral health, and the dorsum of the tongue carries the largest number of microbial species, which is an important part of oral health [9]. Inflammation and oxidative stress in the body caused by alterations in the oral microbiome are associated with the risk of developing heart failure [10]. Observing the oral microenvironment of individuals with HFpEF can reflect their lifestyle, dietary habits, and intestinal health [11]. An automated device or system that allows doctors to perform a quick tongue-based diagnosis would be helpful in clinical practice.

Light, which is a single color before dispersion, becomes a pattern of colors arranged in order of wavelength after dispersion; this is called the optical spectrum or optical band [12]. Hyperspectral imaging (HSI) consists of narrower bands, is a technique that captures and analyzes the details of each band in a region, can therefore detect substances that are visually indistinguishable to humans, and has a wide range of applications in many fields [13]. In recent years, many studies have explored the application of HSI technology in medicine [14]. The emergence of handheld HSI cameras, which allow users to quickly capture spectral images, has boosted the application of spectral imaging in medicine [15]. HSI technology has been used in patient information acquisition, medical image analysis, and disease diagnosis [16,17]. HSI is expected to promote the management of patient health. Notably, numerous innovative heart failure prediction models have been developed using common clinical indicators and advanced machine learning techniques [18-23]. These studies primarily focused on the prognosis of heart failure or enhanced existing diagnostic modalities such as echocardiography [24-26]. There exists a significant need for the development of noninvasive, easily accessible diagnostic tools specifically targeting HFpEF, which is a subtype of heart failure characterized by subtle clinical manifestations and complex pathophysiology.

In this study, we collected HSI information of the oral environment, and multiple algorithms were used to select the most characteristic spectral bands of individuals with HFpEF. The best model was selected for internal and external testing, and we used the Shapley Additive Explanations (SHAP) model to additively interpret the best model. The digital HSI HFpEF diagnostic model constructed in this study can help with early detection and management of individuals with HFpEF.

Methods

Study Populations

We included 196 patients who visited Renmin Hospital Wuhan University from April 2023 to August 2023; they comprised the training group and internal testing groups. We included 53 patients who visited Yichang Central People's Hospital from August 2024 to September 2024 as the external testing group (Figure 1). For all participants, HSI and clinical information were collected, and routine tests were conducted. Participants were selected according to the inclusion and exclusion criteria. HFpEF was diagnosed using international standards [4], including (1) typical signs and symptoms of heart failure; (2) left ventricular ejection fraction ≥50%, as assessed by echocardiography performed by a proficient physician; and (3)

XSL•FO
RenderX

brain natriuretic peptide >35 ng/L or amino-terminal pro-brain natriuretic peptide >125 ng/L. In addition to these criteria, participants also had to have at least left ventricular hypertrophy or left atrial enlargement with abnormal diastolic function. Individuals were excluded for any of the following reasons: previous cosmetic surgery and facial aesthetics, severe hepatic or renal organ insufficiency, mental or legal incapacity, malignancy-related diseases, other diseases such as psychiatric or neurological disorders, and unable to complete the study activities. Participants were consecutively included in the training and internal testing groups (140 of 196 participants) before randomization using a ratio of 3:1, while participants in the external testing group (35 of 53 participants) were also consecutively included [19]. The manuscript was written in strict accordance with the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [20,21].

Figure 1. Study protocol. AI: artificial intelligence; HFpEF: heart failure with preserved ejection fraction; HSI: hyperspectral imaging; SHAP: Shapley Additive Explanations.

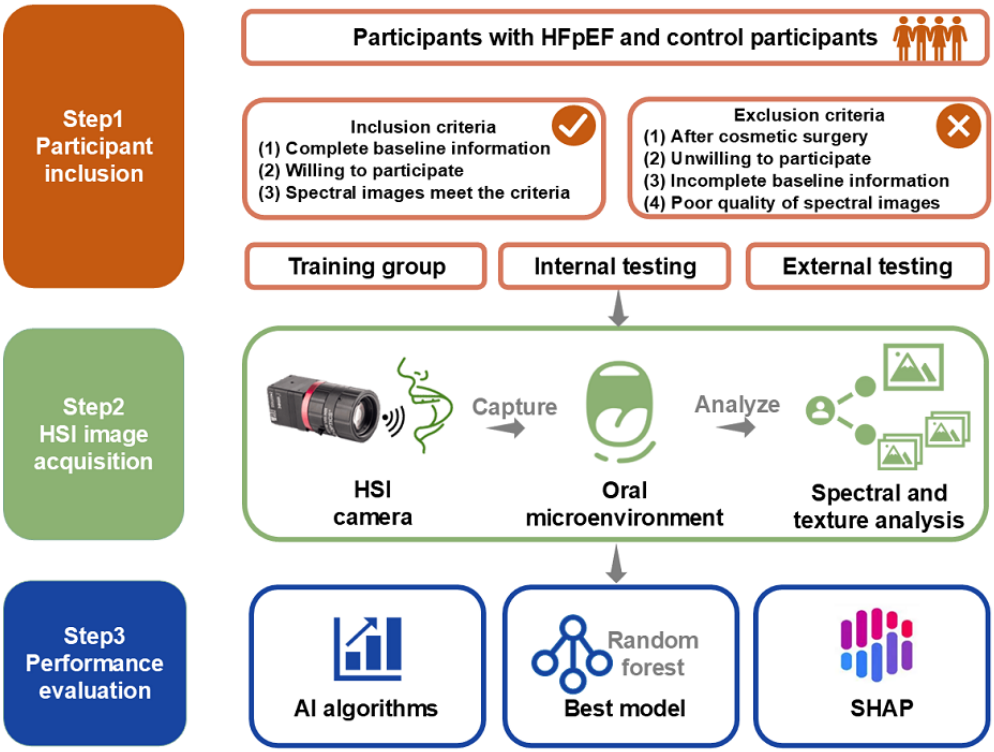

Ethical Considerations

All participants were informed about the study and signed an informed consent document. The study protocol was reviewed and approved by the Ethics Committee of Renmin Hospital of Wuhan University (number WDRM2023-K174) and Yichang Central People's Hospital (number 2024-216-01).

HSI Collection

We prepared a room with good light avoidance conditions and created a dark environment using curtains, blackboards, and other items to avoid external light interference as much as possible to ensure that all participants were in the same light environment for image acquisition. A halogen lamp was chosen as the only light source. The participant's head and face were fixed to ensure that there were no accessories nor hair to obscure facial features. The distance between the HSI camera and the participants was maintained to ensure that the size of the facial features was in the same area of each spectral image and that all participants had the same image size and clarity. We asked participants to expose their tongues for HSI acquisition (Figure 2A). Each participant was provided with a single hyperspectral photomicrograph, and only 1 oral spectral image meeting the criteria was captured.

XSL•FO
RenderX

Figure 2. (A) Hyperspectral imaging (HSI) acquisition environment, (B) processing of a spectral image for a control participant, (C) processing of a spectral image for a participant with heart failure with preserved ejection fraction (HFpEF), (D) processing of a texture image for a control participant, (E) processing of a texture image for a participant with HFpEF, (F) representative HSI and echocardiogram of a control participant, and (G) representative HSI and echocardiogram of a participant with HFpEF.

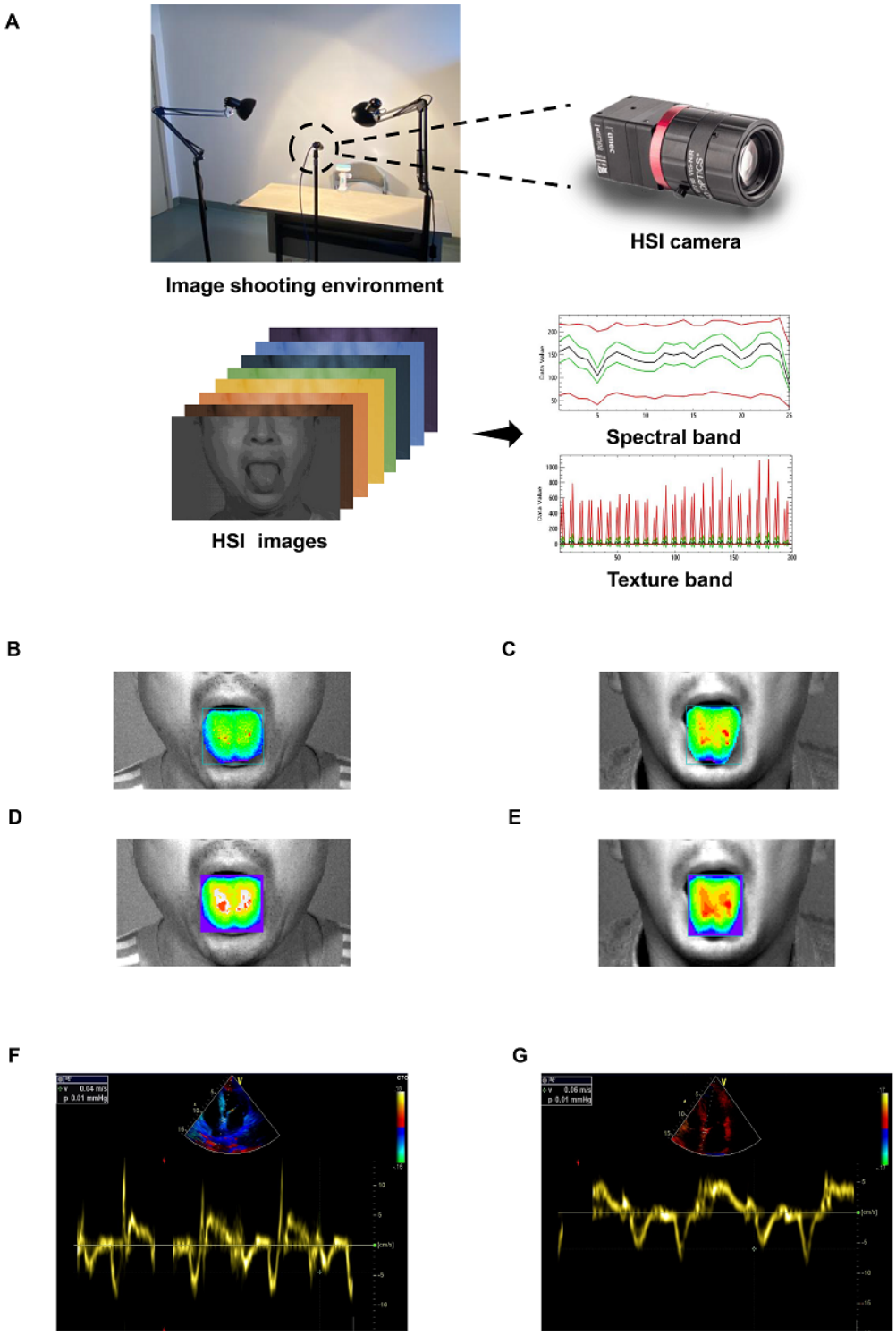

An HSI camera (MQ0220HG-IM-SM4X4-VIS; XIMEA) captured 25 spectral bands in the 665 nm to 960 nm spectral range. The camera has a pixel resolution of 644 × 484 (black and white) or 640 × 480 (color), pixel size of 7.4 × 7.4 µm, active area size of 4.8 × 3.6 mm, sensor diagonal of 5.9 mm, and dynamic range of 60 dB. The camera has an exposure time of 54 µs to 1 s with a step of 7.56 µs, with a capture time less than 1 s per image (Figure 2A). We checked image quality after image capture to ensure that all images had the same luminance, were in sharp focus, and were of good quality.

HSI Analysis

We preprocessed the obtained images [22,23]. HSI preprocessing methods are of great importance in HSI analysis. Effective preprocessing methods can minimize or even eliminate the influence of extraneous information (eg, sample background,

XSL•FO
RenderX

electrical noise, stray light). We first checked the image quality of all oral images, and when poor image quality was found, the participant's oral images were taken repeatedly in the same environment to ensure that the images could be qualified and used for further analysis. We preprocessed the images using the median filtering method to reduce any noise. HSI was preprocessed using a normalized method to reduce redundant information from the original bands and improve the precision of the HSI [27].

All images were edited using image analysis software (ENVI [Environment for Visualizing Images], version 5.3; NV5 Geospatial Solutions Inc). Backgrounds and clothing were cropped out, and the tongue images were retained. The spectral data of the features were initially extracted from the cropped images to obtain the mean and SD for each spectral photograph. Each spectral image was capable of extracting information for 25 spectral bands. Subsequently, texture analysis was conducted on the image, resulting in the generation of 8 texture features based on the mean and SD values for each band. For each participant, their spectral image could be obtained with 50 original band values and 400 band texture values.

Machine Learning Methods

In order to find the most suitable algorithm for the model, we used as many of the currently known machine learning algorithms as possible, resulting in a total of 28 machine learning algorithms. The single algorithms included linear models, such as linear regression, logistic regression, least absolute shrinkage and selection operator (LASSO) regression, ridge regression, and ElasticNet regression. Other types of single algorithms were also used such as decision tree algorithms, neural network algorithms, and support vector machine (SVM) algorithms. The ensemble models included both a boosting algorithm and bagging algorithm; the boosting algorithms included XGBoost algorithms, and the bagging models included random forest algorithms. For probabilistic models, we used Bayesian algorithms, Bernoulli naïve Bayes algorithms, and Gaussian naïve Bayes algorithms.

We performed 5-fold cross-validation in the training group, which was randomly divided into 5 subsets, each with the same sample size. The model was constructed using 4 subsets, and the model performance was assessed using the remaining data. Next, the model was constructed with another combination of the 4 subsets, and the model performance was assessed using the remaining data. The 5 cycles were performed in sequence, and the results obtained for the model performances were summarized. After constructing the model in the training group, model performance was evaluated in the testing group. We evaluated the importance of features using a random forest variable importance ranking method, where the importance of a feature is typically measured by calculating how often the feature is used in a decision tree: the more often a feature is used in the decision tree or the more it contributes to dividing the data, the higher its importance score. The results of all the decision trees' assessments of feature importance were summarized to obtain an importance score for each feature in the entire random forest model. Each machine learning model was constructed by incorporating these rankings until the performance of the model could not be improved; the incorporation of features was then stopped, enabling the construction of the model [28].

Evaluation of Model Performance

Based on the previously mentioned diagnostic criteria for HFpEF, we categorized all participants using a bivariate category. During the model construction process, we evaluated the ability of each model to be able to distinguish between participants with HFpEF and control participants. We evaluated the performance of each model by comparing the ability of the model to accurately identify participants with HFpEF. We assessed the performance of all algorithms; the accuracy, sensitivity, specificity, F_1-score, positive predictive value (PPV), and negative predictive value (NPV) of each algorithm were calculated separately. The results are presented in a table. Receiver operating characteristic (ROC) curves of the 5 best performing models for the internal testing and external testing groups were drawn, and the area under the ROC curve (AUC) was calculated. Decision curve analysis (DCA) of the 5 best performing models was implemented to assess the clinical usefulness of the models. Calibration curves of the 5 best performing models were plotted to assess the calibration.

SHAP Model Interpretation

We used the SHAP model to explain the best machine learning algorithm [29]. SHAP is a method of interpreting the output of a machine learning model and assigns weights to the optimal indexes using the Shapley values derived from the analysis; we used it to quantify the contribution of different features to the predicted values [30]. The SHAP value allows visual identification of the impact of different features on the model prediction results. In addition, the SHAP value explains the prediction results for each individual in the training group, helping to understand why the model made a particular prediction. It is also possible to perform an aggregated global interpretation of Shapley values for multiple data points in order to provide a total interpretation of the model and to demonstrate the interconnections between different features. SHAP analysis was implemented using Python software, and the results are presented using visualization methods. We first demonstrated the contribution of the bands selected by the best model and ranking by the contribution of different features. We then ranked the contribution of each individual in the study to the model and showed the SHAP analysis results for 1 participant with HFpEF and 1 control participant using intuitive visualization methods to reveal the contribution of different features.

Statistical Analysis

The data in this study were analyzed using SPSS 23.0 (IBM Corp). Model construction and graphic drawing were completed using R version 3.6.1 (R Foundation for Statistical Computing). We used the Shapiro-Wilk normality test to check the distribution of the data. A 1-way ANOVA was used to compare continuous variables that had a normal distribution, and the results are shown as mean (SD). The Kruskal-Wallis *H* test was used to compare continuous variables with nonnormal distributions, and the results are shown as median (IQR). The Fisher exact test was used to compare categorical data, and the

XSL•FO
RenderX

results are reported as counts and percentages. $P<.05$ was considered statistically significant.

Results

Study Population

Individuals who visited Renmin Hospital of Wuhan University from April 2023 to July 2023 were enrolled. A total of 196 participants were included in this study after excluding individuals who did not meet the inclusion criteria, had incomplete baseline data, did not cooperate with spectral acquisition, or had poor image quality. Data were collected for a total of 140 participants in the training and testing groups (Figure 1). Individuals who visited Yichang Central People's Hospital from August 2024 to September 2024 were also enrolled. A total of 53 participants were included in this study after excluding individuals who did not meet the inclusion criteria, had incomplete baseline data, did not cooperate with spectral acquisition, or had poor image quality. Data were collected for a total of 35 patients in the external testing groups (Figure 1). The participants were divided into training (n=105), internal testing (n=35), and external testing (n=35) groups. The baseline information, which included basic information, previous medical history, and basic examination and test results, for the training and testing groups was compared (Table 1). There were no significant differences in the baseline data between the training, internal testing, and external testing groups.

XSL•FO
RenderX

Table 1. Participants' baseline characteristics.

Characteristic	Training group (n=105)	Internal testing group (n=35)	External testing group (n=35)	*P* value
Age (years), mean (SD)	62 (11)	64 (11)	59 (15)	.20
Sex, n (%)				.11
Male	64 (61)	28 (80)	24 (68.6)	
Female	41 (39.1)	7 (20)	11 (31.4)	
Current smoker, n (%)				.32
No	39 (37.1)	18 (51.4)	15 (42.9)	
Yes	66 (62.9)	17 (48.6)	20 (57.1)	
Current drinker, n (%)				.94
No	24 (22.9)	8 (22.9)	9 (25.7)	
Yes	81 (77.1)	27 (77.1)	26 (74.3)	
Hypertension				.35
No	53 (50.5)	20 (57.1)	14 (40)	
Yes	52 (49.5)	15 (42.9)	21 (60)	
Diabetes				.61
No	26 (24.8)	9 (25.7)	6 (17.1)	
Yes	79 (75.2)	26 (74.3)	29 (82.9)	
CK-MB[a] (ng/mL), median (IQR)	1.34 (0.59-2.49)	1.62 (0.75-2.34)	1.75 (1.33-2.49)	.22
Cardiac troponin I (ng/mL), median (IQR)	1.45 (0.06-2.90)	1.92 (0.78-4.08)	2.05 (0.31-2.76)	.18
TG[b] (mmol/L), median (IQR)	1.43 (1.10-1.87)	1.57 (1.23-2.01)	1.68 (1.29-2.20)	.14
TC[c] (mmol/L), mean (SD)	3.68 (1.17)	3.59 (1.17)	3.57 (1.04)	.85
HDL-C[d] (mmol/L), median (IQR)	1.30 (1.06-2.00)	1.20 (0.92-1.66)	1.67 (1.25-2.10)	.08
LDL-C[e] (mmol/L), median (IQR)	2.21 (1.46-2.70)	2.11 (1.53-2.34)	2.58 (1.80-3.27)	.052
TSH[f] (μIU/mL), median (IQR)	1.97 (1.32-3.02)	1.78 (1.09-2.71)	1.77 (1.20-2.46)	.34
FT3[g] (pg/mL), median (IQR)	3.25 (2.96-3.58)	3.17 (2.72-3.58)	3.30 (3.09-3.54)	.62
FT4[h] (ng/dL), median (IQR)	1.19 (1.04-1.34)	1.23 (1.07-1.37)	1.19 (1.08-1.29)	.48
BMI (kg/m^2), median (IQR)	24.6 (22.6-26.7)	25.4 (23.7-27.5)	24.7 (22.1-26.8)	.47
NT-proBNP[i] (pg/mL), median (IQR)	95 (60-375)	153 (74-271)	86 (52-114)	.23
HFpEF[j], n (%)				.69
No	74 (70.5)	24 (68.6)	27 (77.1)	
Yes	31 (29.5)	11 (31.4)	8 (22.9)	
LAD[k] (mm), median (IQR)	37.0 (34.0-41.0)	36.0 (31.5-39.0)	36.0 (33.0-39.0)	.38
RAD[l] (mm), median (IQR)	35.0 (33.0-38.0)	34.0 (31.5-36.5)	35.0 (33.5-38.0)	.24
LVEF[m] (%), median (IQR)	55.0 (51.3-60.0)	58.0 (52.0-60.0)	56.0 (51.0-58.0)	.07
LVDD[n] (mm), median (IQR)	49.0 (46.0-52.0)	48.0 (45.5-50.0)	49.0 (45.0-51.5)	.25
E/e′ ratio, median (IQR)	10.7 (8.3-15.3)	11.6 (9.5-16.3)	10.2 (7.9-14.2)	.26

[a]CK-MB: creatine kinase-MB.

[b]TG: triglyceride.

[c]TC: total cholesterol.

XSL•FO
RenderX

[d]HDL-c: high-density lipoprotein cholesterol.
[e]LDL-c: low density lipoprotein cholesterol.
[f]TSH: thyroid-stimulating hormone.
[g]FT3: free triiodothyronine.
[h]FT4: free thyroxine.
[i]NT-proBNP: amino-terminal pro-brain natriuretic peptide.
[j]HFpEF: heart failure with preserved ejection fraction.
[k]LAD: left atrium diameter.
[l]RAD: right atrium diameter.
[m]LVEF: left ventricular ejection fraction.
[n]LVDD: left ventricular end diastolic diameter.

Image Processing and Machine Learning Algorithm Comparison

HSI data were collected for the training, internal testing, and external testing groups. Poor image quality for 4 participants in the training group, 1 participant in the internal testing group, and 2 participants in the external testing group made analysis difficult. Therefore, we acquired oral HSI in the same environmental conditions for those participants again and used those images for the next analysis after ensuring that the image quality met the standards. After preprocessing the images, spectral values and texture values of the images were extracted. We present the characteristic tongue HSI and echocardiography images of control participants and participants with HFpEF in Figures 2B-2G. We used 28 algorithms to filter the characteristic spectral bands of images from participants with HFpEF. We used 5-fold cross-validation in the training group, and the model performance was evaluated in the testing group. The accuracy, F_1-score, PPV, NPV, sensitivity, specificity, and AUC of the different algorithms were calculated separately, and the results for the 28 algorithms are listed in Table S1 in Multimedia Appendix 1. Among all the models constructed, the Tweedie, SVM, partial least squares, Huber, and random forest algorithms were the top 5 in terms of performance, showing good ability to identify participants with HFpEF. The ROC curves (Figure 3A), calibration curves (Figure 3B), and DCA (Figure 3C) of the top 5 algorithms were plotted, and the AUCs were calculated (Table S1 in Multimedia Appendix 1). The AUCs in the internal testing group were 0.884 (95% CI 0.769-1.000) for random forest, 0.795 (95% CI 0.633-0.958) for Tweedie, 0.814 (95% CI 0.657-0.972) for SVM, 0.803 (95% CI 0.659-0.947) for Huber, and 0.799 (95% CI 0.640-0.959) for partial least squares.

Figure 3. (A) Receiver operating characteristic (ROC) curve, (B) calibration curve, and (C) decision curve analysis (DCA) for the internal testing group and (D) ROC curve, (E) calibration curve, and (F) DCA for the external testing group.

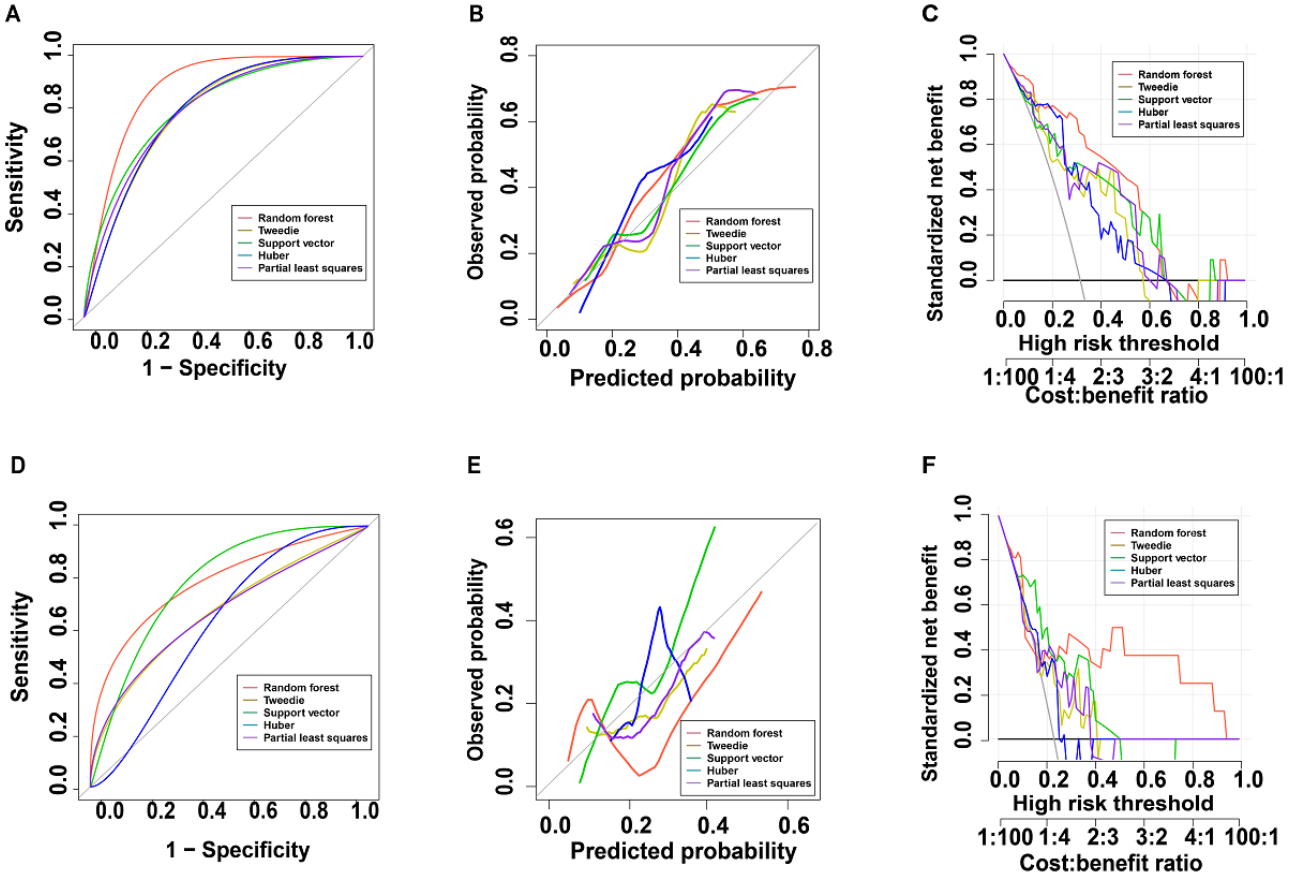

XSL•FO
RenderX

External Validation of the Top 5 Algorithms

The HSI collected in the external testing group were preprocessed in the same way, and the spectral values and texture values were extracted and analyzed. We used the top 5 best performance models in the internal testing group to analyze the external testing data. The accuracy, F_1-score, PPV, NPV, sensitivity, specificity, and AUC of the different algorithms were also calculated separately, and the results for the 5 algorithms are listed in Table S2 in Multimedia Appendix 1. The random forest algorithm still had the best model performance among the 5 algorithms. The ROC curves (Figure 3D), calibration curves (Figure 3E), and DCA (Figure 3F) of the 5 algorithms were plotted, and the AUCs were calculated (Table S2 in Multimedia Appendix 1). The AUCs in the external testing group were 0.812 (95% CI 0.633-0.992) for random forest, 0.676 (95% CI 0.438-0.914) for Tweedie, 0.792 (95% CI 0.632-0.951) for SVM, 0.634 (95% CI 0.428-0.841) for Huber, and 0.671 (95% CI 0.428-0.915) for partial least squares. We extracted the features selected by random forest for further interpretation.

SHAP Interpretation of the Best Algorithms

We used the SHAP model for additive interpretation. The Shapley value was calculated to assign the benefit each characteristic brings to the overall model, showing the contribution of each characteristic to the model's predicted results. We present the characteristics of the top 5 best performing models in order of their contribution to the overall models in Table S3 in Multimedia Appendix 1. The summary plot (Figure 4A) shows the feature importance ranking and distribution via the Shapley value of each spectral band, where the blue bar indicates that the eigenvalue positively affected the model and the red bar indicates that the eigenvalue negatively affected the model. The Shapley value represents the magnitude of each feature's impact on the predicted results, with the point farthest from the centerline indicating a greater influence on the model output. The SHAP feature importance map ranks each characteristic by their contribution (Figure 4B), with features at the top having a greater impact on the model output and those near the bottom having a lesser total impact. SHAP waterfall plots and SHAP bar charts (Figure 4C-4F) were used to visualize the Shapley values of individual samples and their individual results. We visualized the model's ability to recognize HFpEF using a waterfall plot (Figure 4C) and bar chart (Figure 4E) for a control participant and a waterfall plot (Figure 4D) and bar chart (Figure 4F) for a participant with HFpEF. In the waterfall plot, the contribution of each feature is represented by a bar. The length of the bar indicates the magnitude of the feature's influence on the predicted value. A blue bar indicates that the feature increased the predicted value, and a red bar indicates that the feature decreased the predicted value. There was an intuitive difference between the SHAP results for control participants and those for participants with HFpEF. The HFpEF diagnostic model constructed in this study was able to distinguish control participants from participants with HFpEF.

XSL•FO
RenderX

Figure 4. Shapley Additive Explanations (SHAP) of the best algorithm: (A) summary plot, (B) feature importance map, (C) waterfall plot for a control participant, (D) waterfall plot for a participant with heart failure with preserved ejection fraction (HFpEF), (E) bar chart for a control participant, (F) bar chart for a participant with HFpEF.

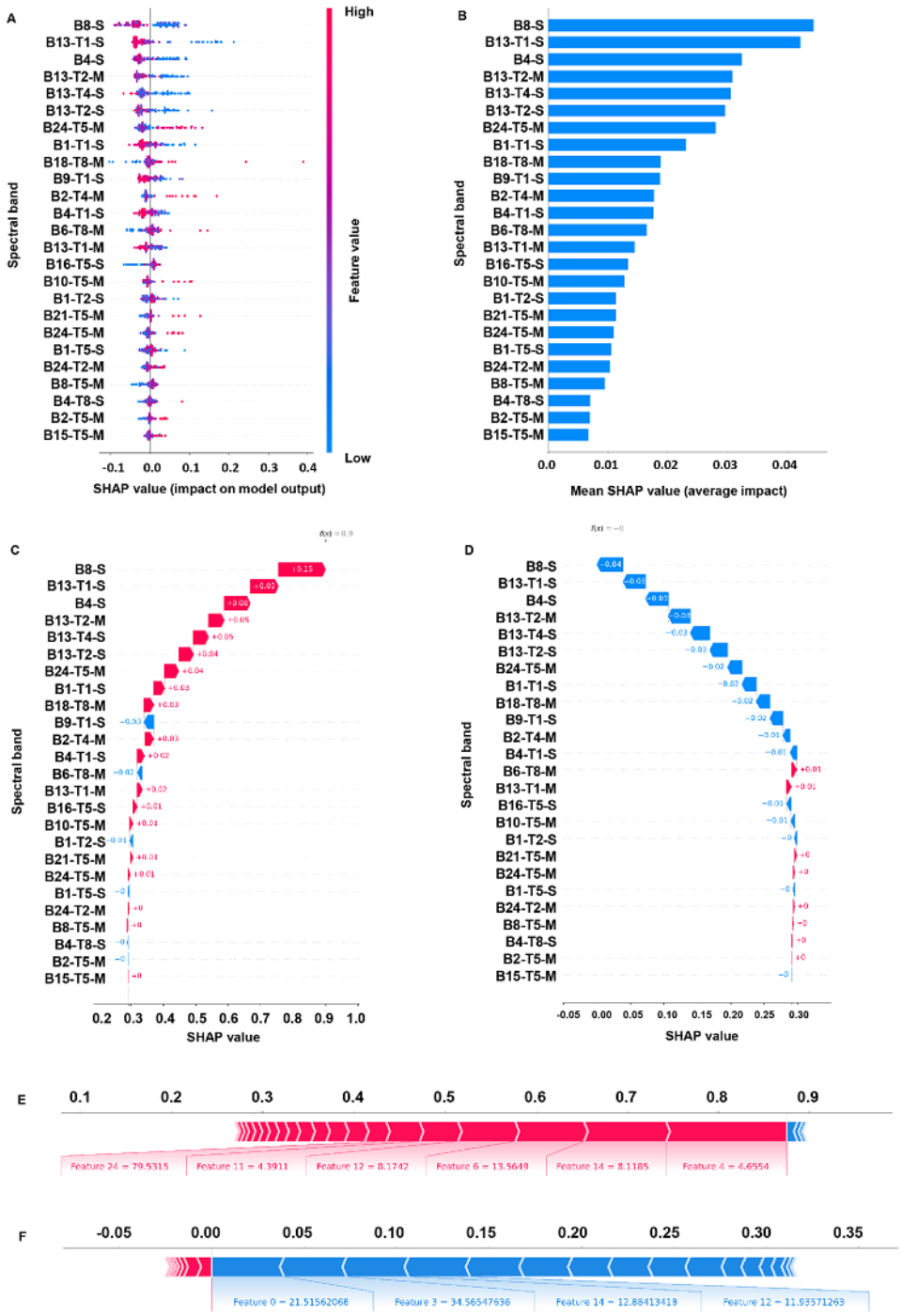

Discussion

Principal Findings

In this study, we innovatively used HSI to acquire oral images of control participants and participants with HFpEF, extracted the spectral and textural information of the HSI, and characterized the HSI using a variety of machine learning algorithms. The optimal algorithm was selected for the construction of the HFpEF diagnostic model, the model performance was validated in the internal and external testing groups, and the SHAP model was used for additive interpretation.

Digital health care is the future of medicine, introducing considerable convenience to the management of people's health [31]. As artificial intelligence and machine learning continue to advance, enabling automated interpretation and classification of HSI will allow researchers to focus more on deep analysis and decision-making [32]. A large amount of data redundancy is present for HSI data, and traditional image data processing methods struggle to meet processing and analyzing needs; therefore, we generated a series of feature extraction and

XSL•FO
RenderX

classification methods for spectral information. The use of multiple artificial intelligence algorithms helps to identify the characteristic bands and textures in large amounts of HSI information [33]. The use of artificial intelligence and machine learning can increase the interpretability of HSI data [34]. To select the most suitable machine learning algorithms for oral spectral images of individuals with HFpEF, we incorporated as many of the current mainstream machine learning models as possible [22]. We analyzed the obtained HSI using 28 machine learning algorithms and filtered the top 5 algorithms for external testing. The random forest algorithm showed good prediction performance in both the internal and external testing groups, so we used the spectral features screened by the random forest algorithm for further analysis. We also performed subgroup analysis based on age and gender in both internal and external testing groups, and the random forest algorithm showed good performance (Figure S1 in Multimedia Appendix 1). Although studies have investigated the use of artificial intelligence for HFpEF diagnosis [19,35,36], these studies primarily focused on constructing echocardiography-based diagnostic models for HFpEF. However, the reliance on specialty physicians to interpret echocardiography exams hinders the early identification of individuals with HFpEF [37]. Moreover, considering the existing diagnostic challenges and limited availability of imaging data during the initial stages of HFpEF, attention should be directed toward noninvasive imaging techniques for diagnosis and management. Therefore, it is imperative to integrate machine learning approaches to develop predictive models for early screening. In our study, we identified spectral bands and textural features that exhibit a close association with the pathophysiological changes observed in HFpEF. This condition is characterized by microvascular dysfunction, inflammation, and oxidative stress leading to alterations in optical properties of oral tissues [3]. HSI effectively captures these subtle changes by detecting variations in light absorption and scattering at different wavelengths. Our model significantly contributes to identifying HFpEF at an early stage, enabling clinicians to promptly initiate appropriate therapies while potentially slowing down disease progression and improving patient health care.

To effectively integrate our diagnostic model into existing health care management for HFpEF, we used an additive interpretation of this digital diagnostic model using the SHAP model. This allowed us to evaluate the contribution of each feature to the prediction, quantify the capability of these features to contribute to the overall model, and demonstrate their impact on the final predictive and diagnostic performance [28]. The SHAP model allows each parameter to be analyzed individually, which can be useful in understanding the decisions made by the model and to improve and rationalize the results of the model [38]. In the macrointerpretation of the SHAP model, the ranking of the contribution of the 25 features screened by random forest to the overall model and the ability of each feature to explain the overall model performance can be clearly seen. In the microinterpretation of the SHAP model, we show the SHAP values for each feature during model evaluation separately for control participants and participants with HFpEF. In the random forest HFpEF diagnostic model, there was a relatively straightforward difference between control participants and participants with HFpEF. The SHAP model demonstrates more intuitively the discriminatory power of the random forest algorithm for participants with HFpEF. Hence, through the provision of intelligent diagnoses, our model effectively bridges the gap between health care professionals and patients, thereby addressing the prevalent issue of frequent misdiagnosis or underdiagnosis and ultimately enhancing the overall quality of care. Furthermore, users can independently access data about oral characteristics at home and transmit the data to health care professionals for expert evaluation through a smartphone app, facilitating continuous monitoring of cardiac health [39]. In addition, in underserved remote areas with limited medical resources, this portable device can be used by mobile medical teams to offer diagnostic screening services to local residents [40], thereby enhancing public health in these regions. Precision medicine and personalized health care represent the future of medical development [41], and our research can also contribute to formulating individualized treatment plans whereby the medical team can gain comprehensive insights into the patient's condition through remotely collected digital data (Figure 5).

XSL•FO
RenderX

Figure 5. Future application scenarios. HFpEF: heart failure with preserved ejection fraction; HSI: hyperspectral imaging.

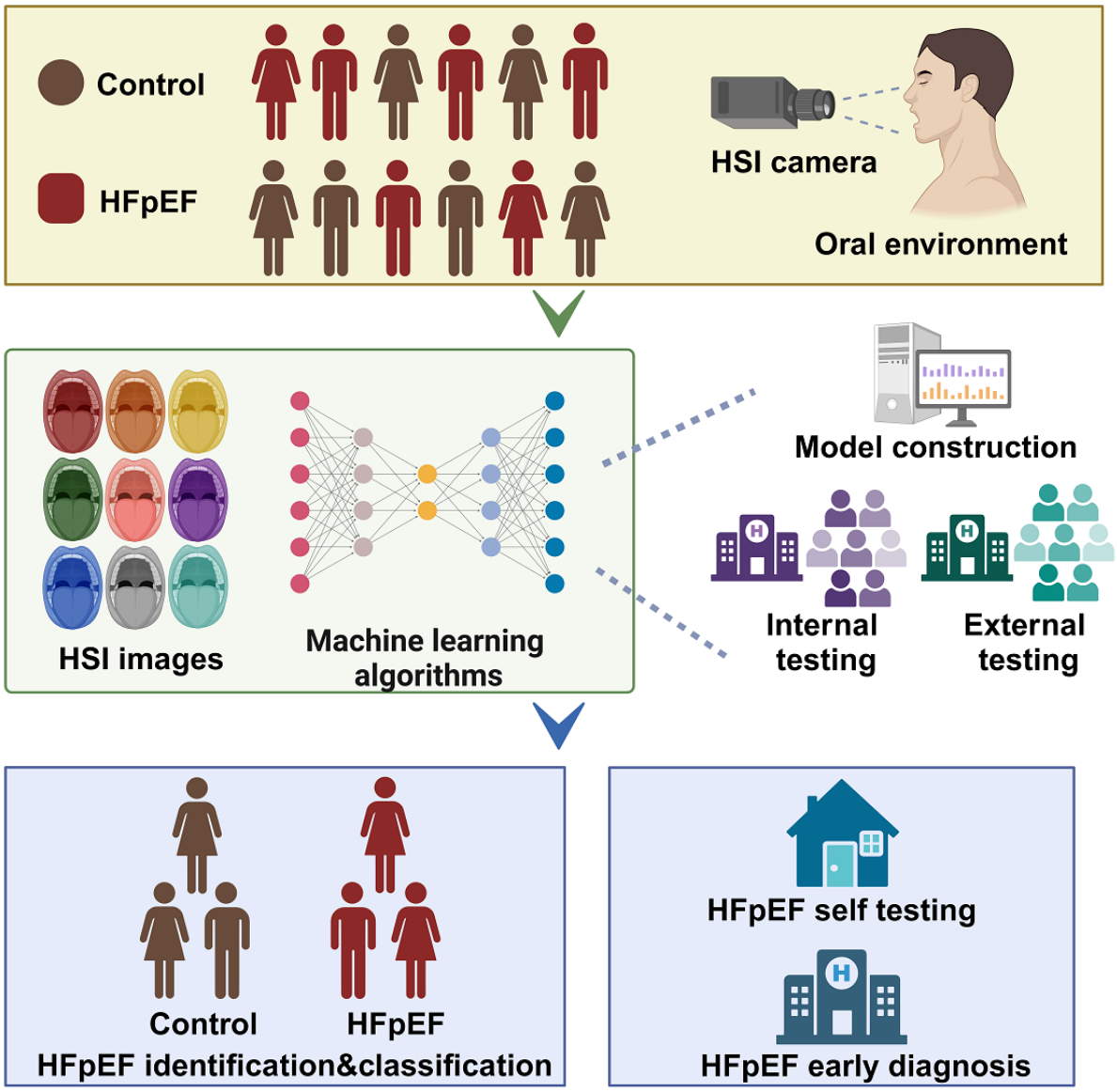

Limitations

First, in this study, model construction and internal testing were conducted in only 1 center, and external testing was conducted in 1 center. Further improvement of the model for the diagnosis of HFpEF should be conducted in multiple centers. Second, this study used multiple algorithms to select the characteristic hyperspectral bands; however, we did not identify images using deep learning algorithms, which could be used for direct image analysis for HFpEF diagnosis. Third, although we used some means to reduce data redundancy and overfitting, these may still exist. Fourth, different demographic information may affect the results. In future model optimization studies, the applicability of the model in different subgroups and including easily accessible clinical information should be considered. The promotion of HSI technology in HFpEF diagnosis has prospective, broader clinical applications that need to be further explored and developed.

Conclusion

This study demonstrates the innovative use of HSI technology to capture oral images and machine learning algorithms to construct a digital model to diagnose HFpEF. This technology was validated to have excellent performance in both internal and external testing groups. This study offers novel insights into the development of portable devices for rapid identification of HFpEF, thereby facilitating the advancement of digital diagnosis and treatment approaches for HFpEF and ultimately leading to improved patient outcomes and reduced health care costs.

XSL•FO
RenderX

Acknowledgments

We thank all the authors for their joint efforts in the completion of this study and many thanks to the School of Remote Sensing and Information Engineering and State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing of Wuhan University for providing the hyperspectral imaging camera.

This work was supported by grants from the National Ten Thousand Talent Program, National Natural Science Foundation of China (82241057, 82270532); Foundation for Innovative Research Groups of Natural Science Foundation of Hubei Province (2021CFA010); Central University Basic Scientific Research Business Expenses Special Funds (2042024kf0041, 2042022kf1211); Taikang Life Medicine Center PI Program of Wuhan University; and Chutian Talent Program Science and Technology Innovation Team Project, Cross-Innovation Talent Program of Renmin Hospital, Wuhan University (JCRCYG-2022-001).

Data Availability

Individual participant data underlining the results reported in this manuscript after deidentification can be obtained from the corresponding author upon request.

Authors' Contributions

LY, YZ, JY, and XW conceptualized and supervised the study and acquired funding. XY, ZL, LL, XS, FZ, WL, TX, XL, and SW curated the data and performed the formal analysis. XY and ZL wrote the original manuscript draft. LY, YZ, JY, QY, and XW reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional performance evaluations of the algorithms.
[DOCX File , 284 KB - jmir_v27i1e67256_app1.docx]

References

1. Dunlay SM, Roger VL, Redfield MM. Epidemiology of heart failure with preserved ejection fraction. Nat Rev Cardiol 2017 Oct 11;14(10):591-602. [doi: 10.1038/nrcardio.2017.65] [Medline: 28492288]
2. Campbell P, Rutten FH, Lee MM, Hawkins NM, Petrie MC. Heart failure with preserved ejection fraction: everything the clinician needs to know. The Lancet 2024 Mar;403(10431):1083-1092. [doi: 10.1016/s0140-6736(23)02756-3] [Medline: 38367642]
3. Borlaug BA. The pathophysiology of heart failure with preserved ejection fraction. Nat Rev Cardiol 2014 Sep 24;11(9):507-515. [doi: 10.1038/nrcardio.2014.83] [Medline: 24958077]
4. Pieske B, Tschöpe C, de Boer RA, Fraser A, Anker S, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J 2019 Oct 21;40(40):3297-3317 [FREE Full text] [doi: 10.1093/eurheartj/ehz641] [Medline: 31504452]
5. Guazzi M, Wilhelm M, Halle M, Van Craenenbroeck E, Kemps H, de Boer RA, et al. Exercise testing in heart failure with preserved ejection fraction: an appraisal through diagnosis, pathophysiology and therapy - A clinical consensus statement of the Heart Failure Association and European Association of Preventive Cardiology of the European Society of Cardiology. Eur J Heart Fail 2022 Aug 31;24(8):1327-1345 [FREE Full text] [doi: 10.1002/ejhf.2601] [Medline: 35775383]
6. Molinsky RL, Yuzefpolskaya M, Norby FL, Yu B, Shah AM, Pankow JS, et al. Periodontal status, C-reactive protein, NT-proBNP, and incident heart failure: the ARIC study. JACC Heart Fail 2022 Oct;10(10):731-741 [FREE Full text] [doi: 10.1016/j.jchf.2022.05.008] [Medline: 36175058]
7. Tonelli A, Lumngwena EN, Ntusi NAB. The oral microbiome in the pathophysiology of cardiovascular disease. Nat Rev Cardiol 2023 Jun 09;20(6):386-403. [doi: 10.1038/s41569-022-00825-3] [Medline: 36624275]
8. Carrizales-Sepúlveda EF, Ordaz-Farías A, Vera-Pineda R, Flores-Ramírez R. Periodontal disease, systemic inflammation and the risk of cardiovascular disease. Heart Lung Circ 2018 Nov;27(11):1327-1334. [doi: 10.1016/j.hlc.2018.05.102] [Medline: 29903685]
9. Wilbert SA, Mark Welch JL, Borisy GG. Spatial ecology of the human tongue dorsum microbiome. Cell Rep 2020 Mar 24;30(12):4003-4015.e3 [FREE Full text] [doi: 10.1016/j.celrep.2020.02.097] [Medline: 32209464]
10. Yuzefpolskaya M, Bohn B, Ladanyi A, Khoruts A, Colombo PC, Demmer RT. Oral and gut microbiome alterations in heart failure: Epidemiology, pathogenesis and response to advanced heart failure therapies. J Heart Lung Transplant 2023 Mar;42(3):291-300. [doi: 10.1016/j.healun.2022.12.009] [Medline: 36586790]

XSL•FO
RenderX

11. Beale AL, O'Donnell JA, Nakai ME, Nanayakkara S, Vizi D, Carter K, et al. The gut microbiome of heart failure with preserved ejection fraction. JAHA 2021 Jul 06;10(13):1. [doi: 10.1161/jaha.120.020654]
12. Goetz AFH, Vane G, Solomon JE, Rock BN. Imaging spectrometry for Earth remote sensing. Science 1985 Jun 07;228(4704):1147-1153. [doi: 10.1126/science.228.4704.1147] [Medline: 17735325]
13. Manifold B, Men S, Hu R, Fu D. A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. Nat Mach Intell 2021 Apr 11;3(4):306-315 [FREE Full text] [doi: 10.1038/s42256-021-00309-y] [Medline: 34676358]
14. Shi W, Koo DES, Kitano M, Chiang HJ, Trinh LA, Turcatel G, et al. Pre-processing visualization of hyperspectral fluorescent data with Spectrally Encoded Enhanced Representations. Nat Commun 2020 Feb 05;11(1):726 [FREE Full text] [doi: 10.1038/s41467-020-14486-8] [Medline: 32024828]
15. Cui R, Yu H, Xu T, Xing X, Cao X, Yan K, et al. Deep learning in medical hyperspectral images: a review. Sensors (Basel) 2022 Dec 13;22(24):9790 [FREE Full text] [doi: 10.3390/s22249790] [Medline: 36560157]
16. Lin S, Li Z, Fu B, Chen S, Li X, Wang Y, et al. Feasibility of using deep learning to detect coronary artery disease based on facial photo. Eur Heart J 2020 Dec 07;41(46):4400-4411. [doi: 10.1093/eurheartj/ehaa640] [Medline: 32818267]
17. Zhao Y, Pilvar A, Tank A, Peterson H, Jiang J, Aster JC, et al. Shortwave-infrared meso-patterned imaging enables label-free mapping of tissue water and lipid content. Nat Commun 2020 Oct 23;11(1):5355 [FREE Full text] [doi: 10.1038/s41467-020-19128-7] [Medline: 33097705]
18. Wang J, Wang Y, Duan S, Xu L, Xu Y, Yin W, et al. Multimodal data‐driven prognostic model for predicting long‐term prognosis in patients with ischemic cardiomyopathy and heart failure with preserved ejection fraction after coronary artery bypass grafting: a multicenter cohort study. JAHA 2024 Nov 27:1. [doi: 10.1161/jaha.124.036970] [Medline: 39604036]
19. Chiou Y, Hung C, Lin S. AI-assisted echocardiographic prescreening of heart failure with preserved ejection fraction on the basis of intrabeat dynamics. JACC Cardiovasc Imaging 2021 Nov;14(11):2091-2104 [FREE Full text] [doi: 10.1016/j.jcmg.2021.05.005] [Medline: 34147456]
20. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. J Med Internet Res 2023 Aug 31;25:e48763 [FREE Full text] [doi: 10.2196/48763] [Medline: 37651179]
21. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
22. Leon R, Fabelo H, Ortega S, Cruz-Guerrero IA, Campos-Delgado DU, Szolna A, et al. Hyperspectral imaging benchmark based on machine learning for intraoperative brain tumour detection. NPJ Precis Oncol 2023 Nov 14;7(1):119 [FREE Full text] [doi: 10.1038/s41698-023-00475-9] [Medline: 37964078]
23. Hadoux X, Hui F, Lim JKH, Masters CL, Pébay A, Chevalier S, et al. Non-invasive in vivo hyperspectral imaging of the retina for potential biomarker use in Alzheimer's disease. Nat Commun 2019 Sep 17;10(1):4227 [FREE Full text] [doi: 10.1038/s41467-019-12242-1] [Medline: 31530809]
24. Guo Y, Yu F, Jiang F, Yin S, Jiang M, Li Y, et al. Development and validation of novel interpretable survival prediction models based on drug exposures for severe heart failure during vulnerable period. J Transl Med 2024 Aug 06;22(1):743 [FREE Full text] [doi: 10.1186/s12967-024-05544-6] [Medline: 39107765]
25. Pan X, Wang C, Yu Y, Reljin N, McManus DD, Darling CE, et al. Deep cross-modal feature learning applied to predict acutely decompensated heart failure using in-home collected electrocardiography and transthoracic bioimpedance. Artif Intell Med 2023 Jun;140:102548 [FREE Full text] [doi: 10.1016/j.artmed.2023.102548] [Medline: 37210152]
26. Cen H, Chen S, Feng S, Chen X, Zhu H, Jiang W, et al. Establishing a novel model to assess exercise capacity in chronic heart failure based on stress echocardiography. ESC Heart Fail 2024 Dec;11(6):4335-4347 [FREE Full text] [doi: 10.1002/ehf2.15038] [Medline: 39219247]
27. Amigo JM, Babamoradi H, Elcoroaristizabal S. Hyperspectral image analysis. A tutorial. Anal Chim Acta 2015 Oct 08;896:34-51. [doi: 10.1016/j.aca.2015.09.030] [Medline: 26481986]
28. Hu J, Xu J, Li M, Jiang Z, Mao J, Feng L, et al. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. EClinicalMedicine 2024 Feb;68:102409 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102409] [Medline: 38273888]
29. Deng H, Eftekhari Z, Carlin C, Veerapong J, Fournier KF, Johnston FM, et al. Development and validation of an explainable machine learning model for major complications after cytoreductive surgery. JAMA Netw Open 2022 May 02;5(5):e2212930 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.12930] [Medline: 35612856]
30. Gao Z, Liu X, Kang Y, Hu P, Zhang X, Yan W, et al. Improving the prognostic evaluation precision of hospital outcomes for heart failure using admission notes and clinical tabular data: multimodal deep learning model. J Med Internet Res 2024 May 02;26:e54363 [FREE Full text] [doi: 10.2196/54363] [Medline: 38696251]
31. Huang T, Xu H, Wang H, Huang H, Xu Y, Li B, et al. Artificial intelligence for medicine: Progress, challenges, and perspectives. The Innovation Medicine 2023;1(2):100030 [FREE Full text] [doi: 10.59717/j.xinn-med.2023.100030]

XSL•FO
RenderX

32. Xue X, Sun H, Yang M, Liu X, Hu H, Deng Y, et al. Advances in the application of artificial intelligence-based spectral data interpretation: a perspective. Anal Chem 2023 Sep 19;95(37):13733-13745. [doi: 10.1021/acs.analchem.3c02540] [Medline: 37688541]
33. Hu W, Huang Y, Wei L, Zhang F, Li H. Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors 2015:1-12 [FREE Full text] [doi: 10.1155/2015/258619]
34. Hong D, He W, Yokoya N, Yao J, Gao L, Zhang L, et al. Interpretable hyperspectral artificial intelligence: when nonconvex modeling meets hyperspectral remote sensing. IEEE Geosci. Remote Sens. Mag 2021 Jun;9(2):52-87. [doi: 10.1109/mgrs.2021.3064051] [Medline: 39604036]
35. Akerman AP, Porumb M, Scott CG, Beqiri A, Chartsias A, Ryu AJ, et al. Automated echocardiographic detection of heart failure with preserved ejection fraction using artificial intelligence. JACC Adv 2023 Aug;2(6):100452 [FREE Full text] [doi: 10.1016/j.jacadv.2023.100452] [Medline: 38939447]
36. Kittleson M, Panjrath G, Amancherla K, Davis L, Deswal A, Dixon D, et al. 2023 ACC Expert Consensus Decision Pathway on Management of Heart Failure With Preserved Ejection Fraction: a report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 2023 May 09;81(18):1835-1878 [FREE Full text] [doi: 10.1016/j.jacc.2023.03.393] [Medline: 37137593]
37. Shah SJ, Borlaug BA, Kitzman DW, McCulloch AD, Blaxall BC, Agarwal R, et al. Research priorities for heart failure with preserved ejection fraction. Circulation 2020 Mar 24;141(12):1001-1026. [doi: 10.1161/circulationaha.119.041886] [Medline: 32202936]
38. Rodríguez-Pérez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. J Med Chem 2020 Aug 27;63(16):8761-8777. [doi: 10.1021/acs.jmedchem.9b01101] [Medline: 31512867]
39. Kokkonen J, Mustonen P, Heikkilä E, Leskelä RL, Pennanen P, Krühn K, et al. Effectiveness of telemonitoring in reducing hospitalization and associated costs for patients with heart failure in Finland: nonrandomized pre-post telemonitoring study. JMIR Mhealth Uhealth 2024 Feb 07;12:e51841 [FREE Full text] [doi: 10.2196/51841] [Medline: 38324366]
40. Craig W, Ohlmann S. The benefits of using active remote patient management for enhanced heart failure outcomes in rural cardiology practice: single-site retrospective cohort study. J Med Internet Res 2024 Nov 26;26:e49710 [FREE Full text] [doi: 10.2196/49710] [Medline: 39589775]
41. Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. Cell 2021 Mar 18;184(6):1415-1419 [FREE Full text] [doi: 10.1016/j.cell.2021.01.015] [Medline: 33740447]

Abbreviations

AUC: area under the curve
DCA: decision curve analysis
ENVI: Environment for Visualizing Images
HFpEF: heart failure with preserved ejection fraction
HSI: hyperspectral imaging
LASSO: least absolute shrinkage and selection operator
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating characteristic
SHAP: Shapley Additive Explanations
SVM: support vector machine

Edited by A Coristine; submitted 07.10.24; peer-reviewed by B Wu, GH He; comments to author 13.11.24; revised version received 04.12.24; accepted 20.12.24; published 07.01.25.

Please cite as:
Yang X, Li Z, Lei L, Shi X, Zhang D, Zhou F, Li W, Xu T, Liu X, Wang S, Yuan Q, Yang J, Wang X, Zhong Y, Yu L
Noninvasive Oral Hyperspectral Imaging–Driven Digital Diagnosis of Heart Failure With Preserved Ejection Fraction: Model Development and Validation Study
J Med Internet Res 2025;27:e67256
URL: https://www.jmir.org/2025/1/e67256
doi:10.2196/67256
PMID:

XSL•FO
RenderX

©Xiaomeng Yang, Zeyan Li, Lei Lei, Xiaoyu Shi, Dingming Zhang, Fei Zhou, Wenjing Li, Tianyou Xu, Xinyu Liu, Songyun Wang, Quan Yuan, Jian Yang, Xinyu Wang, Yanfei Zhong, Lilei Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Enhancing Diagnostic Accuracy of Lung Nodules in Chest Computed Tomography Using Artificial Intelligence: Retrospective Analysis

Weiqi Liu[1,2], DrPH; You Wu[3], PhD; Zhuozhao Zheng[4], PhD; Mark Bittle[1], DrPH; Wei Yu[5], PhD; Hadi Kharrazi[1], MD, PhD

[1]Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[2]Department of Research, Sophmind Technology (Beijing) Co Ltd, Beijing, China

[3]Institute for Hospital Management, School of Medicine, Tsinghua University, Beijing, China

[4]Department of Radiology, Beijing Tsinghua Changgung Hospital, Tsinghua University, Beijing, China

[5]Department of Radiology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

Corresponding Author:
Wei Yu, PhD
Department of Radiology
Beijing Anzhen Hospital
Capital Medical University
2 Anzhen Road, Chaoyang District
Beijing, 100029
China
Phone: 86 10 84005287
Email: nxyw1969@163.com

Abstract

Background: Uncertainty in the diagnosis of lung nodules is a challenge for both patients and physicians. Artificial intelligence (AI) systems are increasingly being integrated into medical imaging to assist diagnostic procedures. However, the accuracy of AI systems in identifying and measuring lung nodules on chest computed tomography (CT) scans remains unclear, which requires further evaluation.

Objective: This study aimed to evaluate the impact of an AI-assisted diagnostic system on the diagnostic efficiency of radiologists. It specifically examined the report modification rates and missed and misdiagnosed rates of junior radiologists with and without AI assistance.

Methods: We obtained effective data from 12,889 patients in 2 tertiary hospitals in Beijing before and after the implementation of the AI system, covering the period from April 2018 to March 2022. Diagnostic reports written by both junior and senior radiologists were included in each case. Using reports by senior radiologists as a reference, we compared the modification rates of reports written by junior radiologists with and without AI assistance. We further evaluated alterations in lung nodule detection capability over 3 years after the integration of the AI system. Evaluation metrics of this study include lung nodule detection rate, accuracy, false negative rate, false positive rate, and positive predictive value. The statistical analyses included descriptive statistics and chi-square, Cochran-Armitage, and Mann-Kendall tests.

Results: The AI system was implemented in Beijing Anzhen Hospital (Hospital A) in January 2019 and Tsinghua Changgung Hospital (Hospital C) in June 2021. The modification rate of diagnostic reports in the detection of lung nodules increased from 4.73% to 7.23% (χ^2_1=12.15; *P*<.001) at Hospital A. In terms of lung nodule detection rates postimplementation, Hospital C increased from 46.19% to 53.45% (χ^2_1=25.48; *P*<.001) and Hospital A increased from 39.29% to 55.22% (χ^2_1=122.55; *P*<.001). At Hospital A, the false negative rate decreased from 8.4% to 5.16% (χ^2_1=9.85; *P*=.002), while the false positive rate increased from 2.36% to 9.77% (χ^2_1=53.48; *P*<.001). The detection accuracy demonstrated a decrease from 93.33% to 92.23% for Hospital A and from 95.27% to 92.77% for Hospital C. Regarding the changes in lung nodule detection capability over a 3-year period following the integration of the AI system, the detection rates for lung nodules exhibited a modest increase from 54.6% to 55.84%, while the overall accuracy demonstrated a slight improvement from 92.79% to 93.92%.

XSL•FO
RenderX

Conclusions: The AI system enhanced lung nodule detection, offering the possibility of earlier disease identification and timely intervention. Nevertheless, the initial reduction in accuracy underscores the need for standardized diagnostic criteria and comprehensive training for radiologists to maximize the effectiveness of AI-enabled diagnostic systems.

(J Med Internet Res 2025;27:e64649) doi:10.2196/64649

KEYWORDS

artificial intelligence; diagnostic accuracy; lung nodule; radiology; AI system

Introduction

In the rapidly evolving field of medical imaging, the integration of artificial intelligence (AI) systems is increasingly disrupting common methods of interpreting radiological images [1]. Traditionally, radiologists have been at the forefront of interpreting complex details within medical images, using their clinical expertise and experience [2]. The growing volume and complexity of medical imaging studies, however, have highlighted the limitations of human cognition [3,4]. In turn, AI systems, undeterred by the data volume, offer the potential solution for efficient, consistent, and accurate image analysis, marking a transformative change in radiological practice [5]. In practice, the China National Medical Products Administration and the US Food and Drug Administration have approved some of the AI image-assisted diagnostic systems to be marketed with the aim of improving efficiency and performance [6]. Therefore, there is an urgent need to evaluate the clinical impact before and after the implementation of AI systems in order to gather sufficient clinical evidence for their broader adoption.

In medical imaging, accurate diagnoses are crucial as such data inform clinical decisions, guide treatment plans, and significantly impact patient outcomes [7]. Radiologists are tasked with detecting abnormalities ranging from early-stage diseases to minor structural changes [8]. Any decrease in diagnostic accuracy can lead to delayed interventions, misdiagnoses, and undesired health outcomes [7]. Thus, given the impact of radiological findings on clinical decisions, improving the diagnostic accuracy of human-generated imaging reports has been a constant endeavor [9].

Several radiological studies have shown the efficacy of AI system applications in enhancing sensitivity (true positive rate) of detecting lung-related diseases [10-12], breast cancer [13,14], thyroid nodules [15], and fractures [16]. These studies, however, have been less consistent in evaluating the effect of AI systems in improving the specificity (true negative rate) of diagnosis. Lung nodule screening, a routine medical screening service, is critical for early lung cancer detection [17]. Various AI-based computer-aided diagnosis systems have been reported to substantially enhance radiologists' performance when used as a second reader [18-20]. However, prioritizing sensitivity improvement over specificity could lead to a high number of false positives, causing unnecessary stress for patients and potentially wasting medical resources due to overuse of interventions [21].

To address the lack of studies measuring the improvement of diagnosis specificity in radiology using AI systems, this study aims to evaluate the impact of an AI-assisted lung nodule diagnostic system on the diagnostic accuracy of junior radiologists examining chest computed tomography (CT) scans. The results of this study could influence the future development of AI-assisted diagnostic systems to advance the accuracy of radiological diagnosis and treatment of lung nodules [22].

Methods

Study Design

This study was carried out at 2 tertiary care facilities in China, Hospital A, and Hospital C, from April 2018 to March 2022. Hospital A, in close collaboration with the Capital Medical University, offers a comprehensive radiology department consisting of 27 radiologists conducting and reviewing over 200,000 CT exams annually. Hospital C, affiliated with Tsinghua University, maintains a radiology department with over 20 radiologists. Both health care institutions provide an extensive spectrum of medical imaging services, comprehensive whole-body CT and magnetic resonance imaging scans, specialized breast imaging, as well as an array of interventional diagnostic and therapeutic procedures.

The AI-assisted diagnostic system used in this research embodies cutting-edge technology crafted to assist radiologists in the interpretation of medical images, with a specific emphasis on lung nodules. These lung nodule AI systems were developed and integrated by Care.ai [23] (Yitu, integrated into Deepwise's framework in 2021) and Dr.Wise [24] (Deepwise) to operate at Hospital A (since January 2019) and Hospital C (since June 2021), respectively. These systems harness advanced machine learning algorithms and deep neural networks to analyze radiological images, with a primary focus on the automated detection, categorization, and characterization of abnormalities, particularly lung nodules [25].

The comprehensive and seamless integration of the AI tool within the clinical platform of either hospital enables a thorough assessment of lung nodules integrated into the radiologist's diagnostic workflow (Figure 1), the AI system processes image data, generates annotations or highlights specific regions of interest, such as lung nodules, nodule location, and quantitative data, including nodule size, density, and other relevant parameters [24] (Figure 2). Junior and senior radiologists can scrutinize these findings and seamlessly integrate them into their diagnostic evaluations, thus enabling a comprehensive assessment of lung nodules (Figure 1).

XSL•FO
RenderX

Figure 1. Radiologist's diagnostic workflow with or without the artificial intelligence (AI) system. Note: AI medical devices are regulated to aid in diagnosis. All diagnoses should be made by radiologists and then checked by AI. Some radiologists may rely on the results given by the AI to double-check and modify the results. PACS: picture archiving and communication system; RIS: radiology information system.

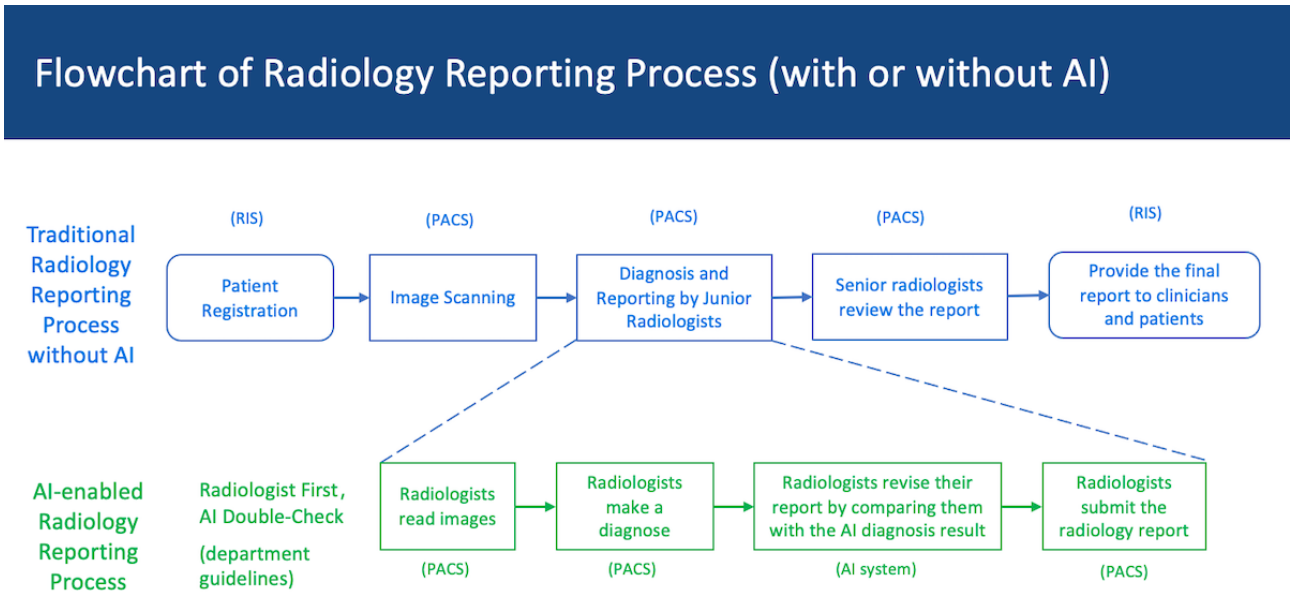

Figure 2. Working window of the artificial intelligence (AI) system.

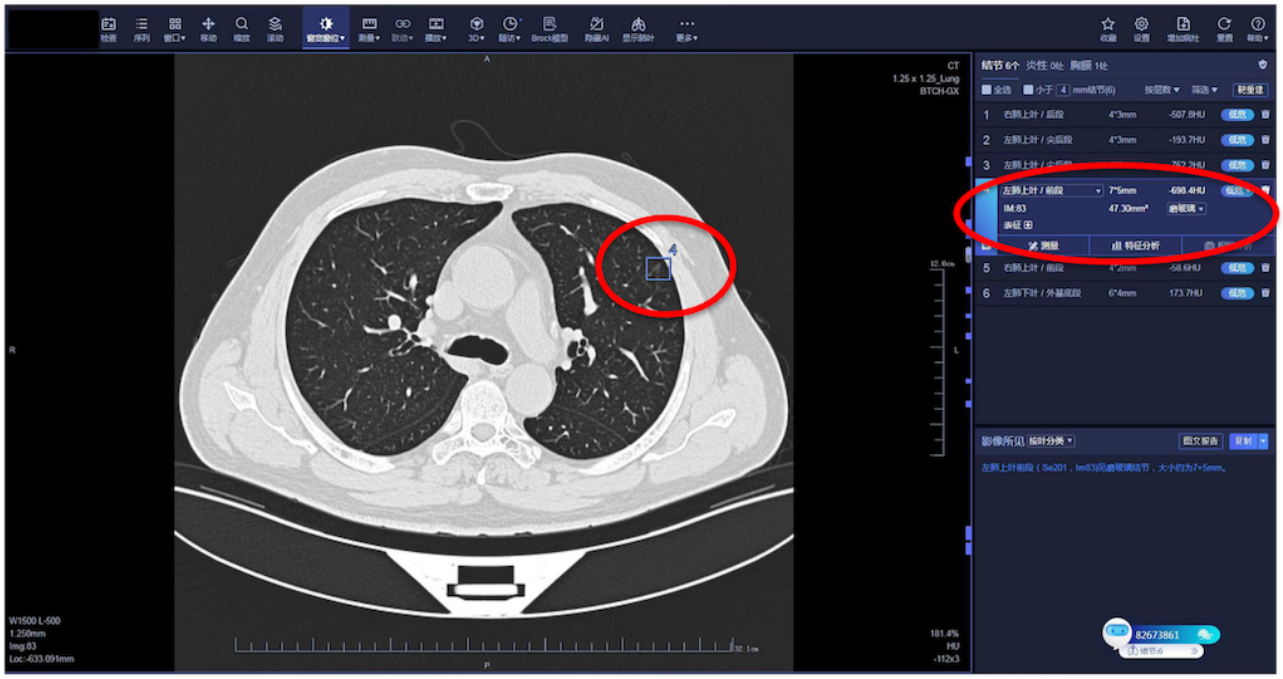

To enhance efficiency and ensure accuracy, hospitals in China typically implement a 2-tiered system for radiology reporting (Figure 1). This process involves a less experienced radiologist, often a junior radiologist, who initially examines the medical images and drafts the report. This preliminary report is then passed on to a more experienced senior radiologist for further evaluation and revision. Once the report has been carefully reviewed and amended as necessary, it is finally forwarded to the physician responsible for diagnosis and also provided to the patient. In some instances, the same senior radiologist is responsible for both composing and reviewing the report. However, given that this study aimed to assess the impact of AI system introduction on the diagnostic accuracy of junior radiologists, those types of reports were excluded from the experimental sample.

We conducted a patient-specific evaluation of the reporting accuracy of junior radiologists for CT imaging. The analysis encompassed data collected from both Hospital A and Hospital C and spanned from April 2018 to March 2022. However, specific intervals, encompassing January 2019 to March 2019, January 2020 to August 2020, and April 2021 to August 2021, were excluded from the analysis due to the substantial impact of the COVID-19 pandemic [26], the implementation of the AI

XSL•FO
RenderX

system, and the Spring Festival vacation on patterns of health-seeking behavior.

Data Collection

The sample size for this study was computed with the assistance of the University of California, San Francisco calculator [27]. The aim was to detect a 30% increase (estimated risk ratio=1.3) in the efficacy of the AI intervention. To achieve 80% statistical power and significance (α=.05, 2-tailed), the study necessitated a sample size of 1594. To adhere to these criteria, the researchers adopted a systematic approach. Each month, they used Microsoft Excel for Mac (version 16.77.1), specifically using the random number sampling function to select approximately 200 patients from the complete pool of individuals who had undergone lung CT examinations at both hospitals during the study period. This methodology ensured that the selection process was unbiased and representative.

A meticulous quality control approach was followed to ensure data validity. The validation entailed the assignment of two researchers to conduct data collection and analysis, with all data undergoing a thorough anonymization process. Data were primarily collected from diagnostic reports written by junior and senior radiologists and captured in the picture archiving and communication system. The picture archiving and communication system allowed for transparent documentation of modifications made by senior radiologists to the reports initially generated by junior radiologists. Since these reports could not be directly exported from the system, photographs were taken, and the researchers summarized diagnostic report modification records. One researcher transcribed the original reports based on the photographic documentation and made adjustments pertaining to the detection of lung nodules and changes in lung lesion morphology, including modifications regarding the number, size, location, and diagnosis of lung nodules. Subsequently, another researcher thoroughly reviewed these entries to ensure accuracy. In cases of disagreement between the two entries, consultation with an expert radiologist who had more than a decade of clinical experience was sought. Due to the 2-tiered system for radiology reporting, the diagnostic results of senior radiologists were used as the reference standard to assess whether the diagnostic accuracy of junior radiologists improved after the AI system was launched. Senior radiologists set the rules for entering data on lung nodules and lung lesion morphology changes. The inclusion criteria for lung nodules included reports describing nodular shadow, punctate hyperdense shadow, and sclerotic foci. For lung lesion morphology changes, reports describing striated hyperdense shadow, patchy shadow, and ground-glass density shadow were included.

Statistical Analysis

The primary outcome measures of this study included lung nodule detection rate, rate of missed diagnoses, rate of misdiagnoses, sensitivity, and positive predictive value. The detection rate was defined as the percentage of all patients in the study who were ultimately reported to have a lung nodule. The rate of missed diagnoses referred to the percentage of lung nodules that were not identified by the junior radiologists but were reported after review by the senior radiologists. The rate of misdiagnoses referred to instances where the junior radiologists incorrectly reported a lung nodule. Sensitivity was defined as a junior radiologist's ability to correctly identify a true positive case, as compared with the diagnosis of a senior radiologist, which was considered the reference standard. The positive predictive value was described as the proportion of correctly identified positive cases in the diagnosis.

This study used descriptive statistics, specifically frequencies and percentages, to concisely present categorical data, particularly diagnostic outcomes. To assess statistical significance, we used the chi-square, Cochran-Armitage, and Mann-Kendall tests. The chi-square test was used to evaluate changes in both the lung nodule's detection modification rate and detection ability as well as the lung's morphological lesions before and after AI launch. The Cochran-Armitage and Mann-Kendall tests were used to compare the directional trend in the detection modification rate and detection ability of lung nodules and lung morphological lesions in relation to the duration of AI's online presence. The chi-square test was used to assess relationships between categorical variables, while the Cochran-Armitage and Mann-Kendall tests addressed trend analysis. Together, these methods effectively met our study's goals of examining intergroup relationships and temporal trends. The raw variables met the distributional assumptions of the applied tests (eg, the chi-square test does not require normality, while the Cochran-Armitage and Mann-Kendall tests are suited to nonparametric data). Therefore, raw values were used for statistical analysis. Significance was established when *P* values fell below the .05 threshold. Data analysis was carried out using R version 4.1.2 (R Development Core Team) [28], and the resulting findings were visualized using tables and graphs for enhanced interpretability.

Ethical Considerations

This study was approved by the Institutional Review Board of the Johns Hopkins Bloomberg School of Public Health (reference number FWA 00000287). The use of data was also approved by the institutional review boards of the Tsinghua Changgung Hospital (reference number 23352-4-01) and the Beijing Anzhen Hospital (reference number 2023078X). The data used in this study were anonymous and did not contain any personally identifiable information. Furthermore, no patients can be identified in any images contained within the manuscript or its supplementary materials.

Results

Patient Characteristics

This study analyzed patient data collected from both hospitals before and after the implementation of the AI system. The study population included a total of 12,889 patients, with 6439 patients having an encounter with Hospital C and 6450 patients from Hospital A. The demographic characteristics of the patients are presented in Table 1.

XSL•FO
RenderX

Table 1. Demographic characteristics of the study population before and after the AI launch.

Variables and categories	Hospital C (n=6439)		Hospital A (n=6450)	
	Before AI[a] (n=4830), n (%)	After AI (n=1609), n (%)	Before AI (n=1606), n (%)	After AI (n=4844), n (%)
Age (years)				
<18	24 (0.5)	16 (0.99)	4 (0.25)	41 (0.85)
18-54	1747 (36.17)	887 (55.13)	499 (31.07)	1939 (40.03)
≥55	3059 (63.33)	706 (43.88)	1103 (68.68)	2864 (59.12)
Gender				
Male	2542 (52.63)	907 (56.37)	864 (53.8)	2626 (54.21)
Female	2288 (47.37)	702 (43.63)	742 (46.2)	2218 (45.79)
Patient source				
Physical examination	514 (10.64)	231 (14.36)	5 (0.31)	188 (3.88)
Outpatient	2187 (45.28)	451 (28.03)	1036 (64.51)	3510 (72.46)
Emergency room	868 (17.97)	699 (43.44)	12 (0.75)	443 (9.15)
Inpatient	1261 (26.11)	228 (14.17)	553 (34.41)	703 (14.51)
Examination time				
2018.4-2018.12	1613 (33.4)	0 (0)	1606 (100)	0 (0)
2019.4-2019.12	1608 (33.29)	0 (0)	0 (0)	1608 (33.2)
2020.9-2021.3	1609 (33.31)	0 (0)	0 (0)	1608 (33.2)
2021.9-2022.3	0 (0)	1609 (100)	0 (0)	1628 (33.61)

[a]AI: artificial intelligence.

Diagnostic Accuracy

Alterations in diagnostic accuracy for the identification of lung nodules and the assessment of lung lesion morphology were measured by comparing the reports of junior and senior radiologists (Table 2). The modification rate was defined as the ratio of changes made by senior radiologists to the diagnostic reports authored by junior radiologists, divided by the total number of radiology reports. Concerning lung nodule detection, both hospitals experienced an increase in their report modification rates after the introduction of the AI system, with Hospital A showing a statistically significant rise from 4.73% to 7.23% (χ^2_1=12.15; P<.001) in the modification rate. Conversely, in the case of lung lesion morphology, serving as a negative control, no significant differences were observed between the two hospitals before and after the implementation of the diagnostic AI system.

Table 2. Diagnostic accuracy of lung nodule detection and lung lesion morphology detection before and after artificial intelligence (AI) launch.

Modified or not	Hospital C (n=6439)				Hospital A (n=6450)			
	Before AI[a] (n=4830), n (%)	After AI (n=1609), n (%)	Chi-square (*df*)	*P* value	Before AI (n=1606), n (%)	After AI (n=4844), n (%)	Chi-square (*df*)	*P* value
Lung nodule			2.27 (1)	.13			12.15 (1)	<.001
Modified	322 (6.67)	125 (7.77)			76 (4.73)	350 (7.23)		
Not modified	4508 (93.33)	1484 (92.23)			1530 (95.27)	4494 (92.77)		
Lung lesion morphology			0.01 (1)	.91			0.27 (1)	.6
Modified	307 (6.36)	101 (6.28)			59 (3.67)	192 (3.96)		
Not modified	4523 (93.64)	1508 (93.72)			1547 (96.33)	4652 (96.04)		

[a]AI: artificial intelligence.

Lung Nodule Detection Ability

Alteration in lung nodule detection proficiency after the introduction of the AI system was also measured by comparing the reports of junior and senior radiologists (Table 3). At Hospital C, a notable and statistically significant rise in the detection rate was observed, increasing from 46.19% to 53.45% (χ^2_1=25.48; P<.001) after adopting the AI-assisted system. While a slight increase in the false negative rate and a decrease in accuracy were observed, the positive predictive value remained relatively consistent. Conversely, at Hospital A, the

XSL•FO
RenderX

AI system led to a significant enhancement in the detection rate of lung nodules, elevating the rate from 39.29% to 55.22% (χ^2_1=122.55; *P*<.001). Finally, the rollout of the AI system resulted in a reduction of the false negative rate from 8.4% to 5.16% (χ^2_1=9.85; *P*=.002) but led to an increase in the false positive rate from 2.36% to 9.77% (χ^2_1=53.48; *P*<.001). Nonetheless, the detection accuracy exhibited a decline, leading to a decrease in the positive predictive value, the rate from 96.17% to 92.29% (χ^2_1=11.41; *P*<.001).

Table 3. Lung nodule detection ability before and after artificial intelligence (AI) launch.

Hospital	Diagnostic method	Detection rate, %	False negative rate, %	False positive rate, %	Accuracy, %	Positive predictive value[a], %
Hospital C	CT[b] examination	46.19	10.98	2.96	93.33	96.27
	AI-assisted[c] system + CT examination	53.45	11.28	3.74	92.23	96.46
Hospital A	CT examination	39.29	8.4	2.36	95.27	96.17
	AI-assisted system + CT examination	55.22	5.16	9.77	92.77	92.29

[a]Positive predictive value: The proportion of correctly identified positive cases in the diagnosis.

[b]CT: computed tomography.

[c]AI-assisted: artificial intelligence–assisted.

Subgroup Analysis of Different AI Launch Times

To examine the patterns of diagnostic accuracy and detection ability of the AI system over time, we conducted subgroup analyses of different AI launch times (Table 4). Regarding lung nodules, the percentage of modified diagnoses increased from 7.21% in the first year to 8.40% in the second year but then decreased to 6.08% in the third year. Based on the Mann-Kendall test, the *z* score of –1.25 indicated a nonsignificant trend (*P*=.21). As for lung lesion morphology, the percentage of modified diagnoses remained relatively stable, with a slight increase to 4.42% in the third year. The *z* score of 0.83 also indicated a nonsignificant trend (*P*=.41). Most diagnoses remained unchanged during the 3 years following the launch of the AI system.

Table 4. Time trend analysis of lung nodule diagnostic accuracy from different artificial intelligence (AI) launch times at Hospital A.

Modified or not	1 year after AI[a] launch (n=1608), n (%)	2 years after AI launch (n=1608), n (%)	3 years after AI launch (n=1628), n (%)	*z*[b] score	*P* value
Lung nodule				–1.25	.21
Modified	116 (7.21)	135 (8.4)	99 (6.08)		
Not modified	1492 (92.79)	1473 (91.6)	1529 (93.92)		
Lung lesion morphology				0.83	.41
Modified	62 (3.86)	58 (3.61)	72 (4.42)		
Not modified	1546 (96.14)	1550 (96.39)	1556 (95.58)		

[a]AI: artificial intelligence.

[b]The test statistic for the Mann-Kendall test.

Alterations in lung nodule detection capability were measured over 3 years after the integration of the AI system (Table 5). The detection rates for lung nodules displayed a slight increase from 54.6% to 55.84%. In parallel, the false negative rates decreased from 7.06% to 4.07%. Nevertheless, the false positive rates exhibited fluctuations, reaching a peak of 13.33% in the second year. Despite these variations, the overall accuracy remained notably high, with a slight increase from 92.79% to 93.92%. Similarly, the positive predictive value demonstrated a comparable trend, with a minor decline in the second year.

XSL•FO
RenderX

Table 5. Time trend analysis of lung nodule detection ability from different artificial intelligence (AI) launch times at Hospital A.

Examination time	Detection rate, %	False negative rate, %	False positive rate, %	Accuracy, %	Positive predictive value[a], %
1 year after AI[b] launch	54.6	7.06	7.4	92.79	93.79
2 years after AI launch	55.22	4.39	13.33	91.6	89.84
3 years after AI launch	55.84	4.07	8.62	93.92	93.36

[a]Positive predictive value: The proportion of correctly identified positive cases in the diagnosis.

[b]AI: artificial intelligence.

Discussion

Enhanced Detection of Lung Nodules

Our study noted a substantial improvement in the detection rate of lung nodules following the implementation of the AI system. This improvement can be attributed to the AI's capability to augment radiologists' ability to spot small nodules that might indicate an underlying disease. AI algorithms are designed to methodically examine a wide range of medical images, thereby highlighting potential abnormalities that the human eye might miss. Furthermore, small lung nodules are often linked to early-stage diseases, which provide additional time for effective interventions, thus increasing the value of AI diagnostic systems in improving the detection of lung nodules in CT scans as well as the overall outcome of such patients [10].

Initial Decrease in Diagnostic Accuracy

Interestingly, compared with the diagnostic results of radiologists without AI assistance, our study observed a decline in the diagnostic accuracy of lung nodules within 3 years of the introduction of the AI system, mainly due to a significant rise in misdiagnosis rates. This intriguing observation calls for a more detailed investigation into the complex relationship between AI technology and radiologists' clinical decision-making. The increase in misdiagnosis rates could be due to the following reasons. First, the inherent characteristics of lung nodules present challenges for accurate detection. Specifically, CT imaging can be difficult when nodules are small, faint, or located in endobronchial or hilar regions [29]. In addition, pulmonary vasculature and artifacts were two of the main causes of false positive findings [30]. In addition, varying clinical attitudes of radiologists on reporting very small, and often microscopic, nodules may affect the detection accuracy. Radiologists may differ in their evaluation of clinical significance, resulting in varied reporting practices [31-33]. Furthermore, while the AI system is proficient in detecting nodules, it may also lead to a surge in cases with minimal clinical relevance, contributing to higher misdiagnosis rates. In order to reduce the false positives brought about by the introduced AI system, it is necessary to take comprehensive measures to maximize the effectiveness of the AI system in the future. In terms of algorithms, some algorithms have attempted to use a 2-step strategy, with the first step being the detection of candidate nodules with high specificity, followed by a false-positive reduction step in recent years [34]. For radiologists, before formally applying AI in the clinic, radiologists should receive training [35,36], which requires standardized diagnostic criteria to avoid diagnostic alterations due to differing attitudes of radiologists towards microscopic nodules. Meanwhile, promoting the standardized application of Chinese expert consensus on the diagnosis and treatment of pulmonary nodules (2024) in clinical practice is essential [6].

Long-Term Use and Temporal Trends

Our study revealed intriguing temporal trends in the diagnostic accuracy of lung nodules. Over time, we noticed a steady improvement in accuracy, indicating a dynamic process of adaptation and refinement. Several factors could potentially explain this phenomenon. One likely reason could be the ongoing optimization of AI algorithms through machine learning and feedback from actual clinical use [37,38]. As AI systems gain more experience in interpreting medical images and learning from radiologists' diagnostic patterns, their performance becomes increasingly refined. Furthermore, the growing familiarity and acceptance of AI systems among physicians could enhance their ability to work in tandem with these tools, thereby leading to improved diagnostic accuracy [39,40]. These pieces of evidence suggest that the AI system is expected to identify the unique features of lung nodules that differentiate them from other anatomies, as well as develop the capability for 3D reconstruction.

Strengths and Limitations

This study has several strengths. First, we conducted a statistical analysis of 12,889 diagnostic reports before and after the implementation of the AI system in 2 tertiary hospitals in China. The large sample size enabled us to draw reliable conclusions about the impact of the AI system on radiologists' performance. Second, we specifically selected the widely used lung nodule AI system to analyze detection rates and diagnostic accuracy, aiming to improve the quality of care and patient diagnosis in a real-world scenario. Finally, to evaluate the long-term application of the AI system, we included a random sample of diagnostic reports spanning 4 years, thus allowing us to conclude the temporal trends in the ongoing optimization process.

This study also has limitations. First, the research was conducted exclusively in 2 tertiary hospitals situated in Beijing, China. It is essential to recognize that junior radiologists, despite having relatively limited diagnostic experience, boast a high educational background and medical proficiency. Therefore, the influence detected by the introduction of the AI system in our study might not be generalizable to other institutions, specifically primary care settings. Furthermore, it is pertinent to acknowledge that our comparative benchmark was based on the diagnostic outcomes of senior radiologists. While the diagnoses rendered by senior radiologists are undoubtedly a robust reference, it is important to acknowledge that distinctions between clinical and

XSL•FO
RenderX

pathological assessments may exist, and senior radiologists' evaluation and revision can be influenced by the implementation of AI systems too.

For future investigations, it is advisable to contemplate the collection of prospective data and the incorporation of pathological results to enable a more comprehensive evaluation.

Conclusion

In summary, the integration of AI systems has yielded significant enhancements in lung nodule detection rates, particularly in the case of small nodules. This advancement, however, has been accompanied by a temporary decrease in diagnostic accuracy, primarily attributed to increased misdiagnosis rates, potentially arising from the influence of varying diagnostic criteria following the integration of AI systems and the performance of AI proficiency in detecting tiny lung nodules while ignoring clinical significance. Nonetheless, our research reveals encouraging trends over time, with diagnostic precision gradually ameliorating. This improvement can be ascribed to the continual refinement of AI algorithms and more effective collaboration among radiologists. Overall, our study underscores the promising role of AI in clinical settings, thus presenting opportunities for early disease identification and personalized patient care.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to restricted usage imposed by the patient privacy protection policy from Beijing Anzhen Hospital (Hospital A) and Tsinghua Changgung Hospital (Hospital C). The dataset applied in this study can be accessed with permission from the institutional review boards of Hospital A and Hospital C.

Authors' Contributions

WL contributed to study concepts and design, literature research, experimental studies, data analysis, statistical analysis, and manuscript preparation. YW was involved in study concepts and design, statistical analysis, and manuscript editing. ZZ participated in clinical studies and experimental studies. MB contributed to the study concepts and design. WY was involved in study concepts and design, clinical studies, experimental studies, and manuscript editing. HK acted as the guarantor of the integrity of the entire study and contributed to study concepts and design and manuscript editing.

Conflicts of Interest

None declared.

References

1. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69S:S36-S40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
2. Bonekamp D, Schlemmer HP. Artificial intelligence (AI) in radiology? : Do we need as many radiologists in the future? [Article in German]. Urologe A 2022;61(4):392-399. [doi: 10.1007/s00120-022-01768-w] [Medline: 35277758]
3. McDonald RJ, Schwartz KM, Eckel LJ, Diehn FE, Hunt CH, Bartholmai BJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Acad Radiol 2015;22(9):1191-1198. [doi: 10.1016/j.acra.2015.05.007] [Medline: 26210525]
4. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. JAMA 2012;307(22):2400-2409 [FREE Full text] [doi: 10.1001/jama.2012.5960] [Medline: 22692172]
5. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
6. Chinese Thoracic Society, Chinese Medical Association, Chinese Alliance Against Lung Cancer Expert Group. Chinese expert consensus on diagnosis and treatment of pulmonary nodules(2024) [Article in Chinese]. Zhonghua Jie He He Hu Xi Za Zhi 2024;47(8):716-729. [doi: 10.3760/cma.j.cn112147-20231109-00300] [Medline: 39069848]
7. Saber Tehrani AS, Lee H, Mathews SC, Shore A, Makary MA, Pronovost PJ, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. BMJ Qual Saf 2013;22(8):672-680. [doi: 10.1136/bmjqs-2012-001550] [Medline: 23610443]
8. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021;4(1):65 [FREE Full text] [doi: 10.1038/s41746-021-00438-z] [Medline: 33828217]
9. Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. Nat Commun 2020;11(1):3923 [FREE Full text] [doi: 10.1038/s41467-020-17419-7] [Medline: 32782264]
10. Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, et al. Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. JAMA Netw Open 2022;5(8):e2229289 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.29289] [Medline: 36044215]

XSL•FO
RenderX

11. Hong W, Hwang EJ, Lee JH, Park J, Goo JM, Park CM. Deep learning for detecting pneumothorax on chest radiographs after needle biopsy: clinical implementation. Radiology 2022;303(2):433-441. [doi: 10.1148/radiol.211706] [Medline: 35076301]
12. Li J, Zhou L, Zhan Y, Xu H, Zhang C, Shan F, et al. How does the artificial intelligence-based image-assisted technique help physicians in diagnosis of pulmonary adenocarcinoma? A randomized controlled experiment of multicenter physicians in China. J Am Med Inform Assoc 2022;29(12):2041-2049 [FREE Full text] [doi: 10.1093/jamia/ocac179] [Medline: 36228127]
13. Yi C, Tang Y, Ouyang R, Zhang Y, Cao Z, Yang Z, et al. The added value of an artificial intelligence system in assisting radiologists on indeterminate BI-RADS 0 mammograms. Eur Radiol 2022;32(3):1528-1537. [doi: 10.1007/s00330-021-08275-0] [Medline: 34528107]
14. Calisto FM, Santiago C, Nunes N, Nascimento JC. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. International Journal of Human-Computer Studies 2021;150:102607. [doi: 10.1016/j.ijhcs.2021.102607]
15. Ye FY, Lyu GR, Li SQ, You JH, Wang KJ, Cai ML, et al. Diagnostic performance of ultrasound computer-aided diagnosis software compared with that of radiologists with different levels of expertise for thyroid malignancy: a multicenter prospective study. Ultrasound Med Biol 2021;47(1):114-124. [doi: 10.1016/j.ultrasmedbio.2020.09.019] [Medline: 33239154]
16. Nguyen T, Maarek R, Hermann AL, Kammoun A, Marchi A, Khelifi-Touhami MR, et al. Assessment of an artificial intelligence aid for the detection of appendicular skeletal fractures in children and young adults by senior and junior radiologists. Pediatr Radiol 2022;52(11):2215-2226. [doi: 10.1007/s00247-022-05496-3] [Medline: 36169667]
17. Gould MK, Tang T, Liu IA, Lee J, Zheng C, Danforth KN, et al. Recent trends in the identification of incidental pulmonary nodules. Am J Respir Crit Care Med 2015;192(10):1208-1214. [doi: 10.1164/rccm.201505-0990OC] [Medline: 26214244]
18. Jiang B, Li N, Shi X, Zhang S, Li J, de Bock GH, et al. Deep learning reconstruction shows better lung nodule detection for ultra-low-dose chest CT. Radiology 2022;303(1):202-212. [doi: 10.1148/radiol.210551] [Medline: 35040674]
19. Nam JG, Hwang EJ, Kim J, Park N, Lee EH, Kim HJ, et al. AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. Radiology 2023;307(2):e221894. [doi: 10.1148/radiol.221894] [Medline: 36749213]
20. Nam JG, Park S, Hwang EJ, Lee JH, Jin K, Lim KY, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. Radiology 2019;290(1):218-228. [doi: 10.1148/radiol.2018180237] [Medline: 30251934]
21. Jonas DE, Reuland DS, Reddy SM, Nagle M, Clark SD, Weber RP, et al. Screening for lung cancer with low-dose computed tomography: updated evidence report and systematic review for the US preventive services task force. JAMA 2021;325(10):971-987. [doi: 10.1001/jama.2021.0377] [Medline: 33687468]
22. Shelmerdine SC, Martin H, Shirodkar K, Shamshuddin S, Weir-McCall JR, FRCR-AI Study Collaborators. Can artificial intelligence pass the Fellowship of the Royal College of Radiologists examination? Multi-reader diagnostic accuracy study. BMJ 2022;379:e072826 [FREE Full text] [doi: 10.1136/bmj-2022-072826] [Medline: 36543352]
23. Pan F, Li L, Liu B, Ye T, Li L, Liu D, et al. A novel deep learning-based quantification of serial chest computed tomography in coronavirus disease 2019 (COVID-19). Sci Rep 2021;11(1):417 [FREE Full text] [doi: 10.1038/s41598-020-80261-w] [Medline: 33432072]
24. Qi LL, Wu BT, Tang W, Zhou LN, Huang Y, Zhao SJ, et al. Long-term follow-up of persistent pulmonary pure ground-glass nodules with deep learning-assisted nodule segmentation. Eur Radiol 2020;30(2):744-755. [doi: 10.1007/s00330-019-06344-z] [Medline: 31485837]
25. Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. Clin Radiol 2020;75(1):13-19. [doi: 10.1016/j.crad.2019.04.017] [Medline: 31202567]
26. Tan HMJ, Tan MS, Chang ZY, Tan KT, Ee GLA, Ng CCD, et al. The impact of COVID-19 pandemic on the health-seeking behaviour of an Asian population with acute respiratory infections in a densely populated community. BMC Public Health 2021;21(1):1196 [FREE Full text] [doi: 10.1186/s12889-021-11200-1] [Medline: 34158019]
27. Sample size calculators for designing clinical research. URL: https://sample-size.net/ [accessed 2025-01-10]
28. The R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
29. Del Ciello A, Franchi P, Contegiacomo A, Cicchetti G, Bonomo L, Larici AR. Missed lung cancer: when, where, and why? Diagn Interv Radiol 2017;23(2):118-126 [FREE Full text] [doi: 10.5152/dir.2016.16187] [Medline: 28206951]
30. Li L, Liu Z, Huang H, Lin M, Luo D. Evaluating the performance of a deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing lung nodules: comparison with the performance of double reading by radiologists. Thorac Cancer 2019;10(2):183-192 [FREE Full text] [doi: 10.1111/1759-7714.12931] [Medline: 30536611]
31. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. Radiology 2017;284(1):228-243. [doi: 10.1148/radiol.2017161659] [Medline: 28240562]
32. Thoracic Surgery Committee, Department of Simulated Medicine, Wu Jieping Medical Foundation. Chinese experts consensus on artificial intelligence assisted management for pulmonary nodule (2022 Version) [Article in Chinese].

XSL•FO
RenderX

Zhongguo Fei Ai Za Zhi 2022;25(4):219-225 [FREE Full text] [doi: 10.3779/j.issn.1009-3419.2022.102.08] [Medline: 35340198]
33. Bai C, Choi CM, Chu CM, Anantham D, Chung-Man Ho J, Khan AZ, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for Asia. Chest 2016;150(4):877-893. [doi: 10.1016/j.chest.2016.02.650] [Medline: 26923625]
34. de Margerie-Mellon C, Chassagnon G. Artificial intelligence: a critical review of applications for lung nodule and lung cancer. Diagn Interv Imaging 2023;104(1):11-17 [FREE Full text] [doi: 10.1016/j.diii.2022.11.007] [Medline: 36513593]
35. Abuzaid MM, Elshami W, Tekin H, Issa B. Assessment of the willingness of radiologists and radiographers to accept the integration of artificial intelligence into radiology practice. Acad Radiol 2022;29(1):87-94. [doi: 10.1016/j.acra.2020.09.014] [Medline: 33129659]
36. Bergquist M, Rolandsson B, Gryska E, Laesser M, Hoefling N, Heckemann R, et al. Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. Eur Radiol 2024;34(1):338-347 [FREE Full text] [doi: 10.1007/s00330-023-09967-5] [Medline: 37505245]
37. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med 2021;385(3):283-286 [FREE Full text] [doi: 10.1056/NEJMc2104626] [Medline: 34260843]
38. Liu Y, Chen PHC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA 2019;322(18):1806-1816. [doi: 10.1001/jama.2019.16489] [Medline: 31714992]
39. Dendl LM, Pausch AM, Hoffstetter P, Dornia C, Höllthaler J, Ernstberger A, et al. Structured reporting of whole-body trauma CT scans using checklists: diagnostic accuracy of reporting radiologists depending on their level of experience. Rofo 2021;193(12):1451-1460 [FREE Full text] [doi: 10.1055/a-1541-8265] [Medline: 34348402]
40. Chen Y, Stavropoulou C, Narasinkan R, Baker A, Scarbrough H. Professionals' responses to the introduction of AI innovations in radiology and their implications for future adoption: a qualitative study. BMC Health Serv Res 2021;21(1):813 [FREE Full text] [doi: 10.1186/s12913-021-06861-y] [Medline: 34389014]

Abbreviations

AI: artificial intelligence
CT: computed tomography

Edited by A Coristine; submitted 23.07.24; peer-reviewed by J Jagtap, L Wang; comments to author 21.10.24; revised version received 10.12.24; accepted 27.12.24; published 27.01.25.

Please cite as:
Liu W, Wu Y, Zheng Z, Bittle M, Yu W, Kharrazi H
Enhancing Diagnostic Accuracy of Lung Nodules in Chest Computed Tomography Using Artificial Intelligence: Retrospective Analysis
J Med Internet Res 2025;27:e64649
URL: https://www.jmir.org/2025/1/e64649
doi:10.2196/64649
PMID:

©Weiqi Liu, You Wu, Zhuozhao Zheng, Mark Bittle, Wei Yu, Hadi Kharrazi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

MetaAnalysisOnline.com: Web-Based Tool for the Rapid Meta-Analysis of Clinical and Epidemiological Studies

János Tibor Fekete[1,2], PhD; Balázs Győrffy[1,2,3], MD, PhD

[1]Department of Bioinformatics, Semmelweis University, Budapest, Hungary

[2]Cancer Biomarker Research Group, Institute of Molecular Life Sciences, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

[3]Department of Biophysics, Medical School, University of Pecs, Pecs, Hungary

Corresponding Author:
Balázs Győrffy, MD, PhD
Department of Bioinformatics
Semmelweis University
Tuzolto u 7-9
Budapest, H-1095
Hungary
Phone: 36 305142822
Email: gyorffy.balazs@yahoo.com

Abstract

Background: A meta-analysis is a quantitative, formal study design in epidemiology and clinical medicine that systematically integrates and quantitatively synthesizes findings from multiple independent studies. This approach not only enhances statistical power but also enables the exploration of effects across diverse populations and helps resolve controversies arising from conflicting studies.

Objective: This study aims to develop and implement a user-friendly tool for conducting meta-analyses, addressing the need for an accessible platform that simplifies the complex statistical procedures required for evidence synthesis while maintaining methodological rigor.

Methods: The platform available at MetaAnalysisOnline.com enables comprehensive meta-analyses through an intuitive web interface, requiring no programming expertise or command-line operations. The system accommodates diverse data types including binary (total and event numbers), continuous (mean and SD), and time-to-event data (hazard rates with CIs), while implementing both fixed-effect and random-effect models using established statistical approaches such as DerSimonian-Laird, Mantel-Haenszel, and inverse variance methods for effect size estimation and heterogeneity assessment.

Results: In addition to statistical tests, graphical representations including the forest plot, the funnel plot, and the *z* score plot can be drawn. A forest plot is highly effective in illustrating heterogeneity and pooled results. The risk of publication bias can be revealed by a funnel plot. A *z* score plot provides a visual assessment of whether more research is needed to establish a reliable conclusion. All the discussed models and visualization options are integrated into the registration-free web-based portal. Leveraging MetaAnalysisOnline.com's capabilities, we examined treatment-related adverse events in patients with cancer receiving perioperative anti–PD-1 immunotherapy through a systematic review encompassing 10 studies with 8099 total participants. Meta-analysis revealed that anti–PD-1 therapy doubled the risk of adverse events (risk ratio 2.15, 95% CI 1.39-3.32), with significant between-study heterogeneity (I^2=95%) and publication bias detected through the Egger test (*P*=.02). While these findings suggest increased toxicity associated with anti–PD-1 treatment, the *z* score analysis indicated that additional studies are needed for definitive conclusions.

Conclusions: In summary, the web-based tool aims to bridge the void for clinical and life science researchers by offering a user-friendly alternative for the swift and reproducible meta-analysis of clinical and epidemiological trials.

(J Med Internet Res 2025;27:e64016) doi:10.2196/64016

KEYWORDS

statistics; pharmacology; treatment; epidemiology; fixed effect model; random effect model; hazard rate; response rate; clinical trial; funnel plot; z score plot

XSL•FO
RenderX

Introduction

Systematic literature reviews in medicine summarize the best-proven knowledge from a research domain into a single paper [1]. A meta-analysis is the statistical combination of results from at least 2 separate studies which can serve as a powerful tool for summarizing, integrating, and interpreting quantitative research findings from the collected sources. The first potential advantage of a meta-analysis is its capacity to enhance precision. Many studies, when considered in isolation, may be too small to provide convincing evidence about intervention effects, and a meta-analysis of multiple studies can provide the necessary power to investigate whether interventions have an impact on the incidence of rare events.

Second, meta-analyses allow for the exploration of questions not explicitly addressed by individual studies. Primary studies often focus on specific participant types and interventions with well-defined parameters. By incorporating studies with diverse characteristics, it becomes possible to investigate the consistency of effects across a broader range of populations and interventions. Additionally, this approach may shed light on reasons for variations in effect estimates.

Third, meta-analyses can play a role in settling controversies that may arise from apparently conflicting studies or in generating new hypotheses. The statistical synthesis of findings enables formal assessment of conflicting results, providing a means to explore and quantify the reasons for different outcomes.

On the other hand, meta-analyses can be limited by inadequate search strategies, lack of clarity on comparators, insufficient assessment of risk of bias in primary studies, failure to address heterogeneity across studies, and the use of inappropriate or nonstandard methodological approaches, all of which can undermine the reliability and validity of the conclusions drawn [2]. Recent systematic evaluation of meta-analyses focusing on digital biomarker interventions revealed significant methodological inadequacies, with 92% of studies demonstrating critically low quality, highlighting the need for more rigorous methodological approaches in meta-analytic research [3]. Furthermore, meta-analyses often fail to adequately consider important demographic factors and subgroup analyses, as evidenced by a recent overview of systematic reviews on digital technologies in chronic obstructive pulmonary disease where only 1 out of 30 reviews included age-based subgroup analysis, despite the potential impact of such factors on treatment outcomes and clinical applicability [4].

Here, the study aims to establish a new web portal for the simple and rapid execution of meta-analysis studies. In a meta-analysis, first, a summary statistic is computed for each study to describe the observed intervention effect uniformly across all studies. Then, a combined intervention effect estimate is calculated as a weighted average of the intervention effects estimated in the individual studies. We list options used by our web-based tool and for an all-inclusive description of the different approaches we suggest consulting statistical papers like the publications by Haidich [5] and by Tawfik et al [6]. In the last part of our paper, we describe a sample application where we analyzed the correlation between immune checkpoint blockade and adverse events.

Methods

Web-Based Meta-Analysis Portal

Our main goal was to create a web-based application that allows users to perform a complete meta-analysis of clinical trials with minimal effort. For this, we established a user-friendly portal that is accessible in any major web browser. The web application is running on an Ubuntu server (Canonical Ltd) powered by Apache. The application can be used to perform meta-analyses using binary (using total and event numbers), continuous (using mean and SD data), and time-to-event data (using hazard ratio and CI data). The results of each study can be pasted from commonly used spreadsheet applications like Microsoft Excel.

The calculations and plots used by the web-based system are done using *meta*, *RTSA*, and *metafor* packages in R programming environments (version 4.2.2; R Foundation for Statistical Computing). The Shiny user interface was created by using the *shinyjs*, *shinydashboard*, and *rhandsontable* R packages. The portal generates a forest plot to summarize the results of the meta-analysis. A 2-tailed *P* value below .05 indicates significance and is displayed in the analysis results. MetaAnalysisOnline.com uses the DerSimonian-Laird estimator to estimate tau-square [7].

In addition to the forest plot, a funnel plot is displayed to detect potential bias visually, and a *z* score plot is used for assessing the power achieved by the cumulative sample number. While there is commercial software available for performing a meta-analysis [8], the portal we developed can be accessed completely free, without the need for registration or logging in.

The Implemented Statistical Methods

Experimental data are typically classified into 2 main categories: clinical data, which detail the effects of specific treatments on individual participants; and epidemiological data, which unveil patterns of disease or mortality in groups of participants exposed to either single agents or various substances. The most common classification in such studies uses categorical variables. A categorical variable is nonnumerical, relying on qualitative properties such as treatment arm, race, or gender, among others. Unlike numerical variables, categorical variables lack a specific ordering and assume values from a finite set of possibilities.

In MetaAnalysisOnline.com, the 2 study arms can be compared by either the fixed-effect model or the random-effect model. The fixed effect model operates on the assumption that a single true effect size underlies all studies in the meta-analysis, hence the term “fixed effect” or “common-effect” or sometimes “equal-effects.” Any variations in observed effects are attributed to sampling error. The random effects model posits that the true effect may vary across studies due to inherent differences or heterogeneity between them. As studies encompass diverse participant compositions and different interventions, individual studies may yield varying effect sizes. If an infinite number of studies were conducted, the effect estimates from all studies would conform to a normal distribution and the pooled estimate would then represent the mean or average effect. The assumption

XSL•FO
RenderX

is that the effect sizes observed in conducted studies represent a random sample from the universe of all possible effect sizes, hence the term "random effects" [9]. The random effects approach for aggregating evidence from a series of experiments that compare 2 treatments was originally proposed by DerSimonian and Laird [7]. They integrated the heterogeneity of effects into the analysis, providing a comprehensive assessment of the overall treatment efficacy. The heterogeneity can be assessed using the I^2 and ^{2} statistics, where the ^{2} represents the between-trial variance of the effect size. The square root of the ^{2} provides an estimate of the SD of the effects in the analyzed studies. High values of ^{2} indicate substantial heterogeneity among the effect sizes, suggesting that the true effects are not consistent across studies and may be influenced by other factors [10]. The I^2 method, developed by Higgins and Thompson [11], offers a percentage value representing the variability observed in the effect size, independent of sampling error. One of its advantages lies in the ease of interpreting its results: I^2 percentages of 25% correspond to low, 50% to medium, and 75% to high levels of heterogeneity, respectively [12].

There are 4 commonly used methods of meta-analysis for dichotomous outcomes: 4 fixed-effect methods (Mantel-Haenszel [13], Peto, Bakbergenuly [14], and inverse variance) and 1 random-effect method (DerSimonian and Laird inverse variance [7]). The Peto method is limited to combining odds ratios (ORs), while the other 3 methods can combine ORs, risk ratios, or risk differences. The Mantel-Haenszel method is a fixed effect meta-analysis method that uses a different weighting scheme depending on the type of effect measure being used (eg, risk ratio or OR) [13]. A relatively recent development is the Bakbergenuly sample-size-weighted estimator approach, which offers an alternative by using a weighted average where each study's weight is determined by its effective sample size [14]. In the fixed effect model, the weights are calculated as $w_i=1/v_1$ where v_1 represents the variance of the study. In the calculation, also known as the inverse variance method, studies with lower variances are weighted higher. In the random effect model, the calculation of weights is different, and the formula is $_i=1/(\ ^2+v_i)$, where ^{2} represents the square of the variance of the distribution of true effect sizes. The inverse-variance method's key advantage is its wide applicability, as it can be used for all effect measures and not only for OR, risk ratio, or risk difference. For its execution, we need to have the point estimate and the variance of the treatment effect in each study [7].

Quantitative variables can be either discrete or continuous. Discrete variables are constrained to a finite number of values (such as whole numbers), whereas continuous variables can assume any value, including all possible values within a given range (extending to an infinite number of decimal places). Standard methods for meta-analysis of continuous data rely on the assumption that the outcomes follow a normal distribution within each intervention arm in each study. The 3 commonly used summary statistics for a meta-analysis of continuous data include the mean difference (MD), the standardized mean difference (SMD), and the ratio of means. The choice of summary statistics for continuous data primarily depends on whether studies report the outcome using the same scale (in which case MD is used) or different scales (in which case the SMD is typically used). In the MD approach, the SDs and sample sizes are used to calculate the weight assigned to each study and studies with smaller SDs are given relatively larger weights. In the SMD approach, the SDs are used to standardize the MDs to a single scale and are also used in the computation of study weights. When comparing with MD, interpreting the results can be challenging due to the need to categorize the SMD. Typically, we use the categories outlined by Cohen, which indicate that an SMD of approximately 0.2 corresponds to a small effect, around 0.5 signifies a moderate effect, and approximately 0.8 indicates a large effect [15]. In addition to MD and SMD, MetaAnalysisOnline.com also includes the ratio of means, a more recent solution where instead of using the difference between groups, the log-transformed ratio of the means of the groups is used in the analysis. The advantage of this approach is that the output of the result is easily interpretable similar to the odds or risk ratio [16].

Demonstration: The Correlation Between Immune Checkpoint Blockade and Adverse Events

In recent years, the addition of immune checkpoint blockade using anti–PD-1 drugs, such as pembrolizumab, nivolumab, and cemiplimab, into cancer therapy has improved clinical outcomes. However, the safety of the immune checkpoint blockade needs further evaluation. As a demonstration of the functionalities of MetaAnalysisOnline.com, a set of previous studies was analyzed to assess how incorporating immune checkpoint blockade into perioperative cancer therapy affects treatment-related adverse events in patients treated with anti–PD-1 therapy.

Results

Graphical Representation of the Overall Effect: The Forest Plot

A comprehensive forest plot displays a set of the following information.

- The name of the original studies used in the meta-analysis. It is crucial to ensure the uniqueness of study names, particularly when an author or study name is listed more than once.
- The number of events and the total number of participants in each group of the study.
- A graph representing the relative risk and 95% CIs. Each square in the plot embodies the study results, centered on the point estimate, with a horizontal line indicating the 95% CI. The diamond represents the overall meta-analysis estimate, with the center signifying the pooled estimate and the prediction interval indicating the confidence limits. The CI depicts the range of intervention effects compatible with the study's result, while the size of the block draws attention to the studies with larger weight, which dominates the calculation of the summary result presented by the diamond at the bottom.

XSL•FO
RenderX

- The risk ratio or hazard ratio and 95% CI for each individual study and the overall meta-analysis.
- The percentage of weight assigned to each study, particularly when presenting pooled results. Studies with higher sample numbers or those with narrower CIs receive higher weights.
- Finally, at the bottom of the plot, the heterogeneity and the overall effect values. A significant *P* value for heterogeneity indicates a substantial difference between studies. However, they heavily rely on the assumption of a normal distribution for the effects across studies and can present challenges when the number of studies is small, potentially leading to artificially wide or narrow intervals.

The features discussed in this section are marked in the plot generated by demo data in Figure 1. In addition, a screenshot of the analysis page with a generated forest plot is shown in Figure 2.

Figure 1. Components of a forest plot.

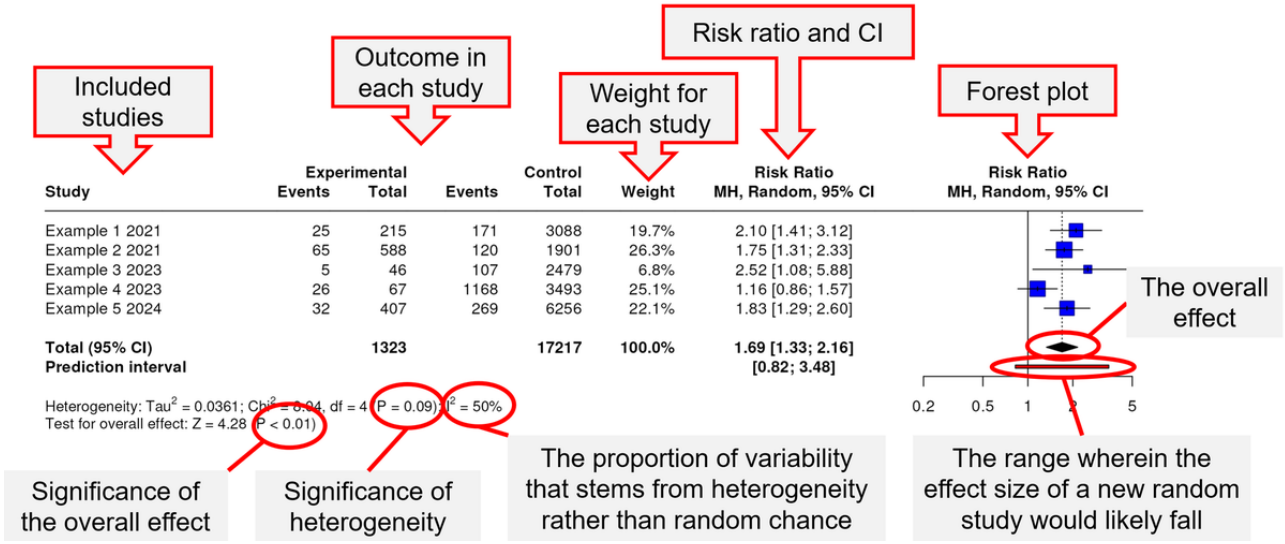

Figure 2. Screenshot of the web-based analysis interface showing the demonstration data and the generated forest plot.

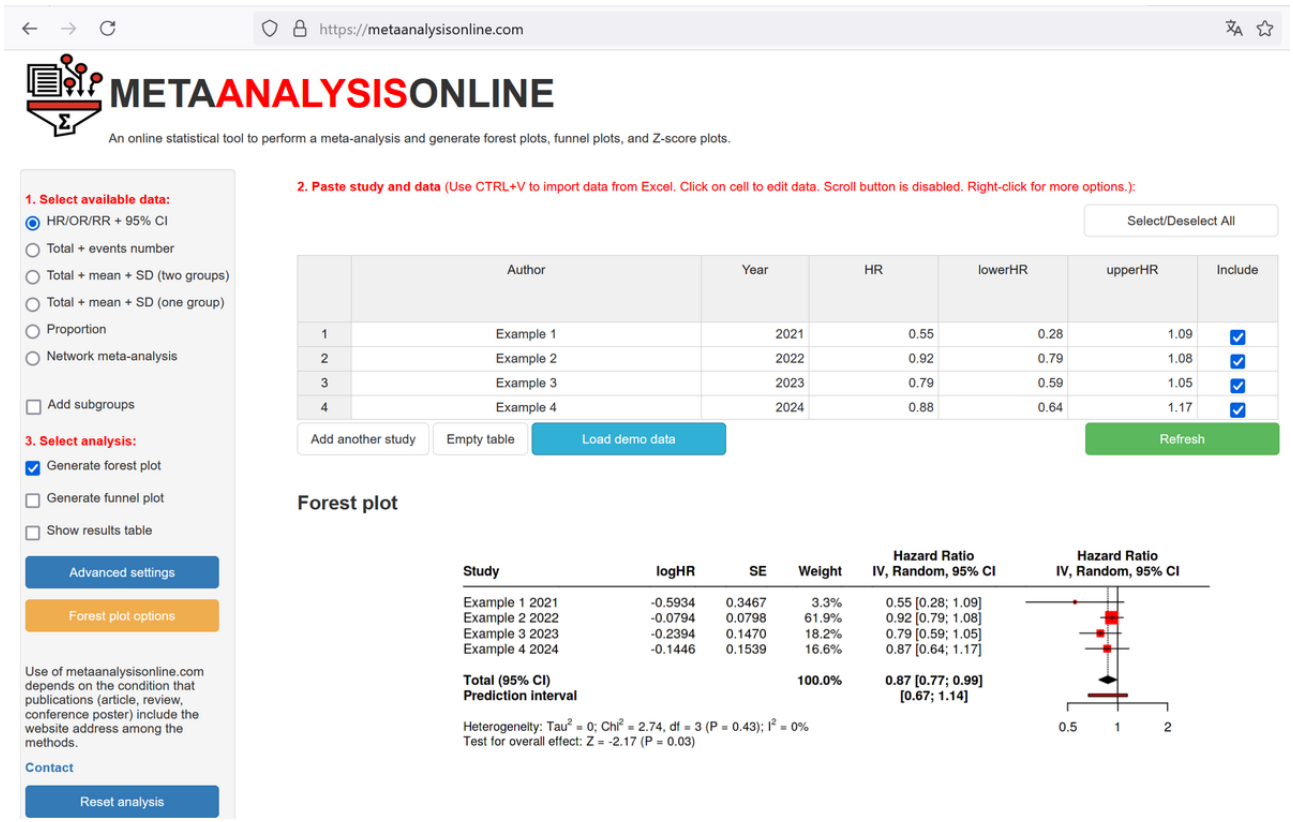

XSL•FO
RenderX

Graphical Representation of Heterogeneity: The Funnel Plot

The second analysis available in MetaAnalysisOnline.com is the funnel plot designed to detect publication bias. Whether the primary studies for pooling are observational studies or randomized controlled trials, heterogeneity between studies is probable. We show the components of a funnel plot used to assess study heterogeneity in Figure 3. Each dot on the plot represents an individual study. The x-axis represents the hazard ratio, or in other cases the log of OR, indicating the effect size of the treatment in each study. The y-axis represents the standard error, reflecting study precision, with larger studies having better precision (within the green area) and smaller studies having worse precision (within the orange area). In the plot, the vertical solid black or red line serves as the reference line corresponding to no effect or a hazard rate of 1 (or log OR=0). The overall effect and 95% CIs are depicted by the dashed black lines.

Figure 3. Components of a funnel plot. HR: hazard ratio.

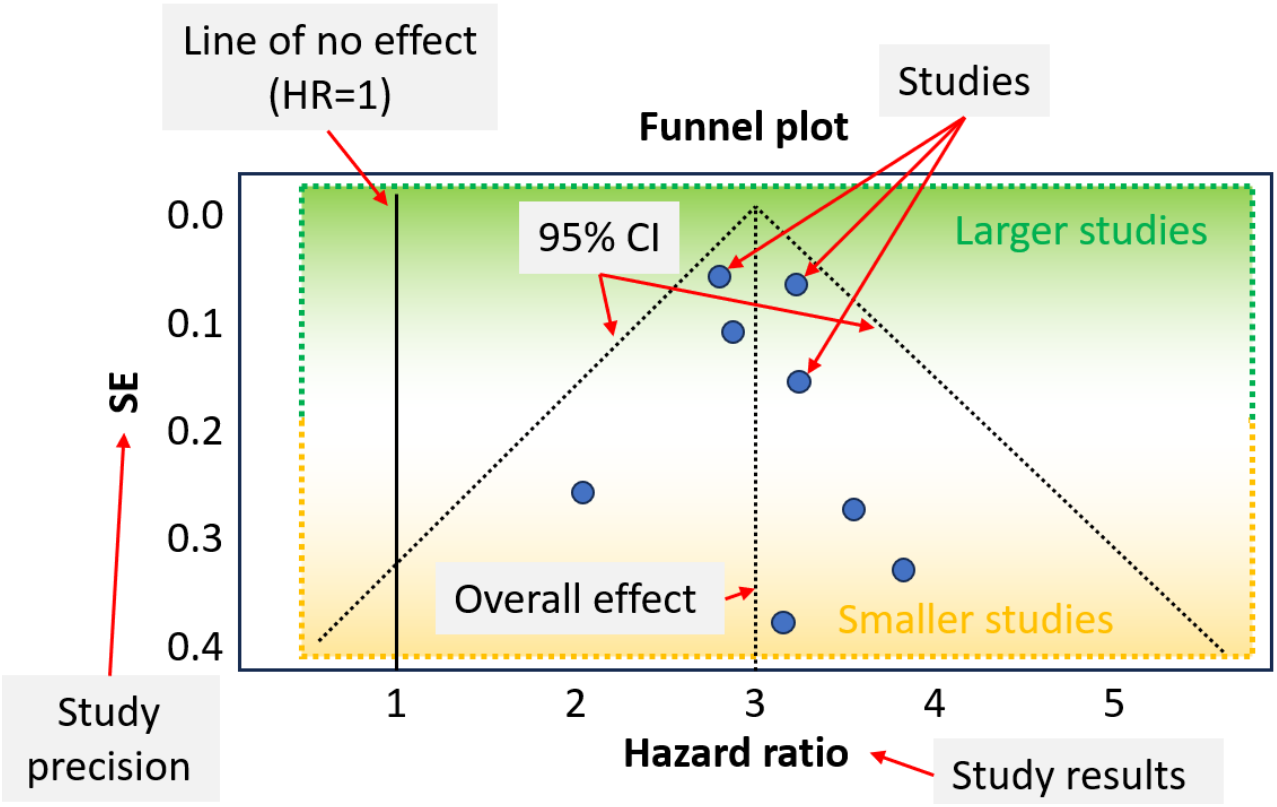

If the funnel plot suggests the presence of publication bias, the trim-and-fill method can be used to estimate the number of missing studies and impute their effect sizes. This allows the meta-analysis to be “trimmed” of asymmetric studies, and “filled” with the estimated missing studies to create a more symmetric funnel plot. The web-based platform automatically displays a button to generate trim-and-fill plots in case of a significant Egger test.

Trial Sequential Analysis and the z Score Plot

MetaAnalysisOnline.com performs a trial sequential analysis by drawing the *z* score plot to illustrate the reliability of the sample size. The y-axis of a *z* score plot depicts the cumulative *z* score and the x-axis the cumulative sample size. The cumulative *z* curve is built by sequentially adding studies based on chronological order. The plot depicts the relationship between the collective sample size and the achieved significance. When the ideal sample size is attained, a decisive conclusion regarding the examined condition can be made. The components of a *z* score plot are visualized in Figure 4.

XSL•FO
RenderX

Figure 4. The z score plot visualizes the robustness of the sample number. The plot displays the correlation between the cumulative sample number and significance after the addition of each new study. Once the optimal sample number is reached, a definitive conclusion can be drawn about the investigated condition. As the displayed z curve includes more and more patients, it seems to favor outcome B by the conventional cut-offs—however, this sample number is not yet sufficient for statistical significance (red asterisk). As further studies are included, the cumulative outcome reaches statistical significance (green asterisk). Finally, after adding the last study, the z curve surpasses the optimal sample number indicating that no further research is needed.

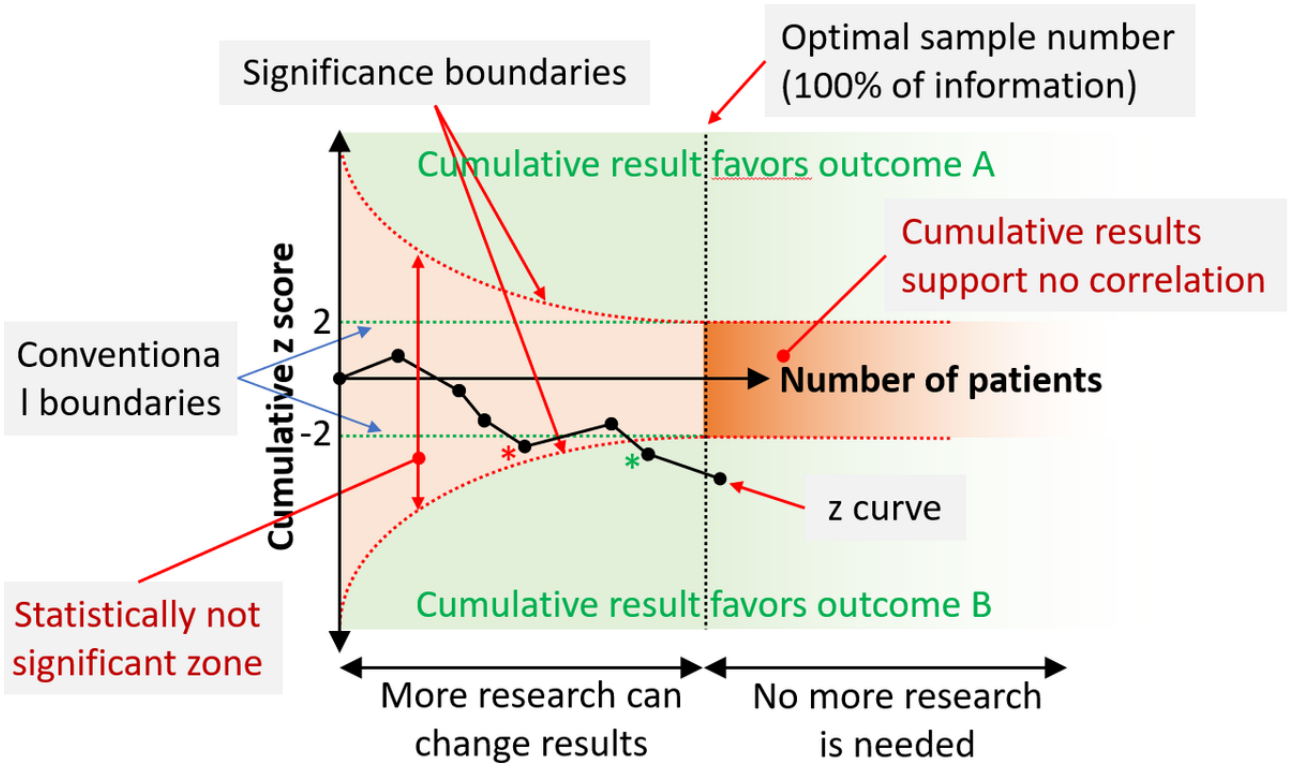

At the end of the *z* curve lies the most recently added study, falling into one of the following zones: "favoring outcome A," "favoring outcome B," "no correlation," or "statistically not significant." The first 2 zones indicate statistically significant results, which may vary until the optimal sample size is reached, but after this, a reliable final conclusion can be determined. Placement in the "no correlation" signifies strong evidence suggesting that subsequent studies are unlikely to alter the no-effect outcome significantly. Placement within the "not statistically significant" zone indicates the definitive requirement for further studies as neither correlation nor missing correlation could be determined.

Independent Validation

To verify the output of the digital portal, a set of 10 samples was drawn from the dataset of a simulated randomized trial developed for the *metafor* package and a meta-analysis was performed. Analysis was repeated using IBM SPSS (version 29). The following outputs were compared: OR, the calculated weight, overall effect, as well as measures of heterogeneity (I^2 and τ^2). For all parameters, the corresponding 95% CIs were also checked. The output results of the 2 methods were identical, indicating sufficient reliability of the implemented procedures.

Demonstration: Immune Checkpoint Blockade and Adverse Events

To show the entire analysis pipeline, the association between immune checkpoint blockade using anti–PD-1 agents and adverse events was analyzed. Altogether 10 studies were included with a total of 4379 participants in the anti–PD-1–treated cohort and 3720 participants in the control cohort [17-26]. Based on the analysis performed using the random effects model with the Mantel-Haenszel method to compare the risk ratio, there is a statistical difference between the 2 cohorts, the overall risk ratio is 2.15 with a 95% CI of 1.39-3.32. The test for overall effect reached significance as depicted in Figure 5A.

XSL•FO
RenderX

Figure 5. (A) Immune checkpoint blockade by anti-PD-1 drug class significantly increases adverse events. However, the funnel plot shows a (B) potential publication bias and (C) the z score plot indicates insufficient patient number of patients to draw a definitive conclusion (at significance=.05, type II error level=.2, and minimal clinically relevant outcome=30%) [30-39]. AIS: actual information size; HARIS: heterogeneity adjusted recommended information size; MH: Mantel-Haenszel method.

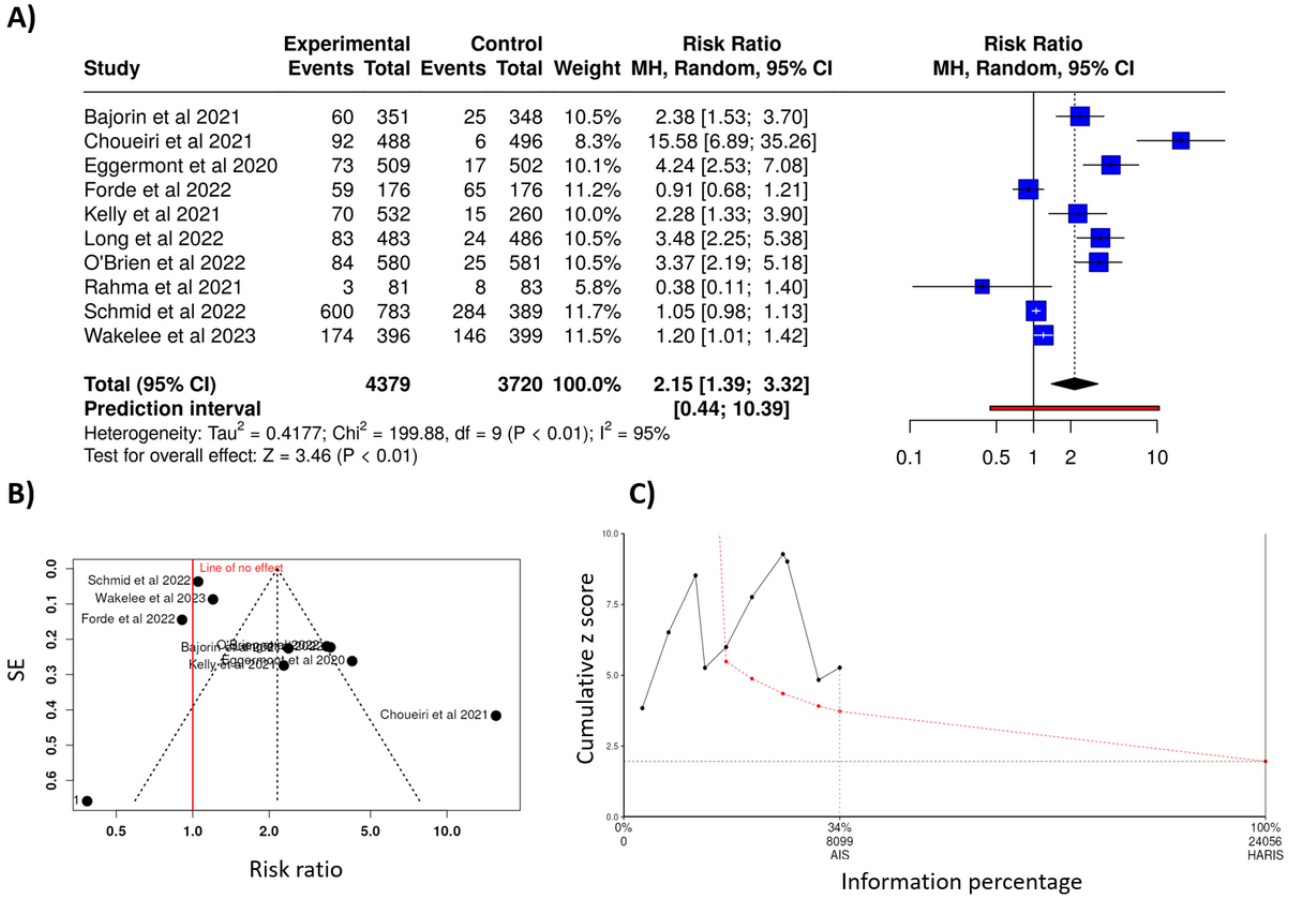

Notably, substantial heterogeneity was detected, suggesting inconsistent effects in magnitude and/or direction. The I^2 value indicates that 95% of the variability among studies stems from heterogeneity rather than random chance.

The funnel plot indicates a potential publication bias. The Egger test supports the presence of funnel plot asymmetry (intercept: 3.85, 95% CI 1.39-6.3; t_8=3.072; P=.02; visualized in Figure 5B).

Based on the *z* score plot, the total number of samples did not reach the optimal number, and more research is needed to establish a conclusion (Figure 5C).

Discussion

Here, we have established a new aid for those performing systematic reviews and meta-analyses. A major advantage of our tool is its direct design for life sciences researchers and medical professionals who generally lack extensive statistical and bioinformatic skills. The user-friendly interface enables the execution of the analysis in real-time. Entering the data is possible by directly copying from a database with a single click. The tool generates the 3 most important graphical representations–a forest plot, a funnel plot, and a *z* score plot—in real-time. The most important advantage of the tool over other software lies in its extremely user-friendly setup—neither installation nor registration is necessary for the use and the input page was set up to enable prompt copying of data from popular spreadsheet applications.

The most important graphical result provided by MetaAnalysisOnline.com is a forest plot, which serves as a visual depiction of both the outcomes from individual studies and the overall results of the analysis [27]. Additionally, it illustrates overall effect estimates and the heterogeneity of studies, reflecting the variation in results among individual studies [28]. The plot also includes the prediction interval, which is widely used to express the amount of heterogeneity in a meta-analysis [29]. Notably, many reviews in the literature use slightly different forest plots as there is no uniform standardization on their appearance. A previous study investigating over 2000 forest plots from the literature identified highly standardized plots in Cochrane reviews but missing key elements in non-Cochrane reviews. Certain formats were not optimal for information exchange, and a significant number of plots contained insufficient data to be considered useful [30].

The second implemented visual result is the funnel plot. In this, results from smaller studies will scatter broadly at the bottom of the graph, while the dispersion is likely to decrease among larger studies. In the absence of bias, the plot will take on the appearance of a symmetrical, inverted funnel, with asymmetry suggesting the presence of a risk of publication bias [31]. A

XSL•FO
RenderX

significant Egger test result means that the points on the funnel plot are not symmetric. It is important to acknowledge that in practical scenarios, the source of heterogeneity cannot be determined and is often unknown. It is important to distinguish between various types of heterogeneity. Clinical heterogeneity refers to variability in participants, interventions, and outcomes studied, while methodological heterogeneity pertains to variability in study design, outcome measurement tools, and risk of bias. Statistical heterogeneity arises from variability in the intervention effects being evaluated in different studies, which can be a result of clinical or methodological diversity, or both. I^2 and ^{2} statistics can be used to assess study heterogeneity [32] and although the funnel plot is primarily used to detect publication bias, it can also be used to visualize the heterogeneity [33]. To provide a meaningful summary, it is crucial to ensure that the group of studies is sufficiently homogeneous in terms of participants, interventions, and outcomes before conducting a meta-analysis.

A fundamental question omitted in many performed meta-analyses is assessing the robustness of the sample number reached [34]. In particular, meta-analyses face the risk of yielding misleading outcomes due to significant results that may be false positives (type I errors; α) or falsely insignificant findings (type II errors; β). These errors can stem from various sources such as low-quality or underpowered trials, publication bias, and excessive significance testing. To address this issue, trial sequential analysis (TSA) has been devised as a cumulative meta-analysis technique [35]. TSA aims to account for both type I and type II errors, providing a means to estimate the point at which the observed effect size becomes robust enough to resist further influence from additional studies. In other words, TSA offers guidance on the necessity of further studies and helps clinicians prevent unnecessary trials [36].

Our tool was established to incorporate a simple way to conduct subgroup analysis as well—to examine the potential variability of treatment effects among different subgroups of patients or trials [37]. Subgroup analysis involves dividing all participant data in the meta-analysis into subsets based on patient characteristics (eg, age) or trial characteristics (eg, location), and then performing a meta-analysis on one or more of these subsets. These analyses can help to estimate treatment effects for clinically relevant subgroups of patients or to identify sources of heterogeneity [38]. The decision to conduct such analyses may be based on prior research suggesting that treatment effects could differ among different patient subgroups.

Two considerations should be kept in mind when running the statistical analysis using MetaAnalysisOnline.com. Sometimes the data points scatter over a wide range and truncating the data is advisable (to constrain the length of the follow-up time, the applied treatment dose, or any other descriptive characteristics of the participants) at a selected threshold in case of scarcity of data above the threshold value. A second issue refers to the inherent differences between the studies due to the varying number of included participants. To achieve the most robust results, studies should be weighed for sample size to better reflect the characteristics of the general population.

We demonstrate the use of MetaAnalysisOnline.com by performing a meta-analysis of studies examining the association between immune checkpoint blockade for anti-PD-1 agents and adverse events [39]. Although immune checkpoint blockade significantly increased the rate of adverse events, we observed a significant heterogeneity, and we can conclude that more research is needed to establish an accurate correlation.

Despite the meta-analysis's aim to identify and evaluate all relevant studies meeting inclusion criteria, this goal may not always be fully realized. Some studies might be overlooked, particularly if they are not published in English or if they report nonsignificant results, making them less likely to be published. Understanding and accounting for these factors is essential for a comprehensive and critical appraisal of a meta-analysis.

Finally, we have to discuss some limitations of our tool. In order to retain user-friendliness, we had to make some compromises. The maximal number of studies in 1 analysis is limited to 100. We have established a default format for the plots, and the graphical parameters (eg, font type and size, spacing, and line thickness) cannot be changed. In addition, we used Shiny to enable seamless visualization and scaling. Because of this, the server behind the platform can sometimes slow down when a large number of studies are analyzed in 1 setting.

In summary, MetaAnlysisOnline.com can be used for a quick and comprehensive meta-analysis in clinical and epidemiological trials. When presenting the results of a meta-analysis, 3 key aspects need to be considered. First, we have to examine the pooled result, representing the overall combined outcome obtained by pooling the individual studies in a forest plot. While providing a synthesized perspective on the collective findings, we have to consider heterogeneity, which relates to variations in results, methodology, or study populations across the included studies. Second, we must assess the potential presence of the risk of publication bias. Finally, the *z* score plot can be used to determine whether the cumulative sample number is sufficient for a definitive conclusion, or whether additional studies are still needed.

Acknowledgments

The 4.0 version of ChatGPT, developed by OpenAI, was used as a language tool to refine our writing and enhance the clarity of our work. This project was supported by the National Research, Development, and Innovation Office (PharmaLab, RRF-2.3.1-21-2022-00015). A5 Genetics Ltd (Kutaso, Hungary) provided computational infrastructure for the study.

Data Availability

All data generated or analyzed during this study are included in this published article.

XSL•FO
RenderX

Authors' Contributions

FJT contributed to the study by developing the conceptual framework and designing the methodology. He curated and analyzed the data, ensuring the accuracy and validity of the findings. His expertise in software development and visualization allowed for the effective presentation of results. Throughout the research process, FJT played a key role in validating the data and refining the study through critical review and editing. BG provided the overall direction of the project, securing funding and managing resources to support its execution. He played a central role in conceptualizing the study, overseeing methodological development, and ensuring the validity of the findings. BG took the lead in drafting the original manuscript and also contributed to data visualization.

Conflicts of Interest

None declared.

References

1. Paul J, Barari M. Meta‐analysis and traditional systematic literature reviews—What, why, when, where, and how? Psychol Mark 2022;39(6):1099-1115. [doi: 10.1002/mar.21657]
2. Kajal F. Clarity on the type of review. Comment on "Value Cocreation in Health Care: Systematic Review". J Med Internet Res 2022;24(7):e38457 [FREE Full text] [doi: 10.2196/38457] [Medline: 35816377]
3. Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, Abid MM, Péntek M, Gulácsi L, et al. Digital biomarker-based interventions: systematic review of systematic reviews. J Med Internet Res 2022;24(12):e41042 [FREE Full text] [doi: 10.2196/41042] [Medline: 36542427]
4. Matthias K, Honekamp I, Heinrich M, De Santis KK. Consideration of sex, gender, or age on outcomes of digital technologies for treatment and monitoring of chronic obstructive pulmonary disease: overview of systematic reviews. J Med Internet Res 2023;25:e49639 [FREE Full text] [doi: 10.2196/49639] [Medline: 38019578]
5. Haidich AB. Meta-analysis in medical research. Hippokratia 2010;14:29-37 [FREE Full text] [Medline: 21487488]
6. Tawfik GM, Dila KAS, Mohamed MYF, Tam DNH, Kien ND, Ahmed AM, et al. A step by step guide for conducting a systematic review and meta-analysis with simulation data. Trop Med Health 2019;47:46 [FREE Full text] [doi: 10.1186/s41182-019-0165-6] [Medline: 31388330]
7. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-188 [FREE Full text] [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
8. Borenstein M. Comprehensive meta-analysis software. In: Systematic Reviews in Health Research. Hoboken, NJ: Wiley; 2022:535-548.
9. Dettori JR, Norvell DC, Chapman JR. Fixed-effect vs random-effects models for meta-analysis: 3 points to consider. Global Spine J 2022;12(7):1624-1626 [FREE Full text] [doi: 10.1177/21925682221110527] [Medline: 35723546]
10. Ruppar T. Meta-analysis: How to quantify and explain heterogeneity? Eur J Cardiovasc Nurs 2020;19(7):646-652. [doi: 10.1177/1474515120944014] [Medline: 32757621]
11. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
12. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006;11(2):193-206. [doi: 10.1037/1082-989X.11.2.193] [Medline: 16784338]
13. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22(4):719-748. [Medline: 13655060]
14. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Methods for estimating between-study variance and overall effect in meta-analysis of odds ratios. Res Synth Methods 2020;11(3):426-442. [doi: 10.1002/jrsm.1404] [Medline: 32112619]
15. Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y, Furukawa TA. Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? BMC Med Res Methodol 2014;14:30 [FREE Full text] [doi: 10.1186/1471-2288-14-30] [Medline: 24559167]
16. Friedrich JO, Adhikari NKJ, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. BMC Med Res Methodol 2008;8:32 [FREE Full text] [doi: 10.1186/1471-2288-8-32] [Medline: 18492289]
17. Bajorin DF, Witjes JA, Gschwend JE, Schenker M, Valderrama BP, Tomita Y, et al. Adjuvant nivolumab versus placebo in muscle-invasive urothelial carcinoma. N Engl J Med 2021;384(22):2102-2114 [FREE Full text] [doi: 10.1056/NEJMoa2034442] [Medline: 34077643]
18. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson VG, Dalle S, et al. Longer follow-up confirms recurrence-free survival benefit of adjuvant pembrolizumab in high-risk stage III melanoma: updated results from the EORTC 1325-MG/KEYNOTE-054 trial. J Clin Oncol 2020;38(33):3925-3936 [FREE Full text] [doi: 10.1200/JCO.20.02110] [Medline: 32946353]

XSL•FO
RenderX

19. Choueiri TK, Tomczak P, Park SH, Venugopal B, Ferguson T, Chang YH, et al. Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. N Engl J Med 2021;385(8):683-694 [FREE Full text] [doi: 10.1056/NEJMoa2106391] [Medline: 34407342]
20. Forde PM, Spicer J, Lu S, Provencio M, Mitsudomi T, Awad MM, et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. N Engl J Med 2022;386(21):1973-1985 [FREE Full text] [doi: 10.1056/NEJMoa2202170] [Medline: 35403841]
21. Kelly RJ, Ajani JA, Kuzdzal J, Zander T, Van Cutsem E, Piessen G, et al. Adjuvant nivolumab in resected esophageal or gastroesophageal junction cancer. N Engl J Med 2021;384(13):1191-1203. [doi: 10.1056/NEJMoa2032125] [Medline: 33789008]
22. Long GV, Luke JJ, Khattak MA, de la Cruz Merino L, Del Vecchio M, Rutkowski P, et al. Pembrolizumab versus placebo as adjuvant therapy in resected stage IIB or IIC melanoma (KEYNOTE-716): distant metastasis-free survival results of a multicentre, double-blind, randomised, phase 3 trial. Lancet Oncol 2022;23(11):1378-1388. [doi: 10.1016/S1470-2045(22)00559-9] [Medline: 36265502]
23. O'Brien M, Paz-Ares L, Marreaud S, Dafni U, Oselin K, Havel L, et al. Pembrolizumab versus placebo as adjuvant therapy for completely resected stage IB-IIIA non-small-cell lung cancer (PEARLS/KEYNOTE-091): an interim analysis of a randomised, triple-blind, phase 3 trial. Lancet Oncol 2022;23(10):1274-1286. [doi: 10.1016/S1470-2045(22)00518-6] [Medline: 36108662]
24. Rahma OE, Yothers G, Hong TS, Russell MM, You YN, Parker W, et al. Use of total neoadjuvant therapy for locally advanced rectal cancer: initial results from the pembrolizumab arm of a phase 2 randomized clinical trial. JAMA Oncol 2021;7(8):1225-1230 [FREE Full text] [doi: 10.1001/jamaoncol.2021.1683] [Medline: 34196693]
25. Schmid P, Cortes J, Dent R, Pusztai L, McArthur H, Kümmel S, et al. Event-free survival with pembrolizumab in early triple-negative breast cancer. N Engl J Med 2022;386(6):556-567. [doi: 10.1056/NEJMoa2112651] [Medline: 35139274]
26. Wakelee H, Liberman M, Kato T, Tsuboi M, Lee SH, Gao S, et al. Perioperative pembrolizumab for early-stage non-small-cell lung cancer. N Engl J Med 2023;389(6):491-503 [FREE Full text] [doi: 10.1056/NEJMoa2302983] [Medline: 37272513]
27. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322(7300):1479-1480 [FREE Full text] [doi: 10.1136/bmj.322.7300.1479] [Medline: 11408310]
28. Dettori JR, Norvell DC, Chapman JR. Seeing the forest by looking at the trees: how to interpret a meta-analysis forest plot. Global Spine J 2021;11(4):614-616 [FREE Full text] [doi: 10.1177/21925682211003889] [Medline: 33939533]
29. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;342:d549. [doi: 10.1136/bmj.d549] [Medline: 21310794]
30. Schriger DL, Altman DG, Vetter JA, Heafner T, Moher D. Forest plots in reports of systematic reviews: a cross-sectional study reviewing current practice. Int J Epidemiol 2010;39(2):421-429. [doi: 10.1093/ije/dyp370] [Medline: 20093305]
31. Afonso J, Ramirez-Campillo R, Clemente FM, Büttner FC, Andrade R. The perils of misinterpreting and misusing "Publication Bias" in meta-analyses: an education review on funnel plot-based methods. Sports Med 2024;54(2):257-269 [FREE Full text] [doi: 10.1007/s40279-023-01927-9] [Medline: 37684502]
32. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One 2012;7(7):e39471 [FREE Full text] [doi: 10.1371/journal.pone.0039471] [Medline: 22848355]
33. Sterne JAC, Harbord RM. Funnel plots in meta-analysis. Stata J 2004;4(2):127-141. [doi: 10.1177/1536867x0400400204]
34. Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JPA, Thabane L, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? Int J Epidemiol 2009;38(1):276-286. [doi: 10.1093/ije/dyn179] [Medline: 18824467]
35. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol 2008;61(1):64-75. [doi: 10.1016/j.jclinepi.2007.03.013] [Medline: 18083463]
36. Kang H. Trial sequential analysis: novel approach for meta-analysis. Anesth Pain Med 2021;16(2):138-150 [FREE Full text] [doi: 10.17085/apm.21038] [Medline: 33940767]
37. Chang Y, Phillips MR, Guymer RH, Thabane L, Bhandari M, Chaudhary V, R.E.T.I.N.A. study group. The 5 min meta-analysis: understanding how to read and interpret a forest plot. Eye 2022;36(4):673-675 [FREE Full text] [doi: 10.1038/s41433-021-01867-6] [Medline: 34987196]
38. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. Clin Epidemiol Glob Health 2019;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]
39. Fujiwara Y, Horita N, Adib E, Zhou S, Nassar AH, Asad ZUA, et al. Treatment-related adverse events, including fatal toxicities, in patients with solid tumours receiving neoadjuvant and adjuvant immune checkpoint blockade: a systematic review and meta-analysis of randomised controlled trials. Lancet Oncol 2024;25(1):62-75. [doi: 10.1016/S1470-2045(23)00524-7] [Medline: 38012893]

Abbreviations

MD: mean difference

XSL•FO
RenderX

OR: odds ratio
SMD: standardized mean difference
TSA: trial sequential analysis

Edited by X Ma; submitted 05.07.24; peer-reviewed by P Wu, H Nguyen; comments to author 14.10.24; revised version received 30.10.24; accepted 18.01.25; published 06.03.25.

Please cite as:
Fekete JT, Győrffy B
MetaAnalysisOnline.com: Web-Based Tool for the Rapid Meta-Analysis of Clinical and Epidemiological Studies
J Med Internet Res 2025;27:e64016
URL: https://www.jmir.org/2025/1/e64016
doi:10.2196/64016
PMID:39928123

©János Tibor Fekete, Balázs Győrffy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Introducing Novel Methods to Identify Fraudulent Responses (Sampling With Sisyphus): Web-Based LGBTQ2S+ Mixed-Methods Study

Kinnon Ross MacKinnon[1,2], MSW, PhD; Naail Khan[3], MA; Katherine M Newman[3], MA; Wren Ariel Gould[2], MSW; Gin Marshall[1], MES; Travis Salway[4], MSW, PhD; Annie Pullen Sansfaçon[5,6], MSW, PhD; Hannah Kia[7], MSW, PhD; June SH Lam[8,9], MD, PhD

[1]School of Social Work, York University, Toronto, ON, Canada
[2]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Department of Psychology, York University, Toronto, ON, Canada
[4]Faculty of Health Sciences, Simon Fraser University, Vancouver, BC, Canada
[5]School of Social work, University of Montreal, Montreal, QC, Canada
[6]University of Stellenbosch, Stellenbosch, South Africa
[7]School of Social Work, The University of British Columbia, Vancouver, BC, Canada
[8]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[9]Gender Identity Clinic, Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Kinnon Ross MacKinnon, MSW, PhD
School of Social Work
York University
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100
Email: kinnonmk@yorku.ca

Abstract

Background: The myth of Sisyphus teaches about resilience in the face of life challenges. Detransition after an initial gender transition is an emerging experience that requires sensitive and community-driven research. However, there are significant complexities and costs that researchers must confront to collect reliable data to better understand this phenomenon, including the lack of a uniform definition and challenges with recruitment.

Objective: This paper presents the sampling and recruitment methods of a new study on detransition-related phenomena among lesbian, gay, bisexual, transgender, queer, and 2-spirit (LGBTQ2S+) populations. It introduces a novel protocol for identifying and removing bot, scam, and ineligible responses from survey datasets and presents preliminary descriptive sociodemographic results of the sample. This analysis does not present gender-affirming health care outcomes.

Methods: To attract a large and heterogeneous sample, 3 different study flyers were created in English, French, and Spanish. Between December 1, 2023, and May 1, 2024, these flyers were distributed to >615 sexual and gender minority organizations and gender care providers in the United States and Canada, and paid advertisements totaling >CAD $7400 (US $5551) were promoted on 5 different social media platforms. Although many social media promotions were rejected or removed, the advertisements reached >7.7 million accounts. Study website visitors were directed from 35 different traffic sources, with the top 5 being Facebook (3,577,520/7,777,218, 46%), direct link (2,255,393/7,777,218, 29%), Reddit (1,011,038/7,777,218, 13%), Instagram (466,633/7,777,218, 6%), and X (formerly known as Twitter; 233,317/7,777,218, 3%). A systematic protocol was developed to identify scam, nonsense, and ineligible responses and to conduct web-based Zoom video platform screening with select participants.

Results: Of the 1377 completed survey responses, 957 (69.5%) were deemed eligible and included in the analytic dataset after applying the exclusion protocol and conducting 113 virtual screenings. The mean age of the sample was 25.87 (SD 7.77; median 24, IQR 21-29 years). A majority of the participants were White (Canadian, American, or of European descent; 748/950, 78.7%),

XSL•FO
RenderX

living in the United States (704/957, 73.6%), and assigned female at birth (754/953, 79.1%). Many participants reported having a sexual minority identity, with more than half the sample (543/955, 56.8%) indicating plurisexual orientations, such as bisexual or pansexual identities. A minority of participants (108/955, 11.3%) identified as straight or heterosexual. When asked about their gender-diverse identities after stopping or reversing gender transition, 33.2% (318/957) reported being nonbinary, 43.2% (413/957) transgender, and 40.5% (388/957) identified as detransitioned.

Conclusions: Despite challenges encountered during the study promotion and data collection phases, a heterogeneous sample of >950 eligible participants was obtained, presenting opportunities for future analyses to better understand these LGBTQ2S+ experiences. This study is among the first to introduce an innovative strategy to sample a hard-to-reach and equity-deserving group, and to present an approach to remove fraudulent responses.

(J Med Internet Res 2025;27:e63252) doi:10.2196/63252

KEYWORDS

sampling; bots; transgender; nonbinary; detransition; lesbian, gay, bisexual, and transgender; mobile phone

Introduction

Background

In the past several years, the topic of detransition has garnered significant attention in the mainstream media and academic literature as an emerging sociomedical phenomenon in need of further research and care provision [1-3]. Detransition is an umbrella term inclusive of stopping, shifting, or reversing the social, legal, and medical interventions undertaken during an initial gender transition [4-6]. Depending on the social, legal, and medical steps taken for an initial gender transition, detransition may include shifting sexual and gender minority (SGM) identities and gender expression, social detransition such as reverting to using pronouns and a gender expression aligned with the assigned-at-birth gender, legal change of name or gender marker, and the discontinuation or reversal of gender-related medical treatments [4,5,7]. Transgender and gender-diverse (TGD) people may stop taking gender-related hormonal treatments for reasons unrelated to identity shifts, and it is important to recognize that detransition is distinct from—but sometimes overlaps with—the temporary or permanent discontinuation of gender-related medical care. Depending on definitions and measurement decisions, studies conducted in the United States, Canada, the United Kingdom, Australia, and the Netherlands estimate that between 1% and 13.1% of TGD people experience detransition at some point in their lives [4,8-10]. However, due to study designs, it can be difficult to know the extent to which stopping treatments or detransition is temporary or permanent.

Researchers, gender-affirming care guidelines, and TGD scholars themselves have highlighted a need for qualitative and quantitative research to understand long-term outcomes after gender transition, including detransition and gender identity evolution [10-14]. However, several challenges can hamper efforts to produce this knowledge. Topping the list includes the stigmatization and politicization of detransition, concerns that publishing negative or unexpected results about gender transition are being weaponized against TGD populations and access to gender care, and a lack of scholarly agreement about concepts—all of which contribute to limitations in studying this issue [7].

There are also opportunities to overcome these barriers by using comprehensively designed, SGM (eg, lesbian, gay, bisexual, transgender, queer, and 2-spirit [LGBTQ2S+]) community–led research studies. Nonprobability web-based sampling and recruitment offers specific strengths in researching gender minority populations [15]. This approach to recruitment is recommended to periodically explore health and social issues affecting LGBTQ2S+ communities [16], and it has been shown to be effective alongside in-person recruitment [17]. However, there are important considerations and challenges in conducting web-based research generally; for example, web-based research has always presented issues with fraudulent or “impostor” responses to surveys, but this activity has surged since the COVID-19 pandemic [18]. Many research projects have been affected by bots and impostor responders, often motivated by financial incentive [19,20].

The integrity of web-based research faces threats on several fronts; for example, identifying and removing bot, scam, or ineligible responses from datasets is an emerging priority [21,22]. TGD and detransition research, in particular, has been subject to scam responses and sabotage efforts [23-25]. Haverkamp et al [23] identified that 50 out of 349 responses to a web-based survey designed for TGD students included slurs or hate speech. One detransition study received invalid responses to skew results [24]. While web-based survey platforms often include automatic bot detection methods such as CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) [21] and the identification of nonunique IP addresses, these alone are not effective at identifying fraudulent or malicious responses [26]. Additional techniques used to identify and deter scam responses include collecting IP addresses or geolocation data, conducting video screenings, and incorporating scam detection questions that legitimate participants would not answer affirmatively; for example, the study by Hays et al [27] included asking participants about being diagnosed with fictitious conditions. Participants who affirmed such diagnoses were deemed ineligible. Other studies have set exclusion criteria based on mining survey responses for specific nonsense or incoherent responses [25]. A technique that has been used in studies of detransition involves personal video interviews, which remove anonymity and require participants to verify demographic details such as age and location to ensure consistency with their survey responses [24]. Requiring

XSL•FO
RenderX

participants to appear on camera has also been identified as a deterrent to scam responses [24,25].

Objectives

Following from these considerations, we present a novel sampling and recruitment methodology of a study designed to generate knowledge about detransition. Its aims are four-fold: (1) to present survey design, sampling, and recruitment decisions; (2) to illustrate comprehensive study design guidance for future researchers who are studying detransition-related phenomena; (3) to highlight a protocol for identifying and excluding scam, nonsense, and ineligible survey responses; and (4) to present sociodemographic study results that demonstrate that these efforts obtained a heterogeneous SGM sample.

The Detransition Analysis, Representation, and Exploration (DARE) study was launched in December 2023 to understand sociodemographic characteristics, LGBTQ2S+ identities, life histories, minority stressors, and gender care encounters of individuals with experience of stopping, shifting, or reversing an initial gender transition (eg, detransition). This project was designed to overcome challenges in reaching people with these experiences. The study design, sampling, and study promotion decisions were all made purposefully to include individuals connected with—or disconnected from—gender care providers, organizations serving TGD and LGBTQ2S+ populations, and web-based networks for detransitioned (detrans) people. These decisions aimed to mitigate sampling bias and limitations from past research and to include hard-to-reach disparate TGD and detrans communities with the goal of building theoretical, empirical, and practice-oriented knowledge.

Exploratory Theories of Detransition Developed From Community and Clinical Samples

Diverse theories proposed to understand pathways to detransition have been developed via community and clinical samples, and they comprise internal (eg, internal change in SGM identity and medical complications), system-level (eg, health care availability and access), and external (discrimination, stigma, and lack of support for TGD identities) factors. Taken together, two primary experiences seem to occur with detransition: (1) a shift in one's internal self-conceptualization of sex, gender, and/or sexual orientation identity *after* an initial gender transition and (2) discontinuation or reversal of *prior* gender-affirming medical, legal, and social interventions [4,24,28-32].

However, existing theories should be considered preliminary, given the issues of study design and methodological limitations; for instance, disparate conceptualizations of detransition are applied in the literature, and there are issues of sampling and selection bias and potentially unreliable data collection instruments. Notable examples include analyses of medical case notes [32-34], community surveys administered to TGD populations [9], and surveys administered to web-based detrans communities [24,30,35].

Sampling and Selection Bias in Clinical and Community-Derived Samples

Prior research examining detransition has been limited by sampling strategy, often due to restrictive inclusion criteria such as reidentifying with birth-assigned gender to be counted as a detransition; for instance, studying detransition with clinical data (eg, medical case notes and recruiting from health care clinics) carries several limitations. North American gender clinics rarely have the resources to conduct long-term follow-up, and some TGD people emigrate; decline to participate in research studies; or move from pediatric to adult services, which can introduce survivorship bias [7]. Case series studies carry the risk of excluding patients who discontinue treatment, detransition, and disconnect from care or avoid care providers due to feelings of shame or fear of judgment—experiences identified in past research [28,35,36]. Short data collection periods in clinical samples can also hinder estimating detransition prevalence and understanding identity evolution that—when these outcomes do occur—often happen several years after initiating a medical transition [5,37]. Community-derived samples obtained via organizations and social media can provide useful data on individuals who are no longer connected to the same care providers or the health care system and who are otherwise considered hard to reach due to systemic marginalization.

One notable example is the 2015 United States Transgender Survey (USTS)—a large cross-sectional nonprobability community-driven study [9]. Finding that detransition among TGD adults is largely driven by external pressures such as lack of support and social stigma, Turban et al [9] analyzed a subsample of the 2015 USTS dataset—17,151 TGD adults who had ever initiated a gender transition. The authors found that 2242 (13.07%) of the 17,151 participants responded positively to the question "Have you ever de-transitioned? In other words, have you ever gone back to living as your sex assigned at birth, at least for a while?" Between the cross-sectional design of the study and the wording of this question, it is difficult to know whether respondents detransitioned temporarily, permanently, or both. The survey, developed by the National Center for Transgender Equality, favored external factors and pressure-related reasons in its predetermined list of reasons for detransition. Participants were shown 10 reasons for detransition that included external factors and pressure-related reasons (eg, "pressure from a parent" and "pressure from family members"), and they were shown only 2 nonexternal reasons (ie, "I realized gender transition was not for me" and "not listed above [please specify]"). Study promotion and recruitment efforts comprehensively targeted 800 TGD-serving, LGBTQ2S+, and allied community organizations [38]. Support groups for detransitioning were likely nonexistent at the time of data collection and were thus not involved in study promotion or recruitment. Moreover, individuals with a history of gender transition who detransitioned and no longer held a TGD identity would have been excluded per the USTS inclusion criteria [38]. A majority of the respondents were TGD adults aged ≥25 years (65.5%) and transfeminine individuals assigned male at birth (AMAB; 55%).

Similarly, cross-sectional community surveys conducted through web-based detrans networks present limitations in sampling and study-related decisions that prioritize detransitioned individuals who stopped identifying as TGD or disconnected from TGD communities. The survey conducted in 2019 [39]

XSL•FO
RenderX

by Vandenbussche [30] aimed to explore the care needs of the web-based detrans community via the Post Trans project [40] and other web-based groups largely serving female detransitioners on Facebook, Instagram, Twitter (subsequently rebranded as X), and Reddit (r/detrans). No TGD or LGBTQ2S+ organizations seem to have been involved in participant recruitment [30]. The survey recruited 237 participants (92% female or assigned female at birth [AFAB] individuals). Of this sample, 65% transitioned both socially and medically, while 31% transitioned only socially. Although the survey provided a question on gender and gender identity ("How do you see yourself now?" with the following response options: "woman," "man," "trans man," "trans woman," "female detransitioner," "male detransitioner," "non-binary," and "other"), no results are presented to discern whether any participants in the sample affirmed a TGD identity when they took the survey—participants are referred to as either "female detransitioners" or "male detransitioners."

Another survey of 100 people (69% female or AFAB individuals) who discontinued or reversed gender-related medical interventions recruited participants from largely web-based sources from December 2016 to April 2017 (4.5 mo) via Twitter, Tumblr, Reddit, and closed detransition groups [35]. This survey was also circulated on professional listserves via the American Psychological Association, the World Professional Association of Transgender Health, and SEXNET. The inclusion criteria required participants to have undergone medical transition (those with only social transition experiences were excluded). While participants in this survey had the option to indicate a current TGD identity, only 39% reported a TGD identity.

In comparison to the 2015 USTS instrument that prioritized external pressure–related reasons for detransition, the surveys administered by Vandenbussche [30] and Littman [35] favored internal, psychological, and physical health–related factors in their predetermined reasons for detransition. However, these surveys also included structural (eg, financial barriers) and externally driven factors, such as a lack of support and discrimination. Both surveys recruited a majority of female or AFAB individuals and favored individual-level driving factors motivating detransition. However, these studies were limited by small sample sizes and selection bias in that the recruitment and inclusion criteria focused primarily on individuals who understood themselves as *detransitioners* or as having been *detransitioned* at the time of data collection—terminology that can be divisive [41].

Some detransitioned people report experiencing rejection from TGD and LGBTQ2S+ communities and organizations [30,42], meaning that these individuals may be hard to reach for investigators whose networks are composed primarily of LGBTQ2S+ contacts. Therefore, sampling and recruitment that takes a broad approach is necessary to include individuals who may have become disconnected from SGM networks.

Methods

Study Population

To overcome past limitations and to understand the life experiences, minority stressors, and SGM identities of people who self-identify with experiences related to detransitioning, we conducted a binational, cross-sectional, mixed methods study in Canada and the United States. Our objective was to target a large and heterogeneous sample of 500 to 1000 participants (including detrans, TGD, and LGBTQ2S+ people). This target sample size of >500 was selected due to the anticipated heterogeneity of the group and requiring a large sample to explore and characterize detransition broadly [43]. The mixed quantitative and qualitative approach was selected to quantify different pathways (reasons for detransitioning) and to understand qualitatively the care needs specific to various pathways.

Inclusion and Exclusion Criteria

To be eligible, participants had to be aged ≥16 years; able to fill out a survey in English, Spanish, or French; and had to report a life experience of stopping, shifting, or reversing a gender transition. Gender transition was defined as inclusive of social, legal, and medical transition, and it was not a requirement to have accessed gender-related medical treatments to be eligible. We included social-only, social+medical, and medical-only transition experiences. This decision was supported by the range of inclusion criteria from past studies [9,30,41,42,44] and aligned with our aim of building knowledge about individuals who self-identify with detransition-related experiences. As has been pointed out, some individuals may detransition socially or medically and continue to affirm a TGD identity, while others may shift from a TGD identity to a detransitioner identity without ever being able to detransition socially or medically [4,30,41]. Recruitment and study promotional materials stated that the study was open to anyone with a history of shifting gender identity; stopping or reversing a gender transition; or detransitioning for any reason, such as a loss of access to gender-affirming health care, medical complications, misdiagnosis, discrimination and lack of support, or an identity change.

To be eligible, participants also had to be living in the United States or Canada during the study period, and IP addresses and geolocation data were collected. The United States was included as a geographic region of interest, in addition to Canada, due to the US political and legislative context surrounding gender-affirming health care restrictions [45], antitransgender rhetoric, and proximity to Canada [46]. As it is likely that some TGD people may be forced to detransition due to the reduced availability of gender-affirming health care, legislative restrictions, and antitransgender rhetoric, one of the DARE study flyers was designed specifically to invite participants who were forced to stop transitioning or detransition due to the loss of gender-affirming health care.

The minimum age for participation was set at 16 years for 3 reasons. First, many Canadian research ethics boards recognize this as the minimum legal age for providing independent informed consent for research studies. Second, we aimed to

XSL•FO
RenderX

examine the unique experiences of detransitioning during the teen years compared to adulthood. Third, given that the time between transition and detransition can span years, setting the minimum age at 16 years meant that we would be likely to include some individuals who began their initial transition in childhood. No maximum age limit was imposed to ensure a diverse sample with broad sociodemographic backgrounds and life experiences.

Survey Development and Measures

Reflexivity and Community Engagement

The English-language survey was developed collaboratively by the full DARE study team—a multidisciplinary group of social work and public health researchers who study the health and well-being of SGM populations. The team includes a majority of TGD (spanning the transfeminine and transmasculine spectrums) and SGM individuals and includes a gender-affirming clinician. Following the recommendations by Hildebrand-Chupp [41] on language considerations for detrans research, we developed the survey using neutral language designed to be sensitive to diverse TGD and detrans populations; for example, after detransition, some individuals feel affirmed by terms and language associated with sex rather than gender identity, and past qualitative studies with detrans young people illustrate that some feel that they do not have a gender identity [28]. To be inclusive of all respondents, the survey posed the question "What term(s) best describe your current *gender*?" rather than asking about *gender identity*. Participants were also asked a series of questions about whether, after stopping or reversing transition, they understand themselves as transgender, nonbinary, cisgender, or detransitioned (the response options were "yes," "no," and "unsure"). These questions were not mutually exclusive.

We used community-engaged research strategies. The survey was reviewed for language accuracy by an English-speaking person in the United States with experience of detransition who received a gift card valued at US $150. The survey was initially pilot-tested in English by members of the DARE study team (n=9). Subsequently, external pilot testers were recruited through personal and professional networks, including TGD and non-TGD detransitioned people, and a gender care provider—including AFAB and AMAB individuals (n=4). Pilot testers were asked to comment on the appropriateness of the language, survey completion time, the flow and organization of the survey, broken links, the coherent ordering of questions, questions that were challenging to understand, and overall comments. After pilot testing, the revised English survey was translated into French and Spanish. The translated French and Spanish surveys were reviewed and pilot-tested by native speakers. External pilot testers received gift cards worth CAD $50 (US $37) or US $50, depending on their country of residence.

Survey Development

The comprehensive survey was programmed in Qualtrics (Qualtrics International Inc) and made available through a Qualtrics platform accessible from a dedicated web page. On the basis of pilot testing, it was anticipated that participants would take 35 to 60 minutes to complete the survey. It was programmed using skip logics that were sensitive to individualized transition experiences reported by participants (those who had engaged in both a social and a medical transition were shown the greatest number of questions; those responding "no" to initiating a medical transition were not shown questions pertaining to medical or surgical interventions). The survey included extensive sociodemographic items and integrated measures from several prior health questionnaires delivered to SGM communities [47] and individuals who had detransitioned [9,35].

To ensure that the survey was community and data driven as well as inclusive of a wide range of detransition pathways, participants were presented with a Likert scale listing 21 possible reasons for stopping or reversing their initial gender transition. The Likert scale was developed from prior empirical studies with TGD and detransitioned people [9,28-31,35,48]. The survey transparently highlighted 4 different thematic categories, prompting participants to indicate to what extent a cluster of various reasons contributed to their decision (Multimedia Appendix 1). Whereas Turban et al [9] retrospectively organized reasons for detransition into "internal" and "external" reasons, the DARE survey explicitly organized reasons a priori into 4 dimensions. Participants were shown grouped factors as follows: (1) mental health or psychological reasons (eg, "My mental health did not improve while transitioning" and "My mental health was worse while transitioning"), (2) physical reasons (eg, "Satisfied with the physical results of transition" and "My physical health was worse while transitioning"), (3) external reasons (eg, "I felt discriminated against," "I did not have enough support in my life to continue transitioning," "I had trouble paying for hormones or surgery," and "Legislative bans on gender care required me to stop transitioning"), and (4) social or internal reasons (eg, "My personal definition of woman or man changed and I became more comfortable with my birth sex" and "My identity changed and I no longer felt a need for medical interventions"). Unlike prior studies, participants were asked to rate these reasons using a Likert scale (eg, "not at all," "a little," "somewhat," and "a lot") rather than categorically ("yes" or "no") to reflect the possibility that participants may cite multidimensional pathways.

Study Promotion and Participant Recruitment

Overview

To provide information about the study and direct prospective participants to the informed consent form and survey, an English-language website was created. This website featured a promotional video about the study and provided buttons linking to the English, French, and Spanish versions of the consent form and survey. The website also contained information about the study funder, the study team, project objectives, and recruitment flyers.

Study promotion materials were intentionally broad and neutral due to narratives of detransition being polarized. As the study objective was to reach heterogeneous populations, including TGD, SGM, and non-TGD detransitioned people, we aimed to be inclusive. To reach a diverse sample, 3 different tailored recruitment flyers were designed with a variety of language

XSL•FO
RenderX

choices and color schemes that were anticipated to attract people with a range of experiences (Multimedia Appendix 2). One flyer used language that focused on attracting TGD people who had experienced shifts in gender identity, and another applied TGD-coded colors (pink and blue) to attract participants who were forced to detransition due to external factors. A third flyer used colors (green and blue) and language choices anticipated to attract detrans populations who no longer identify as TGD. The term *detransition* was included on all recruitment materials, the study website, and in the informed consent form. The French and Spanish versions of these 3 flyers were also advertised widely across social media platforms and distributed to organizations and care providers in the United States and Canada.

LGBTQ2S+ Organizations and Gender Care Providers

Direct invitation emails were sent to >615 TGD or LGBTQ2S+ organizations and gender care providers, most of them serving English-speaking or multilingual populations. Efforts were made to contact organizations, gender clinics, and gender therapists who work with TGD, LGBTQ2S+, and detransitioning people. The direct emails disproportionately focused on English speakers because French and Spanish are spoken by a minority of people in North America. Hence, 21 French-speaking organizations were contacted directly by email, 5 French-speaking organizations were contacted via LinkedIn, and the flyer was shared directly with Francophone groups for parents of TGD individuals. In addition, 9 Spanish-speaking organizations were contacted. The Spanish recruitment flyers were shared on social media by another researcher studying Spanish-speaking TGD and detrans populations.

We offered to meet with 5 LGBTQ2S+ youth groups in the Greater Toronto Area, Ontario, Canada, to engage in direct study promotion and recruitment and answer questions about the study. None of these organizations responded to the offer. One LGBTQ2S+ organization in Toronto; 1 organization serving TGD populations in British Columbia, Canada; and at least 2 gender clinics in the United States and in Canada confirmed posting physical recruitment flyers in high-traffic areas to support recruitment efforts. There may have been additional physical posters publicized by organizational contacts without our awareness.

Webinar

Toward the end of the data collection phase in April 2024, we privately invited select stakeholders from LGBTQ2S+ organizations, gender clinics, and TGD and detransitioned people within our networks to a webinar presenting preliminary data to aid in further study promotion. Of the 83 people who registered for the event, 38 (46%) attended. The webinar presented the objectives and methodology of the study, preliminary results, and ended with a direct request for support with participant recruitment. The webinar was not recorded or posted on the web.

Former LGBTQ2S+ Research Participants

Participants from prior team research studies who had consented to being contacted for future research opportunities were emailed information about the study. These were largely Canadian former participants of 2 different studies related to LGBTQ2S+ affirmation (N=1181) and detransition and identity fluidity (N=27). The majority (1014/1208, 83.9%) were Anglophone and sent English recruitment materials; the rest (194/1208, 16.1%) were Francophone and sent French recruitment materials.

Paid and Unpaid Social Media Advertisements

Between December 1, 2023, and May 1, 2024, flyers in all 3 study languages were widely distributed on the web via paid and unpaid social media posts and advertisements across TikTok, Instagram, Facebook, Tumblr, X, Reddit, Discord, and Grindr (Multimedia Appendices 3-5). Where possible, we also advertised in English-speaking private groups on social networking sites for TGD and detrans people, as well as parents of TGD youth. Of note, the DARE study principal investigator maintains a TikTok account with >29,500 followers, where they discuss TGD and detrans research. To enhance engagement and promote the study using audiovisuals, several TikTok videos about the study, as well as YouTube videos, were produced and shared on the web. TikTok has a powerful algorithm that is able to reach large numbers of people and is recognized as a useful platform in community-engaged TGD and gender care knowledge sharing [49]. To estimate the reach and impact of our advertisements, at the end of the data collection period, we reviewed impressions, likes, and shares across all social media platforms used. We summed analytics data from each platform.

Excluding Scam and Ineligible Survey Responses

To enhance the reliability of the survey data, we followed guidance from prior studies [25,26,50]. Adding 2 fictitious conditions—"chekalism" and "syndomitis"—has been recommended as an effective way of identifying scam responses [27]. These 2 items were included in the survey and shown to all participants (eg, "At any point while transitioning, did you develop chekalism?" with the following response options: "yes," "no," and "unsure"). We further flagged potential scam responses based on nonsense, incoherent, or hateful email addresses, as advised by Pullen Sansfaçon et al [25] and Haverkamp et al [23].

To remove bot, scam, and ineligible responses from the dataset, a protocol was developed using guidelines and recommendations from previous studies. First, we removed any responses from IP addresses that had submitted multiple completed surveys [50], as well as those originating from outside the United States or Canada (based on IP address and geolocation data). Next, to remove any bot responses, we followed the guidelines suggested by Google and Qualtrics developers and excluded any responses with a reCAPTCHA score of <0.5 [26]. We then excluded respondents who indicated that they had neither socially nor medically transitioned (they were deemed ineligible based on the study's inclusion and exclusion criteria). Participants who indicated "yes" to having either of the 2 fictitious conditions were also excluded. In addition, we left out participants who reported no discordance between their past and current gender, as well as those whose reported sex assigned at birth conflicted with the hormone they indicated taking during their initial medical transition (eg, a female or AFAB individual who reported taking estrogen or a male or AMAB individual who

XSL•FO
RenderX

reported taking testosterone). Survey responses with a completion time of <12 minutes were also removed, given that pilot testing estimated a typical completion time of 35 to 60 minutes.

Participants identified as potentially scam or ineligible for the study who had consented to being contacted were invited via email to complete a screening interview over Zoom (Zoom Video Communications, Inc) to verify their survey responses. If the research team could validate their eligibility and the accuracy of their survey responses, their survey data were readded to the dataset. As part of the screening process, interview participants were required to turn on their cameras for identification at the beginning of the interview, following guidelines suggested by previous literature [18,25]. Each screening interview was audio recorded. The screening questions were designed to be easy for genuine participants to answer. Those interviewing potential scam participants were also asked to document additional indicators of scams, such as poor or unusual audio quality and sounds, inconsistencies in daylight relative to the participant's reported time zone, and vague and brief responses [25]. The screening interview included the following six questions:

1. Can you please confirm your age?
2. Can you please confirm the country, province (or state), and city you are currently living in?
3. Can you please tell us the country, province (or state), and city you were in when you accessed the survey?
4. Can you please confirm what terms best describe your current gender?
5. For your initial transition, did you socially transition, and if so, in which year?
6. Could you please describe the main reasons you detransitioned?

Participants who could not answer these questions or provided answers inconsistent with their survey responses were excluded. During the process of screening potential scam responses, team members had participants select an interview time via the Calendly website, which allowed the interviewer to confirm the participant's email address and time zone. This verification process was also applied to 4 gift card raffle winners and participants who were screened before the qualitative interview phase.

We removed from the analytic sample all potential scam and ineligible respondents flagged by this protocol who did not consent to being contacted by the research team, did not provide a follow-up email, or did not respond to an invitation to complete the Zoom interview. During the screening process, 2 recurring anomalies were observed: first, Calendly showed an unexpected number of sign-ups from the West African time zone; second, there were several sign-ups for screening from email addresses that we had not personally contacted. This suggests that our email invitation to the screening process had been shared with additional individuals who had never participated in the DARE survey.

Ethics Approval

The study (protocol #3964) received ethics approval from York University (e2023-298), and all participants provided written informed consent to take the survey. This consent included providing explicit agreement to publishing anonymized data for research. Most survey questions were optional, and participants could exit the survey at any time. At any point while taking the survey, participants could access a list of mental health care resources for youth and adults. After providing informed consent, participants were given the opportunity to enter a raffle for a US $50 or CAD $50 (US $37) gift card, which was also advertised in study promotional materials.

Results

Overview

Between December 1, 2023, and May 1, 2024, we widely distributed 3 different study flyers in English, French, and Spanish via direct emails and social media platforms. Paid and unpaid advertising on social media served as a primary recruitment strategy, and paid advertisements totaling CAD $7494.81 (US $5551.71) were promoted on 5 different platforms that targeted >40 unique groups and networks (Table 1; Multimedia Appendices 4 and 5).

Table 1. Paid advertisements (December 13, 2023-April 25, 2024) per social media platform (total cost: CAD $7494.81 [US $5551.71]).

Social media platforms	Cost (CAD $[a]), n (%)
Grindr (US $ converted to CAD $)	1028.60 (13.7)
Instagram and Facebook	2808.18 (37.5)
Reddit	2166.34 (28.9)
TikTok	659.40 (8.8)
Tumblr (US $ converted to CAD $)	832.29 (11.1)
X (no paid advertising accepted)	0 (0)

[a]A currency exchange rate of CAD $1=US $1.35 is applicable.

According to data provided by each social media platform, paid and unpaid study advertisements reached >7.7 million accounts (Table 2). Although efforts were made to equally promote all study posters in 3 languages, discrepancies arose because some posts seemed to receive more positive or negative attention and engagement, which likely affected engagement levels and overall promotion and views. Analytics data from TikTok indicated that flyer 3 ("forced detransition") received a higher number of shares between users and "likes," which may have amplified its views and impressions (Multimedia Appendix 3).

XSL•FO
RenderX

During the study promotion period, 30 different Reddit subreddits were targeted via paid advertisements totaling CAD $2166.34 (US $1604.69) and unpaid posts from our study team or other users (Table 1; Multimedia Appendix 4). Instagram and Facebook were also frequently reported as referral sources, with 8 private Facebook groups targeted and advertising expenses totaling CAD $2808.18. Additional study promotion took place on Tumblr and Discord. Advertisements on Grindr began in December 2023 but were discontinued after approximately 1 month due to high advertising costs and low recruitment yield (1/957, 0.1%).

Table 2. Social media impressions, views, and likes (N=7,777,218).

Social media platforms	Impressions (n=7,617,166), n (%)	Views (n=154,430), n (%)	Likes (n=5622), n (%)
Facebook	—[a]	—	43 (0.8)
Instagram	5,725,940 (75.2)	—	1447 (25.7)
Reddit (account 1)	235,772 (3.1)	—	—
Reddit (account 2: DARE[b] study)	1,352,388 (17.8)	—	—
TikTok	51,765 (0.7)	51,765 (33.5)	2041 (36.3)
Tumblr	—	—	677 (12)
Tumblr Blaze advertisements	245,690 (3.2)	—	1021 (18.2)
X	—	102,124 (66.1)	375 (6.7)
YouTube	5611 (0.1)	541 (0.4)	18 (0.3)
Discord	—	—	—

[a]Not applicable.

[b]DARE: Detransition Analysis, Representation, and Exploration.

Challenges and Controversies During Recruitment and Data Collection

The DARE study was successful in reaching diverse and disparate web-based networks and was therefore discussed, promoted, and commented on by people with various sociodemographic backgrounds who shared divergent political perspectives as well as expectations about the study itself. These activities may have driven awareness about the study and notified more eligible people. However, it may have also biased various web-based communities' perceptions (eg, negative or positive) of the study objectives; for example, a Tumblr post encouraged participation in the study but noted that, because the principal investigator's past detransition research included nonbinary people, he "has an agenda" [51]. On Reddit, posts within LGBTQ2S+ and TGD subreddits introduced the DARE study as triggering feelings of "disgust" (r/LGBT) [52] or specifically encouraged TGD people who had detransitioned or retransitioned without regret to participate because "anti-trans grifters are trying to skew results of this study" (r/MtF) [53]. Five completed surveys, primarily originating from antidetransition spam email addresses, included hate speech, and these were removed by the scam and nonsense protocol (Figure 1). In private TGD or detrans-focused groups on Facebook, members occasionally questioned the study or hesitated to post recruitment flyers. A few times, Facebook group moderators removed posts about the study; however, in most cases, after the research team clarified the study's goal of building care and support, moderators reversed the decision.

Several barriers to study promotion and data collection were encountered with paid social media advertising campaigns. Some advertising campaigns were rejected outright or removed shortly after launch by Reddit, Tumblr, and X (Figure 2). After 3 weeks of carrying paid advertising for flyers 1 and 2 via the study principal investigator's verified professional account, Reddit removed both campaigns, citing a violation of its restricted policy for promoting "health and wellness," which required special approval [54]. The Reddit advertising division was contacted several times for support in having the advertisements approved, but these emails went unanswered. Later, using a different Reddit user account, the 3 advertisements (in all 3 languages) were successfully promoted via paid advertising campaigns for approximately 4 weeks. However, all 3 posters were eventually removed again, with Reddit citing a violation of "style" policies.

XSL•FO
RenderX

Figure 1. Sankey diagram of the number of survey responses removed for each criterion.

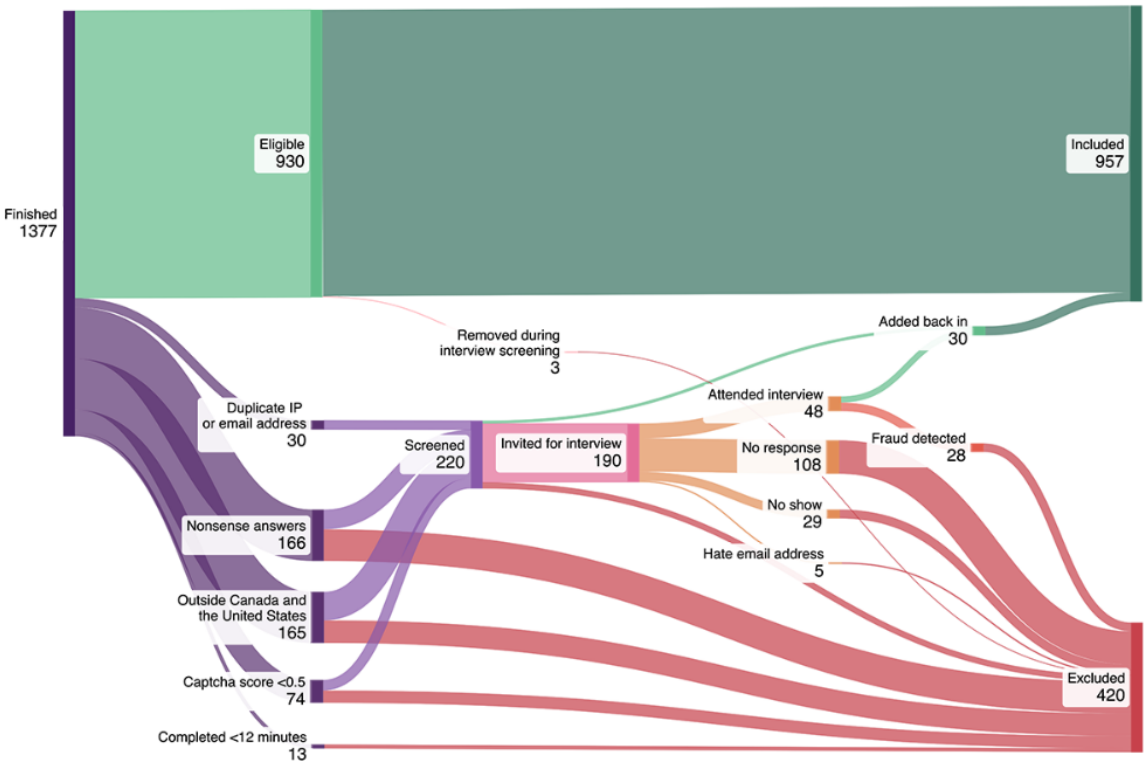

Figure 2. Rejected or removed paid advertisements.

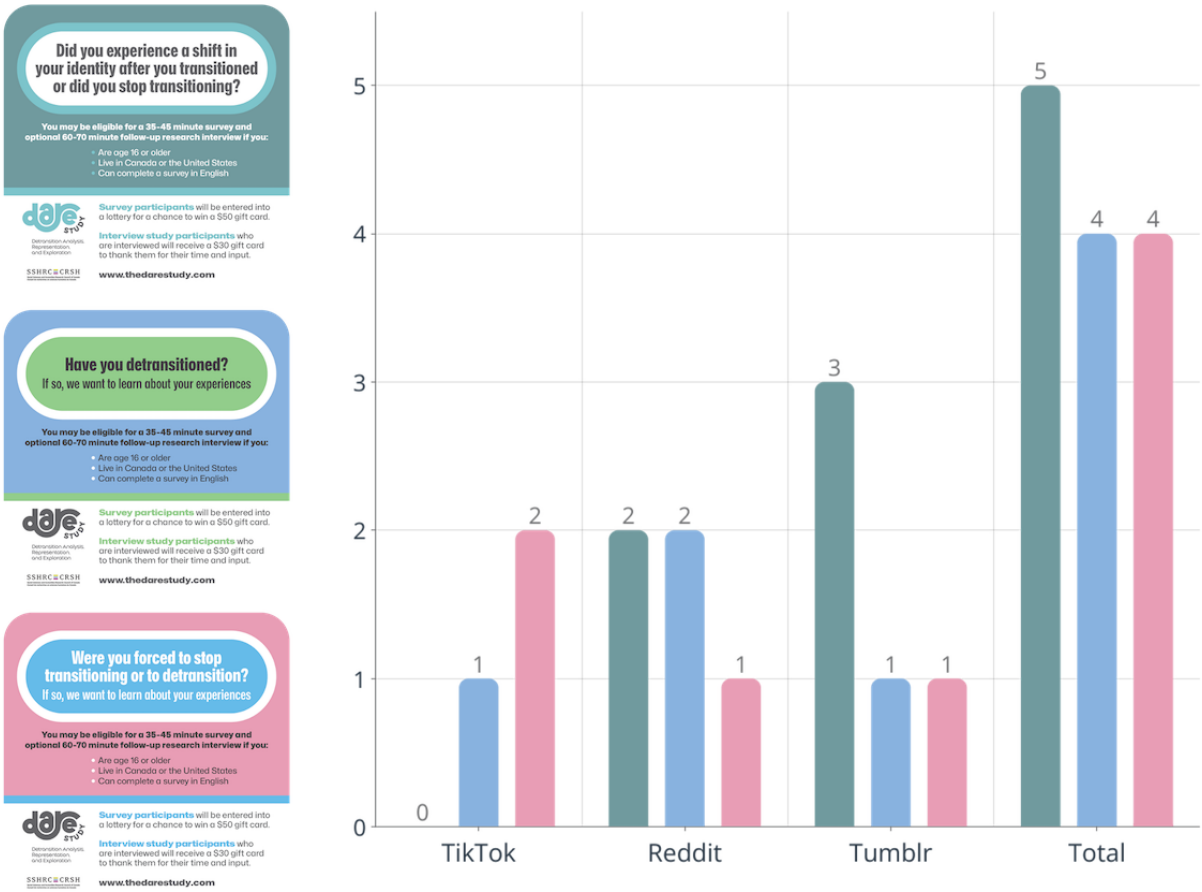

Similar issues were encountered on other platforms. Tumblr rejected several requests for paid “Blaze” advertising campaigns, while, on 2 occasions, TikTok either rejected the application for paid promotion (for flyers 2 and 3) or discontinued promoting a video after accepting payment. TikTok also temporarily suspended the principal investigator’s account after a paid promotion for the study had been purchased. Instagram and Facebook did not remove or reject any paid advertising, to

XSL•FO
RenderX

our knowledge, for any of the flyers. Despite numerous emails, X never approved any advertising campaigns or responded to efforts to purchase promotions.

According to TikTok analytics, the platform's algorithm promoted flyer 3 ("forced detransition") to a greater degree than the other 2 flyers (Multimedia Appendix 3). This may be explained by flyer 3 receiving more "likes" and shares among users, which likely contributed to higher engagement and led the algorithm to pushing the video to a wider audience, a known feature of TikTok's algorithm [49].

Website Traffic, Completed Surveys, and Participant Self-Report About Study Promotion Sources

The study website had 12,272 unique visitors from December 1, 2023, to April 30, 2024, from 78 countries. To access the website, 90.1% (11,167/12,272) of the visitors used a mobile phone, 8% (982/12,272) used a desktop computer, and 1% (123/12,272) used a tablet. Of the 12,272 visitors, 11,412 (93%) were from Canada and the United States, with a majority (n=7609, 66.68%) coming from the United States. The visitors were directed from 35 different traffic sources, with the top 5 being Facebook (3,577,520/7,777,218, 46%), direct link (2,255,393/7,777,218, 29%), Reddit (1,011,038/7,777,218, 13%), Instagram (466,633/7,777,218, 6%), and X (233,317/7,777,218, 3%). These website traffic data contrasted with participants' self-reports. For instance, survey data indicated that Tumblr (286/957, 26.7%) and Reddit (247/957, 23%) were the most frequently reported platforms where participants learned about the study. Notably, Tumblr was not a predefined survey option but was commonly selected as an "other" response. These sources were followed by Meta (198/957, 18.5%), a friend (78/957, 7%), Twitter/X (61/957, 6%), and a referral from a researcher (61/957, 6%). The remaining self-reported survey sources included TikTok, Discord, word of mouth, LGBTQ+ organizations, care providers, Grindr, and other sources (141/957, 13.2%). Despite directly emailing >1800 organizations, care providers, and former research contacts, direct email referrals did not seem to be a popular source: <7% (67/957) of the participants reported a direct email referral source (eg, from a researcher or care provider). Participants completed surveys in the following languages: English (1354/1377, 98.3%), Spanish (5/1377, 0.4%), and French (18/1377, 1%).

Of the 190 participants suspected of scam who shared their e-mail address and were invited to verify their responses via a Zoom interview, 48 (25.3%) attended. After the screenings, we excluded 28 (58%) of these 48 participants. After applying the scam screening exclusion protocol and conducting Zoom screenings, 957 (69.5%) of the 1377 completed survey responses were determined to be eligible and included in the analytic dataset. Figure 1 and Table 3 highlight this process, including the number of, and rationale for, exclusions.

Table 3. Total exclusions before Zoom screenings. The exclusion criteria were applied in the order listed in the table (n=448).

Exclusion criteria	Participants excluded, n (%)
Ineligible and nonsense answers	166 (37.1)
Survey completed outside Canada and the United States	165 (36.8)
reCAPTCHA[a] score of <0.5	74 (17)
Duplicated "identity" (duplicate email or IP address)	30 (7)
Survey completion time <12 min	13 (3)

[a]CAPTCHA: Completely Automated Public Turing Test to Tell Computers and Humans Apart.

Sociodemographics and Geographic Distribution of the Analytic Sample

After exclusions, 957 eligible participants were included in the analytic dataset. The sample was homogenous in terms of race and sex assigned at birth. A large majority identified as White (Canadian, American, or of European descent; Table 4), and a large majority were AFAB (754/953, 79.1%). Out of 957 participants, 4 (0.4%) skipped the question on sex recorded on birth certificate.

Participants were diverse in terms of age distribution and SGM identities (eg, TGD, detransitioned, cisgender, bisexual, lesbian, gay, heterosexual, and asexual; refer to Figure 3 and Table 5 for a complete breakdown of SGM status). As shown in Figure 4, the mean age of the sample was 25.87 (SD 7.77; range 16-74) years.

A majority of the participants were living in the United States (704/957, 73.6%). While the geographic distribution of the sample was fairly representative, with participants from most of the US states and Canadian provinces (Figure 5), there were a few provinces and states with no participants, such as Canada's northern territories (Yukon, Northwest Territories, and Nunavut) and Prince Edward Island, as well as Alaska, North Dakota, Wyoming, and Mississippi in the United States.

XSL•FO
RenderX

Table 4. Self-identified racial and ethnic backgrounds. Of the 957 participants, 7 (0.7%) skipped this question. Participants could choose multiple options.

Racial and ethnic backgrounds	Participants (n=950), n (%)
Black African (eg, Ghana, Kenya, or Somalia)	12 (1.3)
Black Canadian or African American	33 (3.5)
Black Caribbean (eg, Jamaica or Haiti)	13 (1.4)
Central Asian	1 (0.1)
East Asian (eg, China, Japan, Korea, or Taiwan)	36 (3.7)
Indigenous (eg, First Nations, Metis, Inuit, or American Indian)	76 (7.9)
Indo-Caribbean (eg, Guyanese with origins in India)	2 (0.2)
Jewish	98 (10.2)
Latin American (eg, Argentina, Mexico, or Nicaragua)	70 (7.3)
Middle Eastern (eg, Egypt, Iran, Israel, or Saudi Arabia)	26 (2.7)
Pacific Islander or Polynesian	2 (0.2)
South Asian (eg, India, Sri Lanka, or Pakistan)	4 (0.4)
Southeast Asian (eg, Vietnam, Malaysia, or the Philippines)	14 (1.5)
Romany	1 (0.1)
White Canadian or White American	748 (78.2)
White European (eg, United Kingdom, Greece, Sweden, or Russia)	194 (20.3)

Figure 3. Self-identification as cisgender, nonbinary, transgender, and detransitioned. Participants could select “yes” to >1 option.

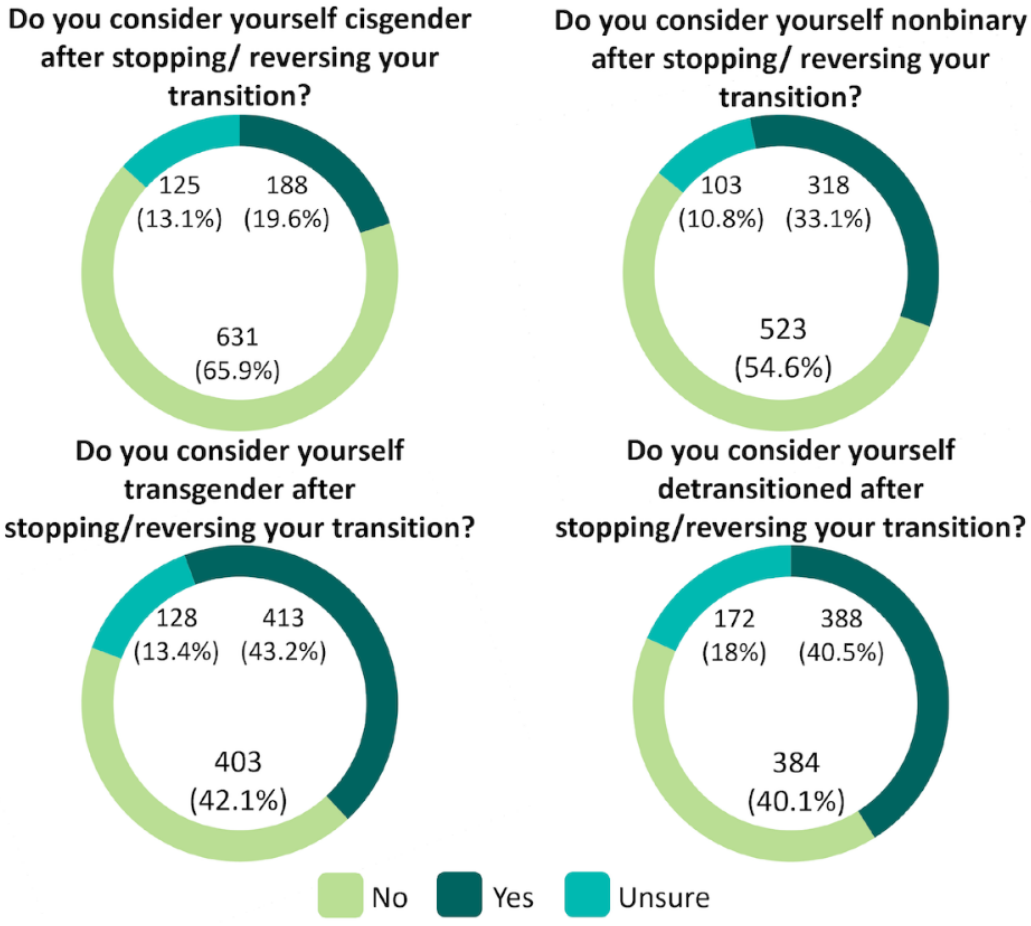

XSL•FO
RenderX

Figure 4. Age distribution of the analytic sample (n=957; mean 25.87, SD 7.77 y; median 24 y).

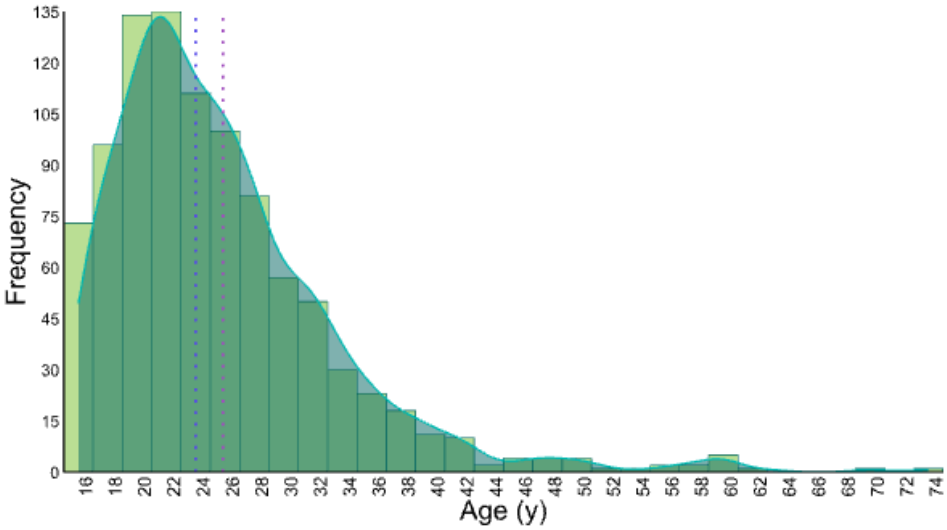

Table 5. Sexual orientation identity. Of the 957 participants, 2 (0.2%) skipped this question. Participants could choose multiple options.

Sexual orientation identity	Participants (n=955), n (%)
Bisexual	429 (44.8)
Queer	277 (28.9)
Lesbian or homosexual	254 (26.5)
Pansexual	114 (11.9)
Straight or heterosexual	108 (11.3)
Asexual	101 (10.6)
Gay or homosexual	94 (9.8)
Not sure or questioning	73 (7.6)
2-spirit	8 (0.9)

Figure 5. Geographic distribution of the analytic sample. (A) Pie diagram illustrating the proportions and number of responses from participants living in Canada and the United States. (B) Number of responses from each province in Canada. (C) Number of responses from each state in the United States (1 participant from Washington, DC, was included in the count for Virginia, and 1 participant was living in Puerto Rico [not included in the image]).

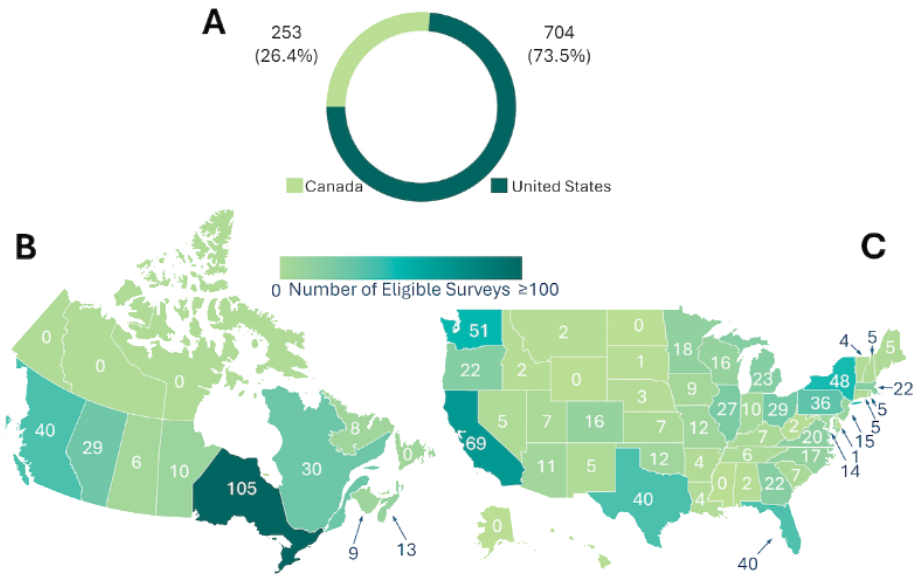

XSL•FO
RenderX

Discussion

Summary

The DARE study attracted a large and diverse US and Canadian sample of 957 LGBTQ2S+ individuals with experience of stopping, shifting, or reversing an initial gender transition. To obtain this nonprobability sample, >1800 organizations, care providers, and former research participants living in the United States or Canada were individually emailed information about the study. A major advertising and recruitment strategy involved paid study promotions (featuring 3 flyers in 3 different languages) totaling >CAD $7400 (US $5551) on popular social media platforms and social networking sites. We reached >7.7 million accounts. These efforts attracted 12,272 unique visitors to our study website, with 1377 (11.22%) completing the survey. Compared to study promotions on other platforms, those on Instagram and Facebook encountered fewer barriers to advertising and were among the top referral sources to the study website, a finding consistent with a prior SGM web-based survey [16]. In light of these results, we discuss and reflect on the Sisyphean efforts to sampling and recruitment procedures for this community-driven study, as well as the strengths and limitations of our novel approach.

Study Promotion and Recruitment With LGBTQ2S+ and Detrans Communities

The DARE study encountered both support and challenges in study promotion and data collection, as reflected by public posts on Reddit and Tumblr as well as difficulties in placing advertisements. Given the prominence of antitransgender rhetoric and the politicization of detransition by right-wing actors [55], SGM people who have detransitioned may be reluctant to participate in studies designed to examine their experiences due to concerns about nefarious actors. This may also explain some of the discussion about the study in the r/LGBT and r/MtF subreddits, as well as several malicious antidetransition responses received. Given this, despite a large majority of our participants identifying as SGM, many LGBTQ2S+ organizations and gender clinics might have felt unable to show outward support for this research [56]. This could explain why only a small minority of participants (15/957, ~2%) reported these referral sources.

As well, some detransitioned people disconnect from LGBTQ2S+ communities or gender clinics, a phenomenon identified in previous research [36,42]. People with fluid gender identities are also more likely to avoid health care than gender-consistent TGD individuals [57]. These challenges were proactively anticipated when designing the study, which informed the decision to use social media advertising and promote multiple study flyers—particularly flyer 2 (“stopping transition or identity shift”) and flyer 3 (“forced detransition”). Although it is possible that even including the term “detransition” in our study title created apprehension, SGM researchers should note that entirely avoiding using “detransition” in favor of alternative, euphemistic terminology risks alienating some people who relate to this language [7]. Due to polarization, particularly within both TGD and detrans community networks, there can be mistrust of researchers. Language choices and advertising designs that prioritize one detransition experience over another may inherently create selection bias. Using multiple, relatively neutral flyers tailored to include disparate perspectives and experiences is one solution that seemed to attract heterogeneous identities.

Bot, Scam, and Nonsense Detection and Exclusion Criteria: Lessons for Researchers

Potentially amplified by the US $50 or CAD $50 (US $37) gift card raffle incentive, the study attracted many scam as well as ineligible and nonsense survey responses. Given that more than a quarter of the completed surveys (420/1377, 30.5%) were deemed ineligible, our results in identifying and excluding scam and nonsense survey responses offer important lessons on internet-based sampling. Early in the data collection phase, we realized that many of the completed survey responses seemed suspicious; therefore, we created a protocol to identify and remove these responses from the dataset (refer to the Methods section). Making these adjustments required a protocol amendment and approval from the research ethics board. The scam screening process, including email communications and virtual video interviews, was both time and resource intensive. Despite these additional time and resource costs, the Zoom screening process enabled us to confirm that, of the 220 responses flagged as suspicious, 30 (13.7%) were valid and could be readded to the dataset. An additional 179 participants who agreed to be contacted for qualitative interviews for the second phase of the study were also invited to complete verification screening. These participants were not flagged as potential scam or fraud and after completing 65 preinterview Zoom screenings, only 3 scam cases were identified. Lower-resourced projects may decide to exclude all suspicious responses, given the relatively low odds of identifying eligible respondents.

Sample Sociodemographics and Geographic Distribution

The final sample was heterogeneous in terms of LGBTQ2S+ identities, age (including adolescents, young adults, and adults aged >35 y), geographic distribution across US states and Canadian provinces, and other sociodemographic variables. Interestingly, ~11% (105/957) of the participants were living in Ontario, Canada. While Ontario is the most populous province in Canada, with 15.6 million residents [58]—of whom >39,000 are TGD individuals aged >15 years [59]—this is still disproportionate representation in comparison to many US states (eg, California).

There are a few possible explanations for this finding. First, much of the in-person study promotion and posting of physical flyers occurred at Ontario-based gender clinics and LGBTQ2S+ support spaces. In addition, a majority of the study team members (5/7, 71%) have extensive LGBTQ2S+ professional and social networks in Ontario. Another interpretation is that social media promotion was largely based out of Toronto and southern Ontario, which may have led algorithms to prioritize sharing content and advertisements locally and through our team’s existing social media networks (eg, TikTok and Instagram). A third explanation is that, in comparison to other US and Canadian regions, a Toronto-based child gender clinic

XSL•FO
RenderX

was among the first in the world to begin referring children and adolescents for gender-affirming medical interventions in the early 2000s [60]. It is plausible that some of those individuals may have detransitioned as adults and participated in the study. Regions with wider access to gender-affirming health care may also see greater distribution of these experiences, whereas areas with a hostile climate toward gender diversity and poor access to gender-affirming health care may see fewer instances of detransition.

Limitations

This study took efforts to sample a large and heterogeneous group in terms of transition-related experiences and pathways to detransition, LGBTQ2S+ identities, assigned sex and gender, race and ethnicity, age, geographic location, and other sociodemographic factors. Evenly distributing the 3 flyers via paid advertisements proved difficult, which may have created selection bias favoring some experiences over others. There were 13 total rejections and removals of the 3 flyers and disparities in views, likes, and engagement. These issues may have arisen because the study advertisements were reported by other users; for instance, TikTok is known to "shadow ban" or censor some LGBTQ2S+ content [49] and sexuality-related topics [61]. X, TikTok, and Reddit posed more substantial barriers to placing paid advertisements than Tumblr, Instagram, and Facebook. On Reddit, the study was advertised via paid or unpaid methods across 30 different subreddits, including those focused on TGD, LGBTQ2S+, detrans, and general topics. The relative success with web-based recruitment channels versus organizational and care provider referrals may have introduced a bias favoring younger participants.

Furthermore, most of the participants were White (Canadian, American, or of European descent) and AFAB (754/953, 79.1%), and were living in the United States (704/957, 73.4%). While this indicates homogeneity in these particular demographic characteristics, these demographics are also roughly consistent with the general TGD adolescent and young adult population diagnosed with gender dysphoria and who have socially or medically transitioned in North America [62,63]. Per Canadian gender clinic and community samples, approximately 80% of TGD adolescents and young adults are estimated to be transmasculine and AFAB, and a large majority are White (75%-78%) [64,65]. A representative US sample of TGD people showed that the increase in the prevalence of TGD identities observed between 2014 and 2022 was disproportionately driven by White, AFAB, young adults [66]. Still, there may be important distinctions between AMAB individuals and racialized LGBTQ2S+ people, which future analyses with this dataset will examine.

While the sample had a wide age range (16-74 y), the mean age of the sample was 25.87 (SD 7.77) years. This may be partly because a majority of the sample (811/957, 84.7%) reported learning about the study through social media and because the survey was only available on the web. In addition, those without access to a computer, smartphone, or tablet would have been unable to take the survey, which may have excluded some older individuals. Another Canadian and US study that examined TGD gender-affirming medical treatment discontinuation found that those who stopped or reversed treatment had a mean age of 22 years [65]. Overall, while many studies on detransition show samples with a majority of participants aged ≤29 years [7], greater efforts to include older adults are warranted.

To identify and remove ineligible and scam responses, we developed a protocol. We encourage other researchers to review and build from this approach because recent surges in bot and scam activity pose a threat to web-based research. We recognize that this approach may have risked excluding individuals who were indeed eligible but who misread questions or incorrectly filled out parts of the survey (eg, by accidentally indicating "yes" to a fictitious disorder question).

Conclusions

Several new studies examining the discontinuation of gender-affirming medical treatments, detransition, and gender identity fluidity have been published recently, and this study extends this body of literature [8,67-70]. However, there is still no mutually agreed-upon or uniform conceptualization for detransition. This, together with the antitransgender politicization of these experiences and the prevalence of bot and scam responses in web-based research, presented many challenges to sampling and recruitment. Persistence in light of these challenges during the study promotion and data collection phases of our study culminated in a Sisyphean LGBTQ2S+ community–engaged project that successfully obtained a large and diverse SGM sample. As writing about detrans people now appears in major publications such as *Nature* [6], future analyses of this dataset can support the development of robust and empirically rich characterization and conceptualization of this understudied SGM experience.

Acknowledgments

The Detransition Analysis, Representation, and Exploration (DARE) study research team appreciates all survey participants for their time and engagement. This study was funded by an Insight Grant (*#435-2023-1099)* provided by the Social Sciences and Humanities Research Council of Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The 21-item Likert scale (pathways to detransition).

XSL•FO
RenderX

[PDF File (Adobe PDF File), 71 KB - jmir_v27i1e63252_app1.pdf]

Multimedia Appendix 2
Study flyers in English, French, and Spanish.
[DOCX File , 905 KB - jmir_v27i1e63252_app2.docx]

Multimedia Appendix 3
TikTok advertisements (breakdown).
[PDF File (Adobe PDF File), 143 KB - jmir_v27i1e63252_app3.pdf]

Multimedia Appendix 4
Reddit paid and unpaid advertising (subreddits).
[PDF File (Adobe PDF File), 66 KB - jmir_v27i1e63252_app4.pdf]

Multimedia Appendix 5
Facebook groups (unpaid advertising).
[PDF File (Adobe PDF File), 114 KB - jmir_v27i1e63252_app5.pdf]

References

1. Brierley J, Larcher V, Hadjipanayis AA, Grossman Z. European Academy of Paediatrics statement on the clinical management of children and adolescents with gender dysphoria. Front Pediatr 2024;12:1298884 [FREE Full text] [doi: 10.3389/fped.2024.1298884] [Medline: 38374879]
2. Campbell D, Gentleman A. Adult transgender clinics in England face inquiry into patient care. The Guardian. 2024 Apr 10. URL: https://www.theguardian.com/society/2024/apr/10/adult-transgender-clinics-in-england-face-inquiry-into-patient-care [accessed 2025-03-06]
3. de Vries AL, Richards C, Tishelman AC, Motmans J, Hannema SE, Green J, et al. Bell v Tavistock and Portman NHS Foundation Trust (2020) EWHC 3274: weighing current knowledge and uncertainties in decisions about gender-related treatment for transgender adolescents. Int J Transgend Health 2021 Apr 05;22(3):217-224 [FREE Full text] [doi: 10.1080/26895269.2021.1904330] [Medline: 34240066]
4. Expósito-Campos P, Salaberria K, Pérez-Fernández JI, Gómez-Gil E. Gender detransition: a critical review of the literature. Actas Esp Psiquiatr 2023 May;51(3):98-118 [FREE Full text] [Medline: 37489555]
5. MacKinnon KR, Gould WA, Enxuga G, Kia H, Abramovich A, Lam JS, et al. Exploring the gender care experiences and perspectives of individuals who discontinued their transition or detransitioned in Canada. PLoS One 2023 Nov 29;18(11):e0293868 [FREE Full text] [doi: 10.1371/journal.pone.0293868] [Medline: 38019738]
6. Ashley F, Brightly-Brown S, Rider GN. Beyond the trans/cis binary: introducing new terms will enrich gender research. Nature 2024 Jun;630(8016):293-295. [doi: 10.1038/d41586-024-01719-9] [Medline: 38858484]
7. MacKinnon KR, Expósito-Campos P, Gould WA. Detransition needs further understanding, not controversy. BMJ 2023 Jun 14;381:e073584 [FREE Full text] [doi: 10.1136/bmj-2022-073584] [Medline: 37315956]
8. Cavve BS, Bickendorf X, Ball J, Saunders LA, Thomas CS, Strauss P, et al. Reidentification with birth-registered sex in a Western Australian pediatric gender clinic cohort. JAMA Pediatr 2024 May 01;178(5):446-453. [doi: 10.1001/jamapediatrics.2024.0077] [Medline: 38436975]
9. Turban JL, Loo SS, Almazan AN, Keuroghlian AS. Factors leading to "detransition" among transgender and gender diverse people in the United States: a mixed-methods analysis. LGBT Health 2021 Jun 01;8(4):273-280. [doi: 10.1089/lgbt.2020.0437]
10. Olson KR, Raber GF, Gallagher NM. Levels of satisfaction and regret with gender-affirming medical care in adolescence. JAMA Pediatr 2024 Dec 01;178(12):1354-1361. [doi: 10.1001/jamapediatrics.2024.4527] [Medline: 39432272]
11. Final report. The Cass Review. URL: https://cass.independent-review.uk/home/publications/final-report/ [accessed 2024-07-09]
12. Coleman E, Radix AE, Bouman WP, Brown GR, de Vries AL, Deutsch MB, et al. Standards of care for the health of transgender and gender diverse people, version 8. Int J Transgend Health 2022 Sep 06;23(Suppl 1):S1-259 [FREE Full text] [doi: 10.1080/26895269.2022.2100644] [Medline: 36238954]
13. de Vries AL, Hannema SE. Growing evidence and remaining questions in adolescent transgender care. N Engl J Med 2023 Jan 19;388(3):275-277. [doi: 10.1056/NEJMe2216191] [Medline: 36652360]
14. Veale JF, Deutsch MB, Devor AH, Kuper LE, Motmans J, Radix AE, et al. Setting a research agenda in trans health: an expert assessment of priorities and issues by trans and nonbinary researchers. Int J Transgend Health 2022 Mar 11;23(4):392-408 [FREE Full text] [doi: 10.1080/26895269.2022.2044425] [Medline: 36324879]
15. Turban JL, Almazan AN, Reisner SL, Keuroghlian AS. The importance of non-probability samples in minority health research: lessons learned from studies of transgender and gender diverse mental health. Transgend Health 2023 Jul 28;8(4):302-306. [doi: 10.1089/trgh.2021.0132] [Medline: 37525831]

XSL•FO
RenderX

16. Delgado-Ron JA, Jeyabalan T, Watt S, Black S, Gumprich M, Salway T. Sampling sexual and gender minority youth with UnACoRN (Understanding Affirming Communities, Relationships, and Networks): lessons from a web-based survey. J Med Internet Res 2023 Jan 12;25:e44175 [FREE Full text] [doi: 10.2196/44175] [Medline: 36633900]
17. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. J Med Internet Res 2018 Jun 18;20(6):e197 [FREE Full text] [doi: 10.2196/jmir.9461] [Medline: 29914861]
18. Santinele Martino A, Perrotta A, McGillion BJ. Who can you trust these days?: dealing with imposter participants during online recruitment and data collection. Qual Res 2024 Jan 22;24(5):1291-1301. [doi: 10.1177/14687941231224591]
19. Chandler JJ, Paolacci G. Lie for a dime: when most prescreening responses are honest but most study participants are impostors. Soc Psychol Pers Sci 2017 Apr 27;8(5):500-508. [doi: 10.1177/1948550617698203]
20. Hanna P. Using internet technologies (such as Skype) as a research medium: a research note. Qual Res 2012 Apr 05;12(2):239-242. [doi: 10.1177/1468794111426607]
21. Loebenberg G, Oldham M, Brown J, Dinu L, Michie S, Field M, et al. Bot or not? Detecting and managing participant deception when conducting digital research remotely: case study of a randomized controlled trial. J Med Internet Res 2023 Sep 14;25:e46523 [FREE Full text] [doi: 10.2196/46523] [Medline: 37707943]
22. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020 Oct 07;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
23. Haverkamp A, Johnson F, Michelle KD, Bothwell, Qwo-Li D, Montfort D. Attack helicopters and white supremacy: interpreting malicious responses to an online questionnaire about transgender undergraduate engineering and computer science student experiences. Bull Appl Transgend Stud 2023 Jun 30;2(1-2):67-94. [doi: 10.57814/qd1y-9b22]
24. Littman L, O'Malley S, Kerschner H, Bailey JM. Detransition and desistance among previously trans-identified young adults. Arch Sex Behav 2024 Jan;53(1):57-76 [FREE Full text] [doi: 10.1007/s10508-023-02716-1] [Medline: 38038854]
25. Pullen Sansfaçon A, Gravel E, Gelly MA. Dealing with scam in online qualitative research: strategies and ethical considerations. Int J Qual Methods 2024 Jan 09;23(3). [doi: 10.1177/16094069231224610]
26. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022;56(4):2841-2852 [FREE Full text] [doi: 10.1007/s11135-021-01252-1] [Medline: 34629553]
27. Hays RD, Qureshi N, Herman PM, Rodriguez A, Kapteyn A, Edelen MO. Effects of excluding those who report having "Syndomitis" or "Chekalism" on data quality: longitudinal health survey of a sample from Amazon's Mechanical Turk. J Med Internet Res 2023 Aug 04;25:e46421 [FREE Full text] [doi: 10.2196/46421] [Medline: 37540543]
28. Pullen Sansfaçon A, Gelly MA, Gravel R, Medico D, Baril A, Susset F, et al. A nuanced look into youth journeys of gender transition and detransition. Infant Child Dev 2023 Jan 17;32(2):e2402. [doi: 10.1002/icd.2402]
29. Narayan SK, Hontscharuk R, Danker S, Guerriero J, Carter A, Blasdel G, et al. Guiding the conversation-types of regret after gender-affirming surgery and their associated etiologies. Ann Transl Med 2021 Apr;9(7):605 [FREE Full text] [doi: 10.21037/atm-20-6204] [Medline: 33987303]
30. Vandenbussche E. Detransition-related needs and support: a cross-sectional online survey. J Homosex 2022 Jul 29;69(9):1602-1620 [FREE Full text] [doi: 10.1080/00918369.2021.1919479] [Medline: 33929297]
31. MacKinnon KR, Kia H, Gould WA, Ross LE, Abramovich A, Enxuga G, et al. A typology of pathways to detransition: considerations for care practice with transgender and gender diverse people who stop or reverse their gender transition. Psychol Sex Orientat Gend Divers 2023. [doi: 10.1037/sgd0000678]
32. Pazos Guerra M, Gómez Balaguer M, Gomes Porras M, Hurtado Murillo F, Solá Izquierdo E, Morillas Ariño C. Transsexuality: transitions, detransitions, and regrets in Spain. Endocrinología Diabetes Nutr (Engl Ed) 2020 Nov;67(9):562-567. [doi: 10.1016/j.endien.2020.03.005]
33. Hall R, Mitchell L, Sachdeva J. Access to care and frequency of detransition among a cohort discharged by a UK national adult gender identity clinic: retrospective case-note review. BJPsych Open 2021 Oct 01;7(6):e184 [FREE Full text] [doi: 10.1192/bjo.2021.1022] [Medline: 34593070]
34. Boyd I, Hackett T, Bewley S. Care of transgender patients: a general practice quality improvement approach. Healthcare (Basel) 2022 Jan 07;10(1):121 [FREE Full text] [doi: 10.3390/healthcare10010121] [Medline: 35052285]
35. Littman L. Individuals treated for gender dysphoria with medical and/or surgical transition who subsequently detransitioned: a survey of 100 detransitioners. Arch Sex Behav 2021 Nov 19;50(8):3353-3369 [FREE Full text] [doi: 10.1007/s10508-021-02163-w] [Medline: 34665380]
36. MacKinnon KR, Kia H, Salway T, Ashley F, Lacombe-Duncan A, Abramovich A, et al. Health care experiences of patients discontinuing or reversing prior gender-affirming treatments. JAMA Netw Open 2022 Jul 01;5(7):e2224717 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.24717] [Medline: 35877120]
37. Dhejne C, Öberg K, Arver S, Landén M. An analysis of all applications for sex reassignment surgery in Sweden, 1960-2010: prevalence, incidence, and regrets. Arch Sex Behav 2014 Nov 29;43(8):1535-1545. [doi: 10.1007/s10508-014-0300-8] [Medline: 24872188]

XSL•FO
RenderX

38. James SE, Herman JL, Rankin S, Keisling M, Mottet L, Anafi M. The report of the 2015 U.S. transgender survey. National Center for Transgender Equality. 2006. URL: https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf [accessed 2024-06-12]
39. Survey on detransition-related needs and support. Post Trans. URL: https://post-trans.com/Survey-on-Detransition-Related-Needs-and-Support [accessed 2024-06-12]
40. Post trans - detransition stories. Post Trans. URL: https://post-trans.com/ [accessed 2024-06-12]
41. Hildebrand-Chupp R. More than 'canaries in the gender coal mine': a transfeminist approach to research on detransition. Sociol Rev 2020 Aug 10;68(4):800-816. [doi: 10.1177/0038026120934694]
42. MacKinnon KR, Gould WA, Ashley F, Enxuga G, Kia H, Ross LE. (De)transphobia: examining the socio-politically driven gender minority stressors experienced by people who detransitioned. Bull Appl Transgend Stud 2022;1(3):235-259. [doi: 10.57814/8nd4-6a89]
43. Sinha P, Calfee CS, Delucchi KL. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. Crit Care Med 2021 Jan 01;49(1):e63-e79 [FREE Full text] [doi: 10.1097/CCM.0000000000004710] [Medline: 33165028]
44. Butler G, Adu-Gyamfi K, Clarkson K, El Khairi R, Kleczewski S, Roberts A, et al. Discharge outcome analysis of 1089 transgender young people referred to paediatric endocrine clinics in England 2008-2021. Arch Dis Child 2022 Oct 19;107(11):1018-1022. [doi: 10.1136/archdischild-2022-324302] [Medline: 35851291]
45. Bans on best practice medical care for transgender youth. Movement Advancement Project. URL: https://www.lgbtmap.org/equality-maps/healthcare_youth_medical_care_bans [accessed 2024-06-12]
46. Dryden J, Lee J. How Alberta's proposed trans youth rules fit into a polarized international landscape. CBC News. 2024 Mar 11. URL: https://www.cbc.ca/news/canada/calgary/alberta-trans-youth-rules-international-landscape-1.7135324 [accessed 2024-06-12]
47. Scheim AI, Coleman T, Lachowsky N, Bauer GR. Health care access among transgender and nonbinary people in Canada, 2019: a cross-sectional survey. CMAJ Open 2021 Dec 21;9(4):E1213-E1222 [FREE Full text] [doi: 10.9778/cmajo.20210061] [Medline: 34933879]
48. Kuiper AJ, Cohen-Kettenis PT. Gender role reversal among postoperative transsexuals. Int J Transgend 1998;23(2) [FREE Full text]
49. MacKinnon KR, Kia H, Lacombe-Duncan A. Examining TikTok's potential for community-engaged digital knowledge mobilization with equity-seeking groups. J Med Internet Res 2021 Dec 09;23(12):e30315 [FREE Full text] [doi: 10.2196/30315] [Medline: 34889739]
50. Miner MH, Bockting WO, Romine RS, Raman S. Conducting internet research with the transgender population: reaching broad samples and collecting valid data. Soc Sci Comput Rev 2012 May 01;30(2):202-211 [FREE Full text] [doi: 10.1177/0894439311404795] [Medline: 24031157]
51. Patriarchy is a momentary aberration. Tumblr. URL: https://terra-feminarum.tumblr.com/post/737541844099072000/i-recommend-people-to-take-part-in-this-study-we [accessed 2024-06-12]
52. I feel disgusted by this. Reddit. URL: https://www.reddit.com/r/lgbt/comments/18lci8u/i_feel_disgusted_by_this/?rdt=51701 [accessed 2024-06-12]
53. Anti-trans grifters are trying to skew the results of this study on detransition in the US and CA. If you detransitioned unwillingly, retransitioned after detransitioning, and/or detransitioned but don't regret transitioning, you are eligible to take it too and make the results more accurate. Reddit. URL: https://www.reddit.com/r/MtF/comments/18nt276/antitrans_grifters_are_trying_to_skew_the_results/ [accessed 2024-06-12]
54. Reddit ads policy: restricted ads. Reddit. URL: https://business.reddithelp.com/helpcenter/s/article/Reddit-Advertising-Policy-Restricted-Advertisements#N9 [accessed 2024-06-12]
55. Astor M. How a few stories of regret fuel the push to restrict gender transition care. The New York Times. 2023 May 16. URL: https://www.nytimes.com/2023/05/16/us/politics/transgender-care-detransitioners.html [accessed 2025-02-20]
56. MacKinnon KR, Sansfaçon AP, Expósito-Campos P. Detransition and gender fluidity: deeper understanding can improve care and acceptance. The Conversation. 2024 Mar 12. URL: https://theconversation.com/detransition-and-gender-fluidity-deeper-understanding-can-improve-care-and-acceptance-225543 [accessed 2025-02-20]
57. Ocasio MA, Isabel Fernández M. The association between gender identity fluidity and health outcomes in transgender and gender diverse people in New Orleans and Los Angeles, USA. Prev Med Rep 2024 Jun;42:102735 [FREE Full text] [doi: 10.1016/j.pmedr.2024.102735] [Medline: 38689888]
58. Population estimates for Ontario, Canada from 2000 to 2023. Statista. URL: https://www.statista.com/statistics/569874/population-estimates-ontario-canada/ [accessed 2024-06-12]
59. Canada's most populous province, Ontario, is home to the highest number of transgender and non-binary people. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/220427/mc-b001-eng.htm [accessed 2024-06-12]
60. Sadjadi S. The vulnerable child protection act and transgender children's health. TSQ 2020;7(3):508-516 [FREE Full text] [doi: 10.1215/23289252-8553202]

XSL•FO
RenderX

61. Tillewein H, Mohon-Doyle K, Cox D. A critical discourse analysis of sexual violence survivors and censorship on the social media platform TikTok. Arch Sex Behav 2024 Nov 24;53(10):3785-3794. [doi: 10.1007/s10508-024-02987-2] [Medline: 39317917]
62. Kahn NF, Sequeira GM, Asante PG, Kidd KM, Coker TR, Christakis DA, et al. Estimating transgender and gender-diverse youth populations in health systems and survey data. Pediatrics 2024 Jun 01;153(6):e2023065197. [doi: 10.1542/peds.2023-065197] [Medline: 38752289]
63. Kahn NF, Sequeira GM, Garrison MM, Orlich F, Christakis DA, Aye T, et al. Co-occurring autism spectrum disorder and gender dysphoria in adolescents. Pediatrics 2023 Aug 01;152(2):e2023061363. [doi: 10.1542/peds.2023-061363] [Medline: 37395084]
64. Bauer GR, Pacaud D, Couch R, Metzger DL, Gale L, Gotovac S, et al. Transgender youth referred to clinics for gender-affirming medical care in Canada. Pediatrics 2021 Nov;148(5):e2020047266. [doi: 10.1542/peds.2020-047266] [Medline: 34620727]
65. MacKinnon KR, Jeyabalan T, Strang JF, Delgado-Ron JA, Lam JS, Gould WA, et al. Discontinuation of gender-affirming medical treatments: prevalence and associated features in a nonprobabilistic sample of transgender and gender-diverse adolescents and young adults in Canada and the United States. J Adolesc Health 2024 Oct;75(4):569-577 [FREE Full text] [doi: 10.1016/j.jadohealth.2024.05.015] [Medline: 38944803]
66. Twenge JM, Wells BE, Le J, Rider GN. Increases in self-identifying as transgender among US adults, 2014–2022. Sex Res Soc Policy 2024 Jul 13. [doi: 10.1007/s13178-024-01001-7]
67. Nieder TO. The intersectionality of detransitioning and being nonbinary in transgender care. J Sex Med 2023 May 27;20(6):712-715. [doi: 10.1093/jsxmed/qdad045] [Medline: 37255300]
68. Real AG, Lobato MI, Russell ST. Trajectories of gender identity and depressive symptoms in youths. JAMA Netw Open 2024 May 01;7(5):e2411322 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.11322] [Medline: 38776085]
69. van der Loos MA, Klink DT, Hannema SE, Bruinsma S, Steensma TD, Kreukels BP, et al. Children and adolescents in the Amsterdam Cohort of Gender Dysphoria: trends in diagnostic- and treatment trajectories during the first 20 years of the Dutch Protocol. J Sex Med 2023 Feb 27;20(3):398-409. [doi: 10.1093/jsxmed/qdac029] [Medline: 36763938]
70. Roberts CM, Klein DA, Adirim TA, Schvey NA, Hisle-Gorman E. Continuation of gender-affirming hormones among transgender adolescents and adults. J Clin Endocrinol Metab 2022 Aug 18;107(9):e3937-e3943. [doi: 10.1210/clinem/dgac251] [Medline: 35452119]

Abbreviations

AFAB: assigned female at birth
AMAB: assigned male at birth
CAPTCHA: Completely Automated Public Turing Test to Tell Computers and Humans Apart
DARE: Detransition Analysis, Representation, and Exploration
LGBTQ2S+: lesbian, gay, bisexual, transgender, queer, and 2-spirit
SGM: sexual and gender minority
TGD: transgender and gender-diverse
USTS: United States Transgender Survey

Edited by SR Ramos; submitted 14.06.24; peer-reviewed by L Littman, CC Kaufman, C Streed; comments to author 19.09.24; revised version received 25.10.24; accepted 26.01.25; published 17.03.25.

Please cite as:
MacKinnon KR, Khan N, Newman KM, Gould WA, Marshall G, Salway T, Pullen Sansfaçon A, Kia H, Lam JSH
Introducing Novel Methods to Identify Fraudulent Responses (Sampling With Sisyphus): Web-Based LGBTQ2S+ Mixed-Methods Study
J Med Internet Res 2025;27:e63252
URL: https://www.jmir.org/2025/1/e63252
doi:10.2196/63252
PMID:

©Kinnon Ross MacKinnon, Naail Khan, Katherine M Newman, Wren Ariel Gould, Gin Marshall, Travis Salway, Annie Pullen Sansfaçon, Hannah Kia, June SH Lam. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

What's Going On With Me and How Can I Better Manage My Health? The Potential of GPT-4 to Transform Discharge Letters Into Patient-Centered Letters to Enhance Patient Safety: Prospective, Exploratory Study

Felix Eisinger[1], MD; Friederike Holderried[2], MD; Moritz Mahling[3], MD; Christian Stegemann–Philipps[2], MSc; Anne Herrmann–Werner[2], MD; Eric Nazarenus[2], BSc; Alessandra Sonanini[2], MD; Martina Guthoff[1], MD; Carsten Eickhoff[4], BS, MS, PhD; Martin Holderried[3], MD

[1]Department of Diabetology, Endocrinology, Nephrology, University of Tübingen, Tübingen, Germany

[2]Tübingen Institute for Medical Education, University of Tübingen, Tübingen, Germany

[3]Department of Medical Strategy, Process and Quality Management, University Hospital Tübingen, Tübingen, Germany

[4]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany

Corresponding Author:
Friederike Holderried, MD
Tübingen Institute for Medical Education
University of Tübingen
Elfriede-Aulhorn-Str. 10
Tübingen, 72076
Germany
Phone: 49 1704848650
Email: Friederike.Holderried@med.uni-tuebingen.de

Abstract

Background: For hospitalized patients, the discharge letter serves as a crucial source of medical information, outlining important discharge instructions and health management tasks. However, these letters are often written in professional jargon, making them difficult for patients with limited medical knowledge to understand. Large language models, such as GPT, have the potential to transform these discharge summaries into patient-friendly letters, improving accessibility and understanding.

Objective: This study aims to use GPT-4 to convert discharge letters into more readable patient-centered letters. We evaluated how effectively and comprehensively GPT-4 identified and transferred patient safety–relevant information from the discharge letters to the transformed patient letters.

Methods: Three discharge letters were created based on common medical conditions, containing 72 patient safety–relevant pieces of information, referred to as "learning objectives." GPT-4 was prompted to transform these discharge letters into patient-centered letters. The resulting patient letters were analyzed for medical accuracy, patient centricity, and the ability to identify and translate the learning objectives. Bloom's taxonomy was applied to analyze and categorize the learning objectives.

Results: GPT-4 addressed the majority (56/72, 78%) of the learning objectives from the discharge letters. However, 11 of the 72 (15%) learning objectives were not included in the majority of the patient-centered letters. A qualitative analysis based on Bloom's taxonomy revealed that learning objectives in the "Understand" category (9/11) were more frequently omitted than those in the "Remember" category (2/11). Most of the missing learning objectives were related to the content field of "prevention of complications." By contrast, learning objectives regarding "lifestyle" and "organizational" aspects were addressed more frequently. Medical errors were found in a small proportion of sentences (31/787, 3.9%). In terms of patient centricity, the patient-centered letters demonstrated better readability than the discharge letters. Compared with discharge letters, they included fewer medical terms (132/860, 15.3%, vs 165/273, 60/4%), fewer abbreviations (43/860, 5%, vs 49/273, 17.9%), and more explanations of medical terms (121/131, 92.4%, vs 0/165, 0%).

Conclusions: Our study demonstrates that GPT-4 has the potential to transform discharge letters into more patient-centered communication. While the readability and patient centricity of the transformed letters are well-established, they do not fully

XSL•FO
RenderX

address all patient safety–relevant information, resulting in the omission of key aspects. Further optimization of prompt engineering may help address this issue and improve the completeness of the transformation.

(J Med Internet Res 2025;27:e67143) doi:10.2196/67143

KEYWORDS

GPT-4; patient letters; health care communication; artificial intelligence; patient safety; patient education

Introduction

Ensuring patient safety is fundamental in health care. A key aspect of patient safety is adherence to treatments and interventions prescribed by medical providers, as it is essential for preventing long-term disease progression, reducing complications, and improving quality of life [1]. However, in clinical practice, nonadherence remains a widespread problem. According to the World Health Organization (WHO), only 50% of patients with chronic diseases in developed countries adhere to their prescribed therapy regimens [2]. A common factor driving nonadherence is patients' lack of understanding of their disease and the underlying principles of therapy [3]. Historically, the patient-health care worker relationship followed a "paternalistic" model, where the patient was a "passive spectator in their own healing process" [4]. Fortunately, this dynamic has changed in recent years. Nonetheless, empowering patients to comprehensively understand their individual health issues remains a promising approach to improving adherence and, consequently, promoting patient safety [3].

For hospitalized patients, the discharge letter is an important source of medical information that complements the conversation with the physician. It also plays a vital role in facilitating communication between hospital doctors and other health care providers, such as primary care physicians [5]. Effective communication during the transition from inpatient to outpatient care is essential for ensuring patient safety [6]. Forster et al [7] found that 59% of preventable adverse events after hospital discharge were attributed to poor communication between hospital caregivers and either the patient or the primary care physician. After discharge, patients with chronic conditions, in particular, face numerous self-management challenges, such as adhering to prescribed medication regimens, maintaining a specific diet, and engaging in physical activity. Failure to adhere to these aspects can result in serious health consequences.

However, in clinical practice, discharge letters are typically addressed to general practitioners or other medical professionals and are often laden with professional jargon, making them inaccessible to patients with limited medical knowledge [8]. Developing patient-centered discharge letters with improved readability has been shown to enhance patient understanding [9,10]. However, in the hospital setting, time for preparing additional, individualized patient letters is often limited, as a significant portion of working hours is devoted to nonpatient-related tasks and documentation [11].

In this context, advances in artificial intelligence (AI) offer a promising approach to providing personalized and scalable support for helping patients understand medical information. Various studies have demonstrated the substantial medical knowledge of large language models (LLMs), such as GPTs [12-15]. Zaretsky et al [16] utilized LLMs to translate discharge summaries into patient-friendly language, addressing common readability metrics and the Patient Education Materials Assessment Tool (PEMAT) scoring. However, they encountered significant limitations in both accuracy and completeness [16]. By contrast, the use of readability scores has been a topic of controversy in the literature [17]. In our own work, we demonstrated GPT-4's ability to answer psychosomatic medicine examination questions [18]. A qualitative analysis of incorrect answers, based on Bloom's revised taxonomy [19,20], revealed that errors varied depending on the cognitive level. It remains unclear whether this effect is also observed in patient letters generated by GPT-4. Additionally, the extent to which GPT-4 addresses comprehensive patient safety–relevant information—a key aspect of discharge letters—has yet to be clarified.

In this study, we further investigated a related topic. We used GPT-4 to transform discharge letters into accessible, patient-centered letters and evaluated its ability to identify and incorporate patient safety–relevant information. To pinpoint potential errors or gaps in the AI-driven transformation process, patient safety–relevant information was categorized and analyzed based on Bloom's revised taxonomy (Remember, Understand) [19,20].

In summary, our study addresses the following questions:

Major objective:

- How comprehensively does GPT-4 identify and transform patient safety–relevant information, as measured by the learning objectives, from discharge letters into patient-centered letters?

Minor objectives:

- How do GPT-4–generated patient-centered letters perform in terms of medical correctness, measured by medical accuracy, case-specific relevance, and the sources of information used in the constituent sentences?
- How well do GPT-4–generated patient-centered letters perform in terms of patient centricity compared with discharge letters, with patient-centered language measured by standard readability scores, word and sentence count, and the use of medical jargon, explanations, abbreviations, repetitions, and direct addressing?

Methods

Study Outline

As the primary goal of patient-centered letters is to convey important information to patients in alignment with specific didactic "learning objectives," we defined learning objectives

XSL•FO
RenderX

for 3 common medical conditions, including the corresponding competence levels based on Bloom's taxonomy. Using these learning objectives and associated medical conditions, we created 3 discharge letters (Figure 1 and Multimedia Appendix 1). We then prompted GPT-4 to generate a patient-centered letter for each discharge letter (Multimedia Appendix 1). To account for variability in GPT-4's output, we repeated the generation of each patient-centered letter 5 times using the same prompt.

The resulting patient-centered letters were analyzed by a team of 2 experienced clinicians (FE and FH) in terms of medical quality, patient centricity, and their potential to convey safety-relevant medical information. The discharge letters, the GPT-4 prompt, and the processing of the prompts by GPT-4 to create the patient-centered letters were all conducted in German. DeepL Translate (DeepL SE) was used to translate the discharge letters, patient letters, and the prompts (Multimedia Appendix 1).

Figure 1. Comprehensive overview of the study's structure and progression. Prompting steps are illustrated in the green box. AI: artificial intelligence.

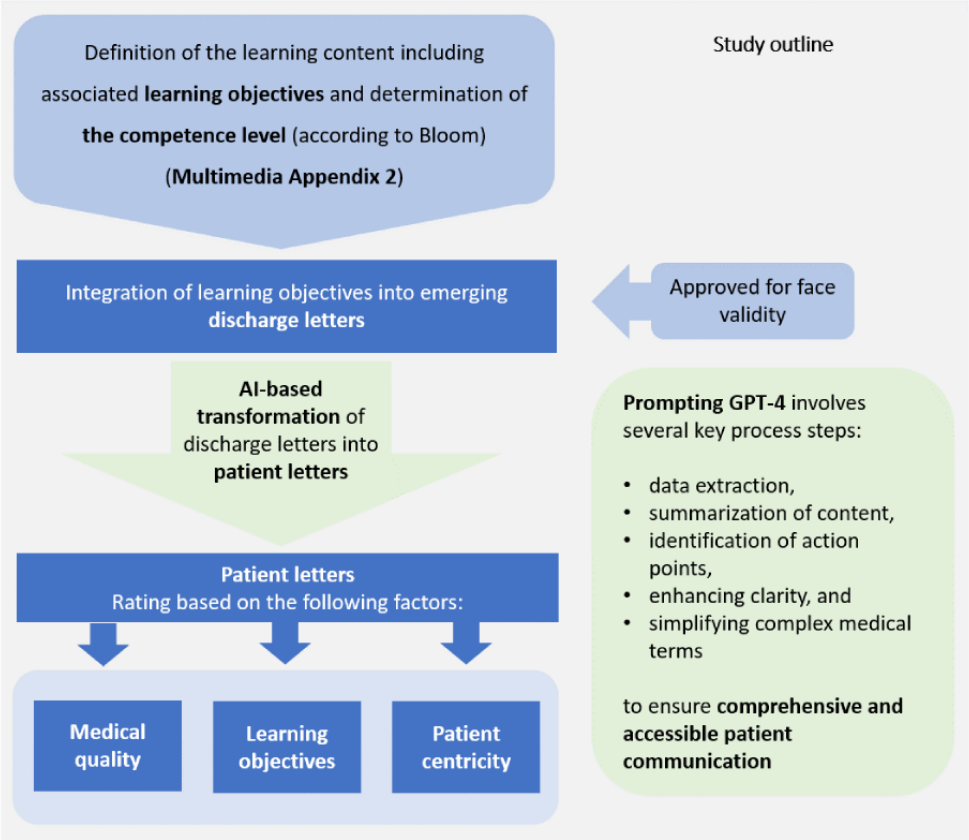

Development of Discharge Letters

For the development of the discharge letters, we meticulously structured a multistep process (Multimedia Appendix 2) to cluster critical patient information, which GPT-4 was prompted to extract from the discharge letters at various structural levels. These levels enabled us to assess GPT-4's effectiveness in retrieving critical patient information during the evaluation process based on multiple criteria. This approach provided a clearer depiction of GPT-4's competence in this specific task area. To demonstrate the transferability of our findings, it was essential to establish the extent to which GPT-4's ability to identify patient-important information was context-independent. Accordingly, we developed 3 distinct scenarios, each with different disease profiles and care settings.

Stepwise approach:

- Step 1: We selected 3 common diseases based on their high prevalence—arterial hypertension (AHT), type 2 diabetes mellitus (DM), and diabetic kidney disease (DKD).
- Step 2: The setting for each scenario was chosen, encompassing either outpatient consultations with diagnostics or inpatient treatment for initial diagnosis or follow-up care.
- Step 3: To construct relevant educational content for each scenario, we aligned all learning content with the cognitive process dimension of factual knowledge in Bloom's revised taxonomy [20].
- Step 4: The educational content was then categorized into 4 distinct fields: (1) organizational aspects, (2) medication protocols, (3) prevention of complications, and (4) disease management and lifestyle changes.
- Step 5: For each piece of learning content, we developed 2 corresponding learning objectives (Multimedia Appendix 1): one at the Remember level (eg, the patient knows they need to conduct a 24-hour urine collection test), and another at the Understand level (eg, the patient understands that the 24-hour urine collection is necessary to rule out hormonal causes of hypertension). Each discharge letter thus comprised 12 pieces of learning content, distributed across

XSL•FO
RenderX

the 4 content fields, resulting in 24 learning objectives per letter.

- Step 6: These learning objectives were then systematically integrated into the format of a typical discharge letter. The discharge letters were developed by 2 experienced physicians (FE and FH) and validated for face validity by 2 additional board-certified internal medicine and nephrology specialists (MM and MG).

Prompt Development and Creation of Patient-Centered Letters

To the best of our knowledge, there is currently no ideal model for the structure of patient-centered letters. Our group focused first on expert-formulated learning objectives based on content deemed important from a medical perspective, and second on patient-centered language. This approach provided us with a basic letter structure to aim for.

We utilized GPT-4 (gpt-4-0613, accessed December 2023) to generate the patient-centered letters, maintaining the model's default parameters and setting the temperature to the default value of 1.0. The original prompt can be found in Multimedia Appendix 1. The AI-generated patient-centered letter development process was divided into multiple stages (Textbox 1; Multimedia Appendix 3), leveraging methodologies involving LLMs as agents.

Further refinement involved adversarial prompting to identify and simplify complex medical terms and jargon, tailoring the content to a comprehension level equivalent to an eighth-grade reading ability [21]. The result of this process was a restructured and more comprehensible summary of actionable points, designed to effectively communicate essential medical information to patients.

Textbox 1. Stages in the artificial intelligence–generated patient-centered letter development process.

1. Stage 1

- Essential descriptions from the original discharge letters were transformed into concise, comprehensible summaries, referred to as "structured info," using a specialized system prompt for data extraction.

2. Stage 2

- A general summary of the original letter was created to provide an overview of the information, using a straightforward system prompt without specific constraints to minimize redundancy.

3. Stage 3

- This final stage focused on extracting detailed "action points" critical for patient understanding and compliance.
- This multistep process, known as "prompt chaining," began with segmenting the original letter to isolate key instructions and reasons for necessary behavioral changes.
- Each segment was processed to reduce redundancy and enhance clarity, using structured prompts to request information in a predefined JSON format.

Ethical Considerations

This study was approved by the local ethics committee (approval number 778/2023BO2) of the University Hospital of Tübingen. As the discharge letters were prepared by a team of medical experts based on general clinical cases and without direct reference to specific patients, informed consent was not required. As the discharge letters do not contain any individual patient data that could be traced back to specific persons, anonymization was also unnecessary. Furthermore, no compensation payments are expected due to the chosen methodological approach.

Analysis of Patient-Centered and Discharge Letters

Rating Process

Two experienced clinicians (FE and FH) independently rated the discharge letters and the patient-centered letters. For the rating process, each letter was divided into sentences. A sentence was defined as a unit of 1 or more words ending with a period, colon, or a new paragraph (typically, but not exclusively, representing complete sentences). Specific titles in the patient-centered letters, such as "Main Diagnosis," were predefined using established terms and excluded from the rating.

Each sentence was assessed individually by both raters. To ensure a standardized rating process, a general rating structure was defined in advance, using an ordinal scale from 0 (not fulfilled) to 2 (fully fulfilled; Tables 1 and 2). In cases of uncertainty, the raters discussed the issues until a consensus was reached. The results are presented as overall comparisons (patient-centered letters vs discharge letters) and as a subgroup analysis (subgroups: letters on AHT, DM, and DKD).

XSL•FO
RenderX

Table 1. Schematic rating scale (learning objectives).[a]

Content field	Common disease, Bloom's category, learning objective, and rating examples		
Medication	**Type 2 diabetes mellitus**		
	Remember	Take new medication atorvastatin 20 mg in the morning	2: "One tablet (20mg) atorvastatin in the morning (...)" 1: "Atorvastatin: Take one tablet every morning."
	Understand	Atorvastatin helps to control lipid levels.	2: "Atorvastatin: (...) can help keep cholesterol low and protect your heart." 1: "You have also been given medication to lower high cholesterol, which is important in reducing the risk of heart disease."
Prevention of complications	**Diabetic nephropathy**		
	Remember	Avoid NSAIDs[b] such as ibuprofen	2: "Ibuprofen: You should avoid this painkiller (…)." 1: "It is important for you to avoid anti-inflammatory pain medications."
	Understand	Avoiding NSAIDs is important to prevent kidney damage	2: "Ibuprofen: (…) as it can impair your kidneys." 1: "You had been taking painkillers regularly, which may have also strained your kidneys."

[a]This is an illustration of the rating scale used to analyze the learning objectives. It includes rating examples of learning objectives that were either fully (2) or partially (1) rendered in the patient-centered letters.

[b]NSAID: nonsteroidal anti-inflammatory drug.

XSL•FO
RenderX

Table 2. Rating scale (medical quality and patient centricity). The table describes the rating scale used to evaluate the patient-centered letters for medical quality and patient centricity. Examples from the patient-centered letters or discharge letters are presented within quotation marks.

Rating scales	Definition	Example from patient letters or discharge letters
Medical correctness		
Medically correct	The sentence is medically correct.	"It is important that you reduce your alcohol consumption, as alcohol can affect blood sugar levels and increase the risk of other health problems."
Medically incorrect	The sentence is medically not correct (normal HbA_{1c}[a] is below 5.7%).	"The HbA_{1c} value, which tells me how my blood sugar has been over the last few weeks, is 14.1%, which is also very high (normal below 6%)."
Case-specific relevance		
Very relevant	Information is related to the primary diagnosis.	"During your stay in hospital, you were diagnosed with type 2 diabetes for the first time, which means that your blood glucose levels are too high."
Rather relevant	Information is related to the secondary diagnosis.	"You had an appendectomy in 1972."
Neither/nor relevant	Not attributable to a primary or secondary diagnosis.	"In plain language, your hospital stay can be summarized as follows:"
Source of information		
Discharge letter	Information is derived from the discharge letter.	"You have chronic kidney disease caused by diabetes and high blood pressure."
Not from discharge letter	Information was added by GPT-4.	"They performed a kidney biopsy in which they obtained a small sample of tissue from your kidney to examine it under a microscope."
Medical terms		
Use of the medical term	The medical term is used in the sentence.	"We ask for the completion of a 24h urine collection for metanephrine to investigate a pheochromocytoma."
No special term	No special term is used in the sentence.	"You should also try to reduce your weight, eat more healthily (lots of fruit and vegetables, less fatty dairy products) and use less salt."
Explanations		
Medical term explained	The medical term is explained in everyday language.	"Sleep apnea syndrome is a condition in which breathing stops during sleep, which can also affect blood pressure."
Medical terms not explained	Medical term is not explained.	"In addition, sleep apnea syndrome should be ruled out in the outpatient setting."
Abbreviations		
Use of abbreviations	An abbreviation is used in the sentence.	"CVRF[b]:"
No use of abbreviations	An abbreviation is not used in the sentence.	"This is called hypertension."
Repetitions		
Use of repetitions	Repetition of a fact in the sentence.	"It is also recommended that you regularly measure your blood pressure at home (…) Measure your blood pressure regularly at home (…)"
No use of repetitions	No repetition of a fact in the sentence.	"Drink about 2 liters of water a day unless your doctor says otherwise."
Addressing		
Directly addressed	The patient is directly addressed.	"You should take one tablet every morning."
Not directly addressed	The patient is not directly addressed.	"We therefore recommend initiation of atorvastatin as indicated above."
Not applicable	Not applicable because there is no person to be addressed in the sentence.	"An ultrasound of the pancreas showed no evidence of a malignancy."

[a]HbA_{1c}: glycated hemoglobin.

[b]CVRF: Cardiovascular Risk Factors.

XSL•FO
RenderX

Learning Objectives

The GPT-4–generated patient-centered letters were rated based on how many of the learning objectives were identified and addressed. To better understand GPT-4's ability to highlight patient safety–relevant information, a descriptive analysis of the learning objectives was performed. This analysis determined whether the learning objectives were fully addressed, partially addressed, or missed entirely (Table 1).

Indicators of Patient Centricity

All letters were rated for readability using the Flesch Reading Ease test and the Swedish Läsbarhetsindex (LIX) and the Readability Index (RIX). The Flesch Reading Ease score ranges from 0 (very difficult to read) to 100 (very easy to read) [22]. The Swedish LIX typically ranges from 20 (very easy to read) to 60 (very difficult to read) [23]. The RIX is a modified version of the LIX, with a higher number indicating a more complex text [24]. Because of the fragmented nature of some discharge letter sentences (Multimedia Appendix 1), only the sections on the procedure and the medical discharge summary were used to calculate the readability score. By contrast, patient-centered letters, written in complete sentences, were evaluated as a whole for readability. To identify potential barriers for nonmedical readers, the use of medical terms, explanations of medical terms, abbreviations, and repetitions were analyzed for each sentence (Multimedia Appendix 4). Additionally, it was assessed whether the recipient of the letter (patient-centered letter: patient; discharge letter: general practitioner) was directly addressed in each sentence, where applicable.

Medical Correctness, Case-Specific Relevance, and Source of Information

All sentences were rated for medical correctness (Multimedia Appendix 4). To gain deeper insights into medical errors in the patient letters, a qualitative analysis of the incorrect sentences was conducted using the Braun-Clarke inductive approach [25]. Sentences were also evaluated for their case-specific relevance: sentences referring to the primary diagnosis were rated as highly relevant, those related to secondary diagnoses were considered moderately relevant, and all other sentences were classified as neither relevant nor irrelevant, as no irrelevant information was found in the letters. Additionally, it was assessed whether the medical information in the patient letters originated from the discharge letter or was generated by GPT-4.

Statistical Analysis

Statistical analysis and figure generation were conducted using R statistical software (version 4.3.1; R Foundation for Statistical Computing). Data were presented as total counts (n) and percentages, or, when not normally distributed, as medians with interquartile ranges (25th and 75th percentiles). Decimal numbers were rounded to whole numbers for clarity.

Results

Quantitative Analysis of Letter Structure

For the discharge letters, the number of sentences per letter and the median word count per sentence showed minimal variation across the 3 disease entities (Table 3).

For the patient-centered letters, GPT-4 generated a total of 952 sentences, with 860 sentences remaining after excluding those predefined in the prompts. The median word count per sentence was 15 (IQR 8-20) words. The number of sentences per patient-centered letter varied slightly between the medical conditions of AHT, DM, and DKD. The highest number of sentences was recorded in the discharge and patient-centered letters for DKD, while the fewest sentences were found in the patient-centered letters on DM.

Table 3. Quantitative analysis of letter structure.

Analysis	Overall	Arterial hypertension	Diabetes mellitus	Diabetic kidney disease
Number of sentences				
Discharge letter, n	273	83	92	98
Patient letter, n	860	278	271	311
Sentences per letter				
Discharge letter, n	92	83	92	98
Patient letter, median (IQR)	56 (54-59)	55 (52-57)	54 (53-58)	58 (56-65)
Words per sentence, median (IQR)				
Discharge letter	8 (3-13)	9 (4-14)	8 (3-13)	7 (3-13)
Patient letter	15 (8-20)	15 (9-21)	14 (9-21)	14 (8-19)

Learning Objectives in GPT-4–Generated Patient-Centered Letters

A detailed illustration of the learning objectives addressed in the patient-centered letters is provided in Figure 2. Out of the 72 learning objectives (24 per medical condition across 3 conditions), 57 were identified and addressed in the majority (≥3/5 letters per disease). However, no patient-centered letter included all the learning objectives present in the corresponding discharge letter. There was no significant difference in the coverage of learning objectives across the 3 diseases (AHT, DM, and DKD).

Figure 3 illustrates the learning objectives categorized according to Bloom's taxonomy and content field. Notably, the figure reveals that learning objectives classified under Bloom's "Remember" category were more frequently addressed in the

XSL•FO
RenderX

patient-centered letters compared with those in the "Understand" category.

Some learning objectives were only partially addressed in the majority (≥3 of 5 letters) of the patient-centered letters (5/72). Examples of these partially addressed learning objectives are provided in Multimedia Appendix 5. The missing information in these cases could be categorized into different areas (eg, responsibility, frequency, dosage). However, no systematic categorical errors were identified in the analysis of the partially addressed learning objectives.

Of the 72 learning objectives, 11 were completely omitted in the majority (≥3 of 5 letters) of the patient-centered letters (Multimedia Appendix 4). Of these, only 2 belonged to the Remember category, while the majority were under the Understand category. Furthermore, most of the missing learning objectives (6/11) were related to the field of "prevention of complications." By contrast, learning objectives related to "lifestyle" and "organizational" aspects were more frequently included in the patient-centered letters by GPT-4.

Figure 2. (A-C) Illustration of the learning objectives described in the discharge letters and their representation in the patient-centered letters. The colors indicate the extent to which each learning objective has been addressed in the patient-centered letters (dark blue: fully listed; light blue: partially listed; and yellow: not listed). The learning objectives are described on the y-axis and the number of patient letters per disease on the x-axis. GLP1: glucagon-like peptide-1; NSAID: nonsteroidal anti-inflammatory drug; SGLT2: sodium-glucose cotransporter 2;.

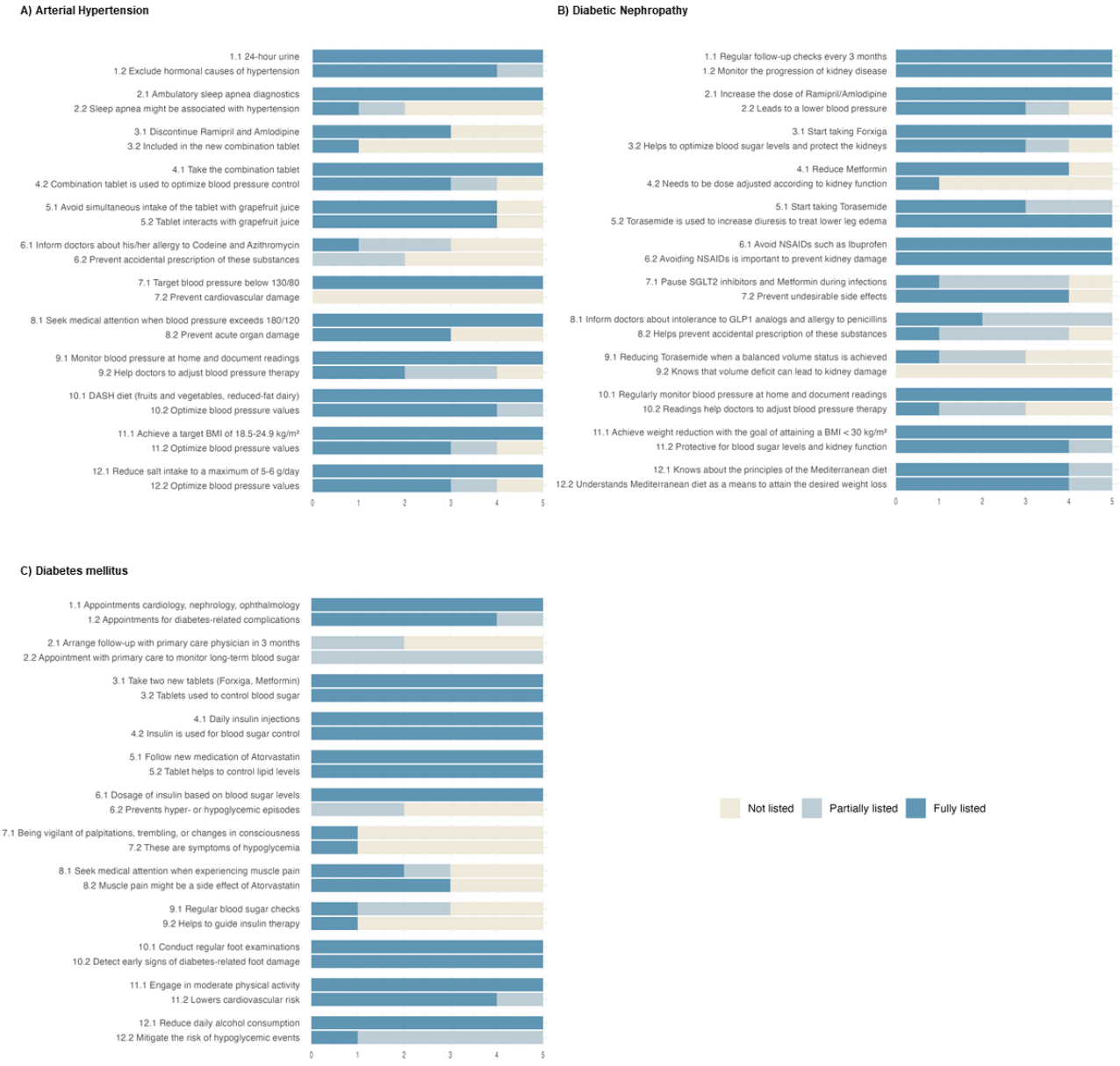

XSL•FO
RenderX

Figure 3. (A) Classification of learning objectives according to the relevant content field (organizational, medication, prevention of complications, and lifestyle). (B) Classification of learning objectives based on Bloom's revised taxonomy (Remember, Understand). The colors indicate the extent to which the learning outcome is included in the patient letter (dark blue: fully included; light blue: partially included; and yellow: not included). AHT: arterial hypertension; DKD: diabetic kidney disease; DM: diabetes mellitus.

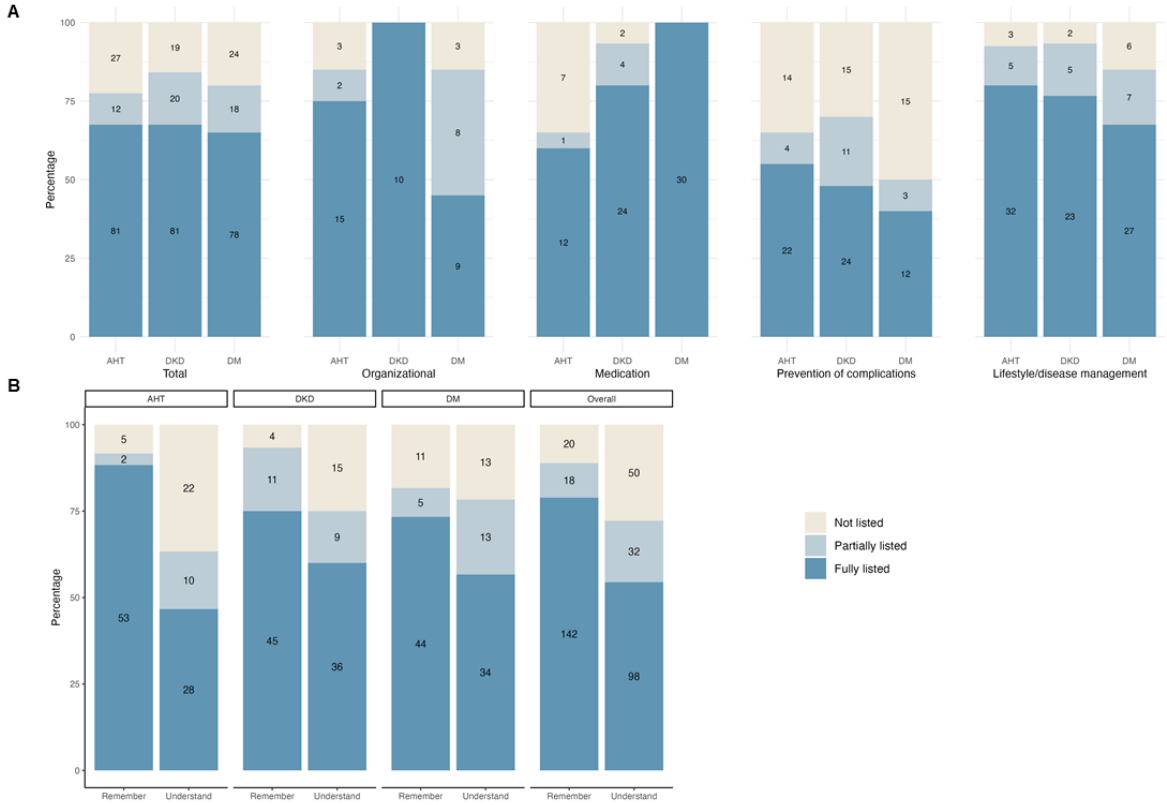

Indicators of Patient Centricity

The Flesch Reading Ease test scores for both the discharge letters and the patient-centered letters are shown in Figure 4. The patient-centered letters scored around 60, which corresponds to a "standard" reading level [26]. By contrast, the discharge letters scored around 40, indicating a "difficult" reading level. As illustrated in Figure 4, the patient-centered letters have an LIX score of approximately 50, while the discharge letters scored around 60, suggesting greater difficulty in reading. The patient-centered letters also scored lower than the discharge letters on the RIX, further demonstrating better readability.

Abbreviations were used in 49 of 273 (17.9%) discharge letter sentences, compared with 43 of 860 (5%) in the patient-centered letters. The abbreviation rate remained consistent across all patient letters (AHT: 14/273, 5.1%; DM: 14/273, 5.1%; and DKD: 15/273, 5.5%).

Repetitions were found in 37 of 273 (13.6%) discharge letter sentences and 207 of 860 (24.1%) patient-centered letter sentences. The number of repetitions varied slightly across the patient-centered letters (AHT: 60/278, 21.6%; DM: 75/271, 27.7%; and DKD: 72/311, 23.2%).

Medical terms appeared in 165 of 273 (60.4%) discharge letter sentences but in only 132 of 860 (15.3%) patient-centered letters. Of the medical terms used in patient-centered letters, 121 of 132 (91.7%) were explained.

Patients were directly addressed in 502 of 860 (58.4%) sentences in the patient-centered letters. This varied slightly between the different patient-centered letters (AHT: 151/278, 54.3%; DM: 149/271, 55%; DKD: 202/311, 65%).

XSL•FO
RenderX

Figure 4. The graph shows the readability scores for the discharge letters and the patient-centered letters. In the Flesch Reading Ease test, a higher score indicates a better readability (green, good readability), whereas in the Läsbarhetsindex (LIX) and Readability Index (RIX), a higher score represents lower readability (red, worse readability). AHT: arterial hypertension; DKD: diabetic kidney disease; DM: diabetes mellitus.

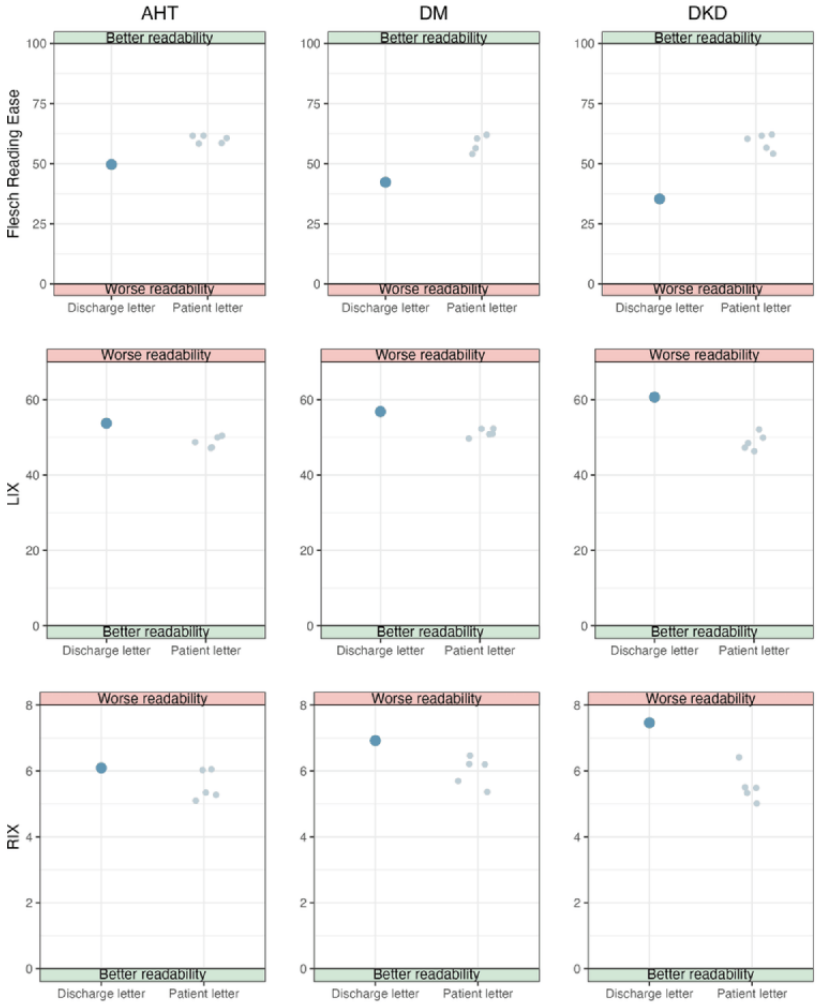

Medical Correctness, Case-Specific Relevance, and Source of Information

Of the 860 patient letters, 756 sentences (87.9%) were medically correct, while 31 were medically incorrect. The rate of medically correct sentences was comparable among the patient letters for AHT (250/278, 89.9%), DM (240/271, 88.6%), and DKD (266/311, 85.5%). Medical errors were found in 7 of 278 (2.5%) sentences in the patient-centered letters on AHT, 10 of 271 (3.7%) letters on DM, and 14 of 311 (4.5%) letters on DKD (Figure 5).

Regarding case-specific relevance, 756 of 860 (87.9%) sentences in the patient-centered letters were very relevant, while 72 of 860 (8.4%) were rather relevant. By contrast, 32 of 860 (3.7%) sentences were not relevant. The proportion of highly relevant sentences was similar across the patient-centered letters on AHT, DM, and DKD (Figure 5).

The information in the patient-centered letters was derived from the discharge letters in 616 of 860 (71.6%) sentences. Additional medical information was provided by GPT-4 in 207 of 860 (24.1%) sentences. However, these proportions varied across the specific conditions. For instance, the patient-centered letters on AHT were based on the discharge letters in 231 of 278 (83.1%) sentences, compared with 190/271 (70.1%) for DM and 195/311 (62.7%) for DKD. Further details are shown in Figure 5.

A qualitative analysis of the medical errors, conducted using thematic analysis based on Braun and Clarke [25], can be found in Multimedia Appendix 6.

No discernible pattern was observed in the errors within the patient-centered letters. Some errors were due to imprecision or incomplete information, while others resulted from incorrect assumptions made by GPT-4. Overall, most of the sentences (182/204, 89.2%) generated by GPT-4 were medically correct. However, of the medically incorrect sentences, the majority (22/31, 71%) contained information provided solely by GPT-4.

XSL•FO
RenderX

Figure 5. (A) Illustration of medically correct (green) and incorrect (red) sentences in the patient-centered letters with the absolute number of sentences displayed within the bars. (B) Representation of the case-specific relevance of the sentences in the patient-centered letters (blue: very relevant; brown: rather relevant; and yellow: neither/nor relevant). The absolute number of sentences is shown in the bars. (C) Proportion of sentences derived from the discharge letter (gray) and from GPT-4 (red) with the absolute number of sentences displayed within the bars. AHT: arterial hypertension; DKD: diabetic kidney disease; DM: diabetes mellitus.

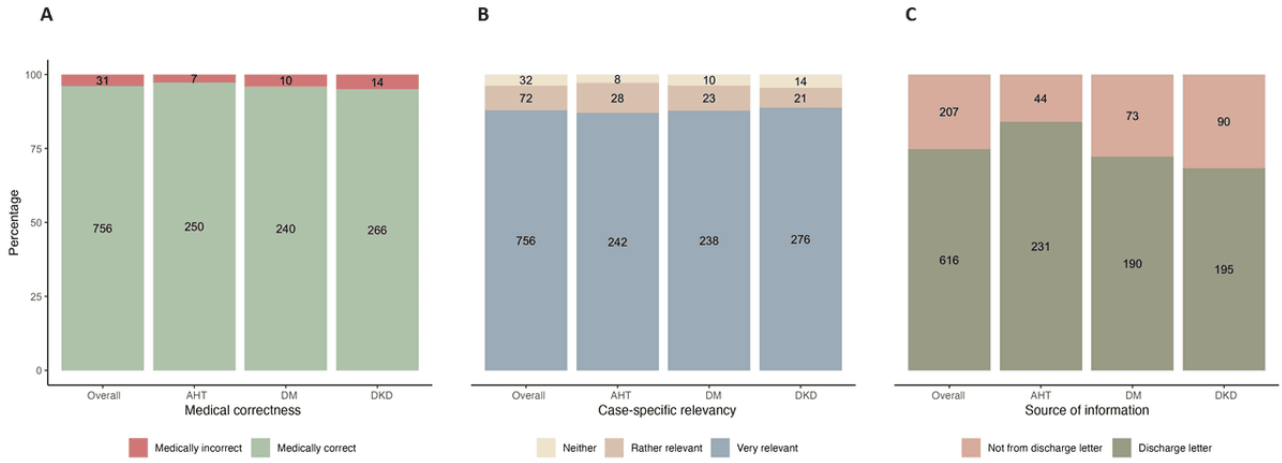

Discussion

Principal Findings

This study demonstrates the potential of GPT-4 to transform discharge letters into more readable, patient-centered letters. While this aligns with previous studies that show GPT-4's ability to generate patient-centered letters based on discharge letters or shorthand clinical instructions [16,27], our study goes beyond simply proving the concept of patient letter generation. We focused specifically on communicating relevant information from a patient-safety perspective. To our knowledge, this is the first study to analyze this aspect in AI-generated patient-centered letters.

Automated Learning Objective Identification and Transformation of Treatment and Patient Safety–Relevant Information

Our primary focus was on identifying and transferring information relevant to patient safety, defined as "learning objectives," in GPT-4–generated patient-centered letters. Overall, GPT-4 effectively identified and addressed 56 of the 72 learning objectives, with 5 partially addressed and 11 omitted entirely.

A key finding was that no significant differences were observed between the 3 diseases (AHT, DM, and DKD) in terms of the quality of GPT-4's identification and transformation of learning objectives.

Another important finding is that GPT-4 fully identified and transformed 56 of the 72 (78%) learning objectives from the discharge letters. In 5 of 72 (7%) cases, GPT-4 identified relevant learning objectives related to treatment and patient safety but omitted certain aspects during the transformation into patient-centered letters. In 11 of 72 (15%) cases, GPT-4 failed to identify any learning objectives, and thus, did not transform or include them in the patient-centered letters.

Consistent with previous studies in the medical field, we observed that the omission of relevant key information in the medical context, as highlighted in AI research, remains a prevalent issue [16,27].

Given that the partial omission of relevant key information in the medical context is highly pertinent to the quality and safety of patient care, we have, to the best of our knowledge, examined this phenomenon in detail for the first time in this study.

From our perspective, it is particularly noteworthy that we observed more frequent omissions by GPT-4 in relation to more complex medical requirements. These omissions were primarily associated with learning objectives that demand a deeper understanding (Bloom's category: Understand), as opposed to simpler objectives (Bloom's category: Remember). Basic topics, such as "lifestyle" and "organizational" aspects of self-management (eg, regularly attending follow-up examinations every 3 months), were more often identified and effectively transformed into patient-friendly language by GPT-4. More complex content, such as "prevention of complications" (eg, interactions between certain foods and medications), was less frequently addressed by GPT-4. This phenomenon was particularly evident in the processing and transformation of complex logical structures involving multiple dependencies.

To address this challenge effectively, strategies such as "few-shot learning," where the AI model is guided by several concise examples, and "chain-of-thought" prompts, which break down complex, multilevel problems into intermediate steps, could be promising. These approaches may enhance GPT-4's ability to translate complex medical information into clear, easily understandable statements and actionable directions [28,29].

In our case, "chain-of-thought" prompting would need to begin with an initial step that clarifies to the AI why the patient should adhere to the respective learning objective. For example, the first step could involve identifying the content relevant to the patient (eg, "pay attention to interactions"). The next step would be for the AI to break this content down into individual components (eg, medication, diet, kidney function). In the final step, the AI should organize these components in the correct

XSL•FO
RenderX

order. Only in the subsequent step should the AI check this order for existing dependencies between the individual information units and adjust it if necessary. This could help prevent the omission of complex learning objectives that are crucial for patient safety and quality of care. It should be noted that this approach extends the AI prompting, which may affect its performance in other areas. Therefore, this potentially promising strategy should be specifically explored in future studies to enhance the complete translation of complex medical information by GPT-4 into more easily understandable information for patients.

Regarding the learning objectives that were only partially transformed by GPT-4 from the discharge letters to the patient letters, we were unable to identify a clear connection with the content. Previous studies have demonstrated GPT-4's ability to extract specific information from medical notes with a high degree of accuracy. However, in those studies, GPT-4 focused exclusively on the task of information extraction.

By contrast, our study involved a multitasking prompt that required both the identification of learning objectives and the subsequent transformation of these medical facts into the correct context and simple language for the creation of patient-centered letters. Similar to the findings related to the complete omission of complex learning objectives, the challenge for AI in processing and reproducing complex logical structures with multiple dependencies was also evident in this case.

Particularly challenging for GPT-4 appeared to be learning objectives from which no immediate or simple logical pattern could be derived. This may be attributed to the multitasking prompts (identification, summarization, and simplification of medical information) and the overall high complexity of medical issues when translated into accessible language for patients. An alternative approach for the prompt could involve implementing "action points" to clearly structure the learning objectives into required actions, success criteria, and potential consequences. This strategy warrants further investigation in future studies.

Quality of Medical Information After Transformation by GPT-4

In this study, it was particularly important for us to assess the medical accuracy of the information transformed by GPT-4 in the patient-centered letters, as this is crucial for patient safety and quality of care. Any inaccuracies must therefore be regarded as a potential risk in the use of LLMs in a medical context.

Our results demonstrate that the patient-centered letters generated by GPT-4 exhibited a high degree of medical accuracy. The low error rate observed in our study may be attributed to the detailed medical information contained in the discharge letters, which GPT-4 accessed during the transformation process. This extensive data set enabled GPT-4 to accurately translate the medical information into patient-friendly language. Previous studies have shown that, in addition to information omissions, factors such as the type of task (eg, text summarization) and the specific field of application (eg, medical writing, question-answer format) influence the accuracy of LLM transformations [30]. Notably, our study demonstrates a significant improvement in medical accuracy and overall result quality compared with prior work examining various scenarios for the use of LLMs in the medical context [27,31].

A critical issue in transforming medical information using LLMs is the phenomenon of "hallucinations," where coherent and grammatically correct, yet factually incorrect or misleading, information is generated [32,33]. These hallucinations can jeopardize patient safety and the quality of care. While previous studies have reported hallucination rates as high as 40%, our study encountered this phenomenon only in isolated cases [34]. Nevertheless, despite the observed downward trend, the risk remains significant with the ongoing development of LLMs, as the following example illustrates:

Discharge letter: *(...) An ultrasound of the pancreas showed no signs of a mass. (...)*

Patient-centered letter: *(...) they also tested to see if your pancreas has any issues, which it does not (...).*

Here, GPT-4 makes a general statement about the unremarkable ultrasound of the pancreas, despite the fact that medically, only a malignant disease was excluded, and other potential pathologies were not examined at all. While these types of errors, resulting from the hallucination phenomenon, occurred only very rarely, they still pose a significant risk to patient safety and the quality of care. Therefore, despite their low frequency, this phenomenon must be considered a substantial limitation to the use of LLMs in medical contexts and should be specifically addressed in future studies.

When using LLMs such as GPT-4 with sensitive medical information, ensuring compliance with data protection and information security requirements is paramount. Before deploying LLMs in health care settings, relevant experts must be consulted for each specific application scenario to define appropriate regulations for anonymizing medical information and to establish a secure and resilient infrastructure.

Patient Centricity

In terms of readability and patient centricity—measured by standard readability scores, word and sentence counts, and the use of specialized terminology, explanations, abbreviations, and direct patient address—the GPT-4–generated patient-centered letters significantly outperformed the conventional discharge letters.

This is consistent with previous studies that analyzed AI-generated patient letters, although using less comprehensive methodologies, which showed high readability ranging from the sixth [16] to the ninth [27] grade level. It is important to note that these studies typically used the Flesch Reading Ease, the Flesch-Kincaid Reading Level, or the PEMAT score for readability assessments [16,35].

The Flesch Reading Ease is one of the most commonly used tools for evaluating readability in medical literature, with scores ranging from 0 (unreadable) to 100 (very easy to read) [22]. However, a major limitation of this approach is that it only considers sentence and word length, failing to adequately account for the level of medical knowledge required by patients or the frequency of medical terms used.

XSL•FO
RenderX

From the authors' perspective, these factors, along with directly addressing patients, are crucial for how discharge letters are perceived and understood. Previous studies have shown that avoiding abbreviations and explaining medical terms in layperson's terms significantly improve the comprehensibility of discharge letters from the patient's viewpoint [36]. Additionally, medical letters are demonstrably better received by patients when they are personally addressed in the letter [37].

For this study, we implemented additional methodological approaches to analyze, for the first time, the patient-centered transformation of medical information by GPT-4. These advanced methods not only accounted for the use of medical terms and abbreviations but also included their explanations and the occurrence of repetitions.

It is noteworthy that the patient-centered letters transformed by GPT-4 demonstrated a significantly higher level of patient centricity than conventional discharge letters, even when accounting for these advanced parameters, which were examined for the first time. The GPT-4–generated letters contained notably fewer medical terms and abbreviations, explained medical terms in simple language, and directly addressed the patients.

Limitations

Our study has several limitations. First, GPT-4's generation of patient-centered letters was based solely on extensive prompting techniques. In the long term—especially with the increasing availability of additional LLMs [38,39]—fine-tuning (both general and domain-specific) and the potential application of specialized LLMs should be considered for the creation of patient-centered letters. Second, our investigation focused on major widespread diseases that occur frequently but did not encompass all medical specialties. As a result, our findings may not be fully transferable to all medical specialties. Another limitation is the exclusive use of the LLM "GPT-4," which prevents us from drawing conclusions about the performance of other LLMs in generating patient-centered letters based on discharge letters.

Furthermore, we used a new multidimensional approach to assess the patient centricity of GPT-4–based communication of medical information. This approach considered factors such as the use of medical terms, abbreviations, and repetitions. While these variables are objectively measurable and part of common readability standards, they have not been sufficiently validated in the context of LLM-based communication. Therefore, further studies are needed to explicitly examine how patients perceive and understand the patient letters created by LLMs, as well as how this innovative form of patient-centered communication impacts patient empowerment and self-management of their health. Additionally, the perspective of the treating physicians on this promising form of patient-centered communication was not explored in our study. This aspect should also be considered in future research.

Additionally, the discharge letters, prompts, and patient letters were written in German, and all content was translated into English for the preparation of the manuscript. As a limitation, it should be noted that the transferability of our study results to the use of GPT in other languages requires further investigation.

Conclusions

In summary, our study demonstrates that GPT-4 has the potential to significantly enhance the patient-centeredness of discharge letters. While we used a detailed prompting technique and GPT-4 generally exhibits a high degree of medical accuracy when transforming discharge letters into more patient-friendly formats, it is not yet fully suitable for patient care without review by medical professionals, particularly due to the noted omissions and hallucinations. Despite the already strong readability and patient orientation, even an advanced language model like GPT-4 did not fully account for all information relevant to patient safety and quality of care in the patient-centered letters. Further advancements in prompting techniques and the targeted development of language models for the medical field could help minimize these limitations. If these challenges are addressed, GPT-4 could offer enhanced potential to support health care professionals in patient-centered communication and improve patients' understanding of their medical conditions. This would mark a significant step toward better patient safety and improved quality of care.

Acknowledgments

We wish to thank the Open Access Publishing Fund of the University of Tübingen for supporting our study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Discharge letters, original prompting, examples of patient letters, and table with learning objectives.
[DOCX File , 151 KB - jmir_v27i1e67143_app1.docx]

Multimedia Appendix 2
Illustration of the multistep process used for the development of discharge letters. Three common diseases were selected as the basis for the discharge letters. To assess GPT-4's ability to identify critical medical information, 24 learning objectives were developed in alignment with Bloom's taxonomy. Examples are given in the figure.
[PNG File , 404 KB - jmir_v27i1e67143_app2.png]

XSL•FO
RenderX

Multimedia Appendix 3
Overview of the multistep process of GPT-4 prompting. The 3 tasks involved in the generation of the patient-centered letters are illustrated in the rectangles. The direction of the arrows indicates the sequential order of the individual steps.
[PNG File , 278 KB - jmir_v27i1e67143_app3.png]

Multimedia Appendix 4
Qualitative analysis of missing learning objectives. This table illustrates learning objectives that were not included in the majority of patient-centered letters (≥3/5 letters). Recommendations are provided on how the prompting can be refined to address the missing information.
[DOCX File , 17 KB - jmir_v27i1e67143_app4.docx]

Multimedia Appendix 5
Analysis of incomplete learning objectives. The table presents examples of learning objectives that have been partially listed in the patient-centered letters including recommendations for further prompt engineering.
[DOCX File , 18 KB - jmir_v27i1e67143_app5.docx]

Multimedia Appendix 6
Qualitative analysis of medical errors. Medical errors made by GPT-4 were categorized and analyzed based on content area. The errors were classified into different error categories: imprecise, incorrect, incomplete, and presumptive.
[DOCX File , 18 KB - jmir_v27i1e67143_app6.docx]

References

1. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
2. Sabaté E. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
3. DiMatteo MR, Haskard-Zolnierek KB, Martin LR. Improving patient adherence: a three-factor model to guide practice. Health Psychology Review 2012 Mar;6(1):74-91. [doi: 10.1080/17437199.2010.537592]
4. Longtin Y, Sax H, Leape LL, Sheridan SE, Donaldson L, Pittet D. Patient participation: current knowledge and applicability to patient safety. Mayo Clin Proc 2010 Jan;85(1):53-62 [FREE Full text] [doi: 10.4065/mcp.2009.0248] [Medline: 20042562]
5. Schwarz CM, Hoffmann M, Schwarz P, Kamolz L, Brunner G, Sendlhofer G. A systematic literature review and narrative synthesis on the risks of medical discharge letters for patients' safety. BMC Health Serv Res 2019 Mar 12;19(1):158-110 [FREE Full text] [doi: 10.1186/s12913-019-3989-1] [Medline: 30866908]
6. Roy CL, Poon EG, Karson AS, Ladak-Merchant Z, Johnson RE, Maviglia SM, et al. Patient safety concerns arising from test results that return after hospital discharge. Ann Intern Med 2005 Jul 19;143(2):121-128. [doi: 10.7326/0003-4819-143-2-200507190-00011] [Medline: 16027454]
7. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med 2003 Feb 04;138(3):161-167. [doi: 10.7326/0003-4819-138-3-200302040-00007] [Medline: 12558354]
8. Harris E, Rob P, Underwood J, Knapp P, Astin F. Should patients still be copied into their letters? A rapid review. Patient Educ Couns 2018 Dec;101(12):2065-2082. [doi: 10.1016/j.pec.2018.06.014] [Medline: 30420045]
9. Smolle C, Schwarz CM, Hoffmann M, Kamolz L, Sendlhofer G, Brunner G. Design and preliminary evaluation of a newly designed patient-friendly discharge letter - a randomized, controlled participant-blind trial. BMC Health Serv Res 2021 May 12;21(1):450 [FREE Full text] [doi: 10.1186/s12913-021-06468-3] [Medline: 33975590]
10. Nguyen DL, Ambinder EB, Jones MK, Hill G, Harvey SC. Improving patient comprehension of screening mammography recall lay letters. J Am Coll Radiol 2019 Dec;16(12):1669-1676. [doi: 10.1016/j.jacr.2019.05.029] [Medline: 31199890]
11. Wolff J, Auber G, Schober T, Schwär F, Hoffmann K, Metzger M, et al. Work-time distribution of physicians at a German university hospital. Dtsch Arztebl Int 2017 Oct 20;114(42):705-711 [FREE Full text] [doi: 10.3238/arztebl.2017.0705] [Medline: 29122102]
12. Garabet R, Mackey BP, Cross J, Weingarten M. ChatGPT-4 performance on USMLE step 1 style questions and its implications for medical education: a comparative study across systems and disciplines. Med Sci Educ 2024 Feb;34(1):145-152. [doi: 10.1007/s40670-023-01956-z] [Medline: 38510401]
13. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 04;47(1):33-38 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
14. Johnson-Laird P. Deductive reasoning. Wiley Interdiscip Rev Cogn Sci 2010 Jan;1(1):8-17. [doi: 10.1002/wcs.20] [Medline: 26272833]

XSL•FO
RenderX

15. Eppler MB, Ganjavi C, Knudsen JE, Davis RJ, Ayo-Ajibola O, Desai A, et al. Bridging the gap between urological research and patient understanding: the role of large language models in automated generation of layperson's summaries. Urol Pract 2023 Sep;10(5):436-443. [doi: 10.1097/UPJ.0000000000000428] [Medline: 37410015]
16. Zaretsky J, Kim JM, Baskharoun S, Zhao Y, Austrian J, Aphinyanaphongs Y, et al. Generative artificial intelligence to transform inpatient discharge summaries to patient-friendly language and format. JAMA Netw Open 2024 Mar 04;7(3):1-10 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.0357] [Medline: 38466307]
17. Mac O, Ayre J, Bell K, McCaffery K, Muscat DM. Comparison of readability scores for written health information across formulas using automated vs manual measures. JAMA Netw Open 2022 Dec 01;5(12):1-8 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.46051] [Medline: 36508219]
18. Herrmann-Werner A, Festl-Wietek T, Holderried F, Herschbach L, Griewatz J, Masters K, et al. Assessing ChatGPT's mastery of Bloom's taxonomy using psychosomatic medicine exam questions: mixed-methods study. J Med Internet Res 2024 Jan 23;26:e52113-e52113. [doi: 10.2196/52113] [Medline: 38261378]
19. Bloom BF, Engelhart M, Hill W, Krathwohl D. Taxonomy of Educational Objectives: The Classification of Educational Goals. London, UK: Longman Green & Co; 1956.
20. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory Into Practice 2010 Jun 24;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
21. Rooney MK, Santiago G, Perni S, Horowitz DP, McCall AR, Einstein AJ, et al. Readability of patient education materials from high-impact medical journals: a 20-year analysis. J Patient Exp 2021;8:1-9 [FREE Full text] [doi: 10.1177/2374373521998847] [Medline: 34179407]
22. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of flesch formula. Educ Health (Abingdon) 2017;30(1):84-88. [doi: 10.4103/1357-6283.210517] [Medline: 28707643]
23. Björnsson C. Läsbarhet. Stockholm, Sweden: Liber; 1968.
24. Anderson R, Freebody P. Reading comprehension and the assessment and acquisition of word knowledge. Advances in Reading/Language Research 1983;2:231-256.
25. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Spadaro DC, Robinson LA, Smith LT. Assessing readability of patient information materials. Am J Hosp Pharm 1980 Feb;37(2):215-221. [Medline: 7361793]
27. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
28. Paranjape B, Lundberg S, Singh S, Hajishirzi H, Zettlemoyer L, Ribeiro M. ART: automatic multi-step reasoning and tool-use for large language models. arXiv. Preprint posted online on March 16, 2023 2023:1-26 [FREE Full text]
29. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New York, NY: Curran Associates Inc; 2023 Presented at: Conference on Neural Information Processing Systems; November 28 to December 9, 2022; New Orleans, LA.
30. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1-5 [FREE Full text] [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
31. Eriksen AV, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. NEJM AI 2024 Jan 11;1(1):1-3. [doi: 10.1056/aip2300031]
32. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR Med Educ 2023 Oct 20;9:1-13 [FREE Full text] [doi: 10.2196/48785] [Medline: 37862079]
33. Masters K. Medical teacher's first ChatGPT's referencing hallucinations: lessons for editors, reviewers, and teachers. Med Teach 2023 Jul;45(7):673-675. [doi: 10.1080/0142159X.2023.2208731] [Medline: 37183932]
34. Chen T, Kaminski E, Koduri L, Singer A, Singer J, Couldwell M, et al. Chat GPT as a Neuro-Score calculator: analysis of a large language model's performance on various neurological exam grading scales. World Neurosurg 2023 Nov;179:e342-e347 [FREE Full text] [doi: 10.1016/j.wneu.2023.08.088] [Medline: 37634667]
35. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can Artificial Intelligence Improve the Readability of Patient Education Materials? Clin Orthop Relat Res 2023 Nov 01;481(11):2260-2267. [doi: 10.1097/CORR.0000000000002668] [Medline: 37116006]
36. Weetman K, Dale J, Scott E, Schnurr S. Adult patient perspectives on receiving hospital discharge letters: a corpus analysis of patient interviews. BMC Health Serv Res 2020 Jun 15;20(1):537-511. [doi: 10.1186/s12913-020-05250-1] [Medline: 32539716]
37. Mahadavan L, Bird NJ, Chadwick M, Daniels IR. Prospective assessment of patient directed outpatient communication from a patient and general practitioner perspective. Postgrad Med J 2009 Aug;85(1006):395-398. [doi: 10.1136/pgmj.2008.068601] [Medline: 19633003]

XSL•FO
RenderX

38. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. NPJ Digit Med 2023 Nov 16;6(1):1-10 [FREE Full text] [doi: 10.1038/s41746-023-00958-w] [Medline: 37973919]
39. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, et al. Health system-scale language models are all-purpose prediction engines. Nature 2023 Jul;619(7969):357-362 [FREE Full text] [doi: 10.1038/s41586-023-06160-y] [Medline: 37286606]

Abbreviations

AHT: arterial hypertension
AI: artificial intelligence
DKD: diabetic kidney disease
DM: diabetes mellitus
LIX: Läsbarhetsindex
LLM: large language model
PEMAT: Patient Education Materials Assessment Tool
RIX: Readability Index
WHO: World Health Organization

Edited by A Coristine; submitted 04.10.24; peer-reviewed by G Tofler, J Walsh; comments to author 04.11.24; revised version received 22.11.24; accepted 28.11.24; published 21.01.25.

Please cite as:
Eisinger F, Holderried F, Mahling M, Stegemann–Philipps C, Herrmann–Werner A, Nazarenus E, Sonanini A, Guthoff M, Eickhoff C, Holderried M
What's Going On With Me and How Can I Better Manage My Health? The Potential of GPT-4 to Transform Discharge Letters Into Patient-Centered Letters to Enhance Patient Safety: Prospective, Exploratory Study
J Med Internet Res 2025;27:e67143
URL: https://www.jmir.org/2025/1/e67143
doi:10.2196/67143
PMID:

©Felix Eisinger, Friederike Holderried, Moritz Mahling, Christian Stegemann–Philipps, Anne Herrmann–Werner, Eric Nazarenus, Alessandra Sonanini, Martina Guthoff, Carsten Eickhoff, Martin Holderried. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Shared Decision-Making Tools Implemented in the Electronic Health Record: Scoping Review

Joni H Pierce[1], MBA, MS; Charlene Weir[1], RN, PhD; Teresa Taft[1], PhD; William Richards II[1], MS; Mary M McFarland[2], BS; Kensaku Kawamoto[1], MD, MHS, PhD; Guilherme Del Fiol[1], MD, PhD; Jorie M Butler[1,3,4,5], PhD

[1]Department of Biomedical Informatics, Spencer Fox Eccles School of Medicine, University of Utah, Salt Lake City, UT, United States

[2]Spencer S. Eccles Health Sciences Library, University of Utah, Salt Lake City, UT, United States

[3]Division of Geriatrics, Department of Internal Medicine, University of Utah, Salt Lake City, UT, United States

[4]Geriatrics Research, Education, and Clinical Center (GRECC), VA Salt Lake City Healthcare System, Salt Lake City, United States

[5]Informatics Decision-Enhancement and Analytic Sciences Center (IDEAS), VA Salt Lake City Healthcare System, Salt Lake City, United States

Corresponding Author:
Joni H Pierce, MBA, MS
Department of Biomedical Informatics
Spencer Fox Eccles School of Medicine
University of Utah
421 Wakara Way
Suite 140
Salt Lake City, UT, 84108
United States
Phone: 1 801 588 9395
Email: jonipierce@comcast.net

Abstract

Background: Patient-centered care promotes the involvement of patients in decision-making related to their health care. The adoption and implementation of shared decision-making (SDM) into routine care are constrained by several obstacles, including technical and time constraints, clinician and patient attitudes and perceptions, and processes that exist outside the standardized clinical workflow.

Objective: We aimed to understand the integration and implementation characteristics of reported SDM interventions integrated into an electronic health record (EHR) system.

Methods: We conducted a scoping review using the methodological framework by Arksey and O'Malley with guidance from the Joanna Briggs Institute. Eligibility criteria included original research and reviews focusing on SDM situations in a real-world clinical setting and EHR integration of SDM tools and processes. We excluded retrospective studies, conference abstracts, simulation studies, user design studies, opinion pieces, and editorials. To identify eligible studies, we searched the following databases on January 11, 2021: MEDLINE, Embase, CINAHL Complete, Cochrane Library including CENTRAL, PsycINFO, Scopus, and Web of Science Core Collection. We systematically categorized descriptive data and key findings in a tabular format using predetermined data charting forms. Results were summarized using tables and associated narratives related to the review questions.

Results: Of the 2153 studies, 18 (0.84%) were included in the final review. There was a high degree of variation across studies, including SDM definitions, standardized measures, technical integration, and implementation strategies. SDM tools that targeted established health care processes promoted their use. Integrating SDM templates and tools into an EHR appeared to improve the targeted outcomes of most (17/18, 94%) studies. Most SDM interventions were designed for clinicians. Patient-specific goals and values were included in 56% (10/18) of studies. The 2 most common study outcome measures were SDM-related measures and SDM tool use.

Conclusions: Understanding how to integrate SDM tools directly into a clinician's workflow within the EHR is a logical approach to promoting SDM into routine clinical practice. This review contributes to the literature by illuminating features of SDM tools that have been integrated into an EHR system. Standardization of SDM tools and processes, including the use of patient decision aids, is needed for consistency across SDM studies. The implementation approaches for SDM applications showed varying levels of planning and effort to promote SDM intervention awareness. Targeting accepted and established clinical processes

XSL•FO
RenderX

may enhance the adoption and use of SDM tools. Future studies designed as randomized controlled trials are needed to expand the quality of the evidence base. This includes the study of integration methods into EHR systems as well as implementation methods and strategies deployed to operationalize the uptake of the SDM-integrated tools. Emphasizing patients' goals and values is another key area for future studies.

(J Med Internet Res 2025;27:e59956) doi:10.2196/59956

KEYWORDS

shared decision-making; patient engagement; decision support; electronic health records

Introduction

Background

Shared decision-making (SDM) is a model of patient-centered care that encourages patients and clinicians to work together to reach medical decisions by weighing the risks and benefits of various options within the context of the values and goals of the patient [1]. SDM is the process of communication, deliberation, and decision-making between clinicians and their patients [2,3]. During this process, treatment options, including risks and benefits, are discussed and patient preferences are explored to inform decision-making. Previous research has shown that SDM increases patient engagement, knowledge, risk comprehension, and participation in decision-making [4-6]. Furthermore, SDM has been shown to improve patient–health care provider communication, patient satisfaction, compliance, and clinical outcomes [7,8]. Notably, the US Preventive Services Task Force considers SDM an ethical right of all patients, independent of health outcomes [9].

Currently, SDM is not routinely required by large health care payers, such as the Centers for Medicare and Medicaid Services, which is a US federal agency providing health care for patients with low income or older adult patients. Medicare is the largest single payer for health care in the United States [10]. The Centers for Medicare and Medicaid Services is slowly adopting SDM standards for limited and specific procedures, conditions, and devices, including lung cancer screening, atrial fibrillation, and implantable defibrillators [11]. US states such as Washington and Vermont are also beginning to mandate SDM for specific conditions and other states such as Connecticut, Massachusetts, Maine, Minnesota, New Hampshire, and Oklahoma are considering legislation [12].

The Health Information Technology for Economic and Clinical Health Act was adopted in the United States in 2009, enabling the widespread adoption of electronic health record (EHR) systems across the US health care system [13]. Leveraging EHR systems to promote the use and uptake of SDM appears to be a logical approach to operationalizing SDM programs. For example, Kuo et al [7] found several positive effects of integrating SDM tools within an EHR system, which include improved clinical outcomes, positive lifestyle behavior changes, more deliberation with clinicians, and less decisional conflict. Despite the interest in incorporating SDM into routine care, current research studies identify a variety of obstacles that limit SDM adoption. Some of these obstacles include technical integration issues; logistical and workflow challenges; and psychological impediments, such as uncertainty and legacy belief systems, which continue to impede progress [14-16]. Integrating SDM tools and processes into EHR systems is often a complex and difficult problem [17-21].

SDM may involve the use of a decision aid, a type of tool that helps patients consider options for medical decision-making by increasing patient knowledge; illuminating treatment options including risks, benefits, and efficacy; and exploring patient preferences and values [22-24]. The Ottawa Decision Support Framework provides recommendations and guidelines for health care decisions and standardized checklists for the development and evaluation of decision aids based on the International Patient Decision Aid Standards [25]. SDM components found within the International Patient Decision Aid Standards include clarifying patient values, coaching and guiding patients when making decisions, and ensuring literacy levels are addressed [26-28]. Decision aids developed using the Ottawa Decision Support Framework as well as other decision aids outperformed the usual care for performance in a large overview of systematic reviews [25,28]. Patient decision aids can play a key role in facilitating the occurrence of SDM at the point of care as clinicians may use a decision aid to support SDM conversations. Furthermore, decision aids may provide a way to standardize SDM experiences. Notably, decision aid research developed before the widespread adoption of EHRs [29-32]. Decision aids are not synonymous with SDM processes as SDM processes may incorporate the use of a decision aid or may use a conversation between patient and clinician without the use of a decision aid. Decision aids typically contain patient education as part of the decision aid content; however, a clinician may choose to verbally educate a patient without the use of a decision aid. Previous research has shown that decision aids integrated into an EHR system show improvement in patient-centered outcomes, including a reduction in decisional conflict. However, there are significant barriers and complexities to overcome when integrating them into commercial EHR systems [33-35].

Both decision aids and SDM tools still face substantial hurdles to clinician uptake [15,32,36-38]. Gaining the attention of health care providers can be difficult as SDM relies on health care provider discretion, time, and required effort [15,39-41]. Integration into the EHR system is recognized as an important strategy as it enables SDM to be part of clinical workflow [42]. However, integration can also pose challenges because unavoidable health care provider alerts run the risk of high levels of alert fatigue or reactance. SDM tool developers and health care systems must be judicious in their choice of tools to develop and integrate. It is important to understand the characteristics of SDM tools that have been integrated into EHR systems.

XSL•FO
RenderX

Objectives

A literature review revealed limited studies describing the integration and implementation of SDM tools within an EHR workflow [43-46]. A number of recent systematic reviews have been published on the general topic of SDM [19,33,47-58]. These studies focused on a range of SDM topics, including SDM within specific surgical specialties, behavioral theories underpinning SDM, organizational characteristics to support implementation, validated SDM measurement tools, SDM key components, and quality of SDM. However, at the time of the protocol registration, we were unable to identify a systematic or scoping review that specifically addressed the characteristics of SDM tools that have been integrated into an EHR. Thus, this study aimed to understand the *characteristics of SDM tools* that have been integrated into an EHR system and implemented within a health care system by conducting a scoping review.

Methods

Overview

This scoping review followed methods described by Arksey and O'Malley [59], Levac et al [60], and the Joanna Briggs Institute (JBI) [61]. The study protocol is published with the Protocol Registry of Evidence Reviews at the University of Utah. Consensus was achieved on the inclusion and exclusion criteria and primary research question. For transparency and reproducibility, we followed the PRISMA-ScR (Preferred Reporting Items For Systematic Reviews And Meta-Analyses Extension For Scoping Reviews) and PRISMA-S (Preferred Reporting Items For Systematic Reviews And Meta-Analyses Search Extension) reporting guidelines for the scoping review protocol and manuscript [62,63].

Study Identification

An information specialist developed the search strategies using database-specific subject headings, keywords, and team feedback for the key concepts of SDM, patient engagement, decision support, and EHRs. Example publications were provided to the information specialist for strategy development and term harvesting. The search strategy was evaluated by a librarian coauthor using the Peer Review of Electronic Search Strategies guidelines [64]. EndNote (version 20; Clarivate) was used to manage citations and remove duplicates. Covidence (Veritas Health Innovation) provided a second pass for duplicate removal. Search strategies are provided in Multimedia Appendix 1.

The following databases were used to identify eligible studies between 2009 and 2021: MEDLINE (via Ovid platform), Embase (via Embase platform), CINAHL Complete (via EBSCOhost platform), Cochrane Library (via Wiley platform) including CENTRAL (via Wiley platform), PsycINFO (via EBSCOhost platform), Scopus (via Scopus platform), and Web of Science Core Collection (via Clarivate Analytics). No filters other than date limits were applied. The date limit from 2009 corresponds to the adoption of the Health Information Technology for Economic and Clinical Health Act in the United States [65].

Study Screening

Covidence software was used for screening. Four researchers independently screened titles and abstracts for inclusion and exclusion criteria and met for discussion. The protocol was modified during this phase to expand inclusion criteria based on discussions among the screeners. For example, after finding a number of design studies and concept papers, we added the requirement that all studies must have SDM tools that have been implemented into a real-world clinical setting. Furthermore, we excluded retrospective studies after identifying studies that focused on a retrospective analysis of clinical notes and activities. Changes to the protocol are listed in Textbox 1. Following the screening process, 303 (10%) of the 3017 articles were identified for full eligibility assessment.

Expanded screening criteria were added to the original inclusion and exclusion criteria identified in the research protocol, as noted in Textbox 1. The additional categories were identified and added as the team recognized additional screening criteria as screening commenced.

Textbox 1. Expanded scoping review inclusion and exclusion screening criteria.

Inclusion criteria

- Original studies, systematic reviews, and meta-analyses
- Decisions made at the point of care
- Electronic health record–integrated shared decision-making tool
- Intervention implemented into a real clinical setting
- Shared decision-making tool outcomes reported
- Studies written in English

Exclusion criteria

- Retrospective studies
- Conference abstracts
- Limited measures

XSL•FO
RenderX

Eligibility Assessment

A team of 5 researchers, including 2 professors with extensive backgrounds in cognitive and sociotechnical informatics (CW and JMB), 1 informatics researcher experienced in sociotechnical design (TT), an informatics student (JHP), and another informatics student (WR II), met on a weekly basis to discuss inclusion and exclusion criteria and extraction goals for research studies. Initially, each researcher was assigned the same 5 studies to assess for eligibility review. Then, the research team discussed and deliberated on whether each study met the established inclusion criteria until the team achieved consensus. A spreadsheet was created with each evaluation component, including basic study characteristics, integration, and implementation methods. Study characteristics included study type, SDM intervention, clinical domain and conditions, measures, outcomes, integration approaches, and implementation processes. The research team discussed the criteria and achieved consensus for spreadsheet components through discussion, debate, and review. The reviewers then evaluated approximately 10 studies per week as a group, using the spreadsheet to determine criteria for analysis against the study aims. After this, the remaining studies were distributed among the team members for a full review of eligibility. We excluded studies that did not report measurable quantitative outcomes, were not integrated into an EHR system, did not involve a shared decision (with or without a decision aid), and were not in person at the point of care. EHR integration for the purpose of this review ranged from scanned PDFs appended to EHR patient records to full technical integration using semantic interoperability methods.

We identified the studies of tools designed to support SDM that were integrated into the EHR system for evaluation and reported a total of 18 studies, which were included for final review. See PRISMA-ScR Figure 1 and Multimedia Appendix 2.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram. EHR: electronic health record; SDM: shared decision-making.

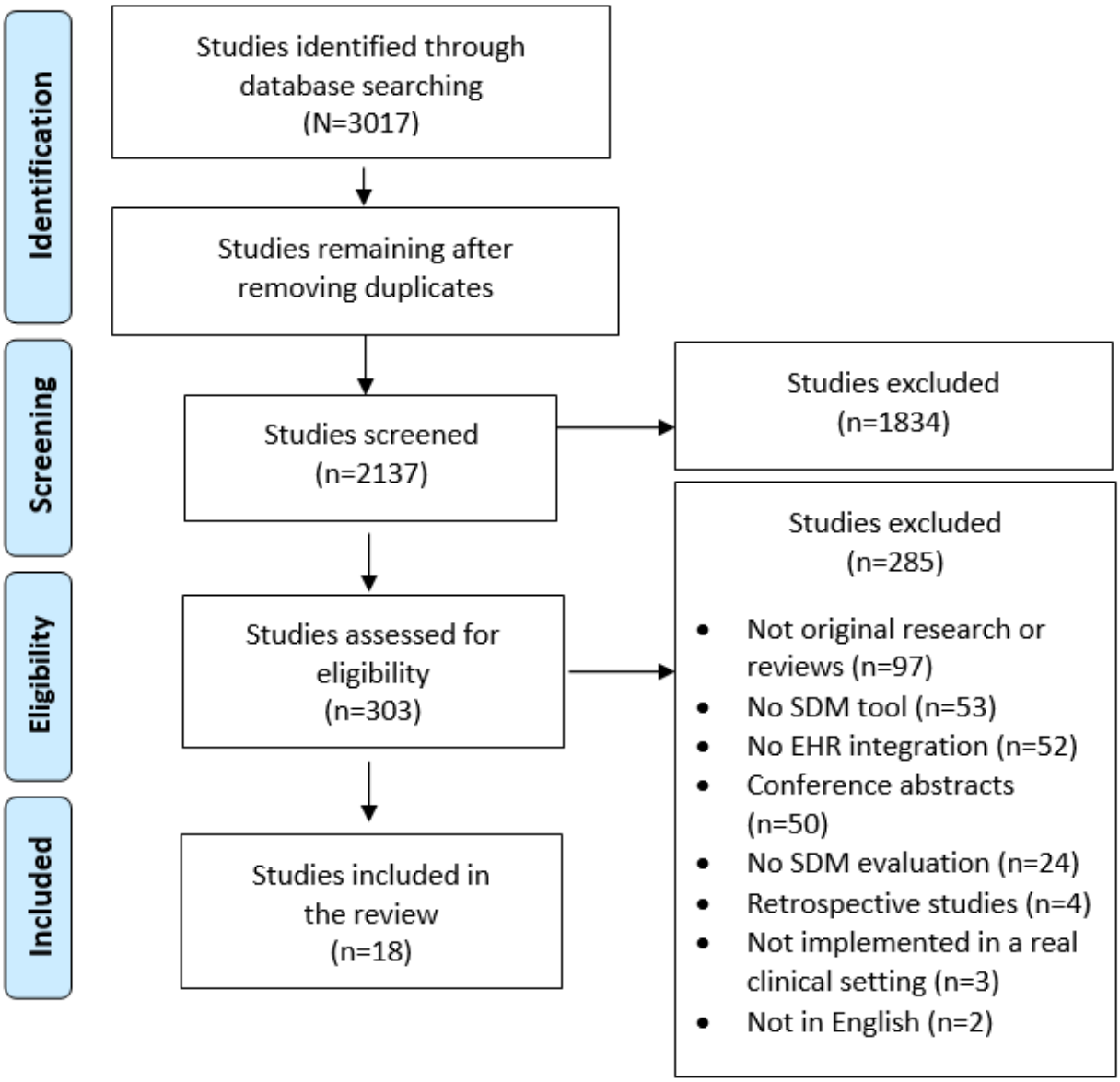

XSL•FO
RenderX

Review Process

Overview

The research team used the following three-step process to interpret and synthesize the final 18 studies included in the scoping review: (1) data charting, (2) analysis of evidence, and (3) data summary and synthesis. During this process, we performed a detailed review and analysis of the integration methods along with the associated clinical decision support features and then mapped them into a table. Implementation strategies and methods were evaluated and mapped according to the narrative descriptions in the studies.

Data Charting

Data charting was performed according to the methods proposed by Arksey and O'Malley [59] and the JBI, as described by Tricco et al [63], Levac et al [60], and Peters et al [66]. The template was piloted by the researchers and adjusted based on discussion, debate, and deliberation. Data items included the following: (1) study design, (2) aims, (3) measures, (4) outcomes, (5) intervention type, (6) clinical domain, (7) clinical processes, (8) clinical conditions, (9) EHR integration methods, (10) implementation strategies, (11) SDM goals, and (12) SDM model and components.

Analysis of Evidence

The analysis of the data was then performed to understand and map themes and clusters of notable data according to the updated methodological guidance for scoping reviews [67]. To do this, the research team discussed the definition along with examples of each extraction category. Following this, one researcher (WR II) performed a preliminary extraction on categories that were unambiguous. Next, the remaining research team members were assigned manuscripts to extract the more complex categories. The team met for iterative discussions, deliberation, and decision-making to achieve consensus on the final extraction items.

Data Summary and Synthesis

We summarized our findings using narrative synthesis and descriptive statistics to describe the characteristics of SDM tools integrated into the EHR systems [68,69]. For general characteristics, we mapped study types, measures, clinical domains and conditions, and goals of the SDM intervention in a table and calculated frequency counts and percentages [69]. We then categorized each study according to integration methods and features using descriptive content analysis to summarize narrative findings [70]. To do this, we analyzed the various methods used in each of the studies to surface and document SDM tools within the EHR systems. We considered integration to have been achieved so long as the SDM tools surfaced in the clinician's workflow within the EHR systems. We mapped EHR clinical decision support features for each SDM tool according to risk, patient data display, health care provider notification, documentation, and EHR workflow presentation. Then, we analyzed the implementation approaches across studies. To do this, we looked for methods of creating awareness, training programs, and SDM tool buy-in and promotion through clinician leaders. We were then able to organize our results by general characteristics, integration approaches, and implementation methods.

Results

Overview

This scoping review included 18 studies that met the inclusion criteria (Figure 1), including 15 (83%) quantitative and 3 (17%) mixed methods studies. The most commonly used SDM outcome measures are listed in Table 1.

Table 1. Shared decision-making (SDM) outcome measures reported in the included studies (N=18).

SDM outcome measures of included studies	Studies, n (%)
SDM measures	7 (39)
Tool use	7 (39)
Clinical outcomes	5 (28)
Patient satisfaction	4 (22)
Documentation	4 (22)
Clinician satisfaction	2 (11)

Of the 18 studies, 7 (39%) measured SDM components, such as patient empowerment, self-efficacy, self-determination, and goal setting. The 7 (39%) studies measuring use rates reported various levels of use; however, predetermined targeted use rate goals were not noted. We found that at least 3 (17%) of the studies achieved a use rate of >50% of the targeted users. In total, 4 (80%) of the 5 studies using clinical outcome measures reported positive improvements based on the SDM intervention [71,72]. Of the 18 studies, 4 (22%) studies that measured patient satisfaction included elements of communication quality, such as helpfulness, respect for the patient's opinion, understandability, confidence in the health care team, and worry reduction [72-75]. Documentation rates increased for all 4 (22%) studies using documentation as the primary outcome measure [76-79]. A total of 2 (11%) studies that measured clinician satisfaction reported clinician satisfaction at >85% and preference for a systematic approach via a tool at 100% [71,80]. In total, 4 (22%) included studies incorporated patient goals and values in the SDM process [72,78,81,82]. Consistent with previous research, 1 (5%) study reported that obstacles to tool use included a lack of awareness of the tool and low confidence in how to use the tool [83].

XSL•FO
RenderX

Goals of SDM Intervention

We categorized the included studies by macrolevel functional goals to include (1) care planning and goal setting (10/18, 56% studies), (2) prevention and screening (4/18, 22% studies), and (3) medical management and treatment (4/18, 22% studies; Table 2). The largest percentage of the care planning and goal setting category was end-of-life planning [76,77,81,82], followed by communications [72,84]. In total, 75% (3/4) of the prevention and screening studies focused on the detection of cardiovascular disease [71,74,85], and 75% (3/4) of the medication management and treatment decision-making studies focused on diabetes and related pharmacotherapy [80,83,86].

XSL•FO
RenderX

Table 2. Characteristics of the included studies.

Study	Study type	SDM[a] intervention	Clinical domain and conditions	Measures	Study outcomes
Ballard et al [83], 2017	Quantitative	SCDA[b] and DMCDA[c]	• Primary care • Diabetes and hyperlipidemia	• Decision aid tool use rates and barriers and facilitators for use	Log file data indicated that 51% of clinicians used the SCDA, and 9% of the clinicians used the DMCDA. Facilitators of decision aid use included clinicians finding them very useful and their impact on treatment decisions. Both facilitators were reported more frequently by clinicians for the SCDA than for the DMCDA (56% vs 30% and 42% vs 22%, respectively). Barriers to use included lack of knowledge of the EMR[d] link, not finding the decision aids helpful, and time constraints. The use of the tool as intended was low, with many clinicians only discussing decision aid topics that they found relevant.
Bose-Brill et al [76], 2018	Quantitative	A previsit planning tool for patients accessed via a patient portal and EHR[e]-guided point-of-care SDM questions	• Primary care • ACPf	• Documentation of ACP	In total, 19.5% (39/200) of the patients who received previsit planning responded to the framework. The intervention site had an improvement in new ACP documentation rates (P<.01) and quality (P<.01) among all eligible patients. ACP documentation rates increased by 105% (19/39 to 39/39), and quality improved among all patients who engaged in the previsit planning framework (n=39). Among eligible patients aged between 50 and 60 years at the intervention site, ACP documentation rates increased by 37% (27/96 to 37/96). ACP documentation rates increased by 34% among patients who sent >10 messages per year in the EHR (27/67 to 36/67).
Choi et al [77], 2019	Quantitative	EHR reminder to prompt discussions about unintended cardiac defibrillator shocks and deactivation options for EOL[g] care	• Intensive care • Cardiology	• Discussion rates and cardiac defibrillator deactivation rates	After the interventions, the rates of discussions regarding the deactivation of ICD[h] improved from 50% to 93% in patients in comfort care and from 32% to 70% in DNR[i] patients. The rates of deactivated ICDs improved from 45% to 73% in patients in comfort care patients and from 29% to 40% in DNR patients.
Chunchu et al [78], 2012	Mixed methods	PCCP[j] tool to promote self-management planning	• Family medicine	• Self-management plan documented in the EHR • Thematic analysis	More frequent documentation was noted in 8 problem-solving elements within the EHR, and the number of concerns for a visit was twice as high as in the control group.
Coylewright et al [85], 2020	Quantitative	iPDAs[k]	• Multiple specialties, stroke prevention in atrial fibrillation, fracture prevention in osteoarthritis, breast cancer screening, and lung cancer screening	• Clinicians' use of SDM tools and their perceptions of the tool's impact on the quality of discussion and time demands	For the 8-year analysis period, 1209 clinicians used iPDAs, with 57,116 unique patients in 81,728 visits. Of the clinicians, 76% were physicians, 16% were APPs[l], and 8% were other clinicians. There were 2607 unique patient-clinician uses in 2010; 7966 in 2014; and 24,384 in 2017. There was an average increase of 151 new clinicians per year (range 99-302). On average, 82 new clinicians used the iPDAs with at least 1 patient each year (range 56-108), with 657 clinicians for the 8-year period. In total, 54.3% of the clinicians used iPDAs with at least 5 patients.
Crosby and Gutierrez [81], 2019	Quantitative	ACP tool with risk identification for patients with late-stage heart failure	• Ambulatory care and congestive heart failure	• Use of EOL template and clinician survey	In total, 35% use of templates was reported over 3 months, 100% of templates were documented in EHR, 100% of clinic staff reported inconsistent processes for ACP discussion, and 100% of clinic staff preferred a systematic approach to initiating ACP processes.

XSL•FO
RenderX

Study	Study type	SDM[a] intervention	Clinical domain and conditions	Measures	Study outcomes
Denig et al [80], 2014	Quantitative	Diabetes decision aid	• Primary care • Type 2 diabetes	• Patient empowerment, diabetes management, and lipid-regulating prescription changes	Patient empowerment score increases were 0.1 on a 5-point scale, and drug treatment was intensified in 25% of intervention participants with increased cholesterol levels.
Fiks et al [73], 2015	Quantitative	My Asthma clinical decision support tool	• Primary care • Asthma	• My Asthma tool use and number of missed days of school or work	In total, 57% of the parents in the intervention group used My Asthma during at least 5 of the 6 study months. Parents of children with moderate to severe persistent asthma used the portal more often than others; 92% were satisfied with My Asthma. Parents reported that portal use improved their communication with the office.
Fossa et al [84], 2018	Quantitative	OpenNotes patient access for SDM	• Primary care • General practice	• A 3-item Collaborate scale, which measures a patient's experience with SDM	A total of 6913 patients responded (28% response rate). Patients reading ≥4 clinical notes in the past 12 months were 17% more likely to have top Collaborate scores when compared to those who had not read a note in the previous 12 months (response rate: 1.17, 95% CI 1.04-1.32).
Fried et al [87], 2017	Quantitative	TRIM[m]	• Primary care • Pharmacotherapy	• Patient involvement in care, patient-clinician communication, and changes in medications	In total, 29.7% of TRIM participants and 15.6% of control participants provided the highest PACIC[n] ratings; this difference was not significant. TRIM improved communication about medications and the accuracy of documentation.
Huang et al [75], 2020	Quantitative	TEA[o] tool with an annual assessment of key transition skills and setting of transition skill goals	• Pediatrics • Gastroenterology	• Health care providers' TEA tool utility ratings and recommendations to patients	All patients (N=53) performed the transition skills self-assessment and practicum and set transition goals with their health care provider. On a scale of 0 (not helpful at all) to 10 (very helpful), patients reported median (IQR) utility scores of 8 (7-10) for the transition readiness assessment, 9 (7-10) for transition resources provided, and 9 (7-10) for the medical history summary. Most (91%) patients would recommend TEA to other patients.
Jouni et al [74], 2017	Quantitative	CHD[p] genetic risk tool, decision aid for CHD risk estimates, and facilitation of SDM for statin use	• Biobank cardiology data • Genetic heart diseases	• Satisfaction with clinical visits, quality of discussion or participation in decision-making, and physician visit satisfaction	There were no statistically significant differences between the 2 groups in the SDM score, satisfaction with the clinical encounter, and perception of the quality of the discussion or participation in decision-making and physician visit satisfaction scores.
Lafata et al [79], 2014	Quantitative	5A[q] framework–guided discussion during primary care visits for CRC[r]	• Primary care • CRC screening	• Documented CRC screening rates	In total, 93% of patients received a recommendation for screening (advise), and 53% were screened in the following year. The likelihood of screening increased as the number of 5A steps increased: compared to patients whose visit contained none of the 5A steps, those whose visit contained 1 to 2 steps (OR[s] 2.96, 95% CI 1.16-7.53), and ≥3 steps (OR 4.98, 95% CI 1.84-13.44) were significantly more likely to use screening.
Litzelman et al [82], 2017	Mixed methods	GW[t] cards and discussions to identify priorities for advanced care planning	• Geriatrics • EOL care	• ACP conversations completed	A total of 86 patients' data indicated that they had engaged in a preferences-for-care process using GW R cards. The top 3 card choices by patients were attending to spirituality and religious concerns, preparing for EOL, and maintaining personal wholeness.
Plößnig et al [86], 2015	Quantitative	EMPOWER[u] SMP[v] tool to foster self-management of diabetes	• Internal medicine • Type 2 diabetes	• Tool use and empowerment measures: (1) meaningfulness, (2) self-efficacy, (3) impact, and (4) self-determination	Patients used functions for setting goals (65%) or planning activities (71%). More than half of the patients reported using the charts, the calendar, and the journals, while review (29%) and information material (24%) were of little interest. Initial testing showed an overall high level of empowerment. Improvements in self-efficacy, impact, and self-determination were observable in the German users and meaningful in the Turkish users.

XSL•FO
RenderX

Study	Study type	SDM[a] intervention	Clinical domain and conditions	Measures	Study outcomes
Schafer et al [88], 2016	Mixed methods	Pharmacist-led SDM template for medication decisions	• Outpatient • Pharmacy • Mental health	• Feasibility of template, type, and number of DTPsw	On average, 2 DTPs were identified per consultation and were most commonly related to appropriate compliance (30%), appropriate indication (26%), medication effectiveness (23%), and safety (21%).
Sona et al [72], 2020	Quantitative	Family communication tools, including education, decision-making, surrogate decision makers, and advanced directives	• Critical care	• Communication quality, helpfulness of family care conferences, confidence in health care team, impact on patient care, and reduced worry about family members	Excellent communication rating was reported by patients at 69% for RNs[x] and 54% for physicians; 43.5% of patients or families reported liking family care conferences, 52.2% thought family care conferences were helpful, 62.5% had confidence in health care team, 62.5% reported that family communication impacted care of the patient, and 62.5% reported that it reduced worry about family members.
Sperl-Hillen et al [71], 2018	Quantitative	CV Wizard, an SDM tool to reduce cardiovascular risk	• Primary care • Cardiovascular disease	• 10-year cardiovascular risk trajectory and clinician satisfaction rates	Health care provider satisfaction was 85% to 98% for various aspects of the intervention, 10-year cardiovascular risk reduction at 12 months was –2.24%, and use rates were 71% to 77% over the intervention.

[a]SDM: shared decision-making.

[b]SCDA: statin choice decision aid.

[c]DMCDA: diabetes medication choice decision aid.

[d]EMR: electronic medical record.

[e]EHR: electronic health record.

[f]ACP: advanced care planning.

[g]EOL: end of life.

[h]ICD: implantable cardioverter defibrillator.

[i]DNR: do not resuscitate.

[j]PCCP: patient-centered care plan.

[k]iPDA: integrated patient decision aid.

[l]APP: advanced practice provider.

[m]TRIM: tool to reduce inappropriate medication.

[n]PACIC: patient assessment of chronic illness care.

[o]TEA: transition electronic health record activity.

[p]CHD: coronary heart disease.

[q]5A: assess, advise, agree, assist, and arrange.

[r]CRC: colorectal cancer.

[s]OR: odds ratio.

[t]GW: Go Wish.

[u]EMPOWER: Support of Patient Empowerment by an intelligent self-management pathway for patients.

[v]SMP: self-management pathway.

[w]DTP: drug therapy problem.

[x]RN: registered nurse.

Tool objectives were generally a direct function of the use of the SDM tool; for example, an end-of-life planning template may have a stated objective to increase documentation rates for communications. Patient satisfaction–related outcomes, such as decision satisfaction because of a shared conversation, were not noted in 11 (61%) of the 18 studies [72,73,75,80-84,86-88].

Promoting SDM Awareness and Adoption

EHR and Workflow Integration

Our study found that workflow integration approaches were highly variable, ranging from scanning paper-based tools into the EHR system to implementing SDM tools with full semantically interoperable data elements surfacing within patient portals and in the clinicians' EHR workflow (Figure 2 and Table 3).

XSL•FO
RenderX

Figure 2. Electronic health record (EHR) integration methods described in the included studies. SDM: shared decision-making.

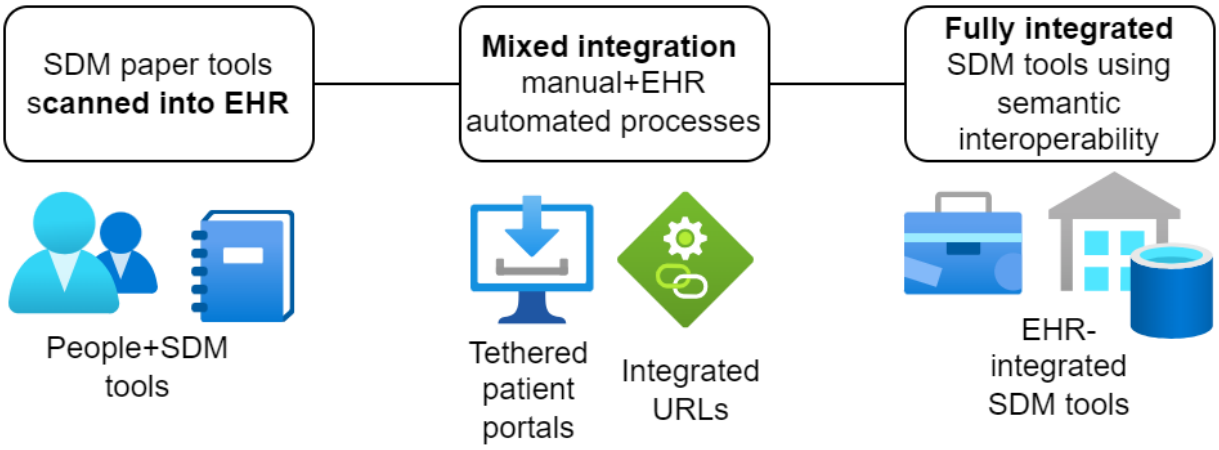

XSL•FO
RenderX

Table 3. Shared decision-making (SDM) interventions and goals, electronic health record (EHR) integration methods, and implementation approaches (N=18).

Functional goals and studies	Clinical processes	SDM interventions	SDM goals	EHR integration method and implementation approach
Care planning and goal setting (n=10, 56%)				
Bose-Brill et al [76], 2018	ACP[a]	A previsit planning tool for patients accessed via a patient portal and EHR-guided point-of-care SDM questions	To increase the rate of ACP documentation	• An EHR-tethered patient portal presents ACP questions before a visit. The algorithm within the EHR guides the communication between clinician and patient. • Physicians, nurses, and other clinical staff at the intervention clinic were collectively involved in developing this ACP previsit planning algorithm that was rolled out practice wide over a 3-month period.
Choi et al [77], 2019	End-of-life planning	EHR clinician alerts and reminders to prompt discussions regarding unintended cardiac defibrillator shocks and deactivation options during end-of-life care	ICD[b] deactivation discussion	• EHR alerts and decision support tools were integrated into the comfort care order set and DNRc document. • A 10-min training with pre- and postknowledge assessment, posttraining surveys with follow-up education, and special training sessions for various specialties were provided, as needed.
Chunchu et al [78], 2012	PCCP[d]	PCCP tool to promote self-management planning	To increase collaborative self-management planning	• PCCP was developed in EHR with prompts. • Clinician training with video demonstration was provided.
Crosby and Gutierrez [81], 2019	ACP	ACP tool with risk identification for patients with late-stage heart failure	Risk stratification, life expectancy of <3 years, conversations about preferences, and scanning into EHR	• ACP tool was scanned into the EHR. • Readiness to change assessment tool was administered to all clinic staff and internal stakeholders. Education was provided about the ACP tool and project goals. Key stakeholders were involved in tool development and project implementation, workflow assessment, and process development.
Fossa et al [84], 2018	Communication	OpenNotes patient access for SDM	Collaborative decision-making	• OpenNotes access was provided via the patient portal.
Fried et al [87], 2017	Medication reconciliation	TRIM[e]	To facilitate clinician and patient medication communications	• TRIM was fully embedded into the EHR and extracted patient data from the EHR for medication reconciliation.
Huang et al [75], 2020	Transition readiness	TEA[f] tool with annual assessment of key transition skills and setting of transition skill goals	Transition readiness decision-making	• The TEA tool was developed fully in the EHR. • Pilot-testing of the TEA tool was performed at an independent clinic.
Litzelman et al [82], 2017	End-of-life planning or ACP	Go Wish cards and discussions to identify priorities for ACP	ACP conversations	• Go Wish cards+ACP discussion took place with EHR data entry of goals and values. • Training was provided on ACP and Go Wish R cards through workshops, simulation sessions, case conferences, and guidance from the CDSg tool.
Schafer et al [88], 2016	Medication reconciliation	Pharmacist-led SDM template for medication decisions	To promote patient and pharmacist SDM with the use of the Ottawa Personal Decision Guide	• Pharmacy SDM conversation template was integrated into EHR and printed for the patient.
Sona et al [72], 2020	Communication	Family communication tools, including education, decision-making, surrogate decision makers, and advanced directives	To increase the rates of critical care family communications	• The communication bundle document template was developed in the EHR system.
Prevention and screening (n=4, 22%)				
Coylewright et al [85], 2020	Risk assessment and prevention	iPDAs[h]	To promote clinician use of iPDAs with patients for SDM	• A suite of decision aids was integrated into the EHR with use tracking. • A clinician adopter survey was conducted.

XSL•FO
RenderX

Functional goals and studies	Clinical processes	SDM interventions	SDM goals	EHR integration method and implementation approach
Jouni et al [74], 2017	Risk assessment and prevention	CHD[i] genetic risk tool; decision aid for CHD risk estimates and facilitation of SDM for statin use	Visual comprehension of CHD risk and associated risk reduction with statins	• GDMSj, a web-based tool, was integrated using a web viewer into the EHR system.
Lafata et al [79], 2014	Risk assessment and prevention	5A[k] framework–guided discussion during primary care office visits for CRC[l] screening	Use of 5As during clinician-patient discussions to support SDM	• Previsit patient survey and clinic visit discussion were guided by EHR prompts.
Sperl-Hillen et al [71], 2018	Risk assessment and prevention	CV Wizard: a point-of-care CDS tool to facilitate discussion and SDM to reduce cardiovascular risk	Treatment options discussion per printed document	• A CDS tool per an EHR-integrated web service
Medical management and treatment (n=4, 22%)				
Ballard et al [83], 2017	Pharmacotherapy	SCDA[m] and DMCDA[n]	SDM for cholesterol reduction treatment options and SDM for diabetes treatment options	• EHR embedded links were sent to decision aids with use tracking.
Denig et al [80], 2014	Pharmacotherapy	Diabetes decision aid	Treatment options discussion	• Decision aid software was integrated into the EHR and populated with individual patient data. • Pre- and postclinic visit surveys and health care provider interviews were conducted.
Fiks et al [73], 2015	Treatment compliance	My Asthma CDS tool	Patient-reported data collection via the My Asthma tool on the patient portal	• My Asthma was fully integrated into the EHR for both patients and health care providers. • All 3 primary care practices received in-person training on the portal from a physician leader.
Plößnig et al [86], 2015	Pharmacotherapy	EMPOWER[o] SMP[p] tool to foster self-management of diabetes	Self-management of diabetes	• EMPOWER SMP was integrated using ISOq 13606 and openEHR information models to support full semantic interoperability.

[a]ACP: advanced care planning.

[b]ICD: implantable cardioverter defibrillator.

[c]DNR: do not resuscitate.

[d]PCCP: patient-centered care plan.

[e]TRIM: tool to reduce inappropriate medication.

[f]TEA: transition electronic health record activity.

[g]CDS: clinical decision support.

[h]iPDA: integrated patient decision aid.

[i]CHD: coronary heart disease.

[j]GDMS: generic disease management system.

[k]5A: assess, advise, agree, assist, and arrange.

[l]CRC: colorectal cancer.

[m]SCDA: statin choice decision aid.

[n]DMCDA: diabetes medication choice decision aid.

[o]EMPOWER: Support of Patient Empowerment by an intelligent self-management pathway for patients.

[p]SMP: self-management pathway.

[q]ISO: International Organization for Standardization.

We found that 8 (44%) of the 18 included studies described implementation strategies in sufficient detail to define the approach to operationalizing the SDM processes [71,73,75,77,78,80,82,85]. In 2 (11%) studies, patient portals were leveraged as part of the SDM process [73,84]. One patient portal was described as a tethered patient portal, which referred to SDM processes that were partially embedded into the clinicians' EHR view. This was accomplished by collecting data and then saving them into the patient's record so that a clinician could view the data from their workflow screen [76]. One intervention simply scanned paper documents and uploaded them into the patient record or other sections predefined within the EHR [81].

XSL•FO
RenderX

Some (5/18, 28%) of the studies noted hybrid approaches to EHR integration, such as a combination of manual data collection with EHR documentation of an SDM process (Table 3). Of the 18 studies, 2 (11%) noted the integration of a URL into the EHR that launched a third-party web-based SDM tool [83,85]. Other (12/18, 67%) studies included health care provider alerts that presented patient risk factors to facilitate discussion and decision-making. Health care provider alerts may also help capture documentation of the SDM interaction [74,78].

Determining when to engage a patient in SDM before a clinical encounter is another element of SDM intervention planning. In 1 (6%) study, the engagement interval with patients was extended to previsit preparatory activities [76]. This included SDM processes that initially engaged the patient through preparatory SDM tools or surveys that were delivered via email, patient portals, or web links or delivered in the waiting room immediately before the clinical encounter [73,74,76,79,83-85,87]. Other SDM interventions were designed to be used during the clinical encounter, including EHR-based tools, patient record sharing, communication and conversation guides, patient health plans, and educational printouts for patients [71,72,75,77,78,80-82,86,88]. None of the included studies investigated postclinical encounter SDM tools for ongoing decision affirmation and goal alignment.

Implementation Strategies

A total of 9 (50%) of the 18 included studies described implementation strategies in sufficient detail to define the approach to operationalizing the SDM processes [71,73,75,77,78,80-82,85]. Methods to make clinicians and patients aware of the tools included email notifications and reminders, links in the EHR to an external third-party SDM tool, paper-based posters displayed in common workspaces, educational presentations, and training (Table 3). Trainings were often conducted in person, but some used prerecorded video. Brief trainings focused on how to use the tool; however, longer trainings included skills for engaging in SDM and sometimes involved role-playing these behaviors. In one case, the required implementation workflow caused users to question the “long-term feasibility” of the SDM intervention due to the additional time and resources required to complete the SDM processes, specifically in the case of advanced care and end-of-life planning for patients with heart failure [81].

Clinical adoption was supported by collaboration with physician and nursing leaders for integrating the tools into the clinical workflow, engaging clinicians who initially engaged patients when starting the SDM process, providing feedback at the health care provider or clinic level, and sponsoring lotteries for clinic leadership when tool use achieved a specified level.

Other implementation strategies used to support patient engagement with SDM tools included email reminders to complete surveys, the provision of kiosks in the waiting room to allow patients to access the patient portal, and the provision of paper-based information for patients to read while waiting in the examination room.

We found that there was variance across the targeted users for the SDM tool interaction. In several (7/18, 39%) studies, both patients and clinicians had roles to play within the SDM processes [73,75,76,78,82,84,86]. In most cases, the tool was targeted toward the clinician to activate the process and conduct conversations with the patient [71,72,74,77,79-81,83,85,87,88].

EHR Clinical Decision Support Features

The SDM interventions identified in this review contained several clinical decision support features. Table 4 describes each of these approaches, along with indicating whether the component was delivered manually or automatically within the EHR system. These features include the following: (1) patient risk alerts, which use patient data to compute personalized risk; (2) patient data display through the retrieval of relevant patient information into an informative display; (3) health care provider notification indicating whether a health care provider has to seek out the SDM tool or whether it surfaces within the EHR automatically based on a set of criteria; (4) presented in the EHR, allowing clinicians to access an SDM tool within the EHR (through several different technical approaches), as opposed to an application that is external to the EHR system; and (5) documentation support, including manual, autogenerated, or template-based clinician notes within an EHR to document that an SDM event occurred. One study used an emoji in the EHR to indicate that the SDM conversation had occurred [72].

XSL•FO
RenderX

Table 4. Electronic health record (EHR) clinical decision support features by study (N=18).

Studies	Patient risk calculator+alert		Patient data display		Health care provider notification		Presented in EHR		Documentation		
	Manual (n=6, 33%)	EHR (n=11, 61%)	Manual (n=7, 39%)	EHR (n=13, 72%)	Manual (n=6, 33%)	EHR alert (n=12, 67%)	No (n=3, 17%)	Yes (n=14, 78%)	Manual (n=10, 56%)	Automatic (n=5, 28%)	Template (n=4, 22%)
Ballard et al [83], 2017		✓		✓		✓		✓	✓		
Bose-Brill et al [76], 2018	✓		✓		✓			✓	✓		
Choi et al [77], 2019		✓		✓		✓		✓	✓		
Chunchu et al [78], 2012	✓		✓		✓			✓	✓		✓
Coylewright et al [85], 2020		✓		✓		✓		✓		✓	
Crosby and Gutierrez [81], 2019		✓	✓		✓		✓		✓		✓
Denig et al [80], 2014	✓			✓		✓		✓	✓		
Fiks et al [73], 2015		✓		✓		✓		✓	✓		
Fossa et al [84], 2018				✓							
Fried et al [87], 2017	✓		✓	✓	✓	✓	✓				
Huang et al [75], 2020		✓	✓	✓		✓		✓		✓	
Jouni et al [74], 2017		✓		✓		✓		✓		✓	
Lafata et al [79], 2014		✓		✓		✓		✓			
Litzelman et al [82], 2017	✓		✓		✓			✓	✓		
Plößnig et al [86], 2015		✓		✓		✓		✓	✓		
Schafer et al [88], 2016		✓		✓		✓		✓	✓		✓
Sona et al [72], 2020	✓		✓		✓		✓		✓		✓
Sperl-Hillen et al [71], 2018		✓		✓		✓		✓		✓	

Discussion

Principal Findings

This review summarizes the findings of 18 studies that investigated approaches to the integration and implementation of SDM tools within EHR systems. Prominent aspects found across studies included functional SDM goals, such as communication, documentation of communication and discussions, and SDM template use. We identified only 4 (22%) randomized controlled trials (RCTs), which may be a reflection of the operational nature of the study measures. In 10 (56%) studies, clinicians were the primary targeted users, whereas 4 (22%) studies targeted patients and 5 (28%) targeted both. Integrating SDM interventions into a clinician's workflow within an EHR system appears to support improved SDM uptake [60,89-91]. Studies reported multiple EHR and workflow integration approaches, which may be an indication that a dynamic model that is scalable could help guide the adoption of SDM tools and processes across different environments. Future research to understand the utility of adjusting SDM model depth and complexity based on the magnitude of the decision,

XSL•FO
RenderX

the complexity of the clinical situation, and the patient preferences is warranted.

Goals of SDM Tools

Most (10/18, 56%) of the included studies focused on SDM tools that were functionally designed to support care planning and goal setting. We also noted 1 (6%) study that used the concept of open medical notes as a precursor to shared care conversations.

Most (4/18, 22%) prevention and screening studies targeted by SDM focused on common conditions, such as the detection of heart disease, colorectal cancer, and other common conditions. Future opportunities for EHR-integrated SDM screening tools include the assessment of social determinants of health for the purpose of supporting individuals with social needs. Furthermore, mental health SDM screening tools offer an important area for prevention and screening to detect anxiety, depression, loneliness, addiction, and suicide risk, as a few examples.

Promoting SDM Awareness and Adoption

Although studies generally lacked sufficient technical detail on the actual integration method into each EHR system, we found considerable variation in how SDM tools are implemented within EHR workflows. The implementation approaches for SDM applications showed varying levels of planning and efforts to promote SDM intervention awareness. For example, 33% (6/18) of the studies did not describe implementation activities for clinicians, whereas other (6/18, 33%) studies described more detailed implementation plans that included training, demonstrations, knowledge queries, and incentives [71,73,77,78,81,82].

In total, 22% (4/18) of the studies included in the review suggested that integrating an SDM intervention into a clinician's EHR workflow may support increased awareness of the intervention [60,72,89,90]. This represents an important step toward supporting clinicians in the adoption and use of SDM tools and processes. Creating awareness, interest, and commitment to SDM tool use is often overlooked and may contribute to waning adoption for otherwise well-designed SDM tools, potentially leading to adoption failure. The implementation plan is a critical component to moving people and the organizational culture toward the adoption of SDM tools. These steps also provide a forum for direct feedback should clinicians lack confidence in the accuracy or efficacy of an SDM tool, as was noted by Ballard et al [83].

This review found that established processes and required patient interactions, such as end-of-life care planning protocols, appear to support greater SDM adoption success because they rely on workflows and processes that are already in practice. Adoption through incremental change, such as an evolutionary approach rather than a revolution, may be more likely to gain momentum over time [85].

Implications for Future Research

Our review found that only a few randomized clinical studies were noted. Future research studies should assess the effect of SDM in multisite pragmatic RCTs.

This review identified a wide range of approaches to how SDM interventions are organized and delivered within the clinical setting. Future research could explore contextually aware tiered levels of intervention. This is important to standardize SDM care processes within specialties in order to support large-scale adoption of validated SDM tools and methods while addressing practical considerations within the constraints of each specialty.

Our review found that 17% (3/18) of the studies attempted to engage patients before a clinical encounter; however, no studies attempted to engage patients in decision-making after a clinical encounter [76,79,84]. Protracted decisions that include a series of small decisions align well with the use of patient portals to begin preparing patients for the SDM process. Another area for future research could be to understand the efficacy of postclinical encounter engagement to address the ongoing maintenance and reinforcement of decisions that have been made in collaboration with the clinician.

We found limited studies that included the identification and measurement of patient values. Thus, patient involvement in establishing their values and alignment of care with those values appear to be areas for additional focus in future research. Involving patients in goal setting is a key element to support patient engagement, which is essential to help patients manage their decisions outside of the clinician encounter [92].

Strengths and Limitations of Approach

Strengths

The review followed rigorous standard scoping review methods set by the JBI, Arksey and O'Malley [59], and others [60,66,93,94]. We carried out a systematic search of multiple databases developed by a medical librarian and independent article screening and data extraction by at least 2 reviewers, with disagreement resolved through consensus. Our research team included investigators from an academic medical setting with extensive research experience in biomedical informatics, cognitive psychology, human factors, and health care.

Limitations

This scoping review may have missed relevant studies that are not indexed by the databases searched. Furthermore, the search retrieved studies published between January 1, 2009, and January 11, 2021; therefore, studies published outside this time frame were not included in our analysis. We did not search the gray literature or unpublished data sources. SDM studies that lacked an SDM tool, EHR implementation, and an actual clinical setting were not included. Implementation details were highly variable in terms of the approach and the level of detail described in the studies. SDM definitions and interpretations were variable, and we found limited references to validated SDM definitions, such as commonly cited definitions by the US National Institutes of Health or other credible sources.

Conclusions

This scoping review revealed a wide range of approaches to the integration and implementation of SDM tools within an EHR workflow. We found that SDM tool integration methods ranged from simple scan and save methods to fully integrated and operable technical integration with an EHR system. Moreover,

XSL•FO
RenderX

implementation approaches were highly variable, ranging from sending an email to clinicians to create awareness to more complete programs involving demonstrations, training, incentives, clinician-led promotion, and feedback sessions. Areas for future research include more RCTs, the development of contextually aware SDM interventions, longitudinal patient engagement with pre- and postvisit interventions, the study of patient-driven goals and values, and the subsequent impact on patient satisfaction levels. Future studies should detail pragmatic and actionable integration and implementation steps to advance the uptake of SDM.

Acknowledgments

The authors would like to acknowledge Hitachi, Ltd (Tokyo, Japan) for partial funding of this project. The funding source played no role in the design, execution, or findings of this study.

During the preparation of this study, the authors used Covidence systematic review software (Veritas Health Innovation) [95] for the screening of studies and EndNote (Clarivate) [96] for management, citation, and storage of references. Following and concurrent with the use of these tools, the authors reviewed, synthesized, and organized the scoping review content. Figure 2 was created using Drawio app (JGraph Ltd) [97].

The views expressed in this paper are those of the authors and do not necessarily represent the position or policy of the US Department of Veterans Affairs or the US government.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KK, GDF, and CW obtained funding for this project. CW and KK conceptualized the research and recommended the research methodology. JHP led the project administration, with supervision by CW, JMB, and GDF. MMM conducted all data and literature searches. WR II, JHP, CW, JMB, and TT conducted data extraction and synthesis for the final studies. JHP, CW, and TT wrote the original draft of the manuscript. JHP created Figures 1 and 2 and Tables 1-3. CW and JHP created Table 4. All authors reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search terms.
[DOCX File , 20 KB - jmir_v27i1e59956_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 151 KB - jmir_v27i1e59956_app2.pdf]

References

1. El-Ghazali S, Khan M, Santhirapala R. Shared decision making. In: Perioperative Quality Improvement. Amsterdam, The Netherlands: Elsevier; 2023.
2. United States. President's commission for the study of ethical problems in medicine and biomedical and behavioral research. U S Code Annot U S 1982;Title 42 Sect. 300v as added 1978. [Medline: 12041401]
3. Alston C, Berger Z, Brownlee S, Elwyn G, Fowler FJJ, Hall LK, et al. Shared decision-making strategies for best care: patient decision aids. National Academy of Medicine. 2014 Sep 18. URL: https://nam.edu/perspectives-2014-shared-decision-making-strategies-for-best-care-patient-decision-aids/ [accessed 2021-01-11]
4. McCarron TL, Moffat K, Wilkinson G, Zelinsky S, Boyd JM, White D, et al. Understanding patient engagement in health system decision-making: a co-designed scoping review. Syst Rev 2019 Apr 18;8(1):97 [FREE Full text] [doi: 10.1186/s13643-019-0994-8] [Medline: 30999950]
5. Hibbard JH, Greene J, Tusler M. Improving the outcomes of disease management by tailoring care to the patient's level of activation. Am J Manag Care 2009 Jun;15(6):353-360 [FREE Full text] [Medline: 19514801]
6. Vaseur RM, Te Braake E, Beinema T, Nijeweme- d'Hollosy WO, Tabak M. Technology-supported shared decision-making in chronic conditions: a systematic review of randomized controlled trials. Patient Educ Couns 2024 Jul;124:108267 [FREE Full text] [doi: 10.1016/j.pec.2024.108267] [Medline: 38547638]

XSL•FO
RenderX

7. Kuo AM, Thavalathil B, Elwyn G, Nemeth Z, Dang S. The promise of electronic health records to promote shared decision making: a narrative review and a look ahead. Med Decis Making 2018 Nov;38(8):1040-1045. [doi: 10.1177/0272989X18796223] [Medline: 30226100]
8. Wickramasekera N, Taylor SK, Lumley E, Gray T, Wilson E, Radley S. Can electronic assessment tools improve the process of shared decision-making? A systematic review. Health Inf Manag 2023 May;52(2):72-86 [FREE Full text] [doi: 10.1177/1833358320954385] [Medline: 33016126]
9. US Preventive Services Task Force, Davidson KW, Mangione CM, Barry MJ, Nicholson WK, Cabana MD, et al. Collaboration and shared decision-making between patients and clinicians in preventive health care decisions and US preventive services task force recommendations. JAMA 2022 Mar 22;327(12):1171-1176. [doi: 10.1001/jama.2022.3267] [Medline: 35315879]
10. Medicare payment. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/cms-guide-medical-technology-companies-and-other-interested-parties/payment#:~:text=Medicare%20is%20the%20single%20largest,services%20in%20the%20United%20States [accessed 2024-08-06]
11. Merchant FM, Dickert NWJ, Howard DH. Mandatory shared decision making by the Centers for Medicare and Medicaid Services for cardiovascular procedures and other tests. JAMA 2018 Aug 21;320(7):641-642. [doi: 10.1001/jama.2018.6617] [Medline: 29868828]
12. Lu Y, Elwyn G, Moulton BW, Volk RJ, Frosch DL, Spatz ES. Shared decision-making in the U.S.: evidence exists, but implementation science must now inform policy for real change to occur. Z Evid Fortbild Qual Gesundhwes 2022 Jun;171:144-149. [doi: 10.1016/j.zefq.2022.04.031] [Medline: 35599230]
13. Clinical decision support. Assistant Secretary for Technology Policy. URL: https://www.healthit.gov/topic/safety/clinical-decision-support [accessed 2021-01-11]
14. Becher S, Holzhüter F, Heres S, Hamann J. Barriers and facilitators of shared decision making in acutely ill inpatients with schizophrenia-qualitative findings from the intervention group of a randomised-controlled trial. Health Expect 2021 Oct;24(5):1737-1746 [FREE Full text] [doi: 10.1111/hex.13313] [Medline: 34258833]
15. Légaré F, Witteman HO. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. Health Aff (Millwood) 2013 Feb;32(2):276-284. [doi: 10.1377/hlthaff.2012.1078] [Medline: 23381520]
16. Légaré F, Ratté S, Gravel K, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. Patient Educ Couns 2008 Dec;73(3):526-535. [doi: 10.1016/j.pec.2008.07.018] [Medline: 18752915]
17. Politi MC, Neuman MD. Understanding when, why and how shared decision making is used in routine care. Patient Educ Couns 2023 Jan;106:1-2. [doi: 10.1016/j.pec.2022.10.352] [Medline: 36333194]
18. Aronson PL, Schaeffer P, Ponce KA, Gainey TK, Politi MC, Fraenkel L, et al. Stakeholder perspectives on hospitalization decisions and shared decision-making in bronchiolitis. Hosp Pediatr 2022 May 01;12(5):473-482 [FREE Full text] [doi: 10.1542/hpeds.2021-006475] [Medline: 35441213]
19. Scholl I, LaRussa A, Hahlweg P, Kobrin S, Elwyn G. Organizational- and system-level characteristics that influence implementation of shared decision-making and strategies to address them - a scoping review. Implement Sci 2018 Mar 09;13(1):40 [FREE Full text] [doi: 10.1186/s13012-018-0731-z] [Medline: 29523167]
20. Cassidy BP, Getchell LE, Harwood L, Hemmett J, Moist LM. Barriers to education and shared decision making in the chronic kidney disease population: a narrative review. Can J Kidney Health Dis 2018 Nov 2;5:2054358118803322 [FREE Full text] [doi: 10.1177/2054358118803322] [Medline: 30542621]
21. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
22. O'Connor AM, Llewellyn-Thomas HA, Flood AB. Modifying unwarranted variations in health care: shared decision making using patient decision aids. Health Aff (Millwood) 2004;Suppl Variation:VAR63-VAR72. [doi: 10.1377/hlthaff.var.63] [Medline: 15471770]
23. Coronado-Vázquez V, Canet-Fajas C, Delgado-Marroquín MT, Magallón-Botaya R, Romero-Martín M, Gómez-Salgado J. Interventions to facilitate shared decision-making using decision aids with patients in primary health care: a systematic review. Medicine (Baltimore) 2020 Aug 07;99(32):e21389 [FREE Full text] [doi: 10.1097/MD.0000000000021389] [Medline: 32769870]
24. Ankolekar A, Dekker A, Fijten R, Berlanga A. The benefits and challenges of using patient decision aids to support shared decision making in health care. JCO Clin Cancer Inform 2018 Dec;2:1-10. [doi: 10.1200/CCI.18.00013] [Medline: 30652607]
25. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
26. Hoefel L, O'Connor AM, Lewis KB, Boland L, Sikora L, Hu J, et al. 20th anniversary update of the Ottawa decision support framework part 1: a systematic review of the decisional needs of people making health or social decisions. Med Decis Making 2020 Jul;40(5):555-581. [doi: 10.1177/0272989X20936209] [Medline: 32659154]

XSL•FO
RenderX

27. Hoefel L, Lewis KB, O'Connor A, Stacey D. 20th anniversary update of the Ottawa decision support framework: part 2 subanalysis of a systematic review of patient decision aids. Med Decis Making 2020 May;40(4):522-539. [doi: 10.1177/0272989X20924645] [Medline: 32522091]
28. Stacey D, Légaré F, Boland L, Lewis KB, Loiselle MC, Hoefel L, et al. 20th anniversary Ottawa decision support framework: part 3 overview of systematic reviews and updated framework. Med Decis Making 2020 Apr;40(3):379-398. [doi: 10.1177/0272989X20911870] [Medline: 32428429]
29. Drug and Therapeutics Bulletin. An introduction to patient decision aids. BMJ 2013 Jul 23;347:f4147. [doi: 10.1136/bmj.f4147] [Medline: 23881944]
30. Lenert L, Dunlea R, Del Fiol G, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014 Nov;34(8):987-995 [FREE Full text] [doi: 10.1177/0272989X14550102] [Medline: 25224366]
31. Gillum RF. From papyrus to the electronic tablet: a brief history of the clinical medical record with lessons for the digital age. Am J Med 2013 Oct;126(10):853-857. [doi: 10.1016/j.amjmed.2013.03.024] [Medline: 24054954]
32. Hofstede SN, Marang-van de Mheen PJ, Wentink MM, Stiggelbout AM, Vleggeert-Lankamp CL, Vliet Vlieland TP, et al. Barriers and facilitators to implement shared decision making in multidisciplinary sciatica care: a qualitative study. Implement Sci 2013 Aug 23;8:95 [FREE Full text] [doi: 10.1186/1748-5908-8-95] [Medline: 23968140]
33. Staszewska A, Zaki P, Lee J. Computerized decision aids for shared decision making in serious illness: systematic review. JMIR Med Inform 2017 Oct 06;5(4):e36 [FREE Full text] [doi: 10.2196/medinform.6405] [Medline: 28986341]
34. Lee CN, Sullivan J, Foraker R, Myckatyn TM, Olsen MA, Phommasathit C, et al. Integrating a patient decision aid into the electronic health record: a case report on the implementation of BREASTChoice at 2 sites. MDM Policy Pract 2022 Oct 08;7(2):23814683221131317 [FREE Full text] [doi: 10.1177/23814683221131317] [Medline: 36225966]
35. Scalia P, Ahmad F, Schubbe D, Forcino R, Durand MA, Barr PJ, et al. Integrating option grid patient decision aids in the epic electronic health record: case study at 5 health systems. J Med Internet Res 2021 May 03;23(5):e22766 [FREE Full text] [doi: 10.2196/22766] [Medline: 33938806]
36. Westerink HJ, Oirbans T, Garvelink MM, van Uden-Kraan CF, Zouitni O, Bart HA, et al. Barriers and facilitators of meaningful patient participation at the collective level in healthcare organizations: a systematic review. Health Policy 2023 Dec;138:104946 [FREE Full text] [doi: 10.1016/j.healthpol.2023.104946] [Medline: 38000333]
37. Tong WT, Ng CJ, Lee YK, Lee PY. What will make patients use a patient decision aid? A qualitative study on patients' perspectives on implementation barriers and facilitators. J Eval Clin Pract 2020 Jun;26(3):755-764. [doi: 10.1111/jep.13161] [Medline: 31115132]
38. Pekala KR, Shill DK, Austria M, Langford AT, Loeb S, Carlsson SV. Shared decision-making before prostate cancer screening decisions. Nat Rev Urol 2024 Jun;21(6):329-338. [doi: 10.1038/s41585-023-00840-0] [Medline: 38168921]
39. Ke Y, Zhou H, Chan RJ, Chan A. Decision aids for cancer survivors' engagement with survivorship care services after primary treatment: a systematic review. J Cancer Surviv 2024 Apr;18(2):288-317 [FREE Full text] [doi: 10.1007/s11764-022-01230-y] [Medline: 35798994]
40. Mhaimeed N, Mhaimeed N, Mhaimeed O, Alanni J, Burney Z, Elshafeey A, et al. Shared decision making with black patients: a scoping review. Patient Educ Couns 2023 May;110:107646 [FREE Full text] [doi: 10.1016/j.pec.2023.107646] [Medline: 36739706]
41. Jones JB, Shah NR, Bruce CA, Stewart WF. Meaningful use in practice using patient-specific risk in an electronic health record for shared decision making. Am J Prev Med 2011 May;40(5 Suppl 2):S179-S186. [doi: 10.1016/j.amepre.2011.01.017] [Medline: 21521593]
42. Schubbe D, Yen RW, Saunders CH, Elwyn G, Forcino RC, O'Malley AJ, et al. Implementation and sustainability factors of two early-stage breast cancer conversation aids in diverse practices. Implement Sci 2021 May 10;16(1):51 [FREE Full text] [doi: 10.1186/s13012-021-01115-1] [Medline: 33971913]
43. Branda ME, Ridgeway JL, Mann D, Wieser J, Gomez Y, Dagoberg A, et al. Healthcare systems collaborating to implement a shared decision-making tool in the electronic health record and build evidence on its adoption and use. Learn Health Syst 2024 Apr 15;8(Suppl 1):e10418 [FREE Full text] [doi: 10.1002/lrh2.10418] [Medline: 38883873]
44. Mohindra NA, Garcia SF, Kircher S, Barnard C, Perry LM, Lyleroehr M, et al. Development of an electronic health record-integrated patient-reported outcome-based shared decision-making dashboard in oncology. JAMIA Open 2024 Jul 24;7(3):ooae056. [doi: 10.1093/jamiaopen/ooae056] [Medline: 39049991]
45. Scalable decision support and shared decision making for lung cancer screening. Agency for Healthcare Research and Quality. URL: https://digital.ahrq.gov/ahrq-funded-projects/scalable-decision-support-and-shared-decisionmaking-lung-cancer-screening [accessed 2025-02-12]
46. Lowery J, Fagerlin A, Larkin AR, Wiener RS, Skurla SE, Caverly TJ. Implementation of a web-based tool for shared decision-making in lung cancer screening: mixed methods quality improvement evaluation. JMIR Hum Factors 2022 Apr 01;9(2):e32399 [FREE Full text] [doi: 10.2196/32399] [Medline: 35363144]
47. Thompson-Leduc P, Clayman ML, Turcotte S, Légaré F. Shared decision-making behaviours in health professionals: a systematic review of studies based on the theory of planned behaviour. Health Expect 2015 Oct;18(5):754-774 [FREE Full text] [doi: 10.1111/hex.12176] [Medline: 24528502]

XSL•FO
RenderX

48. Shinkunas LA, Klipowicz CJ, Carlisle EM. Shared decision making in surgery: a scoping review of patient and surgeon preferences. BMC Med Inform Decis Mak 2020 Aug 12;20(1):190 [FREE Full text] [doi: 10.1186/s12911-020-01211-0] [Medline: 32787950]
49. Raj R, Ahuja KD, Frandsen M, Jose M. Older patient considering treatment for advanced renal disease: protocol for a scoping review of the information available for shared decision-making. BMJ Open 2016 Dec 08;6(12):e013755 [FREE Full text] [doi: 10.1136/bmjopen-2016-013755] [Medline: 27932341]
50. Norful AA, Dillon J, Baik D, George M, Ye S, Poghosyan L. Instruments to measure shared decision-making in outpatient chronic care: a systematic review and appraisal. J Clin Epidemiol 2020 May;121:15-19 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.01.001] [Medline: 31904466]
51. Niburski K, Guadagno E, Mohtashami S, Poenaru D. Shared decision making in surgery: a scoping review of the literature. Health Expect 2020 Oct;23(5):1241-1249 [FREE Full text] [doi: 10.1111/hex.13105] [Medline: 32700367]
52. Doherr H, Christalle E, Kriston L, Härter M, Scholl I. Use of the 9-item shared decision making questionnaire (SDM-Q-9 and SDM-Q-Doc) in intervention studies-a systematic review. PLoS One 2017 Mar 30;12(3):e0173904 [FREE Full text] [doi: 10.1371/journal.pone.0173904] [Medline: 28358864]
53. Corell A, Guo A, Vecchio TG, Ozanne A, Jakola AS. Shared decision-making in neurosurgery: a scoping review. Acta Neurochir (Wien) 2021 Sep;163(9):2371-2382 [FREE Full text] [doi: 10.1007/s00701-021-04867-3] [Medline: 33942189]
54. Bouniols N, Leclère B, Moret L. Evaluating the quality of shared decision making during the patient-carer encounter: a systematic review of tools. BMC Res Notes 2016 Aug 02;9:382 [FREE Full text] [doi: 10.1186/s13104-016-2164-6] [Medline: 27485434]
55. Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, Pieterse AH. Key components of shared decision making models: a systematic review. BMJ Open 2019 Dec 17;9(12):e031763 [FREE Full text] [doi: 10.1136/bmjopen-2019-031763] [Medline: 31852700]
56. Baik D, Cho H, Masterson Creber RM. Examining interventions designed to support shared decision making and subsequent patient outcomes in palliative care: a systematic review of the literature. Am J Hosp Palliat Care 2019 Jan;36(1):76-88 [FREE Full text] [doi: 10.1177/1049909118783688] [Medline: 29925244]
57. Ahmad M, Abu Tabar N, Othman EH, Abdelrahim Z. Shared decision-making measures: a systematic review. Qual Manag Health Care 2020;29(2):54-66. [doi: 10.1097/QMH.0000000000000250] [Medline: 32224789]
58. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2015 Jan;35(1):114-131 [FREE Full text] [doi: 10.1177/0272989X14551638] [Medline: 25351843]
59. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
60. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
61. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
62. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
63. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
64. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
65. Health IT legislation. Assistant Secretary for Technology Policy. URL: https://www.healthit.gov/topic/laws-regulation-and-policy/health-it-legislation [accessed 2021-01-11]
66. Peters M, Godfrey CM, Mcinerney P, Munn Z, Trico A, Khalil H. Chapter 11: Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; Jun 2020.
67. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
68. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. CiteSeerX. 2006 Apr. URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ed8b23836338f6fdea0cc55e161b0fc5805f9e27 [accessed 2021-01-11]
69. Nick TG. Descriptive statistics. Methods Mol Biol 2007;404:33-52. [doi: 10.1007/978-1-59745-530-5_3] [Medline: 18450044]
70. Lindgren BM, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. Int J Nurs Stud 2020 Aug;108:103632. [doi: 10.1016/j.ijnurstu.2020.103632] [Medline: 32505813]

XSL•FO
RenderX

71. Sperl-Hillen JM, Crain AL, Margolis KL, Ekstrom HL, Appana D, Amundson G, et al. Clinical decision support directed to primary care patients and providers reduces cardiovascular risk: a randomized trial. J Am Med Inform Assoc 2018 Sep 01;25(9):1137-1146 [FREE Full text] [doi: 10.1093/jamia/ocy085] [Medline: 29982627]
72. Sona C, Pollard KA, Schallom M, Schrupp A, Wessman BT. Implementation of a standardized patient/family communication bundle. Crit Care Nurs Clin North Am 2020 Jun;32(2):243-251. [doi: 10.1016/j.cnc.2020.02.006] [Medline: 32402319]
73. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
74. Jouni H, Haddad RA, Marroush TS, Brown SA, Kruisselbrink TM, Austin EE, et al. Shared decision-making following disclosure of coronary heart disease genetic risk: results from a randomized clinical trial. J Investig Med 2017 Mar;65(3):681-688 [FREE Full text] [doi: 10.1136/jim-2016-000318] [Medline: 27993947]
75. Huang JS, Yueh R, Wood K, Ma S, Cruz R, Boyd N, et al. Harnessing the electronic health record to distribute transition services to adolescents with inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2020 Feb;70(2):200-204. [doi: 10.1097/MPG.0000000000002516] [Medline: 31978017]
76. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a novel electronic health record patient portal advance care planning delivery system. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
77. Choi DY, Wagner MP, Yum B, Jannat-Khah DP, Mazique DC, Crossman DJ, et al. Improving implantable cardioverter defibrillator deactivation discussions in admitted patients made DNR and comfort care. BMJ Open Qual 2019 Dec 04;8(4):e000730 [FREE Full text] [doi: 10.1136/bmjoq-2019-000730] [Medline: 31922034]
78. Chunchu K, Mauksch L, Charles C, Ross V, Pauwels J. A patient centered care plan in the EHR: improving collaboration and engagement. Fam Syst Health 2012 Sep;30(3):199-209. [doi: 10.1037/a0029100] [Medline: 22866953]
79. Lafata JE, Cooper G, Divine G, Oja-Tebbe N, Flocke SA. Patient-physician colorectal cancer screening discussion content and patients' use of colorectal cancer screening. Patient Educ Couns 2014 Jan;94(1):76-82 [FREE Full text] [doi: 10.1016/j.pec.2013.09.008] [Medline: 24094919]
80. Denig P, Schuling J, Haaijer-Ruskamp F, Voorham J. Effects of a patient oriented decision aid for prioritising treatment goals in diabetes: pragmatic randomised controlled trial. BMJ 2014 Sep 25;349:g5651 [FREE Full text] [doi: 10.1136/bmj.g5651] [Medline: 25255799]
81. Crosby EM, Gutierrez KM. Facilitating timely advance care planning: discussions in the ambulatory setting with a heart failure population. J Hosp Palliat Nurs 2019 Oct;21(5):445-452. [doi: 10.1097/NJH.0000000000000588] [Medline: 31425316]
82. Litzelman DK, Inui TS, Schmitt-Wendholt KM, Perkins A, Griffin WJ, Cottingham AH, et al. Clarifying values and preferences for care near the end of life: the role of a new lay workforce. J Community Health 2017 Oct;42(5):926-934 [FREE Full text] [doi: 10.1007/s10900-017-0336-5] [Medline: 28353007]
83. Ballard AY, Kessler M, Scheitel M, Montori VM, Chaudhry R. Exploring differences in the use of the statin choice decision aid and diabetes medication choice decision aid in primary care. BMC Med Inform Decis Mak 2017 Aug 10;17(1):118 [FREE Full text] [doi: 10.1186/s12911-017-0514-5] [Medline: 28797295]
84. Fossa AJ, Bell SK, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 01;25(9):1153-1159 [FREE Full text] [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
85. Coylewright M, Keevil JG, Xu K, Dodge SE, Frosch D, Field ME. Pragmatic study of clinician use of a personalized patient decision aid integrated into the electronic health record: an 8-year experience. Telemed J E Health 2020 May;26(5):597-602. [doi: 10.1089/tmj.2019.0112] [Medline: 31381477]
86. Plößnig M, Kabak Y, Lamprinos I, Pabst A, Hildebrand C, Mantwill S. EMPOWER--pathways for supporting the self-management of diabetes patients. Stud Health Technol Inform 2015;212:159-166. [Medline: 26063272]
87. Fried TR, Niehoff KM, Street RL, Charpentier PA, Rajeevan N, Miller PL, et al. Effect of the tool to reduce inappropriate medications on medication communication and deprescribing. J Am Geriatr Soc 2017 Oct;65(10):2265-2271 [FREE Full text] [doi: 10.1111/jgs.15042] [Medline: 28804870]
88. Schafer KM, Gionfriddo MR, Boehm DH. Shared decision making and medication therapy management with the use of an interactive template. J Am Pharm Assoc (2003) 2016;56(2):166-172. [doi: 10.1016/j.japh.2015.12.013] [Medline: 27000167]
89. Martin RW, Brogård Andersen S, O'Brien MA, Bravo P, Hoffmann T, Olling K, et al. Providing balanced information about options in patient decision aids: an update from the international patient decision aid standards. Med Decis Making 2021 Oct;41(7):780-800. [doi: 10.1177/0272989X211021397] [Medline: 34196241]
90. Guo KL. DECIDE: a decision-making model for more effective decision making by health care managers. Health Care Manag (Frederick) 2020;39(3):133-141. [doi: 10.1097/HCM.0000000000000299] [Medline: 32701610]
91. Butler JM, Gibson B, Lewis L, Reiber G, Kramer H, Rupper R, et al. Patient-centered care and the electronic health record: exploring functionality and gaps. JAMIA Open 2020 Oct 29;3(3):360-368. [doi: 10.1093/jamiaopen/ooaa044] [Medline: 33215071]

XSL•FO
RenderX

92. Clavel N, Paquette J, Dumez V, Del Grande C, Ghadiri DP, Pomey MP, et al. Patient engagement in care: a scoping review of recently validated tools assessing patients' and healthcare professionals' preferences and experience. Health Expect 2021 Dec;24(6):1924-1935 [FREE Full text] [doi: 10.1111/hex.13344] [Medline: 34399008]
93. Hadie SN. ABC of a scoping review: a simplified JBI scoping review guideline. Educ Med J 2024;16(2):185-197. [doi: 10.21315/eimj2024.16.2.14]
94. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013 Mar 23;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
95. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2021-01-11]
96. EndNote: Office 2013 compatibility. Clarivate. 2013. URL: https://support.clarivate.com/Endnote/s/article/EndNote-Office-20132016-Compatibility?language=en_US [accessed 2021-01-11]
97. Graph DJ. diagrams.net, draw.io, version: 15.5.2. diagrams.net. 2021 Oct 14. URL: https://www.diagrams.net/ [accessed 2025-02-07]

Abbreviations

EHR: electronic health record
JBI: Joanna Briggs Institute
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Search Extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
SDM: shared decision-making

Edited by A Coristine; submitted 26.04.24; peer-reviewed by K Sajeev, C Bonner, A Coe; comments to author 18.06.24; revised version received 30.10.24; accepted 11.01.25; published 21.02.25.

Please cite as:
Pierce JH, Weir C, Taft T, Richards II W, McFarland MM, Kawamoto K, Del Fiol G, Butler JM
Shared Decision-Making Tools Implemented in the Electronic Health Record: Scoping Review
J Med Internet Res 2025;27:e59956
URL: https://www.jmir.org/2025/1/e59956
doi:10.2196/59956
PMID:

©Joni H Pierce, Charlene Weir, Teresa Taft, William Richards II, Mary M McFarland, Kensaku Kawamoto, Guilherme Del Fiol, Jorie M Butler. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Enhancing Doctor-Patient Shared Decision-Making: Design of a Novel Collaborative Decision Description Language

XiaoRui Guo[1], MSE; Liang Xiao[1], PhD; Xinyu Liu[1], BEng; Jianxia Chen[1], MSE; Zefang Tong[1], MSE; Ziji Liu[1], BEng

School of Computer Science, Hubei University of Technology, Wuhan, China

Corresponding Author:
Liang Xiao, PhD
School of Computer Science
Hubei University of Technology
28 Nanli Road, Hongshan District, Hubei Province
Wuhan, 430068
China
Phone: 86 18062500600
Email: lx@mail.hbut.edu.cn

Abstract

Background: Effective shared decision-making between patients and physicians is crucial for enhancing health care quality and reducing medical errors. The literature shows that the absence of effective methods to facilitate shared decision-making can result in poor patient engagement and unfavorable decision outcomes.

Objective: In this paper, we propose a Collaborative Decision Description Language (CoDeL) to model shared decision-making between patients and physicians, offering a theoretical foundation for studying various shared decision scenarios.

Methods: CoDeL is based on an extension of the interaction protocol language of Lightweight Social Calculus. The language utilizes speech acts to represent the attitudes of shared decision-makers toward decision propositions, as well as their semantic relationships within dialogues. It supports interactive argumentation among decision makers by embedding clinical evidence into each segment of decision protocols. Furthermore, CoDeL enables personalized decision-making, allowing for the demonstration of characteristics such as persistence, critical thinking, and openness.

Results: The feasibility of the approach is demonstrated through a case study of shared decision-making in the disease domain of atrial fibrillation. Our experimental results show that integrating the proposed language with GPT can further enhance its capabilities in interactive decision-making, improving interpretability.

Conclusions: The proposed novel CoDeL can enhance doctor-patient shared decision-making in a rational, personalized, and interpretable manner.

(J Med Internet Res 2025;27:e55341) doi:10.2196/55341

KEYWORDS

shared decision-making; speech acts; agent; argumentation; interaction protocol

Introduction

Shared decision-making (SDM) is a collaborative process between health care providers and patients, designed to incorporate the patient's preferences, values, and the best available evidence to discuss and agree on the most suitable care options [1]. Numerous studies have demonstrated the benefits of SDM, including improved patient engagement in decision-making, enhanced quality of care, and reduced medical costs [2]. In recent years, SDM has emerged as the pinnacle of clinical decision-making, particularly in evidence-based and patient-centered health care settings [3-5].

However, opportunities for active patient participation in medical decision-making remain limited [6], and SDM is predominantly applied in specific health care scenarios [7]. Furthermore, considerable disagreement exists regarding the implementation of SDM, primarily due to the absence of a universally accepted set of steps to guide the process [8,9]. To address these challenges, the Tripartite Dialogue Model was developed as an approach to promote SDM by engaging relevant stakeholders, ensuring continuous updates, and fostering broader

XSL•FO
RenderX

participation [10,11]. However, the practical implementation of similar dialogue models remains largely theoretical, with a lack of detailed guidelines for execution. As a result, understanding the process of joint decision-making and addressing the issue of inadequate consultation—often leading to inefficient utilization of medical resources—has proven challenging [12]. Research indicates that reaching an agreement on a practical model would be a crucial step in promoting its adoption. Currently, there is no existing research specifically focused on doctor-patient codecision-making [7].

Communication among agents, akin to the interaction between a doctor and a patient, is a multifaceted and intricate process. Achieving effective communication and seamless information exchange is crucial for agents to collaborate harmoniously and work toward a shared objective. In the field of linguistic analysis, speech acts have gained prominence for capturing expressions of mental states—such as embarrassment, gratitude, or regret—that convey thoughts or ideas and elicit specific behaviors, such as commands, warnings, or requests, to sustain social interaction [13]. Most research on speech acts has primarily focused on areas such as task assignment, knowledge expression, and identifying intrinsic behavioral patterns in language [14,15]. However, the use of speech acts for regulating interaction behavior remains limited. Among the existing studies that leverage language to describe group behavior, a particularly suitable approach is the Lightweight Coordination Calculus (LCC) language. Initially developed within the framework of the Open Knowledge Project [16], LCC was later extended to the Lightweight Social Calculus (LSC) to support the development of social machines [17]. These languages were not specifically designed to address decision problems and therefore require substantial enhancements to be effectively applied in group decision-making scenarios.

Over the years, various interaction mechanisms for group decision-making have been explored. For example, researchers have developed decentralized group decision consensus mechanisms [18] and introduced personalized feedback mechanisms that provide decision makers with varying levels of consensus alongside tailored feedback suggestions [19,20]. However, in the medical domain, there have been relatively few studies focusing on personalized modeling to describe the group decision-making process. The primary reason is that poor decision-making is often influenced by artificial problems, such as cognitive biases, which may stem from social attributes or certain personality traits [21]. To support group decision-making in health care, some researchers have developed clinical decision support systems [22,23]. These systems aim to create human-computer interactive medical platforms that use data or models to assist doctors and patients in clinical decision-making [24]. However, these systems largely depend on third-party knowledge and fail to adequately consider users' value orientations or preferences.

In medical decision-making, treatment choices are often influenced by physicians' preferences and experience rather than being solely based on medical necessity or scientific evidence, partly due to the lack of evidence-based guidance on the optimal treatment order [25]. To encourage greater patient participation in SDM, patient decision aids (PtDAs) have been developed. These aids provide standardized information about available options and their associated outcomes, helping patients implicitly evaluate the value of these outcomes [26]. However, many existing PtDAs are text-based, include instructions and training for health professionals, and are not aligned with clinical guidelines [27].

With advancements in machine learning and the availability of powerful hardware supporting real-time speech recognition—such as high-quality text-to-speech capabilities and semantic understanding of natural language—health dialogue systems are becoming increasingly popular among patients, hospitals, and universities [28]. These systems enable patients to acquire knowledge and engage in conversations without the direct involvement of health care professionals. Among these systems, ChatGPT, developed by OpenAI, stands out as a prominent language model with significant potential to contribute to public health [29]. Leveraging its extensive knowledge base, it can address patients' inquiries, particularly when dealing with uncertain questions. However, it is important to recognize that, despite its vast database, ChatGPT lacks a deep semantic understanding of the messages it processes [30]. This limitation can undermine trust in its responses, especially in the complex field of medicine, where accuracy and reliability are paramount.

Therefore, by integrating the aforementioned aspects, this paper makes the following key contributions:

- We enhance GPT's explainability by integrating our method with GPT, thereby verifying the feasibility of our approach and enriching our argument.
- We introduce Collaborative Decision Description Language (CoDeL), which includes 4 key elements: (1) a decision-maker interaction process protocol based on LSC; (2) the semantic relationships in the interaction process, defined by combining speech acts; (3) arguments in semantics guided by clinical guidelines; and (4) constraints incorporated for personality modeling. This language provides a more comprehensive description of group decision-making problems and addresses the limitations of LSC.
- We have created some reusable interaction patterns that can be applied in other domains. Specifically, we model the field of atrial fibrillation to validate the feasibility of our method and introduce a novel evaluation approach.

Methods

Overview

We first outlined what our LSC inherited and extended from previous studies (the "LSC-Based Decision-Making Protocol" section). Next, we introduced grammatical semantics within our defined CoDeL (the "Syntactic and Semantics of Collaborative Decision Languages [CoDeL]" section) and demonstrated the interactive relationships of speech acts (the "Constraints of Permissible Speech Acts Pairs" section). Based on the communication model for joint decision-making between doctors and patients, we defined 3 distinct decision-making models (the "An Interactive Model of Shared Decision-Making

XSL•FO
RenderX

Between Doctors and Patients" section). We then used CoDeL to establish the protocol for these decision-making models (the "Protocols Specification Using Collaborative Decision Description Language [CoDeL]" section). Additionally, we briefly discussed the arguments guided by clinical guidelines. Finally, we incorporated the personality model constraint into the protocol (the "The Personality Modeling Included in the Protocol" section).

LSC-Based Decision-Making Protocol

LSC is a protocol language designed to regulate social interactions. It provides a framework for creating social machines that build upon existing social networks, such as X (Twitter), while supporting human interactions through formal role binding, protocol enactment, and result interpretation back to human society [31]. By meeting specific conditions, participants can seamlessly integrate into LSC-based systems, with each condition defining a distinct role and corresponding behavior. In practical applications, the LSC protocol comprehensively defines the agents, roles, messages, computations, and information flows involved, ensuring precise control over complex social interactions and guaranteeing system reliability and efficiency. However, the earlier version of LSC lacked a critical decision-making component, as it did not incorporate clinical guidelines or natural language processing capabilities. Specifically, the original LSC lacked mechanisms for processing natural language and did not support dialogues with argumentation structures grounded in logical reasoning.

To address these gaps, we have extended the earlier version of LSC with a new decision-making component that integrates clinical guidelines and natural language processing capabilities. Our primary objective is to develop an integrated framework that facilitates decision-making in medical settings by leveraging evidence-based recommendations and logical reasoning.

Therefore, we propose a CoDeL to provide health care professionals and patients with a more comprehensive decision support system. Specifically, CoDeL contributes in the ways detailed in Textbox 1.

In contrast to the original LSC, our CoDeL is designed to enable more effective and informed decision-making in medical settings. CoDeL, equipped with natural language processing and evidence-based recommendation capabilities, supports the development of comprehensive decision-support systems tailored to the needs of health care professionals and patients. By addressing this critical gap, we aim to improve patient outcomes, enhance health care quality, and empower clinicians with a cutting-edge tool to support clinical decisions.

Textbox 1. Collaborative Decision Description Language (CoDeL) contributions.

1. Semantically linked to abstract-level speech acts

CoDeL formalizes concrete speech acts into an abstract symbolic format, for example, Rebuttal(surgery1, against_argument)), enabling precise linguistic data transmission and ensuring accuracy in communication.

2. Seamless integration of clinical guidelines

CoDeL incorporates the latest medical guidelines into the decision-making process, providing clinicians with access to up-to-date, evidence-based recommendations. This integration supports informed decision-making through logical reasoning and argumentation structures, ensuring decisions are grounded in empirical evidence.

3. Enhanced patient engagement

CoDeL uses user-friendly natural language formats alongside interactive dialogue mechanisms, enabling patients to participate actively in treatment decision-making. This approach fosters clearer communication, empowering patients to make more informed choices.

Syntactic and Semantics of Collaborative Decision Languages (CoDeL)

Group decision problems involve analyzing individual opinions and preferences to understand how decisions are made. These problems often require cooperation and negotiation among multiple individuals, taking into account factors such as opinions, preferences, rights, and interests. To address these challenges effectively, we propose CoDeL as a framework for describing and solving group decision-making problems. Specifically, to enhance the scientific validity of CoDeL, we define distinct categories of communication processes within the context of doctor-patient SDM in medical scenarios. This approach facilitates improved communication and negotiation among the involved agents, enabling a more robust decision-making process, as outlined in Textbox 2.

It is crucial to emphasize that within the context of our framework, the term *rebuttal* allows for negotiation, discussion, and the presentation of counterarguments. By contrast, *refuse* denotes a definitive and explicit rejection, devoid of any accompanying explanation or reasoning. The concept of *refuse* carries a more forceful connotation, as it involves an outright dismissal of the proposed action or treatment without providing further justification or elaboration.

In this research, we introduce CoDeL, a framework that not only encompasses the aforementioned interaction behaviors but also incorporates a group decision protocol. This protocol defines the roles of agents, specifies the messages to be exchanged, and outlines the information to be utilized during interactions. The protocol can be executed either sequentially or as a committed-choice option, ensuring a systematic and coordinated decision-making process. Within each action, agents follow the protocol associated with their roles and interact with other agents accordingly. Before taking any action, agents evaluate whether they are authorized to perform the action by verifying compliance with the corresponding behavioral rules.

To facilitate a better understanding of the CoDeL syntax, we provide an illustrative example in Figure 1, demonstrating how the framework is applied in practice.

XSL•FO
RenderX

- "a(name(K),N)": Indicates the agent's role name, ID, and intrinsic knowledge of existence.
- "Agenti() ⇒ Agentj()" or "Agenti() ⇐ Agentj()": Indicates that the message is being passed to another agent. A message declaration needs to be specified on the sender side and a rule declaration on the receiver side.
- "then": Indicates that the clause before it must continue the clause after it.
- "or": Indicates that there are alternative decision branches available and only 1 should be selected.
- "←()": Indicates that the constraint satisfaction function must resolve to true in order to execute its associated clause.
- "assert(diseases, surgery1 + surgery2) ⇐ a(doctor(),D)": The patient receives an assertion message sent by the doctor, which includes the name of the disease and the doctor's recommended treatment plan.
- "question(disease, surgery1 + surgery2) ⇒ a(doctor(),D)←interested in(disease, surgery1 + surgery2)": If the patient wishes to acquire more knowledge, they can send a message of inquiry to the doctor and await confirmation from the doctor.

Textbox 2. Decision-making process.

1. Assert

Provide information related to diagnoses, medication instructions, treatment plans, or consultation details.

2. Accept

Agree to follow the proposed action plan, treatment plan, or medical advice.

3. Refuse

Reject the proposed action plan, treatment plan, or medical advice.

4. Question

Seek further clarification due to limited understanding of the information provided in an assertion.

5. Justify

Offer honest answers to questions raised and provide reasonable explanations to support assertions made.

6. Rebuttal

Disagree with an assertion and present reasons for the opposition.

7. Persuade

Attempt to influence the beliefs, attitudes, or behaviors of the opposing party to accept or adopt the assertion made.

8. Retract

Withdraw the initial assertion due to the discovery of errors or a change in beliefs based on updated information.

Figure 1. An example of doctor-patient communication is LSC. After receiving a message from the doctor, the patient decides whether to send an inquiry message.

```
a(patient(),P)::=
assert(diseases, surgery1+surgery2) ⇐ a(doctor(),D) then
  question(diease, surgery1+surgery2) ⇒ a(doctor(),D) ←interested_in(diease,surgery1+surgery2)
a(patient(),P)
```

Constraints of Permissible Speech Acts Pairs

Speech acts do not exist in isolation; instead, they intersect and influence each other within the dynamic evolution of dialogue, thereby facilitating communication between agents. Figure 2 illustrates the relationships among speech acts, and Table 1 standardizes and explains the speech act interaction pairs applicable to this study.

For instance, when a doctor makes a diagnostic assertion, the patient's response can take various forms, including acceptance, rejection, questioning, or refutation (1-4). If the patient questions the initial assertion, the doctor must provide detailed arguments and validate the rationality of the assertion (6). Based on the doctor's response, the patient may choose whether to continue questioning (7). If the patient rejects the evidence and refutes it (8), the doctor may use persuasive strategies to reinforce their initial position (9), potentially initiating another feedback loop. During this process, the patient may persist with further questions (10), fostering a cycle of in-depth interaction until all concerns are resolved. At the conclusion of the cycle, the patient, influenced by the doctor's persuasion and verification, ultimately makes 1 of 2 decisions: to accept or reject the doctor's assertion (11-14). As the conversation progresses and both parties exchange their perspectives, if the patient persists in rejecting a specific assertion, the doctor may decide to actively retract the previous statement after considering the pros and cons (16).

XSL•FO
RenderX

The doctor can then determine whether to propose a new assertion based on the available alternatives (5). Notably, certain speech act interaction pairs that violate standard behavioral logic are prohibited during agent interactions. For instance, as described above, an agent sending a *refuse* message demonstrates a strong rejection attitude and will not receive any response message. Consequently, the (*refuse-persuade*) interaction pair is logically invalid. Similarly, the (*assert-justify*) interaction pair, where an agent sends an *assert* message to express its opinion and the receiving agent replies with a *justify* message to provide supporting arguments for the other agent's opinion, also violates normal behavioral logic.

Figure 2. Interaction relationship of speech acts.

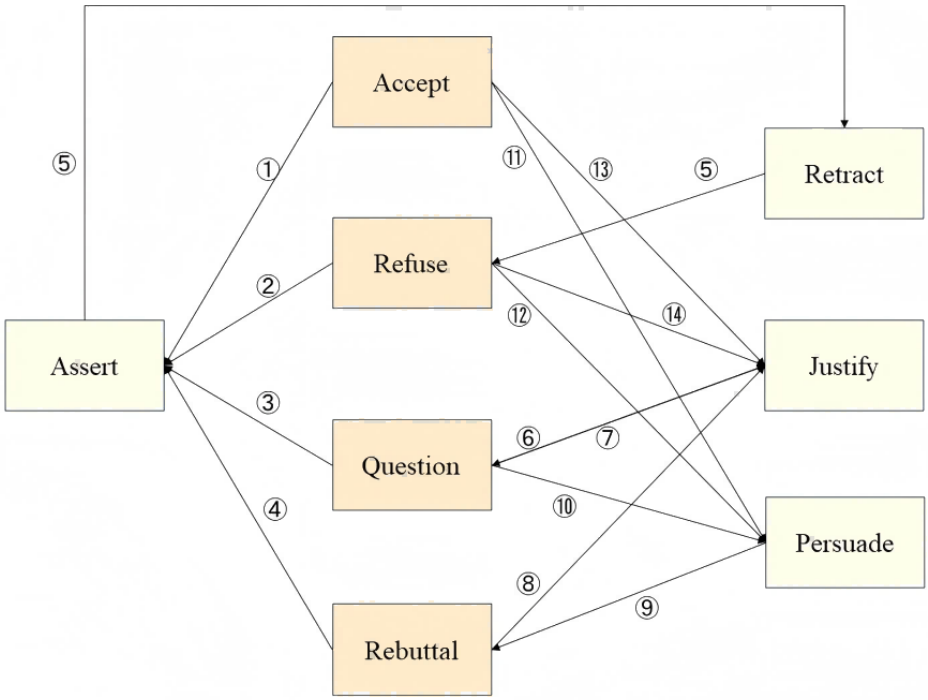

XSL•FO
RenderX

Table 1. Legal speech acts interaction pairs that can be constructed and their meanings.

Number	Speech acts interaction	Description
1	*Assert-accept*	The agent that receives the assertion message evaluates that the assertion is reasonable and feasible, and then sends an acceptance message to explicitly express its agreement with the assertion.
2	*Assert-refuse*	When a proxy receives an assertion message, if it cannot accept the assertion result, it sends a rejection message to explicitly indicate that it does not accept the assertion.
3	*Assert-question*	The agent that receives the assertion message further explores the rationality of the assertion by sending a challenge message.
4	*Assert-rebuttal*	The agent that receives the assertion message sends a rebuttal message if it disagrees with the assertion.
5	*Refuse-retract + assert*	An agent that receives a rejection message should withdraw its previous assertion and choose whether to send a new assertion based on the current alternatives.
6	*Question-justify*	An agent that receives a challenge message should send a confirmation message containing supporting arguments to alleviate the agent's challenge.
7	*Justify-question*	When a confirming message is received but doubts remain unresolved, further inquiries or questions can be sent to obtain a more comprehensive understanding.
8	*Justify-rebuttal*	The agent that receives the confirmation message sends a rebuttal message if it disagrees with the supporting information.
9	*Rebuttal-persuade*	The agent that receives the rebuttal message sends a persuasive message to try to change the other party's mind and guide the other party to accept its assertion.
10	*Persuade-question*	When an agent that receives a persuasion message has doubts about the persuasion argument provided, they may send a questioning message.
11	*Persuade-accept*	An agent that receives a persuasion message and believes that the persuasive argument provided is valid will send an acceptance message indicating that it has been persuaded and agrees to accept the assertion issued.
12	*Persuade-refuse*	An agent that receives a persuasion message and deems the provided persuasive argument to be invalid sends a rejection message indicating that it is not persuaded and refuses to accept the assertion issued.
13	*Justify-accept*	After evaluating the received confirmation message and finding no objection, the agent sends an acceptance message to explicitly indicate its agreement with the assertion.
14	*Justify-refuse*	When an agent evaluates a confirmation message and still cannot accept its contents, it will send a rejection message to explicitly convey that it does not accept the assertion.

An Interactive Model of Shared Decision-Making Between Doctors and Patients

Doctor-patient SDM involves collaborative communication and negotiation between health care providers and patients to reach a consensus on the most suitable treatment plan. To establish a standardized framework for their interaction, we introduce a generic interaction model that outlines the specific flow of this process, as shown in Figure 3. This model serves as a valuable tool for understanding and analyzing the dynamics of doctor-patient SDM, supporting the implementation of consistent and effective practices in health care settings.

- Assert(disease, surgery1 + surgery2): The physician sends an assertive message to the patient, outlining the disease and proposed treatment plan.
- Question(disease, surgery1 + surgery2): After receiving the assertive message from the physician, if the patient seeks further clarification, they send a questioning message to the physician.
- Justify(disease, surgery1 + surgery2, information): Upon receiving the questioning message, the physician sends a confirming message with supporting information in response.
- Accept(surgery1 + surgery2): After receiving the confirming message, the patient accepts the physician's proposed treatment plan.
- Rebuttal(surgery1, against argument): If the patient is dissatisfied with a particular surgery, they express their concerns by sending a rebuttal message, outlining the reasons for their opposition.
- Persuade(surgery1, support argument): Upon receiving the patient's rebuttal, the physician engages in persuasion by offering supportive arguments and explanations to address the patient's concerns.
- Accept(surgery1 + surgery2)/Refuse(surgery1 + surgery2): If the persuasion is successful, the patient accepts the physician's assertion and agrees to the proposed treatment plan. However, if the patient remains unconvinced, they may refuse the assertion.
- Retract(surgery1 + surgery2) and Assert(disease, surgery3 + surgery2): If rejected, the physician retracts the previous assertion and sends a new assertive message, offering an alternative treatment plan to address the patient's concerns.
- Accept(surgery3 + surgery2): The patient carefully considers the new treatment plan and ultimately accepts it as the revised course of action.

XSL•FO
RenderX

The determination of a treatment plan is not solely the doctor's decision; the patient is also involved in the decision-making process. After communication and negotiation between the physician and the patient, the most suitable treatment plan is ultimately decided. Based on this process, we have categorized and modeled the different states held by both doctors and patients regarding the same treatment plan, as outlined below.

Figure 3. Generic interaction timing diagram for patient-provider shared decision-making.

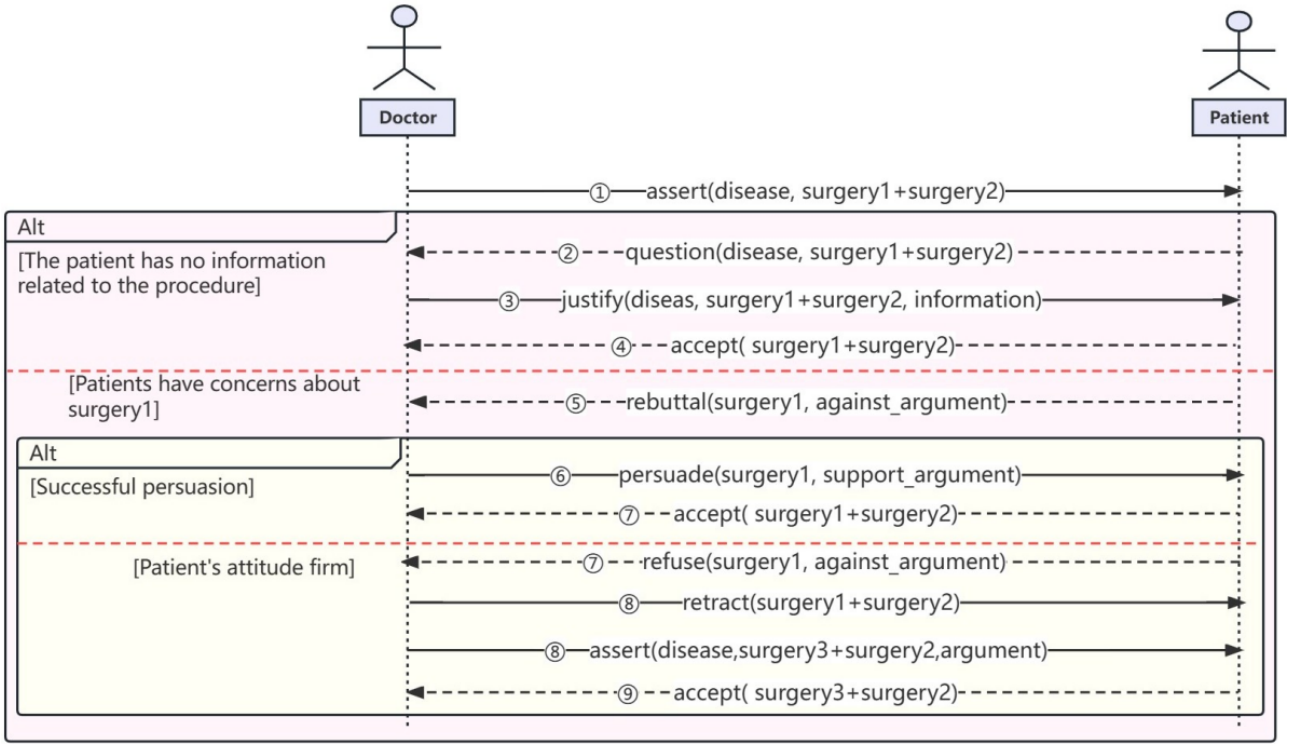

Direct Acceptance Without Consultation

The physician conducts a thorough diagnosis, identifies the specific disease, and proposes a comprehensive treatment plan along with relevant information. After receiving this information, patients who may lack sufficient knowledge about the disease and treatment options may request additional details. As a result, the patient asks a series of questions to seek clarification and further explanation from the physician regarding the information provided. Throughout the communication process, the patient gradually gains a better understanding of the disease and treatment options. Eventually, the patient fully accepts the physician's advice. To describe this process and capture the dynamics of the interaction, we use speech acts to represent the types of messages exchanged during the interaction. As shown in Figure 4, the relevant parameters for each type of interaction and their interpretations are clearly indicated.

Figure 4. Definition of parameters and interaction processes in speech acts.

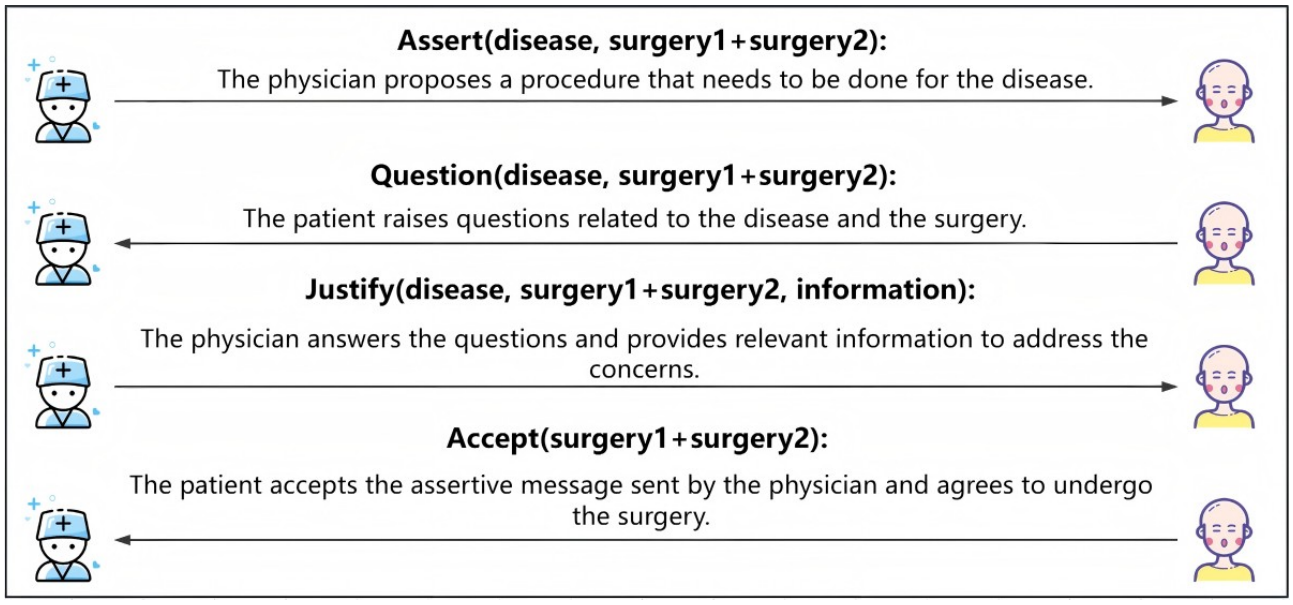

XSL•FO
RenderX

Patient Has Concerns and Agrees After Consultation

The patient opposes the treatment plan proposed by the physician. When the physician presents the treatment plan, the patient directly refutes it. Upon receiving the patient's objection, the physician, considering the treatment plan as a whole, explains its benefits in an attempt to persuade the patient. Ultimately, after the physician's persuasion, the patient changes their attitude and accepts the proposed treatment plan. The corresponding interaction types and parameter definitions are shown in Figure 5.

Figure 5. Definition of parameters and interaction processes in speech acts.

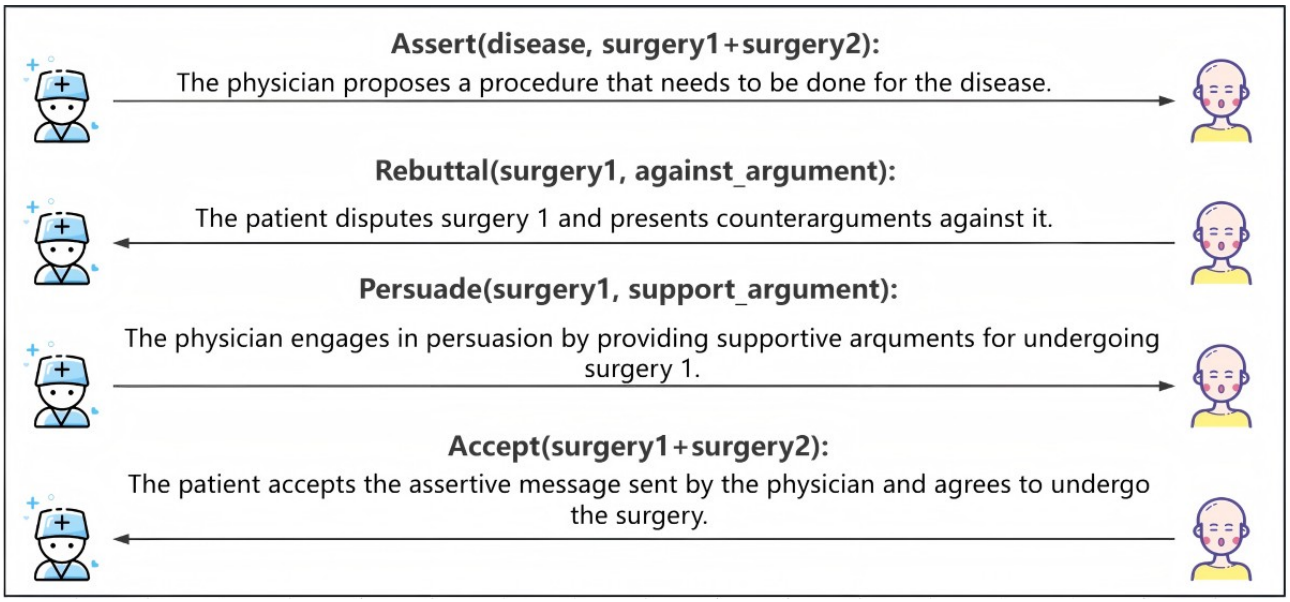

Patient Had Concerns and Altered the Treatment Plan After Consultation

Similar to the previous scenario, the patient expresses opposition to the treatment plan, and the physician attempts to persuade them. However, by contrast, the patient maintains a firm stance, refusing to change their position despite the physician's persuasion. As a result, the physician proposes a new treatment plan, which the patient ultimately accepts. Figure 6 illustrates the interaction definition for the scenario of unsuccessful persuasion, omitting the parts identical to Figure 5 and focusing on the section related to the unsuccessful persuasion.

Figure 6. Definition of parameters and interaction processes in speech acts.

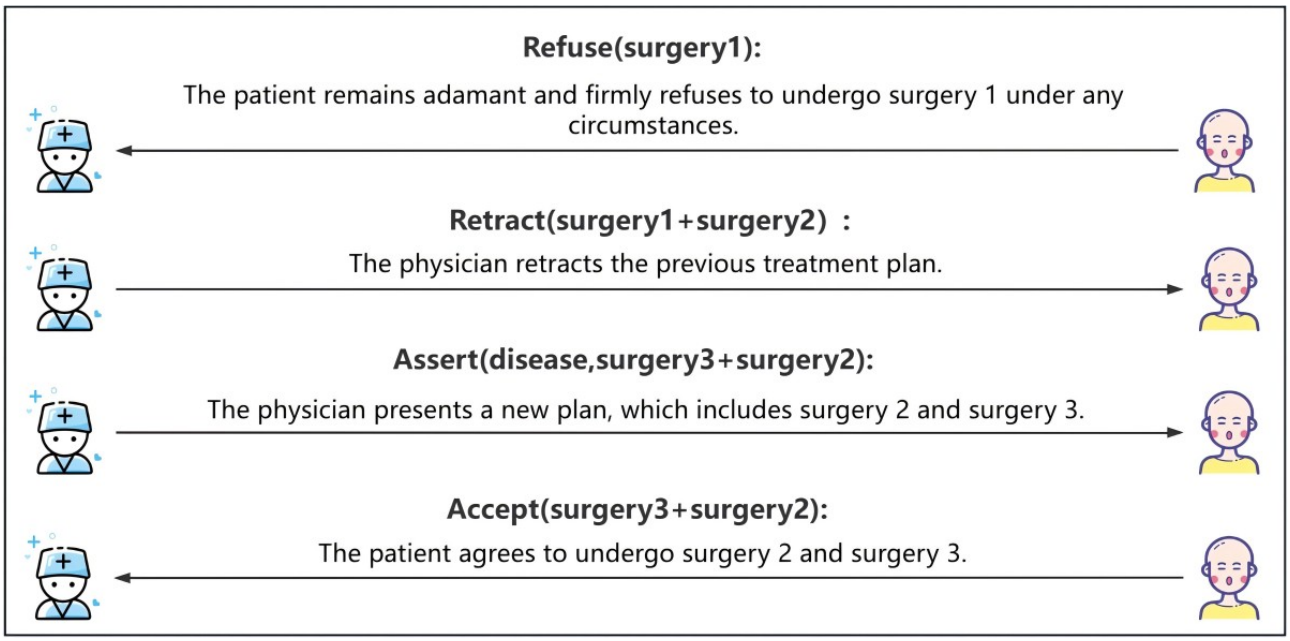

Protocol Specification Using Collaborative Decision Description Language(CoDeL)

By integrating the 3 interaction models for patient-physician SDM, as detailed in the "LSC-Based Decision-Making Protocol" section, we have developed a robust and comprehensive protocol that enhances both interpretability and applicability. The resulting protocols, shown in Figures 7 and 8, define the doctor-patient interaction process, respectively. Within this protocol, there are 2 distinct roles: the "doctor" and the

XSL•FO
RenderX

"patient." These roles are explicitly defined using "a(doctor(), D)" and "a(patient(), P)." To enhance comprehension and maintain consistency, we assign unique identities to each role, represented by the symbols "D" and "P." To provide a comprehensive understanding of the protocol, we carefully explain each sentence, thoroughly examining and clarifying the constraints embedded within it. These constraints, which define the boundaries and requirements of the protocol, are extensively detailed in the "An Interactive Model of Shared Decision-Making Between Doctors and Patients" section of the documentation. To build a comprehensive interaction protocol, we integrate the 3 interaction models defined in the "Syntactic and Semantics of Collaborative Decision Languages (CoDeL)" section, as shown in Figure 8. Each color-coded segment in the figure corresponds to its respective interaction model. By combining these models, we create an overarching framework that encompasses all aspects of the interaction process, facilitating structured and standardized methods of communication and decision-making among the parties involved.

Figure 7. Definition of physician interaction protocol.

a(doctor(),D)::= assert(diseases, surgery1+surgery2) ⟹ a(patient(),P← check(disease) then	Upon meeting the diagnostic criteria for a specific disease, the doctor proceeds to communicate the disease and the treatment plan he has formulated to the patient through an assertion message.
justify(diseases, surgery1+surgery2, information) ⟹ a(patient(),P) ← receive(question(diseases, surgery1+surgery2), a(patient(),P)) then	When the doctor receives a questioning message from the patient, the doctor will respond with a confirming message, providing explanations to address the raised concerns. "Information" is the doctor's confirming evidence in response to the patient's inquiries.
(accept(surgery1+surgery2) ⟸ a(patient(),P) or	Waiting for the patient's acceptance message.
persuade(surgery1, support_argument) ⟹ a(patient(),P) ← receive(rebuttal(surgery1, against_argumen), a(patient(),P))) then	When receiving the patient's rebuttal message, the doctor will engage in persuasion with the patient. "support_argument" refers to the supportive arguments the doctor uses to persuade the patient.
(accept(surgery1+surgery2) ⟸ a(patient(),P) or	Waiting for the patient's acceptance message.
assert(diseases, surgery3+surgery2) ⟹ a(patient(),P) ← receive(refuse(surgery1), a(patient(),P))) then	When receiving the patient's refusal message, the doctor will present a new assertive message. "Surgery3+surgery2" represents the new treatment plan proposed by the doctor.
accept(surgery3+surgery2) ⟸ a(patient(),P) a(doctor(),D)	Waiting for the patient's acceptance message.

XSL•FO
RenderX

Figure 8. Definition of patient interaction protocol.

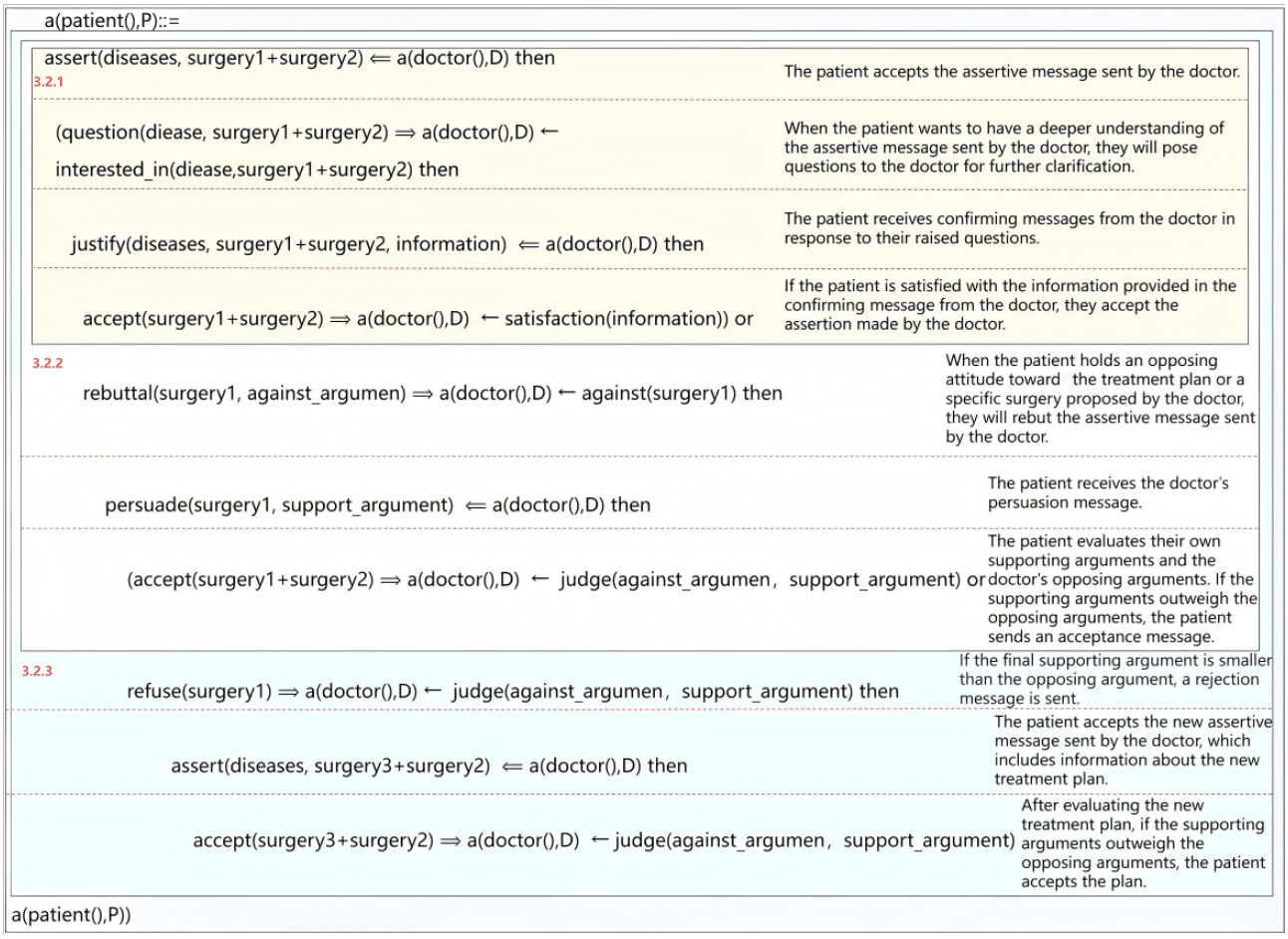

Argumentation Driven by Clinical Guidelines

The CoDeL framework facilitates dynamic decision-making between doctors and patients through natural language interactions. Throughout this process, the system continuously integrates and updates clinical guidelines and other relevant information to inform its decisions.

To illustrate the operational logic and effects of this pattern, let us consider an example based on clinical guidelines for atrial fibrillation. As shown in Figure 9, the doctor proposes an initial treatment plan (cardiac angiography and surgical ablation) through an assertion to the patient. However, rather than merely presenting a decision, the information exchanged between the doctor and patient includes not only the proposed treatment but also the patient's attitude toward it (eg, acceptance, rejection, or hesitation), along with any concerns or doubts they may have regarding the decision. In the figure, the patient raises a rebuttal, opposing cardiac angiography in the treatment plan. CoDeL will then find supporting arguments for cardiac angiography based on the provided clinical guidelines (eg, "This aids in identifying potential abnormalities..."). The doctor can then use this argument to construct their next dialogue, attaching the corresponding *persuade* speech act semantics to it (eg, "Cardiac Angiography is crucial for accurately assessing...").

As doctors and patients engage in interactive conversations, CoDeL identifies specific decision types and searches for relevant arguments to support or refute these decisions. The system considers not only the proposed decisions but also the dynamic attitudes and concerns of both parties, continuously updating its knowledge base with new information. If persuasion fails, the system offers alternative treatment plans based on prior requests. The conversation is considered complete only when CoDeL determines that the patient's decision type has shifted to acceptance, signaling a mutually agreed-upon treatment plan. Throughout this process, clinical guidelines and other relevant knowledge are dynamically integrated into the joint decision-making model, ensuring that both doctors and patients have access to the most up-to-date information.

XSL•FO
RenderX

Figure 9. CoDeL (Collaborative Decision Description Language) framework driven by clinical guidelines in actual scenarios.

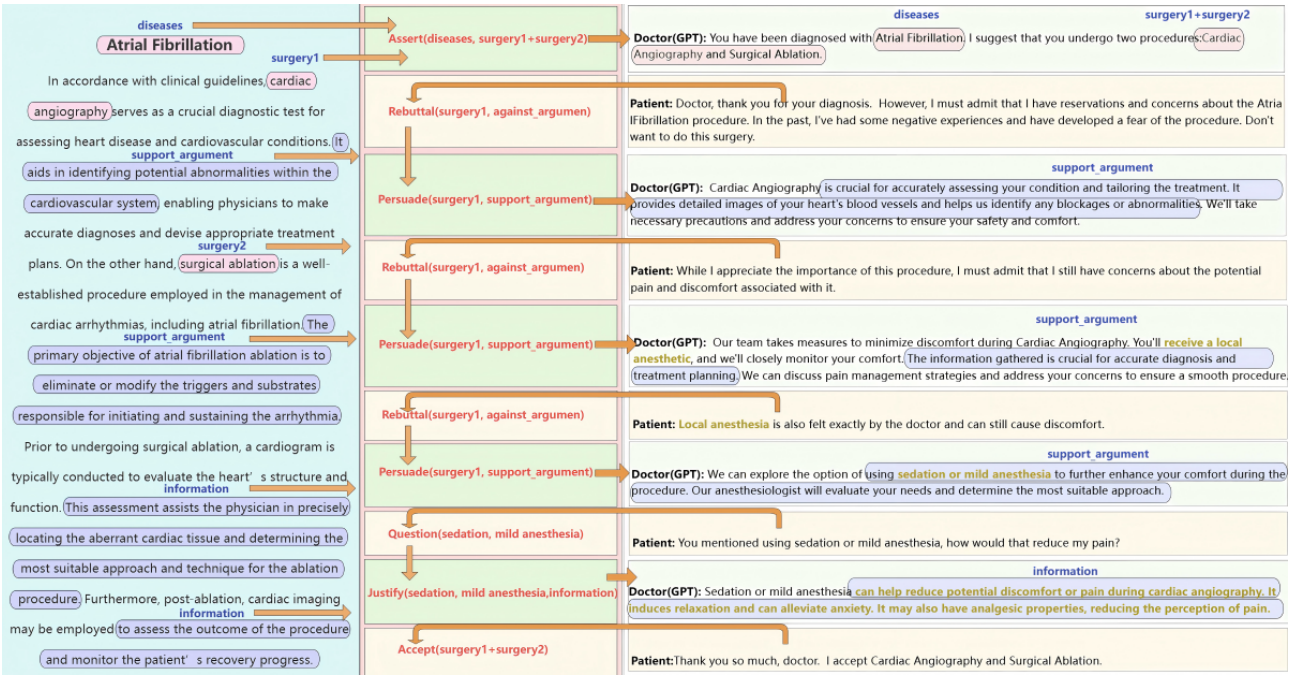

The Personality Modeling Included in the Protocol

To capture the dynamics of decision-making and the influence of individual personalities, we have developed a model that incorporates both arguments and personality factors. This model is represented by the following formula:



In this formula, the decision outcome is determined by the interaction between the arguments presented during the discussion and the unique personality traits of each individual involved. In our model, we define 3 basic personality traits, and by setting parameters for personality modeling, we can explain why patients may exhibit different responses of acceptance or rejection, as outlined in the protocol in the "Constraints of Permissible Speech Acts Pairs" section (Textbox 3).

As depicted in Figure 10, both the doctor and the patient have distinct belief sets, denoted as BDset and BPset, respectively. Each interaction involves a collection of supporting arguments (Arg + set) and opposing arguments (Arg – set). Within the context of their belief sets, individuals encounter both external arguments that challenge their beliefs and supporting arguments that reinforce them. To capture the various manifestations of individuals' cognitive personalities in response to judgment, we represent them in the formula mentioned earlier. The belief sets of both the doctor and the patient have a fixed degree, represented by K. Persistency strengthens the Arg + set, with the degree of persistence determined by K1. The higher the K1, the stronger the persistence, making it less likely to change one's view. The extent to which *openness* enhances the Arg – set is determined by K2, indicating that the greater the willingness to accept opposing views, the more likely it is to change one's own. Critical thinking should be determined based on the current evaluation of the argument. If the evaluation tends to support the argument, it will amplify the effect of the Arg + set, as shown in equations (1) and (3). Conversely, if the evaluation leans toward the opposing view, the impact of the Arg – set will be strengthened, as shown in equations (2) and (4). The exact level of influence is determined by K3. By incorporating these factors into the model, a comprehensive framework is established. Using this model, we can gain a deeper understanding and analysis of the dynamics in doctor-patient interactions, considering the influence of individual personalities and their respective arguments. The calculated values of these factors help to more fully and accurately reflect the decision-making process.

The incorporation of persona modeling constraints is essential to our protocol, as it helps to:

- Capture the differentiation of individual personalities and their impact on decision-making.
- Provide a categorized representation of patient responses to doctor-patient interactions.
- Enhance the overall effectiveness of the protocol by considering the unique characteristics of each individual involved.

Personality modeling offers a more comprehensive understanding of the decision-making process by incorporating individual personality traits and corresponding arguments. By defining 3 basic personality traits—persistence, critical thinking, and openness—we can capture the dynamics of decision-making and the influence of personality. The model considers the unique traits and corresponding arguments of each individual involved, enabling a deeper understanding of how patients respond to physician persuasion during medical consultations.

XSL•FO
RenderX

Textbox 3. Patient responses.

1. Persistency

Individuals with a persistent personality are characterized by strong beliefs and clear goals. They are determined and resilient, continuing to pursue their objectives despite challenges or obstacles. When confronted with the doctor's persuasion, persistent individuals are less likely to change their beliefs easily and may resist external influence. This persistence often leads them to maintain their original position, resulting in a rejection outcome.

2. Critical thinking

Individuals with a critical thinking personality are adept at evaluating and analyzing information objectively. They approach decision-making with caution and rationality, using logic and reasoning to assess different viewpoints. Critical thinkers actively seek evidence and sound arguments to support their beliefs, and they are vigilant about recognizing biases and errors in the decision-making process. When confronted with persuasion, critical thinkers carefully evaluate the arguments presented, weighing them against their own beliefs. Depending on the strength of the arguments and their alignment with personal reasoning, critical thinkers may either accept or reject the suggestions of others.

3. Openness

Individuals with an open-minded personality are receptive to new ideas and diverse perspectives. They are adaptable and flexible in their beliefs, willing to consider different viewpoints. When faced with persuasion, open individuals are more likely to accept new information and adjust their beliefs accordingly.

Figure 10. The relationship between individual personality and belief.

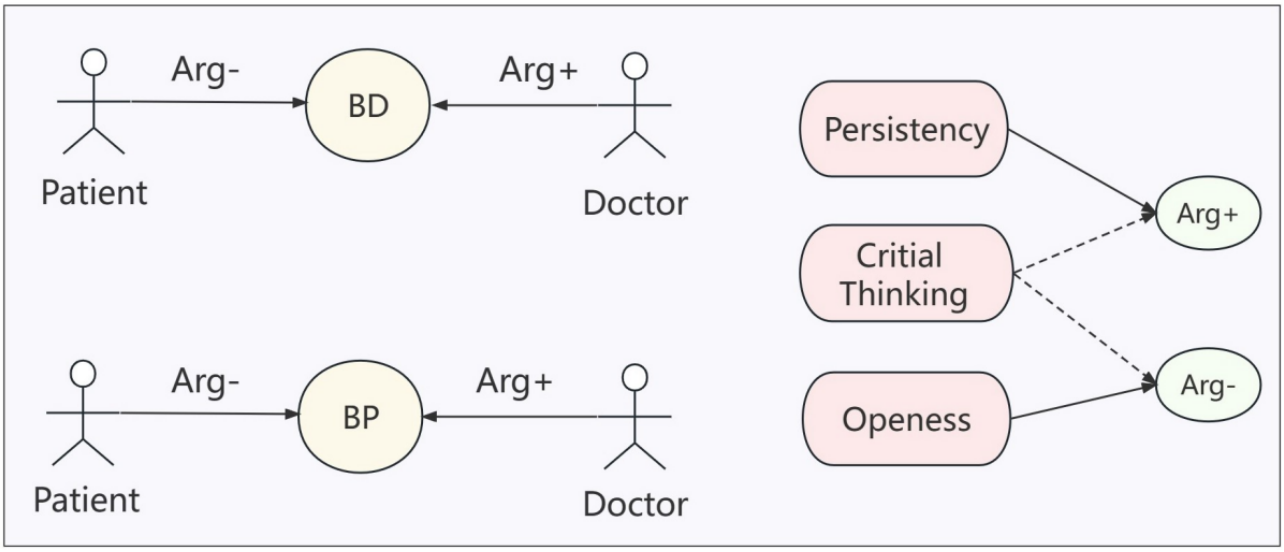

Ethical Considerations

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. This study uses the communication process between the knowledgeable person and the doctor as the research support, without any interest relationship, and the informed consent of the other party. Written informed consent has been obtained from participants regarding the release of their individual clinical details.

In accordance with the second provision of Article 32 of the Ethical Review Measures for Life Sciences and Medical Research Involving Humans issued by the National Health Commission of the People's Republic of China, which states that "research using anonymized information data" is exempt from ethical review for non-interventional studies, this study does not require institutional review board approval [32].

Results

This paper aims to explore the medical decision-making process for a patient diagnosed with atrial fibrillation, with the goal of validating the 3 decision models and interaction patterns presented in the "Methods" section (Textbox 4).

To provide a comprehensive understanding of these diagnostic and therapeutic modalities, Table 2 presents a detailed comparison that includes various aspects, such as the alternative of cardiac ultrasound.

In the context of an appeal, the communication between the patient and physician can result in the following possible scenarios:

- David has no further concerns or objections and readily agrees to the proposed solution after fully understanding it. This indicates that the decision-making process was smooth, with no major obstacles or conflicts.
- David expressed reluctance to undergo the cardiac imaging procedure due to the psychological impact of his previous negative experience. However, after effective consultation and discussion with the physician, David was successfully

XSL•FO
RenderX

persuaded and ultimately agreed to proceed with the recommended protocol. This underscores the importance of doctor-patient communication in addressing patients' concerns and alleviating their anxiety to reach a mutual agreement.

- David strongly opposed undergoing cardiac imaging and remained unconvinced despite the physician's attempts to persuade him. In such cases, the physician explores alternative options and proposes a different procedure or diagnosis that aligns with the patient's preferences. This illustrates the flexibility and adaptability of the decision-making process, ensuring the patient's active participation and respect for their autonomy.

To represent these 3 decision-making processes, we apply the generic model outlined in the "LSC-Based Decision-Making Protocol" section, which offers a comprehensive framework for analyzing and understanding the dynamics of SDM between patients and physicians.

Textbox 4. Case description

1. Case

The following case is based on a real-life scenario: David, a 53-year-old individual with a medical history of diabetes and heart disease, has previously undergone cardiac angiography. Recently, he has been experiencing prolonged episodes of heart palpitations, along with temporary episodes of cardiac arrest. To address these issues, he sought medical attention and visited the hospital for a comprehensive examination. Subsequent medical evaluation revealed that David was diagnosed with atrial fibrillation, a common arrhythmia that, if left untreated, can lead to severe complications such as blood clot formation and an increased risk of stroke. As a result, the medical team recommended that David undergo both cardiac angiography and surgical ablation procedures. It is important to note that during his previous treatment, David underwent only a heart angiogram because he was in a city alone without family support. Additionally, he experienced pain during the procedure, which has had a lasting psychological impact on him.

2. Clinical guideline

In accordance with clinical guidelines [33], cardiac angiography is a critical diagnostic tool for evaluating heart disease and other cardiovascular conditions. It helps identify potential abnormalities within the cardiovascular system, allowing physicians to make precise diagnoses and develop appropriate treatment plans. By contrast, surgical ablation is a widely used procedure for managing cardiac arrhythmias, including atrial fibrillation. The main goal of atrial fibrillation ablation is to eliminate or modify the triggers and substrates that initiate and sustain the arrhythmia. Before undergoing surgical ablation, a cardiogram is typically performed to assess the heart's structure and function. This assessment helps the physician accurately identify the abnormal cardiac tissue and determine the most appropriate approach and technique for the ablation procedure. Additionally, after the ablation, cardiac imaging may be used to evaluate the procedure's outcome and monitor the patient's recovery.

Table 2. Information about the surgery involved in the case.

Surgery	Rationale	Support argument (advantages)	Against argument (disadvantages and side effects)
Cardiac angiography	The type, duration, and frequency of atrial fibrillation can be determined, while the presence of potential cardiovascular problems can be assessed.	• Provides detailed information about the structure of the heart. Make an accurate diagnosis. • Helps to plan surgical procedures. • Improves the success rate of surgery and reduces the risk.	• Invasive surgery. • There are potential complications, such as vascular damage, infection, and contrast agent toxicity to the kidney. • Slight pain (which was aggravated by David's last experience). • Vascular injury. • Infection. • Contrast agent toxicity.
Surgical ablation	Interventional therapy to target and eliminate abnormal atrial conduction tissue.	• Effectively restoring normal heart rhythm. • Eliminates arrhythmia symptoms. • Improves the quality of life.	• Invasive surgery. • There is a risk of bleeding, blood clots, and infection. • It may cause pain. • Bleeding and infection.
Cardiac ultrasound	To assess heart structure and function using noninvasive ultrasound imaging.	• Noninvasive procedures. • Provides real-time imaging of the heart. • Helps assess heart function. • Assists in the development of treatment plans.	• Limited ability to visualize certain structures. • It takes expertise to explain it accurately.

David Talked to the Doctor and Accepted the Decision

Despite David's previous experience with cardiac angiography and the negative outcome, his determination to address his atrial fibrillation remained strong. To make an informed decision, he actively participated in the communication process outlined in Figure 11. Through effective dialogue with his physician, David sought to gain a comprehensive understanding of the proposed treatment plan. As illustrated in Figure 11, the physician-patient communication followed our defined Collaborative Decision Language (CoDeL), ensuring clear communication and a structured decision-making process.

The doctor initially sends an assertion message recommending a cardiac angiogram and surgical ablation for David's atrial

XSL•FO
RenderX

fibrillation. As David is not familiar with the treatment plan, he responds by asking questions about the 2 procedures. The doctor acknowledges David's questions and provides detailed information, including relevant supporting arguments. After being well-informed, David accepts the doctor's recommendations.

Figure 11. The interaction between the doctor and David and the corresponding CoDeL (Collaborative Decision Description Language).

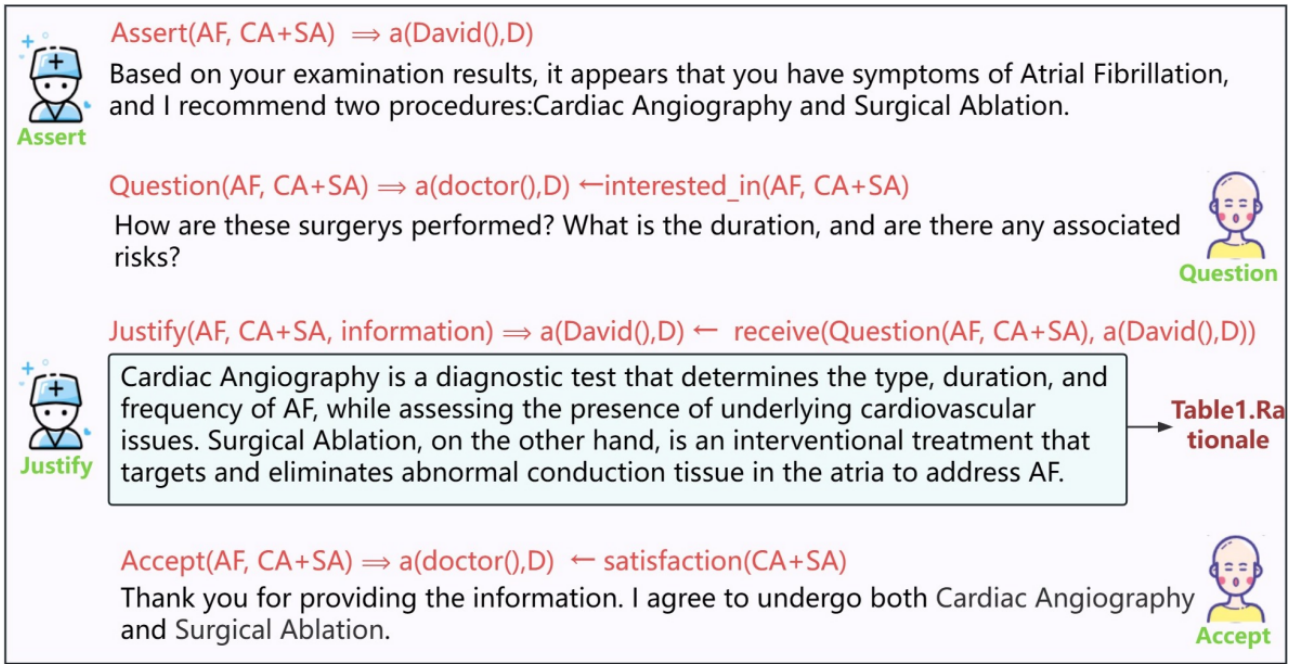

David Has a Concern but Accepts the Decision of the Doctor

Figure 12 presents a concise depiction of the case in which David initially hesitated to undergo the cardiac angiography procedure due to a negative prior experience. However, through the doctor's effective persuasion, David ultimately agreed to proceed with the procedure again.

Initially, the physician delivered an assertive message outlining a comprehensive treatment plan, which included cardiac imaging and surgical ablation for David's atrial fibrillation. Given David's prior experience with cardiac angiography, which had left a lasting impression, he voiced his concerns and objections about the proposed treatment plan. David articulated his reservations, sharing his perspective with the physician. In response to David's rebuttal, the doctor skillfully used persuasive techniques, emphasizing the benefits and advantages of undergoing the cardiac angiography procedure. By carefully explaining the procedure and addressing David's concerns, the doctor successfully persuaded him that the benefits outweighed the potential risks. As a result, David's apprehensions were alleviated, and he ultimately agreed to undergo both cardiac imaging and surgical ablation as part of his treatment plan.

Figure 12. The interaction between the doctor and David and the corresponding CoDeL (Collaborative Decision Description Language).

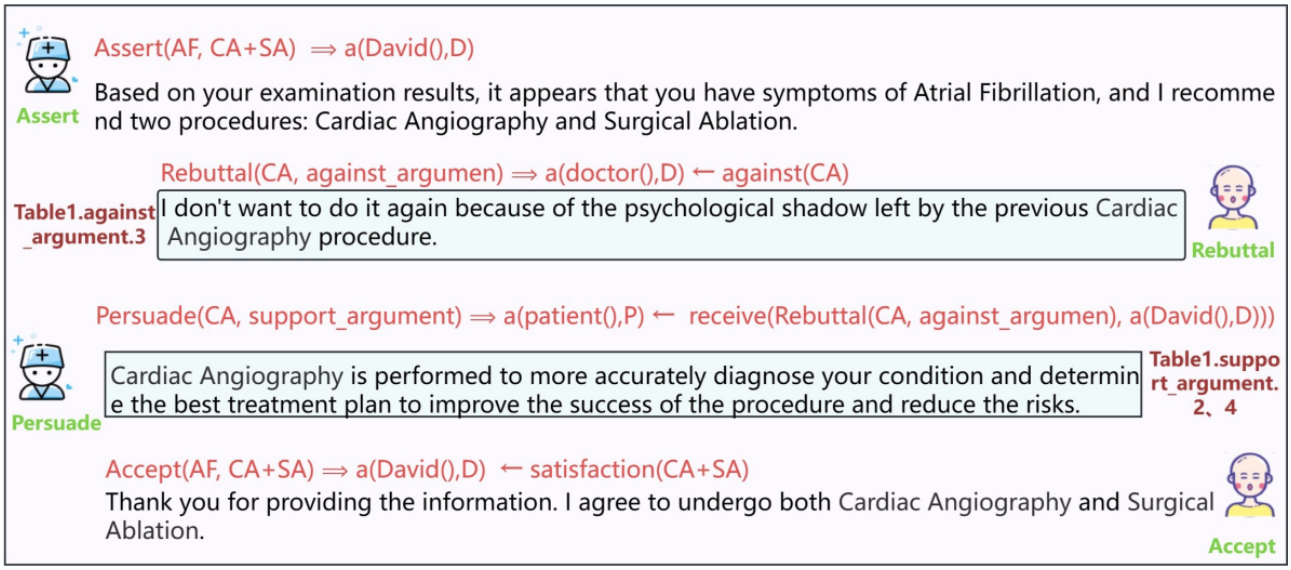

XSL•FO
RenderX

David Had Concerns and the Doctor Accepted His Decision

As described earlier, David expressed a strong aversion to undergoing a cardiac angiogram. However, in this case, despite the doctor's persuasive efforts, David remained resolute and unconvinced, insisting on opposing the procedure. Consequently, the doctor proposed an alternative course of action, suggesting a simultaneous cardiac ultrasound and surgical ablation. Thus, Figure 13 focuses only on the part that was ultimately rejected, which contrasts with the part accepted earlier in Figure 12.

Figure 13. The interaction between the doctor and David and the corresponding CoDeL (Collaborative Decision Description Language).

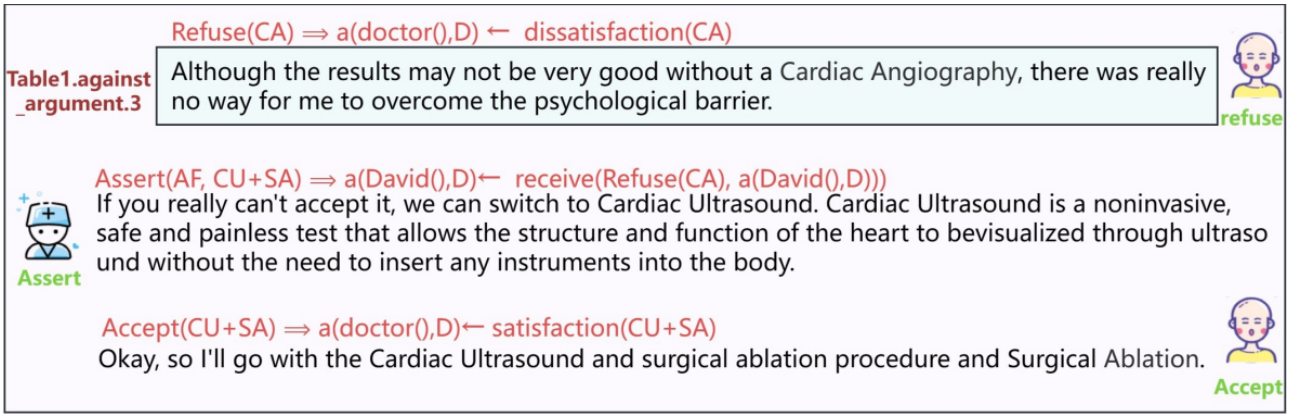

Findings and Insights From Interactions Between Patients and Physicians With Atrial Fibrillation

In the "David Talked to the Doctor and Accepted the Decision" section, the patient raises questions about the 2 surgical procedures due to insufficient understanding. At this stage, the patient's emotions do not express clear opposition, so the system generates a justification based on the clinical guidelines for these procedures, providing relevant arguments to inform the patient. If the patient accepts the treatment plan, the conversation concludes, and the system generates an acceptance to finalize the decision. This combination of the 4 speech acts can be abstracted and recorded as mode A. However, if the patient opposes the surgery or initially rejects it, as shown in the "David Has a Concern but Accepts the Decision of the Doctor" section, the system will generate a Persuade act, utilizing supporting arguments from the clinical guidelines to advocate for the original treatment plan. If the patient accepts the treatment plan, the conversation concludes and the final decision is established (*accept*). Unlike mode A, we can record these 4 speech acts as mode B. If the patient remains unconvinced by the system and continues to oppose the treatment, the system will replace the opposed treatment plan with an alternative from the clinical guidelines. In the "David Had Concerns and the Doctor Accepted His Decision" section, we observe that the patient repeatedly opposes cardiac angiography, and ultimately, the system replaces it with cardiac ultrasound, generating an *assert* to inquire about the patient's acceptance of the new treatment plan. If the treatment plan is not accepted, the process will repeat until the patient ultimately agrees. This creates a dynamic sequence of multiple speech acts, where the total number of speech acts exceeds those in mode A and mode B. We can categorize this process as mode C.

In essence, a variety of distinct decision modes can be constructed based on the combination of speech acts in dialogues. As shown in Figure 14, each mode is abstracted individually. For example, mode A, which aligns with the decision scenario of addressing patient questions, consists of 4 speech acts: *assert*, *question*, *justify*, and *accept*. Within the CoDeL framework, we can explicitly identify and construct reusable decision modes by dynamically combining different speech acts in specific ways. These modes can be applied to various decision scenarios and adapt to all disease problem domains. This process is flexible and extends beyond the 3 modes identified in this study, referred to as mode X.

In this comprehensive process, our approach emphasizes patient-centered decision-making by adapting to their attitudes and concerns. It devises a treatment plan tailored to their specific worries and needs, ensuring personalized care. Simultaneously, the information and treatment plans provided are guided by clinical guidelines, ensuring transparency and evidence-based reasoning in every interaction. By integrating these elements, our approach ultimately enhances communication between doctors and patients.

XSL•FO
RenderX

Figure 14. Reusable decision modes made up of a various kinds of combination of speech acts.

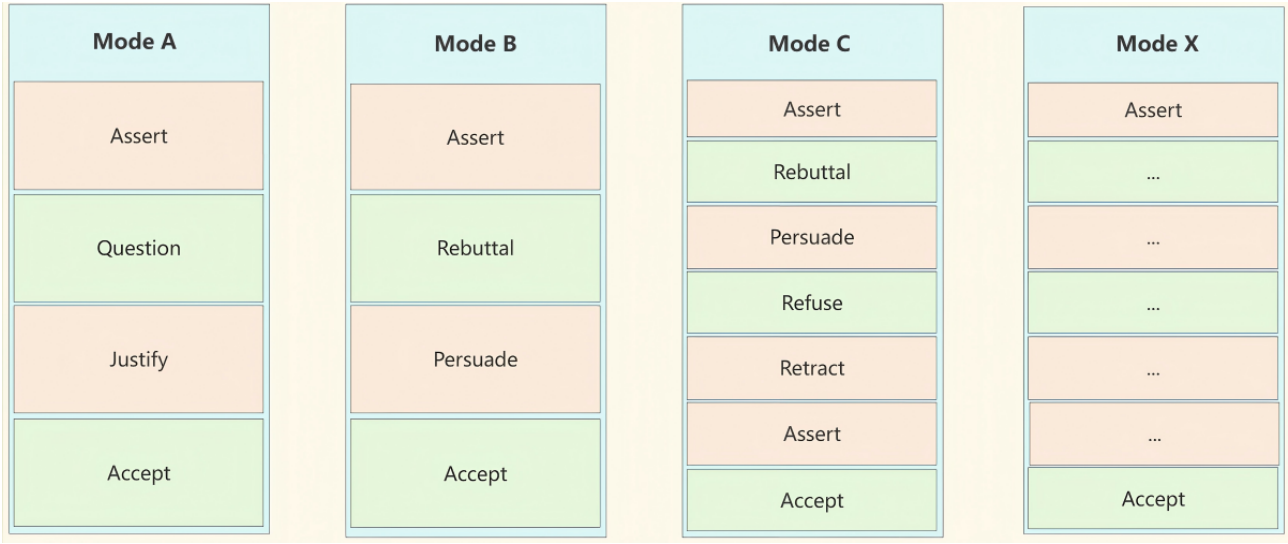

Validation of the Model for Other Cases

To verify the broad applicability of the CoDeL system, this section will use severe depression as an example to fully demonstrate the medical decision-making process for patients with severe depression (Textbox 5).

Based on the above case information, we will provide a complete demonstration of using the Collaborative Decision Description Language (CoDeL). The communication process is shown in Figure 15. Emily, a patient with major depression, discussed her condition with her doctor. The doctor recommended a treatment plan called CBT + antidepressant, which combines cognitive behavioral therapy (CBT) with antidepressant medication. Emily inquired about the treatment process for CBT and the side effects of the antidepressant medication. The doctor explained the process of CBT and the potential risks of antidepressant medication. However, Emily expressed concerns about the side effects of the medication. The doctor then recommended a selective serotonin reuptake inhibitor, which is generally well-tolerated and considered a safe option. Despite this, Emily remained reluctant to take antidepressants and declined the treatment. As a result, the doctor suggested proceeding with CBT alone, without the addition of antidepressant medication. This cautious choice helps mitigate the potential risks of medication. Ultimately, Emily agreed to the plan, decided to proceed with CBT, and was committed to completing the required homework. This marked a new beginning, with the doctor and Emily working together to address her depression.

Textbox 5. Medical decision-making process for patients with severe depression

1. Case description

Emily, a 28-year-old female, has been experiencing intense sadness and a loss of interest in activities she once enjoyed for the past 6 months. These symptoms have progressively worsened, prompting her to seek medical attention and undergo a comprehensive examination at the hospital. A subsequent psychological evaluation led to a diagnosis of major depressive disorder, a common mental health condition marked by prolonged periods of sadness and disinterest in previously enjoyed activities.

During her hospital visit, Emily expressed feelings of hopelessness and worthlessness. She also reported struggling with sleep issues, including both insomnia and oversleeping. Additionally, she mentioned having difficulty concentrating, remembering things, and making decisions. Emily revealed that she has been isolating herself from friends and family, which has further exacerbated her depression.

2. Clinical guideline

According to clinical guidelines, cognitive behavioral therapy is a type of talking therapy that helps individuals manage problems by changing the way they think and behave. It is most commonly used to treat anxiety and depression, but can also be effective for other mental and physical health issues. By contrast, antidepressants are another treatment option for managing the symptoms of depression. Most people with moderate or severe depression benefit from antidepressants, although not everyone responds the same way. If your general practitioner believes you would benefit from an antidepressant, they will typically prescribe a modern medication known as a selective serotonin reuptake inhibitor. These are as effective as older antidepressants but tend to have fewer side effects, though they can cause nausea, headaches, dry mouth, fatigue, sleep problems, and sexual issues. However, these side effects usually improve over time.

XSL•FO
RenderX

Figure 15. The interaction between the doctor and Emily and the corresponding CoDeL (Collaborative Decision Description Language).



Experiment and Evaluation

By instructing GPT to follow the interaction model when communicating with humans, we assume the role of the patient, while GPT takes on the role of the doctor, simulating a realistic doctor-patient interaction. Figure 16 illustrates the exact content and types of conversations that occur during this communication. In this scenario, patients actively challenge doctors, expressing the need for further information and clarification about treatment plans. In Figure 16, we annotate the parameters generated by GPT according to the interaction pattern we defined. The green-marked font in the figure highlights content that differs from the parameters in the interaction pattern. New concepts in the confirmation messages generated by GPT are introduced, which do not include the disease or the treatment proposed by the doctor. Therefore, we will question GPT and seek confirmation from it.

The patient initially showed an aversion to cardiac angiography surgery, influenced by the psychological impact of past experiences. However, through the doctor's persistence and skillful persuasion, the patient was eventually convinced and agreed to the proposed treatment plan. Throughout the persuasion, the doctor emphasized the numerous benefits of cardiac imaging and offered strategies to alleviate the patient's discomfort during the procedure. It is important to note that in

XSL•FO
RenderX

Figure 17, GPT generates new concepts during the persuasion process, prompting us to question them. In turn, GPT confirms our questions and provides additional information.

As the third situation is similar to the one shown in Figure 17, the main difference is that the patient refuses the doctor's persuasion. Therefore, in the interaction shown in Figure 18, the steps before the rejection are omitted, and the rejection message is presented directly from the beginning. In this conversation, GPT, acting as the doctor, introduced a new alternative option—cardiac computed tomography scan—that we had not previously considered. This novel proposal not only expands the range of available options but also highlights the importance of considering different perspectives in the decision-making process.

In many experiments using only GPT dialogues, we found that without CoDeL language as a dialogue framework, GPT struggled to accurately understand the overall relationship of the patient's intentions within the context. As shown in Figure 19, this is a typical experiment using only GPT dialogues. In the figure, we compare the logical markup in the CoDeL language with the doctor-patient dialogue using only GPT. The upper-right corner of each dialogue shows the logical markup in the CoDeL language, with the logical arguments also circled within the dialogue. At the end of the dialogue, the patient only intended to reject the treatment option of cardiac angiography, but GPT rejected both cardiac angiography and surgical ablation, that is, Assert(diseases, surgery3 + surgery4), and chose other treatment methods. GPT was unable to accurately understand the patient's intentions logically, resulting in output text that lacked coherence and explanatory power.

Figure 16. A dialogue process in which the patient agrees directly with the treatment plan proposed by the doctor (GPT).



XSL•FO
RenderX

Figure 17. A dialogue process in which the patient and the doctor negotiate and finally the patient agrees to the treatment plan proposed by the doctor (GPT).

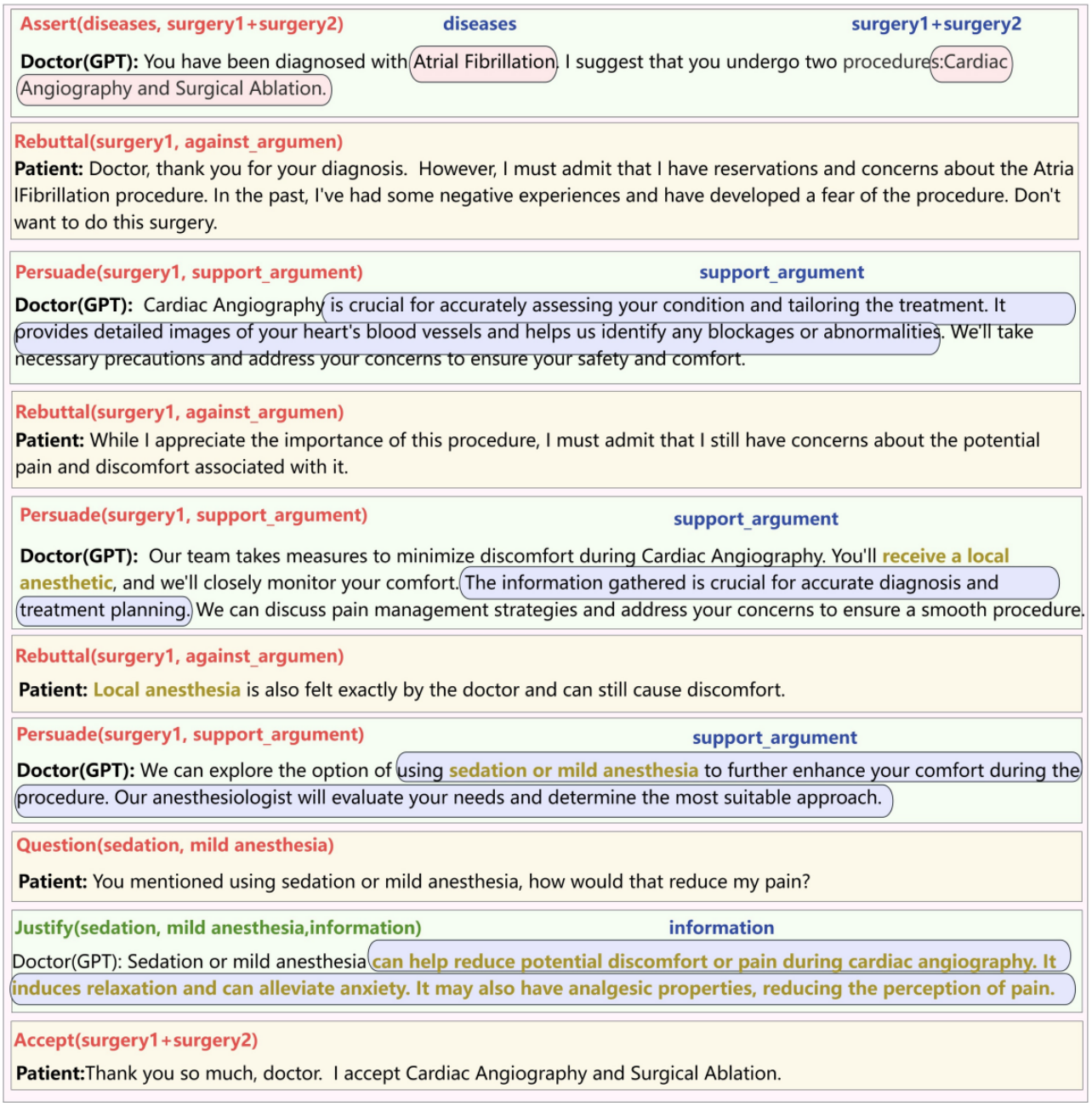

Figure 18. A dialogue process in which the patient and doctor negotiate and ultimately the doctor (GPT) proposes an alternative treatment plan.

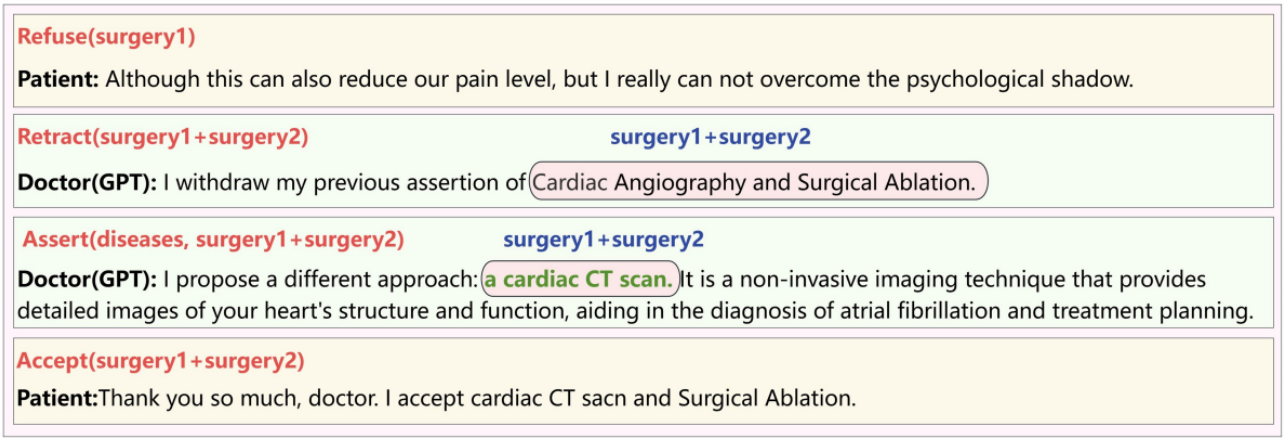

XSL•FO
RenderX

Figure 19. A typical GPT-only interaction experiment without incorporating the CoDel dialogue framework.

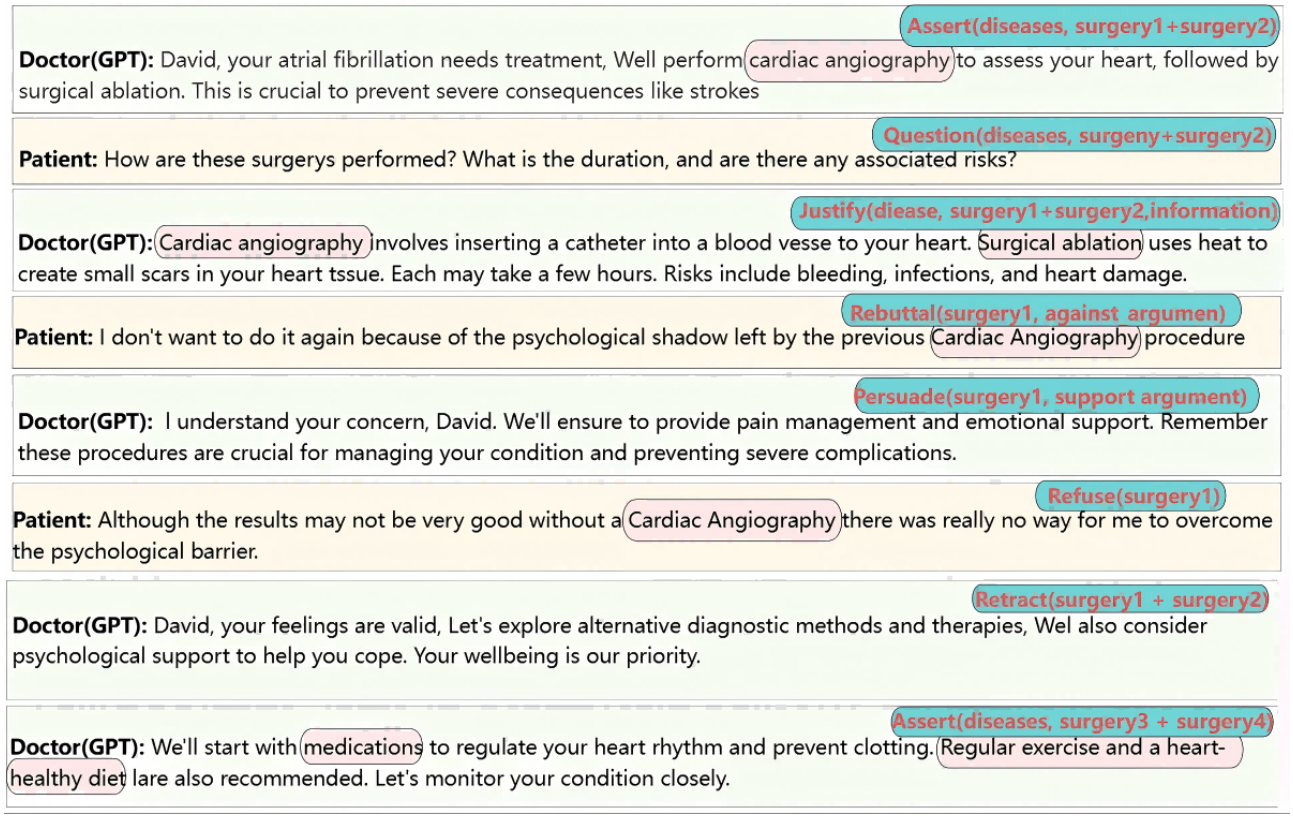

Discussion

Principal Findings

By engaging in a dialogue with GPT using our approach, we can achieve the following outcomes:

1. Verify the efficacy of our method: The interaction with GPT enables us to test and evaluate the effectiveness of our proposed method in generating meaningful and coherent responses.
2. Enhance the comprehensiveness of our argument: Through dialogue, we can identify gaps or limitations in our argument and address them, thereby improving the overall quality and completeness of our approach.
3. Establish new guidelines for CoDeL: The dialogue with GPT provides an opportunity to propose new agreements or principles for CoDeL (Collaborative Decision Description Language) based on insights gained from the interaction. This contributes to the advancement and refinement of CoDeL as a valuable tool for characterizing collaborative decision-making processes.

By using our method to engage in a dialogue with GPT, we can provide the following benefits for GPT.

1. Enhance interpretability: Through interaction, our method allows GPT to gain a deeper understanding and interpretation of the input text, leading to more accurate and contextually relevant responses. This improves the overall performance and interpretability of GPT.
2. Provide a comprehensive set of interaction modes with personalized capabilities: By utilizing our method, GPT gains access to a diverse range of interaction modes, allowing it to engage in meaningful conversations with users across various scenarios and contexts. This broadens the scope of GPT's capabilities and enhances its versatility as a language model.

Based on the CoDeL method, significant improvements can be observed in the communication between doctors and patients. The actual effects are as follows:

1. Enhanced communication between doctors and patients, supported by the SDM protocol, which includes well-defined speech act semantics and arguments explicitly embedded in the interactive reasoning dialogue processes.
2. More effective elicitation of patient concerns throughout the decision-making process, ensuring the delivery of more personalized decision support and care services. This is particularly evident in the interaction between a patient with atrial fibrillation and their doctor, where CoDeL's implementation encouraged the patient to actively participate in deciding whether to choose cardiac angiography as a treatment plan. This ensured that doctors accurately understood the patient's concerns, preventing errors or unforeseen events caused by insufficient communication.

Conclusions and Future Work

In this study, we explored the semantics of patient-physician collaborative decision-making and the issues of information asymmetry to develop a language framework called CoDeL for describing group decisions. We also defined 3 distinct models of physician-patient decision-making. Both these models and CoDeL are designed to offer a clearer understanding of the communication and decision-making processes between

XSL•FO
RenderX

physicians and patients, ultimately leading to more systematic and standardized methods. The main innovations in this study are found in the following 3 aspects:

- Defined a novel shared clinical decision language that utilizes speech act semantics and arguments to enhance communication between doctors and patients, as well as their interactive reasoning process.
- Demonstrated the applicability of the method using a case study of atrial fibrillation, showing that personalized decision support and care services can be delivered.
- Integrated the shared decision language with GPT to enhance its capabilities in interactive decision-making, allowing GPT-generated arguments to be automatically embedded into the decision protocol instead of being manually extracted from clinical evidence.

Although the proposed CoDeL language framework and its applications have the potential to enhance communication and decision-making between patients and physicians, we acknowledge several limitations that may still exist.

The study's limitations are as follows:

- Our study relies on simulated data, meaning that the CoDeL framework has not yet been validated in real-world clinical scenarios.
- While CoDeL has shown potential in multiple clinical domains, its applicability may be limited in certain specific contexts.

In our future work, we will focus on optimizing the decision-making negotiation language, emphasizing conciseness to achieve more efficient and intuitive communication methods. This will significantly accelerate the decision-making process and facilitate smoother integration in future applications. Specifically, we will actively apply CoDeL in clinical practice, leveraging external knowledge sources as evidence-based support, and conduct rigorous scientific research to ensure accurate clinical decision-making. We also aim to provide a common interaction model and negotiation language for collaborative decision-making between doctors and patients, to accommodate a wider variety of cases and enhance the efficiency of the group decision-making process.

Authors' Contributions

LX proposed the general research direction, and XRG organized and drew charts and was a major contributor to writing the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Land V, Parry R, Seymour J. Communication practices that encourage and constrain shared decision making in health-care encounters: systematic review of conversation analytic research. Health Expect 2017 Dec 18;20(6):1228-1247 [FREE Full text] [doi: 10.1111/hex.12557] [Medline: 28520201]
2. Oshima Lee E, Emanuel EJ. Shared decision making to improve care and reduce costs. N Engl J Med 2013 Jan 03;368(1):6-8. [doi: 10.1056/NEJMp1209500] [Medline: 23281971]
3. Fried TR. Shared decision making — finding the sweet spot. N Engl J Med 2016 Jan 14;374(2):104-106. [doi: 10.1056/nejmp1510020]
4. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2014 Oct 28;35(1):114-131. [doi: 10.1177/0272989x14551638]
5. Hoffmann TC, Montori VM, Del Mar C. The connection between evidence-based medicine and shared decision making. JAMA 2014 Oct 01;312(13):1295-1296. [doi: 10.1001/jama.2014.10186] [Medline: 25268434]
6. Seibel K, Valeo SC, Xander C, Adami S, Duerk T, Becker G. Terminally ill patients as customers: the patient's perspective. J Palliat Med 2014 Jan;17(1):11-17. [doi: 10.1089/jpm.2013.0141] [Medline: 24341322]
7. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct 23;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
8. Rosenbaum L. The paternalism preference — choosing unshared decision making. N Engl J Med 2015 Aug 13;373(7):589-592. [doi: 10.1056/nejmp1508418]
9. Légaré F, Thompson-Leduc P. Twelve myths about shared decision making. Patient Educ Couns 2014 Sep;96(3):281-286 [FREE Full text] [doi: 10.1016/j.pec.2014.06.014] [Medline: 25034637]
10. Joseph-Williams N, Lloyd A, Edwards A, Stobbart L, Tomson D, Macphail S, et al. Implementing shared decision making in the NHS: lessons from the MAGIC programme. BMJ 2017 Apr 18;357:j1744 [FREE Full text] [doi: 10.1136/bmj.j1744] [Medline: 28420639]
11. Elwyn G, Durand MA, Song J, Aarts J, Barr PJ, Berger Z, et al. A three-talk model for shared decision making: multistage consultation process. BMJ 2017 Nov 06;359:j4891 [FREE Full text] [doi: 10.1136/bmj.j4891] [Medline: 29109079]

XSL•FO
RenderX

12. Siouta E, Hellström Muhli U, Hedberg B, Broström A, Fossum B, Karlgren K. Patients' experiences of communication and involvement in decision-making about atrial fibrillation treatment in consultations with nurses and physicians. Scand J Caring Sci 2016 Sep 28;30(3):535-546. [doi: 10.1111/scs.12276] [Medline: 26709843]
13. Potts C. Speech acts. In: Cross-Cultural Pragmatics (2021). Oxford, UK: Oxford University Press; Sep 30, 2021.
14. Afzaal A. Identification of speech acts: a linguistic analysis of advertisements in Pakistan. JELLE 2022 Mar 09;3(3):1-14. [doi: 10.54692/jelle.2021.030398]
15. Mády K, Gyuris B, Gärtner H M. Perceived emotions in infant-directed narrative across time and speech acts. 2022 May 23 Presented at: International Conference on Speech Prosody; May 23, 2022; University of Lisbon, Lisbon, Portugal. [doi: 10.21437/speechprosody.2022-120]
16. Murray-Rust D, Robertson D. LSCitter: building social machines by augmenting existing social networks with interaction models. 2014 Apr 7 Presented at: The 23rd International Conference on World Wide Web; April 7-11, 2014; Seoul, Korea. [doi: 10.1145/2567948.2578832]
17. Panisson AR, Farias G, Freitas A, Meneguzzi F, Vieira R, Bordini RH. Planning interactions for agents in argumentation-based negotiation. 2014 Presented at: Eleventh International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2014); May 5, 2014; Paris, France p. 1-15 URL: https://www.mit.edu/~irahwan/argmas/argmas14/w12-05.pdf
18. Yan-rui Z. A flexible job shop scheduling system based on agent negotiation. Microcomput Inf 2011;27(1):14-16.
19. Ateib MT. Agent based negotiation in e-commerce. New York, NY: IEEE; 2010 Jun 15 Presented at: 2010 International Symposium on Information Technology; June 15-17, 2010; Kuala Lumpur, Malaysia p. 861-868. [doi: 10.1109/ITSIM.2010.5561565]
20. Sim KM. Agent-Based Cloud Computing. IEEE Trans. Serv. Comput 2012 Oct 10;5(4):564-577. [doi: 10.1109/TSC.2011.52]
21. Mendes F, Mendes E, Salleh N, Oivo M. Insights on the relationship between decision-making style and personality in software engineering. Information and Software Technology 2021 Aug;136:106586. [doi: 10.1016/j.infsof.2021.106586]
22. Kaplan B. Evaluating informatics applications--clinical decision support systems literature review. Int J Med Inform 2001 Nov;64(1):15-37. [doi: 10.1016/s1386-5056(01)00183-6] [Medline: 11673100]
23. Das S, Kar S. Group decision making in medical system: an intuitionistic fuzzy soft set approach. Applied Soft Computing 2014 Nov;24:196-211. [doi: 10.1016/j.asoc.2014.06.050]
24. Beeler P, Bates D, Hug B. Clinical decision support systems. Swiss Med Wkly 2014 Dec 23;144:w14073 [FREE Full text] [doi: 10.4414/smw.2014.14073] [Medline: 25668157]
25. Schumacher F, Helenowski I, Oswald L, Gonzalez B, Benning J, Morgans A. Treatment decision-making in metastatic prostate cancer: perceptions of locus of control among patient, caregiver, and physician triads. PPA 2022 Jan;Volume 16:235-244. [doi: 10.2147/ppa.s334827]
26. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014 Jan 28(1):CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]
27. Zhang Z, Bickmore T. Medical shared decision making with a virtual agent. 2018 Nov 5 Presented at: The 18th International Conference on Intelligent Virtual Agents; November 5-8, 2018; Sydney, NSW, Australia p. 113-118. [doi: 10.1145/3267851.3267883]
28. Kearns WR, Chi N, Choi YK, Lin S, Thompson H, Demiris G. A systematic review of health dialog systems. Methods Inf Med 2019 Dec 29;58(6):179-193. [doi: 10.1055/s-0040-1708807] [Medline: 32349152]
29. Biswas SS. Role of Chat GPT in public health. Ann Biomed Eng 2023 May 15;51(5):868-869. [doi: 10.1007/s10439-023-03172-7] [Medline: 36920578]
30. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023 Mar 30;388(13):1233-1239. [doi: 10.1056/nejmsr2214184]
31. Xiao L. An agent-oriented group decision architecture. In: Jezic G, Chen-Burger YH, Kusek M, Šperka R, Howlett R, Jain L, editors. Agents and Multi-agent Systems: Technologies and Applications 2019. Smart Innovation, Systems and Technologies, vol 148. Singapore City, Singapore: Springer; Jun 20, 2019:15-33.
32. Ethical Review Measures for Life Science and Medical Research Involving Human Beings. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/qjjys/s7946/202302/c3374c180dc5489d85f95df5b46afaf5.shtml [accessed 2025-03-04]
33. Lévy S, Camm AJ, Saksena S, Aliot E, Breithardt G, Crijns H, Working Group on Arrhythmias, Working Group on Cardiac Pacing of the European Society of Cardiology, North American Society of PacingElectrophysiology. International consensus on nomenclature and classification of atrial fibrillation; a collaborative project of the Working Group on Arrhythmias and the Working Group on Cardiac Pacing of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Europace 2003 Apr;5(2):119-122. [doi: 10.1053/eupc.2002.0300] [Medline: 12633634]

Abbreviations

CBT: cognitive behavioral therapy
CoDeL: Collaborative Decision Description Language

XSL•FO
RenderX

LCC: Lightweight Coordination Calculus
LSC: Lightweight Social Calculus
PtDA: patient decision aid
SDM: shared decision-making

Edited by B Puladi; submitted 10.12.23; peer-reviewed by S Biswas, D Singh; comments to author 04.06.24; revised version received 29.09.24; accepted 30.11.24; published 04.03.25.

Please cite as:
Guo X, Xiao L, Liu X, Chen J, Tong Z, Liu Z
Enhancing Doctor-Patient Shared Decision-Making: Design of a Novel Collaborative Decision Description Language
J Med Internet Res 2025;27:e55341
URL: https://www.jmir.org/2025/1/e55341
doi:10.2196/55341
PMID:40053763

©XiaoRui Guo, Liang Xiao, Xinyu Liu, Jianxia Chen, Zefang Tong, Ziji Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Consumer Engagement With Risk Information on Prescription Drug Social Media Pages: Findings From In-Depth Interviews

Jacqueline B Amoozegar[1], MSPH; Peyton Williams[1], MPH; Kristen C Giombi[1], PhD; Courtney Richardson[1], MPH; Ella Shenkar[1], BA; Rebecca L Watkins[1], BS; Amie C O'Donoghue[2], PhD; Helen W Sullivan[2], MPH, PhD

[1]RTI International, Research Triangle Park, NC, United States

[2]US Food and Drug Administration, Silver Spring, MD, United States

Corresponding Author:
Jacqueline B Amoozegar, MSPH
RTI International
3040 East Cornwallis Road
Research Triangle Park, NC, 27709-2194
United States
Phone: 1 919 541 6000
Email: jamoozegar@rti.org

Abstract

Background: The volume of digital drug promotion has grown over time, and social media has become a source of information about prescription drugs for many consumers. Pharmaceutical companies currently present risk information about prescription drugs they promote in a variety of ways within and across social media platforms. There is scarce research on consumers' interactions with prescription drug promotion on social media, particularly on which features may facilitate or inhibit consumers' ability to find, review, and comprehend drug information. This is concerning because it is critical for consumers to know and weigh drug benefits and risks to be able to make informed decisions regarding medical treatment.

Objective: We aimed to develop an understanding of the user interface (UI) and user experience (UX) of social media pages and posts created by pharmaceutical companies to promote drugs and how UI or UX design features impact consumers' interactions with drug information.

Methods: We conducted in-person interviews with 54 consumers segmented into groups by device type (laptop or mobile phone), social media platform (Facebook or Instagram), and age. Interviewers asked participants to navigate to and review a series of 4 pages and 3 posts on their assigned device and platform. Interviewers encouraged participants to "think aloud," as they interacted with the stimuli during a brief observation period. Following each observation period, participants were asked probing questions. An analyst reviewed video recordings of the observation periods to abstract quantitative interaction data on whether a participant clicked on or viewed risk information at each location it appeared on each page. Participants' responses were organized in a metamatrix, which we used to conduct thematic analysis.

Results: Observational data revealed that 59% of participants using Facebook and 70% of participants using Instagram viewed risk information in at least 1 possible location on average across all pages tested during the observation period. There was not a single location across the Facebook pages that participants commonly clicked on to view risk information. However, a video with scrolling risk information attracted more views than other features. On Instagram, at least half of the participants consistently clicked on the highlighted story with risk information across the pages. Although thematic analysis showed that most participants were able to identify the official pages and risk information for each drug, auto-scrolling text and text size posed barriers to identification and comprehensive review for some participants. Participants generally found it more difficult to identify the drugs' indications than risks. Participants using Instagram more frequently reported challenges identifying risks and indications compared to those using Facebook.

Conclusions: UI or UX design features can facilitate or pose barriers to users' identification, review, and comprehension of the risk information provided on prescription drugs' social media pages and posts.

(J Med Internet Res 2025;27:e67361) doi:10.2196/67361

XSL•FO
RenderX

KEYWORDS

social media; prescription drugs; risk information; safety information; Facebook; Instagram; prescription; risk; information; safety; interview; consumer engagement; digital; drug promotion; user experience; promotion

Introduction

Background

Pharmaceutical companies are present and active on social media. Their multipurpose use of social media platforms for corporate identity, community management, and paid promotion of their products achieves varying levels of measured success [1]. The health care and pharmaceutical industry in the United States spends a significant amount of money on social media promotion. Between July and November 2020, this industry spent US $198.3 million on Facebook promotion and US $151.5 million on Instagram posts [2]. The size of these expenditures is particularly noteworthy, given that the cost of digital promotion is considerably lower than other mediums (eg, television).

Consumers can use social media to find information about prescription drugs in a variety of ways, including searching for information from pharmaceutical companies, reading reviews and testimonials, and interacting with other consumers. A nationally representative survey of 1744 US adults found that 26% of respondents had seen a prescription drug advertised on social media in the past 3 months, and 9% had watched a video on the web about a prescription drug on a social networking website [3]. Another survey of 1000 US consumers found that 9% used social media to evaluate new treatment options, and 7% used social media to look for reliable information about medication side effects [4]. A systematic review found that consumers use social media as a complement to a physician's treatment, particularly around social, emotional, esteem, network, and information support [5].

Despite these increases in the volume of digital drug promotion, the amount of research on consumers' use of and trust in prescription drug promotion on social media is scarce, particularly on user interface (UI) and user experience (UX) features associated with accurate understanding of product claims and risks. This lack of evidence is problematic, given that the information shared on these platforms could have a large impact on the treatment decisions of consumers. For example, when this information is shared on television, one study found that exposure to prescription drug promotion for as little as 5 minutes increased consumers' intentions to search for medications on the web or switch medications [6]. Another study found that reading a prescription drug promotion on Twitter was associated with college students obtaining a prescription medication without first obtaining a prescription from a doctor [7].

Prescription drug promotion is regulated by the Food and Drug Administration [8]. According to Food and Drug Administration regulations, when drug benefits are promoted, drug risks must also be described to achieve a "fair balance" [9]. Ensuring "fair balance" on an interactive medium like social media—where there could be character space limitations, targeted newsfeed advertisements, and consumer interaction through comments, shares, and likes—is challenging.

Pharmaceutical companies currently present risk information (often referred to as "safety information" in promotion) about the prescription drugs they promote in a variety of ways across social media platforms. For example, companies may use links to provide access to risk information, such as a link to the drug's external website or a link to a PDF file with the full prescribing and medication guide. Risk information is also sometimes conveyed in a featured text post, a scrolling video post, a risk information pop-up, or a highlighted story.

Many pharmaceutical companies place the risk information in multiple places on their social media pages and posts. However, even when this information is displayed in various locations, UI or UX features of social media platforms may still inhibit consumers' ability to find, review, and comprehend the risk information. This is concerning, given that it is critical for consumers to know and weigh drug benefits and risks to be able to make informed decisions regarding their medical treatment.

Study Purpose and Value

With the increasing use of social media for promotions, understanding the UI or UX design and its impact on how consumers interact with the pages and posts created by pharmaceutical companies to promote drugs is becoming increasingly important. Accordingly, we examined how different approaches to presenting information about the risks of promoted drugs on social media pages and posts influence how consumers engage with the content. Specifically, we assessed how consumers' exposure to and perceptions of risk information was influenced by where and how on the social media pages and posts the information was conveyed. Further, we sought to identify differences in experiences and perceptions by user age as well as the type of device being used to access the pages and posts.

Methods

Study Design

We developed our plan and instruments for conducting in-person, one-on-one interviews with a total of 54 individuals based on our research questions and a review of relevant peer-reviewed literature. As shown in Table 1, study participants were segmented based on three criteria: (1) device type, (2) social media platform, and (3) age. For device type, we segmented participants into reviewing stimuli on either a mobile device (an Apple iPhone 6s running iOS 15.8) or a laptop computer (an Acer Aspire E5-571 laptop with a 15.6-inch screen running Windows 10). For social media platform, we segmented participants to review stimuli on either Facebook or Instagram. For age, participants in the Facebook segment were further segmented into either "adults" (18-61 years of age) or "older adults" (62 years of age or older). Participants in the Instagram segment were further segmented into either "younger adults"

XSL•FO
RenderX

(18-24 years of age) or "adults" (25 years of age or older). The different definitions of "adults" for Instagram versus Facebook reflect that users of Instagram are typically younger [10].

Table 1. Sample segmentation by device type used, social media platform used, and age.

	Device type		Social media platform		Age cohort		
	Mobile	Laptop	Facebook	Instagram	18+ years	18-24 years	62+ years
Group 1 (n=9)	✓		✓		✓		
Group 2 (n=9)		✓	✓		✓		
Group 3 (n=9)	✓			✓	✓		
Group 4 (n=9)	✓			✓		✓	
Group 5 (n=9)	✓		✓				✓
Group 6 (n=9)		✓	✓				✓

Recruitment and Screening Process

To recruit participants, we partnered with a marketing research firm with access to a consumer panel with a national population. This firm emailed adult panelists an invitation to participate. Interested panelists clicked on a hyperlink within the invitation to complete a web-based screener to determine eligibility. If deemed eligible by the web-based screener, participants were contacted by phone to complete a more comprehensive screening process.

To be eligible, participants needed to use either Facebook or Instagram 4 or 5 days per week on a smartphone or laptop or desktop. Participants also had to self-report a type 2 diabetes diagnosis by a medical professional to increase the likelihood that participants would be motivated to learn about some of the products included in the stimuli, which would in turn increase the realism and thus how generalizable our study findings are to the real world and everyday life. We encountered difficulties recruiting persons with type 2 diabetes for the 18- to 24-year age group assigned to view Instagram stimuli on mobile devices (group 4) and removed the diabetes requirement for this one group as a result.

Because the interviews were conducted in person, participants needed to be able to drive or otherwise be transported to the study location (a private meeting room in an office setting) in Raleigh, North Carolina. To reduce the risk of bias, we excluded participants who worked in marketing, in the pharmaceutical industry, or for the US Department of Health and Human Services. Individuals who had participated in a focus group or interview-based research study in the previous 3 months were also ineligible to participate.

Ethical Considerations

This study was determined to be exempt under category 2ii (tests, surveys, interviews, or observation) by RTI International's institutional review board (STUDY00022581) on August 14, 2023. Eligible participants were sent a consent form in advance of the interview providing information about the protection of their privacy and the confidentiality of their data. Additionally, the interviewer reviewed the information presented in the consent form before starting each interview and obtained the participant's verbal consent to participate and have their voice and screen interactions recorded. Participants were informed that neither the transcripts produced based on interview recordings nor any reporting of findings would include identifying information. No potential participants declined to consent or dropped out after consenting. Upon completion of each interview, the participant was sent a US $75 incentive in the form of a digital gift card.

Stimuli

We identified existing social media pages and posts created by pharmaceutical companies for the prescription drugs they manufacture. We focused on Facebook and Instagram pages and posts based on use trends and to limit the number of combinations of variables to be assessed [10].

We selected pages and posts about prescription drugs with a mix of indications, including some that were relevant, and possibly familiar, to many of our participants (ie, drugs indicated to treat diabetes) and some that were not (eg, drugs indicated to treat high cholesterol). To avoid endorsing specific products, we have anonymized the stimuli and referred to the pages and posts by numbers (ie, rather than the promoted drugs' brand names).

We began each interview by asking all participants to review the same "baseline" stimuli page for one of the selected prescription drugs not indicated to treat diabetes (with which they were unlikely to be familiar) with the hope that they would have to explore the content in more depth to learn about the indications, risks, and benefits of a drug that are novel to them.

In total, 3 pages and 3 posts for each of the 2 platforms (ie, Facebook and Instagram) served as the remainder of our stimuli. Key characteristics of these pages and posts are shown in Tables 2 and 3, respectively. Working within the confines of selecting only real-life pages and posts, we sought to achieve the best possible variation in stimuli. For instance, for Facebook pages, we sought pages where engagement policies varied (ie, whether consumers were allowed to comment on or react to posts). For posts, we strove for variation in how risk information was shown, such as whether it appeared in static text or scrolling text in a video.

XSL•FO
RenderX

Table 2. Key characteristics of the study stimuli: prescription drug–focused pages on Facebook and Instagram.

Characteristic	Facebook				Instagram			
	Baseline page	Page #1	Page #2	Page #3	Baseline page	Page #1	Page #2	Page #3[a]
Condition	Multiple autoimmune disorders	Diabetes	Diabetes	High cholesterol	Multiple autoimmune disorders	Diabetes	Diabetes	High cholesterol
Audience engagement allowed on posts	No	Yes	No	No	No	Yes	Yes	Unknown (content removed)
Verified account	Yes	Yes	Yes	Yes	Yes	No	No	Yes

[a]We excluded this page from our analysis because all content (except the risk information pinned post and an external link) was deleted from the page between when we selected it and when we started testing. With limited places to click, all participants who viewed this page clicked on the risk information highlighted story.

Table 3. Key characteristics of the study stimuli: prescription drug–focused posts on Facebook and Instagram.

Characteristic	Facebook			Instagram		
	Post #1	Post #2	Post #3	Post #1	Post #2	Post #3
Type of post	Video	Text only	Video	Video	Photo	Video
Condition	Diabetes	Diabetes	High cholesterol	Diabetes	Diabetes	Diabetes
Risk information is included	Yes	Yes	Yes	Yes	Yes	Yes
Post focuses exclusively on risk information	Yes	Yes	No	No	No	No
Risk information is scrolling	Yes (entire frame scrolls)	No	Yes (bottom one-third of video scrolls)	No	N/A[a]	Yes (some information appears in static text with voice-over; additional information scrolls with no voiceover)
Risk information is shown immediately in the video	Yes	N/A	Yes	No	N/A	Yes (a note advises viewer that safety info appears at the end of the video)
Risk information is shown in caption	No	N/A	Only a hyperlink is shown in caption	No	Yes	Yes
Must click “See more” to view all warnings	N/A	N/A	N/A	No	Yes	Yes
Allows comments on posts	No	Yes	No	Yes	Yes	Yes

[a]N/A: not applicable.

Data Collection

We conducted 54 in-depth, in-person interviews using a semistructured interview guide during October and November 2023. Each interview lasted approximately an hour and was conducted by 1 of 3 experienced qualitative researchers (PW, KCG, and RLW) employed by an independent, nonprofit research institute.

Based on the participant’s assigned group, the interviewer asked each participant to use the study mobile phone or laptop to navigate to stimuli on the assigned social media platform. A camera was positioned above the phone screen to record the surface of the phone as well as the fingers of the participants and where they were tapping on the phone. For both devices, the screen was shared, recorded, and livestreamed via Zoom (Zoom Video Communications) to enable the other members of the research team to observe.

First, interviewers directed participants to the baseline drug’s page on Facebook or Instagram (depending on the participant’s assigned segment) via the platform’s search function. Using a think-aloud approach, interviewers encouraged participants to narrate their thoughts and feelings on the social media content as they interacted with the study stimuli [11]. The interviewers allowed participants to explore the page and posts on their own for 3 to 5 minutes, which we called the “observation period.” After the observation period, the interviewers asked probing questions.

Second, participants reviewed the 3 subsequent drugs’ pages. Again, the interviewers allowed an observation period for each page before asking probing questions. After participants had looked at all the pages, the interviewers guided participants to navigate to predetermined posts. The interviewers shuffled the order of the stimuli pages and posts to be reviewed between interviews (with the exception of the baseline stimuli page) so that the sequence was sufficiently random to prevent order effects.

XSL•FO
RenderX

Quantitative Analysis

After data collection was complete, an analyst watched the video recordings of the screenshare to abstract interaction data collected during the observation period. We only abstracted interaction data from the observation period, given that this was the only time participants were able to freely explore the pages on their own.

Because risk information could appear in multiple locations, we identified all locations within each of the stimuli where risk information was displayed or linked before beginning data abstraction. Locations varied greatly across stimuli and included approaches such as links to PDFs, videos that displayed risk information, and posts with static written risk information.

During data abstraction, the analyst recorded whether a participant clicked on or viewed risk information at each location. We calculated the following metrics:

- Ever attentive: For a given drug's social media page, risk information could be shown in multiple locations. "Ever attentive" is the percentage of participants who viewed the risk information for a given drug's page in at least 1 of those locations during the observation period. For example, 10% ever attentive means that 10% of participants viewed risk information in at least 1 location for a given drug's page. This metric helped us determine whether consumers see and engage with information about risks.
- Locations viewed: This metric is more granular and helps us understand the exact locations participants viewed the risk information. "Locations viewed" is the percentage of participants who viewed risk information at a particular location or feature. This could have involved actively clicking on a link or been more passive, such as having the risk information appear upon scrolling down. This metric helped us identify the features or characteristics of the platform associated with consumers' review of risk information.

When abstracting data, we did not code the participant's interactions with social media content for the final page that they reviewed. We opted for this approach because we noticed that most participants learned over the course of the interview that we would be asking them about the drugs' risk information and began proactively seeking this information upon switching to a new stimulus in preparation to answer our questions. Similarly, we did not code participants' interactions with the predetermined posts that participants navigated to after reviewing the final page.

Qualitative Analysis

During each interview, we entered deidentified participant responses into a metamatrix that organized responses by participant ID, interview topic, and segment (ie, device type, platform, and user age). We selected a metamatrix approach because this usability study used interview notes (which were most efficiently captured in Microsoft Excel by notetakers) as our unit of analysis (ie, vs using the transcripts themselves). Unlike the transcripts, the notes captured observational data reflecting what notetakers saw when the user interacted with the pages and posts (eg, the user scrolled past relevant information or clicked around to locate information). The metamatrix enabled us to see multiple participant answers across a single topic and question, which facilitated the identification of themes or consistencies in responses [12].

Following data collection, 3 analysts conducted a thematic analysis to identify commonalities and patterns and to summarize the findings for all pages and posts [13]. Organizing the matrix by segment facilitated the identification of any differences between the segments. Given the sample size and the fact that participant responses were not formally coded using computer software, qualitative themes were not associated with participant counts. A general sense of magnitude is included by describing what portion of participants mentioned each theme (eg, most participants, many participants, and a few participants) [14,15].

Results

Participant Characteristics

The 54 consumers who participated in interviews were reasonably varied with regard to demographic characteristics. Most study participants self-identified as male (n=33, 61%). Of the 36 participants assigned to view Facebook stimuli, 44% (n=16) were in the "adult" segment (18-61 years of age), and 56% (n=20) were considered "older adults" (62 years of age or older). The 18 participants assigned to view Instagram stimuli were evenly split into the "younger adult" (18-24 years of age) and "adult" (25 years of age or older) segments. The mean age for participants viewing Facebook stimuli (60.1, SD 12.16 years) was approximately 22 years older than the mean age for participants viewing Instagram stimuli (38.1, SD 18.05 years). Most of the participants were White (n=39, 72%), and nearly one-fourth (n=13, 24%) were Black or African American. Finally, most participants held an undergraduate or postgraduate degree (n=32, 60%).

Key Quantitative Findings

Results of the analysis of our quantitative, observational data are presented for each of the 2 metrics described in our quantitative analysis methods earlier (ie, ever attentive and locations viewed) and further organized by platform. Differences by device type have been noted as appropriate.

Ever Attentive to Risk Information

Facebook

On average, across all the drugs' pages that we tested on Facebook, 59% of participants were ever attentive to risk information during the observation period (Table 4). Of interest, a higher proportion of the participants using the laptop (72%) were attentive to risk information compared to less than half of the participants using the mobile phone (46%).

XSL•FO
RenderX

Table 4. Participants who were ever attentive to risk information by platform, device, and prescription drug page.

Platform and device	Baseline page, n (%)	Page #1, n (%)	Page #2, n (%)	Page #3, n (%)	Average (%) (SD)
Facebook					
Overall	18 (50)	15 (71)	19 (63)	11 (55)	59 (3.99)
Laptop	10 (56)	8 (80)	12 (86)	8 (73)	72 (5.62)
Mobile	8 (44)	7 (64)	7 (44)	3 (33)	46 (5.59)
Instagram					
Mobile	18 (67)	14 (79)	15 (67)	N/A[a]	70 (3.27)

[a]N/A: not applicable.

Instagram

On average, across all the drugs' pages that we tested on Instagram, 70% of participants were ever attentive to risk information. Notably, a higher proportion of participants (70%) viewing Instagram pages on a mobile phone were ever attentive to information about risks as compared to Facebook mobile (46%).

Risk Information of Locations Viewed

Facebook

Each Facebook page we tested had nuances in how and where the pharmaceutical company presented risk information, making it difficult to summarize commonalities and differences among them. We did, however, observe 4 common locations where pharmaceutical companies displayed risk information on Facebook (Figure 1).

Figure 1. Common locations of risk information on Facebook pages about prescription drugs (as it appears on a laptop).

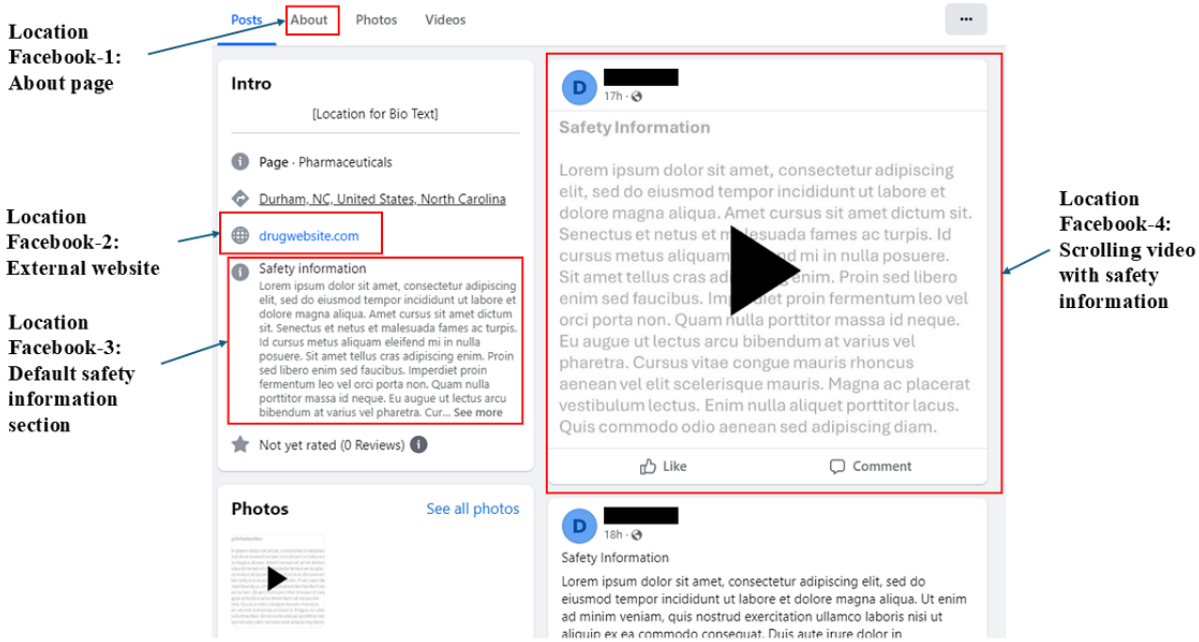

- Location Facebook-1: "About" page. This page was accessed by clicking on "About" from the "Posts" page. If a page is set up as a "Pharmaceuticals" page, Facebook automatically gives the page creator the ability to activate a "Safety Information" section on the "About" page. Risk information was frequently placed here, although it is not required by Facebook.
- Location Facebook-2: external website. On the "Posts" page, pharmaceutical companies often included a link to their drug's external website. Companies frequently display risk information at the bottom of their external website landing page.
- Location Facebook-3: default "Safety Information" section in the introduction pane. By default, once a pharmaceutical company adds the "Safety Information" section to the "About" page, Facebook then adds a section entitled "Safety Information" to the introduction pane on the "Posts" page. All text that is added in location Facebook-1 is automatically duplicated here. If a participant clicks "See More," a pop-up will appear with the full text.
- Location Facebook-4: scrolling video. Pharmaceutical companies frequently create videos of scrolling textual risk information.

XSL•FO
RenderX

We expected location Facebook-3—the "Safety Information" section in the introduction pane—to have a consistently high level of engagement because it is clearly labeled, but we were surprised to find a wide range in the proportion of participants who viewed it across the drugs (Table 5).

A video with scrolling risk information seemed to be the location most often viewed with risk information. For Facebook on a laptop, at least 40% (n=4) of participants consistently viewed risk information when it was presented in a scrolling video, potentially because the motion caught their attention. On the other hand, the percentage of participants who viewed the scrolling video on the mobile device was lower across the 3 pages using this approach (between n=2, 11% and n=4, 36%).

We found that if 2 or more links to external websites were presented together, participants often only clicked the first link. For example, in situations where links to the full prescribing information and risk information appeared one after another, participants who engaged with either link would always click the first one, but not the second. Furthermore, when the link went to the full prescribing information, participants did not find the risk information or medication guide in the PDF.

Table 5. Participants who viewed risk information on Facebook and Instagram by common location on prescription drug–focused pages.

Location and device	Baseline page, n (%)	Page #1, n (%)	Page #2, n (%)	Page #3, n (%)
Facebook				
Location Facebook-1: "About" page				
Laptop	N/A[a]	2 (20)	3 (25)	0 (0)
Mobile	N/A	2 (18)	1 (6)	1 (11)
Location Facebook-2: external website				
Laptop	2 (11)	0 (0)	3 (25)	5 (45)
Mobile	6 (33)	3 (27)	5 (31)	3 (33)
Location Facebook-3: default "Safety Information" section in the introduction pane of the "Posts" page				
Laptop	N/A	2 (20)	8 (67)	1 (9)
Mobile	N/A	1 (9)	8 (50)	0 (0)
Location Facebook-4: scrolling video				
Laptop	10 (56)	4 (40)	N/A	6 (55)
Mobile	2 (11)	4 (36)	N/A	3 (33)
Instagram				
Location Instagram-1: external website link				
Mobile	9 (50)	4 (27)	8 (57)	N/A
Location Instagram-2: highlighted story with risk information				
Mobile	9 (50)	9 (60)	7 (50)	N/A

[a]N/A: not applicable.

Instagram

Unlike Facebook, Instagram does not have a reserved section dedicated to risk information. This means that pharmaceutical companies must find alternative ways of displaying risk information. There were 2 main locations that pharmaceutical companies used to convey this information on the profile page, which is the main page a person reaches when visiting a drug's Instagram page (Figure 2).

XSL•FO
RenderX

Figure 2. Common locations of risk information on Instagram pages about prescription drugs.

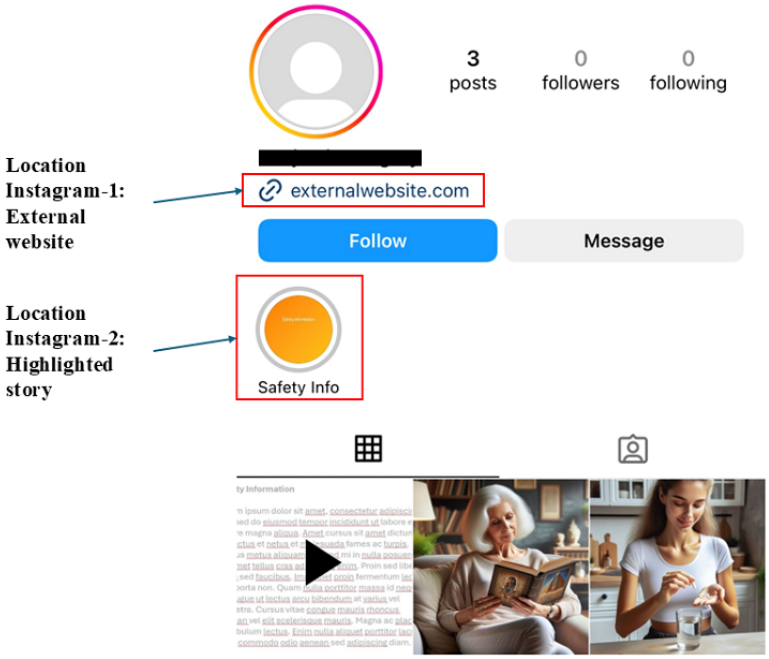

- Location Instagram-1: external website. Instagram allows one or more external links to be included on the profile page.
- Location Instagram-2: highlighted story with risk information. Every drug page on Instagram that we tested created a highlighted story with risk information. A highlighted story is a story that is pinned at the top of the page. These stories can have one or more screens.

Location Instagram-2—highlighted stories—was a common place for participants to view risk information. For all 3 drugs' pages, at least half of the participants consistently clicked on the highlighted story with risk information.

Key Qualitative Findings

Results from our analysis of participants' qualitative interview responses are presented by topic: (1) identification of official pages and verified accounts, (2) ease of locating risk information, (3) factors affecting comprehensiveness of review of risk information, and (4) identification of product risks and indications. Quotations illustrating key themes have been included as appropriate.

Identification of Official Pages and Verified Accounts

Regardless of platform, the majority of participants were able to identify the official page for each drug when searching for it without assistance from the interviewer. Some participants reported that they could tell the page was official because they saw a registered trademark symbol. Participants also reported relying on various other indicators such as the presence of a blue checkmark, the number of followers, the number of posts, and the pharmaceutical company's branding (ie, colors and logos). Participants further mentioned noticing details such as the generic name of the drug, external links to pharmaceutical company websites, and specific medical terms (eg, glucagon and insulin). As one participant said, "It has a Linktree, a lot of followers, and they have 420 posts, which makes me think it's official" [Instagram, mobile, adult]. Finally, another sign participants viewing pages on Facebook noticed was the "Pharmaceuticals" page type in the profile.

When asked about the meaning of the blue checkmark next to the drug's name, participants' responses were mixed. Some participants were unsure about its meaning. Others explained that the blue checkmark meant that the page was real and legitimate or that it confirmed that the page's creator is indeed who they say they are. One participant described their interpretation of the checkmark's implications as: "It's not somebody else with their opinion of it; it's more sort of facts about the drug as the drug company sees it based on the negative reactions that the drug companies had, and they've done their studies" [Facebook, mobile, older adult]. Most participants who reported uncertainty about the meaning of the blue checkmark were older adults viewing Facebook pages.

Participants' responses were also mixed about whether an account being verified influenced their level of trust in the information provided on the page. Some participants reported that account verification has no effect on how much they trust the information. Older adults who viewed Facebook pages were more likely to report that an account being verified influenced their trust of the information compared to adults in the younger age segment. Similarly, none of the young adults who viewed Instagram pages reported that the account being verified affected their trust of the information. One participant viewing Instagram stimuli explained their lack of trust in the blue checkmark: "It doesn't seem like an actual verification since you can pay for it!" [Instagram, mobile, adult].

Ease of Locating Risk Information

As participants finished reviewing each page or post, we asked how difficult it was to locate the risk information for the drug.

XSL•FO
RenderX

Participants commonly reported that it was easy to find the risk information because the information appeared in multiple places or near the top of the page or post. One young adult participant viewing Instagram stimuli explained that it was "not hard, just because it was pinned at the top, and typically things that are pinned are most important" [Instagram, mobile]. However, some participants did find the risk information challenging to locate or at least reported that it took more effort to find information on these drug pages than on other types of pages. We noticed frustration when participants clicked on a link where they expected to find patient safety information but instead found the full prescribing information (ie, information intended for a clinical audience). Furthermore, participants noted at times, especially with third-party multilink services, that they had difficulty finding risk and patient-oriented information.

We asked participants to identify ways this important information could be made easier to find. Participants offered multiple ideas, including changing the formatting of the information by using bold, highlighted, colorful, or larger text; applying section headers or other descriptive labels clearly identifying the topic; using pinned or featured posts; adding risk information as an option in the menu bar (Facebook); and including links in additional places.

Factors Affecting Comprehensiveness of Review of Risk Information

After viewing each of 3 consecutive posts on their assigned platform, participants were asked whether they would normally watch or read the entire post. Participants' responses were mixed across platforms and posts. Participants who reported that they would normally watch or read the entire post commonly clarified that it would depend on whether they were interested in the specific drug highlighted in the post. Some participants who reported that they would not watch or read the entire post indicated that they would just skim it for specific information. Of those who said they would skip the post entirely, a few said they would learn about the drug from an information source other than social media (eg, the drug's website and their physician).

We asked participants who viewed the posts on Facebook whether they were more or less likely to read all the information in the post if it was auto-scrolling and whether they would prefer the post to scroll automatically or not. A few participants reported that they preferred scrolling information (and claimed they would be more likely to read it) or that they did not have any preference about scrolling. However, participants more commonly reported being less likely to fully read scrolling risk information because they prefer to read at their own pace or because they found the speed of scrolling or font size of the scrolling information to be less than ideal. Although most participants knew how to pause a scrolling video, they preferred to avoid the need for pausing or rewinding, especially if they were seeking specific pieces of information. One participant who said that they were less likely to read scrolling information explained, "I don't like the scrolling. I want to go at my speed. If I don't have issues with my gallbladder, I want to skip that information. I don't want to read about gallbladder issues" [Facebook, mobile, adult]. Another participant who reacted negatively to scrolling indicated, "When I saw the scrolling, I saw the pace which I thought was too fast and then I thought, is there something else that could help me? And I saw the links ..." [Facebook, computer, older adult].

After viewing all the Instagram posts and their variety of approaches to showing risk information, participants were asked which they were most and least likely to fully read. Of the posts, 2 presented risk information in static text; participants typically reported one of these 2 posts as being the one they were most likely to read fully. Beyond aversions to scrolling, participants also cited the volume and quality of information as well as how concisely it was written, as factors affecting the likelihood that they would fully review it.

Identification of Product Risks and Indications

After reviewing the baseline drug's page, participants were asked if they felt that they knew the drug's risks and indications. Regardless of whether they viewed the page on Facebook or Instagram, most participants were able to locate the drug's risks without any difficulty. Participants commonly demonstrated that they had successfully found the risks by reading several of them aloud. Some participants clarified that they would need to fully review the risk information to feel like they really "knew" the risks. A few acknowledged that they had not reviewed the page very closely when perusing the drug information.

Similarly, when reviewing the subsequent drugs' pages, nearly all participants indicated that they were able to find the risk information easily and felt that they knew the risks of the drugs. We also asked participants to imagine that they were worried about a specific side effect and to explain whether they would know where to look to find this information. Almost all participants demonstrated that they knew where to find information about the specified side effects.

We asked participants what the risks of a given drug were. Instagram users more frequently reported that they did not know the risks compared to Facebook users. Of the participants who reported that they did not feel that they knew, or could easily identify, the drugs' risks, scrolling and small text size were commonly cited as reasons irrespective of platform. One participant viewing Facebook stimuli remarked, "If I could read this little fine print here, it's probably got it all in there" [Computer, older adult]. Other explanations specific to Instagram included the perception that the drug pages were uninformative and issues with risk information being obscured by other page elements.

Participants were also asked if they felt they knew "what the drug is for" (ie, its indications). Overall, participants generally experienced more difficulty finding this information than they had in finding the risks. However, most participants were ultimately successful in identifying the indications or at least in identifying where they could obtain that information. Participants viewing the baseline drug's page on Facebook were often able to identify information about the drug's indications more quickly and easily than participants viewing it on Instagram. That said, participants viewing the pages on Facebook commonly visited the external website to obtain this

XSL•FO
RenderX

information. Participants viewing Instagram stimuli frequently attempted (and failed) to decipher the indications based on posted images. For example, one participant viewing Instagram stimuli noted that "Just looking at the page, I'd say weight loss and energy ... They look like they're on a road trip or rollercoaster, so maybe it's for the heart or something?" [Mobile, young adult].

In addition to the size of the content, other reasons participants provided for not feeling aware of the drug's risks and indications after their review included the perception that the page's information was not sufficiently comprehensive or credible (ie, that it did not mention all relevant risks). Some participants indicated that they would visit the drug's website for details versus relying solely on the information on any social media page. Several participants (a few on each platform) found scrolling to be a barrier to their efforts to read risk information for understanding. One participant viewing Instagram stimuli explained: "I guess you can also click on the safety story, but I would rather click the link because personally I don't like pausing to stop the scrolling post. I'm sure it's slow enough to read, but just a personal preference" [Mobile, young adult]. All participants viewing Facebook pages who voiced frustration with scrolling risk information were older adults.

Discussion

This unique mixed methods usability study addresses a gap in the literature by enhancing our understanding of consumers' interactions with and perceptions of prescription drug information on social media pages and posts created by pharmaceutical companies, which is a topic of growing importance, given recent, substantial increases in the use of social media for drug promotion.

Key Findings and Implications

Observational data revealed that the majority of participants viewed risk information in at least 1 possible location on average across all pages tested during the observation period. However, there was not a single location across the Facebook pages that participants commonly clicked on to view risk information, likely because of high variability in where risk information was located and how it was labeled. Further, we found that the characteristics of the pages themselves can impede efforts to locate and comprehensively review information. With third-party multilink services or when an external link led participants to the full prescribing information, participants did not find the risk information or medication guide. Furthermore, if 2 or more links to an external website were presented together, participants often only clicked the first link even if they subsequently could not find risk information on that website. More consistency across pages and posts may help consumers find drug information.

Although thematic analysis showed that most participants were able to identify the official pages and risk information for each drug, auto-scrolling text and text size posed barriers to identification and comprehensive review for some participants. Participants sometimes reported that risk information was scrolling at a suboptimal speed or presented in a font size that was too small to be read. Occasionally, part of the text was obscured due to inadequate margins. However, a video with scrolling risk information attracted more views than other features. We found that risk information on Facebook pages was consistently viewed more frequently on a laptop than on a mobile device, suggesting that having more screen real estate may make reviewing risk information easier. On Instagram, at least half of the participants consistently clicked on the highlighted story with risk information across the pages. These findings indicate that the layout and formatting of pages and posts can also affect participants' access to drug information.

Prior research has shown that when consumers do successfully find and read the information on pharmaceutical companies' social media pages, an assortment of other issues can come into play that may impede the use of that information, such as a lack of quality or reliability, information overload, and the inability to correctly apply information found on the web to the users' personal health situation [16]. Along these lines, some participants in our study voiced concerns about the clarity, level of detail, and credibility of the safety information they reviewed during our interviews.

Often social media platforms use a symbol, such as a blue checkmark, to indicate that the account is verified or the page is "official." We found that older adults were less aware of the meaning of the blue checkmark compared to younger participants. This could mean that older adults struggle to locate official drug pages, making them potentially at greater risk of consuming incorrect information from unofficial sources. We also found that older adults generally had a harder time reviewing auto-scrolling risk information and were also less likely than younger participants to know how to pause the scrolling. Part of the reason older adults struggled may be that they are less savvy at navigating social media because they use this information source less than their younger counterparts [16-18]. In summary, our study demonstrates that meaningful barriers remain to consumers' effective use of social media as a source of prescription drug risk information. To maximize the potential of social media to effectively convey information about prescription drugs, it is essential that the information be made as easy as possible to find, read, understand, and use.

Strengths and Limitations

This study had several strengths. First, the study's mixed methods nature enabled us to achieve a comprehensive understanding of users' perspectives by synthesizing quantitative observational data with qualitative data from semistructured discussions, follow-ups, and probes while maximizing reliability and validity. Second, this study's sample was on the larger side for an interview-based usability study, and it included a wide range of age groupings to account for generational differences in digital interactions. Given that the majority of US adults now report using or having used some form of social media [19], a truly representative population for this research should encompass all ages. Finally, we implemented multiple measures to ensure that the stimuli had relevance to the study population and that the interactions were as realistic as possible so that we could assess important differences users may experience, as they navigate this sort of content across different platforms or

XSL•FO
RenderX

with different devices. These measures included assigning participants to use the platform and device they typically use for social media, using actual pages and posts current and live at the time of the interview as stimuli, and including stimuli pertinent to the study's target health condition of diabetes.

This study also had some important limitations. First, the set interview length meant that participants were not able to explore on their own for long periods of time. This likely reduced the opportunity for additional possible insights, especially those concerning discoverability. Second, although our sample included adults of all ages, more research is needed with audiences that vary in terms of other sociodemographic characteristics and levels of health literacy. Third, our use of real-life stimuli means that participants could have had prestudy exposure to promotional content for the selected drugs, either on social media or through some other format, such as television advertisements. Finally, it remains unclear how the findings may be applied to other drugs, medical conditions, and social media platforms beyond the small number included in this study.

Future Research

The exploratory nature of the study provides valuable insights into avenues for further research into the UX of those interacting with prescription drug information on social media. Further exploration with participants who have (or care for those with) other health conditions is likely to provide even more insight into consumer engagement with prescription drug promotion on social media. Since this research tested stimuli only on Facebook and Instagram, supplementary research investigating other widely used platforms could also prove valuable for compiling a well-rounded view of UX with prescription drug promotion in social media.

Additionally, the findings of this study could provide a foundation for developing hypotheses testable by other study types. For instance, experimental stimuli could be developed and tailored to manipulate the way in which the same risk information is presented to enable the quantitative assessment of the level of influence on risk recall and comprehension. Quantitative data such as these could facilitate the development of best practices around UI or UX design features, which could in turn help consumers identify and review important information about drugs on social media more easily. This type of study would circumvent some complexities associated with conducting research in the context of the ever-changing and individualized nature of social media.

Conclusions

This study contributes to existing research by examining how consumers interact with, interpret, and react to prescription drug promotion on social media. Our findings identified UI or UX design features (including where and how drug risk information is presented) that facilitate or pose barriers to users' identification, review, and comprehension of the risk information provided on or via prescription drugs' Facebook and Instagram pages and posts. Based on this study's findings, we have identified future research possibilities to further explore consumer experience regarding prescription drug promotion on social media, as the information shared on these platforms could have a large impact on consumers' treatment decisions.

Acknowledgments

This paper reflects the views of the authors and should not be construed to represent the Food and Drug Administration's (FDA) views or policies. This work was supported by the Office of Prescription Drug Promotion, Office of Medical Policy, Center for Drug Evaluation and Research, US FDA. The authors thank Tierra Butler, Jason Cober, and Ankur Kalola of the US FDA for their help with stimuli development. The authors also thank the following employees of RTI International for their assistance: Stephanie Lane (thoughtful input on our quantitative data collection and analysis plan) and Bridget Kelly (thorough review of the manuscript).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Bulik B. The top pharma companies in social media. Fierce Pharma. 2020. URL: https://www.fiercepharma.com/special-report/top-pharma-companies-social-media [accessed 2024-07-17]
2. statista. Healthcare and pharma industry advertising spending on selected social media in the United States between July and November 2020. 2024. URL: https://www.statista.com/statistics/1251970/healthcare-pharma-industry-social-media-ad-spend-us/ [accessed 2024-07-17]
3. Aikin KJ, Sullivan HW, Berktold J, Stein KL, Hoverman VJ. Consumers' experience with and attitudes toward direct-to-consumer prescription drug promotion: a nationally representative survey. Health Mark Q 2021;38(1):1-11 [FREE Full text] [doi: 10.1080/07359683.2021.1947067] [Medline: 34238134]
4. Hannon C. Health information: what sources do people trust? PatientsLikeMe blog. 2021. URL: https://www.patientslikeme.com/blog/perceptions-of-health-information-sources [accessed 2024-07-17]

XSL•FO
RenderX

5. Mukherjee SK, Kumar J, Jha AK, Rani JR. Role of social media promotion of prescription drugs on patient belief-system and behaviour. Int J e-Collab 2019;15(2):21 [FREE Full text] [doi: 10.4018/ijec.2019040102]
6. Eisenberg MD, Singh Y, Sood N. Association of direct-to-consumer advertising of prescription drugs with consumer health-related intentions and beliefs among individuals at risk of cardiovascular disease. JAMA Health Forum 2022 Aug 05;3(8):e222570 [FREE Full text] [doi: 10.1001/jamahealthforum.2022.2570] [Medline: 36200632]
7. Fogel J, Zhuk A. Direct-to-consumer prescription medication advertisements and obtaining prescriptions with or without a doctor's prescription. Health Mark Q 2019;36(3):220-235. [doi: 10.1080/07359683.2019.1618009] [Medline: 31267822]
8. The Office of Prescription Drug Promotion (OPDP). US Food and Drug Administration. 2024. URL: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/office-prescription-drug-promotion-opdp [accessed 2024-07-17]
9. Drug advertising: a glossary of terms. US Food and Drug Administration. 2020. URL: https://www.fda.gov/drugs/prescription-drug-advertising/drug-advertising-glossary-terms#fair_balance [accessed 2024-07-17]
10. Auxier B, Anderson M. Social media use in 2021. Pew Research Center Report. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-07-17]
11. Van Someren M, Barnard YF, Sandberg J. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. Cambridge, MA: Academic Press; 1994.
12. Miles M, Huberman MA. Qualitative Data Analysis. 2nd Edition. Thousand Oaks, CA: Sage Publications; 1994.
13. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
14. Maxwell JA. Using numbers in qualitative research. Qual Inq 2010 Apr 15;16(6):475-482. [doi: 10.1177/1077800410364740]
15. Salmon P. Assessing the quality of qualitative research. Patient Educ Couns 2013 Jan;90(1):1-3. [doi: 10.1016/j.pec.2012.11.018] [Medline: 23237574]
16. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
17. Chou WYS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
18. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010;15(Suppl 3):216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
19. Gottfried J. Americans' social media use: YouTube and Facebook are by far the most used online platforms among U.S. adults; TikTok's user base has grown since 2021. Pew Research Center. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ [accessed 2025-02-25]

Abbreviations

UI: user interface
UX: user experience

Edited by A Mavragani; submitted 09.10.24; peer-reviewed by H Son, J Walsh; comments to author 12.12.24; revised version received 24.01.25; accepted 13.02.25; published 25.03.25.

Please cite as:
Amoozegar JB, Williams P, Giombi KC, Richardson C, Shenkar E, Watkins RL, O'Donoghue AC, Sullivan HW
Consumer Engagement With Risk Information on Prescription Drug Social Media Pages: Findings From In-Depth Interviews
J Med Internet Res 2025;27:e67361
URL: https://www.jmir.org/2025/1/e67361
doi:10.2196/67361
PMID:

©Jacqueline B Amoozegar, Peyton Williams, Kristen C Giombi, Courtney Richardson, Ella Shenkar, Rebecca L Watkins, Amie C O'Donoghue, Helen W Sullivan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions, Experiences, and Beliefs About Patient Portals Among Women With Limited English Proficiency: Multicultural Qualitative Interview Study

Hae-Ra Han[1,2], RN, PhD; Joyline Chepkorir[1], RN, MSc, PhD; Tina Kim[1], MSN, RN; Martha Zamora[2], BA; Emma Huang[1], MSN, RN

[1]Johns Hopkins University School of Nursing, Baltimore, MD, United States

[2]Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Hae-Ra Han, RN, PhD
Johns Hopkins University School of Nursing
525 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 410 614 2669
Fax: 1 410 502 5481
Email: hhan3@jhu.edu

Abstract

Background: Individuals in the United States with limited English proficiency (LEP) experience a disproportionate disease burden. Patient portals provide patient education, outreach, and linkage to preventive health services. While patient portals have been found to be effective in facilitating the use of preventive services, they have predominantly engaged well-educated, English-speaking, White populations. There is limited research investigating experiences and beliefs about patient portals among populations with LEP.

Objective: This study aims to explore perceptions, experiences, and beliefs about patient portals among women with LEP.

Methods: We used a qualitative semistructured interview design and recruited a purposive sample of women through diverse methods. The interview guide covered topics including experiences with patient portals, the perceived feasibility and relevance of patient portals, and perceptions of patient portals targeted toward women with LEP for health promotion. The interviews were audio recorded for verbatim transcription and analysis. Each bilingual interviewer reached data saturation after interviewing 12 (43%), 9 (32%), and 7 (25%) Korean-, Spanish-, and Swahili-speaking women, respectively, yielding a total of 28 women in the study.

Results: We identified 4 main themes that were common across all linguistic groups: perceived benefits of patient portals, perceived facilitators of patient portal use, perceived barriers to patient portal use, and preferred features and suggested improvements. *Perceived benefits of patient portals* had 5 subthemes: easier communication with health care providers and health systems, getting connected and staying connected with health systems, easier and efficient access to one's health records over time, staying informed of and engaged with one's health and health management, and better patient engagement in medical visits. Subthemes for *perceived facilitators of patient portal use* were availability of time, widespread use and availability of smartphones and the internet in the United States, family support, and parenthood. Subthemes for *perceived barriers to patient portal use* were limited digital literacy and limited access to technology, LEP, lack of awareness and knowledge about patient portals, and illiteracy. Finally, subthemes for *preferred features and suggested improvements* were expanded language access to accommodate non-English speakers, improved accessibility to health information using graphics and patient education materials, and user onboarding education and technical support. Of note, while most subthemes were shared across all 3 groups, the widespread use and availability of smartphones and the internet in the United States and illiteracy subthemes were unique to Swahili-speaking women.

Conclusions: Women with LEP recognized multiple benefits of patient portals; however, several barriers were also identified. These included limited digital literacy, restricted access to technology, LEP, and illiteracy. Barriers to patient portal use were closely tied to social determinants of health, which are commonly experienced by women with LEP. To expedite the attainment of health equity, it is important to promote access to health resources such as patient portals.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e60699) doi:10.2196/60699

KEYWORDS

patient portal; limited English proficiency; immigrant; women; mobile phone

Introduction

Background

Racial and ethnic minority individuals in the United States, particularly those with limited English proficiency (LEP) [1], experience significant health disparities; for example, despite high incidence rates of cervical cancer, early detection through regular screening remains low among these populations, especially among women with LEP [2-4]. As a result, racial and ethnic minority women often present with more advanced disease stages than non-Hispanic White women [5]. This is largely due to language and cultural barriers, as well as low health literacy, which hinder their ability to navigate the US health system effectively [1].

One of the innovations in health care is the patient portal. Patient portals offer numerous benefits, including access to diagnosis, treatment information, and other health care resources, which empower patients by enhancing their knowledge and autonomy; for example, a recent systematic review of 10 studies on asynchronous patient–health care provider communication through patient portals found that electronic messaging with health care providers was associated with clinical benefits for patients with breast cancer, including improved screening behaviors [6]. Similarly, another systematic review of 24 patient portal interventions revealed that patient portals are generally effective in improving psychological outcomes (eg, health knowledge, self-efficacy, and decision-making), medication adherence, and the use of preventive services [7]. However, the review also noted that most patient portal interventions have predominantly engaged highly educated, English-speaking, White populations, highlighting a gap in research regarding other demographic groups.

While the use of patient portals has increased over the years, it remains most common among individuals with high literacy, high education levels, and those who are White; for instance, an analysis of the 2011-2015 National Health Interview Survey showed a 36% increase in the use of eHealth services (eg, web-based medical appointments, prescription refills, and email communication with health care providers) among immigrants compared to an increase of 53% among individuals born in the United States [8]. Notably, immigrants with LEP were 81% less likely to use these eHealth services (adjusted odds ratio 0.19, 95% CI 0.09-0.42) than those who spoke English proficiently [8].

Objectives

Despite these disparities, little is known about how individuals with LEP perceive health IT, particularly patient portals, in terms of relevance, feasibility, and impact. To address this gap, the goal of this study was to better understand immigrant women's experiences and perceptions regarding patient portals. Immigrant women face more challenges than immigrant men in accessing respectful and culturally competent health care [9]. By focusing on some of the common languages spoken in households in the United States [10]—Korean, Spanish, and Swahili—this study aimed to (1) explore the perspectives and beliefs about patient portals among women with LEP and (2) describe potential barriers and facilitators to their use in meeting these women's needs.

Methods

Design and Sample

We used a qualitative semistructured interview design to describe the experiences and perspectives of women with LEP in using patient portals for health promotion. This approach allows researchers to assess what respondents describe as aspects of their culture that influence their use of technology for health, in their own words and context, without having constrained categories [11]. This is an extremely valuable approach in cross-cultural work with ethnic minority groups, in that the interviewer is free to pursue topics in greater depth and gain better understanding of cultural values and norms concerning a given phenomenon [12].

Participants were eligible if they (1) self-identified as women with LEP (ie, difficulty using English); (2) spoke Korean, Spanish, or Swahili; and (3) were aged ≥18 years. Participants were excluded if they had an acute or terminal condition (eg, life expectancy of <6 mo or cancer treatment within the past 5 y) or if they had either a psychiatric diagnosis (eg, schizophrenia or cognitive impairment) or other conditions that precluded participation in study activities. The study team reached data saturation after interviewing 12 (43%), 9 (32%), and 7 (25%) Korean-, Spanish-, and Swahili-speaking women, respectively, yielding a total of 28 women with LEP in the study.

Procedures

Given that the study sample involved diverse linguistic and cultural groups of women with LEP, the study team included trained bilingual research staff for each language group, who recruited and interviewed eligible women in their respective languages. Specifically, the research staff distributed study flyers in the 3 target languages through mobile messaging apps, such as WhatsApp (Meta) for Swahili-speaking women, Kakao Talk (Kakao Corporation) for Korean-speaking women, and ResearchMatch (a web-based national registry to connect volunteers to study teams [13]) for Spanish-speaking women, as well as ethnic churches. We also used word of mouth to recruit participants. Once potential participants were identified, bilingual staff sent them a Qualtrics (Qualtrics International Inc) survey link through email or messaging apps and asked them to complete the survey. The Qualtrics survey verified eligibility for participation, and for those who confirmed eligibility, the link provided pertinent consent information followed by a brief survey addressing sociodemographic characteristics and prior use of patient portals. The brief survey

XSL•FO
RenderX

also allowed women to select tentative days and times for the study interview, which was followed up by bilingual staff.

Interviews were conducted via Zoom (Zoom Video Communications, Inc) or in person, based on participant preference. All interviews were conducted in the participant's native language. We developed a semistructured interview guide that covered topics including past experiences with using patient portals for those with prior patient portal use, perceived feasibility and relevance of patient portals among those without prior patient portal use, and perceptions of patient portals targeted toward immigrant women with LEP for health promotion. Sample questions for individuals with prior use of the patient portal included "Tell me about your experience with the patient portal" and "What/who helped you decide to use the patient portal?" Participants were presented with screenshots of patient portal features throughout the interview to facilitate the discussion. Individuals without prior patient portal use were first presented with screenshots of sample patient portal features and informed about its current functionalities. Sample questions for nonusers included "How useful do you think a patient portal would be to meet your needs?" "What device would you likely use to access the patient portal?" (refer to Textbox 1 for example interview questions). Interview guides were originally developed in English and translated into the 3 target languages by the research staff. Each interview lasted 30 to 45 minutes and was audio recorded to facilitate verbatim transcription and analysis. Each bilingual interviewer reached data saturation after interviewing 12 (43%), 9 (32%), and 7 (25%) Korean-, Spanish-, and Swahili-speaking women, respectively, among the 28 participants.

Textbox 1. Example interview questions.

For women with prior patient portal use

- Tell me about your experience with the patient portal.
  - How was it introduced? Has your provider discussed your health conditions or results via patient portal?
  - Was it available in your language? How did you feel about it?
- Tell me about what you gained from using the patient portal.
  - What are some benefits of having health technology like patient portals?
- What did you like least about the patient portal?
  - In making the patient portal better or more useful for you, what should be added, changed, or removed from it?

For women without prior patient portal use

- What would help you decide to use or to not use a patient portal?
  - What or who would help you feel comfortable using it?
- What do you think about the patient portal?
  - How useful do you think a patient portal would be to meet your needs (eg, looking up medication, finding out diagnosis, obtaining laboratory results, and cancer screening)?
  - How would you want to be informed or educated on how to use a patient portal?

Ethical Considerations

The study was approved by the Johns Hopkins Medicine Institutional Review Board (IRB00243015). Verbal informed consent was obtained from all study participants before data collection, either over the telephone or in person. The consent included information on the risks and benefits of participating in the study. Every potential participant was given a chance to ask questions about their involvement in the study before consenting. To maintain participants' confidentiality, the audio-recorded interviews were exported to a secure cloud storage service (Johns Hopkins OneDrive) immediately after the interviews and deleted from the original recording device. Upon interview completion, participants were sent a US $30 gift code for their participation.

Analysis

We used descriptive statistics to summarize sample characteristics collected from a brief sociodemographic survey. The findings from the qualitative interviews were used to identify descriptors and expressions related to the patient portal experiences of women with LEP. The first author and the coders met regularly to discuss the themes and subthemes and resolve any discrepancies.

The bilingual data collectors transcribed the audio-recorded interviews into Microsoft Word documents and individually coded the data in the original languages to minimize the loss of participants' intended meanings of words, phrases, and concepts [14]. We used a deductive coding based on a priori topics used to develop the interview guide and applied an inductive coding approach, allowing the codes to emerge from the data during the coding process [15]. The coders met regularly to discuss the themes and subthemes that emerged from the data. Data

XSL•FO
RenderX

saturation was determined when no new themes emerged from the analysis. The findings from the qualitative interviews were used to identify common themes and subthemes among prior users and nonusers of patient portals. Finally, all themes, subthemes, and quotes were translated into English and merged in a Microsoft Excel spreadsheet.

Results

Sample Characteristics

Table 1 presents the characteristics of the interview participants. Half (14/28, 50%) of the participants were in their 30s and 40s, with the average ages being 37 (SD 12), 42 (SD 11), and 48 (SD 15) years for Korean-, Spanish-, and Swahili-speaking women, respectively. Korean-speaking participants had the longest average length of stay in the United States (15, SD 11 y), followed by Spanish- (12, SD 9 y) and Swahili-speaking (5, SD 2 y) women. All Spanish-speaking participants (9/9, 100%) and most of the Korean-speaking women (11/12, 92%) had received a high school or college education, whereas more than half of the Swahili-speaking participants (4/7, 57%) had only an elementary-level education. All Korean- and Swahili-speaking women identified themselves as non-Hispanic Asian and African American, respectively, whereas 89% (8/9) of the Spanish-speaking women identified themselves as Hispanic. The interview participants originated from Central America (Dominican Republic, Colombia, and Venezuela), South Korea, and sub-Saharan Africa (Democratic Republic of the Congo and Kenya). Prior use of patient portals varied: 58% (7/12) and 44% (4/9) of the Korean- and Spanish-speaking participants, respectively, had prior experience, whereas all Swahili participants (7/7, 100%) had no prior exposure to patient portals.

Table 1. Sample characteristics (N=28).

Characteristics	Korean-speaking women (n=12)	Spanish-speaking women (n=9)	Swahili-speaking women (n=7)
Age (y), mean (SD; range)	37 (12; 23-61)	42 (11; 23-56)	48 (15; 27-70)
Length of stay in the United States (y), mean (SD; range)	15 (11; 1-36)	12 (9; 1-30)	5 (2; 2-6)
Education, n (%)			
Elementary school	0 (0)	0 (0)	4 (57)
Middle school	0 (0)	0 (0)	1 (14)
High school	1 (8)	5 (56)	1 (14)
College	10 (83)	4 (44)	1 (14)
Graduate school	1 (8)	0 (0)	0 (0)
Race and ethnicity, n (%)			
African	0 (0)	1 (11)	7 (100)
Asian	12 (100)	0 (0)	0 (0)
Hispanic	0 (0)	8 (89)	0 (0)
Country of birth, n (%)			
Democratic Republic of the Congo	0 (0)	0 (0)	4 (57)
Dominican Republic	0 (0)	3 (33)	0 (0)
Colombia	0 (0)	1 (11)	0 (0)
Kenya	0 (0)	0 (0)	3 (43)
South Korea	12 (100)	0 (0)	0 (0)
Venezuela	0 (0)	4 (44)	0 (0)
No response	0 (0)	1 (11)	0 (0)
Prior patient portal use, n (%)	7 (58)	4 (44)	0 (0)

Themes and Subthemes

Overview

Multimedia Appendix 1 summarizes the findings on perceptions and experiences with patient portal use, preferred features, and suggested improvements, organized by themes and subthemes with illustrative quotes. We identified 4 main themes that were common across the linguistic groups: perceived benefits of patient portals, perceived facilitators of patient portal use, perceived barriers to patient portal use, and preferred features and suggested improvements. While most subthemes were shared across all 3 groups, some were unique to specific groups, as detailed in the subsections that follow.

XSL•FO
RenderX

Theme 1: Perceived Benefits of Patient Portals

Participants acknowledged and described multiple benefits of using patient portals. Perceived benefits of patient portals included easier communication with health care providers and health systems, getting connected and staying connected with health systems, easier and efficient access to one's health records over time, staying informed of and engaged with one's health and health management, and better patient engagement in medical visits.

Easier Communication With Health Care Providers and Health Systems

Patient portals were perceived as effective and efficient tools for facilitating communication with the health system. Both users and nonusers noted that these portals increase efficiency and reduce unnecessary in-person visits to health care providers. A Swahili-speaking participant in her late 20s perceived that patient portals might be helpful tools for communicating with health care providers, accessing health information, and ultimately reducing the need for frequent in-person hospital visits:

> *It may help you communicate with the health care provider...it makes everything easier; you don't have to frequently go to the hospital to get information concerning your health from the health care provider.* [SIDSW05]

A Korean-speaking participant in her late 40s noted that patient portals enable women with LEP to take as much time as needed to understand health information, including searching for the meanings of terms using translation tools. She pointed out that some terms can easily be missed or misunderstood during conversations:

> *[I]t's better to use the patient portal so that we [people with LEP] can take our own time to dig into the information. We need more time to understand the information because it's challenging to grasp everything during a conversation. By using the patient portal, we can search for terms and use translation tools. I think it's better to use it rather than feeling pressured to understand everything all at once.* [SIDKO01]

Getting Connected and Staying Connected With Health Systems

The health system felt distant to many participants not only due to various barriers but also because of the difficulty in becoming familiar with a health system different from the one they had experienced in their home countries. Many participants expressed that the patient portal made staying connected with the health system and accessing check-ups a smoother process. A Spanish-speaking participant noted that she visits the patient portal at least once a month to schedule or monitor upcoming appointments:

> *At least once a month I would check if I had any upcoming appointments, if I need to schedule a new appointment, if my annual exam is approaching, or if I need to see my gynecologist.* [SIDSP05]

Similarly, a highly educated Korean woman in her early 50s recalled feeling hesitant when she first encountered the patient portal. However, over time, she gained familiarity and appreciation for it. She stated as follows:

> *As I kept using [the patient portal], it became easier to use and...then I realized it doesn't take that much time to get the check-ups [scheduled]...the scope of my medical examination has expanded.* [SIDKO02]

Easier and Efficient Access to One's Health Records Over Time

One of the main benefits highlighted by the participants was that patient portals serve as a convenient way to keep their health records in one place, allowing easy access to their information. Patient portal users shared their positive experiences, contrasting them with past struggles in tracking health information such as immunization records, laboratory results, or medication history. A Korean-speaking participant in her 40s noted that she finds it convenient to have all her health records in one place, readily available whenever she needs to check her test results:

> *It's good that my medical history or information is stored in one place. I can go back to my records and check when I received certain tests.* [SIDKO01]

Efficiency and ease of information access was echoed by a Spanish-speaking participant:

> *A great benefit of patient portals is that I have access to my information immediately...I think the patient portal is very accessible because it lets me see all of my test results.* [SIDSP05]

Staying Informed of and Engaged With One's Health and Health Management

This subtheme, focusing on managing one's health records in one place, contributed to greater awareness of one's health. Having easy access to their records allowed individuals to become knowledgeable about their health status, motivating them to take a more active role in managing their health. As noted by a Spanish-speaking participant, the patient portal gave patients the autonomy to access and stay informed about their records at their convenience:

> *Accessing my medical records [through the patient portal] gives me a lot of independence. I can use [the portal] whenever I want or whenever I can, knowing that I will always have access.* [SIDSP08]

Parallel to this benefit, a Korean-speaking participant emphasized that the patient portal is a valuable tool for looking up one's test results and proactively managing one's health. She stated as follows:

> *I think it's easier to stay informed and reminded of my health status when I can see my own [test] results. By looking at my results, I gain more awareness and can manage my health better.* [SIDKO03]

In addition, a Swahili-speaking participant highlighted the benefits of patient portal access in improving engagement in care:

XSL•FO
RenderX

> *[I]t would be beneficial in communicating with the health care provider; understanding how to get medication, and how to follow up with the health care provider.* [SIDSW04]

Better Patient Engagement in Medical Visits

Participants viewed the patient portal as a valuable tool for accessing health information before and after medical visits. They described experiencing a higher level of stress when communicating in English about vital health matters compared to their experiences in their home countries. The patient portal allowed them to anticipate discussions, prepare questions, and review results at their convenience. A Korean-speaking participant in her 30s mentioned that reviewing her results before her appointment helped facilitate a productive discussion with her health care provider:

> *I can have a better understanding before the appointment. If I go to see the doctor without knowing anything, there might be parts I don't understand. However, if I already know the results in advance, it would be easier for me to comprehend and discuss with the doctor during the appointment.* [SIDKO10]

Along the same lines, Spanish- and Swahili-speaking participants viewed patient portals as a valuable digital tool for following up with health care providers. A Spanish-speaking participant noted that the patient portal is useful for seeking clarification after appointments:

> *There are times during your appointment that you forget to ask [questions], so having the patient portal and being able to go back and check it whenever is a great advantage.* [SIDSP06]

Theme 2: Perceived Facilitators of Patient Portal Use

Participants identified various factors that facilitated their continued use of patient portals, rather than initial adoption. These included availability of time, widespread use and availability of smartphones and the internet in the United States, family support, and parenthood.

Availability of Time

Spanish- and Swahili-speaking participants affirmed that availability of time played a role in facilitating the use of patient portals. Some Swahili-speaking participants also noted that availability of time may be a good motivator for learning how to use patient portals as a source of health information and for seeking care. No relevant quotes were identified for Korean-speaking participants. A Spanish-speaking participant mentioned that availability of time increased the likelihood of using patient portals to view results or schedule appointments:

> *Having the time to make appointments and access your tests results...is a facilitator of patient portal use.* [SIDSP03]

Widespread Use and Availability of Smartphones and the Internet in the United States

Of note, Swahili-speaking participants perceived that the extensive use and availability of smartphones and the internet in the United States enhanced patient portal use. This subtheme was unique to the Swahili-speaking group. A Swahili-speaking participant stated as follows:

> *There is no one in America who does not have a phone, I was even given a computer when I immigrated here, it's just that I don't know how to use it.* [SIDSW07]

Family Support

Participants perceived that support from family members would facilitate their access to and use of patient portals. A Korean-speaking participant disclosed that she would seek help from her son on how to use the patient portal and was confident that she would know how to use it after a few trials:

> *I'll ask my son whenever there's something I don't know, and then once I practice a couple of times, I can use it on my own.* [SIDKO07]

Notably, most of the Korean-speaking participants in their 20s to 40s expressed concerns that their parents would face difficulties accessing the patient portal. Consequently, they expected to provide support to their parents. Similarly, a Swahili-speaking participant who was illiterate shared that either her children or spouse would help her use the patient portal if needed. Consistent with these perceptions, a Spanish-speaking participant in her 50s (a nonuser) reported that she would like a family member to show her how to use a patient portal:

> *It would be more comfortable if someone in my home taught me [how to use the patient portal], like a family member.* [SIDSP03]

Parenthood

Having school-age children acted as a facilitator for patient portal use among Korean- and Spanish-speaking participants with prior portal experience. No relevant quotes were identified for Swahili-speaking participants or for those without prior patient portal use. All 3 Korean-speaking participants who had school-age children emphasized that convenient access to their children's health care records was a key facilitator. While recent immigration could have been a barrier, parenthood seemed to help them quickly adapt to the US culture and health system:

> *With the patient portal, it seems to be an efficient way to connect with their [children's] health care providers, especially for simple questions and check-ups. That's my first impression of it.* [SIDKO1]

Similarly, a Spanish-speaking participant mentioned that, despite difficulties communicating with staff at the physician's office, she was able to schedule her son's medical appointment using the patient portal:

> *Not having the ability to speak English has made everything more difficult. For example, I needed to schedule an appointment for my son to see the doctor. When I called, everyone spoke English, and I didn't understand anything. So, I used my patient portal to open his account, and through there I was able to accomplish it [schedule her son's appointment].* [SIDSP09]

XSL•FO
RenderX

Theme 3: Perceived Barriers to Patient Portal Use

Participants mentioned several barriers to using the patient portal, including limited digital literacy and limited access to technology; LEP; lack of awareness and knowledge about patient portals; and, additionally, for Swahili-speaking participants, illiteracy.

Limited Digital Literacy and Limited Access to Technology

Digital illiteracy, often stemming from limited access to digital devices and the internet, was mentioned as a potential navigational challenge for interested users. Many participants expressed that a lack of knowledge on how to navigate the patient portal would be a key barrier to its use; for example, some Korean-speaking participants shared that their older adult parents or relatives would not use the patient portal:

> *When I think about my aunt, she's not familiar with using smartphones or websites at all. First of all, she forgets her username and password. She dislikes the process of retrieving passwords...Navigating through websites is also difficult for her, whether it's in English or Korean, because there are too many menus. Our portal also has menus, but when you click one, several options pop up. Also, the fonts are small.* [SIDKO12]

LEP Issue

LEP was identified as a profound barrier to patient portal access and use. Women with prior patient portal experience noted that the lack of patient portal access through a language-concordant medical practitioner is a major barrier for many women with LEP. For some, LEP was an even greater challenge than limited digital literacy. A Korean-speaking participant suggested that LEP can undermine self-efficacy in using the patient portal, even among those who are knowledgeable about it and have access to digital devices:

> *If English proficiency is limited, even with well-functioning devices, websites, or apps that are used in English, there is still some fear and uncertainty because of the fixed idea that 'I don't know' or 'I don't understand.' Even if there is knowledge of the existence of such portals, it's still intimidating to try because of the discomfort with English.* [SIDKO05]

Similarly, a Spanish-speaking participant observed that while the patient portal may be accessible in Spanish, the presentation of results in English limits its usability:

> *Honestly, I think the portal is really good because at least it gives you the option to put it in Spanish. That way, it isn't as hard to follow the steps to create your account. However, your tests results are always in English, which always leaves me thinking 'how do I read these?' Even when I try to translate the results, it is not the same [as having it in Spanish]...The doctor could call you and tell you 'Everything is okay,' but that is not enough.* [SIDSP09]

Patient portals are not yet available in other ethnic languages, such as Swahili, which hinders access and use. This concern was highlighted by a Swahili-speaking participant who had no prior experience using a patient portal**:**

> *Patient portal would be more useful if it is written in both English and Swahili, since not everyone understands English.* [SIDSW05]

Lack of Awareness and Knowledge About Patient Portals

Participants noted that, despite the multiple benefits highlighted, patient portals are generally not well known within their communities. They shared that increasing awareness would likely encourage interest and use. A few Korean-speaking participants emphasized that greater awareness could drive adoption:

> *As long as one knows about the patient portal, it's something that would be definitely utilized. But I didn't know...Since I now know, I'm going to use it.* [SIDKO06]

Illiteracy

Illiteracy was identified as a key barrier to patient portal use among Swahili-speaking participants. No other group noted this as a barrier. A Swahili-speaking participant expressed interest in using the portal but cited limited reading ability as a significant obstacle:

> *If patient portal is written in Swahili, I wouldn't know how to read well, I just know a few words, therefore, having information in audio or video formats would be useful since I don't know how to read.* [SIDSW06]

Theme 4: Preferred Features and Suggested Improvements

Participants also shared suggestions to enhance the patient portal user experience, focusing on language accommodation, device access, features, formatting, and specific health information preferences.

Expanded Language Access to Accommodate Non-English Speakers

Most participants emphasized the need for expanded language access to facilitate patient portal use. Across all language groups, they expressed a preference for patient portals available in their preferred languages. A Spanish-speaking participant noted that since LEP is a major barrier to patient portal use, offering access in preferred languages would be ideal:

> *Not knowing the language [English] is the main reason why people can't use the patient portal. Ideally, the portal should be able to be used in different languages depending on the person's preference.* [SIDSP08]

Her sentiments mirrored those of a Korean-speaking participant:

> *I hope these features can be available in Korean. I wish for patient portals to be accessible in Korean so that even those who struggle with new technologies or have difficulty as they age can still have the opportunity to try it in their native language.* [SIDKO07]

XSL•FO
RenderX

In this context, another participant discussed the issue of understanding medical terminology (ie, health literacy) and noted that it would be helpful if information was provided in Korean because medical terms are more difficult to comprehend in English:

> *Since I'm not familiar with medical terms, if something is written only in English, I wouldn't understand. It would be helpful to have more detailed explanations.* [SIDKO09]

In addition, a Swahili-speaking participant affirmed the need for expanded language access:

> *The patient portal will be more useful if it is written in Swahili.* [SIDSW06]

Improved Accessibility to Health Information Using Graphics and Patient Education Materials

Several participants wished for features that would enhance understanding, such as comparisons of laboratory test results, explanations of medications (eg, precautions and contraindications), and detailed interpretations of laboratory results using graphs or charts. A Korean-speaking participant asserted that displaying laboratory values in graphs or charts would be beneficial:

> *It would be much more convenient if there were a chart or graph showing last year's lab values compared to this year's values. For example, if there were red arrows indicating how much they went up or down compared to last year, that would be really helpful.* [SIDKO05]

Likewise, a Spanish-speaking participant stated that the patient portal should be able to provide basic, patient-centered information about health status:

> *[A]t least check my health status. To see if I am allergic to something; if I have any upcoming appointments; if I am taking any medication, and if so what kind, how much of it, and how many times a day...Basic things like that.* [SIDSP05]

In addition, Swahili-speaking participants expressed a desire for the inclusion of pictures and audio-recorded files to improve accessibility to health information on the patient portal:

> *[P]ictures and audios, with texts written in simple English or even Swahili.* [SIDSW03]

User Onboarding Education and Technical Support

Participants' perceived needs for using the patient portal were similar; many suggested that user onboarding education and technical support are necessary improvements for broader adoption. Regarding the format of onboarding education, some participants expressed a preference for audiovisual tools, while others were comfortable receiving links to instructional content. A Korean-speaking participant mentioned that an internet-based demonstration would be sufficient for her understanding:

> *I think [getting instructions] via email and a link to view the video about it would be helpful.* [SIDKO09]

Considering limitations in technology access and use as well as English proficiency, participants noted that printed instructions in their native language or in-person guidance would be helpful for older adults:

> *I would like to be instructed by someone, for instance, after church service.* [SIDSW06]

Finally, recognizing the constraints of physicians' busy schedules, a Korean-speaking participant stated that a brief demonstration during an appointment would be highly beneficial:

> *It's best if the doctor explains how to use the patient portal, but they're busy. If they open my portal during my visit and let me know 'You can see it on this portal,' I'd be more motivated to use it.* [SIDKO05]

Discussion

Principal Findings

Despite the rapid adoption of patient portals in health care settings [7,8], research on the experiences and perspectives of individuals with LEP is scarce. Our sample of women with LEP from 3 cultural and linguistic groups provided valuable insights into their perceptions, experiences, and beliefs regarding patient portals. The findings highlight the benefits, facilitators, and barriers to using patient portals, as well as preferences for specific features and suggestions for improvements.

Participants with LEP strongly associated patient portals with various perceived benefits, including easier communication with health systems and health care providers and better engagement in medical visits. These perceived benefits are similar to those reported in a recent analysis of comments by predominantly non-Hispanic White female patients in their 50s who had used a patient portal in a health system [16]. According to the analysis, most patients highlighted communicating with health care teams and monitoring health status and care activities as beneficial functions of patient portal use. Especially for women with LEP, the ability to review their records and take as much time as needed to understand health information, including searching for the meanings of terms using translation tools, were considered an important benefit of patient portals. Without this option, understanding health information during medical conversations would be challenging. Patient portals are generally considered beneficial to improving patient engagement in health management [17-19]. As patient engagement is increasingly recognized as an essential component of person-centered care [18], our findings suggest that providing women with LEP access to a patient portal could be a viable approach to promote their engagement in health management and care, potentially leading to positive health outcomes [18].

While the subthemes for facilitators of patient portal use (eg, availability of time, widespread use and availability of smartphones and the internet in the United States, family support, and parenthood) varied across groups, family support emerged as a common subtheme across all 3 groups. A cohort study conducted to assess the prevalence of health care portal use before, during, and after the COVID-19 pandemic and involving middle-aged and older adult primary care patients (2019-2022) [20] found that portal log-in activity was higher during the pandemic compared to the 2019 baseline: higher

XSL•FO
RenderX

portal log-in activity was associated with adequate health literacy and multimorbidity, whereas lower portal log-in activity was associated with older age and female sex. Disparities associated with age and sex decreased as the pandemic progressed, but disparities based on health literacy were exacerbated. The subtheme of family support identified in our study highlights how women with LEP navigate limited health literacy related to patient portals by mobilizing resources within their reach, such as children. Evidence points to the unique role of children in immigrant families as language brokers who "translate and mediate between their heritage language/culture and English/US culture" for their parents with LEP [21]. A similar pattern has been reported in the context of digital health where children help their immigrant parents to access web-based health resources [22]. Nevertheless, research in this area is scarce and has primarily relied on qualitative study designs. Future research is warranted to investigate how health literacy develops and is applied to digital health access in immigrant families, as well as the role bilingual children play in obtaining, processing, and understanding web-based health information to support family members, such as parents with LEP.

While patient portals enable women with LEP to communicate effectively with health care providers and health systems, several barriers identified in this study were closely linked to social determinants of health, including limited digital literacy and limited access to technology, LEP, and illiteracy. Similar barriers to patient portal use have been reported; for example, a recent cross-sectional study of 1850 respondents in the United States found that digital health literacy, home internet access, access to technology devices, and English proficiency were among the correlates of patient portal enrollment and intent to use [23]. Similarly, another study using data from the 2017-2018 Health Information National Trends Survey revealed that LEP, along with no health insurance, limited education, and not having a regular doctor, was significantly associated with lower patient portal use [24]. Of note, several reviews reported concerns about privacy and data security as major barriers to patient portal use. However, these concerns did not emerge as subthemes among the women with LEP in this study. One possible explanation is that our study sample included both women with prior patient portal use and those without, whereas previous research often involved patients already engaged with a health system. It would be important to consider how attitudes and perceptions of patient portals in ethical domains (eg, privacy concerns and data security) might hinder or facilitate patient portal adoption and meaningful use among populations with LEP [25].

Key improvements to patient portals suggested by the women with LEP in this study included expanded language access to accommodate non-English speakers, improved accessibility to health information using graphics and patient education materials, and user onboarding education and technical support. An innovative patient portal that transcends critical barriers such as LEP can be an empowering tool, especially for immigrant populations, to obtain health information, manage care, increase knowledge, and take health care ownership. In addition, the suggestions for the use of graphics and educational materials as well as user onboarding education and technical support underscore the need for improved usability, making patient portals more accessible and user-friendly. This also highlights the importance of more active and direct engagement and endorsement by the health system and health care providers [26,27]. Interactions with patient portals have been widely studied; yet, the influence of usability on patients' decision-making remains understudied [28]. Future research should take a comprehensive approach to examining usability and its effect on patient portal use among the growing and increasingly diverse patient population with LEP.

Limitations

This study has some limitations. First, as a qualitative study, the interpretations are inherently subjective, and it is also possible that social desirability bias during the interviews influenced participants' responses. Second, we did not examine participants' type of insurance, which could be associated with immigration status and may present additional barriers to equitable health care access. Insurance coverage seems to play a role in determining whether participants use patient portals [24]. Third, English proficiency was self-reported and may be subject to self-report bias. In addition, our study sample primarily consisted of younger individuals with more than a decade of residence in the United States and high levels of education, except for the Swahili-speaking group, necessitating careful interpretation of the findings. Their experiences were also specific to the US health care context. In addition, part of the recruitment process relied on technology such as WhatsApp, websites, and Zoom, which may have excluded individuals without such access. Future research should strive to include a more diverse sample to capture a broader range of perspectives and experiences. Finally, the study included women without prior patient portal use, primarily Swahili-speaking participants. The interpretability of the findings, particularly among those without prior exposure to patient portals, may be limited.

Conclusions

Women with LEP acknowledged several benefits of patient portals but also identified barriers to their use, including limited digital literacy, restricted access to technology, LEP, and illiteracy. These barriers are closely linked to social determinants of health, which disproportionately affect women with LEP. As digital health continues to evolve, it is critical for health systems to address digital health disparities and foster meaningful use. This study is novel in its inclusion of women with LEP, providing unique insights into their perceptions, beliefs, and experiences with patient portals. One of the Healthy People 2030 goals is to "eliminate health disparities, achieve health equity, and attain health literacy to improve the health and well-being of all" [29]. Our findings suggest that this goal may be supported by promoting accessible health resources, such as patient portals, with several considerations to enhance usability for populations with LEP. Emerging evidence highlights the significance of presentation formats in improving patient satisfaction and usability [30], and our findings could inform future research to promote engagement with patient portals among individuals with LEP.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all participants for their participation in individual interviews.

Data Availability

Interview data in local languages are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Themes, subthemes, and illustrative translated quotes for each group.
[DOCX File , 21 KB - jmir_v27i1e60699_app1.docx]

References

1. Twersky SE, Jefferson R, Garcia-Ortiz L, Williams E, Pina C. The impact of limited English proficiency on healthcare access and outcomes in the U.S.: a scoping review. Healthcare (Basel) 2024 Jan 31;12(3):364 [FREE Full text] [doi: 10.3390/healthcare12030364] [Medline: 38338249]
2. Fuzzell LN, Perkins RB, Christy SM, Lake PW, Vadaparampil ST. Cervical cancer screening in the United States: challenges and potential solutions for underscreened groups. Prev Med 2021 Mar;144:106400. [doi: 10.1016/j.ypmed.2020.106400] [Medline: 33388330]
3. Nguyen C. Cancer screening & Asian American immigrant women. Stanford University. 2012 Jun 6. URL: https://med.stanford.edu/content/dam/sm/schoolhealtheval/documents/AAWomenCancerScreeningClarice_NguyenNEW.pdf [accessed 2025-02-13]
4. Adunlin G, Cyrus JW, Asare M, Sabik LM. Barriers and facilitators to breast and cervical cancer screening among immigrants in the United States. J Immigr Minor Health 2019 Jun 16;21(3):606-658. [doi: 10.1007/s10903-018-0794-6] [Medline: 30117005]
5. Nardi C, Sandhu P, Selix N. Cervical cancer screening among minorities in the United States. J Nurse Pract 2016 Nov;12(10):675-682. [doi: 10.1016/j.nurpra.2016.08.036]
6. Conroy M, Powell M, Suelzer E, Pamulapati S, Min H, Wright T, et al. Electronic medical record-based electronic messaging among patients with breast cancer: a systematic review. Appl Clin Inform 2023 Jan 15;14(1):134-143 [FREE Full text] [doi: 10.1055/a-2004-6669] [Medline: 36581054]
7. Han HR, Gleason KT, Sun CA, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
8. Wang Y, Do DP, Wilson FA. Immigrants' use of eHealth services in the United States, National Health Interview Survey, 2011-2015. Public Health Rep 2018 Nov 17;133(6):677-684 [FREE Full text] [doi: 10.1177/0033354918795888] [Medline: 30223716]
9. Pillai D, Artiga S, Hamel L, Schumacher S, Kirzinger A, Presiado M, et al. Health and health care experiences of immigrants: the 2023 KFF/LA times survey of immigrants. Kaiser Family Foundation. 2023 Sep 17. URL: https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-and-health-care-experiences-of-immigrants-the-2023-kff-la-times-survey-of-immigrants/ [accessed 2025-02-13]
10. Dietrich S, Hernandez E. Language use in the United States: 2019. United States Census Bureau. 2022 Aug. URL: https://www.census.gov/content/dam/Census/library/publications/2022/acs/acs-50.pdf [accessed 2025-02-13]
11. Magaldi D, Berler M. Semi-structured interviews. In: Zeigler-Hill V, Shackelford T, editors. Encyclopedia of Personality and Individual Differences. Cham, Switzerland: Springer; 2018.
12. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
13. NIH Clinical Center and ResearchMatch. National Institutes of Health Clinical Center. URL: https://www.cc.nih.gov/recruit/research-match [accessed 2025-02-13]
14. Smith HJ, Chen J, Liu X. Language and rigour in qualitative research: problems and principles in analyzing data collected in Mandarin. BMC Med Res Methodol 2008 Jul 10;8(1):44 [FREE Full text] [doi: 10.1186/1471-2288-8-44] [Medline: 18616812]
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

XSL•FO
RenderX

16. Son EH, Nahm ES. Adult patients' experiences of using a patient portal with a focus on perceived benefits and difficulties, and perceptions on privacy and security: qualitative descriptive study. JMIR Hum Factors 2023 Jul 25;10:e46044 [FREE Full text] [doi: 10.2196/46044] [Medline: 37490316]
17. Ammenwerth E, Hoerbst A, Lannig S, Mueller G, Siebert U, Schnell-Inderst P. Effects of adult patient portals on patient empowerment and health-related outcomes: a systematic review. Stud Health Technol Inform 2019 Aug 21;264:1106-1110. [doi: 10.3233/SHTI190397] [Medline: 31438096]
18. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
19. Yousef CC, Thomas A, Alenazi AO, Elgadi S, Abu Esba LC, AlAzmi A, et al. Adoption of a personal health record in the digital age: cross-sectional study. J Med Internet Res 2020 Oct 28;22(10):e22913 [FREE Full text] [doi: 10.2196/22913] [Medline: 32998854]
20. Yoon E, Hur S, Opsasnick L, Huang W, Batio S, Curtis LM, et al. Disparities in patient portal use among adults with chronic conditions. JAMA Netw Open 2024 Feb 05;7(2):e240680 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.0680] [Medline: 38421645]
21. Kim SY, Zhang M, Chen S, Song J, Lopez BG, Rodriguez EM, et al. Bilingual language broker profiles and academic competence in Mexican-origin adolescents. Dev Psychol 2020 Aug;56(8):1582-1595. [doi: 10.1037/dev0001010]
22. Gonzalez C, Bollinger B, Yip J, Pina L, Roldan W, Nieto Ruiz C. Intergenerational online health information searching and brokering: framing health literacy as a family asset. Health Commun 2022 Apr;37(4):438-449. [doi: 10.1080/10410236.2020.1847445] [Medline: 33205663]
23. Philpot LM, Ramar P, Roellinger DL, Njeru JW, Ebbert JO. Individual-level digital determinants of health and technology acceptance of patient portals: cross-sectional assessment. JMIR Form Res 2024 Jun 10;8:e56493 [FREE Full text] [doi: 10.2196/56493] [Medline: 38695754]
24. El-Toukhy S, Méndez A, Collins S, Pérez-Stable EJ. Barriers to patient portal access and use: evidence from the health information national trends survey. J Am Board Fam Med 2020 Nov 20;33(6):953-968 [FREE Full text] [doi: 10.3122/jabfm.2020.06.190402] [Medline: 33219074]
25. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021 Sep 08;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
26. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
27. Powell K. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573 [FREE Full text] [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
28. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. The influence of patient portals on users' decision making is insufficiently investigated: a systematic methodological review. Int J Med Inform 2018 Mar;111:100-111 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.028] [Medline: 29425621]
29. Healthy people 2030 framework. Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Office of the Secretary, U.S. Department of Health and Human Services. URL: https://odphp.health.gov/healthypeople/about/healthy-people-2030-framework [accessed 2025-02-13]
30. van der Mee FA, Schaper F, Jansen J, Bons JA, Meex SJ, Cals JW. Enhancing patient understanding of laboratory test results: systematic review of presentation formats and their impact on perception, decision, action, and memory. J Med Internet Res 2024 Aug 12;26:e53993 [FREE Full text] [doi: 10.2196/53993] [Medline: 39133906]

Abbreviations

LEP: limited English proficiency

Edited by K Williams; submitted 18.05.24; peer-reviewed by N Alomair, M Lee, M Armstrong; comments to author 07.11.24; revised version received 02.01.25; accepted 05.01.25; published 26.02.25.

Please cite as:
Han HR, Chepkorir J, Kim T, Zamora M, Huang E
Perceptions, Experiences, and Beliefs About Patient Portals Among Women With Limited English Proficiency: Multicultural Qualitative Interview Study
J Med Internet Res 2025;27:e60699
URL: https://www.jmir.org/2025/1/e60699
doi:10.2196/60699
PMID:

XSL•FO
RenderX

©Hae-Ra Han, Joyline Chepkorir, Tina Kim, Martha Zamora, Emma Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Virtual Client Experience Survey for Mental Health and Addictions: Revalidation of a Survey to Measure Client and Family Experiences of Virtual Care

Allison Crawford[1], MD, PhD; Anne Kirvan[1], MSW; Marcos Sanches[1], MSc; Amanda Gambin[1], PhD; Denise Canso[1], MSc; Eva Serhal[1], MBA, PhD

Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Allison Crawford, MD, PhD
Centre for Addiction and Mental Health
60 White Squirrel Way
Office 223
Toronto, ON, M6J1H4
Canada
Phone: 1 416 535 8501
Email: allison.crawford@utoronto.ca

Abstract

Background: The onset of the COVID-19 pandemic precipitated a rapid shift to virtual care in health care settings, inclusive of mental health care. Understanding clients' perspectives on virtual mental health care quality will be critical to informing future policies and practices.

Objective: This study aimed to outline the process of redesigning and validating the Virtual Client Experience Survey (VCES), which can be used to evaluate client and family experiences of virtual care, specifically virtual mental health and addiction care.

Methods: The VCES was adapted from a previously validated telepsychiatry survey. All items were reviewed and updated, with particular attention to the need to ensure relevance across mental health care sectors and settings. The survey was then revalidated using the 6 domains of health care quality of the Institute of Medicine (IOM) as a guiding framework. These 6 domains include being safe, effective, patient-centered, efficient, timely, and equitable. The VCES was piloted with a convenience sample of clients and family members accessing outpatient care at the Centre for Addiction and Mental Health (CAMH) in Toronto, Ontario, through video or telephone. A confirmatory factor analysis (CFA) was conducted in MPlus and used to test the factorial structures of the VCES, with minor respecification of the model based on modification indices, factor loadings, reliability, and item-total correlation. The respecifications were checked for alignment with the construct definitions and item interpretation. The reliability of the constructs was estimated by the Cronbach α coefficient.

Results: The survey was completed 181 times. The construct reliability was generally high. Timely was the only subscale with an α lower than 0.7; all others were above 0.8. In all cases, the corrected item-total correlation was higher than 0.3. For the CFA, the model was adjusted after multiple imputations with 20 datasets. The mean chi-square value was 437.5, with df=199 (P<.001). The mean root mean square error of approximation (RMSEA) was 0.08 (SD 0.002), the mean confirmatory fit index (CFI) was 0.987 (SD 0.001), the mean Tucker-Lewis Index (TLI) was 0.985 (SD 0.001), and the mean standardized root mean square residual (SRMR) was 0.04 (SD 0.001).

Conclusions: This study describes the validation of the VCES to evaluate client and family experiences of virtual mental health and addictions care. Given the widespread uptake of virtual care, this survey has broad applicability across settings that provide mental health and addiction care. The VCES can be used to guide targeted quality improvement initiatives across health care quality domains. By effectively addressing challenges as they emerge, it is anticipated that we will continue to move toward hybrid modalities of practice that leverage the strengths and benefits of telephone, video, and in-person care to effectively respond to unique client and family needs and circumstances.

(J Med Internet Res 2025;27:e49844) doi:10.2196/49844

XSL•FO
RenderX

KEYWORDS

virtual care; digital health; mental health; client satisfaction; health care quality; Virtual Client Experience Survey; telehealth; telemedicine; eHealth

Introduction

The onset of the COVID-19 pandemic precipitated a rapid shift to virtual care in health care settings, inclusive of mental health care. Virtual care, often referred to as telehealth or telemedicine, refers to the provision of clinical care using information and communication technologies and can be delivered synchronously (ie, in real-time, at a distance) or asynchronously (ie, separated by time and space) [1,2]. During the early phases of the pandemic, synchronous forms of virtual care, such as video- and telephone-based care, replaced in-person care as the leading modalities of practice in many health care settings [3-6].

Research evaluating virtual mental health care has consistently shown high levels of both client satisfaction [3-7] and provider satisfaction [5,8]. However, recent literature also highlights challenges with virtual care, including inequities in access to and engagement with virtual care [9,10], technology challenges [11], lack of connection between clients and health care providers [4], and concerns regarding privacy and safety [12]. Understanding clients' perspectives on virtual mental health care quality will be critical to inform future policies and practices. In addition, it will be important to understand how these experiences vary across sociodemographic groups, particularly in light of long-standing inequities in health care access and outcomes that risk being reinforced by the widespread uptake of virtual care [10].

To date, there is only 1 known validated measure of client experiences of virtual mental health care, specifically telepsychiatry [6], and there are no known validated measures that can be used to evaluate client experiences across varied types of mental health care. Since the transition to virtual care, a majority of patients across multiple surveys indicate a preference for virtual care to continue to exist as an option [3,4], with many preferring a hybrid approach that combines both in-person and virtual modalities [3]. Validated means to understand client and family experiences of care will be critical to inform evidence-based decision-making regarding virtual care and hybrid modalities of practice moving forward.

This paper outlines the process to redesign and validate the Virtual Client Experience Survey (VCES) that can be used to evaluate client and family experiences of virtual mental health and addictions care.

Methods

Overview

At the outset of the COVID-19 pandemic, the VCES was adapted from a previously validated telepsychiatry survey developed and used within the TeleMental Health program at the Centre for Addiction and Mental Health (CAMH), a mental health and addiction hospital located in Toronto, Ontario [6]. This telepsychiatry survey was updated to respond to the pressing need to evaluate virtual care both within our hospital and externally. Several adaptations were made to extend the use of the tool beyond its original intended use with clients accessing psychiatric consultation and assessment through videoconference [6].

Survey Redesign

The approach to developing and validating the original telepsychiatry survey is outlined in the paper by Serhal et al [6]. Our team of clinicians, researchers, and health care leaders reviewed the telepsychiatry survey and adapted it for use in the current context. All items were reviewed and updated, in particular, with the need to ensure relevance that is accorded with the expansion of virtual care provision beyond physicians and psychiatrists. For example, items on the newly developed VCES ask about the client's experience with the virtual appointment instead of the telepsychiatry appointment. Another item asks whether the health care provider explained the risks and benefits of treatments or interventions rather than the risks and benefits of medications. After consultation with clinicians across multiple programs and services at the hospital, the survey was also updated to ensure applicability not only to clients but to family members as well. This was an important shift given the key role that families play in mental health and addiction care and recovery [13,14] and in supporting access to care through videoconference.

The transition to virtual care during the COVID-19 pandemic also necessitated the examination of virtual care from a digital health equity perspective [10]. Our team consulted with the CAMH Health Equity Office as well as the Provincial System Support Program (PSSP), which administers the Ontario Perception of Care Tool for Mental Health and Addictions (OPOC-MHA) [15], regarding the addition of sociodemographic questions to facilitate health equity analyses. Furthermore, 5 items from the OPOC-MHA were also added to facilitate alignment and comparison between the 2 tools.

The VCES sociodemographic questions ask about gender, age, geographic region, whether the client was born in Canada, racial or ethnic group, illness and disability, and other factors. Clients and families were also asked about their comfort with technology, where they accessed the virtual care appointment from (eg, home, health care organization), the type of device they used to access virtual care (eg, computer, smartphone), and which videoconference platform was used. The telephone was included as an option on the survey, as CAMH, like many health care settings, began providing more services by telephone during the pandemic, particularly for individuals experiencing barriers to video-based care. Finally, as recommended by Serhal et al [6], we added additional survey items to assess access to care, and physical and emotional safety during virtual care appointments. We also added an item about compassionate care, given our research group's interest in compassion as a critical factor in the therapeutic relationship and inpatient and family experience of care [16,17].

XSL•FO
RenderX

The redesigned VCES consists of 22 items to assess the overall quality of care. These items are rated on a 4-point Likert scale (1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree) with additional options including "not applicable" and "prefer not to answer." The final question on the survey is an open-text question to elicit any additional feedback.

Conceptual Framework

Following the adaption of the survey, the survey was revalidated using the 6 domains of health care quality of the Institute of Medicine (IOM) as a guiding framework. These 6 domains include being safe, effective, patient-centered, efficient, timely, and equitable [18]. The IOM domains were selected as a guiding framework to provide a comprehensive evaluation of client satisfaction and experience and to guide targeted quality improvement efforts across the different domains. The heath quality domain definitions [18] have been adapted in Table 1 below for the VCES.

Table 1. Health quality domain definitions.

Health quality domains	Original definition	Adapted definition for the VCES[a] domains
Safe	"Avoiding injuries to patients from the care that is intended to help them [18]."	Virtual care that is physically and psychologically safe and minimizes potential risks.
Effective	"Providing services based on scientific knowledge to all who could benefit and refraining from providing services to those not likely to benefit (avoiding underuse and overuse)" [18]	Effective delivery of virtual care, inclusive of the use of virtual care technologies, to facilitate engagement in virtual care.
Patient-centered	"Providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions" [18]	Virtual care that is collaborative, respectful, and responsive to client and family preferences, needs, and values.
Efficient	"Avoiding waste, in particular, waste of equipment, supplies, ideas, and energy" [18]	Efficiency and ease in accessing virtual care.
Timely	"Reducing waits and sometimes harmful delays for both those who receive and those who give care" [18]	Timely access to virtual care for clients and families, including limited wait times for services.
Equitable	"Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status" [18]	Virtual care that does not vary in quality based on personal characteristics and sociodemographic factors.

[a]VCES: Virtual Client Experience Survey.

Question Validation

Survey Dissemination

The VCES was piloted with a convenience sample of clients (16 years of age and older) and family members accessing a wide range of outpatient mental health and addiction services at CAMH through video or telephone. Survey data was collected using Research Electronic Data Capture (REDCap), a web-based application for securely collecting surveys and managing data [19,20]. The electronic survey took approximately 15 minutes to complete. Survey completion was voluntary, and all survey responses were anonymous. Survey data collection took place in 2021.

Information about the VCES and the process for survey administration was shared with clinicians and administrative staff using multiple channels, including the organizational intranet, email, support and sponsorship from clinical program directors, and program-level presentations by the project team. CAMH clinicians and administrative staff were asked to include information about the VCES, including the REDCap survey links, by email to all CAMH clients and families before their virtual care appointment. A template script for email communication to clients was provided, including a reminder that survey completion was voluntary and that all responses were anonymous. Clients and families were asked to complete a survey once per program or service that they participated in. The survey was made available to clients seeing providers across programs at the hospital, which includes multidisciplinary care providers (eg, physicians, nurses, social workers, occupational therapists, psychologists).

Content Validity

The 22 Likert-scale items on the VCES were independently reviewed by a panel of 5 individuals. including clinicians, researchers, and health care leaders, to determine alignment with the following IOM health quality domains: safe, effective, patient-centered, efficient, and timely [18]. Equity, the sixth IOM domain, is assessed through a comparison of survey response options by sociodemographic group. The panel then met through videoconference to discuss discrepancies until a consensus was reached.

Factorial Structure

To test the factorial structure of the VCES, a confirmatory factor analysis (CFA) was adjusted in Mplus (version 8.2; Muthén & Muthén) [21] using the method of weighted least squares with mean and variance adjusted chi-square test (WLSMV), which is recommended for ordinal outcomes [22]. Multiple imputation using the covariance method [23], with 20 imputed datasets, was used to handle missing values, which were present in 45% of the surveys, with 80% of surveys having 3 or fewer items missing. The evaluation of the model relied on the fit indices RMSEA (root means square error of approximation) [24], CFI (comparative fit index) [25], TLI (Tucker-Lewis index) [26],

XSL•FO
RenderX

and SRMR (standardized root mean square residual) [27], as well as the inspection of standardized factor loadings, variance explained, and modification index relative to each item. The initial factorial structure was defined following expert opinion, with minor respecification of the model based on modification indices, factor loadings, reliability, and item-total correlation. The respecifications were checked for alignment with the construct definitions and item interpretation. Cutoffs for fit indices in CFA and reliability are not universally accepted [28]. That said, based on results from Hu and Bentler [29], TLI and CFI higher than 0.95, SRMR lower than 0.08, and RMSEA lower than 0.06 are considered acceptable.

Reliability

The reliability of the constructs was estimated by the Cronbach α coefficient [30], a measure of internal consistency between the items. The corrected item-total correlation (correlation between the item and the total scale score with the item removed) and the α coefficient for the scale with the item removed were also used as sources of information as to how the items fit the factors. A Cronbach α of 0.7 or higher is considered good [28].

Ethical Considerations

Consent to participate was implied. The survey contained an introductory section explaining the purpose of the survey, informing potential respondents that participation was voluntary and anonymous, and informing them that the findings would be presented as grouped or aggregated data for any presentation or publication. There was no compensation offered to clients and families for completing the survey. This project received ethical approval from CAMH Quality Projects Ethics Review (QPER-2021-000).

Results

Descriptive Analysis

In total, the survey was completed 181 times, with 99 surveys completed by clients, 16 by support persons on behalf of clients, and 22 by family members or caregivers. Respondent information is missing on 44 surveys. Sociodemographic characteristics are included in Table 2.

Most respondents used a computer (113/179, 62.4%) or tablet or smartphone (58/179, 32%). Only 5.5% (10/179) reported using a telephone (audio only). The majority of respondents were either comfortable with technology (64/179, 35.8%) or very comfortable (85/179, 47.5%).

XSL•FO
RenderX

Table 2. Participant characteristics.

Characteristics	Participants, n (%)
Role	
Registered CAMH[a] client	99 (72.3)
Supporting a registered CAMH client	16 (11.7)
Support person	22 (16.1)
Sex	
Female	95 (52.5)
Male	76 (42)
Transgender: female to male	3 (1.7)
Transgender: male to female	2 (1.1)
Other	4 (2.2)
Do not know	1 (0.6)
Age group	
13-18 years	7 (3.9)
19-25 years	52 (28.7)
26-34 years	59 (32.6)
35-44 years	17 (9.4)
45-54 years	24 (13.3)
55-64 years	19 (10.5)
65+	3 (1.7)
Born in Canada	
Yes	121 (67.2)
No	59 (32.6)
Prefer not to answer	1 (0.6)
Ethnicity	
Asian: East (eg, Chinese, Japanese, Korean)	8 (4.4)
Asian: South (eg, Indian, Pakistani, Sri Lankan)	18 (9.9)
Asian: South East (eg, Malaysian, Filipino, Vietnamese)	4 (2.2)
Black: African (e.g., Ghanaian, Kenyan, Somali)	9 (5)
Black: Caribbean (eg, Barbadian, Jamaican)	5 (2.8)
Black: North American (eg, Canadian, American)	2 (1.1)
First Nations	2 (1.1)
Indian: Caribbean (eg, Guyanese with origins in India)	1 (0.6)
Latin American (eg, Argentinean, Chilean, Salvadoran)	7 (3.9)
Métis	1 (0.6)
Middle Eastern (eg, Egyptian, Iranian, Lebanese)	9 (5)
White: European (eg, English, Italian, Portuguese, Russian)	29 (16)
White: North American (eg, Canadian, American)	73 (40.3)
Mixed heritage (eg, Black-African and White-North American)	6 (3.3)
Other(s)	3 (1.7)
Prefer not to answer	4 (2.2)
Illness and disability	
Chronic illness	24 (13.3)

XSL•FO
RenderX

Characteristics	Participants, n (%)
Developmental disability	25 (13.8)
Substance use disorder	21 (11.6)
Learning disability	28 (15.5)
Mental illness	116 (65.1)
Physical disability	7 (3.9)
Sensorial disability	3 (1.7)
None	30 (16.6)
Others	6 (3.3)
Prefer not to answer	14 (7.7)

[a]CAMH: Centre for Addiction and Mental Health.

Reliability

The construct reliability, as measured by Cronbach α, was generally high. Timely was the only subscale with an α lower than 0.7; all others were above 0.8 (more details in Table 3). In all cases, the corrected item-total correlation was higher than 0.3. Considering the items in the model, item 9 (effective) and items 2 and 10 (timely) had lower corrected item-total correlation, with all the other items having correlations around 0.7 or above. Within the "timely" factor, the low item-total correlation was likely reflective of the small number of items associated with this domain.

XSL•FO
RenderX

Table 3. Confirmatory factor analysis results and item statistics.

Factors and survey items	Loading[a]	SE[b]	Missing[c], %	R^2 [d]	Correlation[e]	Mean[f] (SD[g])
Efficient (Cronbach α[h]>=0.87)						
1. It was easy to access virtual care at CAMH[i].	0.88	0.03	12	0.78	0.79	3.47 (0.71)
3. It was easy to book my virtual appointment.	1.00	0.02	9	1.00	0.85	3.51 (0.66)
21. The physical location of where I accessed my virtual appointment was convenient for me.	0.88	0.03	8	0.78	0.72	3.54 (0.72)
Timely (Cronbach α=0.66)						
2. The wait time for services was reasonable for me.[j]	0.72	0.06	23	0.52	0.46	3.26 (0.82)
10. I was able to get a virtual appointment sooner than an in-person healthcare appointment.	0.81	0.06	42	0.66	0.46	3.27 (0.76)
Effective (Cronbach α=0.80)						
4. During my virtual appointment, I was able to see the healthcare provider clearly.	0.89	0.03	10	0.79	0.69	3.54 (0.70)
5. During my virtual appointment, I was able to hear the healthcare provider clearly.	0.91	0.02	6	0.84	0.69	3.43 (0.75)
9. I believe virtual care is just as effective as in-person healthcare.	0.55	0.06	15	0.31	0.38	2.99 (0.89)
22. Overall, I am satisfied with my virtual appointment.	0.96	0.02	9	0.92	0.74	3.51 (0.72)
Patient-centered (Cronbach α=0.96)						
6. I am confident that the healthcare provider at CAMH and my other service providers are working as a team.	0.85	0.03	10	0.72	0.77	3.48 (0.73)
7. I feel that there was an adequate amount of time allotted for the virtual appointment.	0.87	0.03	5	0.75	0.79	3.54 (0.68)
8. I felt comfortable during my virtual appointment.	0.92	0.02	13	0.85	0.81	3.50 (0.77)
12. Staff understood and responded to my needs and concerns.[j]	0.95	0.01	16	0.89	0.83	3.52 (0.72)
13. I was treated with respect by program staff.[j]	0.96	0.01	7	0.92	0.89	3.62 (0.71)
14. I received compassionate virtual care.	0.97	0.01	15	0.94	0.91	3.59 (0.73)
16. The healthcare provider spoke with me about my mental health and/or addiction in a way that I could understand.	0.97	0.01	27	0.95	0.89	3.54 (0.73)
17. I was involved as much as I wanted to be in decisions about my treatment services and supports.[j]	0.97	0.01	29	0.94	0.89	3.54 (0.69)
19. I am confident that I will be able to follow the healthcare provider's recommendations.	0.87	0.03	16	0.75	0.75	3.36 (0.76)
Safe (Cronbach α=0.89)						
11. I was assured my personal information was kept confidential.[j]	0.87	0.03	10	0.76	0.80	3.50 (0.74)
15. I felt safe (emotionally and physically) during my virtual appointment.	0.99	0.01	8	0.98	0.82	3.56 (0.74)
18. The healthcare provider explained to me the benefits and risks of any treatments or interventions that were recommended during my virtual appointment.	0.89	0.02	15	0.79	0.79	3.45 (0.76)
20. I understand what to do if I have a mental health and/or addiction emergency following this appointment.	0.82	0.03	11	0.67	0.69	3.44 (0.73)

[a]Confirmatory factor analysis standardized loadings.

[b]Standard error of the confirmatory factor analysis loading.

[c]Proportion of the subject with missing values in the item. These items were imputed using multiple imputation for the confirmatory factor analysis

XSL•FO
RenderX

analysis.

[d]Variance explained by the confirmatory factor analysis model.

[e]Corrected item-total correlation (correlation between the item and the sum score with the item removed).

[f]Item mean (Item values range from 1=strongly disagree to 4=strongly disagree).

[g]Item standard deviation.

[h]Cronbach α (internal consistency).

[i]CAMH is referenced in the survey that was disseminated internally. The external version refers to "this organization."

[j]Items from the Ontario Perception of Care Tool for Mental Health and Addictions (OPOC-MHA) [15].

Confirmatory Factor Analysis

Confirmatory Factor Analysis loadings are shown in Table 3 above. The model was adjusted after multiple imputations with 20 datasets. We show fit statistics using the mean and SD of these 20 datasets. The mean chi-square value was 437.5, with *df*=199 (*P*<.001). The mean RMSEA was 0.08 (SD 0.002), which is considered reasonable. The mean CFI was 0.987 (SD 0.001), the mean TLI was 0.985 (SD 0.001), and the mean SRMR was 0.04 (SD 0.001). These values are considered to be quite good. With the exception of item 9 in the "effective" factor, all the others have standardized loadings higher than 0.7.

The model shown in Table 3 had 2 respecifications: item 19 was moved from "effective" to "patient-centered," and item 18 was moved from "patient-centered" to "safe." These changes were based on the modification index and discussions with the project team; however, they did not meaningfully change the fit measures. While item 9 does not fit as well in the "effective factor," it was maintained as is, as it represented the best conceptual fit.

The sample size for structural equation models is not well defined because of the high number of parameters of interest and broad range of model complexity. Considering our model with 5 factors with an average of 4 indicators per factor and loadings around 0.8, Wolf et al [31] found that a sample size of 120 tends to be sufficient for power above 80%. These results are negatively affected by the presence of missing values; as such, our sample of 181 is expected to have reasonable power for a CFA analysis.

Discussion

Principal Findings

This paper provides a valid and reliable measure to evaluate client and family experiences of virtual mental health and addiction care, focusing on 6 domains of health care quality. This survey has been used internally at CAMH to guide the next steps and recommendations for virtual mental health care and has been shared externally in both English and French to support quality measurement within other organizations. Information about the survey was shared through a webinar, and the survey was subsequently downloaded hundreds of times across Canada and internationally.

Implications for Virtual Mental Health Care

Given the widespread uptake of virtual care, this survey has broad applicability across settings that provide mental health and addiction care. The VCES can be used to guide targeted quality improvement initiatives across the 6 IOM domains of health care quality, including safe, effective, patient-centered, efficient, timely, and equitable care [18]. It also has the potential to be adapted to other health care specialties and contexts.

Considerations for Administering the VCES

From our experience, we found the use of REDCap (Research Electronic Data Capture; Vanderbilt University) to be an effective means to disseminate the survey, collate results, and ensure anonymity. Some training and guidance regarding survey administration processes were required to ensure that the survey was delivered as part of the standard of care. We found it advantageous to delineate a time-limited window for survey dissemination and completion to limit time demands on both clients and clinicians. In terms of interpreting results, we set a cutoff of 3 (agree) and above as a positive result indicating high quality of care. Finally, there was a high degree of interest in receiving feedback on the survey results. We offered aggregated program-level reports once an adequate sample of surveys was completed in order to preserve respondent anonymity; this is particularly important as demographic questions attached to the survey could render individual results identifiable.

Next Steps

While the findings from the VCES provided invaluable and timely feedback, it is important not to limit patient and family feedback to a survey. The open text option on the survey provided a breadth of feedback not necessarily addressed within the existing survey questions. Analysis of this qualitative data may spur future adaptations to the survey and, on a practical level, may steer quality improvement initiatives resulting from survey findings. Furthermore, ongoing patient and family engagement in interpreting survey results and prioritizing and guiding quality improvement will be essential to making meaningful and impactful improvements.

Patient experience of virtual care should also be considered alongside provider experience of virtual care. To complement findings from the VCES and to capture another key perspective within health care, our team has developed and validated the Virtual Provider Experience Survey (VPES). These complementary tools provide key metrics to inform how health care organizations and systems can move toward the "quadruple aim," a well-recognized framework for health care that seeks to improve population health outcomes, reduce health care costs, and improve patient, family, and provider experiences of care [32]. The consistent use of standardized tools across settings is a valuable way to inform policy and program planning with respect to virtual care [33] and to further our advancement toward the quadruple aim. Further practice-based research is also needed to examine the relationship between the experience of care and health outcomes for clients and families who access

XSL•FO
RenderX

virtual mental health and addiction care and the cost-effectiveness of those services.

Limitations

While the VCES was validated with 181 survey responses, it is possible that some clients or family members completed the survey more than once if they were involved in multiple programs or services during the evaluation period. In addition, when administering surveys, there is always the risk of self-selection and response bias. We sought to mitigate this risk by asking clinicians and administrative staff to send the survey link to all clients and families accessing virtual care at CAMH. A response rate can be beneficial to ascertain the extent of potential bias; however, we were unable to determine the response rate, as it is unknown how many clients and families received the link in their email correspondence from clinicians or administrative staff. It is quite likely that those who have a higher degree of familiarity or comfort with virtual care were more likely to complete an online survey sent by email.

Conclusion

We sought to address a notable gap in the literature by redesigning and validating a measure of client and family experiences of virtual mental health care. Client and family perspectives are needed to evaluate the current state of mental health service delivery and highlight areas that need improvement. By effectively addressing challenges as they emerge, it is anticipated that we will continue to move toward hybrid modalities of practice that leverage the strengths and benefits of telephone, video, and in-person care to effectively respond to unique client and family needs and circumstances [11].

Acknowledgments

The authors would like to acknowledge members of the virtual mental health team at CAMH for supporting the development and dissemination of the VCES, as well as CAMH clinicians and administrative staff for their support in sharing the VCES with clients and families. We would also like to acknowledge the clients and families who took the time to share their feedback on their experiences of virtual care. Generative AI was not used for any section of the manuscript. No funds, grants, or other financial support were received for this project.

Data Availability

The data that support the findings presented in this report are available upon request from the corresponding author.

Authors' Contributions

AC contributed to study conceptualization, data analysis, writing, and review. AK contributed to study conceptualization, data collection and analysis, writing, and review. MS contributed to data analysis, writing, and review of the manuscript. AG contributed to data collection and analysis, review, and editing of the manuscript. DC contributed to the study conceptualization, data collection and analysis, review, and editing of the manuscript. ES contributed to the study conceptualization, data analysis, review, and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Stamenova V, Agarwal P, Kelley L, Fujioka J, Nguyen M, Phung M, et al. Uptake and patient and provider communication modality preferences of virtual visits in primary care: a retrospective cohort study in Canada. BMJ Open 2020;10(7):e037064 [FREE Full text] [doi: 10.1136/bmjopen-2020-037064] [Medline: 32636284]
2. Mechanic O, Persaud Y, Kimball A. Telehealth systems. In: StatPearls. Florida, United States: StatPearls Publishing; 2023.
3. Benudis A, Re'em Y, Kanellopoulos D, Moreno A, Zonana J. Patient and provider experiences of telemental health during the COVID-19 pandemic in a New York city academic medical center. Psychiatry Res 2022;311:114496 [FREE Full text] [doi: 10.1016/j.psychres.2022.114496] [Medline: 35306378]
4. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
5. Schubert NJ, Backman PJ, Bhatla R, Corace KM. Telepsychiatry and patient-provider concordance. Can J Rural Med 2019;24(3):75-82 [FREE Full text] [doi: 10.4103/CJRM.CJRM_9_18] [Medline: 31249155]
6. Serhal E, Kirvan A, Sanches M, Crawford A. Client satisfaction and experience with telepsychiatry: development and validation of a survey using clinical quality domains. J Med Internet Res 2020;22(9):e19198 [FREE Full text] [doi: 10.2196/19198] [Medline: 32755896]
7. Torales J, Vilallba-Arias J, Bogado JA, O'Higgins M, Almirón-Santacruz J, Ruiz Díaz N, et al. Satisfaction with telepsychiatry during the COVID-19 pandemic: patients' and psychiatrists' report from a university hospital. Int J Soc Psychiatry 2023;69(1):156-160 [FREE Full text] [doi: 10.1177/00207640211070762] [Medline: 34991382]

XSL•FO
RenderX

8. Wright J, Dewan S, Hilty D, Dewan NA. Health care providers' perceptions of quality, acceptance, and satisfaction with telebehavioral health services during the COVID-19 pandemic: survey-based study. JMIR Ment Health 2020;7(12):e23245 [FREE Full text] [doi: 10.2196/23245] [Medline: 33180740]
9. Budhwani S, Fujioka J, Thomas-Jacques T, De Vera K, Challa P, De Silva R, et al. Challenges and strategies for promoting health equity in virtual care: findings and policy directions from a scoping review of reviews. J Am Med Inform Assoc 2022;29(5):990-999 [FREE Full text] [doi: 10.1093/jamia/ocac022] [Medline: 35187571]
10. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
11. Budhwani S, Fujioka JK, Chu C, Baranek H, Pus L, Wasserman L, et al. Delivering mental health care virtually during the COVID-19 pandemic: qualitative evaluation of provider experiences in a scaled context. JMIR Form Res 2021;5(9):e30280 [FREE Full text] [doi: 10.2196/30280] [Medline: 34406967]
12. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
13. Kourgiantakis T, Ashcroft R, Mohamud F, Fearing G, Sanders J. Family-focused practices in addictions: a scoping review. Journal of Social Work Practice in the Addictions 2021;21(1):18-53. [doi: 10.1080/1533256x.2020.1870287]
14. Kourgiantakis T, McNeil SR, Hussain A, Logan J, Ashcroft R, Lee E, et al. Social work's approach to recovery in mental health and addiction policies: a scoping review. Social Work in Mental Health 2022;20(4):377-399. [doi: 10.1080/15332985.2021.2019169]
15. Rush DB, Lentinello E, Khalkhali S, Cvetanova Y, Rotondi N, Furlong A, et al. Development of a client perception of care tool for mental health and addictions: qualitative and quantitative psychometric analysis. Centre for Addiction and Mental Health. 2014. URL: https://kmb.camh.ca/eenet/sites/default/files/wp-content/uploads/2016/04/OPOC-Final-Report_July-2014.pdf [accessed 2024-10-22]
16. Kemp J, Zhang T, Inglis F, Wiljer D, Sockalingam S, Crawford A, et al. Delivery of compassionate mental health care in a digital technology-driven age: scoping review. J Med Internet Res 2020;22(3):e16263 [FREE Full text] [doi: 10.2196/16263] [Medline: 32141833]
17. Wiljer D, Strudwick G, Crawford A. Caring in a digital age: exploring the interface of humans and machines in the provision of compassionate healthcare. In: Without Compassion, There Is No Healthcare: Leading with Care in a Technological Age. Ontario: McGill University Press; 2020:33-57.
18. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the quality chasm: A new health system for the 21st century. National Academies Press. 2001. URL: http://www.ncbi.nlm.nih.gov/books/NBK222274/ [accessed 2020-04-02]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
21. Muthén LK, Muthén BO. Mplus User's Guide. Eighth edition. Statmodel. 2017. URL: https://www..com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2023-03-23]
22. Muthén BO. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Statmodel. 1997. URL: https://www.statmodel.com/download/Article_075.pdf [accessed 2023-03-23]
23. Asparouhov T, Muthen B. MPlus Web Notes: Multiple imputation with Mplus. Statmodel. URL: https://tinyurl.com/36d7m9zy [accessed 2024-11-18]
24. Steiger JH. Structural model evaluation and modification: an interval estimation approach. Multivariate Behav Res 1990;25(2):173-180. [doi: 10.1207/s15327906mbr2502_4] [Medline: 26794479]
25. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]
26. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
27. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal 2007;14(3):464-504. [doi: 10.1080/10705510701301834]
28. Lance CE, Butts MM, Michels LC. The sources of four commonly reported cutoff criteria. Organizational Research Methods 2006;9(2):202-220. [doi: 10.1177/1094428105284919]
29. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
30. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297-334. [doi: 10.1007/BF02310555]

XSL•FO
RenderX

31. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;76(6):913-934 [FREE Full text] [doi: 10.1177/0013164413495237] [Medline: 25705052]
32. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
33. Barsom EZ, van Hees E, Bemelman WA, Schijven MP. Measuring patient satisfaction with video consultation: a systematic review of assessment tools and their measurement properties. Int J Technol Assess Health Care 2020:1-7. [doi: 10.1017/S0266462320000367] [Medline: 32624044]

Abbreviations

CAMH: Centre for Addiction and Mental Health
CFA: confirmatory factor analysis
CFI: confirmatory fit index
IOM: Institute of Medicine
OPOC-MHA: Ontario Perception of Care Tool for Mental Health and Addictions
PSSP: Provincial System Support Program
REDCap: Research Electronic Data Capture
RMSEA: root mean square error of approximation
SE: standard error
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
VCES: Virtual Client Experience Survey
VPES: Virtual Provider Experience Survey
WLSMV: weighted least squares with mean and variance adjusted chi-square test

Edited by T de Azevedo Cardoso; submitted 11.06.23; peer-reviewed by J Sharp, SH Hong, A Makai, J Fitzsimon; comments to author 22.03.24; revised version received 23.05.24; accepted 23.09.24; published 03.01.25.

Please cite as:
Crawford A, Kirvan A, Sanches M, Gambin A, Canso D, Serhal E
The Virtual Client Experience Survey for Mental Health and Addictions: Revalidation of a Survey to Measure Client and Family Experiences of Virtual Care
J Med Internet Res 2025;27:e49844
URL: https://www.jmir.org/2025/1/e49844
doi:10.2196/49844
PMID:39752192

©Allison Crawford, Anne Kirvan, Marcos Sanches, Amanda Gambin, Denise Canso, Eva Serhal. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Validation of the Digital Sensitivity Scale for Adults: Cross-Sectional Observational Study

Hae In Park[1*], MA; Minjeong Jeon[1,2*], MA; Ji Seon Ahn[1,3,4], PhD; Kyungmi Chung[1,3,4], PhD; Jin Young Park[1,2,3,4*], MD, PhD

[1]Department of Psychiatry, Yongin Severance Hospital, Yongin, Republic of Korea

[2]Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Center for Digital Health, Yongin Severance Hospital, Yonsei University, Yongin, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jin Young Park, MD, PhD
Department of Psychiatry
Yongin Severance Hospital
363, Dongbaekjukjeon-daero, Giheung-gu
Yongin, 16995
Republic of Korea
Phone: 82 5189 8148
Email: empathy@yuhs.ac

Abstract

Background: The COVID-19 pandemic has accelerated the digitalization of modern society, extending digital transformation to daily life and psychological evaluation and treatment. However, the development of competencies and literacy in handling digital technology has not kept pace, resulting in a significant disparity among individuals. Existing measurements of digital literacy were developed before widespread information and communications technology device adoption, mainly focusing on one's perceptions of their proficiency and the utility of device operation. In the contemporary landscape, digital transformation is evolving within specialized domains, necessitating a comprehensive evaluation of digital competencies, attitudes, and proficiency in technology application to bridge the digital divide and ensure digital compliance.

Objective: This study was designed to address the shortcomings of existing scales and formulate a digital sensitivity scale tailored to the requirements of today's society.

Methods: Initial items of the Yongin Severance Digital Sensitivity Scale (YI-DSS) were collected through a literature review, and expert opinions were gathered to ensure content validity. An exploratory and confirmatory factor analysis included 986 adult participants evaluating 14 digital literacy items and 6 digital efficacy items. The Cronbach α confirmed internal consistency reliability, and 2-tailed *t* tests, ANOVAs, and post hoc tests analyzed demographic differences in digital literacy and efficacy.

Results: A robust 4-factor digital literacy solution was identified: digital application, digital communication, critical thinking, and digital ethics (Kaiser-Meyer-Olkin=0.891; Bartlett × 2=9829.713; *P*<.001; Cronbach α=0.782-0.947). A 2-factor solution defined digital efficacy: digital confidence and digital anxiety (Kaiser-Meyer-Olkin=0.735; Bartlett × 2=3282.217; *P*<.001; Cronbach α=0.787-0.912). Confirmatory factor analysis was conducted for each model (digital literacy model: χ^2_{71}=676.0, comparative fit index=0.938, Tucker-Lewis index=0.921, standardized root mean square residual=0.73, and root mean square error of approximation=0.093; digital efficacy model: χ^2_8=81.9, comparative fit index=0.977, Tucker-Lewis index=0.958, standardized root mean square residual=0.73, and root mean square error of approximation=0.097), which indicated a good fit. The YI-DSS also showed high correlation with the previously developed Digital Literacy Scale (*r*=0.809; *P*<.001).

Conclusions: The YI-DSS, as a self-assessment tool, has the potential to bridge the generational information gap by promoting acceptance, motivation, and adaptation to digital technology. Furthermore, given the remote nature of digital therapeutics, an individual's familiarity with required technologies and digital communication strongly influences their acceptance of digital treatments and the efficacy thereof. This scale can play a pivotal role in enhancing compliance with digital therapeutics by preemptively assessing individuals' technological literacy and competency.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e55828) doi:10.2196/55828

KEYWORDS

information literacy; health literacy; computer literacy; self-efficacy; attitude; digital divide

Introduction

Background

The COVID-19 pandemic has accelerated digitalization in various sectors, with a notable impact in the field of psychiatry. This shift was essential in maintaining psychological evaluations and treatments during physical distancing, demonstrating the effectiveness of digital approaches in improving mental health care. The incorporation of technology into everyday clinical practice has transitioned from being just a necessity to a vital element of psychiatric services. Despite the successful application of digital therapeutics in treating various conditions such as insomnia, attention-deficit/hyperactivity disorder in children, and substance use disorder [1-4], there are significant barriers that hinder their widespread adoption. In a comprehensive review conducted by van Kessel et al [5], the key factors influencing the assimilation of digital therapeutics into health care systems and their practical application by patients and professionals were examined, identifying digital literacy as the most critical factor according to health care professionals. Specifically, they highlighted a significant concern: if the design of digital therapeutics fails to consider the potential exclusion of individuals with limited digital literacy, it could inadvertently aggravate existing disparities in health care access and outcomes. Therefore, given its pivotal role in the assimilation and practical application of digital therapeutics, digital literacy emerges as a crucial element in the field.

Digital literacy, recognized as a key concept within digital competence, was first introduced by Gilster [6], and its conceptual framework has undergone progressive refinement over time. According to Gilster [6], digital literacy goes beyond the ability to simply operate a computer; it refers to the ability to evaluate and judge digitized information (content evaluation) and combine the collected information with new information to achieve specific personal goals (knowledge), thereby ensuring the proper use of information. This concept has evolved into an essential survival skill essential to the members of modern society, and it is presented as a concept that includes the values of digital citizenship, such as social participation and moral attitude in the digital environment [7-11]. In short, digital literacy is a concept that encompasses the cognitive ability to access and understand digital information and the ability to create new and alternative content, as well as the moral responsibility of communicating in a digital environment and solving problems ethically.

Digital literacy is a comprehensive framework for the complex and integrated subfields of technology, knowledge, ethics, and creative production in a digital network environment. In the *Digital literacy across the curriculum* handbook developed for education workers, Futurelab in the United Kingdom presented eight components of digital literacy: (1) functional skills in information and communications technology (ICT), (2) creativity in developing new content and results, (3) collaboration, (4) effective communication using digital technology, (5) the ability to find and select digital information, (6) critical thinking and evaluation, (7) cultural and social understanding, and (8) e-safety in using digital information without violating laws and ethics [12]. Eshet-Alkalai and Chajut [13] proposed six concepts constituting a framework for digital literacy: (1) photovisual literacy, which refers to the ability to work effectively in interfaces such as graphic communication; (2) reproduction literacy, which refers to the ability to manipulate existing digital information to create meaningful new content; (3) information literacy, which refers to the ability to critically consume information; (4) branching literacy, which refers to the ability to construct knowledge through nonlinear search; (5) socioemotional literacy, which refers to the ability to communicate effectively in online communication spaces; and (6) real-time thinking technique, which refers to the ability to process and evaluate large amounts of information in real time.

According to a report by the Korea Education and Research Information Service (2017), in South Korea, the domains of digital literacy are (1) understanding and using digital technologies, (2) digital consciousness and attitudes, (3) digital thinking skills, and (4) digital practice competence. The study by Kang et al [14] on digital literacy measurement in a smart society suggested that there are 3 domains of diagnostic indexes for digital literacy measurement: technology, application, and mind. The technology domain measured the technological application of ICT, and the application domain measured whether digital technology and knowledge are applied to life. The mind domain in particular measured information ethics awareness and norms required in a smart society, which was different from the finding of previous studies that focused on ICT's instrumental ability and technological proficiency. In his study on digital capabilities in the Fourth Industrial Revolution era, Choi [15] presented the five domains of the framework for digital competence: (1) the understanding of digital society and digital citizenship, (2) communication and collaboration using digital technologies, (3) critical thinking and information literacy, (4) computing thinking and problem-solving, and (5) creative and convergence thinking and content creation.

This proposition was meaningful because it emphasized not only digital citizenship and communication skills and ethics in the digital world, such as the impact of digital technology on society, but also computational thinking, such as programming and modeling. It was also noteworthy that the study used the term "digital competence," emphasizing that "competence" refers to the ability to apply knowledge, expertise, and skills required for effective performance in specific situations. This usage was considered more appropriate than "literacy," which primarily focuses on the encoding and decoding of information [15]. Shin and Lee [16] measured the digital literacy of college students required in a software-oriented society by presenting

XSL•FO
RenderX

four factors: (1) basic competence in ICT, (2) software-oriented social adaptation ability, (3) social networking service (SNS) use and collaboration ability, and (4) basic work use ability, where software-based communication and collaboration capabilities were reflected in addition to competence in ICT use.

Although various previous studies, including the ones mentioned previously, deal with and measure the characteristics and components of digital literacy, they have some limitations. Existing scales simply focus on the operational capability of ICT equipment without incorporating ethical values such as correct information ethics, information use attitude, sound communication, and citizenship [17,18]. Furthermore, many studies were either conducted before the popularization of ICT devices or targeted teachers or students for the purpose of promoting digital literacy education. With the emergence of various digital devices, media, and technologies, there is an increasing need to measure literacy in this new environment.

Meanwhile, studies have suggested 3 stages in the development of digital literacy: digital competence, digital use, and digital transformation [19]. The first stage, digital competence, refers to the confident and critical use of ICT for work, leisure, or communication; it consists of basic knowledge, skills, and attitudes about ICT. Individuals successfully use digital competence for their circumstances and apply it to various situations and contexts to achieve various goals. The second stage, digital use, refers to the application of digital competence to specific situations, including the use of digital tools to solve problems or develop problem-solving methods and to share and learn information within a community of practice. In the final stage, digital transformation, the advances in digital use bring innovation and creativity, effecting significant changes in areas of expertise or knowledge. Although each step may not be performed sequentially, in many cases, skills or knowledge of the digital competence level may be required to achieve innovation at a higher level, such as digital transformation. For this developmental process to occur smoothly, intrinsic factors such as interest and motivation that enable the acceptance and use of digital technologies are necessary [19]. This study intended to present self-efficacy as a personal characteristic that is expected to influence the use of digital technology.

Self-efficacy refers to the belief in one's ability to execute and organize the course of action required to achieve a particular outcome [20]. Individuals' self-efficacy affects their ability to undertake tasks and accept challenges, and it is applicable to various situations [21]. For example, changing the way in which education is delivered—from face-to-face education to non–face-to-face education through computers—may affect students' sense of computer-using self-efficacy [22,23]. Students with previous computer or online training reported a higher level of computer self-efficacy [23,24]. With a high level of computer self-efficacy, the students found it easier to engage in online learning and, thus, participated in online learning for a longer time [25]. In other words, self-efficacy in the digital environment is expected to affect individuals' use of digital technology.

Objectives

This study aimed to develop and present a digital literacy scale suitable for current society while also addressing the limitations of existing scales. In this study, *digital technology literacy* refers to the technical ability to properly handle information and communication equipment (eg, computers, smartphones, and tablets) and the various abilities required to perform tasks in a digital environment. The components of digital technology literacy reconstructed in this study based on the concept of digital literacy presented in previous studies are as follows: (1) the ability to create and reproduce new content based on digital technology, (2) the ability to communicate and collaborate to solve problems using digital technology, (3) the ability to verify the reliability and accuracy of digital information and use it critically, and (4) the ability to comply with ethical guidelines and understand ethical issues that arise from using digital technology. Furthermore, to measure personal factors that may influence the use of digital technology, this study also intended to measure the sense of efficacy in using digital technology. Previous research indicates that self-efficacy levels influence technology use, so this study integrated elements of digital efficacy along with digital literacy. Moreover, to gauge digital literacy across a wide age range, we planned to assess from individuals in their 20s to those aged ≥60 years.

Methods

Preliminary Item Development

For the development of the Yongin Severance Digital Sensitivity Scale (YI-DSS), 6 factors were derived through literature review, with 3 to 4 items being developed for each factor, accounting for a total of 20 items. To check the suitability of the developed items for measuring digital literacy and digital efficacy, the goodness of fit was verified by experts to ensure content validity. The evaluation was conducted by 20 experts, including professors in engineering, psychology, or medicine or experts who had worked in the field for >10 years, and the data collected from 18 experts were used for analysis.

Study Design

This was an observational, cross-sectional study conducted through an online survey. The survey was administered to members of a survey site, and data were collected between May 2023 and July 2023. This study adheres to the Checklist for Reporting Results of Internet E-Surveys guidelines for reporting internet survey results [26].

Participants

An online survey was conducted with 1000 adults aged ≥19 years. Participants eligible for inclusion in the study were adults aged ≥19 years, with quotas set to ensure equal distribution by age and gender and proportional distribution by region. Exclusion criteria included individuals who did not complete the survey or provided responses that appeared insincere based on response patterns. The results were obtained based on the responses of 986 participants, excluding the insincere responses of 14 participants.

XSL•FO
RenderX

Setting

This study recruited a sample from a web panel of individuals registered with Macromill Embrain, a professional survey company specializing in web-based survey and research services. Macromill Embrain maintains a diverse participant pool through continuous recruitment efforts across various demographics (eg, age, gender, and region) and identity verification processes. For this study, participants aged ≥19 years residing in South Korea were selected using quota sampling. Age group and gender were equally allocated, whereas region was proportionally allocated based on the population distribution across the country's 17 major provinces. Participants who completed the consent process were asked to answer screening questions on eligibility and basic demographic information. Those who met the eligibility criteria were invited to participate in the main survey, whereas those who did not qualify or exceeded the required quotas received compensation according to Embrain's policy. Information about this compensation was provided at the bottom of the initial survey screen. Eligible participants were then randomly selected from the assigned quota groups and invited through a link containing a brief description of the study and eligibility criteria.

Participation in the survey was entirely voluntary, and participants could withdraw at any time before submitting their responses without penalty. They were also able to review and change their answers before the final submission, ensuring flexibility. Survey responses were automatically recorded and securely stored in a database. Only completed questionnaires were analyzed.

Sample Size Calculation

According to absolute criteria in factor analysis, a sample size of 100 is considered poor, a sample size of approximately 200 is fair, a sample size of approximately 300 is good, a sample size of approximately 500 is very good, and a sample size of ≥1000 is excellent [27]. In terms of the ratio of sample size to the number of measured variables, various scholars suggest different ratios, but it is generally recommended to have a sample size that is at least 20 times the number of factors to be extracted to obtain stable factors [28]. Considering the aforementioned criteria, the total target number of participants was set at 1000.

Statistical Analysis

The data collected were statistically analyzed using SPSS (version 27.0; IBM Corp) and R (R Foundation for Statistical Computing). First, exploratory factor analysis was performed to extract the factors of the YI-DSS. Items with communality values of <0.4 were considered unsuitable for the factor structure and were removed before conducting the analysis. To examine whether the collected data were suitable for analysis, the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy and Bartlett test of identity matrix were performed. Principal component analysis was performed for exploratory factor analysis, and the varimax rotation was used for factor rotation. Second, confirmatory factor analysis was performed to verify whether the items were properly set in the constructs derived through exploratory factor analysis. Third, the Pearson correlation coefficient was used to identify the concurrent validity of the Digital Literacy Scale developed in the previous study. Fourth, the Cronbach α was obtained to verify the internal consistency reliability of the scale. Finally, the differences in digital literacy and digital efficacy according to demographic characteristics were analyzed using 2-tailed *t* tests and ANOVAs. For cases in which significant differences between groups were found, the Scheffé post hoc analysis was conducted. When heteroscedasticity was identified in the test for homogeneity of variances, the Dunnett T3 test, which accounts for unequal variances, was performed.

Data cleaning, exploratory factor analysis, correlation coefficient analysis, *t* tests, and ANOVAs were conducted using SPSS (version 27.0). Confirmatory factor analysis was performed using the *lavaan* package in R.

Ethical Considerations

This study was approved by the institutional review board of Yongin Severance Hospital (9-2022-0199). Before obtaining consent to participate in the study, the purpose and procedures of the research were explained, including the voluntary nature of participation and the right to withdraw at any time. All personal information provided by participants was anonymized, encrypted, and securely managed. In accordance with the Bioethics and Safety Act, the data will be destroyed 3 years after the conclusion of the study. Monetary incentives were offered to participants upon survey completion to encourage participation.

Measure and Scale

For the development of the YI-DSS, the concepts and components of digital literacy and digital efficacy were defined, as shown in Table 1. For digital literacy, 4 factors—digital application, digital communication, critical thinking, and digital ethics—were defined, and for digital efficacy, 2 factors—digital confidence and digital anxiety—were established.

Preliminary items were created for the development of a scale that reflected the concepts of the components to assess digital literacy and digital efficacy. In total, 3 to 4 items were developed for each construct, and a total of 20 preliminary items were selected. Each item was scored on a 7-point Likert scale, from 1 point for "strongly disagree" to 7 points for "strongly agree"—a higher score indicates higher digital literacy. Table 2 shows the selected items.

XSL•FO
RenderX

Table 1. Components of the scale.

Component	Definition
Digital literacy	It refers to the competence to critically understand digital information using ICT[a] equipment and digital technology and create new content through technology convergence to solve problems cooperatively.
Digital application	The ability to create and reproduce new content based on digital technology
Digital communication	The ability to communicate and collaborate to solve problems using digital technology
Critical thinking	The ability to verify the reliability and accuracy of digital information and use it critically
Digital ethics	The ability to comply with ethical guidelines and understand ethical issues that arise from using digital technology
Digital efficacy	It refers to the sense of having the competence to embrace digital technology and confidently apply it to solve everyday problems.
Digital confidence	The confidence to use digital technology to solve problems
Digital anxiety	Distrust, concern, and anxiety about the use of digital technology

[a]ICT: information and communications technology.

Table 2. Items for the development of the scale.

Construct and item ID	Content
Digital application	
DA1	"I can use digital technologies to create new content."
DA2	"I create my own content to convey information accurately."
DA3	"I am skilled in producing digital content in various formats."
DA4	"Producing content directly using digital technology is not challenging for me."
Digital communication	
DCom1	"When I acquire new information, I share it in online spaces."
DCom2	"I find joy in engaging in online conversations and listening to others."
DCom3	"I actively express my views in online communities through posts, comments, etc., or participate in and signing petitions."
Critical thinking	
CT1	"Before utilizing information I've searched for, I verify its accuracy and reliability."
CT2	"I read information from multiple web sources and cross-check its accuracy before using it."
CT3	"Before using online information, I ensure it is up-to-date."
CT4	"Before utilizing online information, I verify the reliability of the information provider."
Digital ethics	
DE1	"I do not leak personal information about others or upload or download photos or videos of others without their permission."
DE2	"I ethically source information on the web and utilize it directly from its original source."
DE3	"I do not create or share materials online that are harmful to others or illegal."
Digital confidence	
DCon1	"I effectively utilize digital information."
DCon2	"I am confident in utilizing digital technology."
DCon3	"I believe I can solve a problem that arises while using digital technologies myself."
Digital anxiety	
DAnx1	"I feel unsafe using digital technologies as I am concerned about device malfunctions, loss of information, hacking, etc."
DAnx2	"I am hesitant to use digital technologies as I am not confident in using them."
DAnx3	"I am afraid of accidentally losing information while using digital technologies."

XSL•FO
RenderX

The concurrent validity between the YI-DSS developed in this study and the previously developed Digital Literacy Scale was confirmed. The Digital Literacy Scale is a self-report scale developed by Lim et al [29] and consists of a total of 24 questions in 5 factors (using digital information and data, creating and expressing digital content, digital communication and collaboration, information protection and compliance with laws and regulations, and computing thinking). The construct validity of the scale was confirmed through exploratory factor analysis and confirmatory factor analysis, and the internal reliability of the scale was found to be 0.941. The responses are rated on a 5-point Likert-type scale.

Results

Demographics

Table 3 presents data on the gender, age, and educational background of the respondents to the questionnaire used to obtain the study results.

Table 3. Demographics (N=986).

Characteristic	Participants, n (%)
Gender	
Male	499 (50.6)
Female	487 (49.4)
Age (y)	
20-29	159 (16.1)
30-39	164 (16.6)
40-49	200 (20.3)
50-59	210 (21.3)
60-64	103 (10.4)
>65	150 (15.2)
Educational level	
Elementary school graduate	1 (0.1)
Middle school graduate	9 (0.9)
High school graduate	205 (20.8)
Graduate	617 (62.6)
Master's degree	78 (7.9)
Doctorate degree	29 (2.9)

Factor Analysis

Table 4 presents the results of the exploratory factor analysis conducted on the Digital Literacy Scale.

All 20 items met the standard (>0.5) in the measure of communality. The results of the KMO and Bartlett test of sphericity were confirmed. A KMO metric value of ≥0.90 was considered excellent, and a value of ≤0.50 was considered unacceptable. In this analysis, the KMO value of the Digital Literacy Scale was found to be good at 0.891, and the Bartlett sphericity was also statistically significant (χ^2_{91}=9829.7; *P*<.001), indicating that the collected data were suitable for factor analysis.

Regarding each domain, digital application consisted of 4 items, with an eigenvalue of 6.289 and a variance explanatory power of 25.56%. The critical thinking factor consisted of 4 items, with an eigenvalue of 2.552 and a variance explanatory power of 21.02%. The digital ethics factor consisted of 3 items, with an eigenvalue of 1.035 and a variance explanatory power of 15.84%. The digital communication factor consisted of 3 items, with an eigenvalue of 1.013 and a variance explanatory power of 15.363%. Finally, the total variance explanatory power of the 4 factors of digital literacy was confirmed to be 77.782%.

The internal consistency (Cronbach α) of each factor in the Digital Literacy Scale was analyzed. Table 5 shows the results. The Cronbach α was 0.974 for digital application, 0.889 for critical thinking, 0.806 for digital ethics, and 0.782 for digital communication, with the overall Cronbach α for the Digital Literacy Scale being 0.900, indicating a desirable level of reliability for all factors and the scale.

XSL•FO
RenderX

Table 4. Exploratory factor analysis on the Digital Literacy Scale (DLS)[a].

Item ID	Communality	DLS			
		Digital application	Critical thinking	Digital ethics	Digital communication
DA1	0.869	0.872	0.256	0.031	0.204
DA3	0.909	0.904	0.227	0.007	0.203
DA2	0.831	0.838	0.147	–0.027	0.326
DA4	0.847	0.874	0.173	0.004	0.232
CT1	0.622	0.414	0.628	0.159	0.180
CT3	0.800	0.178	0.828	0.242	0.156
CT4	0.848	0.181	0.860	0.180	0.207
CT2	0.784	0.185	0.809	0.252	0.176
AT2	0.814	–0.045	0.098	0.896	–0.005
AT1	0.678	0.143	0.351	0.722	0.117
AT3	0.731	–0.056	0.228	0.820	0.065
DCom1	0.746	0.296	0.248	–0.050	0.771
DCom2	0.810	0.208	0.191	0.064	0.852
DCom3	0.601	0.332	0.133	0.185	0.662
Eigenvalue	—[b]	6.289	2.552	1.035	1.013
Variance explanatory power (%)	—	25.56	21.02	15.84	15.363
Cumulative variance explanatory power (%)	—	25.56	46.58	62.42	77.782

[a]Kaiser-Meyer-Olkin=0.891; Bartlett × 2=9829.713; *P*<.001.

[b]Not applicable.

Table 5. Subfactors of the Digital Literacy Scale and their reliability.

	Item ID	Items, N	Cronbach α
Digital application	DA1, DA2, DA3, and DA4	4	0.974
Critical thinking	CT1, CT2, CT3, and CT4	4	0.889
Digital ethics	DE1, DE2, and DE3	3	0.806
Digital communication	DCom1, DCom2, and DCom3	3	0.782
Digital Literacy Scale: total	—[a]	14	0.900

[a]Not applicable.

A total of 6 preliminary items were selected through literature review and a suitability survey by experts to measure the attitude toward using digital technology in the Digital Efficacy Scale. Each item was scored on a 7-point Likert scale from 1 point for "strongly disagree" to 7 points for "strongly agree." Exploratory factor analysis was performed to extract the factors for the Digital Efficacy Scale. Table 6 shows the results.

In this analysis, the KMO value of the Digital Efficacy Scale was found to be acceptable at 0.734, and the Bartlett sphericity was also statistically significant (χ^2=3389.5; *P*<.001), indicating that the collected data were suitable for factor analysis.

Regarding each domain, the digital confidence factor consisted of 3 items, with an eigenvalue of 2.651 and a variance explanatory power of 44.181%. The digital anxiety factor consisted of 3 items, with an eigenvalue of 2.112 and a variance explanatory power of 35.197%. Finally, the total variance explanatory power of the 2 factors of digital efficacy was confirmed to be 79.378%.

The internal consistency (Cronbach α) of each factor in the Digital Efficacy Scale was analyzed. Table 7 shows the results. The Cronbach α was 0.912 for digital confidence and 0.787 for digital anxiety, with the overall Cronbach α for the Digital Efficacy Scale being 0.720, indicating an acceptable level of reliability for all factors and the scale.

Following the analysis of the exploratory factor analysis results, confirmatory factor analysis of the digital literacy and digital efficacy models was conducted using the *lavaan* package in RStudio (Posit PBC). To evaluate the model fit, several indexes were computed, including the chi-square test, Tucker-Lewis index (TLI), comparative fit index (CFI), standardized root

XSL•FO
RenderX

mean square residual (SRMR), and root mean square error of approximation (RMSEA). TLI and CFI values of >0.90 indicate a good model fit, whereas SRMR values of ≤0.08 and RMSEA values of <0.05 suggest a good fit.

As shown in Table 8, neither the digital literacy nor the digital efficacy models demonstrated good fit based on the chi-square statistic (digital literacy model: χ^2_{71}=676.0 and *P*<.001; digital efficacy model: χ^2_8=81.9 and *P*<.001). However, it is well known that the chi-square statistic is highly sensitive to sample size, so it is important to consider other fit indexes as well. The CFI, TLI, SRMR, and RMSEA values all indicated a good fit, supporting the acceptability of each model's fit.

Table 6. Exploratory factor analysis on the Digital Efficacy Scale (DES)[a].

Item ID	Communality	DES	
		Digital confidence	Digital anxiety
DCon1	0.862	0.926	–0.069
DCon2	0.885	0.937	–0.090
DCon3	0.803	0.894	–0.054
DAnx2	0.653	0.179	0.788
DAnx1	0.767	–0.227	0.846
DAnx3	0.792	–0.181	0.871
Eigenvalue	—[b]	2.651	2.112
Variance explanatory power (%)	—	44.181	35.197
Cumulative variance explanatory power (%)	—	44.181	79.378

[a]Kaiser-Meyer-Olkin=0.734; Bartlett × 2=3389.500; *P*<.001.

[b]Not applicable.

Table 7. Subfactors of the Digital Efficacy Scale and their reliability.

	Item ID	Items, N	Cronbach α
Digital confidence	DCon1, DCon2, and DCon3	3	0.912
Digital anxiety	DAnx1, DAnx2, and DAnx3	3	0.787
Digital Efficacy Scale: total	—[a]	6	0.720

[a]Not applicable.

Table 8. Goodness-of-fit indexes for the confirmatory factor analysis models.

Model	Chi-square (*df*)	*P* value	CFI[a]	TLI[b]	SRMR[c]	RMSEA[d]
Digital literacy	676.0 (71)	<.001	0.938	0.921	0.073	0.093
Digital efficacy	81.9 (8)	<.001	0.977	0.958	0.073	0.097

[a]CFI: comparative fit index.

[b]TLI: Tucker-Lewis index.

[c]SRMR: standardized root mean square residual.

[d]RMSEA: root mean square error of approximation.

Covalidation Analysis

Pearson correlation coefficient analysis was conducted to test the covalidation between the YI-DSS and the previously developed Digital Literacy Scale. The total scores on the YI-DSS showed high correlation with the existing Digital Literacy Scale (*r*=0.809; *P*<.001).

Further Analysis

Analysis of digital device use by age group revealed that smartphone use frequency was notably high among all participants (Table 9). Of the 986 participants, 968 (98.2%) reported using a smartphone either "frequently" or "very frequently." Regarding PC use, 78.2% (771/986) of the participants reported frequent use, with approximately 70% of those aged ≥65 years (103/150, 68.7%) also reporting frequent use. In the case of digital communication tools, >73% of the respondents reported frequent use of both SNSs (748/986, 75.9%) and email (720/986, 73%). The use frequency of emerging technologies such as virtual reality, 3D printers (81/986, 8.2%), and voice assistants was low, with <17% (164/986, 16.6%).

XSL•FO
RenderX

An independent-sample *t* test was conducted to examine the difference in digital sensitivity scores by gender (Table 10). The Levene test was used to assess the equality of variances, and the *t* test results were presented based on whether equal variances were assumed. The analysis revealed significant gender differences in both the total digital literacy score and the total digital efficacy score.

In detail, men (mean 66.735, SD 13.008) scored significantly higher than women (mean 62.834, SD 13.775) on total digital literacy (t_{984}=4.574; *P*<.001). Given the significant gender difference in the total digital literacy score, further analysis was conducted on the subfactors. The results indicated significant gender differences across all subfactors. Men scored significantly higher than women in digital application (mean 15.543, SD 5.847 for men vs mean 12.803, SD 6.158 for women; t_{984}=7.167; *P*<.001). Similarly, in the critical thinking factor, men (mean 20.764, SD 4.242) outperformed women (mean 19.762, SD 4.887; t_{984}=–2.658; *P*<.01). The digital communication scores were also significantly higher for men (mean 13.527, SD 3.692) than for women (mean 12.799, SD 3.859; t_{984}=3.029; *P*=.003). Conversely, in the digital ethics factor, women (mean 17.470, SD 3.536) scored significantly higher than men (mean 16.902, SD 3.536; t_{984}=–2.658; *P*=.008).

Regarding digital efficacy, men (mean 25.363, SD 5.152) also scored higher than women (mean 23.895, SD 4.713; t_{984}=4.664; *P*<.001). Analyzing the subfactors, men showed significantly higher scores in digital confidence (mean 14.719, SD 3.656 for men vs mean 12.669, SD 4.153 for women; *P*<.001). However, in the digital anxiety factor, women (mean 11.226, SD 3.543) scored significantly higher than men (mean 10.643, SD 3.997; t_{984}=–2.423; *P*=.02).

To examine whether there were differences in average digital sensitivity across different age groups, a 1-way ANOVA was conducted. A statistically significant difference was found in the total digital literacy score (F_5=18.076; *P*<.001) based on a significance level of .001. Post hoc analysis revealed significant differences in the average total digital literacy scores between individuals in their 20s and 30s and those in their 40s, 50s, and 60s.

The analysis of differences across subfactors by age group is as follows. In the digital application subfactor, individuals in their 20s scored higher than those in their 40s, 50s, and 60s. Individuals in their 30s also scored significantly higher than those in their 50s and 60s, whereas those in their 40s scored higher than those in their 60s. This indicates that individuals in their 60s had the lowest ability to produce information using digital technologies compared to other age groups. In the critical thinking subfactor, individuals in their 20s and 30s scored higher than those in their 50s and 60s, and individuals in their 40s also showed significant differences compared to those in their 50s and 60s. This suggests that individuals in their 60s had a lower ability to critically use digital information compared to other age groups except for those in their 50s. In the digital communication subfactor, individuals in their 20s scored higher than those in their 60s in terms of using digital technologies for communication. Finally, in the digital ethics subfactor, there were no significant differences between age groups.

XSL•FO
RenderX

Table 9. Frequency of digital technology use by age group (N=986).

Type of digital technology and age group (y)	Use frequency among participants, n (%)				
	Never	Rarely	Occasionally	Frequently	Very frequently
Digital device: PC (desktop or laptop)					
20-29	0 (0)	5 (0.5)	16 (1.6)	41 (4.2)	97 (9.8)
30-39	0 (0)	6 (0.6)	20 (2)	32 (3.2)	106 (10.8)
40-49	0 (0)	10 (1)	22 (2.2)	44 (4.5)	124 (12.6)
50-59	0 (0)	15 (1.5)	28 (2.8)	62 (6.3)	105 (10.6)
60-64	0 (0)	9 (0.9)	17 (1.7)	32 (3.2)	45 (4.6)
>65	0 (0)	14 (1.4)	33 (3.3)	55 (5.6)	48 (4.9)
Digital device: smartphone					
20-29	0 (0)	0 (0)	0 (0)	9 (0.9)	150 (15.2)
30-39	0 (0)	1 (0.1)	0 (0)	23 (2.3)	140 (14.2)
40-49	0 (0)	0 (0)	2 (0.2)	34 (3.4)	164 (16.6)
50-59	0 (0)	1 (0.1)	8 (0.8)	40 (4.1)	161 (16.3)
60-64	0 (0)	0 (0)	3 (0.3)	23 (2.3)	77 (7.8)
>65	0 (0)	0 (0)	3 (0.3)	38 (3.9)	109 (11.1)
Digital communication tool: email					
20-29	1 (0.1)	12 (1.2)	24 (2.4)	60 (6.1)	62 (6.3)
30-39	3 (0.3)	9 (0.9)	24 (2.4)	45 (4.6)	83 (8.4)
40-49	3 (0.3)	19 (1.9)	31 (3.1)	49 (5)	98 (9.9)
50-59	0 (0)	19 (1.9)	28 (2.8)	77 (7.8)	86 (8.7)
60-64	1 (0.1)	14 (1.4)	14 (1.4)	33 (3.3)	41 (4.2)
>65	0 (0)	16 (1.6)	48 (4.9)	50 (5.1)	36 (3.7)
Digital communication tool: social networking service					
20-29	0 (0)	1 (0.1)	9 (0.9)	28 (2.8)	121 (12.3)
30-39	5 (0.5)	7 (0.7)	18 (1.8)	39 (4)	95 (9.6)
40-49	10 (1)	5 (0.5)	39 (4)	55 (5.6)	91 (9.2)
50-59	14 (1.4)	14 (1.4)	29 (2.9)	68 (6.9)	85 (8.6)
60-64	9 (0.9)	7 (0.7)	14 (1.4)	32 (3.2)	41 (4.2)
>65	15 (1.5)	15 (1.5)	27 (2.7)	38 (3.9)	55 (5.6)
Emerging technology: VR[a], 3D printer, or AI[b] analysis					
20-29	90 (9.1)	39 (4)	21 (2.1)	8 (0.8)	1 (0.1)
30-39	107 (10.9)	23 (2.3)	23 (2.3)	8 (0.8)	3 (0.3)
40-49	117 (11.9)	37 (3.8)	27 (2.7)	15 (1.5)	4 (0.4)
50-59	114 (11.6)	41 (4.2)	33 (3.3)	16 (1.6)	6 (0.6)
60-64	58 (5.9)	18 (1.8)	18 (1.8)	4 (0.4)	5 (0.5)
>65	93 (9.4)	21 (2.1)	25 (2.5)	8 (0.8)	3 (0.3)
Emerging technology: voice assistant (eg, Siri and Bixby)					
20-29	37 (3.8)	48 (4.9)	39 (4)	20 (2)	15 (1.5)
30-39	43 (4.4)	51 (5.2)	39 (4)	25 (2.5)	6 (0.6)
40-49	47 (4.8)	54 (5.5)	61 (6.2)	30 (3)	8 (0.8)
50-59	64 (6.5)	58 (5.9)	57 (5.8)	21 (2.1)	10 (1)
60-64	37 (3.8)	23 (2.3)	26 (2.6)	14 (1.4)	3 (0.3)

XSL•FO
RenderX

Type of digital technology and age group (y)	Use frequency among participants, n (%)				
	Never	Rarely	Occasionally	Frequently	Very frequently
>65	76 (7.7)	43 (4.4)	19 (1.9)	9 (0.9)	3 (0.3)

[a]VR: virtual reality.

[b]AI: artificial intelligence.

Table 10. Comparison of subset total scores by gender.

	Men (n=499), mean (SD)	Women (n=487), mean (SD)	*t* test (*df*)	*P* value
DLS[a] total	66.735 (13.008)	62.834 (13.775)	4.574 (984)	<.001
DA[b]	15.543 (5.847)	12.803 (6.158)	7.167 (984)	<.001
CT[c]	20.764 (4.242)	19.762 (4.887)	–2.658 (984)	.001
DE[d]	16.902 (3.173)	17.470 (3.536)	–2.658 (984)	.008
Dcom[e]	13.527 (3.692)	12.799 (3.859)	3.029 (984)	.003
DES[f] total	25.363 (5.152)	23.895 (4.713)	4.664 (984)	<.001
Dconf[g]	14.719 (3.656)	12.669 (4.153)	8.219 (984)	<.001
Danx[h]	10.643 (3.997)	11.226 (3.543)	–2.423 (984)	.02

[a]DLS: Digital Literacy Scale.

[b]DA: digital application.

[c]CT: critical thinking.

[d]DE: digital ethics.

[e]Dcom: digital communication.

[f]DES: Digital Efficacy Scale.

[g]Dconf: digital confidence.

[h]Danx: digital anxiety.

In addition, to examine whether there were differences in average digital efficacy across different age groups, a 1-way ANOVA was conducted (Table 11). The results indicated a statistically significant difference (F=18.592; P<.001) based on a significance level of .001. Post hoc analysis revealed that individuals in their 20s and 30s had higher digital efficacy than those in their 50s and 60s.

When examining the subfactors, it was found that individuals in their 20s and 30s scored higher than those in their 50s and 60s in the digital confidence subfactor. In addition, those in their 40s also showed higher confidence than those in their 50s and 60s. However, in the digital anxiety subfactor, the differences across age groups were not statistically significant.

XSL•FO
RenderX

Table 11. Comparison of subfactor and total scores.

Subfactor	Age groups (years), mean (SD)						*F* test (*df*)	*P* value	Post hoc analysis
	20s (a)	30s (b)	40s (c)	50s (d)	60-64 (e)	>65 (f)			
DLS[a] total	70.415 (11.410)	68.220 (12.582)	66.180 (12.716)	62.976 (13.529)	60.369 (13.296)	58.920 (14.233)	18.076 (5)	<.001	a and b>d, e, and f; c>e and f (Scheffé test)
DA[b]	17.038 (5.478)	16.134 (5.917)	14.450 (6.085)	13.305 (6.010)	11.806 (5.610)	11.573 (5.812)	21.578 (5)	<.001	a>c, d, e, and f; b>d, e, and f (Scheffé test)
CT[c]	21.981 (3.566)	21.250 (3.819)	21.085 (4.426)	19.524 (4.657)	19.117 (5.071)	18.127 (4.969)	17.419 (5)	<.001	a and b>d, e, and f; c>e and f (Dunnett T3 test)
DE[d]	17.126 (2.980)	17.232 (3.181)	17.550 (3.226)	17.271 (3.418)	17.107 (3.670)	16.627 (3.801)	1.352 (5)	.24	—[e]
Dcom[f]	14.270 (3.688)	13.604 (3.843)	13.095 (3.578)	12.876 (3.787)	12.340 (3.701)	12.593 (3.922)	5.167 (5)	<.001	a>e and f (Scheffé test)
DES[g] total	26.478 (4.386)	26.329 (5.405)	24.830 (4.853)	24.219 (4.548)	22.738 (4.037)	22.473 (5.144)	18.592 (5)	<.001	a and b>d, e, and f; c>e and f (Scheffé test)
Dconf[h]	15.560 (3.093)	15.073 (3.502)	14.220 (3.915)	13.057 (3.753)	11.806 (4.118)	11.780 (4.427)	26.623 (5)	<.001	a and b>d, e, and f; c>e and f (Dunnett T3 test)
Danx[i]	10.918 (3.847)	11.256 (4.245)	10.610 (3.939)	11.162 (3.564)	10.932 (3.419)	10.693 (3.546)	0.802 (5)	.55	—

[a]DLS: Digital Literacy Scale.

[b]DA: digital application.

[c]CT: critical thinking.

[d]DE: digital ethics.

[e]Not applicable.

[f]Dcom: digital communication.

[g]DES: Digital Efficacy Scale.

[h]Dconf: digital confidence.

[i]Danx: digital anxiety.

Discussion

Principal Findings

The detailed study findings are as follows. First, to develop a digital sensitivity scale, basic items were compiled through a literature review, and 20 items were selected after content validity was verified by 18 experts. In total, 4 factors of digital literacy and 2 factors of digital efficacy were identified. The factors of digital literacy were digital application, critical thinking, digital ethics, and digital communication, collectively accounting for 77.782% of the variance. Digital efficacy was divided into 2 factors: digital confidence and digital anxiety, showing an explanatory power of 79.378%.

The following are the specifics of the YI-DSS, which is described in this study. This scale was developed to assess digital literacy, which is known to have a significant impact on digital technology use, as well as digital efficacy, which is expected to affect intrinsic motivation and attitudes toward the use of digital technology. First, as with the Digital Literacy Scale, the domain of digital application was designed to assess the ability to create and reproduce new content based on digital technology. Next, the digital communication domain aimed to evaluate the ability to communicate and collaborate in solving problems through digital technology. The critical thinking domain was intended to assess the ability to verify the reliability and accuracy of digital information and use it critically. Finally, the digital ethics domain was designed to measure the ability to comply with ethical guidelines and understand ethical issues that arise from using digital technology. The Digital Efficacy Scale, on the other hand, was designed to assess users' confidence and competence in using digital technology. In addition, to identify factors that impede the use of digital technologies, concerns and anxiety about the use of digital technologies were measured.

The additional analyses revealed the following findings. First, the analysis by age group confirmed a generational gap in digital literacy. In the past, when the penetration rate of digital devices was not high, the key factor contributing to the digital divide was the ease of access to digital devices. However, as the gap in physical accessibility to digital devices has been bridged, the ability to effectively use digital technology—referred to as digital literacy—has emerged as the primary cause of the digital divide [30-32].

XSL•FO
RenderX

An analysis of the types and frequency of digital device use among the respondents in this study showed that 98.2% (968/986) of the respondents frequently used smartphones. Even considering that this study was conducted through an online survey, the fact that >98% of respondents reported using their devices “frequently” or “very frequently” indicates that digital technology is used daily across all age groups, suggesting a high level of accessibility to digital devices. When analyzing the characteristics of device use by age group, >70% of respondents in their 20s, 30s, 40s, 50s, and early 60s reported frequent use of email or SNSs. Notably, the proportion of respondents in their 20s who reported frequent SNS use approached 93% (149/159, 93.7%). In addition, more than half (748/986, 75.9%) of the respondents in their late 60s reported frequent use of email and SNSs, with approximately 62% frequently using SNSs. These findings indicate that the penetration of digital devices is consistently high across all age groups and that online information activities and interactions actively take place across all ages.

Despite the improved accessibility to digital technology, there remains a clear generational gap in digital literacy and confidence in using these technologies. This finding suggests the need for age-appropriate assessments and interventions in digital sensitivity. The analysis revealed that the gap between those in their 20s and 60s was particularly pronounced. The older age groups demonstrated lower digital literacy, especially in the subfactors of digital application, practical use, and communication. Interestingly, while the frequency of online interactions for relationship building and information acquisition was high across all age groups, the generational gap in digital literacy was still evident. Older adults showed lower capabilities to verify the sources and reliability of information, critically evaluate it, and create new content and communicate on the internet. This aligns with previous research indicating that older adults, due to lower digital literacy, are more susceptible to exposure to misinformation and are at higher risk of accepting it uncritically [33-36]. Therefore, it is essential to educate and support older adults to become more active and critical consumers of information in the digital world.

Moreover, the study found significant differences in digital efficacy across age groups. The disparity in digital efficacy between individuals in their 20s and those in their 60s was particularly notable, with older adults showing lower confidence in using digital technologies. Specifically, the difference was significant in the subfactor of digital competence, indicating that older adults have lower confidence and perceived proficiency in using these technologies. However, there was no significant difference in digital anxiety, which measures concerns or fears about potential mistakes when using digital technologies. According to Bandura [37], self-efficacy increases engagement in exploratory behaviors. Individuals with high self-efficacy are more likely to take on unfamiliar tasks, actively learn and apply new information, and exhibit openness. Previous studies examining the relationship between self-efficacy and digital technology use have also found that individuals with higher self-efficacy perceive technologies as easier to use and show a higher intention to use them [30]. In this study, the analysis of the use of emerging technologies by age group revealed that, among respondents in their 20s, who exhibited the highest digital efficacy, approximately 22% (35/159) frequently used emerging technologies such as voice assistants. In contrast, only 8% (12/150) of respondents in their late 60s, which showed the lowest digital efficacy, reported frequent use of such technologies. This suggests that the level of digital efficacy not only influences the use of currently available digital technologies but also impacts the adoption of extended digital technologies, such as digital therapeutics and emerging technologies. Consequently, measuring and enhancing digital efficacy alongside digital literacy is crucial as it may affect openness to technology and proficiency and contribute to the digital divide.

Third, this study found no significant differences between age groups in ethical attitudes related to digital device use, which is one of the subfactors of digital literacy. This subfactor assesses attitudes to consider when using digital technologies, including online morality. The questions on this factor asked about awareness of clearly unethical practices, such as using illegal online routes and plagiarism, which might have led to socially desirable responses, potentially biasing the results positively. Nevertheless, measuring awareness of moral values and attitudes in digital environments remains important. According to previous research, the main factors influencing awareness of digital communication ethics are the recognition of the importance of ethical norms and individual morality [38].

As living and interacting in the digital world, which transcends the boundaries of time and space, becomes increasingly important, measuring digital ethics can serve as a guideline to help individuals align their general moral awareness, shaped by their values and beliefs, with the unique and unfamiliar environment of the digital world, thereby guiding them in focusing on and considering ethical principles.

In the current trend of digital-based medical care, the importance of digital sensitivity becomes prominent. However, despite society’s rapid digitization, there is a growing gap in the ability to accept and use digital technology.

A gap in understanding and attitude toward things appears unavoidable in a society in which the older generation, which has transitioned from an analogue or paper-based system to a digital system, coexists with the generation of “digital natives” born in a digitalized society and raised with the use of digital technology [39,40]. Therefore, it is critical to assess the extent of the digital divide and assess and improve digital literacy, a key variable that can influence the level of such a gap [39]. Studies have emphasized that successful experiences in acquiring and using ICTs during the early stages of digital literacy development are essential for applying digital technology to various contexts of life. In other words, it is necessary not only to understand how to use digital technology but also to feel competent and effective when doing so. Identifying the factors that contribute to resistance to the use of digital technology will increase digital competence and improve individual digital adaptation by increasing acceptance of and motivation for digital technology in light of such factors.

XSL•FO
RenderX

Strengths and Limitations

In this study, we developed and validated the YI-DSS, an easy-to-use self-report tool designed to assess an individual's digital literacy and efficacy. This study refined the concept by incorporating the latest insights related to digital literacy and included digital efficacy, which is recognized as a crucial component of digital competence. This scale is expected to enable a multidimensional assessment of an individual's digital competence, facilitating the identification of key areas for intervention and helping bridge the information gap between individuals and generations.

The limitations of this study are as follows. First, this study aimed to create a scale to assess understanding of and competence in commonly used digital technologies. Given the rapid development of ICT equipment and digital technology in today's society, the concept of digital literacy will continue to evolve. As a result, it will be necessary to constantly revise and supplement measurement tools to keep up with the pace of technological development.

Second, this study recruited participants and conducted the survey on the web, making it difficult to rule out the possibility that the sample was biased toward users who are more familiar with online platforms and digital devices. Future studies should consider revalidating the findings through offline or face-to-face surveys.

Finally, the survey was conducted as a one-time assessment, so test-retest reliability was not confirmed. Follow-up studies should aim to secure more in-depth reliability by conducting repeated assessments with the same group.

Conclusions

In conclusion, this study developed and validated the YI-DSS to assess digital literacy and efficacy across various age groups. Content-valid items were analyzed using exploratory and confirmatory factor analyses, confirming a robust factor structure. The scale demonstrated strong internal consistency reliability and concurrent validity, confirming its robustness as a tool for evaluating digital literacy and efficacy. This study emphasized the importance of addressing both digital literacy and efficacy to bridge the digital divide and enhance digital adaptation across generations. While this scale provides a comprehensive tool for measuring digital literacy and efficacy, future research should focus on refining it to keep pace with technological advancements and validating the findings with more diverse populations through offline assessments.

Acknowledgments

This study was supported by a Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and Information and Communications Technology; the Ministry of Trade, Industry, and Energy; the Ministry of Health and Welfare; and the Ministry of Food and Drug Safety; project 1711194226 and RS-2020-KD000143) and the Virtual Patient·Virtual Hospital–based medical technology development project through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (grant HI23C0514).

Authors' Contributions

HIP was responsible for the conceptualization, formal analysis, methodology, data curation, and drafting of the original manuscript, as well as contributing to the review and editing process. MJ contributed to the conceptualization, methodology, and data curation and also participated in the review and editing of the manuscript. JSA was involved in conceptualization and provided significant input during the review and editing of the manuscript. KC contributed to the conceptualization of the study. JYP was responsible for conceptualization, funding acquisition, and overall supervision of the project.

Conflicts of Interest

None declared.

References

1. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RS, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178. [doi: 10.1016/s2589-7500(20)30017-0]
2. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. NPJ Digit Med 2021 Mar 26;4(1):58 [FREE Full text] [doi: 10.1038/s41746-021-00429-0] [Medline: 33772095]
3. Brezing CA, Brixner DI. The rise of prescription digital therapeutics in behavioral health. Adv Ther 2022 Dec 15;39(12):5301-5306 [FREE Full text] [doi: 10.1007/s12325-022-02320-0] [Medline: 36242730]
4. Fast N, van Kessel R, Humphreys K, Ward NF, Roman-Urrestarazu A. The evolution of telepsychiatry for substance use disorders during COVID-19: a narrative review. Curr Addict Rep 2023 Apr 25;10(2):187-197 [FREE Full text] [doi: 10.1007/s40429-023-00480-9] [Medline: 37266192]
5. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, et al. Mapping factors that affect the uptake of digital therapeutics within health systems: scoping review. J Med Internet Res 2023 Jul 25;25:e48000 [FREE Full text] [doi: 10.2196/48000] [Medline: 37490322]

XSL•FO
RenderX

6. Gilster P. Digital Literacy. Hoboken, NJ: John Wiley & Sons; 1997.
7. Soffer O, Eshet-Alkalai Y. Back to the future: an historical perspective on the pendulum-like changes in literacy. Minds Mach 2008 Oct 7;19(1):47-59. [doi: 10.1007/s11023-008-9119-1]
8. Hobbs R. Digital and Media Literacy: A Plan of Action : A White Paper on the Digital and Media Literacy Recommendations of the Knight Commission on the Information Needs of Communities in a Democracy. Washington, DC: Aspen Institute; 2010. URL: https://www.aspeninstitute.org/wp-content/uploads/files/content/docs/Digital_and_Media_Literacy.pdf
9. Hobbs R. News literacy: what works and what doesn't. In: Proceedings of the Association for Education in Journalism and Mass Communication Conference. 2010 Presented at: AEJMC 2010; August 7, 2010; Denver, CO URL: https://mediaeducationlab.com/sites/mediaeducationlab.com/files/Hobbs,%20NEWS%20LITERACY%20AEJMC%202010.pdf
10. Kim YR. Exploring the social and legal issues of internet medical information: focusing on the credibility of internet medical information use and invasion of cyber privacy. J Commun Sci 2010;10(2):179-220 UCI: G704-001590.2010.10.2.003 [FREE Full text]
11. Kwon SH, Kim SM. Reconceptualizing digital literacy in the social media era : based on 'convergence' and 'participatory culture' of Jenkins. Media Educ 2011;1(1):65-82 [FREE Full text]
12. Hague C, Payton S. Digital literacy across the curriculum. Future Lab. 2011. URL: https://www.futurelab.org.uk/resources/digital-literacy-across-curriculum [accessed 2024-12-27]
13. Eshet-Alkalai Y, Chajut E. Changes over time in digital literacy. Cyberpsychol Behav 2009 Dec;12(6):713-715. [doi: 10.1089/cpb.2008.0264] [Medline: 19514816]
14. Kang JM, Song HJ, Kim HS. Development and application of diagnostic tools for measuring digital literacy in the smart era. J Korean Assoc Reg Inf Soc 2014;17(3):143-173 [FREE Full text] [doi: 10.22896/karis.2014.17.3.006]
15. Choi SY. A study on digital capabilities in the era of the fourth industrial revolution. Korean Soc Comput Educ 2018;21(5):25-35. [doi: 10.32431/kace.2018.21.5.003]
16. Shin S, Lee SH. Development and validation of digital literacy measurement tool: a study on development and validity verification of a measurement tool for digital literacy for university students. J Learner Center Curric Instr 2019 Apr;19(7):749-768. [doi: 10.22251/jlcci.2019.19.7.749]
17. Larsson LC. Digital Literacy Checklist. Health Services, University of Washington. 2000. URL: https://api.semanticscholar.org/CorpusID:150605174 [accessed 2024-12-29]
18. Celot P, Tornero JM. Study on assessment criteria for media literacy levels. A comprehensive view of the concept of media literacy and an understanding of how media literacy level in Europe should be assessed. European Commission. 2009 Oct. URL: https://ec.europa.eu/assets/eac/culture/library/studies/literacy-criteria-report_en.pdf [accessed 2024-12-04]
19. Martin A, Grudziecki J. DigEuLit: concepts and tools for digital literacy development. Innov Teach Learn Inf Comput Sci 2015 Dec 15;5(4):249-267. [doi: 10.11120/ital.2006.05040249]
20. Bandura A, Freeman WH, Lightsey R. Self-efficacy: the exercise of control. J Cogn Psychother 1999 Jan 01;13(2):158-166. [doi: 10.1891/0889-8391.13.2.158]
21. Chang CS, Liu EZ, Sung HY, Lin CH, Chen NS, Cheng SS. Effects of online college student's internet self-efficacy on learning motivation and performance. Innov Educ Teach Int 2013 Mar 15;51(4):366-377. [doi: 10.1080/14703297.2013.771429]
22. Hodges CB. Self-efficacy in the context of online learning environments: a review of the literature and directions for research. Perf Improv Q 2008;20(3-4):7-25. [doi: 10.1002/piq.20001]
23. Jan SK. The relationships between academic self-efficacy, computer self-efficacy, prior experience, and satisfaction with online learning. Am J Distance Educ 2015 Mar 11;29(1):30-40. [doi: 10.1080/08923647.2015.994366]
24. Zimmerman WA, Kulikowich JM. Online learning self-efficacy in students with and without online learning experience. Am J Distance Educ 2016 Aug 03;30(3):180-191. [doi: 10.1080/08923647.2016.1193801]
25. Bates R, Khasawneh S. Self-efficacy and college students' perceptions and use of online learning systems. Comput Hum Behav 2007 Jan;23(1):175-191. [doi: 10.1016/j.chb.2004.04.004]
26. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Comrey AL, Lee HB. A First Course in Factor Analysis. New York, NY: Psychology Press; 1992.
28. Arrindell WA, van der Ende J. An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. Appl Psychol Meas 1985 Jun 01;9(2):165-178. [doi: 10.1177/014662168500900205]
29. Lim HS, Oh H, Lee JA. A study on the development and validation of digital literacy scale for adolescents. Korean J Youth Stud 2022;29(8):219-245. [doi: 10.21509/KJYS.2022.08.29.8.219]
30. Oh JH. A study on factors affecting the intention to use the metaverse by applying the extended technology acceptance model (ETAM): focused on the virtual world metaverse. J Korea Contents Assoc 2021 Oct 28;21(10):204-216 [FREE Full text] [doi: 10.5392/JKCA.2021.21.10.204]
31. Loges WE, Jung J. Exploring the digital divide: internet connectedness and age. Commun Res 2001 Aug 01;28(4):536-562. [doi: 10.1177/009365001028004007]
32. Kim MC, Kim JK. Digital divide: conceptual and practical implications. Korean Sociol Assoc 2002;36(4):123-155 [FREE Full text]

XSL•FO
RenderX

33. Wong FH, Liu T, Leung DK, Zhang AY, Au WS, Kwok WW, et al. Consuming information related to COVID-19 on social media among older adults and its association with anxiety, social trust in information, and COVID-safe behaviors: cross-sectional telephone survey. J Med Internet Res 2021 Feb 11;23(2):e26570 [FREE Full text] [doi: 10.2196/26570] [Medline: 33523825]
34. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
35. Guess A, Nagler J, Tucker J. Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci Adv 2019 Jan 04;5(1):eaau4586 [FREE Full text] [doi: 10.1126/sciadv.aau4586] [Medline: 30662946]
36. Seo H, Blomberg M, Altschwager D, Vu HT. Vulnerable populations and misinformation: a mixed-methods approach to underserved older adults' online information assessment. New Media Soc 2020 May 27;23(7):2012-2033. [doi: 10.1177/1461444820925041]
37. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: W. H. Freeman and Company; 1997.
38. Hagelstein J, Volk SK, Zerfass A, Athaydes AS, Macnamara J, Meng J, et al. Ethical challenges of digital communication: a comparative study of public relations practitioners in 52 countries. Int J Commun 2024;18:1072-1093 [FREE Full text]
39. Min Y. The digital divide among internet users: an analysis of digital access, literacy, and participation. J Commun Res 2011 Feb;48(1):150-187. [doi: 10.22174/jcr.2011.48.1.150]
40. Ahn JI, Seo YK. An analysis of sub-factors of digital media literacy divide: focusing on the effects of generation and economic status. J Digit Converg 2014 Feb 28;12(2):69-78. [doi: 10.14400/jdc.2014.12.2.69]

Abbreviations

CFI: comparative fit index
ICT: information and communications technology
KMO: Kaiser-Meyer-Olkin
RMSEA: root mean square error of approximation
SNS: social networking service
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
YI-DSS: Yongin Severance Digital Sensitivity Scale

Edited by T de Azevedo Cardoso; submitted 28.01.24; peer-reviewed by K Lee, CHA Gallasch, P Kislyakov; comments to author 08.07.24; revised version received 03.09.24; accepted 27.11.24; published 10.01.25.

Please cite as:
Park HI, Jeon M, Ahn JS, Chung K, Park JY
Development and Validation of the Digital Sensitivity Scale for Adults: Cross-Sectional Observational Study
J Med Internet Res 2025;27:e55828
URL: https://www.jmir.org/2025/1/e55828
doi:10.2196/55828
PMID:

©Hae In Park, Minjeong Jeon, Ji Seon Ahn, Kyungmi Chung, Jin Young Park. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Psychometric Analysis of the eHealth Literacy Scale in Portuguese Older Adults (eHEALS-PT24): Instrument Development and Validation

Sara Luz[1], MSc; Paulo Nogueira[1,2,3], PhD; Andreia Costa[1,2], PhD; Adriana Henriques[1,2], PhD

[1]Nursing Research, Innovation and Development Centre of Lisbon (CIDNUR), Nursing School of Lisbon (ESEL), Lisbon, Portugal

[2]Enviromental Health Institute (ISAMB), Faculty of Medicine, University of Lisbon, Lisbon, Portugal

[3]Public Health Research Centre, Comprehensive Health Research Center (CHRC), NOVA National School of Public Health, NOVA University of Lisbon, Lisbon, Portugal

Corresponding Author:
Sara Luz, MSc
Nursing Research, Innovation and Development Centre of Lisbon (CIDNUR)
Nursing School of Lisbon (ESEL)
Av. Dom João II, Lote 4.69.01
Lisbon, 1990-096
Portugal
Phone: 351 217913400
Email: sara.luz@esel.pt

Abstract

Background: In this era of digitalization, eHealth interventions are used to engage patients in health care and help them manage their health. Previous studies showed that this can be particularly interesting for chronic disease self-management and self-care in older adults. Despite older adults becoming increasingly active on the internet, they continue to struggle in using eHealth information due to inadequate eHealth literacy. Thus, assessing and monitoring eHealth literacy is critical to support eHealth interventions.

Objective: This study aimed to describe the translation, adaptation, and validation process of the eHealth Literacy Scale (eHEALS) in Portuguese older adults.

Methods: The cross-cultural adaption followed the steps of forward and blinded backward translations, evaluation of the translations by a committee of judges, pilot-testing, and full psychometric testing. We tested the psychometric properties of the eHEALS by carrying out two studies: general psychometric analysis (study 1) and confirmatory factor analysis (study 2). Study 1 included 80 older adults conveniently selected from a Health Family Unit. Data were collected by in-person questionnaires between May and July 2022. Study 2 included 301 older adults randomly selected from two distinct Health Family Units. Data were collected by in-person questionnaires between May and July 2023.

Results: We tested stability, reliability, construct validity (exploratory and confirmatory factor analyses and known groups), and model fit. Study 1 had 58.8% (47/80) male and 41.3% (33/80) female respondents (mean age 71.20, SD 5.26 years). Study 2 had 56.5% (170/301) male and 43.5% (131/301) female respondents (mean age 71.77, SD 5.15 years). Moderate and strong correlations were identified in the scale items (study 1: $0.42 \leq r \leq 0.91$ and study 2: $0.81 \leq r \leq 0.96$; $P<.001$). The scale showed good internal consistency for study 1 ($\alpha=.92$) and study 2 ($\alpha=.98$), with high correlations between items. The exploratory factor analysis yielded a single-factor structure, explaining 58.3% of the variance in study 1 and 86.4% in study 2. In the confirmatory analysis (study 2), the model fit was mixed ($\chi^2_{20}=265$, $P<.001$; comparative fit index=0.94; Tucker-Lewis Index=0.91; root mean square error of approximation=0.20). Thus, we compared 1-, 2-, and 3-factor structures, deciding on the unidimensional one. In study 1, the eHEALS-PT24 (Portuguese version of the eHealth Literacy Scale for older adults) mean score was 27.25 (SD 5.61), with 43.8% (35/80) and 11.3% (9/80) of participants showing low and high eHealth literacy levels, respectively. In study 2, the eHEALS-PT24 mean score was 23.31 (SD 9.53), with 38.2% (115/301) and 23.6% (71/301) of participants showing low and high eHealth literacy levels, respectively. The known-groups analysis showed statistically significant differences between eHealth literacy and demographic variables ($P<.001$).

Conclusions: The findings suggest that the eHEALS-PT24 is a reliable and valid tool to assess eHealth literacy in Portuguese older adults. Therefore, this instrument can be integrated to support the implementation process of eHealth interventions.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e57730) doi:10.2196/57730

KEYWORDS

eHealth; health literacy; older adults; psychometric properties; public health

Introduction

Background

Currently, the internet is the main source of health information. According to the survey Flash Eurobarometer 404 [1], 8 (80%) in 10 Europeans used the internet for private matters, with the majority (59%) using it for health information searching purposes. The increasing number of internet users has been influencing health systems globally, leading to investments in the provision of eHealth services. In recent years, there have been great efforts to use eHealth interventions to engage patients in health care and help them manage their health, particularly among older adults with chronic disease self-management and self-care [2-4]. Nevertheless, despite the broad access to the internet and other electronic sources, there continues to be a lack of skills to access valid and reliable health information, as well as to understand, evaluate, and apply it adequately for decision-making purposes [5-7]. Considering this condition, eHealth literacy has emerged as a new concept in health care and a challenge in public health [8].

According to Norman and Skinner [9], eHealth literacy is "the ability to seek, find, understand, and evaluate information from electronic sources and apply it to knowledge gained to address or solve a health problem." Based on this, individuals should be able to perform basic or advanced information retrieval, distinguish documents from reliable sources (eg, authoritative ones or scientific evidence-based), and understand eHealth terminology. Furthermore, this set of skills requires the ability to use information and communication technology (ICT), think critically about their nature, and efficiently navigate different electronic resources to obtain information for health-related decision-making.

Considering that eHealth literacy is not a static set of skills, instead changing over time, Norman and Skinner [9] consider the importance of assessing and monitoring eHealth literacy as critical to supporting eHealth interventions. In recent years, indeed, older adults have become increasingly active on the internet and interested in digital health care services to manage their health; however, limited attention has been given to measuring eHealth literacy in this population. Therefore, we sought to address this gap by translating, culturally adapting, and validating a scale for Portuguese older adults that measure eHealth literacy levels (Multimedia Appendix 1).

The eHealth Literacy Scale

The eHealth Literacy Scale (eHEALS), developed by Norman and Skinner [10], was the most frequently used instrument to measure eHealth literacy worldwide, covering 18 languages, 26 countries, and diverse populations and contexts (ie, adolescents, adults, elderly, patients, healthy people, caregivers, health school professionals, community, and clinical practice) [11]. Since its development, the eHEALS has been widely translated, adapted, and validated in several countries, such as Brazil [12], China [13-16], Ethiopia [17], Germany [18,19], Greece [20], Hungary [21], Indonesia [22], Iran [23,24], Italy [25,26], the Netherlands [27]; Norway [28,29], Poland [30,31], Portugal [32], Serbia [33], South Korea [34-36], Spain [37], and Sweden [38], as well as in countries with English as the main language [39-47].

The scale's development study [10] assessed the eHealth literacy of Canadian adolescents (n=664) aged between 13 and 21 years from 14 secondary schools. Considering the psychometrics test results, the instrument showed a good internal consistency (α=.88), with moderate to strong correlations between items ($0.51 \leq r \leq 0.76$) and a moderate test-retest reliability ($0.40 \leq r \leq 0.68$). In addition, a unidimensional structure was found in exploratory and confirmatory factor analyses, explaining 56% of the total variance.

In Portugal, and similarly to the original study, the eHEALS was translated, adapted, and validated in 2014 in a sample of adolescents (n=1215) attending secondary education. The results showed a good internal consistency (α=.84) and a 2-factor structure in the exploratory factor analysis (EFA; extraction of 2 factors with α=.81 and α=.73), explaining 61% of the total variance. From a subgroup analysis, the authors found statistically significant differences regarding the level of education. For the variables sex and age, no statistically significant differences were found. No further translation, adaptation, and validation studies of eHEALS considering other target populations or contexts in Portugal have been carried out since then [32].

Studies with samples including older adults [24,36,39,43,44,48] showed good internal consistency, with Cronbach α ranging between 0.87 and 0.99. Regarding factor analysis, 3 studies [24,36,39] obtained 1D structures, and 2 other studies obtained 3D ones [43,44].

The questionnaire itself consists of 8 items assessing the person's perception of knowledge, comfort, and ability to find, evaluate, and apply health ICT. Each of the items is scored at 5 points on a Likert scale ranging between "1" (totally disagree) and "5" (totally agree), with a total score from 8 to 40 points. The higher the score, the higher the eHealth literacy levels. Items 1 and 2 are related to awareness, items 3 and 4 are related to demand, items 6 and 7 are related to the evaluation of health resources, and items 5 and 8 are related to the use of health information. The instrument also contains 2 additional items, not adding to the final score, which assesses the participant's perception of using the internet to access health information and make health decisions in terms of its usefulness and importance. The theoretical basis of the eHEALS was the social cognitive theory of Albert Bandura and the Lily Model, which explains multiple components of the constructs based on 6 components of literacies—traditional (literacy and numeracy skills), health, information, scientific, media, and computer [10].

XSL•FO
RenderX

This paper presents and discusses the results of full psychometric testing of the eHEALS-PT24 (Portuguese version of the eHealth Literacy Scale for older adults), which aimed to:

- Translate, culturally adapt, and validate the eHEALS for assessing eHealth literacy in Portuguese older adults.
- Test full psychometric characteristics of the eHEALS to be used in Portuguese older adults.
- Explore associations between eHealth literacy and sociodemographic variables.

Methods

Design

We followed a methodology for translating, adapting, and validating instruments proposed by Sousa and Rojjanasrirat [49]—a 7-step guideline based on the review of existing recommendations in the scientific literature on the process of translation, cultural adaptation, and validation of instruments for use in cross-cultural health care (Figure 1).

Figure 1. Steps of the protocol followed for translating, cultural adapting, and validating the eHealth Literacy Scale (eHEALS). [a]: N/A: not applicable.

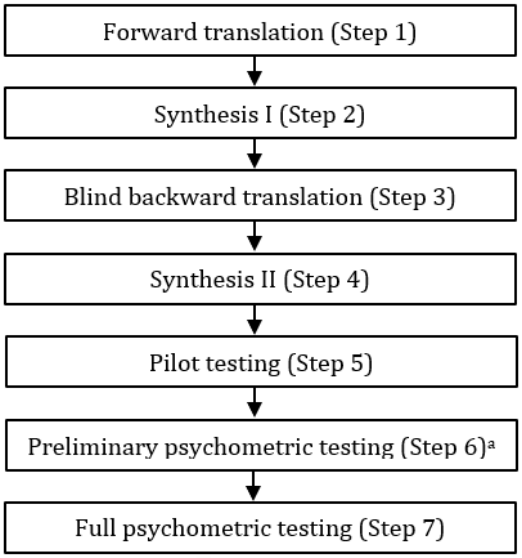

The original version of the questionnaire was independently translated to Portuguese by 2 translators with high proficiency in English: a translator with knowledge of the terminology used in health care and another with knowledge of the cultural and linguistic characteristics of the target language (forward translation; step 1). After this, a third translator compared the two translated versions of the instrument and the original one to detect discrepancies. In addition, a first meeting with a committee of experts with different backgrounds and expertise was held to evaluate discrepancies, composed of the 2 bilingual experts from the first step, the third bilingual translator, academia members, and the research team (step 2). These 2 steps generated the preliminary translated version of the instrument for the Portuguese language.

Independent, blinded backward translation to English was carried out by 2 translators whose first language was English as spoken in the United Kingdom: one translator with knowledge of the terminology used in health care and another with knowledge of the cultural and linguistic characteristics of the target language (step 3). This step generated 2 retroverted versions of the original version of the instrument, which were then compared with the original one (in English) to detect discrepancies by the research team. In addition, a second meeting with a committee of experts with different backgrounds and expertise was held to detect discrepancies, composed of all translators involved in the previous three steps, academia members, and the research team (step 4). These 2 steps generated the prefinal version of the translated instrument, which the original authors validated.

The prefinal version of the instrument was then tested in a pilot study in the target language with a monolingual sample (n=15) to evaluate the instructions, items, and response format clarity (ie, to establish whether the instrument could be satisfactorily understood and completed by people from the target population; step 5). As part of this step, a committee of experts (n=8) was also held to further examine the instrument for clarity of the instructions, items, and response format (content equivalence assessment). The fifth step generated some adjustments in the prefinal version of the instrument.

According to the protocol followed, step 6 (preliminary psychometric testing of the prefinal version of the translated instrument with a bilingual sample) is rarely used, except when a bilingual population is accessible, which justified the option of not carrying it out in this validation process.

Full psychometric testing (step 7) involved reviewing and refining the items of the final version of the eHEALS for Portuguese older adults, as well as establishing stability, internal consistency, validity, and model fit. This step encompassed 2 substudies: general psychometric characteristics of the instrument (n=80, study 1) and confirmatory factor analysis (CFA; n=301, study 2). The sample sizes were defined according to the protocol, namely, at least 10 participants per item of the instrument for general psychometric analysis and 300 to 500 participants for CFA.

XSL•FO
RenderX

Participants and Recruitment

Study 1 included 80 older adults from a Health Family Unit (Primary Care) located in Faro. The sampling procedures were based on a convenience sample. Data collection was conducted by in-person questionnaires between May and July 2022. During that time, patients aged 65 years or older who went to the Health Family Unit (Primary Care) for an appointment were invited to participate in the survey.

Study 2 included 301 older adults from 2 distinct Health Family Units (Primary Care), located in Faro and Lisbon. Participants were randomly selected from the total number of patients of the 2 considered Health Family Units and invited by phone to visit the unit and participate in the survey. Data collection was conducted by in-person questionnaires between May and July 2023.

Inclusion criteria were age (aged 65 years or older) and sex (female and male). Difficulty communicating or using the Portuguese language was an exclusion criterion. The questionnaire removal was performed when the questionnaires were not entirely fulfilled.

Data Analysis

The data analysis was conducted in several stages to ensure a comprehensive understanding of the psychometric properties of the eHEALS-PT24. The following steps outline our analytical approach.

Descriptive Statistics

Initially, we performed descriptive statistical analyses to understand the basic features of the data. This included calculating mean, SD, frequencies, and percentages for all the variables, providing a snapshot of the sample characteristics.

Scale and Items Stability

To assess the scale and item stability, we calculated Pearson correlations. We considered the following Pearson correlation ranges: very weak (0.00-0.19), weak (0.20-0.39), moderate (0.40-0.59), strong (0.60 and 0.79), and very strong (0.80-0.99). Also, statistical significance was set at $P<.001$ [50].

Reliability Analysis

To assess the internal consistency, we calculated Cronbach α coefficients. A Cronbach α value of 0.70 or higher was considered indicative of acceptable reliability. The item-total correlation between the items and the total score was considered high, with values above 0.40 [50].

Construct Validity: Factor Analysis

We conducted an EFA to explore the underlying factor structure of the scale. The Kaiser-Meyer-Olkin (KMO) measure and Bartlett test of sphericity were used to assess the suitability of the data for factor analysis. Factors were extracted using principal component analysis, and a varimax rotation was applied to aid interpretability.

Following EFA, a CFA was performed using structural equation modeling to verify the factor structure obtained from EFA. Model fit was evaluated using different indices, including the comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and chi-square (*df*) ratio. RMSEA values of 0.00 to 0.05 indicate a closer or good fit, 0.05 to 0.08 a fair fit, 0.08 to 0.10 a mediocre fit, and over 0.10 a poor fit. Other fit indexes that we used were the CFI and TLI, both of which should be close to 0.95 [51].

Construct Validity: Known-Groups Validity

The known-groups validity measures an instrument's ability to distinguish among distinct groups (ie, discriminating groups known to differ on the variable of interest) [50]. To evaluate the known-groups validity, we compared eHEALS-PT24 scores across different sociodemographic groups (ie, age, sex, residence area, and education level) using the independent Student *t* test (2-tailed), ANOVA, or Kruskal-Wallis test, as appropriate.

Statistical Software

All analyses were performed using IBM SPSS and Jamovi (The Jamovi Project). By using this multistep analytical process, we aimed to rigorously evaluate the psychometric properties of the eHEALS-PT24, ensuring its reliability and validity in assessing eHealth literacy among Portuguese older adults.

Ethical Considerations

The ethics review assessments were submitted and approved by Portugal's National Data Protection Authority and Regional Health Administrations (Algarve, Lisbon, and Tagus Valley; (approval 4825/2018). Informed consent was applied to participants, according to the model recommended by the Ethics Committee of the Algarve Regional Health Administration, guaranteeing voluntary participation, the possibility of withdrawing at any time, data anonymity, privacy, confidentiality, and no harm to participants. In addition, an authorization request was made to the original authors of the eHEALS [10], which was accepted.

Results

Pilot Study

Before the full psychometric testing, we carried out a pilot study, which we briefly summarized to support the results presented in this subsection. The sampling procedure for piloting was based on snowball sampling, composed of older people aged 65 years or older (n=15). Data collection was conducted using a web-based questionnaire between July and August 2020. The interrater agreement among the sample was above 80% for all items. As part of this step, a committee of experts (n=8) was also held to further examine the instrument for clarity of the instructions, items, and response format; also, to assess the content equivalence, content validity index at the item level (I-CVI=1) and at the scale level (S-CVI=1), as well as the Fleiss κ coefficient of agreement (κ=0.24; $P<.001$), were calculated.

Participant's Characteristics

Study 1

Out of the 80 participants, the sample comprised 47 male (58.8%) and 33 female (41.3%) participants, with a mean age of 71 (SD 5.26) years, ranging from 65 to 88 years. From the total sample (n=80), most of the participants were married or had civil partners (n=63, 78.8%), living with their spouse (n=65,

XSL•FO
RenderX

81.3%) in urban areas (n=70, 87.5%). In addition, most of the respondents were retired (n=69, 86.3%). Regarding formal education, this was analyzed according to the International Standard Classification of Education (ISCED), showing that the most general education level was upper secondary education—ISCED 3 (n=20, 25%), followed by primary education—ISCED 1 (n=17, 21.3%), and bachelor or equivalent degree—ISCED 6 (n=17, 21.3%). Concerning the current health status, most participants reported a diagnosed chronic condition (n=58, 72.5%), mainly high blood pressure (n=29, 36.3%), diabetes (n=25, 31.2%), and dyslipidemia (n=8, 10%). Regarding prescribed medication, 85% (n=68) of the sample answered that they usually take medicines, with 3 being the average number of medicines per person and 9 being the maximum number. Table 1 presents the sociodemographic data.

XSL•FO
RenderX

Table 1. Sociodemographic data.

Characteristics	Study 1 (n=80)	Study 2 (n=301)
Average age (years), mean (SD)	71.20 (5.26)	71.77 (5.15)
Sex, n (%)		
Female	33 (41.3)	131 (43.5)
Male	47 (58.8)	170 (56.5)
Marital status, n (%)		
Single	1 (1.3)	11 (3.7)
Married or civil partner	63 (78.8)	215 (71.4)
Divorced	9 (11.3)	27 (9)
Separated	0 (0)	2 (0.7)
Widowed or surviving civil partner	7 (8.8)	46 (15.3)
Household status, n (%)		
Lives alone	9 (11.3)	52 (17.3)
Lives with spouse	65 (81.3)	203 (67.4)
Lives with spouse and other relatives	3 (3.8)	25 (8.3)
Lives with other relatives	3 (3.8)	21 (7)
Residence area, n (%)		
Urban area	70 (87.5)	267 (88.7)
Rural area	10 (12.5)	34 (11.3)
Formal education, n (%)		
ISCED[a] 1 Primary education	17 (21.3)	116 (38.5)
ISCED 2 Lower secondary education	5 (6.25)	22 (7.3)
ISCED 3 Upper secondary education	20 (25)	51 (17)
ISCED 4 Post-secondary but not tertiary education	8 (10)	22 (7.3)
ISCED 5 Short-cycle tertiary education	8 (10)	27 (9)
ISCED 6 Bachelor or equivalent level	17 (21.3)	50 (16.6)
ISCED 7 Master or equivalent level	2 (12.5)	6 (2)
ISCED 8 Doctoral or equivalent level	3 (3.8)	7 (2.3)
Employment status, n (%)		
Employed	6 (7.5)	15 (5)
Self-employed	3 (3.8)	9 (3)
Retired	69 (86.3)	271 (90)
Unemployed	0 (0)	2 (0.7)
Fulfilling domestic tasks	1 (1.3)	3 (1)
Other	1 (1.3)	1 (0.3)
Diagnosis of a chronic condition, n (%)		
Yes	58 (72.5)	252 (83.7)
No	22 (27.5)	49 (16.3)
Chronic condition (top 3), n (%)		
Diabetes	25 (31.3)	179 (59.5)
Dyslipidemia	8 (10)	45 (15)
High blood pressure	29 (36.3)	116 (38.5)

XSL•FO
RenderX

Characteristics	Study 1 (n=80)	Study 2 (n=301)
Usual prescribed medication, n (%)		
Yes	68 (85)	279 (92.7)
No	12 (15)	22 (7.3)
Number of medicines, mean	3.10	3.10

[a]ISCED: International Standard Classification of Education.

Study 2

The sample comprised 170 male (56.5%) and 131 female (43.5%) participants, with a mean age of 72 (SD 5.15) years, ranging from 65 to 88 years. From the total sample (n=301), most of the participants were married or had civil partners (n=215, 71.4%), living with their spouse (n=267, 88.7%) in urban areas (n=267, 88.7%). In addition, most of the respondents were retired (n=271, 90%). Regarding formal education, this was analyzed according to the ISCED, showing that the most general education level was primary education—ISCED 1 (n=116, 38.5%), followed by lower secondary education—ISCED 2 (n=51, 17%) and bachelor or equivalent degree—ISCED 6 (n=37, 12.3%). Concerning the current health status, most participants reported a diagnosed chronic condition (n=252, 83.7%), mainly diabetes (n=179, 59.5%), dyslipidemia (n=45, 15%), and high blood pressure (n=116, 38.5%). Regarding prescribed medication, 92.7% (n=279) of the sample answered that they usually take medicines, with 3 being the average number of medicines per person and 10 being the maximum number. In this study, 19 questionnaires were excluded since they were not entirely fulfilled. Table 1 presents the sociodemographic data of both studies.

Stability

Study 1

From the stability analysis (Pearson correlation coefficient) of the eHEALS-PT24, correlations between all items of the scale were identified, ranging between moderate and strong (except for one very strong correlation), with statistical significance (P<.001). The lowest correlation value (r=0.42) was between items 1 and 8 (moderate correlation). In turn, the highest value (r=0.91) was a very strong correlation between items 2 and 3 (Table 2).

XSL•FO
RenderX

Table 2. Pearson correlations for study 1.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
Item 1								
r	1	0.66[a]	0.59[a]	0.60[a]	0.59[a]	0.44[a]	0.51[a]	0.42[a]
P value	—[b]	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Item 2								
r	0.66[a]	1	0.91[a]	0.71[a]	0.69[a]	0.61[a]	0.52[a]	0.46[a]
P value	<.001	—	<.001	<.001	<.001	<.001	<.001	<.001
Item 3								
r	0.59[a]	0.91[a]	1	0,63[a]	0.65[a]	0.54[a]	0.45[a]	0.47[a]
P value	<.001	<.001	—	<.001	<.001	<.001	<.001	<.001
Item 4								
r	0.60[a]	0.71[a]	0,63[a]	1	0.76[a]	0.60[a]	0.64[a]	0.42[a]
P value	<.001	<.001	<.001	—	<.001	<.001	<.001	<.001
Item 5								
r	0.59[a]	0.69[a]	0.65[a]	0.76[a]	1	0.55[a]	0.55[a]	0.45[a]
P value	<.001	<.001	<.001	<.001	—	<.001	<.001	<.001
Item 6								
r	0.44[a]	0.61[a]	0.54[a]	0.60[a]	0.55[a]	1	0.70[a]	0.46[a]
P value	<.001	<.001	<.001	<.001	<.001	—	<.001	<.001
Item 7								
r	0.51[a]	0.52[a]	0.45[a]	0.64[a]	0.55[a]	0.70[a]	1	0.47[a]
P value	<.001	<.001	<.001	<.001	<.001	<.001	—	<.001
Item 8								
r	0.42[a]	0.46[a]	0.47[a]	0.42[a]	0.45[a]	0.46[a]	0.47[a]	1
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	—

[a]The correlation is significant at a significance level of .01 (2-tailed).

[b]Not applicable.

Study 2

From the stability analysis (Pearson correlation coefficient) of the eHEALS-PT24, correlations between all items of the scale were identified, ranging between moderate and strong (except for one very strong correlation), with statistical significance (P<.001). The lowest correlation value (r=0.81) was between items 1 and 8 (strong correlation). In turn, the highest value (r=0.96) was a very strong correlation between items 2 and 3 (Table 3).

XSL•FO
RenderX

Table 3. Pearson correlations for study 2.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
Item 1								
r	1	0.91[a]	0.90[a]	0.87[a]	0.87[a]	0.83[a]	0.82[a]	0.81[a]
P value	—[b]	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Item 2								
r	0.91[a]	1	0.96[a]	0.90[a]	0.88[a]	0.86[a]	0.83[a]	0.83[a]
P value	<.001	—	<.001	<.001	<.001	<.001	<.001	<.001
Item 3								
r	0.90[a]	0.96[a]	1	0.90[a]	0.90[a]	0.86[a]	0.83[a]	0.83[a]
P value	<.001	<.001	—	<.001	<.001	<.001	<.001	<.001
Item 4								
r	0.87[a]	0.90[a]	0.90[a]	1	0.93[a]	0.87[a]	0.85[a]	0.83[a]
P value	<.001	<.001	<.001	—	<.001	<.001	<.001	<.001
Item 5								
r	0.87[a]	0.88[a]	0.90[a]	0.93[a]	1	0.87[a]	0.85[a]	0.85[a]
P value	<.001	<.001	<.001	<.001	—	<.001	<.001	<.001
Item 6								
r	0.83[a]	0.86[a]	0.86[a]	0.87[a]	0.87[a]	1	0.87[a]	0.83[a]
P value	<.001	<.001	<.001	<.001	<.001	—	<.001	<.001
Item 7								
r	0.82[a]	0.83[a]	0.83[a]	0.85[a]	0.85[a]	0.87[a]	1	0.85[a]
P value	<.001	<.001	<.001	<.001	<.001	<.001	—	<.001
Item 8								
r	0.81[a]	0.83[a]	0.83[a]	0.83[a]	0.85[a]	0.83[a]	0.85[a]	1
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	—

[a]The correlation is significant at a significance level of .01 (2-tailed).

[b]Not applicable.

Reliability

Study 1

The analysis of the internal consistency of the eHEALS-PT24 showed an adequate Cronbach α coefficient (Cronbach α=0.92). The statistics after excluding 1 of the 8 items did not indicate an increase in reliability: the value of Cronbach α ranged from 0.90 to 0.92. The mean total value in the eHealth literacy for the sample (n=80) was 27.25 (SD 5.61). The average score for each item was 3, ranging between 3.09 (item 8) and 3.55 (item 1). Concerning item-total correlation, coefficients above 0.4 for all items showed that the items were consistent with each other and correlated with the final score (Table 4).

XSL•FO
RenderX

Table 4. eHEALS-PT24[a] means, scale reliability after removing an item, and item-total correlation.

Study and items	Mean (SD)	Mean if item deleted	α if item deleted	Variance of the scale if item deleted	Item-total correlation[b]
Study 1 (n=80)					
Item 1	3.55 (0.80)	23.70	0.907	25.35	0.68
Item 2	3.53 (0.87)	23.73	0.894	23.52	0.83
Item 3	3.51 (0.87)	23.74	0.900	24.09	0.77
Item 4	3.49 (0.94)	23.76	0.898	23.35	0.79
Item 5	3.40 (0.88)	23.85	0.900	24.05	0.77
Item 6	3.34 (0.94)	23.91	0.906	24.11	0.70
Item 7	3.35 (0.89)	23.90	0.907	24.60	0.69
Item 8	3.09 (0.87)	24.16	0.918	25.78	0.55
Sum score, mean (SD)	27.25 (5.61)	—[c]	—	—	—
Study 2 (n=301)					
Item 1	2.99 (1.27)	20.33	0.978	69.58	0.91
Item 2	2.95 (1.30)	20.37	0.977	68.89	0.94
Item 3	2.97 (1.28)	20.34	0.977	69.08	0.94
Item 4	2.99 (1.30)	20.32	0.977	68.99	0.93
Item 5	2.96 (1.28)	20.35	0.977	69.31	0.94
Item 6	2.88 (1.26)	20.43	0.978	70.10	0.91
Item 7	2.87 (1.24)	20.44	0.979	70.74	0.89
Item 8	2.70 (1.22)	20.61	0.980	71.27	0.88
Sum score, mean (SD)	23.31 (9.53)	—	—	—	—

[a]eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults.

[b]All item-to-total correlations were significant as $P<.001$.

[c]Not applicable.

Study 2

The analysis of the internal consistency of the eHEALS-PT24 showed an adequate Cronbach α coefficient (Cronbach α=0.98). The statistics after excluding 1 of the 8 items did not indicate an increase in reliability, with Cronbach α coefficient values remaining stable at 0.98. The mean total value in the eHealth literacy for the sample (n=301) was 23.31 (SD 9.53). The average score for each item was 3, ranging between 2.7 (item 8) and 2.99 (items 1 and 4). Concerning item-total correlation, coefficients above 0.40 for all items showed that the items were consistent with each other and correlated with the final score (Table 4).

Construct Validity: EFA

The Bartlett sphericity test corroborated the factorability of the correlation matrix for both studies (study 1: $\chi^2_{28}=446.87$, $P<.001$; study 2: $\chi^2_{28}=3932.81$, $P<.001$). In addition, the KMO test value demonstrated adequate sampling for both studies (study 1: KMO=0.87; study 2: KMO=0.94). Given the quality of Bartlett and KMO values, the criteria for factor analysis were gathered. By performing the EFA, we considered factors with eigenvalues above 1 for testing the structure of the instrument. In accordance with the original structure of the eHEALS [10], the Jamovi software extracted 1 factor for the structure of the eHEALS-PT24 for both studies (Study 1 and 2).

Study 1

In the EFA of study 1 (n=80), a single factor showed moderate to strong loadings (0.57 to 0.89; Table 5) but poor model fit indicators: RMSEA was high (0.20), TLI was below the threshold (0.79), and there was a significant chi-square test ($\chi^2_{20}=83.7$, $P<.001$). The factor explained 58.3% of the total variance.

XSL•FO
RenderX

Table 5. Factor loadings after varimax rotation.

eHEALS-PT24[a] items	Study 1 (n=80), factor 1	Study 2 (n=301), factor 1
Item 1	0.72	0.92
Item 2	0.89	0.95
Item 3	0.82	0.95
Item 4	0.84	0.95
Item 5	0.81	0.95
Item 6	0.72	0.92
Item 7	0.70	0.90
Item 8	0.57	0.89

[a]eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults.

Study 2

In the EFA of study 2 (n=301), a single factor showed moderate to strong loadings (0.89 to 0.95) but poor model fit indicators: RMSEA was high (0.20), TLI was marginally acceptable (0.91), and there was a significant chi-square test (χ^2_{20}=271, P<.001). The factor explained 86.4% of the total variance.

The single-factor structure of eHEALS-PT24 for both studies (studies 1 and 2) was also empirically confirmed on screen plots (Figures 2 and 3).

Figure 2. Screen plot for the eHEALS-PT24 (study 1). eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults.

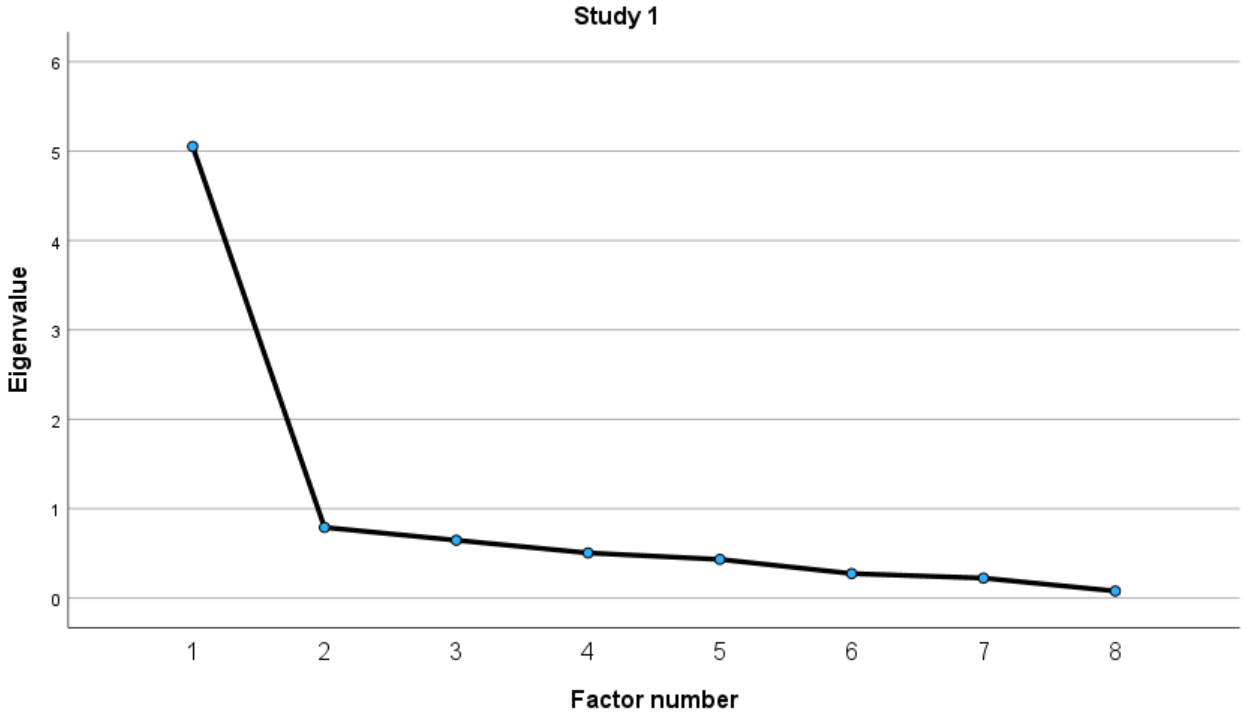

XSL•FO
RenderX

Figure 3. Screen plot for the eHEALS-PT24 (study 2). eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults.

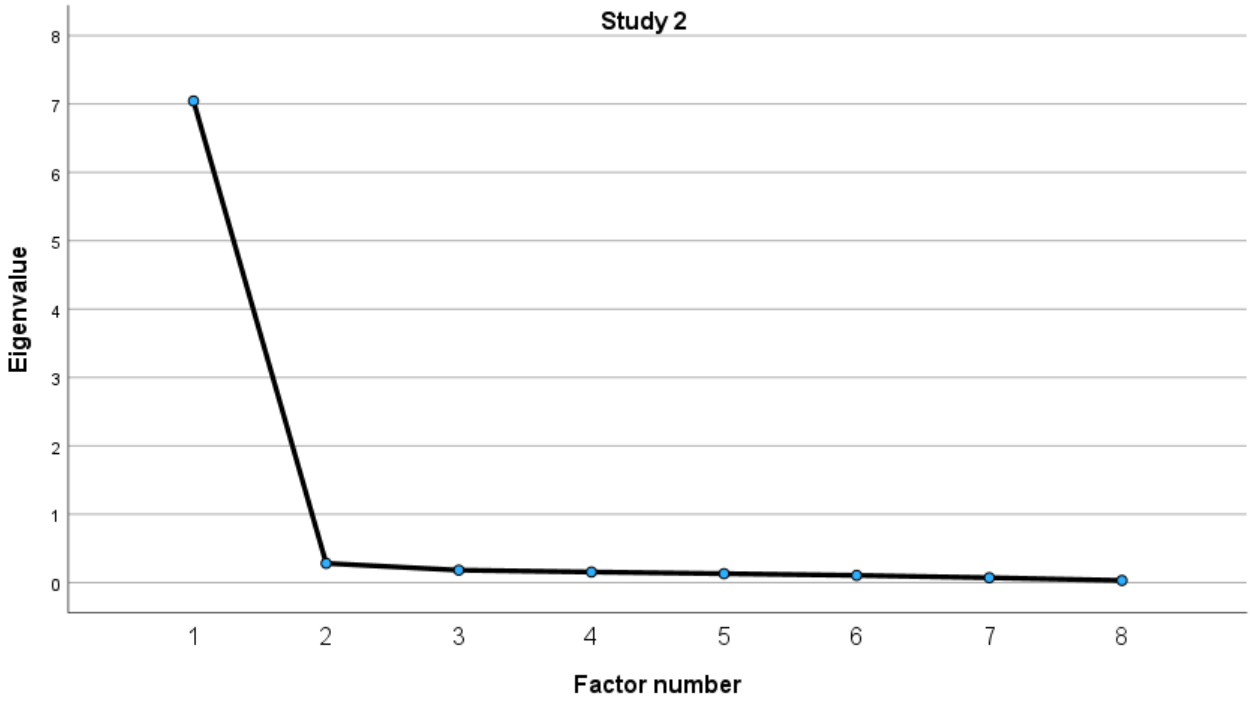

Construct Validity: CFA

Regarding CFA for study 2 (n=301), a single-factor model demonstrated high item loadings (standardized estimates 0.88 to 0.97; Table 6); however, the model fit was mixed: the chi-square test was significant (χ^2_{20}=265, *P*<.001), CFI was acceptable (0.94), while TLI (0.91) was marginally acceptable, and RMSEA (0.20) indicated a poor fit. Given the RMSEA values for one dimension, and due to the variability of factor structures in the literature, we compared the indexes for 1-, 2-, and 3-factor structures in study 2 (Table 7).

Table 6. Factor loadings after varimax rotation.

eHEALS-PT24[a] items	Study 2 (n=301)	Study 2 (n=301)		Study 2 (n=301)		
	Factor 1	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
Item 1	0.93	0.93	—[b]	0.94	—	—
Item 2	0.96	0.97	—	0.98	—	—
Item 3	0.97	0.97	—	—	0.97	—
Item 4	0.94	0.94	—	—	0.94	—
Item 5	0.94	0.94	—	—	0.94	—
Item 6	0.91	—	0.93	—	—	0.93
Item 7	0.89	—	0.93	—	—	0.93
Item 8	0.88	—	0.90	—	—	0.90

[a]eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults.

[b]Not available.

XSL•FO
RenderX

Table 7. Global model fit indices.

Indices	1 Factor	2 Factors	3 Factors
RMSEA[a] (90% CI)	0.20	0.17	0.18
CFI[b]	0.94	0.96	0.96
TLI[c]	0.91	0.94	0.93
Chi-square test, *P* value	<.001	<.001	<.001

[a]RMSEA: root mean square error of approximation.

[b]CFI: comparative fit index.

[c]TLI: Tucker-Lewis Index.

For the 2-factor model, we followed Tomás et al [32], dividing items 1-5 (factor 1) and items 6-8 (factor 2). High item loadings (standardized estimates 0.90 to 0.97) were obtained (Table 6); however, the model fit was mixed: CFI was acceptable (0.96), TLI (0.94) was marginally acceptable, RMSEA (0.17) indicated a poor fit, and the chi-square test was significant (χ^2_{19}=188, *P*<.001; Table 7).

For the 3-factor model, we followed Stellefson et al [43], dividing items 1-2 (factor 1), items 3-5 (factor 2), and items 6-8 (factor 3). High item loadings (standardized estimates 0.90 to 0.98) were obtained (Table 6); however, the model fit was mixed: CFI was acceptable (0.96), TLI (0.93) was marginally acceptable, RMSEA (0.18) indicated a poor fit, and the chi-square test was significant (χ^2_{17}=179, *P*<.001; Table 7).

Interpretation of the eHEALS-PT24 Scores

We used the mean total and SD values to determine high or low eHealth literacy levels. Therefore, participants were divided into 2 groups for each study as follows:

- Study 1: (1) low eHealth literacy levels (≤27 points) and (2) high eHealth literacy levels (>32 points).
- Study 2: (1) low eHealth literacy levels (≤23 points) and (2) high eHealth literacy levels (>31 points).

In study 1 (n=80), 35 (43.8%) participants had low eHealth literacy levels, and 9 (11.3%) had high levels. In study 2 (n=301), 115 (38.2%) participants had low eHealth literacy levels, and 71 (23.6%) had high levels.

Construct Validity: Known-Groups

We performed known-groups validity in study 2, as presented in the following subsections.

Association Between eHealth Literacy and Age (Student t test and Pearson Correlation)

To analyze the association between eHealth literacy and age, participants were distributed into two different groups: (1) 65-79 years and (2) ≥80 years. The first group comprised 269 (89.4%) participants, and the second group comprised 32 (10.6%) participants. The analysis showed differences statistically significant between eHealth literacy levels (total scale value) and age (t_{299}=3.94; *P*<.001). Individuals aged 65-79 years had, on average, 7.0 points higher eHEALS scores than those aged ≥80 years. We also carried out a correlation analysis between both variables (Pearson correlation), which showed that variables are inversely related, which means the greater one variable, the smaller the other (*r*=–0.355, *P*<.001).

Association Between eHealth Literacy and Sex (Student t test)

Female respondents showed an average of 23.68 (SD 9.89) points and male respondents showed an average of 23.03 (SD 9.26) points; however, no statistically significant differences between eHealth literacy levels (total scale value) and sex were found (t_{299}=0.59; *P*=.31).

Association Between eHealth Literacy and Residence Area (Student t test)

Statistically significant differences between eHealth literacy levels (total scale value) and residence area were found (t_{299}=4.19; *P*<.001). Participants living in urban areas showed an average of 24.11 (SD 9.16) points, and those living in rural areas showed an average of 17.03 (SD 10.15) points.

Association Between eHealth Literacy and Education Level (1-Way ANOVA or Kruskal-Wallis Test)

Statistically significant differences between eHealth literacy levels (total scale value) and education level were found (F_8=10.24; *P*<.001), with the highest mean values corresponding to the education levels of doctoral or equivalent level (ISCED 8; mean score of eHealth literacy levels: 31.57, SD 4.20), master's degree or equivalent (ISCED 7; mean 29.17, SD 5.60), and postsecondary but not tertiary education (ISCED 4; mean 27.68, SD 8.52). Since some of the education levels had nonrepresentative samples (ie, ISCEDs 2, 4, 5, 7, and 8), we tested the association between eHealth literacy and education level using Kruskal-Wallis test (H=58.56; *P*<.001). In this analysis, we found statistically significant differences between participants with primary education (ISCED 1) and all other education levels, with an exception for lower secondary education (ISCED 2) and master's degree or equivalent (ISCED 7). No other statistically significant differences were found.

The mean (SD) and significance of eHEALS-PT24 by subgroups (ie, age, sex, residence area, and education level) calculated in study 2 are presented in Table 8.

XSL•FO
RenderX

Table 8. Mean (SD) and significance of eHEALS-PT24[a] score by subgroups (study 2; n=301).

Variable	Frequency, n (%)	eHEALS-PT24 score, mean (SD)	eHEALS-PT24, *P* value
Age group (years)			<.001
65-79	269 (89.37)	24.04 (9.23)	
≥80	32 (10.63)	17.19 (9.99)	
Sex			.56
Female	131 (43.52)	23.68 (9.89)	
Male	170 (56.48)	23.03 (9.26)	
Residence area			<.001
Urban	267 (88.7)	24.11 (9.16)	
Rural	34 (11.3)	17.03 (10.15)	
Education level			<.001[c]
ISCED[b] 1: primary education	116 (38.54)	17.84 (9.91)	
ISCED 2: lower secondary education	22 (7.31)	24.86 (8.67)	
ISCED 3: upper secondary education	51 (16.94)	26.67 (7.25)	
ISCED 4: postsecondary but not tertiary education	22 (7.31)	27.68 (8.52)	
ISCED 5: short-cycle tertiary education	27 (8.97)	27.07 (6.71)	
ISCED 6: bachelor's degree or equivalent level	50 (16.61)	26.08 (7.62)	
ISCED 7: master's degree or equivalent level	6 (1.99)	29.17 (5.60)	
ISCED 8: doctoral or equivalent level	7 (2.33)	31.57 (4.20)	

[a]eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults.

[b]ISCED: International Standard Classification of Education.

[c]ANOVA and Kruskal-Wallis.

Discussion

Principal Findings and Comparison With Previous Work

In this paper, we have presented the translation, cultural adaptation, and validation of the eHEALS for Portuguese older adults (eHEALS-PT24) by carrying out a general psychometric analysis in a sample of 80 participants (study 1) and a CFA including 301 participants (study 2). For this discussion, we highlight the findings obtained in study 2 since it represents the final stage of the validation of eHEALS-PT24, according to the followed protocol [49].

Based on scale and items stability and reliability analysis, we concluded that the eHEALS-PT24 is a reliable tool for measuring eHealth literacy. Correlations between all items of the eHEALS-PT24, ranging between moderate and strong ones, with statistical significance, indicated that the questions were related to the instrument with no redundancy. In addition, the instrument showed good internal consistency. All items were considered reliable to assess eHealth literacy levels in older adults, with high correlations between each other. Compared with the original validation study [10], internal consistency was higher (Cronbach α=0.88 vs 0.98), which is comparable with previous studies among older adults [24,36,39,43,44,48]. The only study showing a higher Cronbach α than ours was a validation in older Hispanic participants [48] (Cronbach α=0.99), but the sample was small (n=20), which was pointed out as a limitation by the authors.

Concerning construct validity, the eHEALS was originally developed with a single-factor structure [10], and in our study, the same structure was yielded in EFA, explaining 86.4% of the total variance. This total variance explained was higher compared with the original validation study [10] (56%), as well as to previous studies among older adults [24,36,39,43,44,48], suggesting the model captured a significant portion of the data's variation. Regarding CFA, a single-factor model demonstrated high item loadings; however, the model fit was mixed: the chi-square test was significant (*P*<.001), CFI (0.94) and TLI (0.91) were both acceptable, and RMSEA (0.20) indicated a poor fit. The finding of a unidimensional structure also aligns with previous studies carried out in older adults [24,36,39]; however, it contradicts the results of the 2 other studies conducted among older adults, which indicated a better fit for the 3-factor structure [43,44]. The variability of factor structures in the literature and the RMSEA value obtained led us to compare the indices (ie, chi-square, CFI, TLI, and RMSEA) for 1-, 2-, and 3-factor structures. All indexes fitted well, but RMSEA remained poor (ie, 0.20 for 1 factor, 0.17 for 2 factors,

XSL•FO
RenderX

and 0.18 for 3 factors), suggesting that the model may not fit the data well. These findings are in line with the eHEALS Korean validation study (CFI=0.95; TLI=0.92; and RMSEA=0.12) [36], which followed a unidimensional structure. Considering the original validation study [10], as well as others who used Rasch analysis (or other analyses under modern test theory) [25,40,41], we also decided on the 1-factor structure.

The eHEALS-PT24 mean score was 23.31, with 38.2% (115/301) of participants showing low eHealth literacy levels and 23.6% (71/301) showing high levels. Compared with other validation studies among older adults, the mean score was higher than in Spain (22.35) [48] but lower than in China (30.94) [39] and in the United States (29.05) [43]. As shown by previous studies [12,15,18,30,32], item 8 of the eHEALS-PT24 had the lowest average among all scale items, indicating that people do not feel so confident in using information from the internet to make health-related decisions compared with other eHealth literacy skills.

To further examine construct validity, known-groups validity was assessed, showing statistically significant differences in demographic variables (age, residence area, and education levels), except for sex. While a previous study among older adults [48] found no differences between eHealth literacy levels and age, we concluded that the greater the age, the smaller the eHealth literacy. Those aged 65-79 years had, on average, 7.0 points higher (eHEALS scores) than those aged ≥80 years. Regarding eHealth literacy levels and sex, Aponte and Nokes [48] pointed out significant differences, where mean values of eHEALS scores for men (13.85) were significantly lower than those in women (25.77). In our study, female respondents had, on average, 0.65 points higher eHEALS scores than male respondents; however, no statistically significant differences were found between eHealth literacy levels and sex (P=.31). We also tested the association of eHealth literacy levels and the residence area, finding statistically significant differences (P<.001), where participants living in urban areas had on average 7.08 points higher (eHEALS scores) than those living in rural areas. Regarding education levels, the highest mean values corresponded to doctoral or equivalent level (ISCED 8; 31.57), master's degree or equivalent (ISCED 7; 29.17), and postsecondary but not tertiary education (ISCED 4; 27.68). Since some education-level cases had nonrepresentative samples (n<30), this result should be read with caution. Therefore, we can only infer participants with primary education (ISCED 1) had significant differences in eHealth literacy levels from the participants with other education levels, except for lower secondary education (ISCED 2) and master's degree or equivalent (ISCED 7).

Strengths and Limitations

There are some limitations in this study. First, there was limited geographic representation since the sample did not include older people from all regions of the country (only Faro and Lisbon). If we consider, however, the sample representativeness, we believe it is suitable for providing us with an adequate snapshot of our population of interest. Second, since the web-based questionnaires were only applied in the pilot study, we did not have any rigorous confirmatory process to determine that participants were aged, in fact, aged 65 years or older. Third, similarly to other validation studies [25,27,30-32,43,44,48], test-retest reliability was not conducted. As such, further research could address reliability and responsiveness over time. Fourth, although a more comprehensive psychometric analysis could be conducted to establish the instrument validity, such as investigating correlations with external measures of eHealth literacy or health outcomes, the results of this study are promising and show that the eHEALS-PT24 is a reliable tool for perceived measurement of eHealth literacy. Fifth, the eHEALS instrument measures a narrow scope of eHealth literacy and may not fully capture the complex concept of eHealth literacy nowadays since it was developed before the social media era (Web 2.0). Nevertheless, the benefits of eHEALS (strong psychometric properties, brevity, and usability) outweigh its shortcomings, and the eHEALS arguably serves as a convenient instrument for health professionals.

Conclusions

In conclusion, the eHEALS-PT24 is a reliable and valid instrument for assessing and monitoring the eHealth literacy of Portuguese older adults. This instrument can be useful in identifying older adults who are able to use eHealth resources and participate in eHealth interventions aimed at engaging them in health care and helping them to manage their health and, by extension, assess the effects of eHealth interventions.

Acknowledgments

We thank Miriam Almeida (RN, MSc) for assistance with the recruitment phase.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Portuguese version of the eHealth Literacy Scale for older adults (eHEALS-PT24).
[PDF File (Adobe PDF File), 95 KB - jmir_v27i1e57730_app1.pdf]

References

XSL•FO
RenderX

1. European citizen's digital health literacy. European Commission. 2014. URL: https://europa.eu/eurobarometer/surveys/detail/2020 [accessed 2025-02-04]
2. Portz JD. A review of web-based chronic disease self-management for older adults. Gerontechnology 2017;16(1):12-20 [FREE Full text] [doi: 10.4017/gt.2017.16.1.002.00] [Medline: 30705614]
3. Buyl R, Beogo I, Fobelets M, Deletroz C, van Landuyt P, Dequanter S, et al. e-Health interventions for healthy aging: a systematic review. Syst Rev 2020;9(1):128 [FREE Full text] [doi: 10.1186/s13643-020-01385-8] [Medline: 32493515]
4. Renzi E, Baccolini V, Migliara G, de Vito C, Gasperini G, Cianciulli A, et al. The Impact of eHealth interventions on the improvement of self-care in Chronic patients: an overview of systematic reviews. Life (Basel) 2022;12(8):1253 [FREE Full text] [doi: 10.3390/life12081253] [Medline: 36013432]
5. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
6. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
7. Jung SO, Son YH, Choi E. E-health literacy in older adults: an evolutionary concept analysis. BMC Med Inform Decis Mak 2022;22(1):28 [FREE Full text] [doi: 10.1186/s12911-022-01761-5] [Medline: 35101005]
8. Health literacy: the solid facts. World Health Organization. URL: https://www.who.int/europe/publications/i/item/9789289000154 [accessed 2013-10-01]
9. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
10. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
11. Lee J, Lee E, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021;23(11):e30644 [FREE Full text] [doi: 10.2196/30644] [Medline: 34779781]
12. Mialhe FL, Moraes KL, Sampaio HADC, Brasil VV, Vila VDSC, Soares GH, et al. Evaluating the psychometric properties of the eHealth literacy scale in brazilian adults. Rev Bras Enferm 2021;75(1):e20201320 [FREE Full text] [doi: 10.1590/0034-7167-2020-1320] [Medline: 34495134]
13. Koo M, Norman C, Hsiao-Mei C. Psychometric evaluation of a Chinese version of the eHealth literacy scale (eHEALS) in school age children. Int Electron J Health Educ 2012;15(1):29-36 [FREE Full text]
14. Chang A, Schulz PJ. The measurements and an elaborated understanding of Chinese eHealth literacy (C-eHEALS) in chronic patients in China. Int J Environ Res Public Health 2018;15(7):1553 [FREE Full text] [doi: 10.3390/ijerph15071553] [Medline: 30041420]
15. Ma Z, Wu M. The psychometric properties of the Chinese eHealth literacy scale (C-eHEALS) in a Chinese rural population: cross-sectional validation study. J Med Internet Res 2019;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
16. Xu RH, Zhou L, Lu SY, Wong EL, Chang J, Wang D. Psychometric validation and cultural adaptation of the simplified Chinese eHealth literacy scale: cross-sectional study. J Med Internet Res 2020;22(12):e18613 [FREE Full text] [doi: 10.2196/18613] [Medline: 33284123]
17. Shiferaw KB. Validation of the ethiopian version of eHealth literacy scale (ET-eHEALS) in a population with chronic disease. Risk Manag Healthc Policy 2020;13:465-471 [FREE Full text] [doi: 10.2147/RMHP.S240829] [Medline: 32547277]
18. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
19. Juvalta S, Kerry MJ, Jaks R, Baumann I, Dratva J. Electronic health literacy in swiss-german parents: cross-sectional study of eHealth literacy scale unidimensionality. J Med Internet Res 2020;22(3):e14492 [FREE Full text] [doi: 10.2196/14492] [Medline: 32167476]
20. Efthymiou A, Middleton N, Charalambous A, Papastavrou E. Adapting the eHealth literacy scale for carers of people with chronic diseases (eHeals-Carer) in a sample of greek and cypriot carers of people with dementia: reliability and validation study. J Med Internet Res 2019;21(11):e12504 [FREE Full text] [doi: 10.2196/12504] [Medline: 31778120]
21. Zrubka Z, Hajdu O, Rencz F, Baji P, Gulácsi L, Péntek M. Psychometric properties of the hungarian version of the eHealth literacy scale. Eur J Health Econ 2019;20(Suppl 1):57-69 [FREE Full text] [doi: 10.1007/s10198-019-01062-1] [Medline: 31098883]
22. Wijaya MC, Kloping YP. Validity and reliability testing of the indonesian version of the eHealth literacy scale during the COVID-19 pandemic. Health Informatics J 2021;27(1):1460458220975466 [FREE Full text] [doi: 10.1177/1460458220975466] [Medline: 33446030]
23. Bazm S, Mirzaei M, Fallahzadeh H, Bazm R. Validity and reliability of Iranian version of eHealth literacy scale. J Community Health Res 2016;5(2):121-130 [FREE Full text]

XSL•FO
RenderX

24. Lin C, Broström A, Griffiths MD, Pakpour AH. Psychometric evaluation of the persian eHealth literacy scale (eHEALS) among elder iranians with heart failure. Eval Health Prof 2020;43(4):222-229. [doi: 10.1177/0163278719827997] [Medline: 30744419]
25. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the eHealth literacy scale using item response and classical test theory methods. J Med Internet Res 2017;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
26. del Giudice P, Bravo G, Poletto M, de Odorico A, Conte A, Brunelli L, et al. Correlation between eHealth literacy and health literacy using the eHealth literacy scale and real-life experiences in the health sector as a proxy measure of functional health literacy: cross-sectional web-based survey. J Med Internet Res 2018;20(10):e281 [FREE Full text] [doi: 10.2196/jmir.9401] [Medline: 30381283]
27. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
28. Brørs G, Wentzel-Larsen T, Dalen H, Hansen TB, Norman CD, Wahl A, et al. Psychometric properties of the norwegian version of the electronic health literacy scale (eHEALS) among patients after percutaneous coronary intervention: cross-sectional validation study. J Med Internet Res 2020;22(7):e17312 [FREE Full text] [doi: 10.2196/17312] [Medline: 32720900]
29. Dale JG, Lüthi A, Fundingsland Skaraas B, Rundereim T, Dale B. Testing measurement properties of the norwegian version of electronic health literacy scale (eHEALS) in a group of day surgery patients. J Multidiscip Healthc 2020;13:241-247 [FREE Full text] [doi: 10.2147/JMDH.S242985] [Medline: 32210568]
30. Duplaga M, Sobecka K, Wójcik S. The reliability and validity of the telephone-based and online polish eHealth literacy scale based on two nationally representative samples. Int J Environ Res Public Health 2019;16(17):1-12 [FREE Full text] [doi: 10.3390/ijerph16173216] [Medline: 31484338]
31. Burzyńska J, Rękas M, Januszewicz P. Evaluating the psychometric properties of the eHealth literacy scale (eHEALS) among polish social media users. Int J Environ Res Public Health 2022;19(7):4067 [FREE Full text] [doi: 10.3390/ijerph19074067] [Medline: 35409753]
32. Tomás C, Queirós P, Ferreira T. Analysis of the psychometric properties of the portuguese version of an eHealth literacy assessment tool. Rev Enf Ref (IV serie) 2014;2:19-28. [doi: 10.12707/RIV14004]
33. Gazibara T, Cakic J, Cakic M, Pekmezovic T, Grgurevic A. eHealth and adolescents in serbia: psychometric properties of eHeals questionnaire and contributing factors to better online health literacy. Health Promot Int 2019;34(4):770-778. [doi: 10.1093/heapro/day028] [Medline: 29800141]
34. Chung S, Park BK, Nahm E. The korean eHealth literacy scale (K-eHEALS): reliability and validity testing in younger adults recruited online. J Med Internet Res 2018;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
35. Gartrell K, Han K, Trinkoff A, Cho H. Three-factor structure of the eHealth literacy scale and its relationship with nurses' health-promoting behaviours and performance quality. J Adv Nurs 2020;76(10):2522-2530. [doi: 10.1111/jan.14490] [Medline: 33463741]
36. Kim H, Yang E, Ryu H, Kim HJ, Jang SJ, Chang SJ. Psychometric comparisons of measures of eHealth literacy using a sample of Korean older adults. Int J Older People Nurs 2021;16(3):e12369. [doi: 10.1111/opn.12369] [Medline: 33527701]
37. Pérez G, Almagro BJ, Hernando Gómez Á, Aguaded Gómez JI. Validácion de la escala eHealth literacy (eHEALS) en población universitaria española. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
38. Wångdahl J, Jaensson M, Dahlberg K, Nilsson U. The swedish version of the electronic health literacy scale: prospective psychometric evaluation study including thresholds levels. JMIR Mhealth Uhealth 2020;8(2):e16316 [FREE Full text] [doi: 10.2196/16316] [Medline: 32130168]
39. Chung S, Nahm E. Testing reliability and validity of the eHealth literacy scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015;33(4):150-156 [FREE Full text] [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
40. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct validity of the eHealth literacy scale (eHEALS) among two adult populations: a rasch analysis. JMIR Public Health Surveill 2016;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
41. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
42. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
43. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: cross-sectional survey. J Med Internet Res 2017;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]

XSL•FO
RenderX

44. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the measurement properties of the eHealth literacy scale (eHEALS) among baby boomers: a multinational test of measurement invariance. J Med Internet Res 2017;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
45. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: measurement invariance study. J Med Internet Res 2018;20(7):e10434. [doi: 10.2196/10434] [Medline: 29986848]
46. Hyde LL, Boyes AW, Evans T, Mackenzie LJ, Sanson-Fisher R. Three-factor structure of the eHealth literacy scale among magnetic resonance imaging and computed tomography outpatients: a confirmatory factor analysis. JMIR Hum Factors 2018;5(1):e6 [FREE Full text] [doi: 10.2196/humanfactors.9039] [Medline: 29459356]
47. Holch P, Marwood JR. EHealth literacy in UK teenagers and young adults: exploration of predictors and factor structure of the eHealth literacy scale (eHEALS). JMIR Form Res 2020;4(9):e14450 [FREE Full text] [doi: 10.2196/14450] [Medline: 32897230]
48. Aponte J, Nokes KM. Electronic health literacy of older Hispanics with diabetes. Health Promot Int 2017;32(3):482-489. [doi: 10.1093/heapro/dav112] [Medline: 26681770]
49. Sousa VD, Rojjanasrirat E. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
50. Clark-Carter D. Quantitative Psychological Research. The Complete Student's Companion. 3rd edition. East Sussex, England: Psychology Press - Taylor & Francis; 2010.
51. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55. [doi: 10.1080/10705519909540118]

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
eHEALS: eHealth Literacy Scale
eHEALS-PT24: Portuguese version of the eHealth Literacy Scale for older adults
ICT: information and communication technology
I-CVI: content validity index at the item level
ISCED: International Standard Classification of Education
KMO: Kaiser-Meyer-Olkin
RMSEA: root mean square error of approximation
S-CVI: content validity index at the scale level
TLI: Tucker-Lewis Index

Edited by A Coristine; submitted 25.02.24; peer-reviewed by A Hassan, M Sarhan, J Mistry, E Vashishtha; comments to author 20.04.24; revised version received 15.06.24; accepted 23.10.24; published 26.02.25.

Please cite as:
Luz S, Nogueira P, Costa A, Henriques A
Psychometric Analysis of the eHealth Literacy Scale in Portuguese Older Adults (eHEALS-PT24): Instrument Development and Validation
J Med Internet Res 2025;27:e57730
URL: https://www.jmir.org/2025/1/e57730
doi:10.2196/57730
PMID:

©Sara Luz, Paulo Nogueira, Andreia Costa, Adriana Henriques. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Short-Video Dependence for e-Mental Health: Development and Validation Study of the Short-Video Dependence Scale

AnHang Jiang[1,2], MAP; Shuang Li[1,2], MAP; HuaBin Wang[1,2], MAP; HaoSen Ni[1,2], MAP; HongAn Chen[1,2], MAP; JunHong Dai[1], MAP; XueFeng Xu[1,2], MAP; Mei Li[1,2*], MAP; Guang-Heng Dong[3*], PhD

[1]Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China

[2]Center for Cognition and Brain Disorders, Affiliated Hospital of Hangzhou Normal University, Hangzhou, China

[3]Department of Psychology, Yunnan Normal University, Kunming, China

[*]these authors contributed equally

Corresponding Author:
Mei Li, MAP
Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments
No. 2318, Yuhangtang Rd, Yuhang District
Hangzhou, 311121
China
Phone: 86 15968849698
Email: maylee530@126.com

Abstract

Background: Short-video dependence (SVD) has become a significant mental health issue around the world. The lack of scientific tools to assess SVD hampers further advancement in this area.

Objective: This study aims to develop and validate a scientific tool to measure SVD levels, ensuring a scientifically determined cutoff point.

Methods: We initially interviewed 115 highly engaged short-video users aged 15 to 63 years. Based on the summary of the interview and references to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria for behavioral addictions, we proposed the first version of the short-video dependence scale (SVDS). We then screened the items through item analysis (second version) and extracted common factors using exploratory factor analysis (third version) and confirmatory factor analysis (final version). Convergent validity was tested with other scales (Chinese Internet Addiction Scale [CIAS] and DSM-5). Finally, we tested the validity of the final version in 16,038 subjects and set the diagnostic cutoff point through latent profile analysis and receiver operating characteristic curve analysis.

Results: The final version of the SVDS contained 20 items and 4 dimensions, which showed strong structural validity (Kaiser-Meyer-Olkin value=0.94) and internal consistency (Cronbach α=.93), and good convergent validity (r_{CIAS}=0.61 and r_{DSM-5}=0.68), sensitivity (0.77, 0.83, 0.87, and 0.62 for each of the 4 dimensions), and specificity (0.75, 0.87, 0.80, and 0.79 for each of the 4 dimensions). Additionally, an SVDS score of 58 was determined as the best cutoff score, and latent profile analysis identified a 5-class model for SVD.

Conclusions: We developed a tool to measure SVD levels and established a threshold to differentiate dependent users from highly engaged nondependent users. The findings provide opportunities for further research on the impacts of short-video use.

(J Med Internet Res 2025;27:e66341) doi:10.2196/66341

KEYWORDS

short-video dependence; problematic short-video use; cutoff point; scale development; mental health; short video; internet addiction; latent profile analysis; exploratory factor analysis; confirmatory factor analysis

XSL•FO
RenderX

Introduction

Short videos are immensely popular digital content, with platforms like TikTok allowing users to create and share videos ranging from a few seconds to a few minutes [1]. These apps engage users by facilitating viewing, creation, and commenting, changing self-presentation and cultural learning. In 2022, TikTok had 1.2 billion monthly active users, with an average viewing time of 168 minutes per day [2]. Hanaysha [3] noted that during the COVID-19 pandemic, short-video apps became key sources of news and entertainment. These apps have impacted various aspects of the world, such as popular culture, content creation and consumption, and the spread of information on political and social issues.

Short-video apps also raised public concerns about their potential negative consequences, such as distraction, poor time management, and mental health problems [1,4,5]. A number of studies have been conducted in this area. For instance, Domoff et al [6] developed a scale to measure the extent of problematic media use considering the need. However, the lack of reliable and valid tools for assessing the corresponding qualities of short-video usage has hindered the progress of research in this field.

Before developing a scale or questionnaire for short-video dependence (SVD), we need to justify not using the term "addiction," which requires careful consideration to establish standards. First, the latest version (fifth edition) of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) recognizes internet gaming disorder (IGD) as a provisional disorder [7], but not short-video disorder. Second, internet-related disorders are often more about specific online activities than excessive internet use [7,8]. Third, some studies have suggested that short-video users show more "dependence" than "addiction," with behavior being more habitual than pathological [9]. Thus, "dependence" is a more appropriate term in this context.

Some researchers have tried to develop internet-related behavioral disorder scales but have faced limitations pertaining to measuring SVD. Güzeller and Coşguner [10] developed a problematic cell phone use scale for Turkish adolescents considering the need for a measurement tool on problematic cell phone use in this population. In addition, these articles were published early, before TikTok was available, and have less applicability to the present context. Domoff et al [6] developed a tool for assessing screen media addiction in children aged 4 to 11 years. The results of the study indicated that the scale has good psychometric properties in terms of factor structure, reliability, and convergent validity for the measurement of problematic media use in children younger than 12 years. However, the limitations of the study include a lack of methodological constraints in assessing screen time and insufficient racial or ethnic diversity of the sample.

The short-video dependence scale (SVDS) was established based on the IGD diagnostic criteria from the DSM-5, since internet addiction has most convincingly been linked to short-video use [11,12], and research addressing dependence behavior is lacking as compared to research on addiction. The DSM-5 is a widely accepted and used classification system for mental disorders that is used to diagnose and categorize mental health problems. It provides standardized definitions and classifications for a variety of mental disorders, including IGD, gambling disorder, etc. According to the DSM-5, a clinical diagnosis of IGD is indicated if 5 or more of the following 9 criteria are met: preoccupation, withdrawal, tolerance, loss of control, loss of interest, continuation, deception, escape, and jeopardizing. By using the DSM-5 as its developmental framework, the SVDS aims to align with the diagnostic criteria and characteristics of IGD, thereby ensuring the validity and comparability of measurements.

Therefore, the DSM-5 criteria for IGD have some advantages in identifying and assessing SVD, but these are not fully applicable or comprehensive. First, SVD is similar to IGD in many ways, such as excessive use, uncontrollable desire, and interference with daily functioning. However, some criteria may not be suitable or sufficient for short-video addiction, such as tolerance, loss of interest, and jeopardizing. For instance, loss of interest may not encompass the full spectrum of motivational factors that affect short-video addiction, such as social comparison, self-presentation, and peer pressure [1]. Furthermore, users may still be able to carry out their daily activities even after spending an excessive amount of time watching short videos, and thus, jeopardizing may not be as severe or uncommon with regard to SVD. Neglecting or compromising other significant facets of life, like relationships, career, or health, might be a more appropriate criterion. The DSM-5 diagnostic criteria for IGD will be improved and refined by the development of the SVDS, which will include extra items identified through in-depth interviews.

Most importantly, to determine the cutoff point for the SVDS among a large sex-balanced sample of Chinese young adults, latent profile analysis (LPA) was used, which is an epidemiological approach to determine the cutoff point when a clinical interview is not available [13-15]. It allows for the identification of distinct groups of short-video users based on their responses in the SVDS, with the group exhibiting the highest levels of SVD being identified as "dependent short-video users" for determining the optimal cutoff point. To date, numerous studies have applied the LPA approach to derive the optimal cutoff points for measurement tools, such as the Bergen Social Media Addiction Scale [16], Online Social Networking Addiction Scale [13], Dimensional Anhedonia Rating Scale [17], and Smartphone Application-Based Addiction Scale [14].

In summary, the field of SVD lacks standardized definitions and measurements for short-video addiction, which may hinder the comparability and generalizability of the results and further hamper the study of its prevalence, impeding crucial future research in the field of SVD. The evolving internet landscape demands a contemporary scale. This study aims to develop and validate a specific SVDS for SVD, drawing on established frameworks and diagnostic criteria from IGD research. The SVDS provides a standardized tool with a clear diagnostic cutoff point to assess SVD and understand its prevalence, correlates, and consequences.

XSL•FO
RenderX

Methods

Ethical Considerations

The experiment conforms to the Code of Ethics of the World Medical Association (Declaration of Helsinki). The Human Investigations Committee of Hangzhou Normal University approved this research (approval number: 20190505). All participants provided written informed consent before experimentation.

Participants

From October 2022 to December 2023, 4 surveys were conducted online, among which, the first 3 surveys employed 2 data collection platforms in China, namely the TClab platform and NAODAO platform. The last survey involved 16,038 participants from Hangzhou Normal University and Zhejiang University of Water Resources and Electric Power, resulting in a total of 17,266 Chinese participants (Table 1; Multimedia Appendix 1). To clarify, participants were recruited using convenience sampling, and all data collection was conducted virtually.

Table 1. Demographic information of the participants for 4 tests.

Version	Additional scales	Valid sample, N	Age (years), mean (SD)	Age (years), minimum	Age (years), maximum	Female percentage
Pretest	—[a]	457	23.67 (6.84)	16	57	39%
SVDS[b] 1.0-3.0	CIAS[c] and DSM-5[d]	402	24.94 (7.21)	17	45	44%
SVDS 4.0	CIAS and DSM-5	369	27.05 (8.18)	17	63	46%
SVDS 4.0	DSM-5	16,038	19.86 (2.14)	15	41	48%

[a]Not applicable.

[b]SVDS: short-video dependence scale.

[c]CIAS: Chinese Internet Addiction Scale.

[d]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

Methodological Overview of the Entire Study

Figure 1 shows the complete process of developing the SVDS. There were several stages. First, in the preparation stage, we screened out people with a high degree of SVD through the questionnaire, obtained the dependence symptoms of these people through interviews, and then combined the findings with the DSM-5 to prepare the first version of the SVDS. Second, SVDS 2.0 was created after internal consistency was examined and item analysis was conducted. Third, through exploratory factor analysis (EFA), we removed the unqualified questions and obtained the third version of the SVDS. Fourth, after confirmatory factor analysis (CFA) and another round of screening, the fourth version was established. Finally, we tested the convergent validity, conducted another round of CFA with a massive sample, and confirmed the cutoff points using receiver operating characteristic (ROC) curve analysis and LPA.

Figure 1. Methodological overview of the entire study. CFA: confirmatory factor analysis; CIAS: Chinese Internet Addiction Scale; DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; EFA: exploratory factor analysis; LPA: latent profile analysis; ROC: receiver operating characteristic; SVDS: short-video dependence scale.

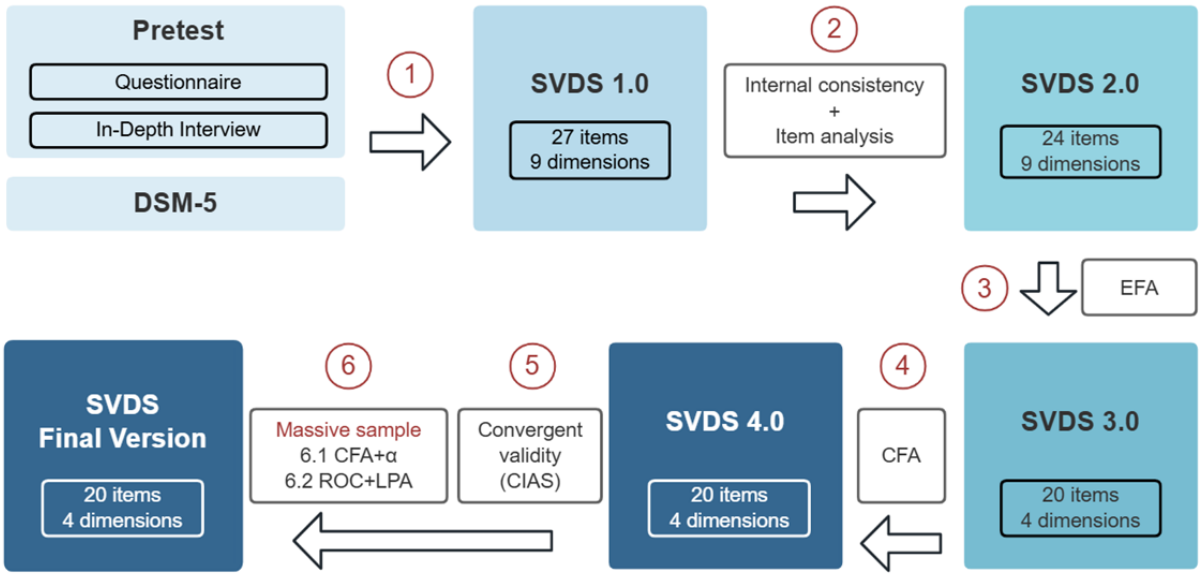

XSL•FO
RenderX

In-Depth Interviews

The in-depth interview method is a fundamental psychological research approach that involves face-to-face conversations between interviewers and participants to understand the psychological and behavioral aspects of the participants [18]. The method is applied in various disciplines; in-depth interviews are also conducted in the form of online conversations [19,20]. As a qualitative research method, in-depth interviews are effective in promptly identifying issues and complementing the questionnaire survey method. In this study, the in-depth interview method was used in conjunction with the questionnaire survey. The process of data collection specifically involved preliminary screening of the pretest questionnaire. The screening criteria were as follows: (1) daily viewing of more than 4 hours, (2) significant impact of short-video viewing on daily life, and (3) voluntary acceptance of the interview and recording. Consequently, semistructured interviews were conducted with 15 short-video users of different ages, users from different regions, and users with different usage levels (Multimedia Appendix 2). The interview content was analyzed using the qualitative analysis software NVivo 12 (Lumivero). The key concepts were derived from the interviews [19], and referring to the 9 criteria of the DSM-5, the initial version of the SVDS with 27 items was developed.

The in-depth interview questions in this study were divided into 5 categories: basic questions, short-video platform usage and dependence, in-depth questions, suggestions and countermeasures, and self-assessment questions with ratings (Multimedia Appendix 3). The main aim was to understand the participants' motivations for using short-video platforms and their levels of dependence and to gain a deeper understanding of the specific symptoms of strong dependence. During the interviews, the interviewer focused on the purpose, motivations, experiences, and dependence of short-video platforms in each round of the interviews. Subsequently, the interview content was coded to determine the dimensions and specific items for the first version of the SVDS. In terms of dimensions, small-scale screening was conducted based on the 9 addiction features of the DSM-5, resulting in the identification of 9 dimensions and 27 items for the SVDS. Moreover, to clarify, we used random sampling to recruit participants. The participants were collected through the NAODAO and TClab platforms.

Encoding In-Depth Interview Data According to the Classical Grounded Theory

The grounded theory establishes a close connection between empirical research and theory construction, enabling researchers to develop concepts and construct theories by systematically analyzing and summarizing empirical data [21,22]. The use of in-depth interviews in this study generates a substantial amount of textual data. By applying the grounded theory, specifically open coding, and aligning it with the DSM-5 scale, key conceptual keywords are abstracted, allowing for a direct reflection of the key points covered in the interview content. This process aids in determining the main dimensions upon which the questionnaire design should be based. The sample saturation point was achieved after the completion of the 15th in-depth interview for this study.

Establishment of the First Version of the SVDS

The first version of the SVDS consisted of 27 items (Multimedia Appendix 4). As for the dimensions, the IGD scale from the DSM-5 consisted of 9 dimensions: preoccupation, tolerance, withdrawal, persistence, escape, problems, deception, displacement, and conflict. These 9 dimensions encompassed most of the situations reported by users with a high dependency on short videos, so we used them as a basis for expansion. Meanwhile, items from multiple related scales were collected for the SVDS question bank for further use [7,23-31]. Additionally, we modified the 9 dimensions by considering the actual situation of users who were highly dependent on short videos in the initial screening. Specifically, we counted the frequency of each dimension mentioned in the narratives of highly dependent users of short videos and removed dimensions that were mentioned zero times. Furthermore, a new dimension, "lie detection," was added. As for the items, we developed 3 items for each dimension based on the information coded in the in-depth interviews, which resulted in the first version of the SVDS.

Developing the Second Version of the SVDS After Conducting Item Analysis

Prior to the EFA, we conducted item analysis [32,33] in sample 2 (n=402) and sample 3 (n=369), which is the process of assessing the quality and effectiveness of each item in the scale. Item analysis mainly includes indicators such as item difficulty, item differentiation, and item relevance. Initially, we removed unqualified samples, which were first preprocessed according to the following two steps: (1) removing samples with missing values and manually removing samples where the length of time was shorter than 2 SDs from the mean and all answers were the same and (2) converting the reversed items to numerical values, recording and scoring them accordingly, and assigning new values to the original items. For further information, sample 2 and sample 3 were also recruited using random sampling. They were primarily gathered through the NAODAO and TClab platforms.

We then screened the items according to 2 indicators during item analysis. The first indicator was the correlation between each item's score and the total score. Based on the Pearson correlation coefficient, the deletion criteria were as follows: (1) nonsignificant correlation between the items and the total scale (P>.05) and (2) correlation coefficient (r) between the item score and the total score below 0.45. The other indicator was the reliability coefficient, and the criteria for deletion were as follows: (1) correlation coefficient between the modified item and the total score lower than 0.45 and (2) increase in the internal consistency coefficient (Cronbach α) after item deletion. This resulted in the second version of the SVDS (Multimedia Appendix 5).

Developing the Third Version of the SVDS After Conducting EFA

The suitability of the items of the first version of the SVDS for factor analysis was assessed using the Kaiser-Meyer-Olkin

XSL•FO
RenderX

(KMO) value and Bartlett Test of Sphericity. Additionally, principal component analysis (PCA) with varimax rotation was conducted on the 21 items of the first version of the SVDS. Items that did not meet the established standards and theoretical expectations were removed. The deletion criteria were as follows: (1) measures of sampling adequacy (MSA) below 0.8, (2) typical values below 0.3, (3) factor loading lower than 0.45, (4) items appearing in two or more factors simultaneously, and (5) factors containing only 1-2 items.

The remaining items were subjected to EFA to obtain the new dimensions of the SVDS based on the factor eigenvalues. This resulted in the third version of the SVDS (Multimedia Appendix 6).

Developing the Fourth Version of the SVDS After Conducting CFA

CFA was conducted to assess the goodness of fit between the proposed model and the collected data [34,35]. Based on the factor solution obtained from the EFA, a model diagram was constructed. CFA was then performed using AMOS24, and model fit was evaluated using the comparative fit index (CFI), Tucker-Lewis Index (TLI), normed fit index (NFI), iterative fit index (IFI), root mean square error of approximation (RMSEA), and 90% CI. When dealing with large samples, the CFI and RMSEA are considered informative fit criteria in structural equation modeling [36]. A good fit is indicated when the CFI, TLI, NFI, and IFI are all greater than 0.9 and the RMSEA is less than 0.08 [34,37].

Convergent Validity and Reliability of the Fourth Version of the SVDS

Convergent validity was determined by examining the correlations between the SVDS scores and the frequency of daily short-video platform use, as well as the DSM-5 and Chinese Internet Addiction Scale (CIAS) scores [24,38,39]. The CIAS is a 19-item self-report scale, with each item rated on a scale of 1-4 (1=completely disagree, 2=somewhat disagree, 3=somewhat agree, and 4=completely agree). The total score ranges from 0 to 76. Participants with a score of 53 or above were classified as having internet addiction. The CIAS has been shown to have satisfactory reliability and validity in the Chinese population [39,40]. In the case of good convergent validity, we expected a strong correlation between SVDS and CIAS scores. Additionally, the internal consistency (reliability) of the scales was assessed using Cronbach α in the 2 sample datasets.

Use of Large Samples to Confirm the Reliability and Validity of the Fourth Version

To assess the final version of the SVDS, we applied it to 16,038 subjects, conducted a CFA to test the validity, and tested the reliability using Cronbach α. The test criteria for the indicators were the same as above. For further clarification, participants in sample 4 were drawn from all first-year students at a college in Hangzhou.

Establishment of Cutoff Points for Dependence and Nondependence

To determine the cutoff point for the population and identify individuals with the disorder, we used the DSM-5 criteria. The group of dependent short-video users according to the DSM-5 (getting more than 5 points out of 9) was considered the "gold standard." Furthermore, ROC curve analysis was adopted to determine the optimal clinical cutoff point based on the gold standard for clinical diagnosis. The diagnostic efficacy of the Bergen Social Media Addiction Scale was assessed through the area under the ROC curve (AUC). Sensitivity analysis was adopted to determine the optimal clinical cutoff point for the SVDS. The sensitivity (the proportion of true positives), specificity (the proportion of true negatives), 95% CI, positive prediction rate (PPR; the proportion of correctly diagnosed positive cases), and negative prediction rate (NPR; the proportion of correctly diagnosed negative cases) were calculated.

LPA was then conducted in the community sample to identify the groups of adolescents with a higher risk of being addicted to short videos among dependent short-video users in MedCalc Version 22.01 [13,14]. Models with 2-6 latent profiles were estimated based on the scores of the 20 items of the SVDS. The Akaike information criterion (AIC), Bayesian information criterion (BIC), sample size–adjusted Bayesian information criterion (SSABIC), and entropy of each model were examined. Finally, the Lo-Mendell-Rubin adjusted likelihood ratio test (LMRT) was used to determine the best-class solution. Considering the probabilistic nature of the LPA classes, the Wald chi-square test of mean equality for latent class predictors in mixture modeling was used for these comparisons.

Results

Coding of In-Depth Interview Data to Derive the First Version of the SVDS

In the preliminary survey, we conducted in-depth interviews with 15 participants who were identified as high-dependence users of short videos from a pool of 457 participants. Subsequently, the interview data were subjected to open coding, taking into account the 9 characteristics from the DSM-5 criteria (Multimedia Appendix 7). This process resulted in the identification of 9 dimensions and 27 specific items for the first version of the SVDS (Multimedia Appendix 4). Open coding focuses on assigning relevant concepts to the data, enabling researchers to categorize the information more clearly. In this study, to present the interview content of the participants more coherently, all original data were systematically numbered and sorted. The in-depth interview data, serving as the primary source of information, were labeled as M1, M2, F3, M4, M5, M5, M7, F8, M9, M10, M11, M12, M13, F14, and M15 (Multimedia Appendix 2). While comprehensive open coding was performed on the primary data, other materials were used for supplementary reference. By conducting open coding on the interview data of the 15 participants, the dimensions and specific items for the first version of the SVDS were determined. In the high-dependence group of short-video users, the most frequently occurring feature was "obsession," while the dimension of "conflict" did not appear at all. Therefore, we eliminated the "conflict" dimension. Considering that the DSM-5 scale consists of only 9 items while the SVDS includes 24, we deemed it

XSL•FO
RenderX

necessary to add a dimension related to deception (comprising 3 items) to screen for valid questionnaires.

Item Analysis of the First Version of the SVDS

Before conducting the EFA, a preliminary item analysis [32,33] was performed. First, 402 samples were left after excluding the unqualified samples. Next, after excluding the polygraph item, item analysis was carried out on a total of 24 items. Initially, the correlation between each item's score and the total score was calculated prior to item screening. Based on the Pearson correlation coefficient, the analysis revealed that items 4 (r=0.43; P=.23), 15 (r=0.32; P=.50), 16 (r=0.14; P=.33), and 24 (r=−0.19; P=.55) did not meet the established criteria, resulting in their deletion.

Furthermore, the reliability coefficient was used for item selection. The internal consistency α coefficient of the 24 items was found to be .90, indicating satisfactory internal consistency of the scale. The results indicated that the correlation coefficients of the modified items in items 15 (r=0.34; α=.90), 16 (r=0.16; α=.91), and 24 (r=0.33; α=.90) were all below 0.4. Thus, the 3 items that fulfilled the condition for an increase in the internal consistency coefficient (α) were retained and considered deleted. Consequently, combined with the above item analysis methods, 3 items (15, 16, and 24) were deleted and 24 items were retained, and thus, the second version of the SVDS was obtained (Multimedia Appendix 5).

EFA of the Second Version of the SVDS and the Establishment of 4 Dimensions

The suitability of the items of the second version of the SVDS for factor analysis was assessed using the KMO value (KMO=0.94) and Bartlett Test of Sphericity (χ^2_{325}=7692.17; P<.001). Additionally, the polygraph item analysis showed that higher scores on the items indicated participants' failure to answer the questions honestly [41]. For our study, 402 valid sample groups were retained in the second sample and 369 valid sample groups were retained in the third sample, meeting the minimum requirement for the number of observations per variable [42].

PCA with varimax rotation was conducted on the 21 items of the second version of the SVDS. Items that did not meet the established standards and theoretical expectations (item 4) were removed. Specifically, the factor loading of item 4 was 0.38 (P<.001), which was the lowest among all items and was below 0.4.

After deleting 4 items, 3 items were not needed for lie detection. The remaining 20 items were subjected to EFA, resulting in the emergence of 4 factors with eigenvalues greater than 1, explaining a cumulative variance of 62.5% (Multimedia Appendix 8). All items demonstrated commonalities ranging from 0.50 to 0.70 and factor loadings ranging from 0.48 to 0.78. Factor 1, named "withdrawal and tolerance," consisted of 5 items (5, 3, 2, 6, and 1), which referred to symptoms of maladaptation following cessation of short-video watching and reduced effects due to repeated or prolonged use. Factor 2, named "preoccupation and persistence," comprised 5 items (8, 7, 10, 12, and 11), reflecting persistent thoughts and behaviors related to short-video use. Factor 3, named "damage and hindrance," encompassed 5 items (9, 20, 21, 18, and 17), indicating the negative impact on physical and mental health, which hindered personal growth and development. Factor 4, named "deception and escape," consisted of 5 items (13, 24, 22, 14, and 23), signifying the concealment of short-video use from others and using short videos as a means of avoiding problems. Consequently, the third version of the SVDS comprised 20 retained items (Multimedia Appendix 6).

CFA and the Internal Consistency of the Third Version of the SVDS

CFA was conducted to assess the goodness of fit between the proposed model and the collected data, and model fit indices were employed. In the evaluation of the goodness of fit, CFI, NFI, IFI, and TLI values should exceed 0.90, while RMSEA should be less than 0.08 for a satisfactory model fit [43]. The results of the goodness-of-fit analysis were as follows: χ^2/df=2.366; RMSEA=0.058; CFI=0.943; NFI=0.906; IFI=0.943; and TLI=0.921. All the data met the acceptable fit index criteria. Moreover, the internal consistency α coefficient of the 20 items was found to be .91. Therefore, the fourth version of the SVDS was identical to the third version (Multimedia Appendix 9).

Convergent Validity of the Final (Fourth) Version of the SVDS

To assess the convergent validity, the scores of the respondents in the first version and the final version of the SVDS were correlated with their self-declared SVD and related constructs, such as the frequency of daily short-video platform use, DSM-5 scores, and CIAS scores. The DSM-5 scale refers to the Short Video Addiction Scale, which is a simple adaptation of the Internet Addiction Scale. As shown in Table 2, all correlations were significant at P<.001 and aligned with the expected directions. Both the first version of the SVDS and the final version of the SVDS demonstrated large positive correlations with self-declared SVD (r>0.5), indicating satisfactory convergent validity.

XSL•FO
RenderX

Table 2. Correlations between the short-video dependence scale and validation constructs.

Variable	SVDS[a] 1.0 (n=402)	SVDS 2.0 (n=369)
Self-declared SVD[b]	0.62	0.58
Frequency of daily use	0.72	0.80
DSM-5[c]	0.68	0.73
CIAS[d]	0.53	0.61

[a]SVDS: short-video dependence scale.

[b]SVD: short-video dependence.

[c]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[d]CIAS: Chinese Internet Addiction Scale.

CFA and Internal Consistency of the Final Version of the SVDS Using a Massive Sample

Similar to the former round of CFA conducted on the third version of the SVDS, a model diagram was constructed using the 20 items and 4 factors. Model fit indices, including RMSEA, CFI, NFI, IFI, and TLI, were employed. Fortunately, the results obtained with the large sample were better than the results of the last CFA. Specifically, the results of the goodness-of-fit analysis were as follows: RMSEA=0.065; CFI=0.928; NFI=0.926; IFI=0.928; and TLI=0.916. All the data met the acceptable fit index criteria. Moreover, the internal consistency α coefficient of the 20 items was .93. The final version of the SVDS is presented in Table 3. The table shows the results after EFA and CFA, with the items rearranged according to the size of factor loading. Factor loading less than 0.4 has not been shown.

Table 3. The final version of the short-video dependence scale.

Index	Item	Dimension
1	Even if advised to watch fewer short videos, I find it difficult to do so.	Preoccupation and persistence
2	In my spare time, I don’t know what to do other than watching short videos.	Preoccupation and persistence
3	If I go without watching short videos for a long time, I fear missing out on popular videos or news.	Preoccupation and persistence
4	Even if I didn’t plan to watch short videos, I would subconsciously open the app.	Preoccupation and persistence
5	While studying or working, I often think about watching short videos.	Preoccupation and persistence
6	I have tried to spend less time watching short videos, but I can’t seem to do it.	Withdrawal and tolerance
7	Compared to last year, I spend more time watching short videos every day.	Withdrawal and tolerance
8	When I try to watch fewer videos, I feel bored or agitated.	Withdrawal and tolerance
9	My life seems uninteresting without short videos.	Withdrawal and tolerance
10	If asked to refrain from watching short videos for a week, I would find it difficult to resist the urge.	Withdrawal and tolerance
11	Watching short videos has had a negative impact on my academic or work performance.	Damage and hindrance
12	It has also negatively affected my physical health, such as eye strain and staying up late.	Damage and hindrance
13	I feel a decreased sense of self-worth due to my inability to control my time spent watching short videos.	Damage and hindrance
14	After watching short videos, I find it harder to focus on self-improvement.	Damage and hindrance
15	Even if I want to go to bed early, I still can’t resist watching short videos.	Damage and hindrance
16	When I watch short videos, it seems as if all of life’s problems disappear.	Deception and escape
17	I have tried to hide the negative effects that watching short videos has had on me from others.	Deception and escape
18	I tend to downplay my short-video viewing time when asked.	Deception and escape
19	After watching short videos, I tend to interact less with family and friends in real life.	Deception and escape
20	The actual amount of time I spend watching short videos is more than I realize.	Deception and escape

Sensitivity and Specificity of the SVDS

The SVDS dimensions were evaluated for sensitivity and specificity. Sensitivity was determined based on positive responses from dependent short-video users in sample 2 (n=107), sample 3 (n=74), and sample 4 (n=16,038), while specificity relied on negative responses from nondependent short-video users in each sample. High sensitivity and specificity are ideal for effective discrimination of false positives and false negatives [44].

XSL•FO
RenderX

Initially, the first SVDS version demonstrated high sensitivity and sufficient specificity in sample 2. Corrected sensitivity and specificity values for each dimension were as follows: preoccupation, 0.87 and 0.45, respectively; tolerance, 0.96 and 0.44, respectively; withdrawal, 0.66 and 0.89, respectively; persistence, 0.88 and 0.61, respectively; escape, 0.98 and 0.32, respectively; problems, 0.85 and 0.84, respectively; deception, 0.68 and 0.85, respectively; and displacement, 0.58 and 0.93, respectively. In sample 4, the final SVDS version showed high sensitivity and sufficient specificity. Sensitivity and specificity values for each dimension were as follows: withdrawal and tolerance, 0.77 and 0.75, respectively; preoccupation and persistence, 0.83 and 0.87, respectively; impairment and hindrance, 0.87 and 0.80, respectively; and deception and escape, 0.62 and 0.79, respectively. Overall, the final SVDS version demonstrated stable sensitivity and specificity, with highly comparable diagnostic accuracy across the samples.

Determination of the Clinical Cutoff Point for the Final Version of the SVDS

According to the SVDS modified based on the IAS in DSM-5, 4693 of the 16,038 participants were identified as dependent short-video users. This clinically diagnosed group was used as the gold standard. The diagnostic efficiency was demonstrated by a high AUC (0.83, 95% CI 0.95-1.00; P<.001).

Table 4 shows the sensitivity, specificity, 95% CI, PPV, and NPV of possible cutoff scores of the SVDS. At a score of 58, the sensitivity was 88.6% and specificity was 96.5%, and the Youden Index achieved its maximum value (85.2%). In this case, the prevalence of dependent short-video users was 4079 out of 16,038 (25.4%). The dependent short-video user group included 53.6% (2188/4079) males and 46.4% (1891/4079) females, while the nondependent short-video user group included 50.1% (5999/11,959) males and 49.8% (5960/11,959) females.

Table 4. Cutoff points for the short-video dependence scale based on clinical diagnostic interviews (N=16,038).

Criterion	Sensitivity[a], value (95% CI)	Specificity[b], value (95% CI)	Youden Index[c]	PPR[d]	NPR[e]
>54	89.58 (87.4-91.5)	86.50 (86.3-87.8)	0.7608	0.79	1.00
>55	89.47 (87.3-91.4)	89.90 (88.5-90.9)	0.7937	0.79	1.00
>56	89.47 (87.3-91.4)	92.68 (91.8-93.0)	0.8215	0.83	1.00
>57	89.47 (87.3-91.4)	95.43 (94.3-96.4)	0.8490	0.86	1.00
>58	88.61 (86.4-90.6)	96.54 (95.5-96.4)	0.8515	0.88	0.99
>59	87.22 (84.9-89.3)	97.41 (96.6-97.4)	0.8463	0.92	0.99
>60	86.40 (84.6-89.0)	97.75 (97.6-98.3)	0.8415	0.93	0.98
>61	85.07 (84.1-88.6)	98.76 (98.5-99.0)	0.8383	0.94	0.94
>62	83.57 (81.0-85.9)	99.03 (99.3-99.7)	0.8260	0.94	0.89

[a]Sensitivity: true negative/true negative and false positive.

[b]Specificity: true positive/true positive and false negative.

[c]Youden index: defined as sensitivity + specificity – 1.

[d]PPR: positive predictive rate (true positive/true positive and false positive).

[e]NPR: negative predictive rate (true negative/true negative and false negative).

LPA Results

The results of the LPA in the last sample are shown in Table 5. The AIC, BIC, and SSABIC values continued to decrease as the number of groups increased. The entropy was adequate for a 2-group solution to a 6-group solution. Despite the small difference in the data, the entropy of 5 classes was the smallest, indicating the highest degree of classification accuracy [45]; therefore, we selected the 5-group solution. The features of the 5 classes are shown in Figure 2. The first class, named “regular users” (6591/16,038, 41.1%), and the second class, named “casual users” (3224/16,038, 20.1%), represent adolescents who generally selected “very rarely” or “rarely” for all 20 items in the final version of the SVDS. The third class, named “low-risk high-engagement users” (2389/16,038, 14.9%), and the fourth class, named “at-risk high-engagement users” (2630/16,038, 16.4%), scored similarly higher on “damage and hindrance” and “deception and escape,” while the third class scored much lower on “preoccupation and persistence” and “withdrawal and tolerance” compared to the fourth class. The fifth class, named “dependent users” (1204/16,038, 7.5%), represents adolescents who generally rated their short-video use as “very often” or “often” for all 20 items in the scale.

XSL•FO
RenderX

Table 5. Results obtained from the latent profile analysis.

Model	Log-likelihood	Replicated log-likelihood	Free parameters, n	AIC[a]	BIC[b]	SSABIC[c]	Entropy	LMRT[d]
2 classes	–178532.773	Yes	13	357091.546	357191.894	357150.581	0.769	26,961.63
3 classes	–173702.934	Yes	18	347441.869	347580.812	347523.609	0.830	9464.91
4 classes	–171682.983	Yes	23	343411.967	343589.506	343516.413	0.861	3958.44
5 classes	–175612.864	Yes	28	341344.701	341560.835	341471.853	0.872	2035.38
6 classes	–169910.753	Yes	33	339887.506	340142.236	340037.364	0.808	1437.61

[a]AIC: Akaike information criterion.

[b]BIC: Bayesian information criterion.

[c]SSABIC: sample size–adjusted Bayesian information.

[d]LMRT: Lo-Mendell-Rubin adjusted likelihood ratio test.

Figure 2. The 5 classes obtained from the latent profile analysis. The latent profile of short-video dependence based on short-video dependence scale (SVDS) scores. The X-axis represents the dimensions of the SVDS. The Y-axis represents the scores of each dimension of the SVDS.

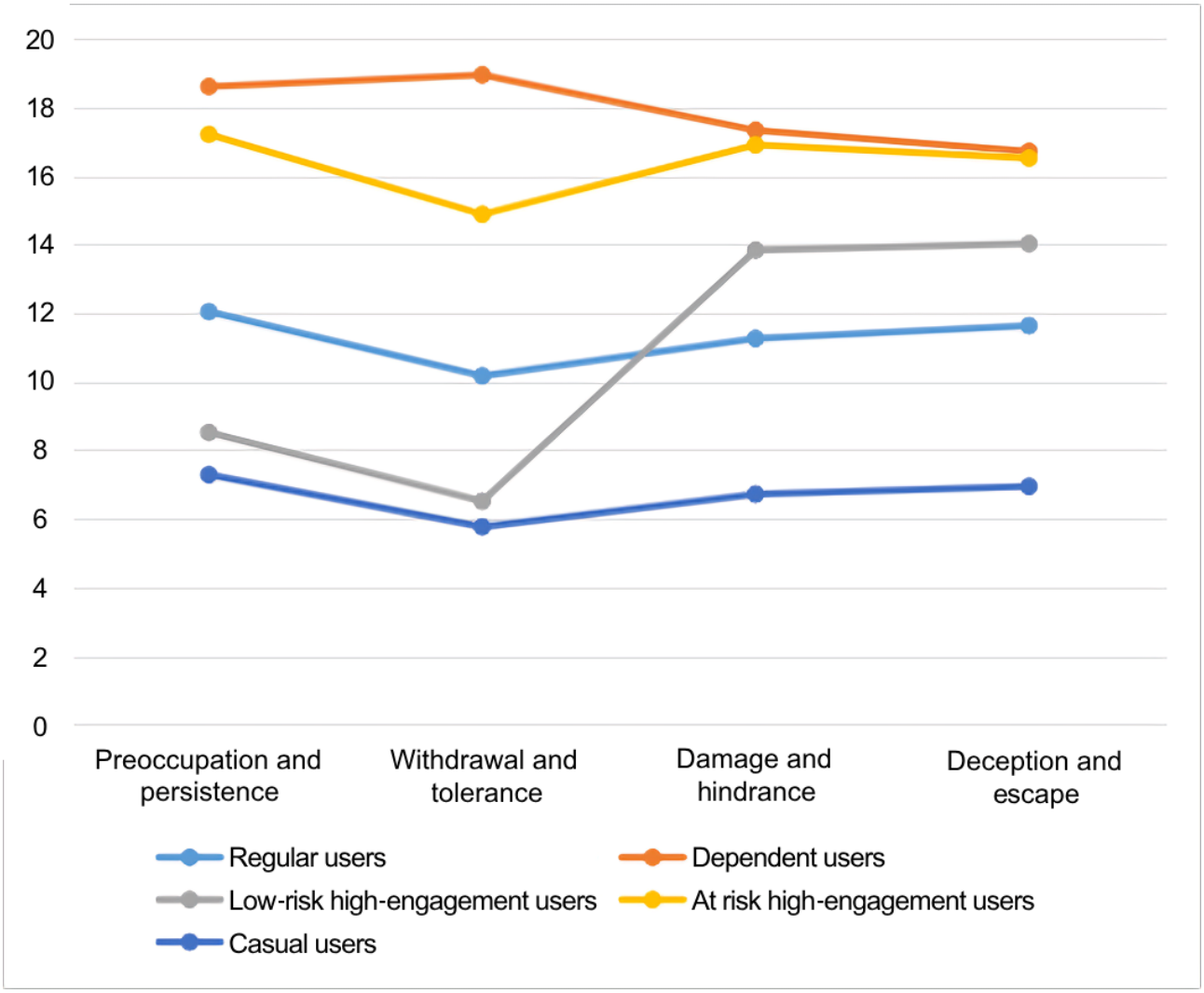

Discussion

Principal Findings

This study aimed to develop and validate the SVDS, which is brief, is easy to administer, and has a clear diagnostic cutoff point to differentiate between dependent and highly engaged nondependent short-video users. Through item analysis, EFA, and CFA, the 20-item SVDS with 4 dimensions was established. To establish the diagnostic cutoff point, ROC curve analysis and LPA were conducted. Consequently, the final version of the SVDS with a clear cutoff point demonstrated strong internal consistency and convergent validity, and high sensitivity and specificity.

XSL•FO
RenderX

Development Processes of the SVDS

We initially identified subjects with higher SVD through pretesting and in-depth interviews, extracting key dimensions and items from the raw data. We also used the 9 diagnostic criteria for IGD from the DSM-5 as a reference, which contributed to both the theoretical and conceptual development of the tool. This process led to the identification of 9 dimensions for the first version of the SVDS, each with 3 items. After item analysis, EFA, and CFA, a 20-item scale with 4 dimensions was finalized, resulting in 3 versions of the SVDS. Finally, using a large sample of 16,038 participants, the final version showed strong internal consistency and convergent validity.

Final Version of the SVDS

The final (fourth) version of the SVDS included 4 dimensions: "withdrawal and tolerance," "preoccupation and persistence," "impairment and hindrance," and "deception and escape." Each dimension consisted of 5 items, and SVD was measured. Furthermore, we employed LPA and ROC curve analysis with a large sample of 16,038 people to detect dependent short-video users and define the cutoff point of 58 points. Five more subtypes of users of short videos were retrieved from the LPA.

Final Version of the SVDS Showed Good Reliability and Validity

Our findings demonstrated that the final version of the SVDS is a reliable and valid measure of SVD. The CFA confirmed a good model fit, indicating strong external validity. Moreover, the final version of the SVDS exhibited satisfactory internal consistency (0.93) as well as good convergent validity. These results have important implications for identifying and preventing SVD and its potential negative consequences. Additionally, based on the diagnostic cutoff point of the final version of the SVDS, the prevalence of SVD in the last sample was 25.4% (4073/16,038).

Convergent validity was established by examining the CFA results of the SVDS and its correlations with similar constructs. The CFA results supported a good model fit for the final version of SVDS items, and as expected, there was a strong correlation between SVDS scores and CIAS scores.

Final Version of the SVDS Showed Good Sensitivity and Specificity

Items of the final version of the SVDS exhibited good sensitivity and specificity. This indicates that the scale can accurately identify and differentiate between the presence and absence of social media dependence. Sensitivity refers to situations in which items correctly identify the target characteristic, while specificity refers to situations in which items correctly exclude the absence of the target characteristic. From the first version to the final version of the scale, there were improvements in its sensitivity and specificity, as outlined below.

Confirmation of the Scale's Dimensions

"Withdrawal and tolerance," "preoccupation and persistence," "impairment and hindrance," and "deception and escape" were the 4 dimensions included in the SVDS.

First, withdrawal refers to discomfort when a substance or behavior is reduced or stopped, while tolerance means needing more to achieve the same effect. These are key addiction concepts, so individuals with high scores are closer to an addictive state and more likely to be "at-risk high-engagement users" or "dependent users." Second, "preoccupation and persistence" involve constant thoughts about the activity and continuing despite harm, with higher scores reflecting stronger dependence. Third, "impairment and hindrance" refer to the negative impact on daily life and the reduced ability to resist addictive behavior. High scores suggest significant life problems and difficulty controlling behavior. Lastly, "deception and escape" involve lying or replacing one addictive behavior with another, with elevated scores indicating stronger dependence. The first 2 dimensions are core symptoms of SVD, as they directly impact the individual, while the latter 2 reflect the indirect effects of SVD on daily life.

Establishment of the Diagnostic Cutoff Point

Using DSM-5–diagnosed dependent short-video users as the gold standard, this study suggests a clinically optimal cutoff score of 58 for diagnosing dependent short-video users with the final version of the SVDS. The best balance between sensitivity (88.6%) and specificity (96.5%) was achieved at this cutoff score. The LPA approach has been performed to identify cutoff points in the absence of a gold standard based on clinical interviews. The adequate diagnostic accuracy of the empirical cutoff point based on the LPA suggests that this approach is a reliable method to identify cutoff points.

We compared the SVDS scores of these dependent short-video users and established a SVDS cutoff point of 58, whereby individuals scoring equal to or higher than 58 were considered to be dependent short-video users.

Comparison With Prior Work

Advantages Compared to DSM-5 Items for Behavioral Addictions

The symptoms of SVD are more closely aligned with the SVDS than with the DSM-5. Some of the parameters of the original SVDS indicated weak sensitivity or specificity because they were largely the same as those of the DSM-5. During the SVDS development process, we were able to improve the scale.

For the first version of the SVDS, the dimensions were identical to those of the DSM-5. Additionally, some items showed lower sensitivity and specificity. To start with, the items measuring displacement showed a lower sensitivity in comparison to the other items of the SVDS and the displacement item of the IGD scale [46]. This suggests that a few high short-video dependents lost interest in their previous hobbies and entertainment. This may be because short-video use can coexist with other activities to some extent. Existing studies in the literature have focused on internet addiction and media use, finding that people who are highly dependent on the internet may reduce offline activities and social interactions but do not necessarily lose interest in previous hobbies altogether [6,9,47]. This implies that people who are highly dependent on short videos may strike a certain balance between online and offline activities, rather than completely abandoning their previous hobbies.

XSL•FO
RenderX

Next, the items measuring escape, preoccupation, and tolerance showed a lower specificity in comparison to the other items. This suggests that most short-video users often use short videos to escape from negative feelings, irrespective of whether they are dependent or nondependent short-video users. This may be because short videos are often light-hearted and entertaining in nature, providing a pleasurable and entertaining experience. For low-dependent individuals, watching short videos may be a relaxing and enjoyable way to spend time, rather than an addictive behavior. Short videos also provide a virtual world where people can temporarily escape from real-life problems, stresses, and challenges. Watching short videos can be an escape mechanism that helps people temporarily avoid unpleasant or difficult situations.

Advantages Compared to Other Similar Questionnaires and Scales

We compared the SVDS to related scales (ie, measurement tools that include internet, dependence, addiction, or other similar keywords), such as the CIAS and the Social Media Disorder Scale (SMDS).

The CIAS is a scale used to assess the core symptoms and problems associated with internet addiction in the Chinese population [24,39]. It does not consider the different types or forms of internet addiction, such as social media addiction or short-form video addiction. Furthermore, it appears inappropriate to use a scale created nearly 20 years ago to gauge how dependent modern society is on short films. Therefore, the SVDS with reference to the content of in-depth interviews conducted in 2023 is more relevant to SVD and can provide more accurate results. Most importantly, we provided statistical evidence for a precise diagnostic cutoff point.

The SMDS is a scale used to assess users' motivation to use social media and the degree of addiction [7]. It does not consider users' communication and interaction with other people or groups on social media platforms, such as liking, commenting, etc, and the psychological and social impacts of such communication and interaction on users. Besides, the scale has only 9 dimensions, which may not cover all aspects of social media addiction, such as motivation, frequency, and duration of social media use. As a result, the SVDS, which has more detailed questions and has questions that are pertinent to the features of short-video platforms, is better suited for measuring SVD.

Limitations and Future Directions

While this study provides distinctive insights into the detection of dependent short-video users and the prevalence of SVD among Chinese university students, it is imperative to acknowledge the presence of certain limitations in the research. First, there may be sampling bias as this study recruited subjects on the NAODAO and TClab platforms and focused on universities in one city. As a significant majority of the participants in our sample had at least a college-level education, there was an imbalance concerning age and educational background. Despite previous studies demonstrating measurement invariance of some scales across education and gender [48-50], the imbalance in the number of participants of different genders or education levels may still have affected our results. To further test the SVDS, future research should collect a well-balanced sample. Additionally, since our sample was primarily drawn from Chinese university students, specifically those in Hangzhou, the findings may have limited generalizability to other cultural or demographic contexts. Future studies are needed to validate the scale in diverse cultural settings and populations. Second, according to recent reports from social surveys, a sizable number of short-video consumers are middle-aged and older [51,52]. The SVDS should be administered to this demographic in subsequent investigations. Third, the self-completion scale might have a social desirability effect. Future studies could cross-validate our results using other research methods (eg, field studies). Fourth, future research should focus on examining the relationship between the degree of short-video dependency and individual characteristics such as personality traits, self-esteem, impulsivity, academic performance or work performance, and other relevant factors. Finally, it is important to examine whether there are differences in SVD among different groups, which could help validate the reliability and validity of the current scale.

Conclusions

We followed the pragmatic approach of developing and validating an administration tool to measure SVD, which enables the investigation of trends and developments in the prevalence of SVD during the current period of rapid changes in the short-video landscape. The final version of the SVDS showed high reliability and validity. It could be applied by subsequent researchers in longitudinal studies to investigate the mechanisms of SVD in terms of behavioral, cognitive, and affective influences, as well as to develop effective interventions to deal with the problem of SVD. The development and validation of a questionnaire with a diagnostic cutoff point are essential for understanding and monitoring SVD, facilitating relevant academic investigations, and informing policy development.

Acknowledgments

This research was supported by the Cultivation Project of Province-leveled Preponderant Characteristic Discipline of Hangzhou Normal University (20JYXK008) and Zhejiang Provincial Natural Science Foundation (LY20C090005). The funding agencies did not contribute to the experimental design or conclusions, and the views presented in the manuscript are those of the authors and may not reflect those of the funding agencies.

XSL•FO
RenderX

Data Availability

The data are stored at our lab-based network attachment system [53] (ID: guests; PIN: dong@123.COM).

Authors' Contributions

AJ contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing-original draft, and writing-review and editing. HN contributed to conceptualization and investigation. SL, HC, JD, XX, and HW assisted in investigation and data curation. ML provided resources. GD contributed to funding acquisition, conceptualization, supervision, and writing-review and editing. All authors contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of participants.
[DOCX File , 14 KB - jmir_v27i1e66341_app1.docx]

Multimedia Appendix 2
Details of the interviewees.
[DOCX File , 18 KB - jmir_v27i1e66341_app2.docx]

Multimedia Appendix 3
In-depth interview questionnaire.
[DOCX File , 14 KB - jmir_v27i1e66341_app3.docx]

Multimedia Appendix 4
Short-video dependence scale 1.0.
[DOCX File , 19 KB - jmir_v27i1e66341_app4.docx]

Multimedia Appendix 5
Short-video dependence scale 2.0.
[DOCX File , 18 KB - jmir_v27i1e66341_app5.docx]

Multimedia Appendix 6
Short-video dependence scale 3.0.
[DOCX File , 17 KB - jmir_v27i1e66341_app6.docx]

Multimedia Appendix 7
Coding interview.
[DOCX File , 17 KB - jmir_v27i1e66341_app7.docx]

Multimedia Appendix 8
Factor analysis of short-video dependence scale 2.0.
[DOCX File , 18 KB - jmir_v27i1e66341_app8.docx]

Multimedia Appendix 9
Short-video dependence scale 4.0.
[DOCX File , 18 KB - jmir_v27i1e66341_app9.docx]

References

1. Zhang X, Wu Y, Liu S. Exploring short-form video application addiction: Socio-technical and attachment perspectives. Telematics and Informatics 2019 Sep;42:101243 [FREE Full text] [doi: 10.1016/j.tele.2019.101243]
2. Agrawal E. Going Viral: An Analysis of Advertising of Technology Products on TikTok. arXiv. URL: https://arxiv.org/abs/2402.00010 [accessed 2025-02-05]
3. Hanaysha JR. Impact of social media marketing features on consumer's purchase decision in the fast-food industry: Brand trust as a mediator. International Journal of Information Management Data Insights 2022 Nov;2(2):100102. [doi: 10.1016/j.jjimei.2022.100102]

XSL•FO
RenderX

4. Qin Y, Omar B, Musetti A. The addiction behavior of short-form video app TikTok: The information quality and system quality perspective. Front Psychol 2022;13:932805 [FREE Full text] [doi: 10.3389/fpsyg.2022.932805] [Medline: 36148123]
5. Ye J, Wu Y, Wu Y, Chen M, Ye J. Effects of short video addiction on the motivation and well-being of Chinese vocational college students. Front Public Health 2022 May 10;10:847672 [FREE Full text] [doi: 10.3389/fpubh.2022.847672] [Medline: 35619803]
6. Domoff SE, Harrison K, Gearhardt AN, Gentile DA, Lumeng JC, Miller AL. Development and validation of the problematic media use measure: a parent report measure of screen media "addiction" in children. Psychol Pop Media Cult 2019 Jan;8(1):2-11 [FREE Full text] [doi: 10.1037/ppm0000163] [Medline: 30873299]
7. van den Eijnden R, Lemmens J, Valkenburg P. The social media disorder scale. Computers in Human Behavior 2016 Aug;61:478-487 [FREE Full text] [doi: 10.1016/j.chb.2016.03.038]
8. Griffiths M. Internet use disorders: What's new and what's not? J Behav Addict 2020 Oct 03;9(4):934-937 [FREE Full text] [doi: 10.1556/2006.2020.00072] [Medline: 33011713]
9. Kovačić Petrović Z, Peraica T, Blažev M, Kozarić-Kovačić D. Association between problematic Internet use and specific Internet activities and COVID-19- and earthquake-related stress, anxiety, and depression symptoms among Croatian young adults. Front Psychiatry 2023;14:1227182 [FREE Full text] [doi: 10.3389/fpsyt.2023.1227182] [Medline: 37645636]
10. Güzeller C, Coşguner T. Development of a problematic mobile phone use scale for Turkish adolescents. Cyberpsychol Behav Soc Netw 2012 Apr;15(4):205-211 [FREE Full text] [doi: 10.1089/cyber.2011.0210] [Medline: 22304426]
11. Li T, Liu H, Liu X. Interdisciplinary Insights into Short Video Addiction: Integrating Psychological and Brain imaging Approaches in a Comprehensive Review. Preprints. 2024. URL: https://www.preprints.org/manuscript/202405.0824/v1 [accessed 2025-02-05]
12. Yang J, Ti Y, Ye Y. Offline and online social support and short-form video addiction among Chinese adolescents: the mediating role of emotion suppression and relatedness needs. Cyberpsychol Behav Soc Netw 2022 May;25(5):316-322 [FREE Full text] [doi: 10.1089/cyber.2021.0323] [Medline: 35486859]
13. Li J, Wu A, Feng L, Deng Y, Li J, Chen Y, et al. Classification of probable online social networking addiction: A latent profile analysis from a large-scale survey among Chinese adolescents. J Behav Addict 2020 Oct 12;9(3):698-708 [FREE Full text] [doi: 10.1556/2006.2020.00047] [Medline: 32829311]
14. Peng P, Chen Z, Ren S, Liu Y, He R, Liang Y, et al. Determination of the cutoff point for Smartphone Application-Based Addiction Scale for adolescents: a latent profile analysis. BMC Psychiatry 2023 Sep 16;23(1):675 [FREE Full text] [doi: 10.1186/s12888-023-05170-4] [Medline: 37716941]
15. van Smeden M, Naaktgeboren C, Reitsma J, Moons K, de Groot J. Latent class models in diagnostic studies when there is no reference standard--a systematic review. Am J Epidemiol 2014 Feb 15;179(4):423-431 [FREE Full text] [doi: 10.1093/aje/kwt286] [Medline: 24272278]
16. Luo T, Qin L, Cheng L, Wang S, Zhu Z, Xu J, et al. Determination the cut-off point for the Bergen social media addiction (BSMAS): Diagnostic contribution of the six criteria of the components model of addiction for social media disorder. J Behav Addict 2021 Jul 15;10(2):281-290 [FREE Full text] [doi: 10.1556/2006.2021.00025] [Medline: 34010148]
17. Lin J, Su Y, Rizvi S, Jagoda J, Li J, Wu Y, et al. Define and characterize the anhedonia in major depressive disorder: An explorative study. J Affect Disord 2022 Sep 15;313:235-242 [FREE Full text] [doi: 10.1016/j.jad.2022.06.082] [Medline: 35788366]
18. Rutledge P, Hogg J. In-Depth Interviews. In: Bulck J, editor. The International Encyclopedia of Media Psychology. Hoboken, NJ: Wiley; 2020.
19. Deterding N, Waters M. Flexible coding of in-depth interviews: a twenty-first-century approach. Sociological Methods & Research 2018 Oct 01;50(2):708-739 [FREE Full text] [doi: 10.1177/0049124118799377]
20. Rutakumwa R, Mugisha J, Bernays S, Kabunga E, Tumwekwase G, Mbonye M, et al. Conducting in-depth interviews with and without voice recorders: a comparative analysis. Qual Res 2020 Oct;20(5):565-581 [FREE Full text] [doi: 10.1177/1468794119884806] [Medline: 32903872]
21. Charmaz K, Thornberg R. The pursuit of quality in grounded theory. Qualitative Research in Psychology 2020 Jun 22;18(3):305-327 [FREE Full text] [doi: 10.1080/14780887.2020.1780357]
22. Fassinger R. Paradigms, praxis, problems, and promise: Grounded theory in counseling psychology research. Journal of Counseling Psychology 2005 Apr;52(2):156-166 [FREE Full text] [doi: 10.1037/0022-0167.52.2.156]
23. Blandón-Castaño J, Puerta-Cortés D, Carmona A, Ivern J, Vilca L, Carbonell X, et al. Cross-cultural invariance of the Internet Gaming Disorder Scale-Short-Form (IGDS9-SF) in seven Latin American countries. Int J Ment Health Addiction 2024 Sep 26:01376 [FREE Full text] [doi: 10.1007/s11469-024-01376-z]
24. Chinese Internet Addiction Scale--Revised. APA PsycTests. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft44491-000 [accessed 2025-02-05]
25. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. Journal of Substance Use 2009 Jul 12;10(4):191-197 [FREE Full text] [doi: 10.1080/14659890500114359]
26. Kırcaburun K. Effects of gender and personality differences on twitter addiction among Turkish undergraduates. Journal of Education and Practice 2016;7(24):33-42 [FREE Full text]

XSL•FO
RenderX

27. Leung L. Linking psychological attributes to addiction and improper use of the mobile phone among adolescents in Hong Kong. Journal of Children and Media 2008 Jul;2(2):93-113 [FREE Full text] [doi: 10.1080/17482790802078565]
28. Shahnawaz M, Rehman U. Social networking addiction scale. Cogent Psychology 2020 Oct 19;7(1):1832032 [FREE Full text] [doi: 10.1080/23311908.2020.1832032]
29. Tejeiro Salguero R, Morán R. Measuring problem video game playing in adolescents. Addiction 2002 Dec;97(12):1601-1606 [FREE Full text] [doi: 10.1046/j.1360-0443.2002.00218.x] [Medline: 12472644]
30. Toda M, Monden K, Kubo K, Morimoto K. Mobile phone dependence and health-related lifestyle of university students. Social Behavior and Personality 2006 Jan 01;34(10):1277-1284 [FREE Full text] [doi: 10.2224/sbp.2006.34.10.1277]
31. Young K. Internet addiction: the emergence of a new clinical disorder. CyberPsychology & Behavior 1998 Jan;1(3):237-244 [FREE Full text] [doi: 10.1089/cpb.1998.1.237]
32. McGahee T, Ball J. How to read and really use an item analysis. Nurse Educ 2009;34(4):166-171 [FREE Full text] [doi: 10.1097/NNE.0b013e3181aaba94] [Medline: 19574855]
33. Sürücü L, Yikilmaz ?, Maslakçi A. Exploratory factor analysis (EFA) in quantitative researches and practical considerations. Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi 2024:947-965 [FREE Full text] [doi: 10.37989/gumussagbil.1183271]
34. Hox JJ. Confirmatory Factor Analysis. In: Barnes JC, Forde DR, editors. The Encyclopedia of Research Methods in Criminology and Criminal Justice. Hoboken, NJ: Wiley; 2021.
35. Hoyle RH. Confirmatory Factor Analysis. In: Tinsley H, Brown S, editors. Handbook of Applied Multivariate Statistics and Mathematical Modeling Internet. Cambridge, MA: Academic Press; 2000:465-497.
36. Montoya A, Edwards M. The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. Educ Psychol Meas 2021 Jun;81(3):413-440 [FREE Full text] [doi: 10.1177/0013164420942899] [Medline: 33994558]
37. Goretzko D, Siemund K, Sterner P. Evaluating model fit of measurement models in confirmatory factor analysis. Educ Psychol Meas 2024 Feb;84(1):123-144 [FREE Full text] [doi: 10.1177/00131644231163813] [Medline: 38250508]
38. Mak K, Lai C, Watanabe H, Kim D, Bahar N, Ramos M, et al. Epidemiology of internet behaviors and addiction among adolescents in six Asian countries. Cyberpsychol Behav Soc Netw 2014 Nov;17(11):720-728 [FREE Full text] [doi: 10.1089/cyber.2014.0139] [Medline: 25405785]
39. Shen Y, Wang L, Huang C, Guo J, De Leon S, Lu J, et al. Sex differences in prevalence, risk factors and clinical correlates of internet addiction among chinese college students. J Affect Disord 2021 Jan 15;279:680-686 [FREE Full text] [doi: 10.1016/j.jad.2020.10.054] [Medline: 33190119]
40. Bai C, Chen X, Han K. Mobile phone addiction and school performance among Chinese adolescents from low-income families: A moderated mediation model. Children and Youth Services Review 2020 Nov;118:105406 [FREE Full text] [doi: 10.1016/j.childyouth.2020.105406]
41. Raskin D, Kircher J. Validity of Polygraph Techniques and Decision Methods. In: Raskin D, Honts C, Kircher J, editors. Credibility Assessment. Cambridge, MA: Academic Press; 2014:63-129.
42. Stevens J, Stevens J. Applied Multivariate Statistics for the Social Sciences. New York, NY: Psychology Press; 2002.
43. Byrne B. Structural Equation Modeling With AMOS. New York, NY: Routledge; 2015.
44. Glaros A, Kline R. Understanding the accuracy of tests with cutting scores: the sensitivity, specificity, and predictive value model. J Clin Psychol 1988 Nov;44(6):1013-1023 [FREE Full text] [doi: 10.1002/1097-4679(198811)44:6<1013::aid-jclp2270440627>3.0.co;2-z] [Medline: 3216006]
45. Spurk D, Hirschi A, Wang M, Valero D, Kauffeld S. Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. Journal of Vocational Behavior 2020 Aug;120:103445 [FREE Full text] [doi: 10.1016/j.jvb.2020.103445]
46. Lemmens J, Valkenburg P, Gentile D. The internet gaming disorder scale. Psychol Assess 2015 Jun;27(2):567-582 [FREE Full text] [doi: 10.1037/pas0000062] [Medline: 25558970]
47. Montag C, Yang H, Elhai J. On the psychology of TikTok use: a first glimpse from empirical findings. Front Public Health 2021;9:641673 [FREE Full text] [doi: 10.3389/fpubh.2021.641673] [Medline: 33816425]
48. Andreassen C, Torsheim T, Brunborg G, Pallesen S. Development of a Facebook addiction scale. Psychol Rep 2012 Apr 01;110(2):501-517 [FREE Full text] [doi: 10.2466/02.09.18.pr0.110.2.501-517]
49. Cheung G, Rensvold R. Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal 2002 Apr;9(2):233-255 [FREE Full text] [doi: 10.1207/s15328007sem0902_5]
50. Kim L, Jörg V, Klassen R. A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. Educ Psychol Rev 2019;31(1):163-195 [FREE Full text] [doi: 10.1007/s10648-018-9458-2] [Medline: 30930595]
51. Bertocchi FM, De Oliveira AC, Lucchetti G, Lucchetti ALG. Smartphone use, digital addiction and physical and mental health in community-dwelling older adults: a population-based survey. J Med Syst 2022 Jun 18;46(8):53. [doi: 10.1007/s10916-022-01839-7] [Medline: 35716194]
52. Weber W, Reinhardt A, Rossmann C. Lifestyle segmentation to explain the online health information-seeking behavior of older adults: representative telephone survey. J Med Internet Res 2020 Jun 12;22(6):e15099 [FREE Full text] [doi: 10.2196/15099] [Medline: 32530433]
53. Quick Connect. URL: https://www.quickconnect.cn/ [accessed 2025-02-12]

XSL•FO
RenderX

Abbreviations

AIC: Akaike information criterion
AUC: area under the receiver operating characteristic curve
BIC: Bayesian information criterion
CFA: confirmatory factor analysis
CFI: comparative fit index
CIAS: Chinese Internet Addiction Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EFA: exploratory factor analysis
IFI: iterative fit index
IGD: internet gaming disorder
KMO: Kaiser-Meyer-Olkin
LPA: latent profile analysis
NFI: normed fit index
PCA: principal component analysis
RMSEA: root mean square error of approximation
ROC: receiver operating characteristic
SMDS: Social Media Disorder Scale
SSABIC: sample size–adjusted Bayesian information criterion
SVD: short-video dependence
SVDS: short-video dependence scale
TLI: Tucker-Lewis Index

Edited by T de Azevedo Cardoso; submitted 10.09.24; peer-reviewed by I Celik, X Sun, AC Lusar; comments to author 29.11.24; revised version received 17.12.24; accepted 29.01.25; published 04.03.25.

Please cite as:
Jiang A, Li S, Wang H, Ni H, Chen H, Dai J, Xu X, Li M, Dong GH
Assessing Short-Video Dependence for e-Mental Health: Development and Validation Study of the Short-Video Dependence Scale
J Med Internet Res 2025;27:e66341
URL: https://www.jmir.org/2025/1/e66341
doi:10.2196/66341
PMID:40053762

©AnHang Jiang, Shuang Li, HuaBin Wang, HaoSen Ni, HongAn Chen, JunHong Dai, XueFeng Xu, Mei Li, Guang-Heng Dong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Trust and Acceptance Challenges in the Adoption of AI Applications in Health Care: Quantitative Survey Analysis

Janne Kauttonen[1], PhD; Rebekah Rousi[2,3], PhD; Ari Alamäki[1], PhD

[1]Digital Transition and AI, Haaga-Helia University of Applied Sciences, Helsinki, Finland

[2]Communication Studies, School of Marketing and Communication, University of Vaasa, Vaasa, Finland

[3]Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

Corresponding Author:
Janne Kauttonen, PhD
Digital Transition and AI
Haaga-Helia University of Applied Sciences
Ratapihantie 13
Helsinki, 00520
Finland
Phone: 358 400 230 404
Email: janne.kauttonen@haaga-helia.fi

Abstract

Background: Artificial intelligence (AI) has potential to transform health care, but its successful implementation depends on the trust and acceptance of consumers and patients. Understanding the factors that influence attitudes toward AI is crucial for effective adoption. Despite AI's growing integration into health care, consumer and patient acceptance remains a critical challenge. Research has largely focused on applications or attitudes, lacking a comprehensive analysis of how factors, such as demographics, personality traits, technology attitudes, and AI knowledge, affect and interact across different health care AI contexts.

Objective: We aimed to investigate people's trust in and acceptance of AI across health care use cases and determine how context and perceived risk affect individuals' propensity to trust and accept AI in specific health care scenarios.

Methods: We collected and analyzed web-based survey data from 1100 Finnish participants, presenting them with 8 AI use cases in health care: 5 (62%) noninvasive applications (eg, activity monitoring and mental health support) and 3 (38%) physical interventions (eg, AI-controlled robotic surgery). Respondents evaluated intention to use, trust, and willingness to trade off personal data for these use cases. Gradient boosted tree regression models were trained to predict responses based on 33 demographic-, personality-, and technology-related variables. To interpret the results of our predictive models, we used the Shapley additive explanations method, a game theory–based approach for explaining the output of machine learning models. It quantifies the contribution of each feature to individual predictions, allowing us to determine the relative importance of various demographic-, personality-, and technology-related factors and their interactions in shaping participants' trust in and acceptance of AI in health care.

Results: Consumer attitudes toward technology, technology use, and personality traits were the primary drivers of trust and intention to use AI in health care. Use cases were ranked by acceptance, with noninvasive monitors being the most preferred. However, the specific use case had less impact in general than expected. Nonlinear dependencies were observed, including an inverted *U*-shaped pattern in positivity toward AI based on self-reported AI knowledge. Certain personality traits, such as being more disorganized and careless, were associated with more positive attitudes toward AI in health care. Women seemed more cautious about AI applications in health care than men.

Conclusions: The findings highlight the complex interplay of factors influencing trust and acceptance of AI in health care. Consumer trust and intention to use AI in health care are driven by technology attitudes and use rather than specific use cases. AI service providers should consider demographic factors, personality traits, and technology attitudes when designing and implementing AI systems in health care. The study demonstrates the potential of using predictive AI models as decision-making tools for implementing and interacting with clients in health care AI applications.

(J Med Internet Res 2025;27:e65567) doi:10.2196/65567

XSL•FO
RenderX

KEYWORDS

artificial intelligence; AI; health care technology; technology adoption; predictive modeling; user trust; user acceptance

Introduction

Background

Artificial intelligence (AI) has shown significant potential in health care and well-being. AI applications play a crucial role in the digital transformation of health care [1]. This includes transformations in areas such as medical data analysis, treatment planning, robotic surgeries, clinical operations support, diagnostics, web-based nursing, and connected health care devices [2]. AI enables new methods for how health diagnoses and treatment recommendations are delivered to patients [3]. In addition, technology adds positive effects to health care such as improving availability, accessibility, and efficiency as well as reducing the cost of health care service delivery [4].

Challenges and the Importance of Trust in AI Adoption

While there are numerous benefits to integrating technology into health care, there are also challenges. Owing to the data-driven and autonomous nature of AI, matters that already posed a concern for health care such as confidentiality (ie, information security and privacy), decision-making, accountability, and responsibility have become more pronounced from a technological perspective. Thus, it is important to examine the ethical and acceptance determinants (including challenges) of AI applications as they are crucial for patient safety and accountability and can significantly boost health consequences [5]. Relating to this, human trust is an *enabler* and a state in which individuals believe that engagement in actions or confidence in others (including objects) will result in a positive outcome [6]. Without assurance that the interests and concerns of humans (patients and professionals) are at the heart of adoption and implementation, human trust—the belief that engagement with a transaction or interaction with the technology will result in positive outcomes [7]—fails to manifest. The current landscape of mass digitalization is rife with scenarios in which individuals are dependent on digitally connected technology without trust (ie, forced digitalization). Trust can be seen as a state of certainty in which the truster believes that the trustee (human, artifact, or service) will act according to promises and expectations [6,8]. Despite the popularity of health care AI scholarship, studies that combine trust with experience of concrete application scenarios are scarce.

Trust can vary dramatically depending on the type of AI system an individual is dealing with, the context, and the level of engagement (ie, queuing for an appointment in a nonemergency situation vs undergoing open-heart surgery). In this domain, human lives and well-being are in question, meaning that the stakes are high for critical AI systems as encountered in medical AI [9]. Trust is closely related to acceptance and intention to use different types of technology, as demonstrated by various studies [10-13]. It is often mediated by perceived usefulness, ease of use, and satisfaction. Applied sociological and economic theories proposed by Pavlou and Gefen [14] revealed that the perceived effectiveness of AI at an institutional level impacted trust at the individual level, both toward the institution and the technology. This perceived effectiveness was based on 2 types of mechanisms—weak (market) and strong (legal). In addition, consideration for behavioral intention based on this perceived effectiveness was incorporated into this study: predicted preference of the technology in the future, the likelihood of using the technology soon, and the intention to use the technology in the future given increased opportunities. Ye et al [15] found that participants with high trust in AI might have high expectations for AI in health care, requiring greater perceived usefulness before engaging with an AI device.

Factors Influencing Trust and Acceptance

Recent studies highlight the influence of various background variables on attitudes toward AI in health care. Khanijahani et al [16] found that the common determinants for the acceptance of AI technologies are *perceived ease of use*, *usefulness*, *performance*, and *expectancy* among health care professionals and patients. However, as we see from the following examples of scholarship, results vary between studies. For instance, Omrani et al [17] found that men and individuals aged >55 years exhibited higher trust in AI. Conversely, Fritsch et al [18] reported that older patients, women, and those with lower education were more cautious regarding AI in health care. Riedl [9] concludes that older people are increasingly concerned by AI systems. The study by Nadarzynski et al [19] on the acceptability of AI-enabled chatbots in health care revealed several emergent themes—understanding: awareness and experience; AI hesitancy: perceived accuracy, premature technology, nonhuman interaction, and cybersecurity; and motivations: anonymity, convenience, and signposting (using words to guide processes). Individuals' opinions toward AI were measured through constructs of *perceived accuracy*, *perceptions of premature technology*, *nonhuman interaction*, *cyber security*, and *accessibility*. The findings highlight hesitancy toward chatbots corresponding with the IT skill levels of participants, as well as dislike for talking to computers. Researchers included multiple demographic factors in their analysis but found no significant association with AI acceptability.

Esmaeilzadeh [20] identified positive effects of gender (man), income, education, employment, technical knowledge, and familiarity with AI on the intention to use AI-based health care tools, while age, race, and general computer skills were not significant. On the other hand, studies such as that by Lambert et al [21] noted no gender effects on AI acceptance among health care professionals. Araujo et al [22] observed that age negatively affected perceived usefulness and positively affected the perceived risk of automated decision-making (ADM) by AI. In this instance, gender was seen to influence perceived usefulness but not risk. A study by Ho et al [23], positioned in a Japanese cultural setting, indicated negative perceptions of emotional AI technology among older adult and male patients. Factors behind these results are attributed to the fear of losing control to AI, while due to its cultural context (Japan), privacy was not a concern as there are high levels of trust regarding the standard

XSL•FO
RenderX

of care in exchange for disclosing medical details. In another cultural setting (Saudi Arabia), Alanzi et al [24] found that individuals in the age group of 43 to 58 years (generation X) were less accepting of AI, mental health, web-based assistants than younger generations, while Park and Woo [25] reported positive associations between age and both emotional and social reactions toward AI. These findings underscore the complexity of factors (cultural background and otherwise) influencing attitudes toward AI in health care, suggesting that demographic variables play significant roles in shaping perceptions and acceptance of AI technologies.

Personality and technical prowess are also relevant factors for attitudes toward AI [9,18,25-27]. Consensus appears that both agreeableness and openness positively affect trust with weaker support for extraversion. High personal technical affinity has a strong positive impact on AI acceptance in health care [18]. In addition, it appears that less-neurotic people exhibit more trust in AI systems [9]. Park and Woo [25] used Big Five personality traits and personal innovativeness in IT (PIIT) trait, which refers to "the willingness of an individual to try out any new information technology." Findings indicated that the Big Five traits and PIIT trait contribute significantly to explaining individuals' attitudes toward AI. For example, those with high extraversion showed rather negative attitudes toward AI, and individuals with high levels of neuroticism displayed more negative emotions toward AI. Those with high agreeableness displayed positive emotions and positive attitudes in the social and functionality dimensions of AI. The PIIT trait had a consistent (positive) effect in predicting all types of attitudes toward AI. For example, those who rated themselves as having poor or moderate IT skills showed lower acceptance of the technology [19]. For patients, higher personal technical affinity increases their perception of AI use in health care [18].

Strategies for AI Adoption in Health Care

Al Badi et al [28] identified 5 main dimensions and 25 subfactors for the main challenges in adopting AI in health care. Their study revealed that accuracy and privacy as well as security criteria are the most important factors toward improving AI adoption in the health care sector. Safety and security were found to be important factors for the acceptance of AI in health care in several other studies [16,21]. The results showed that general knowledge (education) had a positive association with perceptions of ADM usefulness, while domain-specific knowledge (eg, programming and algorithms) had a positive association with perceptions of ADM usefulness and fairness [22]. Characteristics of individuals and medical characteristics of patients also affect AI acceptance and adoption [16]. The use of AI has had negative consequences on trust in data privacy; patient safety; technological maturity (implementing technologies too early in development); and full automation, which has been the result of a rushed transformation from human-human service to human-technology (digital) service [4]. Patients may distrust AI-based data collection methods as part of diagnoses and treatment information [3], which indicates that there might be a difference between AI applications. The lack of available and easily accessible information may operate as a negative factor, slowing down the acceptance of data-driven services in predictive medicine [29].

Previous research has listed several means to increase user acceptance and adoption of AI in health care [4,30,31]. Kwak et al [31] state that to increase nursing students' intent to use AI, promoting performance expectancy, effort expectancy, and self-efficacy are essential interventions. These aid in fostering a positive attitude toward AI. In supporting the adoption of AI in health care, research indicates that focus should be placed more on facilitating early adoption and sustainable implementation in the health system. This particularly applies from the user's perspective [30]. For AI to be successfully adopted in health care, patients and other users must consent to use AI applications [3]. Parvinen et al [29] found that informing potential patients or customers regarding the personal benefits of AI improved their willingness to provide consent to use their genomic data. This shows that a trade-off for benefits would boost the acceptance of using data-driven AI technologies. Positive attitudes toward AI have a substantial impact on the intent to use AI in health care [31]. Positive attitude is decreased by the feeling of losing control to AI systems [23].

Chew and Achananuparp [4] suggest the following improvements to the adoption of AI in health care: enhancing user experience and design, enhancing personalization and customizability, enhancing empathy, and personification of AI-enabled chatbots and education of AI capabilities to the public. Effective governance of AI in health care is a prerequisite to precisely address regulatory, ethical, and trust issues while advancing the acceptance and implementation of AI among patients [5]. Ho et al [23] list the following suggestions that might improve emotional AI perception in patients: explaining how AI tools function (increasing transparency), explaining current legal and ethical safeguards, and describing the role of human professionals in the decision-making process might improve positive attitude. Furthermore, other crucial factors including robust data security, anonymization, and privacy measures as a part of secured data management solutions are crucial for establishing users' acceptance of AI health care applications [32]. In addition, using multimodality and several sensors for health data collection improves the accuracy and trustworthiness of AI applications [33], thus boosting the user acceptance of AI applications in health care.

Frameworks, Ethical AI Models, and Survey Design

Efforts have been made to establish frameworks to create actionable principles that relate to *data management*, *model development*, and *deployment* and *monitoring* [34]. These should exist under the banners of governance (organizational policy) and regulation (legal policy). Such frameworks resonate with the works of Vakkuri et al [35] and Rousi [36]. These can be applied as general guidelines and principles for development, yet context-dependent information must be sought additionally. The priorities and importance of factors affecting acceptance and adoption vary depending on the study. While the previous studies are extensive, we felt that they did not examine the element of trust and the specificities of application context variations to the extent that we are interested in.

For this study, we sought further to find theories and constructs that would enable us to delve deeper and in detail in terms of understanding the varying dynamics of individual perception

XSL•FO
RenderX

of AI in health care according to dimensions and context. We used the ECCOLA model [35] and the Robot Governance Model [36] to establish the ethical AI themes and dimensions under examination. The ECCOLA model was developed on 3 prominent AI ethics studies and guidelines: the systematic mapping of global guidelines by Jobin et al [37], revealing 5 ethical principles (justice and fairness, transparency, responsibility, nonmaleficence, and privacy); Ethically Aligned Design by IEEE Xplore [38]; and the Ethics Guidelines for Trustworthy AI by the European Commission [39]. ECCOLA features 8 themes designed to be practically applicable for increasing ethical awareness in AI development processes. These 8 themes are stakeholder analysis, transparency, data, agency and oversight, safety and security, fairness, well-being, and accountability. The Robot Governance Model adds to this in that it depicts the properties (sense, process, and act) of the AI systems; connects these to specific challenges; and then accounts for layers of governance in relation to accountability and responsibility. In other words, these frameworks were chosen for their thoroughness, multidimensionality, and detail regarding applied settings.

The AI ethics–related themes and dimensions were built-in within the questionnaire in relation to constructs probed that were obtained from studies across the disciplines of human-computer interaction, information systems, and digital health. In this study, themes related to data, privacy, accessibility (fairness), well-being (trust and trade-offs), and safety and security (cybersecurity) are seen as integrated into the questions to *lower* the concepts from high-level principles to actionable propositions. The scenarios were used to tangibly depict the AI applications, while the questions featured were based on the Robot Governance Model, specifically focusing on the "systems and artifacts" layer. The AI acceptability study by Nadarzynski et al [19] was instrumental in formulating the constructs on the propensity to trust AI, also stemming from the foundational work by Pavlou and Gefen [14]. These were also complemented by other constructs obtained from the study by Cheung and To [40] investigating the propensity to trust in-app advertisements.

Using the aforementioned approach with a robust machine learning–based analysis, we established an understanding of the antecedents for trusting and using AI in specific health care use cases. Our main research questions were as follows:

1. How does context affect an individual's willingness to trust and accept AI in health care?
2. How does the perceived level of risk (physical, psychological, and social vulnerability) affect an individual's propensity to trust and accept AI in specific health care settings?

By addressing these questions, we aim to contribute to a deeper understanding of trust dynamics in AI health care applications, ultimately informing strategies for enhancing user acceptance.

Methods

Overview

Our methodological approach aimed to investigate public opinions on AI applications in health care and well-being among consumers. We used a web-based questionnaire and established theoretical frameworks from human-computer interaction, information systems, and digital health. The survey comprised demographic questions, evaluations of 8 AI use cases, and general opinions on AI in health care. After data collection, data preprocessing was conducted to ensure the quality and variability of responses. This was followed by confirmatory factor analysis to identify underlying constructs within the data. For predictive modeling, we used the CatBoost gradient boosted tree regression algorithm to capture nonlinear relationships and complex interactions. We applied the Shapley additive explanations (SHAP) technique to investigate model predictions for each predictor and their interactions. This multifaceted approach enabled us to effectively explore the factors influencing individuals' attitudes toward AI in various health care contexts.

Web-Based Survey Development and Distribution

Our web-based survey contained three parts:

1. Demographic and personal information (31 questions)
2. Opinions toward AI in health care and well-being for 8 use cases, from which 5 were randomly selected per participant to avoid waterfall effects (21 questions per use case)
3. General opinions of AI in health care and well-being (31 questions)

Questions in parts 2 and 3 were considered response variables, while questions in part 1 were considered predictor variables. For the development of the questions, we used frameworks developed by Vakkuri et al [35] and Rousi [36]. Our theoretical constructs were drawn from human-computer interaction [40], information systems [14], and digital health [19]. We adapted the work by Pavlou and Gefen [14] on the *propensity to trust* [19], characterized by *ease to trust*, *tendency to trust*, and *willingness to trust* with the assumption that main respondents would not be of expert level in terms of understanding how AI systems operate. For the intention to use AI, we adapted the Unified Theory of Acceptance and Use of Technology model [41], which has been adapted by various authors [10,11,13,20,24]. The question related to exercising habits was adapted from official well-being surveys by the Finnish Institute for Health and Welfare [42]. For personality traits, we applied a 10-question battery by Gosling et al [43]. Education and working fields were based on official categories by the Statistics Finland department. The remaining questions related to interest and knowledge of technology and the state of health were self-developed. Refer to Multimedia Appendix 1 [10,11,13,14,19,20,24,35,36,40,41] for detailed information on questions-related sources and how we combined them into factor constructs (Tables S1 and S2 and Figure S1 in Multimedia Appendix 1).

Overall, each respondent was asked to answer 167 questions, which typically took 15 to 20 minutes. Quantitative questions were numerical, single, and multiple choice as well as Likert type. In addition, respondents were allowed to respond via open-text questions (eg, feedback). The survey was implemented using LimeSurvey (version 6.0; LimeSurvey GmbH). Data were collected between May 12, 2023, and May 23, 2023, using a survey pool. A total of 1125 people responded to the survey,

XSL•FO
RenderX

out of which a smaller subset (1100/1125, 97.78%) finished the survey fully or partially. A printed version of the survey can be found in Multimedia Appendix 2. Data were collected by a third-party service with access to a large pool of Finnish consumers, with each responder receiving a small monetary reward. The survey was offered both in Finnish and English.

We selected 8 AI use cases, which was considered maximum while still keeping the number of questions and response time reasonable. The criteria for choosing those use cases were to obtain a versatile description of the various AI tools in different contexts, ranging from self-care to hospital environments. As stated by Riedl [9], trust can vary strongly depending on what kind of AI system a user is dealing with (critical vs noncritical). On the basis of the variables of the frameworks [35,36], we selected applications that described different use cases where these aspects appeared differently depending on the use contexts. Thus, variables such as the sensitivity of the data being processed and the level of human agency and oversight associated with the AI-generated advice or actions were considered. The 8 AI use cases featured in the survey were as follows (short form in parenthesis):

1. Activity monitoring AI system to support a healthy lifestyle (*activitymonitor*)
2. Menstrual cycle monitoring and prediction AI system (*menstrual*)
3. AI-controlled robotic surgeon (*robotsurgeon*)
4. Nutrition monitoring and planning AI system (*nutrition*)
5. Real-time health monitoring, analysis, and prediction AI system (*healthmonitor*)
6. Mental health and well-being AI system (*mentalmonitor*)
7. Bioelectronic real-time health monitoring and adjustment AI system (*bioelectric*)
8. AI-controlled robotic nursing and caregiving system (*nursing*)

Use cases are illustrated in Figure 1 with detailed descriptions as present in the survey shown in Multimedia Appendix 3.

Figure 1. The 8 use cases from the health care and well-being domain included in the survey. These include both invasive and noninvasive cases that are either already available (eg, activity monitoring), emerging, or futuristic (eg, robotic surgeon and nursing system). AI: artificial intelligence.

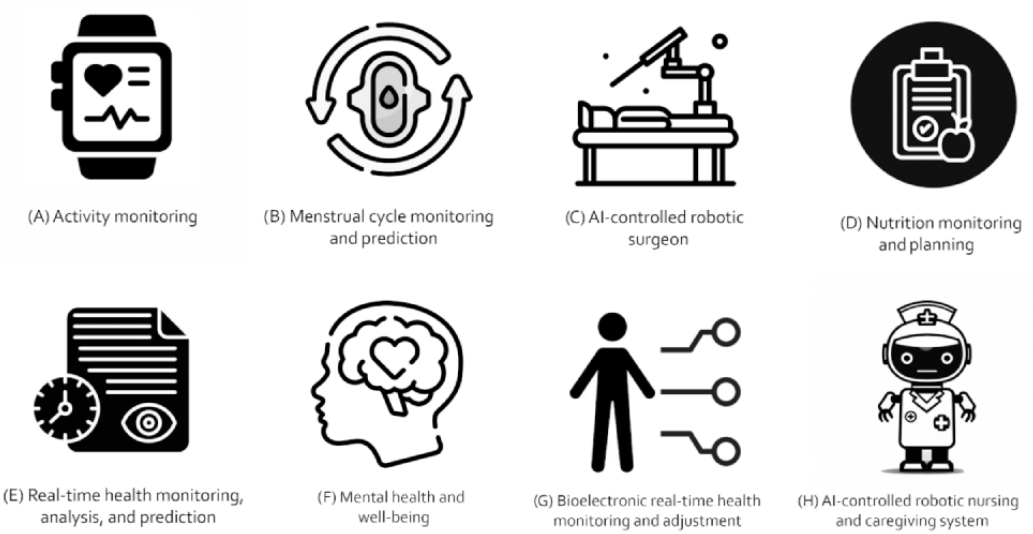

Data Preprocessing and Factor Constructs

Response data were converted into a table with rows representing responses from participants. Responses from part 1 of the survey were converted into the following variables in the 3 main groups. Variable abbreviations with descriptions related to basic demographics and personality types are as follows (variables and variable types are provided in parenthesis):

- *age*: age in years (numerical)
- *gender*: gender identity with 3 options (categorical)
- *education_level*: the highest level of education obtained with 5 options (categorical)
- *education_field_* and work_sector_**: field of education and current work sector including education (*education*); commerce, administration, law, or services (*business_services*); science, information and communication technology, or engineering (*science_ICT_engineer*); health and wellness (*health*); and all others (*other*); and all with yes or no options (binary)
- *persona_**: 10 personality dimensions with questions critical or quarrelsome (*critical_quarrelsome*), dependable or self-disciplined (*dependable_selfdisciplined*), anxious or easily upset (*anxious_easilyupset*), open to new experiences or complex (*newexperiences_complex*), reserved or quiet (*reserved_quiet*), sympathetic or warm (*sympathetic_warm*), calm or emotionally stable (*calm_stable*), conventional or uncreative (*conventional_uncreative*), disorganized or careless (*disorganized_careless*), and extraverted or enthusiastic (*extraverted_enthusiastic*), all with 7 Likert-scale options (ordinal)

Variable abbreviations with descriptions related to technology use, knowledge, and attitudes are as follows:

- *IT_skills*: self-evaluated IT skills compared to others with 5 options (ordinal)

XSL•FO
RenderX

- *technology_adoption*: self-evaluated adoption and use of new technology with 3 options (categorical)
- *technology_usage*: frequency of using technology to monitor own health and well-being with 5 options (ordinal)
- *AI_knowledge*: self-evaluated knowledge and experience of AI, a summary construct (numerical)
- *technology_attitude*: attitude toward technology, a factor construct (numerical)

Variable abbreviations with descriptions related to personal health, opinion on health care services, and physical activity are as follows:

- *health_status*: self-evaluated overall health status in the past 12 months with 5 options (ordinal)
- *health_service_usage*: frequency of health services use in the past 12 months with 5 options (ordinal)
- *health care_services*: how does one feel about the current state of health care services (private and public sector in Finland) with 5 options (ordinal)
- *exercise_level*: the amount of exercise and physical straining of oneself in free time with 4 options (categorical)

In addition, the variable *use_case* included 8 options. Detailed variable listings can be found in Table S1 in Multimedia Appendix 4, including a listing of all response options.

During preprocessing, participants who did not proceed to part 2 of the survey were removed. Furthermore, participants whose data lacked variability were excluded based on the assumption that they were mostly motivated by the reward without putting real effort into their responses or showing signs of survey fatigue. The criteria for the variability were evaluated for Likert-type questions, entailing that the minimum allowed SD was 0.15. Furthermore, we excluded participants whose responses were too extreme without using the Likert scale properly. For this, we set a threshold of 90%, and responses surpassing this ratio were excluded. These criteria were applied individually for use cases (part 2) and part 3 of the survey to avoid the removal of too much data. Ordinal categorical responses (eg, Likert type) were converted into numerical integer values using the conversion rules listed in Multimedia Appendix 4. After this initial preprocessing, *Lavaan* (version 0.617) [44] confirmatory factor analysis for R (R Foundation for Statistical Computing) was applied to create factor constructs from the Likert-type question batteries in parts 2 and 3 of the survey, as well as 1 question from part 1. The "*cfa*" function was applied with a maximum likelihood estimation algorithm. The model's reliability was evaluated by calculating the comparative fit index, Tucker-Lewis Index, and root mean squared error. Internal consistency of factors was evaluated by computing the omega, alpha, and average variance explained [45], making sure our factor constructs were reliable and within acceptable limits [46-48]. See Multimedia Appendix 1 for details.

Regression Modeling With Gradient Boosting

A CatBoost gradient boosted tree regression algorithm with multivariate root mean squared error loss was used to train predictive models [49] for parts 2 (model l) and 3 (model 2) of the survey. CatBoost builds an ensemble of decision trees, where each tree learns to improve prediction accuracy by correcting errors made by the previous trees. CatBoost can natively handle multivariate responses with both numerical and categorical predictor variables, as well as missing values. It can also approximate nonlinear relationships and complex, multilevel interactions, which are not feasible with linear models. Gradient boosting has been successfully applied to modeling in various fields, including behavioral [50] and health sciences [51].

For result interpretations and model analysis, we applied the SHAP method [52]. It is a method for interpreting machine learning models by calculating SHAP values, which quantify the contribution of each feature to the model's predictions, such as tree-based gradient boosting methods, including CatBoost. On the basis of game theory, the SHAP algorithm assigns each feature a contribution score that indicates how much it influences a prediction. For tree models, the SHAP algorithm efficiently calculates these contributions by breaking down the model's complex decision processes into simpler, additive components. This approach helps reveal which features have the greatest impact on predictions, allowing us to see not only the individual contributions but also how features interact with each other in determining outcomes. Tree SHAP analysis provides both local (sample level) and global (average) interpretations. Locally, it explains individual predictions, which can help understand individual participants. Globally, by aggregating these explanations across the dataset, the SHAP algorithm enables understanding of broader model behavior and identifying trends or dominant features [52]. In this study, both the main effects of each variable and the interaction of pairs of variables were investigated. While the main effects estimate how individual variables affect the predictions, the interaction effect is the additional combined feature effect after accounting for the main effects, thus giving more nuanced information. Interactions subtract the main effect of the features to obtain the pure interaction effect after accounting for individual effects. We report both mean and mean absolute SHAP values for features defined as follows:



(1)



(2)

where  . is a (local) SHAP value of sample *i* of feature *j* and *N* is the number of samples.

The Optuna hyperparameter optimization framework [53] was applied to find optimal CatBoost parameters for the data. Optuna is an automatic hyperparameter optimization software framework, particularly designed for machine learning purposes to optimize parameters that control the learning process. For this, the data were randomly split by setting 15% (165/1100) for testing, 15% (165/1100) for development, and the remaining 70% (770/1100) for training. Each participant was present in only 1 single split to maximize predictive power for new participants. Parameters were optimized via Optuna for the development set, and then a model was fitted using both training and development sets. Optuna was again applied in finding

XSL•FO
RenderX

parameters that maximized and minimized predictions to allow illustrative examples. Finally, trained models were applied to the held-out separate test data. The performance of the model was assessed by computing the root mean squared error of the predictions, comparing that against a constant-only model, and computing Pearson R^2 statistics for test data. We applied a permutation testing procedure to evaluate whether R^2 values were notably different from 0 [54]. In this procedure, data were randomly shuffled multiple times to obtain an empirical distribution of null values that were compared against the original (unshuffled) values to estimate *P* values.

Ethical Considerations

This study followed ethical standards and handled data with care. The research involved secondary analysis of survey data purchased from a Finnish third-party company, which is compliant with Finnish legislation. The data were collected anonymously, with no personal identifiers, so additional ethics review was not required. The third-party company obtained informed consent from participants, who willingly completed the survey after being informed of its purpose and use. Consent included provisions for secondary analysis of anonymized data by the researchers. While the third-party company had access to responder identities to ensure sample coverage, they had no access to survey responses and vice versa for the researchers. This system ensured the anonymity of responses. Potentially sensitive questions, that is, those related to personal health, were optional and skippable for the responders. All participants were compensated for their time per the company guidelines and agreements with panel members. No identifiable information or imagery is included in the manuscript or Multimedia Appendices 1-8.

Results

Measurement Constructs for the Use of AI for Use Cases

The number of valid data points was 5146 with 1100 participants. The confirmatory factor analysis for response variables (21 for each use case) resulted in a model that covered 19 items in 7 constructs. For all fit parameters and items, refer to Multimedia Appendix 1 with descriptive analysis. The 7-factor construct and their interpretations were as follows:

- *Intention* (3 items): intention to use the AI system if offered the chance
- *Trust* (3 items): trust that the AI system can make valid and accurate decisions
- *Predictions* (3 items): the importance of knowing how results were produced
- *Data* (3 items): the importance to know what type and size of training data were used
- *Privacy* (3 items): the importance to know how personal data are stored and used
- *Trade-off* (2 items): willingness to share personal data to improve AI predictions
- *Manufacturer* (2 items): importance to know the company developing the AI system

All factors were coded so that higher values indicate higher agreement, that is, toward the *strongly agree* direction in the Likert scale to aid interpretability. Many constructs for responses were highly correlated, particularly intention and trust (0.808), predictions and data (0.840), manufacturer and data (0.794), and manufacturer and predictions (0.752). All correlations are listed in Table 1.

All 33 predictor variables used in the analysis are listed in Multimedia Appendix 4.

XSL•FO
RenderX

Table 1. Correlation coefficients (n=5146 data points with 1100 participants) between 7 responses (factors) for AI use cases (part 2 of the survey).

	Intention	Manufacturer	Trade-off	Privacy	Data	Predictions	Trust
Intention							
r	1	0.022	*0.553*[a]	*–0.106*	*–0.077*	*0.14*	*0.808*
P value	—[b]	.11	<.001	<.001	<.001	<.001	<.001
Manufacturer							
r	0.022	1	*–0.136*	*0.69*	*0.794*	*0.752*	*–0.053*
P value	.11	—	<.001	<.001	<.001	<.001	<.001
Trade-off							
r	*0.553*	*–0.136*	1	*–0.468*	*–0.216*	–0.033	*0.578*
P value	<.001	<.001	—	<.001	<.001	.02	<.001
Privacy							
r	*–0.106*	*0.69*	*–0.468*	1	*0.708*	*0.613*	*–0.228*
P value	<.001	<.001	<.001	—	<.001	<.001	<.001
Data							
r	*–0.077*	*0.794*	*–0.216*	*0.708*	1	*0.84*	*–0.174*
P value	<.001	<.001	<.001	<.001	—	<.001	<.001
Predictions							
r	*0.14*	*0.752*	–0.033	*0.613*	*0.84*	1	*0.044*
P value	<.001	<.001	.02	<.001	<.001	—	.004
Trust							
r	*0.808*	*–0.053*	*0.578*	*–0.228*	*–0.174*	*0.044*	1
P value	<.001	<.001	<.001	<.001	<.001	.004	—

[a]All italicized values were significant at $P<.01$, using a 2-tailed permutation test against 0 (5000 iterations).

[b]Not applicable.

Predictive Modeling With Gradient Boosting

For the predictive model 1 (data from parts 1 and 2 of the survey), we trained a CatBoost model for all 7 response factors. The correlation coefficients for training and testing data were 0.664 and 0.324, corresponding with R^2 values of 0.441 and 0.105. See Table S1 in Multimedia Appendix 5 for detailed information on individual use cases and responses. The difference between training and testing sets indicates some potential overfitting during training. However, our model predicted >10% variance for the testing data and surpassed the chance level ($P<.001$; permutation test with 5000 iterations); thus, we considered it suitable for further analysis. For model 2 (data from parts 1 and 3 of the survey), the procedure was similar. We trained a CatBoost model for intention and trust responses. The final model reached R^2 of 0.225 with 0.235 for intention and 0.220 for trust. See Figures S1-S3 and Table S1 in Multimedia Appendix 6 for full results for model 2.

In the following subsections, SHAP values are reported. The SHAP values were computed using a trained model applied to all available data. First, we investigated the overall importance of features for all 7 responses. Then, we concentrated on intention, trust, and trade-off, for which model 1 reached the best predictive accuracies.

Combined Feature Importance for Use Cases

In Figure 2A, we depict 1 representative sample from the data with its SHAP values computed from the trained model. After combining SHAP values of all samples (n=5146 data points with 1100 participants) by taking their absolute values and means and combining those with all 7 responses, we obtained magnitudes of impact to predictions for all 33 features. This is depicted in Figure 2B, where we have sorted features from the most important (at the top) to the least important (at the bottom) according to mean SHAP magnitudes. The order of features is the same in Figures 2A and 2B. Note that the ranking of features can differ between individual responses; for example, see Figures 3A and 6A for intention and trade-off responses, respectively.

The results indicate that the most important features overall (ranks 1 to 5) were technology_attitude, use_case, technology_adoption, gender, and technology_usage. Other important features covered education_level, age, AI_knowledge, and various personality features. Features related to the education field and work sector were negligible.

XSL•FO
RenderX

Figure 2. Impact of all 33 features on the 7 response variables measured via Shapley additive explanations (SHAP) main effect values. (A) Single representative sample with SHAP values of each feature contributing to the 7 response predictions. Values in parenthesis correspond to numerical Likert encodings. (B) Mean absolute values of SHAP values over all samples (n=5146 data points with 1100 participants) for each response stacked together for all 7 responses. Bars represent the relative impact of each feature for model predictions. B: binary (yes or no); C: categorical; ICT: information and communication technology; N: numerical; O: ordinal categorical.

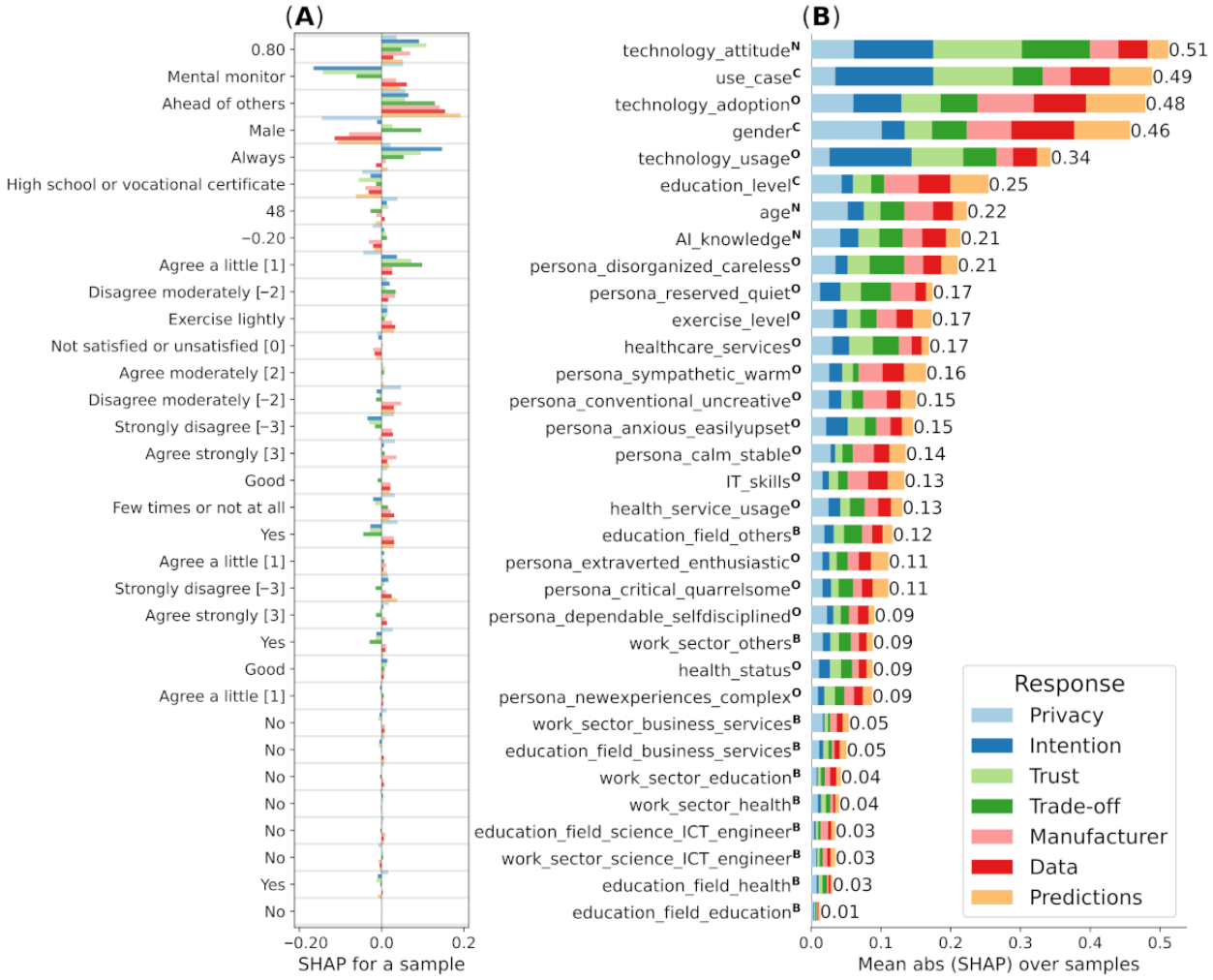

Feature Importance for Intention to Use of Use Cases

Feature importance for intention (ie, intention to use the AI system if offered the chance) are depicted in Figure 3, including the mean of absolute SHAP values and SHAP values for individual use cases and interaction strength for the most important variables. In Figure 3B, we have sorted columns (use cases) according to the similarity of SHAP values for easier visual comparison.

In Figure 4, SHAP estimates are depicted for technology-related variables technology_attitude, technology_usage, and technology adoption, as well as a responders' view of the status of current health care services (variable health care_services). The shaded regions in subplots in Figures 4A, 4B, and 4D correspond with the 75th percentile of data with the mean value at the center, both computed using a sliding-window approach, while box and whiskers in the subplot in Figure 4C correspond with the 50th and 75th percentile of data. These variables remained highly similar for trust and trade-off; hence, separate plots for those are omitted.

XSL•FO
RenderX

Figure 3. Impact of features on the intention to use the AI system if offered the chance (intention) response measured via Shapley additive explanations (SHAP) values (all data points; n=5146 data points from 1100 participants). (A) Mean absolute SHAP main effect values divided between use cases, (B) mean absolute SHAP interaction values with respect to the use case and selected variables, and (C) main effect SHAP values for each use case split between male and female (n=1050, 95.45%) participants.

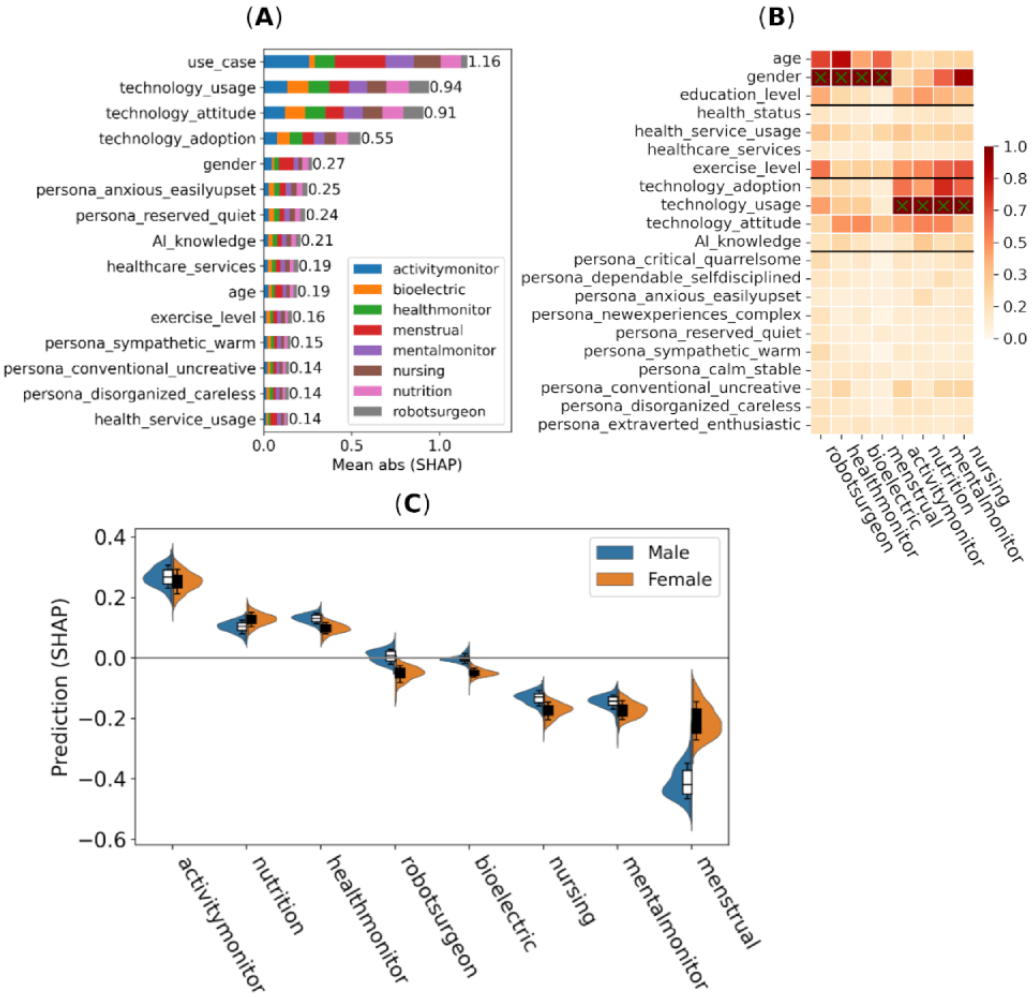

XSL•FO
RenderX

Figure 4. Impact of features on the intention to use the artificial intelligence (AI) system if offered the chance (Intention response) measured via Shapley additive explanation (SHAP) values (all data points; n=5146 data points from 1100 participants). The shaded and boxed regions correspond with the 75th percentile of data. (A) Technology attitude, (B) technology use, (C) technology adoption, and (D) opinion on the status of health care services.

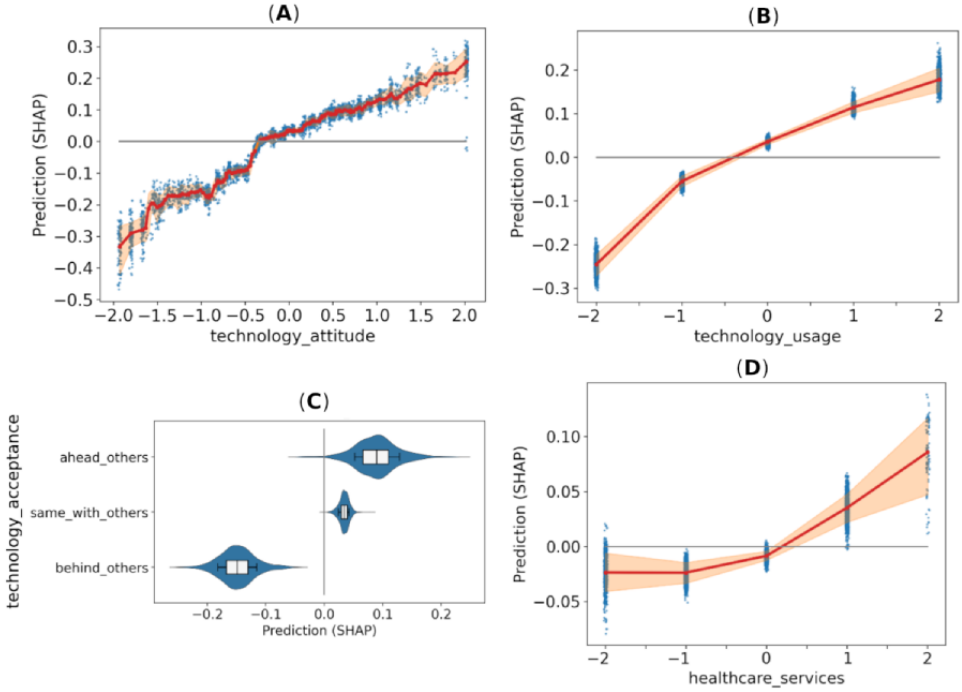

Feature Importance for Trust Toward Use Cases

Feature importance for trust that AI systems can make valid and accurate decisions (trust) are depicted in Figure 5, including the mean of absolute SHAP values and SHAP values for individual use cases and interaction strength for the most important variables.

In distinction from Figure 3B, the importance of the exercise_level variable is elevated for activity and mental monitor, nutrition assistant, and nursing robot, as shown in Figure 5B.

XSL•FO
RenderX

Figure 5. Impact of features on trust that artificial intelligence (AI) systems can make valid and accurate decisions (trust response) measured via Shapley additive explanation (SHAP) values (all data points; n=5146 data points from 1100 participants). (A) Mean absolute SHAP main effect values divided between use cases, (B) mean absolute SHAP interaction values with respect to the use case and selected variables, and (C) Main effect SHAP values for each use case split between male and female (n=1050, 95.45%) participants.

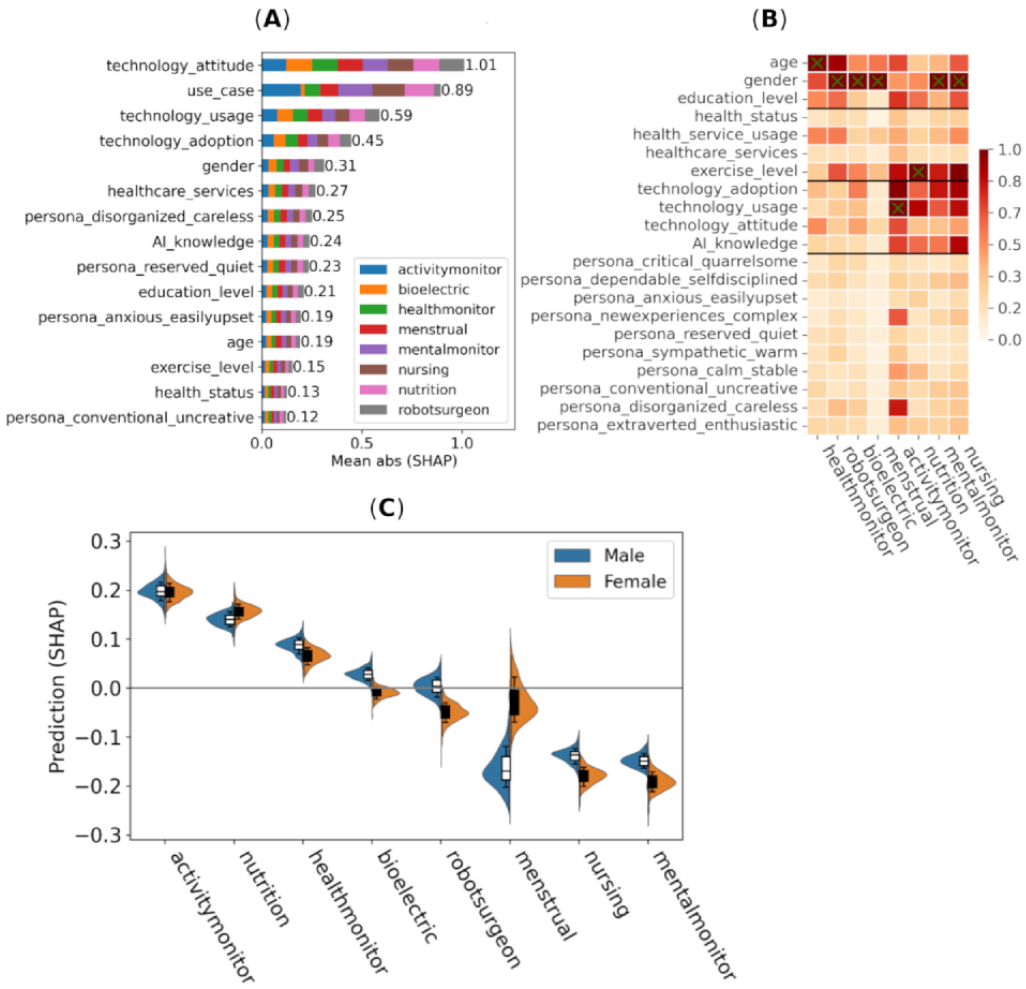

Feature Importance for the Trade-Off Willingness of Use Cases

Feature importance for willingness to share personal data to improve AI predictions (trade-off) is depicted in Figure 6, including the mean of absolute SHAP values and SHAP values for individual use cases and interaction strength for the most important variables.

XSL•FO
RenderX

Figure 6. Impact of features on willingness to share personal data to improve artificial intelligence (AI) predictions (trade-off response) measured via Shapley additive explanation (SHAP) values (all data points; n=5146 data points from 1100 participants). (A) Mean absolute SHAP main effect values divided between use cases, (B) mean absolute SHAP interaction values with respect to the use case and selected variables, and (C) main effect SHAP values for each use case split between male and female (n=1050, 95.45%) participants.

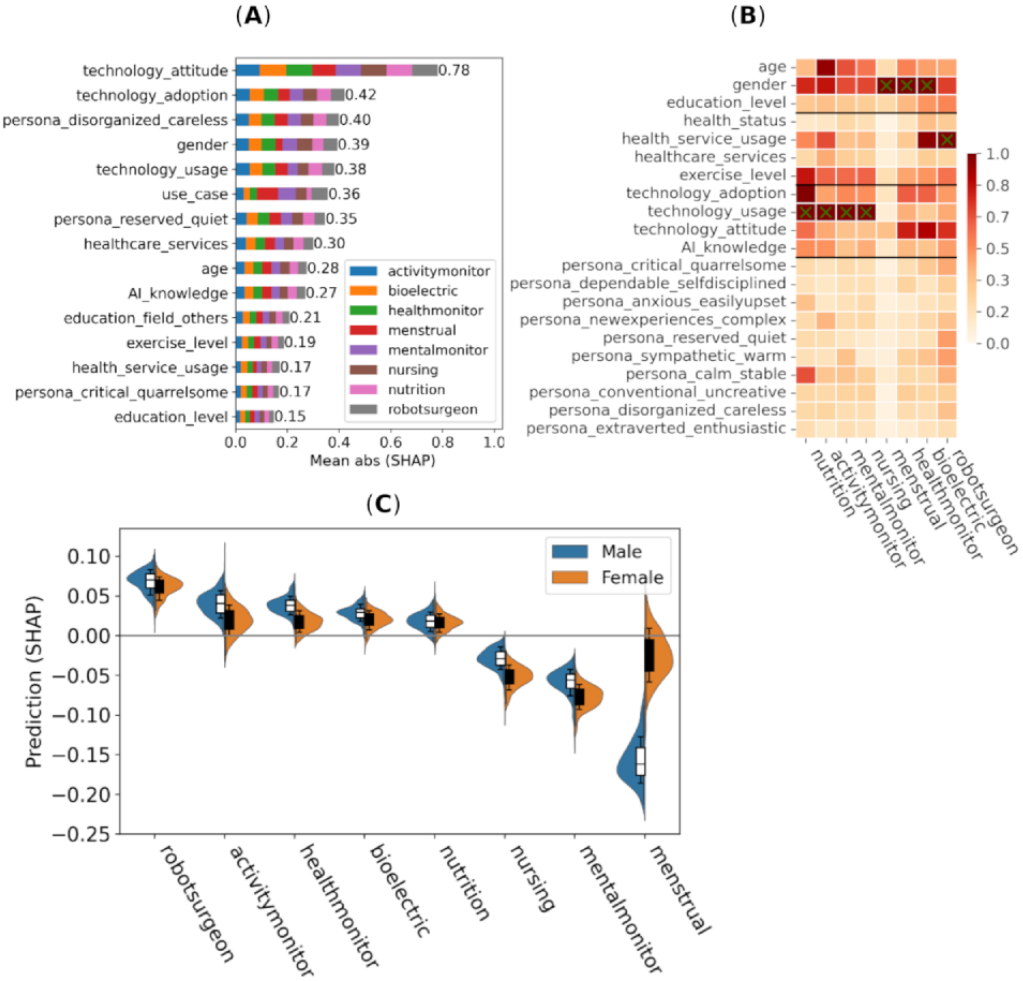

Impact of Age, Knowledge of AI, and Personality

Finally, a side-by-side comparison of the other highly important variables, AI knowledge, and age, is depicted in Figure 7. Positive and negative values correspond with increasing and decreasing contributions to the prediction, respectively. Results indicate that those who were least familiar with AI were also less interested in using or trusting AI. Older people were generally more reserved, at least those aged 65 to 70 years. People aged >70 years were most positive toward AI when considering the trade-off measure.

Figure 8 depicts the effect of personality traits. Each colored cell corresponds with the mean of SHAP values over low (Likert values of –3 and –2), neutral (Likert values from –1 to –1), and high (Likert values of 2 and 3) responses. Positive values correspond with increasing contributions to the prediction and vice versa. The strongest patterns by overall importance remain similar between all 3 responses, namely with the positive relationship for *disorganized and careless* and *anxious and easily upset* and negative for *reserved and quiet*.

XSL•FO
RenderX

Figure 7. Impact of artificial intelligence (AI) knowledge and age on the intention, trust, and trade-off response measured via Shapley additive explanations (SHAP) main effect values (all data points; n=5146 data points from 1100 participants). The shaded region corresponds with the 75th percentile of data.

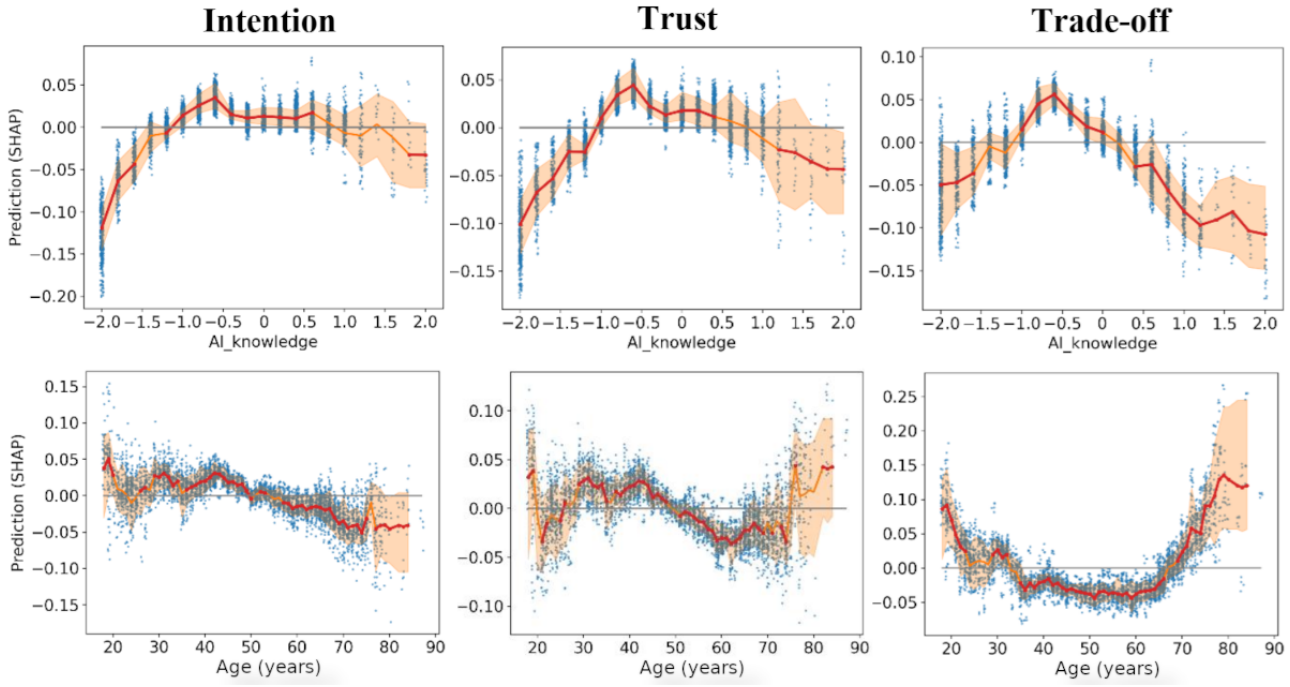

Figure 8. Impact of personality on intention, trust, and trade-off responses measured via Shapley additive explanations (SHAP) main effect mean values (all data points; n=5146 data points from 1100 participants). Plus (+) and minus (–) signs correspond to 75th percentile of SHAP values having positive or negative signs.

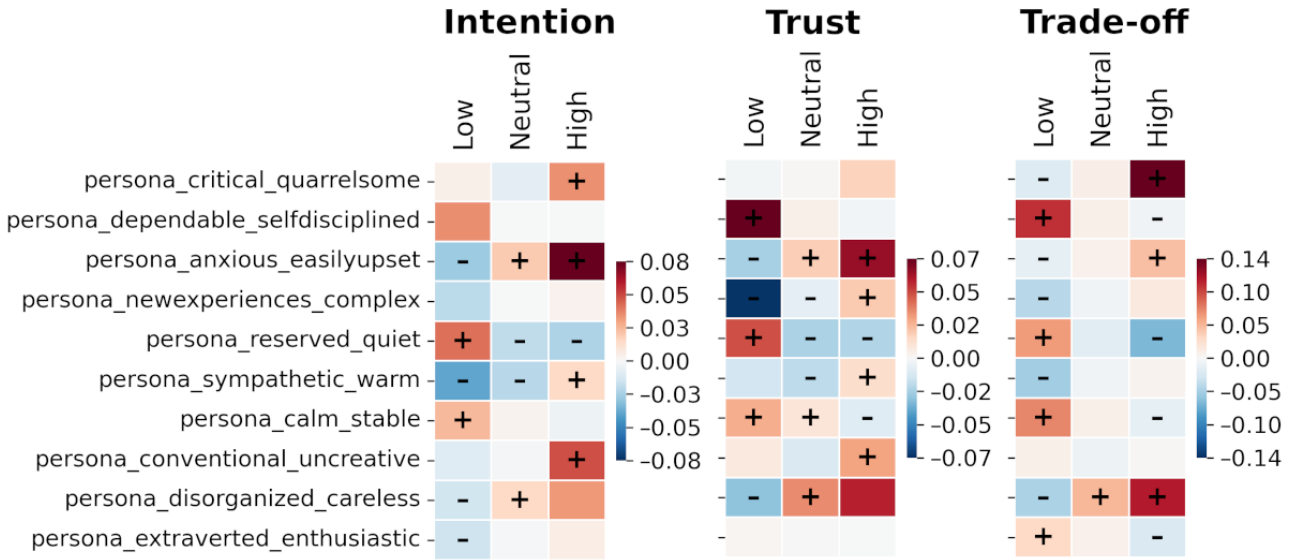

Nonlinear Interactions and Example Predictions

SHAP interaction matrices involving variable pairs are depicted in Figures S1 and S2 Multimedia Appendix 7 for both models. Strongest interactions involve variables age, gender, education_level, exercise_level, health care_services, health_status, technology_adoption, technology_attitude, technology_usage, and AI_knowledge. In addition to the use case, gender had strong interactions with education level, technology attitude, health status, and age. Personality had typically weak interactions with other variables, particularly for model 1. The strongest pairwise interactions are illustrated in Figures S3 (model 1) and S4 (model 2) in Multimedia Appendix 7.

We noticed that low and high levels of technology attitude had different effects depending on education level, gender, and AI knowledge. For example, low values of technology attitude in female participants can increase trust toward AI, while the effect was negative for male participants (Figure S3 in Multimedia Appendix 7). The same was noticed for high school or vocational educational backgrounds versus others (Figure S3 in Multimedia Appendix 7). The main effects showed a general positive relationship between opinion on health care services and acceptance of AI (Figure 3D); however, interactions

XSL•FO
RenderX

revealed that age affects this relationship; opinions toward AI were reversed from between those aged <55 years (positive relationship) and those >55 years (negative relationship; Figure S3 in Multimedia Appendix 7). A similar type of "tipping point" effect occurred between technology attitude and AI knowledge (Figure S4 in Multimedia Appendix 7), technology attitude and gender (Figure S4 in Multimedia Appendix 7), and age and disorganized and careless personality (Figure S4 in Multimedia Appendix 7). Additional findings are listed in Multimedia Appendix 7.

Finally, in Multimedia Appendix 8, we depicted illustrative predictions for individual use cases with either very high or low predictions for trust (Figure S1 in Multimedia Appendix 8) and trade-off (Figure S2 in Multimedia Appendix 8) side by side. Predictions were dominated by technological attitude and adoption, but personality traits and gender also had a notable impact.

Discussion

Overview

In this study, the acceptance, trust, intention, and related views of consumers (health and medical care clients and prospective patients) toward the use of AI in the field of health care were evaluated to map the views of individuals, such as prospective patient, consumer, and user, toward AI use cases. The cases ranged from simple health-related applications to more futuristic cases of invasive AI solutions, such as surgical robotics. Using CatBoost and SHAP allowed us to find nonlinear relationships between predictors and responses. Evaluations were collected about the intention to use (intention), trust for AI (trust), model predictions (predictions), training data (data), privacy issues (privacy), willingness to trade-off information for benefits (trade-off), and manufacturer importance (manufacturer) for each use case.

Principal Findings

Correlations between the 7 properties were statistically different from 0 (P<.01; Table 1) for all except intention and manufacturer and trade-off and predictions. Correlations were highest (>0.7) for intention and trust (0.808), predictions and data (0.840), data and manufacturer (0.794), predictions and manufacturer (0.752), and data and privacy (0.708). For negative correlations, the strongest was privacy and trade-off (–0.468), illustrating that people with a higher demand for privacy are not eager to compromise it for benefits. For the overall case, evaluations were analyzed for intention and trust and concerns over accessibility and cybersecurity in AI. For these, the strongest positive correlation was found for intention and trust (0.768) and a negative correlation for trust and cybersecurity (–0.591; Table S1 in Multimedia Appendix 6). The latter confirms that concern regarding security in AI is strongly associated with decreased trust in technology. For intention and trust, our results were at par with existing knowledge of the strong connection between the two [10,11,13].

For the predictive model 1 (with use cases), the single most important variable was technology attitude (rank 1), followed by related variables technology adoption (ranks 2 and 7), technology use (ranks 5 and 2), and AI knowledge (ranks 8 and 4), as depicted in Figure 2 and Figure S1 in Multimedia Appendix 7. Technology attitude was more important than the use case (rank 2). This indicates that general attitude, view, and knowledge of technology are the primary drivers of how consumers evaluate AI applications. This sets the baseline for the individual evaluation of AI that is further adjusted by particular use cases. This is in line with the previous research that showed how attitudes impact the willingness to accept technology, particularly AI in health care [18,19]. Personality had a major impact on all outputs. Although individual personality traits alone were not of the top importance, most of them (6/10, 60%) were considered more important than health status, work sector, or education field (Figure 3). In fact, the pooled importance of all 10 personality traits (a SHAP value of 1.38) came close to the pooled importance of 4 AI attitude, use, and knowledge-related variables (a SHAP value of 1.54). Thus, this study only slightly supports earlier research findings. In addition, as stated earlier, we critiqued the clarity of the earlier research findings. The findings of this study also display a positive link between attitude toward the state of health care services and trust, intention to use, and willingness to trade-off. People with positive and very positive views toward health care services were also positive toward AI (Figure 4).

With regard to trust and intention, our use cases were divided into 3 tiers of popularity (Figures 3C and 5C):

- Tier 1 (ranks 1 to 3): activity monitor, nutrition monitor, and health monitor (noninvasive)
- Tier 2 (ranks 4 to 5): robot surgeon and health monitor (bioelectric)
- Tire 3 (ranks 6 to 8): nursing robot, mental monitor, and menstrual monitor

However, predicted differences were small, around 0.5 on the numerical scale (ie, half step in the Likert scale). As expected, the gender difference was notable for the menstrual monitor (+0.1 for women) and also for the bioelectric health monitor, nursing robot, and robot surgeon (+0.02 for men). A similar finding was reported by Omrani et al [17] with men having higher levels of trust in AI than women. Looking into interaction effects (Figures 3B and 5B) between use cases and selected demographic and personality traits revealed another distinction between the following use cases:

- Group 1: robot surgeon, health monitors (noninvasive and bioelectric), and menstrual monitor
- Group 2: nursing robot, mental monitor, nutrition monitor, and activity monitor

For group 1, the most important features were gender and age. For group 2, the most important features were technology adaptation, exercise level, technology use, and AI knowledge. There was a negligible effect from health status, health service use, technology attitude, or personality traits, which was important only in setting the baseline opinion toward AI. The results for model 2 (without use cases) remained similar to those of model 1 regarding the most important features, which were technology attitude, technology use, and view on health care services (Figure S2 in Multimedia Appendix 6).

XSL•FO
RenderX

In addition to the use case, gender had notable interactions with education level, technology attitude, health status, and age, as measured via SHAP values. We noticed the following: (1) increasing technology attitude decreased trust and use intention for female participants; (2) with "ahead of others" skills in technology adoption, male participants had a higher willingness for trade-off compared to female participants; (3) among those with good health status, female participants had higher trust and use intention compared to male participants; and (4) increased age for female participants also increased trust, use intention, and willingness for trade-off, while the opposite was true for male participants (Multimedia Appendix 7).

Attitude toward technology was by far the most important factor for trade-off (0.78 in SHAP). At the same time, the use case was the sixth most important factor (0.36 in SHAP; Figure 6). The ranking of use cases was similar to above, however, with the robot surgeon now having ranked first (vs fourth and fifth earlier). This indicates that consumers acknowledge the importance of having the most accurate and detailed data about themselves available for invasive operations, such as surgery, and are willing to share sensitive health data. A robot surgeon could be preferred due to its potential accuracy, while a robot nurse would not be preferred due to the absence of social interaction and perhaps empathy with the patient. On the other hand, the same effect was not present for the mental health case (a SHAP value of –0.12 against the robot surgeon). One explanation for this could be that data related to mental health are considered more sensitive than physical data, even regarding health benefits. For interactions, a similar grouping as before was found regarding gender (group 1) and technology use (group 2) being the most important variables for each group.

For personality traits, we found that out of all 10 traits, the most important were *disorganized and careless* (rank 9) and *reserved and quiet* (rank 10 overall). This corresponds with the findings by Park and Woo [25], where extroverts were shown to demonstrate a more negative disposition toward technology than others. However, the contribution seen in this study is that the importance of individual traits varied depending on the output dimension. Covariance with outputs was mostly monotonous with respect to the Likert scale measured via the SHAP algorithm. Relationships in the 7 dimensions were positive for all except for *dependable and self-disciplined*, *reserved and quiet*, and *calm and stable* (Figure 8). It may be summarized that if a person has high levels of disorganized and careless, anxious and easily upset, and critical and quarrelsome traits and low levels of dependable and self-disciplined as well as reserved and quiet traits, they are likely to have higher levels of trust, intention to use, and willingness to do trade-offs for AI. Despite having strong main effects, interaction effects for personality with other variables were generally weak, particularly for model 1 (Multimedia Appendix 7). However, for model 2, some notable interactions were present for disorganized and careless, reserved and quiet, and extraverted and enthusiastic, particularly with AI knowledge. A high disorganized and careless score was associated with an increasing (increasingly positive) relationship with respect to AI knowledge toward trust and intention. The opposite relationship was present with respect to age.

We also found notable nonlinearities for certain features, highlighting the complex interactions underlying the data and the corresponding need for nonlinear modeling to explain them. For model 1, a decreasing (increasingly negative) relationship was seen toward intention, trust, and trade-off that started around the age of 40 years (Figure 7). However, people aged >70 years had a strong positive relationship for trade-off with a lesser degree of trust. This indicates more willingness toward trade-off, as one's health deteriorates with age, particularly so for female participants (Figure S3 in Multimedia Appendix 7). This group recognizes the acute need for health and medical care, yet it has most likely considered or experienced current societal challenges in the underresourcing of the health and medical sectors. Thus, while trust is not high (ie, as seen in a forced digitalization relationship), the recognized need for technological intervention is understood. For model 2 (Figure S3 in Multimedia Appendix 6), three predictive relationships were found: (1) consumers aged <30 years had the lowest intention to use and trust AI, (2) middle-aged consumers aged 35 to 50 years had increased intention to use AI, and (3) the older consumers aged >65 years had the highest levels of trust in AI.

For AI knowledge, a strong inverted *U* shape was found, in which predictions for intention, trust, and trade-off were negative for those with the least and most knowledge about AI according to their own self-evaluations (Figure 7 and Figure S3 in Multimedia Appendix 6). This maintains a resemblance with the Dunning-Kruger effect [55], where people with the least and most knowledge in certain domains have the highest confidence or expectations for potential outcomes. In this study, the effect was inversed, whereby people with the self-evaluated least or most knowledge of AI possessed the most negative expectations. The effect was similar for both models 1 and 2, although for the latter, the dip for the high-end tail was smaller (Figure S3 in Multimedia Appendix 6). For age, the effect varied depending on the model and specific response. For trust and trade-off, we found a positive relationship for older responders (aged >65 years). For intention, young people (aged <30 years) were at the highest (Figure 7) or lowest (Figure S3 in Multimedia Appendix 6), indicating notable context dependency. Park and Woo [25] reported a curvilinear relationship between age and the functionality of the AI dimension, corresponding to a *U*-shaped dependency similar to our findings for trade-off. The effect of age remains diverse and context dependent.

Finally, for trust, intention, and trade-off responses, strong nonlinear interactions were found between age and health care services and technology attitude and AI knowledge (Figure S3 in Multimedia Appendix 7). For the former, an abrupt change could be observed in SHAP predictions for participants aged between 50 and 60 years: people aged >60 years with negative views on health care services (Likert responses <0) demonstrated a positive impact on outcomes. A similar effect of "flipping" of the views was found for AI knowledge concerning technology attitude (Figure S4 in Multimedia Appendix 7).

Theoretical and Practical Contributions

This study contributes to existing knowledge both methodologically and data wise. Previous studies on the subject routinely used simple statistical methods, typically linear

XSL•FO
RenderX

regression (eg, [17,22,23]) or structural equation modeling (eg, [7,20,25]). These can only consider limited numbers of variables and direct impacts without considering nonlinear interactions. Our study included 33 variables, considering all interactions between them, and a large sample of 1100 participants from all consumer groups. The importance of the variables was quantified using the SHAP method, which allows estimation and separation of main and interaction effects.

Attitude, exposure, use, and interest in technology were the main factors determining *baseline* readiness and interest in adopting AI tools in health care. There was a distinction between use cases according to how invasive (physically, socially, and psychologically) AI was perceived in the contexts. However, use cases had less overall importance on the population level than had been hypothesized at the beginning of the study. Regarding the relationship of perceived risk (vulnerability) between people and the AI systems, we assumed that the robot surgeon, mental health monitor, and menstrual monitor would be the most sensitive and least positive (more skeptical) application contexts for AI. Yet, age, gender, education level, and personality were proven to be relevant factors, while education and professional fields had negligible effects. Considering that most previous studies have been conducted on the younger population (average age ≤40 years [9]), our findings for the older population are relevant. Our results for interaction effects demonstrate that age, gender, education level, use, and attitude toward technology, and domain-specific variables (personal health and view of health care services) contribute to AI acceptance. For example, different combinations of age and education can impact female and male participants differently and nonlinearly.

While technological attitude was the major driving factor, the use case was associated with differences between genders and variable interactions. Hence, there was some context dependency. On the basis of literacy, one could assume that the application context within the use cases would significantly impact the ways in which participants evaluated the AI. Ries [27], for instance, proposed a context-dependent Bayesian model of trust, observing that trust is dependent on a number of factors that may either confirm or reject trust, such as conditions (ie, the quality and integrity of a system or object), other actors (host organization or security threats), or the level of risk (potential direct harm or no perceived risk). This stands to reason in perceptive studies as information driven through metaphors (ie, AI, health, type of procedure, or function) are interpreted via context, that is, their relationship to other elements within the perceived scenario [56]. Yet, similar to the recent findings of Stein et al [26], context does not seem to play as great a role in appraising technology as other factors such as personality type and age. The study by Stein et al [26] revealed that the motivation behind engagement (also reliant on personality type) held a strong relationship with attitude toward technology.

Our findings highlight the importance of considering a wide range of factors when determining the antecedents of trust and the use of AI in health care. These factors include the perception of ethical and privacy risks in AI-driven applications. The findings suggest that the health care context may have a more dominant influence on attitudes than specific use cases within health care. Regardless of the microapplication context (use context), the characteristics of AI in health care have an emotional weight. This would mean that all AI-driven applications are *equal*, if they are applied in the same field or overarching context, such as health care. Thus, the use cases can be interpreted rather as domains or sector-specific issues instead of contexts in and of themselves.

The findings related to AI knowledge and its reversed *U*-shape effect emphasize the importance of education and awareness raising of AI. In addition, it gravitates toward the fundamental findings of pragmatic approaches to ethical AI development [35,57], whereby awareness raising and education are critical. The results show that those with very low AI knowledge (novices) and experts are most critically disposed of AI regarding intention to use and trust the technology in health care settings, with novices showing the strongest negative attitudes. On the basis of the findings of the *U*-shaped effect, we conclude that novices and experts are the most critical. We expect experts to be aware of risks and precautions, whereas novices may be hesitant due to unfamiliarity with the technology. Thus, the majority are aware of risks to some extent, but these are not overestimated or underestimated.

Limitations

This work is not without limitations. First, some use cases were still hypothetical (eg, autonomous robot surgeon and nurse), and respondents were required to imagine their experience of these systems based on partial (textual and pictorial) information. No actual user evaluations of these AI applications could be made available. Second, although the data were a representative sample of the Finnish population and consumers regarding age, gender, and education, it was still biased toward people who chose to participate in the survey and were proficient in digital technology, at least enough to respond to web-based surveys. The fact that all participants had access to computers, phones, and the internet may have skewed the sample toward a higher acceptance of technology. Third, participating in the survey was voluntary, which might cause self-selection bias. Fourth, although our fitted models were statistically valid, their predictive power for unseen testing data was generally low (R^2 values of 0.105 and 0.225 for models 1 and 2). This may have resulted from multiple reasons, including the complexity of the survey, survey fatigue, use cases not being engaging, or some relevant predictor variables missing beyond those included. On the other hand, small R^2 values appear common for web-based survey studies and reported by other researchers as well, for example, values of 0.08 to 0.17 [25], 0.03 to 0.14 [22], and 0.02 to 0.18 [25] for linear models without data splitting and 0.11 to 0.30 for boosting models with splitting [50].

Conclusions

This study aimed to understand the willingness to use and trust AI by measuring overall attitudes to various factors in AI in health care in general, as well as dimensions of use case scenarios. The effect of context on individuals' willingness to trust and accept AI in health care proved nuanced. Instead of a use case, the opinion was mainly driven by the individual's attitude toward AI and use of technology. This carries through

XSL•FO
RenderX

to the subquestion of the relationship between perceived risk and the AI application. Noninvasive and less sensitive AI applications, such as activity, nutrition, and health monitors, were considered the most trustworthy and favored, while the nursing robot and the mental health monitor were the least trustworthy and favored. However, when it came to making trade-offs for benefits, a robot surgeon was ranked highest. Mental health monitoring is perceived with more skepticism than general health monitoring and even menstrual monitoring. In addition, we found the following patterns:

1. An inverted, *U*-shaped relationship between AI knowledge and attitudes: both the least and most knowledgeable individuals (by self-evaluation) showed more negative attitudes for intention, trust, and trade-off. This maintains a resemblance to the Dunning-Kruger effect.
2. A clear personality-attitude relationship: positive attitudes toward AI in health care were associated with higher scores in certain traits (disorganized and careless and anxious and easily upset) and lower scores in others (reserved and quiet).
3. A distinct gender effect: women were more cautious than men toward AI in health care, with these differences becoming more pronounced depending on the use case, education level, technology attitude, and age.

These findings have important implications for designing health care–related AI systems. The results indicate that demographic factors influence attitudes toward AI in health care differently across various application settings. Consumer trust and intention to use AI are driven mainly by overall interest and use of technology, as well as personality traits, age, and gender. Training predictive AI models, such as those used in this study, could be used as a tool to adjust and inform on the correct implementation of other AI tools in specific situations. Such AI models could serve as decision-making tools for deploying AI and communicating the systems to clients (ie, explainable AI). Building on these implications, further research is needed into how developing interventions with AI increases trust in AI applications. This includes investigating how educational programs and hands-on experience with AI systems might influence acceptance formation. In addition, researchers should examine how increasing understanding and experiences about the benefits and limitations of AI affect the acceptance of AI and user trust in health care. Longitudinal research tracking how trust in AI evolves over time with increased awareness and exposure would provide valuable insights into health care organizations implementing AI solutions.

Acknowledgments

This research was funded by “AI Driver” (decision number OKM/108/523/2021) and “AI Forum” (decision number OKM/116/523/2020) projects by Finnish Ministry of Education and Culture and “Emotional experience of privacy and ethics in everyday pervasive systems [BUGGED]” (decision number 348391) and “Multifaceted ripple effects and limitations of AI-human interplay at work, business and society [SYNTHETICA]” (decision number 358714) projects by Research Council of Finland.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JK was involved in the conceptualization, methodology, software, investigation, writing the original draft, formal analysis, data curation, reviewing and editing, visualization, and funding acquisition. RR was involved in conceptualization, methodology, investigation, writing the original draft, visualization, reviewing and editing, and funding acquisition. AA was involved in conceptualization, methodology, investigation, writing the original draft, reviewing and editing, and funding acquisition. All authors reviewed and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of response variables, related references, and confirmatory factor analysis.
[DOCX File , 113 KB - jmir_v27i1e65567_app1.docx]

Multimedia Appendix 2
Copy of the complete web-based survey, including all questions.
[DOCX File , 3454 KB - jmir_v27i1e65567_app2.docx]

Multimedia Appendix 3
Description of 8 artificial intelligence use cases.
[DOCX File , 2563 KB - jmir_v27i1e65567_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
List of predictor variables with descriptions and numerical mapping of ordinal responses.
[DOCX File , 29 KB - jmir_v27i1e65567_app4.docx]

Multimedia Appendix 5
Details of gradient boosting models and parameters and testing R2 values.
[DOCX File , 25 KB - jmir_v27i1e65567_app5.docx]

Multimedia Appendix 6
Predictive modeling results for the overall intention and trust (part 3 of the survey).
[DOCX File , 496 KB - jmir_v27i1e65567_app6.docx]

Multimedia Appendix 7
Pairwise interaction effects for predictive models.
[DOCX File , 2029 KB - jmir_v27i1e65567_app7.docx]

Multimedia Appendix 8
Example model predictions for maximal and minimal trust and trade-off willingness.
[DOCX File , 752 KB - jmir_v27i1e65567_app8.docx]

References

1. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407 [FREE Full text] [doi: 10.3390/ijerph20043407] [Medline: 36834105]
2. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
3. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF, AAO Task Force on Artificial Intelligence. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Transl Vis Sci Technol 2020 Aug 28;9(2):45 [FREE Full text] [doi: 10.1167/tvst.9.2.45] [Medline: 32879755]
4. Chew HS, Achananuparp P. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. J Med Internet Res 2022 Jan 14;24(1):e32939 [FREE Full text] [doi: 10.2196/32939] [Medline: 35029538]
5. Al Kuwaiti A, Nazer K, Al-Reedy A, Al-Shehri S, Al-Muhanna A, Subbarayalu AV, et al. A review of the role of artificial intelligence in healthcare. J Pers Med 2023 Jun 05;13(6):951 [FREE Full text] [doi: 10.3390/jpm13060951] [Medline: 37373940]
6. Saariluoma P, Karvonen H, Rousi R. Techno-trust and rational trust in technology – a conceptual investigation. In: Proceedings of the Human Work Interaction Design. 2018 Presented at: HWID 2018; August 20-21, 2018; Espoo, Finland. [doi: 10.1007/978-3-030-05297-3_19]
7. Robbins BG. Probing the links between trustworthiness, trust, and emotion: evidence from four survey experiments. Soc Psychol Q 2016 Aug 11;79(3):284-308. [doi: 10.1177/0190272516657546]
8. Holton R. Deciding to trust, coming to believe. Australas J Philos 1994 Mar;72(1):63-76. [doi: 10.1080/00048409412345881]
9. Riedl R. Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. Electron Markets 2022 Nov 23;32(4):2021-2051. [doi: 10.1007/s12525-022-00594-4]
10. Choung H, David P, Ross A. Trust in AI and its role in the acceptance of AI technologies. Int J Hum Comput Interact 2022 Apr 20;39(9):1727-1739. [doi: 10.1080/10447318.2022.2050543]
11. Dhagarra D, Goswami M, Kumar G. Impact of trust and privacy concerns on technology acceptance in healthcare: an Indian perspective. Int J Med Inform 2020 Sep;141:104164 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104164] [Medline: 32593847]
12. Lee JH, Song CH. Effects of trust and perceived risk on user acceptance of a new technology service. Soc Behav Pers Int J 2013 May;41(4):587-597. [doi: 10.2224/sbp.2013.41.4.587]
13. Kuen L, Westmattelmann D, Bruckes M, Schewe G. Who earns trust in online environments? A meta-analysis of trust in technology and trust in provider for technology acceptance. Electron Markets 2023 Dec 05;33:61. [doi: 10.1007/s12525-023-00672-1]
14. Pavlou PA, Gefen D. Building effective online marketplaces with institution-based trust. Inf Syst Res 2004 Mar;15(1):37-59. [doi: 10.1287/isre.1040.0015]
15. Ye T, Xue J, He M, Gu J, Lin H, Xu B, et al. Psychosocial factors affecting artificial intelligence adoption in health care in China: cross-sectional study. J Med Internet Res 2019 Oct 17;21(10):e14316 [FREE Full text] [doi: 10.2196/14316] [Medline: 31625950]

XSL•FO
RenderX

16. Khanijahani A, Iezadi S, Dudley S, Goettler M, Kroetsch P, Wise J. Organizational, professional, and patient characteristics associated with artificial intelligence adoption in healthcare: a systematic review. Health Policy Technol 2022 Mar;11(1):100602. [doi: 10.1016/j.hlpt.2022.100602]
17. Omrani N, Rivieccio G, Fiore U, Schiavone F, Agreda SG. To trust or not to trust? An assessment of trust in AI-based systems: concerns, ethics and contexts. Technol Forecast Soc Change 2022 Aug;181:121763. [doi: 10.1016/j.techfore.2022.121763]
18. Fritsch SJ, Blankenheim A, Wahl A, Hetfeld P, Maassen O, Deffge S, et al. Attitudes and perception of artificial intelligence in healthcare: a cross-sectional survey among patients. Digit Health 2022 Aug 08;8:20552076221116772 [FREE Full text] [doi: 10.1177/20552076221116772] [Medline: 35983102]
19. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
20. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020 Jul 22;20(1):170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
21. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello CP, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111 [FREE Full text] [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
22. Araujo T, Helberger N, Kruikemeier S, de Vreese CH. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 2020 Jan 01;35(3):611-623. [doi: 10.1007/s00146-019-00931-w]
23. Ho MT, Le NT, Mantello P, Ho MT, Ghotbi N. Understanding the acceptance of emotional artificial intelligence in Japanese healthcare system: a cross-sectional survey of clinic visitors' attitude. Technol Soc 2023 Feb;72:102166. [doi: 10.1016/j.techsoc.2022.102166]
24. Alanzi T, Alsalem AA, Alzahrani H, Almudaymigh N, Alessa A, Mulla R, et al. AI-powered mental health virtual assistants' acceptance: an empirical study on influencing factors among generations X, Y, and Z. Cureus 2023 Nov;15(11):e49486 [FREE Full text] [doi: 10.7759/cureus.49486] [Medline: 38156169]
25. Park J, Woo SE. Who likes artificial intelligence? Personality predictors of attitudes toward artificial intelligence. J Psychol 2022 Jan 11;156(1):68-94. [doi: 10.1080/00223980.2021.2012109] [Medline: 35015615]
26. Stein JP, Messingschlager T, Gnambs T, Hutmacher F, Appel M. Attitudes towards AI: measurement and associations with personality. Sci Rep 2024 Feb 05;14(1):2909 [FREE Full text] [doi: 10.1038/s41598-024-53335-2] [Medline: 38316898]
27. Ries S. Extending Bayesian trust models regarding context-dependence and user friendly representation. In: Proceedings of the 2009 ACM Symposium on Applied Computing. 2009 Presented at: SAC '09; March 8-12, 2009; Honolulu, HI. [doi: 10.1145/1529282.1529573]
28. Al Badi FK, Alhosani KA, Jabeen F, Stachowicz-Stanusch A, Shehzad N, Amann W. Challenges of AI adoption in the UAE healthcare. Vision J Bus Perspect 2021 Feb 25;26(2):193-207. [doi: 10.1177/0972262920988398]
29. Parvinen L, Alamäki A, Hallikainen H, Mäki M. Exploring the challenges of and solutions to sharing personal genomic data for use in healthcare. Health Informatics J 2023 Jan 18;29(1):14604582231152185 [FREE Full text] [doi: 10.1177/14604582231152185] [Medline: 36651319]
30. Sunarti S, Fadzlul Rahman F, Naufal M, Risky M, Febriyanto K, Masnina R. Artificial intelligence in healthcare: opportunities and risk for future. Gac Sanit 2021;35 Suppl 1:S67-S70 [FREE Full text] [doi: 10.1016/j.gaceta.2020.12.019] [Medline: 33832631]
31. Kwak Y, Seo YH, Ahn JW. Nursing students' intent to use AI-based healthcare technology: path analysis using the unified theory of acceptance and use of technology. Nurse Educ Today 2022 Dec;119:105541. [doi: 10.1016/j.nedt.2022.105541] [Medline: 36116387]
32. Khan UA, Alamäki A. Designing an ethical and secure pain estimation system using AI sandbox for contactless healthcare. Int J Online Biomed Eng 2023 Oct 25;19(15):166-201. [doi: 10.3991/ijoe.v19i15.43663]
33. Khan UA, Xu Q, Liu Y, Lagstedt A, Alamäki A, Kauttonen J. Exploring contactless techniques in multimodal emotion recognition: insights into diverse applications, challenges, solutions, and prospects. Multimed Syst 2024 Apr 06;30:115. [doi: 10.1007/s00530-024-01302-2]
34. Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. AI Ethics 2022 Jul 19;3(1):223-240. [doi: 10.1007/s43681-022-00195-z]
35. Vakkuri V, Kemell KK, Jantunen M, Halme E, Abrahamsson P. ECCOLA — a method for implementing ethically aligned AI systems. J Syst Softw 2021 Dec;182:111067. [doi: 10.1016/j.jss.2021.111067]
36. Rousi R. With clear intention—an ethical responsibility model for robot governance. Front Comput Sci 2022 Apr 1;4. [doi: 10.3389/fcomp.2022.852528]
37. Jobin A, Ienca M, Vayena E. Artificial Intelligence: the global landscape of ethics guidelines. arXiv. Preprint posted online on June 24, 2019 2025 [FREE Full text]
38. Shahriari K, Shahriari M. IEEE standard review — ethically aligned design: a vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In: Proceedings of the IEEE Canada International Humanitarian Technology

XSL•FO
RenderX

Conference. 2017 Presented at: IHTC 2017; July 21-22, 2017; Toronto, ON URL: https://ieeexplore.ieee.org/document/8058187 [doi: 10.1007/978-3-030-12524-0_2]
39. Smuha NA. The EU approach to ethics guidelines for trustworthy artificial intelligence. Comput Law Rev 2019;20(4):97-106. [doi: 10.9785/cri-2019-200402]
40. Cheung MF, To WM. The influence of the propensity to trust on mobile users' attitudes toward in-app advertisements: an extension of the theory of planned behavior. Comput Hum Behav 2017 Nov;76:102-111 [FREE Full text] [doi: 10.1016/j.chb.2017.07.011]
41. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
42. Healthy Finland survey. Finnish Institute for Health and Welfare. 2022 Jan 1. URL: https://thl.fi/en/research-and-development/research-and-projects/healthy-finland-survey?__cf_chl_tk=7UG2RsvQXEkckRKWMpL0zZ.nfLc1sncPjZw4jPjzDqY-1740483756-1.0.1.1-gDOvY2_DMKNHsHRz3VF_R_qYdqcmdzGyzgxYb8uQWNY [accessed 2025-03-11]
43. Gosling SD, Rentfrow PJ, Swann WBJ. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
44. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
45. Brown TA. Confirmatory Factor Analysis for Applied Research Second Edition. New York, NY: The Guilford Press; 2015.
46. Byrne BM. Structural equation modeling: perspectives on the present and the future. Int J Test 2001 Sep;1(3-4):327-334 [FREE Full text] [doi: 10.1080/15305058.2001.9669479]
47. Hair JFJ, Hult GT, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks, CA: SAGE Publications; 2021.
48. Hair JFJ, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. London, UK: Pearson Education; 2009.
49. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018 Presented at: NIPS'18; December 3-8, 2018; Montreal, QC.
50. Khan UA, Kauttonen J, Henttonen P, Määttänen I. Understanding the impact of sisu on workforce and well-being: a machine learning-based analysis. Heliyon 2024 Jan 30;10(2):e24148 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e24148] [Medline: 38293364]
51. Oparina E, Kaiser C, Gentile N, Tkatchenko A, Clark AE, de Neve JE, et al. Machine learning in the prediction of human wellbeing. Sci Rep 2025 Jan 10;15(1):1632 [FREE Full text] [doi: 10.1038/s41598-024-84137-1] [Medline: 39794488]
52. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
53. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: KDD '19; August 4-8, 2019; Anchorage, AK. [doi: 10.1145/3292500.3330701]
54. Pesarin F, Salmaso L. Permutation Tests for Complex Data: Theory, Applications and Software. Hoboken, NJ: John Wiley & Sons; 2010.
55. Dunning D. Chapter five - the Dunning–Kruger effect: on being ignorant of one's own ignorance. In: Advances in Experimental Social Psychology. Cambridge, MA: Academic Press; 2011.
56. Bosch P. Context dependence and metaphor. In: The Ubiquity of Metaphor - Metaphor in Language and Thought. Wahroonga, Australia: North Holland Publisher; 1985:141-176.
57. Vakkuri V, Kemell KK, Kultanen J, Abrahamsson P. The current state of industrial practice in artificial intelligence ethics. IEEE Softw 2020 Jul;37(4):50-57. [doi: 10.1109/ms.2020.2985621]

Abbreviations

ADM: automated decision-making
AI: artificial intelligence
PIIT: personal innovativeness in IT
SHAP: Shapley additive explanations

XSL•FO
RenderX

Edited by A Coristine; submitted 20.08.24; peer-reviewed by B Mesko, N Trakulmaykee; comments to author 02.10.24; revised version received 01.11.24; accepted 16.01.25; published 21.03.25.

Please cite as:
Kauttonen J, Rousi R, Alamäki A
Trust and Acceptance Challenges in the Adoption of AI Applications in Health Care: Quantitative Survey Analysis
J Med Internet Res 2025;27:e65567
URL: https://www.jmir.org/2025/1/e65567
doi:10.2196/65567
PMID:

©Janne Kauttonen, Rebekah Rousi, Ari Alamäki. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Validation of the Digital Health Literacy Questionnaire for Stroke Survivors: Exploratory Sequential Mixed Methods Study

Qin Ye[1*], MSN[‡]; Wei Wang[2*], BSN; Xuan Zeng[3], MSN; Yuxian Kuang[2], MSN; Bingbing Geng[2], BDSc; Song Zhou[4], FSc; Ning Liu[1], DMed

[1]Department of Basic Teaching and Research in General Medicine, Department of Fundamentals, Department of Nursing of Zunyi Medical University Zhuhai Campus, Zhuhai, China

[2]Zhuhai People's Hospital, Zhuhai, China

[3]Guangdong Provincial People's Hospital Zhuhai Hospital（Zhuhai Golden Bay Center Hospital）, Zhuhai, China

[4]The Fifth Affiliated Hospital of Sun Yat-Sen University, Zhuhai, China

[‡]1

[*]these authors contributed equally

Corresponding Author:
Ning Liu, DMed
Department of Basic Teaching and Research in General Medicine, Department of Fundamentals
Department of Nursing of Zunyi Medical University Zhuhai Campus
No. 368, Jinwan Road, Jinwan District
Zhuhai, 519041
China
Phone: 86 13631295806
Email: ln761066906@163.com

Abstract

Background: In China, there is limited research on digital health literacy (DHL) among patients with stroke. This is mainly due to the lack of validated tools, which hinders the precision and sustainability of our country's digital transformation.

Objective: This study aimed to develop and validate a DHL scale specifically for stroke survivors in China.

Methods: We used a sequential, exploratory, mixed methods approach to develop a DHL questionnaire for stroke survivors. This study comprised 418 patients with stroke aged 18 years and older. To evaluate the questionnaire's psychometric qualities, we randomly assigned individuals to 2 groups (subsample 1: n=118, subsample 2: n=300). Construct validity was evaluated through internal consistency analysis, exploratory and confirmatory factor analyses, hypothesis testing for structural validity, measurement invariance assessments using the eHealth Literacy Scale, and Rasch analyses to determine the questionnaire's validity and reliability.

Results: This study underwent 4 stages of systematic development. The initial pool of items contained 25 items, 5 of which were eliminated after content validity testing; 19 items were subsequently retained through cognitive interviews. After an interitem correlation analysis, 2 more items were excluded, leaving 17 items for exploratory factor analysis. Finally, 2 items were excluded by Rasch analysis, resulting in a final version of the questionnaire containing 15 items. The total score range of the scale was 15-75, with higher scores indicating greater DHL competence. Results showed that principal component analysis confirmed the theoretical structure of the questionnaire (69.212% explained variance). The factor model fit was good with χ^2_4=1.669; root mean square error of approximation=0.047; Tucker-Lewis Index=0.973; and Comparative Fit Index=0.977. In addition, hypothesis-testing construct validity with the eHealth Literacy Scale revealed a strong correlation (r=0.853). The internal consistency (Cronbach α) coefficient was 0.937. The retest reliability coefficient was 0.941. Rasch analysis demonstrated the item separation index was 3.81 (reliability 0.94) and the individual separation index was 2.91 (reliability 0.89).

Conclusions: The DHL Questionnaire for Stroke Survivors is a reliable and valid measure to assess DHL among stroke survivors in China.

(J Med Internet Res 2025;27:e64591) doi:10.2196/64591

XSL•FO
RenderX

KEYWORDS

stroke survivors; digital health literacy; validation; reliability; mixed methods study

Introduction

Stroke has become one of the leading causes of death and disability in adults [1]. It is characterized by a high incidence, disability, mortality, and recurrence rate and usually results in physical disability and functional impairment [2]. Approximately 40% of patients experience functional impairment after stroke onset. In particular, more than 85% of patients with stroke experience hemiplegia, which results in impaired upper limb function and decreased motor ability [3]. Many stroke survivors have difficulty accessing outpatient treatment due to mobility issues, transport difficulties, and a lack of necessary support, which in turn affects their outcome and rehabilitation process [4].

Advances in information technology have provided new opportunities for the management of stroke survivors across time and space [5]. A study has shown that web-based health services, including health information dissemination, patient data monitoring, and remote counseling, effectively manage disease risk factors and lower stroke recurrence rates [6]. Furthermore, the internet enables stroke survivors to exercise independently after being discharged from the hospital, therefore boosting the rehabilitative impact [7]. However, the huge number and varying quality of information on the internet hamper stroke survivors' capacity to properly access and use web-based health resources [8]. Digital health literacy (DHL) refers to an individual's entire capacity to search for, acquire, interpret, and evaluate health information from digital sources and use it to solve health problems [9]. DHL and internet access have lately been identified as "super social determinants of health" because of their influence on broader social determinants of health [10]. DHL is required to effectively participate in the digital age and attain optimal health and well-being [11].

DHL is a process skill that develops with personal, social, and environmental settings and is impacted by personal health concerns, educational level, health state at the time of exposure to digital health resources, information-seeking motivations, and technology used [9,12,13]. DHL, like other forms of literacy, is a discursive practice that aims to uncover meaning-construction processes and essentially structures thought and behavior patterns. Its central goal is to improve the ability of individuals to fully use digital health resources for health decision-making [14].

Several eHealth literacy assessment instruments have been developed over the past 2 decades. The eHealth Literacy Scale (eHEALS) has been widely used for self-reported assessments of individuals' abilities to use the internet [15], but it is recognized that there is a need for more comprehensive tools. The Digital Health Literacy Instrument, developed by van der Vaart and Drossaert [16], offers a broader spectrum of health 1.0 and health 2.0 skills measurement. Additionally, a comparative study by Xie and Mo [17] offered valuable information for evaluating the Digital Health Literacy Instrument and eHEALS in the older adult Chinese population. Generic tools can be used to evaluate a construct (like DHL) across a variety of populations, including healthy populations, populations with and without diseases, and populations with various diseases. The limitation is its potential insensitivity in evaluating the construct within a specific illness population, as key disease characteristics may be inadequately addressed. It is recommended that a patient population use a condition-specific instrument that focuses on content regarding clinical conditions [18].

DHL interventions, an emerging management tool for stroke survivors, have the potential to improve the quality of care provided in resource-limited settings [14]. When delivering digital interventions, clinicians must assess the patient's level of DHL [18]. However, there is no condition-specific instrument that measures DHL specific to stroke. The purpose of this study was to develop and validate a DHL questionnaire specifically for Chinese patients with stroke, with the aim of providing scientific guidance and support for digital health interventions for patients with stroke.

Methods

Study Design

Overview

The DHL Questionnaire for Stroke Survivors was developed using a sequential, exploratory, mixed methods approach (Figure 1), and the study period ran from September 15, 2023, to April 28, 2024. The design involves an initial qualitative phase aimed at generating the item pool and a subsequent second phase aimed at testing items using quantitative techniques (psychometric evaluation).

XSL•FO
RenderX

Figure 1. Item generation and testing process. EFA: exploratory factor analysis.

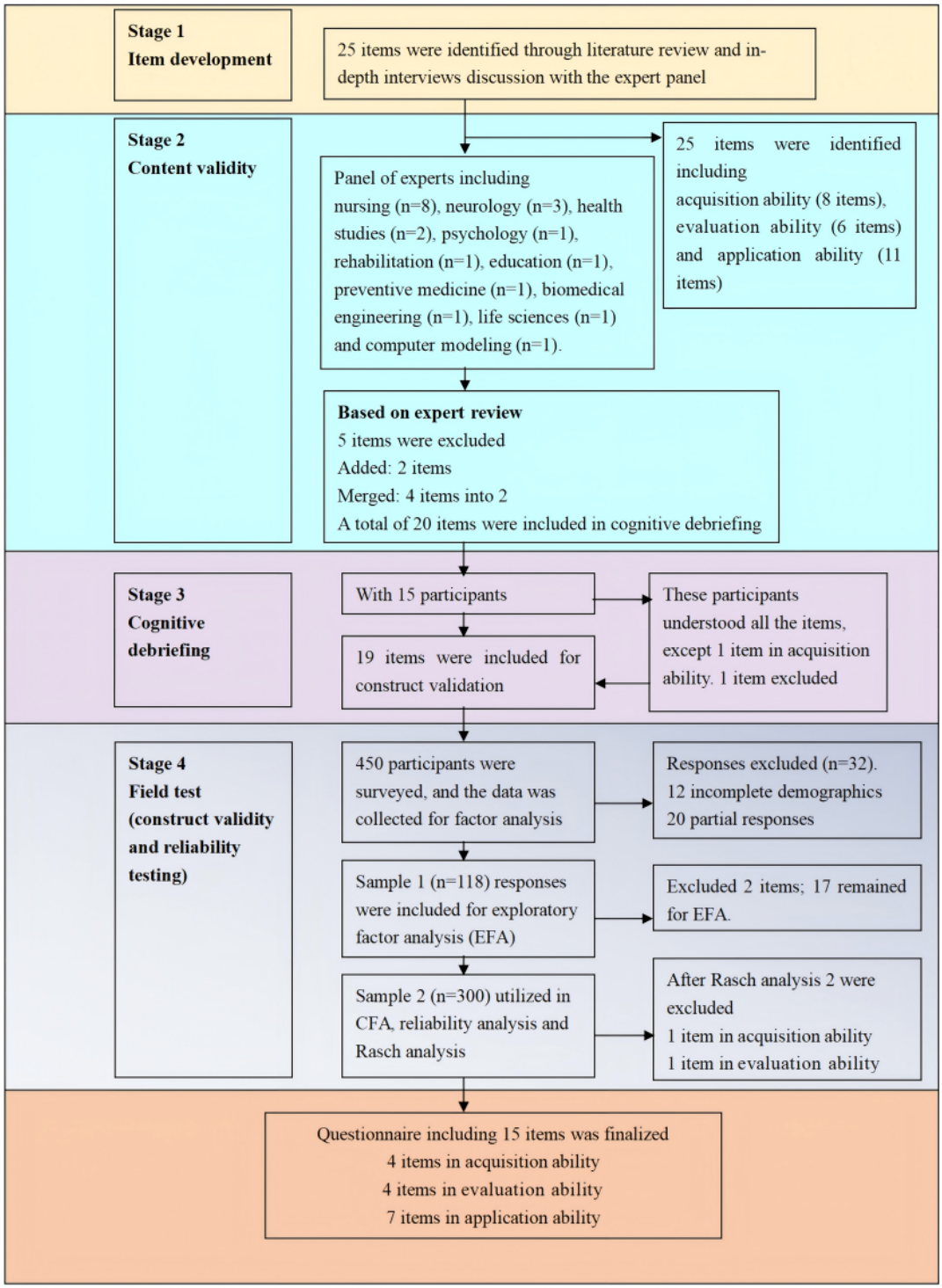

Participants

We used convenience sampling, and all participants gave informed consent at the beginning. Following the survey, professional researchers conducted in-person examinations with participants. Content validation was conducted among stroke survivors recruited from the Department of Neurology and Department of Neurosurgery at the Fifth Affiliated Hospital of Zunyi Medical University, as well as among experts in the field of neurology. The inclusion criteria were defined as being aged 18 years or older, owning a smartphone, and having the ability to read and understand Chinese. Individuals with mental disorders were excluded from the study.

The field test population included survivors with self-reported strokes. Participants were aged 18 years or older and could self-complete the DHL Questionnaire for Stroke Survivors in Chinese. A minimum sample size of 250 respondents was prespecified for field testing. In Rasch analysis, a sample of

XSL•FO
RenderX

250 respondents provides 99% confidence that item calibrations and person measures are stable within a SD of 0.50 logits [19]. Psychometric procedures were conducted to determine the final items and cutoff scores of the DHL Questionnaire for Stroke Survivors.

In this research, the determination of sample size was guided by principles for conducting exploratory factor analysis (EFA). Academic literature [20] offers various heuristics for sample size estimation in EFA, with absolute minimums ranging from 100 to 250 participants [21]. Some sources suggest larger samples of 300, 500, or more [22,23]. A ratio-based approach is often recommended, where the number of participants (N) to the number of variables (p) ideally should be 5:1, meaning 5 participants for each variable [24].

For both EFA and confirmatory factor analysis (CFA), 2 separate, independent samples are necessary. The minimum required sample size for this study was calculated to be 228 participants, considering a 20% allowance for missing data. It is also noted that a sample size ranging from 30 to 200 is considered sufficient for conducting a Rasch analysis [25].

Conceptual Framework Definition

The lily model, grounded in social cognitive and self-efficacy theories, prioritizes enhancing capabilities and confidence over direct skill measurement to foster behavioral change and skill development [9]. We adopted the lily model to identify 3 key domains of eHealth literacy: information acquisition, evaluation, and application abilities. At its heart are 6 core skills (or literacies): traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy. Using the metaphor of a lily, the petals (literacies) feed the pistil (eHealth literacy), and yet the pistil overlaps the petals, tying them together. The lily model of eHealth literacy is used, because it is comprised of multiple literacy types, including an outline of a set of fundamental skills patients require to derive direct benefits from eHealth. A profile of each literacy type with examples of the problems patient-clients might present is provided along with a resource list to aid health practitioners in supporting literacy improvement with their patient-clients across each domain.

Item Generation

For item generation during the development of the new scale, it was important to pool all attributes reflecting the construct being measured. A literature review was used as the source in this study. For the comprehensive literature review, a matrix table was constructed based on the abovementioned DHL conceptualization. The first column on the left of the matrix was dedicated to the posited abilities and skills (acquire, evaluate, and apply). Beginning with the second column, the matrix incorporated specific items. From the third column onward, relevant references were systematically listed.

Content Validation

Content validation was conducted through a web-based expert survey designed to explore the relevance and clarity of the items and to identify items relevant to clinical concerns about DHL. The Item-Level Content Validity Index value for each item was calculated based on the proportion of expert ratings of the item's relevance (modified κ value). Specifically, a modified κ value in the range of 0.4-0.59 indicates fair content validity, 0.60-0.74 indicates good, and ≥0.74 indicates excellent [26]. We relied on the content validity index values of the items to identify those candidates that required further in-depth exploration of their relevance and comprehensibility with stroke survivors.

Cognitive Debriefing

Cognitive debriefing interviews with stroke survivors were conducted face-to-face with the aim of assessing the relevance, comprehensiveness, comprehensibility, and acceptability of the DHL Questionnaire for Stroke Survivors. Verbatim transcripts of the interviews were qualitatively analyzed using thematic analysis. These stroke survivors were also invited to point out other DHL components that may have been omitted from the initial item pool.

Field Testing

A convenience sample of 418 participants was recruited from the Departments of Neurology and Neurosurgery of the Fifth Affiliated Hospital of Zunyi Medical University in China from December 1, 2023, to April 28, 2024.

Test- Retest Reliability

Of the stroke survivors who completed the study survey, 20 were asked to complete the DHL Questionnaire for Stroke Survivors again 2 weeks later.

Measures

The eHEALS, a validated eHealth literacy measure, was used as a comparative tool to evaluate the structural validity of the DHL Questionnaire for Stroke Survivors using a hypothesis-testing method [15]. The eHEALS consists of 10 items scored on a 5-point Likert scale [15]. The eHEALS validation study showed strong internal consistency, with a Cronbach α of .88 and a test-retest dependability intraclass correlation value of 0.49 [15]. Construct validity analysis identified a single component that explained 56% of the variation [15]. The Chinese version of the eHEALS was validated using a sample of 110 high school students [27]. The Cronbach α coefficient was found to be 0.913, indicating excellent internal consistency. Factor analysis revealed loading coefficients ranging from 0.692 to 0.869. In this study, Cronbach α for the instrument was 0.93.

Statistical Analyses

STATA (version 18; StataCorp LLC) was used to clean and prepare the data and for descriptive data analysis. We used SPSS (version 29; IBM Corp) and Amos (version 24.0; IBM Corp) for EFA and CFA. For cross-validation of structural validity, simple random sampling randomly divided the total sample into 2 subsamples. One subsample (subsample 1: n=118) was used for EFA, while the other (subsample 2: n=300) was used for CFA.

In order to determine the applicability of the subsample 1 data to the EFA, the Kaiser-Meyer-Olkin (KMO) test of sample fit and the Bartlett test of sphericity were used to determine the suitability of each category for factor analysis. In EFA, principal

XSL•FO
RenderX

component analysis (PCA) was used to identify common underlying factors. Each factor was identified by eigenvalues greater than 1.0 and rotated orthogonally. A KMO value greater than 0.5 and a Bartlett *P* value less than .05 were used to assess data adequacy. CFA of subsample 2 using diagonally weighted least squares estimation. This study assessed the model fit by root mean square error of approximation<0.08, the Tucker-Lewis Index>0.95, and the Comparative Fit Index>0.95 [28].

Rasch analyses were carried out using Winsteps software (version 3.66.0; WREG). We used standard Rasch analysis procedures. First, we analyzed the fit statistics Then, we deleted items whose response patterns varied from model assumptions. Finally, we re-estimated the item parameters until they were stable.

The Rasch model, a unidimensional measurement framework, was used to assess the likelihood of endorsing items based on their difficulty and the respondents' ability. This model assumes local independence and unidimensionality, meaning that all items are reflections of a single underlying construct [29]. Local independence was confirmed by examining item residual correlations, with correlations below 0.20 supporting the assumption [30]. Infit MnSq statistics should fall within the range of 0.5 to 1.5 for effective measurement [31]. A cutoff of Infit MnSq >1.4 with a standardized z score (Z_{Std}) >2 can be used to identify problematic items [31]. To further evaluate the data, a person-item map was also created. An item separation index of >1.5 was considered indicative of sufficient internal consistency [32].

Measurement reliability was evaluated using person separation indices with defined thresholds: 0.67-0.80 (fair), 0.81-0.90 (good), 0.91-0.94 (very good), and >0.94 (excellent) [33]. The strata were classified as follows: 2 (poor), 2-3 (fair), 3-4 (good), 4-5 (very good), and >5 (magnificent). To confirm model fit, we ran repeated Rasch analyses, modifying items to create new scales and assessing person reliability, separation indices, and scale-to-sample targeting.

Ethical Considerations

The Ethics Committee of Zunyi Medical University's Fifth Affiliated Hospital (2023ZH0084) authorized the study. The study adheres to the principles of the Declaration of Helsinki, and data collection, storage, and analysis are compliant with the European Union General Data Protection Regulation. Informed consent was obtained from all participants, who received a small gift. All images in the manuscript are anonymized to protect participant identities.

Results

Item Development

The literature review extracted an initial pool of 25 attributes that filled the cells of the matrix table constructed in this study (Multimedia Appendix 1).

Content Validation

Content validity indices were calculated for each item based on expert ratings of relevance. The Item-Level Content Validity Index scores for these items ranged from 0.40 to 1, with 20 (80%) of the draft items rated as good or excellent.

The experts' review (n=20) of the item pool informed changes to the wording of items to improve clarity (Multimedia Appendix 2).

Cognitive Debriefing

Cognitive debriefing was established through cognitive debriefing interviews with 15 Chinese stroke survivors aged 18 years or older. The participants' ages ranged from 53 to 72 years, with 7 (46%) participants being male and 8 (53%) participants being older than 65 years. The average monthly income for the participants was between CNY ¥3000 and ¥5000. A total of 7 (47%) participants reported living with a spouse, while 11 (73%) participants had been living with their condition for over a year. In terms of daily activities, 13 (87%) participants demonstrated a mild level of dependence, while 2 (13%) participants were moderately dependent. It is worth noting that 3 (20%) participants were experiencing the illness for the first time.

Thematic analysis was used to classify problems with the relevance, comprehensiveness, clarity, and acceptability of the draft items. The key themes and exemplar quotations from the thematic analysis can be found in Multimedia Appendix 3. The draft DHL Questionnaire for Stroke Survivors consisted of 20 items, organized into 3 domains based on a priori theoretical classification. Each domain was constructed as an independent scale. We used a 5-point Likert scale for assessment (1=strongly disagree, 2=disagree, 3=undecided, 4=agree, and 5=strongly agree). Higher scores indicated a higher level of DHL.

Field Testing

Validation Sample Characteristics

A total of 480 questionnaires were distributed, yielding 450 completed responses, resulting in a response rate of 93.75%. Of these, 32 submissions were deemed incomplete and were therefore excluded from the analyses. Consequently, the final sample comprised 418 complete responses, achieving a completion rate of 100%. The ages of the respondents varied from 33 to 90 years, with a mean age of 65.25 (SD 12.02) years. The sample consisted of 167 female respondents, accounting for 40% of the total. Table 1 provides further details.

XSL•FO
RenderX

Table 1. General characteristics of the study participants.

Characteristics	Subsamples, n (%)		
	1[a] (n=118)	2[b] (n=300)	3[c] (n=20)
Sex			
Male	66 (56)	185 (62)	13 (65)
Female	52 (44)	115 (38)	7 (35)
Age group (years)			
18-45	5 (4)	12 (4)	0 (0)
46-65	54 (46)	11 (39)	4 (20)
66-69	12 (10)	36 (12)	3 (15)
70-79	33 (28)	89 (30)	8 (40)
80	14 (12)	47 (15)	5 (25)
Residence			
Urban	67 (57)	161 (54)	7 (35)
Rural	51 (43)	139 (46)	13 (65)
Education			
Elementary and below	47 (40)	109 (36)	6 (30)
Middle school	27 (23)	91 (30)	5 (25)
Secondary or vocational school	32 (27)	68 (23)	4 (20)
University or college and above	12 (10)	32 (11)	5 (25)
Living with children			
Yes	52 (44)	110 (37)	7 (35)
No	66 (56)	190 (63)	13 (65)
Marital status			
Married	90 (76)	218 (73)	15 (75)
Unmarried, divorced, or widowed	28 (24)	82 (27)	5 (25)
Monthly per capita household income (CNY ¥)			
<3000	24 (20)	46 (15)	1 (5)
3000-5000	46 (39)	150 (50)	10 (50)
>5000	48 (41)	104 (35)	9 (45)
Medical cost payment method			
Self-paid	7 (6)	13 (4)	0 (0)
Medical insurance	111 (94)	28 (96)	2 (100)
First-time illness			
Yes	40 (34)	86 (29)	9 (45)
No	78 (66)	21 (71)	11 (55)
Type of stroke			
Hemorrhagic	22 (19)	14 (5)	0 (0)
Ischemic	96 (81)	28 (95)	20 (100)
Duration of illness			
<3 months	31 (26)	92 (31)	9 (45)
3 months to 1 year	29 (25)	15 (50)	10 (50)
>1 year	58 (49)	58 (19)	1 (5)
Self-care ability			

XSL•FO
RenderX

Characteristics	Subsamples, n (%)		
	1[a] (n=118)	2[b] (n=300)	3[c] (n=20)
Mild dependence	75 (64)	22 (76)	17 (85)
Moderate dependence	43 (36)	70 (23)	1 (5)
Severe dependence	0 (0)	4 (1)	2 (10)

[a]Subsample 1: exploratory factor analysis.

[b]Subsample 2: confirmatory and Rasch analysis.

[c]Subsample 3: test-retest reliability.

During the scale refining process, we first used the critical ratio method, sorted scale scores to identify high and low groups, and calculated critical ratio values using *t* tests. Item 19 was eliminated because its critical ratio value fell below the threshold of 3. Furthermore, Spearman rank correlation analysis was performed, resulting in the elimination of item 18 due to a correlation coefficient of less than 0.4. Following these analyses, the scale was simplified to include 17 items.

EFA Examination

Following an EFA on the 17-item subset, the KMO score was 0.901, which was above the acceptable threshold of 0.7, suggesting that the data were appropriate for component analysis. The Bartlett test of sphericity showed significant results (χ^2_{136}=1372.687, P<.001). PCA showed 3 variables with eigenvalues greater than 1, accounting for 69.212% of the variance, showing excellent explanatory power (Table 2). The scree plot confirmed the retention of 3 factors, as the line leveled off after the fourth factor (Multimedia Appendix 4). The scale's construct validity was confirmed by factor loadings over 0.5 and no cross-loadings above 0.5. The items were divided into 3 categories: "acquisition ability," "evaluation ability," and "application ability."

XSL•FO
RenderX

Table 2. Exploratory factor loadings of the digital health literacy questionnaire for stroke survivors (n=118).

Item	Domains		
	Factor 1	Factor 2	Factor 3
Item 1			
I track stroke news and updates on the internet.	0.123	0.798	0.206
Item 2			
I can search online for in-depth stroke information.	0.162	0.705	0.174
Item 3			
I can find needed stroke-related info online.	0.335	0.857	0.170
Item 4			
I can understand online stroke info.	0.308	0.840	0.268
Item 5			
I can gather stroke data from various online platforms.	0.128	0.788	0.223
Item 6			
I can check the accuracy of stroke information with medical experts.	0.180	0.324	0.761
Item 7			
I verify if online stroke info is current.	0.220	0.154	0.76888
Item 8			
I can validate stroke ad claims with health care staff.	0.101	0.280	0.783
Item 9			
I can review expert-provided stroke information.	0.025	0.067	0.794
Item 10			
I can review stroke health info online and in WeChat.	0.156	0.199	0.773
Item 11			
I can discuss health issues during virtual medical sessions.	0.727	0.259	0.221
Item 12			
I can participate in online discussions focused on stroke.	0.642	0.332	0.060
Item 13			
I can identify stroke warning signs online.	0.771	0.148	0.141
Item 14			
I can use online guides to plan my meals.	0.727	–0.027	0.037
Item 15			
I can get medication info online and take meds as experts say.	0.841	0.194	0.051
Item 16			
I can follow online resources for stroke exercises.	0.866	0.251	0.139
Item 17			
I can record my health stats weekly with digital devices.	0.825	0.183	0.277
Eigenvalue (characteristic root)	7.541	2.521	1.704
Variance Contribution (%)	44.357	14.830	10.026
Cumulative variance contribution rate %	44.357	59.186	69.212
Item 18			
I don’t click on pop-up ads without thinking.	—[a]	—	—
Item 19			
I won’t share my own or others’ health information on social media.	—	—	—

XSL•FO
RenderX

[a]Based on the interitem correlation analysis, items 18 and 19 were considered redundant and were deleted before exploratory factor analysis.

CFA Examination

CFA indicated an excellent model fit, with indices (χ^2_4=1.669; root mean square error of approximation=0.047; Tucker-Lewis Index=0.973; Comparative Fit Index=0.977) aligning with conventional criteria for a good fit. Composite reliability estimates varied from 0.861 to 0.932, and average variance extracted scores were between 0.555 and 0.664, indicating convergent validity. Furthermore, the factor loading values, which are shown in Multimedia Appendix 5, varied from 0.616 to 0.937, supporting convergent validity.

Reliability Testing

The value for Cronbach α, as further evidence of internal consistency reliability, was 0.937. The coefficient of test-retest reliability was 0.941, indicating a high degree of consistency in the measurements over time.

Hypothesis-Testing Construct Validity

The correlation coefficient between the DHL Questionnaire for Stroke Survivors and eHEALS was 0.853 (P<.001).

Rasch Analysis

To evaluate the functionality of the rating scale, we reviewed the Winsteps 3.66.0 output tables and item category probability curves for the 17 items.

After optimization of the response scales, we conducted a Rasch analysis to report the finalized characteristics of the DHL questionnaire for stroke survivors. The DHL Questionnaire for Stroke Survivors consists of 15 items. Multimedia Appendix 6 lists the items included. The Chinese version of the DHL Questionnaire for Stroke Survivors is provided in Multimedia Appendix 7. The scale meets the unidimensional criteria of the Rasch model. The first residual logarithm of the PCA of residuals ranged between 1.4 and 1.8. Item mean squared error values ranged from 0.71 to 1.38 logits, and equipment mean squared error values ranged from 0.67 to 1.75 logits. Items 4 and 7 showed ambiguity in the model. Ultimately, 15 items met the acceptable item fit criteria for the Rasch model.

Four items showed local item dependency, with residual correlation values >0.4 (range 0.44-0.59). In practical terms, a degree of local dependency is always observed in empirical data; therefore, it is necessary to consider the implications for content validity before proceeding with item removal [34]. After a qualitative appraisal of the 4 dependent pairs, we retained all items to ensure a comprehensive assessment of DHL in stroke survivors.

Last, we evaluated item separation, person reliability and separation, and internal consistency reliability. The scale had mean person ability values within a SD of 1.0 logits of mean item difficulty. Person ability ranged from −3.2 to 2.8 logits. Multimedia Appendix 8 shows the item person maps for the DHL Questionnaire for Stroke Survivors. The item separation index was 3.81 (reliability 0.94) and the individual separation index was 2.91 (reliability 0.89).

Discussion

Principal Findings

This study aimed to develop and validate the DHL Questionnaire for Stroke Survivors, comprising of the three components (1) acquisition ability, (2) evaluation ability, (3) and application ability, with a total of 15 items. The validation process included assessing the questionnaire's content validity, construct validity, internal consistency reliability, test-retest reliability, and Rasch analyses. The findings demonstrate the robust psychometric properties of the DHL Questionnaire for Stroke Survivors within the target population in China. Unlike other DHL-related scales [15,16], this study specifically addresses stroke-related competencies in acquisition, evaluation, and application, emphasizing practical applications and interactions with health care providers directly pertinent to stroke management.

The eHEALS developed by Korean scholars for patients with diabetes with comorbid chronic diseases assesses "cognitive behavior" and "digital communication ability," emphasizing the role of "cognition" as a critical component of information health literacy for patients with chronic diseases [18]. Similarly, "accessibility" in this study is closely linked to cognition, as evidenced by behaviors such as tracking stroke news and updates on the internet, searching on the web for in-depth stroke information, and gathering stroke data from various web-based platforms. These actions highlight the cognitive abilities required to identify and filter relevant information, reflecting the mental processes essential for individuals to navigate and process stroke-related information in a digital environment.

In this study, CFA and EFA were performed on 2 separate samples to evaluate the structural validity of the DHL Questionnaire for Stroke Survivors [35]. The CFA results validated the model developed through EFA. Furthermore, the CFA findings demonstrated that each item corresponded to a distinct underlying construct, and the associated item errors were uncorrelated, indicating the reliability of the CFA results [35].

In this study, a χ^2_4<3 was used as the criterion for goodness of fit [36]. The results indicated a χ^2_4=1.669, demonstrating a good model fit. Additionally, the composite reliability of all 3 factors exceeded the criterion of 0.7 [36], ranging from 0.861 to 0.932, which confirmed the good internal consistency of the items within each dimension. Furthermore, the square root of the average variance extracted for factor 1, factor 2, and factor 3 ranged from 0.555 to 0.664, surpassing the correlation values on the nondiagonal line of the discriminant validity scale, thereby further validating the model's discriminant validity. The internal consistency of the final questionnaire was evaluated using Cronbach α, which yielded an overall value of 0.937, demonstrating the reliability and acceptability of the newly developed scale [37]. In terms of overall item performance, the individual reliability was 0.89, and the item reliability was 0.94, reflecting consistency and stability in the results. The person-item separation index was 2.91, indicating a high degree of differentiation between participants' ability levels, while the

XSL•FO
RenderX

item separation index was 3.81, signifying a high degree of difficulty differentiation among the items. The weighted mean square fit statistics (Infit MNSITEM) were 1.06 for persons and 1.07 for items, and the unweighted mean square fit statistics were 1.00 and 1.07, respectively. These values, being close to the ideal value of 1, confirm that the model fits the empirical data well.

The application of the DHL Questionnaire for Stroke Survivors represents a significant advancement in personalized health care by enabling health care providers to tailor interventions according to individual DHL levels. This approach enhances treatment adherence by equipping patients with resources and support that match their digital literacy capabilities. For example, patients with lower literacy may benefit from traditional, face-to-face interactions, whereas those with higher literacy can effectively leverage digital tools and resources [38]. Furthermore, personalized health care facilitated by the DHL Questionnaire for Stroke Survivors fosters patient engagement by ensuring that health information is both accessible and comprehensible, thereby empowering patients to take a more active role in their recovery process [39]. This not only improves the overall efficacy of rehabilitation programs but also optimizes the allocation of health care resources by targeting interventions to meet the specific needs of each patient [40].

Limitations

This study has several limitations. First, the representativeness of the sample is limited, as the respondents were predominantly older individuals with multiple comorbidities. Second, essential items and content for assessing DHL may have been overlooked, given that DHL encompasses a wide range of skills, including complex technology-related health literacy in the digital era. Approaches to DHL have evolved significantly over the past decade and will likely continue to do so. The criterion validity of the DHL Questionnaire for Stroke Survivors was not tested due to the rapidly evolving nature of digital literacy and the absence of a gold standard for assessment. Consequently, the questionnaire should be updated to reflect advancements in this field. Third, the DHL Questionnaire for Stroke Survivors relies on self-reporting, which limits its ability to accurately capture objective proficiency. Future studies should enhance reliability by measuring the temporal stability of individual responses. Finally, the DHL Questionnaire for Stroke Survivors has only been psychometrically tested in Chinese, necessitating cross-national and cross-language validations to confirm its cultural applicability.

Conclusion

With the rapid development of the internet and digital technology, DHL plays an increasingly important role in public health management. For patients with stroke, DHL not only affects their ability to access and understand health information but also directly influences their self-management and quality of life during rehabilitation. Therefore, it is important to develop high-quality measurement tools to assess the level of DHL among patients with stroke in order to better meet the rehabilitation needs of this population. To this end, we developed the DHL Questionnaire for Stroke Survivors. This questionnaire has good psychometric properties and is important for the development of effective health education and interventions.

Acknowledgments

We extend our appreciation to the hospitals that collaborated in the trial, and we are grateful to the entire team involved for their commitment, which was instrumental in the success of this research endeavor. This study was funded by the Implementation Science-Driven Agent Simulation to Develop a Psychological and Behavioral Intervention Protocol for Stroke Survivors Based on a Stepped Wedge Randomized Controlled Trial (82260281); the Science and Technology Fund Project of the Guizhou Provincial Health Commission (gzwkj2023-588); the Scientific Study on the Construction of Shared Medical Model and Optimization Mechanism for Patients with Stroke Depression in Low Resource Area; and Guizhou Science and Technology Plan Project, Guizhou Science and Technology Cooperation (Qiankehe) Foundation (ZK [2024] key Project 069).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Concept coverage matrix for digital health literacy based on a comprehensive literature review.
[DOCX File , 21 KB - jmir_v27i1e64591_app1.docx]

Multimedia Appendix 2
Item-Level Content Validity Index (I-CVI) assessment for each item in the Digital Health Literacy (DHL) Questionnaire for Stroke Survivors.
[DOCX File , 30 KB - jmir_v27i1e64591_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Themes and exemplar quotations from content validation of the Digital Health Literacy (DHL) Questionnaire for Stroke Survivors item pool: cognitive debriefing (n=15) and expert survey (n=20).
[DOCX File , 13 KB - jmir_v27i1e64591_app3.docx]

Multimedia Appendix 4
Scree plot analysis from exploratory factor analysis of the Digital Health Literacy (DHL) Questionnaire for Stroke Survivors.
[DOCX File , 24 KB - jmir_v27i1e64591_app4.docx]

Multimedia Appendix 5
Findings of confirmatory factor analysis for the Digital Health Literacy (DHL) Questionnaire for Stroke Survivors.
[DOCX File , 174 KB - jmir_v27i1e64591_app5.docx]

Multimedia Appendix 6
Final Digital Health Literacy (DHL) Questionnaire for Stroke Survivors—ready for use.
[DOCX File , 17 KB - jmir_v27i1e64591_app6.docx]

Multimedia Appendix 7
Chinese version of the final Digital Health Literacy (DHL) Questionnaire for Stroke Survivors.
[DOCX File , 16 KB - jmir_v27i1e64591_app7.docx]

Multimedia Appendix 8
Person-item map for the Digital Health Literacy (DHL) Questionnaire for Stroke Survivors.
[DOCX File , 37 KB - jmir_v27i1e64591_app8.docx]

References

1. Tu WJ, Wang LD, Special Writing Group of China Stroke Surveillance Report. China stroke surveillance report 2021. Mil Med Res 2023;10(1):33 [FREE Full text] [doi: 10.1186/s40779-023-00463-x] [Medline: 37468952]
2. Zhang W, Zhou W, Ye M, Gao Y, Zhou L. Family resilience of stroke survivors within 6 months after a first-episode stroke: a longitudinal study. Front Psychiatry 2022;13:968933 [FREE Full text] [doi: 10.3389/fpsyt.2022.968933] [Medline: 36311520]
3. Lee KE, Choi M, Jeoung B. Effectiveness of rehabilitation exercise in improving physical function of stroke patients: a systematic review. Int J Environ Res Public Health 2022;19(19):12739 [FREE Full text] [doi: 10.3390/ijerph191912739] [Medline: 36232038]
4. Naqvi IA, Montiel TC, Bittar Y, Hunter N, Okpala M, Johnson C, et al. Internet access and usage among stroke survivors and their informal caregivers: cross-sectional study. JMIR Form Res 2021;5(3):e25123 [FREE Full text] [doi: 10.2196/25123] [Medline: 33683206]
5. Li X, Ren S, Gu F. Medical internet of things to realize elderly stroke prevention and nursing management. J Healthc Eng 2021;2021:9989602 [FREE Full text] [doi: 10.1155/2021/9989602] [Medline: 34326980]
6. Sakakibara BM, Lear SA, Barr SI, Goldsmith CH, Schneeberg A, Silverberg ND, et al. Telehealth coaching to improve self-management for secondary prevention after stroke: a randomized controlled trial of stroke coach. Int J Stroke 2022;17(4):455-464 [FREE Full text] [doi: 10.1177/17474930211017699] [Medline: 33949270]
7. Marwaa MN, Guidetti S, Ytterberg C, Kristensen HK. Use of mobile/tablet and web-based applications to support rehabilitation after stroke: a scoping review. J Rehabil Med 2022;54:jrm00269 [FREE Full text] [doi: 10.2340/jrm.v54.452] [Medline: 35174871]
8. Xiaoyue W. E-health Literacy Status and its Correlation with Quality of Life in Young and Middle-aged Stroke Patients.. master, Shandong University of Chinese Medicine. Beijing, China: Tongfang Knowledge Network; 2024.
9. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
10. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
11. van Kessel R, Wong BLH, Clemens T, Brand H. Digital health literacy as a super determinant of health: more than simply the sum of its parts. Internet Interv 2022;27:100500 [FREE Full text] [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
12. Estrela M, Semedo G, Roque F, Ferreira PL, Herdeiro MT. Sociodemographic determinants of digital health literacy: a systematic review and meta-analysis. Int J Med Inform 2023;177:105124 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105124] [Medline: 37329766]

XSL•FO
RenderX

13. Milanti A, Chan DNS, Parut AA, So WKW. Determinants and outcomes of eHealth literacy in healthy adults: a systematic review. PLoS One 2023;18(10):e0291229 [FREE Full text] [doi: 10.1371/journal.pone.0291229] [Medline: 37792773]
14. Vu MTT, Ho HQ, Lin GH. eHealth interventions of health literacy for stroke survivors: systematic review and meta-analysis. Public Health Nurs 2025;42(1):516-523. [doi: 10.1111/phn.13432] [Medline: 39344209]
15. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
16. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
17. Xie L, Mo PKH. Comparison of eHealth literacy scale (eHEALS) and digital health literacy instrument (DHLI) in assessing electronic health literacy in Chinese older adults: a mixed-methods approach. Int J Environ Res Public Health 2023;20(4):3293 [FREE Full text] [doi: 10.3390/ijerph20043293] [Medline: 36833987]
18. Lee EH, Lee YW, Lee KW, Kim HJ, Hong S, Kim SH, et al. Development and psychometric evaluation of a new brief scale to measure eHealth literacy in people with type 2 diabetes. BMC Nurs 2022;21(1):297 [FREE Full text] [doi: 10.1186/s12912-022-01062-2] [Medline: 36333750]
19. Sample size and item calibration [or person measure] stability. Rasch. URL: https://www.rasch.org/rmt/rmt74m.htm [accessed 2025-02-27]
20. Kyriazos TA. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology 2018;09(08):2207-2230. [doi: 10.4236/psych.2018.98126]
21. Erdos G. The scientific use of factor analysis in behavioral and life sciences. Biol Psychol 1979;8(4):317-318. [doi: 10.1016/0301-0511(79)90014-0]
22. Gudgeon AC, Comrey AL, Lee HB. A first course in factor analysis. J R Stat Soc Ser D (Stat) 1994;43(2):332. [doi: 10.2307/2348352]
23. Tabachnick B, Fidell L. Using multivariate statistics. Semantic Scholar. 1983. URL: https://www.semanticscholar.org/paper/Using-Multivariate-Statistics-Tabachnick-Fidell/ac0a41e75f45bbbb7e7a41d8d38b92bf47386ed6 [accessed 2024-07-13]
24. Johanson GA, Brooks GP. Initial scale development: sample size for pilot studies. Educ Psychol Meas 2009;70(3):394-400. [doi: 10.1177/0013164409355692]
25. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994:328.
26. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? appraisal and recommendations. Res Nurs Health 2007;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
27. Guo S, Yu X, Sun Y, Nie D, Li X, Wang L. Exploring the sinicization and applicability of the eHEALS health literacy scale. Chin J Health Educ 2013;29(2):123-108. [doi: 10.16168/j.cnki.issn.1002-9982.2013.02.019]
28. Wolf MG, McNeish D. dynamic : an R package for deriving dynamic fit index cutoffs for factor analysis. Multivariate Behav Res 2023;58(1):189-194. [doi: 10.1080/00273171.2022.2163476] [Medline: 36787513]
29. Tennant A, McKenna SP, Hagell P. Application of rasch analysis in the development and application of quality of life instruments. Value Health 2004;7 Suppl 1:S22-S26 [FREE Full text] [doi: 10.1111/j.1524-4733.2004.7s106.x] [Medline: 15367240]
30. Lambert S, McCusker J, Belzile E, Yaffe M, Ihejirika C, Richardson J, et al. Using confirmatory factor analysis and Rasch analysis to examine the dimensionality of the patient assessment of care for chronic illness care (PACIC). Qual Life Res 2021;30(5):1503-1512. [doi: 10.1007/s11136-020-02750-9] [Medline: 33492589]
31. Espejo MR. Applying the rasch model: fundamental measurement in the human sciences. J R Stat Soc Ser A 2004;167(2):375-3780. [doi: 10.1111/j.1467-985x.2004.t01-2-.x]
32. Tennant A, Conaghan PG. The rasch measurement model in rheumatology: what is it and why use it? when should it be applied, and what should one look for in a rasch paper? Arthritis Rheum 2007;57(8):1358-1362 [FREE Full text] [doi: 10.1002/art.23108] [Medline: 18050173]
33. Rasch measurement and the R statistics environment. Rasch. 2007. URL: https://www.rasch.org/rmt/rmt211a.htm [accessed 2025-02-27]
34. Local dependency and rasch measures. Rasch. URL: https://www.rasch.org/rmt/rmt213b.htm [accessed 2024-07-11]
35. Roos JM, Bauldry S. Confirmatory Factor Analysis. Washington DC,: SAGE Publications, Inc; 2022.
36. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 8th ed. Boston, MA: Cengage; 2018.
37. Nunnally JC, Bernstein IH. American educational research journal. In: Psychometric Theory. 3rd ed. New York, NY: McGraw-Hill; 1994.
38. Brørs G, Larsen MH, Hølvold LB, Wahl AK. eHealth literacy among hospital health care providers: a systematic review. BMC Health Serv Res 2023;23(1):1144 [FREE Full text] [doi: 10.1186/s12913-023-10103-8] [Medline: 37875882]
39. Chén OY, Roberts B. Personalized health care and public health in the digital age. Front Digit Health 2021;3:595704 [FREE Full text] [doi: 10.3389/fdgth.2021.595704] [Medline: 34713084]
40. Ban S, Kim Y, Seomun G. Digital health literacy: a concept analysis. Digit Health 2024;10 [FREE Full text] [doi: 10.1177/20552076241287894] [Medline: 39381807]

XSL•FO
RenderX

Abbreviations

CFA: confirmatory factor analysis
DHL: digital health literacy
EFA: exploratory factor analysis
eHEALS: eHealth Literacy Scale
KMO: Kaiser-Meyer-Olkin
PCA: principal component analysis

Edited by A Coristine; submitted 21.07.24; peer-reviewed by R Chen, N Godwin; comments to author 20.12.24; revised version received 10.01.25; accepted 18.02.25; published 25.03.25.

Please cite as:
Ye Q, Wang W, Zeng X, Kuang Y, Geng B, Zhou S, Liu N
Development and Validation of the Digital Health Literacy Questionnaire for Stroke Survivors: Exploratory Sequential Mixed Methods Study
J Med Internet Res 2025;27:e64591
URL: https://www.jmir.org/2025/1/e64591
doi:10.2196/64591
PMID:

©Qin Ye, Wei Wang, Xuan Zeng, Yuxian Kuang, Bingbing Geng, Song Zhou, Ning Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Global Use, Adaptation, and Sharing of Massive Open Online Courses for Emergency Health on the OpenWHO Platform: Survey Study

Jamie Sewan Johnston[1], PhD; Nadine Ann Skinner[1], PhD; Anna Tokar[2], PhD; Elham Arabi[2], PhD; Ngouille Yabsa Ndiaye[2], MPH; Matthew Charles Strehlow[3], MD; Heini Utunen[2], PhD

[1]Stanford Center for Health Education, Stanford University, Stanford, CA, United States

[2]Learning and Capacity Development Unit, Health Emergencies Programme, World Health Organization, Geneva, Switzerland

[3]Department of Emergency Medicine, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Jamie Sewan Johnston, PhD
Stanford Center for Health Education
Stanford University
408 Panama Mall
Stanford, CA, 94305
United States
Phone: 1 650 647 0501
Email: jamiejs@stanford.edu

Abstract

Background: The COVID-19 pandemic demonstrated the global need for accessible content to rapidly train health care workers during health emergencies. The massive open access online course (MOOC) format is a broadly embraced strategy for widespread dissemination of trainings. Yet, barriers associated with technology access, language, and cultural context limit the use of MOOCs, particularly in lower-resource communities. There is tremendous potential for MOOC developers to increase the global scale and contextualization of learning; however, at present, few studies examine the adaptation and sharing of health MOOCs to address these challenges.

Objective: The World Health Organization's Health Emergencies Programme Learning and Capacity Development Unit and the Stanford Center for Health Education collaborated to survey learners from 4 emergency health MOOCs on the OpenWHO platform to examine differences in course use by World Bank country income classification across three dimensions: (1) how health education MOOCs are used and shared, (2) how health workers adapt MOOC content to meet local training and information needs, and (3) how content adaptations help frontline health workers overcome barriers to using MOOCs.

Methods: This study draws upon two sources of data: (1) course enrollment data collected from the 4 emergency health MOOCs (N=96,395) and (2) survey data collected from learners who participated in at least 1 of the 4 MOOCs (N=926). Descriptive statistics are used to summarize learner characteristics. Differences in enrollment, sharing, and adaptation by country income classification are examined using Pearson chi-square test.

Results: Of the enrollees who indicated their country of residence, half were from lower-middle-income countries (LMICs; 43,168/85,882, 50%) and another 9% (7146/85,882) from low-income countries. The majority of all respondents shared content (819/926, 88%) and used content in official trainings (563/926, 61%). Respondents were more likely to share and use content for trainings in LMICs than in high-income countries (91% vs 81%; *P*=.001). Learners in LMICs also shared content with more people on average compared with high-income country learners although the difference is not statistically significant (9.48 vs 6.73 people; *P*=.084). Compared with learners in high-income countries, learners in LMICs were more likely to adapt materials to distribute via offline formats or technologies, such as WhatsApp or text message (31% vs 8%; *P*<.001); to address cultural, linguistic, or other contextual needs (20% vs 12%; *P*=.076); and to meet local guidelines (20% vs 9%; *P*=.010). Learners in LMICs indicated greater accessibility challenges due to technological and linguistic barriers.

Conclusions: Learners commonly share content from MOOCs about public health emergencies; this is especially true in low-income countries and LMICs. However, content is often adapted and shared via alternative formats. Our findings identify a critical opportunity to improve MOOC design and dramatically scale the impact of MOOCs to better meet diverse global needs.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e52591) doi:10.2196/52591

KEYWORDS

MOOCs; online learning; global health education; digital health; health worker training; health emergencies; outbreak; COVID-19

Introduction

Overview

First launched in 2017, the World Health Organization's (WHO) Health Emergencies Programme's learning platform, the OpenWHO, uses a massive open access online course (MOOC) format. The platform rapidly expanded during the COVID-19 pandemic, from serving approximately 160,000 learners at the pandemic's start [1] to serving more than 7.5 million learners in 194 countries and 65 languages by the end of 2022, with 46 courses specifically on COVID-19 and more than 85 additional emergency health course topics [1,2]. The MOOCs on the platform are designed to improve equitable access to emergency health topics according to learners' needs and preferred languages [3]. Even with the significant expansion in topics and utilization, technical, cultural, and linguistic barriers limit dissemination of the lifesaving knowledge provided through MOOCs. Preliminary research indicated that local health workers and institutions adapt and share digital content to increase knowledge dissemination [4-6]. This study evaluates how MOOC content on the OpenWHO platform is used, shared, and adapted to better understand the ability of MOOCs to reach frontline health workers during public health emergencies.

Background

Over the past decade, digital education tools have been increasingly incorporated into the training and education of frontline health workers [7-12]. However, the COVID-19 pandemic demonstrated a critical global need to expand easily accessible and trustworthy open-source content for rapid frontline health worker training during health crises. The MOOC format, embraced by the WHO [3,13-16], governments [17,18], and academic institutions worldwide [19,20], is one method for the widespread dissemination of emergency training for frontline health workers.

The MOOC format has been successfully used in public health crises to disseminate information to the public, policy makers, and other stakeholders [21-23] and provide health workers with training and knowledge [15,17,24-30]. While there are different varieties of MOOCs, content is often offered online for learners via readings, wikis, presentation slides, and short videos [31]. MOOCs also typically include learner assessments and interactive components, such as automated quizzes, peer-marked assessments, or online discussion forums, and may be self-paced and asynchronous [31].

While the MOOC format has seen success in public health crises, enrollment in and completion of MOOCs in many low-income and lower-middle-income countries (LICs/LMICs) are disproportionately lower than those in high-income and upper-middle-income countries (HICs/UMICs) [32]. MOOC enrollment and completion vary by region but are substantially lower in countries with limited investment in higher education systems and technology infrastructure [33,34]. Significant technical barriers prevent enrollment and completion of MOOCs in resource-limited and remote settings [6,31-35]. Challenges concerning digital literacy, weak technical support, web access, and broadband cellular coverage persist, limiting the reach and effectiveness of digital health education tools in these regions [27,31,33-38]. Infrastructure barriers can cause challenges when accessing MOOCs, including downloading materials, engaging in interactive elements, and viewing high-resolution videos [31].

There may also be significant cultural and linguistic barriers that prevent MOOC engagement [39-47]. Critics have argued that MOOC content often needs more cultural context for successful curriculum design [31,37,39-48]. Traditional MOOC content has been based on the academic experiences, pedagogical ideas, design concepts, and literature of the Global North [31,39,45,46]. Research has shown that including various learning, assessment, and accreditation activities in MOOC design facilitates learners' progress in one context while inhibiting learners in other contexts [31,40-47]. In particular, Indigenous groups and knowledge have been noted as lacking in MOOC design and research, which has been seen as a limitation in ensuring the relevance of MOOCs for Indigenous populations [47].

In addition to cultural relevance, MOOCs should also be linguistically relevant [3,16] in languages where learners seek out health information [49] and at literacy levels that can be understood by diverse audiences [50,51]. Communities, especially vulnerable communities, have indicated mistrust in digital health materials that are not culturally and linguistically relevant [39,50,52]. Offering MOOCs in multiple languages successfully enables increased access to diverse regional learners with low levels of previous education [2,43], and cultural and linguistic translations can make MOOCs and digital content more relevant for learners [43] and support their ability to learn the material [53].

Culturally and linguistically relevant content has increased positive health beliefs and behaviors [54], which indicates a need for context-specific adaptations for global MOOCs to support learner needs. Yet, initially, most MOOCs were developed in English with more MOOCs now being created in multiple languages [31]. However, research on infectious disease and health MOOCs has found limited resources in languages other than English [20,55,56]. In addition, when MOOCs have been translated from English into other languages, translations include less accessible word choices, especially for low-literacy populations [43,44]. These translations may also reduce the provided health content's complexity [56].

Proponents of open-source digital education have argued that learners can adapt and share well-designed materials to overcome access barriers, leading to what has been termed a "learning multiplier effect" that describes the transformation and sharing of web-based resources into adapted formats to suit learners' needs and local context beyond what can be tracked

XSL•FO
RenderX

by platform analytics [6]. Evidence suggests that during health crises, health workers are motivated to enroll in MOOCs to improve practice [57]. The need for trustworthy information also drives health workers to share MOOC content through personal and professional networks [5,6]. Learners have shared MOOC materials via their networks by downloading materials, providing links to resources, and telling others in their networks about the course and course content [5,57]. They also adapted and integrated MOOC resources into formal training, making MOOCs accessible to even larger audiences [4-6]. Adaptations included efforts to overcome technical barriers, such as downloading or making content available via text message [4-6]. Other adaptations included addressing cultural or local context, such as by adding local health regulations or addressing linguistic relevance with local translations [5,6].

Intervention

The OpenWHO platform is an open access, low-bandwidth requirement online learning platform first launched following the Ebola West Africa outbreak in 2017 [3]. The platform aims to provide lifesaving online knowledge based on the latest WHO technical advice to frontline responders, policy makers, and the public. The platform hosts MOOCs on 33 different infectious diseases in 65 languages aiming to disseminate critical learning materials for any type of health event and outbreak. There are a total of 46 MOOCs related to COVID-19, 5 on Ebola, and a range of courses on different epidemic and pandemic-prone diseases [16].

This study examines 2 MOOCs focused on COVID-19 and 2 on other epidemic and pandemic-prone diseases, Ebola and rabies. These 4 courses were selected because they had large enrollments of health professionals. The 2 COVID-19 courses are “Clinical Management of Patients With COVID-19: General Considerations” and “Clinical Management of Patients With COVID-19: Initial Approach to the Acutely Ill Patient.” The other 2 health emergency courses selected for the study include one of the first courses created for the OpenWHO platform, “Ebola: Clinical Management of Ebola Virus Disease” (developed in response to the 2018 Ebola outbreak in the Équateur Province of the Democratic Republic of the Congo) and a relatively new course, “Rabies and One Health: From Basics to Cross-sectoral Action to Stop Human Rabies Deaths.” The study includes both COVID-19 and other health emergency courses to understand better which course-sharing experiences and adaptations may have been unique to the COVID-19 pandemic context. All of the courses are available in English and 2-4 additional languages. A full description of the MOOCs can be seen in Table 1.

XSL•FO
RenderX

Table 1. Course descriptions.

Course title and description	Languages	Date launched	Course duration	Enrolled learners	Course completion
Clinical Management of Patients With COVID-19: General Considerations: this course gives background on the pandemic, discusses facility operations, and addresses COVID-19 pandemic preparedness at all levels of health care provision, including surge planning, infection prevention and control, palliative care, and transfer of the acutely ill patient. It also discusses ethical issues arising during COVID-19 care.	Albanian, Dutch, English, Indonesian, Kazakh, and Macedonian	October 22, 2020	3 hours	38,941	• Record of achievement: Earning at least 80% of the maximum points from all graded assignments. • Confirmation of participation: Completing at least 80% of the course material.
Clinical Management of Patients With COVID-19: Initial Approach to the Acutely Ill Patient. This course is designed to prepare and support health providers as they provide emergency care to seriously ill COVID-19 patients. The course includes a systematic approach via the WHO/ICRC[a] Basic Emergency Care course content, covering screening and triage, the integrated interagency triage tool, resuscitation area designation, the WHO medical emergency checklist, the ABCDE approach to the acutely ill patient, and the approach to the patient with breathing difficulty, shock, or altered mental status.	English, Portuguese, Russian, Somali, and Spanish	May 5, 2021	6 hours	24,036	• Record of achievement: Earning at least 80% of the maximum points from all graded assignments. • Confirmation of participation: Completing at least 80% of the course material.
Ebola: Clinical Management of Ebola Virus Disease: this comprehensive intermediate-level course is for clinicians caring for patients with suspected or confirmed Ebola virus disease. The course covers screening and triage, infection prevention and control, laboratory diagnostics, organization and clinical care of patients in the Ebola Treatment Centre, and investigational therapeutic agents.	English, French, and Kiswahili	2018	6 hours	23,803	• No certificate available at this time.
Rabies and One Health: From Basics to Cross-Sectoral Action to Stop Human Rabies Deaths: this course provides a general introduction to rabies, and the One Health approach currently taken to prevent it including information on the “Zero by 30” rabies elimination strategy, preventing rabies in people and dogs, awareness and community empowerment, and diagnosis and surveillance. It targets both a general audience and those who would like to learn more about rabies and the pathway to eliminating this disease.	English, French, and Russian	September 2021	3 hours	9615	• Record of achievement: Earning at least 80% of the maximum points from all graded assignments.

[a]WHO/ICRC: World Health Organization/International Committee of the Red Cross.

Objectives

We surveyed the learners from the 4 emergency health MOOCs to understand current trends in sharing and adapting these courses by learners on the OpenWHO platform. This study analyzes the survey responses to examine differences in course use by World Bank country income classification across the following dimensions: (1) how health education MOOCs are being used and shared by learners, (2) how health workers adapt global health education MOOC content to meet local health education training and information needs, and (3) how content adaptations can help frontline health workers overcome the barriers to using health education MOOCs, especially in LICs/LMICs.

Methods

The study leverages 2 sources of data. The first source is course enrollment data collected from the start of each course through the end of March 2022 (N=96,395) from the 4 MOOCs on the

XSL•FO
RenderX

OpenWHO platform. The second data source is surveys of learners who participated in at least 1 of the 4 MOOCs (N=926) during this same time period.

Enrollment Data

The enrollment data describe the demand and the ability to access the 4 MOOCs. The data were collected from learners at the time they registered to enroll in the MOOCs. Enrollment data included information on learners' country of residence, gender, institutional affiliation, age range, primary language, enrollment language, and course completion.

Survey Data

In March 2022, we distributed a survey via email to learners who enrolled in at least 1 of the 4 MOOCs with a link to participate in the survey. The survey was sent in 3 languages: English (all MOOCs had English versions), French (the Ebola and rabies MOOCs had French versions), and Spanish (the COVID-19 "Initial Approach to the Acutely Ill Patient" had a Spanish version).

The 33-question 3-part survey included sections with questions about (1) which MOOCs the learners took and their experiences with the courses, (2) how they shared and adapted course content, and (3) how accessible the learners found course content (see Multimedia Appendix 1 for the full survey). The 7-question section on learners' experiences with the MOOCs asked questions about course completion, timing, and motivation (see survey Section A in Multimedia Appendix 1). The 11-question section on course sharing and adaptation asked questions about course recommendations, sharing modalities, sharing experiences, modification modalities, and modification experiences (see survey Section Bin Multimedia Appendix 1). The 16-question section on accessibility asked questions about learner characteristics and accessibility of course content for professional and patient populations (see survey Section C in Multimedia Appendix 1).

Data Analysis

Enrollment and survey data were summarized using descriptive statistics (mean, SD, and response rates). To understand the generalizability of our survey sample to the overall population of enrollees and course completers, we summarized the learner characteristics of enrollees, course completers, and survey respondents. The learner characteristics we examined include course enrollment, geographic region, country income classification, course language, language preference, professional affiliation, age, and gender. Not all enrollees provided responses to registration questions; we included percentages of those for whom we have data on characteristics.

To investigate the differences in course use across learners globally, we summarized survey data by World Bank country income classifications—HICs, UMICs, LICs, and LMICs. We tested differences in course sharing and adaptation by country income classification using Pearson chi-square test, comparing sharing and adaptation patterns in UMICs, LMICs, and LICs with patterns in HICs.

In an attempt to quantify the course reach that cannot be observed in platform analytics, we calculated a lower-bound estimate of sharing by assuming that each individual shared with the lowest number of people in their respective survey category responses (ie, 0 people, 1-10 people, 11-25 people, 26-50 people, and 50 people). We calculated the lower bound number of recipients reached through sharing, as well as the lower bound number of trainees reached through the use of the course content in official training. We used one-way ANOVA to compare differences in the mean number of recipients and training by country income classification. All statistical analyses were performed using Stata SE version 15.

Ethical Considerations

Informed consent was obtained from all surveyed participants. Approval for this study was obtained from the Stanford University School of Medicine institutional review board (protocol no. 62414). Data were collected anonymously and respondents did not receive compensation.

Results

Enrollment Patterns

A total of 96,395 learners enrolled in the 4 MOOCs from each course's start date until the end of March 2022. Learners could enroll in more than 1 course. Of the enrolled learners who indicated geographic residence when registering for the OpenWHO platform, half indicated that they were from LMICs (43,168/85,882, 50%), and another 9% (7146/85,882) indicated that they were from LICs. For context, roughly 40% of the world's population resides in LMICs, while roughly 9% reside in LICs [58]. More than one-third of enrolled learners indicated that they were from South Asia (29,001/85,882, 34%), with the subsequent largest distribution of learners from Sub-Saharan Africa (14,058/85,882, 16%). Examining enrolled learners' demographics finds that 79% (51,479/64,793) indicated that their preferred language for learning was English, and 88% (84,949/96,395) enrolled in an English language MOOC.

A third of enrolled learners indicated they were students (28,308/87,041, 33%), with the next largest enrolled population being health care professionals (23,773/87,041, 27%). Enrolled learners were generally young, with 85% (50,618/59,896) of learners younger than 40 years. In addition, more than half of the enrolled learners identified as male (35,444/65,430, 54%).

Course Completion Patterns

MOOCs generally have low completion rates, with rates that can be disproportionately lower in LICs/LMICs [31]. These 4 MOOCs have an overall course completion rate of 34% (32,395/96,395), which ranges from 23% (2165/9615) for the Rabies and One Health course to 40% (9376/23,803) for the Ebola: Clinical Management course. Course completion rates were only slightly lower in LICs (2457/7146, 34%) and LMICs (14,942/43,168, 35%) than in UMICs (7207/19,743, 37%) and HICs (6870/15,825, 43%). As shown in Table 2, course completers generally looked similar to the population that enrolled in the MOOCs with few notable differences.

XSL•FO
RenderX

Table 2. Learner characteristics by course enrollment, course completion, and survey completion[a].

Characteristics	Enrolled in course (N=96,395)	Completed course (N=32,935)	Completed survey (N=926)
Course, n (%)			
COVID-19 Acutely Ill Patients	24,036 (25)	7445 (23)	607 (66)
COVID-19 General Considerations	38,941 (40)	13,949 (42)	524 (57)
Ebola Clinical Management	23,803 (25)	9376 (28)	254 (27)
Rabies and One Health	9615 (10)	2165 (7)	283 (31)
Geographic region, n (%)			
East Asia and Pacific	11,972 (14)	4413 (14)	79 (9)
Europe and Central Asia	6693 (8)	2382 (8)	63 (7)
Latin America and Caribbean	9927 (12)	3645 (12)	137 (15)
Middle East and North Africa	6760 (8)	2328 (7)	63 (7)
North America	7471 (9)	3837 (12)	23 (2)
South Asia	29,001 (34)	10,469 (33)	229 (25)
Sub-Saharan Africa	14,058 (16)	4402 (14)	332 (36)
Not specified	10,513	1459	N/A[b]
Country income classification, n (%)			
Low income	7146 (8)	2457 (8)	171 (18)
Lower-middle income	43,168 (50)	14,942 (47)	441 (48)
Upper-middle income	19,743 (23)	7207 (23)	200 (22)
High income	15,825 (18)	6870 (22)	114 (12)
Not specified	10,513	1459	N/A[b]
Course language, n (%)			
English	84,949 (88)	28,499 (87)	769 (83)
French	3864 (4)	1316 (4)	45 (5)
Spanish	4993 (5)	2185 (7)	112 (12)
Other	2589 (3)	935 (3)	N/A[b]
OpenWHO language preference, n (%)			
Arabic	1947 (3)	606 (3)	22 (2)
Chinese	2225 (3)	792 (4)	20 (2)
English	51,479 (79)	17,720 (80)	627 (68)
French	2537 (4)	839 (4)	64 (7)
Portuguese	589 (1)	131 (1)	7 (1)
Russian	454 (1)	153 (1)	2 (0)
Spanish	5562 (9)	1902 (9)	138 (15)
Other	N/A[b]	N/A[b]	46 (5)
Not specified	31,602	10,792	N/A[b]
Professional affiliation, n (%)			
Government/Ministry of Health	8491 (10)	2719 (9)	51 (6)
Health care professional	23,773 (27)	7387 (24)	406 (44)
International organization (eg, WHO, UN)	2532 (3)	974 (3)	23 (2)
Nongovernmental organization	3546 (4)	1186 (4)	58 (6)

XSL•FO
RenderX

Characteristics	Enrolled in course (N=96,395)	Completed course (N=32,935)	Completed survey (N=926)
Student	28,308 (33)	11,370 (37)	84 (9)
Volunteer	4522 (5)	1519 (5)	31 (3)
Other: health care	5568 (6)	1755 (6)	141 (15)
Other:- non–health care	10,301 (12)	3561 (12)	132 (14)
Not specified	9354	2464	N/A[b]
Age (years), n (%)			
<20	9142 (15)	3489 (17)	15 (2)
20-29	27,871 (47)	9976 (47)	248 (27)
30-39	13,505 (23)	4542 (21)	294 (32)
40-49	5938 (10)	2045 (10)	198 (21)
50-59	2470 (4)	810 (4)	111 (12)
60-69	613 (1)	196 (1)	51 (6)
70+	357 (1)	83 (0)	7 (1)
Not specified	36,499	11,794	2
Gender, n (%)			
Female	29,877 (46)	9479 (42)	291 (32)
Male	35,444 (54)	12,921 (58)	615 (67)
Nonbinary/other	109 (0)	39 (0)	11 (1)
Not specified	30,965	10,496	9

[a]This table compares the characteristics (n [%]) of health workers who completed the focal courses and the follow-up survey. A higher proportion of course enrollees did not specify characteristics compared with survey completers. We show the numbers not specified for each but do not include them in the percentage breakdowns. For course completion, geographic region was identified via course platform analytics; however, we were unable to identify a subset, shown as "not specified" in the table. For survey completion, the geographic region was identified through survey reports and IP addresses and was fully specified across respondents. Percentages are shown for those for whom we have data on characteristics. Other course languages include Armenian, Indonesian, Kazakh, Macedonian, Dutch, Portuguese, Russian, Albanian, Kiswahili, and Somali.

[b]N/A: not applicable.

Survey Respondents

A number of survey respondents enrolled in more than 1 course. Nearly twice as many respondents had enrolled in the COVID-19 courses (COVID-19: General Considerations, N=607; COVID-19: Initial Approach, N=534) as compared with the other emergency health courses (Ebola: Clinical Management, N=354; Rabies and One Health, N=283). As shown in Table 2, the characteristics of the survey respondents generally reflect the characteristics of enrollees and course completers, with a few exceptions. A larger proportion of survey respondents indicated that they were from LICs (171/926, 18%). Differing from the overall enrollment and course-completer data, more than one-third of respondents indicated that they were from Sub-Saharan Africa (332/926, 36%). A larger proportion of survey respondents indicated that they were health care professionals (406/926, 44%) and employed in other health care–related professions (141/926, 15%) compared with the overall enrollee and course-completer population.

Sharing Course Content

Overall, Table 3 shows that the vast majority of survey respondents (819/926, 88%) shared the course content. While sharing is high in all countries, respondents in LMICs (402/441, 91%) were significantly more likely to share course content than those in HICs (92/114, 81%; *P*=.001). The majority of learners who reported sharing course content indicated that they shared with 10 or fewer people (542/926, 59%). On average, learners shared content with a lower bound estimate of 8.16 (SD 14.45) people. Compared with sharing among learners in HICs who shared with an average of 6.73 (SD 12.32) people, the number of people with whom learners shared content was higher in LMICs (mean of 9.48, SD 15.76 people; *P*=.08) although not statistically significant.

XSL•FO
RenderX

Table 3. Sharing of course materials and information by country income classificationa.

	Total (N=926)	HIC[b] (N=114)	UMIC[c] (N=200)	*P* value	LMIC[d] (N=441)	*P* value	LIC[e] (N=171)	*P* value
Panel A: informal sharing (shared course materials or information). Sharing, n (%)	819 (88)	92 (81)	176 (88)	.079	402 (91)	.001	149 (87)	.142
Sharing, by number of recipients, n (%)								
1-10 people	542 (59)	61 (54)	122 (61)	.197	253 (57)	.459	106 (62)	.156
11-25 people	136 (15)	15 (13)	25 (13)	.867	73 (17)	.377	23 (13)	.944
26-50 people	67 (7)	11 (10)	18 (9)	.849	30 (7)	.301	8 (5)	.100
>50 people	74 (8)	5 (4)	11 (6)	.667	46 (10)	.047	12 (7)	.360
Lower bound number of recipients, mean (SD)	8.16 (14.45)	6.73 (12.32)	7.13 (13.02)	.789	9.48 (15.76)	.084	6.89 (13.59)	.916
Panel B: Use in official trainings (used course materials in an official training). Used course in training, n (%)	563 (61)	46 (40)	113 (56)	.006	284 (64)	<.001	120 (70)	<.001
Course use in training, by number of trainees, n (%)								
1-10 people	370 (40)	28 (25)	73 (37)	.029	185 (42)	.001	84 (49)	<.001
11-25 people	94 (10)	9 (8)	19 (10)	.633	44 (10)	.501	22 (13)	.188
26-50 people	49 (5)	6 (5)	11 (6)	.929	25 (6)	.867	7 (4)	.644
>50 people	50 (5)	3 (3)	10 (5)	.313	30 (7)	.094	7 (4)	.513
Lower bound number of trainees, mean (SD)	5.65 (12.54)	3.82 (10.00)	5.39 (12.28)	.247	6.46 (13.65)	.054	5.06 (11.20)	.343

[a]This table shows the rate of informal sharing (as measured by a question asking whether learners had recommended the course or informally shared course content with other individuals) and use in an official training (as measured by a question asking whether learners had shared materials from the course as part of an official training). We also show a lower-bound estimate of sharing and number of trainees, calculated assuming learners shared with the lowest number of people in their indicated survey category response. We show differences by World Bank income classifications: HIC, UMIC, LMIC, and LIC. Tests of comparison in course sharing and adaptation by country income classification were conducted using Pearson chi-square test. We used one-way ANOVA to compare differences in the mean number of sharing recipients and trainees by country classification. We show comparisons of UMICs, LMICs, and LICs with HICs.

[b]HIC: high-income country.

[c]UMIC: upper-middle-income country.

[d]LMIC: lower-middle-income country.

[e]LIC: low-income country.

Course content was commonly used in official trainings (61%, 563/926). Use of the content in official training was significantly higher in LICs (120/171, 70%; *P*<.001), LMICs (284/441, 64%; *P*<.001), and UMICs (113/200, 56%; *P*=.006) than in HICs (46/114, 40%). The majority of the trainings in which course content was shared engaged 10 or fewer people (370/563, 67%). The lower bound estimated number of trainees was, on average, 5.65 (SD 12.54) people. Compared with the mean number of trainees in HICs (mean of 3.82, SD 10.00 people), significantly more were trained on average in LMICs (mean of 6.46, SD 13.65 people; *P*=.054).

Characteristics of Course Sharing, Adaptation, and Accessibility

Learners shared course content widely with many in their personal and professional networks, including most notably colleagues (486/819, 59%) and friends (364/819, 44%). As shown in Table 4, this pattern was generally consistent across income country classifications. Likewise, across all income country classifications, sharing was highest among health care professionals (363/819, 44%) and learners working in other health care–related fields (126/819, 15%).

XSL•FO
RenderX

Table 4. Characteristics of course sharing by country income classification[a].

	Total (N=819)	HIC[b] (N=92)	UMIC[c] (N=176)	*P* value	LMIC[d] (N=402)	*P* value	LIC[e] (N=149)	*P* value
With whom did you share? n (%)								
Classmate(s) if a student	197 (24)	10 (11)	43 (24)	.008	107 (27)	.001	37 (25)	.008
Colleague(s)	486 (59	52 (57)	106 (60)	.558	229 (57)	.938	99 (66)	.122
Employee(s) whom you supervise	176 (21)	17 (18)	26 (15)	.433	89 (22)	.440	44 (30)	.055
Employer	108 (13)	12 (13)	17 (10)	.397	49 (12)	.822	30 (20)	.159
Family member(s)	220 (27)	22 (24)	39 (22)	.745	123 (31)	.204	36 (24)	.965
Friend(s)	364 (44)	36 (39)	63 (36)	.591	186 (46)	.214	79 (53)	.036
Patient(s) if a health care worker	180 (22)	14 (15)	35 (20)	.348	98 (24)	.058	33 (22)	.187
Student(s) if an educator	160 (20)	9 (10)	23 (13)	.431	96 (24)	.003	32 (21)	.019
How did you share course information? n (%)								
Shared course website link	489 (60)	51 (55)	102 (58)	.692	240 (60)	.453	96 (64)	.164
Shared course files via offline technology (eg, USB, DVD, CD)	105 (13)	7 (8)	14 (8)	.920	54 (13)	.126	30 (20)	.009
Shared course audio files on the radio	78 (10)	5 (5)	10 (6)	.933	48 (12)	.069	15 (10)	.205
Shared course information via email	242 (30)	23 (25)	31 (18)	.152	132 (33)	.144	56 (38)	.043
Shared course information via text message	108 (13)	9 (10)	15 (9)	.732	57 (14)	.264	27 (18)	.078
Shared course information via WhatsApp	298 (36)	19 (21)	67 (38)	.004	165 (41)	<.001	47 (32)	.065
Shared course information via word of mouth	290 (35)	35 (38)	62 (35)	.649	144 (36)	.689	49 (33)	.414
Shared hard copy course materials (eg, printed handouts)	92 (11)	8 (9)	7 (4)	.111	53 (13)	.238	49 (33)	.100
What components did you share? n (%)								
Videos	284 (35)	22 (24)	51 (29)	.377	156 (39)	.007	55 (37)	.035
Audio files	17 (2)	3 (3)	9 (5)	.486	0 (0)	<.001	5 (3)	.968
Slides	318 (39)	21 (23)	60 (34)	.057	158 (39)	.003	79 (53)	<.001
Quizzes	189 (23)	13 (14)	29 (16)	.616	105 (26)	.015	42 (28)	.012
Text from the course	214 (26)	20 (22)	37 (21)	.892	113 (28)	.214	44 (30)	.183
Transcripts	133 (16)	9 (10)	13 (7)	.497	85 (21)	.012	26 (17)	.101
Downloadable documents	263 (32)	23 (25)	45 (26)	.919	128 (32)	.199	67 (45)	.002
Infographics	111 (14)	12 (13)	20 (11)	.687	65 (16)	.456	14 (9)	.375
Discussions	147 (18)	13 (14)	15 (9)	.154	91 (23)	.071	28 (19)	.349
Photographs	9 (1)	0 (0)	1 (1)	.469	3 (1)	.406	5 (3)	.076
Other	47 (6)	4 (4)	11 (6)	.520	21 (5)	.729	11 (7)	.343
Did you modify any content shared? n (%)								

XSL•FO
RenderX

	Total (N=819)	HIC[b] (N=92)	UMIC[c] (N=176)	*P* value	LMIC[d] (N=402)	*P* value	LIC[e] (N=149)	*P* value
Printed content to be available offline (eg, handouts)	29 (4)	4 (4)	8 (5)	.941	5 (1)	.045	12 (8)	.262
Adapted content to distribute via text message or WhatsApp	216 (26)	7 (8)	42 (24)	.001	125 (31)	<.001	42 (28)	<.001
Added culture or local contextual information	133 (16)	11 (12)	11 (6)	.106	80 (20)	.076	31 (21)	.079
Added additional explanations in local language	178 (22)	18 (20)	18 (10)	.033	99 (25)	.303	43 (29)	.107
Adjusted content to meet local guidelines /regulations	143 (17)	8 (9)	17 (10)	.797	81 (20)	.010	37 (25)	.002
Made changes to course animations and images	73 (9)	5 (5)	8 (5)	.748	42 (10)	.139	18 (12)	.088
Translated all or part of the course into a local language	142 (17)	10 (11)	27 (15)	.314	76 (19)	.067	29 (19)	.078
Other modifications not described above	76 (9)	10 (11)	15 (9)	.531	33 (8)	.414	18 (12)	.776
Did you have any difficulties when sharing? n (%)								
No difficulties	381 (47)	44 (48)	73 (41)	.320	187 (47)	.820	77 (52)	.561
Web-based connectivity	178 (22)	7 (8)	29 (16)	.043	99 (25)	<.001	43 (29)	<.001
Language barriers	120 (15)	8 (9)	18 (10)	.688	60 (15)	.118	34 (23)	.005
Technical difficulty	77 (9)	6 (7)	8 (5)	.490	47 (12)	.148	16 (11)	.270
Reaching intended audience was difficult	61 (7)	4 (4)	5 (3)	.516	35 (9)	.162	17 (11)	.059
Translation difficulty	70 (9)	5 (5)	10 (6)	.933	38 (9)	.217	17 (11)	.118
Other	41 (5)	4 (4)	8 (5)	.941	21 (5)	.729	8 (5)	.723
What are the professions of the sharers? n (%)								
Government	45 (5)	3 (3)	9 (5)	.486	24 (6)	.302	9 (6)	.335
Health care professional	363 (44)	42 (46)	85 (48)	.681	168 (42)	.499	68 (46)	.998
Community health worker	85 (23)	7 (17)	12 (14)	.705	43 (26)	.224	23 (34)	.050
Nurse	119 (33)	23 (55)	28 (33)	.018	51 (30)	.003	17 (25)	.002
Paramedic /emergency medicine technician	18 (5)	4 (10)	7 (8)	.808	6 (4)	.105	1 (1)	.049
Pharmacist	26 (7)	0 (0)	4 (5)	.153	18 (11)	.027	4 (6)	.109
Physician	111 (31)	8 (19)	33 (39)	.025	47 (28)	.239	23 (34)	.094
Traditional /complementary medicine	4 (1)	0 (0)	1 (1)	.480	3 (2)	.383	0 (0)	N/A[f]
International organization (eg, WHO, and UN)	19 (2)	1 (1)	6 (3)	.258	7 (2)	.654	5 (3)	.272
Nongovernmental organization	52 (6)	5 (5)	8 (5)	.748	24 (6)	.844	15 (10)	.205
Student	66 (8)	2 (2)	20 (11)	.009	37 (9)	.024	7 (5)	.315
Volunteer	31 (4)	4 (4)	5 (3)	.516	16 (4)	.872	6 (4)	.903

XSL•FO
RenderX

	Total (N=819)	HIC[b] (N=92)	UMIC[c] (N=176)	*P* value	LMIC[d] (N=402)	*P* value	LIC[e] (N=149)	*P* value
Other: health care	126 (15)	18 (20)	17 (10)	.022	71 (18)	.668	20 (13)	.204
Other: non–health care	117 (14)	17 (18)	26 (15)	.433	55 (14)	.240	19 (13)	.226

[a]This table shows the characteristics of course sharing by World Bank income classifications: HIC, UMIC, LMIC, and LIC. Tests of comparison in course sharing and adaptation by country income classification were conducted using Pearson chi-square test. We show comparisons of UMICs, LMICs, and LICs with HICs.

[b]HIC: high-income country.

[c]UMIC: upper-middle-income country.

[d]LMIC: lower-middle-income country.

[e]LIC: low-income country.

[f]N/A: not applicable.

Across income country classifications, learners were most likely to share course content via website links (489/819, 60%), followed by WhatsApp (298/819, 36%), word of mouth (290/819, 35%), and email (242/819, 30%). Sharing course information via WhatsApp was highest in UMICs (67/176, 38%; *P*=.004) and LMICs (165/402, 41%; *P*<.001) and significantly higher than in HICs (19/92, 21%). Learners in LICs were more likely to share course files via offline technology (30/149, 20%; *P*=.009) than those in HICs (7/92, 8%). The course components shared by learners also varied by country income classification. Learners in LICs or LMICs were more likely to share videos (LMICs, *P*=.007; LICs, *P*=.04), slides (LMICs, *P*=.003; LICs, *P*<.001), quizzes (LMICs, *P*=.02; LICs, *P*=.01), transcripts (LMICs, *P*=.01), and downloadable documents (LICs, *P*=.002) than those learners in HICs.

Learners adapted and made modifications to the MOOCs to share their content with others in all countries, with a few notable variations across income country classification. Over a quarter of learners (216/819, 26%) adapted course content to make it easier to distribute via text message or WhatsApp. This type of adaptation was significantly more likely in UMICs (42/176, 24%; *P*=.001), LMICs (125/402, 31%; *P*<.001), and LICs (42/149, 28%; *P*<.001) than in HICs (7/92, 8%). Compared with learners in HICs, learners in LICs or LMICs were also more likely to make adaptations to add cultural or contextual information (LMICs, *P*=.076; LICs, *P*=.079), to adjust content to meet local guidelines and regulations (LMICs, *P*=.010; LICs, *P*=.002), and to address linguistic barriers by translating all or part of the course into a local language (LMICs, *P*=.067; LICs, *P*=.078).

These adaptations may have been made to address accessibility challenges. Internet connectivity was the biggest challenge to accessing and sharing MOOCs for over a quarter of all learners in LMICs (99/402, 25%; *P*<.001) and LICs (43/149, 29%; *P*<.001) and significantly more learners experienced internet connectivity challenges than learners in HICs (7/92, 8%). In addition, nearly a quarter of learners in LICs (34/149, 23%; *P*=.005) expressed that language barriers presented a challenge to accessing and sharing content, which was significantly higher than the proportion expressing this challenge in HICs (8/92, 9%). Even so, a quarter of learners in LICs pointing to language barriers is relatively low considering the levels of literacy of the common course languages (eg, English), suggesting that courses are not reaching a broader audience in LICs.

Across all survey respondents, learners in UMICs, LMICs, and LICs were more likely to express that they found the OpenWHO MOOC content less accessible than learners in HICs (Table 5). The vast majority of learners in HICs (89/114, 78%) expressed that the MOOC content was accessible to most people in their country, compared with just more than half of learners in UMICs (110/200, 55%; *P*<.001) and LMICs (229/441, 52%; *P*<.001). Less than a third of learners in LICs (53/171, 31%; *P*<.001) expressed that MOOC content was accessible to most people in their country. This pattern holds for accessibility among patients, as indicated by health care professionals, with accessibility particularly low in LICs where 38% (54/171) of learners indicated the MOOC content is accessible to only very few people.

XSL•FO
RenderX

Table 5. Accessibility of OpenWHO courses by country income classification[a].

Accessibility of OpenWHO	HIC[b] (N=114)	UMIC[c] (N=200)	*P* value	LMIC[d] (N=441)	*P* value	LIC[e] (N=171)	*P* value
In your country, n (%)							
Accessible to most	89 (78)	110 (55)	<.001	229 (52)	<.001	53 (31)	<.001
Accessible to some	19 (17)	60 (30)	.009	124 (28)	.013	61 (36)	<.001
Accessible to very few	6 (5)	30 (15)	.009	88 (20)	<.001	57 (33)	<.001
In your personal network, n (%)							
Accessible to most	80 (70)	103 (52)	.001	208 (47)	<.001	45 (26)	<.001
Accessible to some	24 (21)	57 (29)	.147	139 (32)	.029	63 (37)	.005
Accessible to very few	10 (9)	40 (20)	.009	94 (21)	.002	63 (37)	<.001
In your professional network, n (%)							
Accessible to most	81 (79)	123 (66)	.022	232 (56)	<.001	53 (34)	<.001
Accessible to some	16 (16)	50 (27)	.029	124 (30)	.003	71 (45)	<.001
Accessible to very few	6 (6)	14 (7)	.593	57 (14)	.027	34 (22)	.001
Among your patients, n (%)							
Accessible to most	55 (65)	69 (42)	.001	175 (45)	.001	32 (22)	<.001
Accessible to some	15 (18)	56 (34)	.007	128 (33)	.006	58 (40)	<.001
Accessible to very few	14 (17)	38 (23)	.225	85 (22)	.285	54 (38)	.001

[a]This table shows responses to questions about the accessibility of OpenWHO courses by World Bank income classifications: HIC, UMIC, LMIC, and LIC. Tests of comparison in course sharing and adaptation by country income classification were conducted using Pearson chi-square test. We show comparisons of UMICs, LMICs, and LICs with HICs.

[b]HIC: high-income country.

[c]UMIC: upper-middle-income country.

[d]LMIC: lower-middle-income country.

[e]LIC: low-income country.

Discussion

Principal Findings

This study is one of the first to examine the sharing and adaptation of health MOOCs globally. The WHO expedites the production of learning materials as soon as evidence-based science becomes available for any disease outbreak to ensure that communities and health workers have resources that are reliable, readily accessible, and language appropriate [3]. Given this diverse scope and the ongoing challenges of accessing web-based platforms and MOOCs in different countries, we sought to understand how these 4 MOOCs are reaching a global audience. Overall, we found that OpenWHO MOOC content is highly shared across all regions of the world, presenting a tremendous opportunity for the dissemination and scaling of health information. However, adaptation (ie, contextualization and localization) is critical to improve reach in LMICs or LICs.

This study demonstrates that in addition to the official course enrollers and completers, the OpenWHO course content is reaching an even larger audience than that which is observable from platform analytics. The majority of survey respondents indicated that they shared course content or used course content in official training. They were most likely to share videos, share content via WhatsApp, and adapt content to share using technologies such as WhatsApp or text message. While respondents across all country classifications shared course content and materials, learners in LMICs were most likely to share course content. On average, they also reached the largest number of people with their sharing.

The variation in the ways MOOCs are being shared and adapted, as well as the reported accessibility challenges in LICs or LMICs, suggests that MOOCs designed purely for an individual learner in an online format will not optimize knowledge and skill dissemination or most effectively reach geographies with the lowest level of resources. To access this content, learners must rely on others with the access, means, and expertise to adapt and translate the MOOCs to overcome technical limitations. Our findings are consistent with previous research on the OpenWHO platform indicating that although the platform is designed for low-bandwidth use, such as providing downloadable materials, technical limitations remain [16]. In LMICs or LICs, we observed a greater level of effort to distribute content in offline ways to address technical difficulties, such as adapting content to distribute via text message and WhatsApp and by sharing course files via offline technology (eg, USB, DVD, and CD). The lower rate of adaptation in HICs may reflect that the content, made by course developers in HICs, requires less modification to share. It may also reflect that there are more widely available alternative

XSL•FO
RenderX

sources of content in HICs. These findings present a tremendous opportunity for course developers to devise new ways of designing and disseminating courses that could have a dramatic scaling effect, particularly in areas with more access challenges.

In addition to addressing technical limitations, we observed that learners in LMICs or LICs were significantly more likely to modify content to address local guidelines, cultural contexts, and languages. Learners in LICs, in particular, expressed that language barriers created an accessibility challenge for content. Our findings reflect the literature that argues that to be relevant, MOOCs need to be adaptable and linguistically and culturally responsive [3,41,46,47]. Successful content adaptations of this nature likewise require substantial time, resources, and particular linguistic and medical expertise, all of which form barriers to individual learners who wish to adapt and share courses in order to spread knowledge and strengthen the local health care response. These findings require that we intentionally design MOOCs that facilitate and support adaptation and sharing by local health professionals to address needs in LMICs or LICs where the content is critical to promoting information equity during public health emergencies.

Limitations and Future Research

Our study findings are limited to the population who responded to our survey. As a result, our findings do not represent the entire population that accessed the OpenWHO platform or enrolled in the MOOCs and shared content but did not respond to the survey. Our survey respondents are learners with email addresses and the time, interest, and capacity to respond to an online survey. Respondents were also those with the inclination and ability to respond to a survey in English, French, or Spanish. That only a quarter of survey respondents in LICs indicated experiencing language barriers—relatively low considering the levels of literacy of the common course languages, (eg, English, French, and Spanish) overall in LICs—could suggest that MOOCs are not reaching a broader audience in LICs but also likely reflects a limitation with the selected survey response. Furthermore, our findings omit feedback from offline learners exposed to course content through the sharing we describe in this study. We hence cannot speak to the quality of the learning experience among such learners, who would be able to shed light on ways to improve the creation of MOOC content to better facilitate alternative content distribution paths.

Future research must illuminate how best to develop tools and provide the additional supports needed for learners and institutions wishing to adapt and share MOOC content. In addition, it is important that we evaluate the impact of shared trainings on “secondary” learners’ knowledge, skills, and behaviors. Such investigations are essential to ensure that shifts toward online and digital educational approaches during health emergencies do not exacerbate global gaps in access to health care and health knowledge.

Conclusions

There exists a strong demand for OpenWHO MOOCs health emergencies trainings as demonstrated by the high rates of sharing materials across countries of all income levels. Local content sharing presents a tremendous opportunity for organizations and governments to exponentially scale the impact of MOOCs. MOOC developers must design courses for easier distribution via commonly identified formats for sharing such as text message and WhatsApp. Learners in LMICs or LICs were more likely to adapt content to meet local cultural, contextual, and linguistic needs but faced increased barriers to adapting and sharing materials. This study advances our understanding of how to design and disseminate educational materials to promote equitable access to critical health information during times of crisis.

Acknowledgments

This study was made possible by contributions from the Stanford Department of Emergency Medicine, the Stanford Center for Health Education’s Digital Medic initiative, and the Learning and Capacity Development Unit, Health Emergencies Programme at the World Health Organization. Generative artificial intelligence tools were used only for grammatical editing and to improve the clarity of writing in the manuscript. We would like to thank all the institutions and learners who directly contributed to this research, especially those who took the MOOCs and shared their experiences with our team.

Data Availability

Some data are available on reasonable request to the corresponding author.

Authors' Contributions

JSJ and NAS led the conceptualization and design of the study and oversaw study implementation, writing, and editing. JSJ led the quantitative analysis and manuscript writing and editing. NN and AT contributed to the data collection design, and AT led the data collection process. EA led data cleaning. NN, AT, EA, HU, and MS contributed to the study’s design, interpretation of findings, and revision of all drafts. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Learner survey in English.
[PDF File (Adobe PDF File), 445 KB - jmir_v27i1e52591_app1.pdf]

References

1. Utunen H, Staubitz T, George R, Zhao YU, Serth S, Tokar A. Scale up multilingualism in health emergency learning: developing an automated transcription and translation tool. Stud Health Technol Inform 2023;302:408-412. [doi: 10.3233/SHTI230162] [Medline: 37203706]
2. Utunen H, Attias M, Ndiaye N, Tokar A. Expanding multilingual learning as a pathway to equity during the COVID-19 pandemic: a geographical analysis of a WHO online course in 45 languages. Int J Adv Corp Learn 2023;16(1):78-90. [doi: 10.3991/ijac.v16i1.35097]
3. George R, Utunen H, Ndiaye N, Tokar A, Mattar L, Piroux C, et al. Ensuring equity in access to online courses: perspectives from the WHO health emergency learning response. World Med Health Policy 2022;14(2):413-427. [doi: 10.1002/wmh3.492]
4. Ruipérez-Valiente JA, Martin S, Reich J, Castro M. The UnMOOCing process: extending the impact of MOOC educational resources as OERs. Sustainability 2020;12(18):7346. [doi: 10.3390/su12187346]
5. Skinner NA, Job N, Krause J, Frankel A, Ward V, Johnston JS. The use of open-source online course content for training in public health emergencies: mixed methods case study of a COVID-19 course series for health professionals. JMIR Med Educ 2023;9:e42412 [FREE Full text] [doi: 10.2196/42412] [Medline: 36735834]
6. Utunen H, Attias M, George R, O'Connell G, Tokar A. Learning Multiplier Effect of OpenWHO.Org: Use of Online Learning Materials Beyond the Platform.: World Health Organization; 2022. URL: https://www.who.int/publications-detail-redirect/WER9701-02-1-7 [accessed 2022-08-01]
7. Winters N, Patel KD. Can a reconceptualization of online training be part of the solution to addressing the COVID-19 pandemic? J Interprof Care 2021;35(2):161-163. [doi: 10.1080/13561820.2021.1892615] [Medline: 33691565]
8. Otu A, Ebenso B, Okuzu O, Osifo-Dawodu E. Using a mHealth tutorial application to change knowledge and attitude of frontline health workers to Ebola virus disease in Nigeria: a before-and-after study. Hum Resour Health 2016;14:5 [FREE Full text] [doi: 10.1186/s12960-016-0100-4] [Medline: 26872824]
9. Otu A, Okuzu O, Ebenso B, Effa E, Nihalani N, Olayinka A, et al. Introduction of mobile health tools to support COVID-19 training and surveillance in Ogun State Nigeria. Front Sustain Cities 2021;3:638278. [doi: 10.3389/frsc.2021.638278]
10. Burkardt AD, Krause N, Rivas Velarde MC. Critical success factors for the implementation and adoption of e-learning for junior health care workers in Dadaab refugee camp Kenya. Hum Resour Health 2019;17(1):98 [FREE Full text] [doi: 10.1186/s12960-019-0435-8] [Medline: 31818294]
11. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
12. Cory N, Stevens P. Building a global framework for digital health services in the era of COVID-19. Information Technology and Innovation Foundation. 2020. URL: https://itif.org/publications/2020/05/26/building-global-framework-digital-health-services-era-covid-19 [accessed 2021-04-30]
13. World Health Organization. Digital education for building health workforce capacity.: World Health Organization; 2020. URL: https://apps.who.int/iris/handle/10665/331524 [accessed 2023-04-20]
14. Utunen H, Ndiaye N, Piroux C, George R, Attias M, Gamhewage G. Global reach of an online COVID-19 course in multiple languages on OpenWHO in the first quarter of 2020: analysis of platform use data. J Med Internet Res 2020;22(4):e19076 [FREE Full text] [doi: 10.2196/19076] [Medline: 32293580]
15. Goldin S, Kong SYJ, Tokar A, Utunen H, Ndiaye N, Bahl J, et al. Learning from a massive open online COVID-19 vaccination training experience: survey study. JMIR Public Health Surveill 2021;7(12):e33455 [FREE Full text] [doi: 10.2196/33455] [Medline: 34794116]
16. Utunen H, Tokar A, Dancante M, Piroux C. Online learning for WHO priority diseases with pandemic potential: evidence from existing courses and preparing for disease X. Arch Public Health 2023;81(1):61 [FREE Full text] [doi: 10.1186/s13690-023-01080-9] [Medline: 37085922]
17. Magaña-Valladares L, Rosas-Magallanes C, Montoya-Rodríguez A, Calvillo-Jacobo G, Alpuche-Arande CM, García-Saisó S. A MOOC as an immediate strategy to train health personnel in the cholera outbreak in Mexico. BMC Med Educ 2018;18(1):111 [FREE Full text] [doi: 10.1186/s12909-018-1215-1] [Medline: 29769059]
18. Liyanagunawardena TR, Williams SA. Massive open online courses on health and medicine: review. J Med Internet Res 2014;16(8):e191 [FREE Full text] [doi: 10.2196/jmir.3439] [Medline: 25123952]
19. Balsari S, Sange M, Udwadia Z. COVID-19 care in India: the course to self-reliance. Lancet Glob Health 2020;8(11):e1359-e1360 [FREE Full text] [doi: 10.1016/S2214-109X(20)30384-3] [Medline: 32853553]
20. Longhini J, De Colle B, Rossettini G, Palese A. What knowledge is available on massive open online courses in nursing and academic healthcare sciences education? A rapid review. Nurse Educ Today 2021;99:104812 [FREE Full text] [doi: 10.1016/j.nedt.2021.104812] [Medline: 33610051]
21. Attias M, Utunen H, Ndiaye N, Mattar L. Open-access learning as a pathway to equity during health emergencies. In: Guralnick D, Auer ME, Poce A, editors. Innovative Approaches to Technology-Enhanced Learning for the Workplace and

XSL•FO
RenderX

Higher Education. Lecture Notes in Networks and Systems. Cham Switzerland: Springer International Publishing; 2023:573-580.
22. Liyanagunawardena TR, Aboshady OA. Massive open online courses: a resource for health education in developing countries. Glob Health Promot 2018;25(3):74-76. [doi: 10.1177/1757975916680970] [Medline: 28134014]
23. Barteit S, Jahn A, Banda SS, Bärnighausen T, Bowa A, Chileshe G, et al. E-learning for medical education in sub-Saharan Africa and low-resource settings: viewpoint. J Med Internet Res 2019;21(1):e12449 [FREE Full text] [doi: 10.2196/12449] [Medline: 30626565]
24. Caitano AR, Gusmão CMG, Dias-Trindade S, Barbalho IMP, Morais PSG, Caldeira-Silva GJP, et al. Massive health education through technological mediation: analyses and impacts on the syphilis epidemic in Brazil. Front Public Health 2022;10:944213 [FREE Full text] [doi: 10.3389/fpubh.2022.944213] [Medline: 36238258]
25. Evans DP, Luffy SM, Parisi S, Del Rio C. The development of a massive open online course during the 2014-15 Ebola virus disease epidemic. Ann Epidemiol 2017;27(9):611-615. [doi: 10.1016/j.annepidem.2017.07.137] [Medline: 28844550]
26. Narra R, Sobel J, Piper C, Gould D, Bhadelia N, Dott M, et al. CDC safety training course for Ebola virus disease healthcare workers. Emerg Infect Dis 2017;23(13):S217-S224 [FREE Full text] [doi: 10.3201/eid2313.170549] [Medline: 29154748]
27. Nieder J, Nayna Schwerdtle P, Sauerborn R, Barteit S. Massive open online courses for health worker education in low- and middle-income countries: a scoping review. Front Public Health 2022;10:891987 [FREE Full text] [doi: 10.3389/fpubh.2022.891987] [Medline: 35903395]
28. Findyartini A, Greviana N, Hanum C, Husin JM, Sudarsono NC, Krisnamurti DGB, et al. Supporting newly graduated medical doctors in managing COVID-19: an evaluation of a massive open online course in a limited-resource setting. PLoS One 2021;16(9):e0257039 [FREE Full text] [doi: 10.1371/journal.pone.0257039] [Medline: 34506524]
29. Yilmaz Y, Sarikaya O, Senol Y, Baykan Z, Karaca O, Demiral Yilmaz N, et al. RE-AIMing COVID-19 online learning for medical students: a massive open online course evaluation. BMC Med Educ 2021;21(1):303 [FREE Full text] [doi: 10.1186/s12909-021-02751-3] [Medline: 34039344]
30. Strehlow MC, Johnston JS, Aluri KZ, Prober CG, Acker PC, Patil AS, et al. Evaluation of a massive open online course for just-in-time training of healthcare workers. Front Public Health 2024 Oct 01;12:1395931. [doi: 10.3389/fpubh.2024.1395931]
31. King M, Pegrum M, Forsey M. MOOCs and OER in the Global South: problems and potential. IRRODL 2018;19(5). [doi: 10.19173/irrodl.v19i5.3742]
32. Christensen G, Steinmetz A, Alcorn B, Bennett A, Woods D, Emanuel E. The MOOC phenomenon: who takes massive open online courses and why? SSRN. 2013 Nov 06. URL: https://ssrn.com/abstract=2350964 [accessed 2014-04-18]
33. Kizilcec RF, Halawa S. Attrition and achievement gaps in online learning. : Association for Computing Machinery; 2015 Presented at: L@S'15:Proceedings of the Second (2015) ACM Conference on Learning @ Scale; 2015 March 14; New York, NY, United States p. 57-66. [doi: 10.1145/2724660.2724680]
34. Kizilcec RF, Saltarelli AJ, Reich J, Cohen GL. Closing global achievement gaps in MOOCs. Science 2017;355(6322):251-252. [doi: 10.1126/science.aag2063] [Medline: 28104856]
35. Reich J, Ruipérez-Valiente JA. The MOOC pivot. Science 2019;363(6423):130-131. [doi: 10.1126/science.aav7958] [Medline: 30630920]
36. Castaño-Muñoz J, Kreijns K, Kalz M, Punie Y. Does digital competence and occupational setting influence MOOC participation? Evidence from a cross-course survey. J Comput High Educ 2017;29(1):28-46. [doi: 10.1007/s12528-016-9123-z]
37. Dridi MA, Radhakrishnan D, Moser-Mercer B, DeBoer J. Challenges of blended learning in refugee camps: when internet connectivity fails, human connection succeeds. IRRODL 2020;21(3):250-263. [doi: 10.19173/irrodl.v21i3.4770]
38. Winters N, Langer L, Geniets A. Scoping review assessing the evidence used to support the adoption of mobile health (mHealth) technologies for the education and training of community health workers (CHWs) in low-income and middle-income countries. BMJ Open 2018;8(7):e019827 [FREE Full text] [doi: 10.1136/bmjopen-2017-019827] [Medline: 30061430]
39. Karp Gershon S, Ruipérez-Valiente JA, Alexandron G. Defining and measuring completion and assessment biases with respect to English language and development status: not all MOOCs are equal. Int J Educ Technol High Educ 2021;18(1):41. [doi: 10.1186/s41239-021-00275-w]
40. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and facilitators of digital health among culturally and linguistically diverse populations: qualitative systematic review. J Med Internet Res 2023;25:e42719 [FREE Full text] [doi: 10.2196/42719] [Medline: 36853742]
41. Nkuyubwatsi B. Cultural translation in massive open online courses (MOOCs). European Commission for Open Education Europa. 2014. URL: https://figshare.le.ac.uk/articles/journal_contribution/Cultural_translation_in_Massive_Open_Online_Courses_MOOCs_/10127768/1 [accessed 2023-03-23]
42. Perryman LA, Hemmings-Buckler A, Seal T. Learning from TESS-India's approach to OER localisation across multiple Indian States. J Interactive Media Educ 2014;2014(2):7. [doi: 10.5334/jime.af]
43. Stratton C, Grace R. Exploring linguistic diversity of MOOCS: implications for international development. Proc Assoc Info Sci Tech 2016;53(1):1-10. [doi: 10.1002/pra2.2016.14505301071]

XSL•FO
RenderX

44. Grace R, Stratton C, Fonseca F. Content matters: how online language content gives rise to digital divides. Soc Sci Q 2019;100(6):1999-2016. [doi: 10.1111/ssqu.12691]
45. Altbach PG. MOOCs as neocolonialism: who controls knowledge? IHE 2014(75):5-7. [doi: 10.6017/ihe.2014.75.5426]
46. Rizvi S, Rienties B, Rogaten J, Kizilcec RF. Beyond one-size-fits-all in MOOCs: variation in learning design and persistence of learners in different cultural and socioeconomic contexts. Comput Hum Behav 2022;126:106973. [doi: 10.1016/j.chb.2021.106973]
47. Rizvi S, Rienties B, Rogaten J, Kizilcec R. Culturally adaptive learning design: a mixed-methods study of cross-cultural learning design preferences in MOOCs. In: Open World Learning. England, United Kingdom: Routledge; 2022:39-43.
48. Hansen JD, Reich J. Democratizing education? Examining access and usage patterns in massive open online courses. Science 2015;350(6265):1245-1248. [doi: 10.1126/science.aab3782] [Medline: 26785488]
49. Singh PM, Wight CA, Sercinoglu O, Wilson DC, Boytsov A, Raizada MN. Language preferences on websites and in Google searches for human health and food information. J Med Internet Res 2007;9(2):e18 [FREE Full text] [doi: 10.2196/jmir.9.2.e18] [Medline: 17613488]
50. Skinner NA, Sanders K, Lopez E, Sotz Mux MS, Abascal Miguel L, Vosburg KB, et al. Barriers to COVID-19 vaccine acceptance to improve messages for vaccine uptake in indigenous populations in the central highlands of Guatemala: a participatory qualitative study. BMJ Open 2023;13(1):e067210 [FREE Full text] [doi: 10.1136/bmjopen-2022-067210] [Medline: 36707110]
51. Garcia-Retamero R, Dhami MK. Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. Health Expect 2011;14 Suppl 1(Suppl 1):46-57 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00670.x] [Medline: 21323820]
52. Flood D, Rohloff P. Indigenous languages and global health. Lancet Glob Health 2018;6(2):e134-e135 [FREE Full text] [doi: 10.1016/S2214-109X(17)30493-X] [Medline: 29389530]
53. Alidou H, Glanz C. Action research to improve youth and adult literacy?: Empowering learners in a multilingual world. UNESCO Institute for Lifelong Learning. 2015. URL: https://repositorio.minedu.gob.pe/handle/20.500.12799/3535 [accessed 2023-04-15]
54. Abascal Miguel L, Lopez E, Sanders K, Skinner NA, Johnston J, Vosburg KB, et al. Evaluating the impact of a linguistically and culturally tailored social media ad campaign on COVID-19 vaccine uptake among indigenous populations in Guatemala: a pre/post design intervention study. BMJ Open 2022;12(12):e066365 [FREE Full text] [doi: 10.1136/bmjopen-2022-066365] [Medline: 36523220]
55. Bendezu-Quispe G, Torres-Roman JS, Salinas-Ochoa B, Hernández-Vásquez A. Utility of massive open online courses (MOOCs) concerning outbreaks of emerging and reemerging diseases. F1000Res 2017;6:1699 [FREE Full text] [doi: 10.12688/f1000research.12639.2] [Medline: 29259764]
56. Zhao Y, Samo G, Utunen H, Stucke O, Gamhewage G. Evaluating complexity of digital learning in a multilingual context: a cross-linguistic study on WHO's emergency learning platform. Stud Health Technol Inform 2021;281:516-517. [doi: 10.3233/SHTI210222] [Medline: 34042628]
57. Jones J, Johnston JS, Ndiaye NY, Tokar A, Singla S, Skinner NA, et al. Health care workers' motivations for enrolling in massive open online courses during a public health Emergency: evidence from a case study of COVID-19 training. JMIR Med Educ 2024;10(11):e51915 [FREE Full text] [doi: 10.2196/51915] [Medline: 38904474]
58. World development indicators. DataBank. URL: https://databank.worldbank.org/reports.aspx?source=world-development-indicators [accessed 2024-11-02]

Abbreviations

HIC: high-income country
LIC: low-income country
LMIC: lower-middle-income country
MOOC: massive open access online course
UMIC: upper-middle-income country
WHO: World Health Organization

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 20.09.23; peer-reviewed by Z Li, JS Chang, A Hidki; comments to author 12.03.24; revised version received 30.06.24; accepted 23.10.24; published 10.01.25.

Please cite as:
Johnston JS, Skinner NA, Tokar A, Arabi E, Ndiaye NY, Strehlow MC, Utunen H
Global Use, Adaptation, and Sharing of Massive Open Online Courses for Emergency Health on the OpenWHO Platform: Survey Study
J Med Internet Res 2025;27:e52591
URL: https://www.jmir.org/2025/1/e52591
doi:10.2196/52591
PMID:39792445

©Jamie Sewan Johnston, Nadine Ann Skinner, Anna Tokar, Elham Arabi, Ngouille Yabsa Ndiaye, Matthew Charles Strehlow, Heini Utunen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Comparing Virtual Reality–Based and Traditional Physical Objective Structured Clinical Examination (OSCE) Stations for Clinical Competency Assessments: Randomized Controlled Trial

Tobias Mühling[1], MD; Verena Schreiner[1]; Marc Appel[1], PhD; Tobias Leutritz[1], PhD; Sarah König[1], MD

Institute of Medical Teaching and Medical Education Research, University Hospital Würzburg, Würzburg, Germany

Corresponding Author:
Tobias Mühling, MD
Institute of Medical Teaching and Medical Education Research
University Hospital Würzburg
Joseph-Schneider-Straße 2
Würzburg, 97080
Germany
Phone: 49 931 201 70627
Email: muehling_t@ukw.de

Abstract

Background: Objective structured clinical examinations (OSCEs) are a widely recognized and accepted method to assess clinical competencies but are often resource-intensive.

Objective: This study aimed to evaluate the feasibility and effectiveness of a virtual reality (VR)–based station (VRS) compared with a traditional physical station (PHS) in an already established curricular OSCE.

Methods: Fifth-year medical students participated in an OSCE consisting of 10 stations. One of the stations, emergency medicine, was offered in 2 modalities: VRS and PHS. Students were randomly assigned to 1 of the 2 modalities. We used 2 distinct scenarios to prevent content leakage among participants. Student performance and item characteristics were analyzed, comparing the VRS with PHS as well as with 5 other case-based stations. Student perceptions of the VRS were collected through a quantitative and qualitative postexamination online survey, which included a 5-point Likert scale ranging from 1 (minimum) to 5 (maximum), to evaluate the acceptance and usability of the VR system. Organizational and technical feasibility as well as cost-effectiveness were also evaluated.

Results: Following randomization and exclusions of invalid data sets, 57 and 66 participants were assessed for the VRS and PHS, respectively. The feasibility evaluation demonstrated smooth implementation of both VR scenarios (septic and anaphylactic shock) with 93% (53/57) of students using the VR technology without issues. The difficulty levels of the VRS scenarios (septic shock: P=.67; anaphylactic shock: P=.58) were comparable to the average difficulty of all stations (P=.68) and fell within the reference range (0.4-0.8). In contrast, VRS demonstrated above-average values for item discrimination (septic shock: r'=0.40; anaphylactic shock: r'=0.33; overall r'=0.30; with values >0.3 considered good) and discrimination index (septic shock: D=0.25; anaphylactic shock: D=0.26; overall D=0.16, with 0.2-0.3 considered mediocre and <0.2 considered poor). Apart from some hesitancy toward its broader application in future practical assessments (mean 3.07, SD 1.37 for VRS vs mean 3.65, SD 1.18 for PHS; P=.03), there were no other differences in perceptions between VRS and PHS. Thematic analysis highlighted the realistic portrayal of medical emergencies and fair assessment conditions provided by the VRS. Regarding cost-effectiveness, initial development of the VRS can be offset by long-term savings in recurring expenses like standardized patients and consumables.

Conclusions: Integration of the VRS into the current OSCE framework proved feasible both technically and organizationally, even within the strict constraints of short examination phases and schedules. The VRS was accepted and positively received by students across various levels of technological proficiency, including those with no prior VR experience. Notably, the VRS demonstrated comparable or even superior item characteristics, particularly in terms of discrimination power. Although challenges remain, such as technical reliability and some acceptance concerns, VR remains promising in applications of clinical competence assessment.

(J Med Internet Res 2025;27:e55066) doi:10.2196/55066

XSL•FO
RenderX

KEYWORDS

virtual reality; VR; objective structured clinical examination; OSCE; medical education; technological proficiency; assessment; clinical competence; item characteristics; discrimination power; acceptance; technical feasibility; effectiveness; comparative study; physical stations; medical students; standardized patients; cost-effectiveness

Introduction

Objective structured clinical examinations (OSCEs), first described in 1975 [1], have long since been recognized and accepted as a reliable, valid, and objective method to assess clinical competencies in medical education. They are organized in a circuit format and use stations featuring standardized cases and predefined assessment criteria using checklists or global rating scales to objectify the evaluation by the assessor. By breaking down the clinical tasks into multiple subtests, various skills aligned with learning objectives can be evaluated simultaneously [2]. Presently, this examination format is administered using either standardized patients (SPs) or simulators. SPs are trained actors who are provided with specific cases and then present themselves as patients to students [3]. Students using this method can exhibit their proficiency at the third level, “shows how,” of the competency framework by Miller [4]. This surpasses traditional formats such as oral or written examinations, which generally focus on Miller’s second competency level, “knows how,” linked to the application of knowledge. Nevertheless, OSCEs are subject to a number of significant limitations, with one of the primary issues being their resource-intensive nature, both in terms of materials and personnel [2]. The time-intensive training and deployment of SPs also cause significant financial expense. In addition, in terms of content, OSCEs may not comprehensively capture the complexity of emergency scenarios. Using SPs often proves inadequate at accurately simulating the complex pathophysiology found in living organisms. In particular, healthy actors may struggle to depict diseases in all their nuances, and invasive procedures cannot be executed.

Virtual reality (VR) simulation as a supplementary method in medical education has received a high degree of acceptance among learners and shown promising results in terms of learning outcomes [5,6]. Importantly, the use of VR-based scenarios in assessments may potentially overcome the aforementioned limitations: Virtual patients can display symptoms and findings that realistically represent illnesses, and a computed dynamic physiology can replicate appropriate responses to medication or invasive interventions such as endotracheal intubation or administration of catecholamines. Additionally, digitally assisted assessment offers the possibility of relieving examiners through automated and objective recording of the results. Considering these advantages, using VR scenarios of medical emergencies in OSCE stations appears promising, but evidence relating to technical reliability and cost-effectiveness remains limited [7]. Indeed, to ensure smooth examinations, a sufficient level of hardware and software maturity is essential to avoid interruptions resulting from technical issues. Furthermore, the technically available scenarios must align with the learning and assessment objectives of the respective curriculum.

Given these prerequisites, application examples in this domain are limited. A pilot study did showcase the successful use of VR-based training of cardiopulmonary resuscitation in an examination setting [8]. Another study reported that skills assessment using 360° videos delivered via VR head-mounted displays can yield valid outcomes [9]. We also introduced a VR-based simulation training course on complex medical emergencies into the curriculum at our institution in 2020 [10]. It has been continuously refined to meet the aforementioned technical requirements and is now a suitable candidate for use as an examination tool.

Although VR holds promise for medical assessment, research into large-scale OSCEs and technical feasibility is limited. Additionally, only a few studies address the didactic requirements, such as test quality and consistency of results. In this study, we aimed to determine whether the theoretical benefits of VR are realized within the tight schedule of an already established routine OSCE in the curriculum with a full cohort of students.

On this basis, we aimed to address the following: (1) whether it is both organizationally and technically feasible to integrate VR-based stations focused on emergencies within an existing curricular OSCE framework, (2) whether VR-based stations display item characteristics (such as item difficulty, item discrimination, discrimination index) comparable to their physical counterparts that test identical content, and (3) how students perceive and to what extent they accept the VR-based stations.

Methods

VR-Based Simulation Training

STEP-VR (version 0.13b; ThreeDee GmbH) was used as the VR-based simulation of complex emergencies together with the hardware setup and head-mounted displays essentially as described previously [10].

Study Design

The study was conducted at a medical school in Germany (University of Würzburg) at the end of the fifth year of study toward the degree of medicine. The already established curricular OSCE was designed traditionally as a circuit with 10 stations, with 2 circuits running parallel to each other to increase throughput and reduce the total examination duration to 2 days. Of the 10 stations, 5 were case-based focusing on a number of specialties, and 4 were skills-based. Central to the study was the tenth station, the medical emergencies station. This station was the only one available in 2 separate modes: either a VR-based or a real-world mode, which we designated as a VR-based station (VRS) or physical station (PHS), respectively. The VRS and PHS were designed to be visually and functionally as similar as possible, including the case description and task assignment for the students, case dynamics during simulation (expressed through changes in vital parameters), and functionality of the emergency room environment (Figure 1).

XSL•FO
RenderX

At each station, students were given 1 minute to read the case description and task, followed by 9 minutes to complete the examination. To mitigate the consequences of students potentially sharing information, scenarios in all stations were switched at differing intervals.

Figure 1. Representative scenes from the (A) virtual reality–based station (VRS) and (B) the physical station (PHS) counterpart.

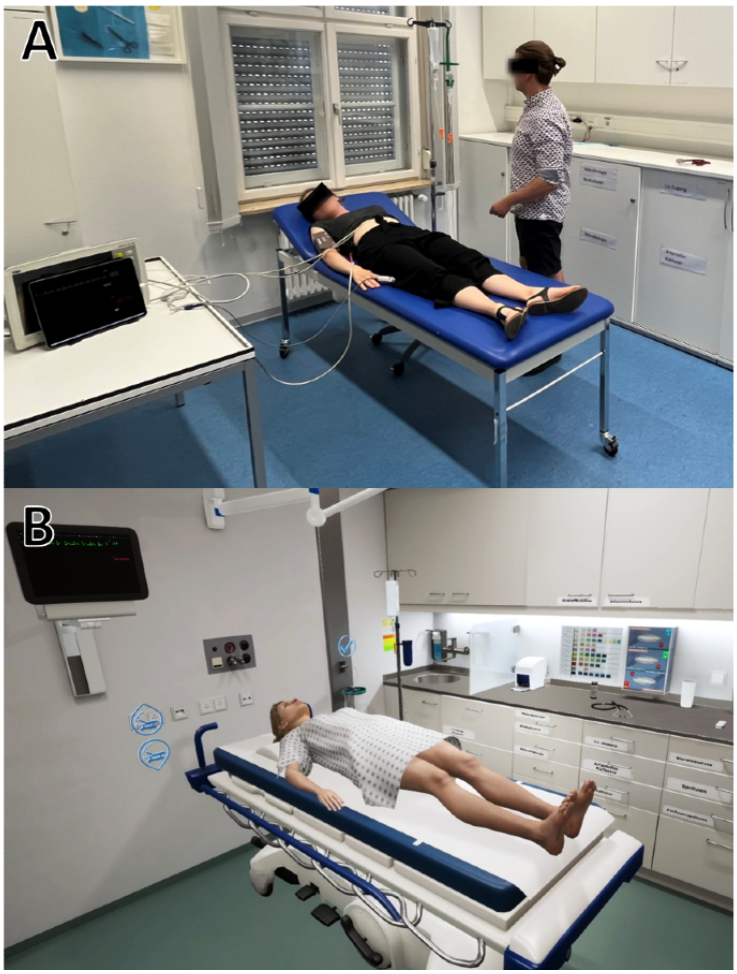

Throughout the semester, students received comprehensive preparation for the OSCE, with particular emphasis on its implementation of emergencies in VR. They were given a script detailing the acute treatment of different types of shock. Additionally, a tutorial video was made available to familiarize students with the VR equipment and software. All students participated in STEP-VR as part of mandatory 3-hour small group sessions. However, during these sessions not all students were active in the VR; some were observing through a screen displaying the first-person perspective. Furthermore, each student had the opportunity to practice using the system beforehand in a voluntary training session.

In the VRS and PHS on day 1, students encountered a patient with an initial diagnosis of fistulizing Crohn disease leading to septic shock. The assessment focused on the managing measures of the “One Hour Bundle” (ie, actions to treat sepsis to be executed within 1 hour) and the decision-making process for either interventional or surgical abscess drainage. On day 2, students faced the challenge of stabilizing a patient experiencing anaphylactic shock triggered by the painkiller metamizole. In addition to managing the associated respiratory distress, they were required to advise the patient on measures to prevent recurrence. All medical content was based on established guidelines and reviewed by experienced faculty members. At the outset of the OSCE, participants were randomly assigned to 1 of the 2 parallel and simultaneous circuits. Ultimately, each student had to undertake 1 of the 2 scenarios either within the VRS or the PHS. A backup VR setup was always on hand to address any technical issues.

In both the VRS and PHS, students were provided with all the essential information for the case in the task description (as a sign on the door prior to entering the station). This encompassed medical history, physical examination, and diagnostic test results. The subject and goal of the examination were focused on acute treatment (taking immediate actions) and making decisions for the next steps (correct indication for intervention or surgery or providing recommendations). Students were provided with assistance when donning the head-mounted display and controllers. The first-person perspective of the students was transmitted to a screen, allowing assessors to view the students’ actions. The performance was rated with standardized candidate assessment forms.

For students assigned to the VRS but who chose not to use VR technology owing to reservations (such as past instances of simulation sickness), a tutor took over the operation of the headset and controllers. The student could observe the scenario

XSL•FO
RenderX

on the screen and had to guide the tutor with the appropriate actions and measures through verbal commands. However, data from these students were not included in the study.

The 5 case-based stations (internal medicine, surgery, family medicine, pediatrics, gynecology) were comparable to the VRS and PHS. In the case-based stations, the scenarios were changed every half day, so that 4 different scenarios were used for each specialty. These also assessed students' management and clinical decision-making, especially with regard to diagnostic and therapeutic measures. The other 4 stations, which focused on procedural competence in highly standardized scenarios (eg, postoperative blood transfusion), were incomparable with the other stations and thus excluded from subsequent analysis in this study.

Randomization

Randomization was performed using computer-generated random numbers. The block size was set to 18 for each session to divide participants into 2 different circuits (including VRS and PHS). The randomization sequence was created prior to the study, and all OSCE stations were conducted within the randomized group allocation.

Blinding

VRS and PHS were substantially different, so that both participants and assessors were aware of the group allocation. Thus, blinding was not possible in this study.

Sample Size Calculation

As the number of participants was constrained by the total number of students available in the semester, sample size was not determined through a specific power analysis. Instead, a full census was conducted, including all available students at the end of their fifth academic year.

Feasibility Evaluation

Throughout the implementation process, various measures of organizational and technical feasibility [11] were evaluated. Building on extensive data from our group and others on the integration of VR scenarios in medical training, along with promising results from pilot studies on partially immersive VR-based assessments (eg, using 360° videos, as mentioned in the Introduction section), we opted to trial the approach with an entire cohort of students. Feasibility outcomes were documented across key areas, including acceptability, demand, implementation, practicality, adaptation, integration, expansion, and limited efficacy. Additionally, a cost comparison between the VRS and PHS modalities was conducted.

Collection of Performance and Evaluation Data

Assessment of Student Performance

To assess students in both the VRS and PHS as equally as possible, identical candidate assessment forms were used. They consisted of 10 items (septic shock) and 13 items (anaphylactic shock) from the categories of (1) additional monitoring and diagnostics, (2) treatment and definitive diagnosis, and (3) subsequent actions and advice. Scoring for each item was either binary (criteria met or not met, corresponding to 1 or 0 points) or ternary (criteria fully met, partially met, or not met, corresponding to 2, 1, or 0 points). When calculating the total sum, all items were equally weighted and normalized to a maximum of 1 point. The candidate assessment forms are outlined in Tables S1 and S2 in Multimedia Appendix 1.

Online Survey for Evaluation and Feedback From Students

Right after the OSCE, students were invited to participate in an online survey accessible via a QR code to be scanned and performed following the examination. The survey encompassed demographic parameters such as age and prior experience with VR technology. It also featured a total of 18 items divided into various topics: stress experience (3 items), usability (2 items), preparation (2 items), acceptance (5 items), subjective performance (3 items), and general rating (3 items to share views on the OSCE's value, its personal relevance, and its future inclusion). Items were rated on a 5-point Likert scale from "strongly agree" (5 points) to "strongly disagree" (1 point). Students could also provide qualitative feedback via 2 open-ended questions.

Analysis of Item Characteristics

The OSCE's item characteristics for both the VRS and PHS across the 2 scenarios as well as for the other 5 case-based stations, each comprising 4 distinct scenarios, were detailed. Parameters, such as difficulty P (average participant score at the station), discrimination r' (correlation of station scores with overall scores excluding that station), and discrimination index D (difference in difficulty between high and low performers based on the top and bottom 33rd percentiles) were computed for all OSCE stations, as suggested by Möltner et al [12]. Regarding item difficulty (P), values between 0.4 and 0.8 were considered ideal [12]. For item discrimination (r'), values above 0.3 were considered good, values between 0.2 and 0.3 were considered acceptable, and values below 0.2 were considered insufficient [12]. For the discrimination index (D), values above 0.4 were considered excellent, values between 0.3 and 0.4 were considered good, values between 0.2 and 0.3 were considered mediocre, and values below 0.2 were considered insufficient [12].

The computation of item statistics was carried out using R 4.3.1 [13]. For the sake of clear presentation, the average was provided for each of the other 5 case-based stations. Reporting was primarily conducted in accordance with the CONSORT-EHEALTH recommendations [14]. The corresponding checklist is provided in Multimedia Appendix 2.

Analysis of Survey Data

Student responses from the VRS and PHS were analyzed using descriptive statistics such as counts (both absolute and percentages), means, and SDs. Nominally scaled variables were compared using the chi-square test. For continuous survey data, such as aggregated Likert data, the Shapiro-Wilk test was initially used to assess normality. Since the assumption of normality was violated for most items, the Wilcoxon rank-sum test was used for comparisons between the groups (VRS vs PHS).

XSL•FO
RenderX

A "thematic analysis" approach was used to summarize the anonymous open-ended responses from VRS participants, following the method outlined in [15]. The free-text comments were reviewed by 2 independent researchers, who identified overarching categories through a consensus process. The comments were then categorized and quantified, with a representative example selected for each category.

Ethical Considerations

Human Subject Ethics Review Approval

The local institutional review and ethics board (Medical Ethics Committee at the University of Würzburg) judged the project as not representing medical or epidemiological research on human subjects and as such adopted a simplified assessment protocol. The project was approved without any reservation under the proposal number 20230323-03.

Informed Consent

Students were informed about the study, and their participation was voluntary. Written informed consent was obtained from all participants, who were also provided with information on data processing for the analysis and the publication of results. Contact details were supplied for participants wishing to withdraw their consent to data processing. The decision to participate or not had no consequences on the students' academic progress.

Privacy and Confidentiality

Survey data from the questionnaires were collected anonymously using the EvaSys platform (Evasys GmbH, Lüneburg, Germany). Data were processed and stored in accordance with local data protection laws.

Compensation Details

No compensation was provided to the participants.

Results

Student Participation and Feasibility Evaluation

In total, 134 students participated in the OSCE examination. We excluded 11 students from the final analysis for various reasons: technical problems with the VR equipment (n=4), VR procedures executed by a tutor instead of the student (n=4), and incomplete assessment forms (n=3). Thus, 123 participants completed either the VRS (n=57) or PHS (n = 66; Figure 2).

The feasibility evaluation (Multimedia Appendix 3) confirmed that the implementation of the VRS proceeded smoothly, without major organizational challenges. From a technical perspective, 53 of 57 students (93%) were able to use the VR technology without any issues. The transition to VR did not negatively impact the overall process nor performance of the OSCE, and the modular design of the OSCE stations facilitated seamless integration of the VRS. Regarding the cost comparison, approximately €750 (US $776.42) per semester was saved with the VRS than with the PHS, primarily due to lower expenses for SPs and consumables. However, these savings must be considered in light of the total development costs, which amounted to approximately €25,000 (US $25,880.50) for the 2 VR scenarios used.

Figure 2. Layout of the objective structured clinical examination (OSCE) and data collection methods (green) used in the study from both the virtual reality–based station (VRS) and physical station (PHS). The number of participants, along with those excluded (yellow), is indicated. ACLS: advanced cardiac life support; ECG: electrocardiogram.

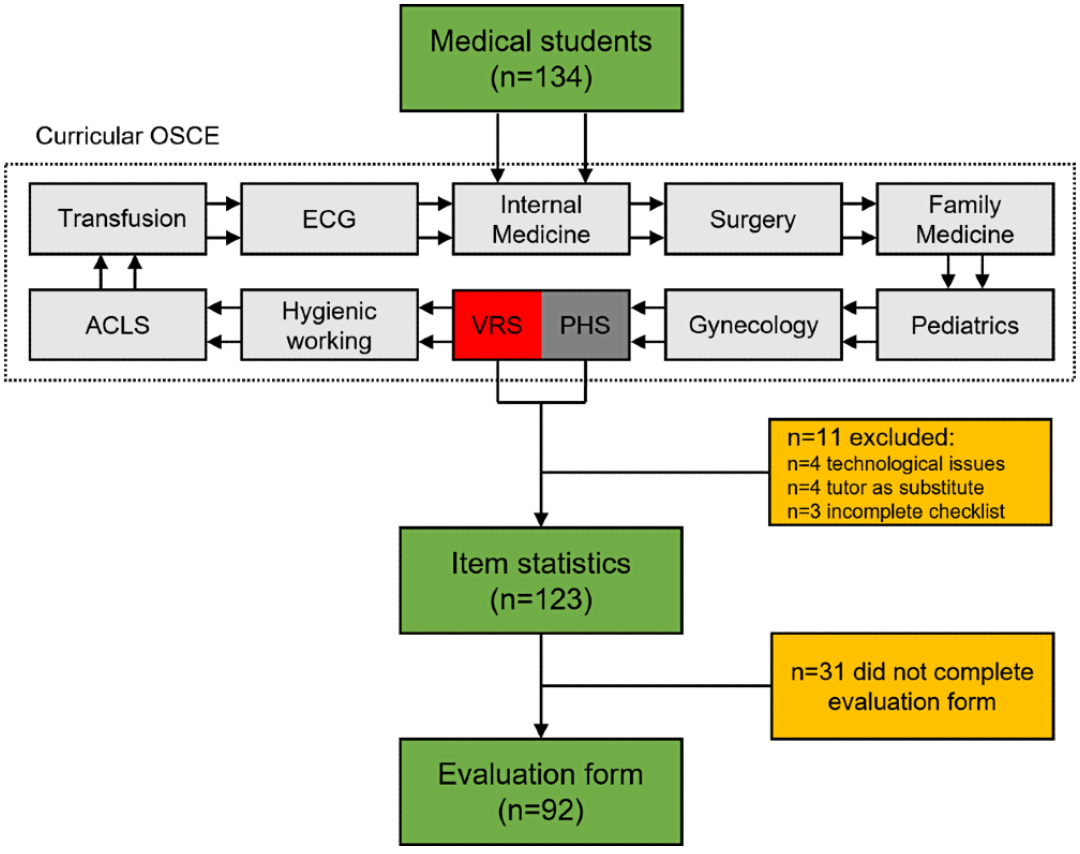

XSL•FO
RenderX

Performance Data and Item Characteristics

The characteristics of items across all stations are presented in Table 1. The overall average item difficulty P was 0.68. In VRS, septic shock and anaphylactic shock exhibited similar item difficulties at 0.67 and 0.58, respectively, which were comparable to the same scenarios in PHS (0.71 and 0.64, respectively). Additionally, the mean item difficulty for scenarios from the other 5 case-based stations consistently aligned at 0.71. Concerning item discrimination, the overall average was r'=0.30. The VRS scenarios proved to be above average with values of 0.40 (septic shock) and 0.33 (anaphylactic shock), in contrast to the PHS scenarios, for which below-average values of 0.12 and 0.25, respectively, were determined. The scenarios from surgery, pediatrics, and gynecology demonstrated values that were comparable or even better than those for the VRS, while others (internal medicine, family medicine) had significantly lower values. When calculating the discrimination index D, an average of 0.16 was observed across all stations. The VRS scenarios outperformed all other stations, achieving the highest values of 0.25 and 0.26. Conversely, the PHS scenarios fell below this average, with values of 0.10 and 0.12. The other case-based scenarios presented a diverse range of values, mirroring the subject-specific trends seen in item discrimination, with an average of around 0.15.

Table 1. Item characteristics of the objective structured clinical examination (OSCE) stations for the virtual reality–based station (VRS) and physical station (PHS) in the specific scenarios, as well as the other 5 case-based stations—each comprising 4 scenarios.

Scenarios	Item difficulty (P)	Item discrimination (r')	Discrimination index (D)
VRS			
Sepsis (n=32)	0.67	0.40	0.25
Anaphylaxis (n=25)	0.58	0.33	0.26
PHS			
Sepsis (n=31)	0.71	0.12	0.10
Anaphylaxis (n=35)	0.64	0.25	0.12
Internal medicine (n=134)	0.68	0.18	0.09
Surgery (n=134)	0.72	0.49	0.22
Family medicine (n=134)	0.78	0.12	0.08
Pediatrics (n=133)	0.67	0.44	0.18
Gynecology (n=134)	0.70	0.40	0.17
All stations	0.68[a]	0.30[a]	0.16[a]

[a]Mean value shown.

Quantitative and Qualitative Results of Survey Data

The online survey was completed by 92 participants following the OSCE, resulting in a 74.8% (92/123) response rate. Table 2 depicts the gender distribution of the participants, as well as their previous experience with emergency medicine and VR. No notable disparities were identified between the VRS and PHS groups.

Students rated their perceptions and attitudes toward the VRS or PHS (Table 3). Moderate scores were recorded for the perception of stress, with no significant difference between the VRS and PHS. Usability was scored favorably, and both formats were viewed as effective in demonstrating acquired skills. Students felt they only had a moderate level of preparation from the general curriculum for the station they completed. However, the specific materials provided enhanced their readiness. Students from both the VRS and PHS groups found their respective scenario to be a realistic portrayal of medical emergencies and clinically relevant. They rated the scenarios as manageable within the given time. Students believed that the station allowed them to demonstrate their skills and felt they handled the station well. Overall, students found the OSCE worthwhile as an examination format in medical school. They perceived the type of examination station they completed as meaningful. Notably, compared with the VRS, there was a significant preference among students for increased use of the PHS in future assessments.

XSL•FO
RenderX

Table 2. Participant gender and experience with emergency medicine, as well as characteristics of their experience with virtual reality (VR) categorized by the total sample, those using the VR-based station (VRS), and those using the physical station (PHS).

Characteristics	Total (n=92), n (%)	VRS (n=43), n (%)	PHS (n=49), n (%)	*P* value[a]
Gender				.86
Male	39 (42)	18 (42)	21 (43)	
Female	52 (57)	25 (85)	27 (55)	
Diverse	1 (1)	0	1 (2)	
Experience in emergency medicine (eg, volunteer service)?				.74
Yes	20 (22)	10 (23)	10 (20)	
No	72 (78)	33 (77)	39 (80)	
Cumulative experience with VR applications (hours)				.29
0	23 (25)	14 (33)	9 (18)	
0-5	67 (73)	28 (65)	39 (80)	
6-10	2 (2)	1 (2)	1 (2)	
>10	0	0	0	
Use of the VR lab for preparation				.97
Yes	28 (30)	13 (30)	15 (31)	
No	64 (70)	30 (70)	34 (69)	

[a]Comparison between the VRS and PHS groups using the Wilcoxon rank-sum test.

XSL•FO
RenderX

Table 3. Students' perceptions and attitudes toward the 2 specific objective structured clinical examination (OSCE) modalities (virtual reality–based station [VRS] and physical station [PHS]).

Theme and items[a]	VRS (n=43), mean (SD)	PHS (n=49), mean (SD)	*P* value[b]
Stress			
1. I felt stressed because many aspects of the scenario were beyond my control.	2.28 (1.26)	2.51 (1.19)	.30
2. I felt stressed because I lacked sufficient medical knowledge to handle the case.	2.56 (1.14)	2.39 (1.06)	.61
3. The presence of the examination staff put pressure on me.	1.61 (0.98)	1.71 (1.17)	.87
Usability			
4. The equipment worked without any issues.	3.63 (1.18)	3.74 (1.17)	.75
5. I was able to perform the medical procedures as I had envisioned.	3.40 (1.20)	3.37 (1.24)	.92
Preparation			
6. I felt adequately prepared for the station through the curriculum lectures and courses.	2.67 (1.06)	2.69 (1.23)	.87
7. I felt adequately prepared for the station through the preparation materials.	3.23 (1.09)	3.61 (1.27)	.06
Acceptance			
8. The scenario felt realistic.	3.21 (1.32)	3.25 (1.32)	.89
9. I found the scenario manageable within the given time.	3.88 (1.12)	4.08 (0.89)	.54
10. The content of the scenario was clinically relevant.	4.47 (0.59)	4.45 (0.82)	.68
11. This type of station should be used more frequently as an examination format.	3.07 (1.20)	3.57 (1.40)	.04
12. Overall, I would rate the station as:[c]	3.58 (1.10)	3.78 (1.10)	.32
Performance			
13. The station provided me with an opportunity to demonstrate my learned skills.	3.61 (0.90)	3.61 (1.22)	.70
14. I was able to handle the examination station well.	3.40 (0.88)	3.53 (1.00)	.35
15. I would rate my performance on this station as:[c]	3.44 (0.80)	3.57 (0.84)	.40
General rating			
16. I believe OSCE examinations in medical school are generally worthwhile.	3.49 (1.32)	3.55 (1.21)	.92
17. I find the type of examination station I completed to be meaningful.	3.05 (1.27)	3.53 (1.24)	.07
18. This type of station should continue to be a part of the OSCE examination.	3.07 (1.37)	3.65 (1.18)	.03

[a]Rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

[b]Comparison between the VRS and PHS groups using the Wilcoxon rank-sum test.

[c]Rated on a 5-point Likert scale ranging from 1 (very poor) to 5 (very good).

We subsequently performed thematic analysis on the answers to the open-ended questions from the VRS participants to provide some context for the quantitative findings and to gain deeper insights into students' perspectives on using VR technology in examinations. From the open-ended responses, we identified both positive and negative themes (Table 4) encompassing usability, difficulty or fairness, preparation, practical relevance, and general feedback. Of note, there were nearly twice as many positive (n=58) as negative comments (n=30). Students predominantly praised the realism and fairness of the examination, with approximately 35% (20/58) and 30% (17/58), respectively, of the positive comments focused on these aspects. These themes also emerged less frequently in a negative context, often from students who faced challenges with time constraints or found some interactions in VR as abstract. Difficulties with the technology were identified by 8 participants, who experienced minor issues or challenges. Other notable positive feedback included a generally positive outlook toward the VRS and a strong appreciation for the relevance of the medical content.

XSL•FO
RenderX

Table 4. Summary of positive and negative feedback on the virtual reality–based station (VRS) from the open-ended questions.

Theme	Responses, n (%)[a]	Example quote
Positive themes (n=58)		
Realism	20 (34)	"The setting was realistic and provided ample room for action."
Fairness	17 (29)	"The station was manageable within the time frame. There was [technical] assistance available when needed."
General	7 (12)	"I had fun during the examination."
Relevance	6 (10)	"Responding in an emergency situation is a highly relevant content."
Usability	5 (9)	"The operation was reliable and intuitive."
Preparation	3 (5)	"The preparation served as good practice."
Negative themes (n=30)		
Realism	9 (30)	"The medications and procedures could be made somewhat more realistic."
Fairness	9 (30)	"I would have needed more time and more support."
General	2 (7)	"The VR[b] station should not be part of a graded examination."
Relevance	0	N/A[c]
Usability	8 (27)	"The handling could be optimized. Additionally, I experienced technical disruptions, which could also be addressed."
Preparation	2 (7)	"Better preparation through teaching is needed for medical emergencies."

[a]Multiple responses were allowed. The percentages refer to responses given in each specific category (positive or negative) and not to the total number of respondents.

[b]VR: virtual reality.

[c]N/A: not applicable.

Discussion

Principal Findings

In our study, we integrated a VRS featuring 2 complex emergency scenarios into the established OSCE for advanced medical students at our institution. We found that the VRS presented a comparable level of difficulty to the traditional PHS stations. Notably, the VRS was more effective in distinguishing between high and low-performing students, as demonstrated by higher item discrimination and discrimination index values. Many students, including first-time VR users, praised the VRS for its realistic portrayal of medical emergencies and the fair assessment conditions provided. However, some expressed hesitancy about the broader application of VR in future practical assessments.

The integration of the VRS into the existing OSCE framework proved feasible both technically and organizationally. This is evidenced by successful participation of 134 students in the OSCE, with 53 students (or their tutor representatives) completing the tasks on the VRS without substantial issues. Our feasibility evaluation suggests that VRS can be implemented both practically and efficiently, even within the constraints of tight examination schedules. Since this is the first description of a VRS being included in a curriculum-based OSCE for a full cohort, it is difficult to compare costs more broadly. A recent review discussed the expenses related to VR-based training, including scenario development costs [16]. Consistent with our findings and those of others [17], these initial development costs can be offset by substantial long-term savings in recurring expenses such as SPs and consumables. Furthermore, once developed, the VR scenarios can be reused in various contexts without additional costs.

Analysis of performance data revealed that the item statistics for the VRS were not only comparable to, but in some cases even superior to, those of the PHS and the physical case-based scenarios from the 5 medical disciplines. Such good item characteristics and high discrimination values align well with observations from other studies. Recently, the efficacy of VR-mediated 360° videos in differentiating various skill levels during assessments was highlighted [9]. Studies of telemedicine examinations during the COVID-19 era also suggest that the validity and reliability of digital examination tools can be strong [18,19]. Of note, the discrimination r' and discrimination index D of the VRS were superior to those of the PHS, despite the randomization of participants and rotation of examiners at the stations. One possible explanation for these findings is the more uniform setting offered by VR. This environment minimizes or even eliminates the variability [20] that might be introduced unintentionally by SPs, which can occur even when SPs are well-trained and experienced in their roles. Furthermore, the possibility exists that operating the VRS itself may constitute an additional task that high-performing students are more adept at handling, which could lead to greater differentiation in performance at these stations. In other words, there might be confounding variables (eg, spatial ability) related to achievement in VR environments [21], which could potentially enhance performance outcomes. This hypothesis warrants further investigation, such as examining the correlation between VR handling skills and overall academic performance. Such insights

XSL•FO
RenderX

could prove valuable in the design of more complex digital examinations and case-based assessments.

The study clearly demonstrates that students generally responded positively to the VRS, indicating a favorable attitude and willingness to engage with and benefit from VR technology. The realism and content relevance of the VR scenarios received above-average ratings and were frequently commended in the qualitative feedback. Fair assessment conditions were emphasized, and even stress experienced during the examinations was considered manageable. This aligns with the findings of recent studies indicating that students exhibit a positive attitude [22] toward VR-based teaching and assessment. Furthermore, they perceive VR as engaging and immersive, affecting learning outcomes positively [9].

Presumably, owing to the lower level of technical refinement in its implementation, another study discovered that students were still more likely to accept VR in classroom settings as opposed to its use in practical assessments [23,24]. Of note, the VRS here was accepted by students across various levels of technological proficiency, including those with no prior VR experience. Interestingly, a substantial majority (64/92, 70%) of the students chose not to use the offers of extra preparation toward the VR examination. Moreover, for 25% (23/92) of the participants, this was their first ever experience with a VR application. These findings highlight the viability of VR as an examination tool, accommodating students with a wide range of familiarity with technology. This study did not concentrate on the viewpoint of assessors; another study has explored the feasibility and benefits of using specific VR scenarios in OSCEs, receiving positive evaluation from assessors [18].

Nevertheless, a degree of hesitancy among participants was recorded regarding whether VRS should be used in future examinations: Agreement with this statement was clearly less than that for the PHS. This aligns with findings from another study, which indicated that students' reservations were primarily due to their lack of experience with VR technology [25]. The referenced study concluded that practical examinations using VR should only be considered once the technology is firmly established and has demonstrated reliability in educational contexts. Notably, some open-ended comments in our study expressed concerns regarding the potential increase in replacing human assessors and SPs with technology, aligning with findings from prior research [26]. A strategy to counteract this could involve clear communication that VR simulation is intended merely as an additional option to complement, not merely replace, existing examination formats. Reservation and reliance on technology during the early stages of implementation were also reasons why we still opted for manual recording during data collection, regardless of the fact that the VR software provided the capability of automated performance evaluation through a checklist. Nonetheless, the potential to use such automated features in the future to aid assessors in their demanding role is promising. This approach could represent a significant advancement in the efficiency and objectivity of future assessments, including approaches for formative feedback [27].

Strengths of the Study

One advantage of the study is the use of hardware and software that have been undergoing continuous evaluation in learning contexts since 2018, providing a high level of realism with minimal simulation sickness. Students were provided enough practice opportunities in the VR environment beforehand to minimalize operational issues (as demonstrated by good usability results). Nevertheless, the technology needs further rigorous development and enhancement to avoid issues in the examination context. Another strength of this study is the relatively large number of participants drawn from an entire semester cohort, which should generally be viewed as representative of the entire medical student population, including those with critical perspectives. The study's greatest strength, however, lies in the comparison of item statistics and questionnaires directly between the VRS and PHS. This approach allows for the separation of effects related to scenario content and those related to the modality, which can influence the overall acceptance of and performance within the examination. Additionally, comparing results with other medical disciplines in the same examination aids in assessing the overall performance of the students.

Limitations of the Study

One clear limitation is the restriction to a single institutional site and medical discipline. This is especially relevant to practical training, for which curricula can vary greatly across different faculties, thus limiting study generalizability. Another weakness is the absence of inferential statistical analysis that would correlate performance data with survey results. This was a consequence of the voluntary and anonymous nature of the questionnaires. Making participation in the survey mandatory may have overwhelmed the students and potentially affected acceptance of the VRS. Nevertheless, this analysis should be conducted in the future to identify students facing specific challenges with the VRS.

Conclusions

Our study successfully demonstrated that complex VR-based assessment scenarios can be integrated into an established OSCE. Compared with the similar PHS, the VRS exhibited a similar level of difficulty but showed more favorable discrimination metrics. Student acceptance was high, with no major significant differences between VR-based and physical OSCE stations. Consequently, we encountered no systematic issues that would hinder the broader adoption of VR in assessing clinical competence.

Acknowledgments

We would like to thank Andrew Entwistle for his critical language review of the manuscript.

XSL•FO
RenderX

Funding for this work was provided by the German nonprofit “Stiftung Innovation in der Hochschullehre” (Foundation for Innovation in Higher Education; grant number FRFMM-776/2022).

No generative AI was used in the writing of this manuscript.

Data Availability

The data sets generated during and analyzed during this study are available in Multimedia Appendix 4 and Multimedia Appendix 5.

Authors' Contributions

TM was responsible for the conceptualization of the study, acquiring funding, supervising the execution, analyzing the survey data, and writing the original draft. VS contributed to the conceptualization, conducted the investigation, and performed analysis of survey data. MA handled the curation of performance data, supervised the overall execution of the objective structured clinical examination (OSCE), and conducted the formal analysis. TL was involved in performance data curation; formal analysis; and the writing, review, and editing of the manuscript. SK also contributed to study conceptualization; performed formal analysis; and participated in writing, review, and editing of the manuscript.

Conflicts of Interest

TM was involved in the software development process for STEP-VR. All other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Candidate assessment form items.
[DOCX File , 22 KB - jmir_v27i1e55066_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist v1.6.1.
[PDF File (Adobe PDF File), 1045 KB - jmir_v27i1e55066_app2.pdf]

Multimedia Appendix 3
Feasibility evaluation and cost comparison.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v27i1e55066_app3.xlsx]

Multimedia Appendix 4
Raw performance data.
[XLSX File (Microsoft Excel File), 1226 KB - jmir_v27i1e55066_app4.xlsx]

Multimedia Appendix 5
Raw survey data.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v27i1e55066_app5.xlsx]

References

1. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975 Feb 22;1(5955):447-451 [FREE Full text] [doi: 10.1136/bmj.1.5955.447] [Medline: 1115966]
2. Barman A. Critiques on the objective structured clinical examination. Ann Acad Med Singap 2005 Sep;34(8):478-482 [FREE Full text] [Medline: 16205824]
3. Peters T, Sommer M, Fritz AH, Kursch A, Thrien C. Minimum standards and development perspectives for the use of simulated patients - a position paper of the committee for simulated patients of the German Association for Medical Education. GMS J Med Educ 2019;36(3):Doc31 [FREE Full text] [doi: 10.3205/zma001239] [Medline: 31211226]
4. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
5. Kim H, Kim E. Effects of medical education program using virtual reality: a systematic review and meta-analysis. Int J Environ Res Public Health 2023 Feb 22;20(5):1 [FREE Full text] [doi: 10.3390/ijerph20053895] [Medline: 36900904]
6. Liu JYW, Yin Y, Kor PPK, Cheung DSK, Zhao IY, Wang S, et al. The effects of immersive virtual reality applications on enhancing the learning outcomes of undergraduate health care students: systematic review with meta-synthesis. J Med Internet Res 2023 Mar 06;25:e39989 [FREE Full text] [doi: 10.2196/39989] [Medline: 36877550]
7. Thakker A, Devani P. Is there a role for virtual reality in objective structured clinical examinations (OSCEs)? MedEdPublish (2016) 2019 Sep 17;8:180 [FREE Full text] [doi: 10.15694/mep.2019.000180.1] [Medline: 38089327]

XSL•FO
RenderX

8. Rodríguez-Matesanz M, Guzmán-García C, Oropesa I, Rubio-Bolivar J, Quintana-Díaz M, Sánchez-González P. A new immersive virtual reality station for cardiopulmonary resuscitation objective structured clinical exam evaluation. Sensors (Basel) 2022 Jun 29;22(13):1 [FREE Full text] [doi: 10.3390/s22134913] [Medline: 35808422]
9. Knudsen MH, Breindahl N, Dalsgaard T, Isbye D, Mølbak AG, Tiwald G, et al. Using virtual reality head-mounted displays to assess skills in emergency medicine: validity study. J Med Internet Res 2023 Jun 06;25:e45210 [FREE Full text] [doi: 10.2196/45210] [Medline: 37279049]
10. Mühling T, Späth I, Backhaus J, Milke N, Oberdörfer S, Meining A, et al. Virtual reality in medical emergencies training: benefits, perceived stress, and learning success. Multimedia Systems 2023 May 12;29(4):2239-2252. [doi: 10.1007/s00530-023-01102-0]
11. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
12. Möltner A, Schellerg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. GMS Zeitschrift für Medizinische Ausbildung 2006;23(3):11 [FREE Full text]
13. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. URL: https://www.R-project.org/ [accessed 2025-01-02]
14. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
15. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18 [FREE Full text] [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
16. Abbas JR, Chu MMH, Jeyarajah C, Isba R, Payton A, McGrath B, et al. Virtual reality in simulation-based emergency skills training: A systematic review with a narrative synthesis. Resusc Plus 2023 Dec;16:100484 [FREE Full text] [doi: 10.1016/j.resplu.2023.100484] [Medline: 37920857]
17. Farra SL, Gneuhs M, Hodgson E, Kawosa B, Miller ET, Simon A, et al. Comparative cost of virtual reality training and live exercises for training hospital workers for evacuation. Comput Inform Nurs 2019 Sep;37(9):446-454 [FREE Full text] [doi: 10.1097/CIN.0000000000000540] [Medline: 31166203]
18. Arrogante O, López-Torre EM, Carrión-García L, Polo A, Jiménez-Rodríguez D. High-fidelity virtual objective structured clinical examinations with standardized patients in nursing students: an innovative proposal during the COVID-19 pandemic. Healthcare (Basel) 2021 Mar 20;9(3):1 [FREE Full text] [doi: 10.3390/healthcare9030355] [Medline: 33804700]
19. Lan Y, Chen W, Wang Y, Chang Y. Development and preliminary testing of a virtual reality measurement for assessing intake assessment skills. Int J Psychol 2023 Jun;58(3):237-246. [doi: 10.1002/ijop.12898] [Medline: 36720650]
20. Wendling AL, Halan S, Tighe P, Le L, Euliano T, Lok B. Virtual humans versus standardized patients: which lead residents to more correct diagnoses? Acad Med 2011 Mar;86(3):384-388 [FREE Full text] [doi: 10.1097/ACM.0b013e318208803f] [Medline: 21248598]
21. Lee EA, Wong KW. Learning with desktop virtual reality: Low spatial ability learners are more positively affected. Computers & Education 2014 Oct;79:49-58. [doi: 10.1016/j.compedu.2014.07.010]
22. Mahling M, Wunderlich R, Steiner D, Gorgati E, Festl-Wietek T, Herrmann-Werner A. Virtual reality for emergency medicine training in medical school: prospective, large-cohort implementation study. J Med Internet Res 2023 Mar 03;25:e43649 [FREE Full text] [doi: 10.2196/43649] [Medline: 36867440]
23. Tsekhmister YV, Konovalova T, Tsekhmister BY, Agrawal A, Ghosh D. Evaluation of virtual reality technology and online teaching system for medical students in Ukraine during COVID-19 pandemic. Int. J. Emerg. Technol. Learn 2021 Dec 08;16(23):127-139 [FREE Full text] [doi: 10.3991/ijet.v16i23.26099]
24. Wang N, Abdul Rahman MN, Lim B. Teaching and curriculum of the preschool physical education major direction in colleges and universities under virtual reality technology. Comput Intell Neurosci 2022;2022:3250986 [FREE Full text] [doi: 10.1155/2022/3250986] [Medline: 35310594]
25. Walter S, Speidel R, Hann A, Leitner J, Jerg-Bretzke L, Kropp P, et al. Skepticism towards advancing VR technology - student acceptance of VR as a teaching and assessment tool in medicine. GMS J Med Educ 2021;38(6):Doc100 [FREE Full text] [doi: 10.3205/zma001496] [Medline: 34651058]
26. Talbot T, Rizzo A. Virtual Human Standardized Patients for Clinical Training. In: Bouchard S, editor. Virtual Reality for Psychological and Neurocognitive Interventions. Virtual Reality Technologies for Health and Clinical Applications. New York, NY: Springer; 2019:387-405.
27. Moore AG, Hu X, Eubanks JC, Aiyaz AA, McMahan RP. A Formative Evaluation Methodology for VR Training Simulations. 2020 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); March 22-26, 2020; Atlanta, GA. [doi: 10.1109/vrw50115.2020.00027]

Abbreviations

OSCE: objective structured clinical examination

XSL•FO
RenderX

SP: standardized patient

Edited by T de Azevedo Cardoso; submitted 01.12.23; peer-reviewed by YH Yin, A DiGiammarino, F Sendra-Portero; comments to author 16.07.24; revised version received 22.08.24; accepted 15.11.24; published 10.01.25.

Please cite as:
Mühling T, Schreiner V, Appel M, Leutritz T, König S
Comparing Virtual Reality–Based and Traditional Physical Objective Structured Clinical Examination (OSCE) Stations for Clinical Competency Assessments: Randomized Controlled Trial
J Med Internet Res 2025;27:e55066
URL: https://www.jmir.org/2025/1/e55066
doi:10.2196/55066
PMID:

©Tobias Mühling, Verena Schreiner, Marc Appel, Tobias Leutritz, Sarah König. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Identifying Research Priorities in Digital Education for Health Care: Umbrella Review and Modified Delphi Method Study

Alison Potter[1], MA; Chris Munsch[2], ChM; Elaine Watson[3], MSc; Emily Hopkins[4], MA; Sofia Kitromili[1], PhD; Iain Cameron O'Neill[5], MA; Judy Larbie[5], MSc; Essi Niittymaki[5], MA; Catriona Ramsay[6], BA; Joshua Burke[7], PhD; Neil Ralph[5], DClinPsy

[1]Technology Enhanced Learning, NHS England, Southampton, United Kingdom
[2]Technology Enhanced Learning, NHS England, Leeds, United Kingdom
[3]Technology Enhanced Learning, NHS England, Oxford, United Kingdom
[4]Knowledge Management Service, NHS England, Manchester, United Kingdom
[5]Technology Enhanced Learning, NHS England, London, United Kingdom
[6]Technology Enhanced Learning, NHS England, Newcastle upon Tyne, United Kingdom
[7]Manchester Foundation Trust, Manchester, United Kingdom

Corresponding Author:
Alison Potter, MA
Technology Enhanced Learning
NHS England
Explorer House
Adanac Drive
Southampton, SO16 0AS
United Kingdom
Phone: 44 01962 690405
Email: alison.potter13@nhs.net

Abstract

Background: In recent years, the use of digital technology in the education of health care professionals has surged, partly driven by the COVID-19 pandemic. However, there is still a need for focused research to establish evidence of its effectiveness.

Objective: This study aimed to define the gaps in the evidence for the efficacy of digital education and to identify priority areas where future research has the potential to contribute to our understanding and use of digital education.

Methods: We used a 2-stage approach to identify research priorities. First, an umbrella review of the recent literature (published between 2020 and 2023) was performed to identify and build on existing work. Second, expert consensus on the priority research questions was obtained using a modified Delphi method.

Results: A total of 8857 potentially relevant papers were identified. Using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, we included 217 papers for full review. All papers were either systematic reviews or meta-analyses. A total of 151 research recommendations were extracted from the 217 papers. These were analyzed, recategorized, and consolidated to create a final list of 63 questions. From these, a modified Delphi process with 42 experts was used to produce the top-five rated research priorities: (1) How do we measure the learning transfer from digital education into the clinical setting? (2) How can we optimize the use of artificial intelligence, machine learning, and deep learning to facilitate education and training? (3) What are the methodological requirements for high-quality rigorous studies assessing the outcomes of digital health education? (4) How does the design of digital education interventions (eg, format and modality) in health professionals' education and training curriculum affect learning outcomes? and (5) How should learning outcomes in the field of health professions' digital education be defined and standardized?

Conclusions: This review provides a prioritized list of research gaps in digital education in health care, which will be of use to researchers, educators, education providers, and funding agencies. Additional proposals are discussed regarding the next steps needed to advance this agenda, aiming to promote meaningful and practical research on the use of digital technologies and drive excellence in health care education.

(J Med Internet Res 2025;27:e66157) doi:10.2196/66157

XSL•FO
RenderX

KEYWORDS

digital education; health professions education; research priorities; umbrella review; Delphi; artificial intelligence; AI

Introduction

In just over two decades, the use of digital technology for delivering education and learning has increased globally [1]. The growth of digital education in health care has seen a similar trajectory internationally [2].

The adoption of digital technologies in health care education forms part of a global strategy to invest in the health workforce [3] and support the United Nations 2023 Agenda for Sustainable Development [4]. This agenda is echoed within the United Kingdom's National Health Service (NHS) Long Term Workforce Plan [5]. In its plan to address projected workforce shortages, the NHS has undertaken to expand health care education and training over the next 15 years. Informed by the Topol Review [6], the Long Term Workforce Plan acknowledges the need to embrace a wide range of learning methods to achieve the scale of education delivery that will be required [5]. These methods include the use of web-based learning, blended learning, hybrid learning, and simulation and immersive technologies. In addition, it is clear that easily accessible and usable artificial intelligence (AI) will have an increasing role in education [5].

The COVID-19 pandemic saw a surge in the use of digital technologies for health care education. Consequently, several bibliometric analyses illustrate how research activity across a variety of digital technologies significantly increased internationally [2,7-13]. Since 2020, research activity has shifted from rapid development during the early stages of the COVID-19 pandemic to a focus on quality improvement and opportunities for sustained use in the postpandemic period [8]. Zhang et al [8] state that the data project a trend of continued growth in digital education in health care globally, including increased application of advanced technologies. In the United Kingdom, the Council of Deans of Health highlighted in their report "Post-Pandemic Progress: Lessons Learnt in Healthcare Education" [14] that attitudes of educators and students toward the use of digital modalities have shifted. Universities UK [15] echoes this sentiment, recommending increased investment and further expansion of digital technologies in health professions education.

To successfully leverage the full potential of existing technologies for education delivery and understand the opportunities presented by advanced technologies, there is a requirement to ensure that learning needs are being effectively met [16] and that there is proven return on the investment needed to deploy these technologies. The Immersive Healthcare Collaboration [17] emphasized the importance of rigorous evaluation when adopting new technologies and highlighted the need for continued investment in digital education research. However, currently it is not clear where these research efforts and resources should be focused.

Recent studies on research priorities in health care digital education have mostly focused on surgery [18-25]. However, a recent literature review by Tudor Car et al [26] explored evidence across all health care disciplines and across multiple digital education modalities. The authors sought to broaden the scope of their inquiry beyond evidence of effectiveness and to identify where gaps in the evidence lay around other factors such as context, infrastructure, education, and learners [26]. In doing so, they constructed an evidence map, a contextual framework, and a potential research agenda. They offered a list of research questions that, at that time, were still to be fully addressed by the scientific literature [26]. While this work is both comprehensive and informative, it was conducted before or early in the COVID-19 pandemic era; all the reviewed papers were published between January 2014 and July 2020. With the massive increase in the use of digital education in response to the COVID-19 pandemic, the authors of this study felt that this work warranted revisiting and updating to uncover additional evidence gaps and new areas of inquiry that may have emerged since 2020.

Therefore, this study aimed to build on work by Tudor Car et al [26] and identify, within the conceptual framework, any new research priorities in health care digital education that have emerged since 2020. Identifying these research priorities would help shape future global research in digital education and guide the determination of priorities for activity and resource allocation within NHS England.

Methods

Overview

To identify research priorities in health care digital education, a 2-stage approach was used. First, an umbrella review was conducted to build on previous work, and second, experts were invited to reach consensus on the relevant research questions using a modified Delphi method. This 2-stage approach was used to ensure a balance between the published literature and test its applicability through expert opinion. Details of each stage are described in the following sections.

Umbrella Review

As a first step to identifying any additional research questions since the study by Tudor Car et al [26], an umbrella review (a review of systematic reviews) of the literature was conducted based on the method used by Tudor Car et al [26]. The systematic review and research syntheses checklist has been used to self-appraise the process on completion (Multimedia Appendix 1) [27].

Inclusion and Exclusion Criteria

Studies eligible for inclusion were peer-reviewed systematic reviews or meta-analyses with a clear focus on digital education for health professions education, which were published between 2020 and 2023 and written in English. The list of digital education technologies as described by Tudor Car et al [26] was used to guide eligibility screening. For convenience, definitions have been shared in Multimedia Appendix 2 (Textbox 1).

XSL•FO
RenderX

Textbox 1. Eligibility criteria.

Included (studies involving one or more of the following digital education technologies)

- Online digital education
- Massive open web-based course
- Mobile education
- Serious gaming and gamification
- Virtual reality
- Virtual patients
- High-fidelity manikins
- Blended education
- Augmented reality
- Extended reality

Excluded

- 3D printing (after considerable discussion, it was decided that 3D printing should not be considered a digital educational intervention because this intervention leads to an analogue product and is used as such)
- Primary research
- Conference abstracts
- Non-English language papers

To further assess eligibility for inclusion, all the following questions had to be answered in the affirmative:

1. Are the participants in the intervention preservice (ie, students) or in-service professionals (ie, after degree completion, including postgraduate trainees and those in independent practice)?
2. Are the participants from disciplines such as medicine, dentistry, nursing and midwifery, medical diagnostic and treatment technology, physiotherapy and rehabilitation, and pharmacy? (practitioners of traditional, alternative, and complementary medicine were excluded)
3. Is the educational intervention using digital technology?

Information Sources

A knowledge specialist (EW) searched the following major bibliographic databases for review papers meeting the inclusion criteria and published between July 2020 and April 2023: MEDLINE (Ovid), Embase (Ovid), CINAHL (EBSCO), and the Cochrane Library. In addition, the knowledge specialist (EW) searched ERIC, Google Scholar, and Research Gate using the following search terms: *systematic review, digital education, eLearning, health professionals*, and *health professions education.* All searches were conducted on April 21, 2023.

Search Strategy

The MEDLINE (Ovid) search strategy was identical to that used by Tudor Car et al [26] (Multimedia Appendix 3). Minor modifications were needed to convert the strategy for CINAHL (EBSCO) and the Cochrane Library.

Study Selection

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology was used to identify studies for inclusion. The search results were captured in RefWorks [28], and duplicates were removed. Pairs of independent reviewers screened the title and abstract of each paper against the eligibility criteria and predefined question set for inclusion. The full-text papers meeting the inclusion criteria were then retrieved and reviewed by pairs of independent reviewers. At both stages of screening, any discrepancy between reviewer pairs was resolved by a third independent reviewer.

Data Extraction

A data capture tool was developed in Jisc Online Surveys [29] and piloted by 8 reviewers (AP, CM, EH, SK, ION, EN and 2 volunteers from the NHS England technology enhanced learning team) using a sample paper, with refinements made to the tool before data extraction from all papers included in the review.

Data extracted were study characteristics and future research recommendations as set out by the authors. Future research recommendations were mapped to the conceptual domain framework devised by Tudor Car et al [26] and any previously identified research questions, where possible. If there was insufficient alignment with existing research questions, a new one was created. All papers with future research recommendations were reviewed a second time, with any discrepancies resolved by a third independent reviewer.

Analysis

AP and EW consolidated the extracted research questions by mapping them to existing questions, removing duplicates, and rephrasing those with similar meanings. New questions were then grouped and categorized by CM into overarching research questions. CM and AP discussed and reached a consensus on the final set of questions.

XSL•FO
RenderX

Delphi Consensus

Overview

A 4-round modified Delphi was conducted through January to March 2024 to arrive at 5 final research questions. The method consisted of a desktop exercise to formulate the Delphi survey from the results of the umbrella review (round 1), assessment (round 2), prioritization (round 3), and ranking (round 4). Each survey was created by AP and tested by members of the research team.

Round 1: Umbrella Review and Preparation of the Survey

The results of the umbrella review provided the material to compile a web-based survey of research questions for consideration by experts in later rounds. Research questions, previously identified in the umbrella review, meeting the inclusion criteria formed the basis of a web-based survey in round 2.

Inclusion criteria were as follows:

- The median number of papers identified and mapped to an existing research question in the study by Tudor Car et al [26] was 7.5, and this was used as an inclusion threshold.
- Following consolidation, all newly identified questions were included in the survey.

Round 2: Assessment Against Criteria

A total of 42 participants were recruited, including digital education professionals and simulation faculty, clinical educators, health and care workforce development and transformation leads, academics, and health care students, using the set criteria (Textbox 2). Recruitment was advertised via the Association for Learning Technologies' weekly news digest and a nationwide digital education mailing list for health professionals maintained by NHS England (Multimedia Appendix 4).

A web-based survey, created in JISC Online Surveys and comprising all research questions from round 1, was distributed to the expert panel.

Using an adapted set of criteria from the Delphi process used by Schneider et al [30], participants were invited to anonymously score each research question against 5 criteria (Table 1).

Research questions that met the consensus threshold formed the basis for round 3.

The consensus thresholds for Round 2 were as follows:

- Inclusion: >75% of respondents provide a positive result (3) on a 3-point Likert scale for all criteria.
- Exclusion: >75% of respondents provide a negative result (2) on a 3-point Likert scale for all criteria.
- Nonconsensus: when the proposed priority research question has met neither the inclusion nor the exclusion consensus thresholds.

Questions that met the inclusion criteria and those with nonconsensus were carried forward into round 3.

Textbox 2. Participant recruitment criteria per expert group.

Stakeholder group and inclusion criteria for expertise

- Technology enhanced learning (TEL) professional and simulation faculty: substantial experience working in TEL or simulation in the National Health Service or academia
- Educator: anyone in clinical practice who trains health care professionals (such as consultant or nurse educator)
- Workforce development and transformation lead: experience or knowledge of TEL in health and care education within their geography or domain
- Academic: working in academia, with at least 1 first or senior author paper in the field of TEL
- Learner: health care student, either final-year undergraduate or postgraduate student or trainee

Table 1. Scoring options for each research question in round 2.

Criteria	Scoring options		
Scientific merit	1=unable to respond	2=disagree	3=agree
Significance for workforce development	1=unable to respond	2=disagree	3=agree
Innovation	1=unable to respond	2=disagree	3=agree
Relevance to the LTWP[a]	1=unable to respond	2=disagree	3=agree
Feasibility for further study	1=unable to respond	2=disagree	3=agree

[a]LTWP: Long Term Workforce Plan.

Round 3: Prioritization

Participants' voting in this round was informed by knowledge of the consensus outcomes for each research question from round 2. A web-based survey was shared with participants, consisting of 2 sections: one with research questions that met the inclusion criteria for round 2 (including the percentage consensus agreement for each question) and another with questions from round 2 that did not reach consensus.

XSL•FO
RenderX

Participants were invited to score each research question again, but this time based on priority, with 1 score assigned to each question. The scoring options were as follows: 1=not to be studied, 2=low priority, 3=medium priority, and 4=high priority.

The consensus thresholds for Round 3 were as follows:

- Inclusion: ≥80% of respondents scored the research question as a 3 or 4
- Exclusion: 100% of respondents scored the research question as a 1 (not to be studied)
- Nonconsensus: when the proposed priority research question met neither the inclusion nor the exclusion consensus thresholds

Those meeting the inclusion criteria were taken forward for ranking in round 4.

Round 4: Ranking

A web-based meeting was conducted with participants, facilitated by CM and AP. Participants were sent results of the previous round in advance (Multimedia Appendix 5). The meeting offered the opportunity to discuss results from round 3 in small groups. Finally, participants were asked to select and rank their top 5 research questions independently via a web-based survey constructed in Slido [31], a web-based audience interaction tool that offers in-meeting and between-meeting polling and surveying. Participants unable to attend the meeting were invited to complete this exercise within 1 week of the meeting date. After this time, for each ranked question, points were assigned in a descending order, where higher ranks received more points, for example, a question ranked first place received 5 points, and each subsequent rank received 1 point less, down to 1 point for a question ranked fifth. The 5 research questions with the highest mean formed the final set of research priorities.

Finally, a narrative synthesis approach was chosen to summarize the diverse range of selected studies in a structured manner.

Ethical Considerations

The NHS Research Ethics screening tool was completed, and approval was not required for this study. However, key ethical requirements were considered when designing and conducting the Delphi. The NHS England Data Protection Impact Assessment was completed. Delphi participant recruitment was voluntary, with opt out being possible. Compensation was not provided. A participant information sheet outlining the purpose of the study, key contacts, their potential involvement with anticipated time commitment, and how data would be handled and stored and by whom was shared. Personal data were stored in a spreadsheet hosted on an organizational SharePoint and were only accessible to relevant members of the research team. Participants were not asked for any personal details in the Delphi surveys; instead, they were assigned a code number. This code number allowed the research team to keep track of response rates and to administrate any follow up.

Results

Umbrella Review

Study Selection

A total of 8857 potentially relevant papers were identified across all information sources. Following removal of duplicates, of the 8857 papers, 6044 (68.24%) were screened by title and abstract. Of these 6044 papers, 441 (7.3%) were assessed for eligibility via full-text review, leading to 217 (3.6%) papers being included in this review (Figure 1). Refer to Multimedia Appendix 6 for a complete list of key study characteristics for all included studies.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram charting identification of studies.

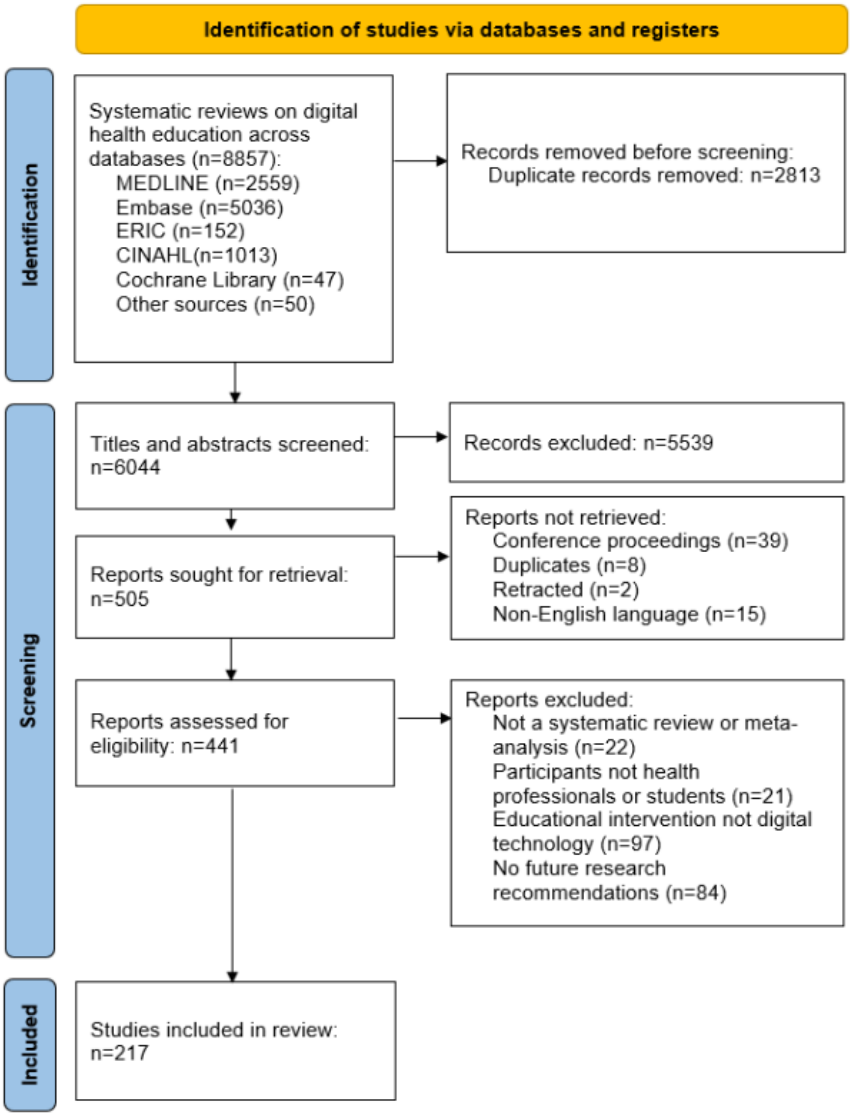

Study Characteristics

All included papers were systematic reviews or meta-analyses. Of the included studies across all the systematic reviews, high-income settings were predominant, with 23% (50/217) exclusively consisting of high-income countries and 33.6% (73/217) including a mix of high- and middle-income countries. In total, 14.3% (31/217) covered high-, middle-, and low-income countries, with only 0.9% (2/217) identifying middle- and low-income countries. The remaining 28.1% (61/217) of systematic reviews did not report the economic setting of their included studies.

In total, 29% (63/217) of the systematic reviews were focused on preservice alone. Of these, 11% (24/217) were nursing students, 5.1% (11/217) were medical students, and 7.8% (17/217) were mixed student populations. The least common population groups to be studied in the included reviews were pharmacy students (4/217, 1.8%), dentistry students (4/217, 1.8%), physiotherapy students (2/217, 1.8%), and health sciences students (1/217, 0.5%). This reflects the number of publications from these disciplines.

Reviews exploring the use of interventions with in-service professionals (44/217, 20.3%) are divided into fewer, less distinct categories: 13.8% (30/217) physicians, 6.5% (14/217) mixed health care professionals, and 0.5% (1/217) health care workers.

Of the systematic reviews comprising studies for both pre- and in-service professionals (110/217, 50.7%), 22.1% (48/217) focused on medical students and physicians, 17.5% (38/217) focused on mixed students and health care professionals, and 6% (14/217) focused on nursing students and nurses. Reviews comprising studies on dentistry, pharmacy, and physiotherapy were less common: 1.4% (3/217) dentistry students and dentists, 0.9% (2/217) pharmacy students and pharmacists, and 0.5% (1/217) physiotherapy students and physiotherapists.

The included reviews investigated digital education interventions for learners at a variety of stages in their career: 45.2% (98/217) were created for both students and graduates, 29.5% (64/217) for students only, 10.1% (22/217) for practicing postgraduate health care professionals, 7.8% (17/217) for a mix of trainees and postgraduate health care professionals, and 7.4% (16/217) for graduate trainees.

XSL•FO
RenderX

Figure 2 displays a chart showing the number of systematic reviews published for each digital modality, organized by publication year. Several additional digital modalities were identified in the systematic reviews over and above the modalities included in the study by Tudor Car et al [26]. These were consolidated into three new categories:

1. AI: this category includes traditional AI and generative AI, as well as novel AI-driven education modalities such as personalized learning, adaptive learning, AI-enabled extended reality (XR), chatbots, and virtual learning coaches. It also covers AI as an enabling technology, for example, when it helps educators produce content or generates data-driven insights.
2. XR and immersive technologies: this category includes augmented reality, XR, mixed reality, and haptic technology. Virtual reality was also reclassified to fall within this definition.
3. Simulation: this category includes high-fidelity simulators and robotic simulation.

Refer to Multimedia Appendix 2 for a full list of definitions.

Most systematic reviews focused on XR and immersive technologies (132/217, 60.8%) and online digital education (99/217, 45.6%). The least focus was given to serious gaming and gamification (16/217, 7.4%), simulation (15/217, 6.9%), massive online open courses (7/217, 3.2%), and AI (6/217, 2.8%).

Publication numbers were highest in 2022, although the annual totals between January 2020 and April 2023 (Figure 2) exceed those published in the original study by Tudor Car et al [26].

Mapped against the conceptual framework of digital health education for health care professionals [26], most systematic reviews were focused on education modality (170/217, 78.3%) and instructional design (99/217, 45.6%). Very few reviews focused on human resource requirements (7/217, 3.2%) and regulatory factors (6/217, 2.8%) in relation to digital education infrastructure, with only 0.9% (2/217) of reviews concerned with quality assurance of digital health education in the setting.

The majority of studies emphasized reporting educational outcomes related to skills (161/571, 28.1%) and knowledge (133/571, 23.3%), followed by studies focusing on satisfaction (77/571, 13.5%), attitude (54/571, 9.5%), and behavior (38/571, 6.7%). Few studies (29/571, 5.1%) reported patient outcomes.

Figure 2. Number of systematic reviews published per year up to April 2023 for each digital modality.

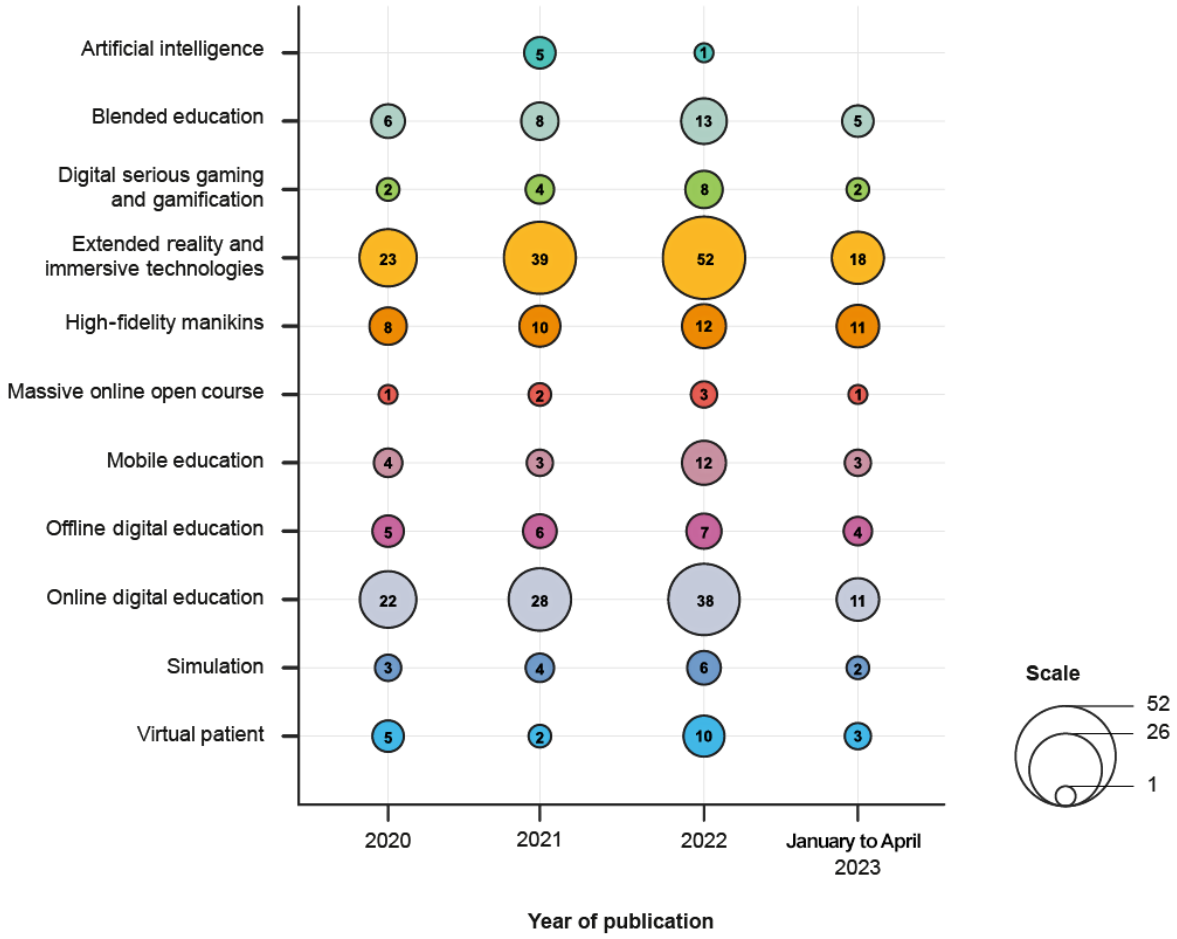

Future Research Recommendations

A total of 151 future research recommendations were extracted from the 217 systematic reviews. These were analyzed, recategorized, and consolidated by 3 authors (AP, CM, and EW) to create a final list of 63 updated research questions. Refer to Tables 2 and 3 for the complete lists.

XSL•FO
RenderX

Across the questions, online digital education (299/1257, 23.7%) and XR and immersive technologies (266/1257, 21.1%) were the most predominant digital modalities requiring further research. Most questions suggest further inquiry within the following domains of education: modality (21 questions), instructional design (16 questions), and assessment (14 questions).

The least predominant modalities are massive online open courses (20/1257, 1.59%) and AI (18/1257, 1.43%).

Table 2. New research questions identified from the included papers.

New research questions identified from the included papers	Evidence source references
Can the positive outcomes observed in studies at levels 2B and 3 of the Kirkpatrick model be further validated through studies at level 4, and what are the potential organizational and participant-level changes that may result from implementing these techniques?	[32]
How do immersive technologies impact learning outcomes?	[32-40]
How do we measure the learning transfer from digital education to the clinical setting?	[41-45]
How does research and evidence inform education commissioning and selection of digital technologies?	[46]
How does the design of VPS[a] impact learning outcomes in health professions education and training curriculum?	[47,48]
How does the use of social media as a pedagogical tool contribute to learner efficacy?	[49,50]
How effective is blended learning in different disciplines, age groups, and learning environments for nursing students?	[51]
How can digital health education be used to offer an inclusive learning experience for staff and students?	[52-55]
How can we optimize the use of AI[b], machine learning, and deep learning to facilitate education and training?	[56,57]
What are health care students' learning needs, and can they be met by the use of digital gamification?	[58-64]
What are the advantages and disadvantages of the recognized standard setting methods in simulation-based training?	[58,65-67]
What are the challenges of setting up digital simulation education in primary care?	[68]
What are the experiences and attitudes of teachers and trainers toward developing and implementing digital learning designs in health care education?	[69-71]
What are the outcomes and implications of high-quality randomized controlled trials evaluating the impact of virtual reality and augmented reality simulators on the entire robotic surgery learning curve?	[37,72]
What are the specific effects of video as a learning delivery method?	[73]
What automated performance metrics might prove effective in remote proctoring and assessment? For example, eye tracking	[61]
What is the comparative efficacy of various e-learning platforms in teaching specific subtopics when standardized guidelines for information delivery and assessment are used?	[74,75]
What is the effectiveness of digital health education on long-term retention of skills and knowledge?	[70,74,76-80]
What is the effectiveness of high-fidelity simulation compared with other forms of simulation training?	[81-84]
What is the effectiveness of immersive technologies compared with other forms of simulation training?	[67,80,85]
What is the effectiveness of synchronous digital education compared with asynchronous digital education to deliver health professions education?	[86-88]
What is the impact of digital health education on clinical outcomes in the short and long term?	[89-93]
What is the long-term impact of the COVID-19 pandemic on the educational gaps of health care learners?	[94-97]
What is the validity of using AI, machine learning, and deep learning to generate automated feedback for procedural skills training?	[98,99]

[a]VPS: virtual patient simulation.

[b]AI: artificial intelligence.

XSL•FO
RenderX

Table 3. Research questions identified from the included papers and mapped to questions originating from the study by Tudor Car et al [26].

Research questions identified from the included papers	Evidence source references
What are the instructional design barriers and facilitators that affect the continued adoption of digital tools in health professions education?	[63,69,88,100-104]
How does the frequency and duration of digital simulation–based psychomotor skills training affect health professionals' skills transfer to the clinical setting?	[105-111]
How does the interactivity of digital education programs affect the learning and clinical outcomes of health professionals?	[89,91,105,112-117]
What are the technical resources needed to deliver digital education to health care professionals?	[36,80,105,107,118-122]
Is spacing digital simulation–based training more or as effective as traditional education in clinical psychomotor skills development?	[41,80,105-109,111,117,123,124]
What are learners' acceptability of digital education with different levels of interactivity?	[87,91,105,107,115,122,125-129]
What is the differential impact of digital education on the clinical performance of trainee or expert surgeons?	[107,120,124-126,130-133]
Can digital education be used to overcome challenges in delivering content-specific topics for health professions education (eg, surgical training in rare pathologic states)?	[33,75,105,107,109,121,125,128,134-137]
Which components of digital health education (eg, interactivity and feedback) contribute to enhanced learning outcomes?	[55,58,61,80,89,115,125,128,129,138-140]
Can digital simulation–based training be used to train nontechnical skills in health professionals?	[32,85,97,111,123,126,128,131,136,141-143]
Is health professions' digital education more time efficient than traditional education?	[95,105,107,115,123-126,144-149]
Which learning theories can be used to inform the development of effective digital health professions education?	[62,63,69,74,81,87,126,150-157]
Does digital simulation–based psychomotor skills training provide any benefit to medical trainees?	[37,107,115,120,123-125,145,147,158-163]
How does health professions' digital education affect individual and health services outcomes and organizational practice?	[87,95,105,107,112,123,124,133,134,146,147,158,164-167]
What is the effectiveness of using digital education to train and assess nontechnical skills in health care professionals?	[42,82,84,88,95,111,117,128,131,143,168-170]
How do cost and cost-related outcomes influence the adoption of digital technology in health professions education?	[36,65,107,115,117,123,135,145,167,171-175]
Which features of digital education (eg, technical features, fidelity, safety, and adaptability) affect the learning outcomes of health professions education?	[36,82,104,105,112,122,124-126,129,130,134,139,142,145,171,176,177]
Is mastery learning via digital education more or as effective as traditional education in terms of clinical psychomotor skills improvement?	[34,35,80,105,108,109,115,123,124,134,147,155,158,178-181]
What are health professionals' attitudes toward digital delivery of education and training programs?	[55,69,70,87,89,91,105,107,112,115,122,125,126,128,146,148,174,182,183]
What are the challenges of digital education for health professionals training in different socioeconomic settings?	[32,68,122,140,171,172,175,184-186]
Can digital education be designed to achieve learning outcomes denoted in the Kirkpatrick model?	[32,40,74,87,107,112,117,124,128,133,134,140,141,147,155,156,164,187,188]
What are health care professionals' learning needs and can they be met using digital simulation training?	[88,104,105,107,115,117,120,122,123,125,126,128,129,138,146,148,158,162,164,183,189,190]
What type of instructional design is used in the effective digital education of health professions education?	[33,42,80,82,87,104,117,119,125,126,128,129,135,137,139,156,162,190-194]
What is the optimal duration, frequency, and intensity of digital health professions education programs to improve the learning and clinical outcomes for health care professionals?	[41,69,80,82,102,106-108,110,111,114,124,127,141,161,175,184,186,195]

XSL•FO
RenderX

Research questions identified from the included papers	Evidence source references
What is the long-term cost-effectiveness of digital education compared with traditional education for health professionals?	[35-37,48,78,81,85,91,95,105,107,115,120,123,148,159,167,169,173-175,196,197]
How does providing access to digital education improve the learning outcomes of health professionals?	[36,87,89,105,107,109,112,113,115,122-126,129,133,134,138,142,145,147,149,162,164,167,174,187]
How should learning outcomes in the field of health professions' digital education be defined and standardized?	[41,45,66,81,87,98,103,104,107,120,121,125,128,129,133,134,141,145,147,149,156,164,165,174,198-202]
What pedagogy should be used in the digital education of health professionals to improve their knowledge and skills?	[74,78,81-83,87,88,95,102,106,114,117,122,126,150,154-156,165,170,191,192,201-204]
Can digital education complement (ie, blended) or substitute traditional education for health professionals?	[34,41,43,56,61,64,80,87,91,95,96,104,115,119,124,126,128,140,145,154,156,187,205-209]
How can digital technology be incorporated into the current health professions education and training curriculum to improve learning outcomes?	[41,80,95,99,105,107,111,112,121,123-127,129,133,134,147,165,168,198,208,210-214]
What is the ideal approach to assessing health professionals' knowledge, skills, attitudes, satisfaction, and clinical outcomes from digital technology–based education and training programs?	[32,51,61,83,87,102-104,117,120,123,128,129,133,140,149,156,164,165,176,178,184,192,193,196,200,201,215-224]
How should studies on digital health professions education be reported?	[37,38,45,46,74,79,82,87,89,91,101,102,104,105,107,114,123,125,126,128,129,134,140,147,148,153,176,178,184,188,201,202,204,211,223,225-227]
Is digital education effective in improving health professionals' knowledge and skills performance in the clinical setting?	[36,44,62,87,95,104,105,107,109,113-115,117,123-126,131,133,134,144,145,147,149,154,158,162,164,176,187,190,192,194,196,209-211,213,215,228]
Which performance metrics or measurement instrument should be used to assess health professionals' knowledge, skills, attitudes, satisfaction, and clinical outcomes from digital technology–based training programs?	[33,41,46,51,58,61,66,82,87,98,102,104,107,109,120,123,124,128,129,133,140,141,145,147,153,156,164,166,169,170,178,179,191,199-201,211,215,220,222,229]
How does the design of digital education interventions (eg, format and modality) in health professions education and training curriculum affect learning outcomes?	[33,37,40-42,47,48,54-56,61,63,64,73,80-82,87,95,96,101-103,112,113,119,126,128,130,131,133,138,142,152,154,196,210,213,230-233]
How should studies of digital health professions education be designed to ensure the generalizability of their findings across different settings?	[38,51,52,62,76,79,80,87,91,92,102,104,105,107,114,115,120,123-126,128,129,133,134,138,140,141,144,145,147-149,164,170,172,176,184,200-202,205,207,211,223,225-227,234-238]
What is the impact of digital simulation–based training on clinical outcomes in the short and long term?	[35,37,40,48,68,70,76,81,87,93,106,108,112,117,123,124,128,131-134,140,143,146,147,155,164,165,174,179,181,185,186,188,196,197,201,218,227,228,236,239-241]
What is the effectiveness of digital education (mixed or single modality) compared with nondigital education to deliver health professions education?	[33,34,36,43,53,56,64,68,75,85,87,89,91,95-97,104,115,119,123,124,126,128,129,131,134,136,138,140,145,147,149,154,159,161-164,181,192,196,198,205-207,209,225,230,242,243]
What are the methodological requirements for high-quality, rigorous studies assessing the outcomes of digital health education?	[35,37-39,42,46,47,51,52,57,60,62,63,70,72,74,79-83,86,87,91,100,102,104,105,107,113-115,123-126,128,129,133,134,140,141,144,145,147-149,152,153,168,176,179,184,185,188-190,195,199,201-204,214,218,222,223,225-227,234,236,238,241,244-247]

Delphi Consensus

A total of 42 experts were enrolled to participate in the Delphi consensus activity. Of these, there were 11 academics, 18 clinical educators, 14 simulation faculty, 6 immersive technology professionals, 6 workforce and transformation professionals, 1 learner, and 13 TEL professionals.

In round 1, a total of 151 research questions identified from the literature review were consolidated into 63 questions. These 63 questions were then used to form the web-based survey for round 2.

In round 2, experts assessed each question against the set criteria (scientific merit, significance on workforce development, innovation, relevance to the Long Term Workforce Plan, and feasibility for further study). Two participants chose to withdraw from the Delphi process after reviewing the survey. There was a 78% (31/40) response rate, with nonconsensus thresholds across all questions. Therefore, all 63 questions went through to the next round.

In round 3, in total, 53% (21/40) experts responded to the survey to determine the priority with which each question ought to be pursued with further research. A total of 18 research questions met the consensus threshold for the final round.

In the fourth and final round, the response rate was 55% (22/40), and consensus was reached on the 5 research questions (Table 4) that the expert group identified as the most important and likely to impact the future delivery of digital education.

XSL•FO
RenderX

Table 4. Top-rated 5 research questions.

Research question	Sum of score
How do we measure the learning transfer from digital education to the clinical setting?	47
How can we optimize the use of AI[a], machine learning, and deep learning to facilitate education and training?	38
What are the methodological requirements for high-quality, rigorous studies assessing the outcomes of digital health education?	32
How does the design of digital education interventions (eg, format and modality) in health professions education and training curriculum affect learning outcomes?	31
How should learning outcomes in the field of health professions' digital education be defined and standardized?	30

[a]AI: artificial intelligence.

Discussion

Principal Findings

This study presents findings from 217 systematic reviews on digital education in health care, published between 2020 and 2023. The findings expand upon previous research conducted by Tudor Car et al [26] and define current gaps in the evidence for the efficacy of digital education. After analysis and consolidation, this study identified 63 future research questions: 24 new questions derived from the umbrella review and 39 from the original study by Tudor Car et al [26]. Most reviews focused on XR and immersive technologies as well as web-based digital education, targeting a mixed audience of students and in-service health care professionals. This was also the case for studies with the least commonly represented population groups (dentistry, pharmacy, and physiotherapy).

The 63 research questions were then prioritized through a consensus process with 40 experts from health care and academia, resulting in 5 key research questions (Table 4). These questions span several domains outlined in the conceptual framework [26]: assessment, engagement, learner, level of education, modality, and research.

The 5 Top-Rated Research Questions

Each of the top-rated questions is supported by a wealth of recommendations for conducting research in digital education. A high-level summary of this evidence is provided for each question.

Question 1: How Do We Measure the Learning Transfer from Digital Education to the Clinical Setting?

The evidence behind this research question points to several related yet distinct avenues of inquiry. A common theme among them is the need for further research to demonstrate how learning outcomes translate into real-world contexts, both for technical and nontechnical skills, particularly in nursing and medical fields [42-44]. One study specifically recommends further research to ensure that the competence and confidence in clinical practice skills demonstrated in a learning environment are maintained when applied in the clinical setting, where additional external factors may come into play [44]. Ensuring that the learning experience is relevant and closely mirrors real-world situations is a key consideration when designing instruction to facilitate effective transfer of knowledge and skills [248].

Furthermore, studies should track and evaluate the effectiveness of learning throughout a health care professional's career [41], using clinical outcomes and patient-reported outcome measures to assess the transfer of learning [45].

Question 2: How Can We Optimize the Use of AI, Machine Learning, and Deep Learning to Facilitate Education and Training?

The use of AI to optimize health education is a rapidly developing field. Current evidence proposes further research on the potential of this technology to elevate and shift performance assessment from benchtop and simulated environments to real-world context, offering more precise and scalable feedback mechanisms for both trainees and practitioners. For instance, machine learning has already shown promise in surgical skills training, where it could provide automated, objective assessments of technical skills, provided that more reliable assessment data were available [57]. Besides enhancing existing approaches to learning and assessment, AI also enables new modalities, such as personalized learning, which can enable more self-directed learning and may also support learner motivation by connecting data-driven insights to learning experiences, and adaptive learning [249]. AI integrated into personalized or adaptive learning could help students practice clinical decision-making by simulating different patient cases customized to their knowledge gaps based on previous performance. Furthermore, AI could generate dialogue, actions, and complex situations that replicate clinical settings, further enabling students to develop critical skills with reduced risk [56]. However, while AI offers substantial benefits, its implementation must be carefully considered to mitigate against potential challenges such as ethical concerns, bias from training data, and over-reliance on technology [249]. To safely and effectively adopt AI into health care education, a careful approach could integrate AI in a way that complements existing strategies that emphasize critical thinking and adaptability as well as technical proficiency.

Question 3: What Are the Methodological Requirements for High-Quality, Rigorous Studies Assessing the Outcomes of Digital Health Education?

Across the spectrum of digital education modalities, several studies highlighted the need for research to adopt appropriate study designs to determine learning gains [35, 63, 80, 87, 102,123,128,129,179,199,203,226]. Designs need to be larger in scale [39,60,105,107,113,124,134,140,188,214,218,246], multicentered [45,134,227,245], and carry out longer-term

XSL•FO
RenderX

follow-up [37,62,81,105,107,128,133,140,149,195,223,241] to ensure retention and transfer of learning. Many of the included studies recommended more randomized controlled trials [37,38,62,81,100,105,115,124,140,188,195,214,222,236,241], although one paper supports the suggestion provided by Tudor Car et al [211] that quasi-experimental designs may be better suited to educational research [144] and another study, in relation to cultural competency, suggests that observational studies and other designs may be preferred [52].

Question 4: How Does the Design of Digital Education Interventions (eg, Format and Modality) Used in Health Professions Education and Training Curriculum Affect Learning Outcomes?

Key recommendations from the data sources that derive this research question are primarily concerned with the integration and sequencing of pedagogical devices (ranging from flipped classrooms, timing of simulation briefing, e-learning, and active vs didactic approaches) in health education training and curricula as well as the quality of the approach used [41,42,82,95,102,119,133,142]. Some studies suggest delving further by closely comparing and evaluating the impact of nuanced difference in the design elements of an intervention or specific educational technology [37,63,82,101,130].

Stepping back from this detail within the intervention design itself, studies also reinforce the need to define educational needs, the context, and the setting to select suitable, appropriate, and effective modalities for specific learner groups [40,80,103,126,232].

Looking to the future, some studies highlight several novel technologies and suggest exploring their potential for designing more advanced educational interventions, such as AI [56], adaptive learning [87], wearable technology [213], eye tracking for assessment [55,61], and gamification to increase learner motivation [61].

While many still recommend further comparison studies between digital and nondigital interventions [33,48,119], others support the recommendation by Tudor Car et al [26] that comparisons between digital-to-digital interventions are now needed and necessary [142,231].

Question 5: How Should Learning Outcomes in the Field of Health Professions Digital Education Be Defined and Standardized?

Principally, the foundations of this research question lie within a body of evidence that urges a more standardized approach to outcome measurement, using validated measurement tools [81,98,120,141,165]. Beyond this, there is little guidance on how to achieve this, although 1 study (specific to surgical skills training) suggests that machine learning could help identify which metrics accurately assess skills [201].

In many ways, each of these research priorities are connected and interdependent. For example, to be able to measure learning transfer, learning outcomes need to be defined from the beginning and integrated into the learning design. This, in turn, will influence the research methodology used.

Comparison With Prior Work

This study builds upon and extends research by Tudor Car et al [26]. By doing so, the study not only updates their findings to include data up to 2023 but also extends them by prioritizing 5 research questions through Delphi consensus, thereby addressing a limitation previously identified by Tudor Car et al [26].

In comparison to the 6-year period covered by Tudor Car et al [26], the 3-year timeframe (2020 to 2023) in this study shows a dramatic increase in the number of publications on digital education in health care (from 77 to 217). This surge of activity is reflective of the rise in uptake of digital education in response to the COVID-19 pandemic.

Research with medical students and physicians has seen the greatest increase (from 12% to 22% of studies) and accounts for the largest portion of studies included in this review. In contrast, most earlier studies were more heterogeneous, with a combination of students and health professionals spanning several professions and specialties [26]. Dentistry staff and students remain one of the least represented groups in the literature, along with pharmacy and physiotherapy. The digital education modalities studied align with the broader outcomes of this study (namely, virtual reality, augmented reality, and web-based digital education), and their recommendations for future research mainly focus on the benefits, effectiveness, and perceptions of these modalities.

This study sees the addition of 3 new modalities: AI (in the form of personalized learning, adaptive learning, AI-enabled XR avatars, chatbots, and virtual learning coaches), XR and immersive technologies, and high-fidelity simulation. There has been an explosion of research around the use of XR technologies (132/217, 60.8% of included studies) as well as the emergence of AI (6/217, 3% of included studies). Aside from virtual reality, these technologies were not present in the earlier study.

There was a slight increase (from 5% to 14%) in the number of studies from high-, medium-, and low-income countries [26]. This could be indicative of greater collaborative relationships during the COVID-19 pandemic, as digital education was used globally to overcome the challenges of social distancing in health education delivery. Although evidence relating to digital education research in low-income countries continues to be sparce, the data from this study suggest that future research should focus more closely on learner audiences and intercultural differences to better understand what works across various socioeconomic settings [175,237].

While there have been noticeable shifts in the research populations and the education technologies studies since the findings shared by Tudor Car et al [26], some aspects remain unchanged. Studies comparing digital education interventions with traditional education remain popular, with a continued emphasis on the randomized control trial study design. Comparison studies between digital and nondigital approaches have made significant progress in recent years. Researchers are now encouraged to focus more on comparisons between digital methods in future studies to further advance digital education

XSL•FO
RenderX

research. However, despite a move toward research focusing on the context of the learner since the study by Tudor Car et al [26], this review has shown that there continues to be a lack of studies relating to quality assurance, human resources, and regulatory aspects of digital education infrastructure.

Strengths and Limitations

The findings of this extensive umbrella review replicated a robust methodology used by Tudor Car et al [26] and expanded it to identify the most pressing research questions currently needing attention in digital education within UK health care. Furthermore, these findings were strengthened by consultation with 40 stakeholders across health care and academia to arrive at consensus for the 5 research questions.

By replicating the methodology used by Tudor Car et al [26], this study was constrained by many of the same limitations set out in their study, particularly the exclusion of original research outside of systematic reviews. In addition, this study excluded conference abstracts and non-English papers. However, the impact of excluding non-English papers is considered minimal on the overall outcome [250], and while there may be advantages in including conference abstracts in a systematic review [251], they did not provide the level of information required for data extraction in this study.

There was very low representation from the learner stakeholder group on the Delphi expert panel. Initially, there was uncertainty among the research team members about whether learners would be classed as "experts" for a Delphi panel, which delayed promotion to this group. According to Nasa et al [252], an "expert can be defined as someone with knowledge and experience on a particular subject matter." In addition, there were similarities drawn with the concept of "experts by experience" [253]. While there are significant differences between being a patient, carer, or service user and a student or learner, the principles of "experts by experience" were considered to be similar. This approach involves gathering the views of people receiving the interventions, in this case, their views and expertise regarding their experience with education. Therefore, a heterogeneous group of stakeholders including final-year undergraduate health care students and postgraduate students or trainees, including doctors in training, was considered applicable to the aims of this study.

Next Steps

Having defined 5 high-level, key research priorities, it remains to determine what the next steps should be. Stefanidis et al [19] observed that, in the field of surgical simulation, little actual research evidence has been generated despite numerous papers attempting to define priorities. To avoid repeating this shortcoming, the next stage should be to identify specific research questions that have the greatest potential to have significant and practical impact across the system. As proposed by Stefanidis et al [19], establishing a number of expert working groups to devise specific, high-impact and feasible research proposals based on these 5 priorities is suggested. These ideas will subsequently be developed into formal research projects including research questions, collaborators, project design, and required resources. It is anticipated that successful completion of these projects will answer important questions about the practice of digital education in health care.

Conclusions

This review provides a list of current research gaps in digital education in health care, presenting them as a set of research questions along with the sources from which they are derived.

While the ongoing focus of digital education research on medical trainees and professionals is acknowledged, there is a need for more research involving nursing and allied health professions. It also highlights that there continues to be a need to address the quality assurance, human resources, and regulatory aspects of digital education infrastructure through research.

The prioritized set of 5 research questions, developed with input from expert representatives across health care and academia, is expected to serve as a valuable guide for researchers, funding agencies, and educators. For researchers, these questions offer a helpful starting point for addressing some of the existing gaps in knowledge. For funding agencies, they provide a framework for allocating resources to the areas identified as top priorities for the health care system. Finally, for educators and education providers, these questions can inform decision-making regarding implementation, curriculum design, and quality.

Acknowledgments

The authors would like to thank their volunteers (Angela Holmes, Claire Beattie, Harbie Rathore-Khatkar, Chris Jacobs, Ian Blacklock, Eleni Keogh, Sarah Gray, and Matt O'Brien) for their time and commitment throughout the process. The authors would also like to thank Nick Peres, Patricia Howe, and Faisal Mushtaq for their valuable advice and for helping to refine the methodological approach. The authors also thank the experts who participated in the Delphi process.

Conflicts of Interest

NR is the Deputy Director of Technology Enhanced Learning at NHS England. This team delivers digital learning services and products directly to the National Health Service (NHS) as well as providing system leadership and transformation programs.

Multimedia Appendix 1
JBI checklist for systematic review and research syntheses.
[DOC File , 125 KB - jmir_v27i1e66157_app1.doc]

XSL•FO
RenderX

Multimedia Appendix 2
Definitions of digital education technologies.
[DOC File , 47 KB - jmir_v27i1e66157_app2.doc]

Multimedia Appendix 3
Search strategy: Ovid MEDLINE.
[DOC File , 51 KB - jmir_v27i1e66157_app3.doc]

Multimedia Appendix 4
Participant recruitment advertisement.
[DOC File , 47 KB - jmir_v27i1e66157_app4.doc]

Multimedia Appendix 5
Round 4 preparatory work.
[DOC File , 103 KB - jmir_v27i1e66157_app5.doc]

Multimedia Appendix 6
Characteristics of the included systematic reviews.
[DOCX File , 426 KB - jmir_v27i1e66157_app6.docx]

References

1. Vaicondam Y, Sikandar H, Irum S, Khan N, Qureshi MI. Research landscape of digital learning over the past 20 years: a bibliometric and visualisation analysis. Int J Online Biomed Eng 2022 Jun 28;18(08):4-22 [FREE Full text] [doi: 10.3991/ijoe.v18i08.31963]
2. Sweileh WM. Global research activity on e-learning in health sciences education: a bibliometric analysis. Med Sci Educ 2021 Mar 02;31(2):765-775 [FREE Full text] [doi: 10.1007/s40670-021-01254-6] [Medline: 33680569]
3. Working for health and growth: investing in the health workforce - high-level commission on health employment and economic growth. World Health Organization. 2016 Dec 03. URL: https://www.who.int/publications/i/item/9789241511308 [accessed 2024-06-04]
4. Transforming our world: the 2030 agenda for sustainable development. United Nations Sustainable Development Goals Knowledge Platform. 2015. URL: https://sustainabledevelopment.un.org/post2015/transformingourworld/publication [accessed 2024-07-01]
5. NHS long term workforce plan. NHS England. URL: https://www.england.nhs.uk/publication/nhs-long-term-workforce-plan/ [accessed 2024-05-16]
6. The Topol review: preparing the healthcare workforce to deliver the digital future. The Topol Review, Health Education England. URL: https://topol.hee.nhs.uk/ [accessed 2024-07-01]
7. Jia K, Wang P, Li Y, Chen Z, Jiang X, Lin CL, et al. Research landscape of artificial intelligence and e-learning: a bibliometric research. Front Psychol 2022 Feb 16;13:795039 [FREE Full text] [doi: 10.3389/fpsyg.2022.795039] [Medline: 35250730]
8. Zhang P, Li X, Pan Y, Zhai H, Li T. Global trends and future directions in online learning for medical students during and after the COVID-19 pandemic: a bibliometric and visualization analysis. Medicine (Baltimore) 2023 Dec 15;102(50):e35377 [FREE Full text] [doi: 10.1097/MD.0000000000035377] [Medline: 38115375]
9. Martins J, Gonçalves R, Branco F. A bibliometric analysis and visualization of e-learning adoption using VOSviewer. Univers Access Inf Soc 2022 Dec 05:1-15 (forthcoming) [FREE Full text] [doi: 10.1007/s10209-022-00953-0] [Medline: 36530863]
10. Fauzi MA. E-learning in higher education institutions during COVID-19 pandemic: current and future trends through bibliometric analysis. Heliyon 2022 May;8(5):e09433 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e09433] [Medline: 35600443]
11. Yao S, Tang Y, Yi C, Xiao Y. Research hotspots and trend exploration on the clinical translational outcome of simulation-based medical education: a 10-year scientific bibliometric analysis from 2011 to 2021. Front Med (Lausanne) 2022 Feb 07;8:801277 [FREE Full text] [doi: 10.3389/fmed.2021.801277] [Medline: 35198570]
12. Li S. Immersive technologies in health professions education: a bibliometric analysis. Comput Educ X Real 2024;4:100051. [doi: 10.1016/j.cexr.2024.100051]
13. Rojas-Sánchez MA, Palos-Sánchez PR, Folgado-Fernández JA. Systematic literature review and bibliometric analysis on virtual reality and education. Educ Inf Technol (Dordr) 2023;28(1):155-192 [FREE Full text] [doi: 10.1007/s10639-022-11167-5] [Medline: 35789766]
14. Report: post-pandemic progress – lessons learnt in healthcare education. Council of Deans of Health. 2024 May 02. URL: https://tinyurl.com/mu6p73xm [accessed 2024-06-17]
15. Universities powering the NHS. Universities UK. URL: https://tinyurl.com/5y8x6we8 [accessed 2024-06-17]

XSL•FO
RenderX

16. Logeswaran A, Munsch C, Chong YJ, Ralph N, McCrossnan J. The role of extended reality technology in healthcare education: towards a learner-centred approach. Future Healthc J 2021 Mar;8(1):e79-e84 [FREE Full text] [doi: 10.7861/fhj.2020-0112] [Medline: 33791482]
17. Mathew RK, Immersive Healthcare Collaboration, Mushtaq F. Three principles for the progress of immersive technologies in healthcare training and education. BMJ Simul Technol Enhanc Learn 2021 May 25;7(5):459-460 [FREE Full text] [doi: 10.1136/bmjstel-2021-000881] [Medline: 35515725]
18. Anton N, Calhoun AC, Stefanidis D. Current research priorities in healthcare simulation: results of a Delphi survey. Simul Healthc 2022 Feb 01;17(1):e1-e7. [doi: 10.1097/SIH.0000000000000564] [Medline: 35104830]
19. Stefanidis D, Lee G, Blair PG, Johnson KA, Sachdeva AK. What are the top research priorities in surgical simulation and how can they be best addressed? Results from a multidisciplinary consensus conference. Ann Surg 2022 Dec 01;276(6):e1052-e1056. [doi: 10.1097/SLA.0000000000004651] [Medline: 33234796]
20. Anton NE, Gardner AK, Stefanidis D. Priorities in surgical simulation research: what do the experts say? Am J Surg 2020 Jul;220(1):95-99. [doi: 10.1016/j.amjsurg.2019.10.017] [Medline: 31672307]
21. Garfinkle R, Petersen RP, DuCoin C, Altieri MS, Aggarwal R, Pryor A, et al. Consensus priority research questions in gastrointestinal and endoscopic surgery in the year 2020: results of a SAGES Delphi study. Surg Endosc 2022 Sep;36(9):6688-6695. [doi: 10.1007/s00464-021-08941-w] [Medline: 35028734]
22. Cook DA, Andersen DK, Combes JR, Feldman DL, Sachdeva AK. The value proposition of simulation-based education. Surgery 2018 Apr;163(4):944-949. [doi: 10.1016/j.surg.2017.11.008] [Medline: 29452702]
23. Johnston MJ, Paige JT, Aggarwal R, Stefanidis D, Tsuda S, Khajuria A, et al. An overview of research priorities in surgical simulation: what the literature shows has been achieved during the 21st century and what remains. Am J Surg 2016 Jan;211(1):214-225. [doi: 10.1016/j.amjsurg.2015.06.014] [Medline: 26343855]
24. Hunt EA, Duval-Arnould J, Chime NO, Auerbach M, Kessler D, Duff JP, et al. Building consensus for the future of paediatric simulation: a novel 'KJ Reverse-Merlin' methodology. BMJ Simul Technol Enhanc Learn 2016 Apr 12;2(2):35-41 [FREE Full text] [doi: 10.1136/bmjstel-2015-000072] [Medline: 35518195]
25. McLaughlin C, Barry W, Barin E, Kysh L, Auerbach MA, Upperman JS, et al. Multidisciplinary simulation-based team training for trauma resuscitation: a scoping review. J Surg Educ 2019;76(6):1669-1680. [doi: 10.1016/j.jsurg.2019.05.002] [Medline: 31105006]
26. Tudor Car L, Poon S, Kyaw BM, Cook DA, Ward V, Atun R, et al. Digital education for health professionals: an evidence map, conceptual framework, and research agenda. J Med Internet Res 2022 Mar 17;24(3):e31977 [FREE Full text] [doi: 10.2196/31977] [Medline: 35297767]
27. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015 Sep;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
28. RefWorks login. ProQuest LLC. URL: https://refworks.proquest.com [accessed 2024-05-21]
29. Online survey. Jisc. URL: https://www.onlinesurveys.ac.uk/ [accessed 2024-08-13]
30. Schneider P, Evaniew N, Rendon JS, McKay P, Randall RL, Turcotte R, et al. Moving forward through consensus: protocol for a modified Delphi approach to determine the top research priorities in the field of orthopaedic oncology. BMJ Open 2016 May 24;6(5):e011780 [FREE Full text] [doi: 10.1136/bmjopen-2016-011780] [Medline: 27221129]
31. Slido homepage. Slido. URL: https://www.slido.com/ [accessed 2024-08-13]
32. Duarte ML, Santos LR, Iared W, Peccin MS. Comparison of ultrasonography learning between distance teaching and traditional methodology. An educational systematic review. Sao Paulo Med J 2022;140(6):806-817 [FREE Full text] [doi: 10.1590/1516-3180.2021.1047.R.19052022] [Medline: 36043680]
33. Ryan GV, Callaghan S, Rafferty A, Higgins MF, Mangina E, McAuliffe F. Learning outcomes of immersive technologies in health care student education: systematic review of the literature. J Med Internet Res 2022 Feb 01;24(2):e30082 [FREE Full text] [doi: 10.2196/30082] [Medline: 35103607]
34. Zhao G, Fan M, Yuan Y, Zhao F, Huang H. The comparison of teaching efficiency between virtual reality and traditional education in medical education: a systematic review and meta-analysis. Ann Transl Med 2021 Feb;9(3):252 [FREE Full text] [doi: 10.21037/atm-20-2785] [Medline: 33708879]
35. Suresh D, Aydin A, James S, Ahmed K, Dasgupta P. The role of augmented reality in surgical training: a systematic review. Surg Innov 2023 Jun 22;30(3):366 [FREE Full text] [doi: 10.1177/15533506221140506] [Medline: 36412148]
36. Shorey S, Ng ED. The use of virtual reality simulation among nursing students and registered nurses: a systematic review. Nurse Educ Today 2021 Mar;98:104662. [doi: 10.1016/j.nedt.2020.104662] [Medline: 33203545]
37. Chahal B, Aydın A, Amin MS, Ong K, Khan A, Khan MS, et al. Transfer of open and laparoscopic skills to robotic surgery: a systematic review. J Robot Surg 2023 Aug;17(4):1207-1225 [FREE Full text] [doi: 10.1007/s11701-022-01492-9] [Medline: 36418717]
38. Dzyuba N, Jandu J, Yates J, Kushnerev E. Virtual and augmented reality in dental education: the good, the bad and the better. Eur J Dent Educ 2022 Nov 06 (forthcoming). [doi: 10.1111/eje.12871] [Medline: 36336847]

XSL•FO
RenderX

39. Sim JJ, Rusli KD, Seah B, Levett-Jones T, Lau Y, Liaw SY. Virtual simulation to enhance clinical reasoning in nursing: a systematic review and meta-analysis. Clin Simul Nurs 2022 Aug;69:26-39. [doi: 10.1016/j.ecns.2022.05.006] [Medline: 35754937]
40. Tolarba JE. Virtual simulation in nursing education: a systematic review. Int J Nurs Educ 2021 Jun 14;13(3):48-54. [doi: 10.37506/ijone.v13i3.16310]
41. Ashokka B, Dong C, Law LS, Liaw SY, Chen FG, Samarasekera DD. A BEME systematic review of teaching interventions to equip medical students and residents in early recognition and prompt escalation of acute clinical deteriorations: BEME Guide No. 62. Med Teach 2020 Jul;42(7):724-737. [doi: 10.1080/0142159X.2020.1763286] [Medline: 32493155]
42. Patel EA, Aydin A, Cearns M, Dasgupta P, Ahmed K. A systematic review of simulation-based training in neurosurgery, part 2: spinal and pediatric surgery, neurointerventional radiology, and nontechnical skills. World Neurosurg 2020 Jan;133:e874-e892. [doi: 10.1016/j.wneu.2019.08.263] [Medline: 31541754]
43. Lei YY, Zhu L, Sa YT, Cui XS. Effects of high-fidelity simulation teaching on nursing students' knowledge, professional skills and clinical ability: a meta-analysis and systematic review. Nurse Educ Pract 2022 Mar;60:103306. [doi: 10.1016/j.nepr.2022.103306] [Medline: 35202957]
44. Stone R, Cooke M, Mitchell M. Undergraduate nursing students' use of video technology in developing confidence in clinical skills for practice: a systematic integrative literature review. Nurse Educ Today 2020 Jan;84:104230. [doi: 10.1016/j.nedt.2019.104230] [Medline: 31689584]
45. Lohre R, Wang JC, Lewandrowski KU, Goel DP. Virtual reality in spinal endoscopy: a paradigm shift in education to support spine surgeons. J Spine Surg 2020 Jan;6(Suppl 1):S208-S223 [FREE Full text] [doi: 10.21037/jss.2019.11.16] [Medline: 32195429]
46. Tori AA, Tori R, Nunes F. Serious game design in health education: a systematic review. IEEE Trans Learning Technol 2022 Dec 1;15(6):827-846. [doi: 10.1109/TLT.2022.3200583]
47. Plackett R, Kassianos AP, Mylan S, Kambouri M, Raine R, Sheringham J. The effectiveness of using virtual patient educational tools to improve medical students' clinical reasoning skills: a systematic review. BMC Med Educ 2022 May 13;22(1):365 [FREE Full text] [doi: 10.1186/s12909-022-03410-x] [Medline: 35550085]
48. Beshir SA, Mohamed AP, Soorya A, Sir Loon Goh S, Moussa El-Labadd E, Hussain N, et al. Virtual patient simulation in pharmacy education: a systematic review. Pharm Educ 2022 Dec 17;22(1):954-970. [doi: 10.46542/PE.2022.221.954970]
49. Fouasson-Chailloux A, Daley P, Menu P, Gross R, Dauty M. Social media in health studies: a systematic review of comparative learning methods. Int J Environ Res Public Health 2022 Feb 15;19(4):2205 [FREE Full text] [doi: 10.3390/ijerph19042205] [Medline: 35206401]
50. Almutairi M, Simpson A, Khan E, Dickinson T. The value of social media use in improving nursing students' engagement: a systematic review. Nurse Educ Pract 2022 Oct;64:103455 [FREE Full text] [doi: 10.1016/j.nepr.2022.103455] [Medline: 36182729]
51. Du L, Zhao L, Xu T, Wang Y, Zu W, Huang X, et al. Blended learning vs traditional teaching: the potential of a novel teaching strategy in nursing education - a systematic review and meta-analysis. Nurse Educ Pract 2022 Aug;63:103354. [doi: 10.1016/j.nepr.2022.103354] [Medline: 35580368]
52. Walshe N, Condon C, Gonzales RA, Burke E, Chianáin LN, Thamanam N, et al. Cultural simulations, authenticity, focus, and outcomes: a systematic review of the healthcare literature. Clin Simul Nurs 2022 Oct;71:65-81. [doi: 10.1016/j.ecns.2022.05.008]
53. Yogeswaran V, El Morr C. Effectiveness of online mindfulness interventions on medical students' mental health: a systematic review. BMC Public Health 2021 Dec 18;21(1):2293 [FREE Full text] [doi: 10.1186/s12889-021-12341-z] [Medline: 34920715]
54. Chen B, Wang Y, Xiao L, Xu C, Shen Y, Qin Q, et al. Effects of mobile learning for nursing students in clinical education: a meta-analysis. Nurse Educ Today 2021 Feb;97:104706. [doi: 10.1016/j.nedt.2020.104706] [Medline: 33360012]
55. Baetzner AS, Wespi R, Hill Y, Gyllencreutz L, Sauter TC, Saveman BI, et al. Preparing medical first responders for crises: a systematic literature review of disaster training programs and their effectiveness. Scand J Trauma Resusc Emerg Med 2022 Dec 24;30(1):76 [FREE Full text] [doi: 10.1186/s13049-022-01056-8] [Medline: 36566227]
56. Chiang FK, Shang X, Qiao L. Augmented reality in vocational training: a systematic review of research and applications. Comput Hum Behav 2022 Apr;129:107125. [doi: 10.1016/j.chb.2021.107125]
57. Lam K, Chen J, Wang Z, Iqbal FM, Darzi A, Lo B, et al. Machine learning for technical skill assessment in surgery: a systematic review. NPJ Digit Med 2022 Mar 03;5(1):24 [FREE Full text] [doi: 10.1038/s41746-022-00566-0] [Medline: 35241760]
58. Noyes JA, Welch PM, Johnson JW, Carbonneau KJ. A systematic review of digital badges in health care education. Med Educ 2020 Jul;54(7):600-615. [doi: 10.1111/medu.14060] [Medline: 31971267]
59. Thangavelu DP, Tan AJ, Cant R, Chua WL, Liaw SY. Digital serious games in developing nursing clinical competence: a systematic review and meta-analysis. Nurse Educ Today 2022 Jun;113:105357. [doi: 10.1016/j.nedt.2022.105357] [Medline: 35429749]
60. Min A, Min H, Kim S. Effectiveness of serious games in nurse education: a systematic review. Nurse Educ Today 2022 Jan;108:105178. [doi: 10.1016/j.nedt.2021.105178] [Medline: 34717098]

XSL•FO
RenderX

61. Chen IH, Ghazi A, Sridhar A, Stoyanov D, Slack M, Kelly JD, et al. Evolving robotic surgery training and improving patient safety, with the integration of novel technologies. World J Urol 2021 Aug;39(8):2883-2893 [FREE Full text] [doi: 10.1007/s00345-020-03467-7] [Medline: 33156361]
62. van Gaalen AE, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma AD, Georgiadis JR. Gamification of health professions education: a systematic review. Adv Health Sci Educ Theory Pract 2021 May;26(2):683-711 [FREE Full text] [doi: 10.1007/s10459-020-10000-3] [Medline: 33128662]
63. Xu Y, Lau Y, Cheng LJ, Lau ST. Learning experiences of game-based educational intervention in nursing students: a systematic mixed-studies review. Nurse Educ Today 2021 Dec;107:105139. [doi: 10.1016/j.nedt.2021.105139] [Medline: 34563963]
64. Lucena-Anton D, Fernandez-Lopez JC, Pacheco-Serrano AI, Garcia-Munoz C, Moral-Munoz JA. Virtual and augmented reality versus traditional methods for teaching physiotherapy: a systematic review. Eur J Investig Health Psychol Educ 2022 Dec 02;12(12):1780-1792 [FREE Full text] [doi: 10.3390/ejihpe12120125] [Medline: 36547026]
65. Kim EJ, Lim JY, Kim GM. A systematic review and meta-analysis of studies on extended reality-based pediatric nursing simulation program development. Child Health Nurs Res 2023 Jan 31;29(1):24-36 [FREE Full text] [doi: 10.4094/chnr.2023.29.1.24] [Medline: 36760110]
66. Pietersen P, Bjerrum F, Tolsgaard MG, Konge LS, Andersen SA. Standard setting in simulation-based training of surgical procedures: a systematic review. Ann Surg 2022 May 01;275(5):872-882. [doi: 10.1097/SLA.0000000000005209] [Medline: 34520423]
67. Moro C, Birt J, Stromberga Z, Phelps C, Clark J, Glasziou P, et al. Virtual and augmented reality enhancements to medical and science student physiology and anatomy test performance: a systematic review and meta-analysis. Anat Sci Educ 2021 May;14(3):368-376. [doi: 10.1002/ase.2049] [Medline: 33378557]
68. Bray L, Krogh TB, Østergaard D. Simulation-based training for continuing professional development within a primary care context: a systematic review. Educ Prim Care 2023 Mar 02;34(2):64-73. [doi: 10.1080/14739879.2022.2161424] [Medline: 36730551]
69. Nowell L, Dhingra S, Carless-Kane S, McGuinness C, Paolucci A, Jacobsen M, et al. A systematic review of online education initiatives to develop students remote caring skills and practices. Med Educ Online 2022 Dec 12;27(1):2088049 [FREE Full text] [doi: 10.1080/10872981.2022.2088049] [Medline: 35694798]
70. Ødegaard NB, Myrhaug HT, Dahl-Michelsen T, Røe Y. Digital learning designs in physiotherapy education: a systematic review and meta-analysis. BMC Med Educ 2021 Jan 13;21(1):48 [FREE Full text] [doi: 10.1186/s12909-020-02483-w] [Medline: 33441140]
71. Shorey S, Pereira TL, Teo WZ, Ang E, Lau TC, Samarasekera DD. Navigating nursing curriculum change during COVID-19 pandemic: a systematic review and meta-synthesis. Nurse Educ Pract 2022 Nov;65:103483 [FREE Full text] [doi: 10.1016/j.nepr.2022.103483] [Medline: 36327596]
72. Srinivasa K, Chen Y, Henning MA. The role of online videos in teaching procedural skills to post-graduate medical learners: a systematic narrative review. Med Teach 2020 Jun;42(6):689-697. [doi: 10.1080/0142159X.2020.1733507] [Medline: 32174211]
73. Anugrahsari S, Chaeruman UA, Abbas H, Suryadi S. Patient safety education for clinical students: a systematic literature review. Open Access Maced J Med Sci 2022 Mar 11;10(F):208-214. [doi: 10.3889/oamjms.2022.8345]
74. Grafton-Clarke C, Uraiby H, Gordon M, Clarke N, Rees E, Park S, et al. Pivot to online learning for adapting or continuing workplace-based clinical learning in medical education following the COVID-19 pandemic: a BEME systematic review: BEME Guide No. 70. Med Teach 2022 Mar;44(3):227-243. [doi: 10.1080/0142159X.2021.1992372] [Medline: 34689692]
75. Savage AJ, McNamara PW, Moncrieff TW, O'Reilly GM. Review article: e-learning in emergency medicine: a systematic review. Emerg Med Australas 2022 Jun 27;34(3):322-332 [FREE Full text] [doi: 10.1111/1742-6723.13936] [Medline: 35224870]
76. Aydın A, Baig U, Al-Jabir A, Sarıca K, Dasgupta P, Ahmed K. Simulation-based training models for urolithiasis: a systematic review. J Endourol 2021 Jul 01;35(7):1098-1117. [doi: 10.1089/end.2020.0408] [Medline: 33198492]
77. Guckian J, Utukuri M, Asif A, Burton O, Adeyoju J, Oumeziane A, et al. Social media in undergraduate medical education: a systematic review. Med Educ 2021 Nov 04;55(11):1227-1241. [doi: 10.1111/medu.14567] [Medline: 33988867]
78. Tursø-Finnich T, Jensen RO, Jensen LX, Konge L, Thinggaard E. Virtual reality head-mounted displays in medical education: a systematic review. Simul Healthc 2023 Feb 01;18(1):42-50. [doi: 10.1097/SIH.0000000000000636] [Medline: 35136005]
79. Clerkin R, Patton D, Moore Z, Nugent L, Avsar P, O'Connor T. What is the impact of video as a teaching method on achieving psychomotor skills in nursing? A systematic review and meta-analysis. Nurse Educ Today 2022 Apr;111:105280 [FREE Full text] [doi: 10.1016/j.nedt.2022.105280] [Medline: 35139443]
80. Foronda CL, Fernandez-Burgos M, Nadeau C, Kelley CN, Henry MN. Virtual simulation in nursing education: a systematic review spanning 1996 to 2018. Simul Healthc 2020;15(1):46-54. [doi: 10.1097/sih.0000000000000411]
81. Piot MA, Dechartres A, Attoe C, Romeo M, Jollant F, Billon G, et al. Effectiveness of simulation in psychiatry for nursing students, nurses and nurse practitioners: a systematic review and meta-analysis. J Adv Nurs 2022 Feb 11;78(2):332-347. [doi: 10.1111/jan.14986] [Medline: 34378236]

XSL•FO
RenderX

82. Lapierre A, Bouferguene S, Gauvin-Lepage J, Lavoie P, Arbour C. Effectiveness of interprofessional manikin-based simulation training on teamwork among real teams during trauma resuscitation in adult emergency departments: a systematic review. Simul Healthc 2020 Dec;15(6):409-421. [doi: 10.1097/SIH.0000000000000443] [Medline: 32218090]
83. Li YY, Au ML, Tong LK, Ng WI, Wang SC. High-fidelity simulation in undergraduate nursing education: a meta-analysis. Nurse Educ Today 2022 Apr;111:105291 [FREE Full text] [doi: 10.1016/j.nedt.2022.105291] [Medline: 35158134]
84. Alshehri FD, Jones S, Harrison D. The effectiveness of high-fidelity simulation on undergraduate nursing students' clinical reasoning-related skills: a systematic review. Nurse Educ Today 2023 Feb;121:105679. [doi: 10.1016/j.nedt.2022.105679] [Medline: 36542870]
85. Hildreth AF, Maggio LA, Iteen A, Wojahn AL, Cook DA, Battista A. Technology-enhanced simulation in emergency medicine: updated systematic review and meta-analysis 1991-2021. AEM Educ Train 2023 Apr 14;7(2):e10848 [FREE Full text] [doi: 10.1002/aet2.10848] [Medline: 36936085]
86. Kim JY, Kim ME. Can online learning be a reliable alternative to nursing students' learning during a pandemic? - A systematic review and meta-analysis. Nurse Educ Today 2023 Mar;122:105710. [doi: 10.1016/j.nedt.2023.105710] [Medline: 36739810]
87. Muirhead K, Macaden L, Smyth K, Chandler C, Clarke C, Polson R, et al. Establishing the effectiveness of technology-enabled dementia education for health and social care practitioners: a systematic review. Syst Rev 2021 Sep 21;10(1):252 [FREE Full text] [doi: 10.1186/s13643-021-01781-8] [Medline: 34548101]
88. Muirhead K, Macaden L, Smyth K, Chandler C, Clarke C, Polson R, et al. The characteristics of effective technology-enabled dementia education: a systematic review and mixed research synthesis. Syst Rev 2022 Feb 23;11(1):34 [FREE Full text] [doi: 10.1186/s13643-021-01866-4] [Medline: 35197109]
89. Abualadas HM, Xu L. Achievement of learning outcomes in non-traditional (online) versus traditional (face-to-face) anatomy teaching in medical schools: a mixed method systematic review. Clin Anat 2023 Jan 25;36(1):50-76 [FREE Full text] [doi: 10.1002/ca.23942] [Medline: 35969356]
90. Arqub SA, Waleed M, Al-Abedalla K, Iverson MG, Uribe F. Insight on the influence of technology-enhanced learning in orthodontics' education: a systematic review. Eur J Dent Educ 2023 Aug 31;27(3):729-745. [doi: 10.1111/eje.12861] [Medline: 36250284]
91. Martinengo L, Yeo NJ, Markandran KD, Olsson M, Kyaw BM, Car LT. Digital health professions education on chronic wound management: a systematic review. Int J Nurs Stud 2020 Apr;104:103512. [doi: 10.1016/j.ijnurstu.2019.103512] [Medline: 32086027]
92. Wirth T, Peters C, Nienhaus A, Schablon A. Interventions for workplace violence prevention in emergency departments: a systematic review. Int J Environ Res Public Health 2021 Aug 10;18(16):8459 [FREE Full text] [doi: 10.3390/ijerph18168459] [Medline: 34444208]
93. Lindhard MS, Thim S, Laursen HS, Schram AW, Paltved CT, Henriksen TB. Simulation-based neonatal resuscitation team training: a systematic review. Pediatrics 2021 Apr;147(4):e2020042010. [doi: 10.1542/peds.2020-042010] [Medline: 33762309]
94. Lee IR, Kim HW, Lee Y, Koyanagi A, Jacob L, An S, et al. Changes in undergraduate medical education due to COVID-19: a systematic review. Eur Rev Med Pharmacol Sci 2021 Jun;25(12):4426-4434 [FREE Full text] [doi: 10.26355/eurrev_202106_26155] [Medline: 34227080]
95. Patano A, Cirulli N, Beretta M, Plantamura P, Inchingolo AD, Inchingolo AM, et al. Education technology in orthodontics and paediatric dentistry during the COVID-19 pandemic: a systematic review. Int J Environ Res Public Health 2021 Jun 04;18(11):6056 [FREE Full text] [doi: 10.3390/ijerph18116056] [Medline: 34199882]
96. Dedeilia A, Sotiropoulos MG, Hanrahan JG, Janga D, Dedeilias P, Sideris M. Medical and surgical education challenges and innovations in the COVID-19 era: a systematic review. In Vivo 2020 Jun 05;34(3 Suppl):1603-1611 [FREE Full text] [doi: 10.21873/invivo.11950] [Medline: 32503818]
97. Brown N, Margus C, Hart A, Sarin R, Hertelendy A, Ciottone G. Virtual reality training in disaster medicine: a systematic review of the literature. Simul Healthc 2023 Aug 01;18(4):255-261. [doi: 10.1097/SIH.0000000000000675] [Medline: 35696131]
98. Castillo-Segura P, Fernández-Panadero C, Alario-Hoyos C, Muñoz-Merino PJ, Delgado Kloos C. Objective and automated assessment of surgical technical skills with IoT systems: a systematic literature review. Artif Intell Med 2021 Feb;112:102007. [doi: 10.1016/j.artmed.2020.102007] [Medline: 33581827]
99. Davids J, Manivannan S, Darzi A, Giannarou S, Ashrafian H, Marcus HJ. Simulation for skills training in neurosurgery: a systematic review, meta-analysis, and analysis of progressive scholarly acceptance. Neurosurg Rev 2021 Aug 18;44(4):1853-1867 [FREE Full text] [doi: 10.1007/s10143-020-01378-0] [Medline: 32944808]
100. Regmi K, Jones L. A systematic review of the factors - enablers and barriers - affecting e-learning in health sciences education. BMC Med Educ 2020 Mar 30;20(1):91 [FREE Full text] [doi: 10.1186/s12909-020-02007-6] [Medline: 32228560]
101. Lockey A, Bland A, Stephenson J, Bray J, Astin F. Blended learning in health care education: an overview and overarching meta-analysis of systematic reviews. J Contin Educ Health Prof 2022 Oct 01;42(4):256-264. [doi: 10.1097/CEH.0000000000000455] [Medline: 36070399]

XSL•FO
RenderX

102. Banks L, Kay R. Exploring flipped classrooms in undergraduate nursing and health science: a systematic review. Nurse Educ Pract 2022 Oct;64:103417. [doi: 10.1016/j.nepr.2022.103417] [Medline: 36113354]
103. Tabatabaeichehr M, Babaei S, Dartomi M, Alesheikh P, Tabatabaee A, Mortazavi H, et al. Medical students' satisfaction level with e-learning during the COVID-19 pandemic and its related factors: a systematic review. J Educ Eval Health Prof 2022 Dec 20;19:37 [FREE Full text] [doi: 10.3352/jeehp.2022.19.37] [Medline: 36579447]
104. Oh SY, Cook DA, van Gerven PW, Nicholson J, Fairbrother H, Smeenk FW, et al. Physician training for electrocardiogram interpretation: a systematic review and meta-analysis. Acad Med 2022 Apr 01;97(4):593-602. [doi: 10.1097/ACM.0000000000004607] [Medline: 35086115]
105. Haiser A, Aydin A, Kunduzi B, Ahmed K, Dasgupta P. A systematic review of simulation-based training in vascular surgery. J Surg Res 2022 Nov;279:409-419 [FREE Full text] [doi: 10.1016/j.jss.2022.05.009] [Medline: 35839575]
106. Chae D, Yoo JY, Kim J, Ryu J. Effectiveness of virtual simulation to enhance cultural competence in pre-licensure and licensed health professionals: a systematic review. Clin Simul Nurs 2021 Jul;56:137-154. [doi: 10.1016/j.ecns.2021.04.013]
107. Mao RQ, Lan L, Kay J, Lohre R, Ayeni OR, Goel DP, et al. Immersive virtual reality for surgical training: a systematic review. J Surg Res 2021 Dec;268:40-58. [doi: 10.1016/j.jss.2021.06.045] [Medline: 34284320]
108. Gomaa AR, Grafton-Clarke C, Saratzis A, Davies RS. The role of high-fidelity simulation in the acquisition of endovascular surgical skills: a systematic review. Ann Vasc Surg 2023 Jul;93:405-427 [FREE Full text] [doi: 10.1016/j.avsg.2023.02.025] [Medline: 36906131]
109. Howlader D, Daga D, Mehrotra D. The scope of computerized simulation in competency-based maxillofacial training: a systematic review. Int J Oral Maxillofac Surg 2022 Aug;51(8):1101-1110. [doi: 10.1016/j.ijom.2021.10.002] [Medline: 35058106]
110. Fahl JT, Duvivier R, Reinke L, Pierie JP, Schönrock-Adema J. Towards best practice in developing motor skills: a systematic review on spacing in VR simulator-based psychomotor training for surgical novices. BMC Med Educ 2023 Mar 13;23(1):154 [FREE Full text] [doi: 10.1186/s12909-023-04046-1] [Medline: 36907871]
111. Schmidt MW, Köppinger KF, Fan C, Kowalewski KF, Schmidt LP, Vey J, et al. Virtual reality simulation in robot-assisted surgery: meta-analysis of skill transfer and predictability of skill. BJS Open 2021 Mar 05;5(2):zraa066 [FREE Full text] [doi: 10.1093/bjsopen/zraa066] [Medline: 33864069]
112. Ortega MA, Marchese VG, Zarro MJ, Film RJ, Shipper AG, Felter C. Digital and blended curriculum delivery in health professions education: an umbrella review with implications for Doctor of Physical Therapy education programs. Phys Ther Rev 2021 Nov 09;27(1):4-24. [doi: 10.1080/10833196.2021.2000286]
113. Aweid B, Haider Z, Wehbe M, Hunter A. Educational benefits of the online journal club: a systematic review. Med Teach 2022 Jan;44(1):57-62. [doi: 10.1080/0142159X.2021.1963424] [Medline: 34403291]
114. Özbay Ö, Çınar S. Effectiveness of flipped classroom teaching models in nursing education: a systematic review. Nurse Educ Today 2021 Jul;102:104922. [doi: 10.1016/j.nedt.2021.104922] [Medline: 33940481]
115. Mao BP, Teichroeb ML, Lee T, Wong G, Pang T, Pleass H. Is online video-based education an effective method to teach basic surgical skills to students and surgical trainees? A systematic review and meta-analysis. J Surg Educ 2022;79(6):1536-1545 [FREE Full text] [doi: 10.1016/j.jsurg.2022.07.016] [Medline: 35933308]
116. Yucel C, Hawley G, Terzioglu F, Bogossian F. The effectiveness of simulation-based team training in obstetrics emergencies for improving technical skills: a systematic review. Simul Healthc 2020 Apr;15(2):98-105. [doi: 10.1097/SIH.0000000000000416] [Medline: 32168287]
117. Jujo S, Nakahira A, Kataoka Y, Banno M, Tsujimoto Y, Tsujimoto H, et al. Transesophageal echocardiography simulator training: a systematic review and meta-analysis of randomized controlled trials. Simul Healthc 2021 Oct 01;16(5):341-352. [doi: 10.1097/SIH.0000000000000537] [Medline: 33428355]
118. Gharib AM, Bindoff IK, Peterson GM, Salahudeen MS. Computer-based simulators in pharmacy practice education: a systematic narrative review. Pharmacy (Basel) 2023 Jan 02;11(1):8 [FREE Full text] [doi: 10.3390/pharmacy11010008] [Medline: 36649018]
119. Lo CK, Hew KF. Design principles for fully online flipped learning in health professions education: a systematic review of research during the COVID-19 pandemic. BMC Med Educ 2022 Oct 13;22(1):720 [FREE Full text] [doi: 10.1186/s12909-022-03782-0] [Medline: 36229820]
120. Polce EM, Kunze KN, Williams BT, Krivicich LM, Maheshwer B, Beletsky A, et al. Efficacy and validity of orthopaedic simulators in surgical training: a systematic review and meta-analysis of randomized controlled trials. J Am Acad Orthop Surg 2020 Dec 15;28(24):1027-1040. [doi: 10.5435/JAAOS-D-19-00839] [Medline: 32398408]
121. Neubauer R, Recker F, Bauer CJ, Brossart P, Schäfer VS. The current situation of musculoskeletal ultrasound education: a systematic literature review. Ultrasound Med Biol 2023 Jun;49(6):1363-1374. [doi: 10.1016/j.ultrasmedbio.2023.02.008] [Medline: 36941182]
122. Santos GN, da Silva HE, Leite AF, Mesquita CR, Figueiredo PT, Stefani CM, et al. The scope of dental education during COVID-19 pandemic: a systematic review. J Dent Educ 2021 Jul 23;85(7):1287-1300 [FREE Full text] [doi: 10.1002/jdd.12587] [Medline: 33759214]

XSL•FO
RenderX

123. Rothschild P, Richardson A, Beltz J, Chakrabarti R. Effect of virtual reality simulation training on real-life cataract surgery complications: systematic literature review. J Cataract Refract Surg 2021 Mar 01;47(3):400-406. [doi: 10.1097/j.jcrs.0000000000000323] [Medline: 32675654]
124. Capitani P, Zampogna B, Monaco E, Frizziero A, Moretti L, Losco M, et al. The role of virtual reality in knee arthroscopic simulation: a systematic review. Musculoskelet Surg 2023 Mar 15;107(1):85-95. [doi: 10.1007/s12306-021-00732-9] [Medline: 34655024]
125. Dickinson KJ, Bass BL. A systematic review of educational mobile-applications (apps) for surgery residents: simulation and beyond. J Surg Educ 2020 Sep;77(5):1244-1256. [doi: 10.1016/j.jsurg.2020.03.022] [Medline: 32307244]
126. Lin IC, Lee A, Mauch JT. Does e-learning improve plastic surgery education?: a systematic review of asynchronous resources. Ann Plast Surg 2021 Jul 01;87(1s Suppl 1):S40-S51. [doi: 10.1097/SAP.0000000000002806] [Medline: 34180865]
127. Naciri A, Radid M, Kharbach A, Chemsi G. E-learning in health professions education during the COVID-19 pandemic: a systematic review. J Educ Eval Health Prof 2021 Oct 29;18:27 [FREE Full text] [doi: 10.3352/jeehp.2021.18.27] [Medline: 34710319]
128. O'Brien B, Bevan K, Brockington C, Murphy J, Gilbert R. Effects of simulation-based cardiopulmonary and respiratory case training experiences on interprofessional teamwork: a systematic review. Can J Respir Ther 2023 Mar 28;59:85-94 [FREE Full text] [doi: 10.29390/cjrt-2022-060] [Medline: 36999003]
129. Plotzky C, Lindwedel U, Sorber M, Loessl B, König P, Kunze C, et al. Virtual reality simulations in nurse education: a systematic mapping review. Nurse Educ Today 2021 Jun;101:104868. [doi: 10.1016/j.nedt.2021.104868] [Medline: 33798987]
130. Gelmini AY, Duarte ML, Silva MO, Guimarães Junior JB, Santos LR. Augmented reality in interventional radiology education: a systematic review of randomized controlled trials. Sao Paulo Med J 2022;140(4):604-614 [FREE Full text] [doi: 10.1590/1516-3180.2021.0606.R2.27122021] [Medline: 35946678]
131. Aditya I, Kwong JC, Canil T, Lee JY, Goldenberg MG. Current educational interventions for improving technical skills of urology trainees in endourological procedures: a systematic review. J Endourol 2020 Jul 01;34(7):723-731. [doi: 10.1089/end.2019.0693] [Medline: 31691593]
132. Orejuela FJ, Aschkenazi SO, Howard DL, Jeppson PC, Balgobin S, Walter AJ, et al. Gynecologic surgical skill acquisition through simulation with outcomes at the time of surgery: a systematic review and meta-analysis. Am J Obstet Gynecol 2022 Jul;227(1):29.e1-2.e24. [doi: 10.1016/j.ajog.2022.01.031] [Medline: 35120886]
133. Clarke E. Virtual reality simulation-the future of orthopaedic training? A systematic review and narrative analysis. Adv Simul (Lond) 2021 Jan 13;6(1):2 [FREE Full text] [doi: 10.1186/s41077-020-00153-x] [Medline: 33441190]
134. Zaed I, Chibbaro S, Ganau M, Tinterri B, Bossi B, Peschillo S, et al. Simulation and virtual reality in intracranial aneurysms neurosurgical training: a systematic review. J Neurosurg Sci 2022 Dec;66(6):494-500. [doi: 10.23736/S0390-5616.22.05526-6] [Medline: 35301837]
135. Cant R, Ryan CM, Kelly MA. Use and effectiveness of virtual simulations in nursing student education: an umbrella review. Comput Inform Nurs 2023 Jan 01;41(1):31-38. [doi: 10.1097/CIN.0000000000000932] [Medline: 35926237]
136. Garnier A, Vanherp R, Bonnabry P, Bouchoud L. Use of simulation for education in hospital pharmaceutical technologies: a systematic review. Eur J Hosp Pharm 2023 Mar 23;30(2):70-76 [FREE Full text] [doi: 10.1136/ejhpharm-2021-003034] [Medline: 34949651]
137. Gelmini AY, Duarte ML, de Assis AM, Guimarães Junior JB, Carnevale FC. Virtual reality in interventional radiology education: a systematic review. Radiol Bras 2021 Aug;54(4):254-260 [FREE Full text] [doi: 10.1590/0100-3984.2020.0162] [Medline: 34393293]
138. Fontaine G, Zagury-Orly I, Maheu-Cadotte MA, Lapierre A, Thibodeau-Jarry N, Denus SD, et al. A meta-analysis of the effect of paper versus digital reading on reading comprehension in health professional education. Am J Pharm Educ 2021 Nov;85(10):8525 [FREE Full text] [doi: 10.5688/ajpe8525] [Medline: 34301544]
139. Margulies IG, Xu H, Henderson PW. Microsurgery training in the digital era: a systematic review of accessible digital resources. Ann Plast Surg 2020 Oct 9;85(4):337-343. [doi: 10.1097/SAP.0000000000002214] [Medline: 31923014]
140. Ozdemir NG, Kaya H. The effect of high-fidelity simulation on experiences of urinary catheterization in nursing students: a mixed-method systematic review. Int J Caring Sci 2022;15(2):1218-1232 [FREE Full text]
141. Berg MN, Ngune I, Schofield P, Grech L, Juraskova I, Strasser M, et al. Effectiveness of online communication skills training for cancer and palliative care health professionals: a systematic review. Psychooncology 2021 Sep 28;30(9):1405-1419. [doi: 10.1002/pon.5702] [Medline: 33909328]
142. Oliveira Silva G, Oliveira FS, Coelho AS, Fonseca LM, Vieira FV, Campbell SH, et al. Influence of simulation design on stress, anxiety and self-confidence of nursing students: systematic review with meta-analysis. J Clin Nurs 2023 Sep 09;32(17-18):5668-5692. [doi: 10.1111/jocn.16681] [Medline: 36894868]
143. Richardson CL, White S, Chapman S. Virtual patient technology to educate pharmacists and pharmacy students on patient communication: a systematic review. BMJ Simul Technol Enhanc Learn 2020 Dec 11;6(6):332-338 [FREE Full text] [doi: 10.1136/bmjstel-2019-000514] [Medline: 35515492]

XSL•FO
RenderX

144. Hao X, Peng X, Ding X, Qin Y, Lv M, Li J, et al. Application of digital education in undergraduate nursing and medical interns during the COVID-19 pandemic: a systematic review. Nurse Educ Today 2022 Jan;108:105183 [FREE Full text] [doi: 10.1016/j.nedt.2021.105183] [Medline: 34741918]
145. Ong CW, Tan MC, Lam M, Koh VT. Applications of extended reality in ophthalmology: systematic review. J Med Internet Res 2021 Aug 19;23(8):e24152 [FREE Full text] [doi: 10.2196/24152] [Medline: 34420929]
146. Tonapa SI, Mulyadi M, Ho KF, Efendi F. Effectiveness of using high-fidelity simulation on learning outcomes in undergraduate nursing education: systematic review and meta-analysis. Eur Rev Med Pharmacol Sci 2023 Jan;27(2):444-458 [FREE Full text] [doi: 10.26355/eurrev_202301_31040] [Medline: 36734697]
147. Woodhead NJ, Mahmud A, Clark J. Effects of simulation for gynaecological ultrasound scan training: a systematic review. BMJ Simul Technol Enhanc Learn 2020 Oct 30;6(6):320-331 [FREE Full text] [doi: 10.1136/bmjstel-2019-000482] [Medline: 35515489]
148. Zhao J, Xu X, Jiang H, Ding Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. BMC Med Educ 2020 Apr 25;20(1):127 [FREE Full text] [doi: 10.1186/s12909-020-1994-z] [Medline: 32334594]
149. Smith SC, Siau K, Cannatelli R, Antonelli G, Shivaji UN, Ghosh S, et al. Training methods in optical diagnosis and characterization of colorectal polyps: a systematic review and meta-analysis. Endosc Int Open 2021 May 22;9(5):E716-E726 [FREE Full text] [doi: 10.1055/a-1381-7181] [Medline: 33937513]
150. de Gagne JC, Koppel PD, Wang EJ, Rushton S, Ledbetter L, Yamane SS, et al. A systematic review of videoconferencing in health professions education: the digital divide revisited in the COVID-19 era. Int J Nurs Educ Scholarsh 2023 Jan 01;20(1) [FREE Full text] [doi: 10.1515/ijnes-2022-0068] [Medline: 36608319]
151. Zhang X, Al-Mekhled D, Choate J. Are virtual physiology laboratories effective for student learning? A systematic review. Adv Physiol Educ 2021 Sep 01;45(3):467-480 [FREE Full text] [doi: 10.1152/advan.00016.2021] [Medline: 34142876]
152. O'Connor S, Wang Y, Cooke S, Ali A, Kennedy S, Lee JJ, et al. Designing and delivering digital learning (e-Learning) interventions in nursing and midwifery education: a systematic review of theories. Nurse Educ Pract 2023 May;69:103635 [FREE Full text] [doi: 10.1016/j.nepr.2023.103635] [Medline: 37060735]
153. Lee J, Kim H, Kim KH, Jung D, Jowsey T, Webster CS. Effective virtual patient simulators for medical communication training: a systematic review. Med Educ 2020 Sep 22;54(9):786-795. [doi: 10.1111/medu.14152] [Medline: 32162355]
154. Qiao J, Huang CR, Liu Q, Li S, Xu J, Li L, et al. Effectiveness of non-immersive virtual reality simulation in learning knowledge and skills for nursing students: meta-analysis. Clin Simul Nurs 2023 Mar;76:26-38. [doi: 10.1016/j.ecns.2022.12.003]
155. Lowe S, Mares K, Khadjesari Z. Immersive technology in ophthalmology education: a systematic review. BMJ Simul Technol Enhanc Learn 2021 Jun 29;7(6):600-604 [FREE Full text] [doi: 10.1136/bmjstel-2021-000906] [Medline: 35520978]
156. Stojan J, Haas M, Thammasitboon S, Lander L, Evans S, Pawlik C, et al. Online learning developments in undergraduate medical education in response to the COVID-19 pandemic: a BEME systematic review: BEME Guide No. 69. Med Teach 2022 Feb;44(2):109-129. [doi: 10.1080/0142159X.2021.1992373] [Medline: 34709949]
157. O'Connor S, Kennedy S, Wang Y, Ali A, Cooke S, Booth RG. Theories informing technology enhanced learning in nursing and midwifery education: a systematic review and typological classification. Nurse Educ Today 2022 Nov;118:105518 [FREE Full text] [doi: 10.1016/j.nedt.2022.105518] [Medline: 36030581]
158. Patel EA, Aydin A, Cearns M, Dasgupta P, Ahmed K. A systematic review of simulation-based training in neurosurgery, part 1: cranial neurosurgery. World Neurosurg 2020 Jan;133:e850-e873. [doi: 10.1016/j.wneu.2019.08.262] [Medline: 31541755]
159. Sun P, Zhao Y, Men J, Ma ZR, Jiang HZ, Liu CY, et al. Application of virtual and augmented reality technology in hip surgery: systematic review. J Med Internet Res 2023 Mar 10;25:e37599 [FREE Full text] [doi: 10.2196/37599] [Medline: 36651587]
160. Davies H, Sundin D, Robinson S, Jacob E. Does participation in extended immersive ward-based simulation improve the preparedness of undergraduate bachelor's degree nursing students to be ready for clinical practice as a registered nurse? An integrative literature review. J Clin Nurs 2021 Oct 19;30(19-20):2897-2911. [doi: 10.1111/jocn.15796] [Medline: 33870592]
161. Al Asmri MA, Ennis J, Stone RJ, Bello F, Haque MS, Parle J. Effectiveness of technology-enhanced simulation in teaching digital rectal examination: a systematic review narrative synthesis. BMJ Simul Technol Enhanc Learn 2021 Sep 01;7(5):414-421 [FREE Full text] [doi: 10.1136/bmjstel-2020-000587] [Medline: 35515732]
162. Susilawati E, Johari A, Marzal J, Anggereini E. Effects of multimedia e-books and augmented reality on knowledge and skills of health sciences students: a systematic review. J Client Centered Nurs Care 2022;8(2):99-110. [doi: 10.32598/JCCNC.8.2.380.2]
163. Taba JV, Cortez VS, Moraes WA, Iuamoto LR, Hsing WT, Suzuki MO, et al. The development of laparoscopic skills using virtual reality simulations: a systematic review. PLoS One 2021 Jun 17;16(6):e0252609 [FREE Full text] [doi: 10.1371/journal.pone.0252609] [Medline: 34138901]

XSL•FO
RenderX

164. Maheu-Cadotte MA, Cossette S, Dubé V, Fontaine G, Lavallée A, Lavoie P, et al. Efficacy of serious games in healthcare professions education: a systematic review and meta-analysis. Simul Healthc 2021 Jun 01;16(3):199-212. [doi: 10.1097/SIH.0000000000000512] [Medline: 33196609]
165. Tudor Car L, Kyaw BM, Teo A, Fox TE, Vimalesvaran S, Apfelbacher C, et al. Outcomes, measurement instruments, and their validity evidence in randomized controlled trials on virtual, augmented, and mixed reality in undergraduate medical education: systematic mapping review. JMIR Serious Games 2022 Apr 13;10(2):e29594 [FREE Full text] [doi: 10.2196/29594] [Medline: 35416789]
166. Brown L, Ilhan E, Pacey V, Hau W, van der Kooi V, Dale M. The effect of high-fidelity simulation–based learning in acute cardiorespiratory physical therapy—a mixed-methods systematic review. J Phys Ther. Educ 2021 Apr 14;35(2):146-158. [doi: 10.1097/jte.0000000000000183]
167. Portelli M, Bianco S, Bezzina T, Abela JE. Virtual reality training compared with apprenticeship training in laparoscopic surgery: a meta-analysis. Ann R Coll Surg Engl 2020 Nov;102(9):672-684 [FREE Full text] [doi: 10.1308/rcsann.2020.0178] [Medline: 32820649]
168. Lee R, Raison N, Lau WY, Aydin A, Dasgupta P, Ahmed K, et al. A systematic review of simulation-based training tools for technical and non-technical skills in ophthalmology. Eye (Lond) 2020 Oct 13;34(10):1737-1759 [FREE Full text] [doi: 10.1038/s41433-020-0832-1] [Medline: 32203241]
169. Jin C, Dai L, Wang T. The application of virtual reality in the training of laparoscopic surgery: a systematic review and meta-analysis. Int J Surg 2021 Mar;87:105859 [FREE Full text] [doi: 10.1016/j.ijsu.2020.11.022] [Medline: 33307244]
170. Engberg M, Taudorf M, Rasmussen NK, Russell L, Lönn L, Konge L. Training and assessment of competence in resuscitative endovascular balloon occlusion of the aorta (REBOA) - a systematic review. Injury 2020 Feb;51(2):147-156. [doi: 10.1016/j.injury.2019.11.036] [Medline: 31810637]
171. Hippe DS, Umoren RA, McGee A, Bucher SL, Bresnahan BW. A targeted systematic review of cost analyses for implementation of simulation-based education in healthcare. SAGE Open Med 2020;8:2050312120913451 [FREE Full text] [doi: 10.1177/2050312120913451] [Medline: 32231781]
172. Youhasan P, Chen Y, Lyndon M, Henning MA. Exploring the pedagogical design features of the flipped classroom in undergraduate nursing education: a systematic review. BMC Nurs 2021 Mar 22;20(1):50 [FREE Full text] [doi: 10.1186/s12912-021-00555-w] [Medline: 33752654]
173. Rangarajan K, Davis H, Pucher PH. Systematic review of virtual haptics in surgical simulation: a valid educational tool? J Surg Educ 2020 Mar;77(2):337-347. [doi: 10.1016/j.jsurg.2019.09.006] [Medline: 31564519]
174. Singh T, Reyes-Portillo JA. Using technology to train clinicians in evidence-based treatment: a systematic review. Psychiatr Serv 2020 Apr 01;71(4):364-377. [doi: 10.1176/appi.ps.201900186] [Medline: 31960775]
175. Umphrey L, Lenhard N, Lam SK, Hayward NE, Hecht S, Agrawal P, et al. Virtual global health in graduate medical education: a systematic review. Int J Med Educ 2022 Aug 31;13:230-248 [FREE Full text] [doi: 10.5116/ijme.62eb.94fa] [Medline: 36057978]
176. Asegid A, Assefa N. Effect of simulation-based teaching on nursing skill performance: a systematic review and meta-analysis. Front Nurs 2021;8(3). [doi: 10.2478/fon-2021-0021]
177. Jones A, Wilson I, McClean S, Kerr D, Breen C. Supporting the learning experience of health ‐ related profession students during clinical placements with technology: a systematic review. Rev Educ 2022 Aug 16;10(2):e3364. [doi: 10.1002/rev3.3364]
178. Khalaf K, El-Kishawi M, Mustafa S, Al Kawas S. Effectiveness of technology-enhanced teaching and assessment methods of undergraduate preclinical dental skills: a systematic review of randomized controlled clinical trials. BMC Med Educ 2020 Aug 28;20(1):286 [FREE Full text] [doi: 10.1186/s12909-020-02211-4] [Medline: 32859187]
179. Olexa J, Cohen J, Alexander T, Brown C, Schwartzbauer GG, Woodworth GF. Expanding educational frontiers in neurosurgery: current and future uses of augmented reality. Neurosurgery 2023 Feb 01;92(2):241-250. [doi: 10.1227/neu.0000000000002199] [Medline: 36637263]
180. Berthold DP, Muench LN, Rupp MC, Siebenlist S, Cote MP, Mazzocca AD, et al. Head-mounted display virtual reality is effective in orthopaedic training: a systematic review. Arthrosc Sports Med Rehabil 2022 Oct;4(5):e1843-e1849 [FREE Full text] [doi: 10.1016/j.asmr.2022.05.009] [Medline: 36312724]
181. Kovoor JG, Gupta AK, Gladman MA. Validity and effectiveness of augmented reality in surgical education: a systematic review. Surgery 2021 Jul;170(1):88-98. [doi: 10.1016/j.surg.2021.01.051] [Medline: 33744003]
182. Pires C. Perceptions of pharmacy students on the e-learning strategies adopted during the COVID-19 pandemic: a systematic review. Pharmacy (Basel) 2022 Feb 15;10(1):31 [FREE Full text] [doi: 10.3390/pharmacy10010031] [Medline: 35202080]
183. Helming AG, Adler DS, Keltner C, Igelman AG, Woodworth GE. The content quality of YouTube videos for professional medical education: a systematic review. Acad Med 2021 Oct 01;96(10):1484-1493. [doi: 10.1097/ACM.0000000000004121] [Medline: 33856363]
184. Gordon M, Patricio M, Horne L, Muston A, Alston SR, Pammi M, et al. Developments in medical education in response to the COVID-19 pandemic: a rapid BEME systematic review: BEME Guide No. 63. Med Teach 2020 Nov;42(11):1202-1215. [doi: 10.1080/0142159X.2020.1807484] [Medline: 32847456]

XSL•FO
RenderX

185. Chou WK, Ullah N, Arjomandi Rad A, Vardanyan R, Shah V, Zubarevich A, et al. Simulation training for obstetric emergencies in low- and lower-middle income countries: a systematic review. Eur J Obstet Gynecol Reprod Biol 2022 Sep;276:74-81. [doi: 10.1016/j.ejogrb.2022.07.003] [Medline: 35820293]
186. Sultana N, Betran AP, Khan KS, Sobhy S. Simulation-based teaching and models for caesarean sections: a systematic review to evaluate the tools for the 'See One, Practice Many, Do One' slogan. Curr Opin Obstet Gynecol 2020 Oct;32(5):305-315. [doi: 10.1097/GCO.0000000000000655] [Medline: 32796165]
187. Berry MC, de M Neto JM, de Souza MI, Figueredo CM, Reher V, Evans JL. Effectiveness of technology-enhanced learning to improve periodontics educational outcomes: a systematic review. J Dent Educ 2020 Jul 18;84(7):830-839. [doi: 10.1002/jdd.12179] [Medline: 32421234]
188. Nielsen CA, Lönn L, Konge L, Taudorf M. Simulation-based virtual-reality patient-specific rehearsal prior to endovascular procedures: a systematic review. Diagnostics (Basel) 2020 Jul 20;10(7):500 [FREE Full text] [doi: 10.3390/diagnostics10070500] [Medline: 32698437]
189. Higgins M, Madan C, Patel R. Development and decay of procedural skills in surgery: a systematic review of the effectiveness of simulation-based medical education interventions. Surgeon 2021 Aug;19(4):e67-e77. [doi: 10.1016/j.surge.2020.07.013] [Medline: 32868158]
190. Asadzandi S, Mojtahedzadeh R, Mohammadi A. What are the factors that enhance metacognitive skills in nursing students? A systematic review. Iran J Nurs Midwifery Res 2022;27(6):475-484 [FREE Full text] [doi: 10.4103/ijnmr.ijnmr_247_21] [Medline: 36712305]
191. Delungahawatta T, Dunne SS, Hyde S, Halpenny L, McGrath D, O'Regan A, et al. Advances in e-learning in undergraduate clinical medicine: a systematic review. BMC Med Educ 2022 Oct 07;22(1):711 [FREE Full text] [doi: 10.1186/s12909-022-03773-1] [Medline: 36207721]
192. Lee J, Campbell S, Choi M, Bae J. Authentic learning in healthcare education: a systematic review. Nurse Educ Today 2022 Dec;119:105596. [doi: 10.1016/j.nedt.2022.105596] [Medline: 36283196]
193. Keinänen AL, Lähdesmäki R, Juntunen J, Tuomikoski AM, Kääriäinen M, Mikkonen K. Effectiveness of mentoring education on health care professionals´ mentoring competence: a systematic review. Nurse Educ Today 2023 Feb;121:105709 [FREE Full text] [doi: 10.1016/j.nedt.2023.105709] [Medline: 36638727]
194. Sezgin MG, Bektas H. Effectiveness of interprofessional simulation-based education programs to improve teamwork and communication for students in the healthcare profession: a systematic review and meta-analysis of randomized controlled trials. Nurse Educ Today 2023 Jan;120:105619. [doi: 10.1016/j.nedt.2022.105619] [Medline: 36343420]
195. Svellingen AH, Søvik MB, Røykenes K, Brattebø G. The effect of multiple exposures in scenario-based simulation-a mixed study systematic review. Nurs Open 2021 Jan 29;8(1):380-394 [FREE Full text] [doi: 10.1002/nop2.639] [Medline: 33318846]
196. Woon AP, Mok WQ, Chieng YJ, Zhang HM, Ramos P, Mustadi HB, et al. Effectiveness of virtual reality training in improving knowledge among nursing students: a systematic review, meta-analysis and meta-regression. Nurse Educ Today 2021 Mar;98:104655. [doi: 10.1016/j.nedt.2020.104655] [Medline: 33303246]
197. Sleiman Z, Bitar R, Christoforou C, Török P, Fichera M, Panella MM, et al. Is there proof of transferability of laparoscopic psychomotor skills from the simulator laboratory to the operating room? Results from a systematic review. J Obstet Gynaecol 2022 Feb 24;42(2):181-187 [FREE Full text] [doi: 10.1080/01443615.2021.1904216] [Medline: 34027781]
198. Lakhani S, Selim OA, Saeed MZ. Arthroscopic simulation: the future of surgical training: a systematic review. JBJS Rev 2021 Mar;9(3):Ae20.00076. [doi: 10.2106/jbjs.rvw.20.00076] [Medline: 33750750]
199. Conte DB, Zancanaro M, Guollo A, Schneider LR, Lund RG, Rodrigues-Junior SA. Educational interventions to improve dental anatomy carving ability of dental students: a systematic review. Anat Sci Educ 2021 Jan 08;14(1):99-109. [doi: 10.1002/ase.2004] [Medline: 32717118]
200. Chawla S, Devi S, Calvachi P, Gormley WB, Rueda-Esteban R. Evaluation of simulation models in neurosurgical training according to face, content, and construct validity: a systematic review. Acta Neurochir (Wien) 2022 Apr 04;164(4):947-966 [FREE Full text] [doi: 10.1007/s00701-021-05003-x] [Medline: 35122126]
201. Iop A, El-Hajj VG, Gharios M, de Giorgio A, Monetti FM, Edström E, et al. Extended reality in neurosurgical education: a systematic review. Sensors (Basel) 2022 Aug 14;22(16):6067 [FREE Full text] [doi: 10.3390/s22166067] [Medline: 36015828]
202. Au ML, Tong LK, Li YY, Ng WI, Wang SC. Impact of scenario validity and group size on learning outcomes in high-fidelity simulation: a systematics review and meta-analysis. Nurse Educ Today 2023 Feb;121:105705. [doi: 10.1016/j.nedt.2022.105705] [Medline: 36599250]
203. Hovgaard LH, Al-Shahrestani F, Andersen SA. Current evidence for simulation-based training and assessment of myringotomy and ventilation tube insertion: a systematic review. Otol Neurotol 2021 Oct 01;42(9):e1188-e1196. [doi: 10.1097/MAO.0000000000003268] [Medline: 34267097]
204. Tokas T, Ortner G, Peteinaris A, Somani BK, Herrmann T, Nagele U, et al. Simulation training in transurethral resection/laser vaporization of the prostate; evidence from a systematic review by the European Section of Uro-Technology. World J Urol 2022 May;40(5):1091-1110. [doi: 10.1007/s00345-021-03886-0] [Medline: 34800135]

XSL•FO
RenderX

205. Vallée A, Blacher J, Cariou A, Sorbets E. Blended learning compared to traditional learning in medical education: systematic review and meta-analysis. J Med Internet Res 2020 Aug 10;22(8):e16504 [FREE Full text] [doi: 10.2196/16504] [Medline: 32773378]
206. Gao J, Yang L, Zou J, Fan X. Comparison of the influence of massive open online courses and traditional teaching methods in medical education in China: a meta-analysis. Biochem Mol Biol Educ 2021 Jul 24;49(4):639-651 [FREE Full text] [doi: 10.1002/bmb.21523] [Medline: 33894023]
207. Baashar Y, Alkawsi G, Ahmad WN, Alhussian H, Alwadain A, Capretz LF, et al. Effectiveness of using augmented reality for training in the medical professions: meta-analysis. JMIR Serious Games 2022 Jul 05;10(3):e32715 [FREE Full text] [doi: 10.2196/32715] [Medline: 35787488]
208. Sinacori B, Williams-Gregory M. The effect of distance learning on knowledge acquisition in undergraduate second-degree nursing students: a systematic review. Nurs Educ Perspect 2021 Dec 16;42(3):136-141. [doi: 10.1097/01.NEP.0000000000000756] [Medline: 33323720]
209. ShahAli S, Shahabi S, Kohan N, Ebrahimi Takamjani I, Ebrahimi R. Using e-learning methods for physiotherapy students learning – a systematic review and meta-analysis of the impact on knowledge, skills, satisfaction and attitudes. Eur J Physiother 2022 Jun 17;25(5):252-264. [doi: 10.1080/21679169.2022.2085789]
210. Chan J, Pangal DJ, Cardinal T, Kugener G, Zhu Y, Roshannai A, et al. A systematic review of virtual reality for the assessment of technical skills in neurosurgery. Neurosurg Focus 2021 Aug;51(2):E15. [doi: 10.3171/2021.5.FOCUS21210] [Medline: 34333472]
211. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 28;21(2):e12945 [FREE Full text] [doi: 10.2196/12945] [Medline: 30816846]
212. Astbury J, Ferguson J, Silverthorne J, Willis S, Schafheutle E. High-fidelity simulation-based education in pre-registration healthcare programmes: a systematic review of reviews to inform collaborative and interprofessional best practice. J Interprof Care 2021 Jun 12;35(4):622-632. [doi: 10.1080/13561820.2020.1762551] [Medline: 32530344]
213. Daniel R, McKechnie T, Kruse CC, Levin M, Lee Y, Doumouras AG, et al. Video-based coaching for surgical residents: a systematic review and meta-analysis. Surg Endosc 2023 Feb 23;37(2):1429-1439 [FREE Full text] [doi: 10.1007/s00464-022-09379-4] [Medline: 35739431]
214. Lee W, Kim GU, Yoon HJ, Kim S. A systematic review of the effect of web-based trauma-education programs for mental-health professionals. J Korean Acad Psychiatr Ment Health Nurs 2020 Dec;29(4):325-338. [doi: 10.12934/jkpmhn.2020.29.4.325]
215. Gosak L, Štiglic G, Budler LC, Félix IB, Braam K, Fijačko N, et al. Digital tools in behavior change support education in health and other students: a systematic review. Healthcare (Basel) 2021 Dec 21;10(1):1 [FREE Full text] [doi: 10.3390/healthcare10010001] [Medline: 35052165]
216. Kanika, Kaur H. Effect of blended learning on academic achievements and attitude of nursing students: a systematic review. Indian J Forensic Med Toxicol 2020 Jul 30;14(3):383-387. [doi: 10.37506/ijfmt.v14i3.10389]
217. Chan SC, Choa G, Kelly J, Maru D, Rashid MA. Implementation of virtual OSCE in health professions education: a systematic review. Med Educ 2023 Sep 20;57(9):833-843. [doi: 10.1111/medu.15089] [Medline: 37080907]
218. Legoux C, Gerein R, Boutis K, Barrowman N, Plint A. Retention of critical procedural skills after simulation training: a systematic review. AEM Educ Train 2021 Jul 16;5(3):e10536 [FREE Full text] [doi: 10.1002/aet2.10536] [Medline: 34099989]
219. Nayar SK, Musto L, Baruah G, Fernandes R, Bharathan R. Self-assessment of surgical skills: a systematic review. J Surg Educ 2020 Mar;77(2):348-361. [doi: 10.1016/j.jsurg.2019.09.016] [Medline: 31582350]
220. Rashidian N, Vierstraete M, Alseidi A, Troisi RI, Willaert W. Surgical education interventions in liver surgery: a systematic review. Updates Surg 2020 Sep 27;72(3):583-594. [doi: 10.1007/s13304-020-00766-x] [Medline: 32342345]
221. Santos VA, Barreira MP, Saad KR. Technological resources for teaching and learning about human anatomy in the medical course: systematic review of literature. Anat Sci Educ 2022 Mar;15(2):403-419. [doi: 10.1002/ase.2142] [Medline: 34664384]
222. Mydin FH, Yuen CW, Othman S. The effectiveness of educational intervention in improving primary health-care service providers' knowledge, identification, and management of elder abuse and neglect: a systematic review. Trauma Violence Abuse 2021 Oct 11;22(4):944-960. [doi: 10.1177/1524838019889359] [Medline: 31823685]
223. Jallad ST, Işık B. The effectiveness of virtual reality simulation as learning strategy in the acquisition of medical skills in nursing education: a systematic review. Ir J Med Sci 2022 Jun 05;191(3):1407-1426. [doi: 10.1007/s11845-021-02695-z] [Medline: 34227032]
224. Ayivi-Vinz G, Bakwa Kanyinga F, Bergeron L, Décary S, Adisso É, Zomahoun HT, et al. Use of the CPD-REACTION questionnaire to evaluate continuing professional development activities for health professionals: systematic review. JMIR Med Educ 2022 May 02;8(2):e36948 [FREE Full text] [doi: 10.2196/36948] [Medline: 35318188]
225. Jhou HJ, Ou-Yang LJ, Lin MH, Chen PH, Ho CL, Lee CH. Different pedagogies for acquisition of knowledge and skill: a systematic review and network meta-analysis. Postgrad Med J 2022 Aug 30;98(1162):604-609. [doi: 10.1136/postgradmedj-2021-140076] [Medline: 33931555]

XSL•FO
RenderX

226. Adewuyi M, Morales K, Lindsey A. Impact of experiential dementia care learning on knowledge, skills and attitudes of nursing students: a systematic literature review. Nurse Educ Pract 2022 Jul;62:103351. [doi: 10.1016/j.nepr.2022.103351] [Medline: 35483217]
227. Tropea J, Bicknell R, Nestel D, Brand CA, Johnson CE, Paul SK, et al. Simulation training in non-cancer palliative care for healthcare workers: a systematic review of controlled studies. BMJ Simul Technol Enhanc Learn 2021 Aug 14;7(4):262-269 [FREE Full text] [doi: 10.1136/bmjstel-2019-000570] [Medline: 35516824]
228. Dromey BP, Peebles DM, Stoyanov DV. A systematic review and meta-analysis of the use of high-fidelity simulation in obstetric ultrasound. Simul Healthc 2021 Feb 01;16(1):52-59 [FREE Full text] [doi: 10.1097/SIH.0000000000000485] [Medline: 32675735]
229. Gawronski O, Thekkan KR, Genna C, Egman S, Sansone V, Erba I, et al. Instruments to evaluate non-technical skills during high fidelity simulation: a systematic review. Front Med (Lausanne) 2022 Nov 3;9:986296 [FREE Full text] [doi: 10.3389/fmed.2022.986296] [Medline: 36405618]
230. Ahmed TM, Hussain B, Siddiqui MA. Can simulators be applied to improve cataract surgery training: a systematic review. BMJ Open Ophthalmol 2020 Sep 08;5(1):e000488 [FREE Full text] [doi: 10.1136/bmjophth-2020-000488] [Medline: 32953997]
231. Corvetto MA, Altermatt FR, Belmar F, Escudero E. Health care simulation as a training tool for epidemic management: a systematic review. Simul Healthc 2023 Dec 01;18(6):382-391. [doi: 10.1097/SIH.0000000000000716] [Medline: 36881436]
232. Chan RJ, Agbejule OA, Yates PM, Emery J, Jefford M, Koczwara B, et al. Outcomes of cancer survivorship education and training for primary care providers: a systematic review. J Cancer Surviv 2022 Apr 24;16(2):279-302 [FREE Full text] [doi: 10.1007/s11764-021-01018-6] [Medline: 33763806]
233. Lin JC, Yu Z, Scott IU, Greenberg PB. Virtual reality training for cataract surgery operating performance in ophthalmology trainees. Cochrane Database Syst Rev 2021 Dec 21;12(12):CD014953 [FREE Full text] [doi: 10.1002/14651858.CD014953.pub2] [Medline: 34931701]
234. Balakrishnan A, Puthean S, Satheesh G, M K U, Rashid M, Nair S, et al. Effectiveness of blended learning in pharmacy education: a systematic review and meta-analysis. PLoS One 2021 Jun 17;16(6):e0252461 [FREE Full text] [doi: 10.1371/journal.pone.0252461] [Medline: 34138880]
235. Chandran VP, Balakrishnan A, Rashid M, Pai Kulyadi G, Khan S, Devi ES, et al. Mobile applications in medical education: a systematic review and meta-analysis. PLoS One 2022 Mar 24;17(3):e0265927 [FREE Full text] [doi: 10.1371/journal.pone.0265927] [Medline: 35324994]
236. Heuer A, Bienstock J, Zhang Y. Simulation-based training within selected allied health professions: an evidence-based systematic review. J Allied Health 2022;51(1):59-71. [Medline: 35239763]
237. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
238. Verville L, Dc PC, Grondin D, Mior S, Moodley K, Kay R, et al. Using technology-based educational interventions to improve knowledge about clinical practice guidelines. J Chiropr Educ 2021 Mar 01;35(1):149-157 [FREE Full text] [doi: 10.7899/JCE-19-17] [Medline: 32931558]
239. Moussa R, Alghazaly A, Althagafi N, Eshky R, Borzangy S. Effectiveness of virtual reality and interactive simulators on dental education outcomes: systematic review. Eur J Dent 2022 Feb 24;16(1):14-31 [FREE Full text] [doi: 10.1055/s-0041-1731837] [Medline: 34428851]
240. Okano H, Mayumi T, Kataoka Y, Banno M, Tsujimoto Y, Shiroshita A, et al. Outcomes of simulation-based education for vascular access: a systematic review and meta-analysis. Cureus 2021 Aug;13(8):e17188 [FREE Full text] [doi: 10.7759/cureus.17188] [Medline: 34414052]
241. Richard O, Jollant F, Billon G, Attoe C, Vodovar D, Piot MA. Simulation training in suicide risk assessment and intervention: a systematic review and meta-analysis. Med Educ Online 2023 Dec 18;28(1):2199469 [FREE Full text] [doi: 10.1080/10872981.2023.2199469] [Medline: 37073473]
242. Ge L, Chen Y, Yan C, Chen Z, Liu J. Effectiveness of flipped classroom vs traditional lectures in radiology education: a meta-analysis. Medicine (Baltimore) 2020 Oct 02;99(40):e22430 [FREE Full text] [doi: 10.1097/MD.0000000000022430] [Medline: 33019421]
243. Le Lous M, Simon O, Lassel L, Lavoue V, Jannin P. Hybrid simulation for obstetrics training: a systematic review. Eur J Obstet Gynecol Reprod Biol 2020 Mar;246:23-28. [doi: 10.1016/j.ejogrb.2019.12.024] [Medline: 31927239]
244. Rueda Esteban RJ, López-McCormick JS, Rodríguez-Bermeo AS, Andrade M, Hernández Restrepo JD, Targarona Soler EM. Face, content, and construct validity evaluation of simulation models in general surgery laparoscopic training and education: a systematic review. Surg Innov 2023 Apr 04;30(2):251-260. [doi: 10.1177/15533506221123704] [Medline: 36062557]
245. Li BZ, Cao NW, Ren CX, Chu XJ, Zhou HY, Guo B. Flipped classroom improves nursing students' theoretical learning in China: a meta-analysis. PLoS One 2020 Aug 27;15(8):e0237926 [FREE Full text] [doi: 10.1371/journal.pone.0237926] [Medline: 32853214]

XSL•FO
RenderX

246. Chen J, Xun H, Abousy M, Long C, Sacks JM. No microscope? No problem: a systematic review of microscope-free microsurgery training models. J Reconstr Microsurg 2022 Feb 23;38(2):106-114. [doi: 10.1055/s-0041-1731761] [Medline: 34425592]
247. Schafer KM, Kremer MJ. Outcomes of simulation-based experiences related to family presence during resuscitation: a systematic review. Clin Simul Nurs 2022 Apr;65:62-81. [doi: 10.1016/j.ecns.2022.01.002]
248. Shank PO. Practice and Feedback for Deeper Learning: 26 Evidence-Based and Easy-to-Apply Tactics That Promote Deeper Learning and Application. Scotts Valley, CA: Createspace Independent Publishing; 2017.
249. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
250. Nussbaumer-Streit B, Klerings I, Dobrescu AI, Persad E, Affengruber L, Gartlehner G. Excluding non-English studies from systematic reviews does not change conclusions: a meta-epidemiological study. In: Proceedings of the Cochrane Colloquium 2019. 2019 Presented at: Cochrane Colloquium 2019; October 22-25, 2019; Santiago, Chile. [doi: 10.1016/j.jclinepi.2019.10.011]
251. Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. Syst Rev 2019 Nov 07;8(1):264 [FREE Full text] [doi: 10.1186/s13643-019-1188-0] [Medline: 31699124]
252. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021 Jul 20;11(4):116-129 [FREE Full text] [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]
253. Experts by experience. Care Quality Commission. URL: https://www.cqc.org.uk/about-us/jobs/experts-experience [accessed 2024-07-08]

Abbreviations

AI: artificial intelligence
NHS: National Health Service
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
XR: extended reality

Edited by A Mavragani; submitted 05.09.24; peer-reviewed by P Dezyani, ST Arasteh; comments to author 04.10.24; revised version received 10.10.24; accepted 29.10.24; published 19.02.25.

Please cite as:
Potter A, Munsch C, Watson E, Hopkins E, Kitromili S, O'Neill IC, Larbie J, Niittymaki E, Ramsay C, Burke J, Ralph N
Identifying Research Priorities in Digital Education for Health Care: Umbrella Review and Modified Delphi Method Study
J Med Internet Res 2025;27:e66157
URL: https://www.jmir.org/2025/1/e66157
doi:10.2196/66157
PMID:39969988

©Alison Potter, Chris Munsch, Elaine Watson, Emily Hopkins, Sofia Kitromili, Iain Cameron O'Neill, Judy Larbie, Essi Niittymaki, Catriona Ramsay, Joshua Burke, Neil Ralph. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning–Based Prediction of Substance Use in Adolescents in Three Independent Worldwide Cohorts: Algorithm Development and Validation Study

Soeun Kim[1,2*], MS; Hyejun Kim[1,3*], BS; Seokjun Kim[1,4*], MD; Hojae Lee[1], MS; Ahmed Hammoodi[5], BBA; Yujin Choi[1,6], BS; Hyeon Jin Kim[1,2], PhD; Lee Smith[7], PhD; Min Seo Kim[8], MD; Guillaume Fond[9], MD, PhD; Laurent Boyer[9], MD, PhD; Sung Wook Baik[10], PhD; Hayeon Lee[1,11], PhD; Jaeyu Park[1,2], PhD; Rosie Kwon[1], MS; Selin Woo[1,4], PhD; Dong Keon Yon[1,2,4,12], MD, PhD

[1]Center for Digital Health, Medical Science Research Institute, Kyung Hee University Medical Center, Seoul, Republic of Korea
[2]Department of Precision Medicine, Kyung Hee University College of Medicine, Seoul, Republic of Korea
[3]Department of Applied Information Engineering, Yonsei University, Seoul, Republic of Korea
[4]Department of Medicine, Kyung Hee University College of Medicine, Seoul, Republic of Korea
[5]Department of Business Administration, Kyung Hee University School of Management, Seoul, Republic of Korea
[6]Department of Korean Medicine, Kyung Hee University College of Korean Medicine, Seoul, Republic of Korea
[7]Centre for Health, Performance and Wellbeing, Anglia Ruskin University, Cambridge, United Kingdom
[8]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, United States
[9]Research Centre on Health Services and Quality of Life, Assistance Publique-Hopitaux de Marseille, Aix Marseille University, Marseille, France
[10]Department of Software, Sejong University College of Electronics and Information Engineering, Seoul, Republic of Korea
[11]Department of Electronics and Information Convergence Engineering, Kyung Hee University, Seoul, Republic of Korea
[12]Department of Pediatrics, Kyung Hee University Medical Center, Kyung Hee University College of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Dong Keon Yon, MD, PhD
Department of Pediatrics
Kyung Hee University Medical Center
Kyung Hee University College of Medicine
23 Kyungheedae-ro, Dongdaemun-gu
Seoul, 02447
Republic of Korea
Phone: 82 269352476
Fax: 82 5044780201
Email: yonkkang@gmail.com

Abstract

Background: To address gaps in global understanding of cultural and social variations, this study used a high-performance machine learning (ML) model to predict adolescent substance use across three national datasets.

Objective: This study aims to develop a generalizable predictive model for adolescent substance use using multinational datasets and ML.

Methods: The study used the Korea Youth Risk Behavior Web-Based Survey (KYRBS) from South Korea (n=1,098,641) to train ML models. For external validation, we used the Youth Risk Behavior Survey (YRBS) from the United States (n=2,511,916) and Norwegian nationwide Ungdata surveys (Ungdata) from Norway (n=700,660). After developing various ML models, we evaluated the final model's performance using multiple metrics. We also assessed feature importance using traditional methods and further analyzed variable contributions through SHapley Additive exPlanation values.

Results: The study used nationwide adolescent datasets for ML model development and validation, analyzing data from 1,098,641 KYRBS adolescents, 2,511,916 YRBS participants, and 700,660 from Ungdata. The XGBoost model was the top performer on the KYRBS, achieving an area under receiver operating characteristic curve (AUROC) score of 80.61% (95% CI 79.63-81.59)

XSL•FO
RenderX

and precision of 30.42 (95% CI 28.65-32.16) with detailed analysis on sensitivity of 31.30 (95% CI 29.47-33.20), specificity of 99.16 (95% CI 99.12-99.20), accuracy of 98.36 (95% CI 98.31-98.42), balanced accuracy of 65.23 (95% CI 64.31-66.17), F_1-score of 30.85 (95% CI 29.25-32.51), and area under precision-recall curve of 32.14 (95% CI 30.34-33.95). The model achieved an AUROC score of 79.30% and a precision of 68.37% on the YRBS dataset, while in external validation using the Ungdata dataset, it recorded an AUROC score of 76.39% and a precision of 12.74%. Feature importance and SHapley Additive exPlanation value analyses identified smoking status, BMI, suicidal ideation, alcohol consumption, and feelings of sadness and despair as key contributors to the risk of substance use, with smoking status emerging as the most influential factor.

Conclusions: Based on multinational datasets from South Korea, the United States, and Norway, this study shows the potential of ML models, particularly the XGBoost model, in predicting adolescent substance use. These findings provide a solid basis for future research exploring additional influencing factors or developing targeted intervention strategies.

(J Med Internet Res 2025;27:e62805) doi:10.2196/62805

KEYWORDS

adolescents; machine learning; substance; prediction; XGBoost; random forest; ML; substance use; adolescents; adolescent; South Korea; United States; Norway; web-based survey; survey; risk behavior; smoking; alcohol; intervention; interventions

Introduction

Substance use among adolescents remains a global concern, often leading to both immediate and long-term health challenges, such as mental health disorders and addiction [1]. When initiated at an early age, these behaviors can escalate to more serious health conditions, including chronic substance dependence and comorbid mental health issues [2]. As globalization and cultural integration continue to expand, substance use patterns vary significantly across regions, making it paramount to understand these patterns across a diverse cultural landscape [3]. Conventional statistical methods have long been used to examine the predictors and outcomes of adolescent substance use [4]. However, these approaches often fall short in capturing complex, nonlinear relationships between variables. Therefore, with recent advancements, machine learning (ML) has introduced powerful tools capable of identifying complex patterns and relationships, offering a deeper understanding of adolescent substance use [5].

Existing studies provide insights into the epidemiology and sociocultural factors associated with adolescent substance use in various contexts, but few have used ML techniques with multinational datasets [4,6]. Such approaches have the potential to yield more precise and globally relevant insights [5,7]. By identifying key predictors of adolescent substance use that remain consistent across diverse cultural contexts, this study aims to develop a prediction model adaptable to global public health initiatives. To validate its generalizability, the model was tested using datasets from two additional countries, highlighting its adaptability to diverse sociocultural environments [8].

This study developed an ML-based prediction model for adolescent substance use, using comprehensive datasets from South Korea, and extended our validation process by using datasets from the United States and Norway [9]. This validation process and refinement process ensured the model's accuracy and applicability across diverse cultural and national contexts. By integrating these global datasets, we developed a predictive model that reflects our collaborative international research. Our novel approach equips stakeholders with a sophisticated tool informed by global data. This helps address and preempt adolescent substance use effectively across different national contexts.

Methods

Study Design and Participants

Adolescents enrolled in middle and high school who completed their respective surveys were included. In the context of educational systems, adolescents are more appropriately categorized by grade level rather than age. To ensure consistency across the three countries, participants were limited to students from middle school (7th grade) to high school (12th grade). This study was primarily designed to develop an ML model for substance use prediction among adolescents using three distinct nationwide datasets: Korea Youth Risk Behavior Web-Based Survey (KYRBS) from South Korea [10], Youth Risk Behavior Survey (YRBS) from the United States [6,11], and Norwegian nationwide Ungdata surveys (Ungdata) from Norway [12]. KYRBS was initially used to train the ML model, followed by the external validation process using the YRBS and Ungdata.

The discovery dataset, KYRBS, was conducted annually by the Korean Disease Control and Prevention Agency (KDCA) from 2008 to 2022, to assess health behaviors among Korean middle and high school students. It began with 1,128,985 participants and ended up with 1,098,641 after exclusions from the missing values in BMI [13], primarily representing a demographic largely comprised of East Asians (Figure 1A). The KYRBS dataset can be accessed through the official website of the KDCA.

XSL•FO
RenderX

Figure 1. Study population. KYRBS: Korea Youth Risk Behavior Web-Based Survey; Ungdata: Norwegian nationwide Ungdata surveys; YRBS: Youth Risk Behavior Survey.

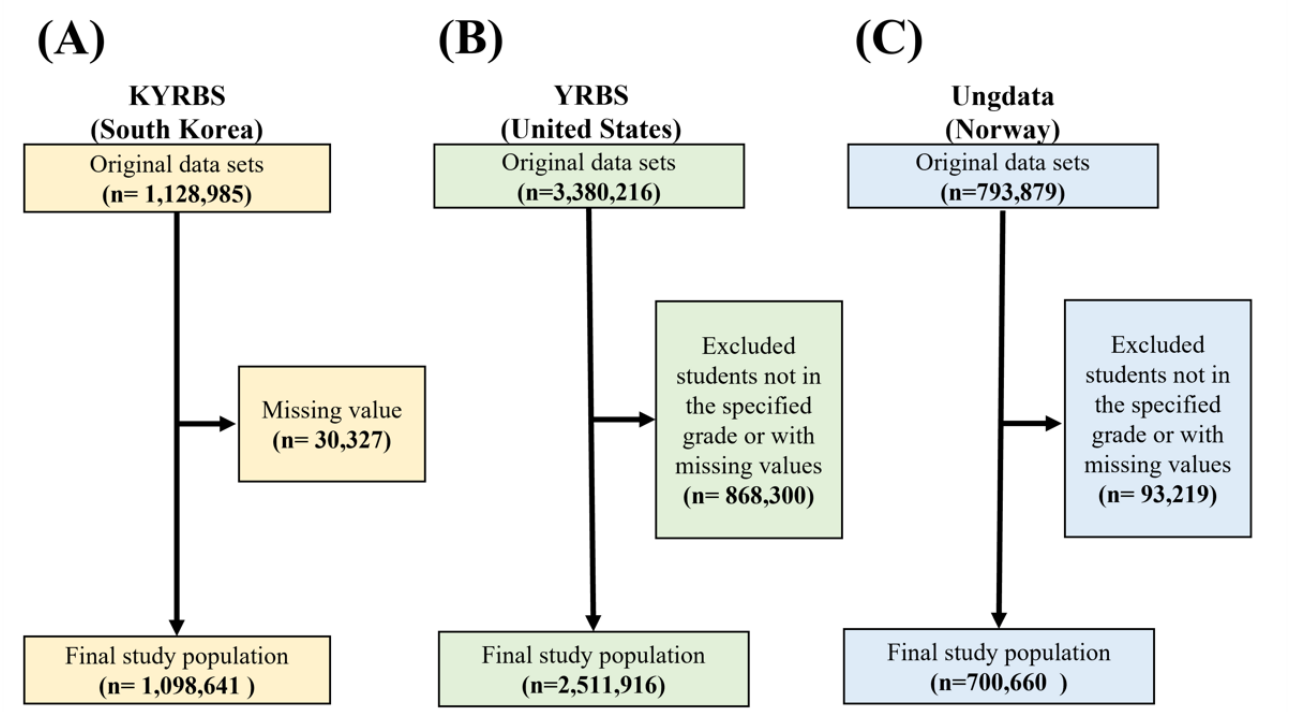

The YRBS, which was conducted by the US Centers for Disease Control and Prevention (CDC), along with state and local education and health agencies, began with 3,380,216 participants, was narrowed down to 2,511,916 after excluding students not enrolled in middle or high school and those with missing data (Figure 1B**)**. YRBS encompasses a diverse racial spectrum of American adolescents. YRBS data are downloaded and available through the US CDC's official website.

Similarly, from the initial count of 793,879 in the Ungdata, a cross-sectional survey conducted by the social research institute, Norwegian Social Research Institute (NOVA), at Oslo Metropolitan University consists of a questionnaire for school pupils throughout, only 700,660 participants were selected after data processing (Figure 1C). Ungdata is offered to all local and county councils in Norway, who administer the questionnaire in collaboration with NOVA and regional centers for substance use rehabilitation. The dataset can be accessed via the official Ungdata website. During the data processing phase, individual missing values in the KYRBS and YRBS datasets were imputed using a random forest regression–based imputation method [14]. For the extravalidation cohorts, YRBS and Ungdata, variables that were entirely absent were imputed using the median values derived from the discovery dataset, KYRBS.

In order to evaluate the generalizability of our model across diverse cultural groups, we used a phased validation approach. The validation process used the YRBS dataset, which includes participants from diverse ethnic backgrounds, as well as the Ung dataset from Norway, representing a markedly different cultural context. By using validation datasets encompassing a wide range of cultural groups, we conducted a rigorous assessment of the model's robustness and versatility across diverse populations [15].

Our primary outcome, substance use, was derived from the question "Have you ever consumed illicit substances at least once in your lifetime." We distinguished smoking and alcohol from other substances in our analysis, recognizing their unique consumption patterns, sociocultural implications, and health effects [16]. Substances other than smoking and alcohol are distinguished primarily due to concerns regarding their potential for misuse and health risks [17]. This decision was made to ensure that our model captures nuances specific to each substance, thereby enhancing the specificity and relevance of our predictions. Integral covariates under consideration spanned across factors including grade, sex, region, BMI, academic achievement, household income, smoking status, alcoholic consumption, stress status, sadness and despair, suicidal thinking, and suicide attempts [13].

Model Development and Validation

Using the KYRBS, we developed a predictive model to extrapolate the behavioral patterns of Korean adolescents regarding substance use. Due to rigorous substance regulations and law enforcement measures in South Korea, accessibility and consumption of substances are notably limited [18]. Consequently, the number of instances representing substance use within the KYRBS was sparse (n=12,803, 1.17%).

Given the intricate nature of the data, we used a comprehensive approach to model development, leveraging 10-fold cross-validation to divide the KYRBS dataset into training and testing subsets. Various tree-based and statistical models, including LightGBM, CatBoost, AdaBoost, random forest, and XGBoost, were evaluated in comparison with logistic regression, a widely used baseline model, to determine the most effective algorithm for predicting substance use [19,20].

XSL•FO
RenderX

After model development using the KYRBS, external validation was conducted with datasets from diverse cultural contexts (YRBS and Ungdata) to evaluate the generalizability of the predictive model. This step ensured the applicability of the modeling approach across heterogeneous adolescent populations [21].

To further strengthen the validity of our results, hyperparameter tuning was performed for each algorithm using GridSearchCV, focusing on maximizing performance metrics such as the area under the receiver operating characteristic curve (AUROC) and precision. The hyperparameter values for the selected model are detailed in Table S1 in Multimedia Appendix 1. The sensitivity and specificity of the model were determined based on a classification threshold of 0.5. Various metrics, including AUROC score, accuracy, sensitivity, specificity, balanced accuracy, precision, F_1-score (ie, the harmonic mean of the precision and recall), and the area under precision-recall curve (AUPRC) were used to evaluate the model's performance across datasets [13].

Performance Assessment

The tools and techniques we used for assessment were consistent. Our evaluation metrics comprised AUROC, accuracy, sensitivity, specificity, balanced accuracy, precision, F_1-score, and AUPRC. These metrics were collectively considered to comprehensively evaluate and compare model performance. To provide a visual representation of the model efficacy, we used visualization techniques, notably the ROC curve [19,20,22-24].

SHapley Additive ExPlanation Value Analysis

To interpret and better understand the model's predictions, we calculated the SHapley Additive explanation (SHAP) values based on the summary model [25]. SHAP is a widely used method for providing local explanations of ML model predictions. Proposed by Lundberg and Lee, SHAP offers a unified framework for explaining the output of any ML model [25]. This technique visualizes the contribution of each feature to the model's prediction for a specific instance, illustrating how each feature shifts the model's output from the base value.

Software and Libraries

All computations, model training, validation, and evaluation processes were executed using Python (version 3.12.4; Python Software Foundation). Key libraries from our toolbox included Scikit-learn (version 1.5.2; Scikit-learn development team), NumPy (version 1.26.4; Python Software Foundation), and Pandas (version 2.2.0; Python Software Foundation) for ML tasks and data wrangling. Visualization was facilitated using Matplotlib (version 3.8.4; Python Software Foundation) and Seaborn (version 0.13.2; Python Software Foundation).

Ethical Considerations

The study protocol was approved by the institutional review board of the KDCA (2014-06EXP-02-P-A), US CDC (#1969.0), and NOVA (18778329) [12], and all participants provided written informed consent. This research followed the guidelines outlined in the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) statement [26].

Results

Demographic Characteristics

This research used a detailed exploration using nationwide adolescent datasets from South Korea, aiming to design and validate an ML model to predict substance use tendencies among adolescents. The primary demographic consisted of middle (7th grade) to high school (12th grade) students (Figure 1).

We collected data from the KYBS, the YRBS, and the Ungdata, and subsequently standardized the covariates for the ML predictive modeling process. Within the primary cohort from the KYRBS to develop the prediction model, the sex distribution was as follows: male (n=566,437, 51.56%) and female (n=532,204, 48.44%). For the initial external validation cohorts of an external validation process, the YRBS features a sex distribution of: male (n=1,233,846, 49.12%) and female (n=1,278,070, 50.88%). In the second validation step, the Ungdata has the following sex distribution: male (n=345,428, 49.30%) and female (n=355,232, 50.70%; Table 1).

XSL•FO
RenderX

Table 1. Demographic characteristics of KYRBS[a] from South Korea (2005-2022), YRBS[b] from the United States (1998-2022), and Ungdata from Norway (2014-2021).

	South Korea (KYRBS; n=1,098,641), n (%)	United States (YRBS; n=2,511,916), n (%)	Norway (Ungdata; n=700,660), n (%)
Region			
Urban	509,058 (46.34)	N/A[c]	N/A
Rural	589,583 (53.66)	N/A	N/A
Grade			
7th grade	190,112 (17.30)	448,865 (17.87)	144,229 (20.58)
8th grade	190,166 (17.31)	402,753 (16.03)	141,806 (20.24)
9th grade	189,842 (17.28)	476,360 (18.96)	143,938 (20.54)
10th grade	182,908 (16.65)	448,865 (17.87)	127,053 (18.13)
11th grade	181,428 (16.51)	402,753 (16.03)	89,637 (12.79)
12th grade	164,185 (14.94)	332,320 (13.23)	53,997 (7.71)
Sex			
Male	566,437 (51.56)	1,233,846 (49.12)	345,428 (49.30)
Female	532,204 (48.44)	1,278,070 (50.88)	355,232 (50.70)
BMI[d]			
Unknown	2940 (0.27)	448,630 (17.86)	N/A
Underweight	88,787 (8.08)	62,540 (2.49)	N/A
Normal	830,683 (75.61)	1,428,266 (56.86)	N/A
Overweight	89,971 (8.19)	308,141 (12.27)	N/A
Obese	86,260 (7.85)	264,339 (10.52)	N/A
Academic achievement			
Low (0-19 percentile)	114,832 (10.45)	N/A	N/A
Lower-middle (20-39 percentile)	255,845 (23.29)	N/A	N/A
Middle (40-59 percentile)	314,304 (28.61)	N/A	N/A
Upper-middle (60-79 percentile)	278,359 (25.34)	N/A	N/A
High (80-100 percentile)	135,301 (12.32)	N/A	N/A
Household income			
Low (0-19 percentile)	43,920 (4.00)	N/A	7968 (1.14)
Lower-middle (20-39 percentile)	159,283 (14.50)	N/A	27,885 (3.98)
Middle (40-59 percentile)	516,595 (47.02)	N/A	121,858 (17.39)
Upper-middle (60-79 percentile)	290,304 (26.42)	N/A	236,136 (33.70)
High (80-100 percentile)	88,539 (8.06)	N/A	306,813 (43.79)
Smoking status			
Nonsmoker	869,907 (79.18)	1,179,077 (46.94)	562,118 (80.23)
Smoker	228,734 (20.82)	1,332,839 (53.06)	138,542 (19.77)
Alcoholic consumption			
Nondrinker	581,345 (52.91)	1,335,962 (53.18)	306,210 (43.70)
More than one time	517,296 (47.09)	1,175,954 (46.82)	394,450 (56.30)
Stress status[e]			
Low	30,456 (2.77)	N/A	N/A
Mild	159,720 (14.54)	N/A	N/A

XSL•FO
RenderX

	South Korea (KYRBS; n=1,098,641), n (%)	United States (YRBS; n=2,511,916), n (%)	Norway (Ungdata; n=700,660), n (%)
Moderate	455,026 (41.42)	N/A	N/A
High	324,847 (29.57)	N/A	N/A
Severe	128,592 (11.70)	N/A	N/A
Sadness and despair in the past year			
Unknown	N/A	1,079,817 (42.99)	N/A
No	749,479 (68.22)	1,005,251 (40.02)	522,702 (74.60)
Yes	349,162 (31.78)	426,848 (16.99)	177,958 (25.40)
Suicidal thinking in the past year			
No	914,543 (83.24)	1,506,410 (59.97)	N/A
Yes	184,098 (16.76)	1,005,506 (40.03)	N/A
Suicide attempts in the past year			
No	1,057,142 (96.22)	1,899,009 (75.60)	N/A
Yes	41,499 (3.78)	612,907 (24.40)	N/A
Substance use			
No	1,085,838 (98.83)	1,310,191 (52.16)	685,866 (97.89)
Yes	12,803 (1.17)	1,201,725 (47.84)	14,794 (2.11)

[a]KYRBS: Korea Youth Risk Behavior Web-Based Survey.

[b]YRBS: Youth Risk Behavior Survey.

[c]N/A: not applicable.

[d]BMI was divided into four groups according to the 2017 Korean National Growth Charts: underweight (0-4 percentile), normal (5-84 percentile), overweight (85-94 percentile), and obese (95-100 percentile).

[e]Stress was defined by the receipt of mental health counseling owing to stress.

We compared the distributions of key variables across the three cohorts (KYRBS, YRBS, and Ungdata), as presented in Table 1. Notable similarities and differences were identified. For instance, the proportion of smokers differed significantly between KYRBS (228,734, 20.82%) and YRBS (n=1,332,839, 53.06%), while the Ungdata cohort exhibited a much lower smoking prevalence (n=138,542, 19.77%). Similarly, alcohol consumption rates were higher in the Ungdata cohort (n=394,450, 56.30%) compared to KYRBS (n=517,296, 47.09%) and YRBS (n=1,175,954, 46.82%). These differences likely reflect cultural and policy variations influencing substance accessibility and behavioral norms in each country. Furthermore, visual comparisons of key baseline characteristics, such as smoking and alcohol consumption, are provided in Figure S1 in Multimedia Appendix 1 to offer additional insights.

ML Model Results

Extensive model evaluations, considering both AUROC and precision, revealed that the XGBoost model was the optimal model for predicting substance use among adolescents (Figures 2 and 3). The primary model, sourced from the KYRBS and assessed disclosed that the XGBoost model notched an AUROC score of 80.61% (95% CI 79.63-81.59) and precision of 30.42 (95% CI 28.65-32.16) with detailed analysis on a sensitivity of 31.30 (95% CI 29.47-33.20), specificity of 99.16 (95% CI 99.12-99.20), accuracy of 98.36 (95% CI 98.31-98.42), balanced accuracy of 65.23 (95% CI 64.31-66.17), F_1-score of 30.85 (95% CI 29.25-32.51), and AUPRC of 32.14 (95% CI 30.34-33.95). Other models exhibited the following AUROC scores: random forest at 81.45 (95% CI 80.51-82.37), LightGBM at 80.35 (95% CI 79.35-81.33), AdaBoost at 80.32 (95% CI 79.29-81.31), CatBoost at 77.84 (95% CI 76.84-78.84), and Logistic 77.38 (95% CI 76.39-78.40). Precision scores were as follows: LightGBM at 35.58 (95% CI 33.63-37.58), AdaBoost at 8.94 (95% CI 8.43-9.45), random forest at 4.61 (95% CI 4.39-4.83), Logistic at 3.07 (95% CI 2.92-3.21), and CatBoost at 2.34 (95% CI 2.23-2.44).

XSL•FO
RenderX

Figure 2. Model architecture. The original KYRBS was partitioned into the original data set for model development, with performance assessed using the AUROC score. Selected high-performing models were further validated. The external validations were generated using YRBS and Ungdata. AUROC: area under the receiver operating characteristic curve; KYRBS: Korea Youth Risk Behavior Web-Based Survey; YRBS: Youth Risk Behavior Survey.

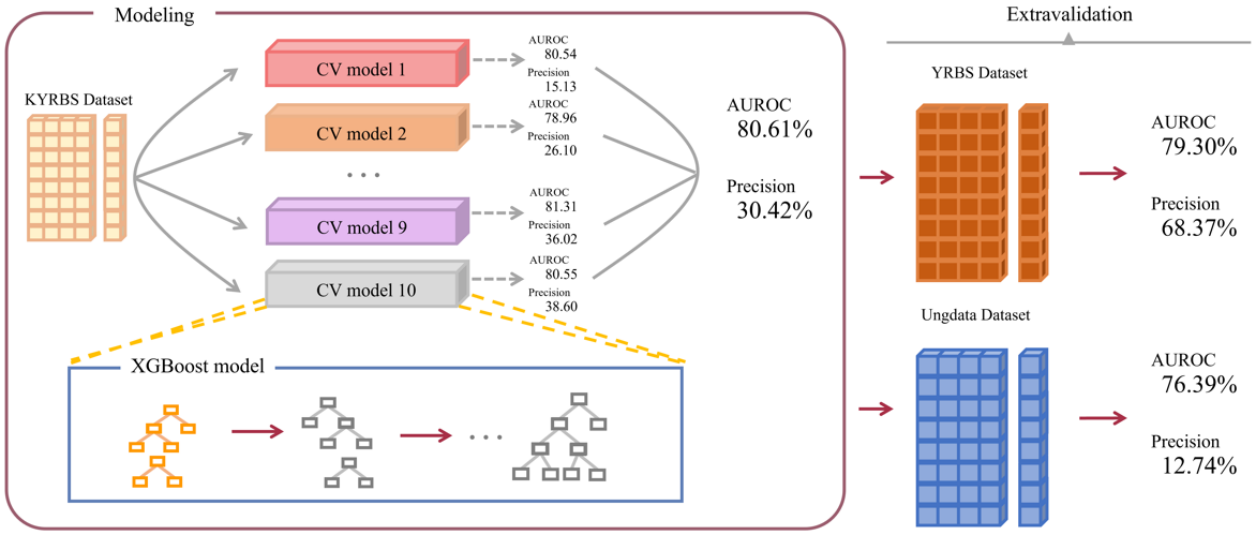

For the initial external validation, the independent YRBS dataset was used. The XGBoost displayed an AUROC score of 79.30%, followed by a precision of 68.37%, sensitivity of 77.77%, specificity of 67%, accuracy of 72.15%, and balanced accuracy of 72.38%, F_1-score of 72.77%, and AUPRC of 74.79%. In the subsequent external validation using the Ung dataset, the XGBoost model yielded an AUROC score of 76.39%, coupled with a precision of 12.74%, sensitivity of 83.02%, specificity of 63.93%, accuracy of 64.34%, balanced accuracy of 73.48%, F_1-score of 2.18, and AUPRC of 7.25 (Figure 3).

Figure 3. The assessment of five different machine learning algorithms using AUROC score and ROC curve for initial model construction with the KYRBS, and external validation with the YRBS and Ungdata. AUROC: area under the receiver operating characteristic curve; KYRBS: Korea Youth Risk Behavior Web-Based Survey; LightGBM: light gradient boosting model; ROC: receiver operating characteristic; XGBoost: extreme Gradient Boosting model; YRBS: Youth Risk Behavior Survey.

Country	Dataset	Model	AUROC	Precision	Sensitivity	Specificity	Accuracy	balanced accuracy	F_1-score	AUPRC
Korea	KYRBS	Xgboost	80.61 (95% CI 79.63-81.59)	30.42 (95% CI 28.65-32.16)	31.30 (95% CI 29.47-33.20)	99.16 (95% CI 99.12-99.20)	98.36 (95% CI 98.31-98.42)	65.23 (95% CI 64.31-66.17)	30.85 (95% CI 29.25-32.51)	32.14 (95% CI 30.34-33.95)
		Randomforest	81.45 (95% CI 80.51-82.37)	4.61 (95% CI 4.39-4.83)	61.63 (95% CI 59.75-63.58)	84.94 (95% CI 84.79-85.09)	84.67 (95% CI 84.52-84.81)	73.28 (95% CI 72.34-74.26)	8.57 (95% CI 8.18-8.98)	20.71 (95% CI 19.15-22.43)
		LightGBM	80.35 (95% CI 79.35-81.33)	35.58 (95% CI 33.63-37.58)	30.52 (95% CI 28.76-32.41)	99.35 (95% CI 99.32-99.38)	98.55 (95% CI 98.49-98.60)	64.93 (95% CI 64.05-65.88)	32.85 (95% CI 31.15-34.57)	32.11 (95% CI 30.34-33.90)
		Adaboost	80.32 (95% CI 79.29 - 81.31)	8.94 (95% CI 8.43-9.45)	44.95 (95% CI 43.03-46.93)	94.60 (95% CI 94.51-94.69)	94.02 (95% CI 93.92-94.12)	69.77 (95% CI 68.82-70.77)	14.92 (95% CI 14.14-15.70)	31.97 (95% CI 30.28-33.84)
		Catboost	77.84 (95% CI 76.84-78.84)	2.34 (95% CI 2.23-2.44)	74.87 (95% CI 73.17-76.49)	63.05 (95% CI 62.86-63.25)	63.19 (95% CI 63.00-63.39)	68.96 (95% CI 68.10-69.79)	4.53 (95% CI 4.34-4.73)	29.38 (95% CI 27.61-31.20)
		Logistic	77.38 (95% CI 76.39-78.40)	3.07 (95% CI 2.92-3.21)	65.64 (95% CI 63.85-67.56)	75.52 (95% CI 75.35-75.71)	75.41 (95% CI 75.23-75.58)	70.58 (95% CI 69.69-71.53)	5.86 (95% CI 5.58-6.12)	6.11 (95% CI 5.55-6.73)
United States	YRBS	Xgboost	79.30	68.37	77.77	67.00	72.15	72.38	72.77	74.79
		Randomforest	78.91	66.69	78.78	63.91	71.02	71.34	72.23	75.39
		LightGBM	70.38	64.22	40.94	79.08	60.83	60.01	50.01	59.84
		Adaboost	80.07	68.05	82.75	64.36	73.16	73.55	74.68	74.9
		Catboost	75.90	65.20	81.65	60.04	70.38	70.84	72.51	69.93
		Logistic	65.93	56.19	74.75	46.54	60.04	60.65	64.15	62.67
Norway	Ungdata	Xgboost	76.39	12.74	83.02	63.93	64.34	73.48	2.18	7.25
		Randomforest	72.96	2.46	96.38	17.46	19.13	56.92	4.79	4.92
		LightGBM	79.45	6.26	7.74	97.5	95.61	52.62	6.92	7.01
		Adaboost	77.42	6.68	35.63	89.26	88.13	62.44	11.25	4.86
		Catboost	75.85	2.21	96.51	7.83	9.70	52.17	4.32	6.68
		Logistic	62.66	2.90	64.99	52.99	53.24	58.99	5.54	3.45

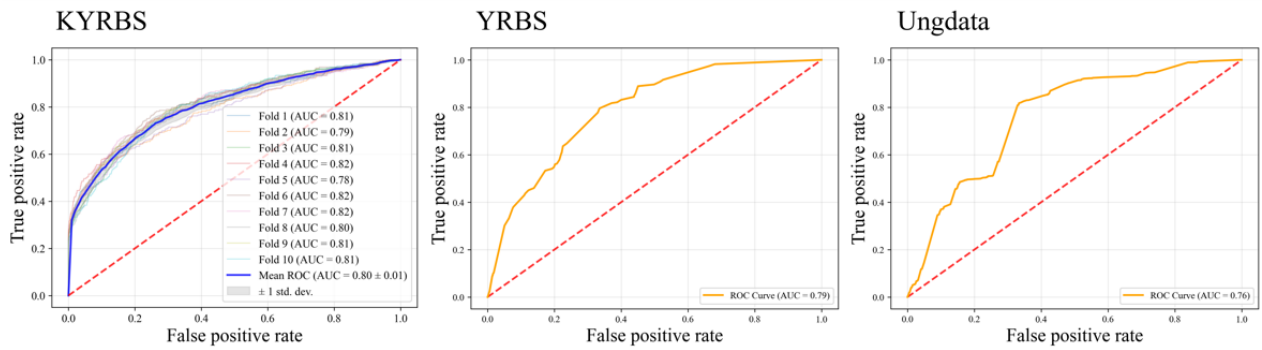

XSL•FO
RenderX

Across all evaluations, both internal and external, the XGBoost model consistently exhibited a predominant performance, particularly in terms of the AUROC score and precision, cementing its adoption for the study objective.

Feature Importance

Table 2 illustrates the importance of various features as determined by the XGBoost model in predicting substance use among adolescents. Specifically, smoking status was identified as the most significant predictor, accounting for 16.62% importance. This was closely followed by BMI with 13.45%, and suicidal thinking at 12.58%. Other notable features include suicide attempts (9.88%), grade (9.83%), stress status (8.99%), academic achievement (7.51%), household income (7.45%), sadness and despair (6.84%), alcoholic consumption (3.78%), sex (2.76%), and region (0.32%).

Table 2. Feature importance of the XGBoost model.

Feature	Importance (%)
Smoking status	16.62
BMI	13.45
Suicidal thinking	12.58
Suicide attempts	9.88
Grade	9.83
Stress status	8.99
Academic achievement	7.51
Household income	7.45
Sadness and despair	6.84
Alcoholic consumption	3.78
Sex	2.76
Region	0.32

SHAP Values

Figure 4 presents the SHAP analysis results for the substance use prediction model, illustrating the contribution of each variable to predicting the likelihood of substance use [25]. The analysis identified smoking status as the most influential variable, with higher smoking levels strongly associated with an increased likelihood of substance use. Similarly, alcoholic consumption and sadness and despair were identified as key factors; alcohol consumption consistently increased the likelihood of substance use, while emotional states such as sadness exhibited both positive and negative effects depending on their levels.

XSL•FO
RenderX

Figure 4. SHAP value of the XGBoost model. SHAP: SHapley Additive exPlanation.

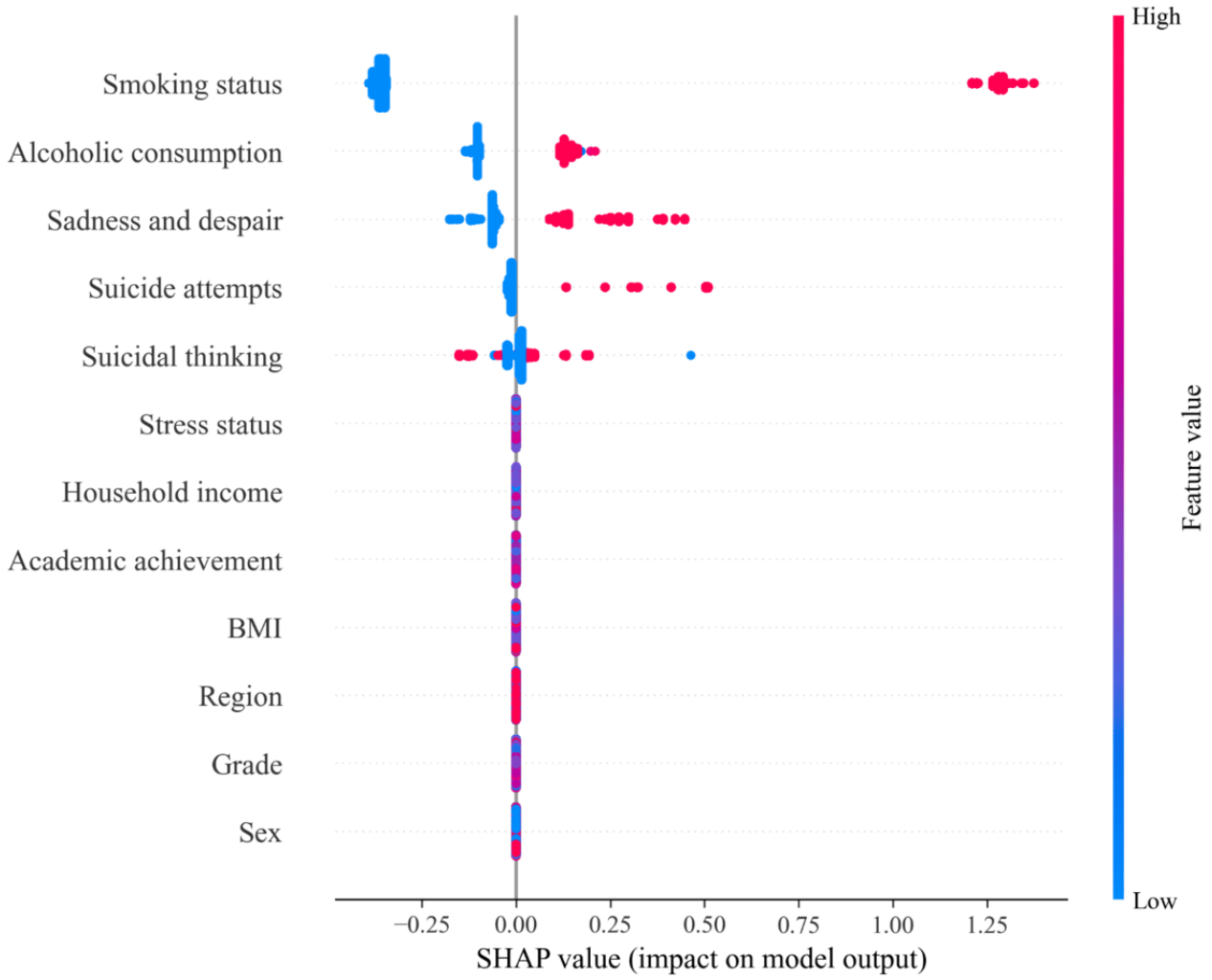

Suicidal thinking and suicide attempts also showed distinct and significant impacts, with higher values substantially increasing the likelihood of substance use. In contrast, variables such as BMI, grade, and sex had relatively minimal contributions, indicating that the model is more sensitive to psychological and behavioral factors than to demographic characteristics. These findings offer critical insights into the primary predictors of substance use and their complex interactions.

Code Availability

Based on the results of the ML model, we established a web-based application for policy implementation or health system management to support their decision-making process for cases involving substance use in adolescents [27]. An example of a web interface and the results are shown in Figure S2 in Multimedia Appendix 1. Custom code for the website is available on the web [28].

Discussion

Key Findings

This study stands out as one of the first comprehensive ML-based approaches to predict adolescent substance use on an international scale. One of our critical insights revolves around the influence of cultural diversity on substance use, drawing datasets from South Korea, the United States, and Norway [21]. Moreover, the outcome revealed that the XGBoost model proves to be commendable. It displayed predictive capabilities with an AUROC score of 80.61% (95% CI 79.63-81.59) and a precision of 30.42% (95% CI 28.65-32.16) in the discovery dataset. Our model consistently exhibited robust performance across external validation sets, achieving AUROC scores of 79.30% and 76.39% and precision of 68.37% and 12.74 in each respective dataset. Upon closer inspection, we discerned pivotal features influencing adolescent substance use predictions. Smoking status emerged as the predominant predictor for substance use, followed by BMI and suicidal thinking. Furthermore, the SHAP value analysis confirmed smoking status as a critical variable, with alcoholic consumption and sadness and despair identified as additional influential factors. To apply our findings to real-world scenarios, we devised a cutting-edge web-based platform. We believe that this tool will serve as an insightful methodology for the public to navigate potential substance-related challenges.

Plausible Mechanism

The influence of smoking status and alcohol consumption on adolescent substance use is noteworthy and warrants consideration. Neurobiological evidence indicates that nicotine, especially when introduced during formative years, can alter

XSL•FO
RenderX

the brain reward pathways, making other substances more appealing [29]. Similarly, alcohol can modify neurotransmitter levels during these formative years, making the brain more susceptible to effects from other substances [30,31]. Smoking and drinking during adolescence often signal risk-taking behaviors [32], leading to further experimentation with other substances.

Both smoking and alcoholic consumption align closely with societal expectations and peer pressures. Societal dynamics and peer interactions, which may normalize or even perceive smoking and drinking as socially acceptable, could act as influential factors in encouraging adolescents to engage in these behaviors [33,34]. In many cultures, both behaviors are viewed as rites of passage that expose adolescents to other available illicit substances [35].

Psychological factors also play a role. Many adolescents resort to smoking or drinking as coping mechanisms for stress or emotional turmoil [36,37]. These initial coping behaviors may prompt adolescents to seek stronger stimuli for more intense experiences, potentially resulting in substance addiction or misuse [38].

This study further emphasized the relationship between BMI and substance use, as another predictor of adolescent substance use. From a physiological perspective, substances can modulate metabolic rates and appetite. Adolescents engaged in substance use might experience weight changes, either due to the direct effects of the substance or inconsistent dietary habits [39]. Additionally, the role of BMI extends beyond mere physiological changes. Adolescents with “nonstandard” BMIs often face societal challenges such as weight-centric bullying and entrenched societal ideals regarding body standards. The prevailing societal standards of an ideal body may influence some adolescents toward substance use, either as a means to conform to societal expectations or to address the associated mental distress [40].

Another noteworthy observation was the understated importance of academic achievement and stress status. Although stress is traditionally considered influential in adolescent behaviors [41,42], its limited representation in this study may be attributed to data constraints. These variables were absent in our extravalidation dataset, and we resorted to imputing these values using the median from our primary training cohort. This modification might have contributed to its reduced importance in our results.

Clinical and Policy Implications

The results of this study offer significant insights into both clinical and political implications. We underscore the vital role of factors such as smoking status, BMI, and alcoholic consumption in predicting substance use among adolescents. These critical determinants enable clinicians to identify and monitor at-risk adolescents more effectively, assisting in their decision-making process [43]. Following further refinement, this model has potential commercial viability [44], especially when combined with a streamlined self-report questionnaire. The existence of multiple models assessing substance use further attests to the commercial potential of our model [45]. Emphasizing characteristics predictive of substance use is essential, suggesting the need for systems to alert parents about potential risks their children might face. Since parental intervention has proven to be effective in preventing adolescent substance use [46], establishing an early detection system becomes paramount.

Strengths and Limitations

Findings from this study must be interpreted in light of several limitations. The external validation datasets contained numerous missing values, which could have impacted the predictive accuracy of the model. Specifically, we were unable to find corresponding data for variables like BMI (due to the absence of height and weight) and academic achievement in external cohorts such as Ungdata. Moreover, stress status, suicidal thinking, and suicide attempts were also missing or unmatched in datasets like YRBS and Ungdata, leading to gaps in these key areas. To address this gap, we applied imputation methods, including median imputation based on the KYRBS dataset for entirely missing variables [47]. While this approach mitigated the issue of missing data, it may have introduced biases, highlighting the need for more harmonized and comprehensive data collection protocols in future studies to enhance model generalizability and robustness. Additionally, this study used a discovery dataset derived from adolescents in South Korea. This biased discovery dataset could unexpectedly reflect the specific racial and cultural features unique to Korean adolescents. While our model underwent external validation from a diverse cultural and demographical landscape, we also acknowledge that it may reduce sample diversity and potentially cause overfitting issues [48]. Furthermore, this study did not pinpoint a definitive causal link between the significant risk factors and adolescent substance use. In other words, it remains unclear whether substance use influences other factors or if those factors stimulate substance use. Thus, further comprehensive studies are needed to elucidate this intricate cause-and-effect relationship. Another limitation of our study is the potential variability in feature importance rankings across different ML algorithms. Different algorithms may prioritize predictors differently due to their inherent characteristics and methods of handling data. This variability suggests that the identified predictors, such as smoking status, BMI, and alcohol consumption, should not be generalized as the sole determinants of adolescent substance use. Instead, these results should be interpreted with caution, and further studies are needed to validate the findings across diverse models and datasets. Finally, this study identifies factors associated with current and past substance use rather than explicitly predicting future trends. While the model provides valuable insights into risk factors and enables early intervention strategies, its predictive performance may be limited by the imbalanced nature of the dataset and the lack of longitudinal data. Future studies should address these limitations by incorporating balanced datasets and temporal data to enhance predictive accuracy and generalizability.

Despite these limitations, this study offers significant contributions. By using extensive datasets from South Korea, the United States, and Norway, our ML model boasts enhanced prediction accuracy, highlighting its global relevance and robustness [49]. Our strategic phased validation approach,

XSL•FO
RenderX

beginning with the YRBS and progressing to the distinct Ungdata from Norway, underlines the model's versatility across diverse sociocultural backgrounds. This phased validation not only ensures consistent model evaluation but also establishes its capability in different cultural contexts [45,50,51]. Moreover, the features incorporated into the model are derived from simple questionnaires. The primary advantage of our model-based platform is its exceptional accessibility, allowing users to gather insights through straightforward surveys. This ease enables swift evaluations and widens its scope of use, equipping both clinicians and individuals with valuable insights conveniently. Our findings provide the relative importance of numerous factors. These results can guide the decision-making process by identifying key areas for the prevention of substance use among adolescents.

Conclusions

This study introduced an ML model using data from three distinct national cohorts to predict adolescent substance use. Among six unique predictive models, the XGBoost model consistently revealed a notable performance (AUROC: KYRBS, 80.61% [discovery]; YRBS, 79.30% [extravalidation]; and Ungdata, 76.39% [extravalidation], and precision: KYRBS, 30.42% [discovery]; YRBS, 68.37% [extravalidation]; and Ungdata, 12.74% [extravalidation]). Feature importance analysis identified smoking status, BMI, and suicidal thinking as significant contributors to the risk of substance use. Further insights into the influence of these variables were derived from SHAP value analysis, which identified smoking status, alcoholic consumption, and sadness and despair as the most impactful factors, in that order. The findings of this study indicate the potential of ML-driven predictive models to swiftly predict the likelihood of substance use among adolescents using a simplistic survey. It is anticipated that with further refinement and development, these models could be broadly used as efficient tools for preventing adolescent substance use.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (RS-2024-00460379), the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (RS-2024-00509257, Global AI Frontier Lab), and the Information Technology Research Center support program (IITP-2024-RS-2024-00438239) supervised by the IITP. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request. The study protocol and statistical code are available from DKY. The dataset can be accessed from the Korean Disease Control and Prevention Agency, US Centers for Disease Control and Prevention, and Norwegian Social Research Institute (NOVA) through a data use agreement.

Authors' Contributions

DKY had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. All authors approved the final version before submission. The study concept and design were contributed by Soeun Kim, HK, Seokjun Kim, SW, and DKY. Data acquisition, analysis, or interpretation were conducted by HK, SW, and DKY. The manuscript was drafted by HK, SW, and DKY. All authors critically revised the manuscript for important intellectual content. Statistical analysis was performed by HK, SW, and DKY. Study supervision was conducted by SW and DKY. DKY supervised the study and is the guarantor for this study. Soeun Kim, HK, and Seokjun Kim contributed equally as cofirst authors. DKY, SW, and RK contributed equally as cocorresponding authors. The corresponding author attests that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional material.
[DOCX File , 498 KB - jmir_v27i1e62805_app1.docx]

References

1. Davidson LL, Grigorenko EL, Boivin MJ, Rapa E, Stein A. A focus on adolescence to reduce neurological, mental health and substance-use disability. Nature 2015;527(7578):S161-S166. [doi: 10.1038/nature16030] [Medline: 26580322]
2. Volkow ND, Wargo EM. Association of severity of adolescent substance use disorders and long-term outcomes. JAMA Netw Open 2022;5(4):e225656 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.5656] [Medline: 35363272]

XSL•FO
RenderX

3. Williams EC, Fletcher OV, Frost MC, Harris AHS, Washington DL, Hoggatt KJ. Comparison of substance use disorder diagnosis rates from electronic health record data with substance use disorder prevalence rates reported in surveys across sociodemographic groups in the veterans health administration. JAMA Netw Open 2022;5(6):e2219651 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.19651] [Medline: 35771574]
4. Chaffee BW, Cheng J, Couch ET, Hoeft KS, Halpern-Felsher B. Adolescents' substance use and physical activity before and during the COVID-19 pandemic. JAMA Pediatr 2021;175(7):715-722 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.0541] [Medline: 33938922]
5. Kim J, Lee H, Lee J, Rhee SY, Shin JI, Lee SW, et al. Quantification of identifying cognitive impairment using olfactory-stimulated functional near-infrared spectroscopy with machine learning: a post hoc analysis of a diagnostic trial and validation of an external additional trial. Alzheimers Res Ther 2023;15(1):127 [FREE Full text] [doi: 10.1186/s13195-023-01268-9] [Medline: 37481573]
6. Lee H, Cho JK, Park J, Lee H, Fond G, Boyer L, et al. Machine learning-based prediction of suicidality in adolescents with allergic rhinitis: derivation and validation in 2 independent nationwide cohorts. J Med Internet Res 2024;26:e51473 [FREE Full text] [doi: 10.2196/51473] [Medline: 38354043]
7. Lam JY, Shimizu C, Tremoulet AH, Bainto E, Roberts SC, Sivilay N, et al. A machine-learning algorithm for diagnosis of multisystem inflammatory syndrome in children and Kawasaki disease in the USA: a retrospective model development and validation study. Lancet Digital Health 2022;4(10):e717-e726 [FREE Full text] [doi: 10.1016/s2589-7500(22)00149-2]
8. Saux P, Bauvin P, Raverdy V, Teigny J, Verkindt H, Soumphonphakdy T, et al. Development and validation of an interpretable machine learning-based calculator for predicting 5-year weight trajectories after bariatric surgery: a multinational retrospective cohort SOPHIA study. Lancet Digital Health 2023;5(10):e692-e702. [doi: 10.1016/s2589-7500(23)00135-8]
9. Clift AK, Dodwell D, Lord S, Petrou S, Brady M, Collins GS, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. BMJ 2023;381:e073800 [FREE Full text] [doi: 10.1136/bmj-2022-073800] [Medline: 37164379]
10. Woo HG, Park S, Yon H, Lee SW, Koyanagi A, Jacob L, et al. National trends in sadness, suicidality, and COVID-19 pandemic-related risk factors among South Korean adolescents from 2005 to 2021. JAMA Netw Open 2023;6(5):e2314838 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.14838] [Medline: 37223902]
11. Boakye E, Erhabor J, Obisesan O, Tasdighi E, Mirbolouk M, Osuji N, et al. Comprehensive review of the national surveys that assess E-cigarette use domains among youth and adults in the United States. Lancet Reg Health Am 2023;23:100528 [FREE Full text] [doi: 10.1016/j.lana.2023.100528] [Medline: 37497394]
12. Leonhardt M, Granrud MD, Bonsaksen T, Lien L. Associations between mental health, lifestyle factors and worries about climate change in Norwegian adolescents. Int J Environ Res Public Health 2022;19(19):12826 [FREE Full text] [doi: 10.3390/ijerph191912826] [Medline: 36232127]
13. Kwon R, Lee H, Kim MS, Lee J, Yon DK. Machine learning-based prediction of suicidality in adolescents during the COVID-19 pandemic (2020-2021): derivation and validation in two independent nationwide cohorts. Asian J Psychiatr 2023;88:103704. [doi: 10.1016/j.ajp.2023.103704] [Medline: 37541104]
14. Pelgrims I, Devleesschauwer B, Vandevijvere S, De Clercq EM, Vansteelandt S, Gorasso V, et al. Using random-forest multiple imputation to address bias of self-reported anthropometric measures, hypertension and hypercholesterolemia in the Belgian health interview survey. BMC Med Res Methodol 2023;23(1):69 [FREE Full text] [doi: 10.1186/s12874-023-01892-x] [Medline: 36966305]
15. Johnston KC, Connors AF, Wagner DP, Haley EC. Predicting outcome in ischemic stroke: external validation of predictive risk models. Stroke 2003 Jan;34(1):200-202 [FREE Full text] [doi: 10.1161/01.str.0000047102.61863.e3] [Medline: 12511774]
16. Nguipdop-Djomo P, Rodrigues LC, Smith PG, Abubakar I, Mangtani P. Drug misuse, tobacco smoking, alcohol and other social determinants of tuberculosis in UK-born adults in England: a community-based case-control study. Sci Rep 2020;10(1):5639 [FREE Full text] [doi: 10.1038/s41598-020-62667-8] [Medline: 32221405]
17. Compton WM, Flannagan KSJ, Silveira ML, Creamer MR, Kimmel HL, Kanel M, et al. Tobacco, alcohol, cannabis, and other drug use in the US before and during the early phase of the COVID-19 pandemic. JAMA Netw Open 2023;6(1):e2254566 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.54566] [Medline: 36719678]
18. Huang J. Drug licensing as evidence of evolution, diffusion and catch-up in East Asia. Nat Biotechnol 2023;41(2):189-192. [doi: 10.1038/s41587-023-01659-1] [Medline: 36792699]
19. Garriga R, Mas J, Abraha S, Nolan J, Harrison O, Tadros G, et al. Machine learning model to predict mental health crises from electronic health records. Nat Med 2022;28(6):1240-1248 [FREE Full text] [doi: 10.1038/s41591-022-01811-5] [Medline: 35577964]
20. Montgomery-Csobán T, Kavanagh K, Murray P, Robertson C, Barry SJE, Ukah UV, et al. Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model): a modelling study. Lancet Digital Health 2024;6(4):e238-e250. [doi: 10.1016/s2589-7500(23)00267-4]
21. Kim H, Son Y, Lee H, Kang J, Hammoodi A, Choi Y, et al. Machine learning-based prediction of suicidal thinking in adolescents by derivation and validation in 3 independent worldwide cohorts: algorithm development and validation study. J Med Internet Res 2024;26:e55913 [FREE Full text] [doi: 10.2196/55913] [Medline: 38758578]

XSL•FO
RenderX

22. Bolo K, Aroca GAA, Pardeshi AA, Chiang M, Burkemper B, Xie X, et al. Automated expert-level scleral spur detection and quantitative biometric analysis on the ANTERION anterior segment OCT system. Br J Ophthalmol 2024;108(5):702-709 [FREE Full text] [doi: 10.1136/bjo-2022-322328] [Medline: 37798075]
23. Hwang SH, Lee H, Lee JH, Lee M, Koyanagi A, Smith L, et al. Machine learning-based prediction for incident hypertension based on regular health checkup data: derivation and validation in 2 independent nationwide cohorts in South Korea and Japan. J Med Internet Res 2024;26:e52794 [FREE Full text] [doi: 10.2196/52794] [Medline: 39499554]
24. Sang H, Lee H, Lee M, Park J, Kim S, Woo HG, et al. Prediction model for cardiovascular disease in patients with diabetes using machine learning derived and validated in two independent Korean cohorts. Sci Rep 2024;14(1):14966 [FREE Full text] [doi: 10.1038/s41598-024-63798-y] [Medline: 38942775]
25. Thorsen-Meyer H, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digital Health 2020;2(4):e179-e191. [doi: 10.1016/s2589-7500(20)30018-2]
26. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594 [FREE Full text] [doi: 10.1136/bmj.g7594] [Medline: 25569120]
27. Probability for substance use. Streamlit. URL: https://predictsubstance.streamlit.app/ [accessed 2025-01-28]
28. CenterForDH/substance. GitHub. URL: https://github.com/centerfordh/predict_substance/ [accessed 2025-01-28]
29. Le Foll B, Piper ME, Fowler CD, Tonstad S, Bierut L, Lu L, et al. Tobacco and nicotine use. Nat Rev Dis Primers 2022;8(1):19. [doi: 10.1038/s41572-022-00346-w] [Medline: 35332148]
30. Yip SW, Lichenstein SD, Liang Q, Chaarani B, Dager A, Pearlson G, et al. Brain networks and adolescent alcohol use. JAMA Psychiatry 2023;80(11):1131-1141. [doi: 10.1001/jamapsychiatry.2023.2949] [Medline: 37647053]
31. Kang J, Kim HJ, Kim T, Lee H, Kim M, Lee SW, et al. Prenatal opioid exposure and subsequent risk of neuropsychiatric disorders in children: nationwide birth cohort study in South Korea. BMJ 2024;385:e077664 [FREE Full text] [doi: 10.1136/bmj-2023-077664] [Medline: 38658035]
32. Kruckow S, Santini ZI, Hjarnaa L, Becker U, Andersen O, Tolstrup JS. Associations between alcohol intake and hospital contacts due to alcohol and unintentional injuries in 71,025 Danish adolescents—a prospective cohort study. EClinicalMedicine 2023;64:102187 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102187] [Medline: 37936661]
33. Zhao Y, Di X, Li S, Zeng X, Wang X, Nan Y, et al. Prevalence, frequency, intensity, and location of cigarette use among adolescents in China from 2013-14 to 2019: findings from two repeated cross-sectional studies. Lancet Reg Health West Pac 2022;27:100549 [FREE Full text] [doi: 10.1016/j.lanwpc.2022.100549] [Medline: 35923777]
34. Kim S, Lee H, Lee J, Lee SW, Kwon R, Kim MS, et al. Short- and long-term neuropsychiatric outcomes in long COVID in South Korea and Japan. Nat Hum Behav 2024;8(8):1530-1544. [doi: 10.1038/s41562-024-01895-8] [Medline: 38918517]
35. Maeng SJ, Lee DJ, Kang JH. First drinking experiences during adolescence in South Korea: a qualitative study focusing on the internal and external factors. Int J Environ Res Public Health 2021;18(15):8200 [FREE Full text] [doi: 10.3390/ijerph18158200] [Medline: 34360493]
36. Meienberg A, Mayr M, Vischer A, Zellweger MJ, Burkard T. Smoking prevention in adolescents: a cross-sectional and qualitative evaluation of a newly implemented prevention program in Switzerland. BMJ Open 2021;11(12):e048319. [doi: 10.1136/bmjopen-2020-048319]
37. Skylstad V, Babirye JN, Kiguli J, Skar AMS, Kühl MJ, Nalugya JS, et al. Are we overlooking alcohol use by younger children? BMJ Paediatr Open 2022;6(1):e001242 [FREE Full text] [doi: 10.1136/bmjpo-2021-001242] [Medline: 36053657]
38. Cheron J, d'Exaerde ADK. Drug addiction: from bench to bedside. Transl Psychiatry 2021;11(1):424 [FREE Full text] [doi: 10.1038/s41398-021-01542-0] [Medline: 34385417]
39. Treasure J, Duarte TA, Schmidt U. Eating disorders. Lancet 2020;395(10227):899-911 [FREE Full text] [doi: 10.1016/s0140-6736(20)30059-3]
40. Bornioli A, Lewis-Smith H, Slater A, Bray I. Body dissatisfaction predicts the onset of depression among adolescent females and males: a prospective study. J Epidemiol Community Health 2020;75(4):343-348. [doi: 10.1136/jech-2019-213033] [Medline: 33288655]
41. Sylvestre MP, Dinkou GDT, Naja M, Riglea T, Pelekanakis A, Bélanger M, et al. A longitudinal study of change in substance use from before to during the COVID-19 pandemic in young adults. Lancet Reg Health Am 2022;8:100168 [FREE Full text] [doi: 10.1016/j.lana.2021.100168] [Medline: 35469267]
42. Liu M, Koh KA, Hwang SW, Wadhera RK. Mental health and substance use among homeless adolescents in the US. JAMA 2022;327(18):1820-1822 [FREE Full text] [doi: 10.1001/jama.2022.4422] [Medline: 35536272]
43. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. BMJ 2019;364:l886 [FREE Full text] [doi: 10.1136/bmj.l886] [Medline: 30862612]
44. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. Lancet Digital Health 2021;3(3):e195-e203 [FREE Full text] [doi: 10.1016/s2589-7500(20)30292-2]

XSL•FO
RenderX

45. Lo-Ciganic WH, Donohue JM, Yang Q, Huang JL, Chang CY, Weiss JC, et al. Developing and validating a machine-learning algorithm to predict opioid overdose in medicaid beneficiaries in two US states: a prognostic modelling study. Lancet Digital Health 2022;4(6):e455-e465 [FREE Full text] [doi: 10.1016/s2589-7500(22)00062-0]
46. Kuntsche S, Kuntsche E. Parent-based interventions for preventing or reducing adolescent substance use—a systematic literature review. Clin Psychol Rev 2016;45:89-101. [doi: 10.1016/j.cpr.2016.02.004] [Medline: 27111301]
47. Pokorney SD, Holmes DN, Thomas L, Fonarow GC, Kowey PR, Reiffel JA, et al. Association between warfarin control metrics and atrial fibrillation outcomes in the outcomes registry for better informed treatment of atrial fibrillation. JAMA Cardiol 2019;4(8):756-764 [FREE Full text] [doi: 10.1001/jamacardio.2019.1960] [Medline: 31268487]
48. Blackburn AM, Vestergren S, COVIDiSTRESS II Consortium. Author Correction: COVIDiSTRESS diverse dataset on psychological and behavioural outcomes one year into the COVID-19 pandemic. Sci Data 2023;10(1):12 [FREE Full text] [doi: 10.1038/s41597-022-01896-0] [Medline: 36599849]
49. Riley RD, Snell KIE, Altman DG, Collins GS. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2019 Jun 25;365:l4379 [FREE Full text] [doi: 10.1136/bmj.l4379] [Medline: 31239248]
50. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. JAMA 2020;324(12):1212-1213 [FREE Full text] [doi: 10.1001/jama.2020.12067] [Medline: 32960230]
51. Jo H, Park J, Lee H, Lee K, Lee H, Son Y, et al. Nationwide trends in sadness, suicidal ideation, and suicide attempts among multicultural and monocultural adolescents in South Korea during the COVID-19 pandemic, 2011-2022. World J Pediatr 2024;20(12):1249-1269. [doi: 10.1007/s12519-024-00858-3] [Medline: 39614994]

Abbreviations

AUPRC: area under precision-recall curve
AUROC: area under receiver operating characteristic curve
CDC: Centers for Disease Control and Prevention
KDCA: Korean Disease Control and Prevention Agency
KYRBS: Korea Youth Risk Behavior Web-Based Survey
ML: machine learning
NOVA: Norwegian Social Research Institute
SHAP: SHapley Additive exPlanation
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
YRBS: Youth Risk Behavior Survey

Edited by F Xie; submitted 31.05.24; peer-reviewed by H Lin, KS Lee, K Kim; comments to author 19.08.24; revised version received 27.12.24; accepted 16.01.25; published 24.02.25.

Please cite as:
Kim S, Kim H, Kim S, Lee H, Hammoodi A, Choi Y, Kim HJ, Smith L, Kim MS, Fond G, Boyer L, Baik SW, Lee H, Park J, Kwon R, Woo S, Yon DK
Machine Learning–Based Prediction of Substance Use in Adolescents in Three Independent Worldwide Cohorts: Algorithm Development and Validation Study
J Med Internet Res 2025;27:e62805
URL: https://www.jmir.org/2025/1/e62805
doi:10.2196/62805
PMID:

©Soeun Kim, Hyejun Kim, Seokjun Kim, Hojae Lee, Ahmed Hammoodi, Yujin Choi, Hyeon Jin Kim, Lee Smith, Min Seo Kim, Guillaume Fond, Laurent Boyer, Sung Wook Baik, Hayeon Lee, Jaeyu Park, Rosie Kwon, Selin Woo, Dong Keon Yon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Patient Simulations Using Social Robotics Combined With Large Language Models for Clinical Reasoning Training in Medical Education: Mixed Methods Study

Alexander Borg[1], MD; Carina Georg[2], PhD; Benjamin Jobs[1]; Viking Huss[3], MD; Kristin Waldenlind[3], MD, PhD; Mini Ruiz[4], MD, PhD; Samuel Edelbring[5,6], PhD; Gabriel Skantze[7], PhD; Ioannis Parodis[1,8], MD, PhD

[1]Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, and Center for Molecular Medicine (CMM), Stockholm, Sweden

[2]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[3]Division of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

[4]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

[5]School of Health Sciences, Örebro University, Örebro, Sweden

[6]School of Education, Culture and Communication, Mälardalen University, Västerås, Sweden

[7]Division of Speech Music and Hearing, Royal Institute of Technology, Stockholm, Sweden

[8]Department of Rheumatology, Faculty of Medicine and Health, Örebro University and Örebro University Hospital, Örebro, Sweden

Corresponding Author:
Ioannis Parodis, MD, PhD
Division of Rheumatology
Department of Medicine Solna
Karolinska Institutet, Karolinska University Hospital, and Center for Molecular Medicine (CMM)
D2:01 Rheumatology Karolinska University Hospital Solna
Stockholm, SE-171 76
Sweden
Phone: 46 722321322
Email: ioannis.parodis@ki.se

Abstract

Background: Virtual patients (VPs) are computer-based simulations of clinical scenarios used in health professions education to address various learning outcomes, including clinical reasoning (CR). CR is a crucial skill for health care practitioners, and its inadequacy can compromise patient safety. Recent advancements in large language models (LLMs) and social robots have introduced new possibilities for enhancing VP interactivity and realism. However, their application in VP simulations has been limited, and no studies have investigated the effectiveness of combining LLMs with social robots for CR training.

Objective: The aim of the study is to explore the potential added value of a social robotic VP platform combined with an LLM compared to a conventional computer-based VP modality for CR training of medical students.

Methods: A Swedish explorative proof-of-concept study was conducted between May and July 2023, combining quantitative and qualitative methodology. In total, 15 medical students from Karolinska Institutet and an international exchange program completed a VP case in a social robotic platform and a computer-based semilinear platform. Students' self-perceived VP experience focusing on CR training was assessed using a previously developed index, and paired 2-tailed *t* test was used to compare mean scores (scales from 1 to 5) between the platforms. Moreover, in-depth interviews were conducted with 8 medical students.

Results: The social robotic platform was perceived as more authentic (mean 4.5, SD 0.7 vs mean 3.9, SD 0.5; odds ratio [OR] 2.9, 95% CI 0.0-1.0; *P*=.04) and provided a beneficial overall learning effect (mean 4.4, SD 0.6 versus mean 4.1, SD 0.6; OR 3.7, 95% CI 0.1-0.5; *P*=.01) compared with the computer-based platform. Qualitative analysis revealed 4 themes, wherein students experienced the social robot as superior to the computer-based platform in training CR, communication, and emotional skills. Limitations related to technical and user-related aspects were identified, and suggestions for improvements included enhanced facial expressions and VP cases simulating multiple personalities.

Conclusions: A social robotic platform enhanced by an LLM may provide an authentic and engaging learning experience for medical students in the context of VP simulations for training CR. Beyond its limitations, several aspects of potential improvement

XSL•FO
RenderX

were identified for the social robotic platform, lending promise for this technology as a means toward the attainment of learning outcomes within medical education curricula.

(J Med Internet Res 2025;27:e63312) doi:10.2196/63312

KEYWORDS

virtual patients; clinical reasoning; large language models; social robotics; medical education; sustainable learning; medical students

Introduction

Virtual patients (VPs) are computed interactive simulations of clinical scenarios for educational or learning purposes [1]. These tools are used to address several learning outcomes in health professions education (HPE) curricula [2]. VP platforms can be designed in a linear fashion, where medical information and treatment options are gathered in a fixed order without consequences, or with a branched design, where interactions also impact the outcome of the case [3]. VPs have been proposed as an effective educational tool for practicing clinical reasoning (CR) in undergraduate medical education [4] while also ensuring that medical students encounter representative patient cases, which might not always be possible to encounter during clinical rotations [5-7]. CR is an essential skill for health care practitioners, and its inadequacy may threaten patient safety, making it a key goal in HPE [8-10]. An essential aspect of VP simulation is the interface, which allows the user to interact with the VP in multiple ways. A common feature of VP platforms is that they can facilitate the obtainment of medical history, the conduct of physical examinations, and referral for laboratory tests and the provision of their findings to culminate in a reasonable diagnosis and management plan [11].

Previous research shows that medical students perceive VP simulation as a satisfactory educational activity for training CR while they acknowledge the lack of complexity and level of interactivity compared with authentic patient encounters [12]. Interestingly, students have been shown to perform better at examinations after exposure to VP simulations compared to other educational activities with similar learning outcomes, but the positive effect of VPs on CR training is still unclear [2,13].

Recent technological developments have enabled the integration of dialogue systems with natural language processing capabilities, embodied conversational agents [14], and social robots [15] to enhance the interactivity and realism of VP interactions. The emergence of large language models (LLMs) and their application in dialogue systems and social robots have introduced novel possibilities within this field [16,17]. However, their use for VPs has been limited to preliminary endeavors [18-21]. To the best of our knowledge, no prior studies have used LLMs in conjunction with a social robot for VP simulation and evaluated their effectiveness in facilitating CR training.

We hypothesized that the integration of an LLM in a social robotic modality for the presentation of VPs would involve several advantages. More specifically, we hypothesized that the physical presence of a social robot, combined with its nonverbal communication capabilities, would enhance engagement and immersion in the learning experience. This integration of visual and social cues, along with advanced verbal interactions, could facilitate a more comprehensive information-gathering process during the VP encounter until a diagnosis has been made and a management plan has been proposed compared with using only LLMs for VP presentation or conventional computer-based platforms.

While VPs can be valuable across many medical disciplines, this study focuses on their application in rheumatology education. Rheumatic diseases affect a substantial proportion of the population and are considered a significant public health concern [22-24]. These conditions often present complex diagnostic challenges that make them particularly suitable for CR training through VP simulations. To ensure that patients with rheumatic diseases receive appropriate care, it is crucial to provide evidence-based training in CR for medical students [25]. The aim of this study was to explore the potential added value of and potential limitations with a social robot combined with an LLM in comparison to a conventional, computer-based, semilinear VP platform presenting rheumatic conditions for medical students' CR training.

Methods

An interventional, explorative, proof-of-concept study was conducted between May and July 2023. We combined quantitative and qualitative methodology to assess medical students' self-perceived CR training using a previously developed index for VP design evaluation [26] as well as in-depth interviews.

Participants

In total, 15 medical students from Karolinska Institutet (KI) were recruited in the study either from a 5-week elective course for third-year students or a 4-week exchange program for international students, both pertaining to inflammation within internal medicine, including rheumatology, gastroenterology, dermatology, nephrology, and infectious diseases. All students had completed the sixth term of their medical education. More specifically, prior to the course, students from KI had completed 2 full years of preclinical medical education (eg, anatomy, physiology, and pathophysiology) and 1 year of clinical rotations in internal medicine. International students had also completed the preclinical phase of their medical education and had varying clinical backgrounds, with at least 1 full year of clinical rotations in internal medicine. While all students had prior experience with patient encounters and were fluent in English, none were native English speakers, and their previous experience of communicating with patients in English varied. In total, 8 students also agreed to participate in in-depth interviews, scrutinizing their experience with the social robotic platform

XSL•FO
RenderX

compared with the conventional computer-based VP platform with regard to CR training.

Ethical Considerations

The study protocol was reviewed and approved by the Swedish Ethical Review Authority (registration 2022-04437-01). Written informed consent was obtained from all study participants prior to enrollment. Participants had the ability to opt out at any time. All data were deidentified prior to analysis. No compensation was provided for participation in this study.

Development and Implementation of VP Cases

A VP case illustrating a common rheumatic disease (ie, polymyalgia rheumatica) was developed for presentation in a social robotic platform from Furhat robotics [27] and a conventional semilinear computer-based platform (ie, the virtual interactive case simulator [VIC]) [28]. The case development in both platforms adhered to distinct principles for VP case development proposed by Posel et al [29].

First, we identified the content of the VP case and selected the design. Between a linear and an explorative design, we anticipated advantages in using an explorative design, as it more closely resembles real-life clinical situations. We then organized the VP content by clarifying the purpose of each step within the VP cases. Next, we tailored the delivery of the VP case by adjusting the level of difficulty to align with the medical students' prior knowledge of rheumatic diseases. The VP cases were continuously evaluated through feedback from students and peers before, during, and after development. Integrity and confidentiality were maintained by creating typical case scenarios without using any information from real patients.

The students were tasked to perform the VP case individually in both platforms to facilitate paired comparisons. The case was considered within a particular situational context to be a primary care encounter. The students were assigned the role of an attending physician. The case was completed when the students had obtained sufficient information to make a diagnosis and suggest a management plan for the patient.

After interaction with each VP platform, students immediately filled out a questionnaire evaluating the VP design with an emphasis on CR training and provided optional free-text evaluation within seven themes: (1) authenticity of patient encounter, (2) professional approach in the consultation, (3) coaching during consultation, (4) learning effect of consultation, (5) overall judgment of case work-up, (6) special strengths of the case, and (7) special weaknesses of the case [26]. The questionnaire consisted of 14 questions, of which the first 12 questions (corresponding to themes 1 to 5) were graded using a 5-point scale, where 1=strong disagreement and 5=strong agreement, whereas responses to the last 2 questions (corresponding to themes 6 and 7) were only provided in the form of free text. This questionnaire was specifically developed for evaluating the design of VP platforms with an emphasis on CR training. Its development combined a theoretical framework for CR training along with the involvement of an international team of VP experts, supporting its content validity [26]. The questionnaire is provided in Figure S1 in Multimedia Appendix 1.

Social Robotic Platform

The social robot serving as the embodiment for the novel VP platform features an animated face back-projected onto a translucent mask and a neck with 3 degrees of freedom, allowing for natural head movements. The social robot consists of a built-in speaker to convey dialogue and a microphone to perceive spoken phrases that are later interpreted using a speech-to-text synthesis. Moreover, the social robot uses computer vision to track multiple users in real time, performing facial expression analysis, head pose estimation, and user distance measurement. It includes a face detection system, multiuser tracking, depth estimation, spatial modeling, and a deep neural network–based face recognition for tracking users over time [27]. The animated face expresses subtle emotions through facial expressions and gaze behavior, crucial for conveying a patient's internal state [30].

The platform combines the robot-steering software development kit FurhatSDK [27] and the LLM OpenAI's GPT-3.5-turbo chat completion model [31]. To maintain authenticity and avoid responses sounding like an assistant, we used a single user message containing the required information to generate a VP response. This prompt included a patient case description, the last 10 dialogue turns, and instructions to generate the next dialogue line. A shortened example of a prompt is provided in Figure S2 in Multimedia Appendix 1.

We also used the LLM to generate facial expressions, essential for reflecting the patient's emotional states during interactions, similar to the methods described by Irfan et al [16]. At the beginning of each utterance, anchor points were inserted at phrase boundaries within the text. The LLM was next prompted to incorporate suitable facial expressions at each anchor point, selecting from the FurhatSDK's available expressions, for example, sad, happy, or surprised facial expressions, illustrating a variety of nonverbal expressions adapted to the context of the conversation.

The LLM calls for generating dialogue responses and facial expressions were separate. First, the dialogue response was generated, and the response was translated using a text-to-speech synthesis. While the response was being synthesized, a second call was made to the LLM to generate the list of facial expressions related to each anchor point. While the robot was speaking, the system kept track of when each word had been spoken, so that the relevant facial expression could be projected.

Using LLMs can cause response delays, complicating turn-taking, as users might not realize that the robot is generating a reply [16]. This issue is worsened when there is no provision for barge-in. To address this, various approaches can be adopted, such as the use of a turn-holding cue to indicate that the robot is actively "thinking" about a response [32]. We adopted a gaze aversion technique [33] combined with LED lights at the base of the robot to communicate its turn-taking status, indicating whether it is speaking or listening.

Virtual Interactive Case Simulator

In VIC, a semilinear approach is used, ensuring a consistent beginning and end for each case while allowing users to explore the case environment in their preferred order [3]. Every case

XSL•FO
RenderX

starts with a brief patient summary and finishes with a diagnostic verdict and management approach, presented in a multiple-choice structure. This interface allows the user to investigate, for example, the patient's medical history, administer a physical examination, and retrieve laboratory test results. Interaction with this platform is facilitated entirely using prewritten clickable options, with no reliance on natural language processing methodologies.

Interviews

We carried out one-on-one, semistructured interviews with medical students following the completion of the VP case sessions using both platforms. This was performed on the same day, immediately after having completed the VP cases in both platforms. We developed an interview guide that consisted of questions concerning the participants' experiences with the VP platforms as tools for learning and CR training as well as their views on potential applications and integration of these platforms within medical curricula. The interview guide is provided in Figure S3 in Multimedia Appendix 1.

Interviews were held at Karolinska University Hospital during clinical placements as described earlier and lasted from 20 to 30 minutes each. The interview data were stored in dedicated spaces at the Department of Medicine Solna, KI. The recorded audio files were transcribed verbatim, and the transcripts were pseudonymized, ensuring the removal of any sensitive or personal identifiable information.

Analysis

To compare responses to the quantitative part of the questionnaire (themes 1 to 5) relating to the robotic platform versus VIC, the paired 2-tailed *t* test was used to compare mean scores between the platforms. This analysis included both separate questions and summary scores within themes. Results from paired 2-tailed *t* tests are presented as the mean score (SD), 95% CI, odds ratio (OR), and *P* value. The statistical analysis was performed using the R statistical software (version 4.2.1; R Foundation for Statistical Computing). Differences generating *P* values <.05 were considered statistically significant.

Interview data were analyzed systematically for the identification of key themes. The thematic analysis was carried out following the approach described by Braun and Clarke [34]. First, the verbatim transcripts of the interviews were reviewed multiple times to ensure familiarity with the content. The interviews were analyzed individually as a first step, and then horizontally, that is, across interviews. International recommendations for integrating CR training into HPE curricula [10] were used as a guideline when coding words and phrases related to CR training. This coding process was iterative, with the researchers collaborating to identify, interpret, and refine the codes until the final themes emerged. The initial stages of the analysis and theme development were undertaken by 2 investigators (AB and BJ), while a third investigator (IP) supervised the process and contributed to the final selection of the themes.

All authors contributed to the interpretation of the results, bringing diverse perspectives to the analysis. AB was at the time of writing a medical intern and PhD student within medical pedagogy, CG was a nurse specialized in intensive care, BJ was a medical student during the fourth year of the KI medical program, VH was a resident physician in rheumatology, KW was a consultant rheumatologist, MR was a consultant endocrinologist, SE was a senior lecturer and associate professor in medical pedagogy, GS was a professor in speech technology, one of the founders of the social robot used in the project, and chief scientist at Furhat Robotics. IP was a senior consultant rheumatologist, associate professor of rheumatology, and educator within the KI medical program. This multidisciplinary team contributed complemental approaches and views during different stages of the analysis, substantially enhancing the overall interpretation of the data.

Results

Questionnaire

In total, 6 (40%) of the 15 medical students were women, and 9 (60%) were men. A total of 11 (73%) students declared no previous experience with VP platforms. The mean age of the students was 23.9 (SD 4.8) years. Response frequencies to all quantitative themes (1 to 5) are provided in Tables S1-S5 in Multimedia Appendix 2.

The students perceived the social robotic platform as more authentic compared with VIC with regard to decision-making (mean 4.4, SD 0.6 vs mean 3.9, SD 0.9; OR 3.2, 95% CI 0.1-1.2; *P*=.03) as well as beneficial regarding the preparation of taking care of a real-life patient with the same complaint (mean 4.4, SD 0.6 vs mean 4.0, SD 0.7; OR 4.2, 95% CI 0.1-0.7; *P*=.009). We documented a numerical difference in favor of the social robotic platform with regard to the appropriateness of the VP case's level of difficulty (mean 4.7, SD 0.5 vs mean 4.3, SD 1.0; OR 2.4, 95% CI –0.1 to 0.9; *P*=.08); however, this difference did not reach statistical significance. Detailed results are shown in Table 1 and Figure 1.

XSL•FO
RenderX

Table 1. Comparison of mean scores using paired 2-tailed *t* test within each question of the clinical reasoning questionnaire (N=15).

Theme and variable	Social robot, mean (SD)	VIC[a], mean (SD)	OR[b] (95% CI)	*P* value
Authenticity of patient encounter				
Question 1	4.6 (0.6)	3.9 (0.9)	3.2 (0.1 to 1.2)	*.03*[c]
Question 2	4.3 (0.9)	3.9 (0.5)	1.9 (–0.2 to 1.0)	.19
Professional approach in the consultation				
Question 1	4.8 (0.4)	4.5 (0.7)	2.1 (–0.1 to 0.8)	.14
Question 2	4.3 (0.6)	4.1 (0.5)	1.9 (–0.1 to 0.5)	.19
Question 3	4.1 (0.9)	3.7 (1.2)	2.0 (–0.2 to 1.0)	.16
Question 4	4.1 (1.0)	4.4 (0.6)	0.5 (–0.8 to 0.2)	.17
Coaching during consultation				
Question 1	4.7 (0.5)	4.3 (1.0)	2.4 (–0.1 to 0.9)	.08
Question 2	4.5 (0.5)[d]	4.4 (0.7)[d]	1.4 (–0.3 to 0.6)	.50
Question 3	4.6 (0.5)[d]	4.3 (0.9)[d]	1.8 (–0.3 to 0.9)	.26
Learning effect of consultation				
Question 1	4.4 (0.6)	4.2 (0.7)	1.9 (–0.1 to 0.5)	.19
Question 2	4.4 (0.6)	4.0 (0.7)	4.2 (0.1 to 0.7)	*.009*
Overall judgment of case work-up				
Question 1	4.7 (0.8)	4.7 (0.5)	1.2 (–0.4 to 0.5)	.75

[a]VIC: virtual interactive case simulator.

[b]OR: odds ratio.

[c]Statistically significant *P* values are in italics format.

[d]n=13.

XSL•FO
RenderX

Figure 1. Violin plots illustrating distributions of scores from students' responses to questions within quantitative themes (1 to 5) with regard to the social robotic platform versus VIC. Lines within violin plots denote median values and IQRs. Numbers at the bottom of the plots refer to mean values. Statistically significant *P* values are in italics format. Q: question; VIC: virtual interactive case simulator.

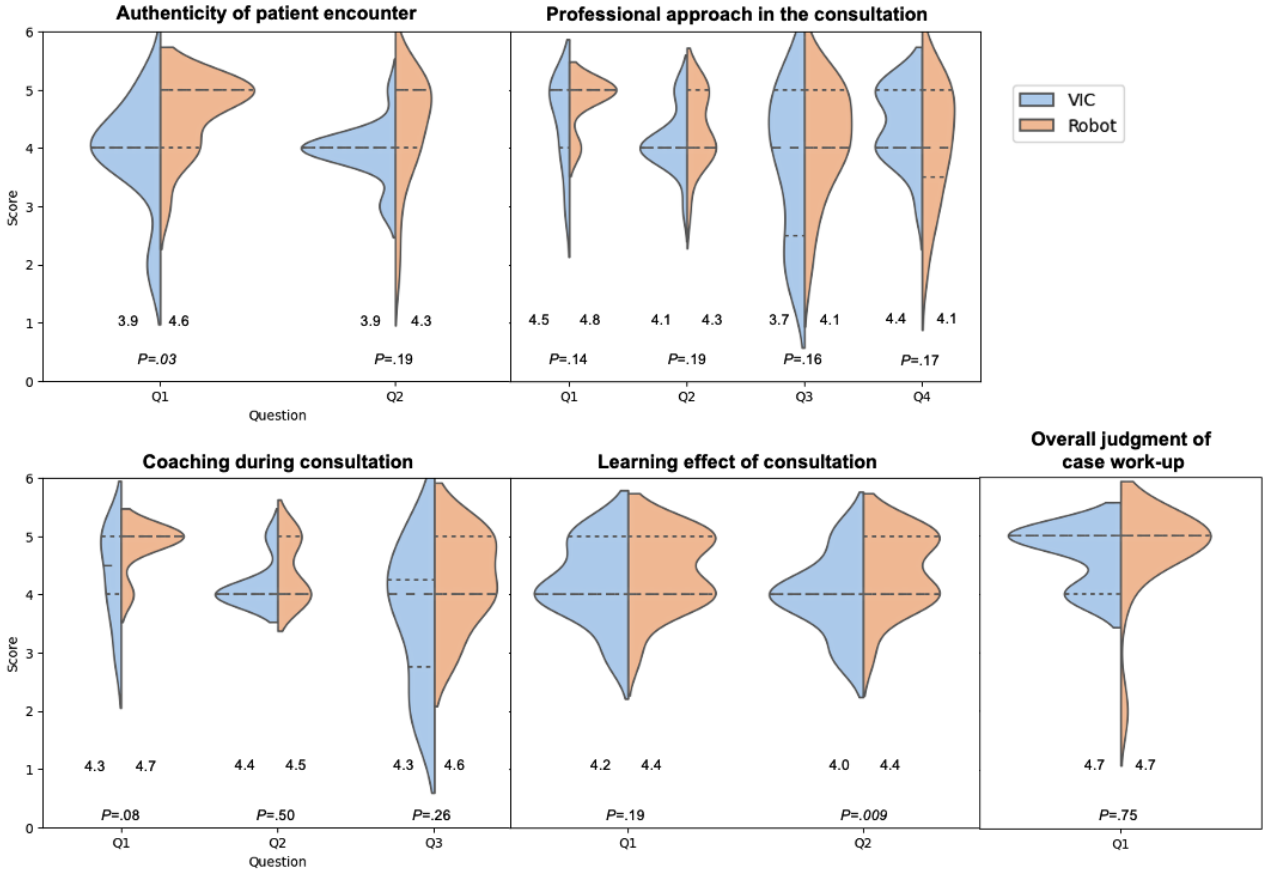

Summary scores within themes demonstrated superiority of the social robotic platform compared with VIC both regarding authenticity (mean 4.5, SD 0.7 vs mean 3.9, SD 0.5; OR 2.9, 95% CI 0.0-1.0; *P*=.04) and learning effect of the consultation by feeling more prepared to care for a real-life patient with a similar complaint (mean 4.4, SD 0.6 vs mean 4.1, SD 0.6; OR 3.7, 95% CI 0.1-0.5; *P*=.01), as illustrated in Figure 2 and Table 2.

Open-ended themes (6 and 7) highlighted the special strengths and limitations with both platforms. Students perceived the social robotic platform as a good way to simulating realistic patient conversations and providing an interactive and engaging experience. Limitations of the social robot included the lack of laboratory results and the lack of modality for training physical examination, as well as that conversations sometimes felt mechanical due to their turn-based structure and due to the fact that students were interrupted by the robot if their verbal pause was too long since this pause was perceived as the end of turn.

By contrast, students perceived the VPs presented through VIC as a good tool for structured learning and revision of the case. VIC provided useful feedback and information regarding various clinical investigations. However, the fixed question format in VIC was a limiting factor, as was the lack of the possibility to ask the VP follow-up questions. Some students noted that VIC did not provide a feeling that resembled a real patient encounter to the same extent as the social robot. Detailed responses are shown in Tables S6 and S7 in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 2. Violin plots illustrating distributions of summary theme scores of quantitative themes (1 to 5) in the social robotic platform versus VIC. Lines within violin plots denote median values and IQRs. Numbers at the bottom of the plots denote mean values. Statistically significant *P* values are in italics format. VIC: virtual interactive case simulator.

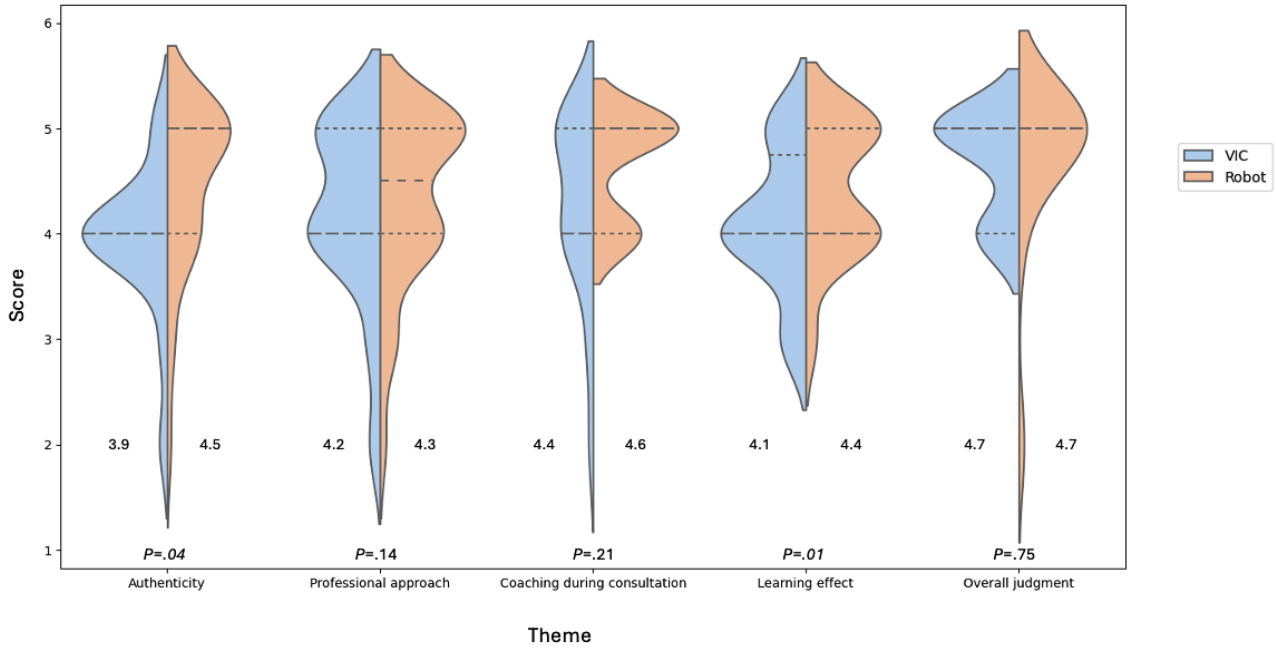

Table 2. Comparisons of summary theme scores from the clinical reasoning questionnaire between the social robotic platform and VIC[a] (N=15).

Theme	Social robot, mean (SD)	VIC, mean (SD)	OR[b] (95% CI)	*P* value
Authenticity of patient encounter	4.5 (0.7)	3.9 (0.5)	2.9 (0.0 to 1.0)	*.04*
Professional approach in the consultation	4.3 (0.5)	4.2 (0.5)	2.0 (–0.1 to 0.4)	.16
Coaching during consultation	4.6 (0.4)[c]	4.4 (0.6)[c]	1.5 (–0.2 to 0.7)	.21
Learning effect of consultation	4.4 (0.6)	4.1 (0.6)	3.7 (0.1 to 0.5)	*.01*[d]
Overall judgment of case work-up	4.7 (0.8)	4.7 (0.5)	1.2 (–0.4 to 0.5)	.75

[a]VIC: virtual interactive case simulator.

[b]OR: odds ratio.

[c]n=12.

[d]Statistically significant *P* values are in italics format.

Interviews

Overview

A total of 8 medical students participated in in-depth interviews. In total, 5 (63%) students were women, and 3 (37%) students were men. The students' mean age was 21.5 (SD 1.9) years. Thematic analysis resulted in four themes: (1) realism and immersion, (2) skill acquisition and development, (3) procedural limitations, and (4) potential for improvement (Table S8 in Multimedia Appendix 2).

Realism and Immersion

The theme "realism and immersion" was divided into two subthemes: (1) authenticity and (2) interactivity and engagement. Students perceived the experience with the LLM-empowered robotic platform as a realistic and authentic patient encounter, outperforming the computer-based platform in these aspects. Within the subtheme "authenticity," students described that the realistic impression conferred from the social robotic platform made them feel like being the treating physician in a real-life scenario, with responsibilities toward the presented patient. Relevant quotes are presented in Textbox 1.

XSL•FO
RenderX

Textbox 1. Relevant quotes by theme "realism and immersion" and subthemes from in-depth interviews.

Subtheme: authenticity

- "It felt close, like if I were with a real patient" [Male student, 22 years].
- "[...] you are asking the questions and in real time there is a response, and it is like a real conversation. [...] I definitely felt like the treating physician with the robot, knowing that it is up to me to ask the questions and to take in what the patient is saying" [Female student, 22 years].
- "Because the patient was expressing how he was in pain, and how, you know, the pain was bothering him. [...] So yeah, it felt quite intimate at that time. And it felt like I was a physician taking care of a patient" [Male student, 25 years].

Subtheme: interactivity and engagement

- "The robot is more interactive compared to the computer. The computer [...] is of course very helpful. But with the robot, it is more fun because you see this robot talking to you and answering your questions in real time" [Female student, 20 years].
- "[...] the computer platform is more passive. You don't have to be so productive" [Male student, 25 years].
- "I think the computer-clicking, the computer platform, is less demanding. I would say it is on a lower level of interactivity than the robot" [Male student, 22 years].

The social robotic platform facilitated a highly interactive and engaging experience, which was described to be superior to VIC. Within the subtheme "interactivity and engagement," students expressed that the robot required active participation, which was experienced as more engaging and immersive than passively clicking through the computer-based case. The face projection of the robot added to the feeling of interactivity (Textbox 1).

Skill Acquisition and Development

The theme "skill acquisition and development" consisted of three subthemes: (1) CR skills, (2) communication skills, and (3) emotional skills. Within the subtheme "CR skills," students perceived the robot as an effective tool for acquirement of CR skills by engaging in an active dialogue, practicing the obtainment of medical history, reasoning across different explorative approaches, and planning for various treatment strategies. Students believed that this was due to them actively having to think about the patient's symptoms and ask relevant questions based on the responses they were given. They acknowledged that the computer-based case was more informative, but in the latter, they lacked the ability to ask follow-up questions, which was possible with the LLM-empowered social robot (Textbox 2).

Students experienced that the robotic platform helped them practice "communication skills," especially by facilitating active communication, engaging them in thinking about which questions to ask and how to phrase them in real time (Textbox 2). While the social robotic platform allowed students to feel closer to VPs compared with the computer-based platform, students perceived that the practice of "emotional skills," such as empathy, was inferior with both platforms compared to real-life patient encounters (Textbox 2).

Textbox 2. Relevant quotes by theme "skill acquisition and development" and subthemes from in-depth interviews.

Subtheme: clinical reasoning skills

- "To direct your questions. Patients often come with something very diffuse, like pain. And then you have to think, 'pain, duration, onset [...]' to quickly orient yourself in the patient's problem and then direct your questions afterwards. It's a great tool for practicing that" [Male student, 25 years].
- "The computer case, I would say, is very comprehensive. It gave us the data on pretty much everything, the lab tests, the patient's history. It has data on everything" [Female student, 20 years].
- "I remember asking 'does it get better with a certain medication?' or 'how long does it last?,' 'when was the last time you had it?' [...] And I felt like that's something I couldn't have found in the computer-based case. It didn't give me that information" [Female student, 21 years].

Subtheme: communication skills

- "With the robot, I felt I got to use my skills of how to speak to a patient. [...] you really have to use your medical knowledge to ask [the right] questions and get answers" [Female student, 21 years].

Subtheme: emotional skills

- "Because you have to not only ask about the medical stuff, but also you have to have a dialogue [...] as you do with a human being. So, yeah, you feel closer [to the patient] with the robot compared with the computer" [Female student, 22 years].
- "[...] Just like having that [...] empathy or when you're just talking to a human, looking them into the eyes and saying, 'oh, I'm very sorry for you that you have had these headaches.' You don't really see the pain in someone through a robot" [Female student, 21 years].
- "Maybe what it lacked is [...] emotions. Not like patients whom you need to reassure" [Female student, 19 years].

XSL•FO
RenderX

Procedural Limitations

The theme "procedural limitations" consisted of two subthemes: (1) technical limitations and (2) user-related challenges. Within the subtheme "technical limitations," students expressed technical problems such as being interrupted by the robot during conversational pauses that were conceived as the end of a spoken sentence. This resulted in the perception of a time constraint while responding to the platform, which sometimes hindered the flow of the interaction (Textbox 3). The subtheme "user-related challenges" sheds light on challenges relating to language barriers as well as a need for adapting to the conversational style of the social robot (Textbox 3).

Textbox 3. Relevant quotes by theme "procedural limitations" and subthemes from in-depth interviews.

Subtheme: technical limitations

- "Well, you have a time when you can talk and another time when you can't talk. Apart from the clinical reasoning, you have to pay attention to when [you should talk]. It can be disturbing" [Male student, 22 years].
- "[...] you have to complete your sentence, or else it might cut you off in the middle and start giving you responses" [Female student, 22 years].

Subtheme: user-related challenges

- "It was a bit difficult. Partly because you're not used to speaking English with patients. But also because of how [the robot's] artificial pauses could disturb" [Female student, 23 years].
- "I guess for some people, who are limited by [...] language barriers in English, it might be difficult for them" [Female student, 20 years].

Potential for Improvement

Within the theme "potential for improvement," students expressed how specific technological aspects could be improved in the social robotic platform. Students envisioned the potential of more realistic facial expressions and a multitude of patient personalities, which both could enhance the degree of interactivity and perception of realism, presumably contributing to increased engagement and increased chance of achieving the intended learning outcomes. This potential was limited to the social robotic platform (Textbox 4).

Textbox 4. Relevant quotes by theme "potential for improvement" from in-depth interviews.

- "Maybe [improvements] on the face. To work with the voice. And to broaden the responses. So that it might be more complex" [Male student, 20 years].
- "Like in real life, all patients won't give you the same amount of information. So, I think it can be interesting, apart from [presenting] different diseases, also present different types of personalities" [Male student, 22 years].
- "[...] various personalities and patients who are not cooperative or might need many questions to receive certain responses need different ways of being handled; one patient is not another alike. I think you have to adapt to whichever type of patient you have in front of you" [Female student, 21 years].

Discussion

Principal Findings

We developed a novel social robotic VP platform in conjunction with an LLM and present herein results from a proof-of-concept study that explored the self-perceived usefulness and effectiveness of this new pedagogical modality for CR training of medical students compared with a conventional, computer-based, semilinear VP platform. The study included a quantitative comparison between the 2 platforms and qualitative in-depth interviews. Results indicated that VP exercise with the social robotic platform was perceived as more authentic and immersive as well as conferring a self-perceived benefit regarding the learning effect compared with the computer-based platform. Overall, students experienced that the social robot could promote training of CR, communication, and emotional skills. Students highlighted the limitations of the social robot relating to technical and user-related aspects and provided suggestions for potential improvements within the context of CR training.

Medical students perceived the social robotic platform as more authentic compared to the computer-based platform, as demonstrated by significantly higher scores in the authenticity theme of the questionnaire. The qualitative analysis of the interviews further supported these findings, with students describing the social robotic platform as providing a realistic and immersive experience that closely resembled a real patient encounter. These results are consistent with previous studies that highlight the role of interactivity and complexity within VP platforms in enhancing learning toward defined outcomes [3,12,35]. It was based on this premise that we designed and developed the social robotic VP platform for enhanced interactivity, integrating an LLM for enhanced complexity.

The quantitative analysis revealed that students perceived the social robotic platform as preferable to the computer-based platform in terms of the learning effect of the consultation, particularly regarding the perceived transferability of skills to real-life clinical scenarios. The qualitative analysis of the interviews provided further insights into the specific skills that students believed they acquired through interaction with the robotic platform. Students reported that the robotic platform facilitated the development of CR skills by engaging them in

XSL•FO
RenderX

active dialogue, requiring them to think critically about the patient's symptoms, and allowing them to ask relevant follow-up questions. Additionally, students experienced that the robotic platform helped them practice communication skills by encouraging them to actively engage in conversation and think about how to phrase their questions, which has previously been highlighted as an important aspect for CR training using VPs [29].

The qualitative analysis of the interviews revealed several limitations and challenges associated with the social robotic platform. Students reported technical limitations, such as being interrupted by the robot during conversational pauses, which sometimes made the conversation feel mechanical and hindered the flow of the interaction. This could negatively affect the perceived authenticity of the social robotic platform, highlighting the need for technical improvements in this area as a priority for future development.

Students also reported the benefits of the computer-based platform compared with the social robot, noting it as more informative. Another advantage, though not directly compared in this study, is that the computer-based platform VIC is more accessible, as students can access these cases on their personal computers, whereas the social robot is only available when students are physically present with the modality.

User-related challenges, such as language barriers and the need to adapt to the conversational style of the robot, were also identified. These findings highlight the importance of addressing technical limitations and providing adequate support to users to ensure a smooth and effective learning experience. Despite these limitations, students also expressed that the social robot introduced potential for additional improvements that cannot be facilitated by a conventional computer-based platform, such as more realistic facial expressions and the incorporation of various patient personalities, which can be anticipated to further enhance interactivity and realism.

Limitations and Strengths

This study has several limitations. First, the sample size was small, which limits the generalizability of the findings. Second, the study was conducted in a single center, and the results may not be representative of the experience of medical students in other settings. Third, CR is a complex skill, and how its training is facilitated cannot be comprehensively assessed by a single questionnaire, whose validation may not encompass all its aspects. The questionnaire used in this study was specifically designed to evaluate VP design with a particular focus on CR, but it did not directly assess the acquisition of CR skills; this aspect was instead explored in the qualitative part of the study through in-depth interviews. Our study also has several strengths. The use of a mixed quantitative and qualitative methodological approach allowed for a comprehensive evaluation of students' perceptions and experiences with the VP platforms. Importantly, the study compared a novel social robotic platform in conjunction with an LLM, to our knowledge the first of its kind for medical pedagogy, with a conventional computer-based platform, providing valuable insights into the added value of social robots and artificial intelligence in VP simulations for CR training in HPE curricula.

Implications

Overall, the social robotic VP platform shows promise for enhancing CR training by bridging theoretical learning and clinical practice. Its heightened authenticity and interactivity bring it closer to real-life patient interactions, enabling medical students to practice CR in a realistic environment before engaging with actual patients. However, further evaluation and comparisons with other LLM-driven modalities are needed to assess the unique benefits of using a physical robot versus an LLM alone for VP simulation. Given current accessibility limitations, the platform may be most effective as part of a hybrid approach, combining traditional computer-based VPs for broader access with scheduled sessions using social robotic VPs to deepen skill development. This approach could optimize educational impact while managing practical resource constraints across various medical education settings.

Conclusions

This study demonstrates that a social robotic platform enhanced by an LLM could provide an authentic and engaging learning experience for medical students in the context of VP simulations for training CR. The social robotic platform was perceived as beneficial compared to a conventional computer-based platform in terms of authenticity and overall learning effect, and students reported that it facilitated training of CR, communication, and emotional skills. Despite the limitations and challenges identified, the potential for improvement in the social robotic platform lends promise for this technology in future medical education. Importantly, however, costs and logistics may constitute limitations for the implementation of this technology. Finally, it is worth noting that this was a proof-of-concept study, and further research with larger sample sizes and in diverse educational settings is warranted to fully understand the value of social robotic VPs in facilitating the attainment of learning outcomes within medical education curricula.

Acknowledgments

The authors would like to thank the medical students who participated in the study as well as the medical staff at the Division of Rheumatology at the Karolinska University Hospital, Stockholm, Sweden.

Data Availability

The data generated or analyzed during this study can be provided by the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

AB, CG, MR, SE, GS, and IP conceived the study and designed the work. AB, BJ, VH, KW, and IP contributed to the acquisition of data. AB, BJ, GS, and IP contributed to the statistical analysis. AB, BJ, and IP contributed to the interpretation of the data. AB and IP wrote the original draft of the manuscript. All authors reviewed and approved the final version of the manuscript prior to submission and agreed to be accountable for all aspects of the work.

Conflicts of Interest

IP has received research funding or honoraria from Amgen, AstraZeneca, Aurinia, BMS, Elli Lilly, Gilead, GSK, Janssen, Novartis, Otsuka, and Roche. GS is a cofounder of and chief scientist at Furhat Robotics; due to this conflict of interest, GS was involved in the technical implementation of the social robotic VP platform but not the acquisition or interpretation of data. The other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
Supplementary figures, including the VP plattform evaluation questionnaire, an LLM prompt example, and the interview guide.
[PDF File (Adobe PDF File), 241 KB - jmir_v27i1e63312_app1.pdf]

Multimedia Appendix 2
Supplementary tables, including responses to themes in the VP plattform evaluation questionnaire and identified themes derived from qualitative thematic analysis coding.
[PDF File (Adobe PDF File), 235 KB - jmir_v27i1e63312_app2.pdf]

References

1. Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. Virtual patients—what are we talking about? A framework to classify the meanings of the term in healthcare education. BMC Med Educ 2015 Feb 01;15:11 [FREE Full text] [doi: 10.1186/s12909-015-0296-3] [Medline: 25638167]
2. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
3. Huwendiek S, de Leng BA, Zary N, Fischer MR, Ruiz JG, Ellaway R. Towards a typology of virtual patients. Med Teach 2009 Aug;31(8):743-748. [doi: 10.1080/01421590903124708] [Medline: 19811212]
4. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
5. Edelbring S, Parodis I, Lundberg IE. Increasing reasoning awareness: video analysis of students' two-party virtual patient interactions. JMIR Med Educ 2018 Feb 27;4(1):e4 [FREE Full text] [doi: 10.2196/mededu.9137] [Medline: 29487043]
6. Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. Acad Med 2007 May;82(5):446-451. [doi: 10.1097/ACM.0b013e31803e8a0a] [Medline: 17457063]
7. Posel N, Mcgee JB, Fleiszer DM. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. Med Teach 2015;37(9):813-818. [doi: 10.3109/0142159X.2014.993951] [Medline: 25523009]
8. Kononowicz AA, Hege I, Edelbring S, Sobocan M, Huwendiek S, Durning SJ. The need for longitudinal clinical reasoning teaching and assessment: results of an international survey. Med Teach 2020 Apr;42(4):457-462. [doi: 10.1080/0142159X.2019.1708293] [Medline: 32017640]
9. Rencic J, Trowbridge RL, Fagan M, Szauter K, Durning S. Clinical reasoning education at US medical schools: results from a National Survey of Internal Medicine Clerkship Directors. J Gen Intern Med 2017 Nov;32(11):1242-1246 [FREE Full text] [doi: 10.1007/s11606-017-4159-y] [Medline: 28840454]
10. Parodis I, Andersson L, Durning SJ, Hege I, Knez J, Kononowicz AA, et al. Clinical reasoning needs to be explicitly addressed in health professions curricula: recommendations from a European consortium. Int J Environ Res Public Health 2021 Oct 25;18(21):18 [FREE Full text] [doi: 10.3390/ijerph182111202] [Medline: 34769721]
11. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
12. Edelbring S, Dastmalchi M, Hult H, Lundberg IE, Dahlgren LO. Experiencing virtual patients in clinical learning: a phenomenological study. Adv Health Sci Educ Theory Pract 2011 Aug;16(3):331-345. [doi: 10.1007/s10459-010-9265-0] [Medline: 21656337]
13. Plackett R, Kassianos AP, Mylan S, Kambouri M, Raine R, Sheringham J. The effectiveness of using virtual patient educational tools to improve medical students' clinical reasoning skills: a systematic review. BMC Med Educ 2022 May 13;22(1):365 [FREE Full text] [doi: 10.1186/s12909-022-03410-x] [Medline: 35550085]
14. Campillos-Llanos L, Thomas C, Bilinski, Zweigenbaum P, Rosset S. Designing a virtual patient dialogue system based on terminology-rich resources: challenges and evaluation. Nat Lang Eng 2019;26(2):183-220. [doi: 10.1017/s1351324919000329]

XSL•FO
RenderX

15. Johansson R, Skantze G, Jönsson A. A psychotherapy training environment with virtual patients implemented using the Furhat robot platform. 2017 Presented at: Intelligent Virtual Agents International Conference, IVA 2017; August 27-30, 2017; Stockholm, Sweden. [doi: 10.1007/978-3-319-67401-8_20]
16. Irfan B, Kuoppamäki SM, Skantze G. Between reality and delusion: challenges of applying large language models to companion robots for open-domain dialogues with older adults. Aut Robots 2023 May 02:2023 [FREE Full text] [doi: 10.21203/rs.3.rs-2884789/v1]
17. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach 2024 Apr;46(4):446-470 [FREE Full text] [doi: 10.1080/0142159X.2024.2314198] [Medline: 38423127]
18. Ng H, Koh A, Foong A, Ong J. Real-time hybrid language model for virtual patient conversations. 2023 Presented at: International Conference on Artificial Intelligence in Education; 2023: Springer; July 3-7, 2023; Tokyo, Japan. [doi: 10.1007/978-3-031-36272-9_71]
19. Bakkum MJ, Hartjes MG, Piët JD, Donker EM, Likic R, Sanz E, et al. Using artificial intelligence to create diverse and inclusive medical case vignettes for education. Br J Clin Pharmacol 2024 Mar;90(3):640-648. [doi: 10.1111/bcp.15977] [Medline: 38016816]
20. Coşkun O, Kıyak YS, Budakoğlu I. ChatGPT to generate clinical vignettes for teaching and multiple-choice questions for assessment: a randomized controlled experiment. Med Teach 2025;47(2):268-274. [doi: 10.1080/0142159X.2024.2327477] [Medline: 38478902]
21. Borg A, Jobs B, Huss V, Gentline C, Espinosa F, Ruiz M, et al. Enhancing clinical reasoning skills for medical students: a qualitative comparison of LLM-powered social robotic versus computer-based virtual patients within rheumatology. Rheumatol Int 2024;44(12):3041-3051. [doi: 10.1007/s00296-024-05731-0] [Medline: 39412574]
22. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis Rheum 2008;58(1):15-25 [FREE Full text] [doi: 10.1002/art.23177] [Medline: 18163481]
23. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58(1):26-35 [FREE Full text] [doi: 10.1002/art.23176] [Medline: 18163497]
24. Pina Gonçalves N, Azeredo S, Sepriano A, Henriques AR, Pires T, Branco JC, et al. Prevalence of undiagnosed rheumatic and musculoskeletal diseases and its association with health-related quality of life and with physical function. ARP Rheumatol 2023;2(1):7-16 [FREE Full text] [Medline: 36739530]
25. Hege I, Adler M, Donath D, Durning SJ, Edelbring S, Elvén M, et al. Developing a European longitudinal and interprofessional curriculum for clinical reasoning. Diagnosis (Berl) 2023;10(3):218-224 [FREE Full text] [doi: 10.1515/dx-2022-0103] [Medline: 36800998]
26. Huwendiek S, de Leng BA, Kononowicz AA, Kunzmann R, Muijtjens AM, van der Vleuten CP, et al. Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning. Med Teach 2015;37(8):775-782. [doi: 10.3109/0142159X.2014.970622] [Medline: 25313931]
27. Al Moubayed S, Beskow J, Skantze G, Granström B. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. 2012 Presented at: Cognitive Behavioural Systems: COST 2102 International Training School; February 21-26, 2011; Dresden, Germany. [doi: 10.1007/978-3-642-34584-5_9]
28. Virtual Interactive Case System. URL: https://pie.med.utoronto.ca/VIC/index.htm [accessed 2024-04-17]
29. Posel N, Fleiszer D, Shore BM. 12 Tips: guidelines for authoring virtual patient cases. Med Teach 2009;31(8):701-708. [doi: 10.1080/01421590902793867] [Medline: 19513927]
30. Mishra C, Verdonschot R, Hagoort P, Skantze G. Real-time emotion generation in human-robot dialogue using large language models. Front Robot AI 2023;10:1271610 [FREE Full text] [doi: 10.3389/frobt.2023.1271610] [Medline: 38106543]
31. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. 2020 Presented at: Advances in Neural Information Processing Systems; December 6-12, 2020; Vancouver, BC p. 1877-1901 URL: https://dl.acm.org/doi/abs/10.5555/3495724.3495883
32. Skantze G, Johansson M, Beskow J. Exploring turn-taking cues in multi-party human-robot discussions about objects. 2015 Presented at: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction; November 9-13, 2015; Seattle, WA. [doi: 10.1145/2818346.2820749]
33. Mishra C, Offrede T, Fuchs S, Mooshammer C, Skantze G. Does a robot's gaze aversion affect human gaze aversion? Front Robot AI 2023;10:1127626 [FREE Full text] [doi: 10.3389/frobt.2023.1127626] [Medline: 37427087]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CPM, Haag M, et al. Design principles for virtual patients: a focus group study among students. Med Educ 2009;43(6):580-588. [doi: 10.1111/j.1365-2923.2009.03369.x] [Medline: 19493183]

XSL•FO
RenderX

Abbreviations

CR: clinical reasoning
HPE: health professions education
KI: Karolinska Institutet
LLM: large language model
OR: odds ratio
VIC: virtual interactive case simulator
VP: virtual patient

Edited by KJ Craig; submitted 16.06.24; peer-reviewed by M Wijnen-Meijer, YS Kiyak; comments to author 22.10.24; revised version received 03.11.24; accepted 10.02.25; published 03.03.25.

Please cite as:
Borg A, Georg C, Jobs B, Huss V, Waldenlind K, Ruiz M, Edelbring S, Skantze G, Parodis I
Virtual Patient Simulations Using Social Robotics Combined With Large Language Models for Clinical Reasoning Training in Medical Education: Mixed Methods Study
J Med Internet Res 2025;27:e63312
URL: https://www.jmir.org/2025/1/e63312
doi:10.2196/63312
PMID:40053778

©Alexander Borg, Carina Georg, Benjamin Jobs, Viking Huss, Kristin Waldenlind, Mini Ruiz, Samuel Edelbring, Gabriel Skantze, Ioannis Parodis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Asynchronous Distance Learning Performance and Knowledge Retention of the National Institutes of Health Stroke Scale Among Health Care Professionals Using Video or e-Learning: Web-based Randomized Controlled Trial

Avinash Koka[1], MD; Loric Stuby[2], EMT-P; Emmanuel Carrera[3], MD; Ahmed Gabr[4], MD; Margaret O'Connor[4], MD; Nathalie Missilier Peruzzo[3], RN; Olivier Waeterloot[3], RN; Friedrich Medlin[5], MD; Fabien Rigolet[6], RN; Thomas Schmutz[7], MD; Patrik Michel[8], MD; Thibaut Desmettre[1], MD; Mélanie Suppan[9], MD, MSc; Laurent Suppan[1], MD

[1]Division of Emergency Medicine, Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals and Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2]Genève TEAM Ambulances, Emergency Medical Services, Geneva, Switzerland

[3]Stroke Centre, Department of Clinical Neurosciences, Geneva University Hospitals and Faculty of Medicine, University of Geneva, Geneva, Switzerland

[4]Department of Ageing and Therapeutics, University of Limerick Hospital Group, University Hospital Limerick, Limerick, Ireland

[5]Stroke Unit, Division of Neurology, Fribourg Cantonal Hospital, Fribourg, Switzerland

[6]Care training center, Fribourg Cantonal Hospital, Fribourg, Switzerland

[7]Division of Emergency Medicine, Fribourg Cantonal Hospital, Fribourg, Switzerland

[8]Stroke Centre, Division of Neurology, Department of Clinical Neurosciences, Lausanne University Hospital, Lausanne, Switzerland

[9]Division of Anesthesiology, Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals and Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
Geneva University Hospitals and Faculty of Medicine, University of Geneva
Rue Gabriel-Perret-Gentil 4
Geneva, 1205
Switzerland
Phone: 41 223723311
Email: laurent.suppan@hug.ch

Abstract

Background: Stroke treatment has significantly improved over the last decades, but the complexity of stroke cases requires specialized care through dedicated teams with specific knowledge and training. The National Institutes of Health Stroke Scale (NIHSS), widely used to assess neurological deficits and make treatment decisions, is reliable but requires specific training and certification. The traditional didactic training method, based on a video, may not adequately address certain NIHSS intricacies nor engage health care professionals (HCPs) in continuous learning, leading to suboptimal proficiency. In the context of time-constrained clinical settings, highly interactive e-learning could be a promising alternative for NIHSS knowledge acquisition and retention.

Objective: This study aimed to assess the efficacy of a highly interactive e-learning module compared with a traditional didactic video in improving NIHSS knowledge among previously trained HCPs. Furthermore, its impact on knowledge retention was also assessed.

Methods: A prospective, multicentric, triple-blind, and web-based randomized controlled trial was conducted in 3 Swiss university hospitals, involving HCPs previously trained in NIHSS. Invitations were sent through email, and participants were randomized to either the e-learning or traditional didactic video group through a fully automated process upon self-registration on the website. A 50-question quiz was administered before and after exposure to the training method, and scores were compared to assess knowledge acquisition. The quiz was repeated after 1 month to evaluate retention. Subjective assessments of learning methods that is, user satisfaction, probability of recommendation, perceived difficulty, and perception of duration, were also

XSL•FO
RenderX

collected through a Likert-scale questionnaire. A sample size of 72 participants were deemed necessary to have an 80% chance of detecting a difference of 2 points in the postcourse quiz between groups at the 5% significance level.

Results: Invitations to participate were sent through email to an estimated 325 HCPs. 174 HCPs enrolled in the study, of which 97 completed the study course. Both learning methods significantly improved NIHSS knowledge, with an improvement of 3.2 (range 2.0-4.3) points in the e-learning group and of 2.1 (1.2-3.1) points in the video group. However, the e-learning group performed better, with higher scores in knowledge acquisition (median score 39.0, IQR 36.0-41.0 vs 37, IQR 34.0-39.0; *P*=.03) and in knowledge retention (mean score 38.2, 95% CI 36.7-39.7 vs 35.8, 95% CI 34.8-36.8; *P*=.007). Participants in the e-learning group were more likely to recommend the learning method (77% vs 49%, *P*=.02), while no significant difference was found for satisfaction (*P*=.17), perceived duration (*P*=.17), and difficulty (*P*=.32).

Conclusions: A highly interactive e-learning module was found to be an effective asynchronous method for NIHSS knowledge acquisition and retention in previously NIHSS-trained HCPs, and may now be considered for inclusion in NIHSS training programs for HCPs.

International Registered Report Identifier (IRRID): RR2-10.3390/healthcare9111460

(J Med Internet Res 2025;27:e63136) doi:10.2196/63136

KEYWORDS

stroke; e-learning; video; medical education; randomized controlled trial; knowledge retention; knowledge acquisition; NIHSS; National Institutes of Health Stroke Scale; learner satisfaction

Introduction

Background

Despite the decline in stroke incidence and mortality since 1990, the absolute number of stroke cases are increasing due to the steep growth in the global population, along with an aging demographic and a greater burden of risk factors across many parts of the world [1]. Hence, stroke remains a major public health issue due to its high morbidity rates [2]. Acute reperfusion therapies, such as intravenous thrombolysis and thrombectomy, improve functional and survival prognosis after stroke [3,4]. However, patients must be carefully screened and selected, and these procedures must be performed within a limited timeframe, due to risks and complications associated with the treatment [3].

The National Institutes of Health Stroke Scale (NIHSS) is widely used around the world to triage, select treatment strategies, monitor and follow-up stroke patients [3,5]. First described in the late 1980s, it gained popularity after the publication of the first successful acute stroke treatment study [5,6]. The reliability of the NIHSS has been demonstrated across diverse populations of health care professionals (HCPs) [7,8]. However, prior training and certification are required [9-14] as it is a complex scale with numerous subtleties. In Switzerland, national stroke center and stroke unit certification procedures specify that all personnel must be appropriately trained. In the French-speaking part of Switzerland, all HCPs working in stroke units should be familiar with NIHSS. However, the training program, usually comprised of formal teaching sessions with a senior neurologist and bedside clinical training, is non standardized and center-specific [15-17].

Constraints of working in these 24-hour acute care units, such as unexpected inflow of patients, and weekend and night shift patterns, can limit HCPs exposure to regular clinical training sessions. The absence of time constraints and the self-paced nature of asynchronous digital learning methods such as electronic learning (e-learning) modules have contributed to their popularity in medical education [18], and have been extensively used during the COVID-19 pandemic [19,20]. Hence, a highly interactive e-learning module was created to teach the NIHSS [21]. Its use has been shown to improve user satisfaction, knowledge acquisition as well as dissemination in NIHSS-naive populations, when compared with the traditional didactic video alone [22,23].

Despite widespread usage of the NIHSS, little literature exists on the actual performance of HCPs using it on a regular basis. Over time, with routine practice and bedside training, some key aspects of NIHSS scoring may be overlooked, improperly practiced, or forgotten. Therefore, we hypothesized that even previously trained HCPs may have suboptimal NIHSS knowledge and that a refresher course using an e-learning module could be more effective than the traditional didactic video at improving their NIHSS proficiency. Further, we also hypothesized that knowledge retention at 1 month would be higher after following this interactive module than after reviewing the standard didactic video.

Given the heterogeneity of training described above, all HCPs authorized by their employers to care for neurological patients on a regular basis were considered eligible to participate in this study.

Objectives

Our primary objective was to assess whether NIHSS-trained HCPs would significantly improve their NIHSS knowledge after completing a highly interactive e-learning module than after following the traditional didactic video. The secondary objective was to determine whether following either training material allowed better retention of knowledge at one month.

Methods

Study Design and Setting

This was a prospective, multicentric, web-based, triple-blind (participants, investigators, and data analyst) randomized

XSL•FO
RenderX

controlled trial. Its main aim was to compare scores to a 50-question quiz, taken before and after exposure to the randomly assigned learning method, to assess the impact on knowledge acquisition. The same quiz was repeated 1 month later to assess the impact on knowledge retention. Figure 1 provides a detailed infographic of the study design.

Figure 1. Study design.

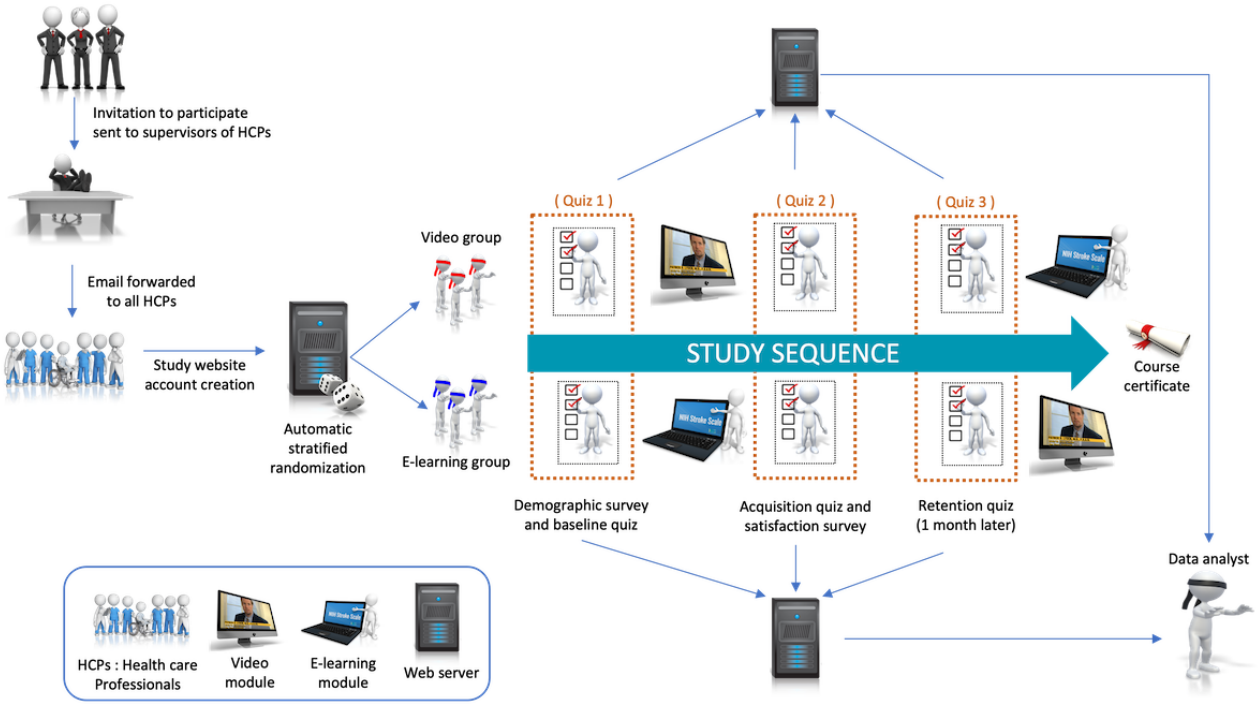

This study was initially designed to include HCPs from the stroke units and neurology wards of 2 university hospitals in the French-speaking part of Switzerland, namely the Geneva University Hospitals and the Lausanne University Hospital. A third center, the Cantonal Hospital of Fribourg, joined the study a few weeks before its launch. Cantonal Hospital of Fribourg supervisors also extended the study invitation to all their personnel who potentially used the NIHSS on a regular basis, which included intensive care unit HCPs and emergency ward nurses.

The medical and nursing supervisors of all 3 centers forwarded an invitation email to their teams, containing a summary of the main points of the study and a link to the online platform. It is estimated that approximately 185 nurses and 140 physicians received this email. One author (L Suppan) sent weekly email updates to the supervisors of each center summarizing the number of participants in each hospital, stratified by profession. This email also detailed the number of participants at each stage of the study, thus enabling supervisors to monitor participation. Reminder emails regarding enrolment were left to the discretion of the supervisors. Automatic reminders were sent to participants who had registered on the platform but had not completed the required and appropriate study steps.

The initial invitation email was sent in July 2022. Enrollment and registration of new participants was disabled on March 31, 2023, and ongoing participants were invited to complete all steps by May 31, 2023. While the study platform remained accessible for scientific purposes such as further data extraction, users were no longer able to log in.

The study protocol has been previously published [24] and protocol deviations are reported in Multimedia Appendix 1.

CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [25] and checklist (Multimedia Appendix 2) were followed, integrating relevant elements from the CHERRIES (Checklist for Reporting Results of Internet E-Survey) [26] and iCHECK-DH (Guidelines and Checklist for Reporting on Digital Health Implementations) guidelines [27].

The RCT was not prospectively registered because registration is not necessary for studies evaluating the effect of interventions on providers rather than patients [28].

Online Platform

The internet-based online study platform, compliant with General Data Protection Regulation, was created using the Joomla 3.10 content management system (Open Source Matters). Participants were assigned to specific user groups according to their progression using Joomla's Access Control List. Custom PHP code was embedded using Sourcerer 9 (Regular Labs) to manage user group assignments.

The self-enrolment process was entirely automated. The first page of the website contained all necessary information regarding the study, and the consent information form was downloadable. Participants were explicitly informed that completing the registration process implied consent to participate in the study. Those who wished to enroll had to click on the logo of their hospital before acknowledging their profession (doctor or nurse) by clicking on a captioned image. This procedure invisibly randomized them either to the video or to

XSL•FO
RenderX

the e-learning group using Gegabyte's Random Article module 2.3. A 1:1 allocation was not possible due to the stratification by profession and center, and an unknown total number of participants. Participants were then asked to self-evaluate their current NIHSS knowledge (limited, moderate, or extended). The Membership Pro 3 component (Joomdonation) was then used to show a simple registration form that only asked for an email address and a secure password.

First Questionnaire

After completing the registration process, participants were asked to fill in a first questionnaire (Multimedia Appendix 3) designed to collect demographic data, administered using the Community Surveys 5.6 component (Shondalai) [29]. A detailed description of the first questionnaire can be found in the study protocol [24].

Participants then completed a 50-question quiz (Multimedia Appendix 4) designed to assess their baseline knowledge (quiz 1). Each question, worth 1 point, pertained to 1 of the 15 NIHSS items. All possible scores for that item were displayed and the participant had to choose the appropriate score. The quiz comprised a full NIHSS evaluation of 3 patients, for a total of 45 points. A total of 5 additional questions pertained to the general principles of NIHSS and certain subtleties, bringing the total to 50 points. All questions were mandatory, with multiple-choice answers to choose from.

Completion of this first quiz granted access to the allocated learning material as per randomization.

Learning Material and Second Questionnaire

The video group was given access to the original video created by Professor Lyden, of a duration of 53:07 minutes. The French subtitled version used for this study is freely available online under a Creative Commons license [30].

The e-learning group was given access to version 21c of the interactive e-learning module created under Storyline 3 (Articulate Global), also available freely online under a Creative Commons license [21]. This web-based module, which contains 184 slides, is divided into chapters following the NIHSS structure. Subtitled excerpts of Professor Lyden's original video are embedded into the different chapters to illustrate clinical testing. This e-learning module has been extensively described in previous publications as well as in the protocol of this study [22-24].

Participants had full control of the allocated learning material, could go back and forth as often as needed, and pause, resume, restart, and review chapters, all without constraints or time limitations. Upon completion of the learning material, participants were then taken through the same 50-question quiz to assess knowledge acquisition (quiz 2).

This quiz was followed by a short satisfaction survey collecting subjective outcomes (ie, satisfaction, difficulty, duration, and likelihood of recommendation) using a 5-point Likert scale (Multimedia Appendix 5).

Knowledge Retention

Four weeks after completion of the satisfaction survey, participants were invited through email to complete the final quiz. They were asked to answer the 50-question quiz a third time to assess knowledge retention (quiz 3). At the end of this quiz, a certificate of participation was automatically generated that could be downloaded and printed. Only the email ID used for registration appeared automatically on the certificate, and participants were asked to write their names down after printing it. Each certificate was embedded with a unique barcode that could be scanned for verification of authenticity.

Outcomes

The primary outcome was knowledge acquisition, assessed by the score on the second quiz, taken after completing the learning material.

Secondary outcomes were knowledge retention, assessed by the score on the third quiz undertaken at least 1 month after completion of the learning material, and subjective assessments of the followed learning method, namely user satisfaction, user perception of the duration, perceived difficulty, and the likelihood of recommending the learning material to a colleague.

Participants and Sample Size

A total of 72 participants were required to have 80% chance of detecting a difference of 2 points in the postcourse 50-question quiz between groups at the 5% significance level. Considering a 40% attrition rate, we aimed to recruit 120 participants. A higher number of participants were accepted as participation did not entail any risk.

Data Curation and Statistical Analysis

Data curation and statistical analysis were carried out using Stata (version 15.1; StataCorp). Data curation was done by 1 author (L Suppan), who assigned neutral names to the e-learning group and the video group. Another author (L Stuby) was then given this curated and blinded DTA file for data analysis. Continuous variables were first graphically described to look for the shape of the distribution. In case of doubtful normality, the Shapiro-Wilk test was applied. Then, depending on the variable's distribution, either the Student *t* test or the Mann-Whitney *U* test were used. Results are reported either using mean (95% CI) or median (IQR).

A sensitivity analysis was performed by excluding those who had previously followed either the e-learning module, the original video from Patrick Lyden [20], or any other official NIHSS training. Categorical variables (eg, Likert scales) were first analyzed graphically, and then by Fisher exact test.

The time interval between the end of quiz 2 and the start of quiz 3 was also analyzed to ensure that the time elapsed before assessing retention was not lower than required as per the original protocol.

Ethical Considerations

The regional ethics committee issued a "Declaration of no objection" in response to a jurisdictional query that we submitted (Req-2021-00543), as this study does not involve patients and falls outside the scope of the Swiss legislation regulating

XSL•FO
RenderX

research on human subjects. Potential participants were provided with a detailed 4-page study information document, downloadable through a link on the study web page. Some of the key points noted in the document were that participation was on a voluntary basis, without any financial compensation, and data was fully anonymized before analysis. Furthermore, the document, and the study page, clearly indicated that registering for the study by creating an account was considered as acceptance to participate in the study. A screenshot of the study web page, and the consent information document, are annexed (Multimedia Appendices 6 and 7).

Results

Participants Characteristics

A total of 174 participants enrolled in the study. The fully automated randomization process attributed 59 participants to the e-learning group, and 115 participants to the video group (Figure 2). Characteristics of participants having completed the study path are described in Table 1. Given the high attrition rates, the characteristics of the participants who dropped out were also extracted and are displayed in Multimedia Appendix 8.

Figure 2. Study flowchart as per CONSORT guidelines. CONSORT: Consolidated Standards of Reporting Trials.

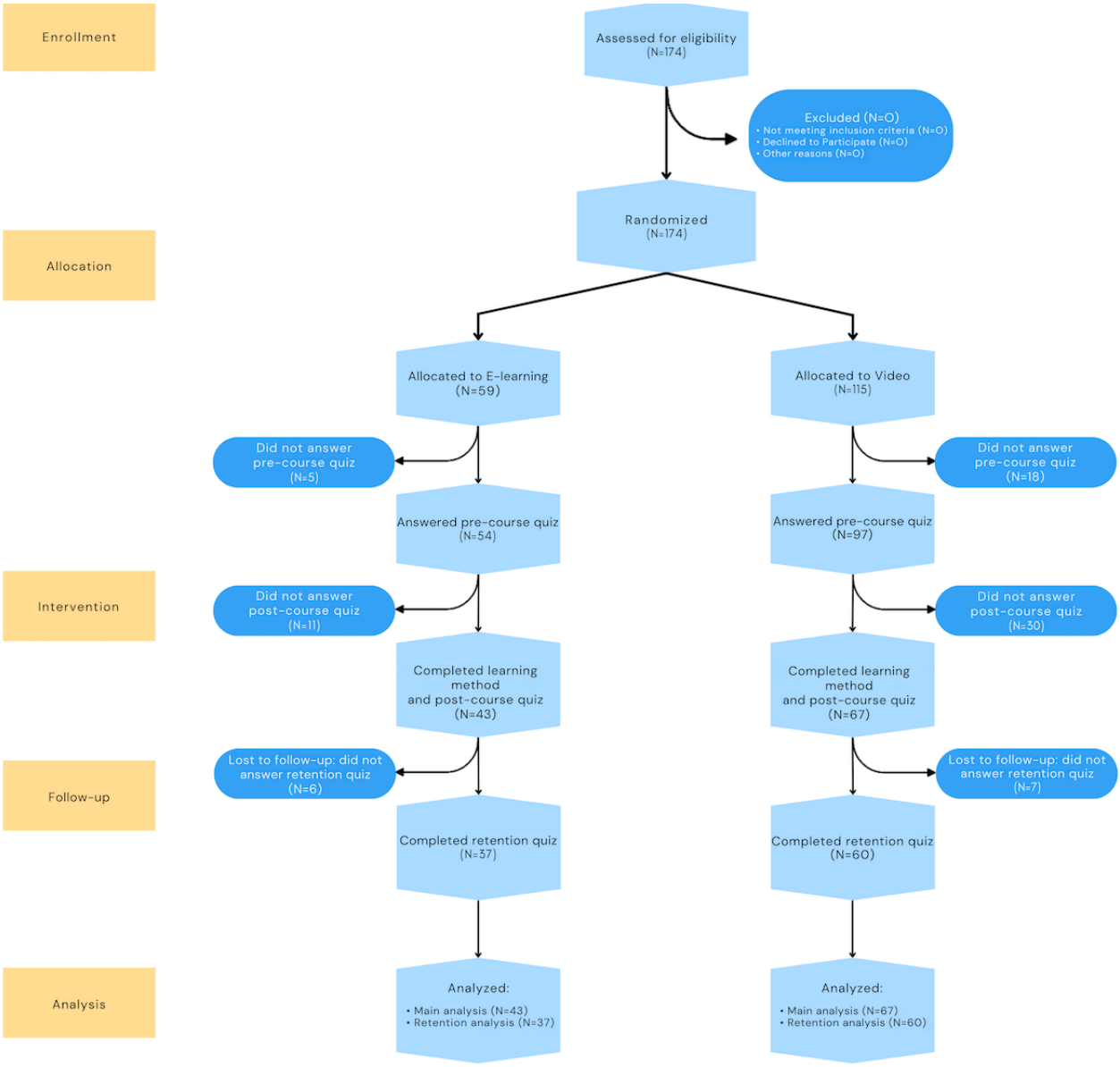

XSL•FO
RenderX

Table 1. Characteristics of participants having completed the study path.

Characteristics[a]	e-Learning group (n=43)	Video group (n=67)
Age (years), median (IQR)	32 (27-41)	32 (29-38)
Gender, n (%)		
Female	33 (76.7)	52 (77.6)
Male	10 (23.3)	15 (22.4)
Profession, n (%)		
Nurse	26 (60.5)	49 (73.1)
Physician	17 (39.5)	18 (26.9)
Time since certification (years), median (IQR)	6 (2-12)	8 (2-13)
Center, n (%)		
HUG[b]	13 (30.2)	30 (44.8)
CHUV[c]	7 (16.3)	6 (9)
HFR[d]	23 (53.5)	31 (46.3)
Service, n (%)		
Ward	10 (23.3)	11 (16.4)
HDU[e]	13 (30.2)	34 (50.8)
ICU[f]	0 (0.0)	3 (4.5)
ED[g]	13 (30.2)	12 (17.9)
Other	7 (16.3)	7 (10.5)
Time in main service (years), median (IQR)	2 (0-5)	3 (1-5)
French mastery, n (%)		
None	0 (0)	0 (0)
Basic	0 (0)	0 (0)
Intermediate	1 (2.3)	0 (0)
Advanced	3 (7)	9 (13.4)
Proficient	39 (90.7)	58 (86.6)
English mastery, n (%)		
None	3 (7)	4 (6)
Basic	10 (23.3)	25 (37.3)
Intermediate	12 (27.9)	24 (35.8)
Advanced	17 (39.5)	14 (20.9)
Proficient	1 (2.3)	0 (0)
NIHSS[h] training		
NIHSS internal training, n (%)	19 (44.2)	38 (56.7)
NIHSS official training, n (%)	2 (4.7)	5 (7.5)
Using NIHSS since, (years), median (IQR)	3 (1-7)	3 (1-5)
NIHSS use frequency, n (%)		
<1/month	6 (14)	3 (4.5)
1×/month	7 (16.3)	10 (14.9)
1×/week	8 (18.6)	14 (20.9)
1×/day	6 (14)	8 (11.9)

XSL•FO
RenderX

Characteristics[a]	e-Learning group (n=43)	Video group (n=67)
>1×/day	16 (37.2)	32 (47.8)
Comfort with NIHSS use, n (%)		
Not comfortable at all	3 (7)	1 (1.5)
Not so comfortable	7 (16.3)	6 (9)
Moderately comfortable	13 (30.2)	17 (25.4)
Quite comfortable	15 (34.9)	32 (47.8)
Very comfortable	5 (11.6)	11 (16.4)
NIHSS expertise, n (%)		
Limited	14 (32.6)	15 (22.4)
Moderate	20 (46.5)	22 (32.8)
Extended	9 (20.9)	30 (44.8)
Baseline performance (quiz 1), mean (95% CI)	34.9 (33.7-36.2)	34.6 (33.8-35.3)

[a]Total may not be exactly 100% due to rounding.
[b]HUG: Geneva University Hospitals.
[c]CHUV: Lausanne University Hospital.
[d]HFR: Fribourg Cantonal Hospital.
[e]HDU: high dependency unit.
[f]ICU: intensive care unit.
[g]ED: emergency department.
[h]NIHSS: National Institutes of Health Stroke Scale.

Participation According to Group Allocation

A difference of 10.5% in the participation rate was found, with 37/59 (62.7%) participants completing the study course in the e-learning group versus 60/115 (52.2%) in the video group.

Knowledge Acquisition

Both learning methods had a positive impact on knowledge acquisition as both groups improved their scores. The improvement in the e-learning group was a mean of 3.2 points (range 2.0-4.3) and of 2.1 points (range 1.2-3.1) in the video group.

Participants who followed the e-learning method performed better in the acquisition quiz than the participants in the video group (median score of 39, IQR 36-41 vs 37, IQR 34-39; *P*=.03; Figure 3).

Figure 3. NIHSS knowledge acquisition scores (quiz 2) among groups. NIHSS: National Institutes of Health Stroke Scale.

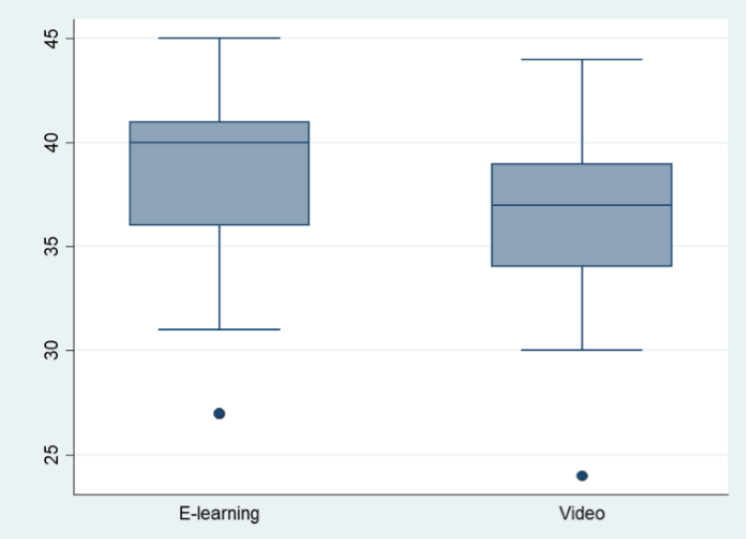

XSL•FO
RenderX

Knowledge Retention

The participants who followed the e-learning module had significantly higher scores on the retention quiz (38.2, 95% CI 36.7-39.7 vs 35.8, 95% CI 34.8-36.8; *P*=.007; Figure 4).

Figure 4. NIHSS knowledge retention scores (quiz 3) among groups. NIHSS: National Institutes of Health Stroke Scale.

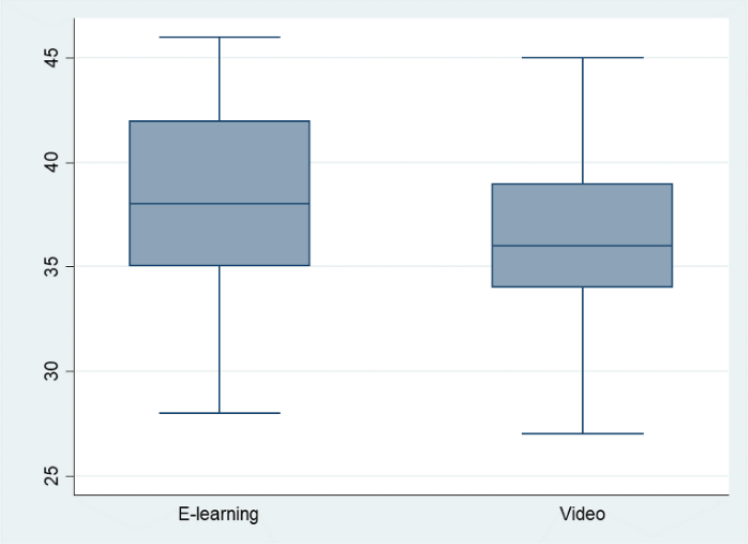

Subjective Assessments

All participants who completed the learning method answered the subjective questionnaire. The likelihood of recommending the learning method was statistically higher in the e-learning group, with 33/43 (77%) participants in the e-learning group and 33/67 (49%) participants in the video group (*P*=.02; Figure 5). There was no significant difference regarding satisfaction (*P*=.17), perceived duration (*P*=.17), and difficulty (*P*=.32) (Multimedia Appendices 9-11).

Figure 5. Likelihood of recommending the learning method to a colleague.

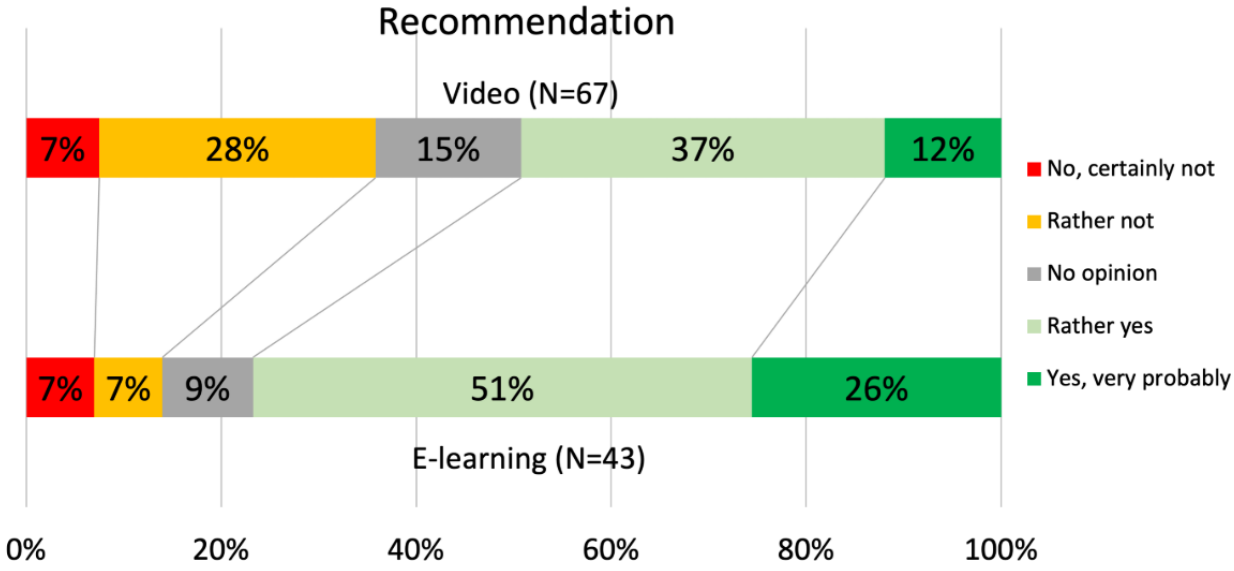

Sensitivity Analysis

The prespecified sensitivity analysis (ie, excluding those who had previously followed either learning method or any official NIHSS training) did not change the results (Table 2).

Table 2. Sensitivity analyses.

	e-Learning	Video	*P* value
Sensitivity analysis of primary outcome (knowledge acquisition), median (IQR), n	39 (37-41), 40	36 (34-39), 59	.02
Sensitivity analysis of knowledge retention, mean (95% CI), n	38.2 (36.6-39.7), 33	35.6 (34.5-36.7), 52	.007

XSL•FO
RenderX

Time Interval Analysis

Time interval analysis confirmed that about a month had elapsed between the start of quiz 2 and the start of quiz 3 in both groups (e-learning versus video median of 36.0, IQR 27.4-49.0 versus 35.3, IQR 28.3-58.9 days).

Discussion

Principal Findings

This study shows that when compared with the traditional didactic video, a highly interactive e-learning module improves NIHSS knowledge acquisition and retention in HCPs caring for patients with stroke.

Decline in knowledge retention has been shown to be nonlinear, with a high percentage of decline measurable after only 4 weeks of acquisition [31]. The higher scores at quiz 3, undertaken at least 4 weeks after quiz 2, demonstrate that this interactive e-learning module also has a significant impact on knowledge retention.

Although both learning methods were considered equally difficult, participants of the e-learning group expressed a higher likelihood of recommending the learning method to colleagues, and thus would probably facilitate knowledge dissemination. Further, it is worth noting that the number of participants who completed the learning process was higher in the e-learning group by more than 10%, also hinting toward higher satisfaction in this group.

Comparison to Previous Work

Previous studies have shown a positive impact of the e-learning module on knowledge acquisition among paramedics and medical students not familiar with the NIHSS [22,23]. However, this is the first study to show similar results in HCPs with prior NIHSS knowledge. This is also the first study to test for impact on knowledge retention as mentioned above.

Strengths and Limitations

Several limitations must be acknowledged. First, attrition rates, as anticipated, were high in both groups. This is probably related to the fact that the entire study path takes a long time, requiring 4 to 5 hours to be completed and that there is currently a very high rate of fatigue among HCPs, with high turnover rates [32]. The study was conducted over a prolonged period of 11 months and participants were allowed to pause and resume the study at any time to try to minimize this issue.

Second, participants were informed that the study was to be completed individually and without any visual support such as a NIHSS form during the quiz. However, given the web-based asynchronous design of the study, we are unable to verify compliance to these points. The randomization however mitigates this issue, should it hold true.

Third, there is no certainty that multiple participants following different learning methods did not collaborate during this study, creating a potential contamination in the learning process. While a cluster-based randomization, with each center being an individual cluster, could have mitigated this risk, the large sample size required for such a design would have seriously impacted the feasibility of this study. In addition, based on their teaching experience with the study participants, the authors and supervisors strongly believe that the risk of such contamination is very low.

Fourth, the improvement of the score at each repetition could be partly induced by a priming effect. However, this effect would be similar across both groups and therefore should not induce any bias in data interpretation when comparing both groups.

Finally, trying to determine NIHSS knowledge by using a 50-question quiz cannot be considered fully representative of the skills in the clinical application of the score. However, this method has been used for the previous 2 decades to test NIHSS competency [11,14,33]. Its usage can therefore be considered valid to compare multiple learning methods.

This study also has some strong points, which are mentioned here. To our knowledge, this is the first paper that assesses the impact of an e-learning method on a population of previously trained HCPs, in terms of knowledge acquisition and retention. The protocol was published before starting this study, and deviations from the protocol have been thoroughly detailed in this paper (Multimedia Appendix 1). The randomization process was fully automated enabling us to guarantee the concealment of allocation, and data blinding was accomplished at the highest standard. The elapsed time between the second and third quizzes was assessed to ensure that the interval was indeed more than 1 month, thus ensuring compliance with the protocol.

Because language proficiency could have an impact on learning performance, participants were asked about their levels of English and French in the initial demographic questionnaire. Although randomization was intended to guarantee an equal balance of these elements in each group, the comparison allowed us to ascertain it.

The e-learning module itself also has certain advantages. Access to the official NIHSS training and certification could be limited by financial constraints and linguistic barriers, as most participating HCPs are French-speaking. This e-learning module is freely available online, its asynchronous learning design is particularly adapted to HCPs constraints, and it can easily be translated into different languages. The intrinsic design of the platform also enables quick updates of individual chapters if needed in the future. All these points allow easier dissemination of NIHSS knowledge and skills through e-learning than with the traditional didactic English video alone.

Future Directions

Compared with previous studies in populations not familiar with the NIHSS [22,23], the overall scores are slightly higher. However, these scores are lower than could be expected given the specific study population of HCPs with previous NIHSS knowledge using the scale on a regular basis. Several hypotheses could explain this finding. First, baseline data analysis shows that 15% of participants declare themselves uncomfortable with the application of NIHSS, 26% consider themselves as having limited NIHSS knowledge, and only 7 participants have followed the official NIHSS training and certification. Furthermore, only about half the HCPs participating in this

XSL•FO
RenderX

study declare having been through the internal training program of their institution. These low numbers could partly be explained by a high turnover currently seen in HCPs in general since the COVID-19 pandemic [34,35], with many new staff members who might have recently started working in these units as reflected by the low number of years the participants have been in their service. Moreover, some of the participants, such as Intensive care unit HCPs and emergency ward nurses, may have less exposure to NIHSS usage in their daily practice. Further studies should investigate the association between self-assessed level of NIHSS expertise, and frequency of exposition to the score, with the actual NIHSS knowledge.

The videos used for the quiz could also have an impact on the overall scores. A previous study [23], using the first edition of the e-learning module, had shown that the lack of video extracts, or graphic examples of neurological deficits, decreased the performance of candidates for certain chapters. An updated version integrating the missing video extracts showed that the performance of the same chapters improved [22]. However, feedback from previous study participants and discussions with experts correlate to say that some of the video excerpts in the e-learning or in the quiz could contain some ambiguity and leave room for some interpretation, for example, due to camera angles. We believe that this could in part explain some incorrect scoring and thus reduce the overall scores, and that replacing the videos with animated models demonstrating the neurological deficits more clearly could improve overall performance at testing.

Conclusion

A highly interactive e-learning module was found to be an effective asynchronous method to improve NIHSS knowledge acquisition and retention, notably in previously NIHSS-trained HCPs. It may now be considered for inclusion in NIHSS training programs for HCPs.

Acknowledgments

The authors would like to thank all participants for their valuable time, and the numerous supervisors for facilitating this study.

Further, the authors would also like to thank the University of Geneva for sponsoring publication fees, as per institutional guidelines.

Data Availability

The datasets generated during and/or analyzed during this study are available in the Yareta repository [36].

Authors' Contributions

L Suppan, MS, EC, PM, L Stuby and AK contributed to conceptualization and methodology. MS, L Suppan, L Stuby, and AK handled software and project administration. L Suppan contributed to supervision. AK, L Suppan, MS, FM, TS, FR, EC, NMP, PM, and L Stuby conducted investigations. L Suppan and MS handled data curation. L Stuby performed formal analysis. AK, L Suppan, and L Stuby contributed to writing-original draft preparation. All the authors performed writing-review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol deviations.
[PDF File (Adobe PDF File), 52 KB - jmir_v27i1e63136_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1798 KB - jmir_v27i1e63136_app2.pdf]

Multimedia Appendix 3
First questionnaire.
[DOC File , 43 KB - jmir_v27i1e63136_app3.doc]

Multimedia Appendix 4
50 Questions quiz.
[XLSX File (Microsoft Excel File), 13 KB - jmir_v27i1e63136_app4.xlsx]

Multimedia Appendix 5
Satisfaction Survey questionnaire.

XSL•FO
RenderX

[DOC File , 40 KB - jmir_v27i1e63136_app5.doc]

Multimedia Appendix 6
Study webpage screenshot.
[PDF File (Adobe PDF File), 158 KB - jmir_v27i1e63136_app6.pdf]

Multimedia Appendix 7
Study consent form.
[PDF File (Adobe PDF File), 449 KB - jmir_v27i1e63136_app7.pdf]

Multimedia Appendix 8
Characteristics of all participants.
[DOC File , 51 KB - jmir_v27i1e63136_app8.doc]

Multimedia Appendix 9
User satisfaction of learning method.
[PNG File , 106 KB - jmir_v27i1e63136_app9.png]

Multimedia Appendix 10
Perceived difficulty of the learning method.
[PNG File , 108 KB - jmir_v27i1e63136_app10.png]

Multimedia Appendix 11
Perceived duration of the learning method.
[PNG File , 104 KB - jmir_v27i1e63136_app11.png]

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet Neurol 2021;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
2. Donkor ES. Stroke in the 21 century: a snapshot of the burden, epidemiology, and quality of life. Stroke Res Treat 2018;2018:3238165 [FREE Full text] [doi: 10.1155/2018/3238165] [Medline: 30598741]
3. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, American Heart Association Stroke Council. 2018 guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American heart association/American stroke association. Stroke 2018;49(3):e46-e110 [FREE Full text] [doi: 10.1161/STR.0000000000000158] [Medline: 29367334]
4. Langhorne P. The Stroke unit story: where have we been and where are we going? Cerebrovasc Dis 2021;50(6):636-643 [FREE Full text] [doi: 10.1159/000518934] [Medline: 34547746]
5. Lyden P. Using the national institutes of health stroke scale: a cautionary tale. Stroke 2017;48(2):513-519. [doi: 10.1161/STROKEAHA.116.015434] [Medline: 28077454]
6. National Institute of Neurological DisordersStroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. N Engl J Med 1995;333(24):1581-1587. [doi: 10.1056/NEJM199512143332401] [Medline: 7477192]
7. Hinkle JL. Reliability and validity of the national institutes of health stroke scale for neuroscience nurses. Stroke 2014;45(3):e32-e34. [doi: 10.1161/STROKEAHA.113.004243] [Medline: 24496393]
8. Dewey HM, Donnan GA, Freeman EJ, Sharples CM, Macdonell RA, McNeil JJ, et al. Interrater reliability of the national institutes of health stroke scale: rating by neurologists and nurses in a community-based stroke incidence study. Cerebrovasc Dis 1999;9(6):323-327. [doi: 10.1159/000016006] [Medline: 10545689]
9. Schmülling S, Grond M, Rudolf J, Kiencke P. Training as a prerequisite for reliable use of NIH Stroke Scale. Stroke 1998;29(6):1258-1259. [doi: 10.1161/01.str.29.6.1258] [Medline: 9626306]
10. André C. The NIH stroke scale is unreliable in untrained hands. J Stroke Cerebrovasc Dis 2002;11(1):43-46. [doi: 10.1053/jscd.2002.123974] [Medline: 17903854]
11. Lyden P, Raman R, Liu L, Grotta J, Broderick J, Olson S, et al. NIHSS training and certification using a new digital video disk is reliable. Stroke 2005;36(11):2446-2449. [doi: 10.1161/01.STR.0000185725.42768.92] [Medline: 16224093]
12. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, et al. Improved reliability of the NIH stroke scale using video training. NINDS TPA stroke study group. Stroke 1994;25(11):2220-2226. [doi: 10.1161/01.str.25.11.2220] [Medline: 7974549]

XSL•FO
RenderX

13. Lyden P. NIH Stroke Scale Training-Part 2-Basic Instruction. URL: https://www.ahajournals.org/doi/10.1161/strokeaha.116.015434 [accessed 2024-12-13]
14. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, et al. Improved reliability of the NIH stroke scale using video training. NINDS TPA stroke study group. Stroke 1994;25(11):2220-2226. [doi: 10.1161/01.str.25.11.2220] [Medline: 7974549]
15. Critères de qualité pour la certification des stroke centers. Swiss Federation of Clinical Neuro-Societies. URL: https://sfcns.ch/certification/stroke/re-certification [accessed 2024-01-30]
16. Décision concernant la planification de la médecine hautement spécialisée (MHS) dans le domaine du traitement des accidents vasculaires cérébraux. Organe de décision MHS. 2011. URL: https://www.gdk-cds.ch/fileadmin/docs/public/gdk/themen/hsm/hsm_spitalliste/bb_dc_stroke_praez_20110520_def_f.pdf [accessed 2024-02-21]
17. Lyrer PA, Engelter S, Gralla J, Andrea Humm PD, Fandino J, Fischer U, Arbeitsgruppen Stroke Unit. Stroke Units et Stroke Centers en Suisse. Swiss Med Forum 2024. [doi: 10.4414/fms.2024.1343858063]
18. Kim KJ, Kim G. Development of e-learning in medical education: 10 years' experience of Korean medical schools. Korean J Med Educ 2019;31(3):205-214 [FREE Full text] [doi: 10.3946/kjme.2019.131] [Medline: 31455050]
19. Suppan M, Abbas M, Catho G, Stuby L, Regard S, Achab S, et al. Impact of a serious game (Escape COVID-19) on the intention to change COVID-19 control practices among employees of long-term care facilities: web-based randomized controlled trial. J Med Internet Res 2021;23(3):e27443 [FREE Full text] [doi: 10.2196/27443] [Medline: 33685854]
20. Suppan L, Abbas M, Stuby L, Cottet P, Larribau R, Golay E, et al. Effect of an E-learning module on personal protective equipment proficiency among prehospital personnel: web-based randomized controlled trial. J Med Internet Res 2020;22(8):e21265 [FREE Full text] [doi: 10.2196/21265] [Medline: 32747329]
21. Stuby L, Koka A, Suppan L, Suppan M. NIHSS E-learning module 21c. URL: https://nihss-study.ch/e-learning [accessed 2023-11-12]
22. Suppan M, Stuby L, Carrera E, Cottet P, Koka A, Assal F, et al. Asynchronous distance learning of the national institutes of health stroke scale during the covid-19 pandemic (E-Learning vs Video): randomized controlled trial. J Med Internet Res 2021;23(1):e23594 [FREE Full text] [doi: 10.2196/23594] [Medline: 33428581]
23. Koka A, Suppan L, Cottet P, Carrera E, Stuby L, Suppan M. Teaching the national institutes of health stroke scale to paramedics (E-Learning vs Video): randomized controlled trial. J Med Internet Res 2020;22(6):e18358 [FREE Full text] [doi: 10.2196/18358] [Medline: 32299792]
24. Koka A, Suppan M, Carrera E, Fraga-Freijeiro P, Massuk K, Imbeault M, et al. Knowledge retention of the NIH stroke scale among stroke unit health care workers using video vs. e-learning: protocol for a web-based, randomized controlled trial. Healthcare (Basel) 2021;9(11):1460 [FREE Full text] [doi: 10.3390/healthcare9111460] [Medline: 34828505]
25. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
26. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Perrin Franck C, Babington-Ashaye A, Dietrich D, Bediang G, Veltsos P, Gupta PP, et al. iCHECK-DH: guidelines and checklist for the reporting on digital health implementations. J Med Internet Res 2023;25:e46694 [FREE Full text] [doi: 10.2196/46694] [Medline: 37163336]
28. ICMJE | About ICMJE | Clinical Trials Registration. URL: https://icmje.org/about-icmje/faqs/clinical-trials-registration/ [accessed 2024-11-13]
29. Community surveys pro. Corejoomla. 2020. URL: https://extensions.joomla.org/extension/community-surveys/ [accessed 2024-12-13]
30. Suppan M, Stuby L, Koka A, Suppan L. NIHSS Video subtitled in french. 2019. URL: https://nihss-study.ch/video [accessed 2021-06-21]
31. Kamuche FU, Ledman RE. Relationship of time and learning retention. J. Coll. Teach. Learn. TLC 2005;2(8):25. [doi: 10.19030/tlc.v2i8.1851]
32. Zou X, Liu S, Li J, Chen W, Ye J, Yang Y, et al. Factors associated with healthcare workers' insomnia symptoms and fatigue in the fight against covid-19, and the role of organizational support. Front Psychiatry 2021;12:652717 [FREE Full text] [doi: 10.3389/fpsyt.2021.652717] [Medline: 33841214]
33. Lyden P, Raman R, Liu L, Emr M, Warren M, Marler J. National institutes of health stroke scale certification is reliable across multiple venues. Stroke 2009;40(7):2507-2511 [FREE Full text] [doi: 10.1161/STROKEAHA.108.532069] [Medline: 19520998]
34. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public - a systematic review and meta-analysis. Psychiatry Res 2020;291:113190 [FREE Full text] [doi: 10.1016/j.psychres.2020.113190] [Medline: 32563745]
35. Raso R, Fitzpatrick JJ, Masick K. Nurses' intent to leave their position and the profession during the COVID-19 pandemic. J Nurs Adm 2021;51(10):488-494. [doi: 10.1097/NNA.0000000000001052] [Medline: 34519700]

XSL•FO
RenderX

36. NIHSS Retention Study Data | Yareta. 2024. URL: https://yareta.unige.ch/archives/c217b916-d1e7-4d65-ba19-72d04cd5a730 [accessed 2024-09-09]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet e-Surveys
CONSORT: Consolidated Standards of Reporting Trials
HCP: health care professional
iCHECK-DH: Guidelines and Checklist for Reporting on Digital Health Implementations
NIHSS: National Institutes of Health Stroke Scale

Edited by N Cahill; submitted 11.06.24; peer-reviewed by RT Pinzon, L Yang; comments to author 23.07.24; revised version received 10.09.24; accepted 09.12.24; published 04.03.25.

Please cite as:
Koka A, Stuby L, Carrera E, Gabr A, O'Connor M, Missilier Peruzzo N, Waeterloot O, Medlin F, Rigolet F, Schmutz T, Michel P, Desmettre T, Suppan M, Suppan L
Asynchronous Distance Learning Performance and Knowledge Retention of the National Institutes of Health Stroke Scale Among Health Care Professionals Using Video or e-Learning: Web-based Randomized Controlled Trial
J Med Internet Res 2025;27:e63136
URL: https://www.jmir.org/2025/1/e63136
doi:10.2196/63136
PMID:40053772

©Avinash Koka, Loric Stuby, Emmanuel Carrera, Ahmed Gabr, Margaret O'Connor, Nathalie Missilier Peruzzo, Olivier Waeterloot, Friedrich Medlin, Fabien Rigolet, Thomas Schmutz, Patrik Michel, Thibaut Desmettre, Mélanie Suppan, Laurent Suppan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Gaming and Subsequent Health and Well-Being Among Older Adults: Longitudinal Outcome-Wide Analysis

Atsushi Nakagomi[1], MD, PhD; Kazushige Ide[1], PhD; Katsunori Kondo[1], MD, PhD; Koichiro Shiba[2], PhD

[1]Department of Social Preventive Medical Sciences, Center for Preventive Medical Sciences, Chiba University, Chiba, Japan

[2]Department of Epidemiology, Boston University School of Public Health, Boston University, Boston, MA, United States

Corresponding Author:
Atsushi Nakagomi, MD, PhD
Department of Social Preventive Medical Sciences
Center for Preventive Medical Sciences
Chiba University
1-33 Yayoicho, Inage-ku
Chiba, 2638522
Japan
Phone: 81 432903177 ext 3177
Email: anakagomi0211@gmail.com

Abstract

Background: Digital gaming has become increasingly popular among older adults, potentially offering cognitive, social, and physical benefits. However, its broader impact on health and well-being, particularly in real-world settings, remains unclear.

Objective: This study aimed to evaluate the multidimensional effects of digital gaming on health and well-being among older adults, using data from the Japan Gerontological Evaluation Study conducted in Matsudo City, Chiba, Japan.

Methods: Data were drawn from 3 survey waves (2020 prebaseline, 2021 baseline, and 2022 follow-up) of the Japan Gerontological Evaluation Study, which targets functionally independent older adults. The exposure variable, digital gaming, was defined as regular video game play and was assessed in 2021. In total, 18 outcomes across 6 domains were evaluated in 2022; domain 1—happiness and life satisfaction, domain 2—physical and mental health, domain 3—meaning and purpose, domain 4—character and virtue, domain 5—close social relationships, and domain 6—health behavior. Furthermore, 10 items from the Human Flourishing Index were included in domains 1-5, with 2 items for each domain. Overall flourishing was defined as the average of the means across these 5 domains. In addition, 7 items related to domains 2, 5, and 6 were assessed. The final sample consisted of 2504 participants aged 65 years or older, with questionnaires containing the Human Flourishing Index randomly distributed to approximately half of the respondents (submodule: n=1243). Consequently, we used 2 datasets for analysis. We applied targeted maximum likelihood estimation to estimate the population average treatment effects, with Bonferroni correction used to adjust for multiple testing.

Results: Digital gaming was not significantly associated with overall flourishing or with any of the 5 domains from the Human Flourishing Index. Although initial analyses indicated associations between digital gaming and participation in hobby groups (mean difference=0.12, *P*=.005) as well as meeting with friends (mean difference=0.076, *P*=.02), these associations did not remain significant after applying the Bonferroni correction for multiple testing. In addition, digital gaming was not associated with increased sedentary behavior or reduced outdoor activities.

Conclusions: This study provides valuable insights into the impact of digital gaming on the health and well-being of older adults in a real-world context. Although digital gaming did not show a significant association with improvements in flourishing or in the individual items across the 5 domains, it was also not associated with increased sedentary behavior or reduced outdoor activities. These findings suggest that digital gaming can be part of a balanced lifestyle for older adults, offering opportunities for social engagement, particularly through hobby groups. Considering the solitary nature of gaming, promoting social gaming opportunities may be a promising approach to enhance the positive effects of digital gaming on well-being.

(J Med Internet Res 2025;27:e69080) doi:10.2196/69080

KEYWORDS

digital gaming; older adults; flourishing; well-being; physical activity; social engagement; mobile phone

XSL•FO
RenderX

Introduction

Digital Gaming for Health

Digital gaming has become increasingly common among older adults. In 2021, 10.7% of people aged 65 years or older in Japan played digital games for hobbies and entertainment [1]. This trend illustrates the growing appeal of digital gaming among older adults, which holds the potential for promoting both physical and cognitive health [2,3]. A systematic review, for example, has demonstrated that video game–based interventions, such as exergames, can support physical health in older adults by improving balance, mobility, strength, physical fitness, and walking performance [3]. In addition, there is a growing body of evidence that computer games with and without physical components may have positive changes in cognitive function including enhanced executive function, memory, and focused attention [4-6]. Such evidence suggests that digital gaming could serve as a valuable tool for supporting active aging and helping older adults maintain physical and cognitive vitality while providing an enjoyable and engaging activity.

Limitations of Current Evidence

While existing research has provided valuable insights into the health impacts of digital gaming, several limitations remain. First, although digital games can enhance the lives of older adults in various ways [7], current evidence is largely focused on physical and cognitive health outcomes [3-6,8,9]. The World Health Organization defines health as “a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity” [10]. This definition highlights the complexity and breadth of health beyond the absence of illness. In recent years, there has been growing recognition of the multidimensional nature of health, leading to increased attention on the concept of human flourishing. Human flourishing refers to “a state in which all aspects of a person’s life are good” and includes five core domains, which are (1) happiness and life satisfaction, (2) physical and mental health, (3) meaning and purpose, (4) character and virtue, and (5) close social relationships [11]. Growing evidence suggests that games can facilitate engaging conversations and serve as a shared activity, helping to promote bridging social capital, foster civic participation, and enhance satisfaction [12,13]. However, the broader social and psychological impacts of digital gaming among older adults remain underexplored [14-16].

Second, much of the existing evidence comes from studies that examine a single or limited number of outcomes, making it difficult to fully understand the range of digital gaming’s effects. The impact of gaming on older adults is inherently multidimensional—potentially beneficial for some outcomes but less so for others. For example, while certain games may encourage physical activity, gaming could also contribute to a more sedentary lifestyle if it predominantly occurs indoors. To gain a holistic understanding of digital gaming’s effects, comprehensive, large-scale studies that assess a wide range of health and well-being outcomes simultaneously are needed.

Third, many previous studies have been conducted in controlled settings, such as short-term exergame interventions lasting 3 months. While these studies provide valuable evidence about the effects of specific gaming interventions, they do not capture the long-term impact of daily gaming activities on health and well-being. To address this gap, studies using well-powered real-world data from community-dwelling older adults are necessary. Such research can offer deeper insights into how regular, everyday digital gaming interacts with the long-term health trajectories of older adults.

The Aim of This Study

This study aimed to expand the current body of evidence on the multidimensional impact of digital gaming on the human flourishing of older adults. As digital gaming represents a novel and increasingly popular activity among older adults, understanding its associations with holistic health outcomes—including human flourishing—can provide valuable insights into how such activities contribute to their overall well-being and support active aging. We used an outcome-wide analytic approach to simultaneously examine the longitudinal associations between digital gaming and 18 subsequent health and well-being outcomes among Japanese older adults. We examined outcomes across domains of happiness and life satisfaction, physical and mental health, meaning and purpose, character and virtue, social well-being, and health behaviors [17].

Methods

Data Sources

We used data from the 2020 (prebaseline), 2021 (baseline), and 2022 (follow-up) waves of the Japan Gerontological Evaluation Study (JAGES), a nationwide survey of adults aged 65 years or older who were not receiving public long-term care insurance benefits in Japan [18]. Our analysis focused on data from participants from one of the JAGES study sites—Matsudo City, as the questionnaire on digital gaming was included only there. Matsudo City is located in Chiba Prefecture, Japan, just northeast of Tokyo. It serves as a suburban area within the Greater Tokyo Area, with a population of around 500,000. Self-administered questionnaires were distributed in 15 districts through random sampling. The study included respondents aged 65 years or older in 2020 who completed the 2020 survey (n=5347; response rate 82.6%). Respondents with inconsistently reported age or gender, as well as those with implausible height (<100 cm or >200 cm) or weight (<30 kg or >100 kg) values, were excluded. Of the 2020 respondents, 73.7% (3941/5060) responded to the 2021 survey. Of the 2021 respondents, 63.8% (2515/3941) responded to the 2022 survey. After excluding 11 participants with inconsistently reported ages, the analytic sample consisted of 2504 participants. For the assessment of human flourishing, the questionnaire version that included the Human Flourishing Index was distributed only to randomly selected half of the respondents (submodule: n=1243). Thus, we used 2 analytic samples—the submodule respondents for flourishing assessments (n=1243) and the full analytic sample for all other outcomes (n=2504). Figure 1 and Figure S1 in Multimedia Appendix 1 provide a detailed flowchart of the data linkage and sample selection process. The participants of the JAGES were informed that their participation was voluntary,

XSL•FO
RenderX

and returning the questionnaire implied their consent to participate in the study.

Figure 1. Flowchart of study participants.

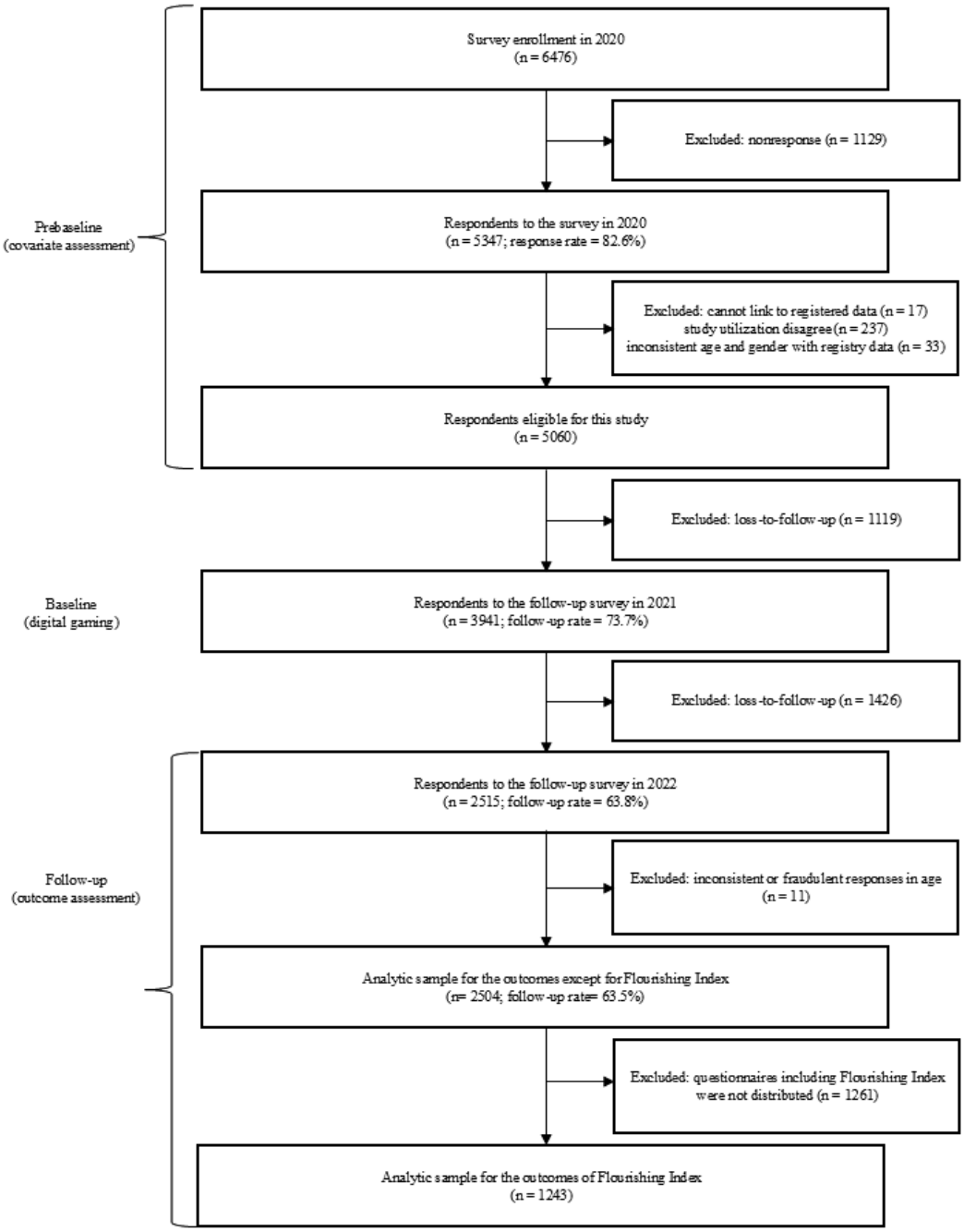

Measures

We measured the covariates, exposures, and outcomes in a temporal sequence using the 3-wave panel structure of the data (2020, 2021, and 2022). This structure allowed us to account for pre-exposure covariates, including previous values of the outcomes before exposure to digital gaming, which helps mitigate the risks of confounding and reverse causality without the need to adjust for potential mediators [17].

Exposure Variable

The exposure variable, digital gaming, was obtained from the 2021 survey. Participants were asked, “Have you played video games with a PC, cell phones, smartphones, tablets, or consoles in the last five years?” The response options included “Playing regularly,” “Have played,” “Know video games, but have not played,” and “Do not know video games.” For analysis, we created a binary variable, where participants who answered “Playing regularly” were coded as 1, indicating regular digital gaming, while all other responses were coded as 0.

Outcomes

We examined 18 outcomes in 2022 across 6 domains of health and well-being, following VanderWeele’s human flourishing framework and a previous outcome-wide study using the JAGES data [11,19]. We first assessed human flourishing based on the Human Flourishing Index. Human flourishing refers to “a state in which all aspects of a person’s life are good,” and the Human

XSL•FO
RenderX

Flourishing Index is a comprehensive measure designed to capture multidimensional aspects of health and well-being, encompassing several domains [11]. This index consists of 10 self-reported items measuring human well-being outcomes across 5 domains (ie, 2 items per domain). Domain 1 consists of happiness and life satisfaction, domain 2 consists of physical and mental health, domain 3 consists of meaning and purpose, domain 4 consists of character and virtue, and domain 5 consists of close relationships. For each domain, we calculated the mean of the responses to 2 domain-specific items, and overall flourishing was defined as the average of these means across the 5 domains.

In addition to these outcomes, we also examined other outcomes based on items available in the JAGES survey. For domain 2, we assessed instrumental activities of daily living (IADL), forgetfulness, and depressive symptoms using the Geriatric Depression Scale. IADL refers to tasks that enable an individual to live independently within their community, such as managing finances, preparing meals, and handling transportation. For domain 3, we assessed ikigai, broadly defined as "what makes life worth living," a well-accepted psychological concept in Japanese culture [20]. For domain 5, we assessed participation in hobby groups, participation in sports groups, meeting with friends, and the number of friends. We also examined several health behaviors as outcomes (domain 6: health behavior), including a sedentary lifestyle, smoking, drinking, and frequency of going out. Detailed information on the definitions and measures of these outcomes is shown in Table S1 in Multimedia Appendix 1.

Covariates

All covariates were obtained from the 2020 survey, conducted a year before the wave in which the exposure levels were assessed.

As confounders in this study, we included several sociodemographic factors, such as age, gender (men or women), years of education (6-9 years vs 10 or more years), self-rated deprivation, marital status (married or not), living arrangement (living alone or not), and employment status (employed or not).

To minimize the potential for reverse causation, we adjusted for previous values of outcomes wherever available. These included pre-exposure levels of happiness, life satisfaction (as a binary variable), self-rated health, IADL, forgetfulness, Geriatric Depression Scale, participation in hobby groups, participation in sports groups, meeting with friends, number of friends, sedentary behavior, smoking, alcohol consumption, and frequency of going out.

Statistical Analysis

This study adopted an outcome-wide analytic approach, which explores the associations between a single exposure and multiple outcomes simultaneously [17,21]. This method allows for a comprehensive assessment of the impact of exposure across a wide range of outcomes. In addition, this approach offers methodological advantages, such as being less susceptible to *P*-hacking and publication bias. It is particularly beneficial for studying the impact of digital gaming, as it captures the multidimensional effects across various domains of health and well-being, providing a holistic understanding of its potential benefits and risks. By examining multiple outcomes within the same framework, this method helps identify patterns and trade-offs that may not be evident when focusing on individual outcomes.

We estimated the population average treatment effects of video gaming on each of the 18 outcomes across the 6 domains, using the doubly robust targeted maximum likelihood estimation. The "doubly robust" method is a statistical approach that combines 2 models—one for the exposure (eg, gaming behavior) and one for the outcome (eg, health or well-being measure). It is termed "doubly robust" because it provides consistent estimates of the average treatment effect as long as at least 1 of these 2 models is correctly specified when the causal identification assumptions are met. Thus, this method is useful as it enhances the robustness of effect estimation by reducing potential biases from model misspecification, compared with methods that rely on a single model (eg, outcome regression and propensity score weighting). To further mitigate the potential bias due to model misspecification, both the exposure and outcome models were fitted data-adaptively using the SuperLearner, an ensemble method that combines weighted estimates from multiple candidate algorithms, including generalized linear models, gradient-boosting machines, and neural networks [22]. The "Super Learner" algorithm is a machine learning technique that data-adaptively selects the optimal weighted combination of estimators from a set of candidate algorithms to improve predictive accuracy. By synthesizing estimates from multiple algorithms, it further reduces the risk of bias due to model misspecification. The targeted maximum likelihood estimation and Super Learning procedures were performed using the ltmle and SuperLearner packages in R (R Foundation for Statistical Computing). For continuous variables, we estimated the difference in the mean levels of the outcomes. All continuous outcomes were standardized (mean 0, SD 1) to ensure that the effect estimates could be interpreted as SD changes in the mean outcomes. For binary outcomes, we estimated risk ratios. To address the issue of multiple testing, we applied the Bonferroni correction that used a highly conservative *P* value threshold for statistical significance of P=.00278 (.05/18).

As a sensitivity analysis, we calculated the e-values for each exposure-outcome association to assess the robustness of the estimated associations against unmeasured confounding [23]. E-values quantify the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and outcome, conditional on the adjusted covariates, for the confounding bias to account for the observed associations.

To address potential bias due to missing data, we used random forest imputation via the R package "missRanger" to impute missing values. All analyses were performed using R (version 4.3.1).

Ethical Considerations

Ethical approval for the study was obtained from the Ethics Committee at Chiba University (approval M10460). The participants of the JAGES were informed that their participation

XSL•FO
RenderX

was voluntary and returning the questionnaire implied their consent to participate in the study.

Results

Table 1 shows the prebaseline characteristics of the submodule sample by the baseline digital gaming. Among the 1243 respondents, 192 (15.45%) and 380 (15.18%) played digital games, respectively. Individuals who played digital games tended to be younger men, have higher educational attainment, experience less deprivation, be married, live with others, and be employed. These trends were consistent between the submodule sample and the full analytic sample (Table S2 in Multimedia Appendix 1; n=2504).

Table 2 shows the estimated effects of digital gaming on overall flourishing and each of its 5 domains based on the Human Flourishing Index. We found no evidence of the association between digital gaming and these outcomes.

Table 3 shows the estimated effects of digital gaming on other outcomes related to human flourishing and health behaviors. Digital gaming was associated with more frequent participation in hobby groups (mean difference=0.124, *P*=.005) and meeting friends (mean difference=0.076, *P*=.02). However, these associations were above the α=.05 cutoff after applying the Bonferroni correction for multiple testing.

Tables S3 and S4 in Multimedia Appendix 1 present E-values, which indicate the robustness of the observed associations between digital gaming and subsequent health and well-being outcomes against potential unmeasured confounders. For example, in the association between digital gaming and participation in hobby groups, an unmeasured confounder would need to have a risk ratio of at least 1.48 with both the exposure and the outcome, conditional on the measured covariates, to fully account for the observed association. In addition, a risk ratio of 1.22 would be required to shift the CI to include the null value.

XSL•FO
RenderX

Table 1. Baseline characteristics of participants who responded to the submodule questionnaires.

Prebaseline characteristics	Digital gaming		
	Total (N=1243)	No (n=1051)	Yes (n=192)
Age (years), mean (SD)	75.7 (5.65)	76.0 (5.68)	74.3 (5.27)
Gender, n (%)			
Men	617 (49.6)	509 (48.4)	108 (56.3)
Women	626 (50.4)	542 (51.6)	84 (43.8)
Education, n (%)			
High education	1035 (83.3)	865 (82.3)	170 (88.5)
Low education	208 (16.7)	186 (17.7)	22 (11.5)
Deprivation, n (%)			
No deprivation	1068 (85.9)	892 (84.9)	176 (91.7)
Deprivation	175 (14.1)	159 (15.1)	16 (8.3)
Marital status, n (%)			
Married	913 (73.5)	754 (71.7)	159 (82.8)
Not married	330 (26.5)	297 (28.3)	33 (17.2)
Living alone, n (%)			
Living with someone	1026 (82.5)	853 (81.2)	173 (90.1)
Living alone	217 (17.5)	198 (18.8)	19 (9.9)
Employment status, n (%)			
Employed	309 (24.9)	255 (24.3)	54 (28.1)
Not employed	934 (75.1)	796 (75.7)	138 (71.9)
Happiness, mean (SD)	7.57 (1.75)	7.56 (1.80)	7.62 (1.49)
Life satisfaction, n (%)			
No	199 (16)	172 (16.4)	27 (14.1)
Yes	1044 (84)	879 (83.6)	165 (85.9)
Self-rated health, n (%)			
Poor	126 (10.1)	110 (10.5)	16 (8.3)
Good	1117 (89.9)	941 (89.5)	176 (91.7)
Instrumental activities of daily living, mean (SD)	4.94 (0.355)	4.93 (0.379)	4.97 (0.174)
Forgetfulness, n (%)			
No	1143 (92)	966 (91.9)	177 (92.2)
Yes	100 (8)	85 (8.1)	15 (7.8)
Depressive symptoms, mean (SD)	2.76 (2.80)	2.82 (2.84)	2.42 (2.55)
Participation in hobby groups, mean (SD)	1.93 (1.39)	1.92 (1.38)	1.98 (1.45)
Participation in sports groups, mean (SD)	2.16 (1.77)	2.11 (1.74)	2.44 (1.91)
Meeting friends, mean (SD)	3.20 (1.65)	3.17 (1.64)	3.37 (1.71)
Number of friends, mean (SD)	2.85 (1.42)	2.81 (1.41)	3.05 (1.45)
Sedentary lifestyle, n (%)			
No	1171 (94.2)	991 (94.3)	180 (93.8)
Yes	72 (5.8)	60 (5.7)	12 (6.3)
Smoking, n (%)			
No	1143 (92)	972 (92.5)	171 (89.1)
Yes	100 (8)	79 (7.5)	21 (10.9)

XSL•FO
RenderX

Prebaseline characteristics	Digital gaming		
	Total (N=1243)	No (n=1051)	Yes (n=192)
Alcohol consumption, n (%)			
No	672 (54.1)	577 (54.9)	95 (49.5)
Yes	571 (45.9)	474 (45.1)	97 (50.5)
Going out, mean (SD)	6.12 (1.09)	6.08 (1.11)	6.31 (0.969)

Table 2. Digital gaming and subsequent flourishing.

Outcomes	Estimate (95% CI)	*P* value
Flourish (mean across the domains 1-5)	–0.001 (–0.107 to 0.105)	.98
Domain 1 (happiness and life satisfaction)	–0.082 (–0.183 to 0.020)	.12
Domain 2 (mental and physical health)	0.000 (–0.136 to 0.136)	>.99
Domain 3 (meaning and purpose)	–0.001 (–0.119 to 0.117)	.99
Domain 4 (character and virtue)	0.033 (–0.117 to 0.184)	.67
Domain 5 (close social relationship)	0.037 (–0.092 to 0.165)	.58

Table 3. Digital gaming and subsequent health and well-being.

Outcomes	Reference (variable type)	Risk ratio or mean difference	
		Estimate (95% CI)	*P* value[a]
Domain 2 mental and physical health			
Instrumental activities of daily living	0.00 (continuous)	0.067 (–0.020 to 0.153)	.13
Forgetfulness	1.00 (binary)	1.175 (0.898 to 1.539)	.24
Depressive symptoms	0.00 (continuous)	0.027 (–0.039 to 0.093)	.42
Domain 3 meaning and purpose			
Ikigai	0.00 (continuous)	–0.092 (–0.199 to 0.016)	.09
Domain 5 close social relationship			
Participation in hobby groups	0.00 (continuous)	0.124 (0.037 to 0.210)	.005[b]
Participation in sports groups	0.00 (continuous)	0.011 (–0.053 to 0.075)	.74
Meeting with friends	0.00 (continuous)	0.076 (0.010 to 0.142)	.02[c]
Number of friends	0.00 (continuous)	0.015 (–0.061 to 0.091)	.70
Domain 6 health behavior			
Sedentary lifestyle	1.00 (binary)	1.055 (0.788 to 1.411)	.72
Smoking	1.00 (binary)	0.991 (0.888 to 1.105)	.87
Drinking	1.00 (binary)	1.047 (0.989 to 1.109)	.12
Going out	0.00 (continuous)	0.026 (–0.081 to 0.133)	.64

[a]*P*<.05 after Bonferroni correction (the *P* value cutoff for Bonferroni correction is *P*=.05/18 outcomes [*P*<.00278]).

[b]*P*<.01 before Bonferroni correction.

[c]*P*<.05 before Bonferroni correction.

Discussion

Principal Findings

The main findings of this study were threefold. First, video gaming was associated with more frequent participation in hobby groups and meeting with friends, although these results should be interpreted with caution as they may be false positive findings due to multiple testing. Second, video gaming was not associated with lower levels of physical activity, as indicated by measures of sedentary lifestyle and frequency of going out. Third, video gaming did not appear to be associated with the flourishing outcomes across the 5 domains, including psychological well-being.

XSL•FO
RenderX

Our findings highlight the potential of digital gaming to positively impact the social well-being of older adults, contributing to meaningful social interactions and engagement. Previous studies have shown that older adults who play digital games experience social benefits, such as reduced loneliness and improved social functioning [15,24]. However, these studies often relied on smaller samples or qualitative methods, limiting the generalizability of their findings. In contrast, our study, based on a larger sample size, provides more robust evidence to support the social benefits of digital gaming.

Digital gaming may enhance social well-being by fostering participation in hobby groups, where gaming can act as a shared interest and a catalyst for engaging conversations. These interactions create opportunities for older adults to build and strengthen social connections, thereby promoting a sense of belonging and community. This underscores the role of digital gaming as not only a recreational activity but also a potential tool for enhancing social engagement and overall well-being among older adults.

Digital gaming, particularly nonphysical types, may reduce physical activity levels among older adults by promoting sedentary behavior, as players often remain seated for extended periods. This sedentary time could potentially replace activities like walking or outdoor socializing. In contrast, exergames—digital games that require physical movement—have been shown to promote physical activity by improving mobility, balance, and fitness levels [2]. Despite concerns about sedentary behavior, our study found no evidence in a real-world context that digital gaming was associated with increased sedentary behavior or decreased outdoor activities among older adults in Japan. This finding suggests that, for many older adults, digital gaming can integrate into a balanced lifestyle without negatively affecting physical activity levels.

Digital gaming may also enhance psychological well-being among older adults by providing relaxation and social interaction [24,25], although we did not find evidence of such improvements in our study. Research indicates that the social aspects of gaming, such as playing together or discussing games with family and friends, are crucial for translating digital gaming experiences into positive well-being outcomes [15,26]. However, many older adults tend to play digital games alone, which can limit the potential social benefits that gaming might otherwise offer. This highlights the need for targeted initiatives to promote and encourage more social forms of digital gaming among older. Facilitating environments where digital games are played together, such as community centers or online platforms designed for older users, could help harness the social benefits of gaming and, in turn, improve their overall well-being.

This study has several limitations. First, the broad definition of gaming behavior used in this study may not adequately reflect the specific characteristics and impacts of individual games. The diversity of digital games, including their types, purposes, and contexts, means that our analysis may have attenuated potential associations. In addition, whether participants played alone or with others could significantly alter the associations. Future research should adopt a more nuanced approach by categorizing gaming behavior to better capture the specific effects of different game types.

Second, the generalizability of our findings is limited, as our data were derived from a single city in Japan. Results may differ in other regions or countries with different cultural contexts and gaming habits.

Third, due to the observational nature of this study, the possibility of unmeasured confounding and reverse causation cannot be completely ruled out. However, to mitigate this issue, we adjusted for numerous covariates, including confounders and prebaseline outcomes, as recommended in the outcome-wide approach [17,21]. Fourth, while our study is larger than previous studies, it is still likely underpowered (eg, n=1243 with an exposure prevalence of approximately 15%). Future studies with larger sample sizes are warranted to provide more robust conclusions.

Fifth, the outcome-wide analytic approach used in this study was designed to provide a comprehensive overview of associations across multiple outcomes; however, it does indeed sacrifice depth for breadth due to the word count limit. The traditional design focusing on a single or a narrow set of outcomes has other limitations (eg, not providing holistic evidence on gaming-health associations, *P*-hacking, and publication bias), which the outcome-wide design seeks to address, at least to some extent. We think this is a trade-off, and the outcome-wide design (which contains more breadth) and traditional study design (which contains more in-depth discussion) are both complementary and play different roles in the advancement of science. Future research on specific outcomes with more detailed analysis (eg, effect heterogeneity across subgroups) is needed.

Finally, our study did not account for the potential overlap between digital gaming and social media usage. Gamers are more likely to use social media, and the impacts of gaming may overlap with those of social media usage. This overlap is particularly relevant as many digital games include social interaction features that resemble social media functions, such as in-game chats or multiplayer platforms. Consequently, some observed associations may reflect the combined effects of gaming and social media rather than gaming alone. Future studies should measure and differentiate these activities to clarify their unique and overlapping impacts on health and well-being.

Conclusion

While digital gaming did not show a significant association with improvements in flourishing or its subdomains, our results indicate that digital gaming could lead to greater social engagement without increasing sedentary behavior and reducing outdoor activities. These results suggest that digital gaming can be integrated into a balanced lifestyle, offering a potential tool for enhancing social well-being in older adults. Furthermore, integrating digital gaming into community and health care settings presents novel opportunities for fostering social connections and supporting healthy aging in this population.

XSL•FO
RenderX

Acknowledgments

This study was supported by the Foundation for the Fusion of Science and Technology.

JAGES was supported by Grant-in-Aid for Scientific Research (20H00557, 20K10540, 21H03153, 21H03196, 21K17302, 22H00934, 22H03299, 22K04450, 22K13558, 22K17409, 23H00449, 23H03117) from JSPS (Japan Society for the Promotion of Science), Health Labour Sciences Research Grants (19FA1012, 19FA2001, 21FA1012, 22FA2001, 22FA1010, 22FG2001), Research Institute of Science and Technology for Society (JPMJOP1831) from the Japan Science and Technology (JST), a grant from Japan Health Promotion & Fitness Foundation, TMDU priority research areas grant, and National Research Institute for Earth Science and Disaster Resilience. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the respective funding organizations.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 124 KB - jmir_v27i1e69080_app1.docx]

References

1. Survey on time use and leisure activities. Statistics Bureau of Japan. 2021. URL: https://www.stat.go.jp/english/data/shakai/index.html [accessed 2025-10-01]
2. Howes SC, Charles DK, Marley J, Pedlow K, McDonough SM. Gaming for health: systematic review and meta-analysis of the physical and cognitive effects of active computer gaming in older adults. Phys Ther 2017;97(12):1122-1137. [doi: 10.1093/ptj/pzx088] [Medline: 29077911]
3. Xu W, Liang H, Baghaei N, Wu Berberich B, Yue Y. Health benefits of digital videogames for the aging population: a systematic review. Games Health J 2020;9(6):389-404. [doi: 10.1089/g4h.2019.0130] [Medline: 32589482]
4. Rosenberg D, Depp CA, Vahia IV, Reichstadt J, Palmer BW, Kerr J, et al. Exergames for subsyndromal depression in older adults: a pilot study of a novel intervention. Am J Geriatr Psychiatry 2010;18(3):221-226 [FREE Full text] [doi: 10.1097/JGP.0b013e3181c534b5] [Medline: 20173423]
5. Basak C, Boot WR, Voss MW, Kramer AF. Can training in a real-time strategy video game attenuate cognitive decline in older adults? Psychol Aging 2008;23(4):765-777 [FREE Full text] [doi: 10.1037/a0013494] [Medline: 19140648]
6. Peretz C, Korczyn AD, Shatil E, Aharonson V, Birnboim S, Giladi N. Computer-based, personalized cognitive training versus classical computer games: a randomized double-blind prospective trial of cognitive stimulation. Neuroepidemiology 2011;36(2):91-99. [doi: 10.1159/000323950] [Medline: 21311196]
7. Ijsselsteijn W, Nap HH, de Kort Y, Poels K. Digital game design for elderly users. 2007 Presented at: Proceedings of the 2007 Conference on Future Play; 2007 November 14; Toronto, Canada.
8. Yu RWL, Chan AHS. Meta-analysis of the effects of game types and devices on older adults-video game interaction: implications for video game training on cognition. Appl Ergon 2021;96:103477. [doi: 10.1016/j.apergo.2021.103477] [Medline: 34107433]
9. Zuo X, Tang Y, Chen Y, Zhou Z. Effects of electronic serious games on older adults with Alzheimer's disease and mild cognitive impairment: systematic review with meta-analysis of randomized controlled trials. JMIR Serious Games 2024;12:e55785 [FREE Full text] [doi: 10.2196/55785] [Medline: 39083796]
10. Constitution of the World Health Organization. Geneva, Switzerland: World Health Organization; 1946.
11. VanderWeele TJ. On the promotion of human flourishing. Proc Natl Acad Sci U S A 2017;114(31):8148-8156 [FREE Full text] [doi: 10.1073/pnas.1702996114] [Medline: 28705870]
12. Bell CS, Fain E, Daub J, Warren SH, Howell SH, Southard KS, et al. Effects of Nintendo Wii on quality of life, social relationships, and confidence to prevent falls. Phys Occup Ther Geriatr 2011;29(3):213-221. [doi: 10.3109/02703181.2011.559307]
13. Lee Y. Older adults' digital gameplay, social capital, social connectedness, and civic participation. Game Studies 2019;19(1) [FREE Full text]
14. Loos E, Kaufman D. Positive impact of exergaming on older adults' mental and social well-being: in search of evidence. 2018 Presented at: Human Aspects of IT for the Aged Population. Applications in Health, Assistance, and Entertainment:

XSL•FO
RenderX

4th International Conference, ITAP 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part II; 2018 July 15; Berlin, Heidelberg. [doi: 10.1007/978-3-319-92037-5_9]
15. Kaufman D, Sauve L. Digital gaming by older adults: can it enhance social connectedness? 2019 Presented at: Human Aspects of IT for the Aged Population. Social Media, Games and Assistive Environments: 5th International Conference, ITAP 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part II; 2019 July 26; Berlin, Heidelberg.
16. Goldstein JH, Cajko L, Oosterbroek M, Michielsen M, Van Houten O, Salverda F. Video games and the elderly. Social Behavior and Personality: An International Journal 1997;25(4):345-352. [doi: 10.2224/sbp.1997.25.4.345]
17. VanderWeele TJ, Mathur MB, Chen Y. Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. Statist. Sci 2020;35(3):437-466. [doi: 10.1214/19-sts728]
18. Kondo K. Progress in aging epidemiology in Japan: the JAGES project. J Epidemiol 2016;26(7):331-336 [FREE Full text] [doi: 10.2188/jea.JE20160093] [Medline: 27349200]
19. Nakagomi A, Tsuji T, Saito M, Ide K, Kondo K, Shiba K. Social isolation and subsequent health and well-being in older adults: a longitudinal outcome-wide analysis. Soc Sci Med 2023;327:115937. [doi: 10.1016/j.socscimed.2023.115937] [Medline: 37167920]
20. Mathews G. What makes life worth living?: How Japanese and Americans make sense of their worlds. Oakland, California: University of California Press; 1996.
21. VanderWeele TJ. Outcome-wide epidemiology. Epidemiology 2017;28(3):399-402 [FREE Full text] [doi: 10.1097/EDE.0000000000000641] [Medline: 28166102]
22. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. Am J Epidemiol 2017;185(1):65-73. [doi: 10.1093/aje/kww165] [Medline: 27941068]
23. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the e-value. Ann Intern Med 2017;167(4):268-274. [doi: 10.7326/M16-2607] [Medline: 28693043]
24. Allaire JC, McLaughlin AC, Trujillo A, Whitlock LA, LaPorte L, Gandy M. Successful aging through digital games: socioemotional differences between older adult gamers and non-gamers. Computers in Human Behavior 2013;29(4):1302-1306. [doi: 10.1016/j.chb.2013.01.014]
25. Zheng H, Li J, Salmon CT, Theng Y. The effects of exergames on emotional well-being of older adults. Comput Hum Behav 2020;110:106383. [doi: 10.1016/j.chb.2020.106383]
26. Osmanovic S, Pecchioni L. Beyond entertainment: motivations and outcomes of video game playing by older adults and their younger family members. Games Cult 2015;11(1-2):130-149. [doi: 10.1177/1555412015602819]

Abbreviations

JAGES: Japan Gerontological Evaluation Study
IADL: instrumental activities of daily living

Edited by A Mavragani; submitted 21.11.24; peer-reviewed by T Yamamoto, R Khanmohammadi; comments to author 05.12.24; revised version received 13.12.24; accepted 27.12.24; published 27.01.25.

Please cite as:
Nakagomi A, Ide K, Kondo K, Shiba K
Digital Gaming and Subsequent Health and Well-Being Among Older Adults: Longitudinal Outcome-Wide Analysis
J Med Internet Res 2025;27:e69080
URL: https://www.jmir.org/2025/1/e69080
doi:10.2196/69080
PMID:

©Atsushi Nakagomi, Kazushige Ide, Katsunori Kondo, Koichiro Shiba. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives

Bryan Cleal[1*], PhD; Yanbing Chen[2*], PhD; Mandy Wäldchen[3], MSc; Hanne Ballhausen[4], MSc; Drew Cooper[5,6,7], MSc; Shane O'Donnell[3], PhD; Christine Knoll[6], MD; Niklas Krug[7]; Klemens Raile[8], MD; Tebbe Ubben[4,9]; Adrian Tappe[4,9], BSc; Dana Lewis[3,10], BA; Ingrid Willaing[1,11], MPH; Timothy Skinner[12], PhD; Katarina Braune[6,7], MD

[1]Diabetes Management Research, Steno Diabetes Center Copenhagen, Herlev, Denmark

[2]Embry-Riddle Aeronautical University, Daytona Beach, FL, United States

[3]School of Sociology, University College Dublin, Belfield, Ireland

[4]#dedoc° Diabetes Online Community, Berlin, Germany

[5]Brain Simulation Section, Department of Neurology with Experimental Neurology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[6]Charité – Universitätsmedizin Berlin, Institute of Medical Informatics, Berlin, Germany

[7]Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany

[8]Department of Pediatrics, Vivantes Klinikum Neukölln, Berlin, Germany

[9]AndroidAPS, Hamilton, New Zealand

[10]OpenAPS, Seattle, WA, United States

[11]Department of Public Health, Copenhagen University, Copenhagen, Denmark

[12]Australian Centre for Behavioural Research in Diabetes, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Bryan Cleal, PhD
Diabetes Management Research
Steno Diabetes Center Copenhagen
Borgmester Ib Juuls Vej 83
Herlev, 2730
Denmark
Phone: 45 39 68 08 00
Email: bryan.richard.cleal@regionh.dk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2025/1/e72519

Abstract

Background: Although commercially developed automated insulin delivery (AID) systems have recently been approved and become available in a limited number of countries, they are not universally available, accessible, or affordable. Therefore, open-source AID systems, cocreated by an online community of people with diabetes and their families behind the hashtag #WeAreNotWaiting, have become increasingly popular.

Objective: This study focused on examining the lived experiences, physical and emotional health implications of people with diabetes following the initiation of open-source AID systems, their perceived challenges, and their sources of support, which have not been explored in the existing literature.

Methods: We collected data from 383 participants across 29 countries through 2 sets of open-ended questions in a web-based survey on their experience of building and using open-source AID systems. Narratives were thematically analyzed, and a coding framework was identified through iterative alignment.

XSL•FO
RenderX

Results: Participants consistently reported improvements in glycemia, physical health, sleep quality, emotional impact on everyday life, and quality of life. Knowledge of open-source AID systems was largely obtained through the #WeAreNotWaiting community, which was also the primary source of practical and emotional support. The acquisition of the components to build an open-source AID system and the technical setup were sometimes problematic.

Conclusions: The #WeAreNotWaiting movement represents a primary example of how informed and connected patients proactively address their unmet needs, provide peer support to each other, and obtain results through impactful, user-driven solutions. Alongside providing evidence on the safety and efficacy of open-source AID systems, this qualitative analysis helps in understanding how patients' experiences and benefits range from psychosocial improvements to a reduction in the burden of managing diabetes.

International Registered Report Identifier (IRRID): RR2-10.2196/15368

(J Med Internet Res 2025;27:e48406) doi:10.2196/48406

KEYWORDS

automated insulin delivery; diabetes technology; type 1 diabetes; insulin pumps; continuous glucose monitoring; peer support; community support; open source; impact; users; diabetes; emotional health; challenges; support; unmet needs; mobile phone

Introduction

Overview

In recent years, much attention has been given to the beneficial impacts that online peer support has on people living with chronic health conditions, yet the exact nature of these impacts may still appear to be somewhat intangible [1]. Through our findings, we provide an example of peer support in which the focus of the interaction is very tangible and where the impacts are profound and wide-ranging. The case in point being type 1 diabetes (T1D), where, in recent years, patients have been taking on the role of innovators in the design and development of technology used for their treatment. People with T1D who have developed and disseminated open-source automated insulin delivery (AID) systems exemplify a potential within online peer-to-peer communities that is only just beginning to be realized, particularly among people with chronic health conditions. To better understand how developments have reached the cutting edge in T1D treatment, we first provide some background about the challenges of everyday diabetes management and how open-source AID technology holds promise to alleviate them.

Background and Challenges in T1D Management

T1D is a lifelong condition caused by the autoimmune-induced loss of insulin-producing cells in the pancreas. Until the discovery of insulin by Banting et al [2] a century ago, T1D inevitably resulted in death by ketoacidosis within months. This changed once pharmaceutically procured insulin was available. Yet, while developments of the last 100 years in pharmaceuticals and technology have improved the physical health and life expectancy of people with T1D from a biomedical perspective, the burden of managing the condition remains a challenge. Life with T1D exists in the center of a continuous data feedback loop, where dosing of exogenous insulin via subcutaneous injections or insulin pumps must be frequently adjusted in accord with glucose levels and predicted trends, carbohydrate intake, physical activity, individual physiology, and behavior, among a variety of other factors.

The #WeAreNotWaiting Movement

Given the complexity of diabetes management, the everyday experience of managing the condition may lead to frustration, a feeling that was also a key driver in the emergence of the movement that has subsequently become known as #WeAreNotWaiting [3]. Initially, this frustration was related to issues with the accessibility of data from continuous glucose monitors (CGMs) in real time. Taking matters into their own hands, individuals reverse engineered commercially available devices to enable uploading of device data to the cloud in real time. In a model of diffusion that has characterized subsequent developments, the source code was published as open-source software, freely available to anyone and linked by the social media hashtag #WeAreNotWaiting. Eventually, a large, global community has united under this banner.

One of the most significant innovations to emerge from the #WeAreNotWaiting movement are open-source AID systems—sometimes referred to as do-it-yourself artificial pancreas systems—such as OpenAPS [4], AndroidAPS [5], iAPS [6], and Loop [7]. These systems link CGM sensors and insulin pumps with predictive control algorithms running on smartphones or microcontrollers. Insulin dosing is automatically adjusted according to predictions based on real-time data from CGMs, individual settings, and user inputs such as meal information. This technology emerged well in advance of the recent availability of commercial AID systems in select countries [8].

Current Evidence and User Perspectives

Evidence from randomized controlled trials and observational studies has demonstrated the safety and efficacy of commercial AID systems [9], as well as a positive impact on the users' lived experience and quality of life [10]. Given the lengthy, complex development and approval processes, only few systems are currently licensed, and their functionality is limited due to regulatory constraints. Even in countries with market availability of commercial AID systems, they are not universally accessible to the user, with reimbursement policies for devices varying considerably between countries [11].

XSL•FO
RenderX

Open-source AID systems work much like commercial AID systems at a basic level, connecting devices and automatically adjusting insulin dosing via predictive algorithms. They differ from commercial systems in terms of device choice, interoperability, transparency, access, customization, and usability. While commercial systems traditionally include manufacturer-designed education for clinicians and users, open-source AID user support and education initially took place via peer support outside of clinical settings [12]. The source code for open-source AID systems is freely available from online portals. In addition, these communities have also created documentation that has been translated into many different languages. Peer support is freely available to help with the setup and use of these systems.

It is estimated that >10,000 individuals worldwide are currently using open-source AID systems, and the uptake continues to increase globally. Evidence from a recent randomized controlled trial [13] and real-world studies based on self- and caregiver-reported outcomes [14,15], in silico [16], user-provided data [17], and observational studies [18-20] point to the safety and effectiveness of these systems, with improvements in clinical parameters such as glycated hemoglobin (HbA_{1c}), time in range, occurrence of hypoglycemia, and glycemic variability. However, until recently, the perspective of the user and their reported outcomes has only been touched upon [21-25], and there remains much to learn about how open-source AID systems impact the lives of those who use them.

Study Objectives

The focus of this study was to establish the physiological, cognitive, and emotional impact of open-source AID system use. We further aimed to identify sources of support and challenges associated with setup, regular use, and maintenance of open-source AID systems. This study was designed and conducted by the OPEN (Outcomes of Patient's Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology) consortium [26], an international, interdisciplinary team of patient innovators, clinicians, and scientists, many of whom also live with T1D and use open-source AID systems.

Methods

The research data were obtained from responses to open-ended questions included in a cross-sectional, web-based survey examining the use of open-source AID systems. The survey, titled DIWHY, was conducted between November 2018 and March 2019 [26].

Survey Design

The survey (Multimedia Appendix 1) was created by the OPEN consortium in collaboration with further open-source AID system users and was piloted with a small group of them before final release. The Checklist for Reporting Results of Internet E-Surveys was used to guide survey development [27]. The survey comprised 39 items in total with 2 sets of open-ended questions, which sought to capture lived experiences with open-source AID systems in the form of narratives. For these questions, participants could provide a free-text response with a maximum length of 1000 words each. The first set of questions assessed sources of information, support, and emotions associated with the preparation and setup process. The second question set addressed the impact of open-source AID system use on everyday life, associated changes, and challenges with respect to the transition to open-source AID systems.

Participants and Recruitment

Participants included in this study were adults (aged >18 y) who are living with diabetes (type 1, type 2, or other) and are current users of open-source AID systems. There were no restrictions in time since diagnosis or commencement of open-source AID systems. In this study, participants were exclusively recruited from the global diabetes online community, leveraging the outreach of the #WeAreNotWaiting movement. This approach enabled us to tap into a highly empowered and informed population of people with diabetes and particularly target those who were users of open-source AID systems. We used several social media channels, including the Facebook (Meta Platforms) groups Looped (with >6000 members) and AndroidAPS users (with >1800 members as of November 2018) and regional subgroups, and posted publicly on Twitter using the hashtags #WeAreNotWaiting and #DIYAPS to engage with the wider diabetes online community. Participation was further promoted through announcements on the OPEN project website. There were no paid promotions or targeted advertising on any platform. Participants were able to choose between 2 language options (English and German).

Data Collection and Analysis

Data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) [28], and the qualitative analysis of the narratives was performed using NVivo (version 12; Lumivero, 2018). The narratives were analyzed with an approach applying the principles of template analysis [29], in which a hierarchical coding structure is recommended to allow researchers to capture the diversity of meaning within broader overarching themes. This approach was deemed necessary due to the framing of the open-ended questions, in which respondents were prompted to reflect on specific aspects of their open-source AID system use experience. Therefore, the initial template and coding were deductively driven, with physical impact, emotional impact, sources of support, and challenges established as the overarching themes. In the subsequent stage of coding, the deductively coded data were analyzed inductively to identify subthemes within the overarching framework. Initially, 3 coders worked (BC, YC, and MW) on a small sample of the deductive data and presented putative inductive themes for evaluation within the author group. These were discussed and refined to ensure that there was equivalence in relation to the levels of abstraction and thus in the hierarchical organization of the template. On the basis of this initial inductive coding, it was agreed that only 1 level of abstraction was necessary and that the overarching themes could be articulated sufficiently by 1 level of subthemes. The resulting comprehensive codebook, which included both deductively and inductively developed codes, descriptions, example quotes, and thematic categories, are detailed in Tables 1-4. All the data were then coded deductively in accordance with the full template by

XSL•FO
RenderX

BC and YC, although emergent themes not established in the initial inductive analysis were also proposed. Finally, 2 coders (HB and DC) reviewed the coding template to assess the external validity of the coding process. Where clarity or consistency was questioned, further group discussions were held until all issues were resolved.

Table 1. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery system user narratives collected from a multinational cohort of 383 participants. This table presents findings related to the physical health benefits and quality of life impact of using open-source automated insulin delivery systems, such as improvements in glycemic control, time in range, and reduced hypoglycemia.

Theme	Description	Example quotes
(A) Physical impact and quality of life impact		
Glycemic outcome improvement (A1)	Refers to improved time in range and HbA_{1c}[a] levels, less glucose variability	• "I purchased the Miaomiao brand Bluetooth transmitter to send my blood glucose levels to my phone, thereby having access to alerts that have undoubtedly saved my life." • "My biggest hope was to control my blood sugar peaks due to gastroparesis and this works 95% of the of time very well. Thanks to this technology, I am a big step closer to my goal." • "He's not T1, but was concerned about my time in range and felt the DIY Loop system was better than approved FDA methods." • "Looping has dramatically improved time in range and how I feel. After 5 weeks Looping, my A_{1C} lowered from 6.8 to 6.3."
Hypo- and hyperglycemia (A2)	Refers to fewer hypo- and hyperglycemia and reduced complications associated	• "Overall, my blood sugar adjustment has improved tremendously. Hypoglycaemia is extremely rare. Times above the limit occur but are limited in duration." • "My values have become immensely better. Hypos with unconsciousness did not occur anymore." • "I've always had problems with hypos. The Freestyle showed me in the morning that I was hypoglycemic, nevertheless, I spent the night—retrospectively—in hypoglycemia. Or even a whole night with levels over 250 mg/dl."
Sleep quality (A3)	Denotes all aspects of improved sleep quality for either caregivers or children, such as increased sleep duration, fewer sleep interruptions, and feeling better rested in the morning	• "Thereby having access to alerts that have undoubtedly saved my life, both while sleeping and while alone in public transportation, among others." • "Only looping for about 2-3 weeks. So far, better quality sleep for my husband and me (no nocturnal hypos which would happen regularly prelooping)." • "I sleep so much better. I no longer fear lows. I understand my body's needs much better." • "It was hard to convince my wife that I was going to turn over control of my diabetes to open source code that I downloaded from the internet. It actually got to a point where I explained to her that I was doing it with or without her approval. When she realized how passionate I was about it, we sat down together and I explained the ins and outs of everything. She sleeps so much better now knowing I'm healthier overall and more safe overnight."
Exercise management (A4)	Refers to the improvements in physical exercise since the use of DIYAPS[b]	• "Better sleep and exercise management, AAPS has very good objectives to work through. Diabetes management is now less hassle and lower risk." • "Exercising and working out is a lot more easier since I no longer have that many low BGs. I overall feel physically and mentally much better since my time in range has increased by 30%."
Quality of life (A5)	Refers to the mentioned improvements in quality of life and describes the degree to which an individual is healthy, comfortable, and able to participate in or enjoy life events	• "But today, at 44, I have a system that already reduces a lot, and ensures very good values and thus a better quality of life." • "Help with stress situations both high and low, preworkout and postworkouts are less drastic, overall better mental health since I don't need to make so many decisions throughout the day about BGs and knowing that I have the ability to have more time in range and a better A_{1C} while still living a pretty normal life." • "A DIY system gives me this flexibility and adaptability. A huge weight has been lifted off my shoulders knowing that I will be safe during the night, and won't be doing undue damage to my organs during stressful situations like exams or during panic episodes." • "Since then, every night was like that and every day I feel more active and fitter...that's what I call Quality of Life."

[a]HbA_{1c}: glycated hemoglobin.

[b]DIYAPS: do-it-yourself artificial pancreas systems.

XSL•FO
RenderX

Table 2. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery (AID) system user narratives collected from a multinational cohort of 383 participants. This table summarizes user-reported experiences regarding the positive and negative emotional effects of open-source AID systems.

Theme	Description	Example quotes
(B) Emotional impact		
Positive (B1): describes positive emotions of participants related to the experience with open-source AID in daily use, including anticipation, enjoyment, excitement, relief, freedom, and inspiration.		
Anticipation	Describes hopeful emotional states of anticipation and great expectations of participants that lie on the open-source AID system for improved diabetes management and hope for improved quality of life	• "Looking forward to next steps: predicting algorithms." • "I had high hopes that everything would be different from now and I do not have to do anything for my diabetes anymore, and that the loop regulates everything on its own." • "Hope and impatience since I was looking for an atomization almost for 20 years."
Enjoyment	Describes that participants enjoy using the open-source AID system or enjoy the satisfied outcomes brought by open-source AID system use	• "Now I enjoy quiet nights and hope for a future without long term complications." • "I am doing well and I am happy to use the loop." • "My CGM alarms are all turned off now and I love the silence." • "In terms of expectations, I had a vague feeling that closed looping could be a solution to my lack of control during the night but I did not have very clear or measurable expectations. I just wanted to try it, I was curious to see what it was going to bring me. And now, of course, soooooooo glad that I did that!" • "Just awesome. By far the biggest impact has been the reduction in time spent "doing" diabetes. I am rarely woken at night even after intense exercise and little tweaks which previously interrupted my day have largely disappeared. It is a revelation."
Excitement	Describes exciting emotions of participants related to the experience with open-source AID systems in daily use, including excitement, happiness, and satisfaction with AID system use and associated results	• "I was just so excited to learn about DIY closed looping, and ordered gear immediately." • "I have been looping for only 4 months now and I am really grateful to all the people that have worked so hard to developed this amazing technology." • "I was very excited learning about how the system worked. I saw it as a challenge to understand and build it. Understanding the algorithm and building it myself gave me a great understanding of the system and mitigated any fears that I might have had about it's functioning." • "To me, loop is a miracle—I can hardly believe it after 35 years to finally have normal blood glucose levels." • "Just awesome. By far the biggest impact has been the reduction in time spent "doing" diabetes. I am rarely woken at night even after intense exercise and little tweaks which previously interrupted my day have largely disappeared. It is a revelation!" • "[I am] able to read the blood sugar over a watch and no longer have to perform finger pricks, should become true! I still remember the exact moment, I immediately had tears in my eyes."
Relief	Describes that participants feel relieved since their diabetes-related complications alleviated and quality of life improved after using open-source AID systems	• "I felt well and relieved, within a few days there was a clear improvement in my control." • "It's the reduction in disease burden, reduction in lows and highs and the overall improvement in time in range that has changed things so much for me." • "Since the change, I feel safer and simply more comfortable." • "My diabetes does not bother me any more, I can accept it and even like it."
Freedom	Describes the feeling of freedom since participants gained more control of their diabetes by using open-source AID systems	• "The essential feeling of freedom and the feeling of being in control of diabetes." • "I feel free for the first time in years!"
Inspiration	Describes that participants were encouraged and motivated during their journey of building open-source AID systems	• "No special 'key events' other than reading many stories of a diverse group of T1's who all seemed to overcome all the I-am-new-to-software-building challenges and reading how happy they now we're that they took this effort. Also, reading questions of people and seeing that they were answered fast and elaborately by more experienced and knowledgeable users made a great impression on me." • "I was immediately thrilled and the hints that the whole thing wouldn't be completely legal somehow motivated me even more." • "I knew that other people could do it so I was convinced I should be able to do it myself as well. I expected to run into issues in the building process, but I wasn't held back as I knew I would also run into solutions."

XSL•FO
RenderX

Theme	Description	Example quotes
Negative (B2): describes negative emotions, such as fear, frustration, worry and anxiety of users, mainly due to the concern of not being able to build and maintain open-source AID systems and difficulties encountered.		
Caution	Describes participants' concerns in regard to building and maintaining open-source AID systems	• "I was quite worried about trying it, there were many things I didn't quite understand and found technically challenging" • "I am still a bit afraid to close the loop." • "I have other fears: will my OTG cable have a loose connection again tonight? will I get to an old replacement pump (combo)? We integrate the Dexcom G6 with xdrip or disassembled xdrip the battery - how was it read today? Will the pharmaceutical industry put more obstacles in our way?" • "I have had some fears of system failure, but i do not have them anymore"
Frustration	Describes participants' frustration when experiencing difficulties in sourcing, building, or maintaining open-source AID systems	• "Unfortunately, I first failed because of the technology (availability of the accessories, order from the USA). It was a bit frustrating, but I did not want to give up." • "I also experienced a sadness for how long I had struggled (43years) with inadequate tools to manage my condition." • "My biggest challenge has been the isolation of it all. Online support is always present very timely, but not the same as having someone to be present to help troubleshoot when loop stops. I live in very rural area and sometimes I feel like I'm stranded on an island. Frustrating at times, but I would not trade my cure for anything!!"
Worry and anxiety	Describes that participants worry about the problems that may arise from the use of open-source AID systems	• "as I have no idea what the programme is doing and every time I keep my fingers crossed that I won't see any screen other than in the tutorial because I have no idea how to fix that." • "The biggest hurdle was I was not confident I could build the loop app on my own." • "And I have concerns about the dependence on Bluetooth/Wifi/Internet/ servers." • "I had heard things about DIY closed loop systems in the various cgm related Facebook groups, but had always thought I wouldn't be able to build & maintain one." • "It was something that I considered at the time, but never thought I would be able to do it as I'm not very good with coding/tech." • "suddenly relying on the phone instead of your own mind is a little strange, I often did not trust the loop and delivered boluses on my own or ignored the TBR suggestions." • "A defective libre sensor the other day (already on the 2nd or 3rd day) has given me a lot of restlessness and effort, because the measurements were incorrect every now and then. I was Worried every now and then a single component of the overall system will fail and cause high adjustment effort."

XSL•FO
RenderX

Table 3. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery (AID) system user narratives collected from a multinational cohort of 383 participants. This table outlines the challenges users face during setup, regular use, and maintenance of open-source AID systems, including technical difficulties and knowledge gaps.

Theme	Description	Example quotes
(C) Challenges		
Accessibility (C1): describes the challenges encountered by the participants regarding accessing the open-source AID systems device, including sourcing hardware, cost, and understanding of the open-source AID systems rationale.		
Sourcing hardware	Describes the problem of sourcing hardware that is compatible with open-source AID systems	• "My main struggle was with losing access to my favorite pump (OmniPod). I switched back to a tube based pump I'm favor of this system, while hoping that the pods are cracked soon. I followed the OpenOmni efforts." • "but then I was unlikely to do it as I didn't have the right pump anymore" • "And my health insurance was vague about the info they needed to get the pump and supplies reimbursed. Denied my request several times. Took a lot of communication between health care providers, myself, insurance which was a frustrating process for me." • "My health care providers initially didn't want to prescribe the (loopable) pump."
Cost	Describes the issue of costs associated with building open-source AID systems	• "The main challenge has been self-funding the CGM which is expensive in New Zealand." • "I started with the pump therapy just before loopin. The health insurance did not supply me with a pump at the time, so I was on my own without diabetic care." • "The Dana RS pump has the one huge advantage of the open interface to control, otherwise the pump looks like a product from the penultimate decade, compared to other systems. The needles are not great either and the counterclockwise luer lock should be replaced. The should be a standard for pump connections to avoid manufacturer tie-ins and make cost reductions possible (the equipment is still too expensive)."
Understanding	Describes problems with understanding the rationale provided by the instructions for open-source AID system setup	• "My initial feeling of starting to build my loop was that it is poorly documented and difficult to diagnose setup issues." • "Understanding the system was both challenging and frustrating. The documentation is poorly explained and the system has its limitations but it is the best we have." • "In the meantime there is a ton of information about closed loops, which was an intensive learning and reading phase in both the medical literature and the posts and documentation of the community. Since at the beginning the overview was missing, this was also a bit confusing because the meaning of individual components was not yet estimable for the own implementation when reading. What is a Wixel, can it be eaten, and if so, how many carbs do I have to expect?. It had to burst a lot of knots until it was clear which software components to make the (actually quite simple) beginning. But maybe that's just too complicated for me, a certain tendency is not to be denied." • "In the lead up to building, I felt overwhelmed by all info." • "Apart from the setup difficulties and my lack of English skills I had no difficulties."
Setting up (C2): describes the difficulties encountered by the participants when maintaining the use of open-source AID systems, including adjusting and fine-tuning, consumption of time and effort, inconvenience in everyday life and technical issues.		
Adjusting and fine-tuning	Describes the process of determining the factors and further adjusting and fine-tuning	• "Determining the factors with decimal place accuracy, calculating the sensitivity factor was/is in first place effortful." • "The only difficulty in the change of therapy (apart from the unfamiliar handling of the technique in general) was to adjust the loop for different types of sports, depending on the time of day and physical condition." • "disillusionment followed immediately - even the updated nightscout / set up is very complicated to operate for non-nerds. It's often not clear what happens and what is shown on night scout."
Time and effort	Describes the amount of time and effort required for open-source AID system setup	• "It takes a lot of time and attention to begin with" • "It has been a tricky process, as we have been learning by doing" • "With the technical implementation, it was an uphill battle. It took a while to work but was worth the effort...I do not want to be without loop..."

XSL•FO
RenderX

Theme		Description	Example quotes
	Everyday life	Describes inconvenience with using open-source AID systems in daily life regarding exercise and diet, etc	• "I had to reduce the overall carb intake in order to achieve better time in Range, and this had a major impact on my and my family lifestyle" • "The regulation of sports is still a bit difficult" • "I discovered Loop app was terrific, but the phone was too bulky. So I purchased a new pump outright and started AAPS with a tiny phone. I am actually about to try AAPS on a stand-alone watch. The bulk of the gear is super important to me, as I can't be carrying a lot of excess stuff at work" • "I always think about if I need to charge anything (with Enlite it's just a bit more complicated than with other systems.) I have to keep reminding myself not to go out of the house without my new hand bag - and not without a smartwatch" • "Social judgment from part of the DT1 Community which [accused] me of obsession with the disease."
	Technical challenges	Describes technical issues with the equipment, such as battery charging	• "The issues I've had have been of a minor technical issue only, like accidentally shorting out my Miao Miao charger, and having difficulty ordering another. So I changed to Dexcom G5, and have learned to rebattery my transmitters, making it actually cheaper than using Libre. I have also had some troubles with my phone updating it's OS and becoming useless. I also have some battery issues with my new pump." • "Biggest challenge was building the app in Android studio, as I have no idea what the programme is doing and every time I keep my fingers crossed that I won't see any screen other than in the tutorial because I have no idea how to fix that." • "The only difficulties that I am sometimes experiencing are technical problems such as connectivity issues between the Riley link and the pump, and blood sugar fluctuations once the insulin sensitivity changes, which are more noticeable now with a tighter BG management than before, where every day with diabetes was just chaos."

XSL•FO
RenderX

Table 4. Comprehensive codebook detailing thematic analysis of open-source automated insulin delivery system user narratives collected from a multinational cohort of 383 participants. This table highlights the role of peer support, online communities, and other resources in facilitating the adoption and use of open-source AID systems, emphasizing the sense of community and shared expertise.

Theme	Description	Example quotes
(D) Support		
Family and friends (D1)	Describes that the participants received support from their family or friends	• "My husband supported and encouraged me, which helped but I'd have gone ahead even without that support." • "Friend who had acquired the components and assisted [me] in building. He was already using [AndroidAPS]."
Online support (D2)	Describes the types of support that participants seek on the web, such as social media, blogs, forums, and other people with diabetes	• "First, I found information about AAPS in Freestyle Libre Forum." • "I felt the whole process was very simple with very comprehensive instructions and support through Facebook." • "March 2017 was the first time I started looking for treatment improvements. I acquired all the information and knowledge in through own research on the internet." • "In our area there is a Whatsapp diabetes group that meets in person now and then. A PwD there works in IT and was contacted about 1,5 years ago. With his technical support I have closed my loop." • "I set up a Nightscout server and the AndroidSeries600Upload app for his Medtronic pump in the hospital - and read a lot - and at first set up OpenAPS for myself. As a technophile I was of course immediately on fire, but I still took 2 months to read about the topic online intensively, trawling through forums, Facebook etc., and my anticipation and enthusiasm grew steadily."
DIY[a] community (D3)	Describes that the participants received support from the DIY community, such as help from author DL	• "I got support from the DIY community via face to face meetings and via the online community." • "When I could connect the technical side (IT) with my diabetes, it all started. Then I met Adrian and Dana [authors AT and DL] in person."
Conference and meeting (D4)	Describes the support that the participants received from conferences or meetings	• "However, due to the great help from the looper.de group and the looper meeting, I already wanted much more at this time and have been able to implement this with a lot of reading and informing and with still some technical problems"
Medical professionals (D5)	Describes that the participants received support from medical professionals, such as doctors, nurses, diabetes educators, endocrinologists, etc	• "I actually first heard about DIY options from my doctor, who referred me to another patient who was already using one."
Self-support (D6)	Describes that the participants learn to build DIYAPS[b] all by themselves, with no direct support from others	• "I like the user manual, which it written clearly step by step." • "[N]o direct support or advice from third parties."

[a]DIY: do-it-yourself.

[b]DIYAPS: do-it-yourself artificial pancreas systems.

Ethical Considerations

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Charité–Universitätsmedizin Berlin (EA2/140/18). Informed consent was obtained electronically from all individual participants included in the study. The deidentified datasets are available from the corresponding author upon request. Participation was anonymous and voluntary. No financial or other compensation was provided.

Results

Participant Characteristics

In total, 383 participants (N=722, 53% of participants of the DIWHY survey) responded to the open-ended questions in the survey, and there were a combined 645 responses to the 2 open-ended items. Characteristics and clinical features of the cohort are shown in Table 5.

XSL•FO
RenderX

Table 5. Demographic and self-reported health characteristics of the participants using open-source automated insulin delivery (AID) systems.

Demographics	Values
Gender (n=383), n (%)	
Men	203 (53)
Women	179 (46.7)
Other	1 (0.3)
Age (y), mean (SD)	43 (12)
Type of diabetes (n=383), n (%)	
Type 1	381 (99.5)
Type 2	0 (0)
Other	2 (0.5)
Duration of diabetes (y), mean (SD)	30 (12)
Duration of open-source AID system use (y), mean (SD)	4 (2)
Most recent self-reported HbA_{1c}[a] level, mean (SD)	5.89 (0.62)
Type of open-source AID system used regularly (n=423), n (%)	
OpenAPS	65 (15.4)
AndroidAPS	245 (57.9)
Loop	110 (26)
Other[b]	3 (0.7)
Region (n=383), n (%)	
Europe	282 (73.6)
Germany	184 (48)
United Kingdom	41 (10.7)
Austria	7 (1.8)
Spain	7 (1.8)
Netherlands	4 (1)
Finland	6 (1.6)
Czech Republic	5 (1.3)
Bulgaria	4 (1)
Sweden	2 (0.5)
Other[c]	22 (5.7)
North America	69 (18)
Canada	15 (3.9)
United States	54 (14.1)
Western Pacific	23 (6)
Australia	12 (3.1)
New Zealand	11 (2.9)
Asia	2 (0.5)
South Korea	2 (0.5)
Africa	2 (0.5)
Algeria	1 (0.3)
South Africa	1 (0.3)
I'd rather not say	5 (1.3)

XSL•FO
RenderX

Demographics	Values
Education (n=379), n (%)	
Doctorate or graduate degree	154 (40.6)
Bachelors, professional or associate degree	129 (34)
Trade, technical or vocational training	27 (7.1)
Some college credits	11 (2.9)
Secondary school	34 (9)
Some secondary or primary school	22 (5.8)
No schooling completed or none of the above	2 (0.5)
Occupational status (n=382), n (%)	
Full time	265 (69.4)
Part time	60 (15.7)
Unemployed	4 (1.1)
Retired	20 (5.2)
Student	24 (6.3)
Other or none of the above	9 (2.4)
Professional background (n=305), n (%)	
Medicine	76 (24.9)
Tech	81 (26.6)
Finance	40 (13.1)
Other	108 (35.4)
Household annual income (US $; n=336), n (%)	
<20,000	34 (10.1)
24,000-34,999	27 (8)
35,000-49,999	48 (14.3)
50,000-74,999	87 (25.9)
75,000-99,999	44 (13.1)
>100,000	79 (23.5)
I'd rather not say	17 (5.1)

[a]HbA_{1c}: glycated hemoglobin.

[b]xDrip, Nightscout, offline uploader for Medtronic 600 series, Hackabetes Artificial Pancreas Project, custom, or own developments.

[c]Belgium, Croatia, France, Hungary, Ireland, Italy, Lithuania, Poland, Romania, Russia, Serbia, Slovakia, and Switzerland.

Template Analysis

Emotional and Quality of Life Impact

"Anticipation" and "curiosity" were emotions mentioned by participants in relation to their first encounter with open-source AID technology. This highlights the intuitive appeal of this solution for diabetes management—"I had envisioned this type of solution for many years and was looking out for the emergence of suitable technology" (Man, age 59 years, United Kingdom)—and why, for many people, initial reservations were quelled by the potential improvements offered by AID systems—"One is a little uncertain, but the curiosity for the improvement of control has won!" (Gender unspecified, age 49 years, Germany).

However, as this sentiment indicates, anxiety and caution were also a part of the emotional responses experienced by our participants. This could be as they confronted the prospect of developing their own system—"I was quite worried about trying it, there were many things I didn't quite understand and found technically challenging" (Woman, age 69 years, United Kingdom)—but was also apparent even once the system had been successfully built—"Originally it felt like a big step to let the algorithm make changes" (Woman, age 42 years, Australia).

Thus, both the challenges—real and anticipated—in setting up the system and the prospect of allowing an algorithm to undertake a life-critical role could have a negative emotional impact.

XSL•FO
RenderX

However, for the most part, initial concerns about the complexity of the technology and ceding control to an algorithm were replaced by a sense of "pride" and "relief"—"I feel very good and proud I did it because it was technically difficult to build it with my pump and CGM" (Man, age 50 years, Germany).

The sense of relief experienced by our participants was 2-fold; relief that the system was built and functioning but also a sense of being partially released from the burden of everyday diabetes management—"The most impressive thing is how little diabetes suddenly plays a role, how simple everything has become, how rarely one suddenly has to wonder about metabolic fluctuations, how well one can sleep, knowing that blood sugar stays in range" (Woman, age 49 years, Germany).

Twisting the concern with automation, some participants also noted that it was precisely because control was given over to an algorithm that improved outcomes could be achieved—"I was happy to hand over control to something which makes fewer irrational decisions and is less emotionally involved in the process" (Woman, age 35 years, United Kingdom).

The relief they experienced did not come without considerable effort and "frustration," and this was also a common emotion in the narratives. Part of this frustration was related to the reliability of technology:

> *While the burden of what to do in reaction to blood glucose has gone down, the tech troubleshooting and figuring out how to fine-tune has increased greatly. Traded one problem for another.* [Woman, age 62 years, United States]

As can be inferred from this comment, frustration was also driven by an expectation that the level of automation would be greater: "A few months into closed looping I am starting to see results, though I was expecting [it] to be easier and thought it was going to handle much more the ups and downs by itself" (Man, age 41 years, Netherlands).

Yet, most participants declared themselves happy to invest the effort when the reward was so tangible and transformative for overall quality of life:

> *...it doesn't just fix everything and that there are still settings to adjust and check, but once these were okay then I've had very few issues. It has allowed me to take a back seat with my diabetes care...It has taken huge amounts of the diabetes burden away from me!* [Woman, age 25 years, United Kingdom]

In fact, the work also served as a source of inspiration, with many participants gaining new insights into important factors influencing glucose fluctuations:

> *Looping has provided me much detailed insight into the inner-workings of my endocrine system and diabetes management. I've learned that my insulin ratios and [basal rates] needed to be greatly adjusted. As I've learned, two bad settings can mask each other and end up appearing to be correct. The learning curve is steep, but very rewarding.* [Woman, age 24 years, France]

Overall, participants indicated that the net gains of open-source AID were extensive and profound. Often, a sense of gratitude was expressed:

> *The community has helped me so much. I can't express my gratitude to all developers, helpers and people in my local community as well who freely give their time and skills to make this possible.* [Woman, age 65 years, Australia]

Source of Support

The community mentioned earlier highlights the particular model of diffusion that has fostered the use of open-source AID systems. To echo a common refrain in this context, *do-it-yourself* does not mean *do-it-alone* [15], so while each user is ultimately responsible for building their own system, the support that they can obtain in doing so is potentially extensive.

For those without preexisting skills in IT, support was at hand, for some, among one's established social network, for example, family and friends:

> *I was very intimidated at first as I have extremely limited coding knowledge. After following along in the group for a while, I began to get more comfortable. My boyfriend also encouraged me and offered to help set it up since he has a bit more tech knowledge.* [Woman, age 22 years, Canada]

Both technical and medical expertise within the individual's network was an important antecedent to the uptake of open-source AID systems:

> *My partner immediately supported me because, as a doctor, it was very clear to him that it had a much better metabolic effect...The support of the Facebook group especially for small logistic things was very important to me.* [Woman, age 37 years, Germany]

This participant's comments also highlighted that direct support from health care professionals (HCPs) may be lacking:

> *The support by my social environment has increased, the support by doctors and the diabetologist's office are completely lost, I consider this to be a risk, I am waiting for the moment when our diabetologists will not only be "not forbidden" but required to inform about the Closed Loop as the gold standard of therapy.* [Woman, age 37 years, Germany]

This is not to say that HCPs were not supportive or positive about AID outcomes, but the support they could provide was more often informal and emotional rather than practical:

> *My two diabetologists know about the loop and are amazed/enthusiastic about my [glucose] levels, but unfortunately cannot support me for legal reasons.* [Woman, age 55 years, Germany]

While an individual's social relations could profoundly impact the building and maintaining an open-source AID system, support was still available even without direct expertise in one's personal network—"I've found the technology almost impossibly difficult to deal with and have had a considerable

XSL•FO
RenderX

amount of personal help from other users" (Woman, age 62 years, United Kingdom).

For some users, this occurred at face-to-face meetings (eg, "build events"), where expert users could guide them, but, for the majority, such support was obtained via online forums—"The biggest (and for me only) help with technical problems or 'fine-tuning' the settings is provided by the Looper online community" (Woman, age 27 years, Germany).

The #WeAreNotWaiting community was the main source of support cited by our participants, and this was multifaceted. In its most basic form, the online documentation developed by users for users was an essential resource and frequently praised for its clarity. Beyond this, in the various social media–based groups connected with open-source AID systems, there was also a wealth of information from reading existing threads and others' posts, where frequently asked questions and troubleshooting topics could guide through challenging aspects of the process. Finally, the online forums also served as a real-time support network, where users could expect rapid and reliable responses to whatever issue they might reach out for:

> *I don't want to finish without mentioning the importance that the support groups are having to me. Both in helping understand and setting the system and managing the everyday life... It is completely amazing being able to be connected to so many people who are also looping and give and get support.* [Woman, age 40 years, Spain]

Also notable was how the encounter with this community and its essentially altruistic spirit could inspire new users to be willing to participate and serve within this support network:

> *What one cannot do, the many can manage. The group has helped me. I'm getting involved as well and spread the knowledge so others can benefit from it.* [Man, age 54 years, Germany]

Challenges

Principally, the challenges reported included (1) accessibility and (2) technical setup and maintenance. There were 3 prevailing issues regarding accessibility. The first one was cost, since essential hardware was not readily available via health care services or insurers:

> *CGM is prohibitively expensive in my country. I only started using it two months before looping as part of preparing to loop. I am trying to hang in there paying for it because of the fantastic benefits but it is a big drain on family income.* [Woman, age 51 years, Australia]

In addition, even in circumstances where hardware was potentially available to users via public health care or insurance, access could still be problematic if potential users were not eligible according to insurers, or HCPs had reservations about recommending devices that could be used for open-source AID systems:

> *My healthcare providers initially didn't want to prescribe the (loopable) pump. And my health insurance was vague about the info they needed to get the pump and supplies reimbursed. Denied my request several times. Took a lot of communication between health care providers, myself, insurance which was a frustrating process for me.* [Woman, age 33 years, country of residence undisclosed]

The final aspect of accessibility was the anticipated complexity of the process and the documentation to be followed to set up an open-source AID system—"I had heard things about DIY closed-loop systems in the various CGM-related Facebook groups, but had always thought I wouldn't be able to build & maintain one" [Woman, age 47 years, United Kingdom].

Thus, the obstacle was sometimes more about expectation than experience, with people deterred from the attempt by the expectation that they would not be able to solve technical issues.

However, for others, it was as much the experience as the expectation that could provide an obstacle to access—"There was a lot to learn. I often sat crying in front of the computer" (Woman, age 35 years, Germany).

Technical challenges associated with building open-source AID systems were prevalent in the narratives, with only a handful indicating that the process was straightforward:

> *I tried about a year before I actually started to build a system and it proved too difficult. After a year of being burned out and things not being any better, I tried again and succeeded.* [Woman, age 36 years, United States]

Participants conveyed difficulties with both hardware and software components, for example, with connectivity loss.

Even with all components fully functioning and connected, other technical challenges remained, though these were more related to the technicalities of diabetes management than technology per se. Users of open-source AID systems take on the role of diabetes experts as much as programmers and are required to fine-tune the settings on their devices in accordance with a selection of parameters:

> *[The] first weeks of looping were a bit hard because my ratios were off and it was hard to understand why Loop is making some decisions. Or, what's even more important, which parameter should be tweaked in order to make it behave better.* [Man, age 32 years, Poland]

Determination of these settings is generally undertaken by HCPs in the context of prescribed, commercial devices, ideally in collaboration with the user, but it may remain more or less opaque to the individual with diabetes. So, although many users were highly engaged with diabetes management previously, there was a learning curve involved for most, not least because the level of control for different parameters allowed by open-source AID systems extends significantly beyond those in commercial systems:

> *So I have to work more on my settings. Nothing is (fully) automatic and runs all by itself. For me, as a technician who believes in the possibilities of self-regulating automation, there is still a lot to be done.* [Man, age 51 years, Germany]

XSL•FO
RenderX

The combination of the different challenges involved in building an open-source AID system evoked another issue for some, inasmuch that considerable time was required to resolve the issues emerging from building and maintaining the system—"My husband has suggested several times that I was doing more work with the system than without, because of the frustration & time (whole weekends) involved in getting my loop back up & running" [Woman, age 51 years, United States].

Yet, while time was undoubtedly a factor to be dealt with, the extent to which it was perceived as a challenge, impinging on everyday life, was often weighed against the time spent "doing diabetes" before transitioning to open-source AID systems:

> *My own personal tight control prior to looping was very time-consuming. Post APS I save more than 1hr every day. Imagine my experience of living 1/24th longer life for the rest of my life because of APS.* [Man, age 46 years, New Zealand]

Physical Impact

For the most part, participants reported marked improvements in physical health in accordance with the measures traditionally used to gauge this in T1D, such as HbA_{1c} and time in range, as reported elsewhere [15]. For some, these improvements occurred soon after their switch to open-source AID systems:

> *Benefits from the first 8 weeks: 80-90% of the time-in-range without changing my lifestyle! Previously that value was 40-45%.* [Man, age 37 years, Germany]

The time regarding changes in HbA_{1c} was longer but still commented upon, often as levels within a reference range for people without diabetes:

> *I knew some that tuning was needed but I was patient. Now I have used a DIY system 24/7 approx 23 months for almost two years! Hba_{1C} is 5.2%, I'm happy.* [Man, age 44 years, Finland]

In addition to these clinical outcome improvements, participants also reported changes in their health based on more immediate, everyday experiences. The experience of hypoglycemia was something alluded to extensively—"The blood glucose fluctuations and the hypos have become much less, I feel much safer and I am doing things again that I used to avoid" (Woman, age 45 years, Germany).

The sense of safety can have a profound effect on an individual's life. Both hyper- and hypoglycemia in their moderate expression can induce physical symptoms that are unpleasant and disruptive but, in their extreme extent, are potentially fatal. Ameliorating the risk of highs and lows, open-source AID systems served to diminish the unpleasant symptoms and, at the same time, reduce the anxiety attached to what might happen in worst-case scenarios. This, aside from its direct benefits, also had follow-on effects on other health-related aspects. Many participants noted benefits of open-source AID systems related to physical activity—"I'm much less afraid of unplanned physical activity because the loop usually regulates it with a few extra carbs" (Woman, age 55 years, Germany).

Again, the point here is not only that glucose levels are within range during exercise but also that the potential fear around exercising was lessened. Fear of exercise and its unpredictable impact on glucose levels represents a clear obstacle removed by open-source AID systems, with potential general health benefits—"I have recently started exercising again after years of sedentary living" (Woman, age 31 years, Australia).

By far, the most frequently mentioned quality of life improvement among our participants was sleep duration and quality:

> *I can sleep and have no alarms from the CGM at night. In the morning I wake up with a value that I do not have to correct. This has a positive effect on the blood sugar during the day.* [Woman, age 38 years, Germany]

Persistently disturbed sleep is by any reckoning something that one would expect to impact health and overall quality of life, but for people living with T1D, it is a given:

> *I SLEEP. That's the most brilliant, life-changing thing. I'd been sleep-deprived for so long I didn't even realize what a difference it would make.* [Woman, age 49 years, United States]

In a similar vein to exercise, open-source AID systems had a dual impact in relation to sleep. It served to alleviate symptoms of hypo- and hyperglycemia that could disturb sleep directly or trigger alarms on devices waking people up. Simultaneously, it helped reduce the fear of nocturnal hypoglycemia, which could result in sleep difficulties due to anxiety and adverse aversion strategies, such as intentionally aiming for higher glucose levels before sleep:

> *I have no anxiety about sleeping alone when my wife is working away from home. I actually sleep through the night. Eating out is no longer a major stress since I know that even if I underestimate carbs [it] will fix my errors overnight and I will wake mostly in range.* [Man, age 42 years, United Kingdom]

As might be expected, improved sleep was also associated with further physical and mental health benefits:

> *Waking up that first morning in a normal range, and every morning thereafter was amazing. It's incredible how much more you can get done in a day when you wake up in a normal blood glucose range.* [Man, age 39 years, United States]

Part of this is obviously about having more energy because of being well-rested. Beyond that, there is the important difference of waking up with glucose in target range and how this resonates through the rest of the day:

> *The almost fully automatic delivery of needed insulin has made life a lot easier and once the factors are set correctly, it is almost possible to live like a 'healthy' person. It's also much easier to start a new day, starting at a value of 90 mg/dl and not 180-200 as before! Working days are much easier than before!* [Man, age 47 years, Germany]

XSL•FO
RenderX

Rather than using considerable time, energy, and resources attempting to reestablish balance in one's glucose levels, open-source AID systems allowed people to concentrate more on the business of the day. So, whereas previously a working day may have felt more like "running the gauntlet," a different experience and outlook on life could be fostered:

> *Since I've used [open-source AID], I was upgraded on my job, I'm mentally faster and sleep like a baby without alarms. I've started several personal projects and [I'm] currently on professional certification. I have plenty of quality time now without hypo or hyper and finally happy.* [Woman, age 40 years, Spain]

Discussion

Principal Findings and Their Implications

The findings we present here concerning the impact of open-source AID systems highlight their immense benefits from the perspective of the user, simultaneously setting the extremely challenging nature of diabetes and the ways it may compromise quality of life into relief. This is the first study to analyze narratives and to examine the emotional and physical health impact of open-source AID systems in adult users. Our findings are in line with our analysis of children and adolescents using open-source AID systems, including their caregivers [30], although there were age-specific findings, for example, navigating diabetes throughout puberty and remote monitoring and control by caregivers that only applied to the child cohort. Our results also align with other, smaller cohort studies examining the user experience with open-source AID systems [22,24,31] with the literature pointing to the importance of setting expectations for both onboarding and the ongoing use of AID systems [32,33]. Studies of users of commercial AID systems found similar results [34,35]. Furthermore, the sense of community and empowerment, often referred to as "paying it forward," was almost exclusively described in open-source AID system users.

In our approach to analyzing the data, we opted to use 4 categories as the basic framework for our template. This was necessary, first and foremost, because these topics were already framed in the wording of the questions. That said, the findings also underline the somewhat fluid nature of the categories, especially with respect to the physical and emotional impact that open-source AID systems can have. So, while we have sought to tease emotional and physical impacts apart for the purposes of our analysis, our findings ultimately serve to highlight how inexorably bound up they are. This is most intuitively illustrated through the example of sleep, where poor quality sleep inevitably impacts emotional well-being, which may, in turn, also impact glucose levels, both directly and indirectly [36].

Peer Support as a Driver of Innovation and Empowerment

Although there is not the same level of symbiosis between the categories "challenges" and "support," the findings did point to a strong link between the two in the sense that many of the challenges associated with initiating and maintaining the use of an open-source AID system were resolved via support from a wider community of users. The sense of community underpinning the development and diffusion of open-source AID systems and peer support as a key resource for practical but also emotional support were predominant topics in other qualitative studies on the lived experience with open-source AID systems [22,31]. For many, the discovery of a peer group that one could identify and engage with was as important and meaningful as building their AID system.

As has previously been noted [37], digital platforms can provide opportunities for peer support and the exchange of experiential knowledge about living with illness. The importance of peer support for people with diabetes in the context of online communities has been clearly highlighted elsewhere [12,22,23,38,39]. Moreover, engagement in these communities has been shown to positively impact HbA_{1c} levels [23,40]; reduce diabetes-related distress [41]; and foster support and connection, advocacy, self-expression, information and education, technical support, and humor as a coping strategy [38,42].

However, it is also evident that the type of peer support upon which the dissemination of open-source AID systems has been based is of a somewhat different order. In part, this reflects something of the nature of T1D, where the prevailing model of care, inasmuch as it requires individuals to be actively engaged in their care, may potentially foster the growth and dissemination of expertise [43]. In this situation, the delineation of expertise into "professional" and "laypeople" seems outmoded and evokes the well-known Shavian aphorism that the former serves the purpose of conspiring against the latter [44].

As AID algorithms are being constantly developed further and new features will be introduced (eg, fully closed-loop systems without bolusing for meals), future research should address the lived experiences of people with diabetes associated with their use in addition to the analysis of clinical outcomes. Their full health impact can only be evaluated if real-world user experiences are included.

Looking back at how innovations in diabetes treatment were perceived by HCPs and the scientific community over the last century, increased autonomy and empowerment of people with diabetes have continuously been regarded with skepticism. Similarly, it was debated if people with diabetes were capable of blood glucose self-monitoring in the 1970s [45], calculating their insulin doses by themselves, or understanding real-time readings of their CGM device [46]—all aspects that are standard of care today. In the IT age, people with diabetes creating their own technological tools might be just another iteration of patient empowerment but accompanied by similar controversy. The "lesson learned" from these controversies should be the urgency to foster collaboration with patients and involve them early as stakeholders—whether in research and development of medical devices or the development of care concepts that will ultimately affect them.

This study is the first large-scale qualitative study assessing the lived experiences of adult open-source AID system users. Moreover, it is a study with a truly multinational scope, and in

XSL•FO
RenderX

its stakeholder engagement via the involvement of the #WeAreNotWaiting community, it strives to remain true to the values of the phenomenon it is investigating.

Of course it should be noted that the survey was undertaken in 2019, and thus the participants can be considered early adopters of open-source AID systems. At one level, this means that the size of our sample represents what, at that time, was a significant proportion of all users of open-source AID systems. However, by the same token, given the subsequent dynamic technical development and rapid expansion of open-source AID system users since that time, the experiences captured in our study may not be reflective of later experiences and current use of open-source AID systems. In addition, many position papers and an international consensus paper have been published to provide guidance to HCPs who wish to support people using open-source AID systems, which may have contributed to increased knowledge and a change of attitudes [47,48]. A selection bias may be present with the survey only being available in German and English.

Conclusions

The efforts of the #WeAreNotWaiting community are changing the landscape of available treatment options and the way we look at the role of patients as initiators rather than as passive recipients of a health care product or service. The online communities that support this movement have not only transformed diabetes care with its technology but have also eased the individual burden for those involved due to the tools and peer support it has fostered. The extensive testimony from users of open-source AID systems acquired in this study provides new insights, highlighting factors inspiring people to adopt such solutions, user experiences in transitioning to open-source AID systems, and the transformative impact of AID systems on the everyday life of people with diabetes. These results may contribute to a better understanding of their unmet needs; the impact of AID systems on physical and emotional health; and some of the current challenges to the uptake of AID technology in terms of access, availability, and usability.

Acknowledgments

This study is part of the OPEN project. OPEN is funded by the European Commission's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Action Research and Innovation Staff Exchange (grant 823902), the Deutsche Forschungsgemeinschaft (DFG)-funded Digital Clinician Scientist Program of the Berlin Institute of Health, and the SPOKES Wellcome Trust Translational Partnership Program.

Data Availability

The datasets generated and analyzed during this study are available from the OPEN project on request. Researchers interested in working with the OPEN data are encouraged to submit inquiries for data access to the project coordinator (SO) via email. The authors welcome the opportunity to collaborate and accommodate research needs where possible.

Authors' Contributions

BC, KB, SO, AT, KR, DL, and TS designed the study. KB and KR were responsible for administration and ethics approval. KB collected the data. KB, MW, and TS processed the data. BC, YC, MW, HB, DC, and SO analyzed the data. BC, YC, and KB drafted the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
DIWHY Questionnaire for Adults.
[PDF File (Adobe PDF File), 85 KB - jmir_v27i1e48406_app1.pdf]

References

1. Hossain SN, Jaglal SB, Shepherd J, Perrier L, Tomasone JR, Sweet SN, et al. Web-based peer support interventions for adults living with chronic conditions: scoping review. JMIR Rehabil Assist Technol 2021 May 25;8(2):e14321 [FREE Full text] [doi: 10.2196/14321] [Medline: 34032572]
2. Banting FG, Best CH, Collip JB, Campbell WR, Fletcher AA. Pancreatic extracts in the treatment of diabetes mellitus: preliminary report. Can Med Assoc J 1922 Mar;12(3):141-146 [FREE Full text] [Medline: 20314060]
3. Lewis D. History and perspective on DIY closed looping. J Diabetes Sci Technol 2019 Jul;13(4):790-793 [FREE Full text] [doi: 10.1177/1932296818808307] [Medline: 30348013]
4. Lewis D. #WeAreNotWaiting to reduce the burden of Type 1 diabetes. OpenAPS.org. URL: https://openaps.org/ [accessed 2023-04-21]
5. Welcome to the AAPS documentation. Android APS. URL: https://androidaps.readthedocs.io/ [accessed 2023-04-21]
6. Artificial-Pancreas / iAPS. GitHub. URL: https://github.com/Artificial-Pancreas/iAPS [accessed 2023-04-21]

XSL•FO
RenderX

7. What is the Loop app? LoopDocs. URL: https://loopkit.github.io/loopdocs/ [accessed 2023-04-21]
8. Peyser T, Dassau E, Breton M, Skyler JS. The artificial pancreas: current status and future prospects in the management of diabetes. Ann N Y Acad Sci 2014 Apr;1311:102-123. [doi: 10.1111/nyas.12431] [Medline: 24725149]
9. Fang Z, Liu M, Tao J, Li C, Zou F, Zhang W. Efficacy and safety of closed-loop insulin delivery versus sensor-augmented pump in the treatment of adults with type 1 diabetes: a systematic review and meta-analysis of randomized-controlled trials. J Endocrinol Invest 2022 Mar;45(3):471-481. [doi: 10.1007/s40618-021-01674-6] [Medline: 34535888]
10. Farrington C. Psychosocial impacts of hybrid closed-loop systems in the management of diabetes: a review. Diabet Med 2018 Apr;35(4):436-449. [doi: 10.1111/dme.13567] [Medline: 29247547]
11. Pfiester E, Braune K, Thieffry A, Ballhausen H, Gajewska KA, O'Donnell S. Costs and underuse of insulin and diabetes supplies: findings from the 2020 T1International cross-sectional web-based survey. Diabetes Res Clin Pract 2021 Sep;179:108996 [FREE Full text] [doi: 10.1016/j.diabres.2021.108996] [Medline: 34363862]
12. Crocket H. Peer mentoring in the do-it-yourself artificial pancreas system community. J Diabetes Sci Technol 2020 Nov;14(6):1022-1027 [FREE Full text] [doi: 10.1177/1932296819883876] [Medline: 31646890]
13. Burnside MJ, Lewis DM, Crocket H, Meier R, Williman J, Sanders OJ, et al. Open-source automated insulin delivery in type 1 diabetes. N Engl J Med 2022 Sep 08;387(10):869-881. [doi: 10.1056/NEJMoa2203913] [Medline: 36069869]
14. Braune K, O'Donnell S, Cleal B, Lewis D, Tappe A, Willaing I, et al. Real-world use of do-it-yourself artificial pancreas systems in children and adolescents with type 1 diabetes: online survey and analysis of self-reported clinical outcomes. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e14087 [FREE Full text] [doi: 10.2196/14087] [Medline: 31364599]
15. Braune K, Gajewska KA, Thieffry A, Lewis DM, Froment T, O'Donnell S, et al. Why #WeAreNotWaiting-motivations and self-reported outcomes among users of open-source automated insulin delivery systems: multinational survey. J Med Internet Res 2021 Jun 07;23(6):e25409 [FREE Full text] [doi: 10.2196/25409] [Medline: 34096874]
16. Toffanin C, Kozak M, Sumnik Z, Cobelli C, Petruzelkova L. In silico trials of an open-source android-based artificial pancreas: a new paradigm to test safety and efficacy of do-it-yourself systems. Diabetes Technol Ther 2020 Feb;22(2):112-120. [doi: 10.1089/dia.2019.0375] [Medline: 31769699]
17. Melmer A, Züger T, Lewis DM, Leibrand S, Stettler C, Laimer M. Glycaemic control in individuals with type 1 diabetes using an open source artificial pancreas system (OpenAPS). Diabetes Obes Metab 2019 Oct;21(10):2333-2337. [doi: 10.1111/dom.13810] [Medline: 31183929]
18. Lum JW, Bailey RJ, Barnes-Lomen V, Naranjo D, Hood KK, Lal RA, et al. A real-world prospective study of the safety and effectiveness of the loop open source automated insulin delivery system. Diabetes Technol Ther 2021 May;23(5):367-375 [FREE Full text] [doi: 10.1089/dia.2020.0535] [Medline: 33226840]
19. Gawrecki A, Zozulinska-Ziolkiewicz D, Michalak MA, Adamska A, Michalak M, Frackowiak U, et al. Safety and glycemic outcomes of do-it-yourself AndroidAPS hybrid closed-loop system in adults with type 1 diabetes. PLoS One 2021;16(4):e0248965 [FREE Full text] [doi: 10.1371/journal.pone.0248965] [Medline: 33819289]
20. Petruzelkova L, Jiranova P, Soupal J, Kozak M, Plachy L, Neuman V, et al. Pre-school and school-aged children benefit from the switch from a sensor-augmented pump to an AndroidAPS hybrid closed loop: a retrospective analysis. Pediatr Diabetes 2021 Jun;22(4):594-604. [doi: 10.1111/pedi.13190] [Medline: 33576551]
21. Hng T, Burren D. Appearance of Do-It-Yourself closed-loop systems to manage type 1 diabetes. Intern Med J 2018 Nov;48(11):1400-1404. [doi: 10.1111/imj.14105] [Medline: 30387315]
22. Schipp J, Skinner T, Holloway E, Scibilia R, Langstrup H, Speight J, et al. How adults with type 1 diabetes are navigating the challenges of open-source artificial pancreas systems: a qualitative study. Diabetes Technol Ther 2021 Aug;23(8):546-554. [doi: 10.1089/dia.2020.0652] [Medline: 33720767]
23. Litchman ML, Lewis D, Kelly LA, Gee PM. Twitter analysis of #OpenAPS DIY artificial pancreas technology use suggests improved A1C and quality of life. J Diabetes Sci Technol 2019 Mar;13(2):164-170 [FREE Full text] [doi: 10.1177/1932296818795705] [Medline: 30198751]
24. Marshall DC, Holloway M, Korer M, Woodman J, Brackenridge A, Hussain S. Do-It-Yourself artificial pancreas systems in type 1 diabetes: perspectives of two adult users, a caregiver and three physicians. Diabetes Ther 2019 Oct;10(5):1553-1564. [doi: 10.1007/s13300-019-00679-y] [Medline: 31440989]
25. Vallis M, Holt RI. User-driven open-source artificial pancreas systems and patient-reported outcomes: a missed opportunity? Diabet Med 2022 May;39(5):e14797. [doi: 10.1111/dme.14797] [Medline: 35092089]
26. O'Donnell S, Lewis D, Marchante Fernández M, Wäldchen M, Cleal B, Skinner T, et al. Evidence on user-led innovation in diabetes technology (the OPEN project): protocol for a mixed methods study. JMIR Res Protoc 2019 Nov 19;8(11):e15368 [FREE Full text] [doi: 10.2196/15368] [Medline: 31742563]
27. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
29. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 03;12(2):202-222 [FREE Full text] [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]

XSL•FO
RenderX

30. Braune K, Krug N, Knoll C, Ballhausen H, Thieffry A, Chen Y, et al. Emotional and physical health impact in children and adolescents and their caregivers using open-source automated insulin delivery: qualitative analysis of lived experiences. J Med Internet Res 2022 Jul 14;24(7):e37120 [FREE Full text] [doi: 10.2196/37120] [Medline: 35834298]
31. Suttiratana SC, Wong JJ, Lanning MS, Dunlap AR, Hanes SJ, Hood KK, et al. Qualitative study of user experiences with loop, an open-source automated insulin delivery system. Diabetes Technol Ther 2022 Jun;24(6):416-423 [FREE Full text] [doi: 10.1089/dia.2021.0485] [Medline: 35099278]
32. Lewis D. Automated Insulin Delivery in Real Life (AID-IRL): real-world user perspectives on commercial AID. J Diabetes Sci Technol 2022 Mar;16(2):500-503 [FREE Full text] [doi: 10.1177/1932296820957360] [Medline: 32935554]
33. Lewis D. Setting expectations for successful artificial pancreas/hybrid closed loop/automated insulin delivery adoption. J Diabetes Sci Technol 2018 Mar;12(2):533-534 [FREE Full text] [doi: 10.1177/1932296817730083] [Medline: 28918653]
34. Hendrieckx C, Poole LA, Sharifi A, Jayawardene D, Loh MM, Horsburgh JC, et al. "It is definitely a game changer": a qualitative study of experiences with in-home overnight closed-loop technology among adults with type 1 diabetes. Diabetes Technol Ther 2017 Jul;19(7):410-416. [doi: 10.1089/dia.2017.0007] [Medline: 28537437]
35. Barnard KD, Wysocki T, Thabit H, Evans ML, Amiel S, Heller S, Angela Consortium. Psychosocial aspects of closed- and open-loop insulin delivery: closing the loop in adults with Type 1 diabetes in the home setting. Diabet Med 2015 May;32(5):601-608. [doi: 10.1111/dme.12706] [Medline: 25615888]
36. Nefs GM, Bazelmans E, Donga E, Tack CJ, de Galan BE. Sweet dreams or bitter nightmare: a narrative review of 25 years of research on the role of sleep in diabetes and the contributions of behavioural science. Diabet Med 2020 Mar;37(3):418-426. [doi: 10.1111/dme.14211] [Medline: 31833083]
37. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
38. Tenderich A, Tenderich B, Barton T, Richards SE. What are PWDs (people with diabetes) doing online? A netnographic analysis. J Diabetes Sci Technol 2019 Mar;13(2):187-197 [FREE Full text] [doi: 10.1177/1932296818813192] [Medline: 30477335]
39. Reidy C, Klonoff DC, Barnard-Kelly KD. Supporting good intentions with good evidence: how to increase the benefits of diabetes social media. J Diabetes Sci Technol 2019 Sep;13(5):974-978 [FREE Full text] [doi: 10.1177/1932296819850187] [Medline: 31096768]
40. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 Apr 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]
41. de Vries L, van der Heijden AA, van 't Riet E, Baan CA, Kostense PJ, Rijken M, et al. Peer support to decrease diabetes-related distress in patients with type 2 diabetes mellitus: design of a randomised controlled trial. BMC Endocr Disord 2014 Mar 04;14:21 [FREE Full text] [doi: 10.1186/1472-6823-14-21] [Medline: 24593296]
42. Hilliard ME, McQuaid EL, Nabors L, Hood KK. Resilience in youth and families living with pediatric health and developmental conditions: introduction to the special issue on resilience. J Pediatr Psychol 2015 Oct;40(9):835-839 [FREE Full text] [doi: 10.1093/jpepsy/jsv072] [Medline: 26275974]
43. Gottlieb SD, Cluck J. "Going rogue": re-coding resistance with type 1 diabetes. Digit Cult Soc 2018;4(2):137-156 [FREE Full text] [doi: 10.14361/dcs-2018-0208]
44. Ashton JR. All professions are conspiracies against the laity? (George Bernard-Shaw). J Epidemiol Community Health 2003:161. [doi: 10.1136/jech.57.3.161-a]
45. Bernstein RK. Blood glucose self-monitoring by diabetic patients: refinements of procedural technique. Diabetes Care 1979;2(2):233-236. [doi: 10.2337/diacare.2.2.233] [Medline: 520127]
46. De Block C, Manuel-y-Keenoy B, Van Gaal L. A review of current evidence with continuous glucose monitoring in patients with diabetes. J Diabetes Sci Technol 2008 Jul;2(4):718-727 [FREE Full text] [doi: 10.1177/193229680800200426] [Medline: 19885251]
47. Braune K, Lal RA, Petruželková L, Scheiner G, Winterdijk P, Schmidt S, OPEN International Healthcare Professional Network and OPEN Legal Advisory Group. Open-source automated insulin delivery: international consensus statement and practical guidance for health-care professionals. Lancet Diabetes Endocrinol 2022 Jan;10(1):58-74 [FREE Full text] [doi: 10.1016/S2213-8587(21)00267-9] [Medline: 34785000]
48. Gawrecki A, Klupa T, Araszkiewicz A, Matejko B, Szadkowska A, Wolnik B, et al. Utilization of do-it-yourself artificial pancreas systems in the management of patients with type 1 diabetes: a position statement of the Pump School Education Initiative by Diabetes Poland. Pol Arch Intern Med 2019 Feb 28;129(2):141-142 [FREE Full text] [doi: 10.20452/pamw.4468] [Medline: 30816103]

Abbreviations

AID: automated insulin delivery
CGM: continuous glucose monitor

XSL•FO
RenderX

HbA_{1C}: glycated hemoglobin
HCP: health care professional
OPEN: Outcomes of Patient's Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes

Edited by A Mavragani; submitted 21.04.23; peer-reviewed by H Crocket, I Degen, M Nagpal; comments to author 20.09.23; revised version received 30.11.23; accepted 30.05.24; published 06.01.25.

Please cite as:
Cleal B, Chen Y, Wäldchen M, Ballhausen H, Cooper D, O'Donnell S, Knoll C, Krug N, Raile K, Ubben T, Tappe A, Lewis D, Willaing I, Skinner T, Braune K
Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives
J Med Internet Res 2025;27:e48406
URL: https://www.jmir.org/2025/1/e48406
doi:10.2196/48406
PMID:

©Bryan Cleal, Yanbing Chen, Mandy Wäldchen, Hanne Ballhausen, Drew Cooper, Shane O'Donnell, Christine Knoll, Niklas Krug, Klemens Raile, Tebbe Ubben, Adrian Tappe, Dana Lewis, Ingrid Willaing, Timothy Skinner, Katarina Braune. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Unraveling Online Mental Health Through the Lens of Early Maladaptive Schemas: AI-Enabled Content Analysis of Online Mental Health Communities

Beng Heng Ang[1], BEng; Sujatha Das Gollapalli[2], PhD; Mingzhe Du[2], MCIT; See-Kiong Ng[2], PhD

[1]Integrative Sciences and Engineering Programme, NUS Graduate School, National University of Singapore, Singapore, Singapore
[2]Institute of Data Science, National University of Singapore, Singapore, Singapore

Corresponding Author:
Beng Heng Ang, BEng
Integrative Sciences and Engineering Programme
NUS Graduate School
National University of Singapore
University Hall, Tan Chin Tuan Wing Level 5
#05-03 21 Lower Kent Ridge Road
Singapore, 119077
Singapore
Phone: 65 92983451
Email: bengheng.ang@u.nus.edu

Abstract

Background: Early maladaptive schemas (EMSs) are pervasive, self-defeating patterns of thoughts and emotions underlying most mental health problems and are central in schema therapy. However, the characteristics of EMSs vary across demographics, and despite the growing use of online mental health communities (OMHCs), how EMSs manifest in these online support-seeking environments remains unclear. Understanding these characteristics could inform the design of more effective interventions powered by artificial intelligence to address online support seekers' unique therapeutic needs.

Objective: We aimed to uncover associations between EMSs and mental health problems within OMHCs and examine features of EMSs as they are reflected in OMHCs.

Methods: We curated a dataset of 29,329 posts from widely accessed OMHCs, labeling each with relevant schemas and mental health problems. To identify associations, we conducted chi-square tests of independence and calculated odds ratios (ORs) with the dataset. In addition, we developed a novel group-level case conceptualization technique, leveraging GPT-4 to extract features of EMSs from OMHC texts across key schema therapy dimensions, such as *schema triggers* and *coping responses*.

Results: Several associations were identified between EMSs and mental health problems, reflecting how EMSs manifest in online support-seeking contexts. Anxiety-related problems typically highlighted *vulnerability to harm or illness* (OR 5.64, 95% CI 5.34-5.96; *P*<.001), while depression-related problems emphasized unmet interpersonal needs, such as *social isolation* (OR 3.18, 95% CI 3.02-3.34; *P*<.001). Conversely, problems with eating disorders mostly exemplified negative self-perception and *emotional inhibition* (OR 1.89, 95% CI 1.45-2.46; *P*<.001). Personality disorders reflected themes of *subjugation* (OR 2.51, 95% CI 1.86-3.39; *P*<.001), while posttraumatic stress disorder problems involved distressing experiences and *mistrust* (OR 5.04, 95% CI 4.49-5.66; *P*<.001). Substance use disorder problems reflected negative self-perception of *failure to achieve* (OR 1.83, 95% CI 1.35-2.49; *P*<.001). Depression, personality disorders, and posttraumatic stress disorder were also associated with 12, 9, and 7 EMSs, respectively, emphasizing their complexities and the need for more comprehensive interventions. In contrast, anxiety, eating disorder, and substance use disorder were related to only 2 to 3 EMSs, suggesting that these problems are better addressed through targeted interventions. In addition, the EMS features extracted from our dataset averaged 13.27 (SD 3.05) negative features per schema, with 2.65 (SD 1.07) features per dimension, as supported by existing literature.

Conclusions: We uncovered various associations between EMSs and mental health problems among online support seekers, highlighting the prominence of specific EMSs in each problem and the unique complexities of each problem in terms of EMSs. We also identified EMS features as expressed by support seekers in OMHCs, reinforcing the relevance of EMSs in these online support-seeking contexts. These insights are valuable for understanding how EMS are characterized in OMHCs and can inform the development of more effective artificial intelligence–powered tools to enhance support on these platforms.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e59524) doi:10.2196/59524

KEYWORDS

early maladaptive schemas; large language models; online mental health communities; case conceptualization; prompt engineering; artificial intelligence; AI

Introduction

Background

Over recent years, mental health problems have become increasingly prevalent worldwide [1-3]. However, access to professional therapy remains limited for many individuals due to high treatment costs and a shortage of trained mental health professionals [4-6]. As a result, a growing number of individuals are seeking support for their problems through online mental health communities (OMHCs) [7-9]. These communities provide a more accessible alternative to professional therapy [10,11], allowing individuals to share their distressing experiences and receive peer support to help them cope [11,12]. These problems, typically shared through posts within OMHCs, not only reveal support seekers' struggles but also reflect the underlying dysfunctional beliefs contributing to their distress. These insights are crucial in shaping more personalized and effective online interventions for support seekers [13,14].

A set of dysfunctional beliefs commonly present in OMHC posts is early maladaptive schemas (EMSs) [15,16]. Central to schema therapy (ST), EMSs are pervasive, self-defeating beliefs formed in childhood and often persist into adulthood. Young et al [15] identified 18 distinct types of EMSs (Table 1) that distort worldviews and are typically associated with mental health problems, such as anxiety [17-21]. Understanding these associations provides a more reliable method for anticipating potential schemas based on a mental health diagnosis, enabling proactive and targeted therapeutic interventions [15,22,23]. Similarly, recognizing the unique features of each schema (eg, "I'm incompetent when it comes to achievement," corresponding to the *failure to achieve* schema) helps inform targeted interventions (eg, setting small, realistic goals to foster a sense of accomplishment). By addressing these schemas, these interventions can enhance therapy effectiveness and improve treatment outcomes.

However, the characteristics of these EMSs can vary across demographic groups. For example, Cámara and Calvete [18] identified a strong association between the *failure* schema and depression in Spanish university students. However, Wright et al [24] did not report such an association in their study of a US cohort. Similarly, Balsamo et al [25] did not observe a relationship between the *failure* schema and depression among young adults in Italy. Likewise, OMHCs likely present a demographically distinct population of online support seekers reflecting unique EMS characteristics [18,24-26]. Gaining a clearer understanding of these characteristics could provide valuable insights for policy making and the design of more effective digital therapeutic tools for this online support-seeking population. However, how EMSs are reflected in these communities remains largely unexplored, despite their growing importance in mental health support.

This research gap largely stems from the lack of tools for accurately identifying EMSs in the asynchronous, text-based environments of OMHCs. In professional practice, therapists use the Young Schema Questionnaire (YSQ) [27] to assess patients' EMSs through clinically validated statements of dysfunctional beliefs, which patients rate on a scale from 1 ("completely untrue of me") to 6 ("describes me perfectly"). Replicating this structured assessment in OMHCs is challenging due to the unstructured nature of online support-seeking interactions [15,28,29]. Overcoming this challenge requires interpreting EMSs directly from the text, which necessitates specialized natural language processing techniques to identify EMSs in OMHC posts. Recent advancements in large language models (LLMs) have demonstrated remarkable performance in complex language tasks, such as text classification, retrieval, and generation, which require nuanced interpretation of unstructured texts [30-32]. These models can capture subtle meanings in textual discourse, such as recognizing that the phrase "dark tunnel" in "I just feel like I'm stuck in a dark tunnel, and there's no way out" metaphorically conveys despair, isolation, and hopelessness [33]. LLMs have also been used to predict mental health problems, such as depression, from online conversations [34-38] and have recently been applied to predict EMSs in OMHC posts [16].

XSL•FO
RenderX

Table 1. The taxonomy proposed by Young et al [15] of 18 early maladaptive schemas (EMSs) with representative examples from the Young Schema Questionnaire (YSQ) [27]. EMSs represent cognitive and emotional patterns that shape perception and behavior, often contributing to mental health problems.

EMS	Description	Examples from YSQ
Abandonment/instability	The perception that others will not be reliable enough to provide emotional support or connection, leading to a sense of instability and insecurity	"I worry that people I feel close to will leave me or abandon me."
Approval seeking/recognition-seeking	A constant desire for approval, recognition, or acceptance from others to the extent that one compromises their authentic self	"Unless I get a lot of attention from others, I feel less important."
Defectiveness/shame	A deep-seated belief that one is flawed, unlovable, or inferior, leading to feelings of shame and inadequacy	"I cannot understand how anyone could love me."
Dependence/incompetence	A belief in one's inability to handle daily responsibilities or make appropriate decisions independently	"I feel that I need someone I can rely on to give me advice about practical issues."
Emotional deprivation	The belief that one's emotional needs will not be adequately met by others	"For much of my life, I haven't felt that I am special to someone."
Emotional inhibition	The suppression or inhibition of emotions, often due to a fear of losing control or being overwhelmed by intense feelings	"People see me as uptight emotionally."
Enmeshment/undeveloped self	An overly intense emotional attachment and proximity to one or more important individuals (typically parents), which hinders the process of becoming an independent individual and impedes normal social development	"I often found myself constantly turning to my parents for approval in every decision I made."
Entitlement/grandiosity	An exaggerated sense of self-importance and a belief that one deserves special treatment	"I get very irritated when people won't do what I ask of them."
Failure to achieve	The belief that one has failed or will inevitably fail in achieving personal and practitioner goals, leading to a sense of underachievement	"I am humiliated by my failures and inadequacies in the work (or school) sphere."
Insufficient self-control/self-discipline	An inability to control impulses, leading to difficulty in achieving long-term goals.	"I often do things impulsively that I later regret."
Mistrust/abuse	A belief that others are likely to deceive, take advantage of, or harm the person emotionally or physically	"I have been physically, emotionally, or sexually abused by important people in my life."
Negativity/pessimism	A pervasive focus on the negative aspects of life while neglecting the positive aspects	"People close to me consider me a worrier."
Punitiveness	The belief that harsh punishment is deserved for perceived mistakes or failures	"I am a bad person who deserves to be punished."
Self-sacrifice	The belief that one must meet the needs of others at the expense of one's own needs and desires	"No matter how much I give, I feel it is never enough."
Social isolation/alienation	The belief that one is different from others, leading to a sense of isolation and difficulty connecting with others	"I always feel on the outside of groups."
Subjugation	A tendency to prioritize others' needs and desires over one's own because one feels coerced	"I let other people have their way, because I fear the consequences."
Unrelenting standards/hypercriticalness	The belief that one must meet high standards of performance, usually to avoid criticism	"I can't let myself off the hook easily or make excuses for my mistakes."
Vulnerability to harm or illness	The belief that catastrophic events are imminent, leading to excessive worry and anxiety about potential harm or illness	"I feel that the world is a dangerous place."

Objectives

Using the EMS classifiers from the recent study on EMS prediction [16] and leveraging recent LLMs, we can identify EMS labels within large volumes of unstructured OMHC texts. This enables us to (1) uncover associations between EMSs and mental health problems within OMHCs and (2) examine how features of EMSs are reflected in online support-seeking populations. These insights will guide the development of online interventions powered by artificial intelligence (AI) and grounded in ST, tailored to meet the unique therapeutic needs of online support seekers [15].

XSL•FO
RenderX

Methods

Curating the Dataset and Identifying EMSs

We curated posts shared by support seekers in 3 highly frequented OMHCs, namely 7-Cups [39], Beyond-Blue [40], and Patient [41], from 2013 to November 2023. Using a custom Python-based web crawler developed for this study, we collected posts from subcommunities focusing on anxiety, depression, posttraumatic stress disorder (PTSD), personality disorders (PDs), eating disorders (EDs), and substance use disorders (SUDs). These mental health problems are prevalent among adults [42,43] and are often treated with ST [15].

To identify EMSs in OMHC posts, we used the entailment-based prediction model (EPM) from a recent study [16]. EPM predicts EMSs from a post by leveraging textual entailment, a sentence-level inference technique from natural language processing [16]. In textual entailment, a statement entails a hypothesis *h* if a human reader would infer that *h* is likely true based on *t* [44]. EPM computes the entailment between sentences in a post and statements from the YSQ. When a sentence entails a YSQ statement, the latter serves as an "explanation" for assigning the corresponding EMS label. For instance, the sentence "Since then I either try to find excuses not to ride or when I do try to ride, I have to push myself and have massive panic attacks." entails the YSQ statement "I often feel that I'm going to have an anxiety attack." This YSQ statement corresponds to the *vulnerability to harm or illness* schema, and the EPM labels the post with this schema. We used the Text-To-Text Transfer Transformer (Google; T5-LLM [45]) in EPM to compute entailment and assign EMS labels to each post.

The 18 EMSs in Table 1 are each represented as a binary categorical variable, where a value *1* indicates the relevance of a schema to an OMHC post, and *0* indicates nonrelevance. In a similar way, we labeled each post with 6 binary categorical variables, each corresponding to one of the mental health problems examined in this study.

Uncovering Associations Between EMSs and Mental Health Problems

Using these categorical variables, we constructed 2×2 contingency tables (Table S1 in Multimedia Appendix 1) and calculated the expected frequency, E_{ij} (equation S1 in Multimedia Appendix 1), and the χ^2 statistic (equation S2 in Multimedia Appendix 1) for each pair of schema and mental health problem [46]. Each χ^2 value was compared to the critical value from the χ^2 distribution table with 1 *df* and a significance level of 0.05. For each χ^2 value that exceeded the critical value, we rejected the null hypothesis to indicate a statistically significant association between a schema and mental health problem. To quantify the strength of association for each statistically significant pair, we calculated the odds ratio (OR), OR_{ems_mhp} (equation S3 in Multimedia Appendix 1), and its 95% CI, $OR_{ems\_mhp\_95\%CI}$ (equation S4 in Multimedia Appendix 1) [47,48]. Following established guidelines for interpreting ORs [48], we considered an association inconclusive if the 95% CI of the OR included the null value of 1.

Examining Features of EMSs With Group-Level Case Conceptualization

We further examined how online support seekers express their schemas by analyzing specific features reflected in their online posts. For this purpose, we used the case conceptualization technique, which is commonly used by therapists to identify key features related to patients' beliefs and mental health problems, thus facilitating personalized therapeutic interventions. In ST, case conceptualization focuses on several dimensions, including *schema triggers* (situations that activate the schema), *emotions* (associated feelings), *negative thoughts* (irrational beliefs), *coping responses* (behaviors and reactions), and *bodily sensations* (physical and physiological experiences) [15].

We adapted this case conceptualization to our dataset using a data-driven AI approach, which we refer to as group-level case conceptualization. First, we prepared the data by using the EPM to identify sentences from OMHC posts that entailed specific YSQ statements, as outlined in the Curating the Dataset and Identifying EMSs section. These sentences reflected the dysfunctional beliefs and respective schemas established in the YSQ. We grouped these sentences according to their schema, resulting in a distribution of sentence counts by EMSs, which ranged from 8494 for *social isolation* to 98 for *subjugation*. To derive group-level features for each schema, we randomly sampled 10% of the sentences (in Multimedia Appendix 2, we demonstrate that the similarity scores between features extracted from different groups of randomly sampled sentences consistently average 0.79 (SD 0.09), which indicates substantial overlap in the features generated across groups. This validates the approach of conducting in-depth analysis on a single group of randomly sampled sentences for each schema and instructed GPT-4 (OpenAI; we used the model gpt-4-1106-preview for GPT-4) [32] through a custom prompt (Textbox 1) to extract common features among the sentences and categorize them along the dimensions of case conceptualization.

XSL•FO
RenderX

Textbox 1. Prompt used to instruct GPT-4 to extract common features from groups of randomly sampled sentences for each schema.

Identify every common features exhibited from the given sentences. The features should belong to one of the following dimension (a) Schema Triggers (What are the situations that trigger the emotions, negative thoughts, coping responses, and bodily sensations?) (feature should start with: Situation where...), (b) Emotions (What are the feelings involved?) (feature should start with: Feeling...), (c) Negative thoughts (What irrational thoughts are going through the mind?) (feature should start with: Thinking that...), (d) Coping Responses (What are the behaviors and reactions?) (feature should start with: Responding by...), (e) Bodily Sensations (What are the physical and physiological experiences with the body?) (feature should start with: Experiencing...). Provide more than 1 relevant samples from the given sentences for each feature.

If a dimension has feature(s), the output should be in the format as this example:

(d) Coping Responses:

** Destructive or impulsive actions when upset.

• "When I'm angry or frustrated, I sometimes lose control of my emotions and destroy things."

• "Im 22, and I have this problem when I get angry I can't control myself I yell hit stuff around me and often I hurt myself."

If a dimension does not have any feature, the output should be in the format as this example:

(d) Coping Responses:

**NO_FEATURES.

Given sentences:

Ethical Considerations

This study aligns with NUS-IRB Review Not Required guidelines by analyzing nonindividually identifiable data from OMHCs to explore the characteristics of EMSs. Specifically, we used textual content that was already anonymized through usernames and we further decoupled these pseudonyms and other identifying information from the data. In addition, the analysis focused exclusively on examining publicly available data, with no interactions with support seekers. This approach minimized the risks of introducing harmful or unsupportive content while also adhering to ethical guidelines and the responsible use of these online platforms [49,50].

Results

Dataset Statistics

Our dataset comprised 29,329 online posts in which support seekers shared their experiences with various mental health problems. Anxiety was the most frequently mentioned problem, appearing in 51.56% (15,149/20,329) of the posts, followed by depression, which appeared in 33.98% (9966/20,329) of the posts. PTSD, SUDs, PDs, and EDs correspondingly represented 8.75% (2565/20,329), 1.96% (576/20,329), 1.87% (548/20,329), and 1.79% (525/20,329) of the dataset. This distribution aligned with general sociodemographic trends in the prevalence of mental health problems. For example, anxiety is typically more prevalent than depression, while both are more common than PTSD, PDs, and EDs [42,51]. This suggests that the mental health problems discussed in OMHCs may reflect their real-world prevalence, suggesting that policies based on these trends could be effectively applied to online settings.

Regarding post length, the median sentence count was 10 (IQR 5-16). Word count showed similar variability, with a median of 214 words (IQR 124-356). This variability likely reflected the open nature of online support forums, where individuals can express themselves freely, unconstrained by the structured environments of in-person therapy. Unlike therapy sessions, where therapists guide conversations and encourage concise, focused responses, OMHCs allow support seekers to share their experiences through detailed, unfiltered narratives. Consequently, the dataset included both succinct and elaborate descriptions of mental distress.

As illustrated in Table 2, the distribution of EMSs, as predicted by EPM [16], was nonuniform across the dataset. Posts more frequently involved schemas related to *social isolation* and *vulnerability to harm or illness* than schemas such as *approval seeking* and *subjugation*. We hypothesize that this skewed distribution resulted from a combination of societal, environmental, biological, and psychological factors. For instance, modern societal pressures, pervasive social media use, and global crises, such as the COVID-19 pandemic, may amplify feelings of isolation and vulnerability [52,53]. In contrast, schemas such as *approval seeking* and *subjugation* may be less common due to cultural factors, including the emphasis on individualism and self-reliance in many societies [54].

XSL•FO
RenderX

Table 2. Distribution of early maladaptive schemas (EMSs) in posts on mental health problems shared by support seekers on 7-Cups, Beyond-Blue, and Patient, between 2013 and 2023 (N=29,329). The distribution is uneven, with certain schemas being more prevalent, reflecting the diverse emotional and cognitive patterns of individuals seeking support in online mental health communities.

EMSs	Posts with EMSs, n (%)
Social isolation/alienation	12820 (43.71)
Vulnerability to harm or illness	10191 (34.75)
Dependence/incompetence	5414 (18.46)
Abandonment/instability	3736 (12.74)
Enmeshment/undeveloped self	2942 (10.03)
Negativity/pessimism	2777 (9.47)
Defectiveness/shame	2451 (8.36)
Emotional inhibition	2006 (6.84)
Emotional deprivation	1916 (6.53)
Mistrust/abuse	1641 (5.6)
Insufficient self-control/self-discipline	1578 (5.38)
Self-sacrifice	1530 (5.22)
Failure to achieve	1316 (4.49)
Unrelenting standards/hypercriticalness	1301 (4.44)
Punitiveness	839 (2.86)
Entitlement/grandiosity	487 (1.66)
Approval seeking/recognition seeking	475 (1.62)
Subjugation	276 (0.94)

Associations Between EMSs and Mental Health Problems

Among all mental health problems, depression had the highest number of statistically significant associations with 12 out of the 18 EMSs (refer to Table S1 in Multimedia Appendix 3 for specific *P* values, ORs, and 95% CI of these associations). This was followed by PDs and PTSD, which had 9 and 7 associations, respectively. In contrast, anxiety, EDs, and SUDs showed fewer associations, with only 2 to 3 EMSs associated with these problems.

In Table S1 in Multimedia Appendix 3, we present the strength of the association between specific schemas and particular mental health problems using ORs. Consistent with previous works [55,56], we used OR<1.5 to indicate small effect sizes (weak associations), OR<2.5 to indicate moderate effect sizes (moderate associations), OR<4.0 to indicate large effect sizes (strong associations), and OR>4.0 to indicate very large effect sizes (very strong associations). Higher ORs suggest stronger associations, indicating that certain EMSs are relatively more likely to be present in specific mental health problems observed in OMHCs compared to others.

For anxiety, the strongest association was observed with the *vulnerability to harm or illness* schema, reflecting a very large effect size. In contrast, *insufficient self-control* and *emotional inhibition* schemas exhibited weak associations with anxiety.

In the case of depression, schemas with strong associations included *social isolation*, *approval-seeking*, and *emotional deprivation*. Moderate associations were observed for *abandonment*, *defectiveness*, *dependence*, *failure to achieve*, and *negativity*. Meanwhile, *unrelenting standards*, *entitlement*, *punitiveness*, and *self-sacrifice* schemas showed weak associations.

For EDs, moderate associations were observed with the *emotional inhibition*, *defectiveness*, and *self-sacrifice* schemas.

In PDs, the *subjugation* schema exhibited a very strong association, followed by moderate associations for the *abandonment*, *entitlement*, *enmeshment*, *defectiveness*, *punitiveness*, and *mistrust* schemas. Weak associations were observed for *emotional deprivation* and *social isolation*.

PTSD showed a very strong association with the *mistrust* schema and strong associations with *punitiveness* and *subjugation*. The *emotional deprivation* schema demonstrated a moderate association with PTSD, while weak associations were observed for *enmeshment*, *self-sacrifice*, and *abandonment*.

For SUDs, a moderate association was observed with the *failure to achieve* schema, while the *dependence* schema exhibited a weak association.

Features of EMSs Extracted From Group-Level Case Conceptualization

Features of each schema were identified using GPT-4, as detailed in Multimedia Appendix 4. On average, our group-level case conceptualization approach identified approximately 13 features per schema across the 5 dimensions of case conceptualization. *Schema triggers*, *emotions*, and *negative*

XSL•FO
RenderX

thoughts each comprised approximately 3 features, followed by *coping responses* with about 2 features, and *bodily sensations* with 1 feature. Most features conveyed negative connotations. For instance, individuals with the *abandonment* schema often perceived themselves as burdens to others and frequently reported feelings of loneliness, anxiety, and emptiness.

These features also highlighted the unique characteristics of each schema. For instance, the *punitiveness* schema, which centers on self-blame, involved features, such as "situations where there is a fear of repeating past mistakes or events," "feelings of guilt or shame over past actions or events," "thinking that one cannot forgive oneself or others for past wrongdoings," and "responding with self-harm or destructive behavior." These features are elaborated on in the Discussion section.

Discussion

Principal Findings

Associations With EMSs Reflect Varying Complexities Among Mental Health Problems

The associations identified in this study highlight the varying complexities of mental health problems among online support-seeking populations. Depression exhibited numerous associations with EMSs, suggesting a considerable overlap with depression symptomatology. This indicates that EMSs may provide a valuable framework for understanding the diverse factors influencing the experience of depression, underscoring its multifaceted nature [57-60]. Similarly, the large number of associations between EMSs and either PDs or PTSD highlight the heterogeneous and complex nature of these mental health problems, emphasizing the need for personalized and comprehensive therapeutic approaches.

In contrast, fewer associations were observed between EMS and anxiety, EDs, and SUDs, respectively. This suggests more uniform schema-related underpinnings for these problems, suggesting that individuals with anxiety, EDs, or SUDs may present a more consistent and predictable symptomatology. Consequently, treatment strategies for these conditions could focus on targeting specific schema patterns rather than adopting broader, more comprehensive approaches.

The Prominence of "Vulnerability to Harm or Illness" in Anxiety Problems

The association observed between anxiety and *vulnerability to harm or illness* schema aligns with findings from previous studies, which have identified this schema in groups with higher anxiety levels, such as individuals affected by COVID-19 and college students [17,18,24,61,62]. The very strong association found in this study suggests that individuals seeking support for anxiety in OMHCs are very likely to experience and express heightened feelings of vulnerability. This observation aligns with the core purpose of OMHCs: creating a supportive safe space for individuals to share their challenges and vulnerabilities.

In contrast, associations between anxiety and the *insufficient self-control* and *emotional inhibition* schemas, though consistent with previous offline studies [62-69], were weak in OMHCs. One possible explanation is that individuals participating in OMHCs may exhibit greater self-discipline, as their active participation reflects motivation for change and growth [70]. In addition, the anonymity of OMHCs likely encourages more open disclosure, mitigating tendencies for emotional inhibition [12].

Overall, these findings suggest that anxiety-related problems discussed in OMHCs are prominently characterized by perceptions of vulnerability to harm and danger rather than deficits in self-control or emotional inhibition. This highlights the central role of vulnerability in the experiences of online support seekers with anxiety and underscores the importance of addressing these feelings in the development of online peer support interventions.

The Prominence of "Prolonged Unmet Interpersonal Needs" in Depression Problems

The associations between depression and schemas of *social isolation*, *approval seeking*, and *emotional deprivation* are consistent with previous research, suggesting that individuals with depression often experience social exclusion, seek external validation, and report unmet emotional needs [71-75]. These schemas showed strong associations in OMHCs, suggesting these schemas are more likely to appear and are central concerns for individuals seeking support for depression-related problems.

In contrast, schemas such as *abandonment*, *defectiveness*, *dependence*, *failure to achieve*, and *negativity* exhibited moderate associations, consistent with studies relating depression to struggles with fears of abandonment, low self-worth, reliance on others for emotional support, and pervasive pessimism [71,76-79]. These schemas are likely to appear in OMHC discussions on depression but may not dominate the discourse to the same extent as those with strong associations.

In contrast, schemas such as *unrelenting standards*, *entitlement*, *punitiveness*, and *self-sacrifice* demonstrated weak associations. These schemas are less likely to emerge in OMHC discourse on depression, although these schemas have been identified as relevant to depression in previous studies [74,77,79].

The varying strengths of association between these schemas indicate a focus on prolonged unmet interpersonal needs in depression-related discourse. Strongly associated schemas, such as *social isolation*, *approval seeking*, and *emotional deprivation*, highlight long-term difficulties in forming social and emotional connections. In addition, the *abandonment*, *defectiveness*, *dependence*, *failure to achieve*, and *negativity* schemas reflect interpersonal aspects, although seemingly primarily on the implications of unmet interpersonal needs. Indeed, depression is known to exacerbate interpersonal challenges, further intensifying fears of rejection, lowered self-esteem, interpersonal dependency, and dissatisfaction with both self and life [80-82]. In contrast, weakly associated schemas, such as *unrelenting standards* and *entitlement*, are less focused on unmet interpersonal needs and seem more related to self-directed or situational aspects of depression.

XSL•FO
RenderX

The Prominence of "Self-Neglect and Negative Self-Perception" in ED Problems

The association between the *emotional inhibition* schema and EDs aligns with previous research suggesting that individuals with disordered eating behaviors often use these behaviors to avoid confronting distressing emotions [83-85]. This avoidance hinders healthy emotional processing and could contribute to neglecting psychological well-being.

The *defectiveness* schema reflects deep-seated feelings of inadequacy and self-criticism, which are particularly prevalent in individuals with binge eating behaviors. These individuals often resort to binge eating as a coping mechanism to manage pervasive feelings of shame, self-criticism, and low self-worth [86-88].

The *self-sacrifice* schema reflects a tendency to prioritize others' needs over one's own well-being. This behavior is closely related to pathological altruism, which involves irrational behaviors aimed at promoting the welfare of others but with negative consequences for the individual [89]. In the context of EDs, previous studies have found that individuals engage in self-sacrifice driven by pathological altruism to cope with negative self-perception of inadequacy and defectiveness, often stemming from their maladaptive eating behaviors [90-93].

These schemas demonstrated moderate associations with EDs in OMHCs, suggesting that themes of self-neglect and negative self-perception are likely to emerge from the distress experienced by individuals seeking support for their EDs within these communities.

The Prominence of "Subjugation" in PD Problems

The strong association between the *subjugation* schema and PDs observed in this study aligns with findings from previous research on avoidant PD and borderline PD (BPD) [19,94,95]. This association suggests that individuals discussing PD-related challenges in OMHCs are more likely to articulate struggles related to feeling oppressed by others. These struggles are often accompanied by fears of retaliation, humiliation, or rejection when expressing their true feelings and needs [15]. These fears can hinder their ability to assert themselves, establish boundaries, and manage conflicts effectively.

Moderate associations were uncovered between PDs and schemas such as *abandonment*, *entitlement*, *defectiveness*, *mistrust*, *enmeshment*, and *punitiveness*. While these schemas likely appear in discussions about PDs, they are less prominent compared to the *subjugation* schema. The *abandonment*, *mistrust*, *defectiveness*, and *enmeshment* schemas are frequently observed in individuals with BPD [19,96-98], whereas *entitlement* is more frequently seen in individuals with narcissistic PD [99], and *punitiveness* is associated with both BPD and avoidant PD [19,96].

This study also found weak associations between PDs and schemas such as *emotional deprivation* and *social isolation*. While consistent with earlier studies on BPD [19,21,100], these findings suggest that these schemas are less emphasized in discussions about PDs within OMHCs.

Overall, themes of *subjugation* emerge as a central focus in PD problems discussed in OMHCs. These themes highlight patterns rooted in power dynamics, emotional oppression, and interpersonal dependency. They appear to overlap with other schemas, such as *abandonment*, *defectiveness*, *entitlement*, *mistrust*, *enmeshment*, and *punitiveness*. For instance, individuals who feel subjugated may develop a fear of abandonment if they challenge others' expectations [15]. In contrast, schemas such as *emotional deprivation* and *social isolation* might be less prominent because they focus more on internal emotional states (eg, feelings of neglect or alienation) and less on the relational and power dynamics that are central to subjugation.

The Prominence of "Traumatizing Experiences" in PTSD Problems

The association between PTSD and the *mistrust* schema in OMHCs is consistent with research showing that individuals typically develop chronic mistrust following traumatic experiences involving betrayal, attachment disruptions, or abuse [101-106]. The very strong association suggests that mistrust is very likely to emerge from OMHC discussions on PTSD.

Similarly, the association between PTSD and the *punitiveness* schema highlights the prominent role of self-blame, which is frequently observed in individuals with PTSD [105,107-109]. The association with the *subjugation* schema indicates that individuals may express feelings of being controlled or oppressed, with the anonymity afforded by these online platforms likely encouraging the disclosure of such sensitive experiences [105,106,110,111]. These strong associations suggest that these schemas are more likely to appear in OMHC posts about PTSD.

The association between PTSD and the *emotional deprivation* schema aligns with findings from previous research on traumatized populations, such as veterans and survivors of interpersonal trauma [105,112]. The moderate association observed in this study suggests that emotional neglect is likely to be reflected in OMHC discussions on PTSD.

Consistent with previous studies that identified relationships between PTSD and schemas such as *enmeshment*, *self-sacrifice*, and *abandonment* [105,106], this study identified similar associations. However, the associations observed in OMHCs are comparatively weak, suggesting that these schemas are less central to PTSD experiences shared within these online platforms.

Overall, the findings suggest that individuals seeking PTSD support in OMHCs are more likely to describe acute traumatizing experiences, reflecting schemas such as *mistrust*, *punitiveness*, *subjugation*, and *emotional deprivation*. In contrast, schemas such as *enmeshment*, *self-sacrifice*, and *abandonment* play a lesser role in these discussions.

The Prominence of "Perceived Failure" in SUD Problems

The association between SUDs and the *failure to achieve* schema is consistent with previous research, suggesting that individuals who perceive themselves as underachievers are more susceptible

XSL•FO
RenderX

to substance misuse, and vice versa [20,113-117]. The moderate association observed in OMHCs indicates that themes of inadequacy and failure are likely to emerge in posts related to SUDs.

In contrast, the weak association between the *dependence* schema and SUDs in OMHCs challenges the traditional understanding of the disorder, which often emphasizes a high degree of helplessness and complete reliance on others for daily functioning [118-122]. The weak association suggests that individuals with SUD seeking support in OMHCs may still retain some autonomy and self-sufficiency despite their struggles.

Extracted Features Reflect Known Characteristics of EMSs

The EMS features (features extracted by GPT-4 for each schema are as listed in Multimedia Appendix 4; the features referenced in this section are enclosed in quotes.) identified in OMHCs reflect the core characteristics of each schema and align with previous research. Notable examples are presented subsequently.

Several features of the *vulnerability to harm or illness* schema suggest that it is triggered when an individual "anticipates or experiences a panic attack," "fears a negative health outcome," "encounters social interactions or public places," and "perceives everything is falling apart." These are consistent with the findings of Young et al [15], who noted that this schema is often activated by fears of uncontrollable external events. Similarly, McGinn et al [123] found that heightened anxiety about bodily symptoms and low perceived control significantly increase vulnerability to panic. The looming vulnerability model by Riskind [124] provides additional support for this, proposing that perceived vulnerability amplifies anxiety, manifesting in hypervigilance and beliefs such as "something bad is going to happen" and "one's health is in serious danger." In addition, Faustino et al [17] related this schema to COVID-19–related anxiety, emphasizing how individuals with heightened vulnerability often interpret health concerns as severe. Dean et al [125] observed that social exclusion intensifies vulnerability, perceived harm, and avoidant behaviors, aligning with the feature of "avoiding situations or places." In addition, features related to somatic symptoms, such as "physical discomfort or pain related to anxiety" and "heavy breathing or heart palpitations," were identified, consistent with the findings of McGinn et al [123] on somatic responses to low perceived control and vulnerability. Another notable feature of this schema is "seeking reassurance or help." Groves et al [126] found that patients awaiting medical diagnoses often sought reassurance from nurses, alleviating feelings of vulnerability and fostering greater optimism about medical outcomes.

Features of the *social isolation* schema indicate an increased sensitivity to "potential rejection," often leading individuals to "self-isolate" and "avoid social interactions or places." This behavior aligns with the findings of Watson and Nesdale [127] on how social rejection exacerbates isolation tendencies. Individuals with this schema often view themselves as "unworthy or unlovable," believe "others are better off without them," and believe they are "fundamentally different and unable to connect with others." These beliefs mirror the low self-esteem associated with social isolation and exclusion, as highlighted by Stanley and Arora [128] and Verkuyten and Thijs [129]. Emotional struggles, such as feelings of "hopelessness or helplessness," and at times, "fear or anxiety," are also prevalent. Johari et al [130] reported similar themes among homeless youth in Iran, who frequently experienced loneliness, abandonment, and fears of social harassment. Despite these challenges, some individuals with this schema expressed a "desire for connection," which is consistent with the observation by Cacioppo and Cacioppo [131] that loneliness, while heightening sensitivity to social threats, can also drive efforts to seek and rebuild social bonds.

The *emotional inhibition* schema is characterized by a persistent fear of having "personal issues or emotions disclosed," particularly in "social interaction" settings. This is consistent with findings by Young et al [15] and Coggins and Fox [68], who noted that individuals with this schema often suppress emotions out of concern that "others will judge or misunderstand them." As a result, they frequently experience feelings of being "overwhelmed by anxiety and worry," "scared or fearful," and at times, "helpless or hopeless." This schema is commonly triggered in situations where "negative outcomes are anticipated," as observed by Iwamitsu et al [132], who found that patients with breast tumor reported higher emotional distress when suppressing emotions after receiving their diagnosis. In addition, individuals with this schema often "avoid discussing personal issues" and "internalize feelings without seeking help." These tendencies mirror the inhibition and withdrawal behaviors observed by Denollet and Duijndam [133] in adults who had been socially inhibited, highlighting an aversion to open emotional expression and a reluctance to seek support from others.

Individuals with the *mistrust* schema often exhibit behaviors such as being "untrusting or suspicious" and perceiving others as "deceitful or have malicious intent," particularly when their "personal safety or well-being feels threatened." This aligns with the findings by Young et al [15], who observed that individuals with this schema commonly view others as potential threats, making them feel vulnerable if they lower their guard. Similarly, LaMotte et al [104] found that mistrust plays a role in the relationship between betrayal in romantic relationships and subsequent experiences of intimate partner violence. Individuals with this schema often believe they "will always be a targeted or wronged," aligning with the concept of victim sensitivity [134-136], which involves a heightened expectation of untrustworthy motives in others and intense reactions to perceived injustices. Pilkington et al [137] further highlighted the link between "mistrust" and accounts of violence, particularly in cases of domestic violence. Furthermore, those with this schema often "push others away" and prefer being "isolated or alone." Smith and Rosen [138] observed similar patterns among older clients receiving methadone treatment, who tended to self-isolate because of their mistrust. This self-isolation undermines the development of healthy relationships that are essential for recovery.

The *failure to achieve* schema reflects a persistent belief that one is "not good enough" and has "ruined everything," leading to cycles of "self-criticism." Young et al [15] similarly noted

XSL•FO
RenderX

that individuals with this schema often feel doomed to failure in both present and future endeavors. Wood et al [139] found that such individuals frequently engage in social comparison, seeking reassurance by comparing themselves to others they perceive as less successful. This aligns with the observation that individuals often "compare themselves to peers or societal standards." However, comparisons with those perceived as more successful can often intensify feelings of "jealousy," "disappointment in themselves," and "helplessness," sometimes causing them to "give up trying" to improve their circumstances. Leahy [140] observed that unfavorable comparisons in competitive environments can lead to envy, humiliation, helplessness, and withdrawal. This withdrawal reinforces feelings of failure, perpetuating the belief in one's inability to succeed and deepening cycles of self-criticism.

Limitations

This study focused exclusively on 3 widely accessed OMHCs that use English as the primary mode of communication. To broaden the cultural and linguistic diversity of the sample, future research could explore OMHCs that use other languages and compare the findings of this study with subsequent research to verify whether our results extend to OMHCs in different sociolinguistic and cultural contexts, where schema expression may vary. In addition, while the anonymity of these platforms encourages openness, it limits access to sociodemographic data (eg, age, gender, and socioeconomic status), which might be important in shaping mental health perspectives and schema characteristics. Future studies could explore methods that allow for the ethical anonymous collection of demographic data or use validated techniques to infer demographic characteristics while maintaining user privacy, which is of utmost importance. These suggestions are essential for assessing the generalizability of the EMS characteristics uncovered in this study. Expanding the demographic and cultural scope in future research would help validate and deepen these findings, contributing to a more comprehensive understanding of EMSs in online support-seeking environments.

Conclusions

In this study, we examined the characteristics of EMSs within OMHCs, situating our findings within the existing literature. By analyzing a large dataset of posts from widely accessed OMHCs, annotated with EMS labels using a state-of-the-art EMS classifier, we identified key EMS features and various associations between EMSs and mental health problems. Notably, we found that depression, PDs, and PTSD were associated with a wide range of EMSs, highlighting the complexity of these multifaceted problems. These findings suggest that interventions for these problems could benefit from addressing multiple EMSs simultaneously. In contrast, mental health problems with fewer EMS associations, such as anxiety, may respond more effectively to more targeted interventions focused on specific schemas. In addition, we identified EMSs as more prominently associated with certain mental health problems in OMHCs. For example, anxiety-related posts typically center around vulnerability to harm, while depression-related posts reflect prolonged unmet interpersonal needs. Negative self-perception and self-neglect are prevalent in posts about EDs, while posts about PDs involve themes of subjugation. Posts about PTSD often describe traumatic experiences, and those concerning SUD are marked by themes of personal failure. A key contribution of this study is the novel group-level case conceptualization technique, which uses recent advanced LLMs and prompt engineering techniques to extract EMS features from OMHC posts. This approach provides valuable insights into how EMSs are expressed in online support-seeking environments and has the potential to inform personalized, ST-focused interventions. Overall, this study provides a foundational understanding of EMS characteristics in online support-seeking contexts and their role in shaping mental health experiences. Our findings highlight the potential for tailored, ST-based interventions in online peer support settings, which could enhance therapeutic outcomes and promote the mental well-being of individuals seeking support in these online communities.

Acknowledgments

This research and project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG-GC-2019-001-2B).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Statistical approach for uncovering associations.
[DOC File , 43 KB - jmir_v27i1e59524_app1.doc]

Multimedia Appendix 2
Consistency of the extracted features.
[DOC File , 98 KB - jmir_v27i1e59524_app2.doc]

XSL•FO
RenderX

Multimedia Appendix 3
Associations between early maladaptive schemas and mental health problems.
[DOC File , 64 KB - jmir_v27i1e59524_app3.doc]

Multimedia Appendix 4
Features of early maladaptive schemas.
[DOC File , 66 KB - jmir_v27i1e59524_app4.doc]

References

1. Mauz E, Walther L, Junker S, Kersjes C, Damerow S, Eicher S, et al. Time trends in mental health indicators in Germany's adult population before and during the COVID-19 pandemic. Front Public Health 2023 Feb 23;11:1065938 [FREE Full text] [doi: 10.3389/fpubh.2023.1065938] [Medline: 36908429]
2. Deng J, Zhou F, Hou W, Heybati K, Lohit S, Abbas U, et al. Prevalence of mental health symptoms in children and adolescents during the COVID-19 pandemic: a meta-analysis. Ann N Y Acad Sci 2023 Feb;1520(1):53-73 [FREE Full text] [doi: 10.1111/nyas.14947] [Medline: 36537131]
3. COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. World Health Organization. 2022 Mar 2. URL: https://tinyurl.com/bdcykcjv [accessed 2024-01-03]
4. Kantor V, Knefel M, Lueger-Schuster B. Perceived barriers and facilitators of mental health service utilization in adult trauma survivors: a systematic review. Clin Psychol Rev 2017 Mar;52:52-68 [FREE Full text] [doi: 10.1016/j.cpr.2016.12.001] [Medline: 28013081]
5. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan 03;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
6. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2010 Dec 07;41(8):1751-1761. [doi: 10.1017/s0033291710002291]
7. Houston TK, Cooper LA, Ford DE. Internet support groups for depression: a 1-year prospective cohort study. Am J Psychiatry 2002 Dec 01;159(12):2062-2068. [doi: 10.1176/appi.ajp.159.12.2062] [Medline: 12450957]
8. Melling B, Houguet-Pincham T. Online peer support for individuals with depression: a summary of current research and future considerations. Psychiatr Rehabil J 2011;34(3):252-254. [doi: 10.2975/34.3.2011.252.254] [Medline: 21208865]
9. Powell J, Clarke A. Investigating internet use by mental health service users: interview study. Stud Health Technol Inform 2007;129(Pt 2):1112-1116. [Medline: 17911888]
10. Liu J, Kong J. Why do users of online mental health communities get likes and reposts: a combination of text mining and empirical analysis. Healthcare (Basel) 2021 Aug 31;9(9):1133 [FREE Full text] [doi: 10.3390/healthcare9091133] [Medline: 34574907]
11. Strand M, Eng LS, Gammon D. Combining online and offline peer support groups in community mental health care settings: a qualitative study of service users' experiences. Int J Ment Health Syst 2020 May 29;14(1):39 [FREE Full text] [doi: 10.1186/s13033-020-00370-x] [Medline: 32514303]
12. Suler J. The online disinhibition effect. Appl Psychoanal Stud 2006 Jan 06;2(2):184-188. [doi: 10.1002/aps.42]
13. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
14. Yazdavar AH, Mahdavinejad MS, Bajaj G, Thirunarayan K, Pathak J, Sheth A. Mental health analysis via social media data. In: Proceedings of the IEEE International Conference on Healthcare Informatics. 2018 Presented at: ICHI 2018; June 4-7, 2018; New York, NY. [doi: 10.1109/ichi.2018.00102]
15. Young JE, Klosko JS, Weishaar ME. Schema Therapy: A Practitioner's Guide. New York, NY: Guilford Publications; 2003.
16. Gollapalli S, Ang B, Ng SK. Identifying early maladaptive schemas from mental health question texts. In: Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023. 2023 Presented at: EMNLP 2023; December 6-10, 2023; Singapore, Singapore. [doi: 10.18653/v1/2023.findings-emnlp.792]
17. Faustino B, Vasco AB, Delgado J, Farinha-Fernandes A, Guerreiro JC. Early maladaptive schemas and COVID-19 anxiety: the mediational role of mistrustfulness and vulnerability to harm and illness. Clin Psychol Psychother 2022 Jul 17;29(4):1297-1308 [FREE Full text] [doi: 10.1002/cpp.2706] [Medline: 34989057]
18. Cámara M, Calvete E. Early maladaptive schemas as moderators of the impact of stressful events on anxiety and depression in university students. J Psychopathol Behav Assess 2011 Oct 9;34(1):58-68. [doi: 10.1007/s10862-011-9261-6]
19. Kunst H, Lobbestael J, Candel I, Batink T. Early maladaptive schemas and their relation to personality disorders: a correlational examination in a clinical population. Clin Psychol Psychother 2020 Nov 17;27(6):837-846 [FREE Full text] [doi: 10.1002/cpp.2467] [Medline: 32358901]
20. Janson DL. Early maladaptive schemas in an Australian adult alcohol dependent clinical sample: differences between men and women. Edith Cowan University. 2015. URL: https://ro.ecu.edu.au/theses_hons/1463/ [accessed 2025-01-27]

XSL•FO
RenderX

21. Barazandeh H, Kissane DW, Saeedi N, Gordon M. A systematic review of the relationship between early maladaptive schemas and borderline personality disorder/traits. Pers Individ Differ 2016 May;94:130-139. [doi: 10.1016/j.paid.2016.01.021]
22. Bishop A, Younan R, Low J, Pilkington PD. Early maladaptive schemas and depression in adulthood: a systematic review and meta-analysis. Clin Psychol Psychother 2022 Jan 16;29(1):111-130. [doi: 10.1002/cpp.2630] [Medline: 34131990]
23. Halvorsen M, Wang CE, Richter J, Myrland I, Pedersen SK, Eisemann M, et al. Early maladaptive schemas, temperament and character traits in clinically depressed and previously depressed subjects. Clin Psychol Psychother 2009;16(5):394-407. [doi: 10.1002/cpp.618] [Medline: 19479673]
24. Wright MO, Crawford E, Del Castillo D. Childhood emotional maltreatment and later psychological distress among college students: the mediating role of maladaptive schemas. Child Abuse Negl 2009 Jan;33(1):59-68. [doi: 10.1016/j.chiabu.2008.12.007] [Medline: 19167067]
25. Balsamo M, Carlucci L, Sergi MR, Klein Murdock K, Saggino A. The mediating role of early maladaptive schemas in the relation between co-rumination and depression in young adults. PLoS One 2015 Oct 21;10(10):e0140177 [FREE Full text] [doi: 10.1371/journal.pone.0140177] [Medline: 26488748]
26. Messman-Moore TL, Coates AA. The impact of childhood psychological abuse on adult interpersonal conflict. J Emot Abuse 2007;7(2):75-92. [doi: 10.1300/j135v07n02_05]
27. Young J, Brown G. YSQ-L3. Schema Therapy Institute. 2017. URL: https://psychology-training.com.au/wp-content/uploads/2017/04/Young-Schema-Questionnaire-L3.pdf [accessed 2024-01-04]
28. Dean J, Potts HW, Barker C. Direction to an internet support group compared with online expressive writing for people with depression and anxiety: a randomized trial. JMIR Ment Health 2016 May 17;3(2):e12 [FREE Full text] [doi: 10.2196/mental.5133] [Medline: 27189142]
29. Gliddon E, Lauder S, Berk L, Cosgrove V, Grimm D, Dodd S, et al. Evaluating discussion board engagement in the MoodSwings online self-help program for bipolar disorder: protocol for an observational prospective cohort study. BMC Psychiatry 2015 Oct 14;15:243 [FREE Full text] [doi: 10.1186/s12888-015-0630-7] [Medline: 26462799]
30. Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, et al. Scaling instruction-finetuned language models. arXiv Preprint posted online on October 20, 2022. [doi: 10.48550/arXiv.2210.11416]
31. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. arXiv Preprint posted online on March 4, 2022. [doi: 10.48550/arXiv.2203.02155]
32. OpenAI. GPT-4 technical report. arXiv Preprint posted online on March 15, 2023.
33. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv Preprint posted online on June 12, 2017. [doi: 10.48550/arXiv.1706.03762]
34. Kjell ON, Sikström S, Kjell K, Schwartz HA. Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. Sci Rep 2022 Mar 10;12(1):3918 [FREE Full text] [doi: 10.1038/s41598-022-07520-w] [Medline: 35273198]
35. Teferra BG, Rose J. Predicting generalized anxiety disorder from impromptu speech transcripts using context-aware transformer-based neural networks: model evaluation study. JMIR Ment Health 2023 Mar 28;10:e44325 [FREE Full text] [doi: 10.2196/44325] [Medline: 36976636]
36. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. Transl Psychiatry 2023 Oct 06;13(1):309 [FREE Full text] [doi: 10.1038/s41398-023-02592-2] [Medline: 37798296]
37. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using informed deep learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
38. Xu X, Yao B, Dong Y, Gabriel S, Yu H, Hendler J, et al. Mental-LLM: leveraging large language models for mental health prediction via online text data. Proc ACM Interact Mob Wearable Ubiquitous Technol 2024 Mar 06;8(1):1-32. [doi: 10.1145/3643540]
39. Community support forum. 7Cups. URL: https://www.7cups.com/community/ [accessed 2023-11-21]
40. Beyond Blue homepage. Beyond Blue. URL: https://forums.beyondblue.org.au/ [accessed 2023-11-21]
41. Patient homepage. Patient. URL: https://patient.info/forums [accessed 2023-11-21]
42. National alliance on mental illness homepage. National Alliance on Mental Illness. URL: https://tinyurl.com/ynb9b3u7 [accessed 2024-01-12]
43. Castaldelli-Maia JM, Bhugra D. Analysis of global prevalence of mental and substance use disorders within countries: focus on sociodemographic characteristics and income levels. Int Rev Psychiatry 2022 Feb;34(1):6-15. [doi: 10.1080/09540261.2022.2040450] [Medline: 35584016]
44. Poliak A. A survey on recognizing textual entailment as an NLP evaluation. In: Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems. 2020 Presented at: Eval4NLP 2020; November 8-12, 2020; Punta Cana, Dominican Republic. [doi: 10.18653/v1/2020.eval4nlp-1.10]
45. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res 2020 Jan 01;21(1):5485-5551 [FREE Full text]

XSL•FO
RenderX

46. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(2):143-149 [FREE Full text] [doi: 10.11613/bm.2013.018] [Medline: 23894860]
47. Lantz B. The large sample size fallacy. Scand J Caring Sci 2013 Jun;27(2):487-492. [doi: 10.1111/j.1471-6712.2012.01052.x] [Medline: 22862286]
48. Szumilas M. Explaining odds ratios. J Can Acad Child Adolesc Psychiatry 2010 Aug;19(3):227-229 [FREE Full text] [Medline: 20842279]
49. Roberts LD. Ethical issues in conducting qualitative research in online communities. Qual Res Psychol 2015 Apr 20;12(3):314-325. [doi: 10.1080/14780887.2015.1008909]
50. Smedley RM, Coulson NS. A practical guide to analysing online support forums. Qual Res Psychol 2018 May 30;18(1):76-103. [doi: 10.1080/14780887.2018.1475532]
51. Mental disorders. World Health Organization. 2022 Jun 08. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [accessed 2024-01-12]
52. Hwang TJ, Rabheru K, Peisah C, Reichman W, Ikeda M. Loneliness and social isolation during the COVID-19 pandemic. Int Psychogeriatr 2020 Oct;32(10):1217-1220. [doi: 10.1017/s1041610220000988]
53. Boyraz G, Legros DN, Tigershtrom A. COVID-19 and traumatic stress: the role of perceived vulnerability, COVID-19-related worries, and social isolation. J Anxiety Disord 2020 Dec;76:102307 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102307] [Medline: 32937259]
54. Li JB, T Vazsonyi A, Dou K. Is individualism-collectivism associated with self-control? Evidence from Chinese and U.S. samples. PLoS One 2018 Dec 19;13(12):e0208541 [FREE Full text] [doi: 10.1371/journal.pone.0208541] [Medline: 30566491]
55. Agopian AJ, Eastcott LM, Mitchell LE. Age of onset and effect size in genome-wide association studies. Birth Defects Res A Clin Mol Teratol 2012 Nov;94(11):908-911 [FREE Full text] [doi: 10.1002/bdra.23066] [Medline: 22933422]
56. Rosenthal JA. Qualitative descriptors of strength of association and effect size. J Soc Serv Res 1996;21(4):37-59. [doi: 10.1300/j079v21n04_02]
57. Knaster P, Estlander AM, Karlsson H, Kaprio J, Kalso E. Diagnosing depression in chronic pain patients: DSM-IV major depressive disorder vs. Beck Depression Inventory (BDI). PLoS One 2016 Mar 23;11(3):e0151982 [FREE Full text] [doi: 10.1371/journal.pone.0151982] [Medline: 27008161]
58. Wiechman SA, Ptacek JT, Patterson DR, Gibran NS, Engrav LE, Heimbach DM. Rates, trends, and severity of depression after burn injuries. J Burn Care Rehabil 2001;22(6):417-424. [doi: 10.1097/00004630-200111000-00012] [Medline: 11761394]
59. O'Neil MF, Moore NA. Animal models of depression: are there any? Hum Psychopharmacol 2003 Jun;18(4):239-254. [doi: 10.1002/hup.496] [Medline: 12766928]
60. Karmen C, Hsiung RC, Wetter T. Screening Internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. Comput Methods Programs Biomed 2015 Jun;120(1):27-36. [doi: 10.1016/j.cmpb.2015.03.008] [Medline: 25891366]
61. Cui L, Lin W, Oei TP. Factor structure and psychometric properties of the Young Schema Questionnaire (short form) in Chinese undergraduate students. Int J Ment Health Addiction 2010 Sep 15;9(6):645-655. [doi: 10.1007/s11469-010-9283-4]
62. Bär A, Bär HE, Rijkeboer MM, Lobbestael J. Early maladaptive schemas and schema modes in clinical disorders: a systematic review. Psychol Psychother 2023 Sep;96(3):716-747. [doi: 10.1111/papt.12465] [Medline: 37026578]
63. Powers JP, Moshontz H, Hoyle RH. Self-control and affect regulation styles predict anxiety longitudinally in university students. Collabra Psychol 2020 Feb 10;6(1):11. [doi: 10.1525/collabra.280]
64. Zhang C, Meng D, Zhu L, Ma X, Guo J, Fu Y, et al. The effect of trait anxiety on bedtime procrastination: the mediating role of self-control. Int J Behav Med 2023 Apr;30(2):260-267. [doi: 10.1007/s12529-022-10089-3] [Medline: 35459983]
65. Bertrams A, Baumeister RF, Englert C. Higher self-control capacity predicts lower anxiety-impaired cognition during math examinations. Front Psychol 2016 Mar 31;7:485 [FREE Full text] [doi: 10.3389/fpsyg.2016.00485] [Medline: 27065013]
66. He M, Zhan X, Liu C, Li L, Zhao X, Ren L, et al. The relationship between self-control and mental health problems among Chinese university students. Front Public Health 2023 Oct 31;11:1224427 [FREE Full text] [doi: 10.3389/fpubh.2023.1224427] [Medline: 38026364]
67. Krause ED, Mendelson T, Lynch TR. Childhood emotional invalidation and adult psychological distress: the mediating role of emotional inhibition. Child Abuse Negl 2003 Feb;27(2):199-213. [doi: 10.1016/s0145-2134(02)00536-7] [Medline: 12615094]
68. Coggins J, Fox JR. A qualitative exploration of emotional inhibition: a basic emotions and developmental perspective. Clin Psychol Psychother 2009;16(1):55-76. [doi: 10.1002/cpp.604] [Medline: 19161207]
69. Roemer L, Salters K, Raffa SD, Orsillo SM. Fear and avoidance of internal experiences in GAD: preliminary tests of a conceptual model. Cogn Ther Res 2005 Feb;29:71-88. [doi: 10.1007/s10608-005-1650-2]
70. Tao S, Jing Y. More sense of self-discipline, less procrastination: the mediation of autonomous motivation. Front Psychol 2023 Nov 23;14:1268614 [FREE Full text] [doi: 10.3389/fpsyg.2023.1268614] [Medline: 38078243]
71. Ahmadpanah M, Astinsadaf S, Akhondi A, Haghighi M, Sadeghi Bahmani D, Nazaribadie M, et al. Early maladaptive schemas of emotional deprivation, social isolation, shame and abandonment are related to a history of suicide attempts

XSL•FO
RenderX

among patients with major depressive disorders. Compr Psychiatry 2017 Aug;77:71-79. [doi: 10.1016/j.comppsych.2017.05.008] [Medline: 28636896]
72. Ge L, Yap CW, Ong R, Heng BH. Social isolation, loneliness and their relationships with depressive symptoms: a population-based study. PLoS One 2017 Aug 23;12(8):e0182145 [FREE Full text] [doi: 10.1371/journal.pone.0182145] [Medline: 28832594]
73. Canlı D, Karaşar B. Predictors of major depressive disorder: the need for social approval and self-esteem. Alpha Psychiatry 2021 Jan 19;22(1):38-42 [FREE Full text] [doi: 10.5455/apd.97683] [Medline: 36426204]
74. Karagöl A, Törenli Kaya Z. Evaluating early maladaptive schemas and depression levels in living kidney and liver donors. Psychol Health Med 2022 Dec;27(10):2161-2170. [doi: 10.1080/13548506.2021.1981409] [Medline: 34545761]
75. Davoodi E, Wen A, Dobson KS, Noorbala AA, Mohammadi A, Farahmand Z. Early maladaptive schemas in depression and somatization disorder. J Affect Disord 2018 Aug 01;235:82-89. [doi: 10.1016/j.jad.2018.04.017] [Medline: 29655079]
76. Matos M, Pinto-Gouveia J, Duarte C. Internalizing early memories of shame and lack of safeness and warmth: the mediating role of shame on depression. Behav Cogn Psychother 2013 Mar 25;41(4):479-493. [doi: 10.1017/s1352465812001099]
77. darvishi F, rahmani MA, Akbari B, Rahbar M. A comparison of relationship between early maladaptive schemas with depression severity in suicidal group and non-clinical sample. Procedia Soc Behav Sci 2013 Jul 09;84:1072-1077. [doi: 10.1016/j.sbspro.2013.06.702]
78. Blatt SJ. Experiences of Depression: Theoretical, Clinical, and Research Perspectives. Washington, DC: American Psychological Association; 2004.
79. Cohen RE, Tennen H. Self-punishment in learned helplessness and depression. J Soc Clin Psychol 1985 Mar;3(1):82-96. [doi: 10.1521/jscp.1985.3.1.82]
80. Segrin C. Interpersonal communication problems associated with depression and loneliness. In: Handbook of Communication and Emotion: Research, Theory, Applications, and Contexts. Cambridge, MA: Academic Press; 1996.
81. Hames JL, Hagan CR, Joiner TE. Interpersonal processes in depression. Annu Rev Clin Psychol 2013;9:355-377. [doi: 10.1146/annurev-clinpsy-050212-185553] [Medline: 23297787]
82. Marcus DK, Nardone ME. Depression and interpersonal rejection. Clin Psychol Rev 1992;12(4):433-449. [doi: 10.1016/0272-7358(92)90126-S]
83. Forbush K, Watson D. Emotional inhibition and personality traits: a comparison of women with anorexia, bulimia, and normal controls. Ann Clin Psychiatry 2006;18(2):115-121. [doi: 10.1080/10401230600614637] [Medline: 16754417]
84. Bekker MH, Spoor ST. Emotional inhibition, health, gender, and eating disorders: the role of (over) sensitivity to others. In: Vingerhoets AJ, Nyklíček I, Denollet J, editors. Emotion Regulation. Boston, MA: Springer; 2008.
85. Espeset EM, Gulliksen KS, Nordbø RH, Skårderud F, Holte A. The link between negative emotions and eating disorder behaviour in patients with anorexia nervosa. Eur Eat Disord Rev 2012 Nov;20(6):451-460. [doi: 10.1002/erv.2183] [Medline: 22696277]
86. Aloi M, Rania M, Caroleo M, Carbone EA, Fazia G, Calabrò G, et al. How are early maladaptive schemas and DSM-5 personality traits associated with the severity of binge eating? J Clin Psychol 2020 Mar;76(3):539-548. [doi: 10.1002/jclp.22900] [Medline: 31733127]
87. Sanftner JL, Barlow DH, Marschall DE, Tangney JP. The relation of shame and guilt to eating disorder symptomatology. J Soc Clin Psychol 1995 Dec;14(4):315-324. [doi: 10.1521/jscp.1995.14.4.315]
88. Kelly AC, Carter JC. Why self-critical patients present with more severe eating disorder pathology: the mediating role of shame. Br J Clin Psychol 2013 Jun;52(2):148-161. [doi: 10.1111/bjc.12006] [Medline: 24215145]
89. Oakley B, Knafo A, Madhavan G, Wilson DS. Pathological Altruism. Oxford, UK: Oxford University Press; Dec 09, 2011.
90. Bachner-Melman R, Zohar AH, Ebstein RP, Bachar E. The relationship between selflessness levels and the severity of anorexia nervosa symptomatology. Eur Eat Disord Rev 2007 May;15(3):213-220. [doi: 10.1002/erv.740] [Medline: 17676691]
91. Buchholz A, Henderson KA, Hounsell A, Wagner A, Norris M, Spettigue W. Self-silencing in a clinical sample of female adolescents with eating disorders. J Can Acad Child Adolesc Psychiatry 2007 Nov;16(4):158-163 [FREE Full text] [Medline: 18392167]
92. Bachner-Melman R, Oakley B. Giving 'til it hurts': eating disorders and pathological altruism. In: Latzer Y, Stein D, editors. Bio-Psycho-Social Contributions to Understanding Eating Disorders. Cham, Switzerland: Springer; 2016.
93. Bachner-Melman R. The relevance of pathological altruism to eating disorders. In: Oakley B, Knafo A, Madhavan G, Wilson DS, editors. Pathological Altruism. Oxford, UK: Oxford University Press; 2012.
94. van der Linde RP, Huntjens RJ, Bachrach N, Rijkeboer MM. Personality disorder traits, maladaptive schemas, modes and coping styles in participants with complex dissociative disorders, borderline personality disorder and avoidant personality disorder. Clin Psychol Psychother 2023 Aug 10;30(6):1234-1245. [doi: 10.1002/cpp.2892] [Medline: 37563773]
95. Carr SN, Francis AJ. Do early maladaptive schemas mediate the relationship between childhood experiences and avoidant personality disorder features? A preliminary investigation in a non-clinical sample. Cogn Ther Res 2009 May 2;34(4):343-358. [doi: 10.1007/s10608-009-9250-1]
96. Bach B, Farrell JM. Schemas and modes in borderline personality disorder: the mistrustful, shameful, angry, impulsive, and unhappy child. Psychiatry Res 2018 Jan;259:323-329. [doi: 10.1016/j.psychres.2017.10.039] [Medline: 29101874]

XSL•FO
RenderX

97. Palihawadana V, Broadbear JH, Rao S. Reviewing the clinical significance of 'fear of abandonment' in borderline personality disorder. Australas Psychiatry 2019 Feb 07;27(1):60-63. [doi: 10.1177/1039856218810154] [Medline: 30403145]
98. Vanwoerden S, Kalpakci A, Sharp C. The relations between inadequate parent-child boundaries and borderline personality disorder in adolescence. Psychiatry Res 2017 Nov;257:462-471. [doi: 10.1016/j.psychres.2017.08.015] [Medline: 28837938]
99. Yakeley J. Current understanding of narcissism and narcissistic personality disorder. BJPsych Adv 2018 Jul 05;24(5):305-315. [doi: 10.1192/bja.2018.20]
100. Esmaeilian N, Dehghani M, Koster EH, Hoorelbeke K. Early maladaptive schemas and borderline personality disorder features in a nonclinical sample: a network analysis. Clin Psychol Psychother 2019 May 05;26(3):388-398. [doi: 10.1002/cpp.2360] [Medline: 30771229]
101. Beaton J, Thielking M. Chronic mistrust and complex trauma: Australian psychologists' perspectives on the treatment of young women with a history of childhood maltreatment. Aus Psychol 2020 Nov 12;55(3):230-243. [doi: 10.1111/ap.12430]
102. Lian AE, Chooi WT, Bono SA. A systematic review investigating the early maladaptive schemas (EMS) in individuals with trauma experiences and PTSD. Eur J Trauma Dissoc 2023 Mar;7(1):100315. [doi: 10.1016/j.ejtd.2023.100315]
103. Ahmadian A, Mirzaee J, Omidbeygi M, Holsboer-Trachsler E, Brand S. Differences in maladaptive schemas between patients suffering from chronic and acute posttraumatic stress disorder and healthy controls. Neuropsychiatr Dis Treat 2015 Jul;11:1677-1684 [FREE Full text] [doi: 10.2147/NDT.S85959] [Medline: 26203249]
104. LaMotte AD, Taft CT, Weatherill RP. Mistrust of others as a mediator of the relationship between trauma exposure and use of partner aggression. Psychol Trauma 2016 Jul;8(4):535-540 [FREE Full text] [doi: 10.1037/tra0000157] [Medline: 27348070]
105. Cockram DM, Drummond PD, Lee CW. Role and treatment of early maladaptive schemas in Vietnam Veterans with PTSD. Clin Psychol Psychother 2010 Mar 29;17(3):165-182 [FREE Full text] [doi: 10.1002/cpp.690] [Medline: 20486158]
106. Greenblatt-Kimron L, Karatzias T, Yonatan M, Shoham A, Hyland P, Ben-Ezra M, et al. Early maladaptive schemas and ICD-11 CPTSD symptoms: treatment considerations. Psychol Psychother 2023 Mar 17;96(1):117-128. [doi: 10.1111/papt.12429] [Medline: 36253922]
107. Ullman SE, Townsend SM, Filipas HH, Starzynski LL. Structural models of the relations of assault severity, social support, avoidance coping, self-blame, and PTSD among sexual assault survivors. Psychol Women Q 2007 Mar 01;31(1):23-37. [doi: 10.1111/j.1471-6402.2007.00328.x]
108. Kline NK, Berke DS, Rhodes CA, Steenkamp MM, Litz BT. Self-blame and PTSD following sexual assault: a longitudinal analysis. J Interpers Violence 2021 Mar 21;36(5-6):NP3153-NP3168. [doi: 10.1177/0886260518770652] [Medline: 29683081]
109. Raz A, Rubinstein R, Shadach E, Chaikin G, Ben Yehuda A, Tatsa-Laur L, et al. Behavioral self-blame in PTSD-etiology, risk factors, and proposed interventions. Int J Environ Res Public Health 2023 Aug 05;20(15):6530 [FREE Full text] [doi: 10.3390/ijerph20156530] [Medline: 37569070]
110. Harding HG, Burns EE, Jackson JL. Identification of child sexual abuse survivor subgroups based on early maladaptive schemas: implications for understanding differences in posttraumatic stress disorder symptom severity. Cogn Ther Res 2011 Aug 10;36(5):560-575. [doi: 10.1007/s10608-011-9385-8]
111. Holmes SC, Facemire VC, DaFonseca AM. Expanding criterion a for posttraumatic stress disorder: considering the deleterious impact of oppression. Traumatology 2016;22(4):314-321. [doi: 10.1037/trm0000104]
112. Karatzias T, Jowett S, Begley A, Deas S. Early maladaptive schemas in adult survivors of interpersonal trauma: foundations for a cognitive theory of psychopathology. Eur J Psychotraumatol 2016 Sep 16;7(1):30713 [FREE Full text] [doi: 10.3402/ejpt.v7.30713] [Medline: 27642181]
113. Fothergill KE, Ensminger ME, Green KM, Crum RM, Robertson J, Juon HS. The impact of early school behavior and educational achievement on adult drug use disorders: a prospective study. Drug Alcohol Depend 2008 Jan 01;92(1-3):191-199 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.08.001] [Medline: 17869029]
114. Henry KL. Academic achievement and adolescent drug use: an examination of reciprocal effects and correlated growth trajectories. J Sch Health 2010 Jan;80(1):38-43. [doi: 10.1111/j.1746-1561.2009.00455.x] [Medline: 20051089]
115. Goslin JC. Underachieves: a curriculum design. Prev Sch Fail 1989 Oct;34(1):22-28. [doi: 10.1080/1045988x.1989.9944548]
116. Adeoye OE, Fasanmi SS, Igo J. Self-concept clarity, learned helplessness and fatigue severity as predictors of psychoactive drug use and abuse among undergraduates. Int J Sci Basic Appl Res 2020;53(2):198-210 [FREE Full text]
117. Bakhshi Bojed F, Nikmanesh Z. Role of early maladaptive schemas on addiction potential in youth. Int J High Risk Behav Addict 2013 Sep 25;2(2):72-76 [FREE Full text] [doi: 10.5812/ijhrba.10148] [Medline: 24971278]
118. Sterling RC, Gottheil E, Weinstein SP, Lundy A, Serota RD. Learned helplessness and cocaine dependence: an investigation. J Addict Dis 1996 Mar 14;15(2):13-24. [doi: 10.1300/j069v15n02_02]
119. Shaghaghy F, Saffarinia M, Iranpoor M, Soltanynejad A. The relationship of early maladaptive schemas, attributional styles and learned helplessness among addicted and non-addicted men. Addict Health 2011;3(1-2):45-52 [FREE Full text] [Medline: 24494116]
120. Koob GF, Powell P, White A. Addiction as a coping response: hyperkatifeia, deaths of despair, and COVID-19. Am J Psychiatry 2020 Nov 01;177(11):1031-1037. [doi: 10.1176/appi.ajp.2020.20091375] [Medline: 33135468]

XSL•FO
RenderX

121. Romano I, Patte KA, de Groh M, Jiang Y, Wade TJ, Bélanger RE, et al. Substance-related coping behaviours among youth during the early months of the COVID-19 pandemic. Addict Behav Rep 2021 Dec;14:100392 [FREE Full text] [doi: 10.1016/j.abrep.2021.100392] [Medline: 34754898]
122. Wardell JD, Kempe T, Rapinda KK, Single A, Bilevicius E, Frohlich JR, et al. Drinking to cope during COVID-19 pandemic: the role of external and internal factors in coping motive pathways to alcohol use, solitary drinking, and alcohol problems. Alcohol Clin Exp Res 2020 Oct;44(10):2073-2083. [doi: 10.1111/acer.14425] [Medline: 32870516]
123. McGinn LK, Nooner KB, Cohen J, Leaberry KD. The role of early experience and cognitive vulnerability: presenting a unified model of the etiology of panic. Cogn Ther Res 2015 Apr 2;39(4):508-519. [doi: 10.1007/s10608-015-9673-9]
124. Riskind JH. Looming vulnerability to threat: a cognitive paradigm for anxiety. Behav Res Ther 1997 Aug;35(8):685-702. [doi: 10.1016/s0005-7967(97)00011-9] [Medline: 9256513]
125. Dean KK, Wentworth G, LeCompte N. Social exclusion and perceived vulnerability to physical harm. Self Identity 2019 Jan 08;18(1):87-102. [doi: 10.1080/15298868.2017.1370389]
126. Groves PS, Bunch JL, Kuehnle F. Increasing a patient's sense of security in the hospital: a theory of trust and nursing action. Nursing Inquiry 2023 Oct;30(4):e12569. [doi: 10.1111/nin.12569]
127. Watson J, Nesdale D. Rejection sensitivity, social withdrawal, and loneliness in young adults. J Appl Soc Psychol 2012 Jun 14;42(8):1984-2005. [doi: 10.1111/j.1559-1816.2012.00927.x]
128. Stanley L, Arora T. Social exclusion amongst adolescent girls their self‐esteem and coping strategies. Educ Psychol Pract 2007 Oct 19;14(2):94-100. [doi: 10.1080/0266736980140204]
129. Verkuyten M, Thijs J. Ethnic discrimination and global self-worth in early adolescents: the mediating role of ethnic self-esteem. Int J Behav Dev 2006 Mar 01;30(2):107-116. [doi: 10.1177/0165025406063573]
130. Johari F, Iranpour A, Dehghan M, Alizadeh S, Safizadeh M, Sharifi H. Lonely, harassed and abandoned in society: the lived experiences of Iranian homeless youth. BMC Psychol 2022 Mar 20;10(1):75 [FREE Full text] [doi: 10.1186/s40359-022-00787-8] [Medline: 35307020]
131. Cacioppo JT, Cacioppo S. Social relationships and health: the toxic effects of perceived social isolation. Soc Personal Psychol Compass 2014 Feb 01;8(2):58-72 [FREE Full text] [doi: 10.1111/spc3.12087] [Medline: 24839458]
132. Iwamitsu Y, Shimoda K, Abe H, Tani T, Kodama M, Okawa M. Differences in emotional distress between breast tumor patients with emotional inhibition and those with emotional expression. Psychiatry Clin Neurosci 2003 Jun 02;57(3):289-294 [FREE Full text] [doi: 10.1046/j.1440-1819.2003.01119.x] [Medline: 12753569]
133. Denollet J, Duijndam S. The multidimensional nature of adult social inhibition: inhibition, sensitivity and withdrawal facets of the SIQ15. J Affect Disord 2019 Feb 15;245:569-579. [doi: 10.1016/j.jad.2018.11.035] [Medline: 30445382]
134. Gollwitzer M, Rothmund T, Süssenbach P. The sensitivity to mean intentions (SeMI) model: basic assumptions, recent findings, and potential avenues for future research. Soc Pers Psychol Compass 2013 Jul 08;7(7):415-426. [doi: 10.1111/spc3.12041]
135. Gollwitzer M, Rothmund T, Pfeiffer A, Ensenbach C. Why and when justice sensitivity leads to pro- and antisocial behavior. J Res Pers 2009 Dec;43(6):999-1005. [doi: 10.1016/j.jrp.2009.07.003]
136. Gollwitzer M, Süssenbach P, Hannuschke M. Victimization experiences and the stabilization of victim sensitivity. Front Psychol 2015 Apr 14;6:439 [FREE Full text] [doi: 10.3389/fpsyg.2015.00439] [Medline: 25926806]
137. Pilkington PD, Noonan C, May T, Younan R, Holt RA. Early maladaptive schemas and intimate partner violence victimization and perpetration: a systematic review and meta-analysis. Clin Psychol Psychother 2021 Sep 05;28(5):1030-1042. [doi: 10.1002/cpp.2558] [Medline: 33527558]
138. Smith ML, Rosen D. Mistrust and self-isolation: barriers to social support for older adult methadone clients. J Gerontol Soc Work 2009 Oct 24;52(7):653-667. [doi: 10.1080/01634370802609049] [Medline: 19787525]
139. Wood JV, Giordano-Beech M, Ducharme MJ. Compensating for failure through social comparison. Pers Soc Psychol Bull 1999 Nov 01;25(11):1370-1386. [doi: 10.1177/0146167299259004]
140. Leahy RL. Cognitive-behavioral therapy for envy. Cogn Ther Res 2020 Jul 29;45(3):418-427. [doi: 10.1007/s10608-020-10135-y]

Abbreviations

AI: artificial intelligence
BPD: borderline personality disorder
ED: eating disorder
EMS: early maladaptive schema
EPM: entailment-based prediction model
LLM: large language model
OMHC: online mental health community
OR: odds ratio
PD: personality disorder
PTSD: posttraumatic stress disorder

XSL•FO
RenderX

ST: schema therapy
SUD: substance use disorder
YSQ: Young Schema Questionnaire

Edited by A Mavragani; submitted 19.04.24; peer-reviewed by A AL-Asadi, F Franza; comments to author 29.08.24; revised version received 08.12.24; accepted 06.01.25; published 07.02.25.

Please cite as:
Ang BH, Gollapalli SD, Du M, Ng SK
Unraveling Online Mental Health Through the Lens of Early Maladaptive Schemas: AI-Enabled Content Analysis of Online Mental Health Communities
J Med Internet Res 2025;27:e59524
URL: https://www.jmir.org/2025/1/e59524
doi:10.2196/59524
PMID:

©Beng Heng Ang, Sujatha Das Gollapalli, Mingzhe Du, See-Kiong Ng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding the Engagement and Interaction of Superusers and Regular Users in UK Respiratory Online Health Communities: Deep Learning–Based Sentiment Analysis

Xiancheng Li[1*], PhD; Emanuela Vaghi[2*], MD; Gabriella Pasi[2], PhD; Neil S Coulson[3], PhD; Anna De Simoni[4*], PhD; Marco Viviani[2*], PhD; AD HOC Group[5]

[1]School of Business and Management, Queen Mary University of London, London, United Kingdom

[2]Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy

[3]School of Medicine, University of Nottingham, Nottingham, United Kingdom

[4]Wolfson Institute of Population Health, Asthma UK Centre for Applied Research, Queen Mary University of London, London, United Kingdom

[5]See Acknowledgments

[*]these authors contributed equally

Corresponding Author:
Xiancheng Li, PhD
School of Business and Management
Queen Mary University of London
Mile End Road, Bethnal Green
London, E14NS
United Kingdom
Phone: 44 2078825555
Email: x.l.li@qmul.ac.uk

Abstract

Background: Online health communities (OHCs) enable people with long-term conditions (LTCs) to exchange peer self-management experiential information, advice, and support. Engagement of "superusers," that is, highly active users, plays a key role in holding together the community and ensuring an effective exchange of support and information. Further studies are needed to explore regular users' interactions with superusers, their sentiments during interactions, and their ultimate impact on the self-management of LTCs.

Objective: This study aims to gain a better understanding of sentiment distribution and the dynamic of sentiment of posts from 2 respiratory OHCs, focusing on regular users' interaction with superusers.

Methods: We conducted sentiment analysis on anonymized data from 2 UK respiratory OHCs hosted by Asthma UK (AUK), and the British Lung Foundation (BLF) charities between 2006-2016 and 2012-2016, respectively, using the Bio-Bidirectional Encoder Representation from Transformers (BioBERT), a pretrained language representation model. Given the scarcity of health-related labeled datasets, BioBERT was fine-tuned on the COVID-19 Twitter Dataset. Positive, neutral, and negative sentiments were categorized as 1, 0, and –1, respectively. The average sentiment of aggregated posts by regular users and superusers was then calculated. Superusers were identified based on a definition already used in our previous work (ie, "the 1% users with the largest number of posts over the observation period") and VoteRank, (ie, users with the best spreading ability). Sentiment analyses of posts by superusers defined with both approaches were conducted for correlation.

Results: The fine-tuned BioBERT model achieved an accuracy of 0.96. The sentiment of posts was predominantly positive (60% and 65% of overall posts in AUK and BLF, respectively), remaining stable over the years. Furthermore, there was a tendency for sentiment to become more positive over time. Overall, superusers tended to write shorter posts characterized by positive sentiment (63% and 67% of all posts in AUK and BLF, respectively). Superusers defined by posting activity or VoteRank largely overlapped (61% in AUK and 79% in BLF), showing that users who posted the most were also spreaders. Threads initiated by superusers typically encouraged regular users to reply with positive sentiments. Superusers tended to write positive replies in threads started by regular users whatever the type of sentiment of the starting post (ie, positive, neutral, or negative), compared to the replies by other regular users (62%, 51%, 61% versus 55%, 45%, 50% in AUK; 71%, 62%, 64% versus 65%, 56%, 57% in BLF, respectively; $P<.001$, except for neutral sentiment in AUK, where $P=.36$).

XSL•FO
RenderX

Conclusions: Network and sentiment analyses provide insight into the key sustaining role of superusers in respiratory OHCs, showing they tend to write and trigger regular users' posts characterized by positive sentiment.

(J Med Internet Res 2025;27:e56038) doi:10.2196/56038

KEYWORDS

social media; online health communities; social network analysis; sentiment analysis; bio-bidirectional encoder representations from transformers; asthma; chronic obstructive pulmonary disease

Introduction

Background

Online health communities (OHCs) have been increasingly explored in recent years as a means of enabling people with long-term conditions (LTCs) to exchange peer self-management support [1-3]. Such communities offer an easily accessible and cost-effective means of sharing experiences, exchanging information, and providing mutual support to one another [4,5]. Participation in OHCs for individuals living with LTCs could address part of the health care service demand and indirectly improve access to health care [6]. The analysis of the role of OHCs in health promotion and management of LTCs indicates that there might be a positive effect on patients' perception of social support, health literacy, clinical outcomes, and behavior change [7,8]. Furthermore, the involvement of patients in these OHCs can improve their engagement with respect to their care and their ability to self-manage, their mental health outcomes [9], and contribute to health equity [3]. However, despite the growing popularity of OHCs, there is still much we do not know about how these communities function [10]. Moreover, the specific nature of regular users' interaction with the so-called superusers—that is, individuals who frequently engage with the community—and the extent to which it supports self-management remains largely unknown [7].

Recent social network analysis performed on 2 active respiratory OHCs has suggested that superusers play a critical role in holding together the community and ensuring timely exchange of support and information [5,10]. These superusers have been shown to contribute more content to the community, initiate more interactions, and respond more often to other users' queries than regular members [8,9]. From a topological point of view, their characteristics are similar to those of hubs, that is, nodes with a disproportionally large number of connections compared to other nodes in the network.

Across a variety of empirical domains, it has been documented that hubs are valuable resources that help facilitate the spread of information widely and amplify information cascades [11], for example, help design effective vaccination campaigns and selective immunization strategies against disease diffusion and epidemics [12,13] and help improve the system's robustness and vulnerability to random failures [14]. However, some hub identification approaches can be very time-consuming and suffer from the possibility that spreaders are so close together that they overlap the sphere of influence. In this context, VoteRank is a simple iterative method to identify a set of decentralized spreaders with the best spreading ability [15]. In this approach, all nodes vote in a spreader in each turn, and the voting ability of neighbors of the elected spreader will be decreased in the subsequent turn. It is, therefore, an effective solution for identifying possible nonoverlapping superusers.

However, the analysis of the network topology alone is not sufficient to fully understand the interactions between regular users and superusers and their impact on the whole community. For this reason, it is necessary to analyze the content of posts and what relationships (if any) exist between the 2 groups with respect to how they react to each other's content. Sentiment analysis (SA), that is, a subfield of natural language processing provides an understanding of the sentiment of posts and whether there is a cause-and-effect relationship between posts in a thread. This approach consists of analyzing digital text to determine its polarity, that is if the emotional tone of the message is positive, negative, or neutral. SA can create structured and actionable knowledge from unstructured text for decision makers [16] in different fields, from marketing to politics and health [17]. In particular, with respect to the health domain, a variety of works have used SA techniques (both lexicon-based and semantic-based) [18] in recent years for different health conditions, for example, assessing the degree of psychological distress linked to COVID-19 [19,20], evaluate the risk of alcoholism in particular categories of users [21], analyze the emotional state of users with diabetes [22], or the role digital platforms in mediating health-related support with respect to specific cancer drugs [23]. In most of these works, a distinction is not made between categories of users, their interactions, and their role in OHCs.

This work is part of a research program that will eventually test whether promoting engagement in OHC improves self-management and clinical outcomes [24]. The primary motivation of this study is to investigate the engagement patterns of different user types, particularly superusers and regular users, and how their interactions influence the overall sentiment of posts. Our hypothesis is that superusers play a pivotal role in community cohesion, offering immediate access to a support network for self-management, as well as emotional and illness-related support. By doing so they foster positive sentiment among regular users, which subsequently may mediate improvements in self-management behaviors [25,26]. By understanding these dynamics, we aim to provide insights that can enhance the effectiveness of OHCs.

Using a semantic approach, this study aims to explore the sentiment of posts in 2 dynamic and active respiratory OHCs; in doing this, regular user interactions with superusers are assessed, in order to shed light on the impact of such interactions on users' sentiment and which may ultimately impact on the self-management of their LTCs. Specifically, we investigate the sentiments of both regular users and superusers expressed in these interactions as well as their patterns over time.

XSL•FO
RenderX

Additionally, we aim to compare the sentiment of superusers' posts, with superusers defined in 2 different ways, one with emphasis on high-posting activity and the other on high-spreading ability, to verify whether they display similar characteristics or represent indeed the same population.

By shedding light on these critical aspects, this study contributes to understanding the mechanisms underlying the effectiveness of OHCs as a tool to facilitate self-management and provides insights into how respiratory OHCs may meet the needs of their users.

Data

As described in our previous work [10], data were collected by HealthUnlocked [27], the platform provider of the Asthma UK (AUK) and British Lung Foundation (BLF) communities. In both communities, registered users can choose to either write posts publicly or send private posts to one another. In the latter case, posts are shared between 2 users only, whereas when posts are written publicly, other users can become connected through threads of posts. For this study, only posts that were shared publicly were considered. Our datasets were stored and analyzed in a Safe Haven space, that is, a secured database held by Queen Mary University. Anonymized user IDs were provided by HealthUnlocked, and no demographic information was available. The datasets included posts and their metadata including the date of posting, the hierarchical level of the post within the corresponding thread, and the dates in which the users joined and left the community. No data were collected on participants' characteristics, though only people declaring themselves to be older than 16 years of age were permitted to create an account and take part in OHCs.

Six different types of data associated with the corresponding user actions were collected for each user including (1) posts followed, (2) users followed, (3) likes, (4) level-0 posts (ie, posts starting new threads), (5) level-1 replies (ie, replies to the level-0 posts), and (6) level-2+ replies (ie, replies to other replies beyond level 2). The original datasets consisted of 32,780 data items associated with 3345 users from 2006 to 2016 for AUK, and 875,151 data items associated with 19,837 users from 2012 to 2016 for BLF. Since in this study, we are interested in analyzing only the textual content associated with posts, and some of them turned out to be without any content, they were removed from the datasets. The final datasets, therefore, contained 12,413 and 369,224 posts for AUK and BLF, respectively. In 2015, HealthUnlocked took over the AUK forum, leading to substantial increases in posting activity and volume of users. Further details are provided in the posting activity section of our previous work [10].

Methods

Study Design

Superusers were first identified using two different methodological approaches: (1) their posting activity and (2) their spreading ability. Next, SA was applied first to all posts and then to interactions between superusers and regular users.

Identification of Superusers

Two ways of identifying superusers were considered. The first method was based on identifying the "top 1% of users characterized by the largest number of posts written in the community over the entire observation period," as previously described in a study by Joglekar et al [10]. The second method approximates being a "spreader" to being a superuser, according to the VoteRank algorithm [15]. This algorithm is implemented in Python (Python Software Foundation) in the *NewtworkX* package [28]. The VoteRank algorithm finds the top-ranked nodes as spreaders according to an influence ranking. The idea behind VoteRank's rank is to choose a set of spreaders one by one according to the voting scores of nodes obtained from the neighbors. The node that gets the most votes in each turn is selected as the most influential node. It is an iterative method where at the beginning all nodes take part in ranking their neighbors. However, when a node is identified as a spreader, it will no longer take part in subsequent iterations and neighboring nodes will have a penalty, so that nodes that are not significant in the transmission of information but exploit proximity to the influencing nodes are not considered as spreaders. To make a fair comparison, we identified and compared the same number of superusers according to the 2 definitions. This was achieved by picking the k top-ranked spreaders by VoteRank, where k is the number of superusers according to the "top 1%" definition.

Sentiment Analysis

SA was carried out by means of a semantic approach based on bidirectional encoder representations from transformers (BERT). BERT is a contextualized word representation pretrained language model [29,30]. Its architecture is a multilayer bidirectional transformer encoder based on the original transformer implementation. Vaswani et al [31] shows further details about the transformer architecture.

BERT is built in 2 steps: pretraining and fine-tuning. The model is trained first on unlabeled data over different pretraining tasks. Later, the model with the pretrained parameters is fine-tuned using labeled data from downstream tasks. Every task has a separate fine-tuned model. However, a unique characteristic of BERT is that it has a unified architecture across different tasks, so the difference between pretrained architecture and the final downstream is small. We used BioBERT (Bio-Bidirectional Encoder Representation from Transformers), which is a pretrained language representation model for the biomedical domain [32].

Fine-Tuning BioBERT for SA

BioBERT was fine-tuned using a COVID-19 Twitter Dataset [33] taken from Kaggle [34]. The choice was opportunistic as there were few open datasets related to health. The COVID-19 Twitter Dataset has a total of 143,903 usable records, with labels associated with neutral, positive, and negative posts. Numeric values have then been associated with sentiment: 0 for neutral, 1 for positive, and –1 for negative posts. Examples of posts with sentiment labels are given in the Multimedia Appendix 1. Normalization and stop-words removal were performed on both datasets.

XSL•FO
RenderX

The considered SA workflow performed on AUK and BLF is illustrated in Figure 1. The BioBERT model was initialized by using the standard configurations and weights from the Hugging Face repository [35]. As a first step, we converted the data to sequences, adding the tags to indicate where the sentence starts and its separator. Then, we tokenized the resulting sequences with the BioBERT tokenizer creating a tensor dataframe of 128 characters. The dataset was then split into training and validation sets, with 85% observations for training and 15% observations for validation. The Adam function was used as an optimizer [36], the sparse categorical cross-entropy as loss function [37], and the sparse categorical accuracy as a metric for training [38]. The training was performed on just 2 epochs with a batch size of 32. Google Colab Pro was used to fine-tune BioBERT [32]. After the fine-tuning, weights, and result configuration files were stored.

Figure 1. Sentiment analysis workflow. BioBert: Bio-Bidirectional Encoder Representation from Transformers.

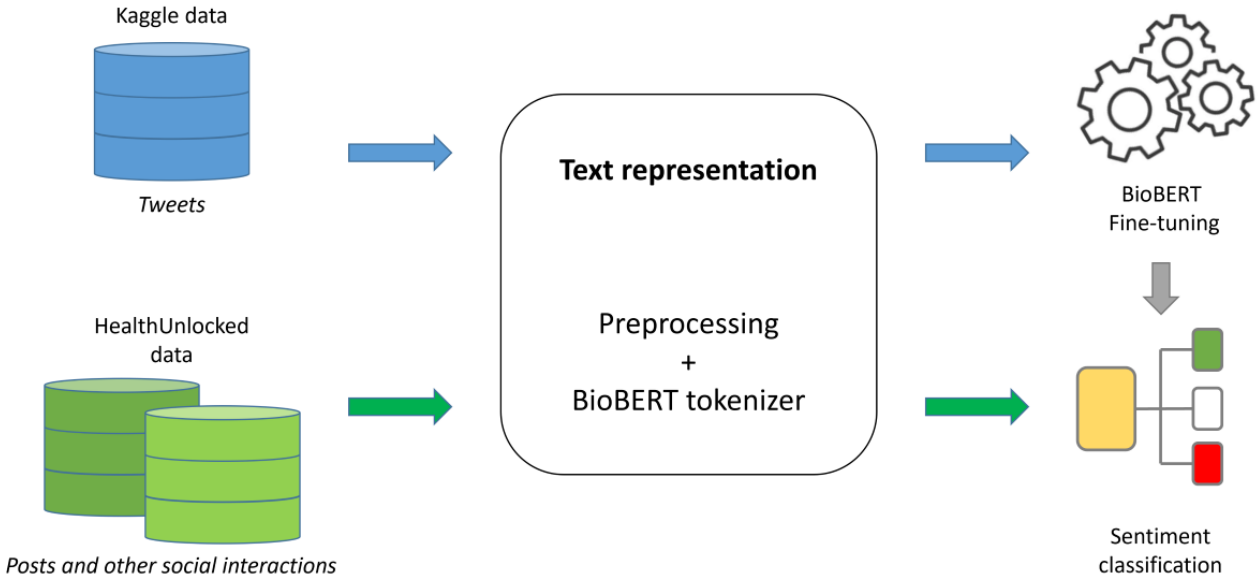

SA on AUK and BLF

By using the earlier fine-tuned model, we performed SA on AUK and BLF data, including all posts (it is important to note that in this paper when we talk about posts, we also generally include replies; when we consider only replies, we refer to them directly). We performed the following analyses.

Average Sentiment

Average sentiment scores (AVSs) for both regular users' and superusers' posts have been computed separately. As sentiment labels are associated with numeric values (ie, 0 to neutral, 1 to positive, and –1 to negative), the average values of the sentiment of a set of posts can be computed and used to capture the general sentiment expressed in those posts. Specifically, the AVSs range in the [–1, 1] interval, where positive values and getting closer to 1 indicate increasing positive sentiment, while negative values and getting closer to –1 indicate increasing negative sentiment. We considered different aggregations of content types as follows: (1) , the AVS of all posts of a set of users in *v*; that is, either regular users (*U*) or superusers (*S*); (2) , the AVS of level-0 posts of users in *v*; (3) , the AVS of all replies given by users in *v*; (4) , the AVS of level-1 replies given by users in *v*; and (5) , the AVS of level-2+ replies given by users in *v*.

Along with the AVSs, we also computed the percentage of posts with different sentiments to show the distribution of sentiment expressed in users' posts. To investigate the trend of the average sentiment over time, we sorted all posts based on their publication time and regrouped them into 15 bins with an equal volume, for which we computed AVSs. We did not aggregate posts based on the month and year of publication because the number of available data was too small in AUK before 2015 (when HealthUnlocked took over the forum), and because the period of analysis in the 2 communities was different. 2-tailed *t* tests were used for statistical significance when comparing 2 AVSs of different users.

User-Superuser Interaction Sentiment

These analyses investigated the interactions between regular users and superusers. We addressed the following.

- Regular users' and superusers' sentiments when replying to each other: To do this, we first identified threads started by regular users and superusers, denoted as TU and TS, and calculated the AVS of replies written by the other group of users. Specifically, we computed , which denotes the average sentiment of regular users in reply to superusers' initiated threads; and , which denotes the average sentiment of superusers in reply to regular users' initiated threads.
- Regular users' sentiments when replying to other regular users and superusers: Here, we focused on regular users' replies only and checked whether they acted differently when replying to other regular users or superusers. Specifically, we compared with respect to which

XSL•FO
RenderX

denotes the average sentiment of regular users in replying to other regular users' threads. A similar approach to that introduced earlier was used to analyze the trend in the value of  over time.

- Superusers' sentiments when replying to regular users: In this case, we only took into consideration the sentiment of the replies superusers give to regular users. Note that superusers also interact with each other, but the study of these interactions is beyond our research questions and is not shown here. We investigate the sentiment of superusers in reply to positive, negative, and neutral level-0 posts of regular users. To do this, we computed  , which denotes the average sentiment of superusers in replying to regular users' positive level-0 posts;  , which denotes the average sentiment of superusers in replying to regular users' negative level-0 posts; and  , which denotes the average sentiment of superusers in replying to regular users' neutral level-0 posts.

Regular users' replies to other regular users' level-0 posts are used as a baseline and compared with AVSs of superusers. This analysis assesses in more detail the superusers' tendency to act as help-givers [10], especially when regular users express negative sentiments. Along with the AVSs, we also computed the percentages of posts with different sentiments to show the distribution of sentiment expressed in users' posts. The 2-sample *z* tests for proportions were used for statistical significance when comparing 2 percentages of different users.

A similar approach to that introduced earlier was used to analyze the trend in these percentages over time.

Ethical Considerations

The study was approved by the Queen Mary University Research Ethics Committee (QMERC 22.279). The research team did not have access to personally identifiable information. The data was anonymized to ensure user privacy, and no demographic information was included in the analysis. The posts analyzed were publicly available, with the users having consented to their use for analytical purposes by choosing to share them publicly. In addition, the research protocol was examined and permission to undertake the research was obtained by AUK and BLF charities, and as well as HealthUnlocked. To further protect the privacy of the users, no posts are directly quoted.

Results

Overview

The ratio of distinct actions over all actions performed with respect to both OHCs is shown in Figure 2. In AUK, the most common action was to "follow" users, while for BLF was to "like" posts or replies. In both communities, the action of generating level-0 posts, that is, starting a new thread, was smaller than generating both level-1 and level-2+ replies, showing that users mostly communicated through replies.

Figure 2. Distribution of different actions in distinct communities. (A) Asthma UK (AUK); (B) British Lung Foundation (BLF).

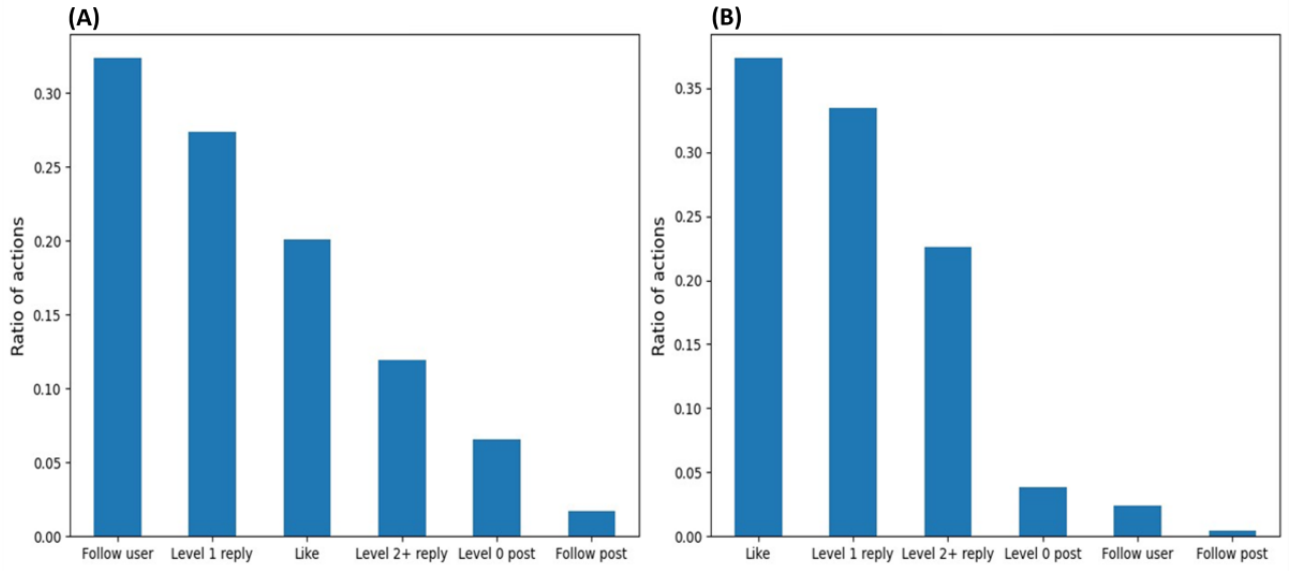

Identification of Superusers

Using the definition of superusers as the "top 1% of users characterized by the largest number of posts written in the community over the entire observation period" yielded 33 and 198 superusers in the AUK and BLF communities, respectively. A detailed description of users' posting activity can be found in our previous work [10].

Then, we used the second definition of superusers with VoteRank to identify the same numbers of superusers. The ratio of overlap of superusers according to the 2 definitions was greater than 60% (20/33 in AUK and 156/198 in BLF), showing that users who posted the most (1% definition) were largely overlapping with the spreaders (VoteRank definition). As the results observed from both definitions were consistent in this study, we only show in the following the results obtained by using the 1% definition and have moved those obtained with the VoteRank definition to Multimedia Appendix 2. In both communities, 60%-70% of registered users never wrote any

XSL•FO
RenderX

posts, which shows that the majority of regular users were passive (ie, the so-called lurkers).

Figure 3 compares the number of posts as well as the length of text written by both categories of users. Specifically, Figures 3A and 3B show that superusers tended to reply more often than starting new posts, as described in our previous work [10]. Figures 3C and 3D show that both superusers and regular users tended to write longer level-0 posts than replies. In addition, we can observe that superusers' level-0 posts and replies are significantly shorter than regular users' ones in AUK, while they are of similar length in BLF.

Figure 3. Comparisons of the number and length of posts written by superusers and regular users. (A,C) Asthma UK (AUK); (B,D) British Lung Foundation (BLF).

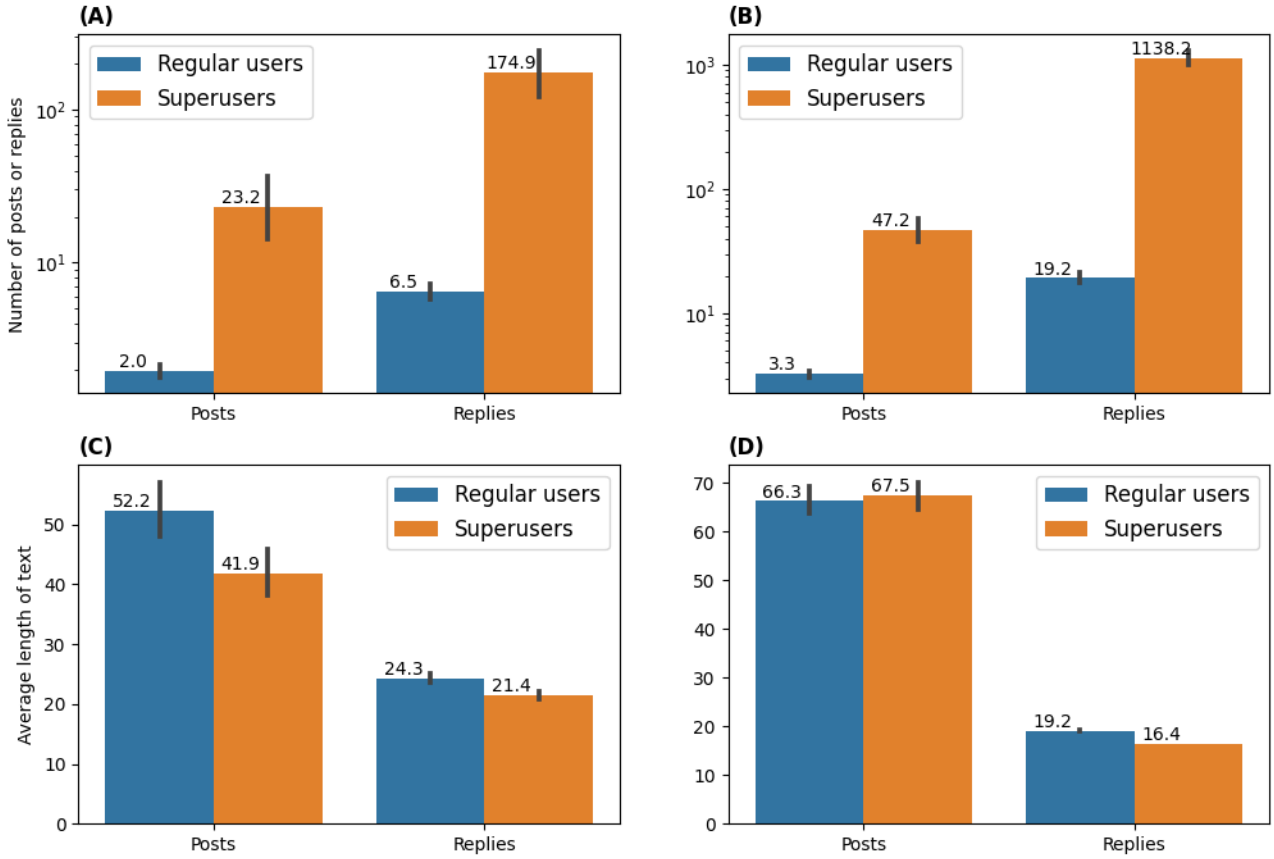

Sentiment Analysis

The fine-tuned BioBERT-based SA with the COVID-19 Twitter Dataset achieved an accuracy of 96% on the validation set.

Average Sentiment

In both communities, most posts were associated with positive sentiment, as illustrated in Figure 4. Figures 4C and 4D show the percentage of posts with positive, negative, and neutral sentiments, across the 15 bins, related in Figures 4A and 4B to sentiment versus total number of posts. In both communities, posts with positive sentiment were always the ones with higher frequency.

Table 1 shows the percentages of sentiment across all posts, along with the overall average sentiment. The average sentiment of superusers' posts was consistently higher than the one for regular users. Both superusers and regular users tended to be significantly more positive in replies compared to their level-0 posts (P<.001).

Figure 5 illustrates the average sentiment in level-0 posts, level-1 replies, and level-2+ replies. Although the patterns of superusers' sentiment were different in the two communities, regular users tended to be the most positive in level-2+ replies and least positive in level-0 posts. This result further confirms that regular users tended to be more positive when having in-depth communications with others in the community.

Figure 6 shows the trend of regular users' and superusers' sentiments over time. There is a trend toward positive sentiment for all users, with superusers being consistently more positive than regular users.

XSL•FO
RenderX

Figure 4. Number of posts and percentage of posts with different sentiments over time. (A,C) Asthma UK (AUK); (B,D) British Lung Foundation (BLF).

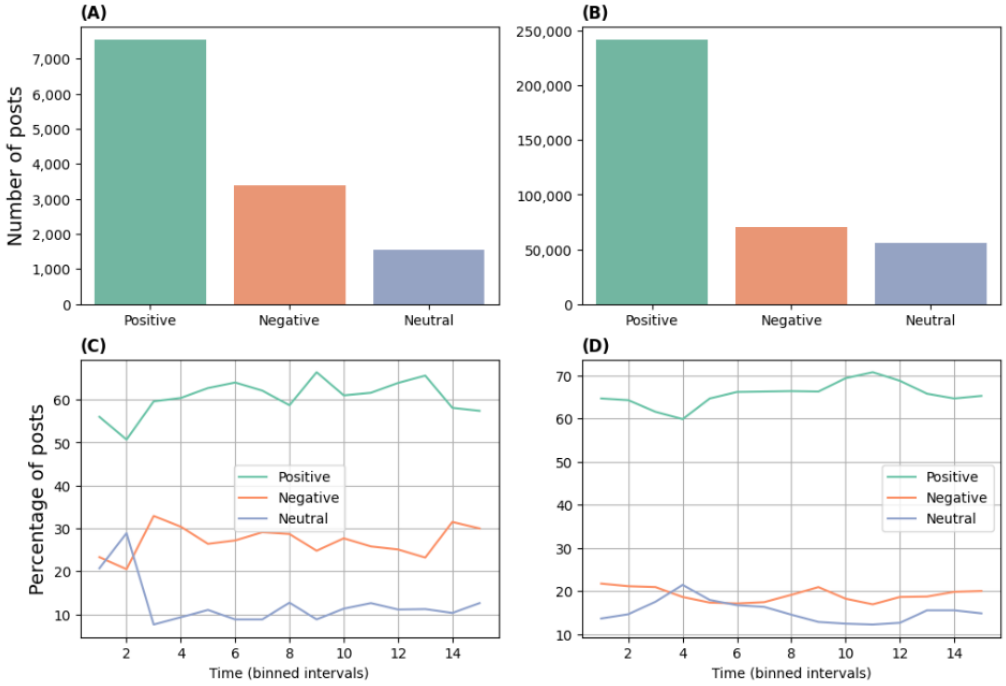

Table 1. Percentages of posts or replies with different sentiments and the average sentiments indicated as AVS[a]. Depending on the target, it refers to either , , or .

	AUK[b]		BLF[c]	
	Superusers	Regular users	Superusers	Regular users
All posts				
Negative, %	22.94	31.46	17.20	22.34
Neutral, %	13.64	11.03	15.92	14.03
Positive, %	63.41	57.50	66.88	63.63
AVS	0.405	0.26	0.497	0.413
Level-0 posts				
Negative, %	30.19	45.92	30.92	33.86
Neutral, %	4.95	8.35	11.06	11.88
Positive, %	64.86	45.73	58.03	54.26
AVS	0.347	-0.002	0.271	0.204
Replies				
Negative, %	22.16	28.42	16.73	21.13
Neutral, %	14.59	11.60	16.09	14.26
Positive, %	63.25	59.98	67.19	64.62
AVS	0.411	0.316	0.505	0.435

[a]AVS: average sentiment score.

[b]AUK: Asthma UK.

[c]BLF: British Lung Foundation.

XSL•FO
RenderX

Figure 5. Average sentiment score (AVS) in different actions. (A) Asthma UK (AUK); (B) British Lung Foundation (BLF).

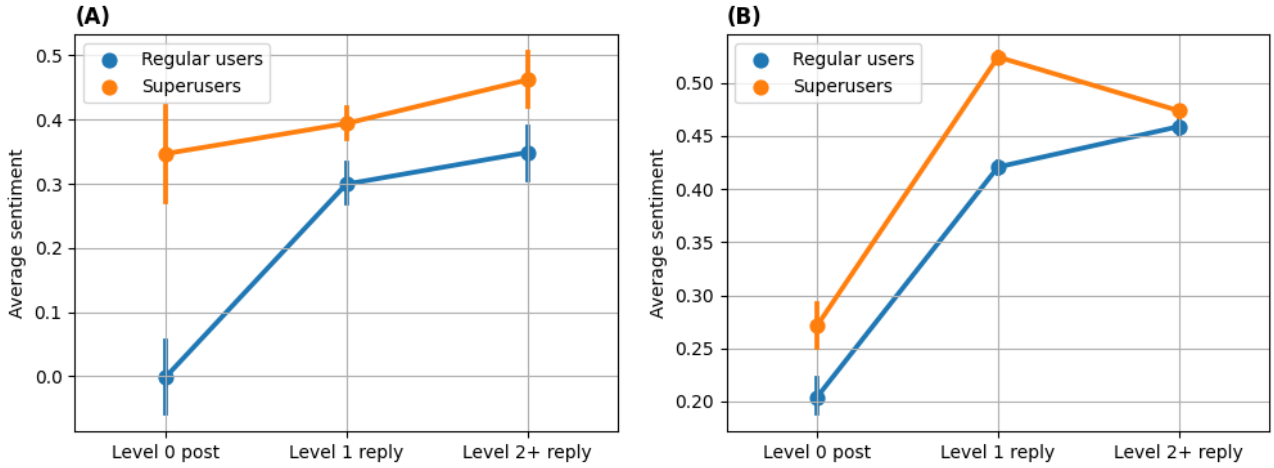

Figure 6. Trend of sentiment. All posts are sorted based on their publication time and regrouped into 15 bins with equal volume. Average sentiment of all posts written by regular users and superusers are calculated separately in each bin. (A) Asthma UK (AUK); (B) British Lung Foundation (BLF).

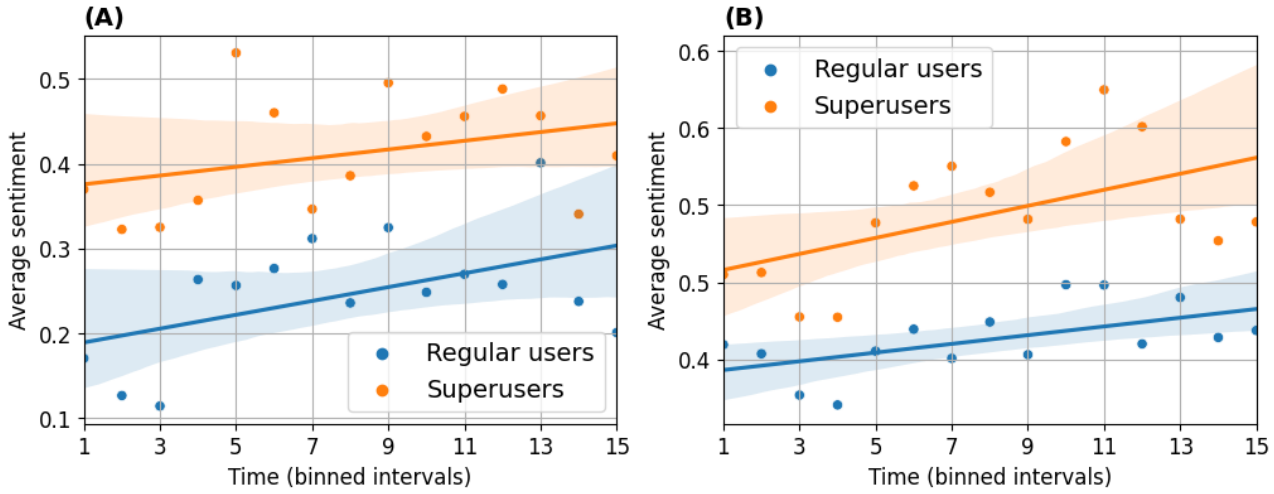

User-Superuser Interaction Sentiment

Regular Users' and Superusers' Sentiments When Replying to Each Other

We first analyzed how regular users reacted to superusers' posts. We specifically identified the threads started by superusers and investigated the sentiments as well as the lengths of replies written by regular users in those threads (Table 2). We compared them with the replies written by superusers in threads started by regular users (Table 2). We showed the number of threads started by different groups of users, the number of threads with replies from the other group of users, AVS, and the average length of text ( ). We calculated the AVS and of (1) posts written by the focal group with replies from the other group, and (2) replies written by the other group of users.

We can observe that, in both communities, superusers started more than a third of all threads (ie, 626/1680 in the AUK and 7831/20,756 in the BLF communities). Regarding the sentiment in posts and replies, the average sentiment of starting posts written by superusers with replies written by superusers at the start of a thread was much more positive than those written by regular users (ie, 0.395 vs –0.004 in AUK and 0.25 vs 0.137 in BLF). Superusers tended to reply in a way that was much more positive than the sentiment expressed in the level-0 posts of regular users. Most interestingly, the sentiment of regular users in reply to superusers' threads was as positive (0.489 vs 0.470 in BLF) or even more positive than superusers' sentiment when replying to regular users (0.447 vs 0.352 in AUK).

In terms of the length of texts, similar to what we found in Figure 3, replies of both superusers and regular users tended to be shorter than level-0 posts. However, superusers tended to write longer texts when replying to regular users than regular users replying to them. This is opposite to what we found in Figure 3, where we show the overall length of superusers' replies tended to be shorter than regular users. This observation suggests that superusers were willing to provide more information as help-givers when replying to regular users.

XSL•FO
RenderX

Table 2. Interactions between superusers and regular users. Depending on the target, the AVS[a] of posts and AVS of replies refer to either [x] and [x], or [x] and [x]. Posts without any replies are excluded from the AVS calculation.

	Threads in AUK[b]		Threads in BLF[c]	
	Superusers	Regular users	Superusers	Regular users
Total threads	626	1054	7831	12,925
Threads with replies	352	719	6088	10,768
AVS of posts	0.395	–0.004	0.250	0.137
AVS of replies	0.447	0.352	0.489	0.470
[x][d] of posts	36.764	47.250	68.836	42.487
[x] of replies	19.266	25.410	16.464	20.029

[a]AVS: average sentiment score.

[b]AUK: Asthma UK.

[c]BLF: British Lung Foundation.

[d][x]: average length of text.

Regular Users' Sentiments When Replying to Other Regular Users and Superusers

To explore how regular users reacted toward superusers' threads, we investigated superusers' posts with different sentiments separately. Table 3 represents the number of posts written by superusers, the percentage of posts with regular users' replies, and the average sentiment and length of text in those replies. In both communities, the sentiment in regular users' replies was largely affected by the sentiment of superusers in their posts. Regular users tended to be less positive and write longer texts when replying to superusers' posts with negative sentiments.

To investigate how regular users' sentiments in replies to superusers' posts change over time and whether they behaved differently when replying to superusers, we show the trend of regular users' sentiments in threads started by superusers and regular users in Figure 7. In both communities, the sentiment of regular users when replying to superusers' threads was becoming more positive while their sentiment in replying to regular users' threads was relatively stable. The results suggest that superusers were somehow instrumental in the trend toward more positive sentiment from regular users over time.

Table 3. Summary of superusers' threads and how regular users replied to them. [x], [x], and [x] are computed for AVS[a] in corresponding columns.

	Sentiment of superusers (AUK[b])			Sentiment of superusers (BLF[c])		
	Negative	Neutral	Positive	Negative	Neutral	Positive
Posts	189	31	406	2421	866	4544
Replies, %	51.32	61.29	58.13	81.91	69.40	77.11
AVS of replies	0.230	0.508	0.495	0.364	0.435	0.503
[x][d] of replies	24.11	19.41	19.14	19.84	14.11	16.76

[a]AVS: average sentiment score.

[b]AUK: Asthma UK.

[c]BLF: British Lung Foundation.

[d][x]: average length of text.

XSL•FO
RenderX

Figure 7. Sentiment trend in interactions between superusers and regular users. Average sentiment of regular users when replying to threads started by other regular users and superusers are calculated separately in each bin. (A) Regular users' sentiments Asthma UK (AUK); (B) Regular users' sentiments British Lung Foundation (BLF).

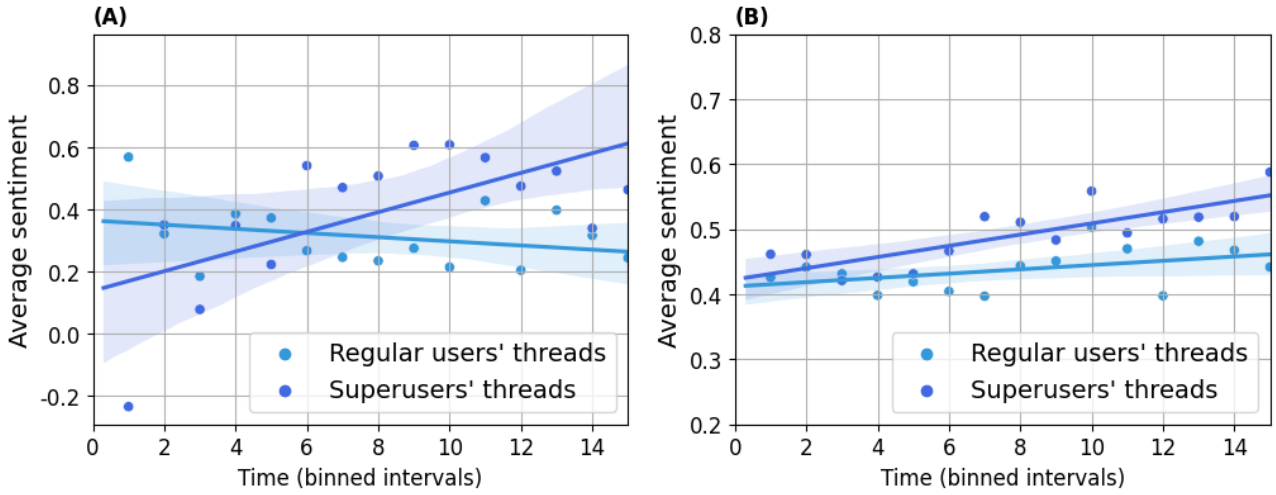

Superusers' Sentiments When Replying to Regular Users

In Table 4, we compared the sentiments of superusers and regular users when replying to threads started by regular users with different sentiments expressed in their level-0 posts. From the sentiment in regular users' replies, we found that regular users' sentiment was largely affected by the sentiment expressed in the level-0 posts. On the other hand, superusers tended to write positive replies in threads started by regular users, whatever the type of sentiment of the starting post, positive, neutral, or negative, compared to regular users (62%, 51%, and 61% vs 55%, 45%, and 50% in AUK and 71%, 62%, and 64% vs 65%, 56%, and 57% in BLF, respectively). $P<.001$ for all pairs except for that in neutral threads in AUK, where $P=.36$. The result suggests superusers replied to regular users in a positive way as help-givers.

In Figure 8, we show the percentage of superusers' replies against the sentiments of the users' posts with different sentiments over time. The percentage of superusers' positive replies only marginally changed over time, and the majority of superusers' replies were consistently associated with positive sentiments, whatever the users' posts' sentiment.

Table 4. Percentage of replies with different sentiments and average sentiments based on the sentiment in the level-0 post in each thread written by regular users. , , and are computed for AVS[a] in corresponding columns. Regular users' replies to other regular users' level-0 posts are used as a baseline and compared with AVSs of superusers.

	Sentiment of regular users (AUK[b])			Sentiment of regular users (BLF[c])		
	Negative	Neutral	Positive	Negative	Neutral	Positive
Regular users' replies						
Negative, %	39.83	21.15	31.93	31.03	22.58	20.80
Neutral, %	9.87	33.65	12.09	11.82	20.48	14.07
Positive, %	50.30	45.19	55.98	57.15	56.94	65.13
Total, %	100	100	100	100	100	100
AVS	0.105	0.24	0.24	0.261	0.344	0.443
Superusers' replies						
Negative, %	30.87	24.72	25.83	25.41	16.75	15.93
Neutral, %	7.83	23.60	11.75	10.07	20.36	12.75
Positive, %	61.30	51.69	62.42	64.53	62.89	71.32
Total, %	100	100	100	100	100	100
AVS	0.278	0.319	0.388	0.395	0.463	0.567

[a]AVS: average sentiment score.

[b]AUK: Asthma UK.

[c]BLF: British Lung Foundation.

XSL•FO
RenderX

Figure 8. Sentiment trend in replies of superusers to regular users. Percentages of replies with corresponding sentiments are computed and shown in each bin. (A,C,E) Asthma UK (AUK); (B,D,F) British Lung Foundation (BLF).

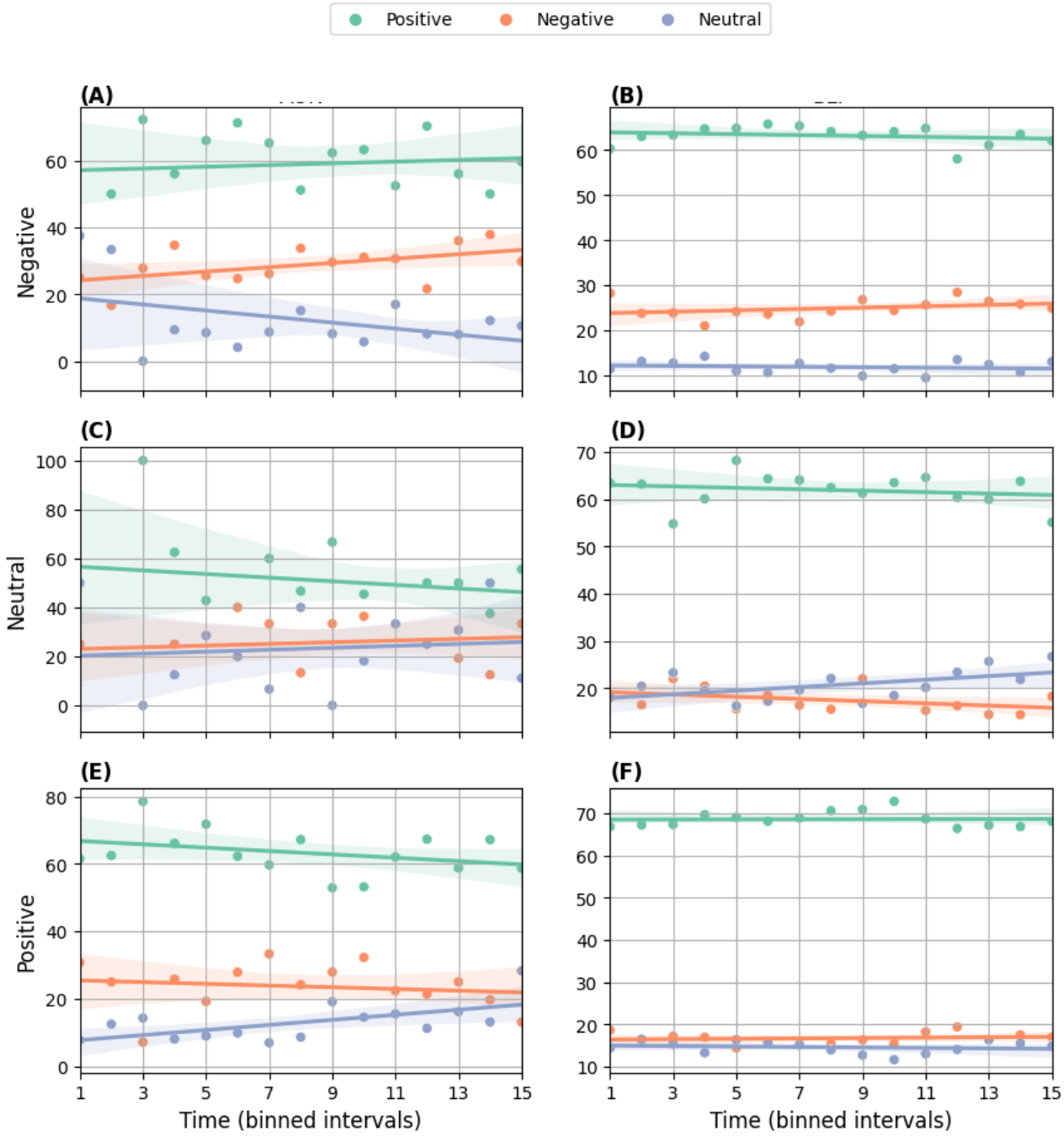

Discussion

Principal Results

This study shed novel light on regular users' and superusers' engagement in OHCs and on the impact of their interaction on the sentiment of communication (ie, posts). First, we showed that superusers who posted the most were largely overlapping with the "spreaders" of information and support. Then we found that superusers had a key role in promoting positive sentiment in OHCs, which could represent one of the mechanisms underlying their OHC "sustaining" role. In both OHCs, the great majority of posts were characterized by positive sentiment. This trend did not change over time. The average sentiment of superusers' posts was consistently higher than that of regular users. Both superusers and regular users tended to be significantly more positive in their replies, especially at level 2+ replies compared to their sentiments in level-0 posts. Although this was likely due to the first posts of threads including a request for help presented within the context of a personal story, this could also be interpreted as regular users becoming more positive as a result of the engagement with the community and interaction with superusers. Indeed, the majority of superusers' replies were consistently characterized by positive sentiments, whatever the users' sentiment (negative, neutral, or positive) at the start of threads. Superusers wrote longer posts when replying to regular users, despite their overall tendency to write short posts. This suggests that superusers were willing

XSL•FO
RenderX

to dedicate more time to providing information in replies to regular users compared to other regular users themselves. Moreover, we found evidence of regular users' change toward a more positive sentiment after interaction with superusers. In fact, there was a trend toward more positive sentiment by regular users when replying to threads started by superusers, while their sentiment in replying to threads started by regular users was relatively stable. These results suggest that the active participation of superusers with consistently positive sentiment can not only enhance community cohesiveness but also foster an encouraging environment conducive to positive interactions and ensure the effective spread of informational and emotional support [4-9].

These findings have significant implications for both researchers and policy makers. Superusers in OHCs may represent a scalable and cost-effective health care workforce, providing a means for health and social care integration [6]. Their significant contributions have the potential to alleviate pressure on formal health services by promoting self-management and enhancing community-based care. As current health care challenges emphasize self-care and local service expansion [10], superusers could provide vital support by increasing the confidence of individuals in managing their conditions, reducing the demand on general practices, emergency care, and hospitals, and ultimately saving money within health care systems and across society.

By understanding the mechanisms of community sustainability and the impact of superuser involvement, policy makers and researchers can leverage these individuals as part of a broader health care team to develop and test more effective interventions. Future research is needed to explore how superusers might be formally recognized as contributors to health care, functioning as allied professionals within digital communities. This approach addresses the need for accessible and effective self-management interventions that integrate peer support into health care systems.

Strengths and Limitations

A key strength of this work is the use of previously characterized datasets from 2 established active and dynamic respiratory OHCs [10]. The paper analyzes extensive data from the 2 significant OHCs with activities of more than 20,000 users in total. Based on SA, this work has started elucidating crucial mechanisms underlying the potential of superusers to affect the sentiment of OHC users. Besides, the paper shows the consistency in results with 2 different definitions of superusers, providing a robust basis for its findings.

A limitation of this study is the lack of a domain-specific labeled dataset for the fine-tuning of BioBERT. The choice to use this dataset for fine-tuning BioBERT was opportunistic as there are few open datasets related to health in social networks, and most are about COVID-19. In our study, BioBERT was fine-tuned on a COVID-19 Twitter Dataset. The brevity of Twitter posts, constrained by character limits, might restrict the richness of health discussions compared to longer, more detailed posts found in OHCs. Additionally, the intense emotional tone surrounding the COVID-19 pandemic may skew SA results toward more extreme expressions of sentiment. This could result in a bias toward stronger sentiment expressions in our model's predictions, potentially differing from sentiment trends in other health-related datasets that are not solely concerned about global health crises. A further limitation is the lack of demographic and clinical information of participants as well as verification and validation of the information shared in the OHCs. In this study, the contents of some posts were analyzed for the validity of our sentiment model, however, findings were not validated through a comprehensive semantic analysis of most posts.

Comparison With Prior Work

The results of this study are consistent with previous work and help illuminate the dynamics within OHCs [10]. These results are also in keeping with previous research on emotional contagion [29], which has shown that one person's mood might fleetingly determine the mood of others. Previous work in social networks suggests that happiness is a network phenomenon, clustering in groups of people that extend up to three degrees of separation [30]. Methodologically, users' sentiment was extracted with semantic-based techniques, which is a common approach used in previous research [19-21]. In most of these works, a distinction is not made between categories of users, their interactions, and their role in OHCs. By investigating the sentiments of regular users and superusers as well as during their interactions, this study provides novel insights into the essential role played by superusers. In agreement with previous research, superusers tended to reply to other users' posts [10] with concise posts in general. In addition, when replying to regular users, they tended to provide more information by writing longer texts. Besides, they contribute more content to the community and respond more often to other users' queries than regular members [8,9]. Comprehending the dynamics of superuser interactions can help us understand more about what effective peer moderation could look like (ie, moderation by superusers). It also helps design strategies for better moderation and engagement in OHCs, potentially improving the support and information exchange in these communities.

Conclusions and Future Research

SA provides insight into the key sustaining role of superusers in respiratory OHCs, showing they tend to write and respond with posts with positive sentiment. Future research is needed to develop an approach able to understand the sequence of responses in a discussion thread within the OHC datasets to understand whether a superuser's response was successful in changing the user's sentiment in the case of negativity. Moreover, the existence of groups of users writing posts with negative sentiments could be explored, whether these users form "echo chambers" [39], therefore reinforcing beliefs and reducing exposure to users with alternative views on the topic of discussion. The evolution of such echo chambers over time could also be characterized. Another interesting development could be topic modeling analysis, in order to test the existence of subcommunities based on topics of discussion, but also to investigate the correlation between topics and sentiment types, and user characteristics associated with different sentiments concerning different topics. In addition, it would be interesting to analyze sentiment associated with topics of discussion, for example, symptoms and treatment, to inform the moderation process and provide tailored information or support. The

XSL•FO
RenderX

ultimate aim would be to develop strategies for better OHC moderation, enhanced effectiveness of engagement, and improved OHC users' safety so that engagement in OHCs could be integrated within health care services for patients with chronic conditions.

Acknowledgments

The authors would like to thank AUK, BLF, and HealthUnlocked for granting them permission to conduct the study. The study was partly funded by the National Institute for Health and Care Research Program Grant for Applied Research (reference 202037; effectiveness and cost-effectiveness of a digital social intervention for people with troublesome asthma promoted by primary care clinicians). The views expressed are those of the author or authors and not necessarily those of the National Health Service, National Institute for Health and Care Research, or the Department of Health and Social Care.

The contributors associated with the AD HOC group are as follows: Pietro Panzarasa, School of Business and Management, Queen Mary University of London, London, UK; Georgios Karampatakis, Helen E Wood, and Chris J Griffiths, Wolfson Institute of Population Health, Asthma UK Centre for Applied Research, Queen Mary University of London, London, UK.

Data Availability

Data sharing is not applicable to this article as no datasets were generated during this study.

Authors' Contributions

XL and EV conceptualized the study, with both taking on leading roles. XL led the data curation, formal analysis, investigation, methodology, project administration, and visualization, while EV contributed equally to data curation and supported the formal analysis and visualization. ADS played a leading role in funding acquisition and shared equal responsibility with MV in resource, supervision, and validation. ADS and MV contributed equally to methodology and project administration. GP supported the methodology and supervision. The writing of the original draft was led by XL, with support from ADS and MV, while the review and editing involved all authors as well as the AD HOC group, in which Pietro Panzarasa contributed to the validation and review of the final manuscript; and Georgios Karampatakis, Helen E Wood, and Chris J Griffiths contributed to the review of the final manuscript.

XL and EV share first authorship, while ADS and MV share last authorship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of random posts with different sentiment classified by BioBERT.
[DOCX File , 16 KB - jmir_v27i1e56038_app1.docx]

Multimedia Appendix 2
Results with superusers identified by VoteRank.
[DOCX File , 418 KB - jmir_v27i1e56038_app2.docx]

References

1. Hossain SN, Jaglal SB, Shepherd J, Perrier L, Tomasone JR, Sweet SN, et al. Web-based peer support interventions for adults living with chronic conditions: scoping review. JMIR Rehabil Assist Technol 2021;8(2):e14321. [doi: 10.2196/14321] [Medline: 34032572]
2. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
3. Petkovic J, Duench S, Trawin J, Dewidar O, Pardo Pardo J, Simeon R, et al. Behavioural interventions delivered through interactive social media for health behaviour change, health outcomes, and health equity in the adult population. Cochrane Database Syst Rev 2021;5(5):CD012932 [FREE Full text] [doi: 10.1002/14651858.CD012932.pub2] [Medline: 34057201]
4. Allen C, Vassilev I, Kennedy A, Rogers A. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res 2016;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
5. De Simoni A, Shah AT, Fulton O, Parkinson J, Sheikh A, Panzarasa P, et al. Superusers' engagement in asthma online communities: asynchronous web-based interview study. J Med Internet Res 2020;22(6):e18185 [FREE Full text] [doi: 10.2196/18185] [Medline: 32573463]

XSL•FO
RenderX

6. De Simoni A, Taylor SJC, Griffiths C, Panzarasa P, Sheikh A. Online "superusers" as allies of the health care workforce. NEJM Catalyst 2018;4(6).
7. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
8. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
9. Bhatti S, Rayner J, Pinto AD, Mulligan K, Cole DC. Using self-determination theory to understand the social prescribing process: a qualitative study. BJGP Open 2021;5(2):BJGPO.2020.0153 [FREE Full text] [doi: 10.3399/BJGPO.2020.0153] [Medline: 33402331]
10. Joglekar S, Sastry N, Coulson NS, Taylor SJ, Patel A, Duschinsky R, et al. How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the Asthma UK and British Lung Foundation online communities. J Med Internet Res 2018;20(7):e238 [FREE Full text] [doi: 10.2196/jmir.9952] [Medline: 29997105]
11. Bakshy E, Hofman JM, Mason WA, Watts DJ. Everyone's an influencer: quantifying influence on twitter. 2011 Presented at: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining; February 09, 2011; Hong Kong, China p. 65-74.
12. Dezso Z, Barabási AL. Halting viruses in scale-free networks. Phys Rev E Stat Nonlin Soft Matter Phys 2002;65:055103. [doi: 10.1103/PhysRevE.65.055103] [Medline: 12059627]
13. Pastor-Satorras R, Vespignani A. Immunization of complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2002;65:036104 [FREE Full text] [doi: 10.1103/PhysRevE.65.036104] [Medline: 11909162]
14. Paul G, Sreenivasan S, Stanley HE. Resilience of complex networks to random breakdown. Phys Rev E Stat Nonlin Soft Matter Phys 2005;72:056130. [doi: 10.1103/PhysRevE.72.056130] [Medline: 16383711]
15. Zhang J, Chen D, Dong Q, Zhao Z. Identifying a set of influential spreaders in complex networks. Sci Rep 2016;6(1):27823 [FREE Full text] [doi: 10.1038/srep27823] [Medline: 27296252]
16. Pozzi FA, Fersini E, Messina E, Liu B. Sentiment Analysis in Social Networks. Cambridge, MA: Morgan Kaufmann; 2016.
17. Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. JMIR Med Inform 2020;8(1):e16023 [FREE Full text] [doi: 10.2196/16023] [Medline: 32012057]
18. Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowl Inf Syst 2018;60(2):617-663. [doi: 10.1007/s10115-018-1236-4]
19. Crocamo C, Viviani M, Famiglini L, Bartoli F, Pasi G, Carrà G. Surveilling COVID-19 emotional contagion on Twitter by sentiment analysis. Eur Psychiatry 2021;64(1):e17 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.3] [Medline: 33531097]
20. Viviani M, Crocamo C, Mazzola M, Bartoli F, Carrà G, Pasi G. Assessing vulnerability to psychological distress during the COVID-19 pandemic through the analysis of microblogging content. Future Gener Comput Syst 2021;125:446-459 [FREE Full text] [doi: 10.1016/j.future.2021.06.044] [Medline: 34934256]
21. Crocamo C, Viviani M, Bartoli F, Carrà G, Pasi G. Detecting binge drinking and alcohol-related risky behaviours from Twitter's users: an exploratory content- and topology-based analysis. Int J Environ Res Public Health 2020;17(5):1510 [FREE Full text] [doi: 10.3390/ijerph17051510] [Medline: 32111047]
22. Gabarron E, Dorronzoro E, Rivera-Romero O, Wynn R. Diabetes on Twitter: a sentiment analysis. J Diabetes Sci Technol 2019;13(3):439-444 [FREE Full text] [doi: 10.1177/1932296818811679] [Medline: 30453762]
23. Cabling ML, Turner JW, Hurtado-de-Mendoza A, Zhang Y, Jiang X, Drago F, et al. Sentiment analysis of an online breast cancer support group: communicating about tamoxifen. Health Commun 2018;33(9):1158-1165 [FREE Full text] [doi: 10.1080/10410236.2017.1339370] [Medline: 28678549]
24. Effectiveness and ost-effectiveness of a digital social intervention for people with troublesome asthma promoted by primary care clinicians (AD-HOC). National Institute for Health and Care Research. URL: https://fundingawards.nihr.ac.uk/award/NIHR202037 [accessed 2024-12-13]
25. Petersen CL, Li X, Stevens CJ, Gooding TL, Carpenter-Song EA, Batsis JA. Adapting natural language processing and sentiment analysis methods for an intervention in older adults: positive perceptions of health and technology. Gerontechnology 2023;22(1):1-6 [FREE Full text] [doi: 10.4017/gt.2023.22.1.824.06] [Medline: 38116325]
26. Pearson JL, Amato MS, Papandonatos GD, Zhao K, Erar B, Wang X, et al. Exposure to positive peer sentiment about nicotine replacement therapy in an online smoking cessation community is associated with NRT use. Addict Behav 2018;87:39-45 [FREE Full text] [doi: 10.1016/j.addbeh.2018.06.022] [Medline: 29940390]
27. Health Unlocked. URL: https://healthunlocked.com/ [accessed 2024-12-13]
28. NetworkX. URL: https://networkx.org/ [accessed 2024-12-13]
29. Naseem U, Razzak I, Khan SK, Prasad M. A comprehensive survey on word representation models: from classical to state-of-the-art word representation language models. ACM Trans Asian Low Resour Lang Inf Process 2021;20(5):1-35. [doi: 10.1145/3434237]
30. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv Preprint posted online on October 11, 2018 [FREE Full text] [doi: 10.48550/arXiv.1810.04805]

XSL•FO
RenderX

31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 04, 2017; Red Hook, NY.
32. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
33. Chakraborty AK, Das S, Kolya AK. Sentiment analysis of COVID-19 tweets using evolutionary classification-based LSTM model. In: Mukherjee A, Pan I, Piuri V, editors. Proceedings of Research and Applications in Artificial Intelligence: RAAI 2020. Singapore: Springer; 2021.
34. COVID19 Tweets. Kaggle. URL: https://www.kaggle.com/datasets/gpreda/covid19-tweets [accessed 2024-12-13]
35. Hugging Face. URL: https://huggingface.co/ [accessed 2024-12-13]
36. Kingma DP, Ba J. Adam: a method for stochastic optimization. ArXiv Preprint posted online on December 22, 2014. [doi: 10.48550/arXiv.1412.6980]
37. Sparse categorical crossentropy. TensorFlow. https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy URL: https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy [accessed 2024-12-13]
38. Sparse categorical accuracy. TensorFlow. URL: https://www.tensorflow.org/api_docs/python/tf/keras/metrics/SparseCategoricalAccuracy [accessed 2024-12-13]
39. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]

Abbreviations

AUK: Asthma UK
AVS: average sentiment score
BERT: bidirectional encoder representations from transformers
BioBERT: bio-bidirectional encoder representation from transformers
BLF: British Lung Foundation
LTC: long-term conditions
OHC: online health community
SA: sentiment analysis

Edited by A Mavragani; submitted 03.01.24; peer-reviewed by J Ray, J Chen; comments to author 17.09.24; revised version received 02.10.24; accepted 06.11.24; published 13.02.25.

Please cite as:
Li X, Vaghi E, Pasi G, Coulson NS, De Simoni A, Viviani M, AD HOC Group
Understanding the Engagement and Interaction of Superusers and Regular Users in UK Respiratory Online Health Communities: Deep Learning–Based Sentiment Analysis
J Med Internet Res 2025;27:e56038
URL: https://www.jmir.org/2025/1/e56038
doi:10.2196/56038
PMID:

©Xiancheng Li, Emanuela Vaghi, Gabriella Pasi, Neil S Coulson, Anna De Simoni, Marco Viviani, AD HOC Group. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Harnessing Social Media Data to Understand Information Needs About Kidney Diseases and Emotional Experiences With Disease Management: Topic and Sentiment Analysis

Hee Jeong Hwang[1], BS; Nara Kim[1], BS; Jeong Yun You[1], BS; Hye Ri Ryu[1], BS; Seo-Young Kim[1], BS; Jung Han Yoon Park[2], PhD; Ki Won Lee[1,2,3,4,5], PhD

[1]Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

[2]Advanced Institutes of Convergence Technology, Seoul National University, Suwon, Republic of Korea

[3]Bio-MAX Institute, Seoul National University, Seoul, Republic of Korea

[4]Institutes of Green Bio Science & Technology, Seoul National University, Pyeongchang, Republic of Korea

[5]Department of Agricultural Biotechnology and Center for Food and Bioconvergence, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Ki Won Lee, PhD
Department of Agricultural Biotechnology
College of Agriculture and Life Sciences
Seoul National University
8th Fl.
1, Gwanak-ro
Seoul, 08826
Republic of Korea
Phone: 82 01099867464
Fax: 82 880 4662
Email: kiwon@snu.ac.kr

Abstract

Background: Kidney diseases encompass a variety of conditions, including chronic kidney disease, acute kidney injury, glomerulonephritis, and polycystic kidney disease. These diseases significantly impact patients' quality of life and health care costs, often necessitating substantial lifestyle changes, especially regarding dietary management. However, patients frequently receive ambiguous or conflicting dietary advice from health care providers, leading them to seek information and support from online health communities.

Objective: This study aimed to analyze social media data to better understand the experiences, challenges, and concerns of patients with kidney disease and their caregivers in South Korea. Specifically, it explored how online communities assist in disease management and examined the sentiment surrounding dietary management.

Methods: Data were collected from *KidneyCafe*, a prominent South Korean online community for patients with kidney disease hosted on the Naver platform. A total of 124,211 posts from 10 disease-specific boards were analyzed using latent Dirichlet allocation for topic modeling and Bidirectional Encoder Representations From Transformers–based sentiment analysis. In addition, Efficiently Learning an Encoder That Classifies Token Replacements Accurately–based classification was used to further analyze posts related to disease management.

Results: The analysis identified 6 main topics within the community: *family health and support*, *medication and side effects*, *examination and diagnosis*, *disease management*, *surgery for dialysis*, and *costs and insurance*. Sentiment analysis revealed that posts related to the *medication and side effects* and *surgery for dialysis* topics predominantly expressed negative sentiments. Both significant negative sentiments concerning worries about kidney transplantation among family members and positive sentiments regarding physical improvements after transplantation were expressed in posts about family health and support. For *disease management*, 7 key subtopics were identified, with inquiries about dietary management being the leading subtopic.

Conclusions: The findings highlight the critical role of online communities in providing support and information for patients with kidney disease and their caregivers. The insights gained from this study can inform health care providers, policy makers,

XSL•FO
RenderX

and support organizations to better address the needs of patients with kidney disease, particularly in areas related to dietary management and emotional support.

(J Med Internet Res 2025;27:e64838) doi:10.2196/64838

KEYWORDS

kidney diseases; online health communities; topic modeling; sentiment analysis; disease management; patient support

Introduction

Background

Kidney diseases primarily manifest in 2 main forms: chronic kidney disease (CKD) and acute kidney injury (AKI), which are 2 significant kidney conditions [1,2]. CKD is characterized by a progressive decline in kidney function lasting >3 months, whereas AKI refers to a sudden decrease in kidney function that occurs over a period of hours to days [1,2]. Both conditions can arise from various underlying causes, including primary glomerular diseases (eg, glomerulonephritis), inherited disorders (eg, polycystic kidney disease), and secondary complications of systemic conditions such as diabetes and hypertension [2,3]. These conditions collectively affect a significant proportion of the global population, with CKD alone estimated to affect 13.4% of people worldwide [4]. Both CKD and AKI, regardless of their underlying etiology, increase the risk of other medical complications and significantly impact patients' quality of life and health care costs [2,5]. Managing these conditions can require substantial lifestyle changes, particularly in dietary habits, which can be challenging for patients [6]. However, maintaining a healthy lifestyle can reduce the risk of kidney disease progression and lower mortality rates [7].

Despite the critical role of lifestyle modifications in managing kidney diseases, patients often report receiving ambiguous or conflicting information from health care providers, leading to confusion and a sense of disempowerment [8]. Time constraints in clinical settings further limit the extent of guidance that health care providers can offer, leading many patients to turn to online health communities for information and support.

The use of web-based platforms for health-related information has become increasingly prevalent, with approximately 80% of patients using the internet to exchange information and over one-third of adults using social media for health information and social support [9,10]. Unlike traditional research methods such as surveys or interviews, social media analysis provides access to authentic, unprompted patient discussions, offering unique insights into their daily challenges and evolving needs [9,10]. This approach has been widely used to understand patient experiences within chronic disease communities. For example, studies have used social media analysis to capture patient narratives about cancer [11] and examine the social media behaviors of patients and health care providers across various conditions, including asthma, mental health disorders, diabetes, and gout [12,13].

While topic modeling has been widely used to extract health-related topics from social media, studies specifically examining public discourse on kidney diseases are scarce, particularly in South Korea. Despite the country's high internet penetration and active online health communities, research that systematically analyzes patient experiences with kidney disease through social media remains limited. Successful applications of social media analysis in South Korea include studies on diabetes-related communication on social networking services [14], evaluations of physician-created thyroid cancer content on YouTube [15], the use of topic modeling for pancreatic cancer information [16], and investigations into cross-cultural perceptions of urological conditions [17]. However, comprehensive studies of kidney disease communities are notably absent. These communities often reflect unique cultural challenges and interactions with the health care system that are specific to South Korean patients.

Recent advances in natural language processing (NLP) have transformed how researchers analyze and interpret unstructured health-related text data. Deep learning approaches in NLP have shown significant success in processing various types of health care text, including clinical narratives, electronic health records, and social media content [18,19]. These computational methods are particularly effective in uncovering patterns and insights that are not readily apparent through traditional analytical methods. Applications of NLP in health care range from extracting medical entities and relationships from clinical texts to analyzing patient sentiments, behaviors, and experiences shared in online health communities. Such approaches are instrumental in bridging the gap between raw data and actionable insights in health care research.

While topic modeling allows for quantitative analysis of discussed topics, it does not fully capture the emotional and experiential dimensions of patients' lives. To address this limitation, our study combined topic modeling with sentiment analysis, using advanced NLP techniques to provide both quantitative insights and a qualitative understanding of patients' emotional challenges. This dual approach bridges the gap between unstructured patient narratives and actionable health care insights. By focusing on the experiences of South Korean patients with kidney disease, our findings can inform the development of culturally appropriate, patient-centered interventions, enabling health care providers to address patient concerns more effectively, design impactful educational materials, and enhance support systems tailored to these patients' unique needs.

Objectives

The main objectives of this study were to (1) analyze social media data using topic modeling to better understand the experiences, challenges, and concerns of patients living with various kidney diseases and their caregivers in South Korea; (2) explore the ways in which online communities can provide assistance, specifically in relation to disease management for

XSL•FO
RenderX

different kidney conditions; and (3) examine the sentiment surrounding disease management among individuals with kidney diseases.

By achieving these objectives, this study aimed to provide valuable insights that can inform health care providers, policy makers, and support organizations in better addressing the needs of patients with a spectrum of kidney diseases and renal conditions.

Methods

Overview

The goals of this study were to identify the challenges faced by patients and determine the key topics related to kidney disease management. We used topic modeling for quantitative analysis and sentiment analysis for qualitative analysis to achieve this. Relevant posts were gathered from online communities discussing kidney disease, preprocessing was conducted, and the topics and sentiments were analyzed. Figure 1 provides an overview of the research methodology.

Figure 1. Overall procedures used in the study. The workflow included data crawling from the Naver KidneyCafe, data preprocessing, latent Dirichlet allocation (LDA) topic modeling, sentiment analysis of the total posts, and manual and Efficiently Learning an Encoder That Classifies Token Replacements Accurately (ELECTRA)–based classification related to the disease management topic. Q1: the safety of specific foods; Q2: dietary management strategies; Q3: dietary recommendations; Q4: water intake management; Q5: weight management; Q6: exercise management.

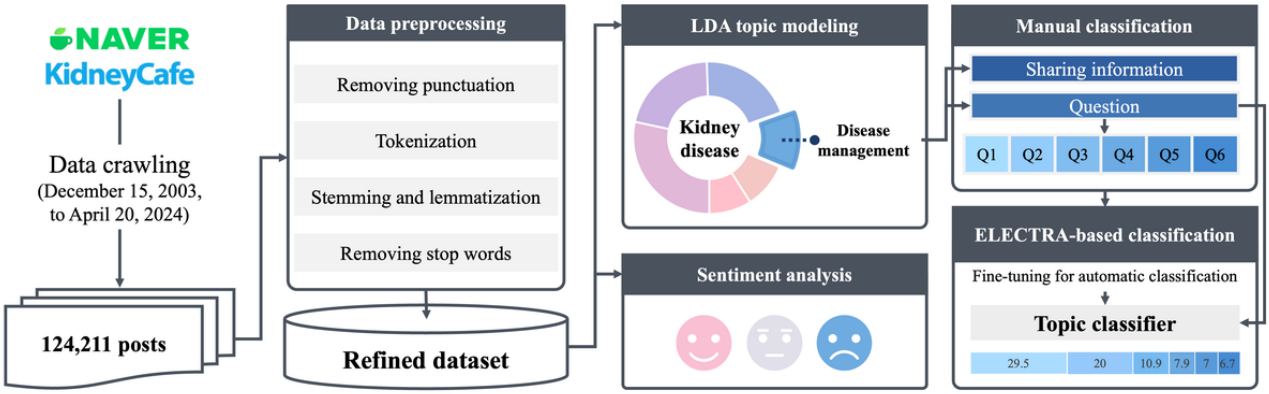

Data Collection and Preprocessing

We collected data from *KidneyCafe*, a prominent South Korean online community for patients with kidney disease and caregivers hosted on the Naver platform [20]. This community has approximately 183,545 members and facilitates discussions across various themed boards, each concentrating on specific kidney-related topics. For our analysis, we focused on the disease-specific boards within the *Member Participation* section, which is dedicated to active discussions among engaged community members. These boards address topics such as kidney transplantation, dialysis, and related medical conditions, enabling us to capture in-depth conversations pertinent to a wide array of kidney disease experiences and management strategies.

The Selenium WebDriver was used to extract 124,211 posts from 10 disease-specific boards, dating from the community's inception on December 15, 2003, to the date of data collection (April 20, 2024). The collected data included post titles and content.

Data preprocessing was conducted using Python (Python Software Foundation), with the *KoNLPy* library used for Korean NLP [21]. This step involved removing punctuation and noninformative words as well as extracting nouns from the text to prepare for subsequent analysis. To handle informal language patterns in user-generated content, we used a custom-built stop word list to exclude frequently appearing, noninformative terms that do not contribute to semantic analysis. This stop word list was created by reviewing existing lists from similar Korean health-related studies [22] and further customizing it to focus on discussions surrounding kidney disease. These preprocessing steps were essential in ensuring the robustness of our analysis and the interpretability of the results.

Topic Analysis Using Latent Dirichlet Allocation

We used latent Dirichlet allocation (LDA) topic modeling [23], an unsupervised machine learning model, to uncover latent topics in the collected posts. Topic modeling is a technique widely used in NLP for discovering topics and extracting meaning from large, unstructured documents. LDA is the most popular topic modeling method and is based on statistical probability inference. It assumes that each document can be represented by a probabilistic distribution over latent topics, and each topic is characterized by a probabilistic distribution over the words [24].

We imported LDA models from the *Gensim* library for this study. When analyzing topics using LDA, the number of topics is typically determined using coherence scores such as U_mass and perplexity [25,26]. The U_mass coherence score, which is based on word co-occurrence statistics within documents, has been validated to closely correspond with human labeling results [27,28]. This score has been effectively used in analyzing health-related social media topics, including obesity, COVID-19, and vaccines [27-30]. Similarly, perplexity has proven to be a valuable metric in recent topic modeling studies [31,32]. While a higher coherence score typically indicates better topic quality and interpretability, recent research shows that statistical optimization alone may not guarantee optimal interpretability of topics [32-34]. To tackle this issue, several studies have

XSL•FO
RenderX

adopted validation methods that integrate statistical metrics with qualitative assessments. For instance, Min et al [31] combined perplexity and coherence metrics with qualitative evaluations in their analysis of occupational accidents, whereas Lyu et al [33] and Li et al [32] emphasized the significance of semantic coherence alongside statistical validation.

Following established approaches [32-34], we first calculated perplexity and U_mass coherence scores across a range of topics (3-40) using 10-fold cross-validation. As shown in Figure 2, both metrics demonstrated better performance with fewer topics, leading us to concentrate on the range of 3 to 10 topics. Within this range, we conducted 10 iterations of the model to ensure consistency and, ultimately, selected 6 topics. This decision was based on a combination of qualitative assessments of topic interpretability and the high U_mass coherence scores observed at both 3 and 6 topics (Figure 3).

Figure 2. Perplexity and U_mass coherence scores across different numbers of topics. This figure shows the perplexity and U_mass coherence scores as the number of topics varies from 3 to 40 using 10-fold cross-validation. Both metrics indicate better performance with a lower number of topics.

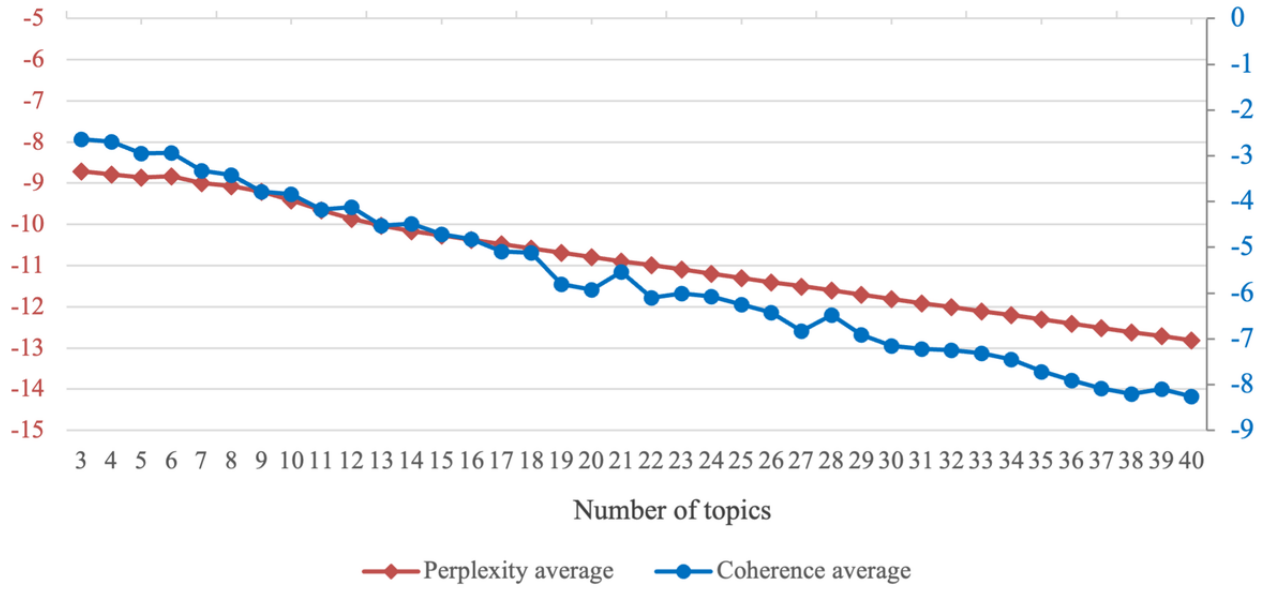

Figure 3. U_mass coherence scores for 3 to 10 topic clusters across 10 trials. This figure shows the U_mass coherence scores as the number of topic clusters increases from 3 to 10 across 10 separate trials. The highest coherence score is observed when there are 3 and 6 topic clusters.

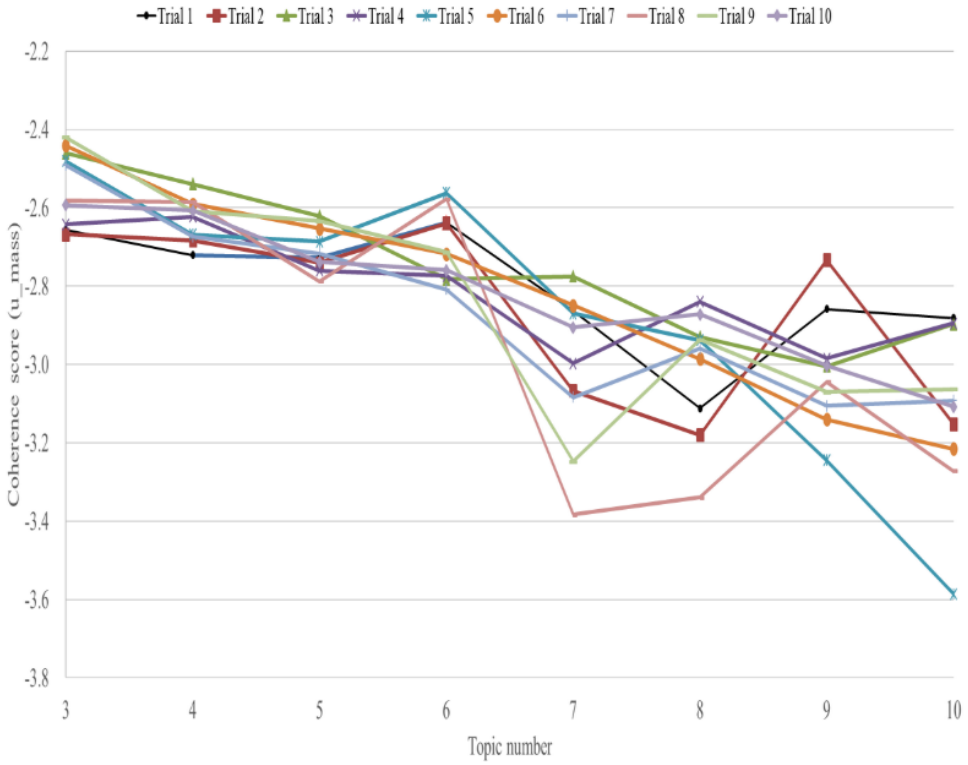

XSL•FO
RenderX

Using the optimal cluster size, we ran the LDA model to obtain lists of representative words for each cluster. The topic-labeling process involved 4 researchers, each with diverse expertise in nutrition and food science, who collaboratively reviewed and interpreted the word lists. We followed established methodologies in social media analysis [33,35,36] using an iterative, discussion-based approach. Initially, each researcher independently analyzed the word lists generated by the topic modeling algorithm, focusing on the most significant terms and their relative weights within each topic to identify potential themes. Next, they reviewed representative posts selected for their strong association with each topic to contextualize the keywords and understand their relevance to the experiences and concerns of patients with kidney disease. This independent review enabled researchers to form initial interpretations and propose tentative labels for further discussion.

After the independent review, the researchers gathered to compare their initial interpretations and collaboratively refine the proposed labels through discussion. When differing interpretations emerged, the team reviewed additional example posts from the topic clusters to gain more context. This process included examining word clouds and representative posts that distinctly exemplified each topic while also considering the specific experiences and information-seeking behaviors of patients with kidney disease. Final topic labels were assigned once all 4 researchers reached a consensus through comprehensive discussion and contextual analysis.

The contribution of each word to a cluster was evaluated using the probability metric derived from the model results. Representative example sentences for each cluster were identified based on the probability that each post belonged to the topic.

Sentiment Analysis

The Bidirectional Encoder Representations From Transformers (BERT) method, a machine learning technique for NLP, was used to identify the sentiment of posts about dietary management—positive, negative, and neutral—to uncover which aspects of dietary management were the most challenging. While word embeddings can be derived from large and unannotated corpora using word co-occurrence statistics, they do not consider the context when creating word vectors [37]. However, BERT is context dependent; it produces dynamic and context-specific representations of the word by accounting for neighboring words, which helps interpret the semantics along with the grammatical structure [38].

The sentiment analysis in this study used the Naver CLOVA Sentiment platform, which uses a pretrained BERT model specifically adapted for analyzing Korean-language texts. This adaptation enables the model to effectively capture context-specific nuances, particularly in the informal and colloquial language common in social media discussions. While the specific parameters and fine-tuning details of the application programming interfaces are proprietary to Naver and cannot be disclosed, the effectiveness of this methodology for Korean sentiment analysis has been validated through several academic studies [39-41]. The platform provides sentiment classification results (ie, which sentiment—positive, negative, or neutral—is best represented in the texts), along with the expected probability of each sentiment.

Analysis of the Disease Management Topic Using Efficiently Learning an Encoder That Classifies Token Replacements Accurately

The *disease management* topic was investigated by conducting a detailed analysis. One author manually classified the posts categorized under *disease management* by the LDA results into questions and nonquestions. In total, 3 additional authors independently annotated a subset of 300 posts to establish the annotation rules. The posts were reviewed for seven key sentiments: (1) general dietary management methods, (2) food recommendations, (3) inquiries about the safety of consuming specific foods, (4) requests for dietary evaluations, (5) water management, (6) exercise management, and (7) weight management. After discussion, the annotators agreed on a finalized set of annotation rules and guidelines.

Next, 2 independent annotators labeled a random set of 1000 posts from the overall dataset. The Cohen κ coefficient for interrater reliability was 0.77, indicating substantial agreement. These annotated data were used for transfer learning. We fine-tuned the Efficiently Learning an Encoder That Classifies Token Replacements Accurately (ELECTRA) model, a pretrained language model that has demonstrated state-of-the-art performance across various Korean-language tasks [42]. Compared to BERT and the A Robustly Optimized BERT Pretraining Approach model, ELECTRA offers both computational efficiency and improved performance on downstream classification tasks [43], making it particularly suitable for handling large-scale datasets such as ours. In addition, ELECTRA's generator-discriminator training mechanism has been shown to outperform masked language models in tasks that require nuanced contextual understanding, as validated in previous Korean NLP studies [42]. We used 1000 manually annotated posts split into training and test sets at a ratio of 80:20. The evaluation metrics on the test set yielded an F_1-score of 0.85, indicating high performance. This fine-tuned ELECTRA model was then used to annotate the question posts related to the topic of *disease management*.

Ethical Considerations

This study used preexisting publicly accessible data with no identifiable information and therefore does not require institutional review board review per Federal Regulations for the Protection of Human Research Subjects (45CFR 46.104(d)(4)) or patient consent [44]. In compliance with these guidelines and established research practices, all identifying information, such as usernames and specific identifiers, was anonymized or removed before analysis. This approach is consistent with methodologies used in similar international studies using publicly accessible data [33,45,46], as well as previous research analyzing public Naver Cafe posts [47-49].

XSL•FO
RenderX

Results

Community Overview

We gathered a comprehensive dataset that included 124,211 posts from various discussion boards in the KidneyCafe community to better understand the experiences, challenges, and concerns of patients living with kidney disease and their caregivers in South Korea. These posts were written by 34,472 unique users, indicating a highly engaged community. Of these users, 34.92% (12,039/34,472) wrote multiple posts, with an average of 3.6 (SD 7.3) posts per user. This high level of repeated engagement suggests that many users found ongoing value and support within the community. Figure 4 illustrates the monthly distribution of posts, showing a significant increase during the beginning of the COVID-19 pandemic in South Korea.

Figure 4. Timeline of the number of posts per month. The red line indicates the first confirmed COVID-19 case in South Korea, whereas the green line marks the World Health Organization's declaration of the end of the COVID-19 emergency. Monthly post counts were aggregated.

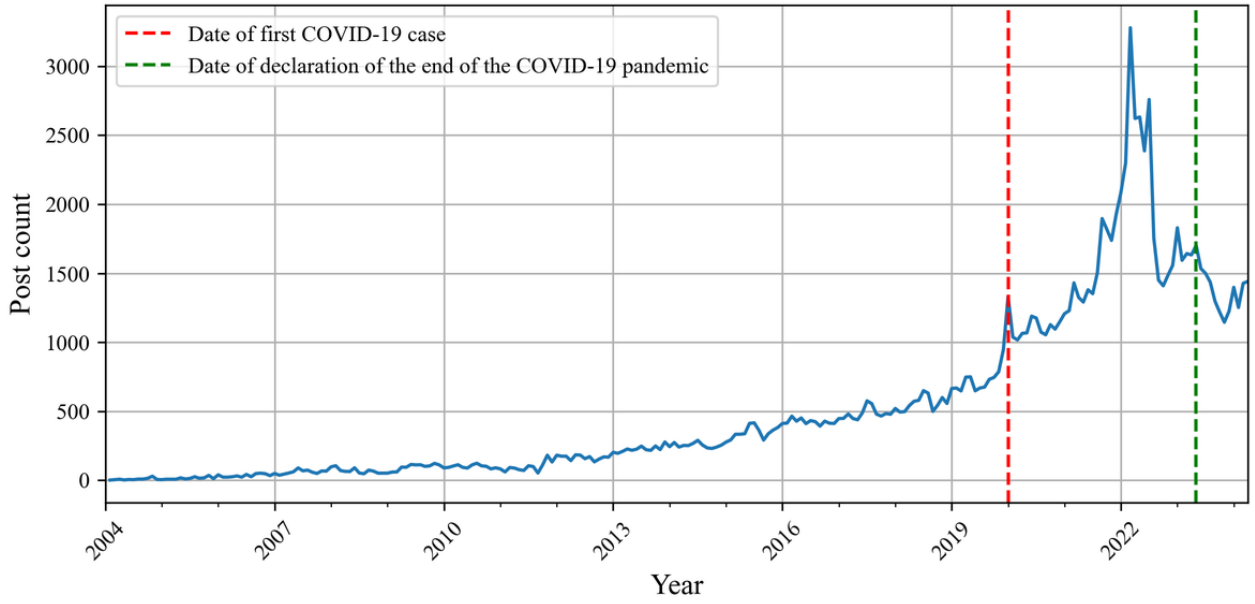

One of the notable advantages of this community is the categorization of discussion boards according to specific diseases. An examination of the number of posts collected per board (Table 1) showed that the highest number of posts was related to kidney transplantation (39,221/124,211, 31.58%), followed by kidney dialysis (27,692/124,211, 22.29%).

Table 1. Number of posts per board (N=124,211).

Board[a]	Crawled posts, n (%)
Hypertension & Diabetes	1422 (1.14)
Pediatric Kidney Disease	3271 (2.63)
Others	3969 (3.2)
Polycystic & Cysts & Lumps	4011 (3.23)
Kidney Cancer	6097 (4.91)
Chronic Kidney Failure	8216 (6.61)
Proteinuria & Hematuria	10,113 (8.14)
Nephrotic Syndrome & Glomerulonephritis	20,199 (16.26)
Hemodialysis & Peritoneal Dialysis	27,692 (22.29)
Kidney Transplant	39,221 (31.58)

[a]KidneyCafe organizes posts into boards according to specific health conditions.

Main Topics Within the KidneyCafe Community

Overview

We used LDA for topic modeling to identify the common topics discussed by patients with kidney disease in KidneyCafe. Our analysis revealed 6 distinct topic clusters that represented important discussions within the community. Each cluster was labeled based on the main keywords, as shown in Table 2. The distribution of discussion topics across the community was as follows: 28.36% (35,226/124,211) *family health and support*,

XSL•FO
RenderX

21.11% (26,221/124,211) *medication and side effects*, 20.12% (24,991/124,211) *examination and diagnosis*, 11.67% (14,495/124,211) *disease management*, 9.71% (12,061/124,211) *surgery for dialysis*, and 9.03% (11,216/124,211) *costs and insurance*.

Table 2. Keywords representative of the 6 discussion clusters in KidneyCafe.

Topic label	Keywords[a]
Family health and support	"Mother," "father," "discharge," "mind," "preparation," "husband," "start," "family," "groom," "tomorrow," "sibling," "call," "nurse," "life," and "body"
Medication and side effects	"Intake," "immunosuppressant," "symptoms," "side effects," "steroids," "prescription," "cold," "treatment," "morning," "rejection reaction," "dose," "injection," "face," "yesterday," and "head"
Examination and diagnosis	"Blood pressure," "tissue examination," "normal," "creatinine," "blood test," "management," "creatinine levels," "outpatient," "professor," "maintenance," "hematuria," "figures," "current," "diagnosis," and "function"
Disease management	"Exercise," "water," "weight," "food," "rice," "meal," "blood sugar," "potassium," "daily," "diet," "eat," "urine volume," "control," "diabetes," and "snack"
Surgery for dialysis	"Blood vessels," "peritoneal dialysis," "removal," "pain," "recommendation," "procedure," "surgical site," "hemodialysis," "anesthesia," "surgery," "peritoneum," "machine," "arteriovenous fistula," "Seoul," and "endoscopy"
Costs and insurance	"Application," "price," "registration," "antibodies," "usage," "insurance," "error," "Korea," "disease," "occurrence," "brain death transplant," "cross-reaction," "benefits," "treatment," and "donation"

[a]The keywords represent the top 15 terms for each topic ranked by their relevance score from the latent Dirichlet allocation results, making them the most representative words for each topic.

Topic 1: Family Health and Support

Topic 1, which accounted for 28.36% (35,226/124,211) of the discussions, mainly focused on family members with kidney disease and concerns regarding their health. The key terms included words related to familial relationships, such as "mother," "father," "husband," "family," "groom," and "sibling." These terms appeared frequently in discussions about kidney transplantation within the family context.

The following is an example of those posts:

> *It seems like my father will donate a kidney to my mother, but I'm worried.*

In addition, concerns and emotions surrounding posttransplant discharge and care were prevalent, with terms such as "discharge," "mind," and "preparation" appearing frequently in the posts, as shown as follows:

> *I work full-time, so how should I take care of my husband after he is discharged?*
>
> *My sibling offered to donate a kidney, but my mind is very uneasy about it.*

Topic 2: Medication and Side Effects

Topic 2 comprised 21.11% (26,221/124,211) of the discussions and included posts about the correct use, precautions, and dosages of medications. Key terms such as "intake," "immunosuppressant," "prescription," "treatment," "morning," and "dose" were frequently mentioned, highlighting discussions on the proper way to take specific prescription medications. The following are examples of those posts:

> *I was prescribed Solondo. Can Solondo and immunosuppressants be taken together?*
>
> *Should immunosuppressants be taken on an empty stomach in the morning?*

The community also engaged in discussions on various side effects and their causes, with terms such as "symptoms," "side effects," "steroids," "injections," and "rejection reaction" being commonly mentioned. The following is an example of such posts:

> *I was prescribed steroid injections and am experiencing side effects.*

Topic 3: Examination and Diagnosis

Topic 3, accounting for 20.12% (24,991/124,211) of the discussions, focused on kidney disease tests and diagnostic content. Terms related to specific test items, such as "blood pressure," "creatinine," and "hematuria," as well as test methods such as "tissue examination" and "blood test," were particularly frequent.

The following are examples of posts using those keywords:

> *My blood pressure was too low during a blood test.*
>
> *Which is more important, creatinine levels or the glomerular filtration rate?*

In addition, discussions often included receiving a diagnosis from a physician based on test results, with terms such as "normal," "management," "outpatient," "doctor," "maintenance," and "current" appearing frequently. Examples of those posts are shown as follows:

> *I came for an outpatient consultation to review my test results.*
>
> *The doctor said that, as of now, my test results showed normal creatinine levels and glomerular filtration rate, but ongoing management will be necessary for maintenance.*

Topic 4: Disease Management

Topic 4, accounting for 11.67% (14,495/124,211) of the discussions, focused on the management of kidney diseases,

XSL•FO
RenderX

including exercise, water intake, and dietary management. Major keywords such as "exercise" and "weight" were prevalent in exercise management discussions such as the following:

> *What exercises should I do to lose weight?*

Keywords such as "water" and "urine volume" were commonly observed in posts related to water intake management, such as "How much *water* do dialysis patients typically consume?"

Keywords such as "food," "snack," "meal," "diet," "eat," "potassium," and "control" were identified in the context of dietary management. Examples of those posts are shown as follows:

> *What are some high-calorie foods or snacks that don't strain the kidneys?*
>
> *Is it okay to eat snacks after dialysis?*
>
> *I'm following a low-protein, low-potassium diet. What should I be cautious of when planning my meals?*

Topic 5: Surgery for Dialysis

Topic 5, which accounted for 9.71% (12,061/124,211) of the posts, focused on dialysis and predialysis procedures. Major keywords such as "hemodialysis," "peritoneal dialysis," "recommendation," and "pain" were prevalent, with many posters seeking advice on the reasons for selecting and transitioning between the 2 types of dialysis. Examples of those posts are shown as follows:

> *Give me a recommendation on whether I should choose hemodialysis or peritoneal dialysis.*
>
> *While undergoing hemodialysis, I experienced pain as the pressure increased.*

In addition, keywords such as "blood vessels," "procedure," "surgical site," "anesthesia," "surgery," "arteriovenous fistula," and "Seoul" frequently appeared in posts related to predialysis surgery, including arteriovenous fistula creation or placement surgery and peritoneal dialysis catheter insertion. The following are some examples:

> *Can you recommend a good hospital in Seoul for arteriovenous fistula surgery?*
>
> *After an arteriovenous fistula surgery, my blood vessel feels enlarged, and the surgical site is swollen. Is it normal for the vessel to enlarge after the procedure?*

Topic 6: Costs and Insurance

In topic 6, accounting for 9.03% (11,216/124,211) of the posts, the users expressed significant concerns regarding financial and insurance-related issues with respect to managing their health conditions. The major keywords were "application," "price," "registration," "insurance," and "benefits." They appeared in sentences such as the following:

> *What documents do I need to prepare for an insurance application and reimbursement for costs?*
>
> *How much compensation do patients undergoing dialysis receive?*
>
> *I am a kidney donor and have registered for insurance. Can I claim insurance benefits?*

Distribution of Topics Across the Boards

We analyzed the distribution of topics across different boards to gain insights into the specific focus areas in each board (Figure 5). Our analysis revealed that the Kidney Transplant board had the highest concentration of posts related to the *family health and support* topic and also showed a significant focus on the *medication and side effects* topic. Of note was that the Hemodialysis & Peritoneal Dialysis board had a higher proportion of transplant-related posts than of dialysis-specific ones. The Nephrotic Syndrome & Glomerulonephritis and Proteinuria & Hematuria boards showed a high concentration of posts related to the *examination and diagnosis* topic. The Pediatric Kidney Disease board notably focused on the *medication and side effects* topic, whereas the *disease management* topic was particularly prominent in the Hypertension & Diabetes board.

XSL•FO
RenderX

Figure 5. Topic distribution per board. The intensity of the colors represents the frequency of a particular topic on a board, with darker shades indicating a higher concentration of posts related to that topic. The topics are ordered from left to right in descending order of overall prevalence across all boards. The boards are arranged from bottom to top in descending order of total post count.

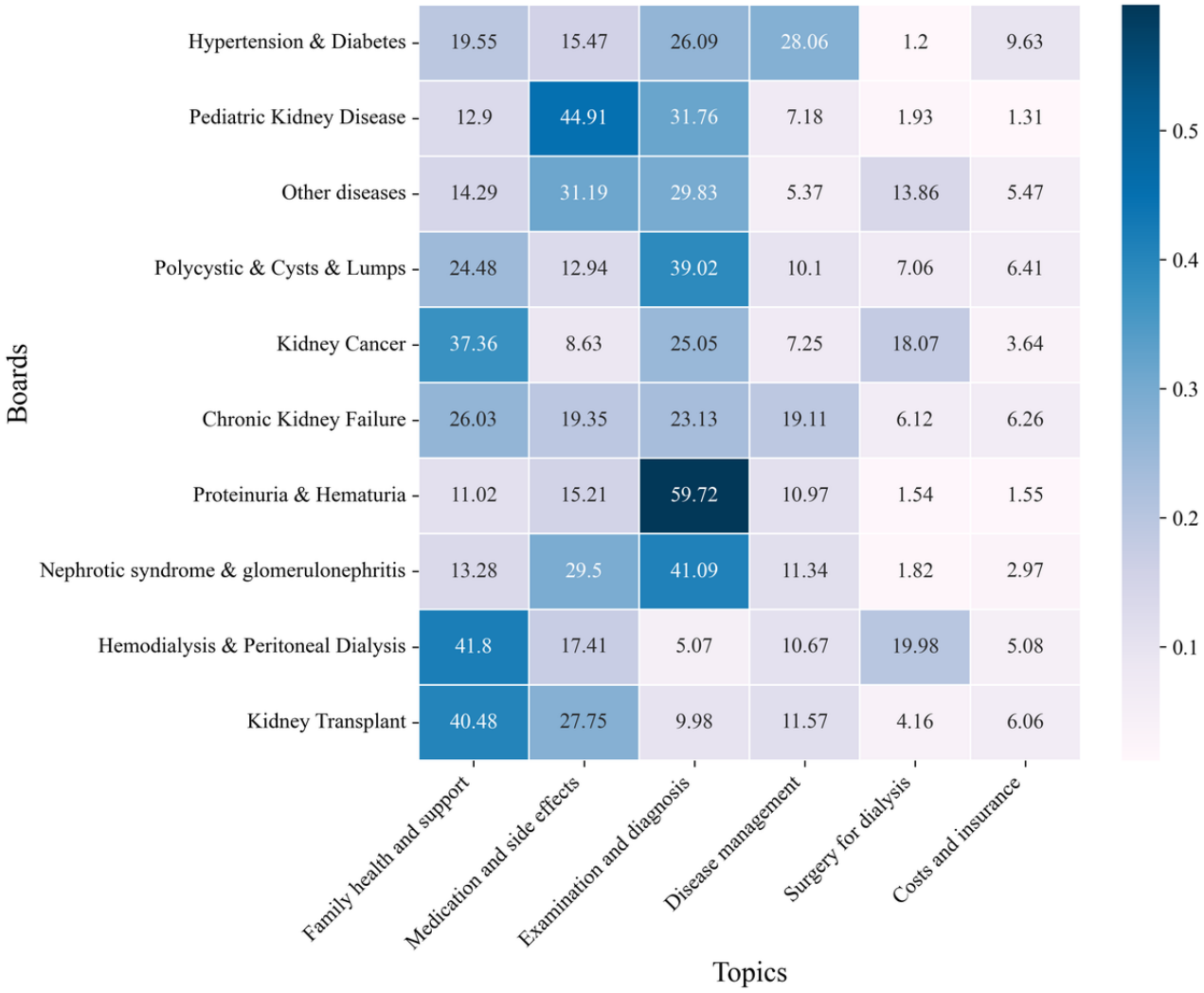

Sentiment Analysis of the Kidney Disease Topics

Posts on each topic cluster were analyzed to determine their sentiment, which was then categorized as positive, neutral, or negative. The results indicated that posts on the *medication and side effects* topic contained the highest proportion of negative sentiments (Figure 6). This cluster often included discussions about adverse reactions to prescribed medications, concerns about the impact of other drugs on kidney health, and general anxiety about the side effects of medications used to manage kidney disease. The following are a few posts from the community expressing these sentiments:

> *Since starting steroids, my eyes feel very dry, and my legs feel heavy. Is this normal?*
>
> *As a polycystic kidney patient, is it safe to be prescribed medication for severe itching from a dermatologist?*

While many posts described challenges, some shared positive experiences, such as temporary relief from side effects:

> *It was incredibly itchy and kept spreading, so I tried applying various remedies. However, the neem oil I bought from Homeplus yesterday had an immediate effect on the itching.*

Similarly, the *surgery for dialysis* and *examination and diagnosis* topics mainly comprised posts with negative sentiments. These posts often discussed difficulties with dialysis, dissatisfaction with diagnostic outcomes related to kidney disease, and other related issues. The following are some examples from each topic:

> *I'm currently in my third year of peritoneal dialysis. I'm not anemic, but I feel extremely dizzy. Does anyone have any advice?*
>
> *Last year, I was diagnosed with stage 4 polycystic kidney disease, which was tough. Now, finding out that my son has also inherited it makes it even harder.*

XSL•FO
RenderX

Figure 6. Sentiment analysis by topic—the distribution of positive, neutral, and negative sentiments across the 6 topics, demonstrating the varied emotional responses within the community discussions.

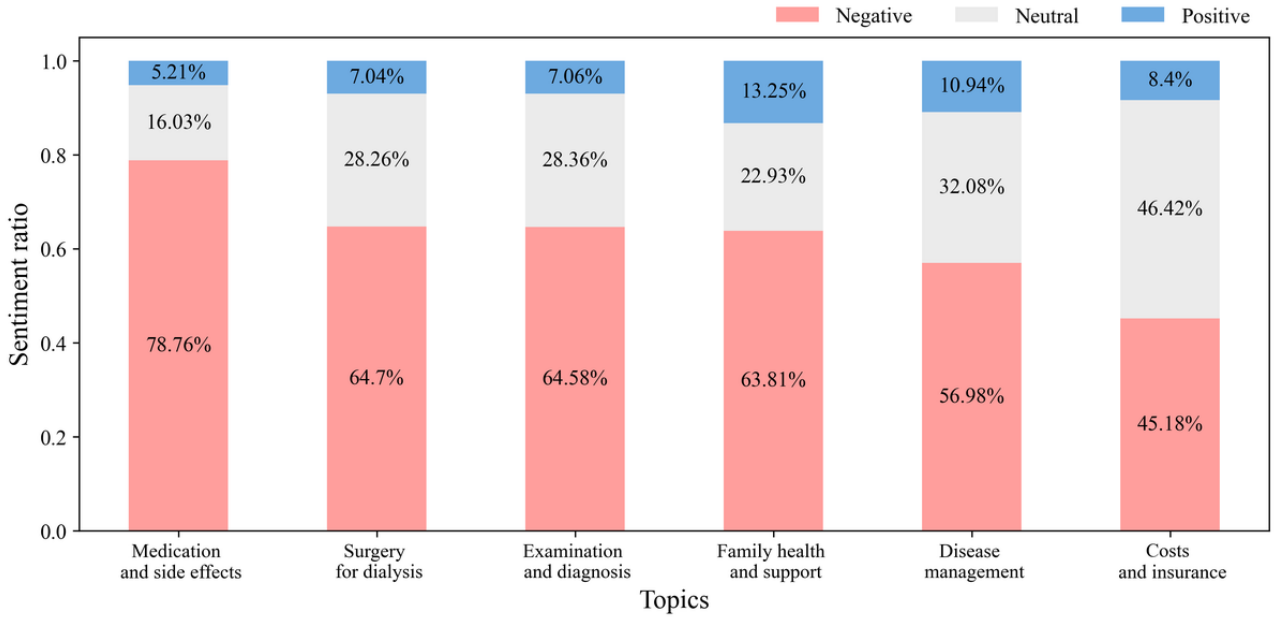

Positive sentiments were observed when participants shared their favorable test results:

> *Last December's blood test was normal, and today's urine test also came back clean. I'm relieved but daunted by the thought of restarting my diet management.*

The *family health and support* topic also comprised posts with many negative sentiments, primarily expressing worries regarding kidney transplantation among family members:

> *My sister is insisting on donating immediately, but I'm too worried about her health.*

However, alongside these concerns, high proportions of posts about physical improvements after transplantation were also notable among the positive sentiments:

> *It's been three years since my simultaneous pancreas-kidney transplant. I just tried a urine test kit, and the results are good, which makes me happy.*
>
> *I'm so happy to receive such warm hearts from everyone who left comments. I sincerely thank both the patient-commenters and the patient-readers.*

While the *disease management* topic also had significant negative sentiments, they were relatively fewer compared to the negative sentiments in other topics. Negative posters often expressed frustration with managing the disease or feeling overwhelmed by uncertainties in disease management:

> *Managing my diet is the hardest part.*
>
> *I really want to drink lots of water, but it's sad that I can't.*

In contrast, many posts inquiring about disease management strategies were classified as expressing neutral sentiments:

> *Is it okay to eat wild mango jelly?*
>
> *Does curry have a lot of potassium?*
>
> *For lunch, I had boiled squid and shiitake mushrooms fried with egg whites, enhanced by the flavorful addition of perilla oil.*

In the *costs and insurance* topic, negative and neutral sentiments were expressed in similar proportions. Negative posts often focused on concerns regarding delays or cost issues in insurance paperwork, such as the following:

> *I applied for reimbursement of my cancer diagnosis fee, but the review process is taking too long.*
>
> *I currently have workplace group supplemental insurance, but I'm worried about what will happen after retirement.*

Satisfaction with the health care system was occasionally noted:

> *Korea's health insurance is exceptional, covering both residential treatment and transplant surgery costs.*

In contrast, neutral posts mainly shared information regarding the handling of costs or insurance, such as the following:

> *Today, when I visited the local district office, they provided me with an informational leaflet regarding the required documents for registering as a person with kidney disease.*
>
> *If my grade 2 renal impairment progresses to grade 5, what changes will occur in my benefits?*

In-Depth Analysis of the Disease Management Topic

Given the complexity and importance of self-management for patients with kidney disease, we conducted an extensive analysis of the *disease management* topic. We carefully sorted the posts in this category into 2 main groups: information sharing and questions, as shown in Figure 7. We used an ELECTRA-based model to further categorize the questions into 6 specific subtopics. This process allowed us to qualitatively examine which issues were being actively discussed. We also analyzed

XSL•FO
RenderX

the sentiments associated with each specific subtopic, which was important for determining which subtopics were causing more difficulties for the patients (Figure 8).

Figure 7. Subtopics in the disease management topic.

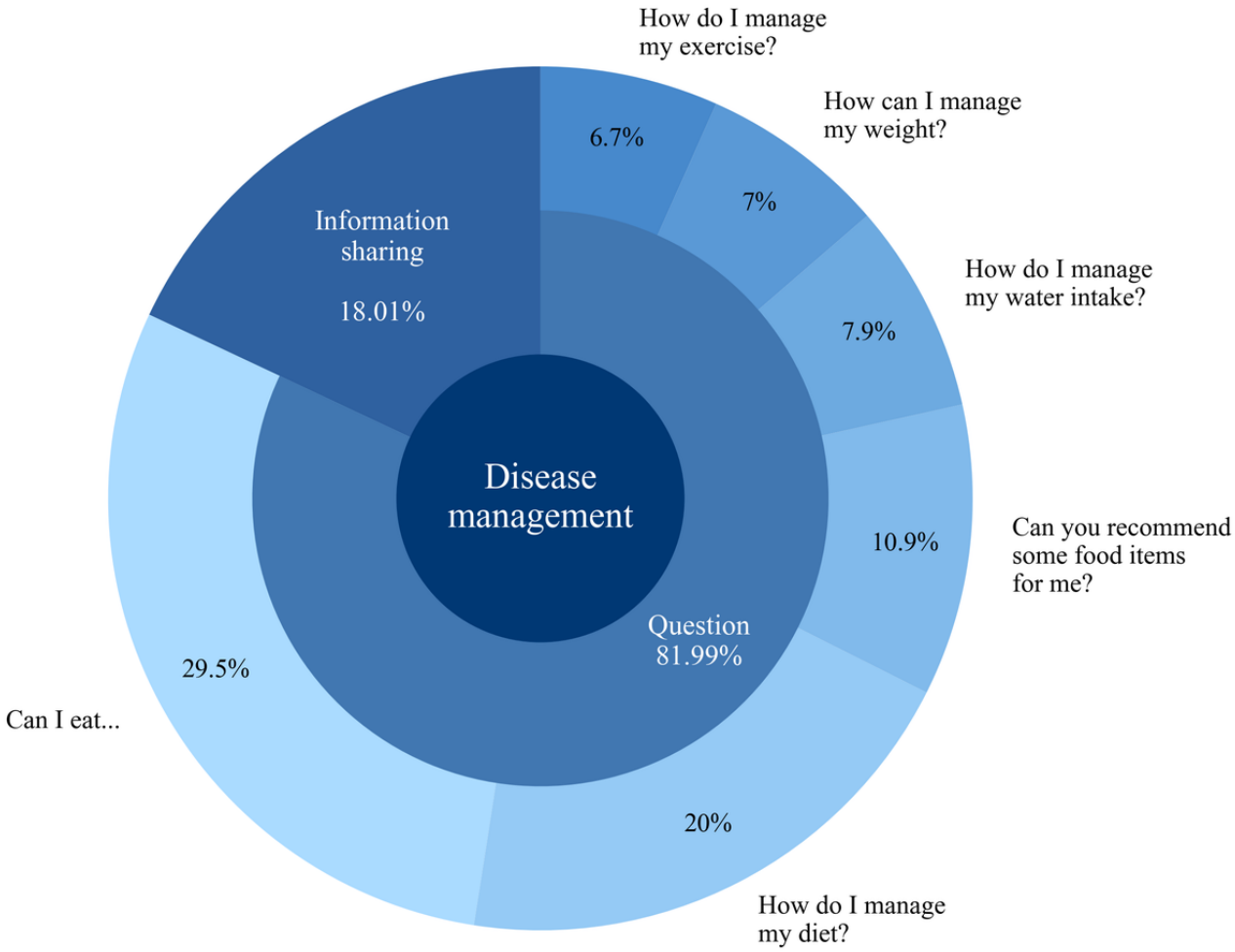

The number of posts under the *questions* category was >4 times higher than that of posts under *information sharing* (Figures 7 and 8). The *information sharing* category included valuable resources such as *Recommended information for post-transplant patients* and *Nephrotic syndrome management methods extracted from the Korean Dietetic Association*. Most posts in the *questions* category were related to dietary management (Figure 7), and these subtopics also comprised posts with more negative sentiments (Figure 8). Specifically, 29.5% (4276/14,495) of the inquiries were about the safety of specific foods, 20% (2899/14,495) sought guidance on dietary management strategies, and 10.9% (1580/14,495) requested dietary recommendations. Examples of inquiries about the safety of specific foods included the following:

Can I eat ginger if I have diabetes or kidney disease?

Are low-sodium salt and low-sodium soy sauce sold in stores okay?

Examples of inquiries seeking guidance on dietary management strategies included the following:

How should a single working person manage their weekday diet?

How should I manage my diet for proteinuria?

Examples of inquiries requesting dietary recommendations included the following:

Can you recommend low-potassium foods?

Other popular subtopics included water intake management (1145/14,495, 9%), weight management (1015/14,495, 7%), and exercise management (971/14,495, 6.7%). Examples of these questions included the following:

Is it okay to do high-intensity exercises like CrossFit if I have polycystic kidney disease?

What should I do if my weight keeps dropping?

How many liters of water do you drink per day?

XSL•FO
RenderX

Figure 8. Sentiment analysis by disease management subtopic—the distribution of positive, neutral, and negative sentiments across the 6 subtopics, demonstrating the varied emotional responses within the disease management topic.

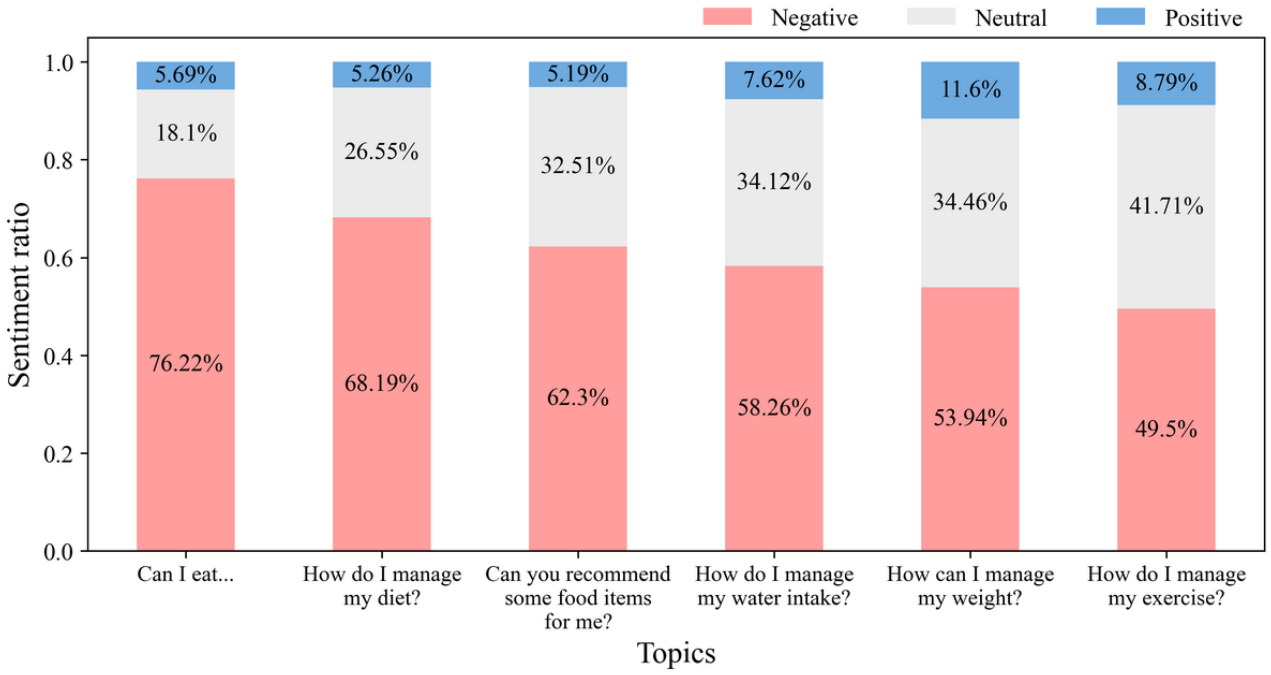

Discussion

Principal Findings

Our study findings provide valuable insights into the concerns and challenges faced by patients with different kidney diseases and conditions. The analysis of online community discussions revealed several key points with important implications for patient care, education, and support.

Impact of the COVID-19 Pandemic

The analysis revealed a significant increase in community activity during the COVID-19 pandemic. This observed increase underscores the heightened health concerns and information needs of patients with CKD during this period. This finding aligns with research highlighting the increased vulnerability of patients with CKD to COVID-19 complications [50-52] and emphasizes the importance of online communities as sources of support and information during health crises.

Topic Distribution and Varied Information Needs Across Disease Stages

Our analysis identified six main discussion topics in the kidney disease community: (1) *family health and support*, focusing on concerns surrounding kidney donation among family members and posttransplantation management; (2) *medication and side effects*, where conversations centered on medication administration methods and the associated side effects; (3) *tests and diagnostics*, where results of blood tests and tissue examinations were predominantly discussed; (4) *disease management*, which involved advice and shared experiences related to exercise, diet, and fluid intake; (5) *surgery for dialysis*, which addressed dialysis methods and surgical procedures; and (6) *costs and insurance*, covering financial burdens related to managing kidney disease and insurance issues.

Our findings demonstrate distinct differences in patient concerns and priorities across various stages of kidney disease. Early-stage patients, as evidenced by discussions on the Nephrotic Syndrome & Glomerulonephritis and Proteinuria & Hematuria boards, primarily focused on examinations and diagnoses. These discussions often included inquiries about test results, such as creatinine levels and glomerular filtration rates, as well as clarifications on diagnostic procedures. This reflects a strong emphasis on understanding their condition and its implications. This aligns with prior work that emphasizes the importance of initial diagnostic understanding in shaping early-stage patient behaviors [53,54].

As the disease progresses, the focus shifts to managing daily life with kidney disease. On the Chronic Kidney Failure board, discussions centered on dietary restrictions, medication adherence, and emotional coping strategies. These findings align with qualitative studies that highlight the barriers to and facilitators of self-management, emphasizing that managing diet, medication regimens, and psychological well-being are critical yet challenging [55,56].

In the later stages of kidney disease, particularly for patients undergoing dialysis, discussions about transplantation became more prominent. On the Hemodialysis & Peritoneal Dialysis board, many patients compared dialysis methods and sought advice on the transplantation process. These findings underscore the transitional nature of this stage as patients undergoing dialysis often consider or prepare for transplantation. Previous research has similarly highlighted the psychological and logistical complexities associated with the transition from dialysis to transplantation [56].

For recipients of transplants, the focus shifted to challenges unique to posttransplantation life. Posts on the Kidney Transplant board were dominated by topics related to *family*

XSL•FO
RenderX

health and support as well as *medication and side effects*. These discussions reflect the complex challenges faced by recipients of transplants, including adherence to strict medication regimens, management of side effects, and navigation of relationships with family members who may have been donors. This aligns with previous qualitative research highlighting the importance of social support, effective communication, and self-management strategies in facilitating successful posttransplant adaptation [56]. Posts also revealed emotional and relational challenges as recipients of transplants grappled with their roles within family dynamics and the impact of their health on loved ones.

In addition to these stage-specific concerns, our analysis highlighted challenges faced by patients with comorbidities and special populations. The Hypertension & Diabetes board prominently featured the *disease management* topic, underscoring the added complexity of simultaneously managing multiple chronic conditions. The findings emphasize the need for integrated care approaches that address the interplay between kidney disease and common comorbidities [57,58]. The Pediatric Kidney Disease board's notable focus on the *medication and side effects* topic reflected the heightened concern about treatment impacts on children. This underscores the importance of providing tailored support and education to meet the specific needs of these populations.

By examining shifts in patient priorities and concerns, our analysis offers valuable insights into the evolving challenges faced by patients with kidney disease at different stages of their journey. These findings underscore the importance of stage-specific, patient-centered approaches in clinical care and support services.

Emotional Challenges and Support

The sentiment analysis revealed that discussions about medication side effects elicited the most negative emotions. The results identified particularly high levels of negative emotions in the topics of *medication and side effects* and *surgery for dialysis*. This underscores the difficulties that patients with kidney disease face regarding pain management as pain is a significant issue for these patients and significantly impacts their quality of life. Pharmacological changes associated with kidney disease increase the risk of side effects from pain medications, complicating pain management [59,60]. Specifically, >58% of patients undergoing dialysis experience pain, with 49% rating it as moderate to severe [61]. Medications typically need to be taken in the long term, causing ongoing concerns about side effects, which can lead to heightened negative emotions. Similarly, posts on the *surgery for dialysis* topic expressed negative emotions as patients began an indefinite period of dialysis.

The *family health and support* topic showed a mix of negative sentiments. Negative sentiments regarding kidney transplants and their impact on families were also prevalent, consistent with studies showing that both recipients of transplants and their families experience significant stress and concerns [62]. However, successful kidney transplants are associated with improved health outcomes and increased life satisfaction, as indicated by the relatively higher proportion of positive sentiments in posts on this subtopic [63,64]. Thus, kidney transplantation, meaning the end of long-term dialysis and dietary restrictions, which can allow patients to return to a more normal lifestyle, can increase positive sentiments.

Financial Concerns

The emergence of *costs and insurance* as a significant topic indicates that financial burden is a major concern for patients with kidney disease. This finding is consistent with previous research highlighting the economic challenges faced by patients with kidney disease and their families, especially those with limited financial resources [65]. The economic strain of managing CKD encompasses expenses for medical supplies, specialized diets, and transportation to medical appointments. These challenges are exacerbated by various factors, including increased costs during hospital stays for transplantation; additional expenses due to delayed graft function, which can cost an additional average of US $18,000 [66]; and the need for multiple dialysis sessions. The severity of the kidney condition typically correlates with the magnitude of the financial impact [65-67].

The prolonged nature of kidney disease often makes it challenging for patients to maintain stable employment, further contributing to their financial difficulties. The strain of managing a long-term illness can lead to interruptions in careers, thereby increasing economic hardships for both patients and their families. Many patients and caregivers report reduced work hours or job loss, leading to further financial strain [68,69]. The financial burden varies between different countries due to differences in local health care systems, yet the impact on work hours and productivity is universally challenging [70].

Disease Management Challenges

Consistent with previous research highlighting the role of online communities in providing a platform for individuals to share information, experiences, and support regarding health issues [71-73], our study found significant engagement in the *disease management* topic. Posts related to disease management often included sharing daily experiences and asking for advice, particularly about dietary management, which appeared more frequently compared to exercise, fluid intake, and weight management. Previous studies [58,73-75] have identified dietary management as one of the most challenging self-care tasks for patients with kidney disease. The dietary regimen for patients with kidney disease is complex and varies based on the stage of kidney disease or remaining kidney function, necessitating comprehensive management. Individual differences in taste, including a preference for salty food, pose another challenge for managing the disease.

Patients with comorbid conditions such as obesity or diabetes experience additional challenges in disease management [57,58]. Because dietary management is different for different kidney diseases, the question of whether a food item is acceptable for their condition was the most popular among the dietary management questions. This highlights the need for precise and tailored patient education regarding disease management from health care providers.

XSL•FO
RenderX

Comparison With Prior Work

Despite the active discussions in kidney disease communities, a notable paucity of research analyzing these communities exists. To our knowledge, the only study that has previously analyzed a kidney disease community focused on the National Kidney Foundation's online community for CKD [73]. However, this study only qualitatively analyzed approximately 20,000 posts without using topic modeling techniques. In contrast, our study used text mining methods to analyze a larger dataset, capturing more comprehensive insights into the perspectives prevalent in kidney disease communities. Another key advantage of our study was the ability to analyze posts based on disease state, which was not possible in the National Kidney Foundation's online community due to the lack of separate boards for different disease states. The previous study did not conduct a sentiment analysis either.

Other studies exploring the challenges faced by patients with kidney disease have relied on surveys or interviews [67,76]. Analyzing online communities offers practical and voluntary insights into consumer sentiment [77], as highlighted by previous studies [78,79]. Our approach captured a more mainstream perspective due to the vast amount of user-generated content available on the web over extended periods [79]. Traditional survey methods are limited by resource constraints in analyzing data within specific spatial and temporal boundaries. In contrast, diverse individuals contribute extensive posts on web-based platforms, making text mining particularly valuable for the qualitative analysis of such large datasets.

Our findings resonate with the themes identified in previous studies. For instance, previous studies [73,76] have highlighted themes related to *managing CKD and symptoms*, *medical tests and results*, and *navigating health care and clinical needs*, which translated to *medication and side effects*, *tests and diagnostics*, and *disease management* in our study. In addition, *family health and support* and *dialysis surgery* have been similarly categorized as themes such as *considering dialysis* and *family and transplantation*, reflecting the dynamics of CKD status. *Support on the forum*, a significant theme in previous research [73], was classified in our study as *sharing information* within *disease management*, emphasizing the role of information exchange and communication in disease management.

Limitations

While our study provides valuable insights into the discussions and challenges in online kidney disease communities, it is not without limitations.

First, the data were limited to posts from the KidneyCafe community, which may not represent all patients with kidney disease. Individuals who do not actively participate in online communities—due to factors such as digital literacy, access to technology, or personal preferences—were likely underrepresented. This limitation may particularly affect perspectives from older adults and individuals in socioeconomically disadvantaged groups. In addition, the specific context of Korean patients with kidney disease, such as differences in race, lifestyle, economic status, diet, and health care systems, may not be fully generalizable to other countries. These factors play a crucial role in shaping patient experiences and discussions. Conducting comparative analyses that use data from various web-based platforms and formal health care databases can offer a more thorough understanding of how health care systems and societal attitudes affect patient conversations. Furthermore, cross-platform comparisons would help validate our findings and reveal any potential biases specific to certain platforms in patient narratives. Expanding future studies to encompass diverse communities and geographical regions would improve the generalizability and comprehensiveness of our findings, providing deeper insights into the experiences and needs of patients with kidney disease worldwide.

Second, while the topic modeling method (LDA) was effective in extracting main topics, it does not inherently determine the optimal number of topics and may not fully capture semantic relationships between topics. The ELECTRA model, while efficient for our classification tasks, could benefit from domain-specific pretraining on medical corpora to enhance its performance on health-related text. These methodological constraints may limit the depth and precision of the topic modeling and classification results.

Third, the manual annotation process for classifying question posts related to disease management achieved substantial interrater agreement (Cohen κ coefficient of 0.77). However, the potential for bias cannot be entirely ruled out. To mitigate this, clear and detailed annotation guidelines were developed and thoroughly discussed by all annotators before the process began. Regular discussions were held to resolve disagreements and ensure consistency in interpretations, considering additional context from posts when necessary to reach consensus. While these measures contributed to the high interrater reliability observed, the limited number of annotators may have introduced some residual bias.

Fourth, the Naver CLOVA Sentiment platform's BERT-based analysis provides significant advantages for Korean text analysis, particularly its ability to accurately interpret nuanced, context-dependent expressions in kidney disease discussions. However, this approach also has its limitations. Domain-specific lexicon-based methods, widely used in English-language sentiment analysis, provide greater transparency by enabling researchers to trace sentiment classifications back to specific words or phrases [80-82]. These methods have demonstrated effectiveness and explainability in both medical [80,81] and public health research [82], achieving meaningful results. However, their application to Korean-language analysis is constrained by the limited scale and domain-specific coverage of existing Korean sentiment lexicons [83]. This limitation highlights the trade-offs inherent in our methodology—while the BERT-based approach effectively captures contextual nuances, it is less interpretable compared to lexicon-based methods. In addition, the Naver CLOVA Sentiment platform's computational intensity and token limit constraints may restrict its ability to process longer posts comprehensively. These choices reflect our decision to prioritize accuracy in contextual understanding, which was crucial for analyzing complex and nuanced patient discussions in Korean medical contexts.

XSL•FO
RenderX

Future research could address these limitations by incorporating data from multiple online communities, formal health care databases, and private discussions through surveys or interviews. Combining qualitative and quantitative methods, such as incorporating expert review processes or using correlated topic models, could further refine topic selection and analysis. Expanding the scope to include diverse communities from different countries would provide broader perspectives, enhancing the comprehensiveness and generalizability of the findings. These efforts would contribute to a deeper understanding of the experiences and needs of patients with kidney disease across varied contexts.

Clinical Practice Implications

The findings of this study have important implications for clinical practice, emphasizing the crucial role that health care providers play in addressing the emotional concerns and information needs of patients with kidney disease, as is prominently reflected in web-based discussions [84].

The high prevalence of negative sentiments in posts about medication side effects and dialysis surgery underscores the need for health care providers to proactively address patients' fears and anxieties. Strategies may include providing detailed explanations of what to expect during treatments, offering practical coping mechanisms, and regularly monitoring patients' emotional well-being. Research indicates that such proactive support can enhance treatment adherence and improve overall clinical outcomes [85].

Our findings underscore the necessity for clearer and more consistent dietary recommendations as diet-related inquiries dominated the *disease management* topic. Dietary adherence is particularly challenging for patients with kidney disease due to the complexity of nutritional restrictions and their impact on daily life [86]. Patients often express confusion and frustration with conflicting or vague dietary advice, which hinders effective disease management. Health care providers should prioritize offering patient-specific, practical, and easy-to-understand dietary recommendations tailored to the disease stage, comorbidities, and individual preferences. Personalized dietary counseling has been shown to improve adherence and clinical outcomes [87]. Regular follow-up consultations are essential for adjusting these recommendations based on patients' experiences and evolving health needs.

In addition, posts expressing frustration, confusion, or anxiety underscore the importance of empathetic communication and psychosocial support from health care providers. Integrating emotional support into routine care can effectively address these concerns. By fostering empathetic patient-provider relationships and offering clear, actionable guidance on disease management—especially regarding diet—health care providers can empower patients to actively participate in their care, ultimately improving their quality of life.

Conclusions

Our study offers an in-depth analysis of web-based discussions among patients with kidney disease, highlighting key concerns and informational needs. The findings underscore the importance of online communities in providing support and education to patients and caregivers. These insights can inform health care providers and policy makers when developing targeted interventions to address the specific needs of patients with kidney disease. Precise and tailored patient education regarding disease management from health care providers is clearly needed and can be significantly enhanced through the insights gathered from online community discussions.

Acknowledgments

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry through the High Value-Added Food Technology Development Program funded by the Ministry of Agriculture, Food, and Rural Affairs (grant 321029-05).

Data Availability

The anonymized dataset and analysis scripts used in this study are available upon reasonable request to the corresponding author.

Authors' Contributions

HJH contributed to conceptualization, data curation, formal analysis, and writing—original draft. NRK contributed to conceptualization and investigation. JYY was involved in investigation. HRR participated in investigation. SYK contributed to conceptualization. JHYP was responsible for writing—review and editing. KWL provided supervision and acquired funding.

Conflicts of Interest

None declared.

References

1. Ronco C, Bellomo R, Kellum JA. Acute kidney injury. The Lancet 2019 Nov;394(10212):1949-1964. [doi: 10.1016/s0140-6736(19)32563-2]
2. Romagnani P, Remuzzi G, Glassock R, Levin A, Jager KJ, Tonelli M, et al. Chronic kidney disease. Nat Rev Dis Primers 2017 Nov 23;3:17088. [doi: 10.1038/nrdp.2017.88] [Medline: 29168475]

XSL•FO
RenderX

3. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. Lancet 2017 Mar 25;389(10075):1238-1252. [doi: 10.1016/S0140-6736(16)32064-5] [Medline: 27887750]
4. Lv JC, Zhang LX. Prevalence and disease burden of chronic kidney disease. Adv Exp Med Biol 2019;1165:3-15. [doi: 10.1007/978-981-13-8871-2_1] [Medline: 31399958]
5. Bello AK, Levin A, Tonelli M, Okpechi IG, Feehally J, Harris D, et al. Assessment of global kidney health care status. JAMA 2017 May 09;317(18):1864-1881 [FREE Full text] [doi: 10.1001/jama.2017.4046] [Medline: 28430830]
6. James G, Nyman E, Fitz-Randolph M, Niklasson A, Hedman K, Hedberg J, et al. Characteristics, symptom severity, and experiences of patients reporting chronic kidney disease in the PatientsLikeMe online health community: retrospective and qualitative study. J Med Internet Res 2020 Jul 15;22(7):e18548 [FREE Full text] [doi: 10.2196/18548] [Medline: 32673242]
7. Ricardo AC, Anderson CA, Yang W, Zhang X, Fischer MJ, Dember LM, et al. Healthy lifestyle and risk of kidney disease progression, atherosclerotic events, and death in CKD: findings from the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 2015 Mar;65(3):412-424 [FREE Full text] [doi: 10.1053/j.ajkd.2014.09.016] [Medline: 25458663]
8. Tran WC, Huynh D, Chan T, Chesla CA, Park M. Understanding barriers to medication, dietary, and lifestyle treatments prescribed in polycystic kidney disease. BMC Nephrol 2017 Jul 05;18(1):214 [FREE Full text] [doi: 10.1186/s12882-017-0641-3] [Medline: 28679364]
9. Chen YY, Li CM, Liang JC, Tsai CC. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
10. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
11. Zhang S, Bantum EO, Owen J, Bakken S, Elhadad N. Online cancer communities as informatics intervention for social support: conceptualization, characterization, and impact. J Am Med Inform Assoc 2017 Mar 01;24(2):451-459 [FREE Full text] [doi: 10.1093/jamia/ocw093] [Medline: 27402140]
12. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
13. de Angelis G, Wells GA, Davies B, King J, Shallwani SM, McEwan J, et al. The use of social media among health professionals to facilitate chronic disease self-management with their patients: a systematic review. Digit Health 2018;4:2055207618771416 [FREE Full text] [doi: 10.1177/2055207618771416] [Medline: 29942633]
14. Kim HS, Jeong YJ, Baik SJ, Yang SJ, Kim TM, Kim H, et al. Social networking services-based communicative care for patients with diabetes mellitus in Korea. Appl Clin Inform 2016 Sep 28;7(3):899-911 [FREE Full text] [doi: 10.4338/ACI-2016-06-RA-0088] [Medline: 27679839]
15. Kang E, Ju H, Kim S, Choi J. Contents analysis of thyroid cancer-related information uploaded to YouTube by physicians in Korea: endorsing thyroid cancer screening, potentially leading to overdiagnosis. BMC Public Health 2024 Apr 02;24(1):942 [FREE Full text] [doi: 10.1186/s12889-024-18403-2] [Medline: 38566004]
16. Jo W, Kim Y, Seo M, Lee N, Park J. Online information analysis on pancreatic cancer in Korea using structural topic model. Sci Rep 2022 Jun 23;12(1):10622 [FREE Full text] [doi: 10.1038/s41598-022-14506-1] [Medline: 35739151]
17. Yu JE, Ko KJ, Kim JY. Public perceptions of enuresis: insights from online communities in South Korea and the United States. Int Neurourol J 2024 Sep;28(3):239-249 [FREE Full text] [doi: 10.5213/inj.2448318.159] [Medline: 39363415]
18. Hao T, Huang Z, Liang L, Weng H, Tang B. Health natural language processing: methodology development and applications. JMIR Med Inform 2021 Oct 21;9(10):e23898 [FREE Full text] [doi: 10.2196/23898] [Medline: 34673533]
19. Elbattah M, Arnaud É, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and applications. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-Up. 2021 Presented at: BIOSTEC 2021; February 11-13, 2021; Virtual event. [doi: 10.5220/0010414508250832]
20. Kidney disease patients' group. KidneyCafe. URL: https://cafe.naver.com/tlswkd [accessed 2023-04-20]
21. Park EL, Cho S. KoNLPy: Korean natural language processing in Python. In: Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology. 2014 Presented at: HCLT 2014; October 10-11, 2014; ChunCheon, Korea URL: https://koreascience.kr/article/CFKO201408355727285.page
22. Park S, Han S, Kim J, Molaie MM, Vu HD, Singh K, et al. COVID-19 discourse on Twitter in four Asian countries: case study of risk communication. J Med Internet Res 2021 Mar 16;23(3):e23272 [FREE Full text] [doi: 10.2196/23272] [Medline: 33684054]
23. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
24. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018 Nov 28;78(11):15169-15211. [doi: 10.1007/s11042-018-6894-4]
25. Chen Z, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R. Discovering coherent topics using general knowledge. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. 2013 Presented at: CIKM '13; October 27-November 1, 2013; San Francisco, CA. [doi: 10.1145/2505515.2505519]
26. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational

XSL•FO
RenderX

Natural Language Learning. 2012 Presented at: EMNLP-CoNLL '12; July 12-14, 2012; Jeju Island, Korea URL: https://aclanthology.org/D12-1087/
27. Danesh F, Dastani M, Ghorbani M. Retrospective and prospective approaches of coronavirus publications in the last half-century: a latent Dirichlet allocation analysis. Library Hi Tech 2021 Apr 01;39(3):855-872. [doi: 10.1108/LHT-09-2020-0216]
28. Yeruva VK, Junaid S, Lee Y. Contextual word embeddings and topic modeling in healthy dieting and obesity. J Healthc Inform Res 2019 Jun 10;3(2):159-183 [FREE Full text] [doi: 10.1007/s41666-019-00052-5] [Medline: 35415426]
29. Batool A, Byun YC. Enhanced sentiment analysis and topic modeling during the pandemic using automated latent Dirichlet allocation. IEEE Access 2024;12:81206-81220. [doi: 10.1109/access.2024.3411717]
30. Duraivel S, Lavanya L, Augustine A. Understanding vaccine hesitancy with application of latent Dirichlet allocation to Reddit Corpora. Ind J Sci Technol 2022 Oct 7;15(37):1868-1875. [doi: 10.17485/ijst/v15i37.687]
31. Min KB, Song SH, Min JY. Topic modeling of social networking service data on occupational accidents in Korea: latent Dirichlet allocation analysis. J Med Internet Res 2020 Aug 13;22(8):e19222 [FREE Full text] [doi: 10.2196/19222] [Medline: 32663156]
32. Li Z, Wu X, Xu L, Liu M, Huang C. Hot topic recognition of health rumors based on anti-rumor articles on the WeChat official account platform: topic modeling. J Med Internet Res 2023 Sep 21;25:e45019 [FREE Full text] [doi: 10.2196/45019] [Medline: 37733396]
33. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
34. Hagen L. Content analysis of e-petitions with topic modeling: how to train and evaluate LDA models? Inf Process Manag 2018 Nov;54(6):1292-1307. [doi: 10.1016/j.ipm.2018.05.006]
35. Janmohamed K, Soale AN, Forastiere L, Tang W, Sha Y, Demant J, et al. Intersection of the web-based vaping narrative with COVID-19: topic modeling study. J Med Internet Res 2020 Oct 30;22(10):e21743 [FREE Full text] [doi: 10.2196/21743] [Medline: 33001829]
36. Yoo R, Kim SY, Kim DH, Kim J, Jeon YJ, Park JH, et al. Exploring the nexus between food and veg*n lifestyle via text mining-based online community analytics. Food Qual Prefer 2023 Mar;104:104714. [doi: 10.1016/j.foodqual.2022.104714]
37. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: a survey. WIREs Data Mining Knowl Discov 2018;8(4):e1253. [doi: 10.1002/widm.1253]
38. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on October 11, 2018 2025 [FREE Full text]
39. Kim S, An M, Lee H, Kang J. Materiality-based online complaint classification: an analytical framework for efficient public service using text mining. ICIC Express Lett 2024;15(1):51-60 [FREE Full text]
40. Hur J, Yang J. South Korean newspaper coverage of Yemeni refugees: analysis of topics and sentiments using machine learning techniques. Asian J Commun 2023 Sep 13;34(1):57-72 [FREE Full text] [doi: 10.1080/01292986.2023.2257230]
41. Kum J, Jung S, Lee M. The effect of eye contact in multi-party conversations with virtual humans and mitigating the Mona Lisa effect. Electronics 2024 Jan 19;13(2):430. [doi: 10.3390/electronics13020430]
42. Yang K. Transformer-based Korean pretrained language models: a survey on three years of progress. arXiv. Preprint posted online on November 25, 2021 2025 [FREE Full text] [doi: 10.48550/arXiv.2112.03014]
43. Clark K, Luong MT, Le QV, Manning CD. ELECTRA: pre-training text encoders as discriminators rather than generators. arXiv Preprint posted online on March 23, 2020. [doi: 10.48550/arXiv.2003.10555]
44. Code of Federal Regulations: Additional Protections for Children Involved as Subjects in Research. U.S. Department of Health and Human Services. 2018 Jun 19. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-d/index.html#46.404
45. Ni C, Song Q, Malin B, Song L, Commiskey P, Stratton L, et al. Examining online behaviors of adult-child and spousal caregivers for people living with Alzheimer disease or related dementias: comparative study in an open online community. J Med Internet Res 2023 Nov 17;25:e48193 [FREE Full text] [doi: 10.2196/48193] [Medline: 37976095]
46. Kong D, Chen A, Zhang J, Xiang X, Lou WQ, Kwok T, et al. Public discourse and sentiment toward dementia on Chinese social media: machine learning analysis of Weibo posts. J Med Internet Res 2022 Sep 02;24(9):e39805 [FREE Full text] [doi: 10.2196/39805] [Medline: 36053565]
47. Roh GH, Lee S, Kim JY, Lee S. SNSMiner_VAC: analyzing vaccination based on social network service data for safety surveillance. Expert Syst Appl 2024 Dec;255:124684. [doi: 10.1016/j.eswa.2024.124684]
48. Kim NR, Hong SG. Text mining for the evaluation of public services: the case of a public bike-sharing system. Serv Bus Preprint posted online June 30, 2020. [doi: 10.1007/s11628-020-00419-4]
49. Jung HS, Yoon HH, Song MK. A study on dining-out trends using big data: focusing on changes since COVID-19. Sustainability Preprint posted online October 18, 2021. [doi: 10.3390/su132011480]
50. Weiss A, Hendrickx R, Stensgaard E, Jellingsø M, Sommer MO. Kidney transplant and dialysis patients remain at increased risk for succumbing to COVID-19. Transplantation 2023 May 01;107(5):1136-1138 [FREE Full text] [doi: 10.1097/TP.0000000000004462] [Medline: 36584380]

XSL•FO
RenderX

51. Trbojevic-Stankovic J, Marjanović Z. #4153 Outcomes of COVID-19 infection among maintenance hemodialysis patients – experience from the first COVID-19 hospital in Serbia. Nephrol Dial Transplant 2023 Jun;38(Supplement_1). [doi: 10.1093/ndt/gfad063c_4153]
52. Silberzweig J, Wu S, Sinclair M, Watson T, Welder N, Concepcion D, et al. Response to COVID-19: the outpatient dialysis setting. Clin J Am Soc Nephrol 2023 Jan 30;18(7):949-952. [doi: 10.2215/cjn.0000000000000091]
53. Levey AS, Coresh J. Chronic kidney disease. The Lancet 2012 Jan;379(9811):165-180. [doi: 10.1016/s0140-6736(11)60178-5]
54. Jamieson NJ, Hanson CS, Josephson MA, Gordon EJ, Craig JC, Halleck F, et al. Motivations, challenges, and attitudes to self-management in kidney transplant recipients: a systematic review of qualitative studies. Am J Kidney Dis 2016 Mar;67(3):461-478. [doi: 10.1053/j.ajkd.2015.07.030] [Medline: 26372087]
55. Cardol CK, Meuleman Y, van Middendorp H, van der Boog PJ, Hilbrands LB, Navis G, et al. Psychological distress and self-management in CKD: a cross-sectional study. Kidney Med 2023 Oct;5(10):100712 [FREE Full text] [doi: 10.1016/j.xkme.2023.100712] [Medline: 37753249]
56. Memory KE, Wilkinson TJ, Smith AC, Lightfoot CJ. A qualitative exploration of the facilitators and barriers to self-management in kidney transplant recipients. J Nephrol 2022 Sep 25;35(7):1863-1872 [FREE Full text] [doi: 10.1007/s40620-022-01325-w] [Medline: 35467326]
57. Potrykus M, Czaja-Stolc S, Małgorzewicz S, Proczko-Stepaniak M, Dębska-Ślizień A. Diet management of patients with chronic kidney disease in bariatric surgery. Nutrients 2022 Dec 29;15(1):165 [FREE Full text] [doi: 10.3390/nu15010165] [Medline: 36615822]
58. Kalantar-Zadeh K, Fouque D. Nutritional management of chronic kidney disease. N Engl J Med 2017 Nov 02;377(18):1765-1776. [doi: 10.1056/nejmra1700312]
59. Lambourg E, Colvin L, Guthrie G, Walker H, Bell S. Analgesic use and associated adverse events in patients with chronic kidney disease: a systematic review and meta-analysis. Br J Anaesth 2022 Mar;128(3):546-561 [FREE Full text] [doi: 10.1016/j.bja.2021.08.035] [Medline: 34763813]
60. Messaoud FB, Sellami N, Saïed S, Miladi S, Abderrahim E. #2901 Pain management in patients with chronic kidney disease: beware of inertia!. Nephrol Dial Transplant 2023(Supplement_1) [FREE Full text] [doi: 10.1093/ndt/gfad063c_2901]
61. Davison SN. Pain, analgesics, and safety in patients with CKD. Clin J Am Soc Nephrol 2015 Mar 06;10(3):350-352 [FREE Full text] [doi: 10.2215/CJN.00600115] [Medline: 25710800]
62. Frontini R, Sousa H, Ribeiro Ó, Figueiredo D. "What do we fear the most?": exploring fears and concerns of patients, family members and dyads in end-stage renal disease. Scand J Caring Sci 2021 Dec 07;35(4):1216-1225. [doi: 10.1111/scs.12940] [Medline: 33615525]
63. Kobayashi S, Akaho R, Omoto K, Shirakawa H, Shimizu T, Ishida H, et al. Post-donation satisfaction in kidney transplantation: a survey of living donors in Japan. BMC Health Serv Res 2019 Oct 26;19(1):755 [FREE Full text] [doi: 10.1186/s12913-019-4556-5] [Medline: 31655578]
64. Kamran F. Does perceived health status influence quality of life after renal transplantation. Am J Appl Psychol 2014;2(4):99-103 [FREE Full text] [doi: 10.12691/ajap-2-4-4]
65. Ganji S, Ephraim PL, Ameling JM, Purnell TS, Lewis-Boyer LL, Boulware LE. Concerns regarding the financial aspects of kidney transplantation: perspectives of pre-transplant patients and their family members. Clin Transplant 2014 Oct 25;28(10):1121-1130 [FREE Full text] [doi: 10.1111/ctr.12428] [Medline: 25066730]
66. Kim DW, Tsapepas D, King KL, Husain SA, Corvino FA, Dillon A, et al. Financial impact of delayed graft function in kidney transplantation. Clin Transplant 2020 Oct 11;34(10):e14022 [FREE Full text] [doi: 10.1111/ctr.14022] [Medline: 32573812]
67. Garcia-Sanchez JJ, Kularatne T, West B, Rao N, Wright J, Reichel H, et al. FC005: pace CKD: qualitative and quantitative insights into the economic burden of CKD on patients and carers. Nephrol Dial Transplant 2022 May;37(Supplement_3) [FREE Full text]
68. Cruz-Flores MA. Employment status and coping strategies of renal patients on hemodialysis. MOJ Public Health 2023 May 8;12(2):88-93. [doi: 10.15406/mojph.2023.12.00415]
69. Alma MA, van der Mei SF, Brouwer S, Hilbrands LB, van der Boog PJ, Uiterwijk H, et al. Sustained employment, work disability and work functioning in CKD patients: a cross-sectional survey study. J Nephrol 2023 Apr 31;36(3):731-743 [FREE Full text] [doi: 10.1007/s40620-022-01476-w] [Medline: 36315355]
70. Rajkumar R, Baumgart A, Martin A, Tong A, Evangelidis N, Manera KE, et al. Perspectives on ability to work from patients' receiving dialysis and caregivers: analysis of data from the global SONG initiative. J Nephrol 2022 Jan 09;35(1):255-266. [doi: 10.1007/s40620-021-01105-y] [Medline: 34241814]
71. Vasilica CM, Brettle A, Ormandy P. A co-designed social media intervention to satisfy information needs and improve outcomes of patients with chronic kidney disease: longitudinal study. JMIR Form Res 2020 Jan 27;4(1):e13207 [FREE Full text] [doi: 10.2196/13207] [Medline: 32012040]
72. Lightfoot CJ, Wilkinson TJ, Hadjiconstantinou M, Graham-Brown M, Barratt J, Brough C, et al. The codevelopment of "My Kidneys and Me": a digital self-management program for people with chronic kidney disease. J Med Internet Res 2022 Nov 14;24(11):e39657 [FREE Full text] [doi: 10.2196/39657] [Medline: 36374538]

XSL•FO
RenderX

73. Du Y, Dennis B, Ramirez V, Li C, Wang J, Meireles CL. Experiences and disease self-management in individuals living with chronic kidney disease: qualitative analysis of the National Kidney Foundation's online community. BMC Nephrol 2022 Mar 04;23(1):88 [FREE Full text] [doi: 10.1186/s12882-022-02717-7] [Medline: 35246060]
74. Langley-Evans S, Thomas N. The challenge of nutritional management in people with kidney disease. J Ren Care 2017 Dec 30;43(4):195-196. [doi: 10.1111/jorc.12227] [Medline: 29098807]
75. Cupisti A, Avesani CM, D'Alessandro C, Garibotto G. Nutritional management of kidney diseases: an unmet need in patient care. J Nephrol 2020 Oct 12;33(5):895-897. [doi: 10.1007/s40620-020-00829-7] [Medline: 32785842]
76. Montalescot L, Dorard G, Speyer E, Legrand K, Ayav C, Combe C, et al. Patient perspectives on chronic kidney disease and decision-making about treatment. Discourse of participants in the French CKD-REIN cohort study. J Nephrol 2022 Jun 13;35(5):1387-1397 [FREE Full text] [doi: 10.1007/s40620-022-01345-6] [Medline: 35696043]
77. Chen Y, Dong T, Ban Q, Li Y. What concerns consumers about hypertension? A comparison between the online health community and the Q and A Forum. Int J Comput Intell Syst 2021;14(1):734-743. [doi: 10.2991/ijcis.d.210203.002]
78. Park Y, Park S, Lee M. Analyzing community care research trends using text mining. J Multidiscip Healthc 2022 Jul;Volume 15:1493-1510. [doi: 10.2147/jmdh.s366726]
79. Shi J, Chen Y. User and context integrated experience mining in online health communities. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020 Presented at: CIKM '20; October 19-23, 2020; Virtual event, Ireland. [doi: 10.1145/3340531.3417410]
80. Gore R, Safaee MM, Lynch CJ, Ames CP. A spine-specific lexicon for the sentiment analysis of interviews with adult spinal deformity patients correlates with SF-36, SRS-22, and ODI scores: a pilot study of 25 patients. Information 2025 Jan 24;16(2):90. [doi: 10.3390/info16020090]
81. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
82. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
83. Cho H. Comparison of three Korean sentiment/emotion word dictionaries. In: Proceedings of the IEEE International Conference on Big Data and Smart Computing. 2023 Presented at: BigComp 2023; February 13-16, 2023; Jeju, Republic of Korea. [doi: 10.1109/bigcomp57234.2023.00073]
84. Morton RL, Devitt J, Howard K, Anderson K, Snelling P, Cass A. Patient views about treatment of stage 5 CKD: a qualitative analysis of semistructured interviews. Am J Kidney Dis 2010 Mar;55(3):431-440. [doi: 10.1053/j.ajkd.2009.11.011] [Medline: 20116914]
85. Davis KM, Eckert MC, Hutchinson A, Harmon J, Sharplin G, Shakib S, et al. Effectiveness of nurse-led services for people with chronic disease in achieving an outcome of continuity of care at the primary-secondary healthcare interface: a quantitative systematic review. Int J Nurs Stud 2021 Sep;121:103986. [doi: 10.1016/j.ijnurstu.2021.103986] [Medline: 34242979]
86. Kim SM, Jung JY. Nutritional management in patients with chronic kidney disease. Korean J Intern Med 2020 Nov;35(6):1279-1290 [FREE Full text] [doi: 10.3904/kjim.2020.408] [Medline: 32872726]
87. Ko GJ, Kalantar-Zadeh K, Goldstein-Fuchs J, Rhee CM. Dietary approaches in the management of diabetic patients with kidney disease. Nutrients 2017 Jul 31;9(8):824 [FREE Full text] [doi: 10.3390/nu9080824] [Medline: 28758978]

Abbreviations

AKI: acute kidney injury
BERT: Bidirectional Encoder Representations From Transformers
CKD: chronic kidney disease
ELECTRA: Efficiently Learning an Encoder That Classifies Token Replacements Accurately
LDA: latent Dirichlet allocation
NLP: natural language processing

Edited by T de Azevedo Cardoso; submitted 02.08.24; peer-reviewed by A Mallett, R Gore, M Elbattah, FJ Yang; comments to author 04.11.24; revised version received 08.12.24; accepted 29.01.25; published 25.02.25.

Please cite as:
Hwang HJ, Kim N, You JY, Ryu HR, Kim SY, Yoon Park JH, Lee KW
Harnessing Social Media Data to Understand Information Needs About Kidney Diseases and Emotional Experiences With Disease Management: Topic and Sentiment Analysis
J Med Internet Res 2025;27:e64838
URL: https://www.jmir.org/2025/1/e64838
doi:10.2196/64838
PMID:39998877

XSL•FO
RenderX

©Hee Jeong Hwang, Nara Kim, Jeong Yun You, Hye Ri Ryu, Seo-Young Kim, Jung Han Yoon Park, Ki Won Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Women's Educating and Coping Strategies for Cultivating Supportive Web-Based Spaces for Discussing Sexual and Reproductive Health: Co-Design Study

Hyeyoung Ryu[1], MS; Wanda Pratt[1], PhD

Information School, University of Washington, Seattle, WA, United States

Corresponding Author:
Hyeyoung Ryu, MS
Information School
University of Washington
Mary Gates Hall 370
Box 352840
Seattle, WA, 98195-2840
United States
Phone: 1 206 543 6653
Email: hyryu115@uw.edu

Abstract

Background: Stigma surrounding women's sexual and reproductive health (SRH) often prevents them from seeking essential care. In South Korea, unmarried women face strong cultural taboos, increasing their risk for conditions such as pelvic inflammatory disease, infertility, and cervical cancer. While many unmarried women turn to web-based communities for support, these spaces frequently expose them to microaggressions, further discouraging their access to health care and worsening their health risks.

Objective: We aimed to encourage a safe space for seeking support on the culturally taboo topic of SRH by counteracting and reducing web-based microaggressions. We sought to make these last-resort safe spaces supportive by reducing and preventing microaggressions, fostering coping strategies, and educating rather than solely punishing perpetrators.

Methods: We conducted co-design sessions with 14 unmarried Korean women. In the first co-design session, we introduced the term microaggression and collaborated with participants to create base design components aimed at countering and preventing microaggressions. In the second co-design session, participants initially viewed examples of microaggression comments, then designed using the provided base design templates inspired by their suggestions from the first session and finally designed for a scenario where they would be seeking support. We analyzed co-design session transcripts using inductive and deductive methods.

Results: Our analysis revealed 6 goals addressing coping strategies, educational approaches, and cultural characteristics shaping participants' designs. Reflective coping strategies were supported through designs that numerically indicate positive support and provide holistic views of diverse perspectives, helping participants reassess provocative situations with cognitive clarity. Suppressive coping strategies were fostered by encouraging less-emotional responses, empowering participants to address microaggressions logically without self-blame. Educational approaches emphasized fostering shared awareness of microaggressions and providing respectful education for perpetrators about the harm their words can cause. Participants suggested counterspeech mechanisms, including rephrasing suggestions and public educational resources, to balance education with freedom of expression. They also proposed that forum-approved experts guide discussions to ensure accurate, empathetic responses and support users in addressing nuanced situations effectively. Cultural characteristics heavily influenced these goals. Participants noted the nebulous nature of microaggressions, their reluctance to burden their social support network, and societal perceptions of women as overly emotional—all of which shaped their desire for designs that enhance logical justification. For example, participants preferred tools such as expert-led discussions and comprehensive perspectives to rationalize their experiences while reducing stigma.

Conclusions: Our work advocates for prioritizing educational and explanatory approaches over punitive detection and deletion measures to create supportive web-based spaces for individuals discussing stigmatized SRH. By integrating culturally informed coping strategies, counter speech mechanisms, and educational designs, these tools empower microaggression targets and allies while fostering reflection and behavior change among perpetrators. Our work provides a first step toward counteracting microaggressions and ultimately encouraging women to seek the needed SRH care.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e62716) doi:10.2196/62716

KEYWORDS

sexual and reproductive health; women's health; stigma; microaggression; coping; counternarrative; social media; web-based communities; support; co-design

Introduction

Background

In many cultures, including South Korea (hereafter referred to as Korea), unmarried women grapple with the stigma surrounding sexual and reproductive health (SRH), often opting to suffer in silence rather than seek necessary care [1-6]. Conversations about SRH topics—such as contraception, menstruation, sexual discomfort, or sexual pleasure—remain taboo, hindering access to essential services, such as screening and prevention of sexually transmitted infections (STIs) [3,7]. This reluctance to seek care poses substantial health risks, including pelvic inflammatory disease, ectopic pregnancy, and infertility [8]. Despite efforts to challenge traditional beliefs and decouple female sexuality from marriage [8,9], many women remain reticent to openly discuss SRH or seek related care, such as cervical cancer screening and prevention [6]. The consequences of this stigma can be dire. For example, Korean women who face many microaggressions regarding SRH [10] also experience disproportionately high rates of cervical cancer, with both native and immigrant populations being affected [6]. Despite the availability of preventive measures such as the human papillomavirus vaccine, the persisting stigma surrounding SRH care hampers efforts to address these health disparities [10]. Stigmatized SRH refers to topics considered taboo for unmarried women, such as STIs, sexual pleasure or discomfort, pregnancy, and contraception. However, the stigma also extends beyond these issues, limiting access to care for non-STI or pregnancy-related concerns, such as cervical cancer, and underscoring the broader barriers unmarried women face in SRH care.

For such stigmatized health topics, many people turn to web-based communities for support [11,12], but those web-based spaces also often contain microaggressions—communications that put down, insult, or invalidate them. These microaggressions not only harm people's physical and psychological health [13-17] but also discourage people from seeking necessary health care [10,18,19]. This problem is particularly salient for women seeking SRH care, as it is a taboo topic in many cultures [1,20,21]. The perpetrators of web-based microaggressions include users who intentionally or unintentionally harm support seekers, including within-group perpetrators [10,15,22]. In this study, we focused on the web-based microaggressions that South Korean women face, often from other women, when seeking support for the recommended SRH care.

To jointly imagine designs that could help counteract web-based microaggressions, we co-designed with unmarried women—the targets of these microaggressions—who grew up in Korea and have used web-based forums to seek SRH care. In this study, we define the term *used* as not solely the active use of posting or commenting but also the passive use of viewing posts and comments of the web-based forum user's interest. In addition, we define "targets" as the women who experience microaggressions related to their SRH. This definition aligns with the concept of *targets* in microaggression literature by Sue et al [23], which describes individuals who are subjected to prejudice and discrimination. While the study by Sue et al [23] focused on racial microaggressions, we expanded the term to include gender-based prejudice and discrimination.

In this study, we focused on designing solutions to counteract web-based microaggressions to cultivate a resilient and safe space for support on stigmatized SRH. We sought to make these last-resort safe spaces truly supportive by reducing the negative effects of microaggressions on the targets and allies' health and well-being and educating rather than solely punishing perpetrators. We refer to allies as individuals who challenge bias within their own communities and actively support others, even within marginalized groups [24,25]. Examples include anonymous women users in web-based health communities, as well as friends who, while also facing stigma related to SRH, work to counteract the microaggressions they experience or witness others experiencing.

Our work focuses on presenting the findings from 2 co-design sessions with unmarried Korean women. We identified 6 goals participants had for counteracting web-based microaggressions to create a resilient and safe space for seeking and providing support on stigmatized SRH. On the basis of these 6 goals, we uncovered that, unlike most web-based content-moderation approaches that focus on detecting and deleting offensive content, most women would prefer systems that help them cope with web-based microaggressions and help perpetrators reflect on the ramifications of their words. In this paper, we have contributed designs that support these goals and preferences.

Related Work

In this section, we connect with 3 bodies of work related to our research: rethinking content moderation, supporting vulnerable users' coping, and designing without compromising vulnerable users' agency.

Rethinking Content Moderation: The Shortcomings of Detection and Deletion Approaches

The predominant content-moderation strategy on platforms such as Facebook and Twitter (subsequently rebranded X) heavily relies on detection and deletion of harmful content [26-28]. These methods focus on removing content after it has been flagged, which makes them predominantly reactive. This approach can leave users feeling unsupported and frustrated, especially when it comes to content that falls into a "gray area," such as microaggressions, that may not explicitly violate platform policies but still causes harm [29,30].

Ma et al [30] highlight this gap, noting that the current detection and deletion model fosters an adversarial relationship between

XSL•FO
RenderX

users and platforms, where moderation feels punitive rather than supportive. Myers West [31] also emphasizes the need for more nuanced moderation practices, particularly in ambiguous cases where platform guidelines are unclear, stressing the importance of building a culture of responsible web-based citizenship that goes beyond mere content removal. Similarly, Sharevski et al [29] demonstrate the ineffectiveness of warning tags and labels, which often fail to engage users meaningfully or prevent the harm caused by problematic content. These limitations underscore the weaknesses of the detection-first approach.

Despite advancements in moderation technologies, the focus remains on swiftly identifying and removing harmful content [32-35]. This method overlooks deeper issues of user discontent, as Jhaver et al [32] report, where users express frustration over content removal without clear explanations or guidance, often feeling that their content was unjustly deleted. These issues demonstrate that the current detection and deletion approach may exacerbate tension rather than address the root causes of harmful behavior [36-38].

In this study, we aim to address these gaps by exploring alternatives to detection and deletion, focusing on strategies that offer more nuanced, supportive solutions to handling harmful content—particularly in complex cases such as microaggressions.

Supporting Vulnerable Users' Coping

The current narrow focus on detection and deletion often results in harm to the very communities that we are trying to support or protect [39,40]. For instance, Thach et al [39] demonstrate how the invisible process of web-based content moderation makes marginalized social media users face disproportionate content moderation and removal (eg, computational model could flag a person of color's post about racial justice or racist experience as hate speech). They state that the content removal or banning of the poster from the web-based community would bring further marginalization of the already marginalized users. Thus, in an attempt to better support and protect marginalized social media users, researchers from the University of Michigan School of Information developed Online Identity Health Center—a website that helps marginalized users (1) understand different social media platforms' policies and rules about content takedowns and (2) provide easy-to-understand resources on different social media guidelines and what to do if your content is taken down [41].

In support of alternative approaches that extend the web-based content-moderation task beyond detection and permanent removal, studies on dealing with the negative consequences of microaggressions [14,33-36] have also emphasized the importance of coping in reframing web-based content moderation to better support marginalized users. Coping with negative emotions and stress induced by microaggressions is crucial to reducing the negative effects on the targets' health and well-being [14]. Microaggressions are perpetrated and experienced subtly and often unconsciously; the target often questions the reality of the oppression [42,43]. Thus, targets of microaggressions frequently blame themselves for being overly sensitive and dismiss, if not excuse, the behavior of the perpetrators. Targets of microaggressions also show a pattern of quietly shouldering the impact of such behavior, sometimes out of fear of retaliation from the perpetrator or others involved. They also worry about reinforcing negative stereotypes (eg, women are emotionally needy and overly sensitive) if they choose to confront the perpetrator [44]. Targets of microaggressions often experience negative impacts on their well-being (eg, feelings of guilt, embarrassment, shame, regret, and remorse) after unconfronted microaggressions. Furthermore, internalizing these feelings has long-lasting effects on the target's self-confidence and mental and physical health [45,46].

In contrast, if targets or allies choose confrontation, they must take on additional risks for themselves. Targets frequently need to circumvent fears of either being further invalidated or reinforcing negative stereotypes about their group if they were to confront microaggressions. Empirical evidence supports this notion; targets report that using confrontation opened them up to further microaggressions (eg, being further invalidated and being called "oversensitive" and "paranoid") [47,48]. Thus, coping is essential for targets and allies in dealing with microaggressions.

Coping can be carried out through 2 emotion regulation strategies: cognitive reappraisal and expressive suppression [49,50]. Cognitive reappraisal refers to re-evaluating the meaning of a given situation to reduce its emotional impact [49,50], whereas expressive suppression refers to inhibiting an emotional response [49,51]. Although suppression may help a person avoid undesirable interpersonal consequences that can follow from expressing negative emotions, it has generally been found to be ineffective in reducing the negative emotions themselves [52-55]. Thus, Juang et al [51] suggested combining the use of both emotion regulation strategies when people encounter microaggressions. With the combined use, they indicated that expressive suppression will help in avoiding further negative interpersonal interactions, and cognitive reappraisal will help ease the emotional burden of denigration. This combined use is further corroborated by the study of Hernández and Villodas [56], where the sole use of reactive and suppressive coping was discouraged. Only reflective coping (ie, which relates to cognitive reappraisal) was associated with more positive mental health, whereas reactive and suppressive coping (ie, which relates to expressive suppression) were associated with poorer mental health after experiencing microaggressions [56]. Thus, in this study, we aimed to design features that help counteract microaggressions and do not merely detect and delete content but rather support people in reflective coping.

Designing Without Compromising Vulnerable Users' Agency

The population we are designing for are microaggression targets within an oppressive social structure, where support for SRH care is difficult to find [10]. Thus, understanding how to design within oppressive social structures without compromising the agency of vulnerable users was pivotal to our work. Sultana et al [24] conducted a study on designing within a patriarchal society and investigated the economic, social, and cultural challenges faced by rural Bangladeshi women. They then

XSL•FO
RenderX

presented 2 user-centered strategies for design to empower the vulnerable users. The first strategy was to try to empower women within the structures of their society, instead of trying to destroy those structures [24]. This strategy was also supported by Alsheikh et al [57] who suggested designing for openness, which supports Islamic feminist values that posit women as having agency but also operating under a set of Islamic cultural assumptions. They defined supporting agency as allowing women in Arabic culture to enact particular cultural roles in the context of their relationships rather than designing for Western feminist values (eg, demanding equality) or the patriarchal values underlying the Arabic sociocultural context.

The second strategy was to shift focus from the problems that women face to the tactics they already use to cope and look for opportunities to support these tactics [24]. Rabaan et al [58] resonated with this strategy and emphasized the importance of understanding the interwoven complexity of the sociocultural dimensions of the problem, considering historical developments, and accepting and respecting the different choices people make while understanding why they make them. With an in-depth understanding of the scarce structural support in protecting Saudi women from domestic abuse, they uncovered that resiliency was a valuable strategy for these women to sustain moving forward and even to take action to end the violence. To design for a customized interactive system that supports resiliency, Rabaan et al [58] suggested designing a screening tool to measure where a woman is at in her options, resources, needs, and concerns and what she perceives as expected or abusive.

In this study, we incorporated the 2 strategies into our design considerations. First, we designed within existing structures using the web-based forums that unmarried Korean women often use to not only seek SRH care support but also reflect the oppressive social structure in Korea [10]. Second, we focused on the tactics that unmarried Korean women already use to cope and looked for opportunities to support those tactics.

Methods

In this section, we describe our study's ethical considerations; co-design process, which includes co-design session 1—the creation of a base design template for improvisational enactment and co-imagination of new ideas—and co-design session 2; and co-design session analysis process.

Ethical Considerations

This study received an exempt status from the University of Washington Human Subjects Division under Category 2, as it involves only interviews with data recorded in a manner that either does not allow for participant identification or poses no risk of harm if disclosed. The institutional review board ID is STUDY0013885. This co-design study received institutional review board approval from the University of Washington. Informed consent was obtained from all participants before their involvement in this study. Participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits, as well as their rights to confidentiality and the option to withdraw from the study at any time without penalty. In addition, participants were given the opportunity to ask questions and receive clarification before providing their consent. All study data were deidentified, with participants referred to by their ID (eg, P [participant ID number]) in co-design transcripts and session designs. Personal information, such as names and ages, was stored in a separate file. Participants were compensated with a US $25 Amazon gift card for each co-design session, totaling US $50 for their participation. To protect participants' privacy, we ensured that no personal identifiers were linked to any images included in the manuscript.

Co-Design Process

Overview

Co-design, defined by Sanders and Stappers [59] and Kleinsmann and Valkenburg [60], involves collective creativity and shared knowledge among participants, leading to a comprehensive understanding of design content and process [59-61]. This process, also known as joint inquiry and imagination, encourages ethical engagement and starts with abduction, suggesting potential solutions [61,62]. Our study used a modified approach for remote co-design sessions, aiming to stimulate creativity and facilitate discussions on sensitive topics such as SRH care [10,63]. In the following sections, we describe our participants, details for each of the co-design procedures, and our analysis approach.

Participant Recruitment and Consent

We recruited unmarried Korean women between the ages of 18 and 29 years, a demographic most susceptible to microaggressions around SRH care due to the assumption that they are unmarried. These younger women often seek support on the web for SRH care, and we wanted to capture a holistic understanding of the challenges they face. Among the participants, only 7 had sought clinical SRH care (eg, obstetrics and gynecology office, women's health clinic, and public health office), while others solely relied on web-based forums. Despite these differences, both groups used web-based spaces to navigate SRH care support, so we interviewed both to inform designs that address microaggressions targeting unmarried Korean women.

Participants were required to have received 9 to 12 years of compulsory education in Korea and to be currently unmarried. Those who had received foreign curriculum–based education during their compulsory education (eg, US curriculum–based schools) were excluded from the study. Recruitment was conducted through university board flyers at Yonsei University in Korea, web-based SRH support communities for Korean women such as women-only forums in Nate Pann, and snowball sampling.

Participants for this study were recruited using a multistep process designed to ensure their informed and voluntary participation. Potential participants initially expressed interest by completing a screening questionnaire. This questionnaire was used to verify eligibility based on the study's inclusion and exclusion criteria.

XSL•FO
RenderX

Once eligibility was confirmed, participants received a consent form via email, which provided comprehensive details about the study. The consent form outlined the study's purpose, the inclusion and exclusion criteria, the procedures involved in participation, the potential benefits and risks, the compensation offered, and the measures taken to protect participant confidentiality and privacy. Participants were encouraged to review the consent form independently, and a minimum of 48 hours was provided for this review to ensure they had adequate time to make an informed decision.

Following this review period, we scheduled individual calls with each participant to discuss the consent form further. During these calls, participants had the opportunity to ask questions, seek clarification about the study, and address any concerns they might have. After ensuring that participants fully understood the consent form and its implications, we proceeded with obtaining their signatures. Once consent was obtained, we scheduled the first co-design session with each participant.

Co-Design Session 1 Procedure

We conducted co-design session 1 as contextual research to provide the necessary background for co-design session 2 (ie, support the creation of the base design components for the co-design sessions). Co-design session 1 was conducted in a web-based setting, and each session lasted approximately 50 to 60 minutes. In co-design session 1, we asked questions about who should be responsible for counteracting microaggressions and how technology could support them. As the topic of gynecological care is taboo, we told participants that they could answer for themselves or someone they know. We first asked participants about their SRH care support–seeking experiences. Then, we explained to them what microaggressions are and showed them examples of microaggressions derived from the study by Ryu and Pratt [10]. Afterward, we asked them to think of microaggression counteraction or prevention design ideas based on their web-based SRH care support–seeking experiences.

After we received participant-generated design suggestions without specific design scenarios, we asked them about specific features (eg, flagging comments, blocking content, and content alerts) derived from currently implemented measures of major social media platforms such as Facebook [26,28] and Twitter [27,28]. Furthermore, we asked them about specific situations that social media platforms have not been able to resolve, such as "gray area" cases of moderation [30]—instances where the removed content was not wildly in violation of explicit guidelines and required some interpretation [31] (eg, when the perpetrator or target is unaware of the harm caused by microaggressions). During co-design session 1, we told participants to ignore current technological or financial limitations when ideating their designs.

Co-Design Session 2 Procedure

On the basis of co-design session 1, we created a base design template that we showed to the participants during co-design session 2. We made a post comment sample using a hypothetical scenario with a post topic commonly seen in the web-based forums that unmarried Korean women seek SRH care support on. The hypothetical scenario focused on vaginitis and clinical visit concerns (Figure 1) because it is one of the most common concerns shared by women in web-based forums [10]. On the basis of our reviews of existing forum posts, we included comments that reflect those that showed genuine support or microaggressions from other users. We also included the ally's response to the microaggressions and showed the microaggression perpetrators attacking the ally to reflect the current state of web-based support and microaggressions [10,42,44,64]. For the format and features (eg, post and comment like and dislike functions with the number of likes and dislikes displayed and comment filter options), we made the post comment sample reflect those of web-based forums that unmarried Korean women typically seek support from. To provide an initial design that participants could react to and provide feedback on, we presented templates for each participant-suggested design found in co-design session 1 next to the post comment sample.

XSL•FO
RenderX

Figure 1. Design session sample scenario. It mimics the format and features of forums that unmarried Korean women currently go to seek sexual and reproductive health care support. The features that the design session sample scenario reflects from the original forum are the (1) post structure (ie, title, author ID, date, post views, post content, and number of post recommendations and oppositions); (2) comment structure (ie, comment filtering options and replier name, reply date, number likes and dislikes, and a comment flagging button for each comment or replies to head comment); and (3) reply and comment posting structure (ie, reply writing box and reply posting button).



Co-design session 2 was divided into three sections: (1) viewing only the post comment sample (Figure 1) without the base design template; (2) designing after viewing the base design template, which were reenactments of the participant-suggested designs; and (3) designing for a situation where the participant themself would be the support-seeking post author. Each design session was held with each participant individually using Lucidchart software (Lucid Software Inc) as the co-design tool and lasted approximately 50 to 60 minutes. Lucidchart is a design tool that facilitates remote, synchronous, and collaborative design. As our participants took part in the co-design sessions remotely, we used Lucidchart to build, replicate, and share post comment samples and design templates developed during co-design session 1. In co-design session 2, participants used Lucidchart to design web-based microaggression counteraction strategies collaboratively and synchronously.

XSL•FO
RenderX

We used the same group of participants from the initial design needs interviews because we wanted to empower them by actualizing their own designs and reflecting on not only their own but also others' designs to counteract microaggressions. Participants received US $25 for participating in co-design session 2. We audio-recorded the interviews and used a professional transcription service to have the recordings transcribed verbatim.

We wanted to learn about the participants' design ideas when they saw the post comment sample, so we first asked the participants to review only the post comment sample and tell us about their thoughts on how they would design the forum to counteract web-based microaggressions. Then, we asked them the same question after showing the base design template for counteracting web-based microaggressions and prototyped each design and their suggested use. We encouraged them to talk aloud as they created their designs and prompted them to explain why they wanted the microaggression counteraction feature to be designed a certain way. We asked the participants to make edits on their own but assisted them when they were experiencing difficulties.

Before and during co-design session 2, we explained that they did not have to use the design in its suggested form or at all if they did not want to use the feature. Also, we emphasized that they could alter the design to better reflect their intended use of the design (eg, adding and removing options and changing the phrases). At the end of the session, we reviewed the designed web-based microaggression tool together to check if there had been any misinterpretations and to give them the opportunity to edit their designs.

Finally, to understand how the web-based microaggression counteraction design would change if they had designed it only for themselves, we asked the participants if they would change anything if they were the ones writing the posts and receiving the microaggressions.

Co-Design Session Analysis

A total of 2 authors participated in the analysis. The first author acknowledges her standpoint as an educated, unmarried Korean woman. She is a native Korean who received primary education in Korea and received additional intensive English language education in middle and high school. Although she is not an avid participant in the comment sections of the web-based spaces where unmarried Korean women seek support, she has investigated numerous interactions of other users, advocated for safe web-based spaces free of microaggressions, and is intrigued by the use of language to mark and protect microaggression targets. She acknowledges that her positionality influenced this study to some extent but also helped her interpret the participants' design ideas and influences from the Korean sociocultural context. The second author is from the United States and has done extensive research on web-based health communities as well as design research with people from marginalized communities. Her cultural background has made her much more aware of microaggressions and their negative effects on her life over decades, which has helped to complement the first author's more recent exposure to the concept of microaggressions. This complementary background also helps them consider the broader implications of this work. Because much of the second author's research portrays the benefits of web-based health communities, she may perceive more positive possibilities for web-based health communities in this study.

The first author completed the analysis of both co-design sessions 1 and 2, with support from the second author using the ATLAS.ti analysis software (Lumivero). We analyzed 28 transcripts in total, consisting of 14 (100%) transcripts from co-design session 1 and 14 (100%) transcripts from co-design session 2. Both co-design sessions were analyzed using the same procedure. The first author cycled between field notes kept during the co-design sessions and the co-design session transcripts in an approach meant to build on themes in real time. To begin, the first author coded 5 co-design session transcripts using an open coding approach [65], labeling what she saw in the data. Then, she discussed the codes with the second author, refined these into a revised codebook, and finished coding all the remaining transcripts. Following another round of discussion and clarification, the authors finalized the codebook. As open coding was completed, the first author conducted a round of deductive coding that focused on sentences about each microaggression counteraction feature. The 2 authors then discussed interpretations and addressed any discrepancies in the application of the code set of features to counteract microaggressions and adjusted their coded data as needed. We reached saturation after coding 6 (42%) co-design session 1 transcripts and 8 (57%) co-design session 2 transcripts.

Results

Overview

First, we report on the participant information and participant-suggested designs and how they envisioned their designs to be enacted from co-design session 1. We further show how the designs were enacted as a design template for them to use during co-design session 2. Second, we discuss 6 goals that participants had for creating a safe and resilient web-based community that would allow them to seek support for the culturally taboo topic of SRH.

Participants

As the co-design participants were asked to share their experiences that are not frequently or overtly discussed and are considered a sensitive topic in the culture, we used snowball sampling to recruit a total of 14 participants (Table 1); the mean age was 24 (SD 2.73) years.

XSL•FO
RenderX

Table 1. Participant demographics.

ID	Age (y)	Grew up in Korea	Currently living in Korea	Knowledge of microaggression before this study	Experience seeking SRH[a] care in a clinical setting[b]
P1	24	Yes	Yes	Yes	No
P2	24	Yes	Yes	Yes	No
P3	24	Yes	Yes	No	Yes
P4	24	Yes	Yes	No	Yes
P5	23	Yes	Yes	No	No
P6	23	Yes	Yes	Yes	No
P7	26	Yes	Yes	No	Yes
P8	24	Yes	Yes	Yes	No
P9	20	Yes	No	Yes	No
P10	25	Yes	Yes	No	Yes
P11	31	Yes	No	No	No
P12	25	Yes	No	No	Yes
P13	20	Yes	No	No	Yes
P14	27	Yes	Yes	No	Yes

[a]SRH: sexual and reproductive health.

[b]For example, obstetrics and gynecology clinic, women's health clinic, and public health office.

All participants had physical and financial access to SRH clinical services, but the women who reported "no" to seeking SRH in various clinical settings stated that they did not seek SRH care due to stigma. We explored whether there were differences in results between participants who had sought SRH clinical care and those who had not, but we found no noteworthy differences in the designs they created.

Participants who had no knowledge of microaggressions before this study elucidated that they had not heard of microaggressions, so even if they had experienced microaggressions, they were unable to name what they were experiencing. We explored whether having prior knowledge of microaggressions influenced the designs participants put forth, but we found no notable differences between those with and without prior knowledge. Participants who had prior knowledge of microaggressions had only a basic understanding, primarily from skimming social media posts. All participants noted that they would not be able to categorize microaggression types or provide a clear definition.

Initial Design Results

In co-design session 1, we identified the features participants wanted and whom they intended the features to be used by (Table 2).

Table 2. Participant-suggested features from co-design session 1.

Feature	Actor using feature
Labeling type of support	Woman seeking support[a]
Choosing percentage of microaggression to view	Woman seeking support
Alerting about microaggression content with extra step of consent to view content	Woman seeking support
Microaggression comment rephrasing suggestions	Potential microaggression perpetrator or ally[b]
Microaggression flagging explanations	Woman seeking support
Messages to microaggression perpetrators or targets	Woman seeking support and potential microaggression perpetrator or ally
Professional certification badge	Woman seeking support

[a]Post authors.

[b]Commenters.

First, to ensure that post authors receive the type of support they want, participants suggested 3 design features or functionalities that could be used while people are writing their posts. The first feature would allow post authors to label their posts with the type of support sought. Participants identified five types of support: (1) emotional support, (2) informational support, (3) women-only support, (4) expert-only support, and (5) debate. Second, participants suggested choosing the percentage of

XSL•FO
RenderX

comments with microaggressions post authors were willing to view. Participants suggested the inclusion of viewing all microaggression comments (100%) and no microaggression comments (0%) as options. Finally, inspired by the idea of content blocking on Instagram, participants wanted a microaggression content alert to initially block comments with microaggressions but allow people to view the comment if they click the button to view the comment.

To reduce the microaggressions received and to educate people writing comments, participants wanted support for people while they were writing comments. Participants envisioned features that would suggest rephrasing a microaggression with explanations for the suggestions. They either wanted the whole comment or specific parts of the comment to be highlighted. For the highlighted parts, participants wanted features that would provide rephrasing suggestions that could be applied with a click of a button. Some participants also wanted features that would provide the comment writer with explanations for why the rephrasing suggestion would make their comment not a microaggression. Other participants did not want a feature with rephrasing suggestions because they wanted to protect comment writers' freedom of speech.

Participants suggested three types of features to identify microaggressions in comments for the original poster, commenters, and readers after a comment has been posted. First, participants wanted to be able to flag comments that contain microaggressions along with an explanation for why it was a microaggression. Participants explained that the current social media platforms allow them to report or flag comments but that their choices for indicating why a comment should be flagged are too general to help commenters understand why their comments contained microaggressions. In particular, participants wanted the flagged reason to be shown in the public explanations for each flagged comment, which would help the community and the commenter learn about microaggressions. Second, participants wanted features that would send messages to the microaggression perpetrators or the targets of the microaggressions—who could be either the original poster or another commenter. To provide reassurance and validation of the targeted person's thoughts and feelings, participants wanted to send messages letting them know that the microaggressions were noticed. For the perpetrators of the microaggressions, participants thought that perpetrators might continue to write microaggressions if they were merely given warnings that their comments had been flagged. However, participants were uncertain about the exact content they would include in such messages or the flagging explanation, and the interviews did not include time for them to think through and compose a sample. Thus, in the design session, we provided an opportunity for participants to compose their explanations or messages if they wished to.

As a justification for their final post comment design suggestion, participants wanted health professionals to have more accountability for their words because the participants placed higher trust and reliance on these professionals. Participants suggested a feature that would flag comments from health professionals. Participants wanted health professionals to make their status as a health professional public, thus putting their status and reputation on the line and making them more responsible for what they say. Participants wanted to create a professional certification badge based on the badges stamped next to the professionals' replies in Naver Jisikin—a widely used question-and-answer forum in Korea.

With the participant-suggested designs and their descriptions of how they should be envisioned, we made templates for each design (Figure 2). These designs allowed participants to experience their designs' enactment in co-design session 2.

XSL•FO
RenderX

Figure 2. Co-design session 2 base design template. It shows the set of participant-suggested designs found in co-design session 1. For support during post writing, participants had the option to design with labeling the type of support post authors needed (a-1) which would be shown with the support type badge on the post (a-2), microaggression percentage (b), and microaggression content alert (c-1) which would be shown on the comments as (c-2). For during comment writing support, participants were able to design reply rephrase suggestions by highlighting the rephrase needed parts (g-1) and offering the rephrase suggestions with or without explanations (g-2). For support after a comment has been posted, participants could choose to implement the professional certification badge (e) and design the flagging options (d-1), flagging explanation messages (d-2), and messages to microaggression targets and perpetrators (f).

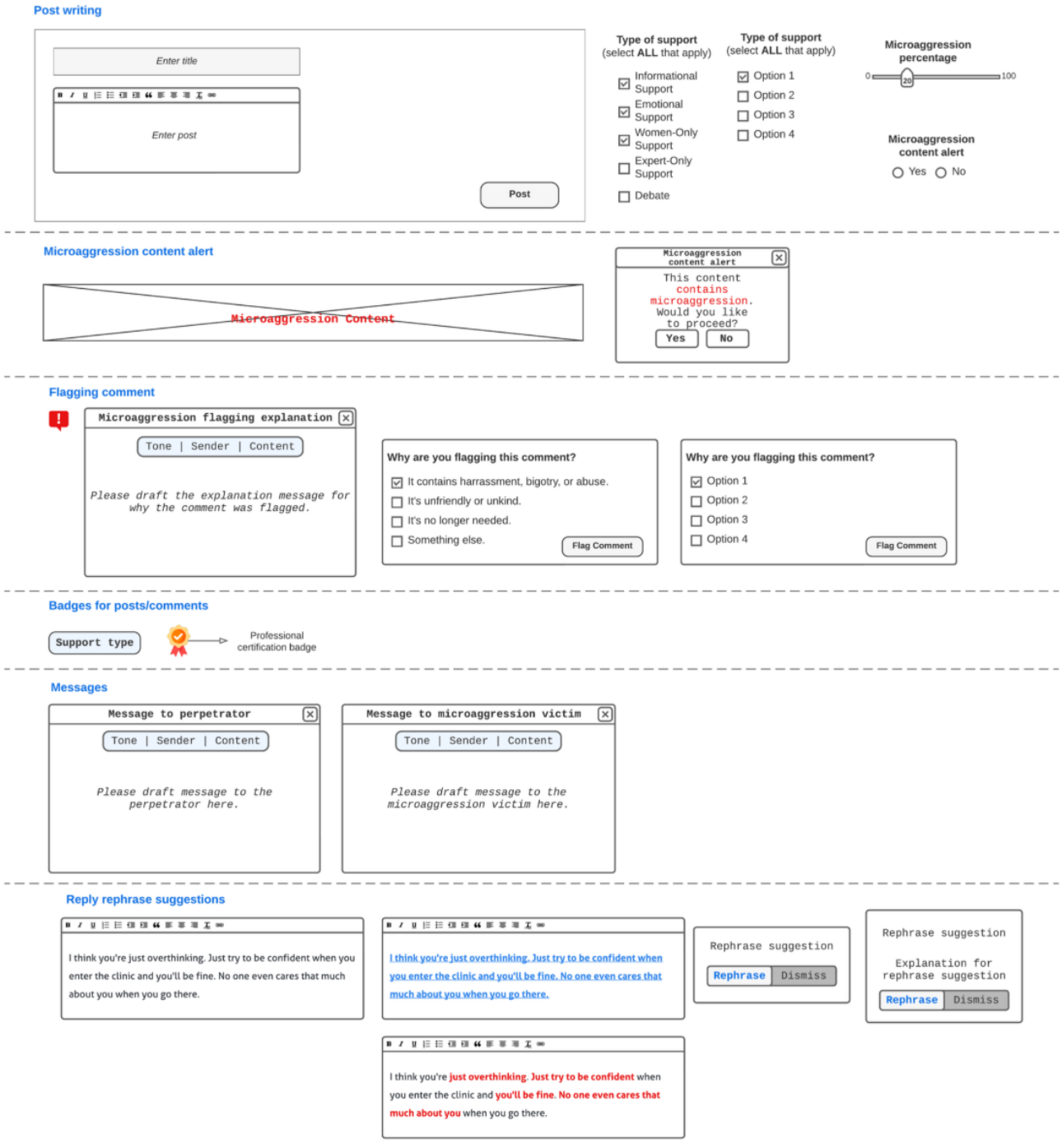

Design Goals

Overview

Through the analysis of the co-design sessions, we found that participants had 6 goals in counteracting web-based microaggressions to create a safe and resilient web-based community for seeking support on the stigmatized SRH. These goals arose in co-design session 1 but were further expanded upon in co-design session 2 after they had learned more about microaggressions and were able to tangibly interact with their envisioned designs. We first describe each of the goals and participants' design ideas for achieving these 6 goals. The figures of the design ideas shared in this paper were created by the participants using Lucidchart software.

- Goal 1: having shared knowledge of microaggressions among all users.

XSL•FO
RenderX

- Goal 2: indicating positive support numerically.
- Goal 3: responding with less emotion to microaggressions.
- Goal 4: having forum-approved experts guide the discussions.
- Goal 5: respecting perpetrators' freedom of speech while educating them to understand their words' negative impact.
- Goal 6: knowing the "reality."

Furthermore, we describe how the 6 goals were affected by 3 cultural characteristics: the nebulous nature of microaggressions, reluctance to burden their social support network, and questioning the validity of their own thoughts due to the social perception that women are too emotional.

Goal 1: Having Shared Knowledge of Microaggressions Among All Users

Participants advocated for providing people with educational resources about microaggressions and supporting people harmed by microaggressions. They wanted unmarried Korean women post authors to be more informed about microaggressions by adding links to microaggression-related educational resources and support groups under the microaggression content alert option (Figure 3):

Figure 3. Microaggression content alert design by P6. On the microaggression content alert, links to a mental health support page and an educational page about microaggression are added.

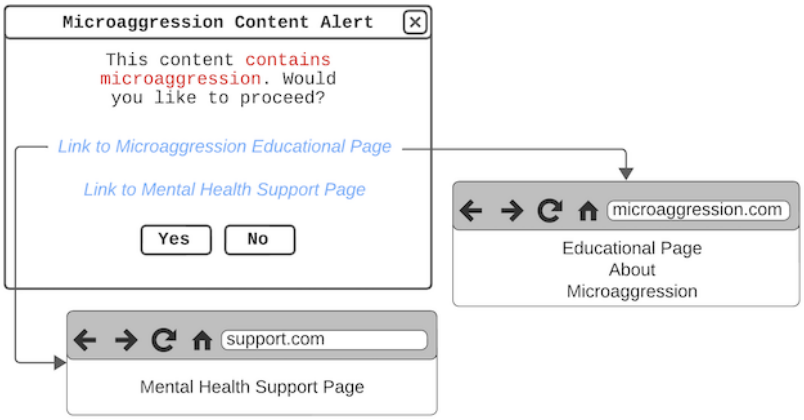

> *First of all, we have to understand that there's a lack of shared knowledge here. People don't know what microaggressions are. How are we supposed to defend ourselves without knowing what to defend ourselves from? When we see the [microaggression] content alert, we need to have additional resources or groups we can access to get support or a better understanding of microaggressions.* [P3]

Beyond the post authors, participants also wanted other web-based forum users to learn about microaggressions and their impacts. Thus, participants suggested adding microaggression educational resources to explanations next to flagged comments (Figure 4):

Figure 4. Educational flagged reason explanation design by P11. When the flagged button next to a flagged comment is clicked, the user can see a message that the comment contains microaggression content and links to microaggression support pages.

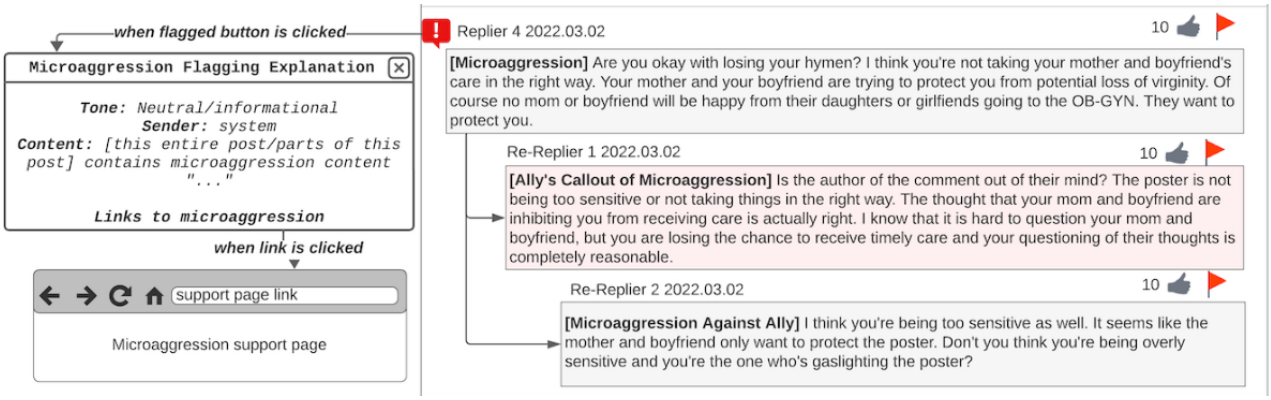

> *It's not just people who post on the [online] forums that need to know about microaggressions. Other people using the [online] forums also do too. I think the best way to do that is to have information [about microaggressions] or [website] links to microaggression information right next to the comments, the ones that are flagged, so everyone can see it.* [P11]

Furthermore, they wanted to ensure that perpetrators were educated on microaggressions by sending them a private message with additional microaggression educational resources attached to the explanation of why the comment was flagged (Figure 5):

XSL•FO
RenderX

Figure 5. Message to perpetrator design by P8. The message to the microaggression perpetrator is designed to have a neutral tone and shows the message that the perpetrator's comment has been flagged because it may contain microaggressions. The message also has links to help the perpetrator understand what microaggressions are.

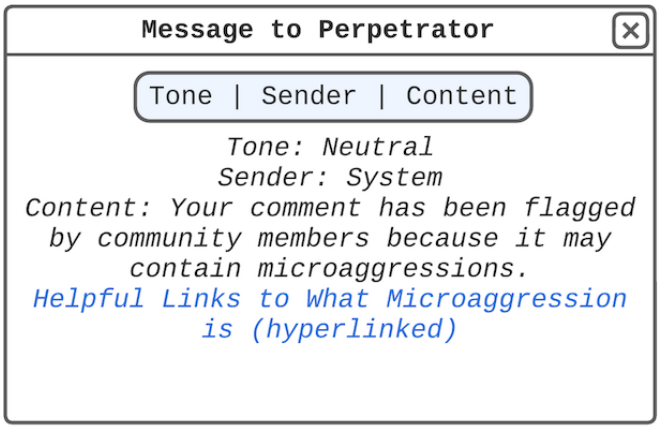

*I don't want the so-called perpetrators to be confused and be like what the f***? when they get a message after writing something that they think in their head is supportive. I don't want them [perpetrators] to be mad. I want them to understand. Having the message [sent to microaggression perpetrators privately] would point them to the right direction of understanding their "unintentional" wrongdoing.* [P4]

Goal 2: Indicating Positive Support Numerically

Participants advocated for emphasizing positive support through numerical indicators and deleting negative support indicators, such as thumbs-down or dislike options for posts and comments. They suggested transforming the like button into more supportive options (eg, text forms: helpful supportive; icon form: heart; Figure 6):

Figure 6. Support button design by P13. Instead of the like and dislike button, only the support button was made available with a heart icon to indicate support. The number of people who pressed the support button was displayed next to the support button.

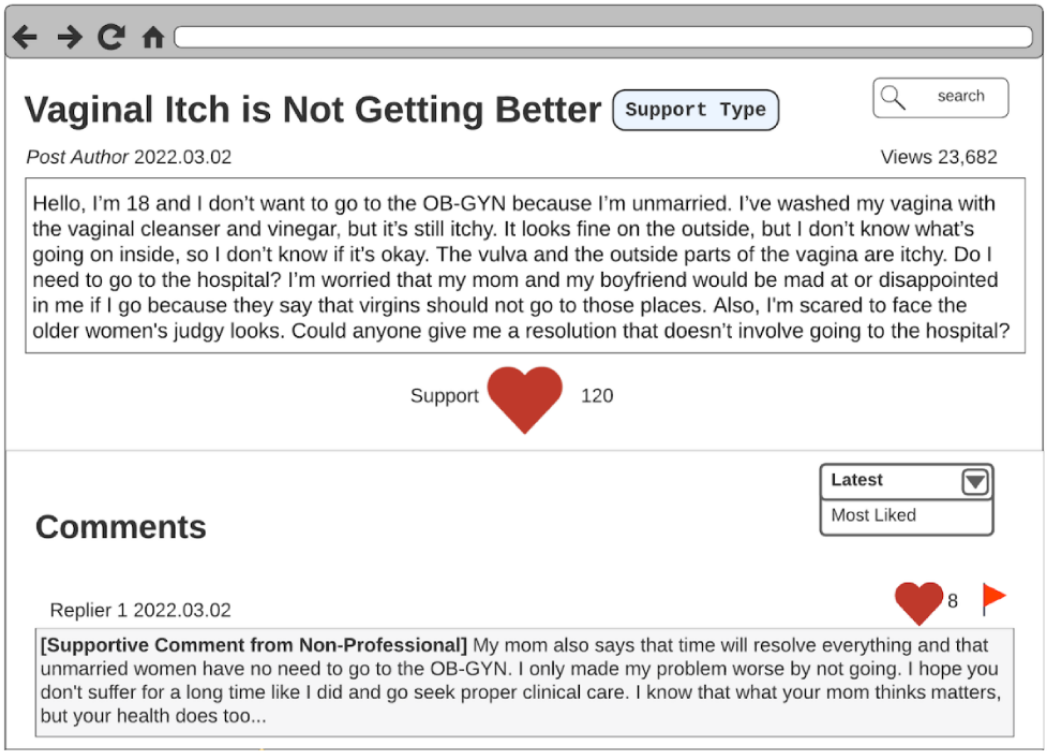

Why does the like button have to be a thumbs-up button? I think it could be words like "supportive". It shows more of why people clicked the button. People have to put in another layer of thought before pressing the button. They would have to think like "Oh was this actually supportive?" [P8]

All participants except one suggested the number of clicks on the transformed like button to be displayed next to each post or comment (Figure 6). With the number of likes, participants stated that they could organically find the comments most people found most supportive, which they would also likely find helpful:

XSL•FO
RenderX

> *Numbers are produced naturally. That sounds weird saying it out loud but it's something you can't refute. The more numbers, the more support (for the comment).* [P6]

As participants perceived that the numerical indicators would be helpful in finding truly supportive comments, many of the participants did not choose to send the messages to post authors who have faced microaggressions in the comment sections. Participants thought that sending the messages to post authors would be "unnecessarily coddling" [P6] when they could be given the agency to learn how to sort out truly supportive comments on their own with the help of numerical indicators:

> *I want to give women the strength to defend and sort out helpful information on their own. I don't know if I would describe the internet as a jungle, but it sometimes is. You got to know what's poison and what's not. Like the color of mushrooms, the brighter they are, the more poisonous they are. The numbers [of support] next to a comment is the color [of mushrooms].* [P1]

Goal 3: Responding With Less Emotion to Microaggressions

After evaluating widely used flagging options in social media, participants perceived the currently used flagging options to be too emotional. Participants stated that emotive language, such as unfriendly or unkind, is too subjective and suggested changing it to less-emotive words (eg, it contains misinformation and it promotes bullying or harassment; Figure 7):

Figure 7. Less-emotive flagging option design by P2. Less-emotive flagging options (eg, it contains misinformation and it promotes bullying or harassment) were used to replace the more emotive current flagging options (eg, it is unfriendly and it is unkind).

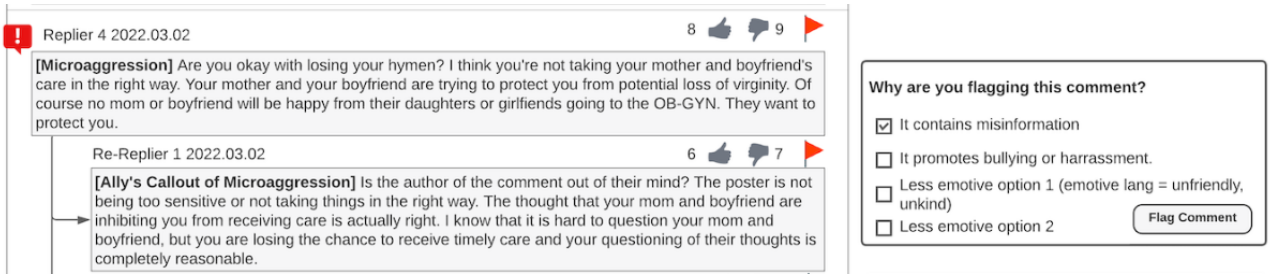

> *Words like unfriendly, unkind are too whatever you make of it. Microaggressions are already subjective or at least in how I understood it [microaggressions are subjective]. I think it amplifies that subjectiveness in the flagging explanation if we continue with the emotive language. We need to be less emotive using phrases like "It promotes bullying or harassment."* [P5]

Participants who wanted private messages to be sent to microaggression perpetrators thought that using a neutral tone stating why the comment was flagged would be sufficient. Participants did not think a more threatening or emotionally intense response would be needed. They stated that guiding microaggression perpetrators to the microaggression-related educational resources and providing opportunities for microaggression perpetrators to learn the impacts of their microaggression comments on their own would be more valuable than penalizing them:

> *Neutral [tone for the private messages sent to microaggression perpetrators] is good. We don't want to threaten them. What if they retaliate? I don't want them to keep causing harm. We want them on our side. In that case, opening their eyes to what they are not seeing is better. Make them learn more and realize on their own.* [P14]

Goal 4: Having Forum-Approved Experts to Guide the Discussions

Participants wanted experts to help provide guidance in sorting out helpful comments since they expected the forum-approved experts only to post helpful and truthful comments. To ensure that the forum-approved experts were posting only helpful and truthful comments, participants requested web-based forums to include the process of professional certification in the forum description page. In addition, they requested more information about the forum-approved professionals, such as their names, the clinics they work at, their comment histories, and a summary of the reviews about the professionals on their user page (Figure 8). With this information, participants wanted forum-approved experts to help guide users to disregard comments with microaggression as well as comments containing misinformation:

Figure 8. Professional certification badge and information page about professional design by P4. Professionals who are certified by the forum would have a badge next to their names when they post a comment. When other users click on the professional's ID, they would be able to see a hovering information page with information about the professional's name, the clinic they work at, comment history, and a summary of reviews about the professional.

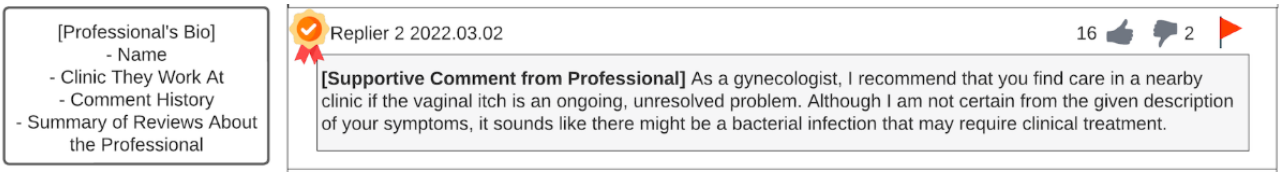

XSL•FO
RenderX

> *People can lie about being professionals and go around spreading misinformation, so for people who are seeking support, the badge would be helpful to differentiate between certified and non-certified professionals.* [P2]

Participants also requested certifications and badges for active forum users (Figure 9). Participants wanted the active forum users to be differentiated from people who were inexperienced about the topic or new to the forum. They wanted the active forum users to act as the "older sister" [P9] role by guiding the discussion in the right way for topics they were more knowledgeable about. With both types of certified experts, participants thought that microaggression comments that are disrespectful and contain misinformation could be weeded out and counteracted:

Figure 9. Active or experienced user badge design by P9. Active or more experienced users who could act as the older sister guide for less experienced users were identified by the forum with a crown badge.

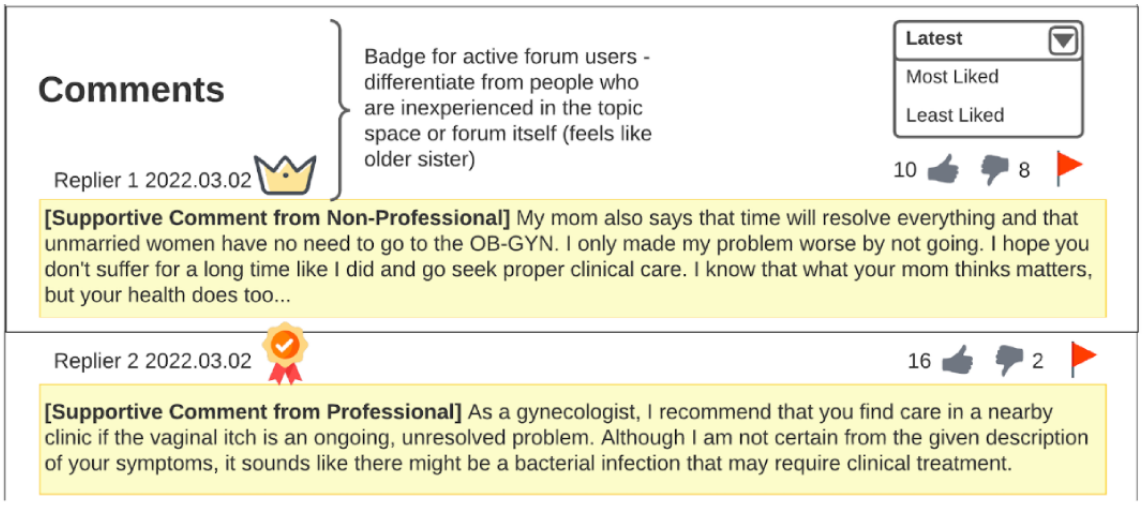

> *Kind, emphasis on the kind, experienced, or more knowledgeable users should be differentiated from people who are inexperienced in the topic space or forum itself. They know how to deal with people who are acting disrespectfully and fend them off, you know like an older sister protecting her younger sister at a playground.* [P9]

Goal 5: Respecting Perpetrators' Freedom of Speech While Educating Them to Understand Their Words' Negative Impact

Most participants did not think that microaggression perpetrators would be willing to change their microaggression comments when given a choice. Thus, they chose not to implement the microaggression comment rephrase suggestions:

> *It's rarely incidental when people are writing microaggressions. Deep down in their hearts, they probably do really know that they are causing harm or at least that what they're posting could be harmful. These people know what they are doing, so rephrasing suggestions or explanations for them is only helpful if they want to know.* [P2]

However, participants also wanted to respect the perpetrators' and potential allies' freedom of speech rather than add a burden on people who were merely offering help or support. They wanted microaggression perpetrators who had the potential to become allies to be less burdened by this "big brother monitoring" [P9]:

> *I like the idea of a little fixing [of microaggressions] but having it still be optional. Something just doesn't sit well with me if I am forcing the person writing the comment to change it before they even post it. I hope that they are smart enough to make the right decision and I want to believe that they could change the parts they want and make it less harmful on their own.* [P7]

Although the microaggression comment rephrasing feature was considered an unnecessary burden by most participants, the few who wanted to provide potential microaggression perpetrators or allies the option to rephrase their comments saw potential in it. The few participants who wanted to implement the microaggression comment rephrasing feature suggested providing both the rephrasing suggestions and explanations of why the suggestions would make the comment genuinely supportive (Figure 10). These participants firmly believed that potential allies "have a good heart" [P12] and would see the value of choosing to rephrase with the rephrasing suggestion when they see the explanation attached to the suggestion:

XSL•FO
RenderX

Figure 10. Microaggression comment rephrase suggestion and explanation design by P10. While writing the comment, the user is notified which parts of the comment contain microaggression. Then, the user receives a rephrase suggestion and an explanation for why the rephrase suggestion would make the current comment genuinely supportive. The user could choose to rephrase the comment by clicking on the "Rephrase" button or choose to not rephrase by clicking on the "Dismiss" button.

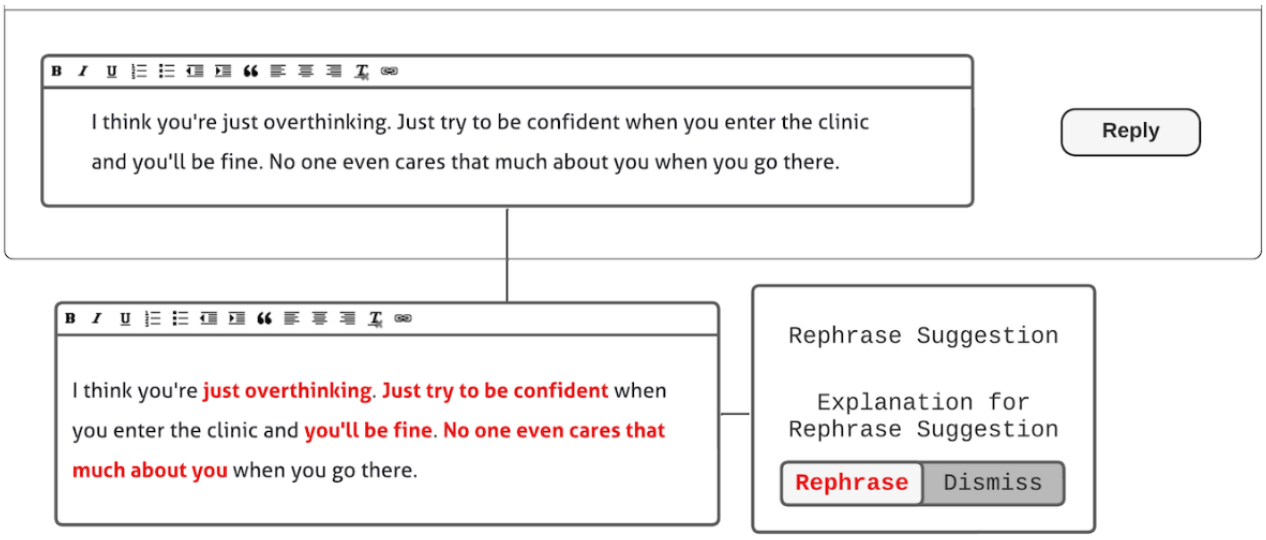

Commenters may not have had any intention to cause microaggressions, so to further decrease microaggressions that were not intentional, having explanations [for why their comments are microaggressive and how the suggested changes make them less microaggressive] would be helpful. They don't know the why, and people need to know the why, if there was no ill-heartedness behind their actions, to make changes to them. [P4]

Goal 6: Knowing the "Reality"

This section is not about findings on microaggression counteraction, but rather the byproduct of participants' evaluation of the effects of counteracting or blocking all microaggressions. During co-design session 1 ideation and at the beginning of co-design session 2, participants were focused on finding ways to counteract and combat microaggressions, as shown in the earlier subsections. However, near the end of the co-design session 2, when they were asked if there was anything they would like to change if they were the sole users of the web-based community and the microaggression counteraction features, participants stated that they had the right to know the truth or reality they were facing, even if it meant that they had to directly interact with microaggressions. Participants were concerned that showing a rosy environment where "everything is unicorn and rainbows" [P2] is deceptive and thought that users should be aware of the reality. Thus, they suggested the implementation of the debate and seeking opinions option for the support type labels and controversiality as a comment sorting option when they were designing microaggression counteraction measures for solely themselves (Figure 11):

XSL•FO
RenderX

Figure 11. Support type label and comment sorting option design by P2. The "Debate" option was added as a type of support that post authors could indicate on their posts. In the comment sorting options, the "Controversial" option was added to see comments that had the most controversy among users.



> *I can honestly handle whatever it is because I want to get the whole picture – both sides, not just one cuddly side. I want to be able to face reality. It would be even better to not only have the true opinions accessible, but also easier to access. Maybe by sorting the comments here [the comment sorting button] with a controversiality sorting.* [P3]

The change in design for themselves showed that the participants found value in incorporating all opinions, even negative, discriminatory, and disheartening ones, to help them make a decision about their SRH care with the information gathered. However, they still wanted to have other microaggression counteraction measures put into effect, which showed that even with the features suggested to know the reality, microaggression counteraction was still deemed by the participants as pivotal and necessary:

> *I still want everything else [the other microaggression counteraction features P2 applied to the design for other unmarried Korean women users] there. I still want to be protected. I just also want to know the controversial opinions to make sure that I have all the information I need to make the most right? I'm not sure if that's the right word, but you know what I mean. The most optimal decision.* [P2]

Cultural Characteristics Affecting Microaggression Counteraction Goals

We uncovered cultural characteristics that affected the 6 microaggression counteraction goals and participants' designs and ideas to support those goals. First, microaggressions are nebulous in nature, particularly in the South Korean culture, where the language does not even have a term for microaggression. Second, participants rejected approaches that would burden their social support network. Third, participants questioned the validity of their own thoughts due to the social perception that women are too emotional and sensitive. The first 2 cultural characteristics jointly influenced goals 1 (ie, having "shared" knowledge of microaggressions among all users) and 5 (ie, respecting the perpetrators' freedom of speech while educating them to understand their words' negative impact). The third cultural characteristic influenced goals 2 (ie, indicating positive support numerically), 3 (ie, responding with less emotion to microaggressions), 4 (ie, having forum-approved experts guide the discussions), and 6 (ie, knowing the "reality"; Figure 12).

XSL•FO
RenderX

Figure 12. Cultural characteristics (CC) affecting microaggression (MA) counteraction goals (G). CC1 is that the microaggressions are nebulous in nature, particularly in the South Korean culture, where the language does not even have a term for microaggression. CC2 is that participants rejected approaches that would burden their social support network. CC3 is that participants questioned the validity of their own thoughts due to the social perception that women are too emotional and sensitive. CC1 and CC2 jointly influenced G1 and G5. CC3 influenced G2, G3, G4, and G6.

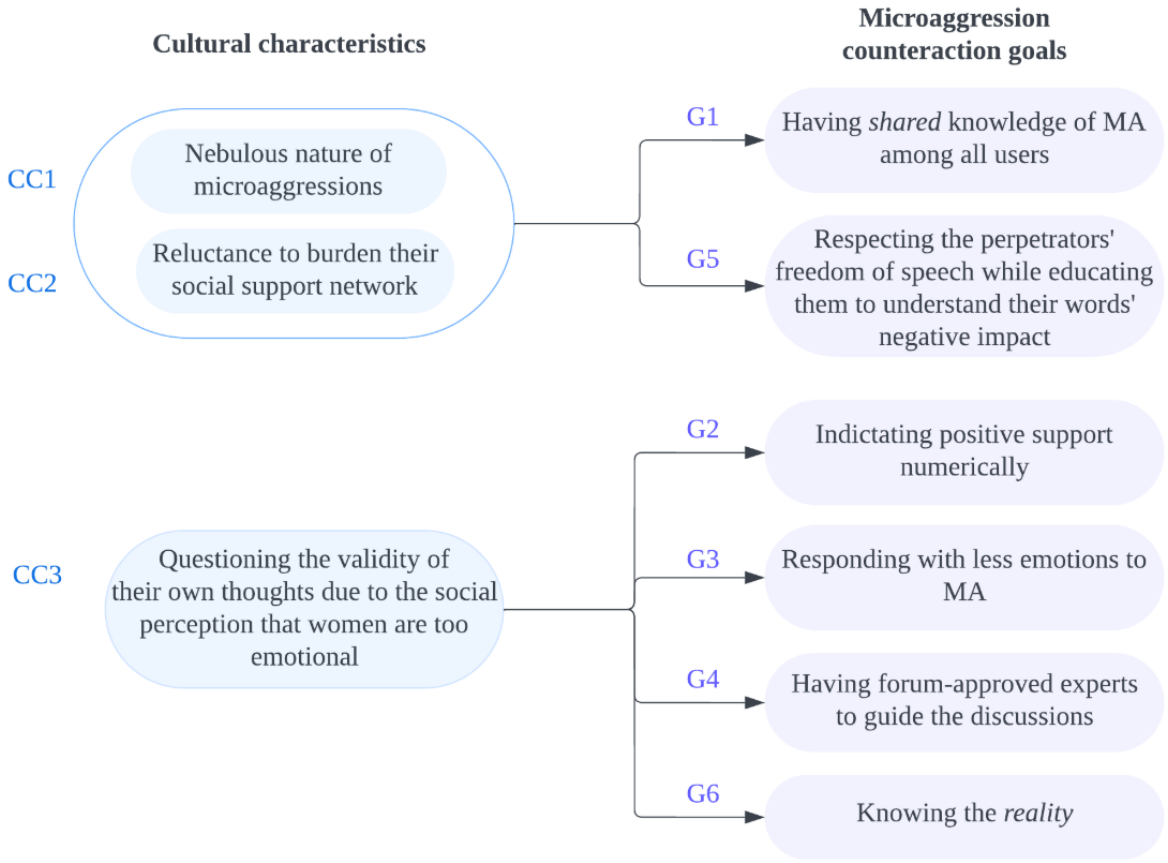

During the discussions for goal 1 (ie, having "shared" knowledge of microaggressions among all users), participants brought up that they did not want to burden their social support network, but due to the nebulous and subtle nature of microaggressions, they were inclined to have a shared knowledge of microaggressions among all users:

> *You might not know if you have just said already harmful words in your daily life, so I don't think it's a very terrible thing. Maybe it depends, but I don't think it should be taken that seriously. So it's okay to give some experience and time for people to get through the flags and learn through all the comments that they said that are harmful without removing them. Support the women who could be supporters!* [P7]

Participants were concerned that acquiring the knowledge of what microaggressions are and how they affect the targets would be cumbersome for male perpetrators who were not interested in putting in the time and effort to see from the target's perspective:

> *I would give all of them [male microaggression perpetrators] vaginas and uteruses and have them go get a pap smear done or understand the process of seeking SRH care so they can become more empathetic towards the situations and experiences that women go through every day.* [P6]

However, they thought that in the long term, it would help their social support network (ie, fellow female users who have the intent to help them) be less burdened to provide the support they were trying to offer:

> *This microaggression thing is hard to pinpoint and I just don't want to make everyone who is trying to help feel like it's too hard to help. I think that having a communal sense of knowledge would be better. I know that it's going to burden women who are only trying to help at first, so I'm kind of iffy about the idea, but I think in the long term, it would help much more [to counteract microaggressions].* [P7]

While ideating the design ideas to support goal 5 (ie, respecting the perpetrators' freedom of speech while educating them to understand their words' negative impact), participants further reflected on how making judgments about users with supportive intentions based on microaggressions which are so subtle in nature would encumber users who were merely trying to be supportive. They elucidated that in the Korean culture, support is sometimes provided in the form of hard-hearted advice rather than a warm and sympathetic embracement, which would often be associated with microaggressions. Although most participants

XSL•FO
RenderX

did not know what microaggressions were before this study, they acknowledged that the current delivery of support is not ideal and could be ameliorated:

> *You know how we [Koreans] can be. Words of affirmation are definitely not our [culture's] love language. We [Koreans] give it to you straight and don't want to just tell you things that will make you feel better at the moment. Although it might sound uncaring, giving you that push to snap out of it and solve your problems is our way of showing that we care. Though I agree that it doesn't always make things better. It sometimes just hurts, and it could be [delivered] better.* [P6]

During the discussions for goals 2, 3, 4, and 6, participants constantly stated that sociocultural context shaped them to question their own thoughts and depend on others' judgments to determine what is right for them. However, the same design-influencing cultural characteristic affected the designs of each goal-supporting features in different ways. For goal 3 (ie, responding with less emotion to microaggressions) supporters, the need to respond with less emotion to microaggressions stemmed from them wanting to be treated less as an emotional or sensitive person and their want to be treated as a person who can justify their opinions or decisions with logic:

> *[Unmarried Korean] women are just framed to be these emotional beings who are just too da** sensitive. In our society, whenever women bring up discomforts or problems, we're told that we're being sensitive and somewhat irrational, even by other women. I think we need the preset options for why a comment is a microaggression to be as coherent and rational as possible to help our case.* [P13]

Similarly, goal 2 (ie, indicating positive support numerically) and goal 6 (ie, knowing the "reality") advocates indicated unmarried Korean women users need to be more logical in their justifications of their thoughts. However, goal 2 and goal 6 supporters had different opinions on what was required for logical justifications. Goal 6 proponents claimed that to provide a logical rationale, they needed to comprehend the diverse perspectives that other users hold to better defend themselves:

> *I need to know what I'm dealing with as a whole to better understand how I should be taking something. Without the whole picture, how do I know that I am making a sensible judgment of what I am reading and the people who are writing what I am reading? After I see the complete picture, I will know who I agree with more and who I agree with less.* [P9]

On the other hand, goal 2 supporters argued that logical rationales for one's own thoughts could be supported by the number of support that a comment or claim has. Participants who advocated for goal 2 claimed that showing the number of negative responses to a comment would only make the questioning of their own thoughts worse rather than help them get the evidence they need to make themselves not be seen as overly emotional people:

> *I am uncomfortable with the idea of just having two options (thumbs-up button and thumbs-down button). Well, I'm more uncomfortable with the thumbs-down button. Seeing the number of thumbs-downs would just make me feel uneasy when I'm trying to make sure that what I'm thinking is right. It would just make me question my decisions when I don't have to. I shouldn't have to [question my decisions], but it's making me do that.* [P8]

In contrast to goals 2, 3, and 6 advocates, goal 4 (ie, having forum-approved experts guide the discussions) proponents expanded the process of finding logical justification for their own thoughts to include the opinions of verified experts. Goal 4 proponents' proposed ideas should not be evaluated as ideas that comply with the social perception that women are too emotional. Instead, they should be regarded as expanding the sources of knowledge they could use in strengthening their logical justifications, as shown in the following quote:

> *We just cannot do it all - not be sensitive or emotional and find all the medical, factual, and experiential knowledge. We need help, and that's where experts, of course those who are certified, could help.* [P12]

Discussion

Overview

In this section, we elucidate how the goals identified from the participants' co-design could be leveraged to build a more resilient and safe web-based space to discuss and seek support on SRH care issues. In the first part of our discussion, we describe how the goals that we identified from participants' co-design activities would also support known psychological strategies for coping with microaggressions. In the second part of the discussion, we detail how participants' goals support educational and explanatory approaches to counteracting microaggressions.

Coping Strategies for Web-Based Microaggressions

The combined use of reflective coping (ie, re-evaluating the meaning of a given situation to reduce its emotional impact [50]) and suppressive coping (ie, inhibiting of an emotional response [51]) have been proven to be effective in coping with microaggressions [50,51,56]. Participants' goals uncovered in the co-design sessions show how to achieve reflective and suppressive coping strategies in web-based microaggression counteraction measures.

Goals 2 (ie, indicating positive support numerically) and 6 (ie, knowing the "reality") supported the microaggression targets to achieve reflective coping. Individuals using reflective coping styles examine causal relationships, reflect, and engage in behaviors that are intended to produce changes in their affective states and cognitive processes [66-68]. Indicating positive support numerically and seeing the whole picture by understanding both sides was suggested by participants to help them reassess the given situation with an objective standard, which is pivotal to cognitive reappraisal or picturing the provocative situation from a third-person perspective [48,50]. By understanding how many people explicitly support an

XSL•FO
RenderX

opinion or argument, participants thought that they would be able to evaluate whose opinions they should trust. In addition, by seeing the whole picture and not just the positive support numerical indicators, they thought that they would not be making a judgment based on viewing only one side and be able to change their affective states and cognitive processes with a holistic understanding of the provocative situation [67,68].

For suppressive coping, participants suggested goal 3 (ie, responding with less emotion to microaggressions) with an added level of agency. In a notable study that investigated strategies to counteract web-based ableist microaggressions, the coping mechanisms for ableist microaggressions were found to align with traditional suppressive coping mechanisms (eg, deleting comments and letting it go, blocking rather than reporting, and taking a break from social media) [69]. However, in this study, we found that responding with less emotion to microaggressions did not necessarily mean to "avoid" dealing with microaggressions. Rather, suppressive coping was transformed into active responses to microaggressions but with less-emotive words. By using less emotion in response to microaggressions, either in private messages or through flagging options, participants thought their claims that a comment contained microaggression would be less subjective. Making their claim seem less subjective was pivotal because microaggression targets tend to blame themselves for being overly sensitive and worry about reinforcing negative stereotypes [10,42,44,47,48].

Thus, with tools to support the stated goals, we hope to expand the ways to incorporate reflective and suppressive coping mechanisms for the targets of web-based microaggressions. However, we advise designers to implement the participant-suggested designs with caution, especially for reflective coping–supporting designs. The goals that were identified to support reflective coping—goals 2 (ie, indicating positive support numerically) and 6 (ie, knowing the "reality")—were influenced by the cultural characteristic that unmarried Korean women are often exposed to situations where they question their own thoughts and value the public's account of their opinions because they are socially perceived to be overly emotional and sensitive. Thus, the implementation of tools to support reflective coping would need to be provided with constant exposure to microaffirmations: subtle verbal remarks and storytelling microaggression targets engage to affirm each other's dignity, integrity, and shared humanity [70]. Jones and Rolon-Dow [38] support the notion that constant exposure to microaffirmations helps targets understand what microaggressions and microaffirmations are, how they are perpetuated, and how they impact the target. With the benefits of repeated listening to microaffirmations, women would be able to achieve deeper reflective coping [38,70,71]. Targets of these microaggressions would be able to better construct a narrative about the microaggressions they experienced and develop more effective action plans to change their affective states and cognitive processes [38,71].

Educational and Explanatory Approaches to Counteracting Microaggressions to Increase Awareness of Microaggressions Among All Users

Goals 1 (ie, having shared knowledge among all users) and 5 (ie, respecting perpetrators' freedom of speech while educating them about the negative impact of their words) can be achieved by using approaches that focus on educating people about microaggressions and their harmful effects, as well as explaining why certain comments are considered microaggressions. By highlighting the harmful consequences of microaggressions, such approaches can help perpetrators evaluate the ramifications of their comments and reconsider causing future harm.

Explanations of what microaggressions are and their harmful impacts make current perpetrators aware of the problem and provide them with a chance to become proper allies by reexamining their own and others' roles and responsibilities [25,31,32]. Potential allies or perpetrators, especially those of the same gender who are experiencing similar oppressions or have overcome them, need to be more aware of their roles and responsibilities in reducing web-based in-group microaggressions. When people normalize oppression rather than acknowledging it, they unintentionally reinforce it with further microaggressions [10,22]. In contrast, when those who have overcome oppression offer empathy, support, or acknowledgment of the harm, they help others recover and learn [10,22]. As suggested by our study participants, we deem the rephrasing suggestions feature helpful in preventing unintentional harm or discouragement of support, even by those experiencing similar oppressions.

Support for encouraging perpetrators or potential allies to rethink their messages can be found in counter speech research [36-38,72]. Unlike content moderation, counter speech does not suppress free expression but aims to reduce hate through persuasion [37]. Counter speech involves 2 main approaches to responding to and deflecting hate speech. The first approach is a nonaggressive response to abusive content, deconstructing stereotypes and misinformation with thoughtful reasoning and fact-bound arguments [38,72]. The second approach, sometimes termed counternarratives, involves creating positive stories to encourage tolerance and disprove the assumptions underlying hate speech [36-38]. Counternarratives focus on triggering positive feelings, such as empathy for targets that could lead to attitude changes [36-38].

Participants were concerned that perpetrators might feel unappreciated and think moderation was unjust or frustrating, potentially leading to retaliation. While participants did not strongly advocate for counter speech explicitly, they thought educational explanations would support users "who have a good heart" [P12] to become allies and adopt rephrasing suggestions. This supports using counternarratives, which aim to foster empathy toward microaggression targets and prompt perpetrators to rethink their actions without feeling wronged.

Furthermore, goals 1 and 5 were influenced by participants' cultural characteristics, rejecting approaches that would burden their social support network. Participants' definition of social support includes current allies and microaggression perpetrators

XSL•FO
RenderX

who have supportive intentions. Participants noted that the fine line between harmful comments and hard-hearted advice can be hard to identify. Through thoughtful reasoning and positive narratives, counter speech can support unmarried people's extended support network, such as microaggression perpetrators with supportive intentions. Moreover, counter speech would help perpetrators feel empathy for targets and reconsider their delivery of support, which is currently difficult to achieve for "gray area" moderation cases [30,31].

Goal 4 (ie, having forum-approved experts guide the discussions) offers further insight into who should lead counternarratives. By having forum-approved medical experts and experienced users lead discussions, participants thought counternarratives could help re-evaluate factually wrong, disrespectful, or invalidating content. Leading critical and logical reactions to microaggressions by experienced users is supported by strategies for dealing with racial microaggressions [72]. Eschmann [72] suggested web-based counter spaces—where "activisty" students of color challenge dominant paradigms—highlight and critically react to experiences with microaggressions, providing ample support and counteraction for targets. In these counter spaces, "activisty" students lead critical reactions using counternarratives, similar to the "older sister" [P9] role suggested by our participants.

To nudge authors to revise or delete offensive text, recent work identifies offensive parts of a message [73] and suggests less-offensive alternatives through paraphrasing and style transfer [74,75]. Style transfer aims to translate abusive text into nonabusive while preserving its nonoffensive meaning [33,75,76]. Our study participants also advocated for low-effort approaches. This appreciation of reducing effort for potential allies is supported by other studies on web-based microaggression counteraction strategies [69,73]. Heung et al [69] recommended reactive approaches providing corrective information in response to microaggressions, similar to counternarratives [5,72]. However, participants also expressed concern about making microaggression perpetrators (ie, potential allies) feel unjustly attacked. Thus, public or private rephrasing through style transfer should be further studied to determine whether it could be interpreted as harmful in certain contexts.

Most approaches limit education and explanations to perpetrators, but our participants suggested that all users be educated on microaggressions, how to cope with or address them, and how to find and receive support after experiencing them. They recommended that explanations for flagged comments be made public and include pointers to support groups and resources. They also suggested that educational resources and support groups be added under the microaggression content alert feature for those who do not want to be exposed to microaggressions at all. Thus, social media platforms should extend their educational approaches to benefit *all* users and not just perpetrators.

Limitations and Future Directions

One potential limitation stems from the interviewer also being an unmarried Korean woman and the study participants being attained using snowball sampling. In investigating the design needs of the population that the interviewer is in and may have weak ties to, the participants may have felt a connection with the interviewer and enabled thoughtful responses. However, they may also have felt pressure to be portrayed as strong, independent women who did not need emotional support, especially when asked how they would alter or not alter their designs if they were the only post author users. Future studies should explore how having a design session facilitator who is not an unmarried Korean woman and does not have weak ties with the participants affects the participants' designs. In addition, we acknowledge that the findings of our study are based on a small number of participants. Although we reached saturation with 14 participants, the microaggression counteraction designs from this study could still be expanded upon. Thus, we encourage future studies to explore co-designing with a larger number of participants to reach further consensus on the microaggression counteraction designs.

Moreover, we suggest future studies delve into how the proposed microaggression counteraction designs could potentially be weaponized against those who are already susceptible. In this study, the focus was to co-design with unmarried Korean women to counteract the microaggressions suppressing their health care and to explore how the corrosive harm of microaggressions and social and cultural norms were reflected in their needs and preferences. We acknowledge the possibility of weaponization of the microaggression counteraction features. For example, automatically deleting a comment after a set number of negative reactions from the community may easily enable orchestrated actions to take nonharmful comments down. In addition, individualized messages to targets could be used to annoy the target if sent automatically by triggering multiple messages. Furthermore, an expert with a badge who writes microaggressions may encourage higher acceptance of the microaggressions. We suggest subsequent studies investigate how to protect people from the potential weaponization of their designs to ensure their safety.

Conclusions

In this study, we co-designed with unmarried Korean women participants with the aim to create a resilient and safe space for support on stigmatized SRH by preventing and counteracting web-based microaggressions that suppress women's health care. During the co-design sessions, we identified 6 goals participants had for counteracting web-based microaggressions that discourage SRH care support seeking. Connecting back to these goals, we also encourage educational, rather than the more common punitive detection and deletion designs that counteract microaggressions. We should help targets and allies cope with microaggressions, and help perpetrators reflect on the consequences of their words. With these design strategies and examples, we hope to encourage new moderation approaches within web-based communities. These approaches should support women in seeking essential care for stigmatized SRH, while reducing the occurrence and detrimental impact of microaggressions on their health and well-being.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the participants whose experiences formed the basis for this research and the University of Washington iMed Research Group for their insightful feedback on the development of the co-design session guide.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to them being transcripts with personally identifiable information and sensitive disclosures about their sexual and reproductive health support–seeking experiences.

Authors' Contributions

HR and WP were responsible for conceptualization. Data curation was done by HR. Formal analysis was done by HR and WP. WP was responsible for funding acquisition. Investigation was done by HR. HR and WP were responsible for methodology. HR was responsible for project administration, resources, and supervision. HR and WP were responsible for writing the original draft. HR was responsible for review and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Naseem A, Majed M, Abdallah S, Saleh M, Lirhoff M, Bazzi A, et al. Exploring Muslim women's reproductive health needs and preferences in the emergency department. West J Emerg Med 2023 Sep 13;24(5):983-992 [FREE Full text] [doi: 10.5811/westjem.58942] [Medline: 37788041]
2. Hall KS, Manu A, Morhe E, Dalton VK, Challa S, Loll D, et al. Bad girl and unmet family planning need among Sub-Saharan African adolescents: the role of sexual and reproductive health stigma. Qual Res Med Healthc 2018 May 30;2(1):55-64 [FREE Full text] [doi: 10.4081/qrmh.2018.7062] [Medline: 30556052]
3. Sychareun V, Vongxay V, Houaboun S, Thammavongsa V, Phummavongsa P, Chaleunvong K, et al. Determinants of adolescent pregnancy and access to reproductive and sexual health services for married and unmarried adolescents in rural Lao PDR: a qualitative study. BMC Pregnancy Childbirth 2018 Jun 08;18(1):219 [FREE Full text] [doi: 10.1186/s12884-018-1859-1] [Medline: 29884139]
4. Shin HY, Park B, Suh M, Choi KS, Jun JK. Association of late marriage and low childbirth with cervical cancer screening among korean women: results from a nationwide survey. Cancers (Basel) 2022 Jan 10;14(2):327 [FREE Full text] [doi: 10.3390/cancers14020327] [Medline: 35053489]
5. Lai RY, Choi SY. Premarital sexuality, abortion, and intergenerational dynamics in China. Modern China 2019 Oct 17;47(3):239-265. [doi: 10.1177/0097700419880524]
6. Tiwari A, Wu WJ, Citrin D, Bhatta A, Bogati B, Halliday S, et al. "Our mothers do not tell us": a qualitative study of adolescent girls' perspectives on sexual and reproductive health in rural Nepal. Sex Reprod Health Matters 2021;29(2):2068211 [FREE Full text] [doi: 10.1080/26410397.2022.2068211] [Medline: 35695251]
7. Lee HY, Lee MH. Barriers to cervical cancer screening and prevention in young Korean immigrant women: implications for intervention development. J Transcult Nurs 2017 Jul 18;28(4):353-362. [doi: 10.1177/1043659616649670] [Medline: 27193752]
8. Dong Y. "Premature" girls with "illegitimate" diseases: a qualitative exploration of the stigma of gynecological diseases for unmarried young women in China. J Health Commun 2023 Feb 01;28(2):111-120. [doi: 10.1080/10810730.2023.2185322] [Medline: 36855814]
9. Min JW, Cha HW. The effects of fear and social norms on unmarried women's intention to visit obstetricians and gynecologists: an application of the extended theory of planned behavior and interpersonal communication. Korean J Journalism Commun Stud 2017 Feb;61(1):217-250. [doi: 10.20879/kjjcs.2017.61.1.007]
10. Ryu H, Pratt W. Microaggression clues from social media: revealing and counteracting the suppression of women's health care. J Am Med Inform Assoc 2022 Jan 12;29(2):257-270 [FREE Full text] [doi: 10.1093/jamia/ocab208] [Medline: 34741511]
11. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
12. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
13. Sue DW. Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation. Hoboken, NJ: John Wiley & Sons; 2010.
14. Sue DW, Capodilupo CM, Torino GC, Bucceri JM, Holder AM, Nadal KL, et al. Racial microaggressions in everyday life: implications for clinical practice. Am Psychol 2007 May;62(4):271-286. [doi: 10.1037/0003-066x.62.4.271]

XSL•FO
RenderX

15. Wong-Padoongpatt G, Zane N, Okazaki S, Saw A. Decreases in implicit self-esteem explain the racial impact of microaggressions among Asian Americans. J Couns Psychol 2017 Oct;64(5):574-583. [doi: 10.1037/cou0000217] [Medline: 28425723]
16. Schmitt MT, Branscombe NR, Postmes T, Garcia A. The consequences of perceived discrimination for psychological well-being: a meta-analytic review. Psychol Bull 2014 Jul;140(4):921-948. [doi: 10.1037/a0035754] [Medline: 24547896]
17. Costa PL, McDuffie JW, Brown SE, He Y, Ikner BN, Sabat IE, et al. Microaggressions: mega problems or micro issues? A meta-analysis. J Community Psychol 2023 Jan 31;51(1):137-153. [doi: 10.1002/jcop.22885] [Medline: 35640014]
18. Nuru AK, Arendt CE. Not so safe a space: women activists of color's responses to racial microaggressions by White women allies. South Commun J 2018 Sep 26;84(2):85-98. [doi: 10.1080/1041794x.2018.1505940]
19. Keum BT, Brady JL, Sharma R, Lu Y, Kim YH, Thai CJ. Gendered racial microaggressions scale for Asian American women: development and initial validation. J Counsel Psychol 2018 Oct;65(5):571-585. [doi: 10.1037/cou0000305]
20. Aniebue UU, Onyeka TC. Ethical, socioeconomic, and cultural considerations in gynecologic cancer care in developing countries. Int J Palliat Care 2014 Jan 19;2014:1-6. [doi: 10.1155/2014/141627]
21. Shultz CG, Chu MS, Yajima A, Skye EP, Sano K, Inoue M, et al. The cultural context of teaching and learning sexual health care examinations in Japan: a mixed methods case study assessing the use of standardized patient instructors among Japanese family physician trainees of the Shizuoka Family Medicine Program. Asia Pac Fam Med 2015 Oct 7;14(1):8 [FREE Full text] [doi: 10.1186/s12930-015-0025-4] [Medline: 26451130]
22. David EJ, Derthick AO. The Psychology of Oppression. Cham, Switzerland: Springer; 2017.
23. Sue DW, Alsaidi S, Awad MN, Glaeser E, Calle CZ, Mendez N. Disarming racial microaggressions: microintervention strategies for targets, white allies, and bystanders. Am Psychol 2019 Jan;74(1):128-142. [doi: 10.1037/amp0000296] [Medline: 30652905]
24. Sultana S, Guimbretière F, Sengers P, Dell N. Design within a patriarchal society: opportunities and challenges in designing for rural women in Bangladesh. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3174110]
25. Ruths D, Saleem HM, Dillon KP, Wright L, Benesch S. Counterspeech on Twitter: a field study. Issue Lab. 2016. URL: https://www.issuelab.org/resources/34066/34066.pdf [accessed 2025-02-05]
26. Rogers R. Deplatforming: following extreme internet celebrities to telegram and alternative social media. Eur J Commun 2020 May 06;35(3):213-229. [doi: 10.1177/0267323120922066]
27. Yaqub W, Kakhidze O, Brockman ML, Memon N, Patil S. Effects of credibility indicators on social media news sharing intent. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376213]
28. Giansiracusa N. Social spread. In: How Algorithms Create and Prevent Fake News. Berkeley, CA: Apress; 2021.
29. Sharevski F, Devine A, Pieroni E, Jacnim P. Meaningful context, a red flag, or both? Preferences for enhanced misinformation warnings among US Twitter users. In: Proceedings of the 2022 European Symposium on Usable Security. 2022 Presented at: EuroUSEC '22; September 29-30, 2022; Karlsruhe, Germany. [doi: 10.1145/3549015.3555671]
30. Ma R, You Y, Gui X, Kou Y. How do users experience moderation?: a systematic literature review. Proc ACM Hum Comput Interact 2023 Oct 04;7(CSCW2):1-30. [doi: 10.1145/3610069]
31. Myers West S. Censored, suspended, shadowbanned: user interpretations of content moderation on social media platforms. New Media Soc 2018 May 08;20(11):4366-4383. [doi: 10.1177/1461444818773059]
32. Jhaver S, Chen QZ, Knauss D, Zhang AX. Designing word filter tools for creator-led comment moderation. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3517505]
33. Kiritchenko S, Nejadgholi I, Fraser KC. Confronting abusive language online: : a survey from the ethical and human rights perspective. J Artif Intell Res 2021 Jul 15;71:431-478. [doi: 10.1613/jair.1.12590]
34. Chancellor S, Lin Z, De Choudhury M. "This post will just get taken down": characterizing removed pro-eating disorder social media content. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 7-12, 2016; San Jose, CA. [doi: 10.1145/2858036.2858248]
35. Katsaros M, Yang K, Fratamico L. Reconsidering tweets: intervening during tweet creation decreases offensive content. Proc Int AAAI Conf Web Soc Media 2022 May 31;16:477-487. [doi: 10.1609/icwsm.v16i1.19308]
36. Hangartner D, Gennaro G, Alasiri S, Bahrich N, Bornhoft A, Boucher J, et al. Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. Proc Natl Acad Sci U S A 2021 Dec 14;118(50):e2116310118 [FREE Full text] [doi: 10.1073/pnas.2116310118] [Medline: 34873046]
37. Baider F. Accountability issues, online covert hate speech, and the efficacy of counter‐speech. Politics Gov 2023 Mar 20;11(2). [doi: 10.17645/pag.v11i2.6465]
38. Jones JM, Rolón-Dow R. Multidimensional models of microaggressions and microaffirmations. In: Torino GC, Rivera DP, Capodilupo CM, Nadal KL, Sue DW, editors. Microaggression Theory: Influence and Implications. Hoboken, NJ: John Wiley & Sons; 2018.
39. Thach H, Mayworm S, Delmonaco D, Haimson O. (In)visible moderation: a digital ethnography of marginalized users and content moderation on Twitch and Reddit. New Media Soc 2022 Jul 18;26(7):4034-4055. [doi: 10.1177/14614448221109804]

XSL•FO
RenderX

40. Jhaver S, Bruckman A, Gilbert E. Does transparency in moderation really matter?: user behavior after content removal explanations on Reddit. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-27. [doi: 10.1145/3359252]
41. Mayworm S, Li S, Thach H, Delmonaco D, Paneda C, Wegner A, et al. The online identity help center: designing and developing a content moderation policy resource for marginalized social media users. Proc ACM Hum Comput Interact 2024 Apr 26;8(CSCW1):1-30. [doi: 10.1145/3637406]
42. Johnson VE, Nadal KL, Sissoko DR, King R. "It's not in your head": gaslighting, 'splaining, victim blaming, and other harmful reactions to microaggressions. Perspect Psychol Sci 2021 Sep;16(5):1024-1036. [doi: 10.1177/17456916211011963] [Medline: 34498522]
43. David EJ, Derthick AO. What is internalized oppression, and so what? In: Internalized Oppression: The Psychology of Marginalized Groups. Berlin, Germany: Springer; 2013.
44. Sue DW, Capodilupo CM, Holder AM. Racial microaggressions in the life experience of Black Americans. Prof Psychol Res Pract 2008;39(3):329-336. [doi: 10.1037/0735-7028.39.3.329]
45. Holder AM, Jackson MA, Ponterotto JG. Racial microaggression experiences and coping strategies of Black women in corporate leadership. Qual Psychol 2015;2(2):164-180. [doi: 10.1037/qup0000024]
46. Lewis JA, Mendenhall R, Harwood SA, Browne Huntt M. Coping with gendered racial microaggressions among Black women college students. J Afr Am Stud 2012 Jun 21;17:51-73. [doi: 10.1007/s12111-012-9219-0]
47. Nadal KL, Hamit S, Lyons O, Weinberg A, Corman L. Gender microaggressions: perceptions, processes, and coping mechanisms of women. In: Psychology for Business Success, Vol. 1. Juggling, Balancing, and Integrating Work and Family Roles and Responsibilities, Vol. 2. Institutional Equity and Compliance, Vol. 3. Managing, Leading, and Developing Employees, Vol. 4. Implementing Best Practices in Human Resources. Santa Barbara, CA: ABC-Clio; 2013.
48. Sue DW, Lin AI, Torino GC, Capodilupo CM, Rivera DP. Racial microaggressions and difficult dialogues on race in the classroom. Cultur Divers Ethnic Minor Psychol 2009 Apr;15(2):183-190. [doi: 10.1037/a0014191] [Medline: 19364205]
49. Wilson TK, Gentzler AL. Emotion regulation and coping with racial stressors among African Americans across the lifespan. Dev Rev 2021 Sep;61:100967. [doi: 10.1016/j.dr.2021.100967]
50. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol 2003 Aug;85(2):348-362. [doi: 10.1037/0022-3514.85.2.348] [Medline: 12916575]
51. Juang LP, Moffitt U, Kim SY, Lee RM, Soto JA, Hurley E, et al. Cognitive reappraisal and expressive suppression: links to racial-ethnic discrimination and adjustment among Latino/a and Asian-heritage college students. J Adolesc 2016 Dec;53:21-33 [FREE Full text] [doi: 10.1016/j.adolescence.2016.08.012] [Medline: 27598799]
52. Park IJ, Kim PY, Cheung RY, Kim M. The role of culture, family processes, and anger regulation in Korean American adolescents' adjustment problems. Am J Orthopsychiatry 2010 Apr;80(2):258-266 [FREE Full text] [doi: 10.1111/j.1939-0025.2010.01029.x] [Medline: 20553519]
53. Gong L, Li W, Zhang D, Rost DH. Effects of emotion regulation strategies on anxiety during job interviews in Chinese college students. Anxiety Stress Coping 2016 May;29(3):305-317. [doi: 10.1080/10615806.2015.1042462] [Medline: 25930079]
54. Su JC, Park IJ, Chang J, Kim SY, Dezutter J, Seol KO, et al. Differential Links Between Expressive Suppression and Well-Being Among Chinese and Mexican American College Students. Asian Am J Psychol 2015 Mar;6(1):15-24 [FREE Full text] [doi: 10.1037/a0036116] [Medline: 34567433]
55. Cheung RY, Park IJ. Anger suppression, interdependent self-construal, and depression among Asian American and European American college students. Cultur Divers Ethnic Minor Psychol 2010 Oct;16(4):517-525 [FREE Full text] [doi: 10.1037/a0020655] [Medline: 21058815]
56. Hernández RJ, Villodas MT. Overcoming racial battle fatigue: the associations between racial microaggressions, coping, and mental health among Chicana/o and Latina/o college students. Cultur Divers Ethnic Minor Psychol 2020 Jul;26(3):399-411. [doi: 10.1037/cdp0000306] [Medline: 31556630]
57. Alsheikh T, Rode JA, Lindley SE. (Whose) value-sensitive design: a study of long-distance relationships in an Arabic cultural context. In: Proceedings of the ACM 2011 conference on Computer supported cooperative work. 2011 Presented at: CSCW '11; March 19-23, 2011; Hangzhou, China. [doi: 10.1145/1958824.1958836]
58. Rabaan H, Young AL, Dombrowski L. Daughters of men: Saudi women's sociotechnical agency practices in addressing domestic abuse. Proc ACM Hum Comput Interact 2021 Jan 05;4(CSCW3):1-31. [doi: 10.1145/3432923]
59. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008;4(1):5-18. [doi: 10.1080/15710880701875068]
60. Kleinsmann M, Valkenburg R. Barriers and enablers for creating shared understanding in co-design projects. Des Stud 2008 Jul;29(4):369-386. [doi: 10.1016/j.destud.2008.03.003]
61. Steen M. Co-design as a process of joint inquiry and imagination. Des Issues 2013;29(2):16-28. [doi: 10.1162/DESI_a_00207]
62. Dorst K. The core of ‘design thinking’ and its application. Des Stud 2011 Nov;32(6):521-532. [doi: 10.1016/j.destud.2011.07.006]
63. Bertran FA, Pometko A, Gupta M, Wilcox L, Banerjee R, Isbister K. Designerly tele-experiences: a new approach to remote yet still situated co-design. ACM Trans Comput Hum Interact 2022 Oct 20;29(5):1-30. [doi: 10.1145/3506698]

XSL•FO
RenderX

64. Vasquez R. You just need more resilience: racial gaslighting as “othering”. J Crit Thought Prax 2022;11(3). [doi: 10.31274/jctp.12586]
65. Lee J, Kim H, Woo J, Chang SM, Hong JP, Lee DW, et al. Impacts of remaining single above the mean marriage age on mental disorders and suicidality: a nationwide study in Korea. J Korean Med Sci 2020 Sep 21;35(37):e319 [FREE Full text] [doi: 10.3346/jkms.2020.35.e319] [Medline: 32959544]
66. Heppner PP, Cook SW, Wright DM, Johnson WC. Progress in resolving problems: a problem-focused style of coping. J Counsel Psychol 1995;42(3):279-293. [doi: 10.1037//0022-0167.42.3.279]
67. Wei M, Heppner PP, Ku TY, Liao KY. Racial discrimination stress, coping, and depressive symptoms among Asian Americans: a moderation analysis. Asian Am J Psychol 2010;1(2):136-150. [doi: 10.1037/a0020157]
68. Pennebaker JW. Telling stories: the health benefits of narrative. Lit Med 2000;19(1):3-18. [doi: 10.1353/lm.2000.0011] [Medline: 10824309]
69. Heung S, Phutane M, Azenkot S, Marathe M, Vashistha A. Nothing micro about it: examining ableist microaggressions on social media. In: Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility. 2022 Presented at: ASSETS '22; October 23-26, 2022; Athens, Greece. [doi: 10.1145/3517428.3544801]
70. Bardzell S. Feminist HCI: taking stock and outlining an agenda for design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; April 10-15, 2010; Atlanta, GA. [doi: 10.1145/1753326.1753521]
71. Pérez Huber L, Robles G, Solórzano DG. “Life was brought back into my body”: a Critical Race Feminista analysis of racial microaffirmations. Race Ethn Educ 2023 Jan 20;26(6):701-718. [doi: 10.1080/13613324.2023.2165514]
72. Eschmann R. Digital resistance: how online communication facilitates responses to racial microaggressions. Sociol Race Ethn 2020 Jun 25;7(2):264-277. [doi: 10.1177/2332649220933307]
73. Carton S, Mei Q, Resnick P. Extractive adversarial networks: high-recall explanations for identifying personal attacks in social media posts. arXiv. Preprint posted online on September 1, 2018 2025. [doi: 10.48550/arXiv.1809.01499]
74. Prabhumoye S, Tsvetkov Y, Salakhutdinov R, Black AW. Style transfer through back-translation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018 Presented at: ACL 2018; July 15-20, 2018; Melbourne, Australia. [doi: 10.18653/v1/p18-1080]
75. Tran M, Zhang Y, Soleymani M. Towards a friendly online community: an unsupervised style transfer framework for profanity redaction. In: Proceedings of the 28th International Conference on Computational Linguistics. 2020 Presented at: COLING 2020; December 8-13, 2020; Online. [doi: 10.18653/v1/2020.coling-main.190]
76. dos Santos CN, Melnyk I, Padhi I. Fighting offensive language on social media with unsupervised text style transfer. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2018 Presented at: ACL 2018; July 15-20, 2018; Melbourne, Australia. [doi: 10.18653/v1/p18-2031]

Abbreviations

SRH: sexual and reproductive health
STI: sexually transmitted infection

Edited by A Mavragani; submitted 03.06.24; peer-reviewed by T Albritton, N Levitz; comments to author 27.09.24; revised version received 28.10.24; accepted 21.12.24; published 26.02.25.

Please cite as:
Ryu H, Pratt W
Women’s Educating and Coping Strategies for Cultivating Supportive Web-Based Spaces for Discussing Sexual and Reproductive Health: Co-Design Study
J Med Internet Res 2025;27:e62716
URL: https://www.jmir.org/2025/1/e62716
doi:10.2196/62716
PMID:

©Hyeyoung Ryu, Wanda Pratt. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining BRCA Previvors' Social Media Content Creation as a Form of Self and Community Care: Qualitative Interview Study

Mariah L Wellman[1], PhD; Camilla M Owens[2], MA; Avery E Holton[2], PhD; Kimberly A Kaphingst[3], SCD

[1]Advertising & Public Relations, College of Communication Arts and Sciences, Michigan State University, East Lansing, United States

[2]Department of Communication, University of Utah, Salt Lake City, UT, United States

[3]Department of Communication, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Mariah L Wellman, PhD
Advertising & Public Relations
College of Communication Arts and Sciences
Michigan State University
404 Wilson Road
East Lansing, 48824
United States
Phone: 1 5173553410
Email: mwellman@msu.edu

Abstract

Background: Genetic testing has become a common way of identifying a woman's risk of developing hereditary breast and ovarian cancer; however, not all medical providers have the necessary information to support patients interested in genetic testing, nor do they always have the proper information for patients once they have been diagnosed. Therefore, many "previvors"—the name given to those who have tested positive for the *BRCA* genetic mutation—have taken to social media to inform others about the importance of genetic testing and explain to them how to understand their test results. Historically, those desiring to speak about their medical issues online have sought out structured support groups or chat rooms; however, many previvors today are instead posting on their own personal social media accounts and creating more niche communities.

Objective: This study aimed to examine why *BRCA* previvors are sharing content on their personal social media accounts and how posting online in this way serves a purpose for their larger community.

Methods: A total of 16 semistructured interviews were conducted with individuals who posted about their experience being diagnosed with the *BRCA* genetic mutation and their subsequent treatment on their personal social media accounts, specifically for followers interested in their medical journey. The interviews were recorded, transcribed, and coded by an experienced qualitative researcher and a graduate student using inductive techniques, and a reflexive thematic analysis was applied to the transcripts.

Results: The results suggest *BRCA* previvors want to control the narrative around their personalized medical experiences rather than participating in existing groups or chat rooms. Controlling their own story, rather than adding to existing narratives, gives previvors a sense of control. It also allows them to set boundaries around the types of experiences they have online when sharing their medical journey. Finally, previvors said they feel they are serving the larger *BRCA* community by each sharing their individual journeys, to hopefully avoid stereotyping and homogenizing the experience of patients with *BRCA* genetic mutations.

Conclusions: Research with the objective of understanding the experiences of *BRCA* previvors should include exploring how and why they talk about their journeys, especially due to the lack of knowledge *BRCA* previvors say many of their medical providers have. We suggest further research should examine how other patients with the *BRCA* genetic mutation, especially racial and ethnic minority patients, are navigating their own content creation, especially considering content moderation policies that social media platforms are continuing to implement that directly impact users' ability to share about their medical experiences.

(J Med Internet Res 2025;27:e67794) doi:10.2196/67794

KEYWORDS

BRCA; breast cancer; genetic testing; social media; breast cancer gene; content creation; self care; community care; qualitative interview; qualitative; interview; previvors; cancer previvors; genetic mutations; online; content; interviews; thematic analysis

XSL•FO
RenderX

Introduction

As the world's most commonly diagnosed cancer, the urgency to address breast cancer and support those with an increased risk of developing this disease has become even more important [1]. In 2020, the World Health Organization reported 2.3 million globally diagnosed cases of breast cancer, along with 685,000 deaths attributed to the disease [1]. In many cases, breast cancer occurs sporadically. However, genetic factors do play an influential role in 10% to 15% of all cases [2]. Like breast cancer, the risk of ovarian cancer also increases in women with age, along with having inherited genetic traits such as a *BRCA1* or *BRCA2* gene mutation [3]. In 2020, ovarian cancer was the third most prevalent and lethal gynecological cancer worldwide [4-6].

With continued research and treatment of breast and ovarian cancer, it has become more common for medical providers to encourage genetic testing, especially for those whose relatives who have been diagnosed with cancer [7-9]. Women who test positive for a *BRCA* genetic mutation are at a higher risk of developing hereditary breast and ovarian cancer, which impacts their health, reproductive choices, and identity [10,11]. Within their lifetime, individuals with a *BRCA* mutation have up to a 75% increased risk in developing breast or ovarian cancer [12]. Those who do test positive for a *BRCA* genetic mutation and have not yet been diagnosed with cancer are known as "previvors," a distinction for those at higher risk for developing cancer [10].

Due to the aggressive nature and reduced life expectancy associated with breast and ovarian cancer, it is necessary that the public, especially those who have a family history of breast and ovarian cancer and those who have already tested positive for the genetic mutation, are informed about the risks associated with the mutation that they carry and its impact on one's health and family planning [7]. However, previously published research suggests not all medical providers have the information previvors are looking for [7]. Previvors are looking for medical information and resources from various in-person and online sources regarding the options available to them should they desire to undergo surgery and reconstruction, just undergo surgery, or simply monitor themselves over time [11]. However, they are also looking for emotional support from loved ones, especially those who also have this genetic mutation and can provide advice and social support [7,11]. Social media and the ability to create content is a convenient and valuable tool for such purposes, and thus, researchers are exploring how social media is used by those with a *BRCA* genetic mutation [7,11,13].

BRCA previvors, cancer survivors, and those living with cancer use social media, online blogs, and internet chat rooms to connect with others [7,11,14,15]; however, much of the research on these communities is done through the lens of those seeking information, support, and connection rather than creating it [16,17]. When patients first began searching for information and for others in similar situations, they often did so through online platforms that afford a built-in community like chat rooms or message boards [15]. More recently, however, patients are using their own personal social media accounts to tell their stories on an individual level, hoping to build community on their own terms [7,11,13]. Individuals with breast and ovarian cancer communicate information that is personal or social on their individual accounts, with two-thirds of posts conveying actual experiences or providing support to others [18]. Using social media in this way distracts users from the stresses they may be experiencing that are caused by new, recurring, or terminal illnesses [19].

As a shift has occurred from group-centered platforms to individual storytelling spaces, there remain significantly fewer social media posts engaging with *BRCA* and genetic testing [11]. However, there are social media users increasing the amount of *BRCA* content being created, and health communication and internet scholars must pay attention to these individuals and their content [7,11]. As mentioned earlier, most research on social media and *BRCA* previvors has examined how previvors find information, rather than create it themselves, although the lens through which scholars are examining this content is shifting [7,11]. Considering the limited information available, this study aims to add qualitative insights regarding the creation of content by and for *BRCA* previvors. To accomplish this, we conducted semistructured interviews with *BRCA* previvors who create social media content related to their health condition and identity as previvors. Trends and themes were analyzed across the interview transcripts regarding *BRCA* previvors, social media, content creation, and connection within digital spaces. At the start of this study, we asked the following research questions: (1) Why do *BRCA* previvors create content on their personal social media accounts rather than in the digital communities created for *BRCA* previvors such as Facebook (Meta) groups? and (2) How do *BRCA* previvors perceive their social media content creation serving a purpose to their larger community?

Through qualitative interviews with a selective group of *BRCA* previvors who post personal medical experiences on their own social media accounts, this study offers a novel perspective on the trend of sharing health information online. Through prioritizing the words of our participants, we believe this study provides insight into the thoughts, feelings, and experiences of individual *BRCA* previvors and can help researchers and health care providers understand the ongoing needs of patients with *BRCA* genetic mutations.

Methods

Recruitment and Data Collection

To examine *BRCA* previvors' experiences creating social media content for themselves and for their larger community, we used a qualitative interview approach. We prioritized recruiting individuals living in the United States who have been found to carry a *BRCA1* or *BRCA2* genetic mutation, resulting in an increased risk for developing breast or ovarian cancer, and publicly shared about their medical experiences on social media, specifically Instagram (Meta). First, the primary investigator conducted a purposive sample by searching for *BRCA* previvors active on Instagram through hashtags including "#brca," "#brca1," "#previvor," "#brcagene," and "#breastcancerprevivor." Once the accounts were populated,

XSL•FO
RenderX

the investigator reached out to potential participants who listed their email addresses publicly on their Instagram accounts. We acknowledge that the algorithm on social media platforms like Instagram may have played a role in our sample procedure; however, we were interested in interviewing individuals who acted as influencers, opinion leaders, and public figures in *BRCA* online communities and, therefore, recognized that the most popular accounts may be the most accessible by using this purposive sampling method. Upon finding these individual accounts with publicly listed email addresses, the primary researcher sent an email detailing the study objectives and requesting potential participants to respond if interested. Of the total study participants (N=16), 15 identified as women and 1 identified as nonbinary. The nonbinary participant noted that they are often perceived as female and check "woman" as the gender on their medical paperwork for insurance purposes. All participants were White presenting; however, 2 identified as Hispanic and 2 identified as Ashkenazi Jewish. The latter group noted their ethnicity as important to their *BRCA* status, as 1 in 40 Ashkenazi Jewish women have a *BRCA* genetic mutation [20]. Participants were all over the age of 18 years at the time of the interview, and all confirmed that they are actively posting about their *BRCA* experiences publicly on one or more social media platforms.

Participants who met all eligibility criteria were then scheduled for a virtual, semistructured interview over Zoom (Zoom Video Communications). Consent forms were signed before the interview was scheduled, and verbal consent was confirmed once again before the interview began. All interviews were conducted by the primary investigator, and each participant was provided the opportunity to turn their camera off before the interview began to further protect their identity. The interviews included questions related to the participants' experiences with *BRCA*, their desire to create content on their personal social media accounts, their interactions with the *BRCA* community, their understanding of their previvor identity, and their relationship with health care providers. Each interview lasted between 45 and 60 minutes. The interviews were audio recorded for accuracy, and upon completion of the discussion, each interview audio file was labeled with a participant number. After all personal identifying information was removed from the files and the audio recordings, they were transcribed by a professional transcription service.

Data Analysis

Reflexive thematic analysis [21,22] was applied by the principal investigator and supported by a graduate research assistant. This analysis method has been applied in recent studies of women with *BRCA* genetic mutations and how they share social media content, build community, and understand their identity as *BRCA* previvors [7,10,11]. First, the principal investigator and graduate research assistant split up the 16 transcripts and manually coded 8 each. They then met and discussed overlapping patterns in the data related to why *BRCA* previvors post on social media about their previvor experiences and the role content creation plays in the act of community building and caring for oneself through various *BRCA* experiences. After the first meeting, the principal investigator and graduate research assistant switched transcripts and coded the other set of 8 transcripts. They then met a second time to finalize the themes present that related to the initial research questions. The team elected to follow this 2-stage coding process recommended by previous researchers to maintain the rigor of the method. All direct quotes provided in the results section below are anonymized and referred to by participant number.

Ethical Considerations

As this study involved human participants, the University of Utah's institutional review board approved the study protocol (IRB_00144720) before we began. An informed consent document was sent to each participant for review before they elected to participate. The document described the nature of the study, the types of questions asked, the protocol researchers followed to protect participant anonymity, and information regarding their rights as research participants including the opportunity to decline to answer any question and stop the interview at any time. The primary researcher kept a password-protected document on an external hard drive, which included the demographic information of each participant, including their name and participant number, to protect their identity but provided a reference sheet for use when needed. All personal identifying information was removed from the files and the audio recordings. The participants who completed the interview process were compensated in the form of a US $50 gift card.

Results

Posting on Social Media Fills a Gap in the Narrative of BRCA Previvorship

The *BRCA* previvors we interviewed said they create content on their personal social media accounts rather than in digital communities, such as Facebook groups, because they like to be able to control the narrative around their own personal journey and fill a gap in society's understanding of what it means to be someone living with a *BRCA* genetic mutation. *BRCA* previvors desire connection beyond existing previvor and survivor communities, and many noted they create content because they want to share every step in their journey with others, regardless of whether those watching have *BRCA* genetic mutations or not. One previvor explained that in her opinion, nothing is off limits when you share about *BRCA* on your own personal page.

> *I've shared every step of the way and there's nothing, in my opinion, that I shouldn't share.* [P11]

Similarly, many of the previvors interviewed have shared in-depth videos and photos of surgery scars, drain removals, breast reconstruction, nipple tattoos, and more.

Sharing whatever they choose, whenever they choose results in previvors feeling a deeper sense of autonomy over their lives and control over when and how they talk about *BRCA* genetic mutations. They do not want to talk strictly about the clinical aspects of their genetic mutation; they want to talk about sex, desire, hormones, body image, and how testing positive for *BRCA* genetic mutations impacts their identity as a wife, daughter, sister, mother, and friend. One previvor explained that she wanted other women with *BRCA* genetic mutations who follow her to understand that they may change, both inside

XSL•FO
RenderX

and out, but they can still be the same woman they were before being diagnosed and they do not have to allow their *BRCA* genetic mutation to take away who they once were.

> *I took people on my journey to show them that you can be strong enough to go through it and come out on top of it and no matter what you look like, it's still you.* [P14]

When discussing *BRCA* genetic mutations from the clinical perspective, previvors say they sometimes feel like their condition is overgeneralized, which results in feelings of homogeneity, especially in the eyes of their health care providers. In previous studies, previvors noted some providers make assumptions about the care they desire and do not always leave space for individuals to request a treatment plan that works best for them [7]. One previvor shared her experience defending her decision to forego reconstructive surgery in favor of a flat chest, something she did not know was possible from only speaking to her providers. She found other options through social media and now posts on her Instagram account to help others.

> *I hope people see (my chest) and they're like 'Oh, we can go so many different directions for our BRCA decisions. We don't just have to have reconstruction or like just a flat chest like'—there's things where you can make it feel more like you or learning how to make your body feel like yours.* [P15]

Showing a personalized perspective of the previvor experience on social media, previvors say, offers others a chance to see what daily living with *BRCA* genetic mutations is like, especially for those diagnosed at a young age. Previvors interviewed noted that many young patients with *BRCA* genetic mutations have different goals regarding their lifestyle, and Facebook groups and online chat rooms do not reflect those previvors and their understanding of *BRCA* genetic mutations. One previvor noted:

> *Facebook is just a different type of platform; people are posting pictures or questions or lamenting some situation. Whereas Instagram, it's more of a story of my life and not just a way to complain. So, I do think it was more helpful to be able to control the content on Instagram and not have anybody disqualify my experience. I think that was my biggest irritation with the Facebook groups. I felt like I was being questioned for whatever decision I was leaning towards. And this way I could just be like, 'This is what I'm doing. These are my risks. Just wanted you to know.' Not open for comment or questions or your opinions.'* [P17]

Younger previvors have turned to Instagram and TikTok (ByteDance) to find others with *BRCA* genetic mutations who have similar experiences and opinions regarding living with a genetic mutation and seeking possible treatments. For example, many younger previvors do not want to take the surgical route immediately after being diagnosed, as they are concerned about infertility, losing the opportunity to breastfeed a future child, and how their future may be dictated by decisions they are making from a young age. One previvor shared that she began posting because she felt like nobody was discussing *BRCA* genetic mutations from the perspective of a younger woman, especially a woman who wanted to have kids one day.

> *I started relying on social media to not only like see what other people were saying, especially people my age, but then being able to start talking about it from my personal perspective because I know my friends especially before 30 are like not thinking about (having kids) yet.* [P18]

For this previvor, posting on her social media accounts specifically on what young previvors need to be thinking about is important because she may be able to help another young patient with *BRCA* genetic mutations make a decision that was not offered by their health care provider.

Previvors View Content Creation as an Act of Self and Community Care

For *BRCA* previvors, creating content on their own social media accounts focused on their experiences helps them build a sense of autonomy over having a *BRCA* genetic mutation while taking care of themselves and offering support to others with a genetic mutation. This act of self-care and care for the community was a common explanation for why our interviewees started posting on their personal social media accounts. As one previvor said:

> *I wanted women to see that they're not the only ones going through it, that there are other women out there and that it will be okay and that you're beautiful even with all the scars and that you're still a woman.* [P14]

Many previvors said their understanding of themselves as a woman shifted once they were diagnosed, and they believe this shift is common among patients with *BRCA* genetic mutations. *BRCA* previvors expressed feeling afraid when initially diagnosed and unsure of what the future would hold for them as women, mothers, wives, daughters, and friends. Posting on social media allowed them to discuss their understanding of womanhood after diagnosis, which became an act of self-care and a way to give back to others in their community. Later in the interview, the same previvor shared how she felt upon being diagnosed and how her social media content creation helped her feel more like herself.

> *For a long time, you don't know who you are anymore. BRCA is—you were dealt this card, and you try to deal with it, but everything that makes you a woman is taken away from you. You don't know—it's like I know my body or my breasts or my ovaries don't define me as a person or a woman, but that is what makes us a woman and when that's taken away from you, you have to kind of learn to love yourself again in a different way.* [P14]

Posting on social media provided an outlet for previvors while they learned to love their new bodies, minds, and spirits after diagnosis. In the interviews, some participants said they felt like their *BRCA* genetic mutation initially stole their voice and that they no longer had a say over their own lives. However, sharing on social media gave them the opportunity to speak about whatever they were going through whenever they felt like they needed to. For one previvor, the posting was a form of therapy.

XSL•FO
RenderX

> *For me, it was almost like a journal, kind of therapeutic to put it out there in the world and be very open about it and then have—I was surprised how many of my friends and people that follow me that have had breast cancer, how supportive they were. I really expected them to be like the nursing staff, like 'Oh, you don't have real cancer,' but they were all super supportive.* [P03]

The therapeutic nature of posting on individual social media accounts was more helpful for these creators than sharing in chat rooms or Facebook groups, many said. Previvors felt like some chat rooms and online groups have become spaces where toxic positivity runs rampant but, at the same time, can be a place where previvors are judged by others regarding their care and treatment decisions. In the interviews, some previvors described the breast cancer survivor community, and more recently, the previvor community, as a space where media and large organizations "pink-wash" the disease and, thus, take away from the authentic lived experiences of breast cancer survivors and previvors [23]. These actions, while they can result in financial gain for some, do not provide the kind of support many previvors say they are looking for. Another previvor explained that the lack of authentic experiences publicized through media outlets and large organizations does a disservice to all people.

> *There was just a lack of representation and of candid, honest representation. I wanted someone that I could relate to. It wasn't just like this sob story because like parts of it were funny and weird and so full of love. I know that my story is not everyone else's, but like somebody's going to be able to relate to that, right? And also…(I wanted to) talk about sex and like all these things that I'm so used to talking about openly, but other people don't. So, I was like 'All right, I'll just write it all. Hopefully, it will help somebody.'*

Others agreed, saying:

> *it's so amazing to have the chance to share your voice, no matter what it be about," and "I post the ugly. I post the good and the bad. I want people to know there is the ugly part of it, but there can be a good part to it too.* [P16]

Ultimately, previvors agreed that sharing on their personal social media accounts is a form of community care and a way to continue building a supportive community. As one previvor noted:

> *I can take the fear that some women have, that I can show them that 'You can do it. You will come out on top of this. Don't let the BRCA run your life. You run BRCA. You can do something'* [P04]

Previvors on social media want others to know they are not alone and that there are spaces where they can be exactly who they are without judgment, even if they must create it themselves. As one previvor summarized:

> *I think if we keep (our journeys) to ourselves, then we're not doing any good* [P14]

Discussion

Principal Findings

Studies describing the content creation of *BRCA* previvors are a relatively understudied area of health communication and medical internet research. The previvors we interviewed noted that they post on their personal social media accounts rather than online spaces created specifically for group communication like chat rooms and Facebook groups because it allows them to control the narrative around their *BRCA* experience. This finding provides greater nuance to research that claims when patients interact online with others, they feel a sense of social support [7,11]. In addition, this finding provides a greater understanding of those individuals who not only seek health information online [16,17] but also create it themselves.

For previvors, controlling the narrative allowed them to speak on the topics they deemed most important and gave them the power to remove or block those who judged or critiqued their choices. This gave our participants a greater sense of power over their situation and reduced their fears around the possibility of pushback from other previvors. Some said they received negative responses from others with *BRCA* genetic mutations in spaces like Facebook groups but did not receive that feedback on their own social media accounts. *BRCA* previvors also believed that posting on their personal social media accounts served a greater purpose beyond themselves. They envisioned their content to serve the larger *BRCA* community in their quest to take care of each other, provide guidance regarding treatment plans and surgery options, and increase the visibility of *BRCA* genetic mutations itself. This finding extends previous studies that posit the discourse within online spaces has the potential to extend offline and impact the care patients receive [7,11,13,16,17].

While our study participants did not bring up any potential risks or negative consequences to sharing their personal health stories online, it is important to consider privacy concerns when one shares personal health information on social media platforms [24,25]. In addition, there is potential for misinformation to spread in these online health spaces, especially as patients continue to share their experiences on individual accounts that are more difficult to monitor [26,27]. While our participants did not mention misinformation as being common within *BRCA* spaces, as the communities grow, the issue will be important for researchers to consider when evaluating the benefits and risks of these individual storytelling spaces.

Conclusion

This study was strengthened by the in-depth interviews given by each participant, but the study was limited by the number of interviews conducted and the interview process itself. Due to COVID-19 limitations, all interviews were conducted through videoconference, often reducing the opportunity for the interviewer to analyze participants' body language. In addition, the reliance on virtual interviews means that while we were able to communicate with previvors across the country, each participant's experiences with their local health care system may differ [28]. In addition, this study relied on the most popular *BRCA* social media accounts to recruit participants. On platforms

XSL•FO
RenderX

like TikTok and Instagram, the algorithm may dictate whose content is promoted and whose is not, which results in a skewed sample, but this does not necessarily mean these stories are less worthy of being told. However, the algorithmic bias present on social media platforms can often result in a list of potential participants who are neither racially nor ethnically diverse. Our study skewed primarily toward White women, which mirrors results found in other recent studies of *BRCA* previvors on social media [7,11]. We recognize the lack of scholarship exploring racial and ethnic minority patients who have been diagnosed with the *BRCA* genetic mutation, especially analysis of their online presence and their desire to create digital spaces for the needs of their specific community, and we urge scholars to continue prioritizing these individuals.

Future research should also consider the practice of moderating and censoring of medical conditions based on platform policies, as some of our participants noted their concerns around content being removed or accounts being closed altogether because platforms may deem their content regarding reconstruction and recovery as sexual in nature. Social media platforms are evolving, and the results of their impact on health behaviors and outcomes remain mixed. Therefore, we believe researchers should continue to examine the therapeutic and supportive role of personal social media content across acute and chronic health conditions [29]. In addition, social media literacy continues to be key in improving health decisions, and research suggests platforms themselves can help in this process [30], especially as more patients share about their health conditions on their personal accounts. As more *BRCA* previvors take to social media platforms to share their stories and experiences with genetic testing, treatment, patient-provider interactions, and more, researchers should consider conducting studies that rely on quantitative, qualitative, and critical rhetorical methodologies to provide a larger body of work in this specific subfield.

Acknowledgments

The authors would like to acknowledge the financial support provided by the National Human Genome Research Institute that made this study possible. The authors would also like to thank the anonymous reviewers and editors for their helpful feedback in revising this manuscript. Generative artificial intelligence was not in any portion of the study or in the creation of this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standards for Reporting Qualitative Research (SRQR) Checklist.
[DOCX File , 765 KB - jmir_v27i1e67794_app1.docx]

References

1. Breast cancer now most common form of cancer: WHO taking action. World Health Organization. 2021. URL: https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action [accessed 2023-06-20]
2. Mittal A, Pramanik R, Gogia A, Batra A, Jha A, Kumar L, et al. 155P profile of pathogenic mutations and evaluation of germline genetic testing criteria in consecutive breast cancer patients treated at a North Indian tertiary care center. Annals of Oncology 2021;32:S87-S88. [doi: 10.1016/j.annonc.2021.03.169]
3. Momenimovahed Z, Tiznobaik A, Taheri S, Salehiniya H. Ovarian cancer in the world: epidemiology and risk factors. Int J Womens Health 2019;11:287-299 [FREE Full text] [doi: 10.2147/IJWH.S197604] [Medline: 31118829]
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
5. Mallen AR, Townsend MK, Tworoger SS. Risk factors for ovarian carcinoma. Hematol Oncol Clin North Am 2018;32(6):891-902. [doi: 10.1016/j.hoc.2018.07.002] [Medline: 30390763]
6. Huang J, Chan WC, Ngai CH, Lok V, Zhang L, Lucero-Prisno DE, et al. Worldwide burden, risk factors, and temporal trends of ovarian cancer: a global study. Cancers (Basel) 2022;14(9):2230 [FREE Full text] [doi: 10.3390/cancers14092230] [Medline: 35565359]
7. Wellman ML, Holton AE, Kaphingst KA. "Where do I go? Who do I go to?": BRCA previvors, genetic counselors and family planning. PEC Innov 2023;2:100157 [FREE Full text] [doi: 10.1016/j.pecinn.2023.100157] [Medline: 37214521]
8. Hesse-Biber S. The genetic testing experience of BRCA-positive women: deciding between surveillance and surgery. Qual Health Res 2014;24(6):773-789. [doi: 10.1177/1049732314529666] [Medline: 24747286]

XSL•FO
RenderX

9. Jacobs C, Patch C, Michie S. Communication about genetic testing with breast and ovarian cancer patients: a scoping review. Eur J Hum Genet 2019;27(4):511-524. [doi: 10.1038/s41431-018-0310-4] [Medline: 30573802]
10. Getachew-Smith H, Ross AA, Scherr CL, Dean M, Clements ML. Previving: how unaffected women with a mutation navigate previvor identity. Health Commun 2020;35(10):1256-1265. [doi: 10.1080/10410236.2019.1625002] [Medline: 31163995]
11. Wellman ML, Holton AE, Kaphingst KA. Previvorship posting: why breast cancer previvors share their stories on social media. Health Commun 2023;38(11):2441-2449. [doi: 10.1080/10410236.2022.2074780] [Medline: 35582752]
12. Hereditary breast and ovarian cancer. American Society of Clinical Oncology. 2020. URL: https://www.cancer.net/cancer-types/hereditary-breast-and-ovarian-cancer [accessed 2025-02-19]
13. Basch CH, Hillyer GC, Yalamanchili B, Morris A. How TikTok is being used to help individuals cope with breast cancer: cross-sectional content analysis. JMIR Cancer 2022;8(4):e42245 [FREE Full text] [doi: 10.2196/42245] [Medline: 36472899]
14. Lazard AJ, Collins MKR, Hedrick A, Varma T, Love B, Valle CG, et al. Using social media for peer-to-peer cancer support: interviews with young adults with cancer. JMIR Cancer 2021;7(3):e28234 [FREE Full text] [doi: 10.2196/28234] [Medline: 34473063]
15. Kent EE, Smith AW, Keegan THM, Lynch CF, Wu X, Hamilton AS, et al. Talking about cancer and meeting peer survivors: social information needs of adolescents and young adults diagnosed with cancer. J Adolesc Young Adult Oncol 2013;2(2):44-52 [FREE Full text] [doi: 10.1089/jayao.2012.0029] [Medline: 23781400]
16. Tai E, Buchanan N, Townsend J, Fairley T, Moore A, Richardson LC. Health status of adolescent and young adult cancer survivors. Cancer 2012;118(19):4884-4891 [FREE Full text] [doi: 10.1002/cncr.27445] [Medline: 22688896]
17. Lang MJ, David V, Giese-Davis J. The age conundrum: a scoping review of younger age or adolescent and young adult as a risk factor for clinical distress, depression, or anxiety in cancer. J Adolesc Young Adult Oncol 2015;4(4):157-173 [FREE Full text] [doi: 10.1089/jayao.2015.0005] [Medline: 26697266]
18. Biancovilli P, Makszin L, Csongor A. Breast cancer on social media: a quali-quantitative study on the credibility and content type of the most shared news stories. BMC Womens Health 2021;21(1):202 [FREE Full text] [doi: 10.1186/s12905-021-01352-y] [Medline: 33992111]
19. Wolfers L, Utz S. Social media use, stress, and coping. Curr Opin Psychol 2022;45:101305 [FREE Full text] [doi: 10.1016/j.copsyc.2022.101305] [Medline: 35184027]
20. Jewish woman and BRCA gene mutations. Hereditary breast cancer and BRCA genes. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/cancer/breast/young_women/bringyourbrave/hereditary_breast_cancer/jewish_women_brca.htm#:~:text=One%20in%2040%20Ashkenazi%20Jewish,cancer%20at%20a%20young%20age [accessed 2023-03-21]
21. Clarke V, Braun V. Thematic analysis. The Journal of Positive Psychology 2016;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
22. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
23. Lubitow A, Davis M. Pastel injustice: the corporate use of pinkwashing for profit. Environmental Justice 2011;4(2):139-144 [FREE Full text] [doi: 10.1089/env.2010.0026]
24. Liao Y. Sharing personal health information on social media: balancing self-presentation and privacy. 2019 Presented at: SMSociety '19: Proceedings of the 10th International Conference on Social Media and Society; July 19-21, 2019; Toronto, ON p. 194-204 URL: https://doi.org/10.1145/3328529.3328560 [doi: 10.1145/3328529.3328560]
25. LaMonica HM, Roberts AE, Lee GY, Davenport TA, Hickie IB. Privacy practices of health information technologies: privacy policy risk assessment study and proposed guidelines. J Med Internet Res 2021;23(9):e26317 [FREE Full text] [doi: 10.2196/26317] [Medline: 34528895]
26. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
27. Loeb S, Langford AT, Bragg MA, Sherman R, Chan JM. Cancer misinformation on social media. CA Cancer J Clin 2024;74(5):453-464 [FREE Full text] [doi: 10.3322/caac.21857] [Medline: 38896503]
28. Healthy People 2030 (2020). Health care access and quality. U.S. Department of Health and Human Services. URL: https://health.gov/healthypeople/objectives-and-data/browse-objectives/health-care-access-and-quality, [accessed 2023-10-17]
29. Kanchan S, Gaidhane A. Social media role and its impact on public health: a narrative review. Cureus 2023;15(1):e33737 [FREE Full text] [doi: 10.7759/cureus.33737] [Medline: 36793805]
30. Cho H, Cannon J, Lopez R, Li W. Social media literacy: a conceptual framework. New Media Soc 2024;26(2):941-960 [FREE Full text] [doi: 10.1177/14614448211068530] [Medline: 38328341]

XSL•FO
RenderX

Edited by A Mavragani; submitted 21.10.24; peer-reviewed by N Garrepalli, O Enaworu, AK Ramachandran Sumangala Devi; comments to author 29.11.24; revised version received 07.01.25; accepted 07.02.25; published 03.03.25.

Please cite as:
Wellman ML, Owens CM, Holton AE, Kaphingst KA
Examining BRCA Previvors' Social Media Content Creation as a Form of Self and Community Care: Qualitative Interview Study
J Med Internet Res 2025;27:e67794
URL: https://www.jmir.org/2025/1/e67794
doi:10.2196/67794
PMID:40053732

©Mariah L Wellman, Camilla M Owens, Avery E Holton, Kimberly A Kaphingst. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Lived Experience of Participating in Online Peer-To-Peer Groups After Acquired Brain Injury: Phenomenological Study

Malin Tistad[1,2], PhD; Lill Hultman[1,3,4], PhD; Annica Wohlin Wottrich[1], PhD; Lena von Koch[1,5], PhD

[1]Care Sciences and Society, Department of Neurobiology, Karolinska Institutet, Huddinge, Sweden

[2]School of Health and Welfare, Dalarna University, Falun, Sweden

[3]School of Social Science, Södertörn University, Huddinge, Sweden

[4]Marie Cederschiöld University Collage, Stockholm, Sweden

[5]Theme of Heart and Vascular and Neuro, Karolinska University Hospital, Stockholm, Sweden

Corresponding Author:
Malin Tistad, PhD
Care Sciences and Society
Department of Neurobiology
Karolinska Institutet
Alfred Nobels Allé 23
Huddinge, 141 83
Sweden
Phone: 46 23778554
Email: malin.tistad@ki.se

Abstract

Background: Stroke and other acquired brain injuries (ABIs) can present challenging experiences for individuals, both in recovery of functions affected by visible or invisible impairments and in learning to live with the new situation. Research has shown that sharing experiences face-to-face in peer groups can be beneficial during recovery. However, there is limited knowledge about the lived experiences of people with ABI who participate in online peer-to-peer groups.

Objective: The aim of our study was to explore the lived experiences of participating in online peer-to-peer groups for people with ABI, where participants themselves set the agenda.

Methods: Members of 2 Facebook groups (FBGs) for people with ABI were invited to participate in this study, and 20 individuals were included (14 women and 6 men; age range 24-74 years). One FBG focused on stroke and the other on fatigue caused by ABI. One group was private, and the other group was public. Data were collected through semistructured interviews, in which participants were encouraged to describe their experiences of engaging in FBGs in detail. The interviews were conducted over telephone or Zoom and digitally recorded. The audio recordings were then transcribed verbatim, resulting in 224 pages of text, and analyzed using the empirical phenomenological psychological method.

Results: The analysis presented a common meaning structure with 1 main characteristic that is, “validating self,” common for all 20 participants, and 3 subcharacteristics, that is, “learning—having one’s own experiences confirmed,” “adjusting self—building competence and self-compassion,” and “supporting others—becoming a valued lived-experience expert/authority.” Together, the subcharacteristics reflected a process of validating self from newcomer to lived-experience expert or authority. In this process, members of FBGs moved from being newcomers with pronounced needs for support and to learn and to have their experiences confirmed by others with similar experiences. Thus, participants were building competence and developing self-compassion. Gradually, they assumed the role of advisors, mentors, or coaches, acknowledging their experiences and competence as valuable to others, thereby validating themselves as compassionate lived-experience experts or authorities in supporting others.

Conclusions: Participation in online peer-to-peer groups can offer unique opportunities for individuals with ABI to validate self through processes that involve learning, developing self-compassion and compassion for others, and offering support to others with similar experiences. Given that rehabilitation after an ABI is often of limited duration and that positive experiences can be achieved over time through involvement in digital peer-to-peer support, health care professionals should assist patients by providing information and directing them to digital networks for people with ABI. However, when recommending the use of online peer-to-peer support, impairments and insufficient digital competence that may complicate or prevent the use of social media should be assessed and support provided when relevant.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e67658) doi:10.2196/67658

KEYWORDS

compassion; experiential knowledge; fatigue; self-compassion; stroke; social media; meaning; interview; normalization

Introduction

An acquired brain injury (ABI) is brain damage that occurs after birth and can be either due to a trauma or a nontraumatic cause, for example, a stroke [1]. ABI can impact body functions and lead to impairments that are either visible (eg, reduced motor control) or invisible (eg, impaired memory, fatigue [1], ie, a profound sense of tiredness and a lack of mental and physical energy). Fatigue has been reported as one of the most common and challenging invisible consequences to manage [2]. Stroke is a leading cause of disability, and the disability-adjusted life-years for stroke indicate that it increases both mortality and the risk of living with lifelong disabilities [3]. Life after a stroke can entail challenging experiences for the stroke survivor, both in the recovery of functions as well as in learning to live with visible or invisible disabilities, which may persist and impact their own lives [4-6] as well as those of their loved ones [7]. Stroke survivors frequently report feelings of emptiness, abandonment [8], and a lack of information about stroke, its consequences, recovery, secondary prevention, and medication [9,10]. They also express a need for recurrent support from health care and community services [9,11]. A synthesis of qualitative studies on embodied experiences after stroke describes a 3-phase process: an initial phase of disconnection between oneself, others, and the world; followed by a transitional period exploring and negotiating; and a third phase of reconnecting with oneself, others, and the world [12]. These poststroke recovery processes can be long-lasting [5,13,14].

Support after a stroke can come from different stakeholders, for example, health services, family, social networks, and peers [15]. Peer support is defined as a system of giving and receiving help founded on key principles of respect, shared responsibility, and mutual agreement of what is helpful [16]. It has been proposed that peer support can address affirmational, emotional, and informational needs by validating concerns, offering a place of belonging and providing relevant knowledge to solve everyday problems [17,18]. Being among peers has been reported to contribute to feelings of safety, understanding, and community [19,20].

Traditionally, peer support is organized through physical meetings, often involving health care professionals working together with peer leaders [17,18,21-23]. However, groups led entirely by stroke survivors have also been reported [24]. A meta-analysis of peer support groups for individuals with ABI found limited evidence of their effectiveness but highlighted positive experiences among participants. These experiences were summarized as “being connected, interacting with others, and providing and receiving support” [25].

Over the past decade, an increasing number of studies have examined peer support organized through social media (eg, for young mothers [26] or related to illness such as multiple sclerosis [27], cancer [28], or Parkinson disease [29]). Similar to in-person peer support groups, online peer support groups can be led either by health care professionals or by individuals with lived experience of the topic in focus. In addition, self-formed online communities are emerging on social media, addressing diverse health concerns. Such communities are referred to as online peer-to-peer support [30,31]. However, there is a lack of knowledge regarding the experiences and meaning of participating in online peer-to-peer support networks for individuals with ABI. Hence, our study aims to explore the lived experiences of participating in online peer-to-peer groups for people with ABI on social media, where participants themselves set the agenda.

Methods

Study Design

This study was conducted as part of the research project “Participation in Society After an Acquired Brain Injury,” with the overall purpose of exploring how individuals with ABI share experience-based knowledge in different contexts. The project had a participatory approach, involving 3 individuals with lived experiences of ABI as co-researchers. These co-researchers identified Facebook as an important arena for people with ABI to share their experiences. A phenomenological approach [32] was chosen, as the aim was to explore the phenomenon, meaning, or essence of participating in online peer-to-peer support groups on social media, specifically related to stroke or fatigue after an ABI. The phenomenological approach allowed the researchers to explore the lived experiences of participants and gain a deeper understanding of the phenomenon. The method requires researchers to be open to the original and immediate experiences of the participants while reflecting on their own previous knowledge. The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist is provided in Multimedia Appendix 1.

Participants

All participants in this study were recruited among members of 2 Facebook groups (FBGs). The FBGs were suggested by the co-researchers and deemed relevant to the study. One FBG focused on stroke, and the other on fatigue caused by ABI. One group was private, and the other group was public. The private group was initiated and moderated by individuals with lived experience of stroke. The public group was initiated and administrated by a researcher/health care professional, who occasionally posted information about upcoming lectures or events relevant to the group but did not otherwise interact with members or respond to posts.

The researchers MT and LH contacted the administrators of both FBGs and informed them about the study. The administrators granted permission for MT and LH to recruit participants within their respective groups and posted information about the study provided by the researchers. The posted information asked members interested in being

XSL•FO
RenderX

interviewed regarding their experiences in the FBG community to contact the researchers by email or telephone. Members were informed that, if interested, they would receive further details and be given the opportunity to schedule an interview.

Altogether, 12 members of the stroke FBG and 8 members of the fatigue FBG—all with lived experience of stroke or fatigue, respectively—expressed interest in being interviewed. None of the members were known to the researchers. All those who had shown an interest were contacted by the researchers, and they provided oral and written information about the study. Participants were then asked if they preferred to be interviewed by telephone or by videoconference software (Zoom) and to schedule a time for the interview. When the interview started, participants were once again informed about the study and how data were to be handled. Thereafter, oral informed consent was obtained and recorded.

Data Collection

Data were collected through semistructured interviews [33] using an interview-guide prepared beforehand. The open-ended questions (Multimedia Appendix 2) were intended to encourage participants to describe their experiences of participating in the FBG community in a narrative manner [34]. They were asked to provide concrete details regarding their participation, including how they became members, how they interacted with FBG, and to describe their experiences and feelings in the beginning, over time, and at the time of the interview. Prompts and follow-up questions (eg, could you please tell a bit more?) were used to encourage further elaboration.

MT and LH, both PhD holders, females, and experienced researchers, performed all the interviews. They are both experienced in working professionally with patients/clients with ABI and in collaborating with individuals with ABI in research projects. LH interviewed all the participants from the stroke FBG, while MT interviewed all the participants from the fatigue FBG. Fifteen interviews were conducted by telephone and 5 over Zoom, and all interviews were digitally recorded. The audio recordings were then transcribed verbatim, resulting in 224 pages of text.

Ethics Approval

All participants gave informed consent after having received detailed information about the study. Data were deidentified. This study explores the meaning of participating in FBGs that were created to allow members to support each other and share experiences for their benefit and well-being. Thus, it was essential to conduct the study in a way that considered both the integrity of the participants and the wider FBG community. To minimize the risk of identification, he/him and she/her have been replaced with he/she and him/her, respectively, in the results section. This study was approved by the Swedish Ethical Review Authority (2018/407-31, 2021-03548, 2022-03207-02).

Data Analysis

The transcribed text was analyzed using the empirical phenomenological psychological method [32], which aims to describe the essence, structure, and character of the phenomenon under study. To discover essential descriptions during the analysis, the authors set aside any theoretical frameworks outside phenomenology that could plausibly explain or account for the phenomenon. Thus, the texts were handled unconditionally, allowing the phenomenon to present itself in the participants' narratives. The analysis was performed in 5 steps. Steps 1 through 4 were performed for each interview separately.

In the first step, the authors read through the transcribed interview to gain a general understanding of the concrete facts, events, and feelings pertaining to the participants' original experiences.

In the second step, based on their understanding of the whole transcribed interview, the authors divided the text into smaller segments, that is, meaning units. A new meaning unit was identified each time there was a shift in meaning within the text.

In the third step, the authors interpreted each meaning unit while considering the whole interview and the phenomenon under study. The hidden meaning, often referred to as the transformed meaning [32], was the focus of the interpretation, specifically in relation to the experience of participating in an FBG.

The fourth step entailed synthesizing the transformed meaning units into a situated structure of meaning [32] in a separate summary for each interview. In the situated structure, the authors arranged the features of the phenomenon in a phenomenologically meaningful way based on their interpretation of the meaning of participating in the FBG community.

In the fifth and final step, the authors analyzed the summaries from step 4 across participants to identify shared experiences of participating in an FBG. Hence, the analysis moved from the situated structure of meaning to a general structure of meaning [32] that spanned all participants.

The authors revisited and reviewed the original interviews throughout the analysis to validate their interpretations. The most valid interpretation was chosen to represent the meaning of participating in an FBG, as experienced by the participants. The analyses were performed by AWW, LvK, and MT. AWW and LvK, both PhDs, have professional experience working with patients/clients with ABI, and are experienced in conducting phenomenological research on individuals with ABI. Together, they conducted all 4 steps of the analyses for 2 interviews. For the remaining 18 interviews, steps 1 through 4 were conducted separately by AWW, MT, or LvK. Step 5 was performed by AWW, MT, and LvK together. All steps of the analysis were recorded in web-based Word documents, which were shared between the authors via a web platform accessible only to them.

To enhance trustworthiness throughout the analysis, the authors reflected iteratively on their interpretations, while considering their previous knowledge, clinical experience, understanding, and theories about the phenomenon under study. The findings were presented and discussed with the co-researchers, who had their own experiences of being members of Facebook communities for individuals with ABI, to ensure credibility, that is, that the descriptions of the characteristics made sense and were recognized as plausible interpretations. To further safeguard trustworthiness, the findings were reviewed and

XSL•FO
RenderX

discussed with LH. The analysis was then revised based on the feedback received, with the authors repeatedly returning to the original interviews and revisiting the different steps of the analyses.

Results

Participant Characteristics and Analysis of the Lived Experiences

Our study was performed with 20 participants, aged between 24 and 74 years. Fourteen participants were women, and 6 were men. In this study, participants in the stroke FBG are referred to as participants 1-12, and those in the fatigue FBG are referred to as participants 13-20.

The analysis of the lived experiences of participating in an FBG for individuals with stroke or fatigue after an ABI presented a common meaning structure, with 1 main characteristic, that is, "validating self" and 3 subcharacteristics, that is, "learning—having one's own experiences confirmed," "adjusting self—building competence and self-compassion," and "supporting others—becoming a valued lived-experience expert/authority."

The meaning structure and the main characteristic "validating self" were seen across all participants. Together, the subcharacteristics reflected a process of validating self from newcomer to lived-experience expert or authority. In this process, members of FBGs moved from being newcomers with pronounced needs for support to learn and to have their experiences confirmed by others with similar experiences. Thus, participants were building competence and developing self-compassion. Gradually, they assumed the role of advisors, mentors, or coaches, acknowledging their experiences and competence as valuable to others and thereby validating themselves as compassionate lived-experience experts or authorities in supporting others. Learning processes were present in all subcharacteristics—moving from seeking information to the ability to apply and relate knowledge, both drawn from lived experiences and gained within the FBG community. Furthermore, challenges related to navigating health care and social support systems were present in all subcharacteristics. Here, participants moved from asking for information pertaining to their own situations to using their growing expertise to take initiatives for changes at societal or political levels.

Most participants had found FBGs when googling, while others were recommended to join by peers. A few participants in the FBG community for fatigue had been recommended to join the community by health care professionals. Interaction with the FBG community varied among participants, but the main characteristic "validating self" was present, regardless of how individuals interacted with the group. Most participants described how their level of interaction had changed over time. Nearly all were or had been actively engaged in writing posts and comments since joining the FBG community, and most of them had been more active in the beginning. A few participants, however, had mostly read posts and comments and had rarely, if ever, written anything themselves.

Learning: Having One's Own Experiences Confirmed

A prominent aspect of this subcharacteristic of the lived experience of participating in an FBG was the participants' feelings of being lost, alone, and longing for understanding and guidance. Initially, they searched for information pertaining primarily to their own health and condition. The issues they addressed were driven by individual needs and focused on symptoms, emotions after stroke, life in the context of fatigue, available health care, and support.

> *...I thought, where can I connect and see how others live, how others did? That's where I went to the Facebook group.* [Participant 11]

Participants described feeling different from their previous selves.

> *...You get a new personality when something happens in your head, so you can share different experiences and how to deal with everyday life in general.* [Participant 4]

Participants followed posts and comments made by others to get acquainted with the FBG community, experiencing a sense of connectedness and belonging to a community of people with similar experiences. Some participants emphasized the importance of the friendly atmosphere and the opportunity to laugh together with each other rather than at each other. The community became a safe haven where they felt comfortable expressing their concerns and received valued and comforting responses.

> *...It is a kind of community around a common topic. Specifically, not feeling alone in a situation like this. There are others who have similar problems. I think you can feel that there is a sense of participation in it all. I think it is good. Yes, otherwise, I wouldn't be there myself.* [Participant 13]

Participants had previously used Facebook to interact socially for pleasure, but following their illness, it took on a different meaning and more significant role in meeting the basic needs in their daily lives.

> *…But before I was ill, I used Facebook too. But then, it wasn't very important. Then it was just for fun. But now, it was different when I started looking for information like this, connected to stroke.* [Participant 7]

Issues raised in the FBG community had sometimes been discussed in other contexts but without satisfactory answers. Participants were also grateful for the guidance they received on how to access services they needed and were entitled to. They also expressed appreciation for the 24/7 accessibility of the FBG community, the possibility of receiving valuable answers rapidly and around the clock, and for some, the option to participate anonymously.

> *…I get an answer, instead of having to constantly search for everything myself. It is actually very comforting to have someone in the same boat, someone who is awake and can just write back, where*

XSL•FO
RenderX

> *you can get a small, embracing hug if that is what you need.* [Participant 1]

The meaning of participating in the FBG community was rooted in the participants' experiences of connecting with others who truly understood their situation, that is, a community of individuals with similar experiences. They shared experiences and knowledge gained from their lived experiences, which differed from the information, advice, and support provided by those without similar experiences (eg, health care professionals). It was comforting to learn through comments from others who really knew what it was like and who could offer an insider's perspective. Many participants expressed that before joining the FBG community, they had felt as though no one else was like them.

> *…You can calm each other down in many ways. And there are so many who write the same thing. So, it's not just one person who says, "yes, but it's completely normal." Instead, everyone writes, "yes that is my experience too."* [Participant 18]

Many participants emphasized the importance of experiencing a sense of normality.

> *...There is confirmation that you're not crazy. It's really normal not to be normal. Still, it's normal. No, but I think, it means a lot to encourage each other.* [Participant 17]

Some participants rarely or never wrote their own posts or comments but still learned about themselves by comparing their experiences with those presented by others. One participant who had reading difficulties also found the information in the FBG community easier to understand than information from health care.

> *…When I read the hospital brochures, then they were too long. And then I couldn't read them. So, that's why I use Facebook instead; I was able to take in the information bit by bit.* [Participant 7]

Participants also expressed that certain members of the FBG community who had long-standing experience were particularly valued and respected.

> *...In these groups, those who answer know what they are talking about, many times.* [Participant 14]

Adjusting Self: Building Competence and Self-Compassion

This subcharacteristic illustrates how participants in the FBG communities compared and related their own experiences of managing life after a stroke or with fatigue. They exchanged experiential knowledge through posts and comments with others. Interacting and exchanging information in the community helped them apply this knowledge, build competence, and to look at themselves with kindness and self-compassion as well as at others with compassion.

> *…And sometimes, I hope, that if I write something kind to someone, or share a piece of advice, that it means something. And it has meant a lot to me too when I read about others. Sometimes I am touched, and sometimes I am happy. I think it is valuable.* [Participant 2]

In comparing and relating their experiences with those of others, the participants had their own experiences and competence confirmed. They came to recognize that they had, so far, achieved what could be expected of them and at times more than what could be expected of them, giving them a sense of control. One participant had overcome many obstacles to be where he/she now was in life and valued his/her experiences highly, which was also his/her reason for joining the FBG community.

> *…Then, my motive for joining the group was precisely wanting to see, what did I do differently to get this far when others are so limited.* [Participant 11]

This participant did not actively interact with the FBG community, but by following posts and comments in the community, he/she experienced him/herself as competent, which appeared to strengthen his/her self-compassion. The FBG community was used as a reference point to compare his/her approach to managing challenges in rehabilitation and everyday life, reinforcing his/her self-perception and belief that he/she had managed the situation better than others. The participant expressed that he/she had a lot to offer to others in the community but did not need support from others in similar situations.

> *…My only need, if I should even call it need, is precisely that I want to help people. I want to help people and show them that, unfortunately, they have to do the hard work. That's the only way they will get their life back. Otherwise, they won't. It's not that I expect anything in return, or that they should help me. That's why I say, unfortunately, a bit cocky, that I don't feel I need much from others.* [Participant 11]

Relating personal experiences of success and building self-compassion could also evoke mixed emotions for some FBG members.

> *...I notice that, among us younger, there is a kind of envy. I know I've felt that way myself at some point. But I can still rejoice in the person's success, even if I myself am not there yet.* [Participant 13]

Despite feelings of envy toward those who had made more progress, participants could also distance themselves, express compassion, and rejoice with others. Overall, FBG was experienced as a meaningful supportive community where participants could express both self-compassion and compassion for others.

> *...Yes, it's still great to know that there are others who get to stand up for their own, new personality.* [Participant 4]

Another participant with aphasia expressed the importance of being a member of the community not only by reading but also by contributing through writing their own posts or comments. The participant longed for the moment when he/she would be able to write a comment or post independently.

XSL•FO
RenderX

> *…I hope I'll get a little better at writing, I've gotten better. But I haven't gotten that cocky yet. I might, I think, I will dare to write a little.* [Participant 7]

Some participants, on the other hand, may not have realized that there were others in the FBG community with communication difficulties due to stroke or ABI, and they expressed a dislike when comments were limited to emojis.

Supporting Others: Becoming a Valued Lived-Experience Expert/Authority

A prominent aspect of this subcharacteristic in the lived experience of participating in an FBG was the participants' empathetic understanding and compassion for the situations of other members in the Facebook community. The participants expressed that they had come to realize and understand that there could be large variations in challenges after a stroke or living with fatigue, and they considered this knowledge when interacting in the FBG community. They were careful to acknowledge that their knowledge stemmed from their lived experience, as opposed to medical expertise.

> *...Above all, I feel that many others feel that I am very credible, that I am worth listening to. And that makes what I say actually have an effect. At the same time, it means that I am very careful with what I say.* [Participant 10]

The participants had their own experiences to share as well as knowledge they had gained from other members in the Facebook community who had different lived experiences from their own. They could also share insights based on a long-term perspective, and they often conveyed compassion and hope to support the self-compassion of others.

> *...Then I wrote, if I won't be able to drive again, then I got tips and advice there. And many who had to wait, and some did not get it (driver's license) back at all.* [Participant 8]

The participants' experiences could also help newcomers find words to describe their experiences.

> *...They get explanations and vocabulary for things they have not been able to, things they have known and experienced but have not been able to explain or express.* [Participant 14]

Furthermore, participants could use their experiential knowledge to describe, explain, or provide arguments for the challenges they faced (eg, how devastating fatigue could be in everyday life). Such information could be shared within the FBG community to help describe the lived experience to an employer or family, in hopes that they could gain a better understanding. Even though the family was often very supportive, the participants expressed that it was impossible for someone who did not have similar experiences to fully understand their situation.

The knowledge they shared was deemed valuable by others and validated their sense of self as knowledgeable individuals and authorities within the Facebook community. Several participants described a few members in their FBG who had taken on a role to organize and support others, regularly presenting topics to stimulate discussion, which was highly appreciated. A researcher would occasionally present new research findings regarding fatigue, which the participants found valuable.

The lived experience expertise in this subcharacteristic was also recognized and appreciated outside the FBG community. For instance, a few participants with this expertise was engaged in medical education.

Although the participants had been part of the FBG community for quite a while, they expressed that it was both interesting and rewarding, and they were still learning. They enjoyed being part of the Facebook community, following posts and comments, and sharing their own experience-based knowledge, even though they participated less frequently than at the beginning.

> *…It's almost six years since I had my stroke, but I enjoy seeing people who have just had a stroke and have just been able to get up from the wheelchair and take their first steps. And then you send hearts, and I remember I was like that too. It's moving forward, and keep it up, and well done! So, I like to see when people step it up a notch.* [Participant 9]

This subcharacteristic also describes how participants were cautious when interacting within the FBG community, choosing with whom and how to write. They carefully chose which posts to comment on by reflecting on whether their input could be of value to the person who had written the post. Some participants had chosen to comment on posts made by one particular individual and to specifically follow that person taking on the role of a mentor or coach. One participant described how he/she had continued his/her mentoring outside the FBG community in a private email conversation. Nevertheless, participants reported having few real life contacts or friendships with members of the FBG community outside the digital community.

The participants were thoughtful and considerate about how they wrote their comments to posts made by others—whether newcomers or more experienced members. Creating their comment was experienced as a balancing act; they aimed to offer hope, encouragement, and practical tips on the one hand and being realistic on the other. This required them to draw on their experience-based knowledge of how people react after a stroke/ABI or in living with fatigue. They also exercised patience and tolerance, avoiding negative comments, even when, for example, a member of the FBG community repeatedly asked the same question. In their comments to posts, they assumed the role of mentors with compassion or humble lived-experience authorities, always reflecting on how to avoid being perceived as condescending experts.

Experiences and insights gained through interactions in the FBG community inspired participants to act beyond themselves and the FBG community, as ombudsmen for people with stroke or fatigue. One example was that through the sharing of experiences in the FBG groups, participants realized there were large differences in the availability and access to services between municipalities and regions in the country. The inequalities in access to care and the unmet needs expressed within the FBG community inspired some participants to act for future improvements. One participant, for example, took on

XSL•FO
RenderX

a leadership role in a patient organization and even become involved in local politics.

Discussion

Principal Findings

Our study shows that people who had experienced a stroke or who had developed fatigue after an ABI and had participated in online peer-to-peer groups on social media revealed 1 main characteristic, that is, "validating self," common for all 20 participants, and 3 subcharacteristics, that is, "learning—having one's own experiences confirmed," "adjusting self—building competence and self-compassion," and "supporting others—becoming a valued lived-experience expert/authority." A learning process was integral to the overall process of validating self. Our findings are discussed in relation to theories of self-validation, self-compassion, compassion, and experiential knowledge. Although all analyses were performed with plausible theories set aside, we acknowledge that other applicable theories may also be relevant to the findings of our study.

The meaning of participating in an FBG for people with stroke or for people with fatigue after an ABI presented 1 common main characteristic: validating self. The self-validation theory focuses on the subjective sense that one's thoughts are valid or appropriate to use [35], as opposed to the objective truth of those thoughts. According to this theory, individuals tend to rely on their thoughts more when they perceive them as likely to be true (cognitive validation) or because when they feel good about the thought (affective validation). Self-validation is furthermore described as a process in which a wide array of factors can influence the perceived validity of one's thoughts. These factors include individual aspects such as self-esteem and situational factors such as body posture at a particular time [35]. The theory of self-validation is proposed to apply to any type of thought, for example, self-efficacy [36], which refers to the belief in one's ability to produce outcomes or achieve goals [37]. By participating in FBGs, all participants reported having their experiences and thought of themselves after a stroke living with fatigue confirmed and validated, both cognitively (ie, recognizing that their experiences were normal) and affectively (ie, not feeling alone). However, the meaning of "what validated the self" varied across the subcharacteristics and differed between participants who recently experienced a stroke and participants with longer lived experiences of stroke or fatigue after an ABI.

Together, the participants described a process in which they initially sought answers for their troubling experiences and lack of information. By posting their concerns in the FBG community and receiving comments and answers, they appeared to learn and find comfort, as their experiences and thoughts were validated by others. The participants gained an understanding of themselves and were able to view themselves with kindness and self-compassion as well as view others with compassion. Nevertheless, we acknowledge that the outlined process could also be supported in other contexts and that entry into the FBG community can take place at different points in time after an ABI, still supporting self-validation. At least one participant appeared to have joined the FBG community at a later stage, where his/her lived experiences were related to the subcharacteristic "adjusting self." Moreover, it is uncertain whether all participants will eventually experience aspects of the subcharacteristic "supporting others."

The development of self-compassion and compassion was prominent in 1 subcharacteristic. Self-compassion is described as being supportive toward oneself when experiencing suffering or pain [38]. The experiences of the study participants resonate with Neff's theory of self-compassion, which entails being present in one's own suffering, understanding and supporting oneself, and feeling connected to others who are also suffering [38]. Self-compassion has 3 interrelated elements that work together to alleviate suffering: the emotional response to suffering, how it is understood cognitively, and how the person pays attention to their suffering. It is conceivable that all elements, along with the development of self-compassion, were supported through experiences of being connected to and learning from others in the Facebook community. One example is that participants seemed to be supported in both standing up for themselves and practicing self-compassion when they received confirmation that others also needed to stand up for their new selves. The theory postulates that when we respond to suffering with kindness toward ourselves, feelings of unworthiness decrease in a similar manner to how we feel when we receive kindness from others [38].

Compassion is described as feelings that arise when witnessing others' suffering and a subsequent desire to help and is different from empathy, which refers to the vicarious experience of another's emotion. Unlike feeling pity or sympathy for someone, compassion includes feelings of connection with others who are suffering [39]. Hence, compassion arises when individuals are connected to others with similar experiences, which was a prominent experience for the study participants.

Our results show that the main characteristic, that is, validating self and its subcharacteristics involve learning processes. These learning processes align well with the 4 epistemic patterns of experiential knowledge pertaining to the lived experience of illness, as previously described by Halloy et al [40]. In the first pattern, the authors suggest that the individual experiences the disability or illness. Our study participants' initial experiences of disability led to feelings of being lost and alone, which appeared to be the reason they sought more knowledge, to learn, and to better understand themselves. They learned that they were not alone, and as proposed by Halloy et al [40], in the second pattern of experiential knowledge, they gained a know-how based on their individual experiences. The third epistemic pattern of experiential knowledge was also evident among the study participants. In this pattern, the experiential knowledge, according to Halloy et al [40], becomes more formalized and tends to be developed at a collective level. Hence, the FBG community appeared to provide the arena for the participants to reflect on their experiences and knowledge, thereby developing experiential knowledge, whether they were interacting with other members of the FBG community or simply reading posts and comments made by others. Undoubtedly, some participants in this study had developed experiential expertise, which corresponds to the proposed fourth pattern of experiential knowledge. Experiential expertise entails the ability

XSL•FO
RenderX

to transfer knowledge between different epistemic worlds, as exemplified in the findings by the transfer from the microlevel within the FBG community to the macro and metalevels, such as in the roles of politicians and contributors to education.

The advantages of online peer-to-peer support over face-to-face peer groups, previously suggested in the literature [30,31], were confirmed in the findings of this study. For example, participants were able to validate themselves simply by following posts and comments made by others, without necessarily interacting with the community or providing/receiving support. This highlighted the possibility of controlling the level of interaction with the Facebook community, which is one of the key advantages of online peer-to-peer support [30]. Another highly appreciated advantage reported by several participants was the around the clock accessibility of the community. Participants felt they could reach out to the Facebook community for support at any time and that their concerns would be addressed promptly. Furthermore, the unrestricted access to the FBG communities allowed participants with long and vast lived life experiences after stroke or with fatigue after an ABI to remain active members of the Facebook community, even if most of them interacted less frequently. They were highly appreciated sources of knowledge and understanding, which were not readily available elsewhere.

The number of participants and the wide variation in age and experiences upon which the results are based is a strength of this study. All participants who contacted the researchers were interviewed, and the large amount of data gathered provided rich, detailed, and relevant descriptions of the phenomenon supporting meaning saturation [41]. However, the recruitment procedure, that is, asking members of FBG communities to sign up voluntarily, may have led to the recruitment of participants with mostly positive experiences. Those who were disappointed had negative experiences and hence did not find meaning in participating in the FBG community may not have signed up for the study or may no longer have been active in the FBGs under study. Nevertheless, the results align with previously reported stages of the embodied recovery process after stroke, which includes an initial phase of disconnection between oneself, others, and the world, followed by exploring and negotiating, and then reconnecting with oneself, others, and the world [12]. They also align with the way experiences of peer group support have been summarized previously, that is, “being connected, interacting with others, and providing and receiving support” [25]. The explicit aim of this study was to explore the meaning of participating in an online peer-to-peer support group. Additional studies are needed to explore how, when, and why experiences may be perceived as negative, and how the process described in this study can be interrupted or fail to take place.

Implication for Practice

Given that rehabilitation after an ABI is often of limited duration and that positive experiences can be achieved over time through involvement in digital peer-to-peer support, health care professionals should assist patients by providing information and directing them to digital networks for people with ABI. When recommending the use of online peer-to-peer support, impairments and insufficient digital competence that may complicate or prevent access to such support for individuals with ABI must be evaluated. Given the increasing digitalization of the society and the need to interact through online activities in everyday life, rehabilitation professionals should regard online activities equivalent to other daily activities and consequently assess and support individuals’ ability to perform such activities. Hence, ensuring that patients can engage in online peer-support according to their preferences can be one way of equipping them with tools to manage their new life situation after an ABI by being part of a context with the potential to support learning processes and the recovery process beyond the rehabilitation phase.

Conclusion

We conclude that participation in an FBG can offer unique opportunities to validate self through processes that involve learning, the development of self-compassion, compassion for others, and supporting others with similar experiences of life after a stroke or living with fatigue after an ABI. Furthermore, this process contributes to the building of experiential knowledge and expertise, which may be applicable beyond the online peer-to-peer support groups.

Acknowledgments

A special thanks to Fredrik Sandman, Jeanette Nelson, and Mikael Åkerlund for suggesting Facebook as an important context for sharing experiences and knowledge, identifying the relevant Facebook groups, and providing feedback during the discussion of the manuscript draft. No generative artificial intelligence has been used in any part of this study. This study was funded by Forte (2020-00685).

Data Availability

The data that support the findings of this study are available on reasonable request from the corresponding author (MT). The data are not publicly available because of information that could compromise research participant privacy/consent.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

COREQ checklist.
[PDF File (Adobe PDF File), 440 KB - jmir_v27i1e67658_app1.pdf]

Multimedia Appendix 2
Interview guide.
[DOCX File , 16 KB - jmir_v27i1e67658_app2.docx]

References

1. Goldman L, Siddiqui EM, Khan A, Jahan S, Rehman MU, Mehan S, et al. Understanding acquired brain injury: a review. Biomedicines 2022 Sep 02;10(9):2167 [FREE Full text] [doi: 10.3390/biomedicines10092167] [Medline: 36140268]
2. Elf M, Eriksson G, Johansson S, von Koch L, Ytterberg C. Self-reported fatigue and associated factors six years after stroke. PLoS One 2016;11(8):e0161942 [FREE Full text] [doi: 10.1371/journal.pone.0161942] [Medline: 27575043]
3. GBD 2021 Stroke Risk Factor Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024 Oct;23(10):973-1003 [FREE Full text] [doi: 10.1016/S1474-4422(24)00369-7] [Medline: 39304265]
4. Erikson A, Ranner M, Guidetti S, von Koch L. In search of self after stroke: a longitudinal qualitative study in the context of client-centred rehabilitation. Int J Qual Stud Health Well-being 2023 Dec;18(1):2282513 [FREE Full text] [doi: 10.1080/17482631.2023.2282513] [Medline: 38010156]
5. Erikson A, Karlsson G, Tham K. Living with the long-term consequences 11-13 years after stroke: A phenomenological study. J Rehabil Med 2016 Nov 11;48(10):847-852 [FREE Full text] [doi: 10.2340/16501977-2161] [Medline: 27797394]
6. Woodman P, Riazi A, Pereira C, Jones F. Social participation post stroke: a meta-ethnographic review of the experiences and views of community-dwelling stroke survivors. Disabil Rehabil 2014;36(24):2031-2043. [doi: 10.3109/09638288.2014.887796] [Medline: 24597937]
7. Ytterberg C, von Koch L, Erikson A. Abandoned to the strains of daily life: a qualitative study of the long-term experiences in partners to persons after a mild to moderate stroke. Disabil Rehabil 2019 Mar;41(6):649-655. [doi: 10.1080/09638288.2017.1401674] [Medline: 29132222]
8. Röding J, Lindström B, Malm J, Ohman A. Frustrated and invisible--younger stroke patients' experiences of the rehabilitation process. Disabil Rehabil 2003 Aug 05;25(15):867-874. [doi: 10.1080/0963828031000122276] [Medline: 12851097]
9. Pindus DM, Mullis R, Lim L, Wellwood I, Rundell AV, Abd Aziz NA, et al. Stroke survivors' and informal caregivers' experiences of primary care and community healthcare services - A systematic review and meta-ethnography. PLoS One 2018;13(2):e0192533 [FREE Full text] [doi: 10.1371/journal.pone.0192533] [Medline: 29466383]
10. Kjörk EK, Gunnel C, Lundgren-Nilsson ?, Sunnerhagen KS. Experiences, needs, and preferences for follow-up after stroke perceived by people with stroke and healthcare professionals: A focus group study. PLoS One 2019;14(10):e0223338 [FREE Full text] [doi: 10.1371/journal.pone.0223338] [Medline: 31574135]
11. Tistad M, Tham K, von Koch L, Ytterberg C. Unfulfilled rehabilitation needs and dissatisfaction with care 12 months after a stroke: an explorative observational study. BMC Neurol 2012 Jun 18;12:40 [FREE Full text] [doi: 10.1186/1471-2377-12-40] [Medline: 22708545]
12. Lo TLT, Lee JLC, Ho RTH. Recovery beyond functional restoration: a systematic review of qualitative studies of the embodied experiences of people who have survived a stroke. BMJ Open 2023 Feb 09;13(2):e066597 [FREE Full text] [doi: 10.1136/bmjopen-2022-066597] [Medline: 36759032]
13. Arntzen C, Borg T, Hamran T. Long-term recovery trajectory after stroke: an ongoing negotiation between body, participation and self. Disabil Rehabil 2015;37(18):1626-1634. [doi: 10.3109/09638288.2014.972590] [Medline: 25318537]
14. Nilsson M, Fugl-Meyer K, von Koch L, Ytterberg C. Experiences of sexuality six years after stroke: a qualitative study. J Sex Med 2017 Jun;14(6):797-803 [FREE Full text] [doi: 10.1016/j.jsxm.2017.04.061] [Medline: 28479131]
15. Pallesen H, Aadal L, Moe S, Arntzen C. Gateway to recovery: a comparative analysis of stroke patients' experiences of change and learning in Norway and Denmark. Rehabil Res Pract 2019;2019:1726964 [FREE Full text] [doi: 10.1155/2019/1726964] [Medline: 30775038]
16. Mead S, Hilton D, Curtis L. Peer support: a theoretical perspective. Psychiatr Rehabil J 2001;25(2):134-141. [doi: 10.1037/h0095032] [Medline: 11769979]
17. Kessler D, Egan M, Kubina L. Peer support for stroke survivors: a case study. BMC Health Serv Res 2014 Jun 16;14:256 [FREE Full text] [doi: 10.1186/1472-6963-14-256] [Medline: 24935460]
18. Wijekoon S, Wilson W, Gowan N, Ferreira L, Phadke C, Udler E, et al. Experiences of occupational performance in survivors of stroke attending peer support groups. Can J Occup Ther 2020 Jun;87(3):173-181. [doi: 10.1177/0008417420905707] [Medline: 32115988]
19. Christensen ER, Golden SL, Gesell SB. Perceived benefits of peer support groups for stroke survivors and caregivers in rural North Carolina. North Carolina Medical Journal 2019 May 09;80(3):143-148. [doi: 10.18043/ncm.80.3.143]

XSL•FO
RenderX

20. Pedersen SG, Anke A, Aadal L, Pallesen H, Moe S, Arntzen C. Experiences of quality of life the first year after stroke in Denmark and Norway. A qualitative analysis. Int J Qual Stud Health Well-being 2019 Dec;14(1):1659540 [FREE Full text] [doi: 10.1080/17482631.2019.1659540] [Medline: 31547779]
21. Kennedy A, Turner B, Kendall M. Growth in a 'new world': Case studies of peer leader experiences in the steps program for people with acquired brain injury. Brain Impairment 2012 Feb 21;12(2):152-164. [doi: 10.1375/brim.12.2.152]
22. Morris R, Morris P. Participants' experiences of hospital-based peer support groups for stroke patients and carers. Disabil Rehabil 2012;34(4):347-354. [doi: 10.3109/09638288.2011.607215] [Medline: 21992465]
23. Wan X, Chau JPC, Mou H, Liu X. Effects of peer support interventions on physical and psychosocial outcomes among stroke survivors: A systematic review and meta-analysis. Int J Nurs Stud 2021 Sep;121:104001. [doi: 10.1016/j.ijnurstu.2021.104001] [Medline: 34246069]
24. May C, Bieber K, Chow D, Mortenson WB, Schmidt J. Experiences of adults with stroke attending a peer-led peer-support group. Brain Impairment 2023;24(3):443. [doi: 10.1071/ib23051]
25. Hughes R, Fleming P, Henshall L. Peer support groups after acquired brain injury: a systematic review. Brain Inj 2020 Jun 06;34(7):847-856. [doi: 10.1080/02699052.2020.1762002] [Medline: 32421382]
26. Sjöberg M, Lindgren S. Challenging the roles of "skilled" professionals and "risky" young mothers: peer support, expertise, and relational patterns in Facebook groups. Journal of Technology in Human Services 2017 Sep 13;35(3):247-270. [doi: 10.1080/15228835.2017.1367350]
27. Della Rosa S, Sen F. Health topics on Facebook groups: content analysis of posts in multiple sclerosis communities. Interact J Med Res 2019 Feb 11;8(1):e10146 [FREE Full text] [doi: 10.2196/10146] [Medline: 30741640]
28. Holdren J, Surkan K, Downing A. Perspectives of people with cancer or hereditary cancer risk on the use and value of online peer support. J Patient Cent Res Rev 2023;10(2):58-67 [FREE Full text] [doi: 10.17294/2330-0698.1968] [Medline: 37091112]
29. Gerritzen EV, Lee AR, McDermott O, Coulson N, Orrell M. Online peer support for people with Parkinson disease: narrative synthesis systematic review. JMIR Aging 2022 Jul 27;5(3):e35425 [FREE Full text] [doi: 10.2196/35425] [Medline: 35896025]
30. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
31. Rayland A, Andrews J. From social network to peer support network: opportunities to explore mechanisms of online peer support for mental health. JMIR Ment Health 2023 Feb 28;10:e41855 [FREE Full text] [doi: 10.2196/41855] [Medline: 36853738]
32. Karlsson G. Psychological Qualitative Research From a Phenomenological Perspective. Stockholm, Sweden: Almqvist & Wiksell International; 1995.
33. Kvale S, Brinkman S. Interviews: Learning the Craft of Qualitative Research Interviewing, 2nd Edition. Thousand Oaks, CA, USA: Sage Publications, Inc; 2009.
34. Mattingly C. Healing Dramas and Clinical Plots: The Narrative Structure of Experience. Cambridge, United Kingdom: Cambridge University Press; 1998.
35. Briñol P, Petty RE. Self-validation theory: An integrative framework for understanding when thoughts become consequential. Psychol Rev 2022 Mar;129(2):340-367. [doi: 10.1037/rev0000340] [Medline: 34968131]
36. Briñol P, Petty RE, Moreno L, Gandarillas B. Self-validation theory: Confidence can increase but also decrease performance in applied settings. Span. J. Psychol 2023 Apr 20;26:e5. [doi: 10.1017/sjp.2023.5]
37. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychological Review 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
38. Neff KD. Self-compassion: theory, method, research, and intervention. Annu Rev Psychol 2023 Jan 18;74(1):193-218 [FREE Full text] [doi: 10.1146/annurev-psych-032420-031047] [Medline: 35961039]
39. Goetz JL, Keltner D, Simon-Thomas E. Compassion: an evolutionary analysis and empirical review. Psychol Bull 2010 May;136(3):351-374 [FREE Full text] [doi: 10.1037/a0018807] [Medline: 20438142]
40. Halloy A, Simon E, Hejoaka F. Defining patient's experiential knowledge: Who, what and how patients know. A narrative critical review. Sociol Health Illn 2023 Feb;45(2):405-422 [FREE Full text] [doi: 10.1111/1467-9566.13588] [Medline: 36443986]
41. Sebele-Mpofu FY. Saturation controversy in qualitative research: Complexities and underlying assumptions. A literature review. Cogent Social Sciences 2020 Oct 23;6(1):1. [doi: 10.1080/23311886.2020.1838706]

Abbreviations

ABI: acquired brain injury
COREQ: Consolidated Criteria for Reporting Qualitative Research
FBG: Facebook group

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.10.24; peer-reviewed by M Kendall, C Papadimitriou; comments to author 22.01.25; revised version received 10.02.25; accepted 25.02.25; published 25.03.25.

Please cite as:
Tistad M, Hultman L, Wohlin Wottrich A, von Koch L
The Lived Experience of Participating in Online Peer-To-Peer Groups After Acquired Brain Injury: Phenomenological Study
J Med Internet Res 2025;27:e67658
URL: https://www.jmir.org/2025/1/e67658
doi:10.2196/67658
PMID:40131323

©Malin Tistad, Lill Hultman, Annica Wohlin Wottrich, Lena von Koch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Major Adverse Kidney Events in Hospitalized Older Patients With Acute Kidney Injury: Machine Learning–Based Model Development and Validation Study

Xiao-Qin Luo[1], MD; Ning-Ya Zhang[2], MS; Ying-Hao Deng[3], MD; Hong-Shen Wang[4], MD; Yi-Xin Kang[5], MD; Shao-Bin Duan[5], MD

[1]Department of Geriatrics, The Second Xiangya Hospital of Central South University, Changsha, China

[2]Information Center, The Second Xiangya Hospital of Central South University, Changsha, China

[3]Health Management Center, National Clinical Research Center for Metabolic Diseases, Hunan Provincial Clinical Medicine Research Center for Intelligent Management of Chronic Disease, The Second Xiangya Hospital of Central South University, Changsha, China

[4]Department of Critical Care Medicine, The Second Xiangya Hospital of Central South University, Changsha, China

[5]Department of Nephrology, Hunan Key Laboratory of Kidney Disease and Blood Purification, The Second Xiangya Hospital of Central South University, Changsha, China

Corresponding Author:
Shao-Bin Duan, MD
Department of Nephrology
Hunan Key Laboratory of Kidney Disease and Blood Purification
The Second Xiangya Hospital of Central South University
139 Renmin Road
Changsha, 410011
China
Phone: 86 73185295100
Email: duansb528@csu.edu.cn

Abstract

Background: Acute kidney injury (AKI) is a common complication in hospitalized older patients, associated with increased morbidity, mortality, and health care costs. Major adverse kidney events within 30 days (MAKE30), a composite of death, new renal replacement therapy, or persistent renal dysfunction, has been recommended as a patient-centered endpoint for clinical trials involving AKI.

Objective: This study aimed to develop and validate a machine learning–based model to predict MAKE30 in hospitalized older patients with AKI.

Methods: A total of 4266 older patients (aged ≥ 65 years) with AKI admitted to the Second Xiangya Hospital of Central South University from January 1, 2015, to December 31, 2020, were included and randomly divided into a training set and an internal test set in a ratio of 7:3. An additional cohort of 11,864 eligible patients from the Medical Information Mart for Intensive Care Ⅳ database served as an external test set. The Boruta algorithm was used to select the most important predictor variables from 53 candidate variables. The eXtreme Gradient Boosting algorithm was applied to establish a prediction model for MAKE30. Model discrimination was evaluated by the area under the receiver operating characteristic curve (AUROC). The SHapley Additive exPlanations method was used to interpret model predictions.

Results: The overall incidence of MAKE30 in the 2 study cohorts was 28.3% (95% CI 26.9%-29.7%) and 26.7% (95% CI 25.9%-27.5%), respectively. The prediction model for MAKE30 exhibited adequate predictive performance, with an AUROC of 0.868 (95% CI 0.852-0.881) in the training set and 0.823 (95% CI 0.798-0.846) in the internal test set. Its simplified version achieved an AUROC of 0.744 (95% CI 0.735-0.754) in the external test set. The SHapley Additive exPlanations method showed that the use of vasopressors, mechanical ventilation, blood urea nitrogen level, red blood cell distribution width-coefficient of variation, and serum albumin level were closely associated with MAKE30.

Conclusions: An interpretable eXtreme Gradient Boosting model was developed and validated to predict MAKE30, which provides opportunities for risk stratification, clinical decision-making, and the conduct of clinical trials involving AKI.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200061610; https://tinyurl.com/3smf9nuw

XSL•FO
RenderX

(J Med Internet Res 2025;27:e52786) doi:10.2196/52786

KEYWORDS

major adverse kidney events within 30 days; older; acute kidney injury; machine learning; prediction model

Introduction

Acute kidney injury (AKI) is a clinical syndrome characterized by a rapid decline in renal function [1]. The incidence of AKI has been reported to be about 10%-15% in hospitalized patients and over 50% in critically ill patients [2-4]. AKI is common in older individuals, associated with increased morbidity, mortality, and health care costs [5-7]. The prevalence of multiple comorbidities as well as age-related changes in the kidneys, systemic vasculature, and immune system render older patients more susceptible to kidney damage and less likely to recover [5,8]. Given the already high and increasing incidence of AKI in the older, there is a tremendous need to develop clinical tools for risk stratification of postAKI outcomes in this age group.

Major adverse kidney events within 30 days (MAKE30), a composite of death, new renal replacement therapy (RRT), or persistent renal dysfunction (PRD), has been recognized as a key metric of post-AKI outcomes [9,10]. Patients with no renal recovery during hospitalization show a higher risk of long-term adverse outcomes than those who recovered [11,12]. Furthermore, the initiation of RRT after AKI onset is associated with long-term renal dysfunction and death [13,14]. Previous studies have confirmed MAKE30 as a common, feasible to measure, and clinically meaningful endpoint in clinical trials involving AKI [15-17]. The prediction of MAKE30 can be critical for evaluating patient prognosis, guiding clinical decision-making, and facilitating the conduct of clinical trials.

Clinical prediction models are mathematical tools that are primarily intended to assist physicians in their clinical decision-making [18]. Recently, the rapid development in medical big data and advances in computer science have sparked a growing interest in applying machine learning techniques to develop clinical prediction models [19]. Compared with traditional statistical methods, advanced machine learning algorithms can better integrate large amounts of clinical data, fit complex nonlinear relationships, and analyze high-order interactions. The eXtreme Gradient Boosting (XGBoost) algorithm, an efficient implementation of the gradient boosting framework, represents one of the notable advances in machine learning and is widely used in the medical field [20-23]. XGBoost excels in preventing overfitting during the training process, enhancing predictive performance and robustness in complex data scenarios.

Therefore, the objective of this study was to develop and validate a clinical prediction model for MAKE30 in hospitalized older patients with AKI using the machine learning XGBoost algorithm.

Methods

Study Design

This study identified older patients (aged ≥ 65 years) with AKI who were admitted to the Second Xiangya Hospital of Central South University from January 1, 2015, to December 31, 2020. AKI was defined according to the Kidney Disease: Improving Global Outcomes criteria as an increase in serum creatinine (SCr) by ≥26.5 μmol/L within 48 hours or to ≥1.5 times baseline within 7 days [24]. Baseline SCr was defined as the lowest creatinine in the past 7 days. AKI was determined every time a SCr measurement occurred based on the changes in SCr over the past 48 hours and 7 days. Patients were excluded if they were diagnosed with end-stage renal disease, had a hospital stay of less than 48 hours, had an initial SCr≥353.6 μmol/L at admission, or required RRT before the diagnosis of AKI.

In addition, a cohort containing eligible patients from the Medical Information Mart for Intensive Care Ⅳ (MIMIC-Ⅳ) database was used for model external validation [25]. MIMIC-Ⅳ is a relational database containing comprehensive information for intensive care unit (ICU) admissions at the Beth Israel Deaconess Medical Center (BIDMC) from 2008 to 2019.

Data Collection

Data on demographics, diagnoses, laboratory tests, interventions, and medications were collected. A total of 53 candidate predictor variables were identified that were considered clinically relevant to MAKE30, were available in the electronic health records (EHRs), and had a proportion of missing values less than 30% (Multimedia Appendix 1). Comorbidities were determined based on diagnoses encoded in the *ICD-9* (*International Classification of Diseases, Ninth Edition*) or *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*). The burden of comorbidities was assessed by the Charlson comorbidity index [26,27]. Laboratory tests were recorded as the measurements within 48 hours of and closest to the diagnosis of AKI. Medications were identified according to the Anatomic Therapeutic Chemical classification system and included if administered within 48 hours after the diagnosis of AKI [28]. The use of mechanical ventilation within 48 hours after the diagnosis of AKI was also collected.

Outcome Measures

The primary outcome was MAKE30, defined as a composite of death, new RRT, or PRD at hospital discharge or at 30 days after the diagnosis of AKI, whichever occurred first [9,10]. PRD was defined as a final inpatient SCr value ≥2 times baseline. Secondary outcomes were length of hospital stay as well as death within 30 days and 1 year after discharge. Survival data were extracted from the Chinese Center for Disease Control and Prevention cause-of-death reporting system.

XSL•FO
RenderX

Statistical Analysis

Statistical analyses were performed using R software (version 4.1.2; R Foundation for Statistical Computing). Categorical data are presented as numbers and percentages and were analyzed with the chi-square test or Fisher exact test, as appropriate. Continuous data are described with median (IQR) and were compared using the Mann-Whitney *U* test. Survival data were analyzed with the Kaplan-Meier method, and differences in survival were evaluated by the log-rank test. The 2-sided α level was set at .05. Missing data were analyzed by the missForest method, which is a nonparametric missing value imputation for mixed-type data [29]. The description of missing data can be found in Multimedia Appendix 2.

Patients from the Second Xiangya Hospital were randomly divided into a training set and an internal test set in a ratio of 7:3. The cohort of patients from the MIMIC-Ⅳ database was used as an external test set.

Feature Selection

To build a parsimonious model, feature selection was performed on the training set. First, we removed the predictor variables with near-zero variance. Near-zero variance was defined as a situation where a variable exhibits very little variation or almost no variability across its values, characterized by over 95% of its values being identical. Then the Boruta algorithm was used to select the most important predictor variables [30]. Boruta is an all-relevant feature selection wrapper algorithm, capable of working with classification methods that output variable importance measures; in this study, the Random Forest model was used. There were 30 variables ultimately selected for inclusion within the machine learning model (Multimedia Appendix 1).

Model Development and Validation

A machine learning XGBoost model was established on the training set to predict the development of MAKE30 [31]. XGBoost is an optimized implementation of the gradient boosting framework, which sequentially adds weak models (decision trees) to iteratively improve the overall prediction. XGBoost improves upon traditional gradient boosting algorithms by offering scalability, regularization techniques, optimized tree construction, and customization options. A set of model hyperparameters were optimized by running 5 random shuffles of 5-fold cross-validation, including eta, max_depth, min_child_weight, gamma, colsample_bytree, and subsample. The descriptions and search ranges for the hyperparameters are listed in Multimedia Appendix 3. The XGBoost model can be accessed at [LuoXiaoqin123/MAKE30-in-elderly-AKI] (GitHub).

The performance of the XGBoost model was evaluated on the internal and external test sets. Model discrimination was assessed by the area under the receiver operating characteristic curve (AUROC), which is an assessment metric that remains unaffected by incidence or threshold selection. Model calibration was evaluated using the calibration plot aggregated by deciles and the Brier score. The area under the precision-recall curve (AUPRC) was calculated, considering that it is a useful metric for class-imbalanced data in a problem setting where finding the adverse outcome is crucial. A decision curve analysis was performed to illustrate the clinical use of the model. In addition, we determined the optimal cutoff by the receiver operating characteristic (ROC) curve and the maximum Youden index (sensitivity + specificity – 1) in the training set. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), and negative likelihood ratio (NLR) at the optimal cutoff were calculated.

In the external validation set, there were 9 predictor variables with a proportion of missing values exceeding 50% (Multimedia Appendix 2). In this case, we simplified the model to include 21 variables, with model development performed as described above using the same training set.

Model Interpretations

We used the SHapley Additive exPlanations (SHAP) method to interpret model predictions [32]. SHAP is a game theoretic approach that helps in explaining the output of any machine learning model. It can provide consistent and locally accurate attribution values, the SHAP values, for each feature within a model. A positive SHAP value represents a higher risk of the outcome, whereas a negative SHAP value represents a lower risk of the outcome. The contribution of a feature to the predicted risk of the outcome can be explained by the cumulative effects of feature attribution in each observation.

Sensitivity Analyses

In order to mitigate the potential impact of age on our findings, we further conducted sensitivity analyses to assess the model's robustness across patients stratified by age. We evaluated the model's performance in distinct age groups: <70 years, 70 to 80 years, and ≥80 years.

Ethical Considerations

The medical ethics committee of the Second Xiangya Hospital of Central South University approved the study protocol (2022-K031) and waived the requirement for informed consent. This project has been registered in Chinese Clinical Trial Registry (ChiCTR2200061610). The institutional review boards of BIDMC and the Massachusetts Institute of Technology approved the project and waived the requirement for informed consent. The study followed the Declaration of Helsinki and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement [33].

Results

Patient Characteristics

A total of 4266 patients from the Second Xiangya Hospital were enrolled, including 2973 patients in the training set and 1293 patients in the internal test set (Figure 1). The overall incidence of MAKE30 in this medical center was 28.3% (95% CI 26.9%-29.7%), including 12.3% (95% CI 11.4%-13.4%) with death, 10.2% (95% CI 9.3%-11.1%) with new RRT, and 17.7% (95% CI 16.6%-18.9%) with PRD. The external test set contained 11,864 patients from the MIMIC-Ⅳ database, 26.7% (95% CI 25.9%-27.5%) of whom developed MAKE30 (Multimedia Appendix 4). The incidence of death, new RRT,

XSL•FO
RenderX

and PRD were 18.9% (95% CI 18.2%-19.6%), 4.9% (95% CI 4.5%-5.3%), and 12.8% (95% CI 12.2%-13.4%), respectively.

Table 1 shows the characteristics of patients from the Second Xiangya Hospital stratified by MAKE30. Patients with MAKE30 were older, had a higher burden of comorbidities, and had greater disease severity than patients without MAKE30. Compared with the nonMAKE30 group, the MAKE30 group had shorter length of hospital stay as well as higher 30-day and 1-year mortality after discharge. Multimedia Appendix 5 shows the Kaplan-Meier survival curves of death within 30 days and 1 year for patients alive at discharge stratified by PRD or RRT. Patients with versus without new RRT or PRD had higher 30-day and 1-year mortality after discharge (log-rank P<.001). Multimedia Appendix 6 provides the characteristics of patients from the MIMIC-Ⅳ database stratified by MAKE30. Multimedia Appendix 7 provides the characteristics of patients in the training, internal test, and external test sets.

Figure 1. Flow diagram of patient selection from the Second Xiangya Hospital. AKI: acute kidney injury; MAKE30, major adverse kidney events within 30 days; RRT: renal replacement therapy; SCr: serum creatinine.

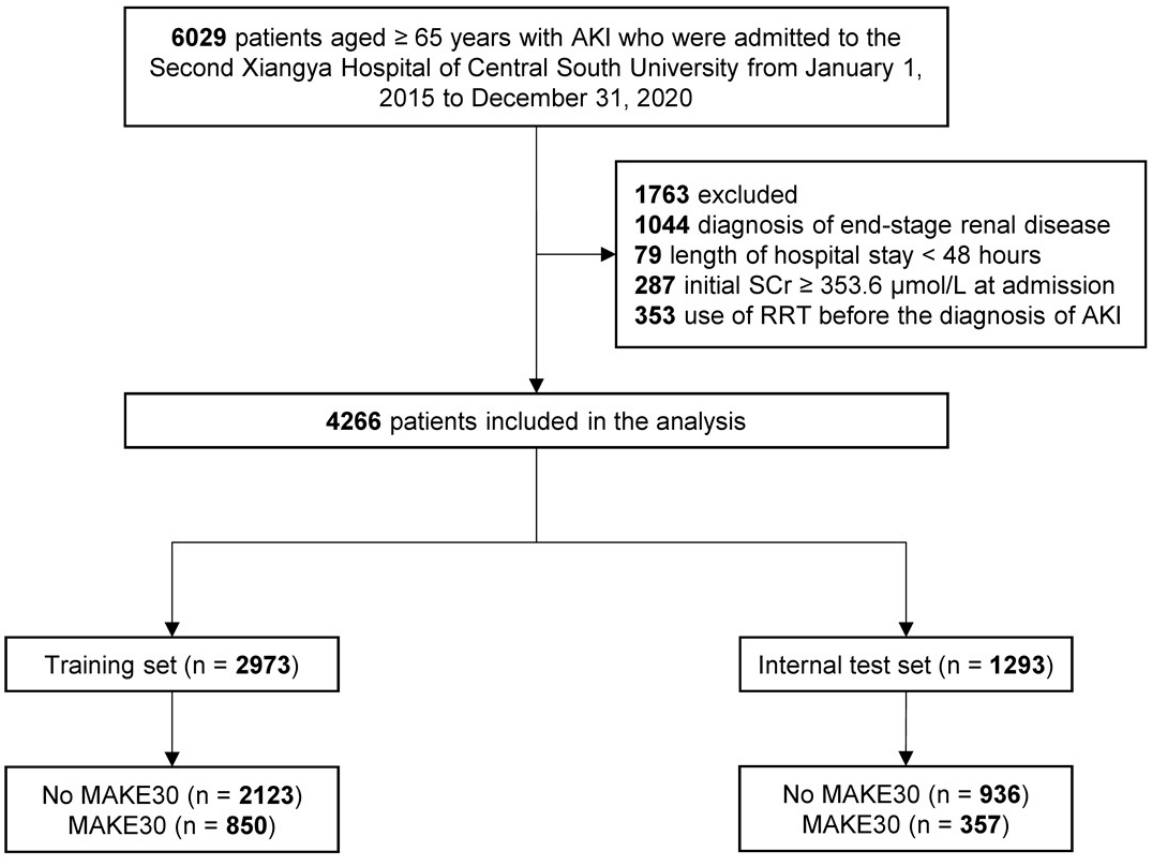

XSL•FO
RenderX

Table 1. Characteristics of patients from the Second Xiangya Hospital.

Variables	No MAKE30[a] (n=3059)	MAKE30[a] (n=1207)	*P* value
Age (years), median (IQR)	72 (68-77)	74 (68-80)	<.001
Sex, male, n (%)	1992 (65.1)	776 (64.3)	.64
Intensive care, n (%)	1284 (42.0)	765 (63.4)	<.001
Comorbidities, n (%)			
Sepsis	127 (4.2)	237 (19.6)	<.001
Hypertension	1685 (55.1)	676 (56.0)	.61
Myocardial infarction	346 (11.3)	170 (14.1)	.01
Congestive heart failure	932 (30.5)	502 (41.6)	<.001
Peripheral vascular disease	455 (14.9)	202 (16.7)	.14
Cerebrovascular disease	611 (20.0)	347 (28.7)	<.001
Dementia	45 (1.5)	41 (3.4)	<.001
Chronic pulmonary disease	515 (16.8)	259 (21.5)	<.001
Rheumatic disease	60 (2.0)	27 (2.2)	.65
Peptic ulcer disease	99 (3.2)	59 (4.9)	.01
Liver disease	630 (20.6)	308 (25.5)	.001
Diabetes	756 (24.7)	349 (28.9)	.005
Hemiplegia or paraplegia	10 (0.3)	3 (0.2)	.91
Renal disease	449 (14.7)	210 (17.4)	.03
Malignancy	913 (29.8)	296 (24.5)	.001
HIV/AIDS	2 (0.1)	0 (0.0)	.92
Charlson Comorbidity Index	2 (1-4)	3 (2-4)	<.001
Laboratory tests			
Red blood cells (×10^9/L), median (IQR)	3.4 (2.9-3.9)	3.2 (2.7-3.8)	<.001
Hemoglobin (g/L), median (IQR)	103 (86-118)	96 (80-114)	<.001
RDW-CV[b] (%), median (IQR)	14.0 (13.1-15.5)	15.0 (13.9-16.9)	<.001
White blood cells (×10^9/L), median (IQR)	9.8 (7.1-13.6)	11.4 (7.5-16.3)	<.001
Neutrophil percentage (%), median (IQR)	83.8 (75.6-89.3)	87.6 (80.8-91.9)	<.001
Lymphocyte percentage (%), median (IQR)	9.2 (5.6-15.1)	7.0 (4.1-11.9)	<.001
Platelets (×10^9/L), median (IQR)	175 (124-236)	147 (90-223)	<.001
Serum total protein (g/L), median (IQR)	58.4 (52.9-63.7)	56.5 (50.5-61.9)	<.001
Serum albumin (g/L), median (IQR)	32.3 (29.1-35.6)	30.0 (26.0-33.7)	<.001
Serum total bilirubin (μmol/L), median (IQR)	11.6 (7.9-18.4)	13.0 (8.0-23.4)	<.001
Serum direct bilirubin (μmol/L), median (IQR)	5.2 (3.3-8.6)	6.7 (3.9-14.1)	<.001
Alanine aminotransferase (U/L), median (IQR)	19.0 (11.5-34.6)	26.1 (13.4-68.2)	<.001
Aspartate aminotransferase (U/L), median (IQR)	26.5 (18.1-48.1)	42.3 (22.8-129.7)	<.001
Serum creatinine (μmol/L), median (IQR)	131.9 (97.7-186.7)	159.0 (109.4-240.2)	<.001
Blood urea nitrogen (mmol/L), median (IQR)	10.36 (7.16-15.36)	15.76 (10.79-22.96)	<.001
Blood uric acid (μmol/L), median (IQR)	331.3 (239.9-443.3)	385.2 (264.7-516.6)	<.001
Potassium (mmol/L), median (IQR)	4.2 (3.8-4.6)	4.3 (3.8-4.9)	<.001
Sodium (mmol/L), median (IQR)	139.5 (136.4-142.6)	141.0 (136.3-146.7)	<.001
Chloride (mmol/L), median (IQR)	103.2 (99.0-106.9)	103.1 (98.5-107.9)	.69

XSL•FO
RenderX

Variables	No MAKE30[a] (n=3059)	MAKE30[a] (n=1207)	*P* value
Calcium (mmol/L), median (IQR)	2.07 (1.96-2.20)	2.03 (1.90-2.17)	<.001
Mechanical ventilation, n (%)	442 (14.4)	564 (46.7)	<.001
Medications, n (%)			
Vasopressors	521 (17.0)	660 (54.7)	<.001
Diuretics	1699 (55.5)	849 (70.3)	<.001
ACEI[c]/ARB[d]	457 (14.9)	149 (12.3)	.03
NSAIDs[e]	1221 (39.9)	310 (25.7)	<.001
Proton pump inhibitors	2308 (75.4)	881 (73.0)	.10
Chemotherapeutic drugs	45 (1.5)	20 (1.7)	.76
Antiepileptic drugs	108 (3.5)	30 (2.5)	.10
Antituberculosis drugs	16 (0.5)	7 (0.6)	>.99
Nephrotoxic antibiotics	223 (7.3)	203 (16.8)	<.001
Antiviral drugs	41 (1.3)	35 (2.9)	.001
Antifungal drugs	186 (6.1)	232 (19.2)	<.001
Iodinated contrast media	159 (5.2)	50 (4.1)	.17
Outcomes			
Length of hospital stay (days), median (IQR)	8 (4-14)	7 (2-14)	<.001
Death within 30 days after discharge[f,g], n (%)	201 (6.8)	137 (21.5)	<.001
Death within 1 year after discharge[f,g], n (%)	637 (21.7)	247 (38.8)	<.001

[a]MAKE30: Major adverse kidney events within 30 days.

[b]RDW-CV: red blood cell distribution width-coefficient of variation.

[c]ACEI: angiotensin converting enzyme inhibitor.

[d]ARB: angiotensin II receptor blocker.

[e]NSAID: non-steroidal anti-inflammatory drug.

[f]There were 165 patients with missing postdischarge survival data.

[g]A total of 526 patients who died at hospital discharge or 30 days, whichever occurred first, were excluded.

Model Performance

Figure 2 shows the ROC curves of the model in the training set and internal test set. The model exhibited good discrimination on both datasets, with the AUROC being 0.868 (95% CI 0.852-0.881) and 0.823 (95% CI 0.798-0.846), respectively. Figure 3 shows the calibration plots of the model in the training set and internal test set. The Brier scores on the 2 datasets were 0.127 and 0.145, respectively. Notably, despite the lower Brier score on the training set, the calibration plot exhibits pronounced deviations from the diagonal, especially at higher probability thresholds. In contrast, the calibration plot for the internal test set appears to demonstrate more consistent alignment with the ideal diagonal line across various probability ranges. Figure 4 shows the precision-recall curves of the model in the training set and internal test set. The model achieved an AUPRC of 0.778 and 0.639 on the 2 datasets, respectively. Figure 5 shows the decision curves of the model in the training set and internal test set. When the threshold probability was between 0.06 and 0.97, the model improved clinical decision-making in the training set. Similarly, within a narrower threshold probability range of 0.09 to 0.76, the model improved clinical decision-making in the internal test set. At the optimal cutoff of 0.350, the model achieved a sensitivity of 67.5%, a specificity of 82.5%, a PPV of 59.5%, and a NPV of 86.9% in the internal test set (Table 2).

Multimedia Appendix 8 presents the performance of the simplified model in the training set, the internal test set, and the external test set. It includes ROC curves, calibration plots, precision-recall curves, decision curves, and model performance at the optimal cutoff. The simplified model achieved an AUROC of 0.851 (95% CI 0.834-0.866) on the training set, 0.818 (95% CI 0.792-0.843) on the internal test set, and 0.744 (95% CI 0.735-0.754) on the external test set.

XSL•FO
RenderX

Figure 2. Receiver operating characteristic curves of the model in the training set (A) and internal test set (B). AUC: area under the curve.

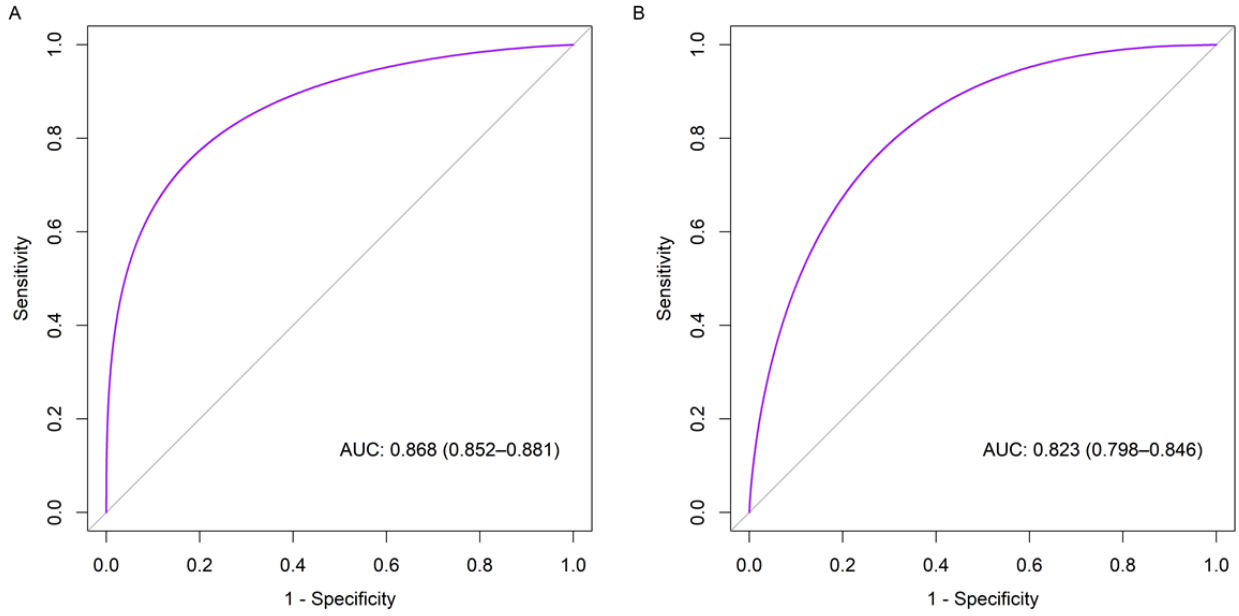

Figure 3. Calibration plots of the model in the training set (A) and internal test set (B).

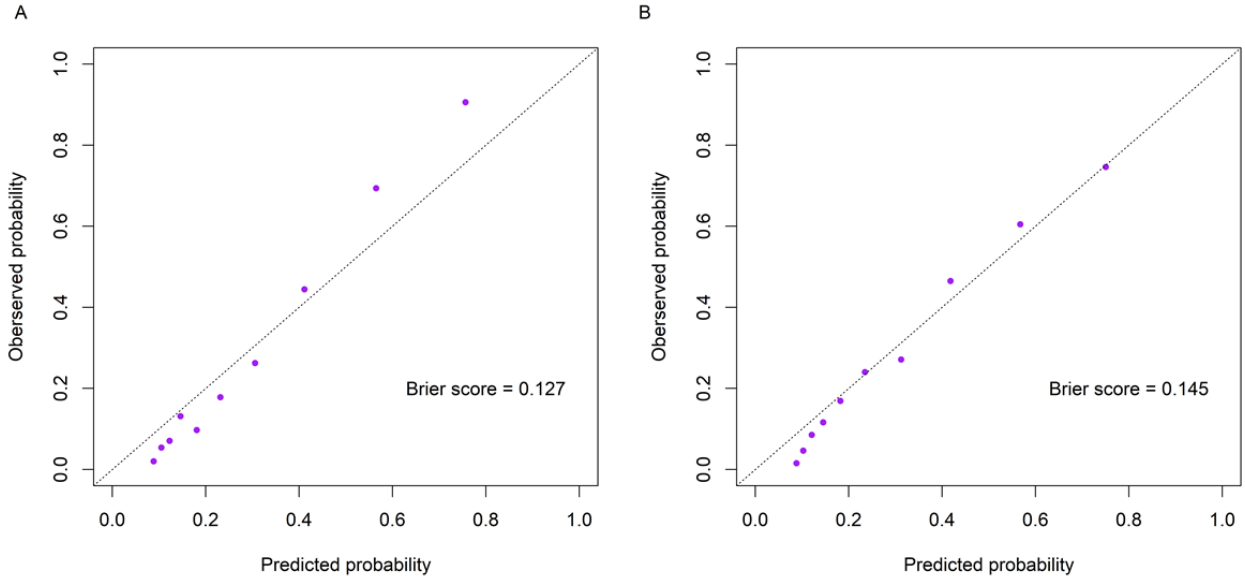

XSL•FO
RenderX

Figure 4. Precision-recall curves of the model in the training set (A) and internal test set (B). AUC: area under the curve.

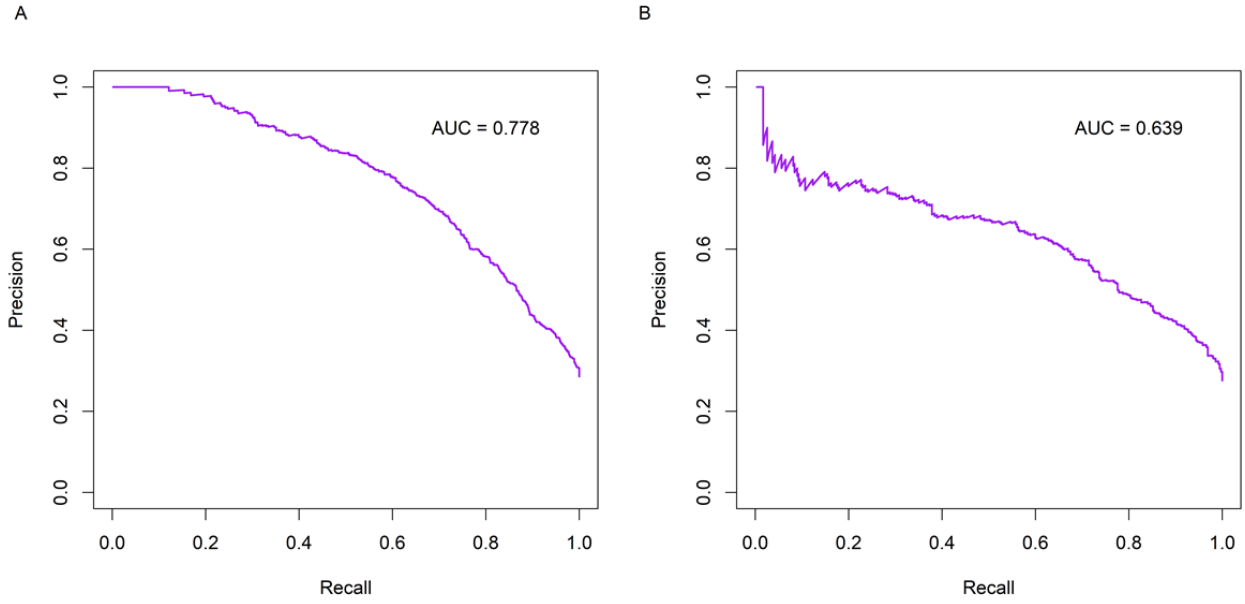

Figure 5. Decision curves of the model in the training set (A) and internal test set (B).

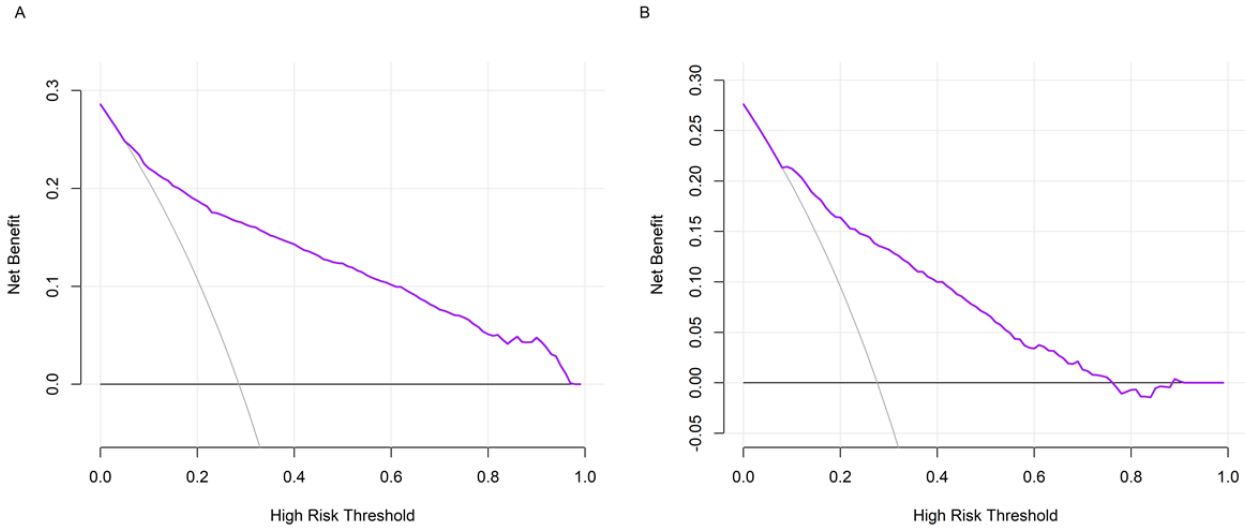

XSL•FO
RenderX

Table 2. Model performance at the optimal cutoff in the training and internal test sets.

Performance metrics	Training set	Internal test set
Cutoff value[a]	0.350	0.350
Sensitivity (%)	71.9	67.5
Specificity (%)	86.6	82.5
PPV[b] (%)	68.3	59.5
NPV[c] (%)	88.5	86.9
PLR[d]	5.37	3.86
NLR[e]	0.32	0.39

[a]The optimal cutoff was determined by the receiver operating characteristic curve and the maximum Youden index (sensitivity + specificity – 1) in the training set.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]PLR, positive likelihood ratio.

[e]NLR, negative likelihood ratio.

Model Interpretations

Figure 6 shows the SHAP summary plots of the model. The use of vasopressors, requirement for mechanical ventilation, increased blood urea nitrogen level, increased red blood cell distribution width coefficient of variation, decreased serum albumin level, elevated aspartate aminotransferase level, use of antifungal drugs, and increased SCr level were identified as important predictor variables associated with an increased risk of MAKE30. Multimedia Appendix 9 provides the SHAP dependence plots of the model that visualize how changes in each variable can affect model output.

Figure 6. SHapley Additive exPlanations summary plot of the model. RDW-CV: red blood cell distribution width-coefficient of variation.

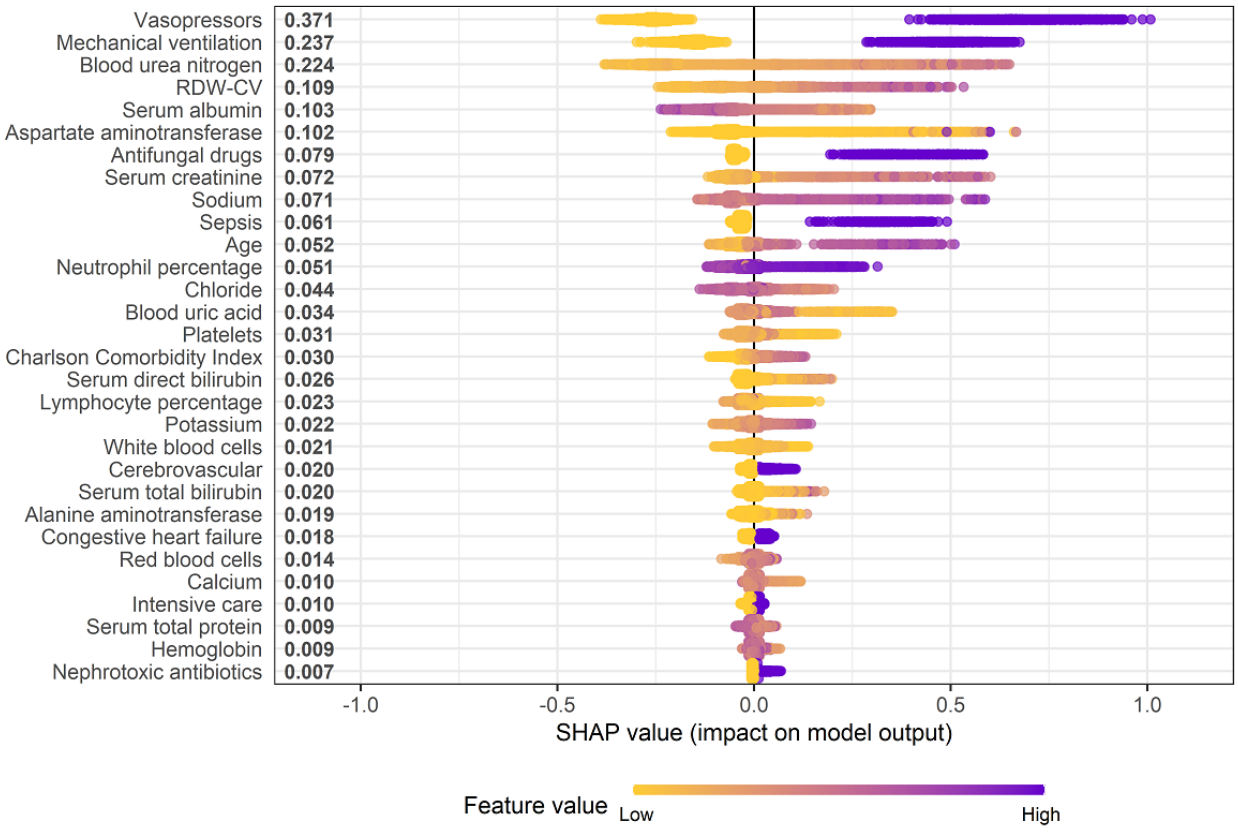

Sensitivity Analyses

Multimedia Appendix 10 shows the ROC curves of the model and the simplified model in distinct age groups. In the internal test set, the model achieved AUROCs of 0.827 (95% CI 0.786-0.862), 0.835 (95% CI 0.794-0.869), and 0.770 (95% CI 0.708-0.837) in patients aged <70 years, 70 to 80 years, and ≥80 years, respectively. When assessing the simplified model

XSL•FO
RenderX

across distinct age groups in the external test set, its performance exhibited overall stability.

Discussion

Principal Findings

This study showed that MAKE30 is common in hospitalized older patients with AKI. The main deliverable of this study is the development and validation of a machine learning–based model to predict MAKE30, representing a pioneering effort to address the critical need in this vulnerable population. The XGBoost model (or its simplified version) achieved adequate predictive performance in both internal and external validation. The model could be a useful tool for prognostic assessment, clinical decisionmaking, and the conduct of clinical trials involving AKI.

A National Institute of Diabetes and Digestive and Kidney Diseases Workshop recommended a composite of death, provision of dialysis, or sustained loss of kidney function at a discrete time point as a meaningful endpoint for trials [9]. This endpoint was later expanded to be MAKE30, which occurred at hospital discharge truncated at 30 days [34,35]. MAKE30 has been shown to be common, easily measurable, and a promising endpoint for patients with AKI. Our study further confirmed that MAKE30 is a prevalent endpoint among hospitalized older patients with AKI, with its nonfatal components being associated with long-term outcomes.

The prediction of MAKE30 is essential for risk stratification and clinical management of hospitalized older patients with AKI. While several prediction models for MAKE30 have been developed, none specifically focus on this population. A recent study developed a logistic regression model specifically for predicting MAKE30 among critically ill adults [36]. Our previous studies established machine learning–based models to predict MAKE30 in hospitalized children with AKI [37] and older patients in critical care [38]. To the best of our knowledge, this study is the first to use machine learning algorithm to predict MAKE30 in hospitalized older patients with AKI. This demographic often experiences impaired recovery of kidney function, resulting in significantly increased morbidity and mortality. By developing a prediction model tailored to this population, we aimed to enhance clinical outcomes and facilitate more effective interventions. The model can be seamlessly integrated into electronic medical record systems, assisting health care providers in risk stratification and informed clinical decision-making. In addition, in clinical trials involving AKI, the model can identify high-risk groups that are more likely to benefit from the intervention.

Compared with the existing studies, this study has several strengths. First, by using state-of-the-art XGBoost machine learning algorithm, our model represented a significant advancement over previous studies that primarily relied on clinical rationale and simpler statistical methods. XGBoost is particularly effective due to its ability to process large amounts of data and identify intricate patterns within complex datasets. Its gradient boosting framework enhances predictive accuracy by combining the strengths of multiple weak learners, leading to more robust and reliable outcomes. Furthermore, XGBoost's flexibility allows for easy tuning of hyperparameters, enabling us to optimize the model for our specific dataset and improve its performance in predicting MAKE30 in hospitalized older patients with AKI.

Second, this study used a larger sample size and conducted external validation to evaluate the model's performance. Notably, the results of the external validation showed that the simplified version of the model exhibited a gap of approximately 10% in AUROC between the training set and the external test set. However, the performance of the simplified model on the internal test set was comparable to that of the full model, suggesting that the observed gap may be attributed to the use of a different test set with distinct characteristics. Compared with the training set, the external test set comprised critically ill patients from various ethnic backgrounds and with distinct patient characteristics and data quality. These findings underscore the necessity for comprehensive testing across diverse datasets to ensure the model's reliability and applicability in real-world settings. Another noteworthy observation is that, although the Brier scores for both the training and test data suggested acceptable calibration, a deeper examination through visual inspection uncovered disparities. Factors contributing to the observed discrepancies may include overfitting during training, differences in underlying data distributions, model complexity, or nuances in evaluation methodology. Further investigation into the factors contributing to the observed discrepancies is warranted.

Finally, this study delved into the interpretability of the model. A major challenge in the clinical implementation of machine learning is how to uncover its "black box" nature. Usually, data come in and decisions go out, but the process between input and output is opaque. The advantage of our study is the use of SHAP method to explain the critical aspects of the data and fully understand the model. The SHAP method identified important predictor variables associated with the development of MAKE30, which could allow early intervention of modifiable factors to mitigate the risk of MAKE30. Another notable finding is the nonlinear relationship between the predictor variables and the risk of MEKE30, as shown in the SHAP dependence plots. This relationship is often ignored by traditional regression analysis, which requires a linearity between the independent variables and the outcome.

Limitations

Our study has several limitations. First, the MIMIC-Ⅳ database only contained data from critically ill patients admitted to the ICU. Model performance awaits further validation in larger samples of patients at different medical centers. Second, it remains unclear whether the model will perform well in individual prognostication and whether its clinical application will improve patient prognosis. Thus, clinical impact studies are needed to confirm the effectiveness of the model. Third, the unavailability of preadmission SCr data could lead to the omission of patients who could have been diagnosed with AKI, that is, community-acquired AKI. Fourth, the urine output criteria were not used to define AKI because hourly urine output data were unavailable for most patients. Finally, the missForest

XSL•FO
RenderX

method, as a single imputation method, might affect the SHAP results by not considering sufficient variation for the imputation of missing data.

Conclusions

MAKE30 is common in hospitalized older patients with AKI. An interpretable machine learning XGBoost model was developed and validated to predict MAKE30. The model exhibited adequate predictive performance, which provides opportunities for risk stratification, clinical decision-making, and the conduct of clinical trials involving AKI. Future studies are needed to support the robustness and clinical effectiveness of the model.

Authors' Contributions

SBD designed and supervised the study and revised the manuscript. XQL performed the data extraction, analyzed and interpreted the data and drafted the manuscript. NYZ performed the data extraction and revised the manuscript critically for important intellectual content. YHD, HSW, and YXK interpreted the data and critically revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The list of 53 candidate predictor variables and 30 variables identified in feature selection.
[DOCX File , 50 KB - jmir_v27i1e52786_app1.docx]

Multimedia Appendix 2
Missing data in predictor variables.
[DOCX File , 55 KB - jmir_v27i1e52786_app2.docx]

Multimedia Appendix 3
The descriptions and search ranges for the hyperparameters of the XGBoost (eXtreme Gradient Boosting) model.
[DOCX File , 52 KB - jmir_v27i1e52786_app3.docx]

Multimedia Appendix 4
Flow diagram of patient selection from the MIMIC-Ⅳ (Medical Information Mart for Intensive Care Ⅳ) database.
[DOCX File , 603 KB - jmir_v27i1e52786_app4.docx]

Multimedia Appendix 5
Kaplan-Meier survival curves of death within 30 days and 1 year for patients alive at discharge from the Second Xiangya Hospital.
[DOCX File , 1248 KB - jmir_v27i1e52786_app5.docx]

Multimedia Appendix 6
Characteristics of patients from the MIMIC-Ⅳ (Medical Information Mart for Intensive Care Ⅳ) database.
[DOCX File , 55 KB - jmir_v27i1e52786_app6.docx]

Multimedia Appendix 7
Characteristics of patients in the training set, internal test set, and external test set.
[DOCX File , 56 KB - jmir_v27i1e52786_app7.docx]

Multimedia Appendix 8
Performance of the simplified model in the training set, internal test set, and external test set.
[DOCX File , 2620 KB - jmir_v27i1e52786_app8.docx]

Multimedia Appendix 9
SHAP (SHapley Additive exPlanations) dependence plots of the model.
[DOCX File , 1842 KB - jmir_v27i1e52786_app9.docx]

Multimedia Appendix 10
Receiver operating characteristic curves of the model and the simplified model in distinct age groups.

XSL•FO
RenderX

[DOCX File , 1300 KB - jmir_v27i1e52786_app10.docx]

References

1. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders H. Acute kidney injury. Nat Rev Dis Primers 2021 Jul 15;7(1):52. [doi: 10.1038/s41572-021-00284-z] [Medline: 34267223]
2. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical Decision Support for In-Hospital AKI. J Am Soc Nephrol 2018 Feb;29(2):654-660 [FREE Full text] [doi: 10.1681/ASN.2017070765] [Medline: 29097621]
3. Hoste EAJ, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 2015 Aug;41(8):1411-1423. [doi: 10.1007/s00134-015-3934-7] [Medline: 26162677]
4. Ronco C, Bellomo R, Kellum JA. Acute kidney injury. Lancet 2019 Nov 23;394(10212):1949-1964. [doi: 10.1016/S0140-6736(19)32563-2] [Medline: 31777389]
5. Chronopoulos A, Cruz DN, Ronco C. Hospital-acquired acute kidney injury in the elderly. Nat Rev Nephrol 2010 Mar;6(3):141-149. [doi: 10.1038/nrneph.2009.234] [Medline: 20125094]
6. Stallone G, Infante B, Grandaliano G. Acute kidney injury in the elderly population. J Nephrol 2012;25 Suppl 19:S58-S66. [doi: 10.5301/jn.5000140] [Medline: 22760878]
7. Chang-Panesso M. Acute kidney injury and aging. Pediatr Nephrol 2021 Oct;36(10):2997-3006 [FREE Full text] [doi: 10.1007/s00467-020-04849-0] [Medline: 33411069]
8. Schmitt R, Coca S, Kanbay M, Tinetti ME, Cantley LG, Parikh CR. Recovery of kidney function after acute kidney injury in the elderly: a systematic review and meta-analysis. Am J Kidney Dis 2008 Aug;52(2):262-271. [doi: 10.1053/j.ajkd.2008.03.005] [Medline: 18511164]
9. Palevsky PM, Molitoris BA, Okusa MD, Levin A, Waikar SS, Wald R, et al. Design of clinical trials in acute kidney injury: report from an NIDDK workshop on trial methodology. Clin J Am Soc Nephrol 2012 May;7(5):844-850. [doi: 10.2215/CJN.12791211] [Medline: 22442182]
10. Semler MW, Rice TW, Shaw AD, Siew ED, Self WH, Kumar AB, et al. Identification of Major Adverse Kidney Events Within the Electronic Health Record. J Med Syst 2016 Jul;40(7):167 [FREE Full text] [doi: 10.1007/s10916-016-0528-z] [Medline: 27234478]
11. Heung M, Steffick DE, Zivin K, Gillespie BW, Banerjee T, Hsu CY, Centers for Disease ControlPrevention CKD Surveillance Team. Acute Kidney Injury Recovery Pattern and Subsequent Risk of CKD: An Analysis of Veterans Health Administration Data. Am J Kidney Dis 2016 May;67(5):742-752 [FREE Full text] [doi: 10.1053/j.ajkd.2015.10.019] [Medline: 26690912]
12. Fiorentino M, Tohme FA, Wang S, Murugan R, Angus DC, Kellum JA. Long-term survival in patients with septic acute kidney injury is strongly influenced by renal recovery. PLoS One 2018;13(6):e0198269 [FREE Full text] [doi: 10.1371/journal.pone.0198269] [Medline: 29870535]
13. De Corte W, Dhondt A, Vanholder R, De Waele J, Decruyenaere J, Sergoyne V, et al. Long-term outcome in ICU patients with acute kidney injury treated with renal replacement therapy: a prospective cohort study. Crit Care 2016 Aug 12;20(1):256 [FREE Full text] [doi: 10.1186/s13054-016-1409-z] [Medline: 27520553]
14. Kirwan CJ, Blunden MJ, Dobbie H, James A, Nedungadi A, Prowle JR. Critically ill patients requiring acute renal replacement therapy are at an increased risk of long-term renal dysfunction, but rarely receive specialist nephrology follow-up. Nephron 2015;129(3):164-170. [doi: 10.1159/000371448] [Medline: 25765730]
15. Self WH, Semler MW, Wanderer JP, Wang L, Byrne DW, Collins SP, SALT-ED Investigators. Balanced Crystalloids versus Saline in Noncritically Ill Adults. N Engl J Med 2018 Mar 01;378(9):819-828 [FREE Full text] [doi: 10.1056/NEJMoa1711586] [Medline: 29485926]
16. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, SMART Investigatorsthe Pragmatic Critical Care Research Group. Balanced Crystalloids versus Saline in Critically Ill Adults. N Engl J Med 2018 Mar 01;378(9):829-839 [FREE Full text] [doi: 10.1056/NEJMoa1711584] [Medline: 29485925]
17. Weiss SL, Balamuth F, Thurm CW, Downes KJ, Fitzgerald JC, Laskin BL. Major Adverse Kidney Events in Pediatric Sepsis. Clin J Am Soc Nephrol 2019 May 07;14(5):664-672 [FREE Full text] [doi: 10.2215/CJN.12201018] [Medline: 31000518]
18. Labarère J, Renaud B, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med 2014 Apr;40(4):513-527. [doi: 10.1007/s00134-014-3227-6] [Medline: 24570265]
19. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
20. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care 2019 Apr 08;23(1):112 [FREE Full text] [doi: 10.1186/s13054-019-2411-z] [Medline: 30961662]
21. Chen T, Li X, Li Y, Xia E, Qin Y, Liang S, et al. Prediction and Risk Stratification of Kidney Outcomes in IgA Nephropathy. Am J Kidney Dis 2019 Sep;74(3):300-309. [doi: 10.1053/j.ajkd.2019.02.016] [Medline: 31031086]
22. Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. J Transl Med 2022 May 13;20(1):215 [FREE Full text] [doi: 10.1186/s12967-022-03364-0] [Medline: 35562803]

XSL•FO
RenderX

23. Luo XQ, Kang YX, Duan SB, Yan P, Song GB, Zhang NY, et al. Machine Learning-Based Prediction of Acute Kidney Injury Following Pediatric Cardiac Surgery: Model Development and Validation Study. J Med Internet Res 2023 Jan 05;25:e41142 [FREE Full text] [doi: 10.2196/41142] [Medline: 36603200]
24. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guideline for acute kidney injury. Kidney International Supplements 2012 Mar;2(1):1-138 [FREE Full text] [doi: 10.1038/kisup.2012.1]
25. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV (version 1.0). PhysioNet. 2021. URL: https://physionet.org/content/mimiciv/1.0/ [accessed 2024-11-23]
26. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
27. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [doi: 10.1097/01.mlr.0000182534.19832.83] [Medline: 16224307]
28. Xu X, Nie S, Zhang A, Mao J, Liu HP, Xia H, et al. Acute Kidney Injury among Hospitalized Children in China. Clin J Am Soc Nephrol 2018 Dec 07;13(12):1791-1800 [FREE Full text] [doi: 10.2215/CJN.00800118] [Medline: 30287424]
29. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 01;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
30. Kursa MB, Rudnicki WR. Feature selection with boruta package. J. Stat. Soft 2010;36(11):1-13. [doi: 10.18637/jss.v036.i11]
31. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. : Association for Computing Machinery; 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13-17; San Francisco, California, USA p. 785-794. [doi: 10.1145/2939672.2939785]
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Adv Neural Inf Process Syst. Red Hook, New York: Curran Associates, Inc; 2017.
33. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015 Jan 06;162(1):55-63 [FREE Full text] [doi: 10.7326/M14-0697] [Medline: 25560714]
34. Kashani K, Al-Khafaji A, Ardiles T, Artigas A, Bagshaw SM, Bell M, et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. Crit Care 2013 Feb 06;17(1):R25 [FREE Full text] [doi: 10.1186/cc12503] [Medline: 23388612]
35. Hoste EAJ, McCullough PA, Kashani K, Chawla LS, Joannidis M, Shaw AD, Sapphire Investigators. Derivation and validation of cutoffs for clinical use of cell cycle arrest biomarkers. Nephrol Dial Transplant 2014 Nov;29(11):2054-2061 [FREE Full text] [doi: 10.1093/ndt/gfu292] [Medline: 25237065]
36. McKown AC, Wang L, Wanderer JP, Ehrenfeld J, Rice TW, Bernard GR, et al. Predicting Major Adverse Kidney Events among Critically Ill Adults Using the Electronic Health Record. J Med Syst 2017 Aug 31;41(10):156 [FREE Full text] [doi: 10.1007/s10916-017-0806-4] [Medline: 28861688]
37. Deng YH, Luo XQ, Yan P, Zhang NY, Liu Y, Duan SB. Outcome prediction for acute kidney injury among hospitalized children via eXtreme Gradient Boosting algorithm. Sci Rep 2022 May 27;12(1):8956 [FREE Full text] [doi: 10.1038/s41598-022-13152-x] [Medline: 35624143]
38. Wang L, Duan SB, Yan P, Luo XQ, Zhang NY. Utilization of interpretable machine learning model to forecast the risk of major adverse kidney events in elderly patients in critical care. Ren Fail 2023 Dec;45(1):2215329 [FREE Full text] [doi: 10.1080/0886022X.2023.2215329] [Medline: 37218683]

Abbreviations

AKI: Acute kidney injury
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
BIDMC: Beth Israel Deaconess Medical Center
EHR: electronic health record
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Statistical Classification of Diseases, Tenth Revision
ICU: intensive care unit
MAKE30: Major adverse kidney events within 30 days
MIMIC-Ⅳ: Medical Information Mart for Intensive Care Ⅳ
NLR: negative likelihood ratio
NPV: negative predictive value
PLR: positive likelihood ratio
PPV: positive predictive value

XSL•FO
RenderX

PRD: persistent renal dysfunction
ROC: receiver operating characteristic
RRT: renal replacement therapy
SCr: serum creatinine
SHAP: SHapley Additive exPlanations
XGBoost: eXtreme Gradient Boosting

Edited by A Mavragani; submitted 15.09.23; peer-reviewed by M Rezk, D Nelson, Q Yang, X Zhang; comments to author 26.02.24; revised version received 18.04.24; accepted 12.11.24; published 03.01.25.

Please cite as:
Luo XQ, Zhang NY, Deng YH, Wang HS, Kang YX, Duan SB
Major Adverse Kidney Events in Hospitalized Older Patients With Acute Kidney Injury: Machine Learning–Based Model Development and Validation Study
J Med Internet Res 2025;27:e52786
URL: https://www.jmir.org/2025/1/e52786
doi:10.2196/52786
PMID:

©Xiao-Qin Luo, Ning-Ya Zhang, Ying-Hao Deng, Hong-Shen Wang, Yi-Xin Kang, Shao-Bin Duan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Remote Patient Monitoring Equipped With an Early Warning System in Tertiary Care Hospital Wards: Retrospective Cohort Study

Pavithra Lakshman[1], MBBS, MD; Priyanka T Gopal[1], MBBS; Sheen Khurdi[1], MBBS

Hospital Administration, Ramaiah Memorial Hospital, Bengaluru, Karnataka, India

Corresponding Author:
Pavithra Lakshman, MBBS, MD
Hospital Administration
Ramaiah Memorial Hospital
New BEL Rd, M S Ramaiah Nagar
MSRIT Post
Bengaluru, Karnataka, 560054
India
Phone: 91 9741592241
Email: dr.pavithra.lakshman@gmail.com

Abstract

Background: Monitoring vital signs in hospitalized patients is crucial for evaluating their clinical condition. While early warning scores like the modified early warning score (MEWS) are typically calculated 3 to 4 times daily through spot checks, they might not promptly identify early deterioration. Leveraging technologies that provide continuous monitoring of vital signs, combined with an early warning system, has the potential to identify clinical deterioration sooner. This approach empowers health care providers to intervene promptly and effectively.

Objective: This study aimed to assess the impact of a Remote Patient Monitoring System (RPMS) with an automated early warning system (R-EWS) on patient safety in noncritical care at a tertiary hospital. R-EWS performance was compared with a simulated Modified Early Warning System (S-MEWS) and a simulated threshold-based alert system (S-Threshold).

Methods: Patient outcomes, including intensive care unit (ICU) transfers due to deterioration and discharges for nondeteriorating cases, were analyzed in Ramaiah Memorial Hospital's general wards with RPMS. Sensitivity, specificity, chi-square test for alert frequency distribution equality, and the average time from the first alert to ICU transfer in the last 24 hours was determined. Alert and patient distribution by tiers and vitals in R-EWS groups were examined.

Results: Analyzing 905 patients, including 38 with deteriorations, R-EWS, S-Threshold, and S-MEWS generated more alerts for deteriorating cases. R-EWS showed high sensitivity (97.37%) and low specificity (23.41%), S-Threshold had perfect sensitivity (100%) but low specificity (0.46%), and S-MEWS demonstrated moderate sensitivity (47.37%) and high specificity (81.31%). The average time from initial alert to clinical deterioration was at least 18 hours for RPMS and S-Threshold in deteriorating participants. R-EWS had increased alert frequency and a higher proportion of critical alerts for deteriorating cases.

Conclusions: This study underscores R-EWS role in early deterioration detection, emphasizing timely interventions for improved patient outcomes. Continuous monitoring enhances patient safety and optimizes care quality.

(J Med Internet Res 2025;27:e56463) doi:10.2196/56463

KEYWORDS

continuous vitals monitoring; remote patient monitoring; early warning system; hospital wards; retrospective; cohort study; early deterioration monitoring; patient care; decision making; clinical information

Introduction

Ensuring patient safety is vital in health care, whether in regular wards or critical care units. Swift detection of vital sign changes allows prompt adjustments in care levels. Delayed intervention for patient deterioration is linked to increased morbidity and mortality [1,2]. Hospitals use aggregate score based early warning systems (EWS) to detect vital sign deterioration. However, these rely on instantaneous measurements conducted hours apart [3], lack historical trends, and have sensitivity <80%

XSL•FO
RenderX

for patient deterioration [4-6]. EWS based on continuous monitoring of the vital signs of hospitalized patients is expected to overcome the constraints of conventional, intermittent vital measurements. Multiple studies have shown that continuous vital signs monitoring is pivotal in identifying at-risk patients and facilitating timely interventions [7-14]. An automated system based on continuous monitoring of hospitalized patients' vital signs in combination with EWS besides enhancing patient safety, has the potential to save health care practitioners (HCPs) time [11,15-17] mitigate errors [18], facilitate information sharing with relevant stakeholders for informed decision-making [19], and reduce costs [20].

To ensure patient safety, and enhance health care services, Ramaiah Memorial Hospital (RMH), has implemented an advanced commercial remote patient monitoring system (RPMS) in its wards. This system is designed for continuous monitoring of vital signs, featuring an EWS (R-EWS). Before the introduction of R-EWS, ward monitoring depended on manual spot checks every 4 hours to calculate the modified early warning score (MEWS), while continuous monitoring in intensive care units (ICUs) relied on a threshold-based alerting system. The implementation of the R-EWS significantly increased the time available for patient care by reducing the time nurses spent on routine monitoring, communication, and coordination. Majority of HCPs reported improvements in the level of care and overall patient safety, with many also noting an enhanced patient experience [21].

After several months of implementing R-EWS, the hospital sought to assess the system's performance against previous monitoring practices. To this end, a retrospective analysis was conducted to evaluate its effectiveness in identifying patient deterioration. The performance of R-EWS was also benchmarked against simulated versions of MEWS (S-MEWS) and the ICU's continuous threshold-based alert system (S-Threshold). The primary objective of this study was to determine how well the R-EWS, S-Threshold, and S-MEWS systems ensure patient safety by accurately distinguishing between deteriorating patients needing ICU transfer and stable patients ready for discharge.

Methods

Materials

The RPMS is a continuous, vital parameters monitoring system. It uses sensor sheets placed under the mattress that uses ballistocardiography technology to capture microvibrations from which heart rate (HR), respiratory rate (RR), and systolic blood pressure (SysBP) are computed in a noncontact manner [22-25]. Accessories like a pulse oximeter and temperature probe are used for contact-based measurements of oxygen saturation (SpO_2) and temperature. The system incorporates a configurable and customizable multitiered EWS (R-EWS) for severity assessment, complemented by a user-friendly web dashboard and mobile app for centralized and remote ward monitoring. The hospital leverages this R-EWS for patient monitoring, reporting, and clinical decision-making.

Study Design

This is a retrospective cohort study, conducted without blinding. It uses data from patients under R-EWS monitoring implemented in the general wards of RMH, a tier 1 city's private hospital with over 500 beds. The study covers the period from December 1, 2022, to May 31, 2023. This study retrospectively analyzed the performance of the R-EWS in distinguishing between patients requiring ICU transfer and those stable enough for discharge. The study also retrospectively compared the R-EWS performance against 2 simulated frameworks: the S-MEWS and a S-Threshold. The vital signs data generated by the RPMS served as input for these simulated frameworks for the alert generation. Detailed descriptions of these frameworks are provided in the following subsections. To reduce bias all patients who satisfied the inclusion and exclusion criteria were used for analysis.

Ethical Considerations

This study was approved by the RMH ethics committee (DRP/EFPl034/2023).

R-EWS Framework

The R-EWS uses a dynamic 3-tiered alert (tier 1, tier 2, and tier 3) system to prioritize alerts based on the severity of vital signs deviations. This ensures that alerts are raised only when a patient's condition deteriorates significantly. Each tier in the alert system is defined by specific threshold settings for monitored vital signs, as detailed in Table 1. The R-EWS focuses on observing trends over a set period rather than relying on isolated values. This observation window helps confirm sustained deviations in vital signs before triggering an alert. In addition, the tiers incorporate a cool-down period during which no alerts are generated, unless a higher tier is breached. This mechanism prevents frequent and redundant alerts. For tiers 1 and 2, the cool-down period is set at 3 hours. In contrast, tier 3 has no cool-down period and generates alerts every 10 minutes if the patient's condition continues to meet the threshold criteria. This ensures that critical conditions are continuously monitored, allowing for immediate medical intervention when necessary.

XSL•FO
RenderX

Table 1. Vital signs thresholds for each alert tier in the remote patient monitoring early warning system framework, showing specific values for heart rate, respiratory rate, and blood pressure that trigger tier 1, tier 2, and tier 3 alerts, illustrating a stratified approach to patient monitoring.

Alert level	Tier 1 range	Tier 2 range	Tier 3 range
HR[a]-contactless (beats per minute)			
High	120 < HR[a] ≤ 129	130 < HR[a] ≤ 139	HR[a] ≥ 140
Low	40 >HR[a] ≥ 36	—[b]	HR[a] ≤ 35
RR[c]-contactless (breaths per minute)			
High	29 < RR[c] ≤ 34	35 < RR[c] ≤ 39	RR[c] ≥ 40
Low	8 >RR[c] ≥ 7	—[b]	RR[c] ≤ 6
SysBP[d]-contactless (mm Hg)			
High	157 < SysBP[d] ≤ 178	—[b]	SysBP[d] ≥ 179
Low	102 >SysBP[d] ≥ 82	—[b]	SysBP[d] ≤ 81
SpO_2[e]-contact accessory (%)			
Low	92 >SpO_2[e] ≥ 91	—[b]	SpO_2[e] ≤ 90
T[f]-contact accessory (°C)			
High	38.33 < T[f] ≤ 38.88	—[b]	T[f] > 38.89
Low	33 >T[f] ≥ 33.88	—[b]	T[f] < 33.87

[a]HR: heart rate.

[b]Not applicable.

[c]RR: respiratory rate.

[d]SysBP: systolic blood pressure.

[e]SpO_2: oxygen saturation.

[f]T: temperature.

S-MEWS Framework

The S-MEWS simulations used vital sign values from the ranges defined in Table 2, excluding clinical signs such as urine output and mental status. To reduce bias, S-MEWS used the same vital signs scoring criteria as the MEWS previously used in the hospital, with scores calculated 4 times a day (12 AM, 6 AM, 12 PM, and 6 PM). The hospital-initiated notification and escalation protocols for a threshold of 4 or higher. Similarly, S-MEWS was calculated every 4 hours using a 1-hour median calculation for all vitals, triggering a positive alert for a score of 4 or higher. The vital signs data generated by the RPMS served as input for this simulated framework for alert generation.

Table 2. Vital signs threshold and the scoring system for simulated modified early warning system framework.

Parameters	3	2	1	0	1	2	3
Respiratory rate	—[a]	8≥	8< and ≤14	14< and <21	21≤ and <25	25≤ and <31	≤31
Heart rate	—[a]	40≥	40< and ≤60	60< and ≤100	100< and <111	111≤ and <130	≤131
Blood pressure	69≥	70≤ and ≤80	80< and ≤110	110< and <140	140≤ and <170	170≤ and <200	≤200
Oxygen saturation	91≥	91< and ≤93	93< and ≤95	95< to 100	—[a]	—[a]	—[a]
Temperature	—[a]	35≥	35< and ≤36	36< and ≤36.9	36.9< and ≤38	38< and ≤38.6	≤38.7

[a]—: not available.

S-Threshold Framework

This EWS simulates traditional ICU monitors that provide continuous monitoring and trigger alarms whenever a specific vital sign exceeds its threshold. These single breach point thresholds can typically be muted for up to 2 minutes. The breach points for various vitals are as follows: High HR >120, Low HR <40, High RR >29, Low RR <8, High SysBP >157, Low SysBP <102, Low SpO_2 <92, High *t*>38.33, and Low *t*<36. To ensure comparability with R-EWS, these threshold values correspond to the tier 1 breach points of R-EWS detailed in Table 1. Similar to S-MEWS, the vital signs data generated by

XSL•FO
RenderX

the RPMS served as input for these simulated frameworks for alert generation.

Study Participants

The study examined retrospective data from individuals who met the specific inclusion and exclusion criteria outlined as follows (Textbox 1).

Textbox 1. Eligibility criteria for study participants.

Inclusion criteria

- All adult patients monitored on Remote Patient Monitoring System (RPMS) system from December 01, 2022, to May 31, 2023.
- Patients with complete intake and discharge information from hospital records.

Exclusion criteria

- All adult patients who were not monitored on the RPMS system.
- Patients aged <18 years.
- Patients whose data were less than 4 hours.

Variables

The 2 study groups were deteriorating patients defined as requiring transfer to the ICU and nondeteriorating patients deemed stable for discharge from the hospital without ICU admission. The total, mean, and median number of alerts in the last 24 hours were calculated for each group for the 3 EWS frameworks. In addition, the average time from the first alert to deterioration was calculated for each framework within the final 24 hours. Specifically for the R-EWS framework, the proportions of patients receiving alerts and the number of alerts at various tiers and combination of vital alerts in both groups during the last 24 hours were computed.

Sensitivity and specificity values were calculated by classifying deteriorating participants with even one alert as true positives, deteriorating participants without alerts as false negatives, nondeteriorating participants without alerts as true negatives, and nondeteriorating participants with even one alert as false positives.

Data Analysis

Descriptive statistics were used to provide the demographic alert counts of the study groups. It is expected that there may be some differences in demographics and significant difference in proportion of participants in each group as it represents the real patient population where only 4%-5% are expected to require advanced care.

A chi-square test assumed equal alert frequency distribution in both groups, with statistical significance indicated by $P<.05$. Using 6 degrees of freedom and a 1-tailed threshold of 12.592, the test assessed if the RPMS system effectively discriminates by generating more alerts in deteriorating patients than nondeteriorating ones.

Results

From December 01, 2022, to May 31, 2023, 922 patients in general wards were monitored using RPMS. For this study, 905 patients met the predefined inclusion and exclusion criteria. Among them, 38 patients experienced deteriorations and were transferred to the ICU. The demographic information of the deteriorating and nondeteriorating group is given in Table 3. The age, gender ratio, and the average length of stay between the 2 groups are similar except the proportion of patients in each group. The median length of stay is 3 days.

Table 4 displays the descriptive alert statistics for the 2 groups across various frameworks. In addition, it illustrates the earliness of each framework in generating alerts for deteriorating group. Across different EWS the deteriorating group consistently showed higher average alerts per patient and median alerts compared with the nondeteriorating group. Specifically, in the R-EWS groups, the deteriorating group had 2.64 times higher alerts, S-Threshold had 2.69 times higher alerts, and S-MEWS had 3.4 times higher alerts. Continuous monitoring based on R-EWS and S-Threshold, produced more alerts for both deteriorating and nondeteriorating groups than the intermittent monitoring S-MEWS. The first alert to deterioration in continuous monitoring systems occurred at least 18 hours before ICU transfer, compared with 11 hours in intermittent monitoring.

XSL•FO
RenderX

Table 3. Demographic details of the participants included for analysis.

Parameters	Deteriorating	Nondeteriorating
Total number of patients, n (%)	38 (4)	867 (96)
Age (years), mean (SD) (range)	64.1 (14.28) (29-90)	54.81 (18.17) (18-98)
Gender number, n (%)		
Male	22 (57.89)	527 (60.78)
Female	16 (42.11)	340 (39.22)
Average length of stay (hours), mean (SD) (range)	95.6 (98.03) (5-448)	94 (112.36) (4-1817)
Total hours of monitoring, n	3632	81510

Table 4. Alert frequencies comparison between the frameworks.

Group	Alerts total	Alerts per patients, mean (SD)	Alerts, median (range)	Time of first alert to ICU[a] Transfer (hours), mean (SD)
R-EWS[b]				
D[c]	1402	36.89 (42.41)	17 (0-147)	18.7 (5.78)
ND[d]	12100	13.96 (29.66)	3 (0-257)	—[e]
S-Threshold[f]				
D	7031	185.03 (201.63)	102 (6-803)	20.1 (5.48)
ND	59479	68.6 (98.64)	31 (0-940)	—
S-MEWS[g]				
D	44	1.16 (1.6)	0 (0-5)	11.1 (8.53)
ND	298	0.34 (0.88)	0 (0-6)	N/A[h]

[a]ICU: intensive care unit.

[b]R-EWS: remote patient monitoring early warning system.

[c]D: deteriorating.

[d]ND: nondeteriorating.

[e]Not available.

[f]S-Threshold: simulated threshold-based alert system.

[g]S-MEWS: simulated modified early warning system.

[h]N/A: not applicable.

In Figure 1, for R-EWS, a depiction of 24-hour cumulative alerts is presented, illustrating the generated data from 5 participants in each category, those undergoing deterioration and those experiencing nondeterioration. The cumulative alerts were generated for the last 24 hours that lead to either a discharge or ICU transfer. These alerts were calculated every hour for the past 24 hours. The figure demonstrates the importance and discriminatory nature of the frequency of alert patterns in R-EWS for both groups. It clearly shows that more than 10 alerts in a day are primarily seen in patients whose conditions worsened and were subsequently transferred to the ICU, rather than in those who were discharged home.

XSL•FO
RenderX

Figure 1. Representative 24-hour cumulative alerts generated in 5 participants each of those experiencing deterioration and nondeterioration in the R-EWS.

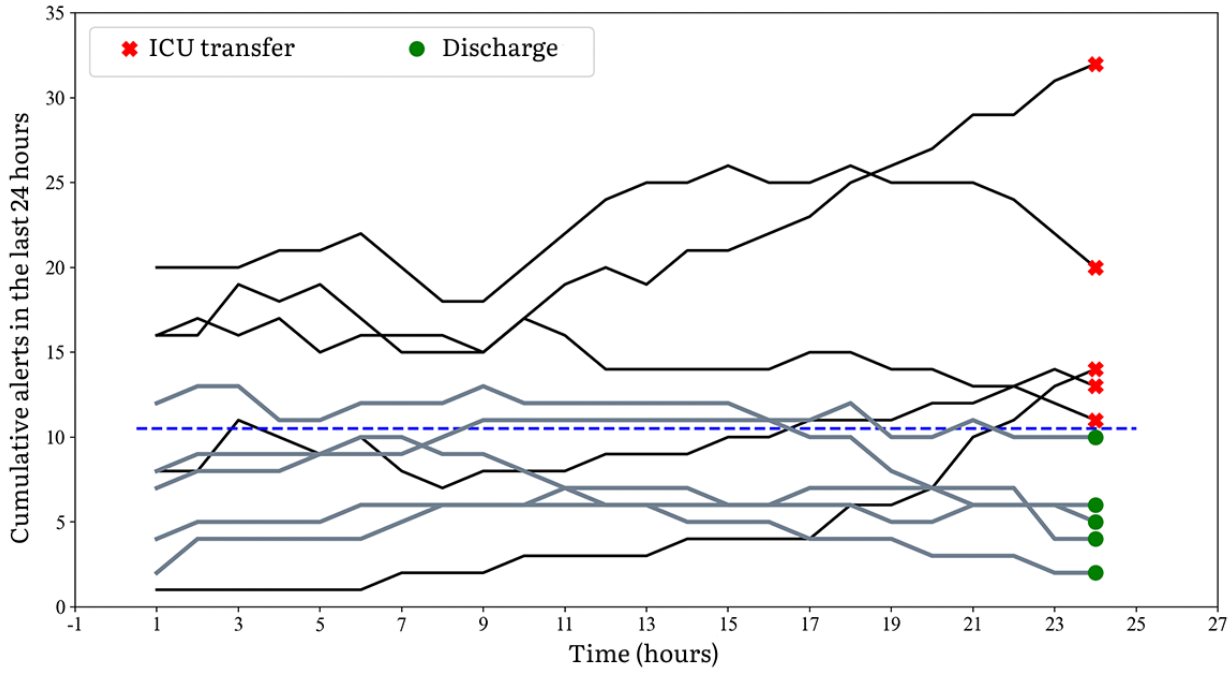

The specificity and sensitivity of the different frameworks across the time points are given and the chi-square analysis is given in Table 5. R-EWS demonstrated high sensitivity (97.37%) but lower specificity (23.41%). S-Threshold achieved perfect sensitivity (100%) but had very low specificity (0.46%). In contrast, S-MEWS had moderate sensitivity (47.37%) and the highest specificity (81.31%), balancing fewer false positives with a lower detection rate. All frameworks showed statistically significant discriminative abilities (P<.001).

Table 5. Sensitivity and specificity and statistical test for discriminatory nature of different early warning systems frameworks and specificity.

Framework	Sensitivity (%) (TP[a]/TP[a]+FN[b])	Specificity (%) (TN[c]/TN[c] + FP[d])	Chi-squared Value	*P* value
R-EWS[e]	97.37 (37/[37+1])	23.41 (203/[203+664])	27.71	<.001
S-Threshold[f]	100 (38/[38+0])	0.46 (4/[4+863])	44.7	<.001
S-MEWS[g]	47.37 (18/[18+20])	81.31 (705/[705+162])	39.2	<.001

[a]TP: true positive.

[b]FN: false negative.

[c]TN: true negative.

[d]FP: false positive.

[e]R-EWS: remote patient monitoring early warning system.

[f]S-Threshold: simulated threshold-based alert system.

[g]S-MEWS: simulated modified early warning system.

Table 6 presents data regarding the tier-wise alert numbers and the proportion of patients who received these alerts in both the deteriorating and nondeteriorating categories in the R-EWS framework. The deteriorating group received a significantly higher proportion of alerts across, especially in tier 2 and 3, compared with the normal discharge group. In tier 1, the deteriorating group had 1.36% high HR alerts, 3.35% high RR alerts, and 14.34% low SpO_2 alerts, whereas the nondeteriorating group had 0.85% high HR alerts, 4.35% high RR alerts, and 26.05% low SpO_2 alerts. In tiers 2 and 3, the deteriorating group shows consistently higher alert proportions, especially for critical parameters like SpO_2. In tier 3, the deteriorating group received 73.61% low SpO_2 alerts, compared with 54.84% in the nondeteriorating group. In addition, 26.32% of the deteriorating participants received high HR alerts in tier 1 versus 5.88% of the nondeteriorating participants, and 42.11% received high RR alerts versus 4.35%. Similarly, in tier 3, 76.32% of the deteriorating participants received low SpO_2 alerts compared with 41.41% of nondeteriorating participants.

XSL•FO
RenderX

Table 6. Alert and patient distribution by tiers of the R-EWS[a].

Study group	High HR[b]	Low HR	High RR[c]	Low RR	High SysBP[d]	Low SysBP	Low SpO_2[e]
Tier 1							
Category: alerts/total (%)							
D[f]	19/1402 (1.36)	0/1402 (0)	47/1402 (3.35)	0/1402 (0)	11/1402 (0.78)	23/1402 (1.64)	201/1402 (14.34)
ND[g]	103/12100 (0.85)	0/12100 (0)	526/12100 (4.35)	18/12100 (0.15)	382/12100 (3.16)	562/12100 (4.64)	3152/12100 (26.05)
Category: patient/total (%)							
D	10/38 (26.32)	0/38 (0)	16/38 (42.11)	0/38 (0)	3/38 (7.89)	8/38 (21.05)	14/38 (36.84)
ND	51/867 (5.88)	0/867 (0)	536/867 (4.35)	16/867 (1.85)	130/867 (14.99)	198/867 (22.84)	306/867 (35.29)
Tier 2							
Category: alerts/total (%)							
D	12/1402 (0.86)	—[h]	10/1402 (1.36)	—	—	—	—
ND	32/12100 (0.26)	—	102/12100 (0.84)	—	—	—	—
Category: patient/total (%)							
D	7/38 (18.42)	—	8/38 (21.05)	—	—	—	—
ND	22/867 (2.54)	—	60/867 (6.92)	—	—	—	—
Tier 3							
Category: alerts/total (%)							
D	41/1402 (2.92)	0/1402 (0)	6/1402 (0.43)	0/1402 (0)	0/1402 (0)	0/1402 (0)	1032/1402 (73.61)
ND	297/12100 (2.45)	0/12100 (0)	73/12100 (0.60)	18/12100 (0.15)	189/12100 (1.56)	21/12100 (0.17)	6636/12100 (54.84)
Category: patient/total (%)							
D	5/38 (13.6)	0/38 (0)	2/38 (5.26)	0/38 (0)	0/38 (0)	0/38 (0)	29/38 (76.32)
ND	13/867 (1.50)	0/867 (0)	19/867 (2.19)	3/867 (0.35)	17/867 (1.96)	6/867 (0.69)	359/867 (41.41)

[a]R-EWS: remote patient monitoring early warning system.

[b]HR: heart rate.

[c]RR: respiratory rate.

[d]SysBP: systolic blood pressure.

[e]SpO_2: oxygen saturation.

[f]D: deteriorating.

[g]ND: nondeteriorating.

[h]Not applicable.

Table 7 displays information on the vital alert numbers, both individually and in combination, along with the percentage of patients who received these alerts in both the deteriorating and nondeteriorating categories within the R-EWS group. Deteriorating group had a higher percentage of alerts across all vital signs compared with the nondeteriorating group. In total, 87.95% of alerts in the deteriorating group were low SpO_2, compared with 80.89% in the nondeteriorating group In addition, combinations of alerts, such as HR and SpO_2 and RR and SpO_2, were significantly more prevalent in the deteriorating group. Notably, 26.32% of ICU transfer patients had HR and SpO_2 alerts, whereas only 5.07% of normal discharge patients experienced these alerts. The presence of all 4 vital signs alerts (HR, RR, SysBP, and SpO_2) is higher (21.05%) in the deteriorating group compared with only 2.31% in the nondeteriorating group. As a single parameter, HR generated the fewest alerts (504 alerts in total) but had the highest discriminatory nature, with alerts in escalated patients being 4.7 times more likely than in nonescalated patients.

XSL•FO
RenderX

Table 7. Alert and patient distribution by vitals for the R-EWS[a] framework.

Parameter	Alerts/total alerts (%)		Patients/total patients (%)	
	D[b]	ND[c]	D	ND
HR[d]	72/1402 (5.14)	432/12100 (3.57)	13/38 (34.21)	63/867 (7.27)
RR[e]	63/1402 (4.49)	726/12100 (6.00)	16/38 (42.11)	248/867 (28.60)
SysBP[f]	34/1402 (2.43)	1154/12100 (9.54)	11/38 (28.95)	328/867 (37.83)
SpO_2[g]	1233/1402 (87.95)	9788/12100 (80.89)	31/38 (81.58)	443/867 (51.10)
HR + RR	135/1402 (9.63)	1158/12100 (9.57)	6/38 (15.79)	38/867 (4.38)
HR + SysBP	106/1402 (7.56)	1586/12100 (13.11)	3/38 (7.89)	34/867 (3.92)
HR + SpO_2	1305/1402 (93.08)	10220/12100 (84.46)	10/38 (26.32)	44/867 (5.07)
RR + SysBP	34/1402 (2.43)	1880/12100 (15.54)	5/38 (13.16)	109/867 (12.57)
RR + SpO_2	1296/1402 (92.44)	10514/12100 (86.89)	12/38 (31.58)	144/867 (16.61)
SysBP + SpO_2	1267/1402 (90.37)	10942/12100 (90.43)	8/38 (21.05)	180/867 (20.76)
HR + RR + SysBP	169/1402 (12.05)	2312/12100 (19.11)	2/38 (5.26)	26/867 (3.00)
HR + RR + SpO_2	1368/1402 (97.57)	10946/12100 (90.46)	4/38 (10.53)	29/867 (3.34)
RR + SysBP + SpO_2	1330/1402 (94.86)	11668/12100 (96.43)	3/38 (7.89)	70/867 (8.07)
HR + RR + SysBP + SpO_2	1402/1402 (100.00)	12100/12100 (100.00)	8/38 (21.05)	20/867 (2.31)

[a]R-EWS: remote patient monitoring early warning system.

[b]D: deteriorating.

[c]ND: nondeteriorating.

[d]HR: heart rate.

[e]RR: respiratory rate.

[f]SysBP: systolic blood pressure.

[g]SpO_2: oxygen saturation.

Figure 2 illustrates trends in vitals and alerts generated by the R-EWS in the last 24 hours for both a deteriorating and a nondeteriorating participant. The deteriorating participant received numerous alerts before ICU transfer. Conversely, the nondeteriorating participant experienced minimal alerts.

XSL•FO
RenderX

Figure 2. Representative 24-hour vital patterns depicting generated alerts in the remote patient monitoring early warning system framework (A) deteriorating and (B) nondeteriorating participants. BP: blood pressure; bpm: beats per minute; brpm: breaths per minute; SpO2: oxygen saturation.

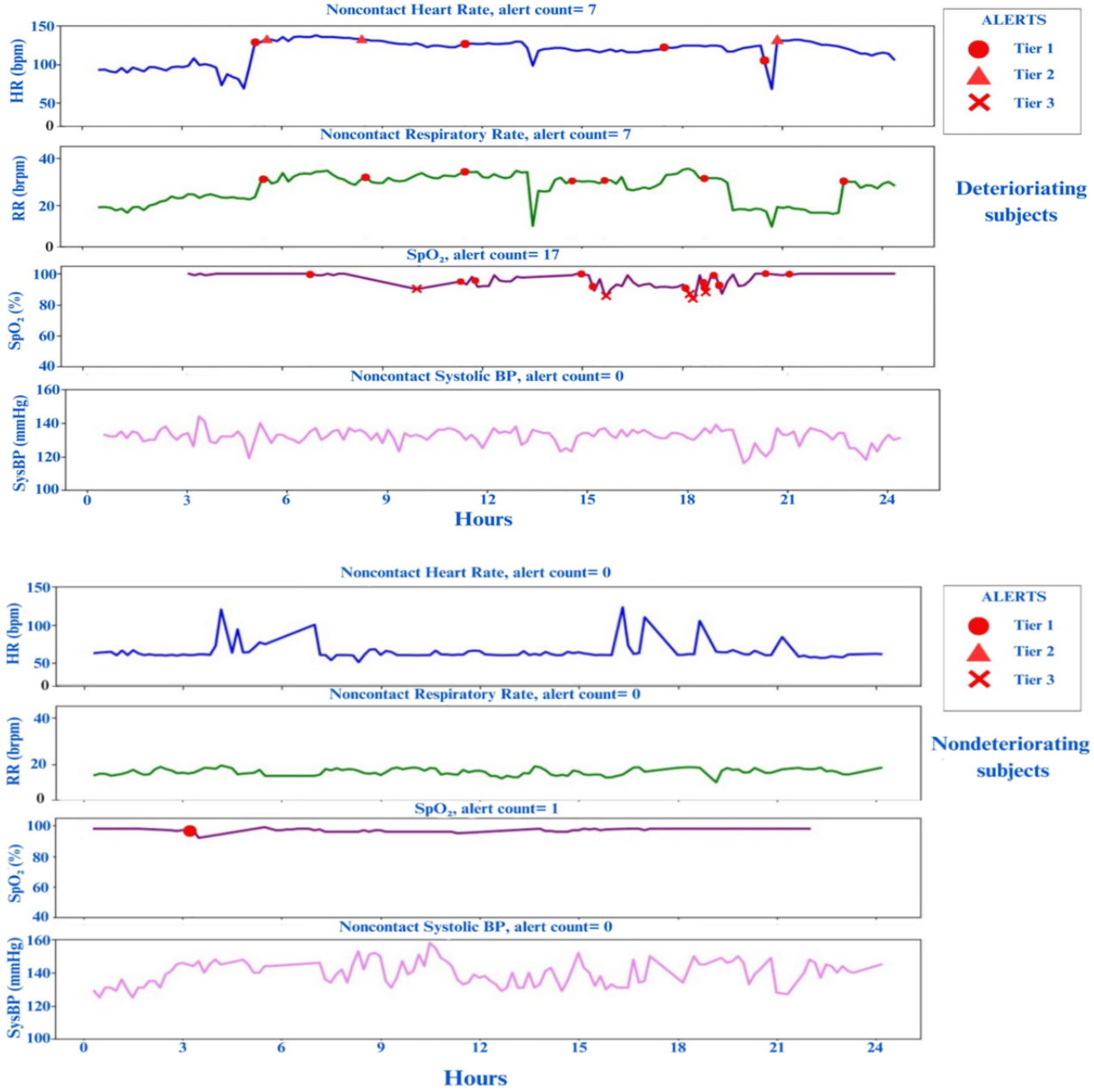

Discussion

Principal Findings

This study aimed to evaluate the effectiveness of the R-EWS in ensuring patient safety by identifying deterioration in patients. Our findings reveal that R-EWS demonstrated a high sensitivity in identifying patients who were deteriorating and in need of ICU transfer. This was achieved with a relatively lower specificity from a single alert. In R-EWS, ICU transfer patients received a higher number of alerts, including alerts classified as critical, compared with their normal discharge counterparts.

As seen in Table 3, both groups had similar demographics though vastly different group sizes (38 deteriorating and 867 nondeteriorating). This is consistent with the expectation of ICU transfers or Death in wards to be 2%-6% [26]. Continuous monitoring identifies more patients developing complications than intermittent monitoring. This is supported by data in Table 5 where both R-EWS and S-Threshold, based on continuous monitoring, demonstrated a high sensitivity in alerting to potential patient deterioration. This is in stark contrast to the S-MEWS system, which, relying on intermittent hourly monitoring, showed a sensitivity of only 47%, a rate not clinically acceptable as it would miss deterioration in the majority of patients. These results align with existing literature indicating that continuous monitoring is more effective than intermittent and manual monitoring in early identification of complications and improving patient outcomes [27-29]. There is also established research showing that EWS based on 3- to 6-hour spot checks, a standard of care practice in most hospitals, can have sensitivity as low as 40% [4,5]. While the sensitivity of S-Threshold was 100% the specificity was nearly zero,

XSL•FO
RenderX

indicating a high rate of false alerts for nondeteriorating patients. Alerting frequently on every patient is as ineffective as alerting on no patient, as these alarms are likely to be ignored and lead to alarm fatigue for health care workers. In this context, the R-EWS, with its balanced approach to minimal alerts with high sensitivity and moderate specificity emerges as a better choice.

The study highlights that continuous monitoring systems generate more alerts than intermittent systems as shown in Table 4 due to their higher frequency of measurements, a trend also observed in the scientific literature [29]. This study found that the intermittent monitoring-based S-MEWS generated 342 alerts which is 39 times less alerts than the 13,502 alerts generated in R-EWS and 194 times less than the 66,510 S-Threshold. By using a tier-based alerting with a cool down period R-EWS produced 5 times less alerts than S-Threshold. This reduction in alert number is important because alarm fatigue can occur when health care professionals become desensitized to the constant stream of alerts, potentially leading to slower response times or even missed alarms. This overload not only increases the workload but also pushes health care professionals toward burnout, a serious issue highlighted by Kristinsson et al [30] in 2022. The findings indicate that, on average, R-EWS generates 15 alerts per patient per 24 hours. This is a more manageable workload for health care professionals compared with the overwhelming 73 alerts per patient per 24 hours generated by the S-Threshold system. Considering it takes approximately 4 mins to assess an alert, addressing every alert would lead to spending 4.9 hours or over 50% of a standard shift on alerts which is impractical. In the case of R-EWS, an average of 15 alerts are generated per patient in 24 hours resulting in HCP spending about an hour collectively which is reasonable. In addition, this time is made up through the time saved from needing to manually check vitals and document and report the same for every patient.

The study also highlights that continuous monitoring systems provide alerts earlier than intermittent systems. As demonstrated in Table 4, one of the significant findings of our study is the ability of the R-EWS and S-Threshold systems to provide alerts an average of at least 18 hours before an ICU transfer. This represents a substantial improvement over the S-MEWS, which provides alerts an average of only 11 hours in advance. These results are consistent with what has been reported in the literature. For instance, continuous vital sign monitoring has been shown to warn of upcoming deterioration 8.3 hours in advance compared with 5.2 hours for periodic EWS [29]. Early detection of clinical deterioration enables timely and effective interventions, which not only improves patient outcomes but also has the potential to reduce associated health care costs [20,31].

The study highlights the effectiveness of the R-EWS system in distinguishing between patients at risk of deterioration and those with stable conditions. The discriminatory nature of the R-EWS is further evidenced by its trend of increased alerts in the deteriorating group compared with the nondeteriorating group seen in Table 4. This increased number of alerts in the deteriorating group aligns well with what is reported in the literature, where a higher frequency of alerts is typically associated with patients experiencing clinical deterioration [22].

R-EWS also demonstrated, as seen in Table 6, a higher proportion of patients receiving critical alerts (tier 2 and tier 3) among those experiencing deterioration compared with those without any escalations in care. Figure 2 shows the importance and discriminatory nature of the frequency of alert patterns in a random sampling from each group (deterioration and nondeterioration). In the last 24 hours health care providers can make more informed decisions based on whether a patient is improving or worsening. As shown in Figures 1 and 2, frequency of greater than 10 alerts in a day is primarily seen in patients whose conditions worsened and were transferred to the ICU, rather than in those who were discharged home. This can serve as a metric for daily evaluations and clinical decision support of discharge, observe further, or transfer or prepare for critical care.

Our findings as detailed in Table 7, highlight the critical importance of not only the volume but also the pattern and type of alerts. SpO_2 alerts combined with HR or RR alerts are twice as likely in deteriorating patients compared with nondeteriorating ones, The integration of SpO_2 alerts, especially in combination with HR or RR alerts, can provide a valuable early warning system for patient deterioration. SpO_2 alerts combined with HR or RR alerts are 5 times and 2 times as likely in deteriorating patients compared with nondeteriorating ones, Furthermore, while all four vital alerts appeared in only 21% of deteriorating patients, they appeared in just 2% of nondeteriorating patients, underscoring this pattern as a critical marker. These findings suggest that monitoring systems should not rely solely on the volume of alerts but should also consider the context and combination of these alerts. Therefore, the development of alert algorithms that take into account the interdependence of various vital signs may enhance the predictive value of these systems.

While the study did not specifically focus on the cost-benefit aspect, it is worth noting the RPMS's potential in reducing direct costs. Manual monitoring incurs high labor expenses and is prone to errors [32,33]. RPMS can lower these costs by automating monitoring and reducing HCP workload, particularly at this hospital with a 1:6 HCP-to-patient ratio. Early detection of patient deterioration, even if it prevents only 2%-5% of ICU admissions, can lead to substantial savings. ICU stays and emergency interventions are costly, with ICU costs (US $132.5 per day) far exceeding general ward costs (US $25 per day [34,35]. Despite the initial investment, RPMS's cost-effectiveness in settings like this hospital is evident. It offers a viable solution for improving patient care and maintaining financial sustainability in Indian health care.

Mobilization is crucial for patient recovery. Traditional monitoring methods, which use wired sensors, restrict movement and are time consuming to set up, making mobilization difficult. The RPMS overcomes this by measuring 3 vital signs namely, HR, RR, and BP remotely, eliminating the need for tethering and allowing patients the freedom to move. During the study period, SpO_2 and temperature were measured with wired sensors; these can be easily removed and reattached, minimizing disruption. Thus, RPMS significantly enhances patient mobility while ensuring continuous monitoring. RPMS is now equipped

XSL•FO
RenderX

with wireless sensors for these measurements, which further enhances mobility without limitations. The contactless measurement of the 3 vital signs saves nursing time by eliminating the need to detach and reattach sensors each time the patient leaves and returns to bed. There is still a limitation in the current system because it only measures vitals when the patient is on the bed and not during ambulatory activities.

There are practical challenges associated with continuous monitoring for every patient that need consideration; high initial setup costs, the necessity for comprehensive staff training, and potential resistance to new technology to name a few. Integration with existing hospital systems can be complex, requiring technical expertise to ensure reliability and accuracy. Patient comfort and compliance are also crucial. Alarm fatigue from frequent alerts is another concern along with data privacy and security [36,37]. There is also a risk of overreliance on automation, which could lead to neglect of thorough patient assessments, deskilling which can lead to missing subtle signs of deterioration [38]. Addressing these challenges and fine-tuning RPMS in the future is necessary for the successful adoption and sustainability of RPMS in health care settings.

Limitations

The limitations of the study include the fact that the study wards had a diverse patient demographic. Such diversity, while reflective of a real-world scenario, could introduce variables that affect the generalizability of the findings. The comparison between the frameworks is unequal due to the inherent nature of each framework, which results in different sets of observations. The simulation of the MEWS only used vital signs, due to the unavailability of data on urine output and consciousness levels. The inclusion of these parameters could potentially have enhanced S-MEWS's sensitivity, though it does speak to the need for a more labor-intensive approach for MEWS compliance.

Conclusions

The study underscores the transformative impact of the RPMS using continuous monitoring and R-EWS alerting system on patient safety, emphasizing early deterioration. Despite certain limitations, the RPMS proves to be effective for improving patient safety, advocating for broader implementation. To further validate these findings, controlled trials are essential. In addition, more detailed costing data are crucial for a comprehensive cost of care analysis and to better understand the system's overall efficiency. Future studies should also involve collaboration with machine learning and artificial intelligence experts to optimize the system's specificity and sensitivity.

Authors' Contributions

Author PL contributed to study design, data analysis, drafting, and review of manuscript. PTG handled support for study design, data collection, data analysis, drafting, and review of manuscript. SK managed support for study design, data collection, data analysis, drafting, and review of manuscript.

Conflicts of Interest

None declared.

References

1. Churpek MM, Wendlandt B, Zadravecz FJ, Adhikari R, Winslow C, Edelson DP. Association between intensive care unit transfer delay and hospital mortality: A multicenter investigation. J Hosp Med 2016 Nov;11(11):757-762 [FREE Full text] [doi: 10.1002/jhm.2630] [Medline: 27352032]
2. Lau F, Bartle-Clar JA, Bliss G. Improving Usability, Safety and Patient Outcomes with Health Information Technology. USA: IOS Press; 2019.
3. Petersen JA, Antonsen K, Rasmussen LS. Frequency of early warning score assessment and clinical deterioration in hospitalized patients: a randomized trial. Resuscitation 2016;101:91-96 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.02.003] [Medline: 26891861]
4. Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. J Biomed Inform 2016 Dec;64:10-19 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.013] [Medline: 27658885]
5. Allarakia J, Felemban T, Alghamdi A, Ashi A, Al Talhi YM, Alsahafi A, et al. Modified Early Warning Score (MEWS) as a predictor of intensive care unit admission in cancer patient on chemotherapy with positive blood culture: A retrospective cohort study. J Infect Public Health 2023 Jun;16(6):865-869 [FREE Full text] [doi: 10.1016/j.jiph.2023.03.012] [Medline: 37031626]
6. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The Value of Modified Early Warning Score (MEWS) in Surgical In-Patients: A Prospective Observational Study. annals 2006 Oct;88(6):571-575. [doi: 10.1308/003588406x130615]
7. van Rossum MC, Bekhuis REM, Wang Y, Hegeman JH, Folbert EC, Vollenbroek-Hutten MMR, et al. Early Warning Scores to Support Continuous Wireless Vital Sign Monitoring for Complication Prediction in Patients on Surgical Wards: Retrospective Observational Study. JMIR Perioper Med 2023 Aug 30;6:e44483 [FREE Full text] [doi: 10.2196/44483] [Medline: 37647104]

XSL•FO
RenderX

8. Eddahchouri Y, Peelen RV, Koeneman M, Touw HRW, van Goor H, Bredie SJH. Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study. Br J Anaesth 2022 May;128(5):857-863 [FREE Full text] [doi: 10.1016/j.bja.2022.01.036] [Medline: 35282866]
9. Verrillo SC, Cvach M, Hudson KW, Winters BD. Using Continuous Vital Sign Monitoring to Detect Early Deterioration in Adult Postoperative Inpatients. J Nurs Care Qual 2019;34(2):107-113. [doi: 10.1097/NCQ.0000000000000350] [Medline: 30095509]
10. Downey CL, Chapman S, Randell R, Brown JM, Jayne DG. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
11. Aasvang EK, Meyhoff CS. The future of postoperative vital sign monitoring in general wards: improving patient safety through continuous artificial intelligence-enabled alert formation and reduction. Curr Opin Anaesthesiol 2023 Dec 01;36(6):683-690. [doi: 10.1097/ACO.0000000000001319] [Medline: 37865847]
12. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. Journal of Clinical Nursing 2015 Nov 04;25(1-2):278-281 [FREE Full text] [doi: 10.1111/jocn.13102] [Medline: 26531215]
13. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
14. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. Am J Med 2014 Mar;127(3):226-232. [doi: 10.1016/j.amjmed.2013.12.004] [Medline: 24342543]
15. Bellomo R, Ackerman M, Bailey M, Beale R, Clancy G, Danesh V, et al. A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards*. Critical Care Medicine 2012;40(8):2349-2361. [doi: 10.1097/ccm.0b013e318255d9a0]
16. Sigvardt E, Grønbaek KK, Jepsen ML, Søgaard M, Haahr L, Inácio A, et al. Workload associated with manual assessment of vital signs as compared with continuous wireless monitoring. Acta Anaesthesiol Scand 2024 Feb;68(2):274-279 [FREE Full text] [doi: 10.1111/aas.14333] [Medline: 37735843]
17. van der Stam JA, Mestrom EHJ, Nienhuijs SW, de Hingh IHJT, Boer AK, van Riel NAW, et al. A wearable patch based remote early warning score (REWS) in major abdominal cancer surgery patients. Eur J Surg Oncol 2023 Jan;49(1):278-284 [FREE Full text] [doi: 10.1016/j.ejso.2022.08.034] [Medline: 36085116]
18. Smith LB, Banner L, Lozano D, Olney CM, Friedman B. Connected care: reducing errors through automated vital signs data upload. Comput Inform Nurs 2009;27(5):318-323. [doi: 10.1097/NCN.0b013e3181b21d65] [Medline: 19726926]
19. Subbe CP, Weichert J, Duller B. Using trends in electronic recordings of vital signs to identify patients stable for transfer from acute hospitals. Acute Med 2019;18(4):216-222. [Medline: 31912052]
20. Javanbakht M, Mashayekhi A, Trevor M, Rezaei Hemami M, L Downey C, Branagan-Harris M, et al. Cost utility analysis of continuous and intermittent versus intermittent vital signs monitoring in patients admitted to surgical wards. J Med Econ 2020 Jul;23(7):728-736 [FREE Full text] [doi: 10.1080/13696998.2020.1747474] [Medline: 32212979]
21. Khurdi S, Mary S. Impact of Remote Patient Monitoring Systems on Nursing Time, Healthcare Providers, and Patient Satisfaction in General Wards. Cureus 2024 Jun;16(6):e61646 [FREE Full text] [doi: 10.7759/cureus.61646] [Medline: 38966455]
22. Kedia S, Ur Rahman I, Kaushal A, Mahizhvannan E, Jain RS, Radhakrishna R. Clinical Validation of an Indigenous Micro-Vibration Vital Parameter Monitor Dozee VS. 2023 Presented at: 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS); January 03-08, 2023; Bangalore, India. [doi: 10.1109/COMSNETS56262.2023.10041343]
23. Saran V, Kumar G, Parchani G. Unsupervised Extraction of Respiration Cycles Through Ballistocardiography. In: Advanced Informatics for Computing Research. United States: Springer, Singapore; 2019:147.
24. Saran V, Kumar R, Kumar G, Chokalingam K, Rawooth M, Parchani G. Validation of dozee, a ballistocardiography-based device, for contactless and continuous heart rate and respiratory rate measurement. 2022 Presented at: Conf Proc IEEE Eng Med Biol Soc; 2022-July 15; United Kingdom p. 1939. [doi: 10.1109/embc48229.2022.9871007]
25. Saran V, Kumar G, Dhawan U, Parchani G. Unsupervised Identification of Cardiac Contraction through Ballistocardiography. 2018 Presented at: IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2018 December 06; Sarawak, Malaysia. [doi: 10.1109/IECBES.2018.8626652]
26. Yadav A, Dandu H, Parchani G, Chokalingam K, Kadambi P, Mishra R, et al. Early detection of deteriorating patients in general wards through continuous contactless vital signs monitoring. Front Med Technol 2024;6:1436034 [FREE Full text] [doi: 10.3389/fmedt.2024.1436034] [Medline: 39328308]
27. Sun L, Joshi M, Khan SN, Ashrafian H, Darzi A. Clinical impact of multi-parameter continuous non-invasive monitoring in hospital wards: a systematic review and meta-analysis. J R Soc Med 2020 Jun;113(6):217-224 [FREE Full text] [doi: 10.1177/0141076820925436] [Medline: 32521195]
28. Cardona-Morrell M, Prgomet M, Turner RM, Nicholson M, Hillman K. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. Int J Clin Pract 2016 Oct;70(10):806-824. [doi: 10.1111/ijcp.12846] [Medline: 27582503]

XSL•FO
RenderX

29. Peelen RV, Eddahchouri Y, Koeneman M, Melis R, van Goor H, Bredie SJH. Comparing Continuous with Periodic Vital Sign Scoring for Clinical Deterioration Using a Patient Data Model. J Med Syst 2023 May 08;47(1):60 [FREE Full text] [doi: 10.1007/s10916-023-01954-z] [Medline: 37154986]
30. Kristinsson EO, Gu Y, Rasmussen SM, Mølgaard J, Haahr-Raunkjær C, Meyhoff CS, et al. Prediction of serious outcomes based on continuous vital sign monitoring of high-risk patients. Comput Biol Med 2022 Aug;147:105559 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.105559] [Medline: 35635901]
31. Dykes PC, Lowenthal G, Lipsitz S, Salvucci SM, Yoon C, Bates DW, et al. Reducing ICU Utilization, Length of Stay, and Cost by Optimizing the Clinical Use of Continuous Monitoring System Technology in the Hospital. Am J Med 2022 Mar;135(3):337-341.e1 [FREE Full text] [doi: 10.1016/j.amjmed.2021.09.024] [Medline: 34717901]
32. Krishnaveni RV, Pandey N, Modh S. Indigenous and disruptive remote patient monitoring devices - a case study on aI in healthcare. SDMIMD 2023;14(2):27-34 [FREE Full text] [doi: 10.18311/sdmimd/2023/32513]
33. Meccariello M, Perkins D, Quigley LG, Rock A. Evaluating the use of an automated vital signs documentation system on a medical/surgical unit JHIM. Vital Time Savings 2010;24:46 [FREE Full text]
34. Chauhan AS, Guinness L, Bahuguna P, Singh MP, Aggarwal V, Rajsekhar K, et al. Cost of hospital services in India: a multi-site study to inform provider payment rates and Health Technology Assessment. BMC Health Serv Res 2022 Nov 14;22(1):1343 [FREE Full text] [doi: 10.1186/s12913-022-08707-7] [Medline: 36376868]
35. Slight SP, Franz C, Olugbile M, Brown HV, Bates DW, Zimlichman E. The return on investment of implementing a continuous monitoring system in general medical-surgical units*. Critical Care Medicine 2014;42(8):1862-1868. [doi: 10.1097/ccm.0000000000000340]
36. Serrano LP, Maita KC, Avila FR, Torres-Guzman RA, Garcia JP, Eldaly AS, et al. Benefits and Challenges of Remote Patient Monitoring as Perceived by Health Care Practitioners: A Systematic Review. Perm J 2023 Dec 15;27(4):100-111 [FREE Full text] [doi: 10.7812/TPP/23.022] [Medline: 37735970]
37. Kuznetsova M, Kim AY, Scully DA, Wolski P, Syrowatka A, Bates DW, et al. Implementation of a Continuous Patient Monitoring System in the Hospital Setting: A Qualitative Study. Jt Comm J Qual Patient Saf 2024 Apr;50(4):235-246 [FREE Full text] [doi: 10.1016/j.jcjq.2023.10.017] [Medline: 38101994]
38. Aquino YSJ, Rogers WA, Braunack-Mayer A, Frazer H, Win KT, Houssami N, et al. Utopia versus dystopia: Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. Int J Med Inform 2023 Jan;169:104903 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104903] [Medline: 36343512]

Abbreviations

EWS: early warning systems
HCP: health care provider
HR: heart rate
ICU: intensive care unit
MEWS: modified early warning score
R-EWS: Remote patient monitoring early warning system
RMH: Ramaiah memorial hospital
RPMS: remote patient monitoring system
RR: respiratory rate
S-MEWS: simulated modified early warning system
S-Threshold: simulated threshold-based alert system
SpO2: oxygen saturation
SysBP: systolic blood pressure

Edited by G Tsafnat; submitted 17.01.24; peer-reviewed by R Zhang, KS Nahla, MD Le Lagadec; comments to author 22.06.24; revised version received 09.08.24; accepted 07.09.24; published 15.01.25.

Please cite as:
Lakshman P, Gopal PT, Khurdi S
Effectiveness of Remote Patient Monitoring Equipped With an Early Warning System in Tertiary Care Hospital Wards: Retrospective Cohort Study
J Med Internet Res 2025;27:e56463
URL: https://www.jmir.org/2025/1/e56463
doi:10.2196/56463
PMID:

XSL•FO
RenderX

©Pavithra Lakshman, Priyanka T Gopal, Sheen Khurdi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Digital Tool for Clinical Evidence–Driven Guideline Development by Studying Properties of Trial Eligible and Ineligible Populations: Development and Usability Study

Shahzad Mumtaz[1,2,3], BSc, MCS, PhD; Megan McMinn[4], BSc, MPH, PhD; Christian Cole[1,2], BSc, PhD; Chuang Gao[1,2], PhD; Christopher Hall[1], BSc; Magalie Guignard-Duff[1], BSc; Huayi Huang[4], BSc, PhD; David A McAllister[5], MD; Daniel R Morales[2], BMSc, MBChB, PhD; Emily Jefferson[1,2,6], BSc, PhD; Bruce Guthrie[4], BChir, MB, MSc, PhD

[1]Health Informatics Centre, School of Medicine, University of Dundee, Dundee, United Kingdom

[2]Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, United Kingdom

[3]School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, United Kingdom

[4]Advanced Care Research Centre, Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[5]School of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom

[6]Health Data Research UK, London, United Kingdom

Corresponding Author:
Emily Jefferson, BSc, PhD
Division of Population Health and Genomics
School of Medicine
University of Dundee
The Health Informatics Centre
Ninewells Hospital and Medical School
Dundee, DD2 1FD
United Kingdom
Phone: 44 01382383943
Email: e.r.jefferson@dundee.ac.uk

Abstract

Background: Clinical guideline development preferentially relies on evidence from randomized controlled trials (RCTs). RCTs are gold-standard methods to evaluate the efficacy of treatments with the highest internal validity but limited external validity, in the sense that their findings may not always be applicable to or generalizable to clinical populations or population characteristics. The external validity of RCTs for the clinical population is constrained by the lack of tailored epidemiological data analysis designed for this purpose due to data governance, consistency of disease or condition definitions, and reduplicated effort in analysis code.

Objective: This study aims to develop a digital tool that characterizes the overall population and differences between clinical trial eligible and ineligible populations from the clinical populations of a disease or condition regarding demography (eg, age, gender, ethnicity), comorbidity, coprescription, hospitalization, and mortality. Currently, the process is complex, onerous, and time-consuming, whereas a real-time tool may be used to rapidly inform a guideline developer's judgment about the applicability of evidence.

Methods: The National Institute for Health and Care Excellence—particularly the gout guideline development group—and the Scottish Intercollegiate Guidelines Network guideline developers were consulted to gather their requirements and evidential data needs when developing guidelines. An R Shiny (R Foundation for Statistical Computing) tool was designed and developed using electronic primary health care data linked with hospitalization and mortality data built upon an optimized data architecture. Disclosure control mechanisms were built into the tool to ensure data confidentiality. The tool was deployed within a Trusted Research Environment, allowing only trusted preapproved researchers to conduct analysis.

Results: The tool supports 128 chronic health conditions as index conditions and 161 conditions as comorbidities (33 in addition to the 128 index conditions). It enables 2 types of analyses via the graphic interface: overall population and stratified by user-defined eligibility criteria. The analyses produce an overview of statistical tables (eg, age, gender) of the index condition population and,

XSL•FO
RenderX

within the overview groupings, produce details on, for example, electronic frailty index, comorbidities, and coprescriptions. The disclosure control mechanism is integral to the tool, limiting tabular counts to meet local governance needs. An exemplary result for gout as an index condition is presented to demonstrate the tool's functionality. Guideline developers from the National Institute for Health and Care Excellence and the Scottish Intercollegiate Guidelines Network provided positive feedback on the tool.

Conclusions: The tool is a proof-of-concept, and the user feedback has demonstrated that this is a step toward computer-interpretable guideline development. Using the digital tool can potentially improve evidence-driven guideline development through the availability of real-world data in real time.

(J Med Internet Res 2025;27:e52385) doi:10.2196/52385

KEYWORDS

multimorbidity; clinical practice guideline; gout; Trusted Research Environment; National Institute for Health and Care Excellence; Scottish Intercollegiate Guidelines Network; clinical practice; development; efficacy; validity; epidemiological data; epidemiology; epidemiological; digital tool; tool; age; gender; ethnicity; mortality; feedback; availability

Introduction

Multimorbidity is the presence of 2 or more long-term health conditions in a person [1]. Typically, multimorbidity was considered a problem for older populations but is increasingly recognized as a challenge in younger people as well [2]. Multimorbidity is associated with worse physical and mental health function, higher service use, and higher mortality [3-5].

Health care services and research often focus on health conditions in isolation, but aging populations mean that both care and research need to better account for multimorbidity. A key point of the intersection of research evidence and care is clinical guideline development, which is where a systematic review of research evidence is used to derive recommendations for clinical practice.

Clinical guideline development preferentially relies on evidence from randomized controlled trials (RCTs). RCT is a gold-standard method to evaluate the efficacy of treatments because they have the highest internal validity [6]. However, the external validity of trial evidence may be limited in the sense that their findings may not always be applicable to or generalizable to clinical populations or population characteristics [7]. This is because RCTs commonly exclude large proportions of patients treated in normal practice. In a systematic review of eligibility for 305 RCTs of treatments for physical conditions, half of the trials excluded more than three-quarters of people with the condition, most commonly older people and people with significant comorbidity (conditions apart from the target of the trial treatment) or coprescribing (of treatments for other conditions) [6,8].

The problem of exclusion from trials is commonly recognized but less commonly quantified. Systematic review guidelines state that generalizability should always be considered and discussed, but rarely is, even in Cochrane systematic reviews [9,10]. The Grades of Recommendation Assessment, Development, and Evaluation (GRADE) recommendations for guideline development also include a method for considering the "directness" of evidence [7]. In principle, guideline developers following GRADE [11] already consider applicability as part of the assessment of indirectness. However, judgments about applicability are often implicit, and the generalizability of estimates (eg, of intervention effect) from RCT evidence is more often assumed than explicitly examined. In practice, clinical guidelines rarely refer to comorbidity or its implications [12], and guidelines rarely mention treatment interactions even though following single-condition guideline recommendations for treatment in a person with multiple conditions commonly leads to potentially serious drug-disease and drug-drug interactions [13], and the cumulative impact of individually rational guideline treatment recommendations can easily be irrational or contradictory in people with multimorbidity [13-17].

Better accounting for multimorbidity and polypharmacy in single-condition guidelines requires robust judgments about the applicability and generalizability of RCT evidence. Such judgments have to be informed by an epidemiological understanding of the clinical population and how it compares to the population studied in RCTs. However, guideline developers are constrained by the lack of tailored epidemiological data analysis designed for this purpose due to data governance, consistency of disease definitions, and reduplicated effort in analysis code. Here, we present a proof-of-concept digital tool using UK health care data to systematically characterize the index condition population, supporting the reuse of existing disease definitions and analysis and integrated disclosure control to ensure governance requirements. A key motivation was to develop an efficient tool that did not require expert data manipulation or technical skills.

Methods

Study Design

We worked with domain experts in guideline development (groups from the National Institute for Health and Care Excellence [NICE] and the Scottish Intercollegiate Guidelines Network [SIGN]) to understand what clinical evidence would be useful for the epidemiological understanding of the disease in terms of overall population and clinical trial eligible or ineligible populations. The Clinical Practice Research Datalink (CPRD) GOLD [18] was used as a primary care data source. The CPRD GOLD is a structured data set extracted from primary care electronic health records (INPS Vision software; In Practice Systems Limited) in the United Kingdom, where key data quality standards are met [19]. In the following subsections, an overview is provided of the data set, details of the requirements

XSL•FO
RenderX

gathering process, the design of the underpinning data structure to support real-time querying, the development of the software tool, deployment of the tool within a Trusted Research Environment (TRE; a secure data and analytics environment), the ethical approval process, and an evaluation of the tool.

Data Set (Primary Care Data Linked with Hospitalization and Mortality)

Data were extracted from the CPRD GOLD data set, consisting of primary care data from general practice electronic health records (EHRs) linked to administrative data for hospital admissions and mortality registration—the organization that manages the CPRD GOLD data extract provided linked data [20]. Data included all permanently registered patients at participating practices on November 30, 2015 (1.17 million), with at least 2 years of prior registration. Clinical conditions were measured any time before the index date, while data on prescriptions were restricted to the previous year due to the size of the data files and clinical relevance for guideline development. Follow-up data for hospital admission and death were available for 3 years (ie, until November 30, 2018). The presence of 161 long-term conditions was ascertained using Read codes and ICD10 (International Classification of Diseases 10th Revision) codes recorded on or before November 30, 2015, using published code lists from the Health Data Research UK phenotype library [21,22]. Read codes and ICD10 coding systems have been used for coding patient data across the UK National Health Services in primary and secondary care services, respectively. For each patient, the electronic frailty index (eFI) [23] and the Charlson Comorbidity Index score [24] were calculated. The raw data set used for this study was provided by the organization that manages CPRD, which explained the data quality and validation checks in place [25]. For study-specific cohort quality and validity, the first and second authors of this paper conducted separate analyses for several conditions, and the cohort sizes generated through the tool and analysis scripts were observed to be the same.

Requirements Gathering (Live Guideline Development Group Consultation)

Tool development was done in conjunction with the guideline development group (GDG) for the NICE gout guideline (subsequently referred to as the NICE gout GDG); therefore, gout was used as an exemplar index condition. Gout is a form of arthritis caused by urate crystal deposition in joints that causes severe joint inflammation and pain, which can be prevented by long-term medication to reduce blood urate levels [26,27]. Initial interviews with GDG were used to define requirements and active engagement with the NICE gout GDG throughout its life cycle, including providing bespoke analysis upon request to help prompt reflection on the use of different kinds of data at different points in guideline development. The bespoke analysis provided is published as an appendix to the guideline [28]. In parallel, a prototype data analysis tool was designed and demonstrated to the NICE gout GDG and other guideline developers in NICE and SIGN to seek their feedback and more detailed requirements in terms of data analysis they wished to have access to as they considered the evidence and drafted recommendations. It was agreed to capture the following two high-level output types:

1. Index condition population analysis: A general view of the population with the underlying condition.
2. Clinical trial population analysis: A modified view allowing explicit comparison of patients eligible or ineligible for a particular clinical trial (or a notional clinical trial), where the user inputs inclusion or exclusion criteria to define the 2 populations: eligible versus ineligible.

Both types of analysis were agreed to produce summaries as tables of numbers for age, gender, ethnicity, and indices of multiple deprivations (IMD) groupings for the selected condition. These summaries were also expanded to demonstrate frailty (measured by the eFI), comorbidity (measured with the Charlson score), the prevalence of comorbidities, and current drug prescription. Additionally, it was highlighted to produce rates of hospitalization and mortality.

Four further requirements included: (1) making the analysis pipeline available for other chronic conditions (because of perceived value across different health conditions by GDGs); (2) supporting the selection of hierarchical groupings of comorbidities and drug prescriptions; (3) statistically disclosure controlled outputs to make egressing of the outputs faster (for instance, by rounding all numbers to nearest 10 and percentages to discrete numbers) to ensure patient confidentiality and to satisfy data governance approvals; and (4) generation of printable reports to use as evidence within the guideline development process, and for subsequent publication as part of the guideline.

Optimizing the Data Structure

A major challenge when developing digital tools that use large data sets is the query processing time. For example, the CPRD GOLD data set has approximately 1.17 million patients in the patient file and approximately 254 million entries in the clinical table (storing codes documenting interactions with the practice). Extracting comorbidities for an index condition based on age, gender, and other features required intensive computation power and storage capacity when dynamic linking the patient and clinical tables. The same challenge was seen for other data analyses. To optimize query processing time, we designed new tables in the Microsoft SQL server, which included only the necessary entries for analysis (eg, for the presence of a condition, then only the first Read code recording its presence and the date recorded is required, rather than all such codes). Similarly, eFI and the Charlson scores were precalculated. However, even with an optimized data structure, the performance of processing queries will increase given the cohort size for a selected index condition unless we increase the computing power and memory. The details of the new data structure design and the rationale for generating each table are given in Multimedia Appendix 1.

Development of the Informatics Tool Using R Shiny

The study has developed an R Shiny (R Foundation for Statistical Computing) proof-of-concept digital informatics tool. R Shiny uses R packages to build graphical, interactive

XSL•FO
RenderX

dashboards for data exploration [29]. Dashboards can be rapidly configured, making R Shiny an ideal tool for prototyping.

The tool dashboard design has considered the novices to analyze and extract useful information without knowledge of writing analysis scripts in R. The dashboard enables analysis of all patients with a chosen index condition (for instance, everyone with gout on November 30, 2015) and clinical trial analysis in terms of eligible and ineligible populations as defined in the requirements gathering process. Further descriptions of the analysis types with input and output of each type are presented in Multimedia Appendix 2.

Comorbidities and Prescriptions Analyses

A key requirement of the tool output was to support a selection of comorbidities and drug prescriptions at different hierarchical grouping levels for the selected index condition of interest. The hierarchical groupings of conditions for comorbidities analysis were determined by 2 expert clinicians (authors) and are provided in Multimedia Appendix 3, whereas the British National Formulary (BNF) [30] was used to hierarchically group drugs. Diseases are classified into 3 hierarchical levels: body system, condition group, and individual conditions. The body system is the top level in the hierarchy of diseases. Further down the hierarchy are condition groups, which are subcategories of the body system and are made up of individual conditions. For example, the "cancers" body system has condition groups such as "hematological cancers," "solid organ cancer (primary)," or "solid organ cancer (secondary)," and "hematological cancer" is a condition group made up of individual conditions such as "Hodgkin lymphoma," "leukemia," "myelodysplastic syndrome," "non-Hodgkin lymphoma," "plasma cell malignancy," and "polycythemia vera."

When presenting comorbidities at the body system level, if a patient has one or more comorbidities, under the hierarchical subcategories of the body system, they will be counted as 1 disease or condition. The tool's design considers a selection of any level of comorbidities in the disease's hierarchy.

For prescription data within CPRD GOLD, drug codes are given as BNF codes so that drug classification was used instead of the World Health Organization (WHO) Anatomical Therapeutic Chemical classification, which is the de facto drug coding standard in clinical trials [31]. In the BNF, drugs are classified into 3 hierarchical levels: drug chapters, drug classes, and individual drugs.

Deployment of the R Shiny Tool

The R Shiny tool was developed and deployed within the Health Informatics Centre (HIC) of the University of Dundee TRE, which was only accessible to authorized members of the study (Figure 1). The tool is an R Shiny application, and it can be accessed simply through a web browser within the TRE. The tool applies disclosure controls to the results displayed by rounding all numbers to the nearest 10 and providing percentages as discrete numbers, thereby avoiding accidental patient reidentification through inference without introducing bias in prevalence estimates. All data presented here fulfills the disclosure control requirements.

Figure 1. A schematic view of the deployed R Shiny tool. A load balancer allows multiple users to interact simultaneously with an R Shiny server. CPRD: Clinical Practice Research Datalink; HIC: Health Informatics Centre.

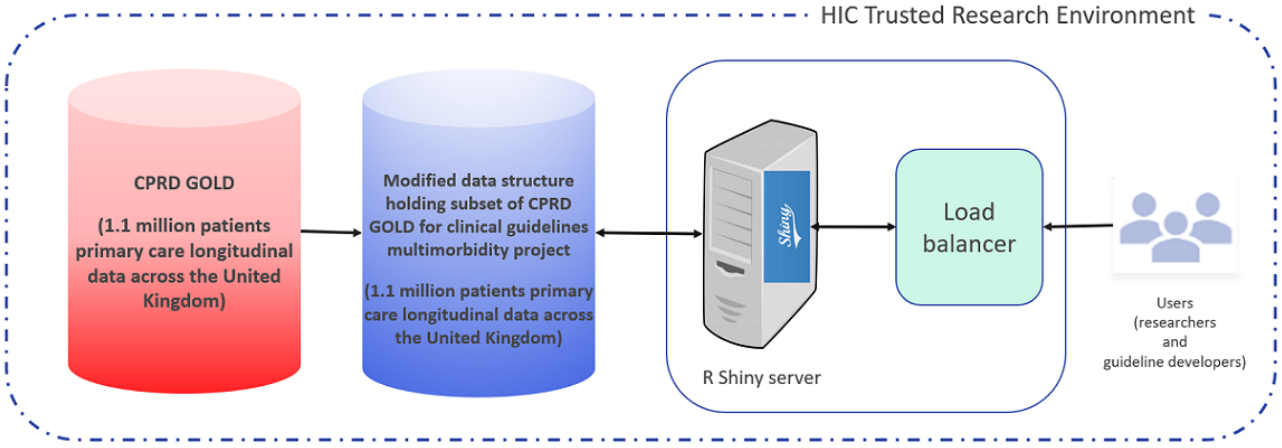

Ethical Considerations

The study only used deidentified data, with integrated disclosure control within the tool to avoid accidental patient reidentification through inference, and was approved by the CPRD Independent Scientific Advisory Committee (ISAC; protocol 20_018) [32]. To ensure patient privacy, CPRD provided anonymized data for this study.

Evaluation of the Tool

Throughout the development of the tool, we performed iterative usability testing by demonstrating updated versions to the domain experts to confirm the applicability of the evidence. The domain experts included gout GDG, the study advisory group (which included representatives from NICE and SIGN), the European Respiratory Society, and public contributors. For the final evaluation, the project team members and external users from NICE and SIGN were given controlled access to trial the tool within the HIC TRE. Simple three-step instructions were shared with users: (1) logging into the HIC TRE virtual machine, (2) typing a given URL in a browser to access the tool, and (3) conducting analysis using the GUI (Figure 2) as described in requirements gathering section earlier.

XSL•FO
RenderX

Figure 2. A screenshot of the R Shiny tool for clinical population analysis. The left sidebar (A) has 2 menu items: index condition population analysis and trial eligible or ineligible population analysis. The top section (B) enables users to choose an index condition of interest from the drop-down list. The bottom right section (C) shows the results of the selected index condition in different tabs focusing on different outputs such as demographic distributions, comorbidities, and coprescriptions.

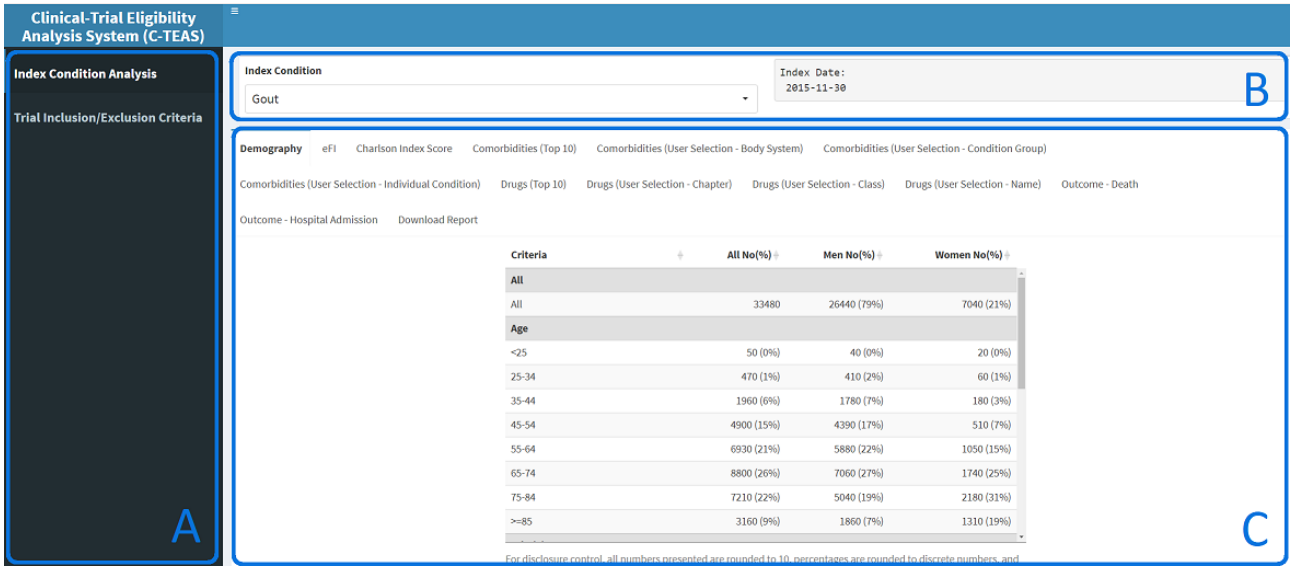

Results

Overview

During the analysis phase of the index conditions of interest, we considered factors in designing and developing an R Shiny tool that can be widely applicable to many conditions in CPRD GOLD. The tool supports the selection of 128 conditions as index conditions and 161 conditions as comorbidities (the 128 index conditions plus 33 other chronic health conditions), which are individual conditions in the hierarchical groupings and the associated higher-level groupings in the hierarchy (details of the supported index conditions and comorbidities are presented in Multimedia Appendix 3). CPRD is broadly representative of the UK population [33], and data quality is high, including gout [34].

Overview of R Shiny Tool

The R Shiny tool supports a 2-step analysis process as described in the methodology section (Figure 3). First, a user performs population-level analysis for the condition of interest from the list of supported index conditions (Multimedia Appendix 3). Second, a user sets the trial inclusion and exclusion criteria to understand trial eligible or ineligible populations. A screenshot of the R Shiny tool is presented in Figure 2, highlighting three areas: (1) Figure 2A has 2 menu options that allow users to switch between analysis types; (2) Figure 2B is the area that allows the user to select the condition of interest from the drop-down list to generate population analysis; and (3) Figure 2C is the results area where multiple tabs are provided to enable switching between different types of outputs.

Figure 3. A workflow of an R Shiny tool. IMD: index of multiple deprivations.

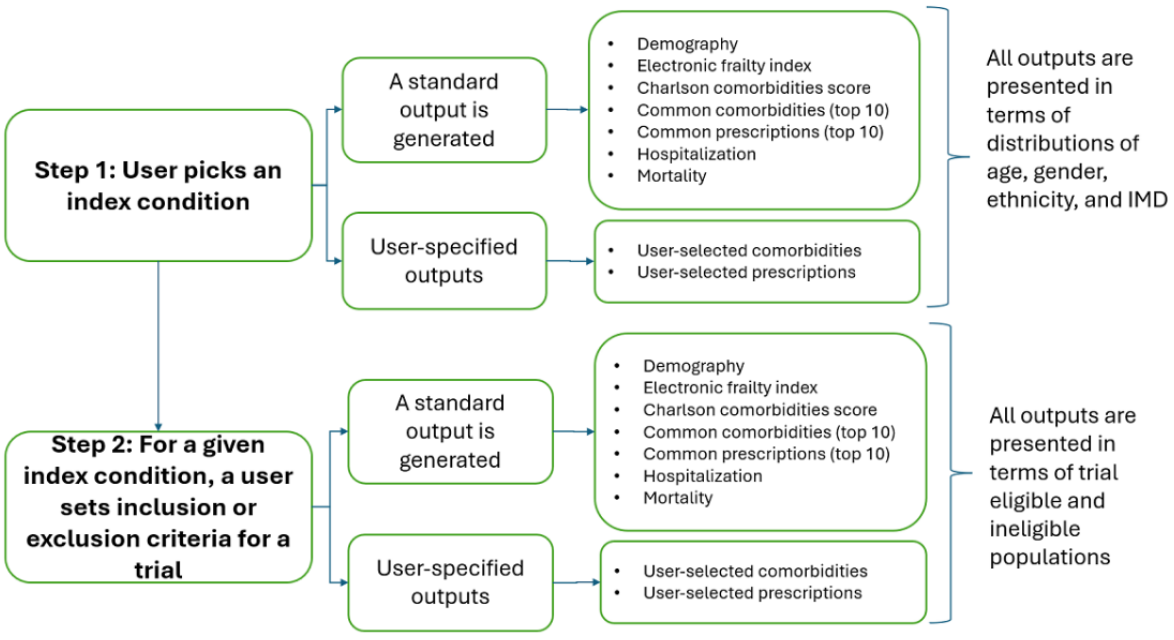

XSL•FO
RenderX

Index Condition Population Analysis in Gout

The tool was designed to be flexible and provide many selection options to the end user to support guideline development; given the many combinations of results with 13 output tabs available, only a few examples are shown here.

We present results for gout as an index condition for demonstration purposes. Table 1 shows an analysis of the whole population with gout: any patient diagnosed with gout on or before the index date (November 30, 2015) is included in the analysis. Details are provided for distributions of patients' characteristics, including age, gender, ethnicity, IMD, and time since diagnosis (prevalent and recently incident populations) at the index date. From a population of 1,168,620 patients from the CPRD GOLD data set, a cohort with gout of 33,480 (2.86%) was identified. The cohort with gout was proportionately older (aged greater than 64 years; 75%). The CPRD GOLD data set is mostly a White ethnic population (68%) and a relatively higher percentage (81%) is observed in the population with gout. For other ethnicity groupings: South Asian (3% vs 2%), Black (2% vs 1%), Chinese or Mixed or Other (2% vs 1%), missing (24% vs 15%). Based upon IMD, a higher percentage of those with gout were richer (26% in the least deprived category) versus poorer (13% in the most deprived category). For many diseases, the opposite trend is observed; however, gout has historically been known as "a rich man's disease" [35-37], a generalization that is evidenced by this data.

Table 1. An example output of population with gout from the R Shiny application describing the distribution of patients' characteristics, including age, gender, ethnicity, IMD[a], and time since diagnosis (prevalent and recently incident populations).[b]

Criteria (feature)	All, n (%)	Men, n (%)	Women, n (%)
All	33,480 (100)	26,440 (79)	7040 (21)
Age (years)			
Less than 25	50 (0)	40 (0)	20 (0)
25-34	470 (1)	410 (2)	60 (1)
35-44	1960 (6)	1780 (7)	180 (3)
45-54	4900 (15)	4390 (17)	510 (7)
55-64	6930 (21)	5880 (22)	1050 (15)
65-74	8800 (26)	7060 (27)	1740 (25)
75-84	7210 (22)	5040 (19)	2180 (31)
85 or older	3160 (9)	1860 (7)	1310 (19)
Ethnicity			
Black	320 (1)	240 (1)	80 (1)
Chinese or Mixed or Other	360 (1)	290 (1)	70 (1)
South Asian	610 (2)	470 (2)	140 (2)
White	27100 (81)	20860 (79)	6250 (89)
Missing	5090 (15)	4590 (17)	500 (7)
IMD quintile			
Q1 (least deprived)	8720 (26)	7130 (27)	1590 (23)
Q2	7400 (22)	5890 (22)	1510 (21)
Q3	7300 (22)	5710 (22)	1580 (22)
Q4	5680 (17)	4330 (16)	1350 (19)
Q5 (most deprived)	4380 (13)	3380 (13)	1010 (14)
Time since diagnosis (years)			
1 or greater	31,210 (93)	24,790 (94)	6420 (91)
Less than 1	2270 (7)	1650 (6)	620 (9)

[a]IMD: index of multiple deprivations.

[b]For disclosure control, all numbers are rounded to 10, percentages are rounded to discrete numbers, and percentages presented are column percentages (except the row titled "All," where the percentages are row percentages). Totals may not add up exactly to 100%.

Once an index condition is defined, the tool presents the most frequently observed comorbidities and prescriptions in separate tabs. Tables 2 and 3 present the top 5 comorbidities and prescriptions as exemplars for demonstration purposes (the default in the tool is the top 10 comorbidities and top 10 prescriptions). The individual counts are stratified by the

XSL•FO
RenderX

demographic characteristics in Table 1 and shown as relative percentages. As can be seen in Table 2, "disease of the circulatory systems" is the most common comorbidity, being present in 68% of the population with gout, followed by "musculoskeletal conditions" (42%), "diseases of the genitourinary system" (39%), "diseases of the digestive system" (38%), and "diseases of the endocrine system" (36%). As can be seen in Table 3, "02 cardiovascular system" is the most common prescription by drug chapter, being present in 67% of the population with gout, followed by "10 musculoskeletal and joint disease" (45%), "01 gastrointestinal system" (36%), "04 central nervous system" (35%), and "06 endocrine system" (27%). For the top 5 comorbidities and top 5 prescriptions, people with gout over the age of 65 years have higher overall comorbidity and prescriptions than younger age groups. For example, overall, 68% of the population with gout had a comorbidity of "disease of the circulatory systems," whereas for age groups such as "65-74" and "75-84" the proportion with the same comorbidity was much higher; 79% and 89% respectively. When compared to the overall percentages, in some cases, women showed greater than 10 percentage points difference for the comorbidity or drug chapter. For example, overall, 68% of people with gout had a comorbidity for "disease of the circulatory system," whereas this was 79% for women. Similarly, this can be seen for "musculoskeletal conditions" (42% vs 63%), "diseases of the digestive system" (38% vs 47%), and "diseases of the endocrine system" (36% vs 51%). An exception of a lower percentage of comorbidity was observed in women with "diseases of the genitourinary system" comorbidity (39% vs 32%).

Table 2. An example of the top 5 comorbidities from the R Shiny application describing the distribution of patients' characteristics, including age, gender, ethnicity, and IMD[a] for the cohort with gout.[b]

Criteria (feature)	Top 5 comorbidities (body systems)				
	Disease of the circulatory system, n (%)	Musculoskeletal conditions, n (%)	Diseases of the genitourinary system, n (%)	Diseases of the digestive system, n (%)	Diseases of the endocrine system, n (%)
All	22,730 (68)	13,920 (42)	13,140 (39)	12,820 (38)	11,980 (36)
Age (years)					
Younger than 25	0 (8)	10 (15)	0 (8)	0 (10)	0 (6)
25-34	50 (10)	50 (11)	50 (10)	70 (15)	60 (14)
35-44	380 (19)	190 (10)	220 (11)	370 (19)	340 (17)
45-54	1820 (37)	930 (19)	1050 (21)	1200 (25)	1150 (23)
55-64	4180 (60)	2260 (33)	2340 (34)	2290 (33)	2310 (33)
65-74	6840 (78)	4040 (46)	3960 (45)	3660 (42)	3540 (40)
75-84	6480 (90)	4270 (59)	3790 (53)	3590 (50)	3260 (45)
85 and older	2980 (94)	2180 (69)	1720 (55)	1620 (51)	1310 (41)
Gender					
Men	17,180 (65)	9500 (36)	10,920 (41)	9480 (36)	8400 (32)
Women	5550 (79)	4420 (63)	2220 (32)	3340 (32)	3340 (47)
Ethnicity					
Black	240 (74)	110 (34)	150 (46)	120 (37)	160 (49)
Chinese or Mixed or Other	240 (68)	110 (30)	150 (42)	110 (32)	120 (35)
South Asian	420 (68)	220 (36)	270 (44)	250 (41)	290 (47)
White	19,820 (73)	12,670 (47)	11,680 (43)	11,680 (43)	10,410 (38)
Missing	2010 (39)	810 (16)	890 (18)	660 (13)	1010 (20)
IMD quintile					
Q1 (least deprived)	5690 (65)	3280 (38)	3400 (39)	3180 (36)	2680 (31)
Q2	4990 (67)	3050 (41)	2930 (40)	2740 (37)	2560 (35)
Q3	5040 (69)	3170 (43)	2850 (39)	2780 (38)	2690 (37)
Q4	3930 (69)	2480 (44)	2230 (39)	2240 (39)	2240 (39)
Q5 (most deprived)	3070 (70)	1950 (44)	1730 (40)	1890 (43)	1800 (41)

[a]IMD: index of multiple deprivations.

[b]For disclosure control, all numbers presented are rounded to 10, percentages are rounded to discrete numbers, and percentages presented are row percentages of the "All" column in Table 1, so totals will not be added exactly.

XSL•FO
RenderX

Table 3. An example of the top 5 prescriptions from the R Shiny application describing the distribution of patients' characteristics, including age, gender, ethnicity, and IMD[a] for the cohort with gout.[b]

Criteria (feature)	Top 5 prescriptions (drug chapters)				
	02 cardiovascular system, n (%)	10 musculoskeletal and joint disease, n (%)	01 gastrointestinal system, n (%)	04 central nervous system, n (%)	06 endocrine system, n (%)
All	22,310 (67)	14,920 (45)	11,980 (36)	11,720 (35)	8980 (27)
Age (years)					
Younger than 25	0 (4)	10 (19)	0 (8)	10 (19)	0 (4)
25-34	40 (9)	120 (26)	50 (10)	70 (14)	30 (6)
35-44	330 (17)	630 (32)	260 (13)	380 (19)	160 (8)
45-54	1660 (34)	1880 (38)	1000 (20)	990 (20)	660 (13)
55-64	4080 (59)	3070 (44)	2010 (29)	2000 (29)	1490 (22)
65-74	6960 (79)	4320 (49)	3370 (38)	3110 (35)	2590 (29)
75-84	6390 (89)	3510 (49)	3540 (49)	3430 (48)	2760 (38)
85 and older	2840 (90)	1380 (44)	1740 (55)	1740 (55)	1300 (41)
Gender					
Men	16,900 (64)	12,030 (45)	8690 (33)	7820 (30)	5960 (23)
Women	5410 (77)	2900 (41)	3290 (47)	3890 (55)	3020 (43)
Ethnicity					
Black	220 (70)	140 (43)	110 (33)	110 (34)	110 (34)
Chinese or Mixed or Other	240 (66)	170 (48)	100 (29)	100 (28)	100 (28)
South Asian	400 (65)	270 (44)	250 (40)	240 (39)	230 (38)
White	19,370 (71)	12,550 (46)	10,930 (40)	10,660 (39)	7940 (29)
Missing	2080 (41)	1790 (35)	590 (12)	620 (12)	600 (12)
IMD quintile					
Q1 (least deprived)	5560 (64)	3660 (42)	2710 (31)	2390 (27)	2000 (23)
Q2	4910 (66)	3270 (44)	2500 (34)	2440 (33)	1980 (27)
Q3	4950 (68)	3260 (45)	2740 (38)	2550 (35)	2060 (28)
Q4	3900 (69)	2670 (47)	2230 (39)	2280 (40)	1620 (29)
Q5 (most deprived)	2990 (68)	2070 (47)	1800 (41)	2060 (47)	1320 (30)

[a]IMD: index of multiple deprivations.

[b]For disclosure control, all numbers presented are rounded to 10, percentages are rounded to discrete numbers, and percentages presented are row percentages of the "All" column in Table 1, so totals will not be added exactly.

Different ethnic groups with gout showed different percentages of comorbidities for different diseases. For example, 47% of the White ethnic group had a musculoskeletal condition comorbidity, whereas this was only 34% of those within the Black ethnic group. However, the comorbidity of "diseases of the endocrine system" was observed in 49% of the Black ethnic group in comparison to 38% of the White ethnic group. There was a general observation that the higher the level of deprivation, the higher the percentage of the population with each comorbidity and prescription.

Such differences are easily highlighted in the tool for guideline developers aiding their workflows regardless of index conditions and parameters specified.

Alternatively, it is possible to view the most frequent comorbidities and prescriptions at the lower level of granularity given in the hierarchical groupings (ie, comorbidities as condition groups and/or individual conditions and prescriptions as BNF classes and/or individual drugs).

In addition to viewing the most common top-level comorbidities and prescriptions for the condition of interest (gout in this case), there is an option for the user to select comorbidities and prescriptions of interest from the complete list of comorbidities and prescriptions. This user selection feature is supported at each hierarchical level of comorbidities (ie, body system, condition group, and individual condition) and prescriptions (ie, drug chapter, drug class, and individual drug). Table 4 demonstrates user selection of comorbidities at the "condition

XSL•FO
RenderX

group" level and prescriptions at "drug class" levels. Selected drug class level statistics were requested by the NICE gout GDG. There were similar trends observed in the data for this lower level of granularity.

Table 4. An example of the user-selected comorbidities (ie, condition groups) and prescriptions (drug classes) from the R Shiny application describing the distribution of patients' characteristics, including age, gender, ethnicity, IMD[a], and time since diagnosis (prevalent and recently incident populations).[b]

Criteria (feature)	User-selected comorbidities (condition groups)				User-selected prescriptions (drug classes)			
	Hypertension, n (%)	Ulcer and upper gastrointestinal acid conditions, n (%)	Chronic lung disease, n (%)	Coronary heart disease, n (%)	Statins, n (%)	Angiotensin receptor blockers, n (%)	Oral anticoagulants, n (%)	Thiazide diuretics, n (%)
All	20,540 (61)	12,120 (36)	7260 (22)	6980 (21)	13,740 (41)	4750 (14)	3380 (10)	3050 (9)
Age (years)								
Younger than 25	0 (4)	0 (8)	20 (35)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
25-34	30 (6)	60 (13)	110 (24)	0 (1)	10 (3)	0 (0)	0 (1)	0 (0)
35-44	300 (16)	340 (17)	360 (18)	30 (2)	120 (6)	40 (2)	10 (1)	20 (1)
45-54	1590 (32)	1090 (22)	840 (17)	240 (5)	820 (17)	280 (6)	90 (2)	150 (3)
55-64	3710 (54)	2100 (30)	1270 (18)	800 (12)	2430 (35)	790 (11)	270 (4)	520 (8)
65-74	6210 (71)	3470 (39)	1970 (22)	2050 (23)	4740 (54)	1560 (18)	920 (10)	990 (11)
75-84	5940 (82)	3470 (48)	1880 (26)	2560 (35)	4120 (57)	1510 (21)	1420 (20)	1000 (14)
85 and older	2760 (87)	1580 (50)	830 (26)	1300 (41)	1500 (48)	570 (18)	660 (21)	360 (11)
Gender								
Men	15,360 (58)	8920 (34)	5230 (20)	5470 (21)	10,812 (41)	3369 (13)	2593 (10)	2059 (8)
Women	5190 (74)	3190 (45)	2030 (29)	1520 (22)	2931 (42)	1380 (20)	791 (11)	989 (14)
Ethnicity								
Black	230 (72)	100 (33)	70 (21)	60 (17)	130 (39)	50 (17)	20 (7)	50 (15)
Chinese or Mixed or Other	210 (58)	100 (29)	60 (16)	70 (20)	160 (43)	50 (14)	20 (6)	30 (8)
South Asian	380 (62)	230 (37)	150 (24)	160 (27)	280 (46)	110 (18)	30 (4)	50 (9)
White	17,850 (66)	11,080 (41)	6380 (24)	6560 (24)	12,000 (44)	4140 (15)	3250 (12)	2610 (10)
Missing	1880 (37)	600 (12)	610 (12)	140 (3)	1180 (23)	400 (8)	60 (1)	310 (6)
IMD quintile								
Q1 (least deprived)	5100 (58)	2990 (34)	1630 (19)	1630 (19)	3320 (38)	1300 (15)	870 (10)	780 (9)
Q2	4500 (61)	2590 (35)	1510 (20)	1470 (20)	2990 (40)	1130 (15)	770 (10)	640 (9)
Q3	4560 (62)	2640 (36)	1570 (22)	1600 (22)	3070 (42)	1030 (14)	750 (10)	700 (10)
Q4	3590 (63)	2110 (37)	1370 (24)	1240 (22)	2380 (42)	760 (13)	570 (10)	530 (9)
Q5 (most deprived)	2800 (64)	1790 (41)	1180 (27)	1040 (24)	1980 (45)	520 (12)	420 (10)	400 (9)

[a]IMD: index of multiple deprivations.

[b]For disclosure control, all numbers presented are rounded to 10, percentages are rounded to discrete numbers, and percentages presented are row percentages of the "All" column in Table 1, so totals will not be added exactly.

As per the requirements, the tool supports the generation of printable reports in HTML format. The individual outcomes to save and print (and then save as PDF) are user selectable, as shown in Figure 4.

XSL•FO
RenderX

Figure 4. A screenshot of the report generation page.

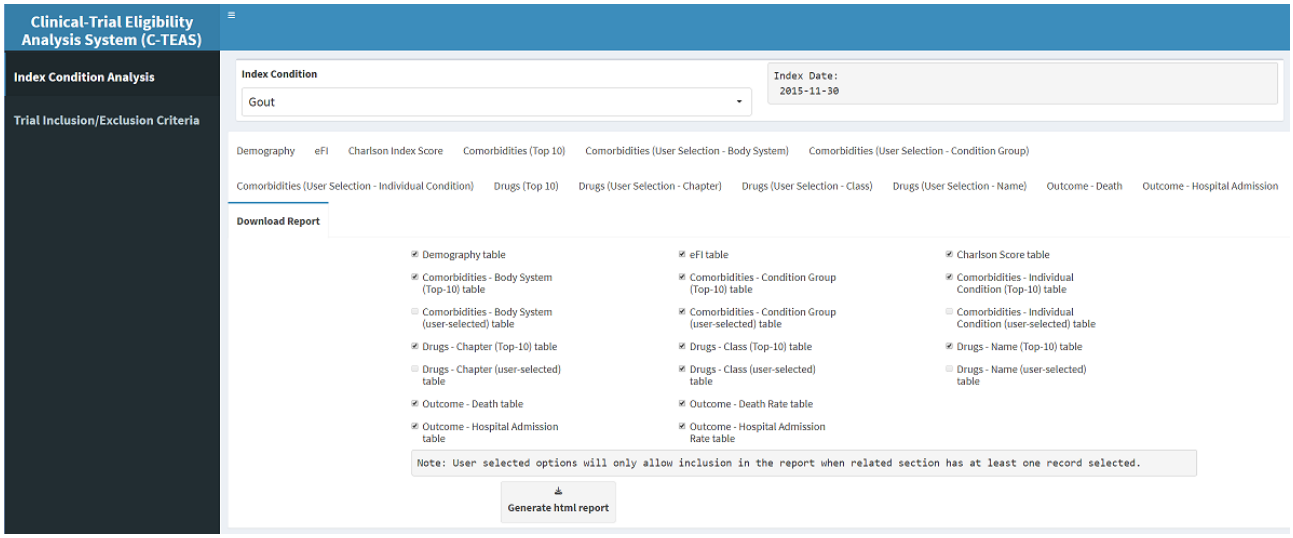

Clinical Trial Population Analysis

The R Shiny tool enables comparisons between trial eligible and ineligible populations using inclusion and exclusion criteria. As illustrated in Figure 5, a user can set inclusion and exclusion criteria based on five main criteria: gender, ethnicity, age, comorbidities (only for individual conditions from hierarchical groupings of conditions), and drug prescription (only for individual drugs).

Figure 5. A screenshot of the R Shiny tool. Setting up inclusion and/or exclusion criteria to analyze eligible and ineligible populations.

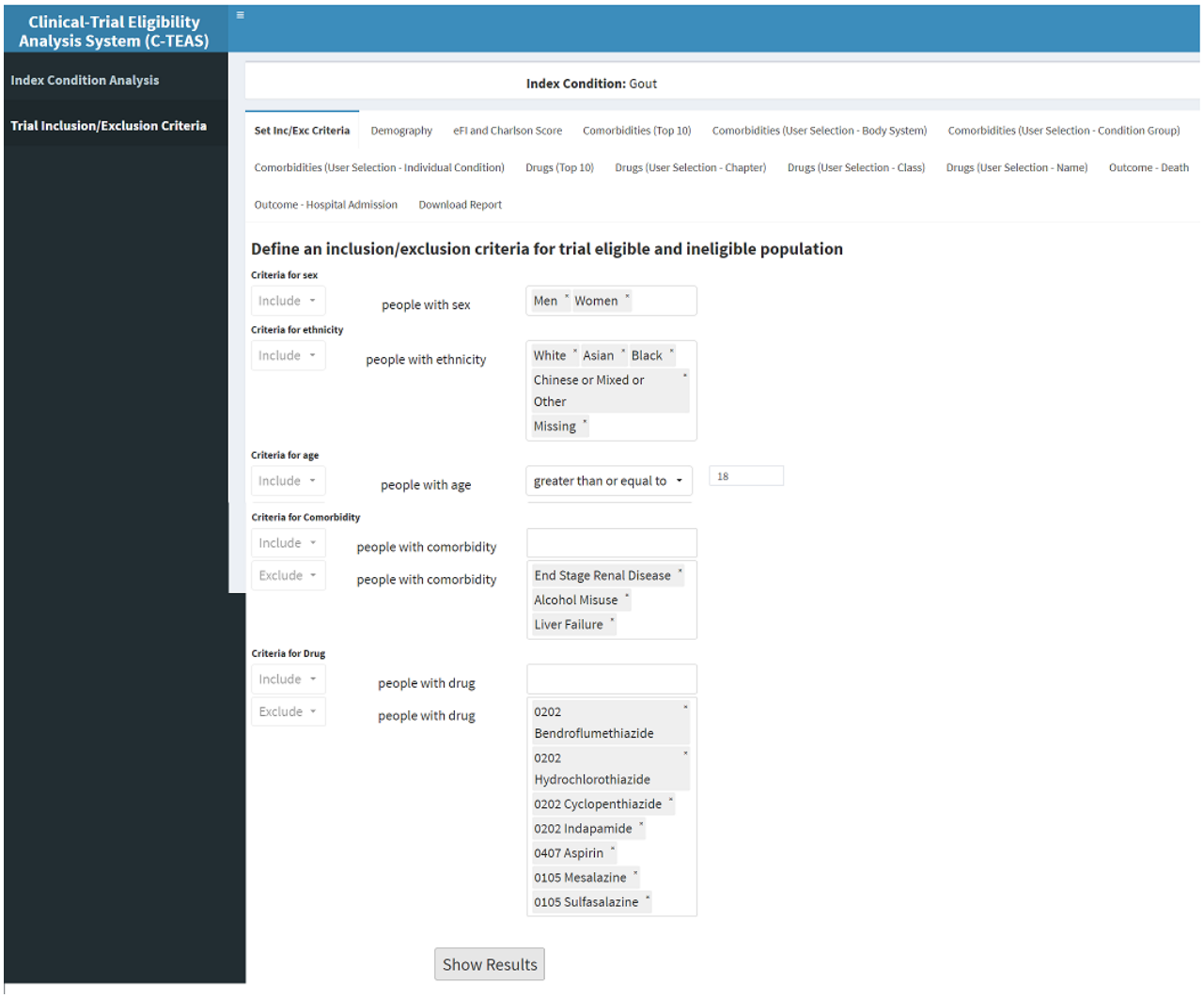

XSL•FO
RenderX

The results of a trial eligible and ineligible query are shown in Table 5. The criteria were adapted as best possible given the different sources of data from a published gout trial in the literature [38].

Table 5. Output generated for an example trial taken from the literature.[a]

Criteria (feature)	All, n (%)	Eligible, n (%)	Ineligible, n (%)
All	33,480	27,080 (81)	6400 (19)
Age (years)			
Younger than 25	50 (0)	40 (85)	10 (15)
25-34	470 (1)	420 (91)	40 (9)
35-44	1960 (6)	1740 (89)	220 (11)
45-54	4900 (15)	4220 (86)	680 (14)
55-64	6930 (21)	5580 (81)	1350 (19)
65-74	8800 (26)	6850 (78)	1960 (22)
75-84	7210 (22)	5610 (78)	1600 (22)
85 and older	3160 (9)	2610 (83)	550 (17)
Gender			
Men	26,440 (79)	21,520 (81)	4920 (19)
Women	7040 (21)	5560 (79)	1480 (21)
eFI[b]			
Fit	13,650 (41)	12,020 (44)	1630 (25)
Mild frailty	10,220 (31)	7850 (29)	2370 (37)
Moderate frailty	6210 (19)	4620 (17)	1590 (25)
Severe frailty	3400 (10)	2600 (10)	810 (13)
Comorbidities (condition group)—user selected			
Hypertension	20,540 (61)	15,320 (57)	5230 (82)
Ulcer and upper gastrointestinal acid conditions	12,120 (36)	9330 (34)	2780 (43)
Chronic lung disease	7260 (22)	5600 (21)	1660 (26)
Coronary heart disease	6980 (21)	5370 (20)	1620 (25)
Prescriptions (drug class)—user selected			
Statins	13,740 (41)	10,400 (38)	3340 (52)
Angiotensin receptor blockers	4750 (14)	3450 (13)	1300 (20)
Oral anticoagulants	3380 (10)	2660 (10)	720 (11)
Thiazide diuretics	3050 (9)	70 (0)	2980 (47)

[a]For disclosure control, all numbers are rounded to 10, percentages are rounded to discrete numbers, and the percentages are column-wise percentages except for a special case of age and gender where percentages in the eligible and ineligible columns are row-wise percentages based on the totals (for each age group or gender type) in the column "All"), and so totals will not add exactly.

[b]eFI: electronic frailty index.

The inclusion criteria are as follows: (1) age (18 years and older); (2) gender (men and women); and ethnicity (any). The exclusion criteria are as follows: (1) comorbidities (end-stage renal disease, alcohol misuse, and liver failure) and (2) drugs (bendroflumethiazide, hydrochlorothiazide, cyclopenthiazide, indapamide, aspirin, mesalazine, and sulfasalazine).

Table 5 illustrates the output collated from multiple output tabs of the tool, capturing demographics, eFI, comorbidities (user selected), and prescriptions (user selected) in terms of trial eligible and ineligible populations. Overall, 81% of the population with gout were eligible, and 19% were ineligible based upon the specified inclusion and exclusion criteria—the current version of the tool is a proof-of-concept and supports a subset of the criteria specified in the published gout trial [38]. In comparing the eligible and ineligible columns, the trialist can assess biases between the cohorts regarding any inherent demographic, comorbidity (commonest and user selected), and prescriptions (commonest and user selected). It should be noted that the total percentage within the eFI, comorbidities, and prescriptions in eligible and ineligible columns do not add up to 100% (as they do for "All," "Age (years)," and "Gender"

XSL•FO
RenderX

groupings) as individuals may have less than 1 or 0 comorbidities or prescriptions, and the results presented are not as a proportion of all of the population with gout (as they are for "All," "Age (years)," "Gender," and "eFI" groupings) against the "All" column instead they are based on total eligible and ineligible populations.

Integrated Disclosure Control and Report Generation

Accidental disclosure of low-count data was avoided by built-in rules, which rounded values to the nearest 10 and percentages to integers. The integrated disclosure-control mechanism was not only helpful in ensuring patients' confidentiality when guideline users were conducting analysis using the tool within the TRE but also enabled rapid egressing of the reports from the TRE—as the TRE output checking staff already knew that outputs generated from the tool were nondisclosive.

Completed analyses can be saved as reports. Printable versions of the completed results across tabs for the overall population and trial eligible or ineligible population analyses are presented in Multimedia Appendices 4 and 5, respectively.

Guideline Users Feedback

To evaluate the tool's functionality, several users from NICE and SIGN guideline developers were given access to the tool within HIC TRE. They found the tool to be very useful while developing guidelines to enable the use of real-world evidence about the nature of the population for whom different treatments were being recommended. They have shown significant interest in the R Shiny tool in its current form due to its simplest navigation functionality for generating analysis and reports at pace and in further development into a production version. The tool supports many different outputs, for example, demography statistics, top comorbidities statistics, user-selected comorbidities, and prescription statistics. There was a general consensus across users testing the functionality that the generation of a range of outputs for the chosen disease/condition within a few minutes is highly advantageous. One of the users (ie, a health economist from NICE) was of the view that generating statistics on comorbidities could be helpful for the study of the burden of comorbidities on the health systems. One of the users of the tool from NICE expressed his views about the tool:

> *The tool enables us to assess the relevance of RCTs and observational studies to real-world UK patients, by comparison between real-world patient characteristics and study eligibility criteria. This provides a new lens through which committees and developers could critically assess studies captured in evidence reviews. There are interesting potential applications of the tool inform surveillance, scoping, and NICE research recommendations. The tool is interactive and returns results immediately. This maximises its potential impact for guideline development, as extended waits for data access and results may mean analyses come too late to be usefully considered by guideline committees working to tight deadlines. The clear format, where most common conditions and prescriptions can be viewed at a level of granularity chosen by the user, is ideal. NICE's 2021-26 Strategy emphasizes increased use of real-world evidence and more rapid and responsive guideline development. This tool aligns well with these ambitions. We hope there will be opportunities to further test the tool in guideline development in future.* [User #1, male]

Key Features of the Tool

Factors such as ease of use of the tool by novice users, the query processing time to extract the required data, integral disclosure control mechanism, and standardized outputs across various chronic conditions were among the challenges addressed when developing the application. The tool is simple to use, with only 2 menu options and minimum input required to generate results for both types of analysis using a graphical interface. To optimize the performance of query processing, we designed an intermediate data structure populated by the data from the CPRD GOLD subset. This resulted in retrieving results from each type of analysis within approximately a couple of minutes compared to a much longer time when running a query against the raw CPRD GOLD data set (eg, generating population-level results at the demography level took more than 3-5 minutes running on the raw data set versus only a few seconds on the optimized data structure). The outputs were designed in consultation with the project advisory group and the gout GDG, but they are generalizable to other chronic conditions and are usable by other GDGs.

Limitations of the Tool

The tool has limitations in its current form: (1) it only allows index condition selection at the lower level in the hierarchical groupings of conditions (such as gout, asthma, and atrial fibrillation); (2) trial criteria selection only allows age, gender, ethnicity, comorbidities (lower level in the hierarchical groupings of conditions), and prescriptions (individual drug names, for instance, bendroflumethiazide, aspirin, and mesalazine); (3) only statistics around hospitalization and mortality are included (and not the patients outcomes from hospitalization and mortality); (4) there are limits to demography-based cross-tabulation across all output for a selected index condition; (5) the tool only includes primary care data for generating evidence; (6) the data it uses are retrospective and relies on Read codes which are being retired in UK primary care electronic records, but with mapping to Systematized Medical Nomenclature for Medicine–Clinical Terminology which is replacing them; (7) it does not allow users to extract information based on user-defined disease definitions (phenotype algorithms); instead it uses predefined disease definitions (phenotype algorithms); (8) users are required to access the tool within a TRE for generating evidence although it generates aggregate level information which meets disclosure control mechanisms in place; and (9) it performs slower for the diseases or conditions that have bigger cohort sizes, even with a subset of around 1 million patient data using an optimized data structure.

XSL•FO
RenderX

Discussion

Principal Findings

This study describes a proof-of-concept tool for analyzing clinical populations with various comorbidities in terms of the overall population and trial analysis. The tool was developed for use by guideline developers to develop clinical evidence–driven guidelines, using the specific example of gout. Typically, clinical guideline recommendations are developed using the evidence available in the published literature, mainly relying on RCT results or, in case of nonexistence of such studies, the experience of the people developing guidelines is used when preparing guideline recommendations [39].

The WHO has envisaged a future that uses clinical data in the development of recommendations [40]. The WHO's Standard-based, Machine-readable, Adaptive, Requirements-based, and Testable framework focuses on transformation from paper-based systems to smart digital systems. Recently, the focus has been shifted to computer-interpretable guideline development, highlighting the need for EHR integration with clinical practice guidelines as one of the emerging themes [41]. However, the challenges of using the computer-interpretable guidelines for patients with multimorbidity would be complex [42].

The tool described here is intended to be used by guideline developers with minimal understanding of the structure or coding of the underlying data, which makes it distinct from tools like Atlas, which are more flexible because users can define cohorts in detail, but require greater baseline knowledge and skills [43]. In the longer term, large guideline developers like NICE might find the flexibility of tools like Atlas helpful, but it is likely that smaller guideline developers with fewer resources would benefit from having access to tools that require less baseline knowledge and skills. The tool developed here addresses the issue of generating evidence from a relevant clinical population by using linked EHR data, thereby assisting guideline developers in better understanding the clinical population and making guideline recommendations. The use of such a tool can enhance the quality of the guidelines, which have often struggled to be generalizable in terms of external validity or lack of evidence from the clinical population.

Future development will focus on: (1) supporting the selection of the index condition at a higher level in the hierarchical groupings (ie, condition group and body system); (2) enhancement of the choice of trial eligibility criteria (eg, laboratory tests), more complex combination of comorbidities and prescription selection such as allowing user to select comorbidities (ie, body system, condition group, and individual condition) and prescriptions (ie, drug chapter, drug class and individual drug) with different hierarchal levels both for inclusion and exclusion criteria; (3) inclusion of outcomes data of hospitalization and mortality; (4) allowing user choice of cross-tabulation as the current form of the tool cross-tabulate everything by demography automatically (for example, data on prescribing drugs for gout be stratified by chronic kidney disease stages 3-5 and non-chronic kidney disease population stage 1-2); (5) for some guidelines, primary care is not the ideal source of data, so enhancement of the tool to query other data sets (eg, rare disease registries or specialist routine care data sets); (6) capability of supporting different coding systems other than Read code/ICD10 for disease or condition definitions using the ontological approach, such as Read codes mapping to Systematized Medical Nomenclature for Medicine-Clinical Terminology, and prescriptions based on BNF (or matching international standards to BNF); (7) enabling the capability of bringing disease or condition cohort definition algorithms other than the listed definitions supported by the tool; (8) enabling the tool to save outputs using and load results offline for analyzing within and outside the TRE after the required governance mechanism in place results from the precomputed instead of using individual patient-level data; (9) making the codebase of the tool publicly available after cleaning it, so researchers interested in generating evidence from the EHR data could benefit. As of now, the codebase in its current form is available on request; and (10) the majority of RCTs in people with physical conditions exclude more than half of the patients with the target condition. The tool is developed as a proof-of-concept to facilitate guideline developers generating evidence from the routinely collected primary care clinical data [8]. This proof-of-concept tool was intended to support guideline developers to better understand how to generalize RCT findings to the wider population, but future work could also explore its use by trialists to understand the implications of the choice of inclusion or exclusion criteria for generalizability.

We have provided access to the tool for assessment by the guideline users from NICE and SIGN. Their qualitative feedback was very encouraging, and they viewed the tool as very useful, but further evaluation of use is needed. They like to use it as an integral part of the guideline development process. The tool was developed as a proof of concept rather than a production tool. We are currently assessing how we can develop the tool further so that it can be provided as a service for guideline development groups across the United Kingdom to make more effective guideline recommendations based on evidence generated from the clinical population.

Conclusions

We have demonstrated a tool for generating evidence from clinical populations that could be used to inform clinical guideline development. The challenge of query performance on the raw CPRD GOLD data source was resolved by designing and implementing an optimized database structure. An exemplary case study of gout demonstrated that large sets of data relevant to and tailored to guideline development can be rapidly generated using the tool. The requirements for the prototype tool were determined with input from guideline developers. The final prototype version was evaluated by the guideline developers from NICE and SIGN. They responded positively and discussed how this could be made available as a service for them, which was not under the project's remit due to governance approvals in place for the study. Despite the mentioned limitations, the tool is already capable of reducing data bottlenecks of guidance development.

XSL•FO
RenderX

Acknowledgments

This work was supported by the Chief Scientist Office (grant HIPS_18_30).

Authors' Contributions

SM wrote the manuscript and developed the software (R Shiny tool); MM wrote the scripts to generate the intermediate data structure and populated it with the study-specific data from the CPRD GOLD dataset; CC, DAM, DRM, EJ, and BG led the project by developing the methodology, and also supervised the team; CG, CH, and MG managed the CPRD GOLD data and helped to conduct analysis and provided infrastructure support; HH and BG contributed to writing an introduction for the manuscript and overseeing the study outcome; EJ and BG also led the funding acquisition. All authors reviewed the final manuscript.

Conflicts of Interest

CC has received research funding from Astra Zeneca for an unrelated project.

Multimedia Appendix 1
Supplementary material S1.
[PDF File (Adobe PDF File), 201 KB - jmir_v27i1e52385_app1.pdf]

Multimedia Appendix 2
Supplementary material S2.
[PDF File (Adobe PDF File), 291 KB - jmir_v27i1e52385_app2.pdf]

Multimedia Appendix 3
Supplementary material S3.
[PDF File (Adobe PDF File), 282 KB - jmir_v27i1e52385_app3.pdf]

Multimedia Appendix 4
Supplementary material S4.
[PDF File (Adobe PDF File), 1011 KB - jmir_v27i1e52385_app4.pdf]

Multimedia Appendix 5
Supplementary material S5.
[PDF File (Adobe PDF File), 202 KB - jmir_v27i1e52385_app5.pdf]

References

1. Multimorbidity.: World Health Organization; 2016. URL: https://apps.who.int/iris/handle/10665/252275 [accessed 2022-10-13]
2. Multimorbidity: a priority for global health research. The Academy of Medical Sciences. 2018. URL: https://acmedsci.ac.uk/file-download/82222577 [accessed 2022-07-17]
3. Willadsen TG, Siersma V, Nicolaisdóttir DR, Køster-Rasmussen R, Jarbøl DE, Reventlow S, et al. Multimorbidity and mortality: a 15-year longitudinal registry-based nationwide Danish population study. J Comorb 2018;8(1):2235042X18804063 [FREE Full text] [doi: 10.1177/2235042X18804063] [Medline: 30364387]
4. Agrawal U, Azcoaga-Lorenzo A, Fagbamigbe AF, Vasileiou E, Henery P, Simpson CR, et al. Association between multimorbidity and mortality in a cohort of patients admitted to hospital with COVID-19 in Scotland. J R Soc Med 2022;115(1):22-30 [FREE Full text] [doi: 10.1177/01410768211051715] [Medline: 34672832]
5. Pearson-Stuttard J, Ezzati M, Gregg EW. Multimorbidity—a defining challenge for health systems. Lancet Public Health 2019;4(12):e599-e600 [FREE Full text] [doi: 10.1016/S2468-2667(19)30222-1] [Medline: 31812234]
6. He J, Morales DR, Guthrie B. Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review. Trials 2020;21(1):228 [FREE Full text] [doi: 10.1186/s13063-020-4139-0] [Medline: 32102686]
7. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. J Clin Epidemiol 2011;64(12):1303-1310. [doi: 10.1016/j.jclinepi.2011.04.014] [Medline: 21802903]
8. He J, Guthrie B, Morales D. A systematic review of the applicability of clinical trial evidence to real-world patients. PROSPERO. 2016. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016042282 [accessed 2022-09-06]
9. Missiou A, Tatsioni A. Systematic reviews do not comment on applicability for primary care. J Clin Epidemiol 2015;68(10):1152-1160. [doi: 10.1016/j.jclinepi.2015.06.003] [Medline: 26093310]

XSL•FO
RenderX

10. Ahmad N, Boutron I, Dechartres A, Durieux P, Ravaud P. Applicability and generalisability of the results of systematic reviews to public health practice and policy: a systematic review. Trials 2010;11:20 [FREE Full text] [doi: 10.1186/1745-6215-11-20] [Medline: 20187938]
11. Kavanagh BP. The GRADE system for rating clinical guidelines. PLoS Med 2009;6(9):e1000094 [FREE Full text] [doi: 10.1371/journal.pmed.1000094] [Medline: 19753107]
12. Lugtenberg M, Burgers JS, Clancy C, Westert GP, Schneider EC. Current guidelines have limited applicability to patients with comorbid conditions: a systematic analysis of evidence-based guidelines. PLoS One 2011;6(10):e25987. [doi: 10.1371/journal.pone.0025987] [Medline: 22028802]
13. Dumbreck S, Flynn A, Nairn M, Wilson M, Treweek S, Mercer SW, et al. Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines. BMJ 2015;350:h949 [FREE Full text] [doi: 10.1136/bmj.h949] [Medline: 25762567]
14. Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. JAMA 2006;294(6):716-724 [FREE Full text] [doi: 10.1001/jama.294.6.716]
15. Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ 2012;345:e6341 [FREE Full text] [doi: 10.1136/bmj.e6341] [Medline: 23036829]
16. Hughes LD, McMurdo MET, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age Ageing 2013;42(1):62-69. [doi: 10.1093/ageing/afs100] [Medline: 22910303]
17. Guthrie B, Thompson A, Dumbreck S, Flynn A, Alderson P, Nairn M, et al. Better guidelines for better care: accounting for multimorbidity in clinical guidelines—structured examination of exemplar guidelines and health economic modelling. Health Serv Deliv Res 2017;5(16). [Medline: 28459515]
18. Clinical Practice Research Datalink. URL: https://www.cprd.com/ [accessed 2022-10-13]
19. CPRD GOLD data specification. Clinical Practice Research Datalink. 2024. URL: https://www.cprd.com/sites/default/files/2024-08/CPRD%20GOLD%20Full%20Data%20Specification%20v2.6.pdf [accessed 2024-08-26]
20. Using linked data. Clinical Practice Research Datalink. 2024. URL: https://www.cprd.com/using-linked-data [accessed 2024-08-26]
21. Kuan V, Denaxas S, Gonzalez-Izquierdo A, Direk K, Bhatti O, Husain S, et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. Lancet Digit Health 2019;1(2):e63-e77 [FREE Full text] [doi: 10.1016/S2589-7500(19)30012-3] [Medline: 31650125]
22. HDR UK Phenotype Library. URL: https://phenotypes.healthdatagateway.org/ [accessed 2022-10-13]
23. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing 2016;45(3):353-360 [FREE Full text] [doi: 10.1093/ageing/afw039] [Medline: 26944937]
24. Metcalfe D, Masters J, Delmestri A, Judge A, Perry D, Zogg C, et al. Coding algorithms for defining Charlson and Elixhauser co-morbidities in Read-coded databases. BMC Med Res Methodol 2019;19(1):115 [FREE Full text] [doi: 10.1186/s12874-019-0753-5] [Medline: 31170931]
25. Data quality. Clinical Practice Research Datalink. URL: https://www.cprd.com/data-quality [accessed 2024-08-06]
26. Gout. National Health Service. 2017. URL: https://www.nhs.uk/conditions/gout/ [accessed 2024-08-26]
27. Ye J, Zhang Y, Zhang W, Doherty M, Lu N, Zeng C, et al. Secular trend of gout incidence in the UK: an age-period-cohort analysis. BMJ Open 2024;14(1):e079665 [FREE Full text] [doi: 10.1136/bmjopen-2023-079665] [Medline: 38286703]
28. Appendix 1—Multi-morbidities project. National Institute for Health and Care Excellence. 2022 Jun. URL: https://www.nice.org.uk/guidance/ng219/evidence/appendix-1-multimorbidities-project-pdf-11132545357 [accessed 2022-12-01]
29. Shiny. URL: https://shiny.rstudio.com/ [accessed 2022-10-14]
30. British National Formulary (BNF). URL: https://bnf.nice.org.uk/
31. Hanlon P, Hannigan L, Rodriguez-Perez J, Fischbacher C, Welton NJ, Dias S, et al. Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: an individual-level participant data analysis. BMC Med 2019;17(1):201 [FREE Full text] [doi: 10.1186/s12916-019-1427-1] [Medline: 31711480]
32. ISAC minutes and annual reports. Clinical Practice Research Datalink. URL: https://www.cprd.com/isac-minutes-and-annual-reports [accessed 2022-09-21]
33. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. Br J Clin Pharmacol 2010;69(1):4-14 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03537.x] [Medline: 20078607]
34. Meier CR, Jick H. Omeprazole, other antiulcer drugs and newly diagnosed gout. Br J Clin Pharmacol 1997;44(2):175-178 [FREE Full text] [doi: 10.1046/j.1365-2125.1997.00647.x] [Medline: 9278205]
35. Rivard C, Thomas J, Lanaspa MA, Johnson RJ. Sack and sugar, and the aetiology of gout in England between 1650 and 1900. Rheumatology 2013;52(3):421-426. [doi: 10.1093/rheumatology/kes297] [Medline: 23175570]
36. Savica V, Santoro D, Ricciardi B, Ricciardi CA, Calo LA, Bellinghieri G. Morbus dominorum: gout as the disease of lords. J Nephrol 2013;26:113-116. [doi: 10.5301/jn.5000349] [Medline: 24375354]

XSL•FO
RenderX

37. Bowen-Davies Z, Muller S, Mallen CD, Hayward RA, Roddy E. Gout severity, socioeconomic status, and work absence: a cross-sectional study in primary care. Arthritis Care Res 2018;70(12):1822-1828. [doi: 10.1002/acr.23562] [Medline: 29579363]
38. Becker MA, Schumacher HR, Wortmann RL, MacDonald PA, Eustace D, Palo WA, et al. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. N Engl J Med 2005;353(23):2450-2461. [doi: 10.1056/NEJMoa050373] [Medline: 16339094]
39. Factsheet 1: How NICE develops clinical guidelines and what documents we publish. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/media/default/About/NICE-Communities/Public-involvement/Developing-NICE-guidance/Factsheet-1-contribute-to-developing-clinical-guidelines.pdf [accessed 2022-10-15]
40. SMART Guidelines. World Health Organization. URL: https://www.who.int/teams/digital-health-and-innovation/smart-guidelines [accessed 2022-10-15]
41. Peleg M. Computer-interpretable clinical guidelines: a methodological review. J Biomed Inform 2013;46(4):744-763 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.009] [Medline: 23806274]
42. Bilici E, Despotou G, Arvanitis TN. The use of computer-interpretable clinical guidelines to manage care complexities of patients with multimorbid conditions: a review. Digit Health 2018;4:2055207618804927 [FREE Full text] [doi: 10.1177/2055207618804927] [Medline: 30302270]
43. OHDSI/Atlas. GitHub. 2024. URL: https://github.com/OHDSI/Atlas [accessed 2024-08-26]

Abbreviations

BNF: British National Formulary
CPRD: Clinical Practice Research Datalink
eFI: electronic frailty index
EHR: electronic health record
GDG: guideline development group
GRADE: Grades of Recommendation Assessment, Development and Evaluation
HIC: Health Informatics Centre
ICD10: International Classification of Diseases 10th Revision
IMD: index of multiple deprivations
NICE: National Institute for Health and Care Excellence
RCT: randomized controlled trial
SIGN: Scottish Intercollegiate Guidelines Network
TRE: Trusted Research Environment
WHO: World Health Organization

Edited by F Xie; submitted 01.09.23; peer-reviewed by S Jameela, S Bonn, O Kumpf, W Zhang, S Sun, H Liu, S Sun; comments to author 08.07.24; revised version received 10.09.24; accepted 25.09.24; published 16.01.25.

Please cite as:
Mumtaz S, McMinn M, Cole C, Gao C, Hall C, Guignard-Duff M, Huang H, McAllister DA, Morales DR, Jefferson E, Guthrie B
A Digital Tool for Clinical Evidence–Driven Guideline Development by Studying Properties of Trial Eligible and Ineligible Populations: Development and Usability Study
J Med Internet Res 2025;27:e52385
URL: https://www.jmir.org/2025/1/e52385
doi:10.2196/52385
PMID:

©Shahzad Mumtaz, Megan McMinn, Christian Cole, Chuang Gao, Christopher Hall, Magalie Guignard-Duff, Huayi Huang, David A McAllister, Daniel R Morales, Emily Jefferson, Bruce Guthrie. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Machine Learning for the Early Prediction of Delayed Cerebral Ischemia in Patients With Subarachnoid Hemorrhage: Systematic Review and Meta-Analysis

Haofuzi Zhang[1*], MD, PhD; Peng Zou[1*], MD, PhD; Peng Luo[1*], MD, PhD; Xiaofan Jiang[1], MD, PhD

Department of Neurosurgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Xiaofan Jiang, MD, PhD
Department of Neurosurgery
Xijing Hospital
Fourth Military Medical University
No. 127, Changle West Road
Xincheng District
Xi'an,
China
Phone: 86 186 0298 0377
Email: jiangxf@fmmu.edu.cn

Abstract

Background: Delayed cerebral ischemia (DCI) is a primary contributor to death after subarachnoid hemorrhage (SAH), with significant incidence. Therefore, early determination of the risk of DCI is an urgent need. Machine learning (ML) has received much attention in clinical practice. Recently, some studies have attempted to apply ML models for early noninvasive prediction of DCI. However, systematic evidence for its predictive accuracy is still lacking.

Objective: The aim of this study was to synthesize the prediction accuracy of ML models for DCI to provide evidence for the development or updating of intelligent detection tools.

Methods: PubMed, Cochrane, Embase, and Web of Science databases were systematically searched up to May 18, 2023. The risk of bias in the included studies was assessed using PROBAST (Prediction Model Risk of Bias Assessment Tool). During the analysis, we discussed the performance of different models in the training and validation sets.

Results: We finally included 48 studies containing 16,294 patients with SAH and 71 ML models with logistic regression as the main model type. In the training set, the pooled concordance index (C index), sensitivity, and specificity of all the models were 0.786 (95% CI 0.737-0.835), 0.77 (95% CI 0.69-0.84), and 0.83 (95% CI 0.75-0.89), respectively, while those of the logistic regression models were 0.770 (95% CI 0.724-0.817), 0.75 (95% CI 0.67-0.82), and 0.71 (95% CI 0.63-0.78), respectively. In the validation set, the pooled C index, sensitivity, and specificity of all the models were 0.767 (95% CI 0.741-0.793), 0.66 (95% CI 0.53-0.77), and 0.78 (95% CI 0.71-0.84), respectively, while those of the logistic regression models were 0.757 (95% CI 0.715-0.800), 0.59 (95% CI 0.57-0.80), and 0.80 (95% CI 0.71-0.87), respectively.

Conclusions: ML models appear to have relatively desirable power for early noninvasive prediction of DCI after SAH. However, enhancing the prediction sensitivity of these models is challenging. Therefore, efficient, noninvasive, or minimally invasive low-cost predictors should be further explored in future studies to improve the prediction accuracy of ML models.

Trial Registration: PROSPERO (CRD42023438399); https://tinyurl.com/yfuuudde

(J Med Internet Res 2025;27:e54121) doi:10.2196/54121

KEYWORDS

machine learning; subarachnoid hemorrhage; delayed cerebral ischemia; systematic review

XSL•FO
RenderX

Introduction

Subarachnoid hemorrhage (SAH) is the third most common type of stroke and is usually associated with the rupture of an aneurysm [1]. The incidence of SAH is approximately 7.9 (95% CI 6.9-9.0) per 100,000 person-years globally, 8.3 (95% CI 7.2-9.5) per 100,000 person-years in Europe, 10.4 (95% CI 5.9-18.6) per 100,000 person-years in Asia, and 8.5 (95% CI 7.1-10.2) per 100,000 person-years in North America [2]. Recently, a discernible decline has been observed in global SAH incidence, which is potentially attributable to effective public health initiatives and the promotion of healthy lifestyles, especially effective management of certain modifiable risk factors (eg, reduction in smoking prevalence, management of hypertension) [3]. However, aneurysmal SAH continues to pose significant challenges, as evidenced by an in-hospital mortality rate of nearly 20% [4]. Of these, delayed cerebral ischemia (DCI) is the leading cause of death, with a prevalence of approximately 29% (95% CI 26%-32%) [5,6].

The early identification of DCI risk continues to pose significant challenges in clinical practice, with a notable absence of recognized and effective early predictive tools. Previous studies have explored diverse methods for the early prediction of DCI, such as transcranial Doppler (TCD) ultrasonography [7], computed tomography perfusion parameters [8], and continuous cranial electroencephalography monitoring [9]. Recent advancements in computational technologies and the ongoing improvement of statistical theory have facilitated the integration of artificial intelligence into clinical practice, mainly for the diagnosis of disease states, assessment of disease courses, and prediction of prognosis [10]. In this context, some studies have attempted to develop machine learning (ML) models for the early prediction of DCI after SAH; however, its accuracy remains debatable [11-14]. In addition, there is a lack of systematic investigations on effective predictors for the early prediction of DCI. Therefore, we conducted this meta-analysis to systematically discuss the accuracy of ML models in DCI prediction after SAH and to offer evidence for advancing the application of artificial intelligence in this field.

Methods

Study Registration

This study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [15]. This study protocol was registered in the International Prospective Register of Systematic Reviews on July 6, 2023 (CRD42023438399).

Eligibility Criteria

The inclusion criteria were as follows: (1) patients with SAH, with the outcome event being DCI; (2) complete construction of an ML model for DCI prediction; (3) primary studies even without independent external validation or any kind of independent validation set; (4) published studies that constructed different ML models based on the same dataset; (5) cohort studies, case-control studies, or cross-sectional studies; and (6) studies published in English. The exclusion criteria were as follows: (1) meta-analysis, guidelines, review, or expert comments; (2) primary studies that analyzed risk factors/independent predictors only, without complete construction of an ML model; (3) absence of any outcome indicators for estimating the prediction accuracy of ML models, including the area under the receiver operating characteristic curve, concordance index (C index), accuracy, sensitivity, specificity, precision, confusion matrix, and F_1-score; and (4) studies with sample size <20.

Data Sources and Search Strategy

PubMed, Cochrane, Embase, and Web of Science databases were systematically searched up to May 18, 2023, by using Medical Subject Headings (MeSH) and free-text words, with no restrictions on publication year or country. The search strategy is described in Multimedia Appendix 1.

Study Selection and Data Retrieval

The retrieved studies were imported into EndNote, and duplicates were excluded through software automarking and manual marking. Then, the titles or abstracts of the remaining studies were scanned, followed by full-text reviews to obtain the final eligible studies. A standardized table was established before data retrieval, and the extracted information included title, first author, publication year and country, study type, DCI cases, total cases, DCI cases in the training and validation sets, cases in the training and validation sets, generation method of the validation set, methods for tackling overfitting and missing data, variable screening methods, model type, and predictors. Literature screening was implemented independently by 2 researchers (HZ and PZ) and then cross-checked. Debatable studies were discussed with a third researcher (PL) to make a final decision.

Assessment of Study Quality

The risk of bias in the primary studies was appraised using PROBAST (Prediction Model Risk of Bias Assessment Tool), which contains many questions on participants, predictors, outcomes, and statistical analyses, reflecting the overall risk of bias and applicability [16]. These domains had 2, 3, 6, and 9 specific questions, with 3 responses (yes/probably yes, no/probably no, and no information) for each question. A domain was regarded as having a high risk of bias if it had at least one no/probably no answer or a low risk of bias if all answers were yes or probably yes. The overall risk was assessed as low with all domains at low risk and high with at least one domain at high risk. Two researchers (HZ and PZ) assessed the risk of bias independently based on PROBAST and cross-checked at the end. If there was a debatable result, a third researcher (PL) was asked to assist in the assessment.

Outcome Indicators

Currently, the C index is an important outcome index for assessing the prediction accuracy of ML models. However, in case of severe imbalance in case number (grossly too few cases for DCI), the C index can hardly explain the specific accuracy of the model for both positive and negative events, and thus, our primary outcome indicators also included sensitivity and specificity at the optimal thresholds of the model.

XSL•FO
RenderX

Statistical Analysis

In some primary studies, when 95% CIs and standard errors were absent in the C index, we referred to Debray et al [17] to assess its standard error. As ML models contained different variables and inconsistent parameters, we gave priority to random-effects models in the meta-analysis of C index. The C index is currently an important end point metric for evaluating ML performance, especially in binary classification and survival analysis [18,19]. Many current binary classification ML studies have used the C index as the main metric for assessing model performance. In addition, ML obtains the probability of positive events. We also need to understand the sensitivity and specificity of ML prediction of positive events under the optimal risk probability cutoff threshold. Therefore, we need to perform a meta-analysis of the sensitivity and specificity based on the diagnostic 2×2 table.

A bivariate mixed-effects model was applied for the meta-analysis of sensitivity and specificity based on the diagnostic fourfold table. Most primary studies did not report the diagnostic fourfold table; therefore, the following 2 ways were used to calculate the diagnostic fourfold table: (1) using sensitivity, specificity, precision, and case number and (2) using sensitivity and specificity extracted based on the optimal Youden index as well as case number. In the meta-analysis of the C index, sensitivity and specificity evaluation were performed in Stata 15.0, and the program packages involved were metan and midas. The meta-analysis of the incidence of DCI was performed in R4.4.1 (R development Core Team). Because of the meta-analysis of the incidence, we need to transform the distribution of the incidence, and the R language packages involved are metafor and meta.

Results

Study Selection

We searched 3486 studies (97 from Cochrane, 266 from PubMed, 522 from Embase, and 2601 from Web of Science) and eliminated 586 duplicates. After reading the title or abstract, we excluded 2831 studies. Among the remaining 69 studies, 48 eligible primary studies [11-14,20-63] were finally included after reviewing the full texts (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) literature screening flowchart.

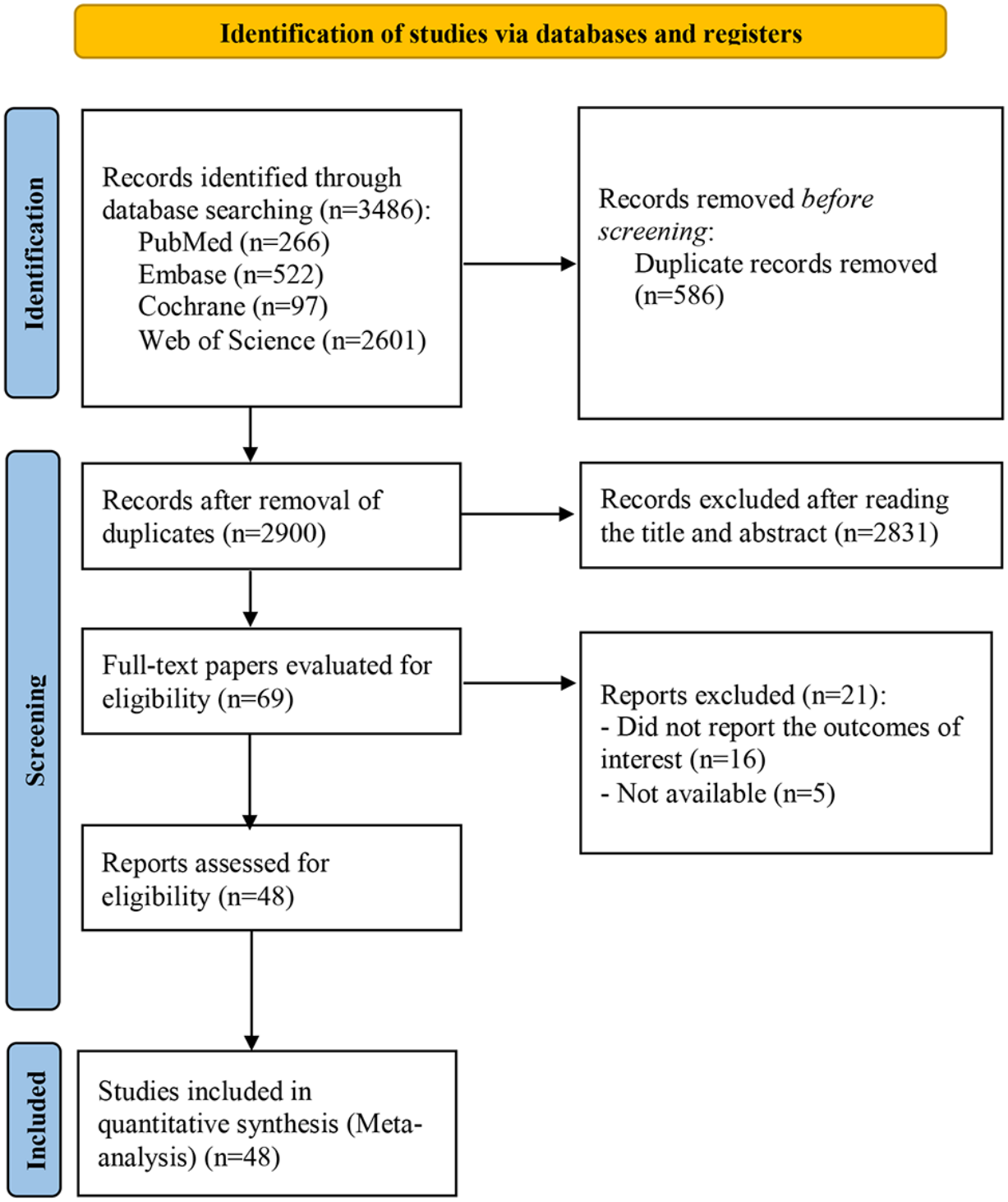

Study Characteristics

The 48 primary studies were published between 2013 and 2023, including 15,346 patients with SAH, of which 3863 patients had DCI. These primary studies were from 13 countries, including the United States [11-13, 21,24, 28, 31, 33, 36, 37, 46, 47, 49-51, 55, 57], China [20, 22, 23, 25, 26, 30, 35, 39, 41, 44, 53, 59], and the Netherlands [40,48,58,62]. In addition to one case-control study [11], the others were cohort studies. Ten studies were multicentered [24-26,28,37,39,40,42,58,61], and the remaining were one-centered. Besides, 21 studies [14, 21, 22, 27, 28, 30, 32, 33, 43, 45-47, 50, 52, 56-61, 63] lacked an independent validation set, and 25 studies [12, 20, 21, 23, 25, 26, 31, 35-42, 45, 47-51, 53-55, 62] reported the methods to avoid overfitting. Nineteen studies [12, 14, 20, 22, 24-26, 30, 31, 33, 34, 36-38, 48-51, 62] had missing data, which were mainly resolved by deletion, mean value imputation, and missing value imputation. In terms of variable screening, only 1 study [50] used a single-factor screening method (Multimedia Appendix 2).

XSL•FO
RenderX

Risk of Bias

The 48 primary studies contained 71 ML models. In the assessment of participants and predictors, a high risk of bias was mainly derived from the ML models constructed in 1 case-control study. In the assessment of outcomes, only 1 model in the case-control study was rated as unclear risk. As for the statistical analyses, the models were mainly rated at high risk since most models did not satisfy modeling sample size estimation, that is, events per variable were >20 in the training set (DCI cases were greater than 20 times that were ultimately included in the model) and needed an independent validation set with over 100 validated cases. In addition, the methods for tackling overfitting and underfitting were not clearly described. The methods for handling missing data during quality assessment remained a concern, as PROBAST only recognized the multiple imputation method as low risk of bias (Figure 2).

Figure 2. Risk of bias of the included machine learning models.

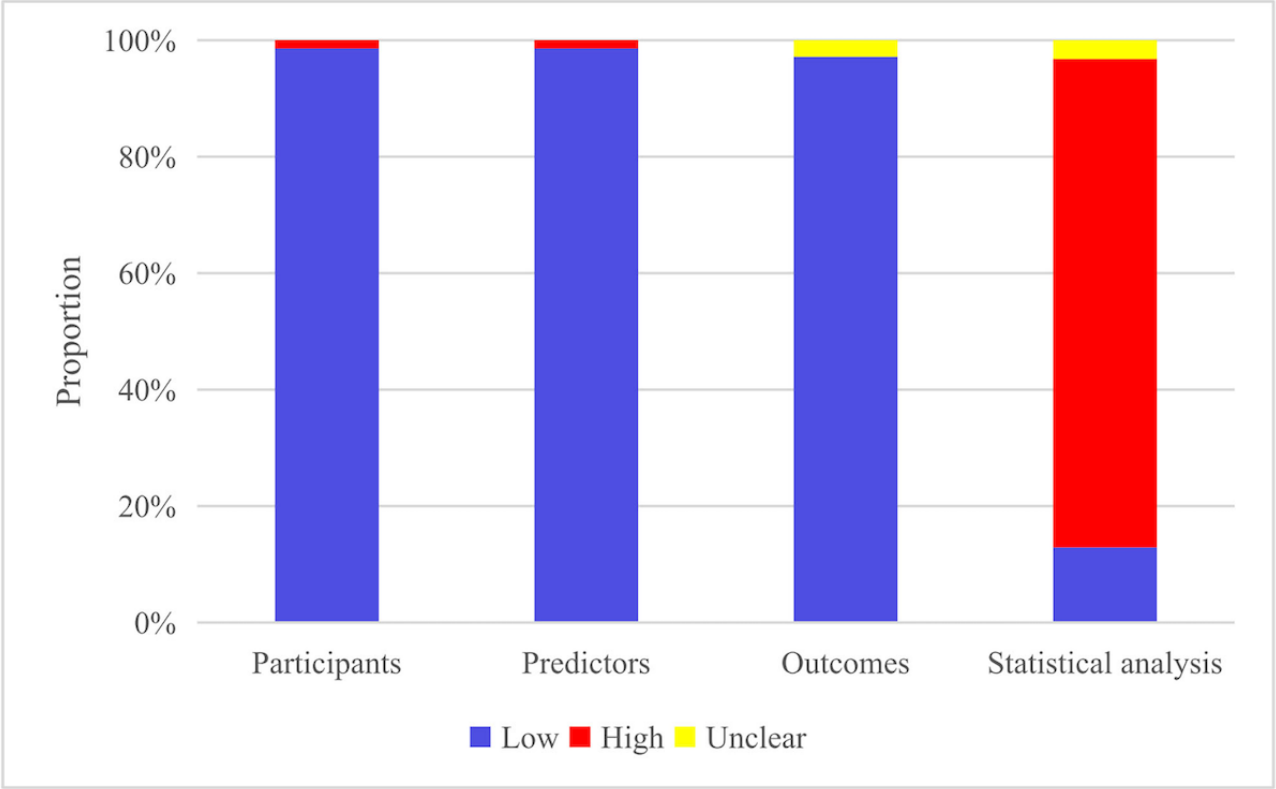

Meta-Analysis

Of the 48 studies included, 45 explicitly reported the numbers of DCI and total SAH, and the prevalence of DCI, which was pooled using a random-effects model, was 27.7% (95% CI 24.7%-30.7%) (Figure 3 [11-14,20-39,41-50,52-59,61-63]).

XSL•FO
RenderX

Figure 3. Forest plot of delayed cerebral ischemia incidence in the training set. IV: independent variable.

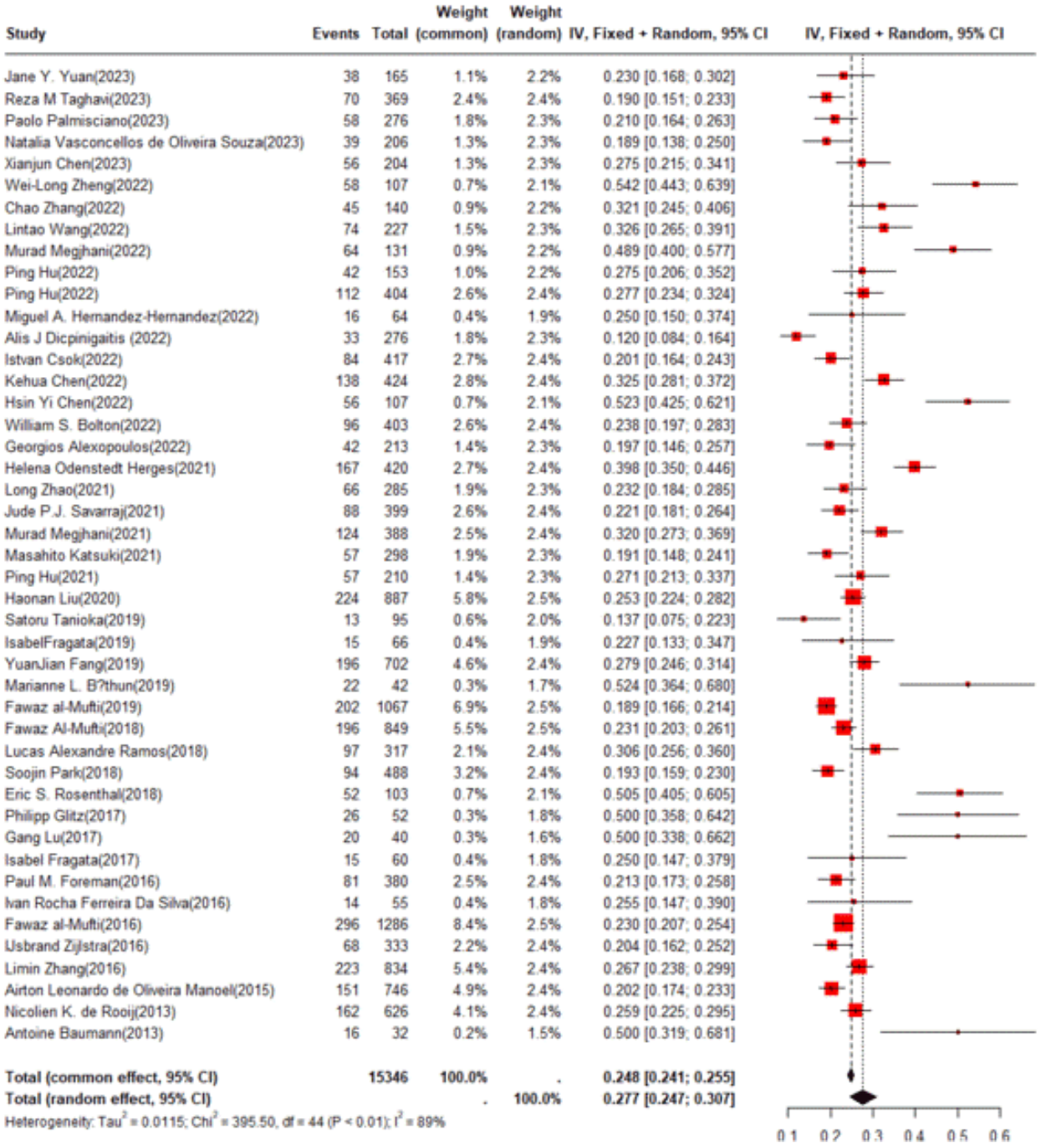

Newly Developed ML Models

C Index

In the training set, C index was reported in 45 ML models, with an overall C index of 0.786 (95% CI 0.737-0.835). The 31 independent validation cohorts manifested an overall C index of 0.767 (95% CI 0.741-0.793). Among them, the logistic regression–based predictive nomogram was the main model, with an overall C index of 0.770 (95% CI 0.724-0.817) in the training set and 0.757 (95% CI 0.715-0.800) in the validation set (Table 1, Figure 4 [1,26,27,32,37,38,41,42,44,49,50,54], Figure 5 [12,20,26,27,30,34-37,39-43,50,52]).

XSL•FO
RenderX

Table 1. Meta-analysis of the concordance index of machine learning models for delayed cerebral ischemia prediction in the training and validation sets.

Machine learning model	Training set		Validation set	
	n	Concordance index (95% CI)	n	Concordance index (95% CI)
Artificial neural network	4	0.721 (0.660-0.783)	4	0.756 (0.676-0.835)
Logistic regression	24	0.770 (0.724-0.817)	11	0.757 (0.715-0.800)
Decision tree	2	0.829 (0.782-0.877)	2	0.705 (0.610-0.799)
Ensemble classifier	1	0.830 (0.795-0.865)	—[a]	—
K-nearest neighbor	1	0.992 (0.986-0.999)	1	0.792 (0.683-0.901)
LASSO (least absolute shrinkage and selection operator)	1	0.793 (0.692-0.893)	2	0.817 (0.661-0.973)
Random forest	5	0.770 (0.646-0.894)	5	0.784 (0.730-0.838)
Support vector machine	5	0.789 (0.619-0.959)	3	0.684 (0.606-0.761)
Extreme gradient boosting	2	0.906 (0.860-0.952)	3	0.826 (0.770-0.882)
Overall	45	0.786 (0.737-0.835)	31	0.767 (0.741-0.793)

[a]Not available.

Figure 4. Forest plot of the concordance index of other machine learning models in the training set.

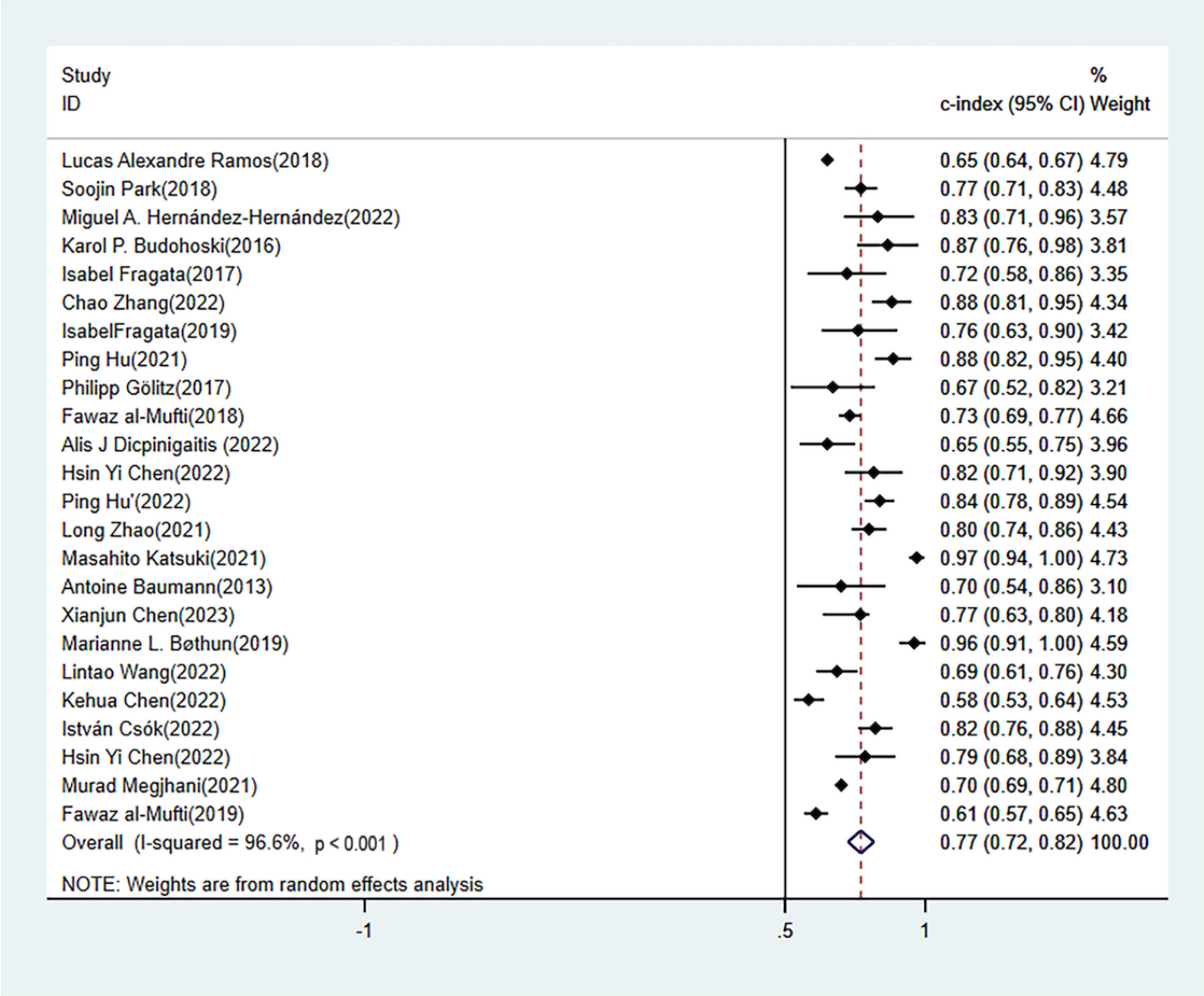

XSL•FO
RenderX

Figure 5. Forest plot of the concordance index of the machine learning models in the validation set. LASSO: least absolute shrinkage and selection operator; XGBoost: extreme gradient boosting.

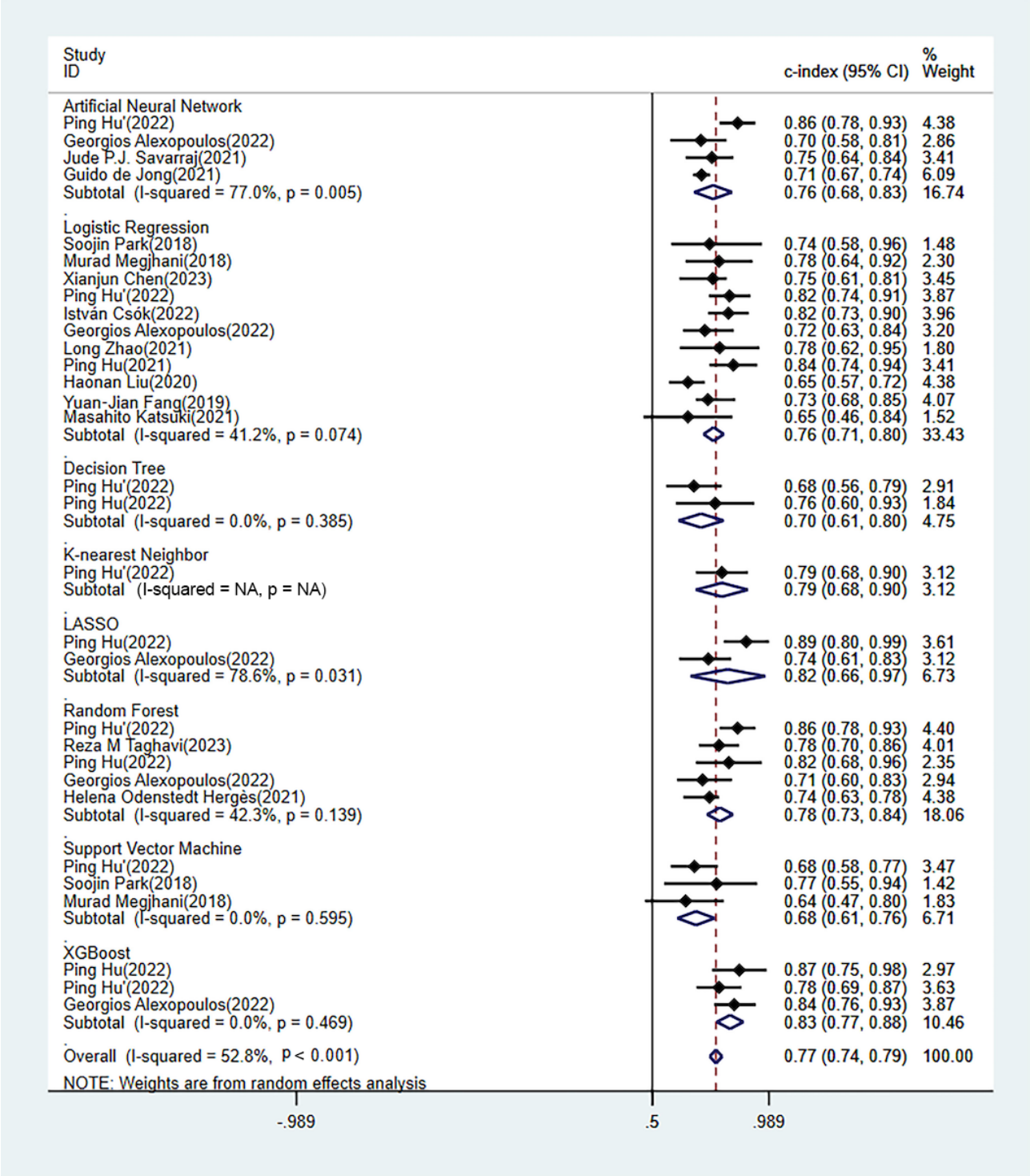

Sensitivity and Specificity

In the training set, we extracted diagnostic 2×2 tables from 37 ML models, with overall sensitivity of 0.77 (95% CI 0.69-0.84) and overall specificity of 0.83 (95% CI 0.75-0.89). The diagnostic 2×2 tables were also extracted from 23 independent validation cohorts, with overall sensitivity of 0.66 (95% CI 0.53-0.77) and overall specificity of 0.78 (95% CI 0.71-0.84). Among them, the logistic regression–based predictive nomogram had overall sensitivity of 0.75 (95% CI 0.67-0.82) and overall specificity of 0.71 (95% CI 0.63-0.78) in the training set and 0.59 (95% CI 0.57-0.80) and 0.80 (95% CI 0.71-0.87), respectively, in the validation set (Table 2, Figure 6 [12,20-22,24-32,36,40,43,46-49,53-55,63], Figure 7 [13,20,26,27,30,34-37,40,45]).

XSL•FO
RenderX

Table 2. Meta-analysis of the sensitivity and specificity of the machine learning models for delayed cerebral ischemia prediction in the training and validation sets.

Machine learning model	Training set			Validation set		
	n	Sensitivity (95% CI)	Specificity (95% CI)	n	Sensitivity (95% CI)	Specificity (95% CI)
Artificial neural network	2	0.55-0.64	0.72-0.91	2	0.37-0.82	0.72-0.84
Logistic regression	20	0.75 (0.67-0.82)	0.71 (0.63-0.78)	5	0.59 (0.57-0.80)	0.80 (0.71-0.87)
Decision tree	2	0.62-0.78	0.82-0.88	2	0.44-0.67	0.80-0.81
K-nearest neighbor	1	0.66	1	1	0.41	0.95
LASSO (least absolute shrinkage and selection operator)	1	0.48	0.89	2	0.54-0.55	0.74-0.78
Random forest	6	0.93 (0.28-1.00)	0.97 (0.66-1.00)	4	0.81 (0.46-0.96)	0.78 (0.63-0.88)
Support vector machine	3	0.63-0.95	0.73-1.00	1	0.26	0.96
Extreme gradient boosting	2	0.93-0.98	0.60-0.61	4	0.78 (0.27-0.97)	0.55 (0.37-0.71)
Overall	37	0.77 (0.69-0.84)	0.83 (0.75-0.89)	23	0.66 (0.53-0.77)	0.78 (0.71-0.84)

Figure 6. Forest plot of the sensitivity and specificity of the machine learning models in the training set.

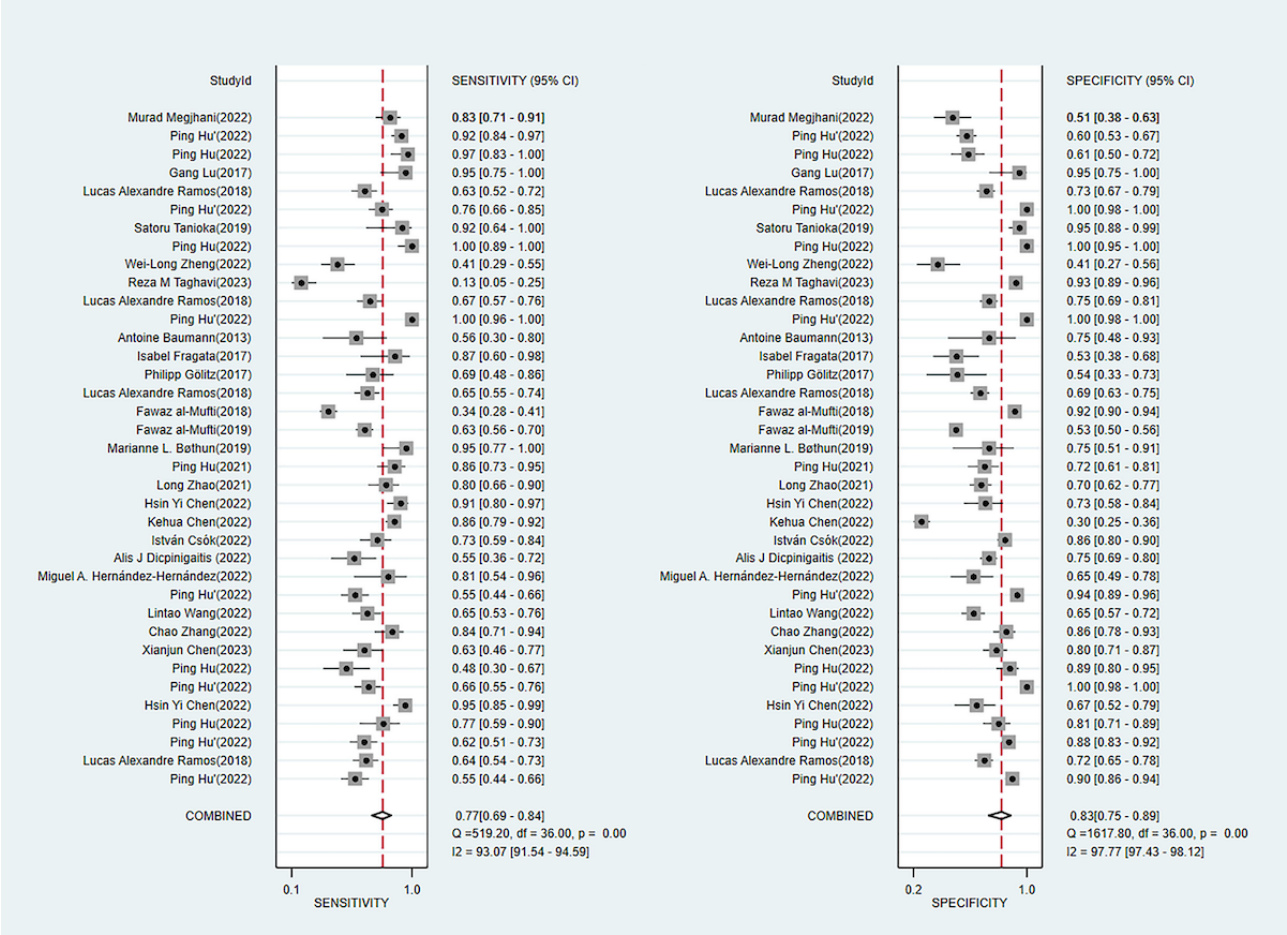

XSL•FO
RenderX

Figure 7. Forest plot of the sensitivity and specificity of the machine learning models in the validation set.

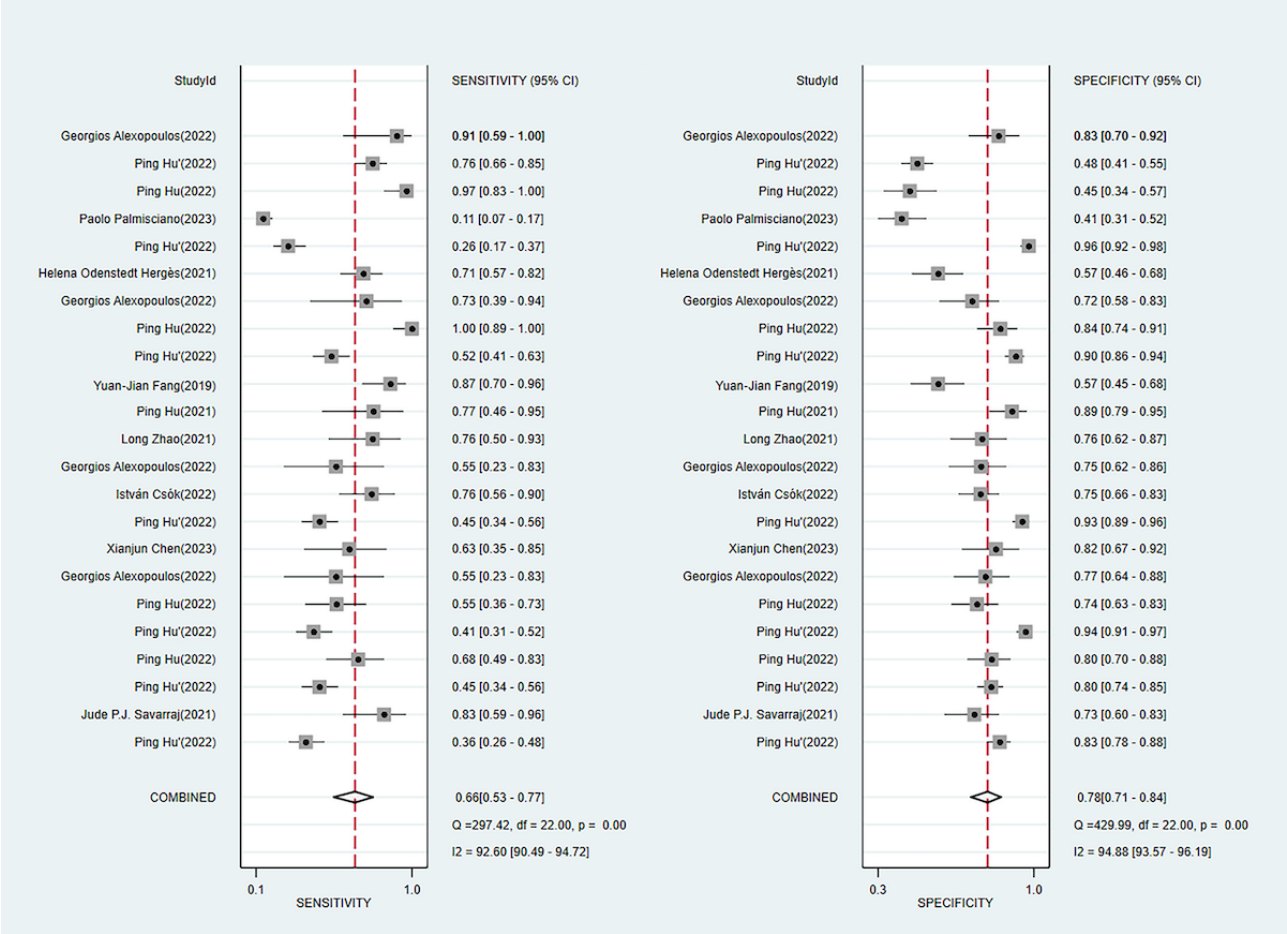

Validation of Previous Scales

The previous models mainly contained VASOGRADE and Practical Risk Chart, with an overall C index of 0.661 (95% CI 0.627-0.696), sensitivity of 0.65 (95% CI 0.59-0.87), and specificity of 0.64 (95% CI 0.52-0.74) (Figure 8 [14,41,56,62,63], Figure 9 [14,56,62,63]).

XSL•FO
RenderX

Figure 8. Forest plot of the concordance index of previous scales in the training set.

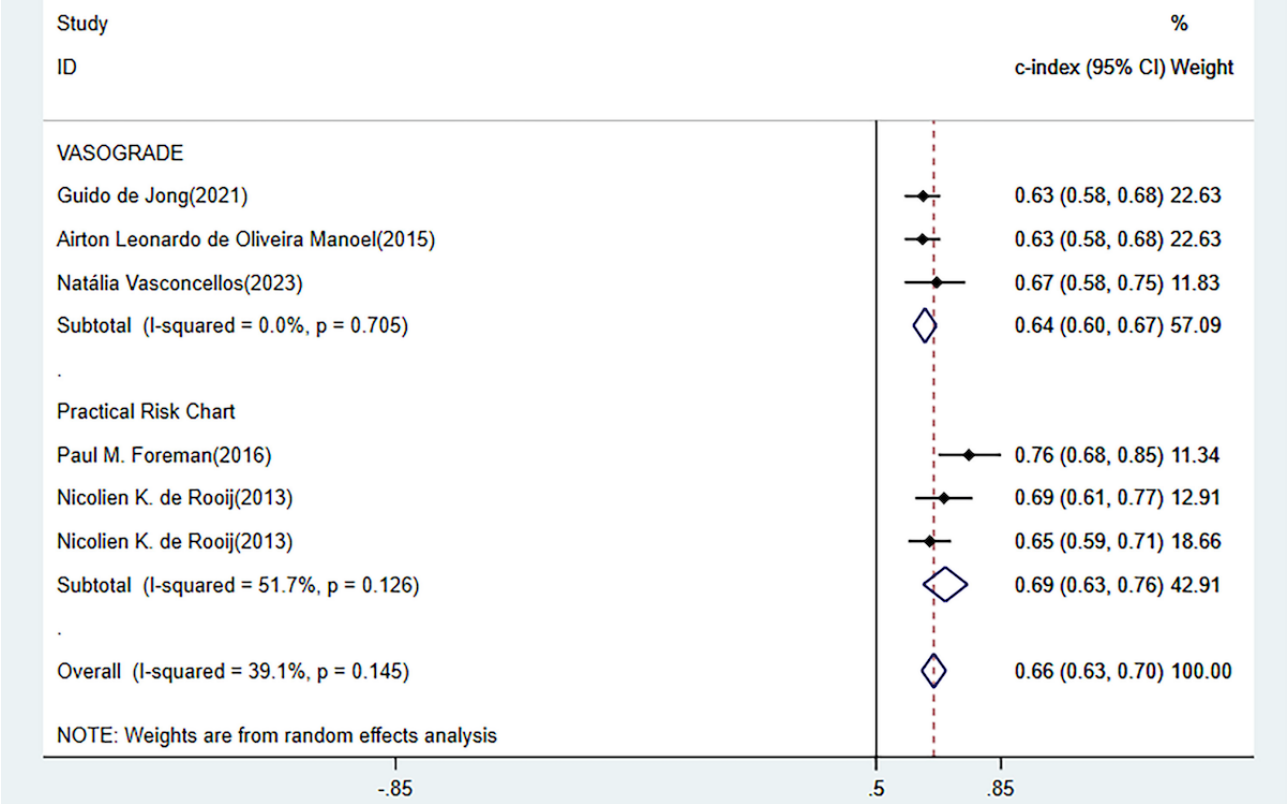

Figure 9. Forest plot of the sensitivity and specificity of previous scales in the training and validation sets.

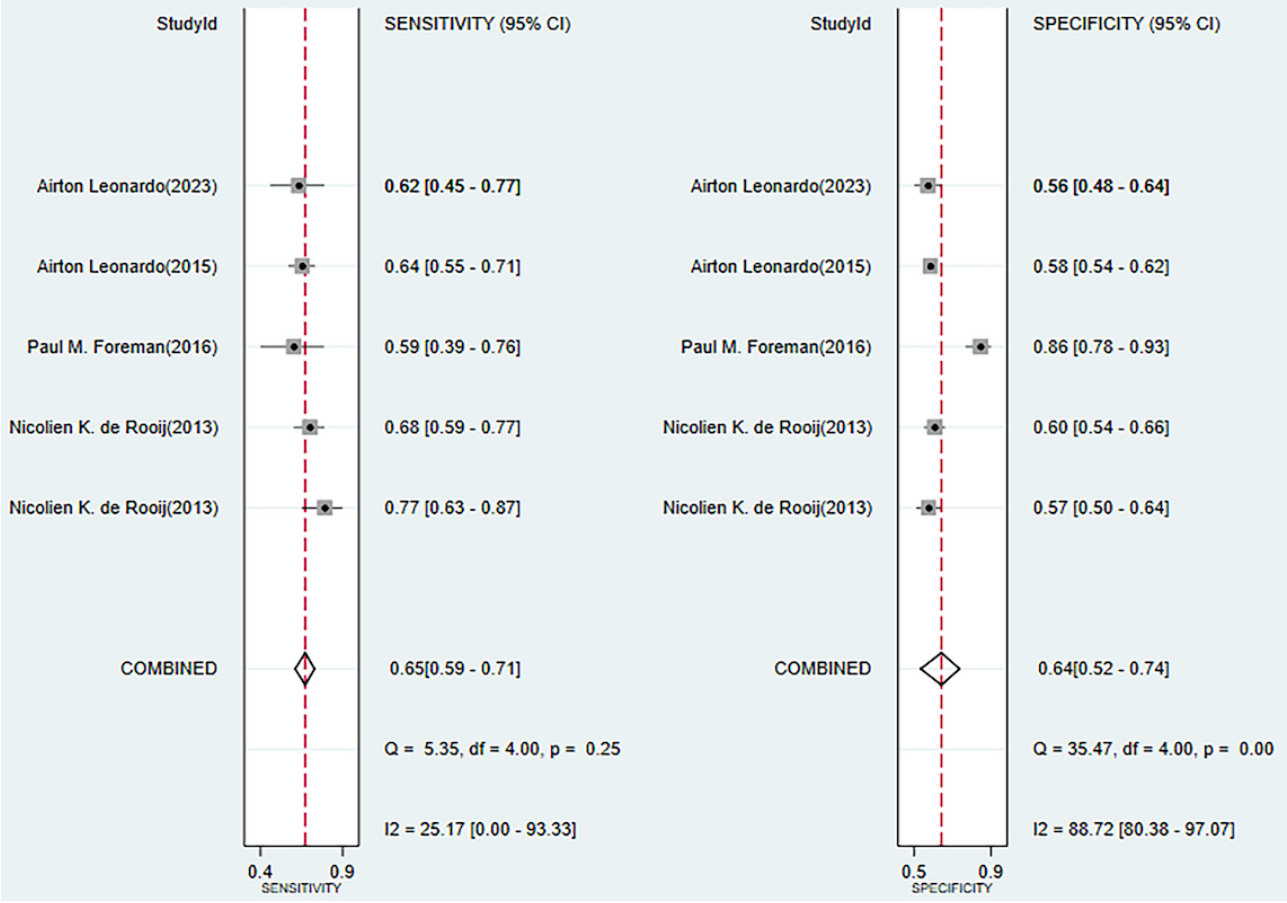

XSL•FO
RenderX

Predictors

A total of 59 predictors were reported in the included ML models, and the top 10 were age (53 models), gender (46 models), Hunt and Hess score (41 models), modified Fisher scale (40 models), World Federation of Neurological Surgeons (35 models), aneurysm location (32 models), hypertension (29 models), coiling or clipping (28 models), smoking/drinking history (28 models), diabetes (23 models), and laboratory results (21 models). All modeling predictors are shown in Multimedia Appendix 3.

Discussion

Main Findings

Our study systematically explores the accuracy of ML for DCI prediction after SAH, and the results showed that ML had a relatively ideal C index for DCI prediction; however, the sensitivity still needs improving. The logistic regression–based nomogram was the main model. The predictors in the 71 models were mainly age, gender, Hunt and Hess score, modified Fisher scale, World Federation of Neurological Surgeons, aneurysm location, hypertension, coiling or clipping, smoking/drinking history, diabetes, and laboratory results.

Comparison With Previous Studies

Continuous investigations have been undertaken to find reliable approaches for the early detection of DCI risk, mainly focusing on imaging techniques. Kumar et al [7] conducted a systematic review of 17 primary studies and reported the accuracy of TCD for predicting vasospasm for the diagnosis of DCI, with an overall 90% (95% CI 77%-96%) sensitivity and 71% (95% CI 51%-84%) specificity. Although their study demonstrated high sensitivity and relatively good specificity, we found that they included many retrospective studies, which may exaggerate the results. Therefore, more randomized diagnostic trials are also needed for validation. Han et al [8] provided evidence for the early prediction of DCI by using computed tomography perfusion parameters, with 97% sensitivity and 89% specificity. A systematic review by Scherschinski et al [9] described the prediction accuracy of continuous cranial electroencephalography monitoring for DCI (sensitivity 11.1%, 95% CI 3.1%-26.1% and specificity 82.9%, 95% CI 66.4%-93.4%). In addition, a systematic review by Kumar et al [64] described the prediction of DCI by cerebral angiography (sensitivity 57%, 95% CI 40%-81%; specificity 68%, 95% CI 61%-76%). Of these imaging tools, Doppler ultrasound appears to have high prediction accuracy, especially with high sensitivity.

Furthermore, Santana et al [65] reported the accuracy of ML in DCI prediction after SAH through a systematic review, which, however, only included 6 primary studies. Their results demonstrated that the overall sensitivity of logistic regression was 0.71 (95% CI 0.57-0.84) and specificity was 0.63 (95% CI 0.42-0.85), and other ML models had a pooled sensitivity of 0.74 (95% CI 0.61-0.86) and specificity of 0.78 (95% CI 0.71-0.86). However, the training and validation set results did not seem to be considered in their studies. In our systematic review incorporating 48 primary studies, the overall C index in the validation set was 0.767 (95% CI 0.741-0.793), with sensitivity of 0.66 (95% CI 0.53-0.77) and specificity of 0.78 (95% CI 0.71-0.84). Meanwhile, logistic regression was the main model, with low sensitivity in the validation set.

In clinical practice, the interpretability of the model and the prediction accuracy for outcomes should be considered; however, it is hard to choose between them [66]. Some models with high interpretability, that is, logistic regression (converting linear combinations of predictor variables into probability values for positive events via logistic functions), decision trees (generating rules by recursively splitting variables in the dataset), and LASSO (least absolute shrinkage and selection operator) regression (selecting variables and reducing model complexity by L1 regularization), which adds a penalty term proportional to the absolute value of coefficients of the features for feature selection and model construction, often have too low prediction accuracy for outcomes. Others with too little explainability, that is, extreme gradient boosting (XGBoost; combining multiple weak learners to form a strong learner with efficient computational power and excellent prediction performance), artificial neural network (an algorithm that simulates the neuronal structure of the human brain and learns complex patterns of data through multilayer nonlinear transformations for various ML tasks such as classification, regression, and clustering), and support vector machines (a supervised learning method to maximize the spacing between different classes by finding optimal hyperplanes for classification and regression tasks) have very high prediction accuracy, especially deep learning models [67]. Among ML models constructed based on common clinical features, models with better interpretability are preferred, which can explicitly explain the relationship between common clinical features and outcomes and thus provide real-time interventions. This may also explain why logistic regression was most frequently used in the included studies. However, in image-based artificial intelligence, models with higher prediction efficiency are preferred, mainly since medical images have many confounding factors and types of imaging devices, as well as a high level of device parameterization [10,68]. Our study shows that XGBoost in the validation set had the best C index of 0.826 (95% CI 0.770-0.882), with sensitivity and specificity of 0.78 (95% CI 0.27-0.97) and 0.55 (95% CI 0.37-0.71), respectively. Although XGBoost can rank the variables' importance, its interpretability remains a serious challenge. In addition, random forests seem to exhibit desirable accuracy, with C index, sensitivity, and specificity of 0.784 (95% CI 0.730-0.838), 0.81 (95% CI 0.46-0.96), and 0.78 (95% CI 0.63-0.88), respectively. The c-indices of logistic regression, LASSO regression, and decision tree based on models with better interpretability were 0.757 (0.715-0.800), 0.705 (0.610-0.799), and 0.817 (0.661-0.973), respectively. This shows the differences in the prediction accuracy of different models, where the less interpretable XGBoost and random forest seem to have better accuracy, especially random forest, which has desirable sensitivity and specificity.

Our study shows that existing ML has a more favorable C index for the prediction of DCI; however, the overall sensitivity seems to be unfavorable. Although random forest has high sensitivity

XSL•FO
RenderX

and specificity, the included models are limited, which may limit the interpretation of results. In constructing ML models, we need to consider data balance. For unbalanced modeling, some results tend to show a high C index, which is due to negative results. Our pooled incidence of DCI for the included studies was 27.7% (95% CI 24.7%-30.7%), which may have some implications for low sensitivity. In addition, some models obtain the probability of DCI, and the probability thresholds are often chosen based on the optimal Youden index. The effect of sensitivity is often not considered when choosing the optimal Youden index, which may also have some effect on low sensitivity. Therefore, the balance of modeling data and the reasonableness of positive probability thresholds need to be considered in future studies.

When constructing clinical prediction models, the number of included cases and the selection of predictors are often neglected. Models derived from a restricted number of cases exhibited low stability, and some included studies had a limited sample size. In addition, efficient predictors are key to enhancing model accuracy, and our summary of the current modeling predictors provides important references for subsequent studies. A systematic review by Jabbarli et al [69] explored the biomarkers for the early prediction of DCI, which greatly differed from the markers we summarized. Efficient predictors should be further explored in subsequent studies.

Limitations of This Study

Our study has the following limitations. First, there may be some differences in the definition of DCI among studies, which may impose some limitations on the interpretation of our results to some extent. Second, the included studies contain a diversity of models. However, except for logistic regression, the other types of models are limited in number. Third, there is a high risk of bias in primary studies, mainly owing to the limited cases for modeling and the avoidance of overfitting. Fourth, the validation methods for the models in the primary studies are mainly random sampling or k-fold cross-validation, with rare external validation. Fifth, in a very small number of studies, validation sets can be used to try out different hyperparameters to improve the algorithm; test sets are essentially raw data. It is not clear in the original studies we included whether the models were optimized in this way. Therefore, we were not able to discuss this in greater depth in our review. Sixth, even though we conducted a systematic search and literature screening, no studies were found on image-based ML for predicting DCI in patients with SAH. Therefore, deeper exploration should be conducted in future studies.

Conclusion

The ML constructed in the original studies that we included was primarily based on the baseline characteristics of patients at admission. Our meta-analysis shows that ML had a relatively desirable predictive value. In subsequent studies, more efficient predictors can be explored and better-performing prediction tools can be developed based on the ML approach to enable treating physicians to enhance daily TCD monitoring. Prioritizing TCD monitoring in resource-limited settings lowers the threshold for performing more invasive diagnostic procedures such as catheter angiography.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82171458, 82171363, 82371381).

Data Availability

The original contributions presented in this study are included in this paper and the multimedia appendices. Further inquiries can be directed to the corresponding author.

Authors' Contributions

HZ and PL conceptualized this study, contributed to the methodology, and conducted formal analysis and investigation. HZ prepared the original draft and arranged all the resources. PL and XJ reviewed and edited this paper, supervised this study, and acquired the funds for this study. All authors commented on the previous versions of this manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy.
[DOCX File , 16 KB - jmir_v27i1e54121_app1.docx]

Multimedia Appendix 2
Basic study characteristics.
[DOCX File , 39 KB - jmir_v27i1e54121_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Modeling predictors.
[DOCX File , 17 KB - jmir_v27i1e54121_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 78 KB - jmir_v27i1e54121_app4.pdf]

References

1. Feigin VL, Lawes CMM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurol 2009 Apr;8(4):355-369. [doi: 10.1016/S1474-4422(09)70025-0] [Medline: 19233729]
2. Etminan N, Chang H, Hackenberg K, de Rooij NK, Vergouwen MDI, Rinkel GJE, et al. Worldwide incidence of aneurysmal subarachnoid hemorrhage according to region, time period, blood pressure, and smoking prevalence in the population: a systematic review and meta-analysis. JAMA Neurol 2019 May 01;76(5):588-597 [FREE Full text] [doi: 10.1001/jamaneurol.2019.0006] [Medline: 30659573]
3. Claassen J, Park S. Spontaneous subarachnoid haemorrhage. Lancet 2022 Sep 10;400(10355):846-862 [FREE Full text] [doi: 10.1016/S0140-6736(22)00938-2] [Medline: 35985353]
4. Spears WE, Greer DM, Nguyen TN. Comment on the 2023 guidelines for the management of patients with aneurysmal subarachnoid hemorrhage. Stroke 2023 Oct;54(10):2708-2712. [doi: 10.1161/STROKEAHA.123.043541] [Medline: 37581267]
5. Francoeur CL, Mayer SA. Management of delayed cerebral ischemia after subarachnoid hemorrhage. Crit Care 2016 Oct 14;20(1):277 [FREE Full text] [doi: 10.1186/s13054-016-1447-6] [Medline: 27737684]
6. Rigante L, van Lieshout JH, Vergouwen MDI, van Griensven CHS, Vart P, van der Loo L, et al. Time trends in the risk of delayed cerebral ischemia after subarachnoid hemorrhage: a meta-analysis of randomized controlled trials. Neurosurg Focus 2022 Mar;52(3):E2. [doi: 10.3171/2021.12.FOCUS21473] [Medline: 35231892]
7. Kumar G, Shahripour RB, Harrigan MR. Vasospasm on transcranial Doppler is predictive of delayed cerebral ischemia in aneurysmal subarachnoid hemorrhage: a systematic review and meta-analysis. J Neurosurg 2016 May;124(5):1257-1264. [doi: 10.3171/2015.4.JNS15428] [Medline: 26495942]
8. Han H, Chen Y, Li R, Lin F, Lu J, Chen X, et al. The value of early CT perfusion parameters for predicting delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage: a systematic review and meta-analysis. Neurosurg Rev 2022 Aug;45(4):2517-2531. [doi: 10.1007/s10143-022-01779-3] [Medline: 35377027]
9. Scherschinski L, Catapano JS, Karahalios K, Koester SW, Benner D, Winkler EA, et al. Electroencephalography for detection of vasospasm and delayed cerebral ischemia in aneurysmal subarachnoid hemorrhage: a retrospective analysis and systematic review. Neurosurg Focus 2022 Mar;52(3):E3. [doi: 10.3171/2021.12.FOCUS21656] [Medline: 35231893]
10. Beam AL, Drazen JM, Kohane IS, Leong T, Manrai AK, Rubin EJ. Artificial intelligence in medicine. N Engl J Med 2023 Mar 30;388(13):1220-1221. [doi: 10.1056/NEJMe2206291] [Medline: 36988598]
11. Yuan JY, Chen Y, Jayaraman K, Kumar A, Zlepper Z, Allen ML, et al. Automated quantification of compartmental blood volumes enables prediction of delayed cerebral ischemia and outcomes after aneurysmal subarachnoid hemorrhage. World Neurosurg 2023 Feb;170:e214-e222 [FREE Full text] [doi: 10.1016/j.wneu.2022.10.105] [Medline: 36323345]
12. Taghavi RM, Zhu G, Wintermark M, Kuraitis GM, Sussman ES, Pulli B, et al. Prediction of delayed cerebral ischemia after cerebral aneurysm rupture using explainable machine learning approach. Interv Neuroradiol 2023 Apr 17:15910199231170411. [doi: 10.1177/15910199231170411] [Medline: 37070145]
13. Palmisciano P, Hoz SS, Johnson MD, Forbes JA, Prestigiacomo CJ, Zuccarello M, et al. External validation of an extreme gradient boosting model for prediction of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. World Neurosurg 2023 Jul;175:e108-e114. [doi: 10.1016/j.wneu.2023.03.036] [Medline: 36914029]
14. Oliveira Souza NVD, Rouanet C, Solla DJF, de Lima CVB, de Souza CA, Rezende F, et al. The role of VASOGRADE as a simple grading scale to predict delayed cerebral ischemia and functional outcome after aneurysmal subarachnoid hemorrhage. Neurocrit Care 2023 Feb;38(1):96-104. [doi: 10.1007/s12028-022-01577-1] [Medline: 36002635]
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
16. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan 01;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]
17. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2019 Sep;28(9):2768-2786 [FREE Full text] [doi: 10.1177/0962280218785504] [Medline: 30032705]

XSL•FO
RenderX

18. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996 Feb 28;15(4):361-387. [doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4] [Medline: 8668867]
19. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138 [FREE Full text] [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
20. Chen X, Ou Y, Leng J, Wang C, Min F, Xia Y, et al. Construction and verification of a risk prediction model for the occurrence of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage requiring mechanical ventilation. Biomed Res Int 2023;2023:7656069 [FREE Full text] [doi: 10.1155/2023/7656069] [Medline: 36845638]
21. Zheng W, Kim JA, Elmer J, Zafar SF, Ghanta M, Moura Junior V, et al. Automated EEG-based prediction of delayed cerebral ischemia after subarachnoid hemorrhage. Clin Neurophysiol 2022 Nov;143:97-106 [FREE Full text] [doi: 10.1016/j.clinph.2022.08.023] [Medline: 36182752]
22. Zhang C, Feng L, You F, Zhao X, Fang X, Zhou Y. Whole-brain permeability analysis on admission improves prediction of delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage. J Stroke Cerebrovasc Dis 2022 Apr;31(4):106312. [doi: 10.1016/j.jstrokecerebrovasdis.2022.106312] [Medline: 35093628]
23. Wang L, Zhang Q, Zhang G, Zhang W, Chen W, Hou F, et al. Risk factors and predictive models of poor prognosis and delayed cerebral ischemia in aneurysmal subarachnoid hemorrhage complicated with hydrocephalus. Front Neurol 2022;13:1014501 [FREE Full text] [doi: 10.3389/fneur.2022.1014501] [Medline: 36353134]
24. Megjhani M, Weiss M, Kwon SB, Ford J, Nametz D, Kastenholz N, et al. Vector angle analysis of multimodal neuromonitoring data for continuous prediction of delayed cerebral ischemia. Neurocrit Care 2022 Aug;37(Suppl 2):230-236. [doi: 10.1007/s12028-022-01481-8] [Medline: 35352273]
25. Hu P, Liu Y, Li Y, Guo G, Su Z, Gao X, et al. A comparison of LASSO regression and tree-based models for delayed cerebral ischemia in elderly patients with subarachnoid hemorrhage. Front Neurol 2022;13:791547 [FREE Full text] [doi: 10.3389/fneur.2022.791547] [Medline: 35359648]
26. Hu P, Li Y, Liu Y, Guo G, Gao X, Su Z, et al. Comparison of conventional logistic regression and machine learning methods for predicting delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage: a multicentric observational cohort study. Front Aging Neurosci 2022;14:857521 [FREE Full text] [doi: 10.3389/fnagi.2022.857521] [Medline: 35783143]
27. Hernández-Hernández MA, Cherchi MS, Torres-Díez E, Orizaola P, Martín-Láez R, Fernández-Torre JL. Bispectral index monitoring to detect delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. J Crit Care 2022 Dec;72:154154. [doi: 10.1016/j.jcrc.2022.154154] [Medline: 36152563]
28. Dicpinigaitis AJ, Galea VP, Sursal T, Al-Shammari H, Feldstein E, Ali S, et al. Low serum albumin as a risk factor for delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage: eICU collaborative research database analysis. J Neurosurg Sci 2024 Jun;68(3):287-293. [doi: 10.23736/S0390-5616.22.05604-1] [Medline: 35766203]
29. Csók I, Grauvogel J, Scheiwe C, Bardutzky J, Wehrum T, Beck J, et al. Basic surveillance parameters improve the prediction of delayed cerebral infarction after aneurysmal subarachnoid hemorrhage. Front Neurol 2022;13:774720 [FREE Full text] [doi: 10.3389/fneur.2022.774720] [Medline: 35309593]
30. Chen K, Huang G, Cai C, Yan C, Zhang F, Yao M, et al. Higher cystatin C level increases the risk of delayed cerebral ischemia after endovascular treatment of aneurysmal subarachnoid hemorrhage: a case-control study. Neurosurg Rev 2022 Jun;45(3):2407-2415. [doi: 10.1007/s10143-022-01765-9] [Medline: 35247129]
31. Chen HY, Elmer J, Zafar SF, Ghanta M, Moura Junior V, Rosenthal ES, et al. Combining transcranial doppler and EEG data to predict delayed cerebral ischemia after subarachnoid hemorrhage. Neurology 2022 Feb 01;98(5):e459-e469 [FREE Full text] [doi: 10.1212/WNL.0000000000013126] [Medline: 34845057]
32. Bolton WS, Gharial PK, Akhunbay-Fudge C, Chumas P, Mathew RK, Anderson IA. Day 2 neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios for prediction of delayed cerebral ischemia in subarachnoid hemorrhage. Neurosurg Focus 2022 Mar;52(3):E4. [doi: 10.3171/2021.12.FOCUS21642] [Medline: 35231894]
33. Alexopoulos G, Zhang J, Karampelas I, Khan M, Quadri N, Patel M, et al. Applied forecasting for delayed cerebral ischemia prediction post subarachnoid hemorrhage: Methodological fallacies. Informatics in Medicine Unlocked 2022;28:100817 [FREE Full text] [doi: 10.1016/j.imu.2021.100817]
34. Odenstedt Hergès H, Vithal R, El-Merhi A, Naredi S, Staron M, Block L. Machine learning analysis of heart rate variability to detect delayed cerebral ischemia in subarachnoid hemorrhage. Acta Neurol Scand 2022 Feb;145(2):151-159. [doi: 10.1111/ane.13541] [Medline: 34677832]
35. Zhao L, Chen T, Yan H, Liu C, Cao Y, Zhang Y, et al. Development and validation of an early predictive nomogram for delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. Ann Transl Med 2021 Nov;9(22):1664 [FREE Full text] [doi: 10.21037/atm-21-5200] [Medline: 34988173]
36. Savarraj JPJ, Hergenroeder GW, Zhu L, Chang T, Park S, Megjhani M, et al. Machine learning to predict delayed cerebral ischemia and outcomes in subarachnoid hemorrhage. Neurology 2021 Jan 26;96(4):e553-e562 [FREE Full text] [doi: 10.1212/WNL.0000000000011211] [Medline: 33184232]

XSL•FO
RenderX

37. Megjhani M, Terilli K, Weiss M, Savarraj J, Chen LH, Alkhachroum A, et al. Dynamic detection of delayed cerebral ischemia: a study in 3 centers. Stroke 2021 Apr;52(4):1370-1379 [FREE Full text] [doi: 10.1161/STROKEAHA.120.032546] [Medline: 33596676]
38. Katsuki M, Kawamura S, Koh A. Easily created prediction model using automated artificial intelligence framework (Prediction One, Sony Network Communications Inc., Tokyo, Japan) for subarachnoid hemorrhage outcomes treated by coiling and delayed cerebral ischemia. Cureus 2021 Jun;13(6):e15695 [FREE Full text] [doi: 10.7759/cureus.15695] [Medline: 34277282]
39. Hu P, Xu Y, Liu Y, Li Y, Ye L, Zhang S, et al. An externally validated dynamic nomogram for predicting unfavorable prognosis in patients with aneurysmal subarachnoid hemorrhage. Front Neurol 2021;12:683051 [FREE Full text] [doi: 10.3389/fneur.2021.683051] [Medline: 34512505]
40. de Jong G, Aquarius R, Sanaan B, Bartels RHMA, Grotenhuis JA, Henssen DJHA, et al. Prediction models in aneurysmal subarachnoid hemorrhage: forecasting clinical outcome with artificial intelligence. Neurosurgery 2021 Apr 15;88(5):E427-E434. [doi: 10.1093/neuros/nyaa581] [Medline: 33548918]
41. Liu H, Xu Q, Li A. Nomogram for predicting delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage in the Chinese population. J Stroke Cerebrovasc Dis 2020 Sep;29(9):105005. [doi: 10.1016/j.jstrokecerebrovasdis.2020.105005] [Medline: 32807421]
42. Tanioka S, Ishida F, Nakano F, Kawakita F, Kanamaru H, Nakatsuka Y, pSEED group. Machine learning analysis of matricellular proteins and clinical variables for early prediction of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. Mol Neurobiol 2019 Oct;56(10):7128-7135. [doi: 10.1007/s12035-019-1601-7] [Medline: 30989629]
43. Fragata I, Alves M, Papoila AL, Nunes AP, Ferreira P, Diogo M, et al. Computed tomography perfusion as a predictor of delayed cerebral ischemia and functional outcome in spontaneous subarachnoid hemorrhage: A single center experience. Neuroradiol J 2019 Jun;32(3):179-188 [FREE Full text] [doi: 10.1177/1971400919829048] [Medline: 30777482]
44. Fang Y, Mei S, Lu J, Chen Y, Chai Z, Dong X, et al. New risk score of the early period after spontaneous subarachnoid hemorrhage: For the prediction of delayed cerebral ischemia. CNS Neurosci Ther 2019 Oct;25(10):1173-1181 [FREE Full text] [doi: 10.1111/cns.13202] [Medline: 31407513]
45. Bøthun ML, Haaland, Moen G, Logallo N, Svendsen F, Thomassen L, et al. Impaired cerebrovascular reactivity may predict delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. J Neurol Sci 2019 Dec 15;407:116539 [FREE Full text] [doi: 10.1016/j.jns.2019.116539] [Medline: 31669725]
46. Al-Mufti F, Amuluru K, Damodara N, Dodson V, Roh D, Agarwal S, et al. Admission neutrophil-lymphocyte ratio predicts delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage. J Neurointerv Surg 2019 Nov;11(11):1135-1140. [doi: 10.1136/neurintsurg-2019-014759] [Medline: 30979846]
47. Al-Mufti F, Misiolek KA, Roh D, Alawi A, Bauerschmidt A, Park S, et al. White blood cell count improves prediction of delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage. Neurosurgery 2019 Feb 01;84(2):397-403. [doi: 10.1093/neuros/nyy045] [Medline: 29528448]
48. Ramos LA, van der Steen WE, Sales Barros R, Majoie CBLM, van den Berg R, Verbaan D, et al. Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. J Neurointerv Surg 2019 May;11(5):497-502. [doi: 10.1136/neurintsurg-2018-014258] [Medline: 30415227]
49. Park S, Megjhani M, Frey H, Grave E, Wiggins C, Terilli KL, et al. Predicting delayed cerebral ischemia after subarachnoid hemorrhage using physiological time series data. J Clin Monit Comput 2019 Feb;33(1):95-105 [FREE Full text] [doi: 10.1007/s10877-018-0132-5] [Medline: 29556884]
50. Rosenthal ES, Biswal S, Zafar SF, O'Connor KL, Bechek S, Shenoy AV, et al. Continuous electroencephalography predicts delayed cerebral ischemia after subarachnoid hemorrhage: A prospective study of diagnostic accuracy. Ann Neurol 2018 May;83(5):958-969 [FREE Full text] [doi: 10.1002/ana.25232] [Medline: 29659050]
51. Megjhani M, Terilli K, Frey H, Velazquez AG, Doyle KW, Connolly ES, et al. Incorporating high-frequency physiologic data using computational dictionary learning improves prediction of delayed cerebral ischemia compared to existing methods. Front Neurol 2018;9:122 [FREE Full text] [doi: 10.3389/fneur.2018.00122] [Medline: 29563892]
52. Gölitz P, Hoelter P, Rösch J, Roessler K, Knossalla F, Doerfler A. Ultra-early detection of microcirculatory injury as predictor of developing delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. Clin Neuroradiol 2018 Dec;28(4):501-507. [doi: 10.1007/s00062-017-0616-6] [Medline: 28812108]
53. Lu G, Wong MS, Xiong MZQ, Leung CK, Su XW, Zhou JY, et al. Circulating micro RNAs in delayed cerebral infarction after aneurysmal subarachnoid hemorrhage. J Am Heart Assoc 2017 Apr 25;6(4):e005363 [FREE Full text] [doi: 10.1161/JAHA.116.005363] [Medline: 28442458]
54. Fragata I, Alves M, Papoila AL, Nunes AP, Ferreira P, Canto-Moreira N, et al. Early prediction of delayed ischemia and functional outcome in acute subarachnoid hemorrhage: role of diffusion tensor imaging. Stroke 2017 Aug;48(8):2091-2097. [doi: 10.1161/STROKEAHA.117.016811] [Medline: 28667021]
55. Foreman PM, Chua MH, Harrigan MR, Fisher WS, Tubbs RS, Shoja MM, et al. External validation of the Practical Risk Chart for the prediction of delayed cerebral ischemia following aneurysmal subarachnoid hemorrhage. J Neurosurg 2017 May;126(5):1530-1536. [doi: 10.3171/2016.1.JNS152554] [Medline: 27177181]

XSL•FO
RenderX

56. Da Silva IRF, Gomes JA, Wachsman A, de Freitas GR, Provencio JJ. Hematologic counts as predictors of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. J Crit Care 2017 Feb;37:126-129 [FREE Full text] [doi: 10.1016/j.jcrc.2016.09.011] [Medline: 27718411]
57. Al-Mufti F, Roh D, Lahiri S, Meyers E, Witsch J, Frey H, et al. Ultra-early angiographic vasospasm associated with delayed cerebral ischemia and infarction following aneurysmal subarachnoid hemorrhage. J Neurosurg 2017 May;126(5):1545-1551. [doi: 10.3171/2016.2.JNS151939] [Medline: 27231975]
58. Zijlstra IA, Gathier CS, Boers AM, Marquering HA, Slooter AJ, Velthuis BK, et al. Association of automatically quantified total blood volume after aneurysmal subarachnoid hemorrhage with delayed cerebral ischemia. AJNR Am J Neuroradiol 2016 Sep;37(9):1588-1593 [FREE Full text] [doi: 10.3174/ajnr.A4771] [Medline: 27102313]
59. Zhang L, Qi S. Electrocardiographic abnormalities predict adverse clinical outcomes in patients with subarachnoid hemorrhage. J Stroke Cerebrovasc Dis 2016 Nov;25(11):2653-2659. [doi: 10.1016/j.jstrokecerebrovasdis.2016.07.011] [Medline: 27476337]
60. Budohoski K, Czosnyka M, Smielewski P, Varsos G, Kasprowicz M, Brady K, et al. Monitoring cerebral autoregulation after subarachnoid hemorrhage. Acta Neurochir Suppl 2016;122:199-203. [doi: 10.1007/978-3-319-22533-3_40] [Medline: 27165906]
61. de Oliveira Manoel AL, Jaja BN, Germans MR, Yan H, Qian W, Kouzmina E, SAHIT collaborators. The VASOGRADE: A simple grading scale for prediction of delayed cerebral ischemia after subarachnoid hemorrhage. Stroke 2015 Jul;46(7):1826-1831. [doi: 10.1161/STROKEAHA.115.008728] [Medline: 25977276]
62. de Rooij NK, Greving JP, Rinkel GJE, Frijns CJM. Early prediction of delayed cerebral ischemia after subarachnoid hemorrhage: development and validation of a practical risk chart. Stroke 2013 May;44(5):1288-1294. [doi: 10.1161/STROKEAHA.113.001125] [Medline: 23512975]
63. Baumann A, Devaux Y, Audibert G, Zhang L, Bracard S, Colnat-Coulbois S, et al. Gene expression profile of blood cells for the prediction of delayed cerebral ischemia after intracranial aneurysm rupture: a pilot study in humans. Cerebrovasc Dis 2013;36(3):236-242. [doi: 10.1159/000354161] [Medline: 24135536]
64. Kumar G, Dumitrascu OM, Chiang C, O'Carroll CB, Alexandrov AV. Prediction of delayed cerebral ischemia with cerebral angiography: a meta-analysis. Neurocrit Care 2019 Feb;30(1):62-71. [doi: 10.1007/s12028-018-0572-2] [Medline: 29998428]
65. Santana LS, Diniz JBC, Rabelo NN, Teixeira MJ, Figueiredo EG, Telles JPM. Machine learning algorithms to predict delayed cerebral ischemia after subarachnoid hemorrhage: a systematic review and meta-analysis. Neurocrit Care 2023 Sep 05:1171-1181. [doi: 10.1007/s12028-023-01832-z] [Medline: 37667079]
66. Reddy S. Explainability and artificial intelligence in medicine. Lancet Digit Health 2022 Apr;4(4):e214-e215 [FREE Full text] [doi: 10.1016/S2589-7500(22)00029-2] [Medline: 35337639]
67. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9(4):e1312 [FREE Full text] [doi: 10.1002/widm.1312] [Medline: 32089788]
68. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med 2023 May 25;388(21):1981-1990. [doi: 10.1056/NEJMra2301725] [Medline: 37224199]
69. Jabbarli R, Pierscianek D, Darkwah Oppong M, Sato T, Dammann P, Wrede KH, et al. Laboratory biomarkers of delayed cerebral ischemia after subarachnoid hemorrhage: a systematic review. Neurosurg Rev 2020 Jun;43(3):825-833. [doi: 10.1007/s10143-018-1037-y] [Medline: 30306357]

Abbreviations

DCI: delayed cerebral ischemia
LASSO: least absolute shrinkage and selection operator
MeSH: Medical Subject Headings
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
SAH: subarachnoid hemorrhage
TCD: transcranial Doppler
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Edited by A Coristine; submitted 31.10.23; peer-reviewed by MO Khursheed, L He, S Mahmud, Y Kim, JJ van der Harst; comments to author 21.08.24; revised version received 14.10.24; accepted 26.11.24; published 20.01.25.

Please cite as:
Zhang H, Zou P, Luo P, Jiang X
Machine Learning for the Early Prediction of Delayed Cerebral Ischemia in Patients With Subarachnoid Hemorrhage: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e54121
URL: https://www.jmir.org/2025/1/e54121
doi:10.2196/54121
PMID:

©Haofuzi Zhang, Peng Zou, Peng Luo, Xiaofan Jiang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Combining a Risk Factor Score Designed From Electronic Health Records With a Digital Cytology Image Scoring System to Improve Bladder Cancer Detection: Proof-of-Concept Study

Sandie Cabon[1], PhD; Sarra Brihi[2], MS; Riadh Fezzani[2], PhD; Morgane Pierre-Jean[1], PhD; Marc Cuggia[1], MD, PhD; Guillaume Bouzillé[1], MD, PhD

[1]Univ Rennes, CHU Rennes, INSERM, LTSI - UMR 1099, F-35000 Rennes, France

[2]R&D, VitaDX International, Paris, France

Corresponding Author:
Sandie Cabon, PhD
Univ Rennes, CHU Rennes, INSERM, LTSI - UMR 1099
Campus de Beaulieu
Bâtiment 22, 263 Avenue du General Leclerc
F-35000 Rennes
France
Phone: 33 2 23 23 62 20
Email: sandie.cabon@univ-rennes.fr

Abstract

Background: To reduce the mortality related to bladder cancer, efforts need to be concentrated on early detection of the disease for more effective therapeutic intervention. Strong risk factors (eg, smoking status, age, professional exposure) have been identified, and some diagnostic tools (eg, by way of cystoscopy) have been proposed. However, to date, no fully satisfactory (noninvasive, inexpensive, high-performance) solution for widespread deployment has been proposed. Some new models based on cytology image classification were recently developed and bring good perspectives, but there are still avenues to explore to improve their performance.

Objective: Our team aimed to evaluate the benefit of combining the reuse of massive clinical data to build a risk factor model and a digital cytology image–based model (VisioCyt) for bladder cancer detection.

Methods: The first step relied on designing a predictive model based on clinical data (ie, risk factors identified in the literature) extracted from the clinical data warehouse of the Rennes Hospital and machine learning algorithms (logistic regression, random forest, and support vector machine). It provides a score corresponding to the risk of developing bladder cancer based on the patient's clinical profile. Second, we investigated 3 strategies (logistic regression, decision tree, and a custom strategy based on score interpretation) to combine the model's score with the score from an image-based model to produce a robust bladder cancer scoring system.

Results: We collected 2 data sets. The first set, including clinical data for 5422 patients extracted from the clinical data warehouse, was used to design the risk factor–based model. The second set was used to measure the models' performances and was composed of data for 620 patients from a clinical trial for which cytology images and clinicobiological features were collected. With this second data set, the combination of both models obtained areas under the curve of 0.82 on the training set and 0.83 on the test set, demonstrating the value of combining risk factor–based and image-based models. This combination offers a higher associated risk of cancer than VisioCyt alone for all classes, especially for low-grade bladder cancer.

Conclusions: These results demonstrate the value of combining clinical and biological information, especially to improve detection of low-grade bladder cancer. Some improvements will need to be made to the automatic extraction of clinical features to make the risk factor–based model more robust. However, as of now, the results support the assumption that this type of approach will be of benefit to patients.

(J Med Internet Res 2025;27:e56946) doi:10.2196/56946

XSL•FO
RenderX

KEYWORDS

bladder cancer; clinical data reuse; multimodal data fusion; clinical decision support; machine learning; risk factors; electronic health records; detection; mortality; therapeutic intervention; diagnostic tools; digital cytology; image-based model; clinical data; algorithms; patient; biological information

Introduction

Bladder cancer ranks as the 9th most prevalent cancer worldwide, and it is the 13th leading cause of cancer-related deaths, claiming over 13,000 lives annually [1,2]. In 2017, bladder cancer was the 5th most frequently diagnosed cancer in men, with an incidence 4 times higher than that in women [3]. Detecting bladder cancer at an early stage significantly enhances the chance of successful treatment, while late-stage detection presents more challenges. Current treatments involve surgical procedures and auxiliary chemotherapy but come with a relatively poor 5-year prognosis [4]. More precisely, the 5-year survival is 69.5% for localized disease. The survival rate decreases to 36.3% for regional disease and only 5% for metastatic bladder cancer [5]. Today, periodic cystoscopy associated with urine cytology is the gold standard for the diagnosis of bladder cancer or to monitor for recurrence [6]. However, cystoscopy is an expensive and invasive procedure usually associated with discomfort. Urine cytology is a test to detect abnormal cells in urine samples. Although urine cytology is the most widely used noninvasive test, it has poor sensitivity, especially for low-grade tumors [7,8].

Thus, to reduce the mortality associated with bladder cancer, new diagnostic tools need to be proposed for early detection of the disease [3]. At the same time, newly proposed methods should be noninvasive and inexpensive to allow large-scale dissemination. Research is currently focusing on this objective through better identification of risk factors or development of more effective diagnostic tools.

Recently, new diagnostic tools based on the detection of specific biomarkers in urine tests were proposed [9-11]. ADXBLADDER is a urine test to detect the mini chromosome maintenance 5 protein in voided urine [9]. It has a sensitivity of 83% and specificity of 77%. Urodiag is based on genetic and epigenetic analysis using multiplex quantitative real-time polymerase chain reaction and consists of detecting 4 mutations of the fibroblast growth factor receptor 3 gene [10]. It has a sensitivity of 95% and specificity of 76%. URO17 is an immunocytochemical test designed to evaluate keratine 17 expression and has a sensitivity of 100% and specificity of 96% [11]. These tests show promising performance but have only been studied in small cohorts (from 26 to 71 patients).

Increasing interest in deep learning algorithms has also led to the emergence of new diagnostic strategies based on image processing (see [12,13] for reviews). Although most of them have been developed from cystoscopy images, some aim to propose noninvasive techniques and have exploited images obtained from urine cytology [14-16]. The authors report good to very good performance, including a sensitivity of 79.5% and specificity of 84.5% [14]; area under the receiver operating characteristic (ROC) curve (AUC) of 0.83 [15]; and sensitivity of 81.0% and specificity of 98.0% [16]. However, studies were conducted either with populations composed mainly of patients with high-grade cancer [14] or with a small amount of data and a limited number of patients (ie, only 398 slides in [15] and 68 patients in [16]). Moreover, a lack of generalization power, as expressed by a decrease in performance when applying the model to new data and especially new data sources, hinders deployment in clinics [13]. In addition, the main challenge is detecting patients with low-grade cancer to improve their chance of survival.

On the other hand, several risk factors were identified as strong determinants of bladder cancer. First, demographic information, such as age and sex, was identified as a strong factor [5]. In 2012, Letašiová et al [17] also highlighted that smoking; exposure to arsenic (in water, air, or food); and exposure to aromatic amines and 4,4'-methylenebis in chemicals that are typically contained in dyes, hair dyes, paints, fungicides, cigarette smoke, plastics, metals, and motor vehicle exhaust are the most notable environmental risk factors [17]. An association with diabetes has also been described [18]. In addition, positive associations between bladder cancer and a family history in first- and second-degree relatives [19] and hypertension [20] were found. In a more general context, having experienced stroke [21] or heart failure [22] was also associated with a higher risk of developing cancer.

The widespread adoption of electronic health records now provides an opportunity to capture granular patient information in large populations. The data can be used to develop machine learning models that could be incorporated in decision support systems. By incorporating risk factors, such models have the potential to identify at-risk patients more accurately, enhancing patient profiles for image-based analysis.

In this paper, we proposed combining knowledge about risk factors and cytology processing software to investigate the benefits that this could provide for bladder cancer detection. To do this, our team worked with a model developed by the company VitaDX [23]. This image-based model, called VisioCyt, integrates image cell processing and deep learning algorithms to predict bladder cancer from voided urinary cytology. In this study, our focus was on building a detection model based on risk factors extracted from our local clinical data warehouse (CDW). We hypothesized that a model trained on a large data set, offering a probability of bladder cancer based on a patient's clinical profile, can complement the VisioCyt scoring system and enhance detection, especially for low-grade cases.

Methods

Ethical Considerations

Our study followed the relevant guidelines and regulations, in accordance with the Declaration of Helsinki. The extracted data did not contain any nominative data, and all information was

XSL•FO
RenderX

deidentified in accordance with the protocol established for the improved project. Agreement from the French Commission Nationale de l'Informatique et des Libertés was obtained (agreement number 2206739).

Study Design

The proposed workflow for bladder cancer prediction is depicted in Figure 1. As a first step, 2 distinct prediction models were built: one based on risk factors (risk factor–based model [RFM]) and the other using cytology images (VisioCyt). The scores given by each model were then combined to provide a composite score.

Figure 1. Overview of the designed workflow for the scoring of bladder cancer using risk factors and cytology images, including a description of the (A) data sets and (B) data processing.

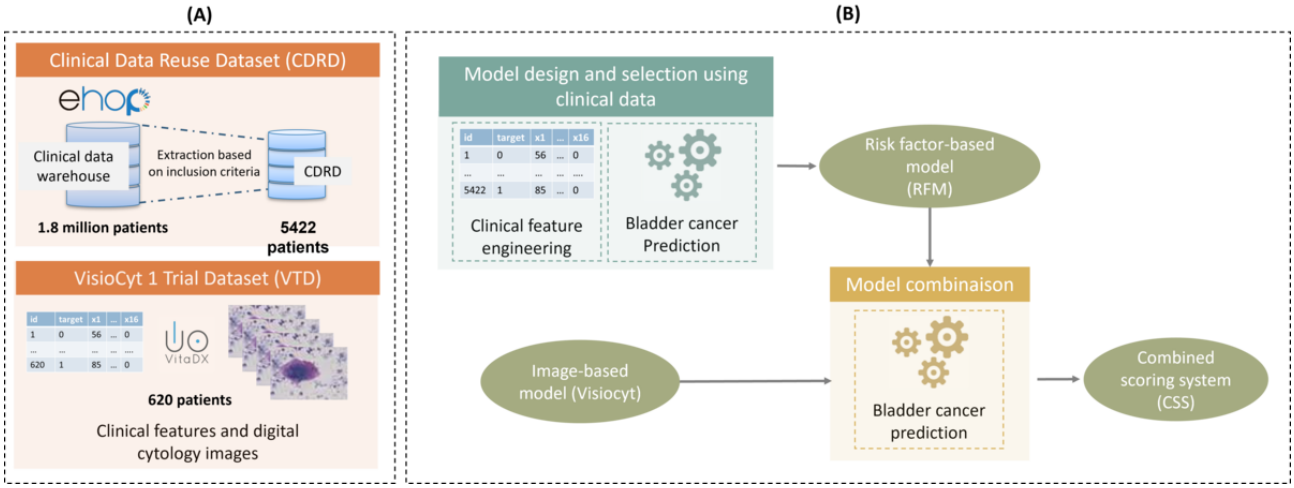

To set up and evaluate our approach, 2 data sets were considered: Clinical Data Reuse Dataset (CDRD) and VisioCyt 1 Trial Dataset (VTD).

CDRD was extracted according to the inclusion criteria from eHOP [24], the CDW developed and deployed at Rennes University Hospital. This database was used to design and select the bladder cancer prediction model based on risk factors (ie, the RFM). For that purpose, clinical features were retrieved and preprocessed to be used as inputs for the machine learning approaches. In the end, several models were compared to identify the best model.

The second data set contained clinical data and cytology images that were acquired during the VisioCyt 1 Trial (NCT02966691). Images from this data set were first used to set up the bladder cancer prediction model based on cytology images and deep learning (ie, VisioCyt). This model was internally developed by VitaDX. In this study, we focused on the combination of this model with an RFM. To evaluate and adopt the best strategy for the combination, the whole data set (ie, images and clinical data from VTD) was used.

In the following sections, each stage of this workflow is described. First, we introduce the study population with the associated inclusion criteria. It is followed by a description of the data sets. Methods to set up the risk factor–based predictor are presented, and a description of the Visiocyt model is given. Finally, the combination strategies to build the final bladder cancer predictor are described.

Study Population

The eligibility criteria were defined in the VisioCyt 1 Trial [23] and used to extract data from the CDW.

The inclusion criteria were the following: older than 18 years, affiliation with a social security system, negative urine culture, programmed bladder endoscopy for a suspicion of bladder cancer (de novo, monitoring, or relapse) or exploration of the lower urinary tract excluding a suspected bladder cancer or prostate cancer.

The exclusion criteria were as follows: ongoing, untreated urinary tract infection; presence of bladder cancer excluding urothelial carcinoma; carcinoma of the high urinary tract, associated; history of lithiasis pathology; had undergone a renal transplantation.

Data Sets

Clinical Data Reuse Data Set

With eHOP, administrative and clinical data from electronic health records are collected, including unstructured (eg, clinical notes) and structured (eg, drugs, laboratory results) data. Data are deidentified, and a unique anonymous identifier allows the linkage among hospital stays of a given patient. The eHOP CDW currently allows users to search 80 million unstructured data elements and 430 million structured elements. All these data cover more than 1.8 million patients. CDRD was extracted from the eHOP database. All data on hospital stays at Rennes University Hospital between 2015 and 2017 were considered.

In this study, we considered people who had undergone endoscopy for vesical tumor resection or had a histology examination that was positive for bladder cancer. People considered negative for bladder cancer were those who had undergone a bladder endoscopy and had a negative histological examination.

XSL•FO
RenderX

For this extraction, we used the following data: demographic data; drugs administered (codification used by the French hospital information system to identify a drug and date of administration); Programme de Médicalisation des Systèmes d'Information (PMSI) data (French diagnosis-related groups) including International Statistical Classification of Diseases, Tenth Revision (ICD-10) codes, procedure codes, mortality, and length of stay; laboratory results described with local terminology; medical reports.

The extraction resulted in 5422 eligible patients: 1351 bladder cancer positives and 4071 bladder cancer negatives.

VisioCyt 1 Trial Data Set

The VisioCyt 1 Trial is registered in clinicaltrials.gov (NCT02966691). This clinical study was designed to evaluate the diagnostic performance of the VisioCyt test, which would improve the early diagnosis of bladder cancer. The diagnostic method used by the VisioCyt device is based on the analysis by transmission of urinary cytology slides prepared according to the protocol outlined for VisioCyt [23].

In addition to the inclusion criteria presented in the Study Population section, other criteria specific to that trial were considered. Additional inclusion criteria were the ability of the patient to understand the protocol and signature of a consent form before beginning the study, while people deprived of liberty or under guardianship were excluded. In addition, patients included in the same hospital as that from which CDRD was extracted were also excluded to avoid double inclusion.

In VTD, patients were considered to have bladder cancer if they had positive urine cytology and positive histology. This resulted in 620 patients, that is, 409 bladder cancer positives and 211 bladder cancer negatives. Among the positive patients, 220 patients with low-grade cancer and 189 patients with high-grade cancer were included. We collected 2 types of data. First, VisioCyt 1 provides urinary cytology digital slide images at a size of 50,000 × 50,000 pixels. They were acquired over 3 focal plans and contain cells of different types (eg, malpighian cells, urothelial cells, neutrophilic polynucleosis, artifacts, blood cells). For each VTD patient, the following demographic variables were retrieved: age, sex, weight (kg), and height (cm). The following risk factors and comorbidities were also collected: smoking status (never, former, or current smoker), diabetes, history of heart failure, hypertension, history of stroke, chronic obstructive pulmonary disease (COPD), family history of cancer, and professional exposure to a carcinogenic factor.

Risk Factor–Based Model (RFM)

Clinical Feature Extraction and Preprocessing

For each CDRD patient, the same features as those collected for VTD patients were retrieved. They were extracted from the data warehouse using structured and unstructured data. For structured data, standardized codes of several terminologies and local codes were used. For unstructured data (ie, medical reports), regular expressions and some more advanced techniques based on embeddings (ie, using spacy modeling from python) were implemented or used [25].

Missing data imputation was performed for weight and height by using k-nearest neighbor identification (k=10) based on age, sex, and weight or height according to the feature to be imputed. Standard scaling was first applied to ensure an equivalent impact on the distance measurement for numerical variables. For binary variables, we assumed that patients did not have the feature if it was not retrieved through the queries used on the data warehouse. A BMI feature was added and computed as weight divided by height squared.

A missing data indicator feature was added for each of the original features except for age and sex, which were complete. In the end, the clinical feature set had 29 clinical features.

Bladder Cancer Prediction From Clinical Data

We investigated 3 supervised classification algorithms, namely random forest (RF), logistic regression (LR), and support vector machine (SVM), to identify patients likely to have bladder cancer based on their clinical data.

A 2-step strategy, depicted in Figure 2, was designed to identify the best modeling approach.

First, the most optimal hyperparameters were identified using 3-fold cross-validation and a random grid search on 80% of CDRD (100 iterations). The best set of hyperparameters was defined as the one leading to the highest averaged AUC on the validation sets. During training, class imbalance was handled using class-weighting to mitigate the risk of bias toward the majority class in splits (ie, no bladder cancer patients). This process was repeated for each machine learning algorithm. This resulted in 3 candidate models composed of a standardizer, missing data imputer, and classifier ready to be applied to new data.

Second, model candidates were applied on 20% of CDRD to verify their generalization power. From there, the best classification approach was identified by comparing the performance using AUC, sensitivity, and specificity. At the end of this step, training using all CDRD data, the best hyperparameter set, and the best approach was conducted and resulted in the RFM.

XSL•FO
RenderX

Figure 2. Identification of the best classification model using 2 steps: (A) The first was training, a selection of the best set of hyperparameters and fit of a model candidate to use for each classification approach, and (B) the second was testing, identification of the best model among the 3 candidates. AUC: area under the receiver operating characteristic curve; CDRD: Clinical Data Reuse Dataset; LR: logistic regression; RF: random forest; SVM: support vector machine.

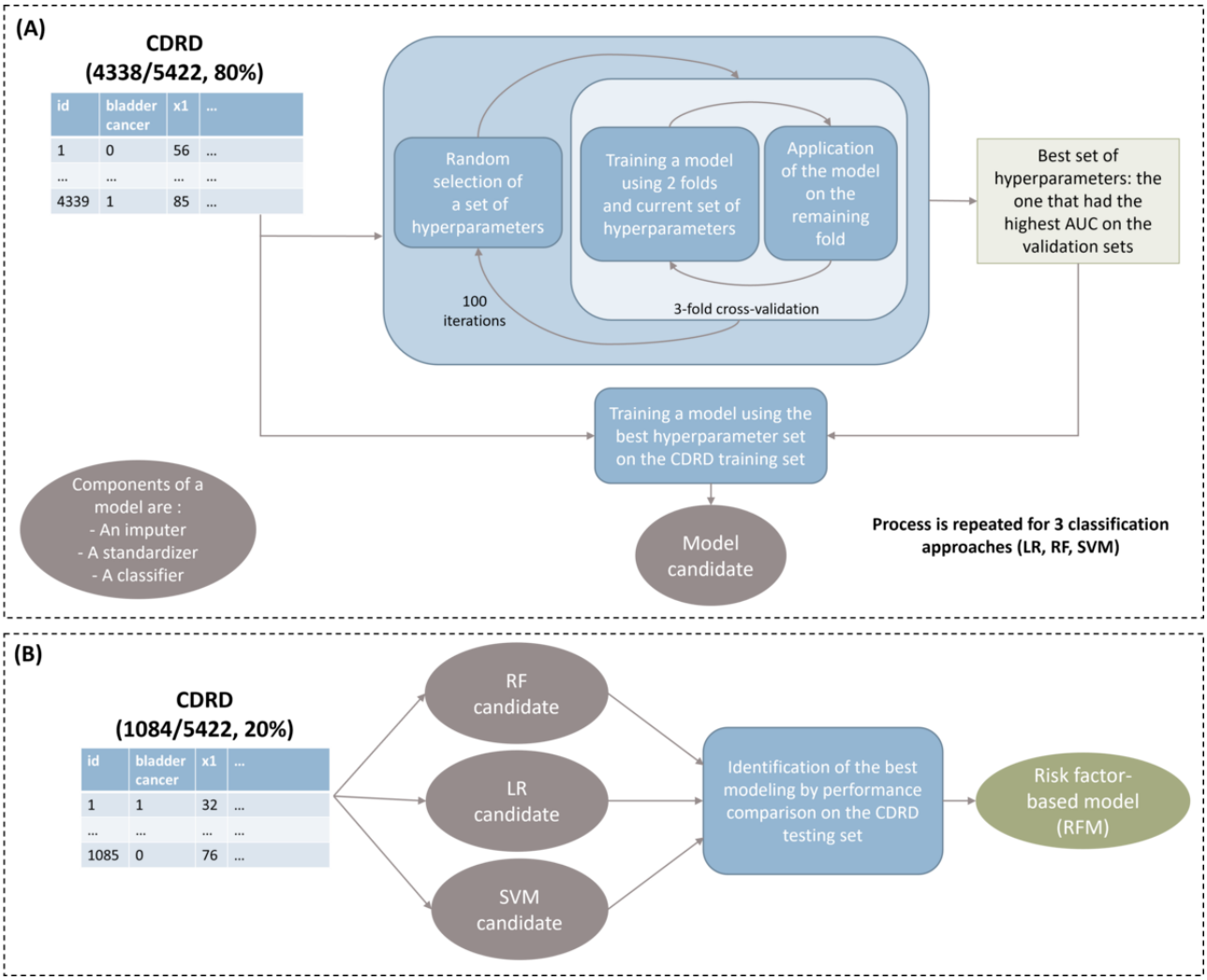

Image-Based Model

The image-based model works in 5 main steps going from checking the quality of the data to prediction of bladder cancer from segmented cell images. The process is depicted in Figure 3.

The process begins with a quality check of the slide, discarding any that are not properly prepared nor digitized. Following this, the biological elements on the slide are detected and classified, retaining only individual urothelial cells and clusters of these cells. A second quality assessment is then carried out to confirm the presence of a sufficient number of urothelial cells.

These cells are subsequently segmented to identify the nucleus and cytoplasm, and the characteristics of the nucleus are calculated based on established criteria and atypia parameters [23]. Urothelial clusters are classified using a detector for atypical urothelial clusters, and the classification score is used as a feature of the cluster.

Features of both individual urothelial cells and urothelial clusters are combined into a vector representing the features of the slide. This vector is ultimately classified by an RF classifier to determine the VisioCyt slide's score.

XSL•FO
RenderX

Figure 3. Workflow of the image-based model.

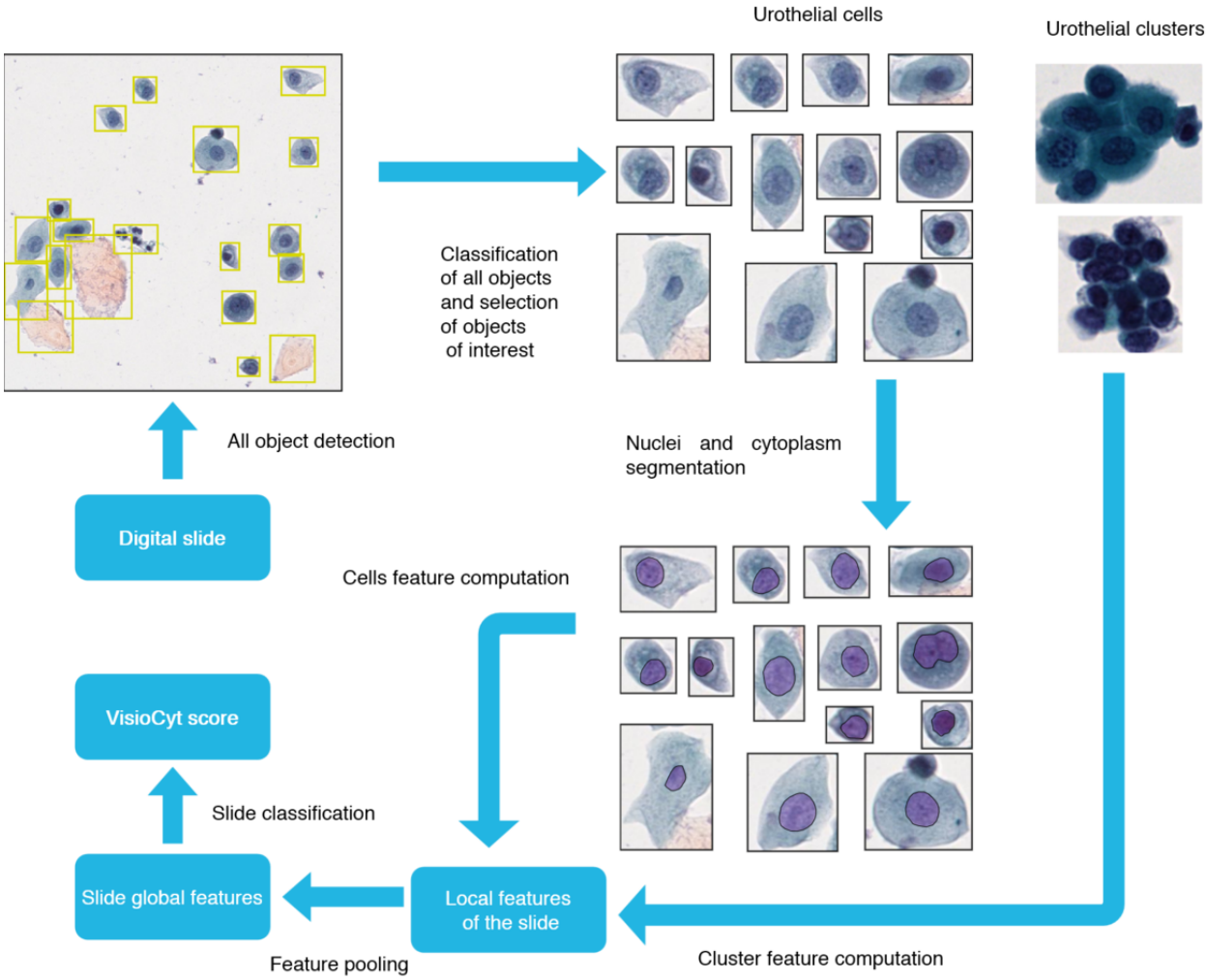

Combination of Image-Based and Risk Factor–Based Models

At this stage, prediction scores obtained from the RFM and VisioCyt are available. To combine these 2 scores and provide the final combined scoring system, 3 strategies were investigated using the output probabilities of both models as inputs: custom, LR, and decision tree (DT).

Experiments and evaluations were performed using the VTD. At first, stratified random selection, based on grade, was performed to allocate 520 patients to a training set and 131 patients to a test set. The most optimal hyperparameters of LR and DT were then identified using 3-fold cross-validation and a random grid search on the training set as described in the Bladder Cancer Prediction From Clinical Data section. The best models were then applied to the test set. For the custom method, the behaviors of the RFM and VisioCyt scores to build a decision rule adapted to our data were examined on the training set.

VisioCyt and RFM were not trained but simply applied on the training and test sets to compare their performance with those of the combination approaches.

Statistical Analysis and Evaluation Metrics

The relationships between the clinical variables and target variable in both CDRD and VTD were investigated. First, the correlations between bladder cancer status (ie, positive or negative) and continuous variables were studied using the point-biserial correlation test. According to Vogt et al [26], only features with point biserial values >0.25 were considered significantly correlated with the target variable. Chi-square tests were conducted for categorical features. This test consists of a null hypothesis that supposes the independence between the binary feature and bladder cancer status. Using Bonferroni correction, we considered that a $P<.003$ indicated a rejection of the null hypothesis and therefore implied a correlation between the binary variable and bladder cancer status.

ROC curve and AUC analyses were carried out to evaluate the performance of models. In addition, commonly used performance metrics such as balanced accuracy, specificity, and sensitivity were assessed for the RFM. They were measured using the optimal threshold given by the ROC as the combination maximizing the difference between the true positive and false positive rates. To better understand how the models made the predictions, the importance of the features on the predictions was investigated. Although intrinsically present in RF and DT models [27], this information is not directly available

XSL•FO
RenderX

for other approaches such as SVMs. In such cases, the permutation feature importance was computed [27].

Software and Platforms

All experiments were implemented in Python, version 3.8.8. Statistical analyses were performed using scipy 1.7.3. Scikit-learn 1.0.2 was used to train and evaluate the classification models.

Results

The results are presented in 3 sections. First, examination of the clinical features was undertaken to assess their univariate predictive capability for bladder cancer and to evaluate their consistency or heterogeneity across the 2 studied data sets. Second, the performances achieved during the design of the RFM are outlined. Last, the results obtained from the combination of the 2 scoring systems are discussed.

Clinical Feature Analysis

First, the distribution of the clinical features can be discussed by comparing either populations (ie, negatives vs positives) or data sets (ie, CDRD vs VTD). The statistics were grouped within these objectives and are reported in Table 1.

Table 1. Distribution of variables across positive and negative bladder cancer status for each data set.

Clinical features	Positive status		Negative status	
	CDRD[a] (n=1351)	VTB[b] (n=409)	CDRD[a] (n=4071)	VTB[b] (n=211)
Continuous variables, mean (SD)				
Age (years)	69.8 (12.9)	70.7 (10.0)	58.8 (17.7)	63.9 (13.1)
Weight (kg)	74.6 (13.8)	76.1 (14.9)	73.4 (33.1)	74.6 (16.1)
Height (cm)	168.2 (7.5)	170.0 (7.1)	167.2 (7.1)	168.0 (10.0)
BMI (kg/m^2)	26.4 (4.5)	26.0 (4.2)	26.3 (11.6)	26.2 (4.7)
Binary variables, n (%)				
Sex (male)	1069 (79.1)	321 (78.5)	2189 (53.8)	118 (55.9)
Smoking status				
Never smoked	482 (48.8)	83 (20.5)	641 (17.7)	114 (54.8)
Active smoker	109 (37.2)	91 (22.5)	384 (25)	27 (13)
Former smoker	301 (15.4)	231 (57)	201 (6.9)	67 (32.2)
Diabetes	243 (21.9)	86 (21)	389 (19.3)	19 (9)
Heart failure	150 (35.1)	91 (22.2)	380 (30.2)	9 (4.3)
Hypertension	311 (69.1)	210 (51.3)	830 (52.1)	55 (26.1)
Stroke	87 (16.3)	22 (5.4)	261 (14.7)	1 (0.5)
COPD[c]	95 (17.9)	83 (20.3)	140 (10.1)	18 (8.5)
Professional exposure	40 (2.7)	101 (24.7)	80 (1.3)	23 (10.9)
Family history of cancer	13 (0.9)	67 (16.4)	35 (0.8)	40 (19)

[a]CDRD: Clinical Data Reuse Dataset.

[b]VTD: VisioCyt 1 Trial Dataset.

[c]COPD: chronic obstructive pulmonary disease.

The distributions of continuous features (age, weight, height, and BMI) were equivalent in both data sets and between negatives and positives. For binary features, 3 scenarios emerged: Some display equivalent percentages and trends (ie, sex and COPD), while others only shared similar trends (active smoker, former smoker, heart failure, and hypertension) and others exhibited differences in both percentages and trends (never smoked, diabetes, stroke, professional exposure, and family history of cancer).

For sex and COPD, percentages and trends were similar in CDRD and VTD. There were more men with bladder cancer than without (with cancer: 1069/1351, 79.1% in CDRD and 321/409, 78.5% in VTD; without cancer: 2189/4071, 53.8% in CDRD and 118/211, 55.9% in VTD), and there was a higher occurrence of COPD among individuals with bladder cancer (95/1351, 17.9% in CDRD and 83/409, 20.3% in VTD) than in individuals without bladder cancer (140/4071, 3.4% in CDRD and 18/439, 8.5% in VTD).

For features with only similar trends, such as being an active smoker, heart failure, and hypertension, the percentages were generally higher in CDRD than in VTD, except for being a former smoker, where lower percentages were observed in CDRD than in VTD (Table 1).

XSL•FO
RenderX

For features considered to have different percentages and trends (never smoked, diabetes, stroke, professional exposure, and family history of cancer), the key observation was the equivalence between positives and negatives in CDRD, while a more pronounced difference existed between positives and negatives in VTD.

Statistical analyses were performed to observe relationships between features and bladder cancer status in each data set. The results are reported in Table 2.

Table 2. Statistical analyses performed to evaluate the correlations between the clinical features and bladder cancer status for each data set.

Clinical features	Clinical Data Reuse Dataset (CDRD)		VisioCyt 1 Trial Dataset (VTD)	
	Statistic	*P* value	Statistic	*P* value
Continuous variables, r_{pb}[a]				
Age	0.276	<.001	0.276	<.001
Weight	0.018	.18	0.051	<.001
Height	0.059	<.001	0.133	<.001
BMI	0.004	.74	–0.025	.54
Binary variables, χ^2 (*df*)[b]				
Sex	108.534 (1)	<.001	10.005 (1)	<.001
Smoking status				
Never smoked	0.659 (1)	.42	49.854 (1)	<.001
Active smoker	53.680 (1)	<.001	6.535 (1)	<.001
Former smoker	81.775 (1)	<.001	17.705 (1)	<.001
Diabetes	3.375 (1)	.04	11.879 (1)	<.001
Heart failure	7.824 (1)	<.001	27.911 (1)	<.001
Hypertension	52.403 (1)	<.001	20.809 (1)	<.001
Stroke	1.710 (1)	.17	9.027 (1)	<.001
COPD[c]	51.868 (1)	<.001	11.822 (1)	<.001
Professional exposure	13.198 (1)	<.001	13.242 (1)	<.001
Family history of cancer	0.073 (1)	.92	0.535 (1)	.46

[a]Values >0.25 considered significant.

[b]*P* values <.003 considered significant.

[c]COPD: chronic obstructive pulmonary disease.

For 8 features (ie, age, sex, active smoker, former smoker, heart failure, hypertension, COPD, and professional exposure), relationships with bladder cancer were highlighted in both CDRD and VTD, while 3 others were only considered related to bladder in VTD (diabetes, never smoked, and stroke). Finally, 4 features (weight, height, BMI, and family history of cancer) were not considered significant in any of the data sets.

Bladder Cancer Prediction Using Clinical Features

This section is dedicated to the investigations of the RFM. First, the best hyperparameters for each classification approach are reported. The performance between model candidates is then compared. To finish, the importance of each feature in the best model is investigated.

Hyperparameter Tuning

During the training phase, hyperparameters were tuned for each of the 3 classification approaches (RF, SVM, LR; Table 3).

In the end, the RF model candidate was composed of 45 trees with a maximum depth of 10 splits, minimum sample split of 10, and minimum sample leaf at 30 samples each. The RF model had an average AUC of 0.76 (SD 0.02). The SVM candidate was a linear kernel with a margin at 10 and 0.01 for gamma and had an average AUC of 0.77 (SD 0.02). The LR candidate embedded a margin at 0.01 with an l1 penalty and performed with an average AUC of 0.77 (SD 0.02). Model candidates were finally trained using these hyperparameters on the CDRD training set.

XSL•FO
RenderX

Table 3. Summary of the tuning of the hyperparameters.

Classifier and tuned hyperparameters	Tested values	Best AUC[a], mean (SD)
Random forest		0.76 (0.02)
number of trees	(10, 14, 18, 23, 27, 32, 41, 45[b], 50)	
maximal depth	(5, 6, 7, 8, 10[b], 13, 20)	
minimal sample split	(5, 10[b])	
minimal sample leaf	(4, 10, 30[b], 50)	
Support vector machine		0.77 (0.02)
kernel	(linear[b], radial basis function)	
margin	(0.001, 0.01, 0.1, 1, 10[b], 100)	
gamma	(0.001, 0.01[b], 0.1, 1, 10, 100)	
Logistic regression		0.77 (0.02)
margin	(0.001, 0.01[b], 0.1, 1, 10, 100)	
penalty	(l1[b], l2, none)	

Performance Comparison

To identify the best candidate to be deployed, performance on the training set (80% of CDRD) and applied once to the test set (20% of CDRD) were compared for each modeling approach. The resulting performances are reported in Figure 4.

Figure 4. Receiver operating characteristic (ROC) curve comparing the performance of risk factor–based classifiers in terms of the area under the ROC curve (AUC), balanced accuracy, specificity, and sensitivity for both the training and test sets.

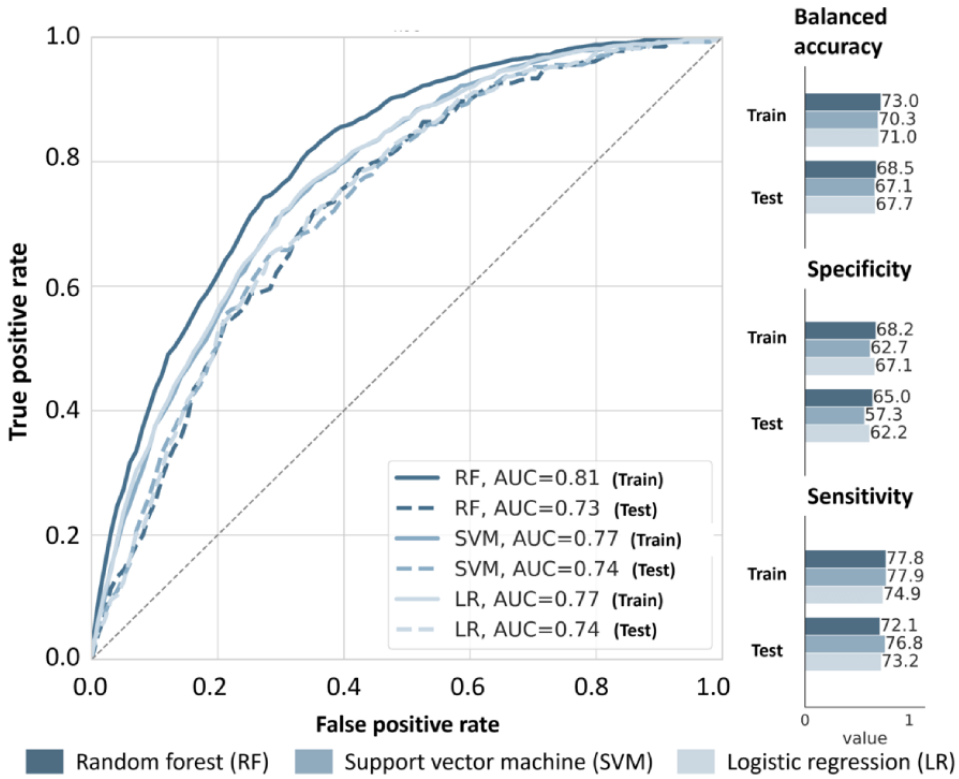

When looking at the AUC, all approaches succeeded, with values above 0.73 on the test set. However, there was a stronger overfitting with the RF candidate (ie, a drop of 0.08 between the training and test sets was observed). The other 2 models had similar AUCs on the training and test sets: 0.77 for SVM and 0.74 for LR. This indicates that these models generalized the problem better. Regarding the other evaluation metrics, the LR candidate had a better sensitivity-specificity balance (ie, sensitivity of 73.2% and specificity of 62.2% on the test set) than the SVM (ie, sensitivity of 76.8% and specificity of 57.3%

XSL•FO
RenderX

on the test set). In the end, the LR candidate was chosen for further investigation as the RFM (ie, feature importance analysis and combination with VisioCyt).

Impact of Clinical Features on Predictions

The individual importance of each feature for the prediction with LR is reported in Table 4.

Table 4. Feature importance of the risk factor–based model.

Feature	Importance model
age	0.14
sex	0.05
BMI_was_missing	0.03
smoking_status_active	0.03
smoking_status_former	0.01
heart_failure	0.01
weight	0.01
BMI	0.01
COPD	<0.01
stroke	<0.01
smoking_status_never	<0.01
professional_expo	<0.01
diabetes	<0.01
height	<0.01
COPD_was_missing	<0.01
hypertension_was_missing	<0.01
heart_failure_was_missing	<0.01
stroke_was_missing	<0.01
professional_expo_was_missing	<0.01
height_was_missing	<0.01
smoking_status_active_was_missing	<0.01
tumor_family_was_missing	<0.01
smoking_status_former_was_missing	<0.01
weight_was_missing	<0.01
sex_was_missing	<0.01
diabetes_was_missing	<0.01
smoking_status_never_was_missing	<0.01
tumor_family	<0.01
hypertension	<0.01

Unsurprisingly, age and sex were associated with the highest importance values (0.14 and 0.05, respectively). Information on patients' smoking status, through *smoking_status_active* and *smoking_status_former* features, had an impact on the predictions. In addition, weight and BMI were taken into consideration along with information on the presence or absence of BMI features (*BMI_was_missing*). Finally, the impact of a history of heart attacks (*heart_failure*) was also considered by the model. All other features had negligible univariate significance.

However, we could assume that the associations between these variables are the important factor in the model rather than when they are considered 1 by 1. Looking at 2-feature partial dependence plots allows us to explore this hypothesis. The combinations of age with features of low importance were investigated (ie, COPD, stroke, *professional_expo*, diabetes, *tumor_family*, and hypertension) and are presented in Figure 5.

XSL•FO
RenderX

Figure 5. Partial dependence plots for 6 feature combinations of age with (A) chronic obstructive pulmonary disorder (COPD), (B) stroke, (C) professional exposure, (D) diabetes, (E) tumor family, and (F) hypertension.

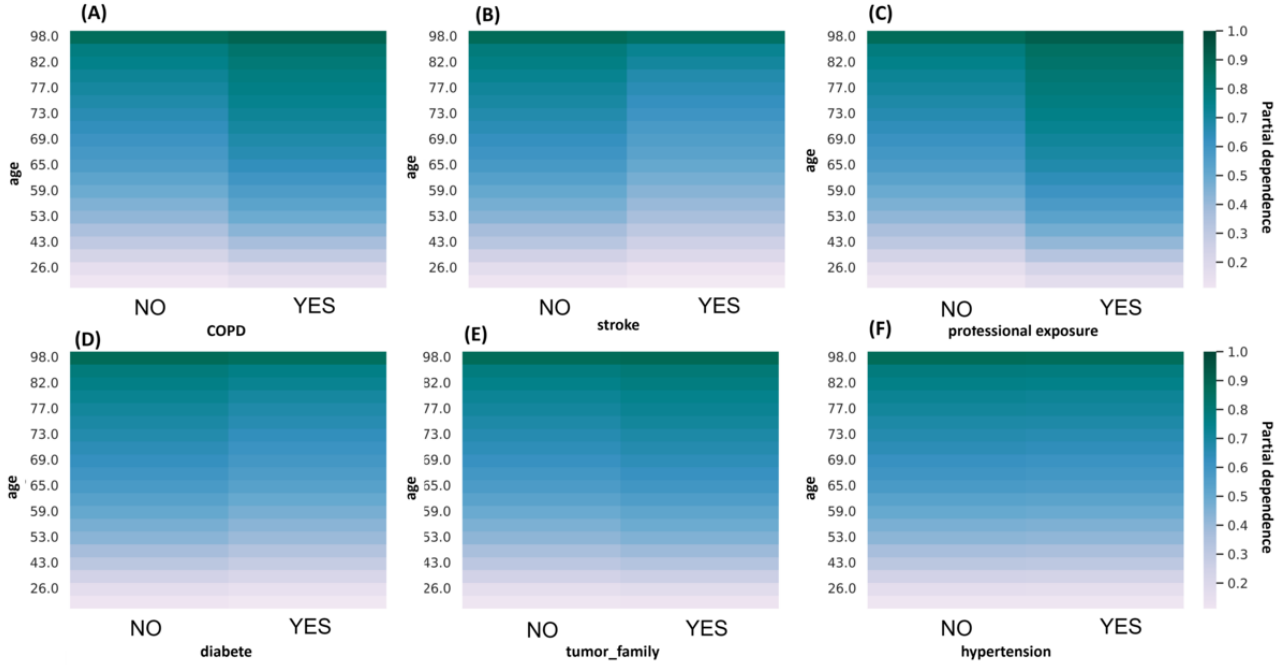

Patient age played an important role. Indeed, the highest probabilities (in dark green) were observed in the oldest patients. Higher risks (darker areas on the right than on the left of figures) were associated with COPD, professional exposure, and a family history of bladder cancer (*tumor_family*). This suggests that the model successfully captured the effects associated with these risk factors commonly identified in the literature. A history of stroke and the presence of diabetes were also considered but with the opposite effect than expected (absence increases risk). No impact of the hypertension feature was observed here.

Bladder Cancer Prediction by Combining RFM and VisioCyt

Building the Custom Method

The resulting scores for the VTD training set are reported in Figure 6.

Figure 6. 2D plot of risk factor–based model (RFM) and VisioCyt scores on the VisioCyt 1 Trial Dataset (VTD) training set.

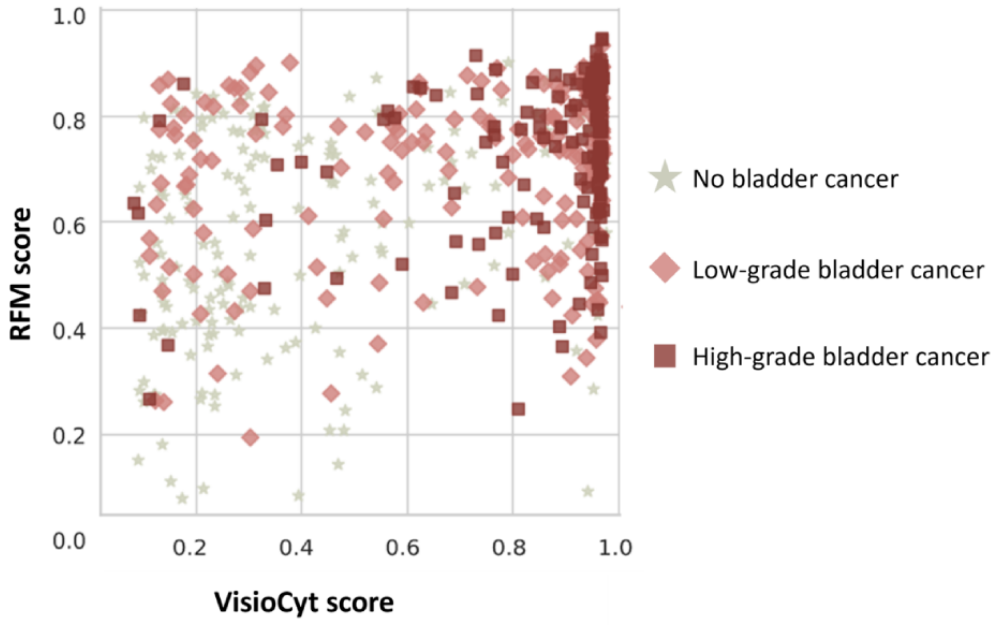

The first possible observation was that the detection of high-grade cancer was broadly identifiable by the high values of both scores. A VisioCyt score >0.6 indicates a high probability of cancer. In fact, very few patients without bladder cancer were in this zone of the scatter plot. Below this threshold, detection of low-grade cancers seems more difficult using the insight from VisioCyt alone. However, the RFM score was often

XSL•FO
RenderX

higher than 0.6 for these patients. From these observations, the custom combination rule was defined as follows:



Identification of the Best Modeling Approach

At first, the LR and DT models were subjected to 3-fold cross-validation on the training set to identify the best hyperparameters. LR with l1 regularization and a margin of 78.8 led to an AUC of 0.81 (SD 0.02) on the validation sets. DT with a minimum sample leaf of 5, maximum depth of 2, and a *Gini* criterion resulted in an AUC of 0.78 (SD 0.02). These hyperparameters were used to retrain the 2 approaches on the VTD training set.

The comparison between the 5 bladder cancer predictors is illustrated in Figure 7. It compiles the AUCs and score distributions according to bladder cancer severities of the image-based model, the RFM, and the 3 strategies for score combination: LR, DT, and custom. Initial scores obtained with VisioCyt and RFM are reported on the left, while scores from the combination using LR, DT, and custom lie on the right. Values obtained on the training and test sets of VTD were included to discuss the generalization of both approaches.

Figure 7. Distribution of the scores by bladder cancer grade for the VisioCyt 1 Trial Dataset (VTD) training set using (A) VisioCyt, (B) the risk factor–based model (RFM), (C) logistic regression (LR), (D) decision tree (DT), and (E) a custom strategy and for the VTD test set using (F) VisioCyt, (G) RFM, (H) LR, (I) DT, and (J) a custom strategy.

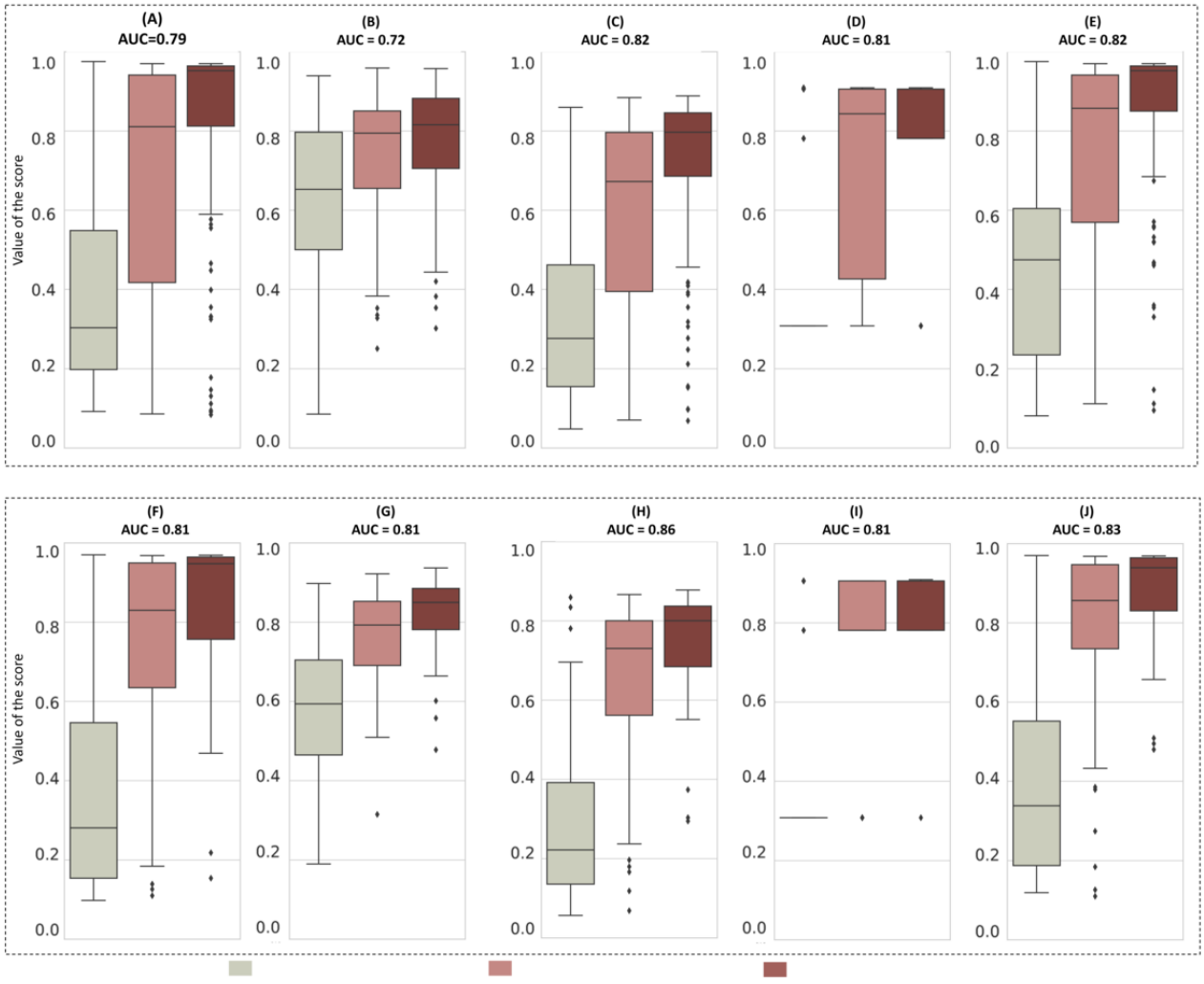

The first element to be noted is that combinations of the image-based model and RFM resulted in better performances in terms of AUC independent of the combination strategy (LR, DT, or custom). Indeed, the AUCs were higher for the training set (0.82 for LR and custom and 0.81 for DT vs 0.79 for VisioCyt and 0.7 for RFM) and higher or equivalent on the test set (0.86 for LR, 0.81 for DT, and 0.83 for custom vs 0.81 for VisioCyt and RFM). In the light of these sole observations, it was difficult to define the most relevant approach between LR, DT, and custom. To explore this further, we proposed studying boxplots of the VisioCyt, CBM, DT, and custom scores on the VTD training and test sets.

The median values of the scores obtained with LR were distinct for the 3 groups: (1) for patients without bladder cancer: 0.28 on the training set and 0.23 on the test set; (2) for patients with low-grade cancer: 0.68 on the training set and 0.72 on the test set; (3) for patients with high-grade cancer: 0.80 on the training set and 0.81 on the test set. The positive effect of this combination was a slight reduction in the overlap of scores

XSL•FO
RenderX

between low-grade patients and those without cancer observed with VisioCyt. However, the IQRs remained quite wide.

The median DT scores showed that DT scores were well separated for patients without (0.30) and with bladder cancer (0.84 for low-grade cancer and 0.78 for high-grade cancer). At first glance, these scores appeared to be more discriminating than the VisioCyt score alone. However, on the test set, IQRs of the DT scores were null for no bladder cancer and equivalent for low-grade and high-grade cancers, which does not allow for continuity in the measurements. This behavior does not allow for accurate interpretation. The doubt about this behavior was confirmed by the fact that the DT scores obtained for these patients during training resulted in a high IQR of 0.42, making it difficult to be sure that patients with low-grade cancer will be correctly associated with a high score. This edge effect also arose because a tree depth of 2 was selected during hyperparameter tuning, limiting the model's capacity to create more granular splits and accurately differentiate between subgroups in the population.

Regarding the custom approach, the scores associated with low-grade cancer (0.86 on the training set and 0.86 on the test set) and high-grade cancer (0.95 on both sets) were higher than the scores for patients without bladder cancer (0.47 on the training set and 0.34 on the test set). The range of scores for high-grade cancer was narrower with custom than with VisioCyt, indicating that several scores were improved and increased. Regarding patients with low-grade cancer, the median scores were higher with custom (eg, 0.89 on the test set) than with VisioCyt (eg, 0.83 on the test set). In addition, on the test and training sets, the overlap between scores for no bladder cancer patients and for patients with low-grade cancer was weaker than for VisioCyt scores for no cancer. The latter scores were also higher but remained distinct from those of the other 2 categories.

In light of these observations, the custom combination seems to be the most effective approach. The 95% CIs for the output scores were 0.33 to 0.52 for no bladder cancer, 0.68 to 0.83 for low-grade cancer, and 0.8 to 0.91 for high-grade cancer in the test set. The main advantage of the custom combination was improved detection of patients with low-grade cancer and increased certainty of the presence of high-grade cancer. One patient in each category was randomly drawn from the test set to illustrate what may be provided to clinician decision-making (Figure 8).

Figure 8. Illustration of the scores given by VisioCyt, the risk factor–based model (RFM), and the custom model for 3 patients in the VisioCyt 1 Trial Dataset (VTD) test set.

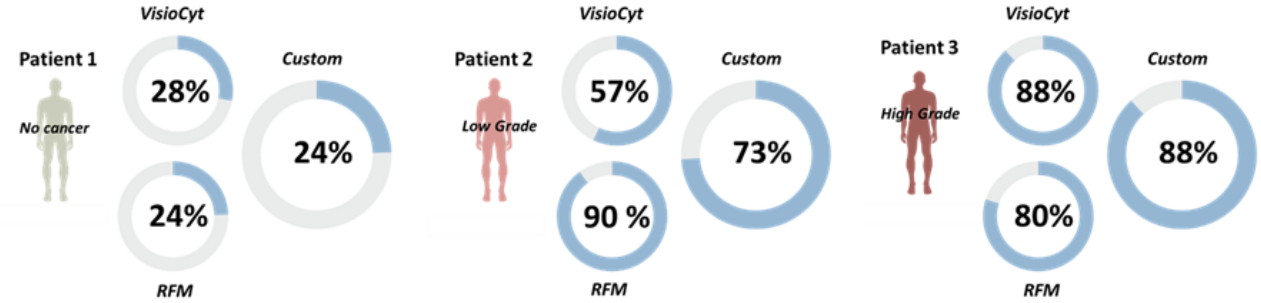

For patient 1, who did not have bladder cancer, the 2 initial scores were quite equivalent: 28% for VisioCyt and 24% for RFM. This also resulted in a low combined risk (28%). For patient 2, who had low-grade bladder cancer, the Visiocyt score was 57%, and the RFM score was 90%, giving a combined score of 73%. For patient 3, who had high-grade bladder cancer, the VisioCyt score was 88%, and the RFM score was 80%, leading to a combined score of 88%. The behaviors of the scores for these 3 patients were well in line with their actual diagnoses.

Although our goal was to create a decision support tool, not a bladder cancer detector, we evaluated the sensitivity and specificity by evaluating the thresholds of the scores obtained with the custom approach. The optimal cut-off point, which maximized the difference between true and false positive rates, was identified through ROC curve analysis. This yielded a threshold of 0.61 which results in 81% sensitivity and 73% specificity on the training set and 76% sensitivity and 83% specificity on the test set.

Discussion

Principal Findings

In this study, our team aimed to reuse large-scale clinical data to improve the performance of an image-based model to predict bladder cancer.

As a first step, an RFM was developed using clinical features from 5422 patients whose data were extracted from the CDW. Clinical features were selected because they were identified as important risk factors in the literature. The best RFM, based on LR, succeeded, with an AUC of 0.74, at classifying patients with or without bladder cancer from clinical features. The most predictive factors were age, sex, and smoking status (active or former). We also saw that risk factors such as COPD, professional exposure, and family history of bladder cancer, when combined with age, impacted the model's predictions. This model is simple and easily interpretable, successfully passing both internal (mean AUC 0.72, SD 0.02 during cross-validation and 0.74 on the CDRD test set) and external (AUC 0.81 on the VTD test set) validation tests.

In the second step, several combinations of this model with an image-based model were proposed. Classical approaches based

XSL•FO
RenderX

on LR and DT were proposed and supplemented by a custom approach grounded in RFM and VisioCyt score interpretation. We saw that all combinations gave a better AUC than RFM or VisioCyt alone. However, to identify the best combination, we stratified the performance by grade and chose the one that improved low-grade bladder cancer detection. The custom method (with an AUC of 0.83 on the VTD test set) provided the best results by assigning a median score of 0.34 to patients with no cancer, 0.86 to patients with low-grade cancer, and 0.95 to patients with high-grade cancer. Finally, we proposed observing the VisioCyt, RFM, and custom scores for 3 randomly selected patients, 1 from each of the 3 categories.

The resulting scores are supplementary and hold potential for enhancing the decision-making process. Indeed, they offer valuable insight into the patient's clinical profile, further improving image-based tools. Although the AUC bears resemblance to the figures observed in the studies conducted by Sanghvi et al [14] and Awan et al [15] (ie, 0.80 and 0.81, respectively), direct comparison was not feasible due to the focus on patients with high-grade cancer in those studies, which did not encompass an evaluation of performance for patients with low-grade cancer. Nevertheless, it is essential to underscore that the RFM is not designed for competitive purposes but rather to serve as a complementary component to these tools, much in the same manner as it did with VisioCyt. Combining RFM with these models may result in an even more robust composite score.

These results show the value of reusing data to propose better bladder cancer scoring strategies. Indeed, the strength of learning on thousands of patients has allowed us to propose a model that enhances the scoring by an image-based model for more accurate identification of patients with low-grade cancer.

Limitations

Some limitations of this work should be noted. First, the multiple missing values during clinical feature extraction raise concerns about the quality of the process. Multimedia Appendix 1 provides insight on the availability of each feature before imputation. This is an inherent problem in data reuse, which is not the primary purpose of the collection, because the information is not collected by default. Indeed, we are uncertain if these concepts are consistently present in all patient documents, as some data are recorded subjectively by clinicians. This is particularly true for variables such as COPD or a family history of cancer, features for which our data collection was limited in this study.

However, sometimes the information may still be hidden in unstructured data, which we are unable to exploit today. We used relatively straightforward extraction methods, possibly resulting in incomplete information retrieval from the reports. To address this, we plan to explore more advanced natural language processing methods like bilateral long short-term memory models or large language models [28], but their implementation and evaluation will be challenging due to limited annotations of the specific clinical concepts we sought to extract and the criticality of the confidentiality of electronic health records.

This last point could explain why we obtained feature distributions in CDRD that were sometimes different from those in VTD. However, they also may be explained by the fact that CDRD is larger than VTD since it includes a broader population. In fact, the percentage of patients with cancer was higher in VTD (66%) than in CDRD (25%). Overall, both data sets had the same characteristics, and most of the features were significant for bladder cancer detection. Some differences were nevertheless observed, and we will have to remain vigilant on these variables (ie, never smoked, stroke). We also observed the opposite behavior to what might be expected for diabetes and stroke in the RFM. Hypertension, though recognized as important in the literature, was not considered by the model. These are elements that deserve to be studied in greater detail once extraction methods are more robust.

Moreover, only 29 features were used, and they were often binary. Obtaining more complete information by either adding other clinical features (ie, biological measurement) or using a finer scale (ie, the type of diabetes or the number of cigarettes smoked per day) is an improvement not to be neglected and may increase the specificity of RFM. The same strategy can be used to refine the target. Indeed, it may be relevant to classify patients according to grade rather than in a binary way. Similarly, the proposed model was only based on features that we found to be linked with bladder cancer in the literature, but it will be interesting to add other comorbidities that may be cofounding factors (such as having another type of cancer).

In addition, we observed that the custom approach worked best. Within one of the defined score intervals, a simple average of the 2 scores was performed. A weighting may be integrated to fuse scores through an optimal fusion strategy [29].

Finally, although we carefully evaluated our models with regards to our available data sets, a concern for generalization may be observed later since only 1 data set with both cytology and clinical data (VTD) was investigated. First, we expect to enrich our RFM model by working on all the Hugo network's health data warehouses [24]. We also intend to deploy the combined model on new data to verify its stability and ensure relevance in routine care. Indeed, at the moment, the proposed combination has not yet been tested and approved by clinicians. However, it is important to consider that the collection of cytology images associated with clinical data requires the implementation of new acquisition protocols and that it will require time.

Conclusions

The results of this study are encouraging. This is the first time that the contribution of a prediction model based on risk factors identified in the literature and on the reuse of massive data has been studied. We observed that it can reinforce recent tools such as urine cytology–based models. Specifically, it may improve low-grade bladder cancer detection.

XSL•FO
RenderX

Acknowledgments

The authors thank VitaDX company for their collaborative contributions to this research.

Data Availability

The clinicobiological data used in this work contain pseudoidentifiable information and cannot be shared directly. However, a collaboration with the Clinical Data Center of Rennes Hospital can be considered to allow the use of these data while adhering to legal and regulatory frameworks. Please contact the corresponding author for more information. However, the conditions under which cytology data were collected do not permit their sharing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Values, type and interpretation of each feature and percentage of missing values across each feature in the Clinical Dataset (CDRD) and the VisioCyt 1 Trial Dataset (VTD) before imputation.
[DOCX File , 517 KB - jmir_v27i1e56946_app1.docx]

References

1. Chang HC, Lin CT, Lin WY, Yu PT. Analysis of the effects of cyproheptadine on bladder cancer through big data. Am J Cancer Res 2020;10(7):2114-2119 [FREE Full text] [Medline: 32775004]
2. Martinez Rodriguez RH, Buisan Rueda O, Ibarz L. Bladder cancer: Present and future. Medicina Clínica (English Edition) 2017 Nov;149(10):449-455. [doi: 10.1016/j.medcle.2017.10.005]
3. Usuba W, Urabe F, Yamamoto Y, Matsuzaki J, Sasaki H, Ichikawa M, et al. Circulating miRNA panels for specific and early detection in bladder cancer. Cancer Sci 2019 Jan 12;110(1):408-419 [FREE Full text] [doi: 10.1111/cas.13856] [Medline: 30382619]
4. Hsieh H, Jou Y, Tung C, Tsai Y, Wang Y, Chi C, et al. Epigenetic silencing of the dual-role signal mediator, ANGPTL4 in tumor tissues and its overexpression in the urothelial carcinoma microenvironment. Oncogene 2018 Feb 01;37(5):673-686. [doi: 10.1038/onc.2017.375] [Medline: 29035390]
5. Saginala K, Barsouk A, Aluru JS, Rawla P, Padala SA, Barsouk A. Epidemiology of bladder cancer. Med Sci (Basel) 2020 Mar 13;8(1):15 [FREE Full text] [doi: 10.3390/medsci8010015] [Medline: 32183076]
6. Babjuk M, Burger M, Compérat EM, Gontero P, Mostafid AH, Palou J, et al. European Association of Urology guidelines on non-muscle-invasive bladder cancer (TaT1 and carcinoma in situ) - 2019 update. Eur Urol 2019 Nov;76(5):639-657. [doi: 10.1016/j.eururo.2019.08.016] [Medline: 31443960]
7. Duan W, Du L, Jiang X, Wang R, Yan S, Xie Y, et al. Identification of a serum circulating lncRNA panel for the diagnosis and recurrence prediction of bladder cancer. Oncotarget 2016 Oct 25;7(48):78850-78858. [doi: 10.18632/oncotarget.12880]
8. Li H, Li C, Wu H, Zhang T, Wang J, Wang S, et al. Identification of Apo-A1 as a biomarker for early diagnosis of bladder transitional cell carcinoma. Proteome Sci 2011 Apr 17;9(1):21 [FREE Full text] [doi: 10.1186/1477-5956-9-21] [Medline: 21496341]
9. Białek Ł, Czerwińska K, Fus Ł, Krajewski W, Sadowska A, Radziszewski P, et al. MCM5 urine expression (ADXBLADDER) is a reliable biomarker of high-risk non- muscle-invasive bladder cancer recurrence: A prospective matched case-control study. Cancer Biomark 2021;30(2):139-143. [doi: 10.3233/CBM-200316] [Medline: 32924986]
10. Roperch J, Hennion C. A novel ultra-sensitive method for the detection of FGFR3 mutations in urine of bladder cancer patients - Design of the Urodiag® PCR kit for surveillance of patients with non-muscle-invasive bladder cancer (NMIBC). BMC Med Genet 2020 May 24;21(1):112 [FREE Full text] [doi: 10.1186/s12881-020-01050-w] [Medline: 32448160]
11. Vasdev N, Hampson A, Agarwal S, Swamy R, Chilvers M, Hampson A, et al. The role of URO17™ biomarker to enhance diagnosis of urothelial cancer in new hematuria patients-First European Data. BJUI Compass 2021 Jan 20;2(1):46-52 [FREE Full text] [doi: 10.1002/bco2.50] [Medline: 35474659]
12. Gandi C, Vaccarella L, Bientinesi R, Racioppi M, Pierconti F, Sacco E. Bladder cancer in the time of machine learning: Intelligent tools for diagnosis and management. Urologia 2021 May 05;88(2):94-102. [doi: 10.1177/0391560320987169] [Medline: 33402061]
13. Borhani S, Borhani R, Kajdacsy-Balla A. Artificial intelligence: A promising frontier in bladder cancer diagnosis and outcome prediction. Crit Rev Oncol Hematol 2022 Mar;171:103601 [FREE Full text] [doi: 10.1016/j.critrevonc.2022.103601] [Medline: 35065220]
14. Sanghvi AB, Allen EZ, Callenberg KM, Pantanowitz L. Performance of an artificial intelligence algorithm for reporting urine cytopathology. Cancer Cytopathol 2019 Oct 14;127(10):658-666 [FREE Full text] [doi: 10.1002/cncy.22176] [Medline: 31412169]

XSL•FO
RenderX

15. Awan R, Benes K, Azam A, Song T, Shaban M, Verrill C, et al. Deep learning based digital cell profiles for risk stratification of urine cytology images. Cytometry A 2021 Jul 20;99(7):732-742 [FREE Full text] [doi: 10.1002/cyto.a.24313] [Medline: 33486882]
16. Sokolov I, Dokukin ME, Kalaparthi V, Miljkovic M, Wang A, Seigne JD, et al. Noninvasive diagnostic imaging using machine-learning analysis of nanoresolution images of cell surfaces: Detection of bladder cancer. Proc Natl Acad Sci U S A 2018 Dec 18;115(51):12920-12925 [FREE Full text] [doi: 10.1073/pnas.1816459115] [Medline: 30509988]
17. Letašiová S, Medveďová A, Šovčíková A, Dušinská M, Volkovová K, Mosoiu C, et al. Bladder cancer, a review of the environmental risk factors. Environ Health 2012;11(Suppl 1):S11. [doi: 10.1186/1476-069x-11-s1-s11]
18. Xu Y, Huo R, Chen X, Yu X. Diabetes mellitus and the risk of bladder cancer: A PRISMA-compliant meta-analysis of cohort studies. Medicine (Baltimore) 2017 Nov;96(46):e8588 [FREE Full text] [doi: 10.1097/MD.0000000000008588] [Medline: 29145273]
19. Yu E, Stern M, Jiang X, Tang L, van den Brandt PA, Lu CM, et al. Family history and risk of bladder cancer: an analysis accounting for first- and second-degree relatives. Cancer Prev Res (Phila) 2022 May 03;15(5):319-326 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-21-0490] [Medline: 35027464]
20. Kok VC, Zhang H, Lin C, Huang S, Wu M. Positive association between hypertension and urinary bladder cancer: epidemiologic evidence involving 79,236 propensity score-matched individuals. Ups J Med Sci 2018 Jun 18;123(2):109-115 [FREE Full text] [doi: 10.1080/03009734.2018.1473534] [Medline: 29911922]
21. Tybjerg AJ, Babore AD, Olsen TS, Andersen KK. Types of occult cancer in stroke and the relation to smoking. Acta Neurol Scand 2020 Nov 13;142(5):486-492. [doi: 10.1111/ane.13299] [Medline: 32562260]
22. Roderburg C, Loosen SH, Jahn JK, Gänsbacher J, Luedde T, Kostev K, et al. Heart failure is associated with an increased incidence of cancer diagnoses. ESC Heart Fail 2021 Oct 27;8(5):3628-3633 [FREE Full text] [doi: 10.1002/ehf2.13421] [Medline: 34180146]
23. Lebret T, Pignot G, Colombel M, Guy L, Rebillard X, Savareux L, et al. Artificial intelligence to improve cytology performances in bladder carcinoma detection: results of the VisioCyt test. BJU Int 2022 Mar 26;129(3):356-363. [doi: 10.1111/bju.15382] [Medline: 33751774]
24. Madec J, Bouzillé G, Riou C, Van Hille P, Merour C, Artigny M, et al. eHOP Clinical Data Warehouse: from a prototype to the creation of an inter-regional clinical data centers network. Stud Health Technol Inform 2019 Aug 21;264:1536-1537. [doi: 10.3233/SHTI190522] [Medline: 31438219]
25. Wajsburt P, Petit-Jean T, Dura B, Cohen A, Jean C, Bey R. EDS-NLP: efficient information extraction from French clinical notes. GitHub. URL: http://aphp.github.io/edsnlp [accessed 2024-12-26]
26. Vogt WP. Point Biserial Correlation. In: Dictionary of Statistics & Methodology. New York, NY: SAGE Publications Inc; 2015:237-238.
27. Breiman L. Random forests. Machine Learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
28. Dumont J, Hernandez AI, Carrault G. Improving ECG beats delineation with an evolutionary optimization process. IEEE Trans. Biomed. Eng 2010 Mar;57(3):607-615. [doi: 10.1109/tbme.2008.2002157]
29. Chair Z, Varshney P. Optimal data fusion in multiple sensor detection systems. IEEE Trans. Aerosp. Electron. Syst 1986 Jan;AES-22(1):98-101. [doi: 10.1109/taes.1986.310699]

Abbreviations

AUC: area under the receiver operating characteristic curve
CDRD: Clinical Data Reuse Dataset
CDW: clinical data warehouse
COPD: chronic obstructive pulmonary disease
DT: decision tree
ICD-10: International Statistical Classification of Diseases, Tenth Revision
LR: logistic regression
PMSI: Programme de Médicalisation des Systèmes d'Information
RF: random forest
RFM: risk factor–based model
ROC: receiver operating characteristic
SVM: support vector machine
VTD: VisioCyt 1 Trial Dataset

XSL•FO
RenderX

Edited by A Coristine; submitted 31.01.24; peer-reviewed by J Luo, Q Wei, J Walsh; comments to author 27.09.24; revised version received 17.10.24; accepted 07.11.24; published 22.01.25.

Please cite as:
Cabon S, Brihi S, Fezzani R, Pierre-Jean M, Cuggia M, Bouzillé G
Combining a Risk Factor Score Designed From Electronic Health Records With a Digital Cytology Image Scoring System to Improve Bladder Cancer Detection: Proof-of-Concept Study
J Med Internet Res 2025;27:e56946
URL: https://www.jmir.org/2025/1/e56946
doi:10.2196/56946
PMID:

©Sandie Cabon, Sarra Brihi, Riadh Fezzani, Morgane Pierre-Jean, Marc Cuggia, Guillaume Bouzillé. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Risk Factors for Gastrointestinal Bleeding in Patients With Acute Myocardial Infarction: Multicenter Retrospective Cohort Study

Yanqi Kou[1*], MD; Shicai Ye[1*], MD; Yuan Tian[1,2*], MD; Ke Yang[3*], MMed; Ling Qin[4], MD; Zhe Huang[5], MD; Botao Luo[1,2], MD; Yanping Ha[1,2], MD; Liping Zhan[1], MMed; Ruyin Ye[1], MMed; Yujie Huang[1], MMed; Qing Zhang[1], MMed; Kun He[1], MMed; Mouji Liang[1], MMed; Jieming Zheng[1], MMed; Haoyuan Huang[1], MMed; Chunyi Wu[1], MBBS; Lei Ge[3*], MD; Yuping Yang[1*], MD

[1]Department of Gastroenterology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

[2]Department of Pathology, Guangdong Medical University, Zhanjiang, China

[3]Department of Gastrointestinal Surgery, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

[4]The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China

[5]Department of Colorectal Surgery, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

[*]these authors contributed equally

Corresponding Author:
Yuping Yang, MD
Department of Gastroenterology
Affiliated Hospital of Guangdong Medical University
No. 2 Wenming East Road, Xiashan, Zhanjiang
Zhanjiang, 524000
China
Phone: 1 13106629993
Email: yangyupingchn@163.com

Abstract

Background: Gastrointestinal bleeding (GIB) is a severe and potentially life-threatening complication in patients with acute myocardial infarction (AMI), significantly affecting prognosis during hospitalization. Early identification of high-risk patients is essential to reduce complications, improve outcomes, and guide clinical decision-making.

Objective: This study aimed to develop and validate a machine learning (ML)–based model for predicting in-hospital GIB in patients with AMI, identify key risk factors, and evaluate the clinical applicability of the model for risk stratification and decision support.

Methods: A multicenter retrospective cohort study was conducted, including 1910 patients with AMI from the Affiliated Hospital of Guangdong Medical University (2005-2024). Patients were divided into training (n=1575) and testing (n=335) cohorts based on admission dates. For external validation, 1746 patients with AMI were included in the publicly available MIMIC-IV (Medical Information Mart for Intensive Care IV) database. Propensity score matching was adjusted for demographics, and the Boruta algorithm identified key predictors. A total of 7 ML algorithms—logistic regression, k-nearest neighbors, support vector machine, decision tree, random forest (RF), extreme gradient boosting, and neural networks—were trained using 10-fold cross-validation. The models were evaluated for the area under the receiver operating characteristic curve, accuracy, sensitivity, specificity, recall, F_1_score, and decision curve analysis. Shapley additive explanations analysis ranked variable importance. Kaplan-Meier survival analysis evaluated the impact of GIB on short-term survival. Multivariate logistic regression assessed the relationship between coronary heart disease (CHD) and in-hospital GIB after adjusting for clinical variables.

Results: The RF model outperformed other ML models, achieving an area under the receiver operating characteristic curve of 0.77 in the training cohort, 0.77 in the testing cohort, and 0.75 in the validation cohort. Key predictors included red blood cell count, hemoglobin, maximal myoglobin, hematocrit, CHD, and other variables, all of which were strongly associated with GIB risk. Decision curve analysis demonstrated the clinical use of the RF model for early risk stratification. Kaplan-Meier survival analysis showed no significant differences in 7- and 15-day survival rates between patients with AMI with and without GIB (P=.83 for 7-day survival and P=.87 for 15-day survival). Multivariate logistic regression showed that CHD was an independent risk factor for in-hospital GIB (odds ratio 2.79, 95% CI 2.09-3.74). Stratified analyses by sex, age, occupation, marital status, and other subgroups consistently showed that the association between CHD and GIB remained robust across all subgroups.

XSL•FO
RenderX

Conclusions: The ML-based RF model provides a robust and clinically applicable tool for predicting in-hospital GIB in patients with AMI. By leveraging routinely available clinical and laboratory data, the model supports early risk stratification and personalized preventive strategies.

(J Med Internet Res 2025;27:e67346) doi:10.2196/67346

KEYWORDS

acute myocardial infarction; gastrointestinal bleeding; machine learning; in-hospital; prediction model

Introduction

Acute myocardial infarction (AMI) remains one of the leading causes of mortality worldwide and is a significant contributor to long-term disability and health care resource use [1]. Nosocomial gastrointestinal bleeding (GIB), as a common complication of patients with AMI, may lead to a large number of morbidity and mortality and remains one of the key factors affecting the prognosis of patients [2,3]. The high incidence and profound clinical implications of GIB have made it a focal point of concern in the management of patients with AMI [4].

Existing studies have established a close association between AMI and the occurrence of GIB, particularly in patients undergoing antiplatelet or anticoagulant therapy [5]. In patients with AMI, the development of GIB not only complicates in-hospital treatment but also markedly increases mortality rates during hospitalization [6]. However, the precise mechanisms underlying GIB in patients with AMI have yet to be fully elucidated. Potential mechanisms include gastrointestinal mucosal injury induced by anticoagulant therapy, stress-related ulceration, and inflammatory responses [7,8]. In addition, underlying comorbidities such as coronary heart disease (CHD), diabetes mellitus, and hypertension may also play significant roles in the development of GIB [9,10].

Although several studies have investigated the incidence and clinical manifestations of GIB in patients with AMI, there is still a lack of systematic analysis regarding its potential risk factors [11]. Previous research has identified common risk factors for GIB in patients with AMI, including anemia, renal insufficiency, advanced age, and polypharmacy [12,13]. However, many of these studies rely on traditional univariate or multivariate statistical methods, which may be limited in addressing the complexity of multifactorial interactions [14]. With the increasing application of machine learning (ML) methods in medical research, there is an opportunity to leverage advanced algorithms for more precise prediction and identification of GIB risk factors in patients with AMI.

In this real-world study, we aimed to integrate ML methods with traditional statistical analysis to comprehensively evaluate the risk factors for in-hospital GIB in patients with AMI. The objectives were to develop and validate an effective predictive model and to assess the impact of GIB on short-term prognosis, ultimately providing more robust guidance for clinical practice.

Methods

Data Source

Participants were retrospectively included based on their admission to the Affiliated Hospital of Guangdong Medical University between January 2005 and June 2024. A total of 10,046 patients with AMI were initially screened for eligibility. Patients were excluded if they had a history of GIB within 1 month before admission or if more than 20% of their data were missing. After applying these criteria, 1910 eligible patients with AMI were included (more details in the patient selection flowchart in the *Results* section). The cohort was divided based on the admission date: patients admitted before January 1, 2023, were assigned to the training cohort (n=1575), and those admitted after this date were included in the testing cohort (n=335).

Data for external validation were extracted from the publicly available, single-center Medical Information Mart for Intensive Care IV (MIMIC-IV) database (version 3.0). A total of 1801 patients with AMI from MIMIC-IV were initially screened. After applying the same exclusion criteria as in the training cohort, a total of 1746 patients with AMI were included in the external validation cohort. The Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology have approved MIMIC-IV. The diagnosis of AMI was based on the *Fourth Universal Definition of Myocardial Infarction*, including both ST-segment elevation (ST elevation) and non–ST-segment elevation (non-ST elevation) AMI. GIB was defined as a clinically significant bleeding event diagnosed by a physician (manifesting as coffee-ground emesis, hematemesis, melena, or hematochezia) or the presence of blood in the upper or lower gastrointestinal tract identified during endoscopic evaluation.

Data and Variables

The extracted variables included demographic characteristics, duration of hospital stay, discharge status, basic vital signs, and laboratory parameters. Cardiac injury biomarkers were recorded at their peak levels. Blood and biochemical test results were collected on the first day of admission. In cases where multiple test results were available for a specific variable, the first measurement was used in the analysis.

Model Construction and Validation

The Boruta algorithm was used in this study to select significant variables from the training dataset. Boruta is a feature selection method based on random forests (RFs) that determines the importance of each variable by comparing its *z* score with that of its “shadow” counterparts [15]. During the algorithm’s execution, all real features were duplicated and randomly

XSL•FO
RenderX

shuffled to generate *z* scores. If a real feature's *z* score consistently exceeded the maximum *z* score of the shadow features across multiple independent tests, it was deemed important and included in subsequent ML model construction.

The important variables screened by the Boruta algorithm were incorporated into 7 different ML algorithms for model construction, including logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), decision tree (DCtree), RF, extreme gradient boosting (XGBoost), and artificial neural network (NNET) [16]. Hyperparameter tuning was conducted using grid search combined with 10-fold cross-validation to optimize the performance of all ML models. Key parameters such as the number of trees, maximum tree depth, and minimum samples per leaf were systematically varied during the grid search. To prevent overfitting, the dataset was divided into 10 folds, with 9 folds used for training and 1 fold for validation. This process was repeated across all folds to identify the combination of hyperparameters that yielded the highest area under the curve (AUC).

The model with the highest area under the receiver operating characteristic (ROC) curve was selected to determine model performance. Model discrimination was further assessed using sensitivity, specificity, recall, accuracy, and F_1-score metrics. Decision curve analysis (DCA) was conducted to evaluate the clinical use of the models [17]. The best-performing model was subsequently used for further interpretive analysis. After training the model on the training cohort, all model parameters were fixed, and the model's performance was further evaluated using the testing and validation cohort.

To better understand the decision-making process of the best-performing model, Shapley additive explanations (SHAP) were used [18]. Based on cooperative game theory's Shapley values, SHAP rationally allocates contributions to the model's output among individual input features. SHAP also reveals feature importance and visually displays the direction and magnitude of each feature's contribution to the predictive outcome, thereby providing a deep understanding of the model's decision-making process. In this study, SHAP analysis was instrumental in identifying risk factors associated with GIB in patients with AMI and evaluating their consistency and accuracy in clinical applications using the testing and validation cohort.

Statistical Analysis

Multiple imputation was used to address missing data, including variables with a missing ratio of less than 20%. Continuous variables were expressed as medians and IQRs, while categorical variables were expressed as total numbers and percentages. Chi-square (χ^2) tests, Wilcoxon rank-sum tests, or Fisher exact tests were used as appropriate. For survival analysis, Kaplan-Meier curves were used. Logistic regression analysis was also performed to evaluate the association between CHD and the occurrence of GIB in patients with AMI. Propensity score matching (PSM) was also conducted using the nearest neighbor method with a 1:1 ratio without replacement and a caliper width of 0.15. The analyses were performed using R software (R Foundation for Statistical Computing; version 4.2.2). Statistical significance was determined by a 2-tailed *P* value of less than .05.

Ethical Considerations

This study was approved by the Institutional Review Board of the Affiliated Hospital of Guangdong Medical University (approval PJKT2024-176). The study used deidentified, retrospective clinical data, and informed consent was waived as the original consent obtained during initial data collection included provisions for secondary analysis. All patient data were anonymized, and confidentiality was strictly maintained. For the external validation cohort, database access was granted upon the completion of an online course and exam. As this portion of the study used a publicly accessible, deidentified database, informed consent was not required. Both data sources complied with all applicable ethical guidelines and privacy protection standards. The data from the training and testing cohorts were anonymized, eliminating the need for informed consent. In addition, the health information obtained from the MIMIC-IV database was deidentified, and therefore, patient consent was not required.

Results

Comparative Analysis of Clinical Profiles in Patients With AMI With and Without In-Hospital GIB

A total of 1575 patients with AMI were included in the training cohort (Figure 1), of whom 775 experienced in-hospital GIB. To minimize potential confounding factors, PSM was performed on variables such as sex, occupation, blood type, marital status, age, and length of hospital stay, achieving a balanced distribution of these selected characteristics between the groups. Table 1 compares the baseline characteristics before and after matching. After PSM, no statistically significant differences were observed between the groups for the matched variables (*P*>.05), indicating that the matching process effectively reduced confounding bias.

XSL•FO
RenderX

Figure 1. Patient selection flowchart for the training, testing, and validation cohorts. A total of 10,046 local patients and 1801 patients from the MIMIC-IV database were screened. After applying exclusion criteria, 1910 patients were included in the local cohort and 1746 patients in the external validation cohort. AMI: acute myocardial infarction; GIB: gastrointestinal bleeding; MIMIC-IV: Medical Information Mart for Intensive Care IV.

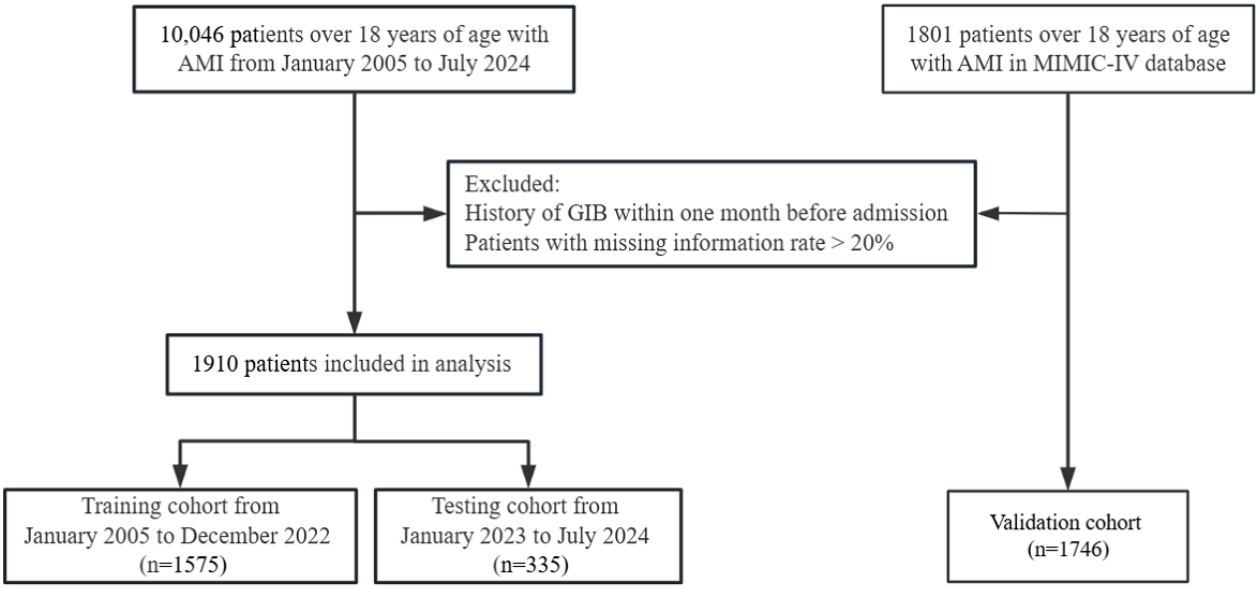

Table 1. Baseline demographic and clinical characteristics of patients with acute myocardial infarction with and without GIBa, before and after propensity score matching (2005-2024, Guangdong Medical University Hospital).

Characteristics	Unmatched			Matched		
	No GIB (n=800)	GIB (n=775)	*P* value	No GIB (n=731)	GIB (n=731)	*P* value
Age (years), mean (SD)	69 (13)	71 (12)	<.001	70 (13)	71 (12)	.19
In-hospital length of stay (days), mean (SD)	11 (9)	11 (9)	.85	11 (10)	11 (9)	.95
Gender, n (%)			.008			.68
Male	543 (68)	573 (7)		524 (72)	531 (73)	
Female	257 (32)	202 (26)		207 (28)	200 (27)	
Marriage status, n (%)			.15			.66
Married	758 (95)	721 (93)		689 (94)	685 (94)	
Single, divorced, or windowed	42 (5)	54 (7)		42 (6)	46 (6)	
Occupation, n (%)			.65			.86
Farmer	147 (18)	135 (17)		139 (19)	134 (18)	
Employee	419 (52)	397 (51)		383 (52)	379 (52)	
Retired or unemployed	234 (29)	243 (31)		209 (29)	218 (30)	
ABO blood type, n (%)			.29			.84
A	202 (25)	189 (24)		183 (25)	170 (23)	
O	301 (38)	318 (41)		287 (39)	299 (41)	
B	250 (31)	214 (28)		215 (29)	213 (29)	
AB	47 (6)	54 (7)		46 (6)	49 (7)	

[a]GIB: gastrointestinal bleeding.

The detailed demographic and baseline clinical characteristics of the PSM-adjusted training cohort are summarized in Table S1 in Multimedia Appendix 1. The results revealed significant differences in several laboratory parameters between patients with and without GIB. Specifically, patients with GIB exhibited significant differences in red blood cell count (RBC), hemoglobin, gamma-glutamyl transferase (GGT), total bilirubin (TBIL), direct bilirubin (DBIL), indirect bilirubin, albumin, total protein (TP), calcium, chloride, phosphorus, and N-terminal pro-brain natriuretic peptide maximum (NT-proBNP max), suggesting distinct clinical profiles related to anemia,

XSL•FO
RenderX

cardiovascular status, liver function, electrolyte imbalance, and coagulation function.

The analysis suggests that patients with AMI who experience in-hospital GIB may have distinct clinical profiles, particularly in terms of anemia, cardiovascular health, liver function, electrolyte balance, and coagulation, which could inform more targeted management strategies.

Key Predictor Identification Using Boruta Algorithm for In-Hospital GIB Risk in Patients With AMI

To identify key variables associated with GIB occurrence in patients with AMI, the Boruta algorithm was used for feature selection. Figure 2 illustrates the *z* scores of each variable, demonstrating how this analysis enhanced model optimization by concentrating on the most relevant features. A total of 20 variables were identified as important, including CHD, white blood cell count, RBC, hematocrit, hemoglobin, GGT, TBIL, DBIL, globulin, TP, total bile acids, creatinine, uric acid, phosphorus, alkaline phosphatase, cystatin C, fibrinogen, NT-proBNP max, maximal myoglobin, and maximal high-sensitivity cardiac troponin T, indicating their strong explanatory power in predicting GIB occurrence.

These variables were incorporated as primary predictors in the subsequent model construction and analysis, providing a solid foundation for developing predictive models.

Figure 2. Feature selection for predicting gastrointestinal bleeding in patients with acute myocardial infarction using the Boruta algorithm. The horizontal axis displays the names of each variable, while the vertical axis displays the corresponding z scores of each variable. The box plot shows the z score of each variable during model calculation. The blue boxes represent important variables, the red represents tentative variables, and the green represents rejected variables. AG: anion gap; ALB: albumin; ALB/GLO: albumin-to-globulin ratio; ALP: alkaline phosphatase; ALT: alanine aminotransferase; Apob: apolipoprotein B; Apoai: apolipoprotein A-I; APTT: activated partial thromboplastin time; APTTR: activated partial thromboplastin time ratio; Baso: basophils; Ca: calcium; CHD: coronary heart disease; ChE: cholinesterase; Cl: chloride; Crea: creatinine; CysC: cystatin C; DBIL: direct bilirubin; Eos: eosinophils; Fbg: fibrinogen; Glu: glucose; GGT: gamma-glutamyl transferase; GLO: globulin; HDLC: high-density lipoprotein cholesterol; Hb: hemoglobin; HCT: hematocrit; HCY: homocysteine; IBIL: indirect bilirubin; K: potassium; LDH: lactate dehydrogenase; LDLC: low-density lipoprotein cholesterol; Lymph: lymphocytes; Mb max: myoglobin maximum; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; Mono: monocytes; Na: sodium; Neut: neutrophils; NT proBNP max: N-terminal pro b-type natriuretic peptide maximum; P: phosphorus; PA: prealbumin; PCT: plateletcrit; PLT: platelets; PTR: prothrombin time ratio; PT: prothrombin time; PTINR: prothrombin time international normalized ratio; RBC: red blood cell; RDWCV: red cell distribution width; TG: triglycerides; TC: total cholesterol; TBIL: total bilirubin; TBA: total bile acids; TT: thrombin time; UA: uric acid; WBC: white blood cell.

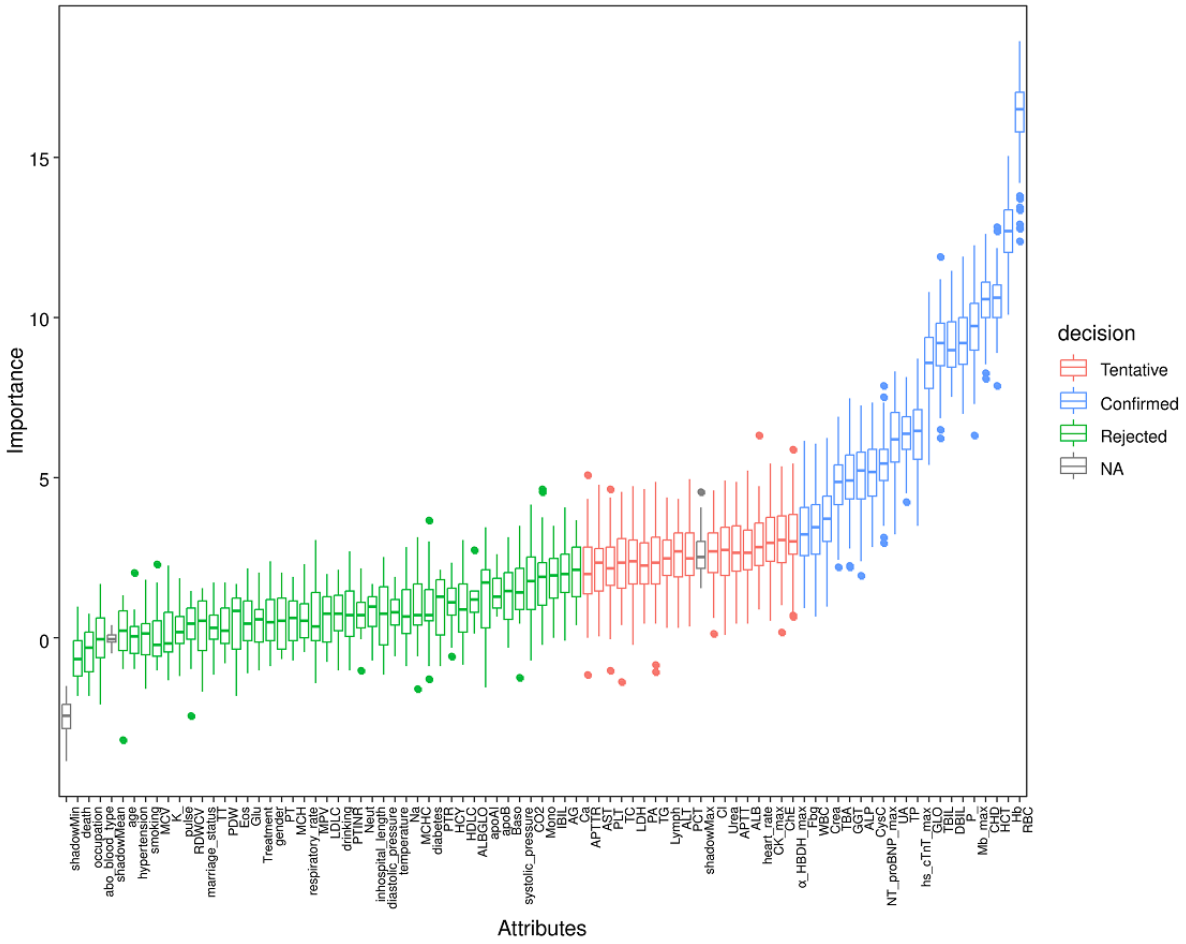

XSL•FO
RenderX

The RF Model Outperforms Other ML Models in Predicting In-Hospital GIB in Patients With AMI

The performance of 7 ML models in predicting GIB in patients with AMI was compared, and their ROC curves and DCA results were presented for the training cohort (Figure 3). The ROC curves (Figure 3A) demonstrate the discriminative ability of the various models, with the RF model achieving the highest AUC value (0.77, outperforming the others. The XGBoost (0.74) and SVM (0.72) models followed, while the DCtree model performed the poorest, with an AUC of 0.66. DCA (Figure 3B) assessed the clinical use of the models, further confirming that the RF model provided the highest net benefit across most threshold ranges, particularly within the intermediate range, highlighting its superiority in predicting GIB. The RF model demonstrated the best overall performance, with the highest AUC and net benefit, indicating its superior predictive capability for GIB. While other models, such as XGBoost and SVM, also performed well in certain scenarios, they were overall less effective than the RF model.

Figure 3. (A) Receiver operating characteristic curve and (B) decision curve analyses for 7 machine learning models predicting gastrointestinal bleeding in patients with acute myocardial infarction (training cohort). DCtree: decision tree; KNN: k-nearest neighbors; LR: logistic regression; NNET: neural network; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting.

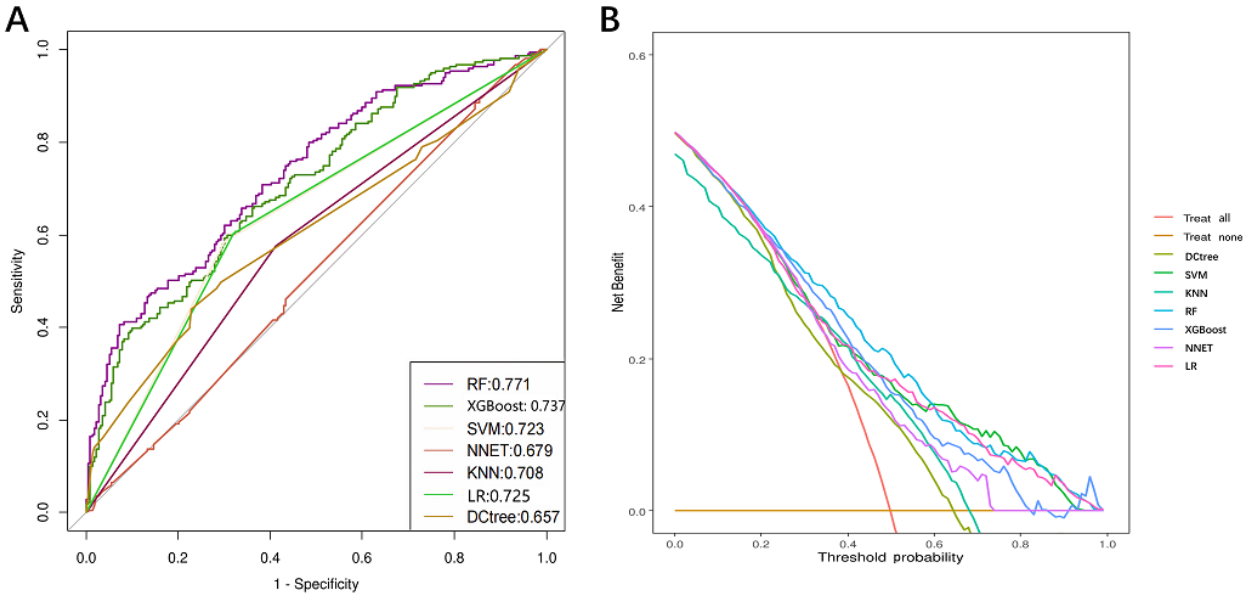

Further detailed performance metrics for each model, including sensitivity, specificity, recall, accuracy, and F_1-score, are provided in Table 2. The RF model emerged as the top performer, with an accuracy of 0.66; an F_1-score of 0.65; and balanced recall and specificity of 0.64 and 0.67, respectively, indicating its strong ability to distinguish between patients with and without GIB. The XGBoost model excelled in specificity, reaching 0.73, but its recall was lower at 0.58, resulting in an overall accuracy of 0.65 and an F_1-score of 0.62. The SVM model achieved the highest recall (0.73) but suffered from low specificity (0.45), yielding an overall accuracy of 0.59 and an F_1-score of 0.64. In contrast, the LR and KNN models exhibited similar performance, with accuracies of 0.61 and 0.60 and balanced recall and specificity around 0.60, reflecting moderate ability in distinguishing patients with GIB. The DCtree model reached a specificity of 0.73 but had lower recall (0.51), leading to an overall accuracy of 0.62 and an F_1-score of 0.57. The NNET model performed the worst, with an accuracy of only 0.55 and an F_1-score of 0.52, suggesting it may not be suitable for GIB prediction in this dataset.

XSL•FO
RenderX

Table 2. Performance comparison of machine learning models for predicting gastrointestinal bleeding in the training cohort.

Model	Accuracy	Sensitivity	Specificity	Recall	F_1_score
RF[a]	0.66	0.64	0.67	0.64	0.65
XGBoost[b]	0.65	0.58	0.73	0.58	0.62
SVM[c]	0.59	0.73	0.45	0.73	0.64
NNET[d]	0.55	0.49	0.62	0.49	0.52
KNN[e]	0.60	0.60	0.60	0.60	0.60
LR[f]	0.61	0.61	0.60	0.61	0.61
DCtree[g]	0.62	0.51	0.73	0.51	0.57

[a]RF: random forest.

[b]XGBoost: extreme gradient boosting.

[c]SVM: support vector machine.

[d]NNET: neural network.

[e]KNN: k-nearest neighbors.

[f]LR: logistic regression.

[g]DCtree: decision tree.

Additional evaluations were conducted using both the testing and validation cohorts to assess the robustness of the models (Figures S1 and S2 in Multimedia Appendix 1; Tables S2 and S3 in Multimedia Appendix 1). The model's performance was consistent, with an ROC value of 0.77 in the testing cohort and 0.75 in the validation cohort, demonstrating its generalizability and reliability across different datasets.

Overall, the RF model consistently outperformed other ML models in both predictive accuracy and clinical use, making it the most reliable tool for predicting in-hospital GIB in patients with AMI across multiple datasets.

SHAP Analysis Quantifies Feature Contributions to In-Hospital GIB Risk in Patients With AMI

The SHAP analysis was used to interpret the predictions of the best-performing model, the RF. SHAP analysis highlighted the impact of 15 key features on the GIB prediction model in patients with AMI, ranked by SHAP values (Figure 4A). Features with positive SHAP values were associated with higher predicted GIB risk, while negative SHAP values indicated a lower risk. Darker colors represent decreases, while lighter colors indicate increases in the respective parameters, with all predictive factors positively correlated with GIB. The SHAP analysis also emphasized the importance of specific variables and provided insights into their directional influence on the prediction outcomes. RBC exhibited the strongest predictive influence, with lower values significantly increasing the predicted risk, followed by myoglobin maximum, CHD, hemoglobin, phosphorous, and NT-proBNP max, which also contributed to the prediction. TBIL and hematocrit further demonstrated their significance in the model. Other influential features included globulin, highly sensitive cardiac troponin T maximum, fasting blood glucose, GGT, DBIL, uric acid, and alkaline phosphatase. With varying SHAP values, these features illustrated their multidimensional impact within the predictive model (Figure 4B).

XSL•FO
RenderX

Figure 4. (A) SHAP (Shapley additive explanations) analysis of the top-15 predictors for gastrointestinal bleeding (GIB) in patients with acute myocardial infarction (AMI) using a random forest model ranked by mean absolute SHAP value, and (B) the impact of SHAP values on the occurrence of GIB in patients with AMI. ALP: alkaline phosphatase; CHD: coronary heart disease; DBIL: direct bilirubin; Fbg: fibrinogen; GGT: gamma-glutamyl transferase; GLO: globulin; HCT: hematocrit; Hb: hemoglobin; Hs cTnT Max: high-sensitivity cardiac troponin T maximum; Mb max: myoglobin maximum; NT proBNP max: N-terminal pro b-type natriuretic peptide maximum; P: phosphorus; RBC: red blood cell; TBIL: total bilirubin; UA: uric acid.

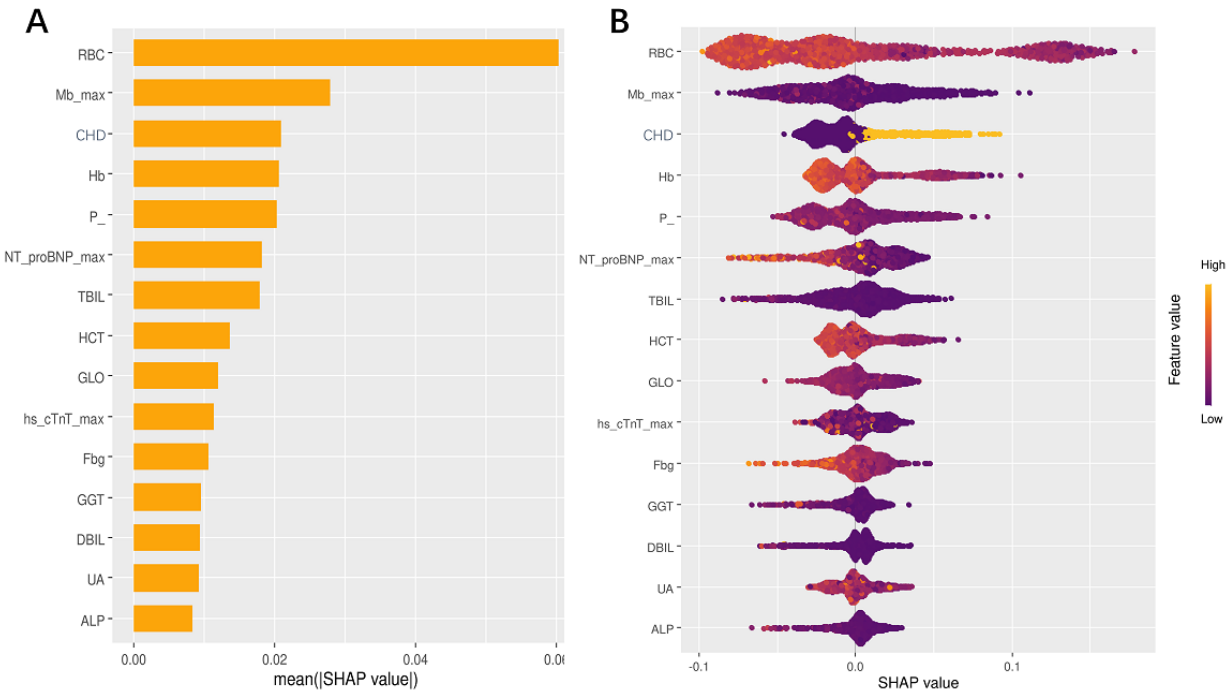

These features provided insights into the driving factors for GIB occurrence in patients with AMI, playing a crucial role in GIB prediction and understanding the underlying mechanisms. Notably, the SHAP values for CHD history indicated that an increase in this variable significantly heightened the risk of GIB, suggesting that CHD may play a crucial role in the pathophysiological mechanisms underlying GIB.

Kaplan-Meier Analysis Shows No Significant Impact of GIB on Short-Term Survival in Patients With AMI

To assess the impact of GIB on short-term survival in patients with AMI, a Kaplan-Meier survival analysis was performed. Figure S3 in Multimedia Appendix 1 shows the survival curves for patients with and without GIB at 7 and 15 days. The results indicated that while overall survival rates were lower in the GIB group compared to the non-GIB group, the differences did not reach statistical significance (P=.83 for 7-day survival and P=.87 for 15-day survival). This suggests that in the study population, the occurrence of GIB did not significantly affect short-term survival in patients with AMI.

These findings suggest that while GIB is associated with lower survival rates, it does not significantly influence short-term mortality in patients with AMI within the study population.

CHD Identified as a Significant Predictor of In-Hospital GIB Risk in Patients With AMI

Multivariate logistic regression was used to analyze further the association between CHD and in-hospital GIB in patients with AMI. The results, presented in Table 3, showed a significant association between CHD and the occurrence of GIB during hospitalization. In the unadjusted crude model, CHD was significantly positively associated with GIB (odds ratio [OR] 2.88, 95% CI 2.29-3.63; P<.001). After PSM, the association remained significant, with the OR slightly reduced to 2.50 (95% CI 1.98-3.17; P<.001). Even after stepwise adjustment for additional confounding variables, including sex, occupation, ABO blood type, marital status, age, and length of hospital stay (Model 1), hypertension, diabetes, smoking, alcohol consumption, body temperature, pulse, heart rate, respiratory rate, blood pressure (Model 2), and all laboratory parameters (Model 3), the significant association between CHD and GIB persisted. The OR values remained consistent across the different models, all retaining statistical significance (P<.001), indicating that CHD is an important risk factor for GIB during hospitalization in patients with AMI.

Furthermore, Figure 5 presents the results of subgroup analysis and interaction tests conducted using the LR model to further explore the moderating effects of various characteristics on the relationship between CHD and GIB. Stratified analyses by sex, age, occupation, marital status, and other subgroups consistently showed that the association between CHD and GIB remained robust across all subgroups. This underscores the generalizability of CHD as a significant risk factor for GIB in patients with AMI, highlighting the importance of considering this risk factor in diverse patient populations.

XSL•FO
RenderX

Table 3. Association between coronary heart disease and risk of gastrointestinal bleeding in patients with acute myocardial infarction.

Variable	Overall		Propensity-score matched	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Crude model[b]	2.88 (2.29-3.63)	<.001	2.50 (1.98-3.17)	<.001
Model 1[c]	2.74 (2.16-3.49)	<.001	2.56 (2.01-3.27)	<.001
Model 2[e]	2.70 (2.12-3.45)	<.001	2.54 (1.98-3.25)	<.001
Model 3[e]	2.92 (2.20-3.86)	<.001	2.79 (2.09-3.74)	<.001

[a]OR odds ratio.

[b]Crude model: No covariates were adjusted.

[c]Model 1: adjusted for gender, occupation, ABO blood type, marriage status, age, and in-hospital length of stay.

[d]Model 2: Model 1 and adjusted for hypertension, diabetes, smoking, drinking, temperature, pulse, heart rate, respiratory rate, diastolic pressure, and systolic pressure.

[e]Model 3: Model 2 and adjusted for all laboratory parameters.

Figure 5. Subgroup analysis of coronary heart disease (CHD) impact on gastrointestinal bleeding (GIB) risk in patients with acute myocardial infarction (AMI). These results confirm that CHD is a consistent and significant risk factor for in-hospital GIB in patients with AMI, reinforcing its importance in risk stratification across diverse patient subgroups. OR: odds ratio.

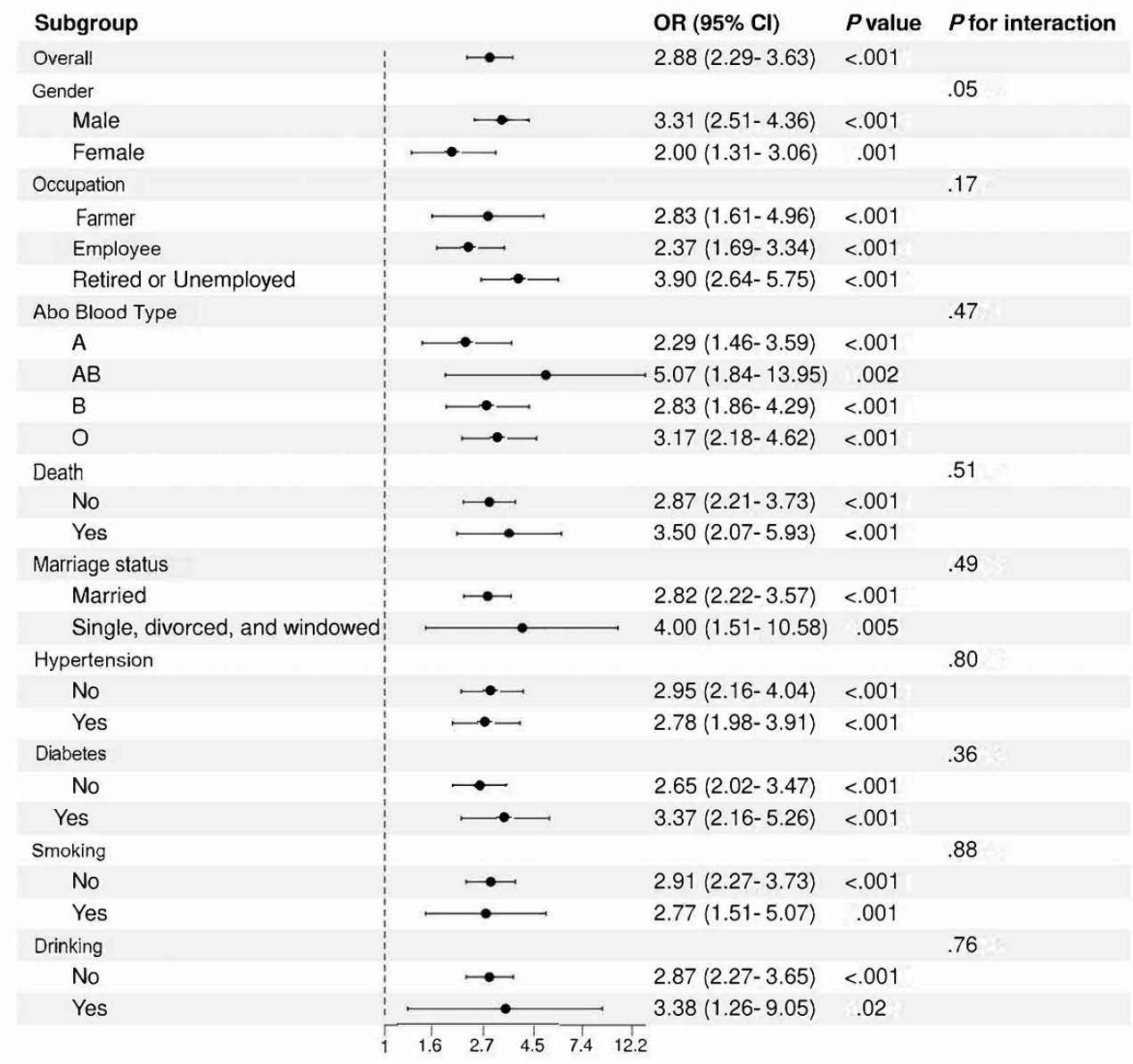

XSL•FO
RenderX

Discussion

Principal Findings

This study systematically analyzes the risk factors associated with GIB in patients with AMI. Using the Boruta algorithm, significant variables were selected and compared across various ML models to construct an effective RF model for predicting GIB occurrence. This research not only enhances understanding of the mechanisms behind GIB in patients with AMI but also provides clinicians with a robust predictive tool aimed at reducing patient complications. In addition, Kaplan-Meier survival analysis was used to assess GIB's impact on short-term survival rates and analyze the association between CHD and GIB, as well as its prognosis implications.

Comparison With Previous Work

In comparison with existing literature, our study exhibits several unique advantages. First, we found that patients with GIB had significantly lower RBC, hemoglobin, and multiple liver function indicators compared to patients without GIB, aligning with previous research findings [19,20]. For instance, Martí et al [21] similarly identified anemia and hepatic dysfunction as pivotal risk factors for GIB in patients with AMI. Furthermore, Holster et al [22], through systematic review and meta-analysis, underscored the heightened risk of GIB associated with anticoagulant use, particularly among patients with underlying hepatic or hematologic conditions. This study further substantiates the significance of these factors through multivariable analysis and enhances prediction accuracy using ML methods. However, models from different studies exhibit variability in predicting GIB occurrence. For example, Chin et al [23] model heavily relies on traditional multivariable regression analysis, identifying significant risk factors such as advanced age, renal impairment, and anticoagulant use, albeit with limited predictive performance [23]. In contrast, our study's adoption of the Boruta algorithm for feature selection and integration of the RF model not only enhances model robustness but also significantly improves predictive performance with an AUC value of 0.771. This underscores the capability of ML methods to offer precise risk prediction when handling complex, high-dimensional data.

Furthermore, through multivariable logistic regression analysis, this study further confirms a significant association between CHD and GIB occurrence, which persists even after adjusting for various confounding factors. Consistent with previous research, this suggests that patients with CHD may face increased GIB risk due to heightened inflammatory response and the necessity for anticoagulant therapy [24]. Subgroup analysis underscores the consistent impact of CHD across different patient demographics, emphasizing the need for heightened management attention, particularly in administering antiplatelet and anticoagulant therapies, which may necessitate stricter monitoring and personalized adjustments [25,26]. Kaplan-Meier survival analysis results indicate that although GIB is a common complication in patients with AMI, it does not significantly impact short-term survival rates. This finding aligns with the observations by Nikolsky et al [27] in patients after coronary artery bypass graft, suggesting minimal short-term prognosis effects among actively treated patients [27]. However, this does not negate the potential long-term impacts of GIB on patient outcomes, warranting further investigation.

Strengths and Limitations

This study represents an innovative application of ML techniques to predict GIB in patients with AMI, integrating both clinical and laboratory data. The multicenter design, coupled with external validation using the MIMIC-IV database, enhances the generalizability of the findings across diverse patient populations. Robust validation methods, including PSM and the use of independent datasets, further strengthen the reliability of the predictive models and mitigate potential biases.

This study has several limitations that should be acknowledged. First, although it used a multicenter retrospective cohort design, the sample size may still be insufficient to fully capture the variability and complexity of GIB in patients with AMI. External validation using the MIMIC-IV database partially mitigates this limitation by increasing the generalizability of the findings. However, larger, multicenter prospective studies are needed to further validate these results across diverse populations. Second, some important variables, such as dietary habits and medication adherence, were not included in the analysis. These unmeasured confounding factors could influence the occurrence of GIB and limit the comprehensiveness and clinical applicability of the model. Incorporating additional predictors, such as dietary habits, medication adherence, and genetic markers, in future studies could enhance the model's accuracy and provide a more holistic understanding of GIB risk. Third, the retrospective nature of the study introduces inherent limitations, including potential selection bias and reliance on existing medical records, which may include missing or incomplete data. In addition, in datasets where the limit of detection is an important factor, alternative imputation methods, such as minimal imputation, may provide more accurate processing of missing data. While this is not the main focus of this study, future studies could explore advanced imputation techniques tailored to specific missing mechanisms to improve the robustness of predictive models. To address this, PSM was applied to minimize confounding factors, and multiple imputation techniques were used to handle missing data. Despite these efforts, prospective studies with real-time clinical data collection would provide stronger evidence, reduce bias, and further validate the findings. In addition, leveraging advanced ML techniques and incorporating real-time clinical data may improve the model's predictive performance and facilitate early intervention for high-risk patients.

Conclusion

In summary, ML models, particularly the RF model, offer reliable tools for predicting GIB in patients with AMI and lay the groundwork for future research into its pathophysiology and prevention.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the staff and clinicians at the Affiliated Hospital of Guangdong Medical University and all individuals involved in the data collection process. Special thanks to the Massachusetts Institute of Technology for providing access to the MIMIC-IV (Medical Information Mart for Intensive Care IV) database. This study was supported by the Big Data Platform of the Affiliated Hospital of Guangdong Medical University, the In-Hospital Funding Clinical Research Project of the Affiliated Hospital of Guangdong Medical University (LCYJ2023A002), and The Clinical-Basic Science and Technology Innovation Special Program of Guangdong Medical University (No. GDMULCJC2024019). No generative artificial intelligence (AI) tools were used in this manuscript. All figures submitted have been created by the authors, who confirm that the images are original with no duplication and have not been previously published in whole or in part.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YK, SY, YT, and KY contributed to conceptualization, validation, formal analysis, data curation, and writing–original draft preparation. LQ and ZH provided supervision, resources, and contributed to writing–review and editing. BL, YH, LZ, and RY were involved in data curation, validation, and formal analysis. YH, QZ, KH, JL, JZ, WC, and HH contributed to methodology, software, and visualization. LG and YY were responsible for the investigation, data curation, and writing–review and editing. All authors reviewed and approved the final manuscript.

LG is the co-corresponding author, they can be reached at:
Department of Gastrointestinal Surgery, Affiliated Hospital of Guangdong Medical University
No. 2 Wenming East Road, Xiashan, Zhanjiang
Zhanjiang, 524000, China
Phone: 19126570562
Email: champane@qq.com

Conflicts of Interest

None declared.

Multimedia Appendix 1
Risk factors for gastrointestinal bleeding (GIB) in patients with acute myocardial infarction (AMI).
[DOCX File , 711 KB - jmir_v27i1e67346_app1.docx]

References

1. Martin SS, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, et al. 2024 heart disease and stroke statistics: a report of US and Global Data From the American Heart Association. Circulation 2024;149(8):e347-e913 [FREE Full text] [doi: 10.1161/CIR.0000000000001209] [Medline: 38264914]
2. Oakland K, Guy R, Uberoi R, Hogg R, Mortensen N, Murphy MF, et al. Acute lower GI bleeding in the UK: patient characteristics, interventions and outcomes in the first nationwide audit. Gut 2018;67(4):654-662. [doi: 10.1136/gutjnl-2016-313428] [Medline: 28148540]
3. Oakland K, Kahan BC, Guizzetti L, Martel M, Bryant RV, Brahmania M, et al. Development, validation, and comparative assessment of an international scoring system to determine risk of upper gastrointestinal bleeding. Clin Gastroenterol Hepatol 2019;17(6):1121-1129.e2. [doi: 10.1016/j.cgh.2018.09.039] [Medline: 30268566]
4. Bilal M, Samuel R, Khalil MK, Singh S, Parupudi S, Abougergi MS. Nonvariceal upper GI hemorrhage after percutaneous coronary intervention for acute myocardial infarction: a national analysis over 11 months. Gastrointest Endosc 2020;92(1):65-74.e2. [doi: 10.1016/j.gie.2020.01.039] [Medline: 32017916]
5. Saltzman JR. The challenge of nonvariceal upper GI bleeding management in patients with acute coronary syndrome receiving dual-antiplatelet therapy. Gastrointest Endosc 2020;92(1):75-77. [doi: 10.1016/j.gie.2020.03.012] [Medline: 32586567]
6. Tersalvi G, Biasco L, Cioffi GM, Pedrazzini G. Acute coronary syndrome, antiplatelet therapy, and bleeding: a clinical perspective. J Clin Med 2020;9(7):2064 [FREE Full text] [doi: 10.3390/jcm9072064] [Medline: 32630233]
7. Sehested TSG, Carlson N, Hansen PW, Gerds TA, Charlot MG, Torp-Pedersen C, et al. Reduced risk of gastrointestinal bleeding associated with proton pump inhibitor therapy in patients treated with dual antiplatelet therapy after myocardial infarction. Eur Heart J 2019;40(24):1963-1970. [doi: 10.1093/eurheartj/ehz104] [Medline: 30851041]
8. Nikolsky E, Stone GW, Kirtane AJ, Dangas GD, Lansky AJ, McLaurin B, et al. Gastrointestinal bleeding in patients with acute coronary syndromes: incidence, predictors, and clinical implications: analysis from the ACUITY (Acute Catheterization

XSL•FO
RenderX

and Urgent Intervention Triage Strategy) trial. J Am Coll Cardiol 2009;54(14):1293-1302 [FREE Full text] [doi: 10.1016/j.jacc.2009.07.019] [Medline: 19778672]
9. Emenike E, Srivastava S, Amoateng-Adjepong Y, al-Kharrat T, Zarich S, Manthous CA. Myocardial infarction complicating gastrointestinal hemorrhage. Mayo Clin Proc 1999;74(3):235-241. [doi: 10.4065/74.3.235] [Medline: 10089991]
10. Cortés P, Zeng JJ, Karime C, Lewis MD, Gharacholou SM, Antwi SO, et al. Validation of prediction tools for GI bleeding in patients on dual anti-platelet therapy after percutaneous coronary intervention. Gastrointest Endosc 2024;99(1):10-20.e6. [doi: 10.1016/j.gie.2023.08.002] [Medline: 37579980]
11. Wärme J, Sundqvist M, Mars K, Aladellie L, Pawelzik S, Erlinge D, et al. Helicobacter pylori screening in clinical routine during hospitalization for acute myocardial infarction. Am Heart J 2021;231:105-109 [FREE Full text] [doi: 10.1016/j.ahj.2020.10.072] [Medline: 33144087]
12. Zhong L, Quan X, Dang P, Tang M, Yu H, Guo F. Clinical characteristics and risk factors of in-hospital gastrointestinal bleeding in patients with acute myocardial infarction. Front Cardiovasc Med 2022;9:933597 [FREE Full text] [doi: 10.3389/fcvm.2022.933597] [Medline: 36237901]
13. Ducrocq G, Puymirat E, Steg PG, Henry P, Martelet M, Karam C, et al. Blood transfusion, bleeding, anemia, and survival in patients with acute myocardial infarction: FAST-MI registry. Am Heart J 2015;170(4):726-734.e2. [doi: 10.1016/j.ahj.2015.07.004] [Medline: 26386796]
14. Kikkert WJ, Hassell ME, Delewi R, van der Laan MH, Baan J, Vis MM, et al. Predictors and prognostic consequence of gastrointestinal bleeding in patients with ST-segment elevation myocardial infarction. Int J Cardiol 2015;184:128-134. [doi: 10.1016/j.ijcard.2015.01.041] [Medline: 25703420]
15. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Soft 2010;36(11):1-13. [doi: 10.18637/jss.v036.i11]
16. Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
17. Zech J, Pain M, Titano J, Badgeley M, Schefflein J, Su A, et al. Natural language-based machine learning models for the annotation of clinical radiology reports. Radiology 2018;287(2):570-580. [doi: 10.1148/radiol.2018171093] [Medline: 29381109]
18. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clin Transl Sci 2024;17(11):e70056. [doi: 10.1111/cts.70056] [Medline: 39463176]
19. Shi W, Fan X, Yang J, Ni L, Su S, Yu M, et al. In-hospital gastrointestinal bleeding in patients with acute myocardial infarction: incidence, outcomes and risk factors analysis from China acute myocardial infarction registry. BMJ Open 2021;11(9):e044117 [FREE Full text] [doi: 10.1136/bmjopen-2020-044117] [Medline: 34493500]
20. Khan R, Lopes RD, Neely ML, Stevens SR, Harrington RA, Diaz R, et al. Characterising and predicting bleeding in high-risk patients with an acute coronary syndrome. Heart 2015;101(18):1475-1484. [doi: 10.1136/heartjnl-2014-307346] [Medline: 26109589]
21. Martí D, Carballeira D, Morales MJ, Concepción R, Del Castillo H, Marschall A, et al. Impact of anemia on the risk of bleeding following percutaneous coronary interventions in patients ≥75 years of age. Am J Cardiol 2020;125(8):1142-1147. [doi: 10.1016/j.amjcard.2020.01.010] [Medline: 32087994]
22. Holster IL, Valkhoff VE, Kuipers EJ, Tjwa ET. New oral anticoagulants increase risk for gastrointestinal bleeding: a systematic review and meta-analysis. Gastroenterology 2013;145(1):105-112.e15. [doi: 10.1053/j.gastro.2013.02.041] [Medline: 23470618]
23. Chin MWS, Yong G, Bulsara MK, Rankin J, Forbes GM. Predictive and protective factors associated with upper gastrointestinal bleeding after percutaneous coronary intervention: a case-control study. Am J Gastroenterol 2007;102(11):2411-2416. [doi: 10.1111/j.1572-0241.2007.01460.x] [Medline: 17850413]
24. Patel P, Nigam N, Sengupta N. Lower gastrointestinal bleeding in patients with coronary artery disease on antithrombotics and subsequent mortality risk. J Gastroenterol Hepatol 2018;33(6):1185-1191. [doi: 10.1111/jgh.14048] [Medline: 29156506]
25. Ng F, Wong S, Lam K, Chang C, Lau Y, Chu W, et al. Gastrointestinal bleeding in patients receiving a combination of aspirin, clopidogrel, and enoxaparin in acute coronary syndrome. Am J Gastroenterol 2008;103(4):865-871. [doi: 10.1111/j.1572-0241.2007.01715.x] [Medline: 18177451]
26. Lee C, Cheng C, Kao Yang Y, Chao T, Chen J, Li Y. Cardiovascular and bleeding risks in acute myocardial infarction newly treated with ticagrelor vs. clopidogrel in Taiwan. Circ J 2018;82(3):747-756 [FREE Full text] [doi: 10.1253/circj.CJ-17-0632] [Medline: 29081474]
27. Nikolsky E, Mehran R, Stone GW. Gastrointestinal bleeding in percutaneous coronary intervention and acute coronary syndromes. Am J Cardiol 2009;104(5 Suppl):22C-29C. [doi: 10.1016/j.amjcard.2009.06.022] [Medline: 19695357]

Abbreviations

AMI: acute myocardial infarction
AUC: area under the curve

XSL•FO
RenderX

CHD: coronary heart disease
DBIL: direct bilirubin
DCA: decision curve analysis
DCtree: decision tree
GGT: gamma-glutamyl transferase
GIB: gastrointestinal bleeding
KNN: k-nearest neighbors
LR: logistic regression
MIMIC-IV: Medical Information Mart for Intensive Care IV
ML: machine learning
NNET: neural network
NT-proBNP max: N-terminal pro-brain natriuretic peptide maximum
OR: odds ratio
PSM: propensity score matching
RBC: red blood cell count
RF: random forest
ROC: receiver operating characteristic
SHAP: Shapley additive explanations
SVM: support vector machine
TBIL: total bilirubin
TP: total protein
XGBoost: extreme gradient boosting

Edited by A Coristine; submitted 09.10.24; peer-reviewed by J Tanna, G Shu; comments to author 17.12.24; revised version received 20.12.24; accepted 21.12.24; published 30.01.25.

Please cite as:
Kou Y, Ye S, Tian Y, Yang K, Qin L, Huang Z, Luo B, Ha Y, Zhan L, Ye R, Huang Y, Zhang Q, He K, Liang M, Zheng J, Huang H, Wu C, Ge L, Yang Y
Risk Factors for Gastrointestinal Bleeding in Patients With Acute Myocardial Infarction: Multicenter Retrospective Cohort Study
J Med Internet Res 2025;27:e67346
URL: https://www.jmir.org/2025/1/e67346
doi:10.2196/67346
PMID:

©Yanqi Kou, Shicai Ye, Yuan Tian, Ke Yang, Ling Qin, Zhe Huang, Botao Luo, Yanping Ha, Liping Zhan, Ruyin Ye, Yujie Huang, Qing Zhang, Kun He, Mouji Liang, Jieming Zheng, Haoyuan Huang, Chunyi Wu, Lei Ge, Yuping Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Easy and Quick Risk-Stratified Early Forewarning Model for Septic Shock in the Intensive Care Unit: Development, Validation, and Interpretation Study

Guanghao Liu[1,2,3*], PhD; Shixiang Zheng[2,4*], MD, PhD; Jun He[1,2,5*], PhD; Zi-Mei Zhang[6], PhD; Ruoqiong Wu[1,2], MSc; Yingying Yu[1,2], MSc; Hao Fu[1,2], MSc; Li Han[1,2], MSc; Haibo Zhu[1,2,3], PhD; Yichang Xu[1,2,7], MSc; Huaguo Shao[1,2], MSc; Haidan Yan[1,2,5], PhD; Ting Chen[1,2,8], PhD; Xiaopei Shen[1,2,3,5], PhD

[1]Department of Bioinformatics, School of Medical Technology and Engineering, Fujian Medical University, Fuzhou, China

[2]Fujian Key Laboratory of Medical Bioinformatics, Institute of Precision Medicine, Fujian Medical University, Fuzhou, China

[3]School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China

[4]Department of Critical Care Medicine, Union Hospital of Fujian Medical University, Fuzhou, China

[5]Key Laboratory of Gastrointestinal Cancer (Fujian Medical University), Ministry of Education, Fuzhou, China

[6]School of Biology and Biological Engineering, South China University of Technology, Guangzhou, China

[7]School of Medical Imaging, Fujian Medical University, Fuzhou, China

[8]Department of Computer Science and Technology, Institute of Artificial Intelligence, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xiaopei Shen, PhD
Department of Bioinformatics
School of Medical Technology and Engineering
Fujian Medical University
1 Xue Yuan Road, University Town
Fuzhou, 350122
China
Phone: 86 17366882017
Email: xshen@fjmu.edu.cn

Abstract

Background: Septic shock (SS) is a syndrome with high mortality. Early forewarning and diagnosis of SS, which are critical in reducing mortality, are still challenging in clinical management.

Objective: We propose a simple and fast risk-stratified forewarning model for SS to help physicians recognize patients in time. Moreover, further insights can be gained from the application of the model to improve our understanding of SS.

Methods: A total of 5125 patients with sepsis from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) database were divided into training, validation, and test sets. In addition, 2180 patients with sepsis from the eICU Collaborative Research Database (eICU) were used as an external validation set. We developed a simplified risk-stratified early forewarning model for SS based on the weight of evidence and logistic regression, which was compared with multi-feature complex models, and clinical characteristics among risk groups were evaluated.

Results: Using only vital signs and rapid arterial blood gas test features according to feature importance, we constructed the Septic Shock Risk Predictor (SORP), with an area under the curve (AUC) of 0.9458 in the test set, which is only slightly lower than that of the optimal multi-feature complex model (0.9651). A median forewarning time of 13 hours was calculated for SS patients. 4 distinct risk groups (high, medium, low, and ultralow) were identified by the SORP 6 hours before onset, and the incidence rates of SS in the 4 risk groups in the postonset interval were 88.6% (433/489), 34.5% (262/760), 2.5% (67/2707), and 0.3% (4/1301), respectively. The severity increased significantly with increasing risk in both clinical features and survival. Clustering analysis demonstrated a high similarity of pathophysiological characteristics between the high-risk patients without SS diagnosis (NS_HR) and the SS patients, while a significantly worse overall survival was shown in NS_HR patients. On further exploring the characteristics of the treatment and comorbidities of the NS_HR group, these patients demonstrated a significantly

XSL•FO
RenderX

higher incidence of mean blood pressure <65 mmHg, significantly lower vasopressor use and infused volume, and more severe renal dysfunction. The above findings were further validated by multicenter eICU data.

Conclusions: The SORP demonstrated accurate forewarning and a reliable risk stratification capability. Among patients forewarned as high risk, similar pathophysiological phenotypes and high mortality were observed in both those subsequently diagnosed as having SS and those without such a diagnosis. NS_HR patients, overlooked by the Sepsis-3 definition, may provide further insights into the pathophysiological processes of SS onset and help to complement its diagnosis and precise management. The importance of precise fluid resuscitation management in SS patients with renal dysfunction is further highlighted. For convenience, an online service for the SORP has been provided.

(J Med Internet Res 2025;27:e58779) doi:10.2196/58779

KEYWORDS

septic shock; early forewarning; risk stratification; machine learning

Introduction

Sepsis is a syndrome or group of symptoms associated with infection rather than a single disease [1]. Reportedly, approximately 48.9 million cases of sepsis occurred globally in 2017, with approximately 11 million sepsis-related deaths, accounting for approximately 19.7% of all deaths worldwide [2]. According to the Third International Consensus on Sepsis and Septic Shock held in 2016 [1], sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection, and septic shock (SS) is defined as an advanced state of sepsis in which severe circulatory, cellular, and metabolic abnormalities lead to a greater risk of death than sepsis alone, with a mortality rate estimated to be as high as 45% [3,4]. Kumar et al [5] found that patients with SS treated within the first hour of diagnosis had a 79.9% survival rate, but each hour of delay in treatment was associated with a 7.6% increase in mortality. Therefore, early forewarning of the onset of SS and timely treatment can help reduce the morbidity, mortality, and length of intensive care unit (ICU) stay. However, early forewarning is challenging due to the complexity of the disease in the clinical context [6,7] and the heterogeneity of the sepsis population [8].

Recently, machine learning techniques have been applied to electronic health record data to develop models for risk prediction [9,10], clinical phenotyping [11-13], and early forewarning [14-19]. While several studies have focused on early forewarning of SS in the ICU, practical limitations remain [20-24]. Many models required extensive clinical features from multiple examinations [21,22], increasing the complexity and time for data collection, which can delay forewarnings. Efforts to reduce the feature number often relied on distributed examinations [23] and left limited time for intervention [24]. Additionally, many studies prioritized predictive performance [25-27] while neglecting clinical interpretability. Specifically, limited attention has been given to the triggers of inconsistent predictions, a crucial factor for improving models and the understanding of disease diagnostic criteria. To enhance clinical applicability, models must offer straightforward actionable insights rather than complex algorithms that burden clinicians.

Therefore, we developed a real-time risk-stratified forewarning model for SS that minimized the need for extensive clinical examinations, enabling physicians to quickly and easily assess the patient's risk and facilitate timely intervention. Furthermore, after stratifying patients based on the risk scores provided by the model, we further analyzed the clinical characteristics of patients in different risk groups, with a particular focus on high-risk (HR) patients. A public web application was deployed that enables easy access and use of our proposed model [28].

Methods

Datasets and Participants

To improve the quality and reproducibility of reporting, we prepared a checklist (Multimedia Appendix 1) according to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines [29].

Data were collected from an open-source dataset released for sizeable critical care databases, including the Medical Information Mart for Intensive Care-IV (MIMIC-IV) [30] and the eICU Collaborative Research Database (eICU) [31]. The MIMIC-IV contains comprehensive and high-quality data on deidentified patients admitted to ICUs at the Beth Israel Deaconess Medical Center (BIDMC), with more than 76,000 admissions between 2008 and 2019. The eICU is a multicenter publicly available database containing deidentified highly granular medical data for more than 200,000 patients in 335 ICUs from 208 hospitals across the United States between 2014 and 2015. Both databases include vital signs (VSs), laboratory results, diagnosis, treatment, care plan documentation, severity of illness measures, and other information.

Patients aged 18 years or older who were admitted for the first time and stayed in the ICU for more than 24 hours were included in the analysis. To reduce the impact of complicating factors in defining SS, patients designated as "do not resuscitate" or "encounter for palliative care" according to the International Classification of Diseases (ICD) were excluded. Moreover, patients in the coronary care unit were excluded to reduce the impact of cardiogenic shock. A final cohort of 5125 patients (including 766 SS patients) in the MIMIC-IV and 2180 patients (480 SS patients) in the eICU was obtained. The MIMIC-IV data were randomly partitioned into training (60%), validation (20%), and testing (20%) sets. The eICU data were used for external validation.

Ethical Considerations

The MIMIC-IV and eICU databases are publicly accessible, deidentified, and transformed, and have been approved for use

XSL•FO
RenderX

by the institutional review boards at BIDMC and the Massachusetts Institute of Technology, in accordance with the Declaration of Helsinki. Both databases received waivers for informed consent due to the deidentification of all protected health information. Access to these databases was granted upon completion of human research ethics training and signing of a data use agreement with PhysioNet (certification number 38539898).

Definitions of Sepsis and SS

The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [1] were applied to determine the onset time of sepsis and SS. Specifically, all episodes of suspected infection ($t_{suspicion}$) were identified by the timestamp of antibiotic administration or blood culture, whichever was earlier, within a specific period. If the antibiotic was administered first, the culture sampling was required to be obtained within 24 hours. If the culture sampling occurred first, the antibiotic was required to be ordered within 72 hours. The onset time of sepsis (t_{sepsis}) was defined as an episode of suspected infection with a change in the Sequential Organ Failure Assessment (SOFA) score of 2 or more points from up to 48 hours before to up to 24 hours after $t_{suspicion}$. Patients with SS were identified by a clinical construct of sepsis with persisting hypotension requiring vasopressors to maintain mean arterial pressure ≥65 mmHg and with a serum lactate level >2 mmol/L (18 mg/dL) despite adequate volume resuscitation [1]. Specifically, adequate fluid resuscitation was defined as a total fluid volume greater than 30 mL/kg or 1000 mL within a 3-hour window before the initiation of vasoactive drugs or hypotension. The first lactate level needed to be greater than 2 mmol/L within 6 hours after adequate resuscitation. The onset time of SS ($t_{septic\ shock}$) was the earliest time when all criteria for SS were met. Due to a lack of information on the starting and ending times of fluid resuscitation, SS, according to Sepsis-3, could not be defined in the eICU data. Instead, we used the ICD code labels in the "diagnosis.csv" table to determine SS onset.

Data Processing

Given the characteristics of time-series data and the need for real-time monitoring for SS risk, this study defined a 6-hour retrospective window, whereby data collected during the 6 hours prior to the current state of the patient were used to provide a forewarning of the SS risk in the next 6 hours. We discretized the features of high-frequency detection, such as VSs, and the features of low-frequency testing, such as laboratory test data, to an hourly value. If multiple measurements were performed in the same hour, their mean value was calculated as the value for this hour of the feature. The forward-filling method was used to fill in missing measurements (with the early measurement), if there was no measurement in this hour.

Next, several time intervals were defined to assess and compare differences among risk groups (derived from the Septic Shock Risk Predictor [SORP] risk scores) in terms of patient pathophysiology (based on VSs and laboratory testing features), clinical characteristics, etc. The time of SS onset is defined as 0. Considering this, –1 indicates 1 hour before onset and +1 indicates 1 hour after onset, and the risk prediction time is –6. Three primary time intervals were delimited, namely, –12 to –6, –6 to 0, and 0 to +6, representing the look-back interval, the preonset interval, and the postonset interval, respectively, where the look-back interval was the time interval used for data collection for risk calculation by the model. For easier comparison among risk groups, the mean value of a clinical feature was used to represent its value within a specified time interval.

Notably, information on patient comorbidities (eg, chronic kidney disease [CKD]) was taken from ICD codes, and invasive procedures (eg, dialysis-continuous renal replacement therapy [CRRT]) were taken from the procedures documented ("procedureevents" table).

Risk Model

To facilitate clinical application, we filtered the features and then developed a scorecard [9,10] to evaluate SS onset risk by logistic regression (LogR). Next, to present the patient's risk more visually, we stratified the risk scoring and assessed the characteristics of patients in each risk group. The workflow is shown in Figure 1.

The scorecard was constructed as follows. The scorecard allowed for transferring continuous variables into multiple bins, and each feature was divided into multiple bins. In this study, we used the weight of evidence (WOE) to measure the probability of illness in each bin and determine the most appropriate number of bins by detecting the change in information value (IV) under each bin. The WOE and IV were calculated according to equations 1 and 2, respectively. Specifically, first, features were divided into 10 bins, making sure that each bin contained both SS patients and nonshock (NS) patients. Subsequently, the chi-square test was performed on 2 neighboring bins, traversing and merging those bins with the largest *P*-value until the number of bins met the expectation. Then, the final number of bins was determined based on the IV. After the binning was completed, the WOE value of each bin was calculated, and then, the WOE was mapped back to the original dataset to build the LogR model.





where i represents each subgroup, $DistributionNS_i$ denotes the proportion of septic NS patients in group i to the NS patients in the entire feature, and $Distribution\ S_i$ denotes the proportion of SS patients in group i to the SS patients in the entire feature.

The scorecard points were calculated using the following equations:







XSL•FO
RenderX

where *P0* and *PDO* (points to double odds) are the score ranges that were set manually, woe_j can be calculated with equation 1, β_i denotes the coefficient of the *i*th feature in LogR, α is the intercept term from LogR, *n* is the number of features, and *k* is the number of bins in each feature.

The scorecard was adjusted to a range of scores between 0 and 100. The final scorecard consisted of a base score and a score for each bin in each feature (Multimedia Appendix 2).

Figure 1. The flowchart of this study. The study flowchart comprises 4 stages. (A) Model development: Electronic health record (EHR) data were collected from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) and eICU Collaborative Research Database (eICU), and multiple forewarning models were constructed. Using Shapley Additive Explanations (SHAP) analysis, we identified key features within the same test batch, which were then used to build the Septic Shock Risk Predictor (SORP) model. The performance of the SORP was compared to that of models based on all available features to evaluate whether the SORP performance loss is acceptable. (B) Risk stratification of septic shock (SS): Patients were stratified into risk groups based on SORP scores. (C) Clinical characteristics of the risk groups: We analyzed clinical characteristics across risk groups, with a focused analysis on high-risk nonshock patients. (D) Model deployment: An online tool was developed for clinical use. HR: high risk; LR: low risk; MR: medium risk; ULR: ultralow risk; WOE: weight of evidence.

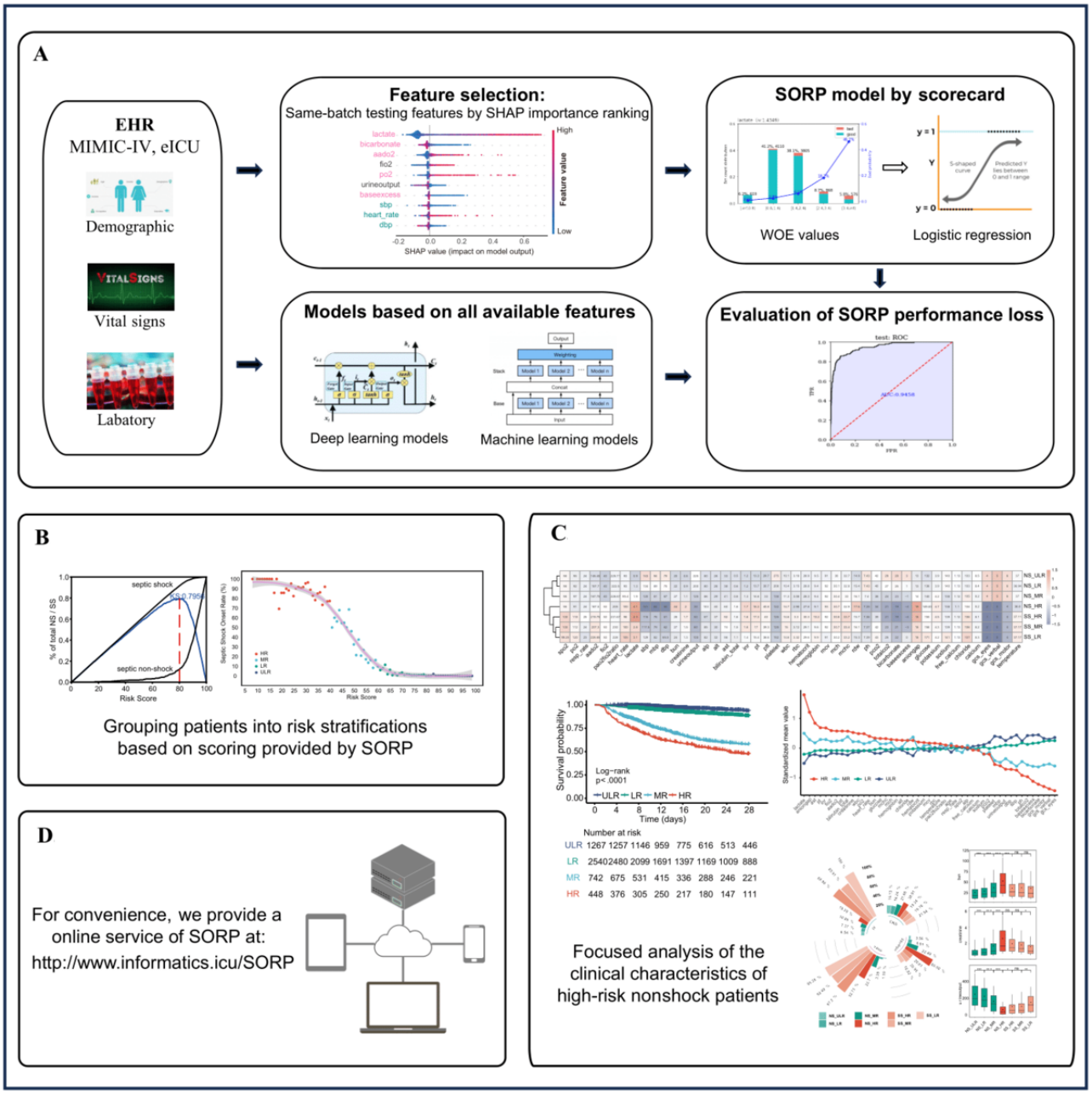

Statistical Methods

We employed the Shapley Additive Explanations (SHAP) [32] method to calculate the importance of clinical features for SS in the look-back interval. SHAP assigns a value to each feature in the given prediction. These values are calculated for each prediction separately and do not provide general information about the entire model. Higher absolute SHAP values indicate higher importance, whereas values close to 0 indicate low significance of a feature.

XSL•FO
RenderX

To facilitate observation and evaluation and determine risk intervals, we employed the Kolmogorov-Smirnov (KS) curve to describe the overall score distribution. The KS curve illustrated the changes in the sample proportion of SS patients, the proportion of septic NS patients, and the trend of differences between SS and NS patients (blue line) with a score from 0 to 100. From the KS curve, it could be seen that the proportion of SS patients in the low score accumulated faster, while the proportion of NS patients in the high score accumulated faster. The blue curve showed the variation process of the difference between SS and NS patients at each score point. At the highest point of the blue curve, the ability to distinguish between SS and NS patients was the strongest. Therefore, to provide a more direct scoring effect, based on the distributions described above, the scores were divided into 4 risk groups: HR, medium risk (MR), low risk (LR), and ultralow risk (ULR), with score intervals of 0-40, 40-60, 60-80, and 80-100, respectively.

The Population Stability Index (PSI) was used to evaluate the stability of risk group distribution across different datasets in the postonset interval. A smaller PSI value indicates a minor difference between 2 distributions and more consistency. In addition, the area under the curve (AUC) was calculated to evaluate the performance of each model. Estimation and comparison of the survival curves among risk groups were performed using Kaplan-Meier survival analysis and log-rank tests, respectively. The Local Polynomial Regression Fitting (locally estimated scatterplot smoothing [LOESS]) method was used to fit the relationship between the risk score and SS onset rates at 6 hours. Hierarchical clustering was used to assess the similarity between risk groups. All statistical analyses and algorithm approaches in this study were performed using Python (Python Software Foundation) and R (R Project for Statistical Computing).

Results

Development of the SORP

Since a particular time window is required for early forewarning of SS, the longer it takes to collect the necessary data for the model, the less time is left for physicians to intervene. Therefore, we calculated the importance of each feature in the early forewarning of SS in the look-back interval and selected the top-ranked feature set that could be accessed in the same testing batch. In particular, we observed that 8 of the 10 most important clinical features for SS prediction were arterial blood gases (ABGs) and VSs (Figure 2A). Notably, point-of-care ABG testing results could be obtained even within 5-10 minutes in the ICU, and VS features could be obtained at any time from the monitor. This result indicated that septic preshock core pathophysiological shifts could be mainly monitored by ABG and VS features, allowing us to quickly obtain essential features for early forewarning of SS and providing a prompt assessment.

Since features obtained simultaneously from an ABG test might have a certain degree of complementarity, we included 8 ABG and 7 VS features collected in the MIMIC-IV and eICU in the subsequent analysis. An SS risk predictor (SORP) based on a scorecard was constructed by combining LogR with WOE in the look-back interval (detailed in the Methods section). According to the scorecard, a patient's SS risk score could be obtained by adding interval scores to the base score (Multimedia Appendix 2). Lower total scores indicated a higher risk of SS onset. In the training, validation, and test datasets, the AUCs were 0.9557, 0.9537, and 0.9458, respectively (Figure 2B). Meanwhile, based on all available clinical features, multi-feature complex models were constructed, including deep learning and machine learning models (Figure 1A; Multimedia Appendix 3). These features are related to acid-base balance, coagulation, renal aspects, hepatic aspects, pulmonary aspects, hematologic aspects, the Glasgow Coma Scale, etc. The predictive performance of the SORP is only slightly lower than that of the optimal complex model. In the test datasets, the AUCs of the SORP and the optimal complex model were 0.9458 and 0.9651, respectively. It indicated that the construction of the SORP was reasonable and had strong early forewarning ability.

To facilitate observation and evaluation and determine risk intervals, according to the KS curve plot (Figure 2C), the scores were divided into 4 risk groups as detailed in the Statistical Methods subsection. The distribution of SS was highly consistent and stable among the training, validation, and test datasets (PSI <0.01; Figure 2D; Multimedia Appendix 4). In total, the incidences of SS for the HR, MR, LR, and ULR groups in the postonset interval were 88.6% (433/489), 34.5% (262/760), 2.5% (67/2707), and 0.3% (4/1301), respectively (Figure 2D; Multimedia Appendix 5). A total of 88.6% (433/489) of HR patients developed SS, and 90.7% (695/766) of SS patients were classified into the HR or MR group. Further, a curve was fitted for the relationship between the risk score and SS onset rate at 6 hours (Figure 2E; Multimedia Appendix 6), and the result showed a clear stratification and monotonicity, with an increase in the SS onset rate with increasing risk (decreasing risk score). These results suggest that the SORP exhibits a great early forewarning performance.

XSL•FO
RenderX

Figure 2. Construction of the Septic Shock Risk Predictor (SORP) and derivation of patient risk stratification. (A) Shapley Additive Explanations (SHAP) summary plot for the top 10 clinical features contributing to the XGBoost model. The position on the y-axis is determined by the feature, in which "hotpink" represents blood gases and "cyan" represents vital signs, and that on the x-axis is determined by the Shapley value. The color from blue to red represents the feature values from low to high. (B) Receiver operating characteristic curve and area under the curve (AUC) for SORP forewarning performance. (C) Kolmogorov-Smirnov (KS) curve shows the difference between the cumulative proportion of septic shock (SS) patients and the cumulative proportion of septic nonshock (NS) patients as the risk score increases (blue line). (D) Stability of SS distribution across datasets. (E) A fitted curve showing the monotonic relationship between the risk score and SS probability in the postonset interval, according to the locally estimated scatterplot smoothing (LOESS) method. HR: high risk; LR: low risk; MR: medium risk; PSI: population stability index; ULR: ultralow risk.

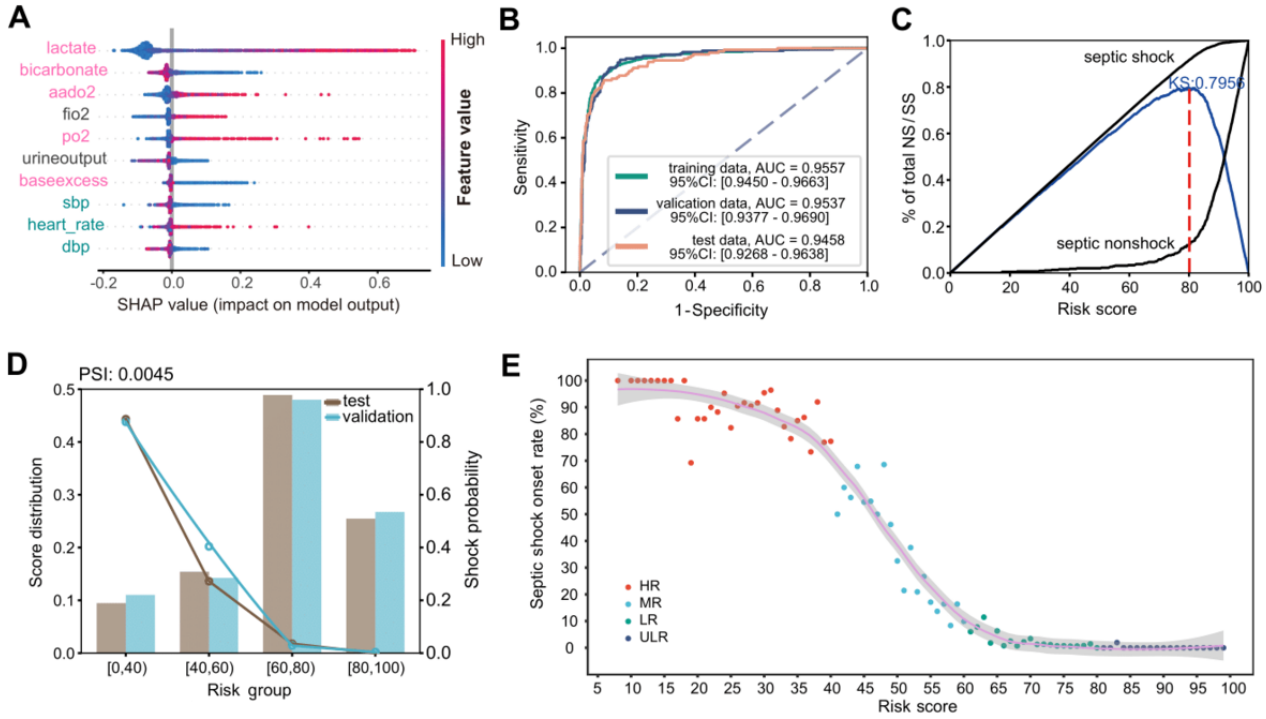

Characteristics of Different Risk Groups After SS Onset

To verify the rationality of risk stratification, we explored the differences in clinical features among risk groups in the postonset interval. Remarkably, increased abnormal values for standardized means of ranking continuous features, especially features related to the SS definition (such as lactate and blood pressure), kidneys, acid-base balance, and coagulation (Figure 3A and 3B; Multimedia Appendix 7), were observed in HR patients when compared with those in other risk groups. The degree of abnormality of these features further increased as risk increased (Figure 3A). For all patients, Kaplan-Meier survival curves progressively declined as patient risk increased (Figure 3C). An even more rapid progressive decline was observed in NS patients in the HR group (Figure 3D), but there was no difference for SS patients (Figure 3E). Furthermore, the length of time from sepsis to SS was significantly reduced with increasing risk (Figure 3F). These results suggest that disease severity increased with increasing risk groups and that the SORP has a reliable risk stratification capability.

XSL•FO
RenderX

Figure 3. Characteristics of different risk groups in the postonset interval. (A) The clinical features are standardized such that all means are scaled to 0 and SDs to 1. A value of 1 for the standardized mean value (y-axis) signifies that the mean value for the risk group was 1 SD higher than the mean value for the 4 risk groups shown in the graph as a whole. (B) Boxplot showing the changing trend for each clinical feature with risk groups. The y-axis shows the standardized value for each clinical feature. (C-F) Kaplan-Meier curves and survival analysis. (C) Overall survival for all patients. (D) Overall survival for septic nonshock patients. (E) Overall survival for septic shock patients in different risk groups. (F) Time from sepsis to septic shock. HR: high risk; LR: low risk; MR: medium risk; ULR: ultralow risk.

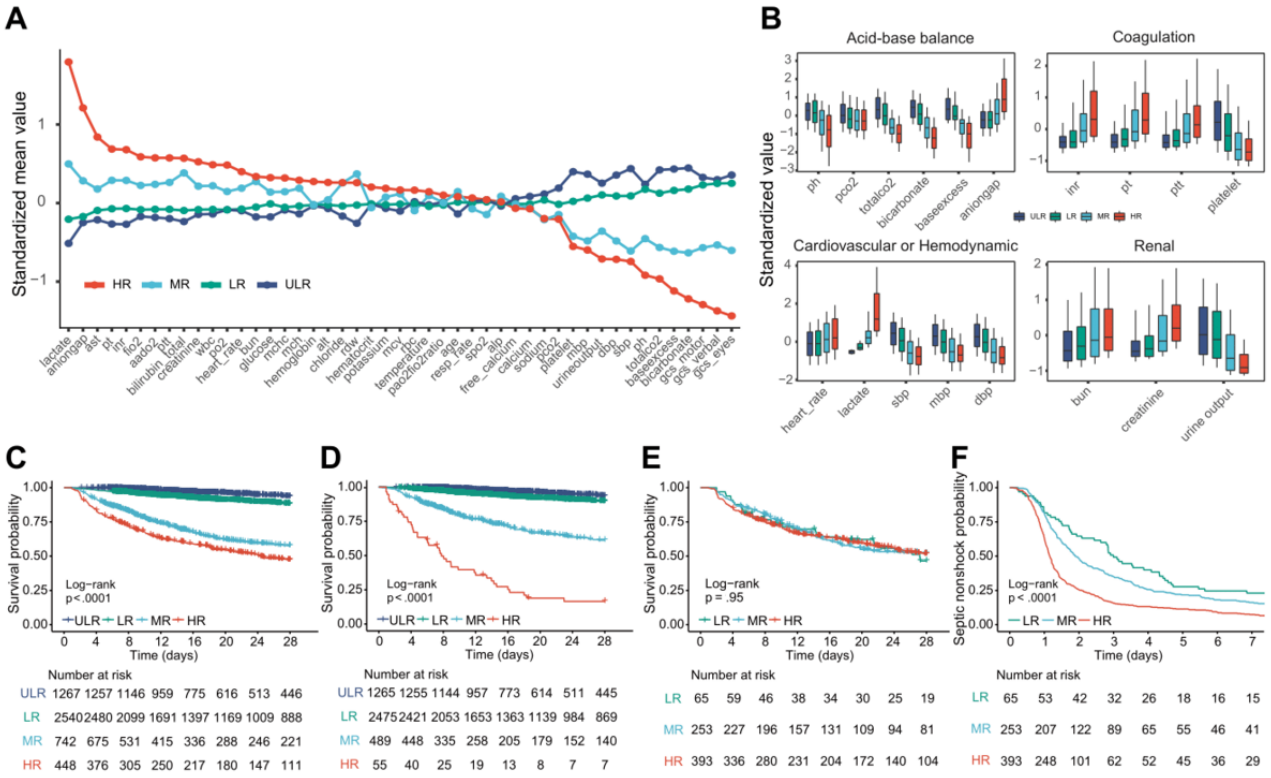

Are NS_HR Patients Truly Septic NS Patients?

Among the HR patients identified by the SORP, 11.5% (56/489) did not meet the SS definition according to the Sepsis-3 criteria [1] and were labeled NS_HR. To further confirm the disease status of these patients, the distribution of features in the postonset interval for NS_HR patients was compared to that for SS and other NS (NS_O) patients. In clustering analysis and heatmap visualization, NS_HR patients clustered with the SS groups but not with the other 3 NS groups (Figure 4A). This result indicates that the NS_HR group was more similar to SS patients in terms of clinical features, such as lactate, the SOFA score, and features related to renal function (creatinine and urine output), coagulation (prothrombin time, international normalized ratio, and platelet count), and acid-base balance (base excess, bicarbonate, total CO_2, and anion gap) (Figure 4B). Compared to NS_O patients, both NS_HR and SS patients displayed a consistently higher degree of variation in clinical features (Figure 4C). The distribution and variation of features demonstrated that both HR_NS and SS patients had similar levels of disorder compared with NS_O patients. Moreover, both NS_HR and SS patients had significantly shorter overall survival than NS_O patients (Figure 4D). In total, NS_HR patients were highly similar to SS patients and distinct from NS_O patients. In addition, we queried the patients' ICD codes, and 35% (19/55) of NS_HR patients were labeled with "septic shock," which implied that a proportion of NS_HR patients might meet the previous criteria for SS.

Notably, the overall survival of NS_HR patients was significantly worse than that of SS patients (Figure 4D), and we explored the potential factors contributing to this survival difference. SS-associated treatments, such as vasopressors and fluid resuscitation, were evaluated. Compared to the SS group, the NS_HR group had a significantly higher incidence of mean blood pressure (MBP) <65 mmHg and a significantly lower rate of vasopressor use and infusion volume (all *P* values <.001; Figure 4E; Multimedia Appendix 8). Furthermore, NS_HR patients showed the highest percentage of CKD (*P*=.006) and the most abnormal kidney-related features (Figure 4E; Multimedia Appendix 8). According to this finding, one of the reasons for high mortality in NS_HR patients might be inadequate treatments, including both vasopressor use and infusion. One of the reasons NS_HR patients did not receive adequate fluid resuscitation might be severe kidney injury, as physicians may reduce fluid infusion in these patients to protect them from further severe kidney damage.

XSL•FO
RenderX

Figure 4. Characteristics of septic nonshock high risk (NS_HR) patients. (A) Clustering heatmap showing the similarity among risk groups. Standardized values were used for clustering (ie, each feature is centered at the sample mean and scaled by its SD), and "Euclidean" and "average" were used for clustering distance and clustering method, respectively. The values in the box are the median raw values of the features. (B) Boxplots showing the distribution of clinical features across risk groups. Red indicates NS_HR patients, orange indicates septic shock (SS) patients, and green indicates septic nonshock (NS) patients. (C) The degree of variation in the features of the risk groups. The y-axis shows the standardized mean for each clinical feature. (D) Overall survival for SS, NS_HR, and NS_O (other risk groups of NS). (E) Bar plot of the polar coordinate showing the proportion of clinical events in each risk group of patients. The boxplot shows the values of kidney-related features during the look-back interval. We transformed the infusion volume (IV), used SS_LR's IV as a reference (3815.25 mL), and expressed the IV in other risk groups as a percentage of that of SS_LR patients. CKD: the rate of patients with chronic kidney disease; HR: high risk; IV: the mean infusion volume of patients from the preonset interval to the postonset interval; LR: low risk; MBP <65: the rate of patients with mean blood pressure (MBP) <65 mmHg from the preonset interval to the postonset interval; MR: medium risk; ns: not significant (P>.05); ULR: ultralow risk; vaso: the rate of patients treated with vasopressors from the preonset interval to the postonset interval. *P≤.05, **P≤.01, ***P≤.001.

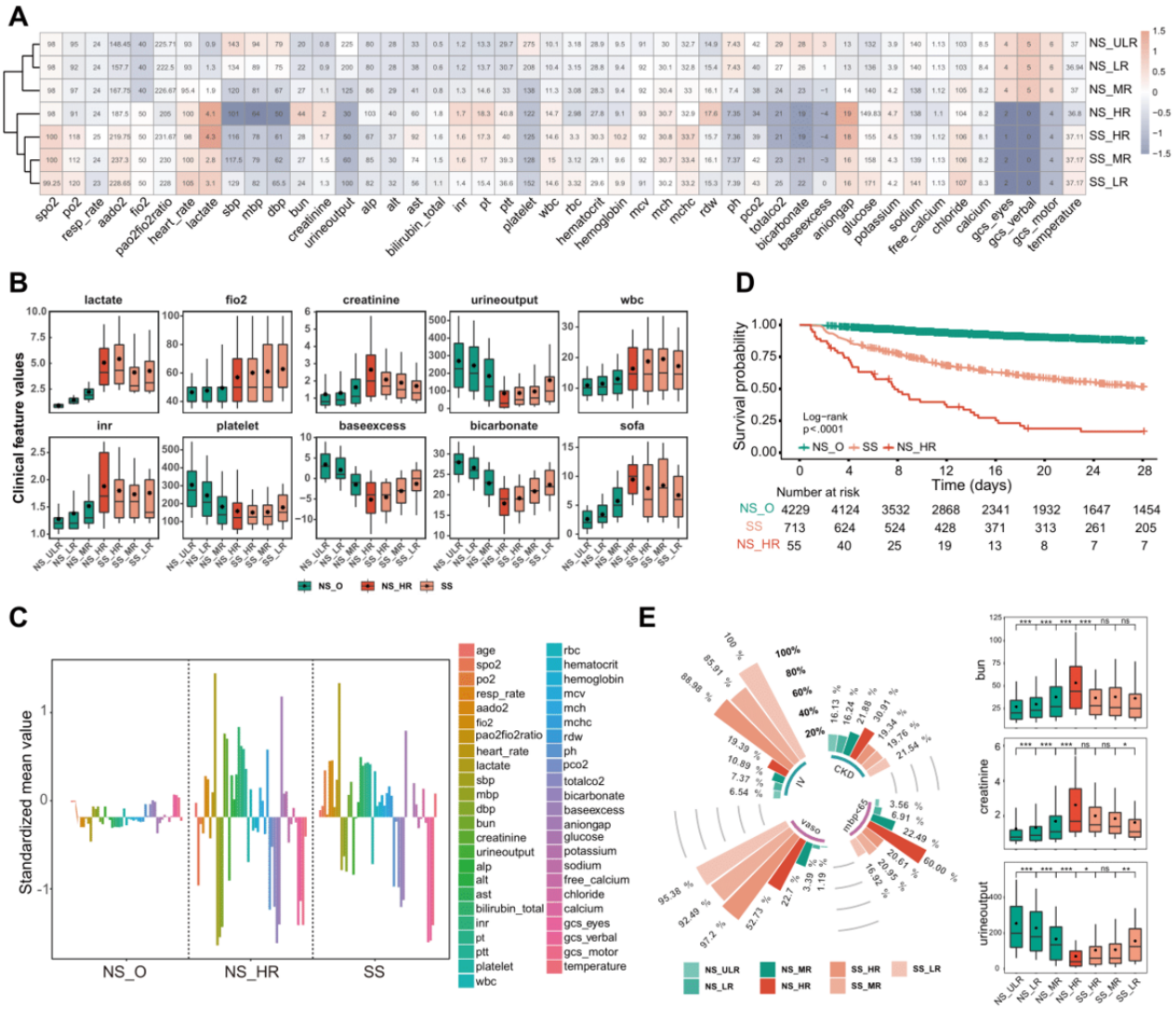

Application of the SORP in Real-Time Risk Monitoring for SS

The SORP showed great performance in sepsis stratification, and we further evaluated its feasibility as a real-time monitoring risk score for patients with sepsis. The hourly risk score for SS onset was calculated by the SORP, and the patient's risk scores and trends throughout the ICU process were fully displayed. Two patients, one with NS and another with SS, are shown in Multimedia Appendix 9. The NS patient had a progressively lower risk (increasing score) nearing discharge. In comparison, the SS patient had an increasingly higher risk until SS onset, with entry into the HR zone 6 hours before onset, and a gradual reduction in risk after onset, which might also reflect the role of treatment. Furthermore, based on continuous real-time scoring, we found that 50.0% (383/766) and 74.9% (574/766) of SS patients with risk scoring entered the HR zone 13 hours and 6 hours before onset (IQR 6-26), respectively. In addition, hourly risk scores were shown for each risk group during the 48-hour period before and after SS onset (Figure 5A). The risk gradually increased (risk score gradually decreased) as septic patients progressed to shock and decreased with treatment intervention (after SS onset), demonstrating that the SORP can monitor the patient's risk for SS in real-time (Figure 5A).

In contrast to the gradual increase in risk in the other 2 groups, SS_LR patients maintained a relatively stable risk during and

XSL•FO
RenderX

before the look-back interval but demonstrated an extremely rapid increase in risk (risk scores decreased) in the preshock interval (Figure 5A). Further analysis demonstrated that SS_LR patients had a higher rate of invasive procedures than other SS patients (Figure 5B; Multimedia Appendix 10), which was significantly higher than that of other LR patients (Figure 5C; Multimedia Appendix 11). Therefore, we speculated that this might be one of the reasons for the rapid progression of SS_LR patients: invasive procedures can introduce additional microorganisms into the body, which can exacerbate the infectious response and lead to rapid deterioration. Analysis showed that 52% (35/67) of SS_LR patients were recognized by the SORP as their risk group increased from LR to MR 3 hours before SS onset. Therefore, we recommend that LR patients exposed to invasive procedures should be assessed more frequently for the timely recognition of SS.

Figure 5. Characteristics of septic shock low risk (SS_LR) patients. (A) Mean value line and unbiased standard error bar for septic shock (SS) patients. Risk groups derived by the Septic Shock Risk Predictor (SORP) 6 hours before SS, and the change in risk over time for each group. The cyan dashed line represents 6 hours before the onset of SS, and the red dashed line represents the onset time of SS. (B) The rate of invasive procedures of each risk group in SS patients. *P* values are derived from SS_LR patients compared to other SS patients (details in Multimedia Appendix 10). (C) The rate of invasive procedures of each risk group in LR patients identified by the SORP. Here, events are measured from 12 hours before the prediction to the time of the prediction. *P* values are derived from SS_LR patients compared to other septic nonshock low risk (NS_LR) patients (details in Multimedia Appendix 11). CRRT: continuous renal replacement therapy; HR: high risk; LR: low risk; MR: medium risk; ns: not significant.

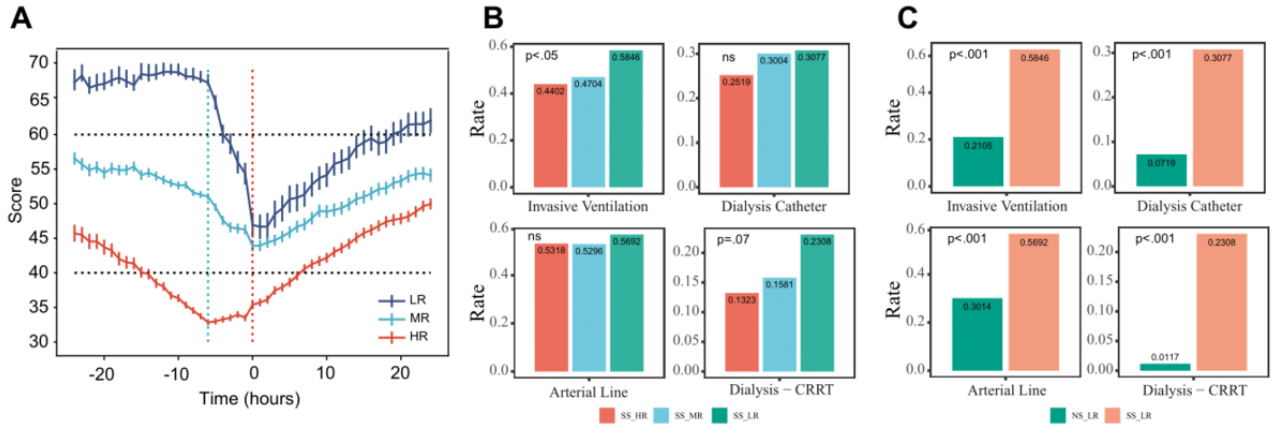

Reliability and Robustness of the SORP

To further verify these findings, we additionally used multicenter eICU data for validation. Consistently, in the look-back interval, eICU patients were grouped into 4 risk groups. Although the proportion of SS patients was reduced in the HR group in the postonset interval (Multimedia Appendix 12), which might be due to different definitions, the distribution of SS was consistent and stable in both the eICU dataset and the MIMIC-IV dataset (PSI <0.05; Multimedia Appendix 13). Similar to the MIMIC-IV, the patients' features became increasingly abnormal as risk increased, especially features related to the SS definition (eg, lactate and blood pressure), kidneys, acid-base balance, and coagulation (Figure 6A; Multimedia Appendix 14). For overall survival, the 4 risk groups also showed significant differences, even in NS patients and SS patients (Figure 6B-D). SS_LR patients, similar to MIMIC-IV patients, underwent significantly more invasive procedures than the other LR patients (Multimedia Appendix 15), but there was no significant difference from the other SS patients (Multimedia Appendix 16). NS_HR patients, similar to MIMIC-IV patients, significantly differed from NS_O patients in clinical features and survival, being more similar to SS patients but with a more severe status (Figure 6E-G). Regarding treatment, as the eICU did not provide start and end times for infusion, we could not obtain accurate infusion volumes for eICU patients within a certain interval, so we focused only on vasopressor use. Compared to SS patients, NS_HR patients showed significant hypotension and less vasopressor use, which were similar to the results of the MIMIC-IV (Figure 6H; Multimedia Appendix 17). Although this analysis was based on a different definition of SS, similar results were still observed, further confirming the robustness of the SORP and the reasonability of the risk stratification system.

In summary, the SORP has a reliable and robust ability to not only provide early forewarning but also stratify patients with potentially different risks of SS. For convenience, we provide an online service of the SORP [28]. The "Help" section in the navigation bar of the website contains detailed instructions on how to use the online tool.

XSL•FO
RenderX

Figure 6. Reproducibility of findings in external independent eICU Collaborative Research Database (eICU) data. (A) The clinical features are standardized with all means scaled to 0 and SDs to 1. A value of 1 for the standardized mean value (y-axis) signifies that the mean value for the risk group was 1 SD higher than the mean value for the 4 risk groups shown in the graph. (B-D) Overall survival for all patients, septic nonshock (NS) patients, and septic shock (SS) patients. (E) Clustering heatmap showing the similarity among risk groups. Standardized values were used for clustering (ie, each feature is centered at the sample mean and scaled by its SD), and "Euclidean" and "average" were used for clustering distance and clustering method, respectively. Red indicates higher levels, and blue indicates lower levels. (F) Characteristics of risk groups. The y-axis shows the standardized mean for each clinical feature. (G) Overall survival for SS, NS_HR, and NS_O (other risk groups of NS). (H) The bar plot of the polar coordinate shows the proportion of clinical events in each risk group of patients. The starting and ending times of patient infusions are not provided in the eICU, and thus, it was not possible to estimate the infusion volume of patients within a specific interval. HR: high risk; LR: low risk; MBP <65: the rate of patients with mean blood pressure (MBP) <65 mmHg from the preonset interval to the postonset interval; MR: medium risk; ULR: ultralow risk; vaso: the rate of patients treated with vasopressors from the preonset interval to the postonset interval.

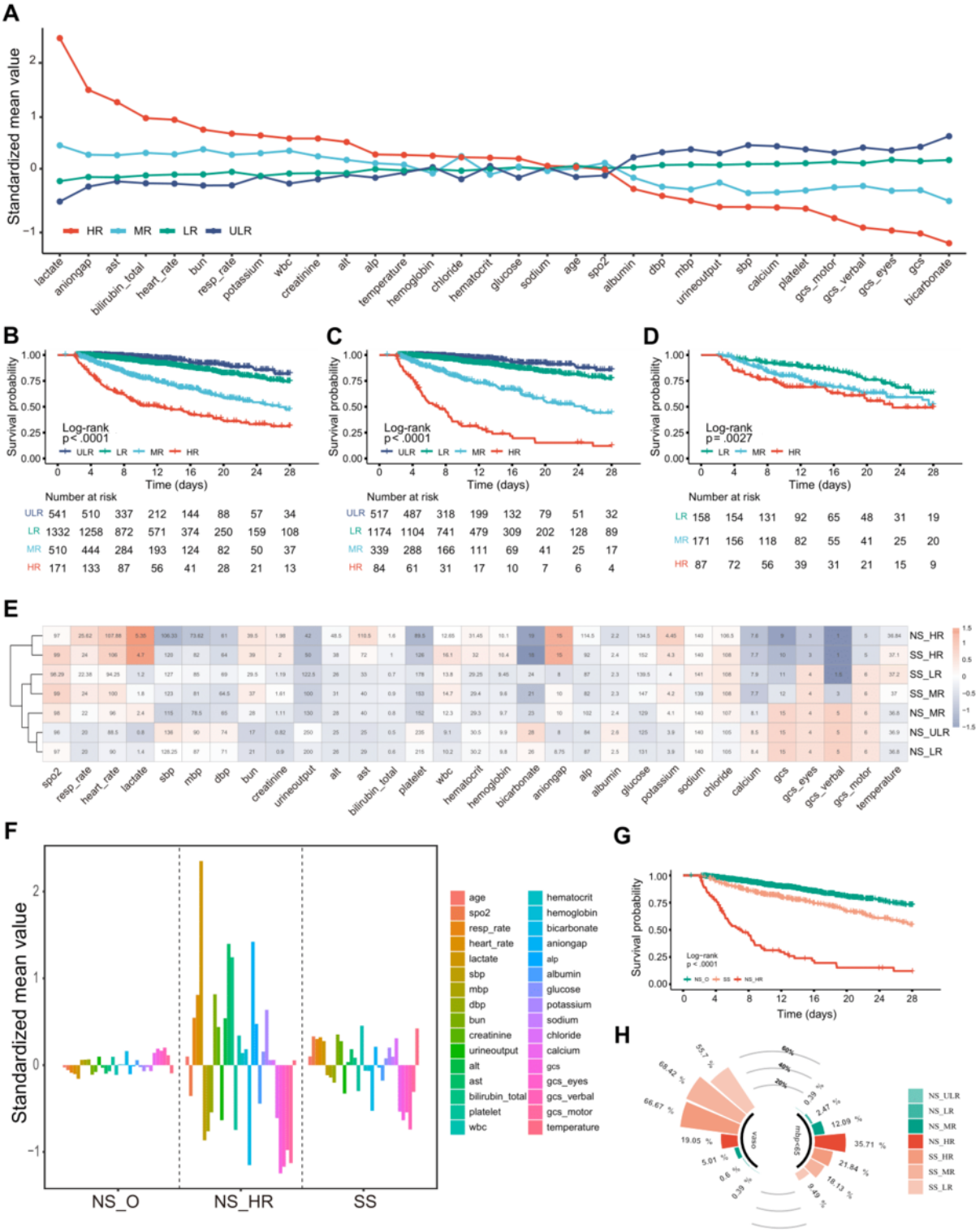

XSL•FO
RenderX

Discussion

Principal Findings

Timely forewarning and early intervention are especially important for patients who are developing SS, and clinically applicable forewarning tools for SS are still lacking. In this study, we constructed the SORP, a simplified model, that not only provides real-time risk stratification for patients but can also easily trace the SS risk of patients during the entire ICU process and forewarn SS onset with a median of 13 hours in advance. The SORP can precisely assess the risk of SS 6 hours in advance by using only ABG and VS features, with an AUC of 0.9458 in the test set. Compared to the findings in previous studies, the SORP remarkably reduces both time and economic costs by using easily available ABG and VS features that can be accessed within minutes and can achieve competitive predictive performance (Multimedia Appendix 18). In contrast, other studies require additional tests, such as biochemical tests, SOFA score calculations, and specific urine output measurements, which extend data collection by several hours. While some studies reduce the number of features through feature engineering methods, these features may still be distributed across test batches, which may not effectively shorten the overall time and costs needed for assessments. Moreover, in addition to being available through the web service [28], the patient's risk score can be simply and quickly calculated manually using a scorecard (Multimedia Appendix 2). The ease of data collection and the simplicity of use will greatly enhance the usefulness of the SORP in clinical applications. Furthermore, we focused more on the clinical characteristics of HR patients, especially septic NS patients, and the assessment of newly discovered NS_HR patients, overlooked by the Sepsis-3 definition, may provide further insights into the pathophysiological processes of SS onset and help to complement its diagnosis and precise management.

In this study, 4 risk groups were identified by the SORP. For these risk groups, we observed distinct clinical features and a significantly stratified risk for SS onset rates and survival status. For SS patients in different risk groups, the time from sepsis onset to SS decreased as their risk increased. Significant survival differences were observed in NS patients across different risk groups, but these same differences were not observed in SS patients (Figure 3E). This may be because of the active guideline-directed resuscitation treatment of SS patients. These findings confirm the effectiveness of the current SS resuscitation treatment guidelines and further affirm the accuracy of risk stratification and the reliability of the SORP.

Interestingly, despite the high agreement between SS patients diagnosed by the Sepsis-3 definition and those identified by the SORP, we still identified 2 relatively distinct groups. One group was SS_LR, where patients were defined as having SS by Sepsis-3 but were identified as LR by the SORP. The overall survival of these patients was not significantly different from that of other SS groups, but rapid progression and an extremely increased risk were demonstrated within 6 hours before SS onset (Figure 5A). In addition, 52% (35/67) of SS_LR patients showed an increase from LR to MR 3 hours before SS onset. Thus, although these patients could not be identified 6 hours in advance, their risk changes were captured in real-time, which highlights the high sensitivity of the SORP. Further analysis showed that SS_LR patients underwent more invasive procedures than other SS patients and much more than other patients in the LR group, such as "invasive ventilation," "dialysis catheter," "arterial line," and "dialysis-CRRT" (Figure 5B and 5C). Thus, we speculate that invasive procedures might be one of the causes of the rapidly exacerbating patient status. We recommend that LR patients undergoing invasive procedures should be evaluated more frequently to recognize rapid exacerbation in a timely manner.

Another notable group was NS_HR, where patients were not defined as having SS by Sepsis-3 but were identified as HR by the SORP. This prompted us to further analyze this particular patient group, and the results showed that these patients had the following characteristics: (1) NS_HR patients were quite similar to SS patients in terms of clinical phenotypes (Figure 4A-C); (2) NS_HR patients had the worst overall survival, which was significantly poorer than even that of SS_HR patients (Figure 3C and 3D; Figure 4D); (3) NS_HR patients did not receive sufficient fluid resuscitation or timely vasopressor administration (Figure 4E), which might be a key factor that resulted in their noncompliance with the SS diagnosis and their higher mortality; and (4) NS_HR patients had more significant abnormalities in renal-related features (Figure 4E), which reflect higher incidences of renal malfunction, such as CKD (Figure 4E). The literature has shown that excessive fluid resuscitation in patients with severe renal disease exacerbates renal injury and leads to poorer outcomes [33,34]. Therefore, cautious and conservative strategies might be adopted by physicians, resulting in a much higher likelihood of insufficient fluid resuscitation in these patients. Hence, precise fluid resuscitation management is crucial in SS patients with renal dysfunction. In total, NS_HR patients share a similar pathophysiologic process with SS patients but experience a more aggressive clinical outcome due to insufficient treatment.

Compared to previous versions, the Sepsis-3 shock definition is more focused on tissue-organ hypoperfusion and metabolic dysfunction and the corresponding therapeutic response, which largely ensures the authenticity and reliability of SS diagnosis, showing a significant increase in specificity but at the cost of sensitivity [35]. Studies have shown that a proportion of patients who met the previous version of the SS criteria, but not the Sepsis-3 shock criteria, had significant organ failure and high mortality [36], with a value of up to 31.6% [37]. In this study, we identified a group of patients (NS_HR) with poor survival and a similar clinical phenotype to SS. In addition to low vasopressor use, the NS_HR group had inadequate fluid resuscitation due to severe renal disease, thus failing the Sepsis-3 shock definition. These patients are missed by the current SS definition, but they are at HR and are worthy of attention. Therefore, for more precise diagnosis and management of SS patients, "adequate fluid resuscitation" in the SS definition may have different criteria in different populations. The need for fluid personalization has emerged as an admirable goal, especially in SS [38].

XSL•FO
RenderX

Limitations

In this study, the designed SS risk early forewarning and stratification model, SORP, can easily and quickly provide ICU physicians with the patient's risk status and its changing trend, which can provide valuable support for the early identification of patient deterioration. However, it should be noted that there are limitations in this study. The current data sources are relatively limited in terms of geographic regions, which to some extent hinders the study of differences in different regions. Future studies should collect data from different regions to test the accuracy of the SORP for early forewarning, analyze the impact of regional differences on risk factors for SS, and establish a more accurate early forewarning system. Moreover, it is crucial to regularly monitor and assess the early forewarning capability and stability of the system and to make necessary adjustments as additional data becomes available. In addition, in the eICU dataset, we were unable to define SS by the Sepsis-3 criteria due to data limitations, which resulted in the incomparability of performance metrics such as AUC. However, based on a different SS definition, we were still able to observe results similar to those with the MIMIC-IV data, which further confirms the robustness of the SORP and the rationality of the risk stratification system.

Summary and Outlook

Early forewarning and diagnosis of SS are essential for reducing mortality. To support this, we developed the SORP early forewarning model, which offers several key benefits: (1) easy and rapid feature collection within minutes; (2) ease of use via online tools or a scorecard (Multimedia Appendix 2); (3) excellent predictive performance (Figure 2B); (4) continuous status monitoring for risk deterioration or improvement; and (5) identification of HR patients, including those with similar symptoms who may not fully meet diagnostic criteria. In clinical practice, the SORP can be easily integrated into sepsis management as an auxiliary monitoring tool (Multimedia Appendix 19). For MR patients whose SORP score is still decreasing, it is recommended to increase the frequency of blood gas testing for timely monitoring. For those identified as HR, the SS determination process can be initiated ahead of time, facilitating the early administration of appropriate treatments, such as fluid resuscitation.

Future research should further explore the applicability of the SORP model in diverse clinical settings, including validation with additional datasets. We have also planned a prospective trial to evaluate its impact on prognosis and survival, particularly in NS_HR patients. These efforts aim to advance early SS identification and management, ultimately improving patient outcomes.

Acknowledgments

This study was supported partly by the Natural Science Foundation of Fujian Province (grant number: 2022J05054, 2023J01321), Fujian Science and Technology Innovation Joint Fund Project (grant number: 2020Y9035), Fujian Key Laboratory of Medical Bioinformatics Open Fund (grant number: FKLMB-202002), and Youth Scientific Research Fund Program of School of Medical Technology and Engineering, Fujian Medical University (2021xy001).

Data Availability

The data generated and analyzed during this study are available from the corresponding author upon request. The electronic health record data used in this study are openly available and can be downloaded at the MIMIC-IV [30] and eICU [39].

Authors' Contributions

GL, JH, TC, SZ, and XS contributed to designing and performing the research, analyzing the data, and writing the manuscript. RW, ZZ, YY, HF, LH, HZ, HS, HY, and YX contributed to performing the research. XS also conceived and supervised the project and was responsible for its overall strategy and management. XS is the guarantor of the manuscript and takes responsibility for the content of the manuscript, including the data and analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist according to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines.
[DOCX File , 53 KB - jmir_v27i1e58779_app1.docx]

Multimedia Appendix 2
Septic shock risk scorecard for the Septic Shock Risk Predictor.
[DOCX File , 52 KB - jmir_v27i1e58779_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Performance of deep learning and machine learning models.
[DOCX File , 59 KB - jmir_v27i1e58779_app3.docx]

Multimedia Appendix 4
Distribution stability of septic shock in Medical Information Mart for Intensive Care-IV (MIMIC-IV) for different risk groups.
[DOCX File , 435 KB - jmir_v27i1e58779_app4.docx]

Multimedia Appendix 5
Medical Information Mart for Intensive Care-IV (MIMIC-IV) data for the rate of patients with septic shock in each risk group.
[DOCX File , 49 KB - jmir_v27i1e58779_app5.docx]

Multimedia Appendix 6
The onset rate of septic shock in each risk score according to the fitted curve.
[DOCX File , 52 KB - jmir_v27i1e58779_app6.docx]

Multimedia Appendix 7
Boxplots showing change trends for each clinical feature by risk groups in Medical Information Mart for Intensive Care-IV (MIMIC-IV) data.
[DOCX File , 883 KB - jmir_v27i1e58779_app7.docx]

Multimedia Appendix 8
Medical Information Mart for Intensive Care-IV (MIMIC-IV) data for the clinical event distribution and significance of risk groups.
[DOCX File , 50 KB - jmir_v27i1e58779_app8.docx]

Multimedia Appendix 9
The changing trends in risk across the whole intensive care unit (ICU) process.
[DOCX File , 312 KB - jmir_v27i1e58779_app9.docx]

Multimedia Appendix 10
Medical Information Mart for Intensive Care-IV (MIMIC-IV) data for the invasive operation distribution and the significance of septic shock risk groups.
[DOCX File , 50 KB - jmir_v27i1e58779_app10.docx]

Multimedia Appendix 11
Medical Information Mart for Intensive Care-IV (MIMIC-IV) data for the invasive operation distribution and the significance of low-risk groups.
[DOCX File , 50 KB - jmir_v27i1e58779_app11.docx]

Multimedia Appendix 12
eICU Collaborative Research Database (eICU) data for the rate of patients with septic shock in each risk group.
[DOCX File , 49 KB - jmir_v27i1e58779_app12.docx]

Multimedia Appendix 13
Distribution stability of septic shock between the eICU Collaborative Research Database (eICU) and Medical Information Mart for Intensive Care-IV (MIMIC-IV) datasets.
[DOCX File , 438 KB - jmir_v27i1e58779_app13.docx]

Multimedia Appendix 14
Boxplots showing change trends for each clinical feature by risk groups in the eICU Collaborative Research Database (eICU) data.
[DOCX File , 584 KB - jmir_v27i1e58779_app14.docx]

Multimedia Appendix 15
eICU Collaborative Research Database (eICU) data for the invasive operation distribution and the significance of low-risk groups.

XSL•FO
RenderX

[DOCX File , 50 KB - jmir_v27i1e58779_app15.docx]

Multimedia Appendix 16
eICU Collaborative Research Database (eICU) data for the invasive operation distribution and the significance of septic shock risk groups.
[DOCX File , 49 KB - jmir_v27i1e58779_app16.docx]

Multimedia Appendix 17
eICU Collaborative Research Database (eICU) data for the clinical event distribution and the significance of risk groups.
[DOCX File , 50 KB - jmir_v27i1e58779_app17.docx]

Multimedia Appendix 18
Current studies for predicting the onset of septic shock.
[DOCX File , 51 KB - jmir_v27i1e58779_app18.docx]

Multimedia Appendix 19
Schematic diagram for auxiliary monitoring of the Septic Shock Risk Predictor (SORP) in clinical practice.
[DOCX File , 144 KB - jmir_v27i1e58779_app19.docx]

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet 2020 Jan 18;395(10219):200-211 [FREE Full text] [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31954465]
3. Kaukonen K, Bailey M, Suzuki S, Pilcher D, Bellomo R. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. JAMA 2014 Apr 02;311(13):1308-1316. [doi: 10.1001/jama.2014.2637] [Medline: 24638143]
4. Daviaud F, Grimaldi D, Dechartres A, Charpentier J, Geri G, Marin N, et al. Timing and causes of death in septic shock. Ann Intensive Care 2015 Dec;5(1):16 [FREE Full text] [doi: 10.1186/s13613-015-0058-8] [Medline: 26092499]
5. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 2006 Jun;34(6):1589-1596. [doi: 10.1097/01.CCM.0000217961.75225.E9] [Medline: 16625125]
6. Vincent J. The clinical challenge of sepsis identification and monitoring. PLoS Med 2016 May 17;13(5):e1002022 [FREE Full text] [doi: 10.1371/journal.pmed.1002022] [Medline: 27187803]
7. Talisa VB, Yende S, Seymour CW, Angus DC. Arguing for adaptive clinical trials in sepsis. Front Immunol 2018;9:1502 [FREE Full text] [doi: 10.3389/fimmu.2018.01502] [Medline: 30002660]
8. de Grooth H, Postema J, Loer SA, Parienti J, Oudemans-van Straaten HM, Girbes AR. Unexplained mortality differences between septic shock trials: a systematic analysis of population characteristics and control-group mortality rates. Intensive Care Med 2018 Mar;44(3):311-322 [FREE Full text] [doi: 10.1007/s00134-018-5134-8] [Medline: 29546535]
9. Yang H, Luo Y, Ma C, Zhang T, Zhou T, Ren X, et al. A gender specific risk assessment of coronary heart disease based on physical examination data. NPJ Digit Med 2023 Jul 31;6(1):136 [FREE Full text] [doi: 10.1038/s41746-023-00887-8] [Medline: 37524859]
10. Yang H, Luo Y, Ren X, Wu M, He X, Peng B, et al. Risk Prediction of Diabetes: Big data mining with fusion of multifarious physical examination indicators. Information Fusion 2021 Nov;75:140-149. [doi: 10.1016/j.inffus.2021.02.015]
11. Zhang Z, Zhang G, Goyal H, Mo L, Hong Y. Identification of subclasses of sepsis that showed different clinical outcomes and responses to amount of fluid resuscitation: a latent profile analysis. Crit Care 2018 Dec 18;22(1):347 [FREE Full text] [doi: 10.1186/s13054-018-2279-3] [Medline: 30563548]
12. Yang T, Chen L, Lin C, Lin H, Fang T, Ge H, et al. Identifying novel clusters of patients with prolonged mechanical ventilation using trajectories of rapid shallow breathing index. Front Med (Lausanne) 2022 Jul 4;9:880896 [FREE Full text] [doi: 10.3389/fmed.2022.880896] [Medline: 35860741]
13. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA 2019 May 28;321(20):2003-2017 [FREE Full text] [doi: 10.1001/jama.2019.5791] [Medline: 31104070]
14. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. Ann Emerg Med 2019 Apr;73(4):334-344. [doi: 10.1016/j.annemergmed.2018.11.036] [Medline: 30661855]

XSL•FO
RenderX

15. Khojandi A, Tansakul V, Li X, Koszalinski RS, Paiva W. Prediction of sepsis and in-hospital mortality using electronic health records. Methods Inf Med 2018 Sep;57(4):185-193. [doi: 10.3414/ME18-01-0014] [Medline: 30248708]
16. Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the MIMIC III database. Comput Biol Med 2019 Oct;113:103395. [doi: 10.1016/j.compbiomed.2019.103395] [Medline: 31480008]
17. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018 Apr;46(4):547-553 [FREE Full text] [doi: 10.1097/CCM.0000000000002936] [Medline: 29286945]
18. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017 Oct 01;89:248-255. [doi: 10.1016/j.compbiomed.2017.08.015] [Medline: 28843829]
19. Henry KE, Adams R, Parent C, Soleimani H, Sridharan A, Johnson L, et al. Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing. Nat Med 2022 Jul;28(7):1447-1454. [doi: 10.1038/s41591-022-01895-z] [Medline: 35864251]
20. Lin C, Zhang Y, Ivy J, Capan M, Arnold R, Huddleston J. Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM. 2018 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); June 04-07, 2018; New York, NY. [doi: 10.1109/ICHI.2018.00032]
21. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015 Aug 05;7(299):299ra122. [doi: 10.1126/scitranslmed.aab3719] [Medline: 26246167]
22. Fagerström J, Bång M, Wilhelms D, Chew MS. LiSep LSTM: A machine learning algorithm for early detection of septic shock. Sci Rep 2019 Oct 22;9(1):15132 [FREE Full text] [doi: 10.1038/s41598-019-51219-4] [Medline: 31641162]
23. Liu R, Greenstein JL, Granite SJ, Fackler JC, Bembea MM, Sarma SV, et al. Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. Sci Rep 2019 Apr 16;9(1):6145 [FREE Full text] [doi: 10.1038/s41598-019-42637-5] [Medline: 30992534]
24. Mollura M, Romano S, Mantoan G, Lehman L, Barbieri R. Prediction of Septic Shock Onset in ICU by Instantaneous Monitoring of Vital Signs. 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, QC, Canada. [doi: 10.1109/embc44109.2020.9176276]
25. Rangan ES, Pathinarupothi RK, Anand KJS, Snyder MP. Performance effectiveness of vital parameter combinations for early warning of sepsis-an exhaustive study using machine learning. JAMIA Open 2022 Dec;5(4):ooac080 [FREE Full text] [doi: 10.1093/jamiaopen/ooac080] [Medline: 36267121]
26. Kim JK, Ahn W, Park S, Lee S, Kim L. Early prediction of sepsis onset using neural architecture search based on genetic algorithms. Int J Environ Res Public Health 2022 Feb 18;19(4):2349 [FREE Full text] [doi: 10.3390/ijerph19042349] [Medline: 35206537]
27. Goh KH, Wang L, Yeow AYK, Poh H, Li K, Yeow JJL, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nat Commun 2021 Jan 29;12(1):711 [FREE Full text] [doi: 10.1038/s41467-021-20910-4] [Medline: 33514699]
28. Septic Shock Risk Predictor (SORP): an online service for early forewarning and risk stratification of septic shock. Department of Bioinformatics, Fujian Medical University. URL: http://www.informatics.icu/SORP/home [accessed 2025-01-09]
29. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378 [FREE Full text] [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
30. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV. PhysioNet. URL: https://physionet.org/content/mimiciv/2.0/ [accessed 2025-01-09]
31. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018 Sep 11;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA. [doi: 10.5555/3295222.3295230]
33. Dodd RL, Archambault ME. Preventing kidney injury by avoiding fluid overload in patients with sepsis. JAAPA 2019 Dec;32(12):40-45. [doi: 10.1097/01.JAA.0000604904.81228.18] [Medline: 31770304]
34. Rice DM, Ratliff PD, Judd WR, Kseibi SA, Eberwein KA. Assessing the impact of CKD on outcomes in septic shock patients receiving standard Vs reduced initial fluid volume. Am J Emerg Med 2020 Oct;38(10):2147-2150. [doi: 10.1016/j.ajem.2020.07.055] [Medline: 33046295]
35. Henning DJ, Puskarich MA, Self WH, Howell MD, Donnino MW, Yealy DM, et al. An emergency department validation of the SEP-3 sepsis and septic shock definitions and comparison with 1992 consensus definitions. Ann Emerg Med 2017 Oct;70(4):544-552.e5 [FREE Full text] [doi: 10.1016/j.annemergmed.2017.01.008] [Medline: 28262318]
36. Sterling SA, Puskarich MA, Glass AF, Guirgis F, Jones AE. The impact of the Sepsis-3 septic shock definition on previously defined septic shock patients. Crit Care Med 2017 Sep;45(9):1436-1442 [FREE Full text] [doi: 10.1097/CCM.0000000000002512] [Medline: 28542029]

XSL•FO
RenderX

37. Vermassen J, Decruyenaere J, De Bus L, Depuydt P, Colpaert K. Characteristics of Sepsis-2 septic shock patients failing to satisfy the Sepsis-3 septic shock definition: an analysis of real-time collected data. Ann Intensive Care 2021 Oct 30;11(1):154 [FREE Full text] [doi: 10.1186/s13613-021-00942-1] [Medline: 34718879]
38. Monnet X, Lai C, Teboul J. How I personalize fluid therapy in septic shock? Crit Care 2023 Mar 24;27(1):123 [FREE Full text] [doi: 10.1186/s13054-023-04363-3] [Medline: 36964573]
39. Pollard T, Johnson A, Raffa J, Celi LA, Badawi O, Mark R. eICU Collaborative Research Database (version 2.0). PhysioNet. URL: https://physionet.org/content/eicu-crd/2.0/ [accessed 2025-01-09]

Abbreviations

ABG: arterial blood gas
AUC: area under the curve
BIDMC: Beth Israel Deaconess Medical Center
CKD: chronic kidney disease
CRRT: dialysis-continuous renal replacement therapy
eICU: eICU Collaborative Research Database
HR: high risk
ICD: International Classification of Diseases
ICU: intensive care unit
IV: information value
KS: Kolmogorov-Smirnov
LogR: logistic regression
LR: low risk
MBP: mean blood pressure
MIMIC-IV: Medical Information Mart for Intensive Care-IV
MR: medium risk
NS: nonshock
NS_O: other nonshock patients, excluding NS_HR
PSI: Population Stability Index
SHAP: Shapley Additive Explanations
SOFA: Sequential Organ Failure Assessment
SORP: Septic Shock Risk Predictor
SS: septic shock
ULR: ultralow risk
VS: vital sign
WOE: weight of evidence

Edited by A Schwartz; submitted 09.04.24; peer-reviewed by BN Beshey, S Liu; comments to author 26.10.24; revised version received 09.11.24; accepted 19.12.24; published 06.02.25.

Please cite as:
Liu G, Zheng S, He J, Zhang ZM, Wu R, Yu Y, Fu H, Han L, Zhu H, Xu Y, Shao H, Yan H, Chen T, Shen X
An Easy and Quick Risk-Stratified Early Forewarning Model for Septic Shock in the Intensive Care Unit: Development, Validation, and Interpretation Study
J Med Internet Res 2025;27:e58779
URL: https://www.jmir.org/2025/1/e58779
doi:10.2196/58779
PMID:

©Guanghao Liu, Shixiang Zheng, Jun He, Zi-Mei Zhang, Ruoqiong Wu, Yingying Yu, Hao Fu, Li Han, Haibo Zhu, Yichang Xu, Huaguo Shao, Haidan Yan, Ting Chen, Xiaopei Shen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions in 3.6 Million Web-Based Posts of Online Communities on the Use of Cancer Immunotherapy: Data Mining Using BERTopic

Xingyue Wu[1*], MSc; Chun Sing Lam[1*], PhD; Ka Ho Hui[1], PhD; Herbert Ho-fung Loong[2], PhD; Keary Rui Zhou[1], PharmD; Chun-Kit Ngan[3], PhD; Yin Ting Cheung[1], PhD

[1]School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

[2]Department of Clinical Oncology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

[3]Data Science Program, Worcester Polytechnic Institute, Worcester, MA, United States

[*]these authors contributed equally

Corresponding Author:
Yin Ting Cheung, PhD
School of Pharmacy, Faculty of Medicine
The Chinese University of Hong Kong
8th Floor, Lo Kwee-Seong Integrated Biomedical Sciences Building, Area 39
Shatin, N.T.
Hong Kong SAR
China
Phone: 852 3943 6833
Email: yinting.cheung@cuhk.edu.hk

Abstract

Background: Immunotherapy has become a game changer in cancer treatment. The internet has been used by patients as a platform to share personal experiences and seek medical guidance. Despite the increased utilization of immunotherapy in clinical practice, few studies have investigated the perceptions about its use by analyzing social media data.

Objective: This study aims to use BERTopic (a topic modeling technique that is an extension of the Bidirectional Encoder Representation from Transformers machine learning model) to explore the perceptions of online cancer communities regarding immunotherapy.

Methods: A total of 4.9 million posts were extracted from Facebook, Twitter, Reddit, and 16 online cancer-related forums. The textual data were preprocessed by natural language processing. BERTopic modeling was performed to identify topics from the posts. The effectiveness of isolating topics from the posts was evaluated using 3 metrics: topic diversity, coherence, and quality. Sentiment analysis was performed to determine the polarity of each topic and categorize them as positive or negative. Based on the topics generated through topic modeling, thematic analysis was conducted to identify themes associated with immunotherapy.

Results: After data cleaning, 3.6 million posts remained for modeling. The highest overall topic quality achieved by BERTopic was 70.47% (topic diversity: 87.86%; topic coherence: 80.21%). BERTopic generated 14 topics related to the perceptions of immunotherapy. The sentiment score of around 0.3 across the 14 topics suggested generally positive sentiments toward immunotherapy within the online communities. Six themes were identified, primarily covering (1) hopeful prospects offered by immunotherapy, (2) perceived effectiveness of immunotherapy, (3) complementary therapies or self-treatments, (4) financial and mental impact of undergoing immunotherapy, (5) impact on lifestyle and time schedules, and (6) side effects due to treatment.

Conclusions: This study provides an overview of the multifaceted considerations essential for the application of immunotherapy as a therapeutic intervention. The topics and themes identified can serve as supporting information to facilitate physician-patient communication and the decision-making process. Furthermore, this study also demonstrates the effectiveness of BERTopic in analyzing large amounts of data to identify perceptions underlying social media and online communities.

(J Med Internet Res 2025;27:e60948) doi:10.2196/60948

XSL•FO
RenderX

KEYWORDS

social media; cancer; immunotherapy; perceptions; data mining; oncology; web-based; lifestyle; therapeutic intervention; leukemia; lymphoma; survival; treatment; health information; decision-making; online community; machine learning

Introduction

In recent years, immunotherapy has emerged as one of the most promising therapeutic approaches for treating cancer. It works by activating the innate immune system to identify and attack cancer cells. Immunotherapy encompasses various strategies, including immune checkpoint inhibitors, T-cell transfer therapy, monoclonal antibodies, treatment vaccines, and immune system modulators [1]. These immunotherapeutic strategies have received approval for the treatment of several cancer types such as lung cancer, prostate cancer, chronic lymphocytic leukemia, and non-Hodgkin lymphoma [2]. Immunotherapy offers substantial benefits in terms of precision, specificity, and long-term survival improvements, representing a significant breakthrough in cancer treatment [3].

Immunotherapy has made remarkable progress and demonstrated clinical value. However, there are drawbacks that may hinder its clinical use and acceptance by patients with cancer. One notable limitation is the variability in individuals' responses to immunotherapy. Although the treatment may be effective in some patients with specific types of cancers, it may not be as effective for others with the same cancer types [4]. Besides, immunotherapy-related adverse events (irAEs) have been observed with the increasing frequency and duration of immunotherapy usage. Patients' decisions to undergo immunotherapy are influenced by a range of factors, including their perceptions of its efficacy, side effects, procedural aspects, costs, their levels of knowledge about the treatment, and the comprehensiveness of advice provided by health care providers [5,6]. These findings highlight the importance of understanding patients' perspectives regarding immunotherapy to manage the uncertainties faced by patients and to make informed decisions on whether to proceed with immunotherapy.

The internet has become an indispensable source of health information for patients [7]. Many patients scour the internet for medical guidance, share their personal experiences, and interact on social media platforms [8]. The internet, therefore, provides a valuable platform for capturing diverse perspectives. Several studies have explored these perspectives by collecting data from social media platforms and forums [9-13]. For example, through analyzing posts from social media, a study revealed that pain and fatigue were the most commonly discussed symptoms among patients with non–small-cell lung cancer regarding the use of immunotherapy [10]. Another study found that 55% of patients' posts and 37% of caregivers' posts expressed positive perceptions of immunotherapy for treating advanced bladder cancer [11]. Notably, these previous studies usually extracted posts from a single site and analyzed only a small portion of the available posts. Moreover, these studies primarily focused on patients with specific cancer types, and thus, the findings may not be generalizable to patients with other cancer types for which immunotherapy has been approved [12].

One challenge in analyzing data from the internet and social media posts is dealing with vast amounts of unstructured text, which necessitates extensive preprocessing measures prior to analysis. Different machine learning techniques have been explored to increase the efficiency of online text analysis. BERTopic, a topic modeling technique that is an extension of the Bidirectional Encoder Representation from Transformers (BERT) machine learning model, has been used to identify the underlying public perceptions within social media posts. It has demonstrated the capability to unveil latent patterns and extract topics from large datasets [14-16]. Compared to other models, BERT models provide a stronger understanding of the contextual meaning of each word in document representations, which is attributable to their bidirectional training of transformers [17]. This enables the generation of more accurate topics than those generated using traditional statistical models. Moreover, the BERT model is continuously trained and updated by researchers, ensuring ongoing improvements [18]. However, no study has employed this technique to explore the viewpoints of the online cancer community on immunotherapy.

With the objective of understanding the perspectives of online cancer communities regarding immunotherapy, this study used BERTopic to analyze a large number of posts extracted from multiple social media and online forums. The findings of this study may offer valuable insights for clinicians, patients, and researchers to enhance decision-making processes when considering immunotherapy as a treatment option.

Methods

Overview

This was a retrospective study that involved analyzing texts collected from social media platforms and online forums. The study proceeded through 4 primary methodological phases: (1) collecting textual data, (2) cleaning and processing the extracted texts, (3) performing and optimizing topic modeling, and (4) conducting thematic analysis. The study process workflow is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart depicting the data mining process. BERTopic: a topic modeling technique that is an extension of the Bidirectional Encoder Representation from Transformers; cTF-IDF: class-based term frequency-inverse document frequency.

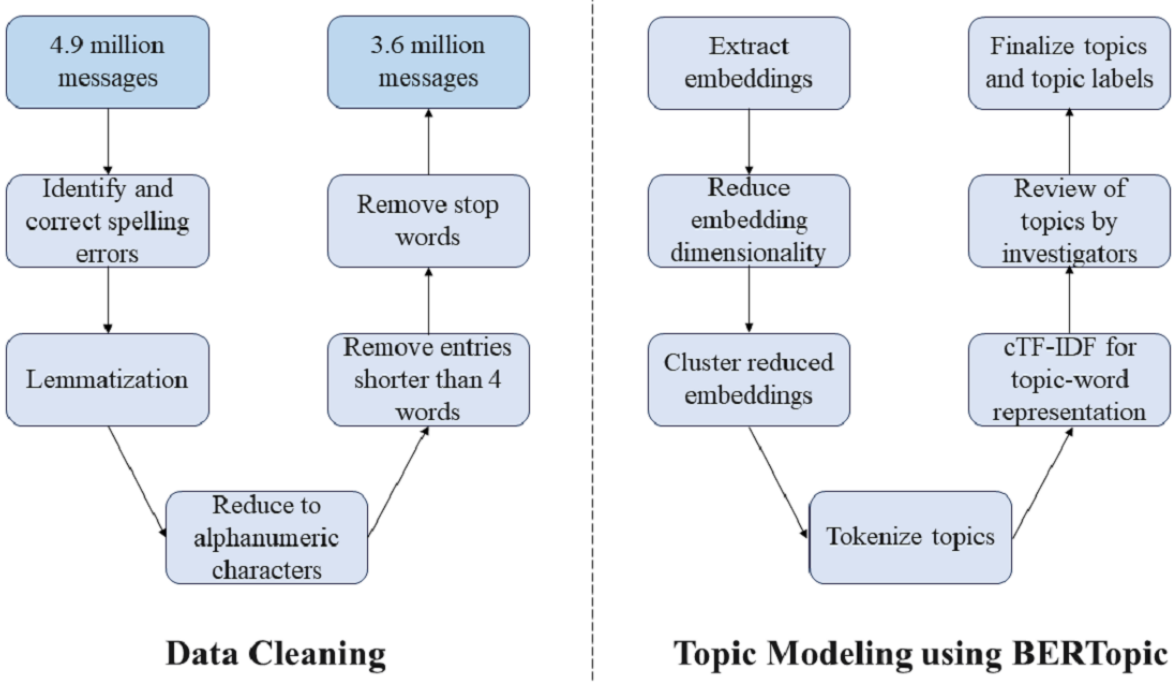

Ethics Approval

This study exclusively utilizes anonymized data from publicly available sources. The quotes presented in this report are modified or paraphrased to prevent any potential identification and direct linkage to the original posts/quotes. All identifiers or pseudo identifiers have been removed from the report. This study was approved by the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong (reference SBRE-20-864).

Data Sources and Collection

For this study, textual data were collected from online forums and social media platforms. For consistency and minimal translation concerns, only English-language texts were included. The posts from these platforms dated before November 15, 2022, were included, with varying start dates. For social media platforms, posts were extracted from Facebook, Twitter, and Reddit. These platforms were chosen for their popularity and the substantial textual content they contain [19]. For cancer-related online forums, the selection criteria have been detailed in a previous study [20]. To describe briefly, we first used the top search engines (Google, Microsoft Bing, Yahoo, and Yandex) to search for health forums by using different combinations of search terms (cancer, tumor/tumour, patient, and forum) [21]. The search results were extracted and examined by the investigators. Web forums were included if they (1) appeared in more than one search result, (2) were open access (ie, no membership or passwords were needed to access the messages), (3) were active for at least the past 5 years, (4) had at least 10 messages posted to the group within the past 30 days from the date of the search, and (5) enabled web scraping of posts or feeds using Python (Python Software Foundation) or R (R Foundation for Statistical Computing). Finally, 16 different online cancer-related forums, including Cancer Chat, Cancer Survivors Network, Cancer Council Online Community, and other forums, were included for the analysis (Table 1). Search terms related to immunotherapy were identified from multiple authoritative websites, including the American Cancer Society, Cancer Research UK, and the National Cancer Institute (Table S1 in Multimedia Appendix 1). Site-provided search engines were utilized to identify related posts.

XSL•FO
RenderX

Table 1. List of the online health forums selected for data extraction [22-37].

Name of web forum	Forum start date[a]
Cancer Chat [22]	2008
Jo's Cervical Cancer Trust Forum [23]	2004
Prostate Cancer UK [24]	2014
Bowel Cancer UK Community [25]	2015
Cancer Survivors Network [26]	2000
Breastcancer.org community [27]	2004
Pancreatic Cancer UK Forum [28]	2007
Breast Cancer Now Forum [29]	2005
Cancer Council Online Community [30]	2009
Irish Cancer Society Community [31]	2008
Navigating Care [32]	2009
TC Cancer.com [33]	2004
HealthBoards (cancer) [34]	2000
Melanoma Patient Forum [35]	2010
Macmillan Cancer Support [36]	2008
HealthUnlocked [37]	2013

[a]In cases where the start dates of certain online health forums were not explicitly provided, an estimation was made based on the published date of the first post available.

Data Preprocessing

Spark NLP, a natural language processing library for Python built on top of Apache Spark, was utilized to preprocess the textual data [38]. The texts extracted from the forums and social media sites were subjected to a thorough cleaning process to enhance their suitability for further analysis. The cleaning process involved multiple steps, including checking and correcting spelling errors, lemmatization, removal of stop words, and elimination of short entries (Figure 1). Each text message was first spell-checked, and the vocabulary was reduced to its base root form by using Spark NLP. After that, each word vocabulary was reduced to alphanumeric characters that contained only letters from “a to z,” “A to Z,” and “0 to 9.” For a precise and semantically meaningful text corpus, any message composed of fewer than 4 words as well as commonly occurring words (eg, the, a, and, in) that carried little or no meaning in the message were removed. Finally, the cleaned text messages were processed using the medical language models from Healthcare Spark NLP, a pretrained pipeline to recognize medical terminologies within the messages [39].

Topic Modeling

The next step involved the extraction of topics from the data. Topic modeling, an unsupervised machine-learning approach, was employed to unveil hidden semantic structures and extract distinct topics from the extracted textual data [40]. In this study, we first evaluated 10 common topic modeling techniques, namely, Latent Dirichlet Allocation, Latent Semantic Analysis, Non-Negative Matrix Factorization, Principal Component Analysis, Random Project, K-Means, Top2Vec, BERT + K-Means, BERT, and Latent Dirichlet Allocation + BERT to determine the appropriate topic modeling approach. Topic diversity (indicating the model’s ability to capture differences between generated topics), coherence (the frequency of descriptive words of the topic within each cluster), and quality (topic diversity multiplied by topic coherence) were used as metrics to evaluate the effectiveness of isolating topics from the texts [41,42]. Finally, we used BERTopic to extract quality topics from the data based on the modelling performance. All texts were passed through the BERT model to create embeddings that converted text messages into numerical representations by using BERTopic. The dimensions of the embedding vectors were reduced due to the “Curse of Dimensionality” [43]. The bisecting K-means algorithm was used to group the reduced-dimension embedding vectors into clusters until optimal silhouette scores were achieved [44]. Various numbers of topics, ranging from 5 to 25, were tested to determine the optimal metrics.

Following the identification of clusters with the highest topic quality, the 20 most frequent words from each topic were extracted. BERTopic modeling was conducted using the Worcester Polytechnic Institute’s Turing cluster, which has high computing power for large volumes of data processing by using multiple physical computers simultaneously. The Turing cluster consists of a 4-node hyperconverged head that controls 79 compute nodes and is located at Worcester Polytechnic Institute. The total central processing unit/random access memory/graphics processing unit counts across all computer nodes were 5224, 49 TB, and 84, respectively [45].

XSL•FO
RenderX

Sentiment and Thematic Analysis

Sentiment analysis was performed using 2 popular libraries, VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob [46,47]. Both libraries were used to enhance the accuracy of the results and minimize potential biases that may arise from relying on a single library. The sentiment scores obtained from each library were then averaged to derive a final sentiment score, which ranged from –1 (indicating a very negative sentiment) to 1 (indicating a very positive sentiment). Based on the topics generated through BERTopic modeling, thematic analysis was conducted by 2 investigators (XW and CSL) to identify themes associated with immunotherapy. The thematic analysis in this study involved the following steps: familiarization with the posts, inductive analysis, theme generation, and theme review among domain experts in the research team (including oncologists and oncology pharmacists) [48]. First, 2 investigators (XW and CSL) familiarized with the posts and encoded them independently. Then, the codes were compared, the same codes were merged, and any discrepancies were addressed through discussion with a third investigator (YTC). The themes were generated through the identification and refinement of the codes, which were grouped by perceived thematic similarities. The final themes were identified by achieving consensus among domain experts. The consensus process involved multiple rounds. The experts rated their level of agreement for each theme: themes with 75% agreement were retained, and themes without consensus were discussed, adjusted, and rated in the next round until 75% agreement was reached.

Results

Overview

A total of 4.9 million posts were gathered from online forums and social media platforms. After data cleaning, 3.6 million posts remained for further analysis (Figure 1).

Number and Quality of the Topics

In comparison with other topic modeling techniques, BERTopic demonstrated the most satisfactory performance, with a topic diversity score of 81.11%, a topic coherence score of 80.48%, and a topic quality score of 65.28% (Table S2 in Multimedia Appendix 1). Table S3 in Multimedia Appendix 1 presents BERTopic's performance across different numbers of topics to determine the optimal value. Fifteen topics achieved the highest overall topic quality, scoring 70.47% (with the highest topic diversity at 87.86% and a high topic coherence of 80.21%). We also used these metrics to assess the quality of individual topics generated by the model, yielding topic quality scores ranging from 78% to 88% (Figure S1 in Multimedia Appendix 2). One topic (Topic 15) was excluded due to the mismatch between its extracted words and the topics related to immunotherapy. Based on the sentiment analysis results, the average sentiment score of around 0.3 across the 14 topics preliminarily suggested generally positive sentiments toward immunotherapy within the online community (Figure 2). A more comprehensive interpretation of the specific topics is presented in the subsequent sections.

Figure 2. Sentiment analysis results. Sentiment score (a measure of how positive or negative the tone of the quotes is) was analyzed for each topic. The sentiment score ranged from –1 (indicating a very negative sentiment) to 1 (indicating a very positive sentiment). N/A: not applicable; T: topic.

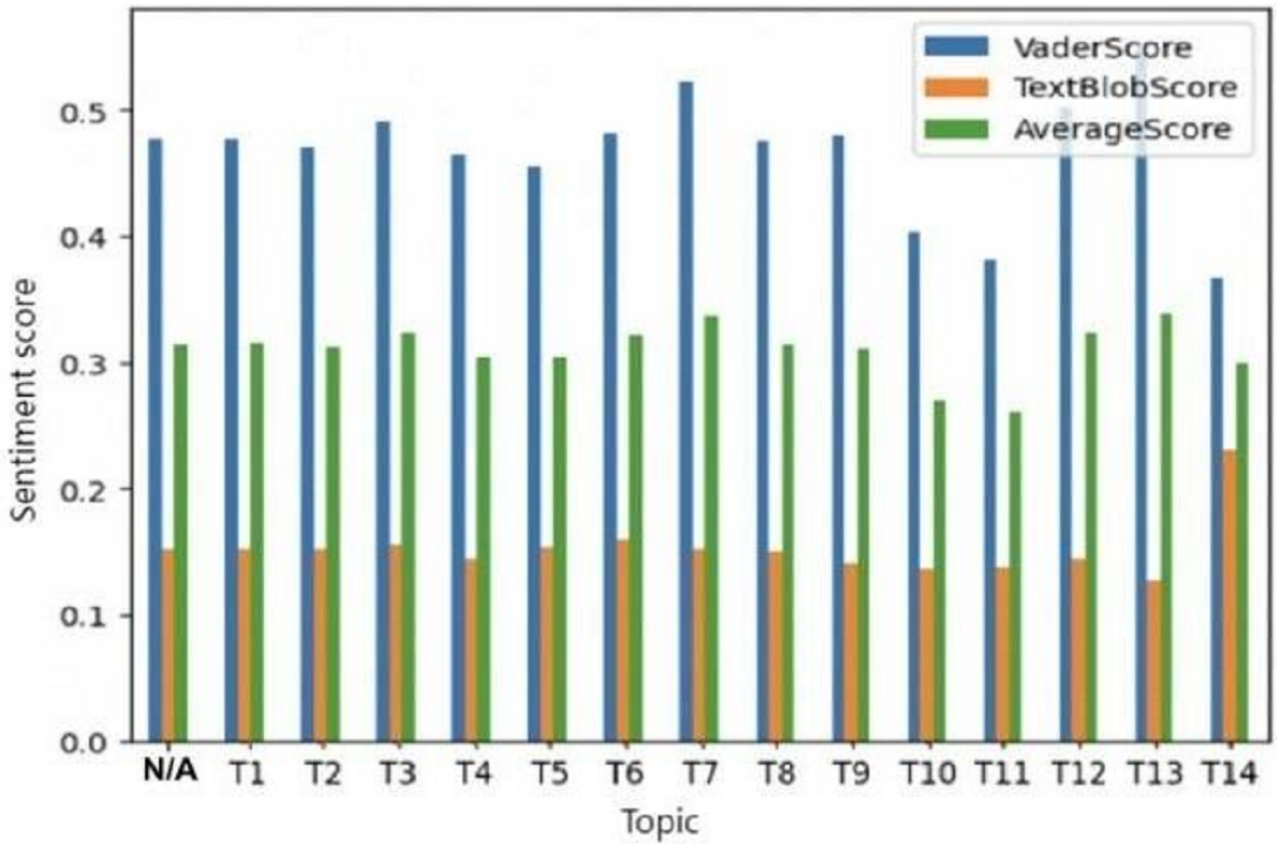

XSL•FO
RenderX

Themes and Topics

A total of 6 themes were identified (Table 2). The quotes in Table 2 have been modified to prevent any potential identification and direct linkage to the original posts. Theme 1 centered on the feelings of hope and positivity among patients treated with immunotherapy. This theme highlighted the belief that immunotherapy offers hope for patients with cancer to lead normal lives and provides a potential cure for cancer with minimal side effects. Theme 2 described immunotherapy as a more effective treatment option after other treatments have failed. Patients may have tried other cancer treatments such as chemotherapy and radiotherapy but only found successful results with immunotherapy. Besides, patients believe combining immunotherapy with other conventional therapies, particularly chemotherapy, can lead to higher efficacy in treating cancer than using a single treatment approach. Theme 3 suggested that patients receiving immunotherapy also explore diets or complementary medicines to manage their symptoms or side effects from treatment or the cancer itself. For example, patients may employ mind-body practices such as massage and meditation or use herbal medicines to alleviate certain physical and psychological side effects caused by immunotherapy. Probiotics were frequently mentioned by users as a potential way to boost the effectiveness of immunotherapy. In Theme 4, the potential ineffectiveness, worse outcomes, and high cost of treatment were frequently mentioned among patients, which acted as hindrances to the use of immunotherapy. Theme 5 suggested that immunotherapy can disrupt patients' normal daily lives, particularly for those enrolled in immunotherapy trials. Additionally, the treatment may negatively impact their activities of daily living, such as taste changes, dietary restrictions, and time consumption. Theme 6 indicated that immunotherapy may cause physical and psychological symptoms and side effects such as rash, inflammation, and nephrotoxicity. These adverse effects may sometimes pose greater challenges compared to those arising from other treatments or the cancer itself. Some patients acknowledged that these symptoms may be unavoidable alongside the curative effects of immunotherapy.

XSL•FO
RenderX

Table 2. Topics and selected posts under each theme.

Theme, topics and related keywords	Samples of modified quotes[a]
Theme 1: Hopeful prospects offered by immunotherapy	
Topic 1: good, time, hope, love, great, feel, work, lot, treatment, start	• …help not to feel alone during treatment. • …hope stop the growth of cancer quickly. • …the new treatment offers new hope.
Topic 3: fight, hope, positive, news, treatment, glad, beat, good, feel, life	• …immunotherapy through clinical trials has shown positive results. • …wish to continue with immunotherapy after completing treatment. • …use the inert immune system to combat cancer cells. • …reduced damage compared with chemotherapy or radiotherapy.
Topic 5: hope, day, love, morning, huge, satisfy, glad, feel, good, love	• …hope of one day curing all cancers without side effects. • …develop immunotherapy tailored to the individual. • …the day when immunotherapy will be available to all. • …eradicate unique cancer cells
Topic 6: saint, faith, family, love, god, strength, miracle, wonderful, hope, glad	• …a miracle for patients to keep the cancer at bay. • …a miracle to help someone with very advanced diseases.
Theme 2: Perceived effectiveness of immunotherapy	
Topic 4: gallbladder, liver, bile, symptom, surgery, scan, remove, problem, ultrasound, test	• …a tremendous response with significant improvements. • …the tumors on my adrenal glands, lungs and liver have shrunk after immunotherapy. • …tried immunotherapy as one last hail Mary pass and it seems to be working.
Topic 13: risk, study, benefit, favorable, evidence, effect, patient, data, high, treat	• …immunotherapy appears to be successful in that the CT scan shows the cancer has shrunk. • …have a favorable response against the melanoma. • …immunotherapy has given lives back.
Topic 8: antioxidant, radiation, chemo, protect, treatment, dose, effect, radical, kill, interfere	• …better results with less side effects. • …combine chemo with immunotherapy gives the potential for long term remission and even a cure for advanced cancers. • …very effective in treating liver tumors.
Theme 3: Complementary therapies or self-treatments	
Topic 2: acupuncture, pain, feel, needle, massage, neuropathy, flash, hot, session, side	• …encouraged to get massage and acupuncture. • …had acupuncture to help with side effects of immunotherapy, such as abdominal pain, muscle pain, and severe neuropathy.
Topic 10: keto, diet, food, organic, probiotic, supplement, eat, weight, sugar, body	• …dependent on good gut bacteria. • …anyone undergoing immunotherapy is recommended to take probiotics concurrently. • …tried anything else, eg, diet, nutrition, supplements, etc.
Theme 4: Financial and mental impact of immunotherapy	
Topic 7: diagnosis, feel, anxiety, traumatic, stress, cost, experience, time, depression, therapist	• …expensive drugs with nasty side effects. • …immunotherapy costs a lot that can’t afford it. • …dealing with progression is profoundly stressful. • …more depressed and weaker when waiting long for treatment.
Theme 5: Impact on lifestyle and time schedules	
Topic 9: simulation, schedule, treatment, appointment, ill, week, start, time, plan, session	• …involve an eight-week program after the previous treatments. • …take a break and/or reduce dosages rather than quit altogether. • …spent 5 weeks to decide on a new treatment plan. • …don’t have too much flexibility.
Theme 6: Side effects due to treatment	
Topic 11: ill, swell, weak, surgery, compression, feel, wear, inflame, tight, back	• …surprised at how the side effects returned so quickly. • …resulted in inflammation of lungs. • …the treatments can result in worst side effects than the disease itself.

XSL•FO
RenderX

Theme, topics and related keywords	Samples of modified quotes[a]
Topic 12: immune, damage, skin, fever, burn, kill, system, therapy, tissue, body	• …developed an acute kidney injury. • …had a very hard time with auto immune reaction • …had several other side effects, but not a rash.
Topic 14: pain, symptom, hope, painful, colitis, arthritis, response, issue, experience, costochondritis	• …back pain mentioned as a side effect. • …have a rash on the back of neck and serious neuropathy. • …found a cream that helps along with pain meds.

[a]The quotes have been modified or paraphrased to prevent any potential identification and direct linkage to the original posts.

Discussion

Principal Findings

We analyzed a large dataset comprising 3.6 million posts from 3 social media platforms and 16 online cancer forums, providing a comprehensive overview of perceptions regarding immunotherapy within English-speaking online cancer communities. A cutting-edge natural language processing technique, BERTopic, was used to generate interpretable topics, ensuring robust data analysis. The identification of different themes underscores the diverse attitudes toward immunotherapy across various aspects. The posts show the positive attitudes of the patients toward immunotherapy and its perceived effectiveness and benefits. However, there are also concerns about the side effects of immunotherapy and potential disruptions to patients' lifestyle. These themes may offer valuable insights into the facilitators and barriers to the adoption of immunotherapy. In addition, the results showed that using machine learning techniques to analyze patient-generated online textual data can enhance our understanding of perspectives on newer cancer treatments, and such information may facilitate communication between patients and clinicians when discussing immunotherapy as a therapeutic option.

Comparison to Current Evidence

The positive factors that encourage the utilization of immunotherapy included its comparatively mild side effect profile and the potential to instill hope in patients by offering a viable treatment option when other methods have failed. Our findings also indicated that online cancer communities seem to be well-informed, with some patients well-versed in the benefits of immunotherapy. Laypeople now access medical information on the internet independently and acquire firsthand information about advances in cancer treatment. Access to online health information assists patients in making health-related decisions, leading to more professional consultation, improved treatment compliance, and better self-care [49]. However, the internet should not be viewed as a replacement for professional health information sources [50]. The guidance and advice of health care professionals remain indispensable in helping patients make complex medical decisions [51]. Online health information cannot provide detailed diagnoses or personalized treatment plans for patients; moreover, the presence of misinformation on websites might mislead patients about treatment [52]. Taken together, our results highlight the rising health literacy of online cancer communities, especially those involving younger generations. Future efforts should focus on enhancing the accuracy of online health information by using reliable rating tools, effective search engine ranking, and progress in crowdsourcing websites.

Although immunotherapy has shown clinical benefits in several trials, there are still potential barriers to its utilization. The high cost associated with immunotherapy emerges as a frequently discussed concern among patients with cancer, emphasizing the need for revising drug valuations and reimbursement models [53]. Many quotes were related to patients' descriptions of their physical symptoms induced by immunotherapy and how it has impeded their activities of daily living and disrupted their normal lives. In addition to physical irAEs, the long-term use of immunotherapy has also been associated with psychiatric side effects such as fatigue, insomnia, anxiety, and depression [54,55]. Hence, it is important for oncology practitioners to manage patients' expectations and communicate with patients regarding the potential side effects of immunotherapy before initiating treatment. This proactive approach allows patients to better understand the benefits and risks of the treatment, which may prevent them from discontinuing treatment because of known adverse reactions.

Relevance to Clinical Practice

Several clinical practice guidelines have been proposed to address the evaluation and management of immunotherapy-induced toxicity [56] and the diagnosis and treatment of toxicity associated with immune checkpoint inhibitor therapy in specific organ systems [57]. The provision of high-quality supportive care from a multidisciplinary team of health care professionals can help identify and manage irAEs [58]. In practice, patients may turn to self-management strategies to cope with these symptoms. We found that some patients mentioned in online posts that they used traditional, complementary, and integrative medicine modalities such as probiotics, herbal remedies, and meditation to self-manage the toxicity induced by immunotherapy. Recently, some studies have evaluated how complementary medicines may enhance the efficacy of immunotherapy and alleviate immune toxicity [59,60]. For example, the treatment of irAEs often necessitates the long-term use of high-dose corticosteroids [61]. Integrating complementary modalities into the management of irAEs may reduce corticosteroid use. Our group has previously reported that patients with cancer frequently seek advice on the use of complementary medicines on social media platforms [20]. At this juncture, clinicians should focus on initiating effective communication regarding the use of complementary medicines and help patients establish realistic expectations while being well-informed about the limited evidence supporting these approaches.

XSL•FO
RenderX

Limitations

Our study has several limitations. First, the data used in this research were collected from English-language online cancer communities, which may limit the generalizability of the findings to data posted in other languages. Social media and cancer forums may be used distinctly in different countries, and the types of immunotherapeutic approaches may also vary among regions. Future studies could conduct a thorough investigation into whether distinct concerns regarding immunotherapy exist among patients from different countries and regions of practice. Second, perceptions regarding immunotherapy on social media and cancer forums are dynamic and may be influenced by current news or events. Therefore, future research can consider exploring the fluctuations in posts on these platforms over varying time periods. Third, although the performance metrics (topic diversity and topic coherence scores) quantified from our models suggest that BERTopic is effective at unveiling latent patterns within large datasets, this approach requires validation. Nevertheless, our study demonstrates that applying artificial intelligence to analyze social media data can facilitate our understanding of patients' perspectives as well as the effectiveness of machine-learning techniques in processing substantial volumes of data retrieved from social media platforms and online forums.

Conclusions

This study captures comprehensive insights and perspectives of online communities regarding immunotherapy through BERTopic modeling. The identified themes and topics may facilitate the understanding of immunotherapy utilization and serve as valuable supporting information when making treatment decisions. The posts have shown the positive attitudes of the patients and their perceived effectiveness and benefits of immunotherapy. Although immunotherapy presents a beacon of hope and a viable treatment option, its side effects and the accompanying potential lifestyle disruptions cannot be ignored and are major concerns perceived by the online communities. It is essential for clinicians to inform patients of the undesirable effects of immunotherapy and establish realistic expectations, thereby enabling patients to make an informed decision considering the benefits and risks associated with the treatment, which may ultimately lead to more optimal treatment outcomes with immunotherapy.

Acknowledgments

We would like to thank Worcester Polytechnic Institute's Turing cluster hosted by the Worcester Polytechnic Institute for the implementation of the computing process.

Data Availability

The dataset used in this study is accessible via the following link [62].

Authors' Contributions

XW and CSL conceptualized this study and contributed to the methodology, formal analysis, writing the original draft, review, and editing. KHH, HHFL, and KRZ contributed to conceptualization and review and editing of this paper. Ngan CK contributed to conceptualization, formal analysis, review, and editing. YTC contributed to conceptualization, formal analysis, supervision, review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional data.
[DOCX File , 27 KB - jmir_v27i1e60948_app1.docx]

Multimedia Appendix 2
The distribution of topic quality scores.
[PNG File , 249 KB - jmir_v27i1e60948_app2.png]

References

1. Immunotherapy to treat cancer. National Cancer Institute. 2019. URL: https://www.cancer.gov/about-cancer/treatment/types/immunotherapy [accessed 2024-03-04]
2. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 2022;72(5):409-436 [FREE Full text] [doi: 10.3322/caac.21731] [Medline: 35736631]
3. Tan S, Li D, Zhu X. Cancer immunotherapy: Pros, cons and beyond. Biomed Pharmacother 2020;124:109821 [FREE Full text] [doi: 10.1016/j.biopha.2020.109821] [Medline: 31962285]
4. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature 2017;541(7637):321-330. [doi: 10.1038/nature21349] [Medline: 28102259]

XSL•FO
RenderX

5. Bien DR, Danner M, Vennedey V, Civello D, Evers SM, Hiligsmann M. Patients' preferences for outcome, process and cost attributes in cancer treatment: a systematic review of discrete choice experiments. Patient 2017;10(5):553-565 [FREE Full text] [doi: 10.1007/s40271-017-0235-y] [Medline: 28364387]
6. Ihrig A, Richter J, Bugaj TJ, Friederich H, Maatouk I. Between hope and reality: How oncology physicians and information providers of a cancer information service manage patients' expectations for and experiences with immunotherapies. Patient Educ Couns 2023;109:107622. [doi: 10.1016/j.pec.2023.107622] [Medline: 36641334]
7. Hesse BW, Greenberg AJ, Rutten LJF. The role of internet resources in clinical oncology: promises and challenges. Nat Rev Clin Oncol 2016;13(12):767-776. [doi: 10.1038/nrclinonc.2016.78] [Medline: 27273045]
8. Iftikhar R, Abaalkhail B. Health-seeking influence reflected by online health-related messages received on social media: cross-sectional survey. J Med Internet Res 2017;19(11):e382 [FREE Full text] [doi: 10.2196/jmir.5989] [Medline: 29146568]
9. Jenei K, Burgess M, Peacock S, Raymakers AJN. Experiences and perspectives of individuals accessing CAR-T cell therapy: A qualitative analysis of online Reddit discussions. J Cancer Policy 2021;30:100303. [doi: 10.1016/j.jcpo.2021.100303] [Medline: 35559799]
10. Booth A, Manson S, Halhol S, Merinopoulou E, Raluy-Callado M, Hareendran A, et al. Using health-related social media to understand the experiences of adults with lung cancer in the era of immuno-oncology and targeted therapies: observational study. JMIR Cancer 2023;9:e45707 [FREE Full text] [doi: 10.2196/45707] [Medline: 37436789]
11. Renner S, Loussikian P, Foulquié P, Marrel A, Barbier V, Mebarki A, et al. Patient and caregiver perceptions of advanced bladder cancer systemic treatments: infodemiology study based on social media data. JMIR Cancer 2023;9:e45011 [FREE Full text] [doi: 10.2196/45011] [Medline: 36972135]
12. Rodrigues A, Chauhan J, Sagkriotis A, Aasaithambi S, Montrone M. Understanding the lived experience of lung cancer: a European social media listening study. BMC Cancer 2022;22(1):475 [FREE Full text] [doi: 10.1186/s12885-022-09505-4] [Medline: 35490223]
13. Chang A, Xian X, Liu MT, Zhao X. Health communication through positive and solidarity messages amid the COVID-19 pandemic: automated content analysis of Facebook uses. Int J Environ Res Public Health 2022;19(10):6159 [FREE Full text] [doi: 10.3390/ijerph19106159] [Medline: 35627696]
14. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. ArXiv 2022:1. [doi: 10.48550/arXiv.2203.05794]
15. Ng QX, Lee DYX, Yau CE, Lim YL, Liew TM. Public perception on 'healthy ageing' in the past decade: An unsupervised machine learning of 63,809 Twitter posts. Heliyon 2023;9(2):e13118 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e13118] [Medline: 36747557]
16. Ng QX, Yau CE, Lim YL, Wong LKT, Liew TM. Public sentiment on the global outbreak of monkeypox: an unsupervised machine learning analysis of 352,182 twitter posts. Public Health 2022;213:1-4 [FREE Full text] [doi: 10.1016/j.puhe.2022.09.008] [Medline: 36308872]
17. Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. : Human Language Technologies; 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7; Minneapolis, Minnesota URL: https://aclanthology.org/N19-1423/
18. Nayak P. Understanding searches better than ever before. Google. URL: https://blog.google/products/search/search-language-understanding-bert/ [accessed 2024-03-25]
19. Wagner J. The most popular social networks of 2021. Ignite Social Media. URL: https://www.ignitesocialmedia.com/twitter-marketing/the-most-popular-social-networks-of-2021/ [accessed 2023-09-28]
20. Lam CS, Zhou K, Loong HH, Chung VC, Ngan C, Cheung YT. The use of traditional, complementary, and integrative medicine in cancer: data-mining study of 1 million web-based posts from health forums and social media platforms. J Med Internet Res 2023;25:e45408 [FREE Full text] [doi: 10.2196/45408] [Medline: 37083752]
21. Search engine market share worldwide. Statcounter GlobalStats. URL: https://gs.statcounter.com/search-engine-market-share [accessed 2024-03-14]
22. Cancer chat. Cancer Research UK. URL: https://www.cancerresearchuk.org/about-cancer/cancer-chat [accessed 2024-01-13]
23. Jo's Cervical Cancer Trust Forum. URL: https://forumstaging.jostrust.org.uk [accessed 2024-01-13]
24. Prostate Cancer UK. URL: https://community.prostatecanceruk.org/alltopics [accessed 2024-01-13]
25. Bowel Cancer UK. URL: https://community.bowelcanceruk.org.uk/forum/ [accessed 2024-01-13]
26. Cancer Survivors Network. URL: https://csn.cancer.org [accessed 2024-01-13]
27. The Breastcancer.org Community Forum. URL: https://community.breastcancer.org/ [accessed 2024-01-13]
28. Pancreatic Cancer UK. URL: https://forum.pancreaticcancer.org.uk/ [accessed 2024-01-13]
29. Breast Cancer Now Forum. URL: https://forum.breastcancernow.org/ [accessed 2024-01-13]
30. Cancer Council Online Community. URL: https://onlinecommunity.cancercouncil.com.au/ [accessed 2024-01-13]
31. Irish Cancer Society Community. URL: https://www.cancer.ie/community [accessed 2024-01-13]
32. Navigating Care. URL: https://www.navigatingcare.com [accessed 2024-01-13]
33. Testicular Cancer Society. URL: http://www.tc-cancer.com/forum/# [accessed 2024-01-13]
34. HealthBoards (cancer). URL: https://www.healthboards.com/boards/cancers/ [accessed 2024-01-13]

XSL•FO
RenderX

35. Melanoma Patient Forum. URL: https://forum.melanoma.org/ [accessed 2024-01-13]
36. Macmillan Cancer Support. URL: https://community.macmillan.org.uk/g [accessed 2024-01-13]
37. HealthUnlocked. URL: https://healthunlocked.com/search/communities?query=cancer [accessed 2024-01-13]
38. Kocaman V, Talby D. Spark NLP: Natural language understanding at scale. Software Impacts 2021;8:100058. [doi: 10.1016/j.simpa.2021.100058]
39. Health care NLP state of the art medical language models. John Snow Labs. URL: https://www.johnsnowlabs.com/health [accessed 2024-03-06]
40. Saracco BH. Data science and predictive analytics: biomedical and health applications using R. Journal of the Medical Library Association 2020;108(2):334. [doi: 10.5195/jmla.2020.901]
41. Dieng AB, Ruiz FJR, Blei DM. Topic modeling in embedding spaces. Transactions of the Association for Computational Linguistics 2020;8:439-453. [doi: 10.1162/tacl_a_00325]
42. Mimno D, Wallach H, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. 2011 Presented at: Proceedings of the conference on empirical methods in natural language processing; July 27-31; Edinburgh, Scotland, UK URL: https://aclanthology.org/D11-1024
43. Grootendorst M. BERTopic. 2022. URL: https://github.com/MaartenGr/BERTopic [accessed 2023-09-01]
44. Krishna BSV, Satheesh P, SuneelKumar R. Comparative study of k-means and bisecting k-means techniques in wordnet based document clustering. Semantic Scholar. 2012. URL: https://api.semanticscholar.org/CorpusID:212499046 [accessed 2024-03-13]
45. High performance computing. Worcester Polytechnic Institute. URL: https://arc.wpi.edu/computing/hpc-clusters/ [accessed 2024-06-17]
46. Bonta V, Kumaresh N, Janardhan N. A comprehensive study on lexicon based approaches for sentiment analysis. AJCST 2019;8(S2):1-6. [doi: 10.51983/ajcst-2019.8.s2.2037]
47. Aljedaani W, Rustam F, Mkaouer MW, Ghallab A, Rupapara V, Washington PB, et al. Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. Knowledge-Based Systems 2022;255:109780. [doi: 10.1016/j.knosys.2022.109780]
48. Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, Almario CV, et al. Using large-scale social media analytics to understand patient perspectives about urinary tract infections: thematic analysis. J Med Internet Res 2022;24(1):e26781. [doi: 10.2196/26781]
49. Thapa DK, Visentin DC, Kornhaber R, West S, Cleary M. The influence of online health information on health decisions: A systematic review. Patient Educ Couns 2021;104(4):770-784. [doi: 10.1016/j.pec.2020.11.016] [Medline: 33358253]
50. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
51. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019;21(4):e12887. [doi: 10.2196/12887] [Medline: 30950796]
52. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
53. Tran G, Zafar SY. Financial toxicity and implications for cancer care in the era of molecular and immune therapies. Ann Transl Med 2018;6(9):166 [FREE Full text] [doi: 10.21037/atm.2018.03.28] [Medline: 29911114]
54. Kovacs D, Kovacs P, Eszlari N, Gonda X, Juhasz G. Psychological side effects of immune therapies: symptoms and pathomechanism. Curr Opin Pharmacol 2016;29:97-103 [FREE Full text] [doi: 10.1016/j.coph.2016.06.008] [Medline: 27456240]
55. Lacouture M, Sibaud V. Toxic side effects of targeted therapies and immunotherapies affecting the skin, oral mucosa, hair, and nails. Am J Clin Dermatol 2018;19(S1):31-39. [doi: 10.1007/s40257-018-0384-3]
56. Haanen J, Obeid M, Spain L, Carbonnel F, Wang Y, Robert C, ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Management of toxicities from immunotherapy: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2022;33(12):1217-1238 [FREE Full text] [doi: 10.1016/j.annonc.2022.10.001] [Medline: 36270461]
57. Brahmer JR, Lacchetti C, Schneider BJ, Atkins MB, Brassil KJ, Caterino JM, National Comprehensive Cancer Network. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2018;36(17):1714-1768 [FREE Full text] [doi: 10.1200/JCO.2017.77.6385] [Medline: 29442540]
58. Rahman MM, Behl T, Islam MR, Alam MN, Islam MM, Albarrati A, et al. Emerging management approach for the adverse events of immunotherapy of cancer. Molecules 2022;27(12):3798 [FREE Full text] [doi: 10.3390/molecules27123798] [Medline: 35744922]
59. Zhang N, Xiao X. Integrative medicine in the era of cancer immunotherapy: Challenges and opportunities. J Integr Med 2021;19(4):291-294. [doi: 10.1016/j.joim.2021.03.005] [Medline: 33814325]
60. Jia W, Wang L. Using Traditional Chinese Medicine to treat hepatocellular carcinoma by targeting tumor immunity. Evid Based Complement Alternat Med 2020;2020:9843486 [FREE Full text] [doi: 10.1155/2020/9843486] [Medline: 32595757]

XSL•FO
RenderX

61. Schneider BJ, Naidoo J, Santomasso BD, Lacchetti C, Adkins S, Anadkat M, et al. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: ASCO guideline update. JCO 2021;39(36):4073-4126. [doi: 10.1200/jco.21.01440]
62. Cheung YT. Perception of online communities towards the use of cancer immunotherapy: a data mining study of 3.6 million web-based posts from social media platforms using BERTopic. The Chinese University of Hong Kong. URL: https://researchdata.cuhk.edu.hk/dataset.xhtml?persistentId=doi:10.48668/0NPBOK [accessed 2025-01-21]

Abbreviations

BERT: Bidirectional Encoder Representation from Transformers
BERTopic: a topic modeling technique that is an extension of the Bidirectional Encoder Representation from Transformers
irAE: immunotherapy-related adverse event
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by N Cahill; submitted 27.05.24; peer-reviewed by D Carvalho, L Wreyford; comments to author 27.09.24; revised version received 17.10.24; accepted 30.12.24; published 10.02.25.

Please cite as:
Wu X, Lam CS, Hui KH, Loong HHF, Zhou KR, Ngan CK, Cheung YT
Perceptions in 3.6 Million Web-Based Posts of Online Communities on the Use of Cancer Immunotherapy: Data Mining Using BERTopic
J Med Internet Res 2025;27:e60948
URL: https://www.jmir.org/2025/1/e60948
doi:10.2196/60948
PMID:

©Xingyue Wu, Chun Sing Lam, Ka Ho Hui, Herbert Ho-fung Loong, Keary Rui Zhou, Chun-Kit Ngan, Yin Ting Cheung. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Automated Process for Monitoring of Amiodarone Treatment: Development and Evaluation

Birgitta I Johansson[1,2], MD, PhD; Jonas Landahl[3], PhD; Karin Tammelin[1,4,5,6], MD; Erik Aerts[7], MSc; Christina E Lundberg[1,8], PhD; Martin Adiels[9], PhD; Martin Lindgren[1,2], MD, PhD; Annika Rosengren[1,2], MD, PhD; Nikolaos Papachrysos[1,2], MD, PhD; Helena Filipsson Nyström[4,5,6,10], MD, PhD; Helen Sjöland[1,2], MD, PhD

[1]Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[2]Department of Medicine, Geriatrics and Emergency Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden
[3]Department of Digital Development, Sahlgrenska University Hospital, Gothenburg, Sweden
[4]Department of Endocrinology, Sahlgrenska University Hospital, Gothenburg, Sweden
[5]Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden
[6]Wallenberg Centre for Molecular and Translational Medicine, Gothenburg, Sweden
[7]Chalmers University of Technology, Gothenburg, Sweden
[8]Department of Food and Nutrition, and Sport Science, Faculty of Education, University of Gothenburg, Gothenburg, Sweden
[9]School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden
[10]Gothenburg's Centre for Person-Centred Care, University of Gothenburg, Gothenburg, Sweden

Corresponding Author:
Helen Sjöland, MD, PhD
Department of Molecular and Clinical Medicine
Sahlgrenska Academy
University of Gothenburg
Diagnosvägen 11
Gothenburg, 41685
Sweden
Phone: 46 763202101
Fax: 46 31196167
Email: helen.sjoland@vgregion.se

Abstract

Background: Amiodarone treatment requires repeated laboratory evaluations of thyroid and liver function due to potential side effects. Robotic process automation uses software robots to automate repetitive and routine tasks, and their use may be extended to clinical settings.

Objective: Thus, this study aimed to develop a robot using a diagnostic classification algorithm to automate repetitive laboratory evaluations for amiodarone follow-up.

Methods: We designed a robot and clinical decision support system based on expert clinical advice and current best practices in thyroid and liver disease management. The robot provided recommendations on the time interval to follow-up laboratory testing and management suggestions, while the final decision rested with a physician, acting as a human-in-the-loop. The performance of the robot was compared to the existing real-world manual follow-up routine for amiodarone treatment.

Results: Following iterative technical improvements, a robot prototype was validated against physician orders (n=390 paired orders). The robot recommended a mean follow-up time interval of 4.5 (SD 2.4) months compared to the 3.1 (SD 1.4) months ordered by physicians (P<.001). For normal laboratory values, the robot recommended a 6-month follow-up in 281 (72.1%) of cases, whereas physicians did so in only 38 (9.7%) of cases, favoring a 3- to 4-month follow-up (n=227, 58.2%). All patients diagnosed with new side effects (n=12) were correctly detected by the robot, whereas only 8 were by the physician.

Conclusions: An automated process, using a software robot and a diagnostic classification algorithm, is a technically and medically reliable alternative for amiodarone follow-up. It may reduce manual labor, decrease the frequency of laboratory testing, and improve the detection of side effects, thereby reducing costs and enhancing patient value.

(J Med Internet Res 2025;27:e65473) doi:10.2196/65473

XSL•FO
RenderX

KEYWORDS

thyroid function; robotics; follow-up studies; disease management; decision support; automated process; monitoring; amiodarone treatment; anti-arrhythmic medication; anti-arrhythmic; development; evaluation; thyroid; liver; side effects; cardiac dysrhythmias; ventricular tachycardia; ventricular fibrillation; arrhythmia; automation; robot; algorithm; clinical decision support system; thyroid gland; heart; atrial fibrillation

Introduction

Amiodarone is a potent antiarrhythmic drug widely used for both severe supraventricular and ventricular arrhythmias. However, careful follow-up is necessary during treatment due to potentially serious adverse effects and toxic reactions in the thyroid gland, liver, lung, and heart. Approximately 15% of patients will experience adverse effects during the first year of amiodarone use and 50% during long-term treatment, most often thyroid and liver dysfunction [1]. Reported frequencies of amiodarone-induced hypothyroidism (AIH) and amiodarone-induced thyrotoxicosis (AIT) vary between 1%-32% and 0.8%-37.8%, respectively [2-4], with hypothyroidism typically occurring sooner than hyperthyroidism. A careful study showed median time-to-onset at 183 days for AIH and 720 days for AIT [2]. In addition, AIT is often more resistant to treatment due to the complex effects of amiodarone on the thyroid involving hyperthyroidism, thyroiditis, and iodine overload. Hepatotoxicity, with slight elevations of serum transaminase levels, occurs in about 25% of patients whereas hepatitis, cirrhosis, or hepatic failure occurs in less than 3% [5,6].

Amiodarone carries electrophysiological characteristics of all Vaughan Williams classes, although most importantly class 3, which consists of prolongation of the third phase of the action potential in myocardial cells [7]. It is an iodinated benzofuran derivative that is structurally similar to thyroid hormones. It contains about 37% organic iodine, resulting in 40 to 100-fold iodine excess into the systemic circulation [4,8], overloading the thyroid. Amiodarone accumulates in adipose tissue, due to its high lipid solubility, and in highly perfused organs such as the liver, lung, and skin [6]. These properties of amiodarone contribute to the considerable excess risk of primarily thyroid and liver side effects. Additionally, due to the prolonged biological half-life of amiodarone, up to 40-142 days, toxicity can occur in the thyroid and lung even after the treatment is stopped, thus requiring follow-up after discontinuation of therapy [3,8].

Monitoring recommendations during amiodarone treatment according to the North American Society of Pacing and Electrophysiology include thyroid and liver function tests at baseline and every 6 months, supplemented with recurrent radiographic and functional testing [9]. The American Thyroid Association recommends analysis of thyroid-stimulating hormone (TSH) and thyroxine levels at baseline and thereafter every 3-6 months [10]. The most recent updated consensus management guide recommendations, by Epstein et al [1], are in accordance with the North American Society of Pacing and Electrophysiology, proposing follow-up every 3-6 months for the first year and every 6 months thereafter.

Repeated notifications to patients on amiodarone for necessary follow-ups with laboratory testing, examinations, evaluations, and prescriptions for continued medication is a time-consuming and repetitive process. To ensure that no patients are missed, clinical controls with patient-doctor meetings and systematic laboratory follow-ups of side effects are organized independently. Thus, if patients fail to appear or postpone doctor's meetings they will still be on the administrative waiting list for amiodarone laboratory controls and be regularly summoned for testing. The overall administrative workload for health care professionals is considerable, contributing to workplace stress and a shortage of available personnel. The development of streamlined automated work processes is therefore an attractive option for amiodarone treatment follow-up.

In this study, we present a model for the automation of repetitive follow-up of patients with amiodarone treatment. First, we constructed an automated solution including the whole surveillance process, using a robotic process automation (RPA) software tool. We then compared the RPA to standard manual amiodarone follow-up as presently practiced.

Methods

Patients

All patients on, or who recently finished amiodarone treatment (N=198), followed at the outpatient clinic of the Department of Cardiology, Sahlgrenska University Hospital/Östra for surveillance, were included in this study. Clinical background data were collected from patient records and stratified by the type of arrhythmia constituting indication for therapy. All patients received annual in-office appointments with a cardiologist for their clinical condition. Complementary follow-up with laboratory testing owing to the amiodarone treatment was carried out according to routine, irrespective of clinical follow-up. Before the start of the therapy, the standard clinical routine at Sahlgrenska University Hospital/Östra was performed, including the following: thyroid function tests with TSH, free thyroxine, free tri-iodothyronine, liver function tests for aspartate aminotransferase, alanine aminotransferase, ECG, chest x-ray, and a pulmonary function test, including diffusion capacity. After initiation of amiodarone treatment, updated thyroid and liver laboratory blood tests were performed at 3 months, 6 months, and thereafter routinely every 6 months as per local guidelines [1,10], or at the discretion of the cardiologist. The routine further prescribed a complementary test of TSH receptor antibodies or thyroid peroxidase antibodies when a thyroid function test was abnormal. Patients who developed AIT were referred to an endocrinologist, whereas patients who presented with AIH were referred to primary care. Side effects already present at baseline as well as those sustained during follow-up were recorded.

XSL•FO
RenderX

Process Mapping in the Design of the Robot

To define the actions that should be included in the RPA procedure, all steps of the laboratory thyroid and liver function surveillance process were mapped by a multiprofessional team (specialized cardiology nurse, cardiologist, administrative personnel, and solution architect; March 2021) and presented in Figure 1. In brief, the manual existing routine in use included the following: (1) patients were manually identified by secretarial personnel from an administrative listing ("amiodarone controls"); (2) orders for blood tests were issued and a time slot for sampling was sent by post to the patients; (3) arrival of completed laboratory analyses were manually monitored from patient records by nurses, and notifications sent to a cardiology specialist; (4) after reviewing the laboratory results the cardiologist decided and recorded the future dose of amiodarone, time interval to next follow-up and possible actions depending on side effects; and (5) the orders were manually transferred to a nurse for execution and notification of the patient. For absent laboratory results the nurse issued a reminder to the patient. This process was typically repeated 2-4 times a year per patient. This routine was maintained during the whole study and included all physicians' orders assembled in the following analysis, that is, the manual routine constituted the real-world follow-up, and was compared to the robot.

Figure 1. Process map of established manual workflow and steps of the laboratory amiodarone surveillance process. lab: laboratory.

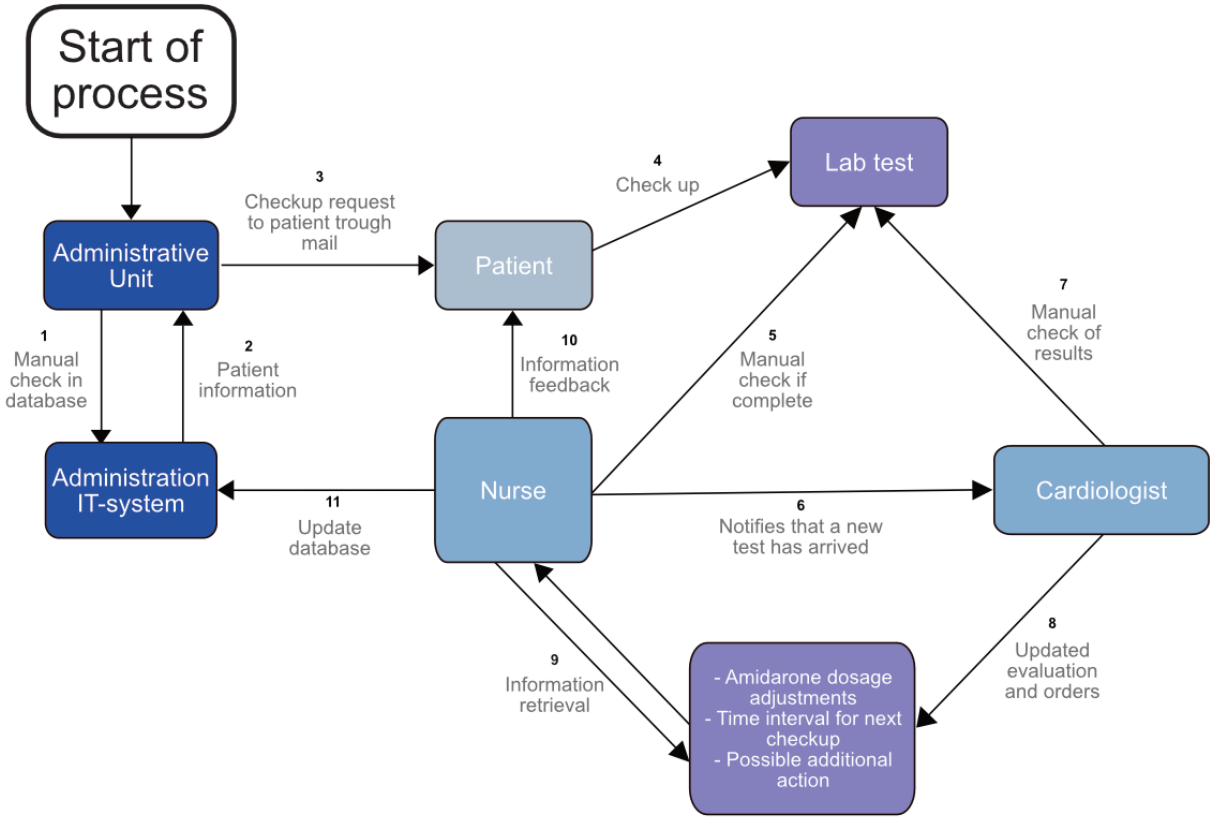

Automation Tool

A commercially available RPA was programmed to perform tasks following a predetermined set of rules and interact with various information systems (UiPath), thereby performing structured and repetitive tasks quickly and reducing the need for manual human work [11]. It thus mimicked human behavior in repetitive and rule-based work [12], and operated by mapping a process in the RPA tool language for the software robot to follow, with runtime allocated to execute the script by a control dashboard [13].

Construction of Initial Prototype

Based on the process map, an initial prototype was constructed. The patient administrative system (ELVIS, Insieme Consulting AB) with the acquisition of listed patients for "amiodarone controls," and the laboratory system (LabBest, CareDialog AB) for access to laboratory results, were integrated. Additionally, an algorithm with reference values for TSH, free thyroxine, free tri-iodothyronine, aspartate aminotransferase, and alanine aminotransferase was built into the process. Three color-coded classes, corresponding to the level of urgency, were created based on reference values. The output was designed as a recommendation for a decision-making cardiologist, serving as a human-in-the-loop, indicating one of three possible actions for each output: class 1: overall normal outcome values (green color) recommending next follow-up test in 6 months, class 2: minor deviation from normal values of uncertain significance (yellow color), recommending follow-up testing in 1 month, and class 3: definite laboratory pathology indicating clear side effect (red color), recommending prompt management by a physician, along with detailed recommendations for management (dose adjustment, discontinuation of amiodarone medication, or consultation of a specific specialist).

XSL•FO
RenderX

Iterative Improvement of Initial Prototype

The first RPA algorithm (robot 1) was tested by comparing the automated output with the real-world manual procedure from April 2021 to December 2021, including iterative adjustments for clinical optimization. Importantly, the new algorithm incorporated expert clinical advice developed by experienced specialists, 2 endocrinologists (HFN and KT) and 1 gastroenterologist (NP), ensuring state-of-the-art recommendations for assessing and managing both normal and pathological findings. Simultaneous technical adjustments optimized performance and reliability. The resulting algorithm was used for validation studies in the final prototype (robot 2) and is presented and explained in detail in Multimedia Appendix 1.

Final Prototype for Validation

For validation and comparative testing, a final prototype, robot 2, was constructed based on the aforementioned adjustments. For comparative analyses, we defined robot 2 as "ground truth," that is, constituting the accurate data used as a reference to validate the system, comparing endocrinologist and gastroenterologist experts' clinical advice against real-world handling by a cardiologist. We considered experienced endocrinologists and gastroenterologists to have more expertise in hormonal or hepatological evaluations compared to the general cardiologist. Furthermore, the output from robot 2 included written instructions for abnormal findings and is presented as supplementary information including specific clinical routines and references to detailed guidelines (Multimedia Appendix 1). The final prototype, robot 2, was tested from February 2022 to February 2023. Notably, the existing routine process remained in operation during the whole period, with the manual routine executed without awareness of the robot, such that the physician's orders constituted the real-world outcome and robot 2 worked in parallel producing outcomes only available to the research team. Paired orders, from the physician and robot 2, were compared.

Comparative Analysis of the Final Prototype for Validation

The comparative analyses' primary outcomes were the following: (1) the time interval between recommended laboratory testing and (2) the detection of side effects. The distribution of recommended time intervals for follow-up and disagreements between robot and physician were analyzed. First, a strict comparison between the number of months to the next follow-up ordered by the robot and the physician was made. Second, each order made by the physician was classified and labeled per robot 2, as follows: after completion of the test period, all dates for the paired test were retrieved from the robot and laboratory results and patient records were examined by one of the authors (HS) and classified as correct or incorrect according to the rules of the algorithm, when compared to the decision made by the physician responsible for managing the case. This evaluation was made without knowledge of the robot classifications. If patient records were ambiguous consultations were made with another author from the team (BIJ). The results from robot classifications were then retrieved for the pairing of outcome classifications and comparative analysis.

We used a confusion matrix as a performance estimate of the physician's classification compared to the robot. A confusion matrix is a statistical tool used to evaluate the performance of a classification model. In this context, it was used to visualize the accuracy of the physician-predicted side effects based on the laboratory test results, comparing the physician-derived orders with the orders by robot 2. The labels applied on classification were the following: (1) true positive (TP) for pathological laboratory results (corresponding to class 2-3, yellow or red), (2) true negative (TN) for normal laboratory results (corresponding to class 1, green), (3) false positive (FP) for follow-up tests ordered by the physician for a time interval >1 month shorter than that ordered by robot 2 (ie, judged as more urgent than ground truth by the physician), and (4) false negative (FN), follow-up tests ordered for a time interval >1 month longer than that ordered by robot 2 (ie, judged as less urgent than ground truth by the physician) or when a definitely pathological test result was not noticed by the physician. Paired orders were compared (n=390 observations). All labels given by the robot and physician were combined into a data frame and presented in the confusion matrix. If the time interval to follow-up test between the algorithm and physician differed ≤1 month the same labels were given. As robot 2 was judged as ground truth, the confusion matrix is limited to true labels (TP and TN) for actual outcomes.

Analysis of Differing Classifications Between the Robot and Physician

To further compare robot 2 and physician estimations, we analyzed the distribution of months to the next follow-up. To investigate which classifications (classes 1-3) contributed the most to the disagreements, that is, the false classifications observed in the confusion matrix, all observations were grouped based on the 3 robot classifications and sorted according to the labels given in the confusion matrix. If the label given by the physician matched the label given by robot 2 the observation was judged as correct. If the physician had given any label other than robot 2, the observation was considered to be incorrect. Classes 1-3 were converted to a month-based scale. Outcome class 3 (red color) was classified as "0 month" since red represented the highest urgency calling for "directly to physician for handling."

Detection of Side Effects

For patients who suffered side effects diagnosed in the patient record during this study, we compared the likelihood of detection between robot 2 and the physician. For robot 2, the correct detection of a side effect was judged as class 3 (red). The correct response to a side effect by the physician (comparable to class 3 [red]), was defined as at least one of the following noted in the patient record: (1) the diagnostic side effect was acknowledged by the cardiologist, (2) the indication for amiodarone was reevaluated, (3) adjustment or termination of medication was undertaken, or (4) referral or consultation with the appropriate specialist was undertaken. Incorrect or questionable action (yellow) was assessed when a new laboratory test was carried out without any of the aforementioned actions.

XSL•FO
RenderX

Statistical Analysis

Data are presented as numbers and percentages, mean (SD) for normally distributed variables tested with the Shapiro-Wilk test or median (IQR), unless otherwise stated. For comparison between groups, the Wilcoxon signed rank test was used. The demographic data includes the population participating during the robot 2 evaluation period.

Ethical Considerations

This study was performed per the principles of the Declaration of Helsinki. The project was approved by the Swedish Ethical Review Authority (2021-01962). This study was observational and informed consent was waived, as the presented robot 2 outcome represents a simulation, and the regular manual routine was upheld.

Results

Demographics of Cohort

Between April 9, 2021, and February 28, 2023, a total of 198 patients, with a mean age of 73 (SD 12) years and 62.6% (n=124) men, were under follow-up routines due to amiodarone prescription and were included in this study. Indications for amiodarone were atrial (n=145, 73.2%) or ventricular (n=53, 26.8%) arrhythmias. The most frequent comorbidities were heart failure (n=142, 71.7%), hypertension (n=127, 64.1%), and coronary artery disease (n=67, 33.8%). Of patients with atrial or ventricular arrhythmias, 23.4% (n=34) and 41.5% (n=22), respectively, had a left ventricular ejection fraction <40%. At baseline, 79.8% (n=156) were on β-blockers, 62.6% (n=124) on novel oral anticoagulants, and 17.2% (n=34) on thyroid hormone (levothyroxine; Table 1).

XSL•FO
RenderX

Table 1. Baseline data for patients with amiodarone treatment at the start of this study[a].

Variables	All patients	Atrial arrhythmia	Ventricular arrhythmia
Population, n (%)	198 (100)	145 (73.2)	53 (26.8)
Demographics			
Male, n (%)	124 (62.6)	87 (60)	37 (69.8)
Age (years), mean (SD)	73 (12)	72 (13)	75 (9)
BMI, median (IQR)	27 (24-31)	27 (24-31)	26 (24-31)
Smoking, n (%)	11 (5.5)	9 (6.2)	2 (3.7)
Comorbidity, n (%)			
Hypertension	127 (64.1)	94 (64.8)	33 (62.3)
Diabetes type I	3 (1.5)	1 (0.7)	2 (3.8)
Diabetes type II	33 (16.7)	20 (13.8)	13 (24.5)
Dyslipidemia	28 (14.1)	25 (17.2)	3 (5.7)
Coronary artery disease	67 (33.8)	42 (29)	25 (47.2)
Congenital heart disease	9 (4.5)	8 (5.5)	1 (1.9)
COPD[b]	11 (5.6)	6 (4.1)	5 (9.4)
Chronic kidney disease	24 (12.1)	17 (12)	7 (13.2)
Stroke	31 (15.7)	19 (13.1)	12 (22.6)
Thyroid disease	17 (8.6)	11 (7.6)	6 (11.3)
Heart failure	142 (71.7)	97 (66.9)	45 (84.9)
HFpEF[c]	54 (27.2)	44 (30.3)	10 (18.9)
HFmrEF[d]	32 (16.2)	19 (13.1)	13 (24.5)
HFrEF[e]	56 (28.3)	34 (23.4)	22 (41.5)
Medication, n (%)			
β-blockers	156 (79.8)	109 (75.2)	47 (88.7)
Digitalis	4 (2)	3 (2.1)	1 (1.9)
Calcium channel blockers	45 (22.7)	32 (22.1)	13 (24.5)
ACEI-ARB[f]	96 (48.5)	73 (50.3)	23 (43.4)
ARNI[g]	41 (20.7)	24 (16.6)	17 (32.1)
Mineralocorticoid receptor antagonist	74 (37.4)	48 (33.1)	26 (49.1)
Acetylsalicylic acid	24 (12.1)	9 (6.2)	15 (28.3)
Platelet inhibitors, other	15 (7.6)	5 (3.4)	10 (18.9)
Novel oral anticoagulant	124 (62.6)	107 (73.8)	17 (32.1)
Waran	33 (16.7)	24 (16.6)	9 (17)
Heparin	3 (1.5)	3 (2.1)	0 (0)
Metformin	22 (11.1)	13 (9)	9 (17)
SGLT2[h]-inhibitor	25 (12.6)	16 (11)	9 (17)
Insulin	9 (4.5)	5 (3.4)	4 (7.5)
Statins	105 (53)	73 (50.3)	32 (60.4)
Long-acting nitroglycerine	16 (8.1)	8 (5.5)	8 (15.1)
Diuretics	75 (37.9)	53 (36.6)	22 (41.5)
Thyroid hormone (levothyroxine)	34 (17.2)	23 (15.9)	11 (20.8)

XSL•FO
RenderX

Variables	All patients	Atrial arrhythmia	Ventricular arrhythmia
X-ray contrast (in prior history)			
Exposure, n (%)	173 (87.4)	122 (84.1)	51 (96.2)
Exposure number of times, median (IQR)	2 (2-4)	2 (2-4)	3 (2-4)
Laboratory values			
Thyroid-stimulating hormone (mIU[i]/L), median (IQR)	1.9 (1.3-2.8)	1.9 (1.2-2.7)	2.1 (1.2-2.7)
Free thyroxin (pmol/L), median (IQR)	15 (13-17)	15 (13-17)	14 (12-17)
AST[j] (µkat/L), median (IQR)	0.44 (0.36-0.56)	0.45 (0.39-0.54)	0.44 (0.34-0.61)
ALT[k] (µkat/L), median (IQR)	0.4 (0.29-0.63)	0.41 (0.3-0.61)	0.39 (0.25-0.64)
Left ventricular ejection fraction			
LVEF[l], median (IQR)	50 (40-59)	55 (45-60)	45 (30-55)
LVEF <40%, n (%)	44 (22.2)	24 (16.6)	20 (37.7)

[a]Values are presented as the number of observations (percentage of population) if not otherwise specified.
[b]COPD: chronic obstructive pulmonary disease.
[c]HFpEF: heart failure with preserved ejection fraction.
[d]HFmrEF: heart failure with midrange ejection fraction.
[e]HFrEF: heart failure with reduced ejection fraction.
[f]ACEI-ARB: angiotensin converter enzyme inhibitors or angiotensin receptor blockers.
[g]ARNI: angiotensin receptor neprilysin inhibitor.
[h]SGLT2: sodium-glucose transport protein 2.
[i]IU: international units.
[j]AST: aspartate aminotransferase.
[k]ALT: alanine aminotransferase.
[l]LVEF: left ventricular ejection fraction.

Construction of the Prototype for Validation

The first prototype, robot 1, was developed from April 2021 to December 2021 through iterative adjustments, eliminating technical errors and improving the medical algorithm until obtaining satisfactory performance as judged by the multiprofessional team, resulting in the final prototype for validation, robot 2 (Figure 2). The technical improvement and reduction of errors over time are presented, with an error defined as an output from the robot not resulting in classification 1, 2, or 3 (Figure 3).

Figure 2. Process map of workflow and procedures included in the prototype robot (robot 2). ALT: alanine aminotransferase; AST: aspartate aminotransferase; PIN: personal identification number; T4: thyroxin; TSH: thyroid stimulating hormone.

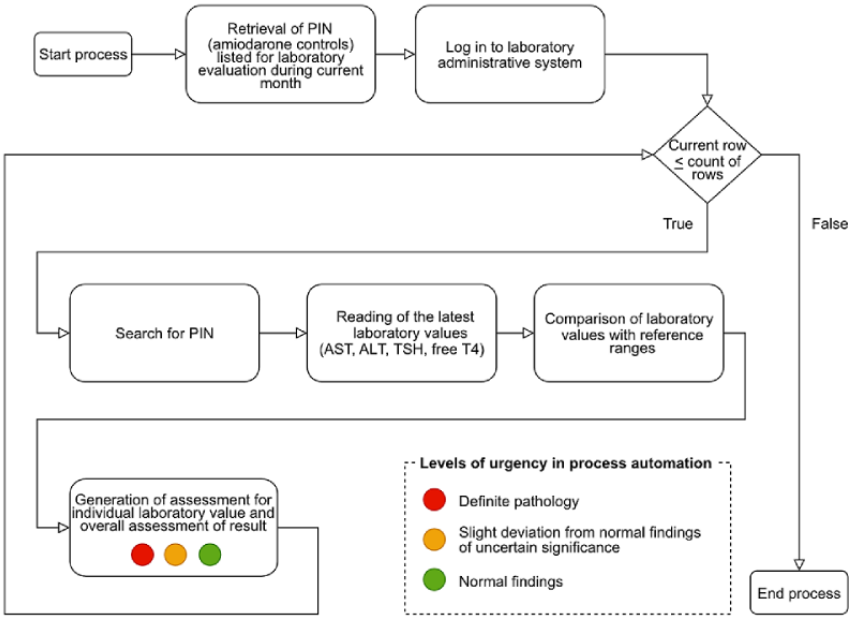

XSL•FO
RenderX

Figure 3. Proportion of technical robot errors over time (checkup time points).

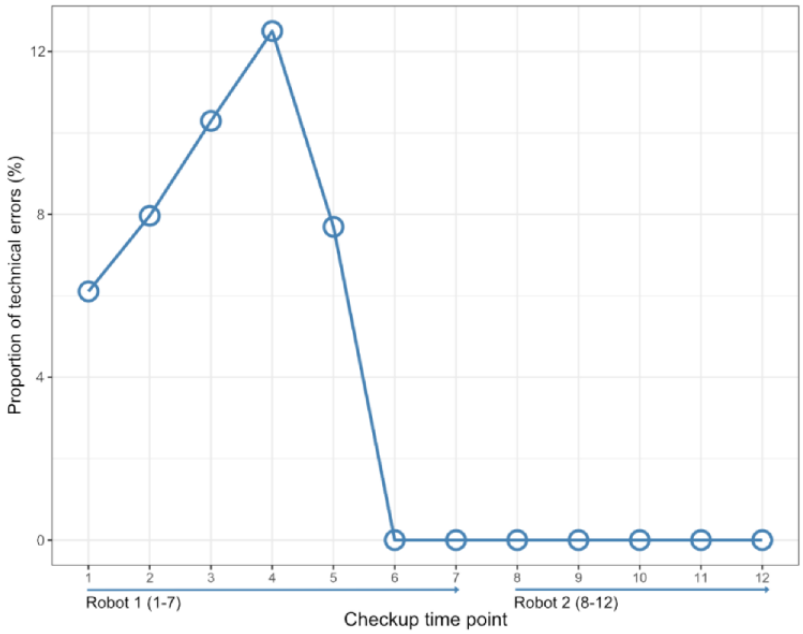

Comparative Analysis of the Final Prototype for Validation

The primary finding of this study (Multimedia Appendix 2) was a significant difference in the ordered time interval to the next follow-up between robot 2 and the physician (robot: mean 4.5, SD 2.4 months; physician: mean 3.1, SD 1.4 months; $P<.001$). To analyze the model's accuracy in predicting each class, paired orders from robot 2 and physicians (n=390 observations) were compared using a confusion matrix, which displays the actual versus predicted classifications. Based on robot 2 output defined as ground truth, the robot labels represent the actual values and the physician's labels the predicted values (Figure 4). Of the observations, 112 (28.7%) corresponded between robot and physician. The most pronounced difference between the robot and the physician occurred for actual TNs according to the robot, which was judged as positive (predicted FP) by the physician in 246 (63.1%) observations.

Figure 4. Confusion matrix of physician (predicted) and robot (actual) labels. Robot labels are judged as ground truth. Percentages of labels are shown in each square and the numbers of paired observations are below. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

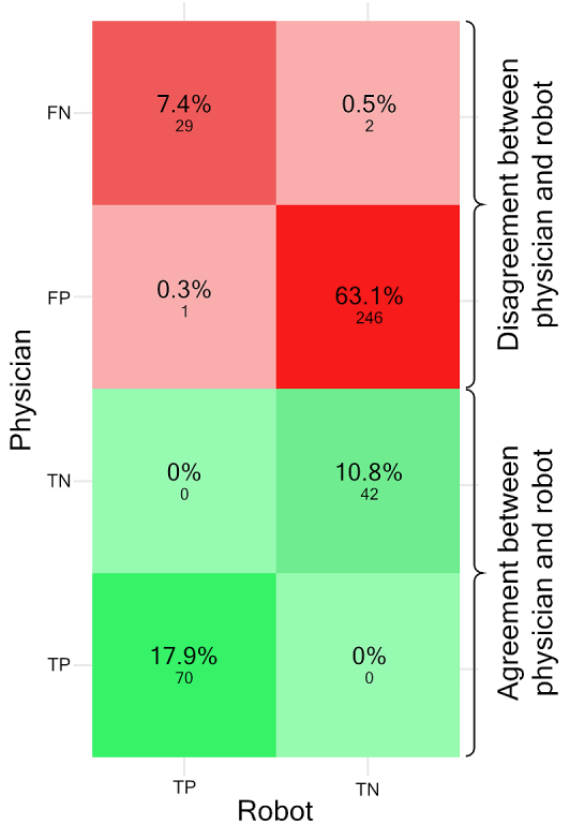

XSL•FO
RenderX

The bar chart, illustrating the difference between robot 2 and the physician's orders, demonstrates the largest visual difference in ordered times for follow-up as the robot ordered 6 months to the next control (ie, corresponding with normal laboratory values) in 281 (72.1%) observations, compared to physician's orders in 38 (9.7%). The remaining robot orders were concentrated in lower time intervals (0 months, ie, recommending immediate physician's action, and 1 month), whereas physician's orders peaked at the 3-month interval (n=118, 30.3%), followed by 4 months (n=109, 27.9%; Figure 5).

Figure 5. Bar chart of frequencies of patients who were ordered a time interval for follow-up by physician or robot, interval ranging from 0 to 6 months, on the x-axis. The absolute counts are shown on the y-axis and the percentage of physician and robot counts, respectively, is marked above the stacks (based on n=390 observations). For the robot, a red outcome implied a referral directly to the physician, which translated to "zero months.".

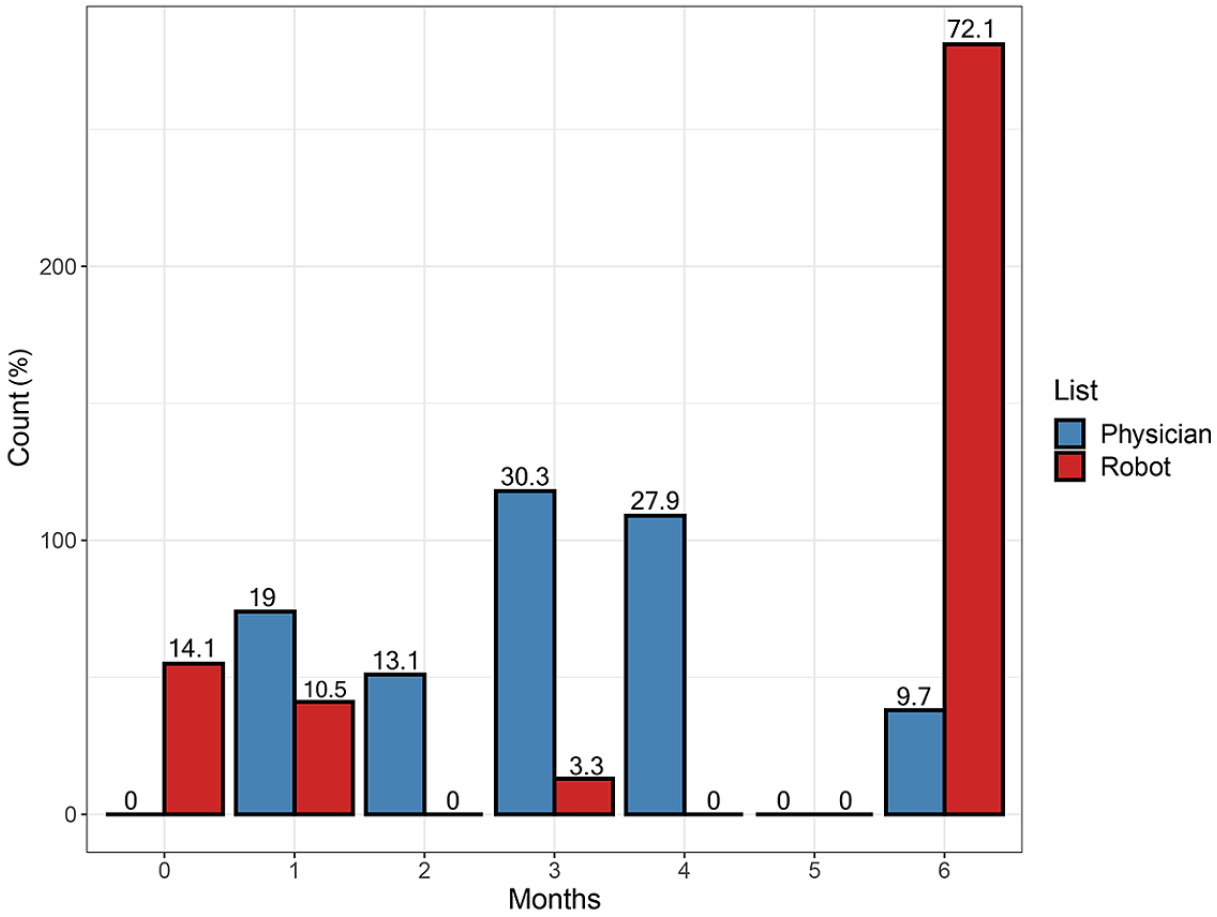

Complementary analysis showed the largest discrepancy for cases with the lowest degree of urgency (class 1: green, follow-up in 6 months) where false decisions greatly outweighed the true ones. When more urgent action was called for (classes 2 and 3: yellow and red colors), the robot and physician orders were more aligned, especially for class red (clearly pathological findings; Figure 6).

Figure 6. Discrepancies in outcome labels by the physician compared to the robot for the 3-color coded classes (representing different levels of aberration vs normality): (1) green class: normal laboratory values, (2) yellow class: minor deviation from normal or uncertain significance, (3) red class: definite laboratory pathology. "True" represents that the physician's classification matched the robot (ground truth). "False" represents that the physician's classification differed from ground truth. The height of the stacks represents the percentage of outcome labels (y-axis), and the absolute number of observations appears below the stacks.

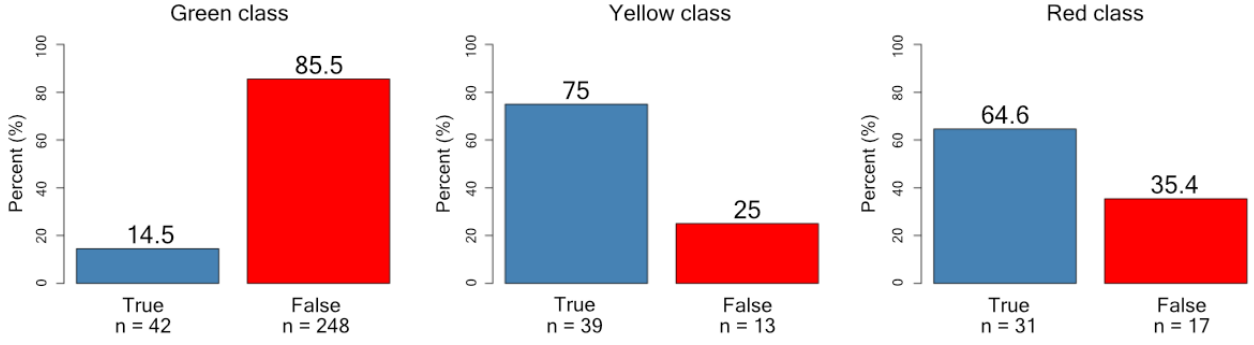

XSL•FO
RenderX

Detection of Side Effects

During the 1-year evaluation, 12 patients developed side effects detectable in laboratory analyses (AIT n=8, AIH n=3, and hepatic side effects n=1). The robot assessed all these as class 3 (red), that is, a correct diagnosis and action. The physician's assessment was red in 8 and yellow in 4 (repeat ordering of follow-up laboratory tests after 1-3 months) of the corresponding cases (Figure 7).

Figure 7. Diagnostic side effects during study of robot 2 and assessment by robot versus physician. Observation of a correct response to a side effect by the physician required that (1) the diagnostic side effect was noted in medical records by the physician, (2) the indication for amiodarone was reevaluated, (3) adjustment or termination of medication was undertaken, or (4) referral or consultation with an appropriate medical specialist was done. Incorrect or questionable action represented that new laboratory tests were ordered without any of these actions. For repeated laboratory tests only, the interval of months ordered is noted in the boxes. Observation of a correct response to a side effect (red) by the robot required that the outcome reading was of the highest alert (red) ordering referral to a physician for action. *The physician reduced the dosage of amiodarone and ordered repeated testing after 1 month but documented no rationale for the action. F: female; M: male; mo: months.

Side effects (age and sex)	**Robot 2**	**Physician**
Hyperthyroidism		
83 F		
83 M		1 mo
76 M		2 mo
73 F		
70 F		
69 M		
63 F		
49 M		
Hypothyroidism		
83 M		3 mo
79 M		
78 M		
Hepatic side effect		
83 M		1* mo

Discussion

Overview

This study demonstrates that an automated process for decision-making based on thyroid and liver laboratory values outperformed a physician-operated manual routine for amiodarone follow-up, leading to reduced frequency of laboratory testing and more rapid detection of side effects. Amiodarone is the most effective antiarrhythmic drug to prevent recurrences of atrial fibrillation (AF) when other antiarrhythmic drugs have failed, are contraindicated, or when ablation of AF or ventricular arrhythmias is not an alternative [14]. In our cohort, AF constituted the most common arrhythmia. The efficiency of amiodarone paired with the considerable risk of side effects supports the rigorous follow-up routines prescribed in international recommendations [1,10], resulting in repetitive and time-consuming manual work in general cardiology practice.

Using automation through a software robot, an RPA, appears as one feasible solution to replace human labor and ensure compliance with guidelines. Software robots work with information systems via the front end, that is, via the presentation layer, and do not impact the underlying data infrastructure [15]. An RPA is suitable for a process that follows a standardized and rule-based structure (does not require cognitive or judgment effort), is conducted both often and manually by humans and requires multiple-system access [16].

XSL•FO
RenderX

Notably, this is not an artificial intelligence tool. The prototype provides recommendations to the physician as a human-in-the-loop. This retains the responsibility as a control mechanism, as all recommendations must be confirmed by physician's orders. Human exception management is necessary to solve problems where cognition, intuition, and situational decisions are needed [17].

Main Findings

Most notably, the robot ordered significantly longer intervals between tests compared to the physician which is especially pronounced in the low-risk cases. Using the robot may thus serve to reduce manual labor, lessen inconvenience and worry in patients, and liberate testing personnel for other care tasks. Additionally, the robot identified all laboratory-diagnosed side effects, as opposed to the physician, who initially missed 1 of 3. Thus, the use of the robot appeared to increase the safety of the procedure, based on more prompt recognition of side effects. Together, these findings support that monitoring recommendations based on appropriate expert clinical advice are superior to handling only by a cardiologist.

Detailed Analysis

Based on the outcomes defined by ground truth, we found that the most prevalent disagreement between the robot and physician occurred when laboratory values were normal (TN), whereas the physician classified normal findings as pathological (FP). This resulted in physicians ordering a significantly shorter time for the next laboratory follow-up than as motivated by guidelines or laboratory findings. Interestingly, the physicians' orders tended to cluster around the mean eligible choice, 3-4 months in almost 60%. This supposedly caused considerably more frequent laboratory follow-ups for patients when guided by the physician, leading to increased testing costs, manual labor, inconvenience, pain, and possible worries for the patients without contributing to a better quality of care.

In contrast, the robot-physician orders were better aligned for the more urgent laboratory results, indicating that clearly pathological findings would not be missed. Still, the detailed analysis showed that the robot was superior as it identified all side effects, compared to 1 of 3 missed at first presentation by the physician. Although the results appear indicative of a more rapid detection of side effects by the robot compared to the physician, this study was not designed or powered to evaluate the performance regarding the detection of side effects. As for the patterns of physician's ordering, we can only speculate but lack empirical support. The shorter control intervals ordered, despite knowledge of guidelines, may reflect a precautionary principle and hesitation to choose intervals at the recommended scale's margins. Other explanations may be decision fatigue and working under time constraints, which may result in excessive caution. The seemingly inferior detection of side effects by the physician may be due to the algorithm being based on expert clinical advice in thyroid and liver disease, and the physicians responsible for monitoring were cardiologists, not mastering the best practices in these disciplines.

The categorization of laboratory values based on an expert-guided algorithm, as created for robot 2 for validation, appeared crucial. Notably, the algorithm was specifically formulated by experienced clinicians with considerable experience in the laboratory effects of amiodarone medication, as no composite guidelines exist for this specific situation. This procedure secured appropriate recommendations for the management of pathological findings and also incorporated a margin for expected deviations due to amiodarone medication per se.

Plans

Following the satisfying outcome, our process for monitoring amiodarone treatment has gained approval as an in-house manufactured medical device for introduction at our health care institution, the Sahlgrenska University Hospital, per Regulation 2017/745 of the European Parliament, Article 5(5) concerning medical devices [18]. We are currently implementing the robot in clinical practice, under follow-up and supervision. However, evolving technology presents new avenues for development, such as the integration of a locally hosted large language model in the robot for increased accuracy and clinical relevance. This would enable both compilation of laboratory outcomes and individual recommendations for the patient at hand, possibly also incorporating consideration of clinical information. Another goal, outside the scope of this study, is to perform a more detailed analysis of physicians' behavior. This could be done through surveys or interviews with physicians, to explore the underlying reasons for not strictly following guidelines, such as stress or decision fatigue involved in decision-making in clinical situations.

Strengths and Limitations

One strength of our study was that the robot was carefully iteratively tested and designed to address technical and medical issues before launching the optimized prototype, robot 2, for validation during a year of final testing. Thus, our study presents detailed information on feasibility and safety in a real-world setting. Notably, the procedure will only identify the most common side effects, detectable in blood laboratory tests, whereas the remaining risks must still be considered in clinical visits. The robot does not replace the clinical consultation but constitutes an automated safety net as clinical consultations are regularly less frequent than the need for follow-up laboratory testing.

One weakness was that the final testing was carried out in parallel with real-world monitoring, providing a simulated outcome as a randomized controlled trial was outside the scope of this study. Additionally, the sample size of 390 paired orders was relatively small, which may limit the generalizability and statistical power of the findings, although including a complete cohort of all patients receiving amiodarone prescriptions at our health care institution. For example, this study was not powered to evaluate the performance between methods regarding the detection of side effects, although the robot appeared to outperform physician recommendations. Additionally, the results imply positive health economic outcomes, which have not been explored in this study.

XSL•FO
RenderX

Conclusion

This study aimed to design and test an automated process for amiodarone treatment monitoring, mastering the manual tasks, currently performed by nurses and physicians, with preserved patient value. After iterative design and technical improvements, a prototype for validation was tested for 1 year. We found that the automated process provided a reliable alternative to the present manual management of amiodarone treatment monitoring. The results indicated that the robot may reduce manual labor and frequency of laboratory testing and detect side effects with increased precision compared to the physician, thereby reducing costs, and enhancing patient value.

Acknowledgments

We would like to acknowledge cardiology nurses Maria Samuelsson, RN, Malin Lindqvist, RN, specialist nurse Inger Martinsson, RN, deputy director Markus Jonebrant, RN, Department of Medicine, Geriatrics and Emergency Medicine, Sahlgrenska University Hospital/Östra, and librarian Emelie Mellbrand, MA, Sahlgrenska Research Library, and the Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, for their valuable and constructive contributions in this work. This study was financed by grants from the Swedish state under the agreement between the Swedish government and the county councils regarding the education of physicians and research, the ALF agreement from the Healthcare Board, Region Västra Götaland (grant ALFGBG-790271) to HFN; ALFGBG-971608 to ML; and the Innovation Platform, Region Västra Götaland, Sweden (grant 993954) to HS. We are grateful to Jamie D Kraft, PhD, for scientific editing and language review.

Data Availability

The data underlying this paper cannot be publicly shared to protect the privacy of this study's participants but will be made available upon reasonable request to the corresponding author. The software code for the RPA is available upon request to JL.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Algorithm for thyroid and liver test results for the robotic process automation in treatment with amiodarone.
[DOCX File , 34 KB - jmir_v27i1e65473_app1.docx]

Multimedia Appendix 2
Brief PowerPoint (Microsoft Corp) description of study including illustration from interactive user interface.
[PDF File (Adobe PDF File), 631 KB - jmir_v27i1e65473_app2.pdf]

References

1. Epstein AE, Olshansky B, Naccarelli GV, Kennedy JI, Murphy EJ, Goldschlager N. Practical management guide for clinicians who treat patients with amiodarone. Am J Med 2016;129(5):468-475. [doi: 10.1016/j.amjmed.2015.08.039] [Medline: 26497904]
2. Kinoshita S, Hosomi K, Yokoyama S, Takada M. Time-to-onset analysis of amiodarone-associated thyroid dysfunction. J Clin Pharm Ther 2020;45(1):65-71. [doi: 10.1111/jcpt.13024] [Medline: 31400296]
3. Trohman RG, Sharma PS, McAninch EA, Bianco AC. Amiodarone and thyroid physiology, pathophysiology, diagnosis and management. Trends Cardiovasc Med 2019;29(5):285-295 [FREE Full text] [doi: 10.1016/j.tcm.2018.09.005] [Medline: 30309693]
4. Martino E, Bartalena L, Bogazzi F, Braverman L. The effects of amiodarone on the thyroid. Endocr Rev 2001;22(2):240-254. [doi: 10.1210/edrv.22.2.0427] [Medline: 11294826]
5. Lewis JH, Ranard RC, Caruso A, Jackson LK, Mullick F, Ishak KG, et al. Amiodarone hepatotoxicity: prevalence and clinicopathologic correlations among 104 patients. Hepatology 1989;9(5):679-685. [doi: 10.1002/hep.1840090504] [Medline: 2785079]
6. Biancatelli RMC, Congedo V, Calvosa L, Ciacciarelli M, Polidoro A, Iuliano L. Adverse reactions of amiodarone. J Geriatr Cardiol 2019;16(7):552-566 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2019.07.004] [Medline: 31447894]
7. Hamilton D, Nandkeolyar S, Lan H, Desai P, Evans J, Hauschild C, et al. Amiodarone: a comprehensive guide for clinicians. Am J Cardiovasc Drugs 2020;20(6):549-558. [doi: 10.1007/s40256-020-00401-5] [Medline: 32166725]
8. Goundan PN, Lee SL. Thyroid effects of amiodarone: clinical update. Curr Opin Endocrinol Diabetes Obes 2020;27(5):329-334. [doi: 10.1097/MED.0000000000000562] [Medline: 32773570]
9. Goldschlager N, Epstein AE, Naccarelli GV, Olshansky B, Singh B, Collard HR, Practice Guidelines Sub-committee, North American Society of PacingElectrophysiology (HRS). A practical guide for clinicians who treat patients with amiodarone: 2007. Heart Rhythm 2007;4(9):1250-1259. [doi: 10.1016/j.hrthm.2007.07.020] [Medline: 17765636]

XSL•FO
RenderX

10. Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, et al. 2016 American thyroid association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. Thyroid 2016;26(10):1343-1421. [doi: 10.1089/thy.2016.0229] [Medline: 27521067]
11. Jonsson J. Robotic process automation in swedish healthcare [Master´s thesis]. Sweden: Lund University; 2021.
12. Van der Aalst W, Bichler M, Heinzl A. Robotic process automation. Bus Inf Syst Eng 2018;60(4):269-272 [FREE Full text] [doi: 10.1007/s12599-018-0542-4]
13. Tornbohm C. Market guide for robotic process automation software. Gartner Research. 2017. URL: https://www.gartner.com/en/documents/3835771 [accessed 2025-01-09]
14. Roy D, Talajic M, Dorian P, Connolly S, Eisenberg MJ, Green M, et al. Amiodarone to prevent recurrence of atrial fibrillation. Canadian trial of atrial fibrillation investigators. N Engl J Med 2000;342(13):913-920. [doi: 10.1056/NEJM200003303421302] [Medline: 10738049]
15. Lacity M, Willcocks L. Robotic Process Automation at Telefonica O2. MIS Q. Exec 2016;15(1):21-35 [FREE Full text]
16. Moffitt K, Rozario A, Vasarhelyi M. Robotic process automation for auditing. J Emerg Technol Account 2018;15(1):1-10 [FREE Full text]
17. Hofmann P, Samp C, Urbach N. Robotic process automation. Electron Markets 2020;30(1):99-106 [FREE Full text] [doi: 10.1007/s12525-019-00365-8]
18. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. European Union. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745 [accessed 2025-01-09]

Abbreviations

AF: atrial fibrillation
AIH: amiodarone-induced hypothyroidism
AIT: amiodarone-induced thyrotoxicosis
FN: false negative
FP: false positive
RPA: robotic process automation
TN: true negative
TP: true positive
TSH: thyroid-stimulating hormone

Edited by A Mavragani; submitted 20.08.24; peer-reviewed by A Pouliakis, C Sun, Y Wu; comments to author 02.12.24; revised version received 18.12.24; accepted 27.12.24; published 19.02.25.

Please cite as:
Johansson BI, Landahl J, Tammelin K, Aerts E, Lundberg CE, Adiels M, Lindgren M, Rosengren A, Papachrysos N, Filipsson Nyström H, Sjöland H
Automated Process for Monitoring of Amiodarone Treatment: Development and Evaluation
J Med Internet Res 2025;27:e65473
URL: https://www.jmir.org/2025/1/e65473
doi:10.2196/65473
PMID:

©Birgitta I Johansson, Jonas Landahl, Karin Tammelin, Erik Aerts, Christina E Lundberg, Martin Adiels, Martin Lindgren, Annika Rosengren, Nikolaos Papachrysos, Helena Filipsson Nyström, Helen Sjöland. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Knowledge and Information-Seeking Behaviors of People Living With Multiple Sclerosis: Cross-Sectional Questionnaire Study

Véronique Duguay[1*], BSciS; Dominique Comeau[2*], PhD; Tiffany Turgeon[3], BSc; Nadia Bouhamdani[1,3,4], PhD; Mathieu Belanger[1,2], PhD; Lyle Weston[5], MD; Tammy Johnson[6], BSN; Nicole Manzer[6], BSN; Melissa Giberson[6], BSN; Ludivine Chamard-Witkowski[1,7], MD

[1]Centre de Formation Médicale du Nouveau-Brunswick, Université de Sherbrooke, Moncton, NB, Canada

[2]Vitalité Health Network, Dr. Georges-L.-Dumont University Hospital Centre, Moncton, NB, Canada

[3]Department of Chemistry and Biochemistry, Faculty of Science, Université de Moncton, Moncton, NB, Canada

[4]Vitalité Health Network, Medical Genetics Department, Dr Georges-L.-Dumont University Hospital Center, Moncton, NB, Canada

[5]Department of Neurology, The Moncton City Hospital, Moncton, NB, Canada

[6]Multiple Sclerosis Clinic, The Moncton City Hospital, Moncton, NB, Canada

[7]Department of Neurology, Dr. Georges-L.-Dumont University Hospital Center, Moncton, NB, Canada

[*]these authors contributed equally

Corresponding Author:
Véronique Duguay, BSciS
Centre de Formation Médicale du Nouveau-Brunswick
Université de Sherbrooke
50 De la Francophonie Street
Moncton, NB, E1A 7R1
Canada
Phone: 1 5062261627
Email: veronique.duguay2@usherbrooke.ca

Abstract

Background: The internet has emerged as a primary source of health-related information for people living with multiple sclerosis (MS). However, given the abundance of misinformation found on the web, this behavior may pose a significant threat to internet users.

Objective: This study aims to explore the knowledge and information-seeking behavior of people living with MS followed at a specialized MS clinic where education is a cornerstone of care.

Methods: This cross-sectional survey–based study comprised 20 true or false statements, covering both scientific facts and popular misinformation about MS treatments. A “scientific fact score” and a “misinformation score” were calculated by attributing a scoring system to each point in the survey: +1 point was attributed to correct answers, –1 point was attributed to incorrect answers, and 0 point was attributed to “I don’t know.” Furthermore, the survey inquired about participants’ health-seeking behaviors.

Results: The mean age of the 69 participants was 48.4 (SD 10.9) years, 78% (54/69) were female, 81% (56/69) were highly educated, 90% (62/69) were receiving a disease-modifying therapy, and 52% (30/58) had experimented with alternative therapies. The mean score for answering the scientific and misinformation questions correctly was 69% (SD 2.4%) and 22% (SD 4.5%), respectively (P<.001). Notably, when questioned about misinformation, answering correctly dropped significantly (P<.001), while indecision (P<.001) and answering incorrectly (P=.02) increased. Sociodemographic factors and medical questions were not significantly associated with scientific and misinformation scores (all P>.05); however, misinformation scores did significantly correlate with levels of education (P=.04). The main sources of health-related information were from expert-led MS websites (48/58, 82%) and health care professionals (34/58, 59%). Low-reliability sources were less used; however, word of mouth seemed to be prevalent (14/58, 24%), followed by Facebook (10/58, 17%). On average, people with MS reported having consulted 3 high- to moderate-quality sources and only 1 low-quality source.

XSL•FO
RenderX

Conclusions: Education at the clinic and consulting primarily moderate- to high-quality sources did not safeguard against misinformation, indicating a need for more misinformation-geared education at the clinic. Notably, there is a need to proactively educate patients about misinformation commonly found on the web, and more importantly, create space for them to discuss the information without prejudice. As novel educational methods may be relatively more time-consuming, implementing change may be challenging. Furthermore, age, sex, education level, and health literacy might not safeguard against misinformation. Herein, we were unable to identify correlations associated with scores obtained on the questionnaire other than educational level. Although the educational level did seem to impact the misinformation score, this did not stop participants from experimenting with alternative therapies. Although studies are exploring novel ways to effectively deal with health misinformation on the web, more research is needed to fully understand this highly complex social phenomenon.

(J Med Internet Res 2025;27:e63763) doi:10.2196/63763

KEYWORDS

multiple sclerosis; chronic illness; misinformation; web-based searches; education; health information; social media; health literacy; patient-doctor relationship; health-related information; information-seeking behavior

Introduction

Although patients can acquire information regarding health from a multitude of sources such as health care professionals, pharmaceutical companies, books, pamphlets, radio, and television, the majority will tend to get their information on the web [1-3]. Given the large number of unreliable sources and inaccurate information currently flooding the internet, patients can easily become misinformed [3]. Misinformation can be defined as information that is not in accordance with the leading scientific consensus [4]. Although misinformation is widespread across the web-based sphere, social media sources including Twitter, YouTube, Facebook, and WhatsApp have been found to be the predominant outlets of misinformation [3,5]. With most of the health information on the web assessed as being of low-quality, inaccurate, and incomplete [6,7], users are at an increased risk of being deceived by misinformation. This is further exacerbated by the fact that false information diffuses significantly farther, faster, deeper, and more broadly than veracious information [8-10]. While web-based searches can empower patients [11] and improve their knowledge, health literacy, capacity to manage their health [12], and the patient-doctor relationship [13], misinformation may also lead individuals to erroneous conclusions and potential harm [4]. Patients may be particularly vulnerable to hopes of recovery driven by misinformation, myths, and unproven therapies [4].

For those living with multiple sclerosis (MS), the internet is becoming an important source of health-related information as well as a source of community support through the exchange of lived experience [14-19]. For people with MS, information found on the internet may help in coping with the disease and provide some reassurance in their chosen therapeutic modality [15]. More than half of people with MS used mass media sources rather than interpersonal information sources such as health care providers and advocacy organizations, among others [14]. Thus, addressing the reliability of the information found on the internet is a critical issue [18] as people with MS are likely to be affected by misinformation [20,21]. Notably, a recent study evaluating the quality of YouTube videos regarding MS found that more than half lacked high-quality content [20]. Nevertheless, studies have indicated that people with MS still believe the most trusted information source to be their physician [14,18], and a large proportion remain skeptical about the quality of the information they gather on the web [14,17], with 40% stating they had concerns [14].

With both web-based information and the global population with MS on the rise [22], a growing number of people with MS are exposed to potential harm through unverified therapies, potentially impacting their security and health [23]. Indeed, most people with MS will use complementary and alternative medicine seen on various social media platforms [24-26]. Although seemingly well tolerated, these unproven therapies may interact adversely with medical treatments, pose harm, and incur financial costs [23,25]. The purpose of this study was to evaluate if people with MS at a specialized MS clinic were able to correctly answer questions about their illness as well as identify misinformation about alternative therapies. Other objectives were to identify where people with MS get their health-related information and identify correlations associated with questionnaire scores.

Methods

Research Design

People living with MS and followed at the MS clinic in Moncton, New Brunswick, Canada, were invited to take part in the survey-based study between April and July 2023. More specifically, participants who agreed to be contacted by the research team were telephoned and informed of the study. To be included, people with MS had to be aged 19 years or older to provide informed consent.

Ethical Considerations

The study was approved by the Vitalité Health Network Research Ethics Board (101729). If the participant decided to participate, they were asked to sign the consent form after which they were invited to fill out the survey questionnaire. Data was de-identified to ensure privacy and confidentiality. Participants were not compensated.

Participant Demographics

Participants included 69 people living with MS followed at a specialized MS clinic providing management, education, and treatment of MS. More specifically, 90 potential participants were contacted for the study and 69 were accepted: giving a

XSL•FO
RenderX

participation rate of 77%. At the clinic, people living with MS are followed by a clinical team comprised of 1 neurologist and 3 nurses as well as neurologists on standby. They have access to the clinical team 5 days per week, which they can readily contact by telephone or by email. Initial consultations are comprised of a thorough educational session that relays information about the disease such as physiopathology, symptoms, disease evolution, as well as medication. During this session, they are given resources to consult for disease-related information; they are also encouraged to contact their clinical team if they have any questions and to be wary of nonexpert sources. Subsequent visits are supplemented with information on relevant symptoms, disease evolution, and medication tailored to their specific needs.

Survey Instrument and Data Collection Procedure

A 20-item "true or false" survey was developed by neurologists and nurses, which comprised 10 scientific facts and 10 popular misinformation statements regarding MS (Multimedia Appendix 1). Scientific facts and misinformation statements were distributed at random in the questionnaire. For each item, participants were asked to respond if they considered the statement to be true, partially true, do not know, partially false, or false. The 10 scientific facts consisted of accurate scientific truths about MS, while the 10 misinformation statements consisted of erroneous statements about the efficacy of alternative medicine in the treatment of MS. The latter was developed based on purportedly effective alternative medicine options found on the web and on social media. Hence, the correct answers for this survey were true for all scientific facts and false for misinformation statements. A "scientific fact score" and a "misinformation score" were calculated by attributing a scoring system to each point in the survey: +1 point was attributed to correct answers (true or partially true for facts and false or partially false for misinformation statements), –1 point was attributed to incorrect answers, and 0 point was attributed to "I don't know." A range of –10 to 10 points could therefore be obtained for each of the scientific fact score and misinformation score. The survey also collected sociodemographic information, the participants' use of disease-modifying therapies and alternative medicine, and the sources used to get health-related information. The sources used to get health-related information were classified as either high to moderate reliability or low reliability. The classification was done by the neurologist and nurses at the clinic and was based on source expertise (ie, the ability of the source to give accurate information) and trustworthiness (ie, the willingness of the source to provide accurate information). Education was classified as either "high" or "low"; high education included all postsecondary education diplomas (ie, university and college diplomas) and low education included high school–level diplomas or less.

Statistical Analysis

ANOVA post hoc Sidak was used to compare means between multiple groups and the nonparametric Mann-Whitney *U* test was used to test between 2 samples. Multiple regression analysis was used to analyze the relationship between scores and several independent variables. All statistical tests were performed using GraphPad Prism (version 10.1.0; GraphPad Software Inc). Parametric tests were used after data passed normality tests (D'Agostino-Pearson omnibus, Anderson-Darling, Shapiro-Wilk, and Kolmogorov-Smirnov).

Results

Overview

The mean age of the 69 participants was 48.4 (SD 10.9) years and the mean time of MS diagnosis for participants was 10.9 (SD 7.8) years (Table 1). A total of 78% (54/69) of participants identified as female, 20% (14/69) as male, and 1.4% (1/69) as nonbinary (Table 1). A total of 81% (54/69) of participants were classified as highly educated (Table 1). Additionally, 90% (62/69) of participants were receiving disease-modifying therapies and 52% (30/58) reported experimenting with alternative medicine (Table 1).

XSL•FO
RenderX

Table 1. Respondent demographics and clinical information (N=69).

Response	Values
Age (years), mean (SD)	48.4 (10.9)
Sex, n (%)	
Female	54 (78)
Male	14 (20)
Nonbinary	1 (1)
Education, n (%)	
High	56 (81)
Low	13 (19)
Time since MS[a] diagnosis (years), mean (SD)	10.9 (7.8)
Receiving disease-modifying therapy, n (%)	
Yes	62 (90)
No	7 (10)
Using alternative medicine[b], n (%)	
Yes	30 (52)
No	28 (48)

[a]MS: multiple sclerosis.

[b]Out of the total 69 patients, only 58 answered this question.

People With MS Were More Knowledgeable About Scientific Fact Than Misinformation

Overall, when asked whether scientific facts about their disease were true or false, most people with MS were able to answer correctly (Figure 1A). Comparatively, relatively fewer participants could accurately identify misinformation. Notably, all participants correctly answered that "individuals with MS experience personal and emotional changes such as anxiety, depression, and difficulties sleeping," and almost all participants correctly answered that "MS is a condition that affects your brain and spinal cord (your central nervous system)." More complex facts, such as "genetic factors don't seem to play a large role in MS," "a variety of viruses have been linked to MS, including Epstein-Barr virus," and "the target of almost all MS-specific drugs are certain cells of the immune system," seemed to be lesser known. The most incorrectly answered misinformation statements were that "cooling methods are thought to be an effective treatment for MS," "adapting a paleolithic diet helps treat MS," and that "marijuana (cannabis) is an effective treatment for MS."

XSL•FO
RenderX

Figure 1. Scientific facts and misinformation identification by people with MS. (A) Answer distribution (correct, indecision, and incorrect) for scientific facts and misinformation per question. (B) Differences in answer distribution between scientific facts and misinformation (Mann-Whitney U test). (C) Average score differences between scientific facts and misinformation (ANOVA, post hoc Sidak). MS: multiple sclerosis.

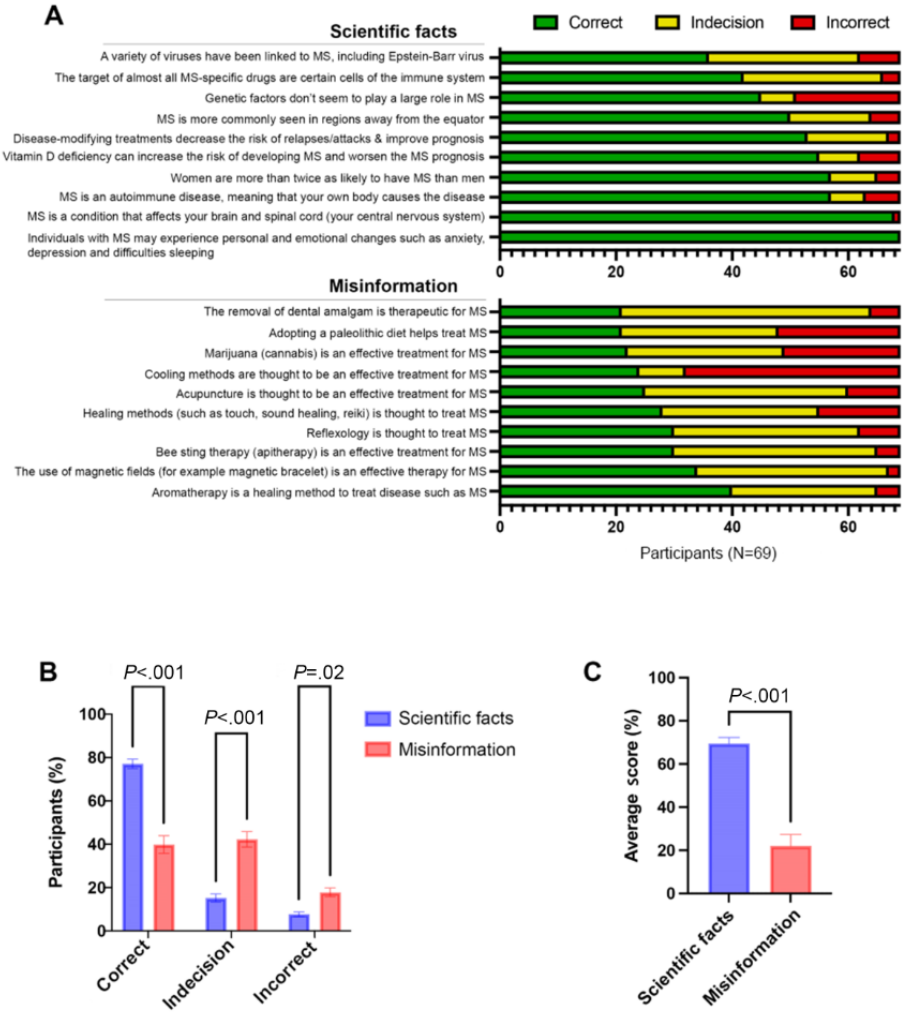

When comparing the scientific fact and misinformation score, participants were better able to correctly answer scientific facts (*P*<.001), indecision (answering "I do not know") was more prevalent for misinformation statements (*P*<.001), and answering incorrectly significantly increased in the case of misinformation (*P*=.02; Figure 1B). Correspondingly, the average scientific fact and misinformation scores of people with MS were 69% (SD 2.4%) and 22% (SD 4.5%), respectively (Figure 1C).

People Living With MS Used High- to Moderate-Reliability Sources for Their Disease-Related Information and Minimal Low-Reliability Sources

People living with MS were asked to share where they got their health-related information about their illness. The sources were classified as either high- to moderate-reliability sources or low-reliability sources (Figure 2). The great majority of participants from the specialized clinic reported getting their information from high- to moderate-reliability sources, notably, from trustworthy and expert-led MS websites (48/58, 82%) and health care professionals, namely, neurologists and nurses (34/58, 59% and 30/58, 52%, respectively). Participants also got their information from scientific documents (22/58, 38%) like peer-reviewed publications, books (20/58, 34%), and pharmaceutical companies (19/58, 33%). Low-reliability sources were less used; however, word of mouth seemed to be prevalent (14/58, 24%), followed by Facebook (10/58, 17%). On average, people with MS reported having consulted 3 high- to moderate-quality sources and only 1 low-quality source.

XSL•FO
RenderX

Figure 2. Sources used to get disease-related health information about MS. Health-related information sources are divided into 2 categories depending on their reliability: low-reliability and high- to moderate-reliability sources. MS: multiple sclerosis.

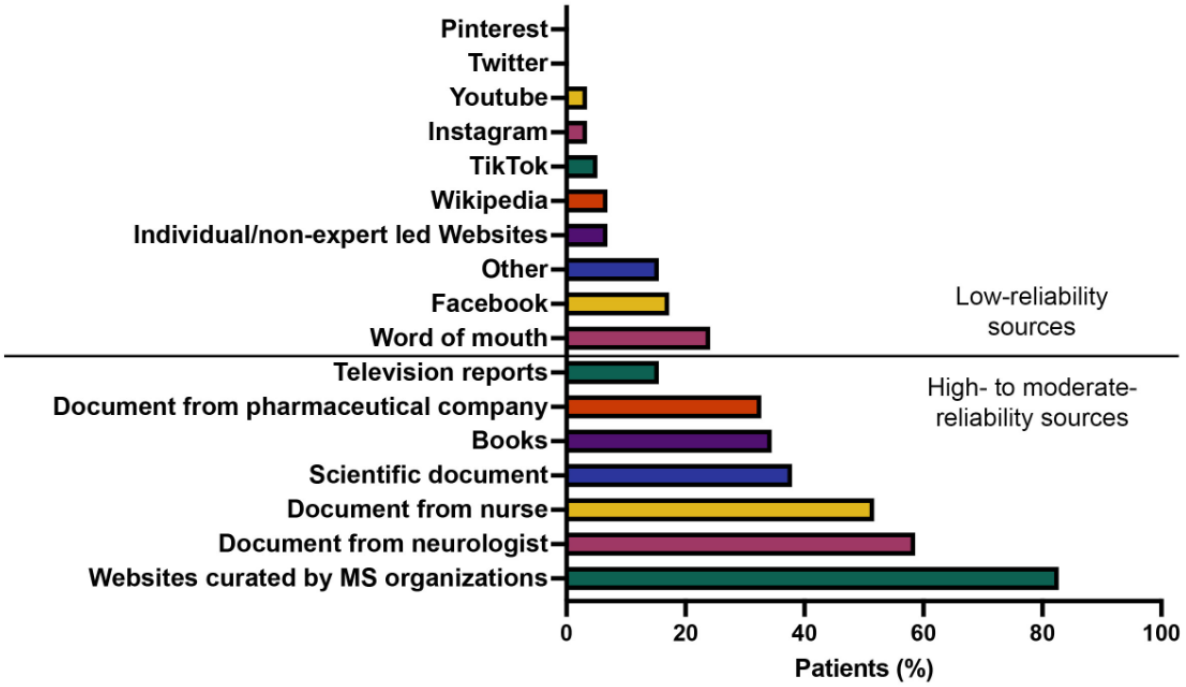

In linear regression analyses, we did not find that scientific score correlated with age, sex, level of education, use of disease-modifying therapies, alternative therapies, nor the use of high- to moderate- or low-reliability sources (Table 2). Misinformation score also did not correlate with age, sex, use of disease-modifying therapies, alternative therapies, or the use of high- to moderate- or low-reliability sources (Table 3). However, we observed that lower levels of education were associated with significantly fewer correct answers in the misinformation score (P=.04).

Table 2. Linear regression between various factors and MS[a] scientific fact score (N=69).

Factors	β (95% CI)	*P* value[b]
Age	–0.3662 (–0.9019 to 0.1695)	.18
Sex[c]	0.037 (–12.23 to 16.52)	.77
Education	–4.093 (–20.26 to 12.08)	.61
Disease-modifying therapy	3.31 (–20.39 to 27.01)	.78
Alternative therapy	6.209 (–5.772 to 18.19)	.30
High to moderate-reliability sources	3.589 (–0.5315 to 7.709)	.09
Low-reliability sources	–0.9313 (–7.487 to 5.624)	.78

[a]MS: multiple sclerosis.

[b]*P* value statistically significant when <.05.

[c]A participant identifying as nonbinary was excluded from the analysis.

XSL•FO
RenderX

Table 3. Linear regression between various factors and MS[a] misinformation score (N=69).

Factors	β (95% CI)	*P* value[b]
Age	–0.1946 (–1.180 to 0.7907)	.69
Sex[c]	0.094 (–37.25 to 16.57)	.45
Education	–30.85 (–60.59 to –1.108)	.04
Disease-modifying therapy	19.07 (–31.77 to 69.91)	.46
Alternative therapy	–4.905 (–30.61 to 20.80)	.70
High to moderate-reliability sources	–0.6062 (–9.445 to 8.233)	.89
Low-reliability sources	0.2161 (–13.85 to 14.28)	.97

[a]MS: multiple sclerosis.

[b]*P* value statistically significant when <.05.

[c]A participant identifying as nonbinary was excluded from the analysis.

Discussion

Principal Findings

We investigated the knowledge and information-seeking behaviors of people living with MS at a specialized MS clinic. Taken together, participants generally correctly identified scientific facts as being true but had greater difficulties identifying misinformation about alternative therapies as false.

Given that participants were followed by a specialized clinic, where education is a cornerstone of care, the high scoring associated with scientific facts was expected. However, even when equipped with education and a team of experts to answer questions and concerns, participants still scored relatively poorly when faced with misinformation. This may indicate a need for more misinformation-geared education but also a certain wariness of health care on the part of participants. Indeed, it was previously reported that although 82% of people with MS gathered information on the web, only 36% discussed this information with their physician [16]. It was proposed that wariness of health care may lead to patients opting out of communicating and discussing information found on the web with health care professionals, which in turn may lead to therapeutic nonadherence. Overall, this may indicate the need to proactively educate patients about misinformation commonly found on the web, and more importantly, talk openly with patients about misinformation in hopes of creating space for them to discuss information found on the web without prejudice. Indeed, people presented with accessible evidence-based corrections can reduce their belief in misinformation [4]. Although participants scored lower on the portion of the questionnaire dealing with misinformation, they more often chose to answer “I do not know.” This cautiousness was also previously documented in people with MS surveyed about web-based searches [14,17]. This healthy skepticism could represent a facilitator when educators are tasked with discrediting misinformation.

In addition, patients often will have difficulty understanding and remembering information communicated by their health care provider [27-29]. Unfortunately, understanding complex health information and applying it to everyday life, that is, health literacy, is pivotal to self-management and improved health outcomes [30,31]. Gaps in understanding may be explained by the overuse of medical terminology while gaps in recall may be due to inadequate educational methods and interventions used by health care professionals [32-34]. Therefore, compounding the issue of misinformation are the challenges that come with educating patients. It has been shown that health care professionals tend to overestimate their own ability to communicate clearly with patients [27-29,35]; thus, being aware of such gaps in communication and implementing new strategies could ameliorate patient understanding of misinformation. There are also many barriers and issues in access to health care. One commonly identified barrier for neurologists in the care of people living with MS is time constraints and availability of appointments [36]. As educational methods may be relatively more time-consuming, implementing change may be challenging [37].

Health literacy is the ability to seek, find, and understand health information on the web to help support decision-making about one’s health [4]. Low health literacy may increase a patient’s susceptibility to misinformation; however, the great majority of people consult low-quality websites for health-related information regardless of their health literacy level [4]. This could be explained by the fact that low-quality sources may be more easily understood or more engaging by eliciting stronger feelings [8]. Health literacy is also correlated with the individual’s educational level [4]. In this study, 81% (56/69) of participants had a high level of education (college or university level diplomas); thus, we could assume that the health literacy of participants was adequate. Furthermore, as MS disproportionally affects female over male individuals, the surveyed population was mostly comprised of the former. In this highly educated group, mostly comprised of female individuals, half had experimented with alternative therapies. In the realm of cancer, those most likely to use alternative medicine are indeed educated female individuals from higher-income households [38]. Furthermore, social media platforms have gained in popularity among health information seekers regardless of age or sex [39]. Consequently, age, sex, education level, and health literacy might not safeguard against web-based misinformation. Herein, we were unable to identify correlations associated with scores obtained on the questionnaire other than educational level. Although the educational level did

XSL•FO
RenderX

seem to impact the misinformation score, this did not stop participants from experimenting with alternative therapies.

The belief that cannabis or diet interventions may be effective treatment alternatives for MS was notable in the surveyed population. This is of some concern as there is little data currently supporting the use of these avenues in the treatment of MS. Although some cannabinoid products may help reduce the severity of spasticity short term, it is still uncertain what the effects are on chronic neurological pain and health-related quality of life [40]. Furthermore, cannabinoids may cause adverse effects leading to treatment discontinuation as well as nervous system and psychiatric disorders [40]. More studies are needed to evaluate the efficacy and safety of using these products. Likewise, the effects of dietary interventions for MS-related outcomes remain uncertain [41]. Misconceptions relating to cannabis use and special diets have previously been identified as prevalent health misinformation for all individuals seeking health information on the web [5], demonstrating that widespread misinformation may affect a broad spectrum of individuals. This may indicate a need for larger-scale antipropaganda government initiatives. Many participants also believed that cooling methods are effective treatments. Although these modalities have demonstrated some effectiveness, no study has assessed comfort and adherence. Moreover, these methods are costly, not accessible, and have not been widely tested [42]**.**

Interestingly, participants in this small cohort mostly used high to moderate web-based sources for their MS-related health information; yet the majority were still subject to misinformation. Other than the lack of misinformation-geared education, this may be explained by the fact that people might not know where they pick up misinformation [4]. For instance, misinformation might cross one's attention while casually scrolling on social media. Belief in misinformation may also be explained by the fact that people with MS will report an improvement in their fatigue and mood with the use of alternative therapies. Hence, it is plausible that study participants who have tried alternative therapies with some positive outcomes may earnestly believe in their efficacy and subsequently be susceptible to misinformation. Unfortunately, these alternative treatments are not without risk, as they may also negatively impact the effectiveness of disease-modifying therapies [25]. Furthermore, although all patients at the specialized MS clinic received disease-related documents to consult with their health care provider, only about half referred to these when searching for information about MS. This may illustrate a need to make health information distributed by health care professionals more engaging and accessible to people with MS, and this for a diverse group of people.

Science and health misinformation can be defined as information that is not in accordance with the leading scientific consensus. This process of delineating what is considered true or false is highly dynamic and is subject to change based on current evidence. MS is a complex disease with many unanswered fundamental questions relating to causation and susceptibility, and although disease-modifying treatments are available for relapsing-remitting MS, there remains a paucity of effective treatments for the progressive forms of MS [43]. Hopes of finding answers to unaddressed questions may increase susceptibility to misinformation, and more importantly, the lack of effective therapeutic avenues in advanced disease may encourage experimenting with alternative medicine. Furthermore, as new data emerges, what is presently considered alternative medicine may become common practice with new research highlighting efficiency and safety. Notably, some recent work has demonstrated that a Mediterranean diet may help alleviate some MS symptoms; however, more robust data are needed to validate these findings [44]. Also, a recent study has demonstrated that the Epstein-Barr virus is linked to MS, but not other previously reported viruses [45]. This highly dynamic process and changing landscape may be challenging not only for people with MS but also for educators. Although studies are exploring novel ways to effectively deal with health misinformation on the web, more research is needed to fully understand this highly complex social phenomenon [4].

Study Limitations

This study is not without its limitations. Notably, the small sample size and the homogenous population may not accurately represent the whole population with MS. Indeed, with only 69 participants, extrapolation of the data to the larger population with MS would be precipitous; larger multicentric cohort studies are needed to better understand this phenomenon. Participants possessed a high level of education, a notable contrast to the 57.2% of Canadians with a similar educational background as reported in 2021, which may indicate that data may not be generalizable [46]. We also assume that education levels equate to health literacy; however, assessing health literacy may have been more informative. Further, participants in this study received education regarding their disease at a specialized MS clinic, suggesting that repeating this study with people with MS who have limited access to a neurologist or clinical team may yield even more worrisome results. Furthermore, some aspects of the questionnaire may have been difficult to understand for people outside of health care or the scientific field. More specifically, delineating the difference between disease-modifying and symptomatic treatments may have been challenging for some. For instance, marijuana and cooling may be considered potentially effective symptomatic treatments but not effective disease-modifying ones, which may have led participants to answer wrongfully. The wording in elements of the questionnaire may also have been confusing; the use of "seem" and "large" in "genetic factors don't seem to play a large role in MS" may have been too vague and thus confusing to respondents. Such misinterpretations may be avoided by using standardized survey tools for assessing susceptibility to misinformation; however, more research is needed to develop such tools in MS. Interestingly, a parallel may also be drawn between questionnaire understandability/clarity and patient education; how do we make information understandable, mitigate the risk of misinterpretation, whilst ensuring that the content remains engaging? In addition, how can we assess if patients have accurately assimilated the content?

In conclusion, we show that people with MS followed at a specialized MS clinic can readily answer questions regarding their disease; however, they are potentially still being impacted by misinformation. As web-based information has the potential

XSL•FO
RenderX

to improve the patient-doctor relationship and aid communication as well as help patients be more actively involved in decision-making, disparaging the use of the internet is not a reasonable solution to mitigate the potentially negative effects of misinformation. Rather, as a first line of defense, we propose that a portion of educational sessions be dedicated to thwarting misinformation as well as opening lines of communication between patients and physicians regarding information found on the web.

Acknowledgments

The authors would like to thank the participants, the Moncton City Hospital Multiple Sclerosis Clinic, and the Centre de formation médicale du Nouveau-Brunswick.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

VD carried out data collection and drafted sections in the manuscript. DC collated and analyzed data, generated figures, and tables, drafted the Results section, and revised the manuscript. TT built the questionnaire and carried out data collection. NB revised and wrote sections of the manuscript. MB helped guide statistical analyses and reviewed the manuscript critically. LW, TJ, NM, and MG facilitated access to participant recruitment. LCW conceptualized and supervised the project and contributed to drafting and revising the manuscript. All authors contributed to the interpretation of results and revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant questionnaire.
[PDF File (Adobe PDF File), 227 KB - jmir_v27i1e63763_app1.pdf]

References

1. Warner D, Procaccino JD. Toward wellness: women seeking health information. J Am Soc Inf Sci 2004;55(8):709-730. [doi: 10.1002/asi.20016]
2. Jacobs W, Amuta A, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
3. Zanatta ET, Wanderley GPDM, Branco IK, Pereira D, Kato LH, Maluf EMCP. Fake news: the impact of the internet on population health. Rev Assoc Med Bras 2021;67(7):926-930 [FREE Full text] [doi: 10.1590/1806-9282.20201151] [Medline: 34817501]
4. Swire-Thompson B, Lazer D. Public health and web-based misinformation: challenges and recommendations. Annu Rev Public Health 2020;41:433-451 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
5. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
6. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
7. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. Asso Info Sci Tech 2015;66(10):2071-2084. [doi: 10.1002/asi.23311]
8. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
9. Sommariva S, Vamos C, Mantzarlis A, Đào LU, Martinez Tyson D. Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study. Am J Health Educ 2018;49(4):246-255. [doi: 10.1080/19325037.2018.1473178]
10. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
11. Rennis L, McNamara G, Seidel EJ, Shneyderman Y. Google it!: urban community college students' use of the internet to obtain self-care and personal health information. Coll Stud J 2013;49(12):414-426 [FREE Full text]
12. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. BMJ 2007;335(7609):24-27 [FREE Full text] [doi: 10.1136/bmj.39246.581169.80] [Medline: 17615222]

XSL•FO
RenderX

13. Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
14. Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR. Preferred sources of health information in persons with multiple sclerosis: degree of trust and information sought. J Med Internet Res 2013;15(4):e67 [FREE Full text] [doi: 10.2196/jmir.2466] [Medline: 23635393]
15. Lejbkowicz I, Paperna T, Stein N, Dishon S, Miller A. Internet usage by patients with multiple sclerosis: implications to participatory medicine and personalized healthcare. Mult Scler Int 2010;2010:640749 [FREE Full text] [doi: 10.1155/2010/640749] [Medline: 22096625]
16. Hay MC, Strathmann C, Lieber E, Wick K, Giesser B. Why patients go online: multiple sclerosis, the internet, and physician-patient communication. Neurologist 2008;14(6):374-381. [doi: 10.1097/NRL.0b013e31817709bb] [Medline: 19008743]
17. Synnot AJ, Hill SJ, Garner KA, Summers MP, Filippini G, Osborne RH, et al. Online health information seeking: how people with multiple sclerosis find, assess and integrate treatment information to manage their health. Health Expect 2016;19(3):727-737 [FREE Full text] [doi: 10.1111/hex.12253] [Medline: 25165024]
18. Brigo F, Lochner P, Tezzon F, Nardone R. Web search behavior for multiple sclerosis: an infodemiological study. Mult Scler Relat Disord 2014;3(4):440-443. [doi: 10.1016/j.msard.2014.02.005] [Medline: 25877054]
19. Lavorgna L, Russo A, de Stefano M, Lanzillo R, Esposito S, Moshtari F, et al. Health-related coping and social interaction in people with multiple sclerosis supported by a social network: pilot study with a new methodological approach. Interact J Med Res 2017;6(2):e10 [FREE Full text] [doi: 10.2196/ijmr.7402] [Medline: 28710056]
20. Altunisik E, Firat YE, Kiyak Keceli Y. Content and quality analysis of videos about multiple sclerosis on social media: The case of YouTube. Mult Scler Relat Disord 2022;65:104024. [doi: 10.1016/j.msard.2022.104024] [Medline: 35841755]
21. Lavorgna L, Lanzillo R, Brescia Morra V, Abbadessa G, Tedeschi G, Bonavita S. Social media and multiple sclerosis in the posttruth age. Interact J Med Res 2017;6(2):e18 [FREE Full text] [doi: 10.2196/ijmr.7879] [Medline: 28954717]
22. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. Mult Scler 2020;26(14):1816-1821 [FREE Full text] [doi: 10.1177/1352458520970841] [Medline: 33174475]
23. Bowling AC. Complementary and alternative medicine and multiple sclerosis. Neurol Clin 2011;29(2):465-480. [doi: 10.1016/j.ncl.2010.12.006] [Medline: 21439454]
24. Apel-Neu A, Zettl UK. Complementary and alternative medicine in multiple sclerosis. J Neurol 2008;255:82-86. [doi: 10.1007/s00415-008-6015-9] [Medline: 19300965]
25. Podlecka-Piętowska A, Sugalska M, Janiszewska K, Wall-Szczech A, Cyganek A, Szejko N, et al. Complementary and alternative medicine in multiple sclerosis: a questionnaire-based study. Neurol Neurochir Pol 2022;56(5):428-434 [FREE Full text] [doi: 10.5603/PJNNS.a2022.0059] [Medline: 36069151]
26. Namjooyan F, Ghanavati R, Majdinasab N, Jokari S, Janbozorgi M. Uses of complementary and alternative medicine in multiple sclerosis. J Tradit Complement Med 2014;4(3):145-152 [FREE Full text] [doi: 10.4103/2225-4110.136543] [Medline: 25161918]
27. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
28. Coran JJ, Koropeckyj-Cox T, Arnold CL. Are physicians and patients in agreement? Exploring dyadic concordance. Health Educ Behav 2013;40(5):603-611. [doi: 10.1177/1090198112473102] [Medline: 23345336]
29. Tongue J, Epps H, Forese L. Communication skills for patient-centered care : research-based, easily learned techniques for medical interviews that benefit orthopaedic surgeons and their patients. J Bone Joint Surg 2005;87(3):652-658. [doi: 10.2106/00004623-200503000-00027]
30. Laws MB, Lee Y, Taubin T, Rogers WH, Wilson IB. Factors associated with patient recall of key information in ambulatory specialty care visits: results of an innovative methodology. PLoS One 2018;13(2):e0191940 [FREE Full text] [doi: 10.1371/journal.pone.0191940] [Medline: 29389994]
31. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
32. Chou L, Ellis L, Papandony M, Seneviwickrama KLMD, Cicuttini FM, Sullivan K, et al. Patients' perceived needs of osteoarthritis health information: a systematic scoping review. PLoS One 2018;13(4):e0195489 [FREE Full text] [doi: 10.1371/journal.pone.0195489] [Medline: 29659609]
33. Richard C, Glaser E, Lussier M. Communication and patient participation influencing patient recall of treatment discussions. Health Expect 2017;20(4):760-770 [FREE Full text] [doi: 10.1111/hex.12515] [Medline: 27868327]
34. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003;96(5):219-222 [FREE Full text] [doi: 10.1177/014107680309600504] [Medline: 12724430]
35. Olson DP, Windish DM. Communication discrepancies between physicians and hospitalized patients. Arch Intern Med 2010;170(15):1302-1307. [doi: 10.1001/archinternmed.2010.239] [Medline: 20696951]

XSL•FO
RenderX

36. Price E, Lucas R, Lane J. Experiences of healthcare for people living with multiple sclerosis and their healthcare professionals. Health Expect 2021;24(6):2047-2056 [FREE Full text] [doi: 10.1111/hex.13348] [Medline: 34480516]
37. Talevski J, Wong Shee A, Rasmussen B, Kemp G, Beauchamp A. Teach-back: A systematic review of implementation and impacts. PLoS One 2020;15(4):e0231350 [FREE Full text] [doi: 10.1371/journal.pone.0231350] [Medline: 32287296]
38. Johnson SB, Park HS, Gross CP, Yu JB. Use of alternative medicine for cancer and its impact on survival. J Natl Cancer Inst 2018;110(1):121-124. [doi: 10.1093/jnci/djx145] [Medline: 28922780]
39. Xiong F, Liu Y. Opinion formation on social media: an empirical approach. Chaos 2014;24(1):013130. [doi: 10.1063/1.4866011] [Medline: 24697392]
40. Filippini G, Minozzi S, Borrelli F, Cinquini M, Dwan K. Cannabis and cannabinoids for symptomatic treatment for people with multiple sclerosis. Cochrane Database Syst Rev 2022;5(5):CD013444 [FREE Full text] [doi: 10.1002/14651858.CD013444.pub2] [Medline: 35510826]
41. Parks NE, Jackson-Tarlton CS, Vacchi L, Merdad R, Johnston BC. Dietary interventions for multiple sclerosis-related outcomes. Cochrane Database Syst Rev 2020;5(5):CD004192 [FREE Full text] [doi: 10.1002/14651858.CD004192.pub4] [Medline: 32428983]
42. Stevens CJ, Singh G, Peterson B, Vargas NT, Périard JD. The effect of cooling garments to improve physical function in people with multiple sclerosis: a systematic review and meta-analysis. Mult Scler Relat Disord 2023;78:104912 [FREE Full text] [doi: 10.1016/j.msard.2023.104912] [Medline: 37499339]
43. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. Lancet 2018;391(10130):1622-1636 [FREE Full text] [doi: 10.1016/S0140-6736(18)30481-1] [Medline: 29576504]
44. Snetselaar LG, Cheek JJ, Fox SS, Healy HS, Schweizer ML, Bao W, et al. Efficacy of diet on fatigue and quality of life in multiple sclerosis: a systematic review and network meta-analysis of randomized trials. Neurology 2023;100(4):e357-e366. [doi: 10.1212/WNL.0000000000201371] [Medline: 36257717]
45. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science 2022;375(6578):296-301 [FREE Full text] [doi: 10.1126/science.abj8222] [Medline: 35025605]
46. [Highest level of education by census year: Canada, provinces and territories, census metropolitan areas and census agglomerations]. Statistics Canada. 2022. URL: https://www150.statcan.gc.ca/t1/tbl1/fr/tv.action?pid=9810038401 [accessed 2023-11-13]

Abbreviations

MS: multiple sclerosis

Edited by A Mavragani; submitted 28.06.24; peer-reviewed by R Haase, E Leray; comments to author 20.09.24; revised version received 08.10.24; accepted 07.11.24; published 25.02.25.

Please cite as:
Duguay V, Comeau D, Turgeon T, Bouhamdani N, Belanger M, Weston L, Johnson T, Manzer N, Giberson M, Chamard-Witkowski L
Evaluating the Knowledge and Information-Seeking Behaviors of People Living With Multiple Sclerosis: Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e63763
URL: https://www.jmir.org/2025/1/e63763
doi:10.2196/63763
PMID:39998866

©Véronique Duguay, Dominique Comeau, Tiffany Turgeon, Nadia Bouhamdani, Mathieu Belanger, Lyle Weston, Tammy Johnson, Nicole Manzer, Melissa Giberson, Ludivine Chamard-Witkowski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Complete Blood Count and Monocyte Distribution Width–Based Machine Learning Algorithms for Sepsis Detection: Multicentric Development and External Validation Study

Andrea Campagner[1*], PhD; Luisa Agnello[2*], PhD; Anna Carobene[3], PhD; Andrea Padoan[4,5], Prof Dr; Fabio Del Ben[6], PhD; Massimo Locatelli[3], Dr med; Mario Plebani[4,5], Prof Dr; Agostino Ognibene[7], Dr med; Maria Lorubbio[7], Dr med; Elena De Vecchi[1], PhD; Andrea Cortegiani[2,8], Prof Dr; Elisa Piva[9], Dr med; Donatella Poz[10], Dr med; Francesco Curcio[10], Prof Dr; Federico Cabitza[1,11*], Prof Dr; Marcello Ciaccio[2,8*], Prof Dr

[1]IRCCS Ospedale Galeazzi Sant'Ambrogio, Milan, Italy
[2]University of Palermo, Palermo, Italy
[3]IRCCS San Raffaele Scientific Institute, Milano, Italy
[4]Department of Medicine, University of Padova, Padova, Italy
[5]Laboratory Medicine Unit, University-Hospital of Padova, Padova, Italy
[6]IRCCS Centro Di Riferimento Oncologico Aviano, Aviano, Italy
[7]San Donato Hospital of Arezzo, Arezzo, Italy
[8]University Hospital Policlinico Paolo Giaccone, Palermo, Italy
[9]Azienda Socio Sanitaria Territoriale di Mantova, Mantova, Italy
[10]University Hospital of Udine, Udine, Italy
[11]Department of Computer Science, Systems and Communication, University of Milano-Bicocca, Milano, Italy
[*]these authors contributed equally

Corresponding Author:
Federico Cabitza, Prof Dr
Department of Computer Science, Systems and Communication
University of Milano-Bicocca
Piazza dell'Ateneo Nuovo, 1
Milano, 20126
Italy
Phone: 39 0264487888
Email: federico.cabitza@unimib.it

Abstract

Background: Sepsis is an organ dysfunction caused by a dysregulated host response to infection. Early detection is fundamental to improving the patient outcome. Laboratory medicine can play a crucial role by providing biomarkers whose alteration can be detected before the onset of clinical signs and symptoms. In particular, the relevance of monocyte distribution width (MDW) as a sepsis biomarker has emerged in the previous decade. However, despite encouraging results, MDW has poor sensitivity and positive predictive value when compared to other biomarkers.

Objective: This study aims to investigate the use of machine learning (ML) to overcome the limitations mentioned earlier by combining different parameters and therefore improving sepsis detection. However, making ML models function in clinical practice may be problematic, as their performance may suffer when deployed in contexts other than the research environment. In fact, even widely used commercially available models have been demonstrated to generalize poorly in out-of-distribution scenarios.

Methods: In this multicentric study, we developed ML models whose intended use is the early detection of sepsis on the basis of MDW and complete blood count parameters. In total, data from 6 patient cohorts (encompassing 5344 patients) collected at 5 different Italian hospitals were used to train and externally validate ML models. The models were trained on a patient cohort encompassing patients enrolled at the emergency department, and it was externally validated on 5 different cohorts encompassing patients enrolled at both the emergency department and the intensive care unit. The cohorts were selected to exhibit a variety of data distribution shifts compared to the training set, including label, covariate, and missing data shifts, enabling a conservative

XSL•FO
RenderX

validation of the developed models. To improve generalizability and robustness to different types of distribution shifts, the developed ML models combine traditional methodologies with advanced techniques inspired by controllable artificial intelligence (AI), namely cautious classification, which gives the ML models the ability to abstain from making predictions, and explainable AI, which provides health operators with useful information about the models' functioning.

Results: The developed models achieved good performance on the internal validation (area under the receiver operating characteristic curve between 0.91 and 0.98), as well as consistent generalization performance across the external validation datasets (area under the receiver operating characteristic curve between 0.75 and 0.95), outperforming baseline biomarkers and state-of-the-art ML models for sepsis detection. Controllable AI techniques were further able to improve performance and were used to derive an interpretable set of diagnostic rules.

Conclusions: Our findings demonstrate how controllable AI approaches based on complete blood count and MDW may be used for the early detection of sepsis while also demonstrating how the proposed methodology can be used to develop ML models that are more resistant to different types of data distribution shifts.

(J Med Internet Res 2025;27:e55492) doi:10.2196/55492

KEYWORDS

sepsis; medical machine learning; external validation; complete blood count; controllable AI; machine learning; artificial intelligence; development study; validation study; organ; organ dysfunction; detection; clinical signs; clinical symptoms; biomarker; diagnostic; machine learning model; sepsis detection; early detection; data distribution

Introduction

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection [1]. The progression from infection to sepsis is influenced by both host and pathogen features and, accordingly, is characterized by a high interindividual variability. Sepsis represents a medical emergency, with a strong time dependency, especially for septic shock [2,3]. Therefore, early detection is fundamental to improving the patient outcome by promptly starting treatment. However, sepsis, especially in the early stage, may be characterized by nonspecific signs and symptoms, mainly when diagnosis of infection is still uncertain. Thus, its early recognition is challenging, particularly in the emergency department (ED) and outside the intensive care unit (ICU), due to the lack of data needed to apply the Sepsis-3 criteria (ie, assessment of organ dysfunction due to Sequential Organ Failure Assessment Score [SOFA], plus probable or certain infection). Therefore, reliable tools to screen sepsis are strongly sought after. The UK National Screening Committee underlines that the aim of screening is to propose a test to a population who "do not necessarily perceive that they are at risk of, or are already affected by, a disease or its complications" [4]. In the scenario of sepsis screening, laboratory medicine could have a crucial role by providing biomarkers mirroring the pathophysiological mechanisms underpinning sepsis and whose alteration could be detected before the onset of clinical signs and symptoms [5].

To date, hundreds of biomarkers have been investigated, but several limitations have hampered the translation from research to clinical practice, such as cost and the request for additional blood samples [6]. Intensive scientific research is ongoing to discover accurate, affordable, easy-to-measure, and rapidly available biomarkers for guiding clinicians to identify patients at higher risk of sepsis, allowing close monitoring or early treatment.

In clinical practice, C-reactive protein (CRP) and procalcitonin are the most used sepsis biomarkers. However, they are not specific to sepsis, and their levels can increase in several clinical, noninfectious conditions. In addition, they are usually ordered when sepsis is clinically suspected, leading to a diagnostic delay [7] and making them unsuitable as a screening tool. The ideal biomarker for sepsis screening should have the following characteristics: (1) easy to measure, (2) high sensitivity and negative predictive value for sepsis, (3) low turn-around time, and (4) always available to clinicians, especially when sepsis is not (yet) suspected. Complete blood count (CBC) parameters fulfill all of these features [8]. Indeed, CBC is the first-level laboratory test most ordered in all clinical wards, from ED to ICU. It is easy to perform, cheap, and provides important clinical information on the health status. Beyond the classic CBC parameters, the new generation of hemocytometers provides cell population data, which reflect the morphological and functional characteristics of neutrophils, lymphocytes, and monocytes. In the last decade, the role for monocyte distribution width (MDW) as a sepsis biomarker has emerged [9]. MDW reflects the dispersion around the mean monocyte volume. It is automatically calculated by the last generation DxH hematology analyzer (Beckman Coulter, Inc) based on volume, conductivity, and scattering technology and is provided together with the basic CBC parameters. Noteworthy, MDW received approval from the Food and Drug Administration and European Community In-Vitro Diagnostic Medical Device as an early sepsis indicator in adult patients in the ED. MDW has great potential to be used in clinical practice because it is timely and fruitfully measured without further samples, additional costs, or even a different request from clinicians. Thus, it represents a cost-effective tool for early sepsis screening. This is an important feature in the health care of the third millennium. To date, several authors have shown good performance of MDW for early diagnosis of patients with sepsis in different clinical wards, especially the ED and ICU [10-17]. All studies have shown excellent specificity and negative predictive value (NPV) but suboptimal sensitivity and positive predictive value (PPV) [9]. Thus, while individual CBC biomarkers and MDW represent

XSL•FO
RenderX

ideal candidates for sepsis screening in general settings, their performance should be improved to be applied for sepsis screening.

Machine learning (ML) techniques offer the promise to potentially offset the limitations mentioned earlier by combining different parameters, beyond the capabilities of rule-based systems and humans, and increasing the performance of single laboratory parameters [18,19]. Indeed, in recent years, ML methods have been increasingly adopted in health care settings [20], including sepsis screening and prediction [21-24]. However, as also noted in the study by Shashikumar et al [25], most existing ML sepsis models have been developed and validated on data from a single hospital only, with no external validation. Thus, their robustness has not been proven, and their performance may degrade when applied in settings that differ from the one in which they are developed, due to either data or missingness shifts [26], limiting their generalizability and adoption in clinical practice. Indeed, widely applied commercially available models (eg, Epic) have been shown to be severely impacted by the aforementioned issues [27-30]. This is especially true for models based on laboratory parameters, wherein state-of-the-art models have not been externally validated [31]. Most notably, no previous work, to the best of our knowledge, combined the use of ML methods with MDW [32].

According to these premises, the aim of this study is to develop and externally validate ML models whose intended use is the early detection and screening of sepsis on the basis of CBC parameters, including MDW and other clinical and laboratory features commonly available in all clinical laboratories. In this multicentric study we used CBC data from an Italian hospital ED (training set) to develop ML models, by adopting both traditional ML models as well as more advanced techniques inspired by controllable artificial intelligence (AI) [33] (ie, cautious classification and explainable AI) that were successively validated by data from multiple settings (encompassing both ED and ICU settings, as well as different forms of data shifts) and different Italian hospitals, to evaluate their robustness and generalization.

Methods

Overview

Herein, we briefly summarize the adopted methodology, which is better detailed in the following sections. The reporting follows the Checklist for Assessment of Medical AI checklist [34] and the "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research" [35].

The study was conducted as a multicentric, retrospective study aimed at developing ML models for the classification problem of sepsis detection from clinical laboratory parameters. To this aim, ML models were trained on data from 1 patient cohort (Palermo, Emergency Department [PA-ED], coming from Palermo, Italy) and evaluated on 5 different patient cohorts (Palermo, Intensive Care Unit [PA-ICU], Padova, Intensive Care Unit [PD-ICU], Udine, Emergency Department [UD-ED], Arezzo, Emergency Department [AR-ED], and Ospedale Galeazzi Sant'Ambrogio, Intensive Care Unit [OGSA-ICU], coming from 5 different hospitals in Italy, respectively from Palermo, Padova, Udine, Arezzo, and Milano [refer to Dataset Collection, Description, and Processing section]), encompassing 5344 records for as many single patients. We decided to consider 5 different external validation cohorts to comprehensively assess the generalizability of the developed models, as well as their robustness to multiple forms of data variation (refer to the subsequent paragraphs). The 6 patient cohorts were collected at 5 different hospitals: 3 cohorts (PA-ED, UD-ED, and AR-ED) were from the ED setting, while the remaining 3 (PA-ICU, PD-ICU, and OGSA-ICU) were from the ICU setting. The PA-ED, PA-ICU, PD-ICU, UD-ED, and AR-ED datasets encompassed 19 features, including the following: demographic information (age and sex), the MDW, other laboratory test results (ie, the CBC, including the differential, and the CRP), together with a derived feature (the neutrophils-lymphocytes ratio). The OGSA-ICU dataset encompassed the same features as in the other cohorts, except for the MDW feature that was missing. The PA-ED cohort was used for training, hyperparameter optimization, and internal validation, while all the other cohorts were used only for external validation. The 5 external datasets allowed for the validation of the robustness of the developed models under different shift settings, including covariate shifts due to different clinical settings (ICU vs ED), label shifts (sepsis-2 vs sepsis-3), missingness shifts [26], and instantial variation [36]. In particular, the AR-ED cohort was used to evaluate the robustness of the developed models in a conservative scenario of label shift (refer to the Dataset Collection, Description, and Processing section), as the population of this dataset only encompassed patients who exhibited clinical signs for suspected sepsis. By contrast, the OGSA-ICU cohort was used to test the performance and robustness of non-MDW systems (that integrate the ML models described earlier) that can also be applied in settings that lack the MDW parameter; indeed, the laboratory parameters for the OGSA-ICU dataset did not contain the MDW feature as they were collected using a Sysmex (rather than the one by Beckmann Coulter, Inc) analyzer. Thus, this cohort was characterized by missingness shift [26] (ie, differences in the distribution of missing values) and differing analytical variation [36] (refer to the Non-MDW Systems Validation section). Notably, no dataset pooling was performed to ensure that the validation sets were entirely blind and guarantee an unbiased estimate of performance and generalizability; hence, each dataset was considered as an entirely separate cohort.

The PA-ED was split into 2 datasets according to a 75/25 split: the largest 75% set was used for training and hyperparameter optimization in cross-validation, while the remaining 25% set was used for internal validation. The split was stratified to preserve the proportion of sepsis and no sepsis cases in the 2 splits of the dataset. Further details on the models' design and development can be found in the Methods section. We considered 5 different classes of models, namely logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), and decision trees (DTs). The performance of the models was evaluated according to different criteria, namely: area under the receiver operating characteristic curve (AUC), sensitivity, specificity, PPV, NPV,

XSL•FO
RenderX

and average PPV (A-PPV; as measures of discrimination power), Brier score (as a measure of calibration), and the standardized net benefit (sNB, as a measure of utility). Following the state-of-the-art development of medical decision support models for sepsis diagnosis [25], and as a way to increase generalizability and reduce overconfident predictions according to the tenets of controllable AI [33], we also considered *cautious classifiers* based on the trained models (refer to Cautious Classification Methods section). Such models can *abstain* from labeling instances, with the aim of implementing a trade-off between coverage (labels for which a prediction is available) and performance metrics (in particular, sensitivity and PPV). The cautious classifiers were evaluated in terms of their coverage, as well as with respect to the high-confidence (HC) metrics, which simply consider the metrics defined earlier only on the subset of the nonabstained-upon cases.

Dataset Collection, Description, and Processing

The PA-ED data refer to patients who were admitted to the ED of the University Hospital "Policlinico Paolo Giaccone" between September, 4, 2019, and November 30, 2019, as described in the study by Agnello [37]. The inclusion criteria were adult (aged ≥18 years) patients presenting to the ED, and whose initial evaluation included a CBC. Exclusion criteria were (1) aged <18 years, (2) incomplete data collection, (3) discharge from the ED within 2 hours from the ED presentation, (4) readmission to the ED within 12 hours, (5) inadequate blood samples (eg, analyzed >4 hours after collection), (6) failure to determine the MDW parameter, and (7) diagnosis of a hematological disorder. The sepsis diagnosis was made according to sepsis-2 criteria [38]. Demographical, clinical, and laboratory data of patients were recorded from the electronic health records of the hospital. Laboratory parameters were measured upon ED admission. MDW and CBC parameters were measured on blood samples collected in the EDTA-K3 tube by a UniCel DxH 900 hematology analyzer (Beckman Coulter, Inc) within 2 hours of collection, as recommended by the manufacturer, after the laboratory analysis ordered by clinicians was performed. CRP and procalcitonin were measured on serum by a fully automated platform (Cobas 8000, Roche Diagnostics).

The PD-ICU data refer to patients who were admitted to the ICU of the University Hospital of Padova and were collected between January 13, 2020 and April 8, 2020. The inclusion criteria were adult (aged >18 years) patients presenting to the ICU, who remained hospitalized for at least 24 hours, enrolled no more than once, with a CBC and differential testing performed at presentation and over the entire course of the length of stay, as part of standard medical care, and procalcitonin or CRP tests ordered at the same time. Exclusion criteria were an incomplete data collection, a failure to determine MDW at least once, and patients with hematological disorders. Sepsis diagnosis was made according to the sepsis-3 criteria, using the SOFA score. Septic shock was recognized in the cases with hypotension requiring vasopressors and with an elevated serum lactate (>2 mmol/L).

The PA-ICU data refer to patients who were admitted to the ICU of the University Hospital "P. Giaccone" between September 10, 2019, and March 19, 2020, as previously described in the study by Agnello et al [8]. The inclusion criteria were all consecutive adult (aged ≥18 years) patients who were admitted to the ICU. Exclusion criteria were (1) individuals aged <18 years; (2) incomplete data collection; (3) failure to determine the MDW parameter; and (4) underlying conditions potentially associated with deregulation of the immune system, including AIDS, organ or bone marrow transplantation, and hematologic diseases. Sepsis diagnosis was made according to sepsis-3 consensus criteria [1]. Laboratory parameters were measured upon ICU admission. MDW and CBC parameters were measured on blood samples collected in the EDTA-K3 tube by a UniCel DxH 900 hematology analyzer (Beckman Coulter, Inc) within 2 hours of collection, as recommended by the manufacturer, after the laboratory analysis ordered by clinicians was performed. CRP and procalcitonin were measured on serum by a fully automated platform (Cobas 8000, Roche Diagnostics).

The AR-ED data refer to patients who were admitted in the ED of the San Donato Hospital and were collected between July 5, 2019, and April 17, 2020. The inclusion criteria were all consecutive adult (aged ≥18 years) patients who were admitted to the ED and had a clinical presentation of suspected sepsis (at least 2 of the following signs: SOFA≥2, body temperature ≥38 °C or ≤36 °C, infection, respiratory rate ≥20 beats/min, heart rate ≥90 beats/min, or altered mental status). Sepsis diagnosis was made according to the sepsis-3 criteria. MDW and CBC parameters were measured by a UniCel DxH 900 hematology analyzer (Beckman Coulter, Inc) within 2 hours of collection and directly extracted from the analyzer. CRP was measured by a Cobas 8000 instrument (Roche Diagnostics) and manually extracted from the hospital's laboratory information system (LIS).

The OGSA-ICU data refer to patients who were admitted to the ICU of the Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Galeazzi Sant'Ambrogio between January 1, 2023 and April 31, 2023. The inclusion criteria were adult (aged ≥18 years) patients who were hospitalized in the ICU for at least 2 days (to exclude patients who were temporarily admitted to the ICU because of specific intervention protocols) and for whom blood culture and procalcitonin were requested. All hematological parameters were measured by a Sysmex XN2000 blood analyzer and manually extracted from the hospital's LIS (BeSimpleLIS, Siemens SpA), using QlikView (Qlik, Inc). CRP and procalcitonin were measured by the Atellica IM Analyzer (Siemens) and manually extracted from the hospital's LIS using QlikView.

The UD-ED data refer to patients who were admitted to the ED of the University Hospital of Udine between November 1, 2019, and December 31, 2019. The inclusion criteria were adult (aged ≥18 years) patients who were enrolled in the ED and had a CBC exam (with differential) requested as part of the standard diagnostic pathway. Sepsis diagnosis was made according to the sepsis-2 criteria. MDW and CBC parameters were measured by a UniCel DxH 900 hematology analyzer (Beckman Coulter, Inc), and directly extracted from the analyzer. CRP was measured by the Cobas C analyzer (Roche Diagnostics) and manually extracted from the hospital's LIS.

XSL•FO
RenderX

Model Design and Development

We considered 5 different classes of statistical machine learning models, namely LR, SVM, RF, XGB, and DTs. These model classes were selected as they exhibited state-of-the-art in-sample [39] and generalization [40] performance on diagnostic tasks from laboratory data in previous work; particularly, we decided to not focus on deep learning models due to several limitations of these latter models on tabular data tasks [41-43], such as requiring extensive hyperparameter optimization and having subpar generalization performance.

The ML models were implemented in Python, using the scikit-learn library (version 1.0.2; Python Software Foundation). Particularly, all models were implemented as 3-step pipelines encompassing a feature scaling step (the scaling technique was selected as a hyperparameter during model training), a feature selection step (using the recursive feature elimination method with an LR model for feature scoring), and the classification step. The full set of hyperparameters, for all the evaluated models, is available in Table S1 in Multimedia Appendix 1. Particularly, except for model-specific hyperparameters, the set of common hyperparameters encompassed the scaling method to be used, the number of features to be selected, and class weights to manage label imbalance.

Model training, hyperparameter selection, and internal model validation were performed on the PA-ED dataset. Specifically, model training and hyperparameter selection were performed on the training fold of the dataset (which encompassed 75% of the total number of cases), while internal validation was performed on the test fold of the same dataset (which encompassed 25% of the total number of cases). Hyperparameter selection was performed by means of randomized search, with a budget of 1000 evaluations, in a 5-fold stratified cross-validation design (RandomizedSearchCV). The training set was split into 5 folds (each containing 20% of the cases in the training data and 15% of the total number of cases in the PA-ED dataset). Following this, for each of the 1000 randomized search evaluations, a random set of hyperparameters was selected and then the models with the specified hyperparameter values were evaluated by means of a 5-fold cross-validation procedure by which they were trained on 4 of the cross-validation folds and evaluated on the remaining fold. Stratified cross-validation was used to preserve the proportion of sepsis and no sepsis cases across all folds. Each evaluated hyperparameter configuration was scored by the average AUC score across the 5 cross-validation folds. For each model class, we selected the hyperparameter configuration that had the highest average AUC score.

After hyperparameter selection, the models were retrained on the entire training set and evaluated on the hold-out test set according to different evaluation criteria, namely sensitivity, specificity, PPV, NPV, AUC, A-PPV, Brier score, and sNB (with a threshold equal to 0.5). On the basis of the internal test set, we compared the developed ML models with 4 different baselines: a fixed MDW threshold (threshold value 23.5 [37]), a fixed CRP threshold (threshold value 80 [44]), a fixed Sepsis Index threshold (threshold value 1 [45]), and the state-of-the-art SVM model (MindraySVM) for sepsis diagnosis based on laboratory biomarkers proposed in the study by Aguirre and Urrechaga [31]. In regard to the latter model, due to the different laboratory instrumentation used in our study, we trained an SVM model using the hyperparameter settings specified in the study by Aguirre and Urrechaga [31] on the same training data used for our models. We also compared the developed models with the state-of-the-art Conformal Multidimensional Prediction of Sepsis Risk (COMPOSER) model proposed in the study by Shashikumar et al [25].

External Validation

The developed ML models were subsequently externally validated, after retraining on the entire PA-ED dataset, on the 4 external validation cohorts (AR-ED, PA-ICU, PD-ICU, and UD-ED), in terms of the same evaluation criteria used for internal validation. As described earlier, the considered external validation allowed us to evaluate the generalizability of the developed ML models in a variety of settings that differed in terms of hospital (4 different hospitals across different regions in Italy), reference population (ED vs ICU), as well as diagnostic criteria (sepsis-2 vs sepsis-3). Therefore, we decided to keep the external validation datasets separate from the internal training set (PA-ED) and did not perform any data pooling. In particular, the choice to consider 4 external validation datasets (as well as an additional external dataset for validation in settings with missing MDW) enabled us to comprehensively assess the robustness of the developed models under a varied selection of distribution shifts. We also evaluated the similarity between the PA-ED cohort and the 5 external validation cohorts, using the degree of correspondence metric (ψ) proposed in the study by Cabitza et al [46]. The ψ metric is a similarity measure used to compare 2 datasets and provides a P value for the hypothesis test that 2 datasets come from the same distribution. The ψ metric has been used as a way to evaluate the robustness of ML models in the context of out-of-domain generalization [40] by assessing the ability of ML models to report good performance also on datasets with low ψ value. We used the implementation of the ψ metric provided [47]. In the paper we used the ψ metric to assess the similarity between the training set (PA-ED) and the external validation cohorts, as well as how the performance of the models varied with respect to the value of ψ.

Validation of Non-MDW Systems

Because the MDW feature was not available in the OGSA-ICU cohort (missingness rate: 100%), we developed non-MDW ML systems for use in settings that lack the MDW parameter. These systems encompassed the ML models trained on the PA-ED cohort, without any retraining, as well as a preprocessing imputation model used to fill in the missing MDW values. The imputation model was derived from the data from the other cohorts (PA-ED, AR-ED, PA-ICU, PD-ICU, and UD-ED), which were used as a training set. The MDW was considered the target feature, while all the other features were used as predictors. The imputation model was an RF regression model with default hyperparameters and was implemented in Python using the scikit-learn library (version 1.0.2). To evaluate the goodness of fit of the imputation model, the training set was split into folds according to a 75/25 split: the 75% was used to

XSL•FO
RenderX

fit the imputation model, while the remaining 25% was used for model evaluation. The performance of the imputation model was assessed by means of the adjusted R^2 score computed between the predicted and true MDW values. After assessment, the imputation model was retrained on the entire training dataset and applied as a preprocessing step to the OGSA-ICU dataset to impute the MDW values for subsequent analysis.

The non-MDW systems were then evaluated by adopting the same methodology described in the Model Design and Development section on the OGSA-ICU cohort. Specifically, we simply applied the developed ML models to the imputed data from the OGSA-ICU cohort.

Cautious Classification Methods

To enable the models to detect a control loss [33] and abstain from providing a diagnosis, to control error rates and reduce misdiagnosis, we implemented cautious classification models based on the ML models trained on the PA-ED cohort. Cautious classification refers to a set of uncertainty quantification techniques that enable an ML model to abstain from providing a prediction, for a given case, whenever the model is not sufficiently confident [48]. Cautious classification has been successfully applied for error rate control in recent work [49,50], also in the context of sepsis diagnosis [25], and provides a simple approach to improve the performance and perceived utility of an ML model. Specifically, we adopted a simple, thresholding-based strategy to implement the cautious classification models: given any new case, the cautious classifiers provided the same diagnosis as the corresponding ML model if the confidence provided by this latter was >0.75; otherwise, it *abstained* from providing a diagnosis. Thus, for a given case, if the ML model would associate with the sepsis class a confidence score larger or equal to 0.75, the cautious classifier would also classify that case as sepsis; similarly, if the confidence score for sepsis was <0.25 (ie, the confidence score for the absence of sepsis was >0.75), the cautious classifier would classify the case as no sepsis; finally, if the confidence score was between 0.25 and 0.75, then the cautious classifier would be considered as abstaining from providing a diagnosis. We decided to apply this cautious classification strategy, rather than other approaches, such as 3-way decision [51] or conformal prediction [52], due to its increased efficiency (the computational complexity cost of the thresholding strategy is O(1), while it is on the order of O(log n), for n being the dataset size, for conformal prediction), ease of interpretation, and its equivalence, in the binary classification setting and under weak assumptions, with the 2 aforementioned methods [51]. Regarding the implementation of the cautious classifiers, we used the HC metrics implemented in the scikit-cautious library [53], and applied them to the ML models trained on the PA-ED dataset, setting the threshold hyperparameter (th) to 0.75. The HC metrics evaluate the performance of the models on the set of nonabstained cases; hence, they select the cases for which either the confidence score of sepsis or that of no sepsis is >0.75. We considered, particularly, the HC metrics for PPV, NPV, sensitivity, specificity, AUC, A-PPV, Brier score, and sNB, as well as the coverage (the percentage of nonabstained-upon cases).

Statistical Analysis

For all continuous variables, we reported mean and SD, while for all categorical variables, we reported the frequency of each value. The sNB was computed at the fixed threshold value of 0.5. CIs for sensitivity, specificity, PPV, and NPV were calculated using the Wald CI formula [54], namely,



(1)

where p is the estimated value, n is the sample size, and $z_{a/2}$ is the *(1- a/2)* quantile for the standard normal distribution. CIs for the AUC, A-PPV, and sNB were calculated based on the formulas defined in the study by Riley et al [55] for estimation of the SE and Wald CI formula, namely,



(2)



(3)

where v is the estimated AUC or A-PPV value, π is the proportion of patients with sepsis in a given cohort, and *sens* (respectively, *spec*) is the estimated sensitivity (resp. specificity). CIs for the Brier score were computed based on Hoeffding inequality [56] for bounding the difference between the sample estimate of a continuous, bounded random variable and its expected value, namely:



(4)

For each pair of developed ML models and each evaluation metric, we evaluated the hypothesis that the 2 classifiers had the same (null hypothesis) or different average performance: we used a 2-tailed chi-square test to test the hypotheses and computed the P values. For each developed ML model and each dataset, we evaluated the hypothesis that the cautious classifier provided an improvement in terms of sensitivity, specificity, PPV, or NPV compared to the corresponding standard ML model: we used a 1-tailed chi-square test to test the hypotheses and compute the P value. Finally, for each developed ML model, each external validation dataset, and each evaluation metric, we evaluated the hypothesis that the performance of the ML model was not worse than what was reported on the internal validation dataset. We used a 2-tailed chi-square test to test the hypotheses and compute the P values. To control the increased type I error (mistaken rejection of a null hypothesis) due to multiple hypothesis testing, we adjusted the computed P values using a false discovery rate correction procedure. Particularly, we used the optimal adaptive procedure by Gavrilov et al [57], which computes the k-th adjusted P value p_k (with the P values sorted in increasing order) as given in the following:



(5)



XSL•FO
RenderX

(6)

where $\alpha=0.05$ is the confidence threshold for null-hypothesis rejection, and $1 \leq k \leq m$, where m was the number of tested hypotheses. All code for statistical analysis was implemented in Python. We used *pandas* and *numpy* for data loading and management, *scipy* for implementation of the chi-square test, and *statsmodels* for implementation of the false discovery rate correction procedure. We used custom Python implementations of the CIs for the AUC, A-PPV, sNB, and Brier score.

Ethical Considerations

The collection of data at the different sites, as well as the multicentric study, was conducted in compliance with the guidelines established by the local ethics committees, relevant ethical and legal regulations, and in agreement with the World Medical Association Declaration of Helsinki. Regarding the ethics review, the study used only secondary data that had been previously collected at different sites in earlier studies, in accordance with an approved ethical framework from the local ethical boards. The present study does not introduce any new risks, burdens, or interactions with participants beyond those already considered and approved in the original ethical reviews. Since the study did not require any blood draws or procedures from the enrolled patients that would not already have been performed as part of standard medical care, and because the data was rendered anonymous before analysis, written informed consent and approval were exempted by the local committees, and no compensation was provided to the enrolled patients.

Results

Overview

The full information about the features' distribution, as well as the sample sizes, in the 6 datasets is reported in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of the populations in the 6 cohorts.

Feature	PA-ED[a]	AR-ED[b]	PA-ICU[c]	PD-ICU[d]	UD-ED[e]	OGSA-ICU[f]
Age (y), mean (SD)	55.1 (20)	69 (21.4)	66.5 (13)	66.3 (14.9)	66.5 (19.2)	64.3 (18.1)
WBC[g] (10^9/L), mean (SD)	10.2 (4.1)	11.3 (6.4)	13.9 (6.6)	11.5 (5.4)	8.9 (4.6)	13.4 (6.2)
Neutrophils(10^9/L), mean (SD)	7.2 (3.9)	8.7 (5.4)	12.7 (6.7)	9.4 (5)	6.4 (3.8)	12.3 (9.4)
Monocytes(10^9/L), mean (SD)	0.8 (0.3)	0.8 (0.4)	0.8 (0.4)	0.9 (0.5)	0.7 (0.4)	0.8 (0.5)
MDW[h], mean (SD)	20.1 (3.2)	25.6 (6.4)	24.6 (7.5)	24.1 (5)	21.2 (4.2)	—[i]
CRP[j] (mg/L), mean (SD)	22.6 (49.3)	83 (93)	81.4 (96.5)	93.9 (81)	25.3 (52.9)	56.6 (80.7)
Lymphocytes (10^9/L), mean (SD)	2.1 (1.1)	1.6 (3.6)	1 (0.5)	1 (0.6)	1.7 (2.3)	1.2 (1.9)
Eosinophils (10^9/L), mean (SD)	0.1 (0.2)	0.1 (0.2)	0.1 (0.5)	0.2 (0.4)	0.1 (0.1)	0.1 (0.1)
Basophils(10^9/L), mean (SD)	0.1 (0.1)	0 (0.1)	0 (0)	0 (0.1)	0 (0)	0.1 (0.1)
RBC[k] (10^{12}/L), mean (SD)	4.6 (0.7)	4.2 (0.8)	3.9 (0.8)	3.5 (0.6)	4.4 (0.7)	3.9 (0.6)
Hemoglobin level (G/100 mL), mean (SD)	13.4 (2.1)	12.3 (2.3)	10.9 (2.3)	10.4 (1.5)	13.1 (2.1)	11.5 (1.8)
Hematocrit test (%), mean (SD)	39.3 (5.9)	37.6 (6.5)	32.6 (7)	31.7 (4.8)	38.8 (5.8)	34.3 (5.1)
MCV[l] (fL), mean (SD)	86.5 (7.9)	91 (6.7)	85.3 (9.5)	90.1 (6.1)	88.9 (6.4)	89.1 (5.3)
MCH[m] (pg), mean (SD)	29.5 (3.2)	29.6 (2.4)	29.3 (7.7)	29.5 (2.2)	30.1 (2.6)	29.9 (2.3)
MCHC[n] (G/100 mL), mean (SD)	34 (1.2)	32.5 (1.1)	33.1 (1.5)	32.7 (0.9)	33.8 (1)	66.1 (327.5)
RDW[o] (%), mean (SD)	14.6 (2.1)	15.8 (2.4)	48 (7.6)	15.7 (2.9)	14.7 (2.1)	14.5 (2.2)
Platelet count (PLT) (10^9/L), mean (SD)	254.4 (90.4)	229.6 (105.2)	243.3 (120.8)	236.7 (145.3)	241.7 (88.5)	196.3 (89.5)
NLR[p], mean (SD)	5 (5.7)	10.3 (11.3)	18 (19.1)	14.4 (17.2)	5.8 (7.1)	16.7 (15.8)
Sex, n (%)						
Male	870 (48.59)	169 (54.9)	46 (61)	1612 (76.25)	441 (47)	50 (50)
Female	930 (51.41)	139 (45.1)	29 (39)	502 (23.75)	497 (53)	50 (50)
Sepsis, n (%)						
Yes	83 (4.59)	76 (24.7)	22 (29)	709 (33.54)	63 (6.7)	6 (6)
No	1726 (95.41)	232 (75.3)	53 (71)	1405 (66.46)	875 (93.3)	94 (94)
Patients, n (%)	1809 (33.85)	308 (5.76)	75 (1.40)	2114 (39.56)	938 (17.55)	100 (1.87)

[a]PA-ED: Palermo, Emergency Department.
[b]AR-ED: Arezzo, Emergency Department.
[c]PA-ICU: Palermo, Intensive Care Unit.
[d]PD-ICU: Padova, Intensive Care Unit.
[e]UD-ED: Udine, Emergency Department.
[f]OGSA-ICU: Ospedale Galeazzi Sant'Ambrogio, Intensive Care Unit.
[g]WBC: white blood cell.
[h]MDW: monocyte distribution width.
[i]Not available.
[j]CRP: C-reactive protein.
[k]RBC: red blood cell.
[l]MCV: mean corpuscular volume.
[m]MCH: mean corpuscular hemoglobin.
[n]MCHC: mean corpuscular hemoglobin concentration.
[o]RDW: RBC distribution width.
[p]NLR: neutrophils-lymphocytes ratio.

XSL•FO
RenderX

Internal Validation Results

The results of the models on the interval validation set are shown in Table 2 and Figures 1 to 3 (Figure S1 in Multimedia Appendix 1). XGB significantly outperformed all other models in terms of AUC, A-PPV, and sNB (Figure S2 in Multimedia Appendix 1). LR and XGB had greater sensitivity and PPV, respectively, than all other models, although not significantly so. XGB, RF, and LR significantly outperformed all the baseline comparison methods (ie, MDW [37], CRP [44], Sepsis Index [45], COMPOSER [25], as well as the SVM model developed in the study by Aguirre and Urrechaga [31]; Figures 1 and 2; Figure S1 in Multimedia Appendix 1) and did so across all possible decision thresholds. The importance of the most relevant features for the models with the highest AUC and PPV (XGB) and sensitivity (LR) are depicted in Figures 4 and 5. It was identified that MDW was the single most predictive feature (for all models), followed by the CRP, the neutrophils-lymphocytes ratio, the leucocytic formula composition, the hematocrit, and red blood cell count, which also had significant predictive power. An interpretable DT model is given in (Figure 5): after discussion with the clinical experts, we derived from the DT a highly interpretable, cautious decision rule model (Partial Decision Rule [PDR]) defined by the following rules: if MDW>24.1, then sepsis; if MDW ≤24.1 and neutrophils ≤8.6 then no sepsis; if MDW ≤20.5 and neutrophils >8.6 then no sepsis; otherwise abstain.

PDR had better PPV and specificity than the developed ML models, although not significantly so. In terms of sensitivity and NPV, it was significantly worse than only LR. Furthermore, among the nonidentified sepsis cases, most of them (14% of the total number of sepsis cases in the internal test set) were classified by the PDR model classified in the abstain group, while only 1 sepsis case was classified in the no sepsis group.

XSL•FO
RenderX

Table 2. The results of the developed machine learning models (with the optimal configuration of hyperparameters) on the internal hold-out test set, together with the corresponding 95% CIs (in parenthesis). The considered models were logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), decision tree (DT), and the cautious decision rule model (PDR). We also report the performance of state-of-the-art baselines monocyte distribution width (MDW), CRP, Sepsis Index, Mindray SVM, and Conformal Multidimensional Prediction of Sepsis Risk (COMPOSER).

	LR (95% CI)	SVM (95% CI)	RF (95% CI)	DT (95% CI)	XGB (95% CI)	PDR (95% CI)	Mindray SVM (95% CI)	MDW (95% CI)	CRP (95% CI)	Sepsis index (95% CI)	COMPOSER (95% CI)
Sensitivity	0.90 (0.03)[a]	0.86 (0.03)[b]	0.81 (0.04)	0.86 (0.03)[b]	0.86 (0.03)[b]	0.81 (0.04)	0.19 (0.04)	0.90 (0.03)[a]	0.67 (0.04)	0.90 (0.03)[a]	0.80 (0.04)
Specificity	0.94 (0.02)	0.96 (0.01)	0.97 (0.02)[b]	0.91 (0.03)[b]	0.97 (0.02)[b]	0.98 (0.01)[a]	0.99 (0.01)	0.89 (0.03)	0.94 (0.02)	0.93 (0.02)	0.93 (0.02)
PPV[c]	0.43 (0.05)	0.59 (0.04)[b]	0.57 (0.05)[b]	0.32 (0.04)	0.60 (0.05)[b]	0.65 (0.04)[a]	0.50 (0.05)	0.28 (0.04)	0.37 (0.04)	0.45 (0.05)	0.30 (0.04)
NPV[d]	1.00 (0.01)[a]	0.99 (0.01)[b]	0.99 (0.01)[b]	0.99 (0.01)[b]	0.99 (0.01)[b]	0.99 (0.01)[b]	0.96 (0.02)	1 (0.01)[a]	0.98 (0.01)	1 (0.01)[a]	0.98 (0.01)
AUC[e]	0.96 (0)	0.95 (0)	0.97 (0)	0.91 (0)	0.98 (0)[a]	—[f]	0.91 (0)	—	—	—	—
Brier	0.05 (0.02)	0.02 (0.01)[a]	0.03 (0.01)[b]	0.05 (0.02)	0.03 (0.01)[b]	—	0.04 (0.02)[b]	—	—	—	—
A-PPV[g]	0.78 (0.01)	0.81 (0)	0.77 (0.01)	0.67 (0.01)	0.83 (0)[a]	—	0.44 (0.01)	—	—	—	—
HC[h] sensitivity	0.90 (0.03)[b]	0.81 (0.04)	0.86 (0.03)	0.85 (0.03)	0.90 (0.03)[b]	0.94 (0.02)[a]	0.11 (0.03)	—	—	—	—
HC specificity	0.96 (0.02)	0.98 (0.01)[b]	0.99 (0.01)[a]	0.99 (0.01)[a]	0.98 (0.01)[b]	0.98 (0.01)[b]	0.99 (0.01)[a]	—	—	—	—
HC PPV	0.51 (0.05)	0.62 (0.04)	0.71 (0.04)[b]	0.67 (0.04)[b]	0.72 (0.04)[a]	0.65 (0.04)[b]	0.33 (0.04)	—	—	—	—
HC NPV	0.99 (0.01)[b]	0.99 (0.01)[b]	0.99 (0.01)[b]	0.99 (0.01)[b]	1 (0.01)[a]	1 (0.01)[a]	0.96 (0.02)	—	—	—	—
Coverage	0.94 (0.02)	0.98 (0.01)[b]	0.92 (0.03)	0.92 (0.03)	0.97 (0.02)[b]	0.90 (0.03)	0.99 (0.01)[a]	—	—	—	—
sNB[i]	0.22 (0.01)	0.60 (0.01)[b]	0.35 (0.01)	0.24 (0.01)	0.62 (0.01)[a]	—	0.16 (0)	—	—	—	—

[a]Best performing model.

[b]Models whose 95% CIs overlap with the best model.

[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

[e]AUC: area under the receiver operating characteristic curve.

[f]Not available.

[g]A-PPV: average positive predictive value.

[h]HC: high confidence.

[i]sNB: standardized net benefit.

XSL•FO
RenderX

Figure 1. Performance (in terms of sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) of the developed models, each represented by a colored circle, on the 6 considered datasets (represented on the y-axis), together with the corresponding 95% CIs. AR-ED: Arezzo, Emergency Department; DT: decision tree; LR: logistic regression; OGSA-ICU: Ospedale Galeazzi Sant'Ambrogio, Intensive Care Unit; PA-ED: Palermo, Emergency Department; PA-ICU: Palermo, Intensive Care Unit; PDR: Partial Decision Rule; PD-ICU: Padova, Intensive Care Unit; RF: random forest; SVM: support vector machine; UD-ED: Udine, Emergency Department; XGB: extreme gradient boosting.

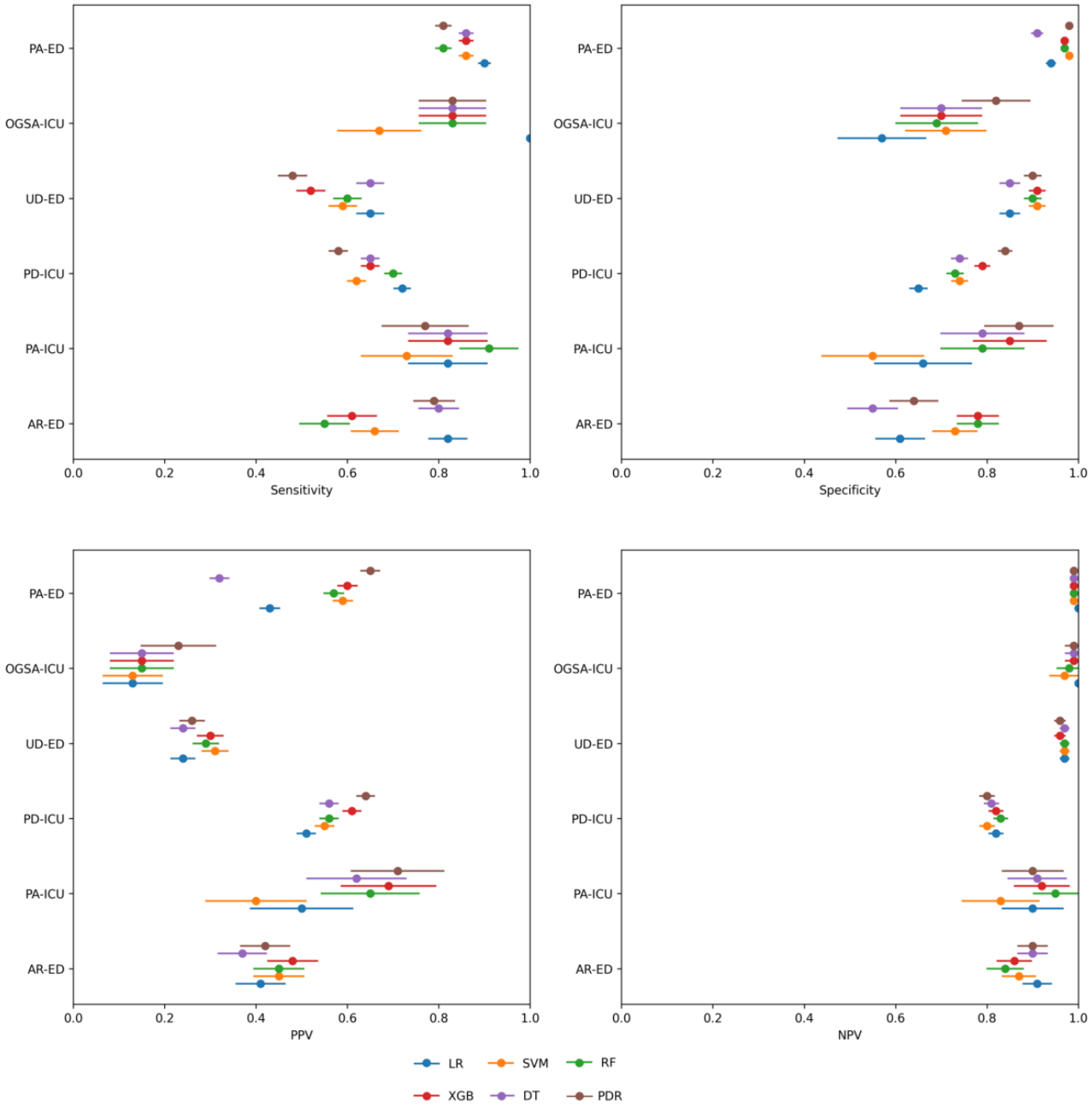

XSL•FO
RenderX

Figure 2. Receiver operating characteristic curves for the models, along with their area under the receiver operating characteristic curve (AUC) values, on the Palermo, Emergency Department (PA-ED) internal validation dataset. In the plot we also report on the sensitivity and specificity of 4 baselines: binary thresholds based on the monocyte distribution width (MDW) [35] and C-reactive protein (CRP) [42], a binary threshold based on the Sepsis-Index parameter [43], the Conformal Multidimensional Prediction of Sepsis Risk model [25], and the support vector machine (SVM) model developed in the study by Aguirre and Urrechaga [31]. DT: decision tree; LR: logistic regression; PDR: Partial Decision Rule; RF: random forest; XGB: extreme gradient boosting.

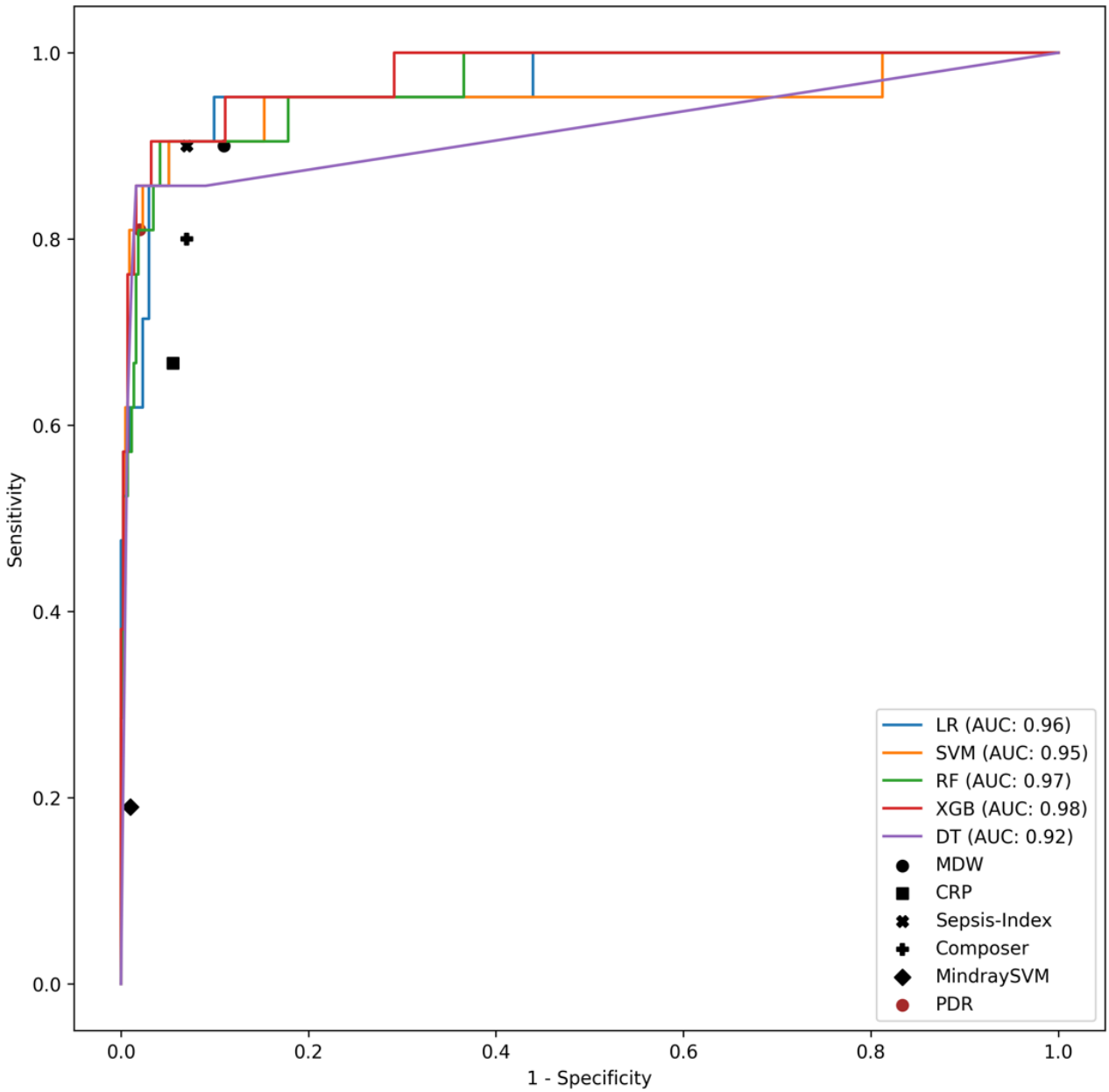

XSL•FO
RenderX

Figure 3. Sensitivity–positive predictive value (PPV) curves for the models, along with their average PPV (A-PPV) values, on the Palermo, Emergency Department (PA-ED) internal validation dataset. The plot also shows the sensitivity and PPV of 4 baselines: binary thresholds based on the monocyte distribution width (MDW) [35] and C-reactive protein (CRP) [42], a binary threshold based on the Sepsis-Index parameter [43], the Conformal Multidimensional Prediction of Sepsis Risk model [25], and the support vector machine (SVM) model developed in the study by Aguirre and Urrechaga [31]. DT: decision tree; LR: logistic regression; PDR: Partial Decision Rule; RF: random forest; XGB: extreme gradient boosting.

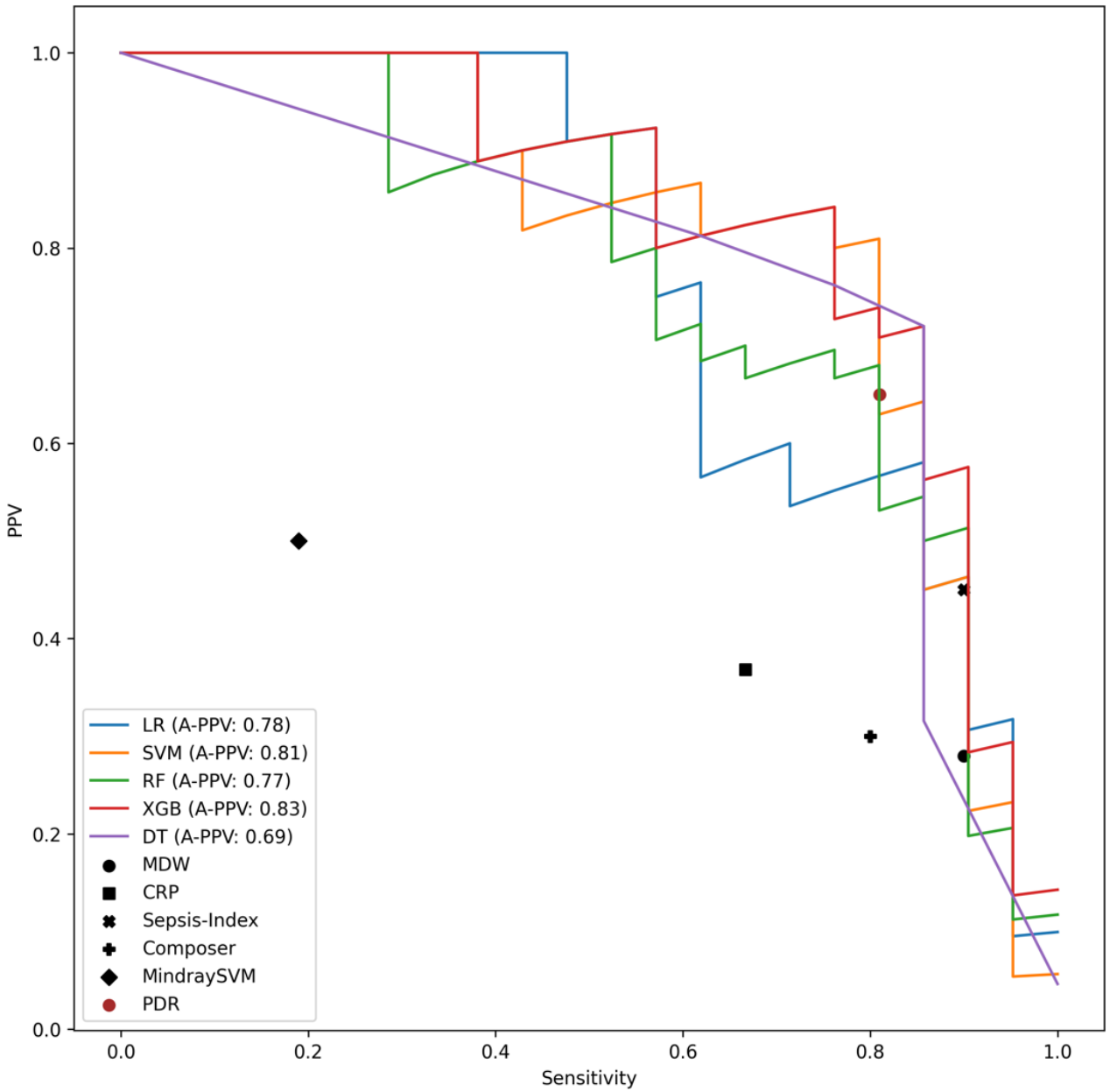

XSL•FO
RenderX

Figure 4. Feature importances for the logistic regression (LR; left) and extreme gradient boosting (XGB; right) models. The feature importance for the LR model represents the coefficients of the induced linear model, shown as bar plots. For each feature, the color represents the sign of the coefficient. The width of the bar represents the magnitude of the coefficient: larger width denotes greater importance. The feature importance for the XGB model was computed through the Shapley Additive Explanations method and is represented in terms of violin plots: for each feature, red denotes high values, while blue denotes low values; values at the right of the middle vertical bar denote an increased confidence score for the positive class (sepsis), while values at the left denote a decreases confidence score. CRP: C-reactive protein; MCHC: mean corpuscular hemoglobin concentration; MCV: mean corpuscular volume; MDW: monocyte distribution width; NLR: neutrophils-lymphocytes ratio; RBC: red blood cell; WBC: white blood cell; HCT: hematocrit test; HGB: hemoglobin; PLT: platelet count.

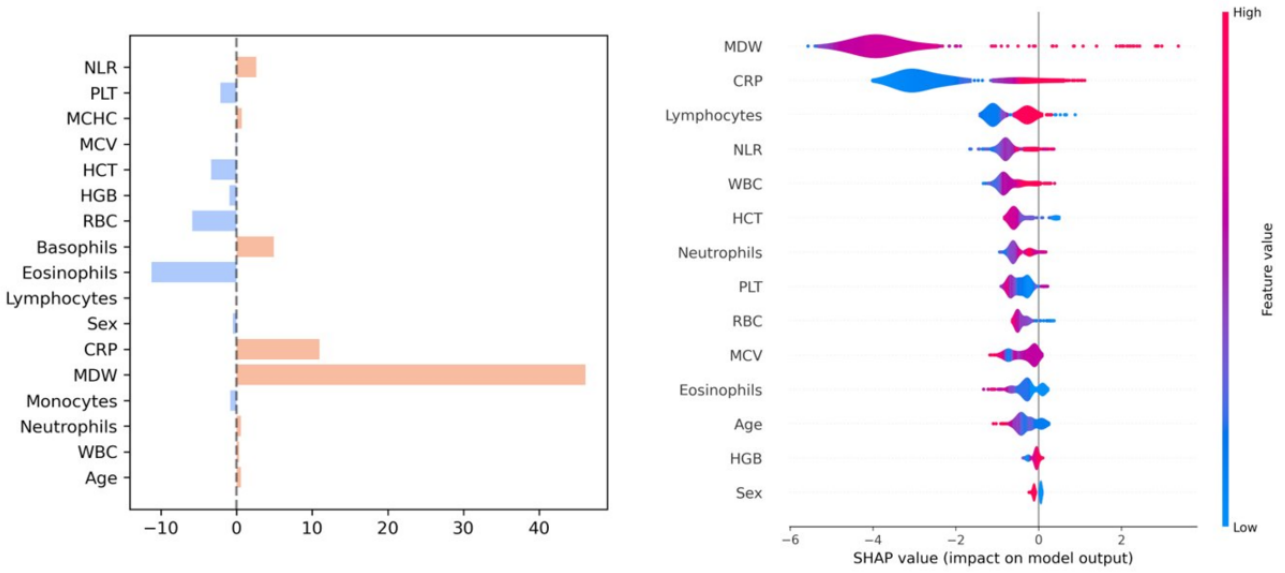

XSL•FO
RenderX

Figure 5. Interpretable decision tree: the color of each node denotes the majority class for the corresponding subset of instances (blue: sepsis and orange: no sepsis). Each non–leaf node contains 3 data elements: a selected feature and corresponding threshold (top); the proportion of samples corresponding to that node (middle); and the proportion of samples belonging to each class (bottom). Leaf nodes contain 2 data elements: the proportion of samples corresponding to that node (top) and the proportion of samples belonging to each class (bottom). MDW: monocyte distribution width; RBC: red blood cell; WBC: white blood cell.

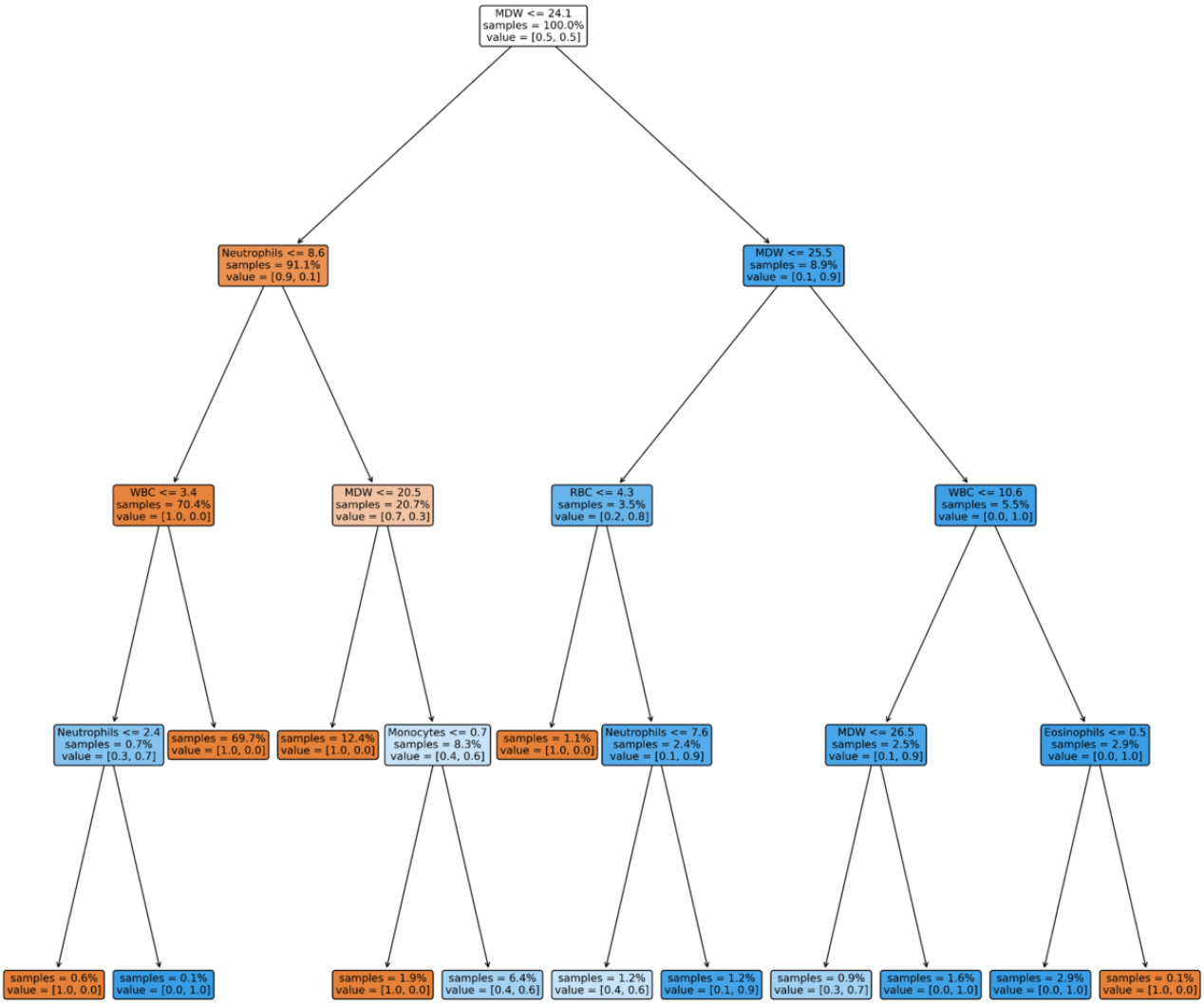

External Validation Results

The results of all models, including the PDR model, on the external validation datasets (PA-ICU, PD-ICU, UD-ED, and AR-ED) are reported in Table S2 in Multimedia Appendix 1 and Figure 1, as well as in Figures S3 to S6 in Multimedia Appendix 1, in terms of receiver operating characteristic and sensitivity-PPV curves. The results of the external validation for the model with the highest AUC on the internal validation (XGB), are represented in Figure 6 in the form of an external performance diagram [40]. The XGB model achieved acceptable-to-excellent performance on all external datasets and with respect to all metrics, except for the sNB on the UD-ED: this dataset was, among the external validation datasets, the most dissimilar to the internal training set (PA-ED). The external performance diagrams for the other models are in Figures S7 to S10 in Multimedia Appendix 1. While the performance of all models worsened compared with the internal test sets, the differences were not significant, as shown in Figure S11 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 6. External performance diagram [40] for the extreme gradient boosting (XGB) model on the external datasets. The diagram illustrates the performance of the XGB model according to 3 different aspects: discrimination power (in terms of area under the receiver operating characteristic curve [AUC]), utility (in terms of standardized net benefit), and calibration (in terms of Brier score). The size of the ellipses associated with the datasets denotes the 95% CIs; the transparency of the ellipses denotes the achievement of the minimum sample size (the lower the transparency, the closer the sample size to the minimum sample size). The diagram has been produced with the tool available on the Metimeter website [58]. AR-ED: Arezzo, Emergency Department; OGSA-ICU: Ospedale Galeazzi Sant'Ambrogio, Intensive Care Unit; PA-ICU: Palermo, Intensive Care Unit; PD-ICU: Padova, Intensive Care Unit; UD-ED: Udine, Emergency Department.

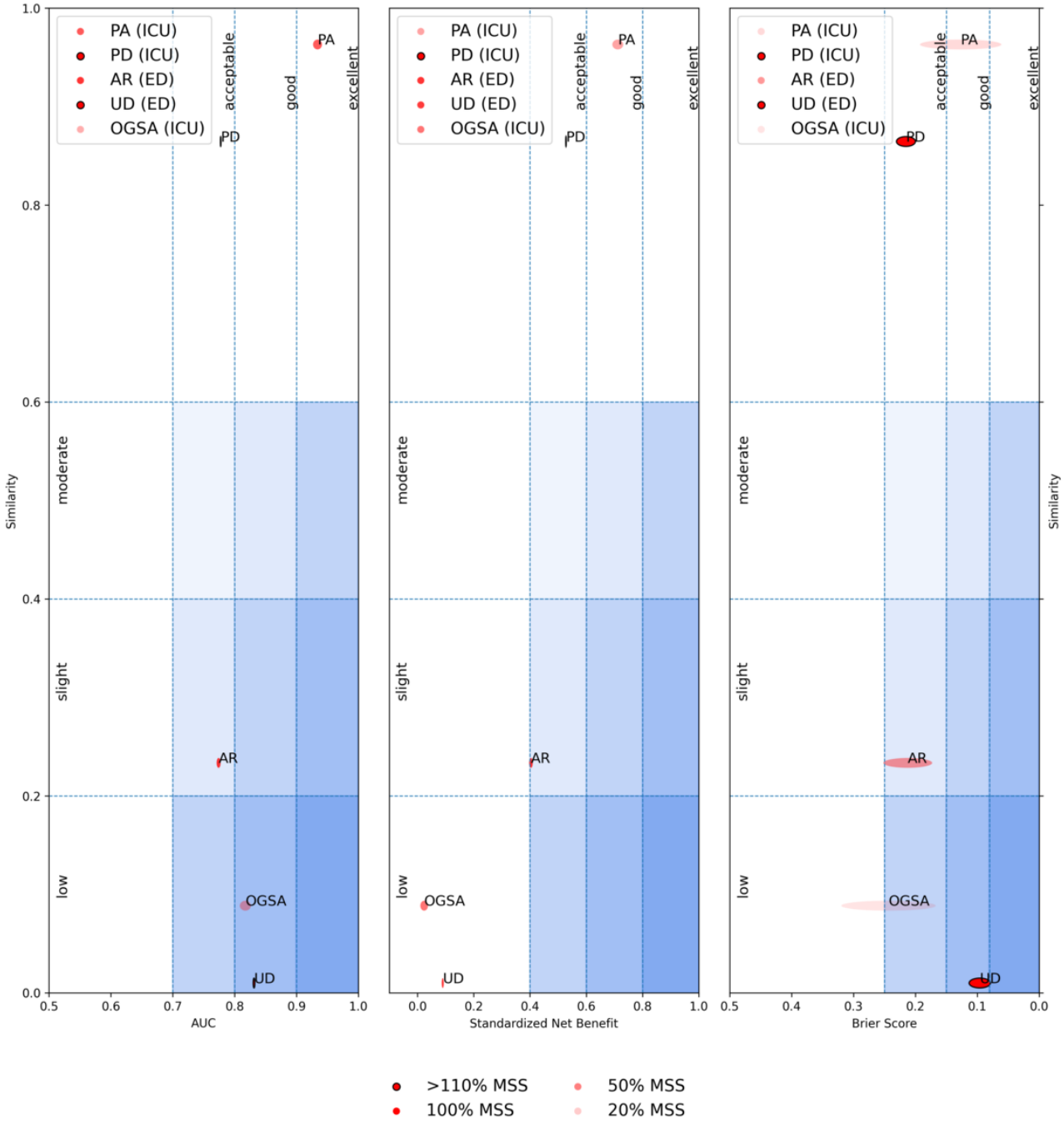

Non-MDW Validation Results

The data from all cohorts except OGSA-ICU were used to develop an imputation model to predict the value of the MDW feature from all the other considered parameters. The developed imputation model was then used to impute the values of MDW in the OGSA-ICU dataset. The results of the imputation process are in Figure S12 in Multimedia Appendix 1, while additional detail about the implementation of the imputation model is in the Methods section.

The results of the non-MDW systems, on the OGSA-ICU cohort, are reported in Figure 1 as well as in Figure S13 in Multimedia Appendix 1, in terms of ROC and sensitivity-PPV curves. The results of the validation, for the system encompassing the model with the highest AUC on the internal validation (XGB), are represented in Figure 6 in the form of an external performance diagram [40]. The XGB-based system achieved low PPV and sNB on the OGSA-ICU dataset, as also shown in Figure S13 in Multimedia Appendix 1, which was also the second most dissimilar to the internal training set (PA-ED).

XSL•FO
RenderX

Analysis of Cautious Classifiers

The use of cautious classifiers reduced the amounts of error for almost all models, datasets, and metrics, as is shown in Figure 7 and Figure S14 in Multimedia Appendix 1. For XGB, PDR, RF, and LR (except for the OGSA-ICU dataset), cautious classification led to an improvement in both sensitivity and PPV and equivalent or better performance in terms of NPV. For all models and datasets, the coverage of the cautious classifiers was >50%, and, in all but 3 cases (all for the RF model), >70%. These results indicate that cautious classification could be useful to reduce false negatives, with minimal deleterious effects in terms of the increase of false positives or the number of undecided cases.

Figure 7. Difference between the performance of cautious models and the corresponding standard models: values greater than 0 denote an improvement in performance in the cautious inference model as compared with the corresponding standard one. The 95% CIs were computed using the pooled SD. Each model is represented by a colored circle. AR-ED: Arezzo, Emergency Department, DT: decision tree; LR: logistic regression; OGSA-ICU: Ospedale Galeazzi Sant'Ambrogio, Intensive Care Unit, PA-ED: Palermo, Emergency Department; PA-ICU: Palermo, Intensive Care Unit; PDR: Partial Decision Rule; PD-ICU: Padova, Intensive Care Unit; RF: random forest; SVM: support vector machine; UD-ED: Udine, Emergency Department; XGB: extreme gradient boosting.

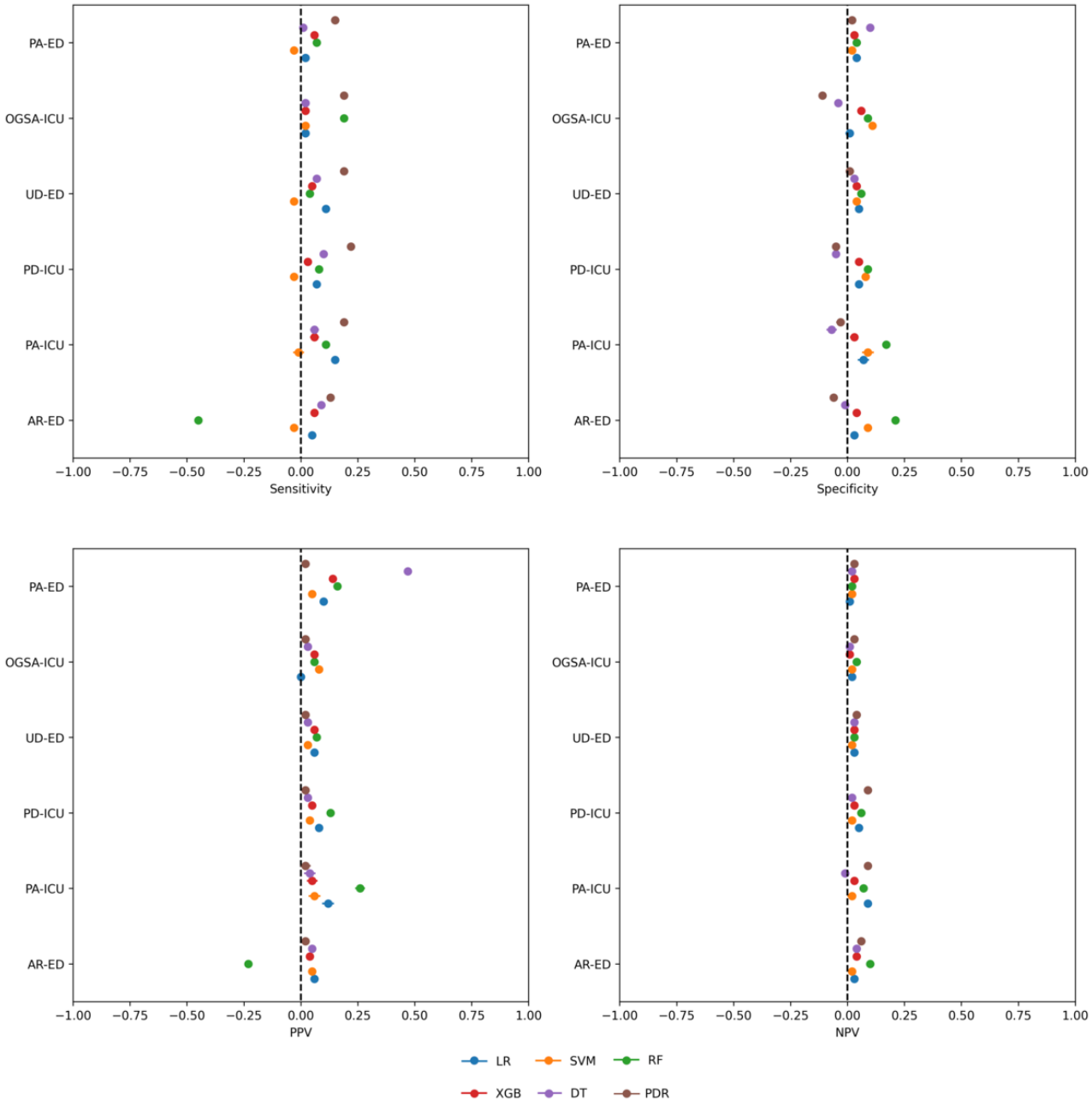

XSL•FO
RenderX

Discussion

Principal Findings

Sepsis recognition, especially in the ED, is challenging due to the potentially limited data that can be timely collected and difficult logistic settings. The recognition of sepsis triggers the application of validated bundles (ie, fluid resuscitation, antibiotic administration, blood culture and collection of other relevant specimens, and lactate dosage) that can reduce mortality if applied timely [2]. However, there is a risk of overtreatment in the case of *false positive* cases, with potential detrimental effects of therapies and waste of resources.

In this study, we developed and validated 5 ML models, as well as an interpretable set of rules derived from the latter, for sepsis screening based on clinical and laboratory features. The main findings of our study can be summarized as follows: (1) XGB and LR models showed the best performance for sepsis screening; (2) the most important features for the XGB and LR model were MDW and CRP; (3) the ML models have been validated on 4 external datasets, showing good generalizability in terms of both discrimination power, calibration, and utility; (4) the interpretable, cautious decision rules had performance comparable to that of black-box ML models, while being more easily understandable by clinicians; (5) cautious classification techniques could further improve screening performance, with minimal impact on coverage; and (6) while the non-MDW ML systems achieved acceptable performance on the non-MDW validation test, the lack of MDW led to a large decrease in PPV. Noteworthy, 5 datasets were obtained from different Italian hospitals, in different periods, and different clinical wards (namely, ED and ICU), and exhibited widely varying characteristics in terms of covariate, label, and missingness distributions. In particular, the datasets exhibited a large variation in terms of sepsis prevalence in the corresponding populations, also due to differences between ED and ICU settings (wherein the latter have typically larger incidence than the former) or population characteristics (as an example, the OGSA-ICU cohort was obtained from a hospital specialized in orthopedics, a setting that typically has much lower sepsis' incidence than general hospitals [59]).

Comparison to Prior Work

In this study, we also compared the developed models against different state-of-the-art approaches, including commonly used clinical thresholds [37,44] and scores [45], as well as ML models [31,32]. Our models reported better performance than all the previous state-of-the-art approaches. In this sense, we believe that the most relevant innovations from the methodological point of view, enabling these results, were the inclusion of MDW in the panel of parameters, which resulted in being the most predictive characteristic for sepsis, and the use of controllable AI methodologies, which enabled us to obtain models that were more robust to various forms of distribution shifts. In this regard, to date, the lack of external validation is one of the most important criticisms of ML studies [40,60]. Indeed, although it is a fundamental step to strengthen model derivation and ensure generalizability, most published articles do not perform it, with potentially negative consequences in terms of robustness, generalizability, and real-world applicability. This issue has been markedly noted also in the context of sepsis screening and prediction, where one of the most commonly adopted proprietary decision support models for sepsis prediction (ie, EPIC) has been shown to be severely impacted by shifts and variabilities associated with varying settings [28-31]. Thus, along with the methodological novelties of our study, one of the main strengths of our study is the external validation on 5 different datasets. On the basis of the results of the external validation, we can conclude that the developed models are able to generalize better than the state-of-the-art, at least in the Italian settings, as they maintain the performance independently from covariate shifts (due to either different patient populations or locations, ED or ICU settings, or different measuring instruments), label shifts (sepsis2 or sepsis3 diagnostic criteria and different incidence of sepsis) as well as missingness shifts scenarios: this latter characteristic, specifically, is of particular interest since, despite its potential impact on model generalization, it has scarcely been investigated in medical ML studies [26]. In this study, we developed ML-based systems that could also be applied in settings where MDW was not available, with the aim of increasing the applicability of the developed ML models as well as of assessing the importance of the MDW parameter (which was the most relevant feature according to all the developed ML models). While, indeed, the obtained results show that acceptable results could be obtained even in the presence of significant missingness shift, the lack of MDW led to a relevant drop in performance, thus highlighting the need for collecting complete data in practical scenarios. By contrast, our models were shown to robustly generalize to different incidence rates, not only due to different clinical settings (ED vs ICU) but also to diagnostic criteria (sepsis-2 vs sepsis-3), allowing us to prove not only their effectiveness and robustness but also that varying incidence rates may have lower influence on performance compared to other variation sources (eg, lack of MDW, as described earlier, or selection criteria, as in the AR-ED cohort).

Strengths and Limitations

Strengths of this study are the external validation, the large sample size, and the use of several ML methods as well as several indicators to assess the model's performance. This study comes from the collaboration and cooperation among different professionals, including clinicians, engineers, biologists, and informatics. Each has made its expertise available with the common aim to develop a tool that could be usable in clinical practice.

Limitations of the study must also be mentioned. First, the use of sepsis-2 criteria for sepsis identification in the database used for derivation. However, the use of sepsis-3 criteria in the ED is costly in terms of resources and time-consuming and may be difficult to apply outside the ICU because full assessment of organ dysfunction is needed. Furthermore, the developed models have been shown to generalize well also in settings where sepsis-3 criteria were used for sepsis identification, showing good performance despite this label shift factor. Thus, despite this potential limitation, we believe that the aid of CBC laboratory biomarkers, easily obtainable and at a low cost, may help improve the specificity of both clinical criteria and initial

XSL•FO
RenderX

screening and support differential diagnosis in case of negative results. Second, we were not able to stratify patients according to the infectious focus (eg, lung, abdomen, central nervous system, soft tissue, blood, and urinary tract) or clinical severity (sepsis vs septic shock) due to the absence of serum lactate levels in our datasets. Both the type of infection and clinical severity may be potential modifiers of model performance and should be investigated in future work. More generally, lactate levels are important information for sepsis, especially in ICU settings. At the same time, lactates generally lack several characteristics of an ideal biomarker for sepsis screening, being more apt as a marker of severity in patients with high pretest probability of sepsis. The objective of the developed ML models was to provide clinicians with quick but accurate sepsis screening tests, especially in cases where sepsis is not yet suspected (patients accessing the ED): because lactates are not frequently requested during initial management, requiring them as features for our models would severely decrease their use (as an example, none of the 3 ED cohorts considered in our study encompassed lactates). Nonetheless, we believe that future studies should consider the combination of lactate levels and MDW (as well as other CBC parameters), which could potentially lead to even better performance. In addition, it should be highlighted that MDW can be obtained only using a particular hematological analyzer, while other available instrumentations do not allow for the calculation of this interesting index. This aspect is particularly relevant in light of the results of the non-MDW validation test, which highlighted the diagnostic power of the MDW parameter. Indeed, we noted that its lack was associated with a relevant drop in performance (especially as regards the PPV), thus showing the potential advantage to use the mentioned hematological analyzer to collect CBC biomarkers (including MDW). Finally, in this study we only validated our models on retrospective data; further research should evaluate the potential application of a clinical algorithm incorporating the results of our models in prospective studies and for clinical timely decision in screening patients for sepsis.

Conclusions and Future Directions

In clinical practice, because of their characteristics in performance and robustness, the ML models we developed and presented in this study may represent a reliable tool for supporting physicians in sepsis screening on patients of both sexes admitted to the ED and ICU. Parameters included in the models can be determined rapidly and cost-effectively. Furthermore, the excellent NPV of the models allows us to rule out sepsis with high accuracy, while they are also characterized by a good sensitivity and specificity. Noteworthy, such models significantly improved the specificity and PPV of MDW and CRP alone, as well as of the other considered state-of-the-art ML models, thus confirming our hypothesis that integrating several CBC parameters through the use of ML could significantly improve the performance of traditional biomarkers. Therefore, we conjecture that the developed models could be especially useful as a first-level test to detect sepsis. Accordingly, an increased value (of the confidence score produced by the model) may represent an alert for clinicians that they should perform further investigations to confirm the sepsis suspicion. Indeed, in the case of nonobvious infectious foci, an increased value may help physicians decide to evaluate further biomarkers, such as procalcitonin [61], both fundamental for the appropriate management of patients with sepsis, specifically for diagnosis and guiding treatment, respectively. To this latter aim, the ability to further isolate a group of predictions with higher PPV and good sensitivity, by using the cautious classification models and the interpretable decision rules, is especially interesting. Indeed, from the clinical point of view, it would enable the administration of aggressive therapies in a small sample of patients, but associated with extremely high confidence (90%), thus minimizing the risk of overtreatment and potentially leading to a significant improvement in the outcome. In our opinion, this is a valuable hypothesis to be validated in properly designed prospective trials. At the same time, the cases in which these models abstain could be useful to alert the clinicians of a possible control loss [33] and, hence, reduce the risk of automation bias or de-skilling [62]. In this sense, we conjecture that the developed interpretable decision rules could be particularly fruitful as a rapid and easy-to-use tool to support clinicians in promptly delivering appropriate treatment to the selected patients. Indeed, a recent line of research in the ML literature has noted how the black-box nature of modern ML models, and the consequent adoption of explainable AI techniques (such as model surrogates or feature attribution methods), could potentially increase automation bias (or also other cognitive biases related to the use of automation technology) [63], hence advocating for the use of simple and interpretable models in high-stake domains [64] where humans adopt AI tools in a human-AI teaming scenario [65]. In this sense, the developed decision rule model could address this need while still providing satisfying screening performance. Furthermore, from the clinical point of view, the decision rule model seems to confirm the important role of MDW and neutrophils in sepsis development [66] and provides a first combined screening model based on these 2 parameters that should be further investigated and validated in future studies.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request. Access to the computer code used in this research is available upon reasonable request to the authors. Code for validation (internal and external), as well as the associated machine learning models, is accessible on the internet [67].

Authors' Contributions

A Campagner, LA, A Carobene, FDB, A Cortegiani, and F Curcio were responsible for conceptualization. A Campagner, LA, A Carobene, and AP were responsible for the investigation. A Campagner, LA, A Carobene, AP, FDB, AO, EDV, and F Curcio

XSL•FO
RenderX

were responsible for methodology. A Campagner, LA, A Carobene, AP, and FDB were responsible for validation. A Campagner, LA, A Carobene, AP, FDB, and A Cortegiani were responsible for writing the original draft and review and editing. LA, AP, AO, M Lorubbio, EDV, DP, F Curcio, and MC were responsible for data curation. M Locatelli, MP, AO, M Lorubbio, EDV, DP, F Curcio, F Cabitza, MC, and EP were responsible for writing, review and editing. M Locatelli, MP, F Cabitza, and MC were responsible for supervision. A Campagner was responsible for formal analysis, software, and visualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional figures and tables reporting the results of the developed machine learning models on both the training and external validation cohorts.
[DOCX File , 4319 KB - jmir_v27i1e55492_app1.docx]

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to treatment and mortality during mandated emergency care for sepsis. N Engl J Med 2017 Jun 08;376(23):2235-2244 [FREE Full text] [doi: 10.1056/NEJMoa1703058] [Medline: 28528569]
3. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021 Nov;47(11):1181-1247 [FREE Full text] [doi: 10.1007/s00134-021-06506-y] [Medline: 34599691]
4. UK NSC: evidence review process. United Kingdom Government. 2024 Jul 26. URL: https://www.gov.uk/government/publications/uk-nsc-evidence-review-process/uk-nsc-evidence-review-process [accessed 2025-01-10]
5. Fan SL, Miller NS, Lee J, Remick DG. Diagnosing sepsis - the role of laboratory medicine. Clin Chim Acta 2016 Sep 01;460:203-210 [FREE Full text] [doi: 10.1016/j.cca.2016.07.002] [Medline: 27387712]
6. Barichello T, Generoso JS, Singer M, Dal-Pizzol F. Biomarkers for sepsis: more than just fever and leukocytosis-a narrative review. Crit Care 2022 Jan 06;26(1):14 [FREE Full text] [doi: 10.1186/s13054-021-03862-5] [Medline: 34991675]
7. Duncan CF, Youngstein T, Kirrane MD, Lonsdale DO. Diagnostic challenges in sepsis. Curr Infect Dis Rep 2021;23(12):22 [FREE Full text] [doi: 10.1007/s11908-021-00765-y] [Medline: 34720754]
8. Agnello L, Giglio RV, Bivona G, Scazzone C, Gambino CM, Iacona A, et al. The value of a complete blood count (CBC) for sepsis diagnosis and prognosis. Diagnostics (Basel) 2021 Oct 12;11(10):1881 [FREE Full text] [doi: 10.3390/diagnostics11101881] [Medline: 34679578]
9. Agnello L, Vidali M, Lo Sasso B, Giglio RV, Gambino CM, Scazzone C, et al. Monocyte distribution width (MDW) as a screening tool for early detecting sepsis: a systematic review and meta-analysis. Clin Chem Lab Med 2022 Feb 15;60(5):786-792 [FREE Full text] [doi: 10.1515/cclm-2021-1331] [Medline: 35166088]
10. Meraj F, Shaikh S, Maqsood S, Kanani F, Khan H, Jamal S. Monocyte distribution width, a novel biomarker for early sepsis screening and comparison with procalcitonin and c-reactive protein. J Lab Physicians 2023 Jan 18;15(2):294-299 [FREE Full text] [doi: 10.1055/s-0042-1758666] [Medline: 37323592]
11. Malinovska A, Hernried B, Lin A, Badaki-Makun O, Fenstermacher K, Ervin AM, et al. Monocyte distribution width as a diagnostic marker for infection: a systematic review and meta-analysis. Chest 2023 Jul;164(1):101-113 [FREE Full text] [doi: 10.1016/j.chest.2022.12.049] [Medline: 36681146]
12. Ciaccio AM, Agnello L, Sasso BL, Giglio RV, Iacona A, Gambino CM, et al. Monocyte Distribution Width (MDW) as a biomarker of sepsis: an evidenced-based laboratory medicine approach. Clin Chim Acta 2023 Feb 01;540:117214. [doi: 10.1016/j.cca.2022.117214] [Medline: 36596354]
13. Jo SJ, Kim SW, Choi JH, Choi SP, Lee J, Lim J. Monocyte distribution width (MDW) as a useful indicator for early screening of sepsis and discriminating false positive blood cultures. PLoS One 2022 Dec 20;17(12):e0279374 [FREE Full text] [doi: 10.1371/journal.pone.0279374] [Medline: 36538555]
14. Cusinato M, Sivayoham N, Planche T. Sensitivity and specificity of monocyte distribution width (MDW) in detecting patients with infection and sepsis in patients on sepsis pathway in the emergency department. Infection 2023 Jun;51(3):715-727 [FREE Full text] [doi: 10.1007/s15010-022-01956-y] [Medline: 36399260]
15. Wu J, Li L, Luo J. Diagnostic and prognostic value of monocyte distribution width in sepsis. J Inflamm Res 2022 Jul 20;15:4107-4117 [FREE Full text] [doi: 10.2147/JIR.S372666] [Medline: 35898818]
16. Laínez Martínez S, González Del Castillo J. Usefulness of monocyte distribution width (MDW) as a sepsis biomarker. Rev Esp Quimioter 2022 Apr;35 Suppl 1(Suppl 1):2-5. [doi: 10.37201/req/s01.01.2022] [Medline: 35488816]

XSL•FO
RenderX

17. Poz D, Crobu D, Sukhacheva E, Rocchi MB, Anelli MC, Curcio F. Monocyte distribution width (MDW): a useful biomarker to improve sepsis management in emergency department. Clin Chem Lab Med 2022 Jan 11;60(3):433-440. [doi: 10.1515/cclm-2021-0875] [Medline: 35001582]
18. Rabbani N, Kim GY, Suarez CJ, Chen JH. Applications of machine learning in routine laboratory medicine: current state and future directions. Clin Biochem 2022 May;103:1-7 [FREE Full text] [doi: 10.1016/j.clinbiochem.2022.02.011] [Medline: 35227670]
19. Zhang Z, Chen L, Xu P, Wang Q, Zhang J, Chen K, et al. Effectiveness of automated alerting system compared to usual care for the management of sepsis. NPJ Digit Med 2022 Jul 19;5(1):101 [FREE Full text] [doi: 10.1038/s41746-022-00650-5] [Medline: 35854120]
20. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
21. Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020 Mar;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]
22. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019 Mar;170:1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]
23. Komorowski M, Green A, Tatham KC, Seymour C, Antcliffe D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. EBioMedicine 2022 Dec;86:104394 [FREE Full text] [doi: 10.1016/j.ebiom.2022.104394] [Medline: 36470834]
24. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. Front Med (Lausanne) 2021 May 28;8:607952 [FREE Full text] [doi: 10.3389/fmed.2021.607952] [Medline: 34124082]
25. Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say "I don't know". NPJ Digit Med 2021 Sep 09;4(1):134 [FREE Full text] [doi: 10.1038/s41746-021-00504-6] [Medline: 34504260]
26. Zhou H, Balakrishnan S, Lipton ZC. Domain adaptation under missingness shift. arXiv Preprint posted online on November 3, 2022. [doi: 10.48550/arXiv.2211.02093]
27. Habib AR, Lin AL, Grant RW. The epic sepsis model falls short-the importance of external validation. JAMA Intern Med 2021 Aug 01;181(8):1040-1041. [doi: 10.1001/jamainternmed.2021.3333] [Medline: 34152360]
28. Lyons PG, Hofford MR, Yu SC, Michelson AP, Payne PR, Hough CL, et al. Factors associated with variability in the performance of a proprietary sepsis prediction model across 9 networked hospitals in the US. JAMA Intern Med 2023 Jun 01;183(6):611-612 [FREE Full text] [doi: 10.1001/jamainternmed.2022.7182] [Medline: 37010858]
29. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized. JAMA Intern Med 2021 Aug 01;181(8):1065-1070 [FREE Full text] [doi: 10.1001/jamainternmed.2021.2626] [Medline: 34152373]
30. Schertz AR, Smith SA, Lenoir KM, Thomas KW. Clinical impact of a sepsis alert system plus electronic sepsis navigator using the epic sepsis prediction model in the emergency department. J Emerg Med 2023 May;64(5):584-595. [doi: 10.1016/j.jemermed.2023.02.025] [Medline: 37045722]
31. Aguirre U, Urrechaga E. Diagnostic performance of machine learning models using cell population data for the detection of sepsis: a comparative study. Clin Chem Lab Med 2022 Nov 11;61(2):356-365. [doi: 10.1515/cclm-2022-0713] [Medline: 36351434]
32. Manetti S, Cumetti M, De Benedictis A, Lettieri E. Adoption of novel biomarker test parameters with machine learning-based algorithms for the early detection of sepsis in hospital practice. J Nurs Manag 2022 Nov;30(8):3754-3764 [FREE Full text] [doi: 10.1111/jonm.13807] [Medline: 36125938]
33. Kieseberg P, Weippl E, Tjoa AM, Cabitza F, Campagner A, Holzinger A. Controllable AI - an alternative to trustworthiness in complex AI systems? In: Proceedings of the 7th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference. 2023 Presented at: CD-MAKE 2023; August 29-September 1, 2023; Benevento, Italy. [doi: 10.1007/978-3-031-40837-3_1]
34. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies. Int J Med Inform 2021 Sep;153:104510 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104510] [Medline: 34108105]
35. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
36. Campagner A, Famiglini L, Carobene A, Cabitza F. Everything is varied: the surprising impact of instantial variation on ML reliability. Appl Soft Comput 2023 Oct;146:110644. [doi: 10.1016/j.asoc.2023.110644]
37. Agnello L, Bivona G, Vidali M, Scazzone C, Giglio RV, Iacolino G, et al. Monocyte distribution width (MDW) as a screening tool for sepsis in the emergency department. Clin Chem Lab Med 2020 Oct 25;58(11):1951-1957. [doi: 10.1515/cclm-2020-0417] [Medline: 32598299]

XSL•FO
RenderX

38. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Intensive Care Med 2003 Apr 28;29(4):530-538. [doi: 10.1007/s00134-003-1662-x] [Medline: 12664219]
39. Cabitza F, Campagner A, Ferrari D, Di Resta C, Ceriotti D, Sabetta E, et al. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. Clin Chem Lab Med 2020 Oct 21;59(2):421-431 [FREE Full text] [doi: 10.1515/cclm-2020-1294] [Medline: 33079698]
40. Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed 2021 Sep;208:106288 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106288] [Medline: 34352688]
41. Fayaz SA, Zaman M, Kaul S, Butt MA. Is deep learning on tabular data enough? An assessment. Int J Adv Comput Sci Appl 2022;13(4). [doi: 10.14569/IJACSA.2022.0130454]
42. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Inf Fusion 2022 May;81:84-90. [doi: 10.1016/j.inffus.2021.11.011]
43. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022 Presented at: NIPS'22; November 28-December 9, 2022; New Orleans, LA.
44. Sierra R, Rello J, Bailén MA, Benítez E, Gordillo A, León C, et al. C-reactive protein used as an early indicator of infection in patients with systemic inflammatory response syndrome. Intensive Care Med 2004 Nov 11;30(11):2038-2045. [doi: 10.1007/s00134-004-2434-y] [Medline: 15378239]
45. Agnello L, Iacona A, Lo Sasso B, Scazzone C, Pantuso M, Giglio RV, et al. A new tool for sepsis screening in the emergency department. Clin Chem Lab Med 2021 Aug 26;59(9):1600-1605 [FREE Full text] [doi: 10.1515/cclm-2021-0208] [Medline: 33851525]
46. Cabitza F, Campagner A, Sconfienza LM. As if sand were stone. New concepts and metrics to probe the ground on which to build trustable AI. BMC Med Inform Decis Mak 2020 Sep 11;20(1):219 [FREE Full text] [doi: 10.1186/s12911-020-01224-9] [Medline: 32917183]
47. Decision Support System Quality Assessment homepage. Decision Support System Quality Assessment. URL: https://dss-quality-assessment.vercel.app/ [accessed 2025-01-13]
48. Ferri C, Flach P, Hernández-Orallo J. Delegating classifiers. In: Proceedings of the Twenty-First International Conference on Machine Learning. 2004 Presented at: ICML '04; July 4-8, 2004; Banff, AB. [doi: 10.1145/1015330.1015395]
49. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med 2021 Jan 05;4(1):4 [FREE Full text] [doi: 10.1038/s41746-020-00367-3] [Medline: 33402680]
50. Megerian JT, Dey S, Melmed RD, Coury DL, Lerner M, Nicholls CJ, et al. Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. NPJ Digit Med 2022 May 05;5(1):57 [FREE Full text] [doi: 10.1038/s41746-022-00598-6] [Medline: 35513550]
51. Campagner A, Cabitza F, Berjano P, Ciucci D. Three-way decision and conformal prediction: isomorphisms, differences and theoretical properties of cautious learning approaches. Inf Sci 2021 Nov;579:347-367. [doi: 10.1016/j.ins.2021.08.009]
52. Vovk V, Gammerman A, Shafer G. Algorithmic Learning in a Random World. New York, NY: Springer; 2005.
53. AndreaCampagner / scikit-cautious. GitHub. URL: https://github.com/AndreaCampagner/scikit-cautious [accessed 2025-01-13]
54. Wallis S. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. J Quant Linguist 2013 Jul 04;20(3):178-208. [doi: 10.1080/09296174.2013.799918]
55. Riley RD, Debray TP, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021 Aug 30;40(19):4230-4251. [doi: 10.1002/sim.9025] [Medline: 34031906]
56. Boucheron S, Lugosi G, Massart P. Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford, UK: Oxford University Press; 2013.
57. Gavrilov Y, Benjamini Y, Sarkar SK. An adaptive step-down procedure with proven FDR control under independence. Ann Stat 2009;37(2):619-629. [doi: 10.1214/07-aos586]
58. Metimeter: an AI/ML DSS Quality Assessment tool. GitHub. URL: https://mudilab.github.io/dss-quality-assessment/ [accessed 2024-12-25]
59. Lakomkin N, Sathiyakumar V, Wick B, Shen MS, Jahangir AA, Mir H, et al. Incidence and predictive risk factors of postoperative sepsis in orthopedic trauma patients. J Orthop Traumatol 2017 Jun 15;18(2):151-158 [FREE Full text] [doi: 10.1007/s10195-016-0437-4] [Medline: 27848054]
60. Steyerberg EW, Harrell FEJ. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016 Jan;69:245-247 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.005] [Medline: 25981519]
61. Wacker C, Prkno A, Brunkhorst FM, Schlattmann P. Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. Lancet Infect Dis 2013 May;13(5):426-435. [doi: 10.1016/s1473-3099(12)70323-7]
62. Ross P, Spates K. Considering the safety and quality of artificial intelligence in health care. Jt Comm J Qual Patient Saf 2020 Oct;46(10):596-599 [FREE Full text] [doi: 10.1016/j.jcjq.2020.08.002] [Medline: 32878718]

XSL•FO
RenderX

63. Vered M, Livni T, Howe PD, Miller T, Sonenberg L. The effects of explanations on automation bias. Artif Intell 2023 Sep;322:103952. [doi: 10.1016/j.artint.2023.103952]
64. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May 13;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
65. Henry KE, Kornfield R, Sridharan A, Linton RC, Groh C, Wang T, et al. Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. NPJ Digit Med 2022 Jul 21;5(1):97 [FREE Full text] [doi: 10.1038/s41746-022-00597-7] [Medline: 35864312]
66. Kovach MA, Standiford TJ. The function of neutrophils in sepsis. Curr Opin Infect Dis 2012;25(3):321-327. [doi: 10.1097/qco.0b013e3283528c9b]
67. AndreaCampagner / SepsisMDW. GitHub. URL: https://github.com/AndreaCampagner/SepsisMDW [accessed 2025-01-13]

Abbreviations

AI: artificial intelligence
A-PPV: average positive predictive value
AR-ED: Arezzo, Emergency Department
AUC: area under the receiver operating characteristic curve
CBC: complete blood count
COMPOSER: Conformal Multidimensional Prediction of Sepsis Risk
CRP: C-reactive protein
DT: decision tree
ED: emergency department
HC: high-confidence
HCT: hematocrit test
HGB: hemoglobin level
ICU: intensive care unit
LIS: laboratory information system
LR: logistic regression
MDW: monocyte distribution width
ML: machine learning
NPV: negative predictive value
OGSA-ICU: Ospedale Galeazzi Sant'Ambrogio, Intensive Care Unit
PA-ICU: Palermo, Intensive Care Unit
PD-ICU: Padova, Intensive Care Unit
PDR: Partial Decision Rule
PLT: platelet count
PPV: positive predictive value
RF: random forest
sNB: standardized net benefit
SOFA: Sequential Organ Failure Assessment
SVM: support vector machine
UD-ED: Udine, Emergency Department
XGB: extreme gradient boosting

Edited by A Coristine; submitted 15.12.23; peer-reviewed by E Kawamoto, Y Zhang; comments to author 15.03.24; revised version received 04.05.24; accepted 09.09.24; published 26.02.25.

Please cite as:
Campagner A, Agnello L, Carobene A, Padoan A, Del Ben F, Locatelli M, Plebani M, Ognibene A, Lorubbio M, De Vecchi E, Cortegiani A, Piva E, Poz D, Curcio F, Cabitza F, Ciaccio M
Complete Blood Count and Monocyte Distribution Width–Based Machine Learning Algorithms for Sepsis Detection: Multicentric Development and External Validation Study
J Med Internet Res 2025;27:e55492
URL: https://www.jmir.org/2025/1/e55492
doi:10.2196/55492
PMID:

XSL•FO
RenderX

©Andrea Campagner, Luisa Agnello, Anna Carobene, Andrea Padoan, Fabio Del Ben, Massimo Locatelli, Mario Plebani, Agostino Ognibene, Maria Lorubbio, Elena De Vecchi, Andrea Cortegiani, Elisa Piva, Donatella Poz, Francesco Curcio, Federico Cabitza, Marcello Ciaccio. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Stroke Diagnosis and Prediction Tool Using ChatGLM: Development and Validation Study

Xiaowei Song[1*], MD, PhD; Jiayi Wang[2*], BS; Feifei He[3], MD; Wei Yin[4], MA; Weizhi Ma[5*], PhD; Jian Wu[1*], MD, PhD

[1]Department of Neurology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

[2]Harbin Institute of Technology, Harbin, China

[3]Department of Neurology, Beijing Geriatric Hospital, Beijing, China

[4]School of Biomedical Engineeering, Tsinghua Medicine, Tsinghua University, Beijing, China

[5]Institute for AI Industry Research, Tsinghua University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Jian Wu, MD, PhD
Department of Neurology
Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University
No.168 of Litang Road
Beijing, 102218
China
Phone: 86 01056118918
Email: wujianxuanwu@126.com

Abstract

Background: Stroke is a globally prevalent disease that imposes a significant burden on health care systems and national economies. Accurate and rapid stroke diagnosis can substantially increase reperfusion rates, mitigate disability, and reduce mortality. However, there are considerable discrepancies in the diagnosis and treatment of acute stroke.

Objective: The aim of this study is to develop and validate a stroke diagnosis and prediction tool using ChatGLM-6B, which uses free-text information from electronic health records in conjunction with noncontrast computed tomography (NCCT) reports to enhance stroke detection and treatment.

Methods: A large language model (LLM) using ChatGLM-6B was proposed to facilitate stroke diagnosis by identifying optimal input combinations, using external tools, and applying instruction tuning and low-rank adaptation (LoRA) techniques. A dataset containing details of 1885 patients with and those without stroke from 2016 to 2024 was used for training and internal validation; another 335 patients from two hospitals were used as an external test set, including 230 patients from the training hospital but admitted at different periods, and 105 patients from another hospital.

Results: The LLM, which is based on clinical notes and NCCT, demonstrates exceptionally high accuracy in stroke diagnosis, achieving 99% in the internal validation dataset and 95.5% and 79.1% in two external test cohorts. It effectively distinguishes between ischemia and hemorrhage, with an accuracy of 100% in the validation dataset and 99.1% and 97.1% in the other test cohorts. In addition, it identifies large vessel occlusions (LVO) with an accuracy of 80% in the validation dataset and 88.6% and 83.3% in the other test cohorts. Furthermore, it screens patients eligible for intravenous thrombolysis (IVT) with an accuracy of 89.4% in the validation dataset and 60% and 80% in the other test cohorts.

Conclusions: We developed an LLM that leverages clinical text and NCCT to identify strokes and guide recanalization therapy. While our results necessitate validation through widespread deployment, they hold the potential to enhance stroke identification and reduce reperfusion time.

(J Med Internet Res 2025;27:e67010) doi:10.2196/67010

KEYWORDS

stroke; diagnosis; large language model; ChatGLM; generative language model; primary care; acute stroke; prediction tool; stroke detection; treatment; electronic health records; noncontrast computed tomography

XSL•FO
RenderX

Introduction

Background

Stroke ranks among the leading causes of morbidity and mortality worldwide. The overall burden of stroke, in terms of the absolute number of individuals affected or left disabled by the condition, has been increasing, particularly in China [1-3]. There are significant disparities in stroke incidence, prevalence, acute care, rehabilitation, risk factor management, and outcomes [4]. In recent years, substantial progress has been made in stroke diagnosis and treatment, with effective recanalization therapies for acute ischemic stroke established to significantly improve clinical prognosis [5,6]. However, treatment for acute ischemic stroke is highly time-dependent. Despite ample evidence supporting intravenous thrombolysis (IVT) and endovascular therapy for acute ischemic stroke, the number of patients who arrive at hospitals and are eligible for IVT remains limited, only 10% in the United States and 6% in China, with the global rate for IVT being lower than 5%. In addition, fewer than 100,000 thrombectomies are performed annually, despite the high incidence of ischemic strokes caused by large vessel occlusion (LVO). Access to recanalization therapies also varies across different regions [7,8]. Furthermore, the limited number of stroke neurologists is insufficient to meet the growing number of new stroke cases each year. Consequently, many patients in underdeveloped areas do not receive adequate care in the event of a sudden stroke. Inequities in stroke care are prevalent, and the costs also vary significantly [4,9-11].

The emergence of artificial intelligence (AI) and its application in health care has shown great promise in bridging existing gaps in medical practice [12,13]. Several studies have demonstrated the potential of large language models (LLMs) in managing various diseases [14,15]. In addition, LLMs have proven their effectiveness in medical diagnostics. For example, the Med-PaLM model illustrates the capabilities of LLMs in enhancing medical decision-making, achieving performance levels comparable to those required for US medical licensing examinations [16]. Furthermore, LLMs are used to extract summaries from electronic health records, making medical data more accessible and manageable for health care providers [17]. The integration of AI in clinical decision-making could reduce interrater variability in routine practice and facilitate the extraction of critical information, thereby improving the identification of patients with stroke, predicting treatment responses, and enhancing patient outcomes [14].

Objective

The objective of this study was to develop an acute stroke diagnosis tool that guides key therapies based on ChatGLM-6B and to verify its accuracy in diagnosing and predicting strokes. The application of this model in primary care could enhance the standard workflow for stroke diagnosis, identify patients who could benefit from recanalization, and facilitate risk prediction.

Methods

Study Design

We used a retrospective cohort for model training and internal validation while using a prospective cohort from various hospitals for external validation.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Beijing Tsinghua Changgung Hospital Ethics Committee (23641-0-01). Written informed consent was waived due to the analysis of electronic health records, and all research data used in this study are anonymous.

Data Acquisition

Training data were obtained from Beijing Tsinghua Changgung Hospital, a comprehensive stroke center and teaching hospital. Subjects admitted to the emergency room (ER) between January 2016 and January 2024 were screened continuously. Of the 341,026 patients admitted to the study hospital's ER during this period, 33,762 presented with focal neurological symptoms and were referred to the neurology ER. Among these, 12,600 were diagnosed with cerebrovascular disease (*ICD-10* [*International Statistical Classification of Diseases and Related Health Problems* 10th Revision]), and 991 were diagnosed with stroke and exhibited symptoms within 24 hours. The control group was randomly selected from patients admitted to the neurology ER who were diagnosed with noncerebrovascular diseases, ensuring a match for age, sex, admission time, and attending physicians after removing duplicate admissions. A focal neurologic deficit refers to a dysfunction in the nerves, spinal cord, or brain that affects a specific area of the body, such as the left side of the face, the right arm, or even a localized region like the tongue. In addition, issues related to speech, vision, and cognitive function are also classified as focal neurological deficits; the number of cases in each symptom category was shown in Multimedia Appendix 1. A total of 1885 cases satisfying the criteria above were included to form the training set. An additional group of patients from the same hospital (n=230) and another hospital (n=105) over the course of another month was used for the external test.

Demographic characteristics, emergency clinical notes, radiological images and reports, as well as other relevant examination records, were obtained from the hospital information system. The subjects were categorized into stroke and nonstroke groups based on the discharge diagnosis. Among the patients with stroke, further classification was adopted to divide them into ischemic and hemorrhagic strokes. Patients with ischemic stroke who received IVT or had LVO, confirmed by computed tomography angiography or magnetic resonance angiography, were also identified, regardless of whether they underwent endovascular therapy. In this study, large vessel occlusion is defined as ischemia resulting from the obstruction of a corresponding large vessel, which includes the anterior cerebral artery, middle cerebral artery, posterior cerebral artery, basilar artery, and vertebral artery. A detailed description of the subject screening process is provided in Figure 1. All stroke

XSL•FO
RenderX

diagnoses and workflows adhered to the recommendations outlined in the acute ischemic stroke management guidelines [18].

Figure 1. Flowchart of training dataset inclusion. BTCH: Bingjing Tsinghua Changgung Hospital; ER: emergency room; IVT: intravenous thrombolysis; LVO: large vessel occlusion.

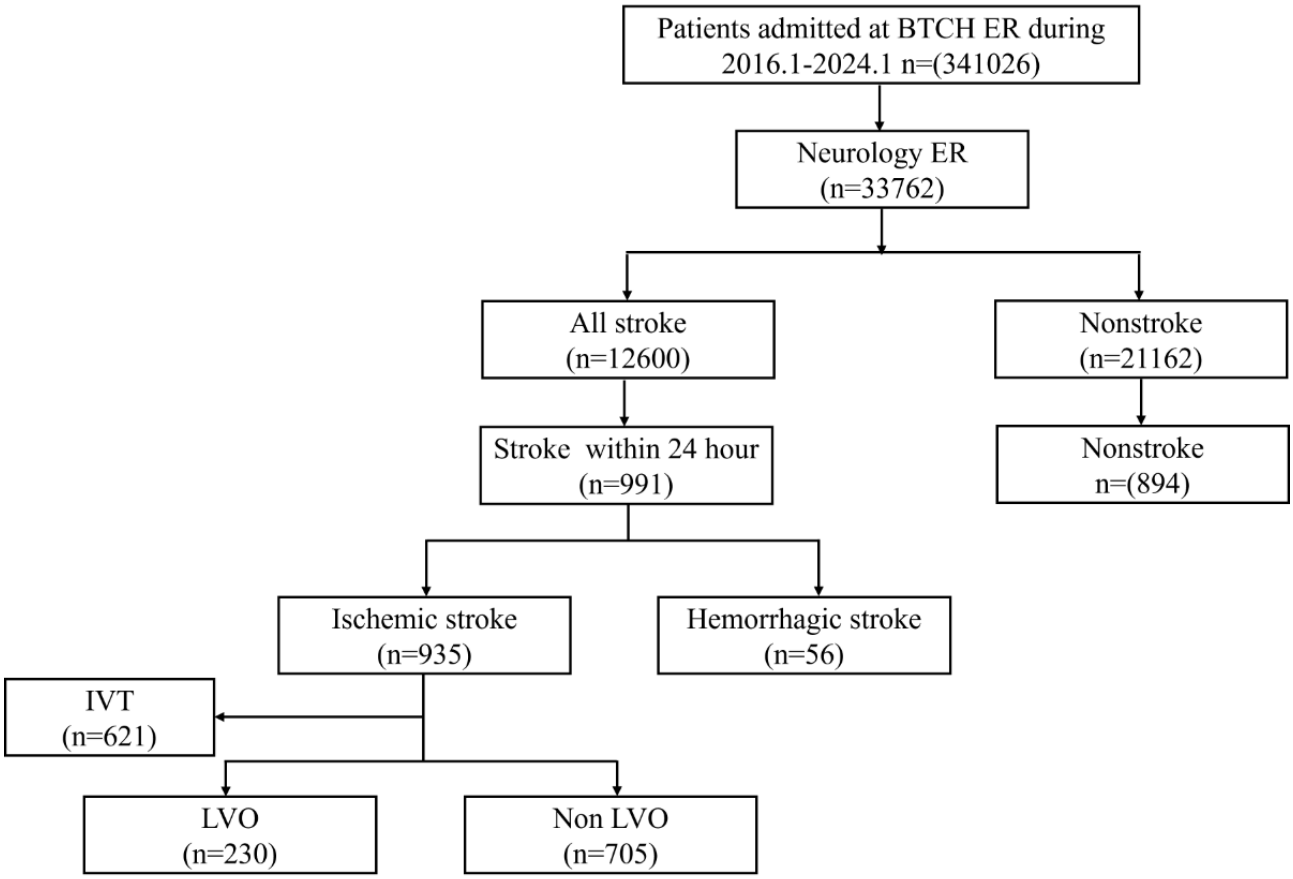

Data Preprocessing

Each record included demographic details such as age, sex, mode of arrival to the hospital (via emergency medical dispatch system or otherwise), main complaints, present medical history related to the current event, previous medical history, neurological examination, auxiliary examinations (including blood tests, electrocardiogram, and brain computed tomography as radiology reports), allergy history, marital and reproductive history, personal history, and family history. Notably, some records contained missing items for specific clinical parameters. We addressed these gaps by indicating "information not provided."

It was observed that some records included details such as past medical history and allergy history within the section on the history of the current illness. These details were extracted using methods such as regular expressions. In addition, information related to COVID-19 epidemiological investigations was present; however, since it was irrelevant to the diagnosis of the patient's condition, it was removed. All cases were randomly divided into training and internal validation datasets in a 9:1 ratio.

Foundation Model: ChatGLM-6B

To address the aforementioned challenges, we used ChatGLM-6B as the foundational model, which serves as the core of our acute stroke diagnosis tool, enabling high-precision analysis of clinical data. The ChatGLM-6B model was chosen due to its relatively low deployment requirements for hospital environments and its outstanding performance capabilities.

Prompt

The following prompt was used:

"As a neurology physician, please review the clinical notes and provide a final diagnosis based on the information provided. Kindly address the following questions: (1) Is the patient experiencing a stroke? (2) If so, is it ischemic or hemorrhagic? (3) If it is an ischemic stroke, does the patient require IVT? (4) If it is an ischemic stroke, is it caused by large vessel occlusion?" (see Figure 2 and Figure 3).

For clarity, when responding to the second, third, and fourth questions, we extracted specific subsets from the original dataset, ensuring that each question presented only two options. For example, when addressing the second question, we exclusively extracted records of stroke cases from the training, validation, and test sets.

XSL•FO
RenderX

Figure 2. Acute ischemic stroke management workflow and key issues. BMT: best medical treatment; CTP: computed tomography perfusion; ER: emergency room; EVT: endovascular thrombectomy; IVT: intravenous thrombolysis; LVO: large vessel occlusion; MRP: magnetic resonance perfusion.

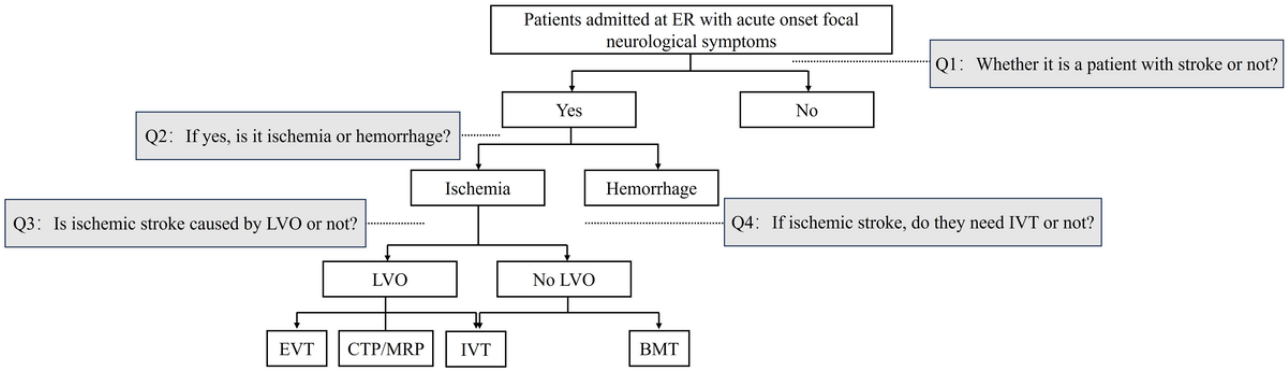

Figure 3. An example of instruction fine-tuning format.

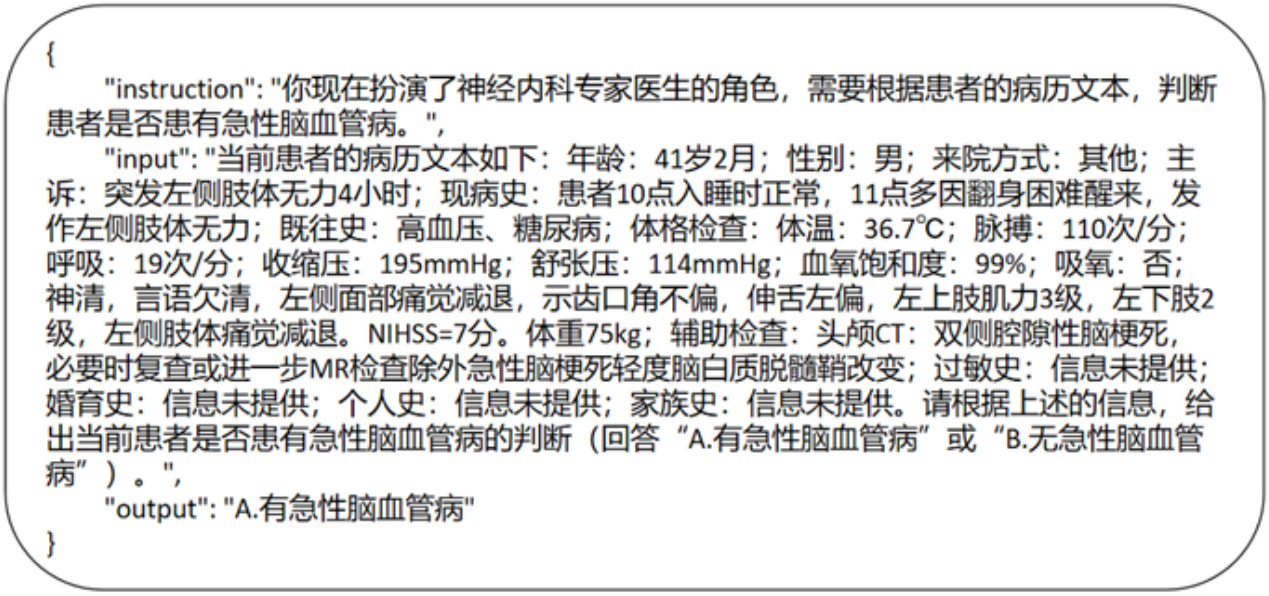

Features of the Medical Records

The electronic medical records contain a wealth of information, including primary complaints, medical history related to the current event, previous medical history, and more. Based on the completeness of various entries, certain combinations of entries are manually configured to select the optimal combination for different tasks, as illustrated in the textbox below (Textbox 1).

Textbox 1. Different electronic health record entry combinations.

Electronic health record entry combinations

- （ALL) Age, sex, mode of arrival, main complaints, medical history for this event, previous medical history, neurological examination, auxiliary examination, allergy history, marital and reproductive history, personal history, and family history.
- Age, sex, mode of arrival, main complaints, medical history for this event, previous medical history, neurological examination, auxiliary examination, allergy history, marital and reproductive history.
- Age, sex, mode of arrival, main complaints, medical history for this event, previous medical history, neurological examination, and auxiliary examination.
- Age, sex, mode of arrival, main complaints, medical history for this event, neurological examination, and auxiliary examination.
- Age, sex, mode of arrival, medical history for this event, previous medical history, neurological examination, auxiliary examination.
- Age, sex, mode of arrival, medical history for this event, neurological examination, auxiliary examination.
- Age, sex, medical history for this event, previous medical history, neurological examination, and auxiliary examination.
- Age, sex, medical history for this event, neurological examination, and auxiliary examination.

XSL•FO
RenderX

In-Context Learning

In-context learning is a key ability of LLMs that allows them to handle new tasks with just a few examples. Few-shot learning is considered one of the most effective methods for in-context learning [19]. By using prompting techniques, foundational models can quickly adapt to a specific domain and learn to follow the task format with just a few demonstrations. This approach uses a mechanism to identify examples based on their similarity to the current case [20].

To identify representative few-shot examples, we followed this procedure: for each test example, we selected k training examples that are semantically similar using k-nearest neighbors clustering within the embedding space. Specifically, we used ChatGLM-6B to embed both training and test questions into vector representations. Then, for each test question (x), we retrieved its k-nearest neighbors from the training set. Unlike fine-tuning, dynamic few-shot learning uses the training data without requiring extensive updates to the model parameters.

Instruction Fine-Tuning and LoRA

In this study, we used instruction tuning and low-rank adaptation (LoRA) techniques to enhance the performance of LLMs on specific medical record tasks [21,22]. Instruction tuning is a training method that provides explicit task instructions to the model during the training process, enabling it to better understand and execute these instructions, thereby improving its performance on targeted tasks. Conversely, LoRA technology adapts the model's weight matrices through low-rank modifications, allowing for fine-tuning of model behaviors without the need to retrain the entire model. This approach improves both training and inference efficiency while preserving the model's complexity.

Specifically, we first developed distinct task instructions based on the four identified issues, which represent critical questions in the stroke diagnosis and treatment workflow (Figure 2). These instructions formed the instruction component of the directive. The primary entries from the medical records were concatenated to create the medical record text, which served as the Input component of the instruction. Finally, the labels corresponding to the questions were designated as the Output component of the instruction. To select the primary features from the medical records, combinations of features were manually chosen based on the completeness of various characteristics within the records (Textbox 1). The best-performing features under a zero-shot scenario, without fine-tuning, were then selected for use during the fine-tuning process. An example is provided below.

Tool Chaining with LLMs

Tool chaining with LLMs involves the integration of various pretrained tools, such as vision models, to enhance the capabilities of LLMs. This process does not necessitate additional model training. For instance, in implementations like Visual ChatGPT and X-GPT, an LLM can assign specific tasks to the appropriate tools to manage subtasks that exceed its inherent capabilities [23,24].

In this study, we provided noncontrast computed tomography scans along with the corresponding radiology reports. By using a toolchain approach, we integrated these examination results into the medical records text and input them collectively into the large model, thereby eliminating the need for any additional tools. Furthermore, we compared the model's performance with and without tool chaining, selecting the more effective option for various tasks.

Parameter

The parameter settings for LoRA fine-tuning are as follows: LoRA_rank is set to 8, batch size to 6, gradient accumulation steps to 1, max steps to 2000, fp16 is enabled, and the learning rate is set at 5e-5.

Results

Overview

A total of 1885 cases were included in the training set, comprising 991 stroke cases and 894 nonstroke cases. Among the stroke cases, there were 935 instances of ischemic stroke; of these, 621 patients received IVT, and 230 presented with large vessel occlusion, which was confirmed by magnetic resonance angiography or computed tomography angiography.

Performance on the Validation Set

We use the performance of the LLM in zero-shot and few-shot (3-shot) settings as the baseline, while the results obtained from fine-tuning serve as the final outcomes.

For zero-shot and few-shot scenarios, we experimented with various combinations of entries and the use of tools, ultimately selecting the most effective approach. For fine-tuning, we adopted the best method identified in the zero-shot scenario for adjustments and evaluated it under this framework. The primary results are presented in Table 1.

XSL•FO
RenderX

Table 1. Performance of large language models on stroke diagnosis on internal validation set using different methods.

Question and method	Approach	Accuracy	Sensitivity	Specificity
Whether it is a patient with stroke or not?				
Zero-shot	No.1[a]+Tool[b]	0.713	0.702	0.722
Few-shot (3)	No.2+Tool	0.784	0.837	0.722
Fine-tune	No.1+Tool	0.99	1	0.978
If yes, is it ischemia or hemorrhage?				
Zero-shot	No.1+Tool	0.606	0.564	1
Few-shot (3)	No.8	0.894	0.883	1
Fine-tune	No.1+Tool	1	1	1
If ischemic stroke, do they need intravenous thrombosis or not?				
Zero-shot	No.8+Tool	0.617	0.776	0.222
Few-shot (3)	No.3	0.649	0.776	0.333
Fine-tune	No.8+Tool	0.894	0.985	0.667
Is ischemic stroke caused by LVO or not?				
Zero-shot	No.8	0.585	0.625	0.571
Few-shot (3)	No.4+Tool	0.713	0.417	0.814
Fine-tune	No.8	0.8	0.5	0.9

[a]No.x: the xth entry combination shown in Textbox 1.

[b]Tool: using the tool chaining, ie, adding imaging examination results.

It can be observed that after few-shot learning, the model's performance significantly improved on questions 1, 2, and 4, with some improvement also noted for question 3. Following the application of LoRA, there was a substantial enhancement in performance across all questions.

Overall Performance Across Different Datasets

Using the fine-tuned weights, the same tests were conducted on the test set. Specifically, because the electronic medical records provided by other hospitals lacked information on the mode of arrival, marital status, and reproductive history, as well as imaging examination results, these components were excluded from the input. Instead, the "nervous system" data, which was available in this dataset but not originally included, was incorporated. The main results are presented in Table 2. For the key procedures in the stroke diagnosis workflow, the model demonstrates relatively satisfactory performance in stroke diagnosis, classification, and prediction.

The model demonstrates exceptionally high accuracy in stroke diagnosis, achieving 99% in the internal validation cohort. Further validation across different cohorts and other stroke centers indicates that the accuracy remains at 95.5% within the same hospital and 79.1% in other hospitals. This model exhibits excellent performance in identifying both ischemic and hemorrhagic strokes, with accuracy reaching as high as 100% in internal validation and 99.1% and 97.1% in external validation. Regarding IVT for ischemic stroke, the model's accuracy ranges from 60% to 89.4%. In addition, the model is capable of identifying large vessel occlusions, with prediction accuracy reaching up to 80% across different cohorts. However, it is important to note that while the model demonstrates high specificity, it has limited sensitivity (see Table 2 and Figure 4).

XSL•FO
RenderX

Figure 4. Confusion matrix of large language models on stroke diagnosis workflow key issues. IVT: intravenous thrombolysis; LVO: large vessel occlusions.

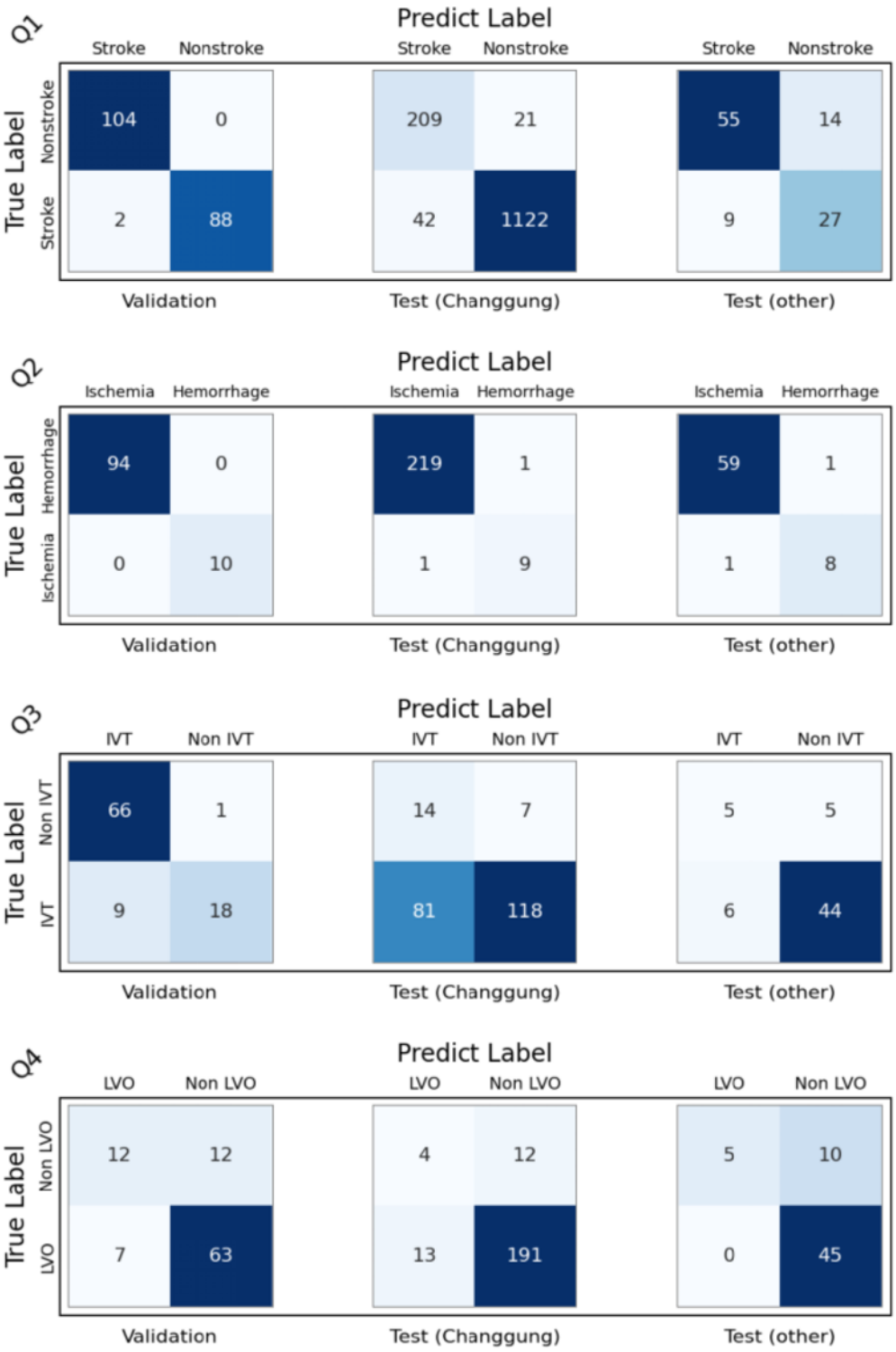

XSL•FO
RenderX

Table 2. Performance of large language models on stroke diagnosis based on low-rank adaptation on both internal validation and external test set.

Question and dataset	Accuracy	Sensitivity	Specificity
Whether it is a patient with stroke or not?			
Validation	0.99	1	0.978
Test			
Changgung (n=230)	0.955	0.909	0.964
Other (n=105)	0.791	0.812	0.75
If yes, is it ischemia or hemorrhage?			
Validation	1	1	1
Test			
Changgung (n=230)	0.991	0.996	0.9
Other (n=105)	0.971	0.983	0.889
If ischemic stroke, do they need intravenous thrombosis or not?			
Validation	0.894	0.985	0.667
Test			
Changgung (n=230)	0.6	0.667	0.593
Other (n=105)	0.817	0.5	0.88
If ischemic stroke, is it caused by LVO or not?			
Validation	0.8	0.5	0.9
Test			
Changgung (n=230)	0.886	0.25	0.936
Other (n=105)	0.833	0.333	1

We also conducted zero-shot and few-shot tests on the test set, using the optimal combinations selected from the validation set for both scenarios. The results were compared with those obtained from LoRA, as shown in Table S3 in Multimedia Appendix 2 and Table S4 in Multimedia Appendix 3.

Verify the Effectiveness of Selecting the Optimal Entry Combinations

Since fine-tuning with various entry combinations is time-consuming, we focus on fine-tuning based on the best combination identified from the zero-shot results. The optimal combinations for questions 1 and 2 consist of all entries combined with medical imaging examination results. We then compare the best combinations for questions 3 and 4 against the all-inclusive combination of entries and medical imaging examination results on the validation set. This comparison aims to demonstrate the effectiveness of selecting specific combinations rather than simply opting for the maximum number of entries (see Table 3). For question 3, the optimal combination yielded significantly better performance than using all available information. In question 4, while the accuracy of the optimal combination was slightly lower than that of the all-inclusive approach, it exhibited a more balanced performance in terms of sensitivity and specificity; the *P* values for the comparison between best and all among questions 3 and 4 were <0.001 and 0.017, respectively.

Table 3. Comparison between selecting the optimal combination and selecting all information on the validation set.

Question and method[a]	Accuracy	Sensitivity	Specificity
If ischemic stroke, do they need intravenous thrombosis or not?			
Best[a]	0.8936	0.9851	0.6667
All[b]	0.8404	0.9701	0.5185
If ischemic stroke, is it caused by LVO[c] or not?			
Best	0.8	0.5	0.9
All	0.8085	0.3333	0.9714

[a]Best: the best entry combinations for questions.

[b]All: all entries and medical imaging examination results.

[c]LVO: large vessel occlusions.

XSL•FO
RenderX

Discussion

Principal Findings

To the best of our knowledge, this is the first LLM developed for stroke diagnosis. Domain-specific models are typically more efficient than general-purpose models. Our model has demonstrated excellent performance in stroke diagnosis and classification. Compared to other LLMs, ChatGLM-6B has an affordable advantage in health care purposes, which is critically important in the medical domain. The easy deployment of our model could undoubtedly facilitate the wide-scale application in primary care.

A Review of Previous Stroke Diagnosis Tool and Their Limitations

Addressing inequities in acute stroke care necessitates a thorough examination of each component of regional workflow, particularly the collaboration among various stroke centers [19]. In addition to raising awareness of stroke symptoms among high-risk populations, the most critical step is to identify patients with stroke as early as possible at primary stroke centers. Although several scales existed for stroke diagnosis, they often exhibited limited sensitivity and specificity. The Cincinnati Prehospital Stroke Scale (CPSS) is a clinical evaluation tool for stroke diagnosis, comprising three items: facial droop, upper limb weakness, and speech difficulties. If a patient presents with one of these three symptoms, there is a 72% likelihood that they are experiencing a stroke; if all three symptoms are present, the probability increases to 85% [20,21]. The 3-Item Stroke Scale (3-ISS) is used to identify proximal vessel occlusion (specifically T-segment or M1-segment occlusion of the middle cerebral artery), with an overall accuracy of 86%, sensitivity of 67%, and specificity of 92% [22]. However, these evaluations are primarily based on clinical symptoms, which can lead to an underestimation of stroke diagnoses in atypical patients and an overestimation in individuals with other neurological conditions. In addition, there are several large vessel occlusion scales based on imaging techniques, such as the Alberta Stroke Program Early CT Score (ASPECTS) and multimodal perfusion imaging. Computed tomography perfusion is the most widely used multimodal imaging method in comprehensive stroke centers, and numerous automated postprocessing workstations, such as Rapid AI and Viz.AI, assist in LVO and penumbra calculations. These technologies can enhance LVO workflow to some extent and reduce door-to-groin time [12,23]. However, there are limitations to this assessment. First, computed tomography perfusion may not be available in some primary care settings. Second, automated postprocessing workstations are not universally implemented across all hospitals. Third, the varying algorithms used by different products can also impact the final diagnosis [24].

Clinical Implications

Our results demonstrated the significant potential of LLMs in stroke diagnosis and management, particularly in differentiating stroke mimics. The relatively low accuracy of stroke diagnosis in external validation may be attributed to discrepancies in the clinical notes template and deficiencies in radiology reports from other hospitals. Our model, which is based on initial visit records and brain computed tomography scans, achieved an accuracy of nearly 80% in identifying LVO and recommending IVT. This level of accuracy is comparable to that of an experienced stroke neurologist and surpasses the performance of the Alberta Stroke Program Early CT Score [25]. Previous studies also support the notion that multimodal LLMs, when used to evaluate the prognosis of hemorrhagic stroke, can significantly enhance predictive performance compared to single modalities, such as clinical notes or imaging [26]. In contrast to other disease-specific language representation models, such as stroke-BERT [27], ChatGLM-6B is a general-purpose language model with extensive capabilities in language understanding and generation. This allows it to capture semantic and contextual information more comprehensively when processing diverse medical records. In addition, the GLM framework supports multimodal data input and can integrate information from images, which is particularly crucial in the diagnosis of cerebrovascular diseases. When compared to traditional machine learning and deep learning models, LLMs often demonstrate superior performance across a range of language-based tasks. For instance, LLMs have been successfully applied in diagnosing medical conditions, assisting in medical education, and even generating medical documentation, where they can outperform more traditional models that require extensive feature engineering and fine-tuning for each specific task [17]. This is the first LLM tool in stroke diagnosis; the model can give some guidance on stroke diagnosis, classification, and management recommendations based on the primary information. As we know, there exist great discrepancies in stroke care across different hospitals. Stroke centers were more likely to receive thrombolytic therapy and had lower 30-day mortality rates [28]. Thus, the deployment of this model in primary care, especially where stroke neurologists are in high demand, could not only improve the ability of stroke diagnosis early and accurately, but also improve reperfusion therapy for patients who are very likely to have LVO, and they can be transferred to advanced stroke center as soon as possible, saving both time and economic costs. A previous study indicated that the use of artificial intelligence decision support in a hyperacute stroke pathway facilitates decision-making and can improve the rate and time of reperfusion therapies in a hub-and-spoke system of care [29]. There are also data supports that transfer-in patients had similar in-hospital mortality rates compared with front-door patients [30].

Limitations and Future Work

This study has several limitations. First, it only included data from two centers in the real world for external validation, and the data were exclusively from the Chinese population, which could introduce some data bias. We still need more data from different stroke centers to validate the accuracy of our model. Second, electronic health records may be fragmented and the clinical notes could also be subjective and depend on different physicians, which could possibly have an effect on the test results. However, the advanced language comprehension capabilities of LLMs enable them to extract crucial information from diverse and complex clinical notes, even when the documentation varies in style and contains incomplete data. This capacity significantly enhances the quality and efficiency

XSL•FO
RenderX

of medical diagnostics. Finally, despite the satisfactory performance of this model in stroke diagnosis, as an LLM, the hallucination is unavoidable. We strive to minimize the possibility by adjusting the temperature, paying more attention to the prompt project, and verifying the results in an external dataset. As an auxiliary diagnostic tool, we must interpret its suggestions with caution.

Conclusion

In this study, we developed an LLM for stroke diagnosis with quite good performance in in-context understanding and reasoning skills by selecting optimal entry combinations, using tools, and fine-tuning with LoRA, proving its effectiveness. Our model also performs well in handling missing values, whether individual data points lack information on certain entries or the dataset itself lacks corresponding entries. In general, AI holds promise as a tool to ease the stroke neurologist's shortage, improve daily workflow, and supply unique diagnostic insights by analyzing data simultaneously from several sources, including neurological history and examination, blood tests, and images.

Acknowledgments

This study is supported by the National Key Research and Development (R&D) Program of China (2023YFC2506600).

Authors' Contributions

XS: conceptualization, resources, and writing the original draft. JW: data curation, formal analysis, investigation, software, and visualization. FH: data curation, resources, validation, review and editing. WY: methodology, project administration, and supervision. WM: methodology, software, validation, and reviewing the manuscript. JW: supervision, funding acquisition, review and final proofreading. XS and JW contributed equally to this work and are co-first authors; WM and JW served as co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Symptom categories and number of cases in training set.
[PNG File , 240 KB - jmir_v27i1e67010_app1.png]

Multimedia Appendix 2
Performance of LLMs on stroke diagnosis on test set (changgung).
[DOCX File , 17 KB - jmir_v27i1e67010_app2.docx]

Multimedia Appendix 3
Performance of LLMs on stroke diagnosis on test set(other).
[DOCX File , 17 KB - jmir_v27i1e67010_app3.docx]

References

1. Feigin VL, Norrving B, Mensah GA. Global burden of stroke. Circ Res 2017;120(3):439-448. [doi: 10.1161/CIRCRESAHA.116.308413] [Medline: 28154096]
2. Tu WJ, Zhao Z, Yin P, Cao L, Zeng J, Chen H, et al. Estimated burden of stroke in China in 2020. JAMA Netw Open 2023;6(3):e231455 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.1455] [Medline: 36862407]
3. Wu S, Wu B, Liu M, Chen Z, Wang W, Anderson CS, China Stroke Study Collaboration. Stroke in China: advances and challenges in epidemiology, prevention, and management. Lancet Neurol 2019;18(4):394-405. [doi: 10.1016/S1474-4422(18)30500-3] [Medline: 30878104]
4. Zheng S, Zhang T, Gu H, Zhang Z, Yang X, Wang C, et al. Adherence to stroke care performance measures in different regions in China, 2015-2019: evidence from the Chinese stroke center alliance. Neurology 2022;99(16):e1767-e1778. [doi: 10.1212/WNL.0000000000201024] [Medline: 35918168]
5. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, ECASS Investigators. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. N Engl J Med 2008;359(13):1317-1329. [doi: 10.1056/NEJMoa0804656] [Medline: 18815396]
6. Berkhemer OA, Fransen PSS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. N Engl J Med 2015;372(1):11-20 [FREE Full text] [doi: 10.1056/NEJMoa1411587] [Medline: 25517348]
7. Saini V, Guada L, Yavagal DR. Global epidemiology of stroke and access to acute ischemic stroke interventions. Neurology 2021;97(20 Suppl 2):S6-S16. [doi: 10.1212/WNL.0000000000012781] [Medline: 34785599]

XSL•FO
RenderX

8. Yuan J, Lu ZK, Xiong X, Li M, Liu Y, Wang L, et al. Age and geographic disparities in acute ischaemic stroke prehospital delays in China: a cross-sectional study using national stroke registry data. Lancet Reg Health West Pac 2023;33:100693 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100693] [Medline: 37181525]
9. Feldmeier M, Kim AS, Zachrison KS, Alberts MJ, Shen Y, Hsia RY. Heterogeneity of state stroke center certification and designation processes. Stroke 2024;55(4):1051-1058. [doi: 10.1161/STROKEAHA.123.045368] [Medline: 38469729]
10. Prust ML, Forman R, Ovbiagele B. Addressing disparities in the global epidemiology of stroke. Nat Rev Neurol 2024;20(4):207-221. [doi: 10.1038/s41582-023-00921-z] [Medline: 38228908]
11. Yang Y, Nicholas S, Maitland E, Huang Z, Chen X, Ma Y, et al. An equity evaluation in stroke inpatients in regard to medical costs in China: a nationwide study. BMC Health Serv Res 2021;21(1):425 [FREE Full text] [doi: 10.1186/s12913-021-06436-x] [Medline: 33952266]
12. Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S, Tariq MB, Abdelkhaleq R, Niktabe A, et al. Automated large vessel occlusion detection software and thrombectomy treatment times: a cluster randomized clinical trial. JAMA Neurol 2023;80(11):1182-1190 [FREE Full text] [doi: 10.1001/jamaneurol.2023.3206] [Medline: 37721738]
13. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney ML, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. JAMA Netw Open 2018;1(5):e182665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2665] [Medline: 30646178]
14. Bivard A, Churilov L, Parsons M. Artificial intelligence for decision support in acute stroke - current roles and potential. Nat Rev Neurol 2020;16(10):575-585. [doi: 10.1038/s41582-020-0390-y] [Medline: 32839584]
15. Elkin PL, Mullin S, Mardekian J, Crowner C, Sakilay S, Sinha S, et al. Using artificial intelligence with natural language processing to combine electronic health record's structured and free text data to identify nonvalvular atrial fibrillation to decrease strokes and death: evaluation and case-control study. J Med Internet Res 2021;23(11):e28946 [FREE Full text] [doi: 10.2196/28946] [Medline: 34751659]
16. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
17. Van Veen D, Van Uden C, Blankemeier L, Delbrouck J, Aali A, Bluethgen C, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med 2024;30(4):1134-1142. [doi: 10.1038/s41591-024-02855-5] [Medline: 38413730]
18. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019;50(12):e344-e418 [FREE Full text] [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
19. Watson T, Tiu J, Clissold B. Addressing inequity in acute stroke care requires attention to each component of regional workflow. Med J Aust 2020;212(1):8-10.e1. [doi: 10.5694/mja2.50440] [Medline: 31785010]
20. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati prehospital stroke scale: reproducibility and validity. Ann Emerg Med 1999;33(4):373-378. [doi: 10.1016/s0196-0644(99)70299-4] [Medline: 10092713]
21. Mohamed GA, Marmarchi F, Fonkeu Y, Alshaer Q, Rangaraju S, Carr M, et al. Cincinnati prehospital stroke scale implementation of an urban county severity-based stroke triage protocol: impact and outcomes on a comprehensive stroke center. J Stroke Cerebrovasc Dis 2022;31(8):106575 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2022.106575] [Medline: 35661542]
22. Singer OC, Dvorak F, du Mesnil de Rochemont R, Lanfermann H, Sitzer M, Neumann-Haefelin T. A simple 3-item stroke scale: comparison with the national institutes of health stroke scale and prediction of middle cerebral artery occlusion. Stroke 2005;36(4):773-776. [doi: 10.1161/01.STR.0000157591.61322.df] [Medline: 15731478]
23. Van Orden K, Meyer DM, Perrinez ES, Torres D, Poynor B, Alwood B, et al. (VISIION-S): Viz.ai implementation of stroke augmented intelligence and communications platform to improve indicators and outcomes for a comprehensive stroke center and network - sustainability. J Stroke Cerebrovasc Dis 2023;32(10):107303. [doi: 10.1016/j.jstrokecerebrovasdis.2023.107303] [Medline: 37572556]
24. Pisani L, Haussen DC, Mohammaden M, Perry da Camara C, Jillella DV, Rodrigues GM, et al. Comparison of CT perfusion software packages for thrombectomy eligibility. Ann Neurol 2023;94(5):848-855. [doi: 10.1002/ana.26748] [Medline: 37584452]
25. Hoelter P, Muehlen I, Goelitz P, Beuscher V, Schwab S, Doerfler A. Automated ASPECT scoring in acute ischemic stroke: comparison of three software tools. Neuroradiology 2020;62(10):1231-1238. [doi: 10.1007/s00234-020-02439-3] [Medline: 32382795]
26. Chen Y, Jiang C, Chang J, Qin C, Zhang Q, Ye Z, et al. An artificial intelligence-based prognostic prediction model for hemorrhagic stroke. Eur J Radiol 2023;167:111081. [doi: 10.1016/j.ejrad.2023.111081] [Medline: 37716178]
27. Lin CH, Hsu KC, Liang CK, Lee TH, Liou CW, Lee JD, et al. A disease-specific language representation model for cerebrovascular disease research. Comput Methods Programs Biomed 2021;211:106446 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106446] [Medline: 34627022]

XSL•FO
RenderX

28. Xian Y, Holloway RG, Chan PS, Noyes K, Shah MN, Ting HH, et al. Association between stroke center hospitalization for acute ischemic stroke and mortality. JAMA 2011;305(4):373-380 [FREE Full text] [doi: 10.1001/jama.2011.22] [Medline: 21266684]
29. Gunda B, Neuhaus A, Sipos I, Stang R, Böjti PP, Takács T, et al. Improved stroke care in a primary stroke centre using AI-decision support. Cerebrovasc Dis Extra 2022;12(1):28-32 [FREE Full text] [doi: 10.1159/000522423] [Medline: 35134802]
30. Ali SF, Singhal AB, Viswanathan A, Rost NS, Schwamm LH. Characteristics and outcomes among patients transferred to a regional comprehensive stroke center for tertiary care. Stroke 2013;44(11):3148-3153. [doi: 10.1161/STROKEAHA.113.002493] [Medline: 24021682]

Abbreviations

3-ISS: 3-Item Stroke Scale
AI: artificial intelligence
ASPECTS: Alberta Stroke Program Early CT Score
CPSS: Cincinnati Prehospital Stroke Scale
ER: emergency room
ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision
IVT: intravenous thrombolysis
LLM: large language model
LoRA: low-rank adaptation
LVO: large vessel occlusion

Edited by A Coristine; submitted 29.09.24; peer-reviewed by X Jia, SF Sung; comments to author 29.11.24; revised version received 17.12.24; accepted 05.02.25; published 26.02.25.

Please cite as:
Song X, Wang J, He F, Yin W, Ma W, Wu J
Stroke Diagnosis and Prediction Tool Using ChatGLM: Development and Validation Study
J Med Internet Res 2025;27:e67010
URL: https://www.jmir.org/2025/1/e67010
doi:10.2196/67010
PMID:

©Xiaowei Song, Jiayi Wang, Feifei He, Wei Yin, Weizhi Ma, Jian Wu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning–Based Prediction of Early Complications Following Surgery for Intestinal Obstruction: Multicenter Retrospective Study

Pinjie Huang[1*], MD; Jirong Yang[1*], MD; Dizhou Zhao[2*], MD; Taojia Ran[1], MM; Yuheng Luo[1], MSc; Dong Yang[3], MSc; Xueqin Zheng[4], MD; Shaoli Zhou[1], MD; Chaojin Chen[1], MD

[1]Department of Anesthesiology, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[2]Department of Anesthesiology, ShenZhen People's Hospital, Shenzhen, China

[3]Guangzhou AI & Data Cloud Technology Co., Guangzhou, China

[4]Department of Anesthesiology, First People's Hospital of Foshan, Foshan, China

[*]these authors contributed equally

Corresponding Author:
Chaojin Chen, MD
Department of Anesthesiology
Third Affiliated Hospital of Sun Yat-sen University
No.600, Tianhe Road, Guangzhou
Guangzhou, 510630
China
Phone: 86 13430322182
Email: chenchj28@mail.sysu.edu.cn

Abstract

Background: Early complications increase in-hospital stay and mortality after intestinal obstruction surgery. It is important to identify the risk of postoperative early complications for patients with intestinal obstruction at a sufficiently early stage, which would allow preemptive individualized enhanced therapy to be conducted to improve the prognosis of patients with intestinal obstruction. A risk predictive model based on machine learning is helpful for early diagnosis and timely intervention.

Objective: This study aimed to construct an online risk calculator for early postoperative complications in patients after intestinal obstruction surgery based on machine learning algorithms.

Methods: A total of 396 patients undergoing intestinal obstruction surgery from April 2013 to April 2021 at an independent medical center were enrolled as the training cohort. Overall, 7 machine learning methods were used to establish prediction models, with their performance appraised via the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, and F_1-score. The best model was validated through 2 independent medical centers, a publicly available perioperative dataset the Informative Surgical Patient dataset for Innovative Research Environment (INSPIRE), and a mixed cohort consisting of the above 3 datasets, involving 50, 66, 48, and 164 cases, respectively. Shapley Additive Explanations were measured to identify risk factors.

Results: The incidence of postoperative complications in the training cohort was 47.44% (176/371), while the incidences in 4 external validation cohorts were 34% (17/50), 56.06% (37/66), 52.08% (25/48), and 48.17% (79/164), respectively. Postoperative complications were associated with 8-item features: Physiological Severity Score for the Enumeration of Mortality and Morbidity (POSSUM physiological score), the amount of colloid infusion, shock index before anesthesia induction, ASA (American Society of Anesthesiologists) classification, the percentage of neutrophils, shock index at the end of surgery, age, and total protein. The random forest model showed the best overall performance, with an AUROC of 0.788 (95% CI 0.709-0.869), accuracy of 0.756, sensitivity of 0.695, specificity of 0.810, and F_1-score of 0.727 in the training cohort. The random forest model also achieved a comparable AUROC of 0.755 (95% CI 0.652-0.839) in validation cohort 1, a greater AUROC of 0.817 (95% CI 0.695-0.913) in validation cohort 2, a similar AUROC of 0.786 (95% CI 0.628-0.902) in validation cohort 3, and the comparable AUROC of 0.720 (95% CI 0.671-0.768) in validation cohort 4. We visualized the random forest model and created a web-based online risk calculator.

XSL•FO
RenderX

Conclusions: We have developed and validated a generalizable random forest model to predict postoperative early complications in patients undergoing intestinal obstruction surgery, enabling clinicians to screen high-risk patients and implement early individualized interventions. An online risk calculator for early postoperative complications was developed to make the random forest model accessible to clinicians around the world.

(J Med Internet Res 2025;27:e68354) doi:10.2196/68354

KEYWORDS

postoperative complications; intestinal obstruction; machine learning; early intervention; risk calculator; prediction model; Shapley additive explanations

Introduction

Early postoperative complications refer to newly occurring situations or events that are detrimental to the patient's health from the first postoperative day until discharge, causing irreversible damage or requiring a change in treatment policy [1]. The incidence of early postoperative complications in intestinal obstruction was reported to be as high as 23%-28%, which contributed to an increased length of hospital stay and mortality [2,3]. Early complications after intestinal obstruction surgery include postoperative infection, wound dehiscence, important organs' dysfunction, intestinal fistula, postoperative bleeding, and other surgery-related complications [3,4]. Surgery is a reliable means for relieving obstruction, but for frail individuals, surgery, anesthesia, and various perioperative interventions pose challenges as well. Accurate prediction of early postoperative complications is essential for proper preoperative selection of surgical patients, determination of the necessary postoperative vigilance level, guidance of perioperative decision-making, and early intervention.

Currently, studies have found that delayed surgery, American Society of Anesthesiologists (ASA) physical status classification, age, and surgical methods may be risk factors for its occurrence [4-6]. A study proposed using the Acute General Emergency Surgical Severity–Small Bowel Obstruction (AGESS-SBO) score [7] to predict the prognosis of small-bowel obstruction (SBO), but the current analysis and early intervention still depend on expert consensus and guidelines. A predictive model for identifying high-risk patients is required to optimize the treatment strategy for intestinal obstruction. However, the intricate etiology, pathophysiology, and perioperative changes of intestinal obstruction render traditional logistic regression models not applicable.

Data-driven machine learning (ML) modeling, a technology used to build data-driven artificial intelligence systems, has the ability to diagnose and predict the prognosis of patients. On the one hand, ML has the advantages of capturing nonlinear relationships more comprehensively and predicting the prognosis accurately [8]; on the other hand, with the increasing complexity and dimensions of modern medical datasets, ML can give full play to the advantage of a rich data volume in big data. These factors make it popular in medicine, including anesthesiology [9], cardiology [10], ophthalmology [11], and so on. ML has been used to predict postoperative complications. Some cohort studies have applied ML to diagnose intestinal obstruction [12], but it has not been applied to predict early postoperative complications. While the performance of ML models is superior to traditional scoring systems, the limitation of these models is that they are mostly designed on the basis of single-center data. What is important is that the ML prediction models have not been transformed into practical applications.

In this multicenter retrospective study, we developed an online risk calculator for early postoperative complications of patients with intestinal obstruction in an independent medical center and performed external validation through two independent medical centers, a publicly available perioperative dataset Informative Surgical Patient dataset for Innovative Research Environment (INSPIRE) and a mixed cohort consisting of the above 3 external validation datasets.

Methods

Source of the Data and Participants

This retrospective study was performed using a multicenter database of patients who underwent surgery for intestinal obstruction: the Third Affiliated Hospital of Sun Yat-sen University, Shenzhen People's Hospital, Foshan First People's Hospital, and the INSPIRE dataset. INSPIRE is a publicly available research dataset for perioperative medicine, which includes approximately 130,000 patients (50% of all surgical patients) who underwent anesthesia for surgery at an academic institution in South Korea between 2011 and 2020 [13,14]. To our knowledge, it is a new dataset that contains data for collaborative research and development in perioperative medicine.

The medical data collected from electronic medical record systems and perioperative databases included demographic data, preoperative comorbidities, laboratory results, drugs, intraoperative data, and diagnosis-related features.

As a result, 396 patients who underwent surgery for intestinal obstruction from April 2013 to April 2021 were extracted. After excluding minors, pregnant women, patients who underwent secondary procedures for intestinal obstruction, and patients who lacked sufficient related records, 371 patients from the Third Affiliated Hospital of Sun Yat-sen University were included in the primary cohort for model development. A total of 116 patients who underwent surgery for intestinal obstruction meeting the same inclusion criteria from Shenzhen People's Hospital (50 patients, 2021-2022) and Foshan First People's Hospital (66 patients, 2020-2022) were exclusively extracted for the validation cohorts 1 and 2. We retrieved 853 patients with preoperative intestinal obstruction from the INSPIRE database of 130,000 patients. After excluding patients without key variables required for the model (678 patients) and

XSL•FO
RenderX

nonintestinal obstruction surgical patients (127 patients), the remaining 48 patients were selected as validation cohort 3. Finally, the patients from 3 external validation cohorts jointly formed the external validation cohort 4 (Figure 1).

Figure 1. Flowchart of patient enrollment. ADA: adaptive boosting classifier; GBM: gradient boosting machine; GNB: Gaussian Bayesian classifier; LR: logistic regression; MLP: multilayer perceptron; RF: random forest; SVM: support vector machine.

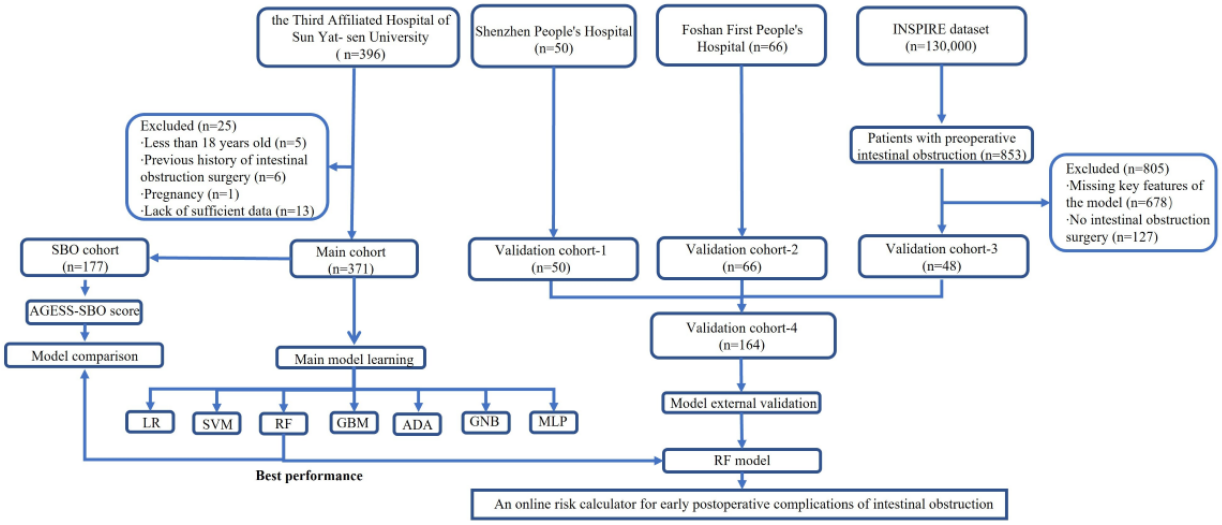

Model Outcome

The primary outcome was to assess early postoperative complications occurring from the first postoperative day until discharge. According to the definition provided by the Association of Surgeons of the Netherlands (ASN), we defined these complications as "a condition or an event that is detrimental to the patient's health, causing irreversible damage or requiring a change in treatment policy." The complications were identified using the *ICD-10* (*International Statistical Classification of Diseases and Related Health Problems 10th Revision*) diagnosis codes and only those labeled as "not present on admission" on the medical record were counted as postoperative complications to ensure that events within the composite outcome were newly developed. We classified these complications in accordance with the Clavien-Dindo [15] (Table S1 in Multimedia Appendix 1) classification and the patients' treatment strategies, to provide more information. Predictions of all levels of complications will be made. Specifically, sepsis was defined as a life-threatening organ dysfunction caused by the host's maladjusted response to infection. The diagnosis of sepsis is based on Sepsis 3.0 diagnostic criteria. Infections or suspected infections include lung infections, intestinal infections, abdominal infections, urinary system infections, and so on, and were manifested as positive pathogen culture or symptoms of infection, but without organ dysfunction. Poor wound healing included wound infection, fat liquefaction, and nonepithelial regeneration of the wound surface accompanied by hematoma, wound dehiscence, and excessive scar hyperplasia.

Predictors, Missing Data, and Selection

Variables with more than 30% missing data and patients with more than 50% missing variables were excluded from further analysis. After doing this, a total of 127 variables were chosen for the univariate analysis, mainly covering demographic characteristics, preoperative comorbidities, laboratory values, cause and complications of intestinal obstruction, as well as intraoperative incidents, shock index (SI), medication, and fluid infusion (Table S2 in Multimedia Appendix 1). Some classification variables were produced by imposing specific rules based on their definitions. According to the statistical analysis of data correlation, only features that had statistical significance ($P<.05$) in the univariate test were selected to minimize the potential overfitting caused by the high dimensions of features. A total of 71 variables were chosen, and then the missing values of these 71 variables were filled. The mode-filling method was used for categorical indicators, and the K-nearest neighbor algorithm was used for continuous indicators. After conducting univariate analysis, we next performed the least absolute shrinkage and selection operator (LASSO) regression analysis on the residual variables. Subsequently, we eliminated variables with LASSO coefficients below 0.02, resulting in the selection of 14 variables. On this basis, we used the wrapper method based on the model score to screen features again. We trained a support vector machine model with 14 variables to calculate their Shapley Additive Explanations (SHAP) values using the training set. We then sorted the variables in descending order according to their SHAP values. These variables were then successively added to the support vector machine model for retraining while monitoring the area under the receiver operating characteristic curve (AUROC) of the model on the validation set. Ultimately, upon selecting the initial 8 indicators, the model's AUROC on the training set reached its maximum (Figure 2).

XSL•FO
RenderX

Figure 2. The process of using a wrapper feature selection method to screen features. Upon the initial selection of the first 8 features, the model achieved its peak area under the receiver operating characteristic (AUROC) curve on the validation set. Subsequent addition of features did not result in a further increase of the AUROC. Consequently, we opted for the top 8 features with the highest Shapley Additive Explanations value.

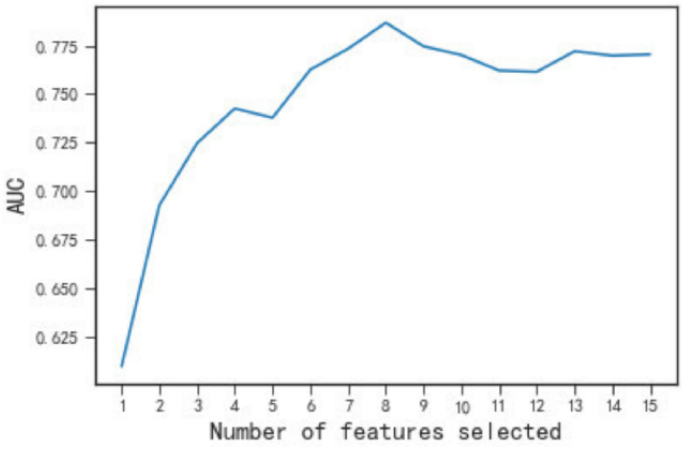

Statistical Analysis

Python (version 3.9.12) was used as the base platform. Table processing was conducted using Pandas (version 1.3.5). The SciPy package (version 1.7.1) was used to analyze the data. The Sklearn software package (version 1.1.2) was used for data preprocessing and building the base models, including logistic regression, support vector machine, random forest (RF), gradient boosting machine implemented by a decision tree, adaptive boosting, the Gaussian Bayesian classifier, and multilayer perceptron.

The main cohort was randomly divided into a 70% development set and a 30% internal validation set. The bootstrap method was used to perform 1000 runs on the internal validation set to calculate confidence intervals for AUROC, accuracy, sensitivity, and specificity. Continuous variables were analyzed and expressed as median values with interquartile intervals and compared using the independent-sample *t* test or Mann-Whitney *U* test. Categorical variables are expressed in terms of quantity and percentage and were compared using a chi-square test. The SHAP method was implemented using the Python SHAP package.

Model Validation

The best-performing model was compared with the AGESS-SBO score, which is calculated by the American Association for the Surgery of Trauma (AAST) anatomical score, physiology score, and comorbidity score as the basis for risk stratification [7]. The ML model was also compared and validated on the data from two other independent cohorts, the INSPIRE dataset, and a mixed cohort to demonstrate its extrapolation and generalization. The prediction performances were compared in terms of AUROC, specificity, sensitivity, accuracy, and F_1-score.

Ethical Considerations

The study protocol followed the principles of the Declaration of Helsinki and was approved by the Institutional Ethics Committee of the Third Affiliated Hospital of Sun Yat-sen University on October 19, 2022 ([2022] 02-004-02). Study data were anonymous or deidentified. The requirements for informed consent and clinical trial registration were waived by the committee. No identification of individual participants in any images of the manuscript or supplementary material was possible.

Results

Univariate Analysis

Among the 396 patients undergoing intestinal obstruction surgery accessed from the EHRs, only 371 patients who met the inclusion criteria were included in the development cohort. The development cohort consisted of a majority of males (242/371, 65.22%), with a median age of 61 (IQR 47-70) years and a median BMI of 20.96 (IQR 19.03-23.81) kg/m^2. The flowchart of the study design is shown in Figure 1. Among the 371 patients included, 176 (47.44%) were diagnosed with early postoperative complications (complication group). The first 3 common postoperative complications were infection (26.95%), sepsis (16.17%), and poor wound healing (7.82%, Table 1). According to Clavien-Dindo's classification of complications [15], grade I had the highest incidence (15.63%, Table 1).

XSL•FO
RenderX

Table 1. The incidence of the 12 most severe and common postoperative complications in the development cohort and the classification of patients in the development cohort based on the Clavien-Dindo classification.

Postoperative complications[a]	Values (N=371), n (%)
Death	16 (4.31)
Sepsis	60 (16.17)
Infection or suspected infection	100 (26.95)
Heart failure	17 (4.58)
Respiratory failure	26 (7.01)
Pulmonary embolism	2 (0.54)
Acute hepatic insufficiency	12 (3.23)
Acute kidney injury	10 (2.7)
Gastrointestinal bleeding	21 (5.66)
Recurrent intestinal obstruction	20 (5.39)
Anastomotic fistula	8 (2.16)
Poor wound healing	29 (7.82)
Clavien-Dindo classification	
Grade I	58 (15.63)
Grade II	55 (14.82)
Grade IIIa	8 (2.16)
Grade IIIb	12 (3.23)
Grade Iva	13 (3.5)
Grade IVb	14 (3.77)
Grade V	16 (4.31)

[a]Some severe and common complications are listed here.

A total of 127 perioperative variables were analyzed by univariate analysis. Finally, only 71 variables with significant differences (P<.05) between the two groups were retained, including 37 preoperative variables and 34 intraoperative variables (Table S2 in Multimedia Appendix 1).

Compared with the patients without early postoperative complications, the time of first postoperative defecation, time of drainage tube retention, total hospitalization cost, total hospitalization days, postoperative hospitalization days, and ICU (intensive care unit) length of stay in the complication group was increased significantly (P<.05, Table S2 in Multimedia Appendix 1).

Internal Validation Performance

After the univariate analysis, the LASSO regression analysis, and the wrapper method based on SHAP, we selected 8 features, including the Physiological Severity Score for the Enumeration of Mortality and Morbidity (POSSUM physiological score), the amount of colloid infusion, SI before anesthesia induction, the ASA classification, the percentage of neutrophils, SI at the end of the surgery, age, and total protein to train our 7 models. To increase the robustness of the model prediction results during the training process, we used the integration idea for modeling [16] and finally established 7 common ML models. In the training set, the RF model achieved a relatively balanced AUROC (0.788, 95% CI 0.709-0.869) and specificity (0.810, 95% CI 0.715-0.884) and the highest sensitivity (0.695, 95% CI 0.570-0.811), accuracy (0.756, 95% CI 0.696-0.831) and F_1-score (0.727, 95% CI 0.633-0.812). Since GNB reached the greatest AUROC (0.805, 95% CI 0.730-0.878), its sensitivity was lowest (0.544, 95% CI 0.433-0.674). We eventually chose the RF model for further analysis and application.

Since the AGESS-SBO score is a scoring system for postoperative complications of SBO, we excluded patients with colorectal obstruction. Then, we validated and compared the performance of this score and ML predictor in a set excluding patients with colorectal obstruction. The AGESS-SBO score presented a higher specificity (0.889, 95% CI 0.750-1.0) with a lower AUROC (0.731, 95% CI 0.585-0.854), accuracy (0.653, 95% CI 0.531-0.776), F_1-score (0.537, 95% CI 0.320-0.718), and sensitivity (0.409, 95% CI 0.222-0.625). For predicting early complications of SBO, the RF model achieved the highest AUROC (0.912, 95% CI 0.809-0.986), accuracy (0.831, 95% CI 0.735-0.918), sensitivity (0.893, 95% CI 0.725-1.0), F_1-score (0.817, 95% CI 0.684-0.929), and balanced specificity (0.786, 95% CI 0.619-0.93; Table 2, Figures 3A and 3B).

XSL•FO
RenderX

Table 2. The performance of machine learning models in training sets and external validation cohorts.

Algorithm	AUROC[a] (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F_1-score (95% CI)
LR[b]	0.797 (0.693-0.881)	0.733 (0.661-0.812)	0.695 (0.565-0.814)	0.769 (0.637-0.877)	0.713 (0.609-0.807)
SVM[c]	0.780 (0.690-0.854)	0.714 (0.638-0.781)	0.615 (0.504-0.731)	0.803 (0.680-0.889)	0.666 (0.570-0.756)
GNB[d]	0.805 (0.730-0.878)	0.698 (0.625-0.777)	0.544 (0.433-0.674)	0.833 (0.743-0.921)	0.625 (0.522-0.723)
RF[e]	0.788 (0.709-0.869)	0.756 (0.696-0.831)	0.695 (0.570-0.811)	0.810 (0.715-0.884)	0.727 (0.633-0.812)
GBM[f]	0.747 (0.667-0.836)	0.683 (0.598-0.777)	0.671 (0.570-0.794)	0.695 (0.568-0.831)	0.666 (0.567-0.774)
ADA[g]	0.668 (0.562-0.745)	0.665 (0.589-0.737)	0.688 (0.571-0.802)	0.644 (0.532-0.759)	0.655 (0.545-0.740)
MLP[h]	0.790 (0.708-0.862)	0.692 (0.598-0.759)	0.611 (0.454-0.724)	0.765 (0.652-0.870)	0.649 (0.520-0.733)
Validation cohort 1	0.755 (0.652-0.839)	0.729 (0.644-0.805)	0.621 (0.444-0.774)	0.772 (0.686-0.859)	0.568 (0.417-0.684)
Validation cohort 2	0.817 (0.695-0.913)	0.742 (0.636-0.848)	0.703 (0.553-0.845)	0.800 (0.625-0.936)	0.754 (0.620-0.853)
Validation cohort 3	0.786 (0.628- 0.902)	0.708 (0.562- 0.833)	0.700 (0.500- 0.870)	0.724 (0.522- 0.875)	0.700(0.529- 0.826)
Validation cohort 4	0.720 (0.671- 0.768)	0.658 (0.602- 0.702)	0.634 (0.566- 0.695)	0.679 (0.617- 0.730)	0.628 (0.575- 0.682)
RF-SBO[i]	0.912 (0.809-0.986)	0.831 (0.735-0.918)	0.893 (0.725-1.0)	0.786 (0.619-0.93)	0.817 (0.684-0.929)
AGESS-SBO[j]	0.731 (0.585-0.854)	0.653 (0.531-0.776)	0.409 (0.222-0.625)	0.889 (0.750-1.0)	0.537 (0.320-0.718)

[a]AUROC: area under the receiver operating characteristic curve.

[b]LR: logistic regression.

[c]SVM: support vector machine.

[d]GNB: Gaussian Bayesian classifier.

[e]RF: random forest.

[f]GBM: gradient boosting machine.

[g]ADA: adaptive boosting.

[h]MLP: multilayer perceptron.

[i]RF-SBO: random forest–small-bowel obstruction.

[j]AGESS-SBO: Acute General Emergency Surgical Severity-Small Bowel Obstruction.

Figure 3. Use of the receiver operating characteristic curve plots to demonstrate the performance of the machine learning model in different cohorts. ADA: adaptive boosting; GBM: gradient boosting machine; GNB: Gaussian Bayesian classifier; LR: logistic regression; MLP: multilayer perceptron; RF: random forest; SVM: support vector machine.

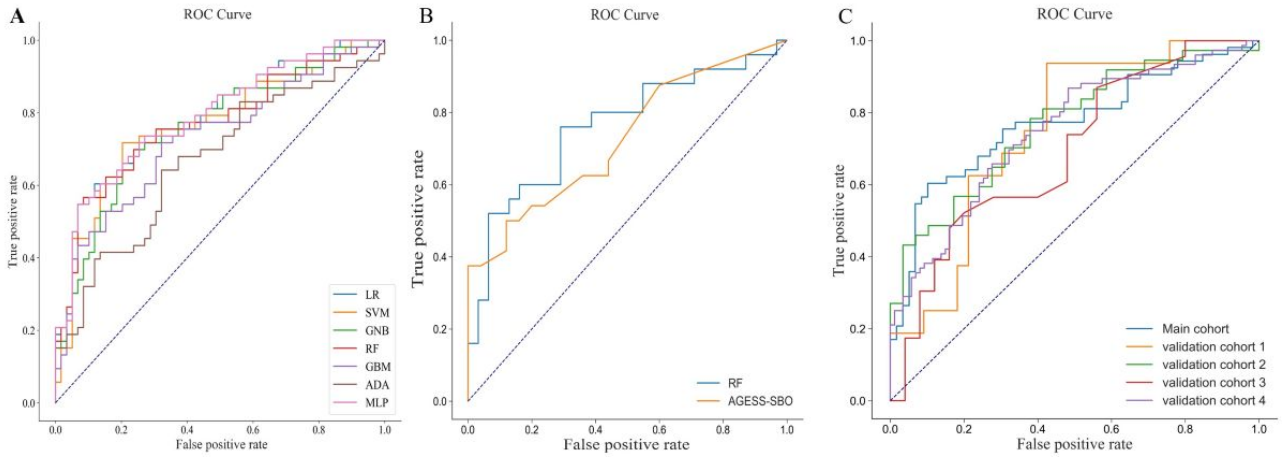

In Figure 3A, the receiver operating characteristic (ROC) curve of all machine learning models in the development set is shown. In Figure 4B, the ROC curve of the random forest model and AGESS-SBO score in the SBO cases of the development set is shown. In Figure 4C, the ROC curve of the random forest model in the validation sets is shown.

XSL•FO
RenderX

Figure 4. Use of Shapley Additive Explanations (SHAP) summary plot and SHAP dependence plot to demonstrate the importance and impact of variables on the random forest model. ASA: American Society of Anesthesiologists; SHAP: Shapley Additive Explanations; NEUT%: neutrophil percentage; POSSUM: Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity.

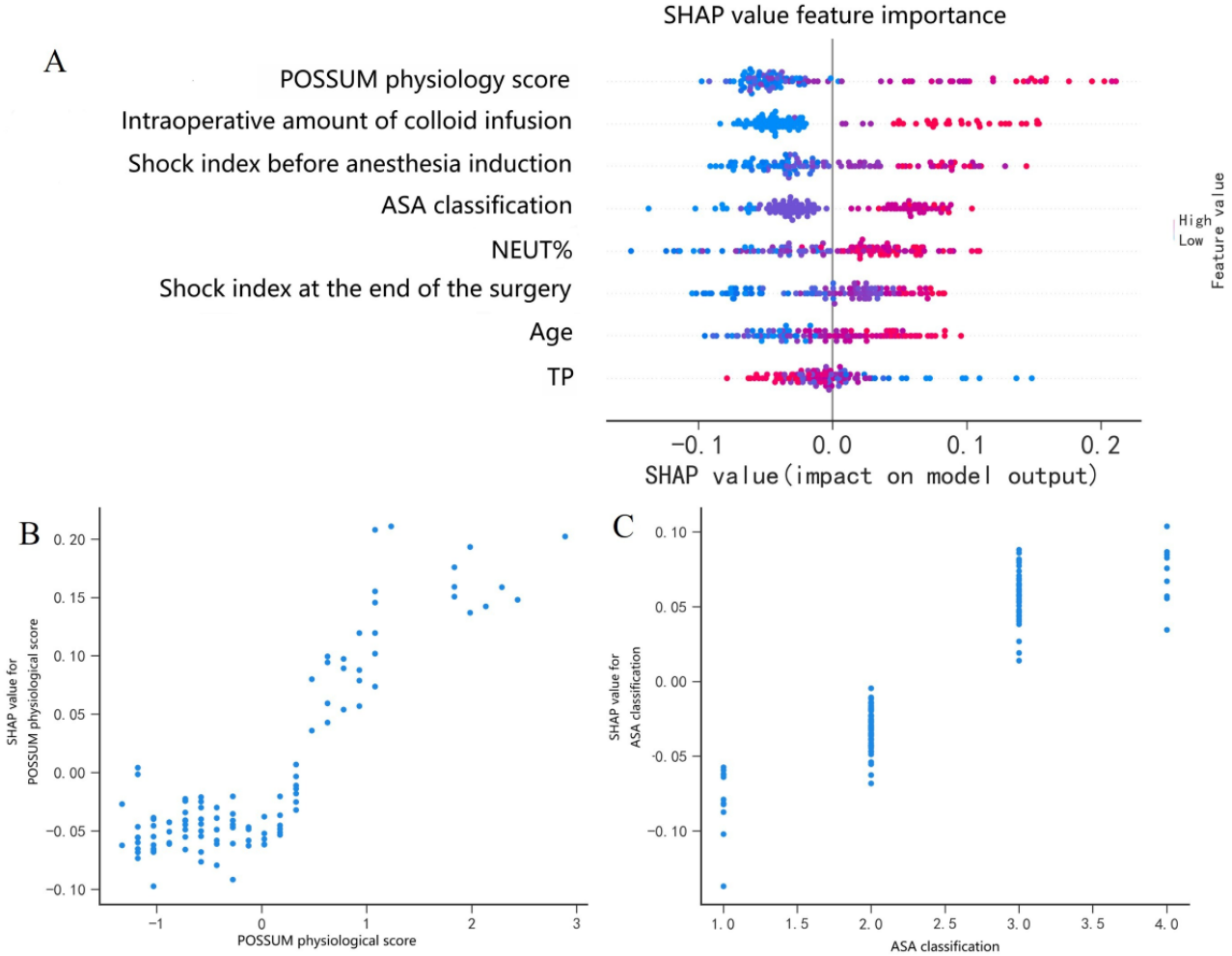

External Validation Performance

The external validation cohorts included 50 patients from Shenzhen People's Hospital (validation cohort 1), 66 patients from Foshan First People's Hospital (validation cohort 2), 48 patients from the INSPIRE dataset (validation cohort 3), and 164 patients from the above 3 datasets (validation cohort 4). In these 4 external validation sets, the incidence of early postoperative complications was 34%, 56.06%, 52.08%, and 48.17%, respectively. In the four validation sets, the RF model achieved a comparable AUROC (0.755, 95% CI 0.652-0.839), a greater AUROC (0.817, 95% CI 0.695-0.913), a similar AUROC (0.786, 95% CI 0.628-0.902) and a comparable AUROC (0.720, 95% CI 0.671-0.768; Table 2 and Figure 3C).

Feature Importance Evaluated by SHAP Values

The baseline for the SHAP value in this study is the average of all predicted early complication incidences in the internal validation set, which was 49.96%. Figure 4A shows the explanation of the ML model by SHAP. The SHAP summary plot demonstrated that the POSSUM physiological score, the amount of colloid infusion, SI before anesthesia induction, the ASA classification, the percentage of neutrophils, SI at the end of the surgery, age, and total protein were ranked as the top 8 important variables for RF. Both kinds of SHAP plots show that a higher POSSUM physiological score, larger amount of colloid infusion, larger SI before anesthesia induction, higher ASA grade, higher percentage of neutrophils, larger SI at the end of the surgery, older age, and lower total protein were associated with a higher SHAP value output in the RF model, indicating higher odds of early postoperative complications after surgery for intestinal obstruction.

We use a SHAP dependence plot to observe how a feature affects the prediction results of the model (Figure S1 in Multimedia Appendix 1; Figures 4B and 4C). For example, when the POSSUM physiological score is high, the SHAP value will change significantly, showing a positive correlation (Figure 4B). From the SHAP dependence plot, it can also be found that there is an obvious "truncation" in the SHAP value between ASA II and ASA III, indicating that ASA II and ASA III have a greater contribution to distinguishing the positive and negative samples (Figure 4C).

An example of correctly classified cases and an example of incorrectly classified cases were demonstrated as a SHAP decision plot (Figures 5A and 5B) and a force plot (Figures 5C and 5D). These plots increase the interpretability and transparency of the predictions made by RF algorithms. The SHAP decision plots showed how the model makes decisions

XSL•FO
RenderX

based on the availability of each feature in the electronic medical record and provided a decision path for each feature. The force plot mainly shows the major factors that contribute to the final model output of a specific individual.

Figure 5. Use the SHAP decision plots and SHAP force plots to demonstrate the decision pathways and feature importance of the RF model in specific patients.

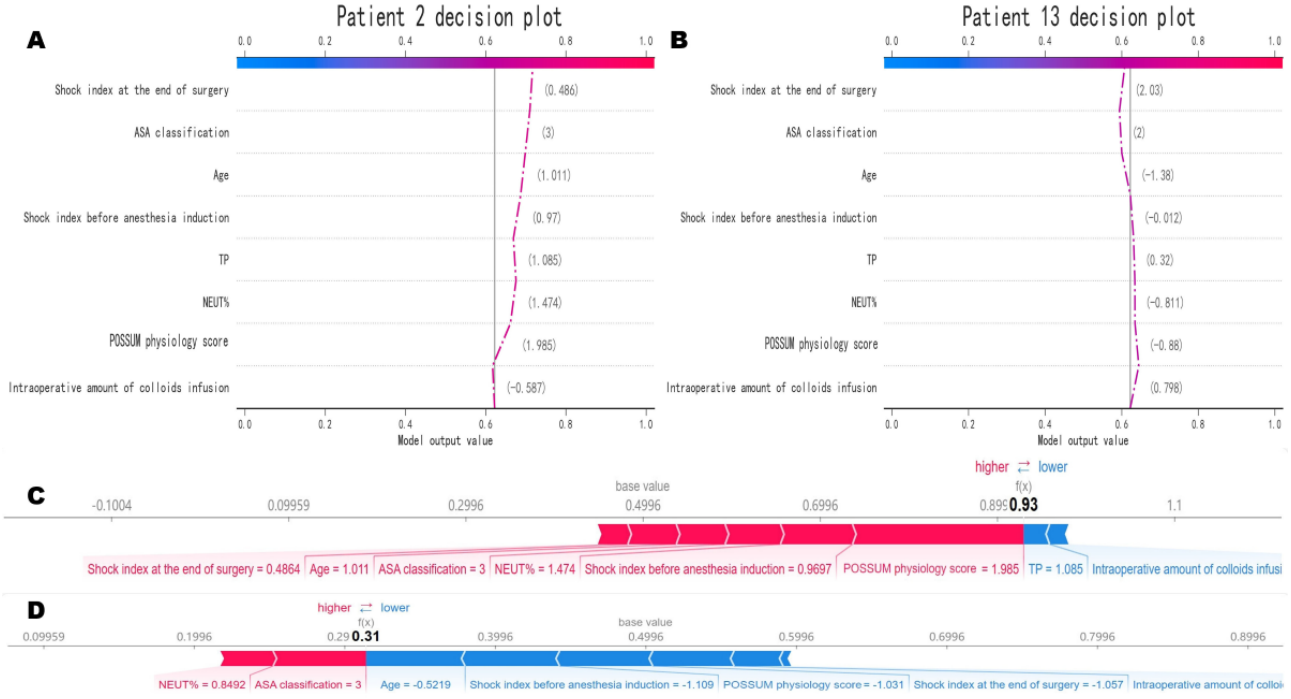

In Figure 4A, the SHAP summary plot illustrates the overall impact distribution of each feature on the RF model. The features are sorted on the basis of the average absolute value of their SHAP values, with the importance of the features decreasing from the top to the bottom. The color scheme can aid in understanding how changes in feature values affect the results. Red indicates high eigenvalues, while blue signifies low eigenvalues. The greater the distance of a point from the zero reference SHAP value, the more significant its impact on the output. In Figure 4B, the SHAP dependence plot illustrates the SHAP values for the POSSUM Physiology Score across all samples, with each point representing an individual sample. From the graph, it can be preliminarily deduced that the POSSUM Physiology Score is positively correlated with the SHAP value. That is, under stable conditions for other features, an increase in the POSSUM Physiology Score corresponds to a larger SHAP value and a higher final prediction probability. In Figure 4C, the SHAP dependence plot illustrates the SHAP values for the ASA classification across all samples. The ASA classification demonstrates a positive correlation with the SHAP value. Specifically, under stable conditions for other features, an increase in the ASA classification corresponds to a larger SHAP value and, accordingly, a higher final prediction probability.

In Figure 5A, the SHAP decision plot of a correctly classified case. At the bottom of the SHAP decision plot lies the baseline value. From the bottom to the top, the contribution of each feature is visualized, with the connecting line denoting the predictive process from the baseline to the ultimate outcome. In Figure 5B, SHAP decision plot of an incorrectly classified case. In Figure 5C, the SHAP force plot of a correctly classified case. The SHAP value denotes the predictive associated features of an individual patient and the respective contributions of each feature to outcome prediction. The prominent numbers represent the probability-predicted values, while the underlying values are the predicted outcomes that have not been incorporated into the model. The red element (on the left) signifies the factor that augments risk, whereas the blue element (on the right) indicates the factor that mitigates risk. The length of the arrow reflects the extent of influence on the prediction, i.e., the longer the arrow, the more pronounced the effect. In Figure 5D, a SHAP force plot of an incorrectly classified case is given.

Model Visualization

We visualized the RF model and created a web-based online risk calculator [17]. The Python code of this online risk calculator is mentioned in Multimedia Appendix 2. By inputting the variables required for the RF model, the incidence of early postoperative complications for a certain patient can be predicted. As shown in Figure 6, after inputting the patient's age of 58 years, SI of 0.75 at the end of surgery, intraoperative colloid infusion volume of 1000 mL, POSSUM physiological score of 22, neutrophil percentage of 0.656, preoperative total protein of 68.10 g/L, SI of 0.68 before anesthesia induction, and ASA classification III, the probability of early postoperative complications for this patient was found to be 86%.

XSL•FO
RenderX

Figure 6. Calculation tool of random forest model to estimate early complications after intestinal obstruction surgery. After inputting 8 variables into the random forest model, the prediction model output for the case was “positive” with a probability of 86%.

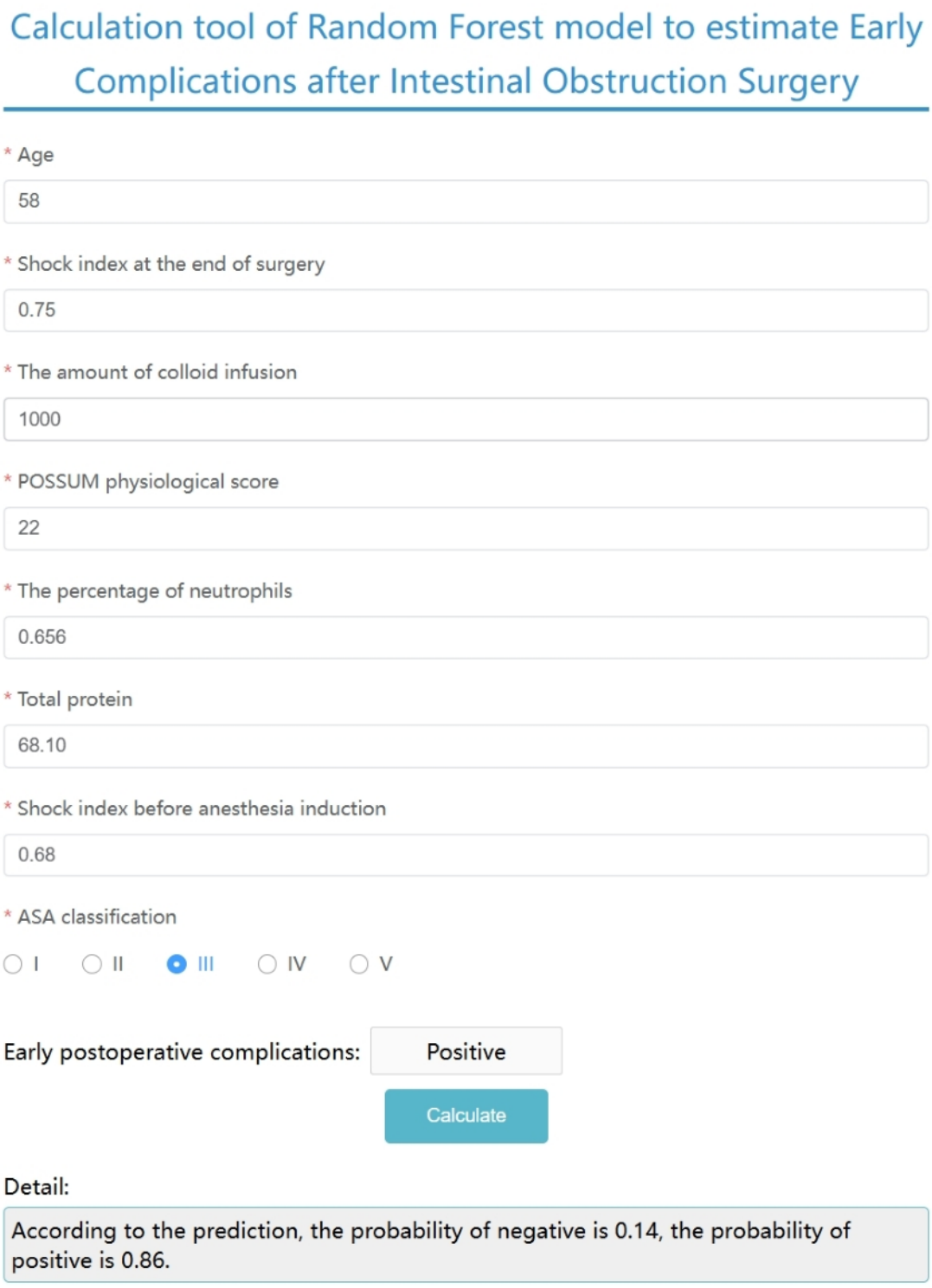

Discussion

Principal Findings

Precisely forecasting the occurrence of postoperative complications in intestinal obstruction is of great significance for enhancing perioperative decision-making. The purpose of this study is to assess the risk of complications after intestinal obstruction surgery and provide an in-depth interpretation of the model’s decisions and their impact on variables using ML algorithms. Finally, we transformed the optimal model into a web-based risk calculator to assist clinicians in identifying high-risk patients. To the best of our knowledge, this is the first study to predict postoperative complications of intestinal obstruction surgery based on the ML algorithm. The findings show that compared with other testing algorithms, RF outperforms the strongest performance, with AUROC,

XSL•FO
RenderX

specificity, sensitivity, and accuracy of 0.788, 0.810, 0.695, and 0.756, respectively. Compared with the AGESS-SBO score, the RF model achieved a better AUROC of 0.912 in predicting postoperative complications of patients with SBO. This attributed to, on the one hand, our study cohort comprised solely of patients undergoing surgical intervention; on the other hand, the ML model incorporated some intraoperative variables, making it more suitable for patients undergoing surgical treatment.

Intestinal obstruction is frequently accompanied by symptoms such as fever, disruption of the internal environment, infection, poisoning, and shock [18], which can further exacerbate the condition in the context of surgery and anesthesia. Consequently, these patients often experience various complications postoperatively. However, there is a scarcity of reliable statistical data available to demonstrate the severity and harmfulness of these complications. Therefore, we first performed a comprehensive statistical analysis and classification. First, nearly half of the patients with intestinal obstruction had early postoperative complications, causing increased hospitalization costs, prolonged hospital stays, a huge medical burden, reduced functional abilities, and even death. The most common complication was an infection because in this kind of patient, it is easy to lose the barrier effect of the intestinal mucosa and wall on bacteria, leading to bacterial translocation [19]. Second, we used the Clavien-Dindo classification to demonstrate the distribution of different degrees of complications. To improve postoperative quality of life, the emergence of disease prediction models may provide a solution for the prevention of early postoperative complications.

In the real world, clinically generated data depend on hospitals and are diversely distributed. Therefore, it is difficult to develop a robust prediction model suitable for multiple institutions or even regions. However, this study showed that with a large amount of data available, a more robust model can be developed with only a few variables. Therefore, we selected many perioperative features to find new associations in the complex relationship between a large data volume and outcome. However, the substantial number of variables within a small cohort with limited events escalates the risk of errors, including overfitting, bias or variance, and data imbalance. To minimize these errors, we have adopted the following methods: (1) performed LASSO regression analysis for variable selection to decrease the number of variables processed by the model, thereby reducing feature dimensions and mitigating the risk of overfitting; (2) preprocessed the data, including filling in missing values and standardization, to ensure effective learning by the model; (3) used cross-validation techniques to assess model performance, which aids in detecting bias or variance issues and provides insights for model improvement; (4) used data from various regions and hospitals across different time periods for external validation to confirm its applicability across different cohorts and test for overfitting; (5) used ensemble learning techniques (RF) to reduce the risk of overfitting; (6) decreased the number of samples with majority negative outcomes to amplify the signal of minority positive outcomes, thereby balancing outcome distribution and enhancing model performance; and (7) selected multiple appropriate evaluation indicators to comprehensively evaluate its performance, such as AUROC, sensitivity, accuracy, F_1-score value, and so on.

Understanding how the RF model prioritizes specific features for prediction is essential to build trust among clinicians and integrate this tool into routine practice. RF is an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of predictions. Each tree in the forest is built from a random sample of the data with replacement, known as bootstrap sampling. At each node of the tree, only a random subset of features is considered for splitting, which helps prevent overfitting and increases diversity among the trees. The final prediction is made by aggregating the results of all trees. Moreover, RF provides a measure of feature importance by evaluating how much each feature contributes to decreasing impurity across all trees. Features causing significant decreases in impurity are considered more important.

The optimization of some potential intervenable variables is highly important in preventing early postoperative complications, such as SI. SI has more advantages than other vital signs in evaluating systemic perfusion, the timing of vasopressin selection, and prognosis [20,21]. Several studies [22,23] found that the incidence of hypotension in patients with a high SI increased after airway establishment. A larger amount of colloid infusion was associated with a higher incidence of early postoperative complications. On the one hand, patients with intestinal obstruction have a high probability of sepsis. Studies have confirmed that colloidal solution is not beneficial to sepsis and even increases the risk of acute renal injury [24]. On the other hand, severely insufficient blood volume during surgery often requires the infusion of more colloids. Several studies have also shown that ASA classification [3,25], age [26], and POSSUM physiological score [6,27] are independent risk factors for early postoperative complications. They have been widely applied to evaluate surgical tolerance and risk stratification. These findings are also supported by the results of this study. The percentage of neutrophils and total protein were found in a previous study to predict the occurrence of intestinal obstruction [28]. Our results indicate that they are sensitive biomarkers for predicting early postoperative complications of intestinal obstruction. Hematological parameters and markers have been widely used to predict the risk of postoperative complications [29,30].

The complex and diverse perioperative pathophysiology of patients with intestinal obstruction poses significant challenges for clinicians. Our constructed online risk calculator based on the RF model aims to facilitate timely decision-making and intervention in disease progression. However, the AUROC of the ML model seldom surpasses 0.8. This might be attributed to the fact that postoperative complications represent a mixed and complex diagnosis, with different complications potentially having specific predictive factors. Consequently, achieving higher performance indicators may be infeasible. Nonetheless, external validation results indicating a higher AUROC demonstrate the model's generalization capability. The inclusion of higher-dimensional information, such as image or video data, in the prediction of complications may potentially enhance the predictive abilities of the model.

XSL•FO
RenderX

In clinical practice, the acceptability and applicability of a scoring system or model are determined not only by its accuracy but also by its simplicity and universality. The online risk calculator, constructed using ML algorithms, possesses these characteristics. First, this online risk calculator incorporates only 8 features, which do not augment the workload or memory burden for clinicians. Second, the calculator's result requirements can be fulfilled during the routine clinical diagnosis and treatment process, obviating the need for additional tests or examinations and thereby avoiding any additional burden on patients. Finally, our findings indicate that the online risk calculator based on the RF model exhibits generalizability and demonstrates strong predictive capabilities across diverse regions, hospitals, and time periods. Facilitating clinicians' decision-making processes and enhancing patient prognoses at nearly zero cost is an essential value of a scoring system. This online risk calculator can help anesthesiologists predict patients' prognoses through partial preoperative and intraoperative information and increase clinical vigilance to make early interventions.

There are still several limitations of our study. First, it is a small sample study, and the predictive model requires a larger sample for further development and verification. Second, the current study is a retrospective study, in which there may be some confounding factors. In addition, the demographic diversity and clinical settings of the development set and validation set were not compared in this study. Future research needs to further compare more characteristics of the development and validation sets to clarify the "Reproducibility" or "Transportability" of the model. Furthermore, the model cannot be used for risk prediction of specific complications. In the future, we will construct a series of submodels for predicting the risk of specific postoperative complications in patients with intestinal obstruction. Finally, our research focused on early complications that may prolong the length of hospitalization and rehabilitation of patients after hospitalization, so there was no survival analysis. We will further explore them in the future.

Conclusions

We constructed an online risk calculator based on the RF model, which includes 8 variables, to assist clinicians in identifying high-risk patients after intestinal obstruction surgery.

Acknowledgments

We would like to thank the research participants, engineers, and other participating centers for their unreserved help. Finally, we thank those who support us directly or indirectly. This study was supported partly by the Natural Science Foundation of Guangdong Province (2021A1515011827), the Special Support Project of Guangdong Province (0720240209), Science and Technology Program of Guangzhou City (2024A04J4246), Joint Funds of the National Natural Science Foundation of China (U22A20276), Science and Technology Planning Project of Guangdong Province-Regional Innovation Capacity and Support System Construction (2023B110006), and the "Five and five" Project of the Third Affiliated Hospital of Sun Yat-Sen University (2023WW501).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

PJH contributed to conceptualization, data curation, methodology, project administration, and writing–review and editing. JRY assisted with conceptualization, data curation, investigation, formal analysis, and writing–original draft. DZZ handled data curation, investigation, validation, and writing–review and editing. TJR performed investigation and writing–review and editing. YHL contributed to formal analysis, software, and visualization. DY assisted with formal analysis, software, and visualization. XQZ handled data curation, project administration, resources, supervision, and writing–review and editing. SLZ contributed to conceptualization, project administration, resources, supervision, and writing–review and editing. CJC assisted with conceptualization, project administration, funding acquisition, resources, supervision, and writing–review and editing.

CC (chenchj28@mail.sysu.edu.cn), SZ (13610272308@139.com), and XZ (1049550363@qq.com) are co-correspondence authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional online content.
[DOCX File , 557 KB - jmir_v27i1e68354_app1.docx]

Multimedia Appendix 2
The Python code used in the online risk calculator.
[ZIP File (Zip Archive), 2 KB - jmir_v27i1e68354_app2.zip]

XSL•FO
RenderX

References

1. Veen EJ, Janssen-Heijnen MLG, Leenen LPH, Roukema JA. The registration of complications in surgery: a learning curve. World J Surg 2005;29(3):402-409. [doi: 10.1007/s00268-004-7358-8] [Medline: 15696399]
2. Amelung FJ, de Beaufort HWL, Siersema PD, Verheijen PM, Consten ECJ. Emergency resection versus bridge to surgery with stenting in patients with acute right-sided colonic obstruction: a systematic review focusing on mortality and morbidity rates. Int J Colorectal Dis 2015;30(9):1147-1155. [doi: 10.1007/s00384-015-2216-8] [Medline: 25935448]
3. Jeppesen MH, Tolstrup MB, Kehlet Watt S, Gögenur I. Risk factors affecting morbidity and mortality following emergency laparotomy for small bowel obstruction: a retrospective cohort study. Int J Surg 2016;28:63-68 [FREE Full text] [doi: 10.1016/j.ijsu.2016.02.059] [Medline: 26912017]
4. Quah GS, Eslick GD, Cox MR. Laparoscopic versus open surgery for adhesional small bowel obstruction: a systematic review and meta-analysis of case-control studies. Surg Endosc 2019;33(10):3209-3217. [doi: 10.1007/s00464-018-6604-3] [Medline: 30460502]
5. Lee MJ, Sayers AE, Drake TM, Marriott PJ, Anderson ID, Bach SP, NASBO steering group and NASBO collaborators. National prospective cohort study of the burden of acute small bowel obstruction. BJS Open 2019;3(3):354-366 [FREE Full text] [doi: 10.1002/bjs5.50136] [Medline: 31183452]
6. Peacock O, Bassett MG, Kuryba A, Walker K, Davies E, Anderson I, National Emergency Laparotomy Audit (NELA) Project Team. Thirty-day mortality in patients undergoing laparotomy for small bowel obstruction. Br J Surg 2018;105(8):1006-1013. [doi: 10.1002/bjs.10812] [Medline: 29603126]
7. Baghdadi YMK, Morris DS, Choudhry AJ, Thiels CA, Khasawneh MA, Polites SF, et al. Validation of the anatomic severity score developed by the American association for the surgery of trauma in small bowel obstruction. J Surg Res 2016;204(2):428-434. [doi: 10.1016/j.jss.2016.04.076] [Medline: 27565079]
8. Tran J, Sharma D, Gotlieb N, Xu W, Bhat M. Application of machine learning in liver transplantation: a review. Hepatol Int 2022;16(3):495-508. [doi: 10.1007/s12072-021-10291-7] [Medline: 35020154]
9. Connor CW. Artificial intelligence and machine learning in anesthesiology. Anesthesiology 2019;131(6):1346-1359 [FREE Full text] [doi: 10.1097/ALN.0000000000002694] [Medline: 30973516]
10. Quesada JA, Lopez-Pineda A, Gil-Guillén VF, Durazo-Arvizu R, Orozco-Beltrán D, López-Domenech A, et al. Machine learning to predict cardiovascular risk. Int J Clin Pract 2019;73(10):e13389. [doi: 10.1111/ijcp.13389] [Medline: 31264310]
11. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018;1:39 [FREE Full text] [doi: 10.1038/s41746-018-0040-6] [Medline: 31304320]
12. Goyal R, Mui LW, Riyahi S, Prince MR, Lee HK. Machine learning based prediction model for closed-loop small bowel obstruction using computed tomography and clinical findings. J Comput Assist Tomogr 2022;46(2):169-174. [doi: 10.1097/RCT.0000000000001269] [Medline: 35297573]
13. Lim L, Lee H, Jung CW, Sim D, Borrat X, Pollard TJ, et al. INSPIRE, a publicly available research dataset for perioperative medicine. Sci Data 2024;11(1):655 [FREE Full text] [doi: 10.1038/s41597-024-03517-4] [Medline: 38906912]
14. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):E215-E220. [doi: 10.1161/01.cir.101.23.e215] [Medline: 10851218]
15. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg 2004;240(2):205-213. [doi: 10.1097/01.sla.0000133083.54934.ae] [Medline: 15273542]
16. Liu XY, Wu J, Zhou ZH. Exploratory undersampling for class-imbalance learning. IEEE Trans Syst Man Cybern B Cybern 2009;39(2):539-550. [doi: 10.1109/TSMCB.2008.2007853] [Medline: 19095540]
17. Pin-Jie H, Ji-Rong Y, Di-Zhou Z, Tao-Jia R, Yu-Heng L, Dong Y, et al. Calculation tool of RF model to estimate early complications after intestinal obstruction surgery. 2024. URL: https://analysis.aidcloud.cn/cn/sunyatsen/?p=ClavienDindo [accessed 2025-02-20]
18. Jackson P, Vigiola Cruz M. Intestinal obstruction: evaluation and management. Am Fam Physician 2018;98(6):362-367 [FREE Full text] [Medline: 30215917]
19. Sagar PM, MacFie J, Sedman P, May J, Mancey-Jones B, Johnstone D. Intestinal obstruction promotes gut translocation of bacteria. Dis Colon Rectum 1995;38(6):640-644. [doi: 10.1007/BF02054126] [Medline: 7774478]
20. Gupta S, Alam A. Shock index is better than conventional vital signs for assessing higher level of care and mortality in severe sepsis or shock. Am J Emerg Med 2021;46:545-549. [doi: 10.1016/j.ajem.2020.11.014] [Medline: 33234359]
21. Rasmy I, Mohamed H, Nabil N, Abdalah S, Hasanin A, Eladawy A, et al. Evaluation of perfusion index as a predictor of vasopressor requirement in patients with severe sepsis. Shock 2015;44(6):554-559. [doi: 10.1097/SHK.0000000000000481] [Medline: 26529657]
22. Lee K, Jang JS, Kim J, Suh YJ. Age shock index, shock index, and modified shock index for predicting postintubation hypotension in the emergency department. Am J Emerg Med 2020;38(5):911-915. [doi: 10.1016/j.ajem.2019.07.011] [Medline: 31345593]

XSL•FO
RenderX

23. Trivedi S, Demirci O, Arteaga G, Kashyap R, Smischney NJ. Evaluation of preintubation shock index and modified shock index as predictors of postintubation hypotension and other short-term outcomes. J Crit Care 2015;30(4):861.e1-861867. [doi: 10.1016/j.jcrc.2015.04.013] [Medline: 25959037]
24. Joannes-Boyau O, Le Conte P, Bonnet MP, Cesareo E, Chousterman B, Chaiba D, et al. Guidelines for the choice of intravenous fluids for vascular filling in critically ill patients, 2021. Anaesth Crit Care Pain Med 2022;41(3):101058 [FREE Full text] [doi: 10.1016/j.accpm.2022.101058] [Medline: 35526312]
25. Margenthaler JA, Longo WE, Virgo KS, Johnson FE, Grossmann EM, Schifftner TL, et al. Risk factors for adverse outcomes following surgery for small bowel obstruction. Ann Surg 2006;243(4):456-464. [doi: 10.1097/01.sla.0000205668.58519.76] [Medline: 16552195]
26. Tabola R, Mantese G, Cirocchi R, Gemini A, Grassi V, Boselli C, et al. Postoperative mortality and morbidity in older patients undergoing emergency right hemicolectomy for colon cancer. Aging Clin Exp Res 2017;29(Suppl 1):121-126. [doi: 10.1007/s40520-016-0643-1] [Medline: 27830519]
27. Fukuda N, Wada J, Niki M, Sugiyama Y, Mushiake H. Factors predicting mortality in emergency abdominal surgery in the elderly. World J Emerg Surg 2012;7(1):12 [FREE Full text] [doi: 10.1186/1749-7922-7-12] [Medline: 22578159]
28. Cao Y, Ke S, Gu J, Mao F, Yao S, Deng S, et al. The value of haematological parameters and tumour markers in the prediction of intestinal obstruction in 1474 Chinese colorectal cancer patients. Dis Markers 2020;2020:8860328 [FREE Full text] [doi: 10.1155/2020/8860328] [Medline: 32855747]
29. Demirjian S, Bashour CA, Shaw A, Schold JD, Simon J, Anthony D, et al. Predictive accuracy of a perioperative laboratory test-based prediction model for moderate to severe acute kidney injury after cardiac surgery. JAMA 2022;327(10):956-964 [FREE Full text] [doi: 10.1001/jama.2022.1751] [Medline: 35258532]
30. Yang Q, Li M, Cao X, Lu Y, Tian C, Sun M, et al. An umbrella review of meta-analyses on diagnostic accuracy of C-reactive protein. Int J Surg 2022;104:106788 [FREE Full text] [doi: 10.1016/j.ijsu.2022.106788] [Medline: 35931353]

Abbreviations

AAST: American Association for the Surgery of Trauma
AGESS-SBO: Acute General Emergency Surgical Severity–Small Bowel Obstruction
ASA: American Society of Anesthesiologists
ASN: Association of Surgeons of the Netherlands
AUROC: area under the receiver operating characteristic curve
ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision
INSPIRE: Informative Surgical Patient dataset for Innovative Research Environment
LASSO: least absolute shrinkage and selection operator
ML: machine learning
POSSUM: Physiological Severity Score for the Enumeration of Mortality and Morbidity
RF: random forest
ROC: receiver operating characteristic
SBO: small-bowel obstruction
SHAP: Shapley Additive Explanations
SI: shock index

Edited by A Coristine; submitted 04.11.24; peer-reviewed by B Yi, J Soldera, R Qiu; comments to author 17.01.25; revised version received 27.01.25; accepted 05.02.25; published 03.03.25.

Please cite as:
Huang P, Yang J, Zhao D, Ran T, Luo Y, Yang D, Zheng X, Zhou S, Chen C
Machine Learning–Based Prediction of Early Complications Following Surgery for Intestinal Obstruction: Multicenter Retrospective Study
J Med Internet Res 2025;27:e68354
URL: https://www.jmir.org/2025/1/e68354
doi:10.2196/68354
PMID:40053794

©Pinjie Huang, Jirong Yang, Dizhou Zhao, Taojia Ran, Yuheng Luo, Dong Yang, Xueqin Zheng, Shaoli Zhou, Chaojin Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the

XSL•FO
RenderX

Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning Models With Prognostic Implications for Predicting Gastrointestinal Bleeding After Coronary Artery Bypass Grafting and Guiding Personalized Medicine: Multicenter Cohort Study

Jiale Dong[1,2,3*], MD, PhD; Zhechuan Jin[1*], MD, PhD; Chengxiang Li[4], MS; Jian Yang[1,4], MD, PhD; Yi Jiang[5], MD, PhD; Zeqian Li[4], MD, PhD; Cheng Chen[6], MD, PhD; Bo Zhang[4,7], MD, PhD; Zhaofei Ye[8], MD, PhD; Yang Hu[3], MD, PhD; Jianguo Ma[9], PhD; Ping Li[8], MD, PhD; Yulin Li[1*], MD, PhD; Dongjin Wang[6*], MD, PhD; Zhili Ji[1,2,3,4*], MD, PhD

[1]Beijing Institute of Heart, Lung and Blood Vessel Diseases, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

[2]Department of Acute Abdomen Surgery, Beijing ChaoYang Hospital, Capital Medical University, Beijing, China

[3]Department of General Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

[4]Department of Hepatobiliary and Pancreaticosplenic Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

[5]Department of Cardiovascular Surgery, Nanjing Drum Tower Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Nanjing, China

[6]Department of Cardiovascular Surgery, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Beijing, China

[7]Department of General Surgery, Beijing Luhe Hospital, Capital Medical University, Beijing, China

[8]Department of Cardiovascular Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

[9]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhili Ji, MD, PhD
Department of General Surgery
Beijing Chaoyang Hospital
Capital Medical University
8 Gongren Tiyuchang Nanlua, Chaoyang District
Beijing, 100020
China
Phone: 86 010 85231610
Fax: 86 010 65060167
Email: anzhenjzl@mail.ccmu.edu.cn

Abstract

Background: Gastrointestinal bleeding is a serious adverse event of coronary artery bypass grafting and lacks tailored risk assessment tools for personalized prevention.

Objective: This study aims to develop and validate predictive models to assess the risk of gastrointestinal bleeding after coronary artery bypass grafting (GIBCG) and to guide personalized prevention.

Methods: Participants were recruited from 4 medical centers, including a prospective cohort and the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. From an initial cohort of 18,938 patients, 16,440 were included in the final analysis after applying the exclusion criteria. Thirty combinations of machine learning algorithms were compared, and the optimal model was selected based on integrated performance metrics, including the area under the receiver operating characteristic curve (AUROC) and the Brier score. This model was then developed into a web-based risk prediction calculator. The Shapley Additive Explanations method was used to provide both global and local explanations for the predictions.

Results: The model was developed using data from 3 centers and a prospective cohort (n=13,399) and validated on the Drum Tower cohort (n=2745) and the MIMIC cohort (n=296). The optimal model, based on 15 easily accessible admission features, demonstrated an AUROC of 0.8482 (95% CI 0.8328-0.8618) in the derivation cohort. In external validation, the AUROC was

XSL•FO
RenderX

0.8513 (95% CI 0.8221-0.8782) for the Drum Tower cohort and 0.7811 (95% CI 0.7275-0.8343) for the MIMIC cohort. The analysis indicated that high-risk patients identified by the model had a significantly increased mortality risk (odds ratio 2.98, 95% CI 1.784-4.978; *P*<.001). For these high-risk populations, preoperative use of proton pump inhibitors was an independent protective factor against the occurrence of GIBCG. By contrast, dual antiplatelet therapy and oral anticoagulants were identified as independent risk factors. However, in low-risk populations, the use of proton pump inhibitors (χ^2_1=0.13, *P*=.72), dual antiplatelet therapy (χ^2_1=0.38, *P*=.54), and oral anticoagulants (χ^2_1=0.15, *P*=.69) were not significantly associated with the occurrence of GIBCG.

Conclusions: Our machine learning model accurately identified patients at high risk of GIBCG, who had a poor prognosis. This approach can aid in early risk stratification and personalized prevention.

Trial Registration: Chinese Clinical Registry Center ChiCTR2400086050; http://www.chictr.org.cn/showproj.html?proj=226129

(J Med Internet Res 2025;27:e68509) doi:10.2196/68509

KEYWORDS

machine learning; personalized medicine; coronary artery bypass grafting; adverse outcome; gastrointestinal bleeding

Introduction

Approximately 1.5 million people worldwide undergo cardiac surgery each year, with coronary artery bypass grafting (CABG) being the most common procedure [1,2]. The incidence of gastrointestinal bleeding (GIB) after CABG (GIBCG) ranges from approximately 0.39% to 5.5% [3-8], increasing between 5% and >10% in patients treated with dual antiplatelet agents [9]. In the study by Yang et al [10], the observed incidence rate reached 22.4%. However, once GIBCG occurs, the mortality rate rises significantly, ranging from 8.8% to 38.0% [3-8], which is notably higher than that of surgical patients without GIB.

GIBCG has delayed and insidious characteristics that make early recognition difficult [4-6]. These patients often experience adverse events such as respiratory or renal failure and cardiac insufficiency, which can mask its clinical signs. Postoperative sedation may also obscure typical abdominal symptoms [6]. Following a GIBCG episode, continued administration of antiplatelet agents can exacerbate bleeding, whereas discontinuing antiplatelet therapy is associated with a higher risk of cardiovascular events and increased mortality rates [11]. Therefore, preventing and diagnosing GIBCG early are essential in clinical practice.

The PRECISE-DAPT (Predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy) score is one of the most commonly used clinical tools for predicting bleeding risk in patients with acute coronary syndrome. It assesses the risk of bleeding within 12 months following percutaneous coronary intervention [12]. Unlike percutaneous coronary intervention, CABG may cause GIB through different mechanisms, such as the effects of extracorporeal circulation and stress [5,13]. However, the applicability of these scores in patients undergoing CABG remains unclear.

Machine learning (ML), a subset of artificial intelligence, effectively identifies nonlinear relationships, deciphers intricate interactions, and manages multicollinearity among predictor variables [14-16]. ML has been widely applied in medicine, leveraging large volumes of patient data to build risk models. These models can predict disease onset, assess condition severity, and evaluate disease prognosis [17-19]. ML outperforms traditional scores in predicting GIB risk after antithrombotic therapy [20]. However, GIB risk prediction after cardiac surgery remains understudied. To date, only 1 study—a single-center, small-sample, traditional risk prediction model using intraoperative and postoperative clinical features—has reported on GIBCG risk prediction, but it does not allow for preoperative risk assessment [10]. In this study, we aimed to develop and evaluate ML-driven risk prediction models using preoperative clinical features to identify individuals at a high risk of GIBCG, thereby enabling proactive preventive measures.

Methods

Ethical Considerations

The Medical Ethics Committee of Beijing Anzhen Hospital, affiliated with Capital Medical University, approved the study protocol (approval number KS2023020). This study is retrospectively registered in the Chinese Clinical Trial Registry (ChiCTR2400086050), which includes GIB as the primary outcome along with other gastrointestinal complications. GIB was the main focus of this study due to its high morbidity and lethality. Informed consent was obtained from all participants. DJ obtained access to and download permission for the Medical Information Mart for Intensive Care IV (MIMIC-IV) database (ID: 59888302). All data have undergone deidentification processing to ensure anonymity.

Patients and Study Design

The study included patients from 4 hospitals in China, with the inclusion criteria being CABG during hospitalization, age over 18 years, and availability of complete key information. Data were also obtained from the American Critical Care database, the MIMIC-IV, which met the same inclusion criteria [21]. Patients hospitalized at Beijing Anzhen Hospital between January 2018 and May 2023, Beijing Luhe Hospital between July 2019 and May 2023, and Beijing Chaoyang Hospital between May 2022 and April 2024 were included in the derivation cohort. Additionally, 3466 patients prospectively recruited at Beijing Anzhen Hospital between May 2023 and April 2024 were included as a prospective cohort within the same derivation cohort. Patients hospitalized at Nanjing Drum Tower Hospital between October 2010 and May 2023, along with data from MIMIC-IV, were included in the external

XSL•FO
RenderX

validation cohort. The exclusion criteria were as follows: patients diagnosed preoperatively with GIB and those whose bleeding cause could not be identified due to comorbidities such as severe liver disease or gastrointestinal malignancy. Figure S1 in Multimedia Appendix 1 details the patient selection process for the derivation and external validation cohorts.

Sample Size Calculation

When developing prediction models, a common practice for determining the required sample size is to ensure at least 10 events per candidate predictor parameter. Additionally, we followed the 4-step procedure proposed by Riley et al [22] to calculate the required sample size. The calculation process, formulas, and results are presented in Table S1 in Multimedia Appendix 1.

Definition of Outcome

The primary outcome was the occurrence of GIB after CABG, defined as (1) hematemesis (including bright red blood or coffee-ground emesis), melena, hematochezia, or positive occult blood tests in either gastric fluid or stool; and (2) bleeding foci identified during endoscopic examination [23-25]. The secondary outcome was the in-hospital mortality rate.

Predictive Features

Admission variables were used as predictive features. Table S2 in Multimedia Appendix 1 provides the definitions of each feature. Data extraction for patients undergoing CABG and the diagnosis of comorbidities were based on International Classification of Diseases 9th/10th Revision (ICD-9/10) disease codes (Table S3 in Multimedia Appendix 1).

Statistical Analysis

Data analysis was performed using SPSS Statistical Software version 26.0(IBM Corp.). Normally distributed continuous variables are presented as means with SDs, while skewed continuous variables are presented as medians with IQRs. For statistical analysis, the independent samples *t* test was used to compare normally distributed continuous variables, and the Mann-Whitney *U* test was used for skewed distributions. Categorical variables are presented as numbers with percentages and were compared using the chi-square or Fisher exact test. Significant variables from the univariate analysis were selected for multivariate unconditional logistic regression. Differences were considered statistically significant at $P<.05$. Missing values were imputed using multiple imputation methods in R version 4.1.0 9 (R Foundation). Subsequent feature screening and model construction were performed using Python version 3.7.0 (Python Foundation). Table S4 in Multimedia Appendix 1 presents the data types and missing values for each feature.

Feature Selection

Figure 1 presents the complete study flowchart. We employed 4 distinct feature selection methods to identify candidate features for model development: least absolute shrinkage and selection operator (LASSO), k-best feature selection (K-Best), random forest recursive feature elimination (RFE), and the mutual information algorithm [26]. A fifth method was a combined approach that selected features appearing in at least three of the four feature selection methods.

XSL•FO
RenderX

Figure 1. Flowchart of the study. The study process is summarized as follows: data were collected from multiple centers, patient admission features were extracted, and the study cohort was defined. A derivation cohort was formed from 3 centers and a prospective cohort, which were used for feature screening and model construction. The optimal model was selected using multiple evaluation metrics and externally validated with the Drum Tower and MIMIC cohorts. Finally, the model was developed into a user-friendly web page to facilitate clinical use and guide treatment decisions. Abbreviations: COM: combined method, including features appearing more than three times in the first four feature selection methods; K-Best: k-best feature selection; LASSO: least absolute shrinkage and selection operator; LR: logistic regression; MI: mutual information; MIMIC IV: medical information mart for intensive care IV; MLP: multilayer perceptron; NB: naive Bayes; RF: random forest; RFE: recursive feature elimination; ROC curve: receiver operating characteristic curve; SHAP: Shapley additive explanations; SVM: support vector machine; XGB: XGBoost (extreme gradient boosting).

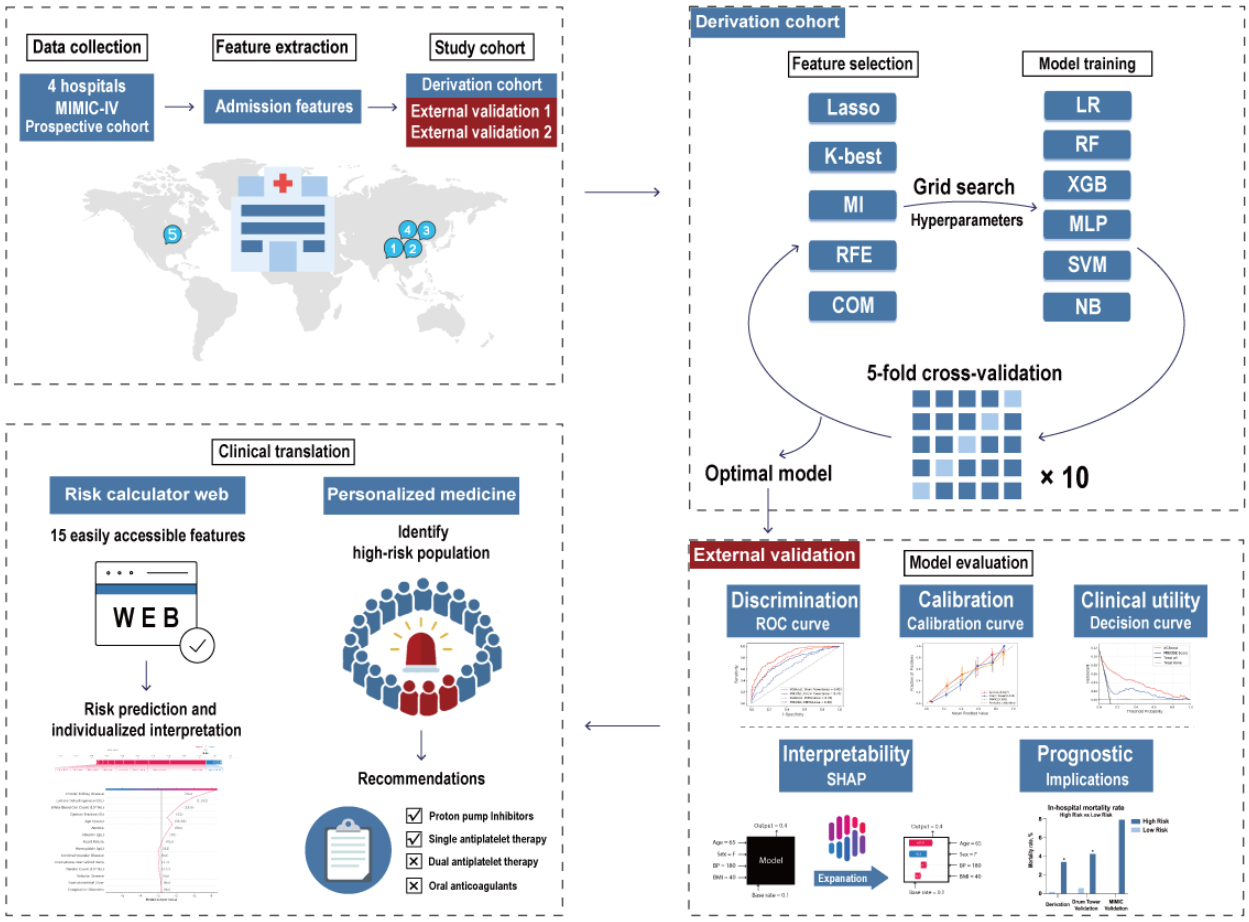

Model Development and Validation

Based on feature engineering, we applied 6 supervised ML algorithms: multilayer perceptron, naive Bayes, random forest, Extreme Gradient Boosting (XGBoost; also called XGB), logistic regression, and support vector machines. By combining 5 feature selection methods with these 6 classifiers, we built 30 models (5 × 6 = 30) and optimized hyperparameters for each model using a 5-fold grid search cross-validation on the derivation cohort. Grid search systematically explores all possible combinations of hyperparameters, while 5-fold cross-validation evaluates each combination's performance across different validation subsets. The optimal hyperparameter combination for each model was selected based on the highest performance across cross-validation folds. The hyperparameter settings, ranges, and rationale for each algorithm are detailed in Table S5 in Multimedia Appendix 1. The final model selection was based on a comprehensive evaluation of multiple performance metrics, including the average and SD of both the area under the receiver operating characteristic curve (AUROC) and the Brier score [27]. AUROC measures the model's ability to distinguish between different classes, which is particularly important in imbalanced data sets. A higher AUROC value indicates better class discrimination. The Brier score, by contrast, assesses both the accuracy and calibration of predicted probabilities by quantifying the difference between predicted probabilities and actual outcomes. Using these metrics provides a well-rounded evaluation of model performance in terms of both discriminative power and predictive accuracy. The average and SD of AUROC and Brier score were calculated based on all validation results from 5-fold, 10-time cross-validation. The ideal model performance is characterized by an average AUROC close to 1, a Brier score close to 0, and low SDs for both metrics. To quantify overall model performance, we first normalized the average AUROC. Next, we normalized the average Brier score and the SDs of both AUROC and Brier score by subtracting them from 1. The comprehensive model score was then obtained by averaging these 4 values. Based on this score, we selected the optimal combination of the feature selection method and ML model for further evaluation in external validation.

Model Evaluation

In the external validation cohort, we further assessed the performance of the final prediction models using AUROC,

XSL•FO
RenderX

calibration plots, and decision curve analysis (DCA). Model discrimination was evaluated using AUROC, and we compared the discriminative performance of the optimal ML model with PRECISE-DAPT scores. Calibration was demonstrated through calibration plots and quantified using the Brier score, while clinical utility was assessed using DCA [28]. We categorized bleeding severity, defining severe GIB as the presence of marked symptoms such as hematemesis, melena, or hematochezia, or the identification of bleeding foci during endoscopic examination. Patients with positive occult blood tests in gastric fluid or stool were classified as having mild bleeding. Additionally, we assessed the model's ability to predict severe bleeding within each study cohort.

Model Interpretations

ML modeling often operates in a "black box" environment, where the complexity and multidimensionality of algorithms make it difficult to clearly elucidate the internal mechanisms driving accurate predictions for specific patient groups. To enhance model interpretability, we used the Shapley Additive Explanations (SHAP) algorithm [29,30]. SHAP is a game theory-based model explanation method that provides consistent and locally precise attribution values—known as SHAP values—for each feature in the model. This approach highlights the significance of individual features and explains the model's decision-making process. Higher SHAP values indicate a greater likelihood of GIBCG.

Clinical Application

We determined the model's optimal cutoff values by identifying the maximum Youden index (sensitivity + specificity – 1) in the derivation cohort. Using these cutoff values, patients were classified into high- and low-risk groups. Subgroup analyses were conducted based on preoperative medication use across different risk groups. Finally, we developed a web-based interface integrating the optimal ML model, which utilizes questionnaire-guided responses for risk assessment.

Results

Patient Characteristics

The calculated sample size was 5667. However, to meet the requirements of ML training, we used a data set that far exceeded this minimum to ensure optimal model performance. In total, we included 16,440 patients, with 13,399 in the derivation cohort and 3041 in the external validation cohort. Baseline characteristics for these groups are presented in Table 1. In the derivation cohort (N=13,399), 803 patients (5.99%) developed GIBCG. The external validation cohort comprised data from Nanjing Drum Tower Hospital and MIMIC, incorporating both Chinese and American centers. The Drum Tower data set (n=2745) was designated as external validation cohort 1, in which 179 patients (6.52%) developed GIBCG. The MIMIC data set (n=296) was designated as external validation cohort 2, with 176 patients (59.5%) developing GIBCG. Baseline characteristics stratified by GIBCG occurrence are compared in Tables S6-S8 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Baseline characteristics of the derivation and external validation cohorts.

Demographic variables	Derivation cohort	External validation cohort	
	Multicenter (n=13,399)	Drum Tower (n=2745)	MIMIC[a] (n=296)
Age, median (IQR), years	63 (11)	70 (14)	74 (15)
Sex, n (%)			
Male	10,108 (75.44)	1921 (69.98)	198 (66.89)
Female	3291 (24.56)	824 (30.02)	98 (33.11)
BMI, median (IQR)	25.59 (4.09)	24.4 (4.3)	30.09 (7.31)
Gastrointestinal bleeding history, n (%)	102 (0.76)	29 (1.06)	30 (10.14)
After percutaneous coronary intervention, n (%)	1578 (11.78)	262 (9.54)	28 (9.46)
Chronic comorbidities, n (%)			
Anemia	1704 (12.72)	319 (11.62)	122 (41.22)
Coagulation disorders	147 (1.10)	17 (0.62)	92 (31.08)
Hypertension	6238 (46.56)	1691 (61.60)	131 (44.26)
Atrial fibrillation	618 (4.61)	258 (9.40)	185 (62.50)
Diabetes	5314 (39.66)	832 (30.31)	100 (33.78)
Heart failure	5114 (38.17)	1114 (40.58)	146 (49.32)
Cerebral vascular disease	2045 (15.26)	484 (17.63)	53 (17.91)
Peripheral vascular disease	347 (2.59)	196 (7.14)	71 (23.99)
Chronic kidney disease	614 (4.58)	133 (4.85)	113 (38.18)
Gastrointestinal ulcer	209 (1.56)	37 (1.35)	11 (3.72)
Gastritis	155 (1.16)	85 (3.10)	10 (3.38)
Hyperlipidemia	8413 (62.79)	1689 (61.53)	201 (67.91)
Valvular disease	2087 (15.58)	433 (15.77)	107 (36.15)
Admission examination, median (IQR)			
White blood cell count ($\times 10^9$/L)	6.89 (2.83)	6.3 (2.4)	11.35 (7.68)
Red blood cell count ($\times 10^{12}$/L)	4.41 (0.77)	4.38 (0.71)	3.25 (0.79)
Platelet count ($\times 10^9$/L)	207 (79)	191 (73.5)	176.5 (136)
Hemoglobin (g/L)	136 (25)	134 (22)	98 (23)
Hematocrit (proportion of 1.0)	0.396 (0.069)	0.398 (0.06)	0.295 (0.063)
Alanine aminotransferase (µkat/L)	0.33 (0.3)	0.36 (0.32)	0.4 (0.56)
Aspartate aminotransferase (µkat/L)	0.32 (0.17)	0.37 (0.24)	0.62 (0.99)
Bilirubin total (µmol/L)	11.01 (6.4)	10.8 (6.9)	10.26 (13.25)
Albumin (g/L)	42.7 (4.8)	39.6 (4)	33.07 (6.09)
Urea (mmol/L)	5.77 (2.28)	6.1 (2.5)	9.64 (8.93)
Creatinine (µmol/L)	75 (21.7)	71 (25.55)	106.08 (79.56)
Prothrombin time (seconds)	11.4 (1.1)	11.5 (1.3)	14.4 (4.58)
Activated partial thromboplastin time (seconds)	31 (4.4)	27.8 (4.1)	31.8 (12.8)
International normalized ratio	1.01 (0.09)	1 (0.12)	1.3 (0.5)
Lactate dehydrogenase (µkat/L)	2.92 (0.77)	3.34 (1.45)	5.06 (3.01)
Admission examination			
Ejection fraction, n (%)			
≥55%	10,611 (79.19)	1572 (57.27)	109 (36.82)

XSL•FO
RenderX

Demographic variables	Derivation cohort	External validation cohort	
	Multicenter (n=13,399)	Drum Tower (n=2745)	MIMIC[a] (n=296)
45%-55%	1772 (13.22)	579 (21.09)	76 (25.68)
30%-44%	968 (7.22)	543 (19.78)	72 (24.32)
<30%	48 (0.36)	51 (1.86)	39 (13.18)
PRECISE-DAPT[b] score, median (IQR)	12.3 (11.36)	14.53 (10.78)	40.48 (22.46)
In-hospital mortality, n (%)	123 (0.92)	69 (2.51)	22 (7.43)

[a]MIMIC: Medical Information Mart for Intensive Care.

[b]PRECISE-DAPT: predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy.

Predictive Features

Five feature selection methods—LASSO, K-Best, RFE, mutual information algorithm, and a combined approach—were used to optimize the feature set (Table S9 in Multimedia Appendix 1). The features identified through these methods were incorporated as predictors of GIBCG and subsequently used in the ML model for further evaluation.

Model Performance

Using grid search and cross-validation, we identified the optimal hyperparameters for each model configuration (Table S10 in Multimedia Appendix 1). Table S11 presents the average AUROC and its 95% CI for each model in the training set, constructed from 4 folds of the 5-fold cross-validation in the derivation cohort. Table S12 displays the AUROC for the internal validation set, which consists of the remaining fold. Similarly, Tables S13 and S14 report the mean Brier score and its 95% CI for the training and internal validation sets, respectively. We constructed 6 models for each of the 5 feature selection methods, yielding 30 (5×6) models. Figure 2A presents a heatmap of comprehensive scores, ranging from 0.1789 to 0.9938. Vertically, models developed using the LASSO method achieved an average score of 0.8709, outperforming those built with other feature selection techniques. Horizontally, XGBoost demonstrated superior performance across feature selection methods, with an average score of 0.9128. Notably, the LASSO-XGBoost combination attained the highest score of 0.9938, making it the top-performing model. Figure 2B illustrates the AUROC of models following LASSO feature selection, with XGBoost achieving the highest AUROC of 0.8482 (95% CI 0.8328-0.8618, bootstrap), outperforming all other models.

Figure 2. Performance of machine learning models. (A) Heatmaps illustrating the predictive performance (model score) of different combinations of feature selection methods (rows) and classification algorithms (columns). (B) Receiver operating characteristic curves of various models following LASSO feature selection. Abbreviations: COM: combined method; K-Best: k-best feature selection; LASSO: least absolute shrinkage and selection operator; LR: logistic regression; MI: mutual information; MLP: multilayer perceptron; NB: naive Bayes; PRECISE: "predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy" score; RF: random forest; RFE: recursive feature elimination; SVM: support vector machine; XGB: XGBoost (extreme gradient boosting).

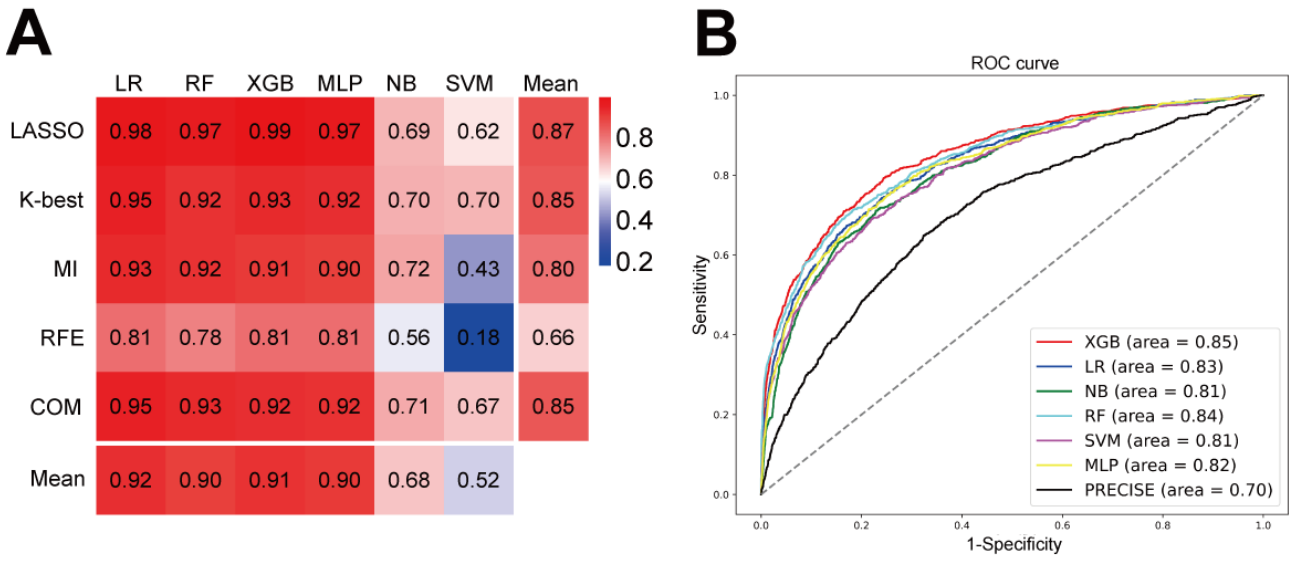

Model Evaluation

The final prediction model was evaluated using 2 independent external cohorts from China and the United States. Figure 3A presents calibration plots for the XGBoost model across the derivation cohort and both external validation cohorts. Given the higher positive rate in the MIMIC cohort, we applied Platt scaling to calibrate predicted probabilities for this data set. In

XSL•FO
RenderX

the derivation cohort, the XGBoost model achieved a Brier score of 0.0446 (95% CI 0.0419-0.0474). The Brier score was 0.0489 (95% CI 0.0431-0.0549) in the Drum Tower validation cohort and 0.2044 (95% CI 0.1787-0.2322) in the MIMIC validation cohort. These results indicate that our model demonstrates the highest calibration in the 2 Chinese cohorts while maintaining satisfactory calibration in the American cohort. Figure 3B compares the AUROC of the XGBoost model and the PRECISE-DAPT score within the Drum Tower validation cohort. The AUROC for XGBoost was 0.8513 (95% CI 0.8221-0.8782), exceeding that of the PRECISE-DAPT score (0.7427, 95% CI 0.7028-0.7836). Similarly, in the MIMIC validation cohort (dashed curves in Figure 3B), XGBoost achieved an AUROC of 0.7811 (95% CI 0.7275-0.8343), outperforming the PRECISE-DAPT score (0.6460, 95% CI 0.5863-0.7123). Figure 3C presents a DCA integrating both external validation cohorts, showing that the net benefit of XGBoost surpasses that of the PRECISE-DAPT score.

Figure 3. Evaluation of the optimal model and its interpretability. (A) Calibration plots of XGBoost (Extreme Gradient Boosting) in the derivation cohort and 2 independent external validation cohorts. The Brier score is reported in the lower-right legend (a smaller value indicates better calibration). (B) Comparison of the area under the receiver operating characteristic curve (AUROC) for XGBoost and the PRECISE-DAPT (predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy) score in the Drum Tower and MIMIC (Medical Information Mart for Intensive Care) validation cohorts. (C) Comparison of clinical decision curves for XGBoost and the PRECISE-DAPT score in the combined external validation cohort. (D) Bar chart of the mean absolute Shapley Additive Explanations (SHAP) value for each predictor.(E) SHAP summary plot. The SHAP value along the x axis serves as a standardized measure of feature influence within the model. Each row of critical features illustrates the contributions of all patients to the outcome, with dots in different colors: red representing high-risk values and blue representing low-risk values.

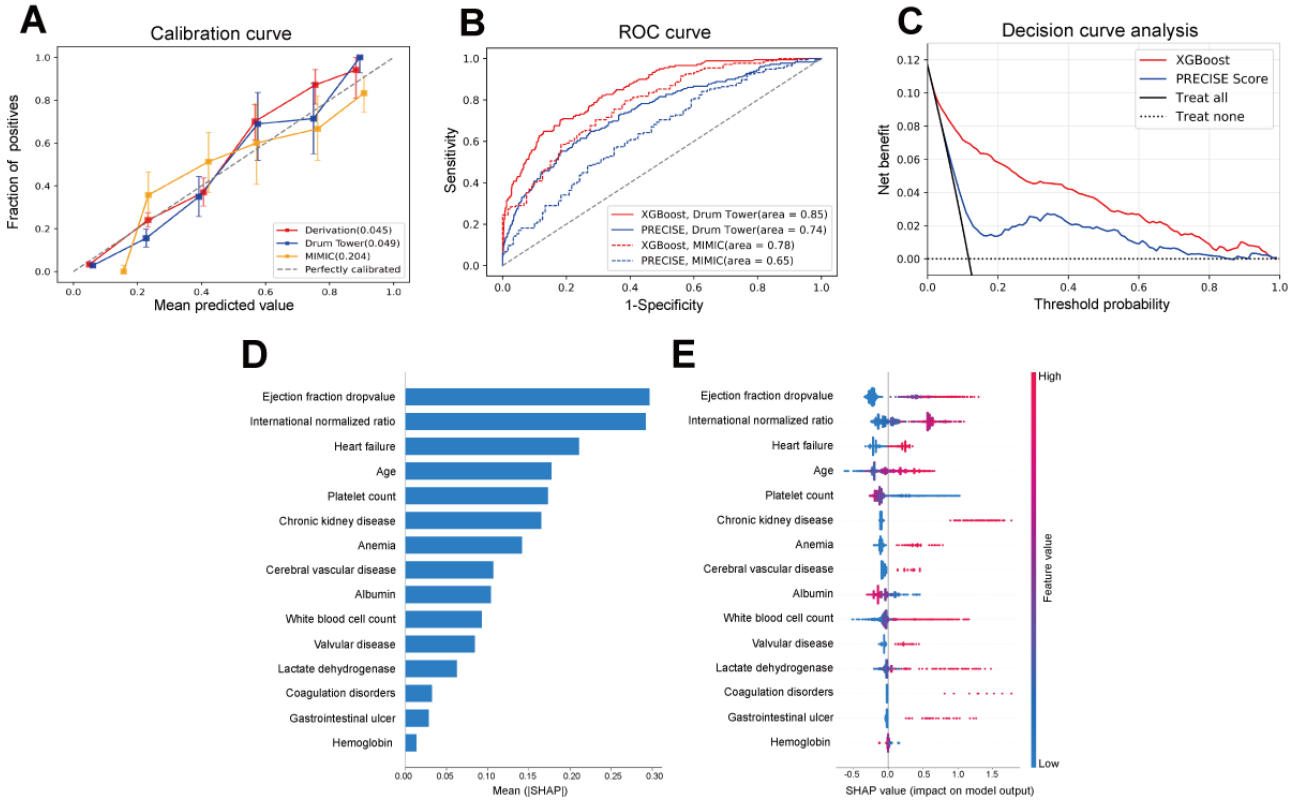

The proportion of patients with severe bleeding among those with GIBCG is presented for each study cohort (Figure S2 in Multimedia Appendix 1). The model demonstrated excellent discriminative ability for severe GIB, with AUROCs of 0.8912 (95% CI 0.8626-0.9196) in the derivation cohort, 0.8817 (95% CI 0.8249-0.9296) in the Drum Tower validation cohort, and 0.8388 (95% CI 0.7749-0.8931) in the MIMIC validation cohort (Figure S3 in Multimedia Appendix 1).

The MIMIC cohort consists of critically ill patients; therefore, we assessed disease severity using intensive care unit stay duration, Sequential Organ Failure Assessment (SOFA) score on the first day, and the Charlson Comorbidity Index for further subgroup analysis (Figure S4 in Multimedia Appendix 1). In the figure, the red line represents the intensive care unit stay duration cutoff, with the model's AUROC calculated for patients below this threshold. The blue and green lines indicate the SOFA score and the Charlson Comorbidity Index, respectively. Line charts illustrate AUROC changes across different cutoff values for these 3 indicators. The results show that as disease severity increases (ie, with higher cutoff values), the AUROC gradually decreases, suggesting that the model performs better in predicting less severe cases than more severe ones.

Model Interpretability Using SHAP Values

Figure 3D presents a bar chart displaying the average absolute SHAP values for each predictive feature, with higher SHAP values indicating greater feature influence. Figure 3E provides a SHAP summary plot illustrating how each predictive feature impacts the XGBoost model's predictions. A decrease in ejection fraction, international normalized ratio (INR), platelet count, albumin, and hemoglobin levels, and the presence of heart

XSL•FO
RenderX

failure, advanced age, chronic kidney disease, anemia, cerebrovascular disease, valvular disease, elevated white blood cell count, increased lactate dehydrogenase levels, coagulation disorders, and gastrointestinal ulcer were associated with a higher predicted probability of GIBCG.

Prognostic Implications

We evaluated the prognostic implications across all 16,440 hospitalized patients from the derivation and external validation cohorts. In the derivation cohort, an optimal cutoff value of 0.0592 was identified based on the maximum Youden index to differentiate between high- and low-risk populations, and the same threshold was applied to the external validation cohort. Univariate and multivariate analyses of risk factors for mortality were conducted across all cohorts (Table S15 in Multimedia Appendix 1). The findings indicate that the model-identified risk levels (high risk vs low risk) serve as an independent risk factor for in-hospital mortality (odds ratio [OR] 2.98, 95% CI 1.784-4.978; *P*<.001). The model's risk probability output was also an independent risk factor for in-hospital mortality (OR 1.017, 95% CI 1.004-1.031; *P*=.009). Compared with the low-risk population, the high-risk population exhibited significantly higher mortality rates in the derivation cohort (106/4534, 2.34% versus 17/8865, 0.19%; χ^2_1=151.91, *P*<.001), Drum Tower validation cohort (60/1354, 4.43% versus 9/1391, 0.65%; χ^2_1=40.09, *P*<.001), and MIMIC validation cohort (22/278, 7.91% versus 0/18, 0%; χ^2_1=0.60, *P*=.44; Figure S5 in Multimedia Appendix 1). Although the PRECISE-DAPT score is an independent risk factor for in-hospital mortality (Table S15 in Multimedia Appendix 1), Figure S6 in Multimedia Appendix 1 presents the receiver operating characteristic curves for predicting in-hospital mortality using both our model and the PRECISE-DAPT score. The AUROC for the risk probability predicted by our model was 0.8642 (95% CI 0.8384-0.8860), compared with 0.7591 (95% CI 0.7266-0.7939) for the PRECISE-DAPT score, demonstrating the superior prognostic performance of our model.

Guiding Personalized Medicine

Given the substantial variations in prognosis across different risk populations, we conducted a subgroup analysis comparing preoperative medications between CABG patients with and without GIB to explore potential intervention strategies. In all cohorts, preoperative use of proton pump inhibitors (PPIs; χ^2_1=53.45, *P*<.001), dual antiplatelet therapy (DAPT; χ^2_1=16.53, *P*<.001), and oral anticoagulants (OACs; χ^2_1=29.87, *P*<.001) was significantly associated with the occurrence of GIBCG. However, in the low-risk population, PPIs (χ^2_1=0.13, *P*=.72), DAPT (χ^2_1=0.38, *P*=.54), and OACs (χ^2_1=0.15, *P*=.69) were not significantly associated (Table S16 in Multimedia Appendix 1). In the high-risk population, univariate analysis identified preoperative use of PPIs and single antiplatelet therapy (SAPT) as protective factors against GIBCG, whereas DAPT and OACs were associated with an increased risk. Furthermore, multivariate analysis confirmed that PPIs, DAPT, and OACs were independent correlates of GIBCG occurrence (Table S17 in Multimedia Appendix 1)

These findings were further validated in the Drum Tower and MIMIC validation cohorts (Table S16 in Multimedia Appendix 1), confirming the universality and robustness of our cutoff value. The analysis of GIBCG risk factors in the Drum Tower validation cohort aligned with those in the derivation cohort, particularly regarding preoperative medication (Table S18 in Multimedia Appendix 1). Although the MIMIC validation cohort, derived from the Critical Care Database, differed from the other 2 cohorts, its analysis yielded largely consistent results. Specifically, preoperative use of PPIs (OR 0.517, 95% CI 0.281-0.95; *P*=.03), SAPT (OR 0.56, 95% CI 0.342-0.915; *P*=.02), and DAPT (OR 2.821, 95% CI 1.12-7.106; *P*=.03) was significantly associated with GIBCG occurrence in the high-risk population (Table S19 in Multimedia Appendix 1).

Clinical Application

The XGBoost model was ultimately selected as the algorithm for the web-based calculator available on our website (Figure 4; also see [31]). Users input 15 admission features into the questionnaire interface, and the web page automatically predicts the patient's GIBCG risk level. The results section displays the predicted risk along with its classification. Simultaneously, the calculator generates a force plot to interpret the prediction, highlighting the features influencing the decision. SHAP values provide insights into the contribution of each feature to the predicted risk of GIBCG: blue features on the right drive the prediction toward "non-GIB," while red features on the left indicate a higher likelihood of "GIB." For example, when we entered the admission features of 2 patients from the Drum Tower cohort, the calculator provided their risk predictions along with corresponding interpretation plots (Figure 4). At the time of admission, patient 1 presented with decreased albumin, elevated white blood cell counts, and a low platelet count within the normal range. These 3 factors were identified as the primary contributors to the increased GIBCG risk and warrant special clinical attention. The model classified patient 1 as high risk and recommended a preventive strategy. Preoperative PPIs use may be considered, and for patients on DAPT or OACs, an early transition to SAPT or alternative therapies is advised. By contrast, although patient 2 is older, most of their predictive factors were within normal ranges, leading to a classification as low risk. For patients with indications, antiplatelet therapy can be continued until preoperative cessation to mitigate the risk of adverse cardiovascular events. This personalized medical approach may help reduce GIBCG risk, improve patient prognosis, and enhance the overall quality of health care services.

XSL•FO
RenderX

Figure 4. Web-based calculator of the XGBoost (Extreme Gradient Boosting) model for risk stratification and personalized medicine. The web-based risk calculator predicts risk accurately online and in real time, providing a convenient tool for clinical doctors. It utilizes 15 features, including patient age, comorbidities, and initial examination results upon admission, to predict the risk of gastrointestinal bleeding in cardiac surgery patients. Additionally, it provides individualized interpretations of risk factors, outputs the patient's risk level, facilitates risk stratification, and guides targeted medication use.

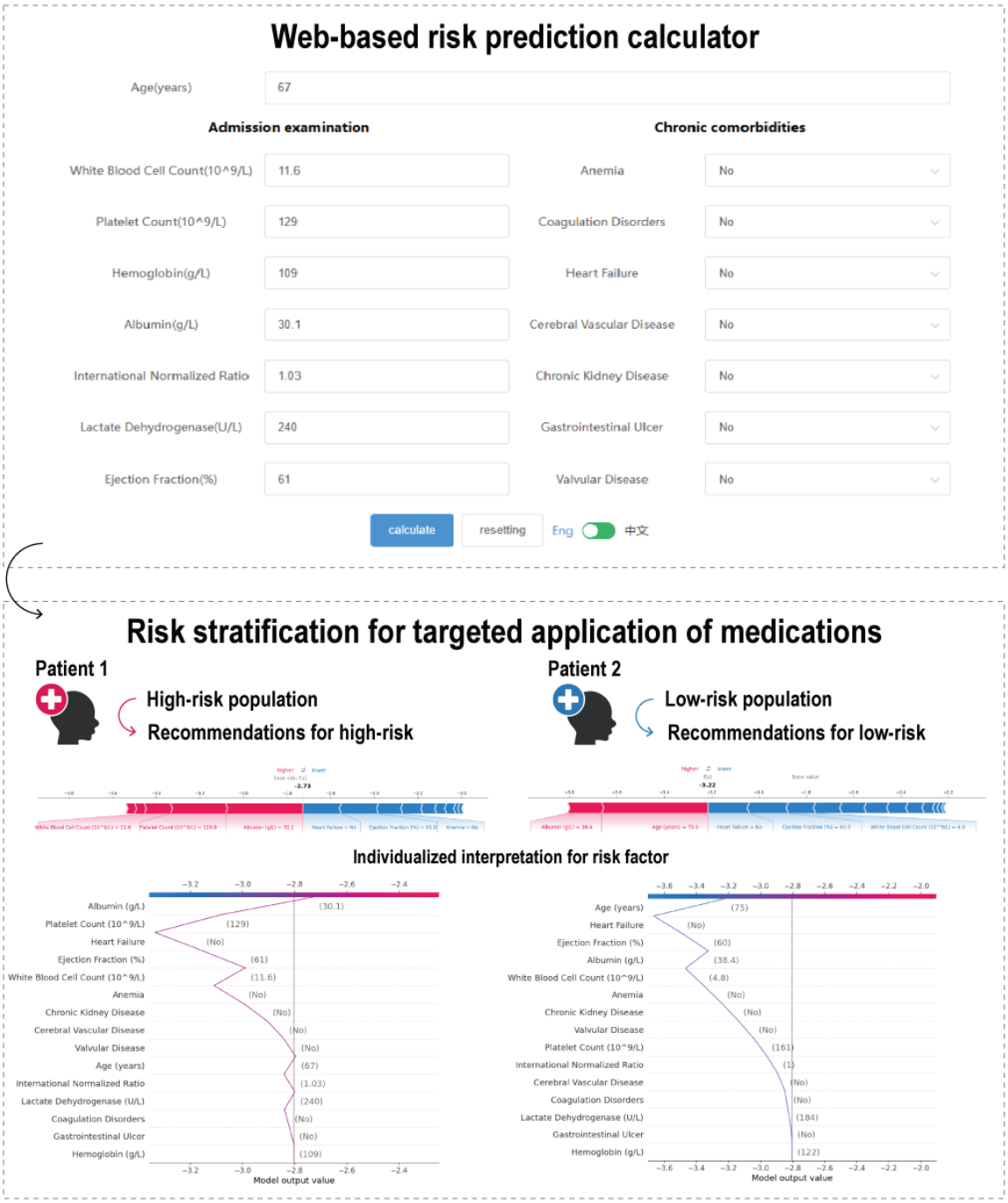

Discussion

Principal Findings

In this multicenter study, we developed and validated ML-based prediction models for GIBCG. After evaluating various feature selection methods and ML algorithms, we identified LASSO for feature selection and XGBoost for model development as the optimal combination. The final model, utilizing 15 admission features, demonstrated strong discriminative performance, calibration, and clinical utility across both the derivation and external validation cohorts. The top 5 predictive features for GIBCG were ejection fraction, INR, chronic heart failure, age, and platelet count, in that order. To our knowledge, this study is the first to develop a model that accurately predicts GIBCG using admission features. Furthermore, our model is prognostically relevant, as it identifies high-risk individuals with an increased postoperative mortality rate, underscoring the need for clinical vigilance. By enabling early risk stratification and targeted prevention, our ML model may help mitigate the risk of GIBCG.

In the derivation and Drum Tower validation cohorts, the prevalence of GIBCG was 803 out of 13,399 (5.99%) patients and 179 out of 2745 (6.52%) patients, respectively, slightly higher than previous epidemiological estimates. This

XSL•FO
RenderX

discrepancy may stem from our inclusion of patients with positive occult blood tests as positive cases. While these patients have no overt bleeding, early identification and preventive measures may help avert the progression to significant bleeding. Including occult blood positivity in the outcome definition enhances the model's sensitivity, enabling earlier risk detection and management. Furthermore, subgroup analysis demonstrated that the model performed better in identifying patients with more severe bleeding, reinforcing its utility for early screening of high-risk patients.

The model developed in the derivation cohort was validated in a large, independent external validation cohort (Drum Tower), consistently demonstrating excellent performance. In intensive care, stress-related GIB remains a significant concern [32,33]. To further assess the model's generalization capability across different populations, we tested it on the MIMIC cohort. This introduced potential biases in the following aspects: (1) the incidence of GIBCG in the MIMIC cohort was higher than in the other 2 cohorts, likely due to the greater severity of illness among these patients and the absence of fecal occult blood test results in some cases, which may have led to selection bias; (2) the AUROC for predicting GIBCG in the MIMIC cohort was lower than in the other cohorts. This discrepancy is primarily due to the higher proportion of critically ill patients in the MIMIC cohort. Analysis reveals that the model performs better for the mild subgroup than for the severe cases (Figure S4 in Multimedia Appendix 1). Furthermore, we conducted a subgroup analysis focusing on severe GIB (Figure S3 in Multimedia Appendix 1). When the outcome was defined as severe GIB, the model's AUROC in the MIMIC cohort improved substantially, reaching 0.8388. These findings demonstrate that the model has predictive value in critically ill patient populations, particularly in identifying patients at risk for severe GIB.

Currently, no dedicated tool exists for assessing GIB risk in CABG patients. To address this gap, we developed a new predictive model and compared it with the PRECISE-DAPT score, a commonly used tool in internal medicine for evaluating bleeding risk in patients receiving DAPT [12]. Across all study cohorts, our model demonstrated superior performance, as evidenced by its higher AUROC and greater net benefit, as shown in the receiver operating characteristic and DCA curves. Furthermore, our model has been validated as a predictor of patient prognosis. In addition, we compared its predictive performance for in-hospital mortality with that of the PRECISE-DAPT score, demonstrating a significantly higher AUROC for our model, underscoring its superior prognostic capability. The PRECISE-DAPT score, developed using Cox regression, is widely used to predict bleeding risk in DAPT patients. However, its reliance on clinical variables that exclude cardiovascular function and gastrointestinal status may limit its applicability in patients undergoing cardiac surgery or in predicting GIB risk. By contrast, ML excels at capturing nonlinear relationships and complex interactions between variables, enabling our model to focus specifically on predicting GIBCG with superior performance. However, our model has limitations. It requires validation in larger-scale prospective cohorts to further establish its reliability. Additionally, its interpretability remains a challenge, making it difficult for clinicians to fully understand the reasoning behind its predictions. To address this, we incorporated SHAP to enhance explainability, as detailed in the following sections.

High-risk patients with GIBCG identified by the model have a significantly increased risk of death and should receive heightened clinical attention. Early identification enables timely interventions, such as preventive medication and optimized perioperative management, to mitigate GIBCG risk and improve prognosis. Strategies include PPIs use, risk factor management, and *Helicobacter pylori* eradication, which have been shown to reduce antithrombotic therapy–associated GIB [34]. Current guidelines recommend prophylactic PPIs use in cardiac surgery patients to minimize gastrointestinal adverse events, with a class II a level of evidence [35].

Additionally, managing antithrombotic therapy in high-risk patients with GIBCG should be optimized. Aspirin and clopidogrel are the most commonly used antiplatelet agents, and patients experiencing acute GIB while on aspirin are generally advised to continue therapy whenever possible [36-38]. During the perioperative period of CABG, aspirin can typically be maintained, except in cases of extremely high bleeding risk [39,40]. For patients scheduled for elective CABG who are receiving P2Y12 receptor blockers (eg, clopidogrel, prasugrel, ticagrelor), guidelines recommend discontinuing these agents 5-7 days before surgery to reduce the bleeding risk [41,42]. In patients with an elevated bleeding risk, discontinuation of antiplatelet therapy may be considered even earlier [36,37,43]. Furthermore, the concomitant use of anticoagulants and antiplatelet agents should be avoided whenever possible due to the significantly increased risk of GIB [44]. Existing evidence suggests that prophylactic PPIs use may be beneficial for high-risk patients undergoing CABG. For those already receiving DAPT or OACs preoperatively, an earlier transition to SAPT or alternative therapies should be considered to mitigate bleeding risk.

Additionally, our findings reinforce this evidence, demonstrating that preoperative PPIs use independently protects high-risk patients from developing GIBCG, whereas DAPT and OACs independently increase the risk. The protective effect of PPIs is primarily attributed to their irreversible binding to proton pumps in the secretory tubules and vesicles of gastric parietal cells, leading to H^+/K^+-ATPase inhibition and subsequent suppression of gastric acid secretion [45]. This helps reduce gastric mucosal damage, prevent ulcer formation, and lower the risk of stress ulcer bleeding [46,47]. Moreover, the suppression of gastric acid secretion by PPIs significantly increases gastric pH, preventing the conversion of pepsinogen to pepsin, which helps maintain clot stability and promotes hemostasis [48,49]. By contrast, antiplatelet agents heighten the risk of GIBCG. Aspirin, for instance, disrupts the gastric mucosal protective layer and induces oxidative stress and apoptosis in epithelial cells, leading to direct gastrointestinal injury [50,51]. It also inhibits cyclooxygenase (COX)-1 and COX-2, reducing prostaglandin production via COX-1, which weakens mucosal protection and contributes to indirect gastrointestinal damage [52]. Additionally, aspirin may impact the gut microbiota,

XSL•FO
RenderX

triggering inflammation by stimulating the immune system through Toll-like receptor 4 [53]. Drugs such as clopidogrel and ticagrelor act indirectly on the gastrointestinal mucosa by inhibiting adenosine diphosphate receptors, reducing the release of vascular endothelial growth factors, inhibiting neovascularization, and hindering gastrointestinal mucosal repair, potentially promoting GIB [54,55]. Furthermore, DAPT has a synergistic effect that can further exacerbate gastrointestinal mucosal damage. OACs increase the risk of GIB through mechanisms such as their local anticoagulation effect, direct corrosive action, and mucosal repair inhibition [56-58]. They prolong clotting time and increase capillary fragility and permeability, thereby elevating the risk of bleeding. Gastrointestinal damage caused by OACs is associated with small intestinal mucosal permeability glycoproteins, which regulate OACs concentration in the gastrointestinal tract. An incompletely absorbed OACs in the gastrointestinal tract may exert a local effect on the mucosa, leading to bleeding [59]. The mechanisms outlined above suggest a potential link, highlighting the need for future experimental studies to further explore the molecular mechanisms underlying the relationship between preoperative medications and GIB risk.

ML technology is often regarded as a “black box,” with its prediction and reasoning processes lacking transparency. Therefore, understanding how ML models make decisions is crucial for clinicians to build trust in the model and identify potential biases. This underscores another key advantage of our study: the use of the SHAP method to provide both global and local explanations of the model [29,60]. In the global explanation, the SHAP summary plot illustrates the overall distribution of each feature’s impact on the model output. Moreover, in GIBCG prediction, the top 5 important admission features include cardiac function–related indicators (such as left ventricular ejection fraction and history of chronic heart failure), age, platelet count, and INR. Notably, left ventricular dysfunction and heart failure are among the most significant independent predictors of mortality and other major adverse events, including bleeding, following CABG [61,62]. The increased risk of postoperative GIB in patients with heart failure may be attributed to reduced cardiac output, which leads to gastrointestinal hypoperfusion, resulting in ischemia and hypoxia of the gastrointestinal mucosa and ultimately weakening its protective barrier function [63]. During ischemia-reperfusion, the release of oxygen-free radicals and inflammatory mediators further exacerbates gastric mucosal injury [64], thereby increasing the risk of ulcers and bleeding. Another important factor is age. Specifically, older adult patients are at a higher risk of GIBCG due to increased vascular fragility [65], impaired gastrointestinal mucosal barrier function [66], multiple comorbidities, and long-term use of gastrointestinal-damaging medications (eg, aspirin). Studies have shown that age over 65 years (OR 2.1) is a risk factor for gastrointestinal complications after CABG [67], while age (OR 1.04) is a predictor of bleeding and transfusion during surgery [68]. These findings align with the interpretation of features in our model, underscoring their clinical significance. Specifically, older adult patients or those with poor preoperative cardiac function are at an increased risk of developing GIB postoperatively and warrant heightened attention. For such populations, greater emphasis should be placed on GIB prevention during preoperative preparation, with enhanced perioperative monitoring, protection of cardiac function, and optimization of blood flow management. Furthermore, coagulation status is closely associated with the occurrence of GIB after cardiac surgery [69-73]. In particular, platelet counts [72] and INR [73] are key indicators of coagulation function that are simple to measure and readily accessible. Abnormal values at admission can help clinicians promptly adjust perioperative antithrombotic treatment strategies, thereby reducing the risk of GIBCG.

Local explanations provide personalized interpretations for each patient’s prediction, while force plots visualize the specific contributions of each feature to the predicted value. For example, in Figure 4, patient 1 is relatively young and has comparatively good cardiac function. However, decreased albumin, elevated white blood cell count, and slightly lower platelet levels (though within the normal range) are identified as the main factors contributing to the increased risk of GIBCG. Previous studies have shown that preoperative hypoalbuminemia is significantly associated with increased mortality [74,75] and postoperative complications, including bleeding [74], in cardiac surgery patients. Notably, an elevated white blood cell count reflects an underlying inflammatory state, which has been associated with increased mortality and complications following CABG [76,77]. Nutritional support, close monitoring of liver function, management of inflammatory responses, and optimization of antithrombotic therapy may help reduce the risk of GIBCG in this patient. Based on readily accessible features, our web platform enables early prediction of GIBCG risk and provides personalized explanations through force plots. This serves as a valuable reference for clinicians in making individualized decisions, thereby supporting the precise prevention of GIBCG.

ML is increasingly applied in personalized medicine and has the potential to improve patient outcomes. However, ethical and practical considerations warrant further discussion [78]. First, patient data privacy is critical. Even deidentified data can be reidentified through available data points [79] or triangulation with other data sets [80]. To mitigate this risk, strategies such as suppression (removing sensitive information), generalization (transforming specific data into ranges), and data minimization (collecting only essential data) are applied [78]. Second, obtaining consent from each patient poses a challenge in acquiring large-scale training data. Alternative methods, such as dynamic consent, broad consent, and implied consent with opt-out options, have been proposed. In our retrospective cohort, broad consent was obtained, allowing for the future use of anonymized data without specific knowledge of the studies [81]. Third, algorithmic bias is a significant concern in ML applications. As ML models are trained on historical data, any biases present in the training data (eg, related to race, gender, or socioeconomic status) may be perpetuated or even amplified in predictions [82]. In addition, experts involved in labeling the data may inadvertently introduce their own biases into the algorithm [83]. Addressing algorithmic bias requires ensuring the diversity and representativeness of the training data set. For instance, our study utilized multicenter data for model training. Moreover, enhancing model interpretability allows health care professionals to understand its decision-making process,

XSL•FO
RenderX

enabling them to identify and mitigate potential biases. Continuous monitoring and evaluation should also be conducted during the model's application. In future research, we plan to validate our model across different populations and refine it based on feedback [84]. Notably, surveys indicate that many American adults are uncomfortable with doctors relying on ML for decision-making [85]. Most ML systems are designed to assist clinicians in decision-making with oversight rather than functioning autonomously [86]. This approach helps alleviate public concerns about ML-driven decisions while ensuring health care quality and patient safety. Future efforts to integrate ML into health care should prioritize the development and implementation of standardized ethical frameworks and guidelines.

Recently, ML has made significant progress in the field of cardiovascular diseases and related surgeries [17,18]. Research has primarily focused on surgical risk assessment [87], prediction of postoperative complications [88-93], evaluation of patient prognosis [94,95], and guidance for personalized medicine [96]. Several ML models have been developed to predict postoperative complications in cardiac surgery, including bleeding [88,89], acute kidney injury [90,91], delirium [92], and major adverse cardiovascular and cerebrovascular events [93]. Furthermore, studies have shown that ML algorithms perform exceptionally well in predicting severe bleeding after CABG [89]. Additionally, ML has been demonstrated to outperform traditional risk-scoring methods in predicting the risk of GIB following antithrombotic therapy [20]. However, in the surgical field, ML models specifically targeting GIBCG remain unavailable. To date, only a single-center study has proposed a traditional risk prediction model [10], which does not enable preoperative prediction of GIBCG risk, thereby limiting early prevention efforts. To address this critical gap, we developed an ML model using multicenter data to predict the risk of GIBCG. The model was validated on 2 independent external cohorts, demonstrating its ability to accurately predict a patient's risk of GIBCG upon admission and guide personalized preventive measures. This approach may help reduce the incidence of this complication and improve patient outcomes. Furthermore, it highlights the significant value of ML in cardiovascular surgery for preventing complications and enhancing the quality of patient care.

Limitations

Our study has several limitations. First, the inclusion of retrospective cohorts may introduce selection bias. However, incorporating a prospective cohort in the model development, along with the use of rigorous selection criteria and a substantial sample size, may help mitigate this limitation. Second, the model was developed based on a Chinese population, and its generalizability to global populations remains uncertain. Nevertheless, its acceptable performance in critical care cohorts in the United States suggests its potential broader applicability. Future prospective validation is necessary to further refine and evaluate the model's performance. Third, the causal relationships between predictive features (risk factors) and GIBCG, as well as between preoperative medication and GIBCG, require further investigation. Randomized controlled trials are essential to determine whether interventions targeting these risk factors and modifications to preoperative medication can effectively prevent GIBCG. Lastly, this study was retrospectively registered, which may introduce biases in design and reporting. To address this, we ensured transparency by adhering to strict inclusion and exclusion criteria and providing a detailed methodology and results. The substantial sample size and multicenter design help mitigate potential biases from retrospective registration and enhance the study's reliability.

Conclusions

In this study, we successfully developed an XGBoost-based ML model for GIBCG risk prediction using 15 readily accessible admission features, facilitating precision prevention and personalized management of GIBCG. For high-risk patients identified by the model—who have an increased risk of in-hospital mortality and require heightened clinical attention—prophylactic PPIs administration and adjustments to antithrombotic therapy may help reduce the incidence of GIBCG. Moreover, the model demonstrated consistent and excellent performance across both the derivation and validation cohorts, confirming its robustness and scalability. To enhance clinical applicability, it has also been integrated into a web-based risk prediction calculator.

Acknowledgments

This research received funding from the National Natural Science Foundation of China (grants 52373294 and 12226005).

Data Availability

Researchers interested in data access should contact the corresponding author (ZJ). Data requests will need to undergo ethical and legal approval by the relevant institutions. The MIMIC-IV data we used are available at [97]. All the codes used for extraction, cleaning, filtering, and modeling are available on GitHub [98].

Authors' Contributions

All the authors approved the final version of the manuscript before submission. JD contributed to conceptualization, data curation, formal analysis, validation, visualization, methodology, and writing, including both the original draft and review and editing. ZJ was responsible for conceptualization, methodology, formal analysis, investigation, and writing, including the original draft and review and editing. CL and JY were involved in project administration, data analysis, interpretation, and writing—review and

XSL•FO
RenderX

editing. YJ contributed to conceptualization, methodology, data curation, and validation. ZL was responsible for data curation and validation. CC contributed to data curation, validation, and writing—review and editing. BZ handled data curation, investigation, and software development. ZY was involved in formal analysis and investigation. YH contributed to data curation, investigation, and supervision. JM worked on software development, visualization, and methodology. PL provided supervision, validation, and writing—review and editing. YL was responsible for project administration and writing—review and editing. DW contributed to conceptualization, methodology, supervision, writing—review and editing, and project administration. ZJ played a key role in the conceptualization, data curation, funding acquisition, methodology, project administration, supervision, resources, formal analysis, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional analysis.
[DOC File , 6751 KB - jmir_v27i1e68509_app1.doc]

References

1. Vervoort D, Lee G, Ghandour H, Guetter CR, Adreak N, Till BM, et al. Global cardiac surgical volume and gaps: trends, targets, and way forward. Ann Thorac Surg Short Rep 2024 Jun;2(2):320-324 [FREE Full text] [doi: 10.1016/j.atssr.2023.11.019] [Medline: 39790140]
2. Alkhouli M, Alqahtani F, Kalra A, Gafoor S, Alhajji M, Alreshidan M, et al. Trends in characteristics and outcomes of patients undergoing coronary revascularization in the United States, 2003-2016. JAMA Netw Open 2020 Feb 05;3(2):e1921326 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21326] [Medline: 32058558]
3. Marsoner K, Voetsch A, Lierzer C, Sodeck GH, Fruhwald S, Dapunt O, et al. Gastrointestinal complications following on-pump cardiac surgery-a propensity matched analysis. PLoS One 2019;14(6):e0217874 [FREE Full text] [doi: 10.1371/journal.pone.0217874] [Medline: 31166962]
4. Elgharably H, Gamaleldin M, Ayyat KS, Zaki A, Hodges K, Kindzelski B, et al. Serious gastrointestinal complications after cardiac surgery and associated mortality. Ann Thorac Surg 2021 Oct;112(4):1266-1274. [doi: 10.1016/j.athoracsur.2020.09.034] [Medline: 33217398]
5. Chor CYT, Mahmood S, Khan IH, Shirke M, Harky A. Gastrointestinal complications following cardiac surgery. Asian Cardiovasc Thorac Ann 2020 Nov;28(9):621-632. [doi: 10.1177/0218492320949084] [Medline: 32777944]
6. Rodriguez R, Robich M, Plate J, Trooskin SZ, Sellke FW. Gastrointestinal complications following cardiac surgery: a comprehensive review. J Card Surg 2010 Mar;25(2):188-197. [doi: 10.1111/j.1540-8191.2009.00985.x] [Medline: 20149010]
7. Chaudhry R, Zaki J, Wegner R, Pednekar G, Tse A, Sheinbaum R, et al. Gastrointestinal complications after cardiac surgery: a nationwide population-based analysis of morbidity and mortality predictors. J Cardiothorac Vasc Anesth 2017 Aug;31(4):1268-1274. [doi: 10.1053/j.jvca.2017.04.013] [Medline: 28800983]
8. Hess NR, Seese LM, Hong Y, Afflu D, Wang Y, Thoma FW, et al. Gastrointestinal complications after cardiac surgery: incidence, predictors, and impact on outcomes. J Card Surg 2021 Mar 11;36(3):894-901. [doi: 10.1111/jocs.15321] [Medline: 33428223]
9. Wolska N, Rozalski M. Blood platelet adenosine receptors as potential targets for anti-platelet therapy. Int J Mol Sci 2019 Nov 03;20(21):5475 [FREE Full text] [doi: 10.3390/ijms20215475] [Medline: 31684173]
10. Yang M, Zhan S, Gao H, Liao C, Li S. Construction and validation of risk prediction model for gastrointestinal bleeding in patients after coronary artery bypass grafting. Sci Rep 2023 Dec 11;13(1):21909 [FREE Full text] [doi: 10.1038/s41598-023-49405-6] [Medline: 38081917]
11. Wang X, Dong B, Hong B, Gong Y, Wang W, Wang J, et al. Long-term prognosis in patients continuing taking antithrombotics after peptic ulcer bleeding. World J Gastroenterol 2017 Jan 28;23(4):723-729 [FREE Full text] [doi: 10.3748/wjg.v23.i4.723] [Medline: 28216980]
12. Costa F, van Klaveren D, James S, Heg D, Räber L, Feres F, PRECISE-DAPT Study Investigators. Derivation and validation of the predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy (PRECISE-DAPT) score: a pooled analysis of individual-patient datasets from clinical trials. Lancet 2017 Mar 11;389(10073):1025-1034. [doi: 10.1016/S0140-6736(17)30397-5] [Medline: 28290994]
13. Krawiec F, Maitland A, Duan Q, Faris P, Belletrutti PJ, Kent WD. Duodenal ulcers are a major cause of gastrointestinal bleeding after cardiac surgery. J Thorac Cardiovasc Surg 2017 Jul;154(1):181-188 [FREE Full text] [doi: 10.1016/j.jtcvs.2017.02.012] [Medline: 28283237]
14. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022 Jan;23(1):40-55. [doi: 10.1038/s41580-021-00407-0] [Medline: 34518686]
15. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023 Mar 30;388(13):1201-1208. [doi: 10.1056/nejmra2302038] [Medline: 36988595]

XSL•FO
RenderX

16. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
17. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018 Jun 12;71(23):2668-2679 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.521] [Medline: 29880128]
18. El Sherbini A, Rosenson RS, Al Rifai M, Virk HUH, Wang Z, Virani S, et al. Artificial intelligence in preventive cardiology. Prog Cardiovasc Dis 2024 Mar;84:76-89. [doi: 10.1016/j.pcad.2024.03.002] [Medline: 38460897]
19. Komuro J, Kusumoto D, Hashimoto H, Yuasa S. Machine learning in cardiology: clinical application and basic research. J Cardiol 2023 Aug;82(2):128-133 [FREE Full text] [doi: 10.1016/j.jjcc.2023.04.020] [Medline: 37141938]
20. Herrin J, Abraham NS, Yao X, Noseworthy PA, Inselman J, Shah ND, et al. Comparative effectiveness of machine learning approaches for predicting gastrointestinal bleeds in patients receiving antithrombotic treatment. JAMA Netw Open 2021 May 03;4(5):e2110703 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.10703] [Medline: 34019087]
21. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 03;10(1):1 [FREE Full text] [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
22. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020 Mar 18;368:m441. [doi: 10.1136/bmj.m441] [Medline: 32188600]
23. Tokar JL, Higa JT. Acute gastrointestinal bleeding. Ann Intern Med 2022 Feb;175(2):ITC17-ITC32. [doi: 10.7326/AITC202202150] [Medline: 35130044]
24. Raju GS, Gerson L, Das A, Lewis B, American Gastroenterological Association. American Gastroenterological Association (AGA) Institute medical position statement on obscure gastrointestinal bleeding. Gastroenterology 2007 Nov;133(5):1694-1696. [doi: 10.1053/j.gastro.2007.06.008] [Medline: 17983811]
25. Rockey DC, Altayar O, Falck-Ytter Y, Kalmaz D. AGA technical review on gastrointestinal evaluation of iron deficiency anemia. Gastroenterology 2020 Sep;159(3):1097-1119 [FREE Full text] [doi: 10.1053/j.gastro.2020.06.045] [Medline: 32828801]
26. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. Front Bioinform 2022;2:927312 [FREE Full text] [doi: 10.3389/fbinf.2022.927312] [Medline: 36304293]
27. Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol 2010 Aug;63(8):938-9; author reply 939. [doi: 10.1016/j.jclinepi.2009.11.009] [Medline: 20189763]
28. Zhao L, Leng Y, Hu Y, Xiao J, Li Q, Liu C, et al. Understanding decision curve analysis in clinical prediction model research. Postgrad Med J 2024 Jun 28;100(1185):512-515. [doi: 10.1093/postmj/qgae027] [Medline: 38453146]
29. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
30. Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian A. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid Anal Prev 2020 Mar;136:105405. [doi: 10.1016/j.aap.2019.105405] [Medline: 31864931]
31. Calculation tool for predicting gastrointestinal bleeding after coronary artery bypass grafting. XSmart Analysis. URL: https://www.xsmartanalysis.com/model/beijing [accessed 2025-02-24]
32. Halling CMB, Møller MH, Marker S, Krag M, Kjellberg J, Perner A, et al. The effects of pantoprazole vs. placebo on 1-year outcomes, resource use and employment status in ICU patients at risk for gastrointestinal bleeding: a secondary analysis of the SUP-ICU trial. Intensive Care Med 2022 Apr 05;48(4):426-434. [doi: 10.1007/s00134-022-06631-2] [Medline: 35122105]
33. MacLaren R, Dionne J, Granholm A, Alhazzani W, Szumita PM, Olsen K, et al. Society of Critical Care Medicine and American Society of Health-System Pharmacists guideline for the prevention of stress-related gastrointestinal bleeding in critically ill adults. Crit Care Med 2024 Aug 01;52(8):e421-e430. [doi: 10.1097/CCM.0000000000006330] [Medline: 39007578]
34. Elshaer A, Abraham NS. Management of anticoagulant and antiplatelet agents in acute gastrointestinal bleeding and prevention of gastrointestinal bleeding. Gastrointest Endosc Clin N Am 2024 Apr;34(2):205-216. [doi: 10.1016/j.giec.2023.09.002] [Medline: 38395479]
35. Sousa-Uva M, Milojevic M, Head SJ, Jeppsson A. The 2017 EACTS guidelines on perioperative medication in adult cardiac surgery and patient blood management. Eur J Cardiothorac Surg 2018 Jan 01;53(1):1-2. [doi: 10.1093/ejcts/ezx448] [Medline: 29240913]
36. Gralnek IM, Stanley AJ, Morris AJ, Camus M, Lau J, Lanas A, et al. Endoscopic diagnosis and management of nonvariceal upper gastrointestinal hemorrhage (NVUGIH): European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2021. Endoscopy 2021 Mar 10;53(3):300-332 [FREE Full text] [doi: 10.1055/a-1369-5274] [Medline: 33567467]
37. Oakland K, Chadwick G, East JE, Guy R, Humphries A, Jairath V, et al. Diagnosis and management of acute lower gastrointestinal bleeding: guidelines from the British Society of Gastroenterology. Gut 2019 May 12;68(5):776-789 [FREE Full text] [doi: 10.1136/gutjnl-2018-317807] [Medline: 30792244]

XSL•FO
RenderX

38. Kamada T, Satoh K, Itoh T, Ito M, Iwamoto J, Okimoto T, et al. Evidence-based clinical practice guidelines for peptic ulcer disease 2020. J Gastroenterol 2021 Apr 23;56(4):303-322 [FREE Full text] [doi: 10.1007/s00535-021-01769-0] [Medline: 33620586]
39. Kulik A, Ruel M, Jneid H, Ferguson TB, Hiratzka LF, Ikonomidis JS, et al. Secondary prevention after coronary artery bypass graft surgery. Circulation 2015 Mar 10;131(10):927-964. [doi: 10.1161/cir.0000000000000182] [Medline: 25679302]
40. Jacob M, Smedira N, Blackstone E, Williams S, Cho L. Effect of timing of chronic preoperative aspirin discontinuation on morbidity and mortality in coronary artery bypass surgery. Circulation 2011 Feb 15;123(6):577-583. [doi: 10.1161/CIRCULATIONAHA.110.957373] [Medline: 21282503]
41. Fox KA, Mehta SR, Peters R, Zhao F, Lakkis N, Gersh BJ, Clopidogrel in Unstable angina to prevent Recurrent ischemic Events Trial. Benefits and risks of the combination of clopidogrel and aspirin in patients undergoing surgical revascularization for non-ST-elevation acute coronary syndrome: the Clopidogrel in Unstable angina to prevent Recurrent ischemic Events (CURE) Trial. Circulation 2004 Sep 07;110(10):1202-1208. [doi: 10.1161/01.CIR.0000140675.85342.1B] [Medline: 15313956]
42. McLean DS, Sabatine MS, Guo W, McCabe CH, Cannon CP. Benefits and risks of clopidogrel pretreatment before coronary artery bypass grafting in patients with ST-elevation myocardial infarction treated with fibrinolytics in CLARITY-TIMI 28. J Thromb Thrombolysis 2007 Oct 24;24(2):85-91. [doi: 10.1007/s11239-007-0016-x] [Medline: 17323134]
43. Veitch AM, Radaelli F, Alikhan R, Dumonceau JM, Eaton D, Jerrome J, et al. Endoscopy in patients on antiplatelet or anticoagulant therapy: British Society of Gastroenterology (BSG) and European Society of Gastrointestinal Endoscopy (ESGE) guideline update. Gut 2021 Sep 06;70(9):1611-1628 [FREE Full text] [doi: 10.1136/gutjnl-2021-325184] [Medline: 34362780]
44. Abraham NS, Noseworthy PA, Inselman J, Herrin J, Yao X, Sangaralingham LR, et al. Risk of gastrointestinal bleeding increases with combinations of antithrombotic agents and patient age. Clin Gastroenterol Hepatol 2020 Feb;18(2):337-346 [FREE Full text] [doi: 10.1016/j.cgh.2019.05.017] [Medline: 31108228]
45. Yang Y, Metz DC. Safety of proton pump inhibitor exposure. Gastroenterology 2010 Oct;139(4):1115-1127. [doi: 10.1053/j.gastro.2010.08.023] [Medline: 20727892]
46. Patel P, Sengupta N. PPIs and beyond: a framework for managing anticoagulation-related gastrointestinal bleeding in the era of COVID-19. Dig Dis Sci 2020 Aug 15;65(8):2181-2186 [FREE Full text] [doi: 10.1007/s10620-020-06408-x] [Medline: 32537704]
47. Guo H, Ye Z, Huang R. Clinical outcomes of concomitant use of proton pump inhibitors and dual antiplatelet therapy: a systematic review and meta-analysis. Front Pharmacol 2021 Aug 2;12:694698 [FREE Full text] [doi: 10.3389/fphar.2021.694698] [Medline: 34408652]
48. Kherad O, Restellini S, Martel M, Barkun A. Proton pump inhibitors for upper gastrointestinal bleeding. Best Pract Res Clin Gastroenterol 2019 Oct;42-43:101609. [doi: 10.1016/j.bpg.2019.04.002] [Medline: 31785730]
49. Yu Z, He J, Cao R, Yang Z, Li B, Hong J, et al. Proton pump inhibitor has no effect in the prevention of post-endoscopic sphincterotomy delayed bleeding: a prospective randomized controlled trial. Front Med (Lausanne) 2023 Jun 2;10:1179512 [FREE Full text] [doi: 10.3389/fmed.2023.1179512] [Medline: 37332745]
50. Goddard PJ, Hills BA, Lichtenberger LM. Does aspirin damage canine gastric mucosa by reducing its surface hydrophobicity? Am J Physiol 1987 Mar;252(3 Pt 1):G421-G430. [doi: 10.1152/ajpgi.1987.252.3.G421] [Medline: 3826379]
51. Hernández C, Barrachina MD, Vallecillo-Hernández J, Álvarez Á, Ortiz-Masiá D, Cosín-Roger J, et al. Aspirin-induced gastrointestinal damage is associated with an inhibition of epithelial cell autophagy. J Gastroenterol 2016 Jul 3;51(7):691-701. [doi: 10.1007/s00535-015-1137-1] [Medline: 26525539]
52. Zhang W, Wang M, Hua G, Li Q, Wang X, Lang R, et al. Inhibition of aspirin-induced gastrointestinal injury: systematic review and network meta-analysis. Front Pharmacol 2021 Aug 12;12:730681 [FREE Full text] [doi: 10.3389/fphar.2021.730681] [Medline: 34475825]
53. Otani K, Tanigawa T, Watanabe T, Shimada S, Nadatani Y, Nagami Y, et al. Microbiota plays a key role in non-steroidal anti-inflammatory drug-induced small intestinal damage. Digestion 2017 Jan 5;95(1):22-28. [doi: 10.1159/000452356] [Medline: 28052268]
54. Bhatt D, Scheiman J, Abraham N, Antman E, Chan F, Furberg C, American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents. ACCF/ACG/AHA 2008 expert consensus document on reducing the gastrointestinal risks of antiplatelet therapy and NSAID use: a report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents. J Am Coll Cardiol 2008 Oct 28;52(18):1502-1517 [FREE Full text] [doi: 10.1016/j.jacc.2008.08.002] [Medline: 19017521]
55. Kou N, Xue M, Yang L, Zang MX, Qu H, Wang MM, et al. Panax quinquefolius saponins combined with dual antiplatelet drug therapy alleviate gastric mucosal injury and thrombogenesis through the COX/PG pathway in a rat model of acute myocardial infarction. PLoS One 2018;13(3):e0194082 [FREE Full text] [doi: 10.1371/journal.pone.0194082] [Medline: 29584740]
56. Cheung K, Leung WK. Gastrointestinal bleeding in patients on novel oral anticoagulants: risk, prevention and management. World J Gastroenterol 2017 Mar 21;23(11):1954-1963 [FREE Full text] [doi: 10.3748/wjg.v23.i11.1954] [Medline: 28373761]

XSL•FO
RenderX

57. Desai J, Kolb J, Weitz J, Aisenberg J. Gastrointestinal bleeding with the new oral anticoagulants--defining the issues and the management strategies. Thromb Haemost 2013 Aug;110(2):205-212. [doi: 10.1160/TH13-02-0150] [Medline: 23702623]
58. Thapa N, Shatzel J, Deloughery TG, Olson SR. Direct oral anticoagulants in gastrointestinal malignancies: is the convenience worth the risk? J Gastrointest Oncol 2019 Aug;10(4):807-809 [FREE Full text] [doi: 10.21037/jgo.2019.02.07] [Medline: 31392062]
59. Blech S, Ebner T, Ludwig-Schwellinger E, Stangier J, Roth W. The metabolism and disposition of the oral direct thrombin inhibitor, dabigatran, in humans. Drug Metab Dispos 2008 Feb;36(2):386-399. [doi: 10.1124/dmd.107.019083] [Medline: 18006647]
60. Ge S, Chen J, Wang W, Zhang L, Teng Y, Yang C, et al. Predicting who has delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage using machine learning approach: a multicenter, retrospective cohort study. BMC Neurol 2024 May 27;24(1):177 [FREE Full text] [doi: 10.1186/s12883-024-03630-2] [Medline: 38802769]
61. Fortescue EB, Kahn K, Bates DW. Development and validation of a clinical prediction rule for major adverse outcomes in coronary bypass grafting. Am J Cardiol 2001 Dec 01;88(11):1251-1258. [doi: 10.1016/s0002-9149(01)02086-0] [Medline: 11728352]
62. Yau TM, Fedak PW, Weisel RD, Teng C, Ivanov J. Predictors of operative risk for coronary bypass operations in patients with left ventricular dysfunction. J Thorac Cardiovasc Surg 1999 Dec;118(6):1006-1013. [doi: 10.1016/S0022-5223(99)70094-2] [Medline: 10595971]
63. Polsinelli VB, Sinha A, Shah SJ. Visceral congestion in heart failure: right ventricular dysfunction, splanchnic hemodynamics, and the intestinal microenvironment. Curr Heart Fail Rep 2017 Dec 26;14(6):519-528 [FREE Full text] [doi: 10.1007/s11897-017-0370-8] [Medline: 29075956]
64. Bhattacharyya A, Chattopadhyay R, Mitra S, Crowe SE. Oxidative stress: an essential factor in the pathogenesis of gastrointestinal mucosal diseases. Physiol Rev 2014 Apr;94(2):329-354 [FREE Full text] [doi: 10.1152/physrev.00040.2012] [Medline: 24692350]
65. Lakatta EG, Levy D. Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises: Part I: aging arteries: a "set up" for vascular disease. Circulation 2003 Jan 07;107(1):139-146. [doi: 10.1161/01.cir.0000048892.83521.58] [Medline: 12515756]
66. Tatge L, Solano Fonseca R, Douglas PM. A framework for intestinal barrier dysfunction in aging. Nat Aging 2023 Oct 18;3(10):1172-1174. [doi: 10.1038/s43587-023-00492-0] [Medline: 37723210]
67. Rodriguez F, Nguyen T, Galanko J, Morton J. Gastrointestinal complications after coronary artery bypass grafting: a national study of morbidity and mortality predictors. J Am Coll Surg 2007 Dec;205(6):741-747. [doi: 10.1016/j.jamcollsurg.2007.07.003] [Medline: 18035256]
68. Karkouti K, Cohen MM, McCluskey SA, Sher GD. A multivariable model for predicting the need for blood transfusion in patients undergoing first-time elective coronary bypass graft surgery. Transfusion 2001 Oct 24;41(10):1193-1203. [doi: 10.1046/j.1537-2995.2001.41101193.x] [Medline: 11606816]
69. Karkouti K, McCluskey S, Syed S, Pazaratz C, Poonawala H, Crowther MA. The influence of perioperative coagulation status on postoperative blood loss in complex cardiac surgery: a prospective observational study. Anesth Analg 2010 Jun 01;110(6):1533-1540. [doi: 10.1213/ANE.0b013e3181db7991] [Medline: 20435945]
70. Li X, Wang R, Sun D, Yao Y, Wang T, Luo G, et al. Risk factors for hypocoagulability after cardiac surgery: a retrospective study. Clin Appl Thromb Hemost 2023 Nov 06;29:10760296231209927 [FREE Full text] [doi: 10.1177/10760296231209927] [Medline: 37933155]
71. Bartoszko J, Wijeysundera D, Karkouti K, Callum J, Rao V, Crowther M, Transfusion Avoidance in Cardiac Surgery Study Investigators. Comparison of two major perioperative bleeding scores for cardiac surgery trials: universal definition of perioperative bleeding in cardiac surgery and European coronary artery bypass grafting bleeding severity grade. Anesthesiology 2018 Dec;129(6):1092-1100. [doi: 10.1097/ALN.0000000000002179] [Medline: 29543629]
72. Ozyilmaz I, Öztürk E, Ozalp S, Recep BZT, Tanıdır, Hatemi AC. Assessment of the frequency and risk factors of gastrointestinal bleeding after cardiopulmonary bypass in paediatric cases. Cardiol Young 2024 Dec;34(12):2557-2561. [doi: 10.1017/S1047951124026106] [Medline: 39370982]
73. Wu S, Lv M, Ma F, Feilong Z, Fang G, Zhang J. A new model (Alfalfa-Warfarin-GIB) for predicting the risk of major gastrointestinal bleeding in warfarin patients. Eur J Clin Pharmacol 2023 Sep 01;79(9):1195-1204. [doi: 10.1007/s00228-023-03533-0] [Medline: 37392366]
74. Xu R, Hao M, Zhou W, Liu M, Wei Y, Xu J, et al. Preoperative hypoalbuminemia in patients undergoing cardiac surgery: a meta-analysis. Surg Today 2023 Aug 07;53(8):861-872. [doi: 10.1007/s00595-022-02566-9] [Medline: 35933630]
75. de la Cruz KI, Bakaeen FG, Wang XL, Huh J, LeMaire SA, Coselli JS, et al. Hypoalbuminemia and long-term survival after coronary artery bypass: a propensity score analysis. Ann Thorac Surg 2011 Mar;91(3):671-675. [doi: 10.1016/j.athoracsur.2010.09.004] [Medline: 21352977]
76. Aizenshtein A, Kachel E, Liza G, Hijazi B, Blum A. Effects of preoperative WBC count on post-CABG surgery clinical outcome. South Med J 2020 Jun;113(6):305-310. [doi: 10.14423/SMJ.0000000000001100] [Medline: 32483641]
77. Newall N, Grayson AD, Oo AY, Palmer ND, Dihmis WC, Rashid A, et al. Preoperative white blood cell count is independently associated with higher perioperative cardiac enzyme release and increased 1-year mortality after coronary

XSL•FO
RenderX

artery bypass grafting. Ann Thorac Surg 2006 Feb;81(2):583-589. [doi: 10.1016/j.athoracsur.2005.08.051] [Medline: 16427856]
78. Aggarwal N, Singh A, Garcia P, Guha S. Ethical implications of artificial intelligence in gastroenterology. Clin Gastroenterol Hepatol 2024 Apr;22(4):689-692. [doi: 10.1016/j.cgh.2024.01.017] [Medline: 38278198]
79. Na L, Yang C, Lo C, Zhao F, Fukuoka Y, Aswani A. Feasibility of reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning. JAMA Netw Open 2018 Dec 07;1(8):e186040 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.6040] [Medline: 30646312]
80. Xiang D, Cai W. Privacy protection and secondary use of health data: strategies and methods. Biomed Res Int 2021;2021:6967166 [FREE Full text] [doi: 10.1155/2021/6967166] [Medline: 34660798]
81. van Delden JJM, van der Graaf R. Revised CIOMS international ethical guidelines for health-related research involving humans. JAMA 2017 Jan 10;317(2):135-136. [doi: 10.1001/jama.2016.18977] [Medline: 27923072]
82. Rashid D, Hirani R, Khessib S, Ali N, Etienne M. Unveiling biases of artificial intelligence in healthcare: navigating the promise and pitfalls. Injury 2024 Mar;55(3):111358. [doi: 10.1016/j.injury.2024.111358] [Medline: 38246015]
83. Acerbi A, Stubbersfield JM. Large language models show human-like content biases in transmission chain experiments. Proc Natl Acad Sci U S A 2023 Oct 31;120(44):e2313790120 [FREE Full text] [doi: 10.1073/pnas.2313790120] [Medline: 37883432]
84. Lin S, Pandit S, Tritsch T, Levy A, Shoja M. What goes in, must come out: generative artificial intelligence does not present algorithmic bias across race and gender in medical residency specialties. Cureus 2024 Feb;16(2):e54448 [FREE Full text] [doi: 10.7759/cureus.54448] [Medline: 38510858]
85. Tyson A, Pasquini G, Spencer A. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Pew Research Center. 2023 Feb. URL: https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/ [accessed 2025-02-01]
86. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
87. Shahian DM, Lippmann RP. Commentary: Machine learning and cardiac surgery risk prediction. J Thorac Cardiovasc Surg 2022 Jun;163(6):2090-2092 [FREE Full text] [doi: 10.1016/j.jtcvs.2020.08.058] [Medline: 32951875]
88. Hui V, Litton E, Edibam C, Geldenhuys A, Hahn R, Larbalestier R, et al. Using machine learning to predict bleeding after cardiac surgery. Eur J Cardiothorac Surg 2023 Dec 01;64(6):297. [doi: 10.1093/ejcts/ezad297] [Medline: 37669153]
89. Gao Y, Liu X, Wang L, Wang S, Yu Y, Ding Y, et al. Machine learning algorithms to predict major bleeding after isolated coronary artery bypass grafting. Front Cardiovasc Med 2022 Jul 28;9:881881 [FREE Full text] [doi: 10.3389/fcvm.2022.881881] [Medline: 35966564]
90. Luo X, Kang Y, Duan S, Yan P, Song G, Zhang N, et al. Machine learning-based prediction of acute kidney injury following pediatric cardiac surgery: model development and validation study. J Med Internet Res 2023 Jan 05;25:e41142 [FREE Full text] [doi: 10.2196/41142] [Medline: 36603200]
91. Song Y, Zhai W, Ma S, Wu Y, Ren M, Van den Eynde J, et al. Machine learning-based prediction of off-pump coronary artery bypass grafting-associated acute kidney injury. J Thorac Dis 2024 Jul 30;16(7):4535-4542 [FREE Full text] [doi: 10.21037/jtd-24-711] [Medline: 39144311]
92. Yang T, Yang H, Liu Y, Liu X, Ding Y, Li R, et al. Postoperative delirium prediction after cardiac surgery using machine learning models. Comput Biol Med 2024 Feb;169:107818. [doi: 10.1016/j.compbiomed.2023.107818] [Medline: 38134752]
93. Mahajan A, Esper S, Oo TH, McKibben J, Garver M, Artman J, et al. Development and validation of a machine learning model to identify patients before surgery at high risk for postoperative adverse events. JAMA Netw Open 2023 Jul 03;6(7):e2322285 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.22285] [Medline: 37418262]
94. Tong C, Du X, Chen Y, Zhang K, Shan M, Shen Z, et al. Machine learning prediction model of major adverse outcomes after pediatric congenital heart surgery: a retrospective cohort study. Int J Surg 2024 Apr 01;110(4):2207-2216 [FREE Full text] [doi: 10.1097/JS9.0000000000001112] [Medline: 38265429]
95. Castela Forte J, Yeshmagambetova G, van der Grinten ML, Scheeren TWL, Nijsten MWN, Mariani MA, et al. Comparison of machine learning models including preoperative, intraoperative, and postoperative data and mortality after cardiac surgery. JAMA Netw Open 2022 Oct 03;5(10):e2237970 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.37970] [Medline: 36287565]
96. Xue L, He S, Singla R, Qin Q, Ding Y, Liu L, et al. Machine learning guided prediction of warfarin blood levels for personalized medicine based on clinical longitudinal data from cardiac surgery patients: a prospective observational study. Int J Surg 2024 Oct 01;110(10):6528-6540. [doi: 10.1097/JS9.0000000000001734] [Medline: 38833337]
97. PhysioNet. URL: https://mimic.physionet.org/ [accessed 2025-02-24]
98. GitHub. URL: https://github.com/Dongjiale1996/gastrointestinal-bleeding-after-coronary-artery-bypass-grafting [accessed 2025-02-24]

Abbreviations

AUROC: area under the receiver operating characteristic curve

XSL•FO
RenderX

CABG: coronary artery bypass grafting
COX: cyclooxygenase
DAPT: dual antiplatelet therapy
DCA: decision curve analysis
GIB: gastrointestinal bleeding
GIBCG: gastrointestinal bleeding after coronary artery bypass grafting
ICU: intensive care unit
INR: international normalized ratio
K-Best: K-best feature selection
LASSO: least absolute shrinkage and selection operator
MIMIC-IV: Medical Information Mart for Intensive Care IV
ML: machine learning
OACs: oral anticoagulants
OR: odds ratio
PPIs: proton pump inhibitors
PRECISE-DAPT: predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy
RFE: random forest recursive feature elimination
SAPT: single antiplatelet therapy
SHAP: Shapley Additive Explanations
SOFA: Sequential Organ Failure Assessment
XGBoost: Extreme Gradient Boosting

Edited by T de Azevedo Cardoso; submitted 07.11.24; peer-reviewed by V Gejjegondanahalli Yogeshappa, L Tan; comments to author 09.12.24; revised version received 04.02.25; accepted 13.02.25; published 06.03.25.

Please cite as:
Dong J, Jin Z, Li C, Yang J, Jiang Y, Li Z, Chen C, Zhang B, Ye Z, Hu Y, Ma J, Li P, Li Y, Wang D, Ji Z
Machine Learning Models With Prognostic Implications for Predicting Gastrointestinal Bleeding After Coronary Artery Bypass Grafting and Guiding Personalized Medicine: Multicenter Cohort Study
J Med Internet Res 2025;27:e68509
URL: https://www.jmir.org/2025/1/e68509
doi:10.2196/68509
PMID:40053791

©Jiale Dong, Zhechuan Jin, Chengxiang Li, Jian Yang, Yi Jiang, Zeqian Li, Cheng Chen, Bo Zhang, Zhaofei Ye, Yang Hu, Jianguo Ma, Ping Li, Yulin Li, Dongjin Wang, Zhili Ji. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Oncology Provider and Patient Perspectives on a Cardiovascular Health Assessment Tool Used During Posttreatment Survivorship Care in Community Oncology (Results from WF-1804CD): Mixed Methods Observational Study

Chandylen L Nightingale[1], MPH, PhD; Emily V Dressler[2], PhD; Maura Kepper[3], MPH, PhD; Heidi D Klepin[4], MD; Simon Craddock Lee[5], MPH, PhD; Sydney Smith[2], MS; Aylin Aguilar[6], BS; Kimberly D Wiseman[6], MS; Stephanie J Sohl[1], PhD; Brian J Wells[2], MD, PhD; Joseph A DeMari[7], MD; Alyssa Throckmorton[8], MD; Lindsey W Kulbacki[9], PA-C; Jenny Hanna[10], MSN; Randi E Foraker[11*], MA, PhD; Kathryn E Weaver[1*], MPH, PhD

[1]Department of Social Sciences and Health Policy, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[2]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[3]The Prevention Research Center, Washington University in St. Louis, St. Louis, MO, United States

[4]Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[5]Department of Population Health, School of Medicine, University of Kansas, Kansas City, KS, United States

[6]Qualitative and Patient-Reported Outcomes Shared Resource, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[7]Section on Gynecologic Oncology, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[8]Baptist Memorial Health Care/Mid-South Minority Underserved NCORP, Memphis, TN, United States

[9]ThedaCare Regional Cancer Center, Appleton, WI, United States

[10]Mercy Hospital Fort Smith, Fort Smith, AR, United States

[11]General Medical Sciences, Institute for Informatics, Washington University in St. Louis, St. Louis, MO, United States

[*]these authors contributed equally

Corresponding Author:
Chandylen L Nightingale, MPH, PhD
Department of Social Sciences and Health Policy
Wake Forest University School of Medicine
Medical Center Boulevard
Winston-Salem, NC, 27157
United States
Phone: 1 3367131432
Email: cnightin@wakehealth.edu

Abstract

Background: Most survivors of cancer have multiple cardiovascular risk factors, increasing their risk of poor cardiovascular and cancer outcomes. The Automated Heart-Health Assessment (AH-HA) tool is a novel electronic health record clinical decision support tool based on the American Heart Association's Life's Simple 7 cardiovascular health metrics to promote cardiovascular health assessment and discussion in outpatient oncology. Before proceeding to future implementation trials, it is critical to establish the acceptability of the tool among providers and survivors.

Objective: This study aims to assess provider and survivor acceptability of the AH-HA tool and provider training at practices randomized to the AH-HA tool arm within WF-1804CD.

Methods: Providers (physicians, nurse practitioners, and physician assistants) completed a survey to assess the acceptability of the AH-HA training, immediately following training. Providers also completed surveys to assess AH-HA tool acceptability and potential sustainability. Tool acceptability was assessed after 30 patients were enrolled at the practice with both a survey developed for the study as well as with domains from the Unified Theory of Acceptance and Use of Technology survey (performance expectancy, effort expectancy, attitude toward using technology, and facilitating conditions). Semistructured interviews at the end of the study captured additional provider perceptions of the AH-HA tool. Posttreatment survivors (breast, prostate, colorectal,

XSL•FO
RenderX

endometrial, and lymphomas) completed a survey to assess the acceptability of the AH-HA tool immediately after the designated study appointment.

Results: Providers (n=15) reported high overall acceptability of the AH-HA training (mean 5.8, SD 1.0) and tool (mean 5.5, SD 1.4); provider acceptability was also supported by the Unified Theory of Acceptance and Use of Technology scores (eg, effort expectancy: mean 5.6, SD 1.5). Qualitative data also supported provider acceptability of different aspects of the AH-HA tool (eg, "It helps focus the conversation and give the patient a visual of continuum of progress"). Providers were more favorable about using the AH-HA tool for posttreatment survivorship care. Enrolled survivors (n=245) were an average of 4.4 (SD 3.7) years posttreatment. Most survivors reported that they strongly agreed or agreed that they liked the AH-HA tool (n=231, 94.3%). A larger proportion of survivors with high health literacy strongly agreed or agreed that it was helpful to see their heart health score (n=161, 98.2%) compared to survivors with lower health literacy scores (n=68, 89.5%; *P*=.005).

Conclusions: Quantitative surveys and qualitative interview data both demonstrate high acceptability of the AH-HA tool among both providers and survivors. Although most survivors found it helpful to see their heart health score, there may be room for improving communication with survivors who have lower health literacy.

Trial Registration: ClinicalTrials.gov NCT03935282; http://clinicaltrials.gov/ct2/show/NCT03935282

International Registered Report Identifier (IRRID): RR2-https://doi-org.wake.idm.oclc.org/10.1016/j.conctc.2021.100808

(J Med Internet Res 2025;27:e65152) doi:10.2196/65152

KEYWORDS

cancer; cardiovascular health; cancer survivors; community oncology; electronic health record integration; provider acceptability; patient-provider; assessment tool; electronic health records; clinical decision support; surveys; interviews; survivors; Automated Heart-Health Assessment

Introduction

Survivors of many common early-stage cancers are now more likely to die of cardiovascular disease than cancer, elevating the importance of addressing cardiovascular health (CVH) in routine survivorship care [1-6]. Over 90% of survivors have multiple cardiovascular risk factors [7], increasing their risk of both poor cardiovascular and cancer outcomes [8-16]. Compared to the general population, survivors of cancer have poorer CVH [17,18]. Over 85% of survivors do not meet the American Heart Association's healthy standards in multiple CVH components (BMI, physical activity, diet, smoking, blood pressure, cholesterol, and glucose) [7,19], many of which increase the risk for both cardiovascular disease and cancer [8,20]. Accordingly, better CVH among survivors is associated with improved survival [21] and reduced risk of both cardiovascular disease [20,22,23] and cancer recurrence [12-14].

Despite Institute of Medicine recommendations for prevention efforts and care coordination for survivors of cancer [24,25], up to 20% of survivors of breast and colorectal cancers may not see a primary care provider [26,27], heightening their risk for lack of preventive services and poor comorbidity management [27-29]. Claims data reveal that only 31%-39% of survivors of breast cancer received cholesterol screening, significantly fewer than women without breast cancer matched on age, ethnicity, sex, region, and comorbidity [29]. Together, these findings emphasize the importance of addressing CVH during routine oncology survivorship care. Both the American Society of Clinical Oncology [30] and the National Comprehensive Cancer Network guidelines [31] recommend cardiovascular risk assessment and discussion for patients with cancer. In our prior work [32] with 20 oncologists, 95% (n=19) reported CVH discussions to be "somewhat" or "very" important; however, 58% only "rarely" or "sometimes" discuss CVH with their patients [33]. Further, nearly 35% of survivors of cancer do not receive assistance from a health care provider for CVH-related lifestyle changes [2]. Similarly, fewer survivors who are at increased risk for health complications report provider discussions about CVH-related lifestyle behaviors (ie, physical activity, diet, and smoking) compared to those with no cancer history [34].

To address these gaps in posttreatment survivorship care and promote guideline adherence, our team developed and deployed a novel, easy-to-use, electronic health record (EHR)–embedded CVH assessment tool, the Automated Heart-Health Assessment (AH-HA) tool. This tool was first implemented in primary care and now incorporates EHR data on receipt of cancer treatments with cardiotoxic potential alongside a visual, interactive display of CVH risk factors, automatically populated from the EHR [35-37]. Before proceeding to future implementation trials, it is critical to establish the acceptability of the tool among oncology providers and survivors [38]. As part of a larger pragmatic trial to test and evaluate AH-HA in survivorship care [39], among practices randomized to the AH-HA tool, we assessed the acceptability of the AH-HA tool among both patients and providers during routine oncology care, along with provider perceptions of potential sustainability.

Methods

Ethical Considerations

This study (WF-1804CD) was approved by the National Cancer Institute (NCI) Central institutional review board (IRB). Each participating institution granted authority to the NCI Central IRB to serve as the IRB of record for NCI Community Oncology Research Program (NCORP) studies, in accordance with the National Institute of Health's single IRB policy. All participants provided consent. NCORP is a national network of community oncology practices with infrastructure to support the recruitment

XSL•FO
RenderX

of patients to clinical trials [40]. This study was facilitated through the Wake Forest NCORP Research Base (UG1CA189824). Study data were de-identified. Providers were offered a $10 gift card upon completion of the posttraining survey and a $20 gift card for participating in the qualitative interview. Survivors received a $10 gift card upon completion of the acceptability survey.

Study Eligibility and Recruitment Procedures

Weaver et al [39] show the complete eligibility criteria and methods for the larger randomized trial. NCORP practice eligibility criteria included (1) use of the Epic EHR, (2) willingness to incorporate the AH-HA tool in their EHR, (3) having two or more providers willing to be trained and use AH-HA, and (4) identified combined providers saw 100 or more potentially eligible patients for follow-up in prior 6 months. Providers were recruited and consented by cancer care delivery research leads within their practice. Eligible providers included physicians and advanced practice providers (nurse practitioners and physician assistants) willing to complete the AH-HA provider training. This manuscript focuses on providers within practices randomized to use the AH-HA tool in the pragmatic trial [39]. To identify eligible survivors, staff at NCORP sites screened clinic schedules and reviewed survivors' medical records. Survivors were contacted by phone, patient portal, or in-person and were eligible if they were at least 6 months post potentially curative cancer treatment for breast, prostate, colorectal, or endometrial cancers or Hodgkin and non-Hodgkin lymphomas and scheduled for a routine cancer-related follow-up care visit.

AH-HA Training and Intervention

A full description of the AH-HA tool and provider training is available in the protocol paper [41]. In brief, providers completed two 30-minute video trainings prior to the practice enrolling patients. The training covered (1) the importance of addressing CVH as part of routine posttreatment follow-up care for survivors of cancer, (2) the basics of the American Heart Association's Life's Simple 7 CVH factors [9] and overall CVH metric, (3) navigation of the AH-HA tool within the EHR, and (4) how to use the tool to guide discussions with survivors. The AH-HA tool was launched using a best practice alert for enrolled patients during a routine posttreatment outpatient oncology visit. Providers could choose to use the tool or not in accordance with their clinical judgment; examples of reasons for nonuse may include a competing clinical demand (eg, new symptom or concern for recurrence), patient distress, or perception that the patient would not be receptive to or benefit from a discussion (eg, in the unlikely case that all factors were ideal). Five of the CVH factors were automatically populated from the EHR when available (BMI, smoking, blood pressure, cholesterol, and hemoglobin A_{1c} or blood glucose); physical activity and diet data were collected on paper and entered directly into the tool by the provider. AH-HA color codes each CVH factor as red (poor), yellow (intermediate), or green (ideal) according to Life's Simple 7 classification framework [9] and also provides a total CVH score. Interactive slider bars can be used to demonstrate how improvements in CVH factors can lead to shifts in the categorization and overall CVH score. A second tab included information about the patient's receipt of cancer treatments with cardiotoxic potential (ie, anthracyclines, antimetabolites, hormone therapy, aromatase inhibitors, monoclonal antibodies, antimicrotubule agents, alkylating agents, and radiation) [3,42,43].

Data Collection and Measures

Providers and survivors provided information about sex, age, race, and ethnicity. Survivor cancer type and time since diagnosis were abstracted from the EHR. Survivor's health literacy was also assessed with 1 item ("How confident are you filling out medical forms by yourself?") with response options ranging from not at all confident to extremely confident. Prior research has demonstrated that this 1 item is effective at identifying health literacy skills [44]. Provider items also included provider type (physician, nurse practitioner, and physician assistant), years in current position, time spent providing direct patient care, time spent using the EHR for direct individual patient care, and proficiency with current EHR. Providers completed 2 surveys: one immediately after participating in the initial AH-HA training (posttraining survey) before participant enrollment and one after 30 patients were enrolled at the practice (postenrollment survey). Provider surveys assessed the acceptability of the training and AH-HA tool, and preferences for when and how often to use the AH-HA tool in the cancer treatment trajectory. Survivors completed one survey to assess the acceptability of the AH-HA tool immediately after the designated routine oncology appointment. Data collection occurred from December 2020 to March 2023. Survey items developed by our team are available upon reasonable request.

Provider Perspectives on Training

We developed a 7-item survey for the purpose of this study to assess various aspects of acceptability (eg, "The AH-HA training provided useful information about the importance of addressing CVH with cancer survivors" and "The AH-HA provider training will help me be more effective when discussing cardiovascular health with survivors") with response options ranging from 1=strongly disagree to 7=strongly agree. A composite score was calculated using the average of all 7 items. Also, 1 item assessed the acceptability of the AH-HA training duration (response options ranging from too short to too long). An additional item assessed comfort in discussing CVH with survivors following the training ("Please indicate your level of comfort discussing CVH with your posttreatment, good prognosis patients") with a 5-point Likert-scale (not at all comfortable to very comfortable). In a separate follow-up survey, providers were asked 1 item retrospectively about their preparedness to use the AH-HA tool at the time they completed the training ("Following the provider training, how prepared were you to use the AH-HA tool with patients?"). Response options included not at all prepared, somewhat prepared, and very prepared.

Provider Perspectives on the AH-HA Tool

Six items, used in prior work [32,36], assessed aspects of provider acceptability of the AH-HA tool (eg, "The information AH-HA provides is useful" and "AH-HA helps me be more

XSL•FO
RenderX

effective") with response options ranging from 1=strongly disagree to 7=strongly agree. A composite score was calculated using the average of all 6 items. We further assessed acceptability using items from the Unified Theory of Acceptance and Use of Technology (UTAUT) survey [45]. Specifically, 15 items assessed the performance expectancy (eg, "The AH-HA tool is useful in my job"), Effort expectancy (eg, "I find the AH-HA tool easy to use"), attitude toward using technology ("Using the AH-HA tool is a good idea"), and facilitating conditions (eg, "I have the resources necessary to use the AH-HA tool") domains of the UTAUT survey [5,45]. Response options ranged from 1=strongly disagree to 7=strongly agree. We calculated scores for each domain using the average of domain items [45]. An additional item that was developed for this study to assess potential sustainability asked providers to report the timing and frequency they would like to use the AH-HA tool ("After the study ends, how often would you like to use the AH-HA Tool when providing care to patients during: (1) initial treatment planning, (2) active treatment, and (3) posttreatment survivorship care") with response options including never or almost never, seldom or about half the time, most of the time, and always or almost always.

Survivor Acceptability

Five items previously used in our pilot work assessed overall acceptability of the AH-HA tool (eg, "I liked the heart health tool I used today with my provider" and "It was helpful to see my heart health score") with a 5-point Likert scale ranging from strongly disagree to strongly agree [32].

Providers also participated in a semistructured qualitative interview conducted via telephone at the end of patient enrollment at their practice to further understand perceptions of the AH-HA tool. Examples of interview content include the impact of AH-HA on the provider's practice ("How do you think having access to the tool impacted your practice, if at all?"), patients' responses to the tool ("How did patients respond or react to the tool?"), recommended changes to the tool ("What changes would you make to AH-HA so it will work effectively in your setting?"), impact on care provided to patients ("Overall, do you feel the tool helped improve the care you provide to patients? Why/why not?"), and benefits and drawbacks to continuing to use the tool ("What benefits and drawbacks do you see in continuing to use AH-HA in your practice after the study is complete?"). A full list of interview questions is available upon request.

Interviews were conducted by 2 trained qualitative research team members from the Qualitative and Patient-Reported Outcomes (Q-PRO) Shared Resource of the Atrium Health Wake Forest Baptism Comprehensive Cancer Center. Interviews lasted an average of 20 minutes and were audio recorded.

Analyses

Descriptive statistics were quantified with mean (SD) and frequency (%) for continuous and categorical outcomes respectively. Figures display mean and corresponding 95% CIs for providers' answers on a 1-7 scale. Total scores for scales are quantified with mean (SD) and range. Univariate associations of demographics characteristics (age, sex, race, ethnicity, and health literacy) and cancer type (breast, colorectal, prostate, endometrial, and lymphoma) with acceptance of the AH-HA tool (using the following items: (1) "It was helpful to see my heart health score" and (2) "I would like to use this tool to talk about my heart health with my oncology provider at a future appointment") were tested using Fisher exact tests. *P* values less than .05 were considered statistically significant.

Qualitative interviews were analyzed in collaboration with the Q-PRO Shared Resource of the Atrium Health Wake Forest Baptism Comprehensive Cancer Center. The interview audio was transcribed verbatim and 2 Q-PRO teammates and coauthors (AA and KW) reviewed the transcripts and developed a draft codebook. The study team reviewed the codebook and provided input, which was incorporated into a new version of the codebook. Transcripts were imported into ATLAS.ti [46] and the codebook was tested by coding several transcripts and revised as necessary. All interviews were independently coded by 2 Q-PRO teammates and coauthors (AA and KW) and compared; any discrepancies were discussed and resolved. Once all transcripts were coded, code reports were run and summaries for each report were written. Summaries for provider's perceptions of the AH-HA tool were synthesized and analyzed for patterns and themes.

Results

Overview

In total, 17 providers were recruited for the pragmatic trial to participate in the intervention arm; 1 provider did not use the AH-HA tool and 1 provider did not complete surveys. Thus, we report results on the 15 providers who used the AH-HA tool and completed the surveys from 4 community oncology practice groups (25% of practices located in the Midwest, 75% or practices located in the South; 25% minority or underserved NCORP, and 50% designated critical access hospital). A total of 13 providers (87%) completed the posttraining survey and 15 (100%) completed the postenrollment survey. Among the 15 providers who used the tool, together they saw 296 survivors (46% of survivors participating in the larger randomized trial). Of these, 245 reported seeing the AH-HA tool (33 did not see the tool and 18 were unknown).

Provider and Survivor Characteristics

Overview

Providers included physicians (n=8, 53%; Table 1), nurse practitioners (n=6, 40%), and a physician assistant (n=1, 7%). Most providers (n=10, 67%) reported spending 76%-100% of their time providing direct patient care and more than half (n=8, 53%) reported spending 76%-100% of their time using the EHR for direct patient care. Most providers (n=11, 73%) reported that they were "very proficient" with their current EHR.

XSL•FO
RenderX

Table 1. Provider characteristics (N=15).

Characteristics	Values
Sex, n (%)	
Female	11 (73)
Male	4 (27)
Age (years), n (%)	
26-35	2 (13)
36-45	6 (40)
46-55	3 (20)
65 and older	1 (7)
Unknown	3 (20)
Race, n (%)	
Asian	3 (20)
White or Caucasian	11 (73)
Not reported	1 (7)
Ethnicity, n (%)	
Non-Hispanic	13 (87)
Not reported	2 (13)
Provider role, n (%)	
Physician	8 (53)
Nurse practitioner	6 (40)
Physician assistant	1 (7)
Years in current position, n (%)	
1-5	5 (33)
6-10	5 (33)
11-20	4 (27)
More than 20	1 (7)
Time spent providing direct patient care (%), n (%)	
51-75	5 (33)
76-100	10 (67)
Time spent using EHR[a] for direct individual patient care (%), n (%)	
26-50	3 (20)
51-75	4 (27)
76-100	8 (53)
Proficiency with current EHR, n (%)	
Very proficient	11 (73)
Somewhat proficient	3 (20)
Neutral	1 (7)
Number of survivors that used AH-HA[b] per provider, mean (SD); range	19.7 (17.1); 2-56

[a]EHR: electronic health record.

[b]AH-HA: Automated Heart-Health Assessment.

Survivors (N=245; Table 2) completed treatment for breast (n=230, 93.9%), endometrial (n=1, 0.4%), or colorectal (n=9, 3.7%) cancers, or lymphoma (n=5, 2%) and were mostly female (n=239, 97.5%). Most survivors were White or Caucasian (n=203, 82.9%) and 13.1% (n=32) were Black or African American. Overall, 5.3% were Hispanic or Latino (n=13).

XSL•FO
RenderX

Survivors were an average age of 61 (SD 10.9) years and most commonly married or living as married (n=176, 71.8%). Most survivors had a college degree (n=111, 45.3%) or some college including vocational or technical school (n=82, 33.5%). The median time since diagnosis was 3.6 (IQR 2.1-5.2) years.

Table 2. Survivor characteristics (N=245).

Characteristics	Values
Sex, n (%)	
Female	239 (97.5)
Male	6 (2.5)
Age (years), n (%)	
18-39	9 (3.7)
40-64	131 (53.5)
65-74	84 (34.3)
75 and older	21 (8.6)
Race, n (%)	
American Indian or Alaskan Native	1 (0.4)
Asian	3 (1.2)
Black or African American	32 (13.1)
White, non-Hispanic or Latino	193 (78.8)
White or other or unknown, Hispanic or Latino	13 (5.3)
More than 1 race, not Hispanic or Latino	2 (0.8)
Other or unknown, not Hispanic or Latino	1 (0.4)
Ethnicity, n (%)	
Hispanic or Latino	13 (5.3)
Not Hispanic or Latino	232 (94.7)
Marital status, n (%)	
Married or living as married	176 (71.8)
Single, divorced, separated, or widowed	69 (28.2)
Education, n (%)	
High school or less	52 (21.2)
Some colleges (including vocational or technical)	82 (33.5)
College degree or more	111 (45.3)
Cancer type, n (%)	
Breast	230 (93.9)
Colorectal	9 (3.7)
Endometrial	1 (0.4)
Lymphoma	5 (2.0)
Time since diagnosis (years)	
Median (IQR)	3.61 (2.14-5.22)
Unknown (n=6)	N/A[a]

[a]N/A: not applicable.

Provider Perspectives on AH-HA Training

Figure 1 depicts provider training acceptability findings. Overall, providers reported high acceptability (mean 5.8, SD 1.0), with the highest item acceptability rating (mean 6.1, SD 0.8) for the following item: "The AH-HA training provided useful information about the importance of addressing CVH with cancer survivors." Providers reported the lowest acceptability rating (mean 5.5, SD 1.3) for the following item: "I feel prepared to use the AH-HA tool in clinic with posttreatment, good

XSL•FO
RenderX

prognosis patients." More than half of providers reported that the duration of the AH-HA training was "about right" (n=7, 54%), followed by "a little too long" (n=4, 31%), "a little too short" (n=1, 8%), and "much too long" (n=1, 8%). At the conclusion of the training, all providers reported that they were somewhat (n=9, 69%) or very (n=4, 31%) comfortable discussing CVH with posttreatment patients with a good prognosis. When providers reflected on their preparedness after using the AH-HA tool, most reported they were "very prepared" (n=8, 57%) followed by "somewhat prepared" (n=6, 43%).

Figure 1. Provider acceptability of the AH-HA training session (N=13). AH-HA: Automated Heart-Health Assessment; CVH: cardiovascular health.

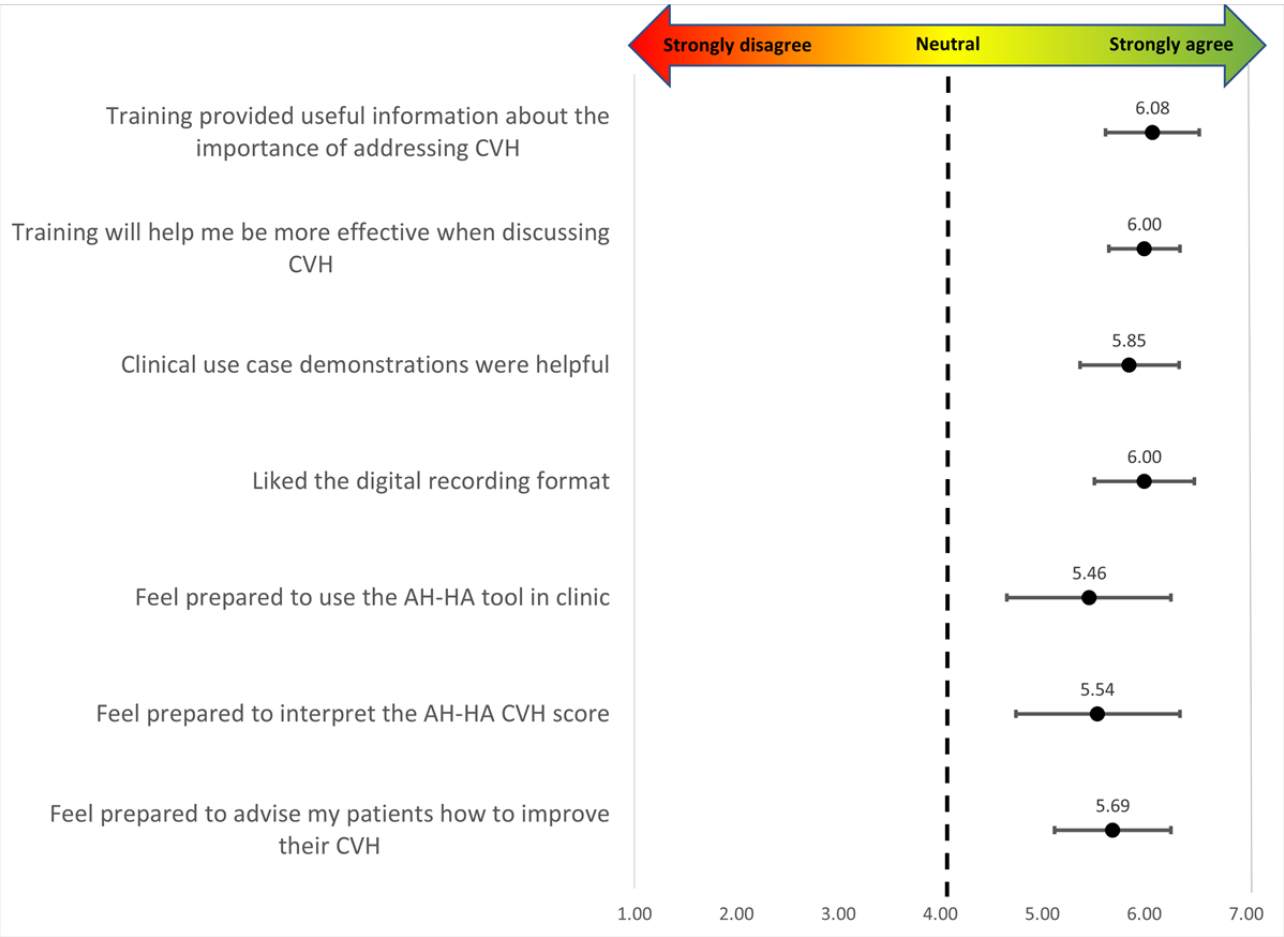

Provider Perspectives of the AH-HA Tool

Quantitative surveys and qualitative interview data converged to demonstrate provider acceptability of the AH-HA tool. Figure 2 shows survey results; providers reported being satisfied with the AH-HA tool (mean 5.5, SD 1.4) with the highest rating for 2 items: "The information AH-HA provides is useful" (mean 5.9, SD 1.2) and "The information in AH-HA is presented in a useful format" (mean 5.9, SD 1.1); and the lowest rating (mean 4.9, SD 1.8) for the following item: "AH-HA makes the information I want easier to access." Providers felt the interactivity and visuals provided in the AH-HA tool were particularly useful for patients. One provider stated, "I think the biggest thing is the visual aspect of the tool is really nice for them and the interactive-ness, the way you can slide the bars and show them if they achieve X, Y, or Z goal, how it can make a difference in their [CVH] score." One provider mentioned that having these data available would allow them to easily track their patients' progress: "it was something that, in a follow up visit, you could look—would be able to look back on to compare and talk with the patient and they can see how they made progress in this area or is there something we can continue to work on. It helps focus the conversation and give the patient a visual of continuum of progress." Providers reported acceptability of AH-HA (Table 3) for the performance expectancy (mean 4.0, SD 2.0), effort expectancy (mean 5.6, SD 1.5), attitude toward using technology (mean 4.8, SD 2.1), and facilitating conditions (mean 5.5, SD 1.5) domains of the UTAUT. Related to performance, providers felt AH-HA helped them have deeper discussions of cardiovascular risk with patients.

XSL•FO
RenderX

Figure 2. Provider acceptability of the AH-HA tool (N=15). AH-HA: Automated Heart-Health Assessment.

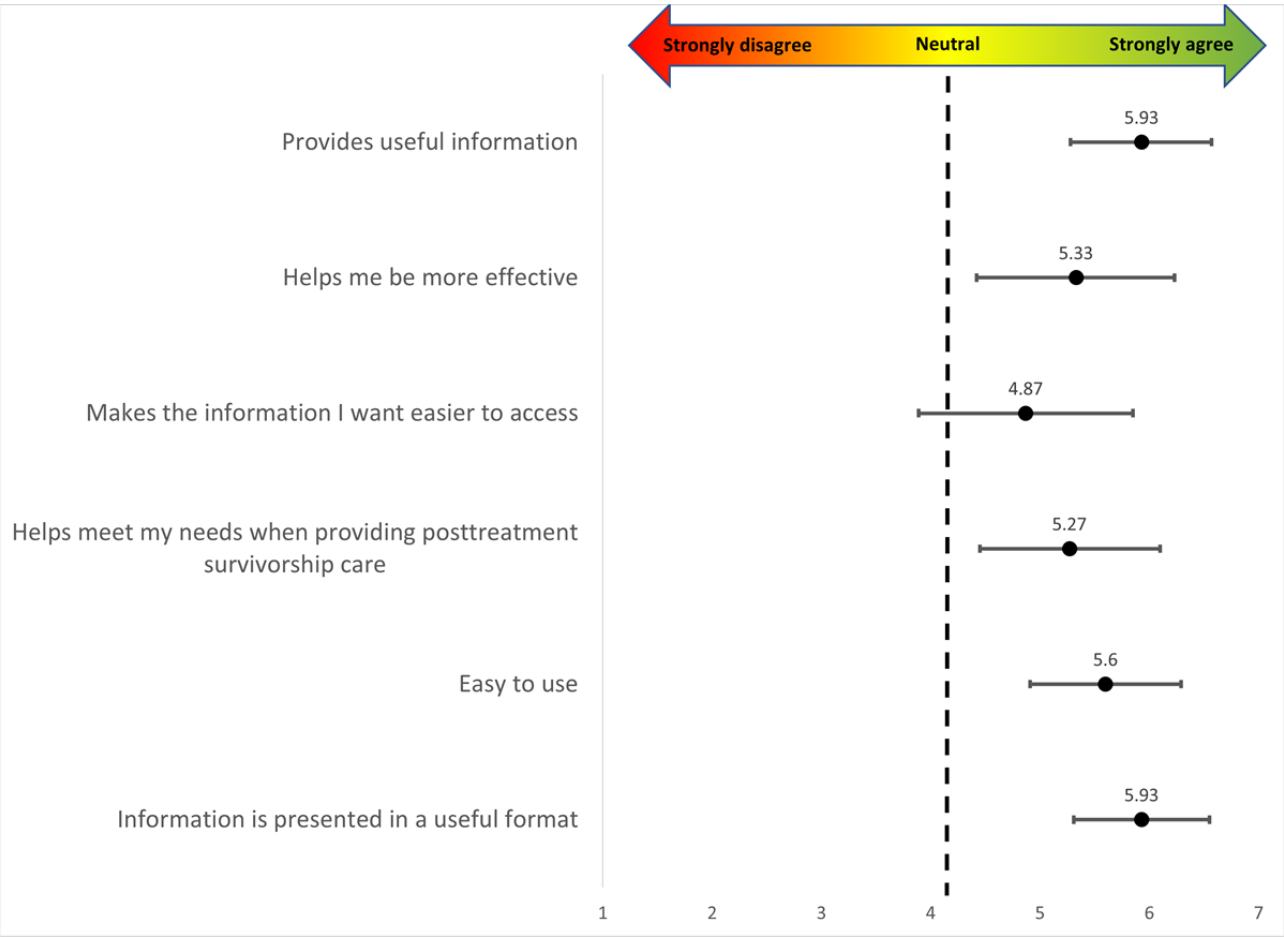

Table 3. Provider acceptability for the performance, effort, attitude, and facilitating conditions domains of the UTAUT[a] survey (N=13).

UTAUT domain	Mean (SD); range
Performance	3.98 (2.04); 1-7
Effort	5.62 (1.49); 2-7
Attitude	4.75 (2.06); 1-7
Facilitating conditions	5.54 (1.51); 2-7

[a]UTAUT: Unified Theory of Acceptance and Use of Technology.

One provider stated, “Before, our CVH approach might have been more of a blanket statement about you are a breast cancer survivor, and you may have increased cardiovascular risks, so you need to optimize your blood pressure, cholesterol with your primary care doctor and what not, but this is a much more thorough tool.”

Some providers noted already having these conversations with patients, which made them less receptive to the tool. For example, 1 provider stated, “because I already do that anyway, I think it was just kind of time-consuming…to actually do it and make extra time in the visit to go through that particular part on the computer and have them ask—or answer very specific questions when we really kind of discuss all of this anyway.” In contrast, favorable effort expectancy was supported qualitatively as some providers noted that the tool was “very simple to use” and “user friendly.” Providers did feel it could be easier to use if it required less “maneuvering” or having to go back and forth” within the EHR.

Potential Sustainability of the AH-HA Tool

When asked about using the tool after the study ended, most providers reported interest in using the AH-HA tool for posttreatment survivorship care (always or almost always: n=3, 21%; most of the time: n=7, 50%; seldom or about half the time: n=2, 14%; and never or almost never: n=2, 14%). There was less interest in using the tool for patients in active treatment or during initial treatment planning for which results were the same (most of the time: n=2, 14%; seldom or about half the time: n=5, 36%; or never or almost never: n=7, 50%).

Survivor Acceptability of the AH-HA Tool

Figure 3 shows results for survivor acceptability of the AH-HA tool. Most survivors reported that they strongly agreed or agreed that they liked the AH-HA tool (n=231, 94.3%), it was helpful

XSL•FO
RenderX

to see their heart health score (n=229, 93.5%), AH-HA was easy to understand (n=228, 93.1%), the picture or diagram (of CVH risk factors) improved their understanding of their heart health (n=204, 83.3%), and they want to use AH-HA to talk about heart health with their oncology provider at a future appointment (n=208, 84.9%).

Figure 3. Survivor acceptability of the AH-HA Tool (N=15). AH-HA: Automated Heart-Health Assessment.

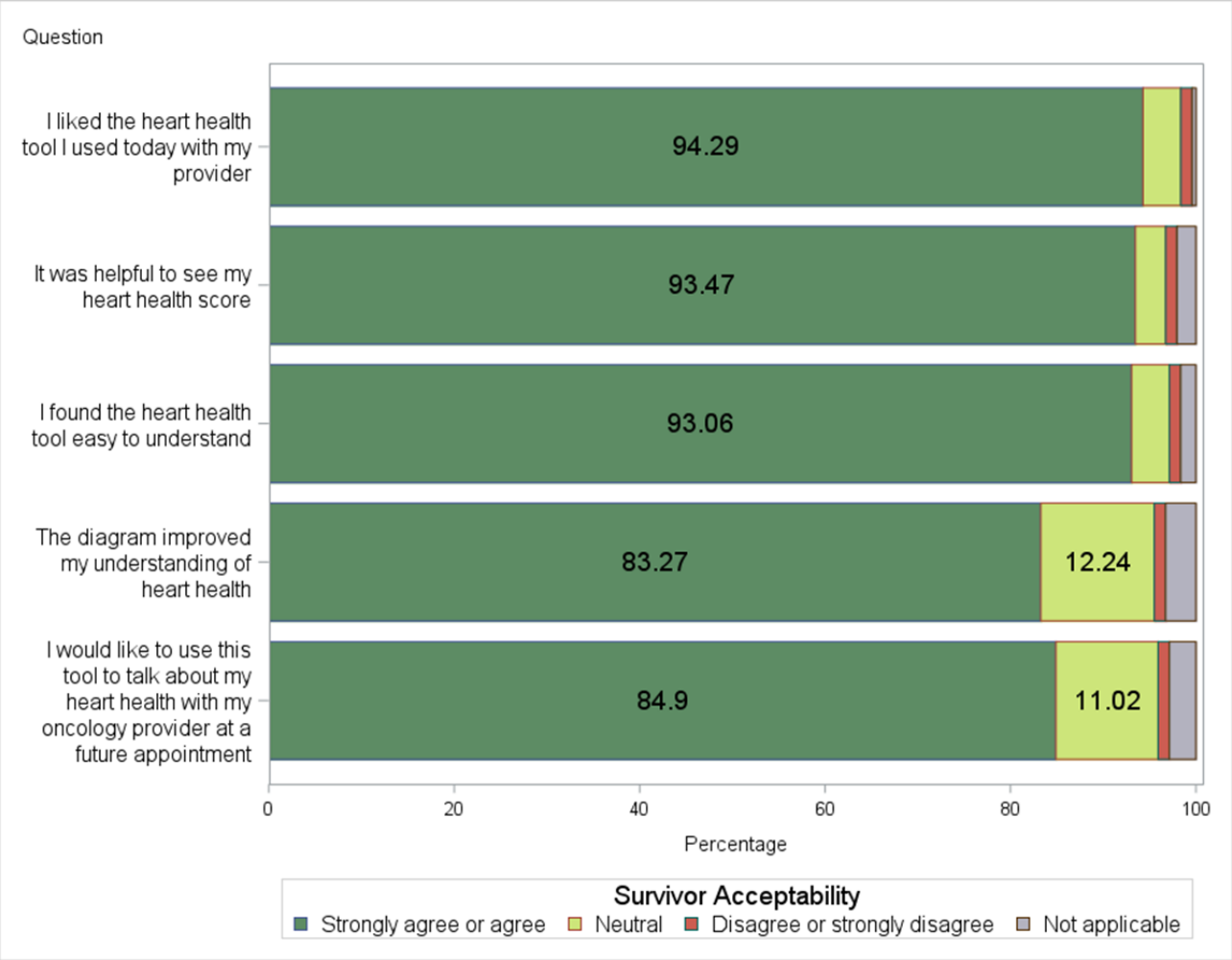

Associations With Survivor Acceptability of the AH-HA Tool

Health literacy was the only survivor demographic characteristic significantly associated with survivors' acceptability of the AH-HA tool. Survivors who indicated they were "extremely confident" filling out medical forms on their own (ie, high health literacy) strongly agreed or agreed that it was helpful to see their heart health score (n=161, 98.2%) compared to survivors with lower health literacy scores (n=68, 89.5%; *P*=.005). Yet, the perceived helpfulness of seeing the heart health score was generally high.

Discussion

Our mixed methods results support the acceptability of the AH-HA CVH assessment tool when used as part of routine posttreatment oncology care in community settings. Survivors of cancer were positive about using the tool in the clinic with their provider. Both oncology physicians and advanced practice providers across 4 community practices reported favorable perceptions of the AH-HA training and use of the tool with survivors of cancer. This suggests AH-HA may be well received in a variety of survivorship care models [47] (eg, advanced practice provider-led survivorship clinics or follow-up with the treating physician).

The overall high acceptability among both patients and providers supports the further implementation of the AH-HA tool, with a continued focus on posttreatment survivors of cancer. Most providers reported they would prefer to use the AH-HA tool for posttreatment survivorship care, and that they would rarely use AH-HA for patients in active treatment or during initial treatment planning. While CVH is important at all points in the cancer treatment trajectory, providers may want to prioritize oncologic treatment during the treatment planning and active treatment phases, and transition to health promotion during the posttreatment survivorship phase. Providers may also perceive that patients are able to more effectively focus on health behavior change without the logistical and psychosocial challenges that are heightened during the treatment phase [48,49]. It is also possible that providers preferred to use the AH-HA tool due to the framing effect of the trial (eg, only posttreatment patients were eligible and the study was focused on survivorship).

XSL•FO
RenderX

Our prior usability assessments considered the tool's appropriateness of the CVH tool for posttreatment survivorship care and we learned that oncology providers also wanted to see potential cardiotoxic treatments received by survivors [33], as incorporated in this study. Moving forward, if the tool is to be used for treatment planning, usability assessments should be repeated with consideration of possible future cardiotoxic effects of treatments incorporated in the design of the tool.

Despite overall positive feedback from both providers and survivors about the AH-HA tool, our results suggest there may be some room for improvement in communicating the heart health score to patients with lower health literacy. Provider training could be augmented to include tips for using the tool with patients who have different levels of health literacy along with scripts to help guide the discussion with patients. Research shows providers often overestimate patient's literacy levels, and patients may be too embarrassed about their limited health literacy to ask questions [50-53]. Increasing the provider's awareness of a survivor's health literacy may be an important step prior to initiating the CVH discussion with survivors. Although we assessed health literacy in this study, this information was not shared with providers. One potential strategy is to include an assessment of survivor health literacy as part of the AH-HA tool to help inform the CVH provider-survivor discussion. Additionally, providing basic information prior to the appointment on the components of CVH and their impact on a patient's overall health may better prepare patients for the upcoming discussion with the provider. Similarly, enhancing patient-facing information in the format of an after-visit summary of CVH recommendations may enhance understanding for patients with lower health literacy [54].

Other potential modifications to the tool may address provider desires for a more streamlined experience. Although most providers found the tool simple to use, others suggested refining AH-HA by requiring less maneuvering within the EHR, and quantitative findings suggested room for improving ease of access to desired information. One way in which the tool could be simplified for providers would be to collect the self-reported diet and physical activity data via the patient portal prior to the visit so that the data would be available in the EHR and callable by the tool. This method would be expected to streamline the use of the tool at the point of care if these data would not need to be manually entered into the tool. Such modifications may also impact providers' perceptions of how AH-HA will impact their job performance and interest in using the AH-HA tool, which corresponds to the UTAUT domains for which providers reported the lowest means.

There were notable strengths to this study. Provider feedback on AH-HA acceptability included both quantitative and qualitative data to provide a fuller picture of both overall acceptability and specific characteristics of the AH-HA tool, consistent with reported strengths of mixed methods research [55]. In this study, these data were complementary and enhanced understanding of provider acceptability. This study was also strengthened by the assessment of perspectives from both providers and survivors as the "end users" from 4 community oncology practices, to inform the next steps, and promote sustainability for AH-HA when implemented widely [56,57]. One limitation of the present study is the predominant enrollment of breast cancer patients despite broad inclusion criteria. This likely reflects the specialization of enrolling providers and the patient mix with respect to cancer type within survivorship programs. Our study team has reported interest in CVH discussions among survivors of gynecologic cancers, yet we acknowledge a more diverse survivor sample is needed to determine the generalizability of these results [58]. Although 15 providers participated and used the AH-HA tool, 2 providers (13%) did not complete the posttraining survey for unknown reasons. Due to the overall high acceptability of AH-HA, there was limited variability in detecting potential differences in patient acceptability by sex, age, or race and ethnicity. Further, although the sample size for our provider key informants was sufficient for theme saturation as our analytic approach [59], it also limited us from making comparisons in acceptability by provider type.

Building upon our strong acceptability findings, the next step for this line of research is to test the AH-HA implementation package to promote guideline-concordant CVH assessment and discussion among a larger and more diverse sample of oncology providers and patients. Tailoring the CVH discussion to meet the needs of patients with higher and lower health literacy will be an important factor to consider in this future direction. It will also be important to assess the sustainability of the AH-HA tool in community practice.

Acknowledgments

The authors would like to acknowledge the following the National Cancer Institute's Community Oncology Research Program (NCORP) sites for their participation: Cancer Research for the Ozarks NCORP, Baptist Memorial Health Care/Mid-South Minority Underserved NCORP, Geisinger Cancer Institute NCORP, Iowa-Wide Oncology Research Coalition NCORP, Wisconsin NCORP, and VCU Massey Cancer Center Minority Underserved NCORP. The authors would also like to thank Wake Forest NCORP Research Base staff members Karen Craver, Jessica Sheedy, William Stanfield, Cheyenne Wagi, and Kimberly Blish. Support for this study was provided by National Cancer Institute (Bethesda, Maryland) grants (R01CA226078, UG1CA189824, and P30CA012197).

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Lajous M, Mozaffarian D, Mozaffarian R, Schrag D, Adami HO. Lifestyle prescriptions for cancer survivors and their communities. J Intern Med 2011;269(1):88-93 [FREE Full text] [doi: 10.1111/j.1365-2796.2010.02273.x] [Medline: 21158981]
2. Patnaik JL, Byers T, DiGuiseppi C, Dabelea D, Denberg TD. Cardiovascular disease competes with breast cancer as the leading cause of death for older females diagnosed with breast cancer: a retrospective cohort study. Breast Cancer Res 2011;13(3):R64 [FREE Full text] [doi: 10.1186/bcr2901] [Medline: 21689398]
3. Shelburne N, Adhikari B, Brell J, Davis M, Desvigne-Nickens P, Freedman A, et al. Cancer treatment–related cardiotoxicity: current state of knowledge and future research priorities. J Natl Cancer Inst 2014;106(9):dju232 [FREE Full text] [doi: 10.1093/jnci/dju232] [Medline: 25210198]
4. van Erning F, van Steenbergen L, Lemmens V, Rutten H, Martijn H, van Spronsen D, et al. Conditional survival for long-term colorectal cancer survivors in the Netherlands: Who do best? Eur J Cancer 2014;50(10):1731-1739. [doi: 10.1016/j.ejca.2014.04.009] [Medline: 24814358]
5. Ward KK, Shah NR, Saenz CC, McHale MT, Alvarez EA, Plaxe SC. Cardiovascular disease is the leading cause of death among endometrial cancer patients. Gynecol Oncol 2012;126(2):176-179. [doi: 10.1016/j.ygyno.2012.04.013] [Medline: 22507532]
6. Stoltzfus KC, Zhang Y, Sturgeon K, Sinoway LI, Trifiletti DM, Chinchilli VM, et al. Fatal heart disease among cancer patients. Nat Commun 2020;11(1):2011 [FREE Full text] [doi: 10.1038/s41467-020-15639-5] [Medline: 32332714]
7. Underwood JM, Townsend JS, Stewart SL, Buchannan N, Ekwueme DU, Hawkins NA, et al. Surveillance of demographic characteristics and health behaviors among adult cancer survivors—Behavioral Risk Factor Surveillance System, United States, 2009. MMWR Surveill Summ 2012;61(1):1-23 [FREE Full text] [Medline: 22258477]
8. Foraker RE, Abdel-Rasoul M, Kuller LH, Jackson RD, Van Horn L, Seguin RA, et al. Cardiovascular health and incident cardiovascular disease and cancer: the women's health initiative. Am J Prev Med 2016;50(2):236-240 [FREE Full text] [doi: 10.1016/j.amepre.2015.07.039] [Medline: 26456876]
9. Folsom AR, Yatsuya H, Nettleton JA, Lutsey PL, Cushman M, Rosamond WD, ARIC Study Investigators. Community prevalence of ideal cardiovascular health, by the American Heart Association definition, and relationship with cardiovascular disease incidence. J Am Coll Cardiol 2011;57(16):1690-1696 [FREE Full text] [doi: 10.1016/j.jacc.2010.11.041] [Medline: 21492767]
10. Rasmussen-Torvik LJ, Shay CM, Abramson JG, Friedrich CA, Nettleton JA, Prizment AE, et al. Ideal cardiovascular health is inversely associated with incident cancer: the Atherosclerosis Risk In Communities study. Circulation 2013;127(12):1270-1275 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.112.001183] [Medline: 23509058]
11. Kulshreshtha A, Vaccarino V, Judd SE, Howard VJ, McClellan WM, Muntner P, et al. Life's Simple 7 and risk of incident stroke: the reasons for geographic and racial differences in stroke study. Stroke 2013;44(7):1909-1914 [FREE Full text] [doi: 10.1161/STROKEAHA.111.000352] [Medline: 23743971]
12. Hurt RD, Ebbert JO, Hays JT, McFadden DD. Preventing lung cancer by treating tobacco dependence. Clin Chest Med 2011;32(4):645-657. [doi: 10.1016/j.ccm.2011.08.004] [Medline: 22054877]
13. Byers T, Sedjo RL. Does intentional weight loss reduce cancer risk? Diabetes Obes Metab 2011;13(12):1063-1072. [doi: 10.1111/j.1463-1326.2011.01464.x] [Medline: 21733057]
14. Thomson CA, Rock CL, Thompson PA, Caan BJ, Cussler E, Flatt SW, et al. Vegetable intake is associated with reduced breast cancer recurrence in tamoxifen users: a secondary analysis from the Women's Healthy Eating and Living Study. Breast Cancer Res Treat 2011;125(2):519-527. [doi: 10.1007/s10549-010-1014-9] [Medline: 20607600]
15. Oeffinger KC, Tonorezos ES. The cancer is over, now what?: Understanding risk, changing outcomes. Cancer 2011;117:2250-2257 [FREE Full text] [doi: 10.1002/cncr.26051] [Medline: 21523742]
16. Davies NJ, Batehup L, Thomas R. The role of diet and physical activity in breast, colorectal, and prostate cancer survivorship: a review of the literature. Br J Cancer 2011;105:S52-S73 [FREE Full text] [doi: 10.1038/bjc.2011.423] [Medline: 22048034]
17. Muhandiramge J, Zalcberg JR, van Londen GJ, Warner ET, Carr PR, Haydon A, et al. Cardiovascular disease in adult cancer survivors: a review of current evidence, strategies for prevention and management, and future directions for cardio-oncology. Curr Oncol Rep 2022;24(11):1579-1592 [FREE Full text] [doi: 10.1007/s11912-022-01309-w] [Medline: 35796941]
18. Armenian SH, Xu L, Ky B, Sun C, Farol LT, Pal SK, et al. Cardiovascular disease among survivors of adult-onset cancer: a community-based retrospective cohort study. J Clin Oncol 2016;34(10):1122-1130 [FREE Full text] [doi: 10.1200/JCO.2015.64.0409] [Medline: 26834065]
19. Weaver KE, Foraker RE, Alfano CM, Rowland JH, Arora NK, Bellizzi KM, et al. Cardiovascular risk factors among long-term survivors of breast, prostate, colorectal, and gynecologic cancers: a gap in survivorship care? J Cancer Surviv 2013;7(2):253-261 [FREE Full text] [doi: 10.1007/s11764-013-0267-9] [Medline: 23417882]
20. Haque R, Prout M, Geiger AM, Kamineni A, Thwin SS, Avila C, et al. Comorbidities and cardiovascular disease risk in older breast cancer survivors. Am J Manag Care 2014;20(1):86-92 [FREE Full text] [Medline: 24512167]

XSL•FO
RenderX

21. de Moor JS, Mariotto A, Parry C, Alfano C, Padgett L, Kent E, et al. Cancer survivors in the United States: prevalence across the survivorship trajectory and implications for care. Cancer Epidemiol Biomarkers Prev 2013;22(4):561-570 [FREE Full text] [doi: 10.1158/1055-9965.EPI-12-1356] [Medline: 23535024]
22. Singla A, Kumar G, Bardia A. Personalizing cardiovascular disease prevention among breast cancer survivors. Curr Opin Cardiol 2012;27(5):515-524. [doi: 10.1097/HCO.0b013e3283570040] [Medline: 22874128]
23. Daher IN, Daigle TR, Bhatia N, Durand J. The prevention of cardiovascular disease in cancer survivors. Tex Heart Inst J 2012;39(2):190-198 [FREE Full text] [Medline: 22740730]
24. National Research Council, Institute of Medicine, National Cancer Policy Board, Committee on Cancer Survivorship, Improving Care and Quality of Life. In: Hewitt M, Greenfield S, Stovall E, editors. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: National Academies Press; 2006.
25. Committee on Improving the Quality of Cancer Care: Addressing the Challenges of an Aging Population;, Board on Health Care Services, Institute of Medicine. In: Levit L, Balogh E, Nass S, Ganz PA, editors. Delivering High-Quality Cancer Care: Charting a New Course for a System in Crisis. Washington DC: National Academies Press; 2013.
26. Snyder CF, Frick KD, Peairs KS, Kantsiper ME, Herbert RJ, Blackford AL, et al. Comparing care for breast cancer survivors to non-cancer controls: a five-year longitudinal study. J Gen Intern Med 2009;24(4):469-474 [FREE Full text] [doi: 10.1007/s11606-009-0903-2] [Medline: 19156470]
27. Snyder CF, Earle CC, Herbert RJ, Neville BA, Blackford AL, Frick KD. Trends in follow-up and preventive care for colorectal cancer survivors. J Gen Intern Med 2008;23(3):254-259 [FREE Full text] [doi: 10.1007/s11606-007-0497-5] [Medline: 18197456]
28. Earle CC, Burstein HJ, Winer EP, Weeks JC. Quality of non-breast cancer health maintenance among elderly breast cancer survivors. J Clin Oncol 2003;21(8):1447-1451. [doi: 10.1200/JCO.2003.03.060] [Medline: 12697865]
29. Snyder CF, Frick KD, Kantsiper ME, Peairs KS, Herbert RJ, Blackford AL, et al. Prevention, screening, and surveillance care for breast cancer survivors compared with controls: changes from 1998 to 2002. J Clin Oncol 2009;27(7):1054-1061 [FREE Full text] [doi: 10.1200/JCO.2008.18.0950] [Medline: 19164212]
30. Armenian SH, Lacchetti C, Barac A, Carver J, Constine LS, Denduluri N, et al. Prevention and monitoring of cardiac dysfunction in survivors of adult cancers: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol 2017;35(8):893-911. [doi: 10.1200/JCO.2016.70.5400] [Medline: 27918725]
31. Denlinger CS, Sanft T, Moslehi JJ, Overholser L, Armenian S, Baker K, et al. NCCN Guidelines insights: survivorship, version 2.2020. J Natl Compr Canc Netw 2020;18(8):1016-1023 [FREE Full text] [doi: 10.6004/jnccn.2020.0037] [Medline: 32755975]
32. Weaver KE, Klepin HD, Wells BJ, Dressler EV, Winkfield KM, Lamar ZS, et al. Cardiovascular assessment tool for breast cancer survivors and oncology providers: usability study. JMIR Cancer 2021;7(1):e18396 [FREE Full text] [doi: 10.2196/18396] [Medline: 33475511]
33. Kelley M, Foraker R, Lin ED, Kulkarni M, Lustberg M, Weaver KE. Oncologists' perceptions of a digital tool to improve cancer survivors' cardiovascular health. ACI open 2019;3(2):e78-e87. [doi: 10.1055/s-0039-1696732] [Medline: 39149692]
34. Arem H, Duan X, Ehlers DK, Lyon ME, Rowland JH, Mama SK. Provider discussion about lifestyle by cancer history: a nationally representative survey. Cancer Epidemiol Biomarkers Prev 2021;30(2):278-285 [FREE Full text] [doi: 10.1158/1055-9965.EPI-20-1268] [Medline: 33268489]
35. Foraker RE, Shoben AB, Kelley MM, Lai AM, Lopetegui MA, Jackson RD, et al. Electronic health record-based assessment of cardiovascular health: the Stroke Prevention in Healthcare Delivery Environments (SPHERE) study. Prev Med Rep 2016;4:303-308 [FREE Full text] [doi: 10.1016/j.pmedr.2016.07.006] [Medline: 27486559]
36. Foraker RE, Kite B, Kelley M, Lai AM, Roth C, Lopetegui MA, et al. EHR-based visualization tool: adoption rates, satisfaction, and patient outcomes. EGEMS 2015;3(2):1159 [FREE Full text] [doi: 10.13063/2327-9214.1159] [Medline: 26290891]
37. Foraker RE, Shoben AB, Lopetegui MA, Lai AM, Payne PRO, Kelley M, et al. Assessment of Life's Simple 7 in the primary care setting: the Stroke Prevention in Healthcare Delivery Environments (SPHERE) study. Contemp Clin Trials 2014;38(2):182-189. [doi: 10.1016/j.cct.2014.03.007] [Medline: 24721482]
38. Sekhon M, Cartwright M, Francis JJ. Acceptability of health care interventions: a theoretical framework and proposed research agenda. Br J Health Psychol 2018;23(3):519-531. [doi: 10.1111/bjhp.12295] [Medline: 29453791]
39. Weaver KE, Dressler EV, Klepin HD, Lee SC, Wells BJ, Smith S, et al. Effectiveness of a cardiovascular health electronic health record application for cancer survivors in community oncology practice: results from WF-1804CD. J Clin Oncol 2025;43(1):46-56. [doi: 10.1200/JCO.24.00342] [Medline: 39571113]
40. About NCORP. National Cancer Institute Community Oncology Research Program. URL: https://ncorp.cancer.gov/about/ [accessed 2025-02-04]
41. Foraker RE, Davidson EC, Dressler EV, Wells BJ, Lee SC, Klepin HD, et al. Addressing cancer survivors' cardiovascular health using the Automated Heart Health Assessment (AH-HA) EHR tool: Initial protocol and modifications to address COVID-19 challenges. Contemp Clin Trials Commun 2021;22:100808 [FREE Full text] [doi: 10.1016/j.conctc.2021.100808] [Medline: 34189339]

XSL•FO
RenderX

42. Brana I, Tabernero J. Cardiotoxicity. Ann Oncol 2010;21 Suppl 7:173-179 [FREE Full text] [doi: 10.1093/annonc/mdq295] [Medline: 20943611]
43. Bird BRJH, Swain SM. Cardiac toxicity in breast cancer survivors: review of potential cardiac problems. Clin Cancer Res 2008;14(1):14-24. [doi: 10.1158/1078-0432.CCR-07-1033] [Medline: 18172247]
44. Wallace LS, Rogers ES, Roskos SE, Holiday DB, Weiss BD. Brief report: screening items to identify patients with limited health literacy skills. J Gen Intern Med 2006;21(8):874-877 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00532.x] [Medline: 16881950]
45. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
46. ATLAS.ti 9 Windows: Quick tour. ATLAS.ti. 2021. URL: https://doc.atlasti.com/QuicktourWin.v9/ATLAS.ti_QuickTourWin.v9.pdf [accessed 2025-04-02]
47. Nekhlyudov L, O'malley DM, Hudson SV. Integrating primary care providers in the care of cancer survivors: gaps in evidence and future opportunities. Lancet Oncol 2017;18(1):e30-e38 [FREE Full text] [doi: 10.1016/S1470-2045(16)30570-8] [Medline: 28049575]
48. Dona AC, Jewett PI, Hwee S, Brown K, Solomon M, Gupta A, et al. Logistic burdens of cancer care: a qualitative study. PLoS One 2024;19(4):e0300852 [FREE Full text] [doi: 10.1371/journal.pone.0300852] [Medline: 38573993]
49. Wang Y, Feng W. Cancer-related psychosocial challenges. Gen Psychiatr 2022;35(5):e100871 [FREE Full text] [doi: 10.1136/gpsych-2022-100871] [Medline: 36311374]
50. Kripalani S, Weiss BD. Teaching about health literacy and clear communication. J Gen Intern Med 2006;21(8):888-890 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00543.x] [Medline: 16881953]
51. Parikh NS, Parker RM, Nurss JR, Baker DW, Williams MV. Shame and health literacy: the unspoken connection. Patient Educ Couns 1996;27(1):33-39. [doi: 10.1016/0738-3991(95)00787-3] [Medline: 8788747]
52. Bass PF, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. Acad Med 2002;77(10):1039-1041. [doi: 10.1097/00001888-200210000-00021] [Medline: 12377684]
53. Powell CK, Kripalani S. Brief report: Resident recognition of low literacy as a risk factor in hospital readmission. J Gen Intern Med 2005;20(11):1042-1104 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0220.x] [Medline: 16307631]
54. Blaes AH, Adamson PC, Foxhall L, Bhatia S. Survivorship care plans and the commission on cancer standards: the increasing need for better strategies to improve the outcome for survivors of cancer. JCO Oncol Pract 2020;16(8):447-450. [doi: 10.1200/JOP.19.00801] [Medline: 32267803]
55. Wasti SP, Simkhada P, van Teijlingen ER, Sathian B, Banerjee I. The growing importance of mixed-methods research in health. Nepal J Epidemiol 2022;12(1):1175-1178 [FREE Full text] [doi: 10.3126/nje.v12i1.43633] [Medline: 35528457]
56. Shelton RC, Cooper BR, Stirman SW. The sustainability of evidence-based interventions and practices in public health and health care. Annu Rev Public Health 2018;39:55-76 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-014731] [Medline: 29328872]
57. Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, Giobbie-Hurder A, et al. Multisite randomized controlled trials in health services research: scientific challenges and operational issues. Med Care 2001;39(6):627-634. [doi: 10.1097/00005650-200106000-00010] [Medline: 11404645]
58. DeMari JA, Dressler EV, Foraker RE, Wells BJ, Smith S, Klepin H, et al. Endometrial cancer survivors' perceptions of their cardiovascular disease risk (results from WF-1804CD AH-HA). Gynecol Oncol 2023;174:208-212 [FREE Full text] [doi: 10.1016/j.ygyno.2023.05.009] [Medline: 37224793]
59. Wutich A, Beresford M, Bernard HR. Sample sizes for 10 types of qualitative data analysis: an integrative review, empirical guidance, and next steps. Int J Qual Methods 2024;23:16094069241296206. [doi: 10.1177/16094069241296206]

Abbreviations

AH-HA: Automated Heart-Health Assessment
CVH: cardiovascular health
EHR: electronic Health Record
IRB: institutional review board
NCI: National Cancer Institute
NCORP: National Cancer Institute Community Oncology Research Program
Q-PRO: Qualitative and Patient-Reported Outcomes
UTAUT: Unified Theory of Acceptance and Use of Technology

XSL•FO
RenderX

Edited by G Tsafnat; submitted 21.08.24; peer-reviewed by Z Ehtesham, J Patrianakos, Y Zhu; comments to author 13.12.24; revised version received 17.01.25; accepted 18.01.25; published 06.03.25.

Please cite as:
Nightingale CL, Dressler EV, Kepper M, Klepin HD, Lee SC, Smith S, Aguilar A, Wiseman KD, Sohl SJ, Wells BJ, DeMari JA, Throckmorton A, Kulbacki LW, Hanna J, Foraker RE, Weaver KE
Oncology Provider and Patient Perspectives on a Cardiovascular Health Assessment Tool Used During Posttreatment Survivorship Care in Community Oncology (Results from WF-1804CD): Mixed Methods Observational Study
J Med Internet Res 2025;27:e65152
URL: https://www.jmir.org/2025/1/e65152
doi:10.2196/65152
PMID:39854647

©Chandylen L Nightingale, Emily V Dressler, Maura Kepper, Heidi D Klepin, Simon Craddock Lee, Sydney Smith, Aylin Aguilar, Kimberly D Wiseman, Stephanie J Sohl, Brian J Wells, Joseph A DeMari, Alyssa Throckmorton, Lindsey W Kulbacki, Jenny Hanna, Randi E Foraker, Kathryn E Weaver. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Risk Stratification in Immunoglobulin A Nephropathy Using Network Biomarkers: Development and Validation Study

Jiaxing Tan[1*], MD, PhD; Rongxin Yang[2*], BS; Liyin Xiao[2], BS; Lingqiu Dong[1], MD, PhD; Zhengxia Zhong[3], MD, PhD; Ling Zhou[4], MD; Wei Qin[1], MD, PhD

[1]Division of Nephrology, Department of Medicine, West China Hospital of Sichuan University, Chengdu, China

[2]College of Computer Science, Sichuan University, Chengdu, China

[3]Division of Nephrology, Department of Medicine, Affiliated Hospital of Zunyi Medical University, Zunyi, China

[4]Division of Nephrology, Zigong Third People's Hospital, Zigong, China

[*]these authors contributed equally

Corresponding Author:
Wei Qin, MD, PhD
Division of Nephrology, Department of Medicine
West China Hospital of Sichuan University
No 37 Guoxue Alley, Wuhou District
Chengdu, 610041
China
Phone: 86 18980602119
Email: qinweihx@scu.edu.cn

Abstract

Background: Traditional risk models for immunoglobulin A nephropathy (IgAN), which primarily rely on renal indicators, lack comprehensive assessment and therapeutic guidance, necessitating more refined and integrative approaches.

Objective: This study integrated network biomarkers with unsupervised learning clustering (k-means clustering based on network biomarkers [KMN]) to refine risk stratification in IgAN and explore its clinical value.

Methods: Involving a multicenter prospective cohort, we analyzed 1460 patients and validated the approach externally with 200 additional patients. Deeper metabolic and microbiomic insights were gained from 2 distinct cohorts: 63 patients underwent ultraperformance liquid chromatography–mass spectrometry, while another 45 underwent fecal 16S RNA sequencing. Our approach used hierarchical clustering and k-means methods, using 3 sets of indicators: demographic and renal indicators, renal and extrarenal indicators, and network biomarkers derived from all indicators.

Results: Among 6 clustering methods tested, the KMN scheme was the most effective, accurately reflecting patient severity and prognosis with a prognostic accuracy area under the curve (AUC) of 0.77, achieved solely through cluster grouping without additional indicators. The KMN stratification significantly outperformed the existing International IgA Nephropathy Prediction Tool (AUC of 0.72) and renal function-renal histology grading schemes (AUC of 0.69). Clinically, this stratification facilitated personalized treatment, recommending angiotensin-converting enzyme inhibitors or angiotensin receptor blockers for lower-risk groups and considering immunosuppressive therapy for higher-risk groups. Preliminary findings also indicated a correlation between IgAN progression and alterations in serum metabolites and gut microbiota, although further research is needed to establish causality.

Conclusions: The effectiveness and applicability of the KMN scheme indicate its substantial potential for clinical application in IgAN management.

(J Med Internet Res 2025;27:e65563) doi:10.2196/65563

KEYWORDS

IgA nephropathy; unsupervised learning; network biomarker; metabolomics; gut microbiota; biomarkers; risk stratification; IgA; immunoglobulin A; renal biopsy; renal; prospective cohort; Berger disease; synpharyngitic glomerulonephritis; kidney; immune system; glomerulonephritis; kidney inflammation; chronic kidney disease; renal disease; nephropathy; nephritis

XSL•FO
RenderX

Introduction

Immunoglobulin A nephropathy (IgAN) is the predominant primary glomerular disorder globally, presenting across diverse demographics. The prognosis of IgAN shows notable heterogeneity, with approximately 40% of patients progressing to end-stage renal disease, posing challenges for clinical judgment and management [1]. Hence, investigating risk factors for IgAN has been a recent research hotspot. Alongside traditional kidney-related markers such as proteinuria [2], hematuria [3], estimated glomerular filtration rate (eGFR) slope [4], uric acid [5,6] and Oxford classification [7], emerging extrarenal indicators like smoking status [8], bilirubin [9], platelet-albumin index [10], and triglyceride-glucose index [11] are increasingly explored for their prognostic value in IgAN. This suggests that clinicians should assess IgAN holistically, not just focusing on the renal biomarkers.

To assess IgAN's prognosis, several stratification models have emerged, with the International Immunoglobulin A Nephropathy Prediction Tool (IIgAN-PT) by Barbour et al [12] being the most widely accepted. This model primarily focuses on traditional kidney-related indicators, such as the Oxford classification, treatment regimens, proteinuria, and blood pressure, and has been validated for prognostic accuracy by other research groups [13,14]. Recently, a Japanese cohort led by Koike et al [15] proposed the renal function–renal histology grading (RF-RG) stratification scheme for IgAN, primarily referencing the clinical grade (proteinuria and eGFR) and histological lesions to predict renal failure risk accurately, making it clinically promising due to its simplicity. Regrettably, these studies, while advancing IgAN risk stratification, are primarily retrospective, focusing mainly on traditional renal markers and often neglecting crucial extrarenal indicators. Consequently, these tools fail to differentiate prognosis among patients with similar renal damage who also have conditions like hepatitis B or diabetes—comorbidities that can impact outcomes. Moreover, while the classification schemes forecast prognosis risks, they provide minimal guidance for treatment, with IgAN guidelines largely relying on proteinuria, renal function, and limited histological data [16]. Therefore, current classification models for patients with IgAN are effective but need improvement.

Recent advancements in artificial intelligence, especially in unsupervised learning, have notably influenced various fields. This technique autonomously uncovers patterns and latent information within datasets, showing early successes in projects like gut microbiota classification and autism spectrum disorder stratification [17,18]. However, these applications often focus narrowly on isolated clinical metrics, missing broader interconnections. Simultaneously, the emergence of network biomarkers in precision medicine provides a deeper understanding of biological systems through the analysis of complex molecular interactions. This advanced approach improves our understanding of disease mechanisms, enhances prognostic accuracy, facilitates personalized treatments, and identifies critical stages in disease progression [19]. Yet, the potential of network biomarkers for clustering in biomedical research remains largely untapped, indicating a significant opportunity for further exploration and application in the field.

To address these issues, our study used a multicenter, prospective, observational cohort design to not only focus on traditional renal damage indicators but also incorporate extrarenal markers closely linked to the disease's progression. By leveraging interdisciplinary techniques such as unsupervised learning and network biomarkers, we derived and externally validated optimal classification models for IgAN risk stratification. Additionally, we investigated whether these models could be used for long-term follow-up and to guide treatment decisions in IgAN. Furthermore, we examined significant differences in the metabolomics and gut microbiota under this stratification model, both of which have been documented in literature as closely related to the onset and progression of IgAN [20-26], aiming to provide preliminary clues into the mechanisms of IgAN progression (Figure 1).

XSL•FO
RenderX

Figure 1. A comprehensive process for developing and applying clustering models. IgA: immunoglobulin A.

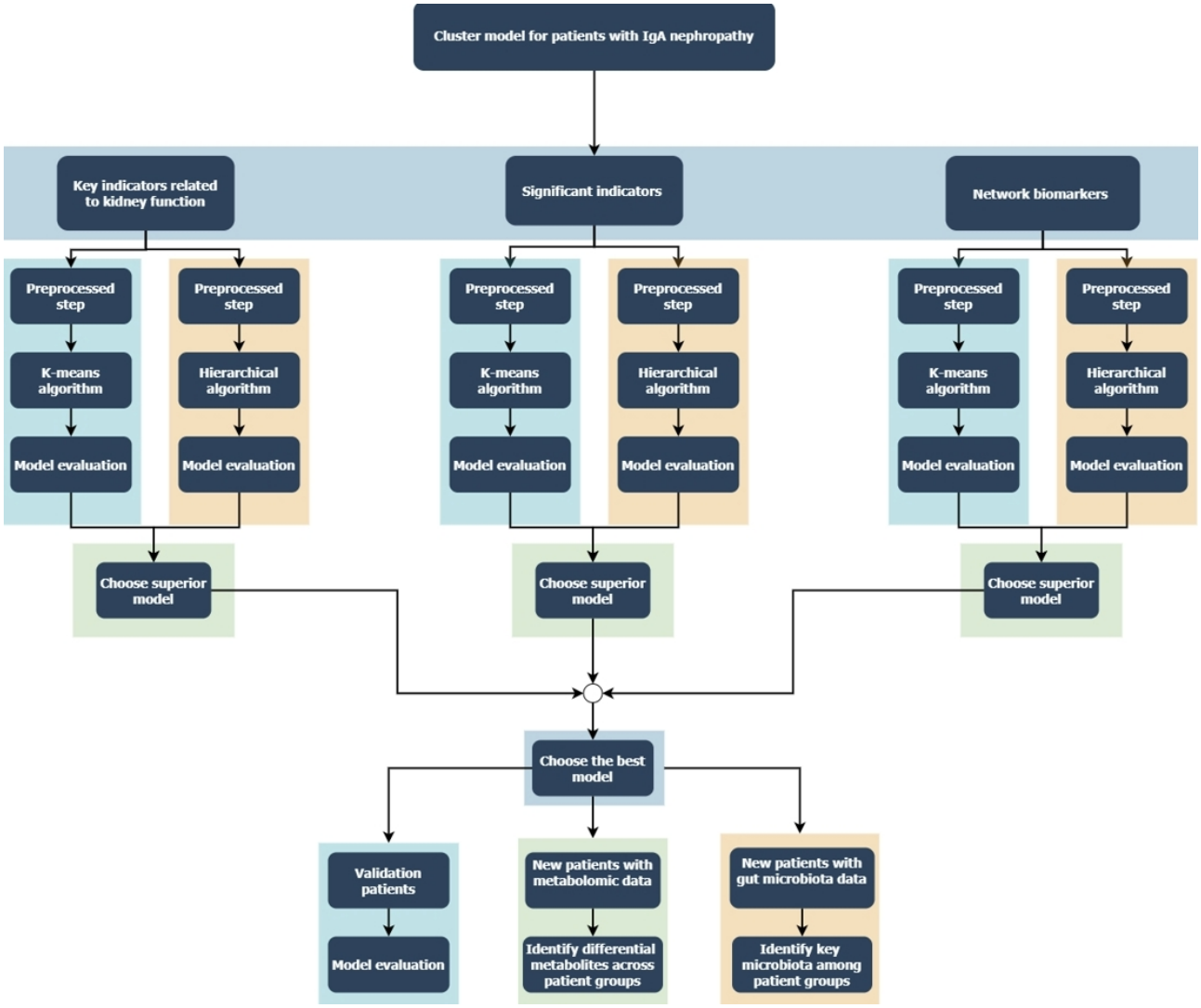

Methods

Participant Recruitment

This multicenter, prospective cohort study included 1768 cases of IgAN diagnosed via renal biopsy between 2009 and 2021 at 4 nephrology centers: West China Hospital of Sichuan University, the Affiliated Hospital of Zunyi Medical University, Zigong Third People's Hospital, and People's Hospital of Mianzhu City. Follow-up data were last updated in October 2023 and are maintained by specialized technicians. The study included patients with a confirmed IgAN diagnosis and complete biopsy and clinical data, with IgAN identified as the primary disease. We excluded patients with insufficient pathological data, biopsies containing fewer than 8 glomeruli, incomplete follow-up data, or diagnosed with secondary IgAN. A total of 1460 patients from West China Hospital of Sichuan University meeting the inclusion criteria underwent unsupervised clustering, with the resulting algorithm subsequently validated on an independent cohort of 200 patients from other centers.

To assess the potential clinical meanings of this clustering, we analyzed separate groups of 63 and 45 patients who met our study criteria. These participants underwent ultraperformance liquid chromatography-mass spectrometry for serum metabolomic profiling and 16S RNA sequencing to identify differences in gut microbiota. Participants were free from metabolic diseases, such as hyperthyroidism and diabetes, and had no history of smoking, alcohol misuse, drug abuse, or infectious diseases. They had not received steroids or immunosuppressants in the past 6 months nor had they taken any lipid-lowering, uric acid–lowering, or metabolism-impacting drugs in the previous week. Additionally, they had not ingested probiotics, antibiotics, or gastrointestinal motility drugs in the last month and showed no significant digestive symptoms or diseases. The study adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational studies, ensuring comprehensive and transparent reporting of methods and results.

Ethical Considerations

Ethics approval for this study was obtained from the Ethics Committee of West China Hospital, Sichuan University (approval 2019-33), and all procedures strictly followed relevant ethical guidelines and regulations. Written informed consent was obtained from all participants, who were fully informed of their rights. Participant data were anonymized to ensure privacy and confidentiality, with securely stored, deidentified datasets accessible only to authorized researchers. No financial or

XSL•FO
RenderX

material compensation was provided to participants. The study did not involve the inclusion of any identifiable individuals in images or Multimedia Appendix 1.

Therapeutic Approaches and End Point

Adhering to KDIGO (Kidney Disease: Improving Global Outcomes) guidelines, our collaborative treatment approach between physicians and patients consists of 3 main protocols: supportive care with angiotensin-converting enzyme inhibitors (ACEIs) or angiotensin receptor blockers (ARBs), glucocorticoid administration starting at 0.5-1 mg/kg of prednisone with subsequent tapering, and a regimen of immunosuppressive agents, used with or without additional glucocorticoids [16]. Considering medication effects on the immune system, we merged glucocorticoid administration and immunosuppressive agents into immunosuppressive therapy (IST). The study aimed to evaluate composite outcomes, including end-stage renal disease, kidney transplantation, mortality, or significant declines in eGFR exceeding 50%.

Data Processing and Unsupervised Clustering

We conducted our analysis using k-means and hierarchical clustering, following data preprocessing that involved imputing missing values with a multilinear interpolation algorithm, standardizing data to zero mean and unit variance, and reducing dimensionality with uniform manifold approximation and projection [27,28]. The k-means algorithm clusters data by first randomly assigning centroids and then iteratively relocating them to minimize intracluster variance, stabilizing when centroid movement ceases. Hierarchical clustering generates a dendrogram through agglomerative or divisive strategies: agglomerative starts with individual points as clusters, merging the most similar until one remains, while divisive begins with a single cluster that is continually split until each data point is isolated.

Network Biomarkers

We conducted a perturbation analysis using a single sample-specific network approach to analyze individual samples, comparing their attributes with a reference group of 50 healthy individuals to identify distinctive features. Using graph feature engineering techniques, we extracted attributes at 3 network levels: node, subgraph, and global graph, to gain comprehensive insights. Node-level analysis included degree analysis to assess connectivity. At the subgraph level, we evaluated significant features like diameter and average clustering coefficient to identify disease-relevant patterns. At the global level, we analyzed node and edge counts and the number of connected components. Using the *Networkx* package (version 2.2) in Python (version 3.1.2; Python Software Foundation), we extracted 46 attributes for each sample's network, which were used for clustering analysis [19,29,30].

Interpretability Assessment of Unsupervised Clustering Algorithms

Unsupervised clustering algorithms, while powerful in detecting latent patterns, often lack interpretability. To address this issue, the subgroups identified by these algorithms were used as surrogate labels to develop a predictive framework through machine learning. This framework was refined using 10-fold cross-validation. An interpretability analysis was subsequently conducted on the proficient machine learning models using the Shapley Additive Explanations method, aiming to elucidate key indicators influencing classification outcomes [31].

External Validation and Subsequent Omics Validation Labels

We applied data standardization and uniform manifold approximation and projection dimensionality reduction to the entire dataset rather than individual data points, using the patient cohort from the cluster set as the reference dataset. Each new patient was then processed sequentially using the standardization and dimensionality reduction models that were initially trained on this cluster set. After dimensionality reduction, the clustering model—previously trained on the cluster set—was used to assign a corresponding cluster label to each patient. This methodology ensures consistent treatment of new data within the established framework.

Processing for Metabolomics Analysis

After fasting for at least 8 hours, about 5 mL of blood was drawn into anticoagulant-free vacutainers in the morning. These were left at room temperature to clot for 30 minutes before centrifuging at 3000 rpm for 10 minutes. The supernatant was then transferred to 1.5-mL Eppendorf tubes and stored at –80 °C for analysis.

Samples were thawed and prepared for ultraperformance liquid chromatography-mass spectrometry analysis, involving protein removal and reconstitution in a specific solvent. Hydrophilic interaction liquid chromatography was performed using a precise column and electrospray ionization for detection in both ionization modes. Mass spectrometry data were processed to identify peaks and extract ion features, with quality control measures to assess the variability of each metabolite, excluding those with a coefficient of variation over 20%. After sum normalization, data analysis included partial least-squares discrimination (PLS-DA) and calculation of variable importance in projection (VIP) values for each metabolite, considering VIP values >1 as statistically significant.

Processing of Gut Microbiota Data

Approximately 4 g of fresh fecal samples were collected in sterile containers and stored at –80 °C for high-throughput sequencing. DNA was extracted from the samples, and its quantity and purity were assessed. The 16S ribosomal ribonucleic acid (rRNA) V3-V4 hypervariable regions were then amplified using specific primers and sequenced on the Illumina MiSeq platform. Raw sequencing reads were quality-filtered and clustered into operational taxonomic units at a 97% similarity threshold using UPARSE. The 16S rRNA sequences were classified using the ribosomal database project classifier algorithm against the Silva (SSU123) 16S rRNA database. Quality-controlled and standardized abundance data were used for subsequent analysis at the genus level.

Statistical Analysis

Survival curves were generated using the Kaplan-Meier method, and differences between groups were assessed using the log-rank test. To estimate hazard ratios (HRs) and their 95% CIs, Cox

XSL•FO
RenderX

proportional hazards models were used, with *P* values less than .05 indicating statistical significance. Additionally, extreme gradient boosting models were used to explore the associations between cluster labels and primary outcomes, with model performance evaluated by the area under the receiver operating characteristic curves. The standardized effect size differences among subtypes were quantified using Cohen *d*. Correlation network figures were constructed based on Spearman or Pearson correlation analyses to examine interactions between various metabolites and gut microbiota; edges were added to the network for any correlation with a *P* value below .05. Furthermore, Mantel tests were applied to analyze the associations between groups of clinical indicators and either metabolites or gut microbiota.

Results

Enhanced Performance of K-Means Clustering With Network Biomarkers in IgAN Stratification Over Other Unsupervised Learning Approaches

We curated a cohort of 1460 patients from West China Hospital of Sichuan University, strictly adhering to stringent inclusion criteria for our clustering set. The average follow-up period was 58.8 (SD 28.8) months. Our comprehensive dataset included demographics, clinical records, pathology, and treatment details such as age, blood pressure, serum creatinine, eGFR, 24-hour urine protein, urine red blood cell count, uric acid, renal immunofluorescence markers, Oxford MESTC classification (M, S, E, T, C), global and arteriolar sclerosis, lifestyle factors (smoking and drinking), liver function, comorbidities (diabetes and hepatitis B), and lipid metabolism markers.

We used hierarchical clustering and k-means methods, which are widely used in the medical field for unsupervised clustering [17,32,33]. These methods used 3 sets of indicators: the first set included demographic and renal indicators, the second set combined renal and extrarenal indicators, and the third set comprised network biomarkers derived from all indicators. The objective was to explore whether classifications based on a broader range of systemic markers could outperform those focusing solely on renal indicators, aiming to address potential shortcomings in existing stratification schemes and guidelines. Network biomarker construction was based on references from our prior research [29]. All 6 clustering methods successfully categorized all patients with IgAN into 4 groups, as demonstrated in Figure 2 and Figures S1-S4 in Multimedia Appendix 1. As cluster numbers rose from 1 to 4, symptom severity increased, reflecting closely with clinical evaluations, as evidenced by declining mean eGFR values and rising mean proteinuria values. Although clinical intuition enabled physicians to categorize patients using all indicators, this experience-based and direct judgment alone was not consistently accurate in dividing patients into 4 groups, especially regarding cluster 2 and cluster 3.

XSL•FO
RenderX

Figure 2. Clustering based on unsupervised learning with renal and extrarenal indicators and network biomarkers. (A) Visual representation of UMAP-reduced datasets integrated with k-means clustering, focusing on renal and extrarenal indicators. Each individual is stratified by subtype and depicted through density plots, highlighting the distribution within each group. (B) Analysis of the subtypes derived from the methodologies described in (A), presented using scatter boxplots and heatmaps. These illustrate the pairwise standardized effect size differences (Cohen d) between subtypes, facilitating a clear comparison of indicator levels across groups. (C) Visual representation of UMAP-reduced datasets combined with k-means clustering, focusing on network biomarkers. Individuals are stratified by subtype and depicted through density plots to showcase the clustering outcome. (D) Analysis of the subtypes derived from the methodologies described in (C), presented using scatter boxplots and heatmaps. These visualizations detail the pairwise standardized effect size differences (Cohen d) between subtypes, providing insights into the variation of network biomarkers across different groups. eGFR: estimated glomerular filtration rate; UMAP: uniform manifold approximation and projection; uPro: urine protein; uRBC: urine red blood cell count.

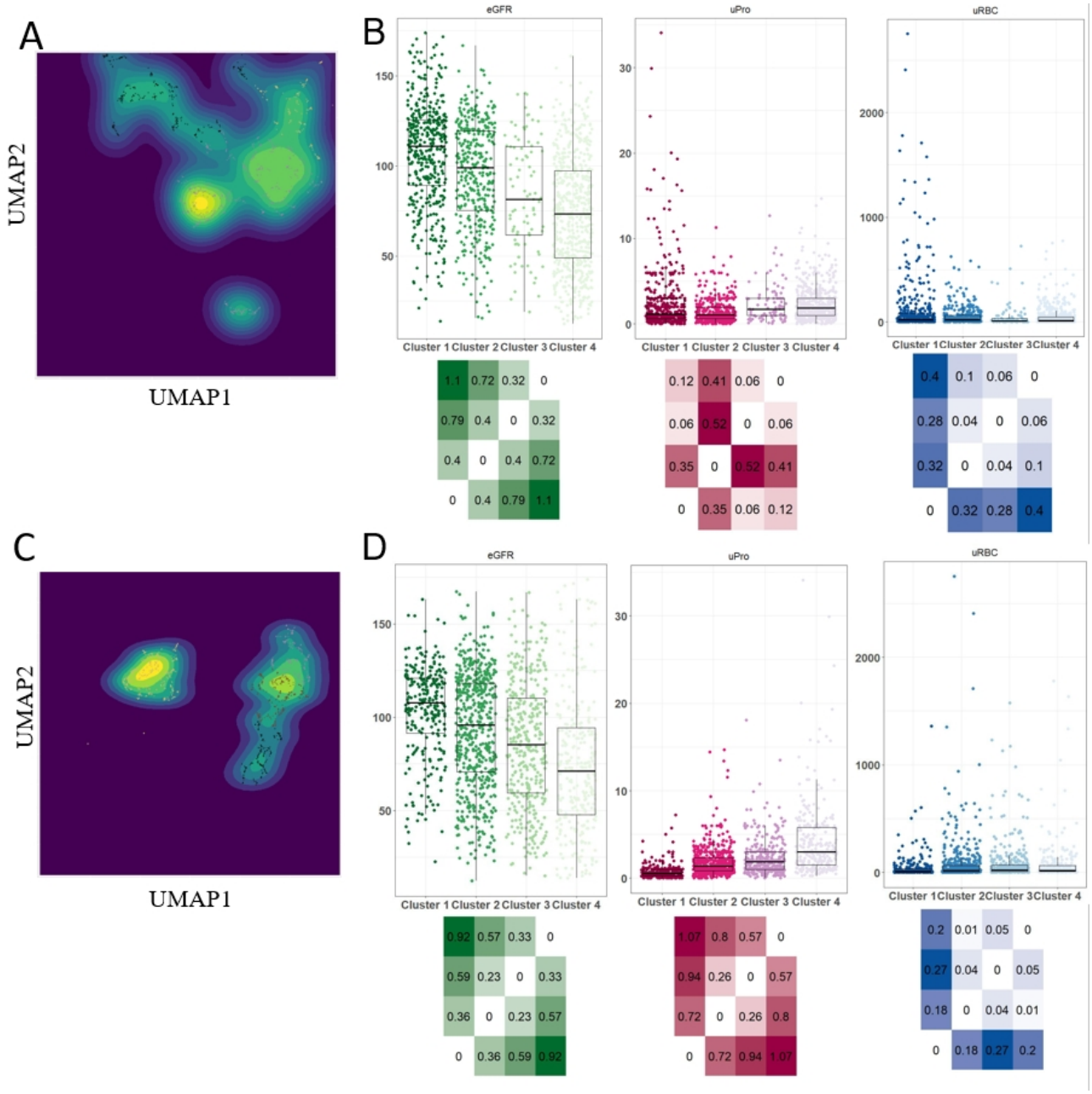

Unsupervised clustering not only accurately differentiated patients but also proved highly effective in evaluating IgAN prognosis. Among the 6 methods used, 3 stood out for effectively reflecting IgAN prognosis: hierarchical clustering based on renal indicators, k-means clustering based on all indicators, and k-means clustering based on network biomarkers (KMN). Kaplan-Meier curves for the 4 clusters showed significant differences (P<.0001; Figure 3A and Figure S5A and B in Multimedia Appendix 1), indicating their prognostic relevance, as supported by a multivariable Cox regression model adjusted for age, gender, urine protein, urine red blood cells, renal function, and blood pressure (P<.0001; Figure 3B and Figure S3C and D in Multimedia Appendix 1). Figure S6 in Multimedia Appendix 1 presents further prognostic parameters, reinforcing previous findings: notably, cluster 1 has the best prognosis, whereas cluster 4 has the worst, consistent with clinical categorizations.

XSL•FO
RenderX

Figure 3. Survival analysis comparisons across different classification methods for IgAN. (A) Kaplan-Meier curves illustrating mortality rates in IgAN, stratified by KMN grouping. (B) Forest plot illustrating the results of a multifactor Cox regression analysis, stratified by KMN grouping. (C). Kaplan-Meier curves showing IgAN mortality rates, stratified by IIGAN grouping. (D) Kaplan-Meier curves detailing IgAN mortality rates, stratified by RF-RG grouping. (E) ROC curve for the KMN grouping, with an AUC value of 0.77. (F) ROC curve for the IIGAN grouping, featuring an AUC value of 0.72. (G) ROC curve for the RF-RG grouping, with an AUC value of 0.69. (H) Comparison of AUCs for classification schemes in IgAN risk stratification. **P*<.05, ***P*<.01, ****P*<.001. AIC: Akaike information criterion; AUC: area under the curve; DBP: diastolic blood pressure; IgAN: immunoglobulin A nephropathy; IIGAN: International Risk-Prediction Tool in Immunoglobulin A Nephropathy; KMN: k-means clustering based on network biomarkers; RF-RG: renal function–renal histology grading; ROC: receiver operating characteristic; SBP: systolic blood pressure; uPro: urine protein; uRBC: urine red blood cell count; XGBoost: extreme gradient boosting.

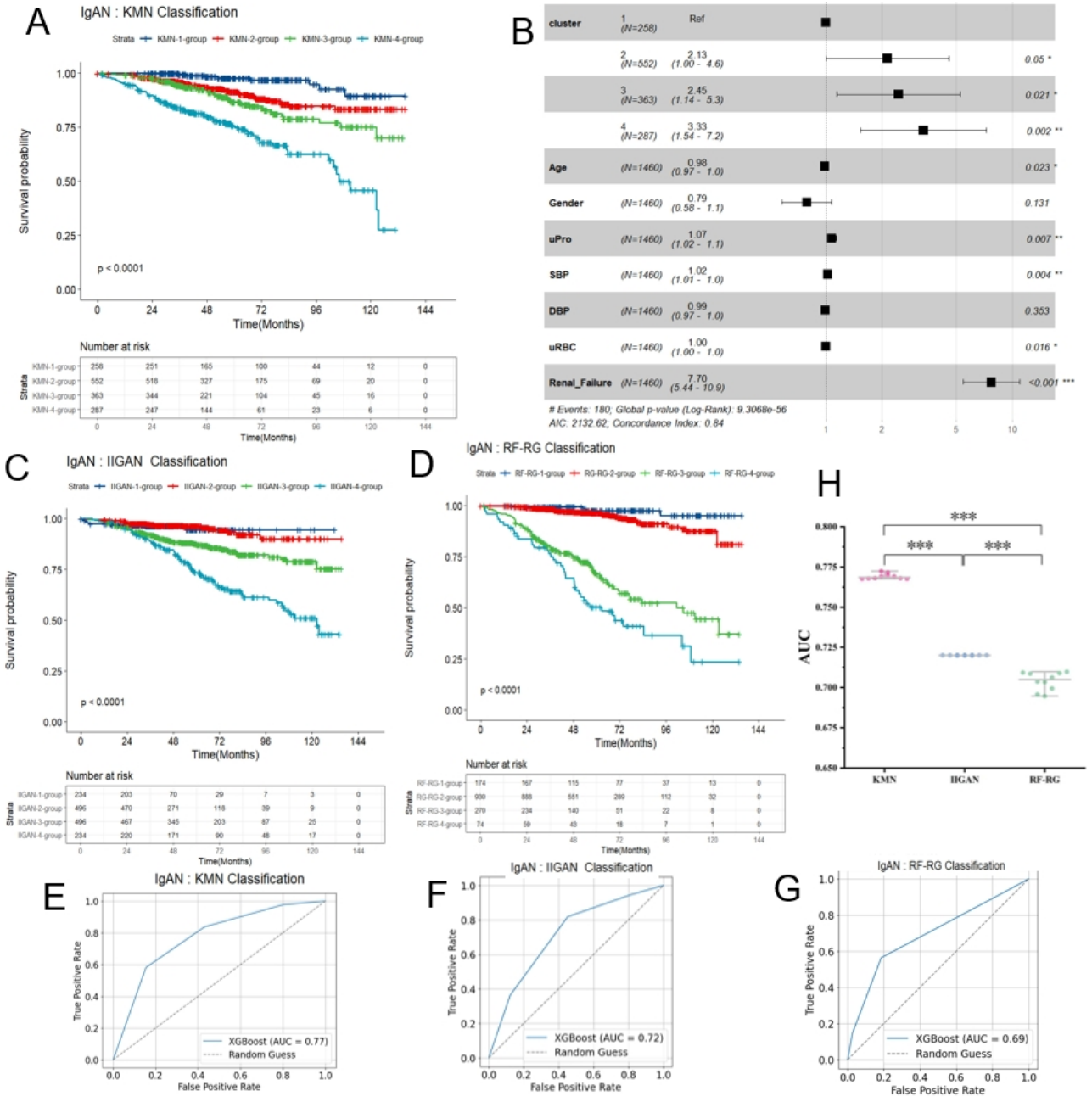

We evaluated the effectiveness of various clustering methods in reflecting IgAN prognosis solely by using cluster labels as input variables and assessing the area under the curve (AUC). This approach holds significant value in assessing the predictive performance of the model. Our study found that KMN exhibited the highest AUC at 0.77 (Figure 3E), outperforming hierarchical clustering based on renal indicators (0.71; Figure S5E in Multimedia Appendix 1) and k-means clustering based on all indicators (0.75; Figure S5F in Multimedia Appendix 1), indicating superior predictive power for KMN and surpassing methods that rely solely on renal indicators by 0.6. Therefore, the k-means clustering approach, based on network biomarkers, was recommended for stratifying IgAN risk.

To further investigate the impact of the KMN stratification on the prognosis of IgAN, we assessed the 3-year trajectories of urinary protein and serum creatinine across different patient groups. Figure 4A and B and Figures S7 and S8 in Multimedia Appendix 1 demonstrate that most patients in cluster 1 exhibited relatively stable average urinary protein levels and normal fluctuations in creatinine levels. Patients in cluster 2 had higher

XSL•FO
RenderX

baseline levels of urinary protein and creatinine than those in cluster 1 but exhibited a declining trend in urinary protein over time. Meanwhile, their creatinine levels increased slightly, with most remaining within chronic kidney disease stages 1, 2, or 3a by the third year. In cluster 3, patients experienced a reduction in urinary protein at the 6-month mark compared to baseline but maintained higher levels (around or above 1 g in 24 hours) throughout the 3 years, with both creatinine and urinary protein showing significant increases by the third year, indicating a potential rapid deterioration of their condition. Patients in cluster 4, having the most severe prognosis, showed a consistent yearly increase in creatinine, quickly progressing toward renal failure, and maintained high levels of urinary protein. These findings suggested that patients in different KMN stratifications had distinct trajectories, indicating that KMN stratification could serve as an effective prognostic marker for IgAN.

Figure 4. Tracking 3-year renal indicators in immunoglobulin A nephropathy and guiding treatment based on KMN grouping. (A) Line graphs depicting the average serum Cr levels of patients in each KMN group over time, illustrating changes in kidney function. (B) Line graphs showing the average uPro levels for each KMN group over time, highlighting trends in proteinuria across different clusters. (C) Kaplan-Meier curves for cluster 1 patients, stratified by treatment type: SC and IST, which includes steroids and other IIs. (D) Kaplan-Meier curves for cluster 1, detailing survival outcomes under different treatments: SC, GS, and IIs. (E and F), (H and I), and (J and K) Kaplan-Meier curves representing survival outcomes for clusters 2, 3, and 4, respectively, under various treatment modalities. (G) A forest plot visualizing the results of a multifactor Cox regression, specifically stratified by the third KMN grouping. ***P<.001. Cr: creatinine; DBP: diastolic blood pressure; GS: glucocorticoid therapy; II: immunosuppressive agents; IST: immunosuppressive therapy; k-means clustering based on network biomarkers; SBP: systolic blood pressure; SC: supportive care; uPro: urine protein; uRBC: urine red blood cell count.

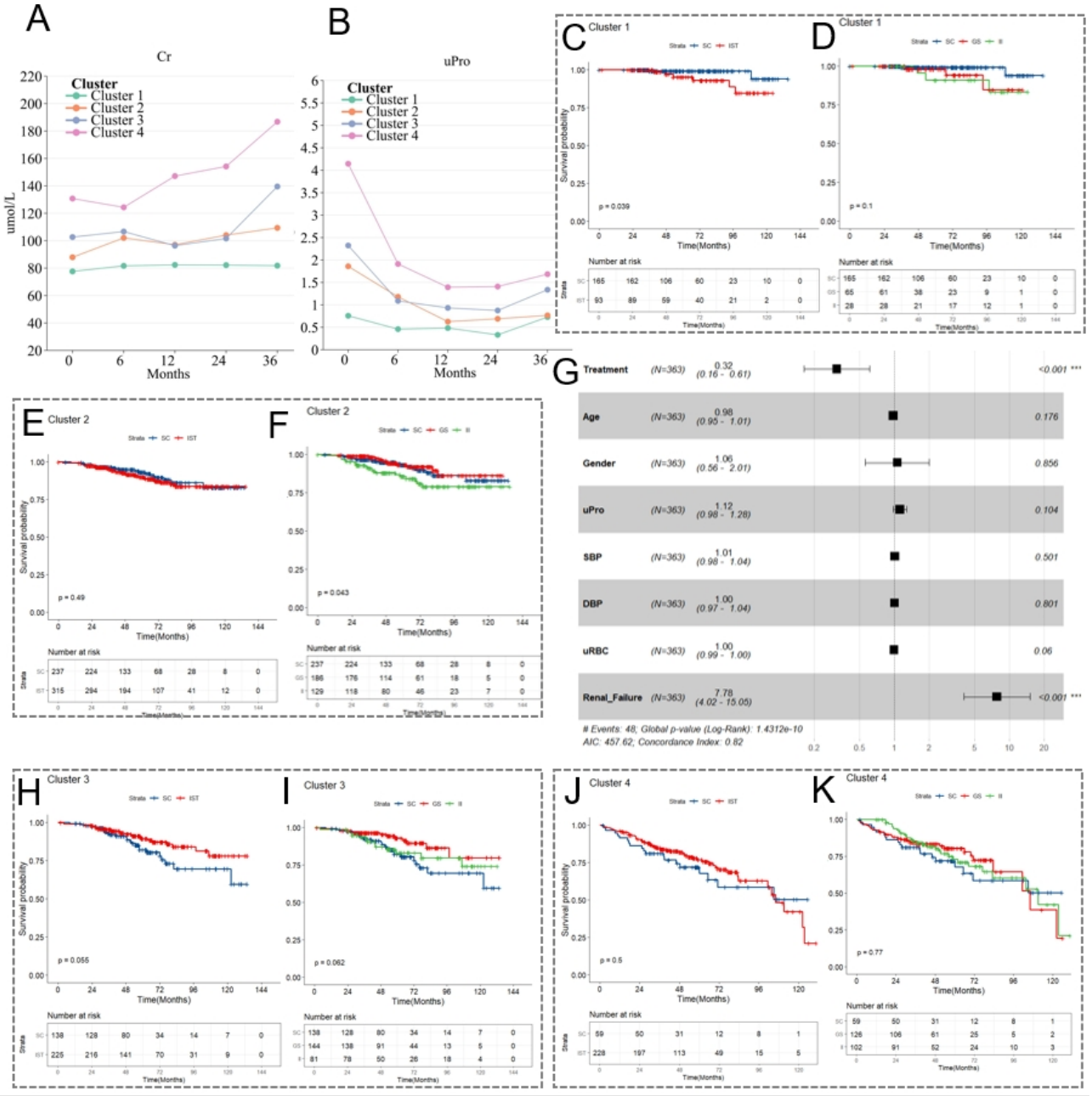

XSL•FO
RenderX

Superiority of KMN Stratification Over IIgAN-PT and RF-RG Schemes

To determine if our KMN scheme outperforms the globally used renal marker–based IIgAN-PT and RF-RG schemes [12,15], we first assessed the applicability of IIgAN-PT and RF-RG to our cohort. As shown in Figure 3C and D, the stratification of our cohort into 4 groups using RF-RG and IIgAN-PT classification methods clearly demonstrates significant disparities in the Kaplan-Meier curves (P<.0001), confirming the prognostic utility of these schemes for our cohort. Multivariable Cox regression analysis (Figure S9 in Multimedia Appendix 1) revealed that in IIgAN-PT, compared to cluster 1, the HRs for clusters 2, 3, and 4 were 0.79 (95% CI 0.35-1.79), 1.69 (95% CI 0.80-3.59), and 3.03 (95% CI 1.41-6.53), respectively. Similarly, for RF-RG, the HRs for clusters 2, 3, and 4 were 2.28 (95% CI 0.81-6.42), 6.60 (95% CI 2.08-20.96), and 8.54 (95% CI 2.55-28.63), respectively. These results suggest that all 6 of our clustering approaches outperformed the RF-RG and IIgAN-PT schemes. Considering potential collinearity between the primary indicators used in these models and the covariates adjusted in our Cox regression, we primarily assessed the models using AUC values. The AUC values for IIgAN-PT and RF-RG were 0.72 (Figure 3F) and 0.69 (Figure 3G), respectively, indicating that IIgAN-PT performs better than RF-RG but is inferior to our KMN model, which achieved an AUC of 0.77 (Figure 3E). We conducted 10 rounds of random sampling to generate AUC values for each classification scheme. A Mann-Whitney U test on the resulting AUC values revealed that the P values for differences between any 2 classification schemes were less than .001. These results confirm statistically significant differences between the methods, with our clustering approach outperforming both RF-RG and IIgAN-PT (Figure 3H).

External Validation of KMN Stratification

To assess whether our risk stratification could be applied to external cohorts, we conducted further validation by including 200 patients with IgAN who met the inclusion and exclusion criteria from the Affiliated Hospital of Zunyi Medical University, Zigong Third People's Hospital, and People's Hospital of Mianzhu City. Our findings indicated that the KMN-based stratification effectively categorized patients in these additional cohorts, with clinical symptoms and predictive performance consistent with previous results (Figure S10 in Multimedia Appendix 1).

Guidance of Treatment Decisions by KMN Stratification Scheme

The current treatment for IgAN predominantly relies on ACEIs or ARBs, with the option of IST for more severe cases. The KIDGO guidelines focus mainly on creatinine and urinary protein levels, offering less guidance on pathological indices or systemic status [16], and do not provide specific recommendations for therapies under the IIgAN-PT or RF-RG schemes [12,15]. Our study aimed to determine if our KMN stratification could effectively guide immunosuppressive treatment for IgAN, a topic currently under debate [34-36]. While recent studies have introduced new therapeutic options such as atacicept [37], telitacicept [38], sparsentan [39], irbesartan [40], and dapagliflozin [41], these were not used in our cohort, which primarily consists of patients followed for over 3-5 years before these drugs were fully available. We explored the utility of KMN stratification in guiding IST through Cox multivariable regression and survival curve analyses as shown in Figure 4 and Figure S11 in Multimedia Appendix 1.

The Kaplan-Meier curve shows significant differences in survival rates within cluster 1 between supportive care, primarily using ACEIs or ARBs, and various ISTs including glucocorticoids and other immunosuppressive agents, underscoring that these therapies might not confer any additional benefit for these patients.

In cluster 2, there was no significant difference between IST and supportive care (P=.49), with supportive care and glucocorticoids performing similarly and better than immunosuppressive agents (P=.04). Given the significant side effects of glucocorticoids and the negative results from multivariable Cox regression for immunosuppressants, supportive care remains the preferred treatment for this group.

For cluster 3, glucocorticoids were found to slightly outperform other immunosuppressants, which themselves were marginally better than supportive care alone (P=.06). Combining glucocorticoids and immunosuppressants demonstrated a clear advantage over supportive care (P=.05). Multivariable Cox regression (Figure 4G) revealed a significant renal protective effect of this combined IST, with an HR of 0.32 (95% CI 0.16-0.61). Given the data showing poor prognosis and rapid renal deterioration typically within 3 years for cluster 3 patients, we strongly recommend immunosuppressive treatment, favoring glucocorticoids or, if intolerable, alternative immunosuppressants such as mycophenolate mofetil.

For cluster 4, which had the poorest prognosis, survival curves and Cox regression analyses showed no significant differences between immunosuppressive treatment and supportive care. However, upon closer examination, immunosuppressive treatment outperformed supportive care for up to 96 months, with intersections thereafter likely due to the small number of patients followed for this duration, which could explain the absence of statistical significance. Given the severity of their condition, immunosuppressive treatment should be considered for these patients if they could manage the side effects, with decisions tailored to individual circumstances.

Interpretable Analysis of KMN Stratification

Unsupervised clustering based on network biomarkers exhibited a "black box" nature; while effective in assessing clinical severity, prognosis, and treatment, it lacked intuitive interpretation. Therefore, we conducted an innovative interpretive analysis of the unsupervised clustering approach, which might offer additional insights for clinical practice. Figure S12 in Multimedia Appendix 1 highlights the key factors influencing the KMN classification, revealing significant network parameters such as diameter, nodes, and edges. Notably, the networks centered on urinary protein, red blood cells, age, and eGFR were particularly crucial for stratifying IgAN. We further analyzed the networks centered on these parameters for patients in clusters 1-4, discovering significant differences

XSL•FO
RenderX

among these groups (Figures S13-S16 in Multimedia Appendix 1). This suggested that not only did urinary protein, urinary red blood cells, age, and eGFR individually impact the prognosis of IgAN, but the relationships between these markers and other indicators also significantly correlated with outcomes. This aspect was often overlooked by traditional statistical methods, which had primarily focused on isolated indicators, especially renal-related ones, neglecting the statistical significance of these networks.

Metabolomic Diversity Analysis Based on KMN Stratification

To identify potential serum metabolomic markers for the progression of IgAN, we used the KMN method to stratify 63 new patients with blood metabolomics data into 4 clusters: cluster 1 with 17 individuals, cluster 2 with 27, cluster 3 with 11, and cluster 4 with 8. After quality control and normalization of the metabolomics data, we conducted a PLS-DA model analysis and observed a distinct separation trend among the 4 clusters (Figure 5A) [42]. Clusters 1 and 2 were then reclassified into a new cohort labeled the low-risk group, and clusters 3 and 4 into a high-risk group. Further PLS-DA analysis of the metabolomic data revealed significant disparities between the high-risk and low-risk classifications (Figure 5B). Additionally, we constructed metabolite interaction networks for the clusters, uncovering distinct intermetabolite relationships across them (Figure S17 in Multimedia Appendix 1). These findings indicated that the differences in serum metabolites between various risk levels of IgAN were significant and that changes in their interaction networks might also contribute to the disease's progression.

XSL•FO
RenderX

Figure 5. Differentiation of serum metabolites and gut microbiota in IgAN progression revealed by KMN grouping. (A) PLS-DA score plots vividly illustrating the separation trends of serum metabolites among the 4 KMN clusters (1, 2, 3, and 4). (B) PLS-DA analysis highlighting significant structural differences in serum metabolites between the high-risk IgAN groups (clusters 3 and 4) and low-risk IgAN groups (clusters 1 and 2). (C) Windrose visualization of variable importance in projection values for metabolites across the 4 grouped PLS-DA models, demonstrating the impact of key metabolites in differentiating the clusters. (D) Bar graph showing the fold changes of differential metabolites in high-risk groups compared to low-risk groups, indicating their increase or decrease in abundance. (E) PLS-DA score plots vividly showcasing the separation trends of gut microbiota among the 4 KMN clusters. (F) PLS-DA analysis indicating significant structural differences in gut microbiota between the high-risk IgAN groups and low-risk IgAN groups. (G) Windrose visualization of variable importance in projection values for gut microbiota across the 4 grouped PLS-DA models, highlighting the impact of key microbiota at the genus level in differentiating the clusters. (H) Bar graph illustrating the fold changes of differential gut microbiota on the genus level in high-risk groups compared to low-risk groups, showing their increase or decrease in abundance. IgAN: immunoglobulin A nephropathy; KMN: k-means clustering based on network biomarkers; PLS-DA: partial least-squares discrimination.

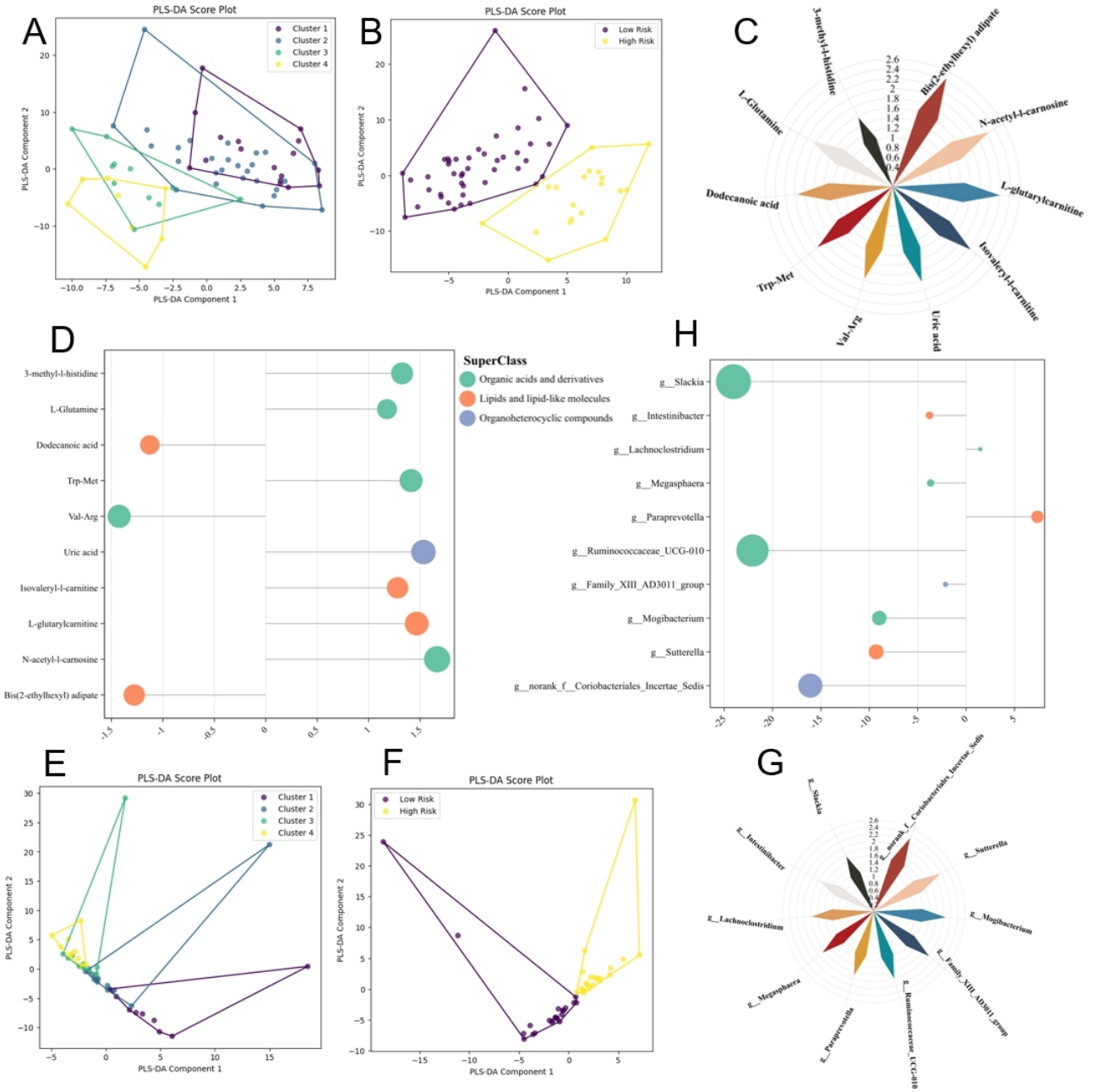

Through VIP analysis of the PLS-DA models for the 4 clusters, we identified the top 10 metabolites that most significantly differentiate the clusters (Figure 5C and D). Notably, levels of bis(2-ethylhexyl) adipate (3-dehydroepiandrosterone sulfate [DEHA]), Val-Arg, and dodecanoic acid were reduced in the high-risk group. DEHA, a plasticizer not typically found in the human body and primarily ingested externally, has no well-documented links to renal diseases, and its metabolic pathways in the body remain unclear. Some studies suggest DEHA could have carcinogenic effects, cause DNA damage, and disrupt steroidogenesis pathways [43-46]. We hypothesize that in high-risk IgAN, abnormal kidney function could decrease the renal excretion of microplastics, leading to their uptake by cells, potentially causing cellular damage, though further experimental evidence is needed to confirm this. Additionally, Val-Arg, a dipeptide involved in protein metabolism, and

XSL•FO
RenderX

dodecanoic acid, a medium-chain fatty acid, may indicate disruptions in protein and fatty acid metabolism in high-risk patients.

Conversely, levels of N-acetyl-L-carnosine (a dipeptide with antioxidant properties) [47], L-glutarylcarnitine and isovaleryl-L-carnitine (both involved in fatty acid metabolism) [48,49], uric acid, Trp-Met (a tryptophan and methionine complex indicating protein metabolism), L-glutamine (an amino acid vital for immune function) [50,51], and 3-methyl-L-histidine (a marker of muscle metabolism) [52] were elevated in the high-risk group, pointing to enhanced oxidative stress and altered protein and energy metabolism in these patients.

An interesting finding from our study was the strong correlation among metabolites, which suggested that their collective interactions might influence the severity of IgAN and potentially lead to poor outcomes (Figure S18A in Multimedia Appendix 1). Additionally, by focusing on these metabolites, we explored the metabolite networks across different clusters and observed distinct network patterns between them (Figures S19-S25 in Multimedia Appendix 1). These variations in metabolite levels and their interactions could provide insights into the mechanisms underlying IgAN progression, particularly in high-risk individuals.

Further investigation into the relationship between these metabolites and clinical indicators revealed that, except for Val-Arg, which demonstrated a statistical correlation with kidney function, the levels of other metabolites did not correlate with traditional renal function markers such as urinary protein, red blood cells, and pathological lesions. This finding suggests that classifications based solely on renal function markers might miss crucial metabolites that play a significant role in the disease's progression. Therefore, our KMN stratification offers a novel approach to understanding the complex mechanisms driving IgAN progression, highlighting the importance of broader biomarker profiles in disease analysis.

Gut Microbiota Diversity Analysis Based on KMN Stratification

Similarly, we recruited an additional 45 patients, stratified into 4 clusters using the KMN method: 9 individuals in cluster 1, 14 in cluster 2, 12 in cluster 3, and 10 in cluster 4. We conducted gut microbiome sequencing on these individuals. Subsequent PLS-DA model analysis of the reclassified cohorts revealed significant disparities between the high- and low-risk groups (Figure 5E and F), indicating substantial differences in the gut microbiomes among patients with varying risk levels. Additionally, the analysis of microbiome interaction networks (Figure S29 in Multimedia Appendix 1) clearly showed distinct intermicrobial relationships across the clusters, suggesting that changes in these interactions may be related to disease progression.

Through VIP analysis of the PLS-DA models for the 4 clusters, significant variances in gut microbiota were identified, delineating distinct profiles between high- and low-risk groups (Figure 5G and H). Elevated levels of *Paraprevotella* and *Lachnoclostridium*—the latter closely linked to kidney function (Figure S18B in Multimedia Appendix 1)—in the high-risk group suggest their role in exacerbating IgAN through gut dysbiosis, given their association with inflammatory processes [53,54]. Conversely, decreased levels of norank_f_Coriobacteriales_Incertae_Sedis, *Sutterella*, *Mogibacterium*, Family_XI AD3011_group, Ruminococcaceae_UCG-010, *Megasphaera*, *Intestinibacter*, and *Slackia* were noted. These genera might be involved in fermenting dietary fibers and producing short-chain fatty acids, essential for colon health and immune modulation [55,56]. Their reduced presence might suggest a loss of protective microbial functions in patients at higher risk of disease progression. Additionally, the interaction patterns between these microbiota batches underscore the potential of gut bacteria as biomarkers for stratifying IgAN risk and highlight the complex relationship between gut health and kidney disease (Figures S26-S34 in Multimedia Appendix 1). However, these causal relationships need to be substantiated by further studies.

Discussion

Principal Findings

Our study used network biomarkers combined with unsupervised clustering to develop a novel stratification scheme for IgAN, offering distinct advantages over existing approaches. By integrating a broad array of variables, including but not limited to renal indicators—such as liver function, lipid metabolism, comorbidities, and lifestyle factors—our method offers a more holistic assessment of patients. Additionally, we examined the interrelationships among these indicators, applying clustering techniques to their interaction vectors, which may more accurately reflect biological interactions. Our stratification scheme has undergone external validation, indicating potential applicability to diverse cohorts. Clinically, this stratification could assist in personalizing treatment, suggesting supportive ACEI or ARB treatment for patients in lower-risk groups (clusters 1 and 2) and potentially considering IST for those in higher-risk groups (clusters 3 and 4).

To our knowledge, our study was the first to integrate k-means clustering with network biomarkers. K-means clustering partitions patients into a predetermined number of clusters based on their similarity to cluster centroids, which is computationally efficient and well-suited for large datasets [32,57]. This makes it particularly effective for managing high-dimensional data in the classification of patients with IgAN. Network biomarkers, offering several advantages over traditional biomarkers, capture the intricate relationships between different biological factors and provide a nuanced understanding of disease heterogeneity [19,30]. They also offer a system-level view that highlights the interconnectedness of various physiological processes involved in IgAN pathogenesis.

By incorporating network biomarkers into k-means classification models, our method outperformed traditional molecular biomarker–based clustering in distinguishing varying degrees of clinical severity in IgAN. This approach not only more accurately categorized patients but also surpassed existing prognostic models, providing deeper insights into disease progression by our interpretable analysis. The innovative

XSL•FO
RenderX

integration of network biomarkers enhances our ability to capture complex biological interactions and systemic effects that traditional markers might overlook. This improves the accuracy of IgAN prognosis and suggests the method's applicability to other diseases with similar complex traits. The scalability and adaptability of this approach could significantly improve disease classification and management across various clinical cohorts, leading to more personalized and effective treatment strategies.

Previous metabolomic and gut microbiome studies have indicated structural changes in the metabolites and microbiota of patients with IgAN compared to healthy individuals [20,24,26]. However, clinical research exploring their impact on disease progression remains preliminary and simplistic. Most investigations into these substances have been conducted through animal or cell experiments to elucidate the changes in the gut-kidney axis and their influence on IgAN [22,24]. Building on this foundation, our study conducted further research, using the KMN stratification to more accurately predict clinical outcomes for these patients, moving beyond simple classifications based on proteinuria and creatinine. We observed significant differences in serum metabolite structures and gut microbiome compositions between high-risk and low-risk patients with IgAN. Importantly, the interaction networks between metabolites and microbiota also varied distinctly, suggesting that changes in these elements and their interactions are closely linked to the progression of IgAN. Preliminary clinical evidence from our study shows that a reduction in beneficial bacteria and an increase in potentially pathogenic microbes, along with environmental pollution, metabolic abnormalities in proteins and lipids, oxidative stress, and immune dysregulation, collectively contribute to the progression of IgAN [58,59]. These findings are consistent with other basic and clinical research, indirectly validating the clinical relevance of our KMN stratification.

Our study has several limitations that warrant discussion. First, our cohort consisted exclusively of Chinese individuals, lacking diversity in racial data. Given the heterogeneity of IgAN across different races, our existing classification scheme may not be directly applicable to other populations. However, other ethnic groups could potentially use our approach of unsupervised clustering based on network biomarkers to develop tailored classification schemes suited to their specific racial characteristics. Second, our analyses of the gut microbiome and metabolomics were essentially cross-sectional, which, despite using the KMN stratification scheme providing prognostic insights, does not establish causality. Additionally, the sample size was relatively small, meaning that our findings represented only a preliminary characterization of the gut microbiome and serum metabolites to validate the clinical value of our KMN stratification. Further in-depth studies are needed to explore the underlying mechanisms of these findings.

Conclusions

This study developed a novel risk stratification model for IgAN using network biomarkers and unsupervised clustering. The KMN scheme outperformed existing models in predicting disease progression and guiding personalized treatment. Our findings also highlight the potential role of systemic biomarkers, including serum metabolites and gut microbiota, in IgAN management. Further research is needed to validate these insights and enhance clinical applicability.

Acknowledgments

This study was partially funded by grants from the National Natural Science Foundation of China (grants 82300797 and 81970612), the Sichuan Science and Technology Program (grants 2024NSFSC1500, 2023YFS0275, and 2022YFS0150), and the clinical incubation project of West China Hospital, Sichuan University (grant 21HXFH015).

Data Availability

All primary data supporting our findings are provided in the main text and in Multimedia Appendix 1. Due to privacy and ethical considerations, secondary datasets generated or analyzed during this study are not publicly accessible. However, these datasets can be obtained from the corresponding author upon reasonable request, subject to strict adherence to privacy regulations and ethical guidelines approved by the ethics committee of West China Hospital, Sichuan University. The software code and methodologies used to derive patient labels are available in our GitHub repository, titled Unsupervised Learning for Risk Stratification in IgA Nephropathy Using Network Biomarkers [60]. This repository provides full transparency of our analytical techniques, facilitating further research and validation.

Authors' Contributions

JT and RY contributed equally to this work as joint first authors. They were involved in the conception and design of the study, acquisition of data, analysis and interpretation of data, and drafting the paper. LX and LD participated in the data collection and analysis and contributed to the revision of the manuscript for important intellectual content. ZZ and LZ joined in the data collection, provided technical and material support, and critically reviewed the final version of the manuscript. WQ served as the corresponding author and was responsible for the overall supervision of the project, confirming the integrity of the data analysis, and ensuring that the analyses were appropriately conducted and interpreted. He also played a key role in revising the manuscript critically for important intellectual content and final approval of the version to be published. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures for evaluating the value of KMN stratification in distinguishing IgAN severity, prognosis, treatment guidance, and multi-omics analysis.
[DOCX File , 8072 KB - jmir_v27i1e65563_app1.docx]

References

1. Pattrapornpisut PC, Avila-Casado C, Reich HN. IgA nephropathy: core curriculum 2021. Am J Kidney Dis 2021;78(3):429-441. [doi: 10.1053/j.ajkd.2021.01.024] [Medline: 34247883]
2. Reich HN, Troyanov S, Scholey JW, Cattran DC, Toronto Glomerulonephritis Registry. Remission of proteinuria improves prognosis in IgA nephropathy. J Am Soc Nephrol 2007;18(12):3177-3183. [doi: 10.1681/ASN.2007050526] [Medline: 17978307]
3. Yu GZ, Guo L, Dong J, Shi S, Liu L, Wang J, et al. Persistent hematuria and kidney disease progression in IgA nephropathy: a cohort study. Am J Kidney Dis 2020;76(1):90-99. [doi: 10.1053/j.ajkd.2019.11.008] [Medline: 32197881]
4. Inker LA, Heerspink HJ, Tighiouart H, Chaudhari J, Miao S, Diva U, et al. Association of treatment effects on early change in urine protein and treatment effects on GFR slope in IgA nephropathy: an individual participant meta-analysis. Am J Kidney Dis 2021;78(3):340-349.e1. [doi: 10.1053/j.ajkd.2021.03.007] [Medline: 33775708]
5. Myllymäki J, Honkanen T, Syrjänen J, Helin H, Rantala I, Pasternack A, et al. Uric acid correlates with the severity of histopathological parameters in IgA nephropathy. Nephrol Dial Transplant 2005;20(1):89-95. [doi: 10.1093/ndt/gfh584] [Medline: 15572382]
6. Qin A, Yang D, Wang S, Dong L, Tan J, Tang Y, et al. Uric acid-based ratios for predicting renal failure in Chinese IgA nephropathy patients. Int J Med Sci 2023;20(12):1584-1591. [doi: 10.7150/ijms.85430] [Medline: 37859692]
7. Trimarchi H, Barratt J, Cattran DC, Cook HT, Coppo R, Haas M, et al. Oxford Classification of IgA nephropathy 2016: an update from the IgA Nephropathy Classification Working Group. Kidney Int 2017;91(5):1014-1021 [FREE Full text] [doi: 10.1016/j.kint.2017.02.003] [Medline: 28341274]
8. Wang S, Qin A, Pei G, Jiang Z, Dong L, Tan J, et al. Cigarette smoking may accelerate the progression of IgA nephropathy. BMC Nephrol 2021;22(1):239 [FREE Full text] [doi: 10.1186/s12882-021-02453-4] [Medline: 34187402]
9. Jiang Z, Tan J, Wang S, Dong L, Han X, Tang Y, et al. Lower serum bilirubin is associated with poor renal outcome in IgA nephropathy patients. Int J Med Sci 2021;18(13):2964-2970 [FREE Full text] [doi: 10.7150/ijms.60111] [Medline: 34220324]
10. Tan J, Song G, Wang S, Dong L, Liu X, Jiang Z, et al. Platelet-to-albumin ratio: a novel IgA nephropathy prognosis predictor. Front Immunol 2022;13:842362. [doi: 10.3389/fimmu.2022.842362] [Medline: 35664006]
11. Qin A, Tan J, Wang S, Dong L, Jiang Z, Yang D, et al. Triglyceride-glucose index may predict renal survival in patients with IgA nephropathy. J Clin Med 2022;11(17):5176. [doi: 10.3390/jcm11175176] [Medline: 36079108]
12. Barbour SJ, Coppo R, Zhang H, Liu Z, Suzuki Y, Matsuzaki K, et al. Evaluating a new international risk-prediction tool in IgA nephropathy. JAMA Intern Med 2019;179(7):942-952. [doi: 10.1001/jamainternmed.2019.0600] [Medline: 30980653]
13. Hu L, Fang Y, Huang J, Liu J, Xu L, He W. External validation of the international prognosis prediction model of IgA nephropathy. Ren Fail 2024;46(1):2313174 [FREE Full text] [doi: 10.1080/0886022X.2024.2313174] [Medline: 38345077]
14. Zhang J, Huang B, Liu Z, Wang X, Xie M, Guo R, et al. External validation of the International IgA Nephropathy Prediction Tool. Clin J Am Soc Nephrol 2020;15(8):1112-1120 [FREE Full text] [doi: 10.2215/CJN.16021219] [Medline: 32616496]
15. Koike K, Kawamura T, Hirano K, Nishikawa M, Shimizu A, Joh K, et al. Clinicopathological prognostic stratification for proteinuria and kidney survival in IgA nephropathy: a Japanese prospective cohort study. Clin Kidney J 2024;17(1):sfad294 [FREE Full text] [doi: 10.1093/ckj/sfad294] [Medline: 38213485]
16. Rovin BH, Adler SG, Barratt J, Bridoux F, Burdge KA, Chan TM, et al. Executive summary of the KDIGO 2021 Guideline for the Management of Glomerular Diseases. Kidney Int 2021;100(4):753-779 [FREE Full text] [doi: 10.1016/j.kint.2021.05.015] [Medline: 34556300]
17. Mandelli V, Landi I, Busuoli EM, Courchesne E, Pierce K, Lombardo MV. Prognostic early snapshot stratification of autism based on adaptive functioning. Nat Mental Health 2023;1(5):327-336. [doi: 10.1038/s44220-023-00056-6]
18. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. Biotechnol Adv 2021;49:107739. [doi: 10.1016/j.biotechadv.2021.107739] [Medline: 33794304]
19. Liu R, Wang X, Aihara K, Chen L. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev 2014;34(3):455-478. [doi: 10.1002/med.21293] [Medline: 23775602]
20. Dong L, Tan J, Zhong Z, Tang Y, Qin W. Altered serum metabolic profile in patients with IgA nephropathy. Clin Chim Acta 2023;549:117561. [doi: 10.1016/j.cca.2023.117561] [Medline: 37722576]
21. Haniuda K, Gommerman JL, Reich HN. The microbiome and IgA nephropathy. Semin Immunopathol 2021;43(5):649-656. [doi: 10.1007/s00281-021-00893-6] [Medline: 34664087]

XSL•FO
RenderX

22. He JW, Zhou X, Lv J, Zhang H. Perspectives on how mucosal immune responses, infections and gut microbiome shape IgA nephropathy and future therapies. Theranostics 2020;10(25):11462-11478. [doi: 10.7150/thno.49778] [Medline: 33052226]
23. Park S, Lee J, Yang SH, Lee H, Kim JY, Park M, et al. Comprehensive metabolomic profiling in early IgA nephropathy patients reveals urine glycine as a prognostic biomarker. J Cell Mol Med 2021;25(11):5177-5190 [FREE Full text] [doi: 10.1111/jcmm.16520] [Medline: 33939273]
24. Tan J, Dong L, Jiang Z, Tan L, Luo X, Pei G, et al. Probiotics ameliorate IgA nephropathy by improving gut dysbiosis and blunting NLRP3 signaling. J Transl Med 2022;20(1):382 [FREE Full text] [doi: 10.1186/s12967-022-03585-3] [Medline: 36038927]
25. Zhao J, Bai M, Ning X, Qin Y, Wang Y, Yu Z, et al. Expansion of in gut is associated with the onset and response to immunosuppressive therapy of IgA nephropathy. J Am Soc Nephrol 2022;33(12):2276-2292. [doi: 10.1681/ASN.2022020189] [Medline: 36041791]
26. Zhong Z, Tan J, Tan L, Tang Y, Qiu Z, Pei G, et al. Modifications of gut microbiota are associated with the severity of IgA nephropathy in the Chinese population. Int Immunopharmacol 2020;89(Pt B):107085. [doi: 10.1016/j.intimp.2020.107085] [Medline: 33068859]
27. Moore A. K-means and hierarchical clustering. CMU School of Computer Science. 2001. URL: https://www.cs.cmu.edu/~./awm/tutorials/kmeans.html [accessed 2025-02-08]
28. McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. J Open Source Softw 2018;3(29):861. [doi: 10.21105/joss.00861]
29. Tan J, Yang R, Xiao L, Xia Y, Qin W. Personalized decision support system for tailoring IgA nephropathy treatment strategies. Eur J Intern Med 2024;124:69-77. [doi: 10.1016/j.ejim.2024.02.014] [Medline: 38443263]
30. Cheng R, Wang L, Le S, Yang Y, Zhao C, Zhang X, et al. A randomized controlled trial for response of microbiome network to exercise and diet intervention in patients with nonalcoholic fatty liver disease. Nat Commun 2022;13(1):2555 [FREE Full text] [doi: 10.1038/s41467-022-29968-0] [Medline: 35538056]
31. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2(1):56-67. [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
32. Sheng W, Wang X, Wang Z, Li Q, Zheng Y, Chen S. A differential evolution algorithm with adaptive niching and k-means operation for data clustering. IEEE Trans Cybern 2022;52(7):6181-6195. [doi: 10.1109/TCYB.2020.3035887] [Medline: 33284774]
33. Zhou F, de la Torre F, Hodgins JK. Hierarchical aligned cluster analysis for temporal clustering of human motion. IEEE Trans Pattern Anal Mach Intell 2013;35(3):582-596. [doi: 10.1109/TPAMI.2012.137] [Medline: 22732658]
34. Rauen T, Wied S, Fitzner C, Eitner F, Sommerer C, Zeier M, et al. After ten years of follow-up, no difference between supportive care plus immunosuppression and supportive care alone in IgA nephropathy. Kidney Int 2020;98(4):1044-1052 [FREE Full text] [doi: 10.1016/j.kint.2020.04.046] [Medline: 32450154]
35. Rauen T, Eitner F, Fitzner C, Sommerer C, Zeier M, Otte B, et al. Intensive supportive care plus immunosuppression in IgA nephropathy. N Engl J Med 2015;373(23):2225-2236. [doi: 10.1056/NEJMoa1415463] [Medline: 26630142]
36. Tan J, Dong L, Ye D, Tang Y, Hu T, Zhong Z, et al. The efficacy and safety of immunosuppressive therapies in the treatment of IgA nephropathy: a network meta-analysis. Sci Rep 2020;10(1):6062 [FREE Full text] [doi: 10.1038/s41598-020-63170-w] [Medline: 32269271]
37. Barratt J, Tumlin J, Suzuki Y, Kao A, Aydemir A, Pudota K, et al. Randomized phase II JANUS study of atacicept in patients with IgA nephropathy and persistent proteinuria. Kidney Int Rep 2022;7(8):1831-1841 [FREE Full text] [doi: 10.1016/j.ekir.2022.05.017] [Medline: 35967104]
38. Lv J, Liu L, Hao C, Li G, Fu P, Xing G, et al. Randomized phase 2 trial of telitacicept in patients with IgA nephropathy with persistent proteinuria. Kidney Int Rep 2023;8(3):499-506 [FREE Full text] [doi: 10.1016/j.ekir.2022.12.014] [Medline: 36938094]
39. Heerspink HJL, Radhakrishnan J, Alpers CE, Barratt J, Bieler S, Diva U, et al. Sparsentan in patients with IgA nephropathy: a prespecified interim analysis from a randomised, double-blind, active-controlled clinical trial. Lancet 2023;401(10388):1584-1594. [doi: 10.1016/S0140-6736(23)00569-X] [Medline: 37015244]
40. Rovin BH, Barratt J, Heerspink HJL, Alpers CE, Bieler S, Chae DW, et al. Efficacy and safety of sparsentan versus irbesartan in patients with IgA nephropathy (PROTECT): 2-year results from a randomised, active-controlled, phase 3 trial. Lancet 2023;402(10417):2077-2090. [doi: 10.1016/S0140-6736(23)02302-4] [Medline: 37931634]
41. Wheeler DC, Toto RD, Stefánsson BV, Jongs N, Chertow GM, Greene T, et al. A pre-specified analysis of the DAPA-CKD trial demonstrates the effects of dapagliflozin on major adverse kidney events in patients with IgA nephropathy. Kidney Int 2021;100(1):215-224 [FREE Full text] [doi: 10.1016/j.kint.2021.03.033] [Medline: 33878338]
42. Lee LC, Liong CY, Jemain AA. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. Analyst 2018;143(15):3526-3539. [doi: 10.1039/c8an00599k] [Medline: 29947623]

XSL•FO
RenderX

43. Boran H, Terzi S. Stress-induced transcriptional changes and DNA damage associated with bis(2-ethylhexyl) adipate exposure in zebrafish (Danio rerio) larvae. Bull Environ Contam Toxicol 2017;99(3):308-314. [doi: 10.1007/s00128-017-2116-4] [Medline: 28536799]
44. Ong HT, Samsudin H, Soto-Valdez H. Migration of endocrine-disrupting chemicals into food from plastic packaging materials: an overview of chemical risk assessment, techniques to monitor migration, and international regulations. Crit Rev Food Sci Nutr 2022;62(4):957-979. [doi: 10.1080/10408398.2020.1830747] [Medline: 33081493]
45. Malarvannan G, Onghena M, Verstraete S, van Puffelen E, Jacobs A, Vanhorebeek I, et al. Phthalate and alternative plasticizers in indwelling medical devices in pediatric intensive care units. J Hazard Mater 2019;363:64-72. [doi: 10.1016/j.jhazmat.2018.09.087] [Medline: 30308366]
46. Duan C, Fang Y, Sun J, Li Z, Wang Q, Bai J, et al. Effects of fast food packaging plasticizers and their metabolites on steroid hormone synthesis in H295R cells. Sci Total Environ 2020;726:138500. [doi: 10.1016/j.scitotenv.2020.138500] [Medline: 32334352]
47. Abdelkader H, Longman M, Alany R, Pierscionek B. Phytosome-hyaluronic acid systems for ocular delivery of L-carnosine. Int J Nanomedicine 2016;11:2815-2827 [FREE Full text] [doi: 10.2147/IJN.S104774] [Medline: 27366062]
48. Colucci S, Mori G, Vaira S, Brunetti G, Greco G, Mancini L, et al. L-carnitine and isovaleryl L-carnitine fumarate positively affect human osteoblast proliferation and differentiation in vitro. Calcif Tissue Int 2005;76(6):458-465. [doi: 10.1007/s00223-004-0147-4] [Medline: 15906015]
49. Tortorelli S, Hahn S, Cowan T, Brewster T, Rinaldo P, Matern D. The urinary excretion of glutarylcarnitine is an informative tool in the biochemical diagnosis of glutaric acidemia type I. Mol Genet Metab 2005;84(2):137-143. [doi: 10.1016/j.ymgme.2004.09.016] [Medline: 15670719]
50. Guo C, You Z, Shi H, Sun Y, Du X, Palacios G, et al. SLC38A2 and glutamine signalling in cDC1s dictate anti-tumour immunity. Nature 2023;620(7972):200-208 [FREE Full text] [doi: 10.1038/s41586-023-06299-8] [Medline: 37407815]
51. Martínez YB, Ferreira FV, Musumeci MA. L-Glutamine-, peptidyl- and protein-glutaminases: structural features and applications in the food industry. World J Microbiol Biotechnol 2022;38(11):204. [doi: 10.1007/s11274-022-03391-5] [Medline: 36002753]
52. White WJ, Lawrie RA. Variations in the levels of 3-methyl-l-histidine of the myosins within the bovine carcass. Meat Sci 1985;15(3):173-181. [doi: 10.1016/0309-1740(85)90035-x]
53. Li F, Wang M, Wang J, Li R, Zhang Y. Alterations to the gut microbiota and their correlation with inflammatory factors in chronic kidney disease. Front Cell Infect Microbiol 2019;9:206 [FREE Full text] [doi: 10.3389/fcimb.2019.00206] [Medline: 31245306]
54. Crossland NA, Beck S, Tan WY, Lo M, Mason JB, Zhang C, et al. Fecal microbiota transplanted from old mice promotes more colonic inflammation, proliferation, and tumor formation in azoxymethane-treated A/J mice than microbiota originating from young mice. Gut Microbes 2023;15(2):2288187 [FREE Full text] [doi: 10.1080/19490976.2023.2288187] [Medline: 38031252]
55. Zhao Y, Zhao M, Zhang Y, Fu Z, Jin T, Song J, et al. Bile acids metabolism involved in the beneficial effects of Danggui Shaoyao San via gut microbiota in the treatment of CCl(4) induced hepatic fibrosis. J Ethnopharmacol 2024;319(Pt 3):117383. [doi: 10.1016/j.jep.2023.117383] [Medline: 37925004]
56. Kaakoush NO. Sutterella species, IgA-degrading bacteria in ulcerative colitis. Trends Microbiol 2020;28(7):519-522. [doi: 10.1016/j.tim.2020.02.018] [Medline: 32544438]
57. Charalampidis D. A modified k-means algorithm for circular invariant clustering. IEEE Trans. Pattern Anal. Machine Intell 2005;27(12):1856-1865. [doi: 10.1109/tpami.2005.230]
58. Lai KN. IgA nephropathy. Nat Rev Dis Primers 2016;2(1):16002. [doi: 10.1038/nrdp.2016.2] [Medline: 27227656]
59. Floege J, Feehally J. The mucosa-kidney axis in IgA nephropathy. Nat Rev Nephrol 2016;12(3):147-156. [doi: 10.1038/nrneph.2015.208] [Medline: 26714580]
60. Yang R. Unsupervised learning for risk stratification in IgA nephropathy using network biomarkers. GitHub. URL: https://github.com/yangrongxinuser/Unsupervised-Learning-for-Risk-Stratification-in-IgA-Nephropathy-Using-Network-Biomarkers [accessed 2025-01-14]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor
ARB: angiotensin receptor blocker
AUC: area under the curve
DEHA: 3-dehydroepiandrosterone sulfate
eGFR: estimated glomerular filtration rate
HR: hazard ratio
IgAN: immunoglobulin A nephropathy
IIgAN-PT: International Immunoglobulin A Nephropathy Prediction Tool
IST: immunosuppressive therapy

XSL•FO
RenderX

KDIGO: Kidney Disease: Improving Global Outcomes
KMN: k-means clustering based on network biomarkers
PLS-DA: partial least-squares discrimination
RF-RG: renal function–renal histology grading
rRNA: ribosomal ribonucleic acid
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VIP: variable importance in projection

Edited by A Schwartz; submitted 19.08.24; peer-reviewed by XJ Zhou, M Oshima; comments to author 30.09.24; revised version received 14.10.24; accepted 29.01.25; published 10.03.25.

Please cite as:
Tan J, Yang R, Xiao L, Dong L, Zhong Z, Zhou L, Qin W
Risk Stratification in Immunoglobulin A Nephropathy Using Network Biomarkers: Development and Validation Study
J Med Internet Res 2025;27:e65563
URL: https://www.jmir.org/2025/1/e65563
doi:10.2196/65563
PMID:40063072

©Jiaxing Tan, Rongxin Yang, Liyin Xiao, Lingqiu Dong, Zhengxia Zhong, Ling Zhou, Wei Qin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Two-Year Hypertension Incidence Risk Prediction in Populations in the Desert Regions of Northwest China: Prospective Cohort Study

Yinlin Cheng[1*], PhD; Kuiying Gu[2*], PhD; Weidong Ji[1], PhD; Zhensheng Hu[1], PhD; Yining Yang[3], PhD; Yi Zhou[1], PhD

[1]Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

[2]School of Public Health, Soochow University, Suzhou, China

[3]People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China

[*]these authors contributed equally

Corresponding Author:
Yi Zhou, PhD
Zhongshan School of Medicine
Sun Yat-sen University
74 Zhongshan 2nd Road
Yuexiu District
Guangzhou, 510080
China
Phone: 86 020 87332139
Email: zhouyi@mail.sysu.edu.cn

Abstract

Background: Hypertension is a major global health issue and a significant modifiable risk factor for cardiovascular diseases, contributing to a substantial socioeconomic burden due to its high prevalence. In China, particularly among populations living near desert regions, hypertension is even more prevalent due to unique environmental and lifestyle conditions, exacerbating the disease burden in these areas, underscoring the urgent need for effective early detection and intervention strategies.

Objective: This study aims to develop, calibrate, and prospectively validate a 2-year hypertension risk prediction model by using large-scale health examination data collected from populations residing in 4 regions surrounding the Taklamakan Desert of northwest China.

Methods: We retrospectively analyzed the health examination data of 1,038,170 adults (2019-2021) and prospectively validated our findings in a separate cohort of 961,519 adults (2021-2023). Data included demographics, lifestyle factors, physical examinations, and laboratory measurements. Feature selection was performed using light gradient-boosting machine–based recursive feature elimination with cross-validation and Least Absolute Shrinkage and Selection Operator, yielding 24 key predictors. Multiple machine learning (logistic regression, random forest, extreme gradient boosting, light gradient-boosting machine) and deep learning (Feature Tokenizer + Transformer, SAINT) models were trained with Bayesian hyperparameter optimization.

Results: Over a 2-year follow-up, 15.20% (157,766/1,038,170) of the participants in the retrospective cohort and 10.50% (101,077/961,519) in the prospective cohort developed hypertension. Among the models developed, the CatBoost model demonstrated the best performance, achieving area under the curve (AUC) values of 0.888 (95% CI 0.886-0.889) in the retrospective cohort and 0.803 (95% CI 0.801-0.804) in the prospective cohort. Calibration via isotonic regression improved the model's probability estimates, with Brier scores of 0.090 (95% CI 0.089-0.091) and 0.102 (95% CI 0.101-0.103) in the internal validation and prospective cohorts, respectively. Participants were ranked by the positive predictive value calculated using the calibrated model and stratified into 4 risk categories (low, medium, high, and very high), with the very high group exhibiting a 41.08% (5741/13,975) hypertension incidence over 2 years. Age, BMI, and socioeconomic factors were identified as significant predictors of hypertension.

Conclusions: Our machine learning model effectively predicted the 2-year risk of hypertension, making it particularly suitable for preventive health care management in high-risk populations residing in the desert regions of China. Our model exhibited excellent predictive performance and has potential for clinical application. A web-based application was developed based on our

XSL•FO
RenderX

predictive model, which further enhanced the accessibility for clinical and public health use, aiding in reducing the burden of hypertension through timely prevention strategies.

(J Med Internet Res 2025;27:e68442) doi:10.2196/68442

KEYWORDS

hypertension; desert; machine learning; deep learning; prevention; clinical applicability

Introduction

Cardiovascular diseases are among the leading causes of chronic noncommunicable diseases worldwide and have become the primary cause of death globally [1]. Hypertension, a major preventable risk factor for cardiovascular disease, accounts for approximately 50% of all cardiovascular-related deaths globally [2]. In China, the prevalence of hypertension has been rising due to rapid urbanization, increasing affluence, and an aging population. Recent surveys estimate that approximately 244.5 million Chinese adults (23.2%) are affected by hypertension, with this number continuing to increase [3,4]. In northwestern China, populations residing in the Taklamakan Desert region face unique public health challenges, as harsh environmental conditions such as extreme temperature fluctuations, frequent sandstorms, particulate matter pollution, and limited greenspace are associated with an elevated prevalence of hypertension [5-7]. Given the increasing burden of hypertension, there is an urgent need for effective tools to identify high-risk individuals early and implement preventive measures.

Numerous hypertension risk prediction models have been developed using electronic health records based on a variety of methodological approaches, including traditional statistical techniques such as Cox regression and logistic regression, as well as machine learning and deep learning methods [8,9]. Although these models have demonstrated promising discriminatory power, several limitations persist. First, many studies have constructed models using small sample sizes or data from a single medical center, thereby limiting the external validity and generalizability of these models [10-14]. Second, a significant proportion of models lack validation on prospective datasets or independent external datasets, which impedes the comprehensive assessment of model robustness [12,15-18]. Additionally, some models incorporate an excessive number of variables or include variables that are not easily accessible in health examinations, thus hindering their practical application [19,20]. Finally, the majority of these models have been developed using data from European and American populations, thereby limiting their applicability to populations in other regions, including Asia.

Therefore, the objective of this study was to develop and prospectively validate a 2-year hypertension incidence risk prediction model by using health examination data collected from populations residing in 4 regions surrounding the Taklamakan Desert between 2019 and 2023. Our model aims to stratify individuals by their risk of developing hypertension, thus facilitating targeted prevention and early intervention strategies.

Methods

Data Source and Study Design

This retrospective cohort study utilized health examination data from adults collected between 2019 and 2021 at 1750 hospitals and community clinics across 4 regions surrounding the Taklamakan Desert of northwest China. Participants' health records were categorized into 3 components: personal and lifestyle information, standard physical examinations, and clinical measurements. Personal and lifestyle data included demographic details, personal and family medical histories, smoking and alcohol consumption habits, dietary patterns, and physical activity levels. Standard physical examinations recorded BMI, waist circumference, heart rate, and blood pressure, while clinical measurements assessed blood biochemical parameters.

A total of 2,872,837 participants were initially included. After applying the following exclusion criteria, that is, (1) individuals younger than 18 years, (2) individuals with a prior diagnosis of hypertension, and (3) participants with incomplete follow-up data, the final study cohort comprised 1,038,170 participants. For external validation, data from a prospective cohort (2021-2023) were employed, following the exclusion criteria outlined in Figure 1. Variables with more than 30% missing data were excluded from the analysis, following the methodology used in a previous study [21]. For variables with fewer missing values, the random forest algorithm was used for imputation. Multimedia Appendices 1 and 2 provide detailed information on the percentage of missing values for each variable in the retrospective and prospective cohorts prior to imputation.

XSL•FO
RenderX

Figure 1. Flowchart illustrating the exclusion criteria, missing data handling, feature selection, model building (2019-2021), and validation (2021-2023) for the 2-year hypertension risk prediction study across 4 Taklamakan Desert–adjacent regions in northwest China.

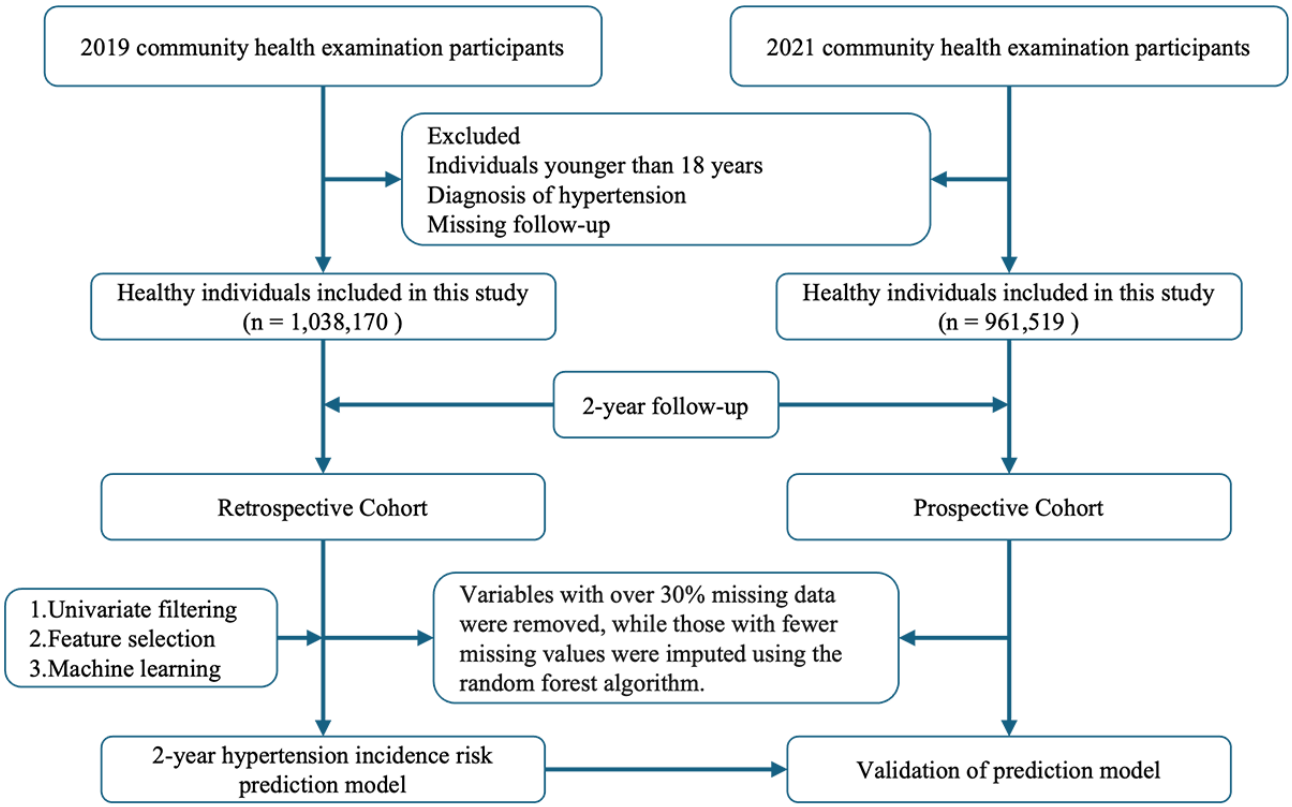

Ethics Approval

This study was approved by the Tsinghua University Science and Technology ethics committee (Medicine; project: 20240123). The data used in this research were collected and managed by the Xinjiang Uygur Autonomous Region Health Commission and were fully anonymized and deidentified before being accessed. According to ethical guidelines in China (Section 3.32), the secondary use of anonymized health data for research purposes is exempt from requiring additional informed consent from individual patients [22]. This exemption aligns with international research practices [23]. Researchers accessed the anonymized data through a designated, secure platform without internet connectivity, ensuring strict data privacy and security protocols. Neither the manuscript nor the supplementary materials contain identifiable information or images of participants, and no financial compensation was involved, as this study was based on secondary data analysis.

Definition of Hypertension

Blood pressure was measured by trained health care professionals by using standardized protocols. The average of 3 consecutive right-arm measurements, taken 30 seconds apart, was used for analysis. Hypertension was defined by any of the following criteria: (1) self-reported diagnosis of hypertension, (2) current use of antihypertensive medication, or (3) an average systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or both [24,25].

Statistical Analysis

The Kolmogorov-Smirnov test was used to assess whether continuous variables followed a normal distribution. To compare baseline characteristics between participants with and without hypertension, we applied the chi-square test for categorical variables, while continuous variables were analyzed using either the independent-sample *t* test (2-sided) or the rank-sum test, depending on the distribution characteristics.

Prediction Modeling and Evaluation

Participants from the retrospective cohort were randomly assigned to training and internal validation sets in a 7:3 ratio. Feature selection commenced with univariate logistic regression analyses to identify independent risk factors. Subsequently, both light gradient-boosting machine (LightGBM)–based recursive feature elimination with cross-validation (RFECV) [26,27] and Least Absolute Shrinkage and Selection Operator (LASSO) methods were applied to further refine the feature set. The final predictive features were determined by selecting the intersection of variables identified by both LightGBM-RFECV and LASSO.

The final model was built using CatBoost, a high-performance gradient boosting algorithm [28]. Additionally, several other models, including logistic regression, random forest, extreme gradient boosting [29], LightGBM, and 2 deep learning methods (Feature Tokenizer + Transformer [30] and SAINT [31]), were constructed to compare predictive performance for hypertension incidence. Bayesian optimization via *Hyperopt* [32] was employed to fine-tune the hyperparameters, maximizing the area under the receiver operating characteristic (AUROC) curve across 5-fold cross-validation. Each model underwent 1000 optimization trials. The detailed parameters for each model are provided in Multimedia Appendix 3.

XSL•FO
RenderX

The optimal model was evaluated using AUROC, average precision, accuracy, sensitivity, specificity, and confusion matrix in the internal validation set. Calibration of the selected model on the training set was performed using isotonic regression [33] to improve predictive accuracy and reliability. The calibrated model was subsequently transformed to both the internal validation set and the prospective cohort. Calibration curves, constructed with 25 evenly spaced bins, were then used to evaluate the model's calibration and predictive performance across both validation sets.

To further refine risk stratification, we calculated the positive predictive values for individuals in the prospective cohort by using the calibrated model. Individuals were ranked by positive predictive values and categorized into 4 risk levels, ranging from low to very high risk. Univariate Cox regression was performed to validate the effectiveness of these risk categories.

Model Interpretation

Shapley Additive Explanations (SHAP) [34] was employed to interpret the contribution of each feature to the model's predictions. For highly weighted features such as age, gender, and BMI, we created subpopulations to analyze the distribution of risk categories within these groups. All analyses were conducted using Python software (version 3.8.0) with packages, including *CatBoost* (version 1.2.5), *Hyperopt* (version 0.2.7), and *SHAP* (version 0.44.1), as well as R software (version 4.0.2; R Foundation for Statistical Computing). A 2-tailed *P* value less than .05 was considered statistically significant.

Results

Clinical Baseline Characteristics

Table 1 outlines the baseline characteristics of the study cohorts. After applying the exclusion criteria, the retrospective cohort included 1,038,170 participants, aged 18-100 years, with baseline characteristics recorded between January and November 2019. The prospective cohort comprised 961,519 participants within the same age range, with baseline characteristics collected from January to November 2021. Over the 2-year follow-up period, the incidence of hypertension was 15.20% (157,766/1,038,170) in the retrospective cohort and 10.50% (101,077/961,519) in the prospective cohort. At baseline, compared to the nonhypertensive group, the hypertensive group had a higher proportion of males, resided in urban areas, and had lower education levels. The hypertensive group also had higher prevalence of hepatic steatosis and type 2 diabetes along with elevated waist circumference, BMI, systolic blood pressure, and diastolic blood pressure. Multimedia Appendices 4 and 5 provide a detailed comparison of the characteristics between individuals with and without incident hypertension in both cohorts.

XSL•FO
RenderX

Table 1. Clinical baseline information of the retrospective and prospective cohorts[a].

Characteristics	Retrospective cohort (n=1,038,170)	Prospective cohort (n=961,519)
Age at baseline (years), mean (SD)	40.11 (14.73)	42.77 (14.45)
Sex, n (%)		
Male	458,964 (44.21)	431,824 (44.91)
Female	579,206 (55.79)	529,695 (55.09)
Residence, n (%)		
Rural	915,254 (88.16)	865,464 (90.01)
Urban	122,916 (11.84)	96,055 (9.99)
Waist circumference (cm), mean (SD)	83.37 (10.65)	85.00 (11.21)
BMI (kg/m^2), mean (SD)	23.80 (3.67)	24.64 (3.82)
Educational level, n (%)		
Illiterate or semiliterate	42,801 (4.12)	40,876 (4.25)
Primary school	381,811 (36.78)	376,961 (39.2)
Junior middle school	480,768 (46.31)	410,493 (42.69)
Senior middle school	87,775 (8.45)	85,923 (8.94)
College degree and above	45,015 (4.34)	47,266 (4.92)
Exercise frequency, n (%)		
Never	1,000,327 (96.35)	922,358 (95.93)
Occasionally	7592 (0.73)	16,839 (1.75)
Often	30,251 (2.92)	22,322 (2.32)
Dietary patterns, n (%)		
Meat and vegetable balanced	939,046 (90.45)	936,106 (97.36)
Meat-based	52,539 (5.06)	13,578 (1.41)
Vegetarian-based	46,585 (4.49)	11,835 (1.23)
Smoking status, n (%)		
Never	889,175 (85.65)	861,763 (89.63)
Smoking	137,212 (13.22)	96,374 (10.02)
Quit smoking	11,783 (1.13)	3382 (0.35)
Alcohol intake, n (%)		
Never	923,357 (88.94)	894,674 (93.05)
Occasionally	102,187 (9.84)	60,485 (6.29)
Often	12,626 (1.22)	6360 (0.66)
Heart rate (bpm), mean (SD)	73.59 (10.13)	74.30 (9.41)
Systolic blood pressure (mm Hg), mean (SD)	109.03 (12.25)	111.55 (11.66)
Diastolic blood pressure (mm Hg), mean (SD)	66.41 (8.45)	67.81 (8.01)
Hemoglobin (g/L), mean (SD)	139.76 (17.51)	141.38 (17.47)
White blood cell count (10^9/L), mean (SD)	6.40 (1.44)	6.44 (1.49)
Alanine aminotransferase (U/L), mean (SD)	21.07 (9.02)	20.94 (9.20)
Aspartate transaminase (U/L), mean (SD)	22.27 (6.51)	21.90 (6.81)
Serum creatinine (μmol/L), mean (SD)	66.08 (17.64)	66.59 (17.78)
Total cholesterol (mmol/L), mean (SD)	4.07 (0.87)	4.14 (0.90)
HDL-C[a] (mmol/L), mean (SD)	1.31 (0.33)	1.28 (0.34)

XSL•FO
RenderX

Characteristics	Retrospective cohort (n=1,038,170)	Prospective cohort (n=961,519)
LDL-C[b] (mmol/L), mean (SD)	2.32 (0.75)	2.25 (0.84)
Hepatic steatosis, n (%)		
No	999,729 (96.3)	932,220 (96.95)
Yes	38,441 (3.7)	29,299 (3.05)
Type 2 diabetes, n (%)		
No	1,002,725 (96.59)	921,947 (95.88)
Yes	35,445 (3.41)	39,572 (4.12)
Family history of hypertension, n (%)		
No	952,605 (91.76)	883,277 (91.86)
Yes	85,565 (8.24)	78,242 (8.14)

[a]HDL-C: high-density lipoprotein cholesterol.

[b]LDL-C: low-density lipoprotein cholesterol.

Development and Validation of Predictive Models

Recursive feature elimination is a widely used method for feature selection. It iteratively trains a model, assesses the importance of each feature, and removes the least important ones until the optimal number of features is identified. In this study, we used an enhanced version of recursive feature elimination, RFECV, which incorporates cross-validation to evaluate model performance at each iteration. Additionally, LASSO regression was employed, using L1 regularization to shrink the coefficients of less important features to zero, thereby further refining the feature set (Figure 2). Through the application of both LightGBM-RFECV and LASSO, we identified 26 and 27 potential risk predictors for hypertension, respectively. By intersecting the results of these 2 methods, we selected 24 key features for the final construction of the 2-year hypertension incidence risk prediction model.

Figure 2. Feature selection of the model. (A) Feature selection results based on light gradient-boosting machine–based recursive feature elimination with cross-validation. (B) Feature selection results based on Least Absolute Shrinkage and Selection Operator. AUROC: area under the receiver operating characteristic; LightGBM: light gradient-boosting machine; RFECV: recursive feature elimination with cross-validation.

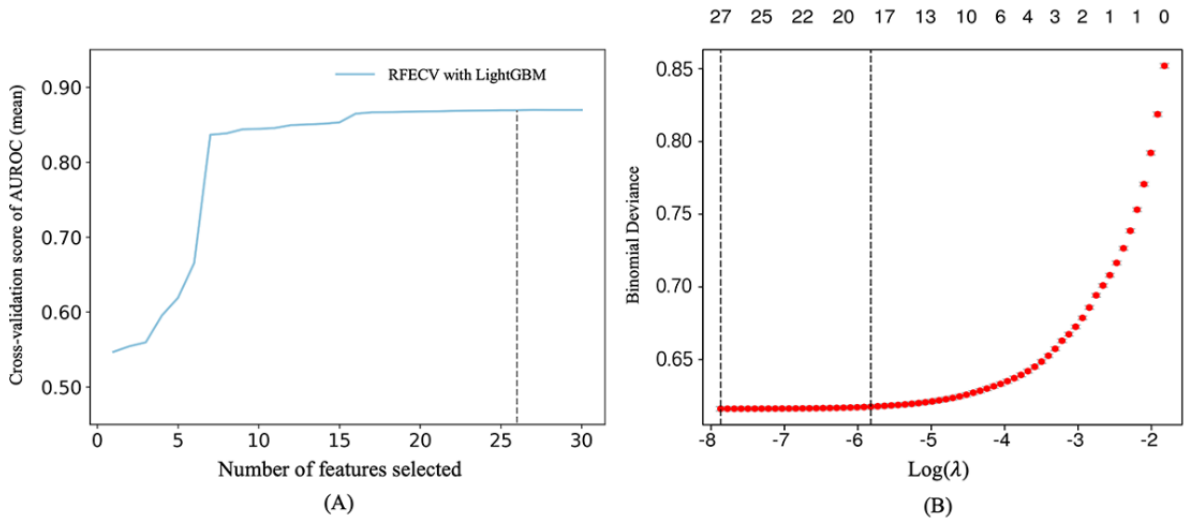

Model Performance

We evaluated the performance of several machine learning and deep learning models. Among the models, CatBoost demonstrated the highest predictive accuracy, with an AUROC curve of 0.888 (95% CI 0.886-0.889). Extreme gradient boosting followed with an AUROC of 0.868 (95% CI 0.866-0.870). Detailed AUROC comparisons for each model are available in Multimedia Appendix 6, while the confusion matrix results are provided in Multimedia Appendix 7. A Mann-Whitney *U* test confirmed that CatBoost significantly outperformed the other algorithms in terms of predictive accuracy (P<.001) (Multimedia Appendix 8) [35]. Further validation of the CatBoost model was performed using the prospective cohort. Compared with the results from the internal validation set, the AUROC for the prospective cohort decreased to 0.803 (95% CI 0.801-0.804) (Figure 3). Isotonic regression was applied to calibrate the output probabilities, improving the accuracy of predicted probabilities in reflecting actual risks. After calibration, the model's performance metrics improved, with a Brier score of 0.090 (95% CI 0.089-0.091) on the internal validation set and 0.102 (95% CI 0.101-0.103) on the prospective cohort.

XSL•FO
RenderX

Figure 3. (A) Receiver operating characteristic curve of the CatBoost model in the retrospective and prospective cohorts. (B) Noncalibrated and calibrated curves of the CatBoost model in the retrospective cohort. (C) Noncalibrated and calibrated curves of the CatBoost model in the prospective cohort. (D) Survival curves for time-to-hypertension across 4 risk categories. AUC: area under the curve; HR: hazard ratio.

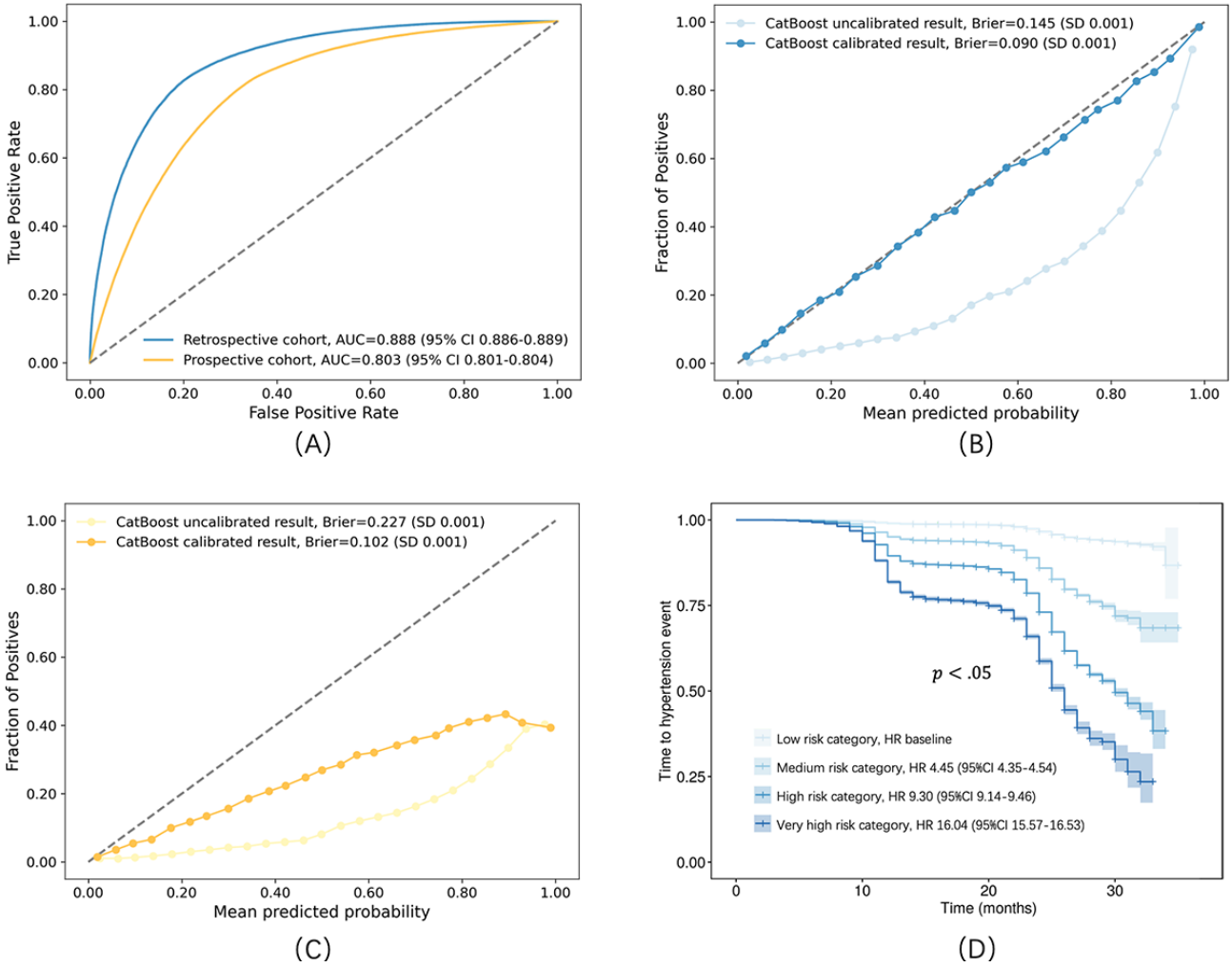

Following calibration, we calculated the positive predictive values for each individual and categorized participants into 4 risk levels: low, medium, high, and very high. In the low-risk group (risk score 0-0.15, n=570,742), only 2.87% (16,392/570,742) developed hypertension over the 2-year period. In contrast, 41.08% (5741/13,975) of the individuals in the very high-risk group (risk score 0.8-1, n=13,975) were diagnosed with hypertension within the same period (Multimedia Appendix 9). To further assess the differences in hypertension onset across the 4 risk categories, we performed univariate Cox regression to estimate time-to-hypertension for each group. The results showed significant differences between the risk categories (P<.05), with the hazard ratio for the very high-risk group being 16.4 (95% CI 15.57-16.53) compared with the low-risk group.

Feature Importance and SHAP Analysis

We used the distribution of SHAP values and the corresponding mean absolute SHAP values within the prospective cohort to interpret the contribution of each feature in the risk prediction model. The analysis revealed that the model's core predictive power derived primarily from basic biometrics (age, BMI, gender, and waist circumference), blood pressure indicators (systolic and diastolic blood pressure), and heart rate (Figure 4).

XSL•FO
RenderX

Figure 4. (A) Complete distribution of Shapley Additive Explanations values for all features in the prospective dataset. (B) Absolute contribution of all features based on Shapley Additive Explanations values for all features in the prospective dataset. AI: alcohol intake; ALT: alanine aminotransferase; AST: aspartate transaminase; DBP: diastolic blood pressure; DP: dietary pattern; EF: exercise frequency; EL: educational level; FHTN: family history of hypertension; Hb: hemoglobin; HDL-C: high-density lipoprotein cholesterol; HR: heart rate; HS: hepatic steatosis; LDL-C: low-density lipoprotein cholesterol; SBP: systolic blood pressure; SCr: serum creatinine; SHAP: Shapley Additive Explanations; SS: smoking status; T2DM: type 2 diabetes; TC: total cholesterol; WBC: white blood cell; WC: waist circumference.

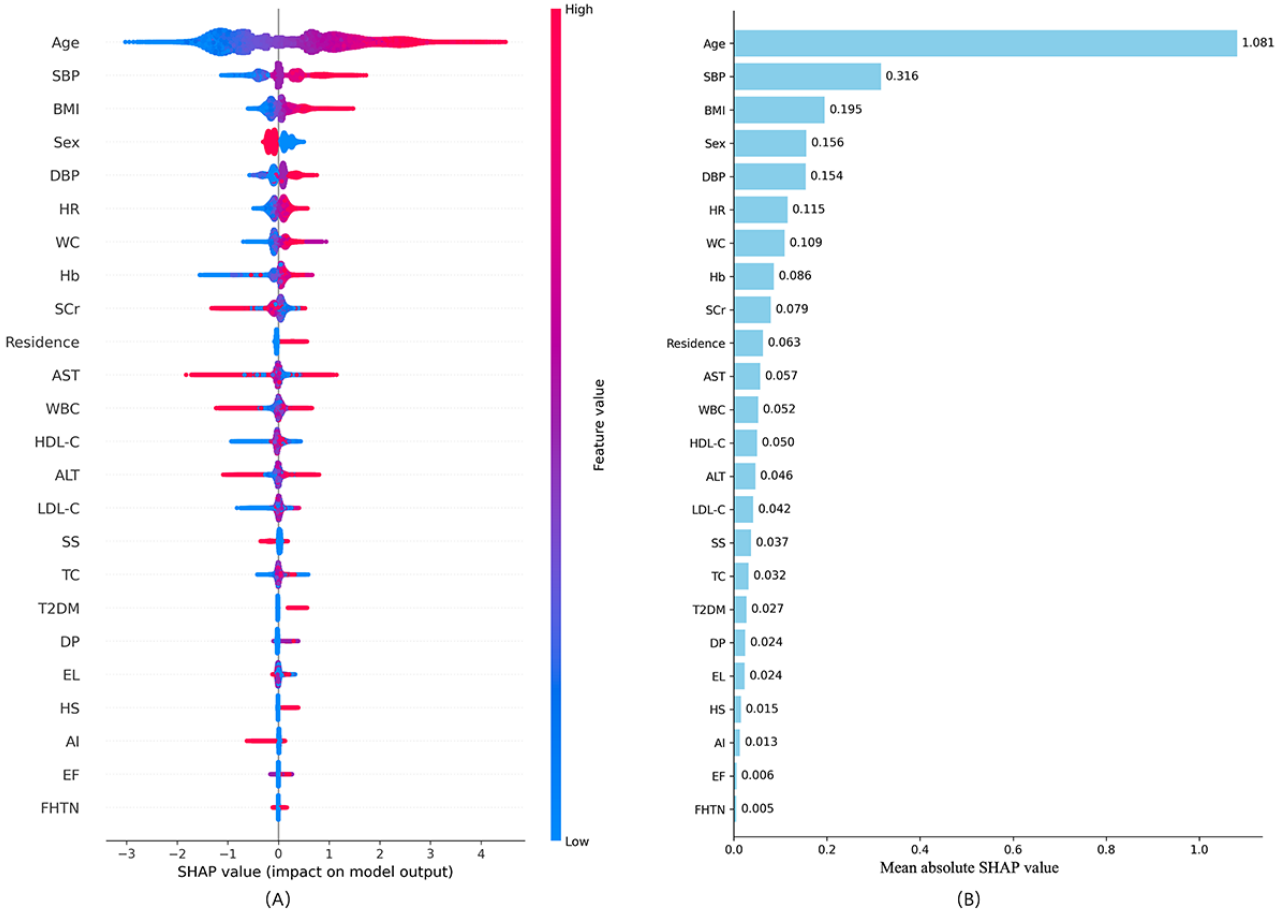

To better understand the impact of individual variables, we analyzed the SHAP values for the top 6 ranked continuous variables (Figure 5). We observed that as these feature values increased, their SHAP values transitioned from negative to positive, indicating that higher values significantly increased the likelihood of being classified as high-risk for hypertension.

XSL•FO
RenderX

Figure 5. (A)-(F) Shapley Additive Explanations value dependence plots for 6 example features in the prospective dataset. DBP: diastolic blood pressure; HR: heart rate; SBP: systolic blood pressure; SHAP: Shapley Additive Explanations; WC: waist circumference.

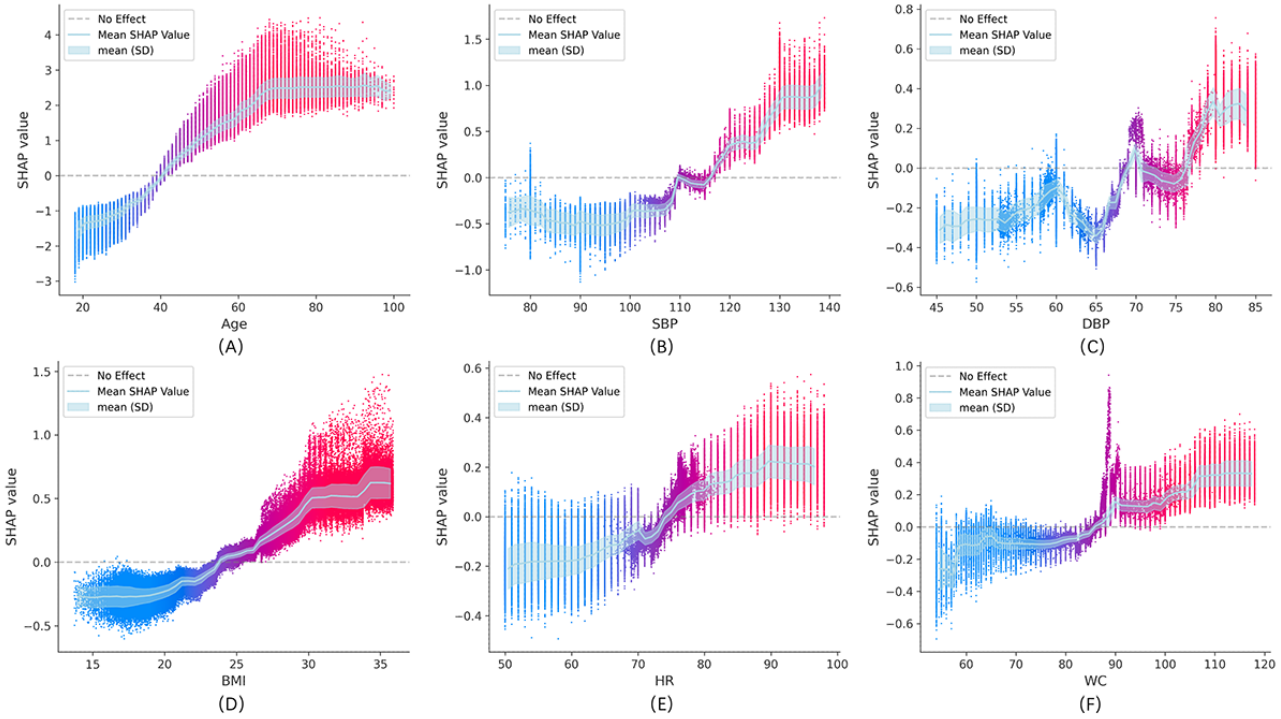

Significant Features

In addition to analyzing SHAP values for individual features, we conducted a grouped analysis of the most influential basic biometrics (age, sex, and BMI) and socioeconomic factors (residence and education level) within the prospective cohort to assess the distribution of the samples across different risk levels and the model's predictive performance within these feature groups.

Based on prior research [19], we divided the population into 4 age categories. The analysis showed that the young population was predominantly in the low-risk category (343,394/347,996, 98.68%), while the older adult population was primarily composed of high-risk (68,238/77,664, 87.86%) and very high-risk (7939/77,664, 10.22%) individuals (Figure 6). Hypertension incidence within 2 years in the older adults (22,277/77,664, 28.68%) was significantly higher than that in the younger population (7154/347,996, 2.05%), confirming age as a key predictor of hypertension onset. Further analysis (Figure 6B) demonstrated that the model performed best in the 35-49 years age group, with an AUROC of 0.734 (95% CI 0.732-0.737), whereas performance was relatively weaker in the older adult population, with an AUROC of 0.589 (95% CI 0.585-0.593). This performance disparity may be attributed to the model's tendency to assign higher risk to older individuals, reducing its discriminatory power in this group.

XSL•FO
RenderX

Figure 6. (A) Proportion of 4 risk categories across different age group subsets. (B) Area under the receiver operating characteristic curve values evaluated for different age group subsets. AUROC: area under the receiver operating characteristic curve; HTN: hypertension.

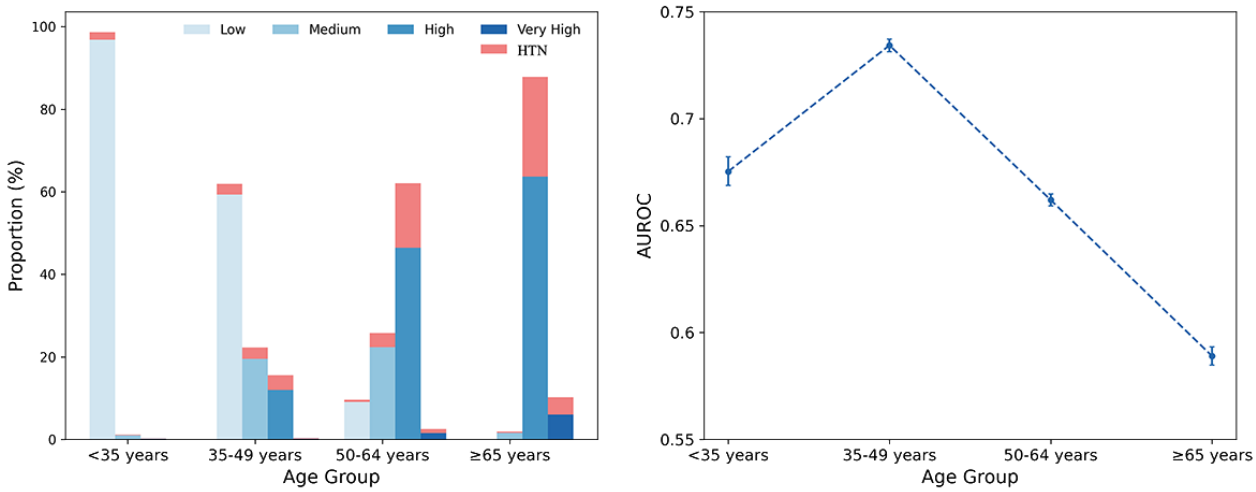

Regarding gender, hypertension incidence within 2 years in women (49,855/529,695, 9.41%) was lower than that in men (51,222/431,824, 11.86%). A higher proportion of women were in the low-risk category, that is, 62.51% (331,122/529,695) of the women were in the low-risk category, while only 55.49% (239,620/431,824) of the men were in the low-risk category (Figure 7). Conversely, men were more likely to be in the high-risk (120,676/431,824, 27.95%) and very high-risk (7821/431,824, 1.81%) categories. Additionally, the model performed better in women, achieving an AUROC of 0.821 (95% CI 0.820-0.823) compared to men with an AUROC of 0.785 (95% CI 0.777-0.781).

Figure 7. (A) Proportion of 4 risk categories across different sex group subsets. (B) Area under the receiver operating characteristic curve values evaluated for male and female subgroups. AUROC: area under the receiver operating characteristic curve; HTN: hypertension.

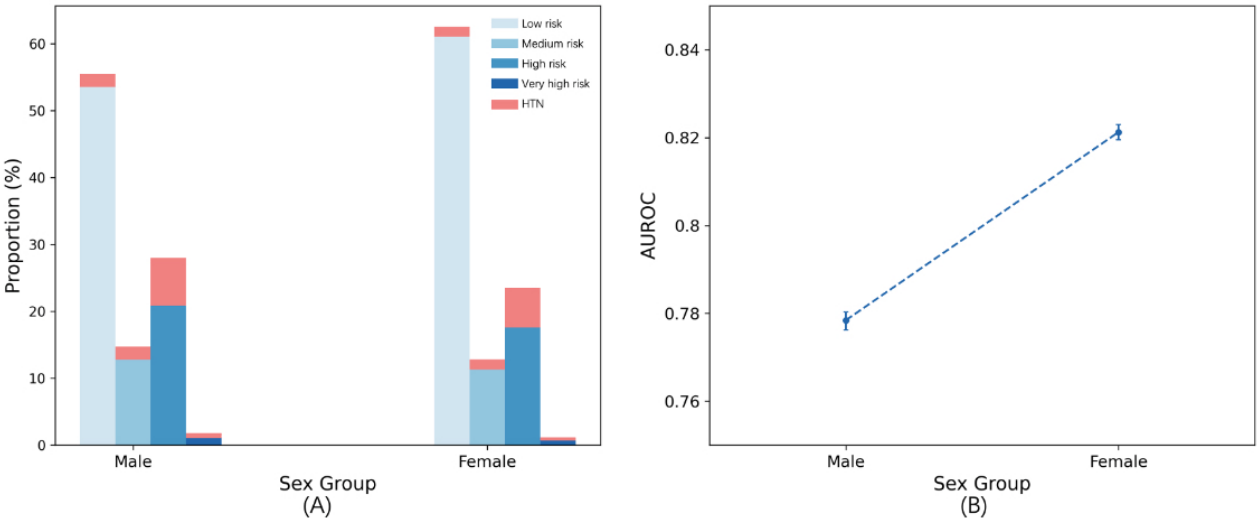

BMI was categorized according to the World Health Organization classification: underweight (BMI<18.5), normal weight (18.5≤BMI<25), overweight (25≤BMI<30), and obesity (BMI≥30). As BMI increased, the proportion of individuals in the high-risk and very high-risk categories also increased along with a corresponding increase in hypertension incidence. This indicates a strong association between obesity and hypertension risk (Figure 8).

XSL•FO
RenderX

Figure 8. (A) Proportion of 4 risk categories across different BMI group subsets. (B) Area under the receiver operating characteristic curve; values evaluated for different BMI group subsets. AUROC: area under the receiver operating characteristic curve; HTN: hypertension.

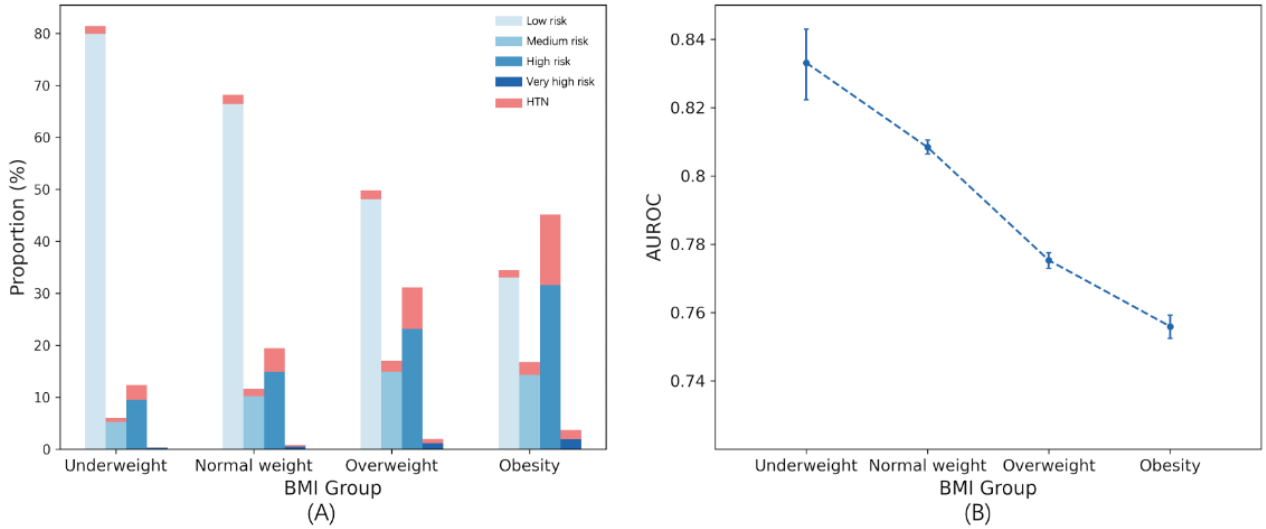

We also analyzed the impact of socioeconomic factors on hypertension risk (Figure 9). Among urban residents, the proportion of individuals classified as high risk (31,764/96,055, 33.07%) and very high-risk (4084/96,055, 4.25%) was greater than that among rural residents, where the high-risk group accounted for 24.66% (213,441/865,464) and the very high-risk group for 1.14% (9891/865,464). Hypertension incidence among the urban population (15,664/96,055, 16.31%) was also higher than that among rural residents (85,414/865,464, 9.87%). Similarly, individuals with lower levels of education (illiterate or semiliterate, primary school) had higher proportions in the high-risk (169,594/417,837, 40.59%) and very high-risk (9104/417,837, 2.18%) categories, with correspondingly higher incidence rates (61,741/417,837, 14.77%). These findings suggest that socioeconomic factors such as living environment and education level are closely linked to hypertension risk.

Figure 9. (A) Proportion of 4 risk categories across different residence group subsets. (B) Proportion of 4 risk categories across different educational level group subsets. HTN: hypertension.

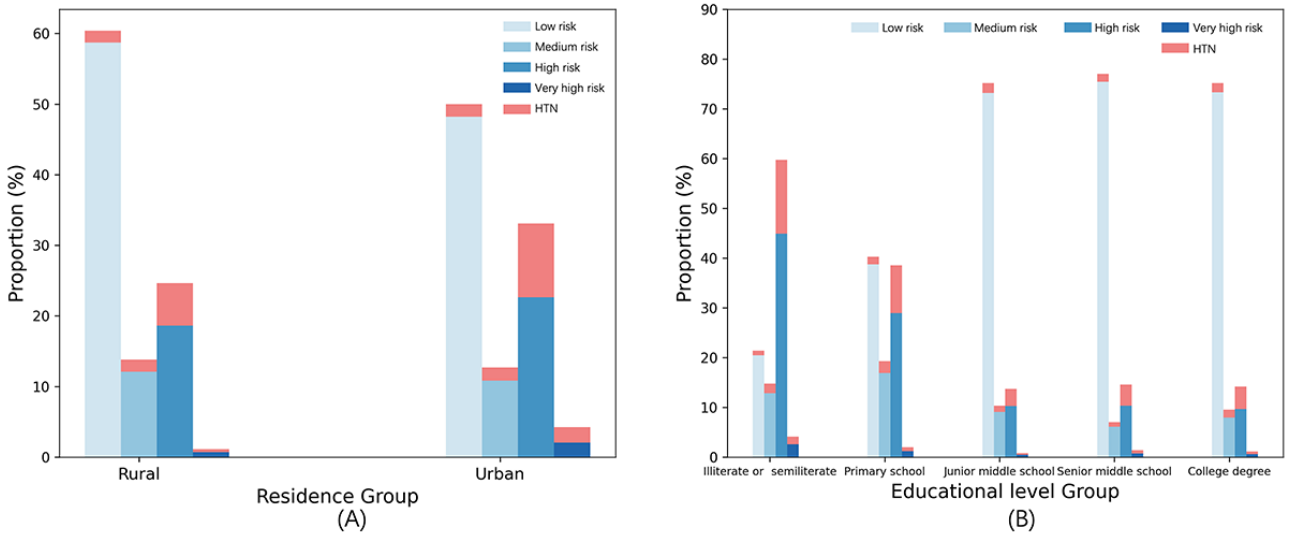

Finally, we developed a web-based application based on our predictive model (see main interface in Multimedia Appendix 10) [36], designed for integration into routine clinical workflows. During health examinations, clinicians can input patient data into the application, which then instantly generates individualized hypertension risk assessments. Each patient is assigned to 1 of the 4 actionable risk levels (low, medium, high, very high), aiding in targeted clinical decision-making. The application also employs interpretable SHAP value visualizations, allowing clinicians to understand the key contributors to each patient's risk profile. The application is publicly accessible to encourage external validation by health care providers [36], and we have also released the corresponding open-source code (Multimedia Appendix 11) [37].

XSL•FO
RenderX

Discussion

Principal Findings

This study shows a 2-year hypertension risk prediction model using a retrospective cohort of over 1 million participants from populations residing in the desert regions of China, with both internal and external validation. Using feature selection techniques, we identified 24 significant variables. The CatBoost model demonstrated superior predictive performance, achieving an area under the curve (AUC) of 0.888 in the retrospective cohort and 0.803 in the prospective cohort, outperforming other machine learning and deep learning models. During the 2-year follow-up in the prospective cohort, the model effectively stratified individuals into 4 distinct risk categories, revealing significant differences in the hazard ratios for hypertension incidence among these groups. Analysis of the model's significant features indicated that the risk of hypertension is strongly influenced by basic biometrics (age, sex, and BMI) and socioeconomic factors (residence and education level). Furthermore, a web-based application was developed and made open-source, providing clinicians with a practical and accessible tool to assess hypertension risk and guide early intervention strategies.

According to the World Health Organization's report on global hypertension prevalence, approximately 1.1 billion adults aged 30-79 years have hypertension, with two-thirds residing in low- and middle-income countries [25]. China, as the largest middle-income country, has an estimated hypertension prevalence of 27.5% [38]. In this study, we observed an incidence rate of 56.30 per 1000 person-years, which is higher than the national average of 48.60 per 1000 person-years reported between 2011 and 2015 [39]. This discrepancy may be attributed to the geographic location of our cohort, composed of residents from 4 regions near the Taklamakan Desert in northwest China. Environmental factors, lifestyle choices, and dietary patterns specific to this area likely contributed to the higher incidence, underscoring a more severe public health challenge in these regions and highlighting the need for targeted prevention strategies.

CatBoost was chosen as the hypertension risk prediction model due to its superior performance compared to other gradient-boosting decision tree models, such as LightGBM and extreme gradient boosting. CatBoost's key advantage lies in its ability to efficiently handle high-dimensional categorical data, thanks to its unique target encoding and ordered boosting algorithms. These methods minimize data leakage and prevent the use of future information in current predictions, enhancing model generalizability and reducing the need for extensive preprocessing. Although recent advancements in deep learning for tabular data prediction have shown promise, the 2 deep learning models (Feature Tokenizer + Transformer and SAINT) evaluated in this study underperformed slightly compared to CatBoost. Feature Tokenizer + Transformer leverages feature tokenization, transforming features into tokens for the Transformer architecture, while SAINT introduces intersample attention mechanisms to capture relationships between samples. Despite their slightly lower performance in this dataset, these models demonstrated comparable results to other gradient-boosting decision tree models, indicating that deep learning approaches have strong potential for future application in tabular data prediction.

When contrasted with previous studies [10-16], this study represents an advancement in both sample size and model performance. Our dataset included data from 1750 hospitals and community clinics, encompassing a total of 2 million individuals, ensuring the reliability and robustness of the results. The CatBoost model achieved an AUC of 0.888 in the retrospective cohort and 0.803 in the prospective cohort. By comparison, López-Martínez et al [16] utilized the 2007-2016 National Health and Nutrition Examination Survey dataset to develop 2 predictive models: a logistic regression model reported in 2018, which achieved an internal validation AUC of 0.73 [15] and an artificial neural network model developed in 2020, which improved the AUC to 0.77.

Consistent with prior studies, our model identified age, gender, and BMI as the major predictors of hypertension [3,40,41]. We found that as age and BMI increased, the proportion of individuals in the high-risk category rose significantly, further emphasizing their importance in hypertension risk prediction. In terms of socioeconomic factors, our study showed that individuals residing in urban areas and those with lower education levels had higher incidences of hypertension over 2 years. This may be due to urban populations engaging in consuming diets high in salt and fat while lacking overall nutritional quality, less physical activity, and being exposed to more severe environmental pollution [42-44], while individuals with lower education levels may have reduced access to health care and limited health literacy, both of which increase their vulnerability to hypertension [45].

The clinical implications of this study are significant. To our knowledge, this is one of the first large-scale, prospective cohort studies focused on hypertension risk prediction in populations residing in regions surrounding a desert. The model relies on 24 basic clinical and biochemical features that are routinely collected during health checkups, making it practical for use in primary health care settings. Given the limited health care resources in desert regions, the model can assist health care providers in identifying high-risk individuals for hypertension up to 2 years in advance. This allows for optimized medical decision-making and resource allocation, promoting earlier intervention and potentially reducing the overall disease burden.

Furthermore, the development of a web-based application extends the model's utility, allowing for broader clinical implementation. This user-friendly tool provides an intuitive way for clinicians to assess hypertension risk predictions and the specific contributions of individual risk factors. Especially in resource-constrained regions, this model offers an opportunity for precision prevention and intervention, potentially reducing long-term health care costs and improving patient outcomes. These contributions collectively aim to address previous methodological and contextual limitations, expand the body of knowledge, and provide potential solutions to pressing public health challenges.

XSL•FO
RenderX

Limitations

This study has several important limitations that may affect the interpretation and generalizability of the findings. First, some of the health examination data contained missing values. Although we addressed this issue by using a random forest imputation method based on the training set, this approach may introduce bias, particularly if the missing data are not missing completely at random. Imputation can improve data completeness, but it may still impact the accuracy of the model's predictions and its external validity. Second, the range of variables included in this study was limited. Important factors such as psychological stress, genetic predisposition, and environmental influences were not incorporated into the model. These factors have been shown to contribute significantly to hypertension risk, and their omission may limit the comprehensiveness of the model. In addition, the model was developed using data from specific geographic region populations surrounding desert areas in northwest China. Although this provides valuable insights for these communities, the model's applicability to other populations, particularly those in different geographic and socioeconomic settings, may be limited. Lastly, certain lifestyle factors such as physical activity and diet were based on self-reported data, which are subject to recall bias and may not accurately reflect actual behaviors.

Conclusions

In conclusion, we developed and validated a 2-year hypertension incidence risk prediction model by using data from nearly 2 million individuals residing in regions surrounding the Taklamakan Desert. The model demonstrated strong predictive performance, with high accuracy in both retrospective and prospective cohorts. By stratifying individuals into distinct risk categories, the model identified significant variations in hypertension incidence rates between these groups, underscoring its potential as a valuable tool for risk stratification in clinical practice. To enhance clinical applicability, we further developed a web-based hypertension prediction application, facilitating early screening and intervention for high-risk populations. This study not only holds significant clinical value but also provides an efficient public health tool for regions with limited health care resources, supporting early prevention and precision management of hypertension.

Acknowledgments

Li Lin provided precise language revision for this paper. No generative artificial intelligence tools were used in this manuscript. All figures submitted have been created by the authors, who confirm that the images are original with no duplication and have not been previously published in whole or in part. This work was supported by the Key Research and Development Program of China (2022YFC3601600), the Province Natural Science Foundation of Guangdong (2024A1515011989), the Guangzhou Science and Technology Plan (202206010028) and Fundamental Research Funds for the Central Universities (24xkjc025).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: YC, KG, YZ, YY
Data curation: YY
Formal analysis: WJ
Funding acquisition: YZ
Investigation: YC, ZH
Methodology: YC, KG
Project administration: YZ
Resources: YY
Software: YC, KG
Supervision: YZ, WJ
Validation: YC, KG
Visualization: YC
Writing – original draft: YC, KG
Writing – review & editing: YC, KG, ZH, YZ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of missing values across variables in the retrospective cohort.
[PDF File (Adobe PDF File), 199 KB - jmir_v27i1e68442_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Distribution of missing values across variables in the prospective cohort.
[PDF File (Adobe PDF File), 207 KB - jmir_v27i1e68442_app2.pdf]

Multimedia Appendix 3
Optimal hyperparameter for each model.
[PDF File (Adobe PDF File), 139 KB - jmir_v27i1e68442_app3.pdf]

Multimedia Appendix 4
Baseline characteristics of the study population in the retrospective cohort.
[PDF File (Adobe PDF File), 164 KB - jmir_v27i1e68442_app4.pdf]

Multimedia Appendix 5
Baseline characteristics of the study population in the prospective cohort.
[PDF File (Adobe PDF File), 157 KB - jmir_v27i1e68442_app5.pdf]

Multimedia Appendix 6
Evaluation of multiple machine learning models to predict the 2-year hypertension incidence risk.
[PDF File (Adobe PDF File), 145 KB - jmir_v27i1e68442_app6.pdf]

Multimedia Appendix 7
Confusion matrices of multiple machine learning models for predicting the 2-year hypertension incidence risk.
[PDF File (Adobe PDF File), 145 KB - jmir_v27i1e68442_app7.pdf]

Multimedia Appendix 8
Statistical significance analysis of area under the receiver operating characteristic curve comparisons between CatBoost and other models.
[PDF File (Adobe PDF File), 144 KB - jmir_v27i1e68442_app8.pdf]

Multimedia Appendix 9
Classification of individuals into 4 risk levels based on the positive predictive values in the prospective dataset.
[PDF File (Adobe PDF File), 155 KB - jmir_v27i1e68442_app9.pdf]

Multimedia Appendix 10
Main interface of the web-based application for 2-year hypertension incidence risk prediction.
[PDF File (Adobe PDF File), 165 KB - jmir_v27i1e68442_app10.pdf]

Multimedia Appendix 11
Hypertension risk prediction web application.
[ZIP File (Zip Archive), 8 KB - jmir_v27i1e68442_app11.zip]

References

1. GBD 2017 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. A global brief on hypertension: silent killer, global public health crisis: World Health Day 2013. World Health Organization. 2013. URL: https://www.who.int/publications/i/item/a-global-brief-on-hypertension-silent-killer-global-public-health-crisis-world-health-day-2013 [accessed 2025-01-26]
3. Wang Z, Chen Z, Zhang L, Wang X, Hao G, Zhang Z, China Hypertension Survey Investigators. Status of hypertension in China: results from the China hypertension survey, 2012-2015. Circulation 2018 May 29;137(22):2344-2356. [doi: 10.1161/CIRCULATIONAHA.117.032380] [Medline: 29449338]
4. Ma S, Yang L, Zhao M, Magnussen CG, Xi B. Trends in hypertension prevalence, awareness, treatment and control rates among Chinese adults, 1991-2015. J Hypertens 2021 Apr 01;39(4):740-748. [doi: 10.1097/HJH.0000000000002698] [Medline: 33186320]

XSL•FO
RenderX

5. Li G, Zhang H, Hu M, He J, Yang W, Zhao H, et al. Associations of combined exposures to ambient temperature, air pollution, and green space with hypertension in rural areas of Anhui Province, China: A cross-sectional study. Environ Res 2022 Mar;204(Pt D):112370. [doi: 10.1016/j.envres.2021.112370] [Medline: 34780789]
6. Zhou W, Wang Q, Li R, Kadier A, Wang W, Zhou F, et al. Combined effects of heatwaves and air pollution, green space and blue space on the incidence of hypertension: A national cohort study. Sci Total Environ 2023 Apr 01;867:161560. [doi: 10.1016/j.scitotenv.2023.161560] [Medline: 36640878]
7. Zhang C, Yan M, Du H, Ban J, Chen C, Liu Y, et al. Mortality risks from a spectrum of causes associated with sand and dust storms in China. Nat Commun 2023 Oct 27;14(1):6867 [FREE Full text] [doi: 10.1038/s41467-023-42530-w] [Medline: 37891159]
8. Echouffo-Tcheugui JB, Batty GD, Kivimäki M, Kengne AP. Risk models to predict hypertension: a systematic review. PLoS One 2013;8(7):e67370 [FREE Full text] [doi: 10.1371/journal.pone.0067370] [Medline: 23861760]
9. Martinez-Ríos E, Montesinos L, Alfaro-Ponce M, Pecchia L. A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. Biomedical Signal Processing and Control 2021 Jul;68:102813. [doi: 10.1016/j.bspc.2021.102813]
10. Golino HF, Amaral LSDB, Duarte SFP, Gomes CMA, Soares TDJ, Dos Reis LA, et al. Predicting increased blood pressure using machine learning. J Obes 2014;2014:637635 [FREE Full text] [doi: 10.1155/2014/637635] [Medline: 24669313]
11. Patnaik R, Chandran M, Lee S, Gupta A, Kim C, Kim C. Predicting the occurrence of essential hypertension using annual health records. 2018 Presented at: Second International Conference on Advances in Electronics, Computers and Communications (ICAECC); October 4; Bangalore, India. [doi: 10.1109/icaecc.2018.8479458]
12. Nour M, Polat K. Automatic classification of hypertension types based on personal features by machine learning algorithms. Mathematical Problems in Engineering 2020 Jan 20;2020:1-13. [doi: 10.1155/2020/2742781]
13. Islam MM, Alam MJ, Maniruzzaman M, Ahmed NAMF, Ali MS, Rahman MJ, et al. Predicting the risk of hypertension using machine learning algorithms: A cross sectional study in Ethiopia. PLoS One 2023;18(8):e0289613 [FREE Full text] [doi: 10.1371/journal.pone.0289613] [Medline: 37616271]
14. Fang M, Chen Y, Xue R, Wang H, Chakraborty N, Su T, et al. A hybrid machine learning approach for hypertension risk prediction. Neural Comput & Applic 2021 May 20;35(20):14487-14497. [doi: 10.1007/s00521-021-06060-0]
15. López-Martínez F, Schwarcz.MD A, Núñez-Valdez ER, García-Díaz V. Machine learning classification analysis for a hypertensive population as a function of several risk factors. Expert Systems with Applications 2018 Nov;110:206-215. [doi: 10.1016/j.eswa.2018.06.006]
16. López-Martínez F, Núñez-Valdez ER, Crespo RG, García-Díaz V. An artificial neural network approach for predicting hypertension using NHANES data. Sci Rep 2020 Jun 30;10(1):10620 [FREE Full text] [doi: 10.1038/s41598-020-67640-z] [Medline: 32606434]
17. LaFreniere D, Zulkernine F, Barber D, Martin K. Using machine learning to predict hypertension from a clinical dataset. 2017 Presented at: IEEE Symposium Series on Computational Intelligence (SSCI); February 13; Athens, Greece. [doi: 10.1109/ssci.2016.7849886]
18. Luo Y, Li Y, Lu Y, Lin S, Liu X. The prediction of hypertension based on convolution neural network. 2019 Presented at: IEEE 4th International Conference on Computer and Communications (ICCC); August 1; Chengdu, China. [doi: 10.1109/compcomm.2018.8780834]
19. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res 2018 Jan 30;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]
20. Tjahjadi H, Ramli K, Murfi H. Noninvasive classification of blood pressure based on photoplethysmography signals using bidirectional long short-term memory and time-frequency analysis. IEEE Access 2020;8:20735-20748. [doi: 10.1109/access.2020.2968967]
21. Huang G, Jin Q, Mao Y. Predicting the 5-year risk of nonalcoholic fatty liver disease using machine learning models: prospective cohort study. J Med Internet Res 2023 Sep 12;25:e46891 [FREE Full text] [doi: 10.2196/46891] [Medline: 37698911]
22. Notice on issuing the ethical review measures for life sciences and medical research involving humans. People's Republic of China. URL: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm [accessed 2025-01-26]
23. Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, et al. Association of short-term exposure to air pollution with mortality in older adults. JAMA 2017 Dec 26;318(24):2446-2456 [FREE Full text] [doi: 10.1001/jama.2017.17923] [Medline: 29279932]
24. Wassertheil-Smoller S, Anderson G, Psaty BM, Black HR, Manson J, Wong N, et al. Hypertension and its treatment in postmenopausal women: baseline data from the Women's Health Initiative. Hypertension 2000 Nov;36(5):780-789. [doi: 10.1161/01.hyp.36.5.780] [Medline: 11082143]
25. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980 [FREE Full text] [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]

XSL•FO
RenderX

26. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine learning 2002 Jan;46:389-422. [doi: 10.1023/A:1012487302797]
27. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W. LightGBM: A highly efficient gradient boosting decision tree. 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4; Red Hook, NY, USA.
28. Prokhorenkova L, Gusev G, Vorobev A, Dorogush A, Gulin A. CatBoost: unbiased boosting with categorical features. 2018 Presented at: Proceedings of the 32nd International Conference on Neural Information Processing Systems; December 3; Red Hook, NY, USA p. 6639-6649.
29. Chen T, Guestrin C, editors. XGBoost: A scalable tree boosting system. 2016 Aug 13 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13; New York, NY, USA. [doi: 10.1145/2939672.2939785]
30. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A. Revisiting deep learning models for tabular data. 2021 Presented at: Proceedings of the 35th International Conference on Neural Information Processing Systems; December 6; Red Hook, NY, USA.
31. Somepalli G, Goldblum M, Schwarzschild A, Bruss C, Goldstein T. SAINT: Improved neural networks for tabular data via row attention and contrastive pre-training. ArXiv Preprint posted online on June 2, 2021. [doi: 10.48550/arXiv.2106.01342]
32. Bergstra J, Yamins D, Cox D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. 2013 Presented at: Proceedings of the 30th International Conference on International Conference on Machine Learning; June 16; Atlanta, GA, USA.
33. Zadrozny B, Elkan C. Transforming classifier scores into accurate multiclass probability estimates. 2002 Presented at: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining; July 23; New York, NY, USA.
34. Scott M, Su-In L. A unified approach to interpreting model predictions. 2017 Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4; Red Hook, NY, USA.
35. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ 2003 Jan 25;326(7382):219 [FREE Full text] [doi: 10.1136/bmj.326.7382.219] [Medline: 12543843]
36. 2-Year Hypertension Risk Prediction Model. URL: http://116.205.138.86:8501/ [accessed 2025-01-26]
37. Hypertension risk prediction web app. GitHub. URL: https://github.com/christinell/webapp [accessed 2025-01-26]
38. Zhang M, Wu J, Zhang X, Hu CH, Zhao ZP, Li C, et al. [Prevalence and control of hypertension in adults in China, 2018]. Zhonghua Liu Xing Bing Xue Za Zhi 2021 Oct 10;42(10):1780-1789. [doi: 10.3760/cma.j.cn112338-20210508-00379] [Medline: 34814612]
39. Luo Y, Xia F, Yu X, Li P, Huang W, Zhang W. Long-term trends and regional variations of hypertension incidence in China: a prospective cohort study from the China Health and Nutrition Survey, 1991-2015. BMJ Open 2021 Jan 13;11(1):e042053 [FREE Full text] [doi: 10.1136/bmjopen-2020-042053] [Medline: 33441360]
40. Zhao M, Bovet P, Xi B. Weight status change from adolescence to young adulthood and the risk of hypertension and diabetes mellitus. Hypertension 2020 Aug;76(2):583-588. [doi: 10.1161/hypertensionaha.120.14882]
41. Kim MS, Kim WJ, Khera AV, Kim JY, Yon DK, Lee SW, et al. Association between adiposity and cardiovascular outcomes: an umbrella review and meta-analysis of observational and Mendelian randomization studies. Eur Heart J 2021 Sep 07;42(34):3388-3403 [FREE Full text] [doi: 10.1093/eurheartj/ehab454] [Medline: 34333589]
42. Kokkinos PF, Giannelou A, Manolis A, Pittaras A. Physical activity in the prevention and management of high blood pressure. Hellenic J Cardiol 2009;50(1):52-59 [FREE Full text] [Medline: 19196621]
43. Liu C, Chen R, Zhao Y, Ma Z, Bi J, Liu Y, et al. Associations between ambient fine particulate air pollution and hypertension: A nationwide cross-sectional study in China. Sci Total Environ 2017 Apr 15;584-585:869-874 [FREE Full text] [doi: 10.1016/j.scitotenv.2017.01.133] [Medline: 28153400]
44. Ranzani OT, Kalra A, Di Girolamo C, Curto A, Valerio F, Halonen JI, et al. Urban-rural differences in hypertension prevalence in low-income and middle-income countries, 1990-2020: A systematic review and meta-analysis. PLoS Med 2022 Aug;19(8):e1004079 [FREE Full text] [doi: 10.1371/journal.pmed.1004079] [Medline: 36007101]
45. Basu S, Millett C. Social epidemiology of hypertension in middle-income countries: determinants of prevalence, diagnosis, treatment, and control in the WHO SAGE study. Hypertension 2013 Jul;62(1):18-26. [doi: 10.1161/HYPERTENSIONAHA.113.01374] [Medline: 23670299]

Abbreviations

AUC: area under the curve
AUROC: area under the receiver operating characteristic
LASSO: Least Absolute Shrinkage and Selection Operator
LightGBM: light gradient-boosting machine
RFECV: recursive feature elimination with cross-validation

XSL•FO
RenderX

SHAP: Shapley Additive Explanations

Edited by A Coristine; submitted 05.11.24; peer-reviewed by R Mpofu, Y Zhang; comments to author 29.11.24; revised version received 13.12.24; accepted 05.02.25; published 12.03.25.

Please cite as:
Cheng Y, Gu K, Ji W, Hu Z, Yang Y, Zhou Y
Two-Year Hypertension Incidence Risk Prediction in Populations in the Desert Regions of Northwest China: Prospective Cohort Study
J Med Internet Res 2025;27:e68442
URL: https://www.jmir.org/2025/1/e68442
doi:10.2196/68442
PMID:40072485

©Yinlin Cheng, Kuiying Gu, Weidong Ji, Zhensheng Hu, Yining Yang, Yi Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning to Assist in Managing Acute Kidney Injury in General Wards: Multicenter Retrospective Study

Nam-Jun Cho[1*], MD; Inyong Jeong[2*], MS; Se-Jin Ahn[2], BS; Hyo-Wook Gil[1], MD; Yeongmin Kim[2], BS; Jin-Hyun Park[2], MS; Sanghee Kang[3], MD; Hwamin Lee[2], PhD

[1]Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Republic of Korea

[2]Department of Biomedical Informatics, Korea University College of Medicine, Seoul, Republic of Korea

[3]Department of Surgery, Korea University Guro Hospital, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hwamin Lee, PhD
Department of Biomedical Informatics
Korea University College of Medicine
Seongbuk-gu
73 Goryeodae-ro
Seoul, 02841
Republic of Korea
Phone: 82 1063205109
Email: hwamin@korea.ac.kr

Abstract

Background: Most artificial intelligence–based research on acute kidney injury (AKI) prediction has focused on intensive care unit settings, limiting their generalizability to general wards. The lack of standardized AKI definitions and reliance on intensive care units further hinder the clinical applicability of these models.

Objective: This study aims to develop and validate a machine learning–based framework to assist in managing AKI and acute kidney disease (AKD) in general ward patients, using a refined operational definition of AKI to improve predictive performance and clinical relevance.

Methods: This retrospective multicenter cohort study analyzed electronic health record data from 3 hospitals in South Korea. AKI and AKD were defined using a refined version of the Kidney Disease: Improving Global Outcomes criteria, which included adjustments to baseline serum creatinine estimation and a stricter minimum increase threshold to reduce misclassification due to transient fluctuations. The primary outcome was the development of machine learning models for early prediction of AKI (within 3 days before onset) and AKD (nonrecovery within 7 days after AKI).

Results: The final analysis included 135,068 patients. A total of 7658 (8%) patients in the internal cohort and 2898 (7.3%) patients in the external cohort developed AKI. Among the 5429 patients in the internal cohort and 1998 patients in the external cohort for whom AKD progression could be assessed, 896 (16.5%) patients and 287 (14.4%) patients, respectively, progressed to AKD. Using the refined criteria, 2898 cases of AKI were identified, whereas applying the standard Kidney Disease: Improving Global Outcomes criteria resulted in the identification of 5407 cases. Among the 2509 patients who were not classified as having AKI under the refined criteria, 2242 had a baseline serum creatinine level below 0.6 mg/dL, while the remaining 267 experienced a decrease in serum creatinine before the onset of AKI. The final selected early prediction model for AKI achieved an area under the receiver operating characteristic curve of 0.9053 in the internal cohort and 0.8860 in the external cohort. The early prediction model for AKD achieved an area under the receiver operating characteristic curve of 0.8202 in the internal cohort and 0.7833 in the external cohort.

Conclusions: The proposed machine learning framework successfully predicted AKI and AKD in general ward patients with high accuracy. The refined AKI definition significantly reduced the classification of patients with transient serum creatinine fluctuations as AKI cases compared to the previous criteria. These findings suggest that integrating this machine learning framework into hospital workflows could enable earlier interventions, optimize resource allocation, and improve patient outcomes.

(J Med Internet Res 2025;27:e66568) doi:10.2196/66568

XSL•FO
RenderX

KEYWORDS

acute kidney injury; machine learning; recovery of function; creatinine; kidney; patient rooms

Introduction

Acute kidney injury (AKI) is an escalating critical health and socioeconomic issue, marked by prolonged hospital stays, elevated medical costs, and high mortality rates [1]. AKI is a secondary condition arising from hospital interventions, such as medication, surgery, and infections, with a prevalence of 10%-15% among general inpatients [2]. Studies have shown that the duration of AKI correlates with increased risks of complications and mortality [3]. AKI and its progression to acute kidney disease (AKD) are associated with significant increases in postdischarge morbidity and mortality. The prolonged duration of AKI in general wards has been associated with higher risks of complications and increased mortality [3-5]. Early prediction of AKI and its progression to AKD using artificial intelligence (AI) models improve patient outcomes through timely intervention and personalized management strategies [5-7]. There are research findings indicating that applying a clinical decision support system for actual AKI occurrences has significantly improved patient outcomes [8-10]. However, there is a lack of standardized operational definitions for AKI, particularly in retrospective studies, where data collection may not be systematic [11,12]. This inconsistency in defining AKI, especially in terms of baseline serum creatinine (SCr) and recovery criteria for AKD, leads to ambiguity in labeling and makes it difficult to compare the performance of different AI models across studies [13,14]. The variability in baseline SCr determination further complicates the issue [14-16]. Studies that investigated the impact of baseline SCr on model performance are limited to the intensive care unit (ICU) setting and do not address real-time early prediction in general wards [17-19], in differing AKI labels for the same patient. Defining the recovery for AKD is even more challenging [4,20-23].

General ward patients typically exhibit milder disease severity, less frequent laboratory measurements, and different baseline characteristics compared with ICU patients [15,20,24-27]. Applying the criterion of baseline SCr being 1.5 times higher requires more caution, especially in cases with low SCr levels. For instance, in a patient with 3 SCr measurements within 48 hours showing values of 0.9 mg/dL, 0.6 mg/dL, and 0.9 mg/dL in chronological order, additional consideration may be needed to determine whether the patient should be classified as experiencing AKI. Furthermore, when the baseline SCr is set below 0.6 mg/dL, even small fluctuations can easily meet the 1.5-fold criterion within 7 days, leading to the classification of AKI occurrence [28]. This study aimed to develop and validate a machine learning-based framework for early prediction of AKI and AKD, applicable to general ward patients, using a refined operational definition of AKI.

Methods

Study Setting and Participants

This multicenter, retrospective study was conducted across the general wards of 3 different hospitals. Data were collected from patients admitted to Korea University Guro Hospital and Anam Hospital between January 1, 2015, and December 31, 2021 (internal cohort), and from Soonchunhyang University Cheonan Hospital between March 1, 2016, and March 31, 2021 (external cohort). Patients younger than 19 years, those with fewer than 3 SCr measurements during their hospital stay, and those with an estimated glomerular filtration rate (eGFR) ≤60 mL/minute on the first day of SCr measurement were excluded.

Operative Definition

AKI was defined based on the Kidney Disease: Improving Global Outcomes (KDIGO) criteria as follows [1].

- An increase in SCr level ≥0.3 mg/dL (or ≥26.5 μmol/L) within 48 h.
- An increase in SCr ≥1.5 times the baseline within 7 days.

The KDIGO criteria are widely used; however, they are sometimes modified based on the characteristics of the cohort or the objectives of the study [28]. Compared with ICU patients, general ward patients are in relatively better condition with less frequent measurements of vital signs and laboratory data. Additionally, individuals with healthy kidneys and low SCr levels can be identified as having AKI due to simple fluctuations [14]. There may be more bias in real-time predictions, as SCr estimation is more frequently performed for patients with missing SCr values. Therefore, to prevent these issues, we restricted the application of baseline SCr levels and refined the labeling. To improve the labeling criteria, baseline SCr was defined as the lowest SCr measured within the previous 7 days. If the patient's baseline SCr was ≥0.3 mg/dL lower than the median and recent SCr values, it was excluded as a baseline to avoid errors due to SCr measurement inaccuracies or temporary fluctuations caused by medication, measurement error, and other factors. AKI was defined as a minimum increase of 0.3 mg/dL to ensure appropriate identification without mislabeling due to minor variations [2,29,30]. AKD was defined as the persistence of AKI for more than 7 days. Cases in which SCr did not return to <1.5 times the baseline SCr within 7 days were identified as AKD [21,29-32]. None of the previous definitions or studies have taken these aspects into consideration.

To develop an early prediction model for AKI occurrence, data from 1 to 3 days before AKI onset were labeled as 1, and the remaining data were labeled as 0. The data from the day of AKI onset were not used. For patients who did not develop AKI, all data were labeled as 0. Days without SCr measurements within 7 days were excluded from training and evaluation. Additionally, patients with ambiguous AKI labeling were excluded from the training and evaluation. Data from the day of AKI onset were used to predict AKD progression. Patients were labeled 0 or 1 based on whether they recovered within 7 days post AKI onset. Patients with insufficient SCr measurements after AKI onset were excluded. Multimedia Appendix 1 illustrates the previous KDIGO criteria, the improved AKI criteria proposed in this study, and examples of AKI and AKD labeling.

XSL•FO
RenderX

Data Preprocessing

The data collected from the electronic health records contained numerous missing values, including measurement values, times, and specific variables. To address this issue, data were summarized at 24-hour intervals, a method validated in previous studies [33]. Vital signs, laboratory data, and variables such as nephrotoxic drugs (eg, nonsteroidal anti-inflammatory drugs, nephrotoxic antibiotics [ANTIs], and cytotoxic chemotherapeutic agents), vascular imaging examinations, surgeries under general anesthesia, contrast-enhanced computed tomography (CECT), and ICU transfers within the past 7 days were collected.

Vital sign data measured multiple times within 24 hours were summarized using mean, maximum, and minimum values, and the number of measurements. The laboratory test results were determined based on recent measurements. Additionally, variables such as nephrotoxic drugs, vascular imaging examinations, surgeries under general anesthesia, CECT, and ICU transfers within the past 7 days were included. eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration 2021 equation [34]. The "BUN/Cr ratio" was defined as blood urea nitrogen (BUN) level divided by serum creatinine level.

Robust scaling was applied to all continuous variables and one-hot encoding was used for categorical variables. Approximately 120 features were extracted from the electronic health records, including basic patient information, vital sign data, laboratory test results, and other factors. Through a comprehensive literature review and consultation with specialists, a feature selection process was undertaken [6,23,35], and 42 features were ultimately selected based on their correlation coefficients and missing value ratios.

To handle outliers, the data distribution for each feature was reviewed along with the individual patient records. Outliers for some numerical variables were determined using histograms and quantiles with input from clinical experts. The missing values were imputed in 2 stages. First, where feasible, missing values were replaced with previous values to maintain data continuity. For variables with low missing values (less than 20%), the multiple imputation by chained equations method was used [36,37]. For variables with a missing rate exceeding 20%, the missing indicator method was used to denote missing values as unknown [38]. Numerical variables were categorized into 2 to 4 groups based on data distribution and expert consultation, and missing values were assigned to the "missing" category. For laboratory test results not subjected to missing indicators, changes in each variable were calculated by subtracting the median of the previous values from the current value (Multimedia Appendix 2).

Model Training and Evaluation

Various traditional machine learning models have been used, including logistic regression [39], random forest [40], eXtreme gradient boosting [41], light gradient boosting machine [42], and categorical boosting (CAT) [43]. For detailed model training procedures (Multimedia Appendix 3). The model evaluation metrics included accuracy, precision, recall, specificity, F_1-score, the area under the receiver operating characteristic (AUROC), and the area under the precision-recall curve. The primary outcomes were the development of models for early AKI prediction within 3 days and its progression to AKD, along with the establishment of a framework. Simulations of the designed framework were repeatedly performed using the external validation cohort by dividing the period into 1-year increments and assessing the generalization performance of the models across different institutions and over time.

Statistical Analysis

Descriptive statistics were used to present the baseline differences between the days with and without AKI. The distribution of continuous and categorical variables was expressed as means and SDs and counts and percentages, respectively. Normally distributed continuous variables were evaluated using 2-tailed *t* tests. For those that did not follow a normal distribution, the Mann-Whitney *U* test was used. The chi-square test was used to analyze categorical variables. Statistical significance was set at $P<.05$. Calibration plots were used to assess the agreement between the predicted probabilities and observed outcomes [44]. The Cox proportional hazards model was used to compare hazard ratios (HRs) between patients with actual AKD and those predicted by the model [45].

Ethical Considerations

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the institutional review boards of Soonchunhyang University Cheonan Hospital, Korea University Anam Hospital, and Guro Hospital (approvals 2019-10-023, 2023AN0145, and 2023GR0425, respectively). The need for individual consent was waived due to the retrospective nature of the study and the use of anonymized clinical data. To ensure privacy and data security, only fully anonymized data were used, and all analyses were conducted within a designated secure environment with restricted access. No financial compensation was provided to participants, as the study was purely observational and used deidentified retrospective data. All processes adhered to the guidelines developed for machine learning model development in the biomedical field [46] and followed the STROE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational studies.

Results

Labeling

Figure 1 is a flowchart of cohort formation and labeling. To compare the previous labeling criteria [7,32,47-49] with the labeling criteria used in this study, we developed and evaluated the model using the same methodology. Additionally, we calculated and presented the HRs for a 30% or 40% reduction in eGFR at the time of AKI onset (Figure 2 and Multimedia Appendix 4). In Table S7 in Multimedia Appendix 4, individuals with a baseline SCr<0.6 mg/dL were included, leading to the classification of AKI in patients with very low changes in SCr or very low baseline SCr levels. Additionally, subtracting the baseline median from the SCr at this point yielded a median of 0.10 (IQR 0.00-0.20).

XSL•FO
RenderX

Figure 2A illustrates this trend. Of the 2509 patients, 2242 (89.36%) had a baseline SCr<0.6 mg/dL at the time of AKI occurrence. The remaining 267 patients experienced decreased SCr levels before being diagnosed with AKI. This trend is illustrated in Figure S5 in Multimedia Appendix 4. Additionally, of the 2072 patients whose AKD status could be determined, only 50 (2.41%) progressed to AKD, and only 1 patient experienced a more than 30% decrease in eGFR within 30 days of AKI onset.

Figure 1. The data composition process. Hospitals A, B, and C refer to Korea University Anam Hospital, Korea University Guro Hospital, and Soonchunhyang University Cheonan Hospital, respectively. Since at least 1 day of data is required for model prediction, patients who developed AKI on the first day of hospitalization were excluded from the final cohort. AKI: acute kidney injury; AKD: acute kidney disease.

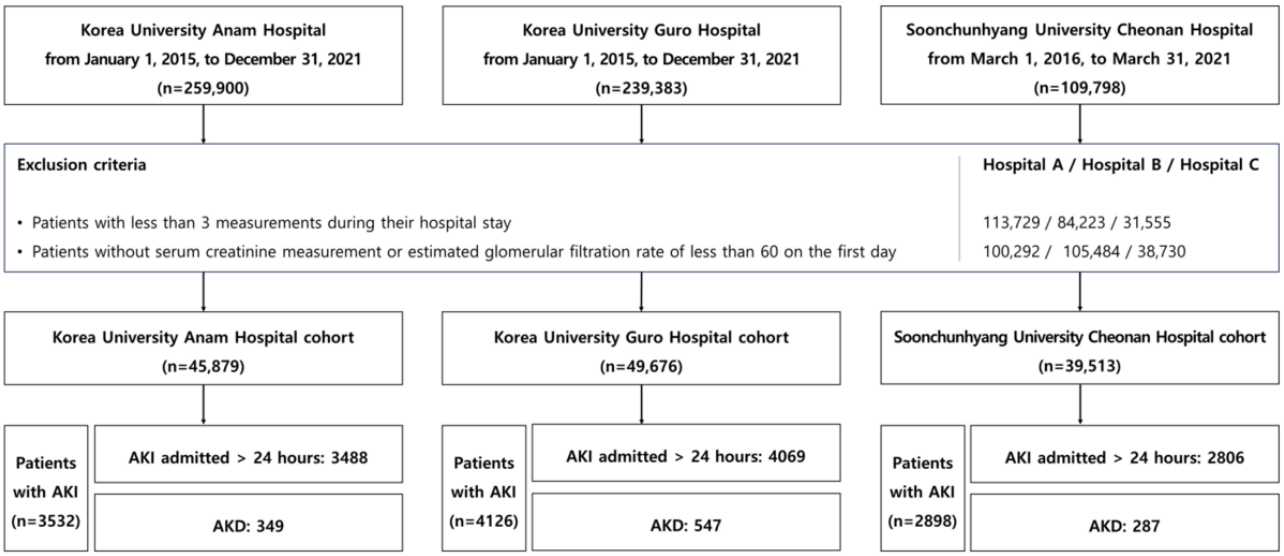

Figure 2. Serum creatinine trends in patients with AKI from the external validation cohort. (A) Results for 2509 patients who meet only the previous AKI criteria. (B) Results for 2898 patients who meet the refined AKI criteria. The blue line represents the median value for patients of each group, while the pink bar graph shows the proportion of serum creatinine measurements across all patients within 7 days before and after the day of acute kidney injury (day 0). The colored lines depict the serum creatinine levels for individual patients. AKI: acute kidney injury.

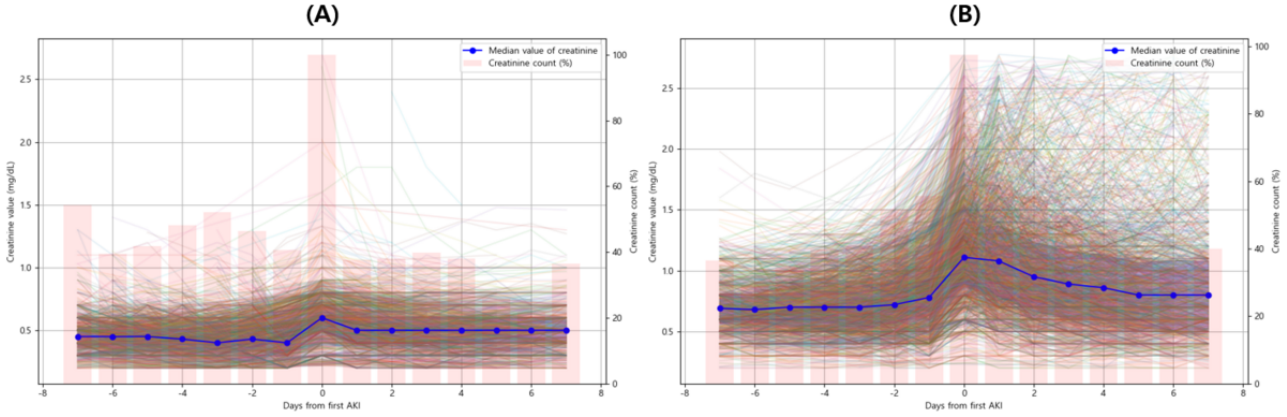

Early AKI Prediction Model

The final analysis included 95,555 and 39,513 cases in the internal and external cohorts, respectively. Within these cohorts, AKI was identified in 7658 (8.1%) and 2898 (7.3%) patients, respectively. The median length of hospital stay was 11 (IQR 7-20) days for the internal cohort and 10 (IQR 6-18) days for the external cohort. The median number of days to AKI occurrence from admission was 8 (IQR 3-17) days for the internal cohort and 7 (IQR 3-16) days for the external cohort (Multimedia Appendix 5). Tables S9 and Table S10 in Multimedia Appendix 5 present the basic statistics of patients with and without AKI. Comparing the days on which AKI occurred to the days on which it did not, there were statistically significant differences in most characteristics between the internal and external cohorts. However, differences in blood pressure and sodium and chloride levels were observed between the internal and external cohorts.

The models were evaluated based on the results of 5-fold cross-validation, with all performance metrics presented at a cutoff of 0.5. The CAT model demonstrated a strong predictive capability (AUROC=0.9134). The logistic regression model performed poorly (AUROC=0.7754). Model evaluation was conducted through both internal and external validations, showing a slight decrease in performance based on the AUROC (range 0.0097-0.0216). Table 1 presents the performance evaluation results of the early AKI prediction model. The hyperparameters used in the model are listed in Table S11 in Multimedia Appendix 5.

XSL•FO
RenderX

Figure 3 shows the Shapley additive explanations (SHAP) analysis for the early AKI prediction model. The interpretation of the CAT model using SHAP values indicated that the most important features were the SCr level and its changes. Other significant features included heart rate, BUN level, activated partial thromboplastin time (aPTT), total bilirubin level, and BMI, all of which increased the AKI risk. Conversely, lower eGFR, platelet count, body temperature, and pH were associated with a higher risk of AKI. Figure S7 in Multimedia Appendix 5 shows the results of testing the probability of the early AKI prediction model. Calibration plots were generated for both the internal and external cohorts. The slope of the calibration plot was 1.15 for the internal cohort and 1.08 for the external cohort, indicating a good calibration of the model's predicted probabilities. Additionally, the model probabilities were evaluated by dividing them into five groups, which demonstrated excellent calibration. A box plot of the probabilities in relation to the AKI status showed a clear distinction between AKI and non-AKI, further validating the predictive value of the model (Figure S8 in Multimedia Appendix 5).

Table 1. Performance evaluation results of the early prediction model for acute kidney injury.

Validation and model	Accuracy	Precision	Recall	F_1-score	AUROC[a]	AUPRC[b]
Cross-validation, mean (SD)						
LR[c]	0.9490 (0.0004)	0.3764 (0.0193)	0.0747 (0.0074)	0.1246 (0.0104)	0.7754 (0.0049)	0.1810 (0.0045)
RF[d]	0.9739 (0.0004)	0.9875 (0.0020)	0.4687 (0.0091)	0.6356 (0.0080)	0.9076 (0.0060)	0.6830 (0.0076)
XGB[e]	0.9720 (0.0006)	0.8149 (0.0148)	0.5485 (0.0091)	0.6555 (0.0066)	0.907 (0.0057)	0.6830 (0.0062)
LGBM[f]	0.9734 (0.0006)	0.8679 (0.0184)	0.5362 (0.0066)	0.6627 (0.0064)	0.9132 (0.0053)	0.6916 (0.0061)
CAT[g]	0.9747 (0.0005)	0.9423 (0.0095)	0.5115 (0.0093)	0.6630 (0.0078)	0.9134 (0.0053)	0.6924 (0.0059)
Internal						
LR	0.9485	0.3728	0.0709	0.1192	0.7584	0.1712
RF	0.9736	0.9902	0.4675	0.6352	0.9021	0.6826
XGB	0.9724	0.8322	0.5490	0.6616	0.9027	0.6834
LGBM	0.9738	0.8863	0.5348	0.6671	0.9058	0.6910
CAT	0.9749	0.9527	0.5142	0.6680	0.9053	0.6890
External						
LR	0.9560	0.3612	0.0865	0.1395	0.7487	0.1630
RF	0.9769	0.9789	0.4502	0.6168	0.8833	0.6303
XGB	0.9707	0.6938	0.5208	0.5950	0.8811	0.6196
LGBM	0.9732	0.7651	0.5071	0.6099	0.8853	0.6238
CAT	0.9755	0.8644	0.4823	0.6191	0.8860	0.6290

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]LR: logistic regression.

[d]RF: random forest.

[e]XGB: eXtreme gradient boosting.

[f]LGBM: light gradient boosting machine.

[g]CAT: categorical boosting.

XSL•FO
RenderX

Figure 3. SHAP values of the early prediction model for acute kidney injury. In the figure, red indicates higher values of the respective features, whereas blue indicates lower values. The position of the dots placed towards the right signifies a greater contribution of the feature value to the model's prediction of nonrecovery for the patient. For categorical variables, red dots represent a value of 1 and blue dots signify a value of 0. CT: computed tomography; SHAP: Shapley additive explanations.

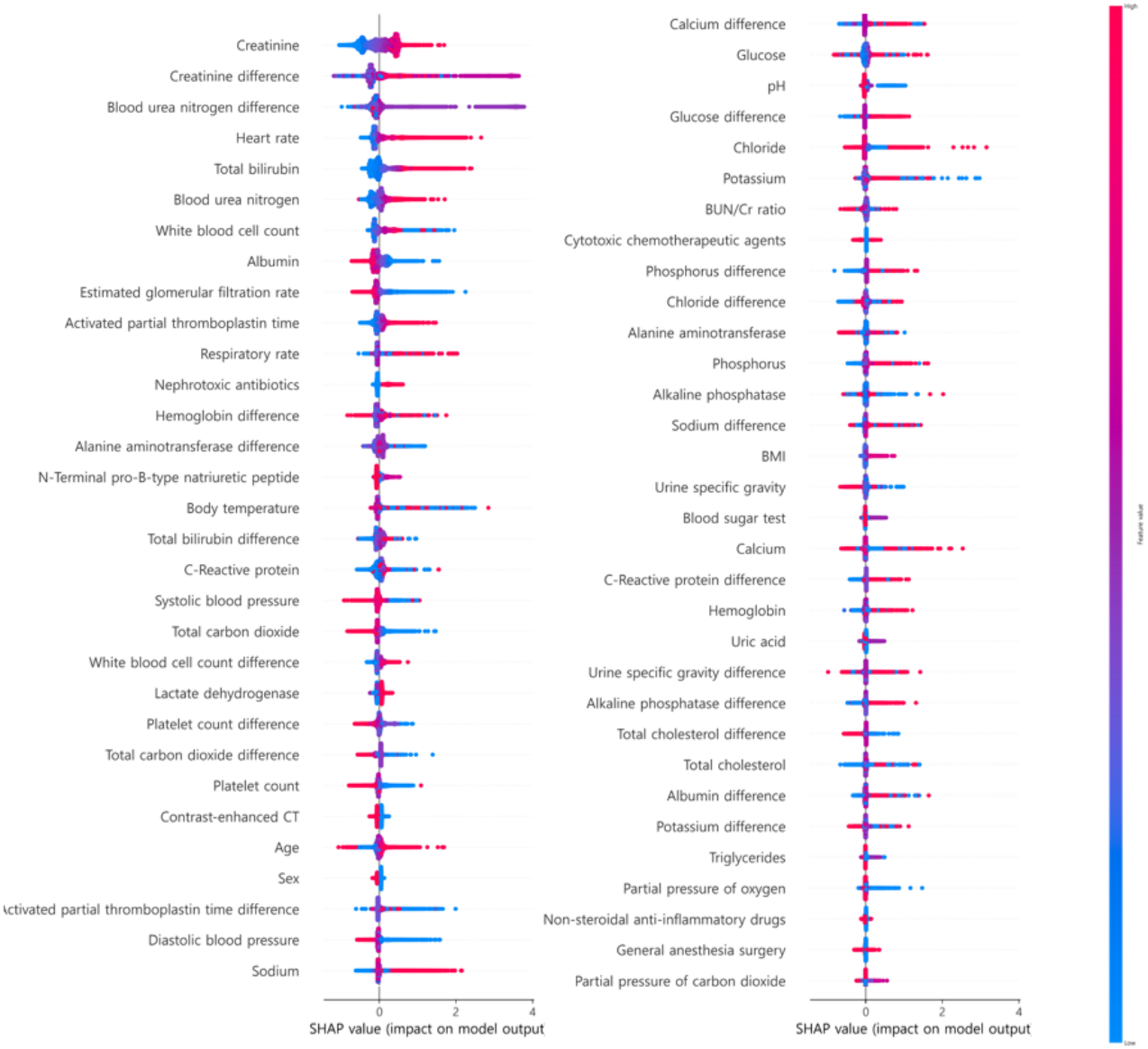

Early AKD Prediction Model

For the analysis of AKD progression, 5429 and 1998 patients in the internal and external cohorts, respectively were included after excluding those with unclear AKD status. Among them, 896 (16.5%) patients in the internal cohort and 287 (14.4%) in the external cohort were diagnosed with AKD. The median duration from AKI to discharge was 10 (IQR 4-23) days in the internal cohort and 9 (IQR 3-21) days in the external cohort. For patients who recovered within 7 days post AKI, the median recovery time was 5 (IQR 3-7) days in both cohorts (Figure S9 in Multimedia Appendix 6). The characteristics of progression to AKD in each cohort are presented in Tables S12 and S13 Multimedia Appendix 6. Variables that demonstrated statistical significance in both cohorts included hemoglobin, albumin, SCr, eGFR, glucose, alanine aminotransferase, blood sugar, calcium, urine specific gravity, C-reactive protein (CRP), ANTIs, and surgery. Some variables such as sex, diastolic blood pressure (DBP), white blood cell (WBC) count, BUN/Cr ratio, and total carbon dioxide showed different trends between the cohorts.

The CAT model outperformed the other models in both the internal (AUROC=0.8202) and external validations (AUROC=0.7833). The external validation showed a slight decrease in performance based on the AUROC (range 0.0262-0.0416). Table 2 presents the performance evaluation results of the early prediction model for AKD. The hyperparameters used in the model are listed in Table S14 in Multimedia Appendix 6.

Figure 4 shows the SHAP analysis of the early-stage AKD prediction model. For AKD prediction, an increase in SCr level at the time of AKI was a strong predictor of delayed recovery. Exposure to ANTIs, high DBP, CRP, aPTT, low eGFR, low pH, and elevated alkaline phosphatase were associated with an

XSL•FO
RenderX

increased risk of AKD. Patients who underwent surgery or CECT tended to recover quicker from AKI.

Figure S10 in Multimedia Appendix 6 shows the results of testing the probability of the early AKD prediction model. Calibration plots for both the internal and external cohorts were generated to assess model performance. The slopes of the calibration plots were 1.27 for the internal cohort and 1.20 for the external cohort, indicating a good calibration of the model's predicted probabilities. The evaluation of the model probabilities by box plots showed excellent calibration, similar to that of the early prediction model for AKI (Figure S11 in Multimedia Appendix 6).

Table 2. Performance evaluation results of the early prediction model for acute kidney disease.

Validation and model	Accuracy	Precision	Recall	F_1-score	AUROC[a]	AUPRC[b]
Cross-validation, mean (SD)						
LR[c]	0.8442 (0.0097)	0.6146 (0.0400)	0.1972 (0.0248)	0.2980 (0.0321)	0.7710 (0.0197)	0.4528 (0.0228)
RF[d]	0.8370 (0.0089)	0.7711 (0.0728)	0.0432 (0.0066)	0.0819 (0.0123)	0.7702 (0.0158)	0.4346 (0.0166)
XGB[e]	0.8439 (0.0165)	0.5972 (0.0848)	0.2372 (0.0327)	0.3382 (0.0406)	0.7622 (0.0250)	0.4490 (0.0409)
LGBM[f]	0.8436 (0.0095)	0.6066 (0.0545)	0.2142 (0.0173)	0.3149 (0.0109)	0.7623 (0.0178)	0.4421 (0.0067)
CAT[g]	0.8402 (0.0113)	0.6873 (0.0779)	0.0942 (0.0110)	0.1655 (0.0186)	0.7806 (0.0181)	0.4695 (0.0255)
Internal						
LR	0.8460	0.7347	0.2182	0.3364	0.7967	0.5376
RF	0.8297	0.9000	0.0545	0.1029	0.8099	0.5302
XGB	0.8547	0.7460	0.2848	0.4123	0.8097	0.5721
LGBM	0.8557	0.7667	0.2788	0.4089	0.7964	0.5598
CAT	0.8341	0.8000	0.0970	0.1730	0.8202	0.5558
External						
LR	0.8475	0.5893	0.1019	0.1737	0.7659	0.3959
RF	0.8446	0.6429	0.0278	0.0533	0.7837	0.4203
XGB	0.8494	0.5761	0.1636	0.2548	0.7682	0.3990
LGBM	0.8475	0.5543	0.1574	0.2452	0.7685	0.3945
CAT	0.8475	0.6786	0.0586	0.1080	0.7833	0.4368

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]LR: logistic regression.

[d]RF: random forest.

[e]XGB: eXtreme gradient boosting.

[f]LGBM: light gradient boosting machine.

[g]CAT: categorical boosting.

XSL•FO
RenderX

Figure 4. SHAP values of the early prediction model for acute kidney disease. In the figure, red indicates higher values of the respective features, whereas blue indicates lower values. The position of the dots placed towards the right signifies a greater contribution of the feature value to the model's prediction of nonrecovery for the patient. For categorical variables, red dots represent a value of 1 and blue dots signify a value of 0. CT: computed tomography; SHAP: Shapley additive explanations.

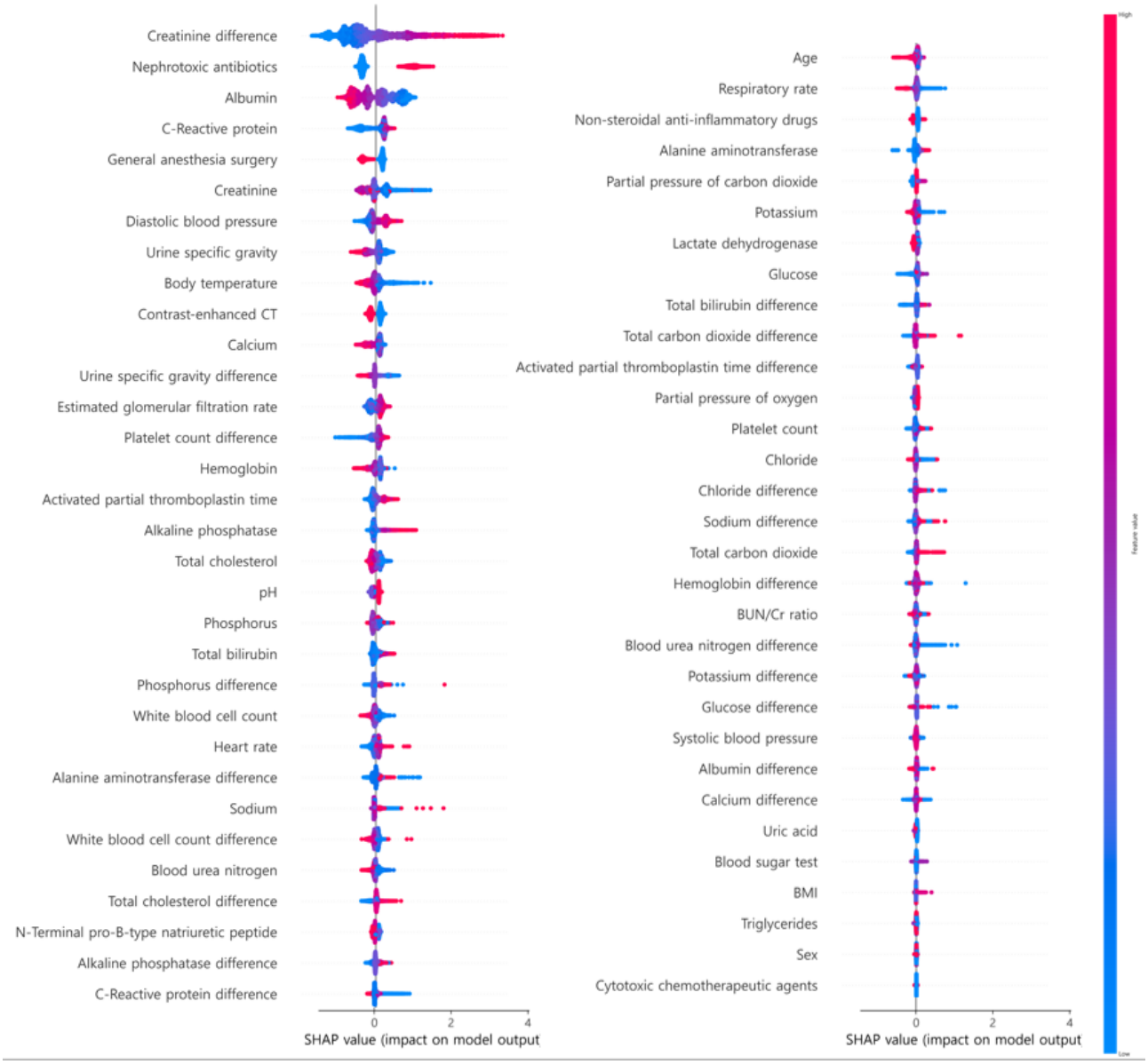

Framework

The framework to assist in managing AKI is shown in Multimedia Appendix 7. Figure 5A presents the framework simulation results for the early AKI prediction model. The designed framework was simulated using an external validation cohort at yearly intervals of 5 years. Among the 7284 patients included annually in the external validation cohort, 501 (6.9%) experienced AKI. Of these, 439 (87.68%) were predicted at least 1 day in advance. The F_1-score showed a decreasing trend after 2018 compared to evaluations between 2016 and 2018, with a minimum of 0.39 and a maximum of 0.46. Figure 5B presents the framework simulation results for the early prediction model for AKD. Excluding patients with insufficient SCr tracking from an average of 410 patients annually, 64 (15.6%) did not recover from AKI within 7 days. Among them, approximately 38 (58.4%) were predicted early.

XSL•FO
RenderX

Figure 5. Simulation results of the framework to assist in the management of AKI using external validation data. (A) Results of the early prediction model for AKI, and (B) results of the early prediction model for AKD. Blue indicates patients included in the yearly prediction, orange indicates event occurrence, green indicates patients successfully predicted by the model, and red indicates events successfully predicted by the model compared with actual events. AKI: acute kidney injury; AKD: acute kidney disease.

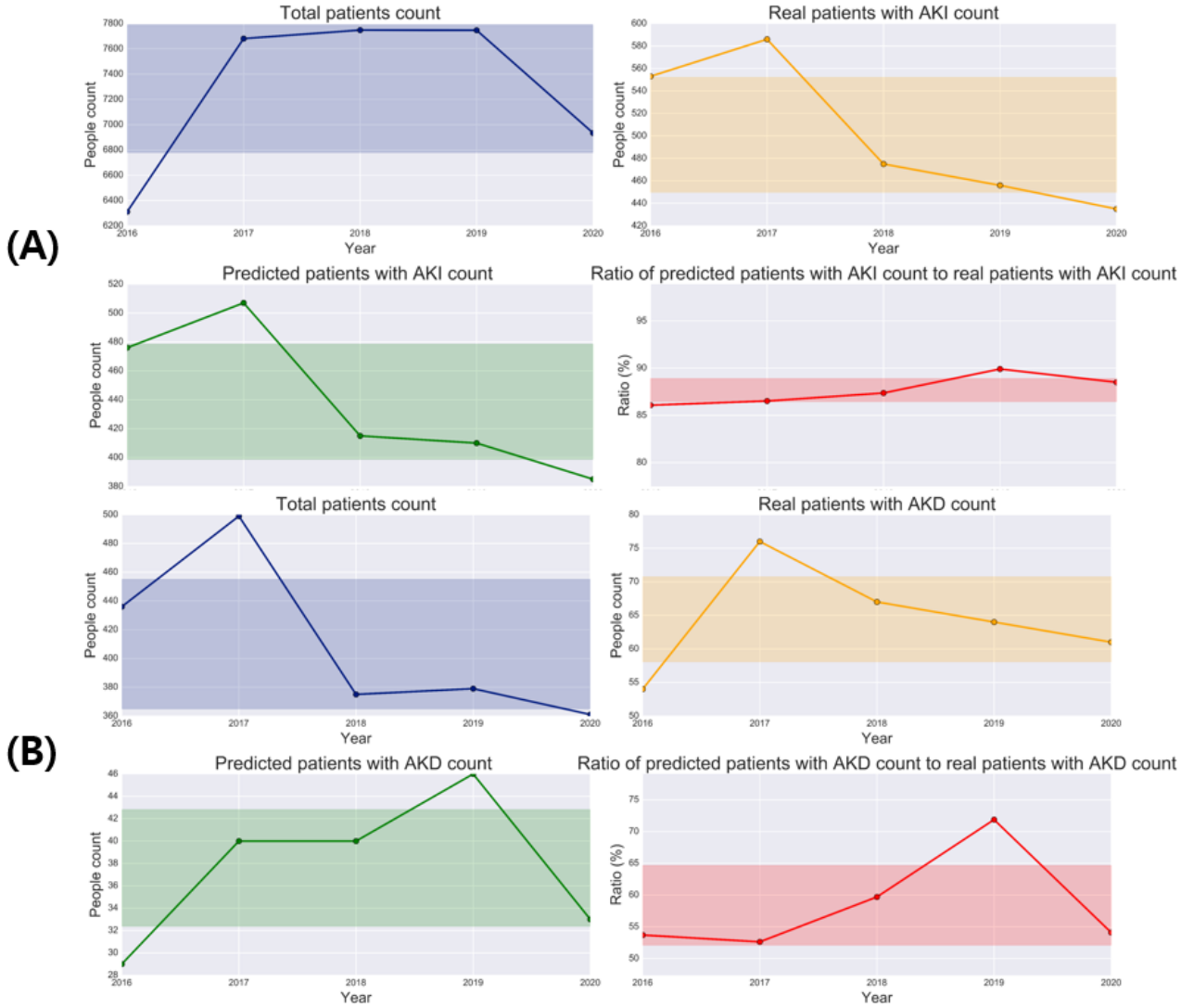

Discussion

Principal Findings

In this study, we developed a framework to assist in the management of AKI. The early AKI prediction model demonstrated high performance with an AUROC of 0.9053, while the early AKD prediction model achieved an AUROC of 0.8202. External validation results showed excellent model performance, although slight variations over time were observed, likely due to changes in disease incidence rates. Our study represents an integrated effort to identify patients with AKI early and, in cases where AKI occurs, classify them into high-risk and low-risk groups. Furthermore, we successfully refined and applied the KDIGO criteria to retrospective data, addressing its limitations and enhancing its applicability for AKI-related research in general ward patients.

Furthermore, our model predicted AKD in patients with AKI and showed an adjusted HR of 2.03 (95% CI 1.38-3.00) for events where eGFR decreased by 30% within 30 days from the onset of AKI. This adjusted HR was higher than that of the model developed using previous criteria, which had an HR of 1.66 (95% CI 1.17-2.34). This indicates that the refined criteria more precisely identify high-risk groups and suggests that the model developed using these criteria performs better. The AKI incidence was higher using the previous criteria, whereas the AKD incidence was higher using the refined criteria. This indicated that the previous criteria identified milder cases. Despite including milder cases, the model developed using the refined criteria showed a higher risk of a poorer prognosis than the previous criteria. Among patients tracked for more than 30 days after AKI, there were 934 and 482 patients before and after criteria advancement, respectively, with 138 and 103 showing poor prognosis. In other words, before the criteria advancement, 35 more patients with a poor prognosis were identified, but 452 additional AKI cases were identified. As shown in Figure S5 in Multimedia Appendix 4, many patients who met only the previous criteria included those whose AKI status was difficult to determine owing to the significant sensitivity to fluctuations. Regardless of how the baseline SCr level is imputed, it is

XSL•FO
RenderX

important in general wards to set a minimum increase criterion when applying a relative standard. Additionally, it is crucial to ensure that the baseline SCr level does not become too low when the SCr decreases. Advancements in labeling criteria should consider both patient prognosis and clinical settings. Excessive identification of AKI can lead to the detection of more patients with poor prognoses but may also result in inefficient allocation of medical resources.

Using the SHAP, factors contributing to AKI and AKD were identified and quantitatively measured. These results are largely aligned with the trends suggested by previous studies on risk factors and variable tendencies. Factors indicating poor patient status, such as low albumin or high alkaline phosphatase levels, were associated with increased AKI and AKD risks. Specifically, albumin is considered a critical biomarker closely related to renal function and may decrease in conditions associated with liver dysfunction [50]. Elevated WBC, heart rate, body temperature, and respiratory rate may indicate infection in patients [51], while the use of nephrotoxic drugs such as ANTIs, nonsteroidal anti-inflammatory drugs, and cytotoxic chemotherapeutic agents increases the risk of AKI [52,53]. Analysis of SHAP for AKD risk factors indicated that patients with increased WBC and high BT tended to recover renal function relatively quickly. Resolving infections appears to reduce the risk of AKD development in patients with AKI. ANTIs have emerged as significant risk factors for AKI and AKD. The difficulty in discontinuing these drugs due to patient conditions may exacerbate negative outcomes. Patients undergoing surgery are closely associated with AKI, although AKI following general anesthesia appears to show transient SCr fluctuations and overall health improvement following surgical resolution [54]. Renal function emerged as a crucial factor, consistent with existing clinical studies that state that baseline renal function is well-known as a major factor in AKI [55]. Additionally, higher age, heart rate, respiratory rate, total T-bil, aPTT, and CRP showed increased AKI risk, indicating that poorer patient condition may correlate with increased AKI incidence.

Similarly, poor renal function has been identified as a major risk factor for AKD. Elevated DBP reflects a tendency toward increased blood vessel volume after AKI, suggesting a possible correlation between increased volume and AKD [56,57]. AKI accompanied by surgery may recover relatively quickly due to transient hemodynamic changes during surgery, while cardiac surgery is known to have a relationship with AKI [58]. Contrast-induced nephropathy often shows peak levels approximately 3-5 days after exposure and often returns to baseline within 7-14 days, indicating relatively good recovery [59,60]. Conversely, ANTIs are associated with intrinsic AKIs such as acute tubular necrosis and slower recovery despite expected renal function impairment [61-63]. A high urine specific gravity suggests dehydration, which can often be corrected through fluid supplementation alone, leading to fast recovery [64,65]. Indicators reflecting infection showed that higher values tended to be associated with good renal recovery, while the significant infection marker CRP yielded ambiguous results. Because the interpretation of our model aligns with that of previous clinical studies, the patterns learned by the model are reasonable. Although biomarkers such as cystatin-C are commonly suggested in various studies, they were not used in this study due to their low measurement frequency. Future research should expand the features used in this study to include such biomarkers [66-68].

Limitations

This study had several limitations. First, owing to the lack of consensus on the AKI recovery criteria, the definitions had to rely on previous research findings. Second, there was a lack of data related to dialysis or kidney transplantation, which was addressed by excluding patients with an eGFR<60 mL/minute. As a result, caution is required when applying and interpreting the model for patients with preexisting chronic kidney disease or poor kidney function from the initial stages of hospitalization. However, the primary aim of this study was to predict “unexpected AKI.” Therefore, the analysis focused on patients with relatively preserved kidney function, who were considered to be at lower risk of AKI. Third, the study did not account for interventions or treatments before or post-AKI, which is crucial because patients’ preexisting conditions are significant factors in AKI and AKD. Fourth, in this study, the model was applied only in cases where AKI occurrences could be clearly identified, specifically when baseline SCr could be estimated at a specific time point and SCr measurements were available at that time. However, the proportion of AKI labels may differ from the actual occurrences. Therefore, despite the frequent measurement of SCr, caution is needed when applying and interpreting the model in situations where SCr measurements have not been conducted. Fifth, although our study used data extracted from different hospital information systems in various regions, it predominantly included data from Korean individuals. Therefore, we could not sufficiently consider racial diversity. Since there may be various differences, including kidney function, depending on ethnicity, future studies should include a multiethnic population.

Conclusions

This study introduces a machine learning framework aimed at assisting in the early management of AKI in general ward patients. To develop the model, we used retrospective data from general wards, refining the operational definition of AKI and externally validating our approach. Our findings demonstrate that AI-driven methods can enhance risk stratification and enable timely interventions. Beyond improving predictive accuracy, this study underscores the potential of AI to streamline clinical workflows, optimize resource allocation, and ultimately reduce the burden of AKI-related complications. Integrating such models into routine hospital practice may support proactive decision-making, allowing physicians to implement tailored interventions based on individual patient risk profiles. Future research should focus on prospective validation, real-time clinical integration, and incorporating additional biomarkers to improve model generalizability and clinical relevance.

XSL•FO
RenderX

Acknowledgments

This research was supported by the Ministry of Science and Information and Communication Technology, Korea, under the ICAN (Information and Communication Technology Challenge and Advanced Network of Human Resource Development) support program (IITP-2025-RS-2022-00156439) supervised by the Institute for Information & Communications Technology Planning & Evaluation and the Korea Institute for Advancement of Technology grant funded by the Korea Government (P0023675, Human Resource Development Program for Industrial Innovation).

Data Availability

The dataset used in this study and the models developed are not publicly available due to ethical restrictions, patient confidentiality, and institutional policies. However, they can be made available on reasonable request to the corresponding author.

Authors' Contributions

HL had full access to all the data in the study and was responsible for data integrity and the accuracy of the data analysis. This manuscript was written and edited entirely by the authors without the use of any generative AI tools. NC played a key role in conceptualizing and designing the study, contributed substantially to data acquisition, conducted statistical analyses, and drafted the manuscript. IJ focused on the study design, performed in-depth data analysis, and contributed to manuscript drafting. SA assisted with manuscript preparation and provided administrative and technical support. HG was instrumental in the study's concept and design, provided critical data, and supervised the research process. HL provided significant data for the study, secured funding, and offered administrative and technical support. All authors participated in data interpretation, critically reviewed the manuscript for important intellectual content, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Labeling-related process.
[DOCX File , 402 KB - jmir_v27i1e66568_app1.docx]

Multimedia Appendix 2
Variable-related information.
[DOCX File , 34 KB - jmir_v27i1e66568_app2.docx]

Multimedia Appendix 3
Model Development and Validation Process.
[DOCX File , 215 KB - jmir_v27i1e66568_app3.docx]

Multimedia Appendix 4
Comparison before and after labeling improvement.
[DOCX File , 527 KB - jmir_v27i1e66568_app4.docx]

Multimedia Appendix 5
Acute Kidney Injury Early Prediction Model Results.
[DOCX File , 212 KB - jmir_v27i1e66568_app5.docx]

Multimedia Appendix 6
Acute Kidney Disease Early Prediction Model Results.
[DOCX File , 237 KB - jmir_v27i1e66568_app6.docx]

Multimedia Appendix 7
Framework to Assist in Managing Acute Kidney Injury.
[DOCX File , 440 KB - jmir_v27i1e66568_app7.docx]

References

1. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin Pract 2012;120(4):c179-c184. [doi: 10.1159/000339789] [Medline: 22890468]

XSL•FO
RenderX

2. Hoste EAJ, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 2015 Aug;41(8):1411-1423. [doi: 10.1007/s00134-015-3934-7] [Medline: 26162677]
3. Kellum JA, Sileanu FE, Bihorac A, Hoste EAJ, Chawla LS. Recovery after Acute Kidney Injury. Am J Respir Crit Care Med 2017 Mar 15;195(6):784-791 [FREE Full text] [doi: 10.1164/rccm.201604-0799OC] [Medline: 27635668]
4. Wang H, Lambourg E, Guthrie B, Morales DR, Donnan PT, Bell S. Patient outcomes following AKI and AKD: a population-based cohort study. BMC Med 2022;20(1):229 [FREE Full text] [doi: 10.1186/s12916-022-02428-8] [Medline: 35854309]
5. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. Crit Care Med 2018;46(7):1070-1077. [doi: 10.1097/CCM.0000000000003123] [Medline: 29596073]
6. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, et al. Internal and external validation of a machine learning risk score for acute kidney injury. JAMA Netw Open 2020;3(8):e2012892 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12892] [Medline: 32780123]
7. Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nat Commun 2020;11(1):5668 [FREE Full text] [doi: 10.1038/s41467-020-19551-w] [Medline: 33168827]
8. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical decision support for in-hospital AKI. J Am Soc Nephrol 2018;29(2):654-660 [FREE Full text] [doi: 10.1681/ASN.2017070765] [Medline: 29097621]
9. Sun H, Depraetere K, Meesseman L, Cabanillas Silva P, Szymanowsky R, Fliegenschmidt J, et al. Machine learning-based prediction models for different clinical risks in different hospitals: evaluation of live performance. J Med Internet Res 2022;24(6):e34295 [FREE Full text] [doi: 10.2196/34295] [Medline: 35502887]
10. Heo S, Kang EA, Yu JY, Kim HR, Lee S, Kim K, et al. Time series aI model for acute kidney injury detection based on a multicenter distributed research network: development and verification study. JMIR Med Inform 2024;12:e47693 [FREE Full text] [doi: 10.2196/47693] [Medline: 39039992]
11. Zhang H, Wang AY, Wu S, Ngo J, Feng Y, He X, et al. Artificial intelligence for the prediction of acute kidney injury during the perioperative period: systematic review and meta-analysis of diagnostic test accuracy. BMC Nephrol 2022;23(1):405 [FREE Full text] [doi: 10.1186/s12882-022-03025-w] [Medline: 36536317]
12. Kamel Rahimi A, Ghadimi M, van der Vegt AH, Canfell OJ, Pole JD, Sullivan C, et al. Machine learning clinical prediction models for acute kidney injury: the impact of baseline creatinine on prediction efficacy. BMC Med Inform Decis Mak 2023;23(1):207 [FREE Full text] [doi: 10.1186/s12911-023-02306-0] [Medline: 37814311]
13. Nateghi Haredasht F, Antonatou M, Cavalier E, Delanaye P, Pottel H, Makris K. The effect of different consensus definitions on diagnosing acute kidney injury events and their association with in-hospital mortality. J Nephrol 2022;35(8):2087-2095 [FREE Full text] [doi: 10.1007/s40620-022-01323-y] [Medline: 35441981]
14. Jeong I, Cho NJ, Ahn SJ, Lee H, Gil HW. Machine learning approaches toward an understanding of acute kidney injury: current trends and future directions. Korean J Intern Med 2024;39(6):882-897 [FREE Full text] [doi: 10.3904/kjim.2024.098] [Medline: 39468926]
15. Li Y, Yao L, Mao C, Srivastava A, Jiang X, Luo Y. Early prediction of acute kidney injury in critical care setting using clinical notes. 2018 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 6, 2018; Madrid, Spain p. 683-686. [doi: 10.1109/bibm.2018.8621574]
16. Zimmerman LP, Reyfman PA, Smith ADR, Zeng Z, Kho A, Sanchez-Pinto LN, et al. Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements. BMC Med Inform Decis Mak 2019;19:16 [FREE Full text] [doi: 10.1186/s12911-019-0733-z] [Medline: 30700291]
17. Sato N, Uchino E, Kojima R, Hiragi S, Yanagita M, Okuno Y. Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Comput Methods Programs Biomed 2021;206:106129 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106129] [Medline: 34020177]
18. Machado GD, Santos LL, Libório AB. Redefining urine output thresholds for acute kidney injury criteria in critically Ill patients: a derivation and validation study. Crit Care 2024;28(1):272 [FREE Full text] [doi: 10.1186/s13054-024-05054-3] [Medline: 39135063]
19. Porschen C, Ernsting J, Brauckmann P, Weiss R, Würdemann T, Booke H, et al. pyAKI-An open source solution to automated acute kidney injury classification. PLoS One 2025;20(1):e0315325 [FREE Full text] [doi: 10.1371/journal.pone.0315325] [Medline: 39752439]
20. Sun S, Annadi RR, Chaudhri I, Munir K, Hajagos J, Saltz J, et al. Short- and long-term recovery after moderate/severe AKI in patients with and without COVID-19. Kidney360 2022;3(2):242-257. [doi: 10.34067/kid.0005342021]
21. Luo XQ, Yan P, Zhang NY, Luo B, Wang M, Deng Y, et al. Machine learning for early discrimination between transient and persistent acute kidney injury in critically ill patients with sepsis. Sci Rep 2021;11(1):20269 [FREE Full text] [doi: 10.1038/s41598-021-99840-6] [Medline: 34642418]
22. Liu CL, Tain YL, Lin YC, Hsu CN. Prediction and clinically important factors of acute kidney injury non-recovery. Front Med 2021;8:789874 [FREE Full text] [doi: 10.3389/fmed.2021.789874] [Medline: 35111778]

XSL•FO
RenderX

23. Cho NJ, Jeong I, Kim Y, Kim DO, Ahn S, Kang S, et al. A machine learning-based approach for predicting renal function recovery in general ward patients with acute kidney injury. Kidney Res Clin Pract 2024;43(4):538-547 [FREE Full text] [doi: 10.23876/j.krcp.23.330] [Medline: 38934029]
24. Rank N, Pfahringer B, Kempfert J, Stamm C, Kühne T, Schoenrath F, et al. Deep-learning–based real-time prediction of acute kidney injury outperforms human predictive performance. NPJ Digit Med 2020;3:139 [FREE Full text] [doi: 10.1038/s41746-020-00346-8] [Medline: 33134556]
25. Jiang X, Hu Y, Guo S, Du C, Cheng X. Prediction of persistent acute kidney injury in postoperative intensive care unit patients using integrated machine learning: a retrospective cohort study. Sci Rep 2022;12(1):17134 [FREE Full text] [doi: 10.1038/s41598-022-21428-5] [Medline: 36224308]
26. Shawwa K, Ghosh E, Lanius S, Schwager E, Eshelman L, Kashani KB. Predicting acute kidney injury in critically ill patients using comorbid conditions utilizing machine learning. Clin Kidney J 2021;14(5):1428-1435 [FREE Full text] [doi: 10.1093/ckj/sfaa145] [Medline: 33959271]
27. Zheng L, Lin Y, Fang K, Wu J, Zheng M. Derivation and validation of a risk score to predict acute kidney injury in critically ill cirrhotic patients. Hepatol Res 2023;53(8):701-712. [doi: 10.1111/hepr.13907] [Medline: 37041695]
28. Sparrow HG, Swan JT, Moore LW, Gaber AO, Suki WN. Disparate outcomes observed within Kidney Disease: Improving Global Outcomes (KDIGO) acute kidney injury stage 1. Kidney Int 2019;95(4):905-913 [FREE Full text] [doi: 10.1016/j.kint.2018.11.030] [Medline: 30819553]
29. Chawla LS, Bellomo R, Bihorac A, Goldstein SL, Siew ED, Bagshaw SM, et al. Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup. Nat Rev Nephrol 2017;13(4):241-257 [FREE Full text] [doi: 10.1038/nrneph.2017.2] [Medline: 28239173]
30. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. Lancet 2012;380(9843):756-766. [doi: 10.1016/S0140-6736(11)61454-2] [Medline: 22617274]
31. He J, Lin J, Duan M. Application of machine learning to predict acute kidney disease in patients with sepsis associated acute kidney injury. Front Med 2021;8:792974 [FREE Full text] [doi: 10.3389/fmed.2021.792974] [Medline: 34957162]
32. Neyra JA, Ortiz-Soriano V, Liu LJ, Smith TD, Li X, Xie D, et al. Prediction of mortality and major adverse kidney events in critically ill patients with acute kidney injury. Am J Kidney Dis 2023;81(1):36-47 [FREE Full text] [doi: 10.1053/j.ajkd.2022.06.004] [Medline: 35868537]
33. Daniels B, Havard A, Myton R, Lee C, Chidwick K. Evaluating the accuracy of data extracted from electronic health records into MedicineInsight, a national Australian general practice database. Int J Popul Data Sci 2022;7(1):1713 [FREE Full text] [doi: 10.23889/ijpds.v7i1.1713] [Medline: 37650032]
34. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New creatinine- and cystatin c-based equations to estimate GFR without race. N Engl J Med 2021;385(19):1737-1749 [FREE Full text] [doi: 10.1056/NEJMoa2102953] [Medline: 34554658]
35. Zhu K, Song H, Zhang Z, Ma B, Bao X, Zhang Q, et al. Acute kidney injury in solitary kidney patients after partial nephrectomy: incidence, risk factors and prediction. Transl Androl Urol 2020;9(3):1232-1243 [FREE Full text] [doi: 10.21037/tau.2020.03.45] [Medline: 32676406]
36. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
37. Mera-Gaona M, Neumann U, Vargas-Canas R, López DM. Evaluating the impact of multivariate imputation by MICE in feature selection. PLoS One 2021;16(7):e0254720 [FREE Full text] [doi: 10.1371/journal.pone.0254720] [Medline: 34320016]
38. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019;572(7767):116-119 [FREE Full text] [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
39. Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen 1936;7(2):179-188. [doi: 10.1111/j.1469-1809.1936.tb02137.x]
40. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
41. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. 2016 Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
42. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. Unknown: Curran Associates, Inc; 2017 Presented at: NeurIPS; December 4-9, 2017; Long Beach Convention Center, Long Beach.
43. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. ArXiv Preprint posted online on June 28, 2017. [doi: 10.48550/arXiv.1706.09516]
44. Haredasht FN, Vanhoutte L, Vens C, Pottel H, Viaene L, De Corte W. Validated risk prediction models for outcomes of acute kidney injury: a systematic review. BMC Nephrol 2023;24(1):133 [FREE Full text] [doi: 10.1186/s12882-023-03150-0] [Medline: 37161365]

XSL•FO
RenderX

45. Deo SV, Deo V, Sundaram V. Survival analysis-part 2: Cox proportional hazards model. Indian J Thorac Cardiovasc Surg 2021;37(2):229-233 [FREE Full text] [doi: 10.1007/s12055-020-01108-7] [Medline: 33642726]
46. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
47. Lei G, Wang G, Zhang C, Chen Y, Yang X. Using machine learning to predict acute kidney injury after aortic arch surgery. J Cardiothorac Vasc Anesth 2020;34(12):3321-3328. [doi: 10.1053/j.jvca.2020.06.007] [Medline: 32636105]
48. Ko S, Jo C, Chang CB, Lee YS, Moon Y, Youm JW, et al. A web-based machine-learning algorithm predicting postoperative acute kidney injury after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2022;30(2):545-554. [doi: 10.1007/s00167-020-06258-0] [Medline: 32880677]
49. Bredt LC, Peres LAB, Risso M, Barros LCAL. Risk factors and prediction of acute kidney injury after liver transplantation: logistic regression and artificial neural network approaches. World J Hepatol 2022;14(3):570-582 [FREE Full text] [doi: 10.4254/wjh.v14.i3.570] [Medline: 35582300]
50. Cheng T, Wang X, Han Y, Hao J, Hu H, Hao L. The level of serum albumin is associated with renal prognosis and renal function decline in patients with chronic kidney disease. BMC Nephrol 2023;24(1):57 [FREE Full text] [doi: 10.1186/s12882-023-03110-8] [Medline: 36922779]
51. Riley LK, Rupert J. Evaluation of patients with leukocytosis. Am Fam Physician 2015;92(11):1004-1011 [FREE Full text] [Medline: 26760415]
52. Naughton CA. Drug-induced nephrotoxicity. Am Fam Physician 2008;78(6):743-750. [doi: 10.1002/9780470372531.ch10]
53. Zhu W, Barreto EF, Li J, Lee HK, Kashani K. Drug-drug interaction and acute kidney injury development: a correlation-based network analysis. PLoS One 2023;18(1):e0279928 [FREE Full text] [doi: 10.1371/journal.pone.0279928] [Medline: 36607965]
54. Mikkelsen TB, Schack A, Oreskov JO, Gögenur I, Burcharth J, Ekeloef S. Acute kidney injury following major emergency abdominal surgery—a retrospective cohort study based on medical records data. BMC Nephrol 2022;23(1):94 [FREE Full text] [doi: 10.1186/s12882-022-02708-8] [Medline: 35247976]
55. Hu L, Gao L, Zhang D, Hou Y, He LL, Zhang H, et al. The incidence, risk factors and outcomes of acute kidney injury in critically ill patients undergoing emergency surgery: a prospective observational study. BMC Nephrol 2022;23(1):42 [FREE Full text] [doi: 10.1186/s12882-022-02675-0] [Medline: 35065624]
56. Jiang YJ, Xi XM, Jia HM, Zheng X, Wang M, Li W, et al. Risk factors, clinical features and outcome of new-onset acute kidney injury among critically ill patients: a database analysis based on prospective cohort study. BMC Nephrol 2021;22(1):289 [FREE Full text] [doi: 10.1186/s12882-021-02503-x] [Medline: 34433442]
57. Mercado MG, Smith DK, Guard EL. Acute kidney injury: diagnosis and management. Am Fam Physician 2019;100(11):687-694 [FREE Full text] [Medline: 31790176]
58. Cheruku SR, Raphael J, Neyra JA, Fox AA. Acute kidney injury after cardiac surgery: prediction, prevention, and management. Anesthesiology 2023;139(6):880-898. [doi: 10.1097/ALN.0000000000004734] [Medline: 37812758]
59. Isaka Y, Hayashi H, Aonuma K, Horio M, Terada Y, Doi K, et al. Guideline on the use of iodinated contrast media in patients with kidney disease 2018. Jpn J Radiol 2020;38(1):3-46. [doi: 10.1007/s11604-019-00850-2] [Medline: 31709498]
60. Kaliyaperumal Y, Sivadasan S, Aiyalu R. Contrast-induced nephropathy: an overview. Dr Sulaiman Al Habib Med J 2023;5(4):118-127. [doi: 10.1007/s44229-023-00040-3]
61. Paquette F, Bernier-Jean A, Brunette V, Ammann H, Lavergne V, Pichette V, et al. Acute kidney injury and renal recovery with the use of aminoglycosides: a large retrospective study. Nephron 2015;131(3):153-160 [FREE Full text] [doi: 10.1159/000440867] [Medline: 26389593]
62. Dalfino L, Puntillo F, Ondok MJM, Mosca A, Monno R, Coppolecchia S, et al. Colistin-associated acute kidney injury in severely ill patients: a step toward a better renal care? A prospective cohort study. Clin Infect Dis 2015;61(12):1771-1777. [doi: 10.1093/cid/civ717] [Medline: 26354965]
63. Molitoris BA, Berns JS. Pathogenesis and Prevention of Aminoglycoside Nephrotoxicity and Ototoxicity. Waltham, MA: UpToDate; 2022.
64. Yang X, Wu H, Li H. Dehydration-associated chronic kidney disease: a novel case of kidney failure in China. BMC Nephrol 2020;21(1):159 [FREE Full text] [doi: 10.1186/s12882-020-01804-x] [Medline: 32366263]
65. Mohsenin V. Practical approach to detection and management of acute kidney injury in critically ill patient. J Intensive Care 2017;5:57 [FREE Full text] [doi: 10.1186/s40560-017-0251-y] [Medline: 28932401]
66. Farrington DK, Surapaneni A, Matsushita K, Seegmiller JC, Coresh J, Grams ME. Discrepancies between cystatin C-based and creatinine-based eGFR. Clin J Am Soc Nephrol 2023;18(9):1143-1152. [doi: 10.2215/CJN.0000000000000217] [Medline: 37339177]
67. Nakano FK, Åkesson A, de Boer J, Dedja K, D'hondt R, Haredasht FN, et al. Comparison between the EKFC-equation and machine learning models to predict glomerular filtration rate. Sci Rep 2024;14(1):26383 [FREE Full text] [doi: 10.1038/s41598-024-77618-w] [Medline: 39487227]

XSL•FO
RenderX

68. Nateghi Haredasht F, Viaene L, Vens C, Callewaert N, De Corte W, Pottel H. Comparison between cystatin C- and creatinine-based estimated glomerular filtration rate in the follow-up of patients recovering from a stage-3 AKI in ICU. J Clin Med 2022;11(24):7264 [FREE Full text] [doi: 10.3390/jcm11247264] [Medline: 36555881]

Abbreviations

AI: artificial intelligence
AKD: acute kidney disease
AKI: acute kidney injury
ANTI: nephrotoxic antibiotic
aPTT: activated partial thromboplastin time
AUROC: area under the receiver operating characteristic
BUN: blood urea nitrogen
CAT: categorical boosting
CECT: contrast-enhanced computed tomography
CRP: C-reactive protein
DBP: diastolic blood pressure
eGFR: estimated glomerular filtration rate
HR: hazard ratio
ICU: intensive care unit
KDIGO: Kidney Disease: Improving Global Outcomes
SCr: serum creatinine
SHAP: Shapley additive explanations
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WBC: white blood cell

Edited by A Mavragani; submitted 17.09.24; peer-reviewed by K Shawwa, F Nateghi Haredasht; comments to author 26.12.24; revised version received 10.01.25; accepted 14.02.25; published 18.03.25.

Please cite as:
Cho NJ, Jeong I, Ahn SJ, Gil HW, Kim Y, Park JH, Kang S, Lee H
Machine Learning to Assist in Managing Acute Kidney Injury in General Wards: Multicenter Retrospective Study
J Med Internet Res 2025;27:e66568
URL: https://www.jmir.org/2025/1/e66568
doi:10.2196/66568
PMID:

©Nam-Jun Cho, Inyong Jeong, Se-Jin Ahn, Hyo-Wook Gil, Yeongmin Kim, Jin-Hyun Park, Sanghee Kang, Hwamin Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Influencing Primary Care Physicians' Intent to Refer Patients With Hypertension to a Digital Remote Blood Pressure Monitoring Program: Mixed Methods Study

Jennifer J Wu[1*], MD, MSED; Ross Graham[1,2*], MS; Julie Çelebi[1*], MD; Kevin Fraser[1*], MD; Geneen T Gin[1*], DO; Laurel Dang[1*], MD; Esmatullah Hatamy[1*], MD; Amanda Walker[1*], MS; Courtney Barbato[1*], MS; Ottar Lunde[3], MD; Lisa Coles[4], MD; Parag Agnihotri[3], MD; Cassandra Morn[1], MD; Ming Tai-Seale[1,3*], PhD

[1]Department of Family Medicine, University of California, San Diego, San Diego, CA, United States

[2]Department of Sociology, University of California, San Diego, San Diego, CA, United States

[3]Department of Medicine, University of California, San Diego, San Diego, CA, United States

[4]Community Care, University of California, San Diego, San Diego, CA, United States

[*]these authors contributed equally

Corresponding Author:
Jennifer J Wu, MD, MSED
Department of Family Medicine
University of California, San Diego
200 West Arbor Drive
San Diego, CA, 92103 - 8201A
United States
Phone: 1 6194719260
Fax: 1 6194719300
Email: j1wu@health.ucsd.edu

Abstract

Background: Primary care physicians' (PCP) referral rates to digital health programs are highly variable. This study explores whether knowledge of the digital remote blood pressure monitoring (RBPM) program and information on referral patterns influence PCPs' intention to refer patients.

Objective: This study aims to examine the relationship between PCPs' knowledge of the digital RBPM program and information on their own prior referral rates versus their own with their peers' referral rates and their likelihood to refer patients to the digital RBPM program.

Methods: This is a mixed methods study integrating quantitative analysis of electronic health record data regarding the frequency of PCPs' referrals of patients with hypertension to a digital health program and quantitative and qualitative analyses of survey data about PCPs' knowledge of the program and their intention to refer patients. PCPs responded to a clinical vignette featuring an eligible patient. They were randomized to either receive their own referral rate or their own plus their peers' referral rate. They were assessed on their intent to refer eligible future patients. Descriptive and multivariable linear regression analyses examined participant characteristics and the factors associated with their intent to refer patients. Narrative reasons for their intention to refer were thematically analyzed.

Results: Of the 242 eligible PCPs invited to participate, 31% (n=70) responded to the survey. From electronic health record data, the mean referral rate of patients per PCP was 11.80% (SD 13.30%). The mean self-reported knowledge of the digital health program was 6.47 (SD 1.81). The mean likelihood of referring an eligible patient (on a scale of 0 to 10, with 0 being not at all, and 10 being definitely) based on a vignette was 8.54 (SD 2.12). The own referral data group's mean likelihood to refer was 8.91 (SD 1.28), whereas the own plus peer prior referral data group was 8.35 (SD 2.19). Regression analyses suggested the intention to refer the vignette patient was significantly associated with their knowledge (coefficient 0.46, 95% CI 0.20-0.73; P<.001), whereas the intention to refer future patients was significantly associated with their intent to refer the patient in the vignette (coefficient 0.62, 95% CI 0.46-0.78; P<.001). No evidence of association was found on receiving own plus peer referral data compared with own referral data and intent to refer future patients (coefficient 0.23, 95% CI –0.43 to 0.89; P=.48).

XSL•FO
RenderX

Conclusions: Physicians' intention to refer patients to a novel digital health program can be extrapolated by examining their intention to refer an eligible patient portrayed in a vignette, which was found to be significantly influenced by their knowledge of the program. Future efforts should engage PCPs to better inform them so that more patients can benefit from the digital health program.

(J Med Internet Res 2025;27:e64933) doi:10.2196/64933

KEYWORDS

digital health; primary care; electronic health records; referral; hypertension; remote monitoring; remote blood pressure; digital technology; mobile phone; mixed method; quantitative analysis; linear regression; clinical information

Introduction

Hypertension is a leading independent and modifiable risk factor for several adverse cardiovascular outcomes including coronary artery disease, stroke, heart failure, and chronic kidney disease, as well as a wide variety of mental and metabolic conditions [1]. Long-term control of hypertension can be achieved through proven interventions including regular blood pressure (BP) monitoring [2]. Accurate BP measurements in the office can be a challenge, particularly if the patient has white coat syndrome [3].

Digital technology including digital remote blood pressure monitoring (RBPM) allows health care to conduct remote monitoring for many chronic diseases [4,5]. RBPM is expanding and shows increasing efficacy [6], cost savings, reduction in health care access disparities, reduced patient and physician burden, and ease in securing the diagnosis of white coat hypertension [7]. RBPM became more widely accepted during the COVID-19 pandemic when patients often avoided coming in for appointments [8] and health care providers were faced with a significant barrier to adequately care for their patients with hypertension and other chronic diseases. Many eligible patients with hypertension were not being seen in the office, and therefore could not be adequately or efficiently managed. Inevitably, there were also many patients developing new and sometimes uncontrolled hypertension that went unrecognized due to missed appointments. With the COVID-19 pandemic, broad legislative permissions facilitated at-home and telemedicine interventions necessary to pragmatically manage the pandemic, which led to an increase in at-home digital health care programs [9-11]. In 2020, the University of California San Diego (UCSD) implemented Project 1000 (P1000) an electronic health record (EHR)-integrated Bluetooth-enabled and team-based telemonitoring RBPM program available free for patients with uncontrolled hypertension. In this study, we explore whether knowledge of and information on referral patterns influence primary care physicians' (PCPs') adoption of the program through intention to refer patients to P1000.

We developed a physician survey to explore the reasons behind diverse referral patterns and analyzed the results. We hypothesize that (1) more knowledge of the digital health program is positively associated with physicians' proclivity to refer eligible patients, and (2) information on their own prior referral rates versus their own and their peers' referral rates could influence their likelihood to refer future patients. Using a mixed methods approach, we examined factors associated with physician intention to refer patients to P1000.

Methods

UCSD P1000 Program

P1000 is an innovative free EHR-integrated, Bluetooth-enabled, and team-based telemonitoring RBPM program piloted to 1000 patients whose in-office BP measured 140/90 or greater. Due to success, the program was expanded. PCPs were initially introduced to the digital RBPM program by the Population Health Services Organization (PHSO) on how to refer eligible patients. The original method of outreach was through automated phone calls, which resulted in low enrollment. An adjustment to the recruitment approach was made to involve patients' PCPs using direct electronic referrals within our Epic (Epic Systems) EHRs system. Once the referral is entered, a digital health program team member reviews eligibility, acknowledges the referral in the EHR, and reaches out to the referred patient. Participating patients are shipped a Bluetooth-enabled digital BP cuff that automatically transmits BP data to the patient portal of the EHR via a smartphone app. Enrollees are supported by a digital health specialist via telephone or in the patient's home for device setup and any technical issues. The devices are provided free to all enrollees and retrieved or replaced only in the case of a defective device, or if the devices are having accuracy issues and used for more than one year. The EHR system stratifies BP measures into 3 risk groups: normal, high or priority, and critical, and presents them on a daily dashboard. Nurse care managers review the dashboard and make treat-to-target adjustments per protocol, encompassing medication adjustments and behavioral change recommendations. Outcomes of the referral are documented by the team member reviewing the patient dashboard and a message is sent back to the referring provider. Since 2020, over 2500 patients have been referred to the program. BP control data analyzed since the inception of this program has shown improvement in BP control for those referred to the program regardless of participation in or continuation with the digital RBPM program [12]. However, analysis of referral patterns suggested heterogeneous referral rates among PCPs. PHSO funded the cost of the program, which remains ongoing at UCSD.

Assessing Factors in Digital Health Adoption

Literature suggests that adoption of new primary care practices is influenced by many factors including knowledge of available resources and programs, implementation of strategies that target the needs of the PCP's practice [13,14], and attitudes and perspectives of the clinician [7,15-17]. In particular, physician uptake of digital health care initiatives hinges upon clearly

XSL•FO
RenderX

demonstrated clinical improvements for their patients, ease of integration with their clinical workflow, and easy-to-use technology that is also amply supported with technical assistance [18]. In addition, the adoption of new treatment options can be influenced by knowledge of or observation of their peer's behavior in the case of pharmaceuticals [19], but not in other instances, for example, the uptake of EHRs [20]. Notably, PCPs are likely to be early adopters of digital health care initiatives [21]. Therefore, performance data on referral rates was deliberately included to study the influence of current performance data on the decision to refer.

We explored the reasons behind such diverse practice patterns by developing a physician survey. We had three goals: (1) assess physicians' knowledge of the digital health program, (2) explore their proclivity to refer patients who clearly meet eligibility criteria, as portrayed in a vignette, and (3) inform them of their own prior referral rates, or their own and their peers' referral rate, and reassess their likelihood to refer future patients.

Design

We used an exploratory, sequential design for this mixed methods study. PCPs were recruited from Family Medicine (FM), General Internal Medicine (GIM), and Community Care (CC) at UCSD Health. Referral data for physicians in these departments were provided by an Epic analyst. Each physician was sent an individual survey link via email.

Various outreach approaches were used. FM physicians received a 10-minute presentation on the survey project and were given time to complete the survey, which was sent prior to the meeting from the FM Clinical Services Chief. GIM, Geriatrics, and CC practice physicians were informed at departmental meetings that an invitation to participate would be emailed to them and encouraged them to reply. They did not receive a formal presentation nor allotted time to answer the survey. As such, outreach to GIM, Geriatrics, and CC was less extensive.

All eligible physicians were randomized using simple randomization via a random number generator in the CRAN package of statistical analysis software "R" (R Core Team) into 1 of 2 study arms, with a link to a personalized survey that included either their own referral data (Arm A) or their own referral data plus their peers' (own+peer) referral rates (Arm B). Two email reminders were sent to PCPs who had not responded. All PCP identifiers were kept confidential.

In the context of a quality improvement (QI) program, the initial introduction of the digital health program to PCPs in 2020 was the QI intervention. The survey in late 2023 assessed what changes they had made to their practice, what impact it had had on their practice, and what they had learned as a result of this QI intervention. Responses to the survey satisfied the requirements for maintenance of certification (MOC) credits for their boards.

The survey contained 5 stages (Figure 1). First, PCPs rated their knowledge of the digital health program on a scale of 0 to 10 (0 being no knowledge and 10 being extremely knowledgeable) and provided a free text response explaining their score (Q1). Second, they assessed their likelihood to refer an eligible patient on a scale of 0 to 10 (0 being not at all and 10 being definitely) by their response to a clinical vignette, a valid method for distinguishing between physician practice [22]. The vignette portrayed a patient that meets the inclusion criteria of the digital health program—a 65-year-old Black male with hypertension whose BP was 150/95, and whom the provider had worked with for 2 years during which time BP remained elevated (Q2). Third, based upon random assignment, PCPs were either shown information on their own referral rate (Arm A) or shown their own and their peers' referral rates concurrently (Arm B). The referral rates presented to respondents covered the period from October 2020 to October 2022 and were pulled from EHR data. In the fourth stage, their intention to refer future patients to the digital health program was assessed again (Q3). Fifth, qualitative reflections were gathered on what they learned, understood, and planned to change in their future practice regarding the digital RBPM program. Survey questions are included in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. The sequence of survey questions, including at what point randomization bifurcates available data (own or own+peer) participants could view when answering survey questions. PCP: primary care physician; QI: quality improvement.

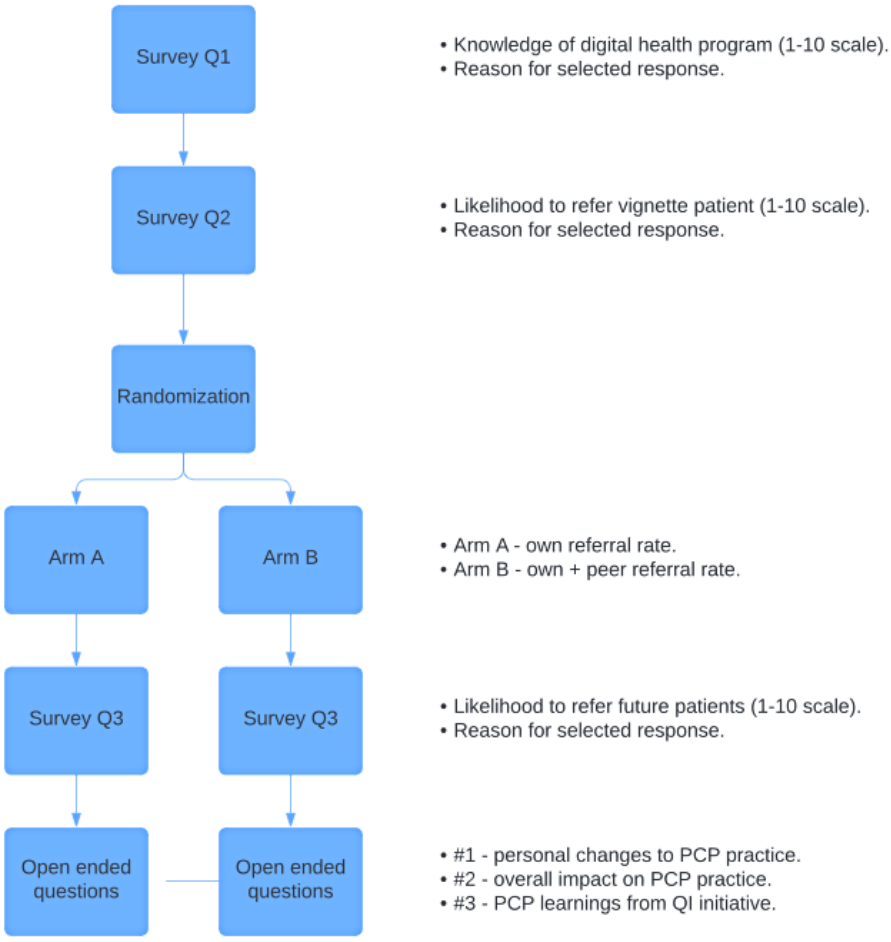

Analysis

Univariate and bivariate descriptive statistics were examined. Multivariable linear regression models were used to assess factors associated with the intention to refer the patient in the vignette and the likelihood to refer future patients after seeing own versus own+peer referral rates. For the qualitative analysis, one member (RG) of the research team developed 4 codebooks (knowledge, vignette, future referral likelihood, and program takeaways) using an inductive grounded theory approach [23]. This involved iteratively refining and condensing themes extracted from a random subset of the qualitative data. In partnership with a second team member (AW), these codes were adjusted and refined. Then, these 2 research team members coded a subset of the data to ensure interpretive alignment (Cohen κ=0.84), before the remaining data were thematically coded.

Ethical Considerations

The project was reviewed by the UCSD Aligning and Coordinating Quality Improvement, Research, and Evaluation (ACQUIRE) Committee. The ACQUIRE Committee approved this project as a QI project and its approval included a decision that the project would satisfy the requirements for maintenance of certification (MOC) credit through UCSD's MOC Portfolio Program. Physicians who completed the survey were eligible for MOC credit. No additional compensation was provided to participants.

All PCP identifiers were kept confidential. The project manager (AW) distributed unique survey links via Qualtrics, ensuring that each physician received their individual or individual-plus-peer referral data. Survey responses were stored in a HIPAA-compliant OneDrive folder accessible only to the project manager. Prior to sharing with the research team, responses were deidentified. An identified dataset was shared solely with the Senior Health Professions Education Specialist, Accreditation and Educational Development, for the purpose of processing MOC credit. Our survey satisfied the Checklist for Reporting the Result of Internet E-Surveys (CHERRIES) [24], shown in Multimedia Appendix 2.

Results

Participant Characteristics and Quantitative Results

Table 1 provides an overview of the sample (N=70). The median range of years in practice was 12.5 (IQR 11-15). In total, 57% (n=40) of respondents were female, 64% (n=45) of respondents were from academic FM, 24% (n=17) of respondents were from academic GIM and Geriatrics, and 11% (n=8) of respondents were from CC (comprised of Internal Medicine or FM physicians). The average prior familiarity with the Digital Health Program was 6.47 (SD 1.81). The average likelihood of referring

XSL•FO
RenderX

the patient from the vignette (on a scale of 0 to 10, with 0 being extremely unlikely and 10 being extremely likely) was 8.54 (SD 2.12). A total of 54% (n=38) of the sample received their own+peers' referral rates, while 46% (n=32) of the sample received their own referral rates. The average referral rate of physicians from EHR data was 11.80% (SD 13.30%), while the average peer referral rate was 11.90% (SD 9.60%). The average likelihood of referring patients in the future was 8.61 (SD 1.83). Participants were randomized into 2 treatment groups—whether they were shown only their own referral rate (Arm A), or their own+peers' referral rates (Arm B).

Table 2 compares the likelihood of referring patients within the 2 treatment groups—whether they were shown only their own referral rate, or their own and their peers' referral rates.

Table 3 displays results from multivariable linear regression analyses to assess factors that could be associated with physicians' likelihood to refer an eligible patient presented in the vignette (model 1) and eligible patients in the future (model 2). Explanatory variables in model 1 included length of practice (dichotomized by ≤10 years, and >10 years), PCP gender (dichotomized by female, and male or N/A or prefer not to answer), prior knowledge of the digital RBPM program, and department (reference group was FM). Prior knowledge of the digital RBPM program was significantly associated with the likelihood of referring the eligible vignette patient (coefficient 0.47; SE 0.13; 95% CI 0.20-0.73; P<.001).

The intention to refer future eligible patients was the dependent variable for model 2 whose explanatory variables included random assignment of information received (reference group was shown own referral rate), the self-rated likelihood to refer the vignette patient, and all explanatory variables from model 1. The likelihood of referring the vignette patient was significantly associated with their intention to refer future patients (coefficient 0.62, SE 0.08, 95% CI 0.46 to 0.78; P<.001).

Table 1. Participant demographic characteristics and department.

	Arm A—own (n=32), n (%)	Arm B—own + peer (n=38), n (%)	Total (n=70), n (%)
Years of practice			
1-5	8 (25)	14 (37)	22 (31)
6-10	6 (19)	5 (13)	11 (16)
11-20	10 (31)	8 (21)	18 (26)
More than 20	8 (25)	7 (18)	15 (21)
Prefer not to answer or N/A[a]	0 (0)	4 (11)	4 (6)
Sex			
Female	19 (59)	21 (55)	40 (57)
Male	11 (35)	14 (37)	25 (36)
Prefer not to answer or N/A	2 (6)	3 (8)	5 (7)
Department			
Family Medicine	19 (59)	26 (68)	45 (64)
Internal Medicine and Geriatrics	7 (22)	10 (26)	17 (24)
Community Care	6 (19)	2 (5)	8 (11)

[a]Not applicable.

Table 2. Quantitative measures of likelihood to refer to the digital RBPM[a] program, familiarity with the program, along with primary care physician referral rates prior to the survey, separated by a randomized arm (receiving own vs own+peer referral data).

	Arm A—own (n=32), mean (SD)	Arm B—own + peer (n=38), mean (SD)	Total (n=70), mean (SD)
Familiarity with digital RBPM program	7.09 (1.42)	5.92 (1.95)	6.47 (1.81)
Likelihood to refer patient in vignette	8.94 (1.61)	8.11 (2.45)	8.54 (2.12)
Own referral rate from EHR[b] data	11.2% (11.6%)	12.3% (14.6%)	11.80% (13.30%)
Peer referral rate from EHR data	12.1% (10.4%)	11.8% (9.0%)	11.90% (9.60%)
Likelihood to refer patients in the future	8.91 (1.28)	8.35 (2.19)	8.61 (1.83)

[a]RBPM: remote blood pressure monitoring.

[b]EHR: electronic health record.

XSL•FO
RenderX

Table 3. Factors (demographics, knowledge of digital RBPM[a] program) associated with vignette referral and likelihood to refer future eligible patients.

	Model 1[b]: likelihood to refer eligible patient in vignette				Model 2[c]: intention to refer future eligible patients			
	Coefficient	SE	*P* value	95% CI	Coefficient	SE	*P* value	95% CI
(Intercept)	6.14	1.02	<.001	4.11 to 8.18	2.21	0.93	.02	0.35 to 4.08
Practice length: >10 years (reference <10 years)	–0.28	0.54	.61	–1.36 to 0.80	0.31	0.34	.37	–0.37 to 0.99
PCP[d] gender: male or N/A[e] or prefer not to answer (reference: female)	–0.63	0.48	.20	–1.60 to 0.33	0.02	0.31	.96	–0.59 to 0.63
Prior knowledge of digital RBPM program	0.46	0.13	<.001	0.20 to 0.73	0.11	0.10	.26	–0.08 to 0.30
Department: CommCare (reference: Fam Med)	–1.09	0.72	.14	–2.53 to 0.36	0.48	0.47	.32	–0.46 to 1.42
Department: IntMed/Ger (reference: Fam Med)	–0.14	0.58	.82	–1.30 to 1.03	0.22	0.36	.56	–0.51 to 0.94
Shown referral rate: own+peer (reference: own)	N/A	N/A	N/A	N/A	0.23	0.33	.48	–0.43 to 0.89
Likelihood to refer patient in vignette	N/A	N/A	N/A	N/A	0.62	0.08	<.001	0.46 to 0.78

[a]RBPM: remote blood pressure monitoring.

[b]Model 1: number of observations=69; R^2=0.14.

[c]Model 2: number of observations=69; R^2=0.53

[d]PCP: primary care physician.

[e]Not applicable.

Qualitative Results

After participants were asked to numerically score their knowledge of the digital RBPM program for hypertension management provided by UCSD Population Health, they were then asked to provide free text comments on their score. A total of 23 participants expressed high perceived knowledge of the program, scoring themselves 8-10 on the 10-point scale; 28 participants expressed moderate perceived knowledge (score 6-7) of the program; and 19 participants expressed low perceived knowledge (score <6) of the program. Of these, 13 participants exhibited low knowledge about the existence or basic function of the PHSO digital program, while 6 participants were aware of the program, but had low knowledge about patient eligibility criteria, which in turn meant they did not use it. Table 4 summarizes the findings.

After participants were asked to numerically score how likely they were to refer an eligible vignette patient with BP 150/95 mm Hg, they were asked to explain their reasons why. Primarily, it was clinical reasons that justified referral to the program, for example, “blood pressure is uncontrolled,” as well as the additional help provided by the program to physicians in monitoring patients of this type, for example, “Having additional oversight for this patient’s blood pressure would be beneficial especially when we as providers can only see the patient on an intermittent basis,” “[patient is] more likely to continue to engage if more people are touching base with him.” Others stated that the program is “not appropriate for all cases” and that the hypothetical patient’s trajectory also “depends on other factors.” Those who suggested they would not refer had hesitations about the program rather than clinical need; one participant said, “I forget about this resource”; while another said, “many older patients are not interested in the extra hassles and intimidated by technology.”

Physicians were randomized to be shown their referral rate to the program (Arm A) versus their own referral rate and the referral rate of their peers in the same clinic (Arm B) and asked to score the likelihood that they would refer patients to the program in the future. After giving their score, they were asked to explain the reason for their score. Nine participants (n=4 Arm A; n=5 Arm B) discussed how they already use the program and intend to continue doing so, for example, “I have been using the program more recently and have seen improvements in my hypertensive patients.” Six participants in Arm B and 2 participants in Arm A professed that their newfound understanding of the utility of the program would lead to greater referrals, for example, “Seeing this information prompts me to recognize that maybe I am not taking advantage of this resource as much as I thought.” However, some expressed hesitations about future referrals. Notably, 9 participants in Arm A and 4 participants in Arm B hesitated about further patient adoption of the program, worrying about technological aptitude: “I find most of my patients do not enroll in the program because they cannot manage the technology involved”; about insurance coverage: “I don’t know about insurance coverage for this service”; or simply refusing: “most patients referred decline.” Five participants from Arm A and 1 participant from Arm B had no intention to refer.

XSL•FO
RenderX

Table 4. Quantitative score of perceived knowledge of the digital program (0-10) alongside qualitative responses.

Knowledge of the digital RBPM[a] program and code, and quote
Very knowledgeable (8-10), n=23
"Used the program frequently"
"[I] Feel confident in how to access their services, how to view the home blood pressures, and how to communicate with the main nurses, pharmacists ... etc."
Somewhat knowledgeable (6-7), n=28
"Refer many patients, see some messages, but not truly aware on logistics and options of services provided"
Lacking knowledge of program features (<6), n=13
"I have no idea on how this program works and how the data collected help me directly manage my patients. Or whether the Population health will manage these patients for me."
Lacking knowledge of patient eligibility (<6), n=6
"Unclear which patients are eligible."

[a]RBPM: remote blood pressure monitoring.

To meet the requirements for MOC, participants were asked to reflect on the QI intervention, and how it would affect their future practice. They were asked about what they would change personally and clinically because of the intervention, as well as what their key learnings and takeaways were. Participants were excited by how the clinical data would improve their practice and patient gaps. Some simply appreciated having more, and more continuous, data: "I have gained valuable data about my patients from their home monitoring, and have been able to, and felt more comfortable to intervene with medical treatment when needed for blood pressure that is uncontrolled." Many highlighted how at-home monitoring clearly demonstrated the extent of overdiagnosis of hypertension stemming from white coat syndrome: "I also found that patients with anxiety and white coat hypertension were less likely to be overtreated with medications when referred to this program." Having contemporaneous data would lead to better clinical outcomes, many felt: "This blood pressure monitoring program and hypertension management program can lead to better cardiovascular outcomes and better overall health for patients because of the real-time monitoring."

A total of 29 (41%) providers expressed a determination to refer to PHSO more often as part of their clinical practice, acknowledging its value in comments, including the following: "I plan to utilize the PHSO HTN program more often. Also to use it when indicated. This improves patient's outcomes." Some went further, noting the availability of collaborative programs within UCSD Health that may improve clinical outcomes and reduce clinician workload: "Better understanding of health systems programs that can address complex cardiovascular care from the primary care perspective." A few noted that the program brought them confidence and assurance: "Participating in the program reassured me about the quality and accuracy of their reading."

Another theme emerging from the reflections was that providers were happy with how the patients were empowered to participate in their own care through the program. One said that patients "were excited when starting digital health monitoring," while another said, "I have learned the benefit of patient participation in collecting data to view where they are in their health." One went further, saying "patients can be more engaged in their health when we make access to decision-making easier." A patient education theme also emerged: "The digital health options have increased patient compliance, participation, and knowledge in the management of hypertension."

Finally, more physicians in Arm B (n=10) than in Arm A (n=6) expressed an increased desire and capacity to harness medical technology in the future, particularly in concert with their patients, due to the QI intervention. One provider suggested they underestimated the ability of their older people, nonnative speaking patient population to do this: "Maybe not underestimating PHSO ability to provide services to my elderly Mandarin-speaking population through the use of interpreters as many of them are still tech-savvy enough to use smartphones so could likely incorporate similar technology at home." By using technology, patients and providers could be in closer contact, but also better connected to the wider medical infrastructure for a more thorough, networked care experience: "It renewed a sense of institutional support for keeping our patients healthy. It also renewed my appreciation that there are teams of health care staff in our system who stand ready to assist us in our work and keep us focused." Finally, the program enabled physicians to practice state-of-the-art medicine: "I feel like the QI project has made my practice more up to date with current medical care trends."

Table 5 summarizes key facilitators and barriers to physician participation in the program. There are 4 key facilitators to ongoing physician participation in the program. First, getting an initial referral underway is important, as the benefits of the program become apparent quickly and diffuse many initial apprehensions. Second, having BP data at more regular intervals provides greater data clarity to inform clinical decision-making, and identifies instances of white coat hypertension. Third, physicians are encouraged to continue in the program by 2 improvements to patients; first, their clinical care, but also the education, empowerment, and agency they receive by participating in their care in collaboration with their doctor. Fourth, physicians can benefit from advances in medical

XSL•FO
RenderX

technology and are well supported by a team while doing so. Four key barriers to physician participation in the program also surfaced. First, PCPs express concerns about patients' ability to adhere to the program, like an inability to use the technology or reluctance to take BP measurements. Second, some PCPs view the program as reducing their control over managing their patient's hypertension, and possibly increasing their workload in the process rather than reducing it. Third, some physicians do not understand the program structure, thinking it costs patients or requires insurance. Fourth, physicians were unclear about patient eligibility to enroll in the program.

Table 5. Qualitative categories of facilitators and barriers to physician use of the digital RBPM[a] program.

Subtheme	Quote
Facilitators to PCP[b] participation in the program	
Getting PCPs started is vital, once they begin using the program they see improvements in patients and understand it better	"I discovered the benefits of home BP readings which directed to manage and control my patients' Hypertension. But, I was not sure that their readings were accurate or the quality of their devices since they were not calibrated. Participating in the program reassured me about the quality and accuracy of their reading."
Continuous clinical oversight improves PCP understanding of hypertension, reduces white coat and other clinical errors	"I also found that patients with anxiety and white coat hypertension were less likely to be overtreated with medications when referred to this program"
Observing 2-fold patient benefit, whereby patient care is improved but also patient experience and agency is heightened	"I have learned the benefit of patient participation in collecting data to view where they are in their health"
Program is a way to harness advancing medical technology and increase patient touchpoints with team-based support	"I'm learning about the power of harnessing remote digital technology to provide another layer of support to patients outside of the office. The more regular the medical contact with a poorly controlled medical problem the better."
Barriers to PCP participation in the program	
Concerns about patient capacity to adhere to the program	"Many older patients are not interested in the extra hassles and intimidated by technology"
Fear over losing control over how patient hypertension is managed	"I feel like it is taking away from my own work for helping patients with their HTN"
Misunderstanding of program funding structure	"I don't know about insurance coverage for this service"
Misunderstanding of eligibility criteria for the program	"I was not aware who qualifies for the program"

[a]RBPM: remote blood pressure monitoring.

[b]PCP: primary care physician.

Discussion

Principal Results

Quantitative analyses suggest knowledge of the program was significantly associated with physicians' likelihood to refer the eligible vignette patient, which itself was significantly associated with the intention to refer future patients. Those with high knowledge scores frequently explained how this knowledge came from frequent use and experience with the program, and those with low knowledge scores frequently expressed reservations about basic eligibility criteria and the clinical and practical utility of the program, which indicates lower use.

Similarly, those more inclined to refer the vignette patient tended to mention their experience receiving additional help referring prior patients with these clinical needs and were less likely to be concerned about whether their patients could use the technology—other studies indicate fears about patients using digital RBPM are often overestimated by physicians [9]. Some of those who were reluctant to refer the vignette patients expressed the need for more clinical context, which, in essence, is what the program offers with minimal additional clinical or administrative burden to PCPs. Additionally, as seen in existing literature [15], some providers are hesitant to refer as they feel it affects the clinician's ownership of the patient's care.

Despite some participants identifying barriers to referring patients, reflections from PCPs about the program were generally positive. Many referenced the improved clinical outcome it produced, and its capacity to provide patients with education and enhanced agency over their care as key reasons why they would use the program in the future. One important clinical insight, that supports other studies of digital monitoring for hypertension [25], is that the program helps clarify the incidence of white coat syndrome—one participant said "I also found that patients with anxiety and white coat hypertension were less likely to be overtreated with medications when referred to this program"—which wastes clinical resources, as well as physician's and patient's time [3]. One previous study discovered as many as 8% of their cohort had white coat syndrome [25] when enrolled in a digital RBPM program; our qualitative data gently supports this clinical justification for digital RBPM programs.

Criteria for hypertension can vary in different clinical contexts, and misalignment between being categorized as hypertensive in a PCP native data environment (>130/80 mm Hg) and being a candidate for the population health program (>140/90 mm

XSL•FO
RenderX

Hg) may have left some physicians questioning the appropriate time to refer. How to best prompt physicians to consider this resource for eligible patients remains a somewhat unanswered question. In addition to PCP awareness of clinical eligibility criteria for the program—which qualitative and quantitative data suggests is mixed—it is important for physicians to be aware of the program's payor-neutral status. Patients face no cost-sharing in this program. Some PCPs incorrectly believed that insurance status and financial limitations might preclude patient participation.

Furthermore, it is vital that health care organizations ensure that team-based services like these are perceived as supportive partners rather than adding to already-excessive administrative burden. In the future, it would be prudent to solicit physician input on determining the most appropriate channels for communication regarding patients comanaged by the PHSO digital health program and PCP. This applies to communicating eligibility, enrollment, and ongoing monitoring. Several physicians who used the program frequently noted that it not only supported and empowered patients, but also actively reduced physician burden, particularly with patients who had poor access to consistent health care for reasons of socioeconomics, geography, or age-related mobility issues. Simple BP follow-up appointments can easily clog a physician's already limited availability, especially if no-shows are frequent, while also placing demands on patients ill-suited to frequent clinic visits due to chronic disease burden. A collaborative emphasis that does not place responsibility solely on the PCP is most appropriate [9].

Some physicians reported concerns regarding patient receptivity to digital monitoring tools, whether it be older patients who are less technologically savvy, patients who are non-English speakers, or those with limited financial resources—highlighting the need to address these potential barriers more proactively to enhance the footprint of this program to those in need. There may be patients with negative attitudes about having others involved in their care, and perhaps that perception reflects PCP practice habits if physician autonomy in care management tends to be prioritized or viewed as superior to referring to other members of the health care team. Though one might intuit that earlier career physicians may be more receptive to team-based care initiatives such as this program as a new way of practice, our data did not show evidence of an effect in the likelihood to refer across years of practice groups. Furthermore, physician concerns regarding their autonomy in patient management did arise in qualitative findings, highlighting the need to emphasize the collaborative nature of such initiatives which recognizes the physician as an essential member of a patient's care team.

This study found no evidence of an effect of seeing own+peer's referral rates compared with seeing only their own referral rate. That said, it is unclear whether awareness of the clinical practices of one's peers might warrant an immediate impact on one's own practices but over time could be a factor of influence. As such, future studies might follow a cohort for a longer time, sharing peer referral data rates, to determine if the effect has a greater influence over time.

Limitations

Outreach approaches varied. FM physicians received a 10-minute presentation about the survey project and were given time to complete the survey, which had been sent prior to the meeting via an email invitation from the FM Clinical Services Chief. GIM, Geriatrics, and CC practice physicians did not receive a formal presentation nor allotted time to complete the survey, which may have contributed to the underrepresentation of these 2 practices.

Patient referrals were attributed to the patient's PCP in the EHR. Patients may not see their attributed PCP at every visit, so a colleague, Resident Physician, or Advanced Practice Provider may have made the referral. Thus, the referral rates as abstracted might not be entirely representative of any given physician's actual personal utilization of the program based upon the data, though we presume that this general foundation of comanaging patients in academic practices likely evens out such that it still is an adequate approximation of their referral patterns. Given our sample size is relatively small and collected at a single institution, this study may be underpowered and limited in its generalizability. Future studies might seek larger PCP populations across multiple institutions situated in diverse socioeconomic and geographic contexts with variable overall chronic disease burden. They might also investigate whether differences exist between FM, Internal Medicine, and other departments, or between academic and community settings, in the acceptance and utilization of digital health for chronic disease care. Finally, the program at the time of the study was limited to English-speaking patients only, limiting the potential patient pool eligible for the study.

Conclusions

This study indicates that PCP knowledge and understanding of digital health monitoring programs, in terms of clinical outcome, practical application, and the reduction of work burden, is crucial to increased use and thus widespread effectiveness of these programs. Contextualizing PCP referrals with their peers' referrals showed no evidence of an effect in furthering their intention to adopt the program. Since remote patient monitoring has an increasing volume of empirical support for improving clinical outcomes, we encourage more robust communication of the effectiveness of these programs, their ease of use, and the availability of support to PCPs to increase enrollment and improve population health.

Acknowledgments

We would like to acknowledge all the University of California San Diego (UCSD) faculty who contributed to our data by completing our survey regarding the remote digital health monitoring program. We appreciate the diligent and dedicated work of our UCSD Family Medicine research team (RG, CB, and AW) without whom the analysis, interpretation of data, and research

XSL•FO
RenderX

paper writing details would not be possible. Finally, special thanks to MT-S for bringing the SPARK team together (Jennifer J Wu, Julie Çelebi, Kevin Fraser, Geneen T Gin, Laurel Dang, and Esmatullah Hatamy) and for igniting the SPARK in all of us!

Data Availability

All data generated or analyzed during this study are included in this published article under Multimedia Appendix 3.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Multimedia survey supplement.
[DOCX File , 21 KB - jmir_v27i1e64933_app1.docx]

Multimedia Appendix 2
CHERRIES framework.
[XLSX File (Microsoft Excel File), 15 KB - jmir_v27i1e64933_app2.xlsx]

Multimedia Appendix 3
Deidentified data.
[XLS File (Microsoft Excel File), 106 KB - jmir_v27i1e64933_app3.xls]

References

1. Oparil S, Acelajado MC, Bakris GL, Berlowitz DR, Cífková R, Dominiczak AF, et al. Hypertension. Nat Rev Dis Primers 2018;4:18014 [FREE Full text] [doi: 10.1038/nrdp.2018.14] [Medline: 29565029]
2. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Erratum: author correction: impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digital Med 2018;1:17 [FREE Full text] [doi: 10.1038/s41746-018-0027-3] [Medline: 31305588]
3. Franklin SS, Thijs L, Hansen TW, O'Brien E, Staessen JA. White-coat hypertension: new insights from recent studies. Hypertension 2013;62(6):982-987. [doi: 10.1161/HYPERTENSIONAHA.113.01275] [Medline: 24041952]
4. Zhang Y, Peña MT, Fletcher LM, Lal L, Swint JM, Reneker JC. Economic evaluation and costs of remote patient monitoring for cardiovascular disease in the United States: a systematic review. Int J Technol Assess Health Care 2023;39(1):e25. [doi: 10.1017/S0266462323000156] [Medline: 37114456]
5. Burckhardt MA, Roberts A, Smith GJ, Abraham MB, Davis EA, Jones TW. The use of continuous glucose monitoring with remote monitoring improves psychosocial measures in parents of children with type 1 diabetes: a randomized crossover trial. Diabetes Care 2018;41(12):2641-2643. [doi: 10.2337/dc18-0938] [Medline: 30377184]
6. Blood AJ, Cannon CP, Gordon WJ, Mailly C, MacLean T, Subramaniam S, et al. Results of a remotely delivered hypertension and lipid program in more than 10,000 patients across a diverse health care network. JAMA Cardiol 2023;8(1):12-21 [FREE Full text] [doi: 10.1001/jamacardio.2022.4018] [Medline: 36350612]
7. Meeker D, Friedberg MW, Knight TK, Doctor JN, Zein D, Cayasso-McIntosh N, et al. Effect of peer benchmarking on specialist electronic consult performance in a Los Angeles safety-net: a cluster randomized trial. J Gen Intern Med 2022;37(6):1400-1407. [doi: 10.1007/s11606-021-07002-1] [Medline: 34505234]
8. Smith M, Sarrazin MV, Wang X, Nordby P, Yu M, DeLonay AJ, et al. Risk from delayed or missed care and non-COVID-19 outcomes for older patients with chronic conditions during the pandemic. J Am Geriatr Soc 2022;70(5):1314-1324 [FREE Full text] [doi: 10.1111/jgs.17722] [Medline: 35211958]
9. Hailu R, Sousa J, Tang M, Mehrotra A, Uscher-Pines L. Challenges and facilitators in implementing remote patient monitoring programs in primary care. J Gen Intern Med 2024;39(13):2471-2477. [doi: 10.1007/s11606-023-08557-x] [Medline: 38653884]
10. Pronovost PJ, Cole MD, Hughes RM. Remote patient monitoring during COVID-19: an unexpected patient safety benefit. JAMA 2022;327(12):1125-1126. [doi: 10.1001/jama.2022.2040] [Medline: 35212725]
11. Drerup B, Espenschied J, Wiedemer J, Hamilton L. Reduced no-show rates and sustained patient satisfaction of telehealth during the COVID-19 pandemic. Telemed e-Health 2021;27(12):1409-1415. [doi: 10.1089/tmj.2021.0002] [Medline: 33661708]
12. Fadlon I, Agnihotri P, Longhurst C, Tai-Seale M. Equity and efficiency in technology adoptionvidence from digital health. NBER Working Pap 2023. [doi: 10.3386/w32992]
13. Dobbins M, Hanna SE, Ciliska D, Manske S, Cameron R, Mercer SL, et al. A randomized controlled trial evaluating the impact of knowledge translation and exchange strategies. Implement Sci 2009;4:61 [FREE Full text] [doi: 10.1186/1748-5908-4-61] [Medline: 19775439]

XSL•FO
RenderX

14. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
15. Mamataz T, Lee DS, Turk-Adawi K, Hajaj A, Code J, Grace SL. Factors affecting healthcare provider referral to heart function clinics: a mixed-methods study. J Cardiovasc Nurs 2024;39(1):18-30. [doi: 10.1097/JCN.0000000000001029] [Medline: 37669639]
16. Keating NL, O'Malley AJ, Onnela JP, Gray SW, Landon BE. Association of physician peer influence with subsequent physician adoption and use of bevacizumab. JAMA Netw Open 2020;3(1):e1918586 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18586] [Medline: 31899533]
17. Borus J, Pitts S, Gooding H. Acceptability of peer clinical observation by faculty members. Clin Teach 2018;15(4):309-313. [doi: 10.1111/tct.12681] [Medline: 28612509]
18. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
19. Donohue JM, Guclu H, Gellad WF, Chang CCH, Huskamp HA, Choudhry NK, et al. Influence of peer networks on physician adoption of new drugs. PLoS One 2018;13(10):e0204826 [FREE Full text] [doi: 10.1371/journal.pone.0204826] [Medline: 30273368]
20. Zheng K, Padman R, Krackhardt D, Johnson MP, Diamond HS. Social networks and physician adoption of electronic health records: insights from an empirical study. J Am Med Inform Assoc 2010;17(3):328-336 [FREE Full text] [doi: 10.1136/jamia.2009.000877] [Medline: 20442152]
21. Zachrison KS, Yan Z, Samuels-Kalow ME, Licurse A, Zuccotti G, Schwamm LH. Association of physician characteristics with early adoption of virtual health care. JAMA Netw Open 2021;4(12):e2141625 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.41625] [Medline: 34967876]
22. Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: a tool for assessing physician practice variation. Am J Med Qual 2005;20(3):151-157. [doi: 10.1177/1062860605274520] [Medline: 15951521]
23. Strauss A, Corbin J. Grounded Theory in Practice. Thousand Oaks, California: Sage Publications, Inc; 1999:288.
24. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
25. Fisher NDL, Fera LE, Dunning JR, Desai S, Matta L, Liquori V, et al. Development of an entirely remote, non-physician led hypertension management program. Clin Cardiol 2019;42(2):285-291 [FREE Full text] [doi: 10.1002/clc.23141] [Medline: 30582181]

Abbreviations

BP: blood pressure
CC: Community Care
CHERRIES: Checklist for Reporting the Result of Internet E-Surveys
EHR: electronic health record
FM: Family Medicine
GIM: General Internal Medicine
MOC: maintenance of certification
PCP: primary care physician
PHSO: Population Health Services Organization
QI: quality improvement
RBPM: remote blood pressure monitoring
UCSD: University of California San Diego

Edited by A Mavragani; submitted 07.08.24; peer-reviewed by H Gandhi, B Green; comments to author 07.10.24; revised version received 29.10.24; accepted 24.11.24; published 24.03.25.

Please cite as:
Wu JJ, Graham R, Çelebi J, Fraser K, Gin GT, Dang L, Hatamy E, Walker A, Barbato C, Lunde O, Coles L, Agnihotri P, Morn C, Tai-Seale M
Factors Influencing Primary Care Physicians' Intent to Refer Patients With Hypertension to a Digital Remote Blood Pressure Monitoring Program: Mixed Methods Study
J Med Internet Res 2025;27:e64933
URL: https://www.jmir.org/2025/1/e64933
doi:10.2196/64933
PMID:40126550

XSL•FO
RenderX

©Jennifer J Wu, Ross Graham, Julie Çelebi, Kevin Fraser, Geneen T Gin, Laurel Dang, Esmatullah Hatamy, Amanda Walker, Courtney Barbato, Ottar Lunde, Lisa Coles, Parag Agnihotri, Cassandra Morn, Ming Tai-Seale. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Detection of Clinically Significant Drug-Drug Interactions in Fatal Torsades de Pointes: Disproportionality Analysis of the Food and Drug Administration Adverse Event Reporting System

Huanhuan Ji[1], MS; Meiling Gong[1,2], MS; Li Gong[3], MS; Ni Zhang[4], MS; Ruiou Zhou[1], BS; Dongmei Deng[1], MS; Ya Yang[1], MS; Lin Song[1*], PhD; Yuntao Jia[1*], MS

[1]Department of Pharmacy, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, China International Science and Technology Cooperation base of Child development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China

[2]School of Pharmacy, Chongqing Medical University, Chongqing, China

[3]Department of Phase I Clinical Trial Ward, Chongqing Key Laboratory of Translational Research for Cancer Metastasis and Individualized Treatment, Chongqing University Cancer Hospital, Chongqing, China

[4]Department of Pharmacy, The Daping Hospital of Army Medical University, Chongqing, China

[*]these authors contributed equally

Corresponding Author:
Yuntao Jia, MS
Department of Pharmacy
National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, China International Science and Technology Cooperation base of Child development and Critical Disorders
Children's Hospital of Chongqing Medical University
No. 136, Zhongshan 2nd Road
Yuzhong District
Chongqing, 400014
China
Phone: 86 13883533546
Email: jiayuntaomail@hospital.cqmu.edu.cn

Abstract

Background: Torsades de pointes (TdP) is a rare yet potentially fatal cardiac arrhythmia that is often drug-induced. Drug-drug interactions (DDIs) are a major risk factor for TdP development, but the specific drug combinations that increase this risk have not been extensively studied.

Objective: This study aims to identify clinically significant, high-priority DDIs to provide a foundation to minimize the risk of TdP and effectively manage DDI risks in the future.

Methods: We used the following 4 frequency statistical models to detect DDI signals using the Food and Drug Administration Adverse Event Reporting System (FAERS) database: Ω shrinkage measure, combination risk ratio, chi-square statistic, and additive model. The adverse event of interest was TdP, and the drugs targeted were all registered and classified as "suspect," "interacting," or "concomitant drugs" in FAERS. The DDI signals were identified and evaluated using the Lexicomp and Drugs.com databases, supplemented with real-world data from the literature.

Results: As of September 2023, this study included 4313 TdP cases, with 721 drugs and 4230 drug combinations that were reported for at least 3 cases. The Ω shrinkage measure model demonstrated the most conservative signal detection, whereas the chi-square statistic model exhibited the closest similarity in signal detection tendency to the Ω shrinkage measure model. The κ value was 0.972 (95% CI 0.942-1.002), and the $P_{positive}$ and $P_{negative}$ values were 0.987 and 0.985, respectively. We detected 2158 combinations using the 4 frequency statistical models, of which 241 combinations were indexed by Drugs.com or Lexicomp and 105 were indexed by both. The most commonly interacting drugs were amiodarone, citalopram, quetiapine, ondansetron, ciprofloxacin, methadone, escitalopram, sotalol, and voriconazole. The most common combinations were citalopram and quetiapine, amiodarone and ciprofloxacin, amiodarone and escitalopram, amiodarone and fluoxetine, ciprofloxacin and sotalol, and amiodarone and citalopram. Although 38 DDIs were indexed by Drugs.com and Lexicomp, they were not detected by any of the 4 models.

XSL•FO
RenderX

Conclusions: Clinical evidence on DDIs is limited, and not all combinations of heart rate–corrected QT interval (QTc)–prolonging drugs result in TdP, even when involving high-risk drugs or those with known risk of TdP. This study provides a comprehensive real-world overview of drug-induced TdP, delimiting both clinically significant DDIs and negative DDIs, providing valuable insights into the safety profiles of various drugs, and informing the optimization of clinical practice.

(J Med Internet Res 2025;27:e65872) doi:10.2196/65872

KEYWORDS

torsades de pointes; FAERS database; drug-drug interactions; QTc-prolonging drugs; adverse drug events

Introduction

Torsades de pointes (TdP) is a rare but life-threatening cardiac arrhythmia characterized by a prolonged heart rate–corrected QT interval (QTc) on an electrocardiogram. TdP is often induced by QTc-prolonging drugs, which can further increase the risk when used in combination. According to the Arizona Center for Education and Research on Therapeutics, 268 drugs are known to be QTc-prolonging drugs, including antiarrhythmics, antibiotics, antipsychotics, antidepressants, and antineoplastic drugs. These drugs are classified into 3 categories: 68 drugs with a known risk of TdP, 148 drugs with a possible risk of TdP, and 52 drugs with a conditional risk of TdP [1]. Chien et al [2] conducted a large-scale, case-crossover study demonstrating that QT-prolonging drug use and related drug-drug interactions (DDIs), particularly “known risk” or “conditional risk” drugs, are significantly associated with QTc prolongation. This risk is further exacerbated by the concurrent administration of these drugs with other QT-prolonging agents or potent cytochrome P450 inhibitors. The resulting exponential increase in potential DDIs is considerable and warrants careful consideration when these medications are used concurrently. In clinical settings, DDI alerts are implemented to warn of potential QTc DDI risks. However, their unclear clinical relevance often leads to indiscriminate alerts, resulting in over 90% being ignored, limiting the effectiveness of QTc DDI alerts [3,4]. Wasserman et al [4] found an average of 3 unique QTc DDI alerts presented per patient admission (range 1-44), with an average of 6 overridden QTc DDI alerts (range 1-432), ultimately resulting in 29% of admissions experiencing a drug-induced QTc prolongation with a ≥60-point increase in QTc value from baseline or a post-QTc value of ≥500 ms, representing a significant increase in TdP risk. Although DDIs are recognized as a significant risk factor for TdP development, the specific drug combinations contributing to this risk remain poorly characterized.

In clinical practice, the assessment of DDI risks often relies on multiple information sources, such as Micromedex, UpToDate, Drugs.com, and Medscape.com. However, significant variability exists among these resources, as demonstrated by Shariff et al [5], who reported poor intersource reliability scores ($\kappa<0.20$, $P<.05$) [5]. This inconsistency complicates decision-making for health care professionals and underscores the importance of using real-world data to detect clinically significant DDIs. The challenge is particularly acute with older adults, patients with cancer, and critically ill patients with multiple comorbidities, who often receive complex drug regimens that extend beyond the 2-drug combinations typically considered in DDI reports [6-8].

Literature data serve as valuable real-world evidence, with numerous cases of drug-induced TdP documented. However, there is a notable paucity of evidence concerning DDIs in these reports. Krumpholz et al [9] conducted a comprehensive review of literature-reported cases of drug-induced TdP from 1980 to 2021, encompassing 424 papers of 634 case reports, which primarily documented suspected and concomitant drugs without adequately considering interacting drugs. This highlights the need for studies focusing on identifying clinically significant DDIs.

The Food and Drug Administration Adverse Event Reporting System (FAERS) database is a spontaneous reporting system that provides information about adverse events (AEs) in clinical settings and has emerged as a valuable resource for pharmacovigilance research. Although the database has limitations, including potential inaccuracies, underreporting, and missing data, it provides unique insights into postmarketing drug safety. In the FAERS database, 2 or more suspected, interacting, or concomitant drugs for each event can be given, which is particularly useful for identifying DDIs that result in severe or fatal AEs. However, significant uncertainty exists regarding the association between drugs and AEs in the given report, as these reports encompass all potential drugs and AEs for a patient. Recent methodological advances have enhanced the utility of FAERS for DDI detection, particularly through the application of multiple statistical models that can improve signal detection reliability [10,11]. This study aimed to integrate multiple data sources, including the FAERS database, DDI information resources, and relevant literature, to detect clinically significant DDIs associated with TdP risk. By using advanced analytical methods and cross-validating findings across multiple sources, we sought to provide more reliable and actionable insights for clinical decision-making.

Methods

Data Sources

A retrospective, disproportionality, pharmacovigilance study was conducted from the first quarter of 2004 to the second quarter of 2023 using the FAERS database to detect potential DDIs that increase the incidence of TdP in a large-scale population. In the FAERS database, each report is coded using the preferred terms (PTs) provided by the Medical Dictionary for Regulatory Activities (MedDRA), and each drug is assigned a code according to its association with AEs: suspect drugs, concomitant drugs, or interacting drugs. OpenVigil

XSL•FO
RenderX

2.1-MedDRA-v24 [12], an open tool for data mining and analysis of pharmacovigilance data using cleansed FDA AE reporting data, was used to detect DDIs.

Definition of AE

In this study, the targeted AE was TdP, which was extracted from FAERS using the PT of TdP (PT code: 10044066). In these cases, all drugs classified as suspected drugs, interacting drugs, and concomitant drugs were included in the process to investigate DDIs. For signal detection, we focused only on combinations of 2 drugs with 3 or more cases.

Suspect Drugs of TdP Cases

Single-drug TdP signals were assessed for disproportionality using the established pharmacovigilance index, the reporting odds ratio (ROR). ROR values were calculated as (a × d)/(b × c) and expressed as point estimates with a 95% CI. In our study, an event was considered significant when the ROR value was greater than 2 and the lower limit of the 95% CI of the ROR was greater than 1, and at least 3 cases were required to define a positive signal for the ROR.

DDI Signal Detection Statistical Models and Standards

Several statistical algorithms have been developed for exploring DDI signals, yet no de facto standard exists for DDI signal detection in spontaneous reporting systems [10,11]. Each method has inherent advantages and limitations. Among them, the Ω shrinkage measure model, proposed by Norén et al [13] and used by the World Health Organization Uppsala Monitoring Center, calculates an observed-to-expected ratio for disproportionality measurement of potential DDIs. This method demonstrates a conservative signal detection trend among frequentist-based approaches, effectively minimizing false positives by handling sparse data, adjusting for deviations in expected frequencies, and shrinking the extreme ratios toward the overall mean. It is particularly well-suited for monitoring rare AEs (eg, TdP) in noisy databases like FAERS but requires clinical validation to confirm associations between drug combinations and adverse reactions. The additive model, as described by Thakrar et al [14], estimates co-medication risks by evaluating target AE incidences across different exposure scenarios. Although highly sensitive for detecting potential drug synergies, it is prone to spurious associations due to noise and biases in spontaneous reporting systems. Similarly, the chi-square statistic model with Yates correction, proposed by Gosho et al [15], is widely used but vulnerable to false negatives or false positives when applied to sparse data, limiting its utility for rare events like TdP. The combination risk ratio model, proposed by Noguchi et al [16], offers a theoretical framework for assessing the joint risk of co-medicated drugs by assuming that the occurrence of AEs represents a combined risk of both drugs. This approach is valuable for exploring potential interactions but relies heavily on sufficient co-medication frequency, limiting its effectiveness with rare drug combinations.

To address these limitations, a multimethod cross-validation approach was used, integrating the Ω shrinkage measure model, additive model, chi-square statistic model, and the combination risk ratio model (calculation methods are detailed in Multimedia Appendix 1). This ensures that signals detected by multiple methods are less likely to be statistical artifacts. Ω shrinkage was prioritized as the baseline model due to its robustness at handling sparse data and its ability to minimize false positives. Additive models were further screened using Ω shrinkage to reduce spurious associations. For signal confirmation, all 4 statistical models were applied, with a positive signal defined as one consistently detected by all 4 methods. We constructed 2 contingency tables—a 4-by-2 table and a 2-by-2 table—to facilitate the analysis (detailed in Multimedia Appendix 1). Additionally, the specific criteria for signal detection applied to each model are detailed in Table 1.

Table 1. The signal detection criteria for each statistical model.

Statistical models	Criteria for positive signals
Ω shrinkage measure model	$\Omega_{025} > 0$, $n_{111} \geq 3$
Chi-square statistic model	$\chi > 2$, $n_{111} \geq 3$
Combination risk ratio model	$PRR^{a}_{drug\ D1 \cap drug\ D2} > 2$, $\chi^2_{drug\ D1 \cap drug\ D2} > 4$, $CRR^{b} > 2$, $n_{111} \geq 3$
Additive model	$P_{11} - P_{10} - P_{01} + P_{00} > 0$, $n_{111} \geq 3$

[a]PRR: proportional reporting ratio.

[b]CRR: combination risk ratio.

To ensure robustness and minimize false positives, positive DDI signals underwent external validation by cross-referencing with established DDI databases and a comprehensive literature review of relevant case reports to verify their clinical significance.

Evaluation of Commonality of Signals Detected

The commonality of the signals detected by each statistical model was evaluated using the Cohen kappa coefficient (κ), the proportionate agreement for positive rating ($P_{positive}$), and the proportionate agreement for negative rating ($P_{negative}$), as reported in a previous study (calculation methods are detailed in Multimedia Appendix 1). The κ provides a measure that adjusts the observed agreement for the chance agreement, which ranges from –1 to 1, and indicates the level of agreement between observed and expected values [16]. A κ value of 0 to 0.20 indicates slight agreement, 0.21 to 0.4 indicates fair agreement, 0.41 to 0.60 indicates moderate agreement, 0.61 to 0.80 indicates substantial agreement, 0.81 indicates upward excellent agreement, and the absolute agreement is 1 [17]. Conversely, κ<0 means that the observed agreement rate is less

XSL•FO
RenderX

than the chance agreement rate, which rarely occurs in practical research.

Signal Verification

As positive controls, we downloaded a list of 268 drugs with known (n=68), possible (n=148), and conditional (n=52) risks of TdP from CredibleMed, an online compendium of QTc drugs [1], and 217 drugs with the highest (n=32), moderate (n=70), and low (n=115) risks of interacting from UpToDate [18].

To ensure the accuracy and reliability of DDI results, the Lexicomp point-of-care database, known for its comprehensive scope, was used to validate DDI signals. Additionally, the open-access online drug interaction checker Drugs.com was used to cross-validate the DDI signals. The severity classifications of DDIs in the 2 drug databases are detailed in Multimedia Appendix 1 [18,19]. Subsequently, a search was conducted in the PubMed database for literature pertaining to drug-induced TdP, and an open-access database of literature-derived drug-related TdP cases, containing 624 TdP cases from 424 papers, was used to further evaluate the clinical relevance of the DDIs [9].

Sensitivity Analyses

Given that nonprofessional reporters may not have the expertise to accurately distinguish between AEs and the symptoms of the disease itself, this could potentially affect the interpretability of aggregated results, particularly in large-scale data sets like FAERS. To improve the clarity and reliability of the findings, we excluded data reported by nonprofessionals from the validated positive signal data, and conducted sensitivity analysis to assess robustness.

Ethical Considerations

This study uses the publicly accessible and anonymized FAERS database, which contains no personally identifiable information and does not involve human subject experimentation. As a result, it is exempt from review and approval by an institutional review board.

Results

Basic Information on TdP Cases

As of September 2023, a total of 4313 TdP cases from 11,439,756 AE cases were included in this study. Of the 4313 TdP cases, 2403 (55.7%) were female, and 1398 (32.4%) were male. The average age was 55.5 (range 0-97) years, and 1653 (38.3%) were older than 60 years. TdP is serious and fatal, with many of the reported cases leading to hospitalization (initial or prolonged; n=1310, 30.4%), life-threatening events (n=1705, 39.5%), or death (n=363, 8.4%). Notably, reports submitted by nonprofessionals, such as consumers and lawyers, accounted for only 290 (6.7%) of the total. This proportion is significantly lower than the 20% to 30% typically reported, likely due to the rarity and severity of TdP, which often requires hospitalization, thereby limiting the likelihood of reports from consumers. Geographically, 1784 (41.4%) of these 4313 cases originated from the United States (Table 2).

XSL•FO
RenderX

Table 2. Characteristics of patients with torsades de pointes in the Food and Drug Administration Adverse Event Reporting System database (n=4313).

Characteristic	Patients, n (%)
Gender	
Male	1398 (32.4)
Female	2403 (55.7)
Unknown	512 (11.9)
Age (years)	
0-19	180 (4.2)
20-39	625 (14.5)
40-59	1011 (23.4)
60-79	1196 (27.7)
≥80	457 (10.6)
Unknown	844 (19.6)
Severity	
Died	363 (8.4)
Life threatening	1705 (39.5)
Disabled	6 (0.1)
Hospitalized: initial or prolonged	1310 (30.4)
Required intervention	19 (0.4)
Other outcomes	882 (20.4)
Unknown	28 (0.6)
Reporter	
Other health professionals[a]	1638 (38)
Medical doctor	1585 (36.7)
Pharmacist	467 (10.8)
Not obtained	333 (7.7)
Consumer	218 (5.1)
Lawyer	72 (1.7)
Reporter country	
United States	1784 (41.4)
Japan	201 (4.7)
China	80 (1.9)
Other countries	1923 (44.6)
Unknown	325 (7.5)

[a]Other health professionals included 1 registered nurse and 396 health professionals.

Figure 1 depicts a cocurrent network of AEs associated with TdP, providing a comprehensive overview of the patterns of TdP occurrence. The size of a node is proportional to the number of neighboring AEs, while the width of an edge is proportional to the number of unique AEs. Among the 4313 TdP cases analyzed, the most common presentations were electrocardiogram QT prolonged (n=1736), long QT syndrome (n=298), electrocardiogram QT-corrected interval prolonged (n=61), cardiac arrest (n=892), ventricular tachycardia (n=738), ventricular fibrillation (n=622), syncope (n=492), ventricular extrasystoles (n=224), loss of consciousness (n=214), cardiorespiratory arrest (n=205), atrial fibrillation (n=144), arrhythmia (n=140), ventricular arrhythmia (n=135), tachycardia (n=115), palpitations (n=114), and dizziness (n=109). The risk factors of hypokalemia, hypomagnesemia, bradycardia, electrolyte imbalance, decreased blood magnesium, and hypocalcemia were observed in 524, 235, 225, 65, 63, and 50 cases, respectively. The drug-related risk factors of drug interactions, overdose, drug abuse, intentional overdose, and toxicity to various agents were present for 513, 271, 168, 164, and 162 cases, respectively.

XSL•FO
RenderX

Figure 1. Cocurrent network diagram of adverse events related to torsades de pointes.

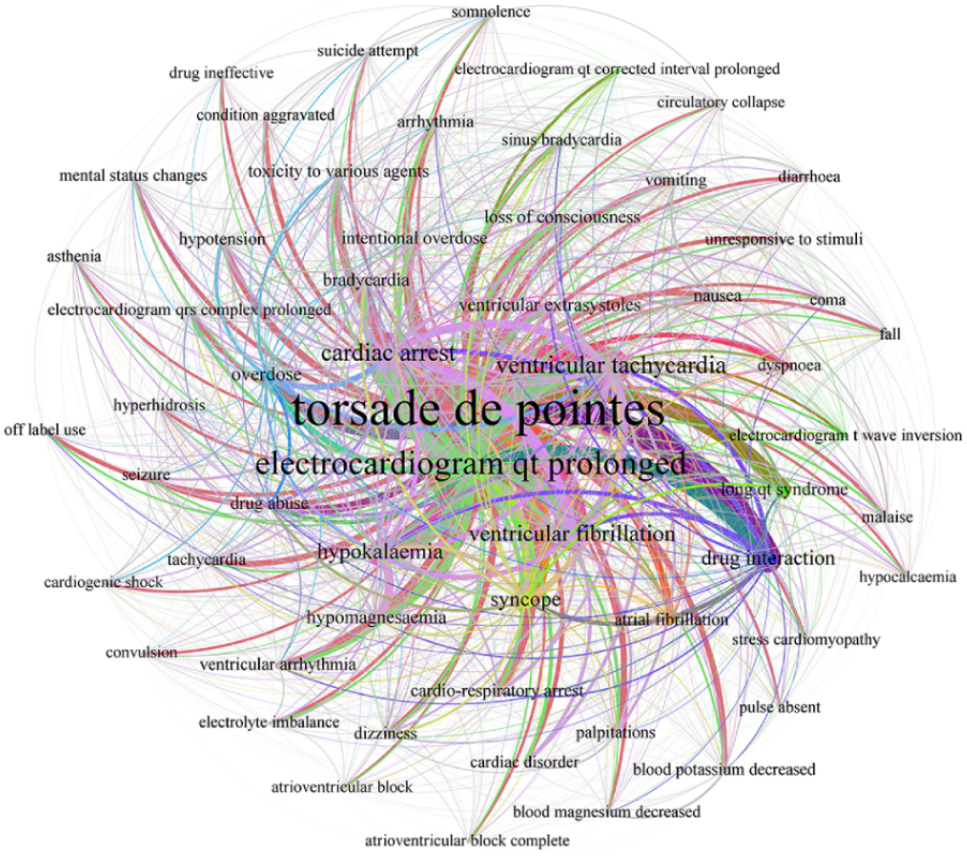

Figure 2. Distribution of torsades de pointes–related drug classes.

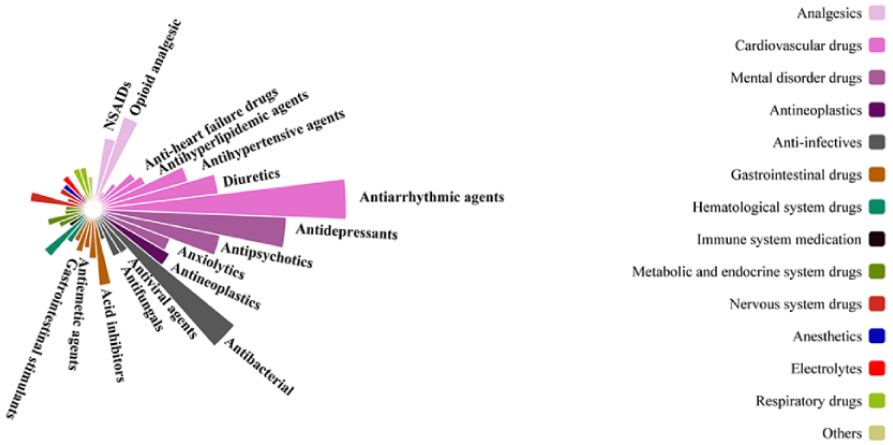

Suspect Drugs of TdP Cases

Of the 4313 analyzed TdP cases, 721 drugs, including 136 QT-prolonging drugs, were implicated in 3 or more cases. The most reported drugs were antiarrhythmics, antidepressants, antibacterials, antipsychotics, analgesics, antineoplastics, and others (Figure 2). Amiodarone had the most reports (n=524), followed by furosemide (n=448), methadone (n=312), loperamide (n=301), and citalopram (n=299). However, loperamide (n=260; ROR 53.19, 95% CI 46.87-60.36) was the most frequently reported primary suspect drug, followed by amiodarone (n=233; ROR 58.48, 95% CI 51.18-66.82), citalopram (n=143; ROR 23.72, 95% CI 20.06-28.04), methadone (n=159; ROR 53.91, 95% CI 45.95-63.26), furosemide (n=141; ROR 30.93, 95% CI 26.13-36.62), and sotalol (n=117; ROR 255.06, 95% CI 210.53-309.02). Ibutilide (n=5; ROR 1206.58, 95% CI 419.04-3474.16) had the strongest signal for TdP, followed by halofantrine (n=3; ROR 884.41, 95% CI 239.35-3267.99), cisapride (n=69; ROR 325.59, 95% CI 253.16-418.74), procainamide (n=5; ROR 270.86, 95% CI 107.87-680.12), and sotalol (n=117; ROR 255.06, 95% CI 210.53-309.02), and these drugs had a known TdP risk (Figure 3).

XSL•FO
RenderX

Figure 3. Scatterplot of the number of reports and reporting odds ratio (ROR) for torsades de pointes–related primary suspect drugs. CR: conditional risk; KR: known risk; PR: possible risk.

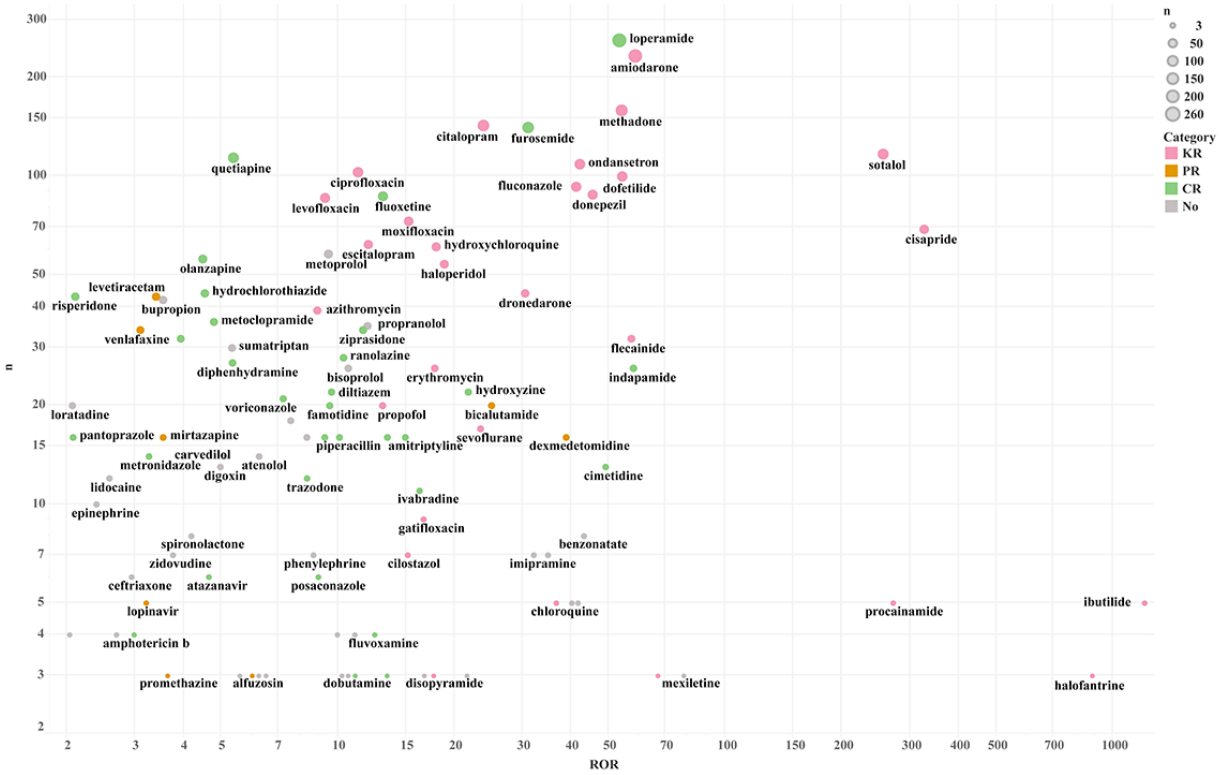

DDI Signals of Statistical Models

We evaluated the reports of TdP for concomitant use of 2 of the 721 drugs, totaling 8264 drug combinations, of which 4230 combinations were reported for at least 3 cases. The additive model detected the most signals with 3296 (77.9%) pairs, followed by the combination risk ratio model with 2574 (60.8%) pairs, the chi-square statistic model with 2243 (53%) pairs, and the Ω shrinkage measure model with 2236 (52.9%) pairs.

Commonality of DDI Signals Detected

Table 3 shows the κ coefficients and proportionate agreements for $P_{positive}$ and $P_{negative}$ among the 4 frequency statistical models. The chi-square statistics showed the greatest similarity to the Ω shrinkage measure, with κ, $P_{positive}$, and $P_{negative}$ values of 0.972 (95% CI 0.942-1.002), 0.987, and 0.985, respectively. In contrast, the κ, $P_{positive}$, and $P_{negative}$ values for the Ω shrinkage measure and additive models were 0.482 (95% CI 0.451-0.513), 0.808, and 0.638, respectively, with moderate similarity.

Table 3. The Cohen kappa coefficient (κ) and proportionate agreements for positive rating (Ppositive) and negative rating (Pnegative) among 4 frequency statistical models

Models	$P_{positive}$	$P_{negative}$	κ (95% CI)
Ω shrinkage measure model vs chi-square statistic model	0.987	0.985	0.972 (0.942-1.002)
Ω shrinkage measure model vs combination risk ratio model	0.904	0.874	0.780 (0.749-0.810)
Ω shrinkage measure model vs additive model	0.808	0.638	0.482 (0.451-0.513)
Chi-square statistic model vs combination risk ratio model	0.910	0.881	0.793 (0.762-0.823)
Chi-square statistic model vs additive model	0.810	0.640	0.485 (0.454-0.516)
Combination risk ratio model vs additive model	0.877	0.721	0.612 (0.578-0.646)

DDI Signal Verification

All 4 statistical models detected 2158 combinations (96.2% of the Ω shrinkage measure model), with 141 combinations indexed by Lexicomp, 205 combinations indexed by Drugs.com, 241 combinations indexed by either, and 1918 combinations indexed by neither as of October 8, 2023. In addition, 105 combinations were indexed by both, with 52 combinations being present in 10 or more cases (Figures 4 and 5).

XSL•FO
RenderX

Figure 4. Process of drug-drug interaction signal monitoring and verification. TdP: torsades de pointes.

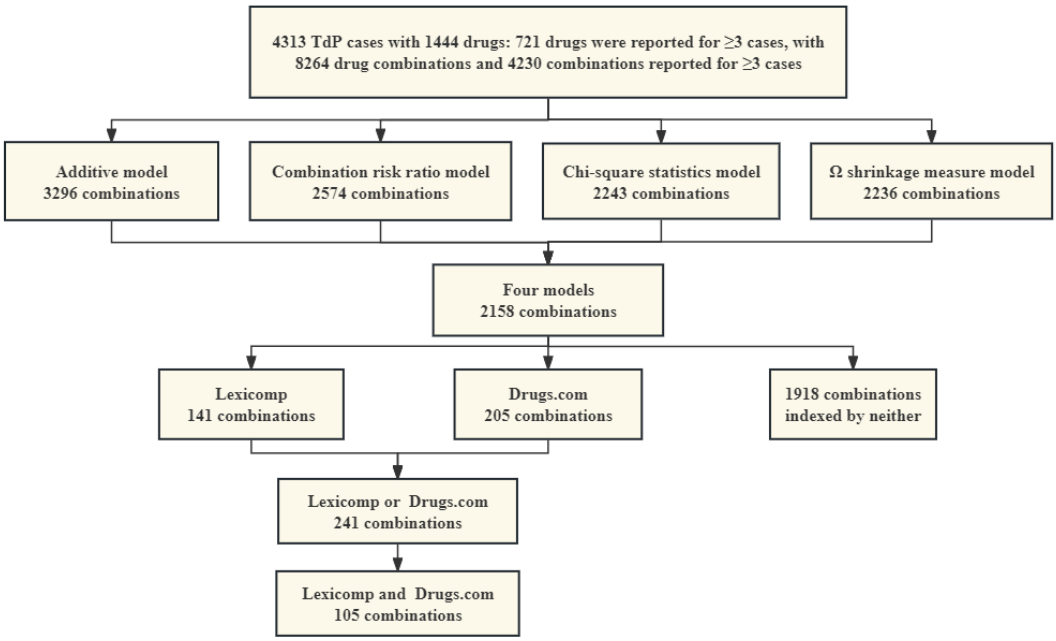

Figure 5. Number of drug-drug interaction signal intersections.

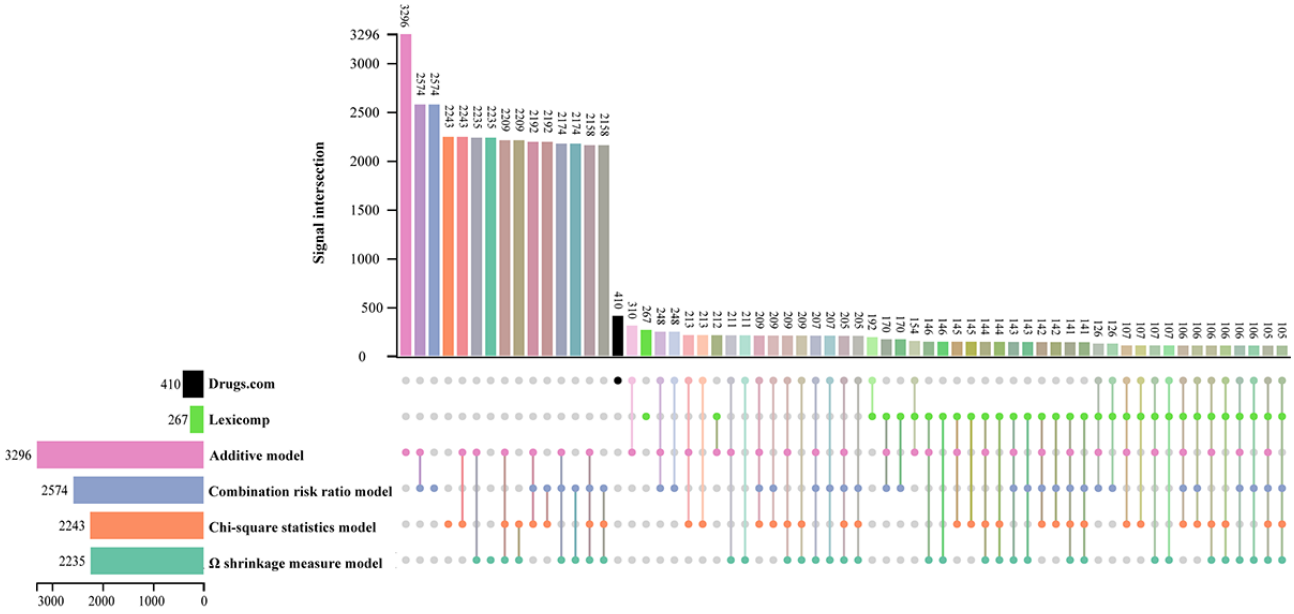

DDI Assessment

Of 2158 combinations, 409 individual drugs were involved. Categorization in CredibleMeds showed 35 “known risk,” 31 “possible risk,” and 45 “conditional risk” drugs. In Lexicomp, 12 were “high risk,” 29 were “moderate risk,” and 29 were “low risk.” The drug with the highest number of interactions was amiodarone, which interacted with 68 other drugs, followed by methadone (n=51), fluoxetine (n=46), bisoprolol (n=42), ciprofloxacin (n=41), citalopram (n=37), and lorazepam (n=37).

However, the drugs with the most reported interaction cases were amiodarone, citalopram, quetiapine, ondansetron, ciprofloxacin, methadone, escitalopram, sotalol, and voriconazole. The drug combinations are shown in Figures 6 and 7, with citalopram and quetiapine (n_{111}=86) having the highest number of cases, followed by amiodarone and ciprofloxacin (n_{111}=49), amiodarone and escitalopram (n_{111}=34), amiodarone and fluoxetine (n_{111}=33), ciprofloxacin and sotalol (n_{111}=31), and amiodarone and citalopram (n_{111}=30). The size of the nodes is proportional to the number of specific drugs with TdP, while the width of the extension branch is proportional to the number of specific combinations with TdP.

XSL•FO
RenderX

Figure 6. Drug-drug interactions (n=105) for torsades de pointes (detected by 4 models and indexed by Drugs.com and Lexicomp，n111≥3).

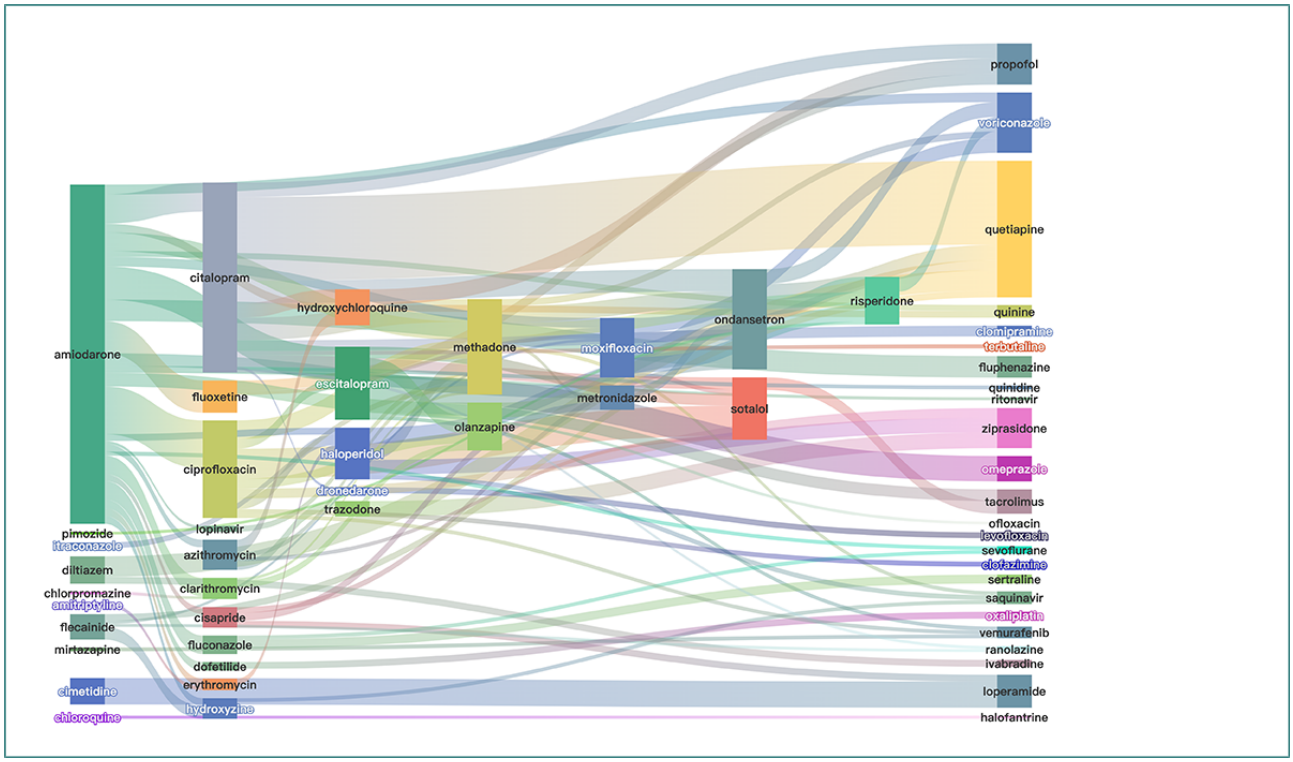

Figure 7. Drug-drug interactions (n=52) for torsades de pointes (detected by 4 models and indexed by Drugs.com and Lexicomp, n111≥10).

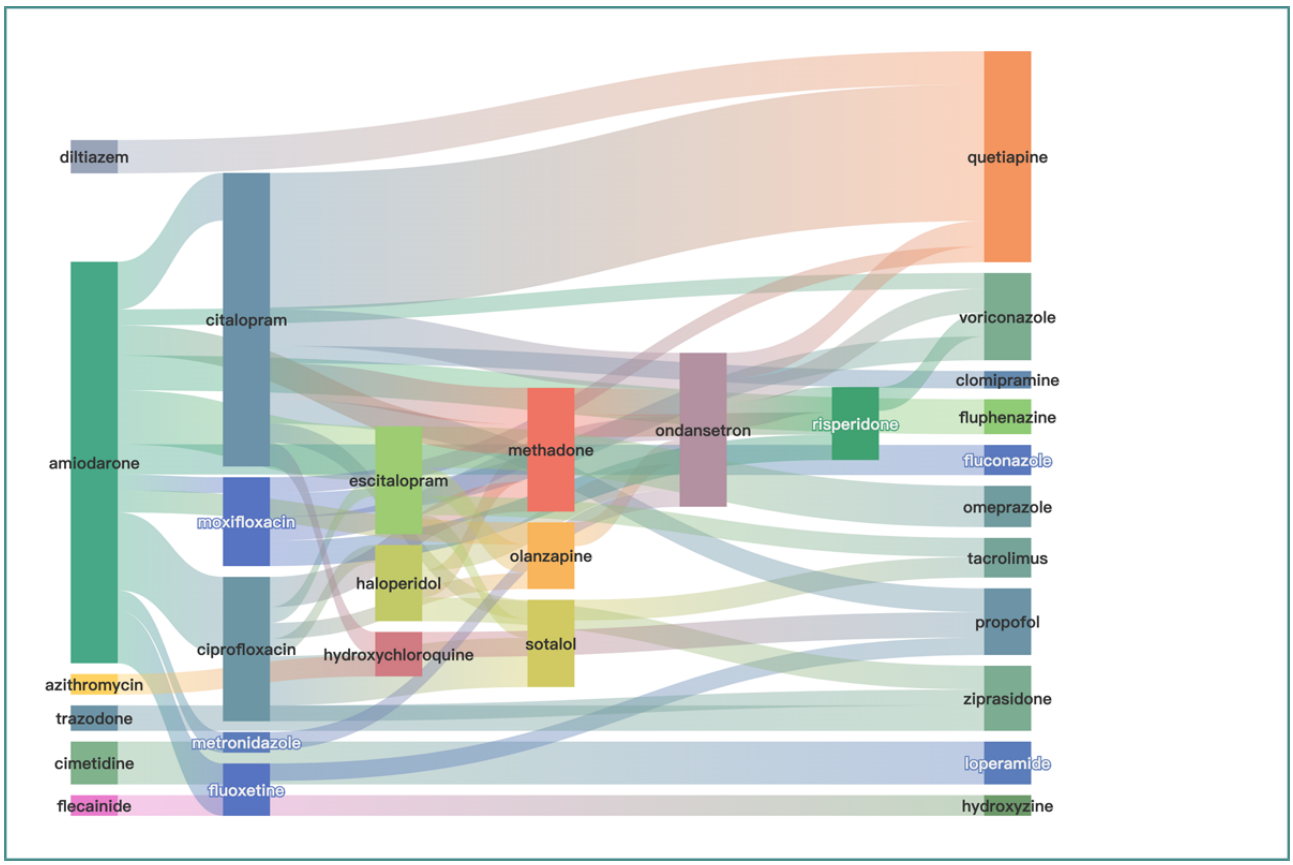

Multimedia Appendix 2 presents the characteristics of 105 drug combinations. Lexicomp classified 20 combinations as “X,” 25 as “D,” 42 as “C,” and 18 as “B.” In contrast, Drugs.com categorized 27 combinations as “Moderate risk” and 78 as “Major risk.” Additionally, 25 combinations were recorded in the literature-derived database of drug-related TdPs with enhanced clinical relevance.

Regrettably, 38 drug combinations indexed in both databases were not detected by any of the 4 models. These included 2 “X”

XSL•FO
RenderX

class (methadone and quetiapine, quetiapine and ziprasidone), 11 “D” class, 14 “C” class, and 9 “B” class drug combinations, with amiodarone and sotalol, chlorpromazine and haloperidol, and amiodarone and dronedarone being the high-risk QT-prolonging drug combinations (Multimedia Appendix 2; Figure 8).

Figure 8. Drug-drug interactions (n=38) for torsades de pointes (indexed by Drugs.com and Lexicomp, n111≥3, but not detected by any of the 4 models).

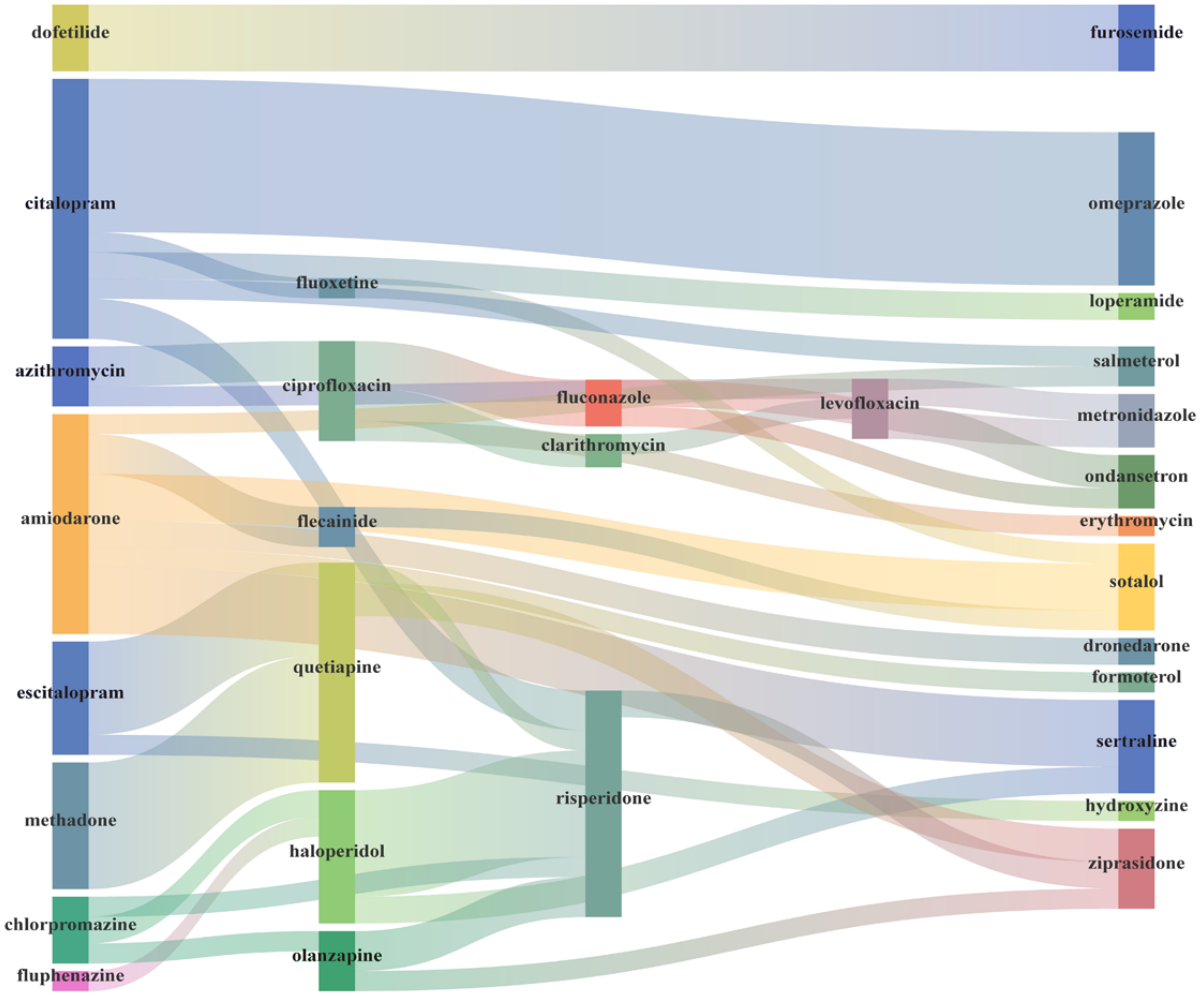

Among the 4313 cases analyzed, 513 involved drug interaction events, with 984 interacting drugs reported for 460 cases. The most commonly interacting drugs were amiodarone, citalopram, fluoxetine, furosemide, escitalopram, ondansetron, omeprazole, moxifloxacin, and metoprolol. By matching the results of signal detection and validation of drug combinations, a total of 189 drug combinations were identified by the 4 models. Of these, 65 combinations were indexed by Lexicomp, 74 were indexed by Drugs.com, 84 were indexed by either, and 55 were indexed by both. Figure 9 illustrates the 55 DDIs classified as “interacting drugs” in FAERS that were detected by the 4 models and indexed by Drugs.com and Lexicomp. These include combinations like amiodarone and fluoxetine (n=25), loperamide and cimetidine (n=23), escitalopram and omeprazole (n=10), and moxifloxacin and ondansetron (n=10). Some combinations, although reported more frequently, were excluded from the analysis as they were either not detected by the model or not indexed in the database. For example, venlafaxine and quetiapine (n=17) and furosemide and metoprolol (n=17) were only detected by the additive model. Additionally, combinations like fluoxetine and fluphenazine (n=16) and fluconazole and ritonavir (n=13) were not indexed either Drugs.com or Lexicomp.

XSL•FO
RenderX

Figure 9. Drug-drug interactions (n=55) for torsades de pointes (drugs classified as "interacting drugs" in the Food and Drug Administration Adverse Event Reporting System, detected by the 4 models, and indexed by Drugs.com and Lexicomp).

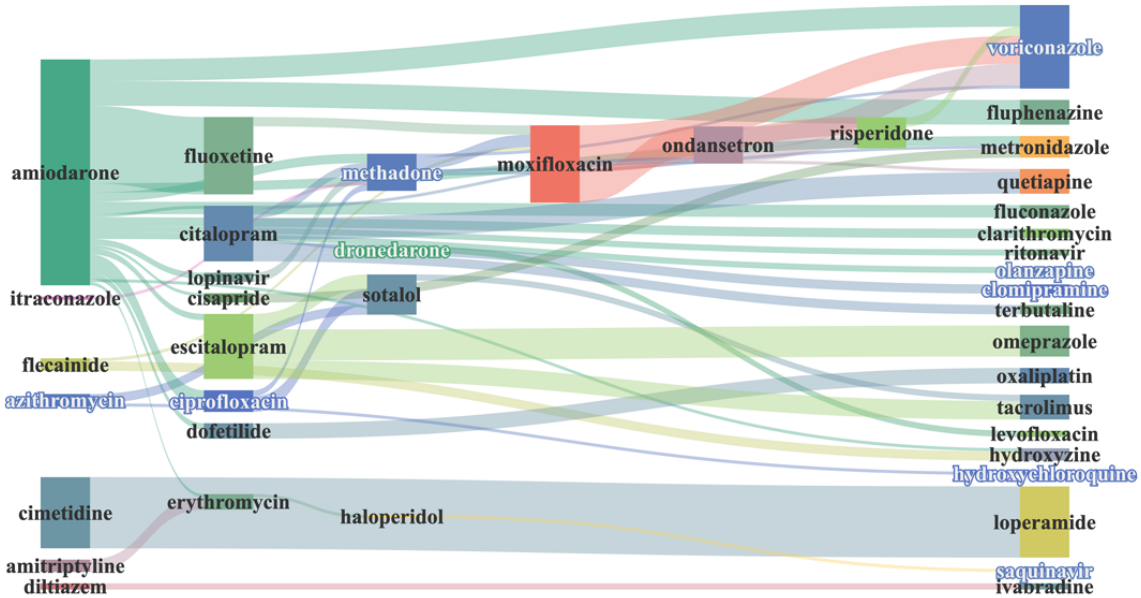

Sensitivity Analyses

Despite the relatively low proportion of reports from nonprofessionals, we conducted an analysis of TdP cases reported by nonprofessionals, such as consumers and lawyers. We evaluated TdP by nonprofessionals for concomitant use of 2 of the drugs, totaling 1862 drug combinations, with 89 combinations reported for at least 3 cases. Among these, 49 combinations were detected by all 4 statistical models, and only 5 combinations were indexed by Lexicomp and Drugs.com when considering all data, including reports from nonprofessionals. Sensitivity analyses revealed that the inclusion of nonprofessional reports did not alter the final determination of the signal detection (Multimedia Appendix 2).

Discussion

Main Findings

This study presents the initial comparative safety evaluation of FAERS data aimed at assessing clinically significant DDIs triggering fatal TdP. Clinical evidence regarding DDIs remains limited, and not all combinations of QTc-prolonging drugs lead to TdP, even when high-risk drugs or those with established TdP risks are involved. This study offers a comprehensive real-world analysis of drug-induced TdP, identifying both clinically significant and nonsignificant DDIs. These findings contribute valuable insights into the safety profiles of various drugs, aiding in the optimization of clinical practice.

Matsuo and Yamaori [20] attempted to use the Japanese Adverse Drug Event Report database to identify and summarize potential DDIs associated with an increased risk of drug-induced long QT syndrome. However, their methodology relied on the ROR, which is typically used for assessing safety signals of individual drugs that may not be appropriate for detecting safety signals of DDIs, potentially leading to false positive signals. To address this limitation, 4 frequency statistical models—namely, the Ω shrinkage measure, combination risk ratio, chi-square statistic, and additive model—were used to detect DDI signals in our study, thereby greatly enhancing the validity and comprehensiveness of the identified signals. Furthermore, the verification of these signals was conducted using the Drugs.com, Lexicomp, and literature databases, thereby ensuring the credibility and reliability of the DDI data.

Consistent with previous studies, the Ω shrinkage measure model demonstrated the most conservative signal detection, while the chi-square statistic model exhibited the closest similarity in signal detection tendency to the Ω shrinkage measure model [16]. In contrast, the additive model detected considerably more signals [5] and had less similarity in signal detection tendency with the other models; therefore, care must be taken when interpreting the signals. For a specific positive signal result, the more models that detected it, the more reliable the result is. In this study, 2158 combinations were detected by the 4 frequency statistical models, of which 241 combinations were indexed by Drugs.com or Lexicomp and 105 were indexed by both. In addition, 171 combinations were indexed by only 1 of the databases, which was related to the poor consistency between the 2 databases [11] due to the various rating criteria and procedures and definitions of different levels of acceptable risk. Drugs.com validated more interactions than Lexicomp, which may be related to the fact that Drugs.com can recognize brand names. Despite using an open-access database of literature-derived, drug-related TdP cases containing 624 TdP cases from 424 papers to evaluate the clinical relevance of DDIs, only concomitant drugs were documented in all cases except for suspected drugs, with no recorded interacting drugs [9].

XSL•FO
RenderX

Unsurprisingly, limited clinical evidence on DDIs is available, owing to the inherent difficulty of judging when interactions lead to AEs. However, DDIs do exist, which highlights the value of detecting clinically significant DDIs based on real-world data.

In our study, a list containing all potential DDIs related to TdP from FAERS was provided. To the best of our knowledge, this is the most comprehensive list derived from a pharmacovigilance database so far. Consistent with the results of existing studies, antiarrhythmics, antidepressants, antipsychotics, antibacterials, analgesics, and antineoplastics drugs, such as amiodarone, citalopram, quetiapine, ondansetron, ciprofloxacin, methadone, escitalopram, sotalol, and voriconazole, were the most common interacting drugs causing TdP [20,21]. Additionally, diuretics were also common interacting drugs causing TdP, and the potential mechanism underlying this may be related to electrolyte disorder caused by diuretics. The most common combinations were citalopram and quetiapine, amiodarone and ciprofloxacin, amiodarone and escitalopram, amiodarone and fluoxetine, ciprofloxacin and sotalol, and amiodarone and citalopram, and previous studies have proven these DDIs are closely related to TdP [22-24]. Trazodone may lead to QTc prolongation and TdP, which is potentially fatal even without risk factors for QTc prolongation [25]. In addition, co-interaction with ziprasidone increases the risk of TdP. Therefore, we can use this report's findings to quickly understand the DDIs that commonly result in TdP in the real world. However, it is worth noting that a high reporting ratio does not always represent a high risk because, for different combinations, the frequency of drug use will vary greatly, which will directly affect the proportion of TdP reports.

The development of TdP in the context of DDIs is likely attributable to two key mechanisms: altered drug concentrations due to metabolic inhibition or induction and superimposed pharmacological effects on cardiac ion channels. For instance, co-administration of drugs that prolong the QT interval may synergistically increase the risk of TdP. Moreover, inhibition of the metabolism of a QT-prolonging drug by a co-administered agent could lead to elevated plasma concentrations, thereby exacerbating the risk. Amiodarone is the most widely interacting drug, with interactions with other 68 drugs, followed by methadone, fluoxetine, bisoprolol, ciprofloxacin, and citalopram [26-28], which interact with 51, 46, 42, 41, and 37 other drugs, respectively. Therefore, additional caution and increased electrocardiogram monitoring may be still warranted when these drugs are combined, especially for patients with risk factors such as increasing age, female sex, bradycardia, heart failure, history of ventricular arrhythmias, use of diuretics, hypothyroidism, hypokalemia, hypomagnesemia, and hypocalcemia.

In addition to offering a complete list of drugs, our study provides novel perspectives and opportunities for further exploration. As real-world negative reporting is missing, counterexamples that could serve as contradictory evidence may exist. Therefore, identifying negative DDI pairs shall enhance the confidence of health care professionals with a level of certainty on DDIs, which in turn will improve medical research and decision-making. It is widely accepted that co-administration of QTc-prolonging drugs, especially those with a known risk or high risk for TdP, heighten the risk of TdP. Our research identified 38 DDIs for QTc prolongation indexed by Drugs.com and Lexicomp but were not detected by any of the 4 models. This finding supports the notion that not all combinations of QTc-prolonging drugs result in TdP, even when involving high-risk drugs or those with a known risk of TdP. There is only a limited increase in QTc prolongation with concurrent use of QTc-prolonging drugs, and that magnitude of TdP risk depends on the specific drugs involved and patient risk factors. If clinical decision support systems generate alerts for QTc DDI whenever 2 QTc-prolonging drugs with a known risk of TdP are combined, irrespective of the clinical relevance of these interactions, a substantial number of false-positive alerts may be produced. The rising incidence of false-positive QTc DDI alerts can contribute to alert fatigue among physicians, potentially leading to the disregard of important warnings. Furthermore, these false-positive alerts may prompt unwarranted discontinuation or substitution of medications, thereby compromising the efficacy of clinical treatments. Consequently, identifying negative QTc DDIs is a rational approach to mitigating the challenge of alert fatigue. Nevertheless, caution must be exercised even when a signal is identified as negative, due to the inherent limitations of the FAERS database.

We acknowledge the inherent limitations of our study. First, our analysis primarily focused on the potential for TdP resulting from the concomitant use of 2 drugs. However, in patients with polypharmacy, there are documented cases of concurrent use of 3 or more suspect drugs. This raises the possibility that a third drug may act as a confounding factor in the relationship between coadministration and TdP risk, which was not accounted for in our analysis. Second, the use of a spontaneous reporting system inherently introduces various biases into the signal detection process, thus rendering the obtained signals as hypotheses rather than definitive conclusions. Moreover, the absence of a universally accepted gold standard for DDI signal detection further impacts the reliability of our findings. Although we devoted significant attention to the interpretation of results in the context of signal research for DDIs, validation efforts have been conducted using resources such as Lexicomp and Drugs.com. Unfortunately, there is a paucity of reliable information on DDIs, and the available data exhibit substantial variability across different drug information resources. Furthermore, there is a scarcity of robust clinical evidence on DDIs. Consequently, it is challenging to use "real" true data for signal verification. Third, although 2 databases were used for signal validation, there exists a risk of overlooking genuine DDIs due to inconsistencies between the databases; when considering only the DDIs present in both as true signals, some valid associations may have been excluded. Fourth, the failure to convert drug trade names to their generic counterparts prior to analysis may have resulted in an elevated number of DDI-positive signals. Nevertheless, this did not substantially affect the outcomes, as the analysis exclusively included DDIs validated by Lexicomp and Drugs.com using generic names. Fifth, we did not initially stratify the analysis based on the reporter population, as reports from nonprofessionals, such as lawyers and consumers, made up only 6.7% of the total. However, despite this small proportion, we could not completely

XSL•FO
RenderX

rule out the potential influence of these nonprofessional reports on the signal detection results due to their limited ability to differentiate between AE symptoms and underlying disease symptoms. Last, the FAERS database lacks sufficient clinical details to exclude comorbidities, such as heart-related diseases, as potential contributing factors to TdP cases. Future research should integrate more comprehensive clinical data or conduct prospective studies to better distinguish between medication-induced and disease-induced TdP, thereby enhancing the reliability of such analyses.

Conclusion

This study provides a preliminary overview of potential culprit DDIs, as well as negative DDIs, for TdP in real-world settings. These findings offer valuable insights for regulators, health care professionals, and stakeholders involved in DDI management. However, it is important to acknowledge the limitations inherent to our study stemming from the reliance on a pharmacovigilance database. It is crucial to recognize that DDI signals only indicate a statistical association between drug combinations and AEs and should not be mistaken for confirmed causal relationships. Further validation through well-designed clinical studies is necessary to establish a causal relationship. Although we validated signals using DDI resources and literature-based evidence, the available clinical evidence is limited and subject to significant inconsistencies. Therefore, in clinical practice, DDI signals should be considered as supplementary evidence rather than a substitute for the expertise of cardiologists and clinical pharmacists.

Although our study provides promising initial findings regarding the identification of high-priority DDIs for TdP, the scalability of these findings should be framed as hypotheses for future research. Specifically, the ability to generalize these results for clinical settings requires further investigation. Future studies should integrate clinical data from electronic health records or prospective trials to validate our findings and provide more actionable insights for clinicians.

Acknowledgments

Funding was provided by the Future Medicine Youth Innovation Team project of Chongqing Medical University (W0081), the Key Project of Chongqing Science and Health Joint Medical Scientific Research Project (2022ZDXM020), Youth project of Chongqing Science and Health Joint Medical Scientific Research Project (2023QNXM033), and the Smart Medicine Project Affiliated to the Chongqing Medical University (ZHYX202216).

Authors' Contributions

HJ and MG contributed equally as co-first authors. YJ (jiayuntaomail@hospital.cqmu.edu.cn) and LS (songlin@hospital.cqmu.edu.cn) contributed equally as co-corresponding authors. HJ, LS, and YJ designed the study. HJ, MG, NZ, RZ, and LS extracted the data. HJ, MG, LG, LS, DD, YY, and YJ analyzed and interpreted the data. All authors were involved in the study, took part in the discussions of the results, and contributed to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detection of clinically significant drug-drug interactions in fatal torasdes de pointes.
[DOCX File , 61 KB - jmir_v27i1e65872_app1.docx]

Multimedia Appendix 2
Characteristics of drug-drug interactions for torsades des pointes.
[DOCX File , 128 KB - jmir_v27i1e65872_app2.docx]

References

1. Woosley RL, Heise CW, Gallo T, Woosley RD, Lambson J, Romero KA. QTdrugs List. CredibleMeds. URL: https://www.crediblemeds.org/druglist [accessed 2023-06-12]
2. Chien HT, Lin FJ, Juang JJ, Lin SW. The impact of QT-prolonging medications and drug-drug interactions on QTc interval prolongation in hospitalized patients: a case-crossover study. Clin Pharmacol Ther 2025 Feb;117(2):495-505. [doi: 10.1002/cpt.3469] [Medline: 39380486]
3. Muylle KM, van Laere S, Pannone L, Coenen S, de Asmundis C, Dupont AG, et al. Added value of patient- and drug-related factors to stratify drug-drug interaction alerts for risk of QT prolongation: development and validation of a risk prediction model. Br J Clin Pharmacol 2023 Apr;89(4):1374-1385. [doi: 10.1111/bcp.15580] [Medline: 36321834]
4. Wasserman RL, Seger DL, Amato MG, Hwang AY, Fiskio J, Bates DW. A calculated risk: evaluation of QTc drug-drug interaction (DDI) clinical decision support (CDS) alerts and performance of the Tisdale Risk Score calculator. Drug Saf 2024 Dec;47(12):1235-1243. [doi: 10.1007/s40264-024-01466-w] [Medline: 38982033]

XSL•FO
RenderX

5. Shariff A, Belagodu Sridhar S, Abdullah Basha NF, Bin Taleth Alshemeil SSH, Ahmed Aljallaf Alzaabi NA. Assessing consistency of drug-drug interaction-related information across various drug information resources. Cureus 2021 Mar 08;13(3):e13766 [FREE Full text] [doi: 10.7759/cureus.13766] [Medline: 33842142]
6. Nicholson K, Liu W, Fitzpatrick D, Hardacre KA, Roberts S, Salerno J, et al. Prevalence of multimorbidity and polypharmacy among adults and older adults: a systematic review. Lancet Healthy Longev 2024 Apr;5(4):e287-e296 [FREE Full text] [doi: 10.1016/S2666-7568(24)00007-2] [Medline: 38452787]
7. Al-Azayzih A, Al-Qerem W, Al-Azzam S, Muflih S, Al-Husein BA, Kharaba Z, et al. Prevalence of medication associated with QTc prolongation used among critically ill patients. Vasc Health Risk Manag 2024;20:27-37 [FREE Full text] [doi: 10.2147/VHRM.S438899] [Medline: 38318252]
8. Agnihotri A, Ramasubbu SK, Bandyopadhyay A, Bidarolli M, Nath UK, Das B. Prevalence, attributes, and risk factors of QT-interval-prolonging drugs and potential drug-drug interactions in cancer patients: a prospective study in a tertiary care hospital. Cureus 2024 May;16(5):e60492 [FREE Full text] [doi: 10.7759/cureus.60492] [Medline: 38882995]
9. Krumpholz L, Wiśniowska B, Polak S. Open-access database of literature derived drug-related torsade de pointes cases. BMC Pharmacol Toxicol 2022 Jan 10;23(1):7 [FREE Full text] [doi: 10.1186/s40360-021-00548-0] [Medline: 35012678]
10. Noguchi Y, Tachi T, Teramachi H. Review of statistical methodologies for detecting drug-drug interactions using spontaneous reporting systems. Front Pharmacol 2019;10:1319 [FREE Full text] [doi: 10.3389/fphar.2019.01319] [Medline: 31780939]
11. Noguchi Y, Tachi T, Teramachi H. Detection algorithms and attentive points of safety signal using spontaneous reporting systems as a clinical data source. Brief Bioinform 2021 Nov 05;22(6):1. [doi: 10.1093/bib/bbab347] [Medline: 34453158]
12. OpenVigil - open tools for data-mining and analysis of pharmacovigilance data. OpenVigil. URL: https://openvigil.sourceforge.net [accessed 2025-01-17]
13. Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug-drug interaction surveillance. Stat Med 2008 Jul 20;27(16):3057-3070. [doi: 10.1002/sim.3247] [Medline: 18344185]
14. Thakrar BT, Grundschober SB, Doessegger L. Detecting signals of drug-drug interactions in a spontaneous reports database. Br J Clin Pharmacol 2007 Oct;64(4):489-495 [FREE Full text] [doi: 10.1111/j.1365-2125.2007.02900.x] [Medline: 17506784]
15. Gosho M, Maruo K, Tada K, Hirakawa A. Utilization of chi-square statistics for screening adverse drug-drug interactions in spontaneous reporting systems. Eur J Clin Pharmacol 2017 Jun;73(6):779-786. [doi: 10.1007/s00228-017-2233-3] [Medline: 28280890]
16. Noguchi Y, Tachi T, Teramachi H. Comparison of signal detection algorithms based on frequency statistical model for drug-drug interaction using spontaneous reporting systems. Pharm Res 2020 Apr 30;37(5):86. [doi: 10.1007/s11095-020-02801-3] [Medline: 32356247]
17. Fayad F, Lefevre-Colau MM, Drapé JL, Feydy A, Chemla N, Quintéro N, et al. Reliability of a modified Modic classification of bone marrow changes in lumbar spine MRI. Joint Bone Spine 2009 May;76(3):286-289. [doi: 10.1016/j.jbspin.2008.09.012] [Medline: 19119042]
18. Facts and Comparisons Drug Interactions Module. Wolters Kluwer. URL: https://www.wolterskluwer.com/en/solutions/uptodate/resources/facts-comparisons/drug-interactions [accessed 2023-10-08]
19. Drug interaction checker. Drugs.com. URL: https://www.drugs.com/drug_interactions.html [accessed 2023-10-08]
20. Matsuo J, Yamaori S. Detecting drug-drug interactions that increase the incidence of long QT syndrome using a spontaneous reporting system. J Clin Pharm Ther 2022 Jan;47(1):70-80. [doi: 10.1111/jcpt.13539] [Medline: 34664726]
21. Yuan N, Oesterle A, Botting P, Chugh S, Albert C, Ebinger J, et al. High-throughput assessment of real-world medication effects on QT interval prolongation: observational study. JMIR Cardio 2023 Jan 20;7:e41055 [FREE Full text] [doi: 10.2196/41055] [Medline: 36662566]
22. Landerholm A, Fedotova NO, Levy-Carrick NC, Chung R, Funk MC. C-L case conference: torsades de pointes in a patient with lifelong medical trauma, COVID-19, remdesivir, citalopram, quetiapine, and hemodialysis. J Acad Consult Liaison Psychiatry 2023;64(2):147-157. [doi: 10.1016/j.jaclp.2022.11.001] [Medline: 36351521]
23. Devaux F, Fillet M, Krzesinski F. [Torsade de pointe resulting from a drug interaction between sotalol and ciprofloxacin]. Rev Med Liege 2019 Jul;74(7-8):382-387 [FREE Full text] [Medline: 31373450]
24. Wei A, Peng J, Gu Z, Li J. QTc prolongation and torsades de pointes due to a coadministration of fluoxetine and amiodarone in a patient with implantable cardioverter-defibrillator: case report and review of the literature. Medicine (Baltimore) 2017 Dec;96(49):e9071 [FREE Full text] [doi: 10.1097/MD.0000000000009071] [Medline: 29245320]
25. Khederlou H, Azimi Pirsaraei V. Torsades de Pointe Associated with Trazodone Consumption. Case Rep Crit Care 2024;2024:5759229. [doi: 10.1155/2024/5759229] [Medline: 38680420]
26. Wang L, Volkow ND, Berger NA, Davis PB, Kaelber DC, Xu R. Cardiac and mortality outcome differences between methadone, buprenorphine and naltrexone prescriptions in patients with an opioid use disorder. J Clin Psychol 2023 Dec;79(12):2869-2883. [doi: 10.1002/jclp.23582] [Medline: 37584532]
27. Nair MK, Patel K, Starer PJ. Ciprofloxacin-induced torsades de pointes in a methadone-dependent patient. Addiction 2008 Dec;103(12):2062-2064. [doi: 10.1111/j.1360-0443.2008.02390.x] [Medline: 19469750]
28. NoorZurani MH, Vicknasingam B, Narayanan S. Itraconazole-induced torsade de pointes in a patient receiving methadone substitution therapy. Drug Alcohol Rev 2009 Nov;28(6):688-690. [doi: 10.1111/j.1465-3362.2009.00128.x] [Medline: 19930027]

XSL•FO
RenderX

Abbreviations

AE: adverse event
DDI: drug-drug interaction.
FAERS: Food and Drug Administration Adverse Event Reporting System
MedDRA: Medical Dictionary for Regulatory Activities
PT: preferred term
QTc: heart rate–corrected QT interval
ROR: reporting odds ratio
TdP: torsades de pointes

Edited by J Sarvestan; submitted 28.08.24; peer-reviewed by S Liu, Y Noguchi; comments to author 26.12.24; revised version received 20.01.25; accepted 10.03.25; published 25.03.25.

Please cite as:
Ji H, Gong M, Gong L, Zhang N, Zhou R, Deng D, Yang Y, Song L, Jia Y
Detection of Clinically Significant Drug-Drug Interactions in Fatal Torsades de Pointes: Disproportionality Analysis of the Food and Drug Administration Adverse Event Reporting System
J Med Internet Res 2025;27:e65872
URL: https://www.jmir.org/2025/1/e65872
doi:10.2196/65872
PMID:

©Huanhuan Ji, Meiling Gong, Li Gong, Ni Zhang, Ruiou Zhou, Dongmei Deng, Ya Yang, Lin Song, Yuntao Jia. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Recruitment of Young Gay, Bisexual, and Other Men Who Have Sex With Men for a Web-Based Human Papillomavirus Vaccination Intervention: Differences in Participant Characteristics and Study Engagement by Recruitment Source in a Randomized Controlled Trial

Daniel J Marshall[1], MSc; Amy L Gower[2], PhD; Mira L Katz[1,3], PhD; José A Bauermeister[4], PhD; Abigail B Shoben[1], PhD; Paul L Reiter[1,3], PhD

[1]College of Public Health, The Ohio State University, Columbus, OH, United States

[2]Division of General Pediatrics and Adolescent Health, Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, United States

[3]Comprehensive Cancer Center, The Ohio State University, Columbus, OH, United States

[4]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Paul L Reiter, PhD
College of Public Health
The Ohio State University
1841 Neil Ave.
Columbus, OH, 43210
United States
Phone: 1 6142924803
Email: reiter.36@osu.edu

Abstract

Background: Young gay, bisexual, and other men who have sex with men have been referred to as a “hard-to-reach” or “hidden” community in terms of recruiting for research studies. With widespread internet use among this group and young adults in general, web-based avenues represent an important approach for reaching and recruiting members of this community. However, little is known about how participants recruited from various web-based sources may differ from one another.

Objective: This study aimed to determine how young gay, bisexual, and other men who have sex with men recruited from various web-based sources differ from one another in terms of participant characteristics and study engagement.

Methods: Data were collected as part of a randomized controlled trial of Outsmart HPV, a web-based human papillomavirus (HPV) vaccination intervention for young gay, bisexual, and other men who have sex with men. From 2019 to 2021, we recruited young gay, bisexual, and other men who have sex with men in the United States who were aged 18-25 years and not vaccinated against HPV (n=1227) through various web-based avenues. We classified each participant as being recruited from either (1) social media (eg, Facebook, Instagram, Snapchat), (2) a dating app (eg, Grindr, Scruff), or (3) some other digital recruitment source (eg, existing research panel, university-based organization). Analyses compared participants from these 3 groups on demographic and health-related characteristics and metrics involving study engagement.

Results: Most demographic and health-related characteristics differed by web-based recruitment source, including race or ethnicity (P<.001), relationship status (P<.001), education level (P<.001), employment status (P<.001), sexual self-identity (P<.001), health insurance status (P<.001), disclosure of sexual orientation (P=.048), and connectedness to the LGBTQ (lesbian, gay, bisexual, transgender, queer) community (P<.001) The type of device used by participants during study enrollment also differed across groups, with smartphone use higher among participants recruited via dating apps (n=660, 96.6%) compared to those recruited via social media (n=318, 78.9%) or other digital sources (n=85, 60.3%; P<.001). Participants recruited via social media were more likely than those recruited via dating apps to complete follow-up surveys at 3 different timepoints (odds ratios 1.52-2.09, P=.001-.008). These participants also spent a longer amount of time viewing intervention content about HPV vaccination (3.14 minutes vs 2.67 minutes; P=.02).

XSL•FO
RenderX

Conclusions: We were able to recruit a large national sample of young gay, bisexual, and other men who have sex with men for a web-based HPV vaccination intervention via multiple methodologies. Participants differed on a range of demographic and health-related characteristics, as well as metrics related to study engagement, based on whether they were recruited from social media, a dating app, or some other digital recruitment source. Findings highlight key issues and considerations that can help researchers better plan and customize future web-based recruitment efforts of young gay, bisexual, and other men who have sex with men.

Trial Registration: ClinicalTrials.gov NCT04032106; https://clinicaltrials.gov/study/NCT04032106

International Registered Report Identifier (IRRID): RR2-10.2196/16294

(J Med Internet Res 2025;27:e64668) doi:10.2196/64668

KEYWORDS

study recruitment; gay and bisexual men; human papillomavirus; vaccination promotion; digital intervention; social media; dating apps; recruitment; young adults; adolescents; gay; bisexual; men who have sex with men

Introduction

Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States [1], with infections having the potential to cause multiple types of cancer (ie, anal, oropharyngeal, and penile cancers) and genital warts in men [2]. Gay, bisexual, and other men who have sex with men tend to have higher rates of HPV infection and HPV-related disease compared to other men [3-5]. Routine HPV vaccination is currently recommended for those aged 11-12 years in the United States, while also recommended for everyone through the age of 26 years who has not already been vaccinated [6]. However, a recent review paper suggests that fewer than 40% of young gay, bisexual, and other men who have sex with men who are age-eligible for HPV vaccination have received any doses of the vaccine series [5]. In response, HPV vaccination interventions for young gay, bisexual, and other men who have sex with men have recently been developed to improve knowledge and increase vaccination rates [7-10].

One issue central to these interventions is reaching and recruiting young gay, bisexual, and other men who have sex with men as study participants [11]. Gay, bisexual, and other men who have sex with men have previously been referred to as a "hard-to-reach" or "hidden" population for study recruitment [12,13], and traditional recruitment approaches for this community often included snowball sampling, use of "gatekeeper" organizations associated with the community, and venue-based recruitment at community events, health clinics, or other settings [11,14,15]. These traditional approaches faced recruitment challenges including potential participants not being comfortable identifying as gay, bisexual, and other men who have sex with men, lack of inclusive language in recruitment materials, and concerns about stigmatization associated with participation [11-13,15,16].

Web-based recruitment methods, including the use of social media and dating apps, present an alternative approach that has gained popularity for recruiting young gay, bisexual, and other men who have sex with men [17-21]. Social media use is ubiquitous among young adults in the United States, with nearly all reporting prior use [22]. Further, about 80% of young gay, bisexual, and other men who have sex with men report using a dating app on at least a monthly basis [23]. Past studies have shown that digital venues can successfully reach and recruit large numbers of participants, including young gay, bisexual, and other men who have sex with men, and that the content of study advertisements on these platforms can affect recruitment metrics [17,24-29]. Studies have also provided a great deal of information about the costs associated with web-based recruitment [24,28,29] and shown that participants recruited via web-based methodologies are demographically different from those recruited through more traditional approaches [13,30-33]. For example, gay, bisexual, and other men who have sex with men participants recruited via web-based methodologies may differ in terms of age, race or ethnicity, and socioeconomic status (eg, education level) compared to those recruited via more traditional approaches [13,30-32]. Furthermore, a few studies have shown that gay, bisexual, and other men who have sex with men participants tend to be more diverse when multiple recruitment sources are used (eg, using Facebook and Craigslist) [18,21].

Even with our current amount of knowledge about the digital recruitment of young gay, bisexual, and other men who have sex with men, important research gaps remain. One key area to examine is how participants recruited from various web-based sources (eg, social media vs dating apps) may differ from one another. This is true not only for demographic characteristics but also for participants' study experience (eg, retention, engagement with study content). The latter is particularly relevant to studies that include the delivery of interventions or involve longitudinal data collection. In this report, we address these research gaps by analyzing data from Outsmart HPV, an HPV vaccination intervention study that recruited young gay, bisexual, and other men who have sex with men through various web-based venues. Results will inform and help guide future web-based recruitment efforts of young gay, bisexual, and other men who have sex with men for research studies.

Methods

Participants

All data were collected as part of a randomized controlled trial (RCT) of Outsmart HPV. The methods of the RCT have been described previously [7] and are briefly summarized below. We recruited a convenience sample of young gay, bisexual, and other men who have sex with men via social media sites, dating

XSL•FO
RenderX

apps, and other web-based avenues. Advertisements on these digital platforms included a combination of images (eg, men in the targeted age range) and brief text about the study (eg, information about HPV, the study being web-based). Although we standardized advertisement content where possible, some content did differ across platforms due to varying advertisement requirements and options (eg, amount of text allowed, advertisement placement). For example, some of our advertisements through dating apps featured content that was sent directly to users' in-app inboxes over a 24-hour period, which was an approach that was not available across all platforms.

Interested individuals were linked via advertisements to a mobile-friendly project website to complete an eligibility screener. Eligibility criteria included self-identifying as (1) cisgender male; (2) aged 18-25 years; (3) either gay, bisexual, or queer; ever having oral or anal sex with another male; or being sexually attracted to other males; (4) living in the United States; (5) not having received any doses of HPV vaccine; and (5) not having previously participated in Outsmart HPV. Eligible individuals provided informed consent and created a project website account.

Participants next completed a baseline survey ("T1 survey") and were then immediately randomized using a 1:1:1 allocation scheme to 1 of 3 study groups. The 3 study groups included 2 groups that received intervention content and 1 control group [7]. Participants then viewed intervention content about HPV vaccination on the project website, with participants in both intervention groups viewing Outsmart HPV content about HPV vaccination and participants in the control group viewing standard information about HPV vaccination. After viewing the intervention content, participants completed a second survey ("T2 survey"). Additional follow-up surveys occurred 3 and 9 months later ("T3 survey" and "T4 survey," respectively). All study surveys were completed on the project website. A total of 1227 participants were randomized from October 2019 to June 2021.

Measures

The main independent variable for this report was the type of web-based recruitment source, with each participant categorized as being recruited from either: (1) social media (eg, Facebook, Instagram, Snapchat); (2) a dating app (eg, Grindr, Scruff); or (3) some other recruitment source (eg, existing research panel, university-based organization).

The T1 survey collected information on a range of demographic and health-related characteristics. This included participants' disclosure of their sexual orientation (3 items; α=.77; possible range 1-5) [34], concealment of their sexual orientation (3 items; α=.75; possible range 1-5) [34], connectedness to the LGBTQ (lesbian, gay, bisexual, transgender, queer) community (2 items; α=.72; possible range 1-4), and electronic health literacy (4 items; α=.83; possible range 1-5) [35]. Items were coded so that higher values indicate greater levels of a given construct. We examined the type of device that participants used during study enrollment, with each participant categorized as using either a smartphone, tablet device, or a personal computer (eg, desktop, laptop).

To examine study retention, we determined whether participants completed the T2, T3, and T4 surveys. For each survey separately, participants were categorized as having either completed the survey or not completed the survey. Among participants in the 2 intervention groups, we also examined additional metrics of study engagement by calculating the total amount of time (in minutes) spent viewing Outsmart HPV content about HPV vaccination and the total number of logins to the project website. Both of these metrics were winsorized so that the top 10% of values were set to equal the value corresponding to the 90th percentile.

Data Analysis

We first calculated descriptive statistics for all variables. We examined differences by recruitment source in participants' demographic and health-related characteristics and device type. To make these comparisons, we used the Pearson chi-square test for categorical dependent variables and 1-way ANOVA (analysis of variance) for continuous dependent variables. We made post hoc pairwise comparisons between the 3 web-based recruitment source groups and used the Bonferroni adjustment to account for multiple comparisons.

We then used regression models to examine differences by recruitment source in study retention and other study engagement metrics. Logistic regression models produced odds ratios (ORs) and 95% CIs for completion of the T2, T3, and T4 surveys. Linear regression models produced standardized β coefficients for the other study engagement metrics. The recruitment source of dating apps served as the referent group in regression models. These statistical tests were 2-tailed with a critical α value of .05. Data were analyzed using R (version 4.2.1; R Foundation for Statistical Computing), and Stata (version 15.0; StataCorp).

Ethical Considerations

The institutional review board at The Ohio State University approved this study (IRB 2019C0028), and the RCT is registered at ClinicalTrials.gov (NCT04032106). We based the reporting of results for this manuscript on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [36] since most of the included data were collected at baseline. Informed consent was obtained from all participants, and all data were deidentified. Participants could earn up to US $95 in gift cards for completing study surveys.

Results

Participant Characteristics

Overall, 52.8% (n=648) of participants indicated a minoritized racial or ethnic identity, 64.4% (n=790) were between 22 and 25 years of age, 69% (n=847) had at least some college education, and 66.4% (n=815) self-identified as gay. Most participants (n=974, 79.4%) had some form of health insurance, but more than half (n=666, 54.3%) had not had a preventive health visit in the prior year.

A total of 403 (32.8%) participants were recruited via social media, 683 (55.7%) participants via dating apps, and 141 (11.5%) participants via other recruitment sources. Most

XSL•FO
RenderX

demographic and health-related characteristics differed by recruitment source (Table 1). Generally, a higher proportion of participants recruited via dating apps reported a minoritized racial or ethnic identity, were single and having sex or casually dating, had less education, reported having sex with a male partner in the past, and had public or no health insurance. Conversely, a higher proportion of participants recruited via social media indicated not being currently employed full-time or part-time, self-identified as gay, and reported being HIV-negative. Participants recruited via social media also reported a higher level of connectedness to the LGBTQ community and a higher level of disclosure of their sexual orientation. Table 1 shows the results of all pairwise comparisons.

The type of device used by participants during study enrollment also differed by recruitment source (P<.001; Table 1). Smartphone use was higher among participants recruited via dating apps (n=660, 96.6%) compared to those recruited via social media (n=318, 78.9%) or other web-based sources (n=85, 60.3%). Conversely, personal computer use was higher among participants recruited via social media (n=81, 20.1%) or other web-based sources (n=55, 39%) compared to those recruited via dating apps (n=17, 2.5%).

XSL•FO
RenderX

Table 1. Participant characteristics of young gay, bisexual, and other men who have sex with men in the United States by recruitment source for a web-based human papillomavirus vaccination intervention, 2019-2021 (N=1227)[a].

Characteristics	Social media (n=403)	Dating app (n=683)	Other online sources (n=141)	Test statistic	*P* value
Demographic characteristics					
Age (years), n (%)				1.9	.39
18-21	144 (35.7)	250 (36.6)	43 (30.5)		
22-25	259 (64.3)	433 (63.4)	98 (69.5)		
Race or ethnicity, n (%)				55.6	<.001[b,c]
Hispanic	84 (20.8)	244 (35.7)	24 (17)		
Non-Hispanic White	224 (55.6)	280 (41)	75 (53.2)		
Non-Hispanic Black	30 (7.4)	85 (12.5)	14 (9.9)		
Another race or ethnicity	65 (16.1)	74 (10.8)	28 (19.9)		
Relationship status, n (%)				50.5	<.001[b,c,d]
Single and not having sex	57 (14.1)	83 (12.2)	30 (21.3)		
Single and having sex or casually dating	239 (59.3)	492 (72)	62 (44)		
In a relationship	107 (26.6)	108 (15.8)	49 (34.8)		
Education level, n (%)				38.0	<.001[b,c]
High school or less	90 (22.3)	261 (38.2)	29 (20.6)		
Some college or more	313 (77.7)	422 (61.8)	112 (79.4)		
Employment status, n (%)				22.9	<.001[b,d]
Employed full time or part time	253 (62.8)	520 (76.1)	105 (74.5)		
Other	150 (37.2)	163 (23.9)	36 (25.5)		
Region of residence, n (%)				12.1	.06
Northeast	90 (22.3)	113 (16.5)	25 (17.7)		
Midwest	77 (19.1)	114 (16.7)	33 (23.4)		
South	136 (33.8)	245 (35.9)	48 (34)		
West	100 (24.8)	211 (30.9)	35 (24.8)		
Sexual identity, n (%)				40.2	<.001[b,c,d]
Gay	296 (73.5)	449 (65.7)	70 (49.7)		
Bisexual	76 (18.9)	194 (28.4)	47 (33.3)		
Another identity	31 (7.7)	40 (5.9)	24 (17.0)		
Ever had sex with a male, n (%)				43.3	<.001[b,c,d]
No	35 (8.7)	24 (3.5)	26 (18.4)		
Yes	368 (91.3)	659 (96.5)	115 (81.6)		
Sexually attracted to males, n (%)				5.9	.051[c]
No	12 (3.0)	17 (2.5)	9 (6.4)		
Yes	391 (97.0)	666 (97.5)	132 (93.6)		
Connectedness to LGBTQ[e] community, mean (SD)[f]	2.7 (0.8)	2.4 (0.9)	2.3 (0.8)	29.3	<.001[b,d]
Disclosure of sexual orientation, mean (SD)[g]	3.3 (1.6)	3.0 (1.6)	3.1 (1.5)	3.9	.048[b]
Concealment of sexual orientation, mean (SD)[h]	2.5 (1.2)	2.7 (1.2)	2.5 (1.1)	0.4	.53
Health-related characteristics					

XSL•FO
RenderX

Characteristics	Social media (n=403)	Dating app (n=683)	Other online sources (n=141)	Test statistic	*P* value
Health insurance, n (%)				34.1	<.001[b,c]
Private insurance	285 (70.7)	379 (55.5)	104 (73.8)		
Public insurance	53 (13.2)	134 (19.6)	19 (13.5)		
None or do not know	65 (16.1)	170 (24.9)	18 (12.8)		
Last preventive health visit, n (%)				0.7	.72
Within the last year	187 (46.4)	306 (44.8)	68 (48.2)		
More than a year ago	216 (53.6)	377 (55.2)	73 (51.8)		
HIV status, n (%)				7.7	.02[b]
Negative	392 (97.3)	638 (93.4)	133 (94.3)		
Positive	11 (2.7)	45 (6.6)	8 (5.7)		
Electronic health literacy, mean (SD)[i]	4.0 (0.7)	3.9 (0.8)	3.9 (0.8)	2.8	.10
Device information					
Device type, n (%)				175.9	<.001[b,c,d]
Smartphone	318 (78.9)	660 (96.6)	85 (60.3)		
Tablet device	4 (1)	6 (0.9)	1 (0.7)		
Personal computer	81 (20.1)	17 (2.5)	55 (39)		

[a]Due to rounding, percentages may not total 100%. Totals may not equal N=1227 due to missing data. Reported test statistics and *P* values are from Pearson chi-square test for categorical variables and 1-way ANOVA (analysis of variance) for continuous variables. Superscript letters b, c, and d indicate statistically significant differences of post hoc pairwise comparisons using the Bonferroni adjustment to account for multiple comparisons.

[b]Social media compared to dating app.

[c]Dating app compared to other web-based source.

[d]Social media compared to other web-based sources.

[e]LGBTQ: lesbian, gay, bisexual, transgender, queer.

[f]2-item scale; items had a 4-point response scale ranging from 1="not at all" to 4="a lot."

[g]3-item scale; items had a 5-point response scale ranging from 1="none" to 5="all."

[h]3-item scale; items had a 5-point response scale ranging from 1="never" to 5="always."

[i]4-item scale; items had a 5-point response scale ranging from 1="strongly disagree" to 5="strongly agree."

Study Engagement

Participant retention overall was 90% for the T2 survey (n=1001), 70% for the T3 survey (n=858), and 69% for the T4 survey (n=842). Compared to participants recruited via dating apps, retention was higher among participants recruited via social media for the T2 survey (94% vs 87%; OR 2.09, 95% CI 1.32-3.30), the T3 survey (75% vs 66%; OR 1.52, 95% CI 1.15-2.00), and the T4 survey (75% vs 65%; OR 1.66, 95% CI 1.26-2.19; Figure 1). There were no statistically significant differences in retention for participants recruited via other web-based sources.

Among participants in the 2 intervention groups (n=815), the average amount of time spent viewing intervention content about HPV vaccination was 2.83 minutes. Participants recruited via social media spent a longer amount of time viewing this content compared to those recruited via dating apps (3.14 minutes vs 2.67 minutes; 95% CI 0.03-0.33; *P*=.02; Table 2). The average number of project website logins was 4.94, and there were no differences in the number of logins by recruitment source (*P*=.83 for social media, *P*=.54 for other web-based sources; Table 2).

XSL•FO
RenderX

Figure 1. Survey retention of young gay, bisexual, and other men who have sex with men in the United States by recruitment source for a web-based human papillomavirus vaccination intervention (N=1227). Error bars represent SE.

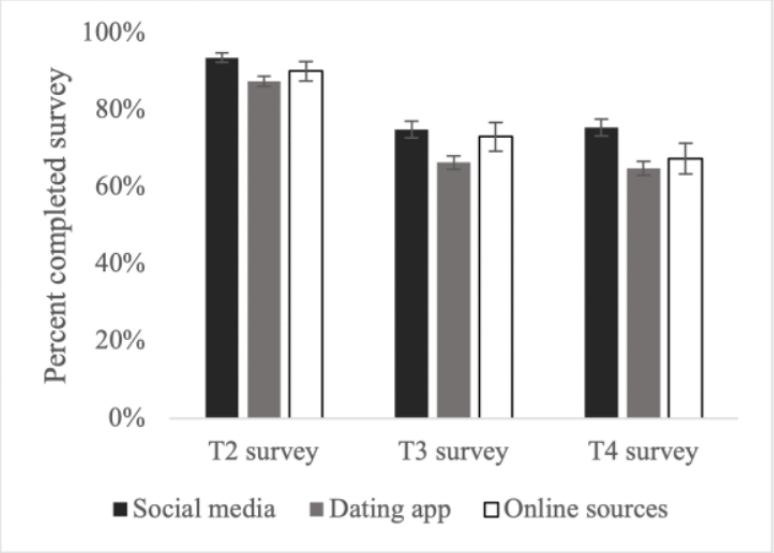

Table 2. Study engagement of young gay, bisexual, and other men who have sex with men in the United States by recruitment source for a web-based human papillomavirus vaccination intervention (n=815)[a,b].

	Time spent on intervention content (minutes)		Number of project website log-ins	
	Mean (SD)	β (95% CI)	Mean (SD)	β (95% CI)
Dating app	2.67 (2.44)	Reference	4.98 (2.40)	Reference
Social media	3.14 (2.80)	0.18[b] (0.03-0.33)	4.94 (2.42)	–0.02 (–0.17 to 0.13)
Other web-based sources	2.74 (2.56)	0.02 (–0.20 to 0.25)	4.81 (2.17)	–0.07 (–0.29 to 0.15)

[a]Table reports data for the 815 participants in the 2 intervention groups. β represents standardized regression coefficients from linear regression models.
[b]Statistically significant at the α=.05 level.

Discussion

Principal Results

With digital methodologies becoming increasingly popular for recruiting young gay, bisexual, and other men who have sex with men for research studies, [17-21] it is important to examine if different web-based recruitment sources yield varying types of participants. The Outsmart HPV study provided an opportunity to address this research topic by recruiting a national sample of young gay, bisexual, and other men who have sex with men through multiple web-based sources, and there are 2 main findings. First, most of the demographic and health-related characteristics examined differed by recruitment sources. Participants recruited via social media tended to be less diverse (eg, race or ethnicity, sexual self-identity), of higher socioeconomic status (eg, education level, health insurance status), and have higher levels of outness (eg, disclosure of sexual orientation, connectedness to the LGBTQ community) compared to participants recruited via dating apps. This pattern differs from a past study that found participants recruited from a social media site (Facebook) were highly similar to those recruited from a dating app (Grindr) in terms of race or ethnicity [18]. Interestingly, the characteristics of participants recruited via social media in our study better align with the characteristics of other national samples of young gay, bisexual, and other men who have sex with men and the larger gay, bisexual, and other men who have sex with men community in the United States than participants recruited via dating apps in our study [37-40]. A potential explanation for this pattern is that users of dating apps tend to be more demographically diverse compared to the general population [41]. Interestingly, some characteristics (eg, socioeconomic status) of participants that were recruited via social media for our study were more comparable to gay, bisexual, and other men who have sex with men recruited through in-person venues in past research than via dating apps in our study [42].

Second, study engagement also differed by recruitment source. Participants who were recruited via dating apps tended to interact less with study content and were less likely to complete follow-up surveys compared to participants recruited via social media. This may be attributable to users of dating apps having an emphasis on immediate gratification [43] and impulsivity [44] as a justification for their app usage (and study enrollment), which in turn may affect their awareness of and interest in engaging with study content and completing study-related activities (eg, follow-up surveys). Related, given that dating app users are more impulsive than nonusers, [44] young gay, bisexual, and other men who have sex with men who use dating apps might experience an initial interest in joining a research study and earning an initial study incentive. However, interest in completing study-related activities at later timepoints may

XSL•FO
RenderX

not be as strong as their general preference for some level of anonymity [45] might dissuade them from wanting to continue discussing their health behaviors (eg, sexual history, vaccination status).

We think these 2 main findings can help researchers better plan and customize future digital recruitment efforts of young gay, bisexual, and other men who have sex with men by highlighting key issues and considerations. For example, researchers will now be able to better anticipate how the selection of web-based recruitment sources may affect the representativeness and diversity (and the potential balance between the two) of their study sample. Recruiting through social media may lead to a sample of participants that better reflect the larger young gay, bisexual, and other men who have sex with men population in the United States, but dating apps may be more effective in reaching more diverse young gay, bisexual, and other men who have sex with men, including those from communities that are often underrepresented in research studies [46]. Recruitment goals differ across studies in terms of representativeness and diversity, and our data on differences in demographic and health-related characteristics can help inform future efforts. Furthermore, studies involving interventions and/or longitudinal data collection should consider how various web-based recruitment sources can impact study retention and participants' engagement with study content. Some studies may want to maximize participant retention and engagement (ie, suggesting that recruitment primarily through social media may be appropriate), whereas other studies may be willing to have slightly lower study engagement metrics in exchange for being able to include additional recruitment sources, such as dating apps. Future studies that do include recruitment from dating apps may also want to consider prioritizing brief study-related activities and other strategies to promote engagement.

One final implication involves the types of devices that were used by participants in our study. Overall, a majority of participants used a smartphone to enroll in the study, including nearly all participants who were recruited via dating apps. This pattern is similar to national data showing that young adults use smartphones much more frequently than other device types [47,48]. Thus, it is important that future digital interventions and other research studies for young adults, including young gay, bisexual, and other men who have sex with men, ensure that study content and data collection instruments are optimized for use on smartphones. For example, to increase usability, it has been suggested that the interface of smartphone-based interventions have a minimal number of screens, limited manual data entry, limited pop-ups and notifications, and personal identification number-based entry (as compared to full passwords). The content of such interventions should be succinct with easy-to-read graphics [49]. Data collection instruments (eg, surveys) should have limited open-ended questions and different "pages" in the survey, rather than having participants scroll through one continuous page. Lastly, and perhaps most generally, mobile surveys should be pilot-tested prior to the start of data collection activities [50]. As smartphone usage continues to be a common approach for participants enrolling in and interacting with digital interventions, such recommendations can help optimize intervention delivery and data collection.

Limitations

Although we were able to compare different groupings of web-based recruitment sources, we were not able to examine potential differences between individual platforms (eg, Facebook vs Twitter vs Grindr) due to modest sample sizes for some of the platforms. These sample sizes also did not allow us to examine how differences between recruitment sources may have changed over time during the recruitment period for our study. Examining potential differences between individual platforms and potential temporal changes are important areas for future research. Our recruitment focused on the social media and dating apps that were among the most popular among young adults in the United States around the time of the start of our study [51,52]. It is likely that the use and popularity of various platforms will evolve over time, and future efforts should examine new and emerging platforms. We were not able to examine advanced metrics related to recruitment efficiency and cost due to a lack of available data. However, past studies have reported such information for various web-based platforms [17,27]. Study recruitment occurred both prior to and during the COVID-19 pandemic, and it is not known how the pandemic may have affected the interest of young gay, bisexual, and other men who have sex with men to enroll in a study focusing on vaccination (which would subsequently affect the characteristics of study participants). Fraudulent accounts with inaccurate self-reported data occur in digital studies, but we used several recommended strategies for detecting such accounts and minimizing this risk [53,54].

Conclusions

We were able to recruit a large national sample of young gay, bisexual, and other men who have sex with men for an HPV vaccination intervention via multiple web-based methodologies. Participants differed on a range of demographic and health-related characteristics based on whether they were recruited from social media, a dating app, or some other web-based recruitment source. Study retention and engagement with study content also differed by web-based recruitment source. These findings and patterns highlight key issues and potential tradeoffs that can help researchers better plan and customize future web-based recruitment efforts of young gay, bisexual, and other men who have sex with men.

Acknowledgments

The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (R37CA226682). Additional support was provided by the Center for Health Communications Research at the University of Michigan (P30CA046592) and the Recruitment, Intervention, and Survey Shared Resource at The Ohio State University

XSL•FO
RenderX

Comprehensive Cancer Center (P30CA016058). Generative artificial intelligence was not used at any point in the development of this manuscript. The authors would also like to acknowledge the contributions of Annie-Laurie McRee to this research project.

Data Availability

The data for this study may be made available from the corresponding author on reasonable request.

Authors' Contributions

DJM, ALG, and PLR created the original concept and design of this manuscript, with DJM writing the initial draft and performing the initial analyses under the supervision of ALG and PLR. DJM, ALG, MLK, JAB, ABS, and PLR all contributed to the final analysis and interpretation of the data, with all authors also providing critical revision and feedback on the final drafts. All authors reviewed the final manuscript. All authors approve of the final version to be published and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. About genital HPV infection. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/sti/about/about-genital-hpv-infection.html [accessed 2024-04-15]
2. HPV and cancer. National Cancer Institute. 2023. URL: https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-and-cancer [accessed 2023-04-17]
3. Wei F, Gaisa MM, D'Souza G, Xia N, Giuliano AR, Hawes SE, et al. Epidemiology of anal human papillomavirus infection and high-grade squamous intraepithelial lesions in 29900 men according to HIV status, sexuality, and age: a collaborative pooled analysis of 64 studies. Lancet HIV 2021;8(9):e531-e543 [FREE Full text] [doi: 10.1016/S2352-3018(21)00108-9] [Medline: 34339628]
4. Machalek DA, Poynten M, Jin F, Fairley CK, Farnsworth A, Garland SM, et al. Anal human papillomavirus infection and associated neoplastic lesions in men who have sex with men: a systematic review and meta-analysis. Lancet Oncol 2012;13(5):487-500. [doi: 10.1016/S1470-2045(12)70080-3] [Medline: 22445259]
5. Meites E, Wilkin TJ, Markowitz LE. Review of human papillomavirus (HPV) burden and HPV vaccination for gay, bisexual, and other men who have sex with men and transgender women in the United States. Hum Vaccin Immunother 2022;18(1):2016007 [FREE Full text] [doi: 10.1080/21645515.2021.2016007] [Medline: 35294325]
6. Meites E, Szilagyi PG, Chesson HW, Unger ER, Romero JR, Markowitz LE. Human papillomavirus vaccination for adults: updated recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 2019;68(32):698-702 [FREE Full text] [doi: 10.15585/mmwr.mm6832a3] [Medline: 31415491]
7. Reiter PL, Gower AL, Kiss DE, Malone MA, Katz ML, Bauermeister JA, et al. A web-based human papillomavirus vaccination intervention for young gay, bisexual, and other men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020;9(2):e16294 [FREE Full text] [doi: 10.2196/16294] [Medline: 32130192]
8. Gerend MA, Madkins K, Crosby S, Korpak AK, Phillips GL, Bass M, et al. Evaluation of a text messaging-based human papillomavirus vaccination intervention for young sexual minority men: results from a pilot randomized controlled trial. Ann Behav Med 2021;55(4):321-332 [FREE Full text] [doi: 10.1093/abm/kaaa056] [Medline: 32914838]
9. Fontenot HB, White BP, Rosenberger JG, Lacasse H, Rutirasiri C, Mayer KH, et al. Mobile app strategy to facilitate human papillomavirus vaccination among young men who have sex with men: pilot intervention study. J Med Internet Res 2020;22(11):e22878 [FREE Full text] [doi: 10.2196/22878] [Medline: 33146621]
10. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Increasing human papillomavirus vaccination among young gay and bisexual men: a randomized pilot trial of the Outsmart HPV intervention. LGBT Health 2018;5(5):325-329 [FREE Full text] [doi: 10.1089/lgbt.2018.0059] [Medline: 29979642]
11. Meyer IH, Wilson PA. Sampling lesbian, gay, and bisexual populations. J Counsel Psychol 2009;56(1):23-31. [doi: 10.1037/a0014587]
12. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. J Med Internet Res 2018;20(6):e197 [FREE Full text] [doi: 10.2196/jmir.9461] [Medline: 29914861]
13. Iott BE, Veinot TC, Loveluck J, Kahle E, Golson L, Benton A. Comparative analysis of recruitment strategies in a study of men who have sex with men (MSM) in Metropolitan Detroit. AIDS Behav 2018;22(7):2296-2311. [doi: 10.1007/s10461-018-2071-z] [Medline: 29470808]
14. López DJ, García A, Engel R, Jackson JL. Challenges and opportunities to recruiting and engaging with gay male Latino sexual assault survivors. Field Methods 2022;35(1):48-63. [doi: 10.1177/1525822x221103736]

XSL•FO
RenderX

15. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant. Int J Qual Methods 2015;14(5):160940691562142. [doi: 10.1177/1609406915621420]
16. Sanchez T, Smith A, Denson D, Dinenno E, Lansky A. Internet-based methods may reach higher-risk men who have sex with men not reached through venue-based sampling. Open AIDS J 2012;6:83-89 [FREE Full text] [doi: 10.2174/1874613601206010083] [Medline: 23049657]
17. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting young gay and bisexual men for a human papillomavirus vaccination intervention through social media: the effects of advertisement content. JMIR Public Health Surveill 2017;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
18. Buckingham L, Becher J, Voytek CD, Fiore D, Dunbar D, Davis-Vogel A, et al. Going social: success in online recruitment of men who have sex with men for prevention HIV vaccine research. Vaccine 2017;35(27):3498-3505 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.002] [Medline: 28526330]
19. Wilkerson JM, Shenk JE, Grey JA, Rosser BRS, Noor SW. Recruitment strategies of methamphetamine-using men who have sex with men into an online survey. J Subst Use 2015;20(1):33-37 [FREE Full text] [doi: 10.3109/14659891.2013.868936] [Medline: 25642143]
20. Brown JR, Reid D, Howarth AR, Mohammed H, Saunders J, Pulford CV, et al. Difficulty accessing condoms because of the COVID-19 pandemic reported by gay, bisexual and other men who have sex with men in the UK: findings from a large, cross-sectional, online survey. Int J STD AIDS 2023;34(8):541-547 [FREE Full text] [doi: 10.1177/09564624231160804] [Medline: 36943893]
21. Sun CJ, Sutfin E, Bachmann LH, Stowers J, Rhodes SD. Comparing men who have sex with men and transgender women who use Grindr, other similar social and sexual networking apps, or no social and sexual networking apps: implications for recruitment and health promotion. J AIDS Clin Res 2018;9(2):757 [FREE Full text] [doi: 10.4172/2155-6113.1000757] [Medline: 29593933]
22. Social media fact sheet. Pew Research Center. 2024 Nov 13. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2024-11-13]
23. Badal HJ, Stryker JE, DeLuca N, Purcell DW. Swipe right: dating website and app use among men who have sex with men. AIDS Behav 2018;22(4):1265-1272. [doi: 10.1007/s10461-017-1882-7] [Medline: 28884248]
24. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
25. Merchant RC, Romanoff J, Clark MA, Liu T, Rosenberger JG, Bauermeister J, et al. Variations in recruitment yield and characteristics of participants recruited across diverse internet platforms in an HIV testing study of young adult men-who-have-sex-with-men (YMSM). Am J Mens Health 2017;11(5):1342-1357 [FREE Full text] [doi: 10.1177/1557988317717383] [Medline: 28691552]
26. Moseson H, Kumar S, Juusola JL. Comparison of study samples recruited with virtual versus traditional recruitment methods. Contemp Clin Trials Commun 2020;19:100590 [FREE Full text] [doi: 10.1016/j.conctc.2020.100590] [Medline: 32637722]
27. Parker JN, Hunter AS, Bauermeister JA, Bonar EE, Carrico A, Stephenson R. Comparing social media and in-person recruitment: lessons learned from recruiting substance-using, sexual and gender minority adolescents and young adults for a randomized control trial. JMIR Public Health Surveill 2021;7(12):e31657 [FREE Full text] [doi: 10.2196/31657] [Medline: 34855613]
28. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and efficiency of online and offline recruitment methods: a web-based cohort study. J Med Internet Res 2017;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
29. Sharma P, Lamb LE, Bartolone SN, Ward EP, Janicki JJ, Peters KM, et al. A cost, time, and demographic analysis of participant recruitment and urine sample collection through social media optimization. Urol Pract 2022;9(6):561-566. [doi: 10.1097/upj.0000000000000339]
30. Gibbs JJ, Schrager SM, Goldbach JT. Young men who have sex with men and substance use: a comparison of venue-based sampling and geosocial networking application sampling. J Homosex 2024;71(5):1163-1176. [doi: 10.1080/00918369.2022.2161086] [Medline: 36630486]
31. Lachowsky NJ, Lal A, Forrest JI, Card KG, Cui Z, Sereda P, et al. Including online-recruited seeds: a respondent-driven sample of men who have sex with men. J Med Internet Res 2016;18(3):e51 [FREE Full text] [doi: 10.2196/jmir.5258] [Medline: 26980147]
32. Bezerra DRB, Jalil CM, Jalil EM, Coelho LE, Netto EC, Freitas J, et al. Comparing web-based venues to recruit gay, bisexual, and other cisgender men who have sex with men to a large HIV prevention service in Brazil: evaluation study. JMIR Form Res 2022;6(8):e33309 [FREE Full text] [doi: 10.2196/33309] [Medline: 35925658]
33. Bender JL, Akinnibosun R, Puri N, D'Agostino N, Drake EK, Tsimicalis A, et al. A comparison of the sociodemographic, medical, and psychosocial characteristics of adolescents and young adults diagnosed with cancer recruited in-person and online: a Canadian cross-sectional survey. Digit Health 2023;9:20552076231205278 [FREE Full text] [doi: 10.1177/20552076231205278] [Medline: 37900258]
34. Meidlinger PC, Hope DA. Differentiating disclosure and concealment in measurement of outness for sexual minorities: the Nebraska Outness Scale. P Sex Orientation Gender Divers 2014;1(4):489-497. [doi: 10.1037/sgd0000080]

XSL•FO
RenderX

35. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
36. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
37. LGBT Demographic Data Interactive. The Williams Institute, UCLA School of Law. 2019 Jan. URL: https://williamsinstitute.law.ucla.edu/visualization/lgbt-stats/?topic=LGBT#about-the-data [accessed 2024-03-01]
38. Cole SW, Glick JL, Campoamor NB, Sanchez TH, Sarkar S, Vannappagari V, et al. Willingness and preferences for long-acting injectable PrEP among US men who have sex with men: a discrete choice experiment. BMJ Open 2024;14(4):e083837 [FREE Full text] [doi: 10.1136/bmjopen-2023-083837] [Medline: 38653510]
39. Strutz KL, Herring AH, Halpern CT. Health disparities among young adult sexual minorities in the U.S. Am J Prev Med 2015;48(1):76-88 [FREE Full text] [doi: 10.1016/j.amepre.2014.07.038] [Medline: 25241194]
40. Reiter PL, McRee A, Katz ML, Paskett ED. Human papillomavirus vaccination among young adult gay and bisexual men in the United States. Am J Public Health 2015;105(1):96-102. [doi: 10.2105/AJPH.2014.302095] [Medline: 25393178]
41. Vogels EA, McClain C. Key findings about online dating in the U.S. Pew Research Center. 2023 Feb 02. URL: https://www.pewresearch.org/short-reads/2023/02/02/key-findings-about-online-dating-in-the-u-s [accessed 2023-02-02]
42. Salway TJ, Morgan J, Ferlatte O, Hawkins B, Lachowsky NJ, Gilbert M. A systematic review of characteristics of nonprobability community venue samples of sexual minority individuals and associated methods for assessing selection bias. LGBT Health 2019;6(5):205-215. [doi: 10.1089/lgbt.2018.0241] [Medline: 31135260]
43. Gudelunas D. There's an app for that: the uses and gratifications of online social networks for gay men. Sex Cult 2012;16(4):347-365. [doi: 10.1007/s12119-012-9127-4]
44. Sawyer AN, Smith ER, Benotsch EG. Dating application use and sexual risk behavior among young adults. Sex Res Soc Policy 2017;15(2):183-191. [doi: 10.1007/s13178-017-0297-6]
45. Waldman AE. Navigating privacy on gay-oriented mobile dating applications. In: Bailey J, Flynn A, Henry N, editors. The Emerald International Handbook of Technology-Facilitated Violence and Abuse. Bingley, United Kingdom: Emerald Publishing Limited; 2021.
46. Allison K, Patel D, Kaur R. Assessing multiple factors affecting minority participation in clinical trials: development of the clinical trials participation barriers survey. Cureus 2022;14(4):e24424 [FREE Full text] [doi: 10.7759/cureus.24424] [Medline: 35637812]
47. Cheung M, Lai JSK, Yip J. Influences of smartphone and computer use on health-related quality of life of early adolescents. Int J Environ Res Public Health 2022;19(4):2100 [FREE Full text] [doi: 10.3390/ijerph19042100] [Medline: 35206289]
48. Faverio M, Anderson M. For shopping, phones are common and influencers have become a factor—especially for young adults. Pew Research Center. 2022 Nov 21. URL: https://www.pewresearch.org/short-reads/2022/11/21/for-shopping-phones-are-common-and-influencers-have-become-a-factor-especially-for-young-adults/ [accessed 2022-11-21]
49. Gomolin A, Lebouché B, Engler K, Vedel I. Optimizing smartphone intervention features to improve chronic disease management: a rapid review. Health Informatics J 2020;26(3):1795-1809 [FREE Full text] [doi: 10.1177/1460458219891377] [Medline: 31826703]
50. Mobile survey optimization. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/mobile-survey-optimization/ [accessed 2024-12-04]
51. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-04-07]
52. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019 Apr 10. URL: https://www.pewresearch.org/short-reads/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2019-04-10]
53. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
54. Xu Y, Pace S, Kim J. Threats to online surveys: recognizing, detecting, and preventing survey bots. Soc Work Res 2022;46(4):343-350. [doi: 10.1093/swr/svac023]

Abbreviations

ANOVA: analysis of variance
HPV: human papillomavirus
LGBTQ: lesbian, gay, bisexual, transgender, and queer
RCT: randomized controlled trial
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.07.24; peer-reviewed by K Card, M Lim; comments to author 14.10.24; revised version received 04.11.24; accepted 24.11.24; published 03.01.25.

Please cite as:
Marshall DJ, Gower AL, Katz ML, Bauermeister JA, Shoben AB, Reiter PL
Recruitment of Young Gay, Bisexual, and Other Men Who Have Sex With Men for a Web-Based Human Papillomavirus Vaccination Intervention: Differences in Participant Characteristics and Study Engagement by Recruitment Source in a Randomized Controlled Trial
J Med Internet Res 2025;27:e64668
URL: https://www.jmir.org/2025/1/e64668
doi:10.2196/64668
PMID:

©Daniel J Marshall, Amy L Gower, Mira L Katz, José A Bauermeister, Abigail B Shoben, Paul L Reiter. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Urban-Suburban Differences in Public Perspectives on Digitalizing Pediatric Research: Cross-Sectional Survey Study

Heping Fang[1], MD, PhD; Ruoling Xian[2], MMeD; Juan Li[1], MD, PhD; Yingcun Li[3], MD, PhD; Enmei Liu[1], MD, PhD; Yan Zhao[2*], BD; Yan Hu[4,5*], MD, PhD

[1]Department of Respiratory Medicine, Pediatric Research Institute, Children's Hospital of Chongqing Medical University, Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, Chongqing, China

[2]Department of Pediatrics, Chongqing Health Center for Women and Children, Women and Children's Hospital of Chongqing Medical University, Chongqing, China

[3]Department of General Surgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

[4]Department of Child Health Care, Children's Hospital of Chongqing Medical University, Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, Chongqing, China

[5]Children's Health Care Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yan Hu, MD, PhD
Department of Child Health Care, Children's Hospital of Chongqing Medical University
Chongqing Key Laboratory of Child Rare Diseases in Infection and Immunity, Ministry of Education Key Laboratory of Child Development and Disorders
National Clinical Research Center for Child Health and Disorders
No. 20 Jinyu Avenue
Liangjiang New Area
Chongqing, 401122
China
Phone: 86 02368370551
Email: hy420@126.com

Abstract

Background: Recruiting and retaining participants in pediatric research has always been challenging, particularly in healthy populations and remote areas, leading to selection bias and increased health disparities. In the digital age, medical research has been transformed by digital tools, offering new opportunities to enhance engagement in clinical research. However, public perspectives on digitalizing pediatric research and potential differences between urban and suburban areas remain unclear.

Objective: This study aimed to investigate public perspectives on digitalizing pediatric research and compare differences between urban and suburban areas to help diversify participants and address health disparities.

Methods: A cross-sectional web-based survey targeting caregivers of kindergarten children (aged 2-7 years) in Chongqing was conducted between June and December 2023. A total of 4231 valid questionnaires were analyzed, with 25.1% (n=1064) of the children residing in urban areas and 74.9% (n=3167) in suburban areas. Descriptive statistics and intergroup comparisons were used for data analysis.

Results: Approximately 59.8% (n=2531) of the caregivers had first impressions of pediatric research, with 36.9% (n=1561) being positive and 22.9% (n=970) being negative. A total of 38.3% (n=1621) of caregivers recognized the growing popularity of digital tools, and 36.7% (n=1552) supported their use in pediatric research, but only 25.2% (n=1068) favored online-only research methods. The main concerns regarding the use of software in pediatric research were privacy issues (n=3273, 77.4%) and potential addiction (n=2457, 58.1%). Public accounts of research institutions (n=3400, 80.4%) were the most favored for online recruitment. Telephones (1916/3076, 62.3%) and social media apps (1801/3076, 58.6%) were the most popular for regular contact. Intergroup comparisons revealed that suburban caregivers had more positive first impressions of pediatric research (38.6% vs 32%; *P*<.001; adjusted odds ratio [aOR] 1.27, 95% CI 1.09-1.47) and faced fewer participation barriers: "worry about being an experimental subject" (70.9% vs 76.6%; *P*<.001; aOR 0.79, 95% CI 0.67-0.93), "pose a risk to children's health" (58.6% vs

XSL•FO
RenderX

67.8%; *P*<.001; aOR 0.71, 95% CI 0.61-0.83), "do not have enough background information" (55.2% vs 61.6%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89), and "worry about recommending other products" (48.2% vs 55%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89). They also showed greater support for online-only research methods (26% vs 22.9%; *P*=.045; aOR 1.19, 95% CI 1.01-1.41) and greater openness to unofficial online recruitment sources (social media friends: 24.7% vs 18.9%; *P*<.001; aOR 1.33, 95% CI 1.11-1.59; moments on social media: 15.5% vs 11.1%; *P*<.001; aOR 1.35, 95% CI 1.09-1.67).

Conclusions: In the digital age, enhancing recruitment and retention in pediatric research can be achieved by integrating both official and unofficial social media strategies, implementing a hybrid online-offline follow-up approach, and addressing privacy concerns.

(J Med Internet Res 2025;27:e60324) doi:10.2196/60324

KEYWORDS

pediatrics; pediatric research; digital health; public opinion; research; patient participation; urban; rural; caregiver attitudes; social media; mobile phone

Introduction

Recruitment and retention in pediatric research have consistently posed challenges [1-3]. In the United States, pediatric randomized clinical trials are often discontinued or unpublished, with patient recruitment difficulties (37%) being the primary reason [4]. Our experiences further corroborate this trend. For instance, in a cross-sectional hospital-based survey, we recruited only 65.8% of the caregivers [5]. Similarly, in a longitudinal study on the growth of critically ill children after liver transplantation, follow-up was completed for only 68.6% of the children one year after the procedure [6]. Recruitment and retention challenges frequently prevent studies from achieving their preset objectives, resulting in inefficiency and wasted personnel and financial resources.

While traditional research has focused on children with evident illnesses, emerging concerns such as overweight, internet addiction, and emotional or behavioral issues [7-10] highlight the importance of studying seemingly healthy children [11-13]. However, recruiting and retaining healthy children for research is more challenging than for those with specific diseases, as shown by pediatric reference interval studies [14]. Additionally, children in remote areas are frequently underrepresented in research [15-17], leading to selection bias and exacerbating health disparities. In China, geographical location has been a primary longstanding contributor to child health inequity [18]. Therefore, researchers need to devise strategies to address these challenges to meet the increasing demands of pediatric research. In this context, obtaining participants' perspectives by seeking their opinions may provide more practical solutions.

Caregivers, typically parents, are key decision makers regarding younger children's participation in research [19]. Numerous studies have examined parental decision-making in enrolling children in pediatric research, primarily focusing on traditional research methods [20-27]. In the digital age, medical research has been significantly transformed by digital tools and the process of digitalization [28,29]. Digital medicine has created new opportunities to enhance engagement in clinical research, especially in healthy populations and remote areas, which improves health outcomes for participants [30-32]. This study investigated public perspectives on pediatric research in the digital age and examined potential differences between urban and suburban areas. The goal was to provide insights to advance the digitalization of pediatric research, which may diversify the pool of participants and address health disparities among children.

Methods

Study Design and Participants

This study was a cross-sectional, web-based questionnaire survey. The study targeted caregivers of kindergarten children in Chongqing, a municipality in southwestern China, with a total of 12.58 million households and 995,239 kindergarten children. Chongqing is divided into urban and suburban areas based on geographical location and development level, with 3.07 million (23.9%) households in urban areas and 9.79 million (76.1%) in suburban areas [33]. Kindergarten children are typically between 3 and 6 years old, with exceptions depending on individual circumstances [34]. The inclusion criteria for kindergarten children were (1) residency in Chongqing, (2) kindergarten attendance, (3) aged 2 to 7 years, and (4) voluntary completion of the questionnaire by the caregiver. The exclusion criteria were (1) failure to pass validation questions, (2) logical errors in responses, and (3) self-reported lack of seriousness in answering.

Ethical Considerations

This study was approved by the Ethics Committee of the Children's Hospital of Chongqing Medical University (Protocol #2023-236), with informed consent signatures waived. The collected data were anonymized, and no compensation was given.

Reporting Guidelines

The CHERRIES (Checklists for Reporting Results of Internet E-Surveys) and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines were followed for reporting the results [35,36].

Research Outcomes and Measurements

The primary outcome of this study was public perspectives on pediatric research, with differences between urban and suburban areas as the secondary outcome. These perspectives were divided into 4 key categories: "Facilitators and Barriers to Participation," "Perspectives on Digital Medicine," "Perspectives on Recruitment," and "Perspectives on the Research Process." Each

XSL•FO
RenderX

category comprised single-choice and multiple-choice questions designed to collect data. In the intergroup comparison, participants were classified as urban or suburban (exposure) based on their responses to a residential address question (Multimedia Appendix 1).

Development of the Questionnaire

The questionnaire was initially developed and refined based on the literature and clinical experience [20-27,37-39], then revised and culturally adapted by 7 pediatric doctors with diverse professional backgrounds (see Acknowledgments). The final version was converted to a web-based format using "Lediaocha," a web-based survey platform, with 50 mandatory questions spread across 4 pages. Of these, 11 questions focused on general characteristics, while the rest investigated perspectives on pediatric research, particularly digital medicine. To ensure data reliability, one question on caregivers' roles was repeated at the beginning and end of the questionnaire, along with a self-assessment query on response seriousness before submission. Participation was voluntary, anonymous, and accessible through WeChat (Tencent) on smartphones. Device IDs were recorded to prevent duplicate entries. When caregivers started answering, the system marked them as "in progress." Although it was technically feasible to save "in progress" answers automatically, ethical concerns led to a decision not to do so, as these were not final submissions. The process of developing questionnaires and the methodological considerations for web-based surveys followed our team's established practices [40].

Determination of the Sample Size

Due to the lack of a clear sample size calculation method for multioutcome survey studies, we used the estimation method from quality-of-life studies, which suggests 5-10 questionnaires per question [41]. Therefore, with 50 questions, a minimum of 500 questionnaires were required. To mitigate potential selection bias [42], the minimum sample size was doubled to 1000. Further calculations took into account the number of households in Chongqing, with questionnaires planned to be collected at a 1:3 ratio between urban and suburban areas. As a result, a total of 4000 questionnaires were needed, with at least 1000 from urban areas and 3000 from suburban areas. Given the possibility of a high invalid response rate (approximately 50%) in public surveys [43], at least 8000 households were estimated to be needed. This sample size represented 0.4% of the entire population of kindergarten children in Chongqing, placing the sampling ratio between the national population dynamics survey (0.1%) and the sample survey (1%). Finally, the actual proportions of recruited participants from urban and suburban areas were 25.1% and 74.9%, respectively, which is consistent with the planned sample distribution.

Survey Process

The survey was conducted between June and December 2023. Prior to the official launch, a pilot test with 50 participants was conducted to validate the survey process. A flexible survey was then conducted in urban and suburban areas over 20 days to accommodate different kindergarten schedules. The questionnaire was distributed to kindergartens and forwarded to caregivers, ensuring that recruitment focused on the target audience and minimized nontarget responses. Forwarding questionnaires by kindergarten teachers was not mandatory. Caregivers could view a short introduction and an anonymous, voluntary declaration of the study on the questionnaire homepage to avoid any sense of obligation. After the survey, researchers downloaded and saved the data from the cloud servers and then removed them from the cloud.

Data Processing

The collected questionnaires were assessed for adherence to the inclusion and exclusion criteria. Specifically, 8600 online visits were documented during the survey, with 8266 (96.1%) identified as unique after eliminating duplicates. Among the visitors, 6834 started the questionnaire (response rate of 82.7%), and 5530 completed it (completion rate of 66.9%). After unreliable and nontarget questionnaires were excluded, 4231 questionnaires were ultimately analyzed (inclusion rate of 76.5%), representing 0.43% of the entire population of kindergarten children (Figure 1).

Figure 1. The inclusion and exclusion process.

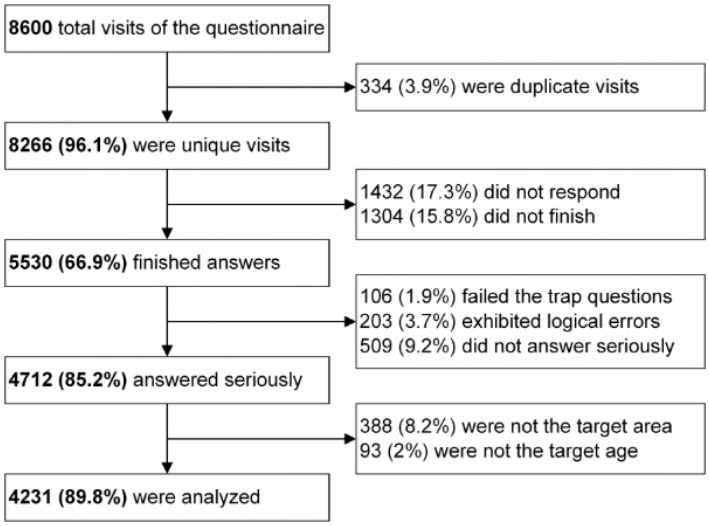

XSL•FO
RenderX

Statistical Analysis

Data analysis was performed using SPSS (version 25; IBM Corp). Qualitative data are presented as frequencies (percentages), while quantitative data are presented as medians (IQRs) after tests for normality. Intergroup comparisons were conducted using chi-square and Mann-Whitney tests. Multivariate logistic regression was used to adjust for potential confounders, including child age and sex, caregiver age, role, education level, and family size, when examining the relationship between region (independent variable) and public perspectives (outcome variables). Stepwise methods were not applied (Multimedia Appendix 2). The results of the logistic regression are presented as adjusted odds ratios (aORs) with 95% CIs. A *P* value less than .05 was considered statistically significant.

Results

General Characteristics

Table 1 presents the general characteristics of the children and caregivers. Among the children, 1064 (25.1%) resided in urban areas, while 3167 (74.9%) children were from suburban areas. Most children (n=3774, 89.2%) had never participated in pediatric research, while only 7.7% (n=326) and 5.4% (n=230) had taken part in offline and online research, respectively. Interestingly, suburban children had a greater proportion of online research participation than urban children (6% vs 3.9%; *P*=.008). Among the caregivers, 3547 (83.8%) were mothers, 2334 (55.2%) had an education level of senior high school or below, and 2297 (54.3%) lived in households with 4 or fewer members.

Table 1. General characteristics of the children and caregivers.

General characteristics	Overall	Urban	Suburban	*P* value
Age of the children (years), median (IQR)	4.8 (4.1-5.6)	4.6 (4.0-5.5)	4.9 (4.2-5.7)	<.001
Sex of the children, n (%)				.44
Male	2167 (51.2)	534 (50.2)	1633 (51.6)	
Female	2064 (48.8)	530 (49.8)	1534 (48.4)	
Ever participated in offline pediatric research, n (%)	326 (7.7)	72 (6.8)	254 (8)	.19
Ever participated in online pediatric research, n (%)	230 (5.4)	41 (3.9)	189 (6)	.008
Age of the caregivers (years), median (IQR)	33 (29-36)	33 (30-36)	32 (29-36)	.81
Role of the caregivers, n (%)				.005
Mother	3547 (83.8)	921 (86.6)	2626 (82.9)	
Father or others	684 (16.2)	143 (13.4)	541 (17.1)	
Education levels of the caregivers, n (%)				<.001
Senior high school or below	2334 (55.2)	455 (42.8)	1879 (59.3)	
College or above	1897 (44.8)	609 (57.2)	1288 (40.7)	
Number of people in the family, n (%)				<.001
2-4 people	2297 (54.3)	707 (66.4)	1590 (50.2)	
≥5 people	1934 (45.7)	357 (33.6)	1577 (49.8)	

Facilitators and Barriers to Participation

When questioned about their perspectives on research participation, 2531 (59.8%) caregivers expressed their first impressions, with 1561 (36.9%) caregivers inclined to participate and 970 (22.9%) caregivers inclined to decline. Suburban caregivers had more positive impressions than urban caregivers (38.6% vs 32%; *P*<.001; aOR 1.27, 95% CI 1.09-1.47). Additionally, caregivers with positive impressions reported higher rates of both offline (14.2% vs 2.9%; *P*<.001) and online (9.8% vs 2.7%; *P*<.001) research participation compared to those with negative impressions.

Table 2 presents the facilitators and barriers to participation, with most facilitators reporting rates lower than the barriers. The most significant facilitator was “be beneficial for children’s health” (n=3579, 84.6%), which was more common among urban caregivers (87.5% vs 83.6%; *P*<.001; aOR 1.45, 95% CI 1.17-1.79). Interestingly, although “be suggested by close people” was deemed less important (n=635, 15%), it had a higher rate among suburban caregivers (16.2% vs 11.5%; *P*<.001; aOR 1.47, 95% CI 1.18-1.79). On the other hand, the most significant barriers were “worry about being an experimental subject” (Guinea pig concerns: n=3060, 72.3%) and “pose a risk to children’s health” (n=2577, 60.9%). These barriers reflected a general apprehension about potential health risks. Furthermore, suburban caregivers had lower reporting rates for most barriers than urban caregivers, including “worry about being an experimental subject” (70.9% vs 76.6%; *P*<.001; aOR 0.79, 95% CI 0.67-0.93), “pose a risk to children’s health” (58.6% vs 67.8%; *P*<.001; aOR 0.71, 95% CI 0.61-0.83), “do not have enough background information” (55.2% vs 61.6%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89), and “worry about recommending other products” (48.2% vs 55%; *P*<.001; aOR 0.78, 95% CI 0.67-0.89).

XSL•FO
RenderX

Table 2. Facilitators and barriers to pediatric research participation.

Questions about facilitators and barriers	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Facilitators				
Be beneficial for children's health	3579 (84.6)	931 (87.5)	2648 (83.6)	<.001[a]
Trust the hospital and researchers	1454 (34.4)	371 (34.9)	1083 (34.2)	.69
Acknowledge the importance of research	1409 (33.3)	358 (33.6)	1051 (33.2)	.78
Increase contact with doctors	1189 (28.1)	295 (27.7)	894 (28.1)	.75
Have small burden on the participants	1139 (26.9)	323 (30.4)	816 (25.8)	.003[a]
Be driven by altruism	1044 (24.7)	261 (24.5)	783 (24.7)	.90
Be suggested by close people	635 (15)	122 (11.5)	513 (16.2)	<.001[a]
Have economic subsidies	390 (9.2)	100 (9.4)	290 (9.2)	.81
Barriers				
Worry about being an experimental subject	3060 (72.3)	815 (76.6)	2245 (70.9)	<.001[a]
Pose a risk to children's health	2577 (60.9)	721 (67.8)	1856 (58.6)	<.001[a]
Do not have enough background information	2404 (56.8)	655 (61.6)	1749 (55.2)	<.001[a]
Worry about personal privacy	2153 (50.9)	574 (53.9)	1579 (49.9)	.02
Worry about recommending other products	2112 (49.9)	585 (55)	1527 (48.2)	<.001[a]
Refuse procedures like venipuncture	1721 (40.7)	468 (44)	1253 (39.6)	.01
Cause disagreements among family members	1106 (26.1)	294 (27.6)	812 (25.6)	.20
Affect the purchase of medical insurance	803 (19)	201 (18.9)	602 (19)	.93

[a]Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Perspectives on Digital Medicine

When questioned about digital medicine (Table 3), 36.4% (n=1542) of caregivers reported following online health-related accounts, while 38.4% (n=1626) had used online information for health-related decisions. Additionally, 38.3% (n=1621) and 36.7% (n=1552) of caregivers recognized the increasing popularity of digital tools in future health care practices and supported their integration into pediatric research. Nearly half of caregivers (n=2073, 49%) were willing to collect data daily using digital tools, and 61.7% (n=2610) supported the use of gamification in pediatric research. However, only 25.2% (n=1068) of caregivers supported online-only research methods, with slightly more support from suburban caregivers (26% vs 22.9%; *P*=.045; aOR 1.19, 95% CI 1.01-1.41). When asked about concerns regarding smartphone data collection software, privacy was the main issue (n=3273, 77.4%), followed by concerns about potential software addiction (n=2457, 58.1%). Multivariate logistic regression revealed privacy concerns as the barrier to supporting digital tools in pediatric research (aOR 0.61, 95% CI 0.52-0.71).

XSL•FO
RenderX

Table 3. Perspectives on digital medicine.

Questions about digital medicine	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Have followed online health-related accounts	1542 (36.4)	410 (38.5)	1132 (35.7)	.10
Can use online information to make health-related decisions	1626 (38.4)	327 (30.7)	1299 (41)	<.001[a]
Knowing that digital tools will become popular in future health care practices	1621 (38.3)	411 (38.6)	1210 (38.2)	.81
Support digital tools in pediatric research	1552 (36.7)	372 (35)	1180 (37.3)	.18
Support gamification of pediatric research	2610 (61.7)	680 (63.9)	1930 (60.9)	.09
Support online-only research methods	1068 (25.2)	244 (22.9)	824 (26)	.045[a]
Be willing to use digital tools daily to collect data	2073 (49)	484 (45.5)	1589 (50.2)	.008
Concerns about using smartphone software in research				
Personal privacy issues	3273 (77.4)	855 (80.4)	2418 (76.3)	.007[a]
Possible addiction of the tools	2457 (58.1)	606 (57)	1851 (58.4)	.39
Transparency of data	2159 (51)	564 (53)	1595 (50.4)	.14
Health impacts of the tools (such as radiation)	1830 (43.3)	439 (41.3)	1391 (43.9)	.13
Possible expenses	1465 (34.6)	367 (34.5)	1098 (34.7)	.92
Ease of use	1279 (30.1)	320 (30.1)	959 (30.3)	.90

[a]Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Perspectives on Recruitment

Regarding recruitment (Table 4), 39.4% (n=1666) and 39.5% (n=1670) of caregivers expressed interest in reviewing recruitment information presented offline and online, respectively. Only 930 (22%) caregivers reported they would check both online and offline recruitment information. Notably, a greater proportion of suburban caregivers preferred checking offline recruitment information (41.2% vs 34%; *P*<.001; aOR 1.28, 95% CI 1.11-1.49). Additionally, 42.9% (n=1813) of caregivers expressed doubts about the reliability of online recruitment, while 58.1% (n=2457) doubted that research participation through online recruitment would lead to different treatments compared to offline methods.

When questioned about preferences, caregivers indicated that doctors with senior titles (n=2157, 51%) and specific researchers (n=2053, 48.5%) were the preferred recruiters, together accounting for 70.8% (n=2996) of caregivers. Research institution public accounts (n=3400, 80.4%) were the most favored source of online recruitment, with higher preferences among urban caregivers (85.9% vs 78.5%; *P*<.001; aOR 1.62, 95% CI 1.33-1.97). Although social media friends (n=984, 23.3%) and moments (n=608, 14.4%) were less popular sources for online recruitment, they were more frequently reported by suburban caregivers (friends: 24.7% vs 18.9%; *P*<.001; aOR 1.33, 95% CI 1.11-1.59; moments: 15.5% vs 11.1%; *P*<.001; aOR 1.35, 95% CI 1.09-1.67). Finally, the majority of caregivers (n=2963, 70%) preferred cartoon-style recruitment advertisements.

XSL•FO
RenderX

Table 4. Perspectives on recruitment.

Questions about recruitment	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Know that informed consent is required before beginning	3549 (83.9)	900 (84.6)	2649 (83.6)	.47
Be willing to review offline recruitment information	1666 (39.4)	362 (34)	1304 (41.2)	<.001[a]
Be willing to review online recruitment information	1670 (39.5)	408 (38.3)	1262 (39.8)	.39
Doubt the reliability of online recruitment information	1813 (42.9)	474 (44.5)	1339 (42.3)	.20
Believe that the treatment between online and offline recruitment is different	2457 (58.1)	603 (56.7)	1854 (58.5)	.29
Preferred role of recruiter				
Doctors (senior title)	2157 (51)	549 (51.6)	1608 (50.8)	.64
Specific researchers	2053 (48.5)	552 (51.9)	1501 (47.4)	.01
Medical students	786 (18.6)	188 (17.7)	598 (18.9)	.38
Doctors (ordinary title)	459 (10.8)	130 (12.2)	329 (10.4)	.10
Nurses	294 (6.9)	75 (7)	219 (6.9)	.88
Preferred source of online recruitment information				
Public accounts of research institutions	3400 (80.4)	914 (85.9)	2486 (78.5)	<.001[a]
Friends on social media	984 (23.3)	201 (18.9)	783 (24.7)	<.001[a]
Moments on social media	608 (14.4)	118 (11.1)	490 (15.5)	<.001[a]
Prefer cartoon version of the recruitment information	2963 (70)	752 (70.7)	2211 (69.8)	.60

[a]Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Perspectives on the Research Process

When questioned about the research process (Table 5), 1445 (34.2%) caregivers reported that they knew they could withdraw from the study, and 79.8% (n=1153) of them would consult with the research team before doing so. Regarding follow-up, 3076 (72.7%) caregivers reported that regular contact was necessary. Telephones (n=1916, 62.3%) were the preferred method, followed by social media apps (n=1801, 58.6%). Social media apps were more common among urban participants (64.7% vs 56.4%; *P*<.001; aOR 1.39, 95% CI 1.17-1.65). For research feedback, 4017 (94.9%) caregivers expressed willingness to receive it, with social media apps (n=2624, 65.3%) being the preferred method, followed by email (n=2535, 63.1%). Additionally, we questioned at the beginning and end of the questionnaire whether caregivers were willing to provide phone numbers in anonymous surveys; a total of 3065 (72.4%) caregivers expressed willingness in offline surveys, a rate significantly higher than that in online surveys (72.4% vs 43.4%; *P*<.001).

XSL•FO
RenderX

Table 5. Perspectives on the research process.

Questions about research process	Overall, n (%)	Urban, n (%)	Suburban, n (%)	*P* value
Know that withdrawing midway is unconditionally allowed	1445 (34.2)	366 (34.4)	1079 (34.1)	.85
Would discuss with the research team before withdrawing	1153 (79.8)	300 (82.0)	853 (79.1)	.23
Believe that regular contact is necessary	3076 (72.7)	785 (73.8)	2291 (72.3)	.36
Preferred method (telephone)	1916 (62.3)	514 (65.5)	1402 (61.2)	.03
Preferred method (social media apps)	1801 (58.6)	508 (64.7)	1293 (56.4)	<.001[a]
Preferred method (video)	715 (23.2)	199 (25.4)	516 (22.5)	.11
Be willing to receive research feedback	4017 (94.9)	1024 (96.2)	2993 (94.5)	.03[a]
Preferred route (social media apps)	2624 (65.3)	687 (67.1)	1937 (64.7)	.17
Preferred route (mail)	2535 (63.1)	694 (67.8)	1841 (61.5)	<.001[a]
Preferred route (telephone)	2071 (51.6)	457 (44.6)	1614 (53.9)	<.001[a]
Be willing to provide phone number in anonymous surveys				
Offline surveys	3065 (72.4)	757 (71.1)	2308 (72.9)	.27
Online surveys	1835 (43.4)	452 (42.5)	1383 (43.7)	.50

[a]Confirmed by multivariate logistic regression that adjusted general characteristics (Multimedia Appendix 2).

Discussion

Principal Findings

This study investigated public perspectives on the digitalization of pediatric research, comparing views between urban and suburban areas. The findings indicate that the current support for digital approaches in pediatric research is suboptimal but still offers valuable strategies for using digital methods to reduce urban-suburban health disparities in children. This study is relatively rare in the literature and explores innovative ways to enhance participation and equity through digital approaches. Previous studies have emphasized the importance of understanding the barriers and facilitators that parents face when considering their children's participation in research [20-27]. As much of the research has already begun digitalization [28-32], it is beneficial to periodically pause and gather insights from participants by considering their perspectives.

Perspectives on Pediatric Research

A previous study on neonatal clinical trials revealed that one-quarter of parents made immediate enrollment decisions [27]. Our research found that 59.8% (n=2531) of caregivers formed initial impressions of pediatric research, with 36.9% (n=1561) expressing positive inclinations and 22.9% (n=970) expressing negative inclinations. These findings emphasize the importance of recruitment strategies during initial contact. Further analysis identified health benefits as the main facilitators and risks as the primary barriers, which aligns with previous studies from other regions [20]. This suggests the results could be applicable beyond Chongqing. The lower reporting rates of facilitators than barriers may explain challenges in pediatric research. Preferences in recruitment leaned toward doctors with senior titles and specific researchers as recruiters, emphasizing the influence of authority on caregivers' decision-making. Additionally, previous research has shown that cartoon images influence children's decision-making [44]. This study found that 70% (n=2963) of caregivers also preferred cartoon-style recruitment materials, possibly because they convey a sense of care for children, which resonates well with caregivers [45].

Perspectives on Digital Medicine

Public perspectives on digital medicine revealed suboptimal acceptance among caregivers. Approximately one-third of caregivers acknowledged the increasing popularity of digital tools and supported their application in pediatric research. This was reflected in a lower willingness to provide phone numbers in online surveys than in offline surveys. The transition from traditional to remote trials, particularly during the COVID-19 pandemic, has been successful [30,46], though some unsatisfactory outcomes have been reported [47,48]. Our results showed that public accounts from research institutions were the most preferred source of online recruitment, aligning with offline sources that emphasize authority. Given that 42.9% (n=1813) and 58.1% (n=2457) of caregivers doubted the reliability and consistency of online recruitment, establishing a credible public account by research institutions could enhance recruitment success on social media, a novel strategy compared to previous studies [49]. On the other hand, the most commonly mentioned methods for regular contact were telephone and social media apps. This may explain the effectiveness of the hybrid follow-up model, which encourages using both phone calls and social media [50]. However, only 25.2% (n=1068) of caregivers supported online-only research methods, highlighting the importance of face-to-face interactions. This raises important questions about the effectiveness of a fully remote model in postpandemic pediatric research. When questioned about gamification in research, 61.7% (n=2610) of caregivers were in favor. Studies on gamified medicine have shown positive outcomes [51,52], but concerns about potential risks remain [53,54]. In this study, 58.1% (n=2457) of caregivers expressed concerns about addiction. Additionally, 77.4% (n=3273) of

XSL•FO
RenderX

caregivers expressed concerns about personal privacy when using smartphone software for research, which served as a barrier to supporting digital tools in pediatric research. Another advantage of digitalizing pediatric research is the availability of personalized real-time feedback, with 94.9% (n=4017) of participants willing to receive it. This meets caregivers' needs and serves as a strategy to increase motivation [55].

Differences Between Urban and Suburban Areas

When comparing urban and suburban areas, demographic differences were not adjusted using methods like propensity score matching, as these differences were inherent to the population. Instead, multivariate logistic regression was used to validate the results. Interestingly, suburban caregivers expressed greater interest in pediatric research and fewer barriers and concerns. These findings contradict common knowledge [15,16,56] but align with a previous study on congenital heart disease [57]. Another recent study also showed that rural caregivers had lower concerns about data privacy and security [58]. A possible explanation for suburban caregivers' higher enthusiasm is that the suburban areas in this study may not accurately represent resource-poor regions. These caregivers demonstrate sufficient health awareness and capabilities, making them value participation in pediatric research. This is evident in their willingness to access offline recruitment information, use digital tools for data collection, and support online-only research methods. Since suburban areas usually have more kindergarten children than urban areas (eg, Chongqing [33]), they represent a valuable potential pool of potential research participants. Further analysis revealed that suburban caregivers more frequently mentioned suggestions from close people. Similarly, suburban caregivers more often cited social media friends and moments as sources of online recruitment information. While these results raise concerns about caregivers' ability to assess the quality of online health information [59], they also highlight the potential for incorporating informal channels in recruitment strategies for suburban participants [60].

Strengths and Limitations

This study has several strengths. First, this study examined perspectives on digital medicine, which is distinct from conventional pediatric research topics [16,20-27] and provides essential data for the digital age. Second, this study included a more diverse sample from the general public, offering a broader representation compared to studies focused on specific illnesses. Third, the sample size was larger than that in previous studies, likely providing a more population-based overview. Finally, we performed a comparative analysis between urban and suburban areas, contributing to the understanding and mitigation of health inequalities among children in different geographical locations.

Despite these strengths, the limitations require careful consideration. First, this study was conducted in Chongqing, China. It is important to consider this when generalizing the findings to other regions. Therefore, further studies across different regions or countries are recommended to expand these findings. Second, although comparisons were made between urban and suburban areas, conditions in rural areas remain unknown and require further investigation. Third, perspectives were assessed using a structured questionnaire but lacked qualitative interview data. Additionally, despite cultural adaptations and a preliminary survey, individuals may still interpret certain questions differently based on their personal experiences. Finally, although smartphones are a prominent tool in digital health, the study's focus on them may introduce bias. However, less common digital tools may not be widely known to the general public, making it difficult to survey participants effectively about unfamiliar technologies.

Conclusions

In the digital age, approximately one-third of caregivers support pediatric research and endorse the use of digital tools in such studies. Caregivers in suburban areas show greater enthusiasm for research and digital medicine. Integrating official and unofficial social media recruitment strategies, implementing a hybrid online-offline follow-up approach, and addressing privacy concerns could help researchers increase participant recruitment and retention rates.

Acknowledgments

YH was affiliated with the Children's Hospital of Chongqing Medical University at the time of the study and is currently affiliated with Beijing Children's Hospital. The authors would like to thank Drs Shuyao Zhu, Keran Ling, Lin Chen, Zehao Li, Wei Zhang, Liyu Chen, and Yuxin Lv for their contributions to revising and culturally adapting the questionnaire. Special thanks are also extended to the kindergartens, teachers, and caregivers for their cooperation in distributing and responding to the questionnaire. This study was supported by the Major Program of the National Clinical Research Center for Child Health and Disorders in Children's Hospital of Chongqing Medical University (NCRCCHD-2023-MP-01) and the Program for Youth Innovation in Future Medicine, Chongqing Medical University, "Clinical Innovation Team for Standardized Management of Food Allergy in Children" (2022-26-W0120). The funder had no role in the preparation, review, or approval of the manuscript or the decision to submit it for publication. No generative artificial intelligence was used to generate the manuscript content.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

HF and RX contributed equally to this work and should be considered joint first authors. YZ and YH contributed equally to this work and should be considered joint corresponding authors. HF and RX conceptualized and designed the study, designed the data collection instruments, carried out the initial analyses, and drafted the initial manuscript. JL designed the data collection instruments, collected the data, carried out the initial analyses, and reviewed and revised the manuscript. YL and EL designed the data collection instruments, obtained funding, and reviewed and revised the manuscript. YZ and YH designed the data collection instruments, coordinated and supervised the data collection, obtained funding, and reviewed and revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire (in Chinese).
[PDF File (Adobe PDF File), 248 KB - jmir_v27i1e60324_app1.pdf]

Multimedia Appendix 2
Multivariate logistic regression.
[PDF File (Adobe PDF File), 172 KB - jmir_v27i1e60324_app2.pdf]

References

1. Wyatt TH, Li X, Fancher S, Mitoubsi AS, Pardue J. Recruitment barriers of an mHealth pediatric asthma pilot study. West J Nurs Res 2024;46(2):152-159. [doi: 10.1177/01939459231217919] [Medline: 38112079]
2. Niemeyer L, Mechler K, Buitelaar J, Durston S, Gooskens B, Oranje B, et al. 'Include me if you can'—reasons for low enrollment of pediatric patients in a psychopharmacological trial. Trials 2021;22(1):178 [FREE Full text] [doi: 10.1186/s13063-021-05119-6] [Medline: 33648579]
3. Fischer HS, Bührer C, Czernik C. Hazards to avoid in future neonatal studies of nasal high-frequency oscillatory ventilation: lessons from an early terminated trial. BMC Res Notes 2019;12(1):237 [FREE Full text] [doi: 10.1186/s13104-019-4268-2] [Medline: 31023363]
4. Pica N, Bourgeois F. Discontinuation and nonpublication of randomized clinical trials conducted in children. Pediatrics 2016;138(3):e20160223 [FREE Full text] [doi: 10.1542/peds.2016-0223] [Medline: 27492817]
5. Fang H, Chen L, Li J, Ren L, Yin Y, Chen D, et al. A web-based instrument for infantile atopic dermatitis identification (electronic version of the Modified Child Eczema Questionnaire): development and implementation. J Med Internet Res 2023;25:e44614 [FREE Full text] [doi: 10.2196/44614] [Medline: 37467020]
6. Fang H, Li Z, Xian R, Yin Y, Wang J, Guo H, et al. Early life growth and developmental trajectory in children with biliary atresia undergoing primary liver transplantation. Front Pediatr 2023;11:1198360 [FREE Full text] [doi: 10.3389/fped.2023.1198360] [Medline: 37377760]
7. Di Cesare M, Sorić M, Bovet P, Miranda JJ, Bhutta Z, Stevens GA, et al. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med 2019;17(1):212 [FREE Full text] [doi: 10.1186/s12916-019-1449-8] [Medline: 31760948]
8. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr 2021;175(11):1142-1150 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.2482] [Medline: 34369987]
9. Christakis DA. Internet addiction: a 21st century epidemic? BMC Med 2010;8(1):61 [FREE Full text] [doi: 10.1186/1741-7015-8-61] [Medline: 20955578]
10. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiatry 2018;5(2):175-186. [doi: 10.1016/S2215-0366(17)30167-0] [Medline: 29033005]
11. Koren G. Healthy children as subjects in pharmaceutical research. Theor Med Bioeth 2003;24(2):149-159. [doi: 10.1023/a:1024694828858] [Medline: 12943269]
12. Postma L, Luchtenberg ML, Verhagen AAE, Maeckelberghe ELM. The attitudes of healthy children and researchers towards the challenges of involving children in research: an exploratory study. Res Involve Engage 2021;7(1):30. [doi: 10.1186/s40900-021-00263-4] [Medline: 33990230]
13. Di Carlo C, Mighton C, Clausen M, Joshi E, Casalino S, Kim THM, et al. Parents' attitudes towards research involving genome sequencing of their healthy children: a qualitative study. Eur J Hum Genet 2024;32(2):171-175. [doi: 10.1038/s41431-023-01476-9] [Medline: 37864046]
14. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: a CALIPER white paper. Crit Rev Clin Lab Sci 2017;54(6):358-413. [doi: 10.1080/10408363.2017.1379945] [Medline: 29017389]

XSL•FO
RenderX

15. Tanner A, Kim SH, Friedman DB, Foster C, Bergeron CD. Barriers to medical research participation as perceived by clinical trial investigators: communicating with rural and African American communities. J Health Commun 2015;20(1):88-96. [doi: 10.1080/10810730.2014.908985] [Medline: 25204763]
16. Watson SE, Smith P, Snowden J, Vaughn V, Cottrell L, Madden CA, et al. Facilitators and barriers to pediatric clinical trial recruitment and retention in rural and community settings: a scoping review of the literature. Clin Transl Sci 2022;15(4):838-853 [FREE Full text] [doi: 10.1111/cts.13220] [Medline: 35037409]
17. Chen Y, Sylvia S, Dill SE, Rozelle S. Structural determinants of child health in rural China: the challenge of creating health equity. Int J Environ Res Public Health 2022;19(21):13845 [FREE Full text] [doi: 10.3390/ijerph192113845] [Medline: 36360724]
18. Wu Q, Pan J. The trends of health equity in children aged 5-12 years and its determinants in China from 1991 to 2015 [in Chinese]. J Preventive Med Inf 2020;36(5):633-640.
19. COMMITTEE ON BIOETHICS. Informed consent in decision-making in pediatric practice. Pediatrics 2016;138(2):e20161484. [doi: 10.1542/peds.2016-1484] [Medline: 27456514]
20. Nathe JM, Oskoui TT, Weiss EM. Parental views of facilitators and barriers to research participation: systematic review. Pediatrics 2023;151(1):e2022058067 [FREE Full text] [doi: 10.1542/peds.2022-058067] [Medline: 36477217]
21. Engster SA, Fascetti C, Daw K, Reis EC. Parent perceptions of and preferences for participation in child health research: results from a pediatric practice-based research network. J Am Board Fam Med 2019;32(5):685-694 [FREE Full text] [doi: 10.3122/jabfm.2019.05.190030] [Medline: 31506364]
22. Tromp K, Zwaan CM, van de Vathorst S. Motivations of children and their parents to participate in drug research: a systematic review. Eur J Pediatr 2016;175(5):599-612 [FREE Full text] [doi: 10.1007/s00431-016-2715-9] [Medline: 27041121]
23. Sollo N, Ahlers-Schmidt CR, Davis AM, Smith TR, Dedeaux JA, McCulloh RJ. Perceived barriers to clinical trials participation: a survey of pediatric caregivers. Kans J Med 2022;15:139-143 [FREE Full text] [doi: 10.17161/kjm.vol15.16220] [Medline: 35646248]
24. Tu WJ, Li Y, Shi XD. The views of Chinese parents on research participation and informed consent. J Paediatr Child Health 2012;48(2):186-187. [doi: 10.1111/j.1440-1754.2011.02404.x] [Medline: 22320277]
25. McKenzie PL, Siegel DH, Cullen D, Castelo-Soccio L. Strategies to enhance pediatric clinical trial participation: a systematic review with narrative synthesis. Pediatr Dermatol 2021;38(6):1515-1522. [doi: 10.1111/pde.14817] [Medline: 34647357]
26. Weiss EM, Olszewski AE, Guttmann KF, Magnus BE, Li S, Shah AR, et al. Parental factors associated with the decision to participate in a neonatal clinical trial. JAMA Netw Open 2021;4(1):e2032106 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.32106] [Medline: 33433595]
27. Weiss EM, Guttmann KF, Olszewski AE, Magnus BE, Li S, Kim SY, et al. Parental enrollment decision-making for a neonatal clinical trial. J Pediatr 2021;239:143-149.e3 [FREE Full text] [doi: 10.1016/j.jpeds.2021.08.014] [Medline: 34400207]
28. Harmon DM, Noseworthy PA, Yao X. The digitization and decentralization of clinical trials. Mayo Clin Proc 2023;98(10):1568-1578. [doi: 10.1016/j.mayocp.2022.10.001] [Medline: 36669937]
29. Marra C, Chen JL, Coravos A, Stern AD. Quantifying the use of connected digital products in clinical research. NPJ Digital Med 2020;3:50 [FREE Full text] [doi: 10.1038/s41746-020-0259-x] [Medline: 32285011]
30. Portnoy JM, Waller M, De Lurgio S, Dinakar C. Telemedicine is as effective as in-person visits for patients with asthma. Ann Allergy Asthma Immunol 2016;117(3):241-245. [doi: 10.1016/j.anai.2016.07.012] [Medline: 27613456]
31. Alnasser Y, Proaño A, Loock C, Chuo J, Gilman RH. Telemedicine and pediatric care in rural and remote areas of middle-and-low-income countries: narrative review. J Epidemiol Glob Health 2024;14(3):779-786 [FREE Full text] [doi: 10.1007/s44197-024-00214-8] [Medline: 38478166]
32. Tan RKJ, Wu D, Day S, Zhao Y, Larson HJ, Sylvia S, et al. Digital approaches to enhancing community engagement in clinical trials. NPJ Digital Med 2022;5(1):37 [FREE Full text] [doi: 10.1038/s41746-022-00581-1] [Medline: 35338241]
33. Chongqing statistical yearbook 2022. Chongqing Municipal Statistics Bureau. URL: https://tjj.cq.gov.cn/zwgk_233/tjnj [accessed 2022-12-26]
34. Kindergarten work regulations. Ministry of Education of the People's Republic of China. URL: http://www.moe.gov.cn/jyb_xxgk/xxgk/zhengce/guizhang/202112/t20211206_585104.html [accessed 2016-01-05]
35. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
36. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
37. Jin H, Cui M, Liu J. Factors affecting people's attitude toward participation in medical research: a systematic review. Curr Med Res Opin 2020;36(7):1137-1143. [doi: 10.1080/03007995.2020.1760807] [Medline: 32329364]
38. El Achi D, Al Hakim L, Makki M, Mokaddem M, Khalil PA, Kaafarani BR, et al. Perception, attitude, practice and barriers towards medical research among undergraduate students. BMC Med Educ 2020;20(1):195 [FREE Full text] [doi: 10.1186/s12909-020-02104-6] [Medline: 32552801]

XSL•FO
RenderX

39. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Making 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
40. Fang H, Lv Y, Chen L, Zhang X, Hu Y. The current knowledge, attitudes, and practices of the neglected methodology of web-based questionnaires among Chinese health workers: web-based questionnaire study. J Med Internet Res 2023;25:e41591 [FREE Full text] [doi: 10.2196/41591] [Medline: 36533302]
41. Lu Y, Fang J, Tian L, Jin H. Advanced Medical Statistics. Singapore: World Scientific Publishing CO Pte Ltd; 2015:1472.
42. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb) 2021;31(1):010502 [FREE Full text] [doi: 10.11613/BM.2021.010502] [Medline: 33380887]
43. Blumenberg C, Barros AJD. Response rate differences between web and alternative data collection methods for public health research: a systematic review of the literature. Int J Public Health 2018;63(6):765-773. [doi: 10.1007/s00038-018-1108-4] [Medline: 29691594]
44. Tzeng YF, Gau BS. Suitability of asthma education materials for school-age children: Implications for health literacy. J Clin Nurs 2018;27(5-6):e921-e930. [doi: 10.1111/jocn.13993] [Medline: 28793368]
45. Levetown M, American Academy of Pediatrics Committee on Bioethics. Communicating with children and families: from everyday interactions to skill in conveying distressing information. Pediatrics 2008;121(5):e1441-e1460. [doi: 10.1542/peds.2008-0565] [Medline: 18450887]
46. Simmons LA, Phipps JE, Whipps M, Smith P, Carbajal KA, Overstreet C, et al. From hybrid to fully remote clinical trial amidst the COVID-19 pandemic: strategies to promote recruitment, retention, and engagement in a randomized mHealth trial. Digit Health 2022;8:20552076221129065 [FREE Full text] [doi: 10.1177/20552076221129065] [Medline: 36185388]
47. Zapata KA, Wang-Price SS, Fletcher TS, Johnston CE. Factors influencing adherence to an app-based exercise program in adolescents with painful hyperkyphosis. Scoliosis Spinal Disord 2018;13:11 [FREE Full text] [doi: 10.1186/s13013-018-0159-x] [Medline: 30027121]
48. Haynes SC, Kamerman-Kretzmer R, Khan SS, Crossen S, Lieng MK, Marcin JP, et al. Telemedicine use for pediatric asthma care: a mixed methods study. J Asthma 2022;59(12):2431-2440. [doi: 10.1080/02770903.2021.2019265] [Medline: 34913803]
49. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022;24(8):e38015 [FREE Full text] [doi: 10.2196/38015] [Medline: 35925655]
50. Qiao S, Tang L, Zhang W, Tian S, Liu M, Yang L, et al. Nurse-led follow-up to outpatients with cancer pain treated with opioids at home-telephone calls plus WeChat versus telephone calls only: a quasi-experimental study. Patient Prefer Adherence 2019;13:923-931 [FREE Full text] [doi: 10.2147/PPA.S203900] [Medline: 31239650]
51. Pimentel-Ponce M, Romero-Galisteo RP, Palomo-Carrión R, Pinero-Pinto E, Merchán-Baeza JA, Ruiz-Muñoz M, et al. Gamification and neurological motor rehabilitation in children and adolescents: a systematic review. Neurologia (Engl Ed) 2024;39(1):63-83 [FREE Full text] [doi: 10.1016/j.nrleng.2023.12.006] [Medline: 38065433]
52. Suleiman-Martos N, García-Lara RA, Membrive-Jiménez MJ, Pradas-Hernández L, Romero-Béjar JL, Dominguez-Vías G, et al. Effect of a game-based intervention on preoperative pain and anxiety in children: a systematic review and meta-analysis. J Clin Nurs 2022;31(23-24):3350-3367 [FREE Full text] [doi: 10.1111/jocn.16227] [Medline: 35075716]
53. Boendermaker WJ, Maceiras SS, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
54. Holtz BE, Murray K, Park T. Serious games for children with chronic diseases: a systematic review. Games Health J 2018;7(5):291-301. [doi: 10.1089/g4h.2018.0024] [Medline: 29957082]
55. Wu Q, Wang X, Zhang J, Zhang Y, van Velthoven MH. The effectiveness of a WeChat-based self-assessment with a tailored feedback report on improving complementary feeding and movement behaviour of children aged 6-20 months in rural China: a cluster randomized controlled trial. Lancet Reg Health West Pac 2023;37:100796 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100796] [Medline: 37273963]
56. Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, et al. eHealth tools that assess and track health and well-being in children and young people: systematic review. J Med Internet Res 2022;24(5):e26015 [FREE Full text] [doi: 10.2196/26015] [Medline: 35550285]
57. Lin WH, Chen YK, Lin SH, Cao H, Chen Q. Parents' understanding and attitudes toward the use of the WeChat platform for postoperative follow-up management of children with congenital heart disease. J Cardiothorac Surg 2023;18(1):66 [FREE Full text] [doi: 10.1186/s13019-023-02153-0] [Medline: 36755263]
58. Chow E, Virani A, Pinkney S, Abdulhussein FS, van Rooij T, Görges M, et al. Caregiver and youth characteristics that influence trust in digital health platforms in pediatric care: mixed methods study. J Med Internet Res 2024;26:e53657 [FREE Full text] [doi: 10.2196/53657] [Medline: 39467279]
59. Li P, Chen B, Deveaux G, Luo Y, Tao W, Li W, et al. Cross-verification of COVID-19 information obtained from unofficial social media accounts and associated changes in health behaviors: web-based questionnaire study among Chinese netizens. JMIR Public Health Surveill 2022;8(5):e33577 [FREE Full text] [doi: 10.2196/33577] [Medline: 35486529]

XSL•FO
RenderX

60. Vos SC, Adatorwovor R, Roberts MK, Sherman DL, Bonds D, Dunfee MN, et al. Community engagement through social media: a promising low-cost strategy for rural recruitment? J Rural Health 2024;40(3):467-475. [doi: 10.1111/jrh.12809] [Medline: 37985592]

Abbreviations

aOR: adjusted odds ratio
CHERRIES: Checklists for Reporting Results of Internet E-Surveys
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by T de Azevedo Cardoso; submitted 07.05.24; peer-reviewed by L Giacomelli, SL Coven; comments to author 09.09.24; revised version received 15.09.24; accepted 21.11.24; published 07.01.25.

Please cite as:
Fang H, Xian R, Li J, Li Y, Liu E, Zhao Y, Hu Y
Urban-Suburban Differences in Public Perspectives on Digitalizing Pediatric Research: Cross-Sectional Survey Study
J Med Internet Res 2025;27:e60324
URL: https://www.jmir.org/2025/1/e60324
doi:10.2196/60324
PMID:

©Heping Fang, Ruoling Xian, Juan Li, Yingcun Li, Enmei Liu, Yan Zhao, Yan Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Recruiting Young People for Digital Mental Health Research: Lessons From an AI-Driven Adaptive Trial

Wu Yi Zheng[1], PhD; Artur Shvetcov[1], PhD; Aimy Slade[1], PhD; Zoe Jenkins[1], BDes; Leonard Hoon[2], PhD; Alexis Whitton[1], PhD; Rena Logothetis[2], PhD; Smrithi Ravindra[1], BPsych; Stefanus Kurniawan[2], MIT; Sunil Gupta[2], PhD; Kit Huckvale[3], PhD; Eileen Stech[1], PhD; Akash Agarwal[2], MIT; Joost Funke Kupper[2], BCS; Stuart Cameron[2], BAppSc; Jodie Rosenberg[1], MCP; Nicholas Manoglou[2], BCS; Manisha Senadeera[2], BEng; Svetha Venkatesh[2], PhD; Kon Mouzakis[2], PhD; Rajesh Vasa[2], PhD; Helen Christensen[1], PhD; Jill M Newby[1], PhD

[1]Black Dog Institute, University of New South Wales, Sydney, Australia

[2]Applied Artificial Intelligence Institute, Deakin University, Melbourne, Australia

[3]Centre for Digital Transformation of Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Wu Yi Zheng, PhD
Black Dog Institute
University of New South Wales
Hospital Road, Randwick
Sydney, 2031
Australia
Phone: 61 0422510718
Email: wuyi.zheng@blackdog.org.au

Abstract

Background: With increasing adoption of remote clinical trials in digital mental health, identifying cost-effective and time-efficient recruitment methodologies is crucial for the success of such trials. Evidence on whether web-based recruitment methods are more effective than traditional methods such as newspapers, media, or flyers is inconsistent. Here we present insights from our experience recruiting tertiary education students for a digital mental health artificial intelligence–driven adaptive trial—Vibe Up.

Objective: We evaluated the effectiveness of recruitment via Facebook and Instagram compared to traditional methods for a treatment trial and compared different recruitment methods' retention rates. With recruitment coinciding with COVID-19 lockdowns across Australia, we also compared the cost-effectiveness of social media recruitment during and after lockdowns.

Methods: Recruitment was completed for 2 pilot trials and 6 minitrials from June 2021 to May 2022. To recruit participants, paid social media advertising on Facebook and Instagram was used, alongside mailing lists of university networks and student organizations or services, media releases, announcements during classes and events, study posters or flyers on university campuses, and health professional networks. Recruitment data, including engagement metrics collected by Meta (Facebook and Instagram), advertising costs, and Qualtrics data on recruitment methods and survey completion rates, were analyzed using RStudio with R (version 3.6.3; R Foundation for Statistical Computing).

Results: In total, 1314 eligible participants (aged 22.79, SD 4.71 years; 1079, 82.1% female) were recruited to 2 pilot trials and 6 minitrials. The vast majority were recruited via Facebook and Instagram advertising (n=1203; 92%). Pairwise comparisons revealed that the lead institution's website was more effective in recruiting eligible participants than Facebook (z=3.47; P=.003) and Instagram (z=4.23; P<.001). No differences were found between recruitment methods in retaining participants at baseline, at midpoint, and at study completion. Wilcoxon tests found significant differences between lockdown (pilot 1 and pilot 2) and postlockdown (minitrials 1-6) on costs incurred per link click (lockdown: median Aus $0.35 [US $0.22], IQR Aus $0.27-$0.47 [US $0.17-$0.29]; postlockdown: median Aus $1.00 [US $0.62], IQR Aus $0.70-$1.47 [US $0.44-$0.92]; W=9087; P<.001) and the amount spent per hour to reach the target sample size (lockdown: median Aus $4.75 [US $2.95], IQR Aus $1.94-6.34 [US $1.22-$3.97]; postlockdown: median Aus $13.29 [US $8.26], IQR Aus $4.70-25.31 [US $2.95-$15.87]; W=16044; P<.001).

Conclusions: Social media advertising via Facebook and Instagram was the most successful strategy for recruiting distressed tertiary students into this artificial intelligence–driven adaptive trial, providing evidence for the use of this recruitment method for this type of trial in digital mental health research. No recruitment method stood out in terms of participant retention. Perhaps

XSL•FO
RenderX

a reflection of the added distress experienced by young people, social media recruitment during the COVID-19 lockdown period was more cost-effective.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12621001092886; https://tinyurl.com/39f2pdmd; Australian New Zealand Clinical Trials Registry ACTRN12621001223820; https://tinyurl.com/bdhkvucv

(J Med Internet Res 2025;27:e60413) doi:10.2196/60413

KEYWORDS

recruitment; Facebook; retention, COVID-19; artificial intelligence

Introduction

Background

Advances in digital clinical trial infrastructure have made fully remote clinical trials more commonplace, particularly in the field of digital mental health. In such trials, participants are screened, assessed, treated, and followed up entirely via remote digital methods. Pivotal to the scientific validity of these trials is the timely and adequate recruitment of target participants who meet a set of prespecified eligibility criteria. Due to the paucity of research into best practice methods for efficiently recruiting trial samples, in this paper, we present insights from our experiences recruiting young participants for a digital mental health trial that used an artificial intelligence (AI)–driven adaptive trial design. This type of trial design has the potential to minimize time, costs, and the sample size needed to achieve sufficient statistical power [1]. Additionally, this method facilitates the optimization of treatment plans, which can enhance adherence and lower attrition [2]. However, compared to traditional randomized controlled trials (RCTs), AI-adaptive trials face an increased risk of insufficient and untimely enrollment due to the need for frequent, short-spanned recruitment drives (eg, monthly recruitment drives over a year). Hence, identifying recruitment methodologies that are both cost-effective and time-efficient is crucial for the success of such trials.

What recruitment methods are currently used and how do they compare with each other? A review by Lane et al [3] on online recruitment methods for web-based and mHealth studies found that successful recruitment of participants into a trial varied widely depending on the population, budget, intervention, costs, and study design. Of the studies included in their review, less than half (42%) found Facebook advertisements to be an effective method of recruitment and a quarter found Google advertisements to be the most effective method [3]. Only 1 study reported better outcomes using traditional methods of recruitment (eg, print advertisements and flyers) [3]. Overall, there was no consistent evidence on whether web-based recruitment methods were more effective than traditional methods such as newspapers, media, or flyers [3]. Similar mixed findings were echoed by a systematic review of recruitment strategies in mental health clinical trials [4], and a scoping review that examined the role of social media in clinical trials recruitment [5].

Specific to online recruitment methods, there is also contradictory evidence regarding the most cost-effective platform. Morgan et al [6] used a range of internet-based methods to recruit for the Mood Memos depression prevention study, including Facebook, emails, search engine advertisements, and forum posts, and reported that advertising on Google was more cost-effective than Facebook advertising (Aus $12 [US $7.46] per recruited participant vs Aus $19.89 [US $12.37] per recruited participant) in signing up participants. Similarly, recruitment costs for a large RCT of web-based smoking interventions were highest for Facebook (US $40.51) per randomized participant, compared to Google (US $34.71), traditional sources such as press releases, newspaper, radio, and television interviews (US $20.30), and using an online survey panel company (US $13.95) [7]. In contrast, Facebook advertising was found to be more cost-effective for recruiting a large sample of general community participants for suicide prevention research (Aus $2.81 [US $1.75] per completed survey) [8]. Similarly, Batterham [9] reported that advertising on a Facebook page that featured links to a survey on suicidal ideation was more effective in yielding completions (Aus $1.51 [US $0.94] per completion), compared to advertisements linked to an external website (Aus $9.82 [US $6.11] per completion). Both were more cost-effective than postal surveys, which averaged Aus $19.10 [US $11.88] per completion [9]. In a review of 176 studies that used social media for mental health research recruitment, Facebook was found to be the most preferred platform, at a median cost of US $19.47 per final recruited study participant [10]. The authors also found that in 2 out of 3 studies included in the review, social media recruitment performed as well as or better than traditional methods in the number and cost of final enrolled participants [10].

Differences in effectiveness between recruitment methods can often be difficult to interpret. In their systematic review, Liu et al [4] found that it was difficult to assess the overall effectiveness of any particular recruitment strategy as some strategies that worked well for a particular population may not work as well for others. For example, the greater costs of Facebook advertising for the Mood Memos study relative to the other studies described may be due to study requirements to target adults with subthreshold depression symptoms as opposed to recruiting a nonclinical sample. In terms of recruitment of young participants, evidence suggests that online advertisements are effective at appealing to this particular demographic. For example, Ford et al [11] examined paid advertising on Instagram, Snapchat, and Facebook to recruit youths (aged 13-20 years) for 2 cross-sectional surveys. They concluded that these platforms appear to be a modern and cost-effective approach to reaching youths with surveys on sensitive health topics [11]. Similar findings were obtained from

XSL•FO
RenderX

an Australian health survey study that assessed the feasibility of recruiting young females using targeted advertising on Facebook [12]. In another study, Lee et al [8] found that male participants were more difficult to recruit compared to female participants but were able to use targeted, gender-specific advertisements (different language and imagery compared to gender-neutral advertisements) on Facebook to recruit more males. This ability to target different subgroups and easily make changes to targeting strategies are significant benefits of Facebook recruitment. The impact of different targeting strategies can be monitored (eg, costs per click), and further tweaks can be made to determine the best way to reach the target population. In addition, marketing collaterals such as images, headlines, and copy can be changed in order to find the most effective study advertisement [7].

Overall, these studies suggest that online advertising via platforms such as Facebook and Instagram is promising for the effective recruitment of young adults for mHealth survey studies. However, evidence supporting using these recruitment strategies for treatment studies is scarcer. A recent Danish study demonstrated the feasibility of using Facebook and Instagram to recruit young Danes (aged 15-25 years) for an RCT to evaluate a national online youth mental health promotion service [13]. A study that assessed the effectiveness of social media (Facebook), targeted mailing, and in-person recruitment of young adults for diabetes self-management clinical trials found that leveraging a variety of recruitment strategies produced a more representative sample of young adults [14]. To our knowledge, no single study has investigated the effectiveness of various recruitment methods to enlist distressed young adults in digital intervention studies.

The Vibe Up Trial

Vibe Up is an AI-driven adaptive trial, in which a series of "mini" trials are performed instead of 1 large trial. AI methods are used to update an underlying model of the interventions' effectiveness and alter the proportion of participants allocated to each intervention in the next minitrial. Under this scheme, progressively fewer participants are allocated to less effective interventions in later minitrials, with the aim of identifying the most effective intervention as quickly as possible. This trial delivered brief interventions, including mindfulness (a set of five 3-5 minute instructor-guided audio meditations), physical activity (an evidence-based 7-minute high-intensity circuit training protocol, or self-selected daily physical activity), sleep hygiene (a program of 4 infographic-based information modules), and ecological momentary assessment (blended protocol consisting of signal contingent and event-contingent), to ease psychological distress in Australian tertiary education students [15]. A mobile app, Vibe Up, was purpose-built for these trials, allowing eligible students to participate in a "minitrial" by downloading the app from the App Store or Google Play. To ensure that data was captured as intended and to detect fraudulent activities, data verification was carried out by the data manager after the completion of each minitrial. Recruitment for Vibe Up coincided with COVID-19 lockdowns in Australia. This provided us with the opportunity to assess the cost-effectiveness of social media recruitment during a pandemic compared to its aftermath. Research has shown significant increases in distress among young people as a result of the COVID-19 pandemic [16-18]. Transition to online learning for students and restrictions on routine daily activities affected the well-being of young people [19-21]. With access to in-person mental health support no longer possible, we examined whether the cost-effectiveness of recruitment differed between the lockdown and the postlockdown period.

Aims

This study evaluated recruitment data of an AI-driven adaptive digital mental health trial aimed at reducing psychological distress in Australian tertiary education students [15]. This study aims to (1) evaluate the effectiveness of recruitment via Facebook and Instagram compared to traditional methods for a treatment study, (2) compare the cost-effectiveness of online recruitment during and after COVID-19 lockdown, and (3) compare retention rates of different recruitment methods used in this study.

Methods

Study Design

This study evaluated recruitment data collected from the Vibe Up study [15]. Data included engagement metrics collected by Meta (Facebook and Instagram), advertising costs, and Qualtrics data [22] on recruitment methods and rates of survey completion. To recruit participants, paid social media advertising on Facebook and Instagram was used, alongside mailing lists of university networks and student organizations or services, media releases, announcements during classes and events, study posters or flyers on university campuses, and our health professionals' network. Recruitment for 2 pilot trials and 6 minitrials occurred between June 2021 and May 2022.

Participants

Participants were Australian tertiary education students. No restrictions were placed on the state or territory of origin. In order to participate in the study, participants must (1) be aged 18 years or older; (2) be currently enrolled as a student at an Australian university, technical and further education, or other higher education institution; (3) be currently residing in Australia and planning to be a resident throughout the study period; (4) be an advanced, fluent, or native English speaker; (5) own an eligible personal smartphone (iPhone 6S or Android 5 or later) with active mobile number and internet access; and (6) score 20 or more on the Kessler Psychological Distress Scale, 10-item version [23].

Participants are excluded if they (1) score 21 or more on the Suicide Ideation Attributes Scale [24]; (2) report a current active diagnosis of psychosis or bipolar disorder; (3) have already participated in a "minitrial" in the study; (4) indicate major disruptions or events in the next 2 months which may make it difficult to take part in the study; (5) indicate plans to travel outside of Australia in the next 2 months; and (6) indicate that they would be unable to safely undertake a physical activity intervention if allocated to receive this treatment.

XSL•FO
RenderX

Procedure

Recruitment was conducted by the research team in collaboration with the lead institution's in-house marketing and communications department. A social media targeting strategy was developed (Table 1). Clicking on the Meta advertisements (Facebook and Instagram) took potential participants to a landing page, hosted on the lead research institution's website, which contained study information and a direct link to the screening survey that was administered on the Qualtrics XM survey platform [22]. The recruitment window for each Vibe Up trial remained open for up to 7 days only; however, to not risk losing any potential participants between recruitment drives, the survey link was replaced by an expression of interest form on the website to capture the name, email address, and education institution of interested students. These students were contacted at the beginning of the following recruitment drive. Two pilot trials were carried out before recruitment began for 6 minitrials.

Table 1. Profile of individuals targeted on social media.

Category	Detail
Location	Australia
Age (years)	18-40
Sex	Male or female
Education	At university (undergraduate) or at university (postgraduate)
Interests	Science, artificial intelligence, mindfulness, computer science, *DSM-5*[a], mental health foundation, Headspace, Beyond Blue or Lifeline (crisis support service), and international students

[a]DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition).

Active recruitment was conducted via paid advertising on Facebook and Instagram. Traditional methods for recruiting students such as mailing lists, study flyers or posters, class or event announcements, and posts on the Black Dog Institute website and university media were also used. Meta's "dynamic creative" function was used to upload a variety of images, copies, and headlines. Meta delivered a combination of these elements until it determined the top-performing combination and the combination that delivered on the objective at the cheapest cost.

We trialed different advertising objectives, using Meta's (Facebook and Instagram) preset choices. For example, the use of link clicks allowed Meta to deliver study advertisements to people most likely to click on them, with the aim of obtaining the cheapest cost per click. We then unsuccessfully trialed a "landing page view objective" which focused on getting people to not just click on the advertisement, but to load and engage with the content on the page. To improve performance, for minitrials 5 and 6, the Meta pixel was used to set up an "initiate checkout" conversion on the "GET STARTED" button which appeared at the bottom of the recruitment website and took participants to the Qualtrics screening survey. In this instance, Meta optimized study advertisements toward people most likely to take this action.

Measures

Metrics used to measure recruitment in this study were derived from Facebook analytics data and Qualtrics survey data. For ease of comparison, metrics were largely based on previous research in this area [8,9] and are included in Table 2.

XSL•FO
RenderX

Table 2. Study measures and definitions.

Variable name	Definition
Amount spent on advertising per trial (Aus $)	Facebook and Instagram advertising costs per trial in Australian dollars.
Impressions	Number of times study advertisements were viewed.
Reach	Number of people that viewed study advertisements.
Link clicks	Number of times that links to study advertisements were clicked on.
Unique link clicks	Number of people that clicked on study advertisement links, regardless of how many times they clicked on a link.
Click through rate	The percentage of people who viewed and then clicked on the study advertisement. This was calculated by dividing the total number of clicks by the total number of impressions multiplied by 100. Click-through rates measure how successful the study advertisements are in capturing people's attention.
Number of students screened	Number of people who completed screening.
Number of eligible students	Number of people who were eligible to take part in the study.
Number of students who downloaded the study app	Number of people who downloaded the study app from Google Play or the App Store.
Number of students who completed a trial	Number of people who completed the 3 phases of the study (baseline, midpoint, and post).
Costs per screened	Costs per person who completed the screening questionnaires in Australian dollars.
Costs per eligible	Costs per person who was eligible to participate in the study in Australian dollars.
Costs per downloaded	Costs per person who downloaded the study app from Google Play or the App Store in Australian dollars.
Costs per completed	Costs per person who completed all three phases of the study in Australian dollars.

Statistical Analysis

Analyses were performed in RStudio with R (version 3.6.3; R Foundation for Statistical Computing). Descriptive analyses were conducted using Facebook and Instagram analytics data to assess how study advertisements performed and using Qualtrics survey data to determine rates of screened, eligible, and completed participants. A general linear model with pairwise post hoc estimates was used to compare the effectiveness of different methods in recruiting and retaining participants. For each recruitment pathway (Facebook and Instagram advertisement or post, other, and website of lead institution), Fischer exact tests were conducted to determine if gender, student status (domestic or international), and student level (undergraduate or postgraduate) of participants changed significantly as a result of dropout across different stages of the trial. To test for differences in age, Kruskal-Wallis tests were performed. For the purpose of comparing cost-effectiveness between COVID-19 lockdown and post–COVID-19 lockdown, pilot 1 and pilot 2 were classified in the former category and minitrials 1-6 fell into the latter. Inferential statistics were performed using Wilcoxon tests.

Ethical Considerations

Ethical approval was sought and obtained from the University of New South Wales Sydney Human Research Ethics Committee (HC200466). The trial sponsor is the University of New South Wales Sydney. The trial is registered with the Australian New Zealand Clinical Trials Registry (Pilot trials: ACTRN12621001092886; Mini-trials 1-6: ACTRN12621001223820). Informed consent was obtained from each trial participant, and the collected data were stored in a deidentified format. Each participant received Aus $30 [US $18.65] in gift cards.

Results

Demographics

In total, 1314 participants were recruited for 2 pilot trials and 6 minitrials. Participants had a mean age of 22.79 (SD 4.71) years, predominately female (n=1079, 82.1%), domestic (n=1226, 93.3%), and undergraduate students (n=1045, 79.5%). Participants were recruited from 72 tertiary education institutions from across Australia (for the top 10 institutions; Table S1 in Multimedia Appendix 1).

Breakdown by Recruitment Methods

For the 2 pilot trials and 6 minitrials, 3638 students attempted the screening survey. The majority of students who screened for the trials were recruited via social media advertising or posts via Facebook (n=1690, 47%) and Instagram (n=922, 25%). Of the 1314 students (36%) eligible to participate in the study, the vast majority were recruited via social media advertising or posts (n=1203, 92%). Of the 1058 students (81% of those eligible) who provided baseline data, similar proportions were recruited via social media advertising or posts. Of the 982 students (75% of those eligible) who provided midpoint data, recruitment via social media remained the most fruitful (n=889, 91%). Of the 862 students (66% of those eligible) who completed a trial, 90 percent of students were recruited via social media advertising or posts. In terms of participant retention, see Table 3 for a detailed breakdown by recruitment method.

XSL•FO
RenderX

Table 3. Breakdown of participants by recruitment pathway and trial retention.[a]

Recruitment pathway	Number of students retained in the trial				
	Attempted screening (N=3638), n	Screened eligible (n=1314), n (%)	Completed baseline (n=1058), n (%)	Completed midpoint (n=982), n (%)	Completed a trial (n=862), n (%)
Facebook advertisement or post	1690	810 (47.9)	644 (38.1)	597 (35.3)	518 (30.7)
Instagram advertisement or post	922	393 (42.6)	316 (34.3)	292 (31.7)	257 (27.9)
Other	107	61 (57)	55 (51.4)	52 (48.6)	49 (45.8)
Website of lead institution	72	50 (69.4)	43 (59.7)	41 (56.9)	38 (52.7)
Survey not completed	847	—[b]	—	—	—

[a]Other included class or event, email, health professional, poster or flyer, press or media. Retention percentages were calculated based on the number of participants who attempted screening.

[b]Not applicable.

Pairwise comparisons revealed that the lead institution's website was more effective in recruiting eligible participants compared to Facebook (z=3.47; P=.003) and Instagram (z=4.23; P<.001). No statistically significant difference was found between website recruitment and other methods (eg, email, poster or flyer) in recruiting eligible participants. No statistically significant differences were found between recruitment methods in retaining participants at baseline, at mid-point, and at study completion (Tables S2-S9 in Multimedia Appendix 1). To assess possible sample selection effects by recruitment pathway, characteristics of participants who dropped out at each trial stage were examined. No statistically significant age, sex, and student level (undergraduate vs postgraduate) differences were found for recruitment by Facebook, Instagram, others, and the website of the lead institution (Tables S10-S13 in Multimedia Appendix 1). For recruitment by Facebook, statistically significant differences were found for student status (domestic vs international; P=.001; Table S10 in Multimedia Appendix 1). No such differences were found for Instagram, others, and the website of the lead institution (Tables S11-S13 in Multimedia Appendix 1).

Facebook and Instagram Recruitment

Across the 2 pilot trials and 6 minitrials, Aus $7571.72 (US $4707.72) was spent on social media recruitment via Facebook and Instagram advertising, at an average of Aus $946.47 (US $588.47; SE Aus $160.99 [US $100.10]) per trial. A total of 1,119,661 impressions were counted, at an average of 139,958 (SE 27,927) impressions per trial. The performance of study advertisements was measured by the number of clicks generated that led potential participants from the social media platform to the Vibe Up recruitment page. See Figures S1 and S2 in Multimedia Appendix 2 for the top-performing Vibe Up advertisements. During the recruitment drives, 825,635 people were reached, at an average of 103,204 people (SE 17,569) per trial. 8138 clicks were registered on the study website (mean 1017, SE 71), with 7858 being unique link clicks (mean 982, SE 68). The overall click-through rate was 0.73. See Table 4 for the detailed breakdown of participant engagement by trial, during and post–COVID-19 lockdown.

Table 4. Breakdown of trial costs and engagement metrics.

Trial (stage)	Amount spent (Aus $[a])	Impressions, n	Reach, n	Clicks, n	Unique clicks, n	Click-through rate
Pilot 1(lockdown)	445.98	82,266	65,654	1227	1155	1.49
Pilot 2 (lockdown)	254.11	39,709	31,592	699	669	1.76
Trial 1 (postlockdown)	850.00	91,322	71,994	1142	1112	1.25
Trial 2 (postlockdown)	999.70	237,201	150,766	1197	1156	0.50
Trial 3 (postlockdown)	1100.00	247,540	175,161	1203	1168	0.49
Trial 4 (postlockdown)	1750.80	207,799	146,423	967	937	0.47
Trial 5 (postlockdown)	1095.12	99,285	81,652	850	824	0.86
Trial 6 (postlockdown)	1076.01	114,539	102,393	853	837	0.74
Total	7571.72	1,119,661	825,635	8138	7858	0.73

[a]Aus $1=US $0.62.

XSL•FO
RenderX

Cost-Effectiveness of Trial Recruitment

On average, it costs Aus $2.28 (US $1.42; SE Aus $0.49 [US $0.30]) per screened participant, Aus $6.59 (US $4.10; SE Aus $1.39 [US $0.86]) per eligible participant, Aus $7.78 (US $4.84; SE Aus $1.64 [US $1.02]) per participant who downloaded the study mobile app, and Aus $10.01 (US $6.22; SE Aus $2.20 [US $1.37]) per completion. See Table S16 in Multimedia Appendix 1 for the breakdown of costs by trial.

Impacts of COVID-19 Lockdowns

Wilcoxon tests were carried out to compare costs incurred per link click between COVID-19 lockdown (pilot 1 and pilot 2), and post–COVID-19 lockdown (minitrials 1-6). Significant differences were found in the cost per click (lockdown: median Aus $0.35 [US $0.22], IQR Aus $0.27-$0.47 [US $0.17-$0.29]; postlockdown: median Aus $1.00 [US $0.62], IQR Aus $0.70-$1.47 [US $0.44-$0.92]; W=9087; P<.001; Table S14 in Multimedia Appendix 1). As recruitment drives occurred in short bursts (3-7 days as opposed to weeks or months for other recruitment campaigns), we compared lockdown and postlockdown costs using hourly spending on Facebook and Instagram advertising. Differences between lockdown and postlockdown on the amount spent per hour to reach the target sample size were also statistically significant (lockdown: median Aus $4.75 [US $2.95], IQR Aus $1.94-$6.34 [US $1.22-$3.97]; postlockdown: median Aus $13.29 [US $8.26], IQR Aus $4.70-$25.31 [US $2.95-$15.87]; W=16044; P<.001; Table S15 in Multimedia Appendix 1).

Discussion

Principal Findings

With the increasing adoption of remote clinical trials in digital mental health research, and the uptake of the AI-driven adaptive trial methodology, strategies to optimize timely recruitment of suitable participants warrant investigation. Advancements in the way online social media platforms can be manipulated to reach target audiences faster and cheaper provided further impetus for this study. The effectiveness of participant recruitment via Facebook and Instagram was compared with traditional methods such as mailing lists and study flyers or posters. Results suggest that Facebook and Instagram are effective in recruiting young adults for remote digital mental health trials. Social media advertising outperformed more traditional methods in the total number of eligible participants recruited. However, we also found that compared to Facebook and Instagram, recruitment via the lead institution's website was more effective in yielding eligible participants from those who completed screening. It may be the case that participants recruited while browsing the institution's website (a mental health research institute) were actively seeking information to manage mental health concerns, and thus more likely to be eligible for the study than someone shown the study advertisement via social media because they fit the target age group. In addition, although not statistically significant, the completion rate was higher for participants recruited via the website. This could be explained by the familiarity, trust, and support of the mental health research carried out by the lead institution and suggests that this recruitment pathway can be effective for trial enrollment and importantly, trial completion. Given that the vast majority of eligible participants were recruited via social media advertising (92%) compared to website recruitment (4%), if funds are available, it would still make sense to include social media in the recruitment strategy to recruit young people for digital mental health trials. In terms of participant retention from study baseline to completion, there was no stand-out recruitment method. Similar to the number of eligible participants, the sheer volume of completers recruited via social media (90%) compared to traditional methods, provides a rationale for including this method in participant recruitment strategies. We did not find evidence of a sample selection effect due to the recruitment methods used. However, similar trials in the future should identify strategies to attract more international students. For example, targeted social media advertising for this student group and liaising with student leaders to disseminate trial information to club or society members.

In terms of cost-effectiveness, this study provided further evidence to support the benefits of using social media to recruit young adults for mental health research participation. Our results (Aus $7.78 [US $4.84] per participant app download and Aus $10.01 [US $6.22] per trial completion) compared favorably with the Mood Memos Study (Aus $19.89 [US $12.37] per recruited participant) [6], a health survey study targeting young females (US $20 per compliant participant) [12], a health intervention study of 334 young Australian adults (Aus $105.77 [US $65.76] per recruited participant) [25], diabetes self-management clinical trial for young adults (US $334 per enrolled participant) [14], and a review that calculated estimated median recruitment costs for RCTs in mental health research (US $42.82) [10]. Even though costs per completion for our trial were higher than studies conducted by Lee et al [8] (Aus $2.81 [US $1.75] per completion), Ash et al [26] (Aus $6.07 [US $3.77] per completion) [26], and Batterham [9] (Aus $1.51 [US $0.94] per completion), additional costs are explained by the nature of research activities involved; one-off survey completions, as opposed to completing a 29-day trial involving brief mental health interventions in Vibe Up. In comparison to an RCT that recruited 560 Danes aged 15-25 years to evaluate an online youth mental health promotion service [13], US $4.20 per baseline survey completion reported for this trial was almost identical to costs per eligible participant for Vibe Up (Aus $6.59 [US $4.10]).

More than 1.1 million impressions were counted during our recruitment drives, reaching 825,635 people. These are numbers difficult to achieve in a short period of time using flyers or posters, class announcements, and mailing lists. The ability to track the number of clicks on the study website (n=8138), and determine if clicks were by different individuals (7858 unique clicks), in addition to the ability to alter targeting setup (eg, tailoring advertisements to male students), makes social media recruitment via Facebook and Instagram an attractive proposition for researchers looking to recruit a large sample within a limited period of time. Further, click-through rates provided valuable insights into how engaging the study advertisements are to the target audience, allowing fine-tuning of recruitment materials during recruitment drives to maximize participant sign-up.

XSL•FO
RenderX

According to WordStream [27], a good click-through rate for Facebook advertising averages approximately 0.90%. This study achieved an overall click-through rate of 0.73%, with the highest rates achieved during the COVID-19 lockdown (pilot 1: 1.49 and pilot 2: 1.76). It is important to acknowledge that the success in recruiting young people for our trial also leveraged the well-established and trusted brand of the lead institution in the mental health space in Australia.

A spike in costs for minitrial 4 warranted our attention. Up until then, a "link click" objective was used wherein Facebook and Instagram directed the study advertisements toward individuals most likely to click on them. A change to the "landing page view" objective for trial 4, where individuals more likely to load the page content were targeted, resulted in higher costs. It was a tactic that did not pay off as Facebook and Instagram needed time and more money to relearn the change in objective. For subsequent trials, a conversion pixel on the "GET STARTED" button was implemented, targeting people most likely to participate. This was a cost-saving strategy, homing in on people more likely to take action. Although the ability to change strategies demonstrated advancements in social media recruitment to reach specific target samples, it does come with additional financial costs that researchers should be aware of. In addition, fraudulent participants have increasingly infiltrated research recruitment via social media, especially trials offering monetary reimbursement. The current study implemented several safeguards including the use of time-based one-time password for app downloads and manual checking of participant details before inclusion in the study.

The first 2 recruitment drives coinciding with the COVID-19 lockdowns provided us with the unique opportunity to assess whether social media recruitment was more cost-effective during the lockdown period compared to postlockdown. Comparisons between the 2 periods found that fewer funds were needed during lockdown to trigger interest from prospective participants and result in sign-ups. The COVID lockdown may have facilitated recruitment, with students taking the opportunity to learn new, evidence-based coping strategies for psychological distress offered by the trial. The planned recruitment period of 2 weeks was reduced to 7 days, with almost 300 participants recruited for the first pilot trial. Adjustments were made for pilot 2 (7-day recruitment period); however, intake was closed after 3 days with sign-ups exceeding the target of 120-150 participants. COVID-19 lockdown was the most cost-effective recruitment period, with fewer funds required per screened participant, per eligible participant, per app download, and per completed participant. The psychological impact of the lockdown may have prompted students to seek ways to improve their mental well-being. Even though recruitment periods for subsequent trials (postlockdown) remained consistent (3-5 days), more funds were expended to reach the target sample size. Costs per completed participant increased to between Aus $8 (US $4.97) and Aus $21 (US $13.06; compared to under Aus $3 [US $1.87] for the pilot trials). These opportunistic findings suggest that recruitment of young people may yield better results during their "downtime," for example, during semester breaks or festive seasons, when they have the capacity to contribute to research. In addition, recruitment materials need to include information on benefits to participants (for example, providing an opportunity to learn evidence-based coping strategies), and screening needs to exclude those unable to commit for the entire study (for example, asking participants to indicate major disruptions or events in the next 2 months which may make it difficult to take part in the study) to help boost retention rates.

Limitations

First, we were unable to accurately assign a monetary value to the traditional recruitment methods used for our trials (eg, poster or flyer and classroom announcements) so no direct comparisons were made on cost-effectiveness with social media recruitment. As a result, we are unable to comment on whether one recruitment method is more cost-effective than another in this study. Instead, the costs of social media recruitment were compared with previously published data to help determine if this method should be pursued in future mental health AI-driven adaptive trials. Secondly, we struggled to correct the sex imbalance within our sample. This is consistent with the recruitment literature [28,29]. Even though more funds were allocated to the recruitment of male participants from main trial 4 onwards, we were unable to correct the skew. For future studies, user testing of tailored advertisements for male participants should be considered before the start of recruitment campaigns. Next, given that our participants were recruited from all over Australia, it was difficult to isolate which participants were in lockdown, as lockdown periods varied across the country. We have categorized our trials based on dates published online by the Australian government [30]. Finally, the young people referred to in our study are all tertiary education students. Their characteristics and inclination to participate in research may differ from others in this group (eg, in full-time employment). Hence, our findings may not be generalizable to all young people in Australia but are likely to reflect those currently attending an Australian tertiary education institution.

Conclusions

Social media advertising on Facebook and Instagram is a productive method to recruit young people for AI-driven adaptive digital mental health intervention trials. Given that repeated recruitment drives are necessary for this trial methodology (in our case, once every month for 2 pilot trials and 6 minitrials), we found social media to be more effective in reaching potential participants all over Australia than traditional recruitment methods such as mailing lists and flyers or posters. In addition, advertising costs from this study compared favorably with other large-scale traditional trials, providing evidence that social media advertising can be effective for participant recruitment in future AI-driven adaptive trials. Perhaps a reflection of the added distress experienced by young people, online advertising costs were low during COVID-19 lockdowns and increased post lockdown. Based on our experience, we encourage other researchers interested in running AI-driven adaptive trials to consider the use of social media advertising to recruit participants.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge contributions from Nicola Livingstone and Simon Goh to the study.

Authors' Contributions

WYZ, A Shvetcov, A Slade, ZJ, LH, AW, RL, SG, KH, ES, SV, KM, RV, JMN, and HC played key roles in the design and development of study processes. ES, KH, WYZ, ZJ, LH, RL, SV, KM, JMN, and HC developed the recruitment strategies with support from AW, SG, JR, SR, A Shvetcov, and A Slade. LH led the development of the technical trial platform, with support from SK, AA, JFK, SC, MS, and NM. All authors provided critical input throughout participant recruitment and the development of the manuscript. All authors reviewed the manuscript and provided comments. All authors approved the manuscript prior to publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File , 43 KB - jmir_v27i1e60413_app1.docx]

Multimedia Appendix 2
Vibe Up study advertisements.
[DOCX File , 211 KB - jmir_v27i1e60413_app2.docx]

References

1. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med 2018;16(1):29 [FREE Full text] [doi: 10.1186/s12916-018-1017-7] [Medline: 29490655]
2. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. Trends Pharmacol Sci 2019;40(8):577-591 [FREE Full text] [doi: 10.1016/j.tips.2019.05.005] [Medline: 31326235]
3. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
4. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials—a systematic review. PLoS One 2018;13(8):e0203127 [FREE Full text] [doi: 10.1371/journal.pone.0203127] [Medline: 30157250]
5. Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M. The role of social media in enhancing clinical trial recruitment: scoping review. J Med Internet Res 2020;22(10):e22810 [FREE Full text] [doi: 10.2196/22810] [Medline: 33104015]
6. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
7. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote eHealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
8. Lee S, Torok M, Shand F, Chen N, McGillivray L, Burnett A, et al. Performance, cost-effectiveness, and representativeness of Facebook recruitment to suicide prevention research: online survey study. JMIR Ment Health 2020;7(10):e18762 [FREE Full text] [doi: 10.2196/18762] [Medline: 33090115]
9. Batterham PJ. Recruitment of mental health survey participants using internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014;23(2):184-191 [FREE Full text] [doi: 10.1002/mpr.1421] [Medline: 24615785]
10. Sanchez C, Grzenda A, Varias A, Widge AS, Carpenter LL, McDonald WM, et al. Social media recruitment for mental health research: a systematic review. Compr Psychiatry 2020;103:152197 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152197] [Medline: 32992073]
11. Ford KL, Albritton T, Dunn TA, Crawford K, Neuwirth J, Bull S. Youth study recruitment using paid advertising on Instagram, Snapchat, and Facebook: cross-sectional survey study. JMIR Public Health Surveill 2019;5(4):e14080 [FREE Full text] [doi: 10.2196/14080] [Medline: 31599739]
12. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]

XSL•FO
RenderX

13. Hoffmann SH, Paldam Folker A, Buskbjerg M, Paldam Folker M, Huber Jezek A, Lyngsø Svarta D, et al. Potential of online recruitment among 15-25-year olds: feasibility randomized controlled trial. JMIR Form Res 2022;6(5):e35874 [FREE Full text] [doi: 10.2196/35874] [Medline: 35612877]
14. Salvy SJ, Carandang K, Vigen CL, Concha-Chavez A, Sequeira PA, Blanchard J, et al. Effectiveness of social media (Facebook), targeted mailing, and in-person solicitation for the recruitment of young adult in a diabetes self-management clinical trial. Clin Trials 2020;17(6):664-674 [FREE Full text] [doi: 10.1177/1740774520933362] [Medline: 32627589]
15. Huckvale K, Hoon L, Stech E, Newby JM, Zheng WY, Han J, et al. Protocol for a bandit-based response adaptive trial to evaluate the effectiveness of brief self-guided digital interventions for reducing psychological distress in university students: the Vibe Up study. BMJ Open 2023;13(4):e066249 [FREE Full text] [doi: 10.1136/bmjopen-2022-066249] [Medline: 37116996]
16. Wilkins R, Vera-Toscano E, Botha F, Wooden M, Trinh TA. The household, income and labour dynamics in Australia survey: selected findings from waves 1 to 20. Melbourne Institute: Applied Economic & Social Research. 2022. URL: https://melbourneinstitute.unimelb.edu.au/__data/assets/pdf_file/0011/4382057/HILDA_Statistical_Report_2022.pdf [accessed 2024-12-12]
17. Bower M, Smout S, Donohoe-Bales A, O'Dean S, Teesson L, Boyle J, et al. A hidden pandemic? An umbrella review of global evidence on mental health in the time of COVID-19. Front Psychiatry 2023;14:1107560 [FREE Full text] [doi: 10.3389/fpsyt.2023.1107560] [Medline: 36970258]
18. Batra K, Sharma M, Batra R, Singh TP, Schvaneveldt N. Assessing the psychological impact of COVID-19 among college students: an evidence of 15 countries. Healthcare 2021;9(2):222 [FREE Full text] [doi: 10.3390/healthcare9020222] [Medline: 33671363]
19. Alesi M, Giordano G, Gentile A, Caci B. The switch to online learning during the COVID-19 pandemic: the interplay between personality and mental health on university students. Int J Environ Res Public Health 2023;20(7):5255 [FREE Full text] [doi: 10.3390/ijerph20075255] [Medline: 37047875]
20. Butnaru GI, Haller AP, Dragolea LL, Anichiti A, Tacu Hâr an GD. Students' wellbeing during transition from onsite to online education: Are there risks arising from social isolation? Int J Environ Res Public Health 2021;18(18):9665 [FREE Full text] [doi: 10.3390/ijerph18189665] [Medline: 34574589]
21. Liu C, McCabe M, Dawson A, Cyrzon C, Shankar S, Gerges N, et al. Identifying predictors of university students' wellbeing during the COVID-19 pandemic—a data-driven approach. Int J Environ Res Public Health 2021;18(13):6730 [FREE Full text] [doi: 10.3390/ijerph18136730] [Medline: 34206579]
22. Qualtrics LLC. URL: https://www.qualtrics.com/en-au/ [accessed 2024-12-14]
23. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002;32(6):959-976. [doi: 10.1017/S0033291702006074]
24. van Spijker BA, Batterham PJ, Calear AL, Farrer L, Christensen H, Reynolds J, et al. The Suicidal Ideation Attributes Scale (SIDAS): community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]
25. Musiat P, Winsall M, Orlowski S, Antezana G, Schrader G, Battersby M, et al. Paid and unpaid online recruitment for health interventions in young adults. J Adolesc Health 2016;59(6):662-667. [doi: 10.1016/j.jadohealth.2016.07.020] [Medline: 27663927]
26. Ash GI, Robledo DS, Ishii M, Pittman B, DeMartini KS, O'Malley SS, et al. Using web-based social media to recruit heavy-drinking young adults for sleep intervention: prospective observational study. J Med Internet Res 2020;22(8):e17449 [FREE Full text] [doi: 10.2196/17449] [Medline: 32780027]
27. Irvine M. Facebook ad benchmarks for your industry [data]. WordStream. URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks [accessed 2017-02-28]
28. Ryan J, Lopian L, Le B, Edney S, Van Kessel G, Plotnikoff R, et al. It's not raining men: a mixed-methods study investigating methods of improving male recruitment to health behaviour research. BMC Public Health 2019;19(1):814 [FREE Full text] [doi: 10.1186/s12889-019-7087-4] [Medline: 31234825]
29. Knox J, Morgan P, Kay-Lambkin F, Wilson J, Wallis K, Mallise C, et al. Male involvement in randomised trials testing psychotherapy or behavioural interventions for depression: a scoping review. Curr Psychol 2022;42(34):1-16 [FREE Full text] [doi: 10.1007/s12144-022-04017-7] [Medline: 36531196]
30. Department of Education ECEC COVID-19 timeline: Australian Government Department of Education, Skills and Employment. URL: https://tinyurl.com/mvhaa7a3 [accessed 2021-12-19]

Abbreviations

AI: artificial intelligence
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by T McCall; submitted 15.05.24; peer-reviewed by J Bourgaize, K Ishida; comments to author 27.10.24; revised version received 12.11.24; accepted 12.11.24; published 14.01.25.

Please cite as:
Zheng WY, Shvetcov A, Slade A, Jenkins Z, Hoon L, Whitton A, Logothetis R, Ravindra S, Kurniawan S, Gupta S, Huckvale K, Stech E, Agarwal A, Funke Kupper J, Cameron S, Rosenberg J, Manoglou N, Senadeera M, Venkatesh S, Mouzakis K, Vasa R, Christensen H, Newby JM
Recruiting Young People for Digital Mental Health Research: Lessons From an AI-Driven Adaptive Trial
J Med Internet Res 2025;27:e60413
URL: https://www.jmir.org/2025/1/e60413
doi:10.2196/60413
PMID:

©Wu Yi Zheng, Artur Shvetcov, Aimy Slade, Zoe Jenkins, Leonard Hoon, Alexis Whitton, Rena Logothetis, Smrithi Ravindra, Stefanus Kurniawan, Sunil Gupta, Kit Huckvale, Eileen Stech, Akash Agarwal, Joost Funke Kupper, Stuart Cameron, Jodie Rosenberg, Nicholas Manoglou, Manisha Senadeera, Svetha Venkatesh, Kon Mouzakis, Rajesh Vasa, Helen Christensen, Jill M Newby. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Parental Mental Health and Child Maltreatment in the COVID-19 Pandemic: Importance of Sampling in a Quantitative Statistical Study

Lara Engelke[1], BSc, MSc; Claudia Calvano[2,3], PhD; Steffi Pohl[4], MSc, PhD; Sibylle Maria Winter[3,5*], MD; Babette Renneberg[1,3*], PhD

[1]Clinical Psychology and Psychotherapy, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[2]Clinical Child and Adolescent Psychology and Psychotherapy, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[3]German Center for Mental Health (DZPG), Berlin, Germany

[4]Division Methods and Evaluation, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[5]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Lara Engelke, BSc, MSc
Clinical Psychology and Psychotherapy
Department of Education and Psychology
Freie Universität Berlin
Habelschwerdter Allee 45
Berlin, 14195
Germany
Phone: 49 30 838 76522
Email: engelke.l@fu-berlin.de

Abstract

Background: Results on parental burden during the COVID-19 pandemic are predominantly available from nonrepresentative samples. Although sample selection can significantly influence results, the effects of sampling strategies have been largely underexplored.

Objective: This study aimed to investigate how sampling strategy may impact study results. Specifically, we aimed to (1) investigate if outcomes on parental health and child maltreatment during the COVID-19 pandemic from a convenience sample differ from those of a specific representative sample and (2) investigate reasons for differences in the results.

Methods: In 2020, we simultaneously conducted 2 studies: (1) a web-based survey using a convenience sample of 4967 parents of underage children, primarily recruited via social media, and (2) a study using a quota sample representative of the German adult population with underage children (N=1024), recruited through a combination of telephone interviews and computer-assisted web interviews. In both studies, the same questionnaire was used. To evaluate the impact of sampling, we compared the results on outcomes (parental stress, subjective health, parental mental health, general stress, pandemic-related stress, and the occurrence of child maltreatment) between the 2 samples. To explain differences in the results between the 2 studies, we controlled for sociodemographic data, parent-related risk factors, and COVID-19–related experiences.

Results: Compared to parents from the quota sample, parents from the convenience sample reported significantly more parental stress (η^2=0.024); decreased subjective health (η^2=0.016); more anxiety and depression symptoms (η^2=0.055); more general stress (η^2=0.044); more occurrences of verbal emotional abuse (VEA; ϕ=0.12), witnessing domestic violence (WDV; ϕ=0.13), nonverbal emotional abuse (NEA; ϕ=0.03), physical abuse (ϕ=0.10), and emotional neglect (ϕ=0.06); and an increase of child maltreatment (VEA: exp(B)=2.95; WDV: exp(B)=3.19; NEA: exp(B)=1.65). Sociodemographic data, parent-related risk factors, and COVID-19–related experiences explained the differences in parental stress (remaining difference between samples after controlling for covariates: η^2=0.002) and subjective health (remaining difference between samples after controlling for covariates: η^2=0.004) and partially explained differences in parental mental health (remaining: η^2=0.016), general stress (remaining: η^2=0.014),

XSL•FO
RenderX

and child maltreatment (remaining: VEA: exp(B)=2.05 and WDV: exp(B)=2.02) between the 2 samples. The covariates could not explain the difference in NEA (exp(B)=1.70). We discuss further factors that may explain the unexplained differences.

Conclusions: Results of studies can be heavily impacted by the sampling strategy. Scientists are advised to collect relevant explaining variables (covariates) that are possibly related to sample selection and the outcome under investigation. This approach enables us to identify the individuals to whom the results apply and to combine findings from different studies. Furthermore, if data on the distribution of these explanatory variables in the population are available, it becomes possible to adjust for sample selection bias.

(J Med Internet Res 2025;27:e52043) doi:10.2196/52043

KEYWORDS

COVID-19; parental stress; parental mental health; child maltreatment; data collection methods; web-based surveys; convenience sample; sampling methods

Introduction

Background

The COVID-19 pandemic and the resulting consequences (eg, restrictions, such as school and daycare closings, social distancing, and stay-at-home orders) posed various challenges, especially for families. Parents felt stressed [1-4] and were especially burdened during the COVID-19 pandemic [1,5,6]. Studies showed worsening or significant impairments of parental mental health (depression, anxiety, and stress) [2,6-16], a significant increase in parental stress [3,14,17-20], and an increase in the occurrence of child maltreatment during the COVID-19 pandemic [17,21,22].

The existing trend toward convenience sampling in psychological research [23-25] increased during the COVID-19 pandemic [26]. As is the case for most studies on mental burden and well-being of families during the COVID-19 pandemic, findings on parental outcomes are primarily generated through web-based studies with convenience sampling (eg, [4,7,10,13-17,19,22,27,28]). In a web-based study, convenience sampling is often conducted by recruiting participants through easily accessible methods, such as social media posts, previous study participants, or email lists. Consequently, they are often not representative of the population of interest. Only a few studies evaluating the impact of the COVID-19 pandemic on families recruited a representative sample (eg, [1-3,11]). Therefore, many studies are not representative of the general population or the specific population of interest. In other words, in web-based convenience samples, the general population is often not adequately represented. Differences in sample characteristics may not be a problem, if the sample characteristics that differ are irrelevant for the outcome one wants to assess. However, if the samples differ in variables that are relevant for the outcomes (eg, single parents are less likely to take part in a web-based convenience study on parental burden because they may not have the time; at the same time, single parents will more likely be affected by COVID-19 restrictions), the results will be biased and cannot be generalized to the overall population.

Studies have shown substantial differences in sociodemographic composition from web-based convenience samples in comparison with representative samples [29-32]. In their systematic literature review on recruitment via Facebook for health-related studies, Whitaker et al [33] found that some studies showed an overrepresentation of White, female, younger individuals with higher education and higher income. The overrepresentation of these sample characteristics in web-based convenience samples on parenting and family research is also known [34-39]. In various web-based convenience samples assessing the impact of the COVID-19 pandemic on parents, participants are more likely to be female [4,9,14-17,19,22,27,40-45], in a relationship [4,15,19,22,40,41,43], and have higher education [4,14,16,17,27,41-43,45-47]. Differences in sociodemographic variables (eg, sex, child age, parental age, and socioeconomic status) compared to the general population further influence the outcomes, as some of these factors are associated with poorer parental outcomes and greater family burden during the pandemic [8,11,16,40-42,48-60]. Thus, these study results may not necessarily generalize to the overall population during the COVID-19 pandemic [26,61-66].

The generalizability of results from nonprobability studies with web-based convenience samples was already discussed before the COVID-19 pandemic [29,31,67-69]. In their literature review, Cornesse et al [31] concluded that the generalizability of probability sample surveys (eg, representative samples) is higher than that of nonprobability sample surveys (eg, web-based convenience samples). Similar results were found by other researchers [25,30,32,70-76]. Even before the COVID-19 pandemic, studies assessing mental health indicated that individuals with higher burden were more likely to choose to participate in web-based convenience sample studies, resulting in discrepancies in burden scores compared to representative samples [25,77,78]. This selection effect is also evident in studies assessing the impact of the COVID-19 pandemic, resulting in an overestimation and underestimation of specific outcomes [65,79].

Studies on the generalizability of web-based convenience sampling in parenting and family research are scarce. One Australian study found that mothers surveyed on the web via Facebook, compared to a population-representative sample, showed sociodemographic differences (eg, overrepresentation of younger individuals, those in a relationship, with higher education, and who spoke only English at home) as well as poorer mental health and self-rated health [39]. Similarly, Bennetts et al [80] reported higher psychological distress in

XSL•FO
RenderX

their parent sample recruited through Facebook compared to a representative sample.

Objective

To the best of our knowledge, no published studies have investigated the generalizability of results from web-based convenience samples on parental outcomes during the COVID-19 pandemic, leaving the potential biasing effect unknown. The following research questions were addressed: (1) Are results from a convenience sample during the COVID-19 pandemic comparable to results from a specific representative sample regarding parental outcomes (ie, parental stress, parental health, and the occurrence of child maltreatment)? and (2) To what extent can we explain differences in outcomes between the studies by sample characteristics?

Methods

Study Design

This study is based on the project "Parenting in Pandemic Times." The target population of the project was parents with underage children in the household in Germany (national) with sufficient German language skills. Data presented in this paper refer to 2 separate studies conducted simultaneously. First, a quota sample was collected by the market research company INFO Marktforschungsinstitut between August 3 and August 11, 2020. Computer-assisted telephone interviews (CATIs; n=402) were combined with a computer-assisted web interview (CAWI; n=622) to provide data that are as representative as possible. The CAWI allowed respondents to complete the questionnaire on the web at home on a computer, tablet, cell phone, or other smart device. A dual-frame design was used for the CATI sample to reach respondents via landline and cell phone. Participants in the CAWI sample were recruited from an active web-based access panel of the market research institute. They received incentives for their participation. In the first step, specific representativity was assured by recruiting participants based on quota obtained from the latest microcensus data regarding parental sex, parental age, number of children, federal state, school educational level (with vs without A-level), and income. In the second step, to adjust for the remaining disproportion in these variables due to unit nonresponse, INFO Marktforschungsinstitut additionally used poststratification weighting. While educational level was assessed in 4 categories, representativeness (eg, quota and weighting) was only ensured for the dichotomous variable of having an A-level (high education) versus not having an A-level. For the dichotomous variable on education, specific representativity could be assured. The weighting took place retrospectively to the actual quota sampling. (For more information, refer to the study by Calvano et al [2]; in this study, weighting only slightly altered the strata-variable distribution in the specific representative study. Calvano et al [2] determined that using unweighted as compared to weighted data does not alter the findings or conclusions.) Second, a convenience sample was collected between August 2 and September 16, 2020. The recruitment of this sample (N=4967) took place through requests and emails via an web link through social media (eg, Facebook groups, Instagram, and Twitter; 2561/4967, 51.56%; refer to Multimedia Appendix 1, picture 1), friends and family (769/4967, 15.48%), internet (658/4967, 13.25%), schools (551/4967, 11.09%), daycare facilities (263/4967, 5.29%), work (84/4967, 1.69%), clubs (65/4967, 1.31%), clinics for child and adolescent psychiatry (34/4967, 0.68%), child and adolescent therapists and pediatricians (18/4967, 0.36%), newspapers (32/4967, 0.64%), and child protection outpatient clinics (16/4967, 0.32%). As the research team was based in Berlin, Germany, recruitment primarily took place in Berlin through schools, daycare facilities, child and adolescent therapists and physicians, clubs, and newspapers. Social media requests were placed in groups with German-speaking parents. A total of 42.5% (2111/4967) of the convenience sample came from Berlin. Participants in the convenience sample were able to take part in a raffle for vouchers and iPads at the end of the survey. In both samples, few data were excluded due to an unrealistic short completion time (convenience: 53/5020, 1.06%; specific representative: 66/1090, 6.06%). Noteworthily, the surveys were identical in content, that is, all measures and instructions applied were identical.

Ethical Considerations

Data collection, analysis, evaluation, and reporting were conducted in accordance with the current ethical guidelines and data protection laws. All data are anonymized. The Ethics Committee of Charité – Universitätsmedizin Berlin (EA2/128/20; July 27, 2020) approved the study. Participants received information on the purpose of the research, the scope and duration of the diagnostic measures, possible risks and benefits, the voluntary nature of participation, and all aspects of data protection. For participation in the study, participants agreed to an informed consent declaration. Participants were provided with contact details for support services. Participants in the convenience sample could take part in a prize raffle for cash vouchers. Participants in CAWI received incentives from the market research company.

Measures

We distinguished between explaining variables (sociodemographic data, parent-related risk factors, and pandemic-related experience) and parental outcomes (parental stress, subjective health, parental mental health, general stress, and child maltreatment).

Explaining Variables

Sociodemographic Data

Data on parental sex and age, parental status (biological parent, stepparent, and other), nationality, number of children in the household, children's age and sex, marital status, federal state, parental school and professional education, and parental current employment status were assessed. We calculated the Winkler Index [81] with reference data from the German population to classify the socioeconomic status as low, medium, or high.

Parent-Related Risk Factors

The alcohol abuse module of the Patient Health Questionnaire, German version [82,83] was applied to assess parents' risk of alcohol abuse during the COVID-19 pandemic. Furthermore, the study assessed parents' experiences of sexual or physical

XSL•FO
RenderX

violence in adulthood, their own history of childhood sexual or physical abuse, and the presence of mental disorders. In addition, parents were asked to report the presence of a severe or chronic physical condition and if they belonged to the risk group for COVID-19.

Pandemic-Related Experience

To measure pandemic-related experiences and stress related to COVID-19, the Pandemic Stress Scale (Winter, S, Corona Belastungsskala. Charité - Universitätsklinikum Berlin, Klinik für Psychiatrie, Psychosomatik und Psychotherapie des Kindes- und Jugendalter, Sektion Traumafolgen und Kinderschutz, unpublished data, January 2020) was developed. Questions were answered with respect to household members, the family, or the parents themselves. At first, COVID-19–related personal experiences were assessed, that is, contact with persons with a COVID-19 infection, hospital admission, or death. Afterward, the month or months with the subjectively highest burden (January 2020 to August 2020) were assessed.

Parental Outcomes

Parental Stress

To assess negative and positive perceptions of parenthood, the Parental Stress Scale [84] was implemented for (1) the time with the subjectively highest burden (total sample [convenience and specific representative sample combined]: Cronbach α=0.88) and (2) before the COVID-19 pandemic, that is, January 2020 in retrospect (total sample: Cronbach α=0.86). Parents rated their perception on the 18 items on a 5-point frequency scale. Higher scores expressed higher parental stress.

Subjective Health

To measure parents' perception of their own health (1) at the time of the highest burden and (2) retrospectively in January 2020, a well-established single-item measure [85] with a 10-point scale was implemented. Higher scores indicated better self-rated health.

Parent Mental Health

To assess depression and anxiety symptoms, the Patient Health Questionnaire 4 [86] was applied (total sample: generalized anxiety, Cronbach α=0.70; depression, Cronbach α=0.68; and total score, Cronbach α=0.78). Higher scores on the 4-point scale indicated poorer mental health.

Pandemic-Related Stress

The Pandemic Stress Scale (Winter, S, Corona Belastungsskala. Charité - Universitätsklinikum Berlin, Klinik für Psychiatrie, Psychosomatik und Psychotherapie des Kindes- und Jugendalter, Sektion Traumafolgen und Kinderschutz, unpublished, January 2020) consists of 14 items about specific restrictions (eg, daycare closings and social distancing), which parents rated regarding their subjective burden on a 5-point scale (ranging from not at all burdening to very burdening). The answering format also included a "not applicable" option. The Pandemic Stress Scale refers to the time point with the subjectively highest burden during the COVID-19 pandemic so far, from the beginning to August 2020. The scale had good internal consistency (total sample: Cronbach α=0.85). Higher scores expressed a higher subjective burden.

General Stress

To measure general parental stress at the time of the subjectively highest burden, we implemented the stress module of the Patient Health Questionnaire, German version [83] (total sample: Cronbach α=0.71). It consists of 10 items with several psychosocial stressors on a 3-point scale, and higher scores indicated higher general stress.

Child Maltreatment

Parents provided information about child maltreatment regarding the children in their household. The Pediatric Maltreatment and Abuse Chronology of Exposure interview [87-89] was adjusted to collect data on household dysfunction, child neglect, and abuse. First, parents were asked if the children ever faced severe stressful life experiences (eg, abuse, violence, and neglect). Afterward, 10 items on distinct subtypes of child maltreatment experiences were asked: 5 items on different types of child abuse (nonverbal emotional abuse [NEA], verbal emotional abuse [VEA], physical abuse, witnessing domestic violence [WDV], and sexual abuse), 3 items on different types of neglect (supervisory neglect, physical neglect, and emotional neglect), and 2 items on different types of household dysfunction (problems regarding mental illness or alcohol or substance use in the household). Only the first item explicitly referred to severe levels of maltreatment; the other items on subtypes mirrored comparatively lower severity levels in our surveys (according to the maltreatment classification system [90,91]). First, a dichotomous yes or no item was presented for each subtype asking parents if the children ever experienced a specific subtype of child maltreatment. If the answer was yes, a follow-up question was posed, asking about the change in frequency since the beginning of the COVID-19 pandemic, with response options on a 5-point scale ranging from "significantly more often" to "significantly less often." For more details about the measure, refer to the study by Calvano et al [2]. To simplify the presentation of the results, we dichotomized the responses on child maltreatment during the COVID-19 pandemic into increase (increase or significant increase) versus no increase (staying the same, decrease, or significant decrease). To provide meaningful results, we decided not to further compare subtypes with case numbers smaller <30.

Data Analysis

For investigating the research questions, we aimed to answer the following: To what extent do the 2 studies differ with respect to the parental outcomes and specific sample characteristics? How well can we predict sample membership by these sample characteristics? How well can we explain differences in parental outcomes between the 2 samples by sample characteristics?

Investigating Differences in Explaining Variables Between Samples

We calculated effect sizes describing the differences in explaining variables/covariates (sociodemographic data, parent-related risk factors, and COVID-19–related experiences) between the 2 studies. The significance of differences between the samples was investigated using 2-tailed *t* tests or Welch tests for interval-scaled variables (eg, age, rooms in the household, parental mental health, and parental stress) and

XSL•FO
RenderX

chi-square tests for nominal and ordinal variables (eg, sex, federal state, nationality, garden, belonging to a risk group for COVID-19, and school education). For metric variables, we report Cohen *d*, and for nominal and ordinal-scaled variables, we report the phi coefficient.

After examining differences between studies for each individual variable, the next step was to consider all covariates simultaneously. As such, we used logistic regression to predict sample membership (0=specific representative sample and 1=convenience sample) by the covariates, that is, sociodemographic data, parent-related risk factors, and COVID-19–related experience. Due to estimation requirements, we only reported on the prediction of study membership for categories of categorical outcomes with an absolute frequency >30.

Investigating Differences in Parental Outcomes Between Samples

We calculated effect sizes describing the differences in outcomes between the 2 studies, using ANOVA. We report partial η^2 for effect size. Chi-square tests were used for nominal and ordinal outcomes, with the phi coefficient reported for effect size.

Explaining Differences in Parental Outcomes Between Samples

We conducted analyses of covariance (ANCOVAs) using sample type and all covariates as predictors for the respective outcome variable. We then compared the adjusted effect estimates for sample type from the ANCOVA with the unadjusted effect estimates from an ANOVA that did not include any covariates. When in the ANCOVA the equality of error variances assumption was violated, we calculated parameter estimates with robust standard errors for the ANCOVAs. We could not find any considerable differences in the results, and therefore, we decided to present the results of the common and well-known standard methods. The size of the adjusted difference in outcomes between types of samples (adjusted effect estimate) indicates how much the covariates describe the difference in study results between the 2 samples. If effect estimates of the type of sample on the outcome variable decrease or even disappear after controlling for covariates, the considered covariates explain partly (in case of just decrease) or fully (in case of disappearance) the differences between samples for the respective outcome. Herein, we compared the effect sizes (η^2) of the sample indicator using ANOVA, with the covariate-adjusted effect sizes (η^2) of the sample indicator from the ANCOVA.

For an increase in child maltreatment (0=no increase and 1=increase), binary logistic regressions for predicting each subtype of child maltreatment from study type (and covariates) were used. Odds (exp(B)) of the type of sample adjusted for all covariates were compared to unadjusted odds (exp(B)) with study type being the only predictor. If the odds of sample type influencing the outcome variable decrease or approach 1 after controlling for covariates, it indicates that the covariates partially (if the odds decrease) or fully (if the odds approach 1) explain the differences between the samples for the respective outcome. Due to estimation requirements, we only reported on covariate-adjusted outcome differences between studies for categories of categorical outcomes with an absolute frequency >30. All analyses were performed using SPSS (version 29; IBM Corp).

Results

In the following, the results of the convenience sample are compared with the results of the specific representative sample. Note that results are hardly impacted by the weighting [2].

Investigating Differences in Sample Characteristics

We evaluated differences in sociodemographic data, parent-related risk factors, and COVID-19–related experiences.

Sociodemographic Data

Tables 1 and 2 depict the comparison of sociodemographic characteristics of the 2 samples (convenience vs specific representative). In the convenience sample, 4967 people, primarily women (n=4349, 87.56%), participated with a mean age of 39.11 (SD 6.25; range 20-70) years. The average number of children in the household was 1.82 (SD 0.78; range 1-11). These children were on average aged 6.46 (SD 4.25; range 0-17) years. In 42.5% (2111/4967) of the convenience sample, the families lived in Berlin. The distribution of the participants among the federal states and a comparison between the samples is shown in Table S1 in Multimedia Appendix 1. In most cases, participants worked either part time (2051/4967, 41.29%) or full time (1778/4967, 35.8%). Most participants were highly educated with a high school education (4305/4967, 86.67%) or a university degree (3390/4967, 68.25%), and they had a high socioeconomic status (3165/4967, 63.72%).

The specific representative sample comprised 1024 participants (n=534, 52.14% women) with a mean age of 40.89 (SD 8.17; range 18-73) years. On average, 1.69 (SD 0.77; range 1-6) children aged <18 years lived in the household. The mean age of the children was 8.92 (SD 4.92; range 0-17) years. In total, 4.3% (44/1024) of the participants lived in Berlin. In 55.66% (570/1024) of the cases, the parent worked full time, and in 27.64% (283/1024) of the cases, they worked part time. Most participants had a middle school education (480/1024, 46.88%), had completed an apprenticeship (586/1024, 57.23%), and had a middle socioeconomic status (585/1024, 57.13%).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of the convenience sample and the specific representative sample—parent data.

	Convenience sample 2020, (N=4967), n (%)	Specific representative sample 2020[a], (N=1024), n (%)	Comparison, test statistics	*P* value	Effect size
Parental sex (female), n (%)	4349 (87.6)	534 (52.1)	χ^2_1=706.2	<.001	ϕ=0.34
Biological parent, n (%)	4913 (97.8)	979 (95.6)	χ^2_1=57.1	<.001	ϕ=0.10
Single parent, n (%)	504 (10.1)	123 (12)	χ^2_1=3.2	.08	ϕ=0.02
Nationality, n (%)					
German	4803 (94.7)	1001 (97.8)	χ^2_1=3.1	.08	ϕ=0.02
Other	270 (5.3)	23 (2.2)	χ^2_1=16.3	<.001	ϕ=0.05
Age (years), mean (SD)	39.11 (6.25)	40.89 (8.17)	$t_{1218.10}$=6.58	<.001	*d*=0.27
Marital status: in a relationship, n (%)	4568 (92.9)	917 (89.6)	χ^2_1=13.3	<.001	ϕ=0.05
Federal State of Germany: city-state, n (%)	2263 (45.6)	75 (7.3)	χ^2_1=521.6	<.001	ϕ=0.30
School education, n (%)					
Low (up to 9 years of schooling)	58 (1.2)	109 (10.7)	χ^2_1=281.4	<.001	ϕ=–0.22
Middle (10 years of schooling)	564 (11.3)	480 (42.5)	χ^2_1=744.4	<.001	ϕ=–0.35
High (up to 13 years of schooling)	4305 (86.7)	426 (41.6)	χ^2_1=1038.4	<.001	ϕ=0.42
No or other education	40 (0.7)	9 (5.2)	χ^2_1=0.1	.81	ϕ=–0.003
Professional education, n (%)					
Apprenticeship	1041 (21.0)	586 (57.2)	χ^2_1=564.5	<.001	ϕ=–0.31
Technical school	308 (6.2)	119 (11.6)	χ^2_1=37.7	<.001	ϕ=–0.08
University	3390 (68.3)	247 (24.1)	χ^2_1=693.1	<.001	ϕ=0.34
No or other education	228 (4.6)	72 (7)	χ^2_1=10.6	<.001	ϕ=–0.04
Current employment status: working, n (%)	3983 (80.2)	863 (84.3)	χ^2_1=9.2	<.001	ϕ=–0.04
Socioeconomic status index [b], n (%)			$t_{1268.97}$=–4.05	<.001	*d*=–.17
Low	116 (2.3)	100 (9.8)			
Middle	1686 (33.9)	585 (57.5)			
High	3165 (63.7)	338 (33)			
Household					
Balcony, n (%)	3125 (62.9)	598 (58.4)	χ^2_1=7.4	.007	ϕ=–0.04
Garden, n (%)	2969 (59.8)	768 (75.0)	χ^2_1=83.9	<.001	ϕ=–0.12
Rooms, mean (SD)	*4.31 (3.13)*	*4.87 (4.42)*	$t_{1242.91}$=3.84	<.001	*d*=0.16
Other persons in the household, n (%)	4357 (87.7)	914 (89.3)	χ^2_1=1.9	.17	ϕ=–.02

[a]Weighted data.

[b]Index calculated according to the Winkler Index [81].

XSL•FO
RenderX

Table 2. Sociodemographic characteristics of the convenience sample and the specific representative sample—child data.

	Convenience sample 2020 (N=4967), n (%)	Specific representative sample 2020[a] (N=1024), n (%)	Comparison, test statistics	*P* value	Effect size
Number of children			$t_{1492.82}$=–4.90	<.001	*d*=–0.17
1	1813 (36.65)	474 (46.3)			
2	2417 (48.7)	427 (41.7)			
≥3	737 (14.8)	122 (12)			
Child's age (y)			$t_{1329.24}$=18.06	<.001	*d*=0.70
0-2	1953 (39.3)	230 (22.3)			
3-5	2380 (48)	283 (27.4)			
6-12	3634 (73.1)	713 (69.6)			
13-17	1046 (21.1)	506 (49.5)			
Child's sex (female)	4387 (48.7)	862 (49.9)	—[b]	—	—
Children living in 2 households	155 (3.1)	27 (2.6)	χ^2_1=0.7	.41	ϕ=–.01
Previous contact with social and family services	1197 (24.1)	249 (24.3)	χ^2_1=0.2	.88	ϕ=–.01

[a]Weighted data.

[b]Not applicable.

The 2 samples differed substantially in relevant characteristics. The convenience sample had a significantly higher number of women, the children were younger, participants were more frequently from a city-state (ie, Berlin), and had a higher school and professional education than the specific representative sample (all differences with medium effect sizes). Compared to the specific representative sample, in the convenience sample, we observed small differences in the following variables: participants were significantly younger, had more children aged <18 years in the household, had a higher socioeconomic status, were of another nationality than German, were more frequently biological parents and in a relationship, had fewer rooms in the household, and had less frequently a garden. Samples were comparable regarding being a single parent, child sex, current employment status, having a balcony, children living in 2 households (eg, in case of separated parents), and having others in the household or previous contact with social family services.

Parent-Related Risk Factors

Table 3 shows parent-related risk factors. Significantly fewer parents from the convenience sample reported belonging to the risk group for severe COVID-19. A significantly higher share of parents from the convenience sample reported a current or previous mental disorder and a history of own child abuse or neglect. Furthermore, a significantly higher share of parents from the convenience sample reported March and April 2020 as the most stressful months. In contrast, a significantly higher share of parents from the specific representative sample reported that "all months were equally stressful" and "no month was especially stressful." The effect-sizes were small.

Table 3. Parent-related risk factors: descriptive data and comparison between the samples.

	Convenience sample 2020 (N=4967), n (%)	Specific representative sample 2020[a] (N=1024), n (%)	Comparison		
			χ^2 (*df*)	*P* value	ϕ
Parental risk of alcohol abuse[b]	317 (6.38)	56 (5.47)	1.2 (1)	.27	0.01
Parental mental disorder	864 (17.39)	107 (10.45)	30.2 (1)	<.001	0.07
Chronic severe health condition	604 (12.16)	146 (14.26)	3.4 (1)	.07	–.02
Parental history of child abuse or neglect	1562 (31.45)	238 (23.24)	27.0 (1)	<.001	0.07
Parental experience of violence in adulthood	601 (12.10)	114 (11.13)	0.8 (1)	.39	–.01

[a]Weighted data.

[b]n=3430 (convenience sample) and n=700 (specific representative sample) parents indicated that they regularly drink alcohol. n=317 (convenience sample) and n=56 (specific representative sample) were at risk for alcohol abuse according to Patient Health Questionnaire-D [83].

XSL•FO
RenderX

COVID-19–Related Experiences

Table 4 shows the COVID-19–related experiences. In both samples, nearly a quarter of the participants reported reduced working hours (convenience sample: 1197/4967, 24.1% and specific representative sample: 277/1024, 27.05%), and one-fifth of the participants reported a significant financial loss due to COVID-19 (convenience sample: 1111/4967, 22.37% and specific representative sample: 221/1024, 21.58%). In addition, April 2020 (convenience sample: 3508/4967, 70.62% and specific representative sample: 487/1024, 47.56%) and May 2020 (convenience sample: 2975/4967, 59.89% and specific representative sample:430/1024, 42%) were the most stressful months until August 2020 in both samples.

Table 4. COVID-19–related experiences: descriptive data and comparison.

	Convenience sample 2020 (N=4967), n (%)	Specific representative sample 2020[a] (N=1024), n (%)	Comparison		
			$\chi^2(df)$	*P* value	ϕ
Effects of the COVID-19 pandemic on the health situation					
Family or household member infected with COVID-19	93 (1.87)	22 (2.15)	0.3 (1)	.56	–0.01
A family or household member was hospitalized due to COVID-19	15 (0.30)	8 (0.78)	5.0 (1)	.02	–0.03
Family or household member passed away due to COVID-19	13 (0.26)	4 (0.39)	0.5 (1)	.48	0.01
Parent belongs to the risk group for severe COVID-19	328 (6.60)	103 (10.06)	15.3 (1)	<.001	0.05
Effects of the COVID-19 pandemic on job situation					
Reduced working hours	1197 (24.10)	277 (27.05)	4.0 (1)	.046	0.03
Job loss	235 (4.73)	55 (5.37)	0.8 (1)	.39	0.01
Significant financial loss	1111 (22.37)	221 (21.58)	0.3 (1)	.58	–0.01
Most stressful months					
February	113 (2.28)	21 (2.05)	0.2 (1)	.66	0.01
March	2207 (44.43)	268 (26.17)	116.8 (1)	<.001	0.14
April	3508 (70.63)	487 (47.56)	203.3 (1)	<.001	0.18
May	2975 (59.90)	430 (41.99)	110.9 (1)	<.001	0.14
June	1634 (32.90)	263 (25.68)	20.4 (1)	<.001	0.06
July	582 (11.72)	117 (11.43)	0.1 (1)	.79	0.003
August	307 (6.18)	46 (4.49)	4.4 (1)	.04	0.03
All months were equally stressful	288 (5.80)	125 (12.21)	54.3 (1)	<.001	–0.10
No month was especially stressful	381 (7.67)	186 (18.16)	109.1 (1)	<.001	–0.14

[a]Weighted data.

All 34 covariates were simultaneously considered in a binomial logistic regression for the prediction of sample membership (0=specific representative sample and 1=convenience sample). Explained variance in outcomes by the predictors was large (Nagelkerke R^2=0.56) and statistically significant from 0 (χ^2_{36}=2411.4; *P*<.001). The overall accuracy percentage in classification was 89.6%, with a sensitivity of 96.3% and a specificity of 57.2%. The 5 most important predictors for convenience sample membership were as follows: living in a federal state in Germany (Berlin, Bremen, or Hamburg; odds ratio [OR] 14.75, 95% CI 10.93-19.91), parental sex female (OR 11.13, 95% CI 8.89-13.94), parental mental disorder (OR 2.18, 95% CI 1.63-2.92), being a biological parent (OR 2.05, 95% CI 1.11-3.79), and higher education (professional education [university]: OR 2.02, 95% CI 1.33-3.08 and school education [high school]: OR 2.01, 95% CI 0.77-5.23). All model coefficients and odds ratios can be found in Table S2 in Multimedia Appendix 1.

Investigating Differences in Parental Outcomes Between Samples

Differences in parental outcomes between the convenience sample and the specific representative sample are displayed in Table 5 for parental outcomes and Tables 6 and 7 for child maltreatment outcomes.

XSL•FO
RenderX

Table 5. Descriptive data: investigating and explaining differences between the convenience and the specific representative samples in parental outcomes.

	Convenience sample (N=4967), mean (SD)[a]	Specific representative sample[b] (N=1024 [weighted]), mean (SD)[a]	Comparison between the samples, ANOVAs (unadjusted)			ANCOVAs[c] (adjusted)		
			F test (*df*)	*P* value	η^2	*F* test (*df*)	*P* value	η^2
Parental stress: January 2020	36.07 (8.67)	34.72 (10.63)	18.35 (1)	<.001	0.003	0.33[d]	.56	<0.001
Parental stress: time of the highest burden	41.37 (10.50)	36.93 (10.45)	149.97 (1)	<.001	0.024	14.92[e]	<.001	0.002
Subjective health: January 2020	7.36 (1.76)	7.34 (2.04)	0.03 (1)	.87	<0.001	0.06[f]	.81	<0.001
Subjective health: time of the highest burden	6.06 (2.27)	6.80 (2.21)	101.13 (1)	<.001	0.016	21.69[g]	<.001	0.004
Patient Health Questionnaire-4 (anxiety and depression symptoms)	4.37 (3.0)	2.52 (2.63)	354.94 (1)	<.001	0.055	97.98[h]	<.001	0.016
General stress	7.40 (3.73)	5.28 (4.13)	4023.2 (1)	<.001	0.044	84.21[i]	<.001	0.014
Pandemic-related stress	31.80 (8.03)	31.97 (10.96)	0.56 (1)	.45	<0.001	6.75[j]	.009	0.001

[a]Unadjusted significant covariates.

[b]Weighted data.

[c]ANCOVA: analysis of covariance.

[d]Biological parent, federal state, current employment status, previous contact with social and family services, parental mental disorder, parental history of child abuse or neglect, children's age, parental age, socioeconomic status, number of children in the household, and parental risk of alcohol abuse.

[e]Parental sex, biological parent, federal state, current employment status, previous contact with social and family services, parental mental disorder, parental history of child abuse or neglect, children's age, parental age, socioeconomic status, and number of children in the household.

[f]Nationality: German, previous contact with social and family services, family or household member had a COVID-19 infection, family or household member was hospitalized due to COVID-19, parental chronic severe health condition, parental mental disorder, parent belongs to the risk group for severe COVID-19, parental history of child abuse or neglect, socioeconomic status, and parental risk of alcohol abuse.

[g]Parental sex, federal state, previous contact with social and family services, significant financial loss, parental chronic severe health condition, parental mental disorder, parental history of child abuse or neglect, children's age, and parental risk of alcohol abuse.

[h]Parental sex, federal state, single parent, previous contact with social and family services, the availability of a garden, family or household member admitted to the hospital with COVID-19, significant financial loss, parental mental disorder, parental history of child abuse or neglect, parental experience of violence in adulthood, children's age, parental age, number of children in the household, parental risk of alcohol abuse, and number of persons in the household.

[i]Parental sex, federal state, current employment status, previous contact with social and family services, significant financial loss, reduced working hours, job loss, parental mental disorder, parent belongs to the risk group for severe COVID-19, parental history of child abuse or neglect, parental experience of violence in adulthood, children's age, socioeconomic status, child having another main place of residence, and parental risk of alcohol abuse.

[j]Parental sex, current employment status, previous contact with social and family services, significant financial loss, reduced working hours, job loss, parental mental disorder, parent belongs to the risk group for severe COVID-19, parental experience of violence in adulthood, children's age, number of children in the household, and number of persons in the household.

XSL•FO
RenderX

Table 6. Comparison of the occurrence and frequency of adverse childhood experiences between the convenience sample and the specific representative sample.

	Occurrence				Increase during the COVID-19 pandemic			
	Values, n (%)	χ^2 (*df*)	*P* value	ϕ	n (%)	χ^2(*df*)	*P* value	ϕ
Severe stressful living conditions		0.1 (1)	.75	0.004		0.4 (1)	.52	0.01
Web-based sample	307 (6.2)				129 (2.6)			
Representative sample	66 (6)				23 (2)			
Physical abuse		55.8 (1)	<.001	0.10		38.5 (1)	<.001	0.08
Web-based sample	696 (14.1)				329 (6.6)			
Representative sample	57 (5.6)				17 (2)			
Nonverbal emotional abuse		7.0 (1)	.03	0.03		10.5 (1)	.001	0.04
Web-based sample	470 (10)				356 (7.1)			
Representative sample	82 (8)				45 (4)			
Supervisory neglect		3.1 (1)	.08	–0.02		11.3 (1)	<.001	0.04
Web-based sample	559 (11.3)				326 (6.6)			
Representative sample	96 (9.4)				39 (4)			
Emotional neglect		24.7 (1)	<.001	0.06		0.8(1)	.37	–0.01
Web-based sample	497 (10)				284 (5.7)			
Representative sample	157 (15.3)				66 (6)			
Verbal emotional abuse		91.5 (1)	<.001	0.12		137.83	<.001	0.15
Web-based sample	2423 (48.82)				1585 (31.91)			
Representative sample	332 (32.4)				140 (13.7)			
Witness of domestic violence		107.9 (1)	<.001	0.13		116.91 (1)	<.001	0.14
Web-based sample	2488 (50.22)				1236 (24.91)			
Representative sample	332 (32.4)				97 (10)			
Sexual abuse		15.2 (1)	<.001	0.05		__[a]	—	—
Web-based sample	19 (0.4)				2 (0.04)			
Representative sample	14 (1.4)				3 (0.3)			
Physical neglect		8.4 (1)	.02	0.04		—	—	—
Web-based sample	19 (0.4)				8 (0.2)			
Representative sample	11 (1)				3 (0.3)			

[a]Inference statistics are only reported for cell counts ≥5 in all cells.

XSL•FO
RenderX

Table 7. Logistic regression analysis: investigating and explaining differences in the increase of child maltreatment outcomes between the convenience and the specific representative samples.

	Unadjusted			Adjusted		
	B (SE)	*P* value	OR[a] (95% CI)	B (SE)	*P* value	OR (95% CI)
Verbal emotional abuse	1.08[b] (0.10)	<.001	2.95 (2.44-3.56)	0.72[c] (0.12)	<.001	2.05[d] (1.63-2.57)
Witnessing domestic violence	1.16[e] (0.11)	<.001	3.19 (2.56-3.97)	0.70[f] (0.13)	<.001	2.02[g] (1.55-2.63)
Nonverbal emotional abuse	0.50[h] (0.16)	.002	1.65 (1.20-2.27)	0.53[i] (0.20)	.01	1.70[j] (1.15-2.52)
Emotional neglect	–0.14[k] (0.14)	.33	0.87 (0.66-1.148)	–0.10[l]	0.19 (0.59)	0.90[m] (0.62-1.30)

[a]OR: odds ratio.

[b]Model fit: χ^2_1=153.5; *P*<.001; Nagelkerke R^2=0.04.

[c]Model fit: χ^2_{37}=589.6; *P*<.001; Nagelkerke R^2=0.14; Hosmer and Lemeshow χ^2_8=10.3; *P*=.25.

[d]Covariates with a significant partial regression coefficient: Sample membership, parental sex, number of children in the household, parental age, children's age, federal state, current employment status, previous contact with social and family services, the availability of a balcony, reduced working hours, job loss, parental mental disorder, parental history of child abuse or neglect, school and professional education, and child's sex.

[e]Model fit: χ^2_1=137.1; *P*<.001; Nagelkerke R^2=0.04.

[f]Model fit: χ^2_{37}=522.7; *P*<.001; Nagelkerke R^2=0.13; Hosmer and Lemeshow χ^2_8=6.9; *P*=.54.

[g]Covariates with a significant partial regression coefficient: Sample membership, parental sex, number of children in the household, child's age, single parent, material status, current employment status, previous contact with social and family services, significant financial loss, parental mental disorder, parental history of child abuse or neglect, and professional education: university.

[h]Model fit: χ^2_1=10.8; *P*=.001; Nagelkerke R^2=0.01.

[i]Model fit: χ^2_{37}=248.6; *P*<.001; Nagelkerke R^2=0.11; Hosmer and Lemeshow χ^2_8=11.1; *P*=.20.

[j]Covariates with a significant partial regression coefficient: Number of children in the household, children's age, federal state, current employment status, previous contact with social and family services, the availability of a garden, significant financial loss, parental mental disorder, parental history of child abuse or neglect, parent belongs to the risk group for severe COVID-19, and professional education: apprenticeship.

[k]Model fit: χ^2_1=0.9; *P*=.33; Nagelkerke R^2≤0.001.

[l]Model fit: χ^2_{37}=96.9; *P*<.001; Nagelkerke R^2=0.05; Hosmer and Lemeshow χ^2_8=5.2; *P*=.74. Covariates with a significant partial regression coefficient.

[m]Covariates with a significant partial regression coefficient: Previous contact with social and family services, significant financial loss, and parental mental disorder.

Parental Stress

In January 2020, parents from the convenience sample reported significant but negligibly more parental stress (η^2=0.003). However, at the time of the highest burden, parents from the convenience sample reported significantly higher parental stress than parents from the specific representative sample, with a small effect (η^2=0.024).

Subjective Health

In January 2020, there was no significant difference (η^2<0.001) between parents of the 2 samples regarding subjective health (η^2<0.001). Again, at the time of the highest burden, parents from the convenience sample reported significantly poorer subjective health, with a small effect (η^2=0.016).

Mental Health and General Stress

Parents in the convenience sample reported higher overall anxiety and depression symptoms (Patient Health Questionnaire-4) than parents in the specific representative sample at times of the highest burden, with a medium effect size (η^2=0.055). The same pattern applied to general stress, with a medium effect (η^2=0.044).

Pandemic-Related Stress

Samples did not differ in their overall pandemic-related stress (η^2<0.001).

Child Maltreatment

In both samples, the lifetime occurrence of WDV (convenience sample: 2488/4966, 50.2% and specific representative sample: 332/1024, 32.42%) and VEA (convenience sample: 2423 4965, 48.82% and specific representative sample: 3321023, 32.4%) were the most frequent subtypes (Table 6). In the convenience sample, a higher share of parents reported the lifetime occurrence of VEA (ϕ=0.12), WDV (ϕ=0.13), NEA (ϕ=0.03), physical abuse (ϕ=0.10), and emotional neglect (ϕ=0.06). No differences were found for severe stressful living conditions, NEA, or supervisory neglect. The odds of a child experiencing an increase in one of the following subtypes of child maltreatment were higher in the convenience sample: for VEA (convenience sample: 1585/4966, 31.91% and specific representative sample: 140/1024, 13.7%), it was 2.95 times higher; for WDV (convenience sample: 1236/4961, 24.91% and specific representative sample: 97/1024, 9.47%), it was 3.19 times higher; and for NEA (convenience sample: 356/4966, 7.17% and specific representative sample: 45/1024, 4.39%), it was 1.65 times higher during the COVID-19 pandemic (Table 7). The odds (0.87) of an increase in the frequency of emotional

XSL•FO
RenderX

neglect were not significantly different between the samples (convenience sample: 284/4967, 5.72% and specific representative sample: 66/1024, 6.45%).

Explaining Differences in Parental Outcomes Between the 2 Samples

To explain differences in parental outcomes, we calculated ANCOVAs with sociodemographic data, parent-related risk factors, and COVID-19–related experiences as covariates. Only significant covariates are listed in Tables 5 and 7.

Parental Stress

At the time of the highest burden, the difference in parental stress between the samples remained significant but was negligible in size (η^2=0.002) after controlling for sociodemographic data, parent-related risk factors, and COVID-19–related experiences. The covariates explained 92% of the difference in the outcome between the studies. The difference in parental stress between the samples in January was still not significant, and the effect size was negligible (η^2<0.001). As there were no differences between samples, the covariates cannot explain any of them. It must be noted that the covariates also do not induce any relevant differences.

Subjective Health

The difference between the samples at the time of the highest burden was significant but negligible in size (η^2=0.004) after controlling for sociodemographic data, parent-related risk factors, and COVID-19–related experiences. The covariates explained 75% (at the time of the highest burden) of the difference in outcome between the samples. The difference between the samples in January regarding subjective health was still not significant, and the effect size was negligible (η^2<0.001). As there were no differences between samples, the covariates cannot explain any of them. It must be noted that the covariates also do not induce any relevant differences.

Mental Health and General Stress

The difference between the samples regarding mental health was small (η^2=0.016) but significant after controlling for sociodemographic data, parent-related risk factors, and COVID-19–related experiences. The covariates explained 71% of the difference in the outcome between the samples. The same pattern applied to general stress (η^2=0.014). In total, 68% of the difference in the outcome between the samples was explained by the covariates.

Pandemic-Related Stress

The difference between the samples regarding pandemic-related stress was significant, and the effect size slightly grew but was still negligible (η^2=0.001) after controlling for sociodemographic data, parent-related risk factors, and COVID-19–related experiences. As there are no differences between studies, the covariates cannot explain any of them.

Child Maltreatment

There are large and significant differences in child maltreatment between the 2 samples in all subcategories except emotional neglect. The odds of experiencing VEA, WDV, and NEA were between 1.65 and 3.19 times higher in the convenience sample as compared to the specific representative sample. The difference in VEA and WDV between the samples could partly be explained by the covariates. After controlling for sociodemographic data, parent-related risk factors, and COVID-19–related experiences, the odds ratio of a child experiencing an increase in VEA and WDV in the convenience sample compared to the specific representative sample were 2.05 (as compared to 2.95 before adjustment) and 2.02 (as compared to 3.19 before adjustment), respectively. The differences in NEA between the samples could not be explained by the covariates. As there were no differences in emotional neglect between samples, there were no effect differences to be explained by the covariates. It must be noted that the covariates also do not induce any relevant differences.

Discussion

Principal Findings

So far, research has paid little attention to the representativeness of samples. However, this is crucial for drawing valid inferences on the population of interest. This is especially important in areas in which far-reaching decisions are made based on results of a study. This is specifically prevalent in studies on psychological consequences of the COVID-19 pandemic, which made immediate decisions necessary. In this paper, we showed that sampling strategy can have a great impact on the results of a study. We relied on data investigating parental outcomes in the COVID-19 pandemic of 2 web-based studies that differ in their sampling scheme, one being a typical convenience sample (N=4967) and the other a typical specific representative sample drawn by a survey institute (N=1024).

Our first aim was to investigate the impact of sampling strategy on the results. For this, we examined differences in results between the 2 samples. Our second aim was to investigate reasons for these differences. For this, we assessed (1) to what extent the studies differ in sample characteristics and to what extent the sample characteristics may explain group membership and (2) to what extent differences in parental outcomes between the 2 samples can be explained by sample characteristics (sociodemographic data, parent-related risk factors, and COVID-19–related experiences).

Discussion of Results

Differences in Sample Characteristics

We found that the convenience sample differed from the specific representative sample in sociodemographic composition, parent-related risk factors, and COVID-19–related experiences. Sociodemographic data, parent-related risk factors, and COVID-19–related experiences predicted sample membership well. Thus, this study identified a set of variables that partly explain differences in study results and explain differences in the sample selection.

As expected, the sociodemographic composition of the 2 study samples (convenience and specific representative sample) differed in several ways (eg, overrepresentation of the female

XSL•FO
RenderX

sex, younger parental and child age, a higher number of children, higher socioeconomic status, and higher education level in the convenience sample). The differences in the sociodemographic composition of the convenience sample were in line with previous findings [29-32,34-39]. Whitaker et al [33] also reported an overrepresentation of White, female individuals, and younger individuals with higher income and educational status compared to the general population in some studies in their literature review on Facebook recruitment. The overrepresentation of women in convenience samples is likely to be especially pronounced in studies related to mental and physical health [33,69] as well as parenting and family [35,38,80,92-94]. In various convenience samples evaluating parenthood during the COVID-19 pandemic, mothers constituted the majority [4,9,14-17,19,22,27,40-42,44,45]. However, the effects of this overrepresentation have not yet been empirically examined. Furthermore, in many of these convenience samples, most participants reported higher education [4,14,16,17,27,41,42,45-47]. One reason could be that women [95,96], younger people with younger children, and people with higher educational qualifications may more readily participate in web-based research and are thus more likely to be reached for study purposes [80,95,97]. Individuals with high educational attainment may be more familiar and more interested in research. In addition, the use of personal emails and invitations for recruitment may have resulted in a participant pool primarily from academic circles. The overrepresentation of women, along with younger parents and those with more children, could also be attributed to thematic self-selection. This group of parents, who were particularly stressed during the pandemic [4,14,59], may have been more motivated to share their experiences of hardship [30]. In summary, the sample composition of our convenience sample is comparable to many other samples recruited during the COVID-19 pandemic.

Differences in Parental Outcomes

Results on parental outcomes during the COVID-19 pandemic differed notably between the 2 samples. Overall, parents from the convenience sample were more burdened during the COVID-19 pandemic than parents from the specific representative sample. Compared with parents from the specific representative sample, parents from the convenience sample reported significantly more parental stress at the time of the highest burden, less subjective health, more anxiety and depression symptoms overall (Patient Health Questionnaire-4), more general stress, more often the occurrence of child maltreatment (VEA, WDV, physical abuse, NEA, and emotional neglect), and an increase of child maltreatment during the COVID-19 pandemic (VEA, WDV, and NEA). These results corroborate the claim that sampling has considerable impact on the results [31].

Our results are in line with findings from previous studies. Studies show that people with higher burdens are more likely to participate in web-based surveys than those with lower burdens [25,39,77,78]. The reasons may be a higher motivation to share one's burden and a thematic awareness of the topic due to one's own burden [30]. A higher amount of mental health burden in web-based studies with convenience sampling compared to representative sampling has also been observed in other studies with samples of parents [39,80] before the COVID-19 pandemic. Lawson et al [98] showed that the recruitment of web-based convenience samples influenced the prevalence of child abuse. Joyal-Desmarais et al [79] also found differences between samples in >70% of their variables when comparing their web-recruited convenience samples with representative surveys regarding preventive behaviors.

Explaining Differences in Parental Outcomes Between Studies

Differences in parental outcomes between the 2 samples could (partially) be explained by differences in sociodemographic data, parent-related risk factors, and COVID-19–related experiences between the samples. Consequently, at least for this study, we seem to have identified relevant confounders for the difference in effects between the 2 samples.

On the basis of our results, it is initially not possible to say which covariates are explicitly responsible for the differences. However, we can identify covariates that differed between the 2 samples: parental sex and age, children's age, federal state (ie, city-state), school and professional education, number of children in the household, socioeconomic status, nationality, relationship status, biological parents, rooms and garden in the household, risk group of severe COVID-19, current or previous mental disorder, and history of own child abuse or neglect. Other studies identified sociodemographic variables as correlates for poorer mental health outcome (eg, female sex, younger parental and children age, parental psychological or physical health conditions before the COVID-19 pandemic, and lower financial background and income or financial stress) [8,11,16,40-42,48-60]. Furthermore, sociodemographic variables were also related to higher parental stress (eg, more children in the household, younger age of the children, and being a single parent) [4,14,17,20,41,46,59], and the occurrence of child maltreatment and parent-child conflicts (eg, younger parental and children age, job and financial loss, living in an urban area, and having low income or financial concerns) [1,2,43,46,47,59,98-100] for parents during the COVID-19 pandemic. Our convenience sample showed a higher presence of individuals with sociodemographic risks (such as more female individuals, younger age, and more and younger children), potentially contributing to greater disparities in parental outcomes due to an overrepresentation of these identified risk factors. Thus, these sociodemographic risk factors may explain differences in parental outcomes.

Strengths

Investigation of the Impact of Sampling

So far, research has largely overlooked the impact of sampling on the informativeness of the results. With this paper, we present one of the few studies that highlight and evaluate the impact of sampling on the results of studies in clinical psychology.

Availability of 2 Studies With the Same Measurements

We were able to use 2 samples in which not only the same outcome measures but also the same explaining variables were measured in exactly the same way. This is a rather rare case, but it was essential for answering the research question. This

XSL•FO
RenderX

also allowed us to identify variables explaining these differences. These factors could be assessed in future research on parental outcomes during pandemics to (1) understand which population groups a study provides information for and (2) for understanding differences in study results. Note that while we compared 2 typical sampling approaches—a convenience sample and a specific representative sample, as commonly used by survey institutes—this comparison provided insights into how these approaches lead to different results. However, this type of design and analysis is not limited to these specific sampling methods. The results of any sampling approaches, including 2 different convenience samples, can be compared in this way. The key is that the same variables are assessed in both studies.

Explaining Differences in Results

We not only investigated the differences in results between studies due to different sampling approaches but also aimed to understand why studies yield different results. The collection of a comprehensive set of explaining variables in both studies allowed us to explain most of the differences in study results. This is remarkable, as usually this is often not the case (eg, [31]). In many studies, only a few sociodemographic variables are available that are similarly assessed in the different studies. These usually cannot explain differences in results very well [101]. With sociodemographic data, parent-related risk factors, and COVID-19–related experiences, we had powerful explanatory variables for the differences in study results. If, as in many other studies, we had only limited sociodemographic data, we likely would have explained much less of the differences in the results.

Strategy for Investigating Differences in Results Across Studies

Our paper also presents a strategy for investigating reasons for differences in results across different studies. This strategy can also be applied to compare different sampling strategies in other fields. Thus, researchers may at the same time conduct studies with different types of sampling strategies. This may also be 2 different convenience samples. Relevant explaining variables need to be collected in the same way in both studies to analyze to what extent these can explain differences in outcomes between the 2 studies.

Limitations and Outlook

Limitations of the Comparison of the Studies

Our aim was to investigate the impact of the sampling strategy on the study results on parental outcomes. While we could make use of the rare case of having 2 studies with different sampling strategy that were almost identical in their design and setup, not all design aspects were identical. First, while participants in the convenience sample were interviewed via web surveys (CAWI), parents in the specific representative sample were interviewed either via CAWI or CATIs. This could have led to differences in response behavior, especially with regard to sensitive topics (eg, adverse childhood experiences and parental mental disorder) due to social desirability. Second, the 2 studies partly differed with respect to the incentives they received. Parents in the convenience sample and parents in the specific representative study that were interviewed via CAWI may have been particularly motivated by the prospect of incentives (money, vouchers, and Ipad). In contrast, parents in the specific representative study that were interviewed via telephone did not receive any incentives. This could have also influenced response behavior, leading to differences in study results. In summary, in addition to differences in the sample composition, there are further explanations for differences in the results between the samples, such as different modes of data collection in the representative sample (CATI and CAWI) versus the convenience sample (only CAWI) and different motivations of participants (no incentives for participants in CATI in the quota sample vs incentives for participants in CAWI and the convenience sample). These were not controlled for and must be considered in the final evaluation of the study results. While we could explain a large part of the differences in results between the 2 studies, some unexplained differences remained. These may be due to unobserved explanatory variables, other factors that could account for both sample differences and variations in parental outcomes, or differences in the study designs.

From our study, we cannot necessarily conclude that the specific representative sample results in more valid conclusions on the underlying population than the convenience sample. If selection bias is stronger on relevant variables not controlled for in the specific representative sample as compared to the convenience sample, bias could even be larger in the specific representative study. What we surely can conclude is that the sampling strategy does have a large impact on the study results. For the specific representative study, we can rule out that sample selection due to strata variables biases the results in this study.

Limitations of Single Studies

While both studies were conducted rigorously with the aim of accurately inferring parental outcomes in the population, each study has its own limitations. First, while the 2 studies aimed at drawing inferences on parents with underaged children in Germany, due to practical reasons, they had to restrict the sample to German-speaking parents, which excludes a part of the population [102]. Second, although anonymity was assured by different measures, we cannot rule out that participants answered in a socially desirable way, refraining from reporting, for example, child maltreatment. Though we tried to reduce this, it needs to be kept in mind when interpreting the results. All data were collected and evaluated anonymously. The researchers did not have the contact details of the participants available at any time. This was also explained to the participants in the informed consent form. Overall, our lifetime prevalence rates of child maltreatment are in line with the prevalence rates found in the literature [103]. Third, neither of the 2 samples is globally representative of the general population and is, thus, limited in generalizing to the overall population. While specific representativeness for some variables can be assured for the specific representative study, selection bias may still exist for all other variables. Note that in this specific setting, it would have been possible to weight the convenience sample using census data, thereby achieving representativeness for this sample as well. As compared to the quota sample, specific representativity in the convenience sample would be solely achieved by poststratification. As our aim was to investigate

XSL•FO
RenderX

the impact of sampling on study results, we did not consider this here.

Even if their results cannot necessarily be generalized to the underlying population, both the specific representative as well as the convenience samples may provide valuable insights if the sample is well described with respect to variables relevant to the construct under investigation. Adding up information from many studies (that may all be convenience samples) does help to build theory and to describe for which type of persons the theory has been tested for. For this, it is important that relevant person characteristics of the sample are comprehensively described. These should go beyond just some demographic variables and should consider the variables that are relevant to the phenomena of interest. Further, if we know the distribution of these variables in the population, we can infer population-level results from a convenience sample (eg, by weighting the observations in the convenience sample). This approach is similar to how election results are predicted [104].

Generalizability of Our Results

As with any research study, it is not possible to generalize findings from a single investigation to all other studies. Therefore, the results presented in this paper may not necessarily apply to other studies in clinical psychology or those focused on parental outcomes. As in other research, our study provides another piece on the topic that together with other studies adds up to a broader picture. We presume that the results of our study also hold for studies that use similar outcomes to those of our study and recruitment or sampling strategies that do not differ massively from ours. The results of this study may not necessarily be the same for other research topics (specifically other outcomes) and other sample selection mechanisms. In a different study, the selection mechanism of participants in the sample may be different, which leads to different results and possibly also different explaining variables that explain the differences. Different research questions within a study lead to different outcome measures. These are most likely impacted by different explaining variables (covariates). While it is not possible to generalize the findings of this study to other studies, this study shows (1) that sampling strategy can have a large impact on the results and (2) how one can investigate reasons for differences in the results. Relevant covariates that explain differences in study results will most probably be those that are strongly related to the outcomes (researchers usually have a good theory about that from the literature in their field) as well as are relevant for participating in a study (research on this is usually a bit scarce, but pilot studies may help). Assessing these variables as covariates is a promising strategy as it leaves less room for unexplained variation in the outcome [101]. Note that this is not tied to having a quota (ie, a specific representative) sample; it can be done with any 2 studies. The requirement is that the same covariates are assessed in both studies. Future studies may want to incorporate this into their study planning.

Conclusions

To allow readers to judge the representativity of a study, detailed information on sampling and recruitment as well as on generalizability should be provided (eg, constraints on generality [105]).

Our results show that the results of studies differ depending on the sampling strategy. We were able to investigate this. Scientists need to be aware of the fact that from a single study, one cannot directly generalize to the population of interest. Full (global) representative studies are the gold standard in research. However, these are often not feasible for various reasons. Therefore, scientists should collect as many relevant explaining variables (covariates) as possible to describe the sample selection. Such variables are usually those that may have an impact on the results. Collecting these variables makes it possible to understand for which subgroups in the population conclusions can be drawn. As our study shows, if the same explaining variables are assessed in different studies, it is also possible to explain differences in the study results by differences in the group characteristics.

Even if researchers just perform a single study with a specific sampling strategy, we recommend collecting data on potentially explaining variables. This (1) allows to judge for which part of the population the results hold and (2) may allow future studies that assess the same explaining variables to explain differences in their results. If researchers even have information on the distribution of these variables in the general population, they may adjust for sampling bias using sampling weights [31] (eg, [12,13,106]). The choice of explanatory variables is crucial for this.

If possible, scientists may want to consider assessing explaining variables and covariates that are (1) relevant to the outcome under consideration and (2) possibly related to the selection into the study. These will help to identify which piece of the puzzle is actually added to the overall theory by this study. It also helps to integrate results over multiple studies in the future. If they have information on the distribution of these variables in the target population, they may even draw a quota sample or weight their convenience sample to assure specific representativeness. In most cases, it will not be possible to assess all relevant explaining variables. However, considering some relevant explanatory variables can already enhance the robustness of the results.

Acknowledgments

This study was funded by a grant by the Berlin University Alliance to SMW and BR. The funding source Berlin University Alliance had no role in the study design, data collection, analysis, or interpretation of data; the writing of the article; or decision to submit the article for publication.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

BR and SMW conceived and designed the study. CC, SMW, BR, and LE contributed to the compilation of measures. LE and CC compiled the literature review. LE and SP analyzed the data. LE wrote the first draft of the paper. All authors critically reviewed the manuscript and contributed to data interpretation and discussion. All authors approved the final manuscript for submission.

Conflicts of Interest

BR and SMW have received funding from the Berlin University Alliance for the study. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

The flyer on the recruitment of the convenience sample, a table about the sample composition regarding the Federal State of Germany of the web-based convenience sample and the representative sample, and a detailed table about the logistic regression analysis to predict sample membership through our sociodemographic data, parent-related risk factors, and COVID-19–related experiences.

[DOCX File , 321 KB - jmir_v27i1e52043_app1.docx]

References

1. Gadermann AC, Thomson KC, Richardson CG, Gagné M, McAuliffe C, Hirani S, et al. Examining the impacts of the COVID-19 pandemic on family mental health in Canada: findings from a national cross-sectional study. BMJ Open 2021 Jan 12;11(1):e042871 [FREE Full text] [doi: 10.1136/bmjopen-2020-042871] [Medline: 33436472]
2. Calvano C, Engelke L, Di Bella J, Kindermann J, Renneberg B, Winter SM. Families in the COVID-19 pandemic: parental stress, parent mental health and the occurrence of adverse childhood experiences-results of a representative survey in Germany. Eur Child Adolesc Psychiatry 2022 Jul;31(7):1-13 [FREE Full text] [doi: 10.1007/s00787-021-01739-0] [Medline: 33646416]
3. A L van Tilburg M, Edlynn E, Maddaloni M, van Kempen K, Díaz-González de Ferris M, Thomas J. High levels of stress due to the SARS-CoV-2 pandemic among parents of children with and without chronic conditions across the USA. Children (Basel) 2020 Oct 21;7(10):193 [FREE Full text] [doi: 10.3390/children7100193] [Medline: 33096787]
4. Marchetti D, Fontanesi L, Mazza C, Di Giandomenico S, Roma P, Verrocchio MC. Parenting-related exhaustion during the Italian COVID-19 lockdown. J Pediatr Psychol 2020 Nov 01;45(10):1114-1123 [FREE Full text] [doi: 10.1093/jpepsy/jsaa093] [Medline: 33068403]
5. Rabe JE, Schillok H, Merkel C, Voss S, Coenen M, De Bock F, et al. [Burden in parents of school-aged children during different phases of the COVID-19 pandemic in Germany: an analysis within the COVID-19 snapshot monitoring (COSMO) study]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Dec;64(12):1500-1511 [FREE Full text] [doi: 10.1007/s00103-021-03453-3] [Medline: 34825928]
6. Alonzo D, Popescu M, Zubaroglu Ioannides P. Mental health impact of the Covid-19 pandemic on parents in high-risk, low income communities. Int J Soc Psychiatry 2022 May;68(3):575-581 [FREE Full text] [doi: 10.1177/0020764021991896] [Medline: 33517821]
7. Achterberg M, Dobbelaar S, Boer OD, Crone EA. Perceived stress as mediator for longitudinal effects of the COVID-19 lockdown on wellbeing of parents and children. Sci Rep 2021 Feb 03;11(1):2971 [FREE Full text] [doi: 10.1038/s41598-021-81720-8] [Medline: 33536464]
8. C Fong V, Iarocci G. Child and family outcomes following pandemics: a systematic review and recommendations on COVID-19 policies. J Pediatr Psychol 2020 Nov 01;45(10):1124-1143 [FREE Full text] [doi: 10.1093/jpepsy/jsaa092] [Medline: 33083817]
9. Hamadani JD, Hasan MI, Baldi AJ, Hossain SJ, Shiraji S, Bhuiyan MS, et al. Immediate impact of stay-at-home orders to control COVID-19 transmission on socioeconomic conditions, food insecurity, mental health, and intimate partner violence in Bangladeshi women and their families: an interrupted time series. Lancet Global Health 2020 Nov;8(11):e1380-e1389. [doi: 10.1016/s2214-109x(20)30366-1]
10. Racine N, Hetherington E, McArthur BA, McDonald S, Edwards S, Tough S, et al. Maternal depressive and anxiety symptoms before and during the COVID-19 pandemic in Canada: a longitudinal analysis. Lancet Psychiatry 2021 May;8(5):405-415 [FREE Full text] [doi: 10.1016/S2215-0366(21)00074-2] [Medline: 33773109]
11. Paschke K, Arnaud N, Austermann MI, Thomasius R. Risk factors for prospective increase in psychological stress during COVID-19 lockdown in a representative sample of adolescents and their parents. BJPsych Open 2021 May 03;7(3):e94 [FREE Full text] [doi: 10.1192/bjo.2021.49] [Medline: 33938424]

XSL•FO
RenderX

12. Johnson MS, Skjerdingstad N, Ebrahimi OV, Hoffart A, Johnson SU. Parenting in a pandemic: parental stress, anxiety and depression among parents during the government-initiated physical distancing measures following the first wave of COVID-19. Stress Health 2022 Oct;38(4):637-652. [doi: 10.1002/smi.3120] [Medline: 34902219]
13. Westrupp EM, Bennett C, Berkowitz T, Youssef GJ, Toumbourou JW, Tucker R, et al. Child, parent, and family mental health and functioning in Australia during COVID-19: comparison to pre-pandemic data. Eur Child Adolesc Psychiatry 2023 Feb;32(2):317-330 [FREE Full text] [doi: 10.1007/s00787-021-01861-z] [Medline: 34417875]
14. Li J, Bünning M, Kaiser T, Hipp L. Who suffered most? Parental stress and mental health during the COVID-19 pandemic in Germany. J Fam Res 2022 Apr 01;34(1):281-309. [doi: 10.20377/jfr-704]
15. Kerr ML, Rasmussen HF, Fanning KA, Braaten SM. Parenting during COVID-19: a study of parents' experiences across gender and income levels. Fam Relat 2021 Dec;70(5):1327-1342 [FREE Full text] [doi: 10.1111/fare.12571] [Medline: 34548726]
16. Malkawi SH, Almhdawi K, Jaber AF, Alqatarneh NS. COVID-19 quarantine-related mental health symptoms and their correlates among mothers: a cross sectional study. Matern Child Health J 2021 May;25(5):695-705 [FREE Full text] [doi: 10.1007/s10995-020-03034-x] [Medline: 33201448]
17. Tchimtchoua Tamo AR. An analysis of mother stress before and during COVID-19 pandemic: the case of China. Health Care Women Int 2020;41(11-12):1349-1362. [doi: 10.1080/07399332.2020.1841194] [Medline: 33315545]
18. Kurata S, Hiraoka D, Ahmad Adlan AS, Jayanath S, Hamzah N, Ahmad-Fauzi A, et al. Influence of the COVID-19 pandemic on parenting stress across Asian countries: a cross-national study. Front Psychol 2021;12:782298 [FREE Full text] [doi: 10.3389/fpsyg.2021.782298] [Medline: 34992567]
19. Adams EL, Smith D, Caccavale LJ, Bean MK. Parents are stressed! Patterns of parent stress across COVID-19. Front Psychiatry 2021 Apr 8;12:626456 [FREE Full text] [doi: 10.3389/fpsyt.2021.626456] [Medline: 33897489]
20. Miller JJ, Cooley ME, Mihalec-Adkins BP. Examining the impact of COVID-19 on parental stress: a study of foster parents. Child Adolesc Social Work J 2022 Nov 20;39(2):147-156 [FREE Full text] [doi: 10.1007/s10560-020-00725-w] [Medline: 33235406]
21. Karbasi Z, Safdari R, Eslami P. The silent crisis of child abuse in the COVID-19 pandemic: a scoping review. Health Sci Rep 2022 Sep;5(5):e790 [FREE Full text] [doi: 10.1002/hsr2.790] [Medline: 35989944]
22. Rodriguez CM, Lee SJ, Ward KP, Pu DF. The perfect storm: hidden risk of child maltreatment during the Covid-19 pandemic. Child Maltreat 2021 May 23;26(2):139-151 [FREE Full text] [doi: 10.1177/1077559520982066] [Medline: 33353380]
23. Callegaro M, Baker R, Bethlehem J, Göritz AS, Krosnick JA, Lavrakas PJ. Online panel research. In: Callegaro M, Baker R, Bethlehem J, Göritz AS, Krosnick JA, Lavrakas PJ, editors. Online Panel Research: A Data Quality Perspective. Hoboken, NJ: John Wiley & Sons; 2014.
24. Zhao K. Sample representation in the social sciences. Synthese 2020 Mar 13;198:9097-9115. [doi: 10.1007/s11229-020-02621-3]
25. Lehdonvirta V, Oksanen A, Räsänen P, Blank G. Social media, web, and panel surveys: using non‐probability samples in social and policy research. Policy Internet 2020 Apr 29;13(1):134-155. [doi: 10.1002/poi3.238]
26. Hlatshwako TG, Shah SJ, Kosana P, Adebayo E, Hendriks J, Larsson EC, et al. Online health survey research during COVID-19. Lancet Digit Health 2021 Feb;3(2):e76-e77. [doi: 10.1016/s2589-7500(21)00002-9]
27. Spinelli M, Lionetti F, Setti A, Fasolo M. Parenting stress during the COVID-19 outbreak: socioeconomic and environmental risk factors and implications for children emotion regulation. Fam Process 2021 Jun;60(2):639-653. [doi: 10.1111/famp.12601] [Medline: 32985703]
28. Johnson MS, Skjerdingstad N, Ebrahimi OV, Hoffart A, Urnes Johnson S. Mechanisms of parental distress during and after the first COVID-19 lockdown phase: a two-wave longitudinal study. PLoS One 2021 Jun 24;16(6):e0253087 [FREE Full text] [doi: 10.1371/journal.pone.0253087] [Medline: 34166429]
29. Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. Public Opin Q 2011 Oct 05;75(4):709-747. [doi: 10.1093/poq/nfr020]
30. Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet: comparing sample representativeness and response quality. Public Opin Q 2009 Dec 01;73(4):641-678. [doi: 10.1093/poq/nfp075]
31. Cornesse C, Blom AG, Dutwin D, Krosnick JA, De Leeuw ED, Legleye S, et al. A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. J Surv Stat Methodol 2020 Jan 09;8(1):4-36. [doi: 10.1093/jssam/smz041]
32. Blasius J, Brandt M. Repräsentativität in online-befragungen. In: Weichbold M, Bacher J, Wolf C, editors. Umfrageforschung. Freiburg im Breisgau, Germany: VS Verlag für Sozialwissenschaften; 2009.
33. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
34. Johnson LA, Dias N, Clarkson G, Schreier AM. Social media as a recruitment method to reach a diverse sample of bereaved parents. Appl Nurs Res 2019 Dec;50:151201. [doi: 10.1016/j.apnr.2019.151201] [Medline: 31677928]

XSL•FO
RenderX

35. Oesterle S, Epstein M, Haggerty KP, Moreno MA. Using Facebook to recruit parents to participate in a family program to prevent teen drug use. Prev Sci 2018 May;19(4):559-569 [FREE Full text] [doi: 10.1007/s11121-017-0844-7] [Medline: 29116552]
36. Skeens MA, Sutherland-Foggio M, Damman C, Gerhardt CA, Akard TF. Facebook recruitment for research of children and parents during the COVID-19 pandemic. Appl Nurs Res 2022 Jun;65:151574 [FREE Full text] [doi: 10.1016/j.apnr.2022.151574] [Medline: 35577481]
37. Yaremych HE, Persky S. Recruiting fathers for parenting research: an evaluation of eight recruitment methods and an exploration of fathers' motivations for participation. Parent Sci Pract 2023;23(1):1-32 [FREE Full text] [doi: 10.1080/15295192.2022.2036940] [Medline: 37346458]
38. Dworkin J, Hessel H, Gliske K, Rudi JH. A comparison of three online recruitment strategies for engaging parents. Fam Relat 2016 Oct;65(4):550-561 [FREE Full text] [doi: 10.1111/fare.12206] [Medline: 28804184]
39. Leach LS, Butterworth P, Poyser C, Batterham PJ, Farrer LM. Online recruitment: feasibility, cost, and representativeness in a study of postpartum women. J Med Internet Res 2017 Mar 08;19(3):e61 [FREE Full text] [doi: 10.2196/jmir.5745] [Medline: 28274906]
40. Crescentini C, Feruglio S, Matiz A, Paschetto A, Vidal E, Cogo P, et al. Stuck outside and inside: an exploratory study on the effects of the COVID-19 outbreak on Italian parents and children's internalizing symptoms. Front Psychol 2020 Oct 22;11:586074 [FREE Full text] [doi: 10.3389/fpsyg.2020.586074] [Medline: 33192917]
41. Cusinato M, Iannattone S, Spoto A, Poli M, Moretti C, Gatta M, et al. Stress, resilience, and well-being in Italian children and their parents during the COVID-19 pandemic. Int J Environ Res Public Health 2020 Nov 10;17(22):8297 [FREE Full text] [doi: 10.3390/ijerph17228297] [Medline: 33182661]
42. Mousavi SF. Psychological well-being, marital satisfaction, and parental burnout in Iranian parents: the effect of home quarantine during COVID-19 outbreaks. Front Psychol 2020 Dec 3;11:553880 [FREE Full text] [doi: 10.3389/fpsyg.2020.553880] [Medline: 33343439]
43. Wolf JP, Freisthler B, Chadwick C. Stress, alcohol use, and punitive parenting during the COVID-19 pandemic. Child Abuse Negl 2021 Jul;117:105090 [FREE Full text] [doi: 10.1016/j.chiabu.2021.105090] [Medline: 33975257]
44. Yamaoka Y, Hosozawa M, Sampei M, Sawada N, Okubo Y, Tanaka K, et al. Abusive and positive parenting behavior in Japan during the COVID-19 pandemic under the state of emergency. Child Abuse Negl 2021 Oct;120:105212 [FREE Full text] [doi: 10.1016/j.chiabu.2021.105212] [Medline: 34298262]
45. Thorell LB, Skoglund C, de la Peña AG, Baeyens D, Fuermaier AB, Groom MJ, et al. Parental experiences of homeschooling during the COVID-19 pandemic: differences between seven European countries and between children with and without mental health conditions. Eur Child Adolesc Psychiatry 2022 Apr;31(4):649-661 [FREE Full text] [doi: 10.1007/s00787-020-01706-1] [Medline: 33415470]
46. Brown SM, Doom JR, Lechuga-Peña S, Watamura SE, Koppels T. Stress and parenting during the global COVID-19 pandemic. Child Abuse Negl 2020 Dec;110(Pt 2):104699 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104699] [Medline: 32859394]
47. Lee SJ, Ward KP, Lee JY, Rodriguez CM. Parental social isolation and child maltreatment risk during the COVID-19 pandemic. J Fam Violence 2021 Jan 14;37(5):813-824 [FREE Full text] [doi: 10.1007/s10896-020-00244-3] [Medline: 33462526]
48. Fancourt D, Steptoe A, Bu F. Trajectories of depression and anxiety during enforced isolation due to COVID-19: longitudinal analyses of 59,318 adults in the UK with and without diagnosed mental illness. medRxiv Preprint posted online on November 3, 2020. [doi: 10.1101/2020.06.03.20120923]
49. Fitzpatrick KM, Drawve G, Harris C. Facing new fears during the COVID-19 pandemic: the State of America's mental health. J Anxiety Disord 2020 Oct;75:102291 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102291] [Medline: 32827869]
50. Malesza M, Kaczmarek MC. Predictors of anxiety during the COVID-19 pandemic in Poland. Pers Individ Dif 2021 Feb 15;170:110419 [FREE Full text] [doi: 10.1016/j.paid.2020.110419] [Medline: 33052156]
51. Benassi E, Vallone M, Camia M, Scorza M. Women during the Covid-19 lockdown: more anxiety symptoms in women with children than without children and role of the resilience. Mediterr J Clin Psychol 2020;8(3). [doi: 10.6092/2282-1619/mjcp-2559]
52. Zamarro G, Prados MJ. Gender differences in couples' division of childcare, work and mental health during COVID-19. Rev Econ Househ 2021;19(1):11-40 [FREE Full text] [doi: 10.1007/s11150-020-09534-7] [Medline: 33488316]
53. Cheng Z, Mendolia S, Paloyo AR, Savage DA, Tani M. Working parents, financial insecurity, and childcare: mental health in the time of COVID-19 in the UK. Rev Econ Househ 2021;19(1):123-144 [FREE Full text] [doi: 10.1007/s11150-020-09538-3] [Medline: 33456425]
54. Huebener M, Waights S, Spiess CK, Siegel NA, Wagner GG. Parental well-being in times of Covid-19 in Germany. Rev Econ Househ 2021;19(1):91-122 [FREE Full text] [doi: 10.1007/s11150-020-09529-4] [Medline: 33469413]
55. Mazza C, Ricci E, Marchetti D, Fontanesi L, Di Giandomenico S, Verrocchio MC, et al. How personality relates to distress in parents during the Covid-19 lockdown: the mediating role of child's emotional and behavioral difficulties and the moderating effect of living with other people. Int J Environ Res Public Health 2020 Aug 27;17(17):6236 [FREE Full text] [doi: 10.3390/ijerph17176236] [Medline: 32867313]

XSL•FO
RenderX

56. Patrick SW, Henkhaus LE, Zickafoose JS, Lovell K, Halvorson A, Loch S, et al. Well-being of parents and children during the COVID-19 pandemic: a national survey. Pediatrics 2020 Oct;146(4):e2020016824. [doi: 10.1542/peds.2020-016824] [Medline: 32709738]
57. Yue J, Zang X, Le Y, An Y. Anxiety, depression and PTSD among children and their parent during 2019 novel coronavirus disease (COVID-19) outbreak in China. Curr Psychol 2022;41(8):5723-5730 [FREE Full text] [doi: 10.1007/s12144-020-01191-4] [Medline: 33223783]
58. Ren J, Li X, Chen S, Chen S, Nie Y. The influence of factors such as parenting stress and social support on the state anxiety in parents of special needs children during the COVID-19 epidemic. Front Psychol 2020 Dec 10;11:565393 [FREE Full text] [doi: 10.3389/fpsyg.2020.565393] [Medline: 33362628]
59. Russell BS, Hutchison M, Tambling R, Tomkunas AJ, Horton AL. Initial challenges of caregiving during COVID-19: caregiver burden, mental health, and the parent-child relationship. Child Psychiatry Hum Dev 2020 Oct;51(5):671-682 [FREE Full text] [doi: 10.1007/s10578-020-01037-x] [Medline: 32749568]
60. Mazza M, Marano G, Lai C, Janiri L, Sani G. Danger in danger: interpersonal violence during COVID-19 quarantine. Psychiatry Res 2020 Jul;289:113046 [FREE Full text] [doi: 10.1016/j.psychres.2020.113046] [Medline: 32387794]
61. Bramstedt KA. The carnage of substandard research during the COVID-19 pandemic: a call for quality. J Med Ethics 2020 Dec;46(12):803-807. [doi: 10.1136/medethics-2020-106494] [Medline: 33004545]
62. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
63. Kennedy EB, Jensen EA, Jensen AM. Methodological considerations for survey-based research during emergencies and public health crises: improving the quality of evidence and communication. Front Commun 2022 Feb 15;6:1-7. [doi: 10.3389/fcomm.2021.736195]
64. Singh S, Sagar R. A critical look at online survey or questionnaire-based research studies during COVID-19. Asian J Psychiatr 2021 Nov;65:102850 [FREE Full text] [doi: 10.1016/j.ajp.2021.102850] [Medline: 34534919]
65. Zhao Q, Ju N, Bacallado S, Shah RD. BETS: the dangers of selection bias in early analyses of the coronavirus disease (COVID-19) pandemic. Ann Appl Stat 2021 Mar 1;15(1):363-390. [doi: 10.1214/20-AOAS1401]
66. Pierce M, McManus S, Jessop C, John A, Hotopf M, Ford T, et al. Says who? The significance of sampling in mental health surveys during COVID-19. Lancet Psychiatry 2020 Jul;7(7):567-568. [doi: 10.1016/s2215-0366(20)30237-6]
67. Dutwin D, Buskirk TD. Apples to oranges or gala versus golden delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples. Public Opin Q 2017;81(S1):213-239. [doi: 10.1093/poq/nfw061]
68. Szolnoki G, Hoffmann D. Online, face-to-face and telephone surveys—comparing different sampling methods in wine consumer research. Wine Econ Policy 2013 Dec;2(2):57-66. [doi: 10.1016/j.wep.2013.10.001]
69. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 Apr 27;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
70. Duda MD, Nobile JL. The fallacy of online surveys: no data are better than bad data. Hum Dimens Wildl 2010 Feb 08;15(1):55-64. [doi: 10.1080/10871200903244250]
71. Newman A, Bavik YL, Mount M, Shao B. Data collection via online platforms: challenges and recommendations for future research. Appl Psychol 2021 Jul;70(3):1380-1402. [doi: 10.1111/apps.12302]
72. Kohler U, Kreuter F, Stuart EA. Nonprobability sampling and causal analysis. Annu Rev Stat Appl 2019;6:149-172. [doi: 10.1146/annurev-statistics-030718-104951]
73. Groves RM. Nonresponse rates and nonresponse bias in household surveys. Public Opin Q 2006 Jan 01;70(5):646-675. [doi: 10.1093/poq/nfl033]
74. Martin E, Winters F. Money and motive: effects of incentives on panel attrition in the survey of income and program participation. J Off Stat 2001 Jun;17(2):267-284.
75. Zhang C, Lonn S, Teasley SD. Understanding the impact of lottery incentives on web survey participation and response quality: a leverage-salience theory perspective. Field Methods 2017;29(1):42-60. [doi: 10.1177/1525822x16647932]
76. Bethlehem J. Selection bias in web surveys. Int Stat Rev 2010 Aug 03;78(2):161-188. [doi: 10.1111/j.1751-5823.2010.00112.x]
77. Choi I, Milne DN, Glozier N, Peters D, Harvey SB, Calvo RA. Using different Facebook advertisements to recruit men for an online mental health study: engagement and selection bias. Internet Interv 2017 Jun;8:27-34 [FREE Full text] [doi: 10.1016/j.invent.2017.02.002] [Medline: 30135825]
78. Batterham PJ. Recruitment of mental health survey participants using internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191 [FREE Full text] [doi: 10.1002/mpr.1421] [Medline: 24615785]
79. Joyal-Desmarais K, Stojanovic J, Kennedy EB, Enticott JC, Boucher VG, Vo H, et al. How well do covariates perform when adjusting for sampling bias in online COVID-19 research? Insights from multiverse analyses. Eur J Epidemiol 2022 Dec;37(12):1233-1250 [FREE Full text] [doi: 10.1007/s10654-022-00932-y] [Medline: 36335560]

XSL•FO
RenderX

80. Bennetts SK, Hokke S, Crawford S, Hackworth NJ, Leach LS, Nguyen C, et al. Using paid and free Facebook methods to recruit Australian parents to an online survey: an evaluation. J Med Internet Res 2019 Mar 06;21(3):e11206 [FREE Full text] [doi: 10.2196/11206] [Medline: 30839282]
81. Lampert T, Müters S, Stolzenberg H, Kroll LE, KiGGS Study Group. [Measurement of socioeconomic status in the KiGGS study: first follow-up (KiGGS Wave 1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014 Jul;57(7):762-770. [doi: 10.1007/s00103-014-1974-8] [Medline: 24950825]
82. Finlay I, Gilmore I. Covid-19 and alcohol-a dangerous cocktail. BMJ 2020 May 20;369:m1987. [doi: 10.1136/bmj.m1987] [Medline: 32434792]
83. Löwe B, Spitzer RL, Zipfel S, Herzog W. Gesundheitsfragebogen für Patienten (PHQ-D). Medizinische Universitätsklinik. 2002. URL: https://www.klinikum.uni-heidelberg.de/fileadmin/medizinische_klinik/Abteilung_2/pdf/Komplett_PHQ_Fragebogen.pdf [accessed 2024-12-19]
84. Berry JO, Jones WH. The parental stress scale: initial psychometric evidence. J Soc Pers Relatsh 1995 Aug 01;12(3):463-472. [doi: 10.1177/0265407595123009]
85. Benyamini Y, Leventhal EA, Leventhal H. Self-assessments of health: what do people know that predicts their mortality? Res Aging 1999 May 01;21(3):477-500. [doi: 10.1177/0164027599213007]
86. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
87. Teicher MH, Parigger A. The 'Maltreatment and Abuse Chronology of Exposure' (MACE) scale for the retrospective assessment of abuse and neglect during development. PLoS One 2015 Feb 25;10(2):e0117423 [FREE Full text] [doi: 10.1371/journal.pone.0117423] [Medline: 25714856]
88. Hecker T, Boettcher VS, Landolt MA, Hermenau K. Child neglect and its relation to emotional and behavioral problems: a cross-sectional study of primary school-aged children in Tanzania. Dev Psychopathol 2018 Mar 26;31(1):325-339. [doi: 10.1017/s0954579417001882]
89. Isele D, Teicher MH, Ruf-Leuschner M, Elbert T, Kolassa IT, Schury K, et al. KERF–ein instrument zur umfassenden ermittlung belastender kindheitserfahrungen. Zeitschrift Klinische Psychologie Psychotherapie 2014 Apr;43(2):121-130. [doi: 10.1026/1616-3443/a000257]
90. Runyan DK, Cox CE, Dubowitz H, Newton RR, Upadhyaya M, Kotch JB, et al. Describing maltreatment: do child protective service reports and research definitions agree? Child Abuse Negl 2005 May;29(5):461-477. [doi: 10.1016/j.chiabu.2004.06.015] [Medline: 15970320]
91. Manly JT, Cicchetti D, Barnett D. The impact of subtype, frequency, chronicity, and severity of child maltreatment on social competence and behavior problems. Dev Psychopathol 2008 Oct 31;6(1):121-143. [doi: 10.1017/s0954579400005915]
92. Cabrera NJ, Volling BL, Barr R. Fathers are parents, too! Widening the lens on parenting for children's development. Child Dev Perspectives 2018 Jan 08;12(3):152-157. [doi: 10.1111/cdep.12275]
93. Costigan CL, Cox MJ. Fathers' participation in family research: is there a self-selection bias? J Fam Psychol 2001;15(4):706-720. [doi: 10.1037/0893-3200.15.4.706]
94. Davison KK, Charles JN, Khandpur N, Nelson TJ. Fathers' perceived reasons for their underrepresentation in child health research and strategies to increase their involvement. Matern Child Health J 2017 Feb;21(2):267-274 [FREE Full text] [doi: 10.1007/s10995-016-2157-z] [Medline: 27473093]
95. Bartholomew MK, Schoppe-Sullivan SJ, Glassman M, Kamp Dush CM, Sullivan JM. New parents' Facebook use at the transition to parenthood. Fam Relat 2012 Jul;61(3):455-469 [FREE Full text] [doi: 10.1111/j.1741-3729.2012.00708.x] [Medline: 23671354]
96. Duggan M, Lenhart A, Lampe C, Ellison NB. Parents and social media. Pew Research Center. 2015 Jul 16. URL: https://www.pewresearch.org/internet/2015/07/16/parents-and-social-media/ [accessed 2024-12-19]
97. Leach LS, Bennetts SK, Giallo R, Cooklin AR. Recruiting fathers for parenting research using online advertising campaigns: evidence from an Australian study. Child Care Health Dev 2019 Nov 22;45(6):871-876. [doi: 10.1111/cch.12698] [Medline: 31216596]
98. Lawson M, Piel MH, Simon M. Child maltreatment during the COVID-19 pandemic: consequences of parental job loss on psychological and physical abuse towards children. Child Abuse Negl 2020 Dec;110(Pt 2):104709 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104709] [Medline: 32893003]
99. Beland LP, Brodeur A, Haddad J, Mikola D. Covid-19, family stress and domestic violence: remote work, isolation and bargaining power. IZA Institute of Labor Economics. 2020 Jun 5. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3627031 [accessed 2024-12-19]
100. Machlin L, Gruhn MA, Miller AB, Milojevich HM, Motton S, Findley AM, et al. Predictors of family violence in North Carolina following initial COVID-19 stay-at-home orders. Child Abuse Negl 2022 Aug;130(Pt 1):105376 [FREE Full text] [doi: 10.1016/j.chiabu.2021.105376] [Medline: 34728100]
101. Cook TD, Steiner PM, Pohl S. How bias reduction is affected by covariate choice, unreliability, and mode of data analysis: results from two types of within-study comparisons. Multivariate Behav Res 2009 Nov 30;44(6):828-847. [doi: 10.1080/00273170903333673] [Medline: 26801798]

XSL•FO
RenderX

102. Gelebte vielfalt: familien mit migrationshintergrund in Deutschland. Bundesministerium für Familie, Senioren, Frauen und Jugend. 2017. URL: https://www.bmfsfj.de/resource/blob/116880/83c02ec19dbea15014d7868048f697f2/gelebte-vielfalt--familien-mit-migrationshintergrund-in-deutschland-data.pdf [accessed 2024-12-19]
103. Witt A, Brown RC, Plener PL, Brähler E, Fegert JM. Child maltreatment in Germany: prevalence rates in the general population. Child Adolesc Psychiatry Ment Health 2017 Sep 29;11(1):47 [FREE Full text] [doi: 10.1186/s13034-017-0185-0] [Medline: 28974983]
104. Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. Int J Forecast 2015 Jul;31(3):980-991 [FREE Full text] [doi: 10.1016/j.ijforecast.2014.06.001]
105. Simons DJ, Shoda Y, Lindsay DS. Constraints on generality (COG): a proposed addition to all empirical papers. Perspect Psychol Sci 2017 Nov;12(6):1123-1128. [doi: 10.1177/1745691617708630] [Medline: 28853993]
106. Neuhauser H, Rosario AS, Butschalowsky H, Haller S, Hoebel J, Michel J, et al. Nationally representative results on SARS-CoV-2 seroprevalence and testing in Germany at the end of 2020. Sci Rep 2022 Nov 14;12(1):19492 [FREE Full text] [doi: 10.1038/s41598-022-23821-6] [Medline: 36376417]

Abbreviations

ANCOVA: analyses of covariance
CATI: computer-assisted telephone interview
CAWI: computer-assisted web interview
NEA: nonverbal emotional abuse
VEA: verbal emotional abuse
WDV: witnessing domestic violence

Edited by A Coristine; submitted 24.08.23; peer-reviewed by S Gordon, B Rodgers; comments to author 11.01.24; revised version received 03.07.24; accepted 20.09.24; published 24.01.25.

Please cite as:
Engelke L, Calvano C, Pohl S, Winter SM, Renneberg B
Parental Mental Health and Child Maltreatment in the COVID-19 Pandemic: Importance of Sampling in a Quantitative Statistical Study
J Med Internet Res 2025;27:e52043
URL: https://www.jmir.org/2025/1/e52043
doi:10.2196/52043
PMID:

©Lara Engelke, Claudia Calvano, Steffi Pohl, Sibylle Maria Winter, Babette Renneberg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Detecting Deception and Ensuring Data Integrity in a Nationwide mHealth Randomized Controlled Trial: Factorial Design Survey Study

Krista M Kezbers[1], PhD; Michael C Robertson[1,2], PhD; Emily T Hébert[1,2], DrPH; Audrey Montgomery[1], MEd; Michael S Businelle[1,2], PhD

[1]Tobacco Settlement Endowment Trust Health Promotion Research Center, Stephenson Cancer Center, University of Oklahoma Health Sciences, Oklahoma City, OK, United States

[2]Department of Family and Preventive Medicine, University of Oklahoma Health Sciences, Oklahoma, OK, United States

Corresponding Author:
Krista M Kezbers, PhD
Tobacco Settlement Endowment Trust Health Promotion Research Center
Stephenson Cancer Center
University of Oklahoma Health Sciences
655 Research Parkway
Suite 400
Oklahoma City, OK, 73104
United States
Phone: 1 4052718001 ext 45042
Email: krista-kezbers@ouhsc.edu

Abstract

Background: Social behavioral research studies have increasingly shifted to remote recruitment and enrollment procedures. This shifting landscape necessitates evolving best practices to help mitigate the negative impacts of deceptive attempts (eg, fake profiles and bots) at enrolling in behavioral research.

Objective: This study aimed to develop and implement robust deception detection procedures during the enrollment period of a remotely conducted randomized controlled trial.

Methods: A 32-group (2×2×2×2×2) factorial design study was conducted from November 2021 to September 2022 to identify mobile health (mHealth) survey design features associated with the highest completion rates of smartphone-based ecological momentary assessments (n=485). Participants were required to be at least 18 years old, live in the United States, and own an Android smartphone that was compatible with the Insight app that was used in the study. Recruitment was conducted remotely through Facebook advertisements, a 5-minute REDCap (Research Electronic Data Capture) prescreener, and a screening and enrollment phone call. The research team created and implemented a 12-step checklist (eg, address verification and texting a copy of picture identification) to identify and prevent potentially deceptive attempts to enroll in the study. Descriptive statistics were calculated to understand the prevalence of various types of deceptive attempts at study enrollment.

Results: Facebook advertisements resulted in 5236 initiations of the REDCap prescreener. A digital deception detection procedure was implemented for those who were deemed pre-eligible (n=1928). This procedure resulted in 26% (501/1928) of prescreeners being flagged as potentially deceptive. Completing multiple prescreeners (301/501, 60.1%) and providing invalid addresses (156/501, 31.1%) were the most common reasons prescreeners were flagged. An additional 1% (18/1928) of prescreeners were flagged as potentially deceptive during the subsequent study screening and enrollment phone call. Reasons for exclusion at the screening and enrollment phone call level included having an invalid phone type (6/18, 33.3%), completing multiple prescreeners (6/18, 33.3%), and providing an invalid address (5/18, 27.7%). This resulted in 1409 individuals being eligible after all deception checks were completed. Postenrollment social security number checks revealed that 3 (0.6%) fully enrolled participants out of 485 provided erroneous social security numbers during the screening process.

Conclusions: Implementation of a deception detection procedure in a remotely conducted randomized controlled trial resulted in a substantial proportion of cases being flagged as potentially engaging in deceptive attempts at study enrollment. The results of the deception detection procedures in this study confirmed the need for vigilance in conducting remote behavioral research in

XSL•FO
RenderX

order to maintain data integrity. Implementing systematic deception detection procedures may support study administration, data quality, and participant safety in remotely conducted behavioral research.

Trial Registration: ClinicalTrials.gov NCT05194228; https://clinicaltrials.gov/study/NCT05194228

(J Med Internet Res 2025;27:e66384) doi:10.2196/66384

KEYWORDS

ecological momentary assessment; enrollment; fraud; mHealth; randomized controlled trial; recruitment; deception; data integrity; behavior; social; RCT; factorial design; mobile phone

Introduction

A paradigm shift in methodologies used to conduct randomized controlled trials (RCT) focused on human behavior and behavior change is underway [1]. Specifically, remotely conducted behavioral trials, trials that do not require in-person visits, are increasingly commonplace. This was true before the onset of the COVID-19 pandemic, and this trend has accelerated in recent years due to technological advancements and shifting social norms [2].

By reducing barriers to participation in RCTs (eg, transportation limitations and inability to take off work), remotely conducted trials may have greater generalizability than traditional in-person RCTs and may extend health promotion programs to those who might otherwise lack access to them. Furthermore, the widespread use of smartphones, now owned by 90% of adults in the United States, presents an unprecedented opportunity to provide personally tailored health-promoting content to vast numbers of people [1,3].

There are potential disadvantages to conducting behavioral trials remotely via smartphones. First, obtaining appropriately representative samples to study digitally mediated behavior interventions can be challenging [4,5]. While this issue is not unique to remotely delivered interventions, participation requires a willingness and ability to operate various technologies that may systematically affect inclusion, potentially leaving remotely conducted research prone to limitations imposed by the "digital divide" [6]. For example, lower ownership of smartphones among older adults, Americans with lower incomes, and people who reside in rural environments can limit equitable inclusion of these groups if these realities are not carefully addressed [7].

Second, remotely conducted behavioral trials may be prone to participant deception [8]. Several types of deception among research participants in clinical trials have been identified [8]. Concealment has been defined as the act of failing to disclose relevant information to facilitate study admittance (eg, a potential participant not sharing that they have an exclusionary comorbidity) [8]. *Fabrication* refers to the duplicitous invention on the part of a potential participant to support their inclusion in a research trial (eg, erroneously claiming to be diagnosed with a health-related condition) [8]. *Collusion* is a coordinated sharing of information among potential research participants to gain study admittance [8]. Deception on the part of research participants can lead to considerable waste of study resources, compromised integrity of data resulting from a study, and threats to participant safety [8]. Indeed, severe symptoms and death have been attributed, in part, to deception on the part of research participants [9,10].

Financial gain appears to be a primary reason people engage in deceptive practices to enroll in research studies [11,12]. Participants are commonly paid to participate in research studies to reimburse for out-of-pocket expenses associated with study participation (eg, smartphone data usage), to compensate for time and effort associated with study participation, and to incentivize participation [13]. The prevailing consensus is that paying individuals to participate in research is ethical if it does not undermine informed consent [13]. Qualitative research studies have highlighted the critical role that monetary payment plays in increasing willingness to participate in RCTs [14,15]. However, higher levels of compensation can increase the likelihood of deceptive practices to gain admittance into research studies [15].

Financial incentives may motivate people to serially engage in clinical trials. Devine et al [11] recruited "experienced research participants" (ie, individuals who had participated in more than one study in the past year). Research participants (N=100) reported participating in an average of 12 studies within the previous year (range 2-100 studies), with reported lifetime earnings from research participation to be US $9809 on average (range US $50-US $175,000). A total of 75% of participants in this study reported engaging in concealment to avoid study exclusion, and 25% reported engaging in fabrication to facilitate study inclusion. Qualitative data from this study also characterized instances of collusion, including a description of how "research kingpins" have been known to trade information with individuals to facilitate their admittance into research studies [11]. The researchers suggested that "professional subjects" may be substantially overrepresented in clinical research [11].

Technologies that can greatly extend the reach of remotely conducted behavioral research may leave this type of research especially prone to deception from research participants. The enhanced accessibility of digitally mediated behavioral interventions may disproportionately attract individuals who aim to "game the system" [7]. For example, digital screening processes may be exploited by individuals who repeatedly restart the screening process to change their information to satisfy eligibility requirements [7]. There are reports of individuals accessing ClinicalTrials.gov and other clinical trial repositories to identify study inclusion and exclusion criteria to support deceptive enrollment into RCTs [11]. Remotely conducted behavioral research trials may be especially attractive to experienced study participants (1) because they are often low

XSL•FO
RenderX

risk—this appears to be an especially salient consideration among people who participate in many research studies annually [11], and (2) because considerable remuneration is often used as a strategy to reduce the high rates of attrition that can afflict digitally mediated studies [16].

Although it may be impossible to entirely eliminate this threat, there is a growing catalog of techniques for preventing and minimizing the ill effects of deception from research participants in digitally mediated behavioral interventions. Devine et al [11] suggested that limiting the face validity of items used to screen out participants and concealing specific reasons for nonadmittance may reduce deception. Dahne et al [7] created a visual database to aid members of their research team in the identification of potential participants who may have entered numerous screening records. Abroms et al [17] and Bricker et al [18] used techniques such as a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) to validate potential participants as human (ie, not automated programs [bots]), to evaluate screening IP addresses (eg, removing all duplicated IP addresses or those not of US origin), and to review system usage data (ie, taking less than 90 seconds to complete a screening survey or taking less 10 minutes to complete the baseline survey). For flagged cases, members of these research teams contacted individuals to further evaluate the authenticity of attempts at study enrollment. Resnik and McCann [19] have recommended that any evidence of potentially deceptive activity should result in the exclusion of the prospective participant from the study enrollment.

Verifying information provided by participants via photo identification, objective measures, and social security numbers (SSNs) may reduce deceptive attempts at study enrollment [8,11,19]. Smartphone-mediated screening processes may facilitate submitting photographs of identification cards. Smoking status can be tested biochemically using devices that can be remotely delivered [8,11,19]. Advising participants that their self-report responses will subsequently be verified may dissuade deception. Incentivizing the provision of accurate information may further support data integrity (eg, providing a "bonus" for providing information that is not discrepant with subsequent objective measures) [19].

This study details techniques that we implemented to minimize the effects of deception from individuals who attempted to enroll in the Exemplar study, a remotely delivered, nationwide RCT. We conclude by providing recommendations for future remotely conducted behavioral interventions.

Methods

Study Overview

The Exemplar study used a 2×2×2×2×2 factorial design to determine smartphone survey factors related to high daily survey completion rates. Eligible participants were randomly assigned and completed a baseline survey, 28 days of brief daily ecological momentary assessments (EMAs), and a final survey. The entire study was conducted remotely via REDCap (Research Electronic Data Capture; Vanderbilt University) and the Insight platform [20]. The Insight platform allows researchers to rapidly create, develop, and implement EMA studies [21]. This study was registered with ClinicalTrials.gov (NCT05194228). Details concerning the study design, participants, and outcomes are presented elsewhere [20].

Recruitment Strategy

The target population for this study was adults (ie, 18 years or older) living in the United States who had an active Android smartphone. Participants were recruited via Facebook advertisements from November 2021 to September 2022. The advertisements were targeted to individuals aged at least 18 years old who possessed an Android smartphone. Examples of Facebook advertisements included:

> *We're looking for Android users to complete brief daily surveys on their smartphones. Qualified participants will be compensated up to $152 over 4 weeks. NO in person visits required." and "Have an Android smartphone? See if you're eligible to participate in a 4-week research study. Qualified participants will be compensated up to $152 for their time. NO in person visits required.*

Clicking on a Facebook advertisement initiated an encrypted REDCap prescreener survey that was used to determine initial eligibility for the study. Individuals were not shown their eligibility results until after all prescreener items were complete. Ineligible individuals were informed that they were not eligible for the study, but the specific reason(s) were not provided. If a participant was potentially eligible based on the REDCap prescreener, they selected times for a screening and enrollment call. Study staff texted eligible participants, thanking them for completing the prescreener and confirming the date and time for their screening and enrollment call.

Enrollment and Participant Flow

During the screening and enrollment call, individuals were asked additional screening questions; verified that their smartphone was compatible with the Insight mobile health (mHealth) platform [21]; completed a reading comprehension test, Rapid Estimate of Adult Literacy in Medicine – Short Form [22]; and completed the informed consent process. Individuals were also required to text a picture of their photo ID with a US address. Once fully enrolled, participants downloaded the Insight smartphone app onto their personal Android smartphones. They received instructions on how to use the smartphone app to complete the baseline survey, daily EMAs for 28 days, and the final survey. After the baseline survey was completed, participants were mailed a reloadable Greenphire Mastercard. Baseline compensation (ie, US $10) was loaded onto the Greenphire card after the participant confirmed that they received the card by texting study staff the last 4 digits of their card. The rationale for requiring participants to text the last 4 digits of their card was to confirm when they received their card. Funds were loaded onto participant cards after they verified receipt of the card to ensure payment was received (eg, to overcome the possibility of cards being lost in the mail). Compensation for completing EMAs (ie, up to US $56 for those randomly assigned to complete two daily EMAs, up to US $112 for those randomly assigned to complete four daily EMAs) and

XSL•FO
RenderX

the final survey (US $30) was loaded onto the Greenphire card at the completion of the study.

Measures

REDCap Prescreener

All individuals who clicked a Facebook advertisement were taken to an encrypted REDCap screening survey hosted by the University of Oklahoma Health Sciences Center (OUHSC). Individuals completed demographic questions (eg, age, race, address, and phone number) and questions about their smartphones (eg, phone ownership and type of phone). Individuals who met the study inclusion criteria were asked to select a primary and secondary time for the screening and enrollment phone call. Deception detection steps, along with associated considerations and decisions, for the REDCap prescreener can be seen in Table 1. These steps address the deception detection review that occurs once an individual is determined to be eligible at the REDCap prescreener level. REDCap tracking allowed for easy reporting and periodic checks of the data to determine if there were any prescreener patterns that could indicate new types of attempts at deceptive enrollment. This was initially how it was determined to add reCAPTCHA and the identity statement (Table 1, step 6).

Table 1. Steps taken to detect potential deception in REDCap[a] prescreeners that were completed by individuals that viewed study advertisements.

Deception detection step for the REDCap prescreener	Considerations and decisions
1. Incomplete prescreeners	Incomplete prescreeners were not reviewed.
2. Duplicative prescreeners	Previous prescreeners with identical information, individuals currently or previously enrolled in the study, and previously ineligible individuals were not scheduled for a screening and enrollment call.
3. Home and mailing address	Validated by Google and Google Maps; invalid addresses and business addresses not linked to an individual's residence were not scheduled for a screening and enrollment call.
4. Phone number	Phone number provider search was conducted to determine use of bandwidth phones (ie, a phone not connected to a cellular provider); participants were scheduled for a screening and enrollment call but not excluded; this step was collected for future consideration of exclusion.
5. Email address	Verified by Google search as an exploratory approach to deception detection; participants were scheduled for a screening and enrollment call but not excluded.
6. reCAPTCHA[b] and identity statement	8.5 weeks into the study, reCAPTCHA was added to prevent bots from autofilling the prescreener; 9.5 weeks into the study, an identity statement was added, "Due to a large number of spam accounts, we will verify your identity prior to enrolling you in the study."
7. Periodic data checking	Data were periodically reviewed to determine potential patterns of deception. Cases that passed the deception deterrence checklist but were flagged as suspicious were scheduled for a screening and enrollment phone call.
8. Conservative decision-making	In cases where there was any doubt about whether a prescreener contained deceptive information, the screening and enrollment call was scheduled. This conservative approach allowed for the reduction in deceptive enrollments while ensuring that eligible individuals could participate in the study.

[a]REDCap: Research Electronic Data Capture.

[b]CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart.

Screening and Enrollment Phone Call

During this phone call, individuals reviewed and electronically signed the informed consent through the REDCap link, completed a reading comprehension test [22], provided their SSN, and uploaded a copy of their photo identification. All pictures of identification were deleted after identity verification to reduce potential breaches in confidentiality. Eligible individuals provided a minimum 14-hour window of waking hours for each day of the week (EMAs could only be prompted during waking hours). Table 2 shows the deception detection steps for the screening and enrollment phone call.

XSL•FO
RenderX

Table 2. Steps taken to detect potential deception during the study enrollment phone calls.

Deception detection step for the screening and enrollment phone call	Considerations and decisions
9. Informed consent	• If the signed name differed from the REDCap[a] prescreener, study staff asked the individual to confirm the name that was signed. If the name did not match, the screening and enrollment call was terminated, and the screener was marked as "deceptive." • Shortened names (eg, Chris instead of Christopher) were not considered to be deceptive. • Married versus maiden names were confirmed by uploading a picture of a piece of stamped mail with the correct name and matching mailing address. If a matching piece of stamped mail could not be provided, the screening and enrollment call was terminated, and the individual was not invited to enroll in the study.
10. Photo identification	• The name and address on the individual's photo identification had to match information provided during the REDCap prescreener. If the address did not match, individuals were required to send a picture of a piece of stamped mail that was sent to them at their address. If no other information could be provided to verify their address, the screening and enrollment call was ended, and the individual was not allowed to enroll in the study.
11. Social security number	• SSN[b] was required to pay research participants at our institution and was needed for tax reporting purposes. If an individual declined to provide their SSN, the screening and enrollment call was ended, and the individual was marked as "Opt to not participate" in REDCap
12. Greenphire Mastercards	• Physical Greenphire Mastercards, rather than virtual gift cards, were used to compensate participants for completing study surveys. Greenphire cards were only mailed to the verified mailing address for each participant after they completed the baseline survey in the Insight app.

[a]REDCap: Research Electronic Data Capture.

[b]SSN: social security number.

Baseline Survey, EMAs, and Final Survey

Participants used the Insight app to complete the baseline, daily EMAs (2-4 per day), and follow-up surveys. These surveys asked about affect, health, and health behaviors.

Compensation

Once a participant received their Greenphire card in the mail, they were instructed to contact study staff to confirm the card had been received. Study staff then sent the participant a link to sign that they had received their card. This was an additional step to ensure that participants lived at the address they provided in their prescreener and that they interacted with study staff for an additional time. Individuals who did not confirm their card did not receive any study payments.

Statistical Methods

Descriptive statistics were used to describe the sample and highlight deception detection outcomes. All analyses were conducted using SPSS (IBM Corp, version 29).

Ethical Considerations

The University of Oklahoma Health Sciences Institutional Review Board approved this study (IRB#13684). All participants in this study signed an informed consent via REDCap before engaging in any research study activities. The informed consent provided language that the study is optional and participants could withdraw at any time. All study datasets were deidentified before analyses were conducted. Participants could be compensated up to US $152 depending on the group they were randomized into for the study. Since this paper focused on the steps leading up to the start of the study, no further compensation details are provided, but they can be found in the primary outcomes paper [20].

Results

From November 29, 2021, to September 27, 2022, Facebook advertisements yielded 5236 clicks and openings of the REDCap screener. Of those screeners, 3308 were ineligible (eg, 1158 blank prescreeners), and 1928 REDCap prescreeners indicated initial eligibility for the study. Of the 1928 eligible screeners, 195 (15.3%) out of 1928 were duplicates of previously eligible individuals, and 206 (10.7%) out of 1928 were marked as potentially deceptive (Table 3). All screening outcomes are presented in Table 3. Ultimately, 1409 individuals remained eligible after deception detection steps 1-8 were implemented. The deception detection checklist required, on average, approximately 1-4 minutes to complete for each eligible REDCap prescreener.

On January 27, 2022, reCAPTCHA was added to prevent bots from autocompleting the REDCap prescreener in rapid succession. On February 2, 2022, an identity statement was added in the REDCap prescreener so individuals were aware their identity would be verified before enrollment. A total of 3 (1.5%) inaccurate SSNs were detected out of 202 participants who earned over US $100, the OUHSC minimum reporting requirement for participant compensation. The SSNs of the 283 participants who earned less than US $100 were not checked. During the screening and enrollment call, 65 (3.4%) out of 1928 pre-eligible individuals declined to provide their SSN and opted to not participate in the study.

XSL•FO
RenderX

A total of 1928 nonduplicate, complete REDCap screeners were examined using our deception detection and abatement procedure. The majority of deceptive attempts at enrolling into the study were detected at the REDCap screener level (501/1928, 26%) and a minority of deceptive attempts at study enrollment were detected during the screening and enrollment call (18/1928, 1%; see Table 3). Table 4 displays reported characteristics of deceptive attempts at enrollment at the REDCap screener level, screening and enrollment call level, and those that fully enrolled in the study. After all deception checks were complete, 1409 individuals remained eligible to enroll in the study. Ultimately, 924 individuals did not enroll due to various reasons (ie, unable to contact, canceled screening and enrollment call, or ineligible phone) and 485 individuals were randomly assigned into the study. Participants completed 83.8% (28,948/34,552) of all EMAs, and 96.6% (397/411) completed the follow-up assessment. Additional details are provided elsewhere [20].

Table 3. REDCap[a] screener and screener and enrollment call outcomes.

Variables	Value
Facebook screener clicks, n	5236
Incomplete screeners (step 1), n	1689
Ineligible due to inclusion criteria not met, n	1619
Eligible before deception check, n	1928
REDCap deception check, n/N (%)	501/1928 (26)
Duplicates (step 2), n/N (%)	301/501 (60.1)
Invalid address (step 3), n/N (%)	156/501 (31.1)
Invalid phone number (step 4), n/N (%)	2/501 (0.4)
"Other" or "No reason" listed, n/N (%)	42/501 (8.4)
Screening and enrollment deception check , n/N (%)	18/1928 (1)
Invalid phone	6/18 (33.3)
Duplicates	6/18 (33.3)
Invalid address	5/18 (27.8)
Evidence of an automated program (bot)	1/18 (5.6)
Total eligible to enroll after all deception check, n	1409
Fully enrolled and randomly assigned	485
Poststudy SSN[b] check—inaccurate	3

[a]REDCap: Research Electronic Data Capture.

[b]SSN: social security network.

XSL•FO
RenderX

Table 4. Reported demographics by potential deception status. Demographics characteristics listed in the first two columns may not be accurate as deception was detected.

Variable	Flagged as potentially deceptive in prescreener (n=501)	Flagged as potentially deceptive in enrollment call (n=18)	Fully enrolled participants (n=485)
Age (years), mean (SD)	43.1 (13.5)	32.4 (6.9)	48.2 (12.4)
Ethnicity (n, %)			
Hispanic	58 (11.6)	2 (11.1)	48 (9.9)
Non-Hispanic	443 (88.4)	16 (88.9)	437 (90.1)
Race, n (%)			
American Indian or Alaska Native	9 (1.8)	0 (0.0)	7 (1.4)
Asian	16 (3.2)	0 (0.0)	13 (2.7)
Black or African American	191 (38.1)	10 (55.6)	93 (19.2)
Native Hawaiian or Other Pacific Islander	1 (0.2)	0 (0.0)	1 (0.2)
White	265 (52.9)	7 (38.9)	342 (70.5)
More than one race	19 (3.8)	1 (5.6)	29 (6.0)
Biological sex, n (%)			
Female	289 (57.7)	8 (44.4)	370 (76.3)
Male	212 (42.3)	10 (55.6)	115 (23.7)

Discussion

Principal Findings

This study used a systematic deception detection procedure to attempt to preserve study resources, data integrity, and participant safety in a remotely delivered behavioral study. Given the relatively large scale of our digitally mediated recruitment methods, the low risk involved in study participation, considerable financial compensation, and robust practices for ultimately verifying participants' identity, this study provided unique insight into how deceptive practices may be occurring in large, remotely delivered behavioral studies, as well as how to mitigate their impact. The digitally mediated deception check implemented at the REDCap prescreener level (steps 1-8) resulted in 26% (501/1928) of the eligible sample being excluded. Only 1% (18/1928) of the eligible sample was excluded during the subsequent screening and enrollment call. The systematic deception detection steps detailed in this paper support data integrity and may minimize the risk of deceptive enrollments.

Within the digitally mediated REDCap prescreener steps, duplicate screeners and address checks provided the most benefit to the deception detection process in terms of the number of potentially deceptive instances identified. Study staff sorted and compared the data and metadata associated with various submissions. The practices employed at this step were similar to those that have been used elsewhere [7]. Some of the submissions removed during the REDCap prescreener level appeared to have been initiated by automated programs (ie, "bots"). This led to the decision to implement a reCAPTCHA checkpoint before allowing potentially eligible individuals to proceed with the prescreener survey. This practice has been successfully used in other remotely conducted research studies [17,18].

In contrast to other studies, we did not exclude individuals on the basis of their IP address. The university institutional review board did not allow us to save and track IP addresses. Further, while IP addresses can provide information useful for determining if an individual is submitting multiple prescreeners from the same device, location-based decision rules may limit enrollments unnecessarily. For example, many companies, academic institutions, and individuals use roaming IP addresses. This can entail a limited number of IP addresses being associated with one large organization or IP addresses that regularly change for a given user. The use of roaming IP addresses can increase security and preserve privacy. Thus, while IP address information may be an important part of identifying potentially deceptive practices, caution should be used to avoid unnecessarily prohibitive practices.

While individuals were not excluded from this study if they had a bandwidth or voice over internet provider (VoIP) phone number, we observed the limited abilities of such devices to interact with the study app. Bandwidth and VoIP phone numbers can be used to mask true phone numbers or to create an internet-based phone number without a physical phone. Bandwidth and VoIP phone numbers are generally not considered "smartphones" and are typically not able to download phone apps. In this study, people with bandwidth phone numbers were allowed to proceed to the enrollment phone call, but we have since decided to exclude them from our studies due to incompatibility. Researchers should carefully consider how they will handle problematic smartphone types and phone number-masking services in mHealth studies, as they are increasingly commonplace and have a bearing on app compatibility and how surveys and notifications are received.

Few attempts at deceptive enrollment were detected at the screening and enrollment call level. This may provide evidence in support of the steps we took to prevent deceptive enrollments

XSL•FO
RenderX

at the REDCap prescreener level. During the REDCap prescreener, individuals were informed that identification would be verified before enrollment; the mode of verification, however, was not disclosed. Requiring photo identification and ensuring addresses entered into the prescreener assessment matched the photo identification allowed for a relatively quick and easy way to verify an individual's identity. To remain compliant with institutional policy, individuals were also required to provide their SSN during the screening and enrollment call. While SSNs could not be checked in real time, we ultimately observed that of the 202 of participants who provided this information and had their SSNs checked, 98.5% (n=199) provided accurate information. Systematically checking SSNs before full study enrollment would likely reduce deceptive enrollments into research studies but would come with costs (eg, staff time and effort, some qualified individuals may refuse to share their SSN).

Some of the deception detection methods that were used in this study were previously employed in other studies [7]. However, some deception detection strategies were novel to this study (eg, requiring photo identification in a remotely conducted trial and conducting phone number checks). Study results demonstrated that using multiple deception abatement strategies can mitigate deceptive enrollments in remotely conducted clinical trials. Future studies should implement similar strategies and further innovate to reduce deceptive attempts at enrollment into clinical trials.

Limitations

The true sensitivity and specificity of our deception detection and abatement processes are unknown. It is possible that some individuals may have successfully used deception to enroll in the study, and it is possible that some individuals were unnecessarily excluded. In order to deceptively enroll in the study, individuals would have had to fake their identification, mailing address, phone number, and SSN. Following recommended practices to preserve data integrity [19], we conservatively disqualified individuals who had been flagged as potentially engaging in deceptive practices. It is possible, however, that some qualified individuals were excluded from participating in the study.

Future Directions

One of the goals of this work is to inform the development of more sophisticated, digital solutions to deception detection and abatement. Using automation rather than manual deception detection may increase accuracy and reduce staff burden. Further, sensors (eg, GPS) could be integrated into deception detection procedures to verify home addresses and participant locations. Future studies should calculate study costs and potential cost savings of implementing deception detection and abatement procedures.

Conclusions

We implemented a systematic deception detection procedure to reduce the likelihood of deceptive enrollments into a nationwide RCT. To maintain data integrity and participant safety, we recommend engaging in robust identification practices with study participants in a live enrollment call. Before this phase, personnel and financial resources can be preserved by implementing systematic, digitally mediated processes to flag cases of potential deception. Technologies and study operating procedures centered on analyzing incoming screening survey data and their associated metadata can help to efficiently screen out instances of likely deception. More research is needed to improve upon and automate these processes, and researchers must remain vigilant to the advent of new threats. While deception detection procedures are primarily aimed at reducing the likelihood of enrolling individuals who do not actually meet the study inclusion criteria, they can also positively impact staff time, motivation, job satisfaction, and the study budget.

Acknowledgments

Programming and technical support were provided through the mobile health shared resource of the Stephenson Cancer Center via an NCI Cancer Center Support Grant (P30CA225520). Data analysis and manuscript preparation were additionally supported through R00DA046564 and the Oklahoma Tobacco Settlement Endowment Trust grant R22-02.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

MSB is the primary inventor of the Insight mobile health (mHealth) platform, which was used to develop the Exemplar app. MSB receives royalties related to the Insight platform, but he did not receive royalties in this case because he was the lead investigator for this study.

References

1. Businelle MS, Perski O, Hébert ET, Kendzor DE. Mobile health interventions for substance use disorders. Annu Rev Clin Psychol 2024;20(1):49-76 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080822-042337] [Medline: 38346293]
2. He Y, Xie Z, Li J, Meng Z, Xue D, Hao C. Global trends in mHealth and medical education research: bibliometrics and knowledge graph analysis. JMIR Med Educ 2024;10:e52461 [FREE Full text] [doi: 10.2196/52461] [Medline: 38841983]

XSL•FO
RenderX

3. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-11-13]
4. Lyles CR, Nguyen OK, Khoong EC, Aguilera A, Sarkar U. Multilevel determinants of digital health equity: a literature synthesis to advance the field. Annu Rev Public Health 2023;44:383-405 [FREE Full text] [doi: 10.1146/annurev-publhealth-071521-023913] [Medline: 36525960]
5. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
6. van Dijk JAGM. Digital divide research, achievements and shortcomings. Poetics 2006;34(4-5):221-235. [doi: 10.1016/j.poetic.2006.05.004]
7. Dahne J, Tomko RL, McClure EA, Obeid JS, Carpenter MJ. Remote methods for conducting tobacco-focused clinical trials. Nicotine Tob Res 2020;22(12):2134-2140 [FREE Full text] [doi: 10.1093/ntr/ntaa105] [Medline: 32531046]
8. Lee CP, Holmes T, Neri E, Kushida CA. Deception in clinical trials and its impact on recruitment and adherence of study participants. Contemp Clin Trials 2018;72:146-157 [FREE Full text] [doi: 10.1016/j.cct.2018.08.002] [Medline: 30138717]
9. Kolata GB. The death of a research subject. Hastings Cent Rep 1980;10(4):5-6. [doi: 10.2307/3560443]
10. Apseloff G, Swayne JK, Gerber N. Medical histories may be unreliable in screening volunteers for clinical trials. Clin Pharmacol Ther 1996;60(3):353-356. [doi: 10.1016/s0009-9236(96)90063-6]
11. Devine EG, Waters ME, Putnam M, Surprise C, O'Malley K, Richambault C, et al. Concealment and fabrication by experienced research subjects. Clin Trials 2013;10(6):935-948. [doi: 10.1177/1740774513492917] [Medline: 23867223]
12. Bentley JP, Thacker PG. The influence of risk and monetary payment on the research participation decision making process. J Med Ethics 2004;30(3):293-298 [FREE Full text] [doi: 10.1136/jme.2002.001594] [Medline: 15173366]
13. Gelinas L, Largent EA, Cohen IG, Kornetsky S, Bierer BE, Fernandez Lynch H. A framework for ethical payment to research participants. N Engl J Med 2018;378(8):766-771. [doi: 10.1056/nejmsb1710591]
14. Fry C, Dwyer R. For love or money? An exploratory study of why injecting drug users participate in research. Addiction 2001;96(9):1319-1325. [doi: 10.1046/j.1360-0443.2001.969131911.x] [Medline: 11672496]
15. Abadie R. The Professional Guinea Pig: Big Pharma and the Risky World of Human Subjects. United States: Duke University Press; 2010.
16. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
17. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
18. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
19. Resnik DB, McCann DJ. Deception by Research Participants. N Engl J Med 2015;373(13):1192-1193 [FREE Full text] [doi: 10.1056/nejmp1506985]
20. Businelle MS, Hébert ET, Shi D, Benson L, Kezbers KM, Tonkin S, et al. Investigating best practices for ecological momentary assessment: nationwide factorial experiment. J Med Internet Res 2024;26:e50275 [FREE Full text] [doi: 10.2196/50275] [Medline: 39133915]
21. Insight mHealth. https://healthpromotionresearch.org/Mobile-Health-Technology. 2015. URL: https://healthpromotionresearch.org/Mobile-Health-Technology; [accessed 2025-01-04]
22. Arozullah AM, Yarnold PR, Bennett CL, Soltysik RC, Wolf MS, Ferreira RM, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med Care 2007;45(11):1026-1033. [doi: 10.1097/MLR.0b013e3180616c1b] [Medline: 18049342]

Abbreviations

CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart
EMA: ecological momentary assessment
mHealth: mobile health
OUHSC: University of Oklahoma Health Sciences Center
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SSN: social security number
VoIP: voice over internet protocol

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.09.24; peer-reviewed by R Pozzar; comments to author 23.10.24; revised version received 25.11.24; accepted 21.12.24; published 28.01.25.

Please cite as:
Kezbers KM, Robertson MC, Hébert ET, Montgomery A, Businelle MS
Detecting Deception and Ensuring Data Integrity in a Nationwide mHealth Randomized Controlled Trial: Factorial Design Survey Study
J Med Internet Res 2025;27:e66384
URL: https://www.jmir.org/2025/1/e66384
doi:10.2196/66384
PMID:

©Krista M Kezbers, Michael C Robertson, Emily T Hébert, Audrey Montgomery, Michael S Businelle. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Online-Based Recruitment Methods for Community-Dwelling Older Adults: Scoping Review and Lessons Learned From the PLAN Trial

Deborah Min[1,2], MPH; Ji-Young Yun[3], PhD; Chad Parslow[3], MPP, MSN; Anushka Jajodia[3], MA; Hae-Ra Han[1,3], MSN, RN, PhD

[1]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[2]Grossman School of Medicine, New York University, New York, NY, United States

[3]School of Nursing, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Hae-Ra Han, MSN, RN, PhD
School of Nursing
Johns Hopkins University
525 North Wolfe Street
Baltimore, MD, 21205
United States
Phone: 1 14106142669
Fax: 1 4105025481
Email: hhan3@jhu.edu

Abstract

Background: Despite rapid technological advancement and a considerably aging US population, there remains a gap in the literature pertaining to online-based recruitment strategies for older adults.

Objective: This study aimed to describe the lessons learned from the authors' experience of recruiting a sample for PLAN (Preparing successful aging through dementia Literacy education And Navigation), an ongoing, community-based randomized controlled trial designed to promote the transition of community-dwelling Korean American older adults with probable dementia and their caregivers into the health care system. The authors also present online-based recruitment strategies focused on older adults reported in relevant published studies to compare with their experiences.

Methods: Data sources included PLAN recruitment tracking files, study team meeting minutes, and interviews with community consultants. We also conducted a scoping review of published studies, searching PubMed in July 2021, and updated our search in September 2023. Eligibility criteria included (1) focus on older adults aged more than 65 years, (2) sample recruited from a community setting, and (3) inclusion and description of online-based recruitment strategies. Exclusion criteria (1) did not focus on adults older than 65 years in a community setting, (2) did not include or describe online-based recruitment strategies, or (3) used online-based methods but not for the purpose of recruitment. The review followed the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews). Information was extracted using a data charting table and synthesized by conducting a thematic analysis.

Results: In total, 8 articles were included in the scoping review and primarily addressed health promotion and recruitment strategy evaluation. When compared with PLAN data sources, five key themes emerged as relevant to the online-based recruitment of community-dwelling older adults: (1) unfamiliarity with technology—limited digital literacy, (2) differences in internet access and use across older age groups, (3) providing technological support to promote recruitment, (4) successful and unsuccessful recruitment using social media, and (5) other diverse online-based methods of recruitment. In particular, direct quotes from multiple sources for the PLAN trial revealed technological challenges that were common among immigrant older adults as the study team used various online-based recruitment activities.

Conclusions: The literature was limited in the discussion of online-based recruitment among older participants. Data sources revealed the digital divide and limited digital literacy, particularly among non–English-speaking immigrant older adults and their caregivers. The usefulness of online-based recruitment of older adults is uncertain due, in large part, to limited sociodemographic diversity noted in the samples recruited in the included studies. Future research should explore the role of race and ethnicity and

XSL•FO
RenderX

other characteristics, such as socioeconomic status, sex, education, access to technology, and digital literacy, in relation to online-based recruitment for adequate representation of diverse older adults in research.

Trial Registration: ClinicalTrials.gov NCT03909347; https://clinicaltrials.gov/study/NCT03909347

(J Med Internet Res 2025;27:e55082) doi:10.2196/55082

KEYWORDS

older adults; online; online recruitment; community-dwelling; strategies; America; Americans; technology adoption; digital technologies; COVID-19; pandemic; digital health; dementia; caregivers; healthcare system; community health workers; consultants; mobile phone

Introduction

The US population is rapidly aging and expected to experience continued significant growth. According to the US Census Bureau, the population aged 65 years and more surged by 38% from 40.3 million in 2010 to 55.8 million in 2020, compared with a 7% increase in the total population [1]. More than 1 in 5 (22%) Americans will be aged 65 years and more by 2040, reaching 80.8 million in total [2]. The increasing proportion of older adults in the United States distinctly positions this group as a priority study population.

While American adults have seen overall gains in technology adoption and increased dependence on digital technologies, especially during the COVID-19 pandemic, the digital divide—the unequal access to digital technology and the internet—persists among older adults aged 65 years and older. For example, internet usage in 2021 was nearly ubiquitous among adults aged 18-64 years (18-29 years: 99%; 30-49 years: 98%; and 50-64 years: 96%) meanwhile internet usage trailed among older adults aged 65 years and older (75%) [3]. Nevertheless, the older adults' group has also reported a steady increase in internet use over time, up from 14% in 2001, to 46% in 2011, to 75% in 2021 [3].

The COVID-19 pandemic presented unprecedented challenges to study teams with respect to the recruitment of study participants in community-based clinical trials. Globally, COVID-19–related restrictions were implemented including nationwide lockdowns and social distancing, and study teams had to quickly adjust study protocols to operate in an online-based environment. Engaging online with potential participants for recruitment activities (ie, identification, eligibility verification, informed consent, and enrollment) may be particularly challenging when the study population is older adults due to limited digital literacy [4]. Furthermore, limited digital access is also more common among older adults [5].

Despite rapid technological advancement and a considerably aging US population, there remains a gap in the literature pertaining to online-based recruitment strategies for older adults. Extant reviews related to online-based recruitment strategies included the examination of health research recruitment via Facebook (Meta) among adolescents [6], adults [7], and participants of all ages [8,9], as well as the examination of diverse digital tools for recruitment and retention of participants in randomized controlled trials (RCTs) [10], web-based mobile health studies [11], and recruitment performance of web-based respondent-driven sampling [12]. To our knowledge, a review specific to online-based recruitment methods among older adults has not been conducted.

PLAN (Preparing successful aging through dementia Literacy education And Navigation) is an ongoing, community-based RCT designed to promote the transition of community-dwelling Korean American older adults with probable dementia and their caregivers into the health care system for adequate diagnostic follow-up and care. The original protocol was designed to include community-based in-person outreach as the main recruitment approach. The start of PLAN recruitment coincided with the national lockdown and in-person activity restrictions resulting from the COVID-19 pandemic. Consequently, our study team had to devise online-based strategies to recruit study participants. The purpose of this paper is to describe our lessons learned from the experience of recruiting a sample comprised non–English-speaking older individuals who are cognitively impaired along with their caregivers for the PLAN trial. We also present our findings from a scoping review examining online-based recruitment strategies focused on older adults to compare with our experiences.

Methods

Description of the PLAN Trial

The PLAN trial uses an RCT design (ClinicalTrials.gov NCT03909347). Details of the study design and methods are described elsewhere [13]. Briefly, the study intervention consists of 1-hour dementia literacy education followed by monthly phone counseling for 6 months, all delivered by trained community health workers (CHWs). The study is dyad-based and the inclusion criteria for Korean American older adults are (1) self-identified as first-generation Korean American in the New York-New Jersey or Greater Washington Metropolitan Areas, (2) aged more than 65 years, (3) has probable dementia (1+ on clinical dementia rating), (4) has a caregiver who lives in the same household or has at least weekly interaction, and (5) able to consent or has a proxy available for consent. The study inclusion criteria for caregivers are (1) older than 18 years, (2) able to read and speak Korean, (3) living in the same household with the Korean American older adult or has at least weekly interactions, and (4) written consent to participate in the study and permission granted to the study team to audit medical records for linkage to health care. Eligible dyads (ie, Korean American older adult scoring clinical dementia rating 1+ and caregiver) are asked to fill out the study questionnaire at baseline, 3 months, and 6 months.

XSL•FO
RenderX

The trial was launched in late February 2021, when the COVID-19 pandemic was rampant and in-person recruitment activities were not allowed. Therefore, potential participants were recruited through ethnic mass media advertisements such as newspapers, television, and radio. The study team also began to use other online-based recruitment methods including the web pages of the participating community centers and local ethnic commercial websites as well as digital social platforms including Facebook, Instagram (Meta), Twitter (rebranded to X), YouTube (Google), and KakaoTalk (Kakao Corporation), a mobile messaging app popular among Koreans. The study team continues to use such online-based recruitment strategies to date.

Sources of Data and Study Selection

Data sources used to identify online-based recruitment strategies and challenges for the PLAN trial included study recruitment tracking files and study team meeting minutes from the period of February to August 2021 when PLAN recruitment activities were solely online-based due to restrictions resulting from the COVID-19 pandemic. Additional data sources included interviews that took place in August 2021 immediately following the recruitment restriction period; interviewees included PLAN CHWs and community site coordinators as well as community consultants. We compared key lessons learned from the PLAN trial with a scoping review of published studies using digital technology to recruit community-dwelling older adults.

Our review followed the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) [14]. The review protocol is not registered. In consultation with a health science librarian, we formulated a comprehensive search strategy comprising variations of the following search terms: “digital technology,” “recruitment,” “strategies,” and “older adults.” Additional restrictions applied included peer-reviewed literature published in English and with full-text availability. The detailed search strategy is presented in Multimedia Appendix 2. Our initial literature search involved the electronic database PubMed and was conducted on July 13, 2021, and updated on September 20, 2023.

The initial search yielded 505 nonduplicative records including titles and abstracts. Furthermore, 2 reviewers (among DM, J-YY, CP, and AJ) were assigned to screen each title and abstract independently for inclusion eligibility, and disagreements were resolved through consensus. Records were reviewed based on 3 eligibility criteria, that were (1) focus on older adults aged more than 65 years, (2) sample recruited from a community setting, and (3) inclusion and description of online-based recruitment strategies. Records were excluded if they (1) did not focus on older adults aged more than 65 years in a community setting, (2) did not include or describe online-based recruitment strategies, or (3) used online-based methods but not for the purpose of recruitment. After titles and abstracts were screened, a total of 103 full-text records were assessed for eligibility by 1 reviewer (DM) and a second reviewer (J-YY, CP, or AJ) reviewed every tenth record. In total, 6 studies met the eligibility criteria from the initial search. Using the same strategy, the second search yielded 91 additional titles and abstracts, with no duplicate records found. In addition, 2 reviewers (among DM, J-YY, and AJ) independently screened each title and abstract and 12 records entered the full-text review. One reviewer (DM) assessed each full-text record and a second reviewer (J-YY) reviewed every fourth record. With 2 records meeting the eligibility criteria in the updated search, a total of 8 records were included in this review. Articles were screened using Covidence (Veritas Health Innovation Ltd), a systematic review management platform.

Data Extraction

Using a data charting table in Microsoft Excel, we systematically extracted the following information from each of the articles included in our review: first author’s last name, study design, study goal, country, sample characteristics (including sample size, key inclusion criteria, participant age cut-off or range, percentage of female participants, and percentage of White participants), recruitment methods along with associated yield rates, challenges related to technology and older adults, lessons learned, and identified themes.

Data Analysis

We used a variety of methods to analyze relevant data from the PLAN trial and 8 relevant studies. Specifically, we used descriptive statistics, including frequencies and percentages, to calculate recruitment yield rates by different recruitment sources for the PLAN trial. For text-based data derived from the PLAN study team’s meeting minutes, stakeholder interviews, and recruitment-related information extracted from the 8 studies, we conducted a thematic analysis [15] to identify common themes related to recruiting older adults using online-based methods. To do this, the data were systematically coded to identify themes, which were then examined to identify those that were most common and relevant. The coding process informed the timing of interview completion, as we assessed data saturation by noting when no new themes emerged from the analysis.

Methodological rigor was ensured through the following approaches. First, for interview data collected in Korean from PLAN CHWs, community site coordinators, and Korean American community consultants, only the final results—themes, subthemes, and selected quotes—were translated and presented in English. This strategy minimized methodological challenges associated with repeated translations. Second, trustworthiness was established through several measures: (1) credibility was maximized by leveraging the bilingual research team’s extensive experience with the Korean American population and involving community consultants who are trusted members of the Korean American community; (2) dependability was enhanced through the use of multiple data sources and methods, ensuring triangulation of findings; (3) transferability was achieved by including verbatim transcripts and relevant quotes, enabling readers to assess the applicability of the study findings beyond the current context; and (4) confirmability was strengthened by revisiting the data through collaborative review by the initial coders and coauthors [16].

XSL•FO
RenderX

Ethical Considerations

The Johns Hopkins Medicine Institutional Review Board reviewed and approved procedures of the human subjects research component of this study (IRB00242241). Written consent was not deemed as required by the institutional review board. Study staff members assured anonymity and confidentiality of participant information and responses and participants were not compensated for their participation. For the scoping review, ethics approval was not required as published articles were used.

Results

Overview

The initial and updated searches yielded 505 and 91 nonduplicative records, respectively, including titles and abstracts. Of these, 103 and 12 records from the initial and updated searches met the inclusion criteria for full-text review. A combined total of 8 records (6 from initial search and 2 from updated search) were included in this review (Figure 1).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) flow diagram.

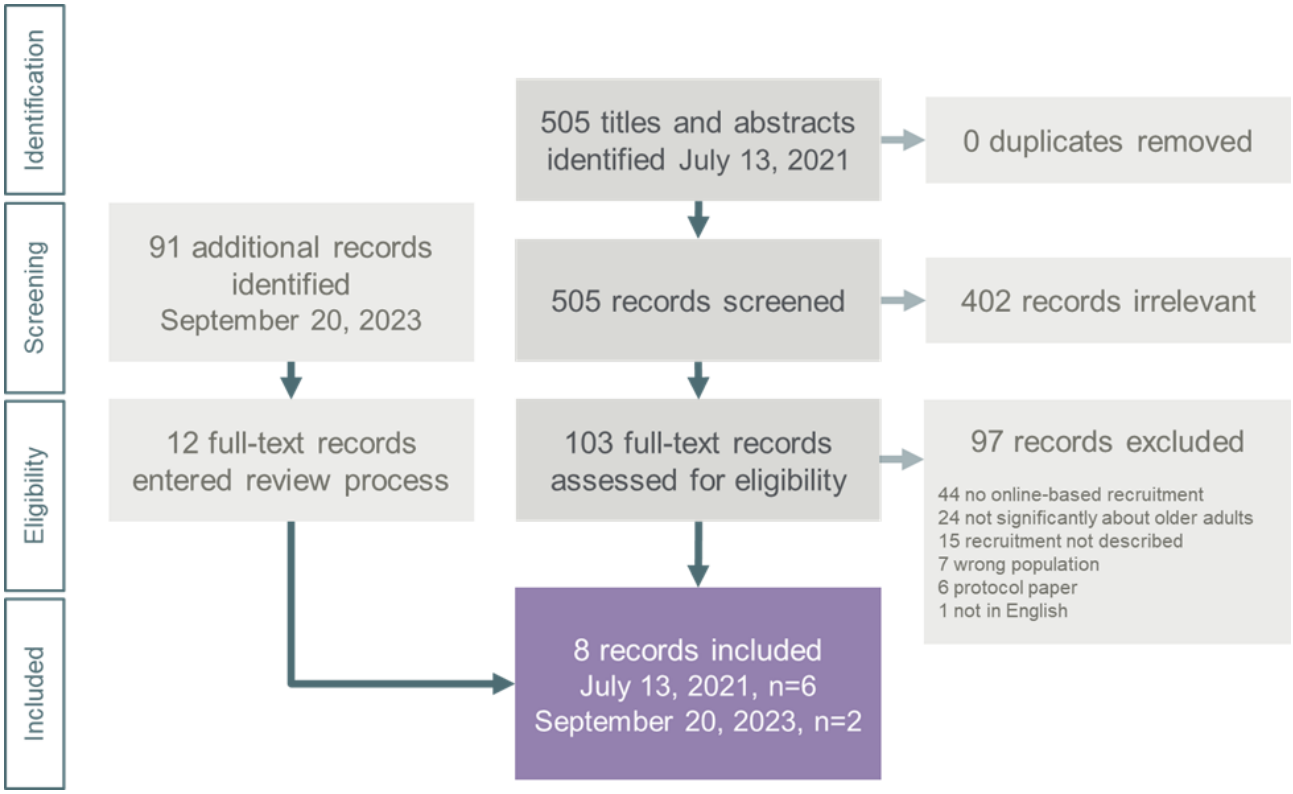

Characteristics of PLAN and Other Relevant Studies

Table 1 details key characteristics of the PLAN trial [13] and 8 relevant published studies [17-24] that used online-based strategies to recruit community-dwelling older adults. The scope of the relevant studies addressed health promotion and recruitment strategy evaluation. Of the 8 relevant studies, there were 3 observational studies [17,18,23], 2 secondary analyses [19,20], 2 RCTs [21,24], and 1 feasibility study [22]; 3 studies were conducted in the United States [18-20], while 2 studies were based in Australia [17,21], 2 studies based in Sweden [23,24], and 1 study based in Israel [22]. The 3 US-based studies involved English-speaking older adults with varying characteristics including 171 former caregivers of individuals with dementia (155/171, 91% White and 156/171, 91% female) [19]; 45 healthy older adults (90% White and 29/45, 64% female) [18]; and 795 adults, comprising 176 (22%) high-risk (having ischemic heart disease or congestive heart failure) older adults, 521 (66%) healthy older adults, and 98 (12%) healthy young adults (759/795 96% White and 504/795, 63% female) [20]. All 8 studies involved participants with normal cognition.

XSL•FO
RenderX

Table 1. Characteristics of studies using online-based recruitment strategies for older adults.

Reference	Study design; goal	Sample	Recruitment methods (% yield rate by source)	
			Online-based	Other
Han et al (PLAN[a]) [13]	RCT[b] (2 arms); to determine the effect of a community-based, CHW[c]-delivered intervention for linkage to dementia medical services	United States; 461 Korean American adults; older adults aged more than 65 years with probable dementia (85%) and their caregivers aged more than 18 years (15%); sex not collected at intake	Contribution[d,e]: Facebook (0%), Instagram (0%), Twitter (0%), YouTube (0%), and KakaoTalk (1%); Korean community center e-newsletters (6%); Korean community websites (5%); local eHealth fair events and seminars (2%); and participation in local dementia caregiver online support groups (1%)	Contribution[d,e]: Adult day care centers (4%); senior centers (3%); churches (4%); newspaper (42%), radio (0%), local Korean television (1%); study flyers mailed to sites such as Korean markets, restaurants, clinics (9%); Meals on Wheels (2%); and word of mouth (20%)
Bajraktari et al [23]	Observational study; to evaluate the reach of a digital fall prevention intervention	Sweden; 173 Swedish older adults older than 70 years who previously experienced a fall or decline in balance in the past year; 70% (n=121) female	Contributions[f]: Local municipality Facebook page (11%) and local municipality homepage (1%)	Contribution[f]: Mailed brochure (47%); older adult meeting place coordinator (13%); newspaper, television, and radio (8%); family or acquaintance (5%); municipality staff (4%); senior center (2%); Friskis&Svettis (2%); patient associations (2%); and not specified (4%)
Bracken et al [17]	Observational study; to evaluate the cost and effectiveness of a range of promotional strategies used to recruit men to a large type 2 diabetes prevention trial	Australia; 1007 men aged 50-74 years with high-risk for developing diabetes; race or ethnicity breakdown not specified	Contribution[e,f]: Facebook page and advertising (2%, 2%); other internet (3 Google AdWords campaigns, study website, links on other websites; 1%, 2%); community promotions (4%, 5%); newspaper ads (0%, 0%); and nearby football club promotion (0%, 0%)	Contribution[e,f]: Mainstream media: radio (42%, 40%), television (20%, 22%), newspaper news (3%, 3%); mass mail-outs (17%, 17%); word of mouth (3%, 3%); health care provider (3%, 2%); radio news story (1%, 1%); and not specified (3%, 2%)
Cowie and Gurney [18]	Observational study; to demonstrate the effectiveness of using targeted Facebook advertising on Facebook to recruit for a Phase 1 clinical trial to assess safety, tolerability, and preliminary cognitive benefit of a compound being developed for the treatment of Alzheimer disease	United States; 45 healthy older adults aged 60-78 years; 64% (n=29) female; 90% White	Contribution[e]: Contract Research Organization website and intranet (10%) and Facebook advertising (73%)	Contribution[e]: Word of mouth, referral, and event (8%); print and newspaper ads (7%); and poster, flyer, direct mail, and billboard (4%)
Corey et al [19]	Secondary data analysis; to describe the use of internet-based recruitment in obtaining a sample inclusive of young and middle aged, young-old, and older-old former dementia caregivers	United States; 171 former dementia caregivers; 18-64 years (n=15, 9%), 65-74 years (n=51, 30%), more than 75 years (n=105, 61%); 91% (n=156) female; 91% (n=155) White	Contribution[f]: Facebook page (31%); caregiving-related websites (15%); and dementia-related websites (11%)	Contribution[f]: Professional referrals (42%) and other (1%)
Dill et al [20]	Secondary data analysis; to examine recruitment methods for a 5-year study comparing immune responses to an inactive influenza vaccine in older adults versus younger adults	United States; 795 adults; high-risk (having ischemic heart disease or congestive heart failure) older adults aged more than 60 years (n=176, 22%); healthy older adults aged more than 60 years (n=521, 66%); healthy young adults (n=98, 12%); 63% (n=504) female; 96% (n=759) White	Contribution[e]: Broadcast emails within research institution (4%)	Contribution[e]: Mailed letters (83%); physical referrals (5%); word of mouth and friend referrals (5%); and others (including newsletters, flyer, and seminars; 4%)

XSL•FO
RenderX

Reference	Study design; goal	Sample	Recruitment methods (% yield rate by source)	
			Online-based	Other
Miller et al [21]	RCT; to describe the recruitment strategies used and the success of each approach in recruiting older adults with type 2 diabetes into a 6-month community-based exercise and nutritional supplementation RCT	Australia; 198 adults aged 50-75 years with type 2 diabetes; 50-59 years (39%), 60-69 years (53%), older than 70 years (11%); 35% (n=70) female; race and ethnicity breakdown not specified	Contribution[e,f]: Facebook advertising (0%, 0%) and website advertising (8%, 10%)	Contribution[e,f]: Targeted mail-outs (41%, 39%); state-wide newspapers (28%, 27%); local newspapers (15%, 14%); physician referrals (2%, 5%); word of mouth (3%, 3%); flyers and community presentations (1%, 2%); radio advertising (1%, 0%); and not specified (0%, 0%)
Pettersson et al [24]	RCT; to describe recruitment strategies, reach, and participant characteristics for a digital fall-prevention intervention	Sweden; 1628 Swedish older adults aged more than 70 years who previously experienced a fall or decline in balance in the past year; 79% (n=1292) female	Contribution[f]: Social media including Facebook pages, paid Facebook advertisements, and Instagram (76%—breakdowns per platform not specified)	Contribution[f]: Newspaper advertisements (7%); family and friends (6%); senior citizen organizations (6%); articles in newspapers or radio (2%); other (2%); and not specified (1%)
Schwartz et al [22]	Feasibility study; to promote physical activity among older adults through live, online group training sessions over 8 weeks among older adults	Israel; 31 Israeli older adults aged 68-76 years medically approved to participate in moderate physical activity by a physician; 65% (n=20) female	Contribution[e,f]: Posts in Facebook groups (100%, 100%)	—[g]

[a]PLAN: Preparing successful aging through dementia Literacy education And Navigation.

[b]RCT: randomized controlled trial.

[c]CHW: community health worker.

[d]Referral sources from February 2021 to August 2021 (restrictions for in-person activities due to the COVID-19 pandemic).

[e]Contribution defined as the percentage of all potential participant expressions of interest resulting from a particular recruitment method.

[f]Contribution defined as the percentage of all participant enrollments resulting from a particular recruitment method.

[g]Not applicable.

Table 1 also summarizes key recruitment methods used in each study along with yield rates, indicated as contributions, by source. Notably, 2 types of yield rates were found across PLAN and the 8 studies, that were (1) the percentage of all potential participant expressions of interest resulting from a particular recruitment method [13,17,18,20-22], and (2) the percentage of all participant enrollments resulting from a particular recruitment method [17,19,21-24]. Online-based recruitment methods for the PLAN trial included YouTube and social media platforms such as Facebook, Instagram, Twitter (rebranded as X), and KakaoTalk, of which only KakaoTalk served as a referral source for potential participants (3/461, 1%). Community-based online strategies were more successful to draw interest; e-newsletters and websites by Korean community organizations (26/261, 6% to 28/461, 6%), local eHealth fair events and seminars (11/461, 2%), and dementia caregiver online support groups (5/461, 1%). Though not online-based, ethnic newspaper advertising was most successful to draw interest (194/461, 42%), followed by word of mouth (90/461, 20%). For the 8 relevant studies, the most widely used (all but 1 study by Corey et al [19]) online-based recruitment method was the social media platform Facebook, with varying yield rates from 0% to 100%, used in various forms such as Facebook pages [17,19,23,24], Facebook advertising [17,18,21,24], and Facebook groups [22]. Additional online-based recruitment strategies included other internet (Google AdWords campaigns, study websites, links on other websites) [17,23], listserve broadcast emails [20,24], and advertising on other and related websites [17-19,21,24] with varying yield rates from 0% to 15%.

Recruitment of Older Adults Using Online-Based Strategies

We identified several key themes and lessons learned relevant to the recruitment of community-dwelling older adults (Table 2). These included unfamiliarity with technology, differences in internet access and use across older age groups, providing technological support to promote recruitment, successful and unsuccessful recruitment using social media, and other diverse online-based methods of recruitment.

XSL•FO
RenderX

Table 2. Key themes for online-based recruitment of older adults.

Theme	Lessons learned
Unfamiliarity with technology—limited digital literacy	• Some participants struggle to understand basic concepts such as an "app" or smartphone versus 2G phone (PLAN[a]) • Korean American older adults may reject the use of new technology, such as Zoom, because they are not used to nor familiar with it (PLAN)
Differences in internet access and use across older age groups	• Online-based recruitment strategies are useful across age groups (young, middle-aged, young-old, and older-old), but provide greater access to young and middle-aged adults than older adults [19] • Older-old (75 years and older) participants recruited only through caregiving- and dementia-related websites and none through Facebook [19]
Providing technological support to promote recruitment	• Triaging participant technological capability through study intake call (PLAN) • Zoom downloading assistance and tutorial through 1:1 precall with CHW[b] (PLAN) • Providing technological support through 1:1 15-minute online-based introductory session with study team member [22]

Successful and unsuccessful recruitment using social media	• Social media yielded no results (Facebook, Instagram, Twitter, and YouTube), with the exception of KakaoTalk (1%; PLAN)
	• Facebook was primary social media platform used across studies and presented mixed results (PLAN) [17-19,21-23]
	  • Successful: Facebook groups (100%) [22], Facebook advertising (73%) [18], and Facebook page (31%) [19]
	  • Unsuccessful: Facebook advertising (0%) [21], Facebook unpaid page and paid advertising (2%) [17], and Facebook page (11%) [23]

Using other diverse online-based methods of recruitment	• Community center e-newsletters, ethnic community websites, e-health fair events or seminar, dementia caregiver online support groups with modest success (PLAN) • Emails more successful with younger individuals; physician-signed letters more successful with older adults [20] • Survey invitation to participate and survey link posted on websites of organizations related to caregivers and dementia less successful than Facebook pages; professional referrals most successful [19]

[a]PLAN: Preparing successful aging through dementia Literacy education And Navigation.

[a]CHW: community health worker.

Unfamiliarity With Technology—Limited Digital Literacy

Unfamiliarity with technology and limited technological proficiency can be a barrier to online-based recruitment for older adults. In the PLAN trial, we saw this among Korean American older adult participants who had varying levels of familiarity and experiences with technology like Zoom (Zoom Communications). Our CHWs reported that some Korean American older adults, when asked if they had a smartphone, could not distinguish between a smartphone and a 2G phone. Similarly, many older adults had difficulty finding the Google Play or the Apple Store to download an app such as Zoom or were unfamiliar with the word "app" or "Zoom." One CHW shared her experience about the level of ability to use technology of PLAN study participants:

> *When I sent a Zoom link to a participant, in most cases, I had to send the link via text message because most KA older adults use phones [as opposed to computers]. Also, for most of them who received the Zoom invitation via text message, it was their first time using Zoom. Only those who use computers were able to communicate via email.*

Community consultants shared a broad spectrum of responses to online-based study participation and affirmed that differences in technology use and familiarity have influenced the decisions of Korean American older adults to participate. A community consultant shared:

> *There are people who want to try new things, and there are people who absolutely reject things that they are not used to. Most older adults [I have encountered] are like that; not limited to Zoom, but because they don't want to do anything that they are not familiar with.*

However, others had previous experience and familiarity with Zoom through personal tasks such as attending telehealth medical visits and church meetings. Such older adults expressed positivity toward participating in PLAN via Zoom, as noted by another community consultant:

> *I did [Zoom] with my doctor about two weeks ago. If Zoom is the only way [to get a doctor's appointment], we would do it...*

Differences in Internet Access and Use Across Older Age Groups

Online-based recruitment may present different challenges and opportunities for older adults across different age groups. While rates of internet use and experiences of different age cohorts of older adults often go unacknowledged in the literature, Corey

XSL•FO
RenderX

et al [19] used a 4-step internet-based approach for the recruitment of previous caregivers of people with dementia to participate in an online-based survey. The approach included the steps, such as (1) creation of study Facebook page, study invitation, and survey link posted to Facebook page; (2) study invitation and survey link posted to caregiver- and dementia-related websites; (3) study invitation and survey link posted to Facebook pages of relevant groups; and (4) study invitation and survey link email to principal investigator's professional contacts [19]. Rates of referral through online-based sources were found to be comparable between the older-old subsample (46.7%) and the young-old subsample (47%), suggesting that internet accessibility and use between young-old and older-old adults is similar [19]. Nevertheless, diverse use of several online-based recruitment strategies was least successful in the recruitment of older-old (aged 75 years and older), with this age cohort only being recruited directly from caregiving- and dementia-related websites and none through Facebook [19].

Providing Technological Support to Promote Recruitment

When the PLAN study team conducted solely online-based research activities, from February through August 2021, early implementation revealed limited use of Zoom and varying levels of familiarity with technology among Korean older adults. One of the most common problems encountered by participants was the downloading process of Zoom. Community feedback supported the utility of technological assistance including that of a community consultant:

> *I don't know how to use [Zoom], but I can do it if someone explains it to me side by side. But I don't know how to go in and navigate through Zoom.*

Responding to community feedback, we integrated an additional component to the PLAN intake call to triage interested potential participants with respect to technological capacity by posing questions related to device (eg, laptop, desktop, or smartphone), WiFi access, previous Zoom experience, and availability of someone to assist with Zoom set up. Based on their responses, individuals were categorized as either green to indicate technological self-sufficiency, or yellow, red, or black to signify lower levels of technological capacity; individuals of the latter groups were scheduled for a 1:1 precall with a CHW for Zoom downloading assistance and a tutorial before the cognitive screening. Similarly, Schwartz et al [22] scheduled online-based introductory sessions through Zoom with each interested potential participant to promote recruitment and provide technological assistance. The 15-minute 1:1 sessions included a Zoom tutorial as well as a brief overview of the study goals and required equipment for participation.

Successful and Unsuccessful Recruitment Using Social Media

All but 2 studies included in the review used social media as one of several recruitment strategies. Among the studies that leveraged social media, Facebook was the primary platform with a mix of targeted, paid advertisements, and unpaid postings [17-19,21-23]. Overall, the success of social media as a recruitment tool for older adults was mixed. For example, in an exercise and nutrition RCT for diabetes control, Miller et al [21] used paid Facebook advertisements with targeted key words based on age and location to capture attention of potential participants. Despite 16,600 clicks, among the 1157 total expressions of interest across recruitment types, only 5 came from social media; among those expressions of interest, no study participants were deemed eligible [21]. Similarly, in a study on diabetes prevention, Bracken et al [17] used both paid Facebook advertisements as well as unpaid Facebook posts. Facebook recruitment yielded a small proportion of expressions of interest as well as participant enrollment (2% and 2%, respectively), slightly more than other internet (combination of Google AdWords campaigns, study website, and links on other websites; 1% and 2%, respectively); radio advertising accounted for the largest proportion of both expressions of interest in addition to participant enrollment (42% and 40%, respectively) in the trial [17]. In a digital fall prevention intervention, Bajraktari et al [23] used the Facebook page of a local municipality, which yielded relatively low enrollment of 11% (19/173) compared with a 47% (81/173) yield from brochures mailed to households with at least 1 community-dwelling person aged 70 years or older as identified by a register provided by the local municipality. Similarly, social media was not an effective means of recruitment for the PLAN trial; relative to other methods, only 1% (3/461) of those screened for the PLAN trial were recruited via Facebook, Instagram, Twitter, YouTube, or Kakao Channel (a feature of a popular messaging app among Koreans—KakaoTalk).

A few studies successfully used social media to identify and enroll older adult participants. For example, using Facebook groups as the primary means of recruitment, 1 study surpassed their goal of 30 participants for an online-based physical activity intervention (using Zoom) for older adults in Israel [22]. In addition, in a study leveraging only social media channels that were unpaid, 24% (12/51) of the study sample aged 65-75 years were recruited via Facebook, although 0% (N=0) among those aged 75 years or older [19]. Recruitment through Facebook (unpaid) consisted of a dedicated page, created and managed by the PI, and posts to Facebook pages of other stakeholder groups [19]. There were no other indications of what other applications or websites were used other than postings on Facebook pages. Another study focused specifically on the effectiveness of social media recruitment to an Alzheimer treatment clinical trial via Facebook, as the study was failing to achieve the recruitment targets from its first phase of traditional recruitment methods (eg, billboards, newspaper advertising, word of mouth, personal referrals, and direct mail) [18]. Compared with the 11-week period of traditional recruitment methods that resulted in 178 inquiries and 6 enrolled subjects, the 8-week period with Facebook advertising resulted in 691 inquiries and 39 enrolled subjects. In the study by Cowie and Gurney [18], older men had a slightly higher engagement rate with the Facebook ads compared with women. However, women were found to have a higher engagement rate with their ad appealing to altruism (the recruitment campaigns targeted "typical" recruits by highlighting financial incentives for participation and "altruistic" recruits by highlighting the need for help) [18].

XSL•FO
RenderX

Using Other Diverse Online-Based Methods of Recruitment

There were other, nonsocial media online-based methods used for the recruitment of older adults with varying ranges of success. In the PLAN trial, other online-based methods used included community center e-newsletters, pop-up advertisements on websites for community sites, and e-health fair events and seminars organized by the trial's 2 partnering community sites, and dementia caregiver online support groups. These diverse online-based methods yielded 16% (73/461) of potential participants referred to the study. Nonetheless, such methods were ineffective relative to traditional ethnic newspaper recruitment through which we identified 42% (194/461) of potential participants. Dill et al [20] used emails to recruit adults aged 20-40 years and older adults aged more than 60 years to compare immune responses with an inactive influenza vaccine between the groups and found that emails were more effective among younger individuals (57/94, 61%) compared with older individuals (25/645, 4%) [20]. Finally, 1 sample of former dementia caregivers was divided into 3 age groups (young and middle-aged: 18-64 years; young-old: 65-74 years; older-old 75+ years) to examine variation in internet-based recruitment and retention [19]. Corey et al [19] used websites of relevant caregiver and dementia organizations (Family Caregiver Alliance and Alzheimer's Association) and found that older-old adults were most frequently directed to the study survey through such relevant websites (3/15, 20% and 4/15, 27%, respectively) compared with the young and middle-aged and young-old groups.

Discussion

Principal Results

Knowledge about online-based recruitment strategies among older adults is limited. This study set out to describe lessons learned from a community-based dementia literacy education intervention in cognitively impaired non–English-speaking older adults and their caregivers, as well as to compare such experiences with what is presented in the literature with respect to recruitment strategies that are online-based and older adult–focused. We learned that the most widely used online-based recruitment method across studies was the social media platform Facebook, with yield rates spanning 0% to 100% [13,17,19,21,22]. Despite an ongoing increase of internet use among older adults [3], recruiting older adults using other online-based methods such as internet (Google AdWords campaigns, study websites, and links on other websites), listserve broadcast emails, and advertising on other and related websites resulted in little to modest success with yield rates ranging from 1% to 15% [17-22].

The digital divide—one of the most significant social determinants of health [25]—was a theme that was particularly relevant to Korean American older adults and their caregivers in the PLAN trial [13]. Through direct recruitment encounters with potential study participants, the CHWs in the trial noted a 2G phone (aka, a "flip phone") as a barrier to working toward online-based eligibility screening. A recent report by the Pew Research Center [26] revealed that in 2021, 97% of Americans owned a cellphone of some kind; 85% owned a smartphone. According to the report, smartphone ownership varied by age (from 96% for those aged 18-29 years to 61% for people aged more than 65 years) and level of education (from 75% for people with high school or less to 93% for those with college education). The rate of smartphone ownership was similar across White, Black, and Hispanic Americans. With the advent of a digital era, smartphone ownership often equates to connectivity to the world and access to the internet [27]. While it is unclear how many immigrant older adults use smartphones, our finding suggests the importance of addressing the issue of limited access to adequate digital devices in this vulnerable group.

Limited digital literacy was also a main barrier frequently reported in our and other studies included in the review [13,22]. Digital literacy is defined as "the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills." In the PLAN trial, our CHWs recognized the challenges associated with limited digital literacy among Korean American older adults who often lacked a basic understanding of key technologies to enable them to follow through instructions for Zoom-based recruitment procedures. In particular, our observation of many Korean American older adults rejecting to learn "new things" (such as Zoom) is similar to the finding reported in a study involving community-dwelling adults [28]. In the study, participants were asked to perform a short task on a tablet involving a brief search with the same number of steps using either a familiar or a less familiar app. The authors found that older participants who used unfamiliar technology felt older after using the technology. Given these findings, trials using online-based recruitment should consider technological support for older adults as part of the recruitment process. Indeed, technological support through diverse methods (eg, triaging potential participants based on technological capabilities or 1:1 precall for assistance) was used to promote recruitment in our and other studies [13,22]. In addition, in a recently published experimental study involving individuals aged 55 and older, a series of in-person workshops incorporating hands-on training activities using the participants' select mobile device, tablet, or smartphone (eg, touch and manual dexterity for mobile use, varying forms of communication such as text, image, voice, video, engaging tasks such as sharing of pictures, video, or other info, and other autonomous activities such as messaging for e-learning) resulted in significant improvement in digital literacy at 1-month follow-up [29]. Taken together, these results highlight some of the promising avenues to promote inclusiveness among older adults. In addition, future study teams may refer to a number of national and state-based resources that have been established to address digital literacy [30-33].

The usefulness of internet-based recruitment methods varied significantly among different age cohorts of older adults. Nevertheless, diverse use of several internet-based recruitment strategies was least successful in the recruitment of older-old (aged 75 years and older), with this age cohort only being recruited directly from caregiving- and dementia-related websites, which suggests the need for supplementary recruitment strategies for this age demographic rather than relying solely on internet-based methods [19]. Also, none of the recruited

XSL•FO
RenderX

older-old participants were directed to the survey from Facebook, which suggests purposeful internet use among this age demographic and that social media use in this age group may not be sufficient for the use of social media–based recruitment methods, suggesting that response bias may be a significant consideration when using the internet in studies including older adults [19].

There was limited sociodemographic diversity in older adult participants recruited through internet-based strategies [17-22]. For example, we found that participant demographics, especially sample composition in terms of race or ethnicity, were not always reported; when reported, it was mainly White samples than non-White samples [18-20]. Furthermore, older participants included in the studies using internet-based recruitment had a high level of education and middle-upper class socioeconomic background [22]. More research is needed to explore the possible relationship between race or ethnicity, education, socioeconomic status, and ability to access and use online-based resources. There is also a demonstrated need to address the feasibility of online-based recruitment and research among older adults from lower socioeconomic backgrounds [19,22].

Success with social media as a recruitment tool for older adults varied. Social media is internet-based digital technology that allows the sharing of information and ideas and includes platforms such as Facebook, YouTube, X (previously called Twitter), Instagram, or WhatsApp. Given its wide reach and efficiency, researchers often find social media as an attractive tool for recruitment. Nonetheless, a recent integrative review [34] of 96 studies on social media use for research participant recruitment revealed that with the exception of 1, all other studies included in the review exclusively involved younger populations (eg, teens, young adults, or middle-aged participants); none specifically targeted older individuals with limited English proficiency. One of the studies included in the current review was a diabetes prevention trial and used Facebook for recruitment [17]. Yet, it only focused on older men and the authors were unable to determine whether lack of engagement via online-based strategies was due to the content of the advertisements and posts or sex-related differences with respect to the social media habits of men over the age of 50 [17]. Given the small number of relevant studies using social media to recruit older adults, more research is necessary to understand the effects of age and sex in relation to the utility of social media platforms for recruitment.

Limitations

Our study has several limitations that warrant consideration. First, due to the timing of our exclusively online-based recruitment (early phase of COVID-19 pandemic), most potential participants who approached the study team had access to the internet. Naturally, we did not find any theme addressing the gap in internet access, although it is important to acknowledge that nationally, the rate of internet access among older adults aged 65 years and older is 75% [3]. The role of social media in our study is also noteworthy. For the most part, earlier articles using social media platforms for recruitment often lacked detailed information compared with more recent articles, providing gaps in reported online-based recruitment methodology and outcome. Furthermore, the articles in our review lacked ethnically diverse samples, and most studies were predominantly female, White, and with higher education, presenting a limitation in terms of generalizability of our findings (eg, 1 study used a Facebook algorithm to target individuals in higher income brackets [18]). Our review included articles published in English only and resulted in a total of 8 studies. The inclusion criteria around language restriction may have limited both the diversity and total number of articles, potentially influencing the level of comprehensiveness of our findings. In addition, due to the time-sensitive nature of our study, we used 1 primary database and selected PubMed given its widespread use and comprehensive coverage. This decision was made to maximize our search results within the constraints of limited time and resources. However, we note this as a limitation as relying solely on PubMed may have excluded relevant studies from other databases. Finally, we acknowledge the possibility of recalling bias. To minimize this, we used data sources that were based on written documentation (ie, recruitment tracking files and study team meeting minutes); in addition, interviews with CHWs, community site coordinators, and community consultants were conducted in August 2021, immediately following the 6-month period from February to August 2021 when PLAN recruitment activities were solely online-based due to restrictions resulting from the COVID-19 pandemic.

Conclusions

Recruiting older adults using online-based strategies had varying success with most studies included in the review reporting more challenges than successes. As our nation continues to age, it is also becoming increasingly diverse [35]. Limited internet access and digital devices were noted as some of the main reasons for unsuccessful recruitment in the published studies of older adults included in our review; the same issue was also observed in our ongoing trial. The persistence of an age-based digital divide in the post–COVID-19 pandemic warrants both careful planning of recruitment activities to ensure technological support during the pre-enrollment period and the analysis of relevant recruitment outcomes as part of the study design. Such endeavors may promote better representation of those in research who are particularly vulnerable to the digital divide such as the older old, the less educated, and non–English-speaking ethnic minorities.

Acknowledgments

The authors would like to acknowledge Ja'Lynn Gray for her contribution to an early version of this work. This study was supported by the National Institute on Aging (R01AG062649).

XSL•FO
RenderX

Data Availability

All data that support this study's findings will be made available from the corresponding author upon reasonable request.

Authors' Contributions

H-RH conceptualized the study and developed the methodology. H-RH and DM led the study. DM, J-YY, CP, and AJ screened the studies and performed the analysis. H-RH, DM, J-YY, CP, and AJ wrote, reviewed, and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) literature review checklist.
[PDF File (Adobe PDF File), 135 KB - jmir_v27i1e55082_app1.pdf]

Multimedia Appendix 2
PubMed search strategy.
[PDF File (Adobe PDF File), 75 KB - jmir_v27i1e55082_app2.pdf]

References

1. US older population grew from 2010 to 2020 at fastest rate since 1880 to 1890. United States Census Bureau. 2023. URL: https://www.census.gov/library/stories/2023/05/2020-census-united-states-older-population-grew.html [accessed 2023-08-25]
2. Administration for community living: U.S. department of health and human services. 2020 profile of older americans. 2021. URL: https://acl.gov/sites/default/files/aging%20and%20Disability%20In%20America/2020Profileolderamericans.final_.pdf [accessed 2022-11-15]
3. Internet/broadband fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2022-10-13]
4. Eyrich NW, Andino JJ, Fessell DP. Bridging the digital divide to avoid leaving the most vulnerable behind. JAMA Surg 2021;156(8):703-704. [doi: 10.1001/jamasurg.2021.1143] [Medline: 33769441]
5. Roberts ET, Mehrotra A. Assessment of disparities in digital access among medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
6. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr 2014;14(5):439-447.e4. [doi: 10.1016/j.acap.2014.05.049] [Medline: 25169155]
7. Reagan L, Nowlin SY, Birdsall SB, Gabbay J, Vorderstrasse A, Johnson C, et al. Integrative review of recruitment of research participants through Facebook. Nurs Res 2019;68(6):423-432. [doi: 10.1097/NNR.0000000000000385] [Medline: 31693547]
8. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
9. Whitaker C, Stevelink S, Fear N. The use of facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
10. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. Trials 2020;21(1):478 [FREE Full text] [doi: 10.1186/s13063-020-04358-3] [Medline: 32498690]
11. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
12. Helms YB, Hamdiui N, Kretzschmar MEE, Rocha LEC, van Steenbergen JE, Bengtsson L, et al. Applications and recruitment performance of web-based respondent-driven sampling: scoping review. J Med Internet Res 2021;23(1):e17564 [FREE Full text] [doi: 10.2196/17564] [Medline: 33448935]
13. Han H, Perrin N, Kwon SC, Joo J, Yun JY, Min D, et al. PLAN-dementia literacy education and navigation for Korean elders with probable dementia and their caregivers: rationale, methods, and design of a community-based, randomized, controlled, multi-site clinical trial. Contemp Clin Trials 2025;148:107771. [doi: 10.1016/j.cct.2024.107771] [Medline: 39622468]
14. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

XSL•FO
RenderX

15. Braun V, Clarke V. Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. Health Psychol Rev 2023;17(4):695-718. [doi: 10.1080/17437199.2022.2161594] [Medline: 36656762]
16. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks (CA): Sage; 1985.
17. Bracken K, Hague W, Keech A, Conway A, Handelsman DJ, Grossmann M, et al. Recruitment of men to a multi-centre diabetes prevention trial: an evaluation of traditional and online promotional strategies. Trials 2019;20(1):366 [FREE Full text] [doi: 10.1186/s13063-019-3485-2] [Medline: 31217024]
18. Cowie JM, Gurney ME. The use of facebook advertising to recruit healthy elderly people for a clinical trial: baseline metrics. JMIR Res Protoc 2018;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]
19. Corey KL, McCurry MK, Sethares KA, Bourbonniere M, Hirschman KB, Meghani SH. Utilizing internet-based recruitment and data collection to access different age groups of former family caregivers. Appl Nurs Res 2018;44:82-87 [FREE Full text] [doi: 10.1016/j.apnr.2018.10.005] [Medline: 30389065]
20. Dill J, Unson C, Kenyon-Pesce L, Kuchel GA, McElhaney J. Recruitment and retention of healthy and high-risk older adults for a multiyear influenza vaccination study. Conn Med 2014;78(8):481-486. [Medline: 25314888]
21. Miller EG, Nowson CA, Dunstan DW, Kerr DA, Solah V, Menzies D, et al. Recruitment of older adults with type 2 diabetes into a community-based exercise and nutrition randomised controlled trial. Trials 2016;17(1):467 [FREE Full text] [doi: 10.1186/s13063-016-1589-5] [Medline: 27669823]
22. Schwartz H, Har-Nir I, Wenhoda T, Halperin I. Staying physically active during the COVID-19 quarantine: exploring the feasibility of live, online, group training sessions among older adults. Transl Behav Med 2021;11(2):314-322 [FREE Full text] [doi: 10.1093/tbm/ibaa141] [Medline: 33447852]
23. Bajraktari S, Zingmark M, Pettersson B, Rosendahl E, Lundin-Olsson L, Sandlund M. Reaching older people with a digital fall prevention intervention in a Swedish municipality context-an observational study. Front Public Health 2022;10:857652 [FREE Full text] [doi: 10.3389/fpubh.2022.857652] [Medline: 35548075]
24. Pettersson B, Bajraktari S, Skelton DA, Zingmark M, Rosendahl E, Lundin-Olsson L, et al. Recruitment strategies and reach of a digital fall-prevention intervention for community-dwelling older adults. Digit Health 2022;8:20552076221126050 [FREE Full text] [doi: 10.1177/20552076221126050] [Medline: 36118253]
25. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
26. Mobile fact sheet. Pew Research Center. 2021. URL: https://tinyurl.com/y98uxvcf [accessed 2023-06-06]
27. Montag C, Błaszkiewicz K, Sariyska R, Lachmann B, Andone I, Trendafilov B, et al. Smartphone usage in the 21st century: who is active on WhatsApp? BMC Res Notes 2015;8:331 [FREE Full text] [doi: 10.1186/s13104-015-1280-z] [Medline: 26238512]
28. Caspi A, Daniel M, Kavé G. Technology makes older adults feel older. Aging Ment Health 2019;23(8):1025-1030. [doi: 10.1080/13607863.2018.1479834] [Medline: 30475067]
29. Quialheiro A, Miranda A, Garcia M, Carvalho ACD, Costa P, Correia-Neves M, et al. Promoting digital proficiency and health literacy in middle-aged and older adults through mobile devices with the workshops for Online Technological Inclusion (OITO) project: experimental study. JMIR Form Res 2023;7:e41873 [FREE Full text] [doi: 10.2196/41873] [Medline: 36753331]
30. eHealth literacy. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/healthliteracy/researchevaluate/eHealth.html [accessed 2023-06-12]
31. Digital health literacy. National Library of Medicine. 2022. URL: https://allofus.nnlm.gov/digital-health-literacy [accessed 2023-06-12]
32. Wisconsin broadband office. Public Service Commision of Wisonsin. 2023. URL: https://psc.wi.gov/Pages/ServiceType/Broadband.aspx [accessed 2023-06-12]
33. The digital navigator model: adding digital equity to our social safety net. National Digital Inclusion Alliance. URL: https://www.digitalinclusion.org/digital-navigator-model/ [accessed 2023-06-12]
34. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022;24(8):e38015 [FREE Full text] [doi: 10.2196/38015] [Medline: 35925655]
35. Colby SL, Ortman JM. Projections of the size and composition of the U.S. population. United States Census Bureau. 2015. URL: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf [accessed 2022-11-15]

Abbreviations

PLAN: Preparing successful aging through dementia Literacy education And Navigation
CHW: community health worker
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Coristine; submitted 01.12.23; peer-reviewed by P Costa, N Santos; comments to author 13.02.24; revised version received 01.03.24; accepted 09.12.24; published 25.02.25.

Please cite as:
Min D, Yun JY, Parslow C, Jajodia A, Han HR
Online-Based Recruitment Methods for Community-Dwelling Older Adults: Scoping Review and Lessons Learned From the PLAN Trial
J Med Internet Res 2025;27:e55082
URL: https://www.jmir.org/2025/1/e55082
doi:10.2196/55082
PMID:39998873

©Deborah Min, Ji-Young Yun, Chad Parslow, Anushka Jajodia, Hae-Ra Han. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Reliability and Quality of Videos as Guidance for Gastrointestinal Endoscopy: Cross-Sectional Study

Jinpei Liu[1*], MD; Yifan Qiu[2*], MD; Yilong Liu[3*], MD; Wenping Xu[3*], MD; Weichen Ning[1], MD; Peimei Shi[3], MD; Zongli Yuan[3], MD; Fang Wang[4], MD; Yihai Shi[1], MD

[1]Department of Gastroenterology, Gongli Hospital of Shanghai Pudong New Area, Shanghai, China

[2]College of Basic Medicine Sciences, Second Military Medical University/Naval Medical University, Shanghai, China

[3]Department of Gastroenterology, Changzheng Hospital, Naval Medical University, Shanghai, China

[4]Department of Respiratory and Critical Care Medicine, Changzheng Hospital, Naval Medical University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Yihai Shi, MD
Department of Gastroenterology
Gongli Hospital of Shanghai Pudong New Area
Pudong New Area 219 Miaopu Road
Shanghai, 200135
China
Phone: 86 5885873
Email: syh01206@163.com

Abstract

Background: Gastrointestinal endoscopy represents a useful tool for the diagnosis and treatment of gastrointestinal diseases. Video platforms for spreading endoscopy-related knowledge may help patients understand the pros and cons of endoscopy on the premise of ensuring accuracy. However, videos with misinformation may lead to adverse consequences.

Objective: This study aims to evaluate the quality of gastrointestinal endoscopy-related videos on YouTube and to assess whether large language models (LLMs) can help patients obtain information from videos more efficiently.

Methods: We collected information from YouTube videos about 3 commonly used gastrointestinal endoscopes (gastroscopy, colonoscopy, and capsule endoscopy) and assessed their quality (rated by the modified DISCERN Tool, mDISCERN), reliability (rated by the *Journal of the American Medical Association*), and recommendation (rated by the Global Quality Score). We tasked LLM with summarizing the video content and assessed it from 3 perspectives: accuracy, completeness, and readability.

Results: A total of 167 videos were included. According to the indicated scoring, the quality, reliability, and recommendation of the 3 gastrointestinal endoscopy-related videos on YouTube were overall unsatisfactory, and the quality of the videos released by patients was particularly poor. Capsule endoscopy yielded a significantly lower Global Quality Score than did gastroscopy and colonoscopy. LLM-based summaries yielded accuracy scores of 4 (IQR 4-5), completeness scores of 4 (IQR 4-5), and readability scores of 2 (IQR 1-2).

Conclusions: The quality of gastrointestinal endoscope-related videos currently on YouTube is poor. Moreover, additional regulatory and improvement strategies are needed in the future. LLM may be helpful in generalizing video-related information, but there is still room for improvement in its ability.

(J Med Internet Res 2025;27:e58855) doi:10.2196/58855

KEYWORDS

gastrointestinal endoscopy; YouTube; patient education; social media gastrointestinal; large language model; LLM; reliability; quality; video; cross-sectional study; endoscopy-related videos; health information; endoscopy; gastroscopy; colonoscopy

Introduction

Gastrointestinal endoscopy is a medical procedure performed by the insertion of an endoscope into the digestive tract through the oral and anal routes. It allows direct visualization of the inner lining of the esophagus, stomach, and intestines and enables the treatment of various conditions, such as ulcers, polyps, and tumors [1]. In recent years, the incidence and

XSL•FO
RenderX

gradual rejuvenation of gastrointestinal diseases have increased [2]. As a significant screening and diagnostic tool for gastrointestinal diseases, gastrointestinal endoscopy is increasing in importance. Gastroscopy and colonoscopy screening have significantly increased the rate of early diagnosis of gastrointestinal cancers, reduced mortality, and improved disease prognosis [3,4]. However, previous studies have shown that 38.1% of people refuse to undergo gastroscopy because they do not sufficiently recognize the knowledge related to this examination [5]. More seriously, some patients are delayed in the diagnosis and treatment of their disease because of misinformation about gastroscopy [6]. Capsule endoscopy is another important gastrointestinal examination technique that has emerged over the past few years. It has the advantages of being painless, noninvasive, and relatively less demanding to operate [7]. However, due to insufficient knowledge of the indications, contraindications, and effects of capsule endoscopy [8], it is difficult for patients to independently make a choice that is appropriate for their condition. Therefore, providing medical education to patients about the indications, contraindications, and key points of preoperative preparation for various gastrointestinal endoscopic techniques can help them receive appropriate treatment in a timely manner and may improve the outcome of their disease.

In recent years, the increasing popularity of mobile electronic devices and the development of internet applications have led to a rising utilization of social media in the medical field [9,10]. Social media platforms that disseminate health information web-based have been proven to enhance the efficiency of doctors' consultations and improve the experience of patients [11-13]. Due to the convenience of accessing information, an increasing number of patients are turning to social media platforms as a source of health-related information [11]. In the field of gastroenterology, social media-based telemedicine is actively involved in patient health care for patients and in doctors' decision-making [12]. It has been reported that communication through social media improves the quality of bowel preparation for patients undergoing colonoscopy [13]. Notably, the advent of large language models, such as ChatGPT (OpenAI), has facilitated faster access to endoscopic information [14]. Interestingly, a video content summarization tool built on the foundation of big language models offers a new option for individuals who are unable or unwilling to watch an entire video [15].

YouTube has emerged as a paramount platform for disseminating web-based content, boasting an impressive global user base of 256 million [16]. Moreover, the amount of medical content on this platform has increased significantly in recent years [17]. Used appropriately, platforms such as YouTube can prompt behavioral change in viewers, enhancing their health [18]. However, YouTube also allows the distribution of unverified and nonprofessional content [19], leading to an increase in low-quality videos, which are created by nonspecialist users with little to no medical training [6]. This compromises the overall quality of medical content on YouTube, providing potentially misleading health information to unsuspecting viewers. Therefore, additional efforts from health care professionals are needed to correct patients' misconceptions from the abovementioned misinformation during their treatments, reducing the efficiency of health care services and potentially escalating doctor-patient conflicts [20,21]. Currently, a substantial number of gastrointestinal endoscopy-related videos have emerged on YouTube. However, there is a lack of research evaluating the quality of these videos.

The aim of the present study was to examine the quality, reliability, and recommendability of videos related to gastroscopy, colonoscopy, and capsule endoscopy available on YouTube. Additionally, we also sought to assess the practicality of a large language model (LLM)–based video content summarization tool for summarizing health-related video content [15].

Methods

Search Strategy and Selection Approach for the Included Videos

All the videos included in the analysis were obtained from the web-based YouTube platform. The keywords used were "(gastroscopy) or (upper gastrointestinal endoscopy)," "(colonoscopy) or (lower gastrointestinal endoscopy)," and "capsule endoscopy." To minimize the impact of the personalized recommendation algorithm on the inclusion of videos, we registered 3 separate brand-new accounts to retrieve the 3 different endoscopies. No filter settings were used during the retrieval process to faithfully simulate the usage scenarios of patients or other regular users. The search process for videos of the 3 gastrointestinal endoscopies was completed between September 28, 2023, and September 30, 2023. Duplicate videos, non-English videos, and videos that were not relevant to the search topic were excluded. Based on the experience from previous studies and the browsing habits of video platform users, namely, they mostly prioritize top search results over extensive browsing on video platforms [22-27], we selected videos that appeared within the top 5 pages of the YouTube interface, which is a total of 60 results for each endoscope for subsequent analysis. The specific retrieval and filtering process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Search strategy, video screening procedure, and flowchart of this research. AI: artificial intelligence.

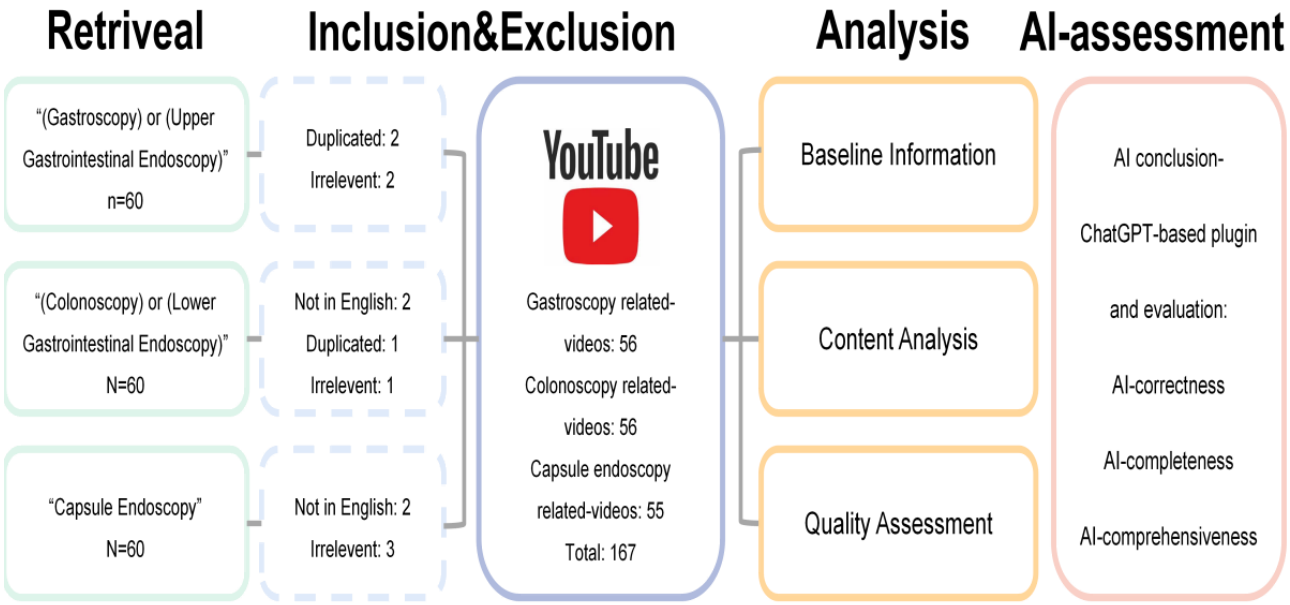

Evaluating Methodologies of the Videos

The reliability and content quality of the videos were evaluated using the Modified DISCERN (mDISCERN), the *Journal of the American Medical Association* (*JAMA*) score, and the Global Quality Scoring (GQS) scale. Before this assessment, the participating physicians familiarized themselves with the endoscopy guidelines and official descriptions of the 3 scales, enabling them to assess the accuracy and completeness of the video content while adhering to the guidelines. The mDISCERN scale was initially devised by Charnock et al [28] for evaluating written health information and was employed to examine video content quality [29]. The scale's modified version contains a 5-question questionnaire (Table S1 in the Multimedia Appendix), providing 1 point for each affirmative response for a maximum of 5 points. Videos scoring 3 or more on the mDISCERN scale were deemed high-quality health information providers [29]. The *JAMA* Rating Scale (Table S2 in the Multimedia Appendix) was used to assess health-related video information reliability [30]. It includes 4 assessment criteria: authorship, attribution, disclosure, and timeliness, with each carrying a 1-point score of up to a total of 4. Higher *JAMA* scores suggest better reliability of health information [30]. The GQS (Table S3 in the Multimedia Appendix) uses a 1-5 scale across 5 criteria, with greater GQS indicating a more highly recommended video [31]. Additionally, the artificial intelligence (AI)–summarized content was evaluated in terms of accuracy, reliability, and readability (Table S4 in the Multimedia Appendix), with accuracy-related elements divided into 6 dimensions from 1 to 6 and completeness and readability divided into 3 dimensions from 1 to 3.

Data Collection and Evaluation of Video Reliability and Content Quality

Essential video information was gathered for subsequent analysis, including video length, release date, number of days since release, play count, likes, steps, comments, uploader's follower count, uploader's identity, and the Video Power Index (VPI). We divided the uploaders into 4 groups based on their status: doctors or hospitals, medical media, other health organizations (including medical associations, medical examination centers, equipment companies, etc), and patients. The VPI, used to measure video popularity, was computed with the formula VPI=([number of likes]/[number of likes+number of dislikes])×100)×([number of views/days after upload]/100) [32]. Additionally, the ratios of (number of likes+number of steps)/plays and number of comments/plays were calculated to evaluate viewer interaction levels. The LLM processes the raw text extracted from video content, including both text and subtitles, and subsequently synthesizes a summary and general insights. To avoid sequence bias, the video order was randomized. All videos included in the analysis and the content synthesized by LLM were assessed by 2 senior gastroenterologists (YL and YF). For scoring items prompting disagreement, a third senior physician (WP) made the final decision.

Statistical Analysis

The nonnormal distribution of the entire data set was initially confirmed using the Shapiro-Wilk test. The median, interquartile range, mean, and variance were calculated for the continuous variables and the frequency of the categorical variables. For continuous variables that deviated from a normal distribution, the Wilcoxon-Mann-Whitney test and the Kruskal-Wallis test were used for between-group differences. All the data analyses were carried out with IBM SPSS Statistics for Mac (version 27; IBM Corp) via a 2-sided test approach. A *P* value of less than 0.05 was considered to indicate statistical significance. Moreover, for comparisons among multiple subgroups, Bonferroni corrections were applied to the *P* values. Finally, data visualization was performed using R software (R Core Team, the R Foundation).

XSL•FO
RenderX

Results

Video Features

A total of 167 relevant videos were included in the analysis; 56 were gastroscopy-related videos, 56 were colonoscopy-related videos, and 55 were capsule endoscopy-related videos. We divided the uploaders into 4 groups based on their status: doctors or hospitals, medical media, other health organizations (including medical associations, medical examination centers, equipment companies, etc), and patients (Figure 2A). Most of the videos were uploaded via platforms with medically related backgrounds, including medical media (n=52, 31.3%), other health organizations (n=49, 29.5%), and doctors or hospitals (n=39, 23.5%; Figure 2A). Colonoscopy-related videos were relatively longer, with a median of 288 (IQR 159-477) seconds, followed by gastroscopy-related videos (median 251, IQR 115-530) and capsule endoscopy-related videos (median 213, IQR 110-487). Capsule endoscopy-related videos were significantly less popular (median VPI 4.45, IQR 1.75-15.59) than gastroscopy-related (median VPI 32.47, IQR 7.77-88.20) and colonoscopy-related (median VPI 31.52, IQR 9.69-129.52) videos were. The trend in the number of views was also similar to that for the VPI; gastroscopy-related videos and colonoscopy-related videos both had significantly more plays than did capsule endoscopy-related videos, while there was no significant difference between the former two (Table 1).

Figure 2. The characteristics and quality of gastrointestinal-related videos on YouTube. (A) Proportion of the 4 groups of the uploaders. (B) Radar charts showing the percentage of each kind of content categorization among videos of different types of endoscopes. (C) mDISCERN score, (D) JAMA score, (E) GQS score of videos of different types of endoscopes. * denotes *P*<.05, and ** denotes *P*<.01 using the Kruskal-Wallis test. GQS: Global Quality Score; JAMA: Journal of the American Medical Association.

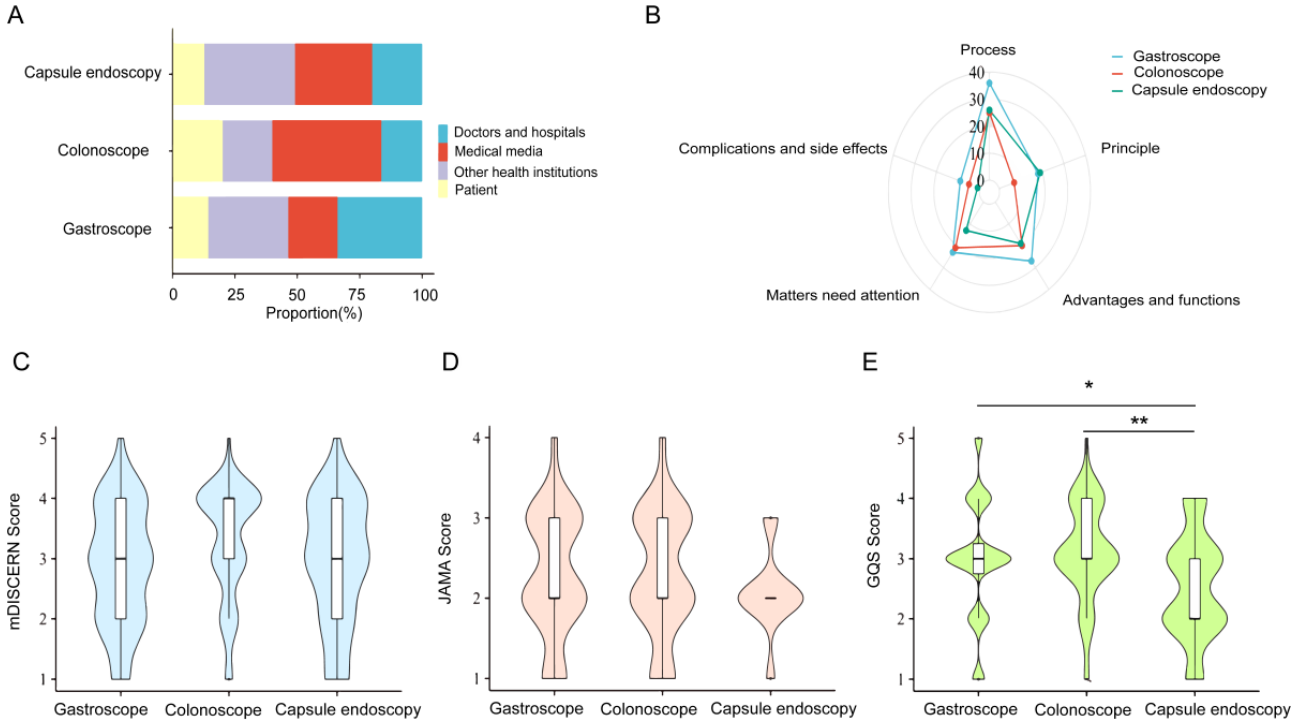

XSL•FO
RenderX

Table 1. Baseline characteristics of videos about gastroscopy, colonoscopy, and capsule endoscopy.

Video features	Gastroscopy		Colonoscopy		Capsule endoscopy	
	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)
VPI[a]	32.47 (7.77-88.2)[b]	267.11 (1318.94)	31.52 (9.69-129.52)[b]	400.16 (1231.56)	4.45 (1.75-15.59)	9.92 (11.90)
Days after upload	1810 (844-3937)	2288.23 (1681.80)	2024 (968-3279)	2213.11 (1410.74)	2225 (697-3442)	2270.27 (1553.14)
Number of views	74,402 (7967-211,364)[b]	402,609.09 (1,244,985.57)	56,967 (21,073-197,575)[b]	776,618.05 (2,775,980.36)	10,003 (1690-22,691)	23,024.84 (33,509.27)
Video length	251 (115-530)	350.41 (314.71)	288 (159-477)	404.41 (426.41)	213 (110-487)	322.35 (284.35)
Number of likes	301 (60-1128)	901.38 (1403.43)	376 (99-1653)	4951.29 (16,833.17)	26 (6-111)	86.93 (128.95)
Number of dislikes	18 (1-65)	75.04 (125.28)	18 (3-60)	236.63 (769.59)	0 (0-5)	4.78 (8.41)
Subscribers of the uploaders	16,900 (1970-130,000)	284,096.68 (1,085,446.40)	48,400 (8580-20,5000)	340,198.88 (1,022,888.32)	3090 (661-23,800)	25,285.58 (48,619.33)
Number of comments	35 (1-126)	96.94 (171.78)	35 (2-202)	350.65 (1137.62)	1 (0-20)	12.91 (22.56)
Number of likes+dislikes or views	49.07 (27.87-83.96)	87.49 (157.30)	62.69 (33.11-122.41)	94.84 (91.92)	44.09 (24.81-79.58)	83.54 (157.11)
Number of comments or views	1.51 (0-10.96)	13.20 (30.21)	1.21 (0-14.89)	13.81 (28.80)	0.66 (0-3.78)	34.41 (173.05)

[a]VPI: Video Power Index.

[b]Significantly different from capsule endoscopy-related videos; $P<.05$ using the Kruskal-Wallis test.

Video Content

Based on the guidelines and the videos included in the analysis, we categorized the content of the 3 types of endoscopy videos. We assessed the comprehensiveness of the included videos' content from the following five perspectives: (1) process: this refers to the process by which a patient undergoes gastrointestinal endoscopy; (2) principle: this refers to the principles of construction and operation of gastrointestinal endoscopes; (3) advantages and benefits: this refers to the roles of gastrointestinal endoscopy techniques, indications, and roles in relation to other investigations; (4) matters that need attention: this means that the patient needs to be aware of the following things before, during, and after the gastrointestinal endoscopic examination; and (5) complications and side effects: possible complications and adverse effects of endoscopic techniques in the digestive system. As shown by the radar chart (Figure 2B), among the 3 types of videos, the content coverage of gastroscopy-related videos was broadest, and the coverage of colonoscopy-related videos and capsule endoscopy-related videos was approximate; both of these videos were narrower than that of gastroscopy-related videos. Of the 167 endoscopy-associated videos analyzed, a substantial majority encompassed depictions of the process and enumerated potential advantages or benefits, accounting for 87 (52.1%) and 66 (39.5%) videos, respectively. In stark contrast, only 15 (9.0%) videos addressed possible complications and adverse reactions associated with gastrointestinal endoscopy.

Information Quality and Reliability

We evaluated the quality and reliability of the videos across different endoscope types and within groups of the same endoscope type (Table 2). On average, colonoscopy-related videos had an mDISCERN of 3.29 (SD 0.94), which was slightly greater than that of both gastroscopy (2.88, SD 0.99) and capsule endoscopy (2.93, SD 1.05), albeit not significantly (Figure 2C). In contrast, capsule endoscopy-related videos significantly underperformed, with a GQS of 1.88 (SD 0.83) when compared to gastroscopy (2.96, SD 0.97; $P=.03$) and colonoscopy (3.13, SD 0.79; $P<.01$; Figure 2D). However, the *JAMA* scores across all 3 endoscopy types remained similar (Figure 2E).

XSL•FO
RenderX

Table 2. mDISCERN, JAMA[a], and GQS[b] scores of the gastroscopy, colonoscopy, and capsule endoscopy.

Category	mDISCERN		*JAMA*		GQS	
	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)
Gastroscopy	3 (2-4)	2.88 (0.99)	2 (2-3)	2.29 (0.78)	3 (2-4)	2.96 (0.97)
Colonoscopy	4 (3-4)	3.29 (0.94)	2 (2-3)	2.29 (0.79)	3 (3-4)	3.13 (0.79)
Capsule endoscopy	3 (2-4)	2.93 (1.05)	2 (2-2)	2.07 (0.50)	2 (2-3)	2.53 (0.88)

[a]JAMA: Journal of the American Medical Association.

[b]GQS: Global Quality Scoring.

In terms of the different sources of the video uploaders (Table 3), we observed that the GQS scores of patient-provided gastroscopy-related videos were significantly lower than those of doctors or hospitals (1.88, SD 0.83 vs 3.05, SD 0.85; *P*=.04), medical media (1.88, SD 0.83 vs 3.27, SD 1.10; *P*<.01), and other professional organizations (1.88, SD 0.83 vs 3.45, SD 0.52; *P*=.02; Figure 3C). Consistent with the GQS results, the mDISCERN score (1.50, SD 0.53) and the *JAMA* score (1.13, SD 0.35) were also lower for patient-provided gastroscopy-related videos than for the other 3 videos (Figures 3A and B). Similarly, when analyzing colonoscopy-related videos, we found that the scores of patient-provided videos were lower than those of the other 3 videos in all 3 categories (Figures 3D-F). Interestingly, while patient-provided capsule endoscopy-related videos were scored lower in quality than videos provided by medical media (mDISCERN: 1.71, SD 0.76 vs 3.18, SD 0.64; *P*<.05) and other professional organizations (mDISCERN: 1.71, SD 0.76 vs 3.10, SD 1.25; *P*<.05), there was no significant difference in quality between patient-provided videos and videos provided by doctors or hospitals according to all the mDISCERN scores, *JAMA* scores and GQS scores (Figures 3G-I).

Table 3. mDISCERN, JAMA[a], and GQS[b] scores of videos from 4 groups of the uploaders.

Source	Videos, n	mDISCERN		*JAMA*		GQS	
		Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)
Gastroscopy							
Doctors or hospitals	19	3 (2-3)	2.74 (0.87)	3 (2-3)	2.58 (0.51)	3 (2-4)	3.05 (0.85)
Medical media	11	2 (2-3)	3.09 (0.70)	2 (2-3)	2.27 (0.79)	3 (3-4)	3.27 (1.10)
Other health institutions	18	4 (3-4)	3.50 (0.79)	3 (2-3)	2.50 (0.71)	3 (3-4)	3.45 (0.52)
Patients	8	2 (1-2)	1.50 (0.53)	1 (1-1)	1.13 (0.35)	2 (1-3)	1.88 (0.83)
Colonoscopy							
Doctors or hospitals	9	4 (4-4)	3.78 (0.44)	3 (2-3)	2.56 (0.53)	3 (3-4)	3.44 (0.73)
Medical media	24	4 (3-4)	3.46 (0.83)	3 (2-3)	2.58 (0.72)	3 (3-4)	3.21 (0.59)
Other health institutions	11	4 (3-4)	3.64 (0.67)	2 (2-3)	2.45 (0.52)	4 (3-4)	3.55 (0.52)
Patients	11	2 (1-3)	2.18 (0.87)	1 (1-2)	1.27 (0.47)	2 (2-3)	2.27 (0.90)
Capsule endoscopy							
Doctors or hospitals	11	3 (3-4)	3.00 (0.89)	2 (2-2)	2.09 (0.54)	3 (2-3)	2.73 (0.79)
Medical media	17	3 (3-4)	3.18 (0.64)	2 (2-3)	2.18 (0.64)	3 (2-3)	2.65 (0.93)
Other health institutions	20	4 (2-4)	3.10 (1.25)	2 (2-2)	2.10 (0.31)	3 (2-3)	2.55 (0.94)
Patients	7	2 (1-2)	1.71 (0.76)	2 (1-2)	1.71 (0.49)	2 (2-2)	1.86 (0.38)

[a]JAMA: Journal of the American Medical Association.

[b]GQS: Global Quality Scoring.

XSL•FO
RenderX

Figure 3. The reliability and quality of gastrointestinal endoscopy from different identities of the uploaders. (A) mDISCERN score, (B) JAMA score, and (C) GQS score of videos about gastroscopy from different identities of the uploaders. (D) mDISCERN score, (E) JAMA score, and (F) GQS score of videos about colonoscopy from different identities of the uploaders. (G) mDISCERN score, (H) JAMA score, and (I) GQS score of videos about capsule endoscopy from different identities of the uploaders. * denotes $P<.05$, and ** denotes $P<.01$ using the Kruskal-Wallis test. GQS: Global Quality Score; JAMA: Journal of the American Medical Association.

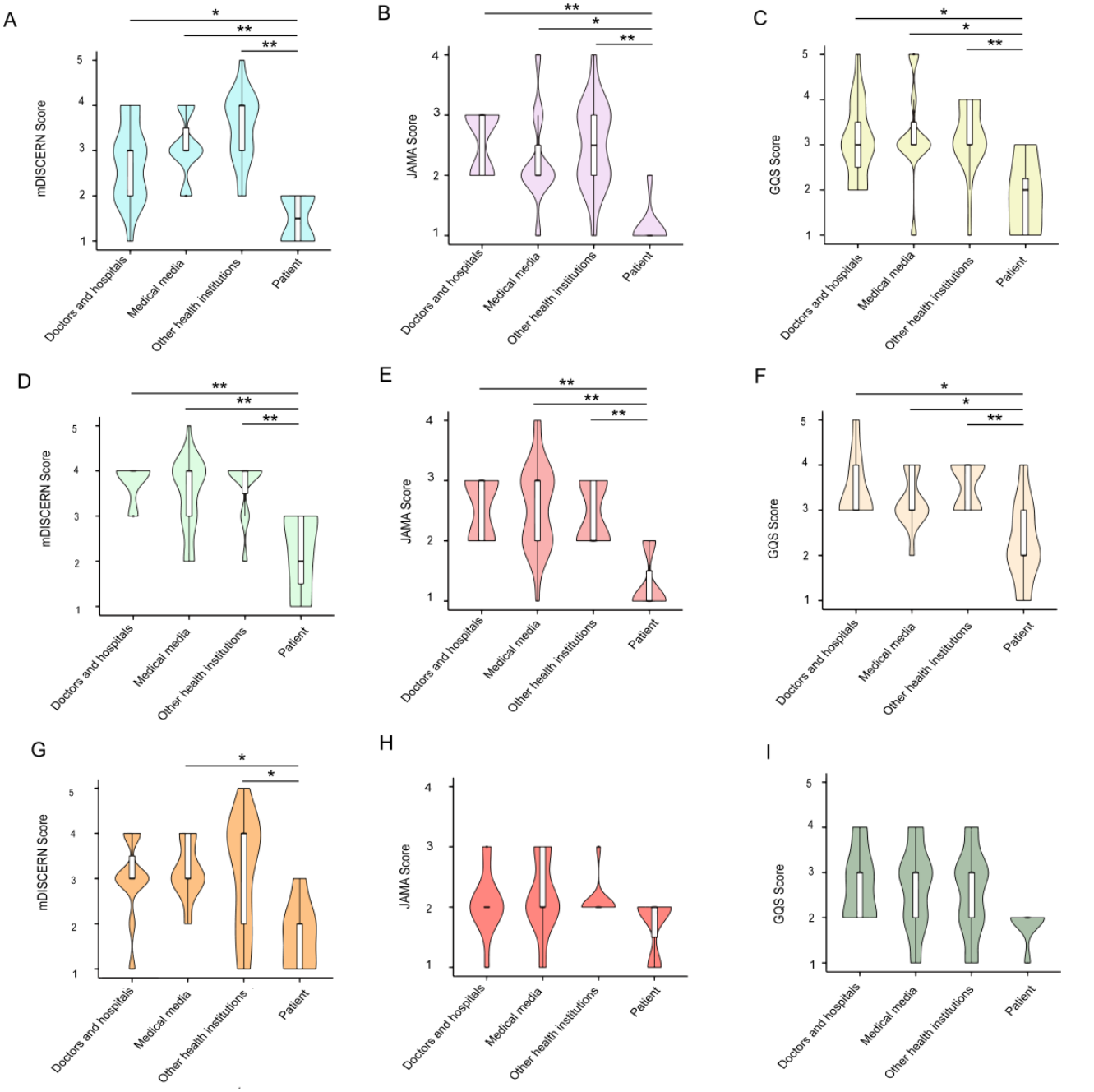

Conclusion and Evaluation

To assess the utility of the LLM-based video content summarization tool for summarizing health videos related to gastrointestinal endoscopy, we imported the 167 videos and collected the content summarized by the video content summarization tool. The summarized content was then evaluated in terms of accuracy, completeness, and readability (Figures 4A-C). LLM-based summaries yielded accuracy scores of 4 (4-5), completeness scores of 2 (1-2), and readability scores of 2 (1-2). None of the 3 scores were significantly different among the 3 gastrointestinal endoscopies.

XSL•FO
RenderX

Figure 4. AI-conclusion and evaluation of the videos. (A) Accuracy score, (B) completeness score, and (C) readability score among different types of endoscopes. AI: artificial intelligence.

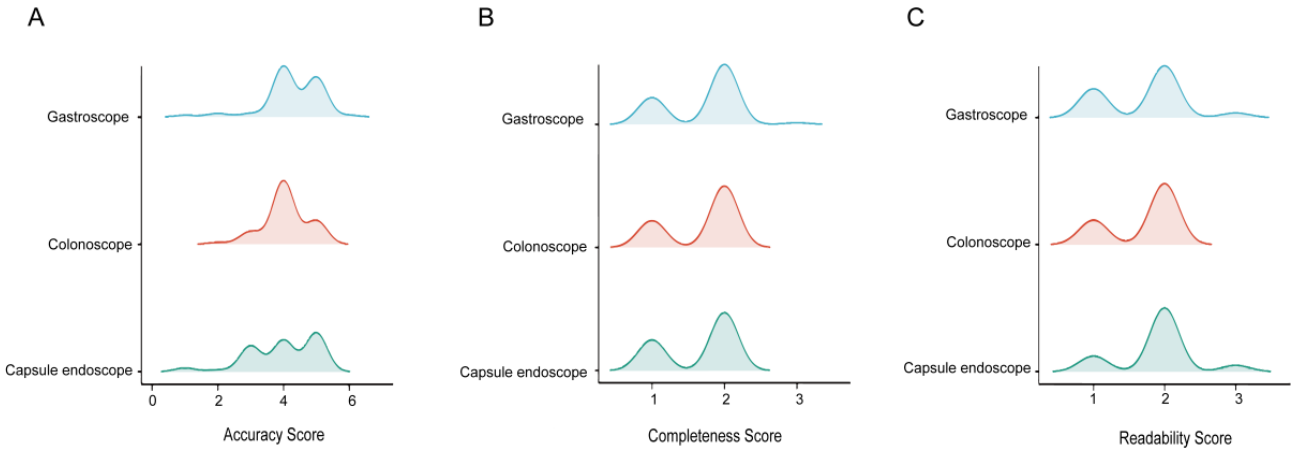

Discussion

Major Findings

In this study, we reviewed the relevant videos about gastroscopy, colonoscopy, and capsule endoscopy posted on YouTube before October 2023. First, we collected basic information such as the number of plays, posting time, number of likes, number of dislikes, and VPI of these videos and categorized and summarized the video content in 5 aspects. Second, we evaluated the quality, reliability, and degree to which the related videos were recommended. Finally, we explored the value of the LLM-based video summarization tool for health-related videos.

Previous research has indicated that mDISCERN scores and GQS scores less than 3 often indicate lower quality and lower levels of recommendation [19]. In our study, the average mDISCERN score was 2.88 for gastroscopy-related videos and 2.93 for capsule endoscopy-related videos. Only 38.9% (67/167) of the videos were assigned an mDISCERN score higher than 3, and for GQS scores, the proportion of videos with a score higher than 3 was even lower. Moreover, only 23.4% (39/167) of the videos were assigned a value above 3 when considering both criteria. Therefore, the overall quality and recommendability of these videos were insufficient. A surprising conclusion is that even professionally produced videos are not always suitable for viewer recommendations. This phenomenon may be attributed to 3 potential reasons: first, a small number of videos were too specialized and not applicable to patients; second, some videos were broad and lacked in-depth information; and third, certain videos were outdated, had lower production quality, and offered a poor viewing experience. In the present study, the mean *JAMA* score for both gastroscopy- and colonoscopy-related videos was 2.29, while for capsule endoscopy-related videos, it was only 2.07. These results suggest that the reliability of gastrointestinal endoscopy videos currently available on YouTube is also questionable. Specifically, we discovered that the majority of video producers did not include references or additional sources of information. Furthermore, some physicians featured in the videos were hesitant to disclose their affiliation unless they were promoting a specific institution. These results are similar to those of previous studies in that some funded physician users did not explicitly disclose conflicts of interest in their YouTube videos [33]. This lack of transparency compromises the videos' reliability and adds to viewer distress. Fortunately, although many videos lacked sufficient information, we did not find any videos that provided clearly incorrect information. When assessing content, our radar chart analysis revealed that only a small number of videos mentioned the potential complications and side effects of endoscopy in the digestive system.

A previous study revealed that the quality of health-related videos varies among authors with different backgrounds [34]. Specifically, in the context of gastrointestinal endoscopy videos, except for the *JAMA* scores of the capsule endoscopy group, videos created by individuals with a medical specialty background demonstrated higher levels of quality and reliability than those produced by nonmedical users. While there was no significant discrepancy in the *JAMA* scores for capsule endoscopy between medical professionals and patients, videos created by professionals still scored higher on average. This outcome can be attributed to 2 potential reasons. First, professionals possess greater familiarity with clinical guidelines and the latest research findings, whereas nonprofessionals are more inclined to share personal experiences and lessons learned. Second, medical professionals are skilled at presenting relevant information coherently and in an organized manner, while videos produced by laypeople tend to be more relaxed and casual, which, to some extent, affects patient quality [35].

Video Education Is an Important Part of Patient Education

Health education plays a pivotal role in patient care [36]. Providing high-quality health education to patients not only enhances patient-physician collaboration and facilitates smoother treatment processes but also contributes to a partial reduction in complications and alleviates the health care burden on society [37]. A previous study demonstrated that physician-led telephone education improved patient preparation prior to gastrointestinal endoscopy, thus enhancing the detection of lesions in the digestive system [38].

With the increasing popularity of short video sharing platforms, increasingly more health-related short videos are becoming known to the public [39]. Compared to traditional text-based information, videos are more interesting and intuitive [40].

XSL•FO
RenderX

High-quality videos play a crucial role in enhancing patients' understanding of the disease and increasing the likelihood of their acceptance of treatment plans [41]. Additionally, they effectively alleviate patient anxiety and enhance satisfaction levels [42]. Therefore, the role of video platforms will become increasingly important in patient education.

The Use of AI in Health-Related Information Dissemination Is Debatable

AI has played a significant role in identifying fake health-related information, as evidenced by studies on the detection of COVID-19 misinformation through machine learning algorithms [43,44]. However, there is a lack of research on AI applications for content summarization and information extraction in health-related videos. The video summarization tool used in this study is based on the LLM model, which summarizes video content using the text and subtitles present within the video [15]. This subtitle-based AI video content summarization tool not only offers a quick summary of the video content but also strives to maintain consistency with the original video. Nonetheless, this approach has several limitations. Upon comparing the text generated by the LLM-based video summarization tool with the actual video content, we observed instances where the tool missed key points and overly emphasized minor details. Moreover, providing a summary solely based on subtitles means that the LLM tool does not truly "watch" the video or refine its content, potentially leading to the omission of valuable information provided by the video creators beyond the subtitles, thereby reducing the comprehensiveness of the summary. The computer graphics and live-action scenes provided in videos allow viewers to create visual impressions, a dimension that the LLM tool lacks in the summarization process. Significantly, the present study indicates that for videos that lack subtitles and rely solely on image presentations, the LLM summarization tool tends to rely on the video title, resulting in discrepancies in accuracy and completeness when compared to the actual video content. Therefore, while summarizing health-related videos using existing LLM-based video summarization tools can enhance viewers' information absorption efficiency to some degree, the nature of this approach suggests that it cannot truly serve as a substitute for the experience of watching a video, which enables viewers to form a more intuitive understanding of diseases and treatment techniques.

Limitations and Future Directions

This study has several limitations. First, only English videos posted on the YouTube platform were analyzed, making it challenging to directly generalize the findings to other platforms or videos in different languages. Second, the video content summarization plugin used in this study was developed based on ChatGPT (version 3.5), and a plugin based on ChatGPT (version 4.0) may yield improved performance.

Many current gastrointestinal endoscopy videos lack details on equipment makers and resolutions. Though not crucial for patient preparation, such information can help viewers better understand the equipment used during procedures. We suggest future videos include this data consistently. We recognize the absence of mandatory policies governing health education videos [40]. As video-sharing platforms continue to gain popularity, it becomes crucial to implement a certain level of censorship and regulation for videos disseminated on the internet. Government agencies and video-sharing platforms should establish a body composed of health care professionals dedicated to screening and monitoring health-related videos uploaded web-based and promptly identifying and flagging videos containing controversial information. Given the intricacy of medical content, we recommend that individuals with professional backgrounds in health care exercise caution when using specialized vocabulary when producing videos and strike a balance between popularity and the scientific nature of educational videos. This approach ensures that a broader audience can comprehend the information presented in the videos. Additionally, labeling the sources of arguments and information in the video will enhance its reliability and further improve viewers' acceptance of the content.

Conclusions

This study evaluated the information quality of videos discussing gastroscopy, colonoscopy, and capsule endoscopy on the YouTube platform. Overall, the videos exhibited an average relatively low quality. Notably, videos created by professionals had a higher level of reliability and quality than did those produced by nonprofessionals. We look forward to the availability of a greater number of high-quality gastroenterology-related videos in the future, as they greatly contribute to providing patients with valuable health-related information. Additionally, we urge the government to swiftly implement policies that curtail the dissemination of low-quality health-related videos.

Acknowledgments

This work is supported by funds from the Young Medical Talents Training Program of Shanghai Pudong New Area Health Commission (PWRq2024-24 to JL), the Pudong Health Committee of Shanghai (NO.2025-PWYC-04 to YS), the Medical Discipline Construction Project of Pudong Health Committee of Shanghai (no. PWYgf2021-08 to YS), the Foundation of Gongli Hospital (no. 2022GPY-A06 to JL), and the Deep Blue Talent Project of Naval Medical University to WX.

Data Availability

The data that support the finding of this study are openly available.

XSL•FO
RenderX

Authors' Contributions

JL involved in concept development, article writing, and data analysis; data collection and data analysis were done by YQ, YL, and WN; PS involved in concept presentation and article revision; ZY involved in concept development and article revision; FW involved in data collection; WX and YS involved in concept presentation and article revision.

Conflicts of Interest

None declared.

References

1. Axon ATR. Fifty years of digestive endoscopy: successes, setbacks, solutions and the future. Dig Endosc 2020;32(3):290-297. [doi: 10.1111/den.13593] [Medline: 31794063]
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
3. Hu CH. Papulosquamous eruptions: manifestation and management. Compr Ther 1988;14:42-46. [doi: 10.1016/b978-0-7020-4449-6.00021-7]
4. Pan J, Xin L, Ma Y, Hu LH, Li ZS. Colonoscopy reduces colorectal cancer incidence and mortality in patients with non-malignant findings: a meta-analysis. Am J Gastroenterol 2016;111(3):355-365 [FREE Full text] [doi: 10.1038/ajg.2015.418] [Medline: 26753884]
5. Liu Q, Zeng X, Wang W, Huang R, Huang YJ, Liu S, et al. Awareness of risk factors and warning symptoms and attitude towards gastric cancer screening among the general public in China: a cross-sectional study. BMJ Open 2019;9(7):e029638 [FREE Full text] [doi: 10.1136/bmjopen-2019-029638] [Medline: 31340970]
6. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
7. Spada C, Hassan C, Galmiche J, Neuhaus H, Dumonceau J, Adler S, European Society of Gastrointestinal Endoscopy. Colon capsule endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy 2012;44(5):527-536 [FREE Full text] [doi: 10.1055/s-0031-1291717] [Medline: 22389230]
8. Velayos B, Fernández L, Aller R, de la Calle F, del Olmo L, Arranz T, et al. Public opinion survey after capsule endoscopy: patient's point of view on its utility. Rev Esp Enferm Dig 2006;98(6):436-448 [FREE Full text] [doi: 10.4321/s1130-01082006000600005] [Medline: 16948543]
9. Barnes SS, Kaul V, Kudchadkar SR. Social media engagement and the critical care medicine community. J Intensive Care Med 2019;34(3):175-182. [doi: 10.1177/0885066618769599] [Medline: 29699469]
10. Gonsalves CA, McGannon KR, Schinke RJ. Social media for health promotion: what messages are women receiving about cardiovascular disease risk by the Canadian heart and stroke foundation? J Health Psychol 2020;25(8):1017-1029. [doi: 10.1177/1359105317743802] [Medline: 29226735]
11. Ayers JW, Goodman AM, Smith DM. #MedEd: medical education and knowledge translation on social media. J Am Med Assoc 2023;330(10):909-910 [FREE Full text] [doi: 10.1001/jama.2023.12465] [Medline: 37642959]
12. Spiegel B. 2015 American Journal of Gastroenterology lecture: how digital health will transform gastroenterology. Am J Gastroenterol 2016;111(5):624-630. [doi: 10.1038/ajg.2016.68] [Medline: 27045930]
13. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016;14(3):429-435.e3. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
14. Lee TC, Staller K, Botoman V, Pathipati MP, Varma S, Kuo B. ChatGPT answers common patient questions about colonoscopy. Gastroenterology 2023;165(2):509-511.e7 [FREE Full text] [doi: 10.1053/j.gastro.2023.04.033] [Medline: 37150470]
15. Glasp. 2023. URL: https://glasp.co/youtube-summary [accessed 2025-01-09]
16. Statista. 2023. URL: https://www.statista.com/topics/2019/youtube/#topicOverview [accessed 2025-01-09]
17. Ahmad T, Sattar K, Akram A. Medical professionalism videos on youTube: content exploration and appraisal of user engagement. Saudi J Biol Sci 2020;27(9):2287-2292 [FREE Full text] [doi: 10.1016/j.sjbs.2020.06.007] [Medline: 32884409]
18. Kite J, Chan L, MacKay K, Corbett L, Reyes-Marcelino G, Nguyen B, et al. A model of social media effects in public health communication campaigns: systematic review. J Med Internet Res 2023;25:e46345 [FREE Full text] [doi: 10.2196/46345] [Medline: 37450325]
19. Libretti A, Vitale SG, Saponara S, Corsini C, Aquino CI, Savasta F, et al. Hysteroscopy in the new media: quality and reliability analysis of hysteroscopy procedures on YouTube™. Arch Gynecol Obstet 2023;308(5):1515-1524 [FREE Full text] [doi: 10.1007/s00404-023-07172-9] [Medline: 37568070]
20. Farag M, Bolton D, Lawrentschuk N. Use of youTube as a resource for surgical education-clarity or confusion. Eur Urol Focus 2020;6(3):445-449. [doi: 10.1016/j.euf.2019.09.017] [Medline: 31606471]

XSL•FO
RenderX

21. Kunze KN. Editorial commentary: youTube videos provide poor-quality medical information: don't believe what you watch!. Arthroscopy 2020;36(12):3048-3049. [doi: 10.1016/j.arthro.2020.07.042] [Medline: 33276891]
22. Solomon P. Looking for information—a survey of research on information seeking, needs, and behavior. Inform Retrieval 2003;6:284-288 [FREE Full text] [doi: 10.1108/s1876-0562(2012)002012b020]
23. Albayrak E, Büyükçavuş MH. Does youTube offer high-quality ınformation? Evaluation of patient experience videos after orthognathic surgery. Angle Orthod 2023;93(4):409-416 [FREE Full text] [doi: 10.2319/042522-315.1] [Medline: 36853060]
24. Aglamis SO, Senel S, Koudonas A. Quality analysis of youTube videos on vulvodynia. Sex Med 2023;11(1):qfac013 [FREE Full text] [doi: 10.1093/sexmed/qfac013] [Medline: 37007847]
25. Aldahlawi SA, Homeida L. YouTube as a source of patient information on oral manifestations of COVID-19: a cross-sectional evaluation of its utility, dependability, and content. Cureus 2023;15(8):e42885 [FREE Full text] [doi: 10.7759/cureus.42885] [Medline: 37664281]
26. Cano-de-la-Cuerda R, Marcos-Antón S, Blázquez-Fernández A, Fernández-Cañas M, Sánchez-Herrera Baeza P, Fernández-González P, et al. Content, educational value and quality analysis of videos about neurorehabilitation in people with multiple sclerosis on YouTube®. Rev Neurol 2024;79(3):77-88 [FREE Full text] [doi: 10.33588/rn.7903.2024091] [Medline: 39007859]
27. Acero N, Herrero E, Foncham J, McIlvaine J, Kayaalp E, Figueora M, et al. Accuracy, quality, and misinformation of YouTube abortion procedural videos: cross-sectional study. J Med Internet Res 2024;26:e50099 [FREE Full text] [doi: 10.2196/50099] [Medline: 39437380]
28. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
29. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
30. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor--let the reader and viewer beware. J Am Med Assoc 1997;277(15):1244-1245. [doi: 10.1001/jama.1997.03540390074039]
31. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
32. Erdem MN, Karaca S. Evaluating the accuracy and quality of the information in kyphosis videos shared on YouTube. Spine (Phila Pa 1976) 2018;43(22):E1334-E1339. [doi: 10.1097/BRS.0000000000002691] [Medline: 29664816]
33. Walradt T, Bilal M, Wadhwa V, Chiang AL, Berzin TM. Confidentiality and conflicts of interest: an assessment of twitter posts in gastrointestinal endoscopy. Am J Gastroenterol 2021;116(7):1542-1544. [doi: 10.14309/ajg.0000000000001238] [Medline: 33767102]
34. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. J Med Internet Res 2021;23(9):e30409 [FREE Full text] [doi: 10.2196/30409] [Medline: 34468327]
35. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. J Am Med Assoc 2001;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
36. Serper M, Gawron AJ, Smith SG, Pandit AA, Dahlke AR, Bojarski EA, et al. Patient factors that affect quality of colonoscopy preparation. Clin Gastroenterol Hepatol 2014;12(3):451-457 [FREE Full text] [doi: 10.1016/j.cgh.2013.07.036] [Medline: 23954647]
37. Lai Y, He Z, Liu Y, Yin X, Fan X, Rao Z, et al. The quality and reliability of TikTok videos on non-alcoholic fatty liver disease: a propensity score matching analysis. Front Public Health 2023;11:1231240 [FREE Full text] [doi: 10.3389/fpubh.2023.1231240] [Medline: 37860803]
38. Liu X, Luo H, Zhang L, Leung FW, Liu Z, Wang X, et al. Telephone-based re-education on the day before colonoscopy improves the quality of bowel preparation and the polyp detection rate: a prospective, colonoscopist-blinded, randomised, controlled study. Gut 2014;63(1):125-130. [doi: 10.1136/gutjnl-2012-304292] [Medline: 23503044]
39. Zheng S, Tong X, Wan D, Hu C, Hu Q, Ke Q. Quality and reliability of liver cancer-related short Chinese videos on TikTok and Bilibili: cross-sectional content analysis study. J Med Internet Res 2023;25:e47210 [FREE Full text] [doi: 10.2196/47210] [Medline: 37405825]
40. He Z, Wang Z, Song Y, Liu Y, Kang L, Fang X, et al. The reliability and quality of short videos as a source of dietary guidance for inflammatory bowel disease: cross-sectional study. J Med Internet Res 2023;25:e41518 [FREE Full text] [doi: 10.2196/41518] [Medline: 36757757]
41. Trinh N, Novice K, Lekakh O, Means A, Tung R. Use of a brief educational video administered by a portable video device to improve skin cancer knowledge in the outpatient transplant population. Dermatol Surg 2014;40(11):1233-1239. [doi: 10.1097/DSS.0000000000000148] [Medline: 25310749]

XSL•FO
RenderX

42. Eley VA, Searles T, Donovan K, Walters E. Effect of an anaesthesia information video on preoperative maternal anxiety and postoperative satisfaction in elective caesarean section: a prospective randomised trial. Anaesth Intensive Care 2013;41(6):774-781. [doi: 10.1177/0310057X1304100613] [Medline: 24180719]
43. Zivkovic M, Stoean C, Petrovic A, Bacanin N, Strumberger I, Zivkovic T. A novel method for COVID-19 pandemic information fake news detection based on the arithmetic optimization algorithm. : IEEE; 2021 Presented at: 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC); 2021 December 7-10; Timisoara, Romania. [doi: 10.1109/synasc54541.2021.00051]
44. Khan Y, Thakur S. Fake news detection of South African COVID-19 related tweets using machine learning. : IEEE; 2022 Presented at: 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD); 2022 August 04-05; Durban, South Africa. [doi: 10.1109/icabcd54961.2022.9856272]

Abbreviations

AI: artificial intelligence
GQS: Global Quality Score
JAMA: Journal of the American Medical Association
LLM: large language model
VPI: Video Power Index

Edited by T de Azevedo Cardoso; submitted 26.03.24; peer-reviewed by F Liao, H Chan; comments to author 09.09.24; revised version received 27.11.24; accepted 20.12.24; published 11.03.25.

Please cite as:
Liu J, Qiu Y, Liu Y, Xu W, Ning W, Shi P, Yuan Z, Wang F, Shi Y
The Reliability and Quality of Videos as Guidance for Gastrointestinal Endoscopy: Cross-Sectional Study
J Med Internet Res 2025;27:e58855
URL: https://www.jmir.org/2025/1/e58855
doi:10.2196/58855
PMID:

©Jinpei Liu, Yifan Qiu, Yilong Liu, Wenping Xu, Weichen Ning, Peimei Shi, Zongli Yuan, Fang Wang, Yihai Shi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mechanism Assessment of Physician Discourse Strategies and Patient Consultation Behaviors on Online Health Platforms: Mixed Methods Study

Menglei Kong[1*], PhD; Yu Wang[2*], PhD; Meixuan Li[1], PhD; Zhong Yao[1], Prof Dr

[1]School of Economics and Management, Beihang University, Beijing, China

[2]School of Public Health, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zhong Yao, Prof Dr
School of Economics and Management
Beihang University
No. 37 Xueyuan Road
Haidian District
Beijing, 100083
China
Phone: 86 13811013409
Email: 05723@buaa.edu.cn

Abstract

Background: Online health platforms are currently experiencing significant growth. Patients can conveniently seek medical consulting services on such platforms. Against the backdrop of the thriving development of digital health care, the patterns of physician-patient communication are undergoing profound changes. It is imperative to focus on physician discourse strategies during online physician-patient interactions, which will improve the efficiency of physician-patient communication and achieve better management of the physician-patient relationship.

Objective: This study aims to explore the influencing mechanism between physician discourse strategies and patient consultation behavior on online health platforms. Additionally, we explore the crucial mediating role of online physician-patient trust and the moderating role of shared decision-making in the online physician-patient communication process.

Methods: We used a mixed research approach to explore the influencing mechanism. Data on physician basic attributes and physician-patient communication text records were collected from the Chunyu Doctor website using a web spider. The study obtained a total of 8628 interaction texts from January 2022 to July 2023. Physician discourse strategies (capacity-oriented strategy, quality-oriented strategy, and goodwill-oriented strategy), online physician-patient trust, and shared decision-making were captured through text mining and a random forest model. First, we employed text mining to extract the speech acts, modal resources, and special linguistic resources of each record. Then, using a well-trained random forest model, we captured the specific discourse strategy of each interaction text based on the learned features and patterns. The study generated 863 groups of physician samples with 17 data fields. The hypotheses were tested using an "ordinary least squares" model, and a stability test was conducted by replacing the dependent variable.

Results: The capacity-oriented strategy, goodwill-oriented strategy, and quality-oriented strategy had significant effects on patient consultation behavior (β=.151, *P*=.007; β=.154, *P*<.001; and β=.17, *P*<.001, respectively). It should be noted that the anticipated strong effect of the capacity-oriented strategy on patient consultation behavior was not observed. Instead, the effects of the quality-oriented strategy and goodwill-oriented strategy were more prominent. Physician notification adequacy from shared decision-making moderated the effect between the goodwill-oriented strategy and patient consultation behavior (β=.172; *P*<.001). Additionally, patient expression adequacy from shared decision-making moderated the effect between the capacity-oriented strategy and patient consultation behavior (β=.124; *P*<.001), and between the goodwill-oriented strategy and patient consultation behavior (β=.104; *P*=.003). Online physician-patient trust played a significant mediating role between physician discourse strategies and patient consultation behavior.

XSL•FO
RenderX

Conclusions: The study findings suggest significant implications for stimulating patient consultation behavior on online health platforms by providing guidance on effective discourse strategies for physicians, thus constructing a trustworthy physician image, improving the physician-patient relationship, and increasing platform traffic.

(J Med Internet Res 2025;27:e54516) doi:10.2196/54516

KEYWORDS

online health consultation; physician discourse strategies; online physician-patient trust; shared decision-making; patient consultation behavior

Introduction

Background

Digital health service resources in China have experienced accelerated expansion, with over 26.7 million people receiving online medical services in 2022 [1]. Notably, online health platforms, such as "Haodf" [2], "Chunyu Doctor" [3], and "JD Health" [4], in China have emerged to offer online medical services, such as health condition assessment and treatment planning. Patients primarily use online health platforms to communicate with suitable physicians about their health conditions. This behavior is commonly called "patient consultation behavior."

Compared with offline consultation, the emergence of online health platforms has helped foster a 2-way, interactive, and democratic environment for both physicians and patients [5], but new problems have emerged. The patient consultation behavior during the online physician-patient communication process is invisible, and patients often exhibit higher levels of anxiety and eagerness to receive valuable health suggestions [6]. Moreover, it is impossible to ensure a constant online presence for both physicians and patients. These aspects may result in the quality of online consultation services for patients being compromised on the platform.

Ancient Greek scholar Hippocrates, the father of Western medicine, said, "There are two elements in the world effective of curing disease: medicine and language." In other words, the language used by physicians plays a crucial role in enhancing the treatment effectiveness and ensuring patients' well-being. At the beginning of online health consultation, patients select the physician who can fulfill their health care demands in terms of their disease, health condition, and preferred communication style. The physician's title and history of peer-patient consultation are important references for potential patients to make consulting decisions. The texts entered by physicians in the chat box often encompass the diagnosis of patients' conditions, their concerns and empathy, and the recommended treatment plans. It is crucial for physicians to communicate with patients smoothly and trustworthily by optimizing their discourse strategies, which is beneficial for stimulating patient consultation behavior. Research on physician discourse strategies in online medical consultation can provide insights into the relationship management between physicians and patients [7]. Some scholars analyzed the construction of the physician image from the perspective of language application by using qualitative research methods. It has been proven that physicians can help patients comprehend their health conditions by using reasonable, logical, and well-intended language, enabling shared decision-making and fostering rapport between both parties [8-10]. However, few studies have specifically explored the impact mechanism of physician discourse strategies on patient consultation behavior from the perspective of pragmatics, with the use of quantitative analysis.

Unlike the interaction process on other platforms, online health platforms require attentive and thorough services at every stage. Maintaining a strong and trustworthy relationship between physicians and patients greatly benefits the health management quality of patients. The "virtualized," "fragmented," and "casual" communication nature during online health consultation can exacerbate tensions in the physician-patient relationship [11]. These platforms and physicians face the critical stage of establishing trust. Thus, it becomes essential to consider patient consultation behavior from the perspective of trust. In addition, shared decision-making is a personalized diagnosis and communication method that was introduced by Veatch [12] in 1972. Veatch argued that shared decision-making would greatly impact the quality of physician-patient communication and medical care. Physician-patient shared decision-making is considered crucial for strengthening trust, improving communication, and enhancing the effectiveness of medical diagnosis [13]. Consequently, shared decision-making could be a significant factor affecting the relationship between discourse strategies and patients' online consultation behavior.

Our research aims to contribute to the existing literature on patient consultation behavior by addressing 2 research objectives. First, we aim to investigate the correlation between physician discourse strategies and patient consultation behavior. Second, we seek to determine the significance of online physician-patient trust and shared decision-making in relation to the interplay between physician discourse strategies and patient consultation behavior. Based on these research objectives, we propose the following three research questions:

1. What is the relationship between physician discourse strategies and patient consultation behavior?
2. How do physician discourse strategies influence patient consultation behavior by influencing the construction of online physician-patient trust?
3. What role does shared decision-making play between physician discourse strategies and patient consultation behavior?

XSL•FO
RenderX

Research Model and Hypotheses

Physician Discourse Strategies and Patient Consultation Behavior

In 1984, Frankel [14] defined physician-patient communication as the "process of exchanging information between a doctor and a patient, which is influenced by various psychological factors and their interactions." Multiple studies have focused on the direct effects of communication style on communication outcomes [15]. We also identified that physicians often use linguistic and discursive tools to convey medical suggestions and express empathy, which can profoundly influence patients' decision-making during consultations. We refer to these communication strategies as "physician discourse strategies."

The SOR (stimulus, organism, and response) model was proposed by Mehrabian and Russell [16] in 1974. This psychological theory explains that environmental factors act as stimuli (S) that first influence an individual's psychological state (O), which in turn affects their behavioral responses (R) [17,18]. In the context of online health platforms, patients encounter a wealth of external information (eg, titles, service volume, prices, etc) before and during interactions, which can stimulate their engagements, including patient consultation behavior [5,18-20]. Aligning with the SOR theory, we examined the discourse strategy employed by physicians as external stimuli received by patients, exploring their potential impact on patient consultation behavior.

Among the essential elements of these texts, language shapes not only the content of the message (ie, what is said?) but also the style (ie, how is it said?) [21]. Most studies addressed how physicians' communication skills during consultation interactions (eg, response speed [22], interaction depth [23], and interaction rounds [24]) can shape patients' perceptions and influence their behavioral outcomes. Unskillful responses to patients' inquiries can not only place extra burden on physicians but also discourage patients from seeking consultation services online [21]. Additionally, scholars found that a physician with high-quality content responses [25-27], proper emotional expressions [26], the right response time [22,25], and in-depth interactions [25,28], can support a patient's willingness to consult on online health platforms. These findings indicate that the appropriate use of discourse strategies is crucial and necessary, especially in the online interaction environment. However, these studies about physician discourse strategies are still in their infancy. Existing research has primarily focused on the superficial characteristics during physician-patient communication, neglecting the complexity of the discourse strategies employed and the foundational structures that inform them. In the study by Wu et al [27], discourse strategies were constructed according to the response time, detailed style, and emotional comfort, with the aim of investigating their effects on patient consultation satisfaction. We have incorporated insights from this study into our own research, but we delve deeper into the application of linguistic resources within physician discourse strategies. Based on the SOR theory and existing research, our research first focuses on the impact of physician discourse strategies on patient consultation behavior. We propose the following hypothesis: different types of physician discourse strategies have varying positive effects on patient consultation behavior (H1).

Trust is crucial for effective physician-patient communication, and online health platforms must prioritize building patients' trust, which helps potential patients transition from curious observers to customers who are willing to use the platform for health consultation [29]. The trust theory proposed by Mayer et al [30] in 1995 identifies trust in 3 main dimensions: capability, quality, and goodwill. In line with the trust theory by Mayer et al [30], we categorized discourse strategies into the following: capacity-oriented strategy, quality-oriented strategy, and goodwill-oriented strategy.

The capacity-oriented strategy indicates that physicians highlight their professional skills and experiences, and show personal authority in the communication process [31]. Demonstrating such professional competencies can foster patient consultation behavior [32]. The goodwill-oriented strategy reflects that physicians are patient-centered, fulfill the needs of patients, patiently guide patients to overcome their negative emotions, and comfort their fragile hearts during communication. Empathetic communication that is easily understood by the patient can help manage expectations, ensure follow-up and treatment adherence, and improve the perception of the entire health care experience [33,34]. Such strategies may be conducive to fostering a willingness among patients to use the platform for health consultations. The quality-oriented strategy emphasizes that physicians focus on their professional code and show sincerity, honesty, and reliability in the communication process. By conveying such a communication style, physicians can foster a sense of professional quality in their interactions, thereby increasing patient trust and willingness to engage [35,36].

In combination with H1, we propose the following hypotheses: the capacity-oriented strategy enhances patient consultation behavior (H1a); the goodwill-oriented strategy enhances patient consultation behavior (H1b); and the quality-oriented strategy enhances patient consultation behavior (H1c).

Physician-Patient Trust

Trust is a complex dynamic that exists among individuals, organizations, and events, involving uncertainties and expectations regarding the future actions of one party by another [37,38]. Most physicians and patients communicate and interact by sending texts or audio in the chat box. Moreover, given the constraints of the online environment and the offline working hours of physicians, it is impossible to ensure a constant online presence for both physicians and patients. Patients need to share their health conditions with physicians, even those involving some private information. The establishment of trust may determine whether the patient consults and honestly shares necessary information with the doctor [39]. Thus, trust could be even more critical in the online health care environment due to individuals' higher levels of information sensitivity and the existence of uncertainty and a lack of regulation [11].

Online trust is generated on an internet platform, where the trustor possesses attributes that benefit the trustee [40]. Online trust (eg, trust in websites, online news, and social networking

XSL•FO
RenderX

site providers) has been studied extensively [41-44]. Notably, some studies have revealed that online trust strongly influences online purchase behavior [45]. In the health care sector, the influencing factors of trust are commonly studied in online health platforms, but there is less discussion on the influence of online trust on patient behavior. The establishment of trust between physicians and patients on online health platforms is often influenced by the physician's title, gender, and reputation [46-48]. A physician's service quality, empathy expression, average word count, and cumulative number of sessions, and the physician-patient session ratio significantly impact physician-patient trust [49,50]. Patient self-disclosure of personal information during consultations leads to increased satisfaction and trust [51,52]. The attributes, quantity, length, sentiment orientation, and composition of negative feedback in physician-patient interactions similarly influence the formation of trust between physicians and patients.

Therefore, in line with the SOR model, we conducted supplementary research on the mediating role of online physician-patient trust in the interaction between physician discourse strategies and patient consultation behavior. Hence, we propose the following hypotheses regarding the mediating role of online trust: online physician-patient trust mediates between physician discourse strategies and patient consultation behavior (H2); online physician-patient trust mediates between the capacity-oriented strategy and patient consultation behavior (H2a); online physician-patient trust mediates between the goodwill-oriented strategy and patient consultation behavior (H2b); and online physician-patient trust mediates between the quality-oriented strategy and patient consultation behavior (H2c).

Shared Decision-Making

The concept of shared decision-making was initially introduced by Veatch [12] in 1972 in the article "A Model for Medical Ethics in a Revolutionary Era." Later, in 1997, Charles et al [53] further elaborated on the meaning of shared decision-making and proposed the following four crucial elements that it should encompass: (1) the participation of 2 parties; (2) the exchange of information between both parties; (3) the achievement of consensus on the treatment plan; and (4) the implementation of specific actions by both parties. Since then, shared decision-making has gained significant attention in the medical field, primarily due to the increasing emphasis on patients' right to access information and their demand for satisfactory medical care [53]. Shared decision-making promotes the transformation of the patient's role from a passive recipient of medical care to an active participant and supervisor, fully embracing the concept of "patient centeredness."

Shared decision-making is a common topic in physician-patient communication and relationship management, which has been widely accepted and applied in the medical field. Some studies have proven the significance of shared decision-making in the quality and process of physician-patient communication by using qualitative methods [54,55]. Veatch [12] argued that shared decision-making would greatly impact the quality of physician-patient communication and medical care. Both patients and physicians tend to overestimate the potential benefits of interventions, and shared decision-making has been proven to be useful in influencing expectations and beliefs [56,57]. Some scholars also developed shared decision-making assessment scales [58-60] and explored the factors influencing participation in shared decision-making to gain better behavioral outcomes [61,62].

Shared decision-making could be a significant factor affecting the relationship between discourse strategies and patient online consultation behavior. However, the research issue of how shared decision-making affects the shaping process of patient behavior has not been solved, especially in the online health care environment. Under the shared decision-making philosophy, 2-way communication helps to prevent the distortion or loss of information that can result from 1-way communication, ensuring that decisions are more aligned with the best interests of the patient. Through shared decision-making, patients can gain a more comprehensive understanding of their condition and treatment options, thereby enhancing their adherence to the prescribed treatment regimen [58] and encouraging high-quality communication and patient well-being. Shared decision-making may indirectly moderate the impact of physician discourse strategies on patient consultation behavior by enhancing patient engagement, regulating physician discourse strategies, and improving medical communication.

In this study, our research objective was to clarify how shared decision-making moderates the influence of physician discourse strategies on patient consultation behavior in online health platforms, and thus, the following hypotheses are proposed: shared decision-making moderates the effect of physician discourse strategies on patient consultation behavior (H3); shared decision-making moderates the effect of the capacity-oriented strategy on patient consultation behavior (H3a); shared decision-making moderates the effect of the goodwill-oriented strategy on patient consultation behavior (H3b); and shared decision-making moderates the effect of the quality-oriented strategy on patient consultation behavior (H3c).

Figure 1 shows our research model. The research explores the relationship between physician discourse strategies and patient consultation behavior on online health platforms (H1 and H1a-H1c), the possible mediating role of online physician-patient trust (H2 and H2a-H2c), and the moderating role of shared decision-making (H3 and H3a-H3c) in the above process.

XSL•FO
RenderX

Figure 1. Research model. H1: different types of physician discourse strategies have varying positive effects on patient consultation behavior; H2: online physician-patient trust mediates between physician discourse strategies and patient consultation behavior; H3: shared decision-making moderates the effect of physician discourse strategies on patient consultation behavior.

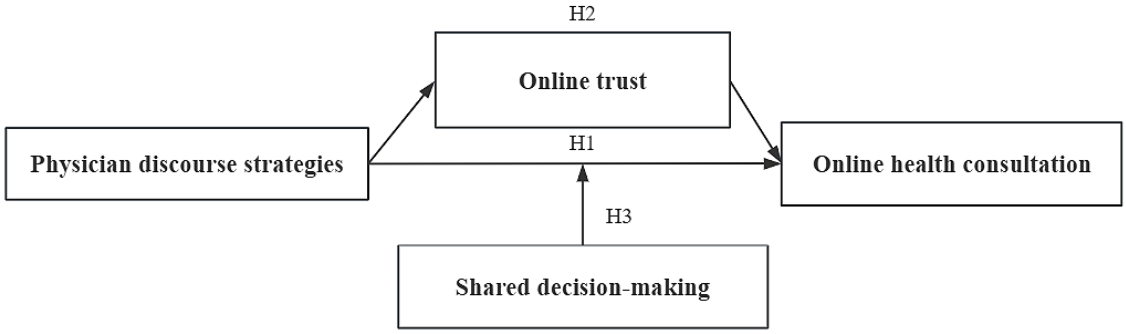

Methods

Data Collection

Registered physicians on the Chunyu Doctor website are the research subjects of this paper. Chunyu Doctor, a platform with over 660,000 physicians from governmental hospitals, offers medical services to over 140,000,000 patients.

Our research explores the relationship between physician discourse strategies and patient consultation behavior while also contemplating the influences of online physician-patient trust and shared decision-making. We collected data in 2 main categories. The first category involved the basic attributes of physicians, such as gender and title. The second category involved the texts extracted from online physician-patient communication records. These anonymized patient consultation records are publicly accessible and serve as typical cases for our research [63]. We used a web spider to crawl and collect physician-patient interaction data from the "Positive Q&A" module on the Chunyu Doctor website from January 2022 to July 2023. These interaction data include information, such as consultation time, topic summary, and specific consultation text records. In total, we collected the basic attribute data of 863 physicians and 10 communication text records displayed by each doctor (8628 groups of physician-patient communication texts in total). These data cover a wide range of medical specialties, including 15 first-level departments and 43 second-level departments, with 201 groups of data per second-level department. Taking the pediatrics department as an example, we collected the attribute data of 40 physicians in the neonatology and pediatrics departments and 400 interactive text records. These texts were very complete, recording the whole physician-patient communication process.

Ethical Considerations

This study has been reviewed and approved by the Institutional Review Board of the School of Economics and Management, Beihang University (BE-202306082). All research procedures comply with the ethical guidelines and standards set by the Institutional Review Board. The data employed in this study were deidentified interaction texts from patients, ensuring that all personal information had been anonymized in accordance with ethical guidelines. Some private data, such as the patient's name, private disease data, and image examination records, were hidden.

Variable Measurements

The main variables examined in this study were discourse strategies, patient consultation behavior, online physician-patient trust, and shared decision-making. The description and measurement of each variable are presented in Table 1.

For the measurement of these variables, we employed text mining to analyze the physician-patient interaction texts that were scraped and applied random forest to predict the specific discourse strategy used by physicians. The data processing for all variables is illustrated in Figure 2, which will be elaborated upon in detail below. Given that the focus of this research is the physician, the data must be matched and integrated individually according to the physician ID. The integration method (summation or averaging) was employed to depict different variables regarding the performance of each physician. Ultimately, a total of 8628 data groups were integrated on a one-to-one correspondence basis with physicians, resulting in a sample dataset of 863 physicians. Each dataset included 17 data fields.

XSL•FO
RenderX

Table 1. Variables and their measurements.

Variables and dimensions	Measurements
Dependent variables	
Patient consultation behavior	
Patient consultation volume	It represents the number of physicians (*i*) offering online consultation service in (*t*) time.
Patient satisfaction	It represents patients rating their physicians based on their consultations.
Independent variables	
Gender	
Male	1
Female	2
Title	
Resident physician	1
Attending physician	2
Associate chief physician	3
Chief physician	4
Department	
First-level department	It is divided into 15 first-level departments, including obstetrics, pediatrics, internal medicine, surgery, etc.
Analyzing variables	
Discourse strategies	
Capacity-oriented strategy[a]	 Physician (*i*) highlights professional skills and experiences, and shows personal authority in (*t*) time.
Goodwill-oriented strategy[b]	 Physician (*i*) is patient-centered, fulfills the needs of patients, patiently guides patients to overcome their negative emotions, and comforts their fragile hearts in the communication process in (*t*) time.
Quality-oriented strategy[c]	 Physician (*i*) emphasizes that physicians focus on their professional code and show sincerity, honesty, and reliability in (*t*) time.
Shared decision-making	
Physician notification adequacy	Physicians should be fully informed about alternative treatment options, including their advantages and disadvantages. Take the value of 1 if physicians exhibit the above characteristics; otherwise, take the value of 0.
Patient expression adequacy	Patients should openly communicate their views, concerns, values, and preferences to the physician. Take the value of 1 if patients exhibit the above characteristics; otherwise, take the value of 0.
Trust	
Online physician-patient trust	The term frequency of words, such as “trustworthy,” “enthusiastic,” “conscientious,” and “effective,” is used to indicate the patient’s recognition of the treatment plan provided by the physician.

[a]The abbreviations for the equation are as follows: COS: capability-oriented strategy; ASA: assertive speech acts; ISA: indicative speech acts; AMR: affirmative modal resources; MT: medical terminology.

[b]The abbreviations for the equation are as follows: GOS: goodwill-oriented strategy; ESA: expressive speech acts; FD: first-person deixis; H: honorifics; E: emojis.

[c]The abbreviations for the equation are as follows: QOS: quality-oriented strategy; RSA: refusal speech acts; CSA: committed speech acts; VMR: vague modal resources; IS: interrogative sentences.

XSL•FO
RenderX

Figure 2. Data processing.

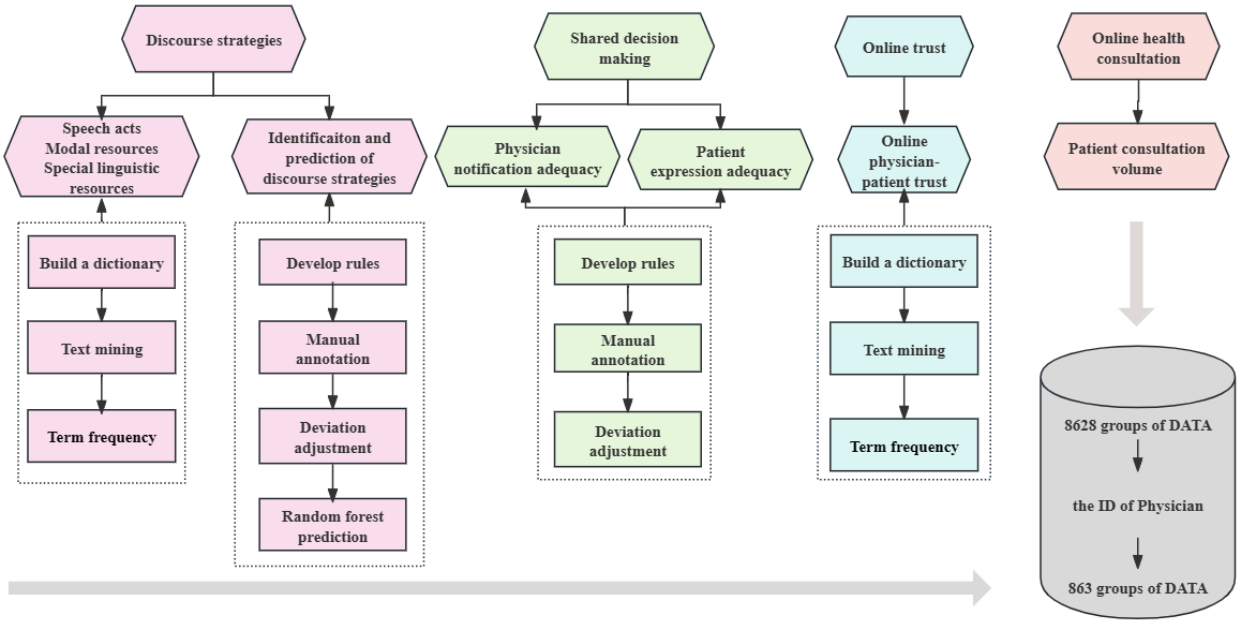

Discourse Strategies

We classified discourse strategies into 3 categories: capacity-oriented strategy, quality-oriented strategy, and goodwill-oriented strategy, which were derived from the trust theory by Mayer et al [30]. To describe discourse strategies by the linguistic characteristics of speech acts, modal resources, and special linguistic resources, we have provided details in Table 2.

In describing discourse strategies, Searle and Vanderveken [64] proposed the speech act theory and categorized speech acts into 5 types based on grammatical indicators, including assertive speech acts, indicative speech acts, refusal speech acts, expressive speech acts, and committed speech acts [64,65]. The use of modal resources can also impact physician-patient communication quality [26]. First-person deixis, emojis, honorifics, and other special linguistic resources often appear in the consultation texts. The inclusion of first-person deixis facilitates the psychological convergence between physicians and patients, effectively shortening the psychological distance between the 2 parties [66]. Additionally, the appropriate use of medical terminology demonstrates the physician's professional proficiency, fostering trust among patients [67,68]. Therefore, this paper proposes that discourse strategies used by physicians are communication strategies that appear during physician-patient interactions, regulating the way physicians exchange information and convey their emotions through the use of speech acts, modal resources, and special linguistic resources.

We calculated term frequencies corresponding to the variables according to dictionaries, which were divided into 2 categories. The first involved summarizing frequently occurring words in the physician-patient interaction texts and using the Chinese word segmentation dictionary to establish the speech act dictionary, modal resource dictionary, and special linguistic resources dictionary. The second was used to calculate the term frequency of medical terminology. We used a Chinese medical vocabulary dictionary with approximately 18,000 words, covering various dimensions, such as symptoms, drugs, medical tests, and diseases.

We then identified the discourse strategies of each text. To achieve this, we used a combination of artificial labeling and random forest for discourse strategy identification. Initially, we formulated the mathematical and semantic features of these discourse strategies. Subsequently, we analyzed the mean value of the data fields grouped by physician ID. To ensure robustness, we enlisted the help of 5 experts, either working in the medical field or studying medical-related majors. They grouped all interaction texts according to first-level departments and randomly selected 100 groups of texts for strategy identification. Each of the 5 experts independently labeled the sampled data for the 15 groups of first-level departments. To gauge the consistency of strategy identification, the identification results were cross-checked against 5 copies of each text, and the identification rules were revised accordingly. The iterative process was repeated until the labeling differences were negligible. Overall, the final average consistency for each first-level department approached over 95%.

XSL•FO
RenderX

Table 2. Linguistic characteristics index system.

Linguistic characteristics and index	Description
Speech acts	
Assertive speech acts	Physicians judge the patient's condition, medication, diagnosis plan, etc.
Indicative speech acts	Physicians indicate the patient to do something.
Refusal speech acts	Physicians express denial and disapproval of something.
Committed speech acts	Physicians promise to do something.
Expressive speech acts	Physicians express empathy and reassurance for the patient's condition, experience, etc.
Modal resources	
Affirmative modal resources	Physicians use words with a strongly positive attitude.
Vague modal resources	Physicians use words with an ambiguous attitude.
Special linguistic resources	
Interrogative sentences	Term frequency of the words used by physicians to express a questioning tone.
First-person deixis	Term frequency of first-person indicative words used by physicians such as "I" and "we."
Honorifics	Term frequency of the words used by physicians to express respect for patients.
Emojis	Frequency of emojis used by physicians.
Medical terminology	Term frequency of specialized words used by physicians such as disease, medication, symptoms, and diagnosis.

Finally, a random forest algorithm was used to predict the remaining unlabeled samples. Overall, 20% of the data items with discourse strategy labels were used as test sets and 80% were used as learning sets. The algorithm had good classification performance and met the classification requirements, with a test accuracy of 85.1% and *F*-score of 0.915. The data items with blank discourse strategies were predicted and filled. Using the well-trained random forest model, we predicted the unlabeled samples in the 8628 interaction texts. The model automatically filled in the corresponding discourse strategy labels for each unlabeled interaction text based on the learned features and patterns. This process not only improved the efficiency of data annotation but also provided a more comprehensive data foundation for subsequent analysis.

Additionally, we have presented examples of physician-patient interaction texts that exemplify each of the 3 main discourse strategies to enhance readers' comprehension of their specific meanings in Multimedia Appendix 1.

Online Physician-Patient Trust

We established an online trust expression word dictionary and calculated the level of online trust between physicians and patients by analyzing the frequency of relevant expression words in patient texts. Words, such as "trustworthy," "enthusiastic," "conscientious," and "effective" were used to indicate the patient's recognition of the treatment plan provided by the physician.

Patient Consultation Behavior

Most of the research on patient consultation often uses patient conversion rates, service volume, and patient satisfaction as feedback [25,69,70]. To measure patient consultation behavior, we used physician consultation volume on the Chunyu Doctor website, which is part of the basic physician attribute data.

Shared Decision-Making

The concept of shared decision-making highlights 2 important aspects. First, physicians should be fully informed about alternative treatment options, including their advantages and disadvantages. Second, patients should openly communicate their views, concerns, values, and preferences to physicians. Therefore, 2 primary indicators of shared decision-making were designed to encompass the above meanings: physician notification adequacy and patient expression adequacy.

The measurement of shared decision-making was mainly conducted through the formulation of label rule development, manual labeling, and bias adjustment. Initially, we proposed the key labeling rules of shared decision-making based on a literature search, medical knowledge, and text comprehension. Subsequently, the texts were organized and grouped within each first-level department. Five experts, either working in the medical field or studying medical-related majors, were selected to assess the 2 indicators of shared decision-making for each text in each department.

Modeling and Statistical Analysis

Ordinary Least Squares Model

Before testing the main effects, moderating effects, and mediating effects, we conducted a general descriptive analysis to describe the basic attributes of physicians and the textual information on physician-patient interactions. Subsequently, a bivariate correlation model was used to test the correlations between patient consultation behavior and other variables.

We then proceeded with the test of main effects, which primarily explored the influence of discourse strategies (capacity-oriented strategy, quality-oriented strategy, and goodwill-oriented strategy) on patient consultation behavior. Additionally, we conducted the test of moderating effects, which explored the influence of shared decision-making on patient consultation

XSL•FO
RenderX

behavior, and the test of mediating effects, which examined the mediating role of online physician-patient trust between discourse strategies and patient consultation behavior. The empirical models pertaining to these tests were as follows:

Main effects model: $Patient\ consultation\ behavior_i = \beta_0 + \beta_1 COS_i + \beta_2 GOS_i + \beta_3 QOS_i + \beta_4 Control + \varepsilon$

Mediating effects model: $Ln\ Patient\ consultation\ behavior_i = \alpha_0 + Ln\ Discourse\ stategies_i$; $Ln\ w = \beta_0 + \beta_1 Ln\ Discourse\ stategies_i + \varepsilon$; $Ln\ tPatient\ consultation\ behavior_i = \beta_0 + \beta_1 Ln\ Discourse\ stategies_i + \beta_2 Ln\ Online\ physician\text{-}patient\ trust_i + \varepsilon$

Moderating effects model: $Patient\ consultation\ behavior_i = \beta_0 + \beta_1 COS_i + \beta_2 GOS_i + \beta_3 QOS + \beta_4 COS_i \times SDM_i + \beta_5 GOS_i \times SDM_i + \beta_6 QOS_i \times SDM_i + \beta_7 Control + \varepsilon$

Control variables include the gender, title, and department of the physician. ε represents the error term. Previous research has shown that physician gender can impact patient interactions, communication styles, and potentially patient satisfaction [46,71,72]. The title of a physician can influence patient perceptions and expectations [72]. Patients may respond differently based on the physician's level of experience or expertise, which could affect their consultation behavior. In addition, physicians from different departments may have varying approaches to communication and may encounter different patient demographics or types of consultations [73].

Next, we analyzed the data using the ordinary least squares regression model to test for main effects, moderating effects, and mediating effects. Stata (StataCorp) was used for the analysis, and all variables were standardized in the regression.

Results

Descriptive Statistics

Our dataset encompassed 863 physicians providing online health consultation services to 77,248 patients. The online health consultations specifically pertain to 15 first-level departments, including internal medicine, surgery, obstetrics, pediatrics, oncology, and prevention and treatment. Among the 863 physicians, there were 637 male physicians and 226 female physicians from 15 departments. The plastic and cosmetic department had the fewest physicians (n=18), while the surgery department had the most physicians (n=197). Attending physicians had the highest count (n=411), while chief physicians had the lowest count (n=83). The 863 physicians had an average of 7847.81 total patient consultations and a mean patient satisfaction rate of 98.53%. Among the 3 types of discourse strategies, the largest number of physicians (n=347) employed the capacity-oriented strategy. Moreover, 411 physicians on the online health platform were primary care physicians. The descriptive statistics are shown in Table 3.

XSL•FO
RenderX

Table 3. Descriptive statistics.

Variable	Value (N=863)
Patient consultation volume, mean (SD)	7847.81 (12,252.30)
Online physician-patient trust, mean (SD)	4.92 (4.30)
Shared decision-making, mean (SD)	
Physician notification adequacy	3.10 (2.72)
Patient expression adequacy	6.70 (3.71)
Physician discourse strategies, mean (SD)	
Capacity-oriented strategy	42.06 (23.94)
Quality-oriented strategy	8.52 (4.92)
Goodwill-oriented strategy	6.28 (4.92)
Gender, n (%)	
Male	637 (73.8)
Female	226 (26.2)
Title, n (%)	
Resident physician	136 (15.8)
Attending physician	411 (47.6)
Associate chief physician	233 (27.0)
Chief physician	83 (9.6)
Department, n (%)	
Obstetrics department	20 (2.3)
Pediatrics department	40 (4.6)
Otolaryngology department	60 (7.0)
Gynecology department	20 (2.3)
Orthopedics department	60 (7.0)
Oral and maxillofacial department	20 (2.3)
Andrology department	20 (2.3)
Internal medicine department	180 (20.9)
Dermatology and sexually transmitted disease department	40 (4.6)
Surgery department	197 (22.8)
Ophthalmology department	20 (2.3)
Nutrition department	20 (2.3)
Plastic and cosmetic department	18 (2.1)
Traditional Chinese medicine department	89 (10.3)
Oncology department	59 (6.8)

Pearson Correlation Analysis and Collinearity Testing

As shown in Table 4, patient consultation volume had the highest positive correlation with the goodwill-oriented strategy (r=0.261; P=.008). Patient consultation satisfaction showed positive correlations with the capacity-oriented strategy (r=0.282; P=.008), quality-oriented strategy (r=0.271; P=.003), goodwill-oriented strategy (r=0.122; P=.005), and physician notification adequacy (r=0.187; P=.003).

When there is a high correlation between 2 independent variables, the existence of one variable may cause the regression coefficient of the other variable to rise sharply. Collinearity becomes a problem in regression analysis, but this does not mean that this variable has no predictive effect on the dependent variable. Because the correlation coefficient between the quality-oriented strategy and capacity-oriented strategy was high (more than 0.7), there might have been a multiple collinearity issue. Thus, this study used variance inflation factor (VIF) testing to examine the collinearity between the 2 variables.

XSL•FO
RenderX

Table 4. Correlations between variables.

Variables	COS[a]	QOS[b]	GOS[c]	PNA[d]	PEA[e]	OPT[f]	PCV[g]	PCS[h]
COS								
r	1	0.765[i]	0.400[i]	0.358	0.065	0.536[i]	0.220[i]	0.282[i]
P value	—[j]	.009	.005	.003	.56	.005	.003	.008
QOS								
r	0.765[i]	1	0.391[i]	0.310[i]	0.032	0.509[i]	0.132[i]	0.271[i]
P value	.009	—	.003	.005	.66	.003	.009	.003
GOS								
r	0.400[i]	0.391[i]	1	0.237[i]	–0.005	0.488[i]	0.261[i]	0.122[i]
P value	.005	.003	—	.005	.58	.003	.008	.005
PNA								
r	0.358	0.310[i]	0.237[i]	1	0.423[i]	0.243[i]	0.243[i]	0.187[i]
P value	.003	.005	.005	—	.005	.005	.008	.003
PEA								
r	0.065	0.032	–0.005	0.423[i]	1	0.036	0.087[k]	0.126[i]
P value	.56	.66	.58	.005	—	.08	.02	.004
OPT								
r	0.536[i]	0.509[i]	0.488[i]	0.243[i]	0.036	1	0.250[i]	0.182[i]
P value	.005	.003	.003	.005	.08	—	.003	.006
PCV								
r	0.220[i]	0.132[i]	0.261[i]	0.243[i]	0.087[k]	0.250[i]	1	0.036
P value	.003	.009	.008	.008	.02	.003	—	.18
PCS								
r	0.282[i]	0.271[i]	0.122[i]	0.187[i]	0.126[i]	0.182[i]	0.036	1
P value	.008	.003	.005	.003	.004	.006	.18	—

[a]COS: capacity-oriented strategy.

[b]QOS: quality-oriented strategy.

[c]GOS: goodwill-oriented strategy.

[d]PNA: physician notification adequacy.

[e]PEA: patient expression adequacy.

[f]OPT: online physician-patient trust.

[g]PCV: patient consultation volume.

[h]PCS: patient consultation satisfaction.

[i]The correlation is significant at a significance level of .001 (2-tailed).

[j]Not applicable.

[k]The correlation is significant at a significance level of .01 (2-tailed).

In the regression model, the VIF can measure the severity of multicollinearity. It is generally believed that if the VIF value is greater than 10, there is a multicollinearity problem (strictly greater than 5) [74]. We conducted collinearity analysis on physician discourse strategies, the capacity-oriented strategy, the quality-oriented strategy, the goodwill-oriented strategy, physician notification adequacy, patient expression adequacy, online physician-patient trust, and patient consultation satisfaction, and found that the VIF values of these variables were 1.398, 4.768, 4.803, 1.667, 1.245, 1.510, 1.096, and 2.083, respectively. All VIF values of the variables were lower than 5, indicating that there was no serious collinearity issue among the independent variables in this study.

Hypothesis Testing

Our research focused on the Chunyu Doctor platform, a leading online health market in China with a significant number of physician and patient registrations. We investigated the impact

XSL•FO
RenderX

of physician discourse strategies on patient consultation behavior in this online health platform. To validate our research models, we first examined the relationship between physician discourse strategies (capacity-oriented strategy, quality-oriented strategy, and goodwill-oriented strategy) and patient consultation behavior. Then, we introduced moderating variables, namely physician notification adequacy and patient expression adequacy, into the model. This allowed us to test our main effects hypotheses (H1, H1a-H1c) as well as the moderating effect of shared decision-making between physician discourse strategies and patient consultation behavior (H3, H3a-H3c). The regression results are summarized in Table 5.

The capacity-oriented strategy, goodwill-oriented strategy, and quality-oriented strategy had significant effects on patient consultation behavior (β=.151, *P*=.003; β=.154, *P*<.001; and β=.170, *P*<.001, respectively). It should be noted that the anticipated strong effect of the capacity-oriented strategy on patient consultation behavior was not observed. Instead, the effects of the quality-oriented strategy and goodwill-oriented strategy were more prominent.

We then conducted a test on the moderating effects. The interaction term between the capability-oriented strategy and physician notification adequacy did not show significance (β=.062; *P*=.62), while the interaction term between the capability-oriented strategy and patient expression adequacy showed significance (β=.124; *P*<.001). This suggests that there is a significant difference in the effect of the capability-oriented strategy on patient consultation behavior at different levels of the moderating variable (patient expression adequacy).

Similarly, the interaction term between the goodwill-oriented strategy and physician notification adequacy was significant (β=.172; *P*<.001), indicating that the effect of the goodwill-oriented strategy on patient consultation behavior varies significantly at different levels of the moderating variable (physician notification adequacy).

In addition, the interaction term between the goodwill-oriented strategy and patient expression adequacy showed significance (β=.104; *P*=.003), suggesting that the effect of the goodwill-oriented strategy on patient consultation behavior differs significantly at different levels of the moderating variable (patient expression adequacy).

The interaction term between the quality-oriented strategy and physician notification adequacy did not show significance (β=.013; *P*=.23). Similarly, the interaction term between the quality-oriented strategy and patient expression adequacy did not show significance (β=.033; *P*=.42). This implies that the effect of the quality-oriented strategy on patient consultation remains consistent at different levels of the moderating variable.

Table 5. Impact of shared decision-making on patient consultation behavior.

Variable	Model 1 (β)	Model 2 (β)	Model 3 (β)	Model 4 (β)	Model 5 (β)
Gender	–0.003	–0.010	–0.003	0.006	–0.009
Title	0.154[a]	0.149[a]	0.143[a]	0.168	0.140[a]
Department	–0.074[b]	–0.050[b]	–0.076[b]	–0.109[a]	–0.066[b]
Capacity-oriented strategy	0.151[a]	0.159[b]	—[c]	—	0.282[a]
Goodwill-oriented strategy	0.154[d]	—	0.160[d]		0.373[a]
Quality-oriented strategy	0.170[d]	—		0.127[a]	0.222[b]
Capacity-oriented strategy × physician notification adequacy	—	0.062	—	—	0.155
Capacity-oriented strategy × patient expression adequacy	—	0.124[d]	—	—	0.780[a]
Goodwill-oriented strategy × physician notification adequacy	—	—	0.172[d]	—	0.704[a]
Goodwill-oriented strategy × patient expression adequacy	—	—	0.104[a]	—	0.065[a]
Quality-oriented strategy × physician notification adequacy	—	—	—	0.013	0.021
Quality-oriented strategy × patient expression adequacy	—	—	—	0.033	0.070
Constant	99.041[d]	99.148[a]	98.958[d]	98.813[d]	91.863[a]
F	18.204[d]	17.214[d]	24.593[d]	10.750[d]	16.907[d]
R^2	0.152	0.108	0.153	0.064	0.193

[a]The correlation is significant at a significance level of .01 (2-tailed).

[b]The correlation is significant at a significance level of .05 (2-tailed).

[c]Not applicable.

[d]The correlation is significant at a significance level of .001 (2-tailed).

Figure 3 depicts the moderating effects of patient expression adequacy on the interplay of the capability-oriented strategy and patient consultation behavior. Figures 4 and 5 describe the moderating roles of physician notification adequacy with the

XSL•FO
RenderX

goodwill-oriented strategy and patient expression adequacy with the goodwill-oriented strategy, respectively. The moderating variables were categorized into 3 levels: mean level, high level (mean + SD), and low level (mean – SD).

Figure 3. Moderating role of physician notification adequacy with the capability-oriented strategy.

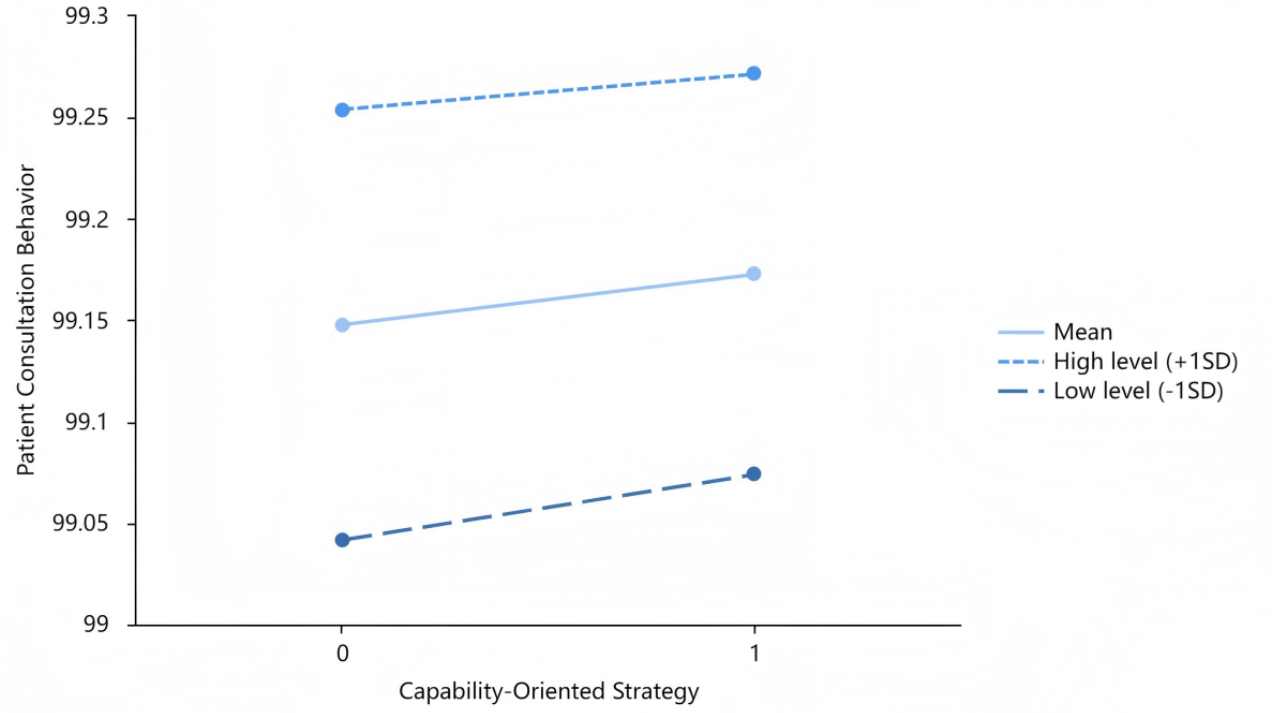

Figure 4. Moderating role of physician notification adequacy with the goodwill-oriented strategy.

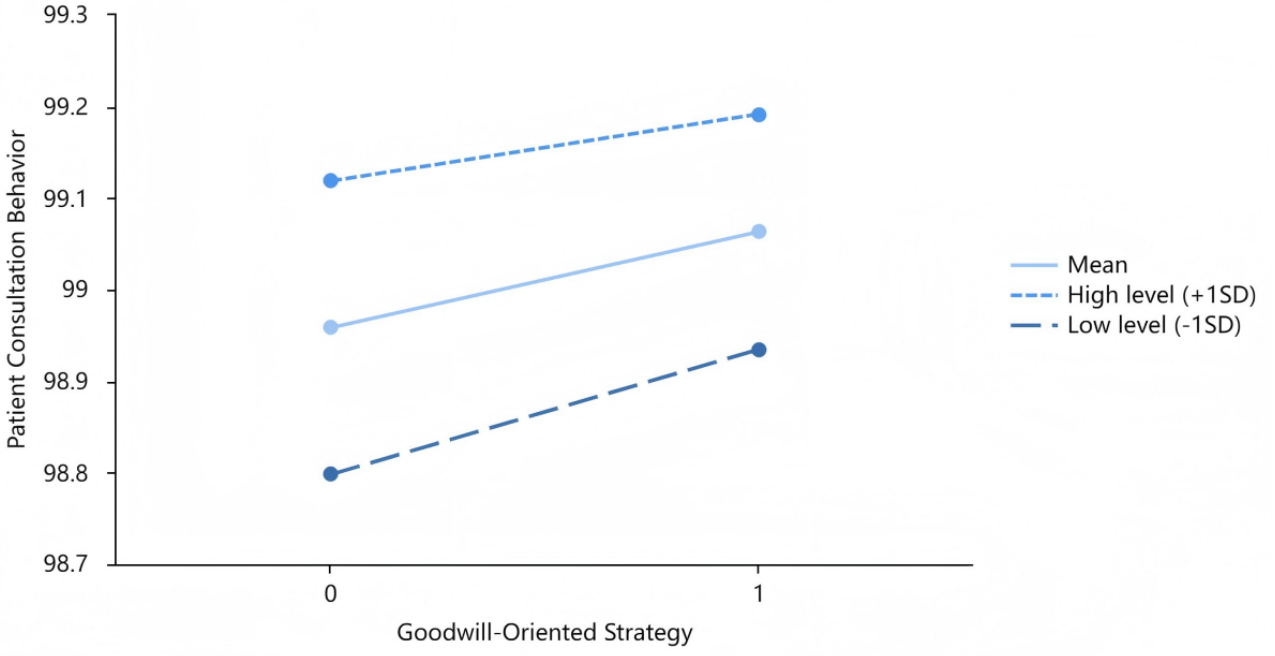

XSL•FO
RenderX

Figure 5. Moderating role of patient expression adequacy with the goodwill-oriented strategy.

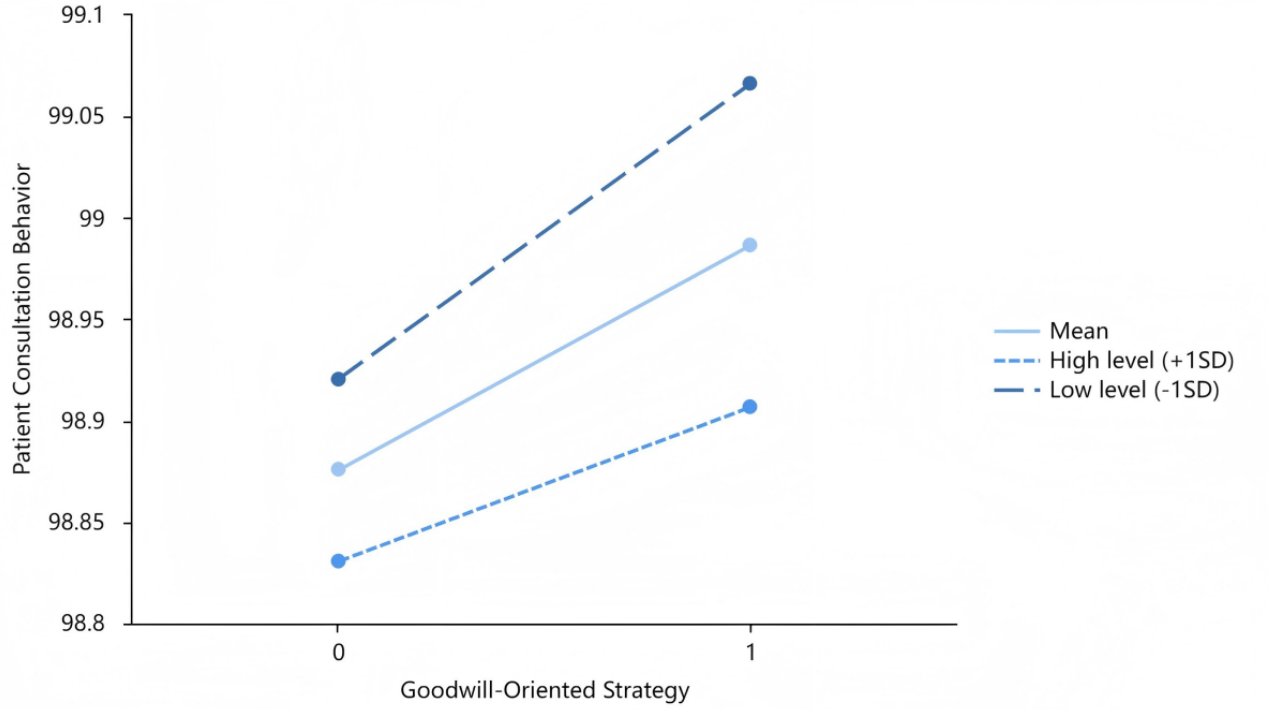

Next, our study examined the mediating role of online physician-patient trust using the nonparametric percentile bootstrap method to test the mediating effects, as shown in Table 6. Physician discourse strategies significantly affected patient consultation behavior (β=.292; *P*<.001). The capacity-oriented strategy significantly affected patient consultation behavior (β=0.260; *P*<.001). The quality-oriented strategy significantly affected patient consultation behavior (β=.223; *P*<.001). The goodwill-oriented strategy significantly affected patient consultation behavior (β=.075; *P*<.001). It involved repeating the sampling 2000 times and estimating 95% CIs. If the CI does not include zero, it indicates a significant mediating effect [75]. In our analysis, online physician-patient trust (95% CI 0.724-0.743) was found to mediate the relationship between physician discourse strategies and patient consultation behavior, with a mediation effect value of 0.348. The mediating effect of online physician-patient trust on the relationship between the capacity-oriented strategy and patient consultation behavior showed a value of 0.031. Additionally, the mediating effect of online physician-patient trust on the relationship between the quality-oriented strategy and patient consultation behavior showed a value of 0.042. Moreover, the mediating effect of online physician-patient trust on the relationship between the goodwill-oriented strategy and patient consultation behavior showed a value of 0.095. Likewise, the goodness-of-fit of the mediating effect model between online physician-patient trust in the 3 dimensions of physician discourse strategies and patient consultation behavior reached the standard (χ^2/df <5; comparative fit index [CFI] >0.9; Tucker-Lewis index [TLI] >0.9; root mean square error of approximation [RMSEA] <0.08; standardized root mean square residual [SRMR] <0.5).

XSL•FO
RenderX

Table 6. Standardized effects of model paths.

Model paths	Estimates	Boot SE	Boot LLCI[a]	Boot ULCI[b]
Direct effect				
Physician discourse strategies→patient consultation behavior	0.292	0.033	0.181	0.214
Capacity-oriented strategy→patient consultation behavior	0.260	0.028	0.017	0.031
Quality-oriented strategy→patient consultation behavior	0.223	0.029	0.165	0.280
Goodwill-oriented strategy→patient consultation behavior	0.075	0.028	0.017	0.129
Mediation effect				
Physician discourse strategies→online physician-patient trust→patient consultation behavior	0.348	0.008	0.724	0.743
Capacity-oriented strategy→online physician-patient trust→patient consultation behavior	0.031	0.205	0.002	0.314
Quality-oriented strategy→online physician-patient trust→patient consultation behavior	0.042	0.018	0.004	0.076
Goodwill-oriented strategy→online physician-patient trust→patient consultation behavior	0.095	0.017	0.061	0.129
Total effect				
Physician discourse strategies→patient consultation behavior	0.640	0.034	0.737	0.753
Capacity-oriented strategy→patient consultation behavior	0.291	0.247	0.021	0.337
Quality-oriented strategy→patient consultation behavior	0.265	0.024	0.219	0.312
Goodwill-oriented strategy→patient consultation behavior	0.170	0.023	0.123	0.215

[a]LLCI: lower limit confidence interval.

[b]ULCI: upper limit confidence interval.

Robustness Testing

This section primarily aims to validate the stability of the derived model and establish the validity of the research conclusions. In this study, patient consultation volume was considered as a variable representing patient consultation behavior. However, patients on online health platforms can rate the physicians they interact with. Thus, patient consultation satisfaction is considered as an alternative variable for patient consultation behavior [25,69,70], which replaced patient consultation volume and was included in the regression analysis model. The robustness testing results are presented in Table 7, reconfirming the earlier findings and indicating the relatively stable nature of this study's results.

The capacity-oriented strategy, goodwill-oriented strategy, and quality-oriented strategy had significant effects on patient consultation behavior (β=.169, *P*=.002; β=.274, *P*<.001; and β=.357, *P*<.001, respectively). Physician notification adequacy from shared decision-making moderated the effect between the goodwill-oriented strategy and patient consultation behavior (β=.153; *P*<.001). Patient expression adequacy from shared decision-making moderated the effect between the capacity-oriented strategy and patient consultation behavior (β=.121; *P*<.001) and between the goodwill-oriented strategy and patient consultation behavior (β=.106; *P*=.03).

Furthermore, we employed the nonparametric percentile bootstrap method to test the significance of the mediating effects. The results are shown in Table 8. This involved repeating the sampling 2000 times and estimating 95% CIs. If the CI does not include zero, it indicates a significant mediating effect. In our analysis, online physician-patient trust (95% CI 0.002-0.061) was found to mediate the relationship between physician discourse strategies and patient consultation behavior, with a mediation effect value of 0.014. The mediating effect of online physician-patient trust on the relationship between the capacity-oriented strategy and patient consultation behavior showed a value of 0.034. Additionally, the mediating effect of online physician-patient trust on the relationship between the quality-oriented strategy and patient consultation behavior showed a value of 0.038. Moreover, the mediating effect of online physician-patient trust on the relationship between the goodwill-oriented strategy and patient consultation behavior showed a value of 0.077. Likewise, the goodness-of-fit of the mediating effect model between online physician-patient trust in the 3 dimensions of physician discourse strategies and patient consultation behavior reached the standard (χ^2/df <5; CFI >0.9; TLI >0.9; RMSEA <0.08; SRMR <0.5).

XSL•FO
RenderX

Table 7. Impact of shared decision-making on online health consultation (patient consultation satisfaction).

Variable	Model 1 (β)	Model 2 (β)	Model 3 (β)	Model 4 (β)	Model 5 (β)
Gender	–0.003	–0.078	–0.072	0.006	–0.036
Title	0.154[a]	0.030	0.045	0.168	0.160[a]
Department	–0.074[b]	–0.112	–0.111	–0.109[a]	0.011
Capacity-oriented strategy	0.169[a]	0.193[c]	—[d]	—	0.208[a]
Goodwill-oriented strategy	0.274[c]	—	0.270	—	0.185[a]
Quality-oriented strategy	0.357[c]	—	—	0.227[a]	0.117[a]
Capacity-oriented strategy × physician notification adequacy	—	0.079	—	—	0.062
Capacity-oriented strategy × patient expression adequacy	—	0.121[c]	—	—	0.106[a]
Goodwill-oriented strategy × physician notification adequacy	—	—	0.153[c]	—	0.193[a]
Goodwill-oriented strategy × patient expression adequacy	—	—	0.106[b]	—	0.097[b]
Quality-oriented strategy × physician notification adequacy	—	—	—	0.013	0.018
Quality-oriented strategy × patient expression adequacy	—	—	—	0.133	0.037
Constant	195.120[c]	286.242[a]	284.934[c]	275.443[a]	263.321[c]
F	15.03[c]	19.450[c]	18.225[c]	7.838[c]	8.844[c]
R^2	0.152	0.114	0.153	0.045	0.119

[a]The correlation is significant at a significance level of .01 (2-tailed).

[b]The correlation is significant at a significance level of .05 (2-tailed).

[c]The correlation is significant at a significance level of .001 (2-tailed).

[d]Not applicable.

Table 8. Standardized effects of model paths (patient consultation satisfaction).

Model paths	Estimates	Boot SE	Boot LL-CI[a]	Boot UL-CI[b]
Direct effect				
Physician discourse strategies→patient consultation satisfaction	0.009	0.019	0.068	0.130
Capacity-oriented strategy→patient consultation satisfaction	0.186	0.027	0.131	0.236
Quality-oriented strategy→patient consultation satisfaction	0.171	0.026	0.119	0.219
Goodwill-oriented strategy→patient consultation satisfaction	0.059	0.023	0.015	0.106
Mediation effect				
Physician discourse strategies→online physician-patient trust→patient consultation satisfaction	0.014	0.012	0.002	0.061
Capacity-oriented strategy→online physician-patient trust→patient consultation satisfaction	0.034	0.015	0.002	0.062
Quality-oriented strategy→online physician-patient trust→patient consultation satisfaction	0.038	0.016	0.003	0.067
Goodwill-oriented strategy→online physician-patient trust→patient consultation satisfaction	0.077	0.014	0.049	0.104
Total effect				
Physician discourse strategies→patient consultation satisfaction	0.113	0.017	0.086	0.140
Capacity-oriented strategy→patient consultation satisfaction	0.220	0.024	0.169	0.261
Quality-oriented strategy→patient consultation satisfaction	0.209	0.022	0.164	0.252
Goodwill-oriented strategy→patient consultation satisfaction	0.136	0.021	0.091	0.176

[a]LLCI: lower limit confidence interval.

[b]ULCI: upper limit confidence interval.

XSL•FO
RenderX

Discussion

Principal Findings

The study confirmed that the 3 primary discourse strategies (capacity-oriented strategy, quality-oriented strategy, and goodwill-oriented strategy) used by physicians have positive impacts on patient consultation behavior. It should be noted that the anticipated strong effect of the capacity-oriented strategy on patient consultation behavior was not observed. Instead, the effects of the quality-oriented strategy and goodwill-oriented strategy were more prominent. Physician notification adequacy from shared decision-making moderated the effect between the goodwill-oriented strategy and patient consultation behavior. Patient expression adequacy from shared decision-making moderated the effect between the capacity-oriented strategy and patient consultation behavior, and between the goodwill-oriented strategy and patient consultation behavior. Online physician-patient trust played significant mediating roles between physician discourse strategies and patient consultation behavior.

Theoretical Implications

This study addresses the gap in research on the influence of physician discourse strategies on patient consultation behavior, providing a new perspective and theoretical basis for the study of the antecedents of patient consultation behavior, especially in the online health environment. Previous studies on patient consultation behavior have mainly focused on influencing factors, such as text amount, response time, negative emotions, empathy, and information disclosure exhibited in physician texts [25,26,69,74]. Few studies have focused on the influence of discourse strategies on patient consultation behavior. Furthermore, this research extends the SOR theory, speech act theory, and trust theory by Mayer et al [30] to the study of patient behavior on online health platforms. Linguistic characteristics of physician texts are used to depict discourse strategies, including speech act features, modal resource features, and special linguistic resource features.

Moreover, the dimensions of research on the characterization of physician discourse strategies have been expanded. The study identified the significant value of managing physician discourse strategies in stimulating patient consultation behavior. Our findings demonstrate that physicians on the platform are required to have comprehensive expressions, in-depth knowledge of the patient's health condition, and accountability to the patient. This is more important than simply showing professional competence to enhance both the quantity and quality of consulting services. The results are counterintuitive, revealing that a quality-oriented strategy has a greater positive impact on patient consultation behavior compared to a goodwill-oriented strategy and capacity-oriented strategy, which have weaker effects. The findings suggest that physician discourse strategies that minimize the use of strong words, focus on comprehensive questioning about the patient's condition, and take responsibility for the patient's well-being are crucial for stimulating potential patients to seek the first consultation and reconsultations. Additionally, the use of emojis, first-person deixis, and honorifics can help physicians build rapport with patients and encourage them to seek consultations. On the other hand, patients do not prefer physicians who solely demonstrate professional competence, create a greater power distance between themselves and their patients, and adopt an assertive approach in their interactions.

Building on the generalized model of online trust proposed by Shankar et al [40] in 2002 and the SOR model [17,18], this study investigated the mediating role of online physician-patient trust between physician discourse strategies and patient consultation behavior. The concept of "online trust" in e-commerce platforms has been extended to online health platforms [44]. Our findings indicate that online physician-patient trust has a mediating effect on the relationship between physician discourse strategies and patient consultation behavior, offering guidance for the platform to strategically position its future competitiveness, particularly in the construction of the physician's image.

We divided shared decision-making into 2 dimensions based on its fundamental concept: physician notification adequacy and patient expression adequacy. Importantly, our study is the first to incorporate physician-patient shared decision-making as a moderating variable in the research of patient consultation behavior. Shared decision-making, which has gained attention in the medical field in recent years, lacks quantitative empirical research in medical practice [61,62]. Our findings revealed that physician notification adequacy only moderated the effect of the goodwill-oriented strategy on patient consultation behavior. It did not play a moderating role in the relationship between the capacity-oriented strategy and patient consultation behavior, as well as the quality-oriented strategy and patient consultation behavior. Patient expression adequacy had moderate effects on the capacity-oriented strategy and goodwill-oriented strategy in relation to patient consultation behavior, but it failed to moderate the relationship between the quality-oriented strategy and patient consultation behavior. It suggests that, when using the capacity-oriented strategy, allowing patients to express their doubts regarding their condition and treatment plan strongly impacts patient consultation behavior. This finding highlights the importance of communication barriers between physicians and patients. Under the capacity-oriented strategy, physicians pay little attention to whether patients can comprehend medical principles or professional considerations and instead adhere to the attitude of "just listen to me." Consequently, the participation and experiences of patients are compromised. For physician-patient communication to be smooth, patients must express their doubts and accurately understand the information conveyed by the physician. Under the goodwill-oriented strategy, there are fewer communication barriers between physicians and patients, creating more opportunities for the shared decision-making dimension to have a positive impact on patient consultation behavior. Conversely, under the quality-oriented strategy, the physician's expression has already provided a greater space for information exchange between the 2 parties. For example, using words, such as "better" and "suggest," and employing interrogative sentences can induce patients to provide a detailed description of their health condition. As a result, the role of shared decision-making may

XSL•FO
RenderX

be partially played by the quality-oriented strategy, and thus, it does not have a notable impact on patient consultation behavior.

Managerial Implications

In terms of contributions to platform management, our results offer guidance for the platform to strategically position its future competitiveness, particularly in terms of physician image construction. This includes the use of physician discourse strategies and shared decision-making, both of which can help increase patient consultation volume. It should be noted that the term "patients" refers to not only current patients but also potential patients, and the term "physicians" also includes artificial intelligence (AI) medical chatbots [76]. If physicians respond to patients in a more appealing and effective way, the online health consultation experiences of patients will be more enjoyable and comfortable. To meet patients' preferences for physician discourse strategies, communication training for physicians or future training of AI medical chatbots can be used to assist in physician-patient interactions, with the corpus in large language models designed accordingly. In addition, the platform should enhance the supervision and evaluation of the quality of physician-patient communication, develop scientifically formulated standards and norms for physicians to respond to patients, and improve physicians' communication skills and abilities [28,32].

Limitations

Regarding the classification of physician discourse strategies, it is important to note that different theories may lead to different categorizations. In this study, our focus was on how physician discourse strategies enhance patient consultation behavior by strengthening online trust between physicians and patients. The categorization of our discourse strategies is rooted in the trust theory. If future research is based on other theoretical frameworks, the understanding of physician discourse strategies will naturally be different.

To identify physician discourse strategies, our study adopted a combination of manual annotation and the random forest algorithm. This means that the "capacity-oriented strategy," "quality-oriented strategy," and "goodwill-oriented strategy" are determined by human judgments based on identification rules. While we designed a bias adjustment mechanism and calculated the level of bias among different judges, it is important to acknowledge that current prevalent large language models, such as the Bidirectional Encoder Representations from Transformers (BERT) algorithm and Generative Pre-trained Transformers (GPT) algorithm, may have a better understanding of semantics and higher efficiency than the manual annotation approach.

In this study, due to regulations regarding patient personal privacy on online health platforms, there were limitations in collecting data and samples. Moreover, there may have been endogenous issues like missing variables and 2-way causality, although we made some effort to solve these issues. In the future, we will conduct thorough research on these issues.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant number 72371007).

Authors' Contributions

MK took charge of the nominal group technique, data analysis, writing of the first draft, and revision of the manuscript. YW contributed to some part of the ordinary least squares model and manuscript revision. ML contributed to the project administration and resources. ZY was responsible for supervision, project administration, and funding acquisition. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of speech acts, modal resources, and special linguistic resources according to physician discourse strategies.
[DOCX File , 15 KB - jmir_v27i1e54516_app1.docx]

References

1. Digital China Development Report (2022). Cyberspace Administration of China. URL: https://www.cac.gov.cn/2023-05/22/c_1686402318492248.htm [accessed 2023-05-22]
2. Haodf. URL: https://www.haodf.com/ [accessed 2023-10-29]
3. Chunyu Doctor. URL: https://www.chunyuyisheng.com/ [accessed 2023-10-29]
4. JD Health. URL: https://ir.jdhealth.com/sc/index.php [accessed 2023-10-29]
5. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. Journal of Management Information Systems 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]

XSL•FO
RenderX

6. Shang-Fa W, Jing-Hua X. A conversation analysis of doctors' identity construction in the doctor-patient discourse. Medicine & Philosophy 2017;38(9A):36-39 [FREE Full text]
7. Markee N. Conversation Analysis. Hoboken, NJ: Wiley-Blackwell; 2000.
8. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. Ann Intern Med 2003 Dec 02;139(11):907-915. [doi: 10.7326/0003-4819-139-11-200312020-00009] [Medline: 14644893]
9. Levinson W, Pizzo PA. Patient-physician communication: it's about time. JAMA 2011 May 04;305(17):1802-1803. [doi: 10.1001/jama.2011.556] [Medline: 21540424]
10. Baker SC, Watson BM. How patients perceive their doctors' communication. Journal of Language and Social Psychology 2015 May 18;34(6):621-639. [doi: 10.1177/0261927x15587015]
11. Yongzhen W, Wenqiang WY. Challenge and coping strategies of the doctor-patient relationship under the new media environment. Chin J Soci Med 2016;33(6):511-513. [doi: 10.3969/j.issn.1673-5625.2016.06.002]
12. Veatch RM. Models for ethical medicine in a revolutionary age. What physician-patient roles foster the most ethical realtionship? Hastings Cent Rep 1972 Jun;2(3):5-7. [Medline: 4679693]
13. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
14. Frankel RM. From sentence to sequence: Understanding the medical encounter through microinteractional analysis. Discourse Processes 2009 Nov 11;7(2):135-170. [doi: 10.1080/01638538409544587]
15. Wang X, Mu C, Li H, Noble A, Wang Q. Product-centered language style in Chinese women online reviews. Front Psychol 2022 May 30;13:839064 [FREE Full text] [doi: 10.3389/fpsyg.2022.839064] [Medline: 35719497]
16. Mehrabian A, Russell JA. An Approach to Environmental Psychology. Cambridge, MA: The MIT Press; 1974.
17. Bitner MJ. Servicescapes: the impact of physical surroundings on customers and employees. Journal of Marketing 1992 Apr 01;56(2):57-71. [doi: 10.1177/002224299205600205]
18. Sah S, Loewenstein G, Cain D. Insinuation anxiety: concern that advice rejection will signal distrust after conflict of interest disclosures. Pers Soc Psychol Bull 2019 Jul 17;45(7):1099-1112. [doi: 10.1177/0146167218805991] [Medline: 30451077]
19. Wu H, Deng Z, Wang B, Wu T. Online service qualities in the multistage process and patients' compliments: A transaction cycle perspective. Information & Management 2020 Jul;57(5):103230. [doi: 10.1016/j.im.2019.103230]
20. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [doi: 10.1016/0277-9536(94)00155-m] [Medline: 7792630]
21. Liu J, Jiang H, Wang S. Physicians' online writing language style and patient satisfaction: the mediator of depth of physician-patient interactions. Healthcare (Basel) 2023 May 26;11(11):1569 [FREE Full text] [doi: 10.3390/healthcare11111569] [Medline: 37297708]
22. He Y, Guo X, Wu T, Vogel D. The effect of interactive factors on online health consultation review deviation: An empirical investigation. Int J Med Inform 2022 Jul;163:104781. [doi: 10.1016/j.ijmedinf.2022.104781] [Medline: 35512623]
23. Liu H, Zhang Y, Li Y, Albright K. Better interaction performance attracts more chronic patients? Evidence from an online health platform. Information Processing & Management 2023 Jul;60(4):103413. [doi: 10.1016/j.ipm.2023.103413]
24. Yang Y, Zhang X, Lee PK. Improving the effectiveness of online healthcare platforms: An empirical study with multi-period patient-doctor consultation data. International Journal of Production Economics 2019 Jan;207:70-80. [doi: 10.1016/j.ijpe.2018.11.009]
25. Jenkins L, Parry R, Pino M. Providing opportunities for patients to say more about their pain without overtly asking: a conversation analysis of doctors repeating patient answers in palliative care pain assessment. Appl Linguist 2021;42(5):990-1013 [FREE Full text] [doi: 10.1093/applin/amaa062]
26. Gemmiti M, Hamed S, Wildhaber J, Pharisa C, Klumb PL. Physicians' speech complexity and interrupting behavior in pediatric consultations. Health Commun 2022 May;37(6):748-759. [doi: 10.1080/10410236.2020.1868063] [Medline: 33441020]
27. Wu DC, Zhao X, Wu J. Online physician-patient interaction and patient satisfaction: empirical study of the internet hospital service. J Med Internet Res 2023 Aug 24;25:e39089 [FREE Full text] [doi: 10.2196/39089] [Medline: 37616031]
28. Tan H, Yan M. Physician-user interaction and users' perceived service quality: evidence from Chinese mobile healthcare consultation. ITP 2020 Jun 30;33(5):1403-1426. [doi: 10.1108/itp-01-2019-0039]
29. Shin SI, Hall DJ. Trust and the Flow Experience on Facebook: What Motivates Social Network Usage? In: Nah FH, Xiao B, editors. HCI in Business, Government, and Organizations. HCIBGO 2018. Lecture Notes in Computer Science, vol 10923. Cham: Springer; 2018:496-512.
30. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. The Academy of Management Review 1995 Jul;20(3):709. [doi: 10.2307/258792]
31. Kramer RM. Trust and distrust in organizations: emerging perspectives, enduring questions. Annu Rev Psychol 1999;50:569-598. [doi: 10.1146/annurev.psych.50.1.569] [Medline: 15012464]

XSL•FO
RenderX

32. Fung CH, Elliott MN, Hays RD, Kahn KL, Kanouse DE, McGlynn EA, et al. Patients' preferences for technical versus interpersonal quality when selecting a primary care physician. Health Serv Res 2005 Aug;40(4):957-977 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00395.x] [Medline: 16033487]
33. Luo H, Liu G, Lu J, Xue D. Association of shared decision making with inpatient satisfaction: a cross-sectional study. BMC Med Inform Decis Mak 2021 Jan 25;21(1):25 [FREE Full text] [doi: 10.1186/s12911-021-01385-1] [Medline: 33494744]
34. Goyal S, Chauhan S, Gupta P. Users' response toward online doctor consultation platforms: SOR approach. MD 2021 Sep 07;60(7):1990-2018. [doi: 10.1108/md-02-2021-0268]
35. Breen K, Cordner S, Thomson C. Good medical practice. Cambridge, United Kingdom: Cambridge University Press; 2010.
36. Patra A, Kumar A. Advancing ethics in medical education: A strategy to prevent medical negligence. J Punjab Acad Forensic Med Toxicol 2024;24(1):5-7. [doi: 10.5958/0974-083x.2024.00002.1]
37. Gilson L. Trust and the development of health care as a social institution. Soc Sci Med 2003 Apr;56(7):1453-1468. [doi: 10.1016/s0277-9536(02)00142-9] [Medline: 12614697]
38. Turner-Stokes L, Rose H, Ashford S, Singer B. Patient engagement and satisfaction with goal planning: Impact on outcome from rehabilitation. International Journal of Therapy and Rehabilitation 2015 May 02;22(5):210-216. [doi: 10.12968/ijtr.2015.22.5.210]
39. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Information Systems Research 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
40. Shankar V, Urban GL, Sultan F. Online trust: a stakeholder perspective, concepts, implications, and future directions. The Journal of Strategic Information Systems 2002 Dec;11(3-4):325-344. [doi: 10.1016/s0963-8687(02)00022-7]
41. Corritore CL, Kracher B, Wiedenbeck S. On-line trust: concepts, evolving themes, a model. International Journal of Human-Computer Studies 2003 Jun;58(6):737-758. [doi: 10.1016/s1071-5819(03)00041-7]
42. Joinson A, Reips U, Buchanan T, Schofield CBP. Privacy, trust, and self-disclosure online. Human-Comp. Interaction 2010 Jan;25(1):1-24. [doi: 10.1080/07370020903586662]
43. Krasnova H, Spiekermann S, Koroleva K, Hildebrand T. Online social networks: why we disclose. Journal of Information Technology 2010 Jun 01;25(2):109-125. [doi: 10.1057/jit.2010.6]
44. Kim DJ. A study of the multilevel and dynamic nature of trust in e-commerce from a cross-stage perspective. International Journal of Electronic Commerce 2014 Dec 05;19(1):11-64. [doi: 10.2753/jec1086-4415190101]
45. Nelson CR. Policy robustness: specification and simulation of a monthly money market model: comment. Journal of Money, Credit and Banking 1982 Nov;14(4):877. [doi: 10.2307/1991569]
46. Atanasova S, Kamin T, Petrič G. The benefits and challenges of online professional-patient interaction: Comparing views between users and health professional moderators in an online health community. Computers in Human Behavior 2018 Jun;83:106-118. [doi: 10.1016/j.chb.2018.01.031]
47. Wu G, Shi R, Zheng J. A Study on the impact of doctor-patient interaction on patient satisfaction with online medical care: based on the two factor theory. Health Econ Res 2023;40(3):52-56 [FREE Full text] [doi: 10.14055/j.cnki.33-1056/f.2023.03.009]
48. Liu X, Guo X, Wu H, Wu T. The impact of individual and organizational reputation on physicians' appointments online. International Journal of Electronic Commerce 2016 Jun 20;20(4):551-577. [doi: 10.1080/10864415.2016.1171977]
49. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
50. Park S, Chung N. Mediating roles of self-presentation desire in online game community commitment and trust behavior of Massive Multiplayer Online Role-Playing Games. Computers in Human Behavior 2011 Nov;27(6):2372-2379. [doi: 10.1016/j.chb.2011.07.016]
51. Zink K, Perry M, London K, Floto O, Bassin B, Burkhardt J, et al. "Let me tell you about my…" Provider self-disclosure in the emergency department builds patient rapport. West J Emerg Med 2017 Jan 19;18(1):43-49 [FREE Full text] [doi: 10.5811/westjem.2016.10.31014] [Medline: 28116007]
52. Hu Y, Zhou H, Chen Y, Yao J, Su J. The influence of patient-generated reviews and doctor-patient relationship on online consultations in China. Electron Commer Res 2021 Aug 16;23(2):1115-1141. [doi: 10.1007/s10660-021-09506-8]
53. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
54. Waldron T, Carr T, McMullen L, Westhorp G, Duncan V, Neufeld S, et al. Development of a program theory for shared decision-making: a realist synthesis. BMC Health Serv Res 2020 Jan 23;20(1):59 [FREE Full text] [doi: 10.1186/s12913-019-4649-1] [Medline: 31973754]
55. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2015 Feb 01;175(2):274-286. [doi: 10.1001/jamainternmed.2014.6016] [Medline: 25531451]
56. Hoffmann TC, Del Mar C. Clinicians' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2017 Mar 01;177(3):407-419. [doi: 10.1001/jamainternmed.2016.8254] [Medline: 28097303]
57. Ren W. Mitigation in Chinese online consumer reviews. Discourse, Context & Media 2018 Dec;26:5-12. [doi: 10.1016/j.dcm.2018.01.001]

XSL•FO
RenderX

58. Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. Patient Educ Couns 2010 Jul;80(1):94-99. [doi: 10.1016/j.pec.2009.09.034] [Medline: 19879711]
59. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
60. Fowler FJ, Sepucha KR, Stringfellow V, Valentine K. Validation of the SDM process scale to evaluate shared decision-making at clinical sites. J Patient Exp 2021 Nov 26;8:23743735211060811 [FREE Full text] [doi: 10.1177/23743735211060811] [Medline: 34869847]
61. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making 2015 Jan;35(1):114-131 [FREE Full text] [doi: 10.1177/0272989X14551638] [Medline: 25351843]
62. Stacey D, Légaré F, Pouliot S, Kryworuchko J, Dunn S. Shared decision making models to inform an interprofessional perspective on decision making: a theory analysis. Patient Educ Couns 2010 Aug;80(2):164-172. [doi: 10.1016/j.pec.2009.10.015] [Medline: 19945813]
63. Austin JL. How to Do Things with Words. Oxford, England: Clarendon Press; 1962.
64. Searle JR, Vanderveken D. Foundations of Illocutionary Logic. Cambridge, United Kingdom: Cambridge University Press; 1985.
65. Brown R, Gilman A. The Pronouns of Power and Solidarity. In: Sebeoki TA, editor. Style in Language. Cambridge, MA: The MIT Press; 1960:253-276.
66. Ellingson LL, Buzzanell PM. Listening to women's narratives of breast cancer treatment: a feminist approach to patient satisfaction with physician-patient communication. Health Commun 1999;11(2):153-183. [doi: 10.1207/s15327027hc1102_3] [Medline: 16370974]
67. Gill V, Maynard D. Explaining illness: patients' proposals and physicians' responses. In: Heritage J, Maynard D, editors. Communication in Medical Care: Interaction between Primary Care Physicians and Patients. Cambridge, United Kingdom: Cambridge University Press; 2006:115-150.
68. Yang H, Zhang X. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: panel data analysis. J Med Internet Res 2019 Mar 22;21(3):e12156 [FREE Full text] [doi: 10.2196/12156] [Medline: 30900997]
69. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: Empirical study from elaboration likelihood perspective. Computers in Human Behavior 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
70. Strekalova YA, Krieger JL, Neil J, Caughlin JP, Kleinheksel AJ, Kotranza A. I understand how you feel. Journal of Language and Social Psychology 2016 Aug 20;36(1):61-79. [doi: 10.1177/0261927x16663255]
71. Bertakis KD. The influence of gender on the doctor-patient interaction. Patient Educ Couns 2009 Sep;76(3):356-360. [doi: 10.1016/j.pec.2009.07.022] [Medline: 19647968]
72. Weisman CS, Teitelbaum MA. Physician gender and the physician-patient relationship: recent evidence and relevant questions. Soc Sci Med 1985 Jan;20(11):1119-1127. [doi: 10.1016/0277-9536(85)90189-3] [Medline: 3895448]
73. Oster A, Wiking E, Nilsson GH, Olsson CB. Patients' expectations of primary health care from both patients' and physicians' perspectives: a questionnaire study with a qualitative approach. BMC Prim Care 2024 Apr 24;25(1):128 [FREE Full text] [doi: 10.1186/s12875-024-02389-2] [Medline: 38658808]
74. Yu H, Jiang S, Land KC. Multicollinearity in hierarchical linear models. Soc Sci Res 2015 Sep;53:118-136. [doi: 10.1016/j.ssresearch.2015.04.008] [Medline: 26188442]
75. MacKinnon DP, Fritz MS, Williams J, Lockwood CM. Distribution of the product confidence limits for the indirect effect: program PRODCLIN. Behav Res Methods 2007 Aug;39(3):384-389 [FREE Full text] [doi: 10.3758/bf03193007] [Medline: 17958149]
76. Gaczek P, Pozharliev R, Leszczyński G, Zieliński M. Overcoming consumer resistance to AI in general health care. Journal of Interactive Marketing 2023 Feb 28;58(2-3):321-338. [doi: 10.1177/10949968221151061]

Abbreviations

AI: artificial intelligence
CFI: comparative fit index
RMSEA: root mean square error of approximation
SOR: stimulus, organism, and response
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
VIF: variance inflation factor

XSL•FO
RenderX

Edited by A Schwartz; submitted 13.11.23; peer-reviewed by Y Liu, H Yang; comments to author 12.02.24; revised version received 07.04.24; accepted 31.01.25; published 19.03.25.

Please cite as:
Kong M, Wang Y, Li M, Yao Z
Mechanism Assessment of Physician Discourse Strategies and Patient Consultation Behaviors on Online Health Platforms: Mixed Methods Study
J Med Internet Res 2025;27:e54516
URL: https://www.jmir.org/2025/1/e54516
doi:10.2196/54516
PMID:

©Menglei Kong, Yu Wang, Meixuan Li, Zhong Yao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Influencing Informed Consent Preferences in Digital Health Research: Survey Study of Prospective Participants

Brian J McInnis[1], PhD; Ramona Pindus[2], BS; Daniah Kareem[2], BSPH; Daniela G Vital[2], BSPH, MPH; Eric B Hekler[2], PhD; Camille Nebeker[2], MS, EdD

[1]School of Information, University of Texas at Austin, Austin, TX, United States

[2]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Camille Nebeker, MS, EdD
Herbert Wertheim School of Public Health and Human Longevity Science
University of California San Diego
9500 Gilman Dr MC 0811
La Jolla, CA, 92093-0811
United States
Phone: 1 6195182875
Email: nebeker@health.ucsd.edu

Abstract

Background: Readability is important to consider when developing informed consent communications for prospective research participants, but not the most important consideration. Other factors to consider relate to learning preferences and literacy needs of people recruited to participate in research, as these factors can influence understanding of consent communications. To promote understanding among prospective participants, researchers should take a human-centered approach to develop consent communications.

Objective: This study aims to explore how factors related to readability, topic, and participant demographic characteristics play into preferences for digital health research consent material. These factors are important to consider as not attending to some details that matter to a specific subgroup of prospective participants may systematically exclude people from research.

Methods: People eligible to participate in a digital health study were recruited to review 31 paragraph length sections of a consent form, referred to as "text snippets," for an existing institutional review board–approved digital health study. Participants (N=79) were surveyed and asked to choose between 2 variations of the text snippets, either indicating a preference for the institutional review board–approved original or a version that was modified to improve readability.

Results: A slim majority of participants provided feedback about the snippets (n=44; 55%). Our qualitative analysis of the feedback found that participants preferred shorter snippets, in general, but the snippets also elicited new questions not addressed by the original consent material. This observation is supported by our quantitative analysis, which found that when the character length of the original was longer, participants were less likely to prefer the original (P<.001) and more likely to prefer the modified text by a factor of 1.20 times (P=.04), and particularly for snippets explaining study risks (P=.03). Our analysis also found significant differences in participant demographic characteristics. For example, older participants tended to prefer the original more than younger participants, by a factor of 1.95 times (P=.004). The results present illustrative examples of how factors related to sex, age, physical activity, and ethnicity all play into preference for consent communication.

Conclusions: The findings point toward new ways of evaluating informed consent communication: (1) for responsiveness to specific prospective participant populations, and (2) effectiveness at eliciting informed questions from people considering participation. We discuss how creating partnerships with prospective participants to prototype informed consent materials, specifically study procedures and risks, can be a way to identify those details before launching a study. Furthermore, future research should go beyond "readability" to explore alternate measures of evaluating consent materials, such as the likelihood that the consent material and communication procedures will elicit "informed questions" for the research team.

(J Med Internet Res 2025;27:e63349) doi:10.2196/63349

XSL•FO
RenderX

KEYWORDS

digital health; research ethics; informed consent; readability; health literacy; human-centered approach; consent communication; text snippet; qualitative analysis; effectiveness; health information; health informatics

Introduction

Background

The process of obtaining and providing informed consent to participate in research is a cornerstone of research ethics. The content of the consent form and its readability has been a topic of concern in health research and is well documented in academic literature [1-3]. Several peer-reviewed studies have focused on improving the ability of prospective research participants to comprehend these documents [4-6]. Specific to promoting the readability of consent materials, institutional review boards (IRBs) in the United States suggest the consent form be written at a sixth- to eighth-grade reading level. Aiming for the sixth- to eighth-grade reading level is a practical guideline with the goal of providing comprehensive information that is, ideally, accessible to most prospective participants. Researchers and IRBs try to follow the reading level guidelines and create consent communications that are clear, concise, and accessible to those who may consider participation. Consent forms that use plain language strategies, including shorter sentences and fewer syllables, appear to improve participant comprehension [7,8]. Consent forms that are not accessible compromise the goal of being informed and undermine the ethical principle of respect for persons [9,10].

Digital health research introduces new complexities to the process of communicating study information to prospective participants. Digital health research can involve data collection strategies and intervention delivery using tools like mobile apps, wearable sensors, social media platforms, and ecological momentary assessment. There is also growing attention on the use of health data repositories, including electronic health records, for training machine learning and other forms of artificial intelligence. These technologies support health research in many ways. For example, health data including sleep, heart rate, and step count can be collected passively in real time. Surveys assessing caloric intake, physical activity, and mental health can be deployed at scheduled or random times using an ecological momentary assessment app, reducing the need for recording data in a journal or relying on self-report [11] and personalized algorithms may be used to test strategies to motivate increased physical activities [12].

There are benefits associated with the use of digital health tools and strategies in health research, but there are also new risks associated with the volumes of granular health data produced and the potential inferences emerging from combining various data types. These risks involve data management protocols and privacy limitations that are important to convey to those considering research participation [10]. Moreover, many of the technologies used in digital health research are commercial products that may have data management processes and privacy protections that put a prospective participant at increased risk of harm. As with any health research, including clinical trials, some specialized knowledge may be needed to understand the study purpose and approach to data collection and management. It is also unlikely that people will have the time to obtain the foundational knowledge needed to facilitate truly informed consent. These tensions create additional challenges for digital health research. While a patient recruited to participate in a study may be somewhat knowledgeable about their specific health condition, they may not understand how the technologies used in digital health research gather, process, and store their data [10]. In fall 2023, the National Institutes of Health (NIH) posted a request for information specific to how informed consent may be improved for digital health research, which included sample language and points to consider [13]. This initiative by the NIH to assist the digital health research community in communicating complex digital health study information is consistent with the goals of our study reported in this paper.

This survey-based study was conducted to identify ways of improving consent communications in digital health research. Specifically, the study investigated the following research question exploring the relationship between the content of consent communications and prospective participant background characteristics on preferences for more and less information in consent material.

Research Question

How do factors related to the content of consent material (eg, readability, length, and lexical diversity) and background characteristics associated with prospective participants (eg, sex, age, physical activity, and tracking physical activity) play into preferences for more and less information in an informed consent process?

This research received support from an NIH bioethics supplement to a parent study called “Your Move.” The “Your Move” study sought to test if a personalized digital health intervention that included several features including a smartwatch, a personalized adaptation algorithm called a controller, and a web-based self-experimentation tool, could help individuals improve their physical activity relative to providing digital health support that is not personalized and represents current deployed wellness program offerings. One challenge, when designing this study, was how to convey to prospective participants that a personalized algorithm would be used to nudge behavior change toward increased physical activity. This paper presents recommendations about how researchers can address these types of concerns, by making practical improvements to informed consent materials and procedures.

Methods

Participant Recruitment

Our study recruited participants who met the eligibility criteria to participate in the “Your Move” digital health study. Potential participants were recruited through a variety of sources, such

XSL•FO
RenderX

as digital research portals (eg, ResearchMatch and Craigslist), via word of mouth through community partnerships with local university and public-school employee networks, as well as through digital advertisements.

Participant eligibility was the same as for the parent digital health study with the following inclusion and exclusion criteria.

Inclusion Criteria

Inclusion criteria were as follows: ability to read English; age of 25 years or older at enrollment; being physically inactive, which in the context of the study means less than 150 minutes (about 2.5 hours) of exercise per week; and the ability to access the internet through a smartphone, computer, or internet-connected device.

Exclusion Criteria

Exclusion criteria were as follows: participants develop a physical or mental health issue that prohibits compliance with the study's protocol, participants become pregnant, and participants do not follow the instructions given to them by the research team.

Ethical Considerations

This study was registered with the University of California San Diego IRB (#201720) and determined to be exempt as defined by the Common Rule (45 CFR 46.104(d)(2)). The consent process involved emailing prospective participants a summary of the study's purpose and procedures, and a link to an interest survey. The digital interest survey included the full text of the consent communication describing study involvement on the first page, with a response item to indicate willingness to participate. Those who did not agree to participate automatically exited the survey. All participants were recruited during June 2022. The study team exported the collected data on a regular basis and stored it in a secure folder with limited access under the University of California San Diego Health Microsoft OneDrive environment. Storage conforms to the Health Insurance Portability and Accountability Act requirements. Completed surveys were manually deidentified by removing names and email addresses. After deidentifying the survey data, the list of participants was randomized by assigning everyone a random number in Excel. This list was shared via University of California San Diego Health Outlook email, utilizing encryption technology. Participants received a US $20 value Amazon.com gift card via email after confirming completion of the survey.

Survey-Based Study Procedures

The survey-based study included three parts to elicit: (1) perspectives on information provided using standard consent processes, in general, (2) ratings to evaluate the clarity and comprehensiveness of specific snippets of consent information, and (3) demographic and background characteristics to consider the participant's eligibility to participate and interest in the digital health technology study (eg, level of physical activity). Part 2 of the study involved reviewing 16 pairs of consent communication "snippets," which were small sections of the consent form text. The snippets differed in terms of their assessed readability, specifically, each pair included a snippet of text from the IRB-approved consent form (hereafter the "original") as well as a version of the original that had been improved for readability (hereafter the "modified").

To create the modified variation of the original consent materials, three members of the research team independently used a free web-based readability analysis software, Readability Calculator, to rewrite the original text, while using the software to monitor how their edits played into specific measures, including the character length, Flesch Kincaid Reading Ease, and lexical density. All researchers then compared their modified text to each original snippet, and associated readability measures, to agree on a final version of the modified text that was the "most readable" based on the readability software (Multimedia Appendix 1). An illustration of a snippet before and after modification is evident in snippet 7 (Textbox 1), which describes the procedure for assessing participant heart rate with moderate to difficult walking conditions.

Textbox 1. Snippet 7 example of the original and modified versions of the consent material presented to participants.

Original text from the study consent form

- "You will complete a treadmill walking test, while wearing a heart monitor on your chest, so that the research team can record your heart rate and blood pressure. The treadmill test will last approximately 5-10 minutes. The treadmill will start at a slow pace and gradually increase in speed and steepness, as though you are gradually walking faster up a hill. If you want to stop the test for any reason—such as if you feel lightheaded, dizzy, or breathless—a member of the research team will be there with you to help."
- Flesch Kincaid Grading level of 10.52.

Modified version of the consent material

- "You will be asked to walk on a treadmill for 5-10 minutes, or as long as you can, while wearing a monitor on your chest to record your heart rate and blood pressure. A member of the research team will be there to help."
- Flesch Kincaid Grading level of 8.92.

To confirm that the modified versions were, in fact, more readable than the original text based on standard measures, Table 1 presents descriptive statistics comparing the original and modified snippets, which indicates significant differences in terms of reading ease and character length. However, lexical diversity was not significantly different, which may be due to challenges in finding synonyms for technical and scientific terminology.

XSL•FO
RenderX

Table 1. Descriptive statistics related to text analysis variables.

	Mean (SD)	Maximum	ANOVA comparison between original and modified, *F* test (*df*)
Original			
Flesch Reading Ease	49.70 (9.31)	64.59	—[a]
Character length	310.53 (151.81)	766.00	—
Lexical diversity	47.65 (5.77)	57.81	—
Modified			
Flesch Reading Ease	55.69 (13.81)	77.41	4.13 (1, 62)[b]
Character length	173.37 (91.57)	480.00	19.51 (1, 62)[c]
Lexical diversity	49.62 (10.21)	80.00	0.89 (1, 62)

[a]No comparison.

[b]$P<.01$.

[c]$P<.001$.

During part 2 of the survey protocol, participants were instructed to compare the original and modified versions of 16 of 31 total snippets, which included 1 comparison that was evaluated by all participants (ie, "Item #01 Purpose: What is this study about?"). Specifically, participants were presented with the original text and asked the following questions.

- How relevant is the content to you (not at all, somewhat, very)?
- How clear is the content to you (not at all, somewhat, very)?

Then participants were presented with the original and modified texts, in a randomized order. Randomization was used to account for any potential ordering effect, which may influence the degree to which participants engage with the materials. For each snippet pair, participants were instructed the following.

"The following are 2 variations on the consent language. Please indicate the variation that you prefer," using the options: (1) prefer version 1; (2) prefer version 2; (3) both versions 1 and 2 look good; and (4) both need improvement.

While options 1 and 2 directly contrast the modified and original options, options 3 and 4 allow for the possibility that the participant did not perceive enough of a difference between the 2 options. For consistency, throughout the paper we refer to preferences for either the original, modified, or both versions (option 3) as "good," and both versions needing improvement (option 4) as "need." From a methodological perspective, including options 3 and 4 also reduces the likelihood of guessing, which otherwise would add noise to the analysis. After comparing the original and modified snippets, participants were asked "What changes would you recommend to the text above?" to capture additional information about their perceptions of the content.

The snippets were labeled as either about the study purpose, procedures, risks, or benefits. All participants (N=79) reviewed a single snippet about the digital health technology study purpose. While all participants reviewed a total of 15 additional snippets, when they initiated the survey, they were randomly assigned to either read a set of snippets that included more risk items (called "survey A") or more procedural items (called "survey B"). The random assignment resulted in 41 (52%) participants receiving survey A and 38 (48%) participants receiving survey B. We found no significant difference in the distribution of participant preferences for the original and modified versions of the snippets ($\chi^2_6=5.9$; $P=.42$; Table 1). This indicates that the responses can be evaluated together in a series of logistic regression analyses. During the logistic regression analysis process, we considered including a variable to capture the effects associated with the survey A and survey B groups, but we found that including this variable did not significantly improve model performance and explainability.

Mixed Methods Data Analysis

To investigate factors that increase the likelihood of participants preferring the original or modified versions of the consent snippets, a series of logistic regressions were fit to evaluate each preference independently: that is, prefer original, prefer modified, both are good, and both need improvement. A logistic regression is used to evaluate the log odds of a condition being true or false, where model A presents the log odds of preferring the original ("true") over all other options ("false").

To investigate how factors related to the content (eg, readability, length, lexical diversity) and background characteristics associated with prospective participants (eg, sex, age, physical activity, and tracking physical activity) play into preferences for the original or modified versions, we applied the survey data and reading analysis tools to construct a series of independent measures for the logistic regression analysis. Each model was fit by beginning with a beyond optimal set of the independent predictors, and then measures of overall "goodness of fit," including a chi-square test of the model with and without predictors, were used in conjunction with tests to evaluate the effect of each predictor on the model, such as the Wald test, to select predictors for each final model. This iterative process helped to identify potential confounding factors in the analysis. As an example, initially, our team considered a broader range of readability measures (eg, Gunning Fog Index, Flesch Kincaid Grade level, Automated Readability Index); however, we found

XSL•FO
RenderX

that many of the measures were highly correlated. To address this, we selected readability measures that reflect a range of important considerations in writing and that are not well correlated (Table 2).

Table 2. Correlations between text analysis variables.

	Original			Modified		
	Flesch Reading Ease	Character length	Lexical diversity	Flesch Reading Ease	Character length	Lexical diversity
Original						
Flesch Reading Ease	1.00	–0.21	–0.12	0.20	0.06	0.09
Character Length	–0.21	1.00	0.08	–0.12	0.54	–0.05
Lexical Diversity	–0.12	0.08	1.00	–0.20	0.08	0.69
Modified						
Flesch Reading Ease	0.20	–0.12	–0.20	1.00	–0.31	–0.14
Character length	0.06	0.54	0.08	–0.31	1.00	–0.09
Lexical diversity	0.09	–0.05	0.69	–0.14	–0.09	1.00

In the findings, the logistic regressions table presents the estimate and standard error for each predictor in each model. To improve the interpretation of the models, the inline description of the analysis in the Results presents odds ratios (ORs), which are the predictor estimates exponentiated. The raw estimates are on a logit scale, which is from negative infinity to positive infinity, but when exponentiated the predictor OR reflects the effect of the predictor on the probability of the dependent measure occurring. For example, in model A the effect of sex male on the preference for the original snippet is –0.68, which when exponentiated as OR is 0.50. As the OR of 0.50 is less than 1, the interpretation is that participants who identify as male are 0.50 times less likely than female participants to prefer the original snippet. As another example, the estimate for sex male in model C is 1.35, which when exponentiated translates to an OR of 3.84, implying that male participants are 3.84 times more likely to feel that both snippets are good enough than female participants. For measures of readability, the interpretation reflects a one standard deviation change in the predictor, for example, in model A an increase in the modified snippet character length by one standard deviation (estimate=–0.39) contributed to a decrease in the likelihood that participants preferred the original snippet by a factor of 0.67 times.

Results

Review of Participant Demographics

In total 79 of 118 (67%) initiating the survey completed the study. Just more than half of the participants identified as female (n=41, 52%), most were under the age of 40 years (n=51, 64%), and 16 participants identified as Hispanic or Latino (20%). The participants reported a range of self-assessed physical activity. The number of people rating themselves as “more active than average” (n=31, 39%) was roughly equivalent to those rating themselves “less active than average” (n=32, 40%). A third of participants also indicated that they track their physical activity, which is important to note as the study context involves a digital health technology intended to promote an active lifestyle (Table 3).

XSL•FO
RenderX

Table 3. Descriptive statistics related to participant backgrounds (N=79 participants).

	Value, n (%)
Sex	
Female	41 (52)
Male	38 (48)
Age (years)	
Under 30	23 (29)
30-39 years	28 (35)
40-49 years	17 (21)
50-59 years	7 (9)
60 years or older	4 (5)
Ethnicity (Hispanic)	
No	63 (80)
Yes	16 (20)
Physical activity	
Average	16 (20)
Less active	32 (40)
More active	31 (39)
Track physical activity	
Yes	25 (31)
No	54 (69)

The majority preferred the original to the modified version improved for readability, then both are good enough. In just 6 comparisons, participants preferred the modified version, which was improved for readability, to the original IRB-approved text (Table 4). To investigate why participants may have preferred the modified version to the original and vice versa, we first reviewed feedback from participants. Of the 79 participants, 44 (55%) provided feedback related to the snippets, and those who provided feedback provided a fair amount (mean 6.5, SD 5.24 responses per participant).

Table 4. Preferences by snippet and consent form topic[a].

Preferences for the consent form snippets	Original, %	Modified. %	Good, %	Need, %
Benefits	36.90	26.19	32.74	4.17
Procedures	35.29	23.11	33.61	7.98
Purpose	34.15	27.44	29.88	8.54
Risks	33.72	26.05	31.16	9.07

[a]Percentages are based on the total number of snippets reviewed per topic. Specifically, the "benefits" section included 4 snippets, "procedures" included 12, "purpose" included 3, and "risks" included 11. Overall, participants tended to prefer the original version of the snippets; however, our qualitative analysis of the feedback and quantitative analysis of preferences by subgroup indicate that there are many additional factors to consider when developing a consent communication.

In snippet comparisons where participants preferred the modified version, participants shared 55 responses (31 unique participants). Most responses indicate that they preferred the modified version because it was shorter and less confusing than the original version, "[original] has too much text about eligibility. I prefer [modified] because it was more concise: If you are eligible..." [Participant 57] and "[original] too many numbers thrown at me-way way too wordy" [Participant 16]. However, participants also requested additional details that were not included in either version. Specifically, the 55 responses included 19 follow-up questions for the researchers, such as "Is it a licensed health care professional conducting the exam?" [Participant 28], "Maybe clarify how points are earned using an example" [Participant 26], and "[...] do I have to take time off work?" [Participant 25]. While participants appreciate brevity, the consent snippets elicited a range of follow-up questions for the research team.

For the most part, participants preferred the original version (in 14 comparisons) and shared a total of 140 responses (made by 40 unique participants) about their preferences. In these cases, many participants felt that the modified text was too short and "[...] leaves too many questions" [Participant 25]; though

XSL•FO
RenderX

participants also made specific suggestions about how to trim the original version, “I think everything after measuring your blood sugar level is not necessary” [Participant 06] and “too much jargon” [Participant 47]. In some instances, the modified text was missing key details, “[modified] does not state that the treadmill will start at a slow pace and gradually increase speed. That is an important detail missing from [the] text” [Participant 23]. In general, participants preferred a trimmed-down version of the original, rather than the modified version.

The desire for more information was especially strong in response to snippets about study procedures and risks. While participants tended to prefer the original version to the modified version, they also identified many missing details, such as, whether they would be disqualified if stopping during a physical fitness examination involving a treadmill. However, participants did not raise any questions about the time commitment, retracting consent, removal from the study, costs associated with participation, and contact information for the research team. For a research team, knowing when to add and remove details from a consent document can be challenging, if they have yet to share drafts with prospective participants.

In 11 of the comparisons, participants indicated that they had no preference between the original and modified, that is, both are good. Participants shared 94 responses about these snippet comparisons (31 unique participants). In these cases, participants felt that both the original and modified versions had value that could be combined, “[modified] is more clear, but does not include the daily survey requirements listed in [original]” [Participant 40] and “[modified] needs to be more specific on what it means to have mental health issues, [but the original] has a clear understanding” [Participant 43].

Several participants suggested ways to rewrite the snippets to promote specific values in the communication, “combine them to remain positive: Our staff will show you how to wear the Fitbit so that it is comfortable. Our staff will show you how to wear it, help you adjust if needed, and provide an (optional) tutorial” [Participant 17]. However, 1 participant was just not sure what to suggest, “This just sounds awkward to me. I don't know how to improve. Maybe even deleting it completely” [Participant 6]. While each version of the consent material could be improved, in most comparisons, participants preferred the original version or felt that the modified and original could be combined.

Risks and Readability, but Inconclusive

To explore how factors related to the consent communication content play into participant preferences for the original or modified versions, we conducted a logistic regression analysis of each preference category: that is, prefer original, prefer modified, prefer both, prefer neither. Presented in Table 5, our analysis evaluated how various factors related to content (eg, total character count including spaces and punctuation, reading level), as well as participant background characteristics (eg, sex and physical activity), may play into consent communication preferences.

XSL•FO
RenderX

Table 5. Logistic regressions to evaluate how aspects of the study and participant backgrounds relate to preferences. Note that all 79 participants responded to snippet 1, then 41 participants responded to 14 items, and 38 participants responded to 15 other items (a total of 15 items for all participants)=1223 responses.

	Model A: prefer original[a]			Model B: prefer modified[b]			Model C: both are good[c]			Model D: neither[d]		
	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
Intercept	0.02	0.85	—[e]	–0.87	0.90	—	0.43	0.93	—	–9.54	1.68	<.001
Consent section (by comparison to "benefits")												
Procedures	–0.06	0.22	—	0.19	0.24	—	–0.23	0.25	—	0.45	0.47	—
Purpose	–0.11	0.28	—	0.18	0.29	—	–0.05	0.31	—	0.28	0.57	—
Risks	–0.28	0.24	—	0.59	0.26	.01	–0.55	0.26	.01	0.89	0.49	<.05
Sex												
Male	–0.68	0.14	<.001	–0.47	0.16	.001	1.35	0.16	<.001	–0.20	0.26	—
Age (years)												
30-39	0.35	0.19	.05	–0.19	0.20	—	–0.38	0.22	.05	0.39	0.40	—
40-49	0.67	0.23	.001	–0.15	0.25	—	–1.49	0.37	<.001	0.72	0.47	—
50-59	0.29	0.31	—	–1.07	0.35	.001	–2.98	1.03	.001	3.44	0.46	<.001
60 or older	–0.30	0.18	.05	–0.95	0.20	<.001	0.77	0.17	<.001	1.08	0.33	<.001
Ethnicity (Hispanic)												
Yes	–0.23	0.16	—	0.22	0.17	—	–0.30	0.19	—	0.88	0.26	<.001
Physical activity												
Less active	–0.13	0.18	—	–0.77	0.20	<.001	0.63	0.24	.001	1.70	0.41	<.001
More active	–0.72	0.19	<.001	–0.26	0.20	—	0.92	0.22	<.001	0.79	0.38	<.01
Track physical activity												
Yes	–0.03	0.15	—	–0.91	0.17	<.001	0.66	0.18	<.001	0.96	0.28	<.001
Original												
Character count (scaled)	0.17	0.08	.01	0.15	0.08	.05	–0.32	0.10	.001	–0.13	0.15	—
Lexical diversity	0.01	0.02	—	–0.02	0.02	—	–0.02	0.02	—	0.05	0.03	—
Reading ease	0.01	0.01	—	0.00	0.01	—	–0.02	0.01	.05	0.00	0.01	—
Modified												
Character count (scaled)	–0.39	0.09	<.001	0.19	0.09	.01	0.19	0.10	.05	0.20	0.15	—
Lexical diversity	–0.02	0.01	.05	0.02	0.01	.05	0.00	0.01	—	0.01	0.02	—
Reading ease	–0.01	0.01	—	0.01	0.01	.05	–0.01	0.01	—	0.02	0.01	<.01

[a]Goodness-of-fit χ^2_{18}=142.0; *P*<.001.

[b]Goodness-of-fit χ^2_{18}=88.9; *P*<.001.

[c]Goodness-of-fit χ^2_{18}=340.9; *P*<.001.

[d]Goodness-of-fit χ^2_{18}=101.2; *P*<.001.

[e]No significant effect.

A primary observation is that character count (scaled) significantly affects a participant's preference for the original version of the consent material. Specifically, for every SD increase in the character length for the modified version, there is a corresponding decrease in the likelihood that participants prefer the original version, by a factor of 0.67 times (*P*<.001), and a corresponding increase in the likelihood that they prefer the modified version, by a factor of 1.20 times (*P*=.04). This directly responds to participant feedback that they tended to prefer the readability of the modified versions, though appreciate when additional details are provided. Other readability factors,

XSL•FO
RenderX

including lexical diversity and Reading Ease, did not significantly affect participant preferences in our analysis.

Our analysis also found that the consent section (eg, purpose, procedures, risks, and benefits) played into participant preferences. Specifically, participants were slightly more likely to choose the modified version when presented with information about study risks, by a factor of 1.8 times (P=.03), and less likely to select that "both versions 1 and 2 look good," by a factor of 0.57 times (P=.04). An initial takeaway is that participants tended to prefer risk communications that are more readable. As the results are marginally significant, future research should investigate the specific study risk topics that prospective participants would like described in more and less detail, rather than simply comparing differences between risks and benefits, as is the case in our analysis.

Sex, Age, Physical Activity, and Tracking Physical Activity

Readability in consent communication, particularly character count, is an important consideration; however, our analysis highlights other factors related to the participants, specifically their demographic and background characteristics, that play into their consent communication preferences. Participants who identified as male were significantly more likely to indicate that "both versions 1 and 2 look good," than participants identifying as female, by a factor of 3.84 (P<.001). This observation is useful to note, as female participants may have had stronger opinions about consent communication. Specifically, female participants were significantly more likely to choose between the original and modified versions, by a factor of 1.96 times (P<.001) and 1.60 times (P=.003). Future studies investigating consent communication content should carefully consider how sex identity may factor into preferences.

Additionally, our analysis found differences by age group. While participants in our study, grouped between the ages of 20-29 and 30-39 years were not statistically different in their preferences for consent communication, participants aged 40 years and older wanted different considerations. Specifically, participants aged 40-49 years tended to prefer the original version more than younger participants, by a factor of 1.95 times (P=.004) and were significantly less likely to state that both versions look good, by a factor of 0.22 times (P<.001). While our study recruited relatively few participants in the age ranges of 50-59 years (n=7, 8%) and 60 years and older (n=4, 5%), they tended to not prefer the modified, rather feeling that both versions needed improvement, in comparison to younger participants. As the specific digital health study used as a context for our research involved physical exercise and cardiovascular risks, prospective participants who are older in age may have preferred the original version of the consent materials because it included more detail. Future research should explore this trend further by recruiting older adult participants.

Study participants were asked about their level of physical activity, as people with already high or already low levels of activity may value different levels of detail about the study. By comparison to people who rate their physical activity as average, our analysis found that participants who perceive their activity as lower-than-average were less likely to prefer the modified version, by a factor of 0.46 times (P<.001), whereas participants with higher-than-average physical activity were less likely to prefer the original versions, by a factor of 0.48 times (P<.001). Our interpretation is that these differences in what participants do not prefer points toward general feelings that both versions were missing details that are relevant to the 60% of study participants (n=48) who rated their physical activity as higher or lower than average; however, these groups may need different levels of study detail.

Participants were also asked about their physical activity and tracking behaviors, as the digital health study context would involve wearing a smartwatch fitness tracking device. Our analysis found that participants who already track their own physical activity were less likely to prefer the modified version than participants who do not track their physical activity, by a factor of 0.40 times (P<.001). By comparison, participants who do not track their physical activity preferred the modified version, by a factor of 2.49 times (P<.001). While recommendations about how to communicate study details tend toward providing as much detail as possible, in our study the participants tended to prefer a pared-down version.

Finally, participants identifying as Hispanic or Latino felt that both versions of the consent communication need improvement, by a factor of 2.40 (P<.001). This is likely due to an oversight in our study design, as both versions of the content were presented in English, and participants were not offered a translated version. Despite this oversight in the study design, the strong signal in our analysis should be taken as a reminder to pay attention to the language needs of participants. In general, preferences related to the option "both need improvement," deserve further consideration as our analysis found a significant effect in the intercept term for model D, which points to additional unexplained variation in the model.

Discussion

Principal Results

Researchers have an ethical obligation to provide information about the technologies used in digital health research, including related risks, benefits, and risk mitigation strategies. This information exchange within the context of health research currently occurs via the informed consent process. Standard recommendations for designing a consent communication encourage researchers to provide considerable details about all aspects of a study, which can yield a consent communication that is lengthy and dense.

Our study investigated the research question: How do factors related to the content of consent material (eg, readability, length, and lexical diversity) and background characteristics associated with prospective participants (eg, sex, age, physical activity, tracking physical activity) play into preferences for more and less information in an informed consent process? While in the aggregate participants in our study preferred the original IRB-approved consent language, our qualitative and quantitative analyses demonstrate that participants often missed information that seemed to matter to specific prospective participant subgroups for the study. Our analysis also found that the snippets

XSL•FO
RenderX

elicited many informed questions about the hypothetical research, underscoring the importance of partnering with prospective participants to plan and finalize study materials.

Partnering with prospective participants to identify how to improve the research informed consent process has led to understanding the study information they most need and want. For instance, Koh et al [5] found that research participants were underinformed and prioritized information about risks or discomforts and study procedures. Similarly, McCarty et al [14] found that the 2 concepts research participants struggled with understanding were the experimental aspects of the study and who was responsible for paying for research-related injuries.

In general, our analysis found that participants prefer brevity but need details that correspond with their background and experiences related to the study. For example, participants who were older preferred the more detailed original version of the consent, whereas participants who identified as being younger than 40 years of age preferred the modified version. These differences may relate to the perception of physical risks associated with study procedures. However, lower-risk study procedures, such as tracking your physical activity with a smartwatch, may not need as much detail, as participants who do not currently track their activity tended to prefer the modified version of the consent. These and other findings point toward different information to offer for each subgroup in the study population, by sex, race, and ethnicity, as well as physical health, and technology adoption.

Prospective participants could help to identify these gaps. A slim majority of participants shared feedback about the consent material (n=44, 55%). While participants by and large preferred the original, many saw value in the way that the content was communicated through the modified consent material and suggested that the two could be combined. Partnering with prospective participants in research could create opportunities for researchers to identify these gaps early in the study design process, by sharing ideas and asking prospective participants for feedback [10]. As participants in our study reflected on ways to improve the snippets, they shared feedback that included questions related to study procedures, expectations, and risks.

The informed consent process is a starting point for people to cultivate a trusting relationship with researchers. Regardless of whether they choose to participate in a study, trust is likely cultivated when people feel as though they can have the opportunity to fully grasp study materials and raise questions with the research team. As algorithmic decision-making plays an increasingly central role in digital health studies, cultivating trust among all the people involved in research becomes ever more important as, without trust and shared understanding, research quality will be compromised. Researchers should consider actively partnering with study participants so that together they can work through any interesting outcomes and misalignments that may emerge through the algorithmic decision-making in a study.

Recommended Practices for Working With Prospective Partners to Develop Trustworthy Consenting Processes

Many of the questions raised by participants focused on details not included in the original consent materials. By listening to prospective participants, research teams could identify barriers to participation that could be mitigated prior to recruitment [10]. While speculative, we contend that this could pay off in the long run-in terms of improved recruitment and retention based on increased confidence in the research team and understanding of what participants are committing to. In this way, the text snippets taken from the consent form offered a foundation for prospective participants to raise informed questions about the study, thereby enabling them to participate as guides of the research.

Researchers can take a few simple steps to incorporate participants as partners in guiding the study and the consenting process. Researchers can present prospective participants with a sample of the study consent materials to identify the unique needs and considerations of specific subgroups within the sample population. Our findings demonstrate that demographic background, susceptibility to possible risks, and study interest all play into a participant's need for more and less detail about a study. Similar findings have been reported in previous studies, which found that ethnicity, level of education, age, and gender were factors impacting participants' comprehension of informed consent information for genetics research, while factors like substance dependency were significantly more likely to increase the level of consent participants accepted [5,15,16].

By taking a 2-step approach to consent material development, digital health researchers can potentially identify these considerations and incorporate them to refine the study materials.

- Step 1: Broadly sample prospective participants for their perspective on snippets of the draft consent materials. At this stage, a sample of the prospective participants is presented with specific sections of the consent materials to comment on to elicit feedback on the language and any questions about the study details. Like our study design, the results should help to shed light on any differences based on demographic, interest-based, and other participant characteristics. Differences may include preferences for level of detail and methods of communication.
- Step 2: Adapt the consent materials for prospective participants. By surveying a sample of prospective participants about their preferences and questions about the consent materials, researchers will learn about any gaps and questions that need to be resolved in general, and for specific participant subgroups (eg, older adults, early technology adopters). This information can be adapted and integrated into broader consent processes for a study.

Our analysis shows how prospective participants in a study are not a monolith but reflect a diverse collection of subgroups that require specific considerations when supporting their learning about a study. Specifically, sex, age, level of physical activity, and ethnicity all played into preferences for more detail in the consent material. Researchers and IRBs should strive to identify and promote these considerations with research sponsors, to

XSL•FO
RenderX

encourage the research funding necessary so that all participants may feel informed through research consent processes. The findings point toward new ways of supporting prospective participants through their process of making an informed decision about whether to participate in a research study. While these recommendations encourage more work on the part of researchers, we contend that the benefits have the real potential to outweigh the added time in the form of improving trustworthiness between researchers and participants. This has the likelihood of producing several benefits overall including improved recruitment, retention, and, likely most critically, the quality of science being conducted as prospective partners can help researchers to understand implicit assumptions of risks, questions, and opportunities that they may not have otherwise been aware of.

Limitations

There are several limitations associated with this study. First, the research involved just 1 digital health study context, which was a low-risk smartwatch intervention to boost physical activity over a 1-year period. Future research should vary the study context to explore how higher risk and more complex study procedures, data management protocols, and digital health systems might play into prospective participant preferences for consent material. In particular, the use of artificial intelligence in health research is still novel and the implications are not fully understood. Researchers should explore best practices for communicating the details of a study involving such novel technologies.

Second, the study involved a survey-based approach to investigate participant perspectives on 31 text-based comparisons related to different aspects of the consent material. In some cases, the points of comparison were more similar than different, in other cases the opposite was true. In some cases, participants clearly agreed that more details were necessary on specific topics. Future research should explore various ways of communicating in more detail on specific topics, such as study risk. Rather than applying a survey-based approach, future research should aim to understand how people process this information by applying approaches like a cognitive interview, in which participants are asked to talk through their thought processes as they review the material. Rather than present 2 points of comparison, researchers should evaluate a wider range of options that include text and non-text-based methods of communicating consent materials [10]. Finally, readability is an important consideration when developing materials, but as demonstrated in this study the measures of readability selected for the analysis did not meaningfully impact prospective participant preferences related to the consent material. There may be other measures of the consent content that would provide meaningful ways of identifying content that is readable, useful, accessible, considerate, and so on. Our own analysis highlights how consent content can elicit informed questions, which may be a future measure of consent communication effectiveness.

Conclusions

Readability has been a primary metric for evaluating the effectiveness of consent communication materials; yet, our study demonstrates that there are other considerations for researchers to weigh. Our analysis highlights that people prefer short and simple text descriptions of study details, but also want answers to questions that relate to their personal circumstances. This finding is based on the breadth of personal questions elicited through the study as well as quantitative analyses demonstrating that factors related to participant sex, age, level of physical activity, and ethnicity all played into preferences for more detail in the consent material. The paper presents a 2-step approach that researchers can take to investigate how such factors may play into consent communication design for a digital health research study. Future research should explore ways to incorporate people and communities into the development of consent communication, to promote understanding and partnership in science.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Snippets of the original and modified versions of the consent material presented to participants.
[PDF File (Adobe PDF File), 197 KB - jmir_v27i1e63349_app1.pdf]

References

1. Emanuel EJ, Boyle CW. Assessment of length and readability of informed consent documents for COVID-19 vaccine trials. JAMA Netw Open 2021;4(4):e2110843 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.10843] [Medline: 33909052]
2. Hadden KB, Prince LY, Moore TD, James LP, Holland JR, Trudeau CR. Improving readability of informed consents for research at an academic medical institution. J Clin Transl Sci 2017;1(6):361-365 [FREE Full text] [doi: 10.1017/cts.2017.312] [Medline: 29707258]
3. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. N Engl J Med 2003;348(8):721-726. [doi: 10.1056/NEJMsa021212] [Medline: 12594317]
4. Glaser J, Nouri S, Fernandez A, Sudore RL, Schillinger D, Klein-Fedyshin M, et al. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: an updated systematic review. Med Decis Making 2020;40(2):119-143 [FREE Full text] [doi: 10.1177/0272989X19896348] [Medline: 31948345]

XSL•FO
RenderX

5. Koh J, Goh E, Yu KS, Cho B, Yang JH. Discrepancy between participants' understanding and desire to know in informed consent: Are they informed about what they really want to know? J Med Ethics 2012;38(2):102-106. [doi: 10.1136/jme.2010.040972] [Medline: 21708828]
6. Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. BMC Med Ethics 2013;14:28 [FREE Full text] [doi: 10.1186/1472-6939-14-28] [Medline: 23879694]
7. Luger E, Moran S, Rodden T. Consent for all: revealing the hidden complexity of terms and conditions. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; April 27, 2013-May 2, 2013; Paris, France. [doi: 10.1145/2470654.2481371]
8. Nebeker C, Gholami M, Kareem D, Kim E. Applying a digital health checklist and readability tools to improve informed consent for digital health research. Front Digit Health 2021;3:690901 [FREE Full text] [doi: 10.3389/fdgth.2021.690901] [Medline: 34713167]
9. Coleman E, O'Sullivan L, Crowley R, Hanbidge M, Driver S, Kroll T, et al. Preparing accessible and understandable clinical research participant information leaflets and consent forms: a set of guidelines from an expert consensus conference. Res Involv Engagem 2021;7(1):31 [FREE Full text] [doi: 10.1186/s40900-021-00265-2] [Medline: 34006326]
10. McInnis BJ, Pindus R, Kareem D, Gamboa S, Nebeker C. Exploring the future of informed consent: applying a service design approach. Proc ACM Hum Comput Interact 2024;8:1-31. [doi: 10.1145/3637330]
11. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):202. [doi: 10.1093/abm/16.3.199]
12. El Mistiri M, Khan O, Rivera DE, Hekler E. System identification and hybrid model predictive control in personalized mHealth interventions for physical activity. Proc Am Control Conf 2023;2023:2240-2245 [FREE Full text] [doi: 10.23919/acc55779.2023.10156652] [Medline: 37426035]
13. NIH. Request for information: developing consent language for research using digital health technologies. National Institute of Health. 2023. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-24-002.html [accessed 2024-04-01]
14. McCarty CA, Nair A, Austin DM, Giampietro PF. Informed consent and subject motivation to participate in a large, population-based genomics study: the Marshfield Clinic Personalized Medicine Research Project. Community Genet 2007;10(1):2-9 [FREE Full text] [doi: 10.1159/000096274] [Medline: 17167244]
15. Ray B, Jackson C, Ducat E, Ho A, Hamon S, Kreek MJ. Effect of ethnicity, gender and drug use history on achieving high rates of affirmative informed consent for genetics research: impact of sharing with a national repository. J Med Ethics 2011;37(6):374-379 [FREE Full text] [doi: 10.1136/jme.2010.037671] [Medline: 21266386]
16. Robinson JO, Slashinski MJ, Wang T, Hilsenbeck SG, McGuire AL. Participants' recall and understanding of genomic research and large-scale data sharing. J Empir Res Hum Res Ethics 2013;8(4):42-52 [FREE Full text] [doi: 10.1525/jer.2013.8.4.42] [Medline: 24169421]

Abbreviations

IRB: institutional review board
NIH: National Institutes of Health
OR: odds ratio

Edited by A Mavragani; submitted 17.06.24; peer-reviewed by P Borry, T Church; comments to author 05.09.24; revised version received 23.09.24; accepted 13.11.24; published 23.01.25.

Please cite as:
McInnis BJ, Pindus R, Kareem D, Vital DG, Hekler EB, Nebeker C
Factors Influencing Informed Consent Preferences in Digital Health Research: Survey Study of Prospective Participants
J Med Internet Res 2025;27:e63349
URL: https://www.jmir.org/2025/1/e63349
doi:10.2196/63349
PMID:39847412

©Brian J McInnis, Ramona Pindus, Daniah Kareem, Daniela G Vital, Eric B Hekler, Camille Nebeker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Finding Consensus on Trust in AI in Health Care: Recommendations From a Panel of International Experts

Georg Starke[1,2], MD, PhD; Felix Gille[3,4], PhD; Alberto Termine[1,2,5], PhD; Yves Saint James Aquino[6], MD, PhD; Ricardo Chavarriaga[7], PhD; Andrea Ferrario[8], PhD; Janna Hastings[4,9], PhD; Karin Jongsma[10], PhD; Philipp Kellmeyer[11,12], MD; Bogdan Kulynych[13], PhD; Emily Postan[14], PhD; Elise Racine[15,16], MPA, MSc; Derya Sahin[17], MD; Paulina Tomaszewska[18], MSc; Karina Vold[19,20], PhD; Jamie Webb[21], PhD; Alessandro Facchini[5], PhD; Marcello Ienca[1,2], PhD

[1]Institute for History and Ethics of Medicine, Technical University of Munich, Munich, Germany

[2]College of Humanities, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[3]Digital Society Initiative, University of Zurich, Zurich, Switzerland

[4]Institute for Implementation Science in Health Care, Faculty of Medicine, University of Zurich, Zurich, Switzerland

[5]Dalle Molle Institute for Artificial Intelligence (IDSIA), The University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Lugano, Switzerland

[6]Australian Centre for Health Engagement, Evidence and Values, University of Wollongong, Wollongong, Australia

[7]Centre for Artificial Intelligence, Zurich University of Applied Sciences (ZHAW), Zurich, Switzerland

[8]Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

[9]School of Medicine, University of St. Gallen, St. Gallen, Switzerland

[10]Bioethics & Health Humanities, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[11]Data and Web Science Group, School of Business Informatics and Mathematics, University of Mannheim, Mannheim, Germany

[12]Department of Neurosurgery, University of Freiburg - Medical Center, Freiburg im Breisgau, Germany

[13]Lausanne University Hospital (CHUV), Lausanne, Switzerland

[14]Edinburgh Law School, University of Edinburgh, Edinburgh, United Kingdom

[15]The Ethox Centre and Wellcome Centre for Ethics and Humanities, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

[16]The Institute for Ethics in AI, Faculty of Philosophy, University of Oxford, Oxford, United Kingdom

[17]Development Economics (DEC), World Bank Group, Washington, DC, United States

[18]Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

[19]Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, ON, Canada

[20]Schwartz Reisman Institute for Technology and Society, University of Toronto, Toronto, ON, Canada

[21]The Centre for Technomoral Futures, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Georg Starke, MD, PhD
Institute for History and Ethics of Medicine
Technical University of Munich
Ismaninger Str. 22
Munich, 81675
Germany
Phone: 49 8941404041
Email: georg.starke@tum.de

Abstract

Background: The integration of artificial intelligence (AI) into health care has become a crucial element in the digital transformation of health systems worldwide. Despite the potential benefits across diverse medical domains, a significant barrier to the successful adoption of AI systems in health care applications remains the prevailing low user trust in these technologies. Crucially, this challenge is exacerbated by the lack of consensus among experts from different disciplines on the definition of trust in AI within the health care sector.

XSL•FO
RenderX

Objective: We aimed to provide the first consensus-based analysis of trust in AI in health care based on an interdisciplinary panel of experts from different domains. Our findings can be used to address the problem of defining trust in AI in health care applications, fostering the discussion of concrete real-world health care scenarios in which humans interact with AI systems explicitly.

Methods: We used a combination of framework analysis and a 3-step consensus process involving 18 international experts from the fields of computer science, medicine, philosophy of technology, ethics, and social sciences. Our process consisted of a synchronous phase during an expert workshop where we discussed the notion of trust in AI in health care applications, defined an initial framework of important elements of trust to guide our analysis, and agreed on 5 case studies. This was followed by a 2-step iterative, asynchronous process in which the authors further developed, discussed, and refined notions of trust with respect to these specific cases.

Results: Our consensus process identified key contextual factors of trust, namely, an AI system's environment, the actors involved, and framing factors, and analyzed causes and effects of trust in AI in health care. Our findings revealed that certain factors were applicable across all discussed cases yet also pointed to the need for a fine-grained, multidisciplinary analysis bridging human-centered and technology-centered approaches. While regulatory boundaries and technological design features are critical to successful AI implementation in health care, ultimately, communication and positive lived experiences with AI systems will be at the forefront of user trust. Our expert consensus allowed us to formulate concrete recommendations for future research on trust in AI in health care applications.

Conclusions: This paper advocates for a more refined and nuanced conceptual understanding of trust in the context of AI in health care. By synthesizing insights into commonalities and differences among specific case studies, this paper establishes a foundational basis for future debates and discussions on trusting AI in health care.

(J Med Internet Res 2025;27:e56306) doi:10.2196/56306

KEYWORDS

expert consensus; trust; artificial intelligence; clinical decision support; assistive technologies; public health surveillance; framework analysis

Introduction

Background

The integration of artificial intelligence (AI) systems into health care is one of the most widely anticipated transformations of health systems worldwide [1]. AI promises improved diagnostics [2,3], optimized treatment strategies [4], and early identification of at-risk patients [5]. Prominent examples also include AI for assistive technologies offered to patients directly [6,7] and for informing public health decision-making beyond the individual [8]. More recently, large language model–based applications promise to revolutionize health care, with applications spanning from clinical research and processes to physician-patient relations [9]. Despite this range of potentially beneficial applications, the broader adoption of AI systems in health care has been struggling due to many inhibiting factors. Problems arise at the level of development, with data bottlenecks impeding the training of machine learning models or a lack of user-centered and value-sensitive design procedures affecting AI acceptability [10,11], and stretch to the level of practical implementation of AI systems [12-14]. Questionable improvements in real-life health care settings [15], a lack of regulatory frameworks [16], and unresolved questions of reimbursement [17] further complicate adoption of AI in health care.

In this paper, we focus on a central inhibitor of successful AI adoption in health care, namely, the low levels of user trust in AI systems [18]. Understanding and fostering trust in AI remains challenging, not only practically but also conceptually. Interpersonal trust constitutes a complex and contested construct in philosophy and social sciences [19-24]. Moreover, accounts diverge on what constitutes being worthy of trust, namely, on the definition of *trustworthiness* [19,20]. Unsurprisingly, there is also no consensus across disciplines such as computer science, philosophy of technology, and social sciences on the definition of trust in AI and the capabilities that a trustworthy AI should maintain [25,26]. As a recent publication assembling existing empirical literature on trust in AI and clinical decision support put it, "Different groups [of researchers], whilst seemingly agreeing in principle that 'trust' and 'trustworthiness' are important, can in fact be referring to very different concepts and talking past one another" [27].

The combination of fragmented conceptual research, practical concerns, and implementation difficulties inhibits fostering warranted trust in AI (ie, trust caused by the trustworthiness of the AI) [25,28], which will be increasingly crucial for delivering health care [29]. As recent empirical work using path analysis has highlighted, trust in AI-based technologies seems to have a significant effect on users' intention to use such systems [30]. However, to build trust in—possibly trustworthy—AI systems in health care applications, we first need a conceptual understanding of trust in AI within the health care sector. This goal requires considering the diverse perspectives, expectations, and limitations that various research disciplines bring to the table in the context of human-AI interactions. Without conceptual clarity, attempts to foster the multifaceted concept of trust in AI—whatever it may *precisely* mean—run the risk of being inefficient or even detrimental, leaving users, including data scientists, physicians, and patients, vulnerable to placing trust in systems that may not warrant it or refraining from it in

XSL•FO
RenderX

situations in which grounds for trusting actually exist, an occurrence of what we might call "unwarranted distrust" [31,32].

Objectives

Bringing together international and multidisciplinary perspectives on the topic of trust in AI, this paper aims to provide common ground for defining trust in AI in health care as a reference point for future debates. To this end, we developed the first consensus statement on trust in AI in health care based on input from international experts on the topic drawing on iterative synchronous and asynchronous discussions of realistic case studies. Our results highlight the need for a more refined and nuanced understanding of trust in the context of AI in health care if the concept is to guide AI design and adoption processes and inform national and international governance.

Methods

Overview

Consensus statements from meetings of experts working on a particular topic constitute a common scientific approach across the fields of medicine [33,34] and ethics [35,36]. While they can take many different forms [37], their shared goal is to identify agreement among people working in a field and assemble multiple perspectives with peer-informed legitimacy [38]. The reporting of our findings follows the Accurate Consensus Reporting Document guidelines [37] and considered applicable elements of guidelines for reporting qualitative research [39,40].

Recruitment

Recruitment for the consensus process took place as part of a 3-day workshop held at the École Polytechnique Fédérale de Lausanne in September 2022 in Lausanne, Switzerland. Authors GS, FG, AT, A Facchini, and MI, who organized the workshop, invited some participants directly based on their contribution to the field and to accommodate different types of expertise and backgrounds. They selected further participants from submissions to a call for abstracts for the workshop. Unanimous agreement concerning participants was achieved among conference organizers GS, FG, AT, A Facchini, and MI.

Consensus Process

Overview

The consensus process comprised 3 sequential steps involving both synchronous (step 1) and asynchronous (steps 2 and 3) activities. We describe them in the following sections. Figure 1 shows the steps of the consensus process and provides an overview of the intended outputs of each step.

Figure 1. The 3-step consensus process and envisioned results.

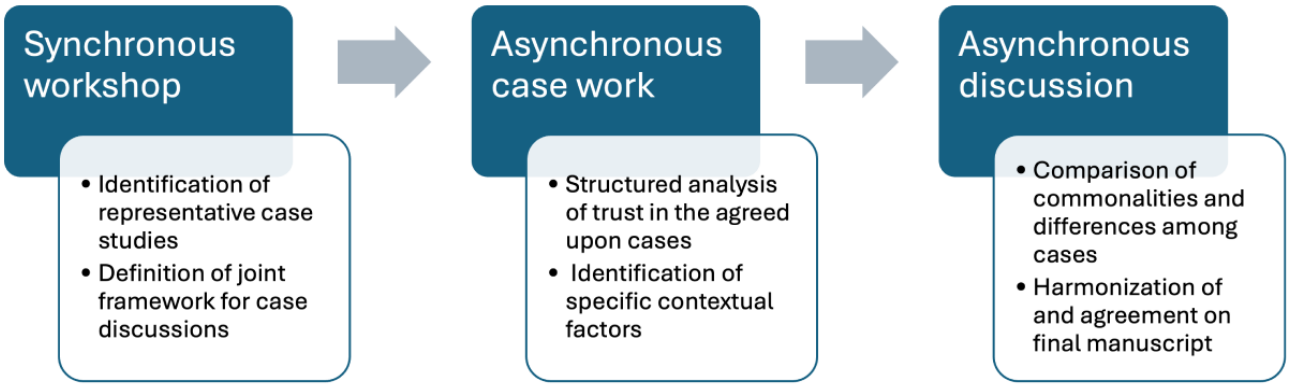

Synchronous Phase: Selection of Case Studies and Framework

We conducted a workshop involving an international group of researchers from various disciplines. To arrive at a consensus regarding trust in AI in health care, we discussed the topic from complementary angles in a series of group activities that iteratively involved talks on specific aspects of trust, moderated group discussions, plenary sessions, and interactive panels. The moderated group discussions were spread over 3 days. They covered the topics of (1) conceptualizing trust in the context of human-AI interactions, (2) conceptualizing trust in AI in the context of health care specifically, (3) requirements for trustworthy AI in health care, and (4) recommendations for implementing requirements for trustworthy AI in health care. To achieve consensus on the case studies, all participants reviewed them. Only case studies assessed as representative yet distinct from other cases by all participants were promoted to the next step.

To guide our analysis, we made use of a conceptual framework. The framework was proposed by FG, critically discussed, and agreed upon by all authors. A conceptual framework represents a phenomenon in the form of a network with interlinked themes that describe how the phenomenon works [41]. Conceptual framework analysis provides an established qualitative method to guide the identification and organization of key aspects and relationships in contexts that draw on multidisciplinary bodies of knowledge [41]. We used framework analysis to provide structure and coherence of discussions for the asynchronous steps, systematically evaluating various dimensions of trust in AI in health care.

Asynchronous Step: Developing Case Studies on Trust in AI in Health Care

To understand how trust unfolds in the context of the agreed upon 5 cases, in the second step, smaller groups of authors expanded on the individual cases agreed upon in step 1. All contributors involved in the writing are listed as authors.

XSL•FO
RenderX

Participants in this step were asked to explicate trust-relevant factors of the cases in accordance with the agreed upon conceptual framework. Responsibilities for coordinating the writing of individual case studies were distributed among the authors, with 1 person taking responsibility for leading each case study. KJ led the consensus for the first case study, JH led the consensus for the second case study, YSJA led the consensus for the third case study, GS led the consensus for the fourth case study, and JW led the consensus for the fifth case study. The results were compiled and shared by GS, FG, and MI.

Asynchronous Step: Comparison and Discussion of Case Studies

Having combined the individual case studies, we compared and discussed relevant aspects to synthesize our findings. We discussed our findings critically in light of existing literature and asked participants again to provide written feedback on the synthesized findings. This process was coordinated by GS, who wrote a first draft of the manuscript, which was then refined by FG, A Facchini, A Ferrario, and MI. All authors provided their feedback on the manuscript, which was incorporated and harmonized by GS and shared with all authors for approval.

Ethical Considerations

In accordance with the existing regulatory framework in Switzerland, such as the Human Research Act and the local rules at the École Polytechnique Fédérale de Lausanne as hosting institution, no ethics approval was required for this collaborative research project involving only consenting experts in the field.

Results

Overview

The results of our study unveil the complexities of understanding trust in AI within the health care domain. Our findings represent consensus among international experts guided by conceptual framework analysis. The following sections provide the characteristics of the involved experts and report the findings from the individual synchronous and asynchronous steps of our consensus process.

Recruitment

The characteristics of the 18 participants in the consensus process are described in Table 1.

Table 1. Participant characteristics—multiple entries possible for location and background (N=18).

Characteristics	Participants, n (%)
Gender	
Women	7 (39)
Men	11 (61)
Nonbinary	0 (0)
Prefer not to say	0 (0)
Geographic location	
Europe	16 (89)
North America	2 (11)
Oceania	1 (6)
Background	
Computer science	6 (33)
Ethics	5 (28)
Medicine	4 (22)
Philosophy	8 (44)
Public health	2 (11)
Social science	3 (17)

Consensus on Trust and Trustworthiness

Synchronous Phase: Selection of Case Studies and Framework

A total of 5 case studies emerged from the synchronous group discussion and were identified as paradigmatic examples of the various applications of AI in health care, namely, diagnosis, clinical risk assessment, public health surveillance, assistive technologies, and health care resource allocation. The involved experts agreed that these cases may not provide an exhaustive sample of all potential AI applications in health care yet allow for a meaningful representation and comparison of trust in AI across largely different health care settings.

To guide the asynchronous analysis of the individual cases, all involved experts identified and defined key components of a conceptual framework for trust in AI in health care. We developed this framework with the backdrop of existing conceptual work on trust in data use in health care, preliminary discussions of the shortcomings of present conceptual work describing what trust in AI is, and existing guidance on conceptual framework development for research and scale development in medicine [29,31,42-49].

XSL•FO
RenderX

Following existing conceptualizations of trust in the context of data use in health care [47], themes were grouped into two main areas: (1) the context of trust and (2) causes and effects of trusting. We considered trust as a context-specific, relational construct between 2 actors that is shaped by the environment in which it develops over time [50]. Therefore, our framework implemented the context specificity of trust by considering (1) the environment within which an AI system is deployed (ie, the health care application in which the AI operates), (2) the actors involved in the specific trust relationship, and (3) the factors that frame the trusting relationship. Here, framing factors influence "the process by which people develop a particular conceptualization of an issue or reorient their thinking about an issue" [51]. Examples of framing factors are historical context, cultural aspects, norms and values, fears, public sentiment, overarching belief systems, and religious beliefs. In addition to context, our framework includes factors that support or inhibit trust as well as the effects of trusting AI systems. Causally important themes for trusting relationships typically relate to factors that make AI trustworthy or untrustworthy in the eyes of those placing trust. Examples are the reliability and accuracy of the AI and, arguably, the level of interpretability of its outcomes [28,52]. In summary, our framework comprised 5 themes: environment, actors, frames, causes, and effects (Textbox 1).

Textbox 1. Conceptual framework for warranted trust in artificial intelligence (AI) in health care.

Context of trust in AI in health care

- Environment: What is the setting for which the AI system is intended?
- Actors: Who are the involved actors in the trust relationship?
- Frame: What frames trust in relation to AI?

Cause and effect of trust in AI in health care

- Cause: What makes AI trustworthy?
- Effect: What is the effect of trust in relation to AI?

Asynchronous Step: Developing Case Studies on Trust in AI in Health Care

Drawing on our framework and previous research, smaller groups of authors examined particular aspects of AI used for diagnostic purposes, clinical risk assessment, public health surveillance, assistive technologies, and health care resource allocation. The results of the individual groups led by KJ, JH, YSJA, GS, and JW, respectively, are reported in the following sections.

Diagnostic AI in the Clinic

Machine learning and, more specifically, deep learning have proven to be particularly suitable for computer vision, especially image processing and pattern recognition. Important preconditions to apply these techniques in medicine, such as suitable infrastructure or availability and storage of digital images, are also being increasingly met, at least in high-income countries. This may explain why most current and proposed AI systems in medicine are used to aid image-based diagnostics in fields such as radiology, ophthalmology, and pathology [53-55]. As a large part of the workload of, for instance, radiologists is to interpret medical images [56], diagnostic AI systems could increase this capacity; support or take over the tasks; and, thereby, change the organization of image-driven diagnostics [57].

Some studies have indicated that, under specific circumstances, AI systems achieve accuracy that is at least equal to that of expert radiologists and pathologists or even outperform them when detecting, classifying, and segmenting tumors in ultrasonography, x-ray imaging, magnetic resonance imaging scans, and digitized microscopy slides [58,59]. These findings have raised massive enthusiasm and fueled the motivation to develop and use algorithms in image-driven diagnostics. As a particular example, we will focus on a recently approved AI system for the diagnostic analysis of chest x-rays [60].

The system aids diagnosis by analyzing chest x-rays, the most frequent radiological examination worldwide. It does so by classifying the scans in nonpathological and potentially pathological cases. For the first group, the program provides a fully automated report, removing the necessity for any further follow-up on the image by radiologists. The scans of the second group are forwarded to trained physicians for further radiological analysis. A test of the software on approximately 10,000 chest x-rays from Finland showed a very high sensitivity (99.8%) with a low specificity (36.4%), yielding a very low probability of missing any critical findings [61].

However, the use of diagnostic AI is associated with technological obstacles as well as epistemic and ethical challenges [62]. Although the potential of self-supervised learning to generate expertlike annotations has been demonstrated, an insufficient number of expertly annotated diverse images is currently often a limiting factor in training AI models on a technical level [63]. Large datasets, for instance, in pathology, can also pose problems if downsampling leads to crucial information being lost, demanding different kinds of representation [64]. Beyond such technical difficulties, there are also more fundamental obstacles, including the fact that diagnostic tasks are not usually choices between 2 distinct outcomes [65]. Another concern is the frequent lack of explainability of medical AI, which limits physicians' ability to recheck the AI's output [66]. It has been argued that this raises particular challenges as decisions based on black-box AI systems are potentially not fully interpretable for the human physician. Moreover, when offering explanations, it is essential

XSL•FO
RenderX

to consider the epistemic requirements of users (eg, radiologists and patients). Explanations that lack a clear connection to clinical knowledge despite being technically accurate fall short in delivering the desired support for clinical practice. Similarly, patients who wish to understand the basis of their diagnosis may find explanations that diverge from their own illness experiences unsatisfactorily [67,68]. That said, there is empirical evidence that a lack of explainability is not necessarily perceived as a problem by physicians if a diagnostic system has been properly validated and other detailed information about the AI system is available [69,70].

The case discussed here takes place in a clinical environment of image-driven diagnostics. It directly involves medical professionals and the AI system, whereas patients do not interact with the system itself; in addition, the developing company and regulatory bodies are indirectly involved. Framing factors are discourses about job security in radiology and the danger of physicians being replaced by AI [71]. Trust leads to acceptance of the system by physicians, potentially at the cost of deskilling, and is arguably fostered by the accuracy of the diagnostic program in question [72].

Predictive AI for Clinical Risk Assessment

AI algorithms can be used to predict individual patient trajectories. Examples include the prediction of COVID-19 severity [73], the prediction of delirium [74], prognostic models of respiratory diseases [75], or systems predicting future lung cancer risk [76]. In addition, predictive models are crucial for determining individualized treatment plans, often particularly relevant in oncology [77]. The difference from the diagnostic use case is that the prognostic case concerns the future (ie, informs decision-making under situations characterized by an inherently larger uncertainty). The prognostic prediction informs choices of action, treatments offered, support, and resource allocation and, therefore, can, in turn, influence the very outcome that has been predicted [78].

As a specific case study, let us consider the prediction of circulatory failure in intensive care units (ICUs) [5]. In ICUs, a great number of machines monitor the state of patients and need to be observed constantly as patient conditions may worsen rapidly. The existing technologies for life support and vital sign monitoring provide information and alerts that clinicians need to oversee. However, the high rates of alerts, including false positives, may cause alert fatigue for practicing clinicians, impairing optimal care. In this context, systems that predict severe deterioration more accurately can obviate alert fatigue and potentially lead to better patient outcomes. The system reported in the study by Hyland et al [5] integrates information from multiple organ function–monitoring systems to alert clinicians to potential circulatory failure 8 hours in advance. Note that the information available to the system is necessarily incomplete, and it is impossible to predict future circulatory failure with 100% accuracy using any algorithm. However, the system in question successfully predicted 90% of cases of circulatory failure in the test set and 82% of them >2 hours before the event [5], so it would likely prove useful in a clinical setting.

The system described here can only be deployed in the environment of hospitals or clinics. Physicians, nurses, and potentially caregivers may interact directly with the AI system, and the social circle of a patient could also potentially receive information derived by the AI system, whereas the patients admitted to an ICU will likely not interact with the system itself. Critical contextual factors are the system's use in stressful situations with the risk of severe consequences, sometimes under time pressure, and the need to synthesize a multitude of complex information. Trust in such a situation can be built by overall improved outcomes in settings where the algorithm is used, ideally by investigating it through a randomized controlled trial. In addition, there should be reasons to believe that the AI system is not biased or will not lead to harmful consequences for any subgroups of the population and that it produces reliable individual decisions that are robust to design choices [79,80] or randomness [79,81] in the AI pipelines. Trust built on these premises will promote AI's acceptance and use. However, as with other medical interventions aiming to change the future, such as screening programs [82,83], trust could also lead to worse outcomes if the system does not actually deserve it (ie, if it is not *trustworthy*).

AI for Public Health Surveillance

Public health surveillance involves the identification of signs of population-level health anomalies and potential disease outbreaks from a heterogeneous collection of data sources [8]. With its intrinsically data-driven nature, public health surveillance has increasingly become a domain for AI applications as AI provides a range of novel methods for data collection and data analysis in large and varied samples [8]. A comprehensive survey of the literature showed that the COVID-19 pandemic highlighted the potential of AI systems in improving surveillance of infectious disease outbreaks [84].

Consider the example of EPIWATCH among the numerous AI-based surveillance systems that provide early signs of disease outbreaks [85]. Developed at the University of New South Wales in Australia, EPIWATCH is an open access web-based tool that offers an interactive dashboard with a sortable, searchable, and filterable global map based on 30 days of data [85]. The tool has been successfully used to model and identify patterns of disease outbreaks, providing crucial information on risk factors and geographic distribution.

For instance, using publicly available data, EPIWATCH has been used to trace global Zika virus outbreaks, tracing transmission modes, affected countries, and complications such as microcephaly [86]. Similarly, a different group of authors used EPIWATCH to model the global epidemiology of hepatitis A, identifying the United States and Europe as major centers of outbreaks, and provided quantitative global data on the most common risk factors, which seem to be homelessness and foodborne outbreaks [87]. Such data can be helpful to inform policy making at both a global and local level.

Let us summarize the contextual factors of this case. Contrary to the previous examples, AI systems for public health are deployed in nonclinical settings. This web-based example is potentially accessible by anyone with a computer or mobile phone with an internet connection interested in data analysis to

XSL•FO
RenderX

predict disease outbreaks. Therefore, relevant actors include the developers as well as a broad spectrum of potential users, ranging from public health practitioners and policy makers to the public. Important framing factors for trust are the severity of the modeled disease as well as the stage of the epidemic or pandemic, impacting the relevance and acceptability of the tool. As the COVID-19 pandemic has demonstrated, antiscience sentiments and conspiracy theories contribute to how public health tools are viewed. Finally, usability aspects further influence whether information will be taken up by practitioners. Trust in the application will likely lead to increased uptake among decision makers. The trustworthiness of the platform developers and endorsement by authorities, such as public health experts and policy makers, will further determine whether the wider public will trust the model's output.

AI for Assistive Neurotechnology

In comparison to the examples discussed previously, a quite different trust relationship can be found in the case of assistive neurotechnologies based on brain-computer interfaces (BCIs). BCIs can, for instance, be used for neurorehabilitation, restoring lost function, or augmenting and enhancing existing cognitive or motor abilities [7]. AI, especially data-driven machine learning with deep neural networks, plays a crucial role in their development and use as these methods enable and facilitate the modeling and decoding of complex neural signals as well as the adaptation to individual users [88]. To delineate differences in the trust relationship between AI-based neurotechnology and other medical AI, we focus on a specific example: a BCI-guided robotic hand orthosis for rehabilitative purposes.

Orthoses are external wearable devices that can be "used to compensate for impairments of the structure and function of the neuromuscular and skeletal systems" [89]. In line with numerous existing research projects at the level of preclinical prototypes [90], let us consider a specific BCI-based robotic orthosis that allows patients with stroke to restore some volitional control of hand-grasping movements. To do so, the system records electrical activity of the dominant motor cortex noninvasively through an electroencephalography cap and extracts features from the measured signals using machine learning–based classifiers to extract patient-specific neural markers of intended hand movements, which are then translated into mechanical movements of the robotic hand orthosis.

There are several aspects of such personalized AI-based assistive neurotechnologies for neurorehabilitation that set them apart from medical AI used for diagnosis, prediction, or public health surveillance. One crucial difference that stands out is that the AI in question is *embodied* in the sense that the AI used for decoding neural signals and for translating them to mechanical movement is inextricably linked to a physical object—namely, the hand orthosis. This implies that physical design aspects play a crucial role on whether trust is expedited to the device [91].

However, there are further differences that impact the trust relationship involving actors, the environment, and framing. First, assistive rehabilitative neurotechnology is designed for patients with impairment of cognitive or motor functions who need to engage with the technology. These patients are also the main trusters of this technology, not health care professionals or public health experts. Second, suppose the neurotechnology also helps patients in their everyday life. In that case, it should not only work under highly constrained and controlled laboratory or clinical conditions but also in a patient's home. This implies additional demands for the devices' ease of use and their security in a more open environment, which would be crucial for the trust of bystanders in these devices. Third, due to the BCIs' physicality, trust in assistive neurotechnology may also be framed differently, invoking widely known images from science fiction [92]. Fourth, users may be concerned that decoding their brain activity could be used not only to control the robotic orthosis but also to infer other types of information that would not be available using other sensing technologies, thus raising potential threats to their privacy. Therefore, addressing potential fears of patients and designing AI that assures human control (in the sense of "human in the loop") at any moment adds to the complexities of building trust in such embodied AI.

To summarize, the assistive AI system discussed here is embedded in an environment of clinical neurorehabilitation. Its trusters are primarily patients and rehabilitation experts who trust in the developing engineers and the pertinent regulatory bodies. In the context of embodied AI, many specific framing factors influence its perception, especially from science fiction literature and cinema. It is also shaped by public attitudes and policies on related technologies such as robotics. To build trust in this environment and prove the accuracy of the AI, a lack of conflicts of interest on the part of the developers; sufficient understanding of the underlying technology by its users and rehabilitation experts; and independent, long-term technical support seem key. Successful trust building will result in wider acceptance and adherence by users, as well as potential inclusion of the technology in public health programs, facilitated regulatory compliance, and improved insurance reimbursement. Taken together, these elements can facilitate access and increase the technology's affordability by expanding the number of its users.

AI for Health Care Resource Allocation

A 2019 study in *Science* revealed that software provided by the health service company Optum, and which was being used to manage the care of >200 million patients at hospital centers across the United States, was significantly—even if unintentionally—biased against Black patients [93]. The machine learning algorithm aimed to predict the future health care needs of patients and direct extra medical care toward the most vulnerable. However, it was shown to systematically underestimate the needs of Black patients. It did this because it used health care costs as a proxy for need in its risk score despite health care costs not being a neutral and reliable proxy for health needs in this context. This led it to assign consistently lower risk scores to Black patients compared to White patients who were equivalently sick. This was despite the model being "race blind" in the sense that race was not specified in the input data [94].

When considering how this case study relates to trust in and trustworthiness of AI in health care, it is important to note that this is not a case of AI being introduced into a dyadic

XSL•FO
RenderX

physician-patient relationship but, rather, of its integration into the operations of complex health systems involving many actors and institutions. Therefore, philosophical accounts of trust that focus on the properties of individuals—these accounts traditionally comprise reliability *plus* some appropriate motivational state on the part of the trustee toward the truster, for example, goodwill [95]—are insufficient to capture the conditions necessary for trustworthy AI in these contexts.

Accounts of institutional trustworthiness are more useful here. They often emphasize features such as transparency as well as competence and reliability [96]. Accounts of trustworthiness that do not necessarily require a phenomenological state such as goodwill may also be useful. For example, the trust responsiveness account by McGeer and Pettit [97] emphasizes the importance of an agent—or, in our case, institutions—responding appropriately to the reasons for doing what they are being relied upon to do. Alternatively, one might provide a deflationary account that either equates trustworthiness with reliability [48] or rejects trust as an appropriate attitude altogether [98].

What does it mean to meet these conditions in this and similar cases of integrating machine learning into resource allocation decisions across an entire hospital or health system? Trust responsiveness demands that systems such as the one provided by Optum be introduced to promote health and well-being and improve patient outcomes and not to meet other goals such as cutting costs to maximize profits for insurance companies and health systems. Using a system such as Optum may not necessarily require internal algorithmic transparency, which is often the focus of work considering the ethics of machine learning algorithms in health care [99], yet may be difficult to translate into a genuine understanding of the AI system by clinicians or patients. However, there needs to be openness regarding design decisions and value choices involved in the algorithm's construction and implementation [100]. Demonstrating these may require engaging in ethical and algorithmic impact assessments during the design process [101,102], welcoming algorithmic auditing after implementation [49], and abiding by regulatory frameworks throughout the product life cycle [103]. Ideally, transparency for AI systems for resource allocation would highlight both global explanations [104], given that the goal of the algorithm is to distribute resources across a population [105], and local explanations for individual decisions as the biased program impacts individual care. Both are potentially useful for promoting fairness [104].

In addition, it is crucial to consider contextual factors impacting such trust relationships. The program was developed by a for-profit health service provider in the US health care system and used to determine resource allocation. Key actors were the health service company itself, which provided the algorithm; the health systems; clinicians interacting with the algorithm's recommendations; patients having their care influenced by the algorithm's recommendations; and, finally, regulatory bodies and (internal and external) algorithmic auditors. Important framing factors include media reporting on algorithms and their impact and, on a conceptual level, theories of institutional trust. Such trust, though unwarranted here, would result in the model's acceptance and use. It is supported by an AI's reliability and competence as well as by the transparency of the developers concerning design choices and the values reflected in them.

Asynchronous Step: Comparison and Discussion of Case Studies

As the 5 cases highlight, medical AI is a multifaceted concept encompassing a diverse group of systems, environments, actors, and framing factors. Talking about trust in medical AI in general can only be a first approximation to the phenomenon. However, some general elements of trust were accepted by all experts. First, they agreed that trust provides a way to deal with complex situations in the face of uncertainty [106,107] and is established in anticipation of a beneficial outcome. Therefore, in a rather minimal fashion, and in line with existing literature, an agreed upon definition of trust in AI in health care focuses on giving discretionary power to an AI system with respect to a specific health care–related task [29,108]. At the same time, it was agreed that situations of trust were characterized by risks, rendering the trusting agent vulnerable to betrayal of trust [21]. To minimize such risks, potentially of life and death in health care, experts agreed that we need to only promote warranted trust in medical AI (ie, trust that is justified, plausible, and well grounded [19,28,49]). In the literature as well as in regulatory contexts, AI systems that warrant such trust have been described as trustworthy [49,109,110]. Therefore, in the eyes of the experts, a complete assessment of trust in AI includes epistemological and practical considerations of trust as much as examining the trustworthiness of the AI in question (ie, the functionalities of an AI system that morally and practically vindicate a relationship of trust [111]).

The involved experts agreed on many practical shortfalls of current AI in health care potentially impacting its trustworthiness. These include the lack of explainability of many AI algorithms [28,66,112]; the difficulties of benchmarking model performance [113]; or the longtail of exceptional situations that will be difficult to anticipate at the development stage, from nonstandard data inputs to the system's deployment in a health care system for which it was not originally trained [114]. In the eyes of the experts, anticipatory difficulty also relates to risks of AI perpetuating, hiding, or reinforcing statistical and social biases as well as other data problems [115,116]. Finally, it was stressed that many medical AI tools so far lack proper engagement with stakeholders in their development, which is crucial for ensuring that they adequately address a real clinical problem and identifying their relevant ethical implications [117].

When looking at the individual case studies, commonalities and differences among them arose at micro and macro levels. Table 2 provides an overview of the contextual factors of each case study.

XSL•FO
RenderX

Table 2. Contextual factors for the 5 case studies.

	Environment	Actors	Framing factors	Causes of trust	Effects of trust
Diagnostic AI[a] (chest x-rays)	Image-driven diagnostics (radiology)	Medical professionals and AI system; patients to a limited extent	Discourse regarding job security and potential AI replacement	Accuracy, design transparency, and human competencies and virtues	Acceptance of systems by physicians, potentially at the cost of deskilling
Predictive AI (ICU[b] setting)	Clinical setting of an ICU	Physicians, nurses, and AI system; patients to a limited extent; and potentially caregivers and family members	Stressful situations potentially with a need to act under time pressure, risk of severe consequences, the need to synthesize too much information, and alert fatigue	Accuracy, transparency, and explainability; fairness; exclusion of harm; and rigorous testing (eg, in the form of an RCT[c])	Acceptance and use of the system, potentially at the risk of erroneous clinical decisions following misleading predictions
Public health AI (disease outbreak model)	Nonclinical setting—publicly accessible web-based tool for the analysis of heterogeneous data	Developers, public health practitioners, policy makers, and the public	Stage and severity of disease outbreak; usability aspects (eg, intuitive interface or data visualization); and, potentially, antiscience sentiments and conspiracy theories with regard to the disease and health service providers	Historical accuracy and endorsement by authorities	Acceptance and use of the system by public decision makers (public health experts and policy makers)
Assistive AI (neurorehabilitation)	Clinical neurorehabilitation; elective use of different technologies for different activities, potentially every day	Patients and their caregivers and social circle, potentially including employers, engineers, and regulators	Clinical setting, science fiction literature and cinema, and public attitudes and policies on related technologies	Accuracy, privacy, lack of conflicts of interest, independence, long-term technical support, and user understanding of the underlying technology	Technology acceptance by users, health care professionals, and health care providers; potentially facilitated reimbursement and increased affordability and accessibility
Resource-allocating AI (predicting costs and needs)	Health service providers and health care system	The developing company providing the algorithm, the health system implementing it, the clinicians interacting with it, the patients having their care influenced by the algorithm, and regulatory bodies and algorithmic auditors	Media reporting on algorithms and their impact and theories of institutional trust	Reliability, accuracy, transparency, design, and model-centric explanations	Acceptance and use in health care systems

[a]AI: artificial intelligence.

[b]ICU: intensive care unit.

[c]RCT: randomized controlled trial.

Most commonalities among the different cases can be found with regard to their causes and effects, some of which arise from the very structure of trust. A common effect of trust was found in it leading to wider acceptance of the technology in question. Such acceptance implies an increased uptake by its intended end users, resulting in a larger role in health care systems, facilitating access, and potentially increasing the systems' affordability. In line with the literature, common aspects with regard to trust-supporting features included both aspects intrinsic to the system, such as transparency and explicability, and extrinsic factors, such as proper external validation and assessment of potential biases [25,29].

Across the discussed cases, some forms of transparency were identified as supportive of establishing trust. Such transparency included openness by the developers about their measures of a well-working system, enabling an alignment with the goals and values of patients and clinicians. Promoting transparency was also assessed as a good way of promoting trust among clinicians using the systems. For instance, there is some empirical evidence that being transparent about the functioning of a triage system makes clinicians more likely to trust it [118]. The extent to which this is possible depends on the exact nature of the system, with clear differences among algorithms that are interpretable by design [119], algorithms that can be explained in terms of their exact functioning on the level of predictors and their weights, and algorithms whose decision process cannot be interpreted causally.

Informational openness was also recognized as central to promoting a system's external trustworthiness (eg, if a system's reliability and performance metrics are publicly available). From the discussed cases, it also emerged that rendering information accessible to all relevant auditing bodies is crucial for mechanisms of accountability and, in doing so, for creating the necessary epistemic basis for trust at an individual level as much as among the public, who can have confidence that the AI systems have been subject to meaningful scrutiny. In this sense,

XSL•FO
RenderX

transparency was deemed to support trust among affected communities and stakeholders by providing the means to evaluate claims made about these systems with regard to bias and discrimination. AI systems reproducing inequalities and exacerbating injustices may, in turn, erode public trust not only in the AI-based tools themselves but also in stakeholders involved in building, deploying, and using such systems. Addressing and testing for such biases was deemed particularly relevant considering that "personal and collective experiences with discrimination or degradation-along lines of race, class, gender, or other personal characteristics especially create reasons for suspicion if not outright distrust" [99]. However, as the experts agreed, such testing needs to go beyond the sole measure of excluding apparently discriminatory information. Deep learning models may predict a patient's race from medical images such as chest and hand x-rays and mammograms [100] or identify patient self-reported race from redacted clinical notes [101] despite human experts being unable to do the same. Therefore, as indicated by the resource allocation case study, discrimination can occur even when models are apparently *race blind* through implicit proxy features, as is well understood and documented in algorithmic fairness research [120]. Moreover, discrimination through differential impact across subpopulations might occur precisely because models are insufficiently sensitive to inequitable distribution of social determinants of health along ethnic or racial lines [121].

Beyond these common themes, our cases highlight important contextual differences among trust in different kinds of medical AI. In particular, these relate to (1) framing factors, (2) previous knowledge of the targeted trusters demanding different levels of explicability and transparency, and (3) the different risk-benefit trade-offs in different environments. Sometimes, the risks in medicine are grave for individual patients, such as in an ICU setting. However, in the resource allocation setting, the risk-benefit trade-offs concern both individual-level consequences of allocating or denying a resource to an individual and group- or population-wide benefits or harms due to better or worse resource use.

Discussion

Principal Findings

Using conceptual framework analysis, our study provides the first consensus-based publication investigating trust in AI in the domain of health care. Commonalities and differences emerged from the examination of our case studies, particularly with respect to causes and effects of user trust. Across all cases, trust was found to enhance acceptance and adoption of AI technologies, and some factors such as a system's accuracy were similarly deemed to be crucial for the trustworthiness of all discussed systems. Developers' openness about a system's internal workings, assumptions, and value judgments underlining the technical choices were also considered crucial across the case studies. Such transparency fosters alignment with the goals and values of health care professionals and patients; supports accountability mechanisms; and may help address biases and discrimination, which is essential for building trust among affected communities and stakeholders. However, our analysis also highlights contextual differences among trust in various forms of AI in health care. These differences relate to a multitude of case-specific aspects, such as framing factors, previous knowledge of the trusters, and risk-benefit trade-offs in different environments, necessitating tailored approaches to establish trust and trustworthiness. This finding has implications for anyone aiming to foster trust in AI within the health care domain, from developers and health care professionals using AI systems to public health experts and regulators—trust-building measures should distinguish among different types of AI systems considering not only implied risk levels and legal responsibilities in case of errors but also the factors outlined in our conceptual framework: context, actors, discourse, and the mechanisms of trust building.

Implications

Our findings have implications for developers, patients and health care professionals, and regulators aiming to increase trust in medical AI systems. The developing companies should communicate openly about the algorithmic design and training of their AI systems and tailor their level of transparency to the communicative needs of their respective audiences [122]. Such needs will be very different depending on whether the AI is used by patients directly (assistive AI case), trained physicians (diagnostic and predictive AI cases), health economists and policy makers (resource allocation AI case), or potentially the interested public (public health AI case). Individual AI end users need to be educated on the technology's limitations and carefully consider who bears the risk and responsibility of involving the AI—especially if it is not themselves.

At a regulatory level, distinguishing among different kinds of AI systems in medicine seems also crucial, regulating them according to the level of implied risk as currently proposed by the European Union AI Act—or possibly at a finer-grained level specific for health care [123]. Regulation is also needed to clarify who bears legal responsibility in the case of specific errors. While there is a clear responsibility by developers to avoid systematic errors as much as possible, for the foreseeable future, it will not be possible to design an AI system without any errors. Therefore, some responsibility also falls on health care providers who use the systems as part of their workflows to use them appropriately and with the right level of human oversight. This requires strengthening both the technical capabilities of health care practitioners to evaluate AI systems and the responsibilities placed on the developers of systems to openly document where a system may be expected to generate errors. Given the limits of both, proper certification of AI systems is furthermore crucial, allowing for trust in an overseeing institution to be extended to the system itself [124].

A final important point is that there should be a focus on institutions implementing these tools in a trustworthy manner, not on increasing trust. There are several reasons for this. First, trust could be increased through marketing and presentational gimmicks targeted at patients and clinicians, which would do nothing to increase trustworthiness [105]. Second, trust may be hard to achieve given prevailing and justified distrust in health service providers such as Optum and private health systems and marginalized communities' distrust in health care systems due

XSL•FO
RenderX

to historic or entrenched disparities in outcomes or experiences of care [125]. In contrast, this may be less of a problem where there are higher levels of trust in health institutions that might implement AI systems [106]. Finally, the size and complexity of health systems complicate questions of where, why, and in whom the individual patient should place trust within these systems [87]. This creates challenges in assessing how far the design and implementation of an algorithmic system for resource allocation impacted the attitudes of trust in affected patients.

Limitations

While our findings aim to provide a starting point for fruitful further work on AI in health care, there are several limitations to our consensus process and its results. Of these limitations, 3 are related to the selection of experts for our consensus process. First, owing to the recruitment of participants working on trust in AI in health care, our work is biased toward a position that considers trust a useful and meaningful concept in the context of AI in the first place. While this position is supported by a substantive part of the literature [28,31,49,52], it should be noted that there are also outspoken critics of using the notion of trust with respect to AI in the first place [98,126,127]. Second, given the location of our workshop, our recruitment focused largely on experts from Europe, with only a small addition of participants from North America and Oceania. While this may reflect the central role that trust and trustworthiness have taken in regulatory debates across the European Union, it limits the generalizability of our findings to other contexts. A recent systematic review of empirical research on trust in AI also revealed a lack of diversity in the discussion surrounding this topic [26]. This review underscored the need for a broader range of perspectives to more comprehensively understand and address the complexities of trust in AI systems [26]. Third, to address the lack of agreement on trust in AI in health care among experts, our work focused on expert agreement, leaving out some other relevant stakeholders. As the backgrounds of our participants highlight, some stakeholders were actively involved, from developers to practicing clinicians, but others were not included, such as patients, nurses, or hospital administrators. Further work is needed to address this gap.

There were 2 additional limitations more conceptual in nature, and we thank an anonymous reviewer for flagging these. First, our framework does not distinguish between trust *before* and trust *during* use of an AI-based system, yet causal and framing factors of trust may evolve with the experience of using a technology over time. In fact, all our case studies focus more on the initial adoption of a technology and less on trust evolving during its use, reflecting the current early-stage integration of AI systems into health care. That being said, we do hold that our framework can adequately reflect the relevant factors of trust in different AI systems in health care—even if they may change over time. Second, our work did not focus on the phenomenon of distrust despite its undebatable influence on user acceptance and its irreconcilability with trust. We consciously omitted distrust as it is often considered to not be a mere absence of trust but a more complex, richer phenomenon [19,23,95]. Therefore, we believe that a consensus on distrust deserves a paper of its own. However, it should be noted that distrust toward a specific technology, or potentially against AI in general, can of course play a crucial role in inhibiting trust building at a causal level.

Finally, when considering wider conceptual work on trust and the focus on warranted trust in this paper, we need to acknowledge a body of literature that understands trust to be motivated by emotions rather than calculated decisions [128]. The conceptual thinking in this paper leans more toward a cognitive approach to trust building. Thereby, the effect of affective states on trust might be somehow undervalued. However, these affective states can play an important role in trusting and accepting AI in medicine, although their contribution to trusting AI can vary greatly across different instances of human-AI interactions [129]. Therefore, it will be necessary to extend the conceptual thinking beyond cognitive approaches to encompass traits of affective trust.

In summary, our findings suggest that achieving trustworthy AI systems in health care requires a multifaceted approach bridging human-centered and technology-centered approaches. While regulatory precision and boundaries help provide the legal basis needed to develop trustworthy AI, and while technological design features are critical to successful AI development, communication and positive lived experiences with AI systems are at the forefront of user trust. From a user perspective, especially for those who are not AI aficionados, trustworthy technological features and legal compliance should be a sure thing, whereas communication and positive experiences point beyond the technical sphere of AI.

Conclusions

This paper examines 5 diverse AI systems in health care, revealing the complex landscape of trust in this field. These cases highlight how AI in health care is influenced by various environments, actors, and framing factors. While discussing trust in medical AI in a broad sense may serve as an initial overview, the specific nuances uncovered in these 5 cases demand a more detailed understanding to characterize trusting AI in health care.

Ultimately, however, focus should shift toward ensuring that institutions implement these tools in a *trustworthy* manner rather than merely aiming to indefinitely increase *trust*—only warranted trust is needed for valuable AI implementation. This approach recognizes the complexity of trust dynamics, the risk of superficial increase in trust without a substantial enhancement of trustworthiness, and the importance of trustworthiness in health care institutions. Despite all domain-specific variations, addressing these challenges and embracing transparency and accountability by design can help build a foundation of trust that will underpin the successful integration of AI into health care for the benefit of patients, physicians, and society at large.

XSL•FO
RenderX

Acknowledgments

This paper was produced within the framework of the workshop titled “To trust or not to trust: When should AI be allowed to make decisions?” funded by the Swiss National Science Foundation (SNSF) under grant IZSEZ0_213480 (applicant: MI; coapplicants: A Facchini and FG) and with local support from the École Polytechnique Fédérale de Lausanne. At the time of the workshop, A Ferrario was affiliated with the Mobiliar Lab for Analytics at Federal Institute of Technology Zürich and gratefully acknowledges their support. FG acknowledges the support of the Digital Society Initiative, University of Zurich. Unrelated and outside of the work leading to this paper, FG acknowledges funding from Novartis International AG, Sanitas Krankenversicherung (Stiftung), the Swiss Academy of Engineering Sciences, and the World Health Organization. RC acknowledges the support of the Digitalization Initiative of the Zurich Higher Education Institutions fellowship program of Zurich University of Applied Sciences digital. KJ acknowledges funding from the Wilhelmina Onderzoeksfonds (grant wkz22040701). ER’s research is supported by a Medical Sciences Graduate School Studentship, issued by the Nuffield Department of Population Health, and the Baillie Gifford-Institute for Ethics in AI Scholarship, both associated with the University of Oxford. PT acknowledges funding from Warsaw University of Technology within the Excellence Initiative – Research University program (grant 1820/97/Z01/2023). A Facchini and AT acknowledge funding from the University of Applied Sciences and Arts of Southern Switzerland Best4EthicalAI research program. MI and GS also acknowledge funding from the ERA-NET NEURON network and the SNSF under grant 32NE30_199436 (HybridMinds). GS acknowledges generous support from Fondation Brocher, Hermance, to finish a first full draft during a research residence. The work reported in this paper was funded by the SNSF (grant 32NE30_199436).

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request. The asynchronous consensus-building process was tracked in repeatedly updated documents.

Authors' Contributions

GS, FG, AT, A Facchini, and MI conceived and planned the workshop and the consensus process. All authors provided input on the choice of case studies and the framework. GS coordinated the consensus process and compiled the individual authors’ contributions for a first full draft of the manuscript. GS, AT, A Facchini, and MI drafted a first introduction, and GS, FG, AT, YSJA, JH, KJ, BK, EP, PT, JW, and A Facchini contributed text to the Methods section. KJ led the consensus for the first case study, JH led the consensus for the second case study, YSJA led the consensus for the third case study, GS led the consensus for the fourth case study, and JW led the consensus for the fifth case study. The discussion was jointly developed by GS, FG, AT, A Facchini, and MI. A Ferrario further contributed substantially to the development of the manuscript by providing extensive critical feedback during the review phase. All authors provided repeated critical feedback and read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118 [FREE Full text] [doi: 10.1038/nature21056] [Medline: 28117445]
3. Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol 2023 Aug;24(8):936-944. [doi: 10.1016/S1470-2045(23)00298-X] [Medline: 37541274]
4. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov 22;24(11):1716-1720 [FREE Full text] [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
5. Hyland SL, Faltys M, Hüser M, Lyu X, Gumbsch T, Esteban C, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med 2020 Mar 09;26(3):364-373. [doi: 10.1038/s41591-020-0789-4] [Medline: 32152583]
6. Ienca M, Fabrice J, Elger B, Caon M, Scoccia Pappagallo A, Kressig RW, et al. Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review. J Alzheimers Dis 2017;56(4):1301-1340. [doi: 10.3233/JAD-161037] [Medline: 28222516]
7. Valeriani D, Santoro F, Ienca M. The present and future of neural interfaces. Front Neurorobot 2022 Oct 11;16:953968 [FREE Full text] [doi: 10.3389/fnbot.2022.953968] [Medline: 36304780]

XSL•FO
RenderX

8. Zeng D, Cao Z, Neill DB. Artificial intelligence–enabled public health surveillance—from local detection to global epidemic monitoring and control. In: Xing L, Giger ML, Min JK, editors. Artificial Intelligence in Medicine: Technical Basis and Clinical Applications. New York, NY: Academic Press; 2021:2021-2053.
9. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug 17;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
10. Shani C, Zarecki J, Shahaf D. The lean data scientist: recent advances toward overcoming the data bottleneck. Commun ACM 2023 Jan 20;66(2):92-102. [doi: 10.1145/3551635]
11. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 05;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]
12. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022 Jan 27;24(1):e32215 [FREE Full text] [doi: 10.2196/32215] [Medline: 35084349]
13. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
14. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
15. Chekroud AM, Hawrilenko M, Loho H, Bondar J, Gueorguieva R, Hasan A, et al. Illusory generalizability of clinical prediction models. Science 2024 Jan 12;383(6679):164-167. [doi: 10.1126/science.adg8538] [Medline: 38207039]
16. Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JFP. Guidelines for artificial intelligence in medicine: literature review and content analysis of frameworks. J Med Internet Res 2022 Aug 25;24(8):e36823 [FREE Full text] [doi: 10.2196/36823] [Medline: 36006692]
17. Maddox TM, Rumsfeld JS, Payne PR. Questions for artificial intelligence in health care. JAMA 2019 Jan 01;321(1):31-32. [doi: 10.1001/jama.2018.18932] [Medline: 30535130]
18. Tucci V, Saary J, Doyle TE. Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. J Med Artif Intell 2022 Mar;5:4. [doi: 10.21037/jmai-21-25]
19. McLeod C. Stanford encyclopedia of philosophy. The Metaphysics Research Lab. URL: https://plato.stanford.edu/archives/fall2021/entries/trust/ [accessed 2024-04-29]
20. McKnight DH, Chervany NL. What is trust? A conceptual analysis and an interdisciplinary model. In: Proceedings of the 2000 Americas Conference on Information Systems. 2000 Presented at: AMCIS '00; August 10-13, 2000; Long Beach, CA p. 382 URL: https://aisel.aisnet.org/amcis2000/382
21. Baier A. Tanner lectures on human values. Princeton University. 1991. URL: https://uchv.princeton.edu/events/tanner-lectures-human-values [accessed 2024-04-29]
22. Luhmann N. Vertrauen : ein Mechanismus der Reduktion sozialer Komplexita¨t. Stuttgart, Germany: F. Enke; 1968.
23. Hawley K. Trust, distrust and commitment. Noûs 2012 Oct 25;48(1):1-20 [FREE Full text] [doi: 10.1111/nous.12000]
24. Castelfranchi C, Falcone R. Trust Theory - A Socio-Cognitive and Computational Model. Hoboken, NJ: John Wiley & Sons; 2010.
25. Jacovi A, Marasović A, Miller T, Goldberg Y. Formalizing trust in artificial intelligence: prerequisites, causes and goals of human trust in AI. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021 Presented at: FAccT '21; March 3-10, 2021; Virtual Event p. 624-635 URL: https://dl.acm.org/doi/10.1145/3442188.3445923 [doi: 10.1145/3442188.3445923]
26. Benk M, Kerstan S, von Wangenheim F, Ferrario A. Twenty-four years of empirical research on trust in AI: a bibliometric review of trends, overlooked issues, and future directions. AI Soc 2024 Oct 02:25 [FREE Full text] [doi: 10.1007/S00146-024-02059-Y]
27. Jones C, Thornton J, Wyatt JC. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. Med Law Rev 2023 Nov 27;31(4):501-520 [FREE Full text] [doi: 10.1093/medlaw/fwad013] [Medline: 37218368]
28. Ferrario A, Loi M. How explainability contributes to trust in AI. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022 Presented at: FAccT '22; June 21-24, 2022; Seoul, Republic of Korea p. 1457-1466 URL: https://dl.acm.org/doi/10.1145/3531146.3533202 [doi: 10.1145/3531146.3533202]
29. Starke G, Ienca M. Misplaced trust and distrust: how not to engage with medical artificial intelligence. Camb Q Healthc Ethics 2022 Oct 20;33(3):1-10. [doi: 10.1017/S0963180122000445] [Medline: 36263755]
30. Choung H, David P, Ross A. Trust in AI and its role in the acceptance of AI technologies. Int J Hum Comput Interact 2022 Apr 20;39(9):1727-1739. [doi: 10.1080/10447318.2022.2050543]
31. Gille F, Jobin A, Ienca M. What we talk about when we talk about trust: theory of trust for AI in healthcare. Intell Based Med 2020 Nov;1-2:100001. [doi: 10.1016/j.ibmed.2020.100001]
32. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]

XSL•FO
RenderX

33. Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, et al. Expert consensus document: the International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat Rev Gastroenterol Hepatol 2017 Aug 14;14(8):491-502 [FREE Full text] [doi: 10.1038/nrgastro.2017.75] [Medline: 28611480]
34. Giustina A, Chanson P, Kleinberg D, Bronstein MD, Clemmons DR, Klibanski A, Acromegaly Consensus Group. Expert consensus document: a consensus on the medical treatment of acromegaly. Nat Rev Endocrinol 2014 Apr 25;10(4):243-248. [doi: 10.1038/nrendo.2014.21] [Medline: 24566817]
35. Ienca M, Fins JJ, Jox RJ, Jotterand F, Voeneky S, Andorno R, et al. Towards a governance framework for brain data. Neuroethics 2022 Jun 03;15(2):20. [doi: 10.1007/S12152-022-09498-8]
36. Ives J, Dunn M, Molewijk B, Schildmann J, Bærøe K, Frith L, et al. Standards of practice in empirical bioethics research: towards a consensus. BMC Med Ethics 2018 Jul 10;19(1):68 [FREE Full text] [doi: 10.1186/s12910-018-0304-3] [Medline: 29986689]
37. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, et al. ACCORD (ACcurate COnsensus reporting document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLoS Med 2024 Jan 23;21(1):e1004326 [FREE Full text] [doi: 10.1371/journal.pmed.1004326] [Medline: 38261576]
38. La Brooy C, Pratt B, Kelaher M. What is the role of consensus statements in a risk society? J Risk Res 2019 Jul 25;23(5):664-677. [doi: 10.1080/13669877.2019.1628094]
39. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
40. Ancker JS, Benda NC, Reddy M, Unertl KM, Veinot T. Guidance for publishing qualitative research in informatics. J Am Med Inform Assoc 2021 Nov 25;28(12):2743-2748 [FREE Full text] [doi: 10.1093/jamia/ocab195] [Medline: 34537840]
41. Jabareen Y. Building a conceptual framework: philosophy, definitions, and procedure. Int J Qual Methods 2009 Dec 01;8(4):49-62. [doi: 10.1177/160940690900800406]
42. U.S. Department of HealthHuman Services FDA Center for Drug EvaluationResearch, U.S. Department of HealthHuman Services FDA Center for Biologics EvaluationResearch, U.S. Department of HealthHuman Services FDA Center for DevicesRadiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006 Oct 11;4(1):79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
43. OECD guidelines on measuring trust. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/en/publications/oecd-guidelines-on-measuring-trust_9789264278219-en.html [accessed 2024-04-29]
44. Gille F, Smith S, Mays N. Why public trust in health care systems matters and deserves greater research attention. J Health Serv Res Policy 2015 Jan 17;20(1):62-64. [doi: 10.1177/1355819614543161] [Medline: 25038059]
45. Ferrario A, Loi M, Viganò E. In AI we trust incrementally: a multi-layer model of trust to analyze human-artificial intelligence interactions. Philos Technol 2019 Oct 23;33(3):523-539. [doi: 10.1007/S13347-019-00378-3]
46. Ferrario A, Loi M, Viganò E. Trust does not need to be human: it is possible to trust medical AI. J Med Ethics 2020 Nov 25;47(6):437-438 [FREE Full text] [doi: 10.1136/medethics-2020-106922] [Medline: 33239471]
47. Gille F. What Is Public Trust in the Health System?: Insights into Health Data Use. Bristool, UK: Policy Press; 2023.
48. Durán JM, Formanek N. Grounds for trust: essential epistemic opacity and computational reliabilism. Minds Mach 2018 Oct 29;28(4):645-666. [doi: 10.1007/S11023-018-9481-6]
49. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics 2021 Mar 18:22. [doi: 10.1136/medethics-2020-106820] [Medline: 33737318]
50. Loi M, Ferrario A, Viganò E. How much do you trust me? A logico-mathematical analysis of the concept of the intensity of trust. Synthese 2023 May 23;201(6):186. [doi: 10.1007/S11229-023-04169-4]
51. Chong D, Druckman JN. Framing theory. Annu Rev Polit Sci 2007 Jun 01;10(1):103-126. [doi: 10.1146/annurev.polisci.10.072805.103054]
52. Starke G, van den Brule R, Elger BS, Haselager P. Intentional machines: a defence of trust in medical artificial intelligence. Bioethics 2022 Feb 18;36(2):154-161. [doi: 10.1111/bioe.12891] [Medline: 34142373]
53. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
54. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
55. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018 Oct 24;2(1):35 [FREE Full text] [doi: 10.1186/s41747-018-0061-6] [Medline: 30353365]
56. McDonald ES, McCarthy AM, Akhtar AL, Synnestvedt MB, Schnall M, Conant EF. Baseline screening mammography: performance of full-field digital mammography versus digital breast tomosynthesis. AJR Am J Roentgenol 2015 Nov;205(5):1143-1148. [doi: 10.2214/AJR.15.14406] [Medline: 26496565]

XSL•FO
RenderX

57. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian association of radiologists white paper on artificial intelligence in radiology. Can Assoc Radiol J 2018 May;69(2):120-135 [FREE Full text] [doi: 10.1016/j.carj.2018.02.002] [Medline: 29655580]
58. Bejnordi BE, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, the CAMELYON16 Consortium, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017 Dec 12;318(22):2199-2210 [FREE Full text] [doi: 10.1001/jama.2017.14585] [Medline: 29234806]
59. Ghafoorian M, Karssemeijer N, Heskes T, Bergkamp M, Wissink J, Obels J, et al. Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. Neuroimage Clin 2017;14:391-399 [FREE Full text] [doi: 10.1016/j.nicl.2017.01.033] [Medline: 28271039]
60. Oxipit awarded CE mark for first autonomous AI medical imaging application. Oxipit.ai. URL: https://oxipit.ai/news/first-autonomous-ai-medical-imaging-application/ [accessed 2024-04-29]
61. Keski-Filppula T, Nikki M, Haapea M, Ramanauskas N, Tervonen O. Using artificial intelligence to detect chest X-rays with no significant findings in a primary health care setting in Oulu, Finland. arXiv Preprint posted online May 17, 2022 [FREE Full text] [doi: 10.48550/arXiv.2205.08123]
62. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics 2020 Mar;46(3):205-211 [FREE Full text] [doi: 10.1136/medethics-2019-105586] [Medline: 31748206]
63. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nat Biomed Eng 2022 Dec 11;6(12):1346-1352. [doi: 10.1038/s41551-022-00914-1] [Medline: 35953649]
64. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. J Pathol Inform 2018;9:38 [FREE Full text] [doi: 10.4103/jpi.jpi_53_18] [Medline: 30607305]
65. Pena GP, Andrade-Filho JS. How does a pathologist make a diagnosis? Arch Pathol Lab Med 2009 Jan;133(1):124-132 [FREE Full text] [doi: 10.5858/133.1.124] [Medline: 19123724]
66. Arbelaez Ossa L, Starke G, Lorenzini G, Vogt JE, Shaw DM, Elger BS. Re-focusing explainability in medicine. Digit Health 2022 Feb 11;8:20552076221074488 [FREE Full text] [doi: 10.1177/20552076221074488] [Medline: 35173981]
67. Starke G, Elger BS, De Clercq E. Machine learning and its impact on psychiatric nosology: findings from a qualitative study among German and Swiss experts. PhiMiSci 2023 Apr 11;4:1-17. [doi: 10.33735/phimisci.2023.9435]
68. Walker MJ, Rogers WA. Diagnosis, narrative identity, and asymptomatic disease. Theor Med Bioeth 2017 Aug 5;38(4):307-321. [doi: 10.1007/s11017-017-9412-1] [Medline: 28681328]
69. Drogt J, Milota M, Vos S, Bredenoord A, Jongsma K. Integrating artificial intelligence in pathology: a qualitative interview study of users' experiences and expectations. Mod Pathol 2022 Nov;35(11):1540-1550 [FREE Full text] [doi: 10.1038/s41379-022-01123-6] [Medline: 35927490]
70. Starke G, Schmidt B, De Clercq E, Elger BS. Explainability as fig leaf? An exploration of experts' ethical expectations towards machine learning in psychiatry. AI Ethics 2022 Jun 07;3(1):303-314. [doi: 10.1007/S43681-022-00177-1]
71. Yang L, Ene IC, Arabi Belaghi R, Koff D, Stein N, Santaguida P. Stakeholders' perspectives on the future of artificial intelligence in radiology: a scoping review. Eur Radiol 2022 Mar 21;32(3):1477-1495. [doi: 10.1007/s00330-021-08214-z] [Medline: 34545445]
72. Aquino YS, Rogers WA, Braunack-Mayer A, Frazer H, Win KT, Houssami N, et al. Utopia versus dystopia: professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. Int J Med Inform 2023 Jan;169:104903 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104903] [Medline: 36343512]
73. Patel D, Kher V, Desai B, Lei X, Cen S, Nanda N, et al. Machine learning based predictors for COVID-19 disease severity. Sci Rep 2021 Feb 25;11(1):4673 [FREE Full text] [doi: 10.1038/s41598-021-83967-7] [Medline: 33633145]
74. Jauk S, Kramer D, Avian A, Berghold A, Leodolter W, Schulz S. Correction to: technology acceptance of a machine learning algorithm predicting delirium in a clinical setting: a mixed-methods study. J Med Syst 2021 Mar 10;45(4):52 [FREE Full text] [doi: 10.1007/s10916-021-01728-5] [Medline: 33740133]
75. van Royen FS, Moons KG, Geersing GJ, van Smeden M. Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. Eur Respir J 2022 Jun 21;60(3):2200250. [doi: 10.1183/13993003.00250-2022]
76. Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, Takigami AK, et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. J Clin Oncol 2023 Apr 20;41(12):2191-2200 [FREE Full text] [doi: 10.1200/JCO.22.01345] [Medline: 36634294]
77. Marinovich ML, Wylie E, Lotter W, Pearce A, Carter SM, Lund H, et al. Artificial intelligence (AI) to enhance breast cancer screening: protocol for population-based cohort study of cancer detection. BMJ Open 2022 Jan 03;12(1):e054005 [FREE Full text] [doi: 10.1136/bmjopen-2021-054005] [Medline: 34980622]
78. King OC, Mertens M. Self-fulfilling prophecy in practical and automated prediction. Ethic Theory Moral Prac 2023 Jan 16;26(1):127-152. [doi: 10.1007/S10677-022-10359-9]
79. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. Perspect Psychol Sci 2016 Sep 29;11(5):702-712. [doi: 10.1177/1745691616658637] [Medline: 27694465]
80. Marx CT, Calmon FD, Ustun B. Predictive multiplicity in classification. In: Proceedings of the 37 th International Conference on Machine Learning. 2020 Presented at: PMLR '20; July 12-18, 2020; Vienna, Austria p. 1-10 URL: https://proceedings.mlr.press/v119/marx20a.html

XSL•FO
RenderX

81. Kulynych B, Hsu H, Troncoso C, Calmon FP. Arbitrary decisions are a hidden cost of differentially private training. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2023 Presented at: FAccT '23; June 12-15, 2023; Chicago, IL p. 1609-1623 URL: https://dl.acm.org/doi/10.1145/3593013.3594103 [doi: 10.1145/3593013.3594103]
82. Kim A, Chung KC, Keir C, Patrick DL. Patient-reported outcomes associated with cancer screening: a systematic review. BMC Cancer 2022 Mar 01;22(1):223 [FREE Full text] [doi: 10.1186/s12885-022-09261-5] [Medline: 35232405]
83. Dickinson JA, Pimlott N, Grad R, Singh H, Szafran O, Wilson BJ, et al. Screening: when things go wrong. Can Fam Physician 2018 Jul;64(7):502-508 [FREE Full text] [Medline: 30002025]
84. Peng Y, Liu E, Peng S, Chen Q, Li D, Lian D. Using artificial intelligence technology to fight COVID-19: a review. Artif Intell Rev 2022 Jan 03;55(6):4941-4977 [FREE Full text] [doi: 10.1007/s10462-021-10106-z] [Medline: 35002010]
85. MacIntyre CR, Lim S, Quigley A. Preventing the next pandemic: use of artificial intelligence for epidemic monitoring and alerts. Cell Rep Med 2022 Dec 20;3(12):100867 [FREE Full text] [doi: 10.1016/j.xcrm.2022.100867] [Medline: 36543103]
86. Bhargavi BS, Moa A. Global outbreaks of zika infection by epidemic observatory (EpiWATCH), 2016-2019. Glob. Biosecurity 2020 Oct 21;2:55. [doi: 10.31646/gbio.83]
87. Lesmanawati DAS, Adam DC, Hooshmand E, Moa A, Kunasekaran MP, MacIntyre CR. The global epidemiology of hepatitis a outbreaks 2016-2018 and the utility of EpiWATCH as a rapid epidemic intelligence service. Glob Biosecurity 2021 Mar 23;3(1):25. [doi: 10.31646/gbio.100]
88. Kuhner D, Fiederer LD, Aldinger J, Burget F, Völker M, Schirrmeister RT, et al. A service assistant combining autonomous robotics, flexible goal formulation, and deep-learning-based brain–computer interfacing. Robot Auton Syst 2019 Jun;116:98-113. [doi: 10.1016/j.robot.2019.02.015]
89. ISO 8549-3:2020: prosthetics and orthotics — vocabulary: part 3: terms relating to orthoses. International Organization for Standardization. 2022. URL: https://www.iso.org/standard/79497.html [accessed 2024-04-29]
90. Baniqued PD, Stanyer EC, Awais M, Alazmani A, Jackson AE, Mon-Williams MA, et al. Brain-computer interface robotics for hand rehabilitation after stroke: a systematic review. J Neuroeng Rehabil 2021 Jan 23;18(1):15 [FREE Full text] [doi: 10.1186/s12984-021-00820-8] [Medline: 33485365]
91. Kellmeyer P, Mueller O, Feingold-Polak R, Levy-Tzedek S. Social robots in rehabilitation: a question of trust. Sci Robot 2018 Aug 15;3(21):eaat1587. [doi: 10.1126/scirobotics.aat1587] [Medline: 33141717]
92. Carmena JM. Brain versus machine control. PLoS Biol 2004 Dec 14;2(12):e430. [doi: 10.1371/journal.pbio.0020430]
93. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453 [FREE Full text] [doi: 10.1126/science.aax2342] [Medline: 31649194]
94. The Lancet Digital Health. There is no such thing as race in health-care algorithms. Lancet Digit Health 2019 Dec;1(8):e375 [FREE Full text] [doi: 10.1016/S2589-7500(19)30201-8] [Medline: 33323212]
95. Jones K. Trust as an Affective Attitude. Ethics 1996 Oct;107(1):4-25. [doi: 10.1086/233694]
96. Sheehan M, Friesen P, Balmer A, Cheeks C, Davidson S, Devereux J, et al. Trust, trustworthiness and sharing patient data for research. J Med Ethics 2020 May 18;47(12):e26. [doi: 10.1136/medethics-2019-106048] [Medline: 32424061]
97. McGeer V, Pettit P. The empowering theory of trust. In: Faulkner P, Simpson T, editors. The Philosophy of Trust. Oxford, UK: Oxford Academic Press; 2017:14-34.
98. Hatherley JJ. Limits of trust in medical AI. J Med Ethics 2020 Jul;46(7):478-481. [doi: 10.1136/medethics-2019-105935] [Medline: 32220870]
99. Wadden JJ. What kind of artificial intelligence should we want for use in healthcare decision-making applications? Can J Bioeth 2021 May 27;4(1):94-100. [doi: 10.7202/1077636ar]
100. Ratti E, Graves M. Explainable machine learning practices: opening another black box for reliable medical AI. AI Ethics 2022 Feb 15;2(4):801-814. [doi: 10.1007/S43681-022-00141-z]
101. McLennan S, Fiske A, Tigard D, Müller R, Haddadin S, Buyx A. Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Med Ethics 2022 Jan 26;23(1):6 [FREE Full text] [doi: 10.1186/s12910-022-00746-3] [Medline: 35081955]
102. Algorithmic impact assessment: a case study in healthcare. Ada Lovelace Instititue. URL: https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/ [accessed 2024-04-29]
103. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform 2021 Jan;113:103655 [FREE Full text] [doi: 10.1016/j.jbi.2020.103655] [Medline: 33309898]
104. Gerke S. Health AI for good rather than evil? The need for a new regulatory framework for AI-based medical devices. Yale J Health Policy Law Ethics 2012;20(2):432 [FREE Full text]
105. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. BMJ 2019 Mar 12;364:l886 [FREE Full text] [doi: 10.1136/bmj.l886] [Medline: 30862612]
106. Möllering G. The nature of trust: from Georg Simmel to a theory of expectation, interpretation and suspension. Sociology 2001;35(2):403-420. [doi: 10.1017/S0038038501000190]
107. Simmel G. Soziologie: Untersuchungen über die Formen der Vergesellschaftung. Berlin, Germany: Duncker und Humblot; 1908.

XSL•FO
RenderX

108. Nickel PJ. Trust in medical artificial intelligence: a discretionary account. Ethics Inf Technol 2022 Jan 24;24(1):e56. [doi: 10.1007/S10676-022-09630-5]
109. Ethics guidelines for trustworthy AI. European Commission: Directorate-General for Communications Networks, Content and Technology. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-en [accessed 2024-04-29]
110. Braun M, Bleher H, Hummel P. A leap of faith: is there a formula for "trustworthy" AI? Hastings Cent Rep 2021 May 19;51(3):17-22. [doi: 10.1002/hast.1207] [Medline: 33606288]
111. O'Neill O. Linking trust to trustworthiness. Int J Philos Stud 2018 Apr 25;26(2):293-300. [doi: 10.1080/09672559.2018.1454637]
112. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
113. Mincu D, Roy S. Developing robust benchmarks for driving forward AI innovation in healthcare. Nat Mach Intell 2022 Nov 15;4(11):916-921. [doi: 10.1038/S42256-022-00559-4]
114. Strickland E. IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. IEEE Spectr 2019 Apr;56(4):24-31. [doi: 10.1109/mspec.2019.8678513]
115. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA 2019 Dec 24;322(24):2377-2378 [FREE Full text] [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
116. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med 2018 Nov;15(11):e1002689 [FREE Full text] [doi: 10.1371/journal.pmed.1002689] [Medline: 30399149]
117. Char DS, Abràmoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. Am J Bioeth 2020 Nov;20(11):7-17 [FREE Full text] [doi: 10.1080/15265161.2020.1819469] [Medline: 33103967]
118. Powell J. Trust me, I'm a chatbot: how artificial intelligence in health care fails the Turing test. J Med Internet Res 2019 Oct 28;21(10):e16222 [FREE Full text] [doi: 10.2196/16222] [Medline: 31661083]
119. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May 13;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
120. Tschantz MC. What is proxy discrimination? In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022 Presented at: FAccT '22; June 21-24, 2022; Seoul, Republic of Korea p. 1993 URL: https://dl.acm.org/doi/fullHtml/10.1145/3531146.3533242 [doi: 10.1145/3531146.3533242]
121. Suleman M, Qureshi Z. Should medicine be colour blind? J Med Ethics 2023 Nov 23;49(11):725-726. [doi: 10.1136/jme-2023-109634] [Medline: 37871944]
122. Starke G. The emperor's new clothes? Transparency and trust in machine learning for clinical neuroscience. In: Friedrich O, Wolkenstein A, Bublitz C, Jox RF, Racine E, editors. Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications. Cham, Switzerland: Springer; 2021:2021-2096.
123. Reddy S. Navigating the AI revolution: the case for precise regulation in health care. J Med Internet Res 2023 Sep 11;25:e49989 [FREE Full text] [doi: 10.2196/49989] [Medline: 37695650]
124. Scharowski N, Benk M, Kühne SJ, Wettstein L, Brühlmann F. Certification labels for trustworthy ai: Insights from an empirical mixed-method study. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. 2023 Presented at: FAccT '23; June 12-15, 2023; Chicago, IL p. 248-260 URL: https://dl.acm.org/doi/10.1145/3593013.3593994 [doi: 10.1145/3593013.3593994]
125. Lee MK, Rich K. Who is included in human perceptions of AI?: Trust and perceived fairness around healthcare AI and cultural mistrust. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-14 URL: https://dl.acm.org/doi/10.1145/3411764.3445570 [doi: 10.1145/3411764.3445570]
126. Ryan M. In AI we trust: ethics, artificial intelligence, and reliability. Sci Eng Ethics 2020 Oct 10;26(5):2749-2767 [FREE Full text] [doi: 10.1007/s11948-020-00228-y] [Medline: 32524425]
127. Metzinger T. Ethics washing made in Europe. Der Tagesspiegel. URL: https://www.tagesspiegel.de/politik/ethics-washing-made-in-europe-5937028.html [accessed 2024-04-29]
128. PytlikZillig LM, Kimbrough CD. Consensus on conceptualizations and definitions of trust: are we there yet? In: Shockley E, Neal TM, PytlikZillig LM, Bornstein BH, editors. Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration. Cham, Switzerland: Springer; 2016:17-47.
129. Kyung N, Kwon HE. Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of AI for preventive care operations. Prod Oper Manag 2022 Jul 31:13785 [FREE Full text] [doi: 10.1111/poms.13785]

Abbreviations

AI: artificial intelligence
BCI: brain-computer interface
ICU: intensive care unit

XSL•FO
RenderX

Edited by N Cahill, T Leung; submitted 12.01.24; peer-reviewed by H Burkhardt, P Nickel, L Weinert; comments to author 29.04.24; revised version received 31.07.24; accepted 28.11.24; published 19.02.25.

Please cite as:
Starke G, Gille F, Termine A, Aquino YSJ, Chavarriaga R, Ferrario A, Hastings J, Jongsma K, Kellmeyer P, Kulynych B, Postan E, Racine E, Sahin D, Tomaszewska P, Vold K, Webb J, Facchini A, Ienca M
Finding Consensus on Trust in AI in Health Care: Recommendations From a Panel of International Experts
J Med Internet Res 2025;27:e56306
URL: https://www.jmir.org/2025/1/e56306
doi:10.2196/56306
PMID:39969962

©Georg Starke, Felix Gille, Alberto Termine, Yves Saint James Aquino, Ricardo Chavarriaga, Andrea Ferrario, Janna Hastings, Karin Jongsma, Philipp Kellmeyer, Bogdan Kulynych, Emily Postan, Elise Racine, Derya Sahin, Paulina Tomaszewska, Karina Vold, Jamie Webb, Alessandro Facchini, Marcello Ienca. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Publishing Identifiable Patient Photographs in the Digital Age: Focus Group Study of Patients, Doctors, and Medical Students

Marija Roguljić[1], DMD, PhD; Dina Šimunović[2], DMD; Ivan Buljan[3], MPsych, PhD; Marija Franka Žuljević[4], MD, PhD; Antonela Turić[5], DMD; Ana Marušić[6], MD, PhD

[1]Department of Periodontology, University of Split, Split, Croatia

[2]Private Dental Practice, Split, Croatia

[3]Department of Psychology, Faculty of Humanities and Social Sciences in Split, University of Split School of Medicine, Split, Croatia

[4]Department of Medical Humanities, University of Split School of Medicine, Split, Croatia

[5]Dental Outpatient Clinic Split, Vis, Croatia

[6]Department of Research in Biomedicine and Health, Center for Evidence-Based Medicine, University of Split School of Medicine, Split, Croatia

Corresponding Author:
Marija Roguljić, DMD, PhD
Department of Periodontology
University of Split
Šoltanska 2A
Split, 21000
Croatia
Phone: 385 21557800
Email: marija.roguljic@mefst.hr

Abstract

Background: The publication of patient photographs in scientific journals continues to pose challenges regarding privacy and confidentiality, despite existing ethical guidelines. Recent studies indicate that key stakeholders—including health care professionals and patients—lack sufficient awareness of the ethical considerations surrounding patient photographs, particularly in the context of digital scientific publishing.

Objective: This qualitative study aims to explore how different stakeholders—patients, medical students, and doctors—understand the challenges of patient privacy and confidentiality in scientific publications. Additionally, it sought to identify key areas for future research, particularly in the context of online, open-access articles.

Methods: We conducted 4 online focus groups due to COVID-19 restrictions: 1 with patients, 2 with final-year medical students, and 1 with head and neck physicians and dentists who regularly handle patient photographs. Participants were invited via email, and those who accepted took part in discussions lasting approximately 1 hour. All interviews were recorded and transcribed for analysis. All 4 focus groups were asked the same set of questions, covering the following topics: (1) consent for publishing patient photographs, (2) information on guidelines and standards for consent to publish patient photographs, (3) the importance of informed consent for various purposes, (4) methods for deidentifying patient photographs, and (5) the use of patient photographs in online, open-access publishing.

Results: Three key themes emerged from the focus group discussions: (1) no definitive resources or practical recommendations available, (2) online publishing of patient images makes them more open to misuse, and (3) anonymization techniques have limitations. All stakeholder groups expressed a lack of knowledge about online publishing in general and concerns about the fate of patient photographs in the digital environment after publication. They emphasized the need for increased awareness among all relevant stakeholders and more stringent procedures for obtaining informed patient consent before publishing photographs. While they recognized the usefulness of image anonymization techniques in protecting patient identity, they were also aware that current methods remain insufficient to ensure complete anonymity.

Conclusions: This qualitative study highlights that publishing patient photographs in open-access scientific journals is an important, serious, and largely unexplored issue, with all stakeholders still uncertain about the best ways to protect patient privacy. Clinicians, publishers, and journal editors should not only implement best practices to ensure fully informed patient consent for publishing identifiable photographs but also develop technical and governance safeguards. Future quantitative studies are needed

XSL•FO
RenderX

to identify the most effective ways to enhance stakeholders' knowledge, policies, and procedures, ultimately guiding the development of practical recommendations for the ethical publication of patient photographs in scientific journals.

(J Med Internet Res 2025;27:e59970) doi:10.2196/59970

KEYWORDS

patient photographs; patient privacy, confidentiality; data protection; ethical publishing; informed consent; open access; scientific journals; focus group.

Introduction

Patient photographs are an essential tool in medical education and are routinely used in daily communication between health professionals to illustrate various clinical problems, procedures, treatment outcomes, and follow-ups. They are ubiquitous on social media platforms, websites, and in scientific journals; however, it is not always clear whether patient privacy is adequately protected in these outlets. As patient photographs are part of confidential medical documentation, their use must adhere to standards consistent with best ethical practices.

According to best practices outlined by the International Committee for Medical Journal Editors (ICMJE), it is recommended to avoid identifying patients in photographs whenever possible and to prioritize patient privacy [1]. The challenge lies in defining the boundary between nonidentifiable and identifiable photographs, particularly in the case of facial images. Traditionally, photographs of body parts or radiographs have been considered nonidentifiable and, therefore, do not require written patient consent. However, when such images are tagged with patient information in an article, they can become identifiable [2]. Additionally, artificial intelligence (AI)–based computer programs have demonstrated the ability to identify individuals from various types of patient photographs previously considered nonidentifiable [3].

Studies suggest that patients generally have a high acceptance of the use of their photographs for various purposes [4-6]. However, these studies also highlight the risk of privacy invasion through the disclosure of an individual's identity. Several studies have shown that patients prefer the use of nonidentifiable photographs [4,6,7]. Additionally, patients are more likely to approve the use of their photographs for follow-up and education rather than for publications, social media, websites, or television broadcasts [8-10].

Although existing ethical standards for protecting patient privacy in the use of photographs are governed by various professional guidelines and legal regulations [1,11-13], a standardized practice has yet to be established. In a recent scoping review, we identified a wide range of disclosure practices for identifiable patient photographs and a lack of awareness of these practices and associated risks among both patients and health professionals [14].

The ethical publication of patient images is particularly relevant in the era of open-access journals and digital publishing [15,16]. For example, some Creative Commons licenses allow anyone to access and reuse parts or entire published scientific articles for any purpose, including commercial use, provided the source and authors are acknowledged [17]. Our previous study found that patients have a high level of trust in their doctors, often believing that verbal consent is sufficient or that permission is not even necessary to publish photographs of their bodies [18]. However, it remains unclear whether both patients and doctors fully understand the potential consequences of publishing identifiable patient photographs in open-access publications.

We conducted a focus group study with patients, students, and doctors—key stakeholders in scientific publishing—to further explore their views on publishing patient photographs in scientific journals, particularly in the context of open-access publishing.

Methods

Study Design

This qualitative study was conducted and reported following the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist for qualitative research (available in Multimedia Appendix 1) [19]. We conducted 4 focus groups using semistructured interviews with different stakeholder groups.

Participants and Recruitment Procedure

Participants in this study included patients, students, and doctors from the University of Split School of Medicine (USSM) and the University Hospital of Split. We used purposive sampling to ensure the representation of different stakeholders across various roles in the research and publishing process:

Patients from the Department of Dentistry, University Hospital of Split, who had experience with diagnostic or therapeutic procedures related to the head and neck area, including participation in research studies.

Doctors with expertise in head and neck conditions from the Departments of Neurology, Dermatology, Maxillofacial Surgery, and Otorhinolaryngology at the University Hospital of Split, Split, Croatia.

Students in the final years of medical school at the USSM. We conducted 2 focus groups with medical students—1 with students from Croatia enrolled in the Croatian-language medical program and another with international students attending the English-language medical program. This distinction was made to capture perspectives from students with experiences in different health care settings (Croatia vs other countries, primarily within the European Union).

We contacted potential participants via email, inviting them to take part in a focus group discussion on research and publication ethics. In total, 30 individuals were invited to participate, and the groups were formed using a purposive sampling method.

XSL•FO
RenderX

Each focus group was designed to achieve a balance in age and gender.

The focus group with non-Croatian students was conducted in English, while the other groups were held in Croatian.

Focus Group Setting and Data Collection

Focus groups were conducted and recorded at the USSM during April and May 2020. The USSM is located in Split, the second-largest city in Croatia, with a metropolitan area population of approximately 350,000.

The focus group interviews were conducted online by 2 lead researchers: 1 female (MR) and 1 male (IB). As a result of COVID-19 restrictions on in-person meetings, we held virtual focus groups using the Zoom platform (Zoom Video Communications, Inc.).

The focus group questions were prepared in advance, and all 4 groups were asked the same set of questions, as outlined in the study protocol (Multimedia Appendix 2). Participants were asked about (1) their opinion on study findings showing that patients trust their doctors to publish identifiable photographs responsibly [18]; (2) sources of information on guidance and standards for obtaining valid informed consent; (3) different purposes for using patient clinical photographs; (4) anonymization of photographs; (5) consequences of publishing identifiable photographs in scientific journals; and (6) open-access publishing and licensing in relation to patient privacy protection.

Each focus group session lasted approximately 1 hour. The discussions were recorded using an audio-recording device and transcribed into a document file by one of the researchers (DŠ). The transcripts were then coded, and the codes were entered into a spreadsheet (Microsoft Excel).

Research Team and Data Analysis

Most participants were familiar with the research team before the study began. The interviewer (MR), who served as the lead investigator, is a certified dental medicine doctor and an Assistant Professor at the USSM. Data analysis involved coding the transcripts, categorizing the initial codes, and identifying themes and patterns. Participant identities were anonymized by 1 of the investigators (DŠ). Each participant was assigned a code based on their focus group and a consecutive number as follows: F1, patients; F2, doctors; F3, medical students in the Croatian-language program; and F4, medical students in the English-language program, with individual participants numbered from P1 to P28.

Two researchers (MR and DŠ) conducted the initial coding of themes based on participants' comments. Preliminary themes were identified during the analysis based on the focus group questions and were iteratively refined using a constant comparative approach [20], allowing for both the refinement of existing themes and the identification of new ones. During the initial coding, 1 researcher (DŠ) defined a large number of codes (>20), which were then reviewed and refined by 2 researchers (MR and AM) to reach a final consensus. Each code was assigned to a single theme category. Data saturation was reached after 3 focus groups, as the analysis of the third group did not yield any new themes. After deriving the themes, summaries of the identified themes were sent to participants for validation. Participants' statements were translated from Croatian into English by 1 author (DŠ), checked for accuracy by another (MR), and finally reviewed by a professional translator. A selection of participant quotations is presented to illustrate the theme categories.

Ethical Considerations

Ethics approval was obtained from the Ethics Committee of the USSM (Class: 003-08/20-03/0005, Reg. No.: 2181-198-04-20-0048) under the project funded by the Croatian Research Foundation, Professionalism in Health: Decision Making in Practice and Research (principal investigator: AM).

Before the focus group interviews, all participants were provided with an information letter and an informed consent form. This documentation outlined the study's purpose, funding, recruitment process, methodology, potential risks or adverse effects, beneficiaries of the research findings, communication of results, data collection, analysis, and protection of personal information. It also informed participants of their right to withdraw from the study at any time and their opportunity to review and, if relevant, comment on interview transcripts and quotations. All participants signed the consent form after reviewing the study information. No compensation was provided for participation.

At the beginning of each focus group, participants were reminded about the confidentiality of any information shared during the discussion and were asked to avoid using identifying characteristics when describing their experiences. They were also informed that only the lead investigator, who conducted all focus groups, would have access to identifying data and that all data would be anonymized for analysis and publication.

Results

Overview

Thirty individuals were initially contacted and agreed to participate; however, 2 later withdrew due to online connectivity issues, resulting in a final total of 28 participants in the focus groups. There was a good balance of gender, age, and education level among participants in each focus group (Table 1), except for a wider age range among the patient group.

Three main topics emerged from the focus group discussions: (1) no definitive resources or practical recommendations available, (2) online publishing of patient images makes them more open to misuse, and (3) anonymization techniques have limitations.

XSL•FO
RenderX

Table 1. Characteristics of the participants in focus groups on publishing identifiable patient photographs

Characteristics	Patients (n=8)	Doctors (n=8)	Students of medical studies in Croatian (n=7)	Students of medical studies in English (n=5)
Gender (female/male), n	4/3	3/5	4/3	2/3
Age (years), median (range)	40.5 (24-61)	37.5 (28-48)	24 (24-25)	25 (24-25)
Country of origin	Croatia	Croatia	Croatia	Croatia (n=1), Germany (n=3), and United States (n=1)
Place of recruitment	Department of Dental Medicine, University Hospital of Split	Departments of Dentistry (n=4), Maxillofacial Surgery (n=1), Neurology (n=1), Dermatology (n=1), and Otorhinolaryngology (n=1) at the University Hospital of Split	University of Split School of Medicine	University of Split School of Medicine
Education level	Master's degree (n=7), PhD degree (n=1)	MD degree (n=2), MD/PhD degree (n=6)	Final (6th) year of the study	Final (6th) year of the study

No Definitive Resources or Practical Recommendations Available

The first theme highlighted the lack of clear best-practice recommendations for obtaining informed consent for the use of patient photographs in online journals. Participants expressed that universal guidelines, applicable to all cases and types of patient photographs, would be ideal. However, they also acknowledged the complexity and variability of different types of patient photographs and the contexts in which they are used, noting that not all images pose the same level of risk to patient privacy.

When asked about the resources they would turn to for best practices, participants expressed considerable uncertainty and gave varied responses. Some felt there should be someone they could consult but were unsure who that would be, while others suggested referring to journal-specific guidelines or seeking advice from their local ethics committee:

> *There should be a body that answers such questions. I am not sure which one it is or if it even exists. But there should definitely be someone to turn to.* [F3/P23, student]
>
> *Maybe on the site of one of the journals in which I want to submit my paper to. I think most of them have some kind of criteria for publishing patients' data, so I would start there.* [F4/P24, student]
>
> *I would definitely check with the journal, check with somebody who had more experience, maybe Google and see what are like general guidelines.* [F4/P27, student]
>
> *...first, there are some instructions from the journal, the ethical commission and the ethical protocol. That is second, and after that, if we had to go further, I don't know, we didn't have the opportunity to have some kind of education about it.* [F2/P16, doctor]

Other participants suggested consulting colleagues or someone within their institution who might have the necessary knowledge, but again, no definitive source was identified.

> *I would also check in with more senior authors, there is someone else on the paper I am working with, maybe my supervisor or anyone who maybe has more experience. They would probably point me to some guidelines but I would also contact someone that I can talk to.* [F4/P26, student]
>
> *Well, we can informally consult with the ethics committee of Clinical Hospital Centre Split, I guess they know, they don't have to give us anything officially, but we can just ask them as colleagues.* [F3/P21, student]
>
> *I guess they should appoint someone to deal with it.* [F3/P22, student]

The uncertainty and variability in responses indicated a general lack of awareness about reliable resources for publishing patient photographs in online journals. Building on this, participants emphasized the need for a clear and comprehensive set of recommendations and best practices to address this issue:

> *Well, maybe there should be some universal form that needs to be filled out and then it's the same for everyone, so that we know, that we can't make a mistake.* [F3/P17, student]
>
> *I think we should have a uniform attitude for any purpose of the image.* [F2/P11, doctor]
>
> *...you never know in the beginning what study might end up in a textbook so I think there should be one form of consent in the beginning of the study and once this is signed it should be OK either to publish into only the study or into the textbook.* [F4/P28, student]

Regarding the idea of definitive or universal best-practice recommendations, participants acknowledged that while desirable, creating such guidelines would be challenging due to the variability in patient photographs and the specific body parts that need to be shown in scientific publications. They provided examples, emphasizing that not all body parts carry the same level of sensitivity when publicly disclosed, potentially placing patients in more vulnerable positions to varying degrees.

XSL•FO
RenderX

> *Because most people don't even care if it's a picture of a tooth, nobody can connect him to it so he doesn't care, on the other hand, when you send him a picture of your face, he is not indifferent.* [F3/P18, student]
>
> *A photo of the face is one thing, and a photo of the genitals is another. Let's say these are two different categories within which a face could be seen according to some criteria.* [F1/P6, patient]

The complexity of the issue was further highlighted by participants who suggested that decisions should be based on the potential audience size for the photograph.

> *wherever that image is shown to a larger number of people, the greater the need for protection and the more detailed the consent should be.* [F1/P6, patient]
>
> *I think it's actually about the number of people who will be seeing the image.* [F1/P4, patient]

Overall, participants' responses suggested a lack of certainty and knowledge on how to obtain patient consent for publicly displayed photographs. One participant explicitly voiced this concern, stating, "I think it is something all of us need to get educated more on" [F4/P26]. This lack of awareness was further compounded by the absence of a universal or easily accessible resource to guide them, as well as the inherent complexity of the issue. Additionally, some doctors in the study believed they were not obligated to seek patient consent for using photographs, arguing that, as health care professionals, they were the most qualified to make such decisions.

> As long as photographs are used in good and positive purposes, there is no need to obtain the consent for education, showing students, publication in journal. Of course, it implies that it will not be any type of abuse. It is mandatory to protect dignity and identity as far as possible. [F2/P14, doctor]

Online Publishing of Patient Images Makes Them More Open to Misuse

Participants viewed the online publication of patient images as a complex issue, where patient privacy and photographs could be vulnerable to unregulated or illegal use. A key concern was the lack of control over how and where these images might be used. Additionally, participants highlighted that patients often lack sufficient knowledge to make fully informed decisions about digital publishing, including how images can be reused or the implications of different copyright licenses. Some proposed solutions included clarifying who is responsible for preventing misuse and finding ways to empower patients in their decision-making. Finally, participants—especially students—emphasized the crucial role of the doctor-patient relationship in ensuring proper informed consent and maintaining ethical integrity.

Despite existing copyright regulations, including requirements to credit the original material (eg, even under the most open Creative Commons license, CC BY 4.0), participants felt that patients and their images remained vulnerable to misuse.

> *People copy them [patient images] and take them for themselves and put them in their presentations when they need them for lectures and seminars.* [F3/P21, student]
>
> *So it makes a little difference to me because it refers to my private life, my private medical case or problem. So if such a picture is misused, I think it's a little worse than the one on the profile picture, which someone can cut out and put on something else.* [F1/P4, patient]
>
> *most of us have social networks and understand that a profile picture can be used for any purpose. Same thing [with pictures in journals].* [F1/P1, patient]

This potential for misuse was seen as higher in open-access journals:

> *Patients are losing their privacy in any case so even if you still say OK, we want to preserve the rights, technically you still end up doing it. The authors are giving up a part of their rights. They are also losing a part of protection of the patient.* [F4/P26, student]
>
> *In my opinion, it should be limited how the images of these scientific papers can be used, whether they are open-access or not. I mean, the man didn't sign to be on some advertisement one day, he signed to be in a scientific magazine, so in my opinion it should be arranged a little, what can be done with that open-access.* [F3/P18, student]
>
> *I think the main difference is whether the journal is published online or not, is it open-access or not. I think that is a huge issue...If this photo is something that someone can find on the internet easily or it is a physical copy of the journal. This really changes their issue.* [F4/P26, student]

Some participants held a contrasting opinion, stating that they saw no distinction between open-access and closed-access publications, as most articles could still be illegally accessed through pirate websites.

> *Theoretically, we can find every locked paper on Sci-Hub and take those pictures, so as XX said, you signed that it can be used anywhere.* [F3/P18, student]

Participants noted that most patients were unaware of the implications of having their images published in an online article, including the meaning and impact of Creative Commons licenses. By contrast, some participants believed that patients might not be concerned with the details of how their images are used.

> *I don't think many people think about this.* [F4/P26, student]
>
> *Patients are not aware of the availability of their images.* [F2/P9, doctor]
>
> *I think it doesn't matter what kind of license it is. I think it means nothing to the patient.* [F2/P11, doctor]
>
> *Patients are generally happy when they realize they have ended up in a scientific journal because it means that their case was really special. And then they brag, I am in a book, I am there, they showed my image at the congress.* [F2/P13, doctor]

XSL•FO
RenderX

Building on the idea that patients may be unaware of potential risks, some participants expressed the view that the responsibility lies with the manuscript author, physician, or the individual obtaining informed consent for image use. Some potential solutions to the risks of image misuse that were suggested were "warn the patient" [F3/P18, student], "to explain as much as possible to people because not everyone has the same amount of knowledge about the Internet" [F1/P4, patient], "communication with the patient, check everything" [F4/P26, student], and "check what the licences are that the journal has and maybe choose those that have a more strict criteria on further usage on the photos" [F4/P24, student].

Others said that the responsibility lies with the party that is culpable for the unauthorized use of the image, and suggested "arrange that it is punishable by law if someone's image is published on an advertisement without permission" [F3/P22, student] and "it should somehow be regulated by law" [F3/P21, student].

A third proposed solution was to enhance patient autonomy by giving them greater control over how their images are used. For example, a participant suggested allowing patients to specify where their images could or could not be shared: "I [could] literally mark where I want and where I don't want [the image to be shared]" [F1/P4, student]. Another offered that one could "allow a certain deadline for withdrawing the image if the patient changes his mind within a week" [F1/P4, student], and a third participant suggested that "In the informed consent, put in a special box for publication for commercial purposes, that it can be published in open-access" [F3/P19, student].

Anonymization Techniques Have Limitations

Building on the previous theme of patient images in online journals being vulnerable to misuse, a related concern emerged regarding anonymization techniques for photographs where the face is fully or partially visible. Participants expressed mixed opinions on whether a patient's face could or should be protected using methods such as a black bar over the eyes. While some viewed these techniques as somewhat helpful, the overall sentiment was that they offer only limited protection and may not fully safeguard patient identities.

Some participants, particularly those from the patient stakeholder group, agreed that a patient's identity should be protected as much as possible. As a result, they viewed anonymization techniques as a positive tool with the potential to at least partially achieve this goal:

> *I would put, I would use that anonymization system. Then I would feel less as if I was exposing the person.* [F3/P19, student]

> *I would be more inclined to mask the eyes, but with that tape, not with those two black dots. So, some kind of wide black band, such an option, but I would not show the patient completely.* [F2/P9, doctor]

> *I agree that I would feel more comfortable if I had black stripe over my face and I actually think [that], even though is not complete anonymization of the person, I feel like it is more abstract.* [F4/P27, student]

> *[With the black bar over the eyes] I am still not anonymous but at least I wouldn't feel totally exposed like not 100-percent exposed but 80% and that 20% of non-exposure would give me some kind of comfort.* [F4/P25, student]

Building on the idea that anonymization is not highly effective in protecting patient identity, some participants viewed these techniques as potentially diminishing the clinical value of the image, thereby undermining the purpose of including the photograph in the first place:

> *I am the first one who doesn't even want to look at the image when I see an image with those things on patient's eyes in books. It's like I can't see anything.* [F3/P18, student]

> *It's a much better display without that anonymization, which doesn't really have a function. And the patient, if he has already agreed, I don't think it matters at all what the topic is because he was informed about the topic of the paper and about everything that is in the picture that he decided that it was OK for him. I don't think there is any reason to cover his eyes.* [F3/P21, student]

> *I would rather publish it without it because it is somehow simpler...and is a better presentation of the case.* [F3/P23, student]

Additionally, not all cases were perceived the same in terms of anonymization techniques. Some participants distinguished between different patients in photographs or the conditions displayed:

> *But again, it depends on what we want to show. If you're trying to show, I don't know, the result of forehead wrinkle correction, then you have to show the eyes as well so that it can be better displayed all together. If we do anything aesthetically, if he already agreed, I would show everything.* [F3/P18, student]

> *It is some degree of anonymisation, not complete but still I would not want my face to be, you know, especially in some disease or something. (...) It depends on the condition.* [F4/P26, student]

Overall, participants' accounts indicated uncertainty and ambivalence regarding the use of anonymization techniques. Few concrete suggestions were identified; instead, we observed a plurality of personal opinions and preferences on the subject. However, one participant suggested a possible way to reconcile the tension between clinical utility and patient protection: "to reveal only the part of the face that is important for it" [F1/P2, patient].

Discussion

Principal Findings

In this study, 3 main topics emerged from the focus group discussions, highlighting the need to improve existing guidelines and enhance stakeholders' knowledge regarding the use of patient photographs in digital academic journals. While all participants recognized the challenges of maintaining patient

XSL•FO
RenderX

privacy and confidentiality when using patient photographs, they were not fully aware of the potential consequences of publishing such images online, particularly in open-access formats. They viewed anonymization techniques as useful for protecting identity but also acknowledged their limitations in ensuring complete anonymity.

To the best of our knowledge, this is the first study to explore the knowledge and opinions of relevant stakeholders through focus group interviews to better understand the emerging issue of accessible identifiable patient photographs in the digital scientific environment. Previous studies were mostly cross sectional and relied on questionnaires with preset answers, which limited deeper insights into the problem [14]. This methodological approach offers a new perspective on the publication process of patient photographs in scientific journals. Our previous cross-sectional study, which examined opinions on obtaining consent for publishing patient photographs in academic journals, found that patients were generally lenient toward consenting to the publication of their facial photographs [18]. However, in this study, after focus group discussions that provided deeper insights into digital publishing and open-access formats, patients expressed a more rigorous stance on informed consent for the publication of their photographs compared with our earlier findings. They demonstrated a heightened awareness of the potential consequences of publishing their photographs in digital scientific journals and emphasized the need for full disclosure before providing written consent. This shift in perspective may, at least in part, be attributed to the higher education levels of the patients in this study. Furthermore, despite recognizing the limitations of fully anonymizing facial photographs by concealing the eyes, patients still considered it a useful method for providing at least some level of protection. Similarly, a study by Oh et al [21], which examined dermatology patients' views on the clinical and nonclinical use of their photographs, found that patients were particularly concerned about privacy protection in nonclinical contexts, such as scientific publication.

When comparing the different stakeholder groups among the participants, we found that students were more aware of potential issues and had greater knowledge about sharing and publishing information online than patients and doctors. This was not unexpected, given their younger age and greater familiarity with digital technology. Recent studies on e-professionalism among medical and dental students and trainees have shown that they frequently use social media to communicate, learn, and share experiences, often by sharing information and patient data [22-24], which could include articles in open-access journals containing patient photographs. Discrepancies between students' self-reported behaviors and their actual practices in social media activities have been noted, highlighting the need for clearer guidelines on ethical behavior in the digital environment [22]. Similarly, students in our study reported a lack of practical knowledge on how to ethically publish identifiable patient photographs in scientific journals, indicating that the future generation of doctors is aware of this issue. They also emphasized the need for clear and practical ethical guidelines on digital publishing in medicine and expressed openness to learning how to effectively protect their patients' privacy in the digital environment.

In contrast to students and patients, medical and dental doctors predominantly believed that patients were unlikely to find or understand academic publications or the publication process in medicine. They considered themselves better positioned to make decisions on behalf of patients regarding the publication of their photographs. However, the doctors themselves demonstrated limited knowledge of available resources on ethical best practices in digital publishing. Although doctors are regarded as key stakeholders in this process [25], some believed that obtaining patient consent for using clinical photographs was not mandatory, and they were uncertain about how to address issues related to the misuse of such images. This raises concerns that doctors may not be adequately equipped to inform patients about the publication of their photographs in online journals or to obtain valid informed consent. The importance of obtaining such consent varies across countries [26,27], suggesting that cultural differences may influence ethical requirements for publishing patient photographs. Previous studies indicate that existing ethical guidelines are not uniformly implemented, highlighting the need for revision and improvement [14-16]. Additionally, publication requirements in scientific journals may sometimes be stricter than local ethical policies, potentially complicating the publication process. Furthermore, online publishing is evolving more rapidly than practical ethical guidelines are being reviewed, updated, and implemented, creating an ongoing challenge that requires continuous scrutiny and engagement. Particularly relevant is the increasing use of AI and deep learning methods in handling patient photographs. Several studies have already demonstrated the successful application of AI in detecting dental caries from patient photographs [28]. At the same time, concerns have been raised about the risks AI poses to patient privacy, as its use on photographs constitutes a form of secondary data usage that may not be adequately covered by existing privacy and patient protection guidelines [29]. A key risk associated with AI tools in dental imaging is the potential for patient reidentification or the inclusion of unprotected patient data in training data sets without proper consent. However, not all implications of AI are negative. An AI tool has been developed to generate a "digital mask" to enhance patient anonymity [30], demonstrating AI's potential to help address some of the privacy and consent challenges associated with patient photographs.

Patient involvement in decision-making regarding diagnostic and treatment procedures is becoming increasingly important [27]. Qualitative studies exploring patients' perspectives on the use of medical images—both for understanding recommended clinical procedures and for nonclinical purposes—have shown that most patients prefer to be involved in such decisions [21,31-33]. Although these studies had different research focuses, their findings align with those of our study, in which patients emphasized the importance of being fully informed before consenting to the publication of their photographs. In our study, participants also expressed concerns about the potential misuse of patient photographs, even when published under most open Creative Commons licenses. This highlights the importance of providing comprehensive information about

XSL•FO
RenderX

all possible consequences. When developing practical guidelines for protecting patients with identifiable photographs in online scientific publishing, it would be advisable to identify key scenarios that require separate, targeted guidelines—considering factors such as the body part exposed and the applicable anonymization techniques. Similar to their involvement in medical treatment and research, patients should also be included as key stakeholders in the development of new guidelines on the ethical publication of their photographs in online scientific journals [34].

A limitation of this study is that the patient focus group included participants with a university degree, whereas survey studies investigating opinions on identifying photographs primarily included participants with a high school education [14]. The educational level can influence individuals' opinions and knowledge about obtaining consent for the publication of identifying photographs in scientific journals. However, our study also highlights that even highly educated individuals, as well as young people—who are generally expected to be more proficient in using online information sources—still have limited awareness and understanding of the consequences of open-access publishing of identifying photographs. We also did not ask participants for their opinions on publishing identifiable photographs of children, which is a significant ethical issue, particularly in cases of rare and complex clinical conditions where sharing valuable information is crucial [35]. Additionally, the focus group interviews were conducted online rather than in person. However, this did not impact the quality of the discussion, as we used reliable technology that ensured high-quality audio transcripts, and participants were already accustomed to this format due to the COVID-19 pandemic [36]. Finally, while the study was conducted at a single medical school in a country, its findings can likely be generalized to many countries with publicly funded health care systems and to most countries in Europe, particularly in Central, Eastern, and Southeastern Europe [37].

Conclusions

All stakeholders in our study—patients, doctors, and students as future doctors—lacked sufficient knowledge about best practices for the ethical publication of patient photographs in scientific journals, particularly when published in open-access formats. They also expressed considerable uncertainty about how to navigate such situations in practice. Even doctors at the university hospital were unfamiliar with the implications of open-access publication under open-access licenses, which suggests that they may be unable to adequately inform patients during the consent process to obtain valid consent for research. Based on these findings, it is crucial to enhance the education of health care professionals on the evolving digital landscape of publishing, different models of patient-doctor relationships, and ethical practices for protecting patient-sensitive data such as photographs—not only at the graduate education level but also throughout their professional careers.

Health care professionals should be equipped to adequately protect patients by discussing potential privacy violations when their photographs are published in open-access journals. Institutional review boards and research ethics committees should ensure that researchers address this critical ethical issue at the protocol stage of their studies. Publishers and journal editors should not only implement best practices to ensure fully informed patient consent for publishing identifiable photographs but also develop technical and governance safeguards. The increasing use of AI in handling patient photographs warrants special attention, as it introduces new and potentially unforeseen risks to patient privacy. However, AI tools also have the potential to enhance privacy protections and should be further explored for their utility in this context. This study may serve as a foundation for future interventional studies and expert panel surveys aimed at finding an appropriate balance between protecting patient privacy and facilitating the dissemination of new medical knowledge.

Acknowledgments

We thank our patients and colleagues who accepted our invitation to participate in the focus group discussions. This study is funded by the grant of Croatian Research Foundation "Professionalism in Health: Decision Making in Practice and Research" (principal investigator: AM [grant IP-2014-09-7672]).

Data Availability

To preserve participants' privacy and confidentiality, only anonymized transcripts of the focus groups are available from the authors on reasonable request.

Authors' Contributions

MR contributed to study planning and organization, recruitment of participants, conducting the interviews, data analysis, interpretation of the results, and writing the manuscript. DŠ contributed to recording and transcribing interviews, data analysis, interpretation of the results, and writing the manuscript. IB contributed to study planning, conducting the interviews, interpretation of the results, and writing the manuscript. MFŽ contributed to data analysis, interpretation of the results, and writing the manuscript. AT contributed to data analysis, interpretation of the results, and writing the manuscript. AM contributed to study planning, data analysis, interpretation of the results, writing the manuscript, and supervised all phases of the study.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 201 KB - jmir_v27i1e59970_app1.pdf]

Multimedia Appendix 2
The Interview guide and interview questions for qualitative, focus group study.
[DOCX File , 18 KB - jmir_v27i1e59970_app2.docx]

References

1. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. ICMJE. 2021. URL: http://www.icmje.org/icmje-recommendations.pdf [accessed 2024-01-06]
2. Roguljić M. Ethical considerations regarding the publication of identifiable patient photographs in academic journals. Doctoral thesis. University of Split School of Medicine. 2020 Dec 01. URL: https://neuron.mefst.hr/docs/graduate_school/tribe/Repozitorij_Doktorata/Roguljic_Marija_thesis.pdf [accessed 2020-01-06]
3. Tegze J. How to identify an unknown person using a photograph. Jan Tegze (Blog post). 2019. URL: https://jantegze.medium.com/how-to-find-out-someones-name-from-a-picture-5472fcc60fe2 [accessed 2024-01-06]
4. Adeyemo W, Mofikoya B, Akadiri O, James O, Fashina A. Acceptance and perception of Nigerian patients to medical photography. Dev World Bioeth 2013 Dec 18;13(3):105-110. [doi: 10.1111/j.1471-8847.2012.00328.x] [Medline: 22708667]
5. Hacard F, Maruani A, Delaplace M, Caille A, Machet L, Lorette G, et al. Patients' acceptance of medical photography in a French adult and paediatric dermatology department: a questionnaire survey. Br J Dermatol 2013 Aug;169(2):298-305. [doi: 10.1111/bjd.12345] [Medline: 23551168]
6. Lau CK, Schumacher HHA, Irwin MS. Patients' perception of medical photography. J Plast Reconstr Aesthet Surg 2010 Jun;63(6):e507-e511. [doi: 10.1016/j.bjps.2009.11.005] [Medline: 19959408]
7. Leger M, Wu T, Haimovic A, Kaplan R, Sanchez M, Cohen D. Patient perspectives on medical photography in dermatology. Dermatol Surg 2014;40(9):1028-1037. [doi: 10.1097/01.dss.0000452632.22081.79]
8. Dumestre DO, Fraulin F. Avoiding breach of patient confidentiality: trial of a smartphone application that enables secure clinical photography and communication. Plast Surg (Oakv) 2020 Feb;28(1):12-18 [FREE Full text] [doi: 10.1177/2292550319880910] [Medline: 32110641]
9. Hopper KD, TenHave TR, Hartzel J. Informed consent forms for clinical and research imaging procedures: how much do patients understand? AJR Am J Roentgenol 1995 Feb;164(2):493-496. [doi: 10.2214/ajr.164.2.7839996] [Medline: 7839996]
10. Wang Y, Tan H, Yang X. Perception and acceptability of medical photography in Chinese dermatologic patients: a questionnaire survey. Dermatol Surg 2017 Mar;43(3):437-442. [doi: 10.1097/DSS.0000000000000984] [Medline: 28099200]
11. Committee on Publication Ethics (COPE). Journals' best practices for ensuring consent for publishing medical case reports: guidance from COPE. COPE. Hampshire, UK: Committee on Publication Ethics (COPE); 2024 Jan 06. URL: https://publicationethics.org/resources/guidelines/journals%E2%80%99-best-practices-ensuring-consent-publishing-medical-case-reports [accessed 2025-02-25]
12. Health Research Authority (HRA). Transparency. HRA. 2018. URL: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/transparency/ [accessed 2024-01-06]
13. Unites States Department of Health and Human Services. Summary of the HIPAA privacy rule: Unites States Department of Health and Human Services. Unites States Department of Health and Human Services. Washington, DC: Unites States Department of Health and Human Services; 2024 Jan 06. URL: https://www.hhs.gov/sites/default/files/privacysummary.pdf [accessed 2025-02-25]
14. Roguljić M, Šimunović D, Poklepović Peričić T, Viđak M, Utrobičić A, Marušić M, et al. Publishing identifiable patient photographs in scientific journals: scoping review of policies and practices. J Med Internet Res 2022 Aug 31;24(8):e37594 [FREE Full text] [doi: 10.2196/37594] [Medline: 36044262]
15. Marshall Z, Brunger F, Welch V, Asghari S, Kaposy C. Open availability of patient medical photographs in Google Images search results: cross-sectional study of transgender research. J Med Internet Res 2018 Feb 26;20(2):e70 [FREE Full text] [doi: 10.2196/jmir.8787] [Medline: 29483069]
16. Marshall Z, Bhattacharjee M, Wang M, Cadri A, James H, Asghari S, et al. Finding medical photographs of patients online: randomized, cross-sectional study. J Med Internet Res 2024 Jun 24;26:e55352 [FREE Full text] [doi: 10.2196/55352] [Medline: 38913416]
17. About the licenses. Creative Commons. 2021. URL: https://creativecommons.org/licenses/ [accessed 2024-01-06]

XSL•FO
RenderX

18. Roguljić M, Peričić TP, Gelemanović A, Jukić A, Šimunović D, Buljan I, et al. What patients, students and doctors think about permission to publish patient photographs in academic journals: a cross-sectional survey in Croatia. Sci Eng Ethics 2020 Jun 20;26(3):1229-1247. [doi: 10.1007/s11948-019-00134-y] [Medline: 31541413]
19. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
20. Boeije HA. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. Quality & Quantity 2002;36:391409-391409. [doi: 10.1023/A:1020808529486]
21. Oh Y, Sun M, Gu L, Salvador T, Sar-Graycar L, Hay JL, et al. Dermatology patient perspectives regarding consent, privacy, security, and identifiability of clinical photography at a tertiary cancer center: a qualitative study. JAAD Int 2023 Sep;12:121-123 [FREE Full text] [doi: 10.1016/j.jdin.2023.05.010] [Medline: 37409316]
22. Kamarudin Y, Mohd Nor NA, Libamin AC, Suriani ANH, Marhazlinda J, Bramantoro T, et al. Social media use, professional behaviors online, and perceptions toward e-professionalism among dental students. J Dent Educ 2022 Aug 05;86(8):958-967. [doi: 10.1002/jdd.12912] [Medline: 35247218]
23. Shenouda JEA, Davies BS, Haq I. The role of the smartphone in the transition from medical student to foundation trainee: a qualitative interview and focus group study. BMC Med Educ 2018 Jul 31;18(1):175-175 [FREE Full text] [doi: 10.1186/s12909-018-1279-y] [Medline: 30064424]
24. Villamizar PJ, Moreno SM, Moreno F. Manejo de las redes sociales electrónicas por parte de los estudiantes de medicina: el caso de la publicación de fotografías con los pacientes y el profesionalismo médico. Biomedica 2015 Sep 09;36(1):140-148. [doi: 10.7705/biomedica.v36i1.2646]
25. Burns K, Belton S. Clinicians and their cameras: policy, ethics and practice in an Australian tertiary hospital. Aust Health Review 2013;37(4):437. [doi: 10.1071/ah12039]
26. Costa ED, Martins LAC, Cral WG, Peroni LV, Freitas DQ, Oliveira ML. Assessment of dentists' behaviour on the use of patients' images. Eur J Dent Educ 2020 Aug 23;24(3):513-517. [doi: 10.1111/eje.12530] [Medline: 32304263]
27. Devakumar D, Brotherton H, Halbert J, Clarke A, Prost A, Hall J. Taking ethical photos of children for medical and research purposes in low-resource settings: an exploratory qualitative study. BMC Med Ethics 2013 Jul 09;14(1):27-27. [doi: 10.1186/1472-6939-14-27] [Medline: 23835013]
28. Moharrami M, Farmer J, Singhal S, Watson E, Glogauer M, Johnson AEW, et al. Detecting dental caries on oral photographs using artificial intelligence: a systematic review. Oral Dis 2024 May;30(4):1765-1783. [doi: 10.1111/odi.14659] [Medline: 37392423]
29. Radiological Society of North America. Protecting patient privacy in the era of artificial intelligence. Radiological Society of North America. 2201 Feb. URL: https://www.rsna.org/news/2021/february/protecting-patient-privacy [accessed 2024-11-20]
30. Yang Y, Lyu J, Wang R, Wen Q, Zhao L, Chen W, et al. A digital mask to safeguard patient privacy. Nat Med 2022 Sep 15;28(9):1883-1892 [FREE Full text] [doi: 10.1038/s41591-022-01966-1] [Medline: 36109638]
31. Price A, Clarke M, Staniszewska S, Chu L, Tembo D, Kirkpatrick M, et al. Patient and pblic involvement in research: a journey to co-production. Patient Educ Couns 2022 Apr;105(4):1041-1047. [doi: 10.1016/j.pec.2021.07.021] [Medline: 34334264]
32. Phelps EE, Wellings R, Kunar M, Hutchinson C, Griffiths F. A qualitative study exploring the experience of viewing three-dimensional medical images during an orthopaedic outpatient consultation from the perspective of patients, health care professionals, and lay representatives. J Eval Clin Pract 2021 Apr 02;27(2):333-343. [doi: 10.1111/jep.13417] [Medline: 32488922]
33. Carlin LE, Smith HE, Henwood F. To see or not to see: a qualitative interview study of patients' views on their own diagnostic images. BMJ Open 2014 Jul 31;4(7):e004999 [FREE Full text] [doi: 10.1136/bmjopen-2014-004999] [Medline: 25082418]
34. Boivin A, Richards T, Forsythe L, Grégoire A, L'Espérance A, Abelson J, et al. Evaluating patient and public involvement in research. BMJ 2018 Dec 06;363:k5147. [doi: 10.1136/bmj.k5147] [Medline: 30522999]
35. Schoeman L, Honey EM, Malherbe H, Coetzee V. Parents' perspectives on the use of children's facial images for research and diagnosis: a survey. J Community Genet 2022 Dec 10;13(6):641-654 [FREE Full text] [doi: 10.1007/s12687-022-00612-0] [Medline: 36214965]
36. Carter M, Shih P, Williams J, Degeling C, Mooney-Somers J. Conducting qualitative research online: challenges and solutions. Patient 2021 Nov;14(6):711-718 [FREE Full text] [doi: 10.1007/s40271-021-00528-w] [Medline: 34114170]
37. International Monetary Fund (IMF). Central, Eastern, and Southeastern Europe: how to get back on the fast track. IMF. 2016 May. URL: https://www.imf.org/external/pubs/ft/reo/2016/eur/eng/pdf/rei0516.pdf [accessed 2024-01-15]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research

XSL•FO
RenderX

ICMJE: International Committee for Medical Journal Editors
USSM: University of Split School of Medicine

Edited by A Mavragani; submitted 27.04.24; peer-reviewed by A Grilo, M Bak; comments to author 07.11.24; revised version received 28.11.24; accepted 16.12.24; published 05.03.25.

Please cite as:
Roguljić M, Šimunović D, Buljan I, Žuljević MF, Turić A, Marušić A
Publishing Identifiable Patient Photographs in the Digital Age: Focus Group Study of Patients, Doctors, and Medical Students
J Med Internet Res 2025;27:e59970
URL: https://www.jmir.org/2025/1/e59970
doi:10.2196/59970
PMID:40053737

©Marija Roguljić, Dina Šimunović, Ivan Buljan, Marija Franka Žuljević, Antonela Turić, Ana Marušić. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Association Between COVID-19 Vaccination Uptake and Information-Seeking Behaviors Using the Internet: Nationwide Cross-Sectional Study

Kazuya Taira[1], PHN, RN, PhD; Misa Shiomi[1], PHN, RN, PhD; Takayo Nakabe[2,3], MPH, PhD; Yuichi Imanaka[2,4], MD, MPH, PhD

[1]Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Healthcare Economics and Quality Management, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[3]The Database Center of the National University Hospitals, University of Tokyo Hospital, Tokyo, Japan

[4]Department of Health Security System, Centre for Health Security, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Corresponding Author:
Kazuya Taira, PHN, RN, PhD
Human Health Sciences
Graduate School of Medicine
Kyoto University
53, Shogoinkawara-cho, Sakyo-ku
Kyoto, 606-8501
Japan
Phone: 81 075 751 3927
Fax: 81 075 751 3927
Email: taira.kazuya.5m@kyoto-u.ac.jp

Abstract

Background: The COVID-19 pandemic, declared in March 2020, profoundly affected global health, societal, and economic frameworks. Vaccination became a crucial tactic in combating the virus. Simultaneously, the pandemic likely underscored the internet's role as a vital resource for seeking health information. The proliferation of misinformation on social media was observed, potentially influencing vaccination decisions and timing.

Objective: This study aimed to explore the relationship between COVID-19 vaccination rates, including the timing of vaccination, and reliance on internet-based information sources in Japan.

Methods: Using a cross-sectional study design using a subset of panel data, this nationwide survey was conducted in 7 waves. A total of 10,000 participants were randomly selected through an internet survey firm, narrowing down to 8724 after applying inclusion and exclusion criteria. The primary outcome was the COVID-19 vaccination date, divided into vaccinated versus unvaccinated and early versus late vaccination groups. The main exposure variable was the use of internet-based information sources. Control variables included gender, family structure, education level, employment status, household income, eligibility for priority COVID-19 vaccination due to pre-existing medical conditions, and a health literacy scale score. Two regression analyses using generalized estimating equations accounted for prefecture-specific correlations, focusing on vaccination status and timing. In addition, chi-square tests assessed the relationship between each information source and vaccination rates.

Results: Representing a cross-section of the Japanese population, the regression analysis found a significant association between internet information seeking and higher vaccination rates (adjusted odds ratio [aOR] 1.42 for those younger than 65 years; aOR 1.66 for those aged 65 years and older). However, no significant link was found regarding vaccination timing. Chi-square tests showed positive associations with vaccination for television, government web pages, and web news, whereas blogs and some social networking sites were negatively correlated.

Conclusions: Internet-based information seeking is positively linked to COVID-19 vaccination rates in Japan, underscoring the significant influence of online information on public health decisions. Nonetheless, certain online information sources, including blogs and some social networks, negatively affected vaccination rates, warranting caution in their use and recognition. The study highlights the critical role of credible online sources in public health communication and the challenge of combating

XSL•FO
RenderX

misinformation on less regulated platforms. This research sheds light on how the digital information landscape influences health behaviors, stressing the importance of accurate and trustworthy health information amidst global health emergencies.

(J Med Internet Res 2025;27:e59352) doi:10.2196/59352

KEYWORDS

COVID-19 vaccines; internet use; information seeking behavior; Japan; vaccine; COVID-19; behavior; panel study; longitudinal; survey; regression analysis; chi-square test; adult; epidemiology; health informatics

Introduction

COVID-19 Pandemic and Vaccination

The COVID-19 pandemic, declared by the World Health Organization in March 2020 [1], has had an unprecedented impact globally, affecting billions of lives and causing widespread health, social, and economic disruptions. Vaccination against COVID-19 emerged as a pivotal strategy in controlling the virus's spread, with over 10 billion vaccine doses administered worldwide by the end of 2022 [2]. The pandemic's toll has been significant, infecting more than 30 million people and resulting in over 70,000 deaths [3]. In addition, 104,736,436 people, or 80.7% of the population, had received their first vaccine dose by this study [4].

The Role of the Internet in Public Health and Theoretical Framework

During the COVID-19 pandemic, the way individuals sought health information underwent significant changes. The internet became an indispensable tool, playing a crucial role in public health by serving as the primary means by which people could register for vaccination [5,6]. This shift in information-seeking behavior is essential for understanding the public's response to health crises [7,8]. Meanwhile, the World Health Organization introduced a framework to address infodemics, characterized by the widespread dissemination of inaccurate information through digital and physical channels [9]. Similarly, the National Academy of Medicine offered recommendations to combat misinformation and disinformation on social media [10]. Indeed, the internet's influence on health behavior is profound, with misinformation notably undermining confidence in COVID-19 vaccinations [11].

This study is based on the Digital Divide Theory [12] and the Health Belief Model [13]. Digital Divide Theory indicates that disparities in Internet access and digital literacy are related to differences in access to health information and health behaviors. This theory encompasses not only the differences in physical access but also the concept of the second-level divide, which includes the necessary skills and literacy of users [14]. In addition, the Health Belief Model emphasizes factors that influence individuals to engage in health behaviors, such as perceived risk, perceived benefits, perceived barriers, and cues to action. Our research is based on the idea that individuals who can gather information through the internet are more likely to understand the risks and benefits of vaccination and, as a result, experience behavior change.

Information-Seeking Behavior Using the Internet and Vaccination

The association between vaccination uptake and information-seeking behaviors using the internet is a critical area of study, particularly in the context of the COVID-19 pandemic. Previous research has demonstrated the significant impact of online health information on vaccination behaviors across different vaccines [15-20]. For instance, individuals who actively seek formal health information from credible online sources and engage with health care providers are more likely to get vaccinated in the influenza vaccine [15]. Intervention studies have been conducted on Tetanus, diphtheria, acellular pertussis, and influenza vaccination during pregnancy, and it has also been reported that web-based vaccination information, including websites and interactive social networking services (SNS), can improve vaccination uptake [16]. Conversely, reliance on SNS for vaccination information has been associated with lower vaccination rates, such as the pertussis vaccine uptake during pregnancy, indicating that informal online sources may contribute to vaccine hesitancy [17]. Other studies have also yielded mixed results, with internet information gathering having a positive effect on pneumococcal vaccination of adults with heart diseases [18] and a negative effect on a mother's decision to vaccinate her daughter with human papillomavirus [19,20]. These findings underscore the complexity of online information's role in vaccination behavior and highlight the necessity of distinguishing between formal and informal sources to understand their differing impacts fully.

For COVID-19 vaccination, a study indicates that the behavior of seeking health information on the internet is positively linked to vaccination uptake in China [21]. Various studies have identified vaccination hesitancy as a concern, noting a reduction in hesitancy when the internet is the primary source of information about COVID-19 [22]. Conversely, some studies have highlighted the negative effects of social media on vaccine hesitancy [23,24]. In Japan, reliance on internet news and video sharing sites was found to exacerbate hesitancy toward COVID-19 vaccination among individuals younger than 65 years [25].

These findings indicate that the effects of internet-based information sources on COVID-19 vaccination uptake are mixed and inconclusive. There is a need for further analysis, as the impact appears to vary depending on the type of internet information sources, such as web news, social media, and video sharing sites. In addition, no studies have reported an association with the timing of vaccination, whether late or early, or with the presence or absence of vaccination.

XSL•FO
RenderX

This study aims to investigate the association between COVID-19 vaccination uptake or timing and the use of internet-based information sources in more detail.

Methods

Study Design and Participants' Setting

This study used a cross-sectional study design using a subset of panel data, conducting nationwide web-based surveys 7 times from the first wave (October to November 2020) to the seventh wave (February to March 2022). The surveys were closed surveys facilitated through an internet research company, Rakuten Insight Inc, using their proprietary web platform and a self-administered, anonymized web form. The data across these panels were linked using user identification numbers known only to the research company. Rakuten Insight is among the largest internet research services, boasting 2.2 million registered active users as of September 2022. In this study, only complete cases with no unanswered items were collected, and multiple responses from the same person were eliminated using panel monitor registration information and IP addresses. Primarily, this study used data from the sixth panel (survey conducted between November 5 and 25, 2021), which coincided with the conclusion of the public's initial COVID-19 vaccination phase.

The survey initially sampled 10,000 participants, ensuring an equal distribution across gender, age, and prefectures through random selection during the first wave. If participants dropped out in subsequent surveys after the second wave, resampling targeted individuals who had responded in any of the previous waves. Should a shortfall persist, new samples were recruited, with replacements selected from the closest matching age group within the same prefecture and gender category. In the event of dropouts during the follow-up survey, new samples were added as necessary to maintain a total of 10,000 participants for each panel survey. The decision to establish a sample size of 10,000 was informed by the need for random sampling to ensure equitable allocation and representativeness, taking into account the distribution and prevalence of monitors. This sample size also facilitates effective stratification across age, gender, and prefectural demographics, enabling robust multivariate analysis.

Inclusion and Exclusion Criteria

The eligibility criteria for the survey, aimed at examining the psychological landscape of ethical values and anxiety, required participants to be adults aged 20 years or older. This age threshold was established to ensure the reliability of the feedback received. Furthermore, to reduce sampling bias inherent in internet-based surveys, professionals working in survey or advertising agencies were excluded from participation, which was in line with the regulations set by internet survey companies. The variable related to medical history was cross-referenced using user IDs from the fifth survey panel, conducted between September and October 2021. As a result, any cases where this variable was missing were systematically excluded from the analysis. Medical and caregiving professionals were also excluded from the analysis because they received vaccinations before the general public and possessed specialized expertise and unique access to information. In addition, participants who reported their vaccination dates as being before April 14, 2021 (the start date of mass vaccination) were excluded, as these entries were considered input errors.

Measures

Main Outcome

The primary outcome variable was the date of COVID-19 vaccination. This variable was converted into two binary variables: (1) one indicating whether individuals were vaccinated or not (vaccination uptake) and (2) the other differentiating between dates before and after the 75th percentile (vaccination timing).

In this study, the dependent variable of vaccination timing is defined based on Everett Rogers' Diffusion of Innovation theory [26]. According to this theory, the adoption of new innovations follows a specific pattern, and the population is categorized into 5 adopter categories that consist of innovators, early adopters, early majority, late majority, and laggards.

Given the unprecedented speed and scale of the COVID-19 vaccination rollout, we aimed to identify the segment of the population that adopted the vaccination last, corresponding to the "Laggards" category in Rogers' theory. This category is generally characterized by lower socioeconomic status, personality challenges, and limited social interaction, making its identification meaningful. Statistical methods such as mean and SD are proposed to determine the Laggards category. Considering the timing that most elderly individuals had completed their vaccinations and younger individuals were starting theirs, we determined that defining this group using the fourth quartile was more appropriate.

Exposure

The exposure variable was identified as the source of information related to COVID-19. Participants were asked to select their sources of information from a list provided in a multiple-choice format (Textbox 1). Based on their responses, the options were categorized into two groups, those that involved information seeking through the internet and those that did not. The "with Internet" category included participants who selected at least one of the following options, Twitter, Facebook (Meta), LINE News (LY Corporation), other social networking services, YouTube (Google), COCOA (COVID-19 Contact-Confirming Application, Ministry of Health, Labour and Welfare), web pages of local governments or the Ministry of Health, Labour and Welfare, web pages of professional associations, blogs by medical professionals, personal blogs (nonprofessional), and web news. Participants who selected only options outside of these were classified as engaging in information seeking without the internet.

XSL•FO
RenderX

Textbox 1. Ways of obtaining COVID-19–related information.

Television, Radio, Magazines/Books, Asahi Shimbun, Mainichi Shimbun, Yomiuri Shimbun, Sankei Shimbun, Nikkei Shimbun, Other newspapers, Twitter, Facebook, LINE news, Other social networking services, YouTube, COCOA (COCOA is an abbreviation for COVID-19 Contact-confirming Application, a smartphone application produced by the Japanese government for COVID-19 measures.), WEB pages of local governments or Ministry of Health, Labour and Welfare, WEB pages of professional organizations, Blogs of medical professional, Blogs of personal (nonprofessional), WEB news, Information from friends and acquaintances, Information from family members, Others

Control Variables

The control variables included gender, family structure, educational background, employment status, household income, pre-existing medical conditions eligible for priority COVID-19 vaccination (Multimedia Appendix 1), and the 14-item health literacy scale for Japanese adults (HLS-14) [27]. Family structure was categorized as either living alone or living with family members. Educational background was classified into three categories, less than a bachelor's degree, associate or bachelor's degree or higher, and unwilling to answer. Household income was defined as no income, less than 3 million yen (approximately US $19,355, at an exchange rate of 155 yen/US dollar), 3 million yen to 5 million yen (approximately US $19,355 to $32,258), 5 million yen to 8 million yen (approximately US $32,258 to $51,613), 8 million yen (approximately US $51,613) or more, or unwilling to answer. The HLS-14 measures health literacy across 3 domains: functional health literacy, communicative health literacy, and critical health literacy. Respondents rate each item on a 1 to 5 scale, with total scores ranging from 14 to 70 points; higher scores indicate greater health literacy.

Statistical Analysis

In Japan, priority for vaccination was given to individuals aged 65 years and older. Therefore, the analysis differentiated between those aged 65 years and older and those younger than 65 years. Descriptive statistics were used to profile the participants and their vaccination statuses. Two regression analyses generalized estimating equations accounted for prefecture-specific correlations, focusing on vaccination uptake and vaccination timing. In these analyses, the main exposure variable was whether information seeking was conducted with or without the internet. The models were adjusted for gender, family structure, educational background, employment status, household income, eligibility for priority COVID-19 vaccination due to pre-existing medical conditions, and HLS-14 scores. In addition, chi-square tests assessed the relationship between the use of each information source and vaccination uptake.

Ethical Considerations

This study was conducted with the approval of the Kyoto University Medical Ethics Review Committee (approval number: R2670). Informed consent was obtained through the initial screen of the online survey, where participants were informed that completion of the questionnaire implied consent. The data provided to the researchers by the commissioned internet research company was anonymized, ensuring no personally identifiable information was included. Participants received compensation in the form of points from the survey company, in accordance with their internal regulations. The exact number of points awarded was not disclosed to the researchers, as this information is confidential.

Results

Participants Included in the Analysis

The sixth survey, which involved 10,000 respondents, was linked with the data from the fifth survey, which led to the exclusion of 751 cases due to missing information on pre-existing medical conditions. In addition, further exclusions were made, 362 individuals working in medical or caregiving professions and 163 participants who reported vaccination dates before April 12, 2021, were deemed ineligible. As a result, the final sample size for analysis was 8724 participants, comprising 6589 individuals younger than 65 years of age and 2135 aged 65 years and older (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of inclusion and exclusion criteria.

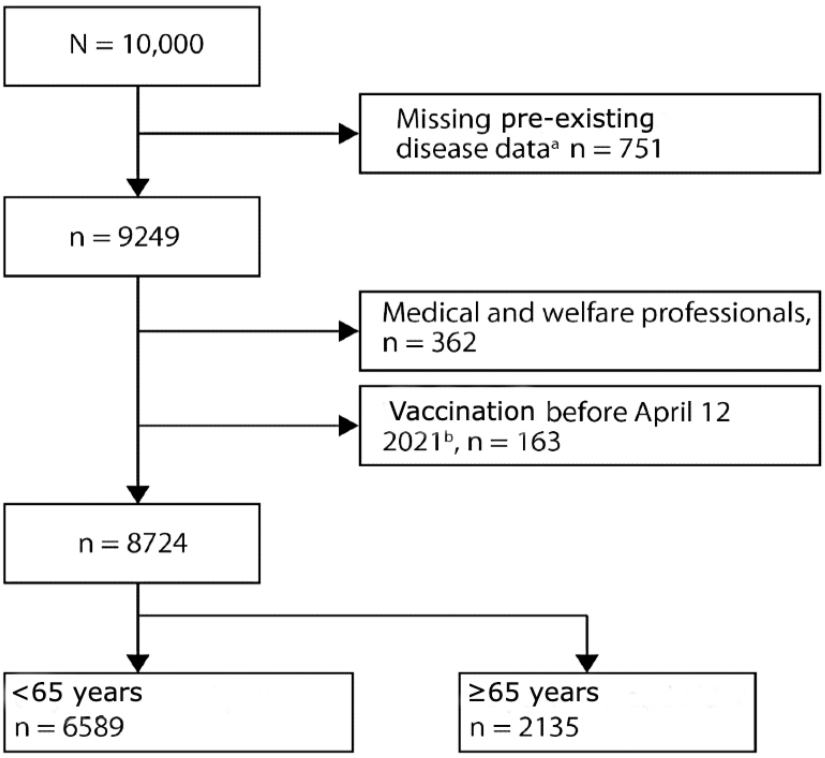

a. Individuals with high-risk pre-existing conditions were prioritized for earlier vaccination in Japanese national measure.
b. The date when the COVID-19 vaccine was approved in Japan and vaccination began.

Demographic Characteristics of the Participants

The analysis of demographic characteristics among the participants showed that of the 6589 individuals younger than 65 years, 3260 (49.5%) were women, with a median age of 44 (IQR 33-55). Among the 2135 respondents aged 65 years and older, 1009 (47.3%) were women, with a median age of 71 (IQR 68-73; Table 1). The use of internet-based information sources did not significantly vary between respondents younger than 65 years and those aged 65 years and older. A total of 42.0% (2765/6589) of participants younger than 65 years and 37.0% (791/2135) of those aged 65 and older used 2 or more sources of information. Television was the most popular information source, used by 86.3% (5684/6589) of respondents younger than 65 years and 94.0% (2006/2135) of the older group. Although there was no significant difference in the types of internet sources used, younger participants were more inclined to use social networking services such as Twitter and LINE news (Multimedia Appendix 2).

XSL•FO
RenderX

Table 1. Participants' demographic characteristics.

	Younger than 65 years			65 years or older		
	Total, (n=6589)	Vaccinated, (n=5325)	Not vaccinated, (n=1264)	Total, (n=2135)	Vaccinated, (n=1921)	Not vaccinated, (n=214)
Female gender, n (%)	3260 (49.5)	2644 (49.7)	616 (48.7)	1009 (47.3)	885 (46.1)	124 (57.9)
Age (years), median (IQR)	44 (33-55)	45 (34-55)	40.5 (29-50)	71 (68-73)	71 (68-73)	71 (68-73)
Health literacy[a], median (IQR)	47 (42-52)	47 (42-53)	43 (42-50)	47 (42-52)	48 (43-53)	45 (42-50.8)
Family structure, n (%)						
With family members	5414 (82.2)	4427 (83.1)	987 (78.1)	1787 (83.7)	1624 (84.5)	163 (76.2)
Living alone	1175 (17.8)	898 (16.9)	277 (21.9)	348 (16.3)	297 (15.5)	51 (23.8)
Educational background, n (%)						
Less than a bachelor's degree	3138 (47.6)	2479 (46.6)	659 (52.1)	1173 (54.9)	1039 (54.1)	134 (62.6)
Associate or bachelor's degree or higher	3205 (48.6)	2679 (50.3)	526 (41.6)	922 (43.2)	848 (44.1)	74 (34.6)
Unanswered	246 (3.7)	167 (3.1)	79 (6.3)	40 (1.9)	34 (1.8)	6 (2.8)
Household income, n (%)						
No income	158 (2.4)	109 (2.0)	49 (3.9)	23 (1.1)	19 (1.0)	4 (1.9)
<3 million yen (approximately US $19,355)	1211 (18.4)	943 (17.7)	268 (21.2)	633 (29.6)	547 (28.5)	86 (40.2)
3-4 million yen (approximately US $19,355 to 32,258)	1426 (21.6)	1113 (20.9)	313 (24.8)	636 (29.8)	580 (30.2)	56 (26.2)
5-7 million yen (approximately US $32,258 to 51,613)	1554 (23.6)	1312 (24.6)	242 (19.1)	285 (13.3)	262 (13.6)	23 (10.7)
≥8 million yen (approximately US $51,613)	1068 (16.2)	945 (17.7)	123 (9.7)	154 (7.2)	145 (7.5)	9 (4.2)
Unanswered	1172 (17.8)	903 (17.0)	269 (21.3)	404 (18.9)	368 (19.2)	36 (16.8)
Risk of medical history for severe COVID-19, n (%)						
Not high risk	5283 (80.2)	4240 (79.6)	1043 (82.5)	1065 (49.9)	934 (48.6)	131 (61.2)
High risk	1306 (19.8)	1085 (20.4)	221 (17.5)	1070 (50.1)	987 (51.4)	83 (38.8)
Employment status, n (%)						
Not employed	1130 (17.1)	883 (16.6)	247 (19.5)	954 (44.7)	844 (43.9)	110 (51.4)
Employed	5459 (82.9)	4442 (83.4)	1017 (80.5)	1181 (55.3)	1077 (56.1)	104 (48.6)
Internet use for information seeking behavior, n (%)						
Does not use	1957 (29.7)	1477 (27.7)	480 (38.0)	700 (32.8)	602 (31.3)	98 (45.8)
Does use	4632 (70.3)	3848 (72.3)	784 (62.0)	1435 (67.2)	1319 (68.7)	116 (54.2)

[a]HLS-14: 14-item health literacy scale.

Summary of Vaccination Uptake as Outcome Variables

Regarding vaccination uptake, 5325 individuals (80.8%, 5325/6589) younger than 65 years and 1921 (90.0%, 1921/2135) aged 65 years and older had been vaccinated. Internet-based information gathering was undertaken by 4632 individuals (70.3%, 4632/5689) younger than 65 years and 1435 (67.2%, 1435/2135) aged 65 or older (Table 1). Furthermore, the quartile values for vaccination dates for those younger than 65 years were: 25th percentile on July 7, 2021, 50th percentile on August 17, 2021, and 75th percentile on September 15, 2021; for those aged 65 years and older, the values were 25th percentile on June 2, 2021, 50th percentile on June 17, 2021, and 75th percentile on July 4, 2021. Consequently, regarding the vaccination timing variable, participants were classified as part of the delayed vaccination group if their vaccination occurred after September 15, 2021, for those younger than 65 and after July 4, 2021, for those aged 65 years and older (Figure 2).

XSL•FO
RenderX

Figure 2. Bar chart showing vaccination timing.

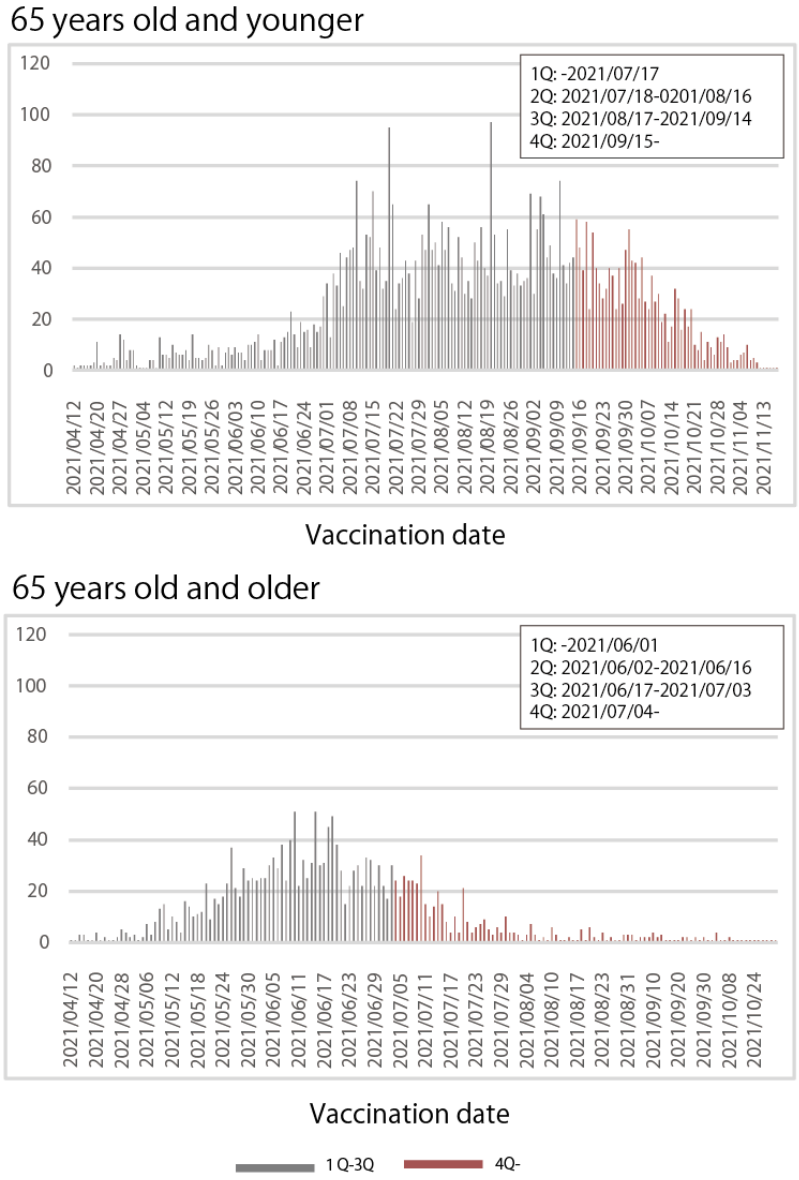

Regression Analysis Results

In the regression analysis with vaccination uptake as the outcome, seeking information through the internet was significantly associated with higher vaccination rates compared with not using the internet for information seeking (adjusted odds ratio [aOR] 1.42, 95% CI 1.27-1.60 for those younger than 65 years; aOR 1.66, 95% CI 1.16-2.37 for those aged 65 years and older). However, when examining vaccination timing as the outcome, there was no significant relationship between using the internet for information and delayed vaccination (aOR 1.12, 95% CI 0.96-1.31 for those younger than 65 years; aOR 1.25, 95% CI 0.98-1.59 for those aged 65 years and older; Table 2).

Chi-square tests assessing the relationship between the use of each information resource and vaccination uptake revealed that TV, government and professional web pages, and web news were more positively associated with COVID-19 vaccination. Conversely, blogs, YouTube, and some social networking sites had a negative association with vaccination uptake (Figure 3).

XSL•FO
RenderX

Table 2. Results of regression analysis using generalized estimating equation model.

Outcomes	Vaccination uptake（Yes=1, No=0）		Vaccination timing（1Q-3Q=1, 4Q=0）	
	Younger than 65 years, aOR[a] (95% CI)	65 years old and older, aOR (95% CI)	Younger than 65 years, aOR (95% CI)	65 years old and older, aOR (95% CI)
Gender, median (IQR)				
Women	Reference	Reference	Reference	Reference
Men	1.19 (1.05-1.34)[b]	0.80 (0.53-1.19)	1.01 (0.90-1.14)	0.99 (0.78-1.24)
Health literacy (HLS-14)[c]	1.03 (1.02-1.03)[b]	1.02 (1.00-1.05)[b]	1.01 (1.00-1.02)[b]	1.01 (1.00-1.03)
Family structure				
With family member	Reference	Reference	Reference	Reference
Living alone	0.8 (0.68-0.94)	0.69 (0.47-1.01)	0.89 (0.74-1.06)	1.04 (0.75-1.45)
Educational background				
Less than a bachelor's degree	Reference	Reference	Reference	Reference
Associate or bachelor's degree or higher	0.83 (0.72-0.96)	0.89 (0.64-1.25)	0.78 (0.69-0.88)[b]	0.85 (0.69-1.06)
Unanswered	1.47 (1.11-1.97)[b]	1.50 (0.62-3.65)	1.35 (0.97-1.88)	1.03 (0.52-2.03)
Household income				
No income	Reference	Reference	Reference	Reference
<3 million yen (approximately US $19,355)	0.76 (0.53-1.08)	0.87 (0.30-2.52)	0.71 (0.43-1.17)	0.62 (0.24-1.59)
3-4 million yen (approximately US $19,355 to 32,258)	0.78 (0.53-1.15)	0.67 (0.23-1.96)	0.82 (0.52-1.3)	0.53 (0.19-1.49)
5-7 million yen (approximately US $32,258 to 51,613)	0.74 (0.5-1.08)	0.63 (0.18-2.23)	0.72 (0.46-1.13)	0.66 (0.23-1.90)
≥8 million yen (approximately US $51,613)	0.41 (0.26-0.64)[b]	0.77 (0.12-4.97)	0.55 (0.34-0.88)[b]	0.56 (0.20-1.61)
Unanswered	0.75 (0.51-1.10)	0.56 (0.19-1.66)	0.66 (0.45-0.97)[b]	0.54 (0.20-1.44)
Risk of medical history for severe COVID-19				
Not high risk	Reference	Reference	Reference	Reference
High risk	1.21 (1.03-1.42)[b]	1.49 (1.17-1.88)[b]	1.78 (1.51-2.11)[b]	1.17 (0.93-1.46)
Employment status				
Not employed	Reference	Reference	Reference	Reference
Employed	0.82 (0.67-1.01)	0.93 (0.66-1.29)	1.20 (1.02-1.42)[b]	1.28 (1.03-1.58)[b]
Internet use for information seeking behavior				
Does not use	Reference	Reference	Reference	Reference
Does use	1.42 (1.27-1.60)[b]	1.66 (1.16-2.37)[b]	1.12 (0.96-1.31)	1.25 (0.98-1.59)

[a]aOR: adjusted odds ratio.

[b]Items with a 95% CI not crossing 1 are considered statistically significant (P<.05).

[c]HLS-14: The 14-item health literacy scale.

XSL•FO
RenderX

Figure 3. Results of chi-square tests between the information resource and vaccination uptake. “Shimbun” refers to a Japanese newspaper. COCOA: COVID-19 Contact-confirming Application; SNS: social networking services.

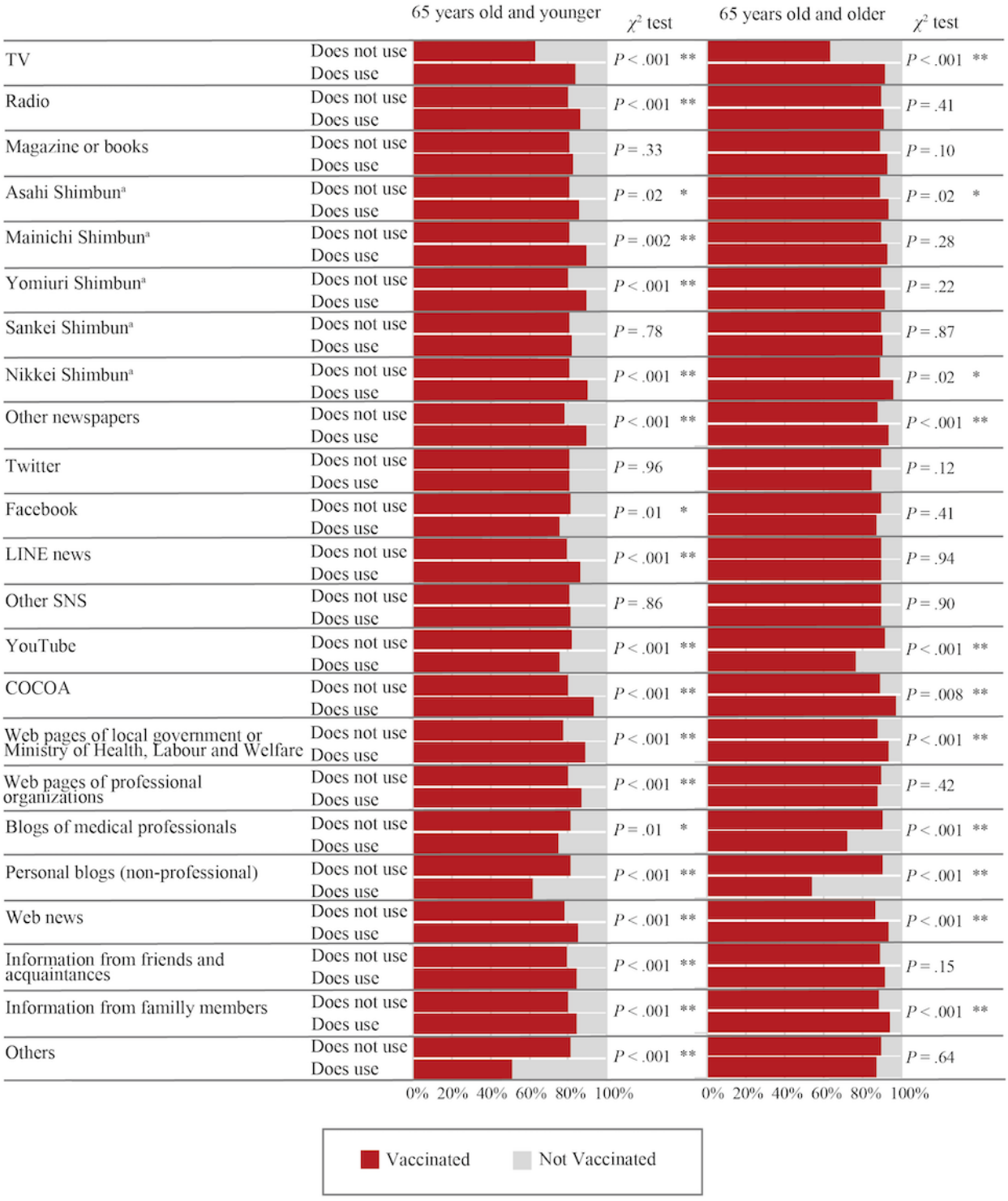

Note: ** refers to $P < .01$, * referes to $P < .05$

Discussion

Principal Results

This study offers a thorough analysis of the relationship between COVID-19 vaccination uptake and the use of various online information resources. Our results indicate that seeking information on the internet is positively associated with higher vaccination rates, highlighting the significant influence of online information on public health decisions. Interestingly, however, our data did not reveal a significant link between online information seeking and the timing of vaccination, suggesting that while online resources contribute to the decision to get vaccinated, they may not affect how promptly individuals act on this decision.

The further analysis presented in Figure 3 showcases the varied impact of different internet resources. Government web pages,

XSL•FO
RenderX

for instance, positively influenced vaccination decisions, emphasizing the critical role of credible, official sources in public health communication. On the contrary, information from blogs and SNS was found to have a negative effect. This highlights concerns about the spread of misinformation on less regulated platforms and underlines the necessity for more efficient measures to counteract misleading health information online. Considering the usage rates of each media type, TV emerged as a predominant information source, with 86.3% (5684/6589) of participants younger than 65 years and 94.0% (2006/2135) of those 65 years and older. This widespread use of TV suggests that even excluding TV's influence, internet usage would likely still be statistically significant in relation to vaccination uptake. Nevertheless, the potential interactive effects between television and specific online sources call for more detailed future investigations.

The insights gained from this study enhance our understanding of how the digital information landscape affects health behaviors and underscore the urgent need to ensure access to accurate and trustworthy health information during global health emergencies.

Comparison With Previous Work

The findings of this study align with those of previous research, indicating that obtaining COVID-19 information on the internet is positively correlated with vaccination uptake [21,28]. Similarly, concerning vaccination hesitancy, the results consistently showed that internet information-seeking behavior plays a positive role in reducing hesitancy [22,23]. While it is plausible that the type of information disseminated across various media, including the internet, may vary by country due to cultural and linguistic differences, these results suggest that the trends observed in Japan mirror those in other nations. However, a study by Hori et al [25] in Japan found that internet web news negatively affected vaccination hesitancy, diverging from some findings of this study. Despite using the same internet research firm for sampling, the exclusion criteria between the two studies significantly differed. Notably, focusing on specific target populations—excluding individuals aged 65 years and older, those who are obese, have pre-existing medical conditions, or have a clear intention regarding vaccination—might yield different impacts on COVID-19 vaccination outcomes. Many of these studies have been conducted in developed economies such as the United States, China, South Korea, Japan, and European Union countries [21-23,25,28], and different results could be obtained if studies were conducted in more diverse countries, including developing economies. Indeed, a paper tracking changes in search trends during the pandemic on Google Trends also reported that search trends were influenced by central government actions and news sources, with some search terms trending differently in different countries [29].

Moreover, reports on information-seeking behavior during the COVID-19 pandemic have indicated that traditional media such as television are more trusted than social media, and Information-seeking behavior through the internet is associated with higher levels of knowledge [30,31]. With regard to SNSs, there is a risk that the sharing of incorrect information may be accelerated because people who are more likely to share information are more confident in their own information and judgment due to a high degree of information sufficiency and trust in the source of information on SNSs [30]. In fact, the outcomes related to social media usage are consistent with previous findings, indicating a negative effect on vaccination rates [23,32]. The growing number of studies highlighting misinformation on social media [33] and the widespread misinformation on critical public health issues [34] advocate for cautious use of social media to prevent reliance on potentially misleading information [32]. However, it is also true that nearly half of people seek COVID-19 information from SNSs [35,36], and it has been shown that vaccination coverage is higher among those who follow reliable information sources on SNSs, so it is important to have a good relationship with SNSs [35]. Moreover, negative impacts have also been reported concerning video sharing sites [25], with some platforms, like YouTube, implementing measures to curb the spread of inaccurate health information by introducing eHealth services [37]. As recommended by the WHO framework [9], a multifaceted approach involving user literacy and ethics, the provision and translation of information by evidence-based professionals, and the analysis and improvement of services by platform operators is essential for fostering a well-informed society in terms of health information.

Therefore, using the internet can be effective in scenarios where vaccination is encouraged as part of a national strategy, such as the COVID-19 response. However, it is crucial to select reliable sources of information and enhance services to ensure the dissemination of accurate information.

Strength and Limitations

The strengths of this study include its close alignment with the demographic distribution of the Japanese population, featuring an aging rate of 24.5% (compared with the national average of 28.9%), a women representation of 49.5% (national average: 51.4%) [38], and a vaccination rate of 83.1% (national average: 75.2% as of November 30, 2021 [6]), ensuring representativeness. A unique aspect of this study is its focus on actual COVID-19 vaccination uptake rather than intentions or hesitancy, analyzing the association with various internet information sources in detail. However, the study is not without limitations. First, it did not thoroughly examine the effect of using multiple information sources on vaccination uptake. Second, the context of the COVID-19 vaccination campaign, being a specific situation with national strategies, including a large-scale vaccination effort, requires caution in generalizing these findings to vaccinations for other infectious diseases. Third, the simultaneous administration of vaccines over a brief period reduced the variance in vaccination timing, potentially making it challenging to identify a relationship between vaccination timing and information sources. While the impact of information sources may extend beyond the context of emerging infectious disease pandemics, as similar outcomes have been reported for other infectious diseases like influenza [39], more comprehensive studies are necessary in the future. Finally, since this survey was conducted through an internet web-form, people who do not use the internet at all were excluded from the survey. As a result, the participants may exhibit higher levels of information-gathering behavior through the internet compared with the general population. However,

XSL•FO
RenderX

this likely represents a conservative effect on the results of this study, suggesting that the true impact of the exposure may be even stronger. Future research using face-to-face or mail survey methods would be desirable to corroborate these findings. Finally, this study might have potential selection bias due to the nature of an internet survey. Individuals, particularly in 65 years old and older, without access to or interest in internet surveys may differ in vaccination behaviors, and unvaccinated individuals may have been less likely to participate. However, over 16.9% (1478/8724) of respondents reported being unvaccinated, suggesting that arrange of vaccination statuses was adequately represented. Despite this limitation, the impact on the study's conclusions is likely minimal.

Conclusions

Seeking information related to COVID-19 on the internet is significantly associated with COVID-19 vaccination uptake but not with the timing of vaccination. However, an analysis differentiated by types of internet media reveals a varied impact across these platforms. Specifically, platforms like blogs and social media networks have been identified as negatively affecting vaccination uptake. Public health policy makers should strategically use online information media during an emerging infectious disease pandemic such as COVID-19, taking into account the unique characteristics of each media type.

Acknowledgments

This work was supported by the Kyoto University FY 2020 GAP Fund Program Type B (to YI), the Japan Society for the Promotion of Science (JSPS KAKENHI; grant number: 19H01075, 23H00448 to YI, 22K17549 to KT), and the Research Promotion Fund (FY2024) from the Japanese Medical Science Federation "Survey of the Impact of COVID-19 on Medicine, Medical Care, and Health (to YI). The funder is not involved in the study design, data collection, analysis, interpretation, writing of the manuscript, or decision to submit it for publication. We would like to thank Editage for English language editing.

During the preparation of this work, the authors used DeepL to improve readability and language. After using this tool or service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Data Availability

The data used in this study are not available for redistribution due to a lack of participant consent for sharing with third parties.

Authors' Contributions

KT was responsible for conceptualizing the study, conducting the formal analysis, and writing the original draft. MS contributed through the validation of the research findings. NT was involved in administering the project, the investigation, and data curation. YI played a critical role in acquiring funding, administering the project, and providing supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Criteria for Prioritizing COVID-19 Vaccination Based on Preexisting Medical Conditions.
[DOCX File , 25 KB - jmir_v27i1e59352_app1.docx]

Multimedia Appendix 2
Usage status of Information Resources.
[DOCX File , 500 KB - jmir_v27i1e59352_app2.docx]

References

1. World Health Organization. WHO director-general's opening remarks at the media briefing on COVID-19. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2024-12-12]
2. COVID-19 vaccine doses administered. Our World in Data. URL: https://ourworldindata.org/covid-vaccinations [accessed 2024-12-12]
3. Visualizing the data: information on COVID-19 infections. Ministry of Health, Labour and Welfare. URL: https://covid19.mhlw.go.jp/en/ [accessed 2024-12-12]
4. COVID-19 vaccines. Prime Minister's Office of Japan. URL: https://japan.kantei.go.jp/ongoingtopics/vaccine.html [accessed 2024-12-12]
5. The internet and the pandemic. Pew Research Center. 2020. URL: https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/ [accessed 2024-12-12]

XSL•FO
RenderX

6. Inazumi Y. Distribution Information and Systems as a Basis for Safety Assessment of COVID 19 Vaccine in Japan. Japanese Journal of Pharmacoepidemiology 2022;27(2):62-70 [FREE Full text] [doi: 10.3820/jjpe.27.62]
7. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
8. Merchant RM, South EC, Lurie N. Public health messaging in an era of social media. JAMA 2021;325(3):223-224. [doi: 10.1001/jama.2020.24514] [Medline: 33393964]
9. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
10. Kington RS, Arnesen S, Chou WYS, Curry SJ, Lazer D, Villarruel AM. Identifying credible sources of health information in social media: principles and attributes. NAM Perspect 2021;2021 [FREE Full text] [doi: 10.31478/202107a] [Medline: 34611600]
11. Lee JJ, Kang KA, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief With COVID-19 knowledge and preventive behaviors: cross-sectional online study. J Med Internet Res 2020;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
12. Van Dijk JAGM. Digital divide: Impact of access. In: The International Encyclopedia of Media Effects. Netherlands: University of Twente; Mar 8, 2017:1-11.
13. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
14. Hargittai E. Second-level digital divide: mapping differences in people's online skills. arXiv 2001 Sep 24 [FREE Full text] [doi: 10.48550/arXiv.cs/0109068]
15. Khanijahani A, Calhoun B, Kiel J. Internet use habits and influenza vaccine uptake among US adults: results from seven years (2012-2018) of the national health interview survey. Public Health 2021;195:76-82. [doi: 10.1016/j.puhe.2021.04.007] [Medline: 34062275]
16. O'Leary ST, Narwaney KJ, Wagner NM, Kraus CR, Omer SB, Glanz JM. Efficacy of a web-based intervention to increase uptake of maternal vaccines: an RCT. Am J Prev Med 2019;57(4):e125-e133. [doi: 10.1016/j.amepre.2019.05.018] [Medline: 31471001]
17. Ford AJ, Alwan NA. Use of social networking sites and women's decision to receive vaccinations during pregnancy: a cross-sectional study in the UK. Vaccine 2018;36(35):5294-5303. [doi: 10.1016/j.vaccine.2018.07.022] [Medline: 30055969]
18. Kesiena O, Atarere J, Singireddy S, Ademiluyi A, Famojuro O. The role of health information technology in pneumococcal vaccination uptake among adults with heart disease. Telemed J E Health 2022;28(5):699-705. [doi: 10.1089/tmj.2021.0333] [Medline: 34515541]
19. Dib F, Mayaud P, Longfier L, Chauvin P, Launay O. Effect of internet use for searching information on vaccination on the uptake of human papillomavirus vaccine in France: a path-analysis approach. Prev Med 2021;149:106615. [doi: 10.1016/j.ypmed.2021.106615] [Medline: 33989671]
20. Dib FH, Mayaud P, Longfier L, Launay O. Online vaccine-related information-seeking in mothers and HPV vaccine uptake in their daughters. European Journal of Public Health 2020;30(Supplement_5):165. [doi: 10.1093/eurpub/ckaa165.1254]
21. Zhang Y, Zhang L, Guan H, Hao R, Liu W. The impact of internet health information seeking on COVID-19 vaccination behavior in China. BMC Public Health 2024;24(1):89 [FREE Full text] [doi: 10.1186/s12889-024-17638-3] [Medline: 38178047]
22. Principe F, Weber G. Online health information seeking and Covid-19 vaccine hesitancy: Evidence from 50+ Europeans. Health Policy 2023;138:104942 [FREE Full text] [doi: 10.1016/j.healthpol.2023.104942] [Medline: 37984096]
23. Lee M, You M. Direct and indirect associations of media use with COVID-19 vaccine hesitancy in South Korea: cross-sectional web-based survey. J Med Internet Res 2022;24(1):e32329 [FREE Full text] [doi: 10.2196/32329] [Medline: 34870605]
24. Benis A, Khodos A, Ran S, Levner E, Ashkenazi S. Social media engagement and influenza vaccination during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2021;23(3):e25977 [FREE Full text] [doi: 10.2196/25977] [Medline: 33651709]
25. Hori D, Takahashi T, Kaneda Y, Ozaki A, Tabuchi T. The influence of information sources on intention changes to receive COVID-19 vaccination: a prospective cohort study in Japan. Environ Health Prev Med 2023;28:10 [FREE Full text] [doi: 10.1265/ehpm.22-00266] [Medline: 36725028]
26. Rogers EM. Innovativeness and adopter categories. In: Diffusion of Innovations, 5th Edition. New York, NY: Free Press; Aug 16, 2003:267-299.
27. Suka M, Odajima T, Kasai M, Igarashi A, Ishikawa H, Kusama M, et al. The 14-item health literacy scale for Japanese adults (HLS-14). Environ Health Prev Med 2013;18(5):407-415 [FREE Full text] [doi: 10.1007/s12199-013-0340-z] [Medline: 23689952]
28. Goel RK, Nelson MA. COVID-19 internet vaccination information and vaccine administration: evidence from the United States. J Econ Finan 2021;45(4):716-734. [doi: 10.1007/s12197-021-09551-x]

XSL•FO
RenderX

29. Mangono T, Smittenaar P, Caplan Y, Huang VS, Sutermaster S, Kemp H, et al. Information-seeking patterns during the COVID-19 pandemic across the United States: longitudinal analysis of google trends data. J Med Internet Res 2021;23(5):e22933 [FREE Full text] [doi: 10.2196/22933] [Medline: 33878015]
30. Yang JZ, Liu Z, Wong JCS. Information seeking and information sharing during the COVID-19 pandemic. Communication Quarterly 2021;70(1):1-21. [doi: 10.1080/01463373.2021.1995772]
31. Superio DL, Anderson KL, Oducado RMF, Luceño MT, Palcullo VEV, Bendalian MVT. The information-seeking behavior and levels of knowledge, precaution, and fear of college students in Iloilo, Philippines amidst the COVID-19 pandemic. Int J Disaster Risk Reduct 2021;62:102414 [FREE Full text] [doi: 10.1016/j.ijdrr.2021.102414] [Medline: 34189029]
32. Cascini F, Pantovic A, Al-Ajlouni YA, Failla G, Puleo V, Melnyk A, et al. Social media and attitudes towards a COVID-19 vaccination: a systematic review of the literature. EClinicalMedicine 2022;48:101454 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101454] [Medline: 35611343]
33. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
34. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
35. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021;23(6):e29802 [FREE Full text] [doi: 10.2196/29802] [Medline: 34043526]
36. Roselina E, Asmiyanto T, Andriany M. Health information-seeking behavior on the COVID-19 pandemic: social media usage by Gen Z in Jakarta, Indonesia. ResearchGate 2021 Feb:1-9 [FREE Full text]
37. eHealth. YouTube. URL: https://www.youtube.com/@eHealthInsurance [accessed 2024-12-12]
38. Current population estimates as of October 1, 2021. Statistics Bureau of Japan. URL: https://www.stat.go.jp/english/data/jinsui/2021np/index.html [accessed 2024-12-12]
39. Kindratt T, Callender L, Cobbaert M, Wondrack J, Bandiera F, Salvo D. Health information technology use and influenza vaccine uptake among US adults. Int J Med Inform 2019;129:37-42. [doi: 10.1016/j.ijmedinf.2019.05.025] [Medline: 31445279]

Abbreviations

aOR: adjusted odds ratio
COCOA: COVID-19 Contact-confirming Application
HLS-14: 14-item health literacy scale for Japanese adults
SNS: social networking services

Edited by A Coristine; submitted 16.04.24; peer-reviewed by Y Sun, DL Superio; comments to author 29.05.24; revised version received 11.07.24; accepted 23.10.24; published 14.01.25.

Please cite as:
Taira K, Shiomi M, Nakabe T, Imanaka Y
The Association Between COVID-19 Vaccination Uptake and Information-Seeking Behaviors Using the Internet: Nationwide Cross-Sectional Study
J Med Internet Res 2025;27:e59352
URL: https://www.jmir.org/2025/1/e59352
doi:10.2196/59352
PMID:39808493

©Kazuya Taira, Misa Shiomi, Takayo Nakabe, Yuichi Imanaka. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cross-Cultural Sense-Making of Global Health Crises: A Text Mining Study of Public Opinions on Social Media Related to the COVID-19 Pandemic in Developed and Developing Economies

Adham Kahlawi[1,2], BSc, PhD; Firas Masri[3], BA, MSc, PhD; Wasim Ahmed[4], BA, MSc, PhD; Josep Vidal-Alaball[5,6,7], MD, PhD

[1]Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy

[2]Lloyds Banking Group (United Kingdom), Edinburgh, United Kingdom

[3]Newcastle Business School, Northumbria University, Newcastle Upon Tyne, United Kingdom

[4]Hull Univeristy Business School, Hull University, Kingston Upon Hull, United Kingdom

[5]Unitat de Recerca i Innovació, Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[6]Intelligence for Primary Care Research Group, Unitat de Suport a la Recerca de la Catalunya Centra, Fundació Institut Universitari per a la recerca, Sant Fruitós de Bages, Spain

[7]Department of Medicine, Faculty of Medicine, University of Vic, Vic, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, PhD
Unitat de Recerca i Innovació
Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central
Institut Català de la Salut
Carrer Pica d'Estats, 36
Sant Fruitós de Bages, 08272
Spain
Phone: 34 936930040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: The COVID-19 pandemic reshaped social dynamics, fostering reliance on social media for information, connection, and collective sense-making. Understanding how citizens navigate a global health crisis in varying cultural and economic contexts is crucial for effective crisis communication.

Objective: This study examines the evolution of citizen collective sense-making during the COVID-19 pandemic by analyzing social media discourse across Italy, the United Kingdom, and Egypt, representing diverse economic and cultural contexts.

Methods: A total of 755,215 social media posts from X (formerly Twitter) were collected across 3 time periods: the virus' emergence (February 15 to March 31, 2020), strict lockdown (April 1 to May 30, 2020), and the vaccine rollout (December 1, 2020 to January 15, 2021). In total, 284,512 posts from Italy, 261,978 posts from the United Kingdom, and 209,725 posts from Egypt were analyzed using the latent Dirichlet allocation algorithm to identify key thematic topics and track shifts in discourse across time and regions.

Results: The analysis revealed significant regional and temporal differences in collective sense-making during the pandemic. In Italy and the United Kingdom, public discourse prominently addressed pragmatic health care measures and government interventions, reflecting higher institutional trust. By contrast, discussions in Egypt were more focused on religious and political themes, highlighting skepticism toward governmental capacity and reliance on alternative frameworks for understanding the crisis. Over time, all 3 countries displayed a shift in discourse toward vaccine-related topics during the later phase of the pandemic, highlighting its global significance. Misinformation emerged as a recurrent theme across regions, demonstrating the need for proactive measures to ensure accurate information dissemination. These findings emphasize the role of cultural, economic, and institutional factors in shaping public responses during health crises.

Conclusions: Crisis communication is influenced by cultural, economic, and institutional contexts, as evidenced by regional variations in citizen engagement. Transparent and culturally adaptive communication strategies are essential to combat

XSL•FO
RenderX

misinformation and build public trust. This study highlights the importance of tailoring crisis responses to local contexts to improve compliance and collective resilience.

(J Med Internet Res 2025;27:e58656) doi:10.2196/58656

KEYWORDS

COVID-19; SARS-CoV-2; pandemic; citizen opinion; text mining; LDA; health crisis; developing economies; Italy; Egypt; UK; dataset; content analysis; social media; twitter; tweet; sentiment; attitude; perception; perspective; machine learning; latent Dirichlet allocation; vaccine; vaccination; public health; infectious

Introduction

Background

According to the World Health Organization, COVID-19 was one of the most significant global public health challenges. Italy, Egypt, and the United Kingdom were among the countries considered to be the worst affected [1,2]. Many studies have noted how the internet and social media are essential in communicating risks and emergencies before, during, and after a crisis [3,4]. Human behavior related to using and processing information has been enormously affected by the orientations of the digital world in the 21st century. This uniquely contrasts previous global deadly disease outbreaks such as the Black Death and the Spanish Influenza. Furthermore, a recent study by Christianson & Barton noted how COVID-19 had created a unique scenario that provides an opportunity to study "sense-making" [4].

Digital platforms allow users to converse with one another for collective sense-making, which involves seeking, generating, processing, and using information [5]. A prevalent platform for seeking and sharing health information alongside breaking news stories is X (formerly known as Twitter) [6].

Specifically, sense-making and social media usage during the COVID-19 pandemic have received much attention from scholars in various fields, including sociology, psychology, communication, and information systems [4,7,8]. Studies in this area have investigated how individuals use social media to make sense of the rapidly changing and complex situation posed by the COVID-19 pandemic. Research has shown that social media can be essential in providing access to information, shaping public opinion, and influencing behavior during a crisis.

One strand of research has investigated how individuals use social media to gather information and make sense of the pandemic. Studies have shown that social media, such as X, can serve as an essential source of information for individuals, especially during a crisis when traditional sources of information may be limited or unreliable. Social media also enables individuals to seek out information from various sources, which can help them form a more nuanced understanding of the situation [9].

Another strand of research has investigated the role of social media in shaping public opinion and behavior during the pandemic. Studies have shown that social media can be an essential factor in shaping public opinion and influencing individuals' decisions about things like their behavior and whether to seek medical treatment or follow recommended guidelines [10,11]. Despite this body of research, there are still some crucial gaps in our understanding of the relationship between sense-making and social media usage during the COVID-19 pandemic. More research is needed to investigate the role of social media in shaping public opinion and behavior, particularly in diverse populations and geographic regions. Additionally, there is a need for more cross-country digital research on how individuals use social media to make sense of the pandemic, including how they navigate and interpret the vast amounts of information available on social media platforms. While studies have explored the impact of social media on public opinion and behavior, more research is needed to understand how X influences public opinion and behavior during uncertainty and crises such as COVID-19 [12].

This study seeks to analyze data on X related to the 3 phases of COVID-19 and the response of citizens across different countries on X. Our study focused on examining data from Italy, the United Kingdom, and Egypt. Italy has a national health system and is a member of the European Union. The United Kingdom is a country that also has a national health system and has recently left the European Union. Egypt does not have a national health system nor is it a member of any allied institutions. For these reasons, these 3 countries provide an interesting comparison point. Our study would uncover the differences in public perceptions and cultural attitudes toward crisis information on social media. The discourse from certain countries may also vary depending on culture, freedom of speech restrictions, and events.

In order to contextualize the developmental differences between Egypt, Italy, and the United Kingdom, we made reference to the Human Capital Index [13]. This is a quantitative measure of development based on health and education outcomes. According to this index, Italy scores 0.73, the United Kingdom 0.78, and Egypt 0.49. The index highlights how there is a substantial developmental gap between Italy and Egypt (0.73 vs 0.49) compared to the gap between Italy and the United Kingdom (0.73 vs 0.78). The significant difference between Egypt's Human Capital Index and that of Italy and the United Kingdom substantiates our classification of Egypt as a developing country relative to these 2 more developed nations. This cross-country variation allows us to explore how diverse sociocultural contexts and government responses influence citizen engagement with COVID-19 information on social media. Italy, the United Kingdom, and Egypt had varied timelines and policies. Italy responded with strict early lockdowns, coupled with economic support programs intended to mitigate financial strain. The United Kingdom initially adopted a more gradual approach, balancing public health measures with economic interests before implementing more

XSL•FO
RenderX

extensive restrictions as the pandemic progressed. Egypt, on the other hand, maintained a comparatively limited governmental response, with religious institutions playing a crucial role in disseminating health guidance and encouraging precautionary behaviors.

Research Aims

This study examines the variation of public attention and interaction with COVID-19 by analyzing data from X. It tries to estimate how COVID-19 has captured the attention of citizens across Italy, Egypt, and the United Kingdom. The theoretical lens of the study is sense-making. More specifically, the research questions (RQs) of this study are as follows (Textbox 1).

Textbox 1. Research aims of this study.

- RQ1: To develop an understanding of public views shared on X toward COVID-19 across Egypt, Italy, and the United Kingdom.
- RQ2: To compare the response of citizens toward COVID-19 on X across Egypt, Italy, and the United Kingdom.
- RQ3: To examine the concept of sense-making concerning community responses to COVID-19 in Egypt, Italy, and the United Kingdom.

This study contributes to the existing body of knowledge by providing a cross-cultural analysis of public sense-making processes and reactions. This remains crucial for understanding how communities adapt to health crises. While the data were collected during the early stages of the pandemic, the lessons learned are still applicable to current and future public health challenges, particularly in designing culturally sensitive communication strategies and mitigating misinformation. Moreover, to the best of our knowledge, this is the first empirical investigation that systematically examines public discourse on X during the COVID-19 pandemic across Egypt, Italy, and the United Kingdom, offering a comparative perspective.

Literature Review: COVID-19 and Social Media Research

A wide range of research on the COVID-19 pandemic, which draws upon X datasets, has been conducted. A recent study proposed several algorithms to estimate work engagement. Other studies have ranged from the identification of misinformation [14], examining public views and opinions related to various issues around the pandemic [15], and research related to epidemiology [16]. In one of the earliest studies examining misinformation on X, the author examined how a conspiracy theory linking COVID-19 to 5G spread on X [16]. Social network analysis was used alongside a content analysis to identify key users, content, and websites related to the conspiracy. Other research has also examined the relationship between fact-checking and misinformation [17]. The study found that fact-checks tend to appear after the spread of misinformation but that their ability to reduce the spread of misinformation is limited. Furthermore, they found that topics such as conspiracies are more resistant to fact-checkers.

Scholars have also analyzed X data to better understand perceptions of face masks [18]. In one study, the authors analyzed the content of the posts that talked about the vaccine and categorized them into these categories (provaccine, antivaccine, and neutral). Consequently, they found that the antivaccine posts use the sarcasm method, and some of them come from bots, while the pro-vaccine posts depend on documented information [19]. It is worth noting that over time, trust in the information spread on the internet is affected by the recipient's political affiliation [20].

Several studies have conducted research looking at COVID-19 in relation to other platforms. Although not a study on social media data itself, Wang et al [21], through a web-based survey posted on Facebook, found that Taiwanese citizens were mostly using the internet for their health information. This is an important finding as it highlights that web-based sources are likely to be the most viewed sources of information for a subset of the population. In another study, Sesagiri Raamkumar et al [22] examined the outreach efforts of public health authorities on Facebook in Singapore, the United States, and England. They found that the most common theme among all posts was those related to prevention and safety measures. It is also important to mention that outside of COVID-19 research, X has also been used across a range of disciplines in innovative ways. For example, X has been used to identify the political affinity of web-based entities through X followers and for the automatic identification of eyewitness messages on X at the time of natural disasters.

Although previous research has examined content from social media related to COVID-19, there is a lack of empirical work that has compared user views across different geographical locations using time as a moderating variable. Our proposed study hopes to address this research gap and contribute to the wider public health literature by developing an understanding of social media user views across 3 countries.

Theoretical Background

We define the term sense-making as a process through which individuals aim to develop an understanding of events that may be confusing, uncertain, and violent by interacting with others [20-25]. In recent years, scholars have studied the concept of sense-making about interactions on social media platforms. For instance, a study by Zolyomi et al [5] examined the #ActuallyAusistic hashtag on X and found that autistic individuals used X to engage with sense-making to understand themselves better. In this context, sense-making was applied to construct identities in relation to individual social groups, employment, education, and the use of technology. Another study [22] focused on X collective sense-making that was examined during the Berlin terrorist attack. The study found that X users made sense of the events that were occurring on X. The study highlighted how collective behavior occurred during the aftermath of a terrorist attack. There are many definitions and discussions of social media within academic

XSL•FO
RenderX

literature [26]. A popular definition notes that they are “a group of internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content” [27].

Previous research has also examined sense-making on social media during other extreme events. For instance, Stieglitz et al [23] analyzed X data related to 3 events: the Sydney Lindt Café siege (2014), the Germanwings plane crash (2015), and the Brussels terror attacks (2016). They found that there were differences in how users communicated with one another across the 3 events. The authors note that citizens may use social media during extreme events to speak with one another and make sense of an event.

The researchers [24] noted that the information a person receives significantly impacts his behavior. Furthermore, they found a strong relationship between misinformation spread on social media and following preventive measures with the COVID-19 epidemic. Therefore, the researchers recommended publishing the correct news and confronting the misinformation to improve the citizens' behavior.

Methods

Study Design

The study was designed and executed in 6 phases, and a summary of these can be found in Figure 1, which provides a visual summary of our research design. Our study was pragmatic, inductive, and exploratory in nature, drawing upon the theoretical lenses of sense-making.

Figure 1. Overview of the 6 phases of our methodology.

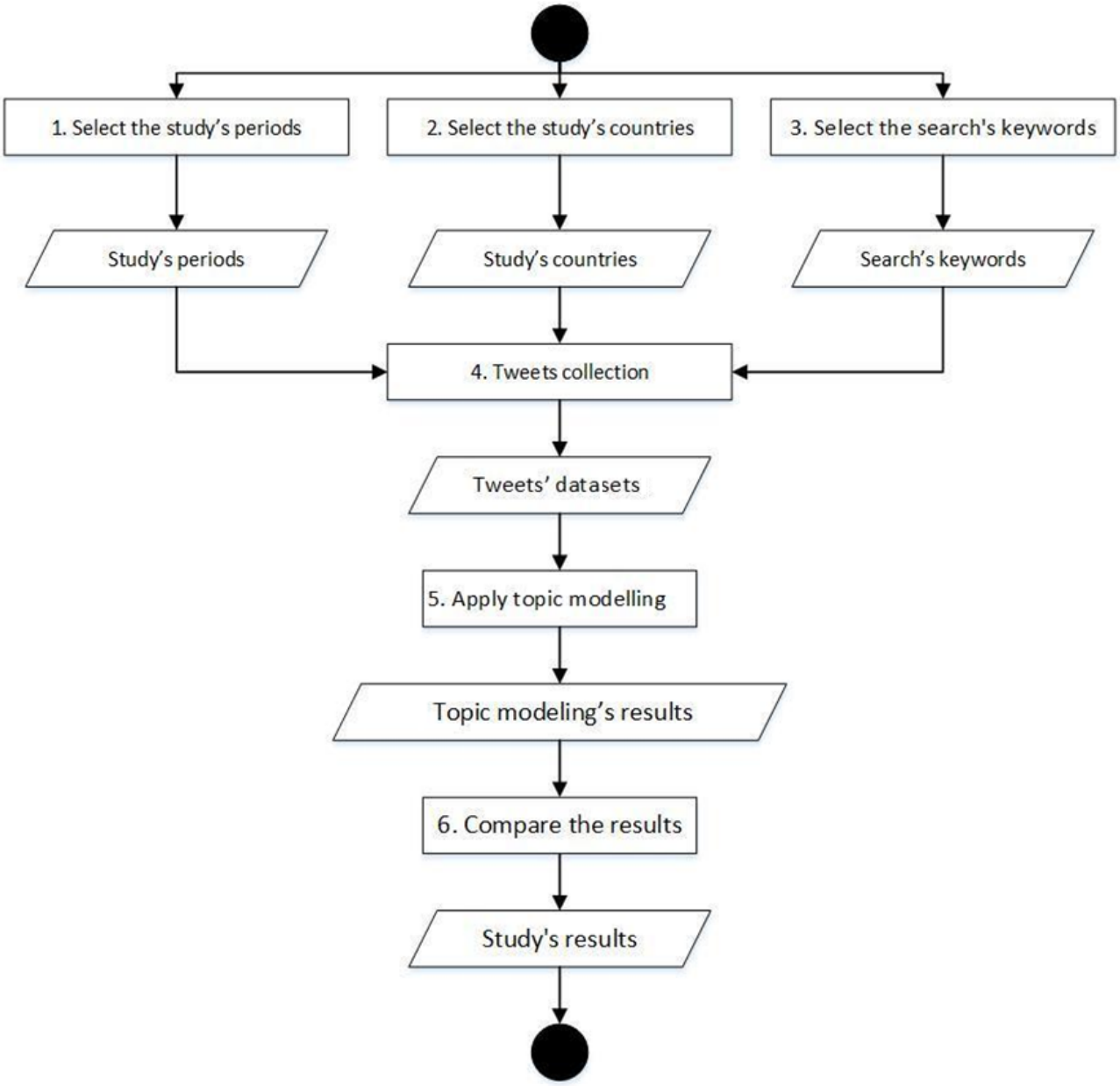

XSL•FO
RenderX

Time Period Selection

COVID-19 appeared toward the beginning of 2020, and up to early 2021, there were 3 key periods for the pandemic, which can be characterized as follows: the emergence phase; during the initial lockdown; and when cases began to reappear during the second winter lockdown.

The difference between the 3 periods has been categorized based on the COVID-19 spreading and the political and governmental procedures and interventions that are undertaken. Therefore, the first period indicates the period of the virus' emergence and spreading, which took almost 2 months to be recognized as a pandemic by the World Health Organization. The second period includes the period of strict decisions that governments have made, which, in some countries, have reached total lockdown. The last period consists of the period of containing the virus and easing the severity of government restrictions. Since countries have responded to the pandemic situation differently, this study has tried to select countries that cover the spectrum of different situations.

All in all, we used the same time frames for Italy, the United Kingdom, and Egypt to analyze the influence of Italy's early pandemic responses on global strategies. This uniform approach allows for comparative insights into how different regions adapted measures pioneered by Italy, facilitating an understanding of both reactive and locally influenced responses. In addition, the gap between the studied periods allows us to examine the adjustments and preparations made after initial lockdowns and before the resurgence of the virus. This helps in understanding the balance countries maintained between economic reopening and health safety, which is crucial for evaluating the effectiveness of pandemic management over time.

The overview of the time intervals across these phases is shown in Table 1, which shows our 3 time periods and their associated time intervals for which data was collected.

Table 1. Overview of the 3 time periods examined.

Periods	Time intervals
The virus' emergence	From February 15 to March 31, 2020
Lockdown	From April 1 to May 30, 2020
Virus spreading	From December 1, 2020, to January 15, 2021

Keyword Strategy

A group of keywords most used by X users when conversing about COVID-19 was chosen using the website Getdaytrends. This website shows the most used hashtags during a specific time, including the number of posts attached to such hashtags for a particular country. This process allows for getting the most posts about COVID-19. Table 2 provides an overview of the keywords used to retrieve data for the respective countries examined.

Table 2. Overview of the keyword strategy to retrieve data.

Hashtag	Country
COVID_19	Italy, the United Kingdom, and Egypt
COVID-19	Italy, the United Kingdom, and Egypt
COVID19	Italy, the United Kingdom, and Egypt
COVID2019	Egypt
coronavirus	Italy, the United Kingdom, and Egypt
19 (en[a]: COVID19)	Egypt
_19 (en: COVID_19)	Egypt
(en: CORONA)	Egypt
(en: CORONA)	Egypt
_ (en: CORONA_VIRUS)	Egypt
_ (en: CORONA_EGYPT)	Egypt

[a]en refers to the English language.

Collecting X Data

Posts were collected in Python via the tweepy library. A requirement when collecting posts was to be written in the country's language and posted by its inhabitants. In order to fulfill these 2 conditions, some parameters were determined in the collection's algorithm. The algorithm uses the "lang" parameter to determine the post's language and the "geocode" parameter to determine the users' position when they have posted. Since the "geocode" parameter requires latitude and longitude, we have collected this information for all cities in our 3 countries via Google Maps. Consequently, the collection steps were executed as a loop for each keyword the algorithm

XSL•FO
RenderX

executes for each city. The posts collected contain at least 1 hashtag of the search's keywords list.

The posts from Italy were filtered based on the language, assuming that the Italian language is unlikely to be used significantly outside of Italy (ie, to represent another county or region). However, filtering the posts based on location and language was used to collect the posts from Egypt and the United Kingdom, as English and Arabic could represent other regions.

Posts that were retrieved contain at least 1 keyword from Table 2. Table 3 shows the number of posts collected for each country across the 3 time periods examined. It is important to note that the number presented in Table 3 represents the final collected posts after applying 3 separate types of filters on the collected posts, which were as follows (Textbox 2).

Table 3. Overview of data retrieved across phases and countries from X.

Country	Final collected posts, n		
	The virus' emergence (phase 1)	Lockdown (phase 2)	Virus spreading (phase 3)
Italy	98,475	100,746	85,291
United Kingdom	84,675	97,842	78,461
Egypt	75,712	91,846	42,167

Textbox 2. Filters applied to collected data.

- All repeated posts were deleted using the post ID.
- All posts that had the same text were deleted.
- All posts with less than 20 characters were deleted.

Application of Topic Modeling

Topic modeling algorithms are frequently used to analyze posts. This method was selected in this study due to the volume of posts being analyzed. One of the most important types of these algorithms is latent Dirichlet allocation (LDA) [25]. The LDA model is a generative model used in natural language study, which allows for extracting topics from a set of source documents. Each document is considered a group of words that form one or more latent topics when combined [26]. A particular distribution of terms characterizes each topic. LDA's generative process is based on analyzing the text's data (text mining). Word combinations are considered as random variables.

The LDA algorithm was used to identify frequently used conversation topics in the corpus of data collected within the COVID-19 context. The selected number of topics (T) were extracted from a collective dataset (C). Probability (Pi) of inclusion of individual posts (Ci) from the dataset can be represented as:

$$\text{Probability } (Ci \in ti) = Pi$$

$$\text{Collective posts } C = \{C1, C2, C3\ldots,Cn\}, \text{ where topic distribution } T = \{t1, t2, t3,\ldots,tn\}$$

Our topic models generated a topic matrix and the likelihood of individual topic strength. The bag of words constructed within the matrix can be designated as n-grams. In our case, a bigram model was selected due to its popularity in strengthening topic cohesion and topic relevance. We experimented with the optimum topic strength value up to the recommended 30. Several rounds of experimentation showed that the thematic importance of collective sense-making could be best captured between 2 and 7 sets of topics using bigram values.

The application of the topic modeling is divided into 3 steps. The first step is preparing the dataset to train the model, where each post is treated as a document. The second step is selecting the number of topics that are identified. The analysis will be conducted more than once, as each time, a different number will be determined for the number of topics that will range between 2 and 10. After each iteration, the coherence value will be extracted, indicating the model's efficiency [26]. Afterwards, the number of topics used in the model with the highest coherence value will be selected, and if all topics represent more than 5% of the dataset, we will decrease the topic numbers to achieve this condition. In the third step, the LDA model will be repeated, confirming the number of topics that are selected in the previous steps. Then, the LDA model will be used to extract a list of words representing each topic; consequently, the word lists will be used to describe each topic's connotation. Moreover, the LDA model classifies posts by their represented topic; consequently, the percentage each topic represents within the database will be calculated.

Identification of Topics

After completing the previous step, the study will have developed results for 9 datasets, 3 for each country. The results for each country are discussed separately and will be compared later. We have obtained 9 independent topic models by applying the LDA model to each country's datasets. Each topic model returns the number of topics representing the dataset and the top words of each one. In the next section, we give a label and a description for each topic.

Finally, the study used LDA and bigrams to analyze large-scale social media posts, aiming to extract meaningful topics. LDA was chosen for its effectiveness in handling vast textual data, while bigrams enhanced the contextual relationships between words. Despite challenges like broad topics and tweet brevity,

XSL•FO
RenderX

LDA proved suitable for capturing public discourse. The paper defines "sense-making" as how individuals interpret public health information, especially during crises. The topics derived through LDA reflect public interpretations of COVID-19, and manual labeling by experts ensured accurate insights into public responses.

Ethical Considerations

This research was covered by ethical approval gained from Newcastle University (ref: 1067/2020). Moreover, our study is conducted on aggregate data, and we do not draw attention to any individual users or publish any identifiable information.

Use of Tools to Support Proofreading

We drew upon large language models such as ChatGPT (version 4o) for basic proofreading as well as tools such as Grammarly to identify any errors in writing and sentence construction. However, no tools were used to generate text of their own.

Results

Table 4 provides an overview of the results of the study.

XSL•FO
RenderX

Table 4. Overview of the key themes, subthemes, countries, and time periods.

Overall theme and topics or subthemes	Country	Period
The general situation		
The general situation in Italy	Italy	The virus' emergence (phase 1)
Region of Lombardy	Italy	Lockdown (phase 3)
General situation	Italy	Virus spreading (phase 3)
General situation	Egypt	Lockdown (phase 3)
General situation	Egypt	Virus spreading (phase 3)
The situation in Europe	United Kingdom	Virus' emergence (phase 1)
General situation	United Kingdom	Lockdown (phase 3)
General situation	United Kingdom	Virus spreading (phase 3)
Protection		
Protection	Italy	The virus' emergence (phase 1)
Quarantine process	Italy	Virus spreading (phase 3)
Religious side	Egypt	Virus spreading (phase 3)
Prepare for a pandemic	United Kingdom	Virus' emergence (phase 1)
Lockdown	United Kingdom	Virus' emergence (phase 1)
Media		
Media	Italy	The virus' emergence (phase 1)
Virus		
Virus	Italy	Lockdown
Virus mutated	Italy	Virus spreading (phase 3)
Economic		
Economic status	Italy	Lockdown (phase 3)
The European role	Italy	The virus' emergence (phase 1)
Economic impact	United Kingdom	Lockdown (phase 2)
Vaccine		
The vaccine	Italy	Lockdown (phase 2)
The vaccine	Italy	Virus spreading (phase 3)
The vaccine	Egypt	Virus spreading (phase 3)
Religion		
Places of worship	Egypt	Virus' emergence (phase 1)
Evoking historical events	Egypt	Virus' emergence (phase 1)
Lack of confidence in the government		
Lack of confidence in the government	Egypt	Virus' emergence (phase 1)
Demanding the dismissal of the Minister of Health	Egypt	Lockdown (phase 2)
Medical staff		
Medical staff	Egypt	Lockdown (phase 2)

Content on X in Italy (Phase 1)

This information in Table 4 can be used to give the following descriptions for content in Italy during phase 1:

- Topic 1: General situation in Italy:
  The virus began to spread increasingly in northern Italy, especially the Lombardy region, which caused an increase in the number of patients in need of intensive care and increased the number of deaths.
- Topic 2: European role:
  With the start of the lockdown, citizens sensed the extent of the economic damage that the virus might cause. They also sensed the Italian government's inability to deal with these damages on its own. For this reason, discussions noted

XSL•FO
RenderX

how the European Union must intervene as soon as possible and help citizens and companies alike.
- Topic 3: Protection:
The virus is unknown by specialists, who cannot predict the mechanism of its spread. Therefore, all preventive measures must be taken, especially wearing masks when leaving the house. Consequently, the government has been asked to secure many masks to meet citizens' needs and distribute them free, as did a region of Tuscany.
- Topic 4: Media:
The media, in their various forms, played a significant role in reporting the stories and sacrifices of medical personnel in the face of this virus; besides, it focuses on the political disagreements between the government and the opposition.

Content on X in Italy (Phase 2)

This information in Table 4 can be used to give the following descriptions for content in Italy for phase 2:

- Topic 1: Region of Lombardy:
Despite quarantine, the number of injuries in the region is still on the rise. This rise has been accompanied by an increase in the number of deaths and increased pressure on the health sector, which has become weak and unable to bear more patients.
- Topic 2: Virus:
Some voices appeared on social media that deny the existence of the virus or claim to reduce its risk. For instance, some noted that the virus can be eliminated by taking certain vitamins and that the lockdown measures were created to destroy countries' economies. The presence of such voices prompted many to respond to them and prove that such claims were incorrect.
- Topic 3: Economic status:
With the increase in the quarantine period, fears of citizens' deteriorating economic situation and the fear of losing their jobs began to appear. A number of citizens noted that they did not have enough money to buy basic necessities.

Content on X in Italy (Phase 3)

This information in Table 4 can be used to give the following descriptions for content in Italy for phase 3:

- Topic 1: Virus mutated:
A new virus mutation in England was discussed, and users hypothesized that there would be an increased transmission. Meanwhile, the arrival of the new mutation to Europe increased fears, and the demand to close borders started.
- Topic 2: Quarantine process:
The government enacted a new law to regulate the quarantine process, giving color to each region according to the data on the epidemic. Each color symbolizes different restrictions on the inhabitants of the areas. This law was difficult for citizens to understand, especially the differences in color restrictions.
- Topic 3: The vaccine:
The government developed a nationwide plan to vaccinate citizens, which began with the arrival of the Pfizer vaccine. X users noted that the Lombardy region should be provided with more doses to cover faster. Concerns resurfaced after Pfizer announced a delay in reaching the agreed quantities.
- Topic 4: General situation:
The daily reports of the number of infections were being actively monitored by citizens. X users also requested that schools remain open regardless of the number of casualties.

Sense-making during the COVID-19 crisis started with wider discussions without a clear focus on selective topics. The trend eventually became more fixated on the cause and evolution of the crisis (SARS-CoV-2) and prevention measures (quarantine and vaccine).

Content on X in Egypt (Phase 1)

This information in Table 4 can be used to give the following descriptions for content in Egypt for phase 1:

- Topic 1: Places of worship:
The first measure to prevent the virus was to close the places of worship, which sparked a great debate among citizens, especially among the Muslim community members, due to the approaching month of Ramadan. Some of the users demanded for mosques not to close, and some insisted on closing the mosques due to the overcrowding during the month of Ramadan.
- Topic 2: Lack of confidence in the government:
The government was not transparent in handling the epidemic, which caused citizens to distrust the reality of the reports, especially the numbers of infections. Besides, many citizens accused the government of spreading false news about the virus.
- Topic 3: Evoking historical events:
Many invitations have appeared for the obligation to adopt the quarantine even if the government did not impose it. Indeed, these invitations were based on epidemics that occurred during the Prophet Muhammad's time and how he ordered people not to go to places where the epidemic was spreading and not leave the place of the epidemic for those in it now of its spread.

Content on X in Egypt (Phase 2)

The information in Table 4 can be used to give the following descriptions for content in Egypt for phase 2:

- Topic 1: Medical staff:
With the increase in the number of cases, hospitals' and medical staff's pressure increased, and they began to work as robots day and night. Meanwhile, the media highlighted the medical staff's sacrifices and the deaths that occurred among them.
- Topic 2: General situation:
Hospitals are no longer able to receive more injured patients. They experience difficulty from shortages in equipment, which has caused terror among citizens who demanded the government to impose a complete lockdown.
- Topic 3: Demanding the dismissal of the Minister of Health:
From many Egyptians' point of view on X, the minister was not up to the responsibility, and because they did not hire more doctors and nurses. And some Egyptians went to accuse the minister of killing patients.

XSL•FO
RenderX

Content on X in Egypt (Phase 3)

This information in Table 4 can be used to give the following descriptions for content in Egypt for phase 3:

- Topic 1: The vaccine:
  Jordan obtained the Chinese vaccine and started the vaccination campaign. Accordingly, there was a campaign beckoning for the government to follow Jordan's steps and to be able to limit the spread of the epidemic.
- Topic 2: General situation:
  The number of injuries increases daily, and the number of deaths increases. Also, problems increase in hospitals. One of the most serious of these problems was the lack of oxygen in Al Hasina Hospital, which caused the death of a group of patients in the intensive care unit.

As opposed to a developed nation like Italy, in Egypt, collective sense-making was narrowly focused during phase 1. In the absence of clear scientific logic, during the initial days of the crisis, people relied on religious and historical analogies. As time progressed, people started to understand the importance of preventative measures like vaccines, yet overall discussions were identified to become broad and dispersed as time progressed.

Content on X in the United Kingdom (Phase 1)

This information in Table 4 can be used to give the following descriptions for content in the United Kingdom during phase 1:

- Topic 1: Prepare for a pandemic:
  The people are sure of the epidemic's arrival; therefore, they declared the importance of preparing for it and strengthening the health sector.
- Topic 2: The situation in Europe:
  Citizens followed the news of the epidemic in Europe with great apprehension. The virus moved quickly between the Union countries; consequently, they were sure that it would reach them.
- Topic 3: Lockdown:
  The government suggested not to close, which caused great fear among the citizens, who immediately demanded the government to apply the lockdown.

Content on X in the United Kingdom (Phase 2)

The information in Table 4 can be used to give the following descriptions for content in the United Kingdom during phase 2:

- Topic 1: General situation:
  The number of injuries increased daily, and the number of deaths also increased. X users also note that problems have increased in hospitals.
- Topic 2: Economic impact:
  Citizens had a real state of terror of the economic effects of the epidemic. Accordingly, they turned to the government to ask it to adopt economic policies that help companies and individuals.

Content on X in the United Kingdom (Phase 3)

This information in Table 4 can be used to give the following descriptions for content in the United Kingdom during phase 3:

- Topic 1: General situation:
  Daily reports are still the focus of citizens' attention, especially the news of virus mutation. In addition, the government's demands for economic aid continued.
- Topic 2: The vaccine:
  The vaccine's arrival marked a significant turning point for citizens, giving them a dose of hope that the virus could spread, which increased significantly.

Collective sense-making in the United Kingdom shows unique characteristics with the transition of time. During the initial days, public opinion was varied and dispersed in terms of preparation for the pandemic. Due to early interventions from the government and health authorities, many people were questioning the preparedness and viability of the public health care system and whether it could withstand the unexpected crisis. With the passage of time, discussions and opinions became more dispersed and polarized in terms of the economy (during phase 2). Finally, during phase 3, collective sense-making topics were largely focused on the vaccine and its scientific viability, along with other general issues in terms of economic and employability support.

Comparing Responses

This section will compare the different phases and the resulting themes that emerged. There were differences and certain similarities in how X users made sense of the unfolding events. Table 4 provided an illustration of all the themes that emerged for the 3 cases within the periods of the study, and further insights can be extracted from it, as provided in Textbox 3.

XSL•FO
RenderX

Textbox 3. Insights from comparing themes.

- It shows how citizens were following posting or reposting the pandemic news, including the number of new cases, deaths, fatality issues, etc.
- It shows how posts in Italy and the United Kingdom covered mainly prevention measures imposed by the governments in their countries, the extent of their validity, and the need to adhere to them. However, posts from Egypt mainly focused on religious arguments to encourage people to take protection measures to cover the void left by the government. This trend has been adopted when the people feel that their government does not implement adequate preventive measures. On the other hand, many posts demanded that the houses of worship not be closed so that the prevention measures could be easily applied according to their description.
- It can be argued that the media has played a crucial role in following up on spreading the virus and discussing preventive procedures and government involvement, especially in the United Kingdom.
- The posts in Italy and the United Kingdom called on their governments to provide economic aid or increase this aid to match the extent of the damage they experienced, while the Egyptians did not demand that might be justified by people's belief in the government's inability to follow the same European economic policy.
- The virus mutation aligned with a great concern for Italian citizens, especially the fear of returning to complete closure again.
- Egyptian posts showed their lack of confidence in the government in its ability to manage the crisis. Many posts claimed that the Minister of Health was unable to protect the doctors working in the first line of defense and questioned how it would be able to protect the people.
- Overall, the posts in the 3 countries found that the only way out of the pandemic would be to use vaccines that had been developed.

Discussion

Principal Findings

In response to RQ1, the results of our analysis have developed an understanding of the key views shared on X toward COVID-19 across Egypt, Italy, and the United Kingdom. Table 4 provides insight into the differing views that X users shared during this time. In response to RQ2, we found that there were differences and some similarities in how X users responded to X across Egypt, Italy, and the United Kingdom. We provided a comparison of the key discussions that took place across different countries and phases. In response to RQ3, it was found that X users used X to seek, generate, and process information collectively. Henceforth, collective sense-making on X was taking place, and it was found that there were differences in how people gave meaning to their experiences. During phase 1, X users conversed about topics and events that were occurring around them, which also continued within phase 2. During phase 3, as global news of the vaccine began to emerge, the collective discussions across all 3 regions appeared to be focused on news of the vaccine. This was significant news, affected all citizens in all countries, and was a potential way to navigate out of the pandemic.

This analysis of social media posts from the United Kingdom, Italy, and Egypt highlights distinct variations in how citizens across different cultural contexts responded to the COVID-19 pandemic. Our study found that this was driven partly by differing levels of governmental trust and state intervention. Previous work highlights that public trust in government (shaped by historical, cultural, and political factors) can dictate the perceived credibility of government-issued health advice [28]. For example, the prevalence of religious discourse in Egypt's response aligns with studies that suggest that when formal institutional trust is low, people may turn to alternative social structures, such as religious communities, to interpret health risks and determine safe behaviors [29]. This has practical implications for policy makers who might need to engage non-governmental institutions. These could include religious leaders who could help to strengthen public adherence to health measures, and this could be particularly impactful in regions where government trust is low.

Furthermore, in Italy and the United Kingdom, the focus of social media posts on government-imposed prevention measures and economic aid highlights an expectation for government responsibility and support. Research has shown that public calls for government assistance in times of crisis often reflect a belief in the government's role as a welfare provider [30]. This observation reflects policy implications in that governments with higher trust levels are likely to see more public demands for direct economic relief, as opposed to countries where citizens may not expect or request such support due to institutional limitations. For instance, economic support policies in Italy and the United Kingdom were closely scrutinized by citizens on social media, suggesting that clear communication on relief measures can help manage public expectations and reinforce trust. Meanwhile, posts in Egypt often revealed skepticism toward government actions, particularly regarding the health sector's capacity to manage the pandemic. This finding highlights the need for targeted interventions to improve transparency and crisis management to build institutional trust, which is important for effective public health responses [31].

Regarding limitations, the study used the X Search API, which meant that certain posts might be missing from our analysis. Moreover, future research could seek to examine the influence of X and conduct an analysis of posts to identify users who were most influential by betweenness centrality across the 3 time periods. Further research could also seek to follow up with certain X users for in-depth qualitative interviews. Moreover, it is important to note that text-based social media data, such as posts from X, can reflect biases influenced by demographics, geographic location, and how users act on a daily basis. These factors may result in varying word choices and sentiment, even when discussing similar topics. For example, geographic differences can affect how happiness is expressed on the web [32], while daily routines and spatiotemporal context can shape sentiments related to specific locations or events [33,34]. It is important to acknowledge these biases because such factors could influence the patterns observed in our analysis.

XSL•FO
RenderX

Additionally, the population of users who geo-tag their posts may not be representative of the larger population, which may introduce further limitations when generalizing results [35,36].

Theoretical and Policy Contribution

Our study found that users on X were observing and being presented with unfolding events that they would assess in relation to their current location and their own situation. These factors would then lead to sense-making among users, and they would repeat the cycle as they were being presented with new information. Figure 2 provides an overview of the process that was taking place.

Figure 2. A visual overview of our proposed cycle of sense-making.

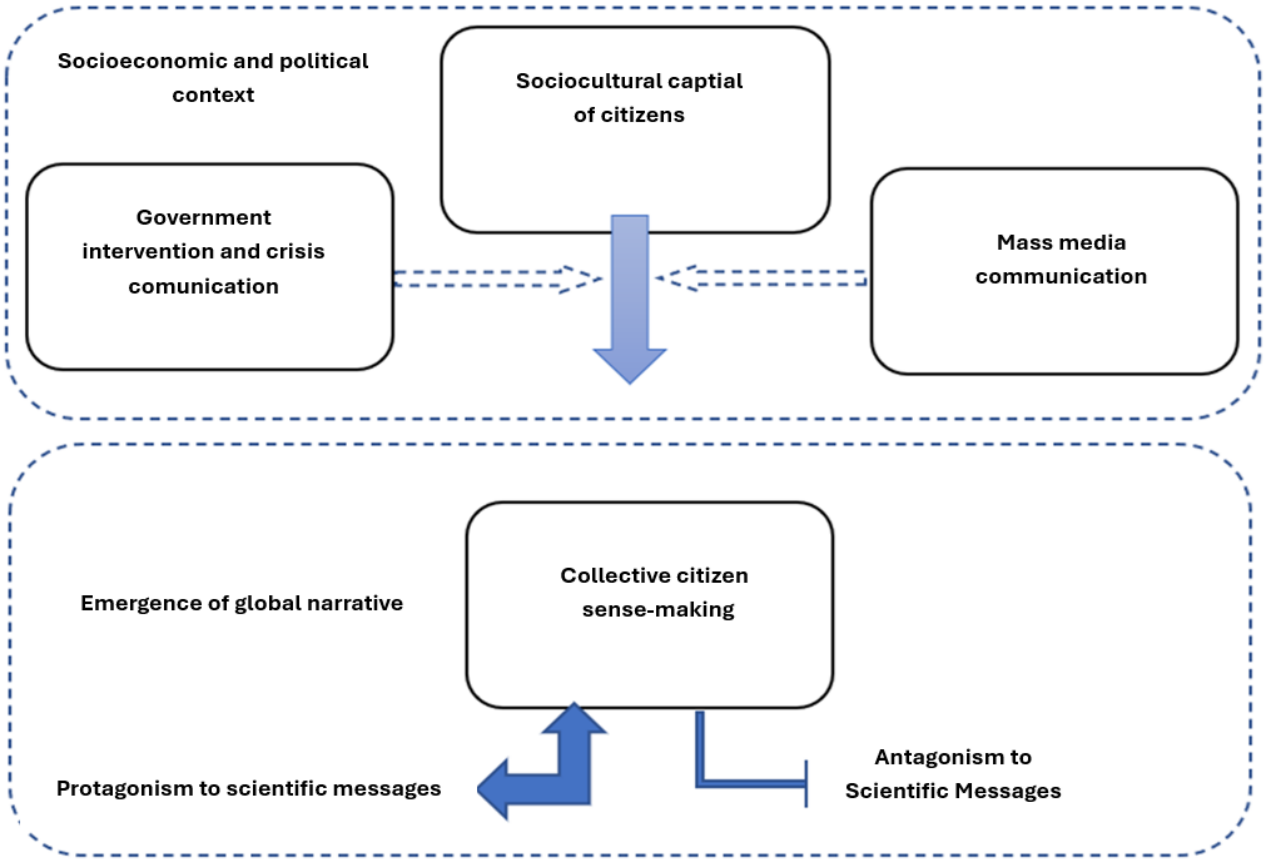

In this study, we interpret "sense-making" as the process by which individuals can interpret and assign meaning to complex events such as the COVID-19 pandemic by engaging with public discourse and personal reflection. Unlike opinion-sharing, sense-making involves actively integrating new information to form coherent narratives [37]. While prior work, such as Stieglitz et al [38], explores sense-making during crises using sentiment and social network analysis, our study is different because it focuses on topic modeling to capture thematic patterns in discourse. Topic modeling, unlike sentiment analysis, enables us to explore the substantive content of conversations, identifying recurring themes beyond emotional valence and offering insights into how users contextualize the pandemic. Zolyomi and Snyder [39] applied a process model and manual coding to study sense-making in the #ActuallyAutistic community and focused on identity formation through qualitative analysis. However, our topic modeling approach is designed for large-scale analysis. This allowed us to reveal dominant narratives across diverse sociocultural contexts rather than process-tracing individual experiences. Social media platforms like X support collective sense-making during crises by providing spaces for users to negotiate meaning and validate perspectives [40]. By analyzing trends and patterns in sampled hashtags, we capture how individuals construct understanding beyond expressing opinions. This allowed us to reveal high-level themes that illuminate public sentiment across distinct institutional settings.

Based on our proposition derived from the findings of this study, "sense-making of events during COVID-19 on X differs based on the location of a user and their relationship with the location," we have several practical insights, recommendations, and suggestions for public health authorities and governments (Textbox 4).

XSL•FO
RenderX

Textbox 4. Practical insights, recommendations, and suggestions for public health authorities and governments.

- Users of social media platforms such as X are likely to have varying information needs at different points of a pandemic.
- Researchers analyzing X data may wish to segregate data based on location.
- The information needs of users are likely to differ based on their location.
- Users' sense-making and reaction process will take place in relation to their past experiences.
- Considering the above, information provided through the likes of X should be highly specific and targeted.
- Topic modeling methods can be used for social listening purposes to identify common queries and concerns for different locations.
- This information can feed into local and international health messaging.

Our research identifies unique patterns in how citizens across different regions use social media during pandemics, which has direct implications for targeted health communication strategies.

For instance, a public health official in Egypt, where our study noted significant religious and cultural influences on information dissemination, could leverage our insights to tailor messaging that resonates with local cultural contexts. During a pandemic, when misinformation can spread rapidly alongside the virus itself, understanding these nuances allows for the development of culturally sensitive, effective public health campaigns. Officials could use localized social listening tools, guided by our findings, to monitor real-time public sentiment and misinformation trends. This enables them to quickly address misconceptions and provide clear, trustworthy information specifically designed to engage and educate distinct audiences effectively.

Moreover, our study's suggestion to segregate data based on geographic and cultural lines could help these officials to identify and respond to emerging local issues more rapidly, potentially curbing the spread of the disease by improving compliance with health directives tailored to the community's prevalent beliefs and behaviors.

This actionable approach not only demonstrates the direct applicability of our findings but also reinforces the value of our research in enhancing public health responses through strategic communication tailored to diverse populations during critical times.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to legal sharing restrictions in place by X.

Conflicts of Interest

None declared.

References

1. COVID - Coronavirus Statistics - Worldometer. URL: https://www.worldometers.info/coronavirus/ [accessed 2023-03-31]
2. Ducharme J. World Health Organization Declares COVID-19 a ‘Pandemic.’ Here’s What That Means. 2020. URL: https://time.com/5791661/who-coronavirus-pandemic-declaration/ [accessed 2023-03-31]
3. Muselli M, Cofini V, Desideri G, Necozione S. Coronavirus (Covid-19) pandemic: how may communication strategies influence our behaviours? Int J Disaster Risk Reduct 2021;53:101982 [FREE Full text] [doi: 10.1016/j.ijdrr.2020.101982] [Medline: 33251100]
4. Christianson MK, Barton MA. Sensemaking in the time of COVID‐19. J Manag Stud 2020;58(2):572-576. [doi: 10.1111/joms.12658]
5. Zolyomi A, Jones R, Kaftan T. #ActuallyAutistic sense-making on Twitter. New York, NY, USA: ACM; 2020 Presented at: Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility; October 29, 2020; Virtual Event, Greece p. 1-4.
6. Ahmed W, Bath PA, Sbaffi L, Demartini G. Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. Health Info Libr J 2019;36(1):60-72 [FREE Full text] [doi: 10.1111/hir.12247] [Medline: 30663232]
7. Rudolph CW, Allan B, Clark M, Hertel G, Hirschi A, Kunze F, et al. Pandemics: implications for research and practice in industrial and organizational psychology. Ind Organ Psychol: Perspect Sci Pract 2021;14(1-2):1-35. [doi: 10.1017/iop.2020.48]
8. Naeem M, Ozuem W. Understanding misinformation and rumors that generated panic buying as a social practice during COVID-19 pandemic: evidence from Twitter, YouTube and focus group interviews. Inform Technol People 2021;35(7):2140-2166. [doi: 10.1108/itp-01-2021-0061]

XSL•FO
RenderX

9. Pine KH, Lee M, Whitman SA, Chen Y, Henne K. Making sense of risk information amidst uncertainty: individuals' perceived risks associated with the COVID-19 pandemic. New York, NY, USA: ACM; 2021 Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 2021 May 07; Yokohama, Japan p. 1-15.
10. Tsoy D, Tirasawasdichai T, Kurpayanidi K. Role of social media in shaping public risk perception during COVID-19 pandemic: a theoretical review. Int J Bus Manag Sci Bus Admin 2021;7(2):35-41.
11. Oh SH, Lee SY, Han C. The effects of social media use on preventive behaviors during infectious disease outbreaks: the mediating role of self-relevant emotions and public risk perception. Health Commun 2021;36(8):972-981. [doi: 10.1080/10410236.2020.1724639] [Medline: 32064932]
12. Hassan MS, Halbusi HA, Najem A, Razali A, Williams KA. Impact of risk perception on trust in government and self-efficiency during COVID-19 pandemic: does social media content help users adopt preventative measures? Res Sq 2022.
13. World Bank. Human capital index (HCI) dataset. World Bank Data Catalog. 2023. URL: https://datacatalog.worldbank.org/search/dataset/0038030/Human-Capital-Index [accessed 2024-12-10]
14. Shahi GK, Dirkson A, Majchrzak TA. An exploratory study of COVID-19 misinformation on Twitter. Online Soc Netw Media 2021;22:100104 [FREE Full text] [doi: 10.1016/j.osnem.2020.100104] [Medline: 33623836]
15. Mittal R, Ahmed W, Mittal A, Aggarwal I. Twitter users' coping behaviors during the COVID-19 lockdown: an analysis of tweets using mixed methods. Inf Discov Deliv 2021 [FREE Full text]
16. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
17. Burel G, Farrell T, Alani H. Demographics and topics impact on the co-spread of COVID-19 misinformation and fact-checks on Twitter. Inf Process Manag 2021;58(6):102732 [FREE Full text] [doi: 10.1016/j.ipm.2021.102732] [Medline: 34511703]
18. Ahmed W, Vidal-Alaball J, Lopez Segui F, Moreno-Sánchez PA. A social network analysis of tweets related to masks during the COVID-19 pandemic. Int J Environ Res Public Health 2020;17(21):8235 [FREE Full text] [doi: 10.3390/ijerph17218235] [Medline: 33171843]
19. Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, et al. COVID-19 vaccine discourse on twitter: a content analysis of persuasion techniques, sentiment and mis/disinformation. J Health Commun 2021;26(7):443-459. [doi: 10.1080/10810730.2021.1955050] [Medline: 34346288]
20. Latkin CA, Dayton L, Strickland JC, Colon B, Rimal R, Boodram B. An assessment of the rapid decline of trust in US sources of public information about COVID-19. In: Vaccine Communication in a Pandemic. London, United Kingdom: Routledge; 2023:22-31.
21. Wang PW, Lu WH, Ko NY, Chen YL, Li DJ, Chang YP, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: facebook survey study in Taiwan. J Med Internet Res 2020;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
22. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on facebook during the COVID-19 pandemic in early 2020: cross-country comparison. J Med Internet Res 2020;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
23. Heverin T, Zach L. Use of microblogging for collective sense‐making during violent crises: a study of three campus shootings. J Am Soc Inf Sci 2011;63(1):34-47. [doi: 10.1002/asi.21685]
24. Maitlis S, Christianson M. Sensemaking in organizations: taking stock and moving forward. Acad Manag Ann 2014;8(1):57-125. [doi: 10.1080/19416520.2014.873177]
25. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
26. Yang P, Yao Y, Zhou H. Leveraging global and local topic popularities for LDA-based document clustering. IEEE Access 2020;8:24734-24745. [doi: 10.1109/access.2020.2969525]
27. Tafesse W, Wien A. A framework for categorizing social media posts. Cogent Bus Manag 2017;4(1):1284390. [doi: 10.1080/23311975.2017.1284390]
28. Siegrist M, Zingg A. The role of public trust during pandemics: implications for crisis communication. Eur Psychol 2014;19(1):23-32 [FREE Full text]
29. Scrimshaw SC, Lane SD, Rubinstein RA, Fisher J. The SAGE Handbook of Social Studies in Health and Medicine. Washington, DC: SAGE Publications Ltd; 2022.
30. Putnam RD. Bowling Alone: The Collapse and Revival of American Community. New York, NY: Simon & Schuster; 2000.
31. Tangcharoensathien V, Mills A, Palu T. Accelerating health equity: the key role of universal health coverage in the sustainable development goals. BMC Med 2018;16(1):102 [FREE Full text]
32. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS One 2013;8(5):e64417 [FREE Full text] [doi: 10.1371/journal.pone.0064417] [Medline: 23734200]
33. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

XSL•FO
RenderX

34. Gore R, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
35. Issa E, Tsou M, Nara A, Spitzberg B. Understanding the spatio-temporal characteristics of Twitter data with geotagged and non-geotagged content: two case studies with the topic of flu and ted (movie). Ann GIS 2017;23(3):219-235 [FREE Full text] [doi: 10.1080/19475683.2017.1343257]
36. Chappell P, Tse M, Zhang M, Moore S. Using GPS geo-tagged social media data and geodemographics to investigate social differences: a twitter pilot study. Sociol Res Online 2017;22(3):38-56 [FREE Full text] [doi: 10.1177/1360780417724065]
37. Weick KE. Sensemaking in Organizations (Foundations for Organizational Science). London: SAGE Publications, Inc; 1995.
38. Stieglitz S, Bunker D, Mirbabaie M, Ehnis C. Sense‐making in social media during extreme events. Conting Crisis Manag 2018;40:85-96 [FREE Full text] [doi: 10.1111/1468-5973.12193]
39. Zolyomi A, Snyder J. #ActuallyAutistic: an exploratory study of autism community dynamics on Twitter. 2018 Presented at: Proceedings of the 10th International Conference on Social Media and Society; July 19-21, 2019; Toronto, Canada p. 139-145 URL: https://doi.org/10.1145/3217804.3217907
40. Oh O, Agrawal M, Rao HR. Community intelligence and social media services: a rumor theoretic analysis of tweets during social crises. MIS Q 2013;37(2):407-426 [FREE Full text]

Abbreviations

LDA: latent Dirichlet allocation
RQ: research question

Edited by A Mavragani; submitted 21.03.24; peer-reviewed by R Gore, M Mackert, L Hemphill; comments to author 11.10.24; revised version received 31.10.24; accepted 27.11.24; published 27.01.25.

Please cite as:
Kahlawi A, Masri F, Ahmed W, Vidal-Alaball J
Cross-Cultural Sense-Making of Global Health Crises: A Text Mining Study of Public Opinions on Social Media Related to the COVID-19 Pandemic in Developed and Developing Economies
J Med Internet Res 2025;27:e58656
URL: https://www.jmir.org/2025/1/e58656
doi:10.2196/58656
PMID:

©Adham Kahlawi, Firas Masri, Wasim Ahmed, Josep Vidal-Alaball. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Description of Weight-Related Content and Recommended Dietary Behaviors for Weight Loss Frequently Reposted on X (Twitter) in English and Japanese: Content Analysis

Fumi Oono[1,2,3], MSc, RD, PhD; Mai Matsumoto[3], MSc, RD, PhD; Risa Ogata[4], RD; Mizuki Suga[5], RD; Kentaro Murakami[5], BEd, MSc, PhD

[1]Department of Social and Preventive Epidemiology, Division of Health Sciences and Nursing, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

[2]Department of Epidemiology & Preventive Medicine, Graduate School of Medicine, Gifu University, Gifu, Japan

[3]Department of Nutritional Epidemiology and Shokuiku, National Institutes of Biomedical Innovation, Health, and Nutrition, Osaka, Japan

[4]Department of Food and Nutrition Science, Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo, Japan

[5]Department of Social and Preventive Epidemiology, School of Public Health, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Mai Matsumoto, MSc, RD, PhD
Department of Nutritional Epidemiology and Shokuiku
National Institutes of Biomedical Innovation, Health, and Nutrition
7 Chome-6-8 Saitoasagi, Ibaraki
Osaka, 567-0085
Japan
Phone: 81 663841124
Email: m-matsumoto@nibiohn.go.jp

Abstract

Background: Both obesity and underweight are matters of global concern. Weight-related content frequently shared on social media can reflect public recognition and affect users' behaviors and perceptions. Although X (Twitter) is a popular social media platform, few studies have revealed the content of weight-related posts or details of dietary behaviors for weight loss shared on X.

Objective: This study aims to describe body weight–related content frequently reposted on X, with a particular focus on dietary behaviors for weight loss, in English and Japanese.

Methods: We collected English and Japanese X posts related to human body weight having over 100 reposts in July 2023 using an application programming interface tool. Two independent researchers categorized the contents of the posts into 7 main categories and then summarized recommended weight loss strategies.

Results: We analyzed 815 English and 1213 Japanese posts. The most popular main category of the content was "how to change weight" in both languages. The Japanese posts were more likely to mention "how to change weight" (n=571, 47.1%) and "recipes to change weight" (n=114, 9.4%) than the English posts (n=195, 23.9% and n=10, 1.2%, respectively), whereas the English posts were more likely to mention "will or experience to change weight" (n=167, 20.5%), "attitudes toward weight status" (n=78, 9.6%), and "public health situation" (n=44, 5.4%) than Japanese posts. Among 146 English and 541 Japanese posts about weight loss strategies, the predominant strategies were diet (n=76, 52.1% in English and n=170, 31.4% in Japanese) and physical activities (n=56, 38.4% and n=295, 54.5%, respectively). The proportion of posts mentioning both diet and physical activity was smaller in Japanese (n=62, 11.5%) than in English (n=31, 21.2%). Among 76 English and 170 Japanese posts about dietary behaviors for weight loss, more than 60% of posts recommended increasing intakes of specific nutrients or food groups in both languages. The most popular dietary component recommended to increase was vegetables in both English (n=31, 40.8%) and Japanese (n=48, 28.2%), followed by protein and fruits in English and grains or potatoes and legumes in Japanese. Japanese posts were less likely to mention reducing energy intake; meal timing or eating frequency; or reducing intakes of specific nutrients or food groups than the English posts. The most popular dietary component recommended to decrease was alcohol in English and confectioneries in Japanese.

Conclusions: This study characterized user interest in weight management and suggested the potential of X as an information source for weight management. Although weight loss strategies related to diet and physical activity were popular in both English

XSL•FO
RenderX

and Japanese, some differences in the details of the strategies were present, indicating that X users are exposed to different information in English and Japanese.

(J Med Internet Res 2025;27:e64739) doi:10.2196/64739

KEYWORDS

social networking service; X, Twitter; web-based health information; dieting; weight loss; content analysis; digital health; weight control; weight; social media; diet; dietary behavior; obesity; eating disorders; public perceptions

Introduction

Obesity and underweight pose critical challenges to public health globally. Obesity or overweight is prevalent in approximately 39% of the world's population [1] and increases the risk of diseases such as type 2 diabetes and cardiovascular diseases [2], with substantial economic cost [3,4]. Simultaneously, being underweight and eating disorders are of particular concern among young women [5-8], which may lead to higher mortality [9,10]. Notably in Japan, more than 20% of women aged 20-29 years are underweight, with a BMI of less than 18.5 kg/m^2 [11]. There is a clear need to establish strategies to maintain a healthy weight, free from obesity and underweight, for promoting health and well-being. In addition, beyond actual weight status, body image (perception of body weight and appearance) is important for mental and physical health. For example, negative body image (body dissatisfaction) can cause eating disorders and depression [12,13]. The estimated prevalence of body dissatisfaction was 11%-72% for women and 8%-61% for men in the United States [14,15], although estimates varied considerably by assessment method and population (including age and sex). Similarly, despite the low obesity rate among young Japanese women [11], 34%-84% of Japanese female adolescents perceived themselves as "fat" [16], with variations by population. Thus, body dissatisfaction is a serious concern in both Western and non-Western countries [16]. Body dissatisfaction is caused by various factors; in particular, sociocultural factors, such as exposure to mass media, play an important role [13,17]. In sociocultural theory, mass media in modern Westernized society presents unrealistic, thin beauty ideals, and individuals (especially women but also men) are encouraged to desire thinness and compare their appearance with these unrealistic ideals, resulting in body dissatisfaction [18].

In addition to traditional media such as television and magazines, the internet and social media have emerged and expanded in recent years. Similar to traditional media, the content of social media has the potential to influence users' behaviors and perceptions. A previous review showed that thin-ideal media images are associated with body dissatisfaction and eating disorders [19]. Additionally, exposure to health-risk behavior content on social media has been associated with unhealthy food intake [20] and a higher desire for junk foods [21]. Meanwhile, the content of social media can often reflect public perceptions or interests, especially when shared among many users. For example, negative attitudes toward obesity, including weight stigmatization, were found to be pervasive on social media [22-25]. Additionally, contents related to obesity prevention and individual-level causes of obesity tended to be shared [26-28], indicating public attention toward the struggle against obesity.

X (formerly Twitter), one of the most popular social media platforms, typically allows users to post messages within a character limit of 140, which are referred to as posts. Since its launch in 2006, X has grown and currently has over 200 million active users accessing it on any given day in 2022 [29]. As of 2023, the United States had the largest number of active users, at 65 million, followed closely by Japan with 52 million, while other countries had fewer than 17 million active users [29]. On X, users can publish posts to their profile pages, subscribe to other users' posts using the "following" function, and share posts of other users with their followers using the "repost" function. The repost function makes sharing content easy and efficient, allowing X to amplify information. Considering these features and the large number of users, X may be an important social media platform for health information, including weight management and body image.

Previous studies have examined obesity-related posts on X and reported a prevalence of negative attitudes toward obesity [22-25], obesity prevention strategies, and obesity causes [26-28]. Most of these studies, however, were limited to search terms such as obesity, overweight, and fat [22-26,28,30]. This resulted in a lack of comprehensiveness, such as a paucity of descriptions that included aspects of underweight and weight loss. Indeed, few studies have examined underweight and weight loss [31-34], and most of these collected posts with hashtags such as #weightloss or #diet [31] or used specific search terms such as "thinspiration" (a word combining "thin" and "inspiration") [33,34]. Additionally, although weight management strategies are major topics of posts about obesity on X [26,28,30] and diets are the predominant strategies mentioned [26,28], the details of recommended dietary behaviors have not been revealed. Given the importance of dietary behaviors in weight management [35], a detailed examination of the dietary behaviors spread on X will help in understanding what information users are exposed to and may accordingly influence their dietary behavior.

Moreover, most previous studies have examined posts that were written in English [22,24,26,28,30-34], as have studies of X in other health-related fields [36,37], despite potential variation in public health and cultural situations and perceptions across different languages. A comparison of the content of posts in multiple languages with different public health concerns may be useful in understanding the differences in public attitudes toward body weight and the information to which users are frequently exposed. In Western countries, body positivity, which encourages individuals to appreciate and accept their bodies regardless of size or appearance, has been popularized on social

XSL•FO
RenderX

media in recent years [38]. Meanwhile, body positivity remains relatively uncommon in Japan [39]. Alongside Western ideals of beauty, in Japan, social and cultural factors, such as traditional gender roles and the "kawaii" (cute) culture, may influence attitudes toward thinness [39,40]. Additionally, differences in dietary habits may also influence recommended dietary behaviors for weight loss. For example, dietary intake in Japan differs from that in Western countries, and includes a higher intake of seafood, white rice, plant food, and sodium, and a lower intake of whole grains, saturated fat, and added sugars [41,42]. The prevalence of excessive intake of saturated fat and added sugars, which can contribute to excessive energy intake and consequent obesity, is lower in Japan than in Western countries [41,43]. Also, Japanese people generally adhere to a stable meal schedule, with relatively rare meal skipping and low frequency of snacking compared to Western countries [44]. These cultural differences in dietary intake may contribute to distinct obesity rates, public attitudes toward body image, and dietary practices for weight loss.

Therefore, this study aimed to examine frequently shared content about body weight on X, with a particular focus on dietary recommendations for weight loss. Specifically, we compared content between English and Japanese, the languages used by the countries with the two largest numbers of X users (United States and Japan), and which have different weight-related concerns and sociocultural contexts.

Methods

Data Collection

The search strings in English and Japanese were determined based on previous studies [22,26,30] and the authors' knowledge, including gain or loss of body weight, obesity, or underweight (Table S1 in Multimedia Appendix 1). We collected posts using Social Insight (User Local Inc), a tool for analyzing social media. It collects posts according to the search strings using an application programming interface and has been used in research for collecting posts [45]. Social Insight does not necessarily collect all posts identified by search strings, and data on the exact collection rate were unavailable. Further, this tool did not collect protected posts (not available to the public).

This study restricted posts to those with more than 100 reposts to analyze frequently shared content. This repost cutoff point was determined considering that the top 0.1%-0.2% of posts in both languages acquired more than 100 reposts in a pilot study in May 2023. We arbitrarily aimed to estimate categories that constitute 10% of the total content with a 95% CI of ±2.5%, which required 554 posts in the final sample in both languages. Assuming half of the posts were excluded, we aimed to collect more than 1108 posts which had been reposted more than 100 times in each language. To collect more than 1108 posts, we assumed that collecting posts made over a 1-month period would be sufficient, based on a sufficient number of posts created in the pilot study (during May 2023).

We collected 1,621,010 English and 1,459,517 Japanese posts created during July 2023 using Social Insight with the search strings. Then, quotes (reposts with some additional comments) were removed, resulting in 720,451 English and 838,028 Japanese original posts (Table S1 in Multimedia Appendix 1). Among these, we extracted 1194 English and 1501 Japanese posts having over 100 reposts. We then excluded posts that: (1) were not related to body weight (eg, "The X algorithm will add weight to posts that *get…*"); (2) were related to animals (eg, "Please find this thin dog"); (3) were related to characters in anime, comic books, or drawings; (4) included sexual content; and (5) were not in one of the targeted languages (English or Japanese; Figure S1 in Multimedia Appendix 1).

Using Social Insight on August 31, 2023, we derived the following characteristics of the posts: username, user ID, user profile, the presence of pictures or videos attached to posts, and the number of following users, followers, reposts, and likes. We assumed that those posts created during July 2023 had a stable number of reposts and likes on August 31, 2023, because Social Insight collected these numbers from up to a few days (for posts with few reposts) to 1 month (for posts with many reposts) after a post was created.

Coding of Content

We manually coded each post according to the type of content [46]. The unit of analysis in this study was the post. The codebook was primarily developed using an inductive approach to content analysis [47,48]. One author (FO) developed a draft of the codebook based on approximately 200 English and 200 Japanese posts with 100 or more reposts in the pilot study. These posts were searched using the same search strings in May 2023. The findings of previous studies [26,28,30] were also used to refine the codebook. Insights from both the pilot data and previous research were used to accurately characterize the content of the posts while maintaining consistency with previous research. The author (FO) then coded sample posts in accordance with the draft codebook to organize and revise it. The other authors (MM, RO, and MS) then categorized the same posts and further revised the codebook by consensus.

We developed seven categories of content: (1) how to change weight (causes, habits, or strategies), (2) will or experience to change weight (including reports of action to change weight), (3) recipes to change weight, (4) attitude about weight status or appearance, (5) results or effects of weight change (including health outcomes, appearance, and social treatment), (6) public health situation (eg, prevalence of obesity), and (7) miscellaneous (including jokes). We then developed subcategories within each category using the same process, resulting in 34 subcategories (Table S2 in Multimedia Appendix 1).

According to the codebook, each post was categorized by 2 of 4 coders independently. One English-native coder holding a PhD in psychology categorized all English posts; another Japanese-native coder (researcher; MS) categorized all Japanese posts. Additionally, both of the other 2 coders (Japanese native researchers with English as a second language; FO and RO) categorized half of the English and Japanese posts. Upon the completion of categorization, disagreements between the 2 independent coders were resolved through discussions among the coder and researchers. Before categorizing the posts, the

XSL•FO
RenderX

English-native coder received 2 hours of instruction in the codebook and coded sample posts.

We further examined the details of the strategies and recipes for weight loss using an inductive approach [47,48]. After overviewing all posts, the researchers (FO, MM, RO, and MS) further modified the codebook by creating categories of strategies and recipes that reflected the frequency of their mention. Dietary habits and physical activity were further categorized into detailed methods. Three independent researchers (FO for posts in both languages, RO for posts in English, and MS for posts in Japanese) categorized the posts in accordance with the codebook, and any disagreements were resolved through discussion with a fourth researcher (MM). Additionally, we examined whether the posts had a thread (a series of connected posts from 1 person to provide additional content by connecting multiple posts together) or not. Information in threads was also used in categorizing the content (main and subcategories) and weight loss strategies.

Users were categorized into the following five types: (1) government and academic institutions (eg, World Health Organization), (2) hospitals or clinics, (3) media outlets, (4) businesses (eg, food or health-related companies, sports teams), and (5) individuals and others (eg, fitness coaches, anonymous accounts). These categories were created in consideration of previous studies [49,50]. The users were categorized independently by 2 researchers (FO and MS), and any disagreements were resolved through discussion with a third researcher (MM).

Ethical Considerations

This study was exempt from ethical approval because of its use of publicly available data without human participants. To maintain X users' anonymity in accordance with recommendations, the study does not provide any identifying information or direct quotes [51].

Statistical Analysis

Data were described using the number and percentages of posts. The 95% CIs for the proportions of each category were calculated using the Clopper-Pearson method. Spearman correlation coefficients were used to examine the correlation between the number of reposts and likes. We compared the content between English and Japanese posts using the chi-square test. When the expected frequency was less than 5 in more than 20% of category values, the Fisher exact test was used instead of the chi-square test. Cohen κ and its 95% CI were calculated to assess interrater agreement in categorization by the 2 independent coders. We considered a κ coefficient between 0.40 and 0.60 as moderate agreement, 0.61 and 0.80 as substantial agreement, and 0.81 and 1.00 as almost complete agreement [52]. All analyses were performed using SAS statistical software (version 9.4; SAS Institute Inc), with 2-tailed *P* values <.05 considered statistically significant.

Results

This study included 815 English and 1213 Japanese posts having over 100 reposts related to human body weight and weight management. Among them, 52% of the English and 22% of the Japanese posts included words related to obesity or weight gain, whereas 56% of the English and 83% of the Japanese posts included words related to underweight or weight loss (Table S1 in Multimedia Appendix 1). Approximately 45% of English and 32% of Japanese posts had less than 200 reposts, whereas 10% of English and 16% of Japanese posts had 1000 or more (Table 1). The number of reposts was correlated with the number of likes (Spearman correlation coefficients of 0.65 in English and 0.73 in Japanese). The collected posts were created by 652 English and 642 Japanese accounts. More than 95% of these accounts were individuals and others in both languages (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of 815 English and 1213 Japanese body weight–related posts with more than 100 reposts during July 2023.

	English, n (%)	Japanese, n (%)
Reposts		
Less than 200	364 (44.7)	393 (32.4)
200 to 999	372 (45.6)	530 (51.9)
1000 to 9999	77 (9.5)	186 (15.3)
10,000 or more	2 (0.3)	4 (0.3)
Likes		
Less than 200	28 (3.4)	64 (5.3)
200 to 999	213 (26.1)	212 (17.5)
1000 to 9999	527 (64.7)	791 (65.2)
10,000 or more	47 (5.8)	146 (12.0)
Account[a]		
Government and academic institutions	3 (0.5)	0 (0)
Hospitals or clinics	0 (0)	1 (0.2)
Media outlets	13 (2)	14 (2.2)
Businesses	8 (1.2)	15 (2.3)
Individuals and others[b]	628 (96.3)	612 (95.3)

[a]n=652 in English and n=642 in Japanese.

[b]Including anonymous accounts.

As shown in Table 2, although the most popular content was "how to change weight" in both languages, the English and Japanese posts differed in the proportion of the main categories of their contents. The Japanese posts were more likely to mention "how to change weight" and "recipes," whereas the English posts were more likely to mention "will or experience," "attitudes toward weight status," and "public health situations." The detailed subcategories of contents also differed between the English and Japanese posts (Table S2 in Multimedia Appendix 1). For example, 5.4% (n=44) of English posts showed negative attitudes toward obesity or gaining weight, versus only 0.9% (n=11) of Japanese posts. Excluding the "miscellaneous" category, subcategories with more than 5% of the posts were: "how to lose weight" and "negative attitudes toward obesity or weight gain" in English; and "how to lose weight," "will to lose weight," and "recipes for weight loss" in Japanese.

Table 2. Main contents of 815 English and 1213 Japanese body weight–related posts with more than 100 reposts during July 2023[a].

	English		Japanese	
	Value, n	Percentage (95% CI)	Value, n	Percentage (95% CI)
How to change weight	195	23.9 (21.0-27.0)	571	47.1 (44.2-50.0)
Will or experience to change weight	167	20.5 (17.7-23.4)	149	12.3 (10.5-14.3)
Recipes to change weight	10	1.2 (0.6-2.0)	114	9.4 (7.8-11.2)
Express negative or positive attitudes toward weight status or appearance	78	9.6 (7.6-11.8)	31	2.6 (2.0-3.6)
Results or effects of weight change	27	3.3 (2.2–4.8)	56	4.6 (3.5-6.0)
Public health situations	44	5.4 (4.0-7.2)	11	0.9 (0.5-1.6)
Miscellaneous	294	36.1 (32.8-39.5)	281	23.2 (20.8-25.6)

[a]$P<.0001$ for differences in content between English and Japanese posts (chi-square test).

Among 146 English and 541 Japanese posts mentioning weight loss strategies, diets and physical activities were frequently mentioned (Table 3). Diets were more likely to be mentioned in English (n=76, 52.1%) than in Japanese posts (n=170, 31.4%), whereas physical activities were more likely to be mentioned in Japanese (n=295, 54.5%) than in English posts (n=56, 38.4%). The proportion of posts mentioning both diet and physical activity was smaller in Japanese (n=62, 11.5%) than in English (n=31, 21.2%). In English posts, other strategies frequently mentioned for weight loss were supplements, medicines, or vaccines (n=25, 17.1%); sleeping (n=24, 16.4%); and drinking water (n=21, 14.4%). In Japanese posts, 11.3% (n=61) included

XSL•FO
RenderX

campaigning for products or apps for weight loss, such as "someone who reposts this post will get our products for weight loss." Although the number of posts was small, bathing and sauna and improving posture were unique to Japanese posts, while motivation, reducing stress, sunlight, and bariatric surgery were unique to English posts. The proportion of posts with a thread was higher in English (n=53, 36.3%) than in Japanese (n=15, 2.8%).

Table 3. Recommended strategies (including habits) for weight loss in 146 English and 541 Japanese posts having more than 100 reposts during July 2023.

	English		Japanese		*P* value[a]
	Value, n	Percentage (95% CI)	Value, n	Percentage (95% CI)	
Dietary intake and habits	76	52.1 (43.6-60.4)	170	31.4 (27.5-35.5)	<.001
Physical activity (including exercise)	56	38.4 (30.4-46.8)	295	54.5 (50.2-58.8)	<.001
Including both diet and physical activity	31	21.2 (14.9-28.8)	62	11.5 (8.9-14.5)	.002
Drinking water	21	14.4 (9.1-21.1)	44	8.1 (6.0-10.8)	.02
Sleeping	24	16.4 (10.8-23.5)	14	2.6 (1.4-4.3)	<.001
Bathing or sauna	1	0.7 (0.02-3.0)	41	7.6 (5.5-10.1)	.002
Supplements, medicines, or vaccines	25	17.1 (11.4-24.2)	23	4.3 (2.7-6.3)	<.001
Campaigning products or apps for weight loss	3	2.1 (0.4-5.9)	61	11.3 (8.7-14.3)	<.001
Motivation	11	7.5 (3.8-13.1)	9	1.7 (0.08-3.1)	<.001
Reducing stress	9	6.2 (2.9-11.4)	3	0.6 (0.1-1.6)	<.001
Sunlight	6	4.1 (1.5-8.7)	1	0.2 (0.0-1.0)	<.001
Bariatric surgery	2	1.4 (0.2-4.9)	0	0 (N/A[b])	.045
Measuring (monitoring) body weight	5	3.4 (1.1-7.8)	12	2.2 (1.2-3.8)	.38
Improving posture	0	0 (N/A)	11	2.0 (1.0-3.6)	.01
Posts with threads[c]	53	36.3 (28.5-44.7)	15	2.8 (1.6-4.5)	<.001

[a]*P* values for the chi-square test. When the expected frequency was less than 5 in more than 20% of category values, a 2-tailed Fisher exact test was used.

[b]N/A: not applicable.

[c]A series of connected posts from 1 person to provide additional content by connecting multiple posts together.

The details of dietary behaviors for weight loss also differed between English and Japanese posts (Table 4). The proportion of posts that recommended reducing energy intake was lower in Japanese (n=21, 12.4%) than in English (n=23, 30.3%). Also, Japanese posts were less likely to mention meal timing or eating frequency (n=37, 21.8%) than English posts (n=26, 34.2%). In both languages, more than 60% of posts recommended increasing the intake of specific nutrients or food groups. The most popular component recommended to be increased was vegetables in both English (n=31, 40.8%) and Japanese (n=48, 28.2%), followed by protein and fruits in English and grains or potatoes and legumes in Japanese. Natto (fermented soy) and miso soup were unique to Japanese posts, whereas "nutrient-dense foods" (including a "nutrient-dense diet" and "high-quality foods") were unique to English posts. English posts were more likely to recommend decreasing the intake of specific nutrients or food groups (n=31, 40.8%) than Japanese posts (n=45, 26.5%). The most popular component recommended to decrease was alcohol in English (n=22, 28.9%) and sweets/confectioneries in Japanese (n=16, 9.4%).

XSL•FO
RenderX

Table 4. Recommended dietary strategies for weight loss in 76 English and 170 Japanese posts having more than 100 reposts during July 2023.

	English		Japanese		*P* value[a]
	Value, n	Percentage (95% CI)	Value, n	Percentage (95% CI)	
Reduce energy intake or amount of eating	23	30.3 (20.3-41.9)	21	12.4 (7.8-18.3)	<.001
Meal timing or eating frequency	26	34.2 (23.7-46.0)	37	21.8 (15.8-28.7)	.04
Mention the name of specific food or drink product(s)	11	14.5 (7.5-24.4)	33	19.4 (13.8-26.2)	.35
Increase intake of specific nutrients or food groups	53	69.7 (58.1-79.8)	105	61.8 (54.0-69.1)	.23
Meat	23	30.3 (20.3-41.9)	29	17.1 (11.7-23.6)	.02
Fish	22	28.9 (19.1-40.5)	25	14.7 (9.8-20.9)	.008
Legumes	8	10.5 (4.7-19.7)	32	18.8 (13.3-25.5)	.10
Natto	0	0 (N/A[b])	19	11.2 (6.9-16.9)	.002
Nuts	9	11.8 (5.6-21.3)	7	4.1 (1.7-8.3)	.02
Eggs	17	22.4 (13.6-33.4)	21	12.4 (7.8-18.3)	.04
Protein drinks	11	14.5 (7.5-24.4)	11	6.5 (3.3-11.3)	.08
Grains or potatoes	13	17.1 (9.4-27.5)	32	18.8 (13.3-25.5)	.75
Whole grains	4	5.3 (1.5-12.9)	22	12.9 (8.3-18.9)	.07
Vegetables (including mushrooms)	31	40.8 (29.7-52.7)	48	28.2 (21.6-35.6)	.051
Seaweeds	5	6.6 (2.2-14.7)	8	4.7 (2.0-9.1)	.54
Fruits	27	35.5 (24.9-47.3)	22	12.9 (8.3-18.9)	<.001
Dairy products	17	22.4 (13.6-33.4)	15	8.8 (5.0-14.1)	.003
Drinks (eg, coffee, tea)	17	22.4 (13.6-33.4)	22	12.9 (0.08-18.9)	.06
Miso soup	0	0 (N/A)	8	4.7 (0.2-9.1)	.06
Other foods (eg, chocolate)	26	34.2 (23.7-46.0)	18	10.6 (6.4-16.2)	<.001
Protein	29	38.2 (27.3-50.0)	18	10.6 (6.4-16.2)	<.001
Dietary fiber	3	3.9 (0.08-11.1)	6	3.5 (1.3-7.5)	.55
Other nutrients	4	5.3 (1.5-12.9)	15	8.8 (5.0-14.1)	.33
Nutrient-dense foods	11	14.5 (7.5-24.4)	1	0.6 (0.01-3.2)	<.001
Decrease intake of specific nutrients or food groups	31	40.8 (29.7-52.7)	45	26.5 (20.0-33.8)	.02
Alcohol	22	28.9 (19.1-40.5)	10	5.9 (2.9-10.6)	<.001
Sugar-sweetened beverages	9	11.8 (5.6-21.3)	8	4.7 (2.1-9.1)	.04
Sweets or confectioneries	3	3.9 (0.8-11.1)	16	9.4 (5.5-14.8)	.14
Fried foods	2	2.6 (0.3-9.2)	8	4.7 (2.1-9.1)	.73
Fast foods or processed foods	10	13.2 (6.5-22.9)	6	3.5 (1.3-7.5)	.009
Other foods	8	10.5 (4.7-19.7)	18	10.6 (6.4-16.2)	.99
Carbohydrates	10	13.2 (6.5-22.9)	5	2.9 (1.0-6.7)	.004
Sugars	13	17.1 (9.4-27.5)	2	1.2 (0.1-4.2)	<.001
Fats	1	1.3 (0.03-7.1)	8	4.7 (2.1-9.1)	.28
Other nutrients (eg, sodium, trans fat)	3	3.9 (0.8-11.1)	9	5.3 (2.5-9.8)	.76
Unspecified diet (eg, “diet is important”)	2	2.6 (0.3-9.2)	7	4.1 (1.7-8.3)	.73

[a]*P* values for the chi-square test. When the expected frequency was less than 5 in more than 20% of category values, a 2-tailed Fisher exact test was used.

[b]N/A: not applicable.

XSL•FO
RenderX

Recipes for weight loss were described in 10 English and 114 Japanese posts (Table S3 in Multimedia Appendix 1). In the English posts, all posts included recipes for drinks (most were water with fruits). In the Japanese posts, recipes for sweets, vegetable dishes, and protein and vegetable dishes accounted for approximately 20% of the posts each.

Among 56 English and 295 Japanese posts that mentioned physical activity for weight loss, the type of recommended physical activities differed between the English and Japanese posts (Table S4 in Multimedia Appendix 1). The English posts were more likely to recommend aerobic exercise and muscle training than the Japanese posts. On the other hand, the Japanese posts were more likely to recommend stretching and massage than the English posts. A total of 200 Japanese posts (67.8%) had attached pictures or videos on how to exercise, versus only 17 English posts (30.4%).

Cohen κ coefficients of the main contents between the 2 independent coders were 0.39 (95% CI 0.35-0.43) for 815 English posts and 0.66 (95% CI 0.63-0.69) for 1213 Japanese posts. Cohen κ coefficients regarding mention of dietary strategies for weight loss were 0.89 (95% CI 0.82-0.96) for 146 English posts and 0.88 (95% CI 0.83-0.92) for 541 Japanese posts.

Discussion

In this study, we described body weight–related content in posts having more than 100 reposts on X in English and Japanese, with a particular focus on dietary behaviors for weight loss. The proportion of main contents differed between posts in English and Japanese. English posts were more likely to mention will or experience to change weight, attitudes toward weight status, and public health situations, whereas Japanese posts were more likely to mention strategies and recipes to change weight. Nevertheless, the most popular content was how to change weight in both languages. The predominant strategies for weight loss involved diet and physical activities in both languages, but the details of the strategies differed between languages. The descriptions identified in our study may be useful in understanding which information X users share and are frequently exposed to.

This study showed that individual-level strategies for weight loss are prevalent in English and Japanese on X. Among them, the most frequently mentioned in both languages were physical activity and diet. These results are consistent with previous studies which reported that obesity-related X posts frequently mentioned physical activity and diet as causes of obesity [26] and strategies to manage obesity [28]. These results and ours suggest that X users, in both English and Japanese, are interested in physical activity and diet in managing their weight status. This in turn indicates that the X has the potential to influence users' behaviors to change weight, especially physical activity and diet. In contrast, less mention was made of public health situations and population-level causes of obesity, especially in Japanese posts. X's users may be more interested in individual-level factors than population-level factors. Considering the importance of socioecological factors in the development of obesity, it may be useful to raise public awareness that individual-level factors play only a limited role and that combating obesity requires a multilevel approach [53]. Additionally, we found that approximately 5% of English posts expressed negative attitudes toward obesity, in accordance with previous studies showing the dissemination of negative attitudes toward obesity in English posts [22,23,25]. There is also a need to address not only obesity prevention and management but also the stigma of obesity.

Although diet and physical activity were the predominant strategies for weight loss shared on X in both English and Japanese, the specific method differed between the languages. English posts were more likely to mention reducing energy intake, and both diet and physical activities than Japanese posts. Energy balance is essential for weight management, and both diet and physical activity are important [54]. Further, a higher proportion of English posts than Japanese posts had threads. English posts may tend to show a comprehensive strategy for weight loss compared to Japanese posts, partly due to the higher seriousness of obesity prevalence [4]. Nevertheless, despite the importance of reducing energy intake for weight loss [54,55], this was not often mentioned even in English posts. Also, posts tended to mention increasing intakes of specific nutrients or food groups rather than decreasing intakes of fatty or sugary foods. Therefore, weight loss strategies frequently shared on X may not include sufficient information to help users make informed decisions about weight management. In Japan, the obesity rate is much lower than in Western countries, and underweight among young women is a concern [11]. Japanese posts were more likely to be limited to showing only videos on how to exercise than English posts and less likely to mention dietary behaviors. Additionally, improving posture was unique to Japanese posts. Japanese posts may focus on body appearance rather than merely reducing body weight, which may affect body image and the desire for thinness. In any case, few posts originated from government or academic institutions. It may be beneficial for government and academic institutions to create posts that provide evidence-based, comprehensive information on weight management, including balanced diets with adequate energy intake, regular exercise, and healthy body image, and aim to get them frequently reposted.

Among posts mentioning dietary habits for weight loss, more than 60% of posts in both languages recommended increasing specific nutrients or foods. Although vegetables were frequently recommended in both languages, other recommended foods and nutrients somewhat differed between English and Japanese posts. English posts were at least 15% more likely to recommend increasing intakes of fruits and protein than Japanese posts. Additionally, nutrient-dense foods were often recommended in English posts (14.5%) but only in 1 Japanese post (0.6%). On the other hand, 11.2% of Japanese posts recommended increasing natto (fermented soy product) intake but no English post did so. These differences in recommended dietary components may reflect differences in dietary culture and user intake and interest. Current evidence supports the advantages of frequently mentioned foods and nutrients in weight management, including vegetables, fruits, and protein [35,56-58]. On the other hand, little evidence exists regarding the effect of natto intake on weight management, despite the

XSL•FO
RenderX

interest in this topic among Japanese X users. Decreasing intake of dietary components such as alcohol, sugar-sweetened beverages, sugars, and fast and processed foods was more likely to be mentioned in English than in Japanese. This may be partly explained by the higher intakes of some of these items in Western countries than in Japan [41,59-61]. Although English X users may be exposed to the importance of reducing these intakes, their intake nevertheless remains high, as is seen in the United States [41,59].

The strength of our study was its collection of X posts using broader search terms, not limited to obesity and related hashtags, across 2 languages with large numbers of users. We found that 48% of the English posts and 78% of the Japanese posts did not contain obesity-related words but were rather words related to weight loss or underweight only. This study accordingly included posts that may have been overlooked in most previous studies using obesity-related search terms only. This wide range of search terms enabled us to collect posts related to body weight on X in a broad and comprehensive manner. However, several limitations should also be mentioned. First, as in many previous studies, we are unable to estimate the collection rate of related posts. Although we collected a large number of posts, we acknowledge the possibility that our sampling strategy was not random [37].

Second, the coding process had a subjective nature despite the use of a predefined codebook. Although 2 independent coders (including at least 1 native speaker in each language) categorized contents, some subjectivity and errors cannot be ruled out. In particular, the interrater agreement did not reach moderate for the main content of English posts. Therefore, the possibility of misclassification and the difficulty of coding a post into 1 specific category should be noted.

Third, the posts were collected during a certain time window (July 2023). A previous study showed that weight loss content was more commonly posted during holidays and after holidays than before holidays [31]. Additionally, social media trends are influenced by global events and prominent figures, which can lead to a temporal increase in posts related to specific topics. Because it is unknown whether the contents of posts differ by time frame, this study should be interpreted as a snapshot of the time it was conducted. Although analyzing past data is an inevitable limitation of any research, future studies could consider examining trends over time to understand dynamic variations in social media content.

Fourth, the study did not examine whether the content of posts affects user's perceptions and behaviors. A randomized controlled trial showed that exposure to junk food-related content on Instagram increased the desire for junk food and reduced the desire for healthy foods [21]. Thus, the highly reposted content was likely to be viewed by many users and may impact at least some users' perceptions and behaviors.

Finally, it should be noted that the posts in English could have been created in any of the various countries that use English. Nevertheless, the United States has the largest number of active users, at more than 3 times the number of active users of any country other than Japan [29], and a previous study showed that the majority of posts about "healthy diet" came from the United States [62]. We therefore speculate that most of the posts in our study also came from the United States.

This study examined the content of human body weight–related posts having more than 100 reposts in English and Japanese. While we found some differences in the content of weight-related X posts between English and Japanese—such as a higher prevalence of posts with negative attitudes toward obesity in English (5%) than in Japanese—the most popular contents were weight loss strategies in both languages. While diet and exercise were predominant weight loss strategies, the proportion of posts mentioning both diet and physical activity was small in both languages. Among posts about dietary strategies for weight loss, 60%-70% recommended increasing intakes of dietary components, such as vegetables, in both languages, whereas only 30% of English and 12% of Japanese posts recommended reducing energy intake. The results of this study suggest users' major interest in weight management in both English and Japanese and the potential of X as an information source on weight management. Additionally, our results identified similarities and differences in information about body weight between the languages, which in turn indicates that challenges in the use of X for weight management differ by language.

Acknowledgments

The authors would like to thank James Anderson of Dmed for coding the content of English posts and Guy Harris DO and Hiromi Inoue of Dmed for their support with the coding. The authors declare that generative artificial intelligence was not used in any part of the manuscript writing. This work was supported by a grant (number 22FA1022) from the Ministry of Health, Labour and Welfare, Japan; by the 2022 Tochigi Prefectural Health and Nutrition Survey Analysis project; and by JSPS KAKENHI grants (JP22K02121 and JP24KJ0098). The funding organizations had no role in the design, analysis, or writing of this paper.

Data Availability

The datasets used or analyzed in this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

All authors conceptualized and designed the study. FO, MM, RO, and MS were involved in coding the posts. FO was involved in formal analysis and writing—original draft preparation. FO and MM did the interpretation of the results. MM, RO, MS, and KM were involved in writing—review and editing. All authors have read and agreed to the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
We have revised the Multimedia Appendix and re-uploaded it. Please publish the revised one.
[PDF File (Adobe PDF File), 443 KB - jmir_v27i1e64739_app1.pdf]

References

1. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism 2019;92:6-10. [doi: 10.1016/j.metabol.2018.09.005] [Medline: 30253139]
2. Hruby A, Manson JE, Qi L, Malik VS, Rimm EB, Sun Q, et al. Determinants and consequences of obesity. Am J Public Health 2016;106(9):1656-1662. [doi: 10.2105/AJPH.2016.303326] [Medline: 27459460]
3. Nagi MA, Ahmed H, Rezq MAA, Sangroongruangsri S, Chaikledkaew U, Almalki Z, et al. Economic costs of obesity: a systematic review. Int J Obes 2024;48(1):33-43. [doi: 10.1038/s41366-023-01398-y] [Medline: 37884664]
4. Kim DD, Basu A. Estimating the medical care costs of obesity in the United States: systematic review, meta-analysis, and empirical analysis. Value Health 2016;19(5):602-613 [FREE Full text] [doi: 10.1016/j.jval.2016.02.008] [Medline: 27565277]
5. Treasure J, Duarte TA, Schmidt U. Eating disorders. Lancet 2020;395(10227):899-911. [doi: 10.1016/S0140-6736(20)30059-3] [Medline: 32171414]
6. Galmiche M, Déchelotte P, Lambert G, Tavolacci MP. Prevalence of eating disorders over the 2000-2018 period: a systematic literature review. Am J Clin Nutr 2019;109(5):1402-1413 [FREE Full text] [doi: 10.1093/ajcn/nqy342] [Medline: 31051507]
7. Sakamaki R, Toyama K, Amamoto R, Liu C, Shinfuku N. Nutritional knowledge, food habits and health attitude of Chinese university students—a cross sectional study. Nutr J 2005;4:4 [FREE Full text] [doi: 10.1186/1475-2891-4-4] [Medline: 15703071]
8. Qian J, Wu Y, Liu F, Zhu Y, Jin H, Zhang H, et al. An update on the prevalence of eating disorders in the general population: a systematic review and meta-analysis. Eat Weight Disord 2022;27(2):415-428 [FREE Full text] [doi: 10.1007/s40519-021-01162-z] [Medline: 33834377]
9. Smink FRE, van Hoeken D, Hoek HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr Psychiatry Rep 2012;14(4):406-414 [FREE Full text] [doi: 10.1007/s11920-012-0282-y] [Medline: 22644309]
10. Berrington de Gonzalez A, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ, et al. Body-mass index and mortality among 1.46 million White adults. N Engl J Med 2010;363(23):2211-2219 [FREE Full text] [doi: 10.1056/NEJMoa1000367] [Medline: 21121834]
11. [National Health and Nutrition Survey in Japan, 2019] [Website in Japanese]. Ministry of Health, Labour and Welfare. 2019. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/eiyou/r1-houkoku_00002.html [accessed 2023-06-15]
12. Richard A, Rohrmann S, Lohse T, Eichholzer M. Is body weight dissatisfaction a predictor of depression independent of body mass index, sex and age? Results of a cross-sectional study. BMC Public Health 2016;16(1):863 [FREE Full text] [doi: 10.1186/s12889-016-3497-8] [Medline: 27558007]
13. Mills JS, Minister C, Samson L. Enriching sociocultural perspectives on the effects of idealized body norms: integrating shame, positive body image, and self-compassion. Front Psychol 2022;13:983534 [FREE Full text] [doi: 10.3389/fpsyg.2022.983534] [Medline: 36506975]
14. Fiske L, Fallon EA, Blissmer B, Redding CA. Prevalence of body dissatisfaction among United States adults: review and recommendations for future research. Eat Behav 2014;15(3):357-365. [doi: 10.1016/j.eatbeh.2014.04.010] [Medline: 25064281]
15. Fallon EA, Harris BS, Johnson P. Prevalence of body dissatisfaction among a United States adult sample. Eat Behav 2014;15(1):151-158. [doi: 10.1016/j.eatbeh.2013.11.007] [Medline: 24411768]
16. Chisuwa N, O'Dea JA. Body image and eating disorders amongst Japanese adolescents. a review of the literature. Appetite 2010;54(1):5-15. [doi: 10.1016/j.appet.2009.11.008] [Medline: 19941921]
17. Paterna A, Alcaraz-Ibáñez M, Fuller-Tyszkiewicz M, Sicilia Á. Internalization of body shape ideals and body dissatisfaction: a systematic review and meta-analysis. Int J Eat Disord 2021;54(9):1575-1600. [doi: 10.1002/eat.23568] [Medline: 34165215]
18. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
19. Grabe S, Ward LM, Hyde JS. The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. Psychol Bull 2008;134(3):460-476. [doi: 10.1037/0033-2909.134.3.460] [Medline: 18444705]

XSL•FO
RenderX

20. BinDhim NF, Althumiri NA, Al-Duraihem RA, Alasmary S, Alkhamaali Z, Alhabeeb AA. Association between daily use of social media and behavioral lifestyles in the Saudi community: a cross-sectional study. Front Public Health 2023;11:1254603 [FREE Full text] [doi: 10.3389/fpubh.2023.1254603] [Medline: 37876715]
21. Zeeni N, Abi Kharma J, Malli D, Khoury-Malhame M, Mattar L. Exposure to Instagram junk food content negatively impacts mood and cravings in young adults: a randomized controlled trial. Appetite 2024;195:107209. [doi: 10.1016/j.appet.2024.107209] [Medline: 38228246]
22. Chou WS, Prestin A, Kunath S. Obesity in social media: a mixed methods analysis. Transl Behav Med 2014;4(3):314-323 [FREE Full text] [doi: 10.1007/s13142-014-0256-1] [Medline: 25264470]
23. Lydecker JA, Cotter EW, Palmberg AA, Simpson C, Kwitowski M, White K, et al. Does this Tweet make me look fat? A content analysis of weight stigma on Twitter. Eat Weight Disord 2016;21(2):229-235. [doi: 10.1007/s40519-016-0272-x] [Medline: 27068174]
24. Lenzi FR, Iazzetta F. Mapping obesity and diabetes' representation on Twitter: the case of Italy. Front Sociol 2023;8:1155849 [FREE Full text] [doi: 10.3389/fsoc.2023.1155849] [Medline: 37397627]
25. Kamiński M, Wieczorek T, Kręgielska-Narożna M, Bogdański P. Tweeting about fatphobia and body shaming: a retrospective infodemiological study. Nutrition 2024;125:112497. [doi: 10.1016/j.nut.2024.112497] [Medline: 38901121]
26. So J, Prestin A, Lee L, Wang Y, Yen J, Chou WS. What do people like to 'share' about obesity? A content analysis of frequent retweets about obesity on Twitter. Health Commun 2016;31(2):193-206. [doi: 10.1080/10410236.2014.940675] [Medline: 26086083]
27. Razzak FA, Saab D, Haddad F, Antoun J. Content analysis of social media regarding obesity as a chronic disease. PeerJ Comput Sci 2023;9:e1321 [FREE Full text] [doi: 10.7717/peerj-cs.1321] [Medline: 37346663]
28. Harris JK, Moreland-Russell S, Tabak RG, Ruhr LR, Maier RC. Communication about childhood obesity on Twitter. Am J Public Health 2014;104(7):e62-e69 [FREE Full text] [doi: 10.2105/AJPH.2013.301860] [Medline: 24832138]
29. Twitter users, stats, data, and trends. DataReportal. 2023. URL: https://datareportal.com/essential-twitter-stats [accessed 2024-04-24]
30. Ghosh DD, Guha R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and geographic information system. Cartogr Geogr Inf Sci 2013;40(2):90-102 [FREE Full text] [doi: 10.1080/15230406.2013.776210] [Medline: 25126022]
31. Turner-McGrievy GM, Beets MW. Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. Transl Behav Med 2015;5(2):160-166 [FREE Full text] [doi: 10.1007/s13142-015-0308-1] [Medline: 26029278]
32. Shadroo S, Yoosefi Nejad M, Bali AO, Hosseinzadeh M, Delghandi MS. A comparison and analysis of the Twitter discourse related to weight loss and fitness. Netw Model Anal Health Inform Bioinforma 2020;9(1):23. [doi: 10.1007/s13721-020-00228-9]
33. Tiggemann M, Churches O, Mitchell L, Brown Z. Tweeting weight loss: a comparison of #thinspiration and #fitspiration communities on Twitter. Body Image 2018;25:133-138. [doi: 10.1016/j.bodyim.2018.03.002] [Medline: 29567619]
34. Harris JK, Duncan A, Men V, Shevick N, Krauss MJ, Cavazos-Rehg PA. Messengers and messages for Tweets that used #thinspo and #fitspo hashtags in 2016. Prev Chronic Dis 2018;15:E01 [FREE Full text] [doi: 10.5888/pcd15.170309] [Medline: 29300696]
35. Hutfless S, Gudzune KA, Maruthur N, Wilson RF, Bleich SN, Lau BD, et al. Strategies to prevent weight gain in adults: a systematic review. Am J Prev Med 2013;45(6):e41-e51. [doi: 10.1016/j.amepre.2013.07.013] [Medline: 24237928]
36. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187. [doi: 10.2196/17187] [Medline: 33470931]
37. Takats C, Kwan A, Wormer R, Goldman D, Jones HE, Romero D. Ethical and methodological considerations of Twitter data for public health research: systematic review. J Med Internet Res 2022;24(11):e40380 [FREE Full text] [doi: 10.2196/40380] [Medline: 36445739]
38. Cohen R, Newton-John T, Slater A. The case for body positivity on social media: perspectives on current advances and future directions. J Health Psychol 2021;26(13):2365-2373. [doi: 10.1177/1359105320912450] [Medline: 32191132]
39. Ando K, Giorgianni FE, Danthinne ES, Rodgers RF. Beauty ideals, social media, and body positivity: a qualitative investigation of influences on body image among young women in Japan. Body Image 2021;38:358-369. [doi: 10.1016/j.bodyim.2021.05.001] [Medline: 34120098]
40. Pike KM, Borovoy A. The rise of eating disorders in Japan: issues of culture and limitations of the model of 'westernization'. Cult Med Psychiatry 2004;28(4):493-531. [doi: 10.1007/s11013-004-1066-6] [Medline: 15847052]
41. Murakami K, Livingstone MBE, Fujiwara A, Sasaki S. Application of the Healthy Eating Index-2015 and the Nutrient-Rich Food Index 9.3 for assessing overall diet quality in the Japanese context: different nutritional concerns from the US. PLoS One 2020;15(1):e0228318 [FREE Full text] [doi: 10.1371/journal.pone.0228318] [Medline: 31999772]
42. Murakami K, Livingstone MBE, Sasaki S. Thirteen-year trends in dietary patterns among Japanese adults in the National Health and Nutrition Survey 2003-2015: continuous Westernization of the Japanese diet. Nutrients 2018;10(8):994 [FREE Full text] [doi: 10.3390/nu10080994] [Medline: 30061492]

XSL•FO
RenderX

43. Fujiwara A, Murakami K, Asakura K, Uechi K, Sugimoto M, Wang H, et al. Association of free sugar intake estimated using a newly-developed food composition database with lifestyles and parental characteristics among Japanese children aged 3-6 years: DONGuRI study. J Epidemiol 2019;29(11):414-423 [FREE Full text] [doi: 10.2188/jea.JE20180036] [Medline: 30344196]
44. Murakami K, Livingstone MBE, Masayasu S, Sasaki S. Eating patterns in a nationwide sample of Japanese aged 1-79 years from MINNADE study: eating frequency, clock time for eating, time spent on eating and variability of eating patterns. Public Health Nutr 2022;25(6):1515-1527 [FREE Full text] [doi: 10.1017/S1368980021000975] [Medline: 33663632]
45. Song C, Fujishiro H. Toward the automatic detection of rescue-request tweets: analyzing the features of data verified by the press. 2019 Presented at: International Conference on Information and Communication Technologies for Disaster Management (ICT-DM); December 18-20, 2019; Paris, France. [doi: 10.1109/ict-dm47966.2019.9032895]
46. Fu J, Li C, Zhou C, Li W, Lai J, Deng S, et al. Methods for analyzing the contents of social media for health care: scoping review. J Med Internet Res 2023;25:e43349 [FREE Full text] [doi: 10.2196/43349] [Medline: 37358900]
47. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
48. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
49. Batheja S, Schopp EM, Pappas S, Ravuri S, Persky S. Characterizing precision nutrition discourse on Twitter: quantitative content analysis. J Med Internet Res 2023;25:e43701 [FREE Full text] [doi: 10.2196/43701] [Medline: 37824190]
50. Murakami K, Shinozaki N, Kimoto N, Onodera H, Oono F, McCaffrey TA, et al. Web-based content on diet and nutrition written in Japanese: infodemiology study based on google trends and Google search. JMIR Form Res 2023;7:e47101 [FREE Full text] [doi: 10.2196/47101] [Medline: 37971794]
51. Ford E, Shepherd S, Jones K, Hassan L. Toward an ethical framework for the text mining of social media for health research: a systematic review. Front Digit Health 2020;2:592237 [FREE Full text] [doi: 10.3389/fdgth.2020.592237] [Medline: 34713062]
52. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174. [Medline: 843571]
53. Malik VS, Willett WC, Hu FB. Global obesity: trends, risk factors and policy implications. Nat Rev Endocrinol 2013;9(1):13-27. [doi: 10.1038/nrendo.2012.199] [Medline: 23165161]
54. Johns DJ, Hartmann-Boyce J, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Diet or exercise interventions vs combined behavioral weight management programs: a systematic review and meta-analysis of direct comparisons. J Acad Nutr Diet 2014;114(10):1557-1568 [FREE Full text] [doi: 10.1016/j.jand.2014.07.005] [Medline: 25257365]
55. Canuto R, Garcez A, de Souza RV, Kac G, Olinto MTA. Nutritional intervention strategies for the management of overweight and obesity in primary health care: a systematic review with meta-analysis. Obes Rev 2021;22(3):e13143. [doi: 10.1111/obr.13143] [Medline: 33006421]
56. Mytton OT, Nnoaham K, Eyles H, Scarborough P, Ni Mhurchu C. Systematic review and meta-analysis of the effect of increased vegetable and fruit consumption on body weight and energy intake. BMC Public Health 2014;14:886 [FREE Full text] [doi: 10.1186/1471-2458-14-886] [Medline: 25168465]
57. Hansen TT, Astrup A, Sjödin A. Are dietary proteins the key to successful body weight management? A systematic review and meta-analysis of studies assessing body weight outcomes after interventions with increased dietary protein. Nutrients 2021;13(9):3193 [FREE Full text] [doi: 10.3390/nu13093193] [Medline: 34579069]
58. van Baak MA, Mariman ECM. Dietary strategies for weight loss maintenance. Nutrients 2019;11(8):1916 [FREE Full text] [doi: 10.3390/nu11081916] [Medline: 31443231]
59. Singh GM, Micha R, Khatibzadeh S, Shi P, Lim S, Andrews KG, et al. Global, regional, and national consumption of sugar-sweetened beverages, fruit juices, and milk: a systematic assessment of beverage intake in 187 countries. PLoS One 2015;10(8):e0124845 [FREE Full text] [doi: 10.1371/journal.pone.0124845] [Medline: 26244332]
60. Shinozaki N, Murakami K, Asakura K, Masayasu S, Sasaki S. Consumption of highly processed foods in relation to overall diet quality among Japanese adults: a nationwide study. Public Health Nutr 2023;26(9):1784-1797 [FREE Full text] [doi: 10.1017/S1368980023000721] [Medline: 37092752]
61. Manthey J, Shield KD, Rylett M, Hasan OSM, Probst C, Rehm J. Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. Lancet 2019;393(10190):2493-2502. [doi: 10.1016/S0140-6736(18)32744-2] [Medline: 31076174]
62. Lynn T, Rosati P, Leoni Santos G, Endo PT. Sorting the healthy diet signal from the social media expert noise: preliminary evidence from the healthy diet discourse on Twitter. Int J Environ Res Public Health 2020;17(22):8557 [FREE Full text] [doi: 10.3390/ijerph17228557] [Medline: 33218105]

XSL•FO
RenderX

Edited by A Mavragani; submitted 26.07.24; peer-reviewed by CA Almenara, YY Kristian; comments to author 28.10.24; revised version received 12.12.24; accepted 17.12.24; published 07.02.25.

Please cite as:
Oono F, Matsumoto M, Ogata R, Suga M, Murakami K
Description of Weight-Related Content and Recommended Dietary Behaviors for Weight Loss Frequently Reposted on X (Twitter) in English and Japanese: Content Analysis
J Med Internet Res 2025;27:e64739
URL: https://www.jmir.org/2025/1/e64739
doi:10.2196/64739
PMID:39918849

©Fumi Oono, Mai Matsumoto, Risa Ogata, Mizuki Suga, Kentaro Murakami. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Data-Sharing Statements Requested from Clinical Trials by Public, Environmental, and Occupational Health Journals: Cross-Sectional Study

Yingxin Liu[1*], MPH; Jingyi Zhang[1*], MM; Lehana Thabane[2,3,4], PhD; Xuerui Bai[5], MPH; Lili Kang[6], MPH; Gregory Y H Lip[7,8], PhD; Harriette G C Van Spall[2,9], PhD; Min Xia[10], PhD; Guowei Li[1,3], PhD

[1]Center for Clinical Epidemiology and Methodology, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, China

[2]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[3]Father Sean O'Sullivan Research Centre, St Joseph's Healthcare Hamilton, Hamilton, ON, Canada

[4]Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa

[5]Department of Epidemiology, School of Medicine, Jinan University, Guangzhou, China

[6]School of Public Health, Guangdong Pharmaceutical University, Guangzhou, China

[7]Liverpool Centre for Cardiovascular Sciences at University of Liverpool, Liverpool Heart & Chest Hospital, Liverpool John Moores University, Liverpool, United Kingdom

[8]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[9]Department of Medicine, McMaster University, Hamilton, ON, Canada

[10]Guangdong Provincial Key Laboratory of Food, Nutrition and Health & Department of Nutrition, School of Public Health, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Guowei Li, PhD
Center for Clinical Epidemiology and Methodology
The Affiliated Guangdong Second Provincial General Hospital of Jinan University
466 Newport Middle Road, Haizhu District
Guangzhou, 510317
China
Phone: 86 02089169546
Email: ligw@gd2h.org.cn

Abstract

Background: Data sharing plays a crucial role in health informatics, contributing to improving health information systems, enhancing operational efficiency, informing policy and decision-making, and advancing public health surveillance including disease tracking. Sharing individual participant data in public, environmental, and occupational health trials can help improve public trust and support by enhancing transparent reporting and reproducibility of research findings. The International Committee of Medical Journal Editors (ICMJE) requires all papers to include a data-sharing statement. However, it is unclear whether journals in the field of public, environmental, and occupational health adhere to this requirement.

Objective: This study aims to investigate whether public, environmental, and occupational health journals requested data-sharing statements from clinical trials submitted for publication.

Methods: In this bibliometric survey of "Public, Environmental, and Occupational Health" journals, defined by the Journal Citation Reports (as of June 2023), we included 202 journals with clinical trial reports published between 2019 and 2022. The primary outcome was a journal request for a data-sharing statement, as identified in the paper submission instructions. Multivariable logistic regression analysis was conducted to evaluate the relationship between journal characteristics and journal requests for data-sharing statements, with results presented as odds ratios (ORs) and corresponding 95% CIs. We also investigated whether the journals included a data-sharing statement in their published trial reports.

Results: Among the 202 public, environmental, and occupational health journals included, there were 68 (33.7%) journals that did not request data-sharing statements. Factors significantly associated with journal requests for data-sharing statements included open access status (OR 0.43, 95% CI 0.19-0.97), high journal impact factor (OR 2.31, 95% CI 1.15-4.78), endorsement of

XSL•FO
RenderX

Consolidated Standards of Reporting Trials (OR 2.43, 95% CI 1.25-4.79), and publication in the United Kingdom (OR 7.18, 95% CI 2.61-23.4). Among the 134 journals requesting data-sharing statements, 26.9% (36/134) did not have statements in their published trial reports.

Conclusions: Over one-third of the public, environmental, and occupational health journals did not request data-sharing statements in clinical trial reports. Among those journals that requested data-sharing statements in their submission guidance pages, more than one quarter published trial reports with no data-sharing statements. These results revealed an inadequate practice of requesting data-sharing statements by public, environmental, and occupational health journals, requiring more effort at the journal level to implement ICJME recommendations on data-sharing statements.

(J Med Internet Res 2025;27:e64069) doi:10.2196/64069

KEYWORDS

data sharing; clinical trial; public health; International Committee of Medical Journal Editors; ICMJE; journal request; clinical trials; decision-making; occupational health; health informatics; patient data

Introduction

Data sharing plays a crucial role in health informatics, contributing to improving health information systems, enhancing operational efficiency, informing policy and decision-making, and advancing public health surveillance including disease tracking [1]. Specifically, sharing individual patient data (IPD) from clinical trials promotes scientific advancement, enhances transparent reporting, and meets ethical obligations to trial participants [1,2]. In 2017, the International Committee of Medical Journal Editors (ICMJE) required that all papers submitted after July 1, 2018, must include a data-sharing statement [3].

Data sharing is of special importance for trials in the field of public, environmental, and occupational health, where findings can translate to policies with population-level impact. Given the fact that many trials incorporate public engagement and are commonly designed as community-based cluster trials, IPD sharing may improve public trust and support by enhancing transparent reporting and reproducibility and helping patients access their health data information and make informed health care decisions [4]. While public perception and willingness to share IPD are generally affected by concerns about privacy, transparency is a critical step toward getting buy-in from the public, particularly since environmental and public health policies may have implications on individual rights [4-6]. Specifically, addressing the ethical and regulatory issues surrounding the confidentiality of shared data is crucial for actual data sharing [7]. Nonetheless, enhancing transparency and providing detailed information about how data will be collected, used, and shared can effectively improve trust among individuals and facilitate actual data sharing. In the setting of public health emergencies such as epidemics, timely data sharing fosters societal-level preparedness, helps advance discovery in disease prevention and therapy, and allays fear of the unknown [8]. Of note, challenges and barriers to data sharing remained for public health emergencies, including poor accessibility, language difficulties, mistrust toward the health system, or concerns about insurance eligibility [4]. Global policymakers and researchers have advocated data sharing for the COVID-19 pandemic and developed publicly available platforms to enhance the sharing of IPD [9-11]. Nevertheless, a study based on trial registration information showed that the proportion of intent to share IPD for COVID-19 clinical trials did not differ from the contemporary non–COVID-19 trials [12].

Some previous studies explored journal requests for data-sharing statements in high-profile journals or specific clinical fields, with no focus on the field of public, environmental, and occupational health [13-16]. It is unclear whether public, environmental, and occupational health journals adhere to ICMJE's requirement for data-sharing statements. Clinical trials are the gold standard for causal inference, and their data sharing is particularly relevant as trials change policy and practice. Therefore, we conducted a survey to investigate whether public, environmental, and occupational health journals request data-sharing statements in clinical trial submissions. We also explored associations between journal characteristics and journal requests for data-sharing statements.

Methods

Study Design

We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline to report this study [17]. According to previous publications exploring the field of public, environmental, and occupational health, we included all 400 journals that were in the category of "Public, Environmental, and Occupational Health" defined by the Journal Citation Reports (as of June 2023) [18-20]. Subsequently, we searched the journal webpages using the keywords "trial" or "clinical trial" to assess whether the journal had published clinical trial reports between January 1, 2019, and December 31, 2022. Journals that did not publish clinical trial reports with IPD between 2019 and 2022 were excluded because the ICMJE required data-sharing statements in clinical trials after July 2018. A total of 202 journals were included for analysis (Figure S1 in Multimedia Appendix 1).

Study Outcome

The study outcome was a journal request for data-sharing statements. We searched journals' webpages about their data-sharing statement request by using the keywords "data sharing," "data availability," "research data," "data accessibility," "data deposition," and "data deposit." Google Translate was used for journals that were in non-English and non-Chinese languages. A journal was deemed to request

XSL•FO
RenderX

data-sharing statements if the journal clearly asked trial authors to provide a data-sharing statement on the paper submission instructions. Table S1 in Multimedia Appendix 1 shows the list of descriptions of data-sharing statement requests from journals' paper submission instructions.

Although some journals did not request statements on the paper submission instructions, it was possible that the journals would request trial authors to upload a statement or ask the authors about their willingness to share data in the journal submission system. These journals were also considered to have requests for data-sharing statements. A random sample of 10 journals that had no request on their submission guidance pages was used for the mockup submission process. Specifically, we evaluated whether these journals without requests from their submission guidance would ask for information on data-sharing statements from trial authors in their submission systems. None of these journals were found to request a data-sharing or availability statement in their submission systems.

The journals with data-sharing statement requests differed in how strongly they worded the request in the description. Specifically, if a journal used the terms "encourage" or "recommend" (eg, we encourage the use of data-sharing statements), we categorized the type of journal request for data-sharing statements as Weak. If a journal worded the terms "require," "request," "should," "mandate," or "must" (eg, all research articles must contain a data-sharing statement), we categorized the journal request as Strong (Table S1 in Multimedia Appendix 1). These terms used to classify the type of journal requests for data-sharing statements were mainly adopted from previous literature [16]. To minimize the potential for misclassification, we performed data extraction and categorized the type of journal request in duplicate (YL and JZ). Any disagreement was resolved by discussion or in consultation with a senior investigator (GL).

Data Extraction

Data extraction was performed by study investigators in duplicate and pairs (YL and JZ, XB, and LK) from September 1, 2023, to March 31, 2024. If there was a disagreement between these study investigators, we tried to reach a consensus by double-checking the extracted information and discussing it with the group. If a consensus could not be reached, we consulted with a senior investigator for a decision (GL). Data on journal characteristics were extracted, including the percentage of open access, its publication region, publisher, publication language, journal impact factor in 2022 (released in June 2023), Journal Citation Reports quartile, whether the journal was on the ICMJE list, whether the journal explicitly endorsed the Consolidated Standards of Reporting Trials (CONSORT) in its submission guidance page, and the total number of trials published between 2019 and 2022 for each journal.

The percentage of open access, as extracted from the open access section in the Journal Citation Reports, denoted the percentage of open access items among all the citable items published in the journal in the previous 3 years.

Both data on publishers and publication regions were collected from the Journal Citation Reports. Journals from the same publisher could belong to different publication regions, thereby potentially providing more information on social and cultural differences. For example, 3 journals (*Disability and Health Journal*, *Cancer Epidemiology*, and *International Journal of Hygiene and Environmental Health*) belonged to Elsevier, which has its headquarters in Amsterdam (Netherlands), while these journals were published in the United States, United Kingdom, and Germany, respectively. Therefore, we collected both the data on the publisher and the publication region of each journal for potential inclusion in analyses.

Statistical Analysis

Continuous characteristics were described using medians with lower and upper quartiles (Q1-Q3), and categorical variables using counts and percentages. Comparisons between journals with and without data-sharing statement requests were conducted by *t* test and chi-square test for continuous and categorical variables, respectively.

Multivariable logistic regression analysis was conducted to evaluate the relationship between journal characteristics and a journal's request for data-sharing statements. Some journal characteristics were considered as covariates in the model including publication region (United Kingdom and United States vs others), publisher, and Journal Citation Reports quartile (Q1-Q2 vs Q3-Q4), in which these characteristics had been reported to associate with journal request in previous studies [15,16]. In this survey, the United States and the United Kingdom had the top 2 numbers of public, environmental, and occupational health journals, far more than the remaining publication regions (Figure 1A). Therefore, we treated the variable of publication region as 3 levels (United States, United Kingdom, and others) and took the other regions as the reference group. Other characteristics were also adjusted for in the model based on our expertise and group discussion, including open access (yes vs no, taking 50% of open access as threshold), journal impact factor (≥ vs <2.9, taking the median of 2.9 as a threshold), whether the journal was on the ICMJE list (yes vs no), whether the journal endorsed the CONSORT (yes vs no), and the number of trials published between 2019 and 2022 (≥ vs <11, taking the median of 11 as a threshold). We assessed the potential multicollinearity in the model by calculating variance inflation factor (VIF) values for each journal characteristic, with a VIF >4 indicating severe multicollinearity [21]. Two variables, the Journal Citation Reports quartile and publisher, were removed because they had a VIF >4 (6.8 for Journal Citation Reports quartile and 4.3 for publisher), leaving all the other journal characteristics remaining in the final model. Results were presented as odds ratios (ORs) with corresponding 95% CIs. An OR >1 indicated that the variable was associated with increased odds of having requests for data-sharing statements.

XSL•FO
RenderX

Figure 1. Characteristics of journals according to data sharing statement request by publication region and publisher. Only publication regions and publishers with ≥5 journals were listed. Proportions were presented as lines. (A) The count and proportion of journals according to data sharing statement request by publication region. (B) The count and proportion of journals according to data sharing statement request by publisher.

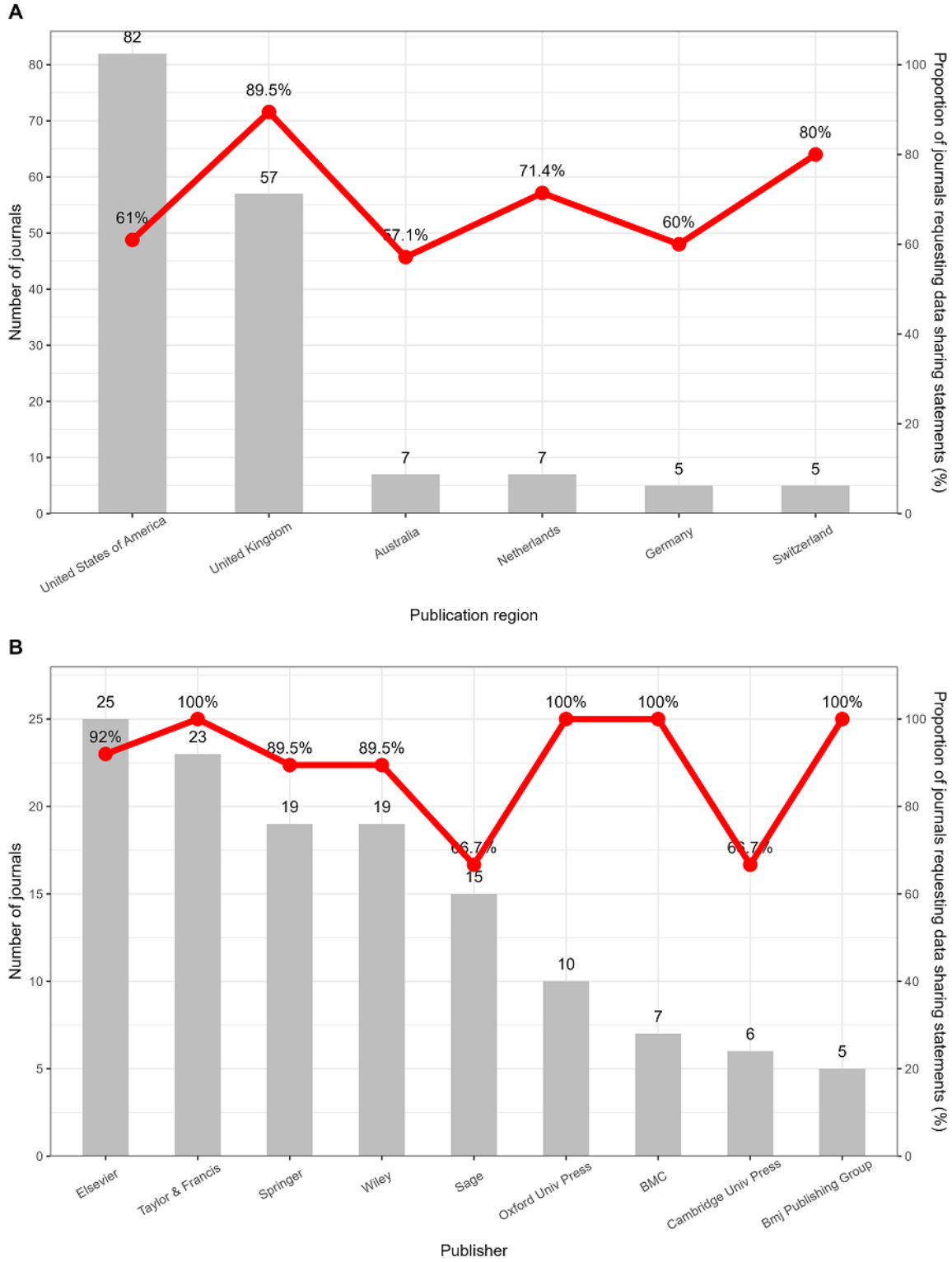

Two sensitivity analyses were conducted to evaluate the robustness of the main findings. First, journals from the same publisher may share a similar or same journal request for data-sharing statements. Given the removal of the publisher from our main analysis due to its VIF, to account for this cluster effect, we performed a sensitivity analysis by using the generalized estimating equations (GEE) approach with an exchangeable correlation structure to model the intraclass correlation of the journals within the same publisher [22]. GEE is specifically designed to handle correlated data from clustered or repeated measures. By using an exchangeable correlation structure, GEE allows for the assumption that the correlation between any 2 observations within the same cluster (eg, publisher) is constant. This could accurately reflect the

XSL•FO
RenderX

underlying data structure and improve the validity of the statistical inferences made [22]. Furthermore, we performed another sensitivity analysis by using the continuous forms for three journal characteristics (percentage of open access, journal impact factor, and number of trials published between 2019 and 2022) for the multivariable regression analysis. Using continuous forms in the model could make full use of the information provided by the continuous variables, which may therefore generate estimates with larger uncertainty when compared with the model using categorical forms [23]. Thus this approach could test the robustness of our main findings from multivariable regression analysis.

We further performed 3 exploratory analyses for this survey. First, we evaluated the relationship between journal characteristics and the type of journal request for data-sharing statements (weak or strong). Multinomial logistic regression analysis was used for this analysis, taking no request as the reference.

Second, to assess whether the journal request for data-sharing statements from the paper submission instructions was in line with their published clinical trial reports, we extracted information on data-sharing statements from 3 recent clinical trial reports published after September 1, 2023. We identified the recent trial reports by searching PubMed in descending order by publication date for each journal. We determined whether there was any data-sharing statement in these clinical trial reports by thoroughly searching the reports, the web pages where the journals published these reports, and their supplemental materials. Journals were therefore categorized into having any or no data-sharing statements in their published trial reports. We used the McNemar test to explore whether there was a significant discordance between journal requests (identified on the paper submission instructions) and the publication of a data-sharing statement (in their published trial reports). Furthermore, multivariable logistic regression was used to evaluate journal characteristics in relation to journals having data-sharing statements in trial reports, taking journals having none of the statements as a reference.

We conducted a third exploratory analysis trying to explore the temporal trend of journal requests for data-sharing statements. The study by Siebert et al [15] included the largest number of journals from the “Clinical Medicine” group defined by Journal Citation Reports in 2018 (n=489), among previous studies assessing journal requests for data-sharing statements. We matched our included journals with this previous publication to evaluate any potential difference in journal requests for data-sharing statements from 2018 to 2023 by using the McNemar mid-*P* test [24]. A significant difference would indicate there was a significant increase or decrease in journal requests for data-sharing statements over the past 5 years.

All statistical tests were 2-sided with a significance level of .05. Analyses were conducted in R software (version 4.1.0; R Core Development Team) and SAS software (version 9.4; SAS Institute).

Ethical Considerations

The present analysis was a secondary analysis based on published materials and website. In accordance with current regulations and the hospital's policy, studies that do not involve human participants or animals are exempt from the requirement for formal ethical approval.

Results

We included a total of 202 public, environmental, and occupational health journals for analysis (Figure S1 in Multimedia Appendix 1). As shown in Table 1, the included journals were mainly non–open access (71.8%) and published in the English language (96.0%). Overall, 18.8% of the journals were on the ICMJE list and 60.4% endorsed CONSORT. The United States (40.6%) and the United Kingdom (27.7%) had the top 2 numbers of journals among all publication regions, while Elsevier, Taylor & Francis, Springer, Wiley, and Sage were the top 5 publishers with the largest number of journals (Figure 1 and Figures S2 and S3 in Multimedia Appendix 1). The median journal impact factor was 2.9 (Q1-Q3: 1.9-4.5), and the median number of trials published between 2019 and 2022 was 11.0 (5.0-26.8).

XSL•FO
RenderX

Table 1. Journals' characteristics and comparisons according to data-sharing statement request. Results are shown as count (%) unless otherwise specified.

Journal characteristics	Overall (n=202)	Whether journal requested data-sharing statements		*P* value
		No (n=68)	Yes (n=134)	
Publication language				
English, n (%)	194 (96)	60 (88.2)	134 (100)	<.001
Non-English, n (%)	8 (4)	8 (11.8)	0 (0)	.98
Percentage of open access (%), median (Q1-Q3)	21.4 (9.4-76.7)	14.0 (0.2-90.4)	23.3 (11.7-48.2)	
Open access[b], n (%)				
No	145 (71.8)	44 (64.7)	101 (75.4)	.11
Yes	57 (28.2)	24 (35.3)	33 (24.6)	
Publication region, n (%)				
United States	82 (40.6)	32 (47.1)	50 (37.3)	<.001
United Kingdom	56 (27.7)	5 (7.4)	51 (38.1)	
Others	64 (31.7)	31 (45.6)	33 (24.6)	
Publisher, n (%)				
Elsevier	25 (12.4)	2 (2.9)	23 (17.2)	<.001
Taylor & Francis	23 (11.4)	0	23 (17.2)	
Springer	19 (9.4)	2 (2.9)	17 (12.7)	
Wiley	19 (9.4)	2 (2.9)	17 (12.7)	
Sage	15 (7.4)	5 (7.4)	10 (7.5)	
Others	101 (50)	57 (83.8)	44 (32.8)	
Journal impact factor, median (Q1-Q3)	2.9 (1.9-4.5)	2.1 (1.3-3.5)	3.3 (2.3-4.8)	.16
Journal impact factor ≥2.9[b]	99 (49.0)	24 (35.3)	75 (56.0)	.01
Journal Citation Reports quartile				
Q1-Q2, n (%)	92 (45.5)	23 (33.8)	69 (51.5)	.02
Q3-Q4, n (%)	110 (54.5)	45 (66.2)	65 (48.5)	
Number of trials published between 2019 and 2022, median (Q1-Q3)	11.0 (5.0-26.8)	8.00 (4.0-15.0)	14.0 (6.0-30.0)	.01
Number of trials published ≥11[c], n (%)	103 (51)	29 (42.6)	74 (55.2)	.09
Whether the journal was on the ICMJE[d] list, n (%)				
No	164 (81.2)	55 (80.9)	109 (81.3)	.94
Yes	38 (18.8)	13 (19.1)	25 (18.7)	
Whether the journal endorsed CONSORT[e], n (%)				
No	80 (39.6)	38 (55.9)	42 (31.3)	<.001
Yes	122 (60.4)	30 (44.1)	92 (68.7)	

[a]Not applicable.

[b]Open access journal was defined as having a percentage of open access ≥50%.

[c]The median journal impact factor was 2.9; the median number of trials published between 2019 and 2022 was 11.

[d]ICMJE: International Committee of Medical Journal Editors.

[e]CONSORT: Consolidated Standards of Reporting Trials.

XSL•FO
RenderX

There were 134 (66.3%) journals requesting data-sharing statements in clinical trials (Table 1). Journals requesting data-sharing statements had a higher journal impact factor and a larger proportion of being in the Journal Citation Report Q1-Q2. They were more likely to endorse CONSORT and publish more clinical trials compared with journals that did not request data-sharing statements. Significant differences were also found in publication regions and publishers between journals with and without request. As displayed in Figure 1A, the United Kingdom (89.5%) had the largest proportion of journals requesting data-sharing statements among all the publication regions.

Figure 2 demonstrates results for the relationship between journal characteristics and journal requests for data-sharing statements. Being open access was significantly associated with lower odds of journal requests for data-sharing statements (OR 0.43, 95% CI 0.19-0.97), while journal impact factor ≥2.9 and endorsement of CONSORT were significantly associated with increased odds (OR 2.31, 95% CI 1.15-4.78 and OR 2.43, 95% CI 1.25-4.79, respectively). Compared with other publication regions, UK journals had significantly higher odds of journal request (OR 7.18, 95% CI 2.61-23.40), while no significant relationship was observed in US journals (OR 0.95, 95% CI 0.43-2.08). Results from sensitivity analyses by using the continuous forms of 3 variables (percentage of open access, journal impact factor, and number of trials published between 2019 and 2022) and the GEE approach were largely in line with our main findings (Figures S4 and S5 in Multimedia Appendix 1).

Figure 2. Multivariable logistic regression assessing the association between journal characteristics and request for data-sharing statements. CONSORT: Consolidated Standards of Reporting Trials; ICMJE: International Committee of Medical Journal Editors.

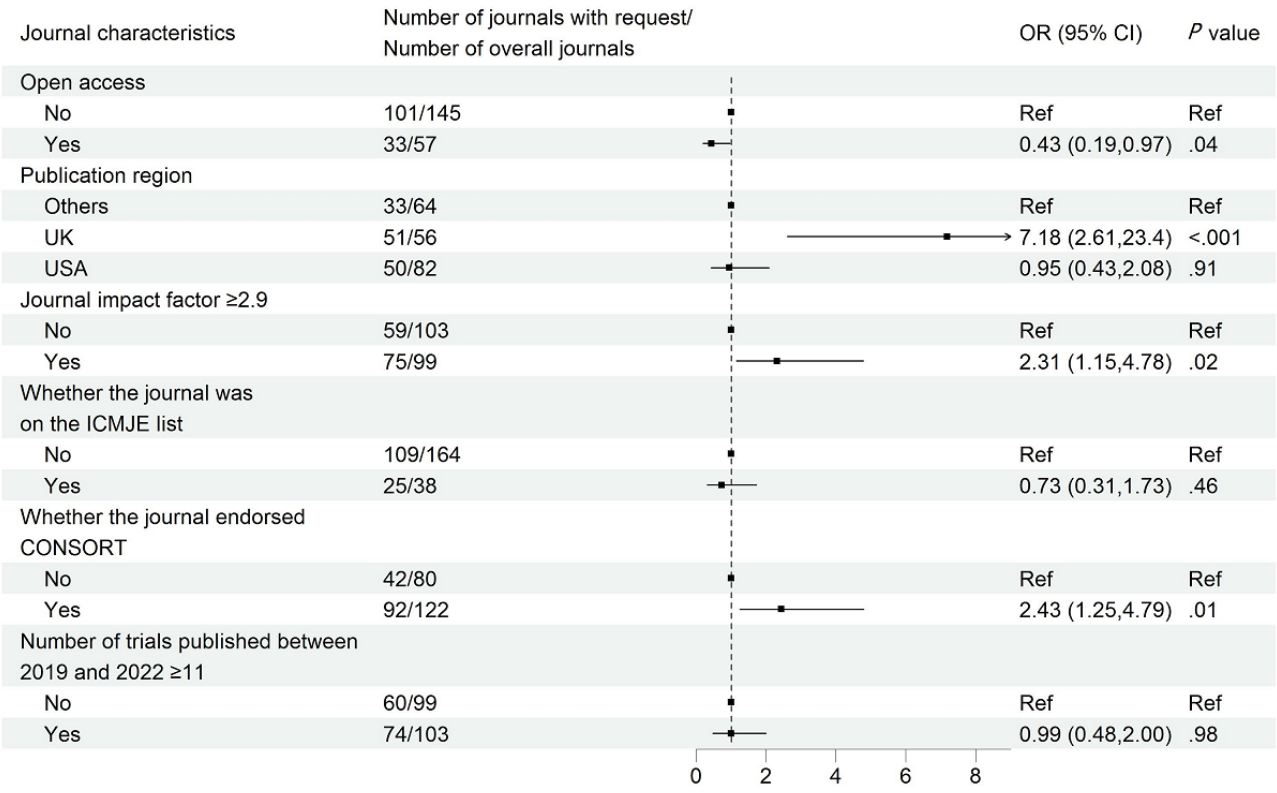

Among the 134 journals requesting data-sharing statements, there were 57 (42.5%) and 77 (57.5%) journals categorized as weak and strong, respectively. Figure 3 shows the results for journal characteristics in relation to the 2 types of journal requests. Both the comparisons between journals with weak requests and those with no requests, and between journals with strong requests and those with no requests, yielded similar results to our main findings.

XSL•FO
RenderX

Figure 3. Multinomial logistic regression assessing the association between journal characteristics and two types of journal requests for data-sharing statements. CONSORT: Consolidated Standards of Reporting Trials; ICMJE: International Committee of Medical Journal Editors.

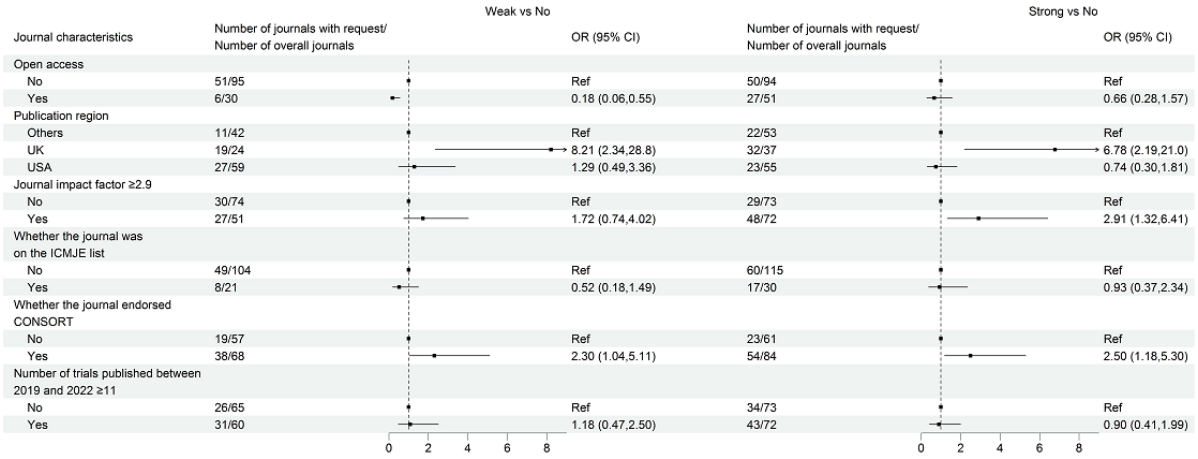

A significant discordance was observed between journal requests identified on the paper submission instructions and the publication of data-sharing statements identified from published clinical trial reports (P<.001, Table 2). Among the 134 journals requesting data-sharing statements, 36 (26.9%) journals did not have statements in the clinical trial reports, while there were 92.6% of the journals without request (63/68) that did not present statements in trial reports. The findings of the association between journal characteristics and journals with any data-sharing statement in their published clinical trial reports were largely consistent with the main results (Figure S6 in Multimedia Appendix 1).

Table 2. Data-sharing statements in published clinical trial reports versus journal requests for data-sharing statements identified on the paper submission instructions. Results are shown as count (%) unless otherwise specified.

	Whether the journal requested data-sharing statements on the paper submission instructions		Total
	No	Yes	
Whether there was any data-sharing statement in clinical trial reports published in the journal, n (%)			
None	63 (31.2)	36 (17.8)	99 (49)
Any	5 (2.5)	98 (48.5)	103 (51)
Total	68 (33.7)	134 (66.3)	202 (100)

There was a total of 7 journals included in both Siebert's study and our own (Table S2 in Multimedia Appendix 1). One journal (out of 7 journals, 14.3%), the *International Journal of Epidemiology*, that did not request data-sharing statements during Siebert's study, was found to have changed to requesting data-sharing statements at the time of our survey. However, the difference in journal requests was not statistically significant (P=.50).

Discussion

Principal Findings

In this survey, we explored the data-sharing statement request of public, environmental, and occupational health journals. The principal findings were as follows: (1) over one-third of journals did not request data-sharing statements in clinical trial reports; (2) being open access was significantly associated with decreased odds of journal request, while publication in the United Kingdom, higher journal impact factor, and endorsement of CONSORT were significantly associated with increased odds of journal requests; and (3) among the journals requesting data-sharing statements, approximately 27% had published trial reports that indeed had no data-sharing statements.

Despite the ICMJE recommendation, our study showed over one-third of the included journals did not request data-sharing statements on their web pages. This data-sharing statement request remained inadequate for public, environmental, and occupational health journals, even though the importance of sharing IPD in public health and health informatics had been prominently highlighted during the COVID-19 pandemic [25]. Ramdjee et al conducted a study to include 148 COVID-19 interventional phase 3 trials registered between September 1, 2020, and March 1, 2021, reporting that the intent to share IPD at registration did not differ between these COVID-19 trials and 296 contemporary non–COVID-19 trials [12]. Furthermore, their subgroup analysis showed that among vaccine trials, COVID-19 trials were significantly related to decreased odds of intent to share IPD when compared with non–COVID-19 trials (OR 0.08, 95% CI 0.01-0.46). Likewise, another study that included 36 COVID-19 vaccine trial reports found that 86% of trials reported data-sharing statements [26]. Even though all

XSL•FO
RenderX

36 trials were from 3 high-impact journals (*New England Journal of Medicine*, *The Lancet*, and *The Lancet Infectious Diseases*), representing the best-case scenario, this again indicated a suboptimal practice in journal requests for data-sharing statements. Vaccines are known as the most effective strategy to safeguard public health especially during a pandemic, while COVID-19 vaccine trials have received major concerns over their data integrity [27-30]. Therefore, more efforts are urgently needed to enhance journal requests for data-sharing statements and thus the actual IPD sharing in trials in the field of public, environmental, and occupational health.

In our exploratory analysis, we observed that 1 journal that did not request data-sharing statements in 2018 changed to requesting data-sharing statements in 2023 [15]. This may reflect a potential improvement toward journal requests in public, environmental, and occupational health journals temporally. Nonetheless, the small sample size may tend to yield spurious findings by chance that should thus be interpreted with caution.

Our study revealed that among the journals on the ICMJE list, more than one-third did not request data-sharing statements. Furthermore, being on the ICMJE list was associated with decreased odds of journal requests for data-sharing statements, although the relationship was not statistically significant (Figure 2). Despite the ICMJE acknowledging that "there may be some listed journals that do not follow all of the many recommendations and policies in the document," more actions would be needed to enhance the request and eventually promote the actual IPD sharing among the ICMJE-listing journals [31]. Open access journals were less likely to request data-sharing statements than subscription journals, in line with a previous study [32]. Some open access journals may be reluctant to have data-sharing requests, because this may further increase trial authors' burden besides costly publication fees [33]. Furthermore, studies suggested that open access journals might lack the resources to enforce data sharing [32,33], which may partly explain the decreased odds of requesting data-sharing statements in open access journals.

Although the United States and the United Kingdom had the greatest numbers of public, environmental, and occupational health journals, our study showed that UK-based journals were more likely to request data-sharing statements. Nevertheless, guidelines on data sharing have been established by the national agencies in both the United Kingdom and the United States [34,35]. The elevated odds of requesting data-sharing statements in UK journals may be due to a combination of regulatory frameworks (ie, UK Research and Innovation Guidelines and General Data Protection Regulation), encouraging attitudes toward open science, and more support from academic institutions. Another possible explanation was that a higher distrust of sharing data had been reported in the United States than in the United Kingdom [4]. Besides, journal endorsement of CONSORT and a higher journal impact factor were associated with increased odds of journal requests for data-sharing statements. Journals with a higher impact factor were found to generally have more stringent peer-review processes and higher-quality publications [36]. Similarly, trial reports published in journals with the requirement of adhering to the CONSORT checklist were typically of higher quality than those trial reports published in journals without the CONSORT requirement [37]. However, these results should be treated with caution, especially given the existence of potential residual confounding effects and unmeasured biases from a nonrandomized survey.

Notably, we found a significant discordance between journal requests on the paper submission instructions and journal practice of including a data-sharing statement in their published trial reports (Table 2). Again, this information emphasized a large room for improving the current practice of journal requests, even in those journals that declared to request trial authors to provide data-sharing statements on the paper submission instructions.

Some previous studies have explored journal requests for data-sharing statements; however, none of these studies focused on public, environmental, and occupational health journals [13-16]. For example, a recent study included top journals in each quartile from 178 categories to examine requests for data-sharing statements in journals of life, health, and physical sciences [16]. While we ran a post hoc analysis by matching their included journals with ours, there were only 5 public, environmental, and occupational health journals included in their study. Similarly, to the best of our knowledge, the survey by Siebert et al involved the largest number of biomedical journals among previous studies, yet it only included 7 public, environmental, and occupational health journals for analysis [15].

Therefore, our survey may provide comprehensive evidence of whether public, environmental, and occupational health journals requested data-sharing statements in clinical trial submissions, thereby potentially generating some new insights into enhancing reporting transparency and eventually improving the actual sharing of IPD in clinical trials. The risk factors identified in this study might be targeted to improve the deficiency of current data-sharing statement request practice. Effective communication between publishers, journals, authors, academic institutions, and funders is encouraged to reach a potential consensus on data-sharing statement requests. Some perspectives on improving data-sharing statement requests may be considered. Publishers and journals may consider applying strict enforcement of their data-sharing statement request and actively monitoring compliance with it. Assigning specific human resources for evaluating data-sharing statements and actual data-sharing could be another possible option. Funders may provide user-friendly data-sharing platforms or tools and specific strategies to encourage data sharing. Editors and peer reviewers need to be tasked with evaluating the adequacy of data-sharing statements provided in submitted trial papers. The societies and associations in the field of public, environmental, and occupational health may consider establishing resource-sharing alliances for scholars and prioritizing publishing incentive policies, aiming to promote a culture of collaboration and openness in human research. Researchers are encouraged to adopt a data management plan at the beginning of a trial and closely adhere to their plans. Overall, joint endeavors from multiple stakeholders would be required to improve the journal request for data-sharing statements and to address the important gap between declared data-sharing statement requests in

XSL•FO
RenderX

journals' webpages and the actual statements provided in their published trial reports, ultimately promoting advancement, information transparency, and reproducibility in public, environmental, and occupational health research.

Strengths and Limitations

This survey was the first to systematically explore whether public, environmental, and occupational health journals requested data-sharing statements in their clinical trial report submissions. Rigorous methodology and detailed analyses strengthened our findings.

Several limitations need to be noted. First, no causal effects of journal characteristics on journal requests for data-sharing statements could be obtained in this observational study. Similarly, potential biases and confounding could not be fully precluded. As Google Translate was used in our study, potential errors due to misinterpretation of non-English or Chinese-language journals (n=8) could not be addressed. In addition, some journals involving multiple fields of health research (eg, *The New England Journal of Medicine* and *The Lancet*) that were not in the category of "Public, Environmental, and Occupational Health" indeed frequently published trials in the field of public, environmental, and occupational health. Even though we aimed to target public, environmental, and occupational health journals by using the definition of Journal Citation Reports for inclusion, this may fail to include a comprehensive list of public, environmental, and occupational health-related journals for analysis. Moreover, we did not contact journal editors or publishers who may determine or at least get involved in making the request for data-sharing statements to further verify our study outcome, mainly due to the low feasibility. Nevertheless, the results of our mockup submission in a random sample of journals without request identified from their submission guidance pages could partly support the accuracy of our outcome measures. Notably, these findings could only reflect the practice of journal requests for data-sharing statements during the time of our study. Given the dynamic change in policy endorsement, future explorations are needed to investigate the latest practice of journal requests for data-sharing statements.

Conclusions

Over one-third of the public, environmental, and occupational health journals did not request data-sharing statements in clinical trial reports. Among those journals that requested data-sharing statements in their submission guidance pages, more than one quarter published trial reports with no data-sharing statements. These results revealed an inadequate practice of requesting data-sharing statements by public, environmental, and occupational health journals, requiring more efforts at the journal level to implement ICJME recommendations on data-sharing statements.

Acknowledgments

We acknowledge Mr Jianfeng Li, Miss Yaoyao Wang, Mr Xibin Sun, and Mr Runquan Zhang for their help with data collection. This study was funded by the Science Foundation of Guangdong Second Provincial General Hospital (YY2018-002), the Young Top Talent Project in Special Support Plan for Training High-level Talents in Guangdong (0720240244), and the National Natural Science Foundation of China (82473612).

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YL contributed to conceptualization, data curation, formal analysis, investigation, methodology, and writing–original draft. JZ assisted with conceptualization, data curation, formal analysis, methodology, writing–review & editing. LT performed conceptualization, methodology, supervision, writing–review & editing. XB contributed to data curation, writing–review & editing. LK performed data curation, writing–review & editing. GYHL assisted handled supervision, writing–review & editing. HGCVS contributed to supervision, writing–review & editing. MX handled conceptualization, investigation, methodology, supervision, writing–original draft, and writing–review & editing. GL performed conceptualization, investigation, methodology, supervision, writing–original draft, and writing–review & editing. YL and JZ contributed equally to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.
[DOCX File , 4427 KB - jmir_v27i1e64069_app1.docx]

References

1. National Academy of Medicine, The Learning Health System Series. Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust. Washington (DC): National Academies Press (US); 2021.

XSL•FO
RenderX

2. Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. JAMA 2016;315(12):1237-1239. [doi: 10.1001/jama.2016.2420] [Medline: 27002444]
3. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. JAMA 2017 Jun 27;317(24):2491-2492. [doi: 10.1001/jama.2017.6514] [Medline: 28586895]
4. Ghafur S, Van Dael J, Leis M, Darzi A, Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. Lancet Digit Health 2020 Sep;2(9):e444-e446 [FREE Full text] [doi: 10.1016/S2589-7500(20)30161-8] [Medline: 32838250]
5. Middleton A, Milne R, Almarri MA, Anwer S, Atutornu J, Baranova EE, et al. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data? Am J Hum Genet 2020;107(4):743-752 [FREE Full text] [doi: 10.1016/j.ajhg.2020.08.023] [Medline: 32946764]
6. Garett R, Young SD. Ethical views on sharing digital data for public health surveillance: analysis of survey data among patients. Front Big Data 2022;5:871236 [FREE Full text] [doi: 10.3389/fdata.2022.871236] [Medline: 35547191]
7. Ugochukwu AI, Phillips PWB. Open data ownership and sharing: challenges and opportunities for application of FAIR principles and a checklist for data managers. Journal of Agriculture and Food Research 2024;16:101157 [FREE Full text] [doi: 10.1016/j.jafr.2024.101157]
8. Littler K, Boon WM, Carson G, Depoortere E, Mathewson S, Mietchen D, et al. Progress in promoting data sharing in public health emergencies. Bull World Health Organ 2017;95(4):243. [doi: 10.2471/blt.17.192096] [Medline: 28479616]
9. Xu B, Kraemer MUG, Open COVID-19 Data Curation Group. Open access epidemiological data from the COVID-19 outbreak. Lancet Infect Dis 2020;20(5):534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30119-5] [Medline: 32087115]
10. Moorthy V, Henao Restrepo AM, Preziosi M, Swaminathan S. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ 2020;98(3):150 [FREE Full text] [doi: 10.2471/BLT.20.251561] [Medline: 32132744]
11. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Mar;579(7798):265-269 [FREE Full text] [doi: 10.1038/s41586-020-2008-3] [Medline: 32015508]
12. Ramdjee B, Husson M, Hajage D, Tubach F, Estellat C, Dechartres A. COVID-19 trials were not more likely to report intent to share individual data than non-COVID-19 trials in ClinicalTrials.gov. J Clin Epidemiol 2023;158:10-17 [FREE Full text] [doi: 10.1016/j.jclinepi.2023.03.015] [Medline: 36965602]
13. Danchev V, Min Y, Borghi J, Baiocchi M, Ioannidis JPA. Evaluation of data sharing after implementation of the International Committee of Medical Journal Editors data sharing statement requirement. JAMA Netw Open 2021;4(1):e2033972 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.33972] [Medline: 33507256]
14. Hamilton DG, Page MJ, Finch S, Everitt S, Fidler F. How often do cancer researchers make their data and code available and what factors are associated with sharing? BMC Med 2022;20(1):438 [FREE Full text] [doi: 10.1186/s12916-022-02644-2] [Medline: 36352426]
15. Siebert M, Gaba JF, Caquelin L, Gouraud H, Dupuy A, Moher D, et al. Data-sharing recommendations in biomedical journals and randomised controlled trials: an audit of journals following the ICMJE recommendations. BMJ Open 2020;10(5):e038887 [FREE Full text] [doi: 10.1136/bmjopen-2020-038887] [Medline: 32474433]
16. Kim J, Kim S, Cho HM, Chang JH, Kim SY. Data sharing policies of journals in life, health, and physical sciences indexed in journal citation reports. PeerJ 2020;8:e9924 [FREE Full text] [doi: 10.7717/peerj.9924] [Medline: 33083109]
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
18. Sweileh WM, Zyoud SH, Al-Jabi SW, Sawalha AF. Public, environmental, and occupational health research activity in Arab countries: bibliometric, citation, and collaboration analysis. Arch Public Health 2015;73(1):1 [FREE Full text] [doi: 10.1186/2049-3258-73-1] [Medline: 25671116]
19. Resnik DB, Konecny B, Kissling GE. Conflict of interest and funding disclosure policies of environmental, occupational, and public health journals. J Occup Environ Med 2017;59(1):28-33 [FREE Full text] [doi: 10.1097/JOM.0000000000000910] [Medline: 28045794]
20. Soteriades ES, Falagas ME. A bibliometric analysis in the fields of preventive medicine, occupational and environmental medicine, epidemiology, and public health. BMC Public Health 2006;6(1):301 [FREE Full text] [doi: 10.1186/1471-2458-6-301] [Medline: 17173665]
21. Lavery MR, Acharya P, Sivo SA, Xu L. Number of predictors and multicollinearity: what are their effects on error and bias in regression. Communications in Statistics - Simulation and Computation 2017;48(1):27-38. [doi: 10.1080/03610918.2017.1371750]
22. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Soft 2005;15(2):1-11. [doi: 10.18637/jss.v015.i02]
23. Barrio I, Arostegui I, Rodríguez-Álvarez MX, Quintana JM. A new approach to categorising continuous variables in prediction models: proposal and validation. Stat Methods Med Res 2017;26(6):2586-2602. [doi: 10.1177/0962280215601873] [Medline: 26384514]

XSL•FO
RenderX

24. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC Med Res Methodol 2013;13:91 [FREE Full text] [doi: 10.1186/1471-2288-13-91] [Medline: 23848987]
25. Maxwell L, Shreedhar P, Dauga D, McQuilton P, Terry RF, Denisiuk A, et al. FAIR, ethical, and coordinated data sharing for COVID-19 response: a scoping review and cross-sectional survey of COVID-19 data sharing platforms and registries. Lancet Digit Health 2023 Oct;5(10):e712-e736 [FREE Full text] [doi: 10.1016/S2589-7500(23)00129-2] [Medline: 37775189]
26. Duan Y, Luo J, Zhao L, Zhang X, Miao J, Moher D, et al. Reporting and data sharing level for COVID-19 vaccine trials: a cross-sectional study. EBioMedicine 2022;78:103962 [FREE Full text] [doi: 10.1016/j.ebiom.2022.103962] [Medline: 35339894]
27. Romero JR, Bernstein HH. COVID-19 vaccines: a primer for clinicians. Pediatr Ann 2020;49(12):e532-e536. [doi: 10.3928/19382359-20201116-01] [Medline: 33290571]
28. Kaur RJ, Dutta S, Bhardwaj P, Charan J, Dhingra S, Mitra P, et al. Adverse events reported from COVID-19 vaccine trials: a systematic review. Indian J Clin Biochem 2021;36(4):427-439 [FREE Full text] [doi: 10.1007/s12291-021-00968-z] [Medline: 33814753]
29. Sathian B, Asim M, Banerjee I, Roy B, Pizarro AB, Mancha MA, et al. Development and implementation of a potential coronavirus disease 2019 (COVID-19) vaccine: a systematic review and meta-analysis of vaccine clinical trials. Nepal J Epidemiol 2021;11(1):959-982 [FREE Full text] [doi: 10.3126/nje.v11i1.36163] [Medline: 33868742]
30. Doshi P, Godlee F, Abbasi K. Covid-19 vaccines and treatments: we must have raw data, now. BMJ 2022;376:o102. [doi: 10.1136/bmj.o102] [Medline: 35045956]
31. Journals stating that they follow the ICMJE recommendations. International Committee of Medical Journal Editors (ICMJE). URL: https://www.icmje.org/journals-following-the-icmje-recommendations/ [accessed 2020-01-09]
32. Vasilevsky NA, Minnier J, Haendel MA, Champieux RE. Reproducible and reusable research: are journal data sharing policies meeting the mark? PeerJ 2017;5:e3208 [FREE Full text] [doi: 10.7717/peerj.3208] [Medline: 28462024]
33. Booth CM, Ross JS, Detsky AS. The changing medical publishing industry: economics, expansion, and equity. J Gen Intern Med 2023;38(14):3242-3246. [doi: 10.1007/s11606-023-08307-z] [Medline: 37438644]
34. Barbui C, Gureje O, Puschner B, Patten S, Thornicroft G. Implementing a data sharing culture. Epidemiol Psychiatr Sci 2016 Aug;25(4):289-290 [FREE Full text] [doi: 10.1017/S2045796016000330] [Medline: 27150624]
35. Guidance on best practice in the management of research data. UK Research and Innovation. URL: https://www.ukri.org/publications/guidance-on-best-practice-in-the-management-of-research-data/ [accessed 2018-06-19]
36. Jalloh MB, Bot VA, Borjaille CZ, Thabane L, Li G, Butler J, et al. Reporting quality of heart failure randomized controlled trials 2000-2020: temporal trends in adherence to CONSORT criteria. Eur J Heart Fail 2024;26(6):1369-1380. [doi: 10.1002/ejhf.3229] [Medline: 38623814]
37. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal 'instructions to authors' in 2014. Trials 2016;17(1):301 [FREE Full text] [doi: 10.1186/s13063-016-1408-z] [Medline: 27343072]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GEE: generalized estimating equations
ICMJE: International Committee of Medical Journal Editors
IPD: individual patient data
OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VIF: variance inflation factor

Edited by Y Zhuang; submitted 08.07.24; peer-reviewed by W Tulvatana, I Mircheva, S Bolore, F Medina; comments to author 29.10.24; revised version received 12.12.24; accepted 12.12.24; published 07.02.25.

Please cite as:
Liu Y, Zhang J, Thabane L, Bai X, Kang L, Lip GYH, Van Spall HGC, Xia M, Li G
Data-Sharing Statements Requested from Clinical Trials by Public, Environmental, and Occupational Health Journals: Cross-Sectional Study
J Med Internet Res 2025;27:e64069
URL: https://www.jmir.org/2025/1/e64069
doi:10.2196/64069
PMID:

XSL•FO
RenderX

©Yingxin Liu, Jingyi Zhang, Lehana Thabane, Xuerui Bai, Lili Kang, Gregory Y H Lip, Harriette G C Van Spall, Min Xia, Guowei Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predicting the Risk of HIV Infection and Sexually Transmitted Diseases Among Men Who Have Sex With Men: Cross-Sectional Study Using Multiple Machine Learning Approaches

Bing Lin[1,2], MD; Jiaxiu Liu[3], MD; Kangjie Li[1,2], MD; Xiaoni Zhong[1,2], MM

[1]School of Public Health, Chongqing Medical University, Chongqing, China

[2]Research Center for Medicine and Social Development, Chongqing Medical University, Chongqing, China

[3]School of Medical Informatics, Chongqing Medical University, Chongqing, China

Corresponding Author:
Xiaoni Zhong, MM
School of Public Health
Chongqing Medical University
No.1 Medical College Road
Yuzhong District
Chongqing, 400016
China
Phone: 86 13527545050
Email: zhongxiaoni@cqmu.edu.cn

Abstract

Background: Men who have sex with men (MSM) are at high risk for HIV infection and sexually transmitted diseases (STDs). However, there is a lack of accurate and convenient tools to assess this risk.

Objective: This study aimed to develop machine learning models and tools to predict and assess the risk of HIV infection and STDs among MSM.

Methods: We conducted a cross-sectional study that collected individual characteristics of 1999 MSM with negative or unknown HIV serostatus in Western China from 2013 to 2023. MSM self-reported their STD history and were tested for HIV. We compared the accuracy of 6 machine learning methods in predicting the risk of HIV infection and STDs using 7 parameters for a comprehensive assessment, ranking the methods according to their performance in each parameter. We selected data from the Sichuan MSM for external validation.

Results: Of the 1999 MSM, 72 (3.6%) tested positive for HIV and 146 (7.3%) self-reported a history of previous STD infection. After taking the results of the intersection of the 3 feature screening methods, a total of 7 and 5 predictors were screened for predicting HIV infection and STDs, respectively, and multiple machine learning prediction models were constructed. Extreme gradient boost models performed optimally in predicting the risk of HIV infection and STDs, with area under the curve values of 0.777 (95% CI 0.639-0.915) and 0.637 (95% CI 0.541-0.732), respectively, demonstrating stable performance in both internal and external validation. The highest combined predictive performance scores of HIV and STD models were 33 and 39, respectively. Interpretability analysis showed that nonadherence to condom use, low HIV knowledge, multiple male partners, and internet dating were risk factors for HIV infection. Low degree of education, internet dating, and multiple male and female partners were risk factors for STDs. The risk stratification analysis showed that the optimal model effectively distinguished between high- and low-risk MSM. MSM were classified into HIV (predicted risk score <0.506 and ≥0.506) and STD (predicted risk score <0.479 and ≥0.479) risk groups. In total, 22.8% (114/500) were in the HIV high-risk group, and 43% (215/500) were in the STD high-risk group. HIV infection and STDs were significantly higher in the high-risk groups (P<.001 and P=.05, respectively), with higher predicted probabilities (P<.001 for both). The prediction results of the optimal model were displayed in web applications for probability estimation and interactive computation.

Conclusions: Machine learning methods have demonstrated strengths in predicting the risk of HIV infection and STDs among MSM. Risk stratification models and web applications can facilitate clinicians in accurately assessing the risk of infection in individuals with high risk, especially MSM with concealed behaviors, and help them to self-monitor their risk for targeted, timely diagnosis and interventions to reduce new infections.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e59101) doi:10.2196/59101

KEYWORDS

HIV; sexually transmitted diseases; men who have sex with men; machine learning; web application; risk stratification

Introduction

HIV Infection and Sexually Transmitted Disease Situation of Chinese Men Who Have Sex With Men

Men who have sex with men (MSM) are at high risk of contracting HIV infection and other sexually transmitted diseases (STDs), and the number of infections is growing fast. The latest report released by the Joint United Nations Program for HIV and AIDS 2023 stated that HIV prevalence was 11 times higher among MSM compared to adults in the general population [1]. According to the results of the current large-scale systematic analysis study in China, the national HIV prevalence rate in the MSM population from 2001 to 2018 was 5.7%, compared to 0.037% in the general population in 2014 [2,3]. At the same time, the burden of other STDs (eg, syphilis, gonorrhea, and condyloma acuminata) among MSM was also high and increasing [4]. The prevalence of syphilis in the MSM population was as high as 35%, comprising both known and newly diagnosed cases [5]. The prevalence of gonorrhea and chlamydia among MSM in Southern China has been reported to be as high as 12.5% and 18.2%, respectively [6,7]. In addition, these STDs have the potential to facilitate HIV transmission and infection [8]. Therefore, given the severe HIV infection and STD situation of Chinese MSM, improving risk perception strategies and developing accurate risk prediction tools can help to prevent and control these infections, thereby improving their quality of life.

Using Machine Learning for Accurate Risk Assessment

Establishing accurate risk assessment and prediction tools for the MSM population is beneficial for clinical services to identify high-risk individuals and prioritize resources. Currently, existing models, such as multivariate logistic regression, Poisson regression, and Cox proportional hazards regression, have been widely used in the context of HIV infection or other STDs [9-11]. However, data structures that were often nonlinear, anomalous, and heterogeneous could lead to poor prediction performance. Despite the existence of traditional statistical methods, research exploring more sophisticated analytical approaches to enhance disease classification and predictive capabilities among the MSM population has been limited. Machine learning provides an innovative perspective on evaluating the risk of infections in population with high risk, allowing more flexibility in handling complex, nonlinear relationships between the variables, thus minimizing error [12]. Furthermore, machine learning is composed of a set of powerful algorithms that have the ability to acquire knowledge, predict outcomes, and review data. The use of these algorithms has attracted great attention in the development of patient-centered predictive models, which can provide valuable clinical insights [13]. However, currently, there are few studies using machine learning methods to predict the risk of HIV infection and STDs in China, and our study extended the literature in this context.

Risk Prediction for Serving Clinical Applications

The ultimate goal of risk prediction is intended to serve clinical applications prioritizing the screening of high-risk individuals in need of intervention. However, if too many predictors were included in the prediction model, it would not be conducive to collecting and applying them in clinical situations. In addition, there were various feature selection methods, each offering unique advantages but also having inevitable drawbacks. If multiple methods were used, the likelihood of screening for excellent predictors could be increased. Therefore, when faced with a large number of variables or when trying to construct a predictive model with a small number of predictors, a *combined* feature screening method was necessary. This method could help to decrease the data dimensions, reduce the amount of computation, and increase the applicability while eliminating some irrelevant or redundant variables and improving the prediction effect [14]. In addition, due to the “black box” of machine learning algorithms, understanding why a specific prediction was made for an individual or how a particular characteristic of an individual influenced a prediction was challenging [15]. The lack of interpretability limited the widespread use of more robust machine learning methods to support medical decision-making [16]. The limited intuitive understanding of machine learning models was still a significant barrier to their implementation in health care due to their inability to reflect the true correlation between individual characteristics and predicted outcomes [17]. As a result, the ability to correctly interpret predictive model outputs is important for generating appropriate user trust, providing insights to improve the model, and supporting an understanding of the modeling process to leverage the potential and productivity benefits of machine learning approaches. Moreover, some studies have failed to effectively address the data imbalance issues in their datasets, and further efforts to address this imbalance may improve model performance [18].

Developing Risk Screening Tools

To address the limitations of previous research, our study comprehensively considered feature selection, model interpretability, and strategies for handling class imbalance in the data. First, we applied a “combined” feature selection process using 3 methods, including univariate analysis, least absolute shrinkage and selection operator (LASSO), and Boruta, to select features and incorporate variables common to each method. At the same time, a series of data processing techniques and machine learning methods were used to identify individuals at risk for HIV infection and STDs in the MSM population. The risk factors for infection were analyzed, and interpretations were supported by the use of partial dependence plots (PDPs), which were designed to provide theoretical support for targeted screening strategies and risk management. Our study had a more comprehensive modeling strategy than previous studies. Moreover, our study did not incorporate difficult-to-obtain clinical information, such as blood-related, immunologic, or

XSL•FO
RenderX

cytologic data. All predictors were relatively easy to assess, making them more suitable for widespread use in screening studies among the MSM population. Finally, regarding clinical applications, the deployment of machine learning predictions into web applications showed good potential and application value. The user-friendly web applications served as pioneering digital tools that have been widely used in managing diabetic blood glucose, analyzing gliomas, and predicting early mortality in patients with bone metastases; however, in MSM populations, the risk of HIV infection and STDs has had limited application [19-21]. At the same time, the MSM population may conceal their sexual behavior due to external pressures from family members, traditional social attitudes, and fear of discrimination and stigmatization [22]. Web applications can help MSM conduct real-time, interactive self-monitoring anytime, anywhere, serving the sexual minority individuals and helping to protect their privacy. These web applications are a simple, practical, and effective risk management tool that can be widely used by clinicians and MSM individuals, and web applications offer advantages such as time efficiency, user-friendliness, and privacy protection. In summary, our study aimed to use machine learning methods to develop accessible, user-friendly web applications and tools to provide valuable insights into the risk factors influencing HIV infection and STDs, as well as to tailor screening models for the MSM population to accurately predict the risk of HIV infection and STDs in this population.

Methods

Study Design and Participants

Our study was a multistage, multicenter cross-sectional survey conducted from 2013 to 2023, based on the key projects of infectious disease research funded by the Chinese Ministry of Science and Technology during the 12th Five-Year Plan and 13th Five-Year Plan periods. Because of the particularity of the MSM population as sexual minority individuals, we used a nonprobability sampling method to recruit MSM in Western China (ie, Chongqing, Xinjiang, Guangxi, and Sichuan) through collaboration with nongovernmental organizations.

Specifically, we advertised on websites, QQ (Shenzen Tencent Computer System), and WeChat (Tencent Holdings Ltd) groups, and we collaborated with local nongovernment organizations to provide information on HIV prevention, counseling, and testing for lesbians, gay men, bisexual, and transgender individuals on each platform. We used this approach to present the details of the study (ie, objectives, procedures, potential benefits, and risks) to the leaders to gain their support and, with their help, recruit participants from the organizations. MSM were included in our study to complete the questionnaire if they were aged ≥18 years, had sexual contact with a man in the past 12 months, and had negative or unknown HIV serostatus. Participants who met the requirements went to local hospitals or community centers with which we had collaborated, where our researchers conducted face-to-face questionnaires and performed the appropriate tests. If participants encountered any unclear terms, trained investigators explained their meaning to them. All questionnaires were subsequently checked for completeness and consistency.

Outcomes

We tested MSM for HIV to determine whether they had experienced HIV infection. Those who were previously infected with STD among MSM were measured by the question, “In the last six months, have you been diagnosed by a doctor with the STD (for example, syphilis, gonorrhea, condyloma acuminatum, genital herpes, non-specific urethritis, and so on)?” On the basis of HIV test results, patients who obtained positive results were labeled as positive outcome indicators, and those who obtained nonpositive results were labeled as negative outcome indicators. Those who were previously infected with an STD were labeled as positive outcome indicators. and those who were never infected with an STD were labeled as negative outcome indicators. Predictive models were developed based on each of the 2 outcome indicators, HIV and STD.

Predictors

A total of 24 variables were selected for the personal characteristics of MSM, including four main aspects: (1) demographic characteristics; (2) HIV-related knowledge, behavior, and risk perception; (3) sexual behavior–related characteristics; and (4) substance use. Among them, demographic characteristics included age, household registration location, ethnicity, degree of education, employment situation, marital status, and monthly disposable income. HIV-related knowledge, behaviors, and risk perception included HIV knowledge scores (≥11 was defined as “high level” [22]), HIV testing and counseling, and HIV risk perceptions. Sexual behavior–related characteristics included sexual role (ie, the insertive partner and the receptive partner), number of sexual partners, high-risk sex, and condom use. Substance use included internet dating (ie, finding sexual partners through the internet), alcohol use, commercial sexual services (eg, offering and receiving sexual services), and recreational drug use. We excluded samples with too much missing data or where the presence of HIV infection or STD could not be determined. After a preliminary analysis, all variables had a missing rate of <10%, so we used the multiple imputation method to fill in the missing values.

Feature Selection

Redundant information in survey data from the MSM population may lead to suboptimal categorization of HIV infection or STD, as excessive dimensions in the data can reduce the accuracy and validity of the models [15]. Therefore, it is promising to use the “combined” feature selection method to select variables for the raw data. Univariate analysis, LASSO, and Boruta were used in our study. In this study, variables with $P<.10$ were included in the univariate analysis; variables with regression coefficients not equal to 0 were included in LASSO; and variables that were classified as confirmed or tentative were included, while rejected variables were excluded in Boruta. Using 3 different methods for feature selection, we ultimately included factors common to all these methods as predictors in the predictive models.

Prediction Models

Our dataset showed a significant imbalance in the number of individuals with and without HIV and STDs. Data-level

XSL•FO
RenderX

resampling techniques have been increasingly used to address unbalanced datasets due to its simplicity and ease of implementation. Therefore, our study used the synthetic minority oversampling technique to deal with the unbalanced dataset. This method oversampled the minority class samples in the boundary region, and it expanded the decision space of the minority class samples with little effect on the decision space of the majority class samples.

Logistic regression, decision tree, random forest, K-nearest neighbor, light gradient boosting machine, and extreme gradient boosting (XGBoost) were used to predict the risk of HIV infection or STDs in the MSM population. To train, construct, and evaluate the 6 predictive models, we divided 75% (1499/1999) of the data into a training set, 25% (500/1999) into a testing set, and used the data from the Sichuan MSM as an external validation set. Using the training data, a cross-validation grid search was conducted to determine the optimal hyperparameters for the predictive model. In this study, we used 7 standard performance metrics, including accuracy, balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve, to evaluate the classification performance of the prediction model. Meanwhile, 10-fold cross-validation was used to assess the stability of several machine learning prediction models. These evaluation metrics were calculated from a binary confusion matrix, and based on the results of previous studies, these metrics were ranked according to their predictive performance [21,23]. On the basis of each evaluation metric, the highest score was assigned to the best-performing model, followed by the next best-performing model. The total score of the model was the sum of the scores for each metric. The data were heat mapped using the *pheatmap* package in R (R Foundation for Statistical Computing).

Model Interpretation

Interpretable analyses of the model included both variable importance ranking and PDP to identify key predictors of HIV infection or STDs in the MSM population. Feature importance refers to quantifying the contribution of each variable in the predictive model to the prediction outcome and ranking them accordingly, which reflects the magnitude of each predictor's influence on the model. Our study used PDP to conduct an explanatory analysis of the predictive model, assessing the degree and direction of influence of each variable on the prediction outcome.

Risk Stratification

The probability of the optimal model output was defined as the predicted risk score (PRS) [24]. The cutoff point could be used as a basis for risk definition, and based on this threshold, MSM were categorized into low-risk and high-risk individuals. MSM with a predicted probability less than the threshold were assigned to the low-risk group, while MSM with a predicted probability greater than or equal to the threshold were assigned to the high-risk group. This approach allowed differentiation of MSM based on predicted risk, facilitating targeted interventions and individualized management strategies.

Web Applications

Using the algorithm of the optimal model in this study, we developed web-based calculators to predict the risk of HIV infection or STDs in the MSM population. The applications were composed of 3 main parts, which were the input box, the prediction result, and the result box. The input box included the predictors of the prediction model, and the user could input the values of the predictors independently according to the individual situation. This interactive panel allowed the user to customize the prediction based on specific individual characteristics. Clicking on the Predict button below the input box displayed the predicted outcome under different conditions, which was the probability of an individual MSM being infected with HIV or STD. This information could provide valuable insights for clinical decision-making, and web applications can predict outcomes under different conditions and save these multiple predictions in the results box.

Sensitivity Analysis

According to the sexual behavior–related characteristics, MSM could be categorized into MSM and women (MSMW), MSM only (MSMO) [25], men who have sex with multiple men (MSMM), and non-MSMM [26]. Previous studies have found differences between MSMW and MSMO in HIV-related high-risk sexual behavior and HIV infection or STD prevalence [27]. At the same time, MSMM had a higher risk of infection than other men [28]. For these subgroups with different risks of infection, we included MSMW and MSMO as well as MSMM and non-MSMM into the optimal model to predict the risk of HIV infection or STDs in these subgroups. This was done to perform sensitivity analyses and validate the stability of the prediction model.

Statistical Analysis

For all statistical analyses, we used SAS (SAS Institute) and R software. Multiple imputation and univariate analysis were accomplished using SAS. For continuous variables that did not follow a normal distribution, we used nonparametric tests for comparison; for comparisons involving categorical variables, we used the chi-square test. If the expected count in 25% of the cells was <5, the chi-square test was deemed ineffective, and the Fisher exact probability test was used. LASSO, Boruta, machine learning algorithms, and plotting were performed using the *glmnet*, *Boruta*, *tidymodels*, and *ggplot* packages in R, respectively. The model interpretability analysis was performed using the *dalex* package for variable importance ranking and PDPs, and the web applications used the *shiny* package in R. The flowchart of the data analysis and machine learning modeling process is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart of the data analysis and machine learning modeling process. DT: decision tree; KNN: K-nearest neighbor; LASSO: least absolute shrinkage and selection operator; LightGBM: light gradient boosting machine; LR: logistic regression; RF: random forest; ROC: receiver operating characteristic curve; STD: sexually transmitted disease; XGBoost: extreme gradient boosting.

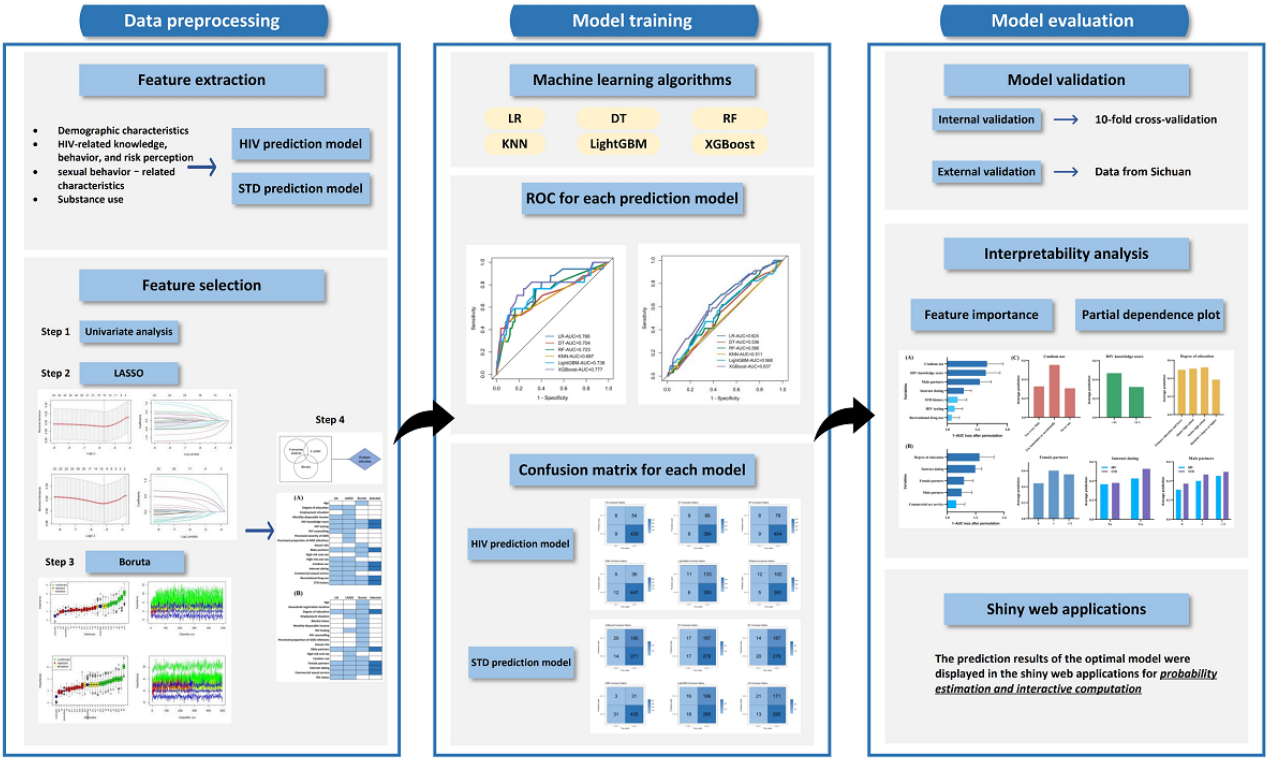

Ethical Considerations

The study was approved by the Chongqing Medical University Ethics Committee (reference numbers 2012010 for 12th Five-Year Plan and 2019001 for 13th Five-Year Plan). Before participating in this research, the participants were fully informed about the purpose, significance, voluntary participation, and confidentiality of the research. Each participant signed an informed consent form and was also compensated with RMB 80 (US $10.90) for transportation and breakfast. The study data were anonymized and deidentified to ensure participant privacy and confidentiality. All images in the manuscript and Multimedia Appendix 1 were reviewed to ensure that individual participants and users cannot be identified.

Results

Overview

In total, we collected individual characteristics from 2345 MSM and excluded those who did not meet the requirements, resulting in 1999 (85.24%) MSM being included in the analyses (refer to Figure S1 in Multimedia Appendix 1 for a flowchart of the participant screening process). The percentage of HIV-positive MSM was 3.6% (72/1999), and the percentage of those who self-reported as previously infected with STD was 7.3% (146/1999). The basic demographic information of the 1999 MSM is shown in Table 1.

XSL•FO
RenderX

Table 1. Description of basic demographic variables for the surveyed population of men who have sex with men (N=1999).

Variables	Values, n (%)
Age (y)	
18-25	147 (7.35)
25-35	732 (36.62)
35-45	689 (34.47)
≥45	431 (21.56)
Household registration location	
Urban	1430 (71.54)
Rural	569 (28.46)
Ethnicity	
Ethnic minority individuals	168 (8.4)
Han	1831 (91.6)
Degree of education	
Primary education and lower	42 (2.1)
Junior high school	170 (8.5)
Senior high school	955 (47.77)
Bachelor's degree or higher	792 (39.62)
Employment	
Unemployed	177 (8.85)
Employed	1591 (79.59)
Student	231 (11.55)
Marital status	
Unmarried	1685 (84.29)
Married	314 (15.71)
Monthly disposable income (RMB)	
1000 (US $137.20) to 3000 (US $411.50)	811 (40.57)
3000 (US $411.50) to 10,000 (US $1371.60)	1088 (54.43)
≥10,000 (US $1371.60)	100 (5)

Feature Selection

LASSO and Boruta were used for feature selection and incorporating factors common to these methods, respectively (refer to Tables S1 and S2 and Figures S2 and S3 in Multimedia Appendix 1 for results of variable screening for each method). On the basis of the results of the statistical analysis, 7 predictors were finally included in the HIV prediction model: HIV knowledge score, HIV testing, number of male partners, condom use, internet dating, recreational drug use, and STD history. A total of 5 predictors were included in the STD prediction model: degree of education, number of male partners, number of female partners, internet dating, and commercial sexual service. The process of variable screening is presented in Figure 2. The results of the selected predictors in the univariate analysis are provided in Table 2.

XSL•FO
RenderX

Figure 2. The process of variable screening (A) in the HIV prediction model and (B) in the sexually transmitted disease (STD) prediction model. LASSO: least absolute shrinkage and selection operator; UA: univariate analysis.

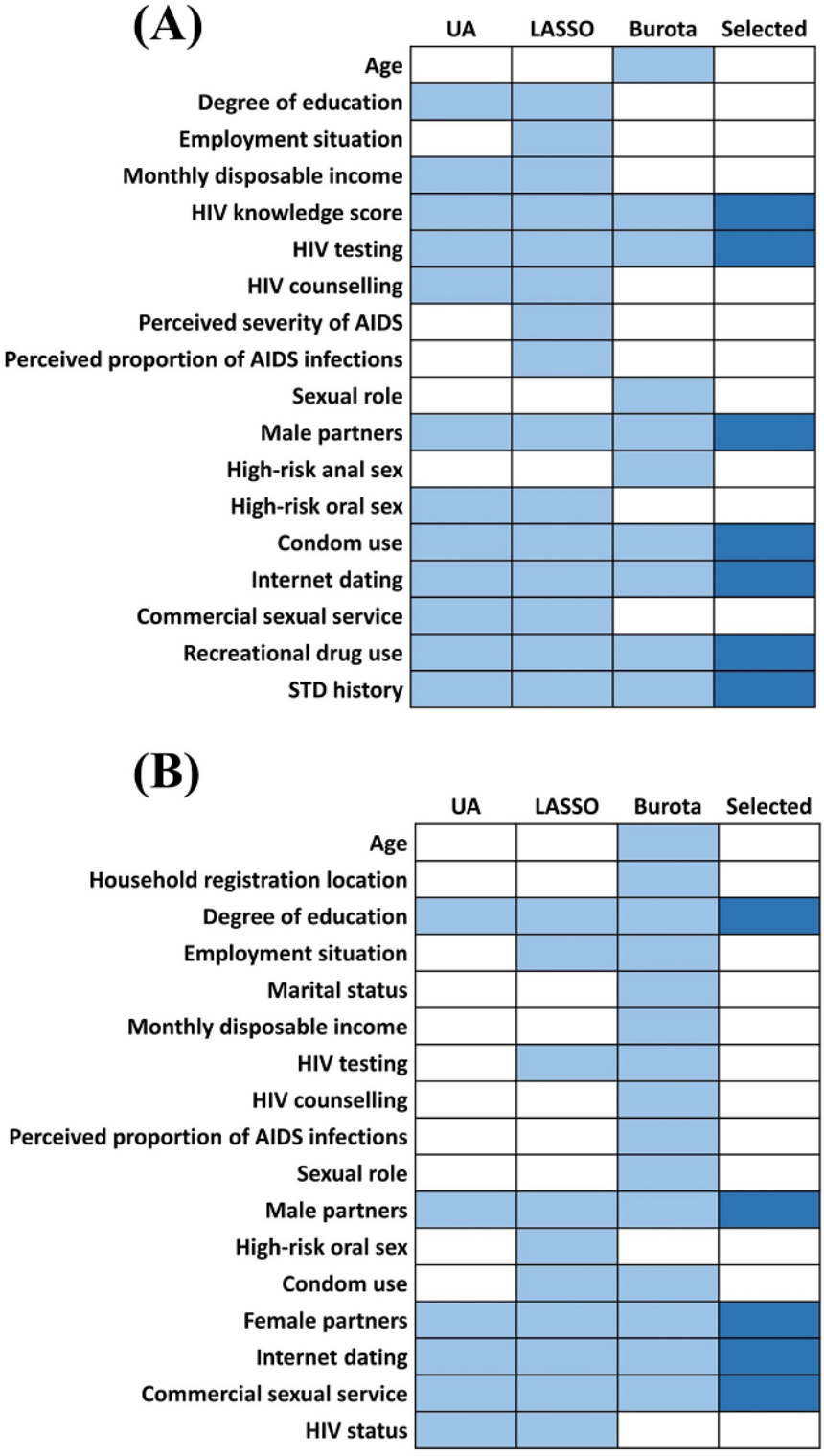

XSL•FO
RenderX

Table 2. Results of selected predictors in HIV and STD[a] prediction models in univariate analysis (N=1999).

Variables	Participants tested positive for HIV (n=72), n (%)	Participants tested negative for HIV (n=1927), n (%)	*P* value	Participants previously infected with STDs (n=146), n (%)	Participants never infected with STDs (n=1853), n (%)	*P* value
Degree of education			—[b]			.046
Primary education and lower	—	—		4 (2.7)	38 (2)	
Junior high school	—	—		16 (10.9)	154 (8.3)	
Senior high school	—	—		84 (57.5)	911 (49.2)	
Bachelor's degree or higher	—	—		42 (28.8)	750 (40.5)	
HIV knowledge scores			<.001	—	—	—
<11	55 (76.4)	1041 (54)				
≥11	17 (23.6)	886 (46)				
HIV testing			<.001	—	—	—
No	20 (27.8)	258 (13.4)				
Yes	52 (72.2)	1669 (86.6)				
Number of male partners			.002			.04
0	5 (77)	293 (15.2)		12 (8.2)	286 (15.4)	
1	34 (47.2)	1098 (57)		84 (57.5)	1048 (56.6)	
≥2	33 (45.8)	536 (27.8)		50 (34.2)	519 (28)	
Condom use			<.001	—	—	—
Use every time	31 (43)	1227 (63.7)				
Use sometimes or occasionally	37 (51.4)	547 (28.4)				
Never use	4 (5.6)	153 (7.9)				
Number of female partners			—			.003
0	—	—		108 (74)	1571 (84.8)	
1	—	—		33 (22.6)	246 (13.3)	
≥2	—	—		5 (3.4)	36 (1.9)	
Internet dating			.01			<.001
No	17 (23.6)	738 (38.3)		33 (22.6)	722 (39)	
Yes	55 (76.4)	1189 (61.7)		113 (77.4)	1131 (61)	
Commercial sexual services			—			.09
No	—	—		135 (92.5)	1771 (95.6)	
Yes	—	—		11 (7.5)	82 (4.4)	
Recreational drug use			.08[c]	—	—	—
No	68 (94.4)	1886 (97.9)				
Yes	4 (5.6)	41 (2.1)				
STD history			<.001	—	—	—
No	58 (80.6)	1795 (93.2)				
Yes	14 (19.4)	132 (6.8)				

[a]STD: sexually transmitted disease.

[b]Not available.

[c]Fisher exact probability test.

XSL•FO
RenderX

Model Construction

We randomly divided the data according to a 75% (1499/1999) training set and a 25% (500/1999) testing set, and comparisons of each predictor between the training and testing sets of the HIV and STD prediction models are provided in Tables S3 and S4 in Multimedia Appendix 1. The synthetic minority oversampling technique was used to balance against the dependent variable of the training set. The balanced data were used to predict the risk of HIV infection and STDs using 6 machine learning algorithms, and the predictions were applied to the testing set. The evaluation indicators of the prediction results of each model obtained through the confusion matrix are provided in Table 3 (refer to Figures S4 and S5 in Multimedia Appendix 1 for the confusion matrix of each prediction model).

Table 3. The prediction results of each model obtained through the confusion matrix.

Outcomes	AUC[a] (95% CI)	Accuracy (%)	Balanced accuracy (%)	Sensitivity (%)	Specificity (%)	PPV[b] (%)	NPV[c] (%)
HIV							
LR[d]	0.766 (0.641-0.890)	0.874	0.679	0.471	0.889	0.129	0.979
DT[e]	0.704 (0.545-0.862)	0.786	0.662	0.529	0.795	0.083	0.979
RF[f]	0.723 (0.591-0.856)	0.824	0.654	0.471	0.836	0.092	0.978
KNN[g]	0.687 (0.562-0.811)	0.904	0.610	0.294	0.925	0.122	0.973
LightGBM[h]	0.738 (0.592-0.884)	0.722	0.686	0.647	0.724	0.076	0.983
XGBoost[i]	0.777 (0.639-0.915)	0.786	0.747	0.706	0.789	0.105	0.987
STDs[j]							
LR	0.624 (0.536-0.711)	0.582	0.625	0.588	0.582	0.093	0.951
DT	0.536 (0.442-0.631)	0.592	0.549	0.500	0.599	0.083	0.943
RF	0.566 (0.473-0.660)	0.586	0.505	0.412	0.599	0.070	0.933
KNN	0.511 (0.455-0.567)	0.876	0.511	0.088	0.933	0.088	0.933
LightGBM	0.568 (0.466-0.671)	0.592	0.536	0.471	0.601	0.079	0.940
XGBoost	0.637 (0.541-0.732)	0.632	0.585	0.618	0.633	0.109	0.958

[a]AUC: area under the curve.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]LR: logistic regression.

[e]DT: decision tree.

[f]RF: random forest.

[g]KNN: K-nearest neighbor.

[h]LightGBM: light gradient boosting machine.

[i]XGBoost: extreme gradient boosting.

[j]STD: sexually transmitted disease.

The area under the curve (AUC) values of each machine learning model and the results of 10-fold cross-validation are presented in Figures 3A-3D. Figures 3E and 3F show heat maps of the predictive performance data in all models for HIV infection and STD. The results indicated that XGBoost had the highest combined predictive performance scores in the HIV infection and STD prediction models, which were 33 and 39, respectively.

XSL•FO
RenderX

Figure 3. Area under the curve (AUC) values for machine learning models, 10-fold cross-validation, and evaluation indicators: (A) AUC of the HIV prediction model, (B) results of 10-fold cross-validation of the HIV prediction model, (C) AUC of the sexually transmitted disease (STD) prediction model, (D) results of 10-fold cross-validation of the STD prediction model, (E) heat map visualization of the predictive performance of each indicator in HIV prediction model, and (F) heat map visualization of the predictive performance of each indicator in STD prediction model. CV: cross-validation; DT: decision tree; KNN: K-nearest neighbor; LightGBM: light gradient boosting machine; LR: logistic regression; NPV: negative predictive value; PPV: positive predictive value; RF: random forest; ROC: receiver operating characteristic curve; XGBoost: extreme gradient boosting.

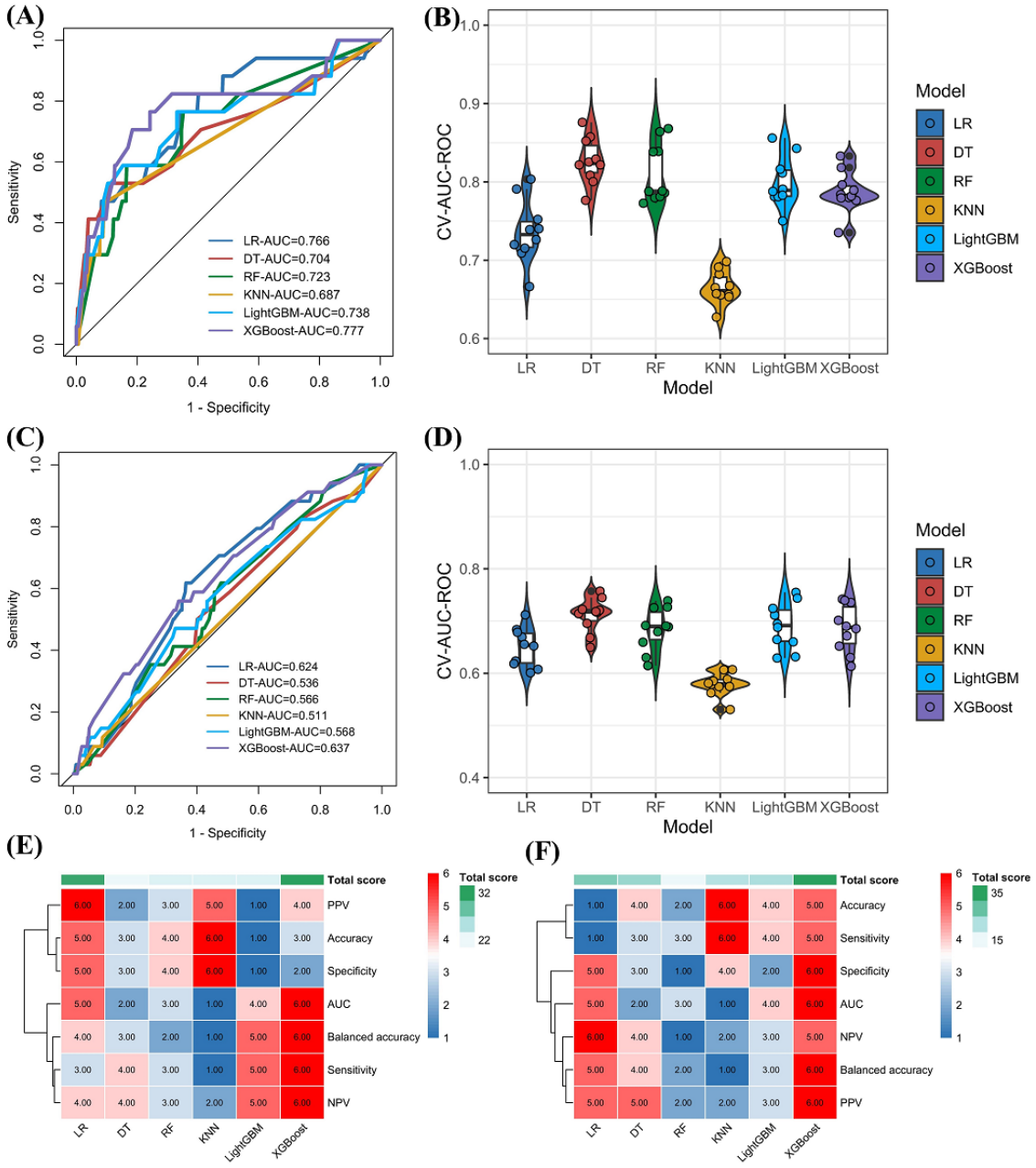

Interpretability Analysis

The XGBoost algorithm was chosen as the optimal model to rank the importance of all predictors, and the top 4 variables were analyzed for interpretability using the PDPs (Figure 4). According to the results of the study, condom use, HIV knowledge score, number of male partners, and internet dating were the more important variables in the HIV risk prediction model. Degree of education, internet dating, number of female partners, and number of male sexual partners were the more important variables in the STD prediction model. According to the PDPs, MSM who sometimes or occasionally use condoms, have an HIV knowledge score of <11, have ≥2 male partners, and use internet dating reported a higher risk of HIV infection than other MSM. MSM with a low degree of education, those using internet dating, having a female partner, and having ≥2 male partners had a higher risk of STDs than other MSM.

XSL•FO
RenderX

Figure 4. Interpretability analysis of optimal model: (A) importance ranking of variables for HIV prediction model, (B) importance ranking of variables for sexually transmitted disease (STD) prediction model, and (C) partial dependence plots for each variable in the top 4 rankings. AUC: area under the curve; STD: sexually transmitted disease.

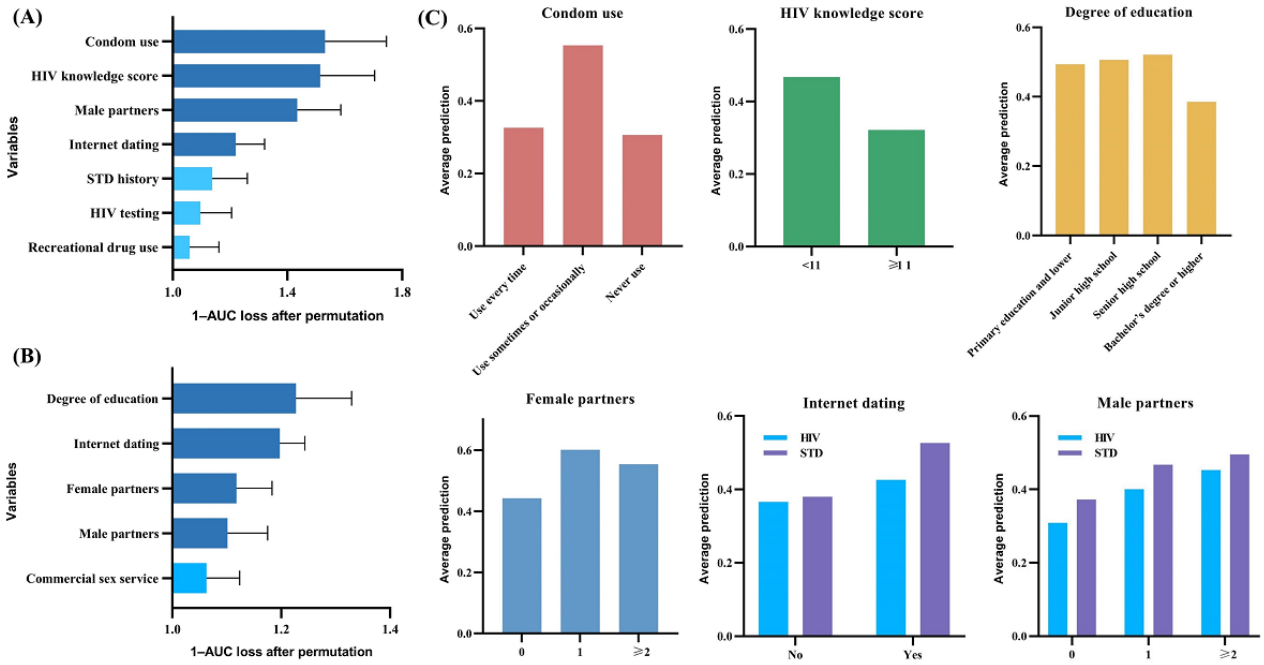

Risk Stratification

On the basis of the cutoff points of the receiver operating characteristic curves of the testing set prediction models, in the HIV prediction model, we defined PRS<0.506 as the low-risk group and PRS ≥0.506 as the high-risk group; in the STD prediction model, we defined PRS<0.479 as the low-risk group and PRS ≥0.479 as the high-risk group. The PRS and risk group distribution of HIV infection and STDs among MSM in the testing set are presented in Figures 5A and 5B. Of the 500 MSM, 386 (77.2%) were in the HIV low-risk group and 114 (22.8%) were in the HIV high-risk group, and 285 (57%) were in the STD low-risk group and 215 (43%) were in the STD high-risk group. The actual number of HIV infections was statistically different in the high- and low-risk groups (P<.001; Figure 5C), and the predicted probability of HIV was significantly higher in the high-risk group than in the low-risk group (P<.001; Figure 5D). In the high-risk group, the actual number of STDs was higher than in the low-risk group (P=.05; Figure 5E), and the predicted probability of STD was significantly higher in the high-risk group than in the low-risk group (P<.001; Figure 5F).

XSL•FO
RenderX

Figure 5. Evaluation of the predicted risk scores of the optimal model in the testing set: (A) distribution of predicted risk scores and risk groups among men who have sex with men (MSM) in HIV prediction model, (B) distribution of predicted risk scores and risk groups among MSM in sexually transmitted disease (STD) prediction model, (C) HIV predicted model's actual number of infections compared among the risk groups, (D) HIV model's predicted probability compared among the risk groups, (E) STD predicted model's actual number of infections compared among the risk groups, and (F) STD model's predicted probability compared among the risk groups.

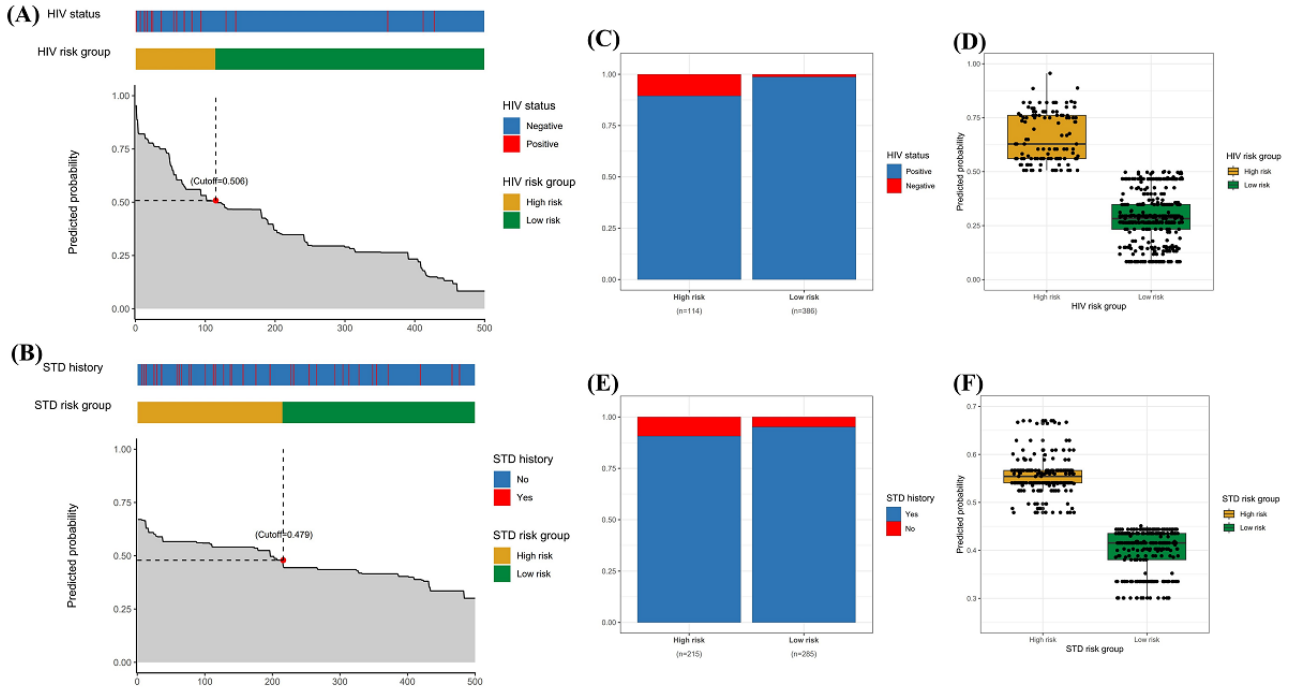

Sensitivity Analysis and Web Applications

Applying the optimal model to an external validation set showed that the model had a fair ability to predict the risk of HIV infection and STDs, with AUC values of 0.627 (95% CI 0.518-0.836) and 0.687 (95% CI 0.528-0.846), respectively. A comparison of differences in demographic characteristics between the external and internal validation sets is provided in Tables S5-S7 in Multimedia Appendix 1. Meanwhile, based on the results of the sensitivity analysis, the optimal model was applied to the MSMM, non-MSMM, MSMO, and MSMW datasets, respectively, and the AUC values ranged from 0.634 to 0.839, indicating that the model was also capable of predicting the risk of HIV infection and STDs in the subgroups (Figure S6 in Multimedia Appendix 1). Finally, the prediction results of the optimal model were displayed in the shiny web applications for probability estimation and interactive computation [29].

Discussion

Principal Findings

Our study demonstrated the accuracy of machine learning methods in the prediction of the risk of HIV infection or STDs among MSM. Among the 6 machine learning methods, the XGBoost model outperformed the others, achieving higher AUC values as well as greater accuracy and sensitivity. Furthermore, the most useful predictors were conventional and easy to collect. Optimization-based models were able to stratify the risk of HIV infection and STDs in MSM well. Sensitivity analyses showed that the prediction models for HIV or STD also demonstrated good performance in predicting MSM subgroups with different risks of infection. In addition, the web-based calculators developed in this study provided a valuable tool for clinicians and MSM to make more informed behavioral decisions. These data could guide prevention and intervention programs, allowing users to tailor interventions to the risk profile of the individual. The routine collection of data, the accuracy of the prediction models, and the convenience of the web applications suggested that machine learning methods may be of value in HIV infection or STD risk assessment services.

Machine learning methods offer flexibility and scalability, making them suitable for a variety of tasks, such as risk stratification, especially when analyzing big data [30]. Risk prediction models are widely used in clinical scenarios, and clinicians may be able to use clinical prediction models to quantify predictions and more intuitively present their risk of infection and intervention options to the MSM population, potentially moderating the information asymmetry between physicians and patients to some extent. In addition, risk prediction models can play a role in the tertiary prevention system of diseases. For example, according to the individual characteristics of high-risk groups, the risk of potential future infections can be quantified, providing a more intuitive and scientific reference for health education and behavioral interventions, and noninvasive, cost-effective, and convenient indicators can be used for risk screening, thus realizing secondary prevention of "early detection, early diagnosis, and early treatment." Clinicians can more accurately screen appropriate study participants, and government departments and health management decision makers can better manage the quality of health care and rationally allocate health care resources. Meanwhile, for MSM, web applications enable self-assessment of infection risk anytime and anywhere based

XSL•FO
RenderX

on their individual situations, which is conducive to protecting their privacy and strengthening awareness of self-management and self-monitoring among high-risk groups.

In our study, for the MSM population in Western China, our prediction results were consistent with previous prediction studies and better predicted the risk of infection. A study conducted to predict future occurrence of syphilis in a high-risk cohort of MSM and transgender women in Peru reported an AUC value of 0.690 [10]. In HIV diagnosis studies of MSM in Australia, the AUC value of their prediction models ranged from 0.698 to 0.763, and for the diagnosis of other STDs, the AUC value ranged from 0.632 to 0.858 [31]. However, the STD prediction model in this study had a slightly lower AUC value (0.777 vs 0.637) compared to the HIV prediction model, which may be related to the way the dependent variables were collected. We objectively measured the HIV infection status of MSM through HIV antigen-antibody testing, whereas the STD history of MSM was obtained through subjective self-reports. MSM may have chosen to conceal their infection history due to social desirability bias, leading to an underestimation of outcomes [32]. Although subjective measurements may lead to a reduction in the efficacy of prediction models, their predictions were generally reliable. At the same time, according to the results of the HIV prediction model, STD history was also an important factor in HIV infection, which was consistent with the results of previous studies [10,33]. In our study, we developed separate prediction models for HIV infection and STDs, designed to function as a *2-step* risk assessment tool. The first step evaluates the risk of STDs in MSM, followed by the second step, which assesses the risk of HIV, offering more comprehensive information for HIV risk prediction.

Our results showed that the XGBoost algorithm performed optimally in predicting the risk of HIV infection and STDs in the MSM population in Western China. XGBoost is an enhanced version of the gradient boosting machine. The optimal performance of XGBoost has been reported in studies predicting HIV infection in Zimbabwe from 2005 to 2015 and in studies predicting timely clinic attendance and acceptance of HIV infection or STD testing [34,35]. The XGBoost algorithm incorporates regular terms to control the complexity of the model, which can effectively prevent overfitting. In addition, XGBoost modeling has been shown to gain signal from missingness without resorting to imputation techniques [36]. In summary, in predicting the risk of HIV infection and STDs in the MSM population in Western China, XGBoost has reported the best prediction performance in terms of discriminative ability and clinical application. XGBoost can be used as a prioritized algorithm in future risk prediction tools.

According to the interpretability analysis of the prediction model, condom use was one of the important factors in predicting the risk of HIV infection in the MSM population, which was also consistent with the results of previous studies [31]. High-risk sexual behaviors, such as unprotected sex and inconsistent condom use, may accelerate HIV transmission between MSM [37,38]. According to our findings, MSM who sometimes or occasionally used condoms were predicted to have a higher mean probability of infection than MSM who used condoms every time. However, the probability of infection among MSM who never used condoms was not high, and we believe that it may be related to the social desirability bias of MSM, which made them reluctant or afraid to admit that they did not use condoms, leading to an underestimation of this outcome.

In addition, our findings showed that MSM who found sexual partners through the internet (ie, internet dating) had a higher risk of both HIV infection and STD than MSM who did not. Considering that individuals have free access to explicit sexual content on the internet, coupled with the increased risk of MSM infection, the role of the internet in sexual partner encounters and its correlation with risky sexual behavior has become an important topic [39]. The internet was known to be a popular place for MSM to find sexual partners; for example, previous studies found that 40% of MSM used the internet to find sexual partners [40,41]. Previous studies have shown that MSM who use internet dating are more likely to engage in psychotropic drug use and unprotected intercourse compared to other MSM [42]. This result was consistent with our findings. It is evident that targeted measures should be taken to improve their attitudes and perceptions and reduce their risk of infection.

A study in the United States showed that MSM with multiple partners are at higher risk of HIV infection [43]. This finding was confirmed by 2 national cross-sectional surveys conducted in Brazil in 2009 and 2016 [44]. In addition, previous studies in China have found that MSM with multiple sexual partners and those having unprotected sex were at increased risk of HIV infection [26]. These were consistent with our findings. It is clear that MSM with multiple sex partners should be given timely attention and education to reduce their risk of infection. Moreover, our prediction model screened for variables such as HIV knowledge score, degree of education, and HIV testing. All of these variables were significant predictors of infection risk, and interventions targeting these variables were also important for HIV infection and STD prevention in the MSM population.

Limitations

In summary, we have developed risk prediction tools that can be widely used in real-life settings. All predictors are self-reported risks, and a user-friendly, web-based self-assessment tool has the potential to help MSM identify their risk of HIV infection or STDs in a nonclinical setting, which enables them to determine their risk of infection and seek timely medical help. Meanwhile, in clinical scenarios, these tools will help clinicians stratify high-risk individuals and be more attentive to individuals who are potentially at high risk of infection. However, there were some limitations in our study. First, we recruited participants using nonprobability sampling, which may have led to bias and limited the generalizability of this study (refer to Multimedia Appendix 1 for the evaluation of the bias in the data relative to the population). In addition, our study relied on voluntary participation by MSM, and the characteristics of participants may have differed from those of nonparticipants, leading to self-selection bias. The validity of risk prediction models depended on the accuracy of self-reported information, which was influenced by MSM recall bias and social desirability bias. In future research on MSM, some

XSL•FO
RenderX

accurate and reliable information gathering tools are necessary, for example, computer-assisted self-interview [31]. Our study revealed significant spatial heterogeneity in the demographic characteristics and the risk of HIV infection or STDs among the MSM population in Western China. This heterogeneity reflects regional differences in population structure, socioeconomic conditions, behavioral patterns, and access to health care services. Considering the epidemiological context, these findings suggest the need to develop tailored prevention and control strategies based on regional characteristics to effectively reduce the risk of HIV and STD transmission in various areas (Figure S7 and Table S8 in Multimedia Appendix 1). Second, the sample size of our study may be smaller than that of other large-scale national survey studies. As the study focused the MSM population, conducting a large-scale national survey was difficult. However, the results of risk prediction models and sensitivity analyses showed that, with the current sample size, the predictions were considerable and could be used in real-world applications. Third, we only used 6 machine learning methods for risk prediction, and others may be explored as part of future research. Fourth, although the optimal prediction model performed moderately well in the MSM in Sichuan (ie, the external validation set), our analysis, presented in Tables S5-S7 in Multimedia Appendix 1, revealed significant differences in demographic variables between the external and internal validation sets. These discrepancies may have contributed to inaccurate predictions, but they also highlighted the authenticity of the external data used to test the risk prediction model, as such differences in the external validation dataset were unavoidable. Future attempts should be made to apply external validation to data from multiple regions to demonstrate the robustness of the prediction model. Finally, our study represented risk projections of outcomes and characteristics from cross-sectional surveys rather than prospective cohort projects, and longitudinal projections will be part of future surveys. As the models were constructed only from Western China, further exploration and validation were needed to generalize the optimal model to the whole country to make it universally applicable.

Conclusions

Our results showed that the machine learning approach demonstrated good performance in predicting the risk of HIV infection or STDs in the MSM population in Western China. The method could be used for the risk stratification of MSM, which would help clinicians and MSM to conduct risk assessment and intervene and prevent the risk in a timely manner. At the same time, user-friendly and easily operated web applications have important public health implications in risk monitoring among the MSM population and can be widely used in reality.

Acknowledgments

The authors would like to thank all the men who have sex with men participants, the members of the nongovernmental organizations, and the physicians working at the First Hospital of Chongqing Medical University. This work was supported by the National Key Project for Infectious Diseases of the Ministry of Science and Technology of China (2012ZX10001007-007 and 2018ZX10721102-005).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

XZ was responsible for the conception and design of the study. BL, JL, and KL were in charge of the analysis and interpretation and the data collection process. All authors were involved in the drafting of the manuscript and critically reviewed it for important intellectual content. All authors gave their approval on the final version to be published and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary analyses, figures, and tables.
[DOCX File , 1353 KB - jmir_v27i1e59101_app1.docx]

References

1. The path that ends AIDS: 2023 UNAIDS global AIDS update. Joint United Nations Program on HIV/AIDS. 2023. URL: https://tinyurl.com/2p93nzaf [accessed 2023-11-27]
2. Dong MJ, Peng B, Liu ZF, Ye QN, Liu H, Lu XL, et al. The prevalence of HIV among MSM in China: a large-scale systematic analysis. BMC Infect Dis 2019 Nov 27;19(1):1000 [FREE Full text] [doi: 10.1186/s12879-019-4559-1] [Medline: 31775654]

XSL•FO
RenderX

3. Lu R, Zhang X, Zhou C, Zhang W, Ouyang L, Xing H, et al. Trends of human immunodeficiency virus, syphilis, and hepatitis c infections among men who have sex with men in Chongqing, China: a serial cross-sectional survey from 2011 to 2018. Sexual Trans Dis 2020 Apr 22;47(7):491-498. [doi: 10.1097/olq.0000000000001173]
4. Chow EP, Grulich AE, Fairley CK. Epidemiology and prevention of sexually transmitted infections in men who have sex with men at risk of HIV. Lancet HIV 2019 Jun;6(6):e396-e405. [doi: 10.1016/S2352-3018(19)30043-8] [Medline: 31006612]
5. Fan L, Yu A, Zhang D, Wang Z, Ma P. Consequences of HIV/syphilis co-infection on HIV viral load and immune response to antiretroviral therapy. Infect Drug Resist 2021 Jul;Volume 14:2851-2862. [doi: 10.2147/idr.s320648]
6. Yang LG, Zhang XH, Zhao PZ, Chen ZY, Ke WJ, Ren XQ, et al. Gonorrhea and chlamydia prevalence in different anatomical sites among men who have sex with men: a cross-sectional study in Guangzhou, China. BMC Infect Dis 2018 Dec 18;18(1):675 [FREE Full text] [doi: 10.1186/s12879-018-3579-6] [Medline: 30563478]
7. Lin XX, Meng SY, Ke WJ, Zhang XH, Wang LY, Liao YY, et al. Community engagement on-site rapid test for chlamydia and gonorrhea among men who have sex with men: a pioneering study in Guangzhou, China. BMC Public Health 2022 May 24;22(1):1036 [FREE Full text] [doi: 10.1186/s12889-022-13460-x] [Medline: 35606713]
8. Marchese V, Tiecco G, Storti S, Degli Antoni M, Calza S, Gulletta M, et al. Syphilis infections, reinfections and serological response in a large Italian sexually transmitted disease centre: a monocentric retrospective study. J Clin Med 2022 Dec 18;11(24):7499 [FREE Full text] [doi: 10.3390/jcm11247499] [Medline: 36556115]
9. Hoenigl M, Weibel N, Mehta SR, Anderson CM, Jenks J, Green N, et al. Development and validation of the San Diego Early Test Score to predict acute and early HIV infection risk in men who have sex with men. Clin Infect Dis 2015 Aug 01;61(3):468-475 [FREE Full text] [doi: 10.1093/cid/civ335] [Medline: 25904374]
10. Allan-Blitz LT, Konda KA, Vargas SK, Wang X, Segura ER, Fazio BM, et al. The development of an online risk calculator for the prediction of future syphilis among a high-risk cohort of men who have sex with men and transgender women in Lima, Peru. Sex Health 2018 Jun;15(3):261-268 [FREE Full text] [doi: 10.1071/SH17118] [Medline: 30021680]
11. Wand H, Reddy T, Naidoo S, Moonsamy S, Siva S, Morar NS, et al. A simple risk prediction algorithm for HIV transmission: results from HIV prevention trials in KwaZulu Natal, South Africa (2002-2012). AIDS Behav 2018 Jan 3;22(1):325-336. [doi: 10.1007/s10461-017-1785-7] [Medline: 28470458]
12. He J, Li J, Jiang S, Cheng W, Jiang J, Xu Y, et al. Application of machine learning algorithms in predicting HIV infection among men who have sex with men: model development and validation. Front Public Health 2022;10:967681. [doi: 10.3389/fpubh.2022.967681] [Medline: 36091522]
13. Ji W, Wang C, Chen H, Liang Y, Wang S. Predicting post-stroke cognitive impairment using machine learning: a prospective cohort study. J Stroke Cerebrovasc Dis 2023 Nov;32(11):107354. [doi: 10.1016/j.jstrokecerebrovasdis.2023.107354] [Medline: 37716104]
14. Yi F, Yang H, Chen D, Qin Y, Han H, Cui J, et al. XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease. BMC Med Inform Decis Mak 2023 Jul 25;23(1):137 [FREE Full text] [doi: 10.1186/s12911-023-02238-9] [Medline: 37491248]
15. Wang X, Qiao Y, Cui Y, Ren H, Zhao Y, Linghu L, et al. An explainable artificial intelligence framework for risk prediction of COPD in smokers. BMC Public Health 2023 Nov 06;23(1):2164 [FREE Full text] [doi: 10.1186/s12889-023-17011-w] [Medline: 37932692]
16. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct;2(10):749-760 [FREE Full text] [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
17. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017 Aug 08;318(6):517-518. [doi: 10.1001/jama.2017.7797] [Medline: 28727867]
18. Leme DE, de Oliveira C. Machine learning models to predict future frailty in community-dwelling middle-aged and older adults: the ELSA cohort study. J Gerontol A Biol Sci Med Sci 2023 Oct 28;78(11):2176-2184 [FREE Full text] [doi: 10.1093/gerona/glad127]
19. Sampa MB, Biswas T, Rahman MS, Aziz NH, Hossain MN, Aziz NA. A machine learning web app to predict diabetic blood glucose based on a basic noninvasive health checkup, sociodemographic characteristics, and dietary information: case study. JMIR Diabetes 2023 Nov 24;8:e49113 [FREE Full text] [doi: 10.2196/49113] [Medline: 37999944]
20. Karabacak M, Jagtiani P, Carrasquilla A, Germano IM, Margetis K. Prognosis individualized: survival predictions for WHO grade II and III gliomas with a machine learning-based web application. NPJ Digit Med 2023 Oct 26;6(1):200 [FREE Full text] [doi: 10.1038/s41746-023-00948-y] [Medline: 37884599]
21. Lei M, Wu B, Zhang Z, Qin Y, Cao X, Cao Y, et al. A web-based calculator to predict early death among patients with bone metastasis using machine learning techniques: development and validation study. J Med Internet Res 2023 Oct 23;25:e47590 [FREE Full text] [doi: 10.2196/47590] [Medline: 37870889]
22. Lin B, Liu J, Ma Y, Zhong X. Factors influencing HIV testing and counselling services among men who have sex with men in Western China: a cross-sectional study based on Andersen's Behavioral Model. Environ Health Prev Med 2022;27:21 [FREE Full text] [doi: 10.1265/ehpm.22-00066] [Medline: 35644545]

XSL•FO
RenderX

23. Cui Y, Shi X, Wang S, Qin Y, Wang B, Che X, et al. Machine learning approaches for prediction of early death among lung cancer patients with bone metastases using routine clinical characteristics: an analysis of 19,887 patients. Front Public Health 2022 Oct 6;10:1019168 [FREE Full text] [doi: 10.3389/fpubh.2022.1019168] [Medline: 36276398]
24. Gao Y, Xin L, Lin H, Yao B, Zhang T, Zhou AJ, et al. Machine learning-based automated sponge cytology for screening of oesophageal squamous cell carcinoma and adenocarcinoma of the oesophagogastric junction: a nationwide, multicohort, prospective study. Lancet Gastroenterol Hepatol 2023 May;8(5):432-445. [doi: 10.1016/s2468-1253(23)00004-3]
25. Hu Y, Zhong XN, Peng B, Zhang Y, Liang H, Dai JH, et al. Comparison of depression and anxiety between HIV-negative men who have sex with men and women (MSMW) and men who have sex with men only (MSMO): a cross-sectional study in Western China. BMJ Open 2019 Jan 03;9(1):e023498 [FREE Full text] [doi: 10.1136/bmjopen-2018-023498] [Medline: 30610021]
26. Liu J, Deng R, Lin B, Pan H, Gao Y, Dai J, et al. Risk management on pre-exposure prophylaxis adherence of men who have sex with multiple men: a multicenter prospective cohort study. Risk Manag Healthc Policy 2021 Apr;Volume 14:1749-1761. [doi: 10.2147/rmhp.s295114]
27. Ramakrishnan L, Ramanathan S, Chakrapani V, Goswami P, Deshpande S, Yadav D, et al. Comparison of sexual risk, HIV/STI prevalence and intervention exposure among men who have sex with men and women (MSMW) and men who have sex with men only (MSMO) in India: implications for HIV prevention. AIDS Behav 2015 Dec;19(12):2255-2269 [FREE Full text] [doi: 10.1007/s10461-015-1058-2] [Medline: 25893657]
28. Yin L, Zhao Y, Peratikos MB, Song L, Zhang X, Xin R, et al. Risk prediction score for HIV infection: development and internal validation with cross-sectional data from men who have sex with men in China. AIDS Behav 2018 Jul 21;22(7):2267-2276 [FREE Full text] [doi: 10.1007/s10461-018-2129-y] [Medline: 29786768]
29. A Shiny web application for predicting HIV among MSM based on machine learning. Shiny App. URL: https://lauralin1104.shinyapps.io/shiny-HIV1/ [accessed 2025-01-24]
30. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/s1470-2045(19)30149-4]
31. Bao Y, Medland NA, Fairley CK, Wu J, Shang X, Chow EP, et al. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. J Infect 2021 Jan;82(1):48-59. [doi: 10.1016/j.jinf.2020.11.007] [Medline: 33189772]
32. Firkey MK, Buckheit KA, Mitzel LD, Maisto SA, Palfai T, Vanable P. Sexual risk and social desirability among Black and White men who have sex with men. J Black Sexuality Relatsh 2020;6(3):29-47. [doi: 10.1353/bsr.2020.0006]
33. Solomon MM, Mayer KH, Glidden DV, Liu AY, McMahan VM, Guanira JV, et al. Syphilis predicts HIV incidence among men and transgender women who have sex with men in a preexposure prophylaxis trial. Clin Infect Dis 2014 Oct;59(7):1020-1026 [FREE Full text] [doi: 10.1093/cid/ciu450] [Medline: 24928295]
34. Birri Makota RB, Musenge E. Predicting HIV infection in the decade (2005-2015) pre-COVID-19 in Zimbabwe: a supervised classification-based machine learning approach. PLOS Digit Health 2023 Jun;2(6):e0000260 [FREE Full text] [doi: 10.1371/journal.pdig.0000260] [Medline: 37285368]
35. Xu X, Fairley CK, Chow EP, Lee D, Aung ET, Zhang L, et al. Using machine learning approaches to predict timely clinic attendance and the uptake of HIV/STI testing post clinic reminder messages. Sci Rep 2022 May 24;12(1):8757 [FREE Full text] [doi: 10.1038/s41598-022-12033-7] [Medline: 35610227]
36. Zhang X, Yan C, Gao C, Malin BA, Chen Y. Predicting missing values in medical data via XGBoost regression. J Healthc Inform Res 2020 Dec;4(4):383-394 [FREE Full text] [doi: 10.1007/s41666-020-00077-1] [Medline: 33283143]
37. Traeger MW, Schroeder SE, Wright EJ, Hellard ME, Cornelisse VJ, Doyle JS, et al. Effects of pre-exposure prophylaxis for the prevention of human immunodeficiency virus infection on sexual risk behavior in men who have sex with men: a systematic review and meta-analysis. Clin Infect Dis 2018 Aug 16;67(5):676-686. [doi: 10.1093/cid/ciy182] [Medline: 29509889]
38. Cao Z, Chen J, Lin B, Zhang C, Zhong X. Factors influencing intention on condom use during sexual intercourse with regular female partners among men who have sex with men in Western China: a structural equation modeling analysis. Sexual Trans Dis 2023 Oct 11;50(12):782-788. [doi: 10.1097/olq.0000000000001876]
39. Xu J, Luo Y, Dong H, Zhao G. The effects of internet exposure on sexual risk behavior among sexually experienced male college students in China: cross-sectional study. JMIR Public Health Surveill 2022 May 02;8(5):e31847 [FREE Full text] [doi: 10.2196/31847] [Medline: 35499864]
40. Phillips G2, Magnus M, Kuo I, Rawls A, Peterson J, Jia Y, et al. Use of geosocial networking (GSN) mobile phone applications to find men for sex by men who have sex with men (MSM) in Washington, DC. AIDS Behav 2014 Sep;18(9):1630-1637. [doi: 10.1007/s10461-014-0760-9] [Medline: 24682866]
41. Holloway IW, Pulsipher CA, Gibbs J, Barman-Adhikari A, Rice E. Network influences on the sexual risk behaviors of gay, bisexual and other men who have sex with men using geosocial networking applications. AIDS Behav 2015 Jun 9;19 Suppl 2(Suppl 2):112-122 [FREE Full text] [doi: 10.1007/s10461-014-0989-3] [Medline: 25572832]
42. Zhao P, Tang S, Wang C, Zhang Y, Best J, Tangthanasup TM, et al. Recreational drug use among Chinese MSM and transgender individuals: results from a national online cross-sectional study. PLoS One 2017 Jan 20;12(1):e0170024 [FREE Full text] [doi: 10.1371/journal.pone.0170024] [Medline: 28107391]

XSL•FO
RenderX

43. Chittamuru D, Icard LD, Jemmott JB3, O'Leary A. Prospective predictors of multiple sexual partners among African American men who have sex with men. Arch Sex Behav 2018 Oct 20;47(7):2081-2090. [doi: 10.1007/s10508-018-1207-6] [Medline: 29926260]
44. Guimarães MD, Kendall C, Magno L, Rocha GM, Knauth DR, Leal AF, et al. Comparing HIV risk-related behaviors between 2 RDS national samples of MSM in Brazil, 2009 and 2016. Medicine (Baltimore) 2018 May;97(1S Suppl 1):S62-S68 [FREE Full text] [doi: 10.1097/MD.0000000000009079] [Medline: 29912816]

Abbreviations

AUC: area under the curve
LASSO: least absolute shrinkage and selection operator
MSM: men who have sex with men
MSMM: men who have sex with multiple men
MSMO: men who have sex with men only
MSMW: men who have sex with men and women
PDP: partial dependence plot
PRS: predicted risk score
STD: sexually transmitted disease
XGBoost: extreme gradient boosting

Edited by A Mavragani; submitted 02.04.24; peer-reviewed by A Onzia, J Wang; comments to author 03.12.24; revised version received 12.12.24; accepted 02.01.25; published 20.02.25.

Please cite as:
Lin B, Liu J, Li K, Zhong X
Predicting the Risk of HIV Infection and Sexually Transmitted Diseases Among Men Who Have Sex With Men: Cross-Sectional Study Using Multiple Machine Learning Approaches
J Med Internet Res 2025;27:e59101
URL: https://www.jmir.org/2025/1/e59101
doi:10.2196/59101
PMID:

©Bing Lin, Jiaxiu Liu, Kangjie Li, Xiaoni Zhong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characterizing Public Sentiments and Drug Interactions in the COVID-19 Pandemic Using Social Media: Natural Language Processing and Network Analysis

Wanxin Li[1], BS; Yining Hua[2,3], MS; Peilin Zhou[4], MS; Li Zhou[3], MD; Xin Xu[1], PhD; Jie Yang[1,5], PhD

[1]School of Public Health, the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[3]Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

[4]Thrust of Data Science and Analytics, Hong Kong University of Science and Technology, Guangzhou, China

[5]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Xin Xu, PhD
School of Public Health, the Second Affiliated Hospital
Zhejiang University School of Medicine
No. 866, Yuhangtang Road
Hangzhou, 310058
China
Phone: 86 13575760802
Email: xuxinsummer@zju.edu.cn

Abstract

Background: While the COVID-19 pandemic has induced massive discussion of available medications on social media, traditional studies focused only on limited aspects, such as public opinions, and endured reporting biases, inefficiency, and long collection times.

Objective: Harnessing drug-related data posted on social media in real-time can offer insights into how the pandemic impacts drug use and monitor misinformation. This study aimed to develop a natural language processing (NLP) pipeline tailored for the analysis of social media discourse on COVID-19–related drugs.

Methods: This study constructed a full pipeline for COVID-19–related drug tweet analysis, using pretrained language model–based NLP techniques as the backbone. This pipeline is architecturally composed of 4 core modules: named entity recognition and normalization to identify medical entities from relevant tweets and standardize them to uniform medication names for time trend analysis, target sentiment analysis to reveal sentiment polarities associated with the entities, topic modeling to understand underlying themes discussed by the population, and drug network analysis to dig potential adverse drug reactions (ADR) and drug-drug interactions (DDI). The pipeline was deployed to analyze tweets related to the COVID-19 pandemic and drug therapies between February 1, 2020, and April 30, 2022.

Results: From a dataset comprising 169,659,956 COVID-19–related tweets from 103,682,686 users, our named entity recognition model identified 2,124,757 relevant tweets sourced from 1,800,372 unique users, and the top 5 most-discussed drugs: ivermectin, hydroxychloroquine, remdesivir, zinc, and vitamin D. Time trend analysis revealed that the public focused mostly on repurposed drugs (ie, hydroxychloroquine and ivermectin), and least on remdesivir, the only officially approved drug among the 5. Sentiment analysis of the top 5 most-discussed drugs revealed that public perception was predominantly shaped by celebrity endorsements, media hot spots, and governmental directives rather than empirical evidence of drug efficacy. Topic analysis obtained 15 general topics of overall drug-related tweets, with "clinical treatment effects of drugs" and "physical symptoms" emerging as the most frequently discussed topics. Co-occurrence matrices and complex network analysis further identified emerging patterns of DDI and ADR that could be critical for public health surveillance like better safeguarding public safety in medicines use.

Conclusions: This study shows that an NLP-based pipeline can be a robust tool for large-scale public health monitoring and can offer valuable supplementary data for traditional epidemiological studies concerning DDI and ADR. The framework presented here aspires to serve as a cornerstone for future social media–based public health analytics.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63755) doi:10.2196/63755

KEYWORDS

COVID-19; natural language processing; drugs; social media; pharmacovigilance; public health

Introduction

The emergence of the COVID-19 pandemic has induced an immediate need for effective pharmacotherapies. While the development and application of such therapies are critically important, they are also influenced by an array of political, economic, and social factors. Meanwhile, an overabundance of drug-related information during the COVID-19 pandemic has rapidly proliferated across social media platforms, drawing significant attention from governments and health organizations. This phenomenon, referred to as an "infodemic," has exacerbated the pandemic's impact, caused additional harm to individuals, and undermined the effectiveness and sustainability of the global health system [1,2]. For example, public pronouncements by high-profile figures, such as former US President Donald Trump's endorsement of hydroxychloroquine, have led to its irrational use and consequential public health crises [3]. Traditional pharmacovigilance mechanisms, reliant on clinical trials and formal reporting systems like MedWatch and DrugBank [4-6], offer valuable but lagged information. These traditional approaches are plagued by inefficiencies, reporting biases, and a lack of timeliness, thereby lacking comprehensive coverage of the population's sentiments and experiences [7-10].

In this context, real-time public comments on pharmacotherapies such as medications on social media provide a valuable resource for complementing research on drug use or repositioning for the COVID-19 pandemic. In addition to the fast accessibility, timeliness, and comprehensive population coverage, social media can also supply real-world evidence on how people respond to different drugs, thus helping researchers mine novel drug potency or side effects [11-13]. Social media also offer data on drugs not typically included in pharmacovigilance datasets, such as over-the-counter drugs [14], herbal remedies [15], and other nontraditional treatments [16]. However, the sheer volume and noise in social media data require robust computational methodologies for effective analysis [17].

Natural language processing (NLP) technologies offer a solution to these challenges. Earlier studies, such as the study conducted by Aramaki et al [18] in 2011, demonstrated that Twitter (subsequently rebranded X) data could be mined to monitor influenza outbreaks using machine learning and rudimentary NLP techniques. Contemporary research in this domain has benefitted immensely from technological advancements, such as deep-learning–based NLP tools specifically for analyzing social media data [19-22]. These have made it increasingly feasible to understand large volumes of colloquial, noisy text for the extraction of meaningful insights on public health.

Substantial efforts such as topic modeling and sentiment analysis have been made to analyze pharmacotherapy-related topics during the COVID-19 pandemic. Notably, existing research lacked of data-driven pipeline with state-of-the-art NLP tools and other big data analysis techniques [23-27] or just involved longitudinal data with a small time span [28]. There is still a gap in how to automatically and accurately extract drug information through social media data for longitudinal monitoring of the drug infodemic.

To address these gaps, this study uses NLP methodologies and network analysis for an extensive assessment of COVID-19 drug-related discourse on social media. We contribute to the existing literature in several ways:

(1) Using deep learning methodologies for named entity recognition (NER), thereby reducing the false positives associated with traditional keyword matching.

(2) Re-examining public sentiments and concerns regarding COVID-19 medications, using target sentiment analysis (TSA) and topic modeling.

(3) Conducting a comprehensive assessment of adverse drug reactions (ADR) and drug-drug interactions (DDI) through network analysis techniques.

We demonstrate that our integrated NLP pipeline can serve as a robust framework for extracting and analyzing drug-related information, thereby enhancing the scope and effectiveness of social media–based pharmacotherapy analysis.

Methods

Overview

As shown in Figure 1, the study workflow is organized into three primary stages: data collection, development of an NLP pipeline, and subsequent data analysis using the constructed pipeline. Initially, we curated a dataset of English tweets related to the COVID-19 pandemic. After a preprocessing phase that excluded tweets with URLs, an NLP pipeline was developed to extract and normalize the drugs and symptoms mentioned in these tweets. Finally, we examined the time trends of drug mentions, public sentiment, and discussion topics toward drugs, as well as the co-occurrence network of drug-drug and drug-symptom pairs [22].

XSL•FO
RenderX

Figure 1. Workflow of drug analysis with natural language processing on Twitter. LDA: latent Dirichlet allocation; METS-CoV: Medical Entity and Targeted Sentiment on COVID-19 Related Tweets; NLP: natural language processing; NER dataset containing medical entities and targeted sentiments from COVID-19–related tweets.

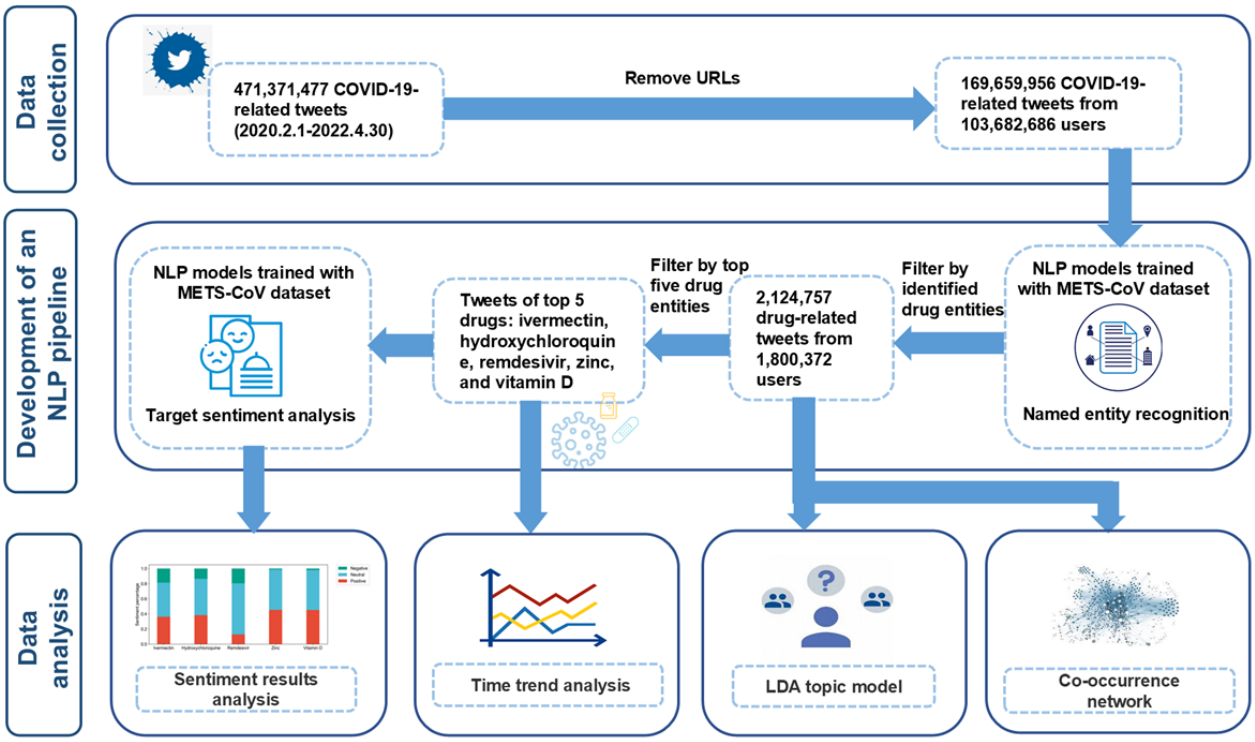

Data Collection and Preprocessing

COVID-19–related tweets from February 1, 2020, to April 30, 2022 were downloaded using Twitter's application programming interface (API) through unique tweet IDs, which were obtained from a public dataset provided by Chen et al [29]. Due to the privacy restrictions of Twitter data, the raw tweets were not publicly available and could only be shared by tweet ID. Therefore, we downloaded tweets by Twitter API based on the provided tweet IDs. The downloaded data included full tweet texts and corresponding metadata such as timestamps and user information. Tweets containing URLs were excluded from the analysis, as they often only contained summaries or quotations of the original tweet. The data collection process adhered to Twitter's privacy and data use management policies.

NLP Pipeline Development

The NLP pipeline consists of 4 principal modules: NER, TSA, topic modeling, and drug network analysis. For the NER and TSA modules, we leveraged state-of-the-art models developed in our previous work "Medical Entity and Targeted Sentiment on COVID-19 Related Tweets (METS-CoV)" [22]. Details on model construction can be found in Figure S1 in the Multimedia Appendix 1.

Named Entity Recognition and Normalization

The NER model aims to extract drug entities from tweets. The model we developed, CT-BERT-NER (COVID Twitter with Bidirectional Encoder Representations from Transformers for Named Entity Recognition) [22], was constructed using the COVID-Twitter-BERT (CT-BERT), a widely adopted language model pretrained on 160 million COVID-19–related tweets. CT-BERT-NER was trained on the entire training set of the NER subset of METS-CoV [22]. Upon evaluation, it showed F_1-scores of 86.35% for drug entity recognition and 81.85% for symptom entity recognition on the corresponding test set, respectively [22]. We used the model trained on all entity types (ie, disease, drug, symptom, vaccine, person, location, and organization) instead of on drug entities only to enable the nuanced differentiation of drug entities from other types of entities.

To standardize colloquial expressions of drugs among the extracted entities, we manually searched Wikipedia for NER-identified drug entities with a frequency of more than 1000 to map colloquial drug expressions and their standardized concepts (ie, drug trade names, chemical names, and generic names). We conducted an accuracy assessment using a random sample of 100 tweets for each of the top 5 most frequently mentioned drugs and symptoms, as identified through 2 methods: NER combined with lexicon-based extraction (NER+lexicon) and lexicon-based extraction alone, with a total of 1000 tweets being manual review. Our results demonstrated that the NER+lexicon method achieved an accuracy rate of 97.8%, significantly surpassing the 89% accuracy achieved by the lexicon-only approach (χ^2_1=61.4, P<.001). Further details on this comparison are available in Table S1 in Multimedia Appendix 1.

Targeted Sentiment Analysis

The TSA module is designed to analyze users' sentiments toward specific drug entities within tweets. Inspired by BERT-SPC [30], we first concatenate the original tweet and the identified drug entity, separated by 1 special token "[SEP]," to

XSL•FO
RenderX

form a combined tweet-entity sentence. The tweet-entity sentence is then fed into a pretrained language model to capture semantic features, which are subsequently passed to a linear layer for 3-class sentiment prediction (positive, neutral, or negative). Notably, instead of using the original BERT model, we used CT-BERT, which has been further trained on 97 million COVID-19–related tweets. This adaptation enhances its understanding of COVID-19–related tweet data [22]. Therefore, although there existed several sentiment-specific embeddings and pretrained models [31-34], we chose the pretrained CT-BERT model as it was trained to understand COVID-19–related tweets. On the TSA test set of METS-COV, the model achieved an F_1-score of 62.67% and an accuracy rate of 75.07% across 4 entity types: person, drug, disease, and vaccine. For our own TSA study, we randomly selected 100 drug-related tweets and assessed their emotional orientation toward drug entities using both model predictions and manual review by a researcher with medical expertise. The results indicated that the model's accuracy, when compared to manual review, was 77% (77/100), aligning closely with the TSA model's original accuracy of 75.07%. Furthermore, both our hand-labeled fine-tuned dataset (METS-CoV) and the final applied dataset (169,659,956 drug-related tweets) were derived from the same source, ensuring the reliability and credibility of the predictions.

Topic Model Analysis

To discern prevailing public interests in the most discussed drugs, we implemented latent Dirichlet allocation (LDA) for topic modeling, using the LdaModel function from the Gensim package [35]. Topic numbers were determined based on conventional evaluation metrics, including low perplexity [36] and high coherence scores [37]. Detailed methodologies are delineated in Figure S2 in Multimedia Appendix 1.

Drug Network Analysis

To illustrate potential relationships among drugs, we constructed a drug network analysis module to generate incidence matrices according to a previous study [38] and visualize co-occurrence networks using Gephi [39] and ForceAtlas2 algorithm [40]. For enhanced comprehensiveness, we incorporated a variant supported by the Anatomical Therapeutic Chemical classification system (ATC) [41], in addition to the Gephi-based visualization. In addition, we used the NER model to extract symptom entities and normalize them through a presummarized lexicon list [42] to extend our analysis to drug-symptom networks. The constructed networks feature nodes represented drugs (Table S3 in Multimedia Appendix 1) or symptoms (Table S4 in Multimedia Appendix 1). Node sizes displayed node degrees (ie, the number of linked entities). Edge weights denoted the cosine similarity score of 2 linked nodes. As our focus is not on causal relationships but rather on the interplay between entities, we used undirected graphs and semantic cosine similarity [43] as the distance metric just as we did in the previous work [38]. Cosine similarity is a widely implemented metric in information retrieval and related studies [44]. In our study, each drug or symptom entity can be represented as a vector, with each dimension of the vector corresponding to 1 tweet text. Details for calculation can be found in Methods in Multimedia Appendix 1.

Pipeline Deployment

Upon completion of the NLP pipeline, we proceeded to its deployment on the preprocessed dataset of COVID-19–related tweets. We first applied the NER and normalization module on the preprocessed dataset (ie, removing URLs) to extract and standardize drug entities to drug concepts. Then we filtered the preprocessed COVID-19–related tweets dataset to get the drug-related tweets dataset according to these drug concepts. Following this standardization, we conducted a distributional analysis of drug mentions to discern time trends, thereby capturing the evolving popularity of these drugs. We also gather related news and the trend of weekly new COVID-19 cases to show a more holistic view of the shift in drug popularity over time. For clarity and simplicity, we only illustrate the top 5 most discussed drugs.

Subsequently, we used the TSA model for drug-related tweets of the top 5 drugs mentioned above to assign each drug entity a sentiment type. To gain a deeper understanding, we also conducted a time-trend analysis on the positive and negative tweets for the 5 drugs and visualized the results. Building upon our understanding of public sentiment, we turned to topic modeling via LDA in all drug-related tweets to explore the thematic concentrations in the discourse surrounding drugs. The model yielded the 20 most probable keywords and bigrams for each identified topic, enabling us to summarize the primary themes. We further analyzed the topic distribution associated with each of the top 5 drugs.

Finally, we constructed co-occurrence networks for drug-drug and drug-symptom interactions to provide a relational overview that complements our earlier analyses. All 67 drugs with more than 1000 mentions and 69 symptoms with more than 250 mentions over time were included in the analysis. Meanwhile, we also zoomed in to analyze the 5 most-discussed drugs.

Statistical Analysis

The chi-square test was used to compare the accuracy differences between NER combined with lexicon and lexicon-based only. We used Python software (version 3.8) to conduct the statistical analyses and chose a P value of .001 as the statistically significant threshold.

Ethical Considerations

Ethical approval for this study was granted by the Institutional Review Board of School of Public Health, Zhejiang University (ZGL202201-2).

Results

Data Summary and Trends of Drug Mention Tweets

This study used a dataset consisting of 471,371,477 COVID-19–related tweets in English, which were collected between February 1, 2020, and April 30, 2022. After excluding tweets containing URLs, the final dataset used for this study consisted of 169,659,956 (36.0%) tweets from 103,682,686 users. Using CT-BERT-NER, we identified 2,124,757

XSL•FO
RenderX

drug-related tweets from 1,800,372 unique Twitter users, accounting for approximately 1.25% of the raw COVID-19–related tweets dataset. Table S2 in Multimedia Appendix 1 provides more detailed statistical results of the medical entity recognition.

Table S3 in Multimedia Appendix 1 presents the 67 most frequently mentioned drugs, each with an occurrence exceeding 1000 times. The most frequent taxonomies are ATC [41] N (nervous system drugs) and J (anti-infective drug). We ranked the total occurrence of all drugs and identified the top 5 most-mentioned drugs: ivermectin, hydroxychloroquine, remdesivir, zinc, and vitamin D to visualize their weekly time trends. Figure 2 presents these temporal trends. The new case counts were collected from the World Health Organization (WHO) [37] on a weekly basis, beginning on February 1, 2020. Given that the dataset is confined to English-language tweets, the scope of new case counts was likewise restricted to the top 4 English-speaking nations with the highest Twitter activity: the United States, the United Kingdom, the Philippines, and Canada [38].

Figure 2. Weekly popularity trends of the top 5 most-mentioned drugs on Twitter examined with COVID-19–related tweets collected between February 1, 2020 and April 30, 2022. The left Y-axis represents the total number of tweets for each drug in a given week (unit: thousand tweets). The right Y-axis represents the weekly new case count (unit: million cases). CDC: Centers for Disease Control and Prevention; FDA: Food and Drug Administration; HCQ: hydroxychloroquine.

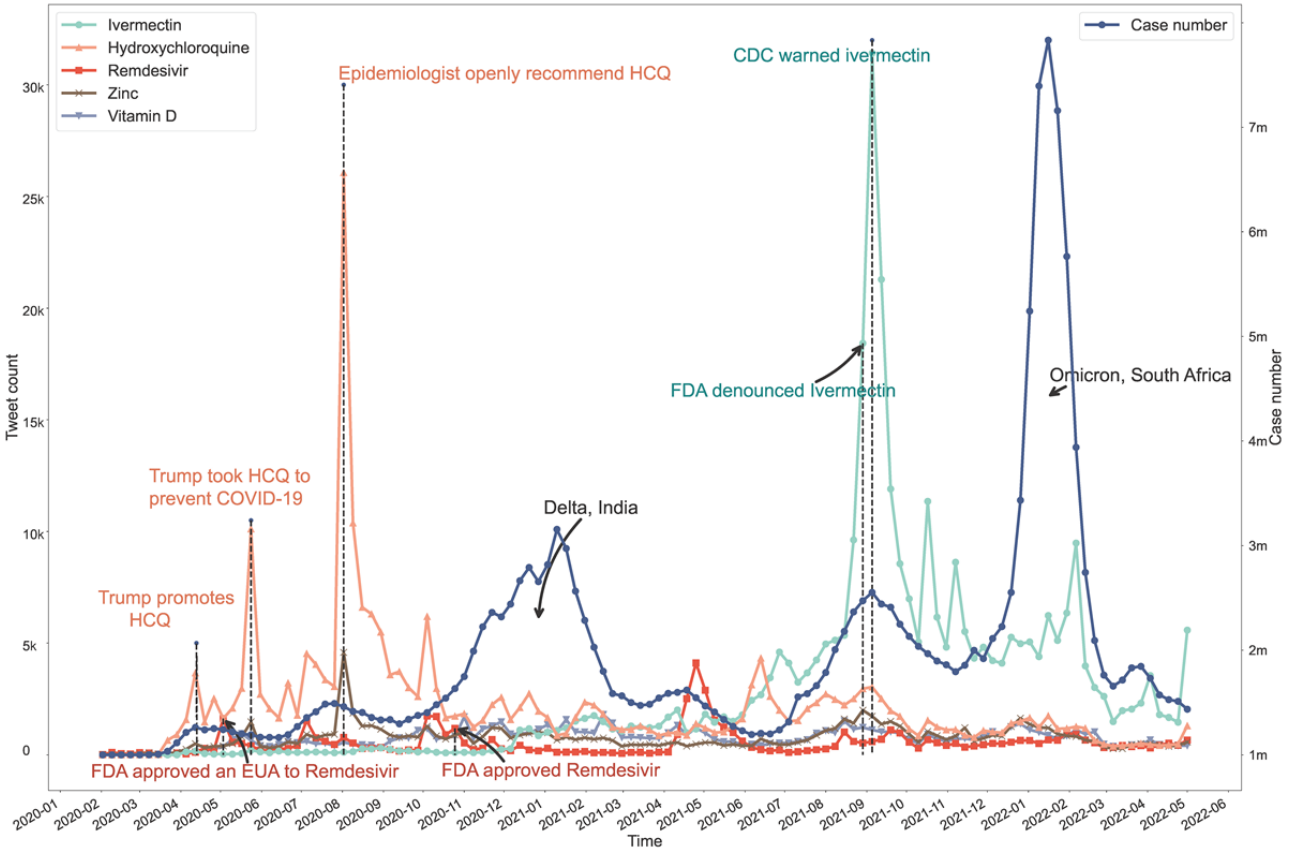

Among the 5 drugs, the public focused mostly on repurposed drugs (ie, hydroxychloroquine and ivermectin), followed by daily supplements (ie, zinc and vitamin D). The only officially approved drug among the 5, remdesivir, received the least attention. The frequency of discussion of hydroxychloroquine and ivermectin fluctuated significantly across time, which seemed to be related to relevant news events or policies (marked in Figure 2). In the early stage of the pandemic, drug-related discussions focused on hydroxychloroquine, with 2 prominent peaks occurring on May 24, 2020, and August 2, 2020. Discussion of ivermectin began to increase in the later stages of the pandemic, with only 1 prominent peak located on September 5, 2021. In contrast, remdesivir received the least public attention, which increased only sporadically throughout the pandemic, with no apparent pattern and a much lower peak on May 3, 2020. As supplements to COVID-19 treatments, vitamin D and zinc elicited much less public interest than ivermectin and hydroxychloroquine, with no significant outbreaks or visible patterns.

Changes in Sentiment for Five Most Frequent Mentioned Drugs

We calculated the sentiment proportion for the 5 drugs and the weekly time trends of positive and negative tweets. Figure 3A shows the visualization of the overall attitude proportions. The public tended to hold positive and neutral attitudes toward the repurposed drugs, ivermectin and hydroxychloroquine. The immune supplements, zinc and vitamin D, were frequently mentioned with positive sentiments. The only COVID-19 drug approved by the Food and Drug Administration (FDA), remdesivir, received the lowest positive attitude, far lower than those of the other drugs.

XSL•FO
RenderX

Figure 3. Sentiment analyses of the 5 top-discussed drugs from February 1, 2020, to April 30, 2022, grouped according to their polarity, including (A) sentiment distribution, (B) weekly ratio of positive tweets, and (C) weekly ratio of negative tweets. The denominator of the percentage was the entities with sentiment. CDC: Centers for Disease Control and Prevention; EUA: emergency use authorization; FDA: Food and Drug Administration; HCQ: hydroxychloroquine.

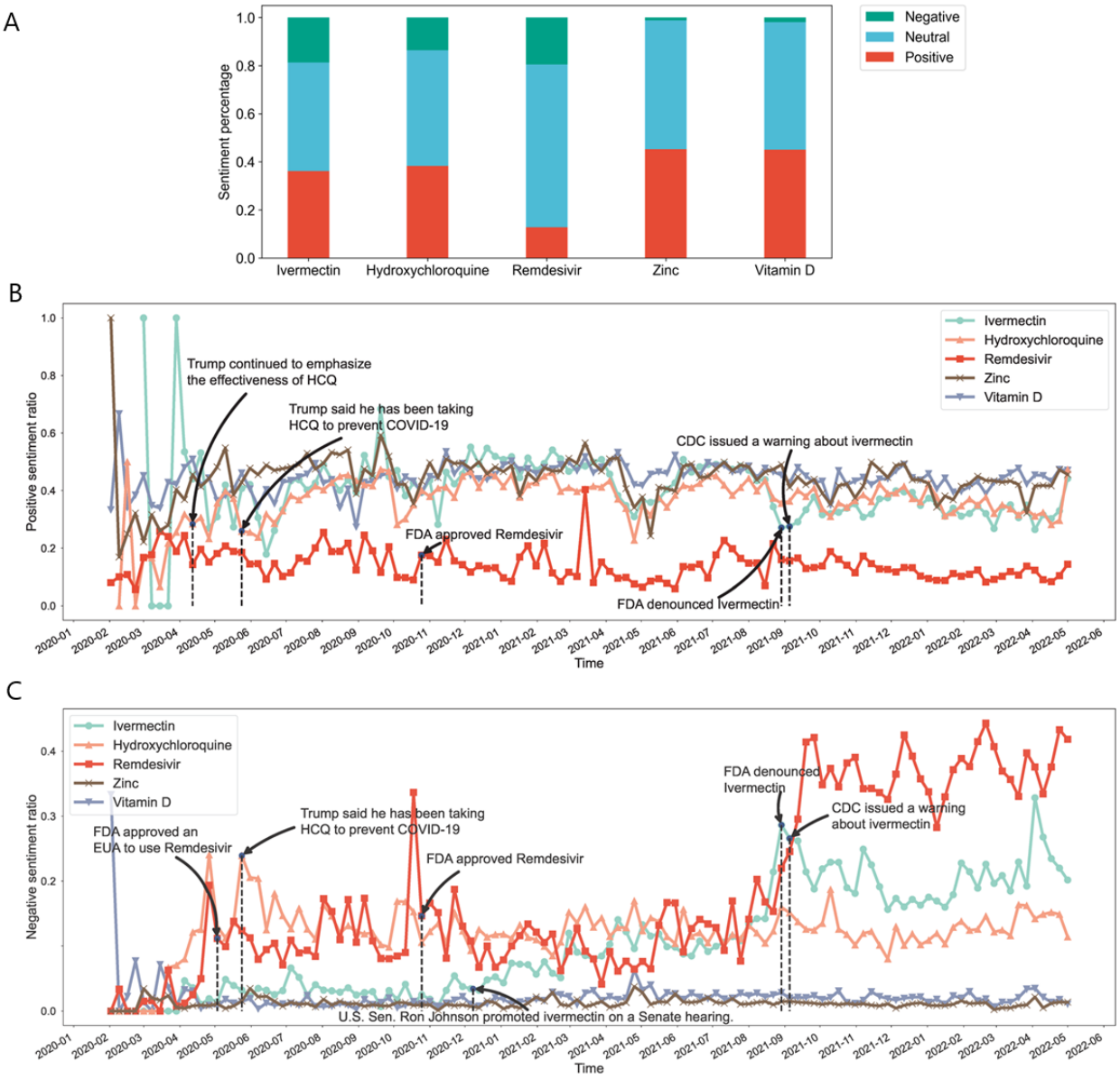

Figures 3B and 3C present weekly trends of tweets expressing positive and negative attitudes, respectively. The major turning points of the trends tend to coincide with new government policies, major social events, and research findings. The criticism of remdesivir (Figure 3C) and ivermectin increased over time since September 2021, and the turning point for remdesivir came at almost the same time as emerging studies showing that the drug is ineffective [45] and has severe side effects [46-48]. For ivermectin, public sentiment was associated with announcements of health authorities and celebrity effects. For example, the FDA denouncing the use of ivermectin for COVID-19 on August 29^{t}, 2021 had simultaneously increasing negative discussions.

Topic Distributions of Drug-Mentioned Tweets

We applied the LDA topic model to all 2,124,757 drug-related tweets and obtained 15 general topics based on their relatively high topic coherence scores and low confusion levels (further discussed in Figure S3 in Multimedia Appendix 1). We displayed the corresponding top 20 most likely keywords in Table 1 and assigned a theme for each topic from these keywords. The topic "clinical treatment effect of drugs" included 288,967 related tweets and dominated the discussions, accounting for 13.6% of all related tweets. In addition, 251,571 (11.84%) were related to "physical symptoms," whereas 220,125 (10.36%), 197,177 (9.28%), 174,868 (8.23%), 172,955 (8.14%), and 154,470 (7.27%) were related to "COVID-19 control," "causes of death," "general treatment," "immune response," and "daily supplement intake." In addition to the overall topic

XSL•FO
RenderX

summary, we explored the distribution of the 15 topics for the 5 drugs. Figure 4 shows a visualization of the distribution. For ivermectin, the prominent theme was “immune response.” In contrast, discussions of remdesivir centered on “hospital care.” Hydroxychloroquine received relatively even attention among the 3 topics “causes of death,” “drug scare,” and “COVID-19 control.” Vitamin D was frequently mentioned in tweets about “daily life,” and the main topics about zinc focused on “hospital care” and “COVID-19 control.”

XSL•FO
RenderX

Table 1. Topic model on drug-related tweets.

Topic	Keywords	Example	Number and Percentage of related tweets, n (%)
Clinical treatment effect of drugs	treatment, covid, study, drug, effective, trial, hydroxychloroquine, prove, safe, covid, evidence, prevent, clinical, vaccine, recommend, cheap, infection, continue, antiviral, efficacy	"@USER[a] Ivermectin is pretty safe but the evidence for it being efficacious against SARS-CoV-2 is lacking."	288,967 (13.60%)
Physical symptoms	day, test, symptom, week, covid, positive, feel, month, steroid, bad, start, med, ago, sick, time, recover, antibiotic, hour, fine, insulin	"When I had covid, it was mild fever for a day, & it was gone"	251,571 (11.84%)
COVID-19 control	cure, vaccine, spread, control, lie, people, drug, covid, push, approve, hydroxychloroquine, claim, force, medium, ban, talk, science, government, experimental, president	"@USER[a] I agree with you re lockdown. Just not on HCQ and vaccines."	220,125 (10.36%)
Causes of death	death, people, die, covid, kill, reason, heart, drug, cancer, cocaine, dead, trust, epidemic, rate, bad, create, attack, sound, result, fentanyl	"@USER[a] Of course, if they died 30 minutes after taking fentanyl but had a positive covid test, guess what their official cause of death is listed as?"	197,177 (9.28%)
General treatment	doctor, treat, patient, steroid, risk, covid, covid, severe, infection, blood, pill, medication, antibiotic, receive, prescribe, illness, lung, hospitalize, prescription, aspirin	"@USER[a] migraines have a very specific causality (i had them for like 20 years), I wonder if the Covid version is one? -- I would try warm compress, NSAIDs, and maybe nasal irrigation with like a neti pot"	174,868 (8.23%)
Immune response	virus, system, body, immune, corona, zinc, fight, antibody, cure, immunity, bleach, deficiency, cell, response, kill, covid, inject, injectingdisinfectant, human, boost	"In addition, mAbs have been shown to improve survival in patients hospitalized with COVID-19 who have not mounted their own immune response."	172,955 (8.14%)
Daily supplement intake	vitamin, people, level, covid, eat, healthy, cold, flu, catch, protect, food, hand, low, stay, vit, survive, bad, chance, common, worry	"@USER[a] regularly take vitamins and a Vitamin D supplement. I started taking the Vitamin D supplement because I wasn't going outside as much at the start of the pandemic. As soon as started taking the Vitamin D supplement, my blood work started to improve."	154,470 (7.27%)
Public panic	live, save, pandemic, life, start, lockdown, real, people, buy, time, basicallystart, money, watch, steroid, hit, hard, normal, deadly, cough, break	"@USER[a] waste of time, politicians have organised orgy's with cocaine, male & female hookers, during lockdown! like they give a fuck about a petition hahaha"	149,583 (7.04%)
Hospital care	patient, care, hospital, remdesivir, treatment, cocktail, covid, covid, injection, require, oxygen, admit, medical, lead, ventilator, pay, provide, health, remove, source	"Both hospitalized and treated immediately with Oxygen & Remdesivir for covid @ the same time. Both went into heavy psychosis."	118,136 (5.56%)
Daily life	drink, stay, hear, lose, family, wait, water, love, leave, pandemic, friend, daily, close, rest, hope, lot, drop, head, play, time	"Interesting situation. Got a call from a Mom, Family of 4 lives in a house. Son and her drinks SOULTOX everyday. Daughter and Dad don't. Dad got Covid, then daughter. Mom and Son tested negative 4xs over the 2 wks. Everyone is vaxed. Now she makes everyone drink SOULTOX now."	97,952 (4.61%)
Political elections	dem, trump, fake, nursinghome, news, pandemic, free, vote, ill, school, election, economy, release, access, truth, guy, forget, sense, deny, hoax	"@realDonaldTrump Hey Captain Covid I VOTED FOR JOE BIDEN AND KAMALA HARRIS! The steroids and tranquilizers are making you more batshit crazy than usual. You should go back to the hospital."	71,179 (3.35%)
Political crisis and discussion	trump, people, person, infect, dangerous, true, covid, condition, shit, happen, stock, send, woman, idea, stupid, destroy, contract, completely, potus, tablet	"@realDonaldTrump @USER[a] At least Putin just poisons his political enemies -- Trump wants all Americans to drink poisons such as hydroxychloroquine and oleandrin -- there just aren't enough dead Americans from COVID-19 for Trump -- he wants to kill more with poison -- Trump is a quack ..."	70,117 (3.30%)

XSL•FO
RenderX

Topic	Keywords	Example	Number and Percentage of related tweets, n (%)
Drug scare	country, low, rate, dexamethasone, cost, covid, drug, recovery, black, base, reduce, sell, supply, explain, mortality, expect, increase, panic, produce, improve	"Once the pandemic hit, stores really upped the prices of aloe vera leaves like i wouldn't notice smh"	60,981 (2.87%)
Personal precautions	mask, wear, medicine, social distance, add, avoid, business, plasma, measure, spray, confirm, skin, drive, advise, campaign, practice, oil, wash hand, air, hourly	"@USER @USER @USER[a] Try working on your immune system. Covid depletes zinc and vitamin D. Try working from prevention instead of fear."	48,444 (2.28%)
Public health care	report, support, health, public, supplement, question, kid, child, datum, stage, issue, american, answer, term, prophylactic, mental, phase, safety, concern, inflammation	"@USER[a] Then, even hcq, zinc, and zithromax had a estimated 50% success and the CDC and WHO said it didnt, so it shouldnt be used. Imagine half of the death tolls because of at least TRYING something instead of shitcanning it just because you are an "authority" that hates Trump."	48,232 (2.27%)

[a]Username and other sensitive information were masked off using @USER. Public figures such as @realdonaldtrump are shown in their usernames.

Figure 4. Topic distribution of 5 top-discussed drugs.

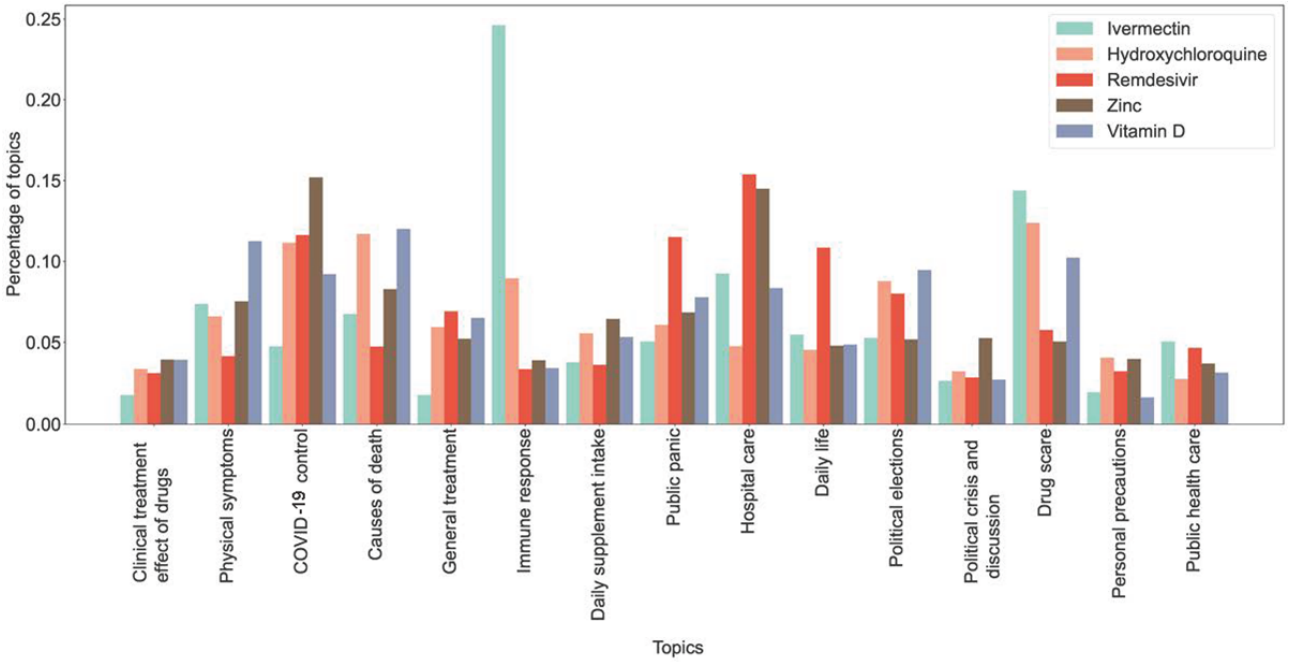

Co-Occurrence Networks

We visualized the co-occurrence network for drug-drug and drug-symptom relations in Figure 5. The nodes represented either drugs (as shown in Table S3 in Multimedia Appendix 1) or symptoms (as shown in Table S4 in Multimedia Appendix 1). The size of each node corresponded to its degree, which referred to the number of connections it has. The weights of the edges indicated the cosine similarity score between two connected nodes.

XSL•FO
RenderX

Figure 5. Visualization of drug-related co-occurrence networks by Gephi, including (A) drug-drug associations based on Gephi clustering (τ=0.005), (B) drug-drug associations based on ATC (τ=0.005), and (C) drug-symptom associations (τ=0.05). The color dots on the lower right of the figure represent the ATC categories for (B). ATC: Anatomical Therapeutic Chemical classification system.

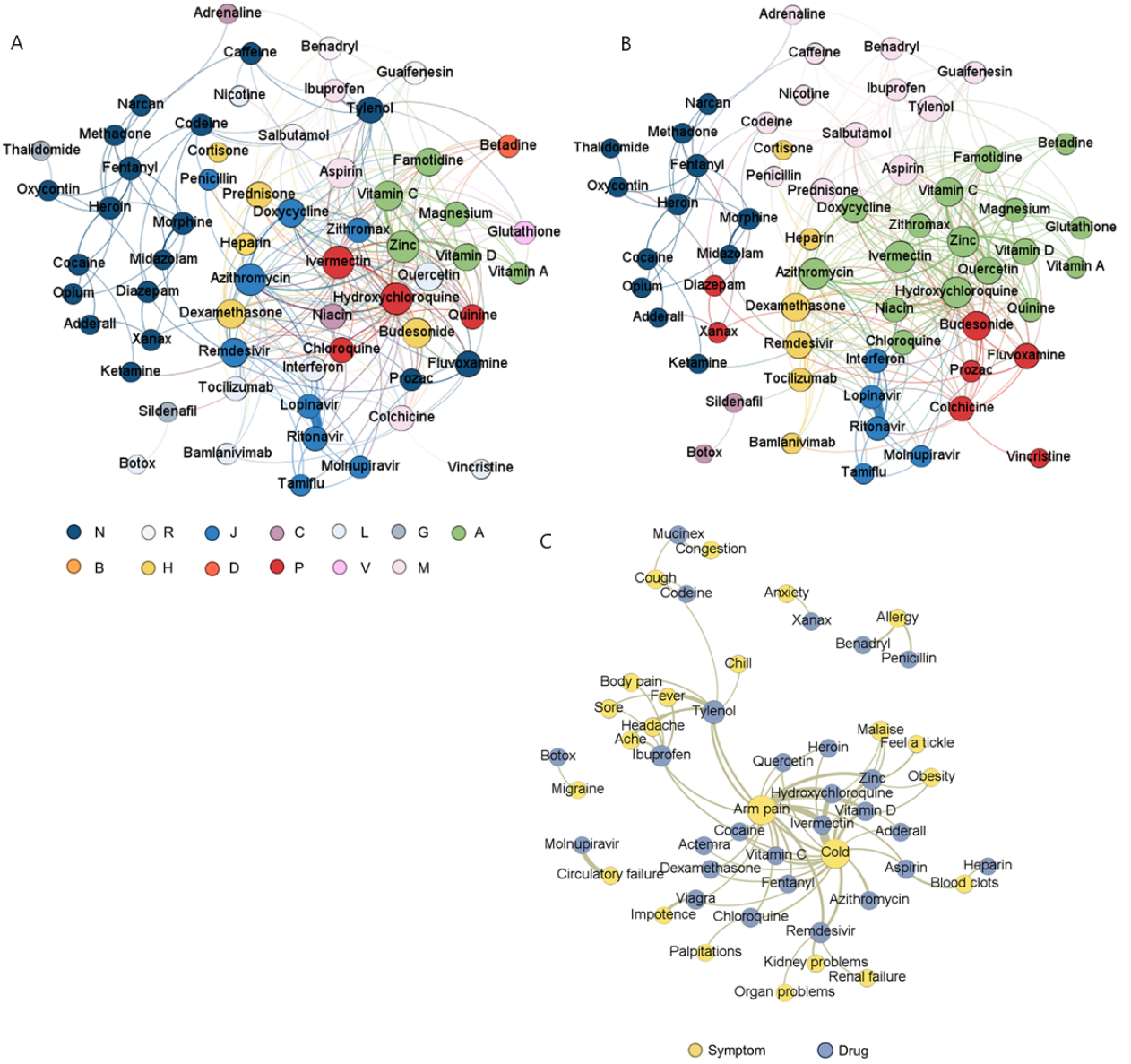

Drug-Drug Network

The origin drug-drug network contained 67 drugs (nodes) with more than 1000 mentions and 1103 relations (edges) among them. A predefined similarity threshold (τ) was established to only visualize relationships with substantial co-occurrence, as measured by cosine similarities exceeding τ. After filtering it with a τ of 0.005, 62 drugs and 317 relations remained in the network. By using the Fast Unfolding (Louvain) algorithm built in Gephi for modularity classification [49], the drugs were clustered into 5 categories and were colored in Figure 5A. The same network with drugs colored by ATC classification (12 types) was shown in Figure 5B for comparison. Drugs in the same group are denoted with the same color. Both figures share similar clustering characteristics, especially in psychotropic drugs ATC-N (Anatomical Therapeutic Chemical classification system, psychotropic drugs; eg, fentanyl, opium, and morphine) and anti-infectious agent ATC-J (Anatomical Therapeutic Chemical classification system, anti-infectious agents; eg, lopinavir, ritonavir, and azithromycin). However, drugs in the ATC-P (Anatomical Therapeutic Chemical classification system, antiparasitic drugs) group (ie, ivermectin, hydroxychloroquine, quinine, and chloroquine) are clustered with the ATC A group in Figure 5A. The reason may partially lie in the fact that most parasites are intestinal [50], so most people who need to take antiparasitic drugs (ie, ATC-P drugs) often present concomitant digestive manifestations [51], thus necessitating the use of digestive medications (ie, ATC A drugs), therefore the 2 drug groups are closely related. Association between some of the significant drug-drug pairs like 2 HIV protease inhibitors ritonavir and lopinavir has been widely studied [52]. In addition, through the co-occurrence network, we observed several unusual drug pairings, such as midazolam and morphine, salbutamol and prednisone, and zinc and quinine. These strong

XSL•FO
RenderX

co-occurrences suggest potential unexplored synergistic effects, adverse reactions, or other public health concerns that warrant further investigation. For instance, we noted a distinct correlation between morphine and midazolam, drugs not typically combined in direct COVID-19 treatment. An analysis of all 376 tweets mentioning both drugs revealed that most discussions focused on end-of-life management for patients with COVID-19 and on conspiracy theories about the intentional misuse of these drugs, leading to deaths attributed to causes other than COVID-19 infection.

Drug-Symptom Network

The original drug-symptom network had 136 nodes (ie, 69 drugs and 67 symptoms) and 3099 edges. After filtering by τ of 0.05, 50 nodes and 71 edges remained and are shown in Figure 5C. We observed that the edges often represented symptoms and corresponding treatments, such as Tylenol for fever medication, suggesting the reliability of our association network. We also observed some side effect relations, such as remdesivir to acute kidney failure [46] and some novel associations receiving no clinical investigation like molnupiravir to circulatory failure, cocaine to chest cold, and vitamin D to malaise. We visualized the top 10 closest drugs and symptoms with co-occurrence relationships to the 5 drugs under investigation (Figure S4 in Multimedia Appendix 1). These networks revealed the great relevance between hydroxychloroquine, ivermectin, and azithromycin from each other. Furthermore, remdesivir was also significantly associated with dexamethasone and tocilizumab.

Discussion

Principal Results

Leveraging new advances in NLP, we constructed a pretrained language model driven by the drug entity recognition model and a new targeted sentiment analysis model for the polarity prediction of target drugs. Based on over 2 years of relevant data, our comprehensive NLP pipeline demonstrates advanced accuracy and completeness in collecting and analyzing data for social media-based drug studies. Our NER model identified the top 5 most-discussed drugs and sentiment and topic analysis revealed that public perception concerning these drugs was predominantly shaped by celebrity endorsements, media hot spots, and governmental directives rather than empirical evidence of drug efficacy. Furthermore, network analysis identified emerging patterns of DDI and ADR (ie, molnupiravir to circulatory failure) that could be critical for public health surveillance like better safeguarding public safety in medicines use. Our pipeline is open-sourced and it can serve as a comprehensive tool to enhance drug safety control, provide crucial guidance for formulating drug usage policies, and support public health decision-making after the outbreak of infectious diseases.

Compared with traditional pharmacovigilance research, the study of drug-related information on social media exhibits distinctive characteristics and advantages. Social media platforms offer real-time and immediate data, enabling the rapid reflection of drug usage patterns and patient feedback, facilitating the prompt identification of potential risks and benefits [53-55]. Furthermore, social media captures the viewpoints and experiences of patients, thus furnishing critical insights for the formulation of patient-centered care [56,57]. For example, understanding patient's preference for drugs and disease burden can improve drug development strategies, enabling pharmaceutical companies to better focus on specific drugs that meet patient needs and preferences [58]. In contrast to previous COVID-19 social media studies [59-62], this work extracted more rigorous data covering a more extended study period and identified the five most discussed drugs to be investigated through a fully data-driven method. The substantial volume of social media data allows for large-scale real-time dynamic analysis, and it also covers a broader population than electronic health records, which are confined to hospitalized individuals and have restricted access [63]. Social media datasets could also provide large-scale samples for the detection of rare events and the examination of specific population responses, which are challenges in electronic health records–based analysis.

Sentiment analysis on drugs can highlight patient misconceptions and disagreements about a specific medication, enabling pharmaceutical companies and public health agencies to address public anxiety and reduce misinformation about drugs. Our results confirmed findings from Hua et al [59] that the public concern and polarity for ivermectin and hydroxychloroquine, which received the most social media attention, are highly correlated with emotional and political factors, such as personal political orientation, presidential elections, and conspiracy theories. For instance, there was a surge of approximately 200% in acquisitions of medication alternatives such as hydroxychloroquine within 2 days after the press briefing conducted by Donald Trump on March 19, 2020 [3]. The topic distribution indicated possible effects or side effects of ivermectin on the immune system and the wide in-hospital treatment use of remdesivir, but the sentiment analysis showed most opposing stances toward remdesivir which climbed significantly as the crisis unfolded. It was due to shortages, emergency needs, inefficiency [64], and potential side effects of remdesivir like bradycardia [65], and increased risk of hepatic, renal, and cardiovascular reactions [66,67]. Some people even hyped up on Twitter that remdesivir was approved solely for the purposes of reaping big profits for Anthony Fauci and the democidal cabal that he fronts, bilking the taxpayers of billions, and all while quietly euthanizing an unwitting public. Furthermore, we also found that daily supplements like zinc and vitamin D did not attract much public attention, but their immune-enhancing properties make them significantly more commended by the public than the other three drugs, especially remdesivir.

Analyzing social media data helps identify patterns of drug abuse, adverse reactions, and epidemics, thereby improving health policy planning and resource allocation to address emerging challenges [68,69]. For example, social media plays a pivotal role in addressing drug-related outbreaks and trends, enabling policy makers to respond swiftly and enhance public safety. Its interactive nature fosters direct engagement with the public, allowing policy makers to better understand community needs and concerns. Since public trust in policy makers is critical, for instance, the successful promotion of drugs and

XSL•FO
RenderX

vaccines relies heavily on public confidence [70,71], tracking public sentiment through social media in real time enables policy makers to align policies with public attitudes, so as to increase their acceptance and effectiveness. In addition, this approach helps in combating misinformation about drugs and vaccines [72]. For public health agencies, timely monitoring of drug-related concerns on social media is especially crucial when managing new drug candidates during pandemics. Specifically, our pipeline allows for real-time monitoring of public opinion on social media, which can be an important tool for public health agencies and organizations to implement clear communication plans, physical and mental health interventions, and a coordinated emergency response [73,74]. It could also help conduct rapid and dynamic screening of special populations [60] during public health emergencies, enable targeted communication [75], and combat public health misinformation [72,76].

Our work found that Twitter discussion topics of drugs during the COVID-19 pandemic were consistent with relevant studies focusing on non-drug COVID-19–related topics [25,77,78]. Similar to them, this study uncovered public concerns about "public health measures" and "treatment and recovery." In addition, by focusing on drugs, we discovered new drug-specific concerns, such as "drug panic" and "immune response." The focus on "drug panic" may reflect societal uncertainty and anxiety about drug use during the epidemic. Understanding these anxieties can be instrumental in enabling mental health professionals and policy makers to take measures to support mental health and implement interventions to alleviate anxiety. Care about the "immune response" may be indicative of public concerns about the immune system, including vaccines and immunotherapies. This can help health agencies better communicate information about vaccinations and immunization support to increase public awareness of immunization.

Many previous studies aimed to detect potential DDI and ADR from social media [79-81] or online literature [82] but largely depended on external vocabulary for keyword-matching and little visualization was performed. This study used advanced pretrained language models to identify drug mentions and classify the corresponding sentiments from social media text, ensuring the accuracy of information extraction and sentiment prediction. As the pretrained language model is the main NLP structure in our pipeline, it can be easily extended by integrating better large language models (LLMs) [83-85] that have a similar deep learning network structure but with larger parameters, given enough computational resources. The visualization module could illustrate associations between drugs, drug-symptoms pairs, and possible clusters or patterns intuitively and clearly, making it easier for researchers to understand and interpret the findings for DDI and ADR [86]. In addition, our co-occurrence network analysis found many widely studied drug-drug and drug-symptom pairs which could verify the reliability of network analysis. The clustering results are consistent with the classification of the general clinic (ie, ATC) to a certain extent, such as the similar clustering characteristics in psychotropic drugs (ie, ATC N) and anti-infectious agent (ie, ATC J), suggesting its potential to capture similarities and associations between drugs. Notably, we also found many drug pairs with not widely examined associations, such as zinc and quercetin. Their complex (Q/Zn) is considered a potential new drug therapy for improving glycemic control and pulmonary dysfunction in diabetes mellitus [87], which needs to be further investigated. We found new drug-related associations, such as rheumatoid drugs (hydroxychloroquine, dexamethasone, etc.) may affect COVID-19 treatment due to drug repositioning. For these novel drug-drug and drug-symptom pairs, researchers interested in further exploration may undertake additional studies, such as cross-study analyses using multiple data sources or more detailed quantitative studies. We expect studies could examine the novel associations and provide more robust evidence in future work. Furthermore, networks of the top 5 drugs revealed the significant associations between them such as the co-medication of ivermectin, hydroxychloroquine, and azithromycin for COVID-19 infection. Our network analysis also indicated the combination of remdesivir and tocilizumab or dexamethasone, and a randomized controlled trial showed their efficacy for the treatment of severe COVID-19 infection [88-90].

In essence, the use of NLP techniques and network analysis in our pipeline to analyze vast amounts of social media data is an emerging research approach in pharmacovigilance [91,92]. It holds immense potential in various areas such as the monitoring of ADR, the analysis of drug usage trends, the prediction of epidemics, and the evaluation of drug treatment effects. This novel method could serve pharmaceutical firms, regulatory agencies, and the health care fields with more precise and timely information to enhance their efforts in safeguarding public health.

Limitations

Certain limitations apply to this study. First, social media users can't represent the general population. For example, Twitter users in the United States are younger, more democratic in their political affiliations, and the most prolific 10% of users create 80% of tweets [93] and older people with lower socioeconomic status may have limited access to social media [94], which may result in bias of our observations. The specific events, geographical contexts, and the dynamic nature of social media usage may also influence the observations. Second, although we tried to automate the information extraction with deep learning, we still relied on an empirical lexicon to cluster different concept representations. This allowed us to effectively reduce false positives but not to avoid false negatives. Third, manual checks for symptom recognition suggested that approximately 2%-3% of the tweets may still be false positive (eg, lexical ambiguity like an American fever dream), which would lead to fake associations, despite the combination of rigorous rules and advanced NLP models based on deep learning. Data accuracy, as well as the reliability of the network analysis, are also limited by the authenticity of social media data and the influence of noisy information and misinformation. However, the primary advantage of social media information is its vast scale and timeliness, which offers opportunities for advancing valuable research directions, such as identifying novel drug interactions. Finally, due to the relatively low accuracy of the TSA module (ie, 75.07%), future work should

XSL•FO
RenderX

develop more effective NLP solutions to facilitate opinion mining.

Comparison With Previous Works

Hua et al [59] used BERT models to examine public perceptions of approved and off-label medications for COVID-19 infection and found these perceptions to be heavily skewed by misinformation and biases. However, the study suffered from methodological limitations, including a narrow and subjectively chosen selection of drugs, manual lexicon-based extraction, and a small time span. Similarly, Wu et al [38] made the very first attempt to construct co-occurrence networks to study symptoms during COVID-19 infection, but their technique was solely based on lexicon matching. Both of them relied solely on lexicons for extraction and, as a result, suffered from insufficient accuracy and a lack of generalizability. In contrast, this study combined advanced deep learning models with lexicon-match to improve the accuracy of entity recognition and sentiment analysis, creating a comprehensive and generalized pipeline to streamline information tracking in public health emergencies.

Conclusion

Our study proposed a pipeline of using social media data and NLP techniques to mine potential drug information, timely track drug-related hot events, facilitate public health stakeholders to conduct reasonable policy enactment, monitor drug public opinion, and avoid malignant events during a public health emergency period. In addition, it can supplement the existing ADR and DDI databases by constructing multiple medical entity co-occurrence networks to provide real-world clues for future research. Our framework applies not only to the COVID-19 pandemic but also to other periods of epidemics or major social events. It can also target other public health care foci such as vaccination.

Data Availability

Due to the privacy restrictions of Twitter, only tweet IDs can be released. Tweet IDs can be obtained from https://github.com/echen102/COVID-19-TweetIDs. The source code and pipeline tutorial of this paper are available at https://github.com/zju-liwanxin/covid-twitter-drug. Datasets and models for NER and TSA models are publicly available at https://github.com/YLab-Open/METS-CoV. All codes are based on Python software (version 3.8) and the NER, and TSA models are developed by PyTorch (version 1.0).

Authors' Contributions

JY, WL, and XX designed the study. WL and JY drafted the manuscript. JY is the senior author. YH collected the data and drafted and revised the manuscript. WL performed data and statistical analysis. PZ built the NER and TSA models. LZ provided critical reviews. All authors reviewed the manuscript. WL takes responsibility for the integrity of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional materials.
[DOCX File , 3509 KB - jmir_v27i1e63755_app1.docx]

References

1. Park HW, Park S, Chong M. Conversations and medical news frames on twitter: infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
2. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
3. Niburski K, Niburski O. Impact of trump's promotion of unproven COVID-19 treatments and subsequent internet trends: observational Study. J Med Internet Res 2020;22(11):e20044 [FREE Full text] [doi: 10.2196/20044] [Medline: 33151895]
4. FDA adverse event reporting system (FAERS). U.S. Food and Drug Administration. URL: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ [accessed 2025-02-12]
5. Correia R, Li L, Rocha L. Monitoring potential drug interactions and reactions via network analysis of instagram user timelines. Biocomputing 2016: World Scientific Publishing Company; 2015 Presented at: Proceedings of the Pacific Symposium; 2024 January 3-7; Big Island of Hawaii p. 492-503. [doi: 10.1142/9789814749411_0045]
6. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther 2012;91(6):1010-1021 [FREE Full text] [doi: 10.1038/clpt.2012.50] [Medline: 22549283]
7. Stephenson WP, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. Pharmacoepidemiol Drug Saf 2007;16(4):359-365. [doi: 10.1002/pds.1323] [Medline: 17019675]

XSL•FO
RenderX

8. Bate A, Hornbuckle K, Juhaeri J, Motsko SP, Reynolds RF. Hypothesis-free signal detection in healthcare databases: finding its value for pharmacovigilance. Ther Adv Drug Saf 2019;10:2042098619864744 [FREE Full text] [doi: 10.1177/2042098619864744] [Medline: 31428307]
9. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf 2009;18(6):427-436. [doi: 10.1002/pds.1742] [Medline: 19358225]
10. Rawlins MD. Spontaneous reporting of adverse drug reactions. I: the data. Br J Clin Pharmacol 1988;26(1):1-5 [FREE Full text] [doi: 10.1111/j.1365-2125.1988.tb03356.x] [Medline: 3203052]
11. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
12. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019;26(6):561-576 [FREE Full text] [doi: 10.1093/jamia/ocz009] [Medline: 30908576]
13. Rees S, Mian S, Grabowski N. Using social media in safety signal management: is it reliable? Ther Adv Drug Saf 2018;9(10):591-599 [FREE Full text] [doi: 10.1177/2042098618789596] [Medline: 30283627]
14. Mekawie N, Hany A. Understanding the factors driving consumers' purchase intention of over the counter medications using social media advertising In Egypt. Procedia Computer Science 2019;164:698-705. [doi: 10.1016/j.procs.2019.12.238]
15. Alshareef M, Alotiby A. Prevalence and perception among Saudi Arabian population about resharing of information on social media regarding natural remedies as protective measures against COVID-19. Int J Gen Med 2021;14:5127-5137 [FREE Full text] [doi: 10.2147/IJGM.S326767] [Medline: 34511995]
16. Lazard AJ. Social media message designs to educate adolescents about E-cigarettes. J Adolesc Health 2021;68(1):130-137 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.030] [Medline: 32654836]
17. Wu J, Wu X, Hua Y, Lin S, Zheng Y, Yang J. Exploring social media for early detection of depression in COVID-19 patients. : Association for Computing Machinery; 2023 Presented at: WWW '23: The ACM Web Conference 2023; 2023 April 30; Austin TX USA p. 3968-3977. [doi: 10.1145/3543507.3583867]
18. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. : Association for Computational Linguistics; 2011 Presented at: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing; July, 2011; Edinburgh, Scotland, United Kingdom p. 1568-1576 URL: https://aclanthology.org/D11-1145/ [doi: 10.1080/07448481.2021.1973480]
19. Nishiyama T, Yada S, Wakamiya S, Hori S, Aramaki E. Transferability based on drug structure similarity in the automatic classification of noncompliant drug use on social media: natural language processing approach. J Med Internet Res 2023;25:e44870 [FREE Full text] [doi: 10.2196/44870] [Medline: 37133915]
20. Helgeson SA, Mudgalkar RM, Jacobs KA, Lee AS, Sanghavi D, Moreno Franco P, National COVID Cohort Collaborative (N3C). Association between X/Twitter and prescribing behavior during the COVID-19 pandemic: retrospective ecological study. JMIR Infodemiology 2024;4:e56675 [FREE Full text] [doi: 10.2196/56675] [Medline: 39556417]
21. Jiang H, Hua Y, Beeferman D, Roy D. Annotating the tweebank corpus on named entity recognition and building NLP models for social media analysis. arXiv:2201.07281 2022 [FREE Full text]
22. Zhou P, Wang Z, Chong D, Guo Z, Hua Y, Su Z. METS-CoV: a dataset of medical entity and targeted sentiment on COVID-19 related tweets. : Curran Associates Inc; 2022 Presented at: NIPS'22: 36th International Conference on Neural Information Processing Systems; 2022 November 28; New Orleans LA USA p. 21916-21932.
23. Satu MS, Khan MI, Mahmud M, Uddin S, Summers MA, Quinn JM, et al. TClustVID: a novel machine learning classification model to investigate topics and sentiment in COVID-19 tweets. Knowl Based Syst 2021;226:107126 [FREE Full text] [doi: 10.1016/j.knosys.2021.107126] [Medline: 33972817]
24. Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online siscussions: NLP using LSTM recurrent neural network approach. IEEE J Biomed Health Inform 2020;24(10):2733-2742. [doi: 10.1109/JBHI.2020.3001216] [Medline: 32750931]
25. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. J Med Internet Res 2020;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
26. Babić K, Petrović M, Beliga S, Martinčić-Ipšić S, Matešić M, Meštrović A. Characterisation of COVID-19-related tweets in the croatian language: framework based on the Cro-CoV-cseBERT model. Applied Sciences 2021;11(21):10442. [doi: 10.3390/app112110442]
27. de Melo T, Figueiredo CMS. Comparing news articles and tweets about COVID-19 in Brazil: sentiment analysis and topic modeling approach. JMIR Public Health Surveill 2021;7(2):e24585 [FREE Full text] [doi: 10.2196/24585] [Medline: 33480853]
28. Beliga S, Martinčić-Ipšić S, Matešić M, Petrijevčanin Vuksanović I, Meštrović A. Infoveillance of the croatian online media during the COVID-19 pandemic: one-year longitudinal study using natural language processing. JMIR Public Health Surveill 2021;7(12):e31540 [FREE Full text] [doi: 10.2196/31540] [Medline: 34739388]

XSL•FO
RenderX

29. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus twitter data set. JMIR Public Health Surveill 2020;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
30. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. : Association for Computational Linguistics; 2018 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); 2025 February 09; Minneapolis, Minnesota p. 181004805.
31. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment embeddings with applications to sentiment analysis. IEEE Trans. Knowl. Data Eng 2016;28(2):496-509. [doi: 10.1109/tkde.2015.2489653]
32. Xu P, Madotto A, Wu C, Park J. Emo2vec: Learning generalized emotion representation by multi-task training. 2018 Presented at: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; 2025 February 09; Brussels, Belgium p. 292-298. [doi: 10.18653/v1/w18-6243]
33. Yu LC, Wang J, Lai KR, Zhang X. Refining word embeddings using intensity scores for sentiment analysis. IEEE/ACM Trans. Audio Speech Lang. Process 2018;26(3):671-681. [doi: 10.1109/taslp.2017.2788182]
34. Wang J, Yu LC, Zhang X. SoftMCL: soft momentum contrastive learning for fine-grained sentiment-aware pre-training. : ELRA and ICCL; 2024 Presented at: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024); 2025 February 09; Torino, Italia p. 15012-15023.
35. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. 2010 Presented at: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks; 2010 July 25; Malta.
36. Guo Y, Zhang Y, Lyu T, Prosperi M, Wang F, Xu H, et al. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. J Am Med Inform Assoc 2021;28(9):2050-2067 [FREE Full text] [doi: 10.1093/jamia/ocab098] [Medline: 34151987]
37. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. 2012 Presented at: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning; 2012 July 12-14; Jeju Island Korea p. 952-961.
38. Wu J, Wang L, Hua Y, Li M, Zhou L, Bates DW, et al. Trend and Co-occurrence Network of COVID-19 symptoms from large-scale social media data: infoveillance study. J Med Internet Res 2023;25:e45419 [FREE Full text] [doi: 10.2196/45419] [Medline: 36812402]
39. About. Gephi. URL: https://gephi.org/about/ [accessed 2025-02-12]
40. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS One 2014;9(6):e98679 [FREE Full text] [doi: 10.1371/journal.pone.0098679] [Medline: 24914678]
41. ATC/DDD index 2025. Norwegian Institute of Public Health. URL: https://www.whocc.no/atc_ddd_index/ [accessed 2025-02-12]
42. Hua Y, Wu J, Lin S, Li M, Zhang Y, Foer D, et al. Streamlining social media information retrieval for public health research with deep learning. J Am Med Inform Assoc 2024;31(7):1569-1577. [doi: 10.1093/jamia/ocae118] [Medline: 38718216]
43. Rahutomo F, Kitasuka T, Aritsugi M. Semantic cosine similarity. 2012 Presented at: The 7th International Student Conference on Advanced Science and Technology ICAST 2012; 2012 October 29-30; Seoul, South Korea.
44. Xia P, Zhang L, Li F. Learning similarity with cosine similarity ensemble. Information Sciences 2015;307:39-52 [FREE Full text] [doi: 10.1016/j.ins.2015.02.024]
45. Ansems K, Grundeis F, Dahms K, Mikolajewska A, Thieme V, Piechotta V, et al. Remdesivir for the treatment of COVID-19. Cochrane Database Syst Rev 2021;8(8):CD014962 [FREE Full text] [doi: 10.1002/14651858.CD014962] [Medline: 34350582]
46. Rahimi MM, Jahantabi E, Lotfi B, Forouzesh M, Valizadeh R, Farshid S. Renal and liver injury following the treatment of COVID-19 by remdesivir. J Nephropathol 2020;10(2):1-4. [doi: 10.34172/jnp.2021.10]
47. Guerra M, Mendoza C, Kandhi S, Sun H, Saad M, Vittorio T. Cardiac arrhythmia related to remdesivir in COVID-19. ISMMS Journal of Science and Medicine 2021;1(2):15. [doi: 10.29024/ijsm.64]
48. Sneij E, Kohli V, Al-Adwan SA, Mealor A. Remdesivir causing profound bradycardia. Journal of the American College of Cardiology 2021;77(18):2037. [doi: 10.1016/s0735-1097(21)03393-3]
49. Blondel V, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J. Stat. Mech 2008;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/p10008]
50. Naveed A, Abdullah S. Impact of parasitic infection on human gut ecology and immune regulations. transl med commun 2021;6(1). [doi: 10.1186/s41231-021-00091-4]
51. Sey ICM, Ehimiyein AM, Bottomley C, Riley EM, Mooney JP. Does malaria cause diarrhoea? A systematic review. Front Med (Lausanne) 2020;7:589379 [FREE Full text] [doi: 10.3389/fmed.2020.589379] [Medline: 33330549]
52. Cvetkovic RS, Goa KL. Lopinavir/ritonavir: a review of its use in the management of HIV infection. Drugs 2003;63(8):769-802. [doi: 10.2165/00003495-200363080-00004] [Medline: 12662125]

XSL•FO
RenderX

53. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
54. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
55. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: a systematic review. J Biomed Inform 2020;108:103500 [FREE Full text] [doi: 10.1016/j.jbi.2020.103500] [Medline: 32622833]
56. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med 2019;17(1):11 [FREE Full text] [doi: 10.1186/s12916-018-1247-8] [Medline: 30646913]
57. Kudchadkar SR, Carroll CL. Using social media for rapid information dissemination in a pandemic: #PedsICU and coronavirus disease 2019. Pediatr Crit Care Med 2020;21(8):e538-e546 [FREE Full text] [doi: 10.1097/PCC.0000000000002474] [Medline: 32459792]
58. Schmidt AL, Rodriguez-Esteban R, Gottowik J, Leddin M. Applications of quantitative social media listening to patient-centric drug development. Drug Discov Today 2022;27(5):1523-1530. [doi: 10.1016/j.drudis.2022.01.015] [Medline: 35114364]
59. Hua Y, Jiang H, Lin S, Yang J, Plasek JM, Bates DW, et al. Using twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. J Am Med Inform Assoc 2022;29(10):1668-1678 [FREE Full text] [doi: 10.1093/jamia/ocac114] [Medline: 35775946]
60. Li M, Hua Y, Liao Y, Zhou L, Li X, Wang L, et al. Tracking the impact of COVID-19 and lockdown policies on public mental health using social media: infoveillance study. J Med Internet Res 2022;24(10):e39676 [FREE Full text] [doi: 10.2196/39676] [Medline: 36191167]
61. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on twitter: topic modeling and sentiment analysis. J Med Internet Res 2021;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
62. Al-Ramahi M, Elnoshokaty A, El-Gayar O, Nasralah T, Wahbeh A. Public discourse against masks in the COVID-19 era: infodemiology study of twitter data. JMIR Public Health Surveill 2021;7(4):e26780 [FREE Full text] [doi: 10.2196/26780] [Medline: 33720841]
63. Wu J, Liu X, Li M, Li W, Su Z, Lin S, et al. Clinical text datasets for medical artificial intelligence and large language models — a systematic review. NEJM AI 2024;1(6). [doi: 10.1056/aira2400012]
64. Hagman K, Hedenstierna M, Widaeus J, Arvidsson E, Hammas B, Grillner L, et al. Effects of remdesivir on SARS-CoV-2 viral dynamics and mortality in viraemic patients hospitalized for COVID-19. J Antimicrob Chemother 2023;78(11):2735-2742 [FREE Full text] [doi: 10.1093/jac/dkad295] [Medline: 37757451]
65. Ishisaka Y, Aikawa T, Malik A, Kampaktsis PN, Briasoulis A, Kuno T. Association of remdesivir use with bradycardia: a systematic review and meta-analysis. J Med Virol 2023;95(8):e29018. [doi: 10.1002/jmv.29018] [Medline: 37539782]
66. Blair HA. Remdesivir: a review in COVID-19. Drugs 2023;83(13):1215-1237 [FREE Full text] [doi: 10.1007/s40265-023-01926-0] [Medline: 37589788]
67. Akhvlediani T, Bernard-Valnet R, Dias SP, Eikeland R, Pfausler B, Sellner J, Infectious Disease Panel of the European Academy of Neurology. Neurological side effects and drug interactions of antiviral compounds against SARS-CoV-2. Eur J Neurol 2023;30(12):3904-3912. [doi: 10.1111/ene.16017] [Medline: 37526048]
68. Lane JM, Habib D, Curtis B. Linguistic methodologies to surveil the leading causes of mortality: scoping review of twitter for public health data. J Med Internet Res 2023;25:e39484 [FREE Full text] [doi: 10.2196/39484] [Medline: 37307062]
69. van Stekelenborg J, Ellenius J, Maskell S, Bergvall T, Caster O, Dasgupta N, et al. Recommendations for the use of social media in pharmacovigilance: lessons from IMI WEB-RADR. Drug Saf 2019;42(12):1393-1407 [FREE Full text] [doi: 10.1007/s40264-019-00858-7] [Medline: 31446567]
70. Rand LZ, Carpenter DP, Kesselheim AS, Bhaskar A, Darrow JJ, Feldman WB. Securing the trustworthiness of the FDA to build public trust in vaccines. Hastings Cent Rep 2023;53 Suppl 2:S60-S68. [doi: 10.1002/hast.1525] [Medline: 37963051]
71. Saechang O, Yu J, Li Y. Public trust and policy compliance during the COVID-19 pandemic: the role of professional trust. Healthcare (Basel) 2021;9(2):151 [FREE Full text] [doi: 10.3390/healthcare9020151] [Medline: 33540762]
72. Pagoto S, Waring ME, Xu R. A call for a public health agenda for social media research. J Med Internet Res 2019;21(12):e16661 [FREE Full text] [doi: 10.2196/16661] [Medline: 31855185]
73. Terry K, Yang F, Yao Q, Liu C. The role of social media in public health crises caused by infectious disease: a scoping review. BMJ Glob Health 2023;8(12):e013515 [FREE Full text] [doi: 10.1136/bmjgh-2023-013515] [Medline: 38154810]
74. Yoo S, Kim D, Yang S, Jeong O. Real-time disease detection and analysis system using social media contents. IJWGS 2020;16(1):22-38. [doi: 10.1504/ijwgs.2020.106103]
75. Murthy BP, Krishna N, Jones T, Wolkin A, Avchen RN, Vagi SJ. Public health emergency risk communication and social media reactions to an errant warning of a ballistic missile threat - Hawaii, January 2018. MMWR Morb Mortal Wkly Rep 2019;68(7):174-176 [FREE Full text] [doi: 10.15585/mmwr.mm6807a2] [Medline: 30789877]
76. Pierri F, DeVerna MR, Yang KC, Axelrod D, Bryden J, Menczer F. One year of COVID-19 vaccine misinformation on twitter: longitudinal study. J Med Internet Res 2023;25:e42227 [FREE Full text] [doi: 10.2196/42227] [Medline: 36735835]

XSL•FO
RenderX

77. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
78. Boukobza A, Burgun A, Roudier B, Tsopra R. Deep neural networks for simultaneously capturing public topics and sentiments during a pandemic: application on a COVID-19 tweet data set. JMIR Med Inform 2022;10(5):e34306 [FREE Full text] [doi: 10.2196/34306] [Medline: 35533390]
79. Yang H, Yang CC. Harnessing social media for drug-drug interactions detection. 2013 Presented at: 2013 IEEE International Conference on Healthcare Informatics; 2013 September 9-11; Philadelphia, PA, USA p. 9-11. [doi: 10.1109/ichi.2013.10]
80. Xia L, Wang GA, Fan W. A deep learning based named entity recognition approach for adverse drug events identification and extraction in health social media. 2017 Presented at: International Conference on Smart Health; 2017 June 26-27; Hong Kong, China p. 237-248. [doi: 10.1007/978-3-319-67964-8_23]
81. Coloma PM, Becker B, Sturkenboom MCJM, van Mulligen EM, Kors JA. Evaluating social media networks in medicines safety surveillance: two case studies. Drug Saf 2015;38(10):921-930 [FREE Full text] [doi: 10.1007/s40264-015-0333-5] [Medline: 26242616]
82. Lu Y, Shen D, Pietsch M, Nagar C, Fadli Z, Huang H, et al. A novel algorithm for analyzing drug-drug interactions from MEDLINE literature. Sci Rep 2015;5:17357 [FREE Full text] [doi: 10.1038/srep17357] [Medline: 26612138]
83. Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. 2020 Presented at: NIPS'20: 34th International Conference on Neural Information Processing Systems; 2020 December 6 - 12; Vancouver BC Canada p. 1877-1901.
84. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
85. Wu J, Wu X, Qiu Z, Li M, Lin S, Zhang Y, et al. Large language models leverage external knowledge to extend clinical insight beyond language boundaries. J Am Med Inform Assoc 2024;31(9):2054-2064. [doi: 10.1093/jamia/ocae079] [Medline: 38684792]
86. Wang R, Li S, Cheng L, Wong MH, Leung KS. Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning. BMC Bioinformatics 2019;20(Suppl 26):628 [FREE Full text] [doi: 10.1186/s12859-019-3283-6] [Medline: 31839008]
87. Refat MS, Hamza RZ, Adam AMA, Saad HA, Gobouri AA, Al-Harbi FS, et al. Quercetin/Zinc complex and stem cells: a new drug therapy to ameliorate glycometabolic control and pulmonary dysfunction in diabetes mellitus: structural characterization and genetic studies. PLoS One 2021;16(3):e0246265 [FREE Full text] [doi: 10.1371/journal.pone.0246265] [Medline: 33661932]
88. Mohiuddin Chowdhury ATM, Kamal A, Abbas KU, Talukder S, Karim MR, Ali MA, et al. Efficacy and outcome of remdesivir and tocilizumab combination against dexamethasone for the treatment of severe COVID-19: a randomized controlled trial. Front Pharmacol 2022;13:690726 [FREE Full text] [doi: 10.3389/fphar.2022.690726] [Medline: 35450050]
89. Marrone A, Nevola R, Sellitto A, Cozzolino D, Romano C, Cuomo G, et al. Remdesivir plus dexamethasone versus dexamethasone alone for the treatment of coronavirus disease 2019 (COVID-19) patients requiring supplemental O2 therapy: a prospective controlled nonrandomized study. Clin Infect Dis 2022;75(1):e403-e409 [FREE Full text] [doi: 10.1093/cid/ciac014] [Medline: 35084022]
90. Gressens SB, Esnault V, De Castro N, Sellier P, Sene D, Chantelot L, Saint-Louis CORE group. Remdesivir in combination with dexamethasone for patients hospitalized with COVID-19: a retrospective multicenter study. PLoS One 2022;17(2):e0262564 [FREE Full text] [doi: 10.1371/journal.pone.0262564] [Medline: 35176057]
91. Alshammari TM, Mendi N, Alenzi KA, Alsowaida Y. Pharmacovigilance systems in arab countries: overview of 22 arab Countries. Drug Saf 2019;42(7):849-868. [doi: 10.1007/s40264-019-00807-4] [Medline: 31006085]
92. Liang L, Hu J, Sun G, Hong N, Wu G, He Y, et al. Artificial intelligence-based pharmacovigilance in the setting of limited resources. Drug Saf 2022;45(5):511-519 [FREE Full text] [doi: 10.1007/s40264-022-01170-7] [Medline: 35579814]
93. Wojcik S, Hughes A. Sizing up twitter users. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2025-02-12]
94. Wojcik S, Hughes A. Sizing up twitter users. PEW Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2025-02-12]

Abbreviations

ADR: adverse drug reactions
API: application programming interface
ATC: Anatomical Therapeutic Chemical classification system
ATC-J: Anatomical Therapeutic Chemical classification system, anti-infectious agents
ATC-N: Anatomical Therapeutic Chemical classification system, psychotropic drugs
ATC-P: Anatomical Therapeutic Chemical classification system, antiparasitic drugs
CT-BERT: COVID-Twitter-Bidirectional Encoder Representations from Transformers

XSL•FO
RenderX

CT-BERT-NER: COVID Twitter with Bidirectional Encoder Representations from Transformers for Named Entity Recognition
DDI: drug-drug interactions
FDA: Food and Drug Administration
LDA: latent Dirichlet allocation
METS-CoV: Medical Entity and Targeted Sentiment on COVID-19 Related Tweets
NER: named entity recognition
NLP: natural language processing
TSA: target sentiment analysis
WHO: World Health Organization

Edited by C Argyropoulos; submitted 28.06.24; peer-reviewed by O Oyelade, Y Zhou, SJ Park, C Duan, A Jang, N Seeman, LC Yu; comments to author 02.11.24; revised version received 19.12.24; accepted 25.01.25; published 05.03.25.

Please cite as:
Li W, Hua Y, Zhou P, Zhou L, Xu X, Yang J
Characterizing Public Sentiments and Drug Interactions in the COVID-19 Pandemic Using Social Media: Natural Language Processing and Network Analysis
J Med Internet Res 2025;27:e63755
URL: https://www.jmir.org/2025/1/e63755
doi:10.2196/63755
PMID:40053730

©Wanxin Li, Yining Hua, Peilin Zhou, Li Zhou, Xin Xu, Jie Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health IT Implementation and the Impact of the COVID-19 Pandemic on Clinician-IT Dynamics: Qualitative Study

Adeola Bamgboje-Ayodele[1], BTech, MSc, PhD; Adrian Boscolo[2], MBBS, MPH; Mitchell Burger[2], BCom, MPH; Owen Hutchings[2], MBBS, BSc; Miranda Shaw[2], MHA; Tim Shaw[1], PhD; Amina Tariq[3], PhD; Sundresan Naicker[3], PhD; Steven McPhail[3,4], PhD; Melissa Baysari[1], PhD

[1]Biomedical Informatics and Digital Health, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia
[2]Sydney Local Health District, Sydney, Australia
[3]Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, Queensland University of Technology, Brisbane, Australia
[4]Digital Health and Informatics Directorate, Metro South Health, Brisbane, Australia

Corresponding Author:
Adeola Bamgboje-Ayodele, BTech, MSc, PhD
Biomedical Informatics and Digital Health, School of Medical Sciences
Faculty of Medicine and Health
University of Sydney
RC Mills Building
Camperdown, 2006
Australia
Phone: 61 02 9351 9644
Email: adeola.ba@sydney.edu.au

Abstract

Background: The COVID-19 pandemic necessitated the rapid development and implementation of health ITs to support health care delivery. Health IT implementation is difficult at the best of times, due to complex sociotechnical challenges that vary across contexts and settings; however, it is currently unclear how the pandemic impacted health IT implementation processes. The aim of this study was to explore the impact of the pandemic on health IT implementation processes, including pre- and postimplementation phases, and identify the sociotechnical factors that shaped health IT implementation during an unprecedented circumstance.

Objective: This study aimed to explore the impact of the pandemic on HIT implementation processes, including pre- and postimplementation phases, and identify the socio-technical factors that shaped health IT implementation during an unprecedented circumstance.

Methods: Participants were from one of two teams: (1) health care staff members (doctors, nurses, nurse unit managers, and support staff members) from a virtual hospital in Australia; and (2) IT professionals within the broader health care organization assigned to the hospital. Participants took part in an interview or focus group from July to November 2022. Participants were asked to describe the process used for rapid health IT design and implementation during the COVID-19 pandemic. Qualitative data were analyzed thematically.

Results: A total of 15 participants took part in the study. Both internal and external team structures, and the communication pathways that underpinned these, were reported to influence the health IT lifecycle, which in turn impacted outcomes, particularly when perceived normal ways of working were challenged during the pandemic. Across the pre-post lifecycle, preimplementation processes were viewed to be most impacted by the COVID-19 pandemic. Participants reported that their roles and responsibilities changed during health IT implementations in the pandemic, impacting co-design processes and highlighting the need for health IT implementation processes to cater for new work and the redistribution of existing work.

Conclusions: Our study uncovered the negative impact of the COVID-19 pandemic on team structures, communication pathways, and health IT preimplementation processes (project management and co-design). While health care organizations are keen to transition beyond the ways of working during the pandemic, it is imperative to learn from the health IT implementation successes and failures that occurred in the pandemic via process evaluations. Our evaluation offers learnings for research (an adapted interdisciplinary team communication framework), practice (the need for health care organizations to review their communication

XSL•FO
RenderX

structures, IT staff skills, and proposed processes), and education (the need for better education and training of IT professionals working in clinical settings on health concepts) on health IT implementations as the world transitions to the "new norm."

(J Med Internet Res 2025;27:e57847) doi:10.2196/57847

KEYWORDS

health IT; implementation; COVID-19 pandemic; process evaluation; sociotechnical factors; virtual hospital; COVID-19

Introduction

The implementation of health IT to improve safety, quality, and efficiency in health systems is increasing globally, more so during the COVID-19 pandemic [1-3]. However, health IT implementations are notoriously difficult due to complex sociotechnical challenges, which vary across contexts and are often hard to navigate and envisage for those managing change [4]. Despite significant efforts to identify and apply the sociotechnical factors that enable successful health IT implementations, this remains an ongoing challenge [5]. One reason for this may be that the commonly used approach to evaluate existing health ITs tends to focus on assessing each aspect of the health IT implementation, rather than using a "whole of system" approach [6]. As new health IT implementations can learn from the evaluation of existing health ITs, there is a strong need for comprehensive evaluations that transcend a disease focus or preliminary adoption, to provide in-depth insights into the various sociotechnical factors that enable successful health IT implementations.

The Technology, People, Organizations, and Macro-environmental factors (TPOM) framework [5] is one model that provides a comprehensive understanding of the socio-technical factors impacting health IT implementations using a "whole of system" approach. TPOM has been used to evaluate a large-scale health IT implementation for the National Health Service [7] before the COVID-19 pandemic, and other health IT implementations in Europe and United States during the pandemic [8]. Unlike other health IT implementation frameworks [9], the TPOM framework is neither illness-specific, nor focused on the likelihood of adoption or spread, but considers both the microcontext of use and macroenvironmental dimensions that impact all stages of implementation [5]. Therefore, we use the TPOM framework as a lens to consider the gaps in evidence regarding the impact of the pandemic on key sociotechnical factors, across the four TPOM domains, that shape the outcomes of health IT implementations.

From a technology use perspective (TPOM domain 1), diverse health systems have leveraged health ITs in response to the COVID-19 pandemic. This has included the development of new devices and applications for remote consultations to minimize transmission of infection, and mobile apps to monitor vital signs data [1]. These health IT implementations experienced mixed success and highlighted that many technology-driven challenges (eg, data sharing across systems) present in the prepandemic times were perpetuated, and potentially magnified, during the pandemic and postpandemic "new-normal" times [10]. Furthermore, these challenges may impact the pre-post-implementation lifecycle, including perpetuating interoperability challenges [10]. However, little is currently known about the impact of the pandemic on these technology-driven challenges across the pre-post implementation lifecycle.

With people at the core of health IT implementations (TPOM domain 2), effective stakeholder engagement including collaboration between clinicians and IT staff is key [11,12]. However, the divide that exists between these 2 groups, such as communication challenges and different viewpoints on system performance and usability, has been well-documented [13-16]. The health system's response to the pandemic was not limited to clinical care delivery, and virtual stakeholder engagement sessions in the context of health IT implementation became increasingly common to support rapid policy changes and social distancing guidelines [17]. As such, the use of flexible locations enabled more efficient use of clinician time in supporting the design and implementation of health ITs [10,18]. While it may be likely that greater use of virtual sessions, rather than face-to-face sessions can exacerbate the clinician IT divide, the impact of virtual engagement on health IT implementations since the pandemic has not been evaluated.

Health service organizations (TPOM domain 3), in response to the COVID-19 pandemic, increasingly aimed to become "learning health systems" intent on learning from what worked and what did not work during pandemic-driven health IT implementations, in an effort to transition to the "new-normal" [19,20]. Despite this intention, there have been limited process evaluations of health IT implementations in real-world health service organizations that span across the pre-post implementation phases, particularly in virtual hospitals. Virtual hospitals substitute in-person consultations with telephone or video consultations and often include asynchronous data collection from the patient via survey tools with or without real-time remote monitoring [21].

The macroenvironmental context (TPOM domain 4) during the COVID-19 pandemic placed unprecedented time pressure on health IT implementations, and other pressures on health care services [1,22]. Yet, the impact of the COVID-19 pandemic on the technology, people, and organizations involved in health IT implementation has not been evaluated.

This study set out to address the identified gaps in literature and knowledge across the 4 TPOM domains highlighted above by interviewing stakeholders involved in rapid implementations of health IT during the pandemic. In particular, we aimed to explore the impact of the pandemic on health IT implementation processes, including pre- and postimplementation phases, and identify the sociotechnical factors that shaped health IT implementation, to guide future successful implementations of health IT.

XSL•FO
RenderX

Methods

Study Design

This was a cross-sectional qualitative study. As a qualitative study design is appropriate for providing insights into people's experiences of a complex phenomenon or activity [23], we conducted semistructured interviews to explore participants' experiences. The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist was used to report the study (Multimedia Appendix 1).

Setting and Recruitment

This study was conducted at a virtual hospital in Australia from July to November 2022. The virtual hospital was equipped with care pods (workspaces) with videoconferencing, telephone facilities, an electronic medical record, and used remote monitoring tools such as wearables, mobile apps, and dashboards. The majority of the videoconferencing and remote monitoring tools were simultaneously implemented during the COVID-19 pandemic. Although multiple teams were involved in health IT implementation at the study site, this study focused on the interdisciplinary relationship and communication processes between clinicians and IT teams. Both teams are part of the same health service but not colocated within the same building.

Using purposive sampling, all clinical (ie, doctors, nurses, and nurse unit managers) and nonclinical staff members from the virtual hospital were invited by email to participate in interviews based on their involvement in health IT implementations, as identified by a clinician-researcher team member who works at the virtual hospital. Participants were eligible for purposive sampling if they were (1) either doctors, nurses, nurse unit managers, or nonclinical support staff members, and (2) took part in any pre or postimplementation activity in health IT implementations at the virtual hospital since the hospital's establishment in March 2020. We conducted semistructured interviews and a focus group discussion to collect qualitative data. Due to scheduling constraints and the limited time availability of participants, we conducted a focus group discussion when individual interviews were not possible.

Data Collection Procedure

The interview questions were broadly guided by the TPOM framework (Multimedia Appendix 2). Participants were asked to describe the process used for rapid health IT design and implementation during the COVID-19 pandemic and to reflect on aspects that worked well and those that did not. Except for one face-to-face interview, other interviews and the focus group were conducted online. For the focus group discussion, participants were fully informed about the data confidentiality responsibilities of both the researchers and the participants and the necessary housekeeping rules. The moderator (AB-A) asked open-ended questions throughout the session to allow participants to share information they were comfortable with and prompted those who spoke less frequently to contribute to the discussion to ensure inclusivity and diversity of voices using probing techniques, tracking of questions for completion, in line with existing research [24]. Data collection ended when no new information emerged, and saturation was achieved. All sessions were recorded and transcribed verbatim. The sessions were conducted by AB-A, a postdoctoral researcher experienced in qualitative research and health IT implementations, and she had no previous relationship with the participants.

Data Analysis and Interpretation

Deidentified transcripts were thematically analyzed independently by 2 researchers (ABA and MB), who first identified and coded the transcribed data through inductive thematic analysis, using a data-driven approach to understand participants' experiences and perceptions [25,26]. Following this, the researchers mapped the identified codes to the interdisciplinary team communication framework of Kuziemsky et al [27] for assessing interdisciplinary teams through the exploration of sociotechnical and other factors that drive teamwork. The framework by Kuziemsky et al [27] is based on Donabedian's [28] meta-concepts of structures, processes, and outcomes, which posits that structures lead to processes, which lead to outcomes [28]. The framework by Kuziemsky et al [27] was selected to provide an understanding of the complex and nonlinear structures, processes, and outcomes involved in pandemic-driven health IT implementation as experienced by the 2 interdisciplinary teams (clinicians and IT staff).

Our analytical approach included data familiarization, coding, generating initial themes, reviewing potential themes, defining and naming themes, and producing the report [29]. Specifically, the 2 researchers (ABA and MB) familiarized themselves with the transcribed data by reading it multiple times before coding independently. After initial code generation, the researchers developed themes by merging codes with a shared meaning. Following this, the researchers reviewed the candidate themes to ensure coherent patterns were formed which contributed to the overall narrative and interpretation of the dataset. Finally, to ensure consistency and improve study rigor, the researchers discussed and reported the themes in relation to the dataset and the research aim. The 2 researchers met frequently throughout data collection and analysis to compare themes. Disagreement in themes was discussed until a consensus was reached. Emerging themes were presented to some participants and key stakeholders, and feedback received informed the final refinement of the themes.

Ethical Considerations

Ethics approval was obtained from the hospital's Human Research Ethics Committee (X21-0362 & 2021/ETH11708). Written informed consent was obtained from all participants, who were informed of their right to withdraw from the study at any time. The data presented in this article have been deidentified.

Results

A total of 15 participants took part in this study, one more person was invited but was too busy to participate. Out of 8 staff at the virtual hospital, including representation from clinical and nonclinical roles, and 7 staff from the IT department (refer to Table 1 for demographic details). The mean duration of the

XSL•FO
RenderX

interviews (n=11) was 58 (range 32-93) minutes and the focus group (n=4 participants) lasted 62 minutes.

Our findings are organized according to the meta-concepts of structures, processes, and outcomes. The concepts and subconcepts that emerged for each of these 3 meta-concepts are described using quotes from the data to illustrate examples of the concepts and subconcepts.

Table 1. Demographic characteristics.

Demographic characteristic	Values
Role, n (%)	
Nurse	2 (13)
Nurse unit manager	3 (20)
Doctor	2 (13)
Nonclinical	1 (7)
IT professionals	7 (47)
Team, n (%)	
Virtual hospital staff	8 (53)
IT staff	7 (47)
Years of experience, n (%)	
0-5	2 (13)
6-10	4 (27)
11-20	6 (40)
>20	3 (20)
Sex or sex assigned at birth, n (%)	
Male	4 (27)
Female	11 (73)
Age (years), mean (range)	41 (28-62)

Structure

Structure represents a meta-concept that describes how the virtual hospital organized their team, culture, and resources internally, and how it interacted with external entities via various communication pathways. Structure was therefore conceptualized as internal and external concepts with communication as a subconcept that enabled the relationship between the internal teams and external teams and agencies (Figure 1).

Figure 1. Ontology of concepts and sub-concepts for the structures of a health care organization during the COVID-19 pandemic.

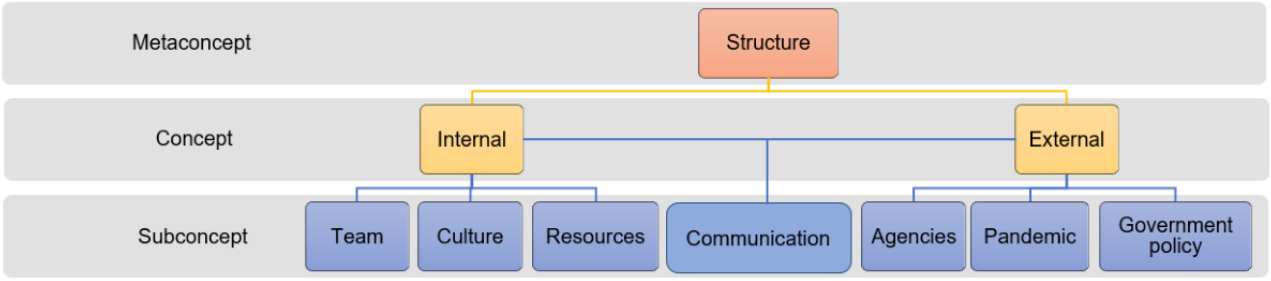

Internal Structure

Internal structure describes how the team functions from within. Out of 3 main subconcepts describing the internal structure emerged from the data. These were team, culture, and resources. Culture and resources were described in relation to dynamics within each of the teams (clinician and IT).

Culture Within Clinician Team

The clinician team comprised doctors, nurses, and nurse unit managers. Within the clinician team, participants reported a culture that fostered the ability of staff members to adapt to rapid changes that frequently occurred during the pandemic.

> *So everyone is working very collaboratively for things to be implemented efficiently, but also quickly …And I think it's purely based on the fact that we got so used to it with the rapid changes with COVID. And*

XSL•FO
RenderX

> *we knew we had to adapt to the change very quickly.* [P005, clinician team]

Also, participants described strong leadership support for health IT implementations, particularly leadership willingness to receive and respond to staff members feedback.

> *I had a lot of experience working on the wards before this. And not much changed, really ever. You know, things really were kind of, this is the way it's done [but here] I think, they [executives] really listen to the feedback of staff and take that on board.* [P004, clinician team]

Culture Within IT Team

Within the IT team, participants reported a learning culture, with IT members reflecting on, and constantly, improving health IT implementation processes.

> *now we've stepped back and gone, hang on, let's just take some time to stop the crisis. Think about what our role is, educate [the clinician team], and give them a process that they can follow, and a process that can follow quite easily…Yes, we have that process now.* [P010, IT team]

Resources Within IT Team

Regarding operational resources, the IT team perceived that the clinician team appeared to have insufficient time to participate in the co-design of health ITs, although it was acknowledged that pandemic patient care is, and should be, their priority. For example, a participant said:

> *…the problem is time, they [clinician team] never have enough time.* [P010, IT team]

High staff turnover within the IT team was identified by participants as another resource concern impacting health IT implementations.

> *And we had people coming and going, so we had business analysts, stop starting. And then we had a six week period where we didn't have a business analyst and we were trying to develop workflow and testing and all of those things… it was a new project team from our side.* [Focus group, P013, IT team]

External Structure

External structure describes the influence of external agencies and factors on health IT implementation. Out of 3 influences were identified including the COVID-19 pandemic, vendors, and government policies.

COVID-19

The onset of the COVID-19 pandemic necessitated the rapid development of virtual health care services to help manage the large numbers of patients affected by the disease and mitigate the risk of shortage of hospital resources. Participants explained that this led to rapid COVID-19 management policy changes, but also unrealistic delivery timelines. The impact was seen to be rapid design without time for adequate reflection.

> *remember policy were changed and we have to roll it out within the week. So we were given one week to do [it]…which caused a lot of problems, because you didn't have time to think about downstream implications. And you really didn't have time to think about, you know what, there's good design required for this. It was just, 'here is what you need to meet, go on ahead and meet that'. So the design was very, very rapid. And it allowed no time to really review.* [P010, IT team]

Vendors

In the context of a virtual hospital where a variety of remote monitoring tools were rapidly and simultaneously implemented during the COVID-19 pandemic, internal teams said that they engaged with multiple vendors to deliver the required technologies. However, participants from the IT team reported that they were unable to themselves configure certain vendor products, which meant that simple configuration tasks were delayed. For example, a participant said:

> *Everything has to go back to the vendor. But that's also one of the features that we wanted to ask them that we could have is to for us to basically be able to configure it ourselves.* [P009, IT team]

Government Policy

Furthermore, it was perceived that the sustainability of the virtual hospital depended on the availability of supportive legislative policies. Participants reported that there was a need for policies to change to recognize virtual hospitals as a viable hospital system.

> *…a lot of systems are built around legislation. And the change in those legislations then mean, we're able to do what we want to do. Because at the moment, the definition of a hospital is not what we are in terms of what we want to do.* [P003, clinician team]

Communication

This is a subconcept that enables the relationship within internal teams and between internal and external teams or agencies for example, state government agencies, vendors so on. Themes emerged surrounding the communication medium and communication approach.

Navigating the Virtual

In terms of the communication medium, virtual co-design sessions became increasingly common during the COVID-19 pandemic to support the rapid design and development of health ITs under time pressure. However, participants explained that virtual co-design sessions often removed the "human" from "human-centered design." Face-to-face co-design sessions were preferred over remote sessions because it was easier to understand each person's role in the project, have informal conversations, and build a social connection.

> *…meetings are online, everything is via [online meeting platform], and that physical element is missing. And it's sort of, you know, watered down the prototyping and the UX [user experience] bit…. Yeah, you don't hear that hallway conversation. … And without these hallway conversations, it seems like*

XSL•FO
RenderX

> *yeah, there's no collaboration that happens.* [P009, IT team]

Some participants described the virtual sessions as transactional and rapid, impacting the interpersonal relationships between project team members as well as the project.

> *A lot of the other projects that we do, we have team forming activities, we get to know each other…We missed all that because there was no time to do it.* [P010, IT team]

Varied Lexicon

Regarding the communication approach within internal teams, participants from both parties reported that clinicians had limited IT literacy and IT teams had limited health literacy, which resulted in miscommunication and key messages being lost in translation. For example:

> *So technically they use English words, and I use English words, but we have different interpretations of what those words mean.* [P002, clinician team]

Another participant said that:

> *we just tell ICT [information and communications technology] and ICT will consult with [the vendor] … and sometimes our request gets lost in translation.* [P005, clinician team]

In addition, participants from the IT team acknowledged that it was important for their team members to simplify complex language and tailor communication to individuals when communicating with clinicians.

> *…like, [staff name]'s great, but [they] use very, very complex language. And sometimes we get lost in that. And so I think that just makes [clinician team] just turn off. And so having to rephrase things all the time, just so they understand is a problem. I think we need to make it simpler, like the development process.* [P010, IT team]

Bridging the Transparency Gap

When reflecting on the communication approach between the internal and external teams, participants from the IT team reported limited transparency and responsiveness from external teams, regarding timelines, strategy, and the continuous improvement of the technology. For example:

> *…there were a lot of problems with [external party] committing to things, also being transparent, showing what they're up to, showing what their intentions were in terms of timeframes in terms of strategy, where the product will be going.* [P011, IT team]

Processes

In this study, Process refers to the health IT development lifecycle and was conceptualized as preimplementation, implementation, and postimplementation phases, as shown in Figure 2.

Figure 2. Ontology of concepts and sub-concepts for health IT implementation processes during the COVID-19 pandemic.

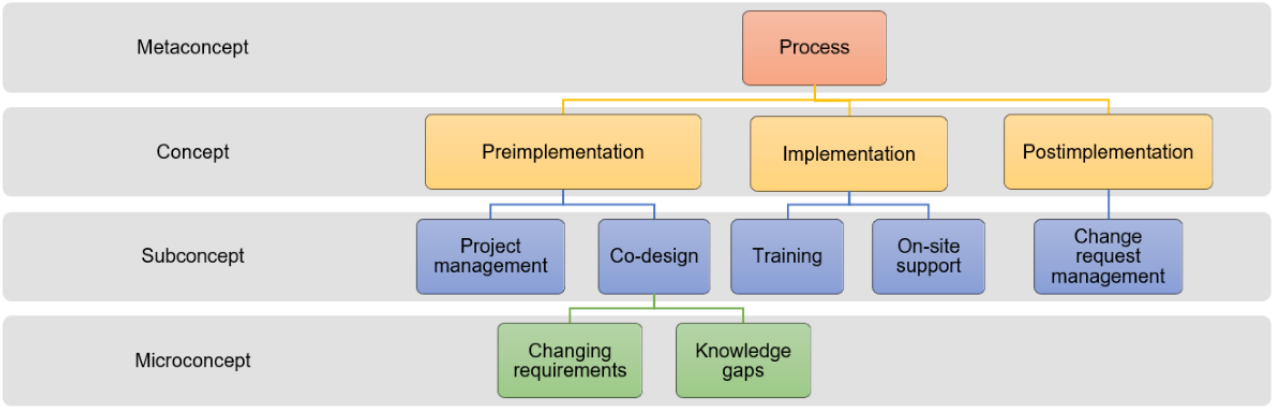

Preimplementation Processes

Preimplementation processes were reported to involve activities undertaken when planning and co-designing HITs. During preimplementation, participants discussed issues related to both project management and co-design.

Optimizing Project Management

In terms of project management, participants raised concerns about how project meetings were planned, managed and documented. For example, regarding time management 1 participant said:

> *Project status meetings with clinicians, we schedule 30 minutes just to say these are the quick updates, but then it goes on for like an hour... I feel like sometimes we don't respect their [clinician team] times in terms of a project team perspective.* [P009, IT team]

Another participant explained that meetings were not well documented because they were conducted virtually, and the minutes were generated electronically:

> *…there were no minutes for almost all the meetings. And then the same questions would come up three months later. And we'd say, well, we discussed this in February. Can we go back to that meeting? No. They're all recorded on [videoconferencing platform], I hate. I hate meetings recorded on teams. That is not minutes. That does not give people an opportunity to reflect on what was said and see if it is what they said.* [P002, clinician team]

XSL•FO
RenderX

Adapting Co-Design to Shifting Realities

Regarding the co-design process, participants described challenges associated with changing requirements over time, and knowledge gaps within both the clinician and IT team.

Adapting to Changing Requirements

Participants reported that fluctuations in COVID-19 patient numbers in different waves of the pandemic led to changes in the system requirements gathered over time:

> *So it was started 12 months prior, and then stopped, and then came into fruition or asked to be activated, once the real COVID number started coming through in that sort of mid-year … So the requirements that were gathered 12 months prior were somewhat different to the requirements that were needed for the COVID dashboard proper, because it had changed.* [P012, IT team, Focus group]

Furthermore, participants stated that the pandemic had placed unprecedented pressure on IT teams resulting in a change in their roles and responsibilities. For example:

> *…the pandemic has put a lot of pressure or more pressure on ICT services staff because you now need to learn a lot more and really quickly to be able to deliver digital solutions not just for clinicians, now it has moved on to patients and now you have to learn AI, clinical decision support systems and all of that. And at a fast pace.* [P009, IT team]

Participants explained that rapid upskilling was needed with user experience (UX) design one of the core skills necessary in the new health IT implementation landscape:

> *…we're not in that UX [user experience] business … we usually just configure things based on what our suppliers have…we don't have that core capability.* [P009, IT team]

Bridging the Knowledge Divide

In addition, some participants reported that the clinician team had knowledge gaps regarding the system development processes, which influenced their expectations around the levels of complexity of some technology requirements and delivery timelines. For example, a participant said:

> *I think their expectations around time to deliver were very short…they don't actually quite understand… that, you've got to do a build, you've got to put it into a non-production environment, you've got to validate it, it's got to go into a staging environment, it's got to be tested, and then it's'got to go into production'environment. And if it's a major change there might be security implications, so there's this whole group of people that are involved, there's an integration activity.* [P010, IT team]

Participants within the clinician team also mentioned that they were unclear about their level of involvement in the co-design process and the level of technical literacy that was required:

> *if I'd known how much work this would be, and how technical it became … I never would have put my hand up to build something. I thought I was just providing some clinical advice about how we use it.* [P002, clinician team]

Conversely, some participants reported that the IT team had knowledge gaps regarding clinical staff roles and levels of seniority, which made it challenging to know when the IT team could engage clinicians; which clinicians should be engaged; and at what stage in the co-design process the clinicians should be engaged. For example, a participant said:

> *So I got asked the other day by the project lead what the difference between a consultant and a registrar is... And for us. That's a really important distinction. And we've worked together for 12 months. And [they] didn't know that difference.* [P002, clinician team]

In addition, participants within the IT team stated that their approach to co-design was not tailored to user needs and rarely included a visual of the user design preferences, noting this to be a knowledge gap within the team.

> *So I think there's a really low level of prototyping or co-design that's happening. ... because people are not comfortable with that level of creating visuals for people… And I guess UX people don't only do visual design, but they try to understand the problems more. So I think we lack a lot in that first part of first understanding people and understanding what they need.* [P009 IT team]

Implementation Processes

Implementation processes involved activities undertaken to facilitate the successful rollout and uptake of health IT. Two main subconcepts describing implementation processes emerged from the data: Training and on-site support.

Tailored Training Strategies

Participants generally viewed training positively and explained that training provided by educators within the clinician team had been tailored for each model of care and was excellent. For example, a participant stated that:

> *Every time there's a new or something goes live, then the staff, the nursing and the medical staff will get training in what are you supposed to do, how are you supposed to assess patients, And they get access to … instructions on what do you do with this patient as part of this trial? …So it's a bit different with each different ones.* [P001, clinician team]

However, some participants identified a need to simplify training to accommodate those who are not technology-savvy and incorporate a more inclusive training approach for older staff.

> *Like, it'll be this huge mass of information. And then there'll be this really essential thing. And I'll think I didn't know that. And then I look at my notes. And I think, why didn't I know that? And I look at my notes, and I didn't know it, because nobody told it to me… I want something I can write on, like, I want something I can look at, because I'm getting older now, and even looking at a screen, I can't see everything, you know, with these presentations. So*

XSL•FO
RenderX

for me to like strain, look and write, it's just messy. [P007, clinician team]

Defining Limits to On-Site Support

During the go-live period, participants from the clinician team reported that the length of time that on-site support was received from the IT team was insufficient. Participants explained that support was available for a predefined length of time, regardless of how many patients were seen, which was problematic when patient numbers were low:

> *So once we went live, we had a small group ICT team that was for support ... That support was supposed to run for 2 weeks. We haven't got [any] patient [for the diverticulitis model of care] for around 3 weeks. So at the end of the 2 weeks. They were like hey guys, so our [diverticulitis model of care] support team is going to finish now. I'm like, we haven't even had our 1st patient. So we don't even know what type of support we will need.* [P005, clinician team]

Postimplementation Processes: Improving Vendor Communication

Postimplementation processes involved activities undertaken to maintain the health IT.

During postimplementation, participants identified one key process that had been challenging, that of issuing requests to the vendor to make technology changes. The change request process, where the IT team requested changes from the vendor, was viewed as ambiguous as it was unclear what tasks required formal requests for change and the associated cost implications. For example, a participant said:

> *So you have to raise change request for that, or you have to raise change requests, everything you want to do, we have to raise change requests for and then they don't tell you if you actually have to pay for the change requested or not. …So they may be doing it for free but they don't tell you that upfront, you only find out afterwards, it makes it really difficult to negotiate anything with them.* [P011, IT team]

Participants also emphasized that it was imperative to involve the IT team in the vendor product release cycle, to allow IT staff to plan and communicate in a timely manner with the stakeholders involved (eg, the clinician team). For example, a participant mentioned that:

> *…if you know there's a bug fix for something coming up, then you can actually communicate that and say, yep, we're on it, rather than keep asking, can we get the colour changed … knowing that in the third quarter release that the colours are all changing.* [P013, IT team]

Focus group Another participant said:

> *And then you can manage your stakeholders better ... And then have a co-design with the [vendor] team on what that product release cycle would look like as well. You know, having input in that.* [P012, IT team, Focus group]

Outcomes

When asked about the performance of the health IT that resulted from the rapid implementation process, participants described impacts on clinician satisfaction and patient experience (Figure 3).

Figure 3. Ontology of concepts for the outcomes of health IT implementation during the COVID-19 pandemic.

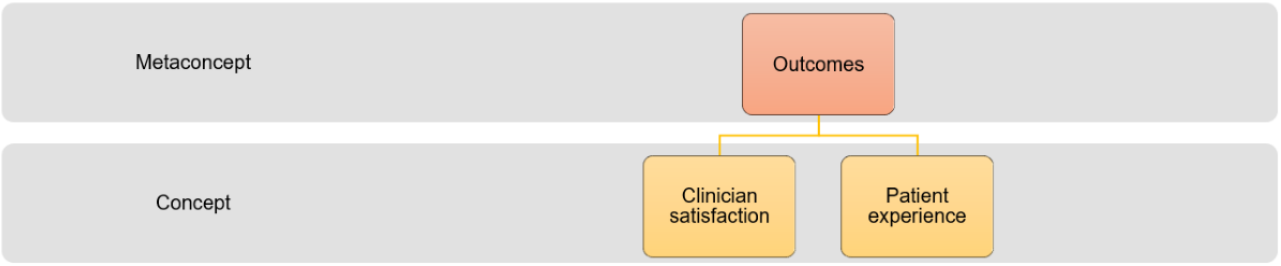

Clinician Satisfaction

Following health IT implementation at the virtual hospital, clinicians reported mixed levels of satisfaction with the technology, but they explained that this improved over time as teething issues with the technology were addressed. Participants raised concerns about negative UX, poor reporting capabilities of the technology, and limited system integration, with the majority of these problems being identified in the first few months following implementation. For example, a participant described a usability problem with one of the dashboards in use:

> *Another issue surrounding alerts is the fact that the dashboard is so big. You need to press… to see more on the dashboard… For us, it's not a scroll. It's a press, it says right here, you have to press it to go all the way to right now… if you don't press all the way towards the right, you can only see half the heart rate.* [P005, clinician team]

Another participant stated that the initial reporting capability did not meet the needs of medical staff members:

> *I actually refused to move the medical workflow to it [dashboard] for two months, because there was no reporting. So, you know, they moved the nursing workflow, and we had medical workflow running on our teams, spreadsheets, and I was like… there's no way of me telling which doctor was asked to see which patient at what time.* [P002, clinician team]

XSL•FO
RenderX

Perceived Patient Experience

As the IT team received direct reports of patient experiences with the remote monitoring mobile phone applications, participants described challenges related to the interface and customizability of the technology. For example, 1 participant said that:

> *Patient facing app…I don't think it's good. I think it doesn't meet the need of the patient…It's interface is terrible. The workflow is not very good. The customization is very, very difficult. And the timeframe for development is way too long for what we want to do.* [P010, IT team]

Also, another participant commented on the ability of the app to present personalized self-management information to patients:

> *Or even just have a link to the website that gives them information about like, heart health or something. They can't do that. So it's like, that's not performing well in terms of the patient's experience, because there's an expectation for you. If you've downloaded an app, you expect more also the things to be contained in that app or will link you to things. Very simple feature set. I don't think the patients are getting that.* [P009, IT team]

Discussion

Principal Findings

In this study, we found that both internal and external team structures and the communication pathways that underpinned these influenced the health IT lifecycle, and pre-to-post implementation, which in turn impacted outcomes, particularly when the normal ways of working were challenged during the pandemic. Preimplementation processes, particularly project management and co-design, were viewed to be most impacted by the COVID-19 pandemic. Specifically, as meetings were held virtually due to social distancing guidelines, and health IT implementations in the pandemic were rapid, participants reported that meeting documentation was automated as it was faster but the outcome of this approach was deemed suboptimal. Also, fluctuations in COVID-19 patient numbers in different waves of the pandemic led to changes in the system requirements gathered over time which led to the need to adapt to shifting realities during co-design. Taken together, our study brings to the fore, the knotty relationship between technology, people, organizational and macroenviromental factors across different health IT implementation phases, aligning with previous research [4].

Comparison With Previous Studies

Effective interdisciplinary communication across the implementation lifecycle was key for successful health IT implementation during the pandemic. Our findings align with existing research demonstrating communication challenges between clinicians and IT teams prepandemic [13,14,30], but showed communication difficulties were exacerbated by time pressures and social distancing guidelines that arose during the pandemic. Our study revealed that the social fabric (ie, web of interactions and connections) that existed within interdisciplinary teams were disrupted due to the urgent need for responsiveness between the internal and external teams, changing system requirements, the communication medium and the varied lexicon between the 2 internal teams. This led to miscommunication, information being lost in translation, and poor interpersonal relationships. While our study adds to the growing body of knowledge on the clinician IT divide, it also highlights effective communication as a necessity for bridging the divide between clinicians and IT experts in time pressured contexts [30]. As the new norm is established, with longer health IT delivery timelines, more stable system requirements, and time for manual meeting documentation and review; factors contributing to miscommunication are minimized but not eliminated as issues such as the varied lexicon between the clinician and IT teams appear enduring.

Regarding the communication medium between teams during health IT implementations, our findings suggest the need to minimize virtual co-design sessions when possible. Virtual co-design sessions became increasingly common during the COVID-19 pandemic to support rapid policy changes and needs, which often occurred with limited experience, preparation or preference of the participants in those sessions. However, it allowed flexibility in the location of co-design and more efficient use of clinician time [18]. Despite these benefits, our interviews revealed that there was limited social connection within co-design teams which contributed to poor interpersonal relationships. This result aligns with that of a simulation study that experimentally manipulated an interdisciplinary co-design process and found that physical separation during co-design led to less social connectedness and misalignment in design outcomes between the subgroups [31]. Another study on virtual co-design with children during the pandemic found that sustaining online sessions over an extended period of time was difficult due to the lack of social connectedness between co-design participants [32]. Also, the clinicians, in our study, reported the limited efficacy of virtual communication, which perhaps reflects the widely experienced videoconferencing fatigue experienced by professionals in health care and other similar service domains [33,34]. Together, these results indicate the need to: identify the types of co-design discussions that suit virtual or face-to-face settings; improve the design and processes around videoconferencing to make it a more acceptable medium for co-design discussions; and reintroduce face-to-face co-design sessions as we transition back to the "new-norm."

The challenges we identified with the preimplementation processes appeared to have a lasting impact on health IT outcomes as they contributed to poor clinician satisfaction and a negative patient experience. For example, in the co-design process, participants from the IT team reported that new work requiring UX expertise was added to their key tasks during the design and implementation of health ITs during the pandemic, thus changing their roles and responsibilities. As their work was previously focused on configuring health ITs from vendors rather than co-designing health ITs with users, the implemented health ITs were sometimes seen to be unsatisfactory to the users. This draws attention to the need for health care organizations to review their IT staff skills and capabilities and align them to the capability needs of a proposed health IT project so that

XSL•FO
RenderX

existing staff members are not required to learn and apply new skills to health IT projects simultaneously. Furthermore, it is imperative for organizations to analyze existing and proposed processes (eg, co-design process) systematically at the organizational level as health IT implementations inevitably lead to new work and redistribution of existing work [35,36]. Consequently, it is imperative for health care organizations to ensure health IT implementation processes cater for work changes within and across clinicians and IT teams rather than the mere purchase of health ITs [37].

In addition to communication and process improvements highlighted in this study, our findings revealed digital health workforce development needs for both clinical and IT students and professionals, to address knowledge gaps and communication challenges. For clinicians, digital competence is typically viewed as knowledge, skills and motivation to use and interact with technologies [38-41]. However, we contend that the perception of *clinicians as users* [42] of digital health technologies must evolve to that of a continuum where clinicians can be at any point of the spectrum that *clinicians as users* to *clinicians as co-designers*. We propose that digital health curriculum should offer clinicians an opportunity to learn the principles of HIT co-design, and the health IT design, development, and implementation lifecycle, to ensure clinicians are equipped with the capabilities and skills to both practice in digital environments and drive digital health innovations. In addition, the suboptimal patient experience reported in our study highlights the need to ensure that patients are also involved as users and co-designers of patient-facing health ITs.

Whilst significant attention has been placed on the development of digital competencies for clinical students, the development of health competencies for IT professionals designing, developing, and implementing health ITs in clinical settings is not as advanced [43-45]. Although clinical informatics programs attempt to bridge this gap, it is not a requirement for IT professionals working in clinical settings, and these skills are sometimes learned on the job [46]. Our findings suggest that IT professionals working in clinical settings may need additional training in the "health sciences" domain, which could encompass knowledge in areas like health concepts, and health sector structures and roles [47].

Strengths, Limitations, and Future Directions

The results from this study have made contributions to the research, practice, and educational requirements for health IT implementations. From a research perspective, we have adapted the interdisciplinary team communication framework (Figures 1-3), incorporating the macro-environmental factor (COVID-19 pandemic) that has significantly impacted health IT implementations since 2020. We tailored the framework to unpack complex structures, processes, and outcomes involved in pandemic-driven health IT implementations between teams. The adapted framework will be useful for researchers undertaking process evaluations for health IT implementations that have occurred since the pandemic.

From a practice perspective, we identified communication structures, including the medium and approach, as key factors that impact health IT implementation processes and outcomes. In our setting, this study resulted in a review of the communication structures available to facilitate better development of interactions and connections that exist within and across internal and external teams. It also led to concrete steps being taken to ensure face-to-face meetings occurred, the creation of office space for IT staff to co-locate in the virtual hospital, and the sustainability of the clinical informatics expertise to work with both the clinician and IT teams.

From the education perspective, we have brought to light the importance of ensuring that digital health education accommodates clinicians as they transition from users to co-designers of health ITs. We further identified the need to better incorporate "health science" knowledge domains in IT curriculum [45] for IT professionals who deliver HITs in clinical settings.

Limitations of our study include that it was an exploratory study that focused on 2 teams within a health service, which may limit the generalizability of the results. The adapted interdisciplinary framework for HIT implementations during the pandemic needs to be validated and studied in the context of other team-based settings. Also, as patients were unable to be recruited for this process evaluation study, our findings may not adequately capture patient perceptions of health ITs during the pandemic. Future work will involve determining the extent to which findings from this study transfer to brick-and-mortar hospital settings and in other interdisciplinary team settings.

Conclusion

The primary purpose of health IT is to facilitate the delivery of quality health care and improve health outcomes [48], but these can only be realized if health IT implementations are successful and health IT is used. Various technological, people, organizational, and macroenvironmental factors have impacted health IT implementation success, more so since the COVID-19 pandemic. Specifically, we found that team structures, communication pathways, and health IT preimplementation processes (project management and co-design) were negatively impacted by the pandemic. While health service organizations are keen to transition to the "new norm," it is imperative to learn from the health IT implementation successes and failures that occurred during the pandemic via process evaluations. Our process evaluation offers learnings for the research (an adapted interdisciplinary team communication framework), practice (the need for health care organizations to review their communication structures, IT staff skills, and proposed processes), and educational requirements (need for better education and training of IT professionals working in clinical settings on health concepts) of health IT implementations as the world transitions to the "new norm."

XSL•FO
RenderX

Acknowledgments

This research was supported by Digital Health CRC Limited (DHCRC). DHCRC is funded under the Australian Commonwealth's Cooperative Research Centres (CRC) Program. SMP is supported by an NHMRC administered fellowship (#1161138).

Data Availability

The datasets generated and analyzed during this study are not publicly available as per ethics approval.

Authors' Contributions

AB-A, M Baysari and SMP contributed to study conceptualization. AB-A and M Baysari contributed to data curation and formal analysis. M Baysari, SMP, TS and M Burger contributed to funding acquisition. AB-A, AB, M Burger, and M Baysari contributed to investigation. AB-A and M Baysari contributed to methodology. AB-A and AB contributed to project administration. AB-A contributed to writing the original draft. All authors contributed to reviewing and editing the paper. All authors reviewed and approved the final version of this manuscript.

Conflicts of Interest

Authors AB, OH and MS are clinical or managerial staff members at the virtual hospital where the study was conducted.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 398 KB - jmir_v27i1e57847_app1.pdf]

Multimedia Appendix 2
Interview Guide.
[DOCX File , 22 KB - jmir_v27i1e57847_app2.docx]

References

1. Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, et al. Health information technology and digital innovation for national learning health and care systems. Lancet Digit Health 2021;3(6):e383-e396 [FREE Full text] [doi: 10.1016/S2589-7500(21)00005-4] [Medline: 33967002]
2. Malden S, Heeney C, Bates DW, Sheikh A. Utilizing health information technology in the treatment and management of patients during the COVID-19 pandemic: lessons from international case study sites. J Am Med Inform Assoc 2021;28(7):1555-1563 [FREE Full text] [doi: 10.1093/jamia/ocab057] [Medline: 33713131]
3. Ye J. Health information system's responses to COVID-19 pandemic in China: a national cross-sectional study. Appl Clin Inform 2021;12(2):399-406 [FREE Full text] [doi: 10.1055/s-0041-1728770] [Medline: 34010976]
4. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
5. Cresswell K, Williams R, Sheikh A. Developing and applying a formative evaluation framework for health information technology implementations: qualitative investigation. J Med Internet Res 2020;22(6):e15068 [FREE Full text] [doi: 10.2196/15068] [Medline: 32519968]
6. Dyda A, Fahim M, Fraser J, Kirrane M, Wong I, McNeil K, et al. Managing the digital disruption associated with COVID-19-driven rapid digital transformation in Brisbane, Australia. Appl Clin Inform 2021;12(5):1135-1143 [FREE Full text] [doi: 10.1055/s-0041-1740190] [Medline: 34852391]
7. Cresswell K, Sheikh A, Franklin BD, Krasuska M, Nguyen HT, Hinder S, et al. Interorganizational knowledge sharing to establish digital health learning ecosystems: qualitative evaluation of a national digital health transformation program in England. J Med Internet Res 2021;23(8):e23372 [FREE Full text] [doi: 10.2196/23372] [Medline: 34420927]
8. Cresswell K, Domínguez Hernández A, Williams R, Sheikh A. Key challenges and opportunities for cloud technology in health care: semistructured interview study. JMIR Hum Factors 2022;9(1):e31246 [FREE Full text] [doi: 10.2196/31246] [Medline: 34989688]
9. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
10. Bamgboje-Ayodele A, Boscolo A, Hutchings O, Shaw M, Burger M, Taggart R, et al. Fighting the same battles on a new battleground: embedding technologies in a virtual care environment. Stud Health Technol Inform 2024;310:1066-1070. [doi: 10.3233/SHTI231128] [Medline: 38269978]

XSL•FO
RenderX

11. Abell B, Naicker S, Rodwell D, Donovan T, Tariq A, Baysari M, et al. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. Implement Sci 2023;18(1):32 [FREE Full text] [doi: 10.1186/s13012-023-01287-y] [Medline: 37495997]
12. Bamgboje-Ayodele A, McPhail SM, Brain D, Taggart R, Burger M, Bruce L, et al. How digital health translational research is prioritised: a qualitative stakeholder-driven approach to decision support evaluation. BMJ Open 2023;13(11):e075009 [FREE Full text] [doi: 10.1136/bmjopen-2023-075009] [Medline: 37931965]
13. Huang ME. IT is from mars and physicians from venus: bridging the gap. PM R 2017;9(5S):S19-S25. [doi: 10.1016/j.pmrj.2017.02.011] [Medline: 28527499]
14. Garcia G, Bertoia A, Cameselle L, Benitez S, Giunta D, Baum A, et al. Communication problems between end-users and technicians through a help desk in a health information system. Stud Health Technol Inform 2015;216:985. [Medline: 26262287]
15. Martikainen S, Korpela M, Tiihonen T. User participation in healthcare IT development: a developers' viewpoint in Finland. Int J Med Inform 2014;83(3):189-200. [doi: 10.1016/j.ijmedinf.2013.12.003] [Medline: 24382475]
16. Shaha JS, El-Othmani MM, Saleh JK, Bozic KJ, Wright J, Tokish JM, et al. The growing gap in electronic medical record satisfaction between clinicians and information technology professionals: issues of most concern and suggested remediations. J Bone Joint Surg Am 2015;97(23):1979-1984. [doi: 10.2106/JBJS.N.01118] [Medline: 26632000]
17. Agulnik A, Boykin D, O'Malley DM, Price J, Yang M, McKone M, et al. Virtual facilitation best practices and research priorities: a scoping review. Implement Sci Commun 2024;5(1):16 [FREE Full text] [doi: 10.1186/s43058-024-00551-6] [Medline: 38365878]
18. Savoy A, Patel H, Shahid U, Offner AD, Singh H, Giardina TD, et al. Electronic co-design (ECO-design) workshop for increasing clinician participation in the design of health services interventions: participatory design approach. JMIR Hum Factors 2022;9(3):e37313 [FREE Full text] [doi: 10.2196/37313] [Medline: 36136374]
19. De Vera K, Challa P, Liu RH, Fuller K, Feroz AS, Gamble A, et al. Virtual primary care implementation during COVID-19 in high-income countries: a scoping review. Telemed J E Health 2022;28(7):920-931. [doi: 10.1089/tmj.2021.0377] [Medline: 34847758]
20. Harrison R, Prokopy M, Perreira T. Virtual care post-pandemic: why user engagement is critical to create and optimise future models of care. Digit Health 2022;8:20552076221131455 [FREE Full text] [doi: 10.1177/20552076221131455] [Medline: 36238755]
21. Hutchings OR, Dearing C, Jagers D, Shaw MJ, Raffan F, Jones A, et al. Virtual health care for community management of patients with COVID-19 in Australia: observational cohort study. J Med Internet Res 2021;23(3):e21064 [FREE Full text] [doi: 10.2196/21064] [Medline: 33687341]
22. Carroll N, Conboy K. Normalising the "new normal": changing tech-driven work practices under pandemic time pressure. Int J Inf Manage 2020;55:102186 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102186] [Medline: 32836643]
23. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Los Angeles, Washington DC: Sage publications; 2017.
24. Morgan DL, Krueger RA, King JA. The Focus Group Kit. Thousand Oaks, California: Sage Publications, Inc; 1998.
25. Clarke V, Braun V. Thematic analysis. In: Encyclopedia of Critical Psychology. Cham: Springer; 2014:1947-1952.
26. Byrne D. A worked example of braun and clarke's approach to reflexive thematic analysis. Qual Quant 2021;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
27. Kuziemsky CE, Borycki EM, Purkis ME, Black F, Boyle M, Cloutier-Fisher D, Interprofessional Practices Team. An interdisciplinary team communication framework and its application to healthcare 'e-teams' systems design. BMC Med Inform Decis Mak 2009;9:43 [FREE Full text] [doi: 10.1186/1472-6947-9-43] [Medline: 19754966]
28. Donabedian A. Evaluating the Quality of Medical Care. Milbank Mem Fund Q 1966;44(3):166-206. [doi: 10.2307/3348969]
29. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
30. Bastian G, Baker GH, Limon A. Bridging the divide between data scientists and clinicians. Intell Based Med 2022;6:100066. [doi: 10.1016/j.ibmed.2022.100066]
31. Mazalek A, Winegarden C, Al-Haddad T, Robinson SJ, Wu CS. Architales: physical/digital co-design of an interactive story table. 2009 Presented at: TEI '09: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction; February 16-18, 2009; Cambridge, United Kingdom p. 241-248. [doi: 10.1145/1517664.1517716]
32. Fails JA, Ratakonda DK, Koren N, Elsayed-Ali S, Bonsignore E, Yip J. Pushing boundaries of co-design by going online: lessons learned and reflections from three perspectives. Int J Child Comput Interact 2022;33:100476 [FREE Full text] [doi: 10.1016/j.ijcci.2022.100476] [Medline: 36034969]
33. Newbould L, Ariss S, Mountain G, Hawley MS. Exploring factors that affect the uptake and sustainability of videoconferencing for healthcare provision for older adults in care homes: a realist evaluation. BMC Med Inform Decis Mak 2021;21(1):13 [FREE Full text] [doi: 10.1186/s12911-020-01372-y] [Medline: 33407437]
34. Anh LET, Whelan E, Umair A. 'You're still on mute'. A study of video conferencing fatigue during the COVID-19 pandemic from a technostress perspective. Behav Inf Technol 2022;42(11):1758-1772. [doi: 10.1080/0144929x.2022.2095304]

XSL•FO
RenderX

35. Carayon P, Hundt AS, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006;15 Suppl 1(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
36. Karsh B, Holden R. New technology implementation in health care. In: Carayon P, editor. Handbook of human factors and ergonomics in health care and patient safety. Mahwah, NJ: Lawrence Erlbaum; 2007:393-410.
37. Walker JM, Carayon P. From tasks to processes: the case for changing health information technology to improve health care. Health Aff (Millwood) 2009;28(2):467-477. [doi: 10.1377/hlthaff.28.2.467] [Medline: 19276006]
38. Jarva E, Oikarinen A, Andersson J, Tuomikoski AM, Kääriäinen M, Meriläinen M, et al. Healthcare professionals' perceptions of digital health competence: a qualitative descriptive study. Nurs Open 2022;9(2):1379-1393 [FREE Full text] [doi: 10.1002/nop2.1184] [Medline: 35094493]
39. Jarva E, Mikkonen K, Andersson J, Tuomikoski AM, Kääriäinen M, Meriläinen M, et al. Aspects associated with health care professionals' digital health competence development – a qualitative study. FinJeHeW 2022;14(1):79-91. [doi: 10.23996/fjhw.111771]
40. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
41. NA. National nursing and midwifery digital health capability framework. In: Australian Digital Health Agency. Sydney, Australia: Australian Government; 2020.
42. McGowan JJ, Passiment M, Hoffman HM. Educating medical students as competent users of health information technologies: the MSOP data. Stud Health Technol Inform 2007;129(Pt 2):1414-1418. [Medline: 17911947]
43. Chatterjee S, LeRouge CM, Chiarini Tremblay M. Educating students in healthcare information technology: is community barriers, challenges, and paths forward. Commun Assoc Inf Syst 2013;33. [doi: 10.17705/1cais.03301]
44. Lee A, Moy L, Kruck SE, Rabang J. The doctor is in, but is academia? Re-tooling IT education for a new era in healthcare. J Inf Syst Educ 2014;25(4):275-282 [FREE Full text]
45. Sukhomlin V, Zubareva E. Analytical review of the current curriculum standards in information technologies. In: International Conference on Modern Information Technology and IT Education. Cham: Springer; 2018.
46. Health Informatics Professional Core Competency. Canada: Canada's Health Informatics Association; 2012.
47. Australian Health Informatics Competency Framework. Australia: Australasian Institute of Digital Health; 2022.
48. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
TPOM: Technology, People, Organizations, and Macro-environmental factors
UX: user experience

Edited by T Leung; submitted 28.02.24; peer-reviewed by S Hajesmaeel Gohari, E Quinn; comments to author 10.04.24; revised version received 08.05.24; accepted 30.11.24; published 11.02.25.

Please cite as:
Bamgboje-Ayodele A, Boscolo A, Burger M, Hutchings O, Shaw M, Shaw T, Tariq A, Naicker S, McPhail S, Baysari M
Health IT Implementation and the Impact of the COVID-19 Pandemic on Clinician-IT Dynamics: Qualitative Study
J Med Internet Res 2025;27:e57847
URL: https://www.jmir.org/2025/1/e57847
doi:10.2196/57847
PMID:

©Adeola Bamgboje-Ayodele, Adrian Boscolo, Mitchell Burger, Owen Hutchings, Miranda Shaw, Tim Shaw, Amina Tariq, Sundresan Naicker, Steven McPhail, Melissa Baysari. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of Virtual Reality Training in Improving Outcomes for Dialysis Patients: Systematic Review and Meta-Analysis

Xin Kang[1*], MScN; Yiping Zhang[1*], MScN; Chaonan Sun[1], MScN; Jiaxin Zhang[1], MScN; Zhe Che[1], MScN; Jinhui Zang[1], MScN; Rongzhi Zhang[1], BN, MSc

Department of Center for Hemodialysis, The Second Hospital of Dalian Medical University, Dalian, China

*these authors contributed equally

Corresponding Author:
Rongzhi Zhang, BN, MSc
Department of Center for Hemodialysis
The Second Hospital of Dalian Medical University
467, Zhongshan Road
Shahekou District, Liaoning Province
Dalian, 116000
China
Phone: 86 17709870601
Email: zhangxinzhi169@163.com

Abstract

Background: Virtual reality (VR) training uses computer-generated simulations that enable users to engage with immersive virtual environments, simulating real-world activities or therapeutic exercises. This technology is increasingly recognized as a promising intervention to address the physical and psychological challenges faced by dialysis patients, who frequently experience diminished physical function, social isolation, and emotional distress associated with prolonged treatment regimens. Given the increasing prevalence of dialysis patients and the limitations of conventional rehabilitation approaches, VR presents a novel, interactive method that has the potential to enhance patient well-being and improve quality of life.

Objective: This meta-analysis aimed to evaluate the effectiveness of VR training interventions for dialysis patients, with a focus on assessing their impact on motor abilities, psychological symptoms (specifically anxiety and depression), social functioning, and self-efficacy. This analysis also explores whether VR can offer comprehensive benefits to support both the physical and mental health of dialysis patients.

Methods: The meta-analysis was conducted following Cochrane guidelines. Comprehensive searches were performed across major databases, including China National Knowledge Infrastructure, Wanfang database, China Science and Technology Journal Database, China Biomedical Literature database, Cochrane library, Web of Science, PubMed, and Embase, encompassing all studies up to December 2023. Inclusion criteria targeted studies assessing VR's impact on motor performance, psychological well-being, social functioning, and self-efficacy in dialysis patients. Two reviewers independently extracted data and assessed methodological quality using Cochrane's risk of bias criteria, ensuring data synthesis reliability.

Results: A total of 12 studies, involving 625 dialysis patients in total, met the inclusion criteria. The meta-analysis demonstrated that VR training led to significant improvements across multiple domains. VR interventions were associated with improved physical capacity, evidenced by higher scores in the 6-minute walk test (standardized mean difference [SD]=29.36, 95% CI 14.32-44.4, P<.001, I^2=46%). VR training was associated with significant reductions in depression (SD=–6.30, 95% CI –7.14 to –5.47, P<.001, I^2=96%) and anxiety (SD=–8.91, 95% CI –9.69 to –8.14, P<.001, I^2=95%). In addition, VR interventions enhanced social functioning (SD=16.20, 95% CI 14.49-17.9, P<.001, I^2=72%), and improved self-efficacy (SD=20.47, 95% CI 18.55-22.39, P<.001, I^2=99%). However, VR training did not yield significant differences in gait speed, balance, or functional tests (Ten Sit-to-Stand Test, Five Sit-to-Stand Test, Sixty Sit-to-Stand Test, Timed Up and Go Test, and fatigue) compared with control groups.

Conclusions: The findings suggest that VR training is a promising intervention for dialysis patients, providing benefits in physical endurance, social engagement, and psychological well-being. Despite these advantages, VR remains underused among peritoneal dialysis patients compared with hemodialysis patients. Further studies with larger sample sizes and more refined

XSL•FO
RenderX

experimental designs are recommended to validate these results and support VR as a complementary tool in the holistic care of dialysis patients.

(J Med Internet Res 2025;27:e58384) doi:10.2196/58384

KEYWORDS

virtual reality; VR; training; dialysis; hemodialysis; peritoneal dialysis; chronic kidney disease; rehabilitation; quality of life; meta-analysis

Introduction

End-stage renal disease (ESRD) signifies a severe impairment in kidney function, rendering the body incapable of meeting normal metabolic demands. This condition is chronic and progressive, with a global prevalence reaching 442.13 per million [1]. Dialysis, encompassing both hemodialysis (HD) and peritoneal dialysis (PD), serves as a common therapeutic approach for ESRD, substituting renal filtration function albeit necessitating frequent hospital visits, significantly impacting patients' lives. Prolonged dialysis treatment often precipitates complications such as anemia, muscle wasting, skeletal disorders, and cardiovascular ailments. Exercise therapy plays a pivotal role in managing patients with ESRD by yielding multifaceted benefits, including enhancements in cardiovascular function [2], alleviation of fatigue and sleep disturbances, bolstering mental health by mitigating anxiety and depression [3], regulating blood pressure, managing anemia [4], and optimizing dialysis adequacy [5]. Furthermore, regular exercise effectively retards the deterioration of muscle strength [6] and maintains superior physical function. Despite extensive research corroborating the merits of exercise in augmenting hemodialysis patients' quality of life, there remains a notable deficiency in exercise awareness, knowledge application, and adherence among patients with ESRD [7] predominantly attributable to inadequate comprehension, deficient self-management skills, and low compliance levels.

Virtual reality (VR) technology represents a computer-generated simulation environment enabling user interaction within virtual realms, engendering an immersive experience that transports individuals into diverse environments [8]. Within the medical sphere, VR has gained traction and is increasingly adopted. VR facilitates immersive encounters that divert patients' attention from the discomfort and anxiety associated with medical procedures, such as dialysis treatment [9]. By creating immersive and interactive virtual environments, VR offers a unique platform to improve operational skills, health knowledge, and psychological resilience among these patients.

The application of VR in dialysis begins with its role in enhancing operational skills and ensuring treatment safety. Patients can engage in simulated environments that replicate dialysis machine operations and scenarios. This allows them to practice procedures and familiarize themselves with equipment without the inherent risks associated with real-world settings. Such training not only reduces procedural errors but also boosts confidence and competence in managing treatments, ultimately contributing to better treatment adherence and patient safety.

In addition, VR serves as a powerful educational platform in dialysis care. It facilitates the delivery of comprehensive health education, covering critical aspects such as medication management, dietary guidelines, and health monitoring techniques [10]. By engaging patients in interactive simulations, VR enables them to acquire a deeper understanding of their treatment regimen and develop essential self-management skills. This educational empowerment is pivotal in encouraging patients to take an active role in their health management, thereby improving treatment outcomes and enhancing overall quality of life.

Furthermore, VR addresses the significant psychological challenges faced by dialysis patients. Long-term treatment often leads to emotional stress and psychological burdens, including conditions such as anxiety and depression. VR environments provide a therapeutic avenue for relaxation and emotional support. Patients can participate in virtual mindfulness exercises, guided relaxation sessions, or social interactions within a controlled setting. These activities help alleviate stress, improve mood, and foster emotional resilience among patients, thereby enhancing their psychological well-being.

VR training offers significant advantages over traditional physical therapy for treating dialysis patients. First, VR technology provides highly immersive experiences that enhance patient engagement and motivation. Dialysis patients often face long-term treatments and physical limitations, and VR environments immerse them in compelling experiences that encourage greater participation in rehabilitation training. Second, VR allows for personalized treatment plans tailored to individual patient needs, including adjusting training difficulty and environment to accommodate varying abilities and therapeutic requirements. This personalized approach enhances treatment efficacy and specificity.

Furthermore, VR training provides a safer therapeutic environment with better control mechanisms. Conducting exercises and rehabilitation in virtual environments reduces real-world safety risks for patients, while health care teams can monitor progress and responses in real time to adjust treatment plans promptly. In addition, VR systems can collect detailed data such as movement skills and physiological responses, offering crucial insights for health care professionals to optimize treatment plans and deliver personalized care.

Conducting a meta-analysis on the application of VR in dialysis patients holds significant clinical implications. Systematically synthesizing data from multiple studies allow for a comprehensive evaluation of VR therapy's actual impact on improving quality of life and enhancing rehabilitation outcomes in dialysis patients. Such meta-analysis not only compares VR

XSL•FO
RenderX

training with traditional physical therapy but also identifies potential biases and inconsistencies in research, providing scientific grounds for designing future clinical trials and refining treatment strategies. Therefore, exploring the effectiveness of VR through meta-analysis in dialysis patients not only deepens understanding of its clinical role but also informs future clinical decisions and rehabilitation strategies based on robust scientific evidence.

Methods

Overview

This meta-analysis was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The PRISMA checklist is provided in Multimedia Appendix 1.

Search Strategy

We conducted a comprehensive literature search across 8 databases, namely China National Knowledge Infrastructure (CNKI), Wanfang database, China Science and Technology Journal Database, China Biomedical Literature database (SinoMed), Cochrane library, Web of Science, PubMed, and Embase. We used a combination of subject terms and free terms tailored to each database. The search was performed for studies published from inception up to December 2023. Two researchers independently executed the searches and managed the retrieved literature using EndNote software. Detailed search strategies for each database are provided in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

Textbox 1 shows the inclusion and exclusion criteria.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Study design: randomized controlled trials following the Population, Intervention, Comparison, Outcome framework.
- Population: adults aged ≥18 years undergoing dialysis (hemodialysis or peritoneal dialysis).
- Intervention: exercise training incorporating virtual reality.
- Comparison: conventional hemodialysis care interventions (including health education and routine exercise) or exercise not based on virtual reality.
- Outcome measures: motor ability, psychological symptoms, social functioning, and self-efficacy.
- Language: studies published in Chinese or English.

Exclusion criteria

- Duplicate publications.
- Non-Chinese or non-English literature.
- Studies with incomplete data (eg, missing outcome indicators).
- Conference abstracts and review articles.

Data Extraction

Two researchers independently extracted data from the included studies, including details such as author, publication year, country, sample size, intervention specifics, and outcome measures. Studies were initially screened based on titles and abstracts, followed by a full-text review to ensure they met the inclusion criteria. Discrepancies between the 2 researchers were resolved through discussion, and if consensus could not be reached, a third researcher was consulted for final arbitration.

To ensure comprehensive data extraction, we contacted study authors by email to request missing or unclear data. Studies with unresolved data issues were excluded from the meta-analysis.

Quality Assessment of Literature

The quality of included studies was assessed using the Cochrane Collaboration's risk of bias tool. We evaluated studies on several criteria, such as random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other potential biases. Each criterion was classified as low risk, unclear risk, or high risk of bias [11].

Statistical Methods

Statistical analyses were performed using RevMan (version 5.4) software and State (version 17) provided by the Cochrane Collaboration. We analyzed both continuous and dichotomous variables, reporting relative risk with 95% CIs. A fixed-effects model was used for the meta-analysis of outcome indicators. We assessed heterogeneity among studies using the chi-square test and *I*-squared (I^2) statistic, with I^2 values of 25%, 50%, and 75% indicating low, moderate, and high levels of heterogeneity, respectively. Publication bias was evaluated through funnel plots and Egger regression test [12]. A 2-tailed *P* value of ≤.05 was considered statistically significant [13]. Due to the limited number of studies, further subgroup and sensitivity analyses were not performed.

In cases where data were incomplete or unclear, we sought additional information from the authors. Studies for which relevant data could not be obtained were excluded from the systematic review and meta-analysis.

XSL•FO
RenderX

Results

Results of Literature Search

A total of 4850 articles were initially retrieved from the literature search, comprising 80 from CNKI, 4 from Wanfang database, 121 from China Science and Technology Journal Database, 38 from SinoMed, 34 from the Cochrane library, 2921 from Web of Science, 656 from PubMed, and 996 from Embase. After a systematic screening process, 12 articles were ultimately included. The detailed screening process and the reasons for exclusion are illustrated in Figure 1.

Figure 1. Flowchart of the study selection process.

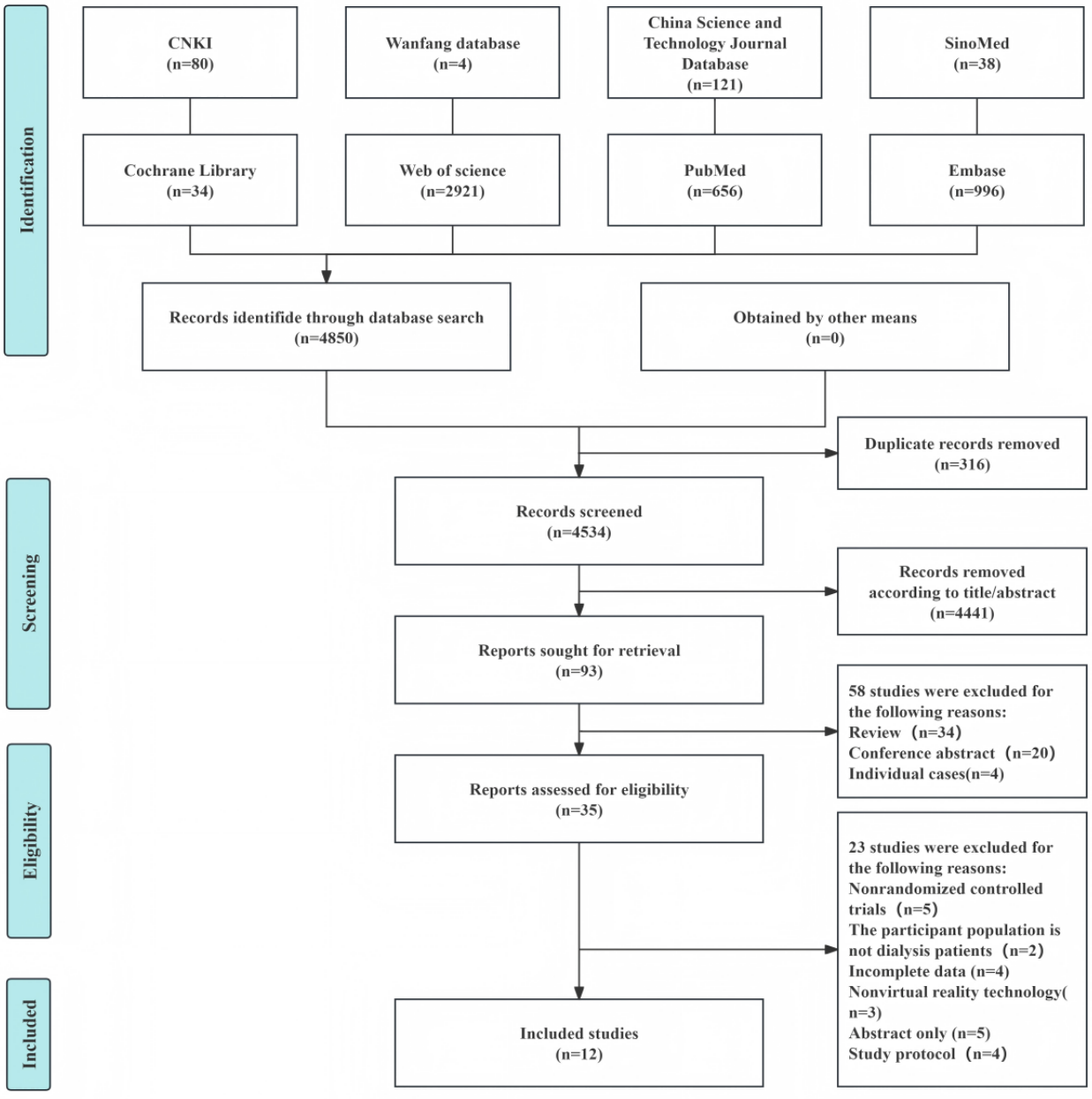

Basic Characteristics of Included Studies

The included studies were conducted across 7 countries: namely, Poland [14], Spain [15,16], China [17-21], Qatar [22], South Korea [23], Germany [24], and Brazil [25]. One study used a 3-arm design [24], while the remaining 11 studies used a 2-arm design. Of the 12 trials, 2 focused on patients undergoing PD, while the remaining 10 involved patients undergoing HD [17,21]. Five trials compared VR training with conventional care [18-21,23], 2 trials included specified interventions in the control group [14,16], and the remaining trials compared VR training with other types of interventions, such as printed materials [17], exercise programs based on daily habits [15,18], and nurse-led exercise programs [16,22]. Summary of included studies are provided in Multimedia Appendix 3. Details regarding the number of participants, interventions, and outcomes are presented in Table 1. Further information on the interventions under experimental and control conditions is detailed in Multimedia Appendix 2.

XSL•FO
RenderX

Table 1. Characteristics of the 12 included studies.

Study	Country	Type	Number (n)			Duration	Time of data collection	Outcomes
			Total	Experimental group	Control group			
Turoń-Skrzypińska et al [14]	Poland	HD[a]	85	39	46	1 year	At the time of entering the research project and following month 3	Depression and anxiety
Lee et al [17]	China	PD[b]	23	12	11	8 times	At baseline (pretreatment), within the first week after recruitment, and at posttreatment	Knowledge proficiency, procedure competence, and self-efficacy
Segur‐Ortí et al [15]	Spain	HD[a]	18	9	9	20 weeks	At baseline, after 16 weeks of intradialysis combined exercise, and by the end of four additional weeks of exercise	STS-10[c], STS-60[d], gait speed, one-leg heel-rise tests, and 6MWT[e]
Martínez-Olmos et al [16]	Spain	HD[a]	56	28	28	12 weeks	At baseline, at 12 weeks, and at 24 weeks	The 4-m gait speed test, SPPB[f], TUG[g] test, OLST[h] for balance, STS-10[c], STS-60[d], 6MWT[e], and adherence to the exercise program
Lei et al [18]	China	HD[a]	42	21	21	6 mouths	Before enrollment, at 3 months, and 6 months after enrollment	Self-efficacy for exercise, risk of falls, and quality of life
Zhou et al [22]	Qatar	HD[a]	73	37	36	4 weeks	At baseline and at 4 weeks	Depression and user experience
Chou et al [19]	China	HD[a]	64	32	32	4 weeks	Before intervention and after 4 weeks of intervention	Fatigue, BUN[i], creatinine, albumin, and hemoglobin
Cho and Sohng [23]	South Korea	HD[a]	46	23	23	8 weeks	At baseline and after intervention	Physical fitness, body composition, and fatigue
Schinner et al [24]	Germany	HD[a]	32	12+9	11	12 weeks	—[j]	Functional capacity, serum biochemistry, muscle strength, muscle circumference, and body composition
Maynard et al [25]	Brazil	HD[a]	40	20	20	12 weeks	After intervention	Functional capacity, quality of life, and depressive symptoms
Xuelian et al [20]	China	HD[a]	70	35	35	—[j]	Before intervention and after intervention	Anxiety, depression, quality of life, and treatment compliance
Lina and Ke [21]	China	PD[b]	76	38	38	3 months	—[j]	Activity level and 6MWT[e]

[a]HD: hemodialysis.

[b]PD: peritoneal dialysis.

[c]STS-10: sit-to-stand 10.

[d]STS-60: sit-to-stand 60.

[e]6MWT: 6-min walking test.

[f]SPPB: short physical performance battery.

[g]TUG: timed up-and-go.

[h]OLST: one-legged stance test.

[i]BUN: blood urea nitrogen.

[j]Not applicable.

XSL•FO
RenderX

Risk of Bias Assessment

According to Cochrane criteria, the risk of bias assessment is illustrated in Figures 2 [14-25] and 3. Of the 12 studies, 7/12 (58.3%) studies [14,15,17,18,21,22,24] were rated as having "some concerns" of bias, and 5/12 (41.7%) studies [16,19,20,23,25] were rated as high risk of bias. In the domain of "random sequence generation," 5/12 (41.7%) studies [16,17,22,24,25] had low bias, 5/12 (41.7%) studies [14,15,18,20,21] had unclear risk of bias, and 2/12 (16.7%) studies [19,23] had high risk of bias. In the domain of "allocation concealment," 6/12 (50%) studies [16-18,22,24,25] had low bias, 4/12 (33.3%) studies [14,15,20,21] had unclear risk of bias, and 2/12 (16.7%) studies [19,23] had high risk of bias. In the domain of "blinding of participants and personnel," 3/12 (25%) studies [21,23,24] had low bias, 7/12 (58.3%) studies [14-19,22] had unclear risk of bias, and 2/12 (16.7%) studies [20,25] had high risk of bias. Blinding participants was not feasible due to the nature of the intervention methods used in all trials. In the domain of "blinding of outcome assessment," 6/12 (50%) studies [16,19-21,23,24] had low bias and 6/12 (50%) studies [14,15,18,20,22,25] had unclear risk of bias. The absence of blinding outcome assessors or data analysts was observed in studies with some bias. In the domain of "incomplete outcome data," 9/12 (75%) studies [14,15,17-23] had low bias, 1/12 (8.3%) studies [24] had unclear risk of bias, and 2/12 (16.7%) studies [16,25] had high risk of bias. In the domain of "selective reporting," 12/12 (100%) studies [14-25] had a low bias.

Figure 2. Risk of bias analysis of each included study [14-25].

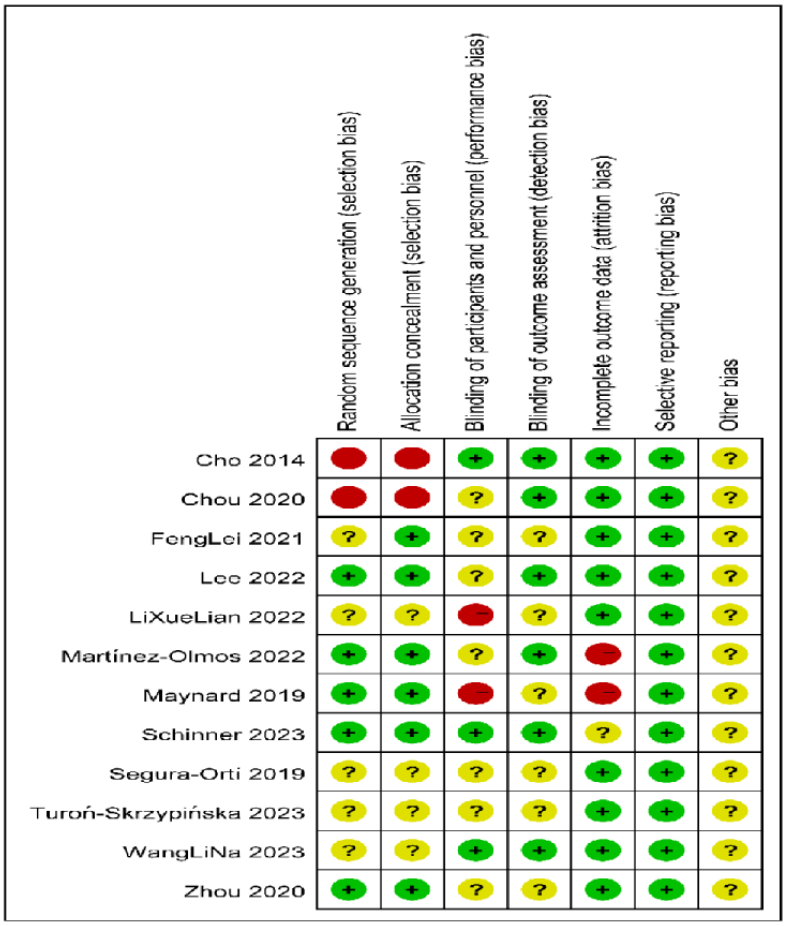

XSL•FO
RenderX

Figure 3. Overall risk of bias analysis of included studies.

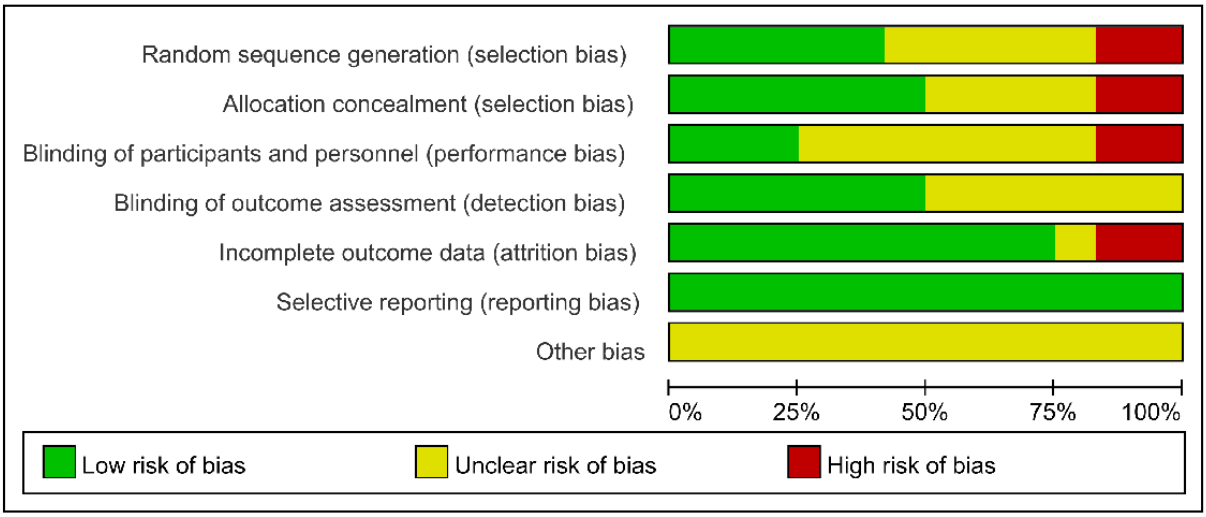

Meta-Analysis Results

Motor Ability

Gait Speed

Two trials used gait speed as a measurement outcome [15,16]. The results indicate no significant difference in gait speed improvement between VR training and other training methods (standardized mean difference [SD]=0.04, 95% CI –0.02 to 0.1, *P*=.16) with low heterogeneity (I^2=0%, *P*=.79; Figure 4 [15,16]). Since there were fewer than 3 studies, publication bias was not assessed.

Figure 4. Forest plot of virtual reality on exercise capacity of gait speed [15,16].

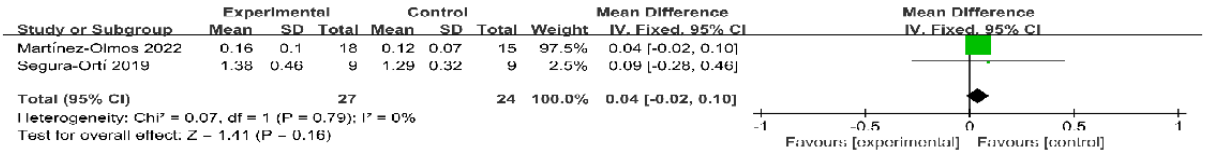

6-Minute Walk Test

Three trials measured walking ability using the 6-minute walk test (6MWT) [15,16,21]. Results showed that VR training improved walking ability significantly more than the control group (SD=29.36, 95% CI 14.32-44.4, *P*<.001) with middle heterogeneity (I^2=46%, *P*=.16; Figure 5 [15,16,21]). Egger regression test and funnel plot showed that there was no significant publication bias between the included studies (*P*=.99; Figure 6).

Figure 5. Forest plot of virtual reality on exercise capacity of 6-minute walk test [15,16,21].

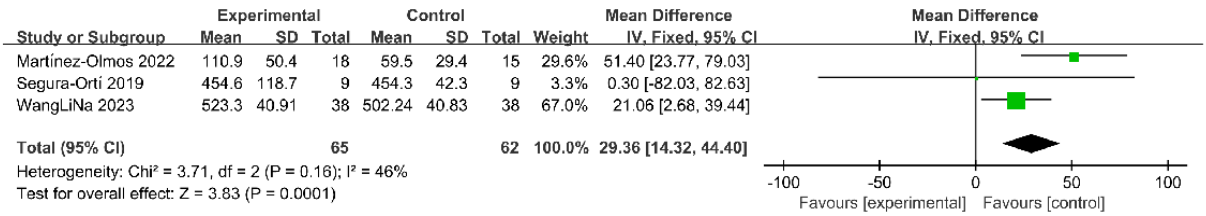

XSL•FO
RenderX

Figure 6. The funnel plot of the effect of virtual reality on dialysis patient.

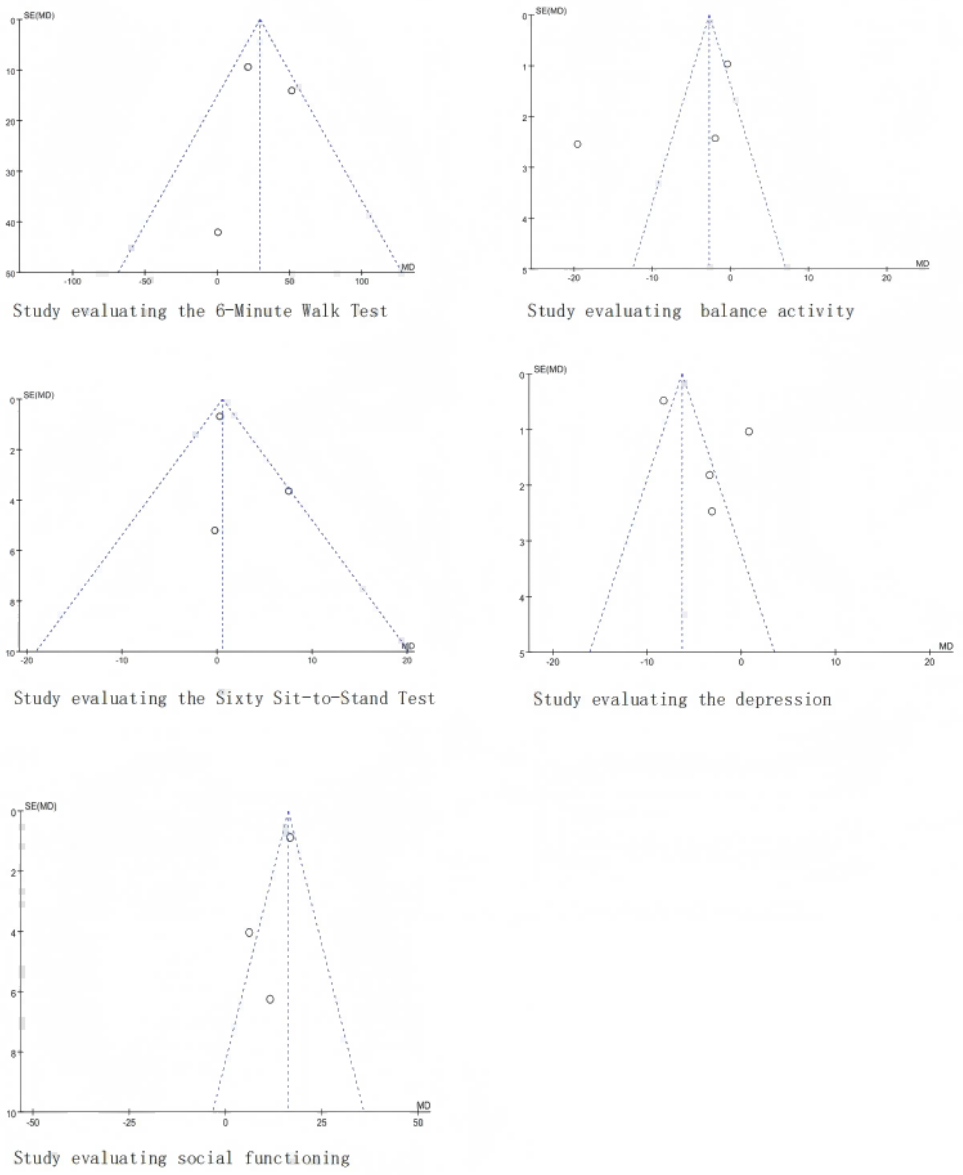

Balance Ability

Three trials included the measurement of balance ability as an outcome [16,18,23]. The findings suggest that VR training did not lead to an improvement in the balance ability of the subjects when compared with the control group (SD=–2.73, 95% CI –4.40 to –1.06, *P*<.001), and considerable heterogeneity was found (I^2=96%, *P*<.001; Figure 7 [16,18,23]). Egger regression test and funnel plot showed that there was no significant publication bias between the included studies (*P*=.48; Figure 6).

Figure 7. Forest plot of virtual reality on exercise capacity of balance [16,18,23].

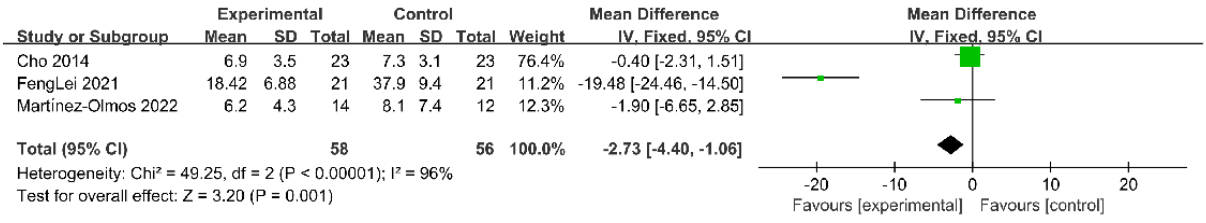

Five Sit-to-Stand Test

Two studies were conducted using the Sit-to-Stand Test 5 (STS-5) as a measurement outcome [15,16]. The results indicate that there was no significant difference in Sit-to-Stand Test performance between VR training and the control group (SD=0.57, 95% CI –0.87 to 2.02, *P*=.44) with low heterogeneity

XSL•FO
RenderX

(I^2=0%, *P*=.67; Figure 8 [15,16]). However, since there were fewer than three studies, publication bias was not assessed.

Figure 8. Forest plot of virtual reality on exercise capacity of Five Sit-to-Stand Test [15,16].

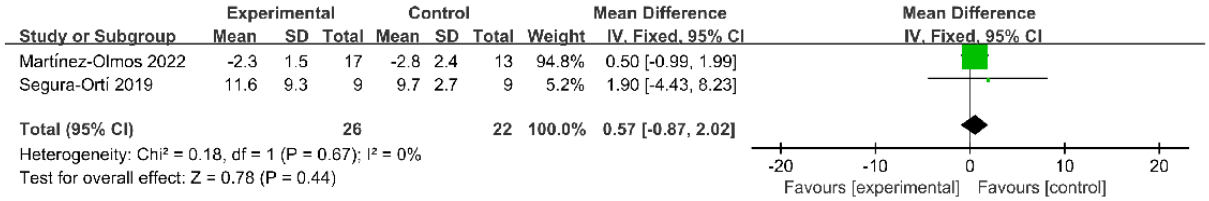

Ten Sit-to-Stand Test

In 2 trials [15,16], Sit-to-Stand Test 10 (STS-10) was used as a measurement outcome to compare the Sit-to-Stand Test performance between the VR training group and the control group. The results did not show any significant difference (SD=0.54, 95% CI –2.02 to 3.09, *P*=.68) with low heterogeneity (I^2=0%, *P*=.64; Figure 9 [15,16]). As there were fewer than 3 studies, publication bias was not assessed.

Figure 9. Forest plot of virtual reality on exercise capacity of Ten Sit-to-Stand Test [15,16].

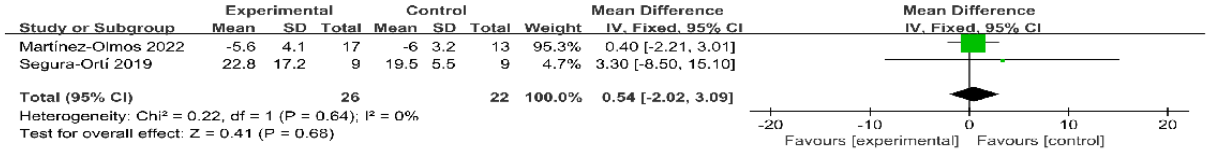

Sixty Sit-to-Stand Test

Three trials were conducted with the Sit-to-Stand Test 60 (STS-60) as the measurement outcome [15,16,24]. The results suggest that there was no significant difference in the STS-60 test performance between VR training and other training methods for dialysis patients (SD=0.55, 95% CI –0.79 to 1.88, *P*=.42) with middle heterogeneity (I^2=47%, *P*=.15; Figure 10 [16,18,24]). Egger regression test and funnel plot showed that there was no significant publication bias between the included studies (*P*=.57; Figure 6).

Figure 10. Forest plot of virtual reality on exercise capacity of Sixty Sit-to-Stand Test [16,18,24].

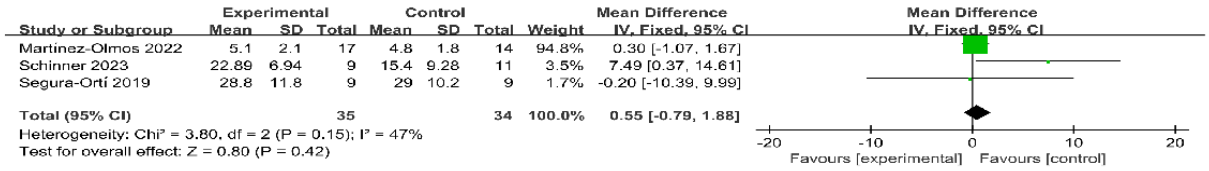

Timed Up and Go Test

Two trials included the Timed Up and Go Test (TUG) test as a measurement outcome [16,25]. The results showed that there was no significant difference between the VR training and control groups (SD=0.24, 95% CI –0.95 to 1.43, *P*=.69) with low heterogeneity (I^2=0%, *P*=.40; Figure 11 [16,25]). However, since there were fewer than 3 studies, publication bias was not assessed.

Figure 11. Forest plot of virtual reality on exercise capacity of Timed Up and Go Test [16,25].

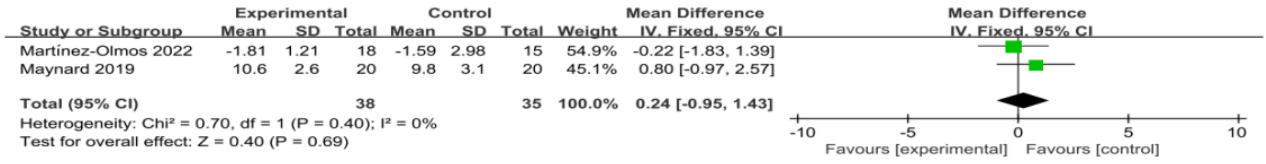

Psychological Outcomes

Depression

Four trials used depression scores as outcome measures [14,20,22,25]. The results show that patients undergoing VR training had lower depression symptoms compared with the control group, indicating that VR was more effective in improving the participants' depressive condition (SD=–6.30, 95% CI –7.14 to –5.47, *P*<.001) and considerable heterogeneity was found (I^2=96%, *P*<.001; Figure 12 [14,20,22,25]). Egger regression test and funnel plot showed that there was no

XSL•FO
RenderX

significant publication bias between the included studies (*P*=.34; Figure 6).

Figure 12. Forest plot of virtual reality on psychological outcome of depression [14,20,22,25].

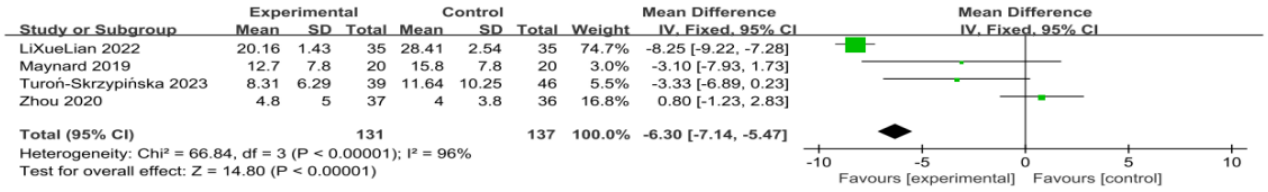

Anxiety

Two trials used anxiety scores as outcome measures [14,20]. The results show that patients undergoing VR training had lower anxiety levels compared with the control group, indicating that VR was more effective in improving the participants' anxiety condition (SD=–8.91, 95% CI –9.69 to –8.17, *P*<.001) with high heterogeneity (I^2=95%, *P*<.001; Figure 13 [14,20]). Since there were fewer than 3 studies, publication bias was not assessed.

Figure 13. Forest plot of virtual reality on psychological outcome of anxiety [14,20].

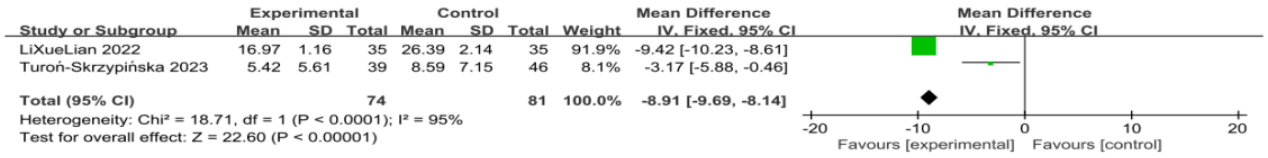

Fatigue

Two trials used fatigue scores as outcome measures [19,23]. The results indicate no significant difference in improving fatigue among dialysis patients between VR training and other training methods (SD=0.42, 95% CI –0.08 to 0.91, *P*=.01) with high heterogeneity (I^2=97%, *P*<.001; Figure 14 [19,23]). Since there were fewer than 3 studies, publication bias was not assessed.

Figure 14. Forest plot of virtual reality on psychological outcome of fatigue [19,23].

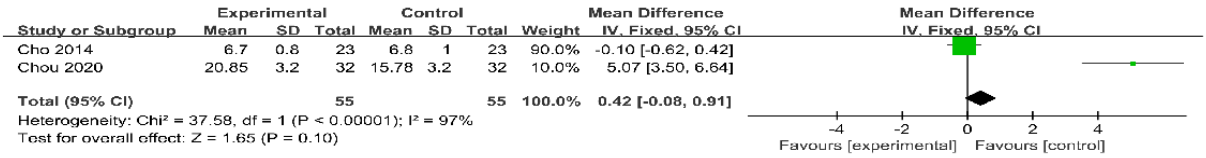

Social Functioning

Three trials used social functioning scores as outcome measures [18,20,25]. The results demonstrate that VR training was more effective than the control group interventions in improving patients' social functioning (SD=16.20, 95% CI 14.49-17.90, *P*<.001) with middle heterogeneity (I^2=72%, *P*=.03; Figure 15 [18,20,25]). Egger regression test and funnel plot showed that there was no significant publication bias between the included studies (*P*=.35; Figure 6).

Figure 15. Forest plot of virtual reality on social function [18,20,25].

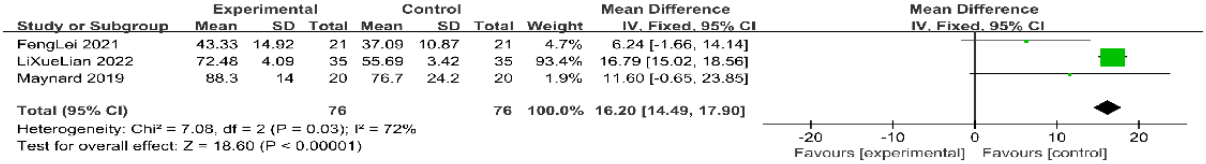

Self-Efficacy

Two trials used self-efficacy scores as outcome measures [17,18]. The results indicate that VR training was more effective than the control group interventions in improving patients' self-efficacy (SD=20.47, 95% CI 18.55-22.39, *P*<.001) with high heterogeneity (I^2=99%, *P*<.001; Figure 16 [17,18]). Since

XSL•FO
RenderX

there were fewer than 3 studies, publication bias was not assessed.

Figure 16. Forest plot of virtual reality on self-efficacy [17,18].

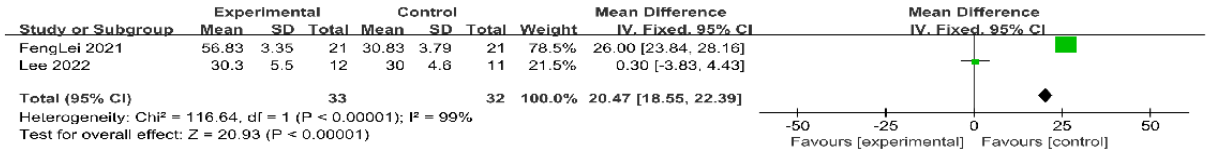

Discussion

Summary of Main Findings

This meta-analysis encompassed 12 trials involving 625 participants, of whom (99/625; 15.8%) were PD patients. All studies used VR training as the intervention, while control groups received conventional methods such as health education, routine care, and exercise adapted to patients' daily routines. Of these 12 studies, 5 reported the use of random sequence generation.

This meta-analysis offers a thorough evaluation of the effectiveness of VR training in comparison with traditional methods across various outcome measures, including motor abilities, balance, and psychological well-being. The findings indicate that VR training significantly enhances walking ability, as measured by the 6MWT, with an SD of 29.36. This effect is notably more pronounced compared with gait speed and other motor ability measures, where VR training did not produce significant improvements. Furthermore, VR training was linked to significant reductions in symptoms of depression and anxiety, along with enhancements in social functioning and self-efficacy. However, VR training did not show significant benefits for balance, Sit-to-Stand tests, or the TUG.

Physical Mobility

The substantial improvement observed in the 6MWT with VR training stands in stark contrast to the absence of significant results in gait speed. The 6MWT evaluates endurance and walking capability over a period, which may be well-suited to the interactive and engaging nature of VR environments. The immersive qualities of VR training may enhance patients' motivation and engagement during prolonged walking tasks, thereby contributing to better performance on the 6MWT. Silva et al [26] demonstrated that visual and auditory stimuli in virtual environments encourage active participation in training, leading to increased exercise, walking distance, and time, which enhances the effectiveness of the 6MWT and improves patients' walking ability. Research indicates that gait rehabilitation programs using VR can aid dialysis patients with muscle weakness and joint pain in improving walking speed and quality by simulating diverse walking environments and obstacles, thereby enhancing their independence in daily life [27]. The inclusion of dynamic, goal-oriented tasks in VR may better simulate real-world walking scenarios, motivating participants to exert greater effort and maintain activity for extended periods. In contrast, gait speed, a measure of immediate walking performance, may not benefit as substantially from the immersive features of VR training. The effectiveness of VR in enhancing gait speed may be influenced by the intensity and duration of training sessions, which varied across studies. Consequently, the lack of significant differences in gait speed improvement may suggest a need for more specific training protocols or supplementary interventions to optimize gait speed outcomes.

Psychological Symptoms

The observed significant improvement in walking ability among dialysis patients underscores the potential of VR training to address key aspects of patient care. Improving walking capability can significantly impact daily functioning and overall quality of life, particularly in individuals with mobility impairments [28]. The capacity to enhance physical function while concurrently addressing psychological factors, such as depression and anxiety, highlights the comprehensive benefits of VR training. Research by Schröder et al [29] further demonstrates that VR training can alleviate patients' negative emotions, including anxiety and depression. By offering a multifaceted therapeutic approach, VR has the potential to enhance both physical performance and mental well-being, essential components of holistic patient care. Furthermore, the effectiveness of VR in reducing depression and anxiety suggests that VR training may serve as a valuable tool for mitigating the broader psychosocial impacts of chronic illness [30]. The immersive and interactive features of VR provide an engaging and motivating environment that fosters positive changes in mood and self-perception, thereby contributing to improved overall patient outcomes [31].

Social Functioning

The findings of this meta-analysis demonstrate that VR training significantly improves social functioning in patients, as evidenced by a SD of 16.2. This result suggests that VR training can substantially enhance social interaction and integration, which are crucial aspects of overall well-being. The positive effect of VR on social functioning may be attributed to various factors inherent to the VR experience. VR environments frequently offer engaging and interactive scenarios that promote social engagement and collaboration, even among individuals with limited mobility or social anxiety. VR training usually includes interactive tasks and group-based activities that encourage active participation, thereby enhancing patients' social skills and interactions. The immersive nature of VR provides opportunities for patients to practice social scenarios in a controlled environment, thereby enhancing their confidence and comfort in real-world social interactions. Furthermore, participation in VR-based social activities may foster a sense of community and support, further enhancing social functioning. Despite these promising results, it is crucial to consider the variability in social functioning outcomes reported across

XSL•FO
RenderX

studies. The observed heterogeneity (I^2=72%) indicates that factors such as the design of VR interventions, the nature of social interactions, and the duration of training may affect the effectiveness of VR in enhancing social functioning. Future research should focus on exploring these variables in detail and identifying the specific components of VR training that most effectively improve social interaction and integration [32].

Self-Efficacy

Virtual reality training demonstrated a significant improvement in self-efficacy, with an SD of 20.47, indicating that VR interventions substantially enhance patients' confidence in their ability to manage various tasks and challenges. Self-efficacy, defined as the belief in one's ability to achieve goals, is a crucial determinant of both physical and psychological health. The observed improvement in self-efficacy with VR training can be attributed to the interactive and engaging nature of VR environments, which frequently include goal-setting, feedback, and achievement [33]. VR training typically integrates elements such as progress tracking, rewards, and virtual challenges, which can enhance patients' confidence in their abilities. By offering immediate feedback and a sense of accomplishment, VR can reinforce positive behaviors and attitudes, thereby contributing to improved self-efficacy. Furthermore, the immersive and supportive features of VR environments may help patients overcome fears and build confidence in a controlled setting, potentially translating into improved self-efficacy in real-world situations [34]. However, the high heterogeneity (I^2=99%) in the self-efficacy outcomes suggests that the effectiveness of VR in enhancing self-efficacy may be influenced by several factors, including the design of the VR interventions and individual patient characteristics. Future research should focus on identifying the most effective VR components for enhancing self-efficacy and investigating how different patient characteristics may influence the outcomes of VR training.

Challenges and Limited Use of VR in PD Patients

The use of VR training among PD patients has been relatively limited compared with other patient populations. Several factors are likely to contribute to this limited use. PD patients face distinct challenges, such as the requirement for specialized equipment and potential complications related to fluid management, which may complicate the integration of VR training. In addition, this patient population may have unique physical and psychological conditions that necessitate tailored interventions. Current research on VR training for this population is still emerging, and the application of VR may need adaptation to address these specific needs effectively.

Furthermore, PD patients may experience varying levels of physical capability and psychological stress compared with other patient groups, which may influence the outcomes of VR interventions. Future research should focus on developing and testing VR programs specifically tailored for this population, considering their unique requirements and limitations to ensure that the interventions are both safe and effective.

Advantages and Limitations

This meta-analysis highlights the potential of VR training as an adjunctive therapy for enhancing walking ability and psychological well-being in dialysis patients. The substantial improvements in walking endurance and psychological outcomes observed in this analysis suggest that VR training may be a valuable addition to existing therapeutic strategies. Nevertheless, the observed variability in results and heterogeneity in certain outcomes underscores the need for further research.

The strengths of this study include (1) it is the first meta-analysis to assess the impact of VR training on dialysis patients and (2) we assessed the effectiveness of VR training using 12 outcome measures: gait speed, balance, 6-minute walk test score, STS-5, STS-10, STS-60, TUG, depression, anxiety, fatigue, social functioning, and self-efficacy, providing a comprehensive reference for future research.

This study also has limitations: (1) it included only published articles in Chinese and English, which may impact the meta-analysis results, (2) some studies did not provide details on sequence generation, allocation concealment, and blinding, and (3) the 12 studies included in the analysis used different control interventions and the intervention durations across the studies varied significantly, which may have contributed to significant heterogeneity among the studies.

Future research should focus on refining VR interventions, exploring their optimal implementation, and addressing the specific needs of diverse patient populations. Larger and more diverse sample sizes, coupled with standardized training protocols, will be crucial for comprehensively evaluating the benefits and limitations of VR.

In this study, we used multiple databases for literature retrieval to ensure comprehensive coverage of relevant research in the field. Our selected databases included CNKI, Wanfang, China Science and Technology Journal Database, SinoMed, Cochrane Library, Web of Science, PubMed, and Embase. These databases offer a wealth of research outputs, particularly the Chinese databases, which aggregate a substantial amount of high-quality studies conducted by Chinese scholars, reflecting the region's active contribution to this field. However, we recognize that our current search strategy may not fully represent the entire research landscape, particularly the lack of studies from North America.

Although our initial search did not yield relevant North American research, this may be attributed to the limitations of our search strategy and keyword selection. We acknowledge that this omission could impact the representativeness of our results and potentially influence the conclusions drawn from the study. Therefore, we emphasize this limitation within the article and recommend that future research consider broader literature retrieval strategies to ensure a more holistic understanding of the field.

This reflection provides important guidance for subsequent studies, ensuring that future efforts can better integrate research findings from diverse regions, thereby enhancing the breadth and depth of analysis.

XSL•FO
RenderX

Conclusions

This meta-analysis offers a thorough evaluation of the use of VR training in dialysis patients. The findings indicate that VR training effectively enhances patients' physical fitness, alleviates anxiety and depression, and improves adherence, thereby facilitating better self-management. However, research focusing on VR training in peritoneal patients is relatively limited compared with that on hemodialysis patients. Future research should involve PD patients and use rigorous, large-scale studies to validate the findings of this review for both hemodialysis and PD populations.

Acknowledgments

The authors would like to acknowledge that all research and manuscript preparation were conducted by the authors themselves, with no external assistance.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1-3.

Authors' Contributions

XK contributed to writing the original draft, data curation, validation, software, formal analysis, and investigation. YZ managed writing the original draft, data curation, visualization, software, formal analysis, and investigation. CS handled validation, investigation, and data curation. ZC and JZ contributed to investigation and data curation. RZ managed conceptualization, data curation, writing—review and editing, supervision, project administration, and visualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 211 KB - jmir_v27i1e58384_app1.pdf]

Multimedia Appendix 2
Search strategies of PubMed, the Cochrane Library, the China National Knowledge Infrastructure (CNKI) and the VIP Journal Database. Supplementary information of intervention in experimental and control groups.
[DOC File , 17 KB - jmir_v27i1e58384_app2.doc]

Multimedia Appendix 3
Summary of included studies.
[DOCX File , 27 KB - jmir_v27i1e58384_app3.docx]

References

1. Zhang L, Zhao MH, Zuo L, Wang Y, Yu F, Zhang H, CK-NET Work Group. China Kidney Disease Network (CK-NET) 2015 Annual Data Report. Kidney Int Suppl (2011) 2019 Mar 01;9(1):e1-e81 [FREE Full text] [doi: 10.1016/j.kisu.2018.11.001] [Medline: 30828481]
2. Li-xia L, Xiao-lei H, Bao-chun G, Jing Q, Lian-yu Z, Xin-zhou Z. EEffects of aerobic exercise on oxidative stress and quality of life in maintenance hemodialysis patients. Chinese Journal of Blood Purification 2013 Feb 26;12(08):419-424. [doi: 10.3969/j.issn.1671-4091.2013.08.004]
3. Orcy R, Antunes MF, Schiller T, Seus T, Böhlke M. Aerobic exercise increases phosphate removal during hemodialysis: a controlled trial. Hemodial Int 2014 Jan 20;18(2):450-458. [doi: 10.1111/hdi.12123] [Medline: 24438516]
4. Henrique DMN, Reboredo MDM, Chaoubah A, Paula RBD. [Aerobic exercise improves physical capacity in patients under chronic hemodialysis]. Arq Bras Cardiol 2010 May 28;94(6):823-828 [FREE Full text] [doi: 10.1590/s0066-782x2010005000043] [Medline: 20428710]
5. QIGE Q, JIANRONG Z. The effects of combined aerobic-resistance training on dialysis adequacy and quality of life in maintenance hemodialysis patients. Chinese Journal of Blood Purification 1976 Mar 18:646-651. [Medline: 1671]
6. de Lima MC, Cicotoste CDL, Cardoso KDS, Forgiarini LA, Monteiro MB, Dias AS. Effect of exercise performed during hemodialysis: strength versus aerobic. Ren Fail 2013 Apr 08;35(5):697-704 [FREE Full text] [doi: 10.3109/0886022X.2013.780977] [Medline: 23560491]

XSL•FO
RenderX

7. JIANYING L, YANLIN H, YANJUAN T. Application progress of exercise therapy in hemodialysis patients with uremia. The Journal of Medical Theory and Practice 2012 Jun 26;35(5):697-704 [FREE Full text] [doi: 10.19381/j.issn.1001-7585.2013.01.012]
8. Peeters D. Virtual reality: A game-changing method for the language sciences. Psychon Bull Rev 2019 Feb 07;26(3):894-900 [FREE Full text] [doi: 10.3758/s13423-019-01571-3] [Medline: 30734158]
9. Colombo D, Díaz-García A, Fernandez-Álvarez J, Botella C. Virtual reality for the enhancement of emotion regulation. Clin Psychol Psychother 2021 Jun 23;28(3):519-537. [doi: 10.1002/cpp.2618] [Medline: 34048621]
10. Ahmadpour N, Randall H, Choksi H, Gao A, Vaughan C, Poronnik P. Virtual Reality interventions for acute and chronic pain management. Int J Biochem Cell Biol 2019 Jul 12;114:105-568. [doi: 10.1016/j.biocel.2019.105568] [Medline: 31306747]
11. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:1-9 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
12. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
13. Weber L, Kamp-Becker I, Christiansen H, Mingebach T. Treatment of child externalizing behavior problems: a comprehensive review and meta-meta-analysis on effects of parent-based interventions on parental characteristics. Eur Child Adolesc Psychiatry 2018 Jun 28;28(8):1025-1036. [doi: 10.1007/s00787-018-1175-3] [Medline: 29948228]
14. Turoń-Skrzypińska A, Tomska N, Mosiejczuk H, Rył A, Szylińska A, Marchelek-Myśliwiec M, et al. Impact of virtual reality exercises on anxiety and depression in hemodialysis. Sci Rep 2023 Aug 01;13(1):12435-12435 [FREE Full text] [doi: 10.1038/s41598-023-39709-y] [Medline: 37528161]
15. Segura-Ortí E, Pérez-Domínguez B, Ortega-Pérez de Villar L, Meléndez-Oliva E, Martínez-Gramage J, García-Maset R, et al. Virtual reality exercise intradialysis to improve physical function: A feasibility randomized trial. Scand J Med Sci Sports 2019 Jan 01;29(1):89-94 [FREE Full text] [doi: 10.1111/sms.13304] [Medline: 30230041]
16. Martínez-Olmos FJ, Gómez-Conesa AA, García-Testal A, Ortega-Pérez-de-Villar L, Valtueña-Gimeno N, Gil-Gómez JA, et al. An intradialytic non-immersive virtual reality exercise programme: a crossover randomized controlled trial. Nephrol Dial Transplant 2022 Jun 23;37(7):1366-1374. [doi: 10.1093/ndt/gfab213] [Medline: 34245292]
17. Lee CMS, Fong KNK, Mok MMY, Lam MK, Kung Y, Chan PPW, et al. Application of virtual reality for peritoneal dialysis exchange learning in patients with end-stage renal disease and cognitive impairment. Virtual Real 2022 Dec 03:1-13 [FREE Full text] [doi: 10.1007/s10055-022-00728-1] [Medline: 36533192]
18. LEI F, YUNSHU L, JIE Y. Virtual reality somatosensory games integrated into the home exercise management and its effects on fall risk and quality of life in dialysis patients [J]. Chinese Journal of Blood Purification 2021 May 12:302-305. [doi: 10.1056/NEJM197603182941206] [Medline: 1671]
19. Chou HY, Chen SC, Yen TH, Han HM. Effect of a Virtual Reality-Based Exercise Program on Fatigue in Hospitalized Taiwanese End-Stage Renal Disease Patients Undergoing Hemodialysis. Clin Nurs Res 2020 Jul 01;29(6):368-374. [doi: 10.1177/1054773818788511] [Medline: 30009636]
20. XUELIAN L, JIANYING J, LING L. Application effect of virtual reality technology on psychological negative emotions in maintenance hemodialysis patients. China Modern Doctor 2022 Jan 25:110-113. [Medline: 110]
21. Lina W, KE L. Intervention effect of exercise mode based on virtual reality technology on physical activity of peritoneal dialysis patients [J]. Chinese Nursing Research 2023 Oct 01;37(20):3727-3731. [doi: 10.12102/j.issn.1009-6493.2023.20.022]
22. Zhou H, Al-Ali F, Kang GE, Hamad AI, Ibrahim RA, Talal TK, et al. Application of Wearables to Facilitate Virtually Supervised Intradialytic Exercise for Reducing Depression Symptoms. Sensors (Basel) 2020 Mar 12;20(6):1571-1571 [FREE Full text] [doi: 10.3390/s20061571] [Medline: 32178231]
23. Cho H, Sohng KY. The effect of a virtual reality exercise program on physical fitness, body composition, and fatigue in hemodialysis patients. J Phys Ther Sci 2014 Oct 28;26(10):1661-1665 [FREE Full text] [doi: 10.1589/jpts.26.1661] [Medline: 25364137]
24. Schinner L, Nagels K, Scherf J, Schmaderer C, Heemann U, Küchle C, et al. Intradialytic neuromuscular electrical stimulation with optional virtual reality distraction improves not only muscle strength and functional capacity but also serum albumin level in haemodialysis patients: a pilot randomized clinical trial. BMC Nephrol 2023 Aug 23;24(1):240-246 [FREE Full text] [doi: 10.1186/s12882-023-03283-2] [Medline: 37608265]
25. Maynard LG, de Menezes DL, Lião NS, de Jesus EM, Andrade NLS, Santos JCD, et al. Effects of Exercise Training Combined with Virtual Reality in Functionality and Health-Related Quality of Life of Patients on Hemodialysis. Games Health J 2019 Oct 01;8(5):339-348. [doi: 10.1089/g4h.2018.0066] [Medline: 31539293]
26. Silva PMM, Silva LWM, Vieira ER, Cavalvanti FAC, Morya E. Go Across Immersive Technology: A Preliminary Study of the Design and Development of a System for Gait Training Using Virtual Reality. Games Health J 2023 Dec 06;12(6):472-479. [doi: 10.1089/g4h.2023.0101] [Medline: 37410502]
27. Ehgoetz Martens KA, Matar E, Phillips JR, Shine JM, Grunstein RR, Halliday GM, et al. Narrow doorways alter brain connectivity and step patterns in isolated REM sleep behaviour disorder. Neuroimage Clin 2022 Feb 07;33:102958-102959 [FREE Full text] [doi: 10.1016/j.nicl.2022.102958] [Medline: 35151040]

XSL•FO
RenderX

28. Gurz D, Coimbatore Dada K, Naga Nyshita V, Aderibigbe FD, Singh M, Yadav KP, et al. The Impact of Virtual Reality (VR) Gaming and Casual/Social Gaming on the Quality of Life, Depression, and Dialysis Tolerance in Patients With Chronic Kidney Disease: A Narrative Review. Cureus 2023 Sep 08;15(9):44904-44910 [FREE Full text] [doi: 10.7759/cureus.44904] [Medline: 37814762]
29. Schröder D, Wrona KJ, Müller F, Heinemann S, Fischer F, Dockweiler C. Impact of virtual reality applications in the treatment of anxiety disorders: A systematic review and meta-analysis of randomized-controlled trials. J Behav Ther Exp Psychiatry 2023 Dec 01;81:101893-101893 [FREE Full text] [doi: 10.1016/j.jbtep.2023.101893] [Medline: 37453405]
30. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan 01;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
31. Kiper P, Przysiężna E, Cieślik B, Broniec-Siekaniec K, Kucińska A, Szczygieł J, et al. Effects of Immersive Virtual Therapy as a Method Supporting Recovery of Depressive Symptoms in Post-Stroke Rehabilitation: Randomized Controlled Trial. Clin Interv Aging 2022 Nov 23;17:1673-1685 [FREE Full text] [doi: 10.2147/CIA.S375754] [Medline: 36447623]
32. Park M, Ko MH, Oh SW, Lee JY, Ham Y, Yi H, et al. Effects of virtual reality-based planar motion exercises on upper extremity function, range of motion, and health-related quality of life: a multicenter, single-blinded, randomized, controlled pilot study. J Neuroeng Rehabil 2019 Oct 24;16(1):3443-3445. [doi: 10.1186/s12984-019-0595-8] [Medline: 31651335]
33. Välimäki M, Hätönen HM, Lahti ME, Kurki M, Hottinen A, Metsäranta K, et al. Virtual reality for treatment compliance for people with serious mental illness. Cochrane Database Syst Rev 2014 Oct 08;2014(10):1-40 [FREE Full text] [doi: 10.1002/14651858.CD009928.pub2] [Medline: 25300174]
34. van der Linde-van den Bor M, Slond F, Liesdek OCD, Suyker WJ, Weldam SWM. The use of virtual reality in patient education related to medical somatic treatment: A scoping review. Patient Educ Couns 2022 Jul 15;105(7):1828-1841 [FREE Full text] [doi: 10.1016/j.pec.2021.12.015] [Medline: 35000833]

Abbreviations

6MWT: 6-minute walk test
CNKI: China National Knowledge Infrastructure
ESRD: end-stage renal disease
HD: hemodialysis
PD: peritoneal dialysis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SD: standardized mean difference
STS-10: Sit-to-Stand Test 10
STS-5: Sit-to-Stand Test 5
STS-60: Sit-to-Stand Test 60
TUG: Timed Up and Go Test
VR: virtual reality
VRT: virtual reality training

Edited by A Coristine; submitted 14.03.24; peer-reviewed by Q Gao, L Pérez-Gisbert; comments to author 18.07.24; revised version received 11.09.24; accepted 13.11.24; published 08.01.25.

Please cite as:
Kang X, Zhang Y, Sun C, Zhang J, Che Z, Zang J, Zhang R
Effectiveness of Virtual Reality Training in Improving Outcomes for Dialysis Patients: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e58384
URL: https://www.jmir.org/2025/1/e58384
doi:10.2196/58384
PMID:

©Xin Kang, Yiping Zhang, Chaonan Sun, Jiaxin Zhang, Zhe Che, Jinhui Zang, Rongzhi Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Extended Reality Interventions for Health and Procedural Anxiety: Panoramic Meta-Analysis Based on Overviews of Reviews

Tom Arthur[1], PhD; GJ Melendez-Torres[1], DPhil; David Harris[1], PhD; Sophie Robinson[1], MA; Mark Wilson[1], PhD; Sam Vine[1], PhD

Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom

Corresponding Author:
Tom Arthur, PhD
Faculty of Health and Life Sciences
University of Exeter
St. Luke's Campus
Heavitree Road
Exeter, EX1 2LU
United Kingdom
Phone: 44 1392 722892
Email: T.T.Arthur@exeter.ac.uk

Abstract

Background: Extended reality (XR) technologies are increasingly being used to reduce health and procedural anxieties. The global effectiveness of these interventions is uncertain, and there is a lack of understanding of how patient outcomes might vary between different contexts and modalities.

Objective: This research used panoramic meta-analysis to synthesize evidence across the diverse clinical contexts in which XR is used to address common outcomes of health and procedural anxiety.

Methods: Review-level evidence was obtained from 4 databases (MEDLINE, Embase, APA PsycINFO, and Epistemonikos) from the beginning of 2013 until May 30, 2023. Reviews that performed meta-analysis of randomized controlled trials relating to patient-directed XR interventions for health and procedural anxiety were included. Studies that analyzed physiological measures, or focused on technologies that did not include meaningful immersive components, were excluded. Furthermore, data were only included from studies that compared intervention outcomes against no-treatment or treatment-as-usual controls. Analyses followed a preregistered, publicly available protocol. Trial effect sizes were extracted from reviews and expressed as standardized mean differences, which were entered into a 3-level generalized linear model. Here, outcomes were estimated for patients (level 1), studies (level 2), and anxiety indications (level 3), while meta-regressions explored possible influences of age, immersion, and different mechanisms of action. Where relevant, the quality of reviews was appraised using the AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews, Revised Instrument) tool.

Results: Data from 83 individual trials were extracted from 18 eligible meta-analyses. Most studies involved pediatric patient groups and focused on procedural, as opposed to general, health anxieties (eg, relating to needle insertion, dental operations, and acute surgery contexts). Interventions targeted distraction-, education-, and exposure-based mechanisms, and were provided via a range of immersive and nonimmersive systems. These interventions proved broadly effective in reducing patient anxiety, with models revealing significant but heterogeneous effects for both procedural (d=–0.75, 95% CI –0.95 to –0.54) and general health (d=–0.82, 95% CI –1.20 to –0.45) indications (when compared with nontreatment or usual-care control conditions). For procedural anxieties, effects may be influenced by publication bias and appear more pronounced for children (vs adults) and nonimmersive (vs immersive) technology interventions, but they were not different by indication.

Conclusions: Results demonstrate that XR interventions have successfully reduced patient anxiety across diverse clinical contexts. However, significant uncertainty remains about the generalizability of effects within various unexplored indications, and existing evidence is limited in methodological quality. Although current research is broadly positive in this area, it is premature to assert that XR interventions are effective for any given health or procedural anxiety indication.

(J Med Internet Res 2025;27:e58086) doi:10.2196/58086

XSL•FO
RenderX

KEYWORDS

virtual reality; exposure therapy; distraction techniques; patient education; fear; phobia; immersive; anxiety; meta-analysis; extended reality; intervention; procedural anxiety; patient anxiety; exposure; distraction

Introduction

Health and procedural anxiety can lead to delayed engagement with treatment, reduced treatment adherence, worsening symptoms, and compromised quality of life. Two primary approaches to alleviate health and procedural anxiety are exposure and distraction therapies. Distraction-based approaches seek to reduce anxiety by diverting attention away from distressing thoughts, situations, or physical sensations. While controversial as a long-term technique [1], these are effective for short-term anxiety reduction [2]. However, distraction methods may not always be readily available or accessible, particularly in moments of acute anxiety/stress. Exposure therapy has been traditionally delivered in vivo. By directly facing the fearful stimuli without resorting to escape, avoidance, or rituals, the patient's connection between the stimulus and memory structures is modified, leading to a potential reduction in the elicited anxiety response [3]. While effective, in vivo exposure has several limitations; it can be difficult to instigate when the anxiety-inducing environment or stimulus is inaccessible and can be time-consuming and expensive to simulate. There is, therefore, a drive toward the use of technologies that can deliver these therapies in a more cost-effective and accessible manner.

Extended reality (XR) technologies—including virtual reality, augmented reality, and mixed reality—are well suited to deliver these therapies [4]. XR immerses users in a digital environment, providing a heightened sense of presence and interaction with synthetic sensory stimuli. These simulations can feel realistic, enabling users to practice skills or experience environments in a controlled and immersive setting that can ultimately aid the therapeutic process. For example, XR can gradually expose patients to anxiety-inducing cues, to elicit realistic emotional responses that can become habituated over time, and they can also supply experiential patient education material about a procedure or condition. Similarly, the highly immersive and interactive nature of XR makes it an ideal distraction tool, which can divert attention toward more pleasurable or relaxing sensory environments. From a practical perspective, XR interventions can be used to address wide-ranging anxiety indications, given the breadth of stimuli that can now be simulated effectively and the capability to be delivered remotely (eg, headsets could be posted to patient's homes before attending appointments). They are also considered highly engaging forms of therapy [5] that can provide personalized patient experiences (eg, by tailoring exposure- and distraction-based cues to individual user data), so they may promote better treatment adherence [6]. XR is therefore a scalable tool that has the potential to enhance the accessibility of therapeutic interventions for patients who may be unwilling or unable to participate in more conventional forms of psychological treatment [7].

Panoramic meta-analyses are a tool to integrate evidence from trials covering the same intervention and the same outcomes over a range of indications. Originally designed for use in the context of overviews of reviews (ie, systematic reviews of systematic reviews), early panoramic meta-analyses addressed topics such as sutures versus staples for wound healing in a range of surgical sites [8]. More recently, panoramic meta-analyses have also addressed mental health outcomes, such as cognitive behavioral therapy for health-related quality of life [9]. When used to evaluate evidence for a given intervention method, panoramic meta-analyses offer a statistical advantage for random effects modeling, in that they generate more stable estimates of between-study variance. But more importantly, they generate "global" estimates of effectiveness that allow for (1) assessment of likely future effects via prediction intervals; (2) assessment of the relative importance of indication in explaining effect heterogeneity via variance partitioning; and (3) tests of effect modification (ie, via meta-regression) that may be better powered by drawing on evidence from multiple indications. Thus, in this research, panoramic meta-analyses can indicate the broad impact that XR interventions have on patient anxiety across diverse clinical contexts, while also offering unique insight into the exchangeability of effects between different intervention types, technologies, populations, and indications. From an applied perspective, one can use this insight to inform future clinical programs (eg, by deciding whether to adopt XR based on pooled data that reflects efficacy or inefficacy over general/specific contexts, or by using findings that indicate whether certain indications/methods are associated with greater efficacy or uncertainty).

Consequently, a panoramic meta-analysis was undertaken to synthesize evidence across the diverse health-related contexts in which XR is used to address common outcomes of health and procedural anxiety. The aim of this analysis was to consider the global effectiveness of XR, to assess the likelihood of effectiveness for other clinical indications, and to explore whether key intervention characteristics are associated with varying levels of effectiveness.

Methods

Overview

A protocol for this review was registered on the Open Science Framework [10], a widely used, publicly available platform that permits the open sharing of research data and material. This analysis is part of a larger evidence-synthesis project focusing on XR interventions for health and procedural anxiety.

Included Studies

A set of systematic reviews was identified, which included meta-analyses of randomized controlled trials relating to patient-directed XR interventions for health and procedural anxiety outcomes published from 2013 onward. In line with the panoramic meta-analysis approach, these reviews informed the sample of trials and effect sizes that were subsequently examined in generalized linear models (see the Statistical Methods section

XSL•FO
RenderX

below). Specific inclusion and exclusion criteria are detailed in Table 1. Procedural anxiety was defined as anxiety both related and proximal to specific medical procedures, generally as a measure of state anxiety [11]. Health anxiety was defined as anxiety relating to broader health conditions, generally as a measure of trait anxiety [12]. Studies that based their analyses on physiological measures of anxiety were excluded (eg, blood pressure and heart rate), and data were only extracted that compared intervention outcomes against no-treatment or treatment-as-usual controls. Eligibility criteria did not impose any further stipulations for the trial-level data, and the extraction of effect sizes was not restricted to studies from any given language, date, or region.

Table 1. Eligibility criteria.

Characteristics	Inclusion criteria	Exclusion criteria
Study design	• Systematic reviews[a] of randomized controlled trials that included meta-analyses or forest plots with reported effect sizes • Reviews published as full-text articles in peer-reviewed journals from 2013 onward[b] • Randomized controlled trials that feature in these systematic reviews	• Reviews that did not perform adequate searches of an electronic database, using the structured search query, with inclusion/exclusion criteria specified • Conference abstracts, dissertations, and theses • Nonrandomized trials
Interventions	• Focus on patient-directed XR[c] interventions[d] in any primary or discrete subgroup analyses	• Focus on technologies that do not include meaningful immersive components (eg, mobile apps)
Comparators	• Data comparisons against no-treatment or treatment-as-usual controls	• None
Outcome	• Patient health or procedural anxiety measures	• Trial data relating to pain outcomes or physiological indicators of anxiety

[a]Definitions of systematic reviews were based on criteria developed for the Database of Abstracts of Reviews of Effects.

[b]Trial eligibility was not restricted to studies from any given date.

[c]XR: extended reality.

[d]XR interventions included virtual reality, mixed reality, or augmented reality systems.

Search and Selection

Searches were carried out on May 30, 2023, in MEDLINE and Embase (on the Ovid platform), APA PsycINFO (in Ovid), and Epistemonikos. This variety of databases was selected to cover general medical literature, specialist psychology sources, and a collection of systematic reviews. Searches used appropriate subject headings and keywords for the intervention (virtual reality), outcome (health anxiety), and study design (systematic reviews), with full terms and methods specified in Multimedia Appendix 1. Records were screened independently and in duplicate, both at the title and abstract stage and at full-text stage, recording reasons for exclusion at full-text and involving a third reviewer in case of disagreement.

Data Extraction, Risk of Bias, and Classification

Informed by included reviews and relevant trials within them, a set of high-level indications were inductively generated as distinct “use cases” of XR. Different classifications were developed for procedural anxiety and general health anxiety. Within each indication and outcome, the highest quality, the most recent review was prioritized for effect size extraction from included forest plots, defining quality using the results of appraisal with the AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews, Revised Instrument) [13]. The analysis then proceeded through the remaining reviews sequentially, organizing reviews by publication year and, within year, by quality, until no additional trials were identified, extracting evidence from randomized controlled trials only and preferring effect estimates for the same trial and outcome from higher-priority reviews. Where available, trial-level risk of bias (RoB) assessments were extracted, from included reviews providing sufficient evidence. Extraction items are provided in the protocol. The original trial was consulted when (1) extracted effect sizes were inappropriate or possibly inaccurate (eg, a standardized mean difference greater than 2) or (2) effect sizes were not presented as standardized mean differences and relevant standard deviations were not presented in included plots. One reviewer led extraction with auditing by a second.

Statistical Methods

All effect estimates were expressed as standardized mean differences, where negative values are positive (ie, greater reduction in anxiety). For each outcome, a 1-step meta-analysis model was estimated using a 3-level generalized linear model, where level 1 (patients within trials) is implied, level 2 is the study-level effect estimate, and level 3 is the indication. All models were estimated using restricted maximum likelihood and random effects at level 2 and level 3, with a common level 2 variance component across all indications to stabilize estimation. Heterogeneity was quantified using a standard variance partitioning method, generating a between-study, within-indication I^2, and a between-indication I^2 based on the Higgins and Thompson method. Statistical analyses were undertaken on Stata (version 18; StataCorp).

XSL•FO
RenderX

Sensitivity and Exploratory Analyses

The registered protocol stated that for any instances containing too few indications in an outcome (fewer than 10), sensitivity analysis would be undertaken with a 2-level model, using fixed effects for indication at level 2. This model was compared with an unconditional meta-analysis model using a likelihood ratio test as an additional assessment of the value of indication as a grouping variable. Additionally, protocol-specified meta-regressions by age group (adult vs children), system immersion (immersive vs nonimmersive), and mechanism of action (relaxation or distraction vs exposure or education) were undertaken, where sufficient variation was present. Finally, funnel plots were drawn as an exploratory analysis, and an Egger test was estimated to examine small-study bias where the number of trials for any one indication and outcome was greater than 10.

Results

Search Results

A total of 18 meta-analyses informed the sample of trials, of which 10 provided procedural anxiety outcomes and 8 provided general health anxiety outcomes. Across these meta-analyses, 53 trial effects for procedural health anxiety outcomes and 30 trial effects for general health anxiety outcomes were examined. Two studies [14,15] reported data from multiple XR interventions, which were examined separately. Analyses excluded 15 trials that other meta-analyses included: 14 did not meet the specified definitions of XR, and one used physiological measures only. Study selection processes are outlined in Figure 1, via a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (see full list of included reviews in Multimedia Appendix 2 [16-33]).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of retrieved, screened, and included studies.

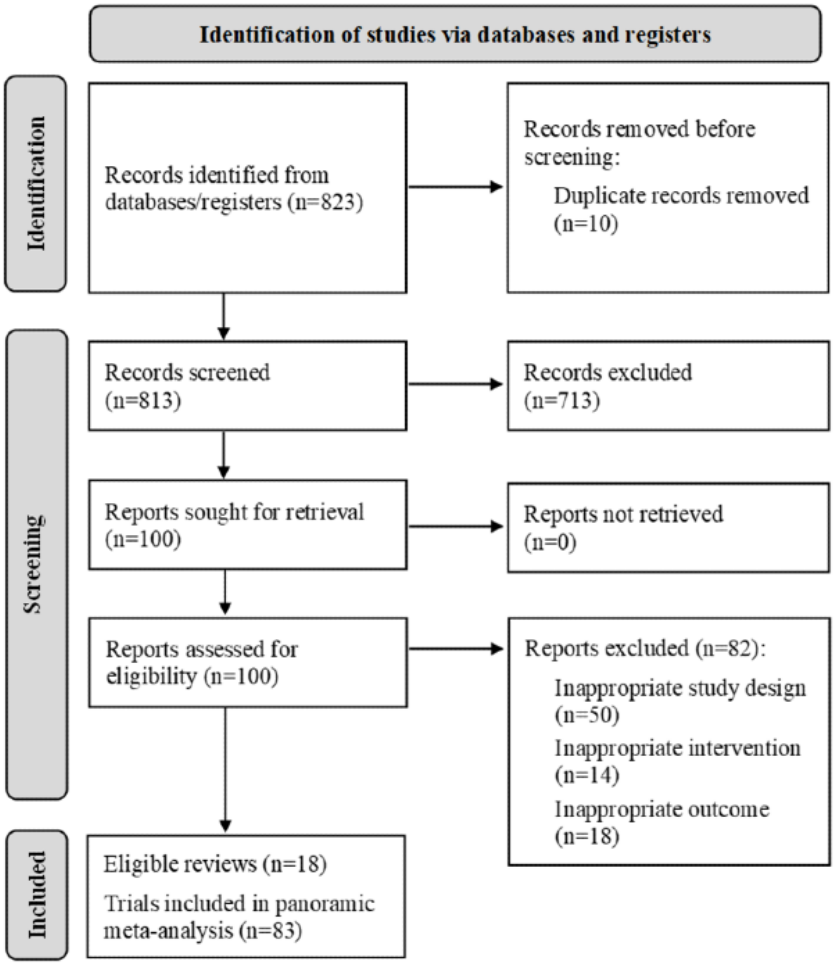

Trial Characteristics

Details of study-level extraction, measures used, and interventions are provided in Multimedia Appendix 3. Reviews examined anxiety indications relating to either a specific medical procedure (n=8) or those additionally relating to general health conditions and prolonged treatment pathways (n=10). The average number of trials examined within each review was 6.67 (SD 3.38), although 18 overlapping trials were detected in multiple meta-analyses. Trial-specific RoB assessments or quality appraisals were reported in all but 3 of the included studies and are presented in Multimedia Appendix 4. These methodological evaluations followed inconsistent assessment criteria (eg, the Cochrane RoB, ROBINS-I, and Delphi List tools); however, only 8 trials were considered to have low RoB overall and 49 were deemed to have a high risk or some concerns. Methods for assessing anxiety notably varied between trials and were dependent on clinical indications. The most common measures were the Modified Dental Anxiety Scale (for

XSL•FO
RenderX

dental studies), the State-Trait Anxiety Inventory (for cancer treatment, cardiac rehabilitation), the Modified Yale Preoperative Anxiety Scale (for nondental surgical operations), the Self-rating Anxiety Scale (for maternity), the Observational Scale of Behavioral Distress for radiographic procedures (for imaging), the Pain Anxiety Symptoms Scale (for chronic pain conditions), and simple visual analog scales (for wound care, needle-related anxieties).

Most of the included trial analyses (44/83, 53%) evaluated anxiety in pediatric patients only, while 36 trials focused on adults only, and 3 trials involved mixed-age samples (ie, children and adults). From an intervention perspective, the vast majority of procedural anxiety interventions (46/53, 55%) included just a single XR-based therapy session, the duration of which varied from 1 minute [34] to 90 minutes [35] (based on the context and population). Interventions for general health anxieties were more heterogeneous in design, ranging from single sessions to repeated activities over a period of months (Multimedia Appendix 3). Immersive forms of XR were adopted in 37 study interventions, whereas 46 interventions used nonimmersive methods. Interventions incorporated a range of XR technologies, including standalone virtual reality systems, portable glasses, mobile phone head-mounted displays, and monitor-based simulation devices.

In total, 66 interventions were classed as distraction-based or relaxation-based. These included natural scenery experiences (28 trial interventions), adventure or rollercoaster games (22 trial interventions), cartoons (12 trial interventions), mindfulness applications (1 trial intervention), and 3D images/videos (2 trial interventions). Conversely, 16 interventions were classified as education- or exposure-based. Separations between education- and exposure-based methods were not made, as numerous interventions incorporated elements of both components, and studies rarely provided sufficient details or rationale about XR procedures. This was also the case for distraction- and relaxation-based approaches (although it is also worth noting that all distraction experiences could be considered somewhat relaxing relative to anxiety-inducing medical stimuli).

Procedural Anxiety

Overview of Trials

Classification of trials measuring procedural anxiety outcomes identified 5 indications: dental anxiety (n=17), imaging (n=1), needle-related anxiety (n=17), nondental surgical operations (n=16), and anxiety related to wounds and burn care (n=2). Indication-specific meta-analyses are presented in Figure 2. In brief, all interventions generated significant effects within indication, except for interventions for wounds and burn care.

XSL•FO
RenderX

Figure 2. Forest plot for procedural anxiety outcomes.

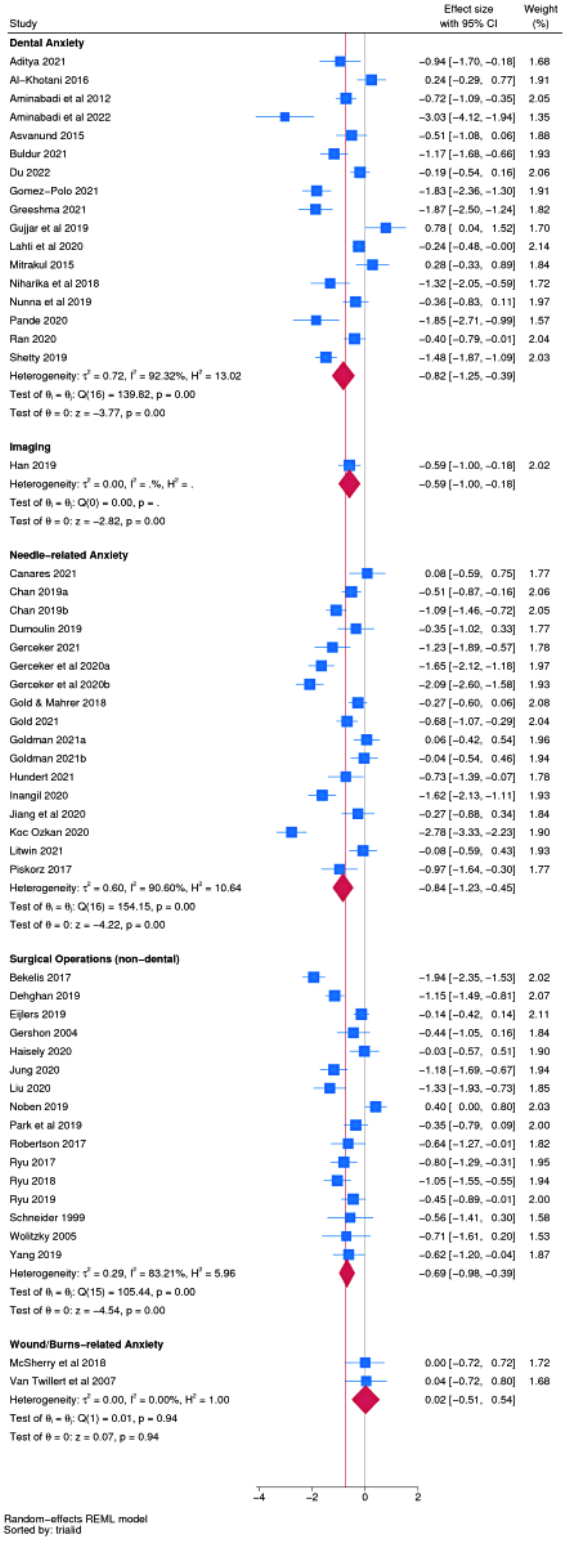

Panoramic Meta-Analysis

A panoramic meta-analysis model nesting trials within indication generated an overall pooled effect (*d*) of –0.75 (95% CI –0.95 to –0.54) (Table 2). The between-indication (level 3) variance parameter for this model was equal to 0, whereas the within-indication, between-trial (level 2) variance parameter was equal to 0.48. Thus, the I^2 attributable to indication was 0%, and the I^2 attributable to between-trial differences was 89.0%. However, a prediction interval for the pooled effect spanned –2.11 to 0.62, indicating that effectiveness in a future trial would be highly uncertain.

XSL•FO
RenderX

Table 2. Panoramic meta-analysis models.

Model	SMD[a] (95% CI)	Between-indication τ^2	Between-trial τ^2
Procedural anxiety			
Overall	–0.75 (–0.95 to –0.55)	0.00	0.48
Age		0.00	0.45
Adults	–0.27 (–0.73 to 0.18)		
Children	–0.86 (–1.07 to –0.64)		
Difference	–0.58 (–1.08 to –0.08)		
Immersive		0.00	0.44
Nonimmersive	–0.97 (–1.22 to –0.71)		
Immersive	–0.46 (–0.75 to –0.16)		
Difference	0.51 (0.12 to 0.90)		
Mechanism		0.00	0.48
Relaxation/distraction	–0.82 (–1.05 to –0.58)		
Education/exposure	–0.53 (–0.95 to –0.11)		
Difference	0.28 (–0.19 to 0.76)		
General health anxiety			
Overall	–0.82 (–1.29 to –0.45)	0.09	0.33
Immersive		0.09	0.35
Nonimmersive	–0.86 (–1.30 to –0.41)		
Immersive	–0.78 (–1.25 to –0.30)		
Difference	0.08 (–0.44 to 0.60)		

[a]SMD: standardized mean difference.

Sensitivity Analyses

Because the number of indications was fewer than 10, a protocol-specified sensitivity analysis was undertaken, treating indication as a fixed effect. A likelihood ratio test comparing this model against a simple 2-level model suggested that including an indication as a fixed effect did not improve model fit (χ^2_4=2.47, *P*=.65).

Three protocol-specified meta-regressions were then undertaken. Because of the size of each indication category, these should be regarded as indicative only. Across indications, interventions were significantly more effective in children versus adults (difference: *d*=–0.58, 95% CI –1.08 to –0.08), and immersive interventions appeared less effective than nonimmersive interventions (difference: *d*=0.51, 95% CI 0.12-0.90). There was no significant difference in the size of effect between interventions using relaxation or distraction and interventions using education or exposure.

An Egger test for the dental anxiety indication implied some evidence of small-study bias (*P*=.048), with a funnel plot reflecting this possibility (Multimedia Appendix 5). While an Egger test for needle-related anxiety did not generate a significant result (*P*=.97), a funnel plot indicated significant unexplained heterogeneity. An Egger test for surgical operations was not significant (*P*=.86), and a funnel plot did not suggest evidence of small-study bias.

General Health Anxiety

Overview of Trials

Classification of trials measuring general health anxiety outcomes identified 5 indications: cancer (n=13), cardiovascular disease (n=7), chronic pain disorders (n=1), maternity (n=6), and anxiety related to wounds and burn care (n=3). Indication-specific meta-analyses are presented in Figure 3. All interventions generated significant effects within the indication, except for interventions for chronic pain disorders.

XSL•FO
RenderX

Figure 3. Forest plot for general health anxiety outcomes.

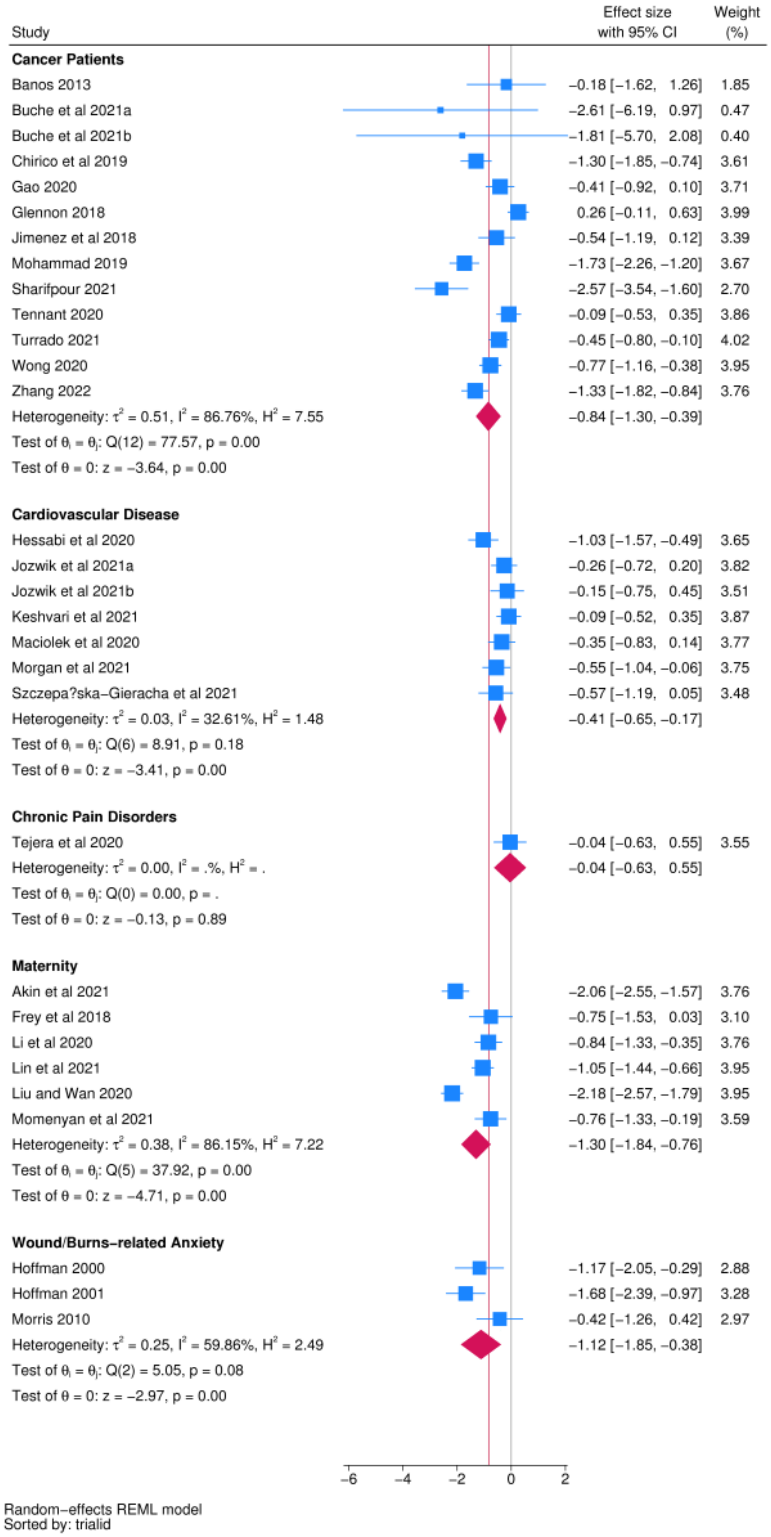

Panoramic Meta-Analysis

A panoramic meta-analysis model nesting trials within indication generated an overall pooled effect (*d*) of –0.82 (95% CI –1.20 to –0.45; Table 2). The between-indication (level 3) variance parameter for this model was equal to 0.09, whereas the within-indication, between-trial (level 2) variance parameter was equal to 0.33. Thus, the I^2 attributable to indication was 17.7%, and the I^2 attributable to between-trial differences was 67.3%. The conditional I^2 using only trial-level and between-trial variance was 81.8%. The prediction interval for the pooled effect spanned –2.09 to 0.45, indicating, as for procedural anxiety, the highly uncertain effectiveness of XR interventions in any future trial.

Sensitivity Analyses

Since the number of indications was fewer than 10, a protocol-specified sensitivity analysis was undertaken, which

XSL•FO
RenderX

treated indication as a fixed effect. A likelihood ratio test comparing this model against a simple 2-level model suggested that including indication as a fixed effect did not improve model fit (χ^2_4=8.57, *P*=.07), albeit with a marginal significance test.

Of the 3 protocol-specified meta-regression analyses, tests for age and mechanism were not undertaken, as only 2 indications had a variation on these meta-regressors. The difference between immersive and nonimmersive interventions was small and not significant (difference: *d*=0.08; 95% CI –0.44 to 0.60). Finally, an Egger test for trials in the cancer indication generated a nonsignificant result (*P*=.19), and a funnel plot did not suggest conclusive evidence of small-study bias (Multimedia Appendix 5).

Discussion

Principal Findings

Using an innovative statistical method, this analysis demonstrated that the effectiveness of XR interventions for both procedural and health anxiety is heterogeneous but conclusive; however, the evidence base for procedural anxiety is limited by publication bias, and effectiveness for procedural anxiety is moderated by age of participant and mechanism of effect. As a distinctive feature of panoramic meta-analysis, it was demonstrated that effects in procedural anxiety generalize across currently tested indications, while indication explains a relatively low amount of variance in general health anxiety outcomes; but that generalizability in a future trial is highly uncertain. Put otherwise, the evidence for the effectiveness of XR interventions in currently assessed indications cannot be used to assert that XR interventions would likely be effective for *any* given indication.

Implications for Research and Practice

Although positive effects emerged across a range of general and procedural health anxieties, various indications remain poorly established. Indeed, most meta-analyses are based on context-dependent intervention data, which relates to specific medical procedures (eg, needle insertions and dental surgery) or long-term conditions (eg, cancer and cardiovascular disease). Panoramic meta-analysis methodologies were specifically undertaken to synthesize evidence from diverse clinical contexts and to establish the exchangeability of effects between different intervention types, technologies, populations, and indications. However, key questions remain about the effectiveness of XR within broader domains. For instance, studies are notably lacking in relation to imaging and diagnostic screening procedures (eg, endoscopies and mammography), where anxiety can have well-documented impacts on the quality of care [36-38]. Moreover, most current trials concern pediatric patient groups, despite health-related anxieties showing significant (and potentially growing) prevalence among adults [39,40]. Given the uncertain generalizability of study data that was highlighted in the present analyses, research needs to ascertain the best-suited clinical contexts for future XR interventions.

For the management of procedural anxieties, meta-regressions showed XR to be more effective in children (vs adults), and when using nonimmersive (as opposed to immersive) technologies. These findings imply that age and technology-related variables could influence the success of future XR programs. However, these findings did not replicate general health anxieties, indicating a role for broader, psychopathological factors in shaping effectiveness. For instance, positive outcomes may be more easily achieved when targeting relatively minor forms of anxiety (eg, a child's worries about needles), compared with more complex and longstanding dispositions (eg, anxieties relating to major surgery or lifelong phobias). Relatedly, nonimmersive XR may be used for implementing simple, distraction-based interventions, whereas immersive devices may be required when attempting to administer more sophisticated or wide-ranging therapeutic solutions that are challenging to deliver successfully. When developing future applications, practitioners must consider the role of these interrelated components, since the value of key technological features (eg, fidelity and face validity) will depend on an intervention's wider situational context.

Future research is needed to advance our mechanistic understanding of the empirical data. Here, studies should not just explore population- and technology-specific variables (eg, the age of patients, system immersion, and types of hardware), but also broader methodological components relating to how the intervention is delivered and how it actually "works". Notably, the present results found no significant differences in the management of procedural anxiety between interventions using relaxation or distraction and interventions using education or exposure. However, there were a limited number of trials that examined exposure- and education-based XR methods in the evaluated review evidence, and there were insufficient data for analyzing general health anxiety outcomes. In addition, the included studies rarely provided sufficient details or rationale about their XR procedures and any cointerventions that were present, which limited capabilities for comparing different mechanisms of action. It is important that the contrasting psychophysiological processes implicated by different XR methods are considered during prospective assessments, as they are likely to impact on system requirements. For instance, a high degree of affective fidelity may be important for exposure therapies aiming to provoke "lifelike" stressors and psychophysiological responses; whereas aspects of user presence and task engagement may be more pertinent in simple distraction-based applications (see [41] for related framework). Nonetheless, without further empirical attention in this area, the relationships between different intervention types, psychophysiological mechanisms, and system features remain somewhat unclear. Overall, it is therefore vital that future work aims to address these "gaps" in theoretical understanding, so that the prospective design of XR programs can be sufficiently optimized for clinical implementation.

Strengths and Limitations

The use of panoramic meta-analysis meant that it was not possible to examine questions not generally addressed in single-indication meta-analyses. While the use of an overview of reviews was efficient in this analysis (and in keeping with the original implementation of this method), it is possible that very recently published trials not captured in existing systematic reviews were missed. In addition, this analysis was reliant on

XSL•FO
RenderX

the quality of underpinning systematic reviews including search, extraction, and appraisal. None of the available reviews were rated as high-quality on AMSTAR-2, and many reviews were rated as critically low. Most trials were not rated as being at low RoB in their "parent" reviews. Relatedly, both the trials and reviews contained inconsistencies and inaccuracies in the reporting of study interventions (eg, in terms of describing levels of system immersion, or conceptualizing procedural and general health anxieties). Though included reviews were able to be traced to "saturation", ensuring the identification of all relevant trials identified across reviews, it is possible that relevant trials were still missed.

Conclusions

While XR interventions are promising for both procedural and health anxiety, a number of unanswered questions remain. Future research should broaden the indications in which XR interventions are used, and should continue to develop emerging approaches, such as immersive technology interventions, to optimize their effectiveness while continuing to refine approaches for which effectiveness is well established.

Acknowledgments

This work was supported by an Innovate UK grant (10051972), which was awarded to Cineon Training in partnership with the University of Exeter. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any author-accepted manuscript version arising from this submission.

Data Availability

The datasets analyzed during this study are available on the Open Science Framework repository [10].

Authors' Contributions

TA contributed to conceptualization, data curation, formal analysis, investigation, visualization, project administration, and writing the original draft. GJM-T was involved in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, supervision, and writing the original draft. DH contributed to conceptualization, investigation, and reviewing and editing of the manuscript. SR was responsible for methodology, investigation, and reviewing and editing the manuscript. MW contributed to the conceptualization and reviewing and editing of the manuscript. SV was involved in conceptualization, investigation, funding acquisition, supervision, and reviewing and editing the manuscript.

Conflicts of Interest

SV is the chief scientific officer and a shareholder of Cineon Training (of whom the Innovate UK grant was awarded) but declares no nonfinancial competing interests. TA and DH have been employed and acted as paid consultants for Cineon Training but declare no nonfinancial competing interests. Neither Cineon Training or Innovate UK played any role in the design of the research, the analysis and interpretation of data, or the writing of the manuscript.

Multimedia Appendix 1
Search methods.
[DOCX File , 24 KB - jmir_v27i1e58086_app1.docx]

Multimedia Appendix 2
List of included reviews.
[DOCX File , 25 KB - jmir_v27i1e58086_app2.docx]

Multimedia Appendix 3
Extracted study data.
[DOCX File , 57 KB - jmir_v27i1e58086_app3.docx]

Multimedia Appendix 4
Quality appraisals.
[DOCX File , 48 KB - jmir_v27i1e58086_app4.docx]

Multimedia Appendix 5
Funnel plots.
[DOCX File , 74 KB - jmir_v27i1e58086_app5.docx]

Multimedia Appendix 6

XSL•FO
RenderX

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 74 KB - jmir_v27i1e58086_app6.pdf]

References

1. Parrish CL, Radomsky AS, Dugas MJ. Anxiety-control strategies: is there room for neutralization in successful exposure treatment? Clin Psychol Rev 2008;28(8):1400-1412. [doi: 10.1016/j.cpr.2008.07.007] [Medline: 18783862]
2. Penfold K, Page AC. The effect of distraction on within-session anxiety reduction during brief in vivo exposure for mild blood-injection fears. Behav Ther 1999;30(4):607-621. [doi: 10.1016/s0005-7894(99)80028-8]
3. Foa EB, Kozak MJ. Emotional processing of fear: exposure to corrective information. Psychol Bull 1986;99(1):20-35. [Medline: 2871574]
4. Emmelkamp PMG, Meyerbröker K. Virtual reality therapy in mental health. Annu Rev Clin Psychol 2021;17:495-519. [doi: 10.1146/annurev-clinpsy-081219-115923] [Medline: 33606946]
5. Bell IH, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice . Dialogues Clin Neurosci 2020;22(2):169-177 [FREE Full text] [doi: 10.31887/DCNS.2020.22.2/lvalmaggia] [Medline: 32699517]
6. Heyse J, Depreeuw B, van Daele T, Daeseleire T, Ongenae F, de Backere F, et al. An adaptation algorithm for personalised virtual reality exposure therapy. Comput Methods Programs Biomed 2022;225:107077. [doi: 10.1016/j.cmpb.2022.107077] [Medline: 36030573]
7. Boeldt D, McMahon E, McFaul M, Greenleaf W. Using virtual reality exposure therapy to enhance treatment of anxiety disorders: identifying areas of clinical adoption and potential obstacles. Front Psychiatry 2019;10:773 [FREE Full text] [doi: 10.3389/fpsyt.2019.00773] [Medline: 31708821]
8. Hemming K, Pinkney T, Futaba K, Pennant M, Morton DG, Lilford RJ. A systematic review of systematic reviews and panoramic meta-analysis: staples versus sutures for surgical procedures. PLoS One 2013;8(10):e75132 [FREE Full text] [doi: 10.1371/journal.pone.0075132] [Medline: 24116028]
9. Fordham B, Sugavanam T, Edwards K, Stallard P, Howard R, das Nair R, et al. The evidence for cognitive behavioural therapy in any condition, population or context: a meta-review of systematic reviews and panoramic meta-analysis. Psychol Med 2021;51(1):21-29 [FREE Full text] [doi: 10.1017/S0033291720005292] [Medline: 33455594]
10. Harris D, Melendez-Torres G, Vine S, Arthur T, Robinson S, Wilson M, et al. Improving service access using virtual exposure (iSAVE). OSF Home. 2023. URL: https://osf.io/nhzf8/ [accessed 2024-03-05]
11. Kain ZN, Mayes LC, Cicchetti DV, Bagnall AL, Finley JD, Hofstadter MB. The yale preoperative anxiety scale: how does it compare with a "gold standard"? Anesth Analg 1997;85(4):783-788. [doi: 10.1097/00000539-199710000-00012] [Medline: 9322455]
12. Asmundson GJG, Abramowitz JS, Richter AA, Whedon M. Health anxiety: current perspectives and future directions. Curr Psychiatry Rep 2010;12(4):306-312. [doi: 10.1007/s11920-010-0123-9] [Medline: 20549396]
13. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
14. Chan E, Hovenden M, Ramage E, Ling N, Pham JH, Rahim A, et al. Virtual reality for pediatric needle procedural pain: two randomized clinical trials. J Pediatr 2019;209:160-167.e4. [doi: 10.1016/j.jpeds.2019.02.034] [Medline: 31047650]
15. Özalp Gerçeker G, Ayar D, Özdemir EZ, Bektaş M. Effects of virtual reality on pain, fear and anxiety during blood draw in children aged 5-12 years old: a randomised controlled study. J Clin Nurs 2020;29(7-8):1151-1161. [doi: 10.1111/jocn.15173] [Medline: 31889358]
16. Kılıç A, Brown A, Aras I, Hui R, Hare J, Hughes LD, et al. Using virtual technology for fear of medical procedures: a systematic review of the effectiveness of virtual reality-based interventions. Ann Behav Med 2021 Oct 27;55(11):1062-1079 [FREE Full text] [doi: 10.1093/abm/kaab016] [Medline: 33821879]
17. Koo CH, Park JW, Ryu JH, Han SH. The effect of virtual reality on preoperative anxiety: a meta-analysis of randomized controlled trials. J Clin Med 2020 Sep 29;9(10):3151 [FREE Full text] [doi: 10.3390/jcm9103151] [Medline: 33003411]
18. Tas FQ, van Eijk CAM, Staals LM, Legerstee JS, Dierckx B. Virtual reality in pediatrics, effects on pain and anxiety: a systematic review and meta-analysis update. Paediatr Anaesth 2022 Dec;32(12):1292-1304 [FREE Full text] [doi: 10.1111/pan.14546] [Medline: 35993398]
19. Simonetti V, Tomietto M, Comparcini D, Vankova N, Marcelli S, Cicolini G. Effectiveness of virtual reality in the management of paediatric anxiety during the peri-operative period: a systematic review and meta-analysis. Int J Nurs Stud 2022 Jan;125:104115. [doi: 10.1016/j.ijnurstu.2021.104115] [Medline: 34781118]
20. Yan X, Yan Y, Cao M, Xie W, O'Connor S, Lee JJ, et al. Effectiveness of virtual reality distraction interventions to reduce dental anxiety in paediatric patients: a systematic review and meta-analysis. J Dent 2023 May;132:104455. [doi: 10.1016/j.jdent.2023.104455] [Medline: 36842625]
21. López-Valverde N, Muriel Fernández J, López-Valverde A, Valero Juan LF, Ramírez JM, Flores Fraile J, et al. Use of virtual reality for the management of anxiety and pain in dental treatments: systematic review and meta-analysis. J Clin Med 2020 Apr 05;9(4):1025 [FREE Full text] [doi: 10.3390/jcm9041025] [Medline: 32260538]

XSL•FO
RenderX

22. Gao Y, Xu Y, Liu N, Fan L. Effectiveness of virtual reality intervention on reducing the pain, anxiety and fear of needle-related procedures in paediatric patients: a systematic review and meta-analysis. J Adv Nurs 2023 Jan;79(1):15-30. [doi: 10.1111/jan.15473] [Medline: 36330583]
23. Wang Y, Guo L, Xiong X. Effects of virtual reality-based distraction of pain, fear, and anxiety during needle-related procedures in children and adolescents. Front Psychol 2022;13:842847 [FREE Full text] [doi: 10.3389/fpsyg.2022.842847] [Medline: 35519646]
24. Turan Kavradim S, Yangöz Ş, Özer Z. Effectiveness of virtual reality interventions on physiological and psychological outcomes of adults with cardiovascular disease: a systematic review and meta-analysis. J Nurs Scholarsh 2023 Sep;55(5):949-966. [doi: 10.1111/jnu.12885] [Medline: 36942847]
25. Lan X, Tan Z, Zhou T, Huang Z, Huang Z, Wang C, et al. Use of virtual reality in burn rehabilitation: a systematic review and meta-analysis. Arch Phys Med Rehabil 2023 Mar;104(3):502-513 [FREE Full text] [doi: 10.1016/j.apmr.2022.08.005] [Medline: 36030891]
26. Gava V, Fialho HRF, Calixtre LB, Barbosa GM, Kamonseki DH. Effects of gaming on pain-related fear, pain catastrophizing, anxiety, and depression in patients with chronic musculoskeletal pain: a systematic review and meta-analysis. Games Health J 2022 May 24. [doi: 10.1089/g4h.2021.0232] [Medline: 35613404]
27. Xu N, Chen S, Liu Y, Jing Y, Gu P. The effects of virtual reality in maternal delivery: systematic review and meta-analysis. JMIR Serious Games 2022 Nov 23;10(4):e36695 [FREE Full text] [doi: 10.2196/36695] [Medline: 36416881]
28. Baradwan S, Khadawardi K, Badghish E, Alkhamis WH, Dahi AA, Abdallah KM, et al. The impact of virtual reality on pain management during normal labor: a systematic review and meta-analysis of randomized controlled trials. Sex Reprod Healthc 2022 Jun;32:100720. [doi: 10.1016/j.srhc.2022.100720] [Medline: 35381438]
29. Wu Y, Wang N, Zhang H, Sun X, Wang Y, Zhang Y. Effectiveness of virtual reality in symptom management of cancer patients: a systematic review and meta-analysis. J Pain Symptom Manage 2023 May;65(5):e467-e482 [FREE Full text] [doi: 10.1016/j.jpainsymman.2023.01.023] [Medline: 36758907]
30. Obrero-Gaitán E, Cortés-Pérez I, Calet-Fernández T, García-López H, López Ruiz MDC, Osuna-Pérez MC. Digital and interactive health interventions minimize the physical and psychological impact of breast cancer, increasing women's quality of life: a systematic review and meta-analysis. Cancers (Basel) 2022 Aug 26;14(17):4133 [FREE Full text] [doi: 10.3390/cancers14174133] [Medline: 36077670]
31. Czech O, Rutkowski S, Kowaluk A, Kiper P, Malicka I. Virtual reality in chemotherapy support for the treatment of physical functions, fear, and quality of life in pediatric cancer patients: a systematic review and meta-analysis. Front Public Health 2023;11:1039720 [FREE Full text] [doi: 10.3389/fpubh.2023.1039720] [Medline: 37124795]
32. Bu X, Ng PH, Xu W, Cheng Q, Chen PQ, Cheng AS, et al. The effectiveness of virtual reality-based interventions in rehabilitation management of breast cancer survivors: systematic review and meta-analysis. JMIR Serious Games 2022 Feb 28;10(1):e31395 [FREE Full text] [doi: 10.2196/31395] [Medline: 35225817]
33. Zeng Y, Zhang JE, Cheng ASK, Cheng H, Wefel JS. Meta-analysis of the efficacy of virtual reality-based interventions in cancer-related symptom management. Integr Cancer Ther 2019;18:1534735419871108 [FREE Full text] [doi: 10.1177/1534735419871108] [Medline: 31441352]
34. Lahti S, Suominen A, Freeman R, Lähteenoja T, Humphris G. Virtual reality relaxation to decrease dental anxiety: immediate effect randomized clinical trial. JDR Clin Trans Res 2020;5(4):312-318 [FREE Full text] [doi: 10.1177/2380084420901679] [Medline: 31962052]
35. Jiang MY, Upton E, Newby JM. A randomised wait-list controlled pilot trial of one-session virtual reality exposure therapy for blood-injection-injury phobias. J Affect Disord 2020;276:636-645. [doi: 10.1016/j.jad.2020.07.076] [Medline: 32871696]
36. Madl J, Janka R, Bay S, Rohleder N. MRI as a stressor: the psychological and physiological response of patients to MRI, influencing factors, and consequences. J Am Coll Radiol 2022;19(3):423-432. [doi: 10.1016/j.jacr.2021.11.020] [Medline: 35131225]
37. Jones MP, Ebert CC, Sloan T, Spanier J, Bansal A, Howden CW, et al. Patient anxiety and elective gastrointestinal endoscopy. J Clin Gastroenterol 2004;38(1):35-40. [doi: 10.1097/00004836-200401000-00009] [Medline: 14679325]
38. Loving VA, Aminololama-Shakeri S, Leung JWT. Anxiety and its association with screening mammography. J Breast Imaging 2021;3(3):266-272. [doi: 10.1093/jbi/wbab024] [Medline: 38424779]
39. Abate SM, Chekol YA, Basu BB. Global prevalence and determinants of preoperative anxiety among surgical patients: a systematic review and meta-analysis. Int J Surg Open 2020;25:6-16. [doi: 10.1016/j.ijso.2020.05.010]
40. Tyrer P, Cooper S, Tyrer H, Wang D, Bassett P. Increase in the prevalence of health anxiety in medical clinics: possible cyberchondria. Int J Soc Psychiatry 2019;65(7-8):566-569 [FREE Full text] [doi: 10.1177/0020764019866231] [Medline: 31379243]
41. Harris DJ, Bird JM, Smart PA, Wilson MR, Vine SJ. A framework for the testing and validation of simulated environments in experimentation and training. Front Psychol 2020;11:605 [FREE Full text] [doi: 10.3389/fpsyg.2020.00605] [Medline: 32296379]

XSL•FO
RenderX

Abbreviations

AMSTAR-2: A Measurement Tool to Assess Systematic Reviews (Revised Instrument)
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RoB: risk of bias
XR: extended reality

Edited by A Coristine; submitted 05.03.24; peer-reviewed by B Fordham, MA Larsen; comments to author 10.08.24; revised version received 27.09.24; accepted 04.11.24; published 08.01.25.

Please cite as:
Arthur T, Melendez-Torres GJ, Harris D, Robinson S, Wilson M, Vine S
Extended Reality Interventions for Health and Procedural Anxiety: Panoramic Meta-Analysis Based on Overviews of Reviews
J Med Internet Res 2025;27:e58086
URL: https://www.jmir.org/2025/1/e58086
doi:10.2196/58086
PMID:

©Tom Arthur, GJ Melendez-Torres, David Harris, Sophie Robinson, Mark Wilson, Sam Vine. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating Whether Nonimmersion Virtual Reality Simulation Training Improves Nursing Competency in Isolation Wards: Randomized Controlled Trial

Dandan Zhang[1,2], MS; MuLi Fu[3], MS; Jianzhong Zhang[1], MS; Yuxuan Li[4], MS; Li Chen[2], MS; Yong-Jun Chen[2], PhD; Zhefeng Zhong[2], MS; Yin-Ping Zhang[1], PhD

[1]School of Nursing, Xi'an Jiaotong University Health Science Center, Xi'an, China

[2]Affiliated Nanhua Hospital, Hengyang Medical School, University of South China, Hengyang, China

[3]The First Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, China

[4]School of Economy & Finance, Xi'an Jiaotong University, Xi'an, China

Corresponding Author:
Yin-Ping Zhang, PhD
School of Nursing, Xi'an Jiaotong University Health Science Center
No.76, Yanta West Road
Xi'an, 7100061
China
Phone: 86 (29) 8265 7015
Email: cathyzh@mail.xjtu.edu.cn

Abstract

Background: During infectious disease outbreaks such as the COVID-19 pandemic, nurses are crucial in patient care and public health safety; however, they face challenges such as inadequate training and high stress in isolation wards. Virtual reality (VR) technology offers innovative training solutions to enhance nurses' clinical skills and preparedness. However, extensive studies on its effectiveness in isolation ward environments are still limited.

Objective: This study aims to develop a nonimmersive VR (NIVR) simulation training program for isolation wards and further validate its feasibility and training effectiveness in aiding nurses in adapting to isolation ward settings.

Methods: This study was a prospective, parallel, open-label, randomized controlled trial. A total of 90 nurses from 3 hospitals in China were randomly assigned to either the control or intervention group, with 45 (50%) individuals in each group. Both groups received training on isolation ward layout and nursing procedures. The control group underwent a 4-hour conventional training session consisting of 2 hours of face-to-face lectures and 2 hours of ward visits. The intervention group received a 4-hour NIVR simulation training session. Subsequently, both groups completed approximately 4 hours of emergency drills and assessments.

Results: After the intervention, there were no significant differences in theoretical test or performance assessment scores between the 2 groups (t_{88}=–0.30, P=.75; Cohen d=–0.06; z score=0.00, P>.99), using a 2-tailed t test. However, the intervention group completed 6 tasks faster than the control group (t_{88}=5.10, P<.001; Cohen d=1.08), with an average reduction of about 3 minutes (control group: mean 43.91, SD 2.99 min; intervention group: mean 40.77, SD 2.85 min). Notably, they completed task 3 (patient reception inward) and task 6 (exiting the isolation area) significantly quicker (t_{88}=3.22, P=.002; Cohen d=0.68; t_{88}=3.03, P=.003; Cohen d=0.64, respectively), with no significant differences for the other tasks.

Conclusions: This study highlights the potential of NIVR simulation training for nurses working in isolation wards. Although NIVR simulation training does not significantly surpass traditional methods in imparting theoretical knowledge, it does reduce task completion time for specific activities. Its capacity for safe, repetitive practice and realistic scenario simulation makes NIVR a valuable tool in medical education. Further research and optimization of VR simulation training programs are recommended to enhance nurses' practical skills and pandemic preparedness.

Trial Registration: Chinese Clinical Trial Registry ChiCTR240083155; https://www.chictr.org.cn/hvshowproject.html?id=250356&v=1.0

(J Med Internet Res 2025;27:e63131) doi:10.2196/63131

XSL•FO
RenderX

KEYWORDS

virtual reality simulation; isolation ward; preparedness; pandemic; nurse

Introduction

Background

The world faced the challenge of outbreaks of new infectious diseases. During the COVID-19 pandemic, nurses played a vital role in ensuring that patients received personalized, high-quality care while also being essential in patient management, infection prevention, and public health safety [1]. However, there were shortcomings in nurses' abilities to respond to the COVID-19 pandemic [2-6]. Therefore, it was necessary for nurses to receive adequate guidance and training to ensure a safe work environment, protect themselves, and effectively manage the pandemic [1]. To address this issue, the *National Nursing Career Development Plan (2021-2025)* released by the National Health Commission of China explicitly proposed targeted training in scarce nursing specialties, such as infectious diseases [7]. With national policy support, training programs for infectious disease specialist nurses are gradually developing. However, existing programs still face challenges, such as insufficient training, incomplete content, unitary training patterns, limited scenario simulations or emergency drills, and insufficient knowledge of ethics and law, particularly in professional training programs for nurses in the COVID-19 pandemic response [2-6,8-12].

When an pandemic breaks out, isolation wards protect health care workers and patients from infection. Studies have shown that the mental stress and occupational safety of frontline health care workers were linked to unfamiliarity with isolation wards [13,14]. Therefore, it is crucial to train nurses to familiarize them with isolation wards before they enter the quarantine area. Building on the COVID-19 pandemic response experience, health care institutions have intensified emergency training programs for professionals [15]. Nurses need to understand the spatial layout and division of the functional regions in isolation wards and critical aspects, such as standard procedures, personnel management, and infection prevention and control. The training and support aimed to help nurses acclimate to the work environment before entering the isolation area, thus reducing stress and anxiety stemming from unfamiliar surroundings, enhancing preparedness and teamwork abilities, and contributing to a more efficient response. However, studies on related topics remain relatively scarce.

Virtual reality (VR) technology uses digital techniques to create 3D VR environments, allowing users to interact with these environments through head-mounted displays, body sensors, and direct input devices [16]. On the basis of the level of immersion, VR is categorized into immersive VR, semi-immersive VR, and nonimmersive VR (NIVR) [17]. NIVR presents a virtual environment via a computer screen or mobile device, allowing users to interact using a keyboard, mouse, or touch devices [18]. Although NIVR offers a relatively diminished sense of immersion, its low cost, ease of deployment, and lack of a requirement for special equipment confer significant application potential across various scenarios, particularly in training [19-21]. Users can easily access these scenarios on PCs or mobile devices, overcoming temporal and spatial limitations.

Studies indicate that in nursing education and clinical practice, the effectiveness of NIVR can be comparable to, or even surpass, traditional simulation training [22-24]. This technology can effectively simulate complex medical procedures and nursing scenarios, allowing trainees to practice repeatedly in a safe, controlled environment, thus enhancing clinical skills, decision-making ability, and critical thinking [22,25-33]. In addition, a study indicates that the design characteristics of VR simulations encompass patient management content, the implementation of various scenarios and NIVR experiences lasting >30 minutes, and the provision of feedback following the conclusion of the scenarios. These strategies can significantly enhance the effectiveness of VR simulations [34]. During the COVID-19 pandemic, VR proved helpful in addressing the training needs of medical students; however, the lack of hands-on practice might impact the development of clinical skills [35,36]. Therefore, it is necessary to explore integrated training that combines VR with hands-on practice. Yang et al [37] demonstrated that combining NIVR and face-to-face simulation training effectively enhances nursing clinical judgment and encourages student progress in comprehensive simulations. Thus, when integrating VR with traditional training methods, it is crucial to understand the strengths and limitations of each technology and adopt the best practices [29,30,38]. Studies should explore how to integrate these technologies more effectively into the training process, driving innovation and progress in medical education.

The COVID-19 pandemic has underscored the necessity of advancing VR-based remote education [33]. NIVR has advantages in remote virtual training and is increasingly being adopted in nursing degree programs and preservice education [13,33,39]. During outbreaks, practical clinical applications of NIVR simulation training cover multiple aspects, such as infection prevention and control, the use of personal protective equipment [40-42]. It also includes training for managing patients in critical condition, such as basic life support, advanced cardiac life support, and extracorporeal membrane oxygenation [43-47]. NIVR training related to simulating clinical scenarios and integrating clinical cases continues to evolve, with management programs for patients with COVID-19 and Ebola virus diseases among the initiatives [42,48-50]. Currently, research on using VR simulations for nurses to familiarize themselves with and adapt to isolation ward environments remains very limited.

Objectives

The primary objective of this study was to develop an NIVR simulation training program specifically designed for isolation wards and to validate its feasibility and effectiveness in assisting nurses in adapting to this unique environment.

XSL•FO
RenderX

Methods

Participants

This study recruited qualified nurses from 3 comprehensive hospitals in Hengyang City, Hunan Province, from April 1 to April 20, 2024. This study included registered nurses aged 18 to 40 years, with ≥1 year of clinical experience, no previous work experience in isolation wards, good physical health, and a willingness to fully participate in the training. Nonvoluntary participants and those who withdrew midway were excluded. Before commencing the training, each participating nurse was provided with an informed consent form outlining the project training process and duration. The study was initiated after the nurses signed the consent form voluntarily.

A total of 90 eligible nurses participated in this study and were randomly divided into control and intervention groups, with 45 (50%) nurses in each group.

Study Design

This prospective, parallel, randomized controlled trial was registered in April 2024 with the Chinese Clinical Trial Registry (registration number: ChiCTR240083155).

Ethical Considerations

This study was approved by the ethics committee of Affiliated Nanhua Hospital, University of South China (2024-ky-044). This study complies with the Declaration of Helsinki. All participants provided written informed consent and were made aware of their right to opt out of the study at any time. All collected data are anonymous and confidential, and only the authors can access them. Upon completing the training and assessment for this study, participants will be awarded a gift worth approximately US $30 as compensation.

Randomization and Blinding

Participant randomization into intervention and control groups was conducted using SPSS software (version 25.0; IBM Corp) to generate random numbers for a 1:1 allocation. To ensure the integrity of the allocation sequence, it was concealed using sequentially numbered, opaque, and sealed envelopes, which were only opened at the time of intervention assignment after all participants had completed their baseline assessments.

Due to the open-label nature of the study, it was not possible to blind participants or research assistants involved in the intervention. However, we implemented several measures to minimize subjective biases and ensure objective data collection. First, we scheduled intervention sessions for the 2 groups at different times to prevent cross-group communication. Second, distinct teams managed different stages of the study, ensuring that data collectors were not aware of participants' group allocations. Moreover, outcome adjudicators remained blinded to group assignments throughout the process.

Interventions

The control group completed the training on May 25, 2024, while the intervention group finished the training on May 26, 2024. Both groups underwent a 4-hour training session on isolation ward layout and nursing workflow in the morning. The intervention group received NIVR simulation training, while the control group received training in the form of conventional lectures and on-site ward visits. After the training, emergency drills and assessments were completed in the afternoon. The specific details of the training program are outlined subsequently.

The control group received conventional training. Participants in the control group initially received 2 hours of theoretical training. An experienced senior nurse who has worked in isolation wards for >5 years and has responded to pandemics multiple times was the lead instructor and used PowerPoint (Microsoft Corporation) presentations, images, videos, and other tools for face-to-face lectures. The topics covered in the training included an overall layout diagram of the isolation ward; the unique environment of the ward; specific distributions of clean areas, semicontaminated areas, contaminated areas, buffer zones, health care provider passages, and patient pathways; the correct personal protective measures; and disinfection. It also covered the work responsibilities and routes to the wards for the nursing shift, including procedures for entering and exiting the ward, patient admission and discharge management, methods for transferring items through pass windows, and transporting collected specimens.

Subsequently, the participants spent 2 hours familiarizing themselves with the isolation ward environment. The infectious disease building at the training site comprises 3 floors of negative pressure isolation wards. Each floor features an identical ward layout and is supervised by 2 senior isolation ward nurses as mentors. The group of 45 participants was further divided into 3 groups, each of 15 participants, with 1 group assigned to tour the first-floor ward. In the first hour, participants toured the layout of the ward under the guidance of a mentor. The following hour was dedicated to individuals or teams independently familiarizing themselves with and exploring the ward's layout. Mentors were available on-site to address any inquiries from the participants.

Next, 4-hour simulation drills and assessments were conducted using a team-based approach in the isolation ward to replicate the management of cases involving class A respiratory infectious diseases. The procedure for the simulation drills is provided in Figure 1. The participants were divided into 3 groups, each consisting of 15 individuals. The group leader assumed the role of the head nurse, while the members performed duties as ward nurses. The team leader scheduled shifts for 15 staff in the unit, assigning each person different shift responsibilities to simulate the management of patients in the isolation ward. The first hour was dedicated to practicing to familiarize oneself with various tasks, followed by a 3-hour Objective Structured Clinical Examination (OSCE) assessment.

XSL•FO
RenderX

Figure 1. Flow diagram of the drills.

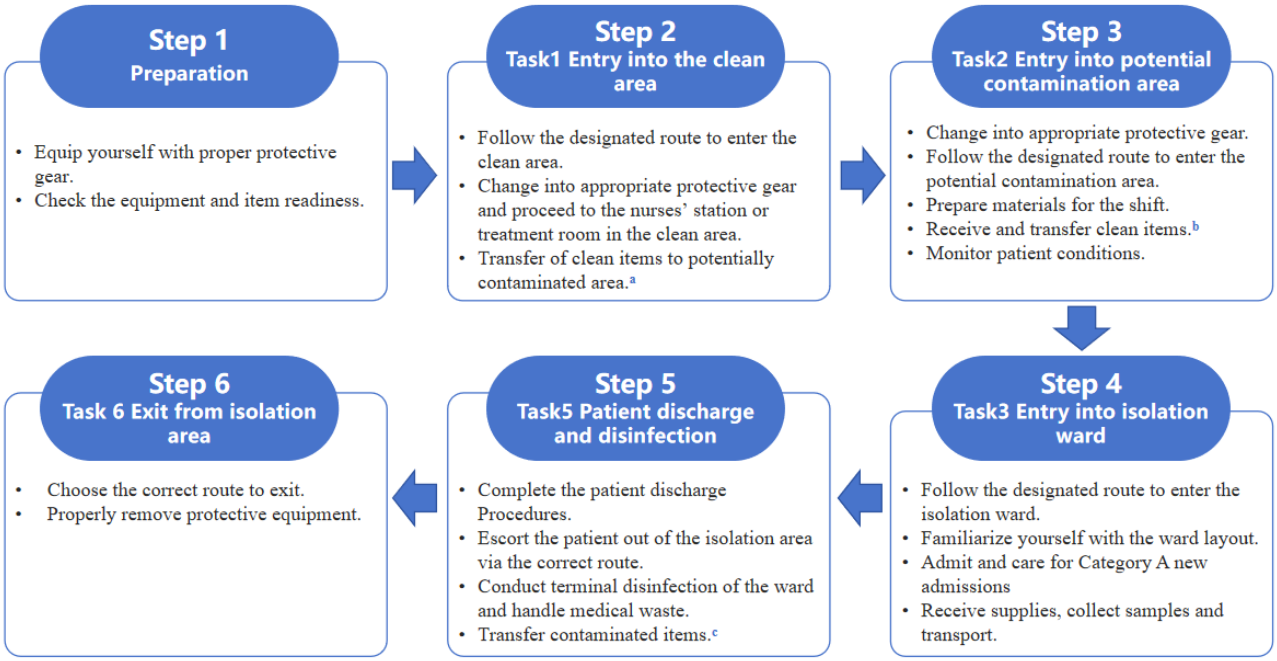

a,b,c Task 4 Use the transfer window to deliver items.

The intervention group received VR simulation training. First, from February to April 2024, our study team of medical experts collaborated closely with a virtual technology team to design and develop a VR project titled "Isolation Ward Layout and Nursing Procedures." The medical experts provided expertise on the layout and procedures of isolation wards, while the technology team used advanced VR technology to transform this guidance into a highly interactive and realistic simulation environment.

In terms of hardware, the project used a NIVR solution that presented a virtual environment via standard PCs or mobile devices, eliminating the necessity for specialized headsets. Participants could engage with the virtual environment on any compatible device using a keyboard, mouse, or touch screen. This approach enhanced accessibility and convenience, rendering NIVR training suitable for large-scale training applications, particularly in contexts where logistical simplicity and widespread access were essential.

For software development, Unreal Engine 5 (Epic Games) was used as the platform. With its powerful graphics rendering capabilities and realistic physics engine, Unreal Engine 5 enabled high-fidelity simulation of the isolation ward environment and layout. Its flexibility and scalability supported complex interactive mechanisms, providing users with an immersive learning and training experience.

This project used a combination of 2D map navigation and interactive 3D dynamic scenes of the isolation ward (Figure 2) to demonstrate specific aspects of high-security (negative pressure) isolation wards. The scenarios presented detailed descriptions of the unique working environments and layouts within these isolation areas, setting out specific tasks for nursing personnel in different shifts. The scenes covered procedures for entering isolation areas, receiving patients with class A infectious disease upon admission, monitoring patient conditions, providing treatment and care, discharging patients, and exiting the isolation ward. Key details included protection and disinfection practices in various areas, the preparation and transfer of supplies, specimen collection, and transportation. This approach aimed to familiarize trainees with distinct area layouts, working environments, and standard procedures. Following practice sessions, assessments were administered to reinforce the learned knowledge and evaluate training effectiveness. The ability to repeat the experiments addressed the challenge of limited opportunities for trainees to engage in on-site field training within infectious disease outbreak isolation areas.

XSL•FO
RenderX

Figure 2. Example of 2D map navigation and 3D layout of isolation ward. The nurse's position is indicated in real-time by a yellow arrow on a 2D navigation map, while a 3D real-life simulation of the current location is provided to help participants become familiar with the ward.



Participants underwent a 4-hour VR simulation training in a computer-simulated laboratory. An instructor provided a 30-minute demonstration and explanation of the VR simulation operations before the participants practiced various simulated tasks. Participants completed virtual simulation tasks, including approximately 3 hours of practice and postpractice evaluation. During the practice, the instructor solved the participants' problems on-site. Subsequently, the instructor led an approximately 30-minute discussion and reporting session to address practice-related queries and share experiences.

For the intervention group, the content of the simulation drills and assessments lasting 4 hours was the same as that of the control group. Figure 3 illustrates the training protocols for the 2 groups.

Figure 3. Training protocols for the 2 groups. VRS: virtual reality simulation.

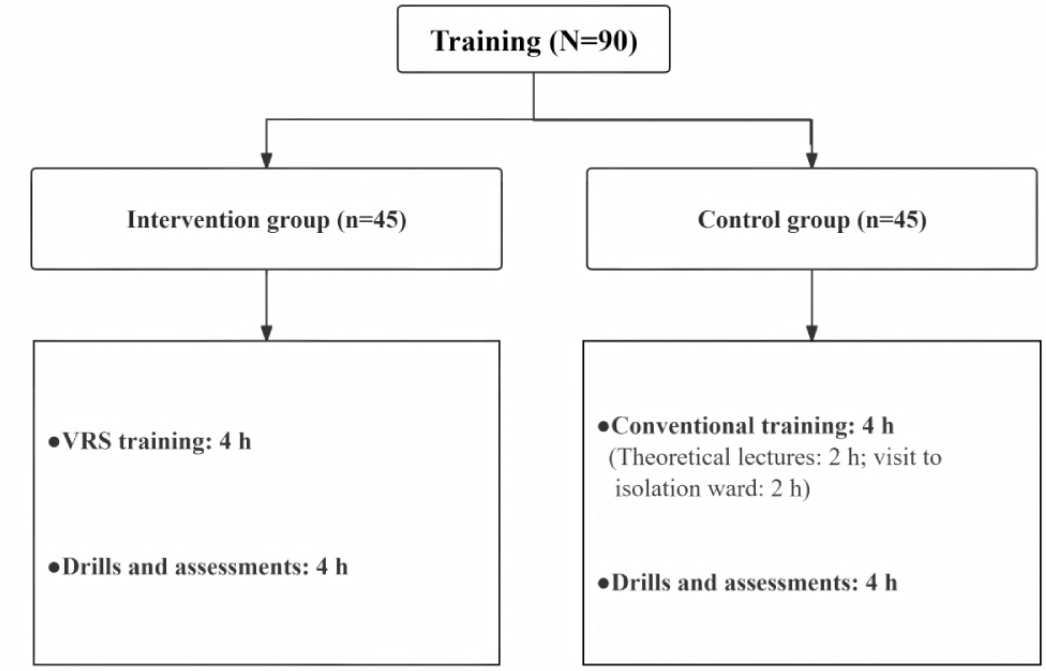

XSL•FO
RenderX

Measurement Indicators

Primary Outcome Measures

Theoretical Test Scores Before and After the Training

Both groups underwent web-based theoretical quizzes simultaneously before and after completing the training. The quizzes consisted of multiple-choice and fill-in-the-blank questions, with a total score of 100 points.

Evaluation of Drill

The assessments were conducted in the afternoon following the completion of training by the 2 groups. The 3 isolation wards were each arranged as assessment venues. Participants were randomly assigned by drawing lots, forming teams of 15 nurses each, who then entered the assessment area of a specific isolation ward. OSCE was used for the evaluation of the test, comprising 6 stations: nurse's entry into the ward to reach the clean nurses' station (10 points), entry into the potentially contaminated area and preparation of therapeutic and caring materials (20 points), entering the patient's room and completing the reception of incoming patients (20 points), using the transfer window to pass items (clean item transfer window, ward transfer window, and contaminated item transfer window; 15 points), handling patient discharge and completing terminal disinfection after discharge (20 points), and nurse's departure from the contaminated area to leave the ward (15 points), with the total score adding up to 100 points. The key scoring aspects included assessing whether the route selection aligned with the optimal path, ensuring the correct implementation of personal protective measures, appropriate use of materials, and adherence to standardized operational procedures. Each participant adhered to the team leader's schedule, arriving at the assigned task stations, sequentially completing tasks at each station, and recording their performance at every station.

Secondary Outcome: Duration of Drill Task Completion

The time each group of participants spent at each station during the OSCE test was simultaneously recorded (excluding the time for wearing and removing protective equipment and performing treatment procedures).

Statistical Analysis

Two researchers entered and verified the data and analyzed it using SPSS software (version 25.0). Continuous data were summarized with mean and SD, while categorical data were summarized using percentages. Student *t* tests (**2-tailed**) were applied to the intergroup comparisons of normally distributed continuous data. In cases of nonnormal distribution, the data were described as median (IQR), and the Mann-Whitney *U* test was used. Group comparisons for count data were conducted using χ^2 tests or the Fisher exact probability method. Statistical significance was denoted by 2-sided *P* values <.05.

Results

Demographic Data

The demographic characteristics of the 2 groups of participants are presented in Table 1. A total of 90 participants were included, with 45 (50%) individuals allocated to the intervention and control groups each (Figure 4). This study primarily included female nurses, with 96% (43/45) of them in the control group and 93% (42/45) in the intervention group. The average age was 27.53 (SD 4.0) years in the intervention group and 28.78 (SD 4.8) years in the control group. The demographic characteristics, including sex, age, highest degree, professional title, and work experience, were comparable between the 2 groups. All participants completed both baseline and postintervention assessments.

XSL•FO
RenderX

Table 1. Demographic data of the 2 groups (N=90).

Variable	Control group (n=45)	Intervention group (n=45)	*t* test (*df*)	χ^2 (*df*)	*P* value
Sex , n (%)			__[b]	0.2 (1)	.65
Male	2 (4)	3 (7)			
Female	43 (96)	42 (93)			
Age (y), mean (SD)[a]	28.78 (4.8)	27.53 (4.0)	1.244 (88)		.19
Clinical experience (y), n (%)			—	4.4 (2)	.11
≤5	20 (44)	29 (64)			
6-10	19 (42)	14 (31)			
≥11	6 (13)	2 (4)			
Highest degree, n (%)			—	6.5 (2)	.09
Diploma or associate degree	6 (13)	13 (29)			
Bachelor’s degree	38 (84)	30 (67)			
Master’s degree	1 (2)	2 (4)			
Professional title, n (%)			—	0.8 (1)	.37
Primary	28 (62)	32 (71)			
Intermediate	17 (38)	13 (29)			
Work experience in infectious disease departments, n (%)			—	1.6 (1)	.21
Yes	13 (29)	8 (18)			
No	32 (71)	37 (82)			
Experience in caring for patients with COVID-19 and SARS-CoV-2 infection, n (%)			—	0.2 (1)	.63
Yes	34 (76)	32 (71)			
No	11 (24)	13 (29)			
Hospital department of work, n (%)			—	5.5 (3)	.24
Internal medicine ward	17 (38)	17 (38)			
Surgical ward	21 (47)	20 (44)			
Emergency and critical care unit	1 (2)	4 (9)			
Pediatric department	5 (11)	1 (2)			
Others	1 (2)	3 (7)			

[a]Age is a continuous variable, expressed as mean (SD).

[b]Not applicable.

XSL•FO
RenderX

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flowchart of the study.

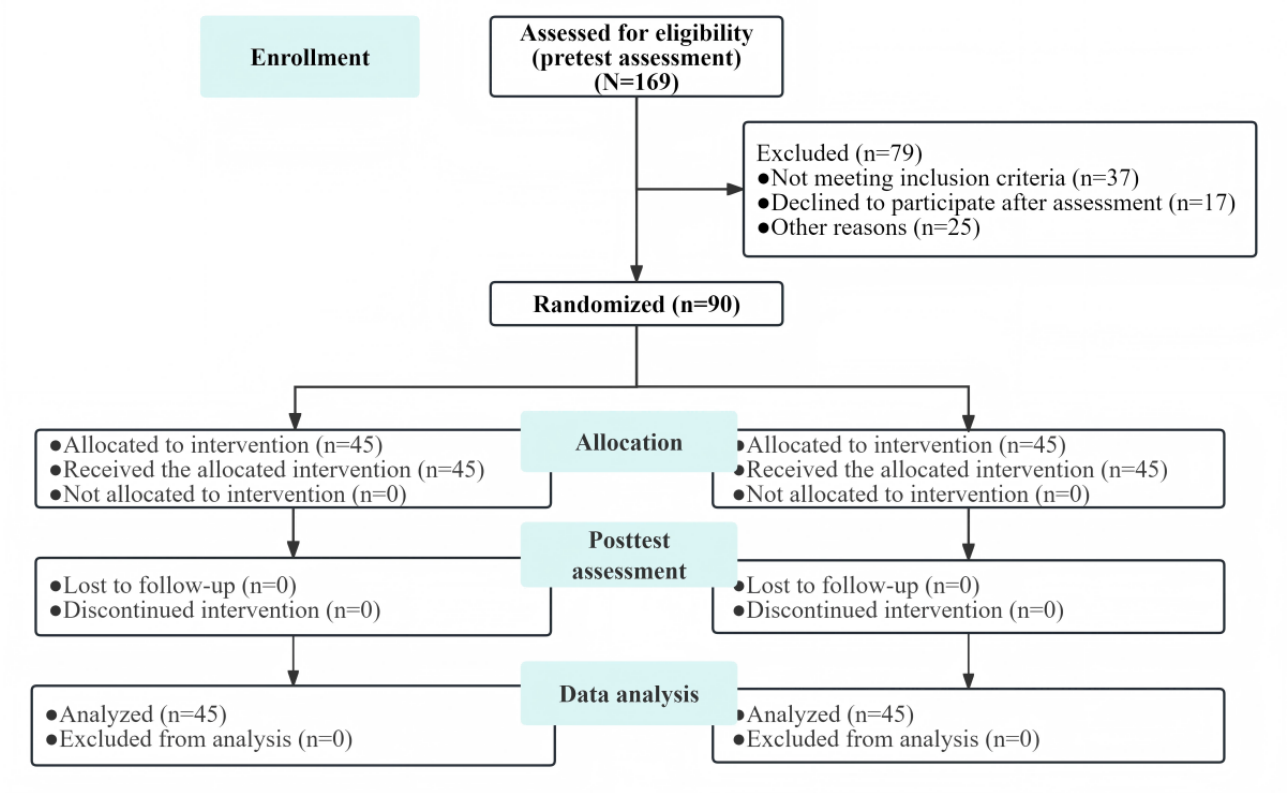

Primary Outcome 1: Comparison of the Theoretical Test Scores Between Groups Before and After Intervention

Before the intervention, no significant difference was found in the score of theoretical tests between the 2 groups (t_{88}=1.15, *P*=.25, Cohen *d*=0.24; Table 2), using a 2-tailed *t* test. However, after the intervention, the total scores of the theoretical tests in both groups significantly improved compared to the scores before the intervention. However, the 2 groups had no statistically significant difference in scores (t_{88}=–0.30, *P*=.75, Cohen *d*=–0.06; Table 2).

Table 2. The theoretical test scores between groups before and after intervention (N=90).

	Control group (n=45), mean (SD)	Intervention group (n=45), mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Preintervention score	49.49 (11.46)	46.47 (13.35)	1.15 (88)	.25	0.24[a]
Postintervention score	85.80 (6.49)	86.24 (7.49)	–0.30 (88)	.75	–0.06[b]

[a]According to Cohen conventions, a small effect size of 0.24 indicates a minor difference between the control and intervention groups before the intervention.

[b]Postintervention score Cohen *d* of –0.06: this effect size is negligible, indicating very little or no practical difference between the groups after the intervention.

Primary Outcome 2 and Secondary Outcomes: Comparison of the Drill Scores and Completion Times Between Groups After Intervention

After the intervention, no statistically significant differences were observed in the drill scores between the 2 groups (all *z* scores=0.00, *P*>.99; Table 3). However, the intervention group demonstrated a notably shorter total duration in completing the 6 target tasks than the control group (t_{88}=5.10,*P*<.001, Cohen *d*=1.08; Table 3). Among the 6 work tasks, the intervention group completed task 3 (patient reception inwards) and task 6 (leaving the isolation area) significantly faster than the control group (t_{88}=3.22, *P*=.002, Cohen *d*=0.68; t_{88}=3.03, *P*=.003, Cohen *d*=0.64; Table 3). In comparison, there were no differences in completion times for the remaining 4 tasks between both groups (*P*>.05; Table 3). The *t* test was 2-tailed.

XSL•FO
RenderX

Table 3. The drill scores and completion times between groups after intervention (N=90).

Variable	Control group (n=45)	Intervention group (n=45)	*t* test (df)	*z* score	*P* value	Cohen *d*
Task 1: the nurse enters the isolation ward and reaches the clean area in the nurses'station						
Scores, median (IQR)[a]	10 (10-10)	10 (10-10)	—[b]	0.00	>.99	—
Time taken (min), mean (SD)	3.88 (0.92)	3.62 (0.94)	1.36 (88)	—	.18	0.28[c]
Task 2: the nurse enters the potentially contaminated area and prepares the necessary supplies						
Scores, median (IQR)[a]	20 (18-20)	20 (18-20)	—	0.00	>.99	—
Time taken (min), mean (SD)	3.79 (0.79)	3.59 (0.80)	1.19 (88)	—	.24	0.25[c]
Task 3: the nurse enters the patient's room and completes the reception of the newly admitted patient						
Scores, median (IQR)[a]	20 (17.5-20)	20 (18-20)	—	0.00	>.99	—
Time taken (min), mean (SD)	11.71 (1.71)	10.76 (1.00)	3.22 (88)	—	.002	0.68[d]
Task 4: the nurse uses the transfer window to deliver items						
Scores, median (IQR)[a]	15 (13.5-15)	15 (13.5-15)	—	0.00	>.99	—
Time taken (min), mean (SD)	2.59 (0.82)	2.48 (0.60)	0.76 (88)	—	.50	0.15[e]
Task 5: the nurse completes the patient discharge procedures and performs terminal disinfection of the ward after the discharge						
Scores, median (IQR)[a]	20 (17-20)	20 (18-20)	—	0.00	>.99	—
Time taken (min), mean (SD)	9.41 (1.13)	9.24 (1.58)	0.60 (88)	—	.55	0.12[e]
Task 6: the nurse exits the isolation ward from the contaminated area						
Scores, median (IQR)[a]	15 (15-15)	15 (10-15)	—	0.00	>.99	—
Time taken (min), mean (SD)	7.00 (2.32)	5.59 (2.09)	3.03 (88)	—	.003	0.64[d]
Total						
Scores, median (IQR)[a]	93 (85-98)	92 (88-97)	—	0.00	>.99	—
Time taken (min), mean (SD)	43.91 (2.99)	40.77 (2.85)	5.10 (88)	—	<.001	1.08[f]

[a]Nonnormally distributed data are presented as median (IQR), using *z* score (Mann-Whitney *U* test).

[b]Not applicable.

[c]Task 1 and task 2: the effect sizes are small (Cohen *d*=0.28 and Cohen *d*=0.25, respectively), indicating minimal to small practical differences in time taken between the 2 groups for these tasks.

[d]Task 3 and task 6: the effect sizes are moderate (Cohen *d*=0.68 and Cohen *d*=0.64, respectively), suggesting that these tasks saw a noticeable improvement in efficiency for the intervention group compared to the control group.

[e]Task 4 and task 5: the effect sizes are very small (Cohen *d*=0.15 and Cohen *d*=0.12, respectively), suggesting very little difference in performance between the groups.

[f]The effect size for the total time taken is large (Cohen *d*=1.08), indicating a significant reduction in overall time taken for completing all tasks by the intervention group, reflecting a substantial practical significance.

Discussion

Principal Findings

In this study, we developed an NIVR simulation training program for isolated ward settings aimed at training nurses and assessing the program's effectiveness through a randomized controlled trial. A total of 90 nurses participated in the training program and underwent pre- and posttraining assessments. The results showed no significant differences between the 2 groups regarding scores of theoretical knowledge and drills after the training (t_{88}=–0.30, *P*=.75, Cohen *d*=–0.06; Table 2 and *z* score=0.00, *P*>.99; Table 3). However, the intervention group that received NIVR simulation training significantly reduced their total task completion time compared to the control group (t_{88}=5.10, *P*<.001, Cohen *d*=1.08; Table 3). It should be noted that this advantage was only evident in task 3 (receiving patients in the ward: t_{88}=3.22, *P*=.002, Cohen *d*=0.68) and task 6 (leaving the isolation ward: t_{88}=3.03, *P*=.003, Cohen *d*=0.64; Table 3), with no significant differences in the remaining 4 tasks (all *P*>.05), and the average total time difference was approximately 3 minutes (control group: mean 43.91, SD 2.99 min; intervention group: mean 40.77, SD 2.85 min). These findings suggest that NIVR training is comparable to traditional training in enhancing knowledge and skills and can reduce the time nurses take to complete tasks to some extent. Although the study did not conclusively demonstrate a significant advantage of NIVR training, the findings indicate that VR training can facilitate the

XSL•FO
RenderX

effective application of knowledge and skills in real-world settings [34,42]. Moreover, it allows nurses to familiarize themselves with ward environments before pandemics, thus improving their response capabilities. However, further studies are required to confirm other potential benefits. Therefore, it is essential to explore the implementation of VR training to better prepare nurses for future pandemics.

VR training has introduced a highly efficient new method for training nurses in isolation wards. Although traditional training methods, such as using simulated laboratories or observing natural ward settings, are effective, they face challenges including labor intensiveness, financial constraints, resource limitations, time constraints, and risks of cross-infection [13]. Several studies [43,49,51-53] have indicated that VR simulations demonstrate significant potential in pandemic management, covering areas such as infection prevention and control, simulated transmission behaviors, clinical skill training, mental health improvement, and emergency response. These studies have highlighted the vital role of VR technology in managing infectious disease outbreaks such as COVID-19 and Ebola virus disease, leading to enhanced health care professionals' skills, reduced stress, and improved patient experiences. Our isolation ward VR simulation training program focuses on training nurses to enhance their adaptation to isolation ward settings within a simulated environment, improving their preparedness and capabilities in managing infectious diseases. The training integrates a 2D ward map with dynamic interactions in 3D simulated ward environments through NIVR technology, reconstructing realistic isolation ward scenarios. This innovative training method allows nurses to operate in a safe, noncontact virtual environment, significantly reducing the risk of cross infection [13,25]. This approach aids nurses in comprehensively understanding the unique layout and standardized protocols of isolation wards, significantly enhancing training safety and effectiveness [49,51]. Customized course design helps nurses familiarize themselves with different working environments before entering the wards, alleviating their stress and anxiety [32]. Meanwhile, this highly interactive training allows nurses to realistically simulate various emergencies, strengthening their ability and resilience in handling unexpected events [29]. Importantly, NIVR simulation training supports repetitive practice to improve proficiency in critical skills [26]. Therefore, there is a need to focus on and promote VR technology to comprehensively improve health care workers' emergency response capabilities and infection control skills.

During outbreaks of infectious diseases, VR training effectively mitigates the challenges posed by contact restrictions encountered in traditional on-site instruction.

In the early stages of the COVID-19 pandemic, restrictions on clinical internships resulted in the widespread adoption of remote virtual simulation technology, particularly in programs where securing face-to-face clinical internships is difficult [39,54-56]. Research by Zhang et al [13] indicates that web-based VR not only meets the training requirements of isolation wards but also effectively enhances nurses' learning motivation and professional safety [13]. NIVR training, due to its low cost and minimal resource requirements, facilitates remote VR and emerges as an ideal choice for postpandemic nursing education [39]. Our isolation ward virtual simulation training provides an effective learning experience by simulating real environments for learners. The study results show that VR training can replace emergency on-site training, assisting nurses in quickly adapting to their work environment [36,42,57]. Particularly under the circumstances of restricted contact time, remote VR training offers nurses a learning platform unrestricted by time, location, and resources [58,59], improving learning efficiency while enhancing confidence and competence in high-risk areas [42]. Looking ahead, with continuous technological advancements and expanded application scenarios, VR training is poised to play a more significant role in nursing education and medical training. It has the potential to enhance health care professionals' skills and provide more practice opportunities while simultaneously ensuring greater safety for both medical personnel and patients.

Studies indicate that interactive VR training can enhance health care professionals' clinical decision-making skills and critical thinking, thereby improving efficiency and quality of care [27,29,30,60]. In high-risk, high-workload isolation wards, effective time management and operational efficiency are particularly crucial for nurses. Our study shows that nurses who underwent VR training required less time to complete specific tasks, such as entering the wards to receive patients and exiting isolation zones (t_{88}=3.22, P=.002, Cohen d=0.68; t_{88}=3.03, P=.003, Cohen d=0.64, respectively; Table 3). This suggests that VR training may offer advantages in managing complex tasks. However, there are still debates regarding the effectiveness of VR. While some studies highlight VR's ability to effectively enhance nursing students' knowledge levels, it has not yet significantly outperformed traditional methods in skills, satisfaction, confidence, and performance time [38,42,61]. Therefore, policy makers should proceed cautiously when promoting VR training and consider integrating various training methods to ensure comprehensive skill improvement for nurses. Future research can further validate the effectiveness of VR training through large-scale, multicenter, randomized controlled trials and explore its application advantages and limitations in specific medical contexts. In addition, establishing task efficiency assessment mechanisms and delving into the long-term effects and cross-disciplinary applications of VR training can further elevate training outcomes.

Cost-Benefit Considerations

In this study, although both groups had the same training duration, the cost of developing VR training programs is an important factor that needs to be considered. While NIVR training significantly enhanced caregivers' coping skills, its initial development costs were relatively higher compared to traditional training methods. However, as the number of trainees increases, these initial investments are progressively amortized, reducing the average cost per trainee. Farra et al [62] noted that despite the average cost per person for VR training being US $327.78 in the first year, which was US $97.99 higher than the US $229.79 for live drills, the cost gradually decreased in the subsequent years. After 3 years, the average cost of VR training had dropped to US $115.43 per person, while live drills remained at US $229.79, resulting in savings of approximately US $114.36 per person for VR training. Due to its reusability,

XSL•FO
RenderX

VR demonstrated greater cost-effectiveness over a three-year period, proving to be more economically advantageous in the long run compared to traditional on-site practice [62,63]. Halfer and Rosenheck [64] further supported this view by discovering that although the development cost of a VR program was higher than that of paper floor plans, it significantly enhanced nurses' efficiency in navigating a new hospital building and reduced the required practice time. Specifically, the total cost for traditional methods was US $570,000, whereas it was US $421,000 for the virtual methods. Therefore, researchers and decision makers need to carefully weigh the cost and impact of VR training, considering its long-term feasibility and effectiveness in medical training. Future cost-benefit analyses will provide crucial support for these evaluations. A comprehensive assessment of VR education's economic efficiency and academic value is essential for determining its long-term viability when considering implementation costs and training outcomes.

Limitations

This study has several limitations. First, the small sample size limits the generalizability of the results, indicating a need for future research with larger samples. Second, the training content and methods are preliminary and require further refinement to address diverse nursing scenarios. Third, the absence of a baseline speed measure before training leaves open the possibility that observed group differences may partly arise from factors unrelated to the training method. This should be considered when assessing our findings. Moreover, the lack of blinding for assessors introduces potential bias in performance evaluations. Future studies should use blinded assessments to enhance reliability. Although VR simulation training appears promising, further research with improved measures of work efficiency and controlled conditions is essential to validate and expand these initial findings.

Conclusions

This study developed an NIVR simulation training program for isolation wards and evaluated its effectiveness through a randomized controlled trial. While the NIVR simulation training demonstrates faster performance in certain tasks, it shows no significant differences in theoretical and drill scores compared to traditional training. Nevertheless, its safety, repeatability, and realistic simulated environment suggest potential for pandemic response training. Despite not significantly outperforming traditional training methods, NIVR simulation training can still facilitate the practical application of knowledge and skills. It helps nurses familiarize themselves with the ward environment in advance, thus improving their response capabilities. Future research should focus on optimizing training programs and validating and enhancing VR training through large-scale, multicenter randomized controlled trials. In addition, establishing mechanisms for task efficiency evaluation and exploring its long-term effects and interdisciplinary applications will enhance pandemic preparedness.

Acknowledgments

The authors would like to thank all participants in this study and the staff from the Affiliated Nanhua Hospital, University of South China; The First Affiliated Hospital, Hengyang Medical School, University of South China; and Xi'an Jiaotong University Health Science Center. They are also deeply grateful to Kun Feng and the virtual technology team at Xi'an Technological University for their technical support on this study. This study was partially supported by grants from the Fund of Hunan Social Science Achievement Appraisal Committee (grant XSP21YBC228 to DZ) and the fund of the innovative program on COVID-19 prevention and treatment from Hengyang Science and Technology Bureau (grant 202010031581 to DZ), Shaanxi Province (grant SGH140537 to YPZ), and the Xi'an Jiaotong University Fund (grant 202107164 to YPZ).

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to participant privacy concerns but are available from the corresponding author upon reasonable request.

Authors' Contributions

DZ, ZZ, MF, and YPZ contributed to the conception and design of the study. DZ, ZZ, YJC, and MF contributed to the design and development of this virtual reality project. DZ generated a random allocation sequence, while YJC and ZZ participated in recruiting and screening participants. JZ, YJC, LC, MF, and YL contributed to the study delivery, data collection, and data interpretation. DZ, ZZ, and YPZ wrote the manuscript, and all authors read the final manuscript and approved its submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 7344 KB - jmir_v27i1e63131_app1.pdf]

References

XSL•FO
RenderX

1. Fawaz M, Anshasi H, Samaha A. Nurses at the front line of COVID-19: roles, responsibilities, risks, and rights. Am J Trop Med Hyg 2020 Oct;103(4):1341-1342 [FREE Full text] [doi: 10.4269/ajtmh.20-0650] [Medline: 32783796]
2. Han J, Liu Y, Gu F, Li J, Wang D, Zhang Y, et al. Nurses' preparedness to respond to COVID-19 and associated factors after the outbreak in China. Nurs Open 2023 Sep;10(9):6320-6325 [FREE Full text] [doi: 10.1002/nop2.1879] [Medline: 37294093]
3. Li H, Gu M, Yang X. Level and determinants of preparedness for infectious disease outbreak among general hospital nurses [in Chinese]. J Nurs Sci 2021;36(4):58-61. [doi: 10.3870/j.issn.1001-4152.2021.04.058]
4. Liu J, Xiang J, Yu W, Lin M, Wu S. Coping ability to infectious disease emergencies among nurses in TCM hospitals and its influencing factors [in Chinese]. J Nurs Sci 2021;36(9):86-83 [FREE Full text] [doi: 10.3870/j.issn.1001-4152.2021.09.051]
5. Shi Y, Wang J, Yang Y, Wang Z, Wang G, Hashimoto K, et al. Knowledge and attitudes of medical staff in Chinese psychiatric hospitals regarding COVID-19. Brain Behav Immun Health 2020 Apr;4:100064 [FREE Full text] [doi: 10.1016/j.bbih.2020.100064] [Medline: 32289123]
6. Song S, Li X, Bell SA, Yang X, Zhang W, Ann Arbor MI. Emergency response: a cross-sectional study of core competencies for nurses regarding major infectious disease outbreaks. J Emerg Nurs 2021 Nov;47(6):902-913. [doi: 10.1016/j.jen.2021.04.010] [Medline: 34183192]
7. Notice of the National Health and Health Commission regarding the issuance of national care council development plan (2021-2025). China National Health Commission. URL: https://www.gov.cn/zhengce/zhengceku/2022-05/09/content_5689354.htm [accessed 2024-04-29]
8. Adatara P, Kuug AK, Nyande FK, Klutsey EE, Johnson BB, Nyefene MK, et al. A qualitative study on frontline nurses' experiences and challenges in providing care for COVID-19 patients in the Volta region of Ghana: implications for nursing management and nursing workforce retention. Healthcare (Basel) 2023 Apr 04;11(7):1028 [FREE Full text] [doi: 10.3390/healthcare11071028] [Medline: 37046955]
9. Alan H, Eskici GT, Sen HT, Bacaksiz FE. Nurses' disaster core competencies and resilience during the COVID-19 pandemic: a cross-sectional study from Turkey. J Nurs Manag 2022 Apr;30(3):622-632. [doi: 10.1111/jonm.13552] [Medline: 35083821]
10. Brewer CA, Hutton A, Hammad KS, Geale SK. A feasibility study on disaster preparedness in regional and rural emergency departments in New South Wales: Nurses self-assessment of knowledge, skills and preparation for disaster management. Australas Emerg Care 2020 Mar;23(1):29-36. [doi: 10.1016/j.auec.2019.12.005] [Medline: 31926956]
11. Yan X, Liu F, Aertai, Li M. Current situation and influencing factors of nurses' response ability to public health emergencies of infectious diseases medical institutions at different levels in medical association. Occup Health 2022;38(1):89-94 [FREE Full text] [doi: 10.12259/j.issn.2095-610X.S20240227]
12. Pérez-Raya F, Cobos-Serrano JL, Ayuso-Murillo D, Fernández-Fernández P, Rodríguez-Gómez JA, Almeida Souza A. COVID-19 impact on nurses in Spain: a considered opinion survey. Int Nurs Rev 2021 Jun;68(2):248-255 [FREE Full text] [doi: 10.1111/inr.12682] [Medline: 34004014]
13. Zhang W, Luo M, Liu Y, Cai S, Yang Q, Huang Y, et al. A pilot study to investigate the role of virtual reality in the preservice training of nursing staff in isolation wards. Comput Inform Nurs 2022 May 01;40(5):307-316 [FREE Full text] [doi: 10.1097/CIN.0000000000000900] [Medline: 35266899]
14. Liu Q, Luo D, Haase JE, Guo Q, Wang XQ, Liu S, et al. The experiences of health-care providers during the COVID-19 crisis in China: a qualitative study. Lancet Glob Health 2020 Jun;8(6):e790-e798 [FREE Full text] [doi: 10.1016/S2214-109X(20)30204-7] [Medline: 32573443]
15. Guidelines for the prevention and control of novel coronavirus infection in medical institutions. China National Health Commission. URL: http://www.nhc.gov.cn/yzygj/s7659/202109/c4082ed2db674c6eb369dd0ca58e6d30.shtml [accessed 2024-04-29]
16. Gerardi M, Cukor J, Difede J, Rizzo A, Rothbaum BO. Virtual reality exposure therapy for post-traumatic stress disorder and other anxiety disorders. Curr Psychiatry Rep 2010 Aug;12(4):298-305. [doi: 10.1007/s11920-010-0128-4] [Medline: 20535592]
17. Bamodu O, Ye X. Virtual reality and virtual reality system components. In: Proceedings of the 2nd International Conference on Systems Engineering and Modeling. 2013 Sep Presented at: ICSEM '13; April 19-20, 2013; Beijing, China p. 1169-1172 URL: https://www.atlantis-press.com/proceedings/icsem-13/5686 [doi: 10.4028/www.scientific.net/amr.765-767.1169]
18. Lo HH, Zhu M, Zou Z, Wong CL, Lo SH, Chung VC, et al. Immersive and nonimmersive virtual reality-assisted active training in chronic musculoskeletal pain: systematic review and meta-analysis. J Med Internet Res 2024 Aug 19;26:e48787 [FREE Full text] [doi: 10.2196/48787] [Medline: 39159449]
19. D'Amico A, Bernardini G, Lovreglio R, Quagliarini E. A non-immersive virtual reality serious game application for flood safety training. Int J Disaster Risk Reduct 2023 Oct;96:103940. [doi: 10.1016/j.ijdrr.2023.103940]
20. Buttussi F, Chittaro L. Effects of different types of virtual reality display on presence and learning in a safety training scenario. IEEE Trans Vis Comput Graph 2018 Feb;24(2):1063-1076. [doi: 10.1109/TVCG.2017.2653117] [Medline: 28092563]
21. Ren Y, Wang Q, Liu H, Wang G, Lu A. Effects of immersive and non-immersive virtual reality-based rehabilitation training on cognition, motor function, and daily functioning in patients with mild cognitive impairment or dementia: a systematic review and meta-analysis. Clin Rehabil 2024 Mar;38(3):305-321. [doi: 10.1177/02692155231213476] [Medline: 38289618]

XSL•FO
RenderX

22. Dhar E, Upadhyay U, Huang Y, Uddin M, Manias G, Kyriazis D, et al. A scoping review to assess the effects of virtual reality in medical education and clinical care. Digit Health 2023;9:20552076231158022 [FREE Full text] [doi: 10.1177/20552076231158022] [Medline: 36865772]
23. Garmaise-Yee J, Houston C, Johnson T, Sarmiento S. Virtual simulation debriefing in health professions education: a scoping review protocol. JBI Evid Synth 2022 Jun 01;20(6):1553-1559. [doi: 10.11124/JBIES-21-00170] [Medline: 35013042]
24. Lie SS, Helle N, Sletteland NV, Vikman MD, Bonsaksen T. Implementation of virtual reality in health professions education: scoping review. JMIR Med Educ 2023 Jan 24;9:e41589 [FREE Full text] [doi: 10.2196/41589] [Medline: 36692934]
25. Plotzky C, Lindwedel U, Sorber M, Loessl B, König P, Kunze C, et al. Virtual reality simulations in nurse education: a systematic mapping review. Nurse Educ Today 2021 Jun;101:104868. [doi: 10.1016/j.nedt.2021.104868] [Medline: 33798987]
26. Pottle J. Virtual reality and the transformation of medical education. Future Healthc J 2019 Oct;6(3):181-185 [FREE Full text] [doi: 10.7861/fhj.2019-0036] [Medline: 31660522]
27. Turrise SL, Thompson CE, Hepler M. Virtual simulation: comparing critical thinking and satisfaction in RN-BSN students. Clin Simul Nurs 2020 Sep;46:57-61. [doi: 10.1016/j.ecns.2020.03.004]
28. Ni L, Yin X, Liu L, Sun K, Zhang W. Research on the effectiveness of the training of nosocomial infection control specialist nurses under the background of the new crown epidemic based on competence-based theory. Altern Ther Health Med (Forthcoming) 2024 May 17 [FREE Full text] [Medline: 38758138]
29. Mistry D, Brock CA, Lindsey T. The present and future of virtual reality in medical education: a narrative review. Cureus 2023 Dec;15(12):e51124 [FREE Full text] [doi: 10.7759/cureus.51124] [Medline: 38274907]
30. Kolcun K, Zellefrow C, Karl J, Ulloa J, Zehala A, Zeno R, et al. Identifying best practices for virtual nursing clinical education: a scoping review. J Prof Nurs 2023;48:128-146. [doi: 10.1016/j.profnurs.2023.07.001] [Medline: 37775227]
31. Kassutto SM, Baston C, Clancy C. Virtual, augmented, and alternate reality in medical education: socially distanced but fully immersed. ATS Sch 2021 Dec;2(4):651-664 [FREE Full text] [doi: 10.34197/ats-scholar.2021-0002RE] [Medline: 35079743]
32. Jallad ST, Işık B. The effectiveness of virtual reality simulation as learning strategy in the acquisition of medical skills in nursing education: a systematic review. Ir J Med Sci 2022 Jun;191(3):1407-1426. [doi: 10.1007/s11845-021-02695-z] [Medline: 34227032]
33. Hong C, Wang L. Virtual reality technology in nursing professional skills training: bibliometric analysis. JMIR Serious Games 2023 Aug 21;11:e44766 [FREE Full text] [doi: 10.2196/44766] [Medline: 37603389]
34. Sim JJ, Rusli KD, Seah B, Levett-Jones T, Lau Y, Liaw SY. Virtual simulation to enhance clinical reasoning in nursing: a systematic review and meta-analysis. Clin Simul Nurs 2022 Aug;69:26-39 [FREE Full text] [doi: 10.1016/j.ecns.2022.05.006] [Medline: 35754937]
35. Kravitz MB, Dadario NB, Arif A, Bellido S, Arif A, Ahmed O, et al. The comparative effectiveness of virtual reality versus e-module on the training of donning and doffing personal protective equipment: a randomized, simulation-based educational study. Cureus 2022 Mar;14(3):e23655 [FREE Full text] [doi: 10.7759/cureus.23655] [Medline: 35510011]
36. Liaw SY, Ooi SW, Rusli KD, Lau TC, Tam WW, Chua WL. Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. J Med Internet Res 2020 Apr 08;22(4):e17279 [FREE Full text] [doi: 10.2196/17279] [Medline: 32267235]
37. Yang J, Zhou WJ, Zhou SC, Luo D, Liu Q, Wang AL, et al. Integrated virtual simulation and face-to-face simulation for clinical judgment training among undergraduate nursing students: a mixed-methods study. BMC Med Educ 2024 Jan 05;24(1):32 [FREE Full text] [doi: 10.1186/s12909-023-04988-6] [Medline: 38183036]
38. Foronda CL, Fernandez-Burgos M, Nadeau C, Kelley CN, Henry MN. Virtual simulation in nursing education: a systematic review spanning 1996 to 2018. Simul Healthc 2020 Feb;15(1):46-54. [doi: 10.1097/SIH.0000000000000411] [Medline: 32028447]
39. Kaminski-Ozturk N, Martin B. Virtual clinical simulation adoption and use by licensed practical nurse/licensed vocational nurse education programs during the COVID-19 pandemic. J Nurs Regul 2023 Apr;14(1):21-29 [FREE Full text] [doi: 10.1016/S2155-8256(23)00065-0] [Medline: 37035779]
40. Nayahangan LJ, Konge L, Russell L, Andersen S. Training and education of healthcare workers during viral epidemics: a systematic review. BMJ Open 2021 May 28;11(5):e044111 [FREE Full text] [doi: 10.1136/bmjopen-2020-044111] [Medline: 34049907]
41. Lee SH, Yang IS. Enhancing competency in infection prevention and control: identifying priorities for clinical nurse educational needs. Nurse Educ Today 2024 Mar;134:106085. [doi: 10.1016/j.nedt.2023.106085] [Medline: 38181491]
42. Zhang D, Liao H, Jia Y, Yang W, He P, Wang D, et al. Effect of virtual reality simulation training on the response capability of public health emergency reserve nurses in China: a quasiexperimental study. BMJ Open 2021 Sep 22;11(9):e048611 [FREE Full text] [doi: 10.1136/bmjopen-2021-048611] [Medline: 34551944]
43. Moll-Khosrawi P, Falb A, Pinnschmidt H, Zöllner C, Issleib M. Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students during COVID-19 pandemic: a randomised controlled trial. BMC Med Educ 2022 Jun 22;22(1):483 [FREE Full text] [doi: 10.1186/s12909-022-03533-1] [Medline: 35733135]

XSL•FO
RenderX

44. Lee DK, Choi H, Jheon S, Jo YH, Im CW, Il SY. Development of an extended reality simulator for basic life support training. IEEE J Transl Eng Health Med 2022;10:4900507 [FREE Full text] [doi: 10.1109/JTEHM.2022.3152365] [Medline: 35937462]
45. Vankipuram A, Khanal P, Ashby A, Vankipuram M, Gupta A, DrummGurnee D, et al. Design and development of a virtual reality simulator for advanced cardiac life support training. IEEE J Biomed Health Inform 2014 Jul;18(4):1478-1484. [doi: 10.1109/JBHI.2013.2285102] [Medline: 24122608]
46. Wolff G, Bruno RR, Reiter M, Kantzow B, Kelm M, Jung C. Virtual reality device training for extracorporeal membrane oxygenation. Crit Care 2020 Jul 02;24(1):390 [FREE Full text] [doi: 10.1186/s13054-020-03095-y] [Medline: 32616025]
47. Castillo J, Rodríguez-Higueras E, Belmonte R, Rodríguez C, López A, Gallart A. Efficacy of virtual reality simulation in teaching basic life support and its retention at 6 months. Int J Environ Res Public Health 2023 Feb 24;20(5):4095 [FREE Full text] [doi: 10.3390/ijerph20054095] [Medline: 36901106]
48. Li K. Virtual simulation program for nursing in the prevention and control of COVD-19. Higher Education Electronic Audiovisual Press. URL: https://www.ilab-x.com/details/page?id=6687&isView=true [accessed 2024-04-29]
49. Ragazzoni L, Ingrassia PL, Echeverri L, Maccapani F, Berryman L, Burkle FM, et al. Virtual reality simulation training for Ebola deployment. Disaster Med Public Health Prep 2015 Oct;9(5):543-546. [doi: 10.1017/dmp.2015.36] [Medline: 25782591]
50. Buyego P, Katwesigye E, Kebirungi G, Nsubuga M, Nakyejwe S, Cruz P, et al. Feasibility of virtual reality based training for optimising COVID-19 case handling in Uganda. BMC Med Educ 2022 Apr 13;22(1):274 [FREE Full text] [doi: 10.1186/s12909-022-03294-x] [Medline: 35418070]
51. Asadzadeh A, Samad-Soltani T, Rezaei-Hachesu P. Applications of virtual and augmented reality in infectious disease epidemics with a focus on the COVID-19 outbreak. Inform Med Unlocked 2021;24:100579 [FREE Full text] [doi: 10.1016/j.imu.2021.100579] [Medline: 33937503]
52. Vlake JH, van Bommel J, Wils E, Bienvenu J, Hellemons ME, Korevaar TI, et al. Intensive care unit-specific virtual reality for critically ill patients with COVID-19: multicenter randomized controlled trial. J Med Internet Res 2022 Jan 31;24(1):e32368 [FREE Full text] [doi: 10.2196/32368] [Medline: 34978530]
53. Nijland J, Veling W, Lestestuiver BP, Van Driel CM. Virtual reality relaxation for reducing perceived stress of intensive care nurses during the COVID-19 pandemic. Front Psychol 2021;12:706527 [FREE Full text] [doi: 10.3389/fpsyg.2021.706527] [Medline: 34659021]
54. Dewart G, Corcoran L, Thirsk L, Petrovic K. Nursing education in a pandemic: academic challenges in response to COVID-19. Nurse Educ Today 2020 Sep;92:104471 [FREE Full text] [doi: 10.1016/j.nedt.2020.104471] [Medline: 32502723]
55. Jeffries PR, Bushardt RL, DuBose-Morris R, Hood C, Kardong-Edgren S, Pintz C, et al. The role of technology in health professions education during the COVID-19 pandemic. Acad Med 2022 Mar 01;97(3S):S104-S109 [FREE Full text] [doi: 10.1097/ACM.0000000000004523] [Medline: 34789662]
56. Luctkar-Flude M, Tyerman J. The rise of virtual simulation: pandemic response or enduring pedagogy? Clin Simul Nurs 2021 Aug;57:1-2. [doi: 10.1016/j.ecns.2021.06.008]
57. Khanal P, Vankipuram A, Ashby A, Vankipuram M, Gupta A, Drumm-Gurnee D, et al. Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles. J Biomed Inform 2014 Oct;51:49-59 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.005] [Medline: 24732098]
58. Jeon J, Kim JH, Choi EH. Needs assessment for a VR-based adult nursing simulation training program for Korean nursing students: a qualitative study using focus group interviews. Int J Environ Res Public Health 2020 Nov 29;17(23):8880 [FREE Full text] [doi: 10.3390/ijerph17238880] [Medline: 33260356]
59. Thompson DS, Thompson AP, McConnell K. Nursing students' engagement and experiences with virtual reality in an undergraduate bioscience course. Int J Nurs Educ Scholarsh 2020 Sep 16;17(1):54 [FREE Full text] [doi: 10.1515/ijnes-2019-0081] [Medline: 32941161]
60. Kiegaldie D, Shaw L. Virtual reality simulation for nursing education: effectiveness and feasibility. BMC Nurs 2023 Dec 19;22(1):488 [FREE Full text] [doi: 10.1186/s12912-023-01639-5] [Medline: 38114940]
61. Chen FQ, Leng YF, Ge JF, Wang DW, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: meta-analysis. J Med Internet Res 2020 Sep 15;22(9):e18290 [FREE Full text] [doi: 10.2196/18290] [Medline: 32930664]
62. Farra SL, Gneuhs M, Hodgson E, Kawosa B, Miller ET, Simon A, et al. Comparative cost of virtual reality training and live exercises for training hospital workers for evacuation. Comput Inform Nurs 2019 Sep;37(9):446-454 [FREE Full text] [doi: 10.1097/CIN.0000000000000540] [Medline: 31166203]
63. Roy MJ, Sticha DL, Kraus PL, Olsen DE. Simulation and virtual reality in medical education and therapy: a protocol. Cyberpsychol Behav 2006 Apr;9(2):245-247. [doi: 10.1089/cpb.2006.9.245] [Medline: 16640488]
64. Halfer D, Rosenheck M. Virtual education: is it effective for preparing nurses for a hospital move? J Nurs Adm 2014 Oct;44(10):535-540. [doi: 10.1097/NNA.0000000000000112] [Medline: 25280076]

XSL•FO
RenderX

Abbreviations

NIVR: nonimmersive virtual reality
OSCE: Objective Structured Clinical Examination
VR: virtual reality

Edited by T Loetscher; submitted 11.06.24; peer-reviewed by WR Yang, R Salehi, J James; comments to author 15.10.24; revised version received 14.11.24; accepted 23.11.24; published 17.01.25.

Please cite as:
Zhang D, Fu M, Zhang J, Li Y, Chen L, Chen YJ, Zhong Z, Zhang YP
Evaluating Whether Nonimmersion Virtual Reality Simulation Training Improves Nursing Competency in Isolation Wards: Randomized Controlled Trial
J Med Internet Res 2025;27:e63131
URL: https://www.jmir.org/2025/1/e63131
doi:10.2196/63131
PMID:

©Dandan Zhang, MuLi Fu, Jianzhong Zhang, Yuxuan Li, Li Chen, Yong-Jun Chen, Zhefeng Zhong, Yin-Ping Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mass Casualty Incident Training in Immersive Virtual Reality: Quasi-Experimental Evaluation of Multimethod Performance Indicators

Anke Sabine Baetzner[1], MSc; Yannick Hill[2,3,4], PhD; Benjamin Roszipal[5], MMS; Solène Gerwann[1], MSc; Matthias Beutel[1], MSc; Tanja Birrenbach[6], Dr Med, PD; Markus Karlseder[5], MA; Stefan Mohr[7,8], Dr Med; Gabriel Alexander Salg[7,9], Dr Med; Helmut Schrom-Feiertag[10], MSc; Marie Ottilie Frenkel[1,11], Prof Dr; Cornelia Wrzus[12], Prof Dr

[1]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany
[2]Department of Human Movement Sciences, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[3]Institute of Brain and Behaviour Amsterdam, Amsterdam, Netherlands
[4]Lyda Hill Institute for Human Resilience, University of Colorado Colorado Springs, Colorado Springs, CO, United States
[5]MindConsole GmbH, Graz, Austria
[6]Department of Emergency Medicine, Inselspital University Hospital, University of Bern, Bern, Switzerland
[7]Medical Faculty, Heidelberg University, Heidelberg, Germany
[8]Department of Anesthesiology, University Hospital Heidelberg, Heidelberg, Germany
[9]General-, Visceral- and Transplantation Surgery, University Hospital Heidelberg, Heidelberg, Germany
[10]AIT Austrian Institute of Technology, Vienna, Austria
[11]Psychology in Health Care, Faculty Health, Safety, Society, Furtwangen University, Freiburg, Germany
[12]Psychological Institute and Network Aging Research, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Anke Sabine Baetzner, MSc
Institute of Sports and Sports Sciences
Heidelberg University
Im Neuenheimer Feld 700
Heidelberg, 69120
Germany
Phone: 49 6221544339
Email: anke.baetzner@issw.uni-heidelberg.de

Abstract

Background: Immersive virtual reality (iVR) has emerged as a training method to prepare medical first responders (MFRs) for mass casualty incidents (MCIs) and disasters in a resource-efficient, flexible, and safe manner. However, systematic evaluations and validations of potential performance indicators for virtual MCI training are still lacking.

Objective: This study aimed to investigate whether different performance indicators based on visual attention, triage performance, and information transmission can be effectively extended to MCI training in iVR by testing if they can discriminate between different levels of expertise. Furthermore, the study examined the extent to which such objective indicators correlate with subjective performance assessments.

Methods: A total of 76 participants (mean age 25.54, SD 6.01 y; 45/76, 59% male) with different medical expertise (MFRs: paramedics and emergency physicians; non-MFRs: medical students, in-hospital nurses, and other physicians) participated in 5 virtual MCI scenarios of varying complexity in a randomized order. Tasks involved assessing the situation, triaging virtual patients, and transmitting relevant information to a control center. Performance indicators included eye-tracking–based visual attention, triage accuracy, triage speed, information transmission efficiency, and self-assessment of performance. Expertise was determined based on the occupational group (39/76, 51% MFRs vs 37/76, 49% non-MFRs) and a knowledge test with patient vignettes.

Results: Triage accuracy (d=0.48), triage speed (d=0.42), and information transmission efficiency (d=1.13) differentiated significantly between MFRs and non-MFRs. In addition, higher triage accuracy was significantly associated with higher triage

XSL•FO
RenderX

knowledge test scores (Spearman ρ=0.40). Visual attention was not significantly associated with expertise. Furthermore, subjective performance was not correlated with any other performance indicator.

Conclusions: iVR-based MCI scenarios proved to be a valuable tool for assessing the performance of MFRs. The results suggest that iVR could be integrated into current MCI training curricula to provide frequent, objective, and potentially (partly) automated performance assessments in a controlled environment. In particular, performance indicators, such as triage accuracy, triage speed, and information transmission efficiency, capture multiple aspects of performance and are recommended for integration. While the examined visual attention indicators did not function as valid performance indicators in this study, future research could further explore visual attention in MCI training and examine other indicators, such as holistic gaze patterns. Overall, the results underscore the importance of integrating objective indicators to enhance trainers' feedback and provide trainees with guidance on evaluating and reflecting on their own performance.

(J Med Internet Res 2025;27:e63241) doi:10.2196/63241

KEYWORDS

prehospital decision-making; disaster medicine; emergency medicine; mass casualty incident; medical education; eye tracking; emergency simulation; virtual reality

Introduction

Overview

Medical first responders (MFRs) are confronted with extreme demands at mass casualty incidents (MCIs), during which they must attend to more patients than their resources allow [1]. Whether those situations are due to natural disasters, accidents, or terrorist attacks, peak performance of MFRs under such demanding circumstances is crucial. However, current MCI training is insufficient to prepare MFRs adequately for MCIs due to low immersion and realism of classroom learning and scarce real-life exercises that are typically used [2-4].

An increasingly popular tool for MCI training is immersive virtual reality (iVR; [3,5]). iVR is typically experienced through a head-mounted display (HMD) that elicits the user's impression of being completely surrounded by a 3D, virtual world [6-8]. In addition, users are often able to move around, explore, and interact with their environment (eg, by walking and using controllers). Especially in the context of MCI training, iVR offers the possibility to train numerous scenarios in a safe, flexible, and resource-efficient way [5]. iVR training applications also allow for high-quality assessments of performance, which are crucial for enabling systematic and structured training [9] and for assessing the degree to which MFRs are prepared for MCIs. Moreover, high-quality performance assessments allow for the evaluation of training effectiveness and the comparison of training methods, contributing to an ongoing improvement of MCI preparedness [10]. However, a systematic evaluation of performance indicators for iVR MCI training is still missing. This study aims to fill this gap by evaluating the usefulness of different performance indicators in virtual MCI scenarios. Specifically, we first provide an overview of the potential of iVR for MCI simulation training. Next, we discuss various performance indicators that cover different aspects of the performance of MFRs who arrive first at an MCI scene (ie, visual attention, triage accuracy, triage speed, and information transmission). Finally, we introduce self-rated performance as a potential tool for a holistic assessment.

MCI Training and iVR

The gold standard for MCI training is to take part in real-life scenarios with patient actors and manikins [11]. However, such exercises are particularly resource-intensive in terms of preparation time and costs. Therefore, technological training solutions have been developed to increase the level of preparedness, such as training in iVR [5].

iVR provides near-realistic training sessions that require fewer personnel and financial resources compared to real-life MCI exercises with patient actors [4]. iVR also allows the creation of an infinite number of scenarios that can be quickly and flexibly adapted to training objectives [8]. By providing a safe training environment where mistakes can be made and scenarios can be repeated until they are handled correctly [12,13], iVR can be used in training for otherwise dangerous scenarios, such as fire or terrorist attacks. The training and its challenges can also be adapted to individual needs and expertise levels, thus providing an opportunity to maintain a balance by being complex enough to trigger learning without leading to overload [8,14]. On the basis of their current skills, MFRs may be placed in virtual scenarios with few patients and distractions or may train in large-scale MCIs with more patients under complex conditions (eg, nighttime and difficult terrain). Another advantage of iVR-based training is the ability to implement automated monitoring of objective performance indicators, which can help improve both the trainees' learning process and their preparedness [15]. Furthermore, iVR facilitates the inclusion of novel performance indicators, such as visual attention assessed with eye-tracking, which integrates seamlessly with the technology [16].

Performance Indicators

Overview

Performance indicators can be validated with the known-groups validation approach [17,18] which tests the expectation that distinct groups (eg, age groups, experts, and novices) differ on a certain measure [19]. Effective indicators should distinguish between experts and novices, as experts, by definition, possess the knowledge and skills to consistently perform at high levels [20]. Previous studies have classified experts and novices based on various measures, including having or not having a specific

XSL•FO
RenderX

profession or certificate, often intertwined with (years of) job experience (eg, medical students vs physicians) [21]. In addition, experts generally possess more knowledge than novices [22].

Because performance assessments should capture multiple dimensions of performance [9], we evaluate three areas for a broad performance assessment of the first MFRs who arrive at an MCI: (1) the process of gaining an overview of the situation when arriving on scene, which could be assessed based on visual attention, (2) the triage process, and (3) the efficiency of information transmission to the control center. Furthermore, to obtain a holistic assessment of performance, self-reports are also included.

Visual Attention During Orientation at the MCI Scene

Visual attention has been suggested as a potential factor contributing to medical performance [17]. Consistent with the eye-mind hypothesis, gaze fixations are usually the focus of thought [23], which is why the interest in gaze behavior is increasing in research on medical decision-making and medical training [24,25]. Previous research demonstrated that perception and recognition skills improve with increasing medical experience [24]. Furthermore, visual attention differs with expertise in several medical tasks, such as surgical simulations and diagnostic decision-making [24,26]. Among other aspects, experts fixate on task-relevant cues more often and fixate less on task-irrelevant areas than novices [27].

Only a few studies have tested visual attention in prehospital settings thus far. Eye tracking while viewing photos of accident scenes revealed that MFRs fixated significantly longer on task-relevant cues, such as injured people or safety hazards, than non-MFRs [28]. However, the association between visual attention and expertise was less clear when tracking the eye movements of physicians and nurses watching a 2D video of an MCI [29]. Regardless of the number of years working in emergency medical services, all participants spent more time observing the patients who were severely injured. In addition, participants navigated through the scenes with less predetermined or structured patterns than initially expected. Another study assessed the visual attention of college students viewing a computer screen with photos of injuries and information boxes about airway, breathing, and circulation [30]. One group was trained in the *simple triage and rapid treatment* (START) triage algorithm, while the active control group was instructed in patient transport. After training, the START triage group fixated on relevant information boxes significantly faster. However, all of these studies relied on 2D stimuli presented on computer screens, and none tested visual attention of MFRs in 3D iVR MCI scenarios, although differences between 2D input and iVR have been documented for many cognitive and behavioral processes [7].

Accuracy and Speed of the Triage Process

Typical objective performance assessments in MCI training address accuracy and speed of the triage process [5,10]. Effective allocation of available resources and a well-decided sequence of patient care and transport are essential for managing MCI situations [4] and require the ability to triage patients correctly even under high stress. Triage accuracy refers to the correct choice of triage levels based on a specific triage algorithm. The START triage algorithm is an internationally known algorithm developed in the United States and used in several emergency medical services organizations [31]. According to START, patients are classified based on their injuries as green (minor priority), yellow (delayed priority), red (immediate priority), and black (dead or fatally injured). In addition to accuracy, triage speed is also considered crucial in the management of MCI situations so that patients with immediate priority receive medical treatment as quickly as possible [32]. In particular, triage accuracy, but also triage speed, have been used as effectiveness indicators for evaluating MCI training in previous studies [5]. Regarding iVR studies, only a few have included objective performance assessments within the virtual environment. These studies suggest that triage accuracy in iVR improves with more iVR training [33]. Furthermore, when comparing iVR to real-life exercises, MFRs had similar triage accuracy scores [4,34], but completed the triage process faster in iVR [4]. However, emergency physicians with more experience did not have higher triage accuracy scores in iVR than those with less experience [35]. This may be because the study did not aim to compare experts and novices, and those with less experience were already practicing residents. Therefore, the difference in practical experience was probably not large enough to be considered a typical expert-novice comparison.

Transmission of Information to the Control Center

One of the main problems commonly found during major incidents is poor communication with the control center [10,36]. For instance, communication problems, such as excessive radio traffic, can lead to several negative consequences, including increased mortality rates and reduced safety for MFRs at the scene [37]. Although information transmission to the control center is crucial to ensure proper coordination of all rescue services, studies measuring the quality [10] and efficiency (ie, accurate information in few words) of such radio messages are still missing. In Germany, the widely used *scene, safety, situation, and support* (SSSS) scheme serves as a standard for the assessment of onsite emergency situations [38]. This scheme provides a structured approach to assessing critical aspects of an incident. Sometimes referred to as the 3S scheme, with *support* not specifically named, the scheme assists MFRs in identifying and communicating all relevant information, especially potential hazards and environmental risks [39]. *Scene* refers to the assessment of the emergency site. In terms of safety, MFRs assess the risk to themselves and others. *Situation* refers to an estimation of the number of patients and evaluation of injury mechanisms. *Support* refers to the possible need for reinforcements and other emergency services, such as police and fire brigades [38]. On the basis of interviews and workshops with European MFRs, communication and the correct use of the scheme were identified to be key performance indicators in training for high-stress situations like MCIs [39]. While no studies seem to have evaluated the efficiency of information transmission, it was found that trained physicians reported more accurately than untrained physicians in a real-life MCI exercise [40].

XSL•FO
RenderX

Subjective Performance Ratings

A large part of MCI training studies evaluates the training effectiveness with self-rated indicators, such as self-ratings in knowledge and skills. For example, emergency medical technicians with more work experience rate their triage skills more favorably than those with less work experience [41]. However, general skill assessment does not necessarily relate to performance in a specific situation, and studies assessing self-rated MCI performance directly related to a previously completed simulation are still missing [5]. Not specific to MCI training but to medical education in general, previous research suggests that self-assessed performance has at least a low to moderate validity [42,43]. Self-rated indicators have the advantage of being easily implemented in all possible training modalities without any technical effort. Although objective performance assessment can be implemented in iVR, such assessment may be more complex in other training modalities, impairing potential comparison studies. If a global subjective performance indicator is accurate, it may be a low-threshold indicator for such studies. Including self-assessment alongside objective performance evaluation could also enhance the debriefing process by encouraging trainees to reflect on their performance and improve self-awareness [44]. However, comparisons between (global) subjective performance and objectively measured behavior in iVR MCI training are still lacking.

Research Aims and Hypotheses

This study aimed to investigate whether different performance indicators that were used in previous MCI training studies can be effectively extended to MCI training in iVR. As a method of validation [17], indicators were tested with a focus on their ability to differentiate between different levels of expertise. For comprehensive testing, 2 indicators of expertise were used, including occupational group as a dichotomous indicator (MFR or non-MFR) and triage knowledge test scores as a continuous indicator. For a broad assessment, we investigated visual attention after arriving on scene, triage accuracy, triage speed, information transmission efficiency, and subjective performance. Overall, this study intends to provide insights into which indicators are suitable for incorporation into the design of effective iVR MCI training programs and MFR performance assessments. The following hypotheses were formulated: hypothesis 1a: with greater expertise, medical staff pays more attention to task-relevant information during MCIs; hypothesis 1b: with greater expertise, medical staff show greater triage accuracy during MCIs; hypothesis 1c: with greater expertise, medical staff complete the triage process faster during MCIs; hypothesis 1d: with greater expertise, medical staff transmit information more efficiently; hypothesis 2: attentional indicators demonstrate incremental value in the discrimination of different levels of expertise beyond performance indicators, such as speed, accuracy, and information transmission during triage; and hypothesis 3: subjective evaluation of one's own performance is better with greater attention to task-relevant information, greater triage accuracy, faster triage time, and more efficient transmission of information.

The wording of 2 hypotheses has changed from the preregistration. Previously, they were "Hypothesis 1d: with greater expertise, medical staff transmits information more quickly and completely" and "Hypothesis 3: subjective evaluation of own performance is better with greater attention to task-relevant information, greater triage accuracy, faster triage time, and faster transmission of information."

Methods

Study Design

In this quasi-experimental, multimethod study, participants (MFRs and non-MFRs) completed virtual MCI scenarios while their performance was assessed.

Procedure

Overview

Interested people visited a website following the link or QR code provided on the recruitment materials. On the website, they received information on the study aims and procedure, provided informed consent, and answered a web-based preliminary questionnaire covering demographics, personal characteristics, and a triage knowledge test. The web-based questionnaire was answered before participants came to their appointment to shorten the time in the laboratory. At the end of the questionnaire, participants scheduled a 2-hour appointment in the VR laboratory. In the laboratory, participants filled out additional questionnaires, were shown the START triage algorithm, and engaged in an iVR familiarization scenario with people who were uninjured and at no accident site. During the familiarization, participants practiced information gathering and navigation within the virtual environment. The familiarization ended when participants announced that they feel sufficiently prepared. The main phase of the study involved 5 virtual MCI scenarios to cover the different performance levels of our sample, thus avoiding floor and ceiling effects and increasing the reliability of the measurements. The scenario order was randomized for each participant using a random number generator.

MCI Scenarios in iVR

The scenarios were built with the XVR software (XVR Simulation BV) and consisted of traffic accidents that varied in scenario complexity (Table 1; Multimedia Appendix 1 gives a detailed description). Tasks included gaining an overview of the situation, performing triage, and transferring all relevant information to the control center via radio messages.

XSL•FO
RenderX

Table 1. Scenario descriptions. After the familiarization scenario, the scenario order was randomized for each participant.

Scenario	Scene description	Number and triage levels of patients	Duration in minutes, mean (SD)
Familiarization	Quiet residential area, no accident, clear vision, middle of the day; 6 healthy people present, 3 of them sitting in cars	0 patients but healthy people to practice the iVR[a] handling	5.78 (1.71)
1 (very low difficulty)	Accident with 2 cars on a countryside road, clear vision, middle of the day	3 green patients and 1 yellow patient	3.32 (1.03)
2 (low difficulty)	Accident with a car, a van, and a motorcyclist on a motorway, clear vision, cloudy day	1 green patient, 3 yellow patients, and 1 red patient	5.33 (1.37)
3 (medium difficulty)	Accident with a car and an SUV[b] in a busy inner-city area with a crowd of bystanders around, clear vision, cloudy day	1 green patient, 3 yellow patients, and 2 red patients	5.06 (1.59)
4 (greater difficulty)	Expressway accident with 3 cars and a van; several bystanders, twilight, end of the day	1 green patient, 3 yellow patients, 3 red patients, and 1 black patient	7.28 (1.97)
5 (highest difficulty)	Expressway accident involving a bus and a truck, nighttime and fog	4 green patients, 6 yellow patients, 4 red patients, and 4 black patients	10.66 (3.08)

[a]iVR: immersive virtual reality.

[b]SUV: sports utility vehicle.

Participants were asked to remain in the starting position for the first 30 seconds during each scenario and were allowed to turn and look around. This time frame was used to assess initial attention processes measured via eye tracking. During the scenario, participants could use visual cues like wounds to select triage levels. Further information about the patients could be acquired verbally and was provided with the help of standardized audio tracks. For example, if participants inquired about the respiration rate, they would hear a prerecorded answer. Obtainable information included whether the patient could walk, whether the airways were clear, respiration rate, recapillarization time, presence of a radial pulse, existence of heavy bleeding, the patient's responsiveness to simple instructions, and localization of injuries (eg, head injury and leg injury). The participants could announce their intention to perform actions, including clearing the airway, stopping heavy bleeding, and conducting triage. After each scenario, they filled out questionnaires and had the option to reread the START algorithm.

Hardware Description

The Varjo Aero HMD had a display resolution of 2880×2720 pixels per eye at 90 Hz. Its field of view was 115° horizontally and 134° diagonally (at 12 mm eye relief), and the gaze data output frequency was 200 Hz. The HMD was connected to a laptop (ROG Strix G with Windows 10, Intel Core i7-9750H central processing unit at 2.60 Hz 2.59 GHz, NVIDIA GeForce TTX 2070 graphical processing unit, 32 GB of RAM) through the standard HMD cable and Varjo laptop adapter. Participants used the Varjo Aero HMD from Varjo Technologies Oy and could freely move within a 3×4-m size area. For greater distances, participants used controllers to teleport (Figure 1).

Figure 1. Immersive virtual reality (iVR) lab (left) and use of teleportation in iVR (right).



Participants

Emergency services personnel, hospital personnel (physicians and nurses), and medical students (all semesters) were recruited to participate in this study. Inclusion criteria were a minimum age of 18 years, proficiency in German, and the absence of a hearing aid. Participants were recruited through social media,

XSL•FO
RenderX

email distribution lists, flyers, and short presentations at local hospitals, emergency medical services, and university courses. Data collection lasted from February to October 2023. Of the 150 participants who filled out the web-based, preliminary questionnaire, 76 (50.7%) scheduled an appointment with our iVR laboratory and participated in the study. Participants were aged between 18 and 49 (mean 25.54, SD 6.01) years, and 59% (n=45) of participants were male. Among the 76 participants, 51% (n=39) were categorized as MFRs (n=36, 92% paramedics and n=3, 8% emergency physicians). The remaining 49% (37/76) non-MFR participants were mainly medical students (n=30, 81%), other medical staff not working in prehospital emergency settings (n=4, 11%), or both at the same time (n=3, 8%).

Measures

Study Data

The data presented in this paper were derived from a larger project, which focused on investigating the stress dynamics and performance of MFRs in virtual MCI training. For detailed information of all measures and instruments used during the study, refer to the construct overview on the web [45].

Web-Based Preliminary Questionnaire

Demographic information included age, gender, profession, years of job experience in the medical sector, and prior MCI training in hours. In the case of profession, participants could choose multiple options from the following: emergency physicians, other physicians, paramedics, emergency medical technicians, medical students, or specify another option through an open text field. Prior MCI training was assessed with the following answering options: none, 1 to 5 hours, 5 to 10 hours, 10 to 20 hours, 20 to 30 hours, 30 to 40 hours, 40 to 50 hours, or >50 hours.

In terms of expertise, participants were classified as MFRs if they were emergency physicians or paramedics; otherwise, they were classified as non-MFRs. The second expertise measure was a triage knowledge test based on Cuttance et al [46] and adapted to the START algorithm. Participants had 10 minutes to assign triage levels to 20 case descriptions, after which the test ended, and the next page opened. One point was awarded for each correctly assigned color, resulting in a score range of 0 to 20.

Triage algorithm or algorithms. Participants were asked which triage algorithm they typically used in the field or in training sessions, including the option to select “none.”

Prior experience with iVR was assessed with the item “How much prior experience do you have with VR?” and a dropdown menu with 9 answering options from “no experience” to “daily use of VR.”

Visual Attention During Orientation in Each iVR MCI Scenario (Onsite)

Attention to task-relevant information was assessed using eye tracking in the first 30 seconds of each scenario. Eye-tracking data were collected using Varjo Base and analyzed with the software iMotions [47]. The gaze behavior was analyzed in terms of average fixation durations and number of fixations of the areas of interest (AOIs; [48,49]). AOIs included patients, vehicle impact zones, safety aspects, and a distractor (Multimedia Appendix 1 gives scenario descriptions). Within the iMotions software, AOIs were defined graphically, and the software automatically computed the fixation parameter [50].

Objective Triage Accuracy and Speed in Each iVR MCI Scenario (Onsite)

Triage accuracy was measured by the number of correctly assigned triage colors per scenario. The accuracy score for each scenario was divided by the number of patients in that scenario. The accuracy scores of all 5 scenarios were then averaged, leading to possible values between 0 and 1 (standardized Cronbach α=0.76).

Speed of triage was measured as the time from the start of the scenario to the completion of triaging the last patient (standardized Cronbach α=0.82). Note, however, that due to the different lengths of the scenarios, averaging with nonstandardized times would lead to a stronger weighting of the more complex scenarios with longer duration and a larger temporal variance. To avoid this, a standardization was performed: values were divided by the average triage speed of the specific scenario before being averaged. Consequently, values <1 signified a triage process faster than the average, while values >1 denoted a slower-than-average triage pace.

Information Transmission to Control Center in Each iVR MCI Scenario (Onsite)

To assess how quickly and completely medical staff transmit information, an efficiency measure was formed as a combination of completeness and speed. Participants’ radio messages during the scenarios were recorded and transcribed according to the content-semantic transcription by Dresing and Pehl [51]. A coding system based on the qualitative content analysis by Mayring [52] was then applied, with categories derived from the SSSS scheme [38] and added to a coding template. Next, 2 independent raters (university students specifically trained for this) analyzed the transcripts with the software MAXQDA (VERBI GmbH), without access to any information on the participants (interrater agreement: 97% for categorization and 93% for assessment of correctness). Discrepancies were resolved through discussion and, if necessary, together with a third rater (ASB). Statements were categorized, with a maximum of 1 point awarded for each of the 4 categories. Half a point was awarded for the category *scene* if participants only mentioned that it was a traffic accident, and a full point if more information was given (eg, the type of street or number of vehicles involved). Furthermore, a penalty for mistakes was implemented: 0.3 points were deducted for each error in the respective category. Regarding the category *situation*, an error was given if the number of patients was not stated correctly. In the 2 most complex scenarios, no penalty was applied if an estimate close to the correct number was reported (+2 or –2; eg, for scenario 4 with 8 patients, “almost 10” or “between 7 and 9”; n=4 cases). No penalty was given for incorrectly stated triage colors, as these are already covered by the triage accuracy variable. The scores of the 4 categories were added up (range: 0-4 points) and standardized based on the number of words in the radio message

XSL•FO
RenderX

(ie, divided by the word count). This standardization accounted for unclear, redundant, and superfluous communication (possibly even in the same time by simply speaking faster), which poses a risk to the management of MCIs [53]. Finally, the average score across all 5 scenarios was calculated (standardized Cronbach α=0.83; range 0.03-0.15).

Subjective Performance of Each iVR MCI Scenario (Onsite)

Subjective performance was assessed with 2 items after each scenario: "How do you rate your performance in the last scenario?" (1=very bad and 10=very good) and "What school grade does your performance correspond to?" (scale: 1-6, with 1 being the best grade). We used 2 items to increase the reliability in capturing the construct. Before the items were averaged, the second item was inverted and transformed to match the 1 to 10 scale of the first item (Cronbach α=0.86):



(1)

Analyses

Analyses were conducted with R (version 4.3.2; R Core Team). Hypotheses 1a to 1d were tested with both expertise measures separately, distinguishing between MFRs and non-MFRs as well as using the triage knowledge test. For hypothesis 1a specifically, a multivariate ANOVA was used because of the multiple AOI categories. According to the Mardia test, the assumption of multivariate normal distribution was violated, but multivariate ANOVAs are robust to this specific violation, particularly in cases of homogeneous covariance matrices, which was given [54]. For hypotheses 1b, 1c, and 1d, independent *t* tests were performed to test for differences between MFRs and non-MFRs. For hypothesis 1b, the Welch *t* test was used because the Levene test indicated a violation of the assumption of homogeneity of variance. Hypotheses 1c and 1d were tested with the student *t* test. According to the Shapiro-Wilk test, the assumption of normality was violated testing hypotheses 1b and 1d. However, *t* tests are largely robust against this violation, especially because both group sample sizes were relatively similar and >30 [55,56]. The Bonferroni-Holm method was used to control for multiple testing in the *t* tests (hypotheses 1b-1d). Spearman rank-order correlations were used to examine the relationship between the knowledge test score and the outcome variables in hypotheses 1a to 1d, as they are robust to distributional violations and outliers. In addition, when correlations were significant, regression analyses were conducted to test the relationship while controlling for age and gender. We selected these 2 control variables because they are potential influencing factors in iVR performance situations [57,58]. For the regression analyses, independent variables and covariates were centered (for gender: –1=female and 1=male). Hypothesis 3 was tested with Spearman correlation analyses.

Outliers were defined as values that were 3 SDs above or below the mean. These values were winsorized (ie, set to mean +3 or –3 SDs, respectively) to reduce the influence of outliers on the results [59]. Analyses were conducted with adjusted outliers and with original variables to examine the robustness of results. When technical problems occurred during the assessment (eg, technical failure of the eye tracking or the microphone for radio message recording), the respective measure was excluded instead of excluding the entire dataset of the participant. Therefore, hypothesis 1a, which addressed eye-tracking data, was tested with a sample of n=71 for the AOIs *patients*, *safety*, and *vehicle impact zone*, and n=57 for the *distractor*. Hypothesis 1b and c, which examined triage behavior, were tested with n=74 and hypothesis 1d, which focused on information transmission, was tested with n=70 participants. There were no missing values for subjective performance (hypothesis 3).

Ethical Considerations

This study was preregistered [60], approved by the ethics committee of the Faculty of Behavioral and Empirical Cultural Sciences at Heidelberg University (AZ Bae 2023 1/1), and carried out in accordance with the Helsinki Declaration on Ethical Principles of Research Involving Humans. To ensure transparency in the analysis process, the R Markdown file, containing the R code and all results, is available on the web [45]. Data was collected in a pseudonymized manner and subsequently anonymized for analysis. As an incentive, the participants received €25 (US $27.48 on Feb 1, 2023).

Results

Analyses

To capture 2 different measures of expertise, analyses were conducted with occupational group as a dichotomous measure (MFRs vs non-MFRs) and the triage knowledge test scores as a continuous measure. Means and SDs of performance indicators per scenario can be found in Table S1 in Multimedia Appendix 1.

Demographics

The classification of expertise based on the profession was supported by the results of the knowledge test. MFRs had a significantly higher knowledge score than non-MFRs, $t_{Welch66.17}$=–3.43, *P*=.001, *d*=0.79. More information on the groups can be found in Table 2. Most participants were not familiar with any triage system before this study (43/76, 57%; 34/43, 79% of those being non-MFRs). Some (27/76, 36%) participants knew the mSTART (modified START) algorithm, 12% (9/76) knew PRIOR (Primäres Ranking zur Initialen Orientierung im Rettungsdienst), and 1% (1/76) the START algorithm. Of the non-MFR group, 65% (24/37) had no prior MCI training, and 30% (11/37) non-MFRs had <5 hours. 3% (1/37) had up to 20 hours, and 3% (1/37) had up to 30 hours of MCI training. Regarding the MFRs' prior MCI training experience, 8% (3/39) reported none, 31% (12/39) had up to 5 hours, 18% (7/39) had up to 10 hours, 18% (7/39) had up to 20 hours, 5% (2/39) had up to 30 hours, 8% (3/39) had up to 40 hours, 5% (2/39) had up to 50 hours, and 8% (3/39) had >50 hours.

XSL•FO
RenderX

Table 2. Description of expertise groups: medical first responders (MFRs) and medical students and other medical staff (non-MFRs).

Characteristics	MFRs (n=39)	Non-MFRs (n=37)
Age (y), mean (SD)	27 (7.15)	24 (4.07)
Gender: female, n (%)	11 (28)	20 (54)
Job experience (y), mean (SD)	6.53 (7.01)	1.62 (1.94)
Experience with iVR[a], none or only 1-2 times, n (%)	33 (85)	34 (92)
Knowledge test score, mean (SD)[b]	13.46 (2.73)	10.89 (3.70)

[a]iVR: immersive virtual reality.

[b]The knowledge test score could range from 0 to 20.

Visual Attention During Orientation and Expertise (Hypothesis 1a)

MFRs and non-MFRs did not differ significantly in their mean durations of average fixation in the 4 AOI categories of patients, vehicle impact zones, safety aspects and distractor, Pillai trace=0.03, $F_{4,52}$=0.40, P=.81. Furthermore, there were no significant correlations between the knowledge test score and the average fixation duration in any of the 4 AOI categories (Table 3).

Regarding the fixation count, MFRs and non-MFRs did not differ across the 4 AOI categories (Pillai trace=0.15; $F_{4, 52}$=2.37; P=.06). Again, the triage knowledge test was not significantly associated with the fixation count in any of the AOI categories (Table 4).

Table 3. Duration of average fixation (DOAF) of specific area of interest (AOI) categories across the 5 scenarios. Time in milliseconds: fixation duration was first averaged per scenario and then averaged across the 5 scenarios.

DOAF AOIs	Patients	Safety	VI[a] zone	Distractor
Examples of AOI cues	All patients and no bystanders	Scenario 1: spilled oil; scenario 2: ongoing traffic; and scenario 5: ongoing traffic, spilled oil, broken glass	VI zones in all scenarios; in scenario 3 and scenario 4 mean value of 2 VI zones	Only in scenario 3: filming bystander
Value per group, mean (SD)				
MFRs[b]	397.41 (105.55)	319.09 (105.69)	314.50 (80.39)	215.13 (92.28)
Non-MFRs	403.18 (106.97)	319.37 (91.20)	315.21 (92.61)	241.35 (118.76)
Correlation with triage knowledge test				
ρ	.06	–.10	.17	.23
P value	.61	.40	.15	.08

[a]VI: vehicle impact.

[b]MFR: medical first responder.

XSL•FO
RenderX

Table 4. Number of fixations of specific area of interest (AOI) categories across 5 scenarios. Fixation count was summed up per scenario and then averaged across the 5 scenarios.

FC[a] AOIs	Patients	Safety	VI[b] zone	Distractor
Examples of AOI cues	All patients and no bystanders	Scenario 1: spilled oil; scenario 2: ongoing traffic; and scenario 5: ongoing traffic, spilled oil, and broken glass	VI zones in all scenarios; in scenario 3 and scenario 4 mean value of 2 VI zones	Only in scenario 3: filming bystander
Value per group, mean (SD)				
MFRs[c]	12.51 (4.43)	8.90 (5.46)	8.84 (3.29)	2.73 (2.13)
Non-MFRs	12.76 (4.30)	11.51 (6.47)	7.71 (2.80)	3.85 (2.36)
Correlation with triage knowledge				
ρ	–.09	–.16	.02	.04
P value	.46	.18	.89	.76

[a]FC: fixation count.

[b]VI vehicle impact.

[c]MFR: medical first responder.

Triage Accuracy and Expertise (Hypothesis 1b)

MFRs had a significantly higher triage accuracy than non-MFRs ($t_{Welch61.62}$=–2.04; *P*=.02 [Bonferroni-Holm corrected *P*=.045]; *d*=0.48). On average, MFRs triaged 84% (SD 12%) of the patients correctly while non-MFRs triaged 77% (SD 18%) correctly (Figure 2A depicts the distributions).

Furthermore, a higher knowledge test score was associated with significantly higher triage accuracy (Spearman ρ=0.40; *P*<.001). A multiple regression analysis was used to test this relationship while controlling for age and gender. More prior triage knowledge significantly predicted higher triage accuracy during the scenarios (*b*=0.02; SE 0.005; *P*=.002), while age (*b*=–0.001; *P*=.67) and gender (*b*=0.005; *P*=.89) were not significant predictors. The model explained 9% of the variance (adjusted R^2) (R^2=.13; $F_{3,70}$=3.45; *P*=.02). However, according to the Shapiro-Wilk test, the assumption of normal distribution of the residuals (*P*<.001) was violated. Therefore, bootstrapping with 5000 draws was used to test the stability of the results. Again, prior triage knowledge significantly predicted triage accuracy during the scenarios (*b*=0.02, SE 0.00, 95% CI 0.01-0.03).

Figure 2. (A) density plots per group for triage accuracy, (B) triage speed, and (C) information transmission efficiency. The dashed lines mark the average per group. MFR: medical first responder.

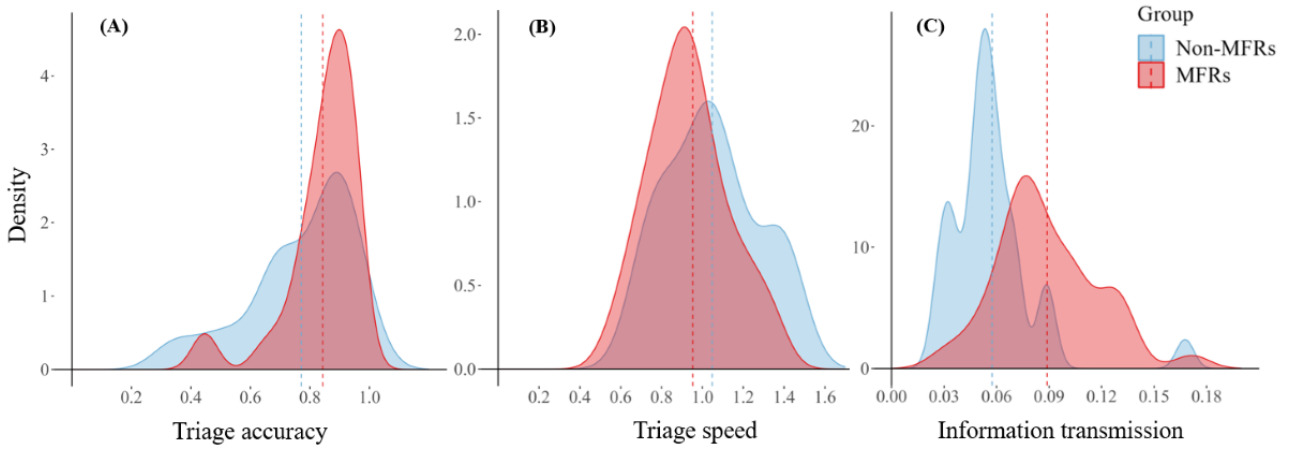

Triage Speed and Expertise (Hypothesis 1c)

MFRs completed the triage process significantly faster than non-MFRs (t_{72}=1.79; *P*=.04 [Bonferroni-Holm corrected *P*=.045]; *d*=0.42). On average, MFRs required 95% of the average triage speed, whereas non-MFRs needed 5% more than the average (MFRs: mean 0.95, SD 0.23; non-MFRs: mean 1.05, SD 0.23; Figure 2B depicts the distributions). The knowledge test score was not significantly associated with triage speed (Spearman ρ=–.04; *P*=.72).

Information Transmission and Expertise (Hypothesis 1d)

MFRs transmitted information significantly more efficiently than non-MFRs (MFRs: mean 0.09, SD 0.03, and non-MFRs: mean 0.06, SD 0.03; t_{68}=–4.74; *P*<.001 [Bonferroni-Holm corrected *P* <0.001]; *d*=1.13) (Figure 2C depicts the distributions). However, there was no significant correlation between the knowledge test score and efficiency of information transmission (Spearman ρ=0.19; *P*=.11).

XSL•FO
RenderX

Explorative analyses were conducted with information transmission scores and word counts separately. Averaged across the scenarios, MFRs did not transmit significantly more correct information (MFRs: mean 2.15, SD 0.80, and non-MFRs: mean 2.07, SD 0.72; maximum score=4; t_{68}=–0.44; P=.33; d=0.11). However, MFRs transmitted information in fewer words (MFRs: mean 32.34, SD 20.99, and non-MFRs: mean 46.09, SD 23.68; t_{68}=2.57; P=.006; d=0.62). Knowledge test scores were not significantly correlated with the number of correct information transmissions (Spearman ρ=–0.05; P=.67) or information transmission length (Spearman ρ=–0.13; P=.29).

Testing the Incremental Value of Attention (Hypothesis 2)

As reported for hypothesis 1a, the attentional indicators assessed with eye tracking did not discriminate between different levels of expertise. Accordingly, testing the incremental value of visual attention indicators beyond the other performance indicators for distinguishing levels of expertise (hypothesis 2) was inapplicable.

Subjective Performance and Objective Performance Indicators (Hypothesis 3)

There were no significant correlations between subjective performance (mean 6.77, SD 1.13) and any of the objective performance indicators (all P>.05; Figure 3). Conversely, significant associations were observed between some objective performance indicators, including between greater triage speed and less efficient information transmission (ρ=–0.33; P=.01), and between less visual attention to safety cues (as measured by fixation count) and greater triage accuracy (ρ=–0.31; P=.01).

Figure 3. Spearman correlation analysis of the performance indicators and immersive virtual reality (iVR) experience (above diagonal: ρ values, below diagonal: P values). iVR experience refers to prior iVR experience as assessed in the preliminary questionnaire; prior iVR experience was not winsorized because of the ordinal data structure. DOAF: duration of average fixation; FC: fixation count; VI: vehicle impact.

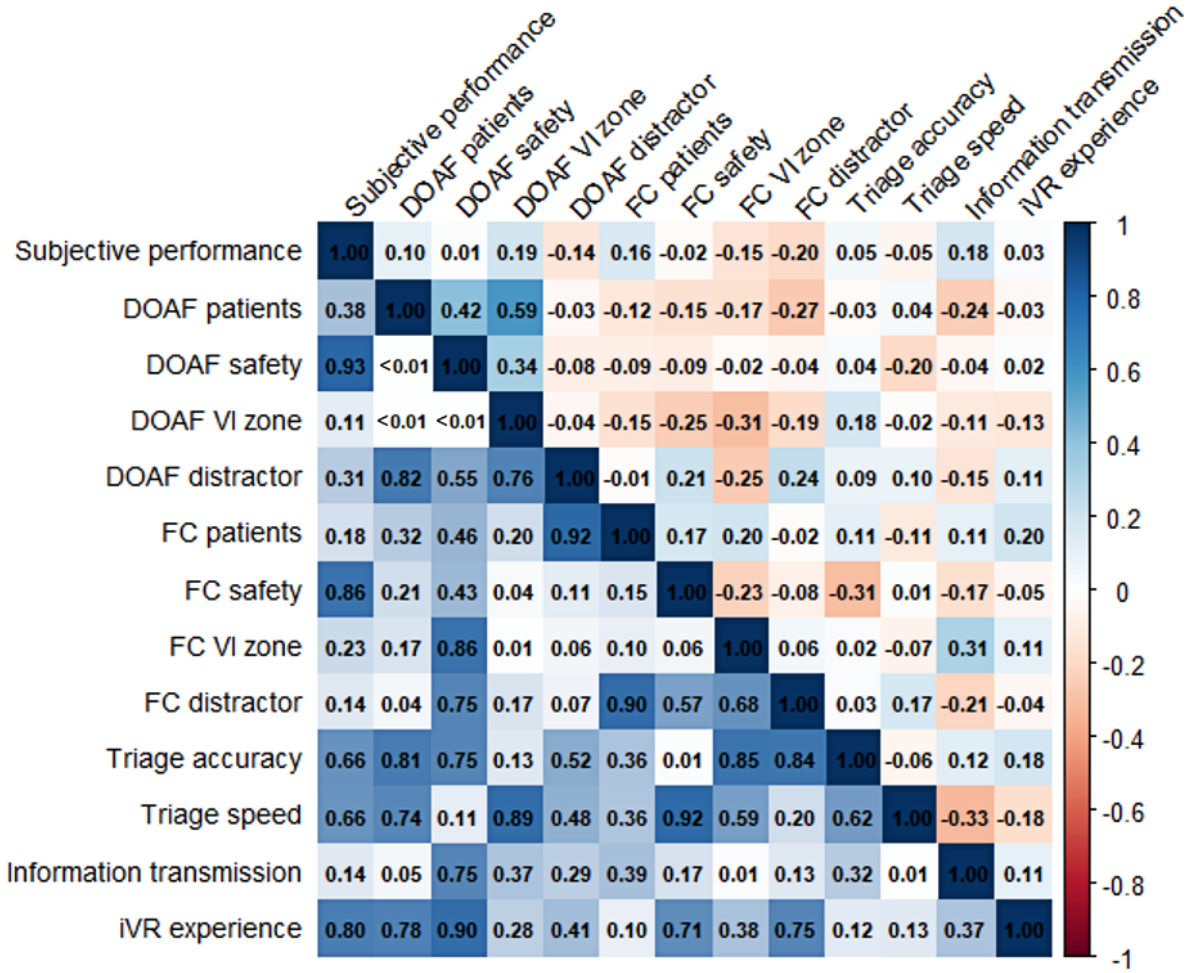

Explorative Analyses

Exploratively, we tested whether prior iVR experience was associated with any of the examined performance indicators. For both MFRs and non-MFRs, the overall median value was 2 (ie, tried iVR 1-2 times). Prior iVR experience did not correlate with any performance indicator (all P>.05; Figure 3).

Discussion

Principal Findings

The prehospital triage process is a central aspect of managing MCIs, and if done correctly, leads to an efficient allocation of treatment and transport. As MCI real-life exercises are resource-intensive and therefore infrequently conducted, iVR

XSL•FO
RenderX

has emerged as a new triage training method [5]. Due to the novelty of such training, systematic evaluations of potential indicators that can be used to assess performance are still needed. This study aimed to provide such an evaluation. The objective performance indicators of triage accuracy and speed, as well as information transmission efficiency, significantly differentiated between MFRs and non-MFRs. In addition, higher triage accuracy was significantly associated with more prior triage knowledge, even when controlling for age and gender. Interestingly, however, visual attention did not differ with the level of expertise. Furthermore, in contrast to the third hypothesis, subjective performance was not correlated with any other performance indicator. Next, we discuss possible explanations as well as future directions for deriving meaningful performance indicators in iVR MCI training.

Visual Attention During Orientation

Visual attention represents a relatively novel potential performance indicator that has received limited attention from research in prehospital contexts. In other medical fields, studies using eye tracking have already identified differences in visual attention between experts and novices [17,24,26]. In this study, we found no significant associations between visual attention and expertise, which aligns with findings based on 2D videos of an MCI [29]. However, the results stand in contrast to 2 studies using 2D, prehospital images as stimuli [28,30]. Loth et al.[30] showed that after START triage training, participants directed their attention more swiftly to triage information boxes. The transferability of such findings to a 3D virtual scenario without information boxes remains uncertain [30]. In the second study, MFRs and novices differed significantly in their visual attention when looking at MCI photos without information boxes [28]. Several factors could explain the discrepancies. First, the former study used a purely 2D examination, which may not translate seamlessly into 3D. In contrast, iVR offers greater ecological validity for natural gaze behavior by providing a larger field of view and the ability to visually explore the environment by moving the head [49]. Moreover, the stereoscopic visualization in iVR enhances depth perception through binocular cues [61]. Previous research has demonstrated that 3D input can significantly alter attentional processes compared to 2D input [62], although it has not yet been directly compared in the context of MCI scenarios. The additional visual cues in the iVR scenarios of this study may have enabled non-MFRs to more quickly identify relevant AOIs, thereby reducing differences between MFRs and non-MFRs. Second, it is also possible that MFRs had to exert more effort in filtering out irrelevant visual stimuli than they would with 2D input, which could diminish potential differences between MFRs and non-MFRs. Furthermore, the former study [28] included true novices, whereas this study included non-MFRs with medical knowledge (primarily medical students), suggesting a possible power issue due to minimal effects between these groups, necessitating a larger sample size. Alternatively, it is plausible that both groups in this study possessed a minimum level of medical expertise necessary for effective visual attention, and effective visual attention remains relatively stable with additional triage knowledge.

Furthermore, for MFRs, visual attention, especially toward nonpatient information, such as safety threats and distractors, may be less critical than in professions such as law enforcement or military, where constant threat awareness is an integral part of training. For instance, police officers with specialized training have exhibited superior visual search behavior for threats and overall performance in real-life exercises compared to their counterparts without such training [63]. Another possibility is that differences in visual attention may only manifest when observing individual patients and their injuries, whereas our study primarily involved participants viewing a broad overview of an accident scene during eye-tracking measurements. Finally, it is worth considering that alternative visual attention indicators may offer more nuanced insights. The selection of appropriate eye-tracking measures is crucial, as certain fixation and saccade metrics or scan paths may be more applicable for specific purposes than others, depending on the task and study population [64]. This study used 2 eye-tracking measures often used in studies on visual attention in medical contexts [21,24], the number and length of fixations in predefined areas of interest (ie, patients, safety aspects, vehicle impact zone, and distractor AOIs). While fixation-derived metrics can be used as a measure for visual attention and processing, saccade-derived metrics could provide further insights into visual search strategies [24]. Furthermore, experts may exhibit distinct eye movement patterns and rapidly develop a holistic view of the scene, as observed in other medical domains [24].

Triage Accuracy and Speed

During real-life MCI exercises, triage accuracy and speed have proved to be useful as performance indicators, although usually not comparing different levels of expertise but prepost improvements after training [65-67]. Other studies used the indicators to compare different training methods [4,68-70]. Triage accuracy and speed have also been used in iVR settings [4,33,35], with evidence suggesting that with more iVR training, participants become better and faster at triaging in virtual MCI situations [33].

The results of this study suggest that accuracy is a suitable performance indicator to differentiate between different levels of expertise, with a medium effect size in terms of profession (MFRs vs non-MFRs) and a medium to large effect size in terms of triage knowledge [71]. These results are consistent with a previous study that found paramedic students to be as good at triage accuracy in a real-world exercise as in an iVR scenario [4]. Conversely, a previous iVR study based on 360° video recordings found a less clear relationship between triage accuracy and expertise, as there were no significant differences in triage scores between residents and attendings or between those with and without prior MCI experience [35]. The medium rather than large effect size in the comparison between MFRs and non-MFRs may be because the non-MFR group were not true novices but also had some medical knowledge. In addition, participants were able to reread the START triage algorithms before starting the scenarios, possibly reducing the difference in triage accuracy performance.

In addition, this study suggests that speed can serve as an indicator to discriminate between MFRs and non-MFRs, as

XSL•FO
RenderX

MFRs completed triage faster with a medium effect size [71]. However, caution should be exercised when comparing this marker to real-life exercises, as previous studies suggest that triage in virtual MCI scenarios may occur faster than in real-life scenarios [4]. This difference is likely due to differences in locomotion and information retrieval. In this study, participants were able to teleport with their controllers, allowing them to quickly cover large distances. Walking, as a method of locomotion, would be closer to real life but would require significantly larger training areas. Future studies could compare MFRs and non-MFRs in iVR systems with natural locomotion and more realistic patient information retrieval to test whether the differences between the groups are similar or greater than in this study.

Information Transmission

The transmission of information to the control center is crucial for the effective coordination of rescue services and thus for the management of large-scale operations. As depicted in the SSSS scheme, the rapid and clear transmission of information regarding the MCI scene, safety hazards, patient information, and the need for additional support is essential for mobilizing necessary reinforcements [38]. To our knowledge, this study was the first to record radio messages and assess information transmission efficiency as a performance indicator differentiating between different levels of expertise in iVR MCI training. MFRs transmitted information more efficiently than non-MFRs (large effect size [71]), yet no correlation with the triage knowledge test score was observed. Furthermore, exploratory tests revealed that MFRs did not transmit significantly more accurate information but managed to do so using fewer words. These results suggest that a certain level of theoretical triage knowledge, that both groups seem to have had, is sufficient to facilitate MFR communication. However, the ability to quickly convey relevant information may improve with practical field experience.

The finding that MFRs transmit information more efficiently is consistent with a previous non-iVR study, which found that trained physicians were better at following a reporting scheme in a real-world simulation than their untrained counterparts [40]. However, the radio messages in the former study were rated by observers only in 3 categories: not performed, incomplete, and complete. In this study, the radio messages were analyzed in more detail. With the proposed method, trainees could receive more specific feedback in addition to their efficiency score, including which SSSS categories they missed or inaccurately reported. Furthermore, information transmission as a performance indicator holds promise for seamless integration into iVR, given the prevalent incorporation of microphones in iVR headsets and the increasing feasibility of automated evaluations through real-time transcription and artificial-intelligence applications. Future research may want to compare radio transmissions from iVR exercises with real-life exercises and actual MCI radio communications, providing further insight into differences in communication effectiveness and situational awareness, including the recognition and prioritization of hazards. In an Austrian study with 7 firefighters and 6 paramedics, participants underwent a 1-minute reporting phase following either an MCI scenario in iVR or nonimmersive VR on a computer screen [72]. In the former study, the information given during the reporting phase was analyzed regarding situational awareness, the anticipation of consequences, and the communication of actions. There was no significant difference between the 2 training modalities. This highlights the utility of verbal information reporting as a performance indicator not only within iVR contexts but also across various training modalities.

Nevertheless, future training should consider the locally used reporting schemes. In the United Kingdom, for instance, the METHANE (a mnemonic for “major incident declared; exact location; type of incident; hazards; access; number and type of casualties; emergency services present and required”) framework is commonly used as a guideline for reporting information [36]. While the SSSS scheme greatly overlaps with METHANE, certain aspects were not particularly explored in this study, such as the exact location and access possibilities. In this study, participants started at the accident site instead of arriving there using a virtual ambulance vehicle and also received no detailed information about the site beforehand. Such aspects must be considered in studies using the METHANE scheme.

Subjective Performance

Self-reported performance did not correlate with the applied objective performance indicators, which contradicts previous research suggesting that self-assessed performance validity is at least low to moderate in medical education [42,43]. This discrepancy may stem from the fact that individuals often base their perception of abilities on comparisons [73]. In terms of self-evaluation, people commonly compare themselves to others (social comparisons) or to their own past performance (temporal comparisons) [73]. However, in this study, the participants did not witness other trainees’ performances, limiting social comparison options. In addition, comparing one’s own progress in performance may have been complicated by the scarcity of MCI training, particularly because the non-MFR group lacked prior MCI training experience. It is also possible that asking about a specific behavior, such as self-rated information transmission performance, might have resulted in an estimate closer to the specific objective indicators. While we assessed global self-rated performance, future studies could test whether specific self-rated performance assessments correlate with objective indicators.

Overall, the findings underscore the significance of integrating objective indicators, providing trainees with benchmarks to assess their performance. Still, self-evaluations remain important because they help trainees reflect on their performance and enhance their self-awareness [44]. Furthermore, comparing objective and subjective performance could potentially mitigate biases, such as the tendency for women to underestimate themselves compared to men [43].

Limitations and Future Research

This study included a German sample of participants that were, on average, relatively young. Consequently, the generalizability of these findings to older populations and those from other countries should be tested in future research. For instance, previous iVR research found that cohorts over the age of 40

XSL•FO
RenderX

years had lower triage accuracy than younger cohorts [35]. This could potentially be attributed to greater familiarity with learning technologies.

This study was based on the START triage algorithm because it is well known internationally, and several triage algorithms are based on it [74]. However, in Germany, the learned triage algorithm varies depending on the rescue organization and region; thus, most of our sample was unfamiliar with START beforehand. Nevertheless, it is quite similar to the mSTART algorithm [74], which was the most frequently known algorithm in our sample. Future studies could program scenarios to initially allow selection of the learned algorithm, through which the scenario would then proceed, as in a previous study [75].

Furthermore, research on visual attention in prehospital settings is still in its infancy. Larger studies considering various attention and eye-tracking indicators, such as gaze movement patterns, would be beneficial in this regard. In this study, we investigated visual attention after arriving on scene and therefore only used the first 30 seconds. For standardization purposes, participants stayed at the starting position during this time. However, visual attention during the exploration of the scene or even during the whole scenario could also be an interesting indicator and more representative of behavior at real MCIs.

In this study, the ability to interact with patients was limited due to technical constraints of the software. In contrast to 360° videos, a major strength of this study was that participants were free to move around, which required them to decide how to navigate the environment, gain a full overview, and ensure that all patients were located. However, other than moving freely and using the controllers to open car doors, other actions, such as obtaining patient information, relied solely on visual and auditory cues as well as verbal information acquisition. Recent advances in iVR MCI software have introduced additional interaction features, including the use of a virtual toolkit with controllers, haptic feedback through vibrating controllers when taking a pulse, and patients responding to simple commands through automated speech recognition [15]. As the need for verbal requests to obtain patient information may have reduced immersion, future training may benefit from the use of iVR MCI tools with more interactive features than those available in this study.

In addition, the assessment of expertise should be examined in a more nuanced manner. We used the participants' professions as well as a triage knowledge test as expertise indicators to cover the expertise construct in a broad manner, including theoretical knowledge and practical experience. In contrast, previous studies often only used one indicator, such as years of experience [29] or profession [28]. Still, measuring expertise is challenging and likely depends on numerous factors, such as years of experience (full or part-time), exposure to MCIs, and the amount of MCI training. Future studies could delve deeper into identifying measures suitable for assessing expertise and perhaps propose a composite measure encompassing multiple factors.

Practical Implications

This study's results highlight the suitability of iVR scenarios for MCI training and performance assessment. With its resource-efficient implementation, iVR training offers a valuable opportunity to supplement current MCI training. Although nonimmersive VR on a computer monitor may be an even more affordable option, immersive training is associated with learning gains, increased enjoyment, and improved concentration [76,77]. While initial costs for hardware, software, and personnel occur, iVR becomes less expensive with repeated use compared to real-life exercises, which require ongoing organizational and financial resources [4]. In addition, iVR software can reduce costs associated with observers by automating performance assessments, facilitating self-directed training, and allowing trainers to focus on higher-level observations. Because previous research has shown that MCI performance declines over time, MCI training should be conducted more frequently than once a year to ensure high-quality triage [67]. In this context, iVR training and performance assessment could provide a valuable supplement to enable regular MCI training for large numbers of first responders. The integration of iVR into standard training curricula could begin with theoretical instruction [67], followed by regular iVR training, which could also be used to prepare MFRs for large-scale, real-life exercises.

In particular, triage accuracy, speed, and information transmission seem to be effective performance indicators for possible incorporation into iVR MCI training, complementing each other by providing insights into various aspects of overall performance. These indicators could be seamlessly integrated into future iVR learning programs to automate performance evaluation, thereby reducing the workload on trainers and allowing them to focus on higher-order evaluations, such as general procedures. The tested indicators are all suitable for individual MCI training but may also be applied in team training and could even be enhanced with novel team performance indicators, such as position and movement tracking [78]. Furthermore, the integration of real-time performance measurement into a (smart) scenario control could be valuable. On the basis of live data, the scenario difficulty could be dynamically adjusted by either the trainer or artificial intelligence to meet the individual needs of trainees.

Conclusions

Overall, iVR proved to be a valuable tool for assessing the performance of MFRs in MCIs scenarios. The performance indicators triage accuracy, triage speed, and information transmission can be extended to MCI training in iVR and capture multiple aspects of MCI performance. While visual attention did not function as a valid performance indicator in this study, future research might further explore visual attention as a potential indicator by examining other aspects, such as gaze patterns. Overall, iVR could be integrated into current MCI training curricula to provide objective and potentially automated performance assessments and allow for more frequent practice.

XSL•FO
RenderX

Acknowledgments

This project was part of the European Union's Horizon 2020 project Medical First Responder Training using a Mixed Reality Approach featuring haptic feedback for enhanced realism (MED1stMR; grant agreement 101021775). Special thanks to Samantha Ullrich for her support with the web-based questionnaires and for reviewing the R code file. The authors would also like to thank Guy Pires Cabrita for his support with the eye-tracking data and Paula Bleier and Julia Adam for their support with the radio messages.

Authors' Contributions

ASB, YH, MOF, and CW designed the study. ASB, YH, CW, BR, and MK planned the scenarios, which were constructed by BR, MK, and ASB. The scenarios and patient information were reviewed by GS, TB, and SM. SG supported the technological setup and developed software for standardized behavioral data collection and audio recordings with triage information. ASB and MB were responsible for data collection. ASB analyzed the data and wrote the manuscript. All authors reviewed and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptions of immersive virtual reality scenarios and performance per scenario.
[PDF File (Adobe PDF File), 498 KB - jmir_v27i1e63241_app1.pdf]

References

1. Lomaglio L, Ansaloni L, Catena F, Sartelli M, Coccolini F. Mass casualty incident: definitions and current reality. In: Kluger Y, Coccolini F, Catena F, Ansaloni L, editors. WSES Handbook of Mass Casualties Incidents Management. Cham, Switzerland: Springer; 2020.
2. Almukhlifi Y, Crowfoot G, Wilson A, Hutton A. Emergency healthcare workers' preparedness for disaster management: an integrative review. J Clin Nurs 2021:1-16. [doi: 10.1111/jocn.15965] [Medline: 34254375]
3. Berndt H, Wessel D, Willer L, Herczeg M, Mentler T. Immersion and presence in virtual reality training for mass casualty incidents. In: Proceedings of the 15th ISCRAM Conference. 2018 Presented at: ISCRAM 2015; May 20-23, 2018; Rochester, NY. [doi: 10.1109/vs-games.2018.8493427]
4. Mills B, Dykstra P, Hansen S, Miles A, Rankin T, Hopper L, et al. Virtual reality triage training can provide comparable simulation efficacy for paramedicine students compared to live simulation-based scenarios. Prehosp Emerg Care 2020;24(4):525-536. [doi: 10.1080/10903127.2019.1676345] [Medline: 31580178]
5. Baetzner AS, Wespi R, Hill Y, Gyllencreutz L, Sauter TC, Saveman BI, et al. Preparing medical first responders for crises: a systematic literature review of disaster training programs and their effectiveness. Scand J Trauma Resusc Emerg Med 2022;30(1):76 [FREE Full text] [doi: 10.1186/s13049-022-01056-8] [Medline: 36566227]
6. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst 1994;77(12):1321-1329.
7. Wrzus C, Frenkel MO, Schöne B. Current opportunities and challenges of immersive virtual reality for psychological research and application. Acta Psychol (Amst) 2024;249:104485 [FREE Full text] [doi: 10.1016/j.actpsy.2024.104485] [Medline: 39244850]
8. Gerwann S, Baetzner AS, Hill Y. Immersive virtual reality and augmented virtuality in sport and performance psychology: opportunities, current limitations, and practical recommendations. Sport Exerc Perform Psychol 2024:1-16. [doi: 10.1037/spy0000367]
9. Salas E, Rosen MA, Held JD, Weissmuller JJ. Performance measurement in simulation-based training: a review and best practices. Simul Gaming 2008;40(3):328-376. [doi: 10.1177/1046878108326734]
10. Markou-Pappas N, Lamine H, Ragazzoni L, Caviglia M. Key performance indicators in pre-hospital response to disasters and mass casualty incidents: a scoping review. Eur J Trauma Emerg Surg 2024;50(5):2029-2037. [doi: 10.1007/s00068-024-02533-8] [Medline: 38990353]
11. Tin D, Hertelendy AJ, Ciottone GR. Disaster medicine training: the case for virtual reality. Am J Emerg Med 2021;48:370-371 [FREE Full text] [doi: 10.1016/j.ajem.2021.01.085] [Medline: 33581938]
12. Bowyer MW, Streete KA, Muniz GM, Liu AV. Immersive virtual environments for medical training. Semin Colon Rectal Surg 2008;19(2):90-97. [doi: 10.1053/j.scrs.2008.02.005]
13. Heldring S, Jirwe M, Wihlborg J, Berg L, Lindström V. Using high-fidelity virtual reality for mass-casualty incident training by first responders – a systematic review of the literature. Prehosp Disaster Med 2024;39(1):94-105. [doi: 10.1017/s1049023x24000049]
14. Hill Y, Kiefer AW, Oudejans RR, Baetzner AS, Den Hartigh RJ. Adaptation to stressors: hormesis as a framework for human performance. New Ideas Psychol 2024;73:101073. [doi: 10.1016/j.newideapsych.2024.101073]

XSL•FO
RenderX

15. Kman NE, Price A, Berezina-Blackburn V, Patterson J, Maicher K, Way DP, et al. First responder virtual reality simulator to train and assess emergency personnel for mass casualty response. J Am Coll Emerg Physicians Open 2023;4(1):e12903 [FREE Full text] [doi: 10.1002/emp2.12903] [Medline: 36817080]
16. Bischof WF, Anderson NC, Kingstone A. A tutorial: analyzing eye and head movements in virtual reality. Behav Res Methods 2024;56(8):8396-8421. [doi: 10.3758/s13428-024-02482-5] [Medline: 39117987]
17. Shinnick MA. Validating eye tracking as an objective assessment tool in simulation. Clin Simul Nurs 2016;12(10):438-446. [doi: 10.1016/j.ecns.2016.06.001]
18. Ziegler M, Bühner M. Grundlagen der Psychologischen Diagnostik. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften; 2012.
19. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52(4):281-302. [doi: 10.1037/h0040957]
20. Ullén F, Hambrick DZ, Mosing MA. Rethinking expertise: a multifactorial gene-environment interaction model of expert performance. Psychol Bull 2016;142(4):427-446. [doi: 10.1037/bul0000033] [Medline: 26689084]
21. Blondon K, Wipfli R, Lovis C. Use of eye-tracking technology in clinical reasoning: a systematic review. Stud Health Technol Inform 2015;210:90-94. [Medline: 25991108]
22. Schmidt HG, Boshuizen HP. On acquiring expertise in medicine. Educ Psychol Rev 1993;5(3):205-221. [doi: 10.1007/bf01323044]
23. Just MA, Carpenter PA. A theory of reading: from eye fixations to comprehension. Psychol Rev 1980;87(4):329-354. [doi: 10.1037//0033-295x.87.4.329]
24. Al-Moteri MO, Symmons M, Plummer V, Cooper S. Eye tracking to investigate cue processing in medical decision-making: a scoping review. Comput Hum Behav 2017;66:52-66. [doi: 10.1016/j.chb.2016.09.022]
25. Brunyé TT, Drew T, Weaver DL, Elmore JG. A review of eye tracking for understanding and improving diagnostic interpretation. Cogn Res Princ Implic 2019;4(1):7 [FREE Full text] [doi: 10.1186/s41235-019-0159-2] [Medline: 30796618]
26. Currie J, Bond RR, McCullagh P, Black P, Finlay DD, Gallagher S, et al. Wearable technology-based metrics for predicting operator performance during cardiac catheterisation. Int J Comput Assist Radiol Surg 2019;14(4):645-657 [FREE Full text] [doi: 10.1007/s11548-019-01918-0] [Medline: 30730031]
27. Gegenfurtner A, Lehtinen E, Säljö R. Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. Educ Psychol Rev 2011;23(4):523-552. [doi: 10.1007/s10648-011-9174-7]
28. Prytz EG, Norén C, Jonson CO. Fixation differences in visual search of accident scenes by novices and expert emergency responders. Hum Factors 2018;60(8):1219-1227 [FREE Full text] [doi: 10.1177/0018720818788142] [Medline: 30102566]
29. Ferrandini Price M, Manzanares Serrano A, Fernández-Pacheco AN, Melendreras Ruiz R, García Collado AJ, Pardo Ríos M. Lack of a standard visual attention pattern for use when health professionals triage mass casualty incidents is revealed by Tobbi eye tracking technology. Emergencias 2022;34(3):233-235 [FREE Full text] [Medline: 35736532]
30. Loth S, Cote AC, Shaafi Kabiri N, Bhangu JS, Zumwalt A, Moss M, et al. Improving triage accuracy in first responders: measurement of short structured protocols to improve identification of salient triage features. World Med Health Policy 2019;11(2):163-176. [doi: 10.1002/wmh3.306]
31. Super G, Groth S, Hook R. START: simple triage and rapid treatment plan. Hoag Memorial Presbyterian Hospital. 1994. URL: https://tinyurl.com/mr2j3pdw [accessed 2024-05-30]
32. Bazyar J, Farrokhi M, Salari A, Khankeh HR. The principles of triage in emergencies and disasters: a systematic review. Prehosp Disaster Med 2020;35(3):305-313. [doi: 10.1017/s1049023x20000291]
33. Vincent DS, Sherstyuk A, Burgess L, Connolly KK. Teaching mass casualty triage skills using immersive three-dimensional virtual reality. Acad Emerg Med 2008;15(11):1160-1165 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00191.x] [Medline: 18699829]
34. Ferrandini Price M, Escribano Tortosa D, Nieto Fernandez-Pacheco A, Perez Alonso N, Cerón Madrigal JJ, Melendreras-Ruiz R, et al. Comparative study of a simulated incident with multiple victims and immersive virtual reality. Nurse Educ Today 2018;71:48-53. [doi: 10.1016/j.nedt.2018.09.006] [Medline: 30241022]
35. Lowe J, Peng C, Winstead-Derlega C, Curtis H. 360 virtual reality pediatric mass casualty incident: a cross sectional observational study of triage and out-of-hospital intervention accuracy at a national conference. J Am Coll Emerg Physicians Open 2020;1(5):974-980 [FREE Full text] [doi: 10.1002/emp2.12214] [Medline: 33145548]
36. Lowes A, Cosgrove JF. Prehospital organization and management of a mass casualty incident. BJA Educ 2016;16(10):323-328. [doi: 10.1093/bjaed/mkw005]
37. Perry O, Jaffe E, Bitan Y. Dynamic communication quantification model for measuring information management during mass-casualty incident simulations. Hum Factors 2022;64(1):228-249 [FREE Full text] [doi: 10.1177/00187208211018880] [Medline: 34275344]
38. Schmid B, Sauer F, Busch HJ. Initial preclinical assessment on-site. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022;65(10):979-986 [FREE Full text] [doi: 10.1007/s00103-022-03582-3] [Medline: 36125539]
39. Zechner O, Schrom-Feiertag H, Wespi R, Pretolesi D, Nguyen Q, Tscheligi M. Enhancing mixed reality simulation training technology with real-time performance visualization: mixed methods study with medical first responders. JMIR XR Spatial Comput 2024;1:e57655 [FREE Full text] [doi: 10.2196/57655]

XSL•FO
RenderX

40. Ingrassia PL, Colombo D, Barra FL, Carenzo L, Corte FD. Impact of training in medical disaster management: a pilot study using a new tool for live simulation. Emergencias 2013;25(6):459-466 [FREE Full text]
41. Soola AH, Mehri S, Azizpour I. Evaluation of the factors affecting triage decision-making among emergency department nurses and emergency medical technicians in Iran: a study based on Benner's theory. BMC Emerg Med 2022;22(1):174 [FREE Full text] [doi: 10.1186/s12873-022-00729-y] [Medline: 36303127]
42. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991;66(12):762-769. [doi: 10.1097/00001888-199112000-00012] [Medline: 1750956]
43. Blanch-Hartigan D. Medical students' self-assessment of performance: results from three meta-analyses. Patient Educ Couns 2011;84(1):3-9. [doi: 10.1016/j.pec.2010.06.037] [Medline: 20708898]
44. Gow KW. Self-evaluation: how well do surgery residents judge performance on a rotation? Am J Surg 2013;205(5):557-562. [doi: 10.1016/j.amjsurg.2013.01.010] [Medline: 23499389]
45. Baetzner AS, Hill Y, Frenkel MO, Wrzus C, Gerwann S. Med1stMR VR study - supplementary study material. OSF Home. 2023. URL: https://doi.org/10.17605/OSF.IO/VHWTP [accessed 2025-01-17]
46. Cuttance G, Dansie K, Rayner T. Paramedic application of a triage sieve: a paper-based exercise. Prehosp Disaster Med 2017;32(1):3-13. [doi: 10.1017/S1049023X16001163] [Medline: 27964769]
47. iMotions homepage. iMotions. URL: https://imotions.com/ [accessed 2024-11-01]
48. Nguyen TH, El-Nasr MS, Isaacowitz DM. Interactive visualization for understanding of attention patterns. In: Proceedings of the Eye Tracking and Visualization. 2015 Presented at: ETVIS 2015; October 25, 2015; Chicago, IL. [doi: 10.1007/978-3-319-47024-5_2]
49. Clay V, König P, König S. Eye tracking in virtual reality. J Eye Mov Res 2019;12(1):1-18 [FREE Full text] [doi: 10.16910/jemr.12.1.3] [Medline: 33828721]
50. Farnsworth B. Eye tracking: the complete pocket guide. iMotions. 2022. URL: https://imotions.com/blog/learning/best-practice/eye-tracking/ [accessed 2024-10-04]
51. Dresing T, Pehl T. Praxisbuch interview, transkription and analyse. Anleitung und Regelsysteme für qualitativ Forschende. 2015. URL: https://d-nb.info/1077320221/34 [accessed 2023-06-01]
52. Mayring P. Qualitative content analysis: theoretical background and procedures. In: Bikner-Ahsbahs A, Knipping C, Presmeg N, editors. Approaches to Qualitative Research in Mathematics Education. Dordrecht, The Netherlands: Springer; 2015.
53. Hutchins SG, Timmons RP. Radio interoperability: addressing the real reason we don’t communicate well during emergencies. Calhoun: The NPS Institutional Archive. 2006. URL: http://www.dodccrp.org/events/12th_ICCRTS/CD/html/presentations/221.pdf [accessed 2024-10-01]
54. Finch H. Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. Methodology 2005;1(1):27-38. [doi: 10.1027/1614-1881.1.1.27]
55. Rasch D, Guiard V. The robustness of parametric statistical methods. Psychol Sci 2004;46:175-208 [FREE Full text]
56. Pagano R. Understanding Statistics in the Behavioral Sciences Tenth Edition. Belmont, CA: Wadsworth Publishing Company; 2012.
57. Felnhofer A, Kothgassner OD, Beutl L, Hlavacs H, Kryspin-Exner I. Is virtual reality made for men only? Exploring gender differences in the sense of presence. In: Proceedings of the Annual Conference of the International Society on Presence Research. 2012 Presented at: ISPR 2012; October 24-26, 2012; Philadelphia, PA. [doi: 10.1016/j.chb.2013.10.045]
58. Parra MA, Kaplan RI. Predictors of performance in real and virtual scenarios across age. Exp Aging Res 2019;45(2):180-198. [doi: 10.1080/0361073x.2019.1586106]
59. Ratcliff R. Methods for dealing with reaction time outliers. Psychol Bull 1993;114(3):510-532. [doi: 10.1037/0033-2909.114.3.510] [Medline: 8272468]
60. Baetzner AS, Wrzus C, Hill Y, Frenkel MO. Med1st XVR RQ2: level of expertise of medical first-responders, attention, decision-making, and triage behavior during MCI training in iVR. OSF Registries. 2023 Feb 21. URL: https://osf.io/amgk8 [accessed 2024-05-01]
61. Lang C, Nguyen TV, Katti H, Yadati K, Kankanhalli M, Yan S. Depth matters: influence of depth cues on visual saliency. In: Proceedings of the 12th European Conference on Computer Vision. 2012 Presented at: ECCV 2012; October 7-13, 2012; Florence, Italy. [doi: 10.1007/978-3-642-33709-3_8]
62. Li G, Anguera JA, Javed SV, Khan MA, Wang G, Gazzaley A. Enhanced attention using head-mounted virtual reality. J Cogn Neurosci 2020;32(8):1438-1454 [FREE Full text] [doi: 10.1162/jocn_a_01560] [Medline: 32286132]
63. Murray NP, Lewinski W, Sandri Heidner G, Lawton J, Horn R. Gaze control and tactical decision-making under stress in active-duty police officers during a live use-of-force response. J Mot Behav 2024;56(1):30-41. [doi: 10.1080/00222895.2023.2229946] [Medline: 37385608]
64. Borys M, Plechawska-Wójcik M. Eye-tracking metrics in perception and visual attention research. Eur J Med Technol 2017;3(16):11-23 [FREE Full text]
65. Cicero MX, Auerbach MA, Zigmont J, Riera A, Ching K, Baum CR. Simulation training with structured debriefing improves residents' pediatric disaster triage performance. Prehosp Disaster Med 2012;27(3):239-244. [doi: 10.1017/S1049023X12000775] [Medline: 22691308]

XSL•FO
RenderX

66. Cicero MX, Whitfill T, Overly F, Baird J, Walsh B, Yarzebski J, et al. Pediatric disaster triage: multiple simulation curriculum improves prehospital care providers' assessment skills. Prehosp Emerg Care 2017;21(2):201-208. [doi: 10.1080/10903127.2016.1235239] [Medline: 27749145]
67. Dittmar MS, Wolf P, Bigalke M, Graf BM, Birkholz T. Primary mass casualty incident triage: evidence for the benefit of yearly brief re-training from a simulation study. Scand J Trauma Resusc Emerg Med 2018;26(1):35 [FREE Full text] [doi: 10.1186/s13049-018-0501-6] [Medline: 29703219]
68. Yanagawa Y, Omori K, Ishikawa K, Takeuchi I, Jitsuiki K, Yoshizawa T, et al. Difference in first aid activity during mass casualty training based on having taken an educational course. Disaster Med Public Health Prep 2018;12(4):437-440. [doi: 10.1017/dmp.2017.99] [Medline: 29151402]
69. Knight JF, Carley S, Tregunna B, Jarvis S, Smithies R, de Freitas S, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. Resuscitation 2010;81(9):1175-1179. [doi: 10.1016/j.resuscitation.2010.03.042] [Medline: 20732609]
70. Luigi Ingrassia P, Ragazzoni L, Carenzo L, Colombo D, Ripoll Gallardo A, Della Corte F. Virtual reality and live simulation: a comparison between two simulation tools for assessing mass casualty triage skills. Eur J Emerg Med 2015;22(2):121-127. [doi: 10.1097/MEJ.0000000000000132] [Medline: 24841770]
71. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New Jersey, USA: Lawrence Erlbaum Associates; 1988.
72. Paletta L, Pszeida M, Schneeberger M, Dini A, Reim L, Kallus KW. Cognitive-emotional stress and risk stratification of situational awareness in immersive first responder training. In: Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics. 2022 Presented at: BHI 2022; September 27-30, 2022; Ioannina, Greece. [doi: 10.1109/bhi56158.2022.9926805]
73. Möller J, Marsh HW. Dimensional comparison theory. Psychol Rev 2013;120(3):544-560. [doi: 10.1037/a0032459] [Medline: 23544443]
74. Bazyar J, Farrokhi M, Khankeh H. Triage systems in mass casualty incidents and disasters: a review study with a worldwide approach. Open Access Maced J Med Sci 2019;7(3):482-494 [FREE Full text] [doi: 10.3889/oamjms.2019.119] [Medline: 30834023]
75. Berndt H, Wessel D, Mentler T, Herczeg M. Human-centered design of a virtual reality training simulation for mass casualty incidents. In: Proceedings of the 10th International Conference on Virtual Worlds and Games for Serious Applications. 2018 Presented at: VS-Games 2018; September 5-7, 2018; Würzburg, Germany. [doi: 10.1109/vs-games.2018.8493427]
76. Mahmoud K, Harris I, Yassin H, Hurkxkens TJ, Matar OK, Bhatia N, et al. Does immersive VR increase learning gain when compared to a non-immersive VR learning experience? In: Proceedings of the 7th International Conference on Learning and Collaboration Technologies. Human and Technology Ecosystems. 2020 Presented at: LCT 2020; July 19-24, 2020; Copenhagen, Denmark. [doi: 10.1007/978-3-030-50506-6_33]
77. Gutiérrez F, Pierce J, Vergara VM, Coulter R, Saland L, Caudell TP, et al. The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. Stud Health Technol Inform 2007;125:155-160. [Medline: 17377256]
78. Wespi R, Birrenbach T, Schauber SK, Manser T, Sauter TC, Kämmer JE. Exploring objective measures for assessing team performance in healthcare: an interview study. Front Psychol 2023;14:1232628 [FREE Full text] [doi: 10.3389/fpsyg.2023.1232628] [Medline: 37941756]

Abbreviations

AOI: area of interest
HMD: head-mounted display
iVR: immersive virtual reality
MCI: mass casualty incident
MFR: medical first responder
mSTART: modified simple triage and rapid treatment
PRIOR: Primäres Ranking zur Initialen Orientierung im Rettungsdienst; Primary Ranking for Initial Orientation in Emergency Medical Services
SSSS: scene, safety, situation, and support
START: simple triage and rapid treatment

XSL•FO
RenderX

Edited by T Loetscher; submitted 14.06.24; peer-reviewed by S Skec, T Mühling, H Farhat; comments to author 21.08.24; revised version received 14.10.24; accepted 22.11.24; published 27.01.25.

Please cite as:
Baetzner AS, Hill Y, Roszipal B, Gerwann S, Beutel M, Birrenbach T, Karlseder M, Mohr S, Salg GA, Schrom-Feiertag H, Frenkel MO, Wrzus C
Mass Casualty Incident Training in Immersive Virtual Reality: Quasi-Experimental Evaluation of Multimethod Performance Indicators
J Med Internet Res 2025;27:e63241
URL: https://www.jmir.org/2025/1/e63241
doi:10.2196/63241
PMID:

©Anke Sabine Baetzner, Yannick Hill, Benjamin Roszipal, Solène Gerwann, Matthias Beutel, Tanja Birrenbach, Markus Karlseder, Stefan Mohr, Gabriel Alexander Salg, Helmut Schrom-Feiertag, Marie Ottilie Frenkel, Cornelia Wrzus. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Use of Augmented Reality for Training Assistance in Laparoscopic Surgery: Scoping Literature Review

Francisco Javier Celdrán[1*], BD; Javier Jiménez-Ruescas[1*], MD; Carlos Lobato[2*], BD; Lucía Salazar[2*], BD; Juan Alberto Sánchez-Margallo[2*], PhD; Francisco M Sánchez-Margallo[3*], PhD; Pascual González[1,4*], Prof Dr

[1]I3A, LoUISE Research Group, University of Castilla-La Mancha, Albacete, Spain

[2]Bioengineering and Health Technologies Unit, Jesús Usón Minimally Invasive Surgery Centre, Cáceres, Spain

[3]Scientific Direction, Jesús Usón Minimally Invasive Surgery Centre, Cáceres, Spain

[4]Computing Systems Department, University of Castilla-La Mancha, Albacete, Spain

[*]all authors contributed equally

Corresponding Author:
Francisco Javier Celdrán, BD
I3A
LoUISE Research Group
University of Castilla-La Mancha
C/ Investigación 2
Albacete, 02071
Spain
Phone: 34 666649104
Email: franciscoj.celdran@uclm.es

Abstract

Background: Laparoscopic surgery training is a demanding process requiring technical and nontechnical skills. Surgical training has evolved from traditional approaches to the use of immersive digital technologies such as virtual, augmented, and mixed reality. These technologies are now integral to laparoscopic surgery training.

Objective: This scoping literature review aimed to analyze the current augmented reality (AR) solutions used in laparoscopic surgery training.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, we conducted a scoping review using 4 databases: Scopus, IEEE Xplore, PubMed, and ACM. Inclusion and exclusion criteria were applied to select relevant articles. Exclusion criteria were studies not using AR, not focused on laparoscopic surgery, not focused on training, written in a language other than English, or not providing relevant information on the topics studied. After selecting the articles, research questions (RQs) were formulated to guide the review. In total, 2 independent reviewers then extracted relevant data, and a descriptive analysis of the results was conducted.

Results: Of 246 initial records, 172 (69.9%) remained after removing duplicates. After applying the exclusion criteria, 76 articles were selected, with 25 (33%) later excluded for not meeting quality standards, leaving 51 (67%) in the final review. Among the devices analyzed (RQ 1), AR video–based devices were the most prevalent (43/51, 84%). The most common information provided by AR devices (RQ 1) focused on task execution and patient-related data, both appearing in 20% (10/51) of studies. Regarding sensorization (RQ 2), most studies (46/51, 90%) incorporated some form of sensorized environment, with computer vision being the most used technology (21/46, 46%) and the trainee the most frequently sensorized element (41/51, 80%). Regarding training setups (RQ 3), 39% (20/51) of the studies used commercial simulators, and 51% (26/51) made use of artificial models. Concerning the evaluation methods (RQ 4), objective evaluation was the most used, featured in 71% (36/51) of the studies. Regarding tasks (RQ 5), 43% (22/51) of studies focused on full surgical procedures, whereas 57% (29/51) focused on simple training tasks, with suturing being the most common among the latter (11/29, 38%).

Conclusions: This scoping review highlights the evolving role of AR technologies in laparoscopic surgery training, although the impact of optical see-through devices remains unclear due to their limited use. It underscores the potential of emerging technologies such as haptic feedback, computer vision, and eye tracking to further enhance laparoscopic skill acquisition. While most relevant articles from other databases were included, some studies may have been missed due to the specific databases and

XSL•FO
RenderX

search strategies used. Moreover, the need for standardized evaluation metrics is emphasized, paving the way for future research into AR's full potential in laparoscopic skill acquisition.

(J Med Internet Res 2025;27:e58108) doi:10.2196/58108

KEYWORDS

laparoscopic surgery; surgical training; surgical simulator; augmented reality–based laparoscopic simulator; AR-based laparoscopic simulator; augmented reality; mobile phone

Introduction

Background

Training in laparoscopic surgery is a long and demanding process that requires extensive theoretical knowledge, along with technical and nontechnical skills. Learning models in surgical training have rapidly evolved from traditional approaches based on the educational philosophy of "*see one, do one, teach one*" to more sophisticated surgical simulators that aim to increase the number of training sessions, thus dramatically enhancing the skills of medical professionals and the safety of patients [1]. Among the formative strategies, there are some based on animal models [2] and cadavers [3]. However, due to the economic and ethical issues involved in some of these solutions, surgical training has rapidly shifted toward the use of simulation-based systems, mainly for the early formative stages [4].

The emergence of immersive digital technologies such as virtual reality, augmented reality (AR), and mixed reality (MR) has led to a paradigm shift in the field of surgical training. Virtual reality allows users to be immersed in a fully digital environment replacing the physical world, whereas AR superimposes virtual elements onto the real world, enhancing or augmenting the user's environment. The latest evolution of these immersive technologies is MR, which merges virtual and real objects, enabling realistic interactions and coexistence. The use of these technologies is becoming an important part of the training process in laparoscopic surgery, enhancing the training experience and content without putting the patient at risk [5,6]. This technique can generate customized 3D models of each patient on which to train. Similarly, during the training process, this technology allows for the inclusion of enriched information, such as holographic images or 3D objects to guide the user during the training process and facilitating a more precise alignment between virtual information and physical objects in a simulator, an experimental model, or a patient. The use of AR technologies is constantly evolving and being integrated into the field of minimally invasive surgery. However, the current level of development of this technology as a tool to assist in the training process in laparoscopic surgery, as well as the available solutions (both commercial and prototypes) and the training functionalities they offer, is not precisely known.

Objectives

This scoping review intended to analyze the current AR and MR solutions used to assist in laparoscopic surgery training. Similarly, we reviewed the types of surgical simulators that make use of this technology, the training assistance information offered, and the main training tasks and procedures in which they have been used. This gave us a more detailed view of the current state of AR in the field of laparoscopic surgery training.

Methods

Search Strategy

A structured bibliographical search was conducted in the Scopus, IEEE Xplore, PubMed, and ACM databases. The same search query adapted to each database query syntax was used. We used a set of keywords related to the topic of this scoping review to identify relevant studies published up to January 1, 2024 (Multimedia Appendix 1). In general terms, the initial search encompassed all articles that included the following terms in the title or abstract or as keywords: "laparoscopic"; AND "augmented reality" OR "mixed reality" OR "extended reality"; AND "training" OR "practice."

Eligibility Criteria

A series of inclusion and exclusion criteria were considered to select the articles that best applied for our objectives. In general, articles on studies that did not make use of AR technologies or did not provide information on the inclusion of such technologies in laparoscopic surgery training, articles that did not focus on laparoscopy or its skill training, articles that were written in a language other than English, reviews, or articles that were not accessible were discarded. In addition, regarding the quality of the articles, those studies that did not provide relevant information on the topics studied were excluded.

Selection of Articles

Results were screened by 2 independent reviewers. In case of discrepancies between reviewers regarding the inclusion or noninclusion of an article, a third reviewer was consulted. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines were followed to carry out this scoping review [7], including the completion of the corresponding checklist (Multimedia Appendix 2).

Research Questions

The aim of this work was to analyze the studies published in scientific literature in relation to the use of AR in laparoscopic surgery training and analyze the relevant aspects that facilitate the development of new investigations in this field. Therefore, with this scoping review, we intended to answer the following research questions (RQs):

1. What type of devices and feedback are used for AR-based laparoscopic surgery training? (RQ 1)
2. What type of sensorization is used for AR-based laparoscopic surgery training? (RQ 2)

XSL•FO
RenderX

3. What type of simulator and setup is used for AR-based laparoscopic surgery training? (RQ 3)
4. What type of evaluation is used to assess skill acquisition in AR-based laparoscopic surgery training? (RQ 4)
5. What type of surgical tasks or procedures are used in AR-based laparoscopic surgery training? (RQ 5)

To do so, we classified every article based on different dimensions, each one related to one of the RQs (Figure 1).

Figure 1. Taxonomy of articles regarding each research question (RQ). AR: augmented reality; ARV: augmented reality video; FLS: Fundamentals of Laparoscopic Surgery; GOALS: Global Operative Assessment of Laparoscopic Skills; OSATS: Objective Structured Assessment of Technical Skills; OST: optical see-through.

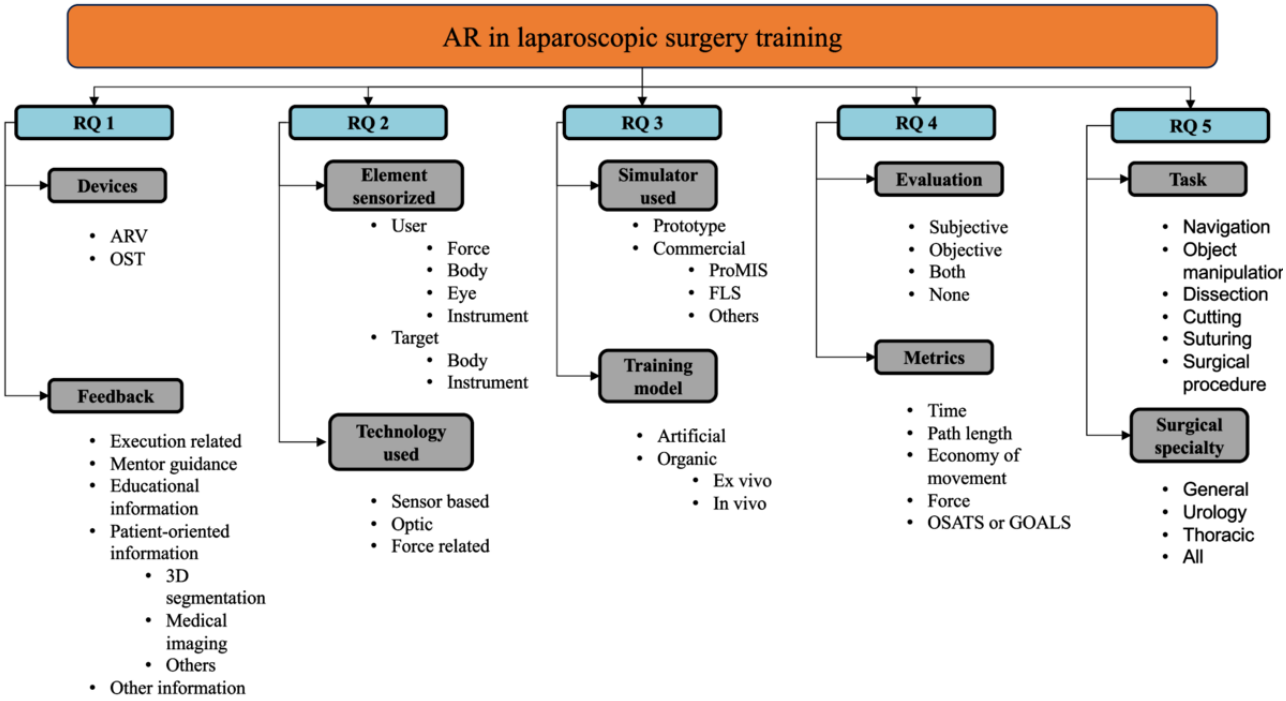

Data Collection and Processing

For each selected study, information on the following aspects was recorded: (1) year of publication, (2) summary of the complete study, (3) modality of teaching (tele-mentoring or conventional), (4) type of device used for AR, (5) type of information provided to the learner, (6) type of sensorization, (7) type of AR simulator (commercial or prototype, and which one if commercial), (8) type of evaluation (objective or subjective), (9) training tasks or procedures performed, and (10) specialty of laparoscopic surgery.

As with the selection of articles, the data collected for each article were screened by 2 independent reviewers. In case of discrepancies between reviewers regarding the data collected for each article, a third reviewer was consulted.

Results

Overview

A total of 246 records were obtained (n=28, 11.4% in IEEE Explore; n=121, 49.2% in Scopus; n=16, 6.5% in ACM; and n=81, 32.9% in PubMed). After removing duplicate records, 69.9% (172/246) remained. Once the exclusion criteria were applied, 44.2% (76/172) of the articles were left. Of these 76 articles, 25 (33%) that did not meet the quality criteria were eliminated so that 51 (67%) articles were finally included in this review Figure 2 illustrates the complete workflow.

The papers included were published between 2002 and 2023, most of them (37/51, 73%) were scientific journal articles [6,8-43], and the rest (14/51, 27%) were papers presented in scientific conferences [44-57].

The evolution of the papers included in this scoping review was heterogeneous (Table 1), although it seems to be related to the launch of new AR devices on the market. Among them, the appearance of the first AR applications in games and smartphones in 2009, the launch of the Google Glass device in 2015, or the launch of the second version of Microsoft's HoloLens in 2019 are highlighted.

In the following sections, we will analyze each of the RQs defined to provide a comprehensive analysis of the current application of AR technology in laparoscopic surgery training.

XSL•FO
RenderX

Figure 2. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram showing the manuscript selection process. AR: augmented reality.

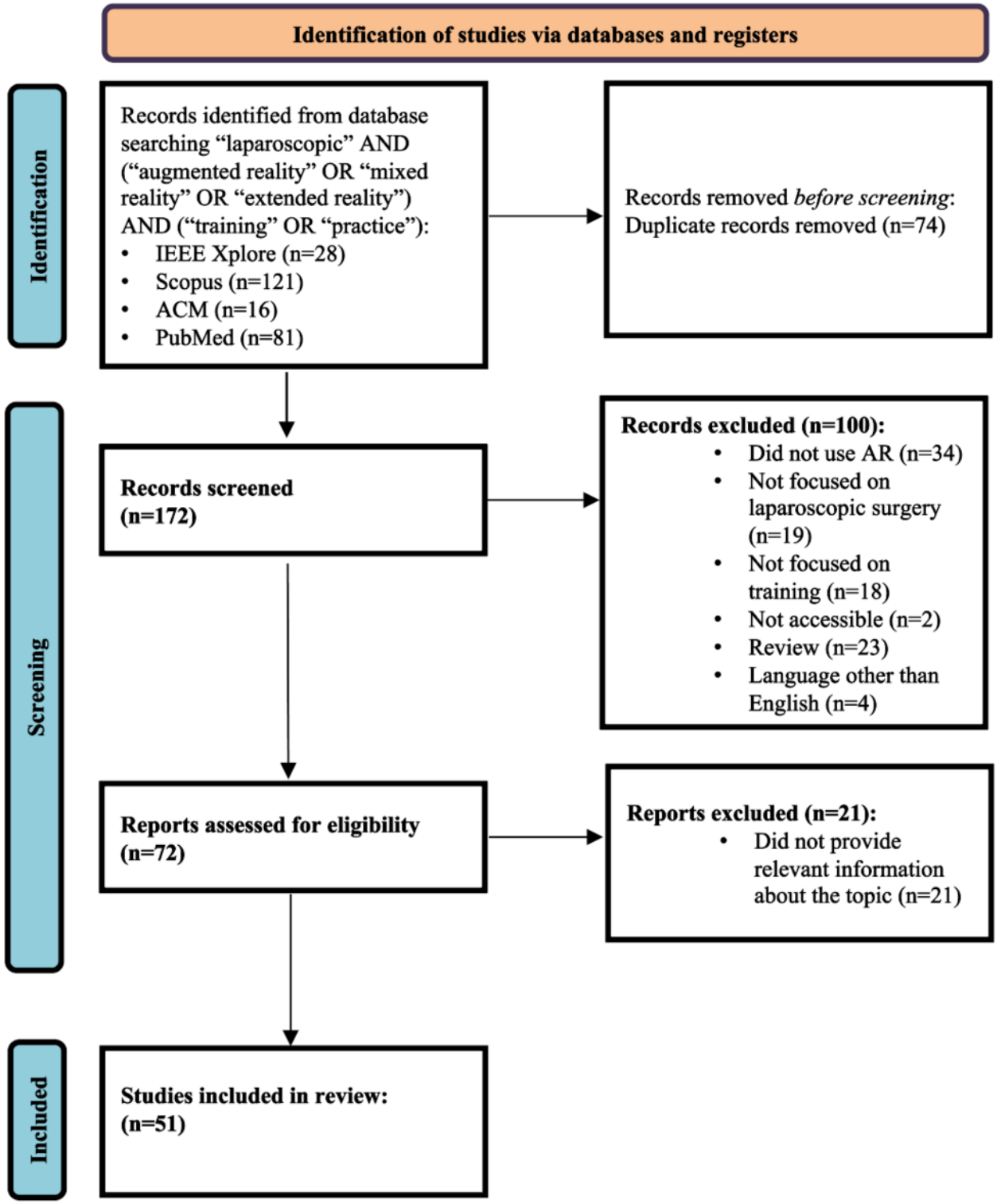

XSL•FO
RenderX

Table 1. Article distribution by year (N=51).

Year	Articles, n (%)
2002	1 (2)
2005	1 (2)
2006	1 (2)
2007	1 (2)
2008	2 (4)
2009	3 (6)
2010	4 (8)
2011	3 (6)
2012	3 (6)
2013	3 (6)
2014	6 (12)
2015	1 (2)
2016	2 (4)
2017	2 (4)
2018	1 (2)
2019	2 (4)
2020	3 (6)
2021	4 (8)
2022	4 (8)
2023	4 (8)

RQ 1: What Types of Devices and Feedback Are Used for AR-Based Laparoscopic Surgery Training?

Our first RQ was about the AR devices used and the information provided. This analysis allowed us to see the evolution of the technology used and how these systems provide relevant information to enhance laparoscopic surgery training.

Regarding the devices, we classified them into 2 main types: optical see-through (OST) and AR video (ARV)–based devices (Table 2). OST refers to those devices that are capable of rendering information on a medium while allowing the real world to be seen through it. Devices such as HoloLens fall into this category as they allow information to be rendered on the screen while superimposing this information on our visual perception of the real world. ARV refers to the use of conventional video devices such as monitors, tablets, or other types of screens in which reality information is displayed (such as a real-time video) and overlaid with expanded information, creating an AR system. OST systems usually include voice commands, eye tracking, and interaction with real-world objects, whereas ARV systems reduce interaction and rely mainly on the display of additional information on the screen. Some articles (2/51, 4%) did not specify the device used, and so they were classified as unspecified devices.

The feedback provided by the system is closely related to the devices being used as it conditions the user-system interaction. We divided the articles by the information given to the user, differentiating among (1) *execution-related* information, which refers to any type of information related to the actions being performed by the trainee, such as the force exerted or time expended; (2) *mentor guidance*, which refers to those systems in which the AR technology allows the mentor to enhance the way in which they provide guidance to the trainee, for instance, using virtual pointers or annotations; (3) *educational information*, which could be general information about the training activity being performed, such as images of the organs and structures involved in the surgical activity or other general information; and, finally, (4) *patient-oriented* information, which corresponds to personalized information about the patient to be operated on. In this case, information was divided into *medical imaging* (as preoperative imaging), *3D segmentation* created from preoperative imaging, or *other* personalized information. *Other information* comprised information that did not fit into any other classification. *Unspecified information* comprised articles that did not indicate the type of information provided to the user of the AR device (Table 2).

XSL•FO
RenderX

Table 2. Type of information provided to the user of augmented reality–based laparoscopic training systems.

	Execution related	Mentor guidance	Educational information	Patient-oriented information			Other information	Unspecified information
				3D segmentation	Medical imaging	Others		
ARV[a]	• Zahiri et al [14] • Nomura et al [15] • Horeman et al [21,26] • Botden et al [34,35] • Preetha et al [44] • Lahanas et al [45] • Rozenblit et al [46]	• Wild et al [8] • Andersen et al [16] • Vera et al [19] • Felinska et al [41] • Cizmic et al [42] • Shabir et al [47]	• Arts et al [12]	• Pessaux et al [20] • Condino et al [48] • Fusaglia et al [49]	• Viglialoro et al [13] • Teber et al [33] • Koehl et al [50]	• Shao et al [43] • Viglialoro et al [51]	• Hughes-Hallett et al [17] • Lahanas et al [18] • Loukas et al [22] • Baumhauer et al [38] • Kim et al [39] • Ivanova et al [52] • Wagner and Rozenblit [53] • Larrarte and Alban [54]	• Nugent et al [23] • Brinkman et al [24] • Pagador et al [25] • Strickland et al [28] • Leblanc et al [29-32] • Botden et al [36,37] • Van Sickle et al [40]
OST[b]	• Cau et al [9]	—[c]	• Simone et al [10] • Doughty et al [55]	• Sánchez-Margallo et al [6]	• Sánchez-Margallo et al [6] • Simone et al [10] • Zorzal et al [11]	—	• Rewkowski et al [56]	—
Unspecified device	—	—	—	—	—	—	—	• Pagador et al [27] • Gupta et al [57]

[a]ARV: augmented reality video.

[b]OST: optical see-through.

[c]Combinations of devices and information types under which no articles were categorized, indicating nonapplicability in the reviewed studies.

Most of the studies analyzed (43/51, 84%) used ARV devices, and only 12% (6/51) used OST devices. A total of 4% (2/51) of the articles did not specify which kind of device was used.

Regarding the ARV studies, although most of them (11/43, 26%) did not specify what information they provided to the user, among those that specified this information, execution-related information was highly common (9/43, 21%), such as in the study by Zahiri et al [14], which showed the time remaining for the trainee to complete the current task in the form of either a numerical timer or a progress bar. Regarding mentor guidance (6/43, 14%), Andersen et al [16] developed a tool for tele-mentoring, allowing the mentor to use different-colored dots for annotation. A common type of information was patient-oriented information (8/43, 19%). Pessaux et al [20] conducted a 3D segmentation from preoperative images to generate a 3D model of the patient's body, which was later superimposed over the real patient's body during a duodenopancreatectomy. Regarding medical imaging, Koehl et al [50] developed a system that allowed trainees to visualize and manipulate preoperative images. Arts et al [12] provided trainees with educational videos with instructions for the ongoing task. Regarding other information, Shao et al [43] used prerecorded ultrasound videos that were played depending on the position of the laparoscopic tools so that their system could be used for training purposes on ultrasound-guided laparoscopic interventions.

Of the 6 studies using OST devices, 3 (50%) used HoloLens version 1 [10,11,56], and 3 (50%) used HoloLens version 2 [6,9,55]. In addition, the year of publication of those articles was 2020, so it could be considered a technology of recent application in laparoscopic training. Although far fewer studies that used OST devices were found, they mainly provided patient-oriented information (4/6, 67%). For instance, Zorzal et al [11] gave users the possibility to obtain access to the preoperative magnetic resonance image of the patient. Others (2/6, 33%) provided educational information, such as the study by Simone et al [10], which used annotations on the patient's body and a mentor's voice instructions.

RQ 2: What Type of Sensorization Is Used for AR-Based Laparoscopic Surgery Training?

To assist in the training activity, it is essential to know in real time the location of the different elements involved in the training environment, such as the training scenario, surgical

XSL•FO
RenderX

instruments, patient, and posture of the surgeon. By monitoring these elements, it is possible to adapt the training process to the trainee's educational needs. In addition, as this task is manipulative, it is relevant to consider the haptic stimuli. Therefore, in this section, we analyze which aspects related to the use of sensors (sensorization) were used in scientific literature. In this regard, we classified systems according to 2 aspects: the element being measured or sensorized (Table 3) and the technology used (Table 4).

Table 3. Classification by type of element being measured and technology.

Element being measured or sensorized	Studies
User tracking	
Force feedback	[26,39,46,52]
Body	[6,8,29,42]
Eye	[11,41]
Instrument	[9,12,15,18,19,21-25,27-40,43,45,46,49,51-54]
Target tracking	
Body	[13,14,16,17,20,22,33,38,48,49,51,53]
Instrument	[13,44,56]
Unspecified	[10,47,50,55,57]

Table 4. Classification of articles by sensorization technology.

Technology	Studies
Sensor based	[6,11,13,18,21,25,27,33,51]
Optic	[8,9,12-20,22-24,28-33,35-38,40-45,48,49,53,54,56]
Force	[26,34,39,46,52]
Unspecified	[10,47,50,55,57]

As we have pointed out, there are different elements in the training environment that can be tracked to know their location, motion, or behavior. We distinguished *user tracking* (ie, monitoring the surgeon and elements related to the surgeon). In this case, we organized the tracked elements into 3 categories: *body*, which were systems that recorded the surgeon's body motion for kinematic analysis; *eye*, to refer to the analysis of the surgeon's gaze and where they focus their attention; and *instrument*, which referred to the tracking of the surgical instruments. In addition to the user, the target can be tracked (*target tracking*), that is, the patient or the simulated model of the patient. This type of analysis was usually conducted using computer vision techniques. In this case, when we talk about *body*, we refer to the patient's body and organs, whereas *instrument* refers to other objects or markers that can be placed inside the patient and can be tracked or recognized, such as the needle used in suturing tasks [44].

Not only the object being tracked is important but also how they were tracked, that is, the technology and techniques used to achieve this tracking. For this reason, we divided the technology into 3 categories: *sensor based*, *optic*, and *force*. *Sensor-based* technology refers to those systems that used a set of sensors specifically located to track and measure certain interactions in those locations. An example could be the use of electromagnetic trackers for recording laparoscopic tool motion [25]. *Optic* technologies refer to the identification and tracking of objects using artificial vision techniques. Felinska et al [41] used iSurgeon, an AR tool that allowed instructors to project their hand gestures in real time onto the laparoscopic screen, enhancing mentoring guidance during laparoscopic training sessions. *Force* refers to those systems that used devices capable of measuring the interaction force of the laparoscopic instrument [26].

Most of the studies (46/51, 90%) used some type of sensorization of the training environment in which the surgical simulation was performed.

In most AR systems, the user saw a real world augmented only with visual information and had no means to interact with virtual objects. If we want to manipulate virtual objects, we need another kind of information—*haptic* [58]. Therefore, together with the display of visual information recorded by tracking devices, the availability of augmented haptic information seems relevant in laparoscopic surgery training, where manipulative tasks are fundamental. In this sense, 8% (4/51) of the studies included some type of haptic feedback in the laparoscopic surgery training environment. For example, Ivanova et al [52] used a virtual 3D model of an organ and a sensorized physical simulator, providing force feedback on the hardness of the tissue being touched and the force exerted.

Regarding the element being measured or sensorized (tracked), the most common was tracking of the surgeon (user tracking). In this case, the most relevant elements tracked were the instruments used by the surgeon (32/51, 63%). Pagador et al [27] used electromagnetic trackers to record the instrument's position during the training activity, whereas Lahanas et al [45]

XSL•FO
RenderX

used fiducial markers attached to the tip of the instrument for tracking instrument position and rotation. There were only a few studies (4/51, 8%) that made use of surgeon posture tracking. For example, Cizmic et al [42] tracked hand position to project the hand movements of the trainer onto the laparoscopic image to enhance communication between trainer and trainee. The use of eye tracking was present in only 4% (2/51) of the studies analyzed. This technique was used by some authors to interact with sets of images and other elements of a user interface by means of gaze [11] or to assess performance and quality of communication [41].

Target tracking was the next most used sensorization in laparoscopic surgery training solutions. In this case, body tracking (tracking objects in the training environment or patient organs) was the category that comprised most of the reviewed articles. For this purpose, most of the studies (35/51, 69%) used artificial vision techniques to identify and track body structures. Viglialoro et al [13] used AR as a training aid in simulator-based laparoscopic cholecystectomy training, mainly during the isolation of the cystic duct and artery (Calot triangle). AR technology allowed trainees to visualize these hidden structures during the training activity [13]. Pessaux et al [20] proposed an AR-based assistance system to superimpose information from preoperative images onto the patient's skin using a beamer, generating body transparency with visualization of deep structures. Finally, only few studies (2/48, 4%) tracked information about artificial markers or objects (instruments) within the surgical workspace inside the patient's body. Preetha et al [44] tracked the surgical needle during a laparoscopic suturing task.

RQ 3: What Type of Simulator and Setup Is Used for AR-Based Laparoscopic Surgery Training?

Regarding the simulator and setup used for laparoscopic training, 2 main characteristics were considered. The first was the origin of the training system, which may be a widely used commercial simulator (eg, ProMIS) or a prototype developed specifically for the research being carried out or pending to be released into the market. The second was the physical characteristics of the simulator and the models used in the training activity (training model). It could be an artificial model created solely for the purpose of training or an organic model (in vivo or ex vivo) obtained from organic tissue (Table 5).

Table 5. Classification of the different training models (artificial training model, ex vivo, and in vivo) according to the type of simulator (commercial or prototype).

	Training model: artificial	Organic		Unspecified
		Ex vivo	In vivo	
Prototype	• Viglialoro et al [13] • Andersen et al [16] • Horeman et al [21,26] • Kim et al [39] • Shao et al [43] • Shabir et al [47] • Condino et al [48] • Viglialoro et al [51] • Wagner and Rozenblit [53]	• Baumhauer et al [38]	—[a]	• Zorzal et al [11] • Loukas et al [22] • Rozenblit et al [46]
Commercial				
ProMIS	• Nomura et al [15] • Nugent et al [23] • Brinkman et al [24] • Botden et al [34,35] • Van Sickle et al [40]	• Strickland et al [28] • Leblanc et al [31,32]	—	• Leblanc et al [29,30] • Botden et al [36,37]
FLS[b]	• Zahiri et al [14] • Vera et al [19]	—	—	—
Others	• Wild et al [8] • Arts et al [12]	• Pagador et al [25] • Cizmic et al [42] • Fusaglia et al [49]	—	—
Unspecified	• Cau et al [9] • Lahanas et al [18] • Preetha et al [44] • Lahanas et al [45] • Larrarte and Alban [54] • Doughty et al [55] • Rewkowski et al [56]	• Felinska et al [41] • Ivanova et al [52]	• Sánchez-Margallo et al [6] • Simone et al [10] • Pessaux et al [20] • Teber et al [33]	• Hughes-Hallett et al [17] • Pagador et al [27] • Koehl et al [50] • Gupta et al [57]

[a]Combinations of devices and information types under which no articles were categorized, indicating nonapplicability in the reviewed studies.

[b]FLS: Fundamentals of Laparoscopic Surgery.

XSL•FO
RenderX

The use of commercial simulators was more widespread (20/51, 39%) than the use of prototypes (14/51, 27%). Commercial simulators mostly used artificial models (10/20, 50%), such as beads for the eye-hand coordination tasks [15]. There were also some studies (6/20, 30%) that used organic models. In this case, some authors compared skill acquisition after training with human cadavers versus using AR-enhanced artificial models [31]. Strickland et al [28] used a lamb liver to which a piece of marshmallow was introduced to simulate a tumor to be removed, and Pagador et al [25] used a porcine stomach for suturing tasks. There were some studies (4/20, 20%) that did not specify the model used by the commercial simulator or that could not be included in any of the defined categories.

Regarding the simulator prototypes, most (8/12, 67%) used artificial models. Lahanas et al [18] used markers inside a box trainer to superimpose images of different interactive objects, such as rings or beads that needed to be manipulated. Regarding the use of organic models, only Baumhauer et al [38] used an ex vivo porcine kidney to test the image overlay capability of the AR-based training system presented. The rest of the papers did not specify the model used or did not fit into any of the categories.

Artificial models were the most used training models (26/51, 51%), especially in simulator prototypes (10/17, 59%), followed by ex vivo training models (8/51, 16%). The latter stand out as 62% (5/8) of the studies that used ex vivo models used commercial simulators. The studies that used in vivo models did not refer to the simulator used as these were real operations (either on humans or animals).

There were studies that did not specify whether they used a commercial simulator or developed a prototype (17/51, 33%). Of these studies, 41% (7/17) used artificial training models, such as the system presented by Doughty et al [55], which did not use a specific laparoscopic simulator as it was intended to be used in any type of surgery. Meanwhile, 6 of the studies that did not specify whether they used a commercial simulator or a prototype used organic models, with 4 (67%) of them being in vivo models. Simone et al [10] used MR smart glasses to enable tele-mentoring in real surgical procedures during the COVID-19 pandemic.

There were a few articles (4/51, 8%) that did not specify either the type of simulator or training model used. Gupta et al [57] mentioned the need for a development framework for this type of simulators in laparoscopic surgery training and which aspects should be considered. However, they did not make use of any training system as it was a theoretical study [57]. Another study tested the impact of AR elements on inattentional blindness during a laparoscopic operation using a prerecorded video [17]. The study by Koehl et al [50] focused on the generation of modular models that could be used for other applications as a 3D visual model or as a force feedback generation system. Pagador et al [27] designed and evaluated a laparoscopic tool tracking system for further assessment purposes.

RQ 4: What Type of Evaluation Is Used to Assess Skill Acquisition in AR-Based Laparoscopic Surgery Training?

In total, 2 main ways of addressing the assessment of laparoscopic skill acquisition in AR-based laparoscopic surgery training solutions were reported. One way of evaluation was through objective evaluation metrics [15-18,22-26,28-30,32-34,38,40-42,44,49,52,54-56] usually calculated by the training system. Another way was subjective evaluation carried out by experts or by means of interviews or self-assessment questionnaires [6,10,11,36,43,47]. Some studies used both ways [8,9,12-14,19,21,31,35,37,48], while others used none [20,27,39,45,46,50,51,53,57].

Objective evaluation was the most used method for the assessment of skill acquisition in laparoscopic surgery. In terms of the metric used, most of the studies (25/51, 49%) evaluated the performance of surgeons to assess whether the system enhanced laparoscopic skill acquisition. A variety of metrics were used, such as execution time, path length, or motion smoothness (Table 6). Time, whether related to task completion or training duration, was a commonly used metric [41] that assessed the time spent on each task as a performance indicator. There were other studies (2/31, 6%) that, instead of measuring the time needed to complete the task, measured the time that the laparoscopic instruments spent in a specific area, for instance, during a suturing task [35].

Table 6. Type of metric used for objective evaluation in each study analyzed.

Metric	Studies
Training time or task time	[8,12,19,23,28,34,35,40,41]
Time in correct area	[34,35]
Path length	[15,18,19,21,24,25,29,30]
Objective Structured Assessment of Technical Skills	[8,31,32,41,42]
Global Operative Assessment of Laparoscopic Skills	[8,42]
Economy of movement	[15,24]
Force	[21,26,34,35]

Path length was another prevalent objective metric (8/31, 26%). This metric records the path followed by the laparoscopic instruments during the performance of the training task. There were another 6% (2/31) of the studies that, apart from measuring path length, also computed the economy of movements of the surgical instruments [15,24]. Force metrics were also commonly used (4/31, 13%). Botden et al [34] evaluated the performance of a suturing task by assessing the strength of the knot. The

XSL•FO
RenderX

Objective Structured Assessment of Technical Skills (OSATS) and the Global Operative Assessment of Laparoscopic Skills (GOALS) formularies are 2 recognized standards for assessing laparoscopic skills that were also used in some of the studies, with the OSATS being used in 14% (5/36) of the studies and the GOALS being used only twice (2/36, 6%).

Subjective evaluation was mostly used for evaluating the usability of the AR tools (6/51, 12%). For example, Arts et al [12] made use of a questionnaire to assess the validity of the ProMIS as an AR laparoscopic surgery simulator. The combination of both types of evaluation (objective and subjective) was used in 22% (11/51) of the studies. Cau et al [9] combined an expert's subjective assessment of a suturing task with metrics obtained by the system itself.

RQ 5: What Types of Surgical Tasks or Procedures Are Used in AR-Based Laparoscopic Surgery Training?

Regarding the laparoscopic training activities (tasks or procedures), in most of the studies (26/51, 51%), trainees performed basic tasks commonly used in the early stages of laparoscopic surgery training programs. These tasks were divided into 5 main categories: navigation tasks focused on the skill needed to explore the surgical workspace using the camera and laparoscopic instruments; object manipulation tasks aimed to train the user to move, rotate, or manipulate objects or organs using the laparoscopic instruments; dissection tasks focused on the ability to separate tissue and anatomical structures without causing any damage to them; cutting tasks focused on training users to make precise cuts in organs and tissues; and suturing tasks focused on training for suturing tissues, including knotting skills. Finally, there were studies that used surgical procedures such as partial nephrectomy on ex vivo models [38] or sigmoid colectomy on cadavers [31].

Similarly, considering its main clinical features, such as the surgical procedure performed or the training model used, the category of surgical specialty into which each study fell or could fall was defined. For those studies in which a training surgical intervention was performed, the surgical specialty was the one which that procedure corresponded to. There were also some studies that included simple tasks. In this case, the surgical specialty was selected based on the training model used. For those cases in which surgical skills did not focus on a particular surgical specialty, the studies were classified as "All" specialties (Table 7).

XSL•FO
RenderX

Table 7. Studies organized by type of training task or procedure and surgical specialty.

	Navigation	Object manipulation	Dissection	Cutting	Suturing	Surgical procedure	Unspecified
General	• Rozenblit et al [46]	• Rozenblit et al [46]	—[a]	—	—	• Wild et al [8] • Simone et al [10] • Viglialoro et al [13] • Pessaux et al [20] • Strickland et al [28] • Leblanc et al [29-32] • Felinska et al [41] • Cizmic et al [42] • Shao et al [43] • Fusaglia et al [49] • Viglialoro et al [51]	• Ivanova et al [52]
Urology	—	—	—	—	—	• Sánchez-Margallo et al [6] • Simone et al [10] • Teber et al [33] • Baumhauer et al [38]	—
Thoracic	• Loukas et al [22]	—	—	—	—	—	—
All	• Lahanas et al [45] • Wagner and Rozenblit [53]	• Zahiri et al [14] • Nomura et al [15] • Andersen et al [16] • Horeman et al [21] • Nugent et al [23] • Botden et al [36] • Rewkowski et al [56]	• Arts et al [12] • Nugent et al [23] • Brinkman et al [24]	• Arts et al [12] • Andersen et al [16] • Brinkman et al [24]	• Cau et al [9] • Vera et al [19] • Pagador et al [25] • Horeman et al [26] • Botden et al [34-36] • Van Sickle et al [40] • Preetha et al [44] • Shabir et al [47]	• Zorzal et al [11] • Koehl et al [50] • Doughty et al [55]	• Pagador et al [27] • Condino et al [48]
Unspecified	• Larrarte and Alban [54]	• Lahanas et al [18]	—	• Lahanas et al [18]	• Botden et al [37]	—	• Hughes-Hallett et al [17] • Kim et al [39] • Gupta et al [57]

[a]Combinations of training task or procedure and clinical specialty under which no articles were categorized, indicating nonapplicability in the reviewed studies.

A large proportion of the studies analyzed (20/51, 39%) used surgical procedures rather than simple tasks to train and assess surgical skills. General surgery was the surgical specialty of most of the studies (14/20, 70%). In the study by Viglialoro et al [13], they presented an AR solution that assisted in the identification of the artery and cystic duct in a training simulator for cholecystectomy. Other studies (4/20, 20%) were included under the specialty of urology, such as the study by Teber et al [33], who presented a system for formative assistance during laparoscopic partial nephrectomy. Other studies (3/20, 15%) performed procedures that were not classified into any specific surgical specialty but could be used in any specialty, such as the study by Doughty et al [55], who developed a context-aware system capable of assisting the surgeon depending on the training task being performed.

Regarding basic laparoscopic surgery training tasks, the most common task used was suturing (11/25, 44%). Botden et al [34], for example, used suturing to analyze the evolution of the trainees' technical skills, assessing how fast trainees performed tasks and the strength of the knotting. Object manipulation tasks was also common in the studies analyzed (9/25, 36%). Lahanas et al [18] presented 3 different object manipulation tasks to assess trainees' laparoscopic skills using their AR-based simulator. The use of navigation tasks was less common in the training settings (5/25, 20%). Fusaglia et al [49] presented an

XSL•FO
RenderX

overlay system to show hidden anatomical structures during the performance of laparoscopic navigation tasks. The first task was to press a set of buttons in a specific order using the laparoscopic instruments, the second task was to transfer an object using the laparoscopic tools, and the last task was to cut a virtual object (using AR technology). Another type of task were dissection tasks, but they were less common that the rest (3/25, 12%). One study proposed the use of 2 balloons (one inside the other) that the trainee had to separate keeping the inner balloon inflated (ie, without being damaged) [12]. Only 16% (4/25) of the studies focused on cutting tasks. In the study by Brinkman et al [24], a full week of laparoscopic surgery training was proposed, having training sessions once a day in which some laparoscopic cutting tasks were included. Finally, there were 12% (6/51) of the articles that did not specify the task performed.

Discussion

Principal Findings

Overview

AR is an emerging technology that is being applied to various fields, including health care. In this sense, applications have been developed mainly oriented toward surgical planning, medical and surgical training, and surgical assistance [59,60]. In this study, a review of scientific literature oriented toward solutions for assistance in laparoscopic surgery training was carried out. We analyzed aspects that we considered critical to innovate and advance in this field, such as the devices used to provide AR to the user, the training environments, the types of training tasks or procedures, and the evaluation metrics used to analyze the evolution of students during the training activities and provide them with educational feedback. For this purpose, we analyzed all the RQs raised in this work with the aim to provide some observations that may help researchers guide their new proposals and innovative solutions to achieve an improvement in the field of laparoscopic surgery training.

The evolution of the number of publications included in this scoping review was heterogeneous and may be associated with the evolution of AR technology and the introduction of new applications and devices. Although there were some studies in the first years of this century associated with the appearance of the first AR applications in games and smartphones, it was in 2009 when a relevant increase in interest in the use of AR was observed, coinciding with the success of applications such as the AR toolkit, an open-source library for the creation of AR applications based on the identification of visual markers. This relevant interest continued until 2015, which could be related to the launch of Google Glass. This device was the first proposal that took AR out of the screen and into the real world using a wearable device. However, it was not until the launch of the second version of Microsoft's HoloLens (2019) when the potential of this type of technological proposals increasingly distanced from solutions linked to screens began to be glimpsed. This fact may be related to the recent increase in interest in the use of this type of AR-linked technologies in laparoscopic surgery training.

In the following sections, we will analyze the results to identify relevant insights that address each of the RQs.

Devices and Feedback (RQ 1)

While video devices such as monitors, laptops, smartphones, and tablets have traditionally been used in the early stages of AR technologies to enhance laparoscopic surgery training, the use of OST devices is a relatively new development due to the recent emergence of devices that facilitate these functions (the first device appeared in 2020). Therefore, only 12% (6/51) of studies were found that explored the application of OST technology in this context [6,9-11,55,56].

OST devices offer greater versatility than traditional video devices (monitors, smartphones, and tablets) for laparoscopic surgery training. As several studies (6/51, 12%) demonstrated, OST devices provide novel interaction methods and functionalities such as eye or gaze tracking, voice commands, and gestural interactions to manipulate virtual objects or add annotations. For instance, the head gaze was used to point at the laparoscopic virtual screen to enhance communication between trainer and trainee [11]. Felinska et al [41] used eye-tracking technology to create a heat map and analyze differences between the visual focus areas of trainees and instructors. In another study using OST technology, the Vuforia library was used to recognize artificial visual markers in the training environment to insert virtual elements to enhance the experience of laparoscopic surgery training [56]. However, this solution was limited to simulated environments, making it difficult to use in real clinical settings as it relied on 5 cameras for marker recognition. Sánchez-Margallo et al [6] and Simone et al [10] focused on the creation of 3D models from real preoperative images to implement AR-based training applications. In addition, this functionality was extended by allowing tutors to add annotations and information to the models [10].

One of the advantages of OST devices is that they can be used in any training environment used in laparoscopic surgery. The trainee always visualizes the real training environment, and it is the augmented information (holograms) that is superimposed onto the real image through the glasses. For instance, in the study by Sánchez-Margallo et al [6], it was not necessary to connect the OST device to the laparoscopic simulator as it did not require the laparoscopic video feed.

However, Zorzal et al [11] accessed the video source from the laparoscopic tower and explored the ergonomic improvements that OST devices could provide by displaying the endoscopic video in any place of the operating room and even following the surgeon's head.

Regarding ARV devices, the analyzed studies highlighted their potential to enhance the experience during laparoscopic surgery training. In the study by Preetha et al [44], a convolutional neural network was used to predict depth from laparoscopic imaging and recognize the manipulation of laparoscopic instruments for use in surgical skill assessment. Another study also used depth prediction to generate and display optimal trajectories for surgical instruments [53]. Solutions for depth perception support were also presented to improve the training process of novice

XSL•FO
RenderX

trainees in laparoscopic surgery [45,49]. The presence of assistive information in the surgeon's field of view is not a major disadvantage as it has been shown that the use of information-enhanced holograms does not have a negative impact on the surgical working field [17]. In another case, ARV technology was used to generate 3D models from preoperative imaging and superimpose them onto laparoscopic images as a support during the performance of the surgical task or formative procedure [20].

Another important aspect to consider during laparoscopic surgery training is to analyze force feedback during the performance of different laparoscopic tasks and procedures. This factor was explored in an ARV training environment, showing that feedback of force exerted improves the acquisition of tissue-handling skills [21,26]. Other studies (6/51, 12%) investigated the overlay of assistive content in the performance of training tasks and procedures by means of ARV applications and concluded that it helps the trainees reduce path length and deviation along the depth axis and improve their orientation skills [18,48,49,52-54].

All in all, we can see that both OST and ARV devices show significant potential as training aids in laparoscopic surgery. Although OST devices are a more recent and less explored development, they offer enhanced versatility compared to traditional video devices used in early AR technologies. The reviewed studies highlight the various applications of OST and ARV, including novel interaction methods, access to preoperative imaging, ergonomic improvements, depth analysis, and instrument motion analysis. This suggests that both OST and ARV have the potential to enhance laparoscopic surgery training, with OST devices offering greater versatility and opportunities for innovation in this field.

Sensorization (RQ 2)

The studies analyzed in this review mainly presented 2 types of sensing technologies: haptics and tracking. The use of haptic technology in laparoscopic surgery training has been a topic of considerable interest, as evidenced by the number of studies found on this topic (9/51, 18% of the studies in total). These studies explored the potential benefits of haptic feedback in enhancing the acquisition of laparoscopic skills and improving surgical performance. The use of haptic feedback in laparoscopic surgery simulation was examined in several studies (4/51, 8%), demonstrating its efficacy in improving surgeons' dexterity [24,26,39,52] The incorporation of haptic feedback in AR-based laparoscopic surgery training applications has been shown to lead to improved performance among novice surgeons [61] and improve surgeons' skills in laparoscopic suturing [23]. The efficacy of haptic feedback in laparoscopic skill acquisition was also studied by other authors (4/51, 8%) [34,36,37,62], who showed that it significantly enhanced surgeons' precision and instrument manipulation skills. These studies highlighted the potential of haptic feedback as a valuable adjunct to laparoscopic surgery training. It seems evident that haptic technology promises further advances in laparoscopic surgery training and proficiency.

Regarding tracking technologies, there were 2 main approaches: a sensor-based approach and an optical technology–based approach. The sensor-based approach uses sensors placed at specific locations to track the object in the training environment. Meanwhile, the optical technology–based approach uses image analysis to perform tracking of the objects of interest. The studies included in this review used artificial vision techniques and optical sensorization, mainly by using the laparoscopic camera feed. For instance, Loukas et al [22] used 2 different algorithms to estimate the tooltip position, one of them using an adaptive model of a color strip attached to the tool and the other one tracking directly the shaft of the tool. Optical technology plays a crucial role in accurate tracking in clinical settings. Zahiri et al [14] used this technology to evaluate the performance of an object transfer task in basic training environments by tracking colored rings. Another study focused on tracking anatomical structures such as the Calot triangle and user interaction during the training procedure [13]. Vision-matching techniques enhance trainees' visualization of hidden anatomical structures by superimposing 3D models onto the laparoscopic camera source, improving trainees' understanding of spatial relationships [33]. Vision-matching techniques were also used for pairing of real and augmented vision, some of them automatically and others semiautomatically. Thus, in the study by Hughes-Hallet et al [17], a semiautomated real-time overlay of preoperative models onto the anatomical field structures was performed, improving accuracy and efficiency in the AR-guided training process, whereas Pessaux et al [20] and Condino et al [48] used manual assistance to align the virtual objects with the patient's anatomy. Overall, the integration of artificial vision techniques and optical technology plays a crucial role in the identification and tracking of the anatomical structures of the patient. These solutions improve the understanding of the steps to be carried out during training activities, as well as the identification of manipulated objects by trainees in laparoscopic surgery training.

The eye-tracking technology present in many of the OST devices has shown great potential in various fields, including human-computer interaction and cognitive sciences. However, in the context of laparoscopic surgery training, the use of this technology remains relatively unexplored. In this scoping review, only 4% (2/51) of the studies investigated the incorporation of eye-tracking technology for assistance in laparoscopic surgery training. In one of the studies, gaze tracking (using head tracking) was performed to focus the laparoscopic camera on the area where the surgeon's gaze was centered [11]. This study made it possible to optimize the way of visualizing the information on the surgical working environment and the surgeon's ergonomics. Another study aimed to use eye-tracking technology to detect convergence between the trainee's and the trainer's gaze, as well as between the trainee and the target area [41]. By monitoring eye movements and fixation points, the system could provide objective feedback on the trainee's visual attention and alignment with the desired gaze targets. This approach has the potential to enhance trainee-trainer communication and improve the trainee's spatial awareness within the surgical field. Considering these aspects, there is a clear need for further investigation on the use of this technology to improve the training process in laparoscopic surgery. Eye tracking offers unique advantages, such as providing the trainee's visual focus, gaze patterns, and cognitive workload in

XSL•FO
RenderX

real time. The integration of eye-tracking technology into laparoscopic simulators and training platforms can enhance the visual perception, spatial orientation, and overall surgical performance of the trainee.

In addition, tracking of laparoscopic instruments and deformations in bodies or organs was also performed using artificial visual markers. In some cases, these were colored markers that were placed around the distal area of the instrument shaft [22,37,40]. Reflective markers were also used, although in this case, it was necessary to use infrared light and camera filters for identification [9]. The use of these techniques helps improve the identification and tracking of the laparoscopic instruments, which is essential to evaluate the surgical performance of trainees and assist them during the training activities. In addition, various studies (3/33, 9%) explored the calculation and modeling of deformations in bodies or organs to be used in advanced AR applications for laparoscopic surgery training. In one study, visual markers were inserted directly on the liver so that they could be identified by the laparoscopic image and allow for the localization and superimposition of preoperative 3D models of the liver and other relevant information on the endoscopic image [33,38]. Another study used retroreflective markers on the laparoscopic instruments and within the formative surgical field to facilitate the superimposition on the laparoscopic image of a preoperative 3D model [49].

Another type of technology used for tracking surgical instruments and organs involved in the formative activity was the use of electromagnetic sensors. Viglialoro et al [51] incorporated these sensors into the laparoscope and bile ducts or arterial tree of an artificial training model for real-time monitoring during the performance of simulator tasks. Electromagnetic sensors inserted into nitinol tubes representing the ducts and arterial trees helped infer possible deformations of the sensed tubular structures. Consequently, the virtual scene was updated in real time, augmenting the laparoscopic images with information on the real-time position of these anatomical structures.

By using the aforementioned techniques for the calculation of organ deformations and the inclusion of 3D models, these studies significantly contributed to the integration of virtual models into the laparoscopic surgery environments, enhancing surgical training scenarios in real time.

Training Simulator and Setup (RQ 3)

Regarding the type of simulator used for AR-based laparoscopic surgery training, most of them were box trainers to which an AR assistance system was added. Of note is the ProMIS simulator (Haptica, Inc), which was one of the first commercial simulators for laparoscopic surgery training with AR functionalities. This simulator included basic training tasks such as dissection and suturing but also more complex procedures such as liver resection and sigmoid colectomy. This is the AR-based simulator with the largest number of scientific publications. Apart from this simulator, we highlight other commercial options to which AR-based laparoscopic surgery training assistance systems were added, such as the Fundamentals of Laparoscopic Surgery simulator (Limbs & Things Ltd) [14,19], the Szabo Pelvic Trainer simulator (ID Trust Medical) together with the iSurgeon system [8,42], the eoSim simulator (eoSurgical) [12], the SIMULAP-IC05 simulator (Jesús Usón Minimally Invasive Surgery Centre) [25], or the Body Torso laparoscopic trainer simulator (Pharmabotics Ltd) [49].

Regarding the training model, the use of artificial models was most common because they are generally more accessible than organic models, although with less realism. However, more and more realistic artificial models have been achieved by simulating the anatomy and behavior of real tissues, as in the case of the study by Viglialoro et al [51], in which they presented a 3D replica of the patient's own liver and gallbladder. Other options presented were the enhancement of the physical model using AR imaging [18].

Evaluation (RQ 4)

In the different studies analyzed, various types of metrics were used to evaluate the quality of surgical performance and the surgical skills of trainees. These metrics, according to their nature, were divided into objective and subjective metrics.

Obtaining objective metrics is one of the main advantages of using AR techniques in laparoscopic surgery training. Some of the training simulators, as is the case of ProMIS, allow for recording the actions executed by the trainee to obtain metrics such as path length or motion smoothness of the laparoscopic instruments, which can help evaluate the trainee's performance and learning curve without the need for constant supervision by an expert evaluator.

Although there are some metrics that were recurrently used in the analyzed studies, so far there is no standardized set of objective metrics used to evaluate performance or technical skills in laparoscopic surgery. The most common metrics used were execution time, path length, economy of movement, and motion smoothness of laparoscopic instruments [15,18,24,25,28-30,40]. Other metrics were also used for specific tasks, such as knotting strength, time spent in the correct suturing and knotting area [34,35], or force exerted by laparoscopic instruments on the target organ or tissue [21,26].

Other types of surgical performance quality assessments based on questionnaires administered by external evaluators were also used, such as the OSATS, a highly standardized surgical assessment tool [8,31,32,41,42]. Horeman et al [21] and Leblanc et al [31] also used similar questionnaires based on peer reviews to determine surgeons' experience in laparoscopic surgery.

Subjective evaluation was commonly used for assessing the validity of the proposed solution. None of the studies used a standard questionnaire for assessing system validity, but many similarities were found in terms of the aspects of the systems they evaluated. For example, usability, realism, and didactic value were included in several studies (7/18, 39%) [12,13,19,35-37,48]. Questionnaires with Likert scales were the most commonly used format in the studies.

Regarding assessment, the size and type of the samples included in the studies were also important. There was a wide variation in sample size, with samples ranging from 10 individuals [13]

XSL•FO
RenderX

to 270 participants [15]. Some studies (2/25, 8%) included >100 participants [12,15], but most of them (23/25, 92%) presented sample sizes of <100 participants [8,17,23,24,29,31,32,37,41,42] or <30 individuals [9,13,14,16,19,28,30,34,35,40,43,47,48].

Finally, the included studies encompassed participants with different levels of laparoscopic surgery experience: novices, intermediates, and experts. Novices, referring to individuals with limited or no previous laparoscopic surgery experience, were frequently included in the studies. Intermediate-level participants, characterized by a moderate level of experience, were present in a few studies (4/25, 16%) [12,28,36,43]. This may be because their inclusion makes it difficult to differentiate results between study groups as there is sometimes no clear difference between the intermediate group and the novice or expert groups. Experts, representing individuals with advanced laparoscopic skills, were also included in several studies (8/25, 32%). This stratification of participants allows for a comprehensive validation of the system's effectiveness across different proficiency levels and helps detect the different needs that these groups may have.

Tasks and Procedures (RQ 5)

Regarding the tasks and procedures used as training activities in the studies analyzed, the most widespread ones were those focused on eye-hand coordination tasks, such as peg transfer, instrument movement, and instrument navigation tasks [14-16,18,21-24,36,45,46,53,54,56]. The next most used type of task was the simulator suturing task [9,19,25,26,34-37,40,44,47]. Although less frequent, some studies (17/51, 33%) also presented more complex models for procedures such as cholecystectomy [8,13,15,42,51]. In this case, models of the liver and biliary anatomy were used to facilitate the training of novice surgeons in gallbladder extraction using laparoscopic techniques. These types of procedures, although still basic, are highly appropriate for learning basic tasks such as the localization of critical anatomical structures (cystic duct and cystic artery) and the performance of dissection and cutting tasks for dissection and cutting of the cystic duct, cystic artery, and gallbladder removal. In a closer-to-reality environment, AR-based assistive tools were used for procedures such as cholecystectomy [41] in ex vivo (porcine) experimental models and sigmoid colectomy in human cadavers [31,32].

Considering that the main objective of AR-based training assistance applications in laparoscopic surgery is to provide the students with support tools during their training activities, the type of task or procedure to be chosen has a great influence on the usefulness of these tools. It seems evident that, in considerably basic tasks, the training assistance systems would not provide significant value to the student, mainly due to the low level of complexity of the activity. However, it is in the more challenging tasks and procedures where training assistance systems may provide remarkable training value. For instance, a task that can present a high level of difficulty for novice laparoscopic surgeons is the performance of suturing. Support tools can enhance reality for the student with visual information to assist in the performance of the suture, such as the proper grip of the needle, the passage of the needle through the tissue, or the process of double and single knotting. In more complex procedures such as laparoscopic cholecystectomy or lymphadenectomy, these training assistance systems could support the students in locating complex or hidden anatomical structures, as well as remotely transmitting instructions from tutors who are reviewing the students' training process, among other possibilities.

Limitations

Although the PRISMA-ScR guidelines were followed to carry out this scoping review to reduce possible biases, we must highlight some limitations encountered in our work. Although most relevant articles found in other databases such as Springer or Google Scholar also appeared in the databases consulted, it is possible that some papers were not included because these databases were not used directly. In addition, the specific search strategies and keywords used in the various databases may have led to the exclusion of articles that used different terminology to refer to the same topics. Although the number of articles affected was small, restricting the selection to English-language publications and excluding articles that we could not access may have led to the omission of studies that could have provided relevant information.

Regarding the devices used, OST devices are relatively new, so there is still much to explore in terms of what they can contribute to the field of laparoscopic surgery training, which we will address in future studies. It should also be noted that there may be interesting technical proposals for the support of laparoscopic surgery training that, as they have not yet been used for such training, were not included in the studies analyzed in our scoping review. Furthermore, the lack of standardized metrics for evaluating the different support systems did not allow for a comprehensive comparison between systems and a comparative analysis of their results.

Conclusions

In conclusion, this scoping review sheds light on the dynamic landscape of AR technologies within laparoscopic surgery training. Although OST devices have recently emerged and their advantages over traditional AR methods are still being explored, it is relevant to indicate that the results obtained are promising and open up new opportunities for the use of AR in this type of training activities. In turn, haptic feedback emerges as a valuable asset for the acquisition of laparoscopic skills. Eye-tracking technology provides relevant information during learning, although its application is in an incipient phase that requires further exploration. The prevalence of commercial simulators and artificial models underscores the delicate balance between safety and realism. Regarding assessments, this review highlights the importance of extending the use of expert assessments, such as OSATS or GOALS, and underlines the need for standardized objective assessment. These assessments are based on the study of participants' behavior in simple tasks such as navigation and object manipulation and in other tasks with a higher level of difficulty, such as suturing. In any case, it seems clear that AR would be more useful for more complex tasks related to complex surgical procedures. These findings beckon future research to analyze the complete potential of AR in enhancing laparoscopic skill acquisition.

XSL•FO
RenderX

Acknowledgments

This study was cofinanced by the Complementary Plan for Biotechnology Applied to Health; the Ministry of Science and Innovation (Spain) with funds from the European Union for the NextGenerationEU recovery, transformation, and resilience plan; the Regional Government of Extremadura and Regional Government of Castilla–La Mancha (PRTR-C17.I1); and the Regional Government of Extremadura (TA18023), and it is also funded by the University of Castilla–La Mancha and the European Regional Development Fund (2022-GRIN-34436).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy.
[DOCX File , 14 KB - jmir_v27i1e58108_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 109 KB - jmir_v27i1e58108_app2.docx]

References

1. Viglialoro RM, Condino S, Turini G, Carbone M, Ferrari V, Gesi M. Augmented reality, mixed reality, and hybrid approach in healthcare simulation: a systematic review. Appl Sci 2021 Mar 06;11(5):2338. [doi: 10.3390/app11052338]
2. DeMasi SC, Katsuta E, Takabe K. Live animals for preclinical medical student surgical training. Edorium J Surg 2016;3(2):24-31 [FREE Full text] [Medline: 28713875]
3. Rocha e Silva R, Lourenção Jr A, Goncharov M, Jatene FB. Low cost simulator for heart surgery training. Braz J Cardiovasc Surg 2016;31(6):449-453 [FREE Full text] [doi: 10.5935/1678-9741.20160089] [Medline: 28076623]
4. Forgione A, Guraya S. The cutting-edge training modalities and educational platforms for accredited surgical training: a systematic review. J Res Med Sci 2017;22:51 [FREE Full text] [doi: 10.4103/jrms.JRMS_809_16] [Medline: 28567070]
5. Sánchez-Margallo FM, Sánchez-Margallo JA, Moyano-Cuevas JL, Pérez EM, Maestre J. Use of natural user interfaces for image navigation during laparoscopic surgery: initial experience. Minim Invasive Ther Allied Technol 2017 Oct;26(5):253-261. [doi: 10.1080/13645706.2017.1304964] [Medline: 28349758]
6. Sánchez-Margallo JA, Plaza de Miguel C, Fernández Anzules RA, Sánchez-Margallo FM. Application of mixed reality in medical training and surgical planning focused on minimally invasive surgery. Front Virtual Real 2021 Oct 28;2:23. [doi: 10.3389/frvir.2021.692641]
7. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
8. Wild C, Lang F, Gerhäuser AS, Schmidt MW, Kowalewski KF, Petersen J, et al. Telestration with augmented reality for visual presentation of intraoperative target structures in minimally invasive surgery: a randomized controlled study. Surg Endosc 2022 Mar 09;36(10):7453-7461. [doi: 10.1007/S00464-022-09158-1]
9. Cau M, Sandoval J, Arguel A, Breque C, Huet N, Cau J, et al. Toward optimal learning of the gesture in laparoscopic surgery: methodology and performance. J Clin Med 2022 Mar 03;11(5):1398 [FREE Full text] [doi: 10.3390/jcm11051398] [Medline: 35268489]
10. Simone M, Galati R, Barile G, Grasso E, De Luca R, Cartanese C, et al. Remote mentoring in laparotomic and laparoscopic cancer surgery during COVID-19 pandemic: an experimental setup based on mixed reality. Med Educ Online 2021 Dec 29;26(1):1996923 [FREE Full text] [doi: 10.1080/10872981.2021.1996923] [Medline: 34713779]
11. Zorzal ER, Campos Gomes JM, Sousa M, Belchior P, da Silva PG, Figueiredo N, et al. Laparoscopy with augmented reality adaptations. J Biomed Inform 2020 Jul;107:103463 [FREE Full text] [doi: 10.1016/j.jbi.2020.103463] [Medline: 32562897]
12. Arts EE, Leijte E, Witteman BP, Jakimowicz JJ, Verhoeven B, Botden SM. Face, content, and construct validity of the take-home EoSim augmented reality laparoscopy simulator for basic laparoscopic tasks. J Laparoendosc Adv Surg Tech A 2019 Nov 01;29(11):1419-1426. [doi: 10.1089/lap.2019.0070] [Medline: 31613679]
13. Viglialoro RM, Esposito N, Condino S, Cutolo F, Guadagni S, Gesi M, et al. Augmented reality to improve surgical simulation: lessons learned towards the design of a hybrid laparoscopic simulator for cholecystectomy. IEEE Trans Biomed Eng 2019 Jul;66(7):2091-2104. [doi: 10.1109/tbme.2018.2883816]
14. Zahiri M, Nelson CA, Oleynikov D, Siu KC. Evaluation of augmented reality feedback in surgical training environment. Surg Innov 2018 Feb 08;25(1):81-87. [doi: 10.1177/1553350617739425] [Medline: 29115184]

XSL•FO
RenderX

15. Nomura T, Matsutani T, Hagiwara N, Fujita I, Nakamura Y, Kanazawa Y, et al. Characteristics predicting laparoscopic skill in medical students: nine years' experience in a single center. Surg Endosc 2018 Jan;32(1):96-104. [doi: 10.1007/s00464-017-5643-5] [Medline: 28639038]
16. Andersen D, Popescu V, Cabrera ME, Shanghavi A, Gomez G, Marley S, et al. Medical telementoring using an augmented reality transparent display. Surgery 2016 Jun;159(6):1646-1653 [FREE Full text] [doi: 10.1016/j.surg.2015.12.016] [Medline: 26804823]
17. Hughes-Hallett A, Mayer EK, Marcus HJ, Pratt P, Mason S, Darzi AW, et al. Inattention blindness in surgery. Surg Endosc 2015 Jan 13;29(11):3184-3189. [doi: 10.1007/S00464-014-4051-3]
18. Lahanas V, Loukas C, Smailis N, Georgiou E. A novel augmented reality simulator for skills assessment in minimal invasive surgery. Surg Endosc 2014 Oct 11;29(8):2224-2234. [doi: 10.1007/S00464-014-3930-Y]
19. Vera AM, Russo M, Mohsin A, Tsuda S. Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. Surg Endosc 2014 Jun 25;28(12):3467-3472. [doi: 10.1007/S00464-014-3625-4]
20. Pessaux P, Diana M, Soler L, Piardi T, Mutter D, Marescaux J. Robotic duodenopancreatectomy assisted with augmented reality and real-time fluorescence guidance. Surg Endosc 2014 Mar 8;28(8):2493-2498. [doi: 10.1007/S00464-014-3465-2]
21. Horeman T, van Delft F, Blikkendaal MD, Dankelman J, van den Dobbelsteen JJ, Jansen FW. Learning from visual force feedback in box trainers: tissue manipulation in laparoscopic surgery. Surg Endosc 2014 Feb 12;28(6):1961-1970. [doi: 10.1007/S00464-014-3425-X]
22. Loukas C, Lahanas V, Georgiou E. An integrated approach to endoscopic instrument tracking for augmented reality applications in surgical simulation training. Int J Med Robot 2013 Dec 25;9(4):e34-e51. [doi: 10.1002/rcs.1485] [Medline: 23355307]
23. Nugent E, Shirilla N, Hafeez A, O'Riordain DS, Traynor O, Harrison AM, et al. Development and evaluation of a simulator-based laparoscopic training program for surgical novices. Surg Endosc 2012 Jul 7;27(1):214-221. [doi: 10.1007/S00464-012-2423-0]
24. Brinkman WM, Buzink SN, Alevizos L, de Hingh IH, Jakimowicz JJ. Criterion-based laparoscopic training reduces total training time. Surg Endosc 2011 Nov 1;26(4):1095-1101. [doi: 10.1007/S00464-011-2005-6]
25. Pagador JB, Sánchez-Margallo FM, Sánchez-Peralta LF, Sánchez-Margallo JA, Moyano-Cuevas JL, Enciso-Sanz S, et al. Decomposition and analysis of laparoscopic suturing task using tool-motion analysis (TMA): improving the objective assessment. Int J Comput Assist Radiol Surg 2012 Mar 14;7(2):305-313. [doi: 10.1007/s11548-011-0650-9] [Medline: 21842396]
26. Horeman T, Rodrigues SP, van den Dobbelsteen JJ, Jansen FW, Dankelman J. Visual force feedback in laparoscopic training. Surg Endosc 2011 Aug 20;26(1):242-248. [doi: 10.1007/S00464-011-1861-4]
27. Pagador JB, Sánchez LF, Sánchez JA, Bustos P, Moreno J, Sánchez-Margallo FM. Augmented reality haptic (ARH): an approach of electromagnetic tracking in minimally invasive surgery. Int J CARS 2010 Jul 2;6(2):257-263. [doi: 10.1007/S11548-010-0501-0]
28. Strickland A, Fairhurst K, Lauder C, Hewett P, Maddern G. Development of an ex vivo simulated training model for laparoscopic liver resection. Surg Endosc 2010 Nov 12;25(5):1677-1682. [doi: 10.1007/S00464-010-1440-0]
29. Leblanc F, Delaney CP, Ellis CN, Neary PC, Champagne BJ, Senagore AJ. Hand-assisted versus straight laparoscopic sigmoid colectomy on a training simulator: what is the difference? A stepwise comparison of hand-assisted versus straight laparoscopic sigmoid colectomy performance on an augmented reality simulator. World J Surg 2010 Dec;34(12):2909-2914. [doi: 10.1007/s00268-010-0765-0] [Medline: 20740283]
30. Leblanc F, Delaney CP, Neary PC, Rose J, Augestad KM, Senagore AJ, Colorectal Surgery Training Group. Assessment of comparative skills between hand-assisted and straight laparoscopic colorectal training on an augmented reality simulator. Dis Colon Rectum 2010 Sep;53(9):1323-1327. [doi: 10.1007/DCR.0b013e3181e263f1] [Medline: 20706077]
31. LeBlanc F, Champagne BJ, Augestad KM, Neary PC, Senagore AJ, Ellis CN, Colorectal Surgery Training Group. A comparison of human cadaver and augmented reality simulator models for straight laparoscopic colorectal skills acquisition training. J Am Coll Surg 2010 Aug;211(2):250-255. [doi: 10.1016/j.jamcollsurg.2010.04.002] [Medline: 20670864]
32. Leblanc F, Senagore AJ, Ellis CN, Champagne BJ, Augestad KM, Neary PC, Colorectal Surgery Training Group. Hand-assisted laparoscopic sigmoid colectomy skills acquisition: augmented reality simulator versus human cadaver training models. J Surg Educ 2010 Jul;67(4):200-204. [doi: 10.1016/j.jsurg.2010.06.004] [Medline: 20816353]
33. Teber D, Guven S, Simpfendörfer T, Baumhauer M, Güven EO, Yencilek F, et al. Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results. Eur Urol 2009 Aug;56(2):332-338. [doi: 10.1016/j.eururo.2009.05.017] [Medline: 19477580]
34. Botden SM, de Hingh IH, Jakimowicz JJ. Suturing training in augmented reality: gaining proficiency in suturing skills faster. Surg Endosc 2009 Sep;23(9):2131-2137 [FREE Full text] [doi: 10.1007/s00464-008-0240-2] [Medline: 19067051]
35. Botden SM, de Hingh IH, Jakimowicz JJ. Meaningful assessment method for laparoscopic suturing training in augmented reality. Surg Endosc 2009 Jan 1;23(10):2221-2228. [doi: 10.1007/S00464-008-0276-3]
36. Botden SM, Buzink SN, Schijven MP, Jakimowicz JJ. ProMIS augmented reality training of laparoscopic procedures face validity. Simul Healthc 2008;3(2):97-102. [doi: 10.1097/SIH.0b013e3181659e91] [Medline: 19088648]

XSL•FO
RenderX

37. Botden SM, Berlage JT, Schijven MP, Jakimowicz JJ. Face validity study of the ProMIS augmented reality laparoscopic suturing simulator. Surg Technol Int 2008;17:26-32. [Medline: 18802880]
38. Baumhauer M, Simpfendörfer T, Müller-Stich BP, Teber D, Gutt CN, Rassweiler J, et al. Soft tissue navigation for laparoscopic partial nephrectomy. Int J Comput Assist Radiol Surg 2008 May 29;3(3-4):307-314. [doi: 10.1007/S11548-008-0216-7]
39. Kim J, Choi C, De S, Srinivasan MA. Virtual surgery simulation for medical training using multi-resolution organ models. Int J Med Robot 2007 Jun 09;3(2):149-158. [doi: 10.1002/rcs.140] [Medline: 17619246]
40. Van Sickle KR, III DA, Gallagher AG, Smith CD. Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surg Endosc 2005 Jul 21;19(9):1227-1231. [doi: 10.1007/S00464-004-8274-6]
41. Felinska EA, Fuchs TE, Kogkas A, Chen ZW, Otto B, Kowalewski KF, et al. Telestration with augmented reality improves surgical performance through gaze guidance. Surg Endosc 2023 May 06;37(5):3557-3566 [FREE Full text] [doi: 10.1007/s00464-022-09859-7] [Medline: 36609924]
42. Cizmic A, Müller F, Wise PA, Häberle F, Gabel F, Kowalewski K, et al. Telestration with augmented reality improves the performance of the first ten ex vivo porcine laparoscopic cholecystectomies: a randomized controlled study. Surg Endosc 2023 Oct;37(10):7839-7848 [FREE Full text] [doi: 10.1007/s00464-023-10360-y] [Medline: 37612445]
43. Shao MY, Aburrous M, Huson D, Parraman C, Hardeberg JY, Clark J. Development and validation of a hybrid simulator for ultrasound-guided laparoscopic common bile duct exploration. Surg Endosc 2023 Sep;37(9):6943-6953 [FREE Full text] [doi: 10.1007/s00464-023-10168-w] [Medline: 37328593]
44. Preetha CJ, Kloss J, Wehrtmann FS, Sharan L, Fan C, Müller-Stich B, et al. Towards augmented reality-based suturing in monocular laparoscopic training. In: Proceedings of the 2020 Conference on Image-Guided Procedures, Robotic Interventions, and Modeling. 2020 Presented at: IGPRIM '20; February 15-20, 2020; Houston, TX p. 5-28 URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11315/113150X/Towards-augmented-reality-based-suturing-in-monocular-laparoscopic-training/10.1117/12.2550830.short [doi: 10.1117/12.2550830]
45. Lahanas V, Loukas C, Nikiteas N, Dimitroulis D, Georgiou E. Psychomotor skills assessment in laparoscopic surgery using augmented reality scenarios. In: Proceedings of 17th International Conference on Digital Signal Processing. 2011 Presented at: ICDSP '11; July 6-8, 2011; Corfu, Greece p. 1-6 URL: https://ieeexplore.ieee.org/document/6004893 [doi: 10.1109/icdsp.2011.6004893]
46. Rozenblit JW, Feng C, Riojas M, Napalkova L, Hamilton AJ, Hong M, et al. The computer assisted surgical trainer: design, models, and implementation. In: Proceedings of the 2014 Conference on Summer Simulation Multiconference. 2014 Presented at: SummerSim '14; July 6-10, 2014; Monterey, CA p. 1 URL: https://dl.acm.org/doi/10.5555/2685617.2685647 [doi: 10.5555/2685617.2685647]
47. Shabir D, Balakrishnan S, Padhan J, Abinahed J, Yaacoub E, Mohammed A, et al. Tele-mentoring using augmented reality: a feasibility study to assess teaching of laparoscopic suturing skills. In: Proceedings of the 36th International Symposium on Computer-Based Medical Systems. 2023 Presented at: CBMS '23; June 22-24, 2023; L'Aquila, Italy p. 61-66 URL: https://ieeexplore.ieee.org/document/10178719 [doi: 10.1109/cbms58004.2023.00193]
48. Condino S, Sannino S, Cutolo F, Giannini A, Simoncini T, Ferrari V. Single feature constrained manual registration method for augmented reality applications in gynecological laparoscopic interventions. Annu Int Conf IEEE Eng Med Biol Soc 2022 Jul;2022:566-571. [doi: 10.1109/EMBC48229.2022.9871263] [Medline: 36086356]
49. Fusaglia M, Gavaghan K, Beldi G, Volonté F, Pugin F, Peterhans M, et al. Endoscopic image overlay for the targeting of hidden anatomy in laparoscopic visceral surgery. In: Proceedings of the 7th International Workshop, AE-CAI 2012, Held in Conjunction with MICCAI 2012 on Augmented Environments for Computer-Assisted Interventions. 2012 Presented at: AE-CAI '12; October 5, 2012; Nice, France p. 9-21 URL: https://link.springer.com/chapter/10.1007/978-3-642-38085-3_3 [doi: 10.1007/978-3-642-38085-3_3]
50. Koehl C, Soler L, Marescaux J. PACS-based interface for 3D anatomical structure visualization and surgical planning. In: Proceedings of the International Conference International Optical Design. 2002 Presented at: CIOD '02; May 17, 2002; San Diego, CA p. 17-24 URL: https://link.springer.com/chapter/10.1007/978-3-642-56168-9_238 [doi: 10.1117/12.466924]
51. Viglialoro RM, Condino S, Gesi M, Ferrari M, Ferrari V. Augmented reality simulator for laparoscopic cholecystectomy training. In: Proceedings of the 1st International Conference on Augmented and Virtual Reality. 2014 Presented at: AVR '14; September 17-20, 2014; Lecce, Italy p. 433 URL: https://link.springer.com/chapter/10.1007/978-3-319-13969-2_33 [doi: 10.1007/978-3-319-13969-2_33]
52. Ivanova V, Boneva A, Vasilev P, Ivanov S, Lekova S. Augmented reality based training of surgical staff to operate a laparoscopic instrument. In: Proceedings of the 2021 Conference on Big Data, Knowledge and Control Systems Engineering. 2021 Presented at: BdKCSE '21; October 28-29, 2021; Sofia, Bulgaria p. 7 URL: https://ieeexplore.ieee.org/document/9627307 [doi: 10.1109/bdkcse53180.2021.9627307]
53. Wagner A, Rozenblit JW. Augmented reality visual guidance for spatial perception in the computer assisted surgical trainer. In: Proceedings of the 2017 Conference on Symposium on Modeling and Simulation in Medicine. 2017 Presented at: MSM '17; April 23-26, 2017; Virginia, NV p. 1-12 URL: https://dl.acm.org/doi/10.5555/3108760.3108765 [doi: 10.22360/springsim.2017.msm.009]

XSL•FO
RenderX

54. Larrarte EA, Alban AV. Virtual markers in virtual laparoscopy surgery. In: Proceedings of the 2016 XXI Symposium on Signal Processing, Images and Artificial Vision. 2016 Presented at: STSIVA '16; November 17, 2016; Bucaramanga, Colombia p. 1 URL: https://ieeexplore.ieee.org/document/7743367 [doi: 10.1109/stsiva.2016.7743367]
55. Doughty M, Singh K, Ghugre NR. SurgeonAssist-Net: towards context-aware head-mounted display-based augmented reality for surgical guidance. In: Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. 2021 Presented at: MICCAI '21; September 27-October 1, 2021; Strasbourg, France p. 667-677 URL: https://link.springer.com/chapter/10.1007/978-3-030-87202-1_64 [doi: 10.1007/978-3-030-87202-1_64]
56. Rewkowski N, State A, Fuchs H. Small marker tracking with low-cost, unsynchronized, movable consumer cameras for augmented reality surgical training. In: Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct. 2020 Presented at: ISMAR '20; November 9-13, 2020; Recife, Brazil p. 90-95 URL: https://ieeexplore.ieee.org/document/9288414 [doi: 10.1109/ISMAR-Adjunct51615.2020.00038]
57. Gupta A, Cecil J, Pirela-Cruz M, Kennison S, Hicks NA. Need for human extended reality interaction (HXRI) framework for the design of extended reality-based training environments for surgical contexts. In: Proceedings of the 10th International Conference on Serious Games and Applications for Health. 2022 Presented at: SeGAH '22; August 10-12, 2022; Sydney, Australia p. 1-8 URL: https://ieeexplore.ieee.org/document/9978563 [doi: 10.1109/segah54908.2022.9978563]
58. Vallino J. Haptics in augmented reality. In: Proceedings IEEE International Conference on Multimedia Computing and Systems. 1999 Presented at: MMCS '99; June 7-11, 1999; Florence, Italy p. 195-200 URL: https://ieeexplore.ieee.org/document/779146 [doi: 10.1109/mmcs.1999.779146]
59. Fan X, Zhu Q, Tu P, Joskowicz L, Chen X. A review of advances in image-guided orthopedic surgery. Phys Med Biol 2023 Jan 05;68(2):02TR01. [doi: 10.1088/1361-6560/acaae9] [Medline: 36595258]
60. Zattoni F, Carletti F, Randazzo G, Tuminello A, Betto G, Novara G, et al. Potential applications of new headsets for virtual and augmented reality in urology. Eur Urol Focus 2024 Jul;10(4):594-598 [FREE Full text] [doi: 10.1016/j.euf.2023.12.003] [Medline: 38160172]
61. Chowriappa A, Raza SJ, Fazili A, Field E, Malito C, Samarasekera D, et al. Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: a multi-institutional randomised controlled trial. BJU Int 2014 Aug 16;115(2):336-345. [doi: 10.1111/bju.12704]
62. Kalavakonda N, Chandra S, Thondiyath A. Development of virtual reality based robotic surgical trainer for patient-specific deformable anatomy. In: Proceedings of the 2015 Conference on Advances in Robotics. 2015 Presented at: AIR '15; July 2-4, 2015; Goa, India p. 1-5 URL: https://dl.acm.org/doi/10.1145/2783449.2783465 [doi: 10.1145/2783449.2783465]

Abbreviations

AR: augmented reality
ARV: augmented reality video
GOALS: Global Operative Assessment of Laparoscopic Skills
MR: mixed reality
OSATS: Objective Structured Assessment of Technical Skills
OST: optical see-through
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RQ: research question

Edited by A Coristine; submitted 06.03.24; peer-reviewed by F Alvarez-Lopez, R Yin, S Laplante; comments to author 05.07.24; revised version received 19.07.24; accepted 31.10.24; published 28.01.25.

Please cite as:
Celdrán FJ, Jiménez-Ruescas J, Lobato C, Salazar L, Sánchez-Margallo JA, Sánchez-Margallo FM, González P
Use of Augmented Reality for Training Assistance in Laparoscopic Surgery: Scoping Literature Review
J Med Internet Res 2025;27:e58108
URL: https://www.jmir.org/2025/1/e58108
doi:10.2196/58108
PMID:

©Francisco Javier Celdrán, Javier Jiménez-Ruescas, Carlos Lobato, Lucía Salazar, Juan Alberto Sánchez-Margallo, Francisco M Sánchez-Margallo, Pascual González. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Virtual Reality Interventions and Chronic Pain: Scoping Review

Michael E Ding[1], MD; Hajar Traiba[1], MD; Hector R Perez[1], MD, MS

Department of Medicine, Division of General Internal Medicine, Montefiore Medical Center, Bronx, NY, United States

Corresponding Author:
Hector R Perez, MD, MS
Department of Medicine
Division of General Internal Medicine
Montefiore Medical Center
3300 Kossuth Avenue
Bronx, NY
United States
Phone: 1 7189205756
Email: heperez@montefiore.org

Abstract

Background: Virtual reality (VR) interventions have demonstrated efficacy for more than a decade for mood and anxiety disorders and emerging evidence suggests they can reduce pain symptoms in both acute and chronic pain. More recently, these interventions have abounded within the commercial and academic sectors, immersing participants within a virtual environment to confer health benefits to users. VR immersion can facilitate the delivery of health interventions by isolating participants from distractors and stressors in a therapeutic environment. While recent studies of VR interventions have exploded, they are not uniform in approach or device type, limiting generalizability. Recent scoping reviews on VR and chronic pain have focused on specific diseases or limited inquiries to specific interventions or study types.

Objective: We conducted a scoping review to generate new knowledge about the sum total of VR studies on chronic pain with specific emphasis on the methods and results of each study, including (1) the type of interventions, (2) outcomes chosen, (3) samples studied, and (4) data generated.

Methods: A scoping review was performed on the literature on VR and chronic pain to describe themes associated with the literature to date and identify important gaps and unanswered questions to guide future research. CINAHL [EBSCO] (Cumulative Index to Nursing and Allied Health Literature) and PubMed were queried for the terms “virtual reality” and “pain,” providing studies of chronic pain adult participants using VR delivered through headset displays. We included English-language manuscripts that had at least one VR intervention arm with adults with chronic pain. For this analysis, we only included VR interventions that were immersive (ie, using headsets). Non–study reports, studies with no specific chronic pain component, those not involving adults, and those using VR as part of a comprehensive rehabilitation program were excluded. A descriptive analytical method was used to extract data, compare studies, and contextualize the presented outcomes. Articles were categorized into several themes including the type of intervention, outcomes chosen, participant characteristics, degree to which immersion was achieved, and adverse effect monitoring and reporting.

Results: A total of 36 articles were included in our analysis. We summarize the literature using 5 themes: (1) heterogeneity of chronic pain types, (2) highly variable intervention types, (3) highly variable secondary and exploratory outcomes, (4) immersion was highly variable between studies and not systemically explored in many articles, and (5) side effect monitoring was limited.

Conclusions: The literature on VR in chronic pain is highly variable and lacks theoretical rigor. While there is emerging evidence that supports VR use in a wide variety of health conditions including chronic pain, future research should focus on producing theoretically rigorous work that focuses on mechanisms and that systematically assesses side effects to generate robust generalizable knowledge.

(J Med Internet Res 2025;27:e59922) doi:10.2196/59922

KEYWORDS

virtual reality; chronic pain; scoping review; pain management; efficacy; anxiety disorders; mood; health condition; health intervention; adults; aging; therapeutic; descriptive–analytical method; monitoring; US; PRISMA

XSL•FO
RenderX

Introduction

Chronic pain affects 50 million US adults [1], and treatments have limited long-term efficacy, significant side effects, or limited accessibility [2-5]. As a result of the continuing opioid overdose crisis, numerous clinical and research guidelines have recommended further study of nonpharmacological options [6-8].

Virtual reality (VR) treatments have the potential to immerse participants within a virtual environment to confer health benefits to users. VR typically uses a headset to display a 3D image or video that participants can watch or interact with. Immersion produced through VR can facilitate the delivery of health interventions by isolating participants from distractors and stressors in a therapeutic environment [9]. VR interventions have demonstrated efficacy for more than a decade for mood and anxiety disorders [10,11], and emerging evidence suggests they can reduce pain symptoms in both acute and chronic pain [12-15].

Studies of VR interventions have exploded in recent years, due to advances in technology and the widespread availability of consumer-facing VR devices. Unfortunately, studies to date of VR in chronic pain have not been uniform in approach or device type, limiting generalizability. Recent scoping reviews on VR and chronic pain have focused on specific diseases or limited inquiries to specific interventions or study types [13,15]. We conducted a scoping review specifically to generate new knowledge about the sum total of VR studies on chronic pain with specific emphasis on the methods and results of each study, including (1) the type of interventions, (2) outcomes chosen, (3) samples studied, and (4) data generated. We expect the results of this scoping review will assist in generating new hypotheses and provide new areas of research inquiry for chronic pain VR researchers.

Methods

We conducted a scoping review of the literature on VR and chronic pain to describe themes associated with the literature to date and identify important gaps and unanswered questions to guide future research.

Eligibility Criteria

To develop the eligibility criteria, we conducted an initial search of the literature on PubMed and CINAHL and reviewed the first 100 articles that had initially resulted from general search terms of “virtual reality” and “pain.” As a first step, reviewers (HP, MD, and HT), independently screened these articles using Covidence software, followed by a meeting to discuss workflow and conflicts. Consensus was made regarding inclusion and exclusion criteria, which will be discussed below.

We decided to include English-language manuscripts and abstracts if at least one intervention arm included patients with chronic pain (even if compared with patients without chronic pain). We included patients with chronic pain of any type, even rare chronic diseases (eg, phantom limb syndrome). Because we noted some manuscripts defined VR as interchangeable with “interactive game” and because we only wished to explore studies about immersive VR, we only included studies of adult participants that used VR delivered through headset displays, not through other modalities (such as projection, computer screens, or television screens). We excluded nonstudy reports (letters, commentaries, and reviews), studies with patients with no specific chronic pain component (eg, acute burns), studies not involving adult patients, and studies using VR as part of a comprehensive rehabilitation program (eg, VR gait training). We did not focus on a particular set of outcomes of interest, and instead aimed to describe relevant outcomes in this research. We included research studies, case studies, and qualitative studies that met our inclusion criteria.

Information Sources

We performed a literature search of 2 electronic databases, CINAHL and PubMed between April 1, 2006, and March 21, 2022. These 2 comprehensive databases reference the nursing, allied health, and medical literature. Search terms included “Virtual reality” and “pain” to be broad and included all publications from the inception of both databases. We elected not to use MeSH (Medical Subject Headings) terms in searches to be sufficiently broad in our initial search. Results were prescreened using Covidence software (Melbourne, Victoria, Australia), with duplicate articles removed. The databases and platforms are outlined in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) diagram (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flowchart.

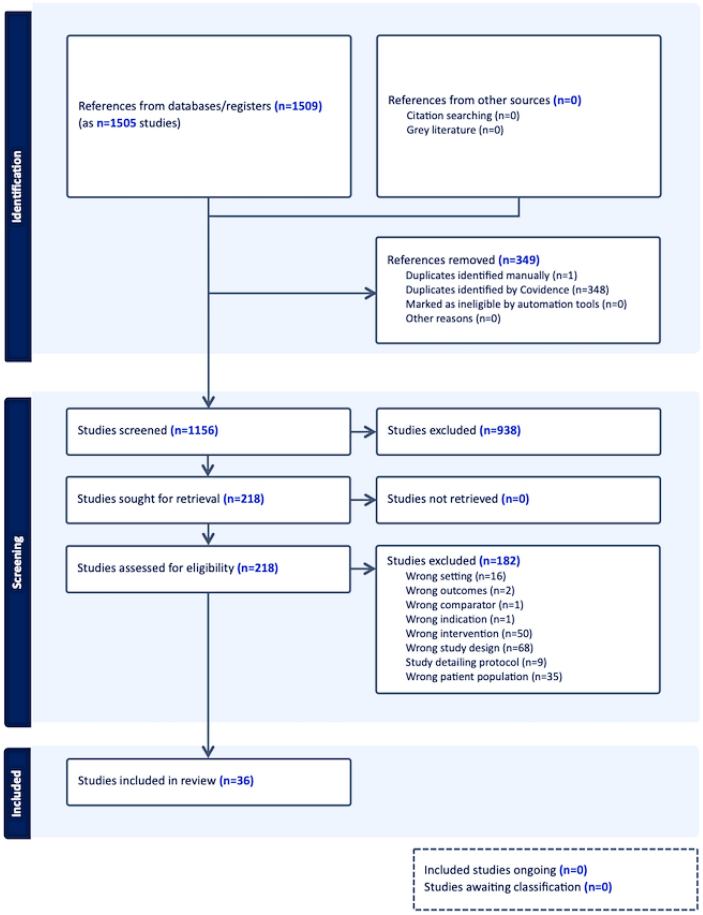

Selection of Sources of Evidence

After an initial prescreen, all citations and abstracts were screened by at least 2 independent reviewers (HP, MD, and HT). Abstracts that passed initial screening were included in a full-text eligibility review conducted by the same reviewers as above, with disagreements resolved by the senior author (HP).

Data Coding and Synthesis of Results

We used a data abstraction form our team designed to extract data using Covidence software. The major categories on the data abstraction form included are (1) study characteristics (year of publication, authors, and country of origin), (2) research setting (population, study design, and research problem), (3) intervention (detailed description of intervention type and immersion quality), and (4) results and conclusions. Two reviewers (MD and HT) and 1 senior reviewer (HP) extracted data; HP resolved discrepancies. Subsequently, we used a descriptive-analytical method to synthesize data, compare studies, and contextualize the presented outcomes [16]. In our review, we categorized articles into several themes to illustrate the diversity of the literature. We focused on some specific elements a priori and allowed for additional themes to emerge upon our reading of the literature. Specific elements categorized included are (1) the type of intervention, (2) outcomes chosen, (3) participant characteristics, (4) the degree to which immersion was achieved, and (5) adverse event monitoring and reporting.

Results

Overview

After removing duplicate citations, we screened 1156 studies using titles and abstracts, and retrieved 218 articles for full-text screens for eligibility. We included 36 articles in the final sample [17-52]. About two-thirds (n=23, 63.9%) were completed in the United States [17,20-24,28,29,32-35,37-41,43,48-52]. About one-fifth (n=7, 19.4%) were randomized controlled trials [22-24,27,30,45,46], 6 (16.7%) were randomized cross-over studies [18,19,31,36,40,41], 16 (44.4%) were single-arm case studies, case series, or cohort studies [17,20,21,26,28,29,32-35,37-39,43,49,52], 6 (16.7%) were qualitative, mixed methods, or included some interview component [19,25,26,33,37,44], 4 (11.1%) were nonrandomized cross-over studies [42,47,50,51], and 1 (2.8%) was a nonrandomized controlled study [48]. We present summaries of all articles in Multimedia Appendix 1.

In our review, we categorized articles into several themes to illustrate the diversity of the literature. We focus here on 5 themes for which we present narrative reviews of pertinent articles. First, while individual articles tended to focus on a particular pain type, conclusions from the literature are limited by the heterogeneity of chronic pain types. Second, intervention types were highly variable. Third, while most articles focused on pain outcomes, secondary and exploratory outcomes were highly variable. Fourth, the degree to which immersion was achieved was highly variable and not systematically explored

XSL•FO
RenderX

in many articles. Fifth, side effect monitoring was limited in most research.

Pain Types

While we expected to see variability in pain diagnoses because of our broad search terms, we found that the heterogeneity of pain types in the research to date limited external generalizability. In our literature search, 10 out of 36 (27.8%) articles focused on chronic pain, not otherwise specified [17,20,26,31,32,34,37,47,50,51]; 9 out of 36 (25%) focused on chronic back pain [22-24,28,29,35,44,45,49]; and 6 out of 36 (16.7%) focused on chronic regional pain syndrome [21,30,36,41,42,52]. Neuropathic pain was the focus of an additional 5 out of 36 (13.9%) articles, [18,36,38,40,48] but 3 were in the context of spinal cord injury [18,40,48]. Two out of 36 (5.6%) articles recruited patients with fibromyalgia [22,27]. Common pain types such as lower extremity osteoarthritis, diabetic neuropathy, abdominal pain, or noncardiac chest pain are absent from the primary literature, though it is possible many participants included in studies had multiple chronic pain types. There were no articles focusing on chronic overlapping chronic pain conditions, such as endometriosis, painful bladder syndrome, chronic fatigue syndrome, and irritable bowel syndrome.

Intervention Types

To understand the breadth of literature, we included articles that ran the spectrum of VR intervention design. We found VR interventions were highly heterogeneous. The plurality (11, 30.6% studies) were on VR behavioral therapy interventions [17,22-26,34,35,38,43,45]. Among these 11 studies, behavioral therapies included related but disparate modalities, including mindfulness techniques (8 out of 11) [17,22-26,34,35] pain education (6 out of 11) [17,22-24,35,45], and hypnosis (2 out of 11) [38,43]. Another 10 (27.8%) studies immersed participants in noninteractive passive environments, such as landscapes or walks [18,19,26,32,33,40,46,47,50,51]. An additional 6 (16.7%) used interactive, and sometimes strenuous, games in VR to attempt to reduce chronic pain [27-29,31,44,45]. VR body swapping or body observation was common in studies of complex regional pain syndrome or neuropathic pain, and constituted an additional 7 (19.4%) studies [21,30,36,41,42,49,52]. VR neurofeedback and "commercially available" VR applications constituted categories with 1 (2.8%) study apiece [20,39].

Outcomes

Almost all non-qualitative studies used pain scales, either numerical rating scales or visual analog scales, as outcome measures (30/32, 96.9% studies). Pain outcomes were not primary outcome measures in 4 studies, all of which were either primarily feasibility or usability studies [28,33,37,47]. Other outcomes varied widely. Changes in mood were assessed in 9 articles [17,18,22-24,33,39,43,52], and changes in sleep were assessed in 5 articles [22-24,39,52]. Changes in kinesophobia or changes in fear avoidance were assessed in 5 articles [27-29,45,46]. Physiological changes, such as heart rate variability or skin temperature, were assessed in a further 3 articles [42,50,51]. Physical activity, quality of life, and opioid use were assessed as outcomes in 2 [27,52], 1 [27], and 1 article, respectively [17].

Immersion

Immersion, or the tendency for a participant to feel the illusion of reality within a VR simulation, was uncommonly assessed. Only 3 studies systematically assessed immersion in the VR world [17,40,45]. Presence, or the tendency to feel one is within a virtual environment, is often used interchangeably with immersion and was assessed in 3 articles [18,19,40]. An additional 2 studies assessed engagement in VR, a related concept, but 1 which relates more to adherence and satisfaction than immersion [20,32].

Side Effect Monitoring

Despite common reports of side effects in commercial VR devices, side effects were variably assessed in studies. There was no reported assessment of side effects in 21 (58.3%) studies [17,18,21,27-31,34-36,38-46,48]. Among studies that systematically reported side effects, most were low frequency, low intensity, or not durable. Most side effects did not adversely affect participation, but 2 studies reported participants dropping out due to side effects (Figure 1) [26,31].

Discussion

Principal Findings

In this scoping review, we describe patterns in the current literature on VR interventions in chronic pain. We identified 36 articles that tested VR interventions through a variety of study designs. We reported 5 themes that encapsulate research to date: (1) pain types are highly heterogenous in the literature to date, (2) VR design was highly variable, (3) secondary and exploratory outcome choice was highly variable, (4) VR immersion was infrequently assessed, and (5) side effect monitoring was limited. These themes are somewhat related and have implications for conclusions to be generated from the research published to date. We provide some recommendations for future research studies.

Heterogeneity Between Studies and Comparison to Previous Work

Studies were heterogeneous in many aspects, including type of VR intervention, patient population, and study design. In the literature we reviewed, even the term VR was used inconsistently and was even used to refer to nonimmersive modalities and platforms such as computers and projector screens. In developing our search strategy, we found that "virtual reality" was not a universal term and had to exclude non immersive devices from our sample of articles; this heterogeneity of study design was also encountered by one other scoping review [53]. The patient population was additionally highly variable between studies. Another scoping review of seven studies on VR-based mindfulness programs for chronic pain noted differences in populations and VR interventions between studies [15]. Such differences pose challenges in making generalizable VR recommendations for patients with chronic pain, as different VR treatments may only ameliorate certain pain types. This in turn, if left unchecked, may negatively

XSL•FO
RenderX

affect the future generalizability and use of VR as a treatment for chronic pain.

There was significant heterogeneity in pain type among those receiving VR interventions. The existing literature on VR for the most common pain types is lacking. For example, it is difficult to find studies on previously mentioned common types such as lower extremity osteoarthritis, diabetic neuropathy, and abdominal pain. Instead, there have been many more studies on VR modalities focusing specifically on rarer types such as phantom limb pain and complex regional pain syndrome. This is likely because there already exist therapeutic elements backed by an evidence base (eg mirror therapy) in treating these painful conditions that can be directly incorporated into VR. Of note, VR behavioral management and VR therapies directly targeting kinesiophobia seem to have an emerging evidence base, stemming from their evidence toward treating relevant pain types in non-VR trials. For example, Lin and colleagues analyzed evidence for pain neuroscience education, a theoretical framework for educating and empowering the patient on his or her pain condition and management options [54]. They found evidence for pain neuroscience education in reducing pain and kinesiophobia in patients with chronic neck pain. Likewise, a recent review on cognitive behavioral therapy (CBT) and chronic pelvic pain has noted the efficacy of CBT in treating many conservative treatment-refractory pain types such as nociplastic pain disorders [55]. Research on employing CBT or CBT elements in VR interventions, and how best to do this without a live, skilled therapist, is a focus of current research, including our own ongoing clinical trials.

Ultimately, which VR interventions to offer which patients and when and how are important and significant questions for which current data is still insufficient. The insufficiency of current data to date was echoed in other recent scoping reviews of VR interventions [12,13,15,18,53,56]. A recent scoping review of 44 studies in chronic pain also concluded that while VR showed promise, the exploratory nature and heterogeneous design of most research limits generalized conclusions [13]. Further, another scoping review of 13 studies in back pain also found that research was highly variable and concluded the field was still nascent [12]. Future research should build on exploratory research conducted to date and should also include comparative effectiveness studies between VR intervention types to move the field forward. In addition, consistent, regular assessment of participant side effects is imperative to producing clear risk and benefit data for clinicians and patients in the future.

Relatedly, we found a lack of theoretical rigor in most studies of VR. Basing study design on theory can help elucidate mechanisms for how interventions work and can help researchers build more robust clinical trials. In particular, such a framework would include identifying and testing the various elements of the VR experience, isolating the aspects of VR that can lead to pain relief, and understanding whether there are mediation or moderation effects that influence the efficacy of VR interventions in chronic pain. For example, immersion is theorized to be a key target for how VR works on chronic pain [57-60]. Nonetheless, as we have shown, it was rarely systematically assessed. Different VR interventions can have markedly different effectiveness in generating immersion, owing to many factors: hardware weight and comfort, graphical fidelity, and the level of involvement and stimulation. Other important mediators, such as mood or fear of physical activity, are also infrequently assessed. Moderators, such as a participant's comfort with technology and video games are also rarely assessed formally. No strong theoretical model yet exists to explain how VR interventions might work in chronic pain [61]; it is important that future research aims to develop and empirically test theoretical models to ensure generalizable knowledge is produced. In addition, and specifically, future studies should measure and compare levels of immersion between various VR types.

Limitations

There were a few limitations to this scoping review. First, we searched 2 large databases, PubMed and CINAHL, for relevant articles because they constitute a significant portion of the clinical literature in chronic pain and we used a broad searching strategy with broad search terms, but it is possible we may have missed some manuscripts due to our search strategy. Future reviews can determine the utility of searching other comprehensive databases. Second, we included only English language articles in our initial search. This may have excluded data from other countries which may alter the conclusions drawn by our study. Future reviews can determine how non-English literature might differ from the English language articles included in this review. Third, we note that this article is unable to make a detailed examination of particular methodologies of VR interventions, partly because of a lack of data but also partly because of lack of comprehensive detail on VR methods in many of the articles reviewed. This is a limitation of the literature at large at this point, and it is very important for the reproducibility of research that there is transparency in research methods. Future VR interventional studies should include more descriptive details about the methodologies of the VR interventions used.

Conclusions

In summary, we found the literature on VR in chronic pain to be highly variable and lacking theoretical rigor. While there is emerging evidence that supports VR use in a wide variety of health conditions including chronic pain, future research should focus on producing theoretically rigorous work that focuses on mechanisms, and transparency in research methods and that systematically assesses side effects in order to generate robust generalizable knowledge.

Acknowledgments

HP acknowledges funding from the National Institutes of Health (RM1DA055437, R01DA060796, and K23DA044327). This played no role in how the research was conducted or the decision to submit it for publication.

XSL•FO
RenderX

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MD was responsible for formal analysis, investigation, writing (original draft), and writing (review and editing); HT was responsible for investigation and writing (review); HP was responsible for conceptualization, formal analysis, investigation, data curation, methodology, and project administration, supervision, writing (review and editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summaries of analyzed articles (n=36).
[PDF File (Adobe PDF File), 318 KB - jmir_v27i1e59922_app1.pdf]

Multimedia Appendix 2
Updated PRISMA ScR checklist.
[PDF File (Adobe PDF File), 305 KB - jmir_v27i1e59922_app2.pdf]

References

1. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. Pain 2022;163(2):e328-e332. [doi: 10.1097/j.pain.0000000000002291] [Medline: 33990113]
2. Bialas P, Maier C, Klose P, Häuser W. Efficacy and harms of long-term opioid therapy in chronic non-cancer pain: systematic review and meta-analysis of open-label extension trials with a study duration ≥26 weeks. Eur J Pain 2020;24(2):265-278. [doi: 10.1002/ejp.1496] [Medline: 31661587]
3. Nadeau SE, Wu JK, Lawhern RA. Opioids and chronic pain: an analytic review of the clinical evidence. Front Pain Res (Lausanne) 2021;2:721357 [FREE Full text] [doi: 10.3389/fpain.2021.721357] [Medline: 35295493]
4. Heyward J, Jones CM, Compton WM, Lin DH, Losby JL, Murimi IB, et al. Coverage of nonpharmacologic treatments for low back pain among US public and private insurers. JAMA Netw Open 2018;1(6):e183044 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3044] [Medline: 30646222]
5. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018;320(23):2448-2460 [FREE Full text] [doi: 10.1001/jama.2018.18472] [Medline: 30561481]
6. Bushnell MC, Frangos E, Madian N. Non-pharmacological treatment of pain: grand challenge and future opportunities. Front Pain Res (Lausanne) 2021;2:696783 [FREE Full text] [doi: 10.3389/fpain.2021.696783] [Medline: 35295445]
7. Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022;71(3):1-95 [FREE Full text] [doi: 10.15585/mmwr.rr7103a1] [Medline: 36327391]
8. The Management of Opioid Therapy for Chronic Pain Working Group. VA/DoD clinical practice guideline for management of opioid therapy for chronic pain. Department of Veterans Affairs, Department of Defense. 2017. URL: https://www.healthquality.va.gov/guidelines/Pain/cot/VADoDOTCPG022717.pdf [accessed 2025-01-09]
9. Arpaia P, D'Errico G, De Paolis LT, Moccaldi N, Nuccetelli F. A narrative review of mindfulness-based interventions using virtual reality. Mindfulness 2021;13(3):556-571. [doi: 10.1007/s12671-021-01783-6]
10. Cieślik B, Mazurek J, Rutkowski S, Kiper P, Turolla A, Szczepańska-Gieracha J. Virtual reality in psychiatric disorders: a systematic review of reviews. Complement Ther Med 2020;52:102480 [FREE Full text] [doi: 10.1016/j.ctim.2020.102480] [Medline: 32951730]
11. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. Harv Rev Psychiatry 2017;25(3):103-113 [FREE Full text] [doi: 10.1097/HRP.0000000000000138] [Medline: 28475502]
12. Nagpal AS, Raghunandan A, Tata F, Kibler D, McGeary D. Virtual reality in the management of chronic low back pain: a scoping review. Front Pain Res (Lausanne) 2022;3:856935 [FREE Full text] [doi: 10.3389/fpain.2022.856935] [Medline: 35295809]
13. Austin PD. The analgesic effects of virtual reality for people with chronic pain: a scoping review. Pain Med 2022;23(1):105-121. [doi: 10.1093/pm/pnab217] [Medline: 34260724]
14. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019;12:2053-2085 [FREE Full text] [doi: 10.2147/JPR.S200498] [Medline: 31308733]

XSL•FO
RenderX

15. O'Connor S, Mayne A, Hood B. Virtual reality-based mindfulness for chronic pain management: a scoping review. Pain Manag Nurs 2022;23(3):359-369 [FREE Full text] [doi: 10.1016/j.pmn.2022.03.013] [Medline: 35491349]
16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int. J. Soc. Res. Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
17. Abd-Elsayed A, Hussain N, Stanley G. Combining virtual reality and behavioral health to promote pain resiliency: analysis of a novel biopsychosocial modality for solving pain in the workplace. Pain Ther 2021;10(2):1731-1740 [FREE Full text] [doi: 10.1007/s40122-021-00333-1] [Medline: 34657235]
18. Austin PD, Craig A, Middleton JW, Tran Y, Costa DSJ, Wrigley PJ, et al. The short-term effects of head-mounted virtual-reality on neuropathic pain intensity in people with spinal cord injury pain: a randomised cross-over pilot study. Spinal Cord 2021;59(7):738-746. [doi: 10.1038/s41393-020-00569-2] [Medline: 33077900]
19. Austin PD, Siddall PJ, Lovell MR. Feasibility and acceptability of virtual reality for cancer pain in people receiving palliative care: a randomised cross-over study. Support Care Cancer 2022;30(5):3995-4005 [FREE Full text] [doi: 10.1007/s00520-022-06824-x] [Medline: 35064330]
20. Baker NA, Baird J, Driscoll M. Integrating virtual reality into a comprehensive chronic pain program. Pain Med 2022;23(7):1342-1343. [doi: 10.1093/pm/pnab170] [Medline: 34086922]
21. Chau B, Phelan I, Ta P, Chi B, Loyola K, Yeo E, et al. Immersive virtual reality for pain relief in upper limb complex regional pain syndrome: a pilot study. Innov Clin Neurosci 2020;17(4-6):47-52 [FREE Full text] [Medline: 32802594]
22. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-administered skills-based virtual reality intervention for chronic pain: randomized controlled pilot study. JMIR Form Res 2020;4(7):e17293 [FREE Full text] [doi: 10.2196/17293] [Medline: 32374272]
23. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
24. Garcia LM, Birckhead BJ, Krishnamurthy P, Mackey I, Sackman J, Salmasi V, et al. Three-month follow-up results of a double-blind, randomized placebo-controlled trial of 8-week self-administered at-home behavioral skills-based Virtual Reality (VR) for chronic low back pain. J Pain 2022;23(5):822-840 [FREE Full text] [doi: 10.1016/j.jpain.2021.12.002] [Medline: 34902548]
25. Garrett BM, Tao G, Taverner T, Cordingley E, Sun C. Patients perceptions of virtual reality therapy in the management of chronic cancer pain. Heliyon 2020;6(5):e03916 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03916] [Medline: 32426540]
26. Garrett B, Taverner T, McDade P. Virtual Reality as an Adjunct Home Therapy in Chronic Pain Management: An Exploratory Study. JMIR Med Inform 2017;5(2):e11 [FREE Full text] [doi: 10.2196/medinform.7271] [Medline: 28495661]
27. Gulsen C, Soke F, Eldemir K, Apaydin Y, Ozkul C, Guclu-Gunduz A, et al. Effect of fully immersive virtual reality treatment combined with exercise in fibromyalgia patients: a randomized controlled trial. Assist Technol 2022;34(3):256-263. [doi: 10.1080/10400435.2020.1772900] [Medline: 32543290]
28. Hennessy R, Rumble D, Sprigner C, Uzochukwu J, Christian M, David B, et al. (380) Preliminary feasibility of a graded, locomotor-enabled, whole-body virtual reality intervention for individuals with chronic low back pain. J. Pain 2019;20(4):S69. [doi: 10.1016/j.jpain.2019.02.076]
29. Hennessy RW, Rumble D, Christian M, Brown DA, Trost Z. A graded exposure, locomotion-enabled virtual reality app during walking and reaching for individuals with chronic low back pain: cohort gaming design. JMIR Serious Games 2020;8(3):e17799 [FREE Full text] [doi: 10.2196/17799] [Medline: 32773381]
30. Jeon B, Cho S, Lee J. Application of virtual body swapping to patients with complex regional pain syndrome: a pilot study. Cyberpsychol Behav Soc Netw 2014;17(6):366-370. [doi: 10.1089/cyber.2014.0046] [Medline: 24892199]
31. Jin W, Choo A, Gromala D, Shaw C, Squire P. A virtual reality game for chronic pain management: a randomized, controlled clinical study. Stud Health Technol Inform 2016;220:154-160. [Medline: 27046570]
32. Jones T, Moore T, Choo J. The impact of virtual reality on chronic pain. PLoS One 2016;11(12):e0167523 [FREE Full text] [doi: 10.1371/journal.pone.0167523] [Medline: 27997539]
33. Kelleher SA, Fisher HM, Winger JG, Miller SN, Amaden GH, Somers TJ, et al. Virtual reality for improving pain and pain-related symptoms in patients with advanced stage colorectal cancer: a pilot trial to test feasibility and acceptability. Palliat Support Care 2022;20(4):471-481 [FREE Full text] [doi: 10.1017/S1478951521002017] [Medline: 35078545]
34. Liu K, Madrigal E, Chung JS, Parekh M, Kalahar CS, Nguyen D, et al. Preliminary study of virtual-reality-guided meditation for veterans with stress and chronic pain. Altern Ther Health Med 2023;29(6):42-49. [Medline: 34559692]
35. Louw A, Louw C, Flynn T. Case report: pain neuroscience education using virtual reality. Pain and Rehabilitation - the Journal of Physiotherapy Pain Association 2019;2019(46):4-7 [FREE Full text]
36. Matamala-Gomez M, Diaz Gonzalez AM, Slater M, Sanchez-Vives MV. Decreasing pain ratings in chronic arm pain through changing a virtual body: different strategies for different pain types. J Pain 2019;20(6):685-697 [FREE Full text] [doi: 10.1016/j.jpain.2018.12.001] [Medline: 30562584]
37. Nakad L, Rakel B. Attitudes of older adults with chronic musculoskeletal pain toward virtual reality. Innovation in Aging 2020;4(Supplement_1):202-203 [FREE Full text]

XSL•FO
RenderX

38. Oneal BJ, Patterson DR, Soltani M, Teeley A, Jensen MP. Virtual reality hypnosis in the treatment of chronic neuropathic pain: a case report. Int J Clin Exp Hypn 2008;56(4):451-462 [FREE Full text] [doi: 10.1080/00207140802255534] [Medline: 18726807]
39. Orakpo N, Vieux U, Castro-Nuñez C. Case report: virtual reality neurofeedback therapy as a novel modality for sustained analgesia in centralized pain syndromes. Front Psychiatry 2021;12:660105 [FREE Full text] [doi: 10.3389/fpsyt.2021.660105] [Medline: 33959057]
40. Putrino D, Tabacof L, Breyman E, Revis J, Soomro Z, Chopra D, et al. Pain reduction after short exposure to virtual reality environments in people with spinal cord injury. Int J Environ Res Public Health 2021;18(17):8923 [FREE Full text] [doi: 10.3390/ijerph18178923] [Medline: 34501513]
41. Solcà M, Krishna V, Young N, Deogaonkar M, Herbelin B, Orepic P, et al. Enhancing analgesic spinal cord stimulation for chronic pain with personalized immersive virtual reality. Pain 2021;162(6):1641-1649. [doi: 10.1097/j.pain.0000000000002160] [Medline: 33259460]
42. Solcà M, Ronchi R, Bello-Ruiz J, Schmidlin T, Herbelin B, Luthi F, et al. Heartbeat-enhanced immersive virtual reality to treat complex regional pain syndrome. Neurology 2018;91(5):e479-e489. [doi: 10.1212/WNL.0000000000005905] [Medline: 29980635]
43. Soltani M, Teeley AM, Wiechman SA, Jensen MP, Sharar SR, Patterson DR. Virtual reality hypnosis for pain control in a patient with gluteal hidradenitis:a case report(). Contemp Hypn Integr Ther 2011;28(2):142-147 [FREE Full text] [Medline: 23205274]
44. Stamm O, Dahms R, Müller-Werdan U. Virtual reality in pain therapy: a requirements analysis for older adults with chronic back pain. J Neuroeng Rehabil 2020;17(1):129 [FREE Full text] [doi: 10.1186/s12984-020-00753-8] [Medline: 32993678]
45. Stamm O, Dahms R, Reithinger N, Ruß A, Müller-Werdan U. Virtual reality exergame for supplementing multimodal pain therapy in older adults with chronic back pain: a randomized controlled pilot study. Virtual Real 2022;26(4):1291-1305 [FREE Full text] [doi: 10.1007/s10055-022-00629-3] [Medline: 35194374]
46. Tejera DM, Beltran-Alacreu H, Cano-de-la-Cuerda R, Leon Hernández JV, Martín-Pintado-Zugasti A, Calvo-Lobo C, et al. Effects of virtual reality versus exercise on pain, functional, somatosensory and psychosocial outcomes in patients with non-specific chronic neck pain: a randomized clinical trial. Int J Environ Res Public Health 2020;17(16):5950 [FREE Full text] [doi: 10.3390/ijerph17165950] [Medline: 32824394]
47. Tong X, Gromala D, Gupta D, Squire P. Usability comparisons of head-mounted vs. stereoscopic desktop displays in a virtual reality environment with pain patients. Stud Health Technol Inform 2016;220:424-431. [Medline: 27046617]
48. Trost Z, Anam M, Seward J, Shum C, Rumble D, Sturgeon J, et al. Immersive interactive virtual walking reduces neuropathic pain in spinal cord injury: findings from a preliminary investigation of feasibility and clinical efficacy. Pain 2022;163(2):350-361. [doi: 10.1097/j.pain.0000000000002348] [Medline: 34407034]
49. Trujillo MS, Alvarez AF, Nguyen L, Petros J. Embodiment in virtual reality for the treatment of chronic low back pain: a case series. J Pain Res 2020;13:3131-3137 [FREE Full text] [doi: 10.2147/JPR.S275312] [Medline: 33269003]
50. Wiederhold BK, Gao K, Kong L, Wiederhold MD. Mobile devices as adjunctive pain management tools. Cyberpsychol Behav Soc Netw 2014;17(6):385-389 [FREE Full text] [doi: 10.1089/cyber.2014.0202] [Medline: 24892202]
51. Wiederhold BK, Gao K, Sulea C, Wiederhold MD. Virtual reality as a distraction technique in chronic pain patients. Cyberpsychol Behav Soc Netw 2014;17(6):346-352 [FREE Full text] [doi: 10.1089/cyber.2014.0207] [Medline: 24892196]
52. Won AS, Barreau AC, Gaertner M, Stone T, Zhu J, Wang CY, et al. Assessing the feasibility of an open-source virtual reality mirror visual feedback module for complex regional pain syndrome: pilot usability study. J Med Internet Res 2021;23(5):e16536 [FREE Full text] [doi: 10.2196/16536] [Medline: 34037530]
53. Carroll J, Hopper L, Farrelly AM, Lombard-Vance R, Bamidis PD, Konstantinidis EI. A scoping review of augmented/virtual reality health and wellbeing interventions for older adults: redefining immersive virtual reality. Front. Virtual Real 2021;2:655338. [doi: 10.3389/frvir.2021.655338]
54. Lin L, Lin T, Chang K, Wu W, Özçakar L. Pain neuroscience education for reducing pain and kinesiophobia in patients with chronic neck pain: a systematic review and meta-analysis of randomized controlled trials. Eur J Pain 2024;28(2):231-243. [doi: 10.1002/ejp.2182] [Medline: 37694895]
55. Lackner JM, Clemens JQ, Radziwon C, Danforth TL, Ablove TS, Krasner SS, et al. Cognitive behavioral therapy for chronic pelvic pain: what is it and does it work? J Urol 2024;211(4):539-550. [doi: 10.1097/JU.0000000000003847] [Medline: 38228093]
56. Lin H, Li Y, Hu W, Huang C, Du Y. A scoping review of the efficacy of virtual reality and exergaming on patients of musculoskeletal system disorder. J Clin Med 2019;8(6):791 [FREE Full text] [doi: 10.3390/jcm8060791] [Medline: 31167435]
57. Triberti S, Repetto C, Riva G. Psychological factors influencing the effectiveness of virtual reality-based analgesia: a systematic review. Cyberpsychol Behav Soc Netw 2014;17(6):335-345. [doi: 10.1089/cyber.2014.0054] [Medline: 24892195]
58. Pourmand A, Davis S, Marchak A, Whiteside T, Sikka N. Virtual reality as a clinical tool for pain management. Curr Pain Headache Rep 2018;22(8):53. [doi: 10.1007/s11916-018-0708-2] [Medline: 29904806]
59. Li H, Zhang X, Wang H, Yang Z, Liu H, Cao Y, et al. Access to nature via virtual reality: a mini-review. Front Psychol 2021;12:725288 [FREE Full text] [doi: 10.3389/fpsyg.2021.725288] [Medline: 34675840]

XSL•FO
RenderX

60. Dumoulin S, Bouchard S, Loranger C, Quintana P, Gougeon V, Lavoie KL. Are cognitive load and focus of attention differentially involved in pain management: an experimental study using a cold pressor test and virtual reality. J Pain Res 2020;13:2213-2222 [FREE Full text] [doi: 10.2147/JPR.S238766] [Medline: 32943914]
61. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. JMIR Serious Games 2018;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]

Abbreviations

CBT: cognitive behavioral therapy
MeSH: Medical Subject Headings
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
VR: Virtual reality

Edited by A Coristine; submitted 11.05.24; peer-reviewed by U Sinha, J Sun; comments to author 05.06.24; revised version received 15.09.24; accepted 09.12.24; published 18.02.25.

Please cite as:
Ding ME, Traiba H, Perez HR
Virtual Reality Interventions and Chronic Pain: Scoping Review
J Med Internet Res 2025;27:e59922
URL: https://www.jmir.org/2025/1/e59922
doi:10.2196/59922
PMID:

©Michael E Ding, Hajar Traiba, Hector R Perez. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Current Clinical and Educational Uses of Immersive Reality in Anesthesia: Narrative Review

Andrew Fleet[1], PhD; Lilia Kaustov[1], MSc, PhD; Elio BR Belfiore[1], MD; Bill Kapralos[2], MSc, PhD; Clyde Matava[3,4], MBChB, MHSc, MMed; Julian Wiegelmann[1,4], MHI, MD; Peter Giacobbe[5,6], MSc, MD; Fahad Alam[1,4], MHSc, MD

[1]Department of Anesthesia, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[2]maxSIMhealth Group, Ontario Tech University, Oshawa, ON, Canada

[3]Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, Toronto, ON, Canada

[4]Department of Anesthesiology & Pain Medicine, University of Toronto, Toronto, ON, Canada

[5]Department of Psychiatry, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, Canada

[6]Harquail Centre for Neuromodulation, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

Corresponding Author:
Fahad Alam, MHSc, MD
Department of Anesthesia
Sunnybrook Health Sciences Centre
Room M3-200
2075 Bayview Avenue
Toronto, ON, M4N 3M5
Canada
Phone: 1 416 480 4864
Email: fahad.alam@sunnybrook.ca

Abstract

Background: The concept of immersive reality (IR), an umbrella term that encompasses virtual reality, augmented reality, and mixed reality, has been established within the health care realm as a potentially valuable tool with numerous applications in both medical education and patient care.

Objective: This review aimed to introduce anesthesiologists to the emerging and rapidly evolving literature on IR, its use in anesthesia education, and its transferability into the clinical context.

Methods: A review of the relevant literature was conducted using the PubMed database from inception to July 5, 2023. Additional references were identified from the reference lists of selected papers.

Results: A total of 51 papers related to the use of IR in anesthesia medical education (including both technical and nontechnical skills) and 63 papers related to applications in clinical practice (eg, preprocedure planning, patient education, and pain management) were included. We present evidence supporting the use of IR in the training and clinical practice of modern anesthesiologists.

Conclusions: IR is useful for a variety of applications in anesthesia medical education and has potential advantages over existing simulation approaches. Similarly, IR has demonstrated potential improvements in patient care across several clinical contexts relevant to practicing anesthesiologists. However, many applications remain in the early stages of development, and robust trials are urgently needed to confirm clinical or educational effectiveness and to assess mechanisms, educational validity, and cost-effectiveness.

(J Med Internet Res 2025;27:e62785) doi:10.2196/62785

KEYWORDS

virtual reality; augmented reality; mixed reality; anesthesia; immersive reality; medical education; artificial intelligence

XSL•FO
RenderX

Introduction

Nomenclature and Definitions

The nomenclature surrounding immersive reality (IR) has evolved, and in this review, we provide some definitions to assist the reader. We use IR as an overarching term encompassing virtual reality (VR), augmented reality (AR), and mixed reality (MR), although other umbrella terms, such as extended reality, can also be found in the literature [1,2]. VR is any type of computer-generated immersive environment that users can interact with, including both 2D and 3D experiences. AR involves overlaying computer-generated elements on a real-world environment, and MR blends aspects of VR and AR [3,4], capturing both real-world and virtual elements in 3D and allowing for interaction across these realms [5].

IR experiences are often facilitated using a head-mounted display (HMD or headset), a piece of hardware worn on the face of the user, surrounding their field of view, with 1 or 2 small displays showing stereoscopic computer-generated imagery [6]. In this form of VR, users wearing an HMD can only perceive images of the virtually generated world displayed on the screens. This allows a user to experience a fully immersive, computer-generated 3D environment and interact with objects using handheld controllers or a set of virtual hands that mimic their own. More recently, with advances in consumer-level hardware, many IR headsets allow users to physically walk around within a defined interactive VR environment. In contrast to VR headsets, AR and MR headsets allow the user to see the real world while immersed in an overlaid computer-generated environment within the field of view of the headset.

Consumer-level HMDs most commonly used for clinical and medical education purposes incorporate variable levels of functionality (Table 1). For example, different devices are capable of tracking various types of motion. Simple systems with 3 degrees of freedom can only detect rotation around a fixed point. Newer, more complex systems, with 6 degrees of freedom can detect rotation and position in 3D space.

Regarding HMDs, 3 degrees of freedom devices can only detect where you are looking (the rotation and tilt of your head) but not your position in space (ie, whether you move forward or backward, left or right, or crouch down). Devices with 6 degrees of freedom can detect all the abovementioned types of motion using either sensors built into the device itself or external sensors, such as tripod or wall-mounted hardware.

XSL•FO
RenderX

Table 1. Technical characteristics and practical considerations for commonly available immersive reality head-mounted displays (HMDs) used in anesthesia.

Device	Commercial availability	Additional component requirement	HMD degrees of freedom	Eye tracking	Hand tracking	Controller degrees of freedom	Example uses in anesthesia
Oculus Rift	Discontinued	Yes, computer desktop	6 (requires external sensors)	No	No	6	• VR[a] distraction [7,8] • Preprocedure education [9] • Immersive VR simulation [10-12]
Oculus Go	Discontinued	No	3	No	No	3	• VR distraction [13-15] • Preprocedure education [16] • Immersive VR simulation [17]
Oculus Quest	Discontinued	No	6	No	Yes	6	• N/A[b]
Oculus Quest 2	Yes	No	6	Yes, with third-party hardware	Yes	6	• N/A
Gear VR (Samsung)	Discontinued	No	3	No	No	3	• VR distraction [7,18-21] • Preprocedure relaxation [22]
Pico G2 4K	Yes	Yes, computer desktop	3	No	No	3	• VR distraction [23]
Pico Neo 3 Pro	Yes	Yes, computer desktop	6	Yes, with additional cost	No	6	• N/A
HTC Vive Pro	Discontinued	Yes, attachments	6	Yes, with third-party hardware	No	6	• Immersive VR simulation [17,24-26]
HTC Vive Pro 2	Yes	Yes, attachments	6	Yes, with third-party hardware	No	6	• —[c]
Valve Index	Yes	Yes, computer desktop	6	Yes, with 3rd party hardware	Yes, using controller	6	• N/A
Prism Pro (Mira)	Yes	Yes, mobile device	N/A	No	No	3	• N/A
PlayStation VR (Sony)	Yes	Yes, PlayStation 4 or PlayStation 5 console	6 (requires external sensors)	No	No	6	• VR distraction game [27]
Google Daydream	Discontinued	Yes, mobile device	3	No	No	3	• N/A
Hololens (Microsoft)	Discontinued	No	6	Yes	Yes	3	• AR[d] simulation [28]
Hololens 2 (Microsoft)	Yes	No	6	Yes	Yes	No controller	• N/A
Magic Leap 1	Discontinued	No	6	Yes	Yes	6	• AR simulation [29,30]
Magic Leap 2	Yes	No	6	Yes	Yes	6	• N/A
Apple Vision Pro	Yes	No	6	Yes	Yes	No controller	• N/A

[a]VR: virtual reality.

[b]N/A: not applicable.

[c]Data not available.

[d]AR: augmented reality.

XSL•FO
RenderX

Similar to headsets, VR controllers with 3 degrees of freedom have limited tracking ability, while newer controllers with 6 degrees of freedom allow a more naturalistic, full range of motion (eg, extending your controller in real life will be replicated in VR). Eye tracking refers to an extra set of lenses within the HMD, which tracks the movement of each eye. This allows the device to track where the user is looking and, in some cases, it allows interaction with the virtual environment through changes in gaze alone [31-33]. Hand tracking uses a front-facing camera on the HMD to track a user's hands and project them into the virtual environment, allowing users to interact with the virtual environment without the need for controllers.

Handheld devices (eg, cell phones and tablets) can also be used for AR, by observing digital images overlaid on the real world using the camera viewer. However, these devices offer limited functionality as they have only a small field of view and are not ergonomic, as they must be held up by the user.

The Spectrum of Immersion

Immersion is the ability of a system to create the sensation of "presence," the subjective feeling of existing in a computer-generated world [34]. This is created in computer-based simulations by providing sensory input, including visual, auditory, and tactile (ie, haptic feedback) stimuli. Placing an individual in an IR system with these stimuli creates the sensation of being physically present in a virtual environment. Advancements in computational processing power and associated decreases in the size of electronic components have led to lower costs and rising availability of consumer-level IR technologies. This increase in availability has been accompanied by a parallel increase in the adoption of virtual-based simulation in medicine and beyond [35].

Historically, 2D worlds, which lack the increased sense of presence provided by newer 3D immersion technology, have also been labeled immersive experiences. In the context of medicine, this has mainly been limited to medical education, where learners view a 2D screen with a virtual image of a simulated task to be practiced. Many examples of simple 2D simulation exist in surgical training [36,37] and anesthesia [38-42], where students view virtual anatomical images on a screen and use partial task trainers or practical tools (eg, endoscopes, bronchoscopes, and surgical instruments), which provide haptic feedback to practice their technical skills.

Affordable HMDs and 360° cameras developed in recent years have allowed educators to increase immersion in IR educational experiences. Instead of watching on a 2D screen, learners can now be immersed in 360° experiences using headsets to experience real or animated clinical environments. As reduced work hours for medical trainees has resulted in fewer clinical learning opportunities, IR has emerged as an alternative method to supplement traditional clinical curricula. In addition, during the global COVID-19 pandemic, where social distancing and remote learning presented challenges for clinical learning, IR enabled students and residents to continue participating in interactive training. However, the literature is just beginning to examine how IR modalities can most effectively be incorporated into clinical teaching curricula.

Another relevant component of teaching curricula is simulation-based training. Simulation is an educational technique that allows interactive, immersive activity by recreating all or part of a clinical experience without exposing patients to associated risks [43]. Historically, full-body manikin simulation has been proposed as an adjunct to clinical learning, complementing clinical experience. Simulation in medical education is a well-established pedagogical practice [44], providing a viable alternative to practice with actual patients, whereby medical trainees are afforded the opportunity to train until they reach a specific competency level. Simulation ranges from decontextualized bench models and VR-based environments to high-fidelity recreations of actual operating rooms [45]. Simulation can engage the trainee in the progressive accumulation of knowledge through deliberate practice, allowing for careful matching of the complexity of the learning environment to the trainees' current level of advancement [46]. Although simulation is widely used in medical education, it is often resource intensive, cost prohibitive [47] and requires learners to physically attend simulation centers. A survey of 154 anesthesiologists found that insufficient time and training opportunities were major barriers to taking part in simulation-based education, with 48% of respondents saying that they would be more likely to attend simulation-based courses if they were hosted at their own hospital [48].

Current advances in IR technology are addressing these limitations through the development of innovative, immersive experiences using cost-effective, consumer-level hardware capable of engaging trainees. A major advantage of IR over manikin-based simulation is increased accessibility, as learners can use the equipment on-site or access IR educational applications remotely. IR equipment can be set up in small spaces (eg, call rooms and conference rooms) with minimal time requirements so that the curricula can be delivered instantaneously in perioperative settings. In addition, although there are costs associated with developing IR applications, the end products can be made modular with the potential for both learning and assessment embedded within. IR curriculum delivery is highly cost-effective and can be easily scaled for additional learners with minimal expense, unlike manikin-based programs. These advantages may result in increased feasibility and reduced costs for IR-based simulation training, although this must still be verified using appropriate cost-benefit studies.

In addition to the use of IR simulation for education and training purposes, IR technologies are increasingly being used directly in patient care, both in patient-centric interventions (eg, as a distraction during painful procedures) and by health care providers (eg, real-time AR-image guidance). In this review, we discuss the theory underlying the use of IR technology in both medical education and clinical care as well as the evidence supporting specific examples relevant to practicing anesthesiologists.

Methods

Overview

This narrative review aimed to identify uses of IR relevant to practicing anesthesiologists. As such, we focused on

XSL•FO
RenderX

perioperative applications and applications related to techniques in which anesthesiologists commonly take part (eg, fiber-optic intubation). Applications of IR focused on addressing pain or anxiety in instances where an anesthesiologist is not normally present (eg, burn wound dressing changes [49]) were excluded. Similarly, while a large amount of research has been done in the area of pediatric anesthesia, we chose to focus on IR applications relevant to those practicing in adult populations. IR applications in pediatric populations have been described elsewhere [50-52]. Finally, we focused on describing existing applications of IR in anesthesia, including up-to-date clinical protocols, while a review focused on theoretical and future applications has been recently published elsewhere [53].

Search Strategy

Relevant publications were identified using PubMed and search terms including "Virtual Reality" AND ("anesthesia" OR "Anaesthesia" OR "Anesthesiology" OR "Anaesthesiology"); "Augmented Reality" AND ("anesthesia" OR "Anaesthesia" OR "Anesthesiology" OR "Anaesthesiology"); vital AND signs AND monitoring AND anesthesia AND display; ("anesth*" OR "anaesthe*") AND (virtual reality OR immersive reality OR augmented reality OR mixed reality OR 360 video) AND (simulation) AND (education) AND (technology) AND (pedagogy OR teaching); ("anesth*" OR "anaesthe*") AND (virtual reality OR immersive reality OR augmented reality OR mixed reality OR 360 video) AND (simulation) AND (education) AND (pedagogy OR Teaching) AND (clinical medicine); ("anesth*" OR "anaesthe*") AND (virtual reality OR immersive reality OR augmented reality OR mixed reality OR 360 video). The literature search included articles published from inception to the date of the search, July 5, 2023. Additional references were identified from the reference lists of the selected papers.

Results

Education Through Immersion: Current and Potential Uses in Medical Education

Overview

Virtual simulation can lead to better learning outcomes for various procedures. Improved health outcomes (eg, patient safety, reduction in costs, and morbidity) have also been reported after the use of computer-enhanced training [54]. Moreover, there are many cases where both technical [55-57] and nontechnical [11] skills acquired during virtual simulation–based training have proven transferrable to subsequent performance in clinical settings.

However, studies aiming to address IR-based educational tools have been heterogeneous in their design and validation, leading to mixed results. This is compounded by the fact that many IR frameworks are designed without using educational theory as a foundation [54,58]. Literature discussing the theory behind the use of immersive technology is limited. However, by taking a broader perspective and viewing the educational literature as a whole, one can begin to form a narrative supporting its use. For example, the learning theory of constructivism outlines that learning occurs by amalgamating previous knowledge and new information to create meaning from an educational activity [59-61]. Both accuracy and real-life relevance of a problem increase learner engagement, and meaningful learning occurs when learners actively interact with the environment through decision-making. IR content is well suited to these tasks as realistic, deeply immersive clinical scenarios can be created, allowing learners to interact with their environment and to have a safe space to make mistakes, free of consequences.

Similarly, medical education has evolved toward self-directed learning and case-based practice [62-64], including the involvement of learning theories, such as self-regulated learning (SRL). SRL is a process by which learners autonomously set goals, monitor their own performance, reflect on outcomes, and then repeat the process to attain desired goals, without an emphasis on instructor-led learning and feedback [65-70]. The versatility of IR learning environments lends itself well to incorporating important components of SRL. For instance, immersive environments based on SRL concepts could provide the learner with educational objectives, allowing them to practice clinical scenarios at their own pace with mechanisms incorporated into the software to assist with progressive difficulty, automatic evaluation, and improved learning.

Many applications in health care, especially those related to nontechnical skills, are still in prototype stages that lack an explicit pedagogical basis [5]. However, this does not imply that IR applications related to clinical decision-making should not be developed. Rather, we are at a critical stage where medical learning theory must inform future IR-based learning applications. Thus, robust research studies examining the effectiveness of these applications, beyond subjective assessments, are urgently needed [71].

IR solutions have the potential to be, and in some cases already are [72,73], effective teaching tools when based on a foundation of learning theory–based design. While this foundation is not always present in current IR-based medical education solutions, we attempted to highlight examples of how IR has been used in anesthesia-based medical education for both technical and procedural teaching and nontechnical skills.

Technical Skills

Traditionally, the most common use of IR in medical education has been to train technical and procedural skills. Training programs in laparoscopic, neurological, endovascular, and orthopedic surgery have led the development of IR with positive learning outcomes [74-77]. This could be considered counterintuitive, as a principal criticism of IR in medical education has been the lack of life-like haptic (the means whereby information is conveyed through touch) technology [78], which is thought to be needed to teach technical skills.

However, educators in heavily procedural medical specialties are faced with a conundrum. On the one hand, trainees require greater autonomy as exposure to clinical settings is reduced. On the other hand, learning technical skills from text alone or by merely watching the corresponding procedure being performed has shown mixed results [79]. IR allows learners to not only visualize procedural skills in the first person but also practice the steps needed for completion in an active manner,

XSL•FO
RenderX

with or without haptic capabilities. The sense of touch has been simulated using joysticks, haptic gloves, and by attaching sensors to common medical tools (eg, syringes).

Existing IR simulators with relevance to anesthesia include simulators for teaching regional anesthesia techniques [39,80-82], bronchoscopy or intubation [40,41,57,83,84], and central vein insertion [85,86] (Table 2; Table S1 in Multimedia Appendix 1) [28,38-42,56,57,80-103]. Often these applications take the form of nonimmersive VR simulators where the trainee uses a clinical tool such as a fiber-optic bronchoscope on a physical model while receiving appropriate clinical feedback (eg, ultrasound or endoscopy images) through a 2D screen. Furthermore, several MR and AR simulators have been developed that combine physical anatomical models with live tracking of handheld tools (eg, ultrasound probes and syringes) translated onto a virtual 3D computer model mimicking real anatomy [80,101,104]. Thus, the user can see, in real time, how their actions performed on the model map onto the 3D virtual anatomy model. By combining physical anatomical models and tools with IR-simulated environments, these models can effectively simulate complex procedures such as vascular access [101,104] and regional anesthesia blocks [80]. In other instances, immersive 360° VR environments have been combined with existing manikin-based simulations to increase immersion and learner engagement [91].

XSL•FO
RenderX

Table 2. Immersive reality (IR) interventions for both technical and nontechnical skills in anesthesia education.

Educational application	Clinical skill
Technical skills	
IR intervention	
AR[a] or MR[b] simulator (nonimmersive)	• Nasogastric tube placement [96] • Central line insertion [85,86,101] • Intravenous insertion [90] • Epidural injection [80] • Transthoracic and transesophageal echocardiography [28] • Lumbar puncture [87] • CPR[c] [92]
VR[d] simulator (nonimmersive)	• Fiber-optic bronchoscopy and intubation [40,41,57,83,84,97,99,102] • Lung isolation [100] • Regional anesthesia [38,39] • Cystoscopy [94] • Ultrasonography [98] • Intravenous insertion [103] • Spinal cord stimulation [42]
VR simulator (immersive)	• Regional and spinal anesthesia [56,81] • CPR and ACLS[e] [17,88,89,91] • Ultrasound guided regional anesthesia [82] • Central line insertion [93]
Nontechnical skills (eg, clinical reasoning and communication)	
IR intervention	
Virtual simulator (nonimmersive)	• Recognition of clinical intraoperative events [105] • Stress response: clinical deterioration [106]
AR simulator	• Communication skills: ACLS scenario [29]
VR simulator (immersive)	• Airway crisis management [26] • Anesthesia crisis management and local anesthetic systemic toxicity [107] • Emergency medicine [108] • Operating room fire safety [10,24,30,109] • Operating room infection prevention [12] • Persuasion training (influenza vaccine hesitancy) [11] • Preparation for an epidural procedure [110] • ACLS team leadership [111] • Surgical unit virtual tour [112] • Therapeutic communication skills [113]
Multiuser VR simulator (immersive)	• Operating room communication [25]
MR simulator (virtual humans)	• Operating room communication [114] • Operating room conflict resolution [115]

[a]AR: augmented reality.

[b]MR: mixed reality.

[c]CPR: cardiopulmonary resuscitation.

[d]VR: virtual reality.

[e]ACLS: advanced cardiac life support.

Nontechnical Skills

IR simulations have also been adapted to teach important nontechnical skills, such as clinical reasoning, communication, and teamwork. These nontechnical skills are crucial to successful clinical encounters, and efficacious uptake of these skills is vital for reducing medical errors and enhancing patient safety. IR simulations can be used to strengthen these skills.

Immersive instructional 360° videos have been developed to teach clinical reasoning skills, allowing for more active learning by incorporating problem-solving elements into the immersive experience. For example, Masson et al [12] developed a 360°

XSL•FO
RenderX

simulation that teaches perioperative infection prevention techniques and allows the learner to make choices during key decision-making points, creating a more "active" observational learning process.

Active learning can also be achieved through computer-generated immersive simulations (Table 2; Table S2 in Multimedia Appendix 1) [10-12,17,24-26,29,30,105-109,111-115]. This is especially useful for teaching comparatively rare or complex clinical events with steep learning curves (eg, airway crisis scenarios). Using commercially available HMDs (Table 1) and tracking sensors, our group [116] (NCT04451590 and NCT04591041) and others [24,26,106,111,117] created simulated crisis scenarios that allowed learners to create objectives and manage patients who are critically ill in virtual operating rooms or trauma bays. Crucially, these IR simulations respond to learners' decisions and provide feedback upon completion. These serve as self-contained simulation scenarios learners can use for deliberate practice to supplement their clinical experiences. Specific clinical scenarios relevant to anesthesiology that have been developed include response to an operating room fire [24,118], airway crisis scenarios (NCT04451590) [26,117], advanced cardiac life support [29,111], and operating room procedures for preventing surgical site infection [12] (Table 2; Table S2 in Multimedia Appendix 1). In addition to specific clinical scenarios, IR can be used to teach and practice "soft skills," such as teamwork and patient communication. Real et al [11] developed a simulator that uses virtual standardized patients to teach health care providers persuasion techniques for supporting patients who are hesitant about vaccines. VR training translated directly to improved clinical communication, as demonstrated by reduced vaccine refusal rates in the 3 months following VR training relative to controls [11].

In another study by Cordar et al [115], virtual humans were used as an operating room team to model conflict resolution skills during critical operating room scenarios. They found that participants were positively influenced by the virtual humans demonstrating ideal conflict resolution techniques and, conversely, were negatively influenced when they modeled negative conflict resolution behavior. These findings demonstrate that IR can be an effective method to teach nontechnical skills, such as team training and communication [115].

We have discussed the theoretical foundation for the use of IR in medical education and specific applications related to anesthesia education for both technical and nontechnical skills. IR allows learners to actively experience the complex anatomy of patients and how they interact with medical tools in a safe, accessible environment. Similarly, IR using HMDs and 360° recordings, facilitated by advances in technology, offers the ability to develop important nontechnical skills. Further information regarding the key outcomes of studies supporting the use of IR in anesthesia medical education is provided in Tables S1 and S2 in Multimedia Appendix 1. Moving forward, it is important to conduct robust trials to assess the educational efficacy of IR, determine its simulator validity, and confirm its cost-effectiveness.

Clinical Uses of IR in Anesthesia

Overview

In addition to the educational applications described in the Education Through Immersion: Current and Potential Uses in Medical Education section, IR technologies are increasingly being used directly in patient care. These include patient-centered interventions, where patients are exposed to IR as an educational tool or as a distraction during painful procedures (Table 3; Table S3 in Multimedia Appendix 1) [7-9,13-16,18-23,27,32,119-144] as well as applications focused on health care providers, where IR technology and 3D visualizations are used to plan or guide clinical care (Table 4; Table S4 in Multimedia Appendix 1) [145-166]. In this Clinical Uses of IR in Anesthesia section, we discuss existing clinical uses of IR technology relevant to practicing anesthesiologists.

XSL•FO
RenderX

Table 3. Patient-based clinical applications of immersive reality (IR) in anesthesia.

Clinical application	Clinical scenario
Intraprocedure distraction	
IR intervention	
Immersive video	• Awake hand surgery [13,19,139] • Orthopedic surgery with regional anesthesia [7,8,138] • Ambulatory surgery [144] • Pregnancy termination [18] • Episiotomy repair [143] • Outpatient hysteroscopy [128] • Lipoma excision [133] • Cystoscopy [129] • Bone marrow aspiration [132] • Transcatheter aortic valve implantation [122] • Gastrointestinal endoscopy [126] • Endovascular aneurysm repair [130] • Breast biopsy [134]
Interactive environment	• Adductor canal catheter insertion [32] • Upper gastrointestinal endoscopy [141,142] • Cystoscopy [135] • Outpatient hysteroscopy [124] • Bone marrow biopsy [121] • Awake hand surgery [127]
Interactive game	• Lumbar puncture [27]
Guided breathing, hypnosis, or relaxation	• Hip or knee arthroplasty [23,125,136] • Upper limb surgery [14,15,123] • Atrial fibrillation ablation [137] • Epidural placement for labor [20] • Endoscopic urological surgery [21]
Preprocedure preparation	
IR intervention	
Operating room tour or procedural education	• Elective surgery with general anesthesia [16] • Elective cranial or spinal procedures [9] • Colorectal cancer surgery [140]
Relaxation	• Elective gynecological surgery [22] • Cardiac surgery [131] • Septorhinoplasty [119] • Gastrointestinal endoscopy [120]

XSL•FO
RenderX

Table 4. Applications of immersive reality (IR) focused on health care providers in clinical anesthesia practice.

Clinical application	Clinical scenario
Clinical data visualization and vital signs monitoring	
IR intervention	
AR[a]- or MR[b]-based HMD[c] or smart glasses	• Simulated operating room scenarios [163,165] • General surgery [166] • Nurse anesthetists anesthesia care [155] • Supervising anesthesiologist (simulated) [156,164]
VR[d] HMD	• Arterial blood gas monitoring [145]
Real-time procedure guidance	
IR intervention	
AR HMD overlay or projection	• Epidural and spinal anesthesia [153,157,161] • Simulated epidural and spinal anesthesia [147-149] • Simulated peripheral nerve block [160,162] • Simulated central venous catheterization [159] • Fine-needle aspiration [158] • Triage decision-making and mass casualty event [154] • Percutaneous rhizotomy [146]
Preprocedure planning	
IR intervention	
Immersive VR HMD	• Difficult airway intubation [150]
Nonimmersive VR simulation	• Regional anesthesia administration [152]
AR or MR HMD	• Awake craniotomy [151]

[a]AR: augmented reality.

[b]MR: mixed reality.

[c]HMD: head-mounted display.

[d]VR: virtual reality.

Clinical Data Visualization

AR-based HMDs allow users to see virtual data overlaid on the real world. For example, an anesthesiologist can view patient vital signs overlaid on the surgical field of view. Since 1995, researchers have been investigating the utility of AR-based HMDs in clinical anesthesia [166]. Several studies of anesthesiologists have reported advantages of using IR, including easier and faster recognition of clinical events [156,163,167] and the ability to spend more time focusing on the patient and the surgical field compared to vital sign monitors [168,169].

Liu et al [163] studied whether 12 anesthesiologists in a simulated environment would notice different events that occurred either on the patient monitor or in the operating room. The HMD did not interfere with participants' ability to detect clinical events and may have allowed participants to detect some events more quickly when they were physically constrained by the need to perform a concurrent clinical procedure [163]. Conversely, Sanderson et al [164] reported that IR displays did not improve the ability to detect clinical events relative to advanced auditory displays [170], and others reported that HMDs might make individuals less likely to detect unexpected clinical events [171]. Thus, further research is needed to confirm how to optimally use HMDs, best ergonomic visualization practices, and whether increased attention leads to improved patient outcomes.

Preanesthesia Planning and Real-Time Guidance

Anesthesiologists often use x-rays, computerized tomography (CT), magnetic resonance imaging (MRI), and ultrasound sonography data to evaluate and plan for complex cases [172]. This provides a foundation for planning the safe management of patients with difficult airways, complex comorbidities, or complex neuraxial anatomy for regional blocks.

The use of AR-based HMDs allows for images obtained via CT, MRI, or ultrasound sonography to be mapped directly onto a patient, offering a window into the patient's anatomy. In a surgical context, systems that generate 3D models of solid organs from CT data and overlay them onto the patient using AR enable precise planning of the optimal approach for abdominal procedures [173,174]. Similarly, our group used IR to augment real-life regional anesthesia procedures using AR holographic projections [175]. Here, live ultrasound scans are converted into holographic projections overlaid on a patient's anatomy to serve as a pathway for an operator's needle to follow. Holographic pathways create reliable needle trajectory guidance that has the potential to facilitate neuraxial block

XSL•FO
RenderX

procedures, reducing unsuccessful attempts, needle repositioning, patient discomfort, hematomas, and failure. These studies, along with findings from simulation or phantom models [147-149], reflect the potential for IR systems in preanesthesia planning and procedure guidance, although further research is needed to confirm clinical findings [153].

Pain Management

IR applications have been widely studied with regard to their use as an adjunct to traditional anesthesia, analgesia, and anxiolytic techniques during painful or uncomfortable procedures [176-178]. These interventions come in many forms (eg, passive 360° videos, interactive games, guided breathing exercises, and hypnosis) and have been studied in diverse patient populations. IR applications of most relevance to anesthesiologists include distraction during minor surgery using local or regional anesthesia [7,14,15,19,125,136] as well as during epidural placement [20,123] and adductor canal catheter insertion [32] (Table 3; Table S3 in Multimedia Appendix 1).

IR is generally well received by patients and health care providers. However, studies have shown conflicting results regarding effectiveness. These contradictory findings can likely be explained by the considerable heterogeneity in study design (eg, the nature of control groups and outcome measures) and in the specific nature of the intervention (eg, passive 360° video vs interactive game) that impede useful comparisons between studies [179]. Thus, while the use of IR as a distraction technique during painful or stressful procedures is well reported [176,177], full-scale, adequately powered studies with appropriate controls are needed to confirm the effectiveness in specific patient populations.

Recent studies have shown that reductions in pain or anxiety can reduce or eliminate sedative requirements, for example, during an orthopedic surgery with spinal anesthesia [125,136] or adductor canal catheter insertion [32]. In a study of endoscopic urological surgery comparing VR distraction without sedation to sedation using intravenous midazolam, participants in the VR group required no rescue sedation and reported higher satisfaction and fewer respiratory adverse events. Despite these promising results, other recent studies have shown no difference in sedation use between IR and control groups [7,137]. Therefore, additional research is needed to clarify the observed effects and determine which populations are most likely to benefit.

Similarly, to optimize the utility of these interventions and better understand how and where they will be useful, further robust studies are needed to determine how IR is able to influence pain and anxiety. For example, are reductions in pain or anxiety the result of distraction via immersion in the virtual environment, due to the content of the intervention (eg, videos and guided relaxation), or due to the novelty of the experience itself? While distraction likely contributes to the analgesic effects of VR, some studies suggest that IR may reduce pain directly through inhibition of ascending pain pathways [180]. These findings have been verified using functional MRI studies, where reductions in brain activity secondary to pain stimuli have been demonstrated with the use of IR. Notably, improved immersion and image quality have been associated with higher levels of pain reduction [181].

Patient Preparation and Education

A significant proportion of patients experience anxiety before anesthesia and surgery, often triggered by fear of the unknown [182-184]. Importantly, preoperative anxiety is associated with negative patient outcomes, including increased intraoperative anesthetic requirements, increased postoperative pain, and longer recovery times [185-187]. IR offers the ability to reduce preoperative anxiety, either through immersive, guided relaxation exercises before the procedure [22] or by providing information that alleviates anxiety through exposure therapy [188,189]. In the latter case, immersive operating room tours can allow patients to witness the steps and processes leading up to the induction of anesthesia and continuing to the recovery room firsthand. While much of the research in this area focuses on pediatric populations [51,190-192], 2 randomized controlled trials have demonstrated reduced preoperative anxiety and higher patient satisfaction using VR relative to standard preoperative procedures in an adult population [9,140].

Overall, IR preprocedure preparation is well received by patients, and the creation of similar 360° content is relatively easy. This, therefore, makes for an accessible entry into IR for anesthesiologists looking to create content at low cost [193].

Further information regarding the key outcomes of studies supporting the use of IR in anesthesia-related patient care is provided in Tables S3 and S4 in Multimedia Appendix 1. However, further research is needed to assess whether these findings are reproducible and generalizable to additional patient populations.

Discussion

Future Applications: Practicality of the Technology and Culture Change

The continued development of low-cost, consumer-level IR technology is facilitating a shift in pedagogy from traditional 2D images and videos to learning through interactive IR environments. This is particularly important in medical education, where knowledge acquisition usually includes greater experiential, self-directed, and hands-on learning than in many other disciplines [4]. As described in this paper, IR technologies have already been used to educate trainees, prepare health care providers, and help patients. Over time, emerging technologies, such as multitouch displays, telepresence (an immersive meeting experience that offers high video and audio quality), 3D environments, natural language processing, and artificial intelligence (AI) software [194], have the potential to transform not only medical education and practice but also teaching and learning in general [195].

Expanding Current Use

The advances in technology discussed above in the Introduction section allow for higher fidelity, increased presence, and creation of new applications in the field of IR, but it is also important to consider how existing IR technologies can extend the impact and reach of experts in the field of anesthesia. For instance, IR

XSL•FO
RenderX

can be used in conjunction with traditional or virtual simulations to facilitate feedback and decrease demands on facilitator time. For example, IR can assist with the initial prebriefing [196,197], wherein participants are oriented to objectives and frameworks for their upcoming experience by facilitating a "flipped classroom" experience [198]. In this case, learners are virtually oriented to the simulation center from home, reducing the time required of facilitators and the overall time spent at the simulation center, allowing more time for participation in the simulated scenarios upon their arrival.

Similarly, telesimulation and telementoring, whereby a skilled practitioner can remotely guide a less experienced colleague through a complex procedure, either for educational purposes [199-201] or in an emergency situation [202], have been successfully used in surgical contexts [199-202]. The most complex versions of these apparatuses allow for annotations and even the movements of surgical tools, manipulated by an instructor, to be projected onto the surgical field of view of a colleague using AR technology [200,203]. This provides a proof of concept for using this technology to extend anesthesia services and aid in training anesthesiologists in underserved and rural communities, where they often work in challenging, underresourced environments to serve populations with high needs. In Ontario, Canada, the government is paying attention to the needs of such populations, especially Indigenous populations [204], as they have higher mortality rates and poorer health outcomes compared to urban populations [205]. IR applications, including telementoring and IR simulation training delivered via remote technology, are tools that can be used to advance health care in these regions and bring it in line with urban centers.

One noteworthy strength of IR over other simulation strategies is the ability to readily reimmerse in a simulation. For example, during feedback stages, IR allows both the facilitator and learner to be reimmersed in the simulation scenario. This can trigger the same emotional cues as when the learner was managing the simulated crisis, heightening recall and potentially leading to more fruitful discussions. This would similarly apply to self-debriefing, where learners are able to visualize the scenario from their own perspective but without the cognitive and emotional burden of managing the case, facilitating focus on the decision-making cues. Finally, during interprofessional training, immersive 360° cameras can be oriented to provide the point of view of each participant (eg, a surgeon could use an HMD to be immersed in the anesthesiologist's point of view), potentially leading to a greater appreciation of one another's roles.

The importance of well-being in health care providers has gained greater attention in recent years and presents another area where IR may have utility. Emerging evidence suggests that IR interventions, such as guided relaxation, could also help reduce stress and burnout among health care practitioners both during training [206] and in high-stress situations, such as a pandemic [207,208]. Intriguingly, a recently published small study of 20 nurses has also provided preliminary evidence that the increased usability of VR patient education compared to traditional approaches may lead to improved health care provider satisfaction and reduced burnout scores [209]. While further studies are needed, this early evidence suggests potential indirect benefits of IR, which have not been fully characterized.

Incorporating Technological Advances

As technology advances, the ways we interact with it also evolve. Instead of interacting through touchscreens or hand controllers, we will be able to interact directly through hand motions and eye movement tracking. This may accelerate as promising design breakthroughs, such as transparent interfaces, are integrated into IR systems. AI is another emerging technology that can be incorporated into medical training and presents a potentially paradigm-shifting advancement. For instance, an early study using AI-based feedback has shown evidence for effectiveness in achieving learning goals [118]. In a simulated operating room fire scenario, this study showed a significant performance improvement when an AI-guided virtual trainer provided timely feedback to users in the virtual operating room.

Further technological advances will occur in the field of haptics. Basic haptic (touch simulation) feedback is already widely available at the consumer level through devices that provide vibration feedback (eg, video game controller "rumble packs" or mobile phone vibrations). Furthermore, pseudohaptics [210,211], where illusory haptic sensations are created using combinations of audio, visual, and kinesthetic cues, has also been used to convey various touch-based properties [212-214]. However, currently available consumer-level haptic devices cannot provide the high level of fidelity and range of motion required to simulate many tasks realistically [215], and high-fidelity haptic solutions are cumbersome, complex, and often cost prohibitive [216,217]. As consumer interest in haptics grows, haptic devices capable of providing force feedback beyond simple vibrations will become more widespread. Notably, integrating haptic feedback into IR enhances immersion by simulating sensations and has been used in medical simulations, including training and simulation of medical procedures, allowing trainees to practice realistic interactions in a safe environment.

Adopting new technological advances in IR systems benefits from the increasing availability of so-called maker spaces. These are collaborative workspaces that contain equipment such as laser cutters, computer numerical control machines, soldering irons, and 3D printers that can easily create innovative, customizable, and cost-effective learning tools. 3D printing, in particular, which is becoming widespread in part due to decreased cost, allows digitally designed objects to be replicated in the real world, facilitating the merger between IR and real physical form, albeit with some limitations [4]. In the context of anesthesia, this creates the potential for faithful recreations of anatomical features (eg, airway or spinal column) to be generated using imaging data for use in training or preprocedure planning [218,219].

Limitations

While the potential benefits of IR simulation are considerable, the potential downsides of IR technology must also be considered. This includes the presence of cybersickness (negative symptoms associated with immersion in a digital

XSL•FO
RenderX

world, including feelings of nausea, dizziness, or headache), which can limit usefulness in both educational and clinical contexts [220,221]. The incidence of cybersickness resulting from the use of IR systems depends on many different factors (eg, frame rate, field of view, sex, and previous game playing experience) and ranges from common to very rare [222]. The heterogeneity of both IR technology and interventions studied likely contributes to this variability [220]. However, both practical factors, such as limiting exposure duration, and technological factors, such as improved latency and higher levels of immersion, are known to reduce cybersickness [177,220,223], with 1 recent review concluding that benefits outweigh risks, at least in some patient populations [221]. Further research is needed to identify factors that can reduce unwanted side effects, such as cybersickness, while preserving the many advantages of IR technology.

Furthermore, it must be acknowledged that potential costs associated with the initial implementation of IR technologies, especially in low-resource settings, could be a barrier and potentially exacerbate existing inequality in health care outcomes. Until recently, IR technology was expensive, cumbersome, and restricted to several large institutions and research centers. However, recent increases in computational processing power and the accompanying decrease in the size of electronic components have led to decreasing costs and rising availability of consumer-level technologies (Table 1). The introduction of freely available game engines and mobile devices (eg, tablets and smartphones) capable of reproducing education-worthy environments with adequate fidelity has made IR technology an exciting and emerging field [4]. Widely available low-cost options, such as Google Cardboard [224] and other smartphone-based IR headsets, are suitable for many of the most common clinical uses of IR, such as patient education and pain distraction [134,225,226]. This suggests that the implementation of IR technology does not need to be limited to high-resource settings.

However, further research is needed to identify potential barriers and secondary costs related to the implementation of IR programs in different settings. Similarly, if supported by evidence from continued research, wide adoption of IR technology facilitated by funding support in low-resource settings could have an important impact on reducing inequality and generating long-term cost savings.

From Virtual to Concrete: Are We There Yet?

Despite the great technological advancements made, particularly with respect to computational processing, it is important to consider the limitations of computer hardware with IR applications, particularly when using VR devices. More specifically, designers and developers of IR applications should consider the fidelity of their virtual environments, that is, the level to which a virtual environment resembles the real world it is meant to represent. It is not currently feasible to create a high-fidelity virtual simulation that includes all the senses (ie, input related to sight, sound, and touch). Furthermore, higher-fidelity environments require greater computer processing requirements, which may not be readily available to the average computer users.

Moreover, we must consider the high variability of users who may be susceptible to motion sickness and address accessibility issues, including those related to visual impairment and other disabilities. Finally, as IR technologies in health care are relatively new, it is essential to address ethical concerns. These include the potential short- and long-term effects of virtual environment exposure and HMD use on both patients and health care providers, as well as whether VR headsets should be classified as medical devices?

Conclusions

In this review, we have presented the theoretical basis and evidence supporting the use of IR in the training and clinical practice of modern anesthesiologists. For those interested in implementing IR technology into their own practice, many ready-made technologies exist (Table 1). The processes and techniques required to create and implement low-cost 360° videos and virtual environments using commercially available cameras and software have been described elsewhere [116,193].

Despite recent advancements and increasing interest in IR as a research field, many applications remain in the early stages of development. Robust trials are urgently needed to confirm its clinical and educational effectiveness and to assess its mechanism, educational validity, and cost-effectiveness. While the use of IR technologies may have practical applications in medical education and for a variety of clinical patient populations, it is important to ensure that enthusiasm for this new technology does not overtake available evidence and that all applications are guided by pedagogical theory and evidence-based practice. Thus, there is a unique opportunity to innovate in a promising area of research to show the validity of IR within the anesthesia specialty and demonstrate its ability to optimize learning and clinical delivery of care. Our group founded the University of Toronto's Collaborative Human Immersive Interaction Laboratory to develop IR solutions for clinical research, medical training, and patient education. The Collaborative Human Immersive Interaction Laboratory also aims to scientifically validate IR as a tool to enhance patient care now and in the future.

Acknowledgments

The authors would like to thank the team at the he authors would like to thank the team at the Sunnybrook Canadian Simulation Centre and Dr. Jordan Tarshis, Dr. Stephen Choi, Dr. Connor Brenna, Dr. Sanch Gupta, Shelly Au, and Monica Caldeira for their help with our research programs, and Kyle Wilcocks for his contributions and discussions during the early stages of this project. The authors would also like to acknowledge funding provided to FA and PG from Physicians' Services Incorporated Foundation

XSL•FO
RenderX

(20-14) and additional funding to support FA and PG's research program via a private donation from the Pasare family to the Sunnybrook Foundation. Funders played no role in the conception, design, or writing of this manuscript.

Authors' Contributions

AF and LK completed the literature search, constructed summary tables, and addressed reviewer comments. AF, LK, BK, CM, and FA were involved in drafting the initial version of the manuscript. EBRB, PG, and JW provided contributions during the editing and revision of the manuscript. All authors provided final approval of the initial and revised manuscript and agreed to be accountable for the published work.

Conflicts of Interest

None of the authors have any financial disclosures that would serve as a conflict in writing this paper. However, some of the immersive reality examples mentioned in this paper were developed in our research laboratory (FA).

Multimedia Appendix 1
Summary tables of educational and clinical applications of immersive reality in anesthesia.
[PDF File (Adobe PDF File), 535 KB - jmir_v27i1e62785_app1.pdf]

References

1. Curran VR, Xu X, Aydin MY, Meruvia-Pastor O. Use of extended reality in medical education: an integrative review. Med Sci Educ 2023 Feb;33(1):275-286 [FREE Full text] [doi: 10.1007/s40670-022-01698-4] [Medline: 36569366]
2. Andrews C, Southworth MK, Silva JN, Silva JR. Extended reality in medical practice. Curr Treat Options Cardiovasc Med 2019 Mar 30;21(4):18 [FREE Full text] [doi: 10.1007/s11936-019-0722-7] [Medline: 30929093]
3. Arnaldi B, Guitton P, Moreau G. Virtual Reality and Augmented Reality: Myths and Realities. Hoboken, NJ: John Wiley & Sons; 2018.
4. Birt J, Stromberga Z, Cowling M, Moro C. Mobile mixed reality for experiential learning and simulation in medical and health sciences education. Information 2018 Jan 31;9(2):31. [doi: 10.3390/info9020031]
5. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. PeerJ 2014;2:e469 [FREE Full text] [doi: 10.7717/peerj.469] [Medline: 25071992]
6. Kapralos B, Uribe-Quevedo A, Dubrowski A. Immersive technologies for medical education. In: Lee N, editor. Encyclopedia of Computer Graphics and Games. Cham, Switzerland: Springer; 2018.
7. Huang MY, Scharf S, Chan PY. Effects of immersive virtual reality therapy on intravenous patient-controlled sedation during orthopaedic surgery under regional anesthesia: a randomized controlled trial. PLoS One 2020 Feb 24;15(2):e0229320 [FREE Full text] [doi: 10.1371/journal.pone.0229320] [Medline: 32092098]
8. Chan PY, Scharf S. Virtual reality as an adjunctive nonpharmacological sedative during orthopedic surgery under regional anesthesia: a pilot and feasibility study. Anesth Analg 2017 Oct;125(4):1200-1202. [doi: 10.1213/ANE.0000000000002169] [Medline: 28598921]
9. Bekelis K, Calnan D, Simmons N, MacKenzie TA, Kakoulides G. Effect of an immersive preoperative virtual reality experience on patient reported outcomes: a randomized controlled trial. Ann Surg 2017 Jun;265(6):1068-1073. [doi: 10.1097/SLA.0000000000002094] [Medline: 27906757]
10. Dorozhkin D, Olasky J, Jones DB, Schwaitzberg SD, Jones SB, Cao CG, et al. OR fire virtual training simulator: design and face validity. Surg Endosc 2017 Sep;31(9):3527-3533 [FREE Full text] [doi: 10.1007/s00464-016-5379-7] [Medline: 28039649]
11. Real FJ, DeBlasio D, Beck AF, Ollberding NJ, Davis D, Cruse B, et al. A virtual reality curriculum for pediatric residents decreases rates of influenza vaccine refusal. Acad Pediatr 2017;17(4):431-435. [doi: 10.1016/j.acap.2017.01.010] [Medline: 28126612]
12. Masson C, Birgand G, Castro-Sánchez E, Eichel VM, Comte A, Terrisse H, et al. Is virtual reality effective to teach prevention of surgical site infections in the operating room? Study protocol for a randomised controlled multicentre trial entitled VIP Room study. BMJ Open 2020 Jun 21;10(6):e037299 [FREE Full text] [doi: 10.1136/bmjopen-2020-037299] [Medline: 32565477]
13. Krish G, Immerman I, Kinjo S. Virtual reality may reduce anxiety and enhance surgical experience during wide-awake local anaesthesia no tourniquet surgery: a report of two cases. J Perioper Pract 2022 Jun;32(6):136-141. [doi: 10.1177/1750458920984048] [Medline: 34190639]
14. Alaterre C, Duceau B, Sung Tsai E, Zriouel S, Bonnet F, Lescot T, et al. Virtual Reality for PEripheral Regional Anesthesia (VR-PERLA Study). J Clin Med 2020 Jan 13;9(1):215 [FREE Full text] [doi: 10.3390/jcm9010215] [Medline: 31941129]
15. Faruki A, Nguyen T, Proeschel S, Levy N, Yu J, Ip V, et al. Virtual reality as an adjunct to anesthesia in the operating room. Trials 2019 Dec 27;20(1):782 [FREE Full text] [doi: 10.1186/s13063-019-3922-2] [Medline: 31882015]

XSL•FO
RenderX

16. Vogt L, Klasen M, Rossaint R, Goeretz U, Ebus P, Sopka S. Virtual reality tour to reduce perioperative anxiety in an operating setting before anesthesia: randomized clinical trial. J Med Internet Res 2021 Sep 01;23(9):e28018 [FREE Full text] [doi: 10.2196/28018] [Medline: 34252034]
17. Semeraro F, Ristagno G, Giulini G, Kayal JS, Cavallo P, Farabegoli L, et al. Back to reality: a new blended pilot course of basic life support with virtual reality. Resuscitation 2019 May;138:18-19. [doi: 10.1016/j.resuscitation.2019.02.034] [Medline: 30844415]
18. Sridhar A, Shiliang Z, Woodson R, Kwan L. Non-pharmacological anxiety reduction with immersive virtual reality for first-trimester dilation and curettage: a pilot study. Eur J Contracept Reprod Health Care 2020 Dec;25(6):480-483. [doi: 10.1080/13625187.2020.1836146] [Medline: 33140989]
19. Hoxhallari E, Behr IJ, Bradshaw JS, Morkos MS, Haan PS, Schaefer MC, et al. Virtual reality improves the patient experience during wide-awake local anesthesia no tourniquet hand surgery: a single-blind, randomized, prospective study. Plast Reconstr Surg 2019 Aug;144(2):408-414. [doi: 10.1097/PRS.0000000000005831] [Medline: 31348351]
20. Kist M, Bekemeyer Z, Ralls L, Carvalho B, Rodriguez ST, Caruso TJ. Virtual reality successfully provides anxiolysis to laboring women undergoing epidural placement. J Clin Anesth 2020 May;61:109635. [doi: 10.1016/j.jclinane.2019.109635] [Medline: 31662229]
21. Moon JY, Shin J, Chung J, Ji SH, Ro S, Kim WH. Virtual reality distraction during endoscopic urologic surgery under spinal anesthesia: a randomized controlled trial. J Clin Med 2018 Dec 20;8(1):2 [FREE Full text] [doi: 10.3390/jcm8010002] [Medline: 30577461]
22. Chan JJ, Yeam CT, Kee HM, Tan CW, Sultana R, Sia AT, et al. The use of pre-operative virtual reality to reduce anxiety in women undergoing gynecological surgeries: a prospective cohort study. BMC Anesthesiol 2020 Oct 09;20(1):261 [FREE Full text] [doi: 10.1186/s12871-020-01177-6] [Medline: 33036555]
23. Ledford CK, VanWagner MJ, Sherman CE, Torp KD. Immersive virtual reality used as adjunct anesthesia for conversion total hip arthroplasty in a 100-year-old patient. Arthroplast Today 2021 Aug 07;10:149-153 [FREE Full text] [doi: 10.1016/j.artd.2021.07.001] [Medline: 34401418]
24. Truong H, Qi D, Ryason A, Sullivan AM, Cudmore J, Alfred S, et al. Does your team know how to respond safely to an operating room fire? Outcomes of a virtual reality, AI-enhanced simulation training. Surg Endosc 2022 May;36(5):3059-3067 [FREE Full text] [doi: 10.1007/s00464-021-08602-y] [Medline: 34264400]
25. Chheang V, Fischer V, Buggenhagen H, Huber T, Huettl F, Kneist W, et al. Toward interprofessional team training for surgeons and anesthesiologists using virtual reality. Int J Comput Assist Radiol Surg 2020 Dec;15(12):2109-2118 [FREE Full text] [doi: 10.1007/s11548-020-02276-y] [Medline: 33083969]
26. Lerner D, Mohr S, Schild J, Göring M, Luiz T. An immersive multi-user virtual reality for emergency simulation training: usability study. JMIR Serious Games 2020 Jul 31;8(3):e18822 [FREE Full text] [doi: 10.2196/18822] [Medline: 32735548]
27. Thomas JJ, Albietz J, Polaner D. Virtual reality for lumbar puncture in a morbidly obese patient with leukemia. Paediatr Anaesth 2018 Nov;28(11):1059-1060. [doi: 10.1111/pan.13505] [Medline: 30284748]
28. Mahmood F, Mahmood E, Dorfman RG, Mitchell J, Mahmood FU, Jones SB, et al. Augmented reality and ultrasound education: initial experience. J Cardiothorac Vasc Anesth 2018 Jun;32(3):1363-1367. [doi: 10.1053/j.jvca.2017.12.006] [Medline: 29452879]
29. Hess O, Qian J, Bruce J, Wang E, Rodriguez S, Haber N, et al. Communication skills training using remote augmented reality medical simulation: a feasibility and acceptability qualitative study. Med Sci Educ 2022 Oct;32(5):1005-1014 [FREE Full text] [doi: 10.1007/s40670-022-01598-7] [Medline: 35966166]
30. Wunder L, Gomez NA, Gonzalez JE, Mitzova-Vladinov G, Cacchione M, Mato J, et al. Fire in the operating room: use of mixed reality simulation with nurse anesthesia students. Informatics 2020 Sep 30;7(4):40. [doi: 10.3390/informatics7040040]
31. Chan E, Hovenden M, Ramage E, Ling N, Pham JH, Rahim A, et al. Virtual reality for pediatric needle procedural pain: two randomized clinical trials. J Pediatr 2019 Jun;209:160-7.e4. [doi: 10.1016/j.jpeds.2019.02.034] [Medline: 31047650]
32. Pandya PG, Kim TE, Howard SK, Stary E, Leng JC, Hunter OO, et al. Virtual reality distraction decreases routine intravenous sedation and procedure-related pain during preoperative adductor canal catheter insertion: a retrospective study. Korean J Anesthesiol 2017 Aug;70(4):439-445 [FREE Full text] [doi: 10.4097/kjae.2017.70.4.439] [Medline: 28794840]
33. Qian K, Arichi T, Price A, Dall'Orso S, Eden J, Noh Y, et al. An eye tracking based virtual reality system for use inside magnetic resonance imaging systems. Sci Rep 2021 Aug 11;11(1):16301 [FREE Full text] [doi: 10.1038/s41598-021-95634-y] [Medline: 34381099]
34. Smith SA, Mulligan NW. Immersion, presence, and episodic memory in virtual reality environments. Memory 2021 Sep;29(8):983-1005. [doi: 10.1080/09658211.2021.1953535] [Medline: 34294002]
35. Reznek M, Harter P, Krummel T. Virtual reality and simulation: training the future emergency physician. Acad Emerg Med 2002 Jan;9(1):78-87 [FREE Full text] [doi: 10.1111/j.1553-2712.2002.tb01172.x] [Medline: 11772675]
36. Nagendran M, Gurusamy KS, Aggarwal R, Loizidou M, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2013 Aug 27;2013(8):CD006575 [FREE Full text] [doi: 10.1002/14651858.CD006575.pub3] [Medline: 23980026]
37. Alaker M, Wynn GR, Arulampalam T. Virtual reality training in laparoscopic surgery: a systematic review and meta-analysis. Int J Surg 2016 May;29:85-94 [FREE Full text] [doi: 10.1016/j.ijsu.2016.03.034] [Medline: 26992652]

XSL•FO
RenderX

38. Grottke O, Ntouba A, Ullrich S, Liao W, Fried E, Prescher A, et al. Virtual reality-based simulator for training in regional anaesthesia. Br J Anaesth 2009 Oct;103(4):594-600 [FREE Full text] [doi: 10.1093/bja/aep224] [Medline: 19706630]
39. O'Sullivan O, Iohom G, O'Donnell BD, Shorten GD. The effect of simulation-based training on initial performance of ultrasound-guided axillary brachial plexus blockade in a clinical setting - a pilot study. BMC Anesthesiol 2014;14:110 [FREE Full text] [doi: 10.1186/1471-2253-14-110] [Medline: 25844062]
40. Jiang B, Ju H, Zhao Y, Yao L, Feng Y. Comparison of the efficacy and efficiency of the use of virtual reality simulation with high-fidelity mannequins for simulation-based training of fiberoptic bronchoscope manipulation. Simul Healthc 2018 Apr;13(2):83-87. [doi: 10.1097/SIH.0000000000000299] [Medline: 29621098]
41. K Latif R, Bautista A, Duan X, Neamtu A, Wu D, Wadhwa A, et al. Teaching basic fiberoptic intubation skills in a simulator: initial learning and skills decay. J Anesth 2016 Feb;30(1):12-19. [doi: 10.1007/s00540-015-2091-z] [Medline: 26493397]
42. White WWJ, Jung MJ. Three-dimensional virtual reality spinal cord stimulator training improves trainee procedural confidence and performance. Neuromodulation 2023 Oct;26(7):1381-1386. [doi: 10.1016/j.neurom.2022.03.005] [Medline: 35570148]
43. Perkins GD. Simulation in resuscitation training. Resuscitation 2007 May;73(2):202-211. [doi: 10.1016/j.resuscitation.2007.01.005] [Medline: 17379380]
44. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. N Engl J Med 2006 Dec 21;355(25):2664-2669. [doi: 10.1056/NEJMra054785] [Medline: 17182991]
45. Kneebone RL. Practice, rehearsal, and performance: an approach for simulation-based surgical and procedure training. JAMA 2009 Sep 23;302(12):1336-1338. [doi: 10.1001/jama.2009.1392] [Medline: 19773572]
46. Guadagnoli M, Morin MP, Dubrowski A. The application of the challenge point framework in medical education. Med Educ 2012 May;46(5):447-453. [doi: 10.1111/j.1365-2923.2011.04210.x] [Medline: 22515752]
47. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: the missing outcome in simulation-based medical education research: a systematic review. Surgery 2013 Feb;153(2):160-176. [doi: 10.1016/j.surg.2012.06.025] [Medline: 22884087]
48. Savoldelli GL, Naik VN, Hamstra SJ, Morgan PJ. Barriers to use of simulation-based education. Can J Anaesth 2005 Nov;52(9):944-950. [doi: 10.1007/BF03022056] [Medline: 16251560]
49. Scapin S, Echevarría-Guanilo ME, Boeira Fuculo Junior PR, Gonçalves N, Rocha PK, Coimbra R. Virtual reality in the treatment of burn patients: a systematic review. Burns 2018 Sep;44(6):1403-1416. [doi: 10.1016/j.burns.2017.11.002] [Medline: 29395400]
50. Wang E, Thomas JJ, Rodriguez ST, Kennedy KM, Caruso TJ. Virtual reality for pediatric periprocedural care. Curr Opin Anaesthesiol 2021 Jun 01;34(3):284-291. [doi: 10.1097/ACO.0000000000000983] [Medline: 33935176]
51. Eijlers R, Utens EM, Staals LM, de Nijs PF, Berghmans JM, Wijnen RM, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353 [FREE Full text] [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
52. Davila U, Price A. Past present and future of simulation in pediatrics. In: StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing; 2023.
53. Alam F, Matava C. A new virtual world? The future of immersive environments in anesthesiology. Anesth Analg 2022 Aug 01;135(2):230-238 [FREE Full text] [doi: 10.1213/ANE.0000000000006118] [Medline: 35839493]
54. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Oct;30(10):4174-4183 [FREE Full text] [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
55. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463. [doi: 10.1097/00000658-200210000-00008] [Medline: 12368674]
56. Kulcsár Z, O'Mahony E, Lövquist E, Aboulafia A, Sabova D, Ghori K, et al. Preliminary evaluation of a virtual reality-based simulator for learning spinal anesthesia. J Clin Anesth 2013 Mar;25(2):98-105. [doi: 10.1016/j.jclinane.2012.06.015] [Medline: 23274077]
57. Wong DT, Mehta A, Singh KP, Leong SM, Ooi A, Niazi A, et al. The effect of virtual reality bronchoscopy simulator training on performance of bronchoscopic-guided intubation in patients: a randomised controlled trial. Eur J Anaesthesiol 2019 Mar;36(3):227-233. [doi: 10.1097/EJA.0000000000000890] [Medline: 30234669]
58. Jiang H, Vimalesvaran S, Wang JK, Lim KB, Mogali SR, Car LT. Virtual reality in medical students' education: scoping review. JMIR Med Educ 2022 Feb 02;8(1):e34860 [FREE Full text] [doi: 10.2196/34860] [Medline: 35107421]
59. Jonassen DH, Duffy TM. Constructivism and the Technology of Instruction: A Conversation. Milton Park, UK: Taylor & Francis; 2013.
60. Whitman N. A review of constructivism: understanding and using a relatively new theory. Fam Med 1993 Sep;25(8):517-521. [Medline: 8405799]
61. Dennick R. Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. Int J Med Educ 2016 Jun 25;7:200-205 [FREE Full text] [doi: 10.5116/ijme.5763.de11] [Medline: 27344115]
62. Graffam B. Active learning in medical education: strategies for beginning implementation. Med Teach 2007 Feb;29(1):38-42. [doi: 10.1080/01421590601176398] [Medline: 17538832]

XSL•FO
RenderX

63. Spencer JA, Jordan RK. Learner centred approaches in medical education. BMJ 1999 May 08;318(7193):1280-1283 [FREE Full text] [doi: 10.1136/bmj.318.7193.1280] [Medline: 10231266]
64. Barrows H. Problem‐based learning in medicine and beyond: a brief overview. New Dir Teach Learn 2006 Aug 17;1996(68):3-12. [doi: 10.1002/tl.37219966804]
65. Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. Med Educ 2012 Jan;46(1):71-79. [doi: 10.1111/j.1365-2923.2011.04100.x] [Medline: 22150198]
66. Rajabi S. Towards self-regulated learning in school curriculum. Procedia Soc Behav Sci 2012;47:344-350. [doi: 10.1016/j.sbspro.2012.06.661]
67. Loyens SM, Magda J, Rikers RM. Self-directed learning in problem-based learning and its relationships with self-regulated learning. Educ Psychol Rev 2008 Jul 22;20(4):411-427. [doi: 10.1007/s10648-008-9082-7]
68. Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory Into Pract 2002 May;41(2):64-70. [doi: 10.1207/s15430421tip4102_2]
69. White CB, Gruppen LD, Fantone JC. Self-regulated learning in medical education. In: Swanwick T, editor. Understanding Medical Education: Evidence, Theory and Practice, Second Edition. Hoboken, NJ: John Wiley & Sons; 2013.
70. Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor-regulated learning in simulation training. Med Educ 2012 Jul;46(7):648-656. [doi: 10.1111/j.1365-2923.2012.04268.x] [Medline: 22691145]
71. Swanwick T, Forrest K, O'Brien BC. Understanding Medical Education: Evidence, Theory, and Practice. Hoboken, NJ: John Wiley & Sons; 2018:1.
72. Huang VW, Jones CB, Gomez ED. State of the art of virtual reality simulation in anesthesia. Int Anesthesiol Clin 2020;58(4):31-35. [doi: 10.1097/AIA.0000000000000298] [Medline: 32815861]
73. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020 Jan 18;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]
74. Brunozzi D, Shakur SF, Kwasnicki A, Ismail R, Charbel FT, Alaraj A. Role of immersive touch simulation in neurosurgical training. In: Alaraj A, editor. Comprehensive Healthcare Simulation: Neurosurgery. Cham, Switzerland: Springer; 2018.
75. Rudarakanchana N, Van Herzeele I, Desender L, Cheshire NJ. Virtual reality simulation for the optimization of endovascular procedures: current perspectives. Vasc Health Risk Manag 2015;11:195-202 [FREE Full text] [doi: 10.2147/VHRM.S46194] [Medline: 25792841]
76. Sugand K, Akhtar K, Khatri C, Cobb J, Gupte C. Training effect of a virtual reality haptics-enabled dynamic hip screw simulator. Acta Orthop 2015;86(6):695-701 [FREE Full text] [doi: 10.3109/17453674.2015.1071111] [Medline: 26168925]
77. Aïm F, Lonjon G, Hannouche D, Nizard R. Effectiveness of virtual reality training in orthopaedic surgery. Arthroscopy 2016 Jan;32(1):224-232. [doi: 10.1016/j.arthro.2015.07.023] [Medline: 26412672]
78. Hannaford B, Okamura AM. Haptics. In: Siciliano B, Khatib O, editors. Springer Handbook of Robotics. Cham, Switzerland: Springer; 2016.
79. Li L, Yu F, Shi D, Shi J, Tian Z, Yang J, et al. Application of virtual reality technology in clinical medicine. Am J Transl Res 2017;9(9):3867-3880 [FREE Full text] [Medline: 28979666]
80. Ameri G, Rankin A, Baxter JS, Moore J, Ganapathy S, Peters TM, et al. Development and evaluation of an augmented reality ultrasound guidance system for spinal anesthesia: preliminary results. Ultrasound Med Biol 2019 Oct;45(10):2736-2746. [doi: 10.1016/j.ultrasmedbio.2019.04.026] [Medline: 31281009]
81. Kim JY, Lee JS, Lee JH, Park YS, Cho J, Koh JC. Virtual reality simulator's effectiveness on the spine procedure education for trainee: a randomized controlled trial. Korean J Anesthesiol 2023 Jun;76(3):213-226 [FREE Full text] [doi: 10.4097/kja.22491] [Medline: 36323305]
82. Chuan A, Qian J, Bogdanovych A, Kumar A, McKendrick M, McLeod G. Design and validation of a virtual reality trainer for ultrasound-guided regional anaesthesia. Anaesthesia 2023 Jun;78(6):739-746 [FREE Full text] [doi: 10.1111/anae.16015] [Medline: 37010989]
83. De Oliveira GSJ, Glassenberg R, Chang R, Fitzgerald P, McCarthy RJ. Virtual airway simulation to improve dexterity among novices performing fibreoptic intubation. Anaesthesia 2013 Oct;68(10):1053-1058 [FREE Full text] [doi: 10.1111/anae.12379] [Medline: 23952805]
84. Cailleau L, Geeraerts T, Minville V, Fourcade O, Fernandez T, Bazin JE, et al. Is there a benefit for anesthesiologists of adding difficult airway scenarios for learning fiberoptic intubation skills using virtual reality training? A randomized controlled study. PLoS One 2023;18(1):e0281016 [FREE Full text] [doi: 10.1371/journal.pone.0281016] [Medline: 36706107]
85. Sappenfield J, Grek S, Cooper LA, Lizdas DE, Lampotang S. Reduced complications of supraclavicular approach in simulated central venous access: applicability to military medicine. Mil Med 2019 Mar 01;184(Suppl 1):329-334 [FREE Full text] [doi: 10.1093/milmed/usy381] [Medline: 30901427]
86. Rochlen LR, Levine R, Tait AR. First-person point-of-view-augmented reality for central line insertion training: a usability and feasibility study. Simul Healthc 2017 Feb;12(1):57-62 [FREE Full text] [doi: 10.1097/SIH.0000000000000185] [Medline: 27930431]
87. Huang X, Yan Z, Gong C, Zhou Z, Xu H, Qin C, et al. A mixed-reality stimulator for lumbar puncture training: a pilot study. BMC Med Educ 2023 Mar 22;23(1):178 [FREE Full text] [doi: 10.1186/s12909-023-04173-9] [Medline: 36949483]

XSL•FO
RenderX

88. Peek JJ, Max SA, Bakhuis W, Huig IC, Rosalia RA, Sadeghi AH, et al. Virtual reality simulator versus conventional advanced life support training for cardiopulmonary resuscitation post-cardiac surgery: a randomized controlled trial. J Cardiovasc Dev Dis 2023 Feb 04;10(2):67 [FREE Full text] [doi: 10.3390/jcdd10020067] [Medline: 36826563]
89. Moll-Khosrawi P, Falb A, Pinnschmidt H, Zöllner C, Issleib M. Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students during COVID-19 pandemic: a randomised controlled trial. BMC Med Educ 2022 Jun 22;22(1):483 [FREE Full text] [doi: 10.1186/s12909-022-03533-1] [Medline: 35733135]
90. Rochlen LR, Putnam E, Levine R, Tait AR. Mixed reality simulation for peripheral intravenous catheter placement training. BMC Med Educ 2022 Dec 17;22(1):876 [FREE Full text] [doi: 10.1186/s12909-022-03946-y] [Medline: 36528576]
91. Issleib M, Kromer A, Pinnschmidt HO, Süss-Havemann C, Kubitz JC. Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial. Scand J Trauma Resusc Emerg Med 2021 Feb 01;29(1):27 [FREE Full text] [doi: 10.1186/s13049-021-00836-y] [Medline: 33526042]
92. Lee DK, Im CW, Jo YH, Chang T, Song JL, Luu C, et al. Comparison of extended reality and conventional methods of basic life support training: protocol for a multinational, pragmatic, noninferiority, randomised clinical trial (XR BLS trial). Trials 2021 Dec 20;22(1):946 [FREE Full text] [doi: 10.1186/s13063-021-05908-z] [Medline: 34930418]
93. Mah E, Yu J, Deck M, Lyster K, Kawchuk J, Turnquist A, et al. Immersive video modeling versus traditional video modeling for teaching central venous catheter insertion to medical residents. Cureus 2021 Mar 02;13(3):e13661 [FREE Full text] [doi: 10.7759/cureus.13661] [Medline: 33824812]
94. Bube S, Dagnaes-Hansen J, Mahmood O, Rohrsted M, Bjerrum F, Salling L, et al. Simulation-based training for flexible cystoscopy - a randomized trial comparing two approaches. Heliyon 2020 Jan;6(1):e03086 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e03086] [Medline: 31922043]
95. Nas J, Thannhauser J, Vart P, van Geuns RJ, Muijsers HE, Mol JQ, et al. Effect of face-to-face vs virtual reality training on cardiopulmonary resuscitation quality: a randomized clinical trial. JAMA Cardiol 2020 Mar 01;5(3):328-335 [FREE Full text] [doi: 10.1001/jamacardio.2019.4992] [Medline: 31734702]
96. Aebersold M, Voepel-Lewis T, Cherara L, Weber M, Khouri C, Levine R, et al. Interactive anatomy-augmented virtual simulation training. Clin Simul Nurs 2018 Feb;15:34-41 [FREE Full text] [doi: 10.1016/j.ecns.2017.09.008] [Medline: 29861797]
97. Casso G, Schoettker P, Savoldelli GL, Azzola A, Cassina T. Development and initial evaluation of a novel, ultraportable, virtual reality bronchoscopy simulator: the computer airway simulation system. Anesth Analg 2019 Nov;129(5):1258-1264 [FREE Full text] [doi: 10.1213/ANE.0000000000003316] [Medline: 29481426]
98. Jensen JK, Dyre L, Jørgensen ME, Andreasen LA, Tolsgaard MG. Simulation-based point-of-care ultrasound training: a matter of competency rather than volume. Acta Anaesthesiol Scand 2018 Jul;62(6):811-819. [doi: 10.1111/aas.13083] [Medline: 29392718]
99. Nilsson PM, Russell L, Ringsted C, Hertz P, Konge L. Simulation-based training in flexible fibreoptic intubation: a randomised study. Eur J Anaesthesiol 2015 Sep;32(9):609-614. [doi: 10.1097/EJA.0000000000000092] [Medline: 24809483]
100. Failor E, Bowdle A, Jelacic S, Togashi K. High-fidelity simulation of lung isolation with double-lumen endotracheal tubes and bronchial blockers in anesthesiology resident training. J Cardiothorac Vasc Anesth 2014 Aug;28(4):865-869. [doi: 10.1053/j.jvca.2013.07.015] [Medline: 24231196]
101. Robinson AR3, Gravenstein N, Cooper LA, Lizdas D, Luria I, Lampotang S. A mixed-reality part-task trainer for subclavian venous access. Simul Healthc 2014 Feb;9(1):56-64. [doi: 10.1097/SIH.0b013e31829b3fb3] [Medline: 24310163]
102. Goldmann K, Steinfeldt T. Acquisition of basic fiberoptic intubation skills with a virtual reality airway simulator. J Clin Anesth 2006 May;18(3):173-178. [doi: 10.1016/j.jclinane.2005.08.021] [Medline: 16731318]
103. Reznek MA, Rawn CL, Krummel TM. Evaluation of the educational effectiveness of a virtual reality intravenous insertion simulator. Acad Emerg Med 2002 Nov;9(11):1319-1325 [FREE Full text] [doi: 10.1111/j.1553-2712.2002.tb01594.x] [Medline: 12414488]
104. Sappenfield JW, Smith WB, Cooper LA, Lizdas D, Gonsalves DB, Gravenstein N, et al. Visualization improves supraclavicular access to the subclavian vein in a mixed reality simulator. Anesth Analg 2018 Jul;127(1):83-89 [FREE Full text] [doi: 10.1213/ANE.0000000000002572] [Medline: 29200069]
105. Erlinger LR, Bartlett A, Perez A. High-fidelity mannequin simulation versus virtual simulation for recognition of critical events by student registered nurse anesthetists. AANA J 2019 Apr;87(2):105-109. [Medline: 31587722]
106. Liaw SY, Sutini, Chua WL, Tan JZ, Levett-Jones T, Ashokka B, et al. Desktop virtual reality versus face-to-face simulation for team-training on stress levels and performance in clinical deterioration: a randomised controlled trial. J Gen Intern Med 2023 Jan;38(1):67-73 [FREE Full text] [doi: 10.1007/s11606-022-07557-7] [Medline: 35501626]
107. Abbas S, Chan V, Jin R, Kim E, Lorello G, Niazi A, et al. ESRA19-0614 Virtual reality simulation for teaching and assessment of anesthesia crisis management. Reg Anesth Pain Med 2019;44:A262 [FREE Full text] [doi: 10.1136/rapm-2019-esraabs2019.480]
108. Knudsen MH, Breindahl N, Dalsgaard TS, Isbye D, Mølbak AG, Tiwald G, et al. Using virtual reality head-mounted displays to assess skills in emergency medicine: validity study. J Med Internet Res 2023 Jun 06;25:e45210 [FREE Full text] [doi: 10.2196/45210] [Medline: 37279049]

XSL•FO
RenderX

109. Sankaranarayanan G, Wooley L, Hogg D, Dorozhkin D, Olasky J, Chauhan S, et al. Immersive virtual reality-based training improves response in a simulated operating room fire scenario. Surg Endosc 2018 Aug;32(8):3439-3449. [doi: 10.1007/s00464-018-6063-x] [Medline: 29372313]
110. Shewaga R, Uribe-Quevedo A, Kapralos B, Alam F. A comparison of seated and room-scale virtual reality in a serious game for epidural preparation. IEEE Trans Emerg Topics Comput 2020 Jan 1;8(1):218-232. [doi: 10.1109/TETC.2017.2746085]
111. Katz D, Shah R, Kim E, Park C, Shah A, Levine A, et al. Utilization of a voice-based virtual reality advanced cardiac life support team leader refresher: prospective observational study. J Med Internet Res 2020 Mar 12;22(3):e17425 [FREE Full text] [doi: 10.2196/17425] [Medline: 32163038]
112. Speidel R, Schneider A, Walter S, Grab-Kroll C, Oechsner W. Immersive medium for early clinical exposure - knowledge acquisition, spatial orientation and the unexpected role of annotation in 360° VR photos. GMS J Med Educ 2023;40(1):Doc8 [FREE Full text] [doi: 10.3205/zma001590] [Medline: 36923314]
113. Hoek KB, van Velzen M, Sarton EY. Patient-embodied virtual reality as a learning tool for therapeutic communication skills among anaesthesiologists: a phenomenological study. Patient Educ Couns 2023 Sep;114:107789 [FREE Full text] [doi: 10.1016/j.pec.2023.107789] [Medline: 37230038]
114. Cordar A, Wendling A, White C, Lampotang S, Lok B. Repeat after me: using mixed reality humans to influence best communication practices. In: Proceedings of the IEEE Virtual Reality. 2017 Presented at: VR 2017; March 18-22, 2017; Los Angeles, CA. [doi: 10.1109/VR.2017.7892242]
115. Cordar A, Robb A, Wendling A, Lampotang S, White C, Lok B. Virtual role-models: using virtual humans to train best communication practices for healthcare teams. In: Proceedings of the 15th International Conference on Intelligent Virtual Agents. 2015 Presented at: IVA 2015; August 26-28, 2015; Delft, The Netherlands. [doi: 10.1007/978-3-319-21996-7_23]
116. Gupta S, Wilcocks K, Matava C, Wiegelmann J, Kaustov L, Alam F. Creating a successful virtual reality-based medical simulation environment: tutorial. JMIR Med Educ 2023 Feb 14;9:e41090 [FREE Full text] [doi: 10.2196/41090] [Medline: 36787169]
117. Putnam EM, Rochlen LR, Alderink E, Augé J, Popov V, Levine R, et al. Virtual reality simulation for critical pediatric airway management training. J Clin Transl Res 2021 Feb 25;7(1):93-99 [FREE Full text] [Medline: 34104812]
118. Truong H, Sullivan AM, Abu-Nuwar MR, Therrien S, Jones SB, Pawlowski J, et al. Operating room team training using simulation: hope or hype? Am J Surg 2021 Dec;222(6):1146-1153. [doi: 10.1016/j.amjsurg.2021.01.044] [Medline: 33933207]
119. Baytar AD, Bollucuo Lu K. Effect of virtual reality on preoperative anxiety in patients undergoing septorhinoplasty. Braz J Anesthesiol 2023;73(2):159-164 [FREE Full text] [doi: 10.1016/j.bjane.2021.08.014] [Medline: 34562488]
120. Kim Y, Yoo SH, Chun J, Kim JH, Youn YH, Park H. Relieving anxiety through virtual reality prior to endoscopic procedures. Yonsei Med J 2023 Feb;64(2):117-122 [FREE Full text] [doi: 10.3349/ymj.2022.0319] [Medline: 36719019]
121. Le Du K, Septans AL, Maloisel F, Vanquaethem H, Schmitt A, Le Goff M, et al. A new option for pain prevention using a therapeutic virtual reality solution for bone marrow biopsy (REVEH Trial): open-label, randomized, multicenter, phase 3 study. J Med Internet Res 2023 Feb 15;25:e38619 [FREE Full text] [doi: 10.2196/38619] [Medline: 36790852]
122. Lind A, Ahsan M, Totzeck M, Al-Rashid F, Haddad A, Dubler S, et al. Virtual reality-assisted distraction during transcatheter aortic valve implantation under local anaesthesia: a randomised study. Int J Cardiol 2023 Sep 15;387:131130. [doi: 10.1016/j.ijcard.2023.131130] [Medline: 37355243]
123. Lopes K, Dessieux T, Rousseau C, Beloeil H. Virtual reality as a hypnotic tool in the management of anxiety during the performance of the axillary block. J Med Syst 2023 Mar 01;47(1):31. [doi: 10.1007/s10916-023-01923-6] [Medline: 36856907]
124. Pelazas-Hernández JA, Varillas-Delgado D, González-Casado T, Cristóbal-Quevedo I, Alonso-Bermejo A, Ronchas-Martínez M, et al. The effect of virtual reality on the reduction of pain in women with an indication for outpatient diagnostic hysteroscopy: a randomized controlled trial. J Clin Med 2023 May 24;12(11):3645 [FREE Full text] [doi: 10.3390/jcm12113645] [Medline: 37297840]
125. Barry KS, Nimma SR, Spaulding AC, Wilke BK, Torp KD, Ledford CK. Perioperative outcomes of immersive virtual reality as adjunct anesthesia in primary total hip and knee arthroplasty. Arthroplast Today 2022 Dec;18:84-88 [FREE Full text] [doi: 10.1016/j.artd.2022.09.015] [Medline: 36312886]
126. Boonreunya N, Nopawong E, Yongsiriwit K, Chirapongsathorn S. Virtual reality distraction during upper gastrointestinal endoscopy: a randomized controlled trial. J Gastroenterol Hepatol 2022 May;37(5):855-860. [doi: 10.1111/jgh.15786] [Medline: 35080055]
127. Faruki AA, Nguyen TB, Gasangwa DV, Levy N, Proeschel S, Yu J, et al. Virtual reality immersion compared to monitored anesthesia care for hand surgery: a randomized controlled trial. PLoS One 2022;17(9):e0272030 [FREE Full text] [doi: 10.1371/journal.pone.0272030] [Medline: 36129891]
128. Fouks Y, Kern G, Cohen A, Reicher L, Shapira Z, Many A, et al. A virtual reality system for pain and anxiety management during outpatient hysteroscopy-a randomized control trial. Eur J Pain 2022 Mar 23;26(3):600-609. [doi: 10.1002/ejp.1882] [Medline: 34748679]

XSL•FO
RenderX

129. Goergen DI, Freitas DM. Virtual reality as a distraction therapy during cystoscopy: a clinical trial. Rev Col Bras Cir 2022;49:e20223138 [FREE Full text] [doi: 10.1590/0100-6991e-20223138-en] [Medline: 35584530]
130. Lame C, Mougin J, Postiglione TJ, Fabre D, Haulon S. Virtual reality to ease endovascular repair of thoracoabdominal aneurysms under local anesthesia. J Endovasc Ther 2023 Apr;30(2):312-315. [doi: 10.1177/15266028221079768] [Medline: 35184605]
131. Rousseaux F, Dardenne N, Massion PB, Ledoux D, Bicego A, Donneau AF, et al. Virtual reality and hypnosis for anxiety and pain management in intensive care units: a prospective randomised trial among cardiac surgery patients. Eur J Anaesthesiol 2022 Jan 01;39(1):58-66 [FREE Full text] [doi: 10.1097/EJA.0000000000001633] [Medline: 34783683]
132. Soret L, Gendron N, Rivet N, Chocron R, Macraigne L, Clausse D, et al. Pain assessment using virtual reality facemask during bone marrow aspiration: prospective study including propensity-matched analysis. JMIR Serious Games 2022 Oct 12;10(4):e33221 [FREE Full text] [doi: 10.2196/33221] [Medline: 36222814]
133. Bozdoğan Yeşilot S, Ciftci H, Yener MK. Using a virtual reality and communication intervention to reduce pain and anxiety in patients undergoing lipoma excision with local anesthesia: a randomized controlled trial. AORN J 2022 May;115(5):437-449. [doi: 10.1002/aorn.13665] [Medline: 35476192]
134. Karaman D, Taşdemir N. The effect of using virtual reality during breast biopsy on pain and anxiety: a randomized controlled trial. J Perianesth Nurs 2021 Dec;36(6):702-705. [doi: 10.1016/j.jopan.2021.04.007] [Medline: 34686402]
135. Łuczak M, Nowak Ł, Chorbińska J, Galik K, Kiełb P, Łaszkiewicz J, et al. Influence of virtual reality devices on pain and anxiety in patients undergoing cystoscopy performed under local anaesthesia. J Pers Med 2021 Nov 16;11(11):1214 [FREE Full text] [doi: 10.3390/jpm11111214] [Medline: 34834565]
136. Peuchot H, Khakha R, Riera V, Ollivier M, Argenson JN. Intraoperative virtual reality distraction in TKA under spinal anesthesia: a preliminary study. Arch Orthop Trauma Surg 2021 Dec;141(12):2323-2328. [doi: 10.1007/s00402-021-04065-x] [Medline: 34292380]
137. Roxburgh T, Li A, Guenancia C, Pernollet P, Bouleti C, Alos B, et al. Virtual reality for sedation during atrial fibrillation ablation in clinical practice: observational study. J Med Internet Res 2021 May 27;23(5):e26349 [FREE Full text] [doi: 10.2196/26349] [Medline: 34042589]
138. Tharion JG, Kale S. Patient satisfaction through an immersive experience using a mobile phone-based head-mounted display during arthroscopic knee surgery under spinal anesthesia: a randomized clinical trial. Anesth Analg 2021 Oct 01;133(4):940-948. [doi: 10.1213/ANE.0000000000005666] [Medline: 34283040]
139. Touil N, Pavlopoulou A, Momeni M, Van Pee B, Barbier O, Sermeus L, et al. Evaluation of virtual reality combining music and a hypnosis session to reduce anxiety before hand surgery under axillary plexus block: a prospective study. Int J Clin Pract 2021 Dec;75(12):e15008. [doi: 10.1111/ijcp.15008] [Medline: 34811860]
140. Turrado V, Guzmán Y, Jiménez-Lillo J, Villegas E, de Lacy FB, Blanch J, et al. Exposure to virtual reality as a tool to reduce peri-operative anxiety in patients undergoing colorectal cancer surgery: a single-center prospective randomized clinical trial. Surg Endosc 2021 Jul;35(7):4042-4047. [doi: 10.1007/s00464-021-08407-z] [Medline: 33683433]
141. Mosso Vázquez JL, Wiederhold BK, Miller I, Wiederhold MD. Virtual reality assisted anaesthesia during upper gastrointestinal endoscopy: report of 115 cases. EMJ Innov 2017 Jan 10;1(1):75-82. [doi: 10.33590/emjinnov/10314545]
142. Mosso JL, Wiederhold BK, Miller I, Lara DM, Wiederhold MD. Virtual Reality Assisted Anesthesia (VRAA) during upper gastrointestinal endoscopy: report of 115 cases— analysis of physiological responses. Surg Res Updates 2017;5:1-11 [FREE Full text]
143. JahaniShoorab N, Ebrahimzadeh Zagami S, Nahvi A, Mazluom SR, Golmakani N, Talebi M, et al. The effect of virtual reality on pain in primiparity women during episiotomy repair: a randomize clinical trial. Iran J Med Sci 2015 May;40(3):219-224 [FREE Full text] [Medline: 25999621]
144. Mosso JL, Gorini A, De La Cerda G, Obrador T, Almazan A, Mosso D, et al. Virtual reality on mobile phones to reduce anxiety in outpatient surgery. Stud Health Technol Inform 2009;142:195-200. [Medline: 19377147]
145. Bergauer L, Akbas S, Braun J, Ganter MT, Meybohm P, Hottenrott S, et al. Visual blood, visualisation of blood gas analysis in virtual reality, leads to more correct diagnoses: a computer-based, multicentre, simulation study. Bioengineering (Basel) 2023 Mar 08;10(3):340 [FREE Full text] [doi: 10.3390/bioengineering10030340] [Medline: 36978731]
146. Berger A, Choudhry OJ, Kondziolka D. Augmented reality-assisted percutaneous rhizotomy for trigeminal neuralgia. Oper Neurosurg (Hagerstown) 2023 Jun 01;24(6):665-669. [doi: 10.1227/ons.0000000000000661] [Medline: 36815787]
147. Hayasaka T, Kawano K, Onodera Y, Suzuki H, Nakane M, Kanoto M, et al. Comparison of accuracy between augmented reality/mixed reality techniques and conventional techniques for epidural anesthesia using a practice phantom model kit. BMC Anesthesiol 2023 May 20;23(1):171 [FREE Full text] [doi: 10.1186/s12871-023-02133-w] [Medline: 37210521]
148. Jun EK, Lim S, Seo J, Lee KH, Lee JH, Lee D, et al. Augmented reality-assisted navigation system for transforaminal epidural injection. J Pain Res 2023 Mar 17;16:921-931 [FREE Full text] [doi: 10.2147/JPR.S400955] [Medline: 36960464]
149. Reinacher PC, Cimniak A, Demerath T, Schallner N. Usage of augmented reality for interventional neuraxial procedures: a phantom-based study. Eur J Anaesthesiol 2023 Feb 01;40(2):121-129 [FREE Full text] [doi: 10.1097/EJA.0000000000001755] [Medline: 36121289]
150. Iliff HA, Ahmad I, Evans S, Ingham J, Rees G, Woodford C. Using virtual reality for difficult airway management planning. Anaesth Rep 2022 Jul 05;10(2):e12175 [FREE Full text] [doi: 10.1002/anr3.12175] [Medline: 35813404]

XSL•FO
RenderX

151. Moon RD, Barua NU. Usability of mixed reality in awake craniotomy planning. Br J Neurosurg 2024 Oct;38(5):1139-1143. [doi: 10.1080/02688697.2022.2152429] [Medline: 36537230]
152. Seong H, Yun D, Yoon KS, Kwak JS, Koh JC. Development of pre-procedure virtual simulation for challenging interventional procedures: an experimental study with clinical application. Korean J Pain 2022 Oct 01;35(4):403-412 [FREE Full text] [doi: 10.3344/kjp.2022.35.4.403] [Medline: 36175339]
153. Tanwani J, Alam F, Matava C, Choi S, McHardy P, Singer O, et al. Development of a head-mounted holographic needle guidance system for enhanced ultrasound-guided neuraxial anesthesia: system development and observational evaluation. JMIR Form Res 2022 Jun 23;6(6):e36931 [FREE Full text] [doi: 10.2196/36931] [Medline: 35737430]
154. Follmann A, Ruhl A, Gösch M, Felzen M, Rossaint R, Czaplik M. Augmented reality for guideline presentation in medicine: randomized crossover simulation trial for technically assisted decision-making. JMIR Mhealth Uhealth 2021 Oct 18;9(10):e17472 [FREE Full text] [doi: 10.2196/17472] [Medline: 34661548]
155. Romare C, Enlöf P, Anderberg P, Jildenstål P, Sanmartin Berglund J, Skär L. Nurse anesthetists' experiences using smart glasses to monitor patients' vital signs during anesthesia care: a qualitative study. PLoS One 2021 Apr 21;16(4):e0250122 [FREE Full text] [doi: 10.1371/journal.pone.0250122] [Medline: 33882100]
156. Schlosser PD, Grundgeiger T, Sanderson PM, Happel O. An exploratory clinical evaluation of a head-worn display based multiple-patient monitoring application: impact on supervising anesthesiologists' situation awareness. J Clin Monit Comput 2019 Dec;33(6):1119-1127. [doi: 10.1007/s10877-019-00265-4] [Medline: 30721389]
157. Hetherington J, Lessoway V, Gunka V, Abolmaesumi P, Rohling R. SLIDE: automatic spine level identification system using a deep convolutional neural network. Int J Comput Assist Radiol Surg 2017 Jul;12(7):1189-1198. [doi: 10.1007/s11548-017-1575-8] [Medline: 28361323]
158. Kaneko N, Tsunoda M, Mitsuhashi M, Okubo K, Takeshima T, Sehara Y, et al. Ultrasound-guided fine-needle aspiration in the neck region using an optical see-through head-mounted display: a randomized controlled trial. J Ultrasound Med 2017 Oct;36(10):2071-2077. [doi: 10.1002/jum.14237] [Medline: 28504313]
159. Kaneko N, Sato M, Takeshima T, Sehara Y, Watanabe E. Ultrasound-guided central venous catheterization using an optical see-through head-mounted display: a pilot study. J Clin Ultrasound 2016 Oct;44(8):487-491. [doi: 10.1002/jcu.22374] [Medline: 27297681]
160. Przkora R, McGrady W, Vasilopoulos T, Gravenstein N, Solanki D. Evaluation of the head-mounted display for ultrasound-guided peripheral nerve blocks in simulated regional anesthesia. Pain Med 2015 Nov;16(11):2192-2194. [doi: 10.1111/pme.12765] [Medline: 25930716]
161. Ashab HA, Lessoway VA, Khallaghi S, Cheng A, Rohling R, Abolmaesumi P. AREA: an augmented reality system for epidural anaesthesia. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:2659-2663. [doi: 10.1109/EMBC.2012.6346511] [Medline: 23366472]
162. Udani AD, Harrison TK, Howard SK, Kim TE, Brock-Utne JG, Gaba DM, et al. Preliminary study of ergonomic behavior during simulated ultrasound-guided regional anesthesia using a head-mounted display. J Ultrasound Med 2012 Aug;31(8):1277-1280. [doi: 10.7863/jum.2012.31.8.1277] [Medline: 22837293]
163. Liu D, Jenkins SA, Sanderson PM, Watson MO, Leane T, Kruys A, et al. Monitoring with head-mounted displays: performance and safety in a full-scale simulator and part-task trainer. Anesth Analg 2009 Oct;109(4):1135-1146. [doi: 10.1213/ANE.0b013e3181b5a200] [Medline: 19762741]
164. Sanderson PM, Watson MO, Russell WJ, Jenkins S, Liu D, Green N, et al. Advanced auditory displays and head-mounted displays: advantages and disadvantages for monitoring by the distracted anesthesiologist. Anesth Analg 2008 Jun;106(6):1787-1797. [doi: 10.1213/ane.0b013e31817325cb] [Medline: 18499611]
165. Ormerod DF, Ross B, Naluai-Cecchini A. Use of an augmented reality display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. Stud Health Technol Inform 2003;94:248-250. [Medline: 15455902]
166. Block FEJ, Yablok DO, McDonald JS. Clinical evaluation of the 'head-up' display of anesthesia data. Preliminary communication. Int J Clin Monit Comput 1995 Feb;12(1):21-24. [doi: 10.1007/BF01142442] [Medline: 7782663]
167. Platt MJ. Heads up display. Br J Anaesth 2004 Apr;92(4):602-603 [FREE Full text] [doi: 10.1093/bja/aeh534] [Medline: 15013967]
168. Liu D, Jenkins SA, Sanderson PM, Fabian P, Russell WJ. Monitoring with head-mounted displays in general anesthesia: a clinical evaluation in the operating room. Anesth Analg 2010 Apr 01;110(4):1032-1038. [doi: 10.1213/ANE.0b013e3181d3e647] [Medline: 20357147]
169. Ormerod DF, Ross B, Naluai-Cecchini A. Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. In: Proceedings of the Sixth International Symposium on Wearable Computers. 2002 Presented at: ISWC 2002; October 10, 2002; Seattle, WA. [doi: 10.1109/iswc.2002.1167229]
170. Sanderson P. The multimodal world of medical monitoring displays. Appl Ergon 2006 Jul;37(4):501-512. [doi: 10.1016/j.apergo.2006.04.022] [Medline: 16759627]
171. Krupenia S, Sanderson PM. Does a head-mounted display worsen inattentional blindness? Proc Hum Factors Ergon Soc Annu Meet 2006 Oct 01;50(16):1638-1642. [doi: 10.1177/154193120605001626]
172. Liu SS, Ngeow JE, Yadeau JT. Ultrasound-guided regional anesthesia and analgesia: a qualitative systematic review. Reg Anesth Pain Med 2009;34(1):47-59. [doi: 10.1097/AAP.0b013e3181933ec3] [Medline: 19258988]

XSL•FO
RenderX

173. Cartucho J, Shapira D, Ashrafian H, Giannarou S. Multimodal mixed reality visualisation for intraoperative surgical guidance. Int J Comput Assist Radiol Surg 2020 May;15(5):819-826 [FREE Full text] [doi: 10.1007/s11548-020-02165-4] [Medline: 32333360]
174. Wen R, Chng CB, Chui CK. Augmented reality guidance with multimodality imaging data and depth-perceived interaction for robot-assisted surgery. Robotics 2017 May 24;6(2):13. [doi: 10.3390/robotics6020013]
175. Wiegelmann J, Choi S, McHardy PG, Matava C, Singer O, Kaustov L, et al. Randomized control trial of a holographic needle guidance technique for thoracic epidural placement. Reg Anesth Pain Med 2024 Dec 02;49(12):861-866. [doi: 10.1136/rapm-2023-104703] [Medline: 38212048]
176. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019 Jul 03;12:2053-2085 [FREE Full text] [doi: 10.2147/JPR.S200498] [Medline: 31308733]
177. Chuan A, Zhou JJ, Hou RM, Stevens CJ, Bogdanovych A. Virtual reality for acute and chronic pain management in adult patients: a narrative review. Anaesthesia 2021 May;76(5):695-704 [FREE Full text] [doi: 10.1111/anae.15202] [Medline: 32720308]
178. Privorotskiy A, Garcia VA, Babbitt LE, Choi JE, Cata JP. Augmented reality in anesthesia, pain medicine and critical care: a narrative review. J Clin Monit Comput 2022 Feb;36(1):33-39. [doi: 10.1007/s10877-021-00705-0] [Medline: 33864581]
179. Small C, Laycock H. Are we near to making virtual reality the new reality in pain medicine? Anaesthesia 2021 May;76(5):590-593 [FREE Full text] [doi: 10.1111/anae.15252] [Medline: 32918754]
180. Hoffman HG, Richards TL, Van Oostrom T, Coda BA, Jensen MP, Blough DK, et al. The analgesic effects of opioids and immersive virtual reality distraction: evidence from subjective and functional brain imaging assessments. Anesth Analg 2007 Dec;105(6):1776-83, table of contents. [doi: 10.1213/01.ane.0000270205.45146.db] [Medline: 18042882]
181. Hoffman HG, Seibel EJ, Richards TL, Furness TA, Patterson DR, Sharar SR. Virtual reality helmet display quality influences the magnitude of virtual reality analgesia. J Pain 2006 Nov;7(11):843-850 [FREE Full text] [doi: 10.1016/j.jpain.2006.04.006] [Medline: 17074626]
182. Wollin SR, Plummer JL, Owen H, Hawkins RM, Materazzo F, Morrison V. Anxiety in children having elective surgery. J Pediatr Nurs 2004 Apr;19(2):128-132. [doi: 10.1016/s0882-5963(03)00146-5] [Medline: 15077211]
183. Kain ZN, Mayes LC, Caldwell-Andrews AA, Karas DE, McClain BC. Preoperative anxiety, postoperative pain, and behavioral recovery in young children undergoing surgery. Pediatrics 2006 Aug;118(2):651-658. [doi: 10.1542/peds.2005-2920] [Medline: 16882820]
184. Ayyadhah Alanazi A. Reducing anxiety in preoperative patients: a systematic review. Br J Nurs 2014 Apr;23(7):387-393. [doi: 10.12968/bjon.2014.23.7.387] [Medline: 24732993]
185. Munafò MR, Stevenson J. Anxiety and surgical recovery. Reinterpreting the literature. J Psychosom Res 2001 Oct;51(4):589-596. [doi: 10.1016/s0022-3999(01)00258-6] [Medline: 11595247]
186. Kagan I, Bar-Tal Y. The effect of preoperative uncertainty and anxiety on short-term recovery after elective arthroplasty. J Clin Nurs 2008 Mar;17(5):576-583. [doi: 10.1111/j.1365-2702.2007.01968.x] [Medline: 18093119]
187. Maranets I, Kain ZN. Preoperative anxiety and intraoperative anesthetic requirements. Anesth Analg 1999 Dec;89(6):1346-1351. [doi: 10.1097/00000539-199912000-00003] [Medline: 10589606]
188. Manyande A, Cyna AM, Yip P, Chooi C, Middleton P. Non-pharmacological interventions for assisting the induction of anaesthesia in children. Cochrane Database Syst Rev 2015 Jul 14;2015(7):CD006447 [FREE Full text] [doi: 10.1002/14651858.CD006447.pub3] [Medline: 26171895]
189. Ryu JH, Park SJ, Park JW, Kim JW, Yoo HJ, Kim TW, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
190. Galvez J, Eisenhower M, England W, Wartman E, Simpao A, Rehman M, et al. An interactive virtual reality tour for adolescents receiving proton radiation therapy: proof-of-concept study. JMIR Perioper Med 2019 Mar 05;2(1):e11259 [FREE Full text] [doi: 10.2196/11259] [Medline: 33393932]
191. Ryu JH, Park JW, Nahm FS, Jeon YT, Oh AH, Lee HJ, et al. The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: a prospective, randomized, and controlled trial. J Clin Med 2018 Sep 17;7(9):284 [FREE Full text] [doi: 10.3390/jcm7090284] [Medline: 30227602]
192. Park JW, Nahm FS, Kim JH, Jeon YT, Ryu JH, Han SH. The effect of mirroring display of virtual reality tour of the operating theatre on preoperative anxiety: a randomized controlled trial. IEEE J Biomed Health Inform 2019 Jan 11;23(6):2655-2660. [doi: 10.1109/jbhi.2019.2892485]
193. O'Sullivan B, Alam F, Matava C. Creating low-cost 360-degree virtual reality videos for hospitals: a technical paper on the dos and don'ts. J Med Internet Res 2018 Jul 16;20(7):e239 [FREE Full text] [doi: 10.2196/jmir.9596] [Medline: 30012545]
194. Mergen M, Junga A, Risse B, Valkov D, Graf N, Marschall B. Immersive training of clinical decision making with AI driven virtual patients - a new VR platform called medical tr.AI.ning. GMS J Med Educ 2023 Apr 17;40(2):Doc18 [FREE Full text] [doi: 10.3205/zma001600] [Medline: 37361242]
195. Huff G, Saxberg B. Full immersion 2025: how will 10-year-olds learn? Educ Next 2009;9(3):79.

XSL•FO
RenderX

196. Stephenson E, Poore J. Tips for conducting the pre-brief for a simulation. J Contin Educ Nurs 2016 Aug 01;47(8):353-355 [FREE Full text] [doi: 10.3928/00220124-20160715-05] [Medline: 27467309]
197. Sharoff L. Simulation: pre-briefing preparation, clinical judgment and reflection. What is the connection? J Contemp Med 2015 Jun;5(2):88-101. [doi: 10.16899/ctd.49922]
198. Dong C, Szarek JL, Reed T. The flipped classroom and simulation: a primer for simulation educators. Med Sci Educ 2020 Dec 03;30(4):1627-1632 [FREE Full text] [doi: 10.1007/s40670-020-01041-9] [Medline: 34457830]
199. Nickel F, Cizmic A, Chand M. Telestration and augmented reality in minimally invasive surgery: an invaluable tool in the age of COVID-19 for remote proctoring and telementoring. JAMA Surg 2022 Feb 01;157(2):169-170. [doi: 10.1001/jamasurg.2021.3604] [Medline: 34705030]
200. Liu P, Li C, Xiao C, Zhang Z, Ma J, Gao J, et al. A wearable augmented reality navigation system for surgical telementoring based on Microsoft Hololens. Ann Biomed Eng 2021 Jan;49(1):287-298. [doi: 10.1007/s10439-020-02538-5] [Medline: 32504141]
201. Rojas-Muñoz E, Cabrera ME, Lin C, Andersen D, Popescu V, Anderson K, et al. The system for telementoring with augmented reality (STAR): a head-mounted display to improve surgical coaching and confidence in remote areas. Surgery 2020 Apr;167(4):724-731. [doi: 10.1016/j.surg.2019.11.008] [Medline: 31916990]
202. Rojas-Muñoz E, Andersen D, Cabrera ME, Popescu V, Marley S, Zarzaur B, et al. Augmented reality as a medium for improved telementoring. Mil Med 2019 Mar 01;184(Suppl 1):57-64. [doi: 10.1093/milmed/usy300] [Medline: 30901394]
203. Shabir D, Abdurahiman N, Padhan J, Trinh M, Balakrishnan S, Kurer M, et al. Towards development of a tele-mentoring framework for minimally invasive surgeries. Int J Med Robot 2021 Oct;17(5):e2305. [doi: 10.1002/rcs.2305] [Medline: 34256415]
204. Martin D, Miller AP, Quesnel-Vallée A, Caron NR, Vissandjée B, Marchildon GP. Canada's universal health-care system: achieving its potential. Lancet 2018 Apr 28;391(10131):1718-1735 [FREE Full text] [doi: 10.1016/S0140-6736(18)30181-8] [Medline: 29483027]
205. Meit M, Knudson A, Gilbert T, Yu AT, Tanenbaum E, Ormson E, et al. The 2014 update of the rural-urban chartbook. Rural Health Reform Policy Research Center. 2014 Oct. URL: https://ruralhealth.und.edu/projects/health-reform-policy-research-center/pdf/2014-rural-urban-chartbook-update.pdf [accessed 2025-02-17]
206. Weitzman RE, Wong K, Worrall DM, Park C, McKee S, Tufts RE, et al. Incorporating virtual reality to improve otolaryngology resident wellness: one institution's experience. Laryngoscope 2021 Sep;131(9):1972-1976. [doi: 10.1002/lary.29529] [Medline: 33764531]
207. Ferrer Costa J, Moran N, Garcia Marti C, Colmenares Hernandez LJ, Radu Ciorba Ciorba F, Ciudad MJ. Immediate impact of an 8-week virtual reality educational program on burnout and work engagement among health care professionals: pre-post pilot study. JMIR XR Spatial Comput 2024;1:e55678. [doi: 10.2196/55678]
208. Croghan IT, Hurt RT, Aakre CA, Fokken SC, Fischer KM, Lindeen SA, et al. Virtual reality for health care professionals during a pandemic: a pilot program. J Prim Care Community Health 2022;13:21501319221086716 [FREE Full text] [doi: 10.1177/21501319221086716] [Medline: 35352605]
209. Kim J, Kim D, Oh SH, Kwon H. Virtual reality for preoperative patient education: impact on satisfaction, usability, and burnout from the perspective of new nurses. World J Clin Cases 2024 Oct 06;12(28):6204-6216 [FREE Full text] [doi: 10.12998/wjcc.v12.i28.6204] [Medline: 39371559]
210. Adolph KE, Hoch JE. Motor development: embodied, embedded, enculturated, and enabling. Annu Rev Psychol 2019 Jan 04;70:141-164 [FREE Full text] [doi: 10.1146/annurev-psych-010418-102836] [Medline: 30256718]
211. Ujitoko Y, Ban Y. Survey of pseudo-haptics: haptic feedback design and application proposals. IEEE Trans Haptics 2021 Oct 1;14(4):699-711. [doi: 10.1109/toh.2021.3077619]
212. Argelaguet F, Hoyet L, Trico M, Lecuyer A. The role of interaction in virtual embodiment: effects of the virtual hand representation. In: Proceedings of the IEEE Virtual Reality. 2016 Presented at: VR 2016; March 19-23, 2016; Greenville, SC. [doi: 10.1109/vr.2016.7504682]
213. Yu R, Bowman DA. Pseudo-haptic display of mass and mass distribution during object rotation in virtual reality. IEEE Trans Visual Comput Graphics 2020 May;26(5):2094-2103. [doi: 10.1109/tvcg.2020.2973056]
214. Pacchierotti C, Prattichizzo D, Kuchenbecker KJ. Cutaneous feedback of fingertip deformation and vibration for palpation in robotic surgery. IEEE Trans Biomed Eng 2016 Feb;63(2):278-287. [doi: 10.1109/TBME.2015.2455932] [Medline: 26186763]
215. D'Abbraccio J, Massari L, Prasanna S, Baldini L, Sorgini F, Airò Farulla G, et al. Haptic glove and platform with gestural control for neuromorphic tactile sensory feedback in medical telepresence. Sensors (Basel) 2019 Feb 03;19(3):641 [FREE Full text] [doi: 10.3390/s19030641] [Medline: 30717482]
216. Junput B, Wei X, Jamone L. Feel it on your fingers: dataglove with vibrotactile feedback for virtual reality and telerobotics. In: Proceedings of the 20th Annual Conference on Towards Autonomous Robotic Systems. 2019 Presented at: TAROS 2019; July 3-5, 2019; London, UK. [doi: 10.1007/978-3-030-23807-0_31]
217. Panariello D, Caporaso T, Grazioso S, Di Gironim G, Lanzotti A, Knopp S, et al. Using the KUKA LBR iiwa robot as haptic device for virtual reality training of hip replacement surgery. In: Proceedings of the 2019 Third IEEE International

XSL•FO
RenderX

Conference on Robotic Computing. 2019 Presented at: IRC 2019; February 25-27, 2019; Naples, Italy. [doi: 10.1109/irc.2019.00094]
218. Shaylor R, Verenkin V, Golden E, Matot I. The use of three-dimensional printing and virtual reality to develop a personalised airway plan in a 7.5-year-old child: a case report. Eur J Anaesthesiol 2020 Jun;37(6):512-515. [doi: 10.1097/EJA.0000000000001184] [Medline: 32379152]
219. Han M, Portnova AA, Lester M, Johnson M. A do-it-yourself 3D-printed thoracic spine model for anesthesia resident simulation. PLoS One 2020 Mar 11;15(3):e0228665 [FREE Full text] [doi: 10.1371/journal.pone.0228665] [Medline: 32160198]
220. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019;10:158 [FREE Full text] [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]
221. Drazich BF, McPherson R, Gorman EF, Chan T, Teleb J, Galik E, et al. In too deep? A systematic literature review of fully-immersive virtual reality and cybersickness among older adults. J Am Geriatr Soc 2023 Dec;71(12):3906-3915. [doi: 10.1111/jgs.18553] [Medline: 37560978]
222. Jensen L, Konradsen F. A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol 2017 Nov 25;23(4):1515-1529. [doi: 10.1007/s10639-017-9676-0]
223. Martirosov S, Bureš M, Zítka T. Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. Virtual Real 2022;26(1):15-32 [FREE Full text] [doi: 10.1007/s10055-021-00507-4] [Medline: 34025203]
224. Google Cardboard. Google. URL: https://arvr.google.com/cardboard/ [accessed 2024-11-06]
225. Patel P, Ivanov D, Bhatt S, Mastorakos G, Birckhead B, Khera N, et al. Low-cost virtual reality headsets reduce perceived pain in healthy adults: a multicenter randomized crossover trial. Games Health J 2020 Apr;9(2):129-136. [doi: 10.1089/g4h.2019.0052] [Medline: 31804853]
226. Pandrangi VC, Gaston B, Appelbaum NP, Albuquerque FCJ, Levy MM, Larson RA. The application of virtual reality in patient education. Ann Vasc Surg 2019 Aug;59:184-189. [doi: 10.1016/j.avsg.2019.01.015] [Medline: 31009725]

Abbreviations

AI: artificial intelligence
AR: augmented reality
CT: computerized tomography
HMD: head-mounted display
IR: immersive reality
MR: mixed reality
MRI: magnetic resonance imaging
SRL: self-regulated learning
VR: virtual reality

Edited by T Loetscher; submitted 31.05.24; peer-reviewed by D Patel, Y Zhang; comments to author 29.10.24; revised version received 12.12.24; accepted 05.01.25; published 11.03.25.

Please cite as:
Fleet A, Kaustov L, Belfiore EBR, Kapralos B, Matava C, Wiegelmann J, Giacobbe P, Alam F
Current Clinical and Educational Uses of Immersive Reality in Anesthesia: Narrative Review
J Med Internet Res 2025;27:e62785
URL: https://www.jmir.org/2025/1/e62785
doi:10.2196/62785
PMID:

©Andrew Fleet, Lilia Kaustov, Elio BR Belfiore, Bill Kapralos, Clyde Matava, Julian Wiegelmann, Peter Giacobbe, Fahad Alam. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Neurobiological Mechanisms of Enhanced Pain-Relieving Transcutaneous Electrical Nerve Stimulation via Visuo-Tactile Stimulation in Immersive Virtual Reality: Randomized Controlled Trial

Chenxi Wang[1], PhD; Lanqi Gao[1], MS; Chuan Zhang[2], PhD; Jun Li[1*], PhD; Jixin Liu[1*], PhD

[1]School of Life Science and Technology, Xidian University, Xi'an, China

[2]Department of Radiology, Affiliated Hospital of North Sichuan Medical College, Nanchong, China

[*]these authors contributed equally

Corresponding Author:
Jixin Liu, PhD
School of Life Science and Technology
Xidian University
266 Xinglong Section of Xifeng Road
Xi'an, 710126
China
Phone: 86 02981891070
Email: liujixin2010@gmail.com

Abstract

Background: Enhancing the effectiveness of current pain relief strategies is a persistent clinical challenge. Although transcutaneous electrical nerve stimulation (TENS) is used in various painful conditions, its effectiveness may decline over time, requiring additional pain management strategies. Immersive virtual reality (VR) with personalized visuo-tactile stimulation has demonstrated analgesic properties. Nevertheless, whether visuo-tactile stimulation can enhance the pain-relieving outcomes of TENS and its underlying neurophysiological mechanisms remains largely unknown.

Objective: The study aims to investigate whether the integration of visuo-tactile stimulation with TENS can enhance the pain-relieving outcomes of TENS alone, and we also aim to explore the brain mechanisms underlying the analgesic effect of this integrated intervention.

Methods: In this study, 75 healthy participants were enrolled and randomly assigned to 1 of 3 groups: congruent TENS-VR (TENS-ConVR) and 2 control groups (incongruent TENS-VR [TENS-InVR] and TENS alone). In the context of TENS-ConVR, we combined TENS and VR by connecting TENS-induced paresthesia with personalized visual bodily feedback. The visual feedback was designed to align with the spatiotemporal patterns of the paresthesia induced by TENS. A pain rating task and a 32-channel electroencephalography were applied.

Results: Two-way ANOVAs showed that TENS-ConVR exhibited a statistically greater reduction in pain rating ($F_{1,48}$=6.84; P=.01) and N2 amplitude ($F_{1,48}$=5.69; P=.02) to high-intensity pain stimuli before and after stimulation than TENS alone. The reduction of brain activity was stronger in participants who reported stronger pain-relieving outcomes. TENS-ConVR reduced the brain oscillation in the gamma band, whereas this result was not found in TENS alone.

Conclusions: This study observed that combining TENS and visual stimulation in a single solution could enhance the pain-relieving effect of TENS, which has the potential to improve the effectiveness of current pain management treatments.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2500098834; https://www.chictr.org.cn/showprojEN.html?proj=254171

(J Med Internet Res 2025;27:e63137) doi:10.2196/63137

KEYWORDS

pain; pain-relieving; transcutaneous electrical nerve stimulation; virtual reality; electroencephalography

XSL•FO
RenderX

Introduction

How to improve the efficacy of existing analgesic strategies has always been a clinical concern. The application of transcutaneous electrical nerve stimulation (TENS) has been widely used for pain management, such as postoperative pain, labor pain, and chronic neck pain [1-3]. Low-intensity and high-frequency (50-100 Hz) electrical pulses, a form of TENS, can activate large-diameter A fibers and inhibit the nociceptive volley transmitted via small-diameter A and C fibers that innervate spatially adjacent skin areas [4]. However, the stand-alone applications of TENS may produce limited analgesic effects [5], necessitating the incorporation of additional pain management strategies.

Virtual reality (VR) is a more recent technology that immerses users in a 3D computer simulation environment [6]. Research indicated that VR could reduce pain through various mechanisms, including offering an effective distraction for users to divert their attention away from pain or regulate their emotions [7,8]. One recent study established a connection between spinal cord stimulation and visual feedback in an immersive VR platform, and they found that the combination of stimulation-induced paresthesia and personalized visual bodily feedback resulted in more potent analgesic effects compared to participants who only had VR [9]. This finding suggested that visuo-tactile stimulation in immersive VR might present an opportunity to optimize analgesic efficacy and enhance the analgesic effects of nerve stimulation, such as the TENS. However, the extent to which visuo-tactile stimulation might improve the analgesic effects of TENS and its underlying neurophysiological mechanisms are still poorly explored.

Electroencephalography (EEG) is a reliable bioassay for evaluating how people experience pain and understanding the underlying neurological processes involved in TENS-induced pain relief [10-12]. For example, Peng et al [11] observed that low-frequency TENS had a long-lasting effect on the brain state, as it increased the alpha oscillations in the primary sensorimotor cortex and enhanced the functional connectivity between S1/M1 and the medial prefrontal cortex. Expanding on the research by Peng et al [11], if the incorporation of immersive VR with TENS (TENS-VR) results in enhanced pain relief, it is expected that the reduction of brain activity evoked by painful stimuli will be more pronounced compared to the use of TENS alone.

In this study, we hypothesized that visuo-tactile stimulation that integrated immersive VR with TENS could enhance the analgesic effects of TENS alone in a model of experimentally induced pain, and we also aimed to explore the brain mechanisms underlying the analgesic effect of this integrated intervention. To achieve our aim, we developed a digital therapeutic platform that integrated TENS and immersive VR into a single solution. This immersive VR platform could be able to record and immerse participants in any 360-degree stereoscopic environment and provide additional visual feedback by manipulating digital scenes in a controlled and realistic manner. Three distinct forms of stimulation were applied to produce analgesia, namely TENS alone, congruent TENS-VR (TENS-ConVR), and incongruent TENS-VR (TENS-InVR). The participants in the TENS-ConVR group experienced spatiotemporal congruency between the visual illumination and the tactile sensations elicited by TENS. However, the participants in the TENS-InVR group had spatiotemporal incongruency.

Methods

Participants

Healthy participants were recruited from the local community (Multimedia Appendix 1 [11,13-15]). Participants were told that the aim of the study was “to investigate the neurophysiological and perceptual effects of some electrical stimulation delivered to the skin.”

Ethical Considerations

The experimental procedures were approved by the institutional review board of the Ethics Committee at the First Affiliated Hospital of Xi'an Jiaotong University (XJTU1AF2025 LSYY-384). For privacy and confidentiality, all data were anonymized before analyzing. All participants signed consent forms including a detailed declaration of study-related benefits and risks before participating in the study. All participants received a compensation of CN ¥200 (~US $27.46) upon completion of the study.

Experimental Design

The experiment consisted of 4 stages conducted on a single day: stimulus intensity correction, a pain rating task, an intervention, and a reperformed pain rating task (Figure 1). Two researchers participated in the experiment: one served as the assessor, while the other administered the intervention. To mitigate bias, the outcome assessor was blinded to the intervention conditions.

XSL•FO
RenderX

Figure 1. Schematic illustration of the experimental procedure. The experimental procedure included stimulus intensity correction, a pain rating task, an intervention, and a reperformed pain rating task. NRS: numerical rating scale; TENS: transcutaneous electrical nerve stimulation; TENS-ConVR: congruent virtual reality with transcutaneous electrical nerve stimulation; TENS-InVR: incongruent virtual reality with transcutaneous electrical nerve stimulation.

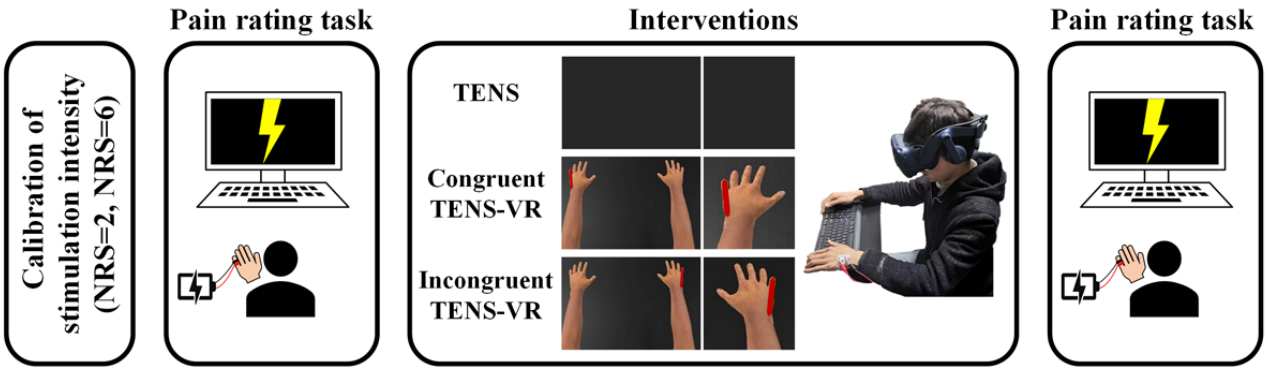

Using a constant-current electrical stimulator to deliver nociceptive stimuli via electrodes on the participant's left little finger (Multimedia Appendix 1), the stimulus intensity is calibrated using an ascending method until the participant rates the intensity of pain as 2 and 6 on a numerical rating scale (NRS; from 0=no pain to 10=unbearable pain). Low-intensity (NRS=2) and high-intensity (NRS=6) electrical stimuli were used in the pain rating task. The determination of the thresholds for low-intensity and high-intensity electrical stimuli is grounded in research conducted by Peng et al [11].

The pain rating task was used to assess the subjective perception of pain in response to pain-inducing stimuli. Participants experienced 2 distinct levels of physical stimulation: low-intensity (NRS=2) and high-intensity (NRS=6). The experiment comprised 4 sessions. Each session consisted of 15 trials, with 10 high-intensity trials and 5 low-intensity trials. These trials were presented in a pseudorandomized order. Each trial began with an 800-1200 ms fixation cross, and participants received either a low-intensity or high-intensity stimulus to the left little finger through a pair of ring electrodes (duration=50 ms). Participants then rated the intensity of the perceived pain on the 0-10 NRS for 2000 ms. The intertrial interval was 8000 ms.

All participants were randomly assigned to receive 1 of 3 interventions: TENS alone, TENS-ConVR, or TENS-InVR. After the pain rating task, participants received a 30-minute intervention according to the randomized stimulation protocols. TENS was generated by a constant current electrical stimulator (Sanxia Technique Inc) and delivered through a pair of surface round electrodes (diameter: 16 mm; interelectrode distance: 3 cm) placed over the ulnar nerve on the dorsum of the left hand. The frequency of stimulation was 85 Hz [3], and the pulse width was 200 µs. The duration of the TENS session was 30 minutes and was divided into 5 blocks. Each block was stimulated for 5 minutes and rested for 1 minute. The stimulus intensity was individually adjusted to elicit a strong but nonpainful tingling sensation (Multimedia Appendix 1).

In order to test whether visuo-tactile stimulation from TENS-ConVR could enhance the analgesic effects of TENS alone, we developed an immersive digital environment referred to as the "VR pain relief experiment platform." A neutral physiotherapy environment was established, comprising a small room with no additional things that could interfere with the experiment. Participants entered the scene using an HTC Vive Pro headset (2880×1600 display resolution; 110 field of view; 90 Hz refreshing rate; and head orientation tracking gravity sensor, gyroscope, magnetometer, and constellation tracking camera), while also being connected to electrical stimulation devices.

The digital arm positions were aligned with respect to the participant's real physical arms. Initially, participants were instructed to identify the area of the hand where they experienced tactile sensations from the TENS. Then, the defined region was illuminated by adjusting visual parameters until it was closely aligned with the subjective experience reported by the participants. In the TENS group, participants observed a dark screen in the immersive digital environment (Figure 1). In the TENS-ConVR group, the immersive TENS-VR platform illuminated the area according to the sensations the individual experienced as induced by TENS (Figure 1). In the TENS-InVR group, VR illuminated the area of the hand contralaterally to the TENS side in real time (Figure 1). The regions of the digital arm matching the electrical stimulation were highlighted during the 5-minute periods of TENS and not highlighted during the 1-minute breaks, aligning with the changes in the electrical stimulation. Following the completion of the intervention, participants were asked to perform the pain rating task again.

EEG Data Collection and Processing

During the pain rating task, a 32-channel EEG system was used to record EEG data synchronously (TMSi SAGA; pass band: 1-100 Hz; sampling rate: 1000 Hz). EEG data were preprocessed in the MATLAB environment using EEGLAB, developed by the Swartz Center for Computational Neuroscience at the University of California, San Diego [13]. The details of the data collection and processing can be found in Multimedia Appendix 1.

XSL•FO
RenderX

Time- and Frequency-Domain Analyses

The event-related potential (ERP) was calculated by averaging the single-trial waveforms in the time domain. The peak latency and amplitude of the N2 and P2 waves were measured from each single-subject average waveform at Cz [11]. Time-frequency distributions of EEG the trials were estimated using a windowed Fourier transform with a fixed 250-ms Hanning window. A frequency-domain analysis was also conducted to evaluate the effect of interventions on spontaneous brain oscillations. Prestimulus EEG signals were extracted from a time window ranging from –1000 ms to 0 ms relative to the stimulus onset (Multimedia Appendix 1).

Statistical Analysis

Two separate 2-way ANOVAs were conducted to evaluate the effects of different interventions on pain ratings and electrophysiological responses. The first ANOVA compared the TENS-ConVR group with the TENS group, and the second ANOVA compared the TENS-ConVR group with the TENS-InVR group. Each ANOVA involved 2 factors: a between-group factor and a within-participants time factor (before and after intervention). Post hoc comparisons were conducted using 2-tailed paired *t* tests with Bonferroni correction when significant main effects or interactions were observed.

Results

Demographics and Questionnaires

A priori power analysis using G*Power software (University of Kiel) [16] indicated that 66 participants were required to achieve a statistical power of 0.95 with an α value of .05 for repeated measures ANOVA designs. To account for potential dropouts or errors, a total of 75 healthy, right-handed participants (mean age 21.52, SD 0.26 years) who never had TENS before were recruited (Figure 2). No significant between-group differences (TENS, TENS-ConVR, and TENS-InVR groups) in demographics were found (Table 1; all $P>.05$). No participants reported any adverse effects related to the experiment.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. TENS: transcutaneous electrical nerve stimulation; TENS-ConVR: congruent virtual reality with transcutaneous electrical nerve stimulation; TENS-InVR: incongruent virtual reality with transcutaneous electrical nerve stimulation.

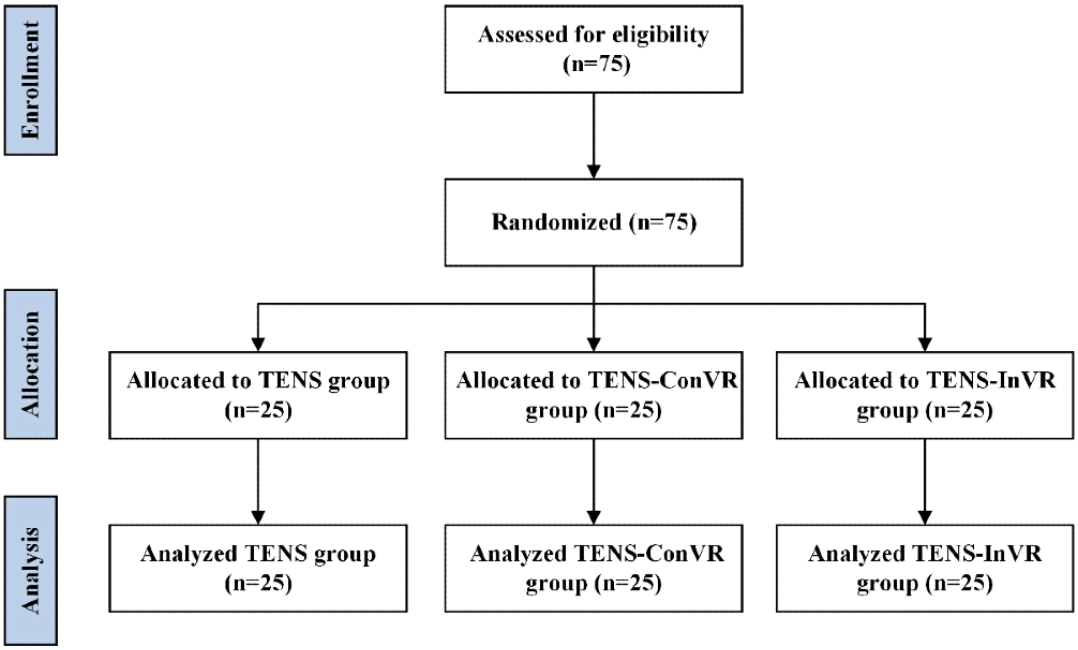

XSL•FO
RenderX

Table 1. The comparison of demographic and psychometrical measurements for TENS[a], TENS-ConVR[b], and TENS-InVR[c] groups[d].

	TENS (n=25)	TENS-ConVR (n=25)	TENS-InVR (n=25)	Statistics	*P* value
Age (years), mean (SEM)	21.84 (0.51)	21.48 (0.43)	21.24 (0.42)	0.44 (2, 72)[e]	.64
Sex, n (%)				0.33 (2)[f]	.85
Female	10 (40)	12 (48)	11 (44)		
Male	15 (60)	13 (52)	14 (56)		
Weight (kg) , mean (SEM)	63.50 (1.75)	62.38 (1.81)	65.97 (3.48)	0.57 (2, 72)[e]	.57
Height (cm) , mean (SEM)	172.60 (1.49)	170.40 (1.69)	170.40 (1.53)	0.57 (2, 72)[e]	.57
Low intensity (µA), mean (SEM)	1171 (74.23)	1191 (126.40)	1376 (110.20)	0.99 (2, 72)[e]	.38
High intensity (µA), mean (SEM)	3074 (267.7)	2915 (275.60)	3444 (253.50)	0.91 (2, 72)[e]	.41

[a]TENS: transcutaneous electrical nerve stimulation.

[b]TENS-ConVR: congruent virtual density with transcutaneous electrical nerve stimulation.

[c]TENS-InVR: incongruent virtual density with transcutaneous electrical nerve stimulation.

[d]Data are expressed using mean (SEM). Statistics were obtained by applying 1-way ANOVA or chi-square test with factors of group (TENS, TENS-ConVR, and TENS-InVR).

[e]*F* test (*df*).

[f]Chi-square (*df*).

Effect of TENS, TENS-ConVR, and TENS-InVR on Pain Rating

For high-intensity trials, significant interactions were observed between the "group" and "time" variables in relation to pain rating in ANOVA analysis (TENS-ConVR group vs TENS group and TENS-ConVR group vs TENS-InVR group; Figure 3A). Specifically, participants in the TENS-ConVR groups had a greater pain reduction compared to those in the TENS group ($F_{1,48}$=6.84; *P*=.01) and TENS-InVR group ($F_{1,48}$=11.54; *P*<.001). After conducting post hoc 2-tailed paired *t* tests on 3 groups, the results showed a significant reduction in pain rating for both the TENS group t_{24}=2.32; 95% CI 0.04-0.63; *P*=.03; Cohen *d*=0.46) and the TENS-ConVR group (t_{24}=5.84; 95% CI 0.57-1.19; *P*<.001; Cohen *d*=1.17). However, the TENS-InVR group did not exhibit a statistically significant difference in pain intensity before and after intervention (t_{24}=0.55; 95% CI –0.27 to 0.46; *P*=.59; Cohen *d*=0.11). For low-intensity trials, no significant interactions were observed between the "group" and "time" when comparing the TENS-ConVR group with the TENS group ($F_{1,48}$=1.72; *P*=.20) or the TENS-ConVR group with the TENS-InVR group ($F_{1,48}$=2.08; *P*=.16; Figure 3B). In the TENS alone group, no significant difference was observed between high-intensity and low-intensity trials ($F_{1,48}$=0.33; *P*=.57; Figure S1A in Multimedia Appendix 2). However, TENS-ConVR–induced pain relieving for high-intensity trials was greater than that for low-intensity trials ($F_{1,48}$=8.78; *P*=.005; Figure S1B in Multimedia Appendix 2).

Figure 3. Effect of TENS, TENS-ConVR, and TENS-InVR on subjective pain rating. (A) Subjective pain rating before and after interventions in high-intensity trials. (B) Subjective pain rating before and after interventions in low-intensity trials. The error bar represents SEM. **P*<.05; ***P*<.01, ****P*<.001. ns: not significant; TENS: transcutaneous electrical nerve stimulation; TENS-ConVR: congruent virtual reality with transcutaneous electrical nerve stimulation; TENS-InVR: incongruent virtual reality with transcutaneous electrical nerve stimulation.

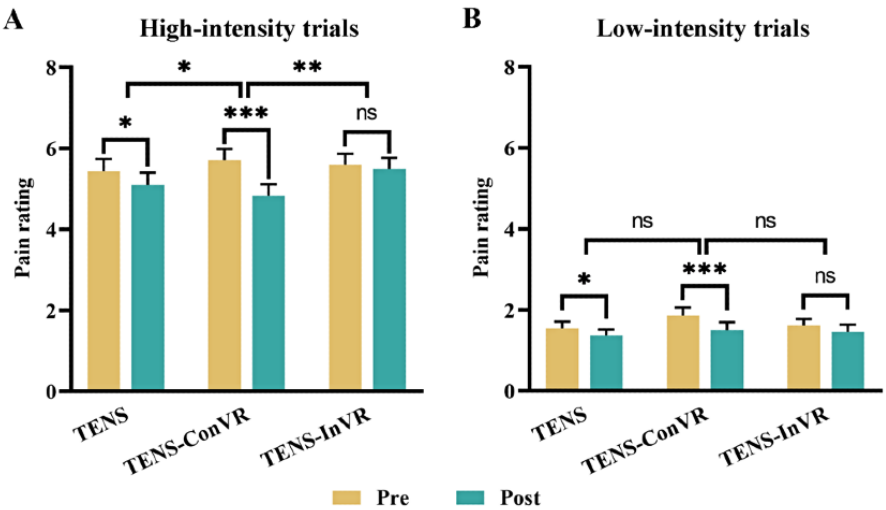

XSL•FO
RenderX

Effect of Different Interventions on ERP

Figure 4 depicts the time-domain scalp morphologies of the ERP waveforms and the N2 pain waves before and after different interventions in high-intensity trials (Figure 4A-C). Significant interactions were observed between the "group" and "time" variables in relation to N2 amplitude in ANOVA analysis (TENS-ConVR group vs TENS group and TENS-ConVR group vs TENS-InVR group; Figure 4D). Specifically, participants in the TENS-ConVR group decreased N2 amplitude more when compared to those in the TENS group ($F_{1,48}$=5.69; *P*=.02) and the TENS-InVR group ($F_{1,48}$=5.41; *P*=.02). After conducting post hoc 2-tailed paired *t* tests on 3 groups, the results showed a significant reduction in N2 amplitude for both the TENS group (t_{24}=3.53; 95% CI –2.93 to –0.77: *P*=.002; Cohen *d*=0.71) and the TENS-ConVR group (t_{24}=8.37, 95% CI –4.29 to –2.59; *P*<.001; Cohen *d*=1.67). However, the TENS-InVR group did not exhibit a statistically significant difference in N2 amplitude before and after intervention (t_{24}=1.19, 95% CI –3 to 0.8; *P*=.25; Cohen *d*=0.24). For P2 amplitude, no significant interactions were observed between the "group" and "time" when comparing the TENS-ConVR group with the TENS group ($F_{1,48}$=0.04; *P*=.85) or the TENS-ConVR group with the TENS-InVR group ($F_{1,48}$=0.11; *P*=.74; Figure S2 in Multimedia Appendix 2). A significant negative correlation was found between the changes in N2 amplitude and pain ratings (*r*=–0.34; *P*=.02; Figure 4E). The time-domain scalp morphologies of the ERP waveforms and the N2 pain waves, both before and after different interventions in low-intensity trials, are detailed in Figure S3A-C in Multimedia Appendix 2. Our analysis found no significant interactions between the "group" and "time" variables in N2 amplitude (TENS-ConVR group vs TENS group: $F_{1,48}$=0.002; *P*=.97 and TENS-ConVR group vs TENS-InVR group: $F_{1,48}$=1.25; *P*=.27; Figure S3D in Multimedia Appendix 2) or P2 amplitude (TENS-ConVR group vs TENS-group: $F_{1,48}$=0.24; *P*=.63 and TENS-ConVR group vs TENS-InVR group: $F_{1,48}$=0.002; *P*=.96; Figure S3E in Multimedia Appendix 2).

Figure 5A illustrates group-level, time-frequency distributions in high-intensity trials, with electrical stimuli eliciting a substantial phase-locked response (ERP:100-500 ms, 1-10 Hz). Significant interactions were observed between the "group" (TENS-ConVR and TENS groups) and "time" variables in relation to ERP magnitude in ANOVA analysis (Figure 5B). Specifically, participants in the TENS-ConVR group had a greater reduction in ERP magnitude compared to those in the TENS group ($F_{1,48}$=4.85; *P*=.03). Post hoc paired *t* tests showed a significant reduction in ERP magnitude for both the TENS group (t_{24}=2.22; 95% CI –0.03 to –0.79; *P*=.04, Cohen *d*=0.44) and the TENS-ConVR group (t_{24}=4.54; 95% CI 0.59-1.56; *P*<.001; Cohen *d*=0.91). This result was not observed in the TENS-InVR group (t_{24}=1.92; 95% CI –0.04 to 1.11; *P*=.07; Cohen *d*=0.38). The group-level, time-frequency distributions in low-intensity trials have been provided in Figure S4A in Multimedia Appendix 2. Our analysis found no significant interactions between the "group" and "time" variables in relation to ERP magnitude for low-intensity trials (TENS-ConVR group vs TENS group: $F_{1,48}$=0.002; *P*=.96 and TENS-ConVR group vs TENS-InVR group: $F_{1,48}$=1.25; *P*=.27; Figure S4B in Multimedia Appendix 2).

Figure 4. Effect of TENS, TENS-ConVR, and TENS-InVR on event-related potential in high-intensity trials. (A-C) Electrical stimulation-evoked responses in the time domain at the group level. For each experimental group, group-level waveforms and scalp topographies of N2 waves (Cz-Avg) are exhibited. At the peak latency of the N2 waves, scale topographies are displayed. (D) N2 amplitude before and after interventions. (E) Correlations between the changes in N2 amplitude and pain rating across TENS and TENS-ConVR conditions. The error bar represents SEM. *P<.05; **P<.01, ***P<.001. ns: not significant; TENS: transcutaneous electrical nerve stimulation; TENS-ConVR: congruent virtual reality with transcutaneous electrical nerve stimulation; TENS-InVR: incongruent virtual reality with transcutaneous electrical nerve stimulation.

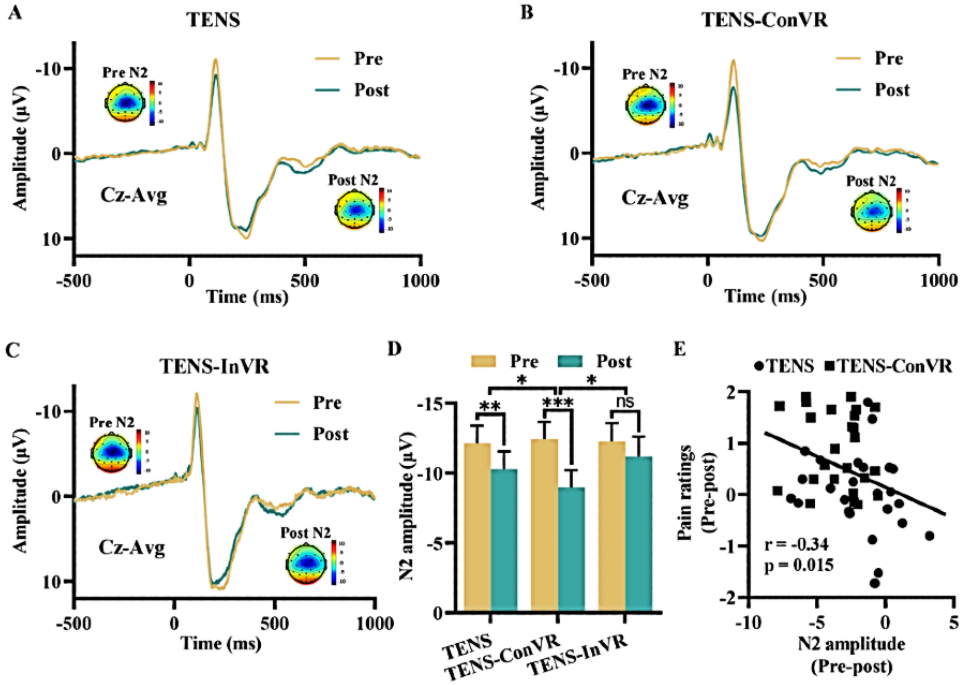

XSL•FO
RenderX

Figure 5. Effect of TENS, TENS-ConVR, and TENS-InVR on time-frequency responses at Cz in high-intensity trials. (A) Group-level time-frequency distributions of ERP responses averaged across intervention groups. (B) ERP magnitude before and after interventions. The error bar represents SEM. *P<.05; **P<.01, ***P<.001. ERP: event-related potential; ns: not significant; TENS: transcutaneous electrical nerve stimulation; TENS-ConVR: congruent virtual reality with transcutaneous electrical nerve stimulation; TENS-InVR: incongruent virtual reality with transcutaneous electrical nerve stimulation.

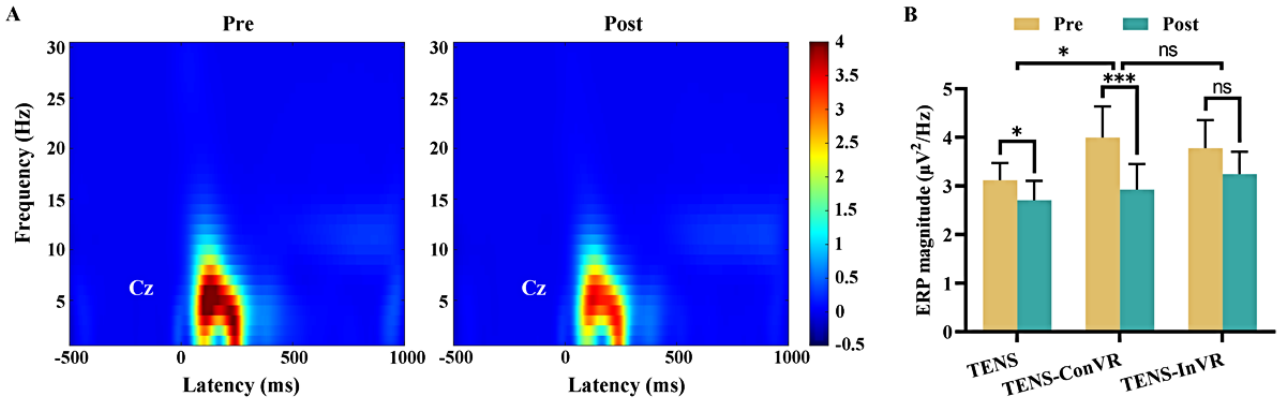

Effect of Different Interventions on Spontaneous Oscillations

Figure S5A-C in Multimedia Appendix 2 depicts the EEG spectra of spontaneous oscillations and scalp topographies before and after different interventions. Participants in the TENS-ConVR and TENS-InVR groups showed a statistically significant reduction in gamma amplitude after stimulation (P=.02 and P=.001, respectively; Figure S5D in Multimedia Appendix 2). No significant pre- and postintervention differences were observed in the remaining frequency bands for all intervention groups (P>.05). A significantly negative correlation was observed between the changes in N2 amplitude and gamma amplitude (r=–0.33; P=.02; Figure S5E in Multimedia Appendix 2).

Discussion

Principal Findings

This study presented a digital therapeutic platform that synergistically combines TENS and VR. Our results demonstrated that the combined TENS and VR intervention, referred to as TENS-ConVR, enhanced the pain-relieving effect of TENS. This integrated approach led to a more pronounced decrease in N2 amplitude compared to TENS alone. Interestingly, participants who reported stronger pain relief also showed a greater decrease in brain activity. Additionally, TENS-ConVR was found to reduce gamma amplitude, which was correlated with the decreased N2 amplitude. These findings suggested that the integration of TENS and visual stimulation in a single solution had the potential to improve the effectiveness of current pain management treatments.

Multiple studies have provided evidence that immersive VR could alleviate pain through distraction or mindfulness-based emotional regulation [7,8,17]. This is achieved by immersing individuals in a computer-generated environment with serene landscapes, soothing music, and other elements [8,18]. However, our findings indicated that the enhanced pain-relieving effect of TENS-ConVR cannot be simply attributed to the distracting effects of VR. This is evidenced by the lack of a statistically significant difference in pain intensity reduction in the TENS-InVR condition, where visual stimuli and tactile stimuli were not spatially congruent. Notably, the analgesic effect of TENS-ConVR was found to be more than double the effect observed in the TENS-InVR condition. Therefore, simply exposing participants to VR or incorporating a visual representation of a participant's body in TENS does not suffice to replicate the pain-relieving effect observed in the TENS-ConVR condition. Notably, we observed an enhanced pain-relieving effect of TENS-ConVR only in high-intensity trials, not in low-intensity trials. Consistent with this, previous studies indicated that VR-based distraction might be more effective at higher levels of pain intensity [19,20]. Therefore, we speculated that the increased pain relief observed with TENS-ConVR may be associated with pain stimuli intensity. Additionally, using a 2-point NRS may lack the sensitivity required to detect minor changes for TENS-ConVR, potentially explaining the absence of an enhanced pain-relieving effect in low-intensity trials.

Studies have demonstrated that pain perception is susceptible to the influence of multisensory inputs and the integration of various body signals [21-23]. This perception is closely tied to any abnormalities in the central representation of painful body parts [24]. The use of congruent multisensory stimulation has the potential to modify this central body representation, thereby providing potential pain relief [9,25]. In a study conducted by Solcà et al [25], participants were shown their digital hands flashing in synchrony or asynchrony with their detected heartbeat. The study found that when the flashing was congruent, it led to reduced pain and increased hand strength in patients with complex regional pain syndrome. This effect was not observed when the flashing was incongruent. Our findings are consistent with the above results. We speculated that the congruent combination of visual and tactile stimulation in the TENS-ConVR condition may foster a sense of ownership over the digital body. This could potentially influence the central representation of the body and consequently enhance the pain-relieving effects of TENS.

Notably, the enhanced pain-relieving effect of TENS-ConVR was further supported through the ERP findings. It showed a

XSL•FO
RenderX

significantly reduced N2 amplitude in the TENS-ConVR condition compared to both TENS alone and TENS-InVR conditions. The N2 component, triggered by painful stimuli and indicating neuronal activation in both the operculo-insular cortices and the contralateral primary somatosensory cortex, is believed to be associated with the processing of pain [26]. According to the gate control theory of pain, applying low-intensity, non-noxious TENS could activate Aβ fibers, thereby inhibiting the transmission of nociceptive signals [27]. Recent studies have used nociceptive stimulation to investigate the analgesic effect of TENS. Researchers measured perceptual and cerebral responses and observed a reduction in N2 amplitude in the active TENS group compared to the sham TENS group [11]. We speculated that the reduced N2 amplitude in the TENS-ConVR condition may be attributed to the congruent visuo-tactile stimulation, which could drive a coherent body representation and potentially enhance somatosensory inhibitory interactions. The discovery of a statistically significant positive connection between the reduction in pain ratings and the decrease in N2 amplitude may provide validity to our inference.

Spontaneous EEG gamma oscillations have been extensively investigated in cognitive processes involved in perception and attention [28]. A study by Hauck et al [29] examined the impact of directed attention on pain-related oscillations and synchronization processes. They found a pronounced increase in induced oscillatory activity within the gamma frequency band when participants focused their attention on painful stimuli. In our results, we found that TENS-ConVR reduced spontaneous gamma oscillations. Additionally, a negative correlation between changes in N2 amplitude and gamma amplitude was observed. We speculated that the reduced spontaneous gamma oscillations may be attributable to decreased attention to pain, facilitated by the multisensory information presented in the immersive VR condition.

TENS is a cost-effective, nonpharmacological intervention that has been widely used for pain management [30]. However, repeated application of TENS has been shown to result in analgesic tolerance within several days in both animal and human studies [31,32]. Unlike other additional therapies for TENS [33-35], such as the combining TENS with local heat and cold applications [35], TENS-VR minimizes the risk of allodynia, a condition where movement or even gentle touch can increase or induce pain, reported by many patients with chronic pain [36]. Furthermore, our digital immersive VR platform enables automatized integration with existing pain treatments, minimizing the active involvement of both patient and therapist. It could enable the implementation of prolonged and repeated stimulation in future TENS-VR investigations, potentially improving its effects.

Limitations

This study has several limitations that should be acknowledged. First, the simplicity of the VR scenario may not adequately induce the desired levels of immersive experience and ownership illusions within the VR environment. Further studies could refine and customize the VR scenario to enhance immersion. Second, the pain relief effect of the TENS-VR was only assessed in healthy participants experiencing experimental pain. Further work is necessary to determine the effectiveness of pain relief in both acute and chronic conditions. Third, there is a weak but statistically significant association between the change in N2 amplitude and the change in NRS score, potentially affected by individual variability and a small sample size. Further studies require validation with a larger sample. Furthermore, the absence of applying a stimulus to the pain threshold and tolerance to assess the effectiveness of TENS, TENS-VR on pain relief represents a limitation in our study. Future studies should include these factors to provide a more comprehensive evaluation of treatment efficacy.

Conclusions

This study observed that combining TENS and visual stimulation in a single solution could enhance the pain-relieving effect of TENS, which has the potential to improve the effectiveness of current pain management treatments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82071879), Xidian University Specially Funded Project for Interdisciplinary Exploration (TZJH2024016), Sichuan Provincial Natural Science Foundation Youth Fund Project (24NSFSC7688), and Sichuan Provincial Health Commission Clinical Research Special Project (23LCYJ019). Full data are not available to be shared due to data protection issues, however, the corresponding author is happy to be contacted. Analysis code can also be shared on request.

Authors' Contributions

CW conducted data analyses, drafted the initial manuscript, and revised and reviewed all drafts of the manuscript. J Li assisted in creating the materials and reviewed and revised the manuscript. LG carried out the data acquisition, assisted with the data analyses, and reviewed and revised the manuscript. CZ assisted in creating the materials, reviewed and provided input on the data analysis, and reviewed and revised the manuscript. J Liu conceptualized and designed the study, coordinated and supervised data collection and analyses, and reviewed and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared. Editorial notice: This randomized study was only retrospectively registered, explained by the authors due to technical problems. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study was considered formative, guiding the development of the application. However,

XSL•FO
RenderX

readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Supplementary methods.
[DOCX File , 24 KB - jmir_v27i1e63137_app1.docx]

Multimedia Appendix 2
Supplementary figures.
[DOCX File , 7497 KB - jmir_v27i1e63137_app2.docx]

Multimedia Appendix 3
CONSORT 2010 checklist.
[DOC File , 230 KB - jmir_v27i1e63137_app3.doc]

References

1. Martimbianco ALC, Porfírio GJ, Pacheco RL, Torloni MR, Riera R. Transcutaneous electrical nerve stimulation (TENS) for chronic neck pain. Cochrane Database Syst Rev 2019;12(12):CD011927 [FREE Full text] [doi: 10.1002/14651858.CD011927.pub2] [Medline: 31830313]
2. Dowswell T, Bedwell C, Lavender T, Neilson JP. Transcutaneous electrical nerve stimulation (TENS) for pain relief in labour. Cochrane Database Syst Rev 2009:CD007214 [FREE Full text] [doi: 10.1002/14651858.CD007214.pub2] [Medline: 19370680]
3. Bjordal JM, Johnson MI, Ljunggreen AE. Transcutaneous electrical nerve stimulation (TENS) can reduce postoperative analgesic consumption. A meta-analysis with assessment of optimal treatment parameters for postoperative pain. Eur J Pain 2003;7(2):181-188. [doi: 10.1016/S1090-3801(02)00098-8] [Medline: 12600800]
4. Pantaleão MA, Laurino MF, Gallego NLG, Cabral CMN, Rakel B, Vance C, et al. Adjusting pulse amplitude during transcutaneous electrical nerve stimulation (TENS) application produces greater hypoalgesia. J Pain 2011;12(5):581-590 [FREE Full text] [doi: 10.1016/j.jpain.2010.11.001] [Medline: 21277840]
5. Johnson M. Transcutaneous electrical nerve stimulation: mechanisms, clinical application and evidence. Rev Pain 2007;1(1):7-11 [FREE Full text] [doi: 10.1177/204946370700100103] [Medline: 26526976]
6. Bruno RR, Wolff G, Wernly B, Masyuk M, Piayda K, Leaver S, et al. Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective. Crit Care 2022;26(1):326 [FREE Full text] [doi: 10.1186/s13054-022-04202-x] [Medline: 36284350]
7. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
8. Li J, Yang H, Xiao Y, Liu X, Ma B, Ma K, et al. The analgesic effects and neural oscillatory mechanisms of virtual reality scenes based on distraction and mindfulness strategies in human volunteers. Br J Anaesth 2023;131(6):1082-1092. [doi: 10.1016/j.bja.2023.09.001] [Medline: 37798154]
9. Solcà M, Krishna V, Young N, Deogaonkar M, Herbelin B, Orepic P, et al. Enhancing analgesic spinal cord stimulation for chronic pain with personalized immersive virtual reality. Pain 2021;162(6):1641-1649. [doi: 10.1097/j.pain.0000000000002160] [Medline: 33259460]
10. Jadidi AF, Jensen W, Zarei AA, Lontis R. Alteration in cortical activity and perceived sensation following modulated TENS. IEEE Trans Neural Syst Rehabil Eng 2023. [doi: 10.1109/TNSRE.2023.3236038] [Medline: 37018675]
11. Peng WW, Tang ZY, Zhang FR, Li H, Kong YZ, Iannetti GD, et al. Neurobiological mechanisms of TENS-induced analgesia. Neuroimage 2019;195:396-408 [FREE Full text] [doi: 10.1016/j.neuroimage.2019.03.077] [Medline: 30946953]
12. Ebrahimian M, Razeghi M, Zamani A, Bagheri Z, Rastegar K, Motealleh A. Does high frequency transcutaneous electrical nerve stimulation (TENS) affect EEG gamma band activity? J Biomed Phys Eng 2018;8(3):271-280 [FREE Full text] [Medline: 30320031]
13. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9-21. [doi: 10.1016/j.jneumeth.2003.10.009] [Medline: 15102499]
14. Hu L, Xiao P, Zhang ZG, Mouraux A, Iannetti GD. Single-trial time-frequency analysis of electrocortical signals: baseline correction and beyond. Neuroimage 2014;84:876-887. [doi: 10.1016/j.neuroimage.2013.09.055] [Medline: 24084069]
15. Mouraux A, Iannetti GD. Across-trial averaging of event-related EEG responses and beyond. Magn Reson Imaging 2008;26(7):1041-1054. [doi: 10.1016/j.mri.2008.01.011] [Medline: 18479877]
16. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]

XSL•FO
RenderX

17. Seabrook E, Kelly R, Foley F, Theiler S, Thomas N, Wadley G, et al. Understanding how virtual reality can support mindfulness practice: mixed methods study. J Med Internet Res 2020;22(3):e16106. [doi: 10.2196/16106] [Medline: 32186519]
18. Bi Y, Liu X, Zhao X, Wei S, Li J, Wang F, et al. Enhancing pain modulation: the efficacy of synchronous combination of virtual reality and transcutaneous electrical nerve stimulation. Gen Psychiatr 2023;36(6):e101164 [FREE Full text] [doi: 10.1136/gpsych-2023-101164] [Medline: 38143714]
19. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
20. Rawlins CR, Veigulis Z, Hebert C, Curtin C, Osborne TF. Effect of immersive virtual reality on pain and anxiety at a veterans affairs health care facility. Front Virtual Real 2021;2:719681. [doi: 10.3389/frvir.2021.719681]
21. Haggard P, Iannetti GD, Longo MR. Spatial sensory organization and body representation in pain perception. Curr Biol 2013;23(4):R164-R176 [FREE Full text] [doi: 10.1016/j.cub.2013.01.047] [Medline: 23428330]
22. Romano D, Pfeiffer C, Maravita A, Blanke O. Illusory self-identification with an avatar reduces arousal responses to painful stimuli. Behav Brain Res 2014;261:275-281 [FREE Full text] [doi: 10.1016/j.bbr.2013.12.049] [Medline: 24412686]
23. Longo MR, Iannetti GD, Mancini F, Driver J, Haggard P. Linking pain and the body: neural correlates of visually induced analgesia. J Neurosci 2012;32(8):2601-2607. [doi: 10.1523/JNEUROSCI.4031-11.2012] [Medline: 22357844]
24. Tsay A, Allen TJ, Proske U, Giummarra MJ. Sensing the body in chronic pain: a review of psychophysical studies implicating altered body representation. Neurosci Biobehav Rev 2015;52:221-232. [doi: 10.1016/j.neubiorev.2015.03.004] [Medline: 25783221]
25. Solcà M, Ronchi R, Bello-Ruiz J, Schmidlin T, Herbelin B, Luthi F, et al. Heartbeat-enhanced immersive virtual reality to treat complex regional pain syndrome. Neurology 2018;91(5):e479-e489. [doi: 10.1212/WNL.0000000000005905] [Medline: 29980635]
26. Iannetti GD, Zambreanu L, Cruccu G, Tracey I. Operculoinsular cortex encodes pain intensity at the earliest stages of cortical processing as indicated by amplitude of laser-evoked potentials in humans. Neuroscience 2005;131(1):199-208. [doi: 10.1016/j.neuroscience.2004.10.035] [Medline: 15680703]
27. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965;150(3699):971-979. [doi: 10.1126/science.150.3699.971] [Medline: 5320816]
28. Kucewicz MT, Berry BM, Kremen V, Brinkmann BH, Sperling MR, Jobst BC, et al. Dissecting gamma frequency activity during human memory processing. Brain 2017;140(5):1337-1350. [doi: 10.1093/brain/awx043] [Medline: 28335018]
29. Hauck M, Lorenz J, Engel AK. Attention to painful stimulation enhances gamma-band activity and synchronization in human sensorimotor cortex. J Neurosci 2007;27(35):9270-9277 [FREE Full text] [doi: 10.1523/JNEUROSCI.2283-07.2007] [Medline: 17728441]
30. Gibson W, Wand BM, Meads C, Catley MJ, O'Connell NE. Transcutaneous electrical nerve stimulation (TENS) for chronic pain—an overview of Cochrane Reviews. Cochrane Database Syst Rev 2019;4(4):CD011890 [FREE Full text] [doi: 10.1002/14651858.CD011890.pub3] [Medline: 30941745]
31. Liebano RE, Rakel B, Vance CGT, Walsh DM, Sluka KA. An investigation of the development of analgesic tolerance to TENS in humans. Pain 2011;152(2):335-342 [FREE Full text] [doi: 10.1016/j.pain.2010.10.040] [Medline: 21144659]
32. Chandran P, Sluka KA. Development of opioid tolerance with repeated transcutaneous electrical nerve stimulation administration. Pain 2003;102(1-2):195-201. [doi: 10.1016/s0304-3959(02)00381-0] [Medline: 12620611]
33. Leemans L, Elma O, Nijs J, Wideman TH, Siffain C, den Bandt H, et al. Transcutaneous electrical nerve stimulation and heat to reduce pain in a chronic low back pain population: a randomized controlled clinical trial. Braz J Phys Ther 2021;25(1):86-96 [FREE Full text] [doi: 10.1016/j.bjpt.2020.04.001] [Medline: 32434666]
34. Adhitya I, Yu WY, Bass P, Kinandana GP, Lin MR. Effects of kinesio taping and transcutaneous electrical nerve stimulation combined with active stretching on hamstring flexibility. J Strength Cond Res 2022;36(11):3087-3092. [doi: 10.1519/JSC.0000000000004088] [Medline: 34224508]
35. Maeda T, Yoshida H, Sasaki T, Oda A. Does transcutaneous electrical nerve stimulation (TENS) simultaneously combined with local heat and cold applications enhance pain relief compared with TENS alone in patients with knee osteoarthritis? J Phys Ther Sci 2017;29(10):1860-1864 [FREE Full text] [doi: 10.1589/jpts.29.1860] [Medline: 29184307]
36. Jensen TS, Finnerup NB. Allodynia and hyperalgesia in neuropathic pain: clinical manifestations and mechanisms. Lancet Neurol 2014;13(9):924-935. [doi: 10.1016/S1474-4422(14)70102-4] [Medline: 25142459]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
EEG: electroencephalography
ERP: event-related potential
NRS: numerical rating scale
TENS: transcutaneous electrical nerve stimulation

XSL•FO
RenderX

TENS-ConVR: congruent virtual reality with transcutaneous electrical nerve stimulation
TENS-InVR: incongruent virtual reality with transcutaneous electrical nerve stimulation
TENS-VR: virtual reality with transcutaneous electrical nerve stimulation

Edited by T Loetscher; submitted 12.06.24; peer-reviewed by L Hu, G Lewis, J Chen, Y Bi; comments to author 20.09.24; revised version received 25.10.24; accepted 13.11.24; published 19.03.25.

Please cite as:
Wang C, Gao L, Zhang C, Li J, Liu J
Neurobiological Mechanisms of Enhanced Pain-Relieving Transcutaneous Electrical Nerve Stimulation via Visuo-Tactile Stimulation in Immersive Virtual Reality: Randomized Controlled Trial
J Med Internet Res 2025;27:e63137
URL: https://www.jmir.org/2025/1/e63137
doi:10.2196/63137
PMID:

©Chenxi Wang, Lanqi Gao, Chuan Zhang, Jun Li, Jixin Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prompt Framework for Extracting Scale-Related Knowledge Entities from Chinese Medical Literature: Development and Evaluation Study

Jie Hao[1*], PhD; Zhenli Chen[1*], BS; Qinglong Peng[2,3], BS; Liang Zhao[4], MD; Wanqing Zhao[1], MS; Shan Cong[2,3], PhD; Junlian Li[1], PhD; Jiao Li[1], PhD; Qing Qian[1], PhD; Haixia Sun[1], PhD

[1]Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Qingdao Innovation and Development Center, Harbin Engineering University, Qingdao, China

[3]College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, China

[4]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Haixia Sun, PhD
Institute of Medical Information/Medical Library
Chinese Academy of Medical Sciences & Peking Union Medical College
No. 3, Yabao Road
Chaoyang District
Beijing, 100020
China
Phone: 86 01052328741
Email: sun.haixia@imicams.ac.cn

Abstract

Background: Measurement-based care improves patient outcomes by using standardized scales, but its widespread adoption is hindered by the lack of accessible and structured knowledge, particularly in unstructured Chinese medical literature. Extracting scale-related knowledge entities from these texts is challenging due to limited annotated data. While large language models (LLMs) show promise in named entity recognition (NER), specialized prompting strategies are needed to accurately recognize medical scale-related entities, especially in low-resource settings.

Objective: This study aims to develop and evaluate MedScaleNER, a task-oriented prompt framework designed to optimize LLM performance in recognizing medical scale-related entities from Chinese medical literature.

Methods: MedScaleNER incorporates demonstration retrieval within in-context learning, chain-of-thought prompting, and self-verification strategies to improve performance. The framework dynamically retrieves optimal examples using a k-nearest neighbors approach and decomposes the NER task into two subtasks: entity type identification and entity labeling. Self-verification ensures the reliability of the final output. A dataset of manually annotated Chinese medical journal papers was constructed, focusing on three key entity types: scale names, measurement concepts, and measurement items. Experiments were conducted by varying the number of examples and the proportion of training data to evaluate performance in low-resource settings. Additionally, MedScaleNER's performance was compared with locally fine-tuned models.

Results: The CMedS-NER (Chinese Medical Scale Corpus for Named Entity Recognition) dataset, containing 720 papers with 27,499 manually annotated scale-related knowledge entities, was used for evaluation. Initial experiments identified GLM-4-0520 as the best-performing LLM among six tested models. When applied with GLM-4-0520, MedScaleNER significantly improved NER performance for scale-related entities, achieving a macro F_1-score of 59.64% in an exact string match with the full training dataset. The highest performance was achieved with 20-shot demonstrations. Under low-resource scenarios (eg, 1% of the training data), MedScaleNER outperformed all tested locally fine-tuned models. Ablation studies highlighted the importance of demonstration retrieval and self-verification in improving model reliability. Error analysis revealed four main types of mistakes: identification errors, type errors, boundary errors, and missing entities, indicating areas for further improvement.

XSL•FO
RenderX

Conclusions: MedScaleNER advances the application of LLMs and prompts engineering for specialized NER tasks in Chinese medical literature. By addressing the challenges of unstructured texts and limited annotated data, MedScaleNER's adaptability to various biomedical contexts supports more efficient and reliable knowledge extraction, contributing to broader measurement-based care implementation and improved clinical and research outcomes.

(J Med Internet Res 2025;27:e67033) doi:10.2196/67033

KEYWORDS

prompt engineering; named entity recognition; in-context learning; large language model; Chinese medical literature; measurement-based care; framework; prompt; prompt framework; scale; China; medical literature; MBC; LLM; MedScaleNER; retrieval; information retrieval; dataset; artificial intelligence; AI

Introduction

Measurement-based care (MBC), which involves the systematic use of standardized scales to assess patient conditions, track progress, and inform clinical decisions, has been shown to enhance patient outcomes and optimize health care processes [1]. Despite its significant benefits, MBC remains underused worldwide, with less than 20% of health practitioners incorporating it into their routine practice [2]. A primary barrier is the lack of accessible and comprehensive knowledge about these scales [3,4]. Clinicians are often unaware of which scales are suitable for specific conditions and lack a clear understanding of the concepts and items within those scales [5]. Acquiring this knowledge requires considerable time and effort in searching and reviewing various unstructured documents, such as technical reports, academic papers, and manuals. This not only adds to the workload of already busy medical professionals [6] but also hinders the widespread adoption of MBC [7]. Transforming unstructured scale-related documents into computable and accessible knowledge systems, such as knowledge graphs, could help alleviate the burden on practitioners and promote MBC adoption [8].

The key to this transformation lies in accurately extracting scale-related knowledge entities from unstructured medical texts. However, several challenges complicate this task. The complexity of medical language, coupled with the variety of scale-related entities, including scale names, measurement concepts, and measurement items, makes accurate extraction difficult [1]. For instance, the entity "scale name" may refer solely to the scale itself or include additional details such as its developer, language, version, or population-specific characteristics. Furthermore, extracting knowledge entities from Chinese medical texts introduces additional difficulties due to linguistic variations and the limited availability of annotated data specific to medical scales in Chinese [9]. Traditional information extraction methods often depend on extensive data annotation and model fine-tuning, which are resource-intensive and struggle to adapt to new tasks or domains.

Addressing these challenges requires innovative solutions that can handle the complexity and variability of medical scale information, particularly in Chinese, with limited resources. Large language models (LLMs) such as GPT [10], GLM [11], LLaMA [12], and Qwen [13] have recently demonstrated impressive capabilities in natural language understanding and in-context learning (ICL) [14], enabling them to perform complex language tasks with minimal fine-tuning and limited annotated data [15,16]. By leveraging prompt engineering, these models can generate targeted outputs from well-crafted inputs, reducing the need for extensive datasets and making them particularly valuable in low-resource scenarios [17-19]. Moreover, LLMs have shown the ability to perform tasks in zero-shot and few-shot settings [20,21], allowing them to generalize from minimal demonstrations (examples) [22]. However, their effectiveness in information extraction tasks is highly dependent on the quality of the prompt design, including the task instructions, example selection, and output formats [23]. In few-shot settings, selecting the most appropriate examples is crucial for guiding the LLM toward accurate information extraction [24]. This selection process is particularly challenging in specialized domains, where the examples should not only be relevant but also representative of the task's complexity.

To improve the performance of LLMs in information extraction tasks, researchers have explored techniques such as demonstration retrieval within the ICL paradigm [25]. In this approach, relevant examples are dynamically selected from a pool of annotated data based on their similarity to the input text, addressing the issue of sample representativeness [26]. Moreover, techniques such as chain-of-thought (CoT) prompting [27] and self-verification [25] have been used to enhance the accuracy and robustness of domain-specific tasks. CoT prompting involves breaking down complex tasks into subtasks, and guiding the LLM step-by-step through multiturn dialogue [27], while self-verification allows the model to review and refine its outputs [28]. Despite these promising strategies, most existing research focuses on extracting common medical entities such as diseases, symptoms, drugs, and procedures [29], with limited attention to the extraction of scale-related entities.

Given the importance of task-oriented prompt design and the linguistic complexities of Chinese medical literature, there is an urgent need to develop a specialized framework for medical scale–related knowledge extraction. In this study, we propose MedScaleNER, a task-oriented prompt framework tailored for named entity recognition (NER) of medical scales in Chinese medical literature. MedScaleNER incorporates demonstration retrieval, CoT prompting, and self-verification strategies to tackle the specific challenges associated with extracting scale-related knowledge entities in Chinese. By dynamically selecting representative examples, the framework enhances the generalization capabilities of LLMs and improves extraction performance in few-shot scenarios. CoT prompting decomposes the scale-related NER task into manageable subtasks, easing

XSL•FO
RenderX

the cognitive load on LLMs, while self-verification ensures output reliability.

This study emphasizes the significance of prompt design in LLM-based information extraction, particularly in specialized domains with limited annotated data. By combining demonstration retrieval with advanced prompt strategies, we aim to overcome the challenges posed by data scarcity and the linguistic variations of Chinese medical literature. To facilitate evaluation, we constructed a manually annotated corpus of Chinese medical scales, covering three key entity types of scale names, measurement concepts, and measurement items. We conducted an in-depth assessment of MedScaleNER's effectiveness on this self-built dataset, examining the impact of the number of demonstrations, the contributions of CoT and self-verification, and the annotated data size required for optimal performance. Our approach contributes to building comprehensive scale knowledge systems, supporting clinicians and researchers in clinical and research efforts, promoting MBC adoption, and ultimately improving patient care.

Methods

Overview

The workflow of the proposed MedScaleNER prompt framework is illustrated in Figure 1 and consists of three main stages: dataset preparation and annotation, design and implementation of the MedScaleNER framework, and in-depth evaluation and comparison. The process begins with the collection of high-quality Chinese journal papers focused on medical scales. These papers are preprocessed and manually annotated to extract three key types of scale-related entities: scale names, measurement concepts, and measurement items. This manually annotated corpus fills the gap caused by the limited availability of annotated data in this area, while also reflecting the complexities unique to Chinese medical literature.

Figure 1. Workflow of the MedScaleNER prompt framework. LLM: large language model.

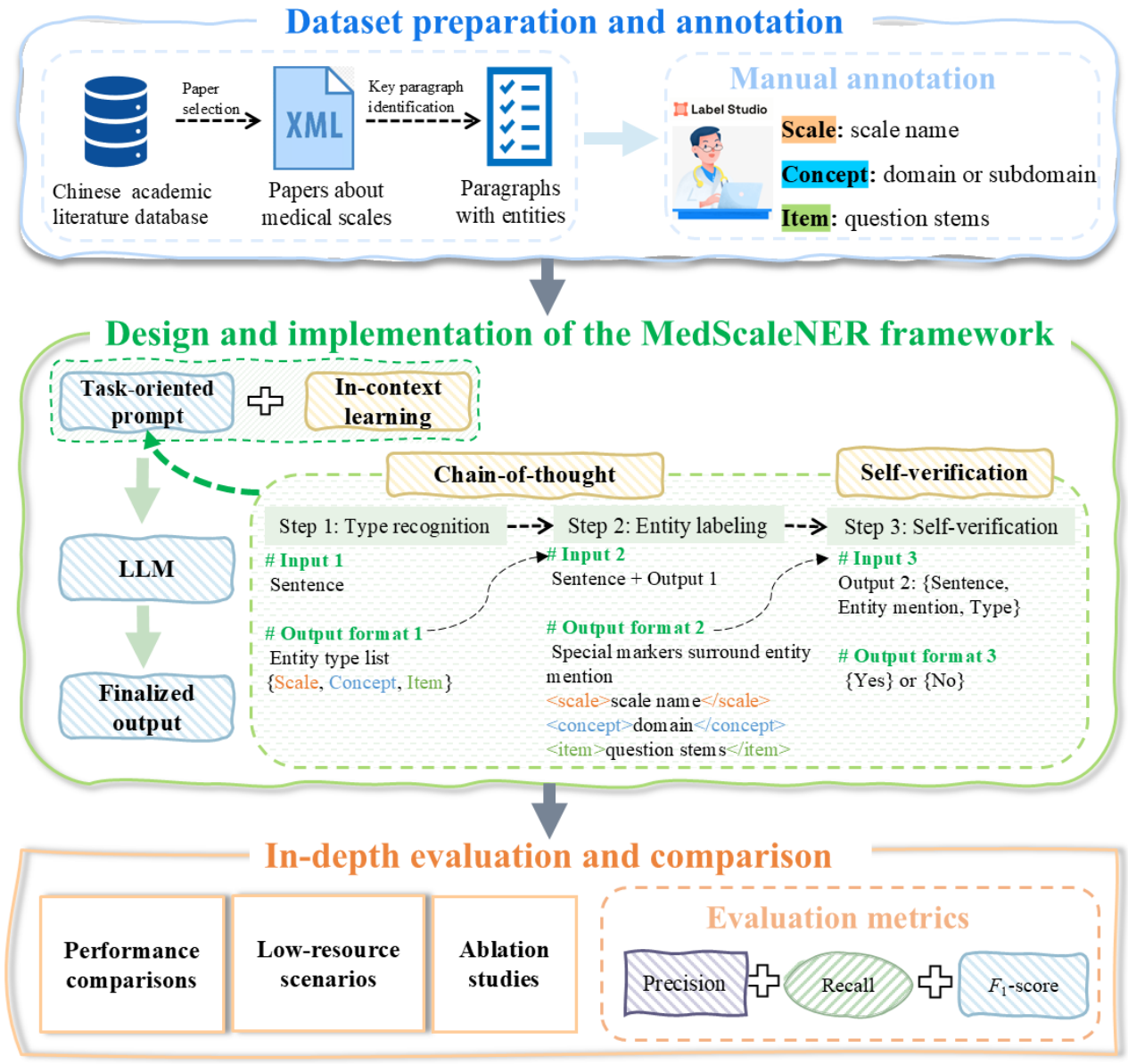

XSL•FO
RenderX

To address the task of medical scale–related NER, we introduce the MedScaleNER framework, which incorporates demonstration retrieval within the ICL paradigm, CoT prompting, and self-verification techniques. The framework selects relevant examples dynamically, breaks down complex NER tasks into manageable subtasks, and improves the reliability of outputs through self-verification. We evaluate MedScaleNER comprehensively, including comparisons of performance with varying numbers of retrieved demonstrations, ablation studies to determine the impact of CoT and self-verification, and assessments of its effectiveness in low-resource scenarios with different training data sizes. Additionally, we benchmark its performance against traditional fine-tuned LLMs in local.

Formally, the task is defined as follows. Given a collection of Chinese academic documents related to medical scales, denoted as D, where each document D_i consists of a sequence of sentences $S = \{s_1, s_2, ..., s_n\}$, and a set of entity types $T = \{\text{scale, concept, item}\}$, the goal of MedScaleNER is to identify all entities e_i within D and assign the appropriate type $t_i \in T$ to each identified entity.

Ethical Considerations

This study used only publicly available published papers from the China National Knowledge Infrastructure, which consist of academic literature and do not contain real patient information. Since the data is publicly accessible and does not involve human participants or private data, ethical approval was not required.

Dataset Preparation and Annotation

Due to the lack of annotated datasets for knowledge entity recognition in Chinese medical scales, we constructed a manually annotated corpus from full-text medical journal papers. The annotation focused on three key types of knowledge entities within medical scales: scale name, measurement concept, and measurement item. The scale name refers to the official or widely recognized title of the medical scale used in MBC, such as “The M. D. Anderson Symptom Inventory.” The measurement concept is defined as the broader theoretical or clinical construct that the scale is designed to assess, such as anxiety or cognitive function. The measurement item, on the other hand, refers to the individual questions within the scale that evaluate specific aspects of the measurement concept.

We began by retrieving abstracts of Chinese core medical journal papers from the China National Knowledge Infrastructure [30], which is a Chinese academic journal full-text database, targeting scale development research. The search was conducted using the “Abstract” and “Chinese Library Classification” criteria. From the retrieved papers, we selected the top three subfields within the Chinese Library Classification R code (Medicine and Health) based on literature frequency. Each abstract was manually reviewed to ensure the inclusion of original research papers, and the corresponding full texts were obtained in XML format. A detailed analysis of these full texts revealed that the Methods, Results, and Discussion sections contained a higher density of mentions related to scale names, measurement concepts, and items. Compared to scale names, mentions of concepts and items were less frequent, with items being particularly sparse.

To improve the balance and density of these entities, we extracted paragraphs specifically from the Methods, Results, and Discussion sections based on their XML structure. We then used key clue words such as “dimension,” “domain,” “variable,” “concept,” “factor,” “item,” and “entry” to identify paragraphs likely to contain the targeted entities. Paragraphs containing these terms were retained for annotation, while others were excluded.

For data annotation, we used the Label Studio tool [31]. Prior to formal annotation, a preannotation phase was conducted to train annotators. During this phase, annotators were introduced to the annotation scheme, guidelines (summarized in Multimedia Appendix 1), and tools. Feedback from this stage was used to refine both the scheme and guidelines through discussions. In the formal annotation phase, each paper was independently annotated by two annotators. A third annotator then checked for consistency, corrected discrepancies based on either annotator’s results, addressed missed annotations, and documented uncertain cases, which were later resolved through group discussions. Cohen κ coefficient was calculated to assess annotation consistency, yielding an overall score of 0.95, indicating a high level of reliability for the constructed dataset [32]. Specifically, the type-specific Cohen κ values were 0.961 for scale entities, 0.950 for concepts, and 0.970 for items.

Design and Implementation of the MedScaleNER Framework

Overview

We developed the MedScaleNER prompt framework to identify scale-related entities in medical texts using LLMs. The framework is designed to optimize entity recognition by incorporating three key stages: zero-shot entity type recognition, few-shot entity labeling, and self-verification. To enhance LLM performance, MedScaleNER integrates CoT prompting, which helps guide the model step by step through complex tasks, reducing the cognitive load. This is achieved by first identifying entity types in a zero-shot setting and then labeling the entities with a few examples. To further improve contextual understanding, the framework dynamically retrieves relevant examples using k-nearest neighbors (KNN) and uses self-verification to minimize hallucinations and overprediction, which are common issues in NER tasks [33]. Figure 2 outlines the MedScaleNER prompt framework, which consists of four main components: demonstration retrieval, entity type recognition, entity labeling, and self-verification.

XSL•FO
RenderX

Figure 2. Design of the MedScaleNER prompt framework. KNN: k-nearest neighbors.

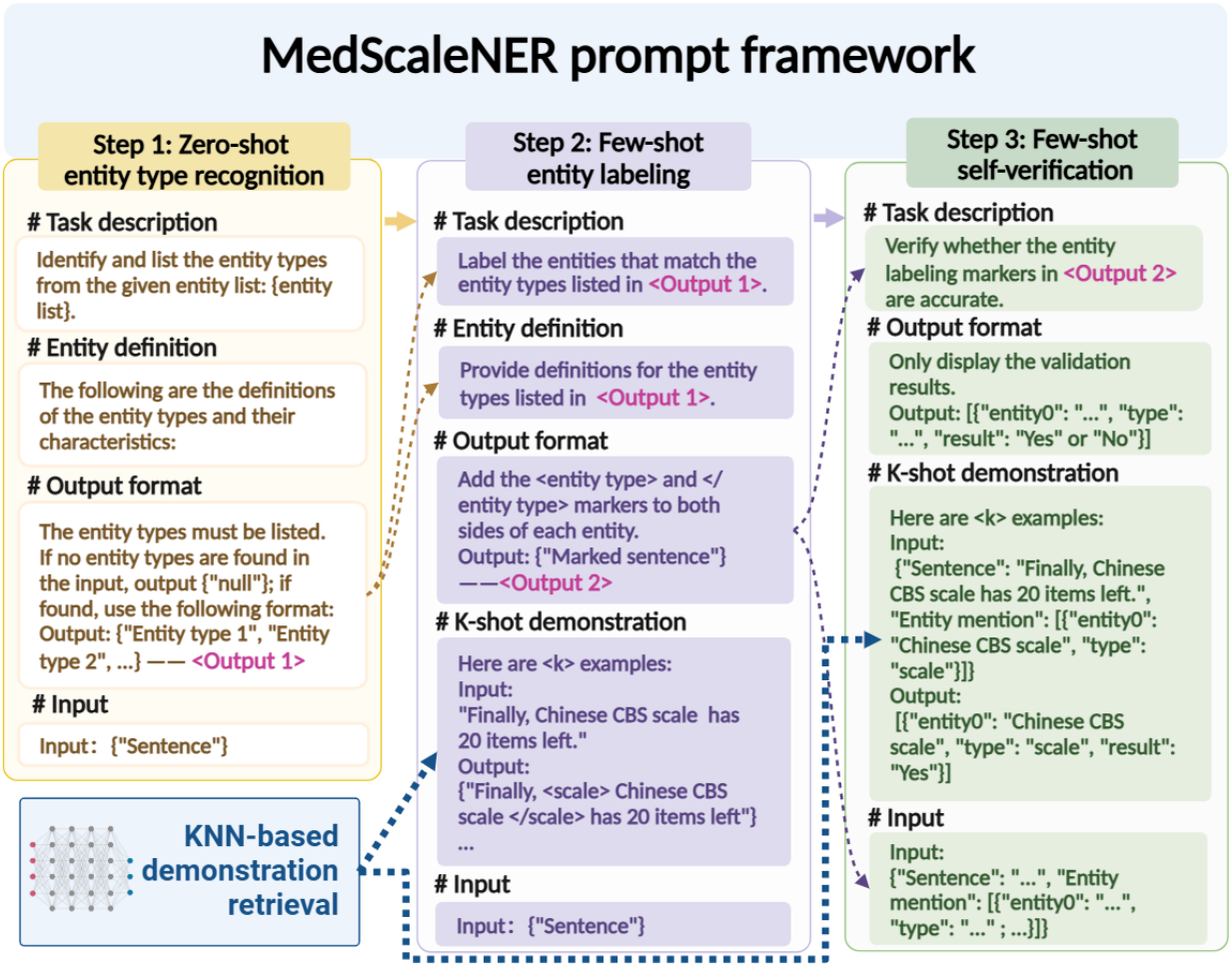

Step 1: Zero-Shot Entity Type Recognition

In the first step, we design a prompt that includes four essential components: task description, entity definitions, output format, and input text [25]. Previous research in medical entity recognition using LLMs emphasizes the critical importance of prompt design, especially in specialized tasks [34]. Therefore, in this step, the model is framed as a "sophisticated linguist and named entity annotation expert" and tasked with identifying and listing entity types present in the input text without examples. This is the zero-shot stage, where the LLM uses only the provided definitions to perform entity-type recognition without any prior labeled examples.

To improve the model's understanding of domain-specific terminology, we provide clear definitions and possible forms for each entity type. This enables the LLM to comprehend and distinguish between different medical scale–related entities. The output is structured as a list of identified entity types, with explicit instructions to avoid unnecessary information, limit responses to the provided entity list, and return "{null}" if no entities are found. The output is structured as a list of identified entity types present in the input text, without repeating types for multiple occurrences. For instance, if the entity type "scale" appears multiple times in the input, it is represented only once in the output to indicate its presence. Step 1 is instructed to limit responses to the provided entity type list and return "{null}" if no entities are found. By guiding the model through these structured prompts, we leverage CoT prompting to break down the task into manageable steps for better performance (detailed in Multimedia Appendix 2).

Step 2: Few-Shot Entity Labeling

The second step builds upon the output of step 1 by introducing few-shot entity labeling. The prompt in this stage includes the task description, entity definitions, examples (demonstrations), output format, and the input sentence. The role of the LLM remains that of a named entity annotation expert. Now, the task is to label the entities that match the entity types identified in step 1 within the input text.

Here, we use few-shot prompting because providing a small number of high-quality examples typically boosts performance, especially in specialized tasks like medical NER [35]. Few-shot prompting often achieves results comparable to those of fine-tuned models trained on hundreds of samples. However, it is sensitive to the representativeness of the examples, as well as the length of the input. To mitigate these challenges, we use KNN retrieval to dynamically select the most relevant examples from the training corpus. These examples, which are semantically similar to the input text, serve as demonstrations for the LLM to follow, guiding it in accurately labeling entities

XSL•FO
RenderX

within the text. Moreover, we incorporate CoT prompting by breaking the task into incremental steps: first identifying entity types (step 1), followed by entity labeling (step 2). Step 1 involves identifying the entity types present in the input text, which informs the candidate pool for KNN retrieval in step 2. For instance, if step 1 determines that the entity types are {scale, concept}, step 2 specifically retrieves examples containing both scale and concept entities. The LLM surrounds the identified entities with appropriate markers in the text [36], as illustrated in Figure 3, with detailed prompts provided in Multimedia Appendix 3.

Figure 3. Output format of entity labeling.

Output format

 # Example 1

Chinese:

<scale> 痴呆态度量表 **</scale>**共有20个条目，**<concept>** 痴呆知识 **</concept>**(dementia knowledge)和**<concept>** 社会舒适度 **</concept>**(social comfort)两个维度。

Translation:

<scale> Dementia Attitude Scale **</scale>** consists of 20 items, covering two dimensions: **<concept>** Dementia Knowledge **</concept>** and **<concept>** Social Comfort **</concept>**.

 # Example 2

Chinese:

公因子4 (**<concept>** 母乳喂养健康和喂养量 **</concept>**)包括2个条目,如**<item>** 母乳喂养的孩子比奶粉喂养的孩子更健康 **</item>**，**<item>** 比奶粉喂养令婴儿吃得更多 **</item>**。

Translation:

Common Factor 4 (**<concept>** Breastfeeding Health and Quantity of Feeding **</concept>**) includes two items, such as **<item>** breastfed children are healthier than formula-fed children **</item>**, and **<item>** infants eat more than formula-fed infants **</item>**.

KNN Demonstration Retrieval

For demonstration retrieval, we use KNN, a widely used method in ICL [37]. The KNN algorithm retrieves semantically similar examples from the annotated dataset to provide the LLM with contextually relevant demonstrations. We begin by generating sentence-level embeddings for both the input texts and candidate examples using the SimCSE [38] method implemented in the acge_text_embedding model [39]. The cosine similarity between the input text and each candidate example is then computed, and the top k examples with the highest similarity scores are selected.

To enable efficient retrieval from large-scale, high-dimensional embedding spaces, we use the FAISS library [40]. FAISS provides specialized data structures and algorithms for fast indexing and searching of embeddings. After indexing the training embeddings, retrieving demonstrations for a given test sentence simply involves extracting its embedding and performing a semantic similarity search against the indexed embeddings. This approach significantly reduces computational overhead by eliminating exhaustive pairwise comparisons.

To determine the optimal value of *k*, we conducted experiments with a range of different *k* values. We selected the value that maximized model performance by balancing the need for enough examples to help the LLM generalize while avoiding performance degradation caused by irrelevant or excessive examples. These examples are fed to the LLM to improve generalization and entity labeling.

Step 3: Few-Shot Self-Verification

The final step incorporates a self-verification mechanism to improve the accuracy and reliability of the labeled entities. After completing the entity labeling, the LLM undergoes a self-verification process through few-shot prompting, where it reviews its own output for correctness. In this step, the model's task in this step is to verify whether the labeled entities are accurate by responding with either "Yes" or "No" for each entity. If the answer is "No," the entity is removed from the output.

The input for this step includes both the labeled entities and their surrounding context, allowing the LLM to validate its previous output holistically. This self-verification step improves the reliability of the final results, ensuring that the identified and labeled entities meet the expected standards of accuracy (detailed prompts are provided in Multimedia Appendix 4).

In-Depth Evaluation and Comparison

To comprehensively assess the performance of MedScaleNER, we conducted an in-depth analysis. Before the formal experiments, we first identified the best-performing LLM for use in MedScaleNER by comparing various LLMs accessed via the application programming interface (API). Following this, we compared MedScaleNER's performance with that of

XSL•FO
RenderX

locally fine-tuned LLMs on the NER task. Additionally, we carried out ablation studies focusing on the two key components of MedScaleNER: CoT prompting in step 1 (zero-shot entity type recognition) and self-verification in step 3. By isolating these components, we examined their individual contributions to the overall framework, specifically their impact on entity recognition accuracy and output robustness. These studies provided valuable insights into the importance of each step in enhancing model reliability and precision.

Furthermore, we evaluated MedScaleNER in low-resource settings by varying the amount of training data and the number of demonstrations in the few-shot setting (steps 2 and 3). This analysis was essential for understanding how the framework performs under limited data conditions and testing its scalability and effectiveness when annotation resources are scarce. By experimenting with different proportions of available data and examples, we gained insights into the adaptability of MedScaleNER in resource-constrained scenarios.

For evaluation, we used precision, recall, and macro F_1-score. Precision represents the proportion of correctly predicted entities out of all entities predicted by the model. Recall is the proportion of correctly predicted entities out of all actual entities present in the dataset. Macro F_1-score is the harmonic mean of precision and recall, averaged across all entity classes to account for imbalanced class distributions. We determined the correctness of entity recognition using exact string matching, meaning only perfect matches between the model's predictions and the ground truth were considered correct. This strict evaluation method ensured a high standard for assessing model performance across all comparisons, providing a clear and objective measure of MedScaleNER's effectiveness.

Results

Summary Statistics of CMedS-NER

We constructed the CMedS-NER (Chinese Medical Scale Corpus for Named Entity Recognition) dataset specifically for the NER task in the context of Chinese medical scales. The dataset consists of 720 full-text Chinese academic papers focused on medical scales, which include 5582 paragraphs and 22,743 sentences. After conducting a concordance test and making necessary emendations, CMedS-NER contained a total of 27,499 entity mentions. These consisted of 12,340 mentions of scales, 11,968 mentions of concepts, and 3191 mentions of items. For evaluation purposes, the dataset was randomly split at the document level into 90% for training and 10% for testing. Detailed characteristic statistics of the training and test data are presented in Table 1.

Table 1. Statistics of training and test data.

Data type	Training set, n (%)	Test set, n (%)	Total, n (%)
Papers	648 (90)	72 (10)	720 (100)
Paragraphs	5055 (90.56)	527 (9.44)	5582 (100)
Sentences	20,454 (89.94)	2289 (10.06)	22,743 (100)
Entities			
Scale	11,106 (90)	1234 (10)	12,340 (100)
Concept	10,836 (90.54)	1132 (9.46)	11,968 (100)
Item	2947 (92.35)	244 (7.65)	3191 (100)
All	24,889 (90.51)	2610 (9.49)	27,499 (100)

LLM Selection and Experimental Setup

To determine the best-performing LLM for the MedScaleNER framework, we conducted preliminary experiments with six generative LLMs: GPT-3.5-turbo, GLM-4-0520, ERNIE-Bot-turbo, Moonshot-v1-8k, AGI Sky-Chat-3.0, and Qwen-turbo-0624. These models were accessed via APIs and evaluated on randomly selected sentences from the CMedS-NER test set, which included ten scale entities. Among the tested models, GLM-4-0520 performed the best, accurately recognizing nine out of ten scale entities, Qwen-turbo-0624 followed, identifying eight entities (complete results are provided in Multimedia Appendix 5). Based on this superior performance, GLM-4-0520 was selected for subsequent experiments. For the GLM-4-0520 setup, we used temperature sampling, setting the temperature parameter to 0.02 and the max_tokens parameter to 2048, while leaving all other hyperparameters at their default values.

In addition to evaluating the API-accessed LLM, we implemented local fine-tuning for four models: GLM-4-9B-Chat [41], Qwen2-7B [42], BiLSTM-CRF [43] (Chinese-BERT-wwm), and W2NER [44] (MacBERT). Fine-tuning was performed using Pytorch 1.12.1+cu11.6 on NVIDIA RTX A6000 graphics processing units. For these locally fine-tuned models, hyperparameters were optimized using empirical tuning methods to achieve the best performance on the CMedS-NER dataset. Detailed hyperparameter settings for each model are provided in Multimedia Appendix 6.

Performance Comparisons

Optimal k-Shot Demonstration Selection

To determine the optimal number of demonstrations (k) for few-shot learning, we tested various k-shot settings (0, 5, 10, 15, 20, and 25) on a randomly selected set of 100 sentences from the test set. As shown in Table 2, overall performance improved as the number of demonstrations increased, with the

XSL•FO
RenderX

highest F_1-score of 81.23% achieved at 20-shot. However, different entity types peaked at different k values. For instance, the F_1-score for concepts peaked at 10-shot (85.07%), while the F_1-score for scales and items reached their highest performance at 20-shot, with scores of 77.27% and 84.21%, respectively.

Table 2. Entity extraction performance with different k-shot values.

k-shot	All, F_1-score (%)	Scale, F_1-score (%)	Concept, F_1-score (%)	Item, F_1-score (%)
0	17.11	23.26	28.07	0
5	68.74	61.87	74.34	70
10	67.59	67.69	*85.07* [a]	50
15	78.55	71.88	83.78	80
20	*81.23*	*77.27*	82.19	*84.21*
25	71.06	75.19	78.73	59.26

[a]The best performance is italicized.

Low-Resource Comparison

To evaluate MedScaleNER's performance in low-resource scenarios, we trained the model using different proportions of the training data (1%, 5%, 10%, 50%, and 100%) and assessed its performance on the test set. As presented in Figure 4A, the overall F_1-score increased as more training data was used. Notably, significant performance gains were observed when increasing the training data from 1% (205 sentences) to 5% (1023 sentences), with the overall F_1-score rising from 48.22% to 58.29%, precision improving from 52.33% to 57.40%, and recall jumping from 45.35% to 60.06%. Beyond this point, improvements plateaued, with only a 1.35% increase in the F_1-score between 5% and 100% of the training data (from 58.29% to 59.64%). A similar trend was observed for precision and recall, although precision dropped slightly at 10% of the training data. When examining scale and concept entities (Figure 4B and C), the same pattern emerged: a significant improvement from 1% to 5% of the training data, followed by minimal gains from 5% to 100%. However, for item entities (Figure 4D), precision, recall, and F_1-scores slightly declined as the training data increased from 5% to 100%.

XSL•FO
RenderX

Figure 4. Low-resource performance of MedScaleNER: precision (P), recall (R), and macro F_1-scores (F_1) across different proportions of training data for (A) all entities, (B) scales, (C) concepts, and (D) items.

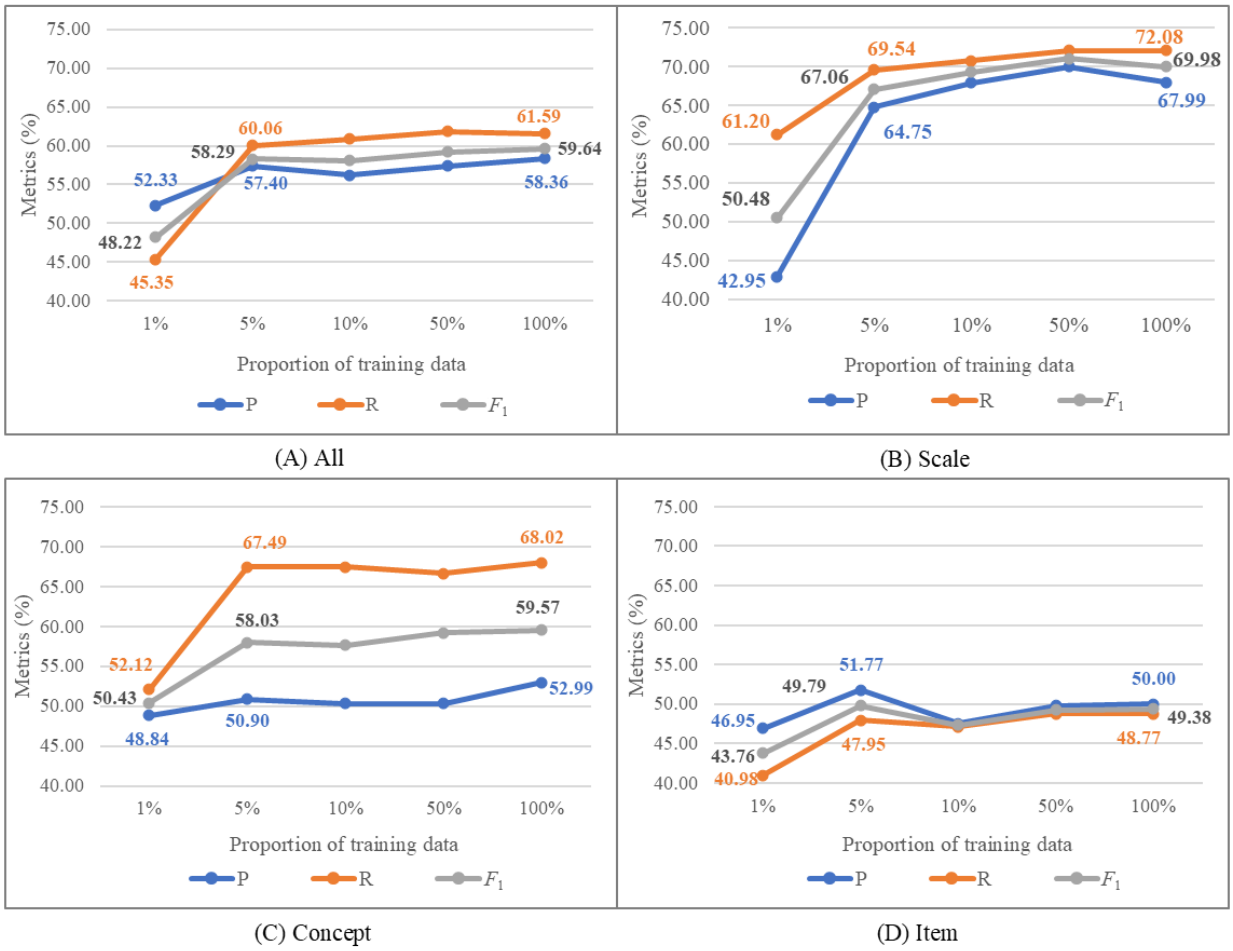

Ablation Studies

To evaluate the contributions of different components within the MedScaleNER framework, we conducted ablation studies to examine the impact of CoT prompting and self-verification under various low-resource settings. We tested the GLM-4-0520's performance with and without these components using different proportions of the training data (1%, 5%, 10%, 50%, and 100%). The baseline model involved directly prompting the task description and labeling entities in the input using 20 examples.

As shown in Table 3, incorporating self-verification consistently improved F_1-scores across all training data sizes. With 100% of the training data, self-verification led to a 0.89% increase in the F_1-score, while breaking down the task into subtasks using CoT resulted in a 2.11% increase. However, in extremely low-resource scenarios (eg, 1% of the training data), adding CoT without self-verification did not enhance performance and slightly decreased the F_1-score. The combination of both CoT and self-verification yielded the best performance across all training data sizes. In low-resource settings, self-verification had a significant impact. It increased the F_1-score of MedScaleNER by 7.92% with 1% of the training data and by 3.27% with 5% of the training data, compared to configurations without it (ie, baseline + CoT).

XSL•FO
RenderX

Table 3. Ablation study results: P[a], R[b], and F_1[c] scores for different components of MedScaleNER across different proportions of training data.

Proportion of training data (n of sentences)	Baseline			Baseline + self-verification			Baseline+ CoT[d]			MedScaleNER		
	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)	P (%)	R (%)	F_1 (%)
1% (205)	48.21	45.70	46.63	51.42	45.59	48.14	36.52	45.46	40.30	52.33	45.35	48.22
5% (1023)	44.21	66.41	52.35	47.61	66.71	55.07	51.10	60.20	55.02	57.40	60.06	58.29
10% (2045)	45.88	66.11	54.37	47.99	65.54	55.03	50.84	60.95	55.13	56.22	60.84	58.11
50% (10,227)	45.12	68.08	54.32	48.86	68.41	56.76	52.49	62.16	56.69	57.41	61.83	59.23
100% (20,454)	45.21	69.50	54.78	47.79	69.31	55.67	53.11	61.73	56.89	58.36	61.59	59.64

[a]P: precision.

[b]R: recall.

[c]F_1: macro F_1-score.

[d]CoT: chain-of-thought.

Comparison With Local Fine-Tuned Models

We compared MedScaleNER with several locally fine-tuned models on the CMedS-NER dataset, including GLM-4-9B-Chat, Qwen2-7B, BiLSTM-CRF (Chinese-BERT-wwm), and W2NER (MacBERT). Both GLM-4-9B-Chat and Qwen2-7B were fine-tuned using the low-rank adaptation method with a parameter-efficient fine-tuning strategy with identical fine-tuning parameters, ensuring a fair comparison. After fine-tuning, we prompted the fine-tuned GLM-4-9B-Chat and Qwen2-7B for the NER task using a similar prompt structure as our Baseline, but without KNN retrieval.

As shown in Figure 5, MedScaleNER achieved an overall F_1-score of 59.64%, which is lower than the fine-tuned Qwen2-7B (79.91%), GLM-4-9B-Chat (80.34%), BiLSTM-CRF (80.99%), and W2NER (81.38%). Notably, under low-resource scenarios (eg, using only 1% of the training data), MedScaleNER significantly outperformed the other fine-tuned models. At 5% of the training data, while MedScaleNER's F_1-score was lower than W2NER and BiLSTM-CRF, it remained substantially higher than Qwen2-7B and stayed competitive with GLM-4-9B-Chat.

XSL•FO
RenderX

Figure 5. F_1-score comparison of MedScaleNER and locally fine-tuned models across different proportions of training data.

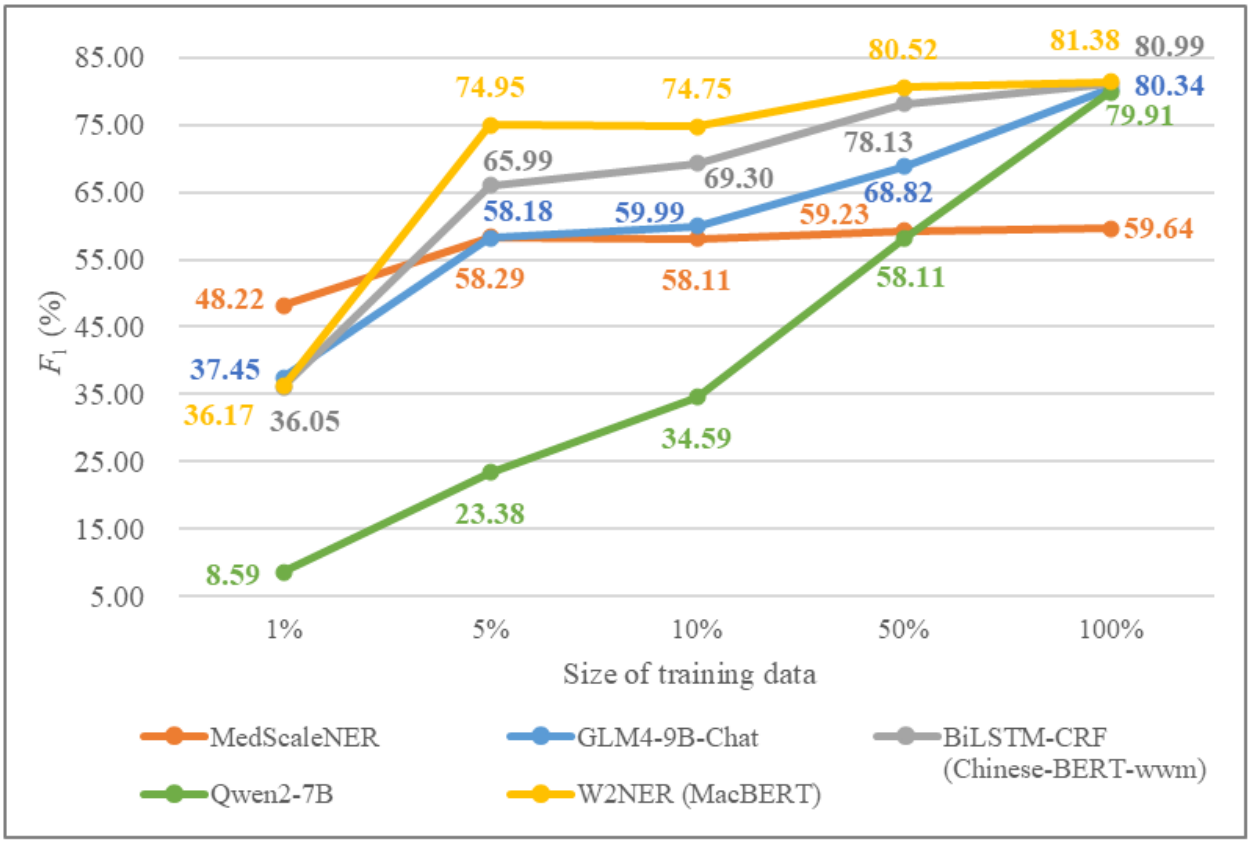

Error Analysis

We conducted an error analysis by manually reviewing 300 randomly selected sentences from the model outputs to identify common types of mistakes and areas for improvement. The errors were classified into four main types: (1) identification errors, where nonentity terms were incorrectly identified as entities; (2) type errors, where entities were correctly identified but assigned the wrong entity type; (3) boundary errors, which involved incorrect determination of the start and end positions of entities; and (4) missing entities, where entities present in the text were not identified by the model.

Figure 6 provides examples of each error type, illustrating the nature of these mistakes. Identification errors were the most common and often resulted from ambiguous entity definitions. For example, generic terms like “item” or “scale” were sometimes misinterpreted as specific entities due to their inclusion in prompt definitions. Type errors occurred when entities were recognized but misclassified. For instance, “overall evaluation of the quality of nursing services” was mistakenly labeled as a concept rather than an item. Boundary errors included incorrect inclusion or exclusion of surrounding text or punctuation, such as parentheses or modifiers that should not be part of the entity span. Finally, missing entities were frequently associated with English names or abbreviations of scales and items, especially in cases involving long or complex strings.

XSL•FO
RenderX

Figure 6. Examples of the four main error types identified in MedScaleNER: identification errors, type errors, boundary errors, and missing entities.

Error types	Chinese	Translation
Identification error	**Ground truth:** 经上述步骤形成的量表包括5个维度,20个条目。[Null] **MedScaleNER:** 经上述步骤形成的量表包括5个维度,20个条目。[Scale]	**Ground truth:** The scale formed by the above steps consists of 5 dimensions and 20 entries. [Null] **MedScaleNER:** The scale formed by the above steps consists of 5 dimensions and 20 entries. [Scale]
Type error	**Ground truth:** 条目21为“护理服务质量总体评价”,采用连续区间标度法评价... [Item] **MedScaleNER:** 条目21为“护理服务质量总体评价”,采用连续区间标度法评价... [Concept]	**Ground truth:** Item 21 is “overall evaluation of the quality of nursing services”, evaluated using the continuous interval scale method... [Item] **MedScaleNER:** Item 21 is “overall evaluation of the quality of nursing services”, evaluated using the continuous interval scale method... [Concept]
Boundary error	**Ground truth:** 本研究选用中文版Cohen-Mansfield激越问卷（CMAI）作为校标量表。[Scale] ; [Scale] **MedScaleNER:** 本研究选用中文版Cohen-Mansfield激越问卷（CMAI）作为校标量表。[Scale]	**Ground truth:** Chinese version of the Cohen-Mansfield Agitation Inventory (CMAI) was used as the calibration scale for this study. [Scale] ; [Scale] **MedScaleNER**: Chinese version of the Cohen-Mansfield Agitation Inventory (CMAI) was used as the calibration scale for this study. [Scale]
Missing	**Ground truth:** 公因子5(工作与饮酒影响母乳喂养)包括2个条目,如母亲恢复工作时奶粉喂养是最佳选择和母亲若偶然饮酒不可以喂母乳。[Concept] ; [Item] ; [Item] **MedScaleNER:** 公因子5(工作与饮酒影响母乳喂养)包括2个条目,如母亲恢复工作时奶粉喂养是最佳选择和母亲若偶然饮酒不可以喂母乳。[Concept]	**Ground truth:** Factor 5 (work and alcohol consumption affects breastfeeding) includes two items, such as formula feeding is the best option when the mother returns to work, and breastfeeding is not possible if the mother is an occasional drinker. [Concept] ; [Item] ; [Item] **MedScaleNER:** Factor 5 (work and alcohol consumption affects breastfeeding) includes two items, such as formula feeding is the best option when the mother returns to work, and breastfeeding is not possible if the mother is an occasional drinker. [Concept]

Discussion

Principal Findings

This study is among the first to explore the use of LLMs and prompt engineering for NER tasks related to Chinese medical scales. We proposed a novel prompt framework, MedScaleNER, which enhances the adaptive learning capabilities of LLMs by dynamically retrieving optimal examples through KNN retrieval. By using a CoT strategy, the framework decomposes the complex task of entity recognition into two sequential steps: first, identifying entity types and then labeling entities. This approach strengthens the logical reasoning ability of LLMs. Additionally, incorporating self-verification mechanisms ensures the accuracy of the final recognition results, improving the reliability of the model's outputs.

Our evaluation of the self-constructed CMedS-NER dataset demonstrated that MedScaleNER effectively recognizes medical scale–related entities. The dataset, comprising 720 full-text Chinese academic papers with 27,499 annotated entities, is a high-quality resource for training and evaluating NER models in this specialized domain. Notably, in low-resource settings with as few as 205 sentences, MedScaleNER outperformed locally fine-tuned models such as BiLSTM-CRF (Chinese-BERT-wwm), W2NER (MacBERT), GLM-4-9B-Chat, and Qwen2-7B. When more annotated data became available (eg, 1023 sentences), MedScaleNER remained competitive. This low-resource performance is particularly significant in biomedical and clinical contexts, where domain-specific annotations are often expensive and time-consuming to produce.

Ablation studies further highlighted that KNN retrieval significantly improved performance in low-resource settings, aligning with previous findings [37] on the benefits of such strategies in ICL. Integrating CoT prompting and self-verification with KNN retrieval boosted the F_1-score by approximately 6% when using 5% of the training data. This suggests that while retrieving representative examples is crucial, the structured CoT and self-verification steps are also important, contributing to more accurate and robust entity annotation than retrieval-based demonstration alone.

Although using high-quality demonstrations improved the LLM's ability to recognize scale-related entities [25,45], performance declined when the number of examples exceeded an optimal threshold. Context length limitations, example ordering [46], and entity-type specific sensitivities influenced this trade-off. For instance, concept entities benefited from fewer examples compared to scale and item entities. It suggests that tailoring demonstration strategies by entity type could maximize performance.

Compared to traditional and fine-tuned NER methods designed for similar biomedical contexts, MedScaleNER offers several advantages. Conventional approaches often require extensive domain adaptation, large annotated corpora, or multiple rounds of fine-tuning to achieve competitive results [47,48]. In contrast,

XSL•FO
RenderX

MedScaleNER excels under low-resource settings by leveraging KNN retrieval, CoT, and self-verification. Its flexible, task-oriented design allows simple modification of entity definitions to adapt to new domains, other languages, or even other LLM backbones. This adaptability supports broader generalizability, enabling MedScaleNER to scale beyond Chinese medical scales to other medical domains and even entirely different biomedical NER tasks.

Moreover, moving toward a human-centered medical scale NER workflow is crucial [23,49]. Allowing domain experts to provide feedback, customize prompt components, control retrieval parameters, and determine when to use self-verification can improve transparency, trust, and overall user satisfaction [50,51]. Such a human-in-the-loop approach ensures that MedScaleNER remains aligned with real-world clinical and research priorities, particularly important in dynamic health care environments.

Limitations

Despite these strengths, there are limitations to this study. First, we primarily focused on the GLM-4 model, and future work should evaluate additional LLMs [52] such as LLaMA, Mistral, GPT, and PaLM, to validate generalizability. Second, our example retrieval strategy relied on KNN based on sentence similarity. Alternative retrieval strategies [53] and more advanced similarity models may further enhance performance. While we focused on three main scale-related entity types: scale names, measurement concepts, and measurement items, future research could extend this framework to other entities, such as functions, targets, and validity measures. Finally, integrating LLMs with traditional NER models could leverage the complementary strengths of both approaches, potentially resulting in more robust and accurate entity recognition systems.

Conclusions

In this study, we introduced MedScaleNER, a task-oriented prompt framework that integrates demonstration retrieval, CoT prompting, and self-verification strategies to enhance the recognition of medical scale-related entities in Chinese medical literature. Evaluated on our self-constructed CMedS-NER dataset, MedScaleNER demonstrates robust performance even with limited annotated data. By allowing simple adjustments to prompt definitions, MedScaleNER readily adapts to diverse biomedical domains, languages, and entity types, making it a resource-efficient solution for broader information extraction challenges. This adaptability supports more efficient and reliable knowledge extraction, ultimately contributing to better clinical and research outcomes in MBC. By continuing to refine and expand MedScaleNER, we aim to advance automated knowledge extraction systems and promote the widespread adoption of MBC in health care.

Acknowledgments

The authors would like to thank our colleagues Zhen Guo and Liu Shen. We are grateful to Zhen Guo for his help with the deployment of the labeling tool, and to Liu Shen for her cooperation in the acquisition of raw data. This work was supported by the National Social Science Fund of China (grant 21BTQ069), the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grants 2021-I2M-1-056 and 2021-I2M-1-057), and the National Key Research and Development Program of China (grant 2022YFC3601005).

Data Availability

The sample data and the source code are available on GitHub [54].

Authors' Contributions

HS, JH, and ZC contributed to the concept and design of the study. HS, LZ, ZC, WZ, and Junlian L prepared, annotated, and curated the dataset. JH, HS, Jiao L, and QQ developed the methodology. JH, ZC, QP, and SC performed the experiments. HS, JH, ZC, and LZ interpreted the results. JH, HS, and ZC drafted the manuscript. LZ, Junlian L, Jiao L, and QQ revised the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Annotation scheme and guidelines.
[DOCX File , 17 KB - jmir_v27i1e67033_app1.docx]

Multimedia Appendix 2
Prompts for step 1.
[DOCX File , 18 KB - jmir_v27i1e67033_app2.docx]

Multimedia Appendix 3
Prompts for step 2.

XSL•FO
RenderX

[DOCX File , 17 KB - jmir_v27i1e67033_app3.docx]

Multimedia Appendix 4
Prompts for step 3.
[DOCX File , 18 KB - jmir_v27i1e67033_app4.docx]

Multimedia Appendix 5
Preliminary experimental results for optimal model selection.
[DOCX File , 17 KB - jmir_v27i1e67033_app5.docx]

Multimedia Appendix 6
Hyperparameters of locally fine-tuned models.
[DOCX File , 16 KB - jmir_v27i1e67033_app6.docx]

References

1. DeVellis RF. Scale Development: Theory and Applications. Thousand Oaks, CA: Sage Publications, Inc; 1991.
2. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
3. Murphy JK, Michalak EE, Liu J, Colquhoun H, Burton H, Yang X, et al. Barriers and facilitators to implementing measurement-based care for depression in Shanghai, China: a situational analysis. BMC Psychiatry 2021;21(1):430 [FREE Full text] [doi: 10.1186/s12888-021-03442-5] [Medline: 34470624]
4. Jensen-Doss A, Haimes EMB, Smith AM, Lyon AR, Lewis CC, Stanick CF, et al. Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. Adm Policy Ment Health 2018 Jan;45(1):48-61 [FREE Full text] [doi: 10.1007/s10488-016-0763-0] [Medline: 27631610]
5. Shoman Y, Majery N, Otelea M, Lambreghts C, Guseva Canu I. How to identify the most suitable questionnaires and rating scales for your clinical practice or research? Int J Clin Pract 2021 Dec;75(12):e14895 [FREE Full text] [doi: 10.1111/ijcp.14895] [Medline: 34541756]
6. Jensen-Doss A, Hawley KM. Understanding clinicians' diagnostic practices: attitudes toward the utility of diagnosis and standardized diagnostic tools. Adm Policy Ment Health 2011 Nov;38(6):476-485 [FREE Full text] [doi: 10.1007/s10488-011-0334-3] [Medline: 21279679]
7. Cheung BS, Murphy JK, Michalak EE, Liu J, Yang X, Wang X, et al. Barriers and facilitators to technology-enhanced measurement based care for depression among Canadian clinicians and patients: results of an online survey. J Affect Disord 2023 Jan 01;320:1-6. [doi: 10.1016/j.jad.2022.09.055] [Medline: 36162664]
8. Sun H, Hao J, Guo Z, Shen L. Construction of a fine-grained knowledge element-based framework for knowledge representation in medical scale documents. Digital Lib Forum 2023;19(12):86-98. [doi: 10.3772/j.issn.1673-2286.2023.12.009]
9. Wenqi C, Pu H, Yule Z, Liang G, Xiaoyan L. Named entity recognition from Chinese medical literature based on deep learning method. 2023 Presented at: China Automation Congress (CAC); November 19, 2023; Chongqing, China. [doi: 10.1109/CAC59555.2023.10451997]
10. Yenduri G, Ramalingam M, Selvi GC, Supriya Y, Srivastava G, Maddikunta, PRK, et al. Generative pre-trained transformer: a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. ArXiv. Preprint posted online on May 11, 2013 2023 [FREE Full text] [doi: 10.1109/access.2024.3389497]
11. Du Z, Qian Y, Liu X, Ding M, Qiu J, Yang Z, et al. GLM: general language model pretraining with autoregressive blank infilling. ArXiv. Preprint posted online on March 18, 2021 2021. [doi: 10.48550/arXiv.2103.10360]
12. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. LLaMA: open and efficient foundation language models. ArXiv. Preprint posted online on February 27, 2023 2023. [doi: 10.48550/arXiv.2302.13971]
13. Bai J, Bai S, Chu Y, Cui Z, Dang K, Deng X, et al. Qwen technical report. ArXiv. Preprint posted online on September 28, 2023 2023. [doi: 10.5260/chara.21.2.8]
14. Rosenbloom L. arXiv. Charleston Adv 2019 Oct 01;21(2):8-10. [doi: 10.5260/chara.21.2.8]
15. Dagdelen J, Dunn A, Lee S, Walker N, Rosen AS, Ceder G, et al. Structured information extraction from scientific text with large language models. Nat Commun 2024 Feb 15;15(1):1418 [FREE Full text] [doi: 10.1038/s41467-024-45563-x] [Medline: 38360817]
16. Hu Y, Chen Q, Du J, Peng X, Keloth VK, Zuo X, et al. Improving large language models for clinical named entity recognition via prompt engineering. J Am Med Inform Assoc 2024 Sep 01;31(9):1812-1820. [doi: 10.1093/jamia/ocad259] [Medline: 38281112]
17. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. ArXiv 2015. [doi: 10.5860/choice.189890]

XSL•FO
RenderX

18. Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, et al. PaLM: scaling language modeling with pathways. ArXiv. Preprint posted online on April 05, 2022 2022. [doi: 10.48550/arXiv.2204.02311]
19. Shi Z, Zhu L, Le X. Material information extraction based on localarge language model and prompt engineering. Data Anal Knowl Discovery 2024;8(7):23-31. [doi: 10.11925/infotech.2096-3467.2023.1119]
20. Jie H, Zhiqiang M, Haixia S, Zhenli C, Jiao L. ChatGPT and zero-shot prompt-based structured information extraction for clinical scale items. Lib Inf Serv 2024;68(22):139-152. [doi: 10.12074/202408.00017V1]
21. Fan W, Ding Y, Ning L, Wang S, Li H, Yin D, et al. A survey on rag meeting LLMs: towards retrieval-augmented large language models. 2024 Presented at: KDD '24: The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; August 25-29, 2024; Barcelona, Spain p. 6491-6501. [doi: 10.1145/3637528.3671470]
22. Wei X, Cui X, Cheng N, Wang X, Zhang X, Huang S, et al. ChatIE: zero-shot information extraction via chatting with ChatGPT. ArXiv. Preprint posted online on February 20, 2023 2023. [doi: 10.48550/arXiv.2302.10205]
23. Zaghir J, Naguib M, Bjelogrlic M, Névéol A, Tannier X, Lovis C. Prompt engineering paradigms for medical applications: scoping review. J Med Internet Res 2024;26:e60501 [FREE Full text] [doi: 10.2196/60501] [Medline: 39255030]
24. Ge Y, Guo Y, Yang YC, Al-Garadi MA, Sarker A. A comparison of few-shot and traditional named entity recognition models for medical text. 2022 Presented at: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI); June 11-14, 2022; Rochester, MN p. 84-89 URL: https://europepmc.org/abstract/MED/37641590 [doi: 10.1109/ichi54592.2022.00024]
25. Wang S, Sun X, Li X, Ouyang R, Wu F, Zhang T, et al. GPT-NER: named entity recognition via large language models. ArXiv. Preprint posted online on April 20, 2023 2023. [doi: 10.48550/arXiv.2304.10428]
26. Xu J, Le H. Generating representative samples for few-shot classification. 2022 Presented at: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 18-24, 2022; New Orleans, LA p. 8993-9003. [doi: 10.1109/cvpr52688.2022.00880]
27. Li M, Zhou H, Yang H, Zhang R. RT: a retrieving and chain-of-thought framework for few-shot medical named entity recognition. J Am Med Inform Assoc 2024;31(9):1929-1938. [doi: 10.1093/jamia/ocae095] [Medline: 38708849]
28. Gero Z, Singh C, Cheng H, Naumann T, Galley M, Gao J, et al. Self-verification improves few-shot clinical information extraction. ArXiv. Preprint posted online on May 30, 2023 2023. [doi: 10.48550/arXiv.2306.00024]
29. Wen C, Chen T, Jia X, Zhu J. Medical named entity recognition from un-labelled medical records based on pre-trained language models and domain dictionary. Data Intell 2021;3(3):417. [doi: 10.1162/dint_a_00105]
30. China National Knowledge Infrastructure. URL: https://www.cnki.net/ [accessed 2024-09-28]
31. Open source data labeling. Label Studio. URL: https://humansignal.com/amplitude/ [accessed 2024-05-01]
32. Artstein R, Poesio M. Inter-coder agreement for computational linguistics. Comput Ling 2008;34(4):555-596. [doi: 10.1162/coli.07-034-r2]
33. Wei Z, Guo D, Huang D, Zhang Q, Zhang S, Jiang K, et al. Detecting and mitigating the ungrounded hallucinations in text generation by LLMs. 2024 Presented at: AISNS '23: Proceedings of the 2023 International Conference on Artificial Intelligence, Systems and Network Security; June 01, 2024; Mianyang, China p. 77-81. [doi: 10.1145/3661638.3661653]
34. Shen Y, Tan Z, Wu S, Zhang W, Zhang R, Xi Y, et al. PromptNER: prompt locating and typing for named entity recognition. 2023 Presented at: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); July, 2023; Toronto, ON p. 12492-12507. [doi: 10.18653/v1/2023.acl-long.698]
35. He K, Mao R, Huang Y, Gong T, Li C, Cambria E. Template-free prompting for few-shot named entity recognition via semantic-enhanced contrastive learning. IEEE Trans Neural Networks Learn Syst 2024;35(12):18357-18369. [doi: 10.1109/tnnls.2023.3314807]
36. Tsujimura T, Yamada K, Ida R, Miwa M, Sasaki Y. Contextualized medication event extraction with striding NER and multi-turn QA. J Biomed Inform 2023;144:104416 [FREE Full text] [doi: 10.1016/j.jbi.2023.104416] [Medline: 37321443]
37. Zhao W, Liu Y, Wan Y, Wang Y, Wu Q, Deng Z, et al. kNN-ICL: Compositional task-oriented parsing generalization with nearest neighbor in-context learning. 2024 Presented at: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers); June, 2024; Mexico City, Mexico p. 326-337. [doi: 10.18653/v1/2024.naacl-long.19]
38. Gao T, Yao X, Chen D. SimCSE: simple contrastive learning of sentence embeddings. ArXiv. Preprint posted online on April 18, 2021 2021. [doi: 10.48550/arXiv.2104.08821]
39. Acge_text_embedding. URL: https://github.com/intsig-textin/acge_text_embedding [accessed 2024-05-20]
40. Douze M, Guzhva A, Deng C, Johnson J, Szilvasy G, Mazaré PE, et al. The Faiss library. ArXiv. Preprint posted online on January 16, 2024 2024(2). [doi: 10.5260/chara.21.2.8]
41. THUDM/glm-4-9b-chat. Hugging Face. URL: https://huggingface.co/THUDM/glm-4-9b-chat [accessed 2024-05-28]
42. Yang A, Yang B, Hui B, Zheng B, Yu B, Zhou C, et al. Qwen2 technical report. ArXiv 2019. [doi: 10.5260/chara.21.2.8]
43. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. ArXiv. Preprint posted online on August 09, 2015 2015 [FREE Full text]
44. Li J, Fei H, Liu J, Wu S, Zhang M, Teng C, et al. Unified named entity recognition as word-word relation classification. 2021 Presented at: Proceedings of the AAAI Conference on Artificial Intelligence; February 20-27, 2024; Vancouver, BC p. 10965-10973. [doi: 10.1609/aaai.v36i10.21344]

XSL•FO
RenderX

45. Qin L, Zhang A, Chen C, Dagar A, Ye W. In-context learning with iterative demonstration selection. ArXiv 2019. [doi: 10.5260/chara.21.2.8]
46. Hao Y, Sun Y, Dong L, Han Z, Gu Y, Wei F. Structured prompting: scaling in-context learning to 1,000 examples. ArXiv. Preprint posted online on December 13, 2022 2022. [doi: 10.48550/arXiv.2212.06713]
47. Divya B, Alejandro SM, Ralf K. The effects of data quality on named entity recognition. 2024 Presented at: Proceedings of the Ninth Workshop on Noisy and User-generated Text (W-NUT 2024); November 15, 2024; San Ġiljan, Malta p. 79-88 URL: https://aclanthology.org/2024.wnut-1.8/
48. Lu RS, Lin CC, Tsao HY. Empowering large language models to leverage domain-specific knowledge in E-learning. Appl Sci 2024;14(12):5264. [doi: 10.3390/app14125264]
49. Intelligence UAB on a governing AI for humanity: interim report. Advisory Body on Artificial Intelligence, UN. URL: https://digitallibrary.un.org [accessed 2024-09-22] [WebCite Cache ID 4062495]
50. Zhang C, Chen J, Li J, Peng Y, Mao Z. Large language models for human–robot interaction: a review. Biomimetic Intell Rob 2023;3(4):100131. [doi: 10.1016/j.birob.2023.100131]
51. Wu T, Terry M, Cai CJ. AI chains: transparent and controllable human-AI interaction by chaining large language model prompts. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29, 2022; New Orleans, LA p. 1-22. [doi: 10.1145/3491102.3517582]
52. Wang X, Salmani M, Omidi P, Ren X, Rezagholizadeh M, Eshaghi A. Beyond the limits: a survey of techniques to extend the context length in large language models. 2024 Presented at: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence Survey Track; August 3-9, 2024; Jeju, Korea. [doi: 10.24963/ijcai.2024/917]
53. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. ArXiv. Preprint posted online on December 18, 2023 2023 [FREE Full text]
54. GitHub. URL: https://github.com/Maizio/JMIR_NER.git [accessed 2025-03-11]

Abbreviations

API: application programming interface
CMedS-NER: Chinese Medical Scale Corpus for Named Entity Recognition
CoT: chain-of-thought
ICL: in-context learning
KNN: k-nearest neighbor
LLM: large language model
MBC: measurement-based care
NER: named entity recognition

Edited by A Schwartz; submitted 30.09.24; peer-reviewed by J Zaghir, G Luo; comments to author 15.11.24; revised version received 06.12.24; accepted 21.02.25; published 18.03.25.

Please cite as:
Hao J, Chen Z, Peng Q, Zhao L, Zhao W, Cong S, Li J, Li J, Qian Q, Sun H
Prompt Framework for Extracting Scale-Related Knowledge Entities from Chinese Medical Literature: Development and Evaluation Study
J Med Internet Res 2025;27:e67033
URL: https://www.jmir.org/2025/1/e67033
doi:10.2196/67033
PMID:

©Jie Hao, Zhenli Chen, Qinglong Peng, Liang Zhao, Wanqing Zhao, Shan Cong, Junlian Li, Jiao Li, Qing Qian, Haixia Sun. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Synthetic Health Care Data to Leverage Large Language Models for Named Entity Recognition: Development and Validation Study

Hendrik Šuvalov[1], MSc; Mihkel Lepson[1], MSc; Veronika Kukk[1], BSc; Maria Malk[1], MS; Neeme Ilves[1], MD; Hele-Andra Kuulmets[1], MSc; Raivo Kolde[1], PhD

Institute of Computer Science, University of Tartu, Tartu, Estonia

Corresponding Author:
Hendrik Šuvalov, MSc
Institute of Computer Science
University of Tartu
Narva mnt 28
Tartu, 51009
Estonia
Phone: 372 7375100
Email: hendrik.suvalov@ut.ee

Abstract

Background: Named entity recognition (NER) plays a vital role in extracting critical medical entities from health care records, facilitating applications such as clinical decision support and data mining. Developing robust NER models for low-resource languages, such as Estonian, remains a challenge due to the scarcity of annotated data and domain-specific pretrained models. Large language models (LLMs) have proven to be promising in understanding text from any language or domain.

Objective: This study addresses the development of medical NER models for low-resource languages, specifically Estonian. We propose a novel approach by generating synthetic health care data and using LLMs to annotate them. These synthetic data are then used to train a high-performing NER model, which is applied to real-world medical texts, preserving patient data privacy.

Methods: Our approach to overcoming the shortage of annotated Estonian health care texts involves a three-step pipeline: (1) synthetic health care data are generated using a locally trained GPT-2 model on Estonian medical records, (2) the synthetic data are annotated with LLMs, specifically GPT-3.5-Turbo and GPT-4, and (3) the annotated synthetic data are then used to fine-tune an NER model, which is later tested on real-world medical data. This paper compares the performance of different prompts; assesses the impact of GPT-3.5-Turbo, GPT-4, and a local LLM; and explores the relationship between the amount of annotated synthetic data and model performance.

Results: The proposed methodology demonstrates significant potential in extracting named entities from real-world medical texts. Our top-performing setup achieved an F_1-score of 0.69 for drug extraction and 0.38 for procedure extraction. These results indicate a strong performance in recognizing certain entity types while highlighting the complexity of extracting procedures.

Conclusions: This paper demonstrates a successful approach to leveraging LLMs for training NER models using synthetic data, effectively preserving patient privacy. By avoiding reliance on human-annotated data, our method shows promise in developing models for low-resource languages, such as Estonian. Future work will focus on refining the synthetic data generation and expanding the method's applicability to other domains and languages.

(J Med Internet Res 2025;27:e66279) doi:10.2196/66279

KEYWORDS

natural language processing; named entity recognition; large language model; synthetic data; LLM; NLP; machine learning; artificial intelligence; language model; NER; medical entity; Estonian; health care data; annotated data; data annotation; clinical decision support; data mining

XSL•FO
RenderX

Introduction

In the era of data-driven health care, the effective extraction of valuable information from electronic health records (EHRs) has become increasingly vital for informed decision-making, medical research, and public health initiatives. Named entity recognition (NER) stands as a cornerstone in the realm of natural language processing, playing a pivotal role in the automated extraction of named entities, such as diseases, medications, symptoms, and medical procedures, from unstructured text [1]. NER facilitates the transformation of text data into structured information, enabling health care professionals and researchers to carry out statistical analyses on the data.

While NER has achieved remarkable success in several languages and domains, the journey to attain proficient NER models for low-resource languages presents intricate challenges [1-3]. Estonian, a Finno-Ugric language spoken by approximately 1.1 million people, is one such language that grapples with limited linguistic resources, especially in specialized domains such as health care, largely because health records are sensitive, personalized data, which makes sharing them for labeling very difficult [4] in addition to the task being already time-consuming and expensive. The same challenges are faced with storing and sharing the annotated data. To tackle this, attempts have been made at removing or masking personally identifiable information [5,6]; however, this can still leave the data vulnerable to reconstruction attacks [6]. Despite its importance, the scarcity of annotated Estonian health care text data has hindered the development of robust NER models tailored to the unique linguistic features and domain-specific nuances of the language. At the same time, large language models (LLMs) such as GPT-4 have been shown to provide impressive results in tasks related to understanding different languages [7].

Many approaches using transformer-based models have been applied to EHRs to extract medical named entities [8-10]; however, these approaches primarily rely on human-annotated data for the training data, which we do not use for training. Most transformer-based approaches rely on using a model pretrained on a large corpus of text and fine-tuning it with the annotated data.

To tackle the problem of low-resource languages lacking annotation data, attempts have been made to improve the model using data from other languages via transfer learning [11,12], which has been shown to improve the model's performance by adding resources from languages with more data.

With the rapid increase in the performance of LLMs, attempts have also been made to leverage these models to extract named entities from texts. It has been shown that using GPT-3 for annotating the texts can be much cheaper and faster than using humans; however, there is room for improvements in the quality of the annotations made by the model [13].

Tang et al [14] used ChatGPT (OpenAI) [15] to generate synthetic medical texts with annotations to train a downstream model for different applications including NER. The performance of the resulting models was promising, considering the approach of using synthetic data. Their approach uses ChatGPT directly to generate the synthetic texts, meaning there is no potential for possible data leakage from the synthetic texts. However, because the GPT models were trained on data that are publicly available, they are not trained to understand the nuances of a specific medical domain in a low-resource language, and the synthetic data can be very different from real EHRs, possibly causing poor performance for the downstream model.

This study introduces a comprehensive pipeline using LLMs to train an NER model, enabling the annotation of targeted entities within health care texts. First, synthetic Estonian EHRs are generated by GPT-2 [16] trained locally on real data. We generate the synthetic data to avoid sharing sensitive information with third-party applications. These are passed to GPT-3.5-Turbo and GPT-4 as payload in a prompt, instructing the model to annotate the text with entities specified by the user. The responses are parsed and converted into training data for token classification with XLM-RoBERTa [17], which can be used on the downstream task of annotating real data. This approach ensures that sensitive data do not get passed to third-party LLM hosts via their application programming interface (API) calls and enables all the training to be carried out on-site.

We showcase the differences in performance between the selected prompts. We then compare the performance of GPT-3.5-Turbo, GPT-4, and a local Llama-based model for the best prompt. We continue with using the best prompt with GPT-4 to annotate generated synthetic data in batches, showcasing the improvement of the model based on the amount of training data. Finally, we discuss the performance of this method and possible room for improvements and conclude the results of this paper.

With this paper, we aim to show how we take advantage of powerful LLMs for training locally usable NER models without risking the privacy of patient records by using synthetically generated data.

Methods

Study Design

Our work follows the pipeline depicted in Figure 1. OpenAI models were selected for annotating the synthetic data due to their availability and their existing wide use in research. The clinical corpus was used for training the GPT-2 model, which we used to generate synthetic data. We then passed the synthetic data to GPT-4 through an API, asking it to annotate the data. The results were parsed and converted into training data for a downstream model, which we used to annotate real-world data.

XSL•FO
RenderX

Figure 1. Illustration of pipeline.

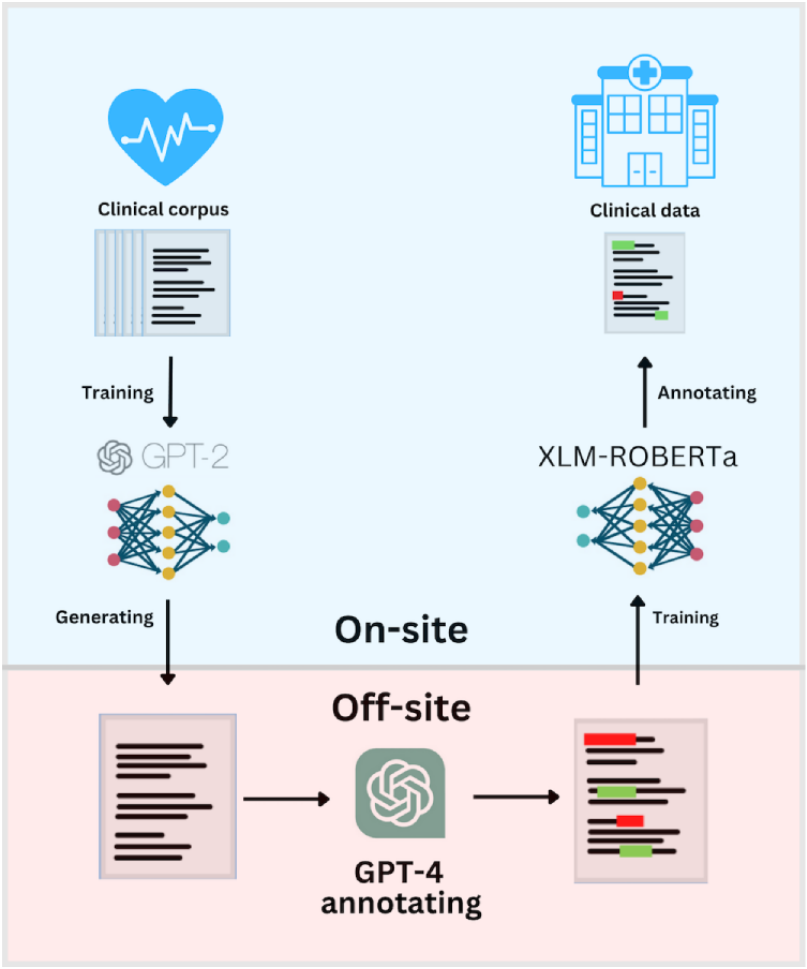

Generating Synthetic Data With GPT-2

As health care data are very sensitive, sharing them with third-party LLM hosts such as OpenAI or Azure (Microsoft Corp) is not possible. For this reason, we used synthetic data for the annotating process, generated by a GPT-2 small model. The dataset we used for training the model consists of around 10 million texts from 200,000 Estonians collected between 2010 and 2020 [18]. The texts are from a national registry of discharge reports covering both primary and secondary care contacts. The documents include anamnesis, summary, test results, objective findings, and procedures. During training, each medical document was prefaced with the document type, patient's age group, patient's gender, and diagnosis code, a similar idea to that used by Keskar et al [19], to generate synthetic health care records based on our specifications.

As the model can generate exact copies of the text it was trained on, it is important to ensure that no generated texts are identical to the training data. One way to verify this is by comparing the generated text directly against all training data. However, this method is unreliable because it fails to detect copies when even a single token differs. Some works [20,21] check if n-grams with a length of 13 in the generated texts are present in the training data, and if there is a match, then the generated texts are considered unoriginal. The problem with such an approach is that it can consider the training text with a length of 13 originals if we switch 1 word in it with a synonym. To address this, we considered using the longest common subsequence (LCS). For each generated text, we calculated the LCS with each training text, selected the longest, and divided by the generated text length, producing a score between 0 (unique) and 1 (identical).



The advantage of LCS is its ability to classify a generated text as a copy of the training data, even if some words have been replaced, deleted, or added. While calculating the LCS for every training instance is computationally expensive, with a training dataset of 10 million texts, we felt it was feasible and critical to ensure that no training text (real medical texts) copies were present in subsequent steps. However, with longer texts, the LCS elements were scattered across hundreds of tokens in the training text not representing continuous expressions. To mitigate this issue, we explored using a penalized LCS, which factors in the distances between matched tokens within the sequence. Given a penalty length k, we subtracted a penalty term *distance/k* for each consecutive LCS element, where the distance between LCS elements is the maximum of distances in x and y_i. If the distance is larger than k, we would subtract 1. If k is 1, then this is the same as finding the longest n-gram. For larger k values, the penalty would be smaller and we would match more texts. We experimented with various penalties and

XSL•FO
RenderX

found 20 to be effective in removing scattered matches without being overly restrictive.

Using LLMs for Annotations

The synthetic texts generated by our custom GPT-2 were passed as payloads into a prompt for Azure Cognitive Services GPT models via their API, which were instructed to annotate the data. We opted out of human review of data in Azure to further ensure the privacy of even the synthetic EHRs. For all experiments, we use the default parameters for the models. Our goal was to find the corresponding text spans for each annotation class; however, LLMs have been shown to make errors when prompted to return indexes for text spans [22], and our experiments showed that both the GPT-3.5-Turbo and GPT-4 models gave us incorrect spans when asked. Therefore, we only asked it to give us the text and the annotation class and map it to the text, finding the spans manually. To make the results as easy to parse as possible, we asked it to return the results in a *JSON* format, with each annotation class as an element and each text as a value of the corresponding element.

Prompt engineering has been shown to largely influence the quality of the output of the models [23]. This led us to select 4 different prompts to test how well they performed so we could choose the best one to annotate the data with. The prompts are listed in Table 1.

First, we used a zero-shot prompt because it is the most simplistic way to ask a model to complete a task. Our zero-shot prompt includes a description of the task, the wanted output format, and the input text.

The second prompt consists of instructions in Estonian to test if the accuracy of the annotations improves if the language of the prompt matches the language of the data. It is the zero-shot prompt translated into Estonian.

As human annotators have annotation guidelines as their instructions, we decided to explore whether adding definitions of entities in the prompts would increase the accuracy of the annotations. By providing definitions, we can specify what counts as an entity in our domain rather than relying on the model's understanding of the terms. The definitions for the classes were written by the same medical expert who annotated the synthetic data against which we tested the accuracy of the GPT model's annotations. Our prompt with definitions was inspired by Li et al [24] because they had promising results using prompts with additional annotation guidelines for NER with a BERT model.

As GPT models have been shown to perform better when provided with examples of task completion [20], we also tried a few-shot prompt. It consists of 5 pairs of example inputs and the expected *JSON* outputs. Few-shot prompts are particularly effective in low-resource settings, as they provide semantically relevant context that helps bridge the language gap between the target language and the high-resource language the model is proficient in [23].

The resulting responses were parsed in Python (Python Software Foundation) to convert them into *.tsv* format for training the downstream model. It was carried out by searching for the annotation entities from the original text, mapping them to spans, and finally aligning the spans into tokens.

XSL•FO
RenderX

Table 1. Contents of prompts.

Prompt name	Prompt
Zero-shot	In the text below, give the list of: drug named entity, procedure named entity, disease named entity, smoking named entity. Words need to be in exactly the same format as in input text. Format the output in JSON with the following keys: DRUG for drug named entity, PROCEDURE for procedure named entity, DISEASE for disease named entity, SMOKING for smoking named entity. Text below:
Estonian	Loetle järgnevas tekstis järgmiseid nimeolemeid: ravimite nimeolemid, protseduuride nimeolemid, haiguste nimeolemid, suitsetamiste nimeolemid. Sõnad päevad olema täpselt samal kujul kui sisendtekstis. Vorminda väljund JSON kujul järgnevate võtmetega: RAVIM ravimite nimeolemite jaoks, PROTSEDUUR protseduuride nimeolemite jaoks, HAIGUS haiguste nimeolemite jaoks, SUITS suitsetamiste nimeolemite jaoks. Tekst on järgmine:
Definitions	In the text below, give the list of drug named entity, procedure named entity, disease named entity, smoking named entity. Words need to be in exactly the same format as in input text. The annotation guidelines are the following: DRUG - any names or active ingredients of drugs, including abbreviations; no quantities PROCEDURE - any mentions of procedures, analyses etc, including also cites (e.g. “ajupiirkonna mrt”); no single blood analyses, only overall mentions of blood analyses DISEASE - names of diseases, ICD-10 codes SMOKING - any mentions of smoking, smoking status, both negative and positive Format the output in JSON with the following keys: DRUG for drug named entity, PROCEDURE for procedure named entity, DISEASE for disease named entity, SMOKING for smoking named entity. Text below:
Few-shot	In the text below, give the list of: drug named entity, procedure named entity, disease named entity, smoking named entity. Words need to be in exactly the same format as in input text. Format the output in JSON with the following keys: DRUG for drug named entity, PROCEDURE for procedure named entity, DISEASE for disease named entity, SMOKING for smoking named entity. Here are a few examples: <Example input 1> <Example output 1> <Example input 5> <Example output 5> Text below:

XSL•FO
RenderX

Evaluating the Performance of the GPT Models' Annotations

To evaluate which prompt to use for annotating the data, we generated 500 synthetic texts with 5 of the most frequent *International Classification of Diseases, Tenth Revision* (*ICD-10*) codes in our dataset. For each code, we generated 67 patient summaries and 33 procedures ranging from all age groups and both genders. This dataset was then annotated by a medical expert for drugs, procedures, diseases, and smoking entities to compare the LLM annotations with human annotations. The numbers of entities for each class are depicted in Table 2.

We compared the annotations made by the LLM word level by taking the human annotations as the ground truth and measuring precision, recall, and F_1-score (harmonic mean of precision and recall) of the annotations. We do this using the *seqeval* [25] library to obtain the classification reports with the default options from the annotations.

After determining the most effective prompt for all the annotation classes, we used the prompt with 3 different models—GPT-3.5-Turbo; GPT-4; and Llammas [26], a Llama-2-7B–based [27], instruction-tuned model adapted to Estonian.

This gave us an estimation of how well the different models annotate the data and which prompt to use for annotating the data that are used for training the downstream model.

Table 2. Entity counts.

Annotation class	Entity count, n
Drug	219
Procedure	643
Smoking	58
Disease	322

Training Downstream XLM-RoBERTa Model

Once the responses from the GPT models were parsed into training data format, we used this as data for fine-tuning a BERT-based model. We chose the XLM-RoBERTa base with 125M parameters as its pretraining data includes Estonian and other languages similar to Estonian. Additionally, because it is multilingual, we could later add annotations from other languages to further improve its performance. For each model we trained, we used 15% of the training data for the validation set.

The parameters for fine-tuning the model were the following:

- batch_size=16
- evaluation_strategy="epoch"
- learning_rate="2e-5"
- per_device_train_batch_size=16
- per_device_eval_batch_size=16
- num_train_epochs=8
- weight_decay=0.01

All the trained models were evaluated at the token level with *seqeval* default settings using 300 documents from real-world data as the test set. We chose 150 texts from patient medical histories and 150 texts from patient procedure texts to make sure we had enough test data for all the annotation classes. We had 2 medical experts annotate these texts for drugs and procedures. The instructions given to the annotators were only the names of the entity classes to ensure that the instructions were similar to the ones given to the models annotating the data. We used one set of annotations as ground truth for our experiments and the second set to calculate the pairwise interannotator agreement [28], which was an F_1-score of 0.73 for drugs and 0.48 for procedures. Once the models were trained on the synthetic data annotated by the GPT models, we could use the models on the test data in a secure local environment and retrieve the results.

Ethical Considerations

The medical texts were accessed according to the Estonian Committee on Bioethics and Human Research approval (1.1-12/3797) that waived the requirement of informed consent for the use of the data. The methodology presented in this study prioritizes patient privacy by exclusively using synthetic health care data for model training instead of augmenting original data. This means there is no direct link between the patients and the generated texts. We furthermore ensured privacy by removing verbatim and close-to-verbatim copies of synthetic texts. All sensitive medical records remain protected, and synthetic data were generated using local models. This ensures that no actual patient data are shared with external platforms, in compliance with privacy laws and ethical standards governing medical data use.

Results

Overview

To test the performance of the pipeline as a whole, we chose 4 annotation classes that we wanted to extract from the texts—drugs, procedures, smoking, and disease. Extracting drugs and smoking is a simple task with relatively fixed vocabularies that we hypothesized would be easy to learn for the model. Procedures and diseases are slightly trickier, as they are often written in shorthand and their representation may vary, making context very important.

Generating the Data

Based on the annotation classes for the experiment, we chose procedures and patient medical history for the types of EHRs we wished to generate. We also added specifications for 13 age

XSL•FO
RenderX

groups, ranging from 15 to 80 years, to generate data for more diverse patients. We chose C50 as the *ICD-10* diagnosis code, generating the texts for patients with malignant neoplasms of the breast. We started with 10,000 generated texts and used our LCS algorithm to filter out texts that were too close to the original data. We ended up with around 4100 documents we could use for annotating with LLMs, out of which 2700 were of medical history and 1400 of procedures.

Prompt Selection

We used all the selected prompts for annotating the 500 synthetically generated texts with GPT-3.5-Turbo to compare their performance. For each prompt, we compared the annotations made by the GPT-3.5-Turbo model with the human annotations. The results for each prompt are depicted in Table 3. For precision and recall, 95% CIs were calculated using the Wilson score method [29]. The parsed responses column displays how many of the responses contained a valid *JSON* element containing the annotation classes.

For drugs, procedures, and smoking, the highest F_1-score was reached by using the few-shot prompt on GPT-3.5-Turbo. In the case of diseases, using the definitions prompt resulted in the highest F_1-score of 0.44, although the Estonian prompt had a higher recall of 0.52 versus 0.43. For smoking, the definitions prompt gave the highest precision of 1 versus 0.92 for the few-shot prompt; however, its recall and F_1-score were lower than those of the few-shot prompt.

We selected the prompt with the highest combined F_1-score for all the classes, which was the few-shot prompt, to continue annotating the rest of the synthetic data. We first used the same training data that GPT-3.5-Turbo annotated and redid it both with GPT-4 and a locally hosted Llammas model to compare the differences in performance. The results are depicted in Table 4.

In all instances, GPT-4 had a higher F_1-score than the GPT-3.5-Turbo and Llammas models; however, for procedure and smoking, GPT-3.5-Turbo had the highest precision.

Table 3. GPT-3.5-Turbo annotation performance compared to human expert annotations.

Annotation class and prompt	Precision (95% CI)	Recall (95% CI)	F_1-score	Parsed responses, n
Drug				
Prompt 1: zero-shot	0.61 (0.55-0.67)	0.75 (0.69-0.80)	0.67	495
Prompt 2: Estonian	0.3 (0.26-0.34)	0.63 (0.56-0.69)	0.41	487
Prompt 3: definitions	0.6 (0.54-0.65)	0.82 (0.76-0.87)	0.7	500
Prompt 4: few-shot	0.63 (0.57-0.68)	0.88 (0.83-0.92)	0.74	495
Procedure				
Prompt 1: zero-shot	0.25 (0.21-0.3)	0.15 (0.12-0.18)	0.19	495
Prompt 2: Estonian	0.24 (0.21-0.28)	0.22 (0.19-0.25)	0.23	487
Prompt 3: definitions	0.25 (0.22-0.28)	0.26 (0.23-0.3)	0.26	500
Prompt 4: few-shot	0.34 (0.3-0.38)	0.26 (0.23-0.3)	0.3	495
Smoking				
Prompt 1: zero-shot	1 (0.65-1)	0.12 (0.06-0.23)	0.22	495
Prompt 2: Estonian	0.21 (0.14-0.31)	0.29 (0.19-0.42)	0.24	487
Prompt 3: definitions	1 (0.85-1)	0.36 (0.25-0.49)	0.53	500
Prompt 4: few-shot	0.92 (0.75-0.98)	0.41 (0.29-0.54)	0.57	495
Disease				
Prompt 1: zero-shot	0.24 (0.21-0.27)	0.5 (0.45-0.55)	0.32	495
Prompt 2: Estonian	0.12 (0.1-0.14)	0.52 (0.47-0.57)	0.2	487
Prompt 3: definitions	0.44 (0.39-0.5)	0.43 (0.38-0.48)	0.44	500
Prompt 4: few-shot	0.4 (0.35-0.45)	0.5 (0.45-0.55)	0.44	495

XSL•FO
RenderX

Table 4. Few-shot annotation performance of different models compared to human annotations.

Annotation class and model	Precision (95% CI)	Recall (95% CI)	F_1-score
Drug			
GPT-3.5-Turbo	0.63 (0.57-0.68)	0.88 (0.83-0.92)	0.74
GPT-4	0.66 (0.6-0.71)	0.88 (0.83-0.92)	0.75
Llammas	0.63 (0.57-0.69)	0.67 (0.61-0.73)	0.65
Procedure			
GPT-3.5-Turbo	0.34 (0.3-0.38)	0.26 (0.23-0.3)	0.3
GPT-4	0.33 (0.3-0.36)	0.42 (0.38-0.46)	0.37
Llammas	0.28 (0.24-0.32)	0.22 (0.19-0.25)	0.25
Smoking			
GPT-3.5-Turbo	0.92 (0.75-0.98)	0.41 (0.29-0.54)	0.57
GPT-4	0.69 (0.55-0.8)	0.55 (0.42-0.67)	0.61
Llammas	0.75 (0.47-0.91)	0.16 (0.09-0.27)	0.26
Disease			
GPT-3.5-Turbo	0.4 (0.35-0.45)	0.5 (0.45-0.55)	0.44
GPT-4	0.4 (0.36-0.44)	0.7 (0.65-0.75)	0.51
Llammas	0.21 (0.18-0.25)	0.35 (0.3-0.4)	0.26

Training the Downstream Model

We used the 2700 medical history texts and 1400 procedure texts resulting from LCS filtering to create batches of training data to see how well the model performs with additional training data. For each batch, we selected two-thirds of the texts from medical history and one-third from procedures, to fully use all the training data. Each model was then validated on 300 real-world documents annotated by a medical expert for drugs and procedures.

Figure 2 shows the performance of the model on the drug and procedure annotation classes. For every amount of data depicted on the x-axis, a downstream model was trained and evaluated on the test data.

For drugs, the performance of the model increased until 750 texts, lowered slightly for 1000 texts, and reached the peak at 2000 texts, achieving a precision of 0.75, a recall of 0.64, and an F_1-score of 0.69. In the case of procedures, the model peaked at 1000 synthetic annotated texts, achieving a precision of 0.47, a recall of 0.32, and an F_1-score of 0.38. For 2000 texts, the F_1-score remained the same; however, the precision increased by 0.04 whereas the recall dropped by 0.01.

We have used the annotations made by the second medical expert as validation data to show how the metrics vary with annotation variability. The results showed similar dynamics to the scores shown in Figure 2, although the absolute F_1-scores were slightly lower for procedures (Multimedia Appendix 1).

Figure 2. Annotation scores on validation data (N=300) for models of different amounts of training data.

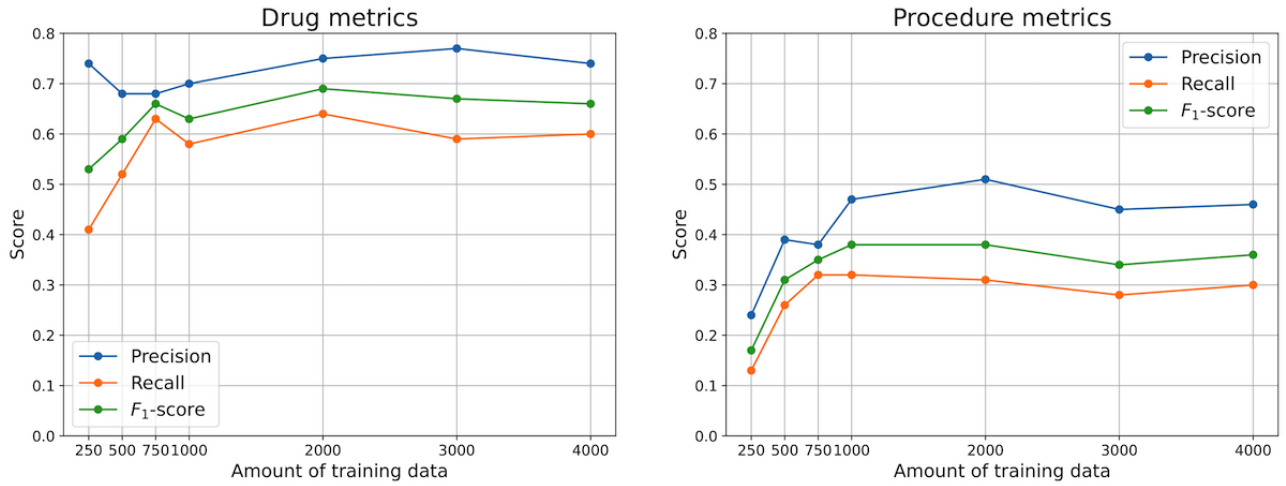

XSL•FO
RenderX

Discussion

Principal Findings

For the final model trained on different batches of data with annotations made by GPT-4, we found that the best-performing model achieved an F_1-score of 0.69 for the drug and an F_1-score of 0.38 for procedures. The result was achieved with 2000 annotated synthetic texts for drugs and 1000 for procedures. Training the models on more texts did not improve the results. Based on these results, it seems there is a cap on how much one can gain from synthetic data using this method. The downstream model seems to learn the task after 2000 texts, and afterward, adding more data just creates more noise and possibly makes the model overfit on similar synthetic texts.

It is important to note that the interannotator agreement showed an F_1-score of 0.73 for drugs and 0.48 for procedures, which means there is inherently a lot of room for disagreement and subjectivity between the entities. In many instances, the disagreement stemmed from how many prefixes to include for the entity (eg, location of the procedure). This result tells us that achieving a perfect F_1-score is not realistic, as ground truth itself can be subjective and depend on the annotators, further strengthening the achieved result of this paper. The disagreement between the medical experts and the LLM can be somewhat mitigated by providing more clear annotation guidelines for both the medical experts and the LLMs annotating the data.

We found that the most effective prompt for generating accurate annotations for our downstream model was the few-shot approach. Notably, while the texts themselves were in Estonian, prompts provided in English consistently produced more precise annotations compared to prompts written in Estonian. Openly available, multilingual LLMs are known to be heavily trained on English text, as English is vastly overrepresented in publicly available datasets. As a result, the model's understanding of English instructions and its ability to follow them are significantly more robust than for low-resource languages such as Estonian [30,31]. The local Llammas model did not perform as well as the GPT-3.5-Turbo, likely because it is a much smaller model, and parsed only 454 of the responses, meaning in some instances it did not return coherent outputs.

We then found that when comparing the performance of the annotations between GPT-3.5-Turbo, GPT-4, and Llammas, GPT-4 performed better per F_1-score for all cases, which was to be expected, because it is a larger and more powerful model. For procedures and smoking, the GPT-3.5-Turbo model achieved a higher precision; however, in the case of procedure, it was only by 0.01, and in the case of smoking, even though the difference was 0.23, there were only 58 entities in the texts, meaning the difference in results is not very reliable.

Comparison to Prior Work

While our model's results do not surpass standard transformer-based models trained on human-annotated data, which achieve F_1-scores of 0.9 for drugs [8,9] and 0.7 for procedures [8], it is crucial to contextualize this comparison. These prior results are derived from English, a high-resource language with abundant training data. In contrast, our approach demonstrates the efficacy of using synthetically generated data for low-resource languages such as Estonian, without relying on human annotations. Considering that our approach does not use any human-annotated data and the models are all trained on synthetically generated data, a good starting point is to use synthetic data for an initial model that can be used to enrich a model trained on real data or help annotators with their tasks on real-world data.

Future Directions

The pipeline of the method described in this paper consists of many pieces, each with room for improvement. First, both the generation and annotation of synthetic data could be enhanced by leveraging more advanced models. The synthetic texts were generated by GPT-2, but newer model architectures with more parameters could outperform this and generate better data, while more specialized models for clinical texts, such as Me-LLaMA [32] and GatorTron [33], could improve annotation accuracy. Second, we saw a noticeable change in performance depending on the prompt used, suggesting that further prompt engineering could lead to more improvements. Third, human annotators in health care settings typically rely on lengthy annotation guidelines, while in contrast, our method provided only 5 examples. A more detailed description of annotation classes could improve the accuracy of model-generated annotations.

Conclusion

In this paper, we took advantage of powerful LLMs GPT-3.5-Turbo and GPT-4 without risking the privacy of the data by using synthetically generated data. We used these models to annotate the synthetic texts, on which we locally trained a BERT-based NER model for use on real-world data. We obtained an F_1-score of 0.69 for drugs and an F_1-score of 0.38 for procedures, which are impressive, given that the data used were synthetic and without human annotations. We then showcased the difference in performance using GPT-3.5-Turbo and GPT-4 for the task, finding that GPT-4 is substantially better for the task. We also showed how the performance of the model is influenced by the amount of synthetic training data used for training it. Altogether, the proposed method allows for developing NER models while removing the bottleneck of using medical experts for data annotation, automating the process of extracting facts from unstructured data while preserving patient privacy.

Acknowledgments

This research was carried out with the support of the Estonian Research Council grants PRG1844 and PRG1604 and the Optimal Treatment for Patients With Solid Tumours in Europe Through Artificial Intelligence (OPTIMA), which is funded through the IMI2 (Innovative Medicines Initiative 2) Joint Undertaking and is listed under grant agreement 101034347. IMI2 receives support

XSL•FO
RenderX

from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA). This work was also supported by the Estonian Centre of Excellence in AI, funded by the Estonian Ministry of Education and Research, by the European Union and Ministry of Education and Research via project TEM-TA72 and by Nordic Innovation (grant 407-7003-P22026).

Data Availability

The medical texts from the Estonian Biobank were used according to the Estonian Committee on Bioethics and Human Research approval (1.1-12/3797). These data cannot be shared publicly because they may contain identifiable information about the participants. To access these data, official procedures outlined in reference [34] have to be followed. The synthetic data used for annotations can be accessed on a case-by-case basis upon contacting author RK. The code for the pipeline is publicly available in GitHub [35].

Authors' Contributions

HŠ conceived and designed this paper, focusing on the primary idea, concept, experiments, and analysis. ML designed and implemented GPT-2 synthetic data generation and longest common subsequence filtering pipelines. VK did the research on prompt selection, MM and NI provided medical annotations, HAK managed the local Llammas model, and RK supervised the project. All authors—HŠ, ML, VK, MM, NI, HAK, and RK—reviewed and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Annotation scores on validation data (N=300) from the second medical expert for models of different amounts of training data.
[DOCX File , 68 KB - jmir_v27i1e66279_app1.docx]

References

1. Li I, Pan J, Goldwasser J, Verma N, Wong W, Nuzumlalı MY, et al. Neural natural language processing for unstructured data in electronic health records: a review. Comput Sci Rev 2022 Nov;46:100511 [FREE Full text] [doi: 10.1016/j.cosrev.2022.100511]
2. Torge S, Politov A, Lehmann C, Saffar B, Tao Z. Named entity recognition for low-resource languages - profiting from language families. : Association for Computational Linguistics; 2023 Presented at: Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023); May 6, 2023; Dubrovnik, Croatia p. 1-10 URL: https://aclanthology.org/2023.bsnlp-1.1 [doi: 10.18653/v1/2023.bsnlp-1.1]
3. Cotterell R, Duh K. Low-resource named entity recognition with cross-lingual, character-level neural conditional random fields. : ACM; 2017 Presented at: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers); November 27 - December 1, 2017; Taipei, Taiwan p. 91-96 URL: https://aclanthology.org/I17-2016 [doi: 10.1109/snlp.2009.5340913]
4. Simpson E, Brown R, Sillence E, Coventry L, Lloyd K, Gibbs J, et al. Understanding the barriers and facilitators to sharing patient-generated health data using digital technology for people living with long-term health conditions: a narrative review. Front Public Health 2021;9:641424 [FREE Full text] [doi: 10.3389/fpubh.2021.641424] [Medline: 34888271]
5. Szarvas G, Farkas R, Busa-Fekete R. State-of-the-art anonymization of medical records using an iterative machine learning framework. J Am Med Inform Assoc 2007;14(5):574-580 [FREE Full text] [doi: 10.1197/j.jamia.M2441] [Medline: 17823086]
6. Olatunji IE, Rauch J, Katzensteiner M, Khosla M. A review of anonymization for healthcare data. Big Data 2024 Dec;12(6):538-555. [doi: 10.1089/big.2021.0169] [Medline: 35271377]
7. GPT-4 technical report. OpenAI. 2023. URL: https://cdn.openai.com/papers/gpt-4.pdf [accessed 2025-02-21]
8. Tian S, Erdengasileng A, Yang X, Guo Y, Wu Y, Zhang J, et al. Transformer-based named entity recognition for parsing clinical trial eligibility criteria. 2021 Aug Presented at: Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics; August 1-4, 2021; Gainesville, FL p. 1-6 URL: https://europepmc.org/abstract/MED/34414397 [doi: 10.1145/3459930.3469560]
9. Alrdahi H, Han L, Šuvalov H, Nenadic G. MedMine: examining pre-trained language models on medication mining. arXiv Preprint posted online on Aug 8, 2023. [doi: 10.48550/arXiv.2308.03629]
10. Du X, Jia Y, Zan H. MRC-based medical NER with multi-task learning and multi-strategies. 2022 Presented at: China National Conference on Chinese Computational Linguistics; October 14, 2022; Nanchang, China p. 149-162 URL: https://aclanthology.org/2022.ccl-1.74 [doi: 10.1007/978-3-031-18315-7_10]
11. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: A transferable clinical natural language processing model for electronic health records. Artif Intell Med 2021 Aug;118:102086. [doi: 10.1016/j.artmed.2021.102086] [Medline: 34412834]

XSL•FO
RenderX

12. Sasikumar N, Mantri K. Transfer learning for low-resource clinical named entity recognition. 2023 Presented at: Proceedings of the 5th Clinical Natural Language Processing Workshop; July 14, 2023; Toronto, ON p. 514-548 URL: https://aclanthology.org/2023.clinicalnlp-1.53 [doi: 10.18653/v1/2023.clinicalnlp-1.53]
13. Ding B, Qin C, Liu L, Chia Y, Joty S, Bing L. Is GPT-3 a good data annotator? 2023 Presented at: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); July 9-14, 2023; Toronto, ON p. 11173-11195. [doi: 10.18653/v1/2023.acl-long.626]
14. Tang R, Han X, Jiang X, Hu X. Does synthetic data generation of LLMs help clinical text mining? arXiv Preprint posted online on April 10, 2023. [doi: 10.48550/arXiv.2303.04360]
15. Introducing ChatGPT. OpenAI. 2022. URL: https://openai.com/blog/chatgpt [accessed 2025-02-21]
16. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. 2019. URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf [accessed 2025-02-21]
17. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, et al. Unsupervised cross-lingual representation learning at scale. 2020 Presented at: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; July 5-10, 2020; Online p. 8440-8451. [doi: 10.18653/v1/2020.acl-main.747]
18. Lepson M. Epikriisi Tekstide Genereerimine GPT-2 Mudeliga. Epicrisis text generation with GPT-2 [Master's thesis]. University of Tartu. 2023. URL: https://dspace.ut.ee/server/api/core/bitstreams/b71d0a9d-0e82-453c-bf53-3a9f7df2a904/content [accessed 2025-02-22]
19. Keskar NS, McCann B, Varshney LR, Xiong C, Socher R. CTRL - a conditional transformer language model for controllable generation. arXiv Preprint posted online on September 20, 2019. [doi: 10.48550/arXiv.1909.05858]
20. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. 2020 Presented at: NIPS'20: 34th International Conference on Neural Information Processing Systems; December 6-12, 2020; Vancouver, BC p. 1877-1901 URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf/> [doi: 10.48550/arXiv.2005.14165]
21. Carlini N, Liu C, Erlingsson U, Kos J, Song D. The secret sharer: evaluating and testing unintended memorization in neural networks. 2019 Presented at: SEC'19: Proceedings of the 28th USENIX Conference on Security Symposium; August 14–16, 2019; Santa Clara, CA p. 267-284 URL: https://www.usenix.org/system/files/sec19-carlini.pdf [doi: 10.5555/3361338.3361358]
22. Han R, Peng T, Yang C, Wang B, Liu L, Wan X. An empirical study on information extraction using large language models. arXiv Preprint posted online on May 23, 2023 [FREE Full text] [doi: 10.48550/arXiv.2305.14450]
23. Grabb D. The impact of prompt engineering in large language model performance: a psychiatric example. J Med Artif Intell 2023;6:20 [FREE Full text] [doi: 10.21037/jmai-23-71]
24. Li X, Feng J, Meng Y, Han Q, Wu F, Li J. A unified MRC framework for named entity recognition. 2022 Presented at: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; July 5-10, 2020; Online p. 5849-5859. [doi: 10.18653/v1/2020.acl-main.519]
25. Nakayama H. seqeval: a python framework for sequence labeling evaluation. Software. 2018. URL: https://github.com/chakki-works/seqeval [accessed 2025-02-21]
26. Kuulmets HA, Purason T, Luhtaru A, Fishel M. Teaching Llama a new language through cross-lingual knowledge transfer. : Association for Computational Linguistics; 2024 Presented at: Findings of the Association for Computational Linguistics: NAACL 2024; June 16-21, 2024; Mexico City, Mexico p. 3309-3325 URL: https://aclanthology.org/2024.findings-naacl.210 [doi: 10.18653/v1/2024.findings-naacl.210]
27. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: open foundation and fine-tuned chat models. arXiv Preprint posted online on July 19, 2023 [FREE Full text]
28. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. J Am Med Inform Assoc 2005;12(3):296-298 [FREE Full text] [doi: 10.1197/jamia.M1733] [Medline: 15684123]
29. Wilson E. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927;22(158):209-212 [FREE Full text] [doi: 10.2307/2276774]
30. Nguyen XP, Aljunied SM, Joty S, Bing L. Democratizing LLMs for low-resource languages by leveraging their English dominant abilities with linguistically-diverse prompts. : Association for Computational Linguistics; 2024 Presented at: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); August 2024; Bangkok, Thailand p. 3501-3516 URL: https://arxiv.org/abs/2306.11372 [doi: 10.18653/v1/2024.acl-long.192]
31. Bang Y, Cahyawijaya S, Lee N, Dai W, Su D, Wilie B. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. 2023 Presented at: Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics; November 1-4, 2023; Nusa Dua, Bali p. 675-718 URL: https://doi.org/10.18653/v1/2023.ijcnlp-main.45 [doi: 10.18653/v1/2023.ijcnlp-main.45]
32. Xie Q, Chen Q, Chen A, Peng C, Hu Y, Lin F, et al. Me-LLaMA: foundation large language models for medical applications. Res Sq 2024 May 22 [FREE Full text] [doi: 10.21203/rs.3.rs-4240043/v1] [Medline: 38826372]

XSL•FO
RenderX

33. Yang X, Chen A, PourNejatian N, Shin H, Smith K, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med 2022;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
34. Estonian Biobank. Institute of Genomics. URL: https://genomics.ut.ee/en/content/estonian-biobank [accessed 2025-02-22]
35. Generating synthetic data with LLM-s for training a downstream model. GitHub. URL: https://github.com/HealthInformaticsUT/annotating_synthetic_data_gpt [accessed 2025-02-22]

Abbreviations

API: application programming interface
EHR: electronic health record
ICD-10: International Classification of Diseases, Tenth Revision
LCS: longest common subsequence
LLM: large language model
NER: named entity recognition

Edited by T de Azevedo Cardoso; submitted 09.09.24; peer-reviewed by Z Ehtesham, T Akindahunsi, A Adewuyi, E Bai; comments to author 29.11.24; revised version received 19.12.24; accepted 31.01.25; published 18.03.25.

Please cite as:
Šuvalov H, Lepson M, Kukk V, Malk M, Ilves N, Kuulmets HA, Kolde R
Using Synthetic Health Care Data to Leverage Large Language Models for Named Entity Recognition: Development and Validation Study
J Med Internet Res 2025;27:e66279
URL: https://www.jmir.org/2025/1/e66279
doi:10.2196/66279
PMID:

©Hendrik Šuvalov, Mihkel Lepson, Veronika Kukk, Maria Malk, Neeme Ilves, Hele-Andra Kuulmets, Raivo Kolde. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Association Between Intergenerational Support, Technology Perception and Trust, and Intention to Seek Medical Care on the Internet Among Chinese Older Adults: Cross-Sectional Questionnaire Study

Hengjiang Jin[1], PhD; Ying Qu[1], MA

School of Journalism, Chongqing University, Chongqing, China

Corresponding Author:
Hengjiang Jin, PhD
School of Journalism
Chongqing University
Number 174, Shazheng Street, Shapingba District
Chongqing
China
Phone: 86 15523596050
Email: jinhj@cqu.edu.cn

Abstract

Background: Avoiding technological innovation does not simplify life. In fact, using internet-based medical services can enhance the quality of life for older adults. In the context of an aging population and the growing integration of information technology, the demand for internet-based medical services among older adults is gaining increased attention. While scholars have highlighted the important role of intergenerational support in promoting digital inclusion for older adults, research on the relationship between intergenerational support and older adults' intentions to seek online care remains limited.

Objective: This study aims (1) to explore the association between intergenerational support, online medical information, and older adults' intention to seek medical care online, and (2) to examine the mediating role of technology perception and trust, as well as the moderating role of eHealth literacy.

Methods: A cross-sectional survey was conducted in China, collecting 958 valid responses from older adults aged 60 years and above. A vast majority of participants were between the ages of 60 and 75 years (771/958, 80.5%). Of the 958 participants, 559 (58.4%) resided in rural areas, while 399 (41.6%) lived in urban areas. The survey included questions on intergenerational support, perceived usefulness, perceived ease of use, trust, online medical information, eHealth literacy, and the intention to seek medical care online. Structural equation modeling and linear regression analysis were applied to explore the relationship between intergenerational support and the intention to seek medical care on the internet.

Results: Intergenerational support was positively associated with perceived ease of use (β=.292, *P*<.001), perceived usefulness (β=.437, *P*<.001), trust (β=.322, *P*<.001), and the intention to seek medical care online (β=.354, *P*<.001). Online medical information also positively affected the intention to seek medical care online among older adults (β=.109, *P*<.001). Perceived ease of use (β=.029, 95% CI 0.009-0.054), perceived usefulness (β=.089, 95% CI 0.050-0.130), and trust (β=.063, 95% CI 0.036-0.099) partially mediated the association between intergenerational support and the intention to seek medical care online. Further analysis found that perceived ease of use, perceived usefulness, and trust played a chain mediating role between intergenerational support and the intention to seek medical care online (β=.015, 95% CI 0.008-0.027; β=.022, 95% CI 0.012-0.036). Additionally, eHealth literacy played a moderating role in the relationship between intergenerational support and perceived ease of use (β=.177, *P*<.001), perceived usefulness (β=.073, *P*<.05), trust (β=.090, *P*<.01), and the intention to seek medical care online (β=.124, *P*<.001).

Conclusions: An integrated model of health communication effects was constructed and validated, providing empirical support for the intention to seek medical care online and for the impact of health communication. This model also helps promote the role of technology in empowering the lives of seniors.

(J Med Internet Res 2025;27:e65065) doi:10.2196/65065

XSL•FO
RenderX

KEYWORDS

intergenerational support; older adults; internet medical intentions; perceived technology; trust

Introduction

Background

The internet has become an important medium for providing various eHealth services and a source of medical information [1]. Smartphones bring the convenience of telemedicine to users' fingertips [2]. According to the 52nd China Internet Network Information Center report [3], China has 364 million internet medical care users, accounting for 33.8% of all internet users. The number of online hospitals continues to grow, and the development of large-scale online medical care platforms shows strong momentum.

The World Health Organization has encouraged member states to actively use digital technologies to improve medical, health, and social services, thereby promoting active aging [4]. Given that older adults are the primary users of health care services [5], there is significant potential for internet-based medical services to enhance geriatric health. Currently, China's population is aging rapidly [6]. As of the end of 2023, the population of older adults in China (aged ≥60 years) has reached 297 million, accounting for 21.1% of the total population. For older adults, proficient use of technology is key to accessing internet-based medical services [7]. However, digital disparities—such as difficulties in acquiring, operating, and using basic equipment, as well as insufficient knowledge—limit their ability to equally benefit from technological advancements. Given the trend of population aging, the challenge of successfully integrating older adults into the digital health ecosystem has become a pressing issue.

Some prior studies have explored this question, noting that older adults are particularly concerned about the quality and risks of internet-based medical services [8]. Additionally, older adults' habits play an important role in explaining their use of such services [9]. However, further discussion is needed to clarify the complexities and mechanisms that influence older adults' intention to seek medical care online.

Seeking Medical Care Online

Intention to adopt internet-based medical services refers to individuals' subjective beliefs and willingness to seek care through such services when facing health problems [10]. It reflects the importance individuals place on their health, as well as their trust in and reliance on internet-based medical services. This intention is influenced by multiple factors, including disease severity, individual perceptions and attitudes toward the illness, and the quality and availability of health care services.

Examinations of the intention to adopt internet-based medical services currently follow 3 approaches. The first approach is based on the Technology Acceptance Model or the Unified Theory of Acceptance and Use of Technology [11], which assesses user acceptance of online health platforms and the willingness of different population groups to use internet-based medical services [12]. The second approach constructs an analytical model centered on trust within the doctor-patient relationship, focusing on factors such as personal trust preferences and website trustworthiness [13]. The third approach is based on Social Exchange Theory, which examines the physicians' perspectives on the potential impact of interactive tools on their careers [14], as well as how regulatory systems, reputation systems, and communication exchanges shape their motivation to use these services [15].

Examining older adults' access to online health care in greater detail, Mansson et al [16] argued that the use of mobile health (mHealth) apps may reduce treatment costs. However, Askari et al [17] identified several factors influencing older adults' willingness to use mHealth apps, including perceived usefulness (PU), ease of use, social connections, social norms, and anxiety.

Intergenerational Support

Intergenerational Support Theory asserts that the flow of resources among family members is bidirectional [18], encompassing economic, technological, and emotional exchanges [19]. In the context of internet-based medical services, intergenerational support focuses on the technical assistance provided by the children to their parents in using these services.

Perceived ease of use (PEOU) measures older adults' ratings of technology's friendliness versus their beliefs about ease of mastery [20]. However, many older adults are unfamiliar with smartphone operation, are often excluded from various internet services, and are unable to benefit equally from digital technologies [21]. Additionally, they perceive apps as complex [22], resulting in low overall levels of PEOU. The lack of digital literacy and the challenges of aging are significant barriers to older adults' intention to use mHealth technology [23]. In the context of online-based medical services, children can assist older adults by guiding them through the process of using online platforms—such as downloading, registering, finding, and consulting internet-based medical services. They can also help resolve technical issues, thereby reducing frustration [24]. This support will make it easier for older adults to use these services, leading to an increase in their assessment of PEOU. Based on this, the following hypotheses are proposed.

- Hypothesis 1a: Intergenerational support is positively associated with perceived ease of use of internet-based medical services among older adults.

PU reflects older people's views that a particular technology can improve their health care experience. It encompasses their belief that internet-based medical services can meet their health management needs, such as online appointment booking, teleconsultation, and e-prescription flow. Studies have shown that older adults are more likely to adopt technology when they gain insight into its usefulness and potential benefits. Through direct operation or demonstration by children, older adults can independently or semi-independently use the internet-based medical service platform to experience its advantages in simplifying clinical services, saving time and energy for medical

XSL•FO
RenderX

treatment, and enhancing their knowledge and acceptance of the platform. Once older adults experience the tangible benefits of internet-based medical services, it may stimulate positive feedback mechanisms that gradually ease the psychological barriers to adopting new technologies. This, in turn, can enhance psychological adaptation and cognitive health [25], thereby deepening the PU of the internet-based medical service platform.

- Hypothesis 1b: Intergenerational support is positively associated with the perceived usefulness of internet-based medical services for older adults.

Trust typically builds up gradually through continuous and intimate interactions, which can enhance patients' positive expectations [26]. However, many older adults have conservative and negative attitudes toward new technologies and are hesitant to use online services. As a result, it can be challenging for services such as internet-based medical services to gain the trust of older adults [27]. According to Ma et al [28], trust serves as a critical criterion for older adults in identifying individuals who can help them enhance their digital skills. Consequently, the younger generation plays a significant role in building trust and security for older adults. This support also involves a sense of being recognized and respected [29]. We hypothesize that greater intergenerational support is more likely to reduce older adults' unfamiliarity and thus increase their trust in these services.

- Hypothesis 1c: Intergenerational support is positively associated with older adults' trust in internet-based medical service.

According to Li and Kostka [30], social support plays a significant role in affecting older adults' digital engagement. Many older adults who maintain close contact with their families tend to receive more effective support for digital learning and engagement [30]. Additionally, given the importance of filial piety in Chinese culture, adult children are expected to assist disadvantaged family members [31]. This includes providing help and counseling regarding digital access, use, and literacy, as well as facilitating bottom-up technology transfer [32] to help their parents accomplish reverse socialization [33]. As a result, support and feedback can effectively enhance the digital competence of older adults [34] and promote the adoption of new technologies.

- Hypothesis 1d: Intergenerational support is positively associated with older adults' intention to seek medical care on the internet.

Perceived Usefulness and Perceived Ease of Use

As the key elements determining users' behavioral intentions, PU indicates the extent to which using a particular system or operation can improve task performance or achieve a specific goal. PEOU reflects the user's subjective evaluation of the convenience and ease provided by using something [35].

Some studies have shown that the PU of social media affects people's trust in the platform as well as the channel [36]. Acharya et al [37] further found that the PU of recommender systems can directly impact consumers' trust. It has also been confirmed that PEOU is positively correlated with consumer trust [38]. This study suggests, therefore, that older adults may trust internet-based medical services if they perceive them as useful or easy to use.

- Hypothesis 2a: Perceived ease of use is positively associated with older adults' trust in internet-based medical service.
- Hypothesis 2b: Perceived usefulness is positively associated with older adults' trust in internet-based medical service.

Currently, technology perception has been studied in various contexts, such as the adoption of smart home technologies [39] and the willingness to use internet applications [40]. Naidoo and Leonard [41] found that the continued willingness to use e-services is entirely dependent on high levels of PU. Further research indicates that the PU of the internet is an important predictor of an individual's use of eHealth solutions and plays a significant role in patients' sustained use of online health care communities [42]. For older adults, the functions of internet-based medical services are far more complex than simple services such as browsing short videos. Older patients tend to consider using these services only if the transition to new medical treatments is easy [43]. This study therefore argues that older adults are more likely to use internet-based medical services if they perceive them to be useful or easy to use.

- Hypothesis 3a: Perceived ease of use is positively associated with older adults' intention to seek medical care on the internet.
- Hypothesis 3b: Perceived usefulness is positively associated with older adults' intention to seek medical care on the internet.

Trust

Schoorman et al [44] suggested that trust in internet-based medical services is the tendency for people to believe that these services can fulfill their health needs and to be willing to use them, reflecting the level of acceptance for such services [45]. Trust is a crucial factor for assessing users' willingness to use internet-based medical services [46], and establishing trust improves the likelihood of older adults using these services [47]. This study therefore suggests that older adults who trust internet-based medical services are more likely to be willing to access such services. We therefore propose the following hypothesis:

- Hypothesis 4: Trust is positively associated with older adults' intention to seek medical care on the internet.

Online Medical Information

Online medical information—encompassing all types of medical-related content delivered and shared through internet platforms such as search engines, medical websites, and online forums—is highly sought after. The causes and treatments of diseases are among the most popular topics for medical information searches [48]. Notably, 60% of individuals view online medical information as equally good or better than information provided by physicians [49]. Online medical information enhances the efficiency of medical knowledge dissemination and is rapidly replacing traditional approaches for seeking advice and information [50]. It helps to improve individuals' knowledge, promote disease prevention, and

XSL•FO
RenderX

facilitate access to appropriate medical services. To pursue more informed health decisions [51], older adults often seek health-related medical information online, which influences their health behaviors [52]. Consequently, older adults may be motivated to seek guidance from physicians practicing online [53]. The following research hypothesis is therefore proposed:

- Hypothesis 5: Online medical information is positively associated with older adults' intention to seek internet-based medical services.

eHealth Literacy

eHealth literacy refers to the ability needed to access, understand, and evaluate health information from digital resources and make informed health decisions [54]. It encompasses active information-seeking, 2-way interactive communication, and information utilization/sharing [55]. Although older adults are increasingly relying on the internet to access health-related services, they often struggle to meet the necessary eHealth literacy requirements [56]. Families play an important role in promoting eHealth literacy among older adults [57]. By strengthening intergenerational support, they can help older adults better learn and apply new technologies. When confronted with a complex internet health care delivery system, families can assist in quickly getting up to speed and becoming proficient in its operation. At this level, internet health care services are easy to use. At the same time, families can help older adults obtain rich health information from the platform [58] and manage their daily health needs, which can enhance their perception of the usefulness of these services. Therefore, we propose the following:

- Hypothesis 6a: eHealth literacy positively moderates the relationship between intergenerational support and perceived ease of use.
- Hypothesis 6b: eHealth literacy positively moderates the relationship between intergenerational support and perceived usefulness.

A study found that social interaction, access to technology, and digital literacy are positively correlated. With the assistance of their children, older adults can significantly improve their ability to access, understand, and apply digital health information [59]. High levels of eHealth literacy can increase older adults' confidence and ability to assess the reliability and validity of internet-based health care services, thereby enhancing their trust in such services.

- Hypothesis 6c: eHealth literacy positively moderates the relationship between intergenerational support and trust.

According to Health Empowerment Theory, eHealth literacy and social support can promote self-care behaviors among older adults [60]. The findings of Hsu et al [61] further indicate that individuals with higher levels of eHealth literacy have a greater potential to make informed health choices, improved health care competence, and ultimately higher quality of life. Therefore, we propose the following:

- Hypothesis 6d: eHealth literacy positively moderates the relationship between intergenerational support and intention to seek medical care on the internet.

Methods

Data Collection and Participants

In this study, a questionnaire (Multimedia Appendix 1) was used to validate the research model. To ensure reliability and validity, all items were borrowed from well-established domestic and international scales and adjusted according to the theme of internet medical care. The questionnaire comprises a total of 7 latent variables, involving 24 items, all measured on a 7-point Likert scale. Before the official survey, a presurvey was conducted with 50 older adults to ensure a high degree of internal consistency. Based on their feedback, questions with ambiguities or semantic problems were adjusted.

Between March 20, 2024, and March 31, 2024, we officially launched the distribution of the questionnaire using a combination of online and offline methods. This approach ensures that the data (Multimedia Appendix 2) collected are broad and representative, providing a more comprehensive understanding of the attitudes and needs of older adults regarding the use of the internet for health care services. The study was conducted among individuals aged 60 years and older who had autonomous behavioral skills and some experience with internet use, encompassing both rural and urban areas. Those who lacked basic knowledge of the internet or were unable to participate due to daily activities or cognitive limitations were excluded.

Online questionnaires offer the advantages of low cost, autonomy, and comprehensive documentation, as well as the ability to effectively expand the range of respondents. In this study, an online self-report questionnaire was designed using Questionnaire Star (Changsha Ranxing Information Technology Co., Ltd.). The questionnaire was distributed through a convenience sample with a snowball approach. An online survey link was shared across various social media platforms and WeChat (Tencent Holdings Limited) app groups, where participation was voluntary. This approach covered older adult groups from regions such as Chongqing, Hunan, Guizhou, and Zhejiang, among others. To address the unfamiliarity of some older adults with the operation of the Questionnaire Star and the limitation of their text reading comprehension ability, this study required them to complete the questionnaire in the presence and under the guidance of their middle-aged or young relatives or friends. This approach ensured the accuracy and reliability of the data by providing assistance with understanding and navigating the questions.

Although many older people already have smart devices, they may be wary of clicking on a link to participate in an online questionnaire. Therefore, face-to-face interviews are particularly necessary. In the offline questionnaire survey, we adopted a dual strategy: on the one hand, the research team personally visited communities and villages in Chongqing and Hunan, directly communicating with older adult groups through on-site visits to ask them in detail about their willingness to use internet health care services. On the other hand, with the support of governmental agencies where the researchers are domiciled, we commissioned local staff to assist in distributing some of the questionnaires. For the older adults interviewed who had the

XSL•FO
RenderX

ability to write, we asked them to fill in the questionnaire themselves; for patients who were unable to fill in the questionnaire by themselves, their primary caregiver or interviewer filled in the questionnaire on their behalf. The questionnaires were collected on-site after completion.

During the data collection phase, a total of 1200 questionnaires were collected. After eliminating invalid questionnaires (including online questionnaires that took less <90 seconds to complete, failed the honesty test, had missing values, or provided the same answer to all questions, as well as offline questionnaires that were obviously invalid, such as choosing "strongly agree" for all of them), a total of 958 valid questionnaires were ultimately obtained for an effective response rate of 79.8%.

Measurements

Intergenerational Support

The measurement items for this dimension were derived from Lang and Schütze's study [62] and consisted of 3 items. It was measured with good reliability (α=.887) [63]. This factor collectively explained 81.6% of the variance (Kaiser-Meier-Olkin [KMO]=0.748). Specific items included "My children encourage me to seek medical care on the internet, and I am willing to try it," "My children guide I would be willing to try," and "When I have problems with internet health care, my children help me solve the problem, so I would be willing to try it."

Online Medical Information

Based on a study by de Boer et al [64], 4 items were used to measure the medical information that older adults sought online. The measurement items included "I would search for medical information on the internet" and "Online medical information has taught me something about health." The results of factor analysis for the 4 items showed a KMO value of 0.843, with the factor explaining a total of 76.2% of the variance; these items also had good reliability (α=.895).

Trust

This dimension is based on the studies by Fogel and Nehmad [65] and Deng et al [66] and consists of 3 items. Specific items included "Most of the doctors on the online medical service platforms are health experts in their fields, and I have no doubt about their professionalism" and "Generally, I trust the health advice or tips from doctors on the internet." The overall performance of these items indicated good reliability (α=.868); this factor collectively explained 79.2% of the variance (KMO=0.735).

Perceived Ease of Use

The 5 items were based on the study by Deng et al [66]. Typical items include "I don't think it is difficult to use the internet for health counseling" or "In general, I think internet health care is easy to use." The results of factor analysis for the 3 items showed a KMO value of 0.885, explaining a total of 72.7% of the variance. The 5 items showed good reliability (α=.906).

Perceived Usefulness

The items for this factor were based on those in the study by Deng et al [66]. The items included "I believe that using the internet for medical care can improve the efficiency of health care" and "Overall, I think online medical care is useful for health management." The results of the analysis showed a KMO value of 0.746, explaining a total of 81.7% of the variance, and good reliability (α=.888).

Intention to Seek Medical Care on the Internet

These 3 items referred to the study by Deng et al [66]., including "When I face health problems, I think I will solve them through internet health care" and "I am willing to use internet health care services for health counseling, such as disease control." These analyses showed a KMO value of 0.751, explaining a total of 82.8% of the variance; all 3 items also had good reliability (α=.895).

eHealth Literacy

This dimension combines health and media literacy and references the study by Norman and Skinner [67]. The 3 items included "I know how to find useful health resources and messages online" or "I know how to use the health care-type information I find on the internet to help myself," with 82.6% variance explained, a KMO value of 0.751, and good reliability (α=.894).

All 24 items were rated on a 7-point Likert scale (1=strongly disagree, and 7=strongly agree). All measurement items can be found in Table S1 in Multimedia Appendix 3.

Data Analysis

Cronbach α was used to determine the reliability of the scales. Construct validity and reliability were examined using exploratory factor analysis with varimax rotation, Bartlett test, and KMO statistics, as well as confirmatory factor analysis (CFA).

For the analysis of structural equation modeling, Amos 26.0 (IBM Corp.) was used to analyze the data and the effect of the model was measured by the fit indices, which included chi-square/degree of freedom (χ^2/df), the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the incremental fit index (IFI), the comparative fit index (CFI), the nonnormed fit index (NNFI), and root mean square error of approximation (RMSEA). Additionally, the bootstrap method was used to test the mediation model with 5000 iterations of repeated sampling. Results were also obtained using SPSS 26.0 (IBM Corp.) to test for interaction effects.

Ethical Considerations

Ethical approval for this study was obtained from the academic committee (acting as the ethics committee) of the School of Journalism and Communication at Chongqing University (approval number 20240320). Informed consent was obtained from all participants before their participation in the survey. Participants were provided with a detailed informed consent form outlining the purpose of the study, the duration of the investigation, the procedures, and the potential risks and benefits. They were informed that their participation was

XSL•FO
RenderX

voluntary and that they had the right to withdraw at any time without facing any negative consequences. Participants were assured of the confidentiality and anonymity of their responses.

Results

Participant Characteristics

The sample demographics are as follows: of the 958 participants, 771 (80.5%) were between the ages of 60 and 75 years, 511 (53.3%) were females, and 447 (46.7%) were males. In terms of residential area, 559 (58.4%) participants resided in a rural area and 399 (41.6%) in an urban area. Most participants (812/958, 84.8%) had between 1 and 3 children, about one-quarter (214/958, 22.3%) of the participants reported an education level of high school and above, nearly half (440/958, 45.9%) rated their self-assessed health status as fair, and 186 (19.4%) rated it as poor or very poor. Detailed demographic information is presented in Table 1.

XSL•FO
RenderX

Table 1. Participant characteristics (N=958).

Characteristics	Values, n (%)
Age (years)	
60-65	444 (46.3)
66-70	183 (19.1)
71-75	144 (15.0)
76-80	123 (12.8)
>80	64 (6.7)
Gender	
Women	511 (53.3)
Male	447 (46.7)
Status of residence	
Residence with spouse	424 (44.3)
Large family living together	188 (19.6)
Living with children	184 (19.2)
Living alone	137 (14.3)
Others	25 (2.6)
Educational level	
Secondary schools	353 (36.8)
Junior high school	243 (25.4)
Not attending school	148 (15.4)
High school/secondary school	138 (14.4)
Undergraduate	53 (5.5)
Graduate students and above	23 (2.4)
Occupation now or before retirement	
Peasants	475 (49.6)
Workers	112 (11.7)
Others	109 (11.4)
Public officials/units	107 (11.2)
Sole trader/freelancer	100 (10.4)
Business managers/office staff	55 (5.7)
Monthly income (CNY[a])	
≤2000	472 (49.3)
2001-4000	286 (29.9)
4001-6000	133 (13.9)
6001-8000	44 (4.6)
≥8001	23 (2.4)
Urban/rural	
Countryside	559 (58.4)
Cities and towns	399 (41.6)
Number of children	
2	452 (47.2)
3	192 (20.0)
1	168 (17.5)

XSL•FO
RenderX

Characteristics	Values, n (%)
4	87 (9.1)
5	30 (3.1)
0	18 (1.9)
≥6	11 (1.1)
Self-assessed health status	
General	440 (45.9)
Better	288 (30.1)
Rather poor	170 (17.7)
Rare	44 (4.6)
Very poor	16 (1.7)

[a]1 CNY=US $0.14.

Structural Model

The CFA model was fit to the data (Table S2 in Multimedia Appendix 3): χ^2_{231}=525.021, GFI=0.958, AGFI=0.945, IFI=0.980, CFI=0.980, NNFI=0.976, and RMSEA=0.036. The model fit appeared to be good, and all items met the criteria proposed by Hu and Bentler [68] and Kline [69].

According to Bagozzi and Youjae [70] and Hair [71], composite reliability (ρ)≥0.60 and average variance extracted ≥0.50 indicate good internal consistency and convergent validity. We therefore used composite reliability to measure the reliability of the constructs, with the following results: intergenerational support=0.887, online medical information=0.896, trust=0.869, PU=0.888, PEOU=0.906, eHealth literacy=0.894, and intention to seek medical care on the internet=0.896. As shown in Table S3 in Multimedia Appendix 3, the factor loadings for all question items were greater than 0.60. Convergent validity was assessed using the average variance extracted, which yielded the following values: 0.723 for intergenerational support, 0.684 for online medical information, 0.690 for trust, 0.726 for PU, 0.659 for PEOU, 0.739 for eHealth literacy, and 0.742 for intention to seek medical care on the internet. These results indicate good convergent validity. As shown in Table S4 in Multimedia Appendix 3, the square root of the average variance extracted for each construct is greater than the correlation coefficients between the variables, indicating good discriminant validity [72].

Based on the final CFA model, a full structural equation modeling was conducted to test the hypotheses. The model fit indices (Table S5 in Multimedia Appendix 3) were as follows: χ^2_{179}=498.30, GFI=0.954, AGFI=0.940, IFI=0.975, CFI=0.975, Tucker-Lewis Index=0.971, and RMSEA=0.043. These indices indicate a good model fit, and the structural equation modeling approach allowed the constructed and hypothesized model to be tested in a more satisfactory manner.

Association of Intergenerational Support and Other Variables and Intention to Seek Medical Care on the Internet

Table 2 shows the results of the path analysis. Intergenerational support was positively associated with PEOU, PU, trust, and older adults' intention to seek medical care on the internet (β=.292, *P*<.001; β=.437, *P*<.001; β=.322, *P*<.001; and β=.354, *P*<.001, respectively); therefore, hypotheses H1a, H1b, H1c, and H1d were supported. PEOU and PU were positively associated with older adults' trust in internet-based medical services (β=.263, *P*<.001 and β=.261, *P*<.001, respectively), supporting H2a and H2b. PEOU and PU were also positively correlated with older adults' intention to seek medical care on the internet (β=.099, *P*=.002 and β=.204, *P*<.001, respectively), supporting H3a and H3b. Trust was positively correlated with older adults' intention to seek medical care on the internet (β=.197, *P*<.001), supporting H4. Online medical information was also positively correlated with older adults' intention to seek medical care on the internet (β=.109, *P*<.001), supporting H5.

XSL•FO
RenderX

Table 2. Results for direct relationships.

Relationship	Path coefficient	SE	Critical ratio	*P* value
PEOU ← IS[a]	0.292	0.037	8.204	<.001
PU ← IS	0.437	0.035	12.388	<.001
T ← IS	0.322	0.034	8.543	<.001
T ← PEOU[b]	0.263	0.029	8.003	<.001
T ← PU[c]	0.261	0.032	7.345	<.001
SMCI[d] ← IS	0.354	0.04	9.265	<.001
SMCI ← T[e]	0.197	0.046	4.961	<.001
SMCI ← PEOU	0.099	0.032	3.088	.002
SMCI ← PU	0.204	0.037	5.809	<.001
SMCI ← OMI[f]	0.109	0.034	3.781	<.001

[a]IS: intergenerational support.

[b]PEOU: perceived ease of use.

[c]PU: perceived usefulness.

[d]SMCI: intention to seek medical care on the internet

[e]T: Trust

[f]OMI: online medical information.

Table 3 shows the mediating relationship between PU, PEOU, and trust. The total effect value of the relationship between intergenerational support and willingness to use internet-based medical services was 0.573, while the direct effect value was 0.354. Both had a positive 95% CI and a *P* value of less than 0.05, indicating that the total and direct effects were significant. The indirect effect of trust was 0.063 (95% CI 0.036-0.099). More specifically, trust partially mediated the association between intergenerational support and older adults' willingness to use internet-based medical services. PEOU, PU, and trust also played a chain mediating role between intergenerational support and willingness to use internet-based medical services, with specific mediation effect sizes of 0.015 and 0.022, respectively. The 95% CI for the PEOU path was 0.008-0.027, and for the PU path, it was 0.012-0.036. The *P* values for both chain mediation paths were less than 0.05, suggesting that intergenerational support can enhance PEOU and PU among older adults. This, in turn, enhances trust, which subsequently increases the willingness to use internet-based medical services.

XSL•FO
RenderX

Table 3. Total, direct, and indirect effects.[a]

Mediation effect and path	Estimate	95% CI	*P* value
Total effect			
IS[b] → SMCI[c]	0.573	0.509-0.632	<.001
Direct effect			
IS → SMCI	0.354	0.278-0.428	<.001
Indirect effect			
IS → T[d] → SMCI	0.063	0.036-0.099	<.001
IS → PEOU[e] → SMCI	0.029	0.009-0.054	0.005
IS → PU[f] → SMCI	0.089	0.050-0.130	<.001
IS → PEOU → T → SMCI	0.015	0.008-0.027	<.001
IS → PU → T → SMCI	0.022	0.012-0.036	<.001

[a]Standardized estimation of 5000 bootstrap samples.

[b]IS: intergenerational support.

[c]SMCI: intention to seek medical care on the internet.

[d]T: trust.

[e]PEOU: perceived ease of use.

[f]PU: perceived usefulness.

Table 4 shows the moderating relationships of eHealth literacy. The interaction term between intergenerational support and eHealth literacy was a significant positive predictor of PEOU (β=.177, *P*<.001) and PU (β=.073, *P*=.018). This suggests that eHealth literacy moderates the effect of intergenerational support on PEOU and PU, thereby supporting H6a and H6b. The interaction term between intergenerational support and eHealth literacy was a significant positive predictor of trust (β=.09, *P*<.01), suggesting that eHealth literacy moderates the effect of intergenerational support on trust, thus supporting H6c. The interaction term between intergenerational support and eHealth literacy had a significant positive predictive effect on the intention to seek medical care on the internet (β=.124, *P*<.001). This indicates that eHealth literacy moderates the influence of intergenerational support on the intention to seek medical care on the internet, thereby supporting H6d. A further simple slope analysis was then conducted (Figure 1). The slope was shallower when eHealth literacy was low (solid line) and steeper when eHealth literacy was high (dashed line). This suggests that intergenerational support was a more significant positive predictor of the intention to seek medical care on the internet, PEOU, PU, and trust at high levels of eHealth literacy.

Table 4. Moderated effect.

Path	Intention to seek medical care on the internet, β (*P* value)	Perceived ease of use, β (*P* value)	Perceived usefulness, β (*P* value)	Trust, β (*P* value)
Intergenerational support	.447 (<.001)	.183 (<.001)	.337 (<.001)	.389 (<.001)
Electronic health literacy[a]	.15 (<.001)	.094 (.002)	.167 (<.001)	.173 (<.001)
IS[b] × eHL[c]	.124 (<.001)	.177 (<.001)	.073 (.018)	.09 (.003)

[a]Electronic health literacy was used as a moderating variable in the model after controlling for the variables age, gender, education, residential status, occupation now or before retirement, monthly income, urban /rural, number of children, self-assessed health status.

[b]IS: intergenerational support.

[c]eHL: eHealth literacy.

XSL•FO
RenderX

Figure 1. Simple slope plot. eHL: eHealth literacy; IS: intergenerational support; PEOU: perceived ease of use; PU: perceived usefulness; SMCI: intention to seek medical care on the internet; T: trust.

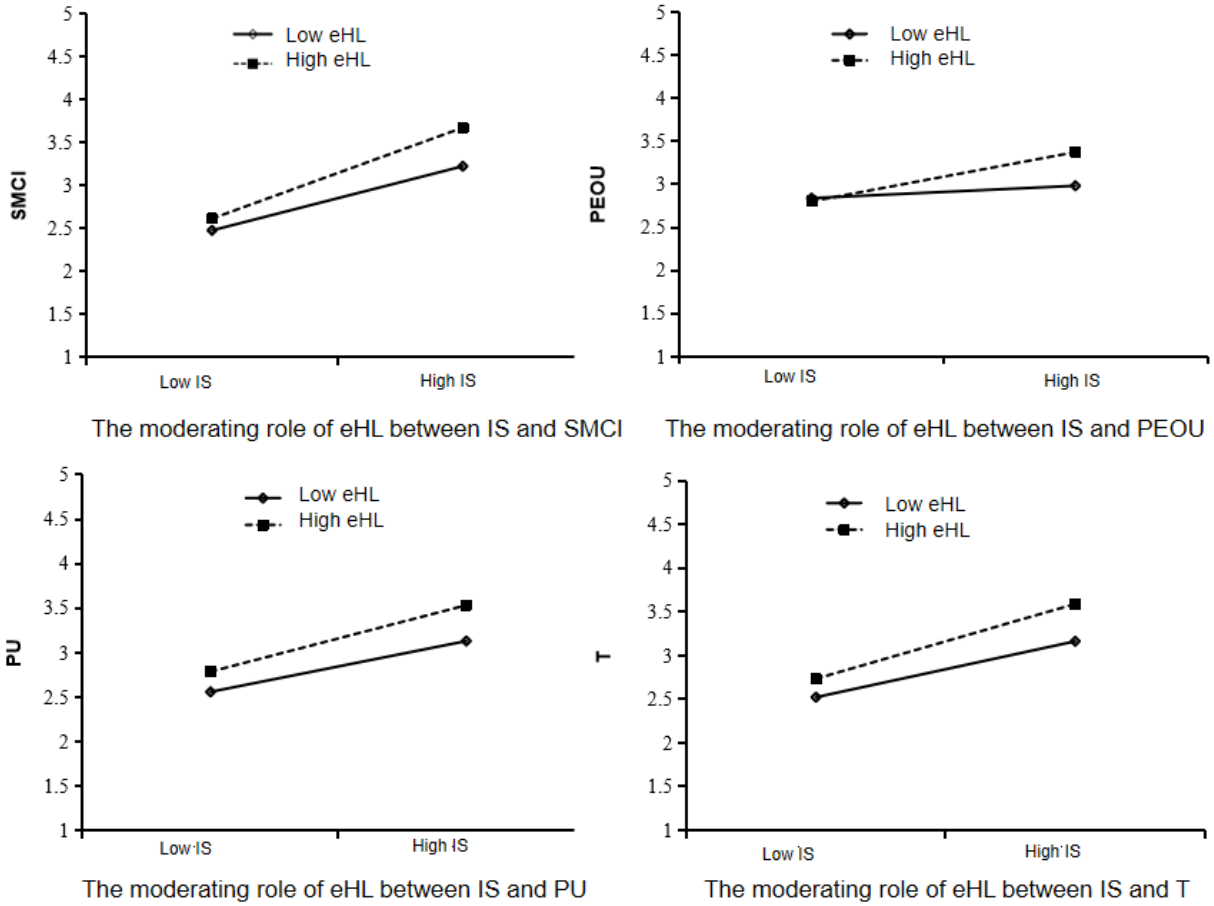

Discussion

Principal Findings

Older people are often referred to as digital refugees [73]. A lack of digital skills may contribute to the risk of widening health disparities and unequal access to services [74]. Studies have shown that many older individuals face numerous barriers to using smartphones, often due to a lack of knowledge about apps and their features [75] and technology anxiety [76]. Most older adult patients are also accustomed to face-to-face outpatient services, making the adoption of internet-based health care not easily achievable for this population. Although the digital divide hinders active aging by reducing the social participation, as well as the physical and mental health, of older adults [77], the active bridging provided by intergenerational support helps older adults overcome their fears, improve their digital skills, and enhance their sense of well-being and accessibility. This, in turn, facilitates their integration into the digital age by fostering a positive and optimistic mindset that promotes effective active aging [78].

Intergenerational support to address the technical aspect occurs when children explain to older adults how to operate the online medical platform, help them understand error messages, and provide tips on various operations such as registration, logging in, booking appointments, and online counseling [79]. This support involves not only technical guidance but also spiritual care and comfort. This kind of benign emotional exchange can promote older people's perception of satisfaction, enhancing their acceptance of new technologies. Therefore, online medical care with intergenerational support is not only a reflection of filial piety but also an endorsement of positive intergenerational relationships. Children's counseling and assistance can improve the cognitive abilities [80] and overall health awareness of older adults, alleviate their fear of technology, reduce digital loneliness, and strengthen their confidence in using internet-based medical services.

The results of this study show that the higher the level of intergenerational support, the greater the PEOU and usefulness of internet health care platforms among older adults, which in turn enhances their willingness to access these services. Through the help and guidance of family members, older adults are able to master internet technologies more quickly and effectively utilize these tools for health management. This support not only increased older adults' PEOU of internet-based health care services but also boosted their confidence and security, leading to a greater willingness to try and use these services. In addition, this support implies encouragement and companionship on an emotional and psychological level [81]. When older adults feel the concern and attention from their family members or later generations regarding their health problems, and the possibility of obtaining a convenient and professional health care experience through internet health care services, they are more likely to develop positive psychological expectations that these services are valuable and can practically solve their health issues [82]. This positive cognitive shift directly enhances older adults'

XSL•FO
RenderX

acceptance of and reliance on internet health care services, thereby strengthening their willingness to seek medical care.

There is also a transformational path between social support factors that generate the intention to adopt internet-based medical services; trust mediates between intergenerational support and the intention to adopt internet-based medical services. Trust, in this context, refers to the extent to which an individual relies on an internet-based medical service to meet health management needs [83], encompassing aspects such as the professionalism of the health care service, platform security, and a positive user experience. Cao et al [84] suggested that mistrust is a common reason for digital disengagement and that trust is crucial for an individual's ongoing behavior when using information technology. This is also true in the internet-based medical service environment [85]. Studies have shown that the realization of internet-based medical services is divided into 2 phases: first, patients develop perceptions about the technical aspects of the health care platform, including its functionality, usefulness, and convenience [86]; second, patients evaluate the trustworthiness of the care provided and the service process [87]. Strong behavioral intentions only occur if older adults have a positive perception of the trustworthiness of the care.

Based on previous studies, this study further clarifies the influence of intergenerational support on older adults' intention to adopt internet-based medical services. It constructs and validates an integrated model of the effect of health communication: intergenerational support → PEOU → PU → trust → intention to adopt internet-based medical services. As a new approach to medical care, internet-based medical services improve convenience and efficiency. With intergenerational support, older adults perceive the technical aspects (PEOU and PU) of these services more positively. This, in turn, enhances their trust and intention to use internet-based medical services, facilitating the integration of online care into health care.

Intergenerational support had a greater effect on the willingness of older adults to use internet-based medical services among those with high eHealth literacy. According to Jung et al [55], eHealth literacy enhances older adults' ability to manage health-related issues effectively. Health literacy has also been positively associated with help-seeking intentions [88], older adults' choice of physician, and their understanding of the physician's recommendations [89]. Older adults with high eHealth literacy are thus more likely to have a higher intention to seek medical care on the internet when they receive support from their children and are more likely to participate in internet-based medical services actively for a better health care experience [90].

Previous studies have shown that PU and trust in online medical information [91] can improve eHealth literacy. However, this study demonstrated that the degree of intergenerational support for perceived technology, as well as trust in internet-based medical services, is deeper when older adults have higher eHealth literacy. According to Nie et al [92], people with higher eHealth literacy better assess the usefulness of searching for health information online. However, it may be that eHealth literacy increases older adults' interest in health, their knowledge, and their expectations of health care [93] while inducing confidence and a sense of the efficacy of internet-based medical services.

Much of the health and medical information provided by internet-based medical service platforms originates from authoritative health care institutions and professionals, ensuring a high degree of accuracy. These platforms can support long-term disease management and health monitoring [94], allowing older people to access the health information they need. Bundorf et al [95] argued that this shift to searching for medical information online has obvious impacts on health behaviors as well as on treatment choices [96], improving older adults' health beliefs and motivating them to engage in healthy activities [97]. There is also evidence that online medical information not only encourages self-initiated participation in health care [98] but also increases the likelihood that older adults will take an active role in disease management. It is thus reasonable to believe, based on the present results, that online medical information can play an active role in disease management and increase the willingness of older adults to use internet-based medical services.

Limitations

Although this study initially verified the proposed theoretical model, certain limitations remain. First, the study focused exclusively on the PEOU and PU of internet-based medical services, leaving other related perceptions (eg, risks and benefits) unexplored. Moreover, it considered only the factors influencing older adults' willingness to use internet-based medical services, without addressing motivation, sustained willingness to use, or actual usage. Finally, this study examined only the influence of the core variable of intergenerational support. Other factors, such as peer influence, social influence (eg, doctors' recommendations), and subjective norms, could be explored in greater depth and incorporated into models in future research.

Implications

Examining technology use among older adults is also a valuable starting point for addressing the range of social issues associated with aging [99]. This study's focus on the needs of older adults regarding internet-based medical services, as well as their intentions to use such services, offers constructive insights for improving health equity and enhancing the inclusiveness and equity of internet-based medical services in particular [100]. Perceptions of and trust in technology are crucial factors in facilitating older adults' adoption of internet-based medical services. Efforts to encourage older adults to use these services may first need to ensure that they recognize such services as not only useful and easy to use but also trustworthy.

Medical institutions and internet-based medical platforms can take the following measures: in platform design, they should fully consider the characteristics and needs of older adults, implement age-adapted designs, lower the barriers to use, and create an age-friendly online environment. Additionally, medical institutions and platforms must establish a highly credible environment by ensuring transparency in the treatment process and maintaining honesty and openness with patients. Protecting the personal information of older adults and preventing data

XSL•FO
RenderX

breaches are essential to enhancing their confidence in using internet-based health care services.

Furthermore, the role of intergenerational support in enhancing older adults' intention to seek medical care online must be emphasized. This highlights the importance of understanding the barriers older adults may face when using internet-based medical services, as well as the critical role children/young adults can play in facilitating the adoption of new technologies by older generations.

Within families, the younger generation should take the initiative to help older adults adapt to digital technology by offering emotional support and technical guidance while leveraging the advantages of the internet for health management.

The use of internet-based medical services by older adults holds significant value for management. First, it improves management efficiency. Internet-based medical services simplify traditional medical processes, enhancing hospital management efficiency. These platforms can allocate medical resources more effectively based on the health status and needs of older adults, optimizing resource utilization. Second, it fosters an innovative service model. Internet-based medical platforms can leverage technologies such as big data and artificial intelligence to deliver more precise medical services. Additionally, they can collaborate with other industries, such as care services for older adults, to innovate and expand service models. Third, it facilitates policy implementation. As a platform for policy dissemination, internet-based medical services can help raise awareness among older adults about national health and pension policies, thereby supporting the effective implementation of these policies. Fourth, it reflects social value. Internet-based medical services address the health needs of older adults, demonstrating social care and respect for older adults. This contributes to the transformation and upgrading of the national health care industry, ultimately improving the overall quality of medical services.

Conclusions

Population aging is a long-term national issue in China, and the availability of medical services is directly linked to the quality of life for older adults. This study provides evidence that intergenerational support is a key factor in promoting the adoption of internet health care services by older adults. Encouraging children to offer guidance and assistance to their parents is essential. To increase older adults' willingness to use internet health care, future efforts should focus not only on intergenerational support but also on improving the PEOU and usefulness of the technology, building trust, and fostering eHealth literacy among older adults.

Acknowledgments

This study was supported by the Project of Humanities and Social Sciences Research of the Ministry of Education (grant 23XJC860002) and the Fundamental Research Fund for the Central Universities (grant 2021CDJSKJC33).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[ZIP File (Zip Archive), 182 KB - jmir_v27i1e65065_app1.zip]

Multimedia Appendix 2
Data (centralized processing).
[ZIP File (Zip Archive), 190 KB - jmir_v27i1e65065_app2.zip]

Multimedia Appendix 3
Additional analysis.
[ZIP File (Zip Archive), 19 KB - jmir_v27i1e65065_app3.zip]

References

1. Takahashi Y, Ohura T, Ishizaki T, Okamoto S, Miki K, Naito M, et al. Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. J Med Internet Res 2011 Dec 14;13(4):e110 [FREE Full text] [doi: 10.2196/jmir.1796] [Medline: 22169526]
2. Zhang C. Smartphones and telemedicine for older people in China: opportunities and challenges. Digit Health 2022;8:20552076221133695 [FREE Full text] [doi: 10.1177/20552076221133695] [Medline: 36353695]
3. China Internet Network Information Center (CNNIC). The 52nd China statistical report on the development of internet in China. CNNIC. 2023 Aug 28. URL: https://www.cnnic.cn/n4/2023/0828/c199-10830.html [accessed 2024-07-14]
4. World Health Organization (WHO). Draft global strategy on digital health 2020-2024. WHO. Geneva, Switzerland: WHO; 2020. URL: https://www.who.int/docs/default-source/documents/gs4dh0c510c483a9a42b1834a8f4d276c6352.pdf [accessed 2024-07-14]

XSL•FO
RenderX

5. Resnick B, Young HM, Fick DM, Kagan SH. Making care for older people the choice of nurses today, tomorrow, and forever. Geriatr Nurs 2022;46:A1-A2 [FREE Full text] [doi: 10.1016/j.gerinurse.2022.04.014] [Medline: 35606212]
6. Sun X, Yan W, Zhou H, Wang Z, Zhang X, Huang S, et al. Internet use and need for digital health technology among the elderly: a cross-sectional survey in China. BMC Public Health 2020 Sep 11;20(1):1386 [FREE Full text] [doi: 10.1186/s12889-020-09448-0] [Medline: 32917171]
7. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
8. Klaver NS, van de Klundert J, van den Broek RJGM, Askari M. Relationship between perceived risks of using mHealth applications and the intention to use them among older adults in the Netherlands: cross-sectional study. JMIR Mhealth Uhealth 2021 Aug 30;9(8):e26845 [FREE Full text] [doi: 10.2196/26845] [Medline: 34459745]
9. Palas JU, Sorwar G, Hoque MR, Sivabalan A. Factors influencing the elderly's adoption of mHealth: an empirical study using extended UTAUT2 model. BMC Med Inform Decis Mak 2022 Jul 24;22(1):191 [FREE Full text] [doi: 10.1186/s12911-022-01917-3] [Medline: 35871682]
10. Maxham JG. Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. Journal of Business Research 2001 Oct;54(1):11-24. [doi: 10.1016/s0148-2963(00)00114-4]
11. Ma Y, Luo M. Older people's intention to use medical apps during the COVID-19 pandemic in China: an application of the Unified Theory of Acceptance and Use of Technology (UTAUT) model and the Technology of Acceptance Model (TAM). Ageing and Society 2022 Oct 17;44(7):1515-1532. [doi: 10.1017/s0144686x22000423]
12. Tu J, Luo SC, Lee Y, Shih M, Chiu S. Exploring usability and patient attitude towards a smart hospital service with the Technology Acceptance Model. Int J Environ Res Public Health 2022 May 16;19(10):6059 [FREE Full text] [doi: 10.3390/ijerph19106059] [Medline: 35627595]
13. Fu Y, Tang T, Long J, Lin B, Li J, Quan G, et al. Factors associated with using the internet for medical information based on the doctor-patient trust model: a cross-sectional study. BMC Health Serv Res 2021 Nov 24;21(1):1268 [FREE Full text] [doi: 10.1186/s12913-021-07283-6] [Medline: 34819044]
14. Ren D, Ma B. Effectiveness of interactive tools in online health care communities: social exchange theory perspective. J Med Internet Res 2021 Mar 12;23(3):e21892 [FREE Full text] [doi: 10.2196/21892] [Medline: 33709940]
15. Ren D, Ma B. Influences of governance mechanisms on patients' usage intention: a study on web-based consultation platforms. Health Informatics J 2023;29(1):14604582231153509 [FREE Full text] [doi: 10.1177/14604582231153509] [Medline: 36657942]
16. Mansson L, Wiklund M, Öhberg F, Danielsson K, Sandlund M. Co-creation with older adults to improve user-experience of a smartphone self-test application to assess balance function. Int J Environ Res Public Health 2020 May 26;17(11):3768 [FREE Full text] [doi: 10.3390/ijerph17113768] [Medline: 32466484]
17. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in the Netherlands: cross-sectional study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]
18. Morgan SP, Hirosima K. The persistence of extended family residence in Japan: anachronism or alternative strategy? American Sociological Review 1983 Apr;48(2):269-281. [doi: 10.2307/2095111]
19. Chen J, Jordan LP. Intergenerational support and life satisfaction of young-, old- and oldest-old adults in China. Aging Ment Health 2018 Mar;22(3):412-420. [doi: 10.1080/13607863.2016.1261798] [Medline: 27918203]
20. Kim SD. Application and challenges of the technology acceptance model in elderly healthcare: insights from ChatGPT. Technologies 2024 May 13;12(5):68. [doi: 10.3390/technologies12050068]
21. McGaughey RE, Zeltmann SM, McMurtrey ME. Motivations and obstacles to smartphone use by the elderly: developing a research framework. IJEF 2013;7(3/4):177. [doi: 10.1504/ijef.2013.058601]
22. Yuan Z, Jia G. Profiling the digital divide of the elderly based on internet big data: evidence from China. Data Science and Management 2021 Sep;3:33-43. [doi: 10.1016/j.dsm.2021.10.001]
23. Almulhem JA. Factors, barriers, and recommendations related to mobile health acceptance among the elderly in Saudi Arabia: a qualitative study. Healthcare (Basel) 2023 Nov 23;11(23):3024. [doi: 10.3390/healthcare11233024] [Medline: 38063592]
24. Wang J, Katz I, Li J, Wu Q, Dai C. Mobile digital divide and older people’s access to ‘Internet plus social work’: implications from the COVID-19 help-seeking cases. Asia Pacific Journal of Social Work and Development 2020 Nov 26;31(1-2):52-58. [doi: 10.1080/02185385.2020.1850332]
25. Sharifi S, Babaei Khorzoughi K, Khaledi-Paveh B, Rahmati M. Association of intergenerational relationship and supports with cognitive performance in older adults: a systematic review. Geriatr Nurs 2023;52:146-151. [doi: 10.1016/j.gerinurse.2023.05.014] [Medline: 37307634]
26. Wei D, Xu A, Wu X. The mediating effect of trust on the relationship between doctor-patient communication and patients' risk perception during treatment. Psych J 2020 Jun;9(3):383-391. [doi: 10.1002/pchj.327] [Medline: 31795011]
27. Lee OE, Kim D. Bridging the digital divide for older adults via intergenerational mentor-up. Research on Social Work Practice 2018 Nov 05;29(7):786-795. [doi: 10.1177/1049731518810798]

XSL•FO
RenderX

28. Ma T, Zhang S, Zhu S, Ni J, Wu Q, Liu M. The new role of nursing in digital inclusion: reflections on smartphone use and willingness to increase digital skills among Chinese older adults. Geriatr Nurs 2022;48:118-126 [FREE Full text] [doi: 10.1016/j.gerinurse.2022.09.004] [Medline: 36155310]
29. Pan Z, Chen J. Association of received intergenerational support with subjective well-being among elderly: the mediating role of optimism and sex differences. Int J Environ Res Public Health 2022 Jun 22;19(13):7614 [FREE Full text] [doi: 10.3390/ijerph19137614] [Medline: 35805273]
30. Li H, Kostka G. Navigating the digital age: the gray digital divide and digital inclusion in China. Media, Culture & Society 2024 Feb 21;46(6):1181-1199. [doi: 10.1177/01634437241229382]
31. Yeh K, Yi C, Tsao W, Wan P. Filial piety in contemporary Chinese societies: a comparative study of Taiwan, Hong Kong, and China. International Sociology 2013 May 14;28(3):277-296. [doi: 10.1177/0268580913484345]
32. Correa T. Bottom-up technology transmission within families: exploring how youths influence their parents' digital media use with dyadic data. J Commun 2013 Dec 02;64(1):103-124. [doi: 10.1111/jcom.12067]
33. Ekström KM. Parental consumer learning or 'keeping up with the children'. J of Consumer Behaviour 2007 Sep 13;6(4):203-217. [doi: 10.1002/cb.215]
34. Ma J, Fang G, Guo K. Bridging the digital divide: the influence of digital feedback on the digital capabilities of the rural elderly. Information Development 2023 Dec 28;39:02666669231222208. [doi: 10.1177/02666669231222208]
35. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
36. Harrigan M, Feddema K, Wang S, Harrigan P, Diot E. How trust leads to online purchase intention founded in perceived usefulness and peer communication. J of Consumer Behaviour 2021 Mar 03;20(5):1297-1312. [doi: 10.1002/cb.1936]
37. Acharya N, Sassenberg A, Soar J. Consumers' behavioural intentions to reuse recommender systems: assessing the effects of trust propensity, trusting beliefs and perceived usefulness. JTAER 2022 Dec 22;18(1):55-78. [doi: 10.3390/jtaer18010004]
38. Martínez-Navalón J, Fernández-Fernández M, Alberto FP. Does privacy and ease of use influence user trust in digital banking applications in Spain and Portugal? Int Entrep Manag J 2023 Feb 22;19(2):781-803. [doi: 10.1007/s11365-023-00839-4]
39. Kim Y, Park Y, Choi J. A study on the adoption of IoT smart home service: using value-based adoption model. Total Quality Management & Business Excellence 2017 Apr 12;28(9-10):1149-1165. [doi: 10.1080/14783363.2017.1310708]
40. Stone R, Baker Eveleth L. User's perceptions of perceived usefulness, satisfaction, and intentions of mobile application. IJMC 2020;18(1):1-18. [doi: 10.1504/ijmc.2020.10025192]
41. Naidoo R, Leonard A. Perceived usefulness, service quality and loyalty incentives: effects on electronic service continuance. SAJBM 2007 Sep 30;38(3):39-48. [doi: 10.4102/sajbm.v38i3.587]
42. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
43. Chen Y, Xu Q. The willingness to use mobile health and its influencing factors among elderly patients with chronic heart failure in Shanghai, China. Int J Med Inform 2021 Dec 16;158:104656. [doi: 10.1016/j.ijmedinf.2021.104656] [Medline: 34933173]
44. Schoorman FD, Mayer RC, Davis JH. An integrative model of organizational trust: past, present, and future. AMR 2007 Apr;32(2):344-354. [doi: 10.5465/amr.2007.24348410]
45. Doney PM, Cannon JP. An examination of the nature of trust in buyer-seller relationships. Journal of Marketing 1997 Apr;61(2):35-51. [doi: 10.2307/1251829]
46. Bu W, Xing L, Xiao M, Chang P, Du Y, Xie H. A survey on the willingness of Ganzhou residents to participate in "internet + nursing services" and associated factors. J Multidiscip Healthc 2022;15:897-906 [FREE Full text] [doi: 10.2147/JMDH.S351071] [Medline: 35509727]
47. Zulman DM, Kirch M, Zheng K, An LC. Trust in the internet as a health resource among older adults: analysis of data from a nationally representative survey. J Med Internet Res 2011 Feb 16;13(1):e19 [FREE Full text] [doi: 10.2196/jmir.1552] [Medline: 21324832]
48. Roshandel D, Rezailashkajani M, Ansari S, Zali MR. Internet use by a referral gastroenterology clinic population and their medical information preferences. Int J Med Inform 2005 Jul;74(6):447-459. [doi: 10.1016/j.ijmedinf.2005.05.003] [Medline: 15963757]
49. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
50. Smith AD, Manna DR. Exploring the trust factor in e‐medicine. Online Information Review 2004 Oct;28(5):346-355. [doi: 10.1108/14684520410564271]
51. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
52. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

XSL•FO
RenderX

53. Peng Y, Yin P, Deng Z, Wang R. Patient-physician interaction and trust in online health community: the role of perceived usefulness of health information and services. Int J Environ Res Public Health 2019 Dec 24;17(1):139 [FREE Full text] [doi: 10.3390/ijerph17010139] [Medline: 31878145]
54. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
55. Jung SO, Son YH, Choi E. E-health literacy in older adults: an evolutionary concept analysis. BMC Med Inform Decis Mak 2022 Jan 31;22(1):28 [FREE Full text] [doi: 10.1186/s12911-022-01761-5] [Medline: 35101005]
56. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
57. Lanxin W, Yan Z, Yutong T, Lixue M, Li L, Ting Z. Potential profiling of self-management skills in older co-morbid patients. BMC Geriatr 2024 Jun 25;24(1):555 [FREE Full text] [doi: 10.1186/s12877-024-05137-4] [Medline: 38918703]
58. Liu S, Zhao H, Fu J, Kong D, Zhong Z, Hong Y, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in Southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996. [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
59. Arora S, Huda RK, Verma S, Khetan M, Sangwan RK. Challenges, barriers, and facilitators in telemedicine implementation in India: a scoping review. Cureus 2024 Aug;16(8):e67388. [doi: 10.7759/cureus.67388] [Medline: 39310647]
60. Wong AKC, Bayuo J, Wong FKY. Investigating predictors of self-care behavior among homebound older adults: the role of self-efficacy, eHealth literacy, and perceived social support. J Nurs Scholarsh 2022 May;54(3):278-285. [doi: 10.1111/jnu.12730] [Medline: 34766694]
61. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
62. Lang FR, Schütze Y. Adult children's supportive behaviors and older parents' subjective well–being—a developmental perspective on intergenerational relationships. Journal of Social Issues 2003 Feb 12;58(4):661-680. [doi: 10.1111/1540-4560.00283]
63. Bland JM, Altman DG. Cronbach's alpha. BMJ 1997 Feb 22;314(7080):572 [FREE Full text] [doi: 10.1136/bmj.314.7080.572] [Medline: 9055718]
64. de Boer MJ, Versteegen GJ, van Wijhe M. Patients' use of the internet for pain-related medical information. Patient Educ Couns 2007 Sep;68(1):86-97. [doi: 10.1016/j.pec.2007.05.012] [Medline: 17590563]
65. Fogel J, Nehmad E. Internet social network communities: risk taking, trust, and privacy concerns. Computers in Human Behavior 2009 Jan;25(1):153-160. [doi: 10.1016/j.chb.2008.08.006]
66. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
67. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
68. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
69. Kline R. Principles and Practice of Structural Equation Modeling. New York, NY: Guilford Publications; 2023.
70. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/bf02723327]
71. Hair J. Multivariate Data Analysis. 7th edition. London, UK: Pearson Prentice Hall; Feb 23, 2009.
72. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]
73. Peine A, Neven L. From intervention to co-constitution: new directions in theorizing about aging and technology. Gerontologist 2019 Jan 09;59(1):15-21. [doi: 10.1093/geront/gny050] [Medline: 29850812]
74. Bakhtiar M, Elbuluk N, Lipoff JB. The digital divide: how COVID-19's telemedicine expansion could exacerbate disparities. J Am Acad Dermatol 2020 Nov;83(5):e345-e346 [FREE Full text] [doi: 10.1016/j.jaad.2020.07.043] [Medline: 32682890]
75. Hsieh M, Ho C, Lee I. Effects of smartphone numeric keypad designs on performance and satisfaction of elderly users. International Journal of Industrial Ergonomics 2022 Jan;87:103236. [doi: 10.1016/j.ergon.2021.103236]
76. Caspi A, Daniel M, Kavé G. Technology makes older adults feel older. Aging Ment Health 2019 Aug;23(8):1025-1030. [doi: 10.1080/13607863.2018.1479834] [Medline: 30475067]
77. Liu L, Wu F, Tong H, Hao C, Xie T. The digital divide and active aging in China. Int J Environ Res Public Health 2021 Dec 01;18(23):12675 [FREE Full text] [doi: 10.3390/ijerph182312675] [Medline: 34886400]
78. Amaral I, Daniel F. The use of social media among senior citizens in Portugal: active ageing through an intergeneration approach. Las Vegas, NV: Champringer International Publishing; 2018 Presented at: 4th International Conference, ITAP 2018, Held as Part of HCI International 2018; July 15-20, 2018; Las Vegas, NV p. 422-434. [doi: 10.1007/978-3-319-92034-4_32]

XSL•FO
RenderX

79. Martínez C, Olsson T. The warm expert—a warm teacher? Learning about digital media in intergenerational interaction. Convergence: The International Journal of Research into New Media Technologies 2022 Feb 03;28(6):1861-1877. [doi: 10.1177/13548565211070409]
80. Li Y, Guo M. Filial piety matters: a study of intergenerational supports and parental health. SSM Popul Health 2022 Jun;18:101096 [FREE Full text] [doi: 10.1016/j.ssmph.2022.101096] [Medline: 35493409]
81. Zhou C, Qian Y, Kaner J. A study on smart home use intention of elderly consumers based on technology acceptance models. PLoS One 2024;19(3):e0300574 [FREE Full text] [doi: 10.1371/journal.pone.0300574] [Medline: 38536849]
82. Wang J, Fu Y, Lou V, Tan SY, Chui E. A systematic review of factors influencing attitudes towards and intention to use the long-distance caregiving technologies for older adults. Int J Med Inform 2021 Sep;153:104536. [doi: 10.1016/j.ijmedinf.2021.104536] [Medline: 34325206]
83. Thom DH, Stanford Trust Study Physicians. Physician behaviors that predict patient trust. J Fam Pract 2001 Apr;50(4):323-328. [Medline: 11300984]
84. Cao X, Zhang H, Zhou B, Wang D, Cui C, Bai X. Factors influencing older adults' acceptance of voice assistants. Front Psychol 2024 Mar 07;15(2):1376207. [doi: 10.3389/fpsyg.2024.1376207] [Medline: 38515974]
85. Zhang X, Yan X, Cao X, Sun Y, Chen H, She J. The role of perceived e-health literacy in users' continuance intention to use mobile healthcare applications: an exploratory empirical study in China. Information Technology for Development 2017 Mar 09;24(2):198-223. [doi: 10.1080/02681102.2017.1283286]
86. Joshi K. A model of users' perspective on change: the case of information systems technology implementation. MIS Quarterly 1991 Jun;15(2):229-242. [doi: 10.2307/249384]
87. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
88. Suka M, Yamauchi T, Sugimori H. Help-seeking intentions for early signs of mental illness and their associated factors: comparison across four kinds of health problems. BMC Public Health 2016 Apr 07;16:301 [FREE Full text] [doi: 10.1186/s12889-016-2998-9] [Medline: 27056546]
89. Xie B. Older adults' health information wants in the internet age: implications for patient-provider relationships. J Health Commun 2009 Sep;14(6):510-524. [doi: 10.1080/10810730903089614] [Medline: 19731124]
90. Waterworth S, Honey M. On-line health seeking activity of older adults: an integrative review of the literature. Geriatr Nurs 2018;39(3):310-317. [doi: 10.1016/j.gerinurse.2017.10.016] [Medline: 29198622]
91. Campbell RJ, Nolfi DA. Teaching elderly adults to use the internet to access health care information: before-after study. J Med Internet Res 2005 Jun 30;7(2):e19 [FREE Full text] [doi: 10.2196/jmir.7.2.e19] [Medline: 15998610]
92. Nie L, Oldenburg B, Cao Y, Ren W. Continuous usage intention of mobile health services: model construction and validation. BMC Health Serv Res 2023 May 05;23(1):442 [FREE Full text] [doi: 10.1186/s12913-023-09393-9] [Medline: 37143005]
93. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and validity of the telephone-based eHealth literacy scale among older adults: cross-sectional survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
94. Yuan Y, Wang S, Tao C, Gu Z, Kitayama A, Yanagihara K, et al. Mapping trends and hotspots regarding the use of telenursing for elderly individuals with chronic diseases: a bibliometric analysis. Medicine (Baltimore) 2024 Mar 01;103(9):e37313 [FREE Full text] [doi: 10.1097/MD.0000000000037313] [Medline: 38428870]
95. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
96. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
97. Dutta-Bergman MJ. Health attitudes, health cognitions, and health behaviors among internet health information seekers: population-based survey. J Med Internet Res 2004 May 28;6(2):e15 [FREE Full text] [doi: 10.2196/jmir.6.2.e15] [Medline: 15249264]
98. Sassenberg K, Greving H. Internet searching about disease elicits a positive perception of own health when severity of illness is high: a longitudinal questionnaire study. J Med Internet Res 2016 Mar 04;18(3):e56 [FREE Full text] [doi: 10.2196/jmir.5140] [Medline: 26944335]
99. Amini R, Chee KH, Mendieta M, Parker S. Online engagement and cognitive function among older adults. Geriatr Gerontol Int 2019 Sep;19(9):918-923 [FREE Full text] [doi: 10.1111/ggi.13749] [Medline: 31368165]
100. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]

Abbreviations

AGFI: adjusted goodness of fit index

XSL•FO
RenderX

CFA: confirmatory factor analysis
CFI: comparative fit index
eHL: eHealth literacy
GFI: goodness of fit index
IFI: incremental fit index
KMO: Kaiser-Meier-Olkin
mHealth: mobile health
NNFI: nonnormed fit index
OMI: online medical information
PEOU: perceived ease of use
PU: perceived usefulness
RMSEA: root mean square error of approximation
SMCI: intention to seek medical care on the internet

Edited by A Schwartz; submitted 04.08.24; peer-reviewed by X Zou, Y Liu; comments to author 07.10.24; revised version received 28.10.24; accepted 29.11.24; published 06.01.25.

Please cite as:
Jin H, Qu Y
Association Between Intergenerational Support, Technology Perception and Trust, and Intention to Seek Medical Care on the Internet Among Chinese Older Adults: Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e65065
URL: https://www.jmir.org/2025/1/e65065
doi:10.2196/65065
PMID:

©Hengjiang Jin, Ying Qu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Jargon and Readability in Plain Language Summaries of Health Research: Cross-Sectional Observational Study

Iain A Lang[1], DPhil; Angela King[1], MA; Kate Boddy[1], MSc; Ken Stein[1†], MD; Lauren Asare[1], MSc; Jo Day[1], PhD; Kristin Liabo[1], PhD

Department of Health and Community Sciences, University of Exeter Medical School, University of Exeter, Exeter, United Kingdom
[†]deceased

Corresponding Author:
Iain A Lang, DPhil
Department of Health and Community Sciences
University of Exeter Medical School
University of Exeter
South Cloisters
St Luke's Campus
Exeter,
United Kingdom
Phone: 44 7500 786180
Email: i.lang@exeter.ac.uk

Abstract

Background: The idea of making science more accessible to nonscientists has prompted health researchers to involve patients and the public more actively in their research. This sometimes involves writing a plain language summary (PLS), a short summary intended to make research findings accessible to nonspecialists. However, whether PLSs satisfy the basic requirements of accessible language is unclear.

Objective: We aimed to assess the readability and level of jargon in the PLSs of research funded by the largest national clinical research funder in Europe, the United Kingdom's National Institute for Health and Care Research (NIHR). We also aimed to assess whether readability and jargon were influenced by internal and external characteristics of research projects.

Methods: We downloaded the PLSs of all NIHR National Journals Library reports from mid-2014 to mid-2022 (N=1241) and analyzed them using the Flesch Reading Ease (FRE) formula and a jargon calculator (the De-Jargonizer). In our analysis, we included the following study characteristics of each PLS: research topic, funding program, project size, length, publication year, and readability and jargon scores of the original funding proposal.

Results: Readability scores ranged from 1.1 to 70.8, with an average FRE score of 39.0 (95% CI 38.4-39.7). Moreover, 2.8% (35/1241) of the PLSs had an FRE score classified as "plain English" or better; none had readability scores in line with the average reading age of the UK population. Jargon scores ranged from 76.4 to 99.3, with an average score of 91.7 (95% CI 91.5-91.9) and 21.7% (269/1241) of the PLSs had a jargon score suitable for general comprehension. Variables such as research topic, funding program, and project size significantly influenced readability and jargon scores. The biggest differences related to the original proposals: proposals with a PLS in their application that were in the 20% most readable were almost 3 times more likely to have a more readable final PLS (incidence rate ratio 2.88, 95% CI 1.86-4.45). Those with the 20% least jargon in the original application were more than 10 times as likely to have low levels of jargon in the final PLS (incidence rate ratio 13.87, 95% CI 5.17-37.2). There was no observable trend over time.

Conclusions: Most of the PLSs published in the NIHR's National Journals Library have poor readability due to their complexity and use of jargon. None were readable at a level in keeping with the average reading age of the UK population. There were significant variations in readability and jargon scores depending on the research topic, funding program, and other factors. Notably, the readability of the original funding proposal seemed to significantly impact the final report's readability. Ways of improving the accessibility of PLSs are needed, as is greater clarity over who and what they are for.

(J Med Internet Res 2025;27:e50862) doi:10.2196/50862

XSL•FO
RenderX

KEYWORDS

readability; jargon; reading; accessibility; health research; science communication; public understanding of science; open science; patient and public involvement; health literacy; plain language summary; health communication

Introduction

In recent years, the idea that science should involve and be accessible to nonscientists has grown. Activities such as patient and public involvement, citizen science, open science, and research coproduction represent different facets of this development and are grounded in both practical and normative motives [1-7]. In health research, one aspect of this involves writing a plain language summary (PLS), sometimes also called a "plain English summary" or lay summary. PLSs are short summaries of a study or project intended to increase its accessibility to nonspecialists. Many regulatory agencies and research funders now require them. For example, the European Union requires a PLS as part of the reporting of all clinical trials [8], and PLSs must be included in all systematic reviews published in the Cochrane Library [9] and all proposals submitted to funders such as the Medical Research Council and the National Institute for Health and Care Research (NIHR), the 2 major state-backed funders of health research in the United Kingdom.

There is evidence to suggest that PLSs do not improve accessibility to the extent we might hope. For example, patients enrolled in clinical trials often have a partial or incorrect understanding of the trial in which they are participating, such as its risks and benefits, despite requirements that they be informed about these issues [10-12]. Pharmaceutical and other industry groups have proposed standards for the preparation of PLS [13-15]. Existing guidelines on how to write a PLS vary and are sometimes contradictory [16] but often highlight issues around readability and avoidance of jargon [17]. Studies of how research findings are disseminated to nonspecialists usually focus on potential users of research and what they need to do to understand research better rather than on examining how characteristics of researchers and research settings influence dissemination to nonacademic audiences [18,19]. Like those who develop public health guidelines, researchers may emphasize internal validity (confidence in the reliability of the results) over external validity (whether and how the results can be applied in other places) [20]. Studies of public health researchers [21] and dissemination and implementation researchers [22,23] found that the context in which research is funded and conducted influences efforts to communicate it to nonspecialists, but we are aware of no studies comparing across fields or subfields. More knowledge about how different characteristics of research influence attempts to communicate the research could help improve communication in the future.

Our aim was to assess the readability and level of jargon in the PLSs of research funded by the United Kingdom's NIHR. The NIHR is the largest single funding body in the United Kingdom and the largest national clinical research funder in Europe [24]. Full reports of all projects funded in its major research programs are published on the web. Since 2014, it has been obligatory for these to include a PLS that sets out a clear, simple summary of research in a way that is accessible to nonspecialists and members of the public. NIHR guidance for researchers on PLSs [25] is that they should follow "a few simple rules" that include "avoid, wherever possible, jargon, abbreviations, and technical terms," "avoid complicated language or uncommon words," and "keep sentences short." In this study, we addressed 3 questions about these PLSs: How readable are they? How much jargon do they contain? Are readability and use of jargon influenced by study characteristics such as topic and size?

Methods

Overview

Our data came from all NIHR Journals Library reports published from mid-2014, when the requirement for inclusion of a PLS was introduced, to mid-2022 (May 30). We downloaded the full text of each report from the NIHR website, where they are publicly available [26]. We then used a purpose-written computer program (written in the computing language called Python) to go through each text and find the PLSs. In a few instances, we could not process the reports in this way, in which case we looked up the PLS manually. We had complete data on 1241 PLSs that were part of reports published in the NIHR Journals Library. Apart from 5 reports that did not have a PLS, we included all the reports published during this period.

Ethical Considerations

Our study was based on open-access data related to academic publications and we did not engage directly with the individuals or groups making or receiving these publications. As such, our study was exempt from ethical review as per the terms of the authors' institutional ethics policy and framework (University of Exeter Research Ethics Policy and Framework, Paragraph 4.3.1).

Outcome Variables

We measured readability using the Flesch Reading Ease (FRE) formula. The FRE is often used in analyzing written health information [27] as well as other scientific texts [28,29] and is based on the idea that longer words and longer sentences make a text less readable. Each text is given a score that gets lower in proportion to the number of longer words and sentences used, so a higher score indicates a text that is easier to read. The formula is as follows:



FRE scores can be categorized as "extremely easy," "very easy," "fairly easy," and so on, down to "very difficult," as well as by approximate reading age (Table 1). We used a short prewritten computer program [30] to calculate readability scores for each PLS. In the rest of this document, when we refer to readability scores, we mean FRE scores calculated in this way.

XSL•FO
RenderX

Table 1. Distribution of summaries by Flesch Reading Ease score classification.

	Approximate reading age (years)	Difficulty	Values, n (%)
Readability score			
≥100	Up to 10	Extremely easy	0
90-100	11	Very easy	0
80-90	12	Easy	0
70-80	13	Fairly easy	3 (0.2)
60-70	14-15	Plain English	32 (2.6)
50-60	16-18	Fairly difficult	169 (13.6)
30-50	Undergraduate: 18-21	Difficult	768 (61.9)
0-30	Postgraduate: ≥21	Very difficult	269 (21.7)

We measured jargon using a calculator called the "De-Jargonizer." It was created to help scientists engage with the public, and it identifies jargon based on the frequency with which words appear in everyday English usage. The developers of the calculator analyzed more than 90 million words used in around 250,000 papers on the British Broadcasting Corporation websites (including news, sports, and science pages) [31] during the years 2012-2015. Based on existing work about how commonly words are used and understood in everyday communication, they categorized the words into high frequency (belonging to the 2000 most common word families, which each appeared more than 1000 times), mid-frequency (appearing between 50 and 1000 times), and jargon (fewer than 50 appearances). Acronyms, which can often be part of jargon, are treated the same way as words, which means that common acronyms such as NHS (National Health Service) or USA (United States of America) fall into the "high frequency" category. Full details of how the calculator was put together, including testing and validation, have been published [32], and a web-based version of the calculator contains a description and additional details [33].

The developers of the calculator also created a score to indicate how suitable a text was for a general audience. If a text uses only common words, the score is 100; lower scores indicate more use of mid-frequency and jargon words. The score is calculated using this formula:



We downloaded the source code [34] for the calculator and used it to work out a jargon score for each PLS. A higher score means that less jargon was used.

We created additional outcome variables to identify PLSs that were better in terms of readability and jargon. For readability scores, we focused on summaries with scores of >50: those classed as "fairly difficult to read" or better. A total of 204 summaries (16.4%, 204/1241) fell into this category. The starting point of this study related to the average reading age in the United Kingdom (see the "Patient and Public Involvement" section). We created an outcome variable showing whether a readability score was suitable for a reading age of 9 years—that is, a readability score of 100 or above (Table 1). However, no summaries fell into this category. Rakedzon and colleagues [32] refer to 2 levels of jargon (2% or 5%) as "recommended for general comprehension." We used the more generous 5% score and categorized scores of more than 95 as having low levels of jargon. A total of 269 (21.7%, 269/1241) summaries fell into this category. Multimedia Appendix 1 shows examples of PLSs with high and low readability and jargon.

Study Characteristics

We used the following information relating to each PLS in our analyses:

1. Research topic: NIHR classifies all research projects using the UK Clinical Research Collaboration Health Research Classification System. This is a system for categorizing health research funding and is used to allow funders and others to assess funding schemes [35]. We used the Health Categories dimension of the Health Research Classification System, which groups each project into 1 or more of 21 categories. In our analysis, we included categories if they had at least 50 projects associated with them, which left us with 12 categories: cancer and neoplasms, cardiovascular, infection, mental health, metabolic and endocrine, musculoskeletal, neurological, oral and gastrointestinal, reproductive health and childbirth, respiratory, stroke, and "generic health relevance."
2. Funding program: Reports published in the National Journals Library relate to 5 NIHR funding programs: Efficacy and Mechanism Evaluation, Health Technology Assessment, Health and Social Care Delivery Research, Programme Grants for Applied Research, and Public Health Research. Of these, the Efficacy and Mechanism Evaluation program is generally considered most "upstream" (closer to basic than applied science), and the Health and Social Care Delivery and Public Health Research programs are most "downstream" (most applied). Details of all NIHR programs are available on the web [36].
3. Project size: We wanted to know whether the size of a project made a difference to the PLS, and we used the amount of funding as an approximate measure for this. Smaller projects tend to be more focused and contained; larger projects may have more resources to move around and support activities such as public engagement (which does not necessarily mean that a "better" PLS will be produced). We categorized projects by size by ranking them

XSL•FO
RenderX

in terms of the amount of funding received and then sorting them into 5 equal groups: top 20%, next 20%, and so on.

4. Length in words: When preparing their report for the National Journals Library, authors are asked to write a PLS of up to 300 words, but some write shorter summaries, and some write longer ones. Studies of patient-information leaflets in trials have found that long [37,38] and short [39] texts each have problems associated with them regarding readability and clarity, and we wanted to see whether the length of the PLS made a difference.
5. Readability and jargon scores of the original funding proposal: Information about all projects funded by the NIHR is publicly available on the web, including a copy of the PLS submitted as part of the original funding proposal. The difference between this and the final National Journals Library PLS is that the original one sets out what the researchers proposed to do; the PLS in the National Journals Library summarizes what they ultimately did and found. NIHR instructs writers of summaries to "follow the same principles and procedures as in writing the plain language summary that accompanied your funding submission" [40], and we wanted to find out whether original and final report summaries were written in similar ways. Just as we did for the National Journals Library PLSs, we calculated readability and jargon scores for each of the funding proposal PLSs. We categorized these scores by ranking them and then sorting them into 5 equal groups.
6. Publication year: PLSs became a requirement in mid-2014, and we downloaded our data in mid-2022 (July 22). We wanted to see whether the readability and use of jargon in PLSs had changed over time and did this by categorizing each one according to the year it was published.

Analysis

We analyzed our data using Stata/SE (version 17.0; StataCorp LLC) and Microsoft Excel (version 2409; Microsoft Corporation). For each of the study characteristics described in the preceding section (research topic, funding program, project size, length in words, readability or jargon used in the original proposal, and publication year), we calculated descriptive statistics on the number of PLSs in each category or, where categories were of equal size, the range covered by the category. For each category, we estimated the average readability and jargon score, the percentage of summaries with readability scores higher than 50, the percentage of summaries with jargon scores higher than 95, and the 95% CIs for these estimates. We looked at the relationship between readability and jargon scores by calculating their pairwise (Pearson) correlation. We estimated incidence rate ratios and 95% CIs of PLSs in the hardest-to-read categories (readability scores of >50 and jargon scores of >95) using a generalized linear model with a modified Poisson approach and robust error variances [41]. We report the results of a model in which all study characteristics were entered simultaneously.

Patient and Public Involvement

This study was prompted by the assertion made at a patient and public involvement meeting, attended by AK (who is a patient and member of the public but not a researcher) and hosted by a leading UK research-funding charity, that the average reading age of the UK population is 9 years. This assertion (along with a few variations) can be found on numerous websites by searching on the web for the term "UK average reading age" (we subsequently found that these figures come from the UK Government's Skills for Life Survey [42]). This raised the question: if this is the case, how accessible to the general population are the PLSs routinely produced in health research funding applications and reports? AK and IAL's discussions about how to address this question led to the writing of this paper. AK has been involved throughout and is a coauthor.

Results

Readability scores in our sample ranged from 1.1 to 70.8. The mean (average) FRE score was 39.0 (95% CI 38.4-39.7), and the median (middle) score was 39.8. The distribution of scores across readability categories is shown in Table 1. Around one-fifth of summaries had a score below 30, "very difficult to read."

Jargon scores in our sample ranged from 76.4 to 99.3. The mean (average) jargon score was 91.7 (95% CI 91.5-91.9), and the median (middle) score was 92.4. The distribution of scores across jargon categories is shown in Table 2. Around one-fifth of summaries had a score of 95 or above, suggesting that they would be suitable for a general audience.

Table 2. Distribution of summaries by jargon score.

Jargon score	Values, n (%)
95-100 (least jargon)	269 (21.7)
90-95	592 (47.7)
85-90	300 (24.2)
85 or lower (most jargon)	80 (5.4)

The pairwise correlation between readability scores and jargon scores was 0.249 (P<.001), which suggests that when one score is higher, the other is also likely to be higher but that the relationship is moderate. Sixty-six summaries (5.3%, 66/1241) were in both the "easier to read" and "least jargon" categories.

Table 3 shows, for each of the study characteristics, the number of summaries in each category (where relevant) and the mean readability and jargon scores for that category. We found statistically significant variations in estimated readability and jargon scores between categories for each study characteristic.

XSL•FO
RenderX

Table 3. Distribution of summaries, mean readability scores, and mean jargon scores in relation to each study characteristic.

	Number in each category, n (%)	Mean readability score (higher=easier to read; scores of ≥60 are "plain English," 50-60 are suitable for people educated to high-school level, 30-50 for undergraduate level, and 30 and below for postgraduate level) (95% CIs)	Mean jargon score (higher=less jargon; scores of ≥95 are suitable for a general audience) (95% CIs)
Research topic[a]			
Cancer and neoplasms	164 (13.2)	39.9 (37.9-41.9)	91.2 (90.6-91.9)
Cardiovascular	165 (13.3)	41.6 (39.8-43.4)	91.6 (91.0-92.1)
Generic health relevance	271 (21.8)	37.7 (36.3-39.1)	94.2 (93.9-94.4)
Infection	93 (7.5)	38.4 (36.2-40.6)	90.3 (89.5-91.1)
Mental health	211 (17.0)	38.2 (36.7-39.7)	93.1 (92.7-93.5)
Metabolic and endocrine	78 (6.3)	42.6 (39.9-45.2)	92.4 (91.6-93.2)
Musculoskeletal	57 (4.6)	37.8 (33.3-40.3)	91.2 (90.2-92.2)
Neurological	81 (6.5)	38.0 (35.3-40.7)	92.4 (91.6-93.2)
Oral and gastrointestinal	106 (8.5)	42.1 (39.9-44.2)	91.7 (91.0-92.5)
Reproductive health and childbirth	96 (7.7)	41.6 (39.6-43.7)	91.2 (90.5-91.9)
Respiratory	56 (4.4)	41.3 (38.2-44.4)	91.0 (90.1-92.0)
Stroke	110 (8.9)	41.7 (39.5-43.9)	92.9 (92.3-93.5)
Funding program[b]			
Efficacy and Mechanism Evaluation	81 (6.5)	38.8 (36.4-41.2)	88.4 (87.7-89.1)
Health Technology Assessment	630 (50.8)	39.4 (38.5-40.3)	90.3 (90.0-90.6)
Health and Social Care Delivery Research	327 (26.4)	37.3 (36.1-38.6)	94.2 (93.9-94.5)
Programme Grants for Applied Research	96 (7.7)	37.5 (35.3-39.6)	92.8 (92.2-93.4)
Public Health Research	107 (8.6)	43.4 (41.3-45.6)	94.0 (93.6-94.5)
Publication year[b]			
2014	91 (7.3)	37.6 (35.4-39.8)	92.5 (91.8-93.2)
2015	170 (13.7)	38.7 (36.8-40.6)	91.5 (90.9-92.0)
2016	168 (13.5)	37.5 (35.5-39.5)	91.2 (90.5-91.9)
2017	147 (11.8)	39.2 (37.6-40.9)	91.8 (91.2-92.4)
2018	142 (11.4)	41.1 (39.4-42.8)	91.9 (91.4-92.5)
2019	153 (12.3)	41.3 (39.6-43.0)	91.7 (91.0-92.3)
2020	154 (12.4)	38.7 (36.7-40.7)	92.1 (91.4-92.7)
2021	145 (11.6)	38.7 (36.8-40.5)	91.7 (91.0-92.4)
2022	71 (5.7)	37.1 (34.2-39.9)	91.5 (90.6-92.5)
Project size			
Smallest 20%		36.9 (35.5-38.4)	91.4 (90.9-91.9)
2		38.1 (36.6-39.7)	92.7 (92.2-93.1)
3		39.5 (38.1-40.9)	93.1 (92.7-93.6)
4		40.2 (38.8-41.6)	90.5 (89.9-91.0)
Largest 20%		40.3 (38.9-41.7)	91.0 (90.5-91.4)
Length of summary			
Shortest 20%		38.6 (37.0-40.1)	91.6 (91.1-92.2)
2		39.2 (37.7-40.7)	92 (91.5-92.4)
3		39.3 (37.9-40.6)	91.2 (90.7-91.7)

XSL•FO
RenderX

	Number in each category, n (%)	Mean readability score (higher=easier to read; scores of ≥60 are "plain English," 50-60 are suitable for people educated to high-school level, 30-50 for undergraduate level, and 30 and below for postgraduate level) (95% CIs)	Mean jargon score (higher=less jargon; scores of ≥95 are suitable for a general audience) (95% CIs)
4		40.9 (39.5-42.2)	92.8 (92.3-93.2)
Longest 20%		37.2 (35.7-38.6)	91.1 (90.5-91.6)
Scores in original proposal[c]			
Lowest 20% of original scores		35.3 (33.8-36.8)	88.0 (87.5-88.4)
2		36.8 (35.3-38.4)	90.5 (90.1-91.0)
3		39.4 (38.0-40.7)	92.2 (91.8-92.6)
4		39.7 (38.4-40.9)	93.3 (92.9-93.6)
Highest 20% of original scores		43.9 (42.5-45.3)	94.7 (94.4-94.9)

[a]Some summaries were associated with more than 1 area of research, so the total percentage does not add up to 100.

[b]Percentages may not add up to 100 because of rounding.

[c]Readability scores are categorized by the readability score of the original funding proposal; jargon scores are categorized by the jargon score of the original funding proposal.

For the research topic, the highest estimated readability score (most readable) was in the "Metabolic and endocrine" category, and the lowest score (least readable) was for "Generic health relevance." However, the "Generic health relevance" category was associated with the highest estimated jargon score (least jargon), and the lowest score (most jargon) was for "Infection."

For the funding program, the highest estimated readability score was associated with the Public Health Research program, and the lowest was in Health and Social Care Delivery Research. In contrast, the Health and Social Care Delivery Research had the highest estimated jargon score (least jargon). The lowest jargon score (most jargon) was associated with the Efficacy and Mechanism Evaluation program.

For project size, mean readability scores but not jargon scores rose as projects got larger. The highest estimated readability score was associated with the largest projects (top 20% of funding), and the lowest readability was associated with the smallest (bottom 20%). There was no clear pattern of variation in jargon scores.

For length of summary, longer summaries were associated with better readability and less jargon, but only up to a point. Scores appeared to rise across the first 80% of summaries when ranked by length but then dipped so that the lowest scores were in the longest 20% of summaries.

For publication year, the lowest readability score was for 2022, but data for that year were incomplete at the time we collected our data. Readability and jargon scores varied by year, but there did not appear to be any trend in either score over time.

Our estimates of scores in summaries in relation to scores in original proposals rose steadily across the categories from lowest to highest proposal scores for both readability and jargon. For both readability and jargon, the highest estimated scores were for PLSs that had the highest scores in the proposals, and the lowest estimated scores were for PLSs that had the lowest scores in the proposals.

The proportion of summaries that have readability scores over 50 ("fairly difficult to read" or better) and jargon scores >95 (low level of jargon) are shown in Figure 1 (by research topic), Figure 2 (by funding program), Figure 3 (by project size), and Figure 4 (by scores in original proposals).

XSL•FO
RenderX

Figure 1. Percentage of summaries with readability scores >50 ("fairly difficult to read" or better) and with jargon scores >95 (low level of jargon), by research topic, with 95% CIs. PLSs: plain language summaries.

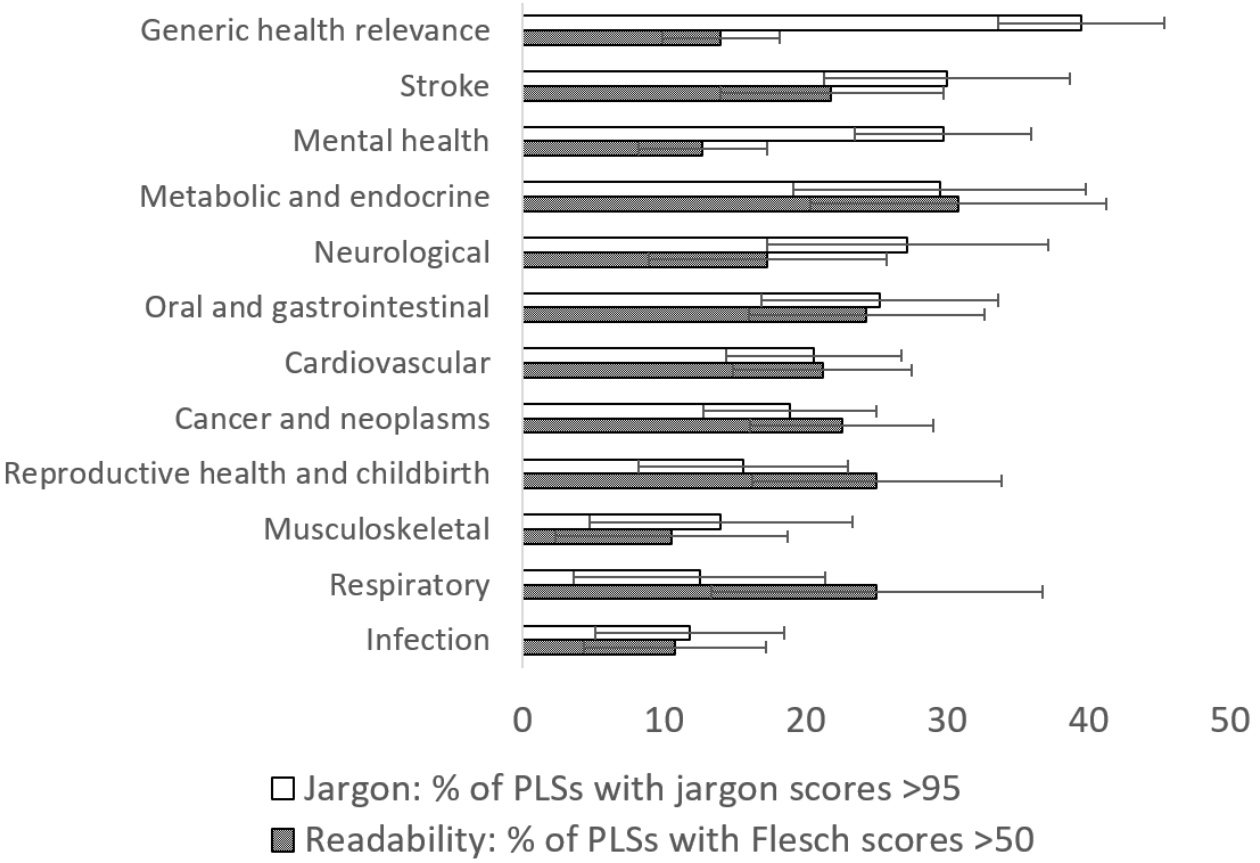

Figure 2. Percentage of summaries with readability scores >50 ("fairly difficult to read" or better) and with jargon scores >95 (low level of jargon), by National Institute for Health and Care Research funding program, with 95% CIs. PLSs: plain language summaries.

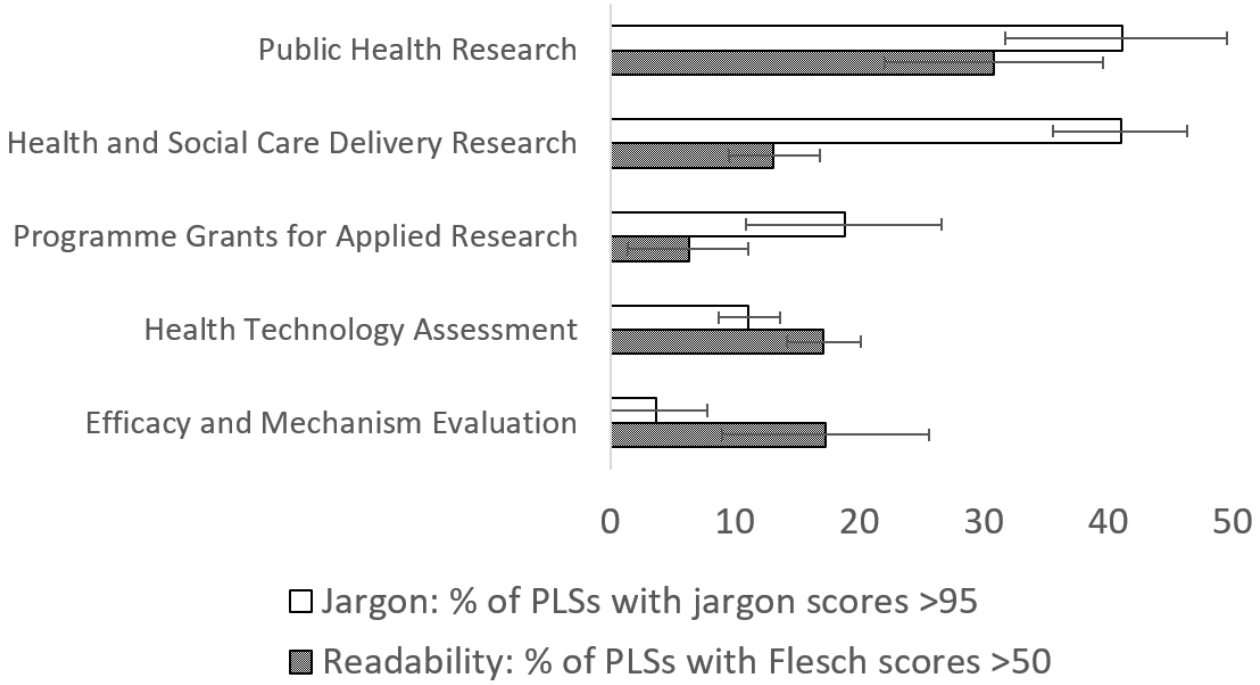

XSL•FO
RenderX

Figure 3. Percentage of summaries with readability scores >50 ("fairly difficult to read" or better) and with jargon scores >95 (low level of jargon), by project size, with 95% CIs. PLSs: plain language summaries.

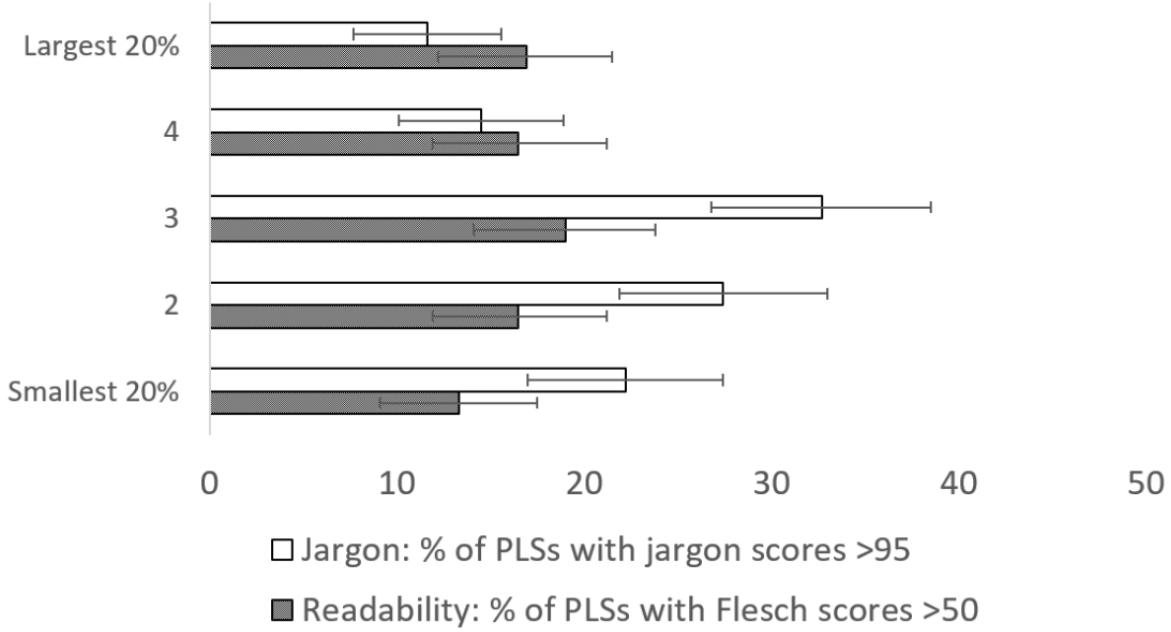

Figure 4. Percentage of summaries with readability scores >50 ("fairly difficult to read" or better) and with jargon scores >95 (low level of jargon), by readability scores or jargon scores (as appropriate: readability scores are categorized in relation to the readability score of the original funding proposal, and jargon scores are categorized in relation to the corresponding original jargon score) in original proposals, with 95% CIs. PLSs: plain language summaries.

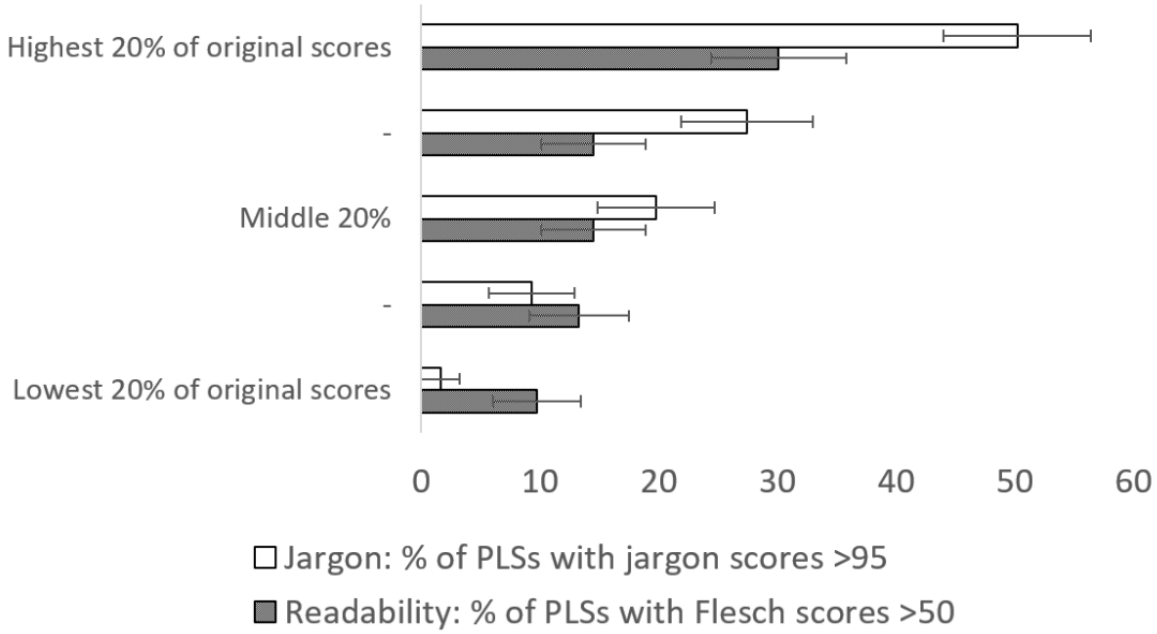

Finally, Table 4 shows the incidence rate ratios of having each of these "better" scores when all study characteristics are included in the same model. This is useful because it allows us to assess whether the differences we have observed are still there when the other differences between summaries are accounted for. It could be, for example, that summaries in the Public Health Research program were more likely to be in the easier-to-read category, compared with Programme Grants for Applied Research, because of differences in project size or research topic. The results of the regression suggest, however, that even when other differences (eg, size, topic, year) are controlled for, Public Health Research summaries are approximately 5 times as likely as Programme Grant summaries to be more readable. Similarly, we can say that projects for which the original proposal had little jargon are markedly more than 10 times as likely to have little jargon in their final reports when compared with those with lots of jargon in the original proposal, even when differences in size, topic, and so on are accounted for.

XSL•FO
RenderX

Table 4. Incidence rate ratios and 95% CIs from regression analyses including all study characteristics for each outcome of interest: readability scores >50 ("fairly difficult to read" or better) and jargon scores >95 (low level of jargon).

	Readability scores ≥50 ("fairly difficult to read" or better), incidence rate ratio (95% CI)	Jargon scores ≥95 (suitable for a general audience), incidence rate ratio (95% CI)
Research topic		
Cancer and neoplasms	1.12 (0.70-1.79)	0.84 (0.45-1.56)
Cardiovascular	0.95 (0.57-1.58)	0.99 (0.55-1.78)
Generic health relevance	0.90 (0.57-1.41)	1.86 (1.32-2.60)
Infection	0.55 (0.31-0.97)	1.25 (0.75-2.09)
Mental health	0.73 (0.47-1.14)	1.71 (1.26-2.32)
Metabolic and endocrine	1.72 (1.08-2.76)	1.26 (0.80-1.98)
Musculoskeletal	0.68 (0.31-1.49)	1.17 (0.63-2.15)
Neurological	1.05 (0.63-1.75)	1.74 (1.21-2.50)
Oral and gastrointestinal	1.10 (0.63-1.93)	1.10 (0.66-1.86)
Reproductive health and childbirth	1.45 (0.95-2.20)	1.32 (0.84-2.09)
Respiratory	1.72 (1.03-2.87)	0.90 (0.44-1.86)
Stroke	0.50 (0.23-1.09)	1.69 (0.98-2.92)
Funding Program		
Efficacy and Mechanism Evaluation	2.59 (1.01-6.61)	1
Health Technology Assessment	3.00 (1.31-6.87)	2.25 (0.74-6.85)
Health and Social Care Delivery Research	2.56 (1.05-6.25)	4.08 (1.33-12.54)
Programme Grants for Applied Research	1	3.27 (1.00-10.66)
Public Health Research	5.35 (2.13-13.4)	3.87 (1.25-11.96)
Publication year		
2014	0.78 (0.4-1.54)	0.62 (0.38-1.00)
2015	0.95 (0.56-1.62)	0.72 (0.47-1.11)
2016	1.00 (0.60-1.67)	0.89 (0.58-1.37)
2017	0.82 (0.47-1.44)	1.05 (0.70-1.58)
2018	1.01 (0.64-1.61)	0.81 (0.54-1.22)
2019	1	1
2020	1.25 (0.82-1.89)	1.22 (0.88-1.70)
2021	0.82 (0.50-1.32)	1.49 (1.05-2.12)
2022	0.65 (0.34-1.25)	1.03 (0.67-1.60)
Project size		
Smallest 20%	1	1
2	1.19 (0.77-1.83)	0.89 (0.67-1.20)
3	1.36 (0.89-2.06)	1.01 (0.76-1.35)
4	1.18 (0.76-1.83)	0.69 (0.49-0.99)
Largest 20%	1.64 (1.06-2.55)	0.59 (0.37-0.97)
Length of summary		
Shortest 20%	1	1
2	1.07 (0.71-1.61)	1.09 (0.81-1.47)
3	0.81 (0.51-1.27)	0.67 (0.47-0.97)
4	1.07 (0.67-1.69)	0.93 (0.65-1.32)
Longest 20%	0.69 (0.42-1.13)	0.74 (0.51-1.10)

XSL•FO
RenderX

	Readability scores ≥50 ("fairly difficult to read" or better), incidence rate ratio (95% CI)	Jargon scores ≥95 (suitable for a general audience), incidence rate ratio (95% CI)
Scores in original proposal[a]		
Lowest 20% of original scores	1	1
2	1.39 (0.85-2.26)	4.53 (1.62-12.68)
3	1.40 (0.87-2.26)	7.26 (2.66-19.79)
4	1.48 (0.91-2.40)	8.74 (3.23-23.63)
Highest 20% of original scores	2.88 (1.86-4.45)	13.87 (5.17-37.2)

[a]We analyzed readability score outcomes in relation to the readability score of the original funding proposal and jargon score outcomes in relation to the jargon score of the original funding proposal.

Discussion

Main Findings

Our findings suggest that the PLSs published in the NIHR's Journals Library are often difficult to read and likely inaccessible to a general audience. Despite the NIHR's advice to avoid jargon and complicated words and to keep sentences short, many published summaries had lots of jargon and poor readability. We analyzed more than 1200 summaries and found none with readability scores suggesting that they would be accessible to people with the average UK reading age of 9 years.

Readability and jargon scores varied significantly in relation to research topic (where "Metabolic and endocrine" projects did best), funding program (where projects in the Public Health Research Programme did best), and, most noticeably, in relation to how readable the original funding proposal was. The relationship between original funding proposal summaries and final report summaries is notable because it suggests that some authors are consistently better (or worse) at writing accessible summaries.

Readability scores and jargon scores were correlated but did not always coincide. For instance, summaries in the "Generic health relevance" category used relatively few jargon words but were not very readable. The opposite was true of larger projects, which, compared with small projects, had more readable summaries and less jargon. The Efficacy and Mechanism Evaluation funding program was also associated with summaries that were moderately readable yet had high levels of jargon. These differences suggest that summaries may be accessible or easy to read in some ways but not in others and that relying on a single measure to assess a text may miss important aspects of readability. We found no pattern over time: looked at year-on-year, there are differences but no upward or downward trend in either readability or jargon scores.

Comparisons With Other Studies

The mean FRE score we found was higher than the mean score of 23.6 reported for PLSs in 2 psychology journals [43] and the mean of 21 found in a sample taken from the Physiotherapy Evidence Database [44]. A study of research on cystic fibrosis found that the FRE score of PLSs in journals was 43.3 and in Cochrane reviews was 46.3 [45]. An analysis of medicine information sheets produced by the professional associations of rheumatologists in 3 countries found that the average FRE score for those from Australia was 50.8, from the United Kingdom was 48.5, and from Canada was 66.1 [46]; in patient information leaflets produced by the British Association of Dermatologists, the mean score was 52.2 [47]. An analysis of Cochrane PLSs, which used a different readability formula, concluded that most would be difficult to read for someone with no medical education [48] and is in line with a previous analysis of Cochrane PLSs that found that they were very heterogeneous and often failed to adhere to standards [49]. Comparing scores across different samples of texts is not straightforward because of differences in the intended audiences for each. What is clear is that the level of readability we found is markedly worse than that recommended for texts in plain language and suitable for a general audience, which is a score of 60 or above (Table 1).

We are aware of only 1 previous use of a jargon calculator to assess PLSs by the developers of the calculator we used [32]. They compared levels of jargon in academic abstracts and lay summaries in 2 journals, *PLOS Computational Biology* and *PLOS Genetics*. Although they found that the PLSs contained less jargon than the abstracts, they—like us—found that jargon use in the summaries was significantly higher than the recommended levels. We used a cutoff of 5% to identify summaries with low levels of jargon and identified that only 21.7% (269/1241) of papers met this criterion; of the summaries in our sample, only 1.0% (13/1241) met the more stringent 2% cutoff proposed by Rakedzon and colleagues [32].

Strengths and Weaknesses of This Study

Previous studies of PLSs have focused on specific research areas (such as physiotherapy) or methods (such as reviews). Our broader approach has enabled us to address a part of what Uphold and colleagues [18] described as a critical gap in the dissemination and implementation literature, the analysis of how researcher characteristics and environmental determinants influence attempts to disseminate research findings. We also looked at both readability and jargon, whereas previous studies have focused on one or the other.

FRE scores are widely used [27], are recommended by NIHR and other funders as a way of assessing the readability of PLSs [25], and—as we have used them here—provide a way of looking at large numbers of texts and identifying trends and tendencies in readability. We recognize, all the same, that readability indices are an imperfect way of assessing texts. They can be misleading and have little to do with how easy a text is to understand [50-52], and scores may be inconsistent when

XSL•FO
RenderX

tested across different pieces of software because of formatting or other differences [27]. Readability indices also capture only 1 aspect of how science is communicated to the public. There are more sophisticated ways of understanding [53,54], conducting [55,56], and assessing [57-59] science communication. Our approach focused on summaries written by research teams and could be extended by looking at the perceptions and responses of readers and evaluating the impact of summaries [60-62]. We had no information on the extent to which teams responded to NIHR's "strong encouragement" to involve a nonacademic member of the public in writing the PLS [25], so we cannot comment on whether this alters their content and presentation.

The way we have assessed jargon focuses on single words and ignores the use of phrases that might otherwise count as jargon. Multimedia Appendix 1 contains some examples of this. For example, "confidence" and "interval" are both classed as mid-frequency words in our jargon calculator, but the phrase "confidence interval," often used in reporting statistical estimates in health studies (as in this paper), would probably be considered jargon. The same applies to longer jargon terms such as "incremental cost-effectiveness ratio," which has a specific technical meaning but which the jargon calculator does not pick up on. We have not accounted for other aspects of summaries that can affect how readable they are, such as the use of numbers and statistics.

The summaries we looked at were all funded by a single large funder (NIHR) and relate to research done in a single country (the United Kingdom). Over the past 3 decades, the United Kingdom has more swiftly embedded public involvement and engagement in health research than most other countries [63,64]. While we may appear critical of NIHR, this study has been possible only because NIHR requires researchers to write summaries and then makes these publicly accessible. NIHR sponsors the UK Standards for Public Involvement, intended as a "description of what good public involvement looks like," and 1 of the 6 standards is "Communication—use plain language for well-timed and relevant communications, as part of involvement plans and activities" [65]. The NIHR also says that it is "the world's first health research funder to publish comprehensive accounts of its funded research within its own publicly and permanently available journals" [26].

The PLSs included in National Journals Library publications are central to these 2 commitments—to make publicly available details of the work it funds and to communicate in plain language—and are "in keeping with the NIHR Journals Library's commitment to accessibility" [26]. They also relate to a substantial investment: the cost of the research represented in the reports at which we looked is not easy to calculate, but NIHR expenditure on these funding streams in financial year 2021/22 was £206.3 million (approximately €241 million/US $268 million), so it seems reasonable to assume that the cost over the years covered here exceeded £1 billion (approximately €1.17 billion/US $1.30 billion). We might expect to see in the work of the NIHR a well-developed set of processes and mechanisms by which to communicate the results of research to the public, and for this reason, we consider the NIHR's flagship publication stream, the Journals Library, to be an apt focus for our enquiry here.

The extent to which the assessment of readability can be regarded as an indication of comprehension is not included in this study, as there is no existing measure that captures this relationship. Assessing comprehensibility usually involves asking people to read 1 or more texts and then measuring their understanding (eg, as Koops et al [66] did), which would be challenging to do with a large sample of texts such as those in the NIHR Journals Library. In the absence of measurable comprehensibility, assumptions on the accessibility of PLSs deserve scrutiny and a willingness to reconsider the methods of their production.

Unanswered Questions and Future Research

We would like this study to contribute to a debate on 2 connected questions: What and who are PLSs for?

From a funder's perspective, PLSs are a relatively cheap and easy way of disseminating research findings. The teams doing the research are responsible for producing a PLS, and this has strengths (the information "straight from the horse's (or scientist's) mouth" [67] is presumably less likely to be incorrect or imprecise) as well as weaknesses (researchers often lack training in science communication). Scientific writing can be exclusive: "The language of science, though forward-looking in its origins, has become increasingly anti-democratic… [it] sets apart those who understand it and shields them from those [who] do not" [68] but fluency in this language is necessary for scientists to do their work and have it accepted by other scientists [69]. PLSs could be an important aspect of a more "engaged" university sector [64], but the intended role and audience for PLSs are unclear, and this limits their potential value.

Communicating in a certain way means communicating to a group of people who can understand that type of communication. Attempts to communicate to "the general public" [25] must address multiple audiences [70], but the members of the public being addressed include people who differ widely in their interest, knowledge, and trust in science [71]. Ahmed [72] argues, drawing on the work of Warner [73], that addressing a public generates a public that can be addressed. Saying something in Swahili implies that you are speaking to people who can understand Swahili, and writing something using lots of scientific jargon and complicated sentences implies that you intend it for an audience who can understand such language. The dissemination of research findings has been challenged as a 1-way form of communication that falls short of the principles and expectations of public engagement [74], can be seen as representing public relations rather than science communication [75,76], and contributes to the exclusion of people from low-income, minority ethnic groups [77]. If we regard the production of PLSs as an aspect of Open Science, with its commitments to social engagement in the conduct and outcomes of research [5], we also need to be aware of what is being made visible and invisible in the process [78] and of who is being included and excluded.

XSL•FO
RenderX

Evidence on what is most effective in PLSs of scientific research, in terms of both form and content, is emerging [16,67,79], but there is still much for us to learn about what works, for whom, and why. The value of training to improve science communication is unclear [80], and attempts to improve PLSs have met with mixed results. For example, Kirkpatrick and colleagues [81] tested 2 approaches: having authors rewrite PLSs using new guidance and having an independent medical writer edit the PLS. In each case, a group of nonspecialists rated the revised versions as easier to read but not easier to understand. A service designed to improve recruitment to studies by having trained patients and carers review research documents was successful in reducing the amount of jargon but not in improving readability [82].

Creative approaches to communicating research findings have the potential to enable 2-way communication and flatten hierarchies between scientists and nonscientists [83]. Again, when putting this to the test, results have been mixed: one recent study found that PLSs of published research were more effective than scientific abstracts or graphical abstracts in terms of comprehension and understanding [84], whereas another study concluded that graphical summaries were one of the most preferred formats [85]. A recent review of instructions for authors on writing PLSs found a lot of inconsistency across journals and suggested that consistent instructions could be developed with members of the public [17]. The creation of common standards for summaries has also been proposed as part of the OpenPharma project [13,14]. Expert consensus conference methods have been used to produce recommendations on maximizing the accessibility of study patient-information leaflets and informed-consent forms [86], and a similar approach could potentially be applied to the preparation of PLSs.

Another approach from which we might learn is citizen science. Communication of all aspects of research is fundamental to citizen science, which has been described as one of the most dramatic developments in science communication in decades [87] It emphasizes multidirectional and ongoing communication [88] and recognizes storytelling and visualization as central to this [89]. At least in aspiration, citizen science has the potential to improve and transform science communication at the same time as it empowers and informs citizen scientists [90]. Other approaches to dissemination have emphasized coproduction [91] and community engagement [92]. As Knowles and colleagues noted, “finding mutually acceptable and valuable ways to express findings is yet another area requiring open discussion and negotiation” [93].

Conclusions

We found that the sample of PLSs that we examined had low readability and contained lots of jargon. Although these things differed in relation to study characteristics, such as topic and size, none of the PLSs had readability scores in line with the average reading age of the UK public. The aims of, and audiences for, these PLSs are unclear, and their place in science communication and public engagement requires further consideration. It is uncertain whether these summaries improve public access to research.

Acknowledgments

This report is independent research supported by the UK National Institute for Health and Care Research Applied Research Collaboration South West Peninsula (grant reference number: NIHR200167). The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the UK Department of Health and Social Care. We did not use generative artificial intelligence in any portion of the writing of this manuscript.

Data Availability

The data we used came from the texts and metadata of reports published in the NIHR’s Journals Library. This is an open-access resource, and all the data we used are publicly available on the web via the Journals Library website: https://www.journalslibrary.nihr.ac.uk/journals/

Authors' Contributions

IAL and AK contributed to conceptualization and methodology. IAL contributed to statistical analysis and writing the original draft. IAL, AK, KB, KS, LA, JD, and KL contributed to the interpretation of the data, critical revision of the manuscript for important intellectual content, and writing—review and editing, and all authors agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exemplar plain language summaries with high and low readability and high and low jargon scores.
[DOCX File , 18 KB - jmir_v27i1e50862_app1.docx]

References

XSL•FO
RenderX

1. Beresford P, Russo J. Patient and Public Involvement In Research. In: Nolte E, Merkur S, Anell A, editors. Achieving Person-Centred Health Systems. Cambridge: Cambridge University Press; 2020:145-172.
2. Liabo K, Boddy K, Bortoli S, Irvine J, Boult H, Fredlund M, et al. Public involvement in health research: what does 'good' look like in practice? Res Involv Engagem 2020;6:11 [FREE Full text] [doi: 10.1186/s40900-020-0183-x] [Medline: 32266085]
3. Duncan S, Oliver S. Editorial: motivations for engagement. Research for All 2017;1(2):229-233. [doi: 10.18546/rfa.01.2.01]
4. Solomon MZ, Gusmano MK, Maschke KJ. The ethical imperative and moral challenges of engaging patients and the public with evidence. Health Aff (Millwood) 2016;35(4):583-589. [doi: 10.1377/hlthaff.2015.1392] [Medline: 27044955]
5. Leonelli S. Data-Centric Biology: A Philosophical Study. Chicago: University of Chicago Press; 2016.
6. Gradinger F, Britten N, Wyatt K, Froggatt K, Gibson A, Jacoby A, et al. Values associated with public involvement in health and social care research: a narrative review. Health Expect 2015;18(5):661-675 [FREE Full text] [doi: 10.1111/hex.12158] [Medline: 24325553]
7. Martin GP. 'Ordinary people only': knowledge, representativeness, and the publics of public participation in healthcare. Sociol Health Illn 2008;30(1):35-54 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01027.x] [Medline: 18254832]
8. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on Clinical Trials on Medicinal Products for Human use, and Repealing Directive 2001/20/EC. 2014. URL: https://health.ec.europa.eu/system/files/2016-11/reg_2014_536_en_0.pdf [accessed 2024-09-26]
9. Pitcher N, Mitchell D, Hughes C. Guidance for writing a Cochrane plain language summary. Cochrane Handbook for Systematic Reviews of Interventions. 2022. URL: https://training.cochrane.org/handbook/current/chapter-iii-s2-supplementary-material [accessed 2024-09-26]
10. Flory J, Wendler D, Emanuel E. Empirical issues in informed consent for research. In: Emanuel E, editor. The Oxford Textbook of Clinical Research Ethics. London: Oxford University Press; 2008:645-660.
11. Mandava A, Pace C, Campbell B, Emanuel E, Grady C. The quality of informed consent: mapping the landscape. a review of empirical data from developing and developed countries. J Med Ethics 2012;38(6):356-365 [FREE Full text] [doi: 10.1136/medethics-2011-100178] [Medline: 22313664]
12. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent in cancer clinical trials: a cross-sectional survey. The Lancet 2001;358(9295):1772-1777. [doi: 10.1016/s0140-6736(01)06805-2]
13. Rosenberg A. Working toward standards for plain language summaries. Sci Editor 2022;45(2):46-50 [FREE Full text] [doi: 10.36591/se-d-4502-46]
14. Rosenberg A, Baróniková S, Feighery L, Gattrell W, Olsen RE, Watson A, et al. Open pharma recommendations for plain language summaries of peer-reviewed medical journal publications. Curr Med Res Opin 2021;37(11):2015 [FREE Full text] [doi: 10.1080/03007995.2021.1971185] [Medline: 34511020]
15. Lobban D, Gardner J, Matheis R. Plain language summaries of publications of company-sponsored medical research: what key questions do we need to address? Current Medical Research and Opinion 2021;38(2):189-200 [FREE Full text] [doi: 10.1080/03007995.2021.1997221]
16. Stoll M, Kerwer M, Lieb K, Chasiotis A. Plain language summaries: a systematic review of theory, guidelines and empirical research. PLoS One 2022;17(6):e0268789 [FREE Full text] [doi: 10.1371/journal.pone.0268789] [Medline: 35666746]
17. Gainey KM, Smith J, McCaffery KJ, Clifford S, Muscat DM. What author instructions do health journals provide for writing plain language summaries? A scoping review. Patient 2023;16(1):31-42 [FREE Full text] [doi: 10.1007/s40271-022-00606-7] [Medline: 36301440]
18. Uphold HS, Drahota A, Bustos TE, Crawford MK, Buchalski Z. “There’s no money in community dissemination”: a mixed methods analysis of researcher dissemination-as-usual. J Clin Trans Sci 2022;6(1):e105. [doi: 10.1017/cts.2022.437]
19. Brownson RC, Fielding JE, Green Lawrence W. Building capacity for evidence-based public health: reconciling the pulls of practice and the push of research. Annu Rev Public Health 2018;39:27-53 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-014746] [Medline: 29166243]
20. Green Lawrence W, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof 2006;29(1):126-153. [doi: 10.1177/0163278705284445] [Medline: 16510882]
21. Tabak RG, Stamatakis KA, Jacobs JA, Brownson RC. What predicts dissemination efforts among public health researchers in the United States? Public Health Rep 2014;129(4):361-368 [FREE Full text] [doi: 10.1177/003335491412900411] [Medline: 24982539]
22. McNeal DM, Glasgow RE, Brownson RC, Matlock DD, Peterson PN, Daugherty SL, et al. Perspectives of scientists on disseminating research findings to non-research audiences. J Clin Transl Sci 2020;5(1):e61 [FREE Full text] [doi: 10.1017/cts.2020.563] [Medline: 33948281]
23. Knoepke CE, Ingle MP, Matlock DD, Brownson RC, Glasgow RE. PLoS One 2019;14(11):e0216971 [FREE Full text] [doi: 10.1371/journal.pone.0216971] [Medline: 31721784]
24. Davies SC, Walley T, Smye S, Cotterill L, Whitty CJ. The NIHR at 10: transforming clinical research. Clin Med (Lond) 2016;16(6):501-502 [FREE Full text] [doi: 10.7861/clinmedicine.16-6-501] [Medline: 27927810]

XSL•FO
RenderX

25. Plain English summaries. National Institute for Health and Care Research. 2021. URL: https://www.nihr.ac.uk/documents/plain-english-summaries/27363 [accessed 2024-09-26]
26. NIHR Journals Library. National Institute of Health and Care Research. URL: https://www.journalslibrary.nihr.ac.uk/#/ [accessed 2024-10-09]
27. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503-516. [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
28. Baram-Tsabari A, Wolfson O, Yosef R, Chapnik N, Brill A, Segev E. Jargon use in Public Understanding of Science papers over three decades. Public Underst Sci 2020;29(6):644-654 [FREE Full text] [doi: 10.1177/0963662520940501] [Medline: 32815790]
29. Plavén-Sigray P, Matheson GJ, Schiffler BC, Thompson WH. The readability of scientific texts is decreasing over time. eLife 2017;6:e27725 [FREE Full text] [doi: 10.7554/elife.27725]
30. DiMascio C. Py-Readability-Metrics [Source code]. GitHub. 2019. URL: https://github.com/cdimascio/py-readability-metrics [accessed 2019-12-09]
31. BBC. URL: https://www.bbc.com/ [accessed 2024-09-26]
32. Rakedzon T, Segev E, Chapnik N, Yosef R, Baram-Tsabari A. Automatic jargon identifier for scientists engaging with the public and science communication educators. PLoS One 2017;12(8):e0181742 [FREE Full text] [doi: 10.1371/journal.pone.0181742] [Medline: 28792945]
33. Jargon Project. De-Jargonizer. 2023. URL: https://scienceandpublic.com/ [accessed 2020-01-31]
34. NoamAndRoy. JargonProject [Source code]. GitHub. 2017. URL: https://github.com/NoamAndRoy/JargonProject [accessed 2024-09-26]
35. HRCS Online - Home. 2023. URL: https://hrcsonline.net/ [accessed 2024-09-26]
36. Funding Programmes. National Institute for Health and Care Research. 2024. URL: https://www.nihr.ac.uk/explore-nihr/funding-programmes/ [accessed 2024-09-26]
37. Ennis L, Wykes T. Sense and readability: participant information sheets for research studies. Br J Psychiatry 2016;208(2):189-194 [FREE Full text] [doi: 10.1192/bjp.bp.114.156687] [Medline: 26382948]
38. Sharp SM. Consent documents for oncology trials: does anybody read these things? Am J Clin Oncol 2004;27(6):570-575. [doi: 10.1097/01.coc.0000135925.83221.b3] [Medline: 15577434]
39. Brierley G, Richardson R, Torgerson DJ. Using short information leaflets as recruitment tools did not improve recruitment: a randomized controlled trial. J Clin Epidemiol 2012;65(2):147-154. [doi: 10.1016/j.jclinepi.2011.06.005] [Medline: 21889304]
40. National Journals Library - Plain language summary. National Institute for Health and Social Care Research. URL: https://www.journalslibrary.nihr.ac.uk/information-for-authors/manuscript-preparation/report-sections/plain-language-summary.htm [accessed 2024-09-26]
41. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159(7):702-706. [doi: 10.1093/aje/kwh090] [Medline: 15033648]
42. The 2011 Skills for Life survey: a survey of literacy, numeracy and ICT levels in England. London, UK; 2012. URL: https://www.gov.uk/government/publications/2011-skills-for-life-survey [accessed 2024-09-26]
43. Stricker J, Chasiotis A, Kerwer M, Günther A. Scientific abstracts and plain language summaries in psychology: a comparison based on readability indices. PLoS One 2020;15(4):e0231160 [FREE Full text] [doi: 10.1371/journal.pone.0231160] [Medline: 32240246]
44. Carvalho FA, Elkins MR, Franco MR, Pinto RZ. Are plain-language summaries included in published reports of evidence about physiotherapy interventions? Analysis of 4421 randomised trials, systematic reviews and guidelines on the Physiotherapy Evidence Database (PEDro). Physiotherapy 2019;105(3):354-361. [doi: 10.1016/j.physio.2018.11.003] [Medline: 30876718]
45. Anderson HL, Moore JE, Millar BC. Comparison of the readability of lay summaries and scientific abstracts published in CF research news and the journal of cystic fibrosis: recommendations for writing lay summaries. J Cyst Fibros 2022;21(1):e11-e14 [FREE Full text] [doi: 10.1016/j.jcf.2021.09.009] [Medline: 34600843]
46. Oliffe M, Thompson E, Johnston J, Freeman D, Bagga H, Wong PKK. Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional health literacy study. BMJ Open 2019;9(2):e024582 [FREE Full text] [doi: 10.1136/bmjopen-2018-024582] [Medline: 30813117]
47. Hunt WTN, Sofela J, Mohd Mustapa MF, British Association of Dermatologists' Clinical Standards Unit. Readability assessment of the British Association of Dermatologists' patient information leaflets. Clin Exp Dermatol 2022;47(4):684-691. [doi: 10.1111/ced.15012] [Medline: 34854104]
48. Banić A, Fidahić M, Šuto J, Roje R, Vuka I, Puljak L, et al. Conclusiveness, linguistic characteristics and readability of Cochrane plain language summaries of intervention reviews: a cross-sectional study. BMC Med Res Methodol 2022;22(1):240 [FREE Full text] [doi: 10.1186/s12874-022-01721-7] [Medline: 36088293]

XSL•FO
RenderX

49. Jelicic Kadic A, Fidahic M, Vujcic M, Saric F, Propadalo I, Marelja I, et al. Cochrane plain language summaries are highly heterogeneous with low adherence to the standards. BMC Med Res Methodol 2016;16(1):61. [doi: 10.1186/s12874-016-0162-y] [Medline: 27216616]
50. Colter A. Assessing the usability of credit card disclosures. Clarity - The Journal of the International Association Promoting Plain Legal Language 2009;62:46-52.
51. Redish J. Readability formulas have even more limitations than Klare discusses. ACM J. Comput. Doc 2000;24(3):132-137. [doi: 10.1145/344599.344637]
52. Zheng J, Yu H. Readability formulas and user perceptions of electronic health records difficulty: a corpus study. J Med Internet Res 2017;19(3):e59 [FREE Full text] [doi: 10.2196/jmir.6962] [Medline: 28254738]
53. Baram-Tsabari A, Lewenstein BV. An instrument for assessing scientists' written skills in public communication of science. Science Communication 2012;35(1):56-85. [doi: 10.1177/1075547012440634]
54. Longnecker N. An integrated model of science communication — more than providing evidence. JCOM 2016;15(05):Y01 [FREE Full text] [doi: 10.22323/2.15050401]
55. Shonkoff JP, Bales SN. Science does not speak for itself: translating child development research for the public and its policymakers. Child Dev 2011;82(1):17-32. [doi: 10.1111/j.1467-8624.2010.01538.x] [Medline: 21291426]
56. Bales SN, Gilliam FD. Communications for Social Good. In: Patrizi P, Sherwood K, Spector A, editors. Practice Matters - The Improving Philanthropy Project. New York, NY, United States: The Foundation Center; 2004.
57. Olesk A, Renser B, Bell L, Fornetti A, Franks S, Mannino I, et al. Quality indicators for science communication: results from a collaborative concept mapping exercise. JCOM 2021;20(03):A06 [FREE Full text] [doi: 10.22323/2.20030206]
58. Pelligrini G. Evaluating science communication: Concepts and tools for realistic assessment. In: Bucchi M, Trench B, editors. Routledge Handbook of Public Communication of Science and Technology, 3rd ed. New York, NY, USA: Routledge; 2021:305-322.
59. Sevian H, Gonsalves L. Analysing how scientists explain their research: a rubric for measuring the effectiveness of scientific explanations. International Journal of Science Education 2008;30(11):1441-1467. [doi: 10.1080/09500690802267579]
60. Spicer S. The nuts and bolts of evaluating science communication activities. Semin Cell Dev Biol 2017;70:17-25. [doi: 10.1016/j.semcdb.2017.08.026] [Medline: 28823946]
61. Jensen E. The problems with science communication evaluation. JCOM 2014;13(01):C04. [doi: 10.22323/2.13010304]
62. Jensen E. Evaluate impact of communication. Nature 2011;469(7329):162-162. [doi: 10.1038/469162c] [Medline: 21228859]
63. Lang I, King A, Jenkins G, Boddy K, Khan Z, Liabo K. How common is patient and public involvement (PPI)? Cross-sectional analysis of frequency of PPI reporting in health research papers and associations with methods, funding sources and other factors. BMJ Open 2022;12(5):e063356 [FREE Full text] [doi: 10.1136/bmjopen-2022-063356] [Medline: 35613748]
64. Wilson P, Mathie E, Keenan J, McNeilly E, Goodman C, Howe A, et al. Research with patient and public involvement: a realist evaluation – the RAPPORT study. Health Serv Deliv Res 2015;3(38):1-176 [FREE Full text] [doi: 10.3310/hsdr03380] [Medline: 26378332]
65. UK standards for public involvement: better public involvement for better health and social care research. UK Public Involvement Standards Development Partnership. 2022. URL: https://drive.google.com/file/d/1U-IJNJCfFepaAOruEhzz1TdLvAcHTt2Q/view [accessed 2024-09-26]
66. Koops van 't Jagt R, Hoeks JCJ, Jansen CJM, de Winter AF, Reijneveld SA. Comprehensibility of health-related documents for older adults with different levels of health literacy: a systematic review. J Health Commun 2016;21(2):159-177. [doi: 10.1080/10810730.2015.1049306] [Medline: 26594852]
67. Kerwer M, Chasiotis A, Stricker J, Günther A, Rosman T. Straight from the scientist's mouth—plain language summaries promote laypeople's comprehension and knowledge acquisition when reading about individual research findings in psychology. Collabra: Psychology 2021;7(1):18898. [doi: 10.1525/collabra.18898]
68. Halliday M, Webster J. Language of Science 5. London: Continuum International Pub. Group; 2004.
69. Prelli LJ. A Rhetoric of Science: Inventing Scientific Discourse. Columbia, SC: University of South Carolina Press; 1989.
70. Schäfer MS, Füchslin T, Metag J, Kristiansen S, Rauchfleisch A. The different audiences of science communication: a segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. Public Underst Sci 2018;27(7):836-856. [doi: 10.1177/0963662517752886] [Medline: 29336218]
71. OST and Wellcome Trust. Science and the public: A review of science communication and public attitudes to science in Britain. London, UK: A Joint Report by the UK Office of Science and Technology and the Wellcome Trust; 2000. URL: https://wellcomecollection.org/works/gqk2wxw6/items [accessed 2024-09-26]
72. Ahmed S. On Being Included: Racism and Diversity in Institutional life. Durham, London: Duke University Press; 2012.
73. Warner M. Publics and Counterpublics. Zone Books. 2005. URL: https://press.princeton.edu/books/paperback/9781890951290/publics-and-counterpublics [accessed 2024-09-26]
74. Weingart P, Joubert M. The conflation of motives of science communication — causes, consequences, remedies. JCOM 2019;18(03):Y01 [FREE Full text] [doi: 10.22323/2.18030401]
75. Entradas M, Bauer MW, O'Muircheartaigh C, Marcinkowski F, Okamura A, Pellegrini G, et al. Public communication by research institutes compared across countries and sciences: building capacity for engagement or competing for visibility? PLoS One 2020;15(7):e0235191 [FREE Full text] [doi: 10.1371/journal.pone.0235191] [Medline: 32639974]

XSL•FO
RenderX

76. Carver RB. Public communication from research institutes: is it science communication or public relations? JCOM 2014;13(03):C01 [FREE Full text] [doi: 10.22323/2.13030301]
77. Dawson E. Reimagining publics and (non) participation: exploring exclusion from science communication through the experiences of low-income, minority ethnic groups. Public Underst Sci 2018;27(7):772-786 [FREE Full text] [doi: 10.1177/0963662517750072] [Medline: 29316863]
78. Levin N, Leonelli S. How does one "open" science? Questions of value in biological research. Sci Technol Human Values 2017;42(2):280-305 [FREE Full text] [doi: 10.1177/0162243916672071] [Medline: 28232768]
79. Maurer M, Siegel JE, Firminger KB, Lowers J, Dutta T, Chang JS. Lessons learned from developing plain language summaries of research studies. Health Lit Res Pract 2021;5(2):e155-e161 [FREE Full text] [doi: 10.3928/24748307-20210524-01] [Medline: 34213994]
80. Rubega MA, Burgio KR, MacDonald AAM, Oeldorf-Hirsch A, Capers RS, Wyss R. Assessment by audiences shows little effect of science communication training. Science Communication 2020;43(2):139-169. [doi: 10.1177/1075547020971639]
81. Kirkpatrick E, Gaisford W, Williams E, Brindley E, Tembo D, Wright D. Understanding plain English summaries. A comparison of two approaches to improve the quality of plain english summaries in research reports. Res Involv Engagem 2017;3(1):17 [FREE Full text] [doi: 10.1186/s40900-017-0064-0] [Medline: 29062542]
82. Jilka S, Hudson G, Jansli SM, Negbenose E, Wilson E, Odoi CM, et al. How to make study documents clear and relevant: the impact of patient involvement. BJPsych Open 2021;7(6):e32 [FREE Full text] [doi: 10.1192/bjo.2021.1040] [Medline: 33427156]
83. Thompson Coon J, Orr N, Shaw L, Hunt H, Garside R, Nunns M, et al. Bursting out of our bubble: using creative techniques to communicate within the systematic review process and beyond. Syst Rev 2022;11(1):56 [FREE Full text] [doi: 10.1186/s13643-022-01935-2] [Medline: 35379331]
84. Bredbenner K, Simon SM. Video abstracts and plain language summaries are more effective than graphical abstracts and published abstracts. PLoS One 2019;14(11):e0224697 [FREE Full text] [doi: 10.1371/journal.pone.0224697] [Medline: 31743342]
85. Martínez Silvagnoli L, Shepherd C, Pritchett J, Gardner J. Optimizing readability and format of plain language summaries for medical research articles: cross-sectional survey study. J Med Internet Res 2022;24(1):e22122 [FREE Full text] [doi: 10.2196/22122] [Medline: 35014966]
86. Coleman E, O'Sullivan L, Crowley R, Hanbidge M, Driver S, Kroll T, et al. Preparing accessible and understandable clinical research participant information leaflets and consent forms: a set of guidelines from an expert consensus conference. Res Involv Engagem 2021;7(1):31 [FREE Full text] [doi: 10.1186/s40900-021-00265-2] [Medline: 34006326]
87. Lewenstein B. Can we understand citizen science? JCOM 2016;15(01):E [FREE Full text] [doi: 10.22323/2.15010501]
88. Delfanti A. Users and peers. From citizen science to P2P science. JCOM 2010;09(01):E. [doi: 10.22323/2.09010501]
89. Hecker S, Luckas M, Brandt M, Kikillus H, Marenbach I, Schiele B, et al. Stories Can Change The World - Citizen Science Communication in Practice. In: Hecker S, Haklay ME, Bowser A, Makuch Z, Vogel J, Bonn A, editors. Citizen Science: Innovation in Open Science, Society and Policy. London: UCL Press; 2018:445-462.
90. Hecker S, Haklay M, Anne B, Makuch Z, Johannes V, Aletta B. Innovation in open science, society and policy - setting the agenda for citizen science. In: Hecker S, Haklay M, Boswer A, Makuch Z, Vogel J, Bonn A, editors. Citizen Science: Innovation in Open Science, Society and Policy. London, UK: UCL Press; 2018:1-23.
91. MacGregor S, Cooper A. Blending research, Journalism, and community expertise: a case study of coproduction in research communication. Science Communication 2020;42(3):340-368. [doi: 10.1177/1075547020927032]
92. Stewart EC, Davis JS, Walters TS, Chen Z, Miller ST, Duke JM, et al. Development of strategies for community engaged research dissemination by basic scientists: a case study. Transl Res 2023;252:91-98. [doi: 10.1016/j.trsl.2022.09.001] [Medline: 36108910]
93. Knowles SE, Allen D, Donnelly A, Flynn J, Gallacher K, Lewis A, et al. Participatory codesign of patient involvement in a learning health system: how can data-driven care be patient-driven care. Health Expect 2022;25(1):103-115 [FREE Full text] [doi: 10.1111/hex.13345] [Medline: 34668634]

Abbreviations

FRE: Flesch Reading Ease
NIHR: National Institute for Health and Care Research
PLS: plain language summary

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 14.07.23; peer-reviewed by M Stoll, C Baur; comments to author 03.11.23; revised version received 04.03.24; accepted 23.09.24; published 13.01.25.

Please cite as:
Lang IA, King A, Boddy K, Stein K, Asare L, Day J, Liabo K
Jargon and Readability in Plain Language Summaries of Health Research: Cross-Sectional Observational Study
J Med Internet Res 2025;27:e50862
URL: https://www.jmir.org/2025/1/e50862
doi:10.2196/50862
PMID:

©Iain A Lang, Angela King, Kate Boddy, Ken Stein, Lauren Asare, Jo Day, Kristin Liabo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of Internet Usage on Depression Among Older Adults: Comprehensive Study

Lin Guo[1*], PhD; Yunwei Li[1*], MA; Kai Cheng[1], BA; Ying Zhao[1], BA; Wenqiang Yin[1], PhD; Ying Liu[2], PhD

[1]School of Management, Shandong Second Medical University, Weifang, China

[2]School of Humanities and Management, Zhejiang Chinese Medical University, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ying Liu, PhD
School of Humanities and Management
Zhejiang Chinese Medical University
No. 260 Baichuan Street, Fuyang District
Hangzhou, 311402
China
Phone: 86 18667013568
Email: skyliu2171@163.com

Abstract

Background: Depression is a widespread mental health issue affecting older adults globally, with substantial implications for their well-being. Although digital interventions have proven effective in high-income countries, research on the potential of internet usage to alleviate depression among older adults in high-income countries remains limited.

Objective: This study aimed to examine the impact of internet usage on depression among older adults in high-income countries by developing a comprehensive theoretical framework and testing key hypotheses.

Methods: Using data from the China Health and Retirement Longitudinal Study (CHARLS), a 2-stage instrumental variable approach was applied to address endogeneity and estimate causal relationships between internet usage and depression.

Results: The findings indicate that internet usage results in a 1.41% reduction in depression levels among older adults. This effect is mediated by four primary mechanisms: (1) enhanced social interaction, (2) increased physical activity, (3) improved intergenerational contact, and (4) expanded access to educational opportunities. A heterogeneity analysis revealed that these effects are more pronounced in urban areas, eastern regions, and regions with superior internet infrastructure.

Conclusions: Internet usage plays a crucial role in alleviating depression among older adults in high-income countries, with regional variations. The findings highlight the need for targeted policy interventions to improve internet access and digital literacy, which can mitigate depression and enhance the mental health of older adults.

(J Med Internet Res 2025;27:e65399) doi:10.2196/65399

KEYWORDS

internet usage; depression; older people; mechanism; heterogeneity; mobile phone

Introduction

Background

Depression is a prevalent psychological and mental health condition. Globally, the prevalence of depression among the older population aged 60 years and above is 5.7%. This issue is even more pronounced in low- and middle-income countries [1]. Data from the United Nations Economic and Social Commission for Asia and the Pacific indicate that the older population in Southeast Asia is growing rapidly. By 2030, individuals aged 65 years and older are projected to make up over 11% of the total population in the region. Mental health issues among this demographic, particularly depression, are becoming a growing concern. A study by the Asian Mental Health Association shows that depression is particularly prevalent among older adults in Southeast Asia. This is especially true for older adults affected by loneliness, physical health issues, and lifestyle changes, which can contribute to higher rates of depression. Common symptoms of depression in older adults include feelings of isolation, sleep disturbances, and low mood [2].

XSL•FO
RenderX

In China, the prevalence of depression among older adults is estimated to be as high as 25.55% [3]. From 2020 to 2050, China's older population aged 60 years and older is projected to increase from 263 million to 522 million, with its proportion of the total population rising from 18.7% to 39.5% [4]. Depression significantly affects the quality of life, daily activities, incidence of geriatric diseases, and suicide rates among older adults [5,6]. It not only causes severe harm to the physical and mental health of older adults but also imposes a significant burden on society. Estimates suggest that health care expenditures for individuals with depression are 50% to 100% higher than for those without depression [7]. Thus, effectively alleviating depression among older adults has become an urgent global health issue.

In recent years, information and communication technology has been recognized as a potential factor for alleviating depression in older adults [8]. Studies in early-aging countries such as the United Kingdom, the United States, Japan, and South Korea have confirmed that digital interventions can effectively alleviate depression among older adults [9-11]. Internet usage in Southeast Asia has experienced a significant upward trend in recent years. Statista reports that the internet penetration rate in the region was about 67.1% in 2024, with most countries exceeding a 70% penetration rate. However, smaller countries like Laos, Myanmar, and Timor-Leste have lower internet access rates, although improvements are ongoing. Digital technologies are increasingly applied in health management and social connectivity, with the widespread use of smartphones and wearable devices offering older adults comprehensive platforms for health monitoring and social interaction [12].

As the world's largest high-income country, China had 119 million older internet users by the end of 2021, with an internet penetration rate of 43.2% among older adults. This raises a critical question: can internet usage effectively alleviate depression among older adults in China, and what are the potential mechanisms at play? Unfortunately, the academic community has paid insufficient attention to this issue, with few studies providing rigorous theoretical analysis and empirical evidence. To fill this gap, this paper investigates the causal relationship between internet usage and depression among older adults from both theoretical and empirical perspectives. The study aims to elucidate the underlying mechanisms and assess the impact of internet usage, thereby providing evidence from high-income countries to improve the mental health of older adults.

The primary challenge of this study is the inherent issue of reverse causality between internet usage and depression among older adults. Internet usage may influence depression among older adults, and conversely, depression may affect internet usage among older adults. This potential endogeneity problem significantly disrupts causal inference, greatly reducing the reliability of the regression coefficients. Existing studies have inadequately addressed this issue. The 2-stage instrumental variable (IV) method is widely regarded as one of the most effective approaches to overcome potential endogeneity issues [13]. By using an exogenous instrument, this method aims to eliminate the endogeneity of the primary variable, thereby enabling precise estimation of the regression coefficients. This approach requires the IV to be correlated with the primary variable while being exogenous to the error term.

This paper innovatively uses the National Big Data Comprehensive Pilot Zone policy as an IV for internet usage, using the 2-stage IV method to achieve more accurate causal identification. The National Big Data Comprehensive Pilot Zone policy, through the development of age-friendly applications and digital literacy training, significantly increases internet usage among older adults, meeting the relevant criterion. Simultaneously, depression among older adults does not influence the National Big Data Comprehensive Pilot Zone policy, satisfying the exogeneity criterion. In addition, to further eliminate endogeneity caused by omitted variables, this paper controls both time-fixed effects and individual-fixed effects in the 2-stage IV approach.

Compared with existing studies, this paper offers several potential contributions: First, this study provides a rigorous analysis of the causal relationship between internet usage and depression among older adults, both theoretically and empirically, thereby contributing to the literature on the mental health of older adults. Second, it is the first to use the 2-stage IV approach for a stringent empirical examination, effectively mitigating endogeneity issues in regression analysis, and thereby enriching existing empirical methodologies. Third, this paper conducts a heterogeneity analysis across different regions, expanding on current research that mostly focuses on heterogeneity based on demographic characteristics, thus broadening the scope of existing studies. Finally, we offer several specific and actionable policy recommendations, providing a basis for decision-making to improve the mental health of older adults in other high-income countries.

The remainder of this paper is organized as follows: (1) section 2 proposes testable theoretical hypotheses; (2) section 3 discusses empirical design, including data sources, econometric models, and variable definitions; (3) section 4 presents a series of empirical tests and provides an in-depth discussion of the results; and (4) section 5 concludes with the findings, policy recommendations, and limitations.

Theoretical Hypotheses

Social Interaction Effect

Social networks refer to stable relationships formed through mutual connections among individuals and social members, encompassing both online and offline networks. Among these, family and friend networks constitute the most crucial offline social networks for older adults. However, due to physical decline, generation gaps, and children moving away or being busy with work, older adults find it challenging to maintain offline social networks. With the proliferation of the digital society, the internet provides a new social platform for older adults, helping them engage in social interactions, learning, and shopping, thereby enhancing their connection with the outside world. As of June 2023, China had approximately 140 million internet users aged 60 years and older, and this proportion continues to expand with increasing internet penetration. Internet usage not only alleviates loneliness among older adults but also

XSL•FO
RenderX

expands their social relationships, prevents social isolation, and increases interaction frequency with family and friends [14].

Enhanced social networks have a significantly positive impact on the psychological well-being of older adults. Social interactions through the internet reduce feelings of loneliness and depression among older adults [15]. Social engagement enhances their sense of achievement, happiness, and security, serving as a crucial pillar for their integration into society and promotion of psychological health. The activity theory suggests that older adults can achieve self-awareness, bridge the gap with society, lessen loneliness, and attain higher life satisfaction through new social engagements. Increasing participation in both conventional and novel social activities is a valuable objective for improving the mental health of the increasingly aging population. The Social Convoy Theory also highlights the protective role of social relationship networks on psychological health. Therefore, by strengthening family and friend networks, the internet alleviates loneliness among older adults, promotes psychological health, and effectively reduces depression, a phenomenon termed here as the social interaction effect. Based on this, hypothesis 1 is proposed.

Hypothesis 1—Internet Usage Mitigates Depression Among Older Adults Through the Social Interaction Effect

Physical Exercise Effect

A series of studies have shown that internet usage among older adults transcends temporal and geographical constraints, enhances interaction with sports service providers, and offers high-quality, convenient, and diverse sports services [16,17]. In the era of big data, the internet also accurately delivers nearby sports information and suitable sports equipment, thus lowering barriers to the participation of older adults in physical activities. Older adults who frequently use the internet engage in physical exercise and moderate- to high-intensity activities more frequently each week. This suggests that the internet significantly boosts older adults' enthusiasm for physical activity by providing convenience, scientific guidance, interactivity, and entertainment.

Increasing physical exercise has a significant positive impact on the psychological health of older adults. Physical exercise can enhance the cardiovascular function and physical fitness of older adults, promote health, delay aging, reduce the incidence of age-related diseases, and improve sleep quality. Research indicates that physical activity enhances subjective well-being [18], effectively treats depression, and reduces anxiety [19]. Based on the active aging theory and health demand model, studies have found that regular physical exercise significantly increases the likelihood of social participation among older adults [20]. Therefore, through increasing physical exercise, older adults not only improve their physical health but also significantly alleviate depression and enhance overall quality of life. We refer to this mechanism as the "physical exercise effect." Based on this, hypothesis 2 is proposed.

Hypothesis 2—Internet Usage Alleviates Depression Among Older Adults Through the Physical Exercise Effect

Intergenerational Contact Effect

The current high population mobility in China often leaves adult children unable to care for their older parents due to work or study commitments, resulting in a shift in social roles and interpersonal relationships for older adults upon retirement. This situation limits their access to information and deepens their sense of social isolation, impacting their psychological and mental health [21]. The development of internet technology has provided convenient means for older individuals to stay in touch with their children, overcoming temporal and spatial barriers. This has significantly increased the frequency and quality of interactions between older parents and their children, thereby enhancing intergenerational support and subsequently improving the older adults' subjective well-being [22].

According to the social support buffering model and main effects model, intergenerational support plays a crucial role in alleviating depression among older adults. On the one hand, intergenerational support helps mitigate the impact of resource loss and buffers the effects of stress on the depression of older adults. On the other hand, intergenerational support provides older adults with positive social evaluations for effective parenting, directly reducing their depression [23]. The cognitive behavioral theory of depression emphasizes that cognitive assessments are crucial variables influencing depression [24]. Intergenerational support meets the primary social goal of emotional management for older adults and fosters a more positive attitude in cognitive assessments, effectively reducing depression among older adults. We refer to this mechanism as the "intergenerational contact effect" and propose hypothesis 3 based on this premise.

Hypothesis 3—Internet Usage Mitigates Depression Among Older Adults Through the Intergenerational Contact Effect

Educational Enhancement Effect

With the widespread adoption of the internet, an increasing number of older individuals are engaging in online education, learning new knowledge and skills through virtual courses. This allows them to find new joys and values in the learning process [25,26]. The internet enables older individuals to access a broader range of information and knowledge, enhancing cognitive abilities and better adapting to social changes, thereby reducing feelings of loneliness caused by information isolation. In summary, the internet provides new learning opportunities for older individuals, expands their knowledge boundaries, and profoundly impacts their education.

Education can effectively mitigate the negative impact of age, gender, and family structure on the psychological health of older individuals [27-29]. As individuals age, their physical fitness gradually declines, but higher levels of education can help older individuals enhance their objective awareness of their health status, thereby influencing their psychological well-being. Studies indicate a positive correlation between cognitive

XSL•FO
RenderX

function and life satisfaction among older individuals [30]. Education enhances the self-awareness and self-care abilities of older individuals, reducing psychological issues, and providing social and recreational opportunities, thereby improving mental health and enhancing overall quality of life. We refer to this mechanism as the educational enhancement effect. Based on this, hypothesis 4 is proposed.

Hypothesis 4—Internet Usage Alleviates Depression Among Older Individuals Through the Educational Enhancement Effect

Figure 1 below presents the conceptual framework of this study.

Figure 1. Conceptual framework.

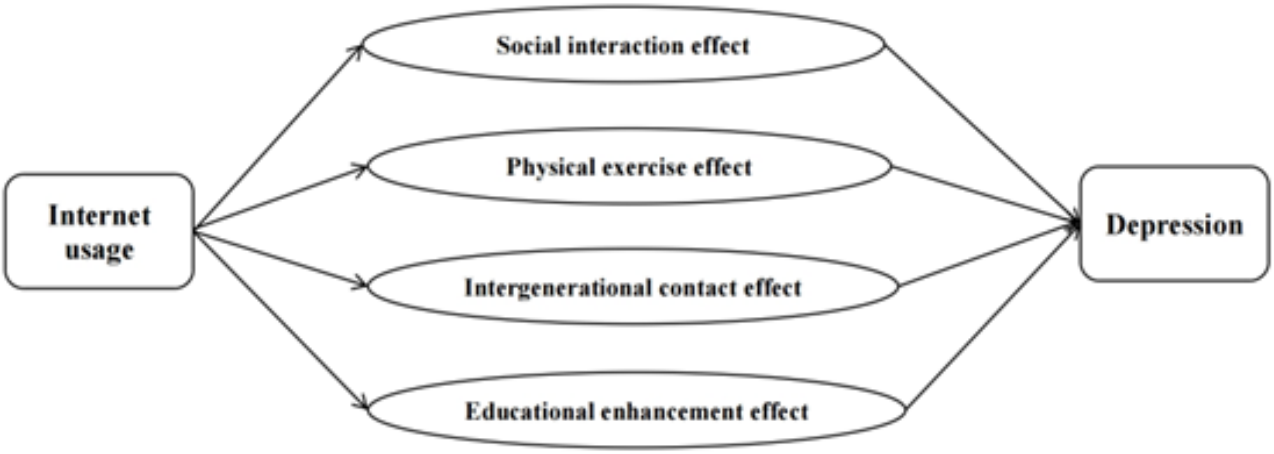

Internet usage serves as the primary exposure variable in this study, while depression is the main outcome variable. In addition, 4 potential mechanisms are explored: the social interaction effect, the physical exercise effect, the intergenerational contact effect, and the educational enhancement effect. These mechanisms are examined to better understand how internet usage might influence depression through various indirect pathways.

Methods

Data Sources

The data for this study primarily originate from the China Health and Retirement Longitudinal Study (CHARLS). CHARLS is a comprehensive, large-scale interdisciplinary survey led by the National School of Development at Peking University, conducted in partnership with the China Social Science Survey Center and the Youth League Committee of Peking University. The survey's objective is to gather household and individual-level data on Chinese adults aged 45 years and older. CHARLS uses a probability proportional to size sampling method, systematically selecting samples from counties, villages, residential areas, households, and individuals. The national baseline survey was conducted in 2011, encompassing 150 counties (districts), 450 villages (residential areas), 10,257 households, and 17,708 individuals, offering a comprehensive representation of China's middle-aged and older population. The CHARLS questionnaire collects data on demographics, family structure, employment and retirement, health status, and health-related behaviors. This study uses longitudinal panel data from 5 waves of CHARLS, spanning 2011 to 2020. The longitudinal data structure facilitates the analysis of individual-level changes over time, offering a more robust framework for examining causal relationships than cross-sectional data. In contrast to cross-sectional data, which captures a snapshot of individuals at a single point in time, longitudinal data enable the observation of changes within the same individuals across multiple time periods. This approach enhances our capacity to control unobserved individual heterogeneity and to more effectively assess temporal dynamics and causal mechanisms. Currently, CHARLS is recognized as the most authoritative and largest representative database for studying the health of older adults in China.

The data cleaning process proceeded as follows: (1) the data from the 5 survey rounds were matched and merged, resulting in a total of 96,616 observations; (2) missing values for the key explanatory and dependent variables were removed, totaling 15,735 (ie, 96,616–80,881); (3) further removal of missing values for control variables amounted to 1179 (ie, 80,881–79,702); and (4) subsequently, 2826 (ie, 79,702–76,876) singleton group values were removed. This resulted in a final sample of 76,876 observations for use in regression analysis.

Econometric Equation

This study uses the IV regression method to address potential endogeneity issues in the data. Endogeneity occurs when an explanatory variable is correlated with the error term, often due to omitted variables, measurement errors, or reverse causality. This endogeneity can result in biased and inconsistent estimates of causal relationships. IV regression addresses this issue by using IVs that are correlated with the endogenous regressor but uncorrelated with the error term, providing a source of exogenous variation to identify causal effects. This method is particularly valuable in panel data analysis, where variable relationships are dynamic and prone to simultaneity bias. Using IVs allows for more reliable and unbiased estimates of causal effects, making this method well-suited for the study design. Thus, the IV approach offers a robust statistical framework to address endogeneity concerns and ensure the validity of causal inferences. The specific econometric model is as follows:

XSL•FO
RenderX





Equation (1) represents the first-stage regression, where  denotes the fitted values of internet usage, $NBDCPZ_{it}$ represents the dummy variable for National Big Data Comprehensive Pilot Zones (NBDCPZ), $X`\theta$ denotes a series of control variables, σ_i represents individual fixed effects to control for individual differences, and γ_t represents time fixed effects to control for unobserved shocks across different years. Controlling for both time- and individual fixed effects aims to further eliminate endogeneity issues caused by omitted variables. The essence of the first-stage regression lies in regressing the endogenous variable $Internet_{it}$ on exogenous variables to obtain the fitted values  . Equation (2) represents the second-stage regression, where denotes depression among the older adults, ε_{it} represents the error term, with other variables as in Equation (1). The essence of the second-stage regression lies in regressing $Lndepress_{it}$ on the instrumented  to obtain accurate estimates of the coefficient α_1.

Variable Definition

Depression (LnDepress)

The level of depression among older adults was assessed using the widely recognized 10-item Center for Epidemiologic Studies Depression Scale (CESD-10). This scale consists of 10 items, each rated on a 4-point Likert scale ranging from 1 to 4, with items 5 and 8 reverse scored. The total score, derived by summing the individual item scores, reflects the level of depression, with higher scores indicating a greater risk of depression. To mitigate the impact of extreme values and facilitate the interpretation of results, the total depression score was logarithmically transformed.

Internet Usage (Internet)

Internet usage was assessed by asking respondents in the CHARLS survey, “Have you used the internet in the past month?” Responses were coded as 1 for “yes” and 0 for “no.” This core explanatory variable exhibits significant potential endogeneity issues with depression; higher levels of depression among older adults may inversely affect their internet usage behavior, potentially compromising the credibility of regression results.

Instrumental Variable (National Big Data Comprehensive Pilot Zone)

On March 1, 2016, China’s first NBDCPZ (Guizhou) received official approval from the National Development and Reform Commission, the Ministry of Industry and Information Technology, and the Cyberspace Administration of China. This pilot zone aims to conduct systematic trials in areas such as data resource management and sharing, data center integration, and data resource application. The plan is to develop a series of advanced big data products and cultivate key big data enterprises over 3 to 5 years, thereby promoting economic transformation and upgrading.

On October 8, 2016, these authorities issued another notice approving the establishment of additional NBDCPZ in 7 regions, including Beijing-Tianjin-Hebei. These newly approved national-level pilot zones comprise 2 interregional comprehensive pilot zones (Beijing-Tianjin-Hebei and the Pearl River Delta), 4 regional demonstration pilot zones (Shanghai, Henan, Chongqing, and Shenyang), and 1 pilot zone focused on the coordinated development of big data infrastructure (Inner Mongolia). The construction of these zones will facilitate trials and exploration in areas such as big data policy innovation, public data openness and sharing, big data innovation applications, big data industry clustering, big data element circulation, data center integration and use, and international cooperation in big data, thereby driving the innovative development of big data in China.

In this study, we creatively use the NBDCPZ as an IV for internet usage for older adults. This approach effectively addresses potential endogeneity issues, thereby enhancing the credibility of the regression results.

Control Variables

The selected control variables include gender, age, household registration, marital status, education, self-rated health, medical insurance, personal income, and alcohol consumption.

Mechanism Variables

In line with the theoretical hypotheses, the mechanism variables encompass the social interaction effect, physical exercise effect, intergenerational contact effect, and educational enhancement effect. Detailed measurement methodologies for these variables are provided in the section “Mechanism Explorations.”

Descriptive Statistics

Table 1 presents the main variables and their descriptive statistics are as follows: (1) the mean score for depression (LnDepress) is 2.847 (SD 0.322); (2) the mean for internet usage (internet) is 0.149 (SD 0.356), indicating that 14.9% of the sample has used the internet; and (3) and the mean for NBDCPZ is 0.1 (SD 0.3), suggesting that 10% of observations are from the NBDCPZ. Details for other variables are omitted for brevity. It should be noted that in Table 1, the number of observations for variables is higher than in the regression tables. This discrepancy is due to the removalx of some singleton groups during the regression process. This deletion was automatic and did not substantively affect the conclusions of this study.

XSL•FO
RenderX

Table 1. Definition and description of main variables.

Variables	Observations, n	Mean (SD)	Definition
LnDepress	79,702	2.847 (0.332)	Using the CESD-10[a] scale, scores were summed and logged
Internet	79,702	0.149 (0.356)	Respondents were asked whether they used the internet in the past month, with responses coded as 1 for "yes" and 0 for "no"
National Big Data Comprehensive Pilot Zone	79,702	0.1 (0.3)	If located in a pilot area, assign a value of 1 to years after the pilot period; otherwise, assign a value of 0
Gender	79,702	1.521 (0.5)	Male=1, female=2
Age	79,702	60.296 (9.731)	Current year: year of birth. Due to the presence of outliers, perform a 1% truncation of the upper and lower ends
Household registration	79,702	0.45 (0.497)	Rural household registration=1, others=0
Marital status	79,702	0.88 (0.325)	Married=1, others=0
Education	79,702	2.398 (4.116)	Not completed primary school=0, primary school=6, junior high school=9, senior high school or technical school=12, junior college=15, bachelor's degree=16, master's degree=19, doctorate=22
Self-rated health	79,702	2.78 (0.933)	Very poor=1, poor=2, fair=3, good=4, very good=5, excellent=6
Medical insurance	79,702	0.95 (0.219)	With medical insurance=1, other=0
Personal income	79,702	0.231 (0.421)	Received wages or bonuses in the past year=1, other=0
Alcohol consumption	79,702	2.375 (0.88)	In the past month, drank alcohol more than once=1, less than once=2, did not drink= 3
Social interaction effect	41,934	0.357 (0.455)	The frequency of social interaction activities was used as a measure, as detailed in the section "Mechanism Explorations"
Physical exercise effect	47,768	2.277 (0.552)	The frequency of physical exercise activities was measured as detailed in the section "Mechanism Explorations"
Intergenerational contact effect	34,020	1.871 (0.909)	The frequency of contact with children was used as a measure, as detailed in the section "Mechanism Explorations"
Educational enhancement effect	32,108	0.013 (0.115)	Measured by participation in school training courses over the past month, as detailed in the section "Mechanism Explorations"

[a]CESD-10: Center for Epidemiologic Studies Depression Scale.

Ethical Considerations

The original data used in this study are derived from the CHARLS, which has been reviewed and approved by the Biomedical Ethics Review Committee of Peking University (approval IRB 00001052–11015). The current analysis is a secondary analysis of deidentified and publicly accessible data. This study does not involve any direct interaction with human participants or observation of individuals. In addition, this study does not involve any procedures or activities that may raise ethical concerns and complies with all applicable laws and regulations regarding research conduct. Therefore, no ethical approval is required.

Results

Benchmark Regression

To present comparative results, columns 1 and 2 in Table 2 show the results of ordinary least squares (OLS) regressions: column 1 indicates that the coefficient of internet usage (internet) on depression (LnDepress) is –0.05, significant at the 1% level. Column 2 presents the regression results further controlling for both individual fixed effects (FE) and year FE; the coefficient of internet usage "Internet" is –0.02, still significant at the 1% level. In Table 2, columns 3 and 4 present the results of the 2-Stage IV method: column 3 shows that the coefficient of internet usage "Internet" on depression (LnDepress) is –0.00, significant at the 1% level, with Kleibergen-Paap rk LM statistic and Cragg-Donald Wald F statistic tests confirming the validity of the instruments. Column 4 presents the regression results further controlling for individual FE and year FE; the coefficient of internet usage "Internet" is –0.01, and the instruments remain valid.

XSL•FO
RenderX

Table 2. Benchmark regression. Robust standard errors are in parentheses.

Variables	LnDepress			
	OLS[a] (1)	OLS-FE[b] (2)	IV[c] (3)	IV-FE[d] (4)
Internet	–0.0483[e] (0.00308)	–0.0151[e] (0.00373)	–0.00160[e] (0.000209)	–0.0141[e] (0.00350)
Gender	0.0745[e] (0.00256)	0.00492 (0.0244)	0.000754[e] (2.59e–05)	0.000501 (0.000398)
Age	0.000898[e] (0.000126)	0.00380 (0.00426)	3.61e–06[f] (1.53e–06)	–0.000119 (8.97e–05)
Household registration	0.0803[e] (0.00228)	0.00178 (0.00369)	0.000779[e] (2.36e–05)	0.000114 (8.08e–05)
Marital status	–0.0780[e] (0.00369)	–0.0595[e] (0.00742)	–0.000775[e] (3.71e–05)	0.000146 (0.000225)
Education	–0.00563[e] (0.000272)	–0.00101[e] (0.000322)	–5.66e–05[e] (2.75e–06)	–7.91e–05[e] (1.85e–05)
Self-rated health	–0.0897[e] (0.00123)	–0.0257[e] (0.00122)	–0.000867[e] (1.35e–05)	–0.000271[e] (2.35e–05)
Medical insurance	–0.0126[f] (0.00507)	–0.00371 (0.00490)	–8.55e–05[g] (5.15e–05)	0.000173[g] (0.000101)
Personal income	–0.0470[e] (0.00277)	–0.0178[e] (0.00317)	–0.000336[e] (3.87e–05)	6.13e–05 (8.69e–05)
Alcohol consumption	0.00788[e] (0.00140)	0.00465[f] (0.00198)	5.29e–05[e] (1.50e–05)	–0.000224[e] (7.79e–05)
KP-LM[h]	No	No	1073.066	20.724
CD-Wald[i]	No	No	2000.060	26.327
Year fixed effect[j]	No	Yes	No	Yes
Individual fixed effect[k]	No	Yes	No	Yes
Robust[l]	Yes	Yes	Yes	Yes
Observations	79,702	79,702	79,702	76,876
R^2	0.127	0.023	No	No

[a]OLS: ordinary least squares.

[b]OLS-FE: ordinary least squares-fixed effects.

[c]IV: instrumental variable.

[d]IV-FE: instrumental variable-fixed effects.

[e]$P<.01$.

[f]$P<.05$.

[g]$P<.1$.

[h]KP-LM: Kleibergen-Paap rk LM statistic, testing the relevance of IV.

[i]CD-Wald: Cragg-Donald Wald F statistic, testing the weak relevance of IV.

[j]Year FE: time fixed effects, controlling for unobserved shocks across different years.

[k]Individual FE: individual fixed effects, controlling for heterogeneity across individuals.

[l]Robust: the use of robust SEs, addressing heteroskedasticity effects.

The baseline regression results indicate the following: first, the coefficient significantly decreases after using IV regression, suggesting that the OLS regression overestimated the impact of internet usage on depression. Second, column 4, which simultaneously controls both individual FEs and year FEs, further mitigates endogeneity issues due to omitted variables. Therefore, column 4 is used as the baseline regression in this study, showing that internet usage significantly reduces depression scores by 1.41%. This finding indicates a significant alleviation of depression among older adults, consistent with the theoretical hypothesis.

Column 4 also reports the impacts of other control variables: education significantly reduces LnDepress, indicating that higher education levels ease depression. Self-rated health significantly reduces LnDepress, suggesting that better self-rated health decreases the likelihood of depression. Medical insurance significantly promotes LnDepress, implying that medical insurance does not alleviate depression. Alcohol consumption significantly reduces LnDepress, indicating that alcohol consumption is beneficial in alleviating depression. The other variables gender, age, household registration, marital status, and personal income do not show significant effects in the baseline regression. It is important to note that the reported R^2 in the IV regression process has no practical significance; hence, it is not reported in this study [31].

XSL•FO
RenderX

Robustness Checks

To examine the robustness of the baseline regression results, this section conducts a series of robustness tests, as detailed in Table 3: column 1 transforms the depression variable into a binary outcome, following the approach of Li and Luo (2024) [32], where depression scores exceeding 10 are assigned a value of 1, and 0 otherwise. Using the IV method based on a Probit model, the results indicate a significant reduction in depression due to internet usage (the coefficient is –2.116 and statistically significant at the 1% level), confirming robustness. Column 2 re-estimates the regression by internet usage frequency instead of binary internet usage. Frequency categories (almost daily=3, almost every week=2, not regularly=1) are derived directly from the CHARLS questionnaire. The results show a significant decrease in depression by 2.01%, reaffirming robustness. Column 3 replaces the IV method with propensity score matching to address potential selection bias in internet usage, where the choice of using the internet may be influenced by other factors, leading to biased regression coefficients. Using a 1:1 nearest-neighbor matching, the results indicate a significant reduction in depression by 0.0761%, demonstrating robust results. Column 4 uses a smart city pilot virtual variable as a new IV for the re-estimation. The results show a significant reduction in depression by 1.35%, confirming robustness. In addition, we conducted other robustness tests, such as re-estimating using the total depression score and applying a negative binomial regression model. The results remained robust across these specifications, and further details are omitted for brevity. In addition, we conducted weighted regressions at the individual, gender, household registration, household, community, and city levels. The results remained robust and are detailed in Multimedia Appendix 1.

Table 3. Robustness regression. Robust SE in parentheses.

	LnDepress			
Variables	Probit model (1)	Frequency (2)	Propensity score matching (3)	Smart city (4)
Internet	–2.116[a] (0.53)	–0.0201[a] (0.00111)	–0.000761[a] (0.0000877)	–0.0135[a] (0.00520)
Controls[b]	Yes	Yes	Yes	Yes
KP-LM[c]	No	182.559	No	6.757
CD-Wald[d]	No	304.661	No	5.604
Year FE[e]	Yes	Yes	Yes	Yes
Individual FE[f]	Yes	Yes	No	Yes
Robust[g]	Yes	Yes	Yes	Yes
Observations	88,359	2691	79,702	76,876
R^2	No	No	0.348	No

[a]$P<.01$.

[b]Controls represent a series of control variables.

[c]KP-LM: Kleibergen-Paap rk LM statistic, testing the relevance of IV.

[d]CD-Wald: Cragg-Donald Wald F statistic, testing the weak relevance of IV.

[e]Year FE: time fixed effects, controlling for unobserved shocks across different years.

[f]Individual FE: individual fixed effects, controlling for heterogeneity across individuals.

[g]Robust: the use of robust SEs, addressing heteroskedasticity effects.

Mechanism Explorations

The empirical results above clearly demonstrate that internet usage significantly reduces depression levels among older adults. Moreover, the theoretical hypotheses proposed above have anticipated potential mechanisms. This section will therefore conduct mechanism tests to validate these theoretical hypotheses. Detailed results are provided in Table 4 below.

XSL•FO
RenderX

Table 4. Mechanism test. Robust standard errors are in parentheses.

Variables	Social interaction effect (1)	Physical exercise effect (2)	Intergenerational contact effect (3)	Educational enhancement effect (4)
Internet	0.492[a] (0.244)	0.680[a] (0.291)	0.374[a] (0.151)	2.183[b] (0.260)
Controls[c]	Yes	Yes	Yes	Yes
KP-LM[d]	6.943	22.750	9.698	No
CD-Wald[e]	8.832	24.261	10.196	No
Year FE[f]	Yes	Yes	Yes	Yes
Individual FE[g]	Yes	Yes	Yes	No
Robust[h]	Yes	Yes	Yes	Yes
Observations	38,860	47,905	30,362	38,985

[a]$P<.05$.

[b]$P<.01$.

[c]Controls represent a series of control variables.

[d]KP-LM: Kleibergen-Paap rk LM statistic, testing the relevance of IV.

[e]CD-Wald: Cragg-Donald Wald F statistic, testing the weak relevance of IV.

[f]Year FE: time fixed effects, controlling for unobserved shocks across different years.

[g]Individual FE: individual fixed effects, controlling for heterogeneity across individuals.

[f]Robust: the use of robust standard errors, addressing heteroskedasticity effects.

Social Interaction Effect

Regarding social interactions, the following activities were quantified: visiting friends in the past week; playing mahjong, chess, cards, or visiting community activity rooms; providing assistance to non-cohabitating relatives, friends, or neighbors; engaging in activities such as dancing, fitness, or qigong with clubs or organizations; volunteering or participating in charitable activities; and caring for non-cohabitating patients or disabled individuals. The total frequency of these activities was aggregated and log-transformed in this study. Column 1 indicates that internet usage increases the frequency of older social interactions by 49.2%, confirming the social interaction effect.

Physical Exercise Effect

Regarding physical exercise, the following activities were quantified: the number of days per week engaging in vigorous, moderate, or light physical activities for more than 10 minutes. The total frequency of these activities was aggregated and log-transformed in this study. Column 2 indicates that internet usage increases the frequency of older adults' physical exercise by 68%, confirming the physical exercise effect.

Intergenerational Contact Effect

Regarding intergenerational contact, activities include contacting all children through phone, message, WeChat, mail, or email, and assigning reverse scores based on the frequency of these contacts as follows: almost every day=10, 2-3 times a week=9, once a week=8, every 2 weeks=7, once a month=6, once every 3 months=5, once every 6 months=4, once a year=3, almost never=2, other=1. The total scores of these frequencies were aggregated and log-transformed in this study. Column 3 indicates that internet usage increases the intergenerational contact scores of older adults by 37.4%, confirming the intergenerational contact effect.

Educational Enhancement Effect

Regarding educational enhancement, the following activity was quantified: whether the individual attended school or training courses in the past month. Clearly, this is a binary variable and should be regressed using the 2-stage IV method based on the Probit model. Column 4 demonstrates that internet usage significantly increases the likelihood of older adults receiving education, confirming the educational enhancement effect.

Heterogeneity Analyses

To further enrich the understanding of the relationship between internet usage and depression, this study conducted a series of heterogeneity analyses, as shown in Table 5.

XSL•FO
RenderX

Table 5. Heterogeneity regression. Robust SEs are in parentheses.

	LnDepress					
Variables	Urban (1)	Village (2)	Eastern (3)	Noneastern (4)	Pilot (5)	Nonpilot (6)
Internet	–0.0174[a] (0.00296)	–0.00774[a] (0.00220)	–0.0185[a] (0.00542)	0.00196 (0.00267)	–0.0190[b] (0.0102)	–0.000707 (0.00393)
Controls[c]	Yes	Yes	Yes	Yes	Yes	Yes
KP-LM[d]	9.438	21.456	13.549	10.199	3.786	4.446
CD-Wald[e]	10.549	39.755	17.314	12.886	3.903	6.327
Year FE[f]	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE[g]	Yes	Yes	No	Yes	Yes	Yes
Robust[h]	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,514	30,083	23,822	53,054	14,082	53,613

[a]$P<.01$.

[b]$P<.1$.

[c]Controls represent a series of control variables.

[d]KP-LM: Kleibergen-Paap rk LM statistic, testing the relevance of IV.

[e]CD-Wald: Cragg-Donald Wald F statistic, testing the weak relevance of IV.

[f]Year FE: time fixed effects, controlling for unobserved shocks across different years.

[g]Individual FE: individual fixed effects, controlling for heterogeneity across individuals.

[h]Robust: the use of robust SEs, addressing heteroskedasticity effects.

Comparison of Urban and Rural Areas

Urbanization exacerbates the hollowing-out problem in rural areas, leading to higher levels of loneliness and more severe depression among rural older adults [33]. The cumulative disadvantage hypothesis suggests that this urban-rural depression disparity stems from the urban-rural divide, where rural residents accumulate disadvantages that peak in old age. The diffusion of the internet has varying effects on alleviating depression among older adults in urban and rural areas. Due to better infrastructure, urban older adults benefit more from internet access, facilitating easier social participation, information access, and entertainment, thereby effectively reducing feelings of loneliness and depression [34]. Moreover, proximity to children's residences and stable family support systems in urban areas, coupled with internet usage, further alleviates their depression. In contrast, rural older adults face limited internet access due to inadequate infrastructure, resulting in fewer opportunities to use internet benefits for social interaction and support. Combined with a lack of effective family support systems and communication channels, the internet's impact in alleviating depression among rural older adults is less pronounced compared with their urban counterparts. Therefore, it is expected that the alleviating effect of internet usage on depression would be more significant in urban areas. The grouped regressions in columns 1 and 2 confirm this hypothesis, indicating that the alleviating effect of internet usage is higher in urban areas (–1.74%) compared with rural areas (–0.774%).

Comparison of Eastern and Noneastern Regions

In eastern China, older adults have more opportunities to engage in social activities, access health information, and participate in online entertainment through the internet, resulting in a more pronounced alleviation of depression [35]. Furthermore, the higher economic development in eastern regions generally correlates with higher levels of education and economic capability among older adults, facilitating easier access to and utilization of the internet for psychological well-being [36,37]. In contrast, the older adults in noneastern regions experience lower internet usage rates and breadth due to limited internet resources, technological barriers, and economic constraints, leading to less significant alleviation of depression compared with their counterparts in eastern regions [38]. Therefore, it is anticipated that the therapeutic effect of internet usage on depression in older adults is more pronounced in eastern China and remains limited in noneastern regions. According to the classification standards of the National Bureau of Statistics, eastern China includes the provinces (and cities) of Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan. The grouped regressions in columns 3 and 4 confirm this hypothesis, indicating that the alleviating effect of internet usage is higher in eastern regions (–1.85%) compared with noneastern regions (0.00196, which is statistically insignificant).

Comparison of Broadband China Pilot and Nonpilot Areas

Since its implementation in 2013, the "Broadband China" strategy has significantly enhanced broadband network coverage and accessibility through advanced network technologies and infrastructure development. In pilot areas, older adults have more frequently used the internet for social interactions and access to health information, resulting in a notable reduction in depression. In contrast, nonpilot areas, constrained by limited internet resources and technical support, offer fewer

XSL•FO
RenderX

opportunities for internet usage by older adults, leading to less significant alleviation of depression compared with pilot areas [39]. Moreover, the widespread availability of broadband networks has facilitated increased older adults' participation in social activities and access to more health services, further mitigating depression [40]. Therefore, this study anticipates that the therapeutic effect of internet usage is greater in the pilot areas of Broadband China and smaller in the nonpilot areas. The grouped regressions in columns 5 and 6 confirm this hypothesis, indicating that the alleviating effect of internet usage is higher in pilot areas (–1.9%) compared with nonpilot areas (–0.000707, which is statistically insignificant).

Discussion

This study's core finding is that internet usage significantly alleviates depression symptoms among older adults, with a reduction of 1.41%. This result suggests that internet usage alleviates depression through four main mechanisms: the social interaction effect, physical exercise effect, intergenerational contact effect, and educational enhancement effect. Specifically, internet usage facilitates social activities among older adults, helping to reduce loneliness and, in turn, alleviating depressive symptoms. In addition, the internet enables older individuals to more easily participate in physical activities, which improves both their physical and mental health. Furthermore, internet usage strengthens communication with children, enhancing intergenerational support and reducing depressive feelings. Finally, internet usage provides older individuals with access to more educational resources, which improves self-awareness and self-management, thereby reducing depression symptoms. In summary, this study provides empirical evidence on how internet usage alleviates depression in older adults through multiple pathways, highlighting the potential of the internet to improve mental health in older adults.

The findings of this study align with existing literature emphasizing the positive impact of internet usage on mental health in older adults. Internet usage reduces depression by alleviating social isolation [41] and enhancing intergenerational relationships [11]. In addition, it reduces loneliness through increased social contact [9,42] and improves psychosocial functions, including physical activity and education [43,44]. Our study expands on this by incorporating the mediating effects of physical activity and education in depression reduction. While excessive internet usage may lead to social media fatigue and worsen mental health [45], our findings suggest that appropriate internet usage fosters mental well-being, as indicated by Lam et al [8]. Furthermore, mobile health apps have been shown to reduce depression in other populations, supporting the efficacy of digital interventions [46]. Despite concerns about internet addiction [47], our results suggest that regulated internet usage can improve both mental and functional health, particularly in older adults [48]. By using a 2-stage IV approach, this study mitigates endogeneity issues, providing robust causal inferences consistent with previous research [49].

The study's findings suggest several policy measures to enhance the mental health of older adults via internet usage, based on the identified mechanisms and observed heterogeneity in the effects of digital engagement: (1) strengthen internet infrastructure in rural and underserved regions: while internet usage enhances mental health through social interaction, physical activity, and educational access, its impact is greater in urban areas. Policy makers should prioritize the expansion of internet infrastructure in rural and noneastern regions to ensure that older adults can fully benefit from digital engagement. (2) Initiate targeted digital literacy programs: considering the influence of digital literacy on the effectiveness of internet usage, targeted training programs for older adults, particularly in rural areas, should be prioritized. These initiatives should emphasize teaching older users to engage in online social activities, access health resources, and participate in educational content, which have proven effective in reducing depression. (3) Create age-appropriate digital mental health platforms: mental health platforms should be customized for different subgroups. In urban areas, platforms should prioritize enhancing social interaction and intergenerational communication, whereas, in rural areas, the emphasis should be on bridging the digital divide by providing simple and accessible health and social resources. (4) Promote the development of age-friendly digital applications: policymakers should facilitate the creation of digital tools that encourage physical activity and cognitive stimulation among older adults. These applications can address mechanisms identified in the study, such as exercise and education, and contribute to reducing depression among older users.

While this study provides valuable insights into the relationship between internet usage and depression among older adults, several limitations must be considered. First, the reliance on self-reported data for depression and internet usage may introduce biases, such as social desirability or recall bias. Future studies could incorporate objective measures, such as log data from internet providers or wearable devices, to improve accuracy. Second, although the 2-stage IV approach addresses endogeneity, the exogeneity of the chosen instrument, the NBDCPZ, may still be questioned, and alternative instruments or experimental designs, like randomized controlled trials, could strengthen causal inference. In addition, the heterogeneity of internet usage effects across subgroups (eg, by age, education, or income) was not fully explored and warrants further investigation. Finally, while this study focused on short-term effects, future research should examine the long-term impacts of internet usage on the mental health of older adults, including potential risks such as internet addiction or over-reliance on digital platforms.

Acknowledgments

This study was supported by the Doctoral Research Innovation Program of Shandong Second Medical University (grant 041174). The authors are thankful to the editors and reviewers for their comments and feedback.

XSL•FO
RenderX

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LG and YWL contributed to the conception or design of the study. KC and YZ contributed to data collection and data analysis. YL drafted the paper while WQY guided the revisions. Wenqiang Yin and Ying Liu contributed equally to this work and should be considered co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weighted regression. Robust standard errors in parentheses.
[DOCX File , 18 KB - jmir_v27i1e65399_app1.docx]

References

1. Depressive Disorder (depression). URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-03-31]
2. Digital Health as an Enabler of Healthy Aging in Southeast Asia. URL: https://www.duke-nus.edu.sg/docs/librariesprovider5/publications/2020_p003_core_-digital-health-as-an-enabler-of-healthy-aging-in-southeast-asia.pdf?sfvrsn=ba1206c1_0 [accessed 2021-09-28]
3. Qiao YJ, Li ZW. Community social capital, family support, confucian pro-setback coping and geriatric depression: evidence from China. Hum Soc Sci Lett 2022;10:64-74. [doi: 10.18488/73.v10i1.2934]
4. Ge YF, Wang LJ, Feng WM, Zhang BZ, Liu SL, Ke YH. The challenge and strategy selection of healthy aging in China. Journal of Management World 2020;36(04):86-96. [doi: 10.19744/j.cnki.11-1235/f.2020.0055]
5. Greenlee K, Hyde P. Suicide and depression in older adults: Greater awareness can prevent tragedy. Generations 2014;38(3):23-26.
6. Hussenoeder FS, Jentzsch D, Matschinger H, Hinz A, Kilian R, Riedel-Heller SG, et al. Depression and quality of life in old age: a closer look. Eur J Ageing 2021;18(1):75-83 [FREE Full text] [doi: 10.1007/s10433-020-00573-8] [Medline: 33746683]
7. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
8. Lam SSM, Jivraj S, Scholes S. Exploring the relationship between internet use and mental health among older adults in England: longitudinal observational study. J Med Internet Res 2020;22(7):e15683 [FREE Full text] [doi: 10.2196/15683] [Medline: 32718913]
9. Escobar-Viera CG, Shensa A, Bowman ND, Sidani JE, Knight J, James AE, et al. Passive and active social media use and depressive symptoms among United States adults. Cyberpsychol Behav Soc Netw 2018;21(7):437-443. [doi: 10.1089/cyber.2017.0668] [Medline: 29995530]
10. Jeon GS, Choi KW, Jang KS. Social networking site usage and its impact on depressive symptoms among older men and women in South Korea. Int J Environ Res Public Health 2020;17(8):2670 [FREE Full text] [doi: 10.3390/ijerph17082670] [Medline: 32295024]
11. Minagawa Y, Saito Y. An analysis of the impact of cell phone use on depressive symptoms among Japanese elders. Gerontology 2014;60(6):539-547. [doi: 10.1159/000363059] [Medline: 24968927]
12. Internet usage in Southeast Asia - statistics & facts. URL: https://www.statista.com/topics/9093/internet-usage-in-southeast-asia/ [accessed 2024-08-26]
13. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. J Health Econ 2008;27(3):531-543 [FREE Full text] [doi: 10.1016/j.jhealeco.2007.09.009] [Medline: 18192044]
14. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
15. Wallinheimo AS, Evans SL. Patterns of internet use, and associations with loneliness, amongst middle-aged and older adults during the COVID-19 pandemic. Healthcare (Basel) 2022;10(7):1179 [FREE Full text] [doi: 10.3390/healthcare10071179] [Medline: 35885706]
16. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
17. Nash S. The Pandemic has Accelerated the Need to Close the Digital Divide for Older Adults. URL: https://longevity.stanford.edu/the-pandemic-has-accelerated-the-need-to-close-the-digital-divide-for-older-adults/ [accessed 2024-12-19]

XSL•FO
RenderX

18. Iwon K, Skibinska J, Jasielska D, Kalwarczyk S. Elevating subjective well-being through physical exercises: an intervention study. Front Psychol 2021;12:702678 [FREE Full text] [doi: 10.3389/fpsyg.2021.702678] [Medline: 34975608]
19. Singh B, Olds T, Curtis R, Dumuid D, Virgara R, Watson A, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. Br J Sports Med 2023;57(18):1203-1209 [FREE Full text] [doi: 10.1136/bjsports-2022-106195] [Medline: 36796860]
20. Gonnord T, Clarys D, Boucard G, Esnard C. Positive impact of social relationships fostered by physical and/or cognitive group activity on older people's quality of life: PRISMA systematic review. Front Psychol 2023;14:1166072 [FREE Full text] [doi: 10.3389/fpsyg.2023.1166072] [Medline: 37767210]
21. Cheng XF, Jiang QB. Social isolation and self-rated health of the older people: the mediation effects of aging attitude. Population and Development 2021;27(06):106-116+150.
22. Huang F, Fu P. Intergenerational support and subjective wellbeing among oldest-old in China: the moderating role of economic status. BMC Geriatr 2021;21(1):252 [FREE Full text] [doi: 10.1186/s12877-021-02204-y] [Medline: 33858343]
23. Zheng R, Yu M, Huang L, Wang F, Gao B, Fu D, et al. Effect of intergenerational exchange patterns and intergenerational relationship quality on depressive symptoms in the elderly: an empirical study on CHARLS data. Front Public Health 2022;10:1009781 [FREE Full text] [doi: 10.3389/fpubh.2022.1009781] [Medline: 36262237]
24. Dobson KS, Quigley L, Dozois DJ. Toward an integration of interpersonal risk models of depression and cognitive-behaviour therapy. Australian Psychologist 2020;49(6):328-336. [doi: 10.1111/ap.12079]
25. Hasan N. Joyful learning: a step towards the positive classroom. International Journal of Education and Management Studies 2016;6(1):133.
26. Hidi S, Harackiewicz JM. Motivating the academically unmotivated: a critical issue for the 21st century. Review of Educational Research 2000;70(2):151-179. [doi: 10.3102/00346543070002151]
27. Burnes D, Sheppard C, Henderson CR, Wassel M, Cope R, Barber C, et al. Interventions to reduce ageism against older adults: a systematic review and meta-analysis. Am J Public Health 2019;109(8):e1-e9. [doi: 10.2105/AJPH.2019.305123] [Medline: 31219720]
28. Matud MP, Bethencourth JM, Ibáñez I, Fortes D. Gender and psychological well-being in older adults. Int Psychogeriatr 2020;32(11):1293-1302. [doi: 10.1017/S1041610220000824] [Medline: 32456745]
29. Lee C, Glei DA, Goldman N, Weinstein M. Children's education and parents' trajectories of depressive symptoms. J Health Soc Behav 2017;58(1):86-101 [FREE Full text] [doi: 10.1177/0022146517690200] [Medline: 28661765]
30. Serra L, Silva L. Cognitive functioning and life satisfaction as predictors of subjective health complaints in elderly people. QAOA 2023;25(1):83-91. [doi: 10.1108/QAOA-08-2023-0055]
31. Negative and Missing R-Squared for 2SLS/IV. URL: https://www.stata.com/support/faqs/statistics/two-stage-least-squares/ [accessed 2024-12-19]
32. Li Z, Luo Y. A Study of the relationship and mechanism between internet use and health of the migrant elderly. Journal of Yunnan Minzu University (Philosophy and Social Sciences Edition) 2024;41(01):95-109. [doi: 10.13727/j.cnki.53-1191/c.20240004.015]
33. Purtle J, Nelson KL, Yang Y, Langellier B, Stankov I, Diez Roux AV. Urban-rural differences in older adult depression: a systematic review and meta-analysis of comparative studies. Am J Prev Med 2019;56(4):603-613. [doi: 10.1016/j.amepre.2018.11.008] [Medline: 30777704]
34. Nan Y, Xie Y, Hu Y. Internet use and depression among Chinese older adults: the mediating effect of interpersonal relationship. Front Public Health 2023;11:1102773 [FREE Full text] [doi: 10.3389/fpubh.2023.1102773] [Medline: 36935716]
35. Yang HL, Zhang S, Zhang SQ, Xie L, Wu YY, Yao YD, et al. Internet use and depressive symptoms among older adults in China. Front Psychiatry 2021;12:739085 [FREE Full text] [doi: 10.3389/fpsyt.2021.739085] [Medline: 34950065]
36. Xu Y, Huang Y. Chinese middle-aged and older adults' internet use and happiness: the mediating roles of loneliness and social engagement. J Appl Gerontol 2021;40(12):1846-1855. [doi: 10.1177/0733464820959168] [Medline: 32954905]
37. Pinquart M, Sörensen S. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. Psychol Aging 2000;15(2):187-224. [doi: 10.1037//0882-7974.15.2.187] [Medline: 10879576]
38. Jiang S, Zhang W, Yang T, Wu D, Yu L, Cottrell RR. Regional internet access and mental stress among university students: a representative nationwide study of China. Front Public Health 2022;10:845978 [FREE Full text] [doi: 10.3389/fpubh.2022.845978] [Medline: 35462833]
39. Aggarwal B, Xiong Q, Schroeder-Butterfill E. Impact of the use of the internet on quality of life in older adults: review of literature. Prim Health Care Res Dev 2020;21:e55 [FREE Full text] [doi: 10.1017/S1463423620000584] [Medline: 33263273]
40. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband internet access is a social determinant of health!. Am J Public Health 2020;110(8):1123-1125. [doi: 10.2105/AJPH.2020.305784] [Medline: 32639914]
41. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
42. Yu K, Wu S, Chi I. Internet use and loneliness of older adults over time: the mediating effect of social contact. J Gerontol B Psychol Sci Soc Sci 2021;76(3):541-550. [doi: 10.1093/geronb/gbaa004] [Medline: 31942629]

XSL•FO
RenderX

43. Mellor D, Firth L, Moore K. Can the internet improve the well-being of the elderly? Ageing Int 2008;32(1):25-42. [doi: 10.1007/s12126-008-9006-3]
44. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
45. Dhir A, Yossatorn Y, Kaur P, Chen S. Online social media fatigue and psychological wellbeing—a study of compulsive use, fear of missing out, fatigue, anxiety and depression. International Journal of Information Management 2018;40(4):141-152 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2018.01.012]
46. Rahayu TB, Supriyanti S, Kasim J, Munazar M, Vandika AY. Use of mobile applications for monitoring adolescent mental health. J. of World Future Medicine, Health and Nursing 2024;2(2):264-275. [doi: 10.70177/health.v2i2.761]
47. Višnjić A, Kök K, Terzić-Šupić Z, Stanković M. Editorial: excessive internet use and its impact on mental health. Front Public Health 2024;12:1473656 [FREE Full text] [doi: 10.3389/fpubh.2024.1473656] [Medline: 39206010]
48. Tajika A, Nakagomi A, Miyaguni Y, Koga C, Kondo K, Ojima T. Internet use and higher-level functional capacity decline suppression in Japanese older adults with low education: JAGES 2016-2019 longitudinal study. JMIR Aging 2024;7:e53384 [FREE Full text] [doi: 10.2196/53384] [Medline: 39303276]
49. Zhang J, Yang Y. Can environmental disclosure improve price efficiency? The perspective of price delay. Finance Research Letters 2023;52:103556. [doi: 10.1016/j.frl.2022.103556]

Abbreviations

CESD-10: Center for Epidemiologic Studies Depression Scale
CHARLS: China Health and Retirement Longitudinal Study
FE: fixed effects
IV: instrumental variable
NBDCPZ: National Big Data Comprehensive Pilot Zone
OLS: ordinary least squares

Edited by A Mavragani; submitted 14.08.24; peer-reviewed by Y Yang, C Zhao, V Rashedi; comments to author 04.11.24; revised version received 09.11.24; accepted 05.12.24; published 31.01.25.

Please cite as:
Guo L, Li Y, Cheng K, Zhao Y, Yin W, Liu Y
Impact of Internet Usage on Depression Among Older Adults: Comprehensive Study
J Med Internet Res 2025;27:e65399
URL: https://www.jmir.org/2025/1/e65399
doi:10.2196/65399
PMID:

©Lin Guo, Yunwei Li, Kai Cheng, Ying Zhao, Wenqiang Yin, Ying Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Measuring Digital Health Literacy in Older Adults: Development and Validation Study

SungMin Kim[1], MA; Choonghee Park[1], MA; Sunghyeon Park[1], MA; Dai-Jin Kim[2], MD, PhD; Ye-Seul Bae[3,4], MD, PhD; Jae-Heon Kang[4], MD, PhD; Ji-Won Chun[1,5], PhD

[1]Department of Medical Informatics, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Republic of Korea

[2]Department of Psychiatry, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Republic of Korea

[3]Big Data Research Center, School of Medicine, Kangbuk Samsung Hospital Sungkyunkwan University, Seoul, Republic of Korea

[4]Department of Family Medicine, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[5]The Catholic Medical Center Institute for Basic Medical Science, The Catholic Medical Center of The Catholic University of Korea, Seoul, Republic of Korea

Corresponding Author:
Ji-Won Chun, PhD
Department of Medical Informatics
Seoul St. Mary's Hospital
The Catholic University of Korea College of Medicine
222 Banpo-daero, Seocho-gu
Seoul, 06591
Republic of Korea
Phone: 82 2 3147 9727
Email: jwchun@catholic.ac.kr

Abstract

Background: New health care services such as smart health care and digital therapeutics have greatly expanded. To effectively use these services, digital health literacy skills, involving the use of digital devices to explore and understand health information, are important. Older adults, requiring consistent health management highlight the need for enhanced digital health literacy skills. To address this issue, it is imperative to develop methods to assess older adults' digital health literacy levels.

Objective: This study aimed to develop a tool to measure digital health literacy. To this end, it reviewed existing literature to identify the components of digital health literacy, drafted preliminary items, and developed a scale using a representative sample.

Methods: We conducted a primary survey targeting 600 adults aged 55-75 years and performed an exploratory factor analysis on 74 preliminary items. Items with low factor loadings were removed, and their contents were modified to enhance their validity. Then, we conducted a secondary survey with 400 participants to perform exploratory and confirmatory factor analyses.

Results: A digital health literacy scale consisting of 25 items was developed, comprising 4 subfactors: use of digital devices, understanding health information, use and decision regarding health information, and use intention. The model fit indices indicated excellent structural validity (Tucker-Lewis Index=0.924, comparative fit index=0.916, root-mean-square error of approximation=0.088, standardized root-mean-square residual=0.044). High convergent validity (average variance extracted>0.5) and reliability (composite reliability>0.7) were observed within each factor. Discriminant validity was also confirmed as the square root of the average variance extracted was greater than the correlation coefficients between the factors. This scale demonstrates high reliability and excellent structural validity.

Conclusions: This study is a significant first step toward enhancing digital health literacy among older adults by developing an appropriate tool for measuring digital health literacy. We expect this study to contribute to the future provision of tailored education and treatment based on individual literacy levels.

(J Med Internet Res 2025;27:e65492) doi:10.2196/65492

KEYWORDS

digital health care; older adults; digital health literacy; exploratory factor analysis; confirmatory factor analysis; mobile phone

XSL•FO
RenderX

Introduction

Background

The proliferation of digital technology and the widespread use of smart devices have facilitated convenient and ubiquitous access to current information and services. The worldwide adoption rate of smartphones is approximately 76%, with South Korea having the highest rate at 95% [1]. The diverse services in the digital realm can immensely benefit individuals using them [2]. Cutting-edge medical services, such as digital therapeutic gadgets, intelligent monitoring, and telemedicine are intricately connected to the digital revolution in health care [3]. The medical device sector, in an attempt to actively identify and meet patient needs, develops digital health care services at medical treatment sites. These digital health care services improve the accessibility of medical care and individualized therapy [4]. For instance, the use of smartphone applications for medication reminders and the implementation of disease monitoring systems that rely on patient-generated health data prove highly beneficial for patients with chronic illnesses [5]. To effectively use digital health care services, patients and those without underlying medical conditions must possess the necessary skills to retrieve and use the desired health information in the digital domain [6,7].

Digital health literacy is a newly developed term that combines 2 distinct concepts: health and digital literacy [8]. Beyond comprehending health information, it includes the ability to efficiently use digital tools and resources for health care. However, the adoption of new technologies varies with the social environment and individual abilities. These disparities contribute to a digital divide and information gap. In South Korea, digitally marginalized groups, including older adults, rural areas residents, individuals with disabilities, and low-income individuals, are unable to keep up with advancements in digital technology [9], with studies recognizing the digital divide as a noteworthy societal problem among these populations [10-12]. The most vulnerable group is the older adult population, who have limited exposure to smart devices and limited opportunities for education in information and communication technology [13]. The COVID-19 pandemic has exacerbated this phenomenon [14].

By 2025, South Korea's population aged 65 years and older is projected to surpass 20% of the total population [15]. Over the next decade, population aging is expected to worsen, with the proportion of older adults steadily increasing [16]. Consequently, South Korea and other countries with aging populations have recognized digital exclusion among older adults as a new social issue. Older adults experience difficulties using everyday technologies such as kiosks and self-checkout counters and medical services such as digital therapeutic devices and telemedicine [17]. Although social support can motivate older adults to use digital technologies [18], those living alone may have limited access to digital health care services owing to a lack of assistance from family members. Therefore, it is crucial to develop a strategy that enhances the digital health literacy level among those lacking access to digital services. Enhancing digital health literacy can offer significant health benefits for individuals [19]. Individuals using digital health services understand which information is most appropriate for their needs, while service providers can deliver optimum services to them [20]. For instance, wearable devices such as smartwatches can transmit real-time vital health signs data including temperature, heart rate, blood pressure, and stress levels, to health care providers, thereby enabling continuous monitoring of patients' conditions [21]. To improve digital health literacy, it is essential to assess the current level of digital health literacy among users.

Despite efforts to assess digital health literacy, there has been a deficiency of thorough, comprehensive research that specifically examines and quantifies its various dimensions [22]. According to the previous systematic review investigating the measurement properties of eHealth literacy instruments, the most widely used scale is the eHealth Literacy Scale [23]. This scale estimated the ability to locate, comprehend, evaluate, and use electronic health information to address health concerns [24]. However, this assessment was insufficient in evaluating digital health literacy comprehensively [25], as it was developed before the emergence of social media and mobile-based health care technologies and was reported to not reflect recent advancements in digital health technologies [23]. van der Vaart and Drossaert [26] developed the Digital Health Literacy Instrument to measure various competencies, emphasizing the ability to use internet information. However, their study's reliance on a highly educated sample might have overestimated digital skills, and the limited platforms used could affect generalizability. Karnoe et al [27] categorized digital health literacy into 7 components, 4 related to health literacy and 3 to digital literacy, and developed a corresponding measurement scale. However, this scale only evaluates the comprehension of health information and the use of digital technologies without considering social literacy elements such as information sharing. Also, the transactional eHealth literacy instrument was evaluated as the best instrument for comprehensive assessments of eHealth literacy; however, its applicability was limited to healthy individuals or patients younger than 40 years [23].

Furthermore, previous studies limited that could not represent certain populations such as older adults or low-income individuals. Age is often cited as a sociodemographic factor that can influence digital health literacy, with evidence generally showing that digital health literacy tends to decline with increasing age [28]. Particularly, the previous study on digital literacy assessment tools for older persons found that further research on the appropriateness of these instruments for this demographic is necessary [29]. This suggests that there may be differences in digital habits and learning potential between older and younger individuals dependent on needs and experiences with digital technology, highlighting the need for specialized measurement tools tailored to the older.

Objective

This study aimed to identify the fundamental components of digital health literacy among older adults and develop a systematic self-report scale that can be used to effectively assess their level of digital health literacy.

XSL•FO
RenderX

Methods

Overview

This study investigated the subjective experiences of digital health literacy among older adults, focusing on their ability to use digital technologies and skills in searching for, understanding, and using health information (Figure 1). We used 2 approaches to create an item pool: a literature review to identify digital health literacy components and focus interviews with health care experts to validate these items. Subsequently, we used a 2-step survey and analysis using exploratory and confirmatory factor analyses.

Figure 1. Factor structure of digital health literacy.

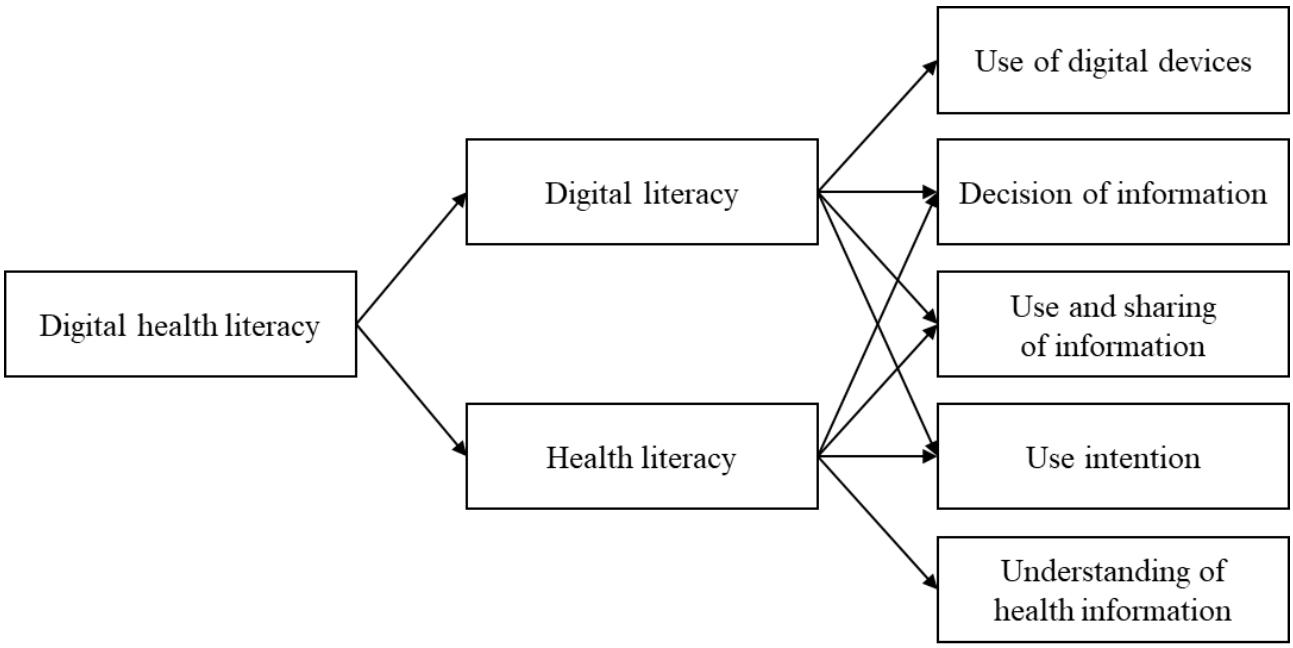

Development of the Digital Health Literacy Scale

We examined previous studies to develop an initial item pool based on pertinent hypotheses regarding motives, functionality, and experiences related to health care technology use among older adults. To achieve this, we searched multiple sources, including the National Library of Medicine PubMed database and Web of Science, using predefined keywords such as “Digital Health Literacy scale,” “Mobile health device,” “eHealth,” “Health Literacy,” and “Digital Literacy.” The inclusion criteria for selecting studies were as follows: (1) studies focusing on digital health literacy or related concepts, (2) studies providing quantitative evaluations of digital health literacy tools, and (3) studies where measurement scales were openly available in the publication or could be obtained upon request. Exclusion criteria included: (1) studies that did not explicitly assess digital health literacy, (2) studies without quantitative data on digital health literacy tools, and (3) studies where measurement scales were unavailable or inaccessible upon request. Two reviewers (SMK and JWC) independently extracted appropriate instruments, considering study design, target population, measured variables, and psychometric properties of the scales, and discrepancies were resolved through consensus. The final selection of items was determined through discussions among the research team, which included medical doctors, medical informatics experts, and psychologists.

The final item pool was informed by a detailed review of five key instruments: (1) the Mobile Device Proficiency Questionnaire (covering basics, communication, data storage, and internet use), (2) European Health Literacy Survey Questionnaire (assessing the ability to understand and process health information), (3) Unified Theory of Acceptance and Use of Technology (evaluating perceived usefulness, ease of use, and attitudes), (4) digital health literacy instrument (focusing on the reliability of health care information), and (5) Transactional eHealth Literacy Instrument (related to health information communication) [27,30-33]. In addition, by reviewing the literature on digital health literacy and older adults [23,29], we categorized validated and available scales, identifying the characteristics of each scale item as well as their limitations. Therefore, the scales used in this study were selected based on two key criteria. First, each scale has been validated for reliability and validity in previous research, ensuring their credibility as assessment tools. Second, the scales were chosen for their ability to comprehensively evaluate various aspects of digital health literacy, such as mobile device proficiency, health information processing, and attitudes toward technology acceptance, allowing for a well-rounded measurement.

Our second approach involved a focus group interview with 10 experts from nursing, social work, public health, education, and medicine. These experts regularly engage with older individuals in digitized medical environments, offering several forms of assistance, which understand the issues and demands typically encountered by seniors. Focus group interviews were conducted with participants in 2 groups of 5, each lasting 2 hours. The interview was audio-recorded, and transcribed immediately after the interview. The data was categorized into 3 overarching themes by the researchers, who independently extracted meaningful units from the transcripts. The initial theme, “Accessibility of Digital Health care Services,” addressed

XSL•FO
RenderX

concerns such as the ability to pay medical expenses through hospital applications and the necessity of locating hospital information during emergencies. The second theme, "Understanding and Utilizing Medical Information," encompassed the capacity to effectively interpret and use health check-up results, medication instructions, and nutritional information. The third theme, "Autonomy through Digital Health Technologies," underscored the ability to use health-related applications independently to access medical services or evaluate personal health without external assistance. Items derived from these themes were developed and refined through expert feedback to ensure their relevance and were subsequently included in the preliminary scale. The guidelines for focus group interviews are included in Multimedia Appendix 1. The initial item pool included 74 items: 59 from the literature review and 15 from focus group interviews (Multimedia Appendix 2).

Participants and Data Collection

The target population comprised Korean adults aged 55 years to 75 years. The target population comprised Korean adults aged 55 years to 75 years. Participants were recruited using a closed survey approach in the platform of the survey agency via emails and website visits. Quota sampling was used to ensure balanced representation by gender and age groups. Data collection was conducted in two phases: the first survey included 600 participants from August to September 2023, and the second survey included 400 participants from October to November 2023. Individuals who participated in the first survey were not eligible to participate in the second survey. To prevent duplicate responses, each participant accessed the survey through a unique URL. IP address checks and cookie-based mechanisms were implemented to ensure data quality and prevent multiple submissions from the same individual.

Data Analysis

Exploratory Factor Analysis

To conduct exploratory factor analysis (EFA), an absolute minimum of 100 participants is required, with 200 participants considered fair and 300 participants deemed good [34,35]. With 600 participants, we performed the first EFA to explore the scale's factor structure and selected items using IBM SPSS Statistics for Windows (version 25.0). Moreover, we conducted frequency analysis to verify demographic information and reliability analysis. Descriptive statistics are summarized as frequencies, percentages, or means (SD). The factor extraction method used was the maximum likelihood method with direct oblimin rotation to calculate factor loadings [36]. We present the number of factors to 5 based on the literature review, which was verified using parallel analysis [37]. To ensure the item structure's validity, we followed factor analysis guidelines, removing items with factor loadings below 0.4 and items with cross-loadings above 0.3 on 2 or more factors [38]. The adequacy of the factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity.

In addition, we conducted a second EFA with a sample of 200 randomly selected individuals from the entire pool of 400 participants in the second survey to validate the final 33 items in the first survey. A determining factor is the ratio of tested items to participants, with a ratio of 5 participants per item considered appropriate [35]. The EFA was conducted using the same method as in the first survey. As before, items with factor loadings below 0.4 and items with cross-loadings above 0.3 on 2 or more factors were removed to select the final items for the confirmatory factor analysis (CFA).

CFA to Validate the Scale's Structure

To validate the scale structure, which consisted of 25 items, we conducted a CFA with 200 participants after the second EFA using IBM SPSS Amos for Windows (version 25.0), and structural equation modeling. It is known that the minimum sample size for CFA is to have at least 200 participants [39] and a ratio of sample size to the number of model parameters is more than 5 in factor analysis [35]. Model fit was evaluated using the chi-square statistic, comparative fit index (CFI), Tucker-Lewis Index (TLI), standardized root-mean-square residual (SRMR), and root-mean-square error of approximation (RMSEA). A model with the ratio of the chi-square statistic to the respective df less than 3, CFI and TLI values greater than 0.90, and SRMR and RMSEA less than 0.08 was considered an acceptable fit. In addition, convergent validity was assessed using average variance extracted (AVE) values, ensuring that they were above 0.5, and construct reliability (CR) with values above 0.7. Discriminant validity was evaluated by comparing the square root of the AVE for each construct with the correlations between the constructs [40].

Ethical Considerations

This study received approval from the institutional review board of the Catholic University of Korea, Songeui Campus (approval number: MIRB 20230825-007). The study adhered to institutional and national ethical guidelines for research involving human participants. Before participation, all participants received detailed information about the study's purpose, procedures, anonymous, data usage, and their rights. Participation was voluntary, and participants could withdraw at any point. Data protection measures were implemented, including the anonymization of all collected data to prevent unauthorized access. No personal identifiers such as names or contact information were collected. Only completed surveys were included in the final analysis. The survey was web-based and conducted through the online platform of the survey agency (Macromill Embrain Co. Ltd.). Participants who completed the survey received a reward of 500 points (approximately US $0.35) redeemable within the online platform.

Results

Participants

The mean age of the sample was 63.63 (SD 5.26) years, and 50% (500/1000) were women. Further, 90.8% (908/1,000) were married, and 4.0% (40/1,000) were single. Regarding the presence of chronic diseases, 48.2% (428/1,000) had been diagnosed with at least one of hypertension, diabetes, or hyperlipidemia, while 51.8% (518/1,000) had not been diagnosed with any chronic disease. Regarding subjective economic status, 51.5% (515/1,000) and 48.5% (485/1,000)

XSL•FO
RenderX

indicated that their economic position was above average and below average, respectively. Regarding subjective health status, 87.5% (875/1,000) and 12.5% (125/1,000) reported their health to be above average and below average, respectively. Participants provided demographic data encompassing gender, age, marital status, chronic disease, subjective economic situation, and subjective health status, along with digital health literacy items (Table 1).

Table 1. The demographic characteristics (N=1000).

Characteristics	First survey (n=600), n (%)	Second survey (n=400), n (%)	Participants, n (%)
Sex			
Male	300 (50)	200 (50)	500 (50)
Female	300 (50)	200 (50)	500 (50)
Age (years)			
55-65	300 (50)	200 (50)	500 (50)
66-74	300 (50)	200 (50)	500 (50)
Marital status			
Married	557 (92.8)	351 (87.8)	908 (90.8)
Not married	17 (2.8)	23 (5.9)	40 (4)
Divorced	26 (4.4)	26 (6.5)	52 (5.2)
Chronic disease (hypertension, diabetes, or hyperlipidemia)			
Yes	276 (46)	206 (51.5)	482 (48.2)
No	324 (54)	194 (48.5)	518 (51.8)
Subjective economic level			
High	3 (0.5)	3 (0.8)	6 (0.6)
Upper middle	50 (8.3)	37 (9.3)	87 (8.7)
Middle	261 (43.5)	161 (40.3)	422 (42.2)
Low middle	233 (38.8)	166 (41.5)	399 (39.9)
Low	53 (8.8)	33 (8.3)	86 (8.6)
Subjective health status			
Very good	21 (3.5)	12 (1)	33 (3.3)
Good	233 (38.8)	127 (11.8)	360 (36)
Average	272 (45.3)	210 (52.5)	482 (48.2)
Not good	72 (12)	47 (31.8)	119 (11.9)
Not very good	2 (0.3)	4 (3)	6 (0.6)

First Validation Survey: EFA

Based on the literature review and focus group interviews (Multimedia Appendix 3), we assembled a pool of 74 items reflecting the 5 factors of older adults' digital health literacy and extracted the underlying structure of items using EFA. In the first survey sample (n=600), all participants were aged 55-75 years, and 50% (n=300) were women. The number of factors was determined using parallel analysis. The eigenvalues of the real data were more significant than those of the random data up to the fourth eigenvalue (real data eigenvalue=0.461; random data eigenvalue=0.334). However, from the fifth eigenvalue onwards, the eigenvalues of the real data were smaller than those of the random data (real data eigenvalue=0.160; random data eigenvalue=0.299). Following the examination of item discrimination and determination of factor loadings, 29 items were retained. The domains of health information decision and use and sharing were consolidated, while other aspects remained. The KMO value, representing the suitability of factor analysis, was 0.981, and Bartlett's test of sphericity yielded a significant result (χ^2_{406}=19365.0, P<.001), confirming the appropriateness of the model. The results also demonstrated a satisfactory level of internal consistency for each factor (Cronbach α values: 0.925-0.944). Table 2 lists the selected components' commonalities and factor loadings.

XSL•FO
RenderX

Table 2. Exploratory factor analysis results for survey 1.

Factors and Items	Communality	Standard Factor Loading
Understanding of Health Information		
Q42. I can understand the instructions for medication provided by the hospital app.	0.861	0.902
Q41. I can understand the emergency manual provided by health-related apps.	0.824	0.770
Q39. I can understand the health check-up results provided by the hospital app.	0.794	0.752
Q48. I can understand the terms of the privacy consent form when registering on the hospital app.	0.781	0.703
Q40. I can understand the nutritional information of food provided by health-related apps.	0.803	0.700
Q49. I am aware of the precautions for online payment when paying medical bills.	0.772	0.636
Q46. I can understand the information about health check-ups (such as target, date, price, fasting requirements, etc.) provided by the hospital app.	0.811	0.566
Use of Digital Devices		
Q7. I can save photos and texts about healthy activities found on the Internet.	0.765	0.856
Q9. I can use the desired services (payment, location search, etc.) through the hospital kiosk.	0.709	0.851
Q13. I can delete health-related apps that I have used.	0.737	0.839
Q8. I can book and confirm medical services through the hospital app.	0.745	0.801
Q6. I can find a suitable hospital for my symptoms using my smartphone.	0.723	0.792
Q3. I can find information about disease symptoms and treatments using my smartphone.	0.661	0.770
Q12. I can use the online “store” (e.g., Apple App Store or Google Play Store) on my device to find health-related apps.	0.711	0.764
Q15. I can use appropriate words or search terms to find the health service information I want on the Internet.	0.788	0.716
Q14. I can access hospital websites through an Internet search.	0.793	0.714
Q17. In an emergency, I can find information about nearby hospitals using my smartphone.	0.751	0.678
Use Intention		
Q32. I believe it is necessary to exchange health information online.	0.692	0.885
Q33. I am willing to use health-related apps to collect health information.	0.753	0827
Q30. I have a lot of interest in health-related apps.	0.678	0.790
Q35. I am interested in learning health knowledge or skills from the Internet.	0.758	0.739
Q29. I find the necessity and convenience of health management through health-related apps.	0.682	0.507
Q25. Using health-related apps improves my ability to manage my health.	0.580	0.431
Use and Decision of Health Information		
Q57. I can judge whether the health information found on my smartphone is trustworthy.	0.784	0.886
Q58. I can evaluate the pros and cons of various treatment methods provided by the hospital app.	0.771	0.785
Q61. I can determine the medical services I need.	0.734	0.707
Q59. I can judge how to use the health information provided by health-related apps.	0.753	0.705
Q51. I can determine if the health information found on the Internet is written for commercial purposes (advertisements).	0.706	0.516
Q64. I can use the health information provided by health-related apps for disease management.	0.774	0.449

In addition, we made minor adjustments to the working scale to align it with the conceptual framework of digital health literacy. First, the health information decision and use and sharing were combined to create a new factor called the use and decision of health information. Second, hospital apps were renamed health-related apps to encompass a broader range of applications. Finally, 4 new items related to digital health services were added. As a result, we created 33 items divided into 4 domains and their associated subscales (Multimedia Appendix 4).

Second Validation Survey: EFA

We randomly selected 200 responses (women 53.0%) from the 400 valid and unique responses in the second survey. Majority

XSL•FO
RenderX

of the respondents were aged 55-65 years (51.5%). Of the 200 participants, 86.5% (173/200) were married, 4% (8/200) unmarried, and 9.5% (19/200) divorced. The subjective economic level was categorized as high (2/200, 1%), upper middle (21/200, 10.5%), low middle (85/200, 42.5%), or low (16/200, 8.0%). Subjective health status was categorized into four groups: (1) high (4/200, 2.0%), (2) upper middle (62/200, 31.0%), (3) low middle (21/200, 10.5%), and (4) low (1/200, 0.5%).

The results of the secondary EFA also supported the feasibility of performing factor analysis (KMO 0.945; Bartlett test of sphericity: χ^2_{300}=4664.2, P<.001). Parallel analysis indicated that the 4 factors were consistent with the primary study. The fifth eigenvalue for the real data was 0.207, while that for the random data was 0.320. The internal reliability of each component and subscale was satisfactory, as indicated by the Cronbach α values for the 4 subscales falling between 0.91 and 0.96, which is above the minimum threshold of 0.7. The first factor (use of digital devices) comprised 10 items and had a Cronbach α value of 0.96. The second factor (understanding health information) had 5 items and exhibited a reliability coefficient of .92. The third (use and decision of health information) and fourth factors (use intention) consisted of 5 items and demonstrated high reliability, with coefficients of 0.91 and 0.93, respectively. The 25-item scale demonstrated a high level of dependability with a value of 0.96. Almost all items had factor loadings above 0.6 (Table 3). Overall, these results suggest that the 4 factors extracted through EFA were appropriate.

XSL•FO
RenderX

Table 3. Exploratory factor analysis results for survey 2.

Factors and items	Communality	Standard factor loading
Use of digital devices		
Q15. I can access the hospital's website through an internet search.	0.757	0.911
Q30. I can register as a member on the hospital website or app.	0.697	0.841
Q11. I can delete health-related apps that I have used.	0.636	0.828
Q17. In an emergency, I can find information about nearby hospitals using my smartphone.	0.762	0.788
Q33. I can make online payments for medical bills through the hospital website or app.	0.745	0.774
Q12. I can use the online "store" (eg, Play Store or App Store) on my device to find health-related apps.	0.724	0.774
Q16. I can use appropriate words or search terms to find the health service information I want on the internet.	0.697	0.741
Q8. I can use the desired services (payment, location search, etc.) through the hospital kiosk.	0.628	0.739
Q9. I can book and confirm medical services through the hospital app.	0.716	0.696
Q31. I can find my test results or prescription details on the hospital website or app.	0.633	0.590
Understanding of health information		
Q1. I can understand the instructions for medication provided by health-related apps.	0.775	0.903
Q3. I can understand the emergency manual provided by health-related apps.	0.737	0.775
Q5. I can understand the nutritional information of food provided by health-related apps.	0.629	0.767
Q7. I can understand the information about health check-ups (such as target, date, price, fasting requirements, etc.) provided by health-related apps.	0.765	0.756
Q4. I can understand the health check-up results provided by health-related apps.	0.594	0.636
Use and decision of health information		
Q24. I can judge whether the health information found on my smartphone is trustworthy.	0.790	0.836
Q25. I can evaluate the pros and cons of various treatment methods provided by health-related apps.	0.719	0771
Q29. I can determine if the health information found on the Internet is written for commercial purposes (advertisements).	0.607	0.681
Q26. I can judge how to use the health information provided by health-related apps.	0.700	0.609
Q27. I can determine the medical services I need.	0.611	0.570
Use intention		
Q18. I believe it is necessary to exchange health information online.	0.794	0.956
Q21. I have a lot of interest in health-related apps.	0.812	0.885
Q19. I am willing to use health-related apps to collect health information.	0.795	0.859
Q22. I find the necessity and convenience of health management through health-related apps.	0.734	0.761
Q23. Using health-related apps improves my ability to manage my health.	0.638	0.619

CFA

We performed CFA on a sample of 200 responses (women 47.0%) that were not included in the previous EFA from the 400 valid and unique responses in the second survey. Majority of respondents were aged 66-74 years, comprising 103 of 200 (51.5%) respondents. Of the 200 participants, 89% (178/200) were married, 7.5% (15/200) unmarried, and 3.5% (7/200) divorced. Subjective economic level was classified into four categories: (1) high (1/200, 0.5%), (2) upper middle (16/200, 8.0%), (3) low middle (81/200, 40.5%), and (4) low (17/200, 8.5%). Subjective health condition was classified into four categories: (1) high (8/200, 4.0%), (2) upper middle (65/200, 32.5%), (3) low middle (26/200, 13.0%), and (4) low (3/200, 1.5%). Following the outcomes of the previous EFA, 4 distinct factors were identified, and all 25 items were associated with each component. The model demonstrated strong fit across key indices. The CFI was 0.924, and the TLI was 0.916, both surpassing the commonly accepted threshold of 0.90, indicating a good fit. The SRMR was 0.044, reinforcing the adequacy of the model. The RMSEA was 0.088, with a 90% CI between 0.080 and 0.096, falling within acceptable limits. These results

XSL•FO
RenderX

confirm the robustness of the 4-factor structure and its alignment with the theoretical framework.

The factor loadings per item in Table 4 were all greater than 0.7, indicating a strong relationship between the items and their respective factors. The AVE values were higher than 0.7 and CR values ranged from 0.93 to 0.96. The high AVE and CR values of this scale indicate that it adequately explains the latent construct it aims to measure and demonstrates high internal consistency. Therefore, the results support the convergent validity. Discriminant validity was also confirmed, as the square roots of the AVE values for each factor were greater than the correlations between the factors. Figure 2 displays the factor structure and standardized factor loadings obtained from the CFA, which demonstrated the validity of the scale developed in this study. The final scale comprised 4 factors and 25 items (Multimedia Appendix 5). Participants assign 4 points for strong agreement and 0 points for severe disagreement to each item. The digital device use component consists of 10 items, contributing up to 40 points. The remaining three components which are (1) understanding health information, (2) use and decision-making of health information, and (3) use intention, each consist of 5 items, contributing up to 20 points per component.

XSL•FO
RenderX

Table 4. Confirmatory factor analysis results for survey 2.

Factors and items	Standard factor loading	Squared multiple correlations	AVE[a]	CR[b]
Use of digital devices			0.734	0.965
Q16. I can access the hospital's website through an Internet search.	0.886	0.78		
Q30. I can register as a member on the hospital website or app.	0.856	0.73		
Q10. I can delete health-related apps that I have used.	0.897	0.80		
Q17. In an emergency, I can find information about nearby hospitals using my smartphone.	0.844	0.71		
Q33. I can make online payments for medical bills through the hospital website or app.	0.837	0.70		
Q14. I can use the online "store" (eg, Play Store or App Store) on my device to find health-related apps.	0.916	0.84		
Q15. I can use appropriate words or search terms to find the health service information I want on the internet.	0.870	0.76		
Q9. I can use the desired services (payment, location search, etc.) through the hospital kiosk.	0.811	0.66		
Q11. I can book and confirm medical services through the hospital app.	0.827	0.68		
Q31. I can find my test results or prescription details on the hospital website or app.	0.814	0.66		
Understanding of health information			0.787	0.949
Q1. I can understand the instructions for medication provided by health-related apps.	0.896	0.80		
Q2. I can understand the emergency manual provided by health-related apps.	0.901	0.81		
Q5. I can understand the nutritional information of food provided by health-related apps.	0.886	0.78		
Q7. I can understand the information about health check-ups (such as target, date, price, fasting requirements, etc.) provided by health-related apps.	0.876	0.77		
Q3. I can understand the health check-up results provided by health-related apps.	0.875	0.77		
Use and decision of health information			0.776	0.945
Q24. I can judge whether the health information found on my smartphone is trustworthy.	0.896	0.80		
Q25. I can evaluate the pros and cons of various treatment methods provided by health-related apps.	0.883	0.78		
Q28. I can determine if the health information found on the internet is written for commercial purposes (advertisements).	0.852	0.73		
Q27. I can judge how to use the health information provided by health-related apps.	0.894	0.80		
Q26. I can determine the medical services I need.	0.878	0.77		
Use Intention			0.719	0.927
Q18. I believe it is necessary to exchange health information online.	0.805	0.65		
Q20. I have a lot of interest in health-related apps.	0.803	0.64		
Q19. I am willing to use health-related apps to collect health information.	0.783	0.61		
Q22. I find the necessity and convenience of health management through health-related apps.	0.936	0.88		
Q23. Using health-related apps improves my ability to manage my health.	0.901	0.81		

[a]AVE: average variance extracted.

[b]CR: construct reliability.

XSL•FO
RenderX

Figure 2. Factor structure with standardized loading for Korean version of the digital health literacy Scale. UDD: use of digital devices; UHI: understanding health information; UDHI: use and decision of health information; UI: use intention.

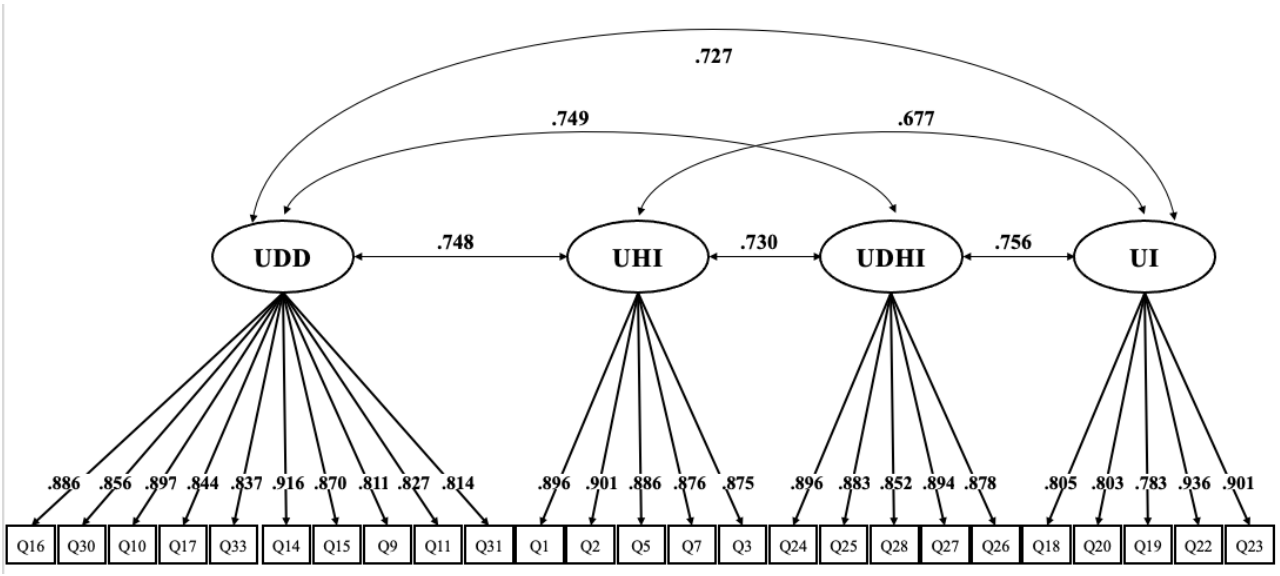

Discussion

Principal Results

This study aimed to develop and validate a scale for measuring digital health literacy. The only validated digital health literacy scale previously translated and validated in Korea was the eHealth Literacy Scale adapted by Chang et al [41]. Despite its extensive translation and use in many countries [42], this scale does not keep pace with current technological development. We developed a Korean version of the Digital Health Literacy Scale designed specifically for older adults who are vulnerable to the digital information divide. This scale has the distinctive feature of being specifically designed for older adults, who are an important but often overlooked population in digital health literacy research. This scale comprises 25 items across 4 main factors: use of digital devices, processing of health information, use and reliability assessment of health information, and evaluation of usage intentions. While other scales primarily focus on the general population or specific health conditions, our scale addresses the unique challenges faced by older adults, such as lower familiarity with technology and specific health information needs. The Korean version of the digital health literacy scale considers various factors, making it a valuable tool for comprehensively assessing the digital health literacy level of older adults. Furthermore, it can serve as an essential foundational resource for developing digital health literacy education programs and applying patient-tailored digital therapeutic devices for older adults.

Use of Digital Devices evaluates the ability to search for and use health information using digital devices such as smartphones, tablets, and computers. Efficient digital device usage is essential for searching and using health information, making it a fundamental element of digital health literacy. Studies have shown that older adults have lower skill levels in using the internet to search for health information compared to younger people [43], with a lack of digital device proficiency cited as a primary cause. Therefore, assessing the use of digital devices on this scale can help develop tailored strategies for enhancing older adults' digital health literacy capabilities.

Understanding of Health Information measures the ability to understand and practically apply health information. Although many applications are available to assist individuals in health management, if users fail to comprehend basic health information, they will struggle to use these tools effectively, resulting in poor health management. Accurately measuring the ability to understand and apply health information based on this factor can serve as a starting point for improving digital health literacy.

Use and Decision of Health Information measures the ability to evaluate the reliability of health information obtained through digital devices and apply it in practice. This is a crucial aspect of digital health literacy, as it includes the ability to use reliable information to engage with personal health information. For instance, it includes the ability to determine whether health information found on the Internet is for commercial purposes. Initially, we hypothesized that "Decision of Health Information" and "Utilization and sharing of Health Information" would form separate factors. However, it revealed that combining these items into a single factor was a more suitable structure that improves the internal consistency of the factor and strengthens its interpretability. Consequently, the results indicated that older persons viewed behaviors associated with decision-making, using, and sharing health information as a cohesive behavioral pattern rather than as separate processes. Compared to younger people, older adults might find it more challenging to discern false information [44], which increases the risk of accepting incorrect health information. Given the vast amount of information available online, it is essential to identify the relevant information and that which contains evidence. Particularly, it is necessary to include items such as digital content creation and digital safety [29]. The questions measuring digital content creation and digital safety were included in the initial items (Multimedia Appendix 2). These items were excluded to enhance the reliability and validity of the scale, but

XSL•FO
RenderX

future studies should focus on developing and validating questions that can assess these aspects.

Finally, Use Intention measures an individual's intention to use digital health applications, that is the motivation and willingness to adopt and use new technologies, which are critical elements for enhancing an individual's digital health literacy level. One of the reasons older adults do not use digital devices is a negative attitude toward new technologies, such as a sense of resistance [45]. Consequently, older adults are slower to adopt new technologies, and even when educational programs are necessary, they may exhibit low learning motivation. For individuals with low intention to use digital health devices, it is critical to first raise awareness of the importance of digital health literacy and present convenient health services. This, in turn, will strengthen digital health literacy.

Digital health literacy significantly improves older adults' ability to manage their health and the effectiveness of digital health care services. This study highlights the importance of digital health literacy in South Korea's medical domain and provides fundamental data for addressing digital exclusion among older adults. Moreover, this study used a distinct approach to assess the digital health literacy of older adults using the 4 key factors, distinguishing it from previous studies. This approach contributes to a more systematic understanding of digital health literacy and a more precise assessment of older adults' actual literacy levels. By measuring overall digital device proficiency as well as the ability to understand and use health information, this scale can be effectively used in countries where smartphones are widely available. In the future, research is needed not only to validate this scale across diverse cultural contexts to enhance its generalizability but also to evaluate its predictive validity, thereby ensuring its utility in identifying individuals' properties with low digital health literacy.

Limitations

First, because data were collected through an online survey, it is possible that participation from older adults unfamiliar with digital devices was limited and resulted in an underestimation of the actual digital health literacy levels in groups with low literacy. Therefore, future studies should consider parallel offline surveys for more accurate data. Second, this study focused on older adults, who are the most vulnerable to digital exclusion [13], which may limit the findings' generalization to other age groups or populations with diverse backgrounds. To achieve broader generalizability, future research should aim to collect samples that include a wider range of ages and backgrounds. Third, we prioritized expert feedback during the item development stage to ensure content validity but did not directly include focus group interviews with older participants to gather insights into their digital device usage for health management. Future studies should consider conducting focus group interviews with older adults to further refine the scale and enhance its relevance to this population. Finally, due to the disparity in total scores among the subscales, caution should be observed while using the overall score.

Conclusions

In conclusion, this study developed and validated a novel scale to assess digital health literacy among older adults, addressing the unique challenges posed by the digital information divide in this population. The Digital Health Literacy Scale, comprising 25 items across four factors, provides a comprehensive framework for evaluating older adults' digital device proficiency, understanding of health information, decision-making, and usage intentions. By highlighting these dimensions, the scale facilitates tailored interventions to enhance digital health literacy, ultimately improving health management outcomes and the adoption of digital healthcare services. This study underscores the necessity of targeted approaches in promoting digital inclusion for older adults, especially in rapidly digitizing societies. Future research should extend this work by validating the scale across diverse cultural and demographic contexts and exploring its utility in predicting health-related behaviors and outcomes.

Acknowledgments

This research was supported by the Ministry of Science and ICT (Information and Communication Technology) of Korea and Korea Medical Device Development Fund (RS-2023-00255145), and the Basic Medical Science Facilitation Program through the Catholic Medical Center of the Catholic University of Korea funded by the Catholic Education Foundation.

Authors' Contributions

JWC and JHK contributed to the conception and design of the study. SMK and JWC wrote the manuscript text and prepared the figures and tables. SMK and CHP performed data analysis. CHP and SHP contributed to the data acquisition and extraction. YSB and JHK contributed to the deliberation of the topic and provided expertise and guidance for the study. DJK assisted with the interpretation of the data and critically revised the manuscript for important intellectual content. JWC was responsible for composing and coordinating the review process, incorporating the comments provided by coauthors. All the authors contributed to the manuscript and approved its final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Focus group interview guidelines for developing a digital health literacy scale.
[DOCX File , 25 KB - jmir_v27i1e65492_app1.docx]

Multimedia Appendix 2
Preliminary Items for Digital Health Literacy Scale. 5-point Likert-scale. 0–4 (Strongly Disagree-Strongly Agree).
[DOCX File , 32 KB - jmir_v27i1e65492_app2.docx]

Multimedia Appendix 3
PRISMA 2020 Checklist.
[DOCX File , 31 KB - jmir_v27i1e65492_app3.docx]

Multimedia Appendix 4
Digital Health Literacy Scale items after first exploratory factor analysis. 5point Likert-scale. 0-4(Strongly Disagree-Strongly Agree).
[DOCX File , 30 KB - jmir_v27i1e65492_app4.docx]

Multimedia Appendix 5
Digital Health Literacy Scale items after confirmatory factor analysis. 5-point Likert-scale. 0–4 (Strongly Disagree-Strongly Agree).
[DOCX File , 30 KB - jmir_v27i1e65492_app5.docx]

References

1. Poushter J, Bishop C, Chwe H. Social media use continues to rise in developing countries, but plateaus across developed ones. Pew Internet and American Life Project. Washington, DC; 2018. URL: https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/ [accessed 2024-03-05]
2. Tang D. What is digital transformation? EDPACS 2021;64(1):9-13. [doi: 10.1080/07366981.2020.1847813]
3. Agarwal RK, Gao GG, DesRoches CM, Jha AK. The digital transformation of healthcare: current status and the road ahead. Information Systems Research 2010 Dec;21(4):796-809. [doi: 10.1287/isre.1100.0327]
4. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. Journal of Business Research 2021;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
5. Milani RV, Bober RM, Lavie CJ. The role of technology in chronic disease care. Prog Cardiovasc Dis 2016;58(6):579-583. [doi: 10.1016/j.pcad.2016.01.001] [Medline: 26772623]
6. Aida A, Svensson T, Svensson AK, Chung U, Yamauchi T. eHealth delivery of educational content using selected visual methods to improve health literacy on lifestyle-related diseases: literature review. JMIR Mhealth Uhealth 2020;8(12):e18316 [FREE Full text] [doi: 10.2196/18316] [Medline: 33295296]
7. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
8. Huhta AM, Hirvonen N, Huotari ML. Health literacy in web-based health information environments: systematic review of concepts, definitions, and operationalization for measurement. J Med Internet Res 2018;20(12):e10273 [FREE Full text] [doi: 10.2196/10273] [Medline: 30567690]
9. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
10. 2022 The Report on the Digital Divide. National Information Society Agency. 2022. URL: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=81623&bcIdx=25353 [accessed 2023-12-12]
11. Anrijs S, Mariën I, De Marez L, Ponnet K. Excluded from essential internet services: examining associations between digital exclusion, socio-economic resources and internet resources. Technology in Society 2023;73:102211. [doi: 10.1016/j.techsoc.2023.102211]
12. Correa T, Pavez I. Digital inclusion in rural areas: a qualitative exploration of challenges faced by people from isolated communities. J Comput-Mediat Comm 2016;21(3):247-263. [doi: 10.1111/jcc4.12154]
13. Mubarak F, Suomi R. Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. Inquiry 2022;59:469580221096272 [FREE Full text] [doi: 10.1177/00469580221096272] [Medline: 35471138]
14. Seifert A, Cotten SR, Xie B. A double burden of exclusion? Digital and social exclusion of older adults in times of COVID-19. J Gerontol B Psychol Sci Soc Sci 2021;76(3):e99-e103 [FREE Full text] [doi: 10.1093/geronb/gbaa098] [Medline: 32672332]
15. Baek JY, Lee E, Jung HW, Jang IY. Geriatrics fact sheet in Korea 2021. Ann Geriatr Med Res 2021;25(2):65-71 [FREE Full text] [doi: 10.4235/agmr.21.0063] [Medline: 34187140]

XSL•FO
RenderX

16. Kim KW, Kim OS. Super aging in South Korea unstoppable but mitigatable: a sub-national scale population projection for best policy planning. Spat Demogr 2020;8(2):155-173 [FREE Full text] [doi: 10.1007/s40980-020-00061-8] [Medline: 34222615]
17. Dong Q, Liu T, Liu R, Yang H, Liu C. Effectiveness of digital health literacy interventions in older adults: single-arm meta-analysis. J Med Internet Res 2023;25:e48166 [FREE Full text] [doi: 10.2196/48166] [Medline: 37379077]
18. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media & Society 2014;18(2):313-331. [doi: 10.1177/1461444814538648]
19. Haleem A, Javaid M, Pratap Singh R, Suman R. Medical 4.0 technologies for healthcare: features, capabilities, and applications. Internet of Things and Cyber-Physical Systems 2022;2:12-30. [doi: 10.1016/j.iotcps.2022.04.001]
20. Arias López MDP, Ong BA, Borrat Frigola X, Fernández AL, Hicklent RS, Obeles AJT, et al. Digital literacy as a new determinant of health: a scoping review. PLOS Digit Health 2023;2(10):e0000279 [FREE Full text] [doi: 10.1371/journal.pdig.0000279] [Medline: 37824584]
21. Kim K, Shin S, Kim S, Lee E. The relation between eHealth literacy and health-related behaviors: systematic review and meta-analysis. J Med Internet Res 2023;25:e40778 [FREE Full text] [doi: 10.2196/40778] [Medline: 36716080]
22. Brørs G, Larsen MH, Hølvold LB, Wahl AK. eHealth literacy among hospital health care providers: a systematic review. BMC Health Serv Res 2023;23(1):1144 [FREE Full text] [doi: 10.1186/s12913-023-10103-8] [Medline: 37875882]
23. Lee J, Lee EH, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021;23(11):e30644 [FREE Full text] [doi: 10.2196/30644] [Medline: 34779781]
24. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
25. Xie L, Mo PKH. Comparison of eHealth literacy scale (eHEALS) and digital health literacy instrument (DHLI) in assessing electronic health literacy in Chinese older adults: a mixed-methods approach. Int J Environ Res Public Health 2023;20(4):3293 [FREE Full text] [doi: 10.3390/ijerph20043293] [Medline: 36833987]
26. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
27. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth literacy assessment toolkit. J Med Internet Res 2018;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
28. Estrela M, Semedo G, Roque F, Ferreira PL, Herdeiro MT. Sociodemographic determinants of digital health literacy: a systematic review and meta-analysis. Int J Med Inform 2023;177:105124 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105124] [Medline: 37329766]
29. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021;23(2):e26145 [FREE Full text] [doi: 10.2196/26145] [Medline: 33533727]
30. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. J Appl Gerontol 2018;37(2):131-156 [FREE Full text] [doi: 10.1177/0733464816642582] [Medline: 27255686]
31. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European health literacy survey questionnaire (HLS-EU-Q). BMC Public Health 2013;13:948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
32. Venkatesh, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
33. Paige SR, Stellefson M, Krieger JL, Miller MD, Cheong J, Anderson-Lewis C. Transactional eHealth literacy: developing and testing a multi-dimensional instrument. J Health Commun 2019;24(10):737-748 [FREE Full text] [doi: 10.1080/10810730.2019.1666940] [Medline: 31583963]
34. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychological Methods 1999;4(1):84-99. [doi: 10.1037/1082-989X.4.1.84]
35. Velicer WF, Fava JL. Affects of variable and subject sampling on factor pattern recovery. Psychological Methods 1998;3(2):231-251. [doi: 10.1037//1082-989x.3.2.231]
36. Gorsuch R. Exploratory factor analysis. In: Handbook of multivariate experimental psychology. Boston, MA: Springer; 1988:231-258.
37. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. Organizational Research Methods 2004;7(2):191-205. [doi: 10.1177/1094428104263675]
38. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation 2005;10(7):1-9. [doi: 10.4135/9781412995627.d8]
39. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. Psychological Bulletin 1988;103(3):391-410. [doi: 10.1037/0033-2909.103.3.391] [Medline: 5637250]
40. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981;18(1):39. [doi: 10.2307/3151312]
41. Chang SJ, Yang E, Ryu H, Kim HJ, Yoon JY. Cross-cultural adaptation and validation of the eHealth literacy scale in Korea. Korean J Adult Nurs 2018;30(5):504. [doi: 10.7475/kjan.2018.30.5.504]

XSL•FO
RenderX

42. Paramio Pérez G, Almagro BJ, Hernando Gómez Á, Aguaded Gómez JI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
43. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California health interview survey. J Med Internet Res 2019;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
44. Brashier NM, Schacter DL. Aging in an era of fake news. Curr Dir Psychol Sci 2020;29(3):316-323 [FREE Full text] [doi: 10.1177/0963721420915872] [Medline: 32968336]
45. Gitlow L. Technology use by older adults and barriers to using technology. Physical & Occupational Therapy In Geriatrics 2014;32(3):271-280. [doi: 10.3109/02703181.2014.946640]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
CR: construct reliability
EFA: exploratory factor analysis
KMO: Kaiser-Meyer-Olkin
RMSEA: root-mean-square error of approximation
SRMR: standardized root-mean-square residual
TLI: Tucker-Lewis index

Edited by A Schwartz; submitted 26.08.24; peer-reviewed by CY Tian; comments to author 04.10.24; revised version received 25.11.24; accepted 08.01.25; published 05.02.25.

Please cite as:
Kim S, Park C, Park S, Kim DJ, Bae YS, Kang JH, Chun JW
Measuring Digital Health Literacy in Older Adults: Development and Validation Study
J Med Internet Res 2025;27:e65492
URL: https://www.jmir.org/2025/1/e65492
doi:10.2196/65492
PMID:

©SungMin Kim, Choonghee Park, Sunghyeon Park, Dai-Jin Kim, Ye-Seul Bae, Jae-Heon Kang, Ji-Won Chun. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying Profiles of Digital Literacy Among Community-Dwelling Korean Older Adults: Latent Profile Analysis

Jiyoung Shin[1,2], PhD; Hun Kang[3], MPH; Seongmi Choi[1,4], PhD; Sang Hui Chu[1], PhD; JiYeon Choi[1,5], PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]Health Insurance Research Institute, National Health Insurance Service, Gangwon, Republic of Korea

[3]Department of Social and Behavioral Sciences, Yale University School of Public Health, New Haven, CT, United States

[4]Research Institute of Nursing Science, College of Nursing, Daegu Catholic University, Daegu, Republic of Korea

[5]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
JiYeon Choi, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3301
Email: jychoi610@yuhs.ac

Abstract

Background: The digital divide is apparent not only between older and younger generations but also within the older adult population itself. Identifying digital literacy profiles among older adults is crucial for developing targeted strategies to narrow this divide.

Objective: This study aimed to identify profiles of digital literacy among community-dwelling older adults and to examine factors associated with these profiles.

Methods: Data were collected from community-dwelling older adults in South Korea through a nationwide cross-sectional survey that assessed digital literacy and related factors. Digital literacy was evaluated across 3 domains: information and communication (9 items), content creation and management (4 items), and safety and security (9 items). Latent profile analysis was used to identify profiles of digital literacy among community-dwelling older adults, and multinomial logistic regression was used to identify predictors of profile membership.

Results: A total of 1016 older adults completed structured questionnaires (average age 68, SD 6.5 years; 486/1016, 47.8% men). Three digital literacy profiles were identified (P<.001): “low level” (346/1016, 34.1%), “middle level” (374/1016, 36.8%), and “high level” (296/1016, 29.1%). With the “middle-level” digital literacy group as the reference group, older adult participants (odds ratio [OR] 1.11, 95% CI 1.07-1.15) with less than a middle school education (vs with a college degree or higher; OR 7.22, 95% CI 2.31-22.54), who needed help with one of the 10 instrumental daily activities (vs ≥2 activities; OR 3.06, 95% CI 1.11-8.40) and who did not engage in in-person social activities (OR 1.81, 95% CI 1.07-3.07), were more likely to be in the “low-level” group. Women were less likely to be in the “high-level” digital literacy group than men (OR 0.45, 95% CI 0.25-0.80). Participants with less than a college education were also less likely to be in the “high-level” group, with those having less than a middle school education showing the lowest OR (OR 0.17, 95% CI 0.07-0.41). Those who had never worked (OR 0.23, 95% CI 0.06-0.90) and those not engaging in regular physical exercise (OR 0.58, 95% CI 0.40-0.84) were also less likely to be in the “high-level” digital literacy group. Participants with greater social support were more likely to be in the “high-level” digital literacy group (OR 1.70, 95% CI 1.22-2.37).

Conclusions: These findings underscore the characteristics linked to lower digital literacy and suggest a tailored approach to meet the needs of diverse groups of older adults in a digitalizing society. To promote digital literacy among older adults, potential strategies include improving access to and guidance for using digital devices, specifically designed for this demographic, as well as promoting social support and encouraging participation in social activities.

(J Med Internet Res 2025;27:e57122) doi:10.2196/57122

XSL•FO
RenderX

KEYWORDS

digital literacy; digital divide; older adults; South Korea; latent profile analysis

Introduction

Background

In a digitalizing world, new technology-supported lifestyles hold the potential to foster healthy and independent aging among older adults. These individuals can manage their health and medical conditions through mobile devices [1] or benefit from telecare and access to online health information [2]. In addition, smart homes powered by the Internet of Things can support the daily activities of older adults [3]. Digital devices also enable older adults to maintain social connections and participate in social activities, even when traditional face-to-face interactions are not possible [4]. Therefore, it is crucial to encourage older adults to embrace digital technology to enhance self-care, functionality, and social connection [5].

Nevertheless, the digital divide remains a substantial challenge in our increasingly digital world [6]. Research consistently shows a gap in digital engagement between older adults and younger generations [7,8]. Furthermore, there is a considerable variation in digital proficiency among older adults themselves. For instance, individuals aged ≥75 years generally demonstrated lower levels of digital use and literacy compared to those aged <75 years [9]. In addition, factors such as race and ethnicity and socioeconomic status significantly influence internet use among older adults [10]. To effectively support older adults in adapting to and participating in the digital society, it is essential to recognize and understand the diversity within this demographic.

To address this knowledge gap, it is necessary to understand the varied characteristics of digital literacy within the population and to identify the factors associated with it. Previous research on digital literacy in older adults has primarily focused on their abilities to search for information and communicate online [11-13]. However, the definition of digital literacy is evolving [14]. It is now described as “the ability to access, manage, understand, integrate, communicate, evaluate, and create information safely and appropriately through digital technologies” [15]. In response, the European Commission has introduced a detailed self-assessment grid within the Digital Competence (DigComp) framework. This framework covers a broader range of skills, including content creation and management as well as safety and security, in addition to information search and online communication [11,16]. Therefore, conducting a comprehensive assessment of digital literacy among older adults is both relevant and timely.

Previous studies have primarily focused on identifying factors associated with digital literacy levels among older adults, treating this population as uniform. These studies revealed that not only sociodemographic factors—such as gender [17], age [18,19], living arrangement [8], education level [18], and region [8,19]—but also health-related factors, including cognitive functioning [20], physical limitations [21], and depressive symptoms [22], as well as social factors, such as social support [23] and social participation [24], are linked to digital literacy. However, this method may overlook the varied expressions of digital literacy within the older adult population. A person-centered approach that recognizes this diversity is essential for developing comprehensive education and training programs aimed at enhancing digital literacy among older adults [25].

Objective

Therefore, this study, grounded in the most recent definition of digital literacy, aims to achieve three objectives: (1) identify the profiles of digital literacy among community-dwelling older adults in South Korea, (2) understand the characteristics of digital literacy profiles, and (3) determine the predictive factors that distinguish these profiles.

Methods

Study Design and Setting

This cross-sectional study aimed to assess the digital literacy levels and associated factors among community-dwelling older adults in South Korea. Data were collected through a nationwide survey conducted from October to November 2022 [26]. The survey used proportional stratified sampling on the basis of region, sex, and age groups, reflecting the population distribution documented in South Korea as of June 2022. The inclusion criteria were (1) being aged ≥60 years, (2) achieving a minimum score of 22 on the Korean version of the Mini-Mental State Examination (second edition), and (3) being proficient in the Korean language. In total, 58 trained interviewers visited the homes of potential participants identified through the sampling process. They provided comprehensive information about the study and obtained informed consent from those who were eligible and willing to participate. A total of 1016 older adults completed structured questionnaires designed by the research team to collect data on digital literacy, sociodemographic characteristics, health status, health behaviors, social activities, and social support. The survey was conducted via one-on-one interviews using tablet PCs, with an average completion time of 26.8 minutes per participant. Interviews were conducted in person by trained interviewers with extensive experience in interviewing older adults. Errors or inaccurate responses identified during the review or verification process of completed surveys were corrected or supplemented by interviewers through follow-up phone inquiries.

Ethical Considerations

In compliance with the Helsinki Declaration, this study received approval from the Institutional Review Board of Yonsei University (4-2022-0396). All participants provided written informed consent before their involvement in the study and received gift vouchers worth 10,000 Korean won (US $6.90) upon completing the survey. The collected data were made accessible only to the research team, and the respondents’ personal information was deidentified before analysis.

XSL•FO
RenderX

Measures

Digital Literacy

In this study, we evaluated the components of digital literacy in older adults using the Everyday Digital Literacy Questionnaire (EDLQ), which consists of 22 items developed from our earlier research based on the European Commission's DigComp framework [26]. The EDLQ has shown satisfactory validity and reliability, and it is organized into 3 domains: information and communication (9 items), content creation and management (4 items), and safety and security (9 items) [26]. Responses to each item were recorded on a Likert scale from 1 (strongly disagree) to 5 (strongly agree), with higher aggregate scores indicating greater digital literacy. In our sample, the Cronbach α value was 0.98, confirming high internal consistency [26].

Predictor Variables of Digital Literacy

Sociodemographic Factors

On the basis of previous studies [8,17-19], several sociodemographic factors were identified as influencing the digital literacy of older adults. These factors included age, sex, educational level, living arrangements, region, and economic activities. Age was treated as a continuous variable, measured in years. Sex was categorized as a binary variable, with male coded as 1 and female coded as 0. Educational levels were divided into 4 categories: below middle school graduation (coded as 0), below high school graduation (coded as 1), below college graduation (coded as 2), and college graduation or higher (coded as 3). Living arrangements were classified into 3 categories: living alone (coded as 0), living with a spouse (coded as 1), and residing in a multigenerational household (coded as 2). Region was split into 2 categories: metropolitan areas (coded as 1), which included Seoul, Incheon, and Gyeonggi province, and nonmetropolitan areas (coded as 0), encompassing cities and provinces outside the metropolitan area. Economic activities were categorized into 3 groups: currently engaged in income-generating activities (coded as 2), previously employed but not currently (coded as 1), and never having worked (coded as 0).

Health Factors

On the basis of previous studies [20-22], variables related to health status and behavior were selected as factors influencing the digital literacy of older adults.

For health status, participants provided information on their cognitive, physical, and psychological health. The cognitive health of the participants was assessed using the Korean version of the Mini-Mental State Examination (second edition) [27]. Following the most widely used cutoff point for screening the extent of cognitive impairment [28], we recoded the scores into a binary variable: 24 or higher (coded as 1) and below 24 (coded as 0).

The physical health of the participants was assessed, covering disabilities, chronic diseases, and physical function. Disability status was determined as a binary variable, indicating whether participants had received an official medical diagnosis from specialty physicians: yes (coded as 1) or no (coded as 0). In South Korea, the Act on Welfare of Persons with Disabilities defines a "person with a disability" as someone whose daily life or social activity is significantly impaired by a physical or mental disability over a long period. This act recognizes 15 medical disability categories, including disabilities of the extremities; vision disability; hearing disability; speech and language disability; disability due to facial deformity; intellectual disability; and disabilities resulting from brain injury, renal failure, heart problems, liver disease, respiratory issues, ostomy, epilepsy, autism, and mental disorders [29]. Chronic diseases were categorized into 3 groups based on the number of chronic conditions that lasted over 3 months as diagnosed by a physician: none (coded as 0); 1 (coded as 1); and ≥2 (coded as 2). Physical function was evaluated from 2 perspectives using the Korean-Activities of Daily Living (K-ADL) and the Korean-Instrumental Activities of Daily Living (K-IADL). The K-ADL assesses the level of assistance needed in 7 basic activities of daily living: dressing, washing face and hands, bathing, eating, transferring, toileting, and continence. The Cronbach α value for K-ADL at the time of its development was 0.94 [30], and in this study, it was 0.75. The K-IADL measures the level of assistance needed in 10 instrumental activities of daily living, including decorating, housework, meal preparation, laundry, short-distance travel, using transportation, shopping, handling money, using the telephone, and managing medication. The Cronbach α value for K-IADL at the time of its development was 0.94 [31], and in this study, it was 0.78. We recoded the number of items requiring assistance in K-ADL and K-IADL into 3 categories for each: none (coded as 0); 1 (coded as 1), and ≥2 (coded as 2).

The psychological health of the participants was assessed through health-related quality of life and depressive symptoms. Health-related quality of life was evaluated using the Short-Form Health Survey, which examines 8 domains: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health. For our study, we used version 2 of the 12-item Short-Form Health Survey, an abbreviated version of the 36-item Short-Form Health Survey [32]. Scores were derived using published algorithms that set the mean score of the US general population at 50 and the SD at 10, which were used to calculate the physical component summary and mental component summary scores [32,33]. Scores ranged from 0 to 100, with higher scores indicating a better quality of life. Depressive symptoms were assessed using the integrated Korean version of the Center for Epidemiological Studies-Depression (CES-D) [34]. Participants rated the frequency of their feelings and behaviors over the past week for a total of 20 items, ranging from "rarely or never (<1 day)" (coded as 0) to "most or all of the time (5-7 days)" (coded as 3). After reverse coding the 4 positively worded items, we calculated the total score from the 20 responses. Using the widely accepted threshold for identifying individuals at risk of clinical depression (ie, ≥16) [35], we categorized participants into 2 groups: those at risk for clinical depression and those not at risk. The Cronbach α values for the original CES-D scale and its Korean version were 0.85 and 0.91, respectively [34,36]. In our study sample, the Cronbach α was determined to be 0.89.

XSL•FO
RenderX

For health behaviors, participants provided information on smoking, alcohol consumption, and physical activity. Participants were categorized into 3 groups based on their smoking habits: nonsmokers (coded as 0), ex-smokers (coded as 1), and current smokers (coded as 2). Regarding alcohol consumption, participants indicated whether they had consumed alcohol at least once in the past year (coded as 1) or not (coded as 0). Physical activity was assessed by determining whether participants engaged in daily physical exercise for 10 minutes or more on a regular basis (coded as 1) or not (coded as 0).

Social Factors

On the basis of prior studies [23,24], variables, such as participation in in-person and digital social activities, along with social support, were included as social factors influencing digital literacy among older adults.

To understand participants' in-person social activities, we inquired about their participation over the past year in various activities, including religious or social gatherings; leisure, culture, or sports activities; alumni meetings; volunteer work; political, civic, interest group activities; or other activities [37]. Participants indicated the annual frequency of their involvement in each category of social activities using a scale ranging from "several times a week" (coded as 1) to "not at all" (coded as 7). For our analysis, we grouped these responses into 2 categories: nonparticipation (coded as 0), which includes only the responses indicating "not at all" for all social activity categories, and participation (coded as 1), which includes all other responses. We also evaluated participants' digital social activities by examining their use of digital devices such as desktops and laptops, mobile phones, tablets, e-books, and wearable devices. Participants were asked whether they had used each of the 5 types of digital devices and to specify the duration of use in months. The longest duration reported by each participant was recorded as a continuous variable to represent their overall digital device use. Responses from 10.5% (107/1016) of participants who had no experience using any digital device were recorded as 0 months.

We used the Multidimensional Scale of Perceived Social Support (MSPSS) to assess perceived social support from family, friends, and significant others. The original scale featured a 7-point Likert scale for its 12 items, ranging from "very strongly disagree" (coded as 1) to "very strongly agree" (coded as 7) [38]. For our study, we used the Korean translated version of the MSPSS, which modified the scale to range from "strongly disagree" (coded as 1) to "strongly agree" (coded as 5) [39]. Higher average scores indicate stronger perceived social support. The Cronbach α values were 0.88 for the original version, 0.89 for the Korean translated version, and 0.94 for our sample.

Data Analysis

We chose latent profile analysis (LPA) as the analytical method to identify profiles of digital literacy among older adults living in the community. LPA is a person-centered clustering approach that classifies individuals into unobserved latent subgroups on the basis of similar profiles of observed continuous variables [40]. Our sample size met the minimum recommended criterion (n≥500) for accurate identification using LPA [40,41]. We used 22 items measuring digital literacy across 3 domains as indicator variables for this analysis.

In the initial step of LPA, we selected the model by using information criteria, classification quality indexes, and relative fit indexes. We calculated the Bayesian information criterion, Akaike information criterion, and sample size–adjusted Bayesian information criterion as our information criteria, with lower values indicating a better-fitting model [42]. To assess the quality of model classification, we analyzed the entropy value, targeting a criterion of ≥0.8 [43]. To determine if significant differences in model fit existed between models with K-1 and K profiles, we evaluated the *P* values ($P<.05$) using the Lo-Mendell-Rubin likelihood ratio test, adjusted Lo-Mendell-Rubin likelihood ratio test, and bootstrapped likelihood ratio test [44]. Subsequently, we chose the most suitable number of profiles on the basis of the goodness-of-fit results and their interpretability. Finally, we incorporated predictor variables that affect digital literacy as auxiliary variables to reduce classification errors among profiles [45]. The LPA process was carried out using Mplus (version 8.8; Muthén & Muthén).

To compare the characteristics of the profiles identified through LPA and determine if there were statistically significant differences, we conducted Kruskal-Wallis *H* tests and chi-square tests. In addition, to identify predictor variables that differentiate each profile, we estimated the odds ratio (OR) and their corresponding 95% CIs for the probability of belonging to a specific profile membership using a multinomial logistic regression model. The analysis for intergroup comparisons was performed using SPSS Statistics (version 26; IBM Corp).

Results

We analyzed data from 1016 community-dwelling older adults in South Korea who completed the survey. The participants' average age was 68.0 (SD 6.5) years, and 47.83% (486/1016) of participants were men.

Identification of Digital Literacy Profiles

Model Selection: 3-Class Model

Table 1 presents the available model selection scenario for our sample. The Bayesian information criterion, Akaike information criterion, and sample size–adjusted Bayesian information criterion values decreased as the number of profiles increased, with the rate of decrease diminishing after the 3-profile point. The entropy values exceeded 0.8 in all simulations, indicating satisfactory classification quality. Relative fit indexes demonstrated significant fit only in the 2-profile, 3-profile, and 4-profile scenarios across all 3 indexes (all $P=.002$ or $<.001$). Overall, the model evaluation results suggested that both the 3-profile and 4-profile models were statistically appropriate. However, given the theoretical interpretability of the profile characteristics, the 3-profile model was deemed the most suitable for our sample.

XSL•FO
RenderX

Table 1. Summary of model fit indexes for latent profile analysis.

Fit indexes	1 profile	2 profile	3 profile	4 profile	5 profile	6 profile
AIC[a]	75,974.05	61,809.16	*57,344.01* [b]	55,771.28	54,597.69	53,771.08
BIC[c]	76,190.69	62,139.05	*57,787.14*	56,327.65	55,267.30	54,553.94
SSABIC[d]	76,050.94	61,926.25	*57,501.29*	55,968.75	54,835.353	54,048.94
Entropy	__[e]	0.982	*0.968*	0.954	0.958	0.954
LMR-LRT[f], *P* value	—	<.001	*<.001*	.002	.18	.049
Adjusted LMR-LRT, *P* value	—	<.001	*<.001*	.002	.18	.051
BLRT[g], *P* value	—	<.001	*<.001*	<.001	<.001	<.001

[a]AIC: Akaike information criterion.

[b]Selected class solution values are italicized.

[c]BIC: Bayesian information criterion.

[d]SSABIC: sample size–adjusted Bayesian information criterion.

[e]Not applicable.

[f]LMR-LRT: Lo-Mendell-Rubin likelihood ratio test.

[g]BLRT: bootstrapped likelihood ratio test.

Describing Digital Literacy Profiles

As shown in Table 2, the indicators of digital literacy revealed statistically significant differences across the 3 profiles (all *P* values <.001). Profile 1, which comprises the second-largest group (346/1016, 34.1%), demonstrated the lowest digital literacy levels, with mean scores <2 (out of 5) for each item. Within profile 1, the items “social networking” (mean 1.78), “find information on the Internet” (mean 1.77), and “evaluate internet information” (mean 1.71) in the information and communication domain showed relatively higher self-reported proficiency. Conversely, “convert document formats” (mean 1.12), “edit and share content” (mean 1.13), and “troubleshoot device issues” (mean 1.13) in the content creation and management domain, along with “block spam/phishing” (mean 1.13) in the security and safety domain, scored relatively lower.

Profile 2, the largest group, comprising 36.8% (374/1016) of the sample, exhibited median scores across all items when compared to the other 2 profiles. The mean scores for each item varied significantly, ranging from 1.86 to 3.56. Notably, within the information and communication domain, the items “find information on the Internet” (mean 3.56) and “social networking” (mean 3.54) achieved higher mean scores, both exceeding a mean of 3.5. Conversely, in the content creation and management domain, items such as “edit and share content” (mean 1.86), “convert document formats” (mean 1.89), “troubleshoot device issues” (mean 1.99), and “create documents” (mean 2.10) recorded lower mean scores.

The smallest group, profile 3 (296/1016, 29.1%), comprised participants with the highest relative level of digital literacy. Although the difference was not substantial compared to profile 2, there was a notable variation in the mean scores among the items within this group. All items in the domains of information and communication and safety and security scored a high mean, around 4 (ranging from 3.83 to 4.33), with the exception of “seeking technical help,” which scored a mean of 3.63. However, similar to the other 2 profiles, items in the content creation and management domain exhibited the lowest relative scores, ranging from 3.10 to 3.56.

Figure 1 illustrates the distribution of mean scores for digital literacy indicators across 3 latent profiles. The x-axis represents the measurement items from the 3 domains, while the y-axis shows response scores ranging from 1 to 5 points. On the basis of the distribution characteristics of the 22-item EDLQ scores that measure digital literacy, the profiles were labeled as follows: profile 1 as “low level” (346/1016, 34.1%), profile 2 as “middle level” (374/1016, 36.8%), and profile 3 as “high level” (296/1016, 29.1%).

XSL•FO
RenderX

Table 2. Mean scores[a] of digital literacy in each latent profile.

Variables	Profile 1: low level (n=346), mean (SD)	Profile 2: middle level (n=374), mean (SD)	Profile 3: high level (n=296), mean (SD)	Kruskal-Wallis *H* test	*P* value
Information and communication					
Find information on the Internet	1.77 (0.96)	3.56 (0.81)	4.33 (0.57)	620.78	<.001
Evaluate internet information	1.71 (0.93)	3.28 (0.89)	4.08 (0.74)	552.68	<.001
Transfer files between devices	1.35 (0.61)	2.56 (1.06)	3.83 (0.85)	575.33	<.001
Save internet files	1.42 (0.69)	3.05 (0.98)	4.12 (0.72)	648.31	<.001
Social networking	1.78 (1.07)	3.54 (0.90)	4.24 (0.63)	539.37	<.001
Email file exchange	1.35 (0.69)	2.66 (1.18)	4.01 (0.90)	559.30	<.001
Video calls or conferences	1.28 (0.59)	2.67 (1.07)	3.97 (0.84)	617.49	<.001
Express opinions	1.30 (0.59)	2.83 (1.00)	4.08 (0.75)	663.83	<.001
Comment on posts	1.30 (0.57)	2.83 (1.04)	4.07 (0.78)	656.01	<.001
Content creation and management					
Create documents	1.23 (0.57)	2.10 (0.94)	3.56 (1.03)	545.76	<.001
Convert document formats	1.12 (0.34)	1.89 (0.82)	3.10 (1.06)	520.24	<.001
Edit and share content	1.13 (0.38)	1.86 (0.81)	3.20 (1.08)	523.36	<.001
Troubleshoot device issues	1.13 (0.38)	1.99 (0.81)	3.34 (0.86)	620.99	<.001
Safety and security					
Copyright awareness	1.33 (0.73)	2.89 (1.06)	3.98 (0.82)	608.33	<.001
Protect copyright	1.32 (0.69)	2.90 (1.00)	3.95 (0.78)	631.26	<.001
Device security	1.16 (0.41)	2.43 (0.92)	4.05 (0.73)	733.66	<.001
File deletion	1.30 (0.70)	3.10 (0.99)	4.19 (0.70)	681.36	<.001
Clear search history	1.14 (0.39)	2.62 (0.93)	4.04 (0.83)	732.09	<.001
Block spam or phishing	1.13 (0.39)	2.68 (0.95)	4.07 (0.80)	730.69	<.001
Physical side effects awareness	1.49 (0.90)	3.05 (1.05)	4.07 (0.73)	569.83	<.001
Mental side effects awareness	1.49 (0.93)	3.08 (1.05)	4.01 (0.71)	548.17	<.001
Seek technical help	1.21 (0.56)	2.39 (1.03)	3.63 (0.80)	598.24	<.001

[a]Item scores range from 1 to 5.

XSL•FO
RenderX

Figure 1. The distribution of digital literacy mean scores among 3 latent profiles.

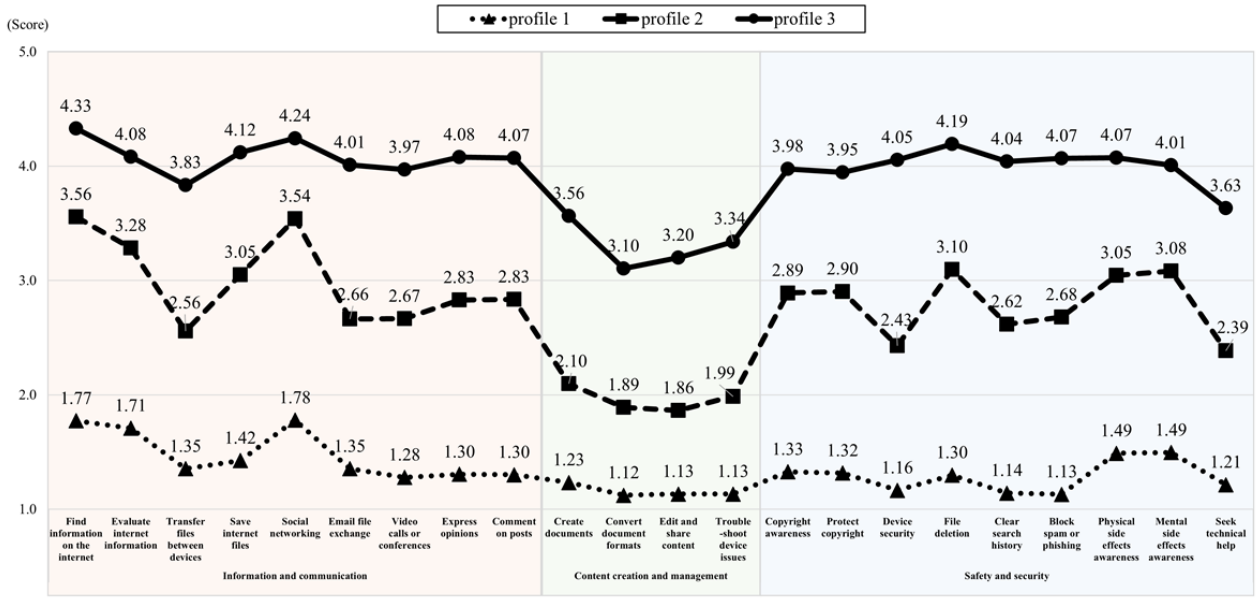

Comparison of Participant Characteristics Among Digital Literacy Profiles

Table 3 provides an overview of the participant characteristics and the differences between the groups in the adopted 3-profile model. The average age of the participants was 68 (SD 6.5) years. Among the total number of participants (N=1016), the majority were female (n=530, 52.2%), had completed at least high school (n=551, 54.2%), lived with a spouse (n=524, 51.6%), resided in nonmetropolitan areas (n=547, 53.8%), and were engaged in income-generating activities (n=622, 61.2%). Most participants had no cognitive impairment (n=969, 95.4%), had been diagnosed with one or more chronic diseases (n=704, 69.3%), and had no diagnosed disabilities (n=967, 95.2%). They were capable of performing independent activities of daily living (n=952, 93.7%) and instrumental activities of daily living (n=839, 82.6%). Regarding health-related quality of life, the mean score for the physical aspect was 50.1 (SD 7.3; out of 100), and for the mental aspect, it was 49.4 (SD 8.4; out of 100). Most participants were not at risk of clinical depression (n=750, 73.8%), engaged in regular physical exercise for more than 10 minutes a day (n=594, 58.5%), had not consumed alcohol in the past year (n=516, 50.8%), and were nonsmokers (n=628, 61.8%). Most participants reported having participated in in-person social activities over the past year (n=853, 84%), and the average perceived social support score was 3.9 (SD 0.7; out of 5). Participants reported an average digital device use history of 82.9 (SD 64.9) months.

When comparing participant characteristics across the 3 profiles, statistically significant differences were observed in sociodemographic, health, and social factors on the basis of digital literacy levels, with the exceptions of region and ADL (both *P* values>.05). In the group with “high-level” digital literacy, the average age was the lowest at 64.7 (SD 4.6) years, and the majority (260/296, 87.8%) had attained at least a high school education. In addition, this group had a higher proportion of male individuals (190/296, 64.2%) and a greater participation in current income-generating activities (217/296, 73.3%) than the other 2 groups. Conversely, in the group with “low-level” digital literacy, a higher percentage of participants had education levels below high school (277/346, 80.1%) and a larger proportion lived alone (117/346, 33.8%), relative to the other 2 groups.

Distinct health patterns were observed among the profiles analyzed. In the “high-level” digital literacy group, a higher proportion of participants exhibited no cognitive impairment (294/296, 99.3%), no diagnosed chronic conditions (120/296, 40.5%), no disabilities (286/296, 96.3%), and no depressive symptoms (244/296, 82.4%) compared to the other 2 groups. This group also scored highest in health-related quality of life, both physically and mentally (physical component summary: mean 53.2, SD 5.0; mental component summary: mean 51.4, SD 7.9), indicating a quality of life that surpasses the average of the reference group. In contrast, the “low-level” digital literacy group displayed poorer outcomes across these health indicators. Although there was no significant difference in ADL among the 3 groups, IADL differed significantly. Notably, the “low-level” digital literacy group had a higher proportion of participants reporting dependency in one or more IADL items (96/346, 27.8%) compared to the other groups. In terms of health behaviors, the “high-level” digital literacy group had a relatively higher proportion of participants who engaged in regular exercise for more than 10 minutes (211/296, 71.3%), consumed alcohol (183/296, 61.8%), and were current smokers (48/296, 16.2%) than the other groups.

Regarding social factors, the “high-level” digital literacy group had the highest proportion of participants engaging in in-person social activities than the other 2 groups, at 92.2% (273/296). In addition, this group reported the longest history of using digital devices, averaging 128.0 (SD 74.8) months, and the highest mean score for perceived social support, at a mean of 4.1 (SD 0.7; out of 5).

XSL•FO
RenderX

Table 3. Participants' characteristics and differences among the 3 latent profiles.

Variables	Total (N=1016)	Profile 1: low level (n=346)	Profile 2: middle level (n=374)	Profile 3: high level (n=296)	*P* value
Age (y), mean (SD)	68.0 (6.5)	72.5 (7.0)	66.5 (4.8)	64.7 (4.6)	<.001
Sex, n (%)					<.001
Male	486 (47.8)	129 (37.3)	167 (44.7)	190 (64.2)	
Female	530 (52.2)	217 (62.7)	207 (55.3)	106 (35.8)	
Education level, n (%)					<.001
Below middle school	240 (23.6)	183 (52.9)	44 (11.8)	13 (4.4)	
Below high school	225 (22.1)	94 (27.2)	108 (28.9)	23 (7.8)	
Below college	451 (44.4)	63 (18.2)	202 (54)	186 (62.8)	
College and above	100 (9.8)	6 (1.7)	20 (5.3)	74 (25)	
Living arrangements, n (%)					<.001
Living alone	199 (19.6)	117 (33.8)	48 (12.8)	34 (11.5)	
Living with a spouse	524 (51.6)	156 (45.1)	209 (55.9)	159 (53.7)	
Multigenerational household	293 (28.8)	73 (21.1)	117 (31.3)	103 (34.8)	
Region, n (%)					.15
Metropolitan area[a]	469 (46.2)	145 (41.9)	180 (48.1)	144 (48.6)	
Nonmetropolitan area[b]	547 (53.8)	201 (58.1)	194 (51.9)	152 (51.4)	
Engaged in economic activities, n (%)					<.001
Never	40 (3.9)	19 (5.5)	18 (4.8)	3 (1)	
Used to	354 (34.8)	157 (45.4)	121 (32.4)	76 (25.7)	
Yes	622 (61.2)	170 (49.1)	235 (62.8)	217 (73.3)	
K-MMSE-2[c], n (%)					<.001
≥24	969 (95.4)	309 (89.3)	366 (97.9)	294 (99.3)	
22-23	47 (4.6)	37 (10.7)	8 (2.1)	2 (0.7)	
Number of diagnosed chronic diseases, n (%)					<.001
0	312 (30.7)	60 (17.3)	132 (35.3)	120 (40.5)	
1	301 (29.6)	90 (26.0)	113 (30.2)	98 (33.1)	
≥2	403 (39.7)	196 (56.6)	129 (34.5)	78 (26.4)	
Disability, n (%)					0.006
No	967 (95.2)	319 (92.2)	363 (97.1)	285 (96.3)	
Yes	49 (4.8)	27 (7.8)	11 (2.9)	11 (3.7)	
Number of dependent items of K-ADL[d], n (%)					.20
0	952 (93.7)	320 (92.5)	352 (94.1)	280 (94.6)	
1	37 (3.6)	11 (3.2)	15 (4)	11 (3.7)	
≥2	27 (2.7)	15 (4.3)	7 (1.9)	5 (1.7)	
Number of dependent items of K-IAD[e], n (%)					<.001
0	839 (82.6)	250 (72.3)	332 (88.8)	257 (86.8)	
1	68 (6.7)	38 (11)	16 (4.3)	14 (4.7)	
≥2	109 (10.7)	58 (16.8)	26 (7)	25 (8.4)	
HRQoL[f] (scores), mean (SD)					
PCS[g]	50.1 (7.3)	46.4 (8.2)	50.9 (6.5)	53.2 (5.0)	<.001

XSL•FO
RenderX

Variables	Total (N=1016)	Profile 1: low level (n=346)	Profile 2: middle level (n=374)	Profile 3: high level (n=296)	*P* value
MCS[h]	49.4 (8.4)	47.5 (8.8)	49.7 (8.0)	51.4 (7.9)	<.001
CES-D[i] scale (scores), n (%)					<.001
1-15	750 (73.8)	219 (63.3)	287 (76.7)	244 (82.4)	
≥16	266 (26.2)	127 (36.7)	87 (23.3)	52 (17.6)	
Exercise, n (%)					<.001
No	422 (41.5)	168 (48.6)	169 (45.2)	85 (28.7)	
Yes	594 (58.5)	178 (51.4)	205 (54.8)	211 (71.3)	
Alcohol consumption, n (%)					<.001
No	516 (50.8)	215 (62.1)	188 (50.3)	113 (38.2)	
Yes	500 (49.2)	131 (37.9)	186 (49.7)	183 (61.8)	
Smoking, n (%)					<.001
Nonsmoker	628 (61.8)	247 (71.4)	228 (61.0)	153 (51.7)	
Ex-smoker	260 (25.6)	70 (20.2)	95 (25.4)	95 (32.1)	
Current smoker	128 (12.6)	29 (8.4)	51 (13.6)	48 (16.2)	
In-person social activities, n (%)					<.001
No	163 (16)	96 (27.7)	44 (11.8)	23 (7.8)	
Yes	853 (84)	250 (72.3)	330 (88.2)	273 (92.2)	
History of digital device use (months), mean (SD)	82.9 (64.9)	43.3 (42.4)	84.0 (47.9)	128.0 (74.8)	<.001
MSPSS[j] (scores), mean (SD)	3.9 (0.7)	3.6 (0.8)	3.9 (0.7)	4.1 (0.7)	<.001

[a]Metropolitan area: Seoul, Incheon, and Gyeonggi province.

[b]Nonmetropolitan area: cities and provinces other than the metropolitan area.

[c]K-MMSE-2: Korean Mini-Mental State Examination, second edition.

[d]K-ADL: Korean Activities of Daily Living.

[e]K-IADL: Korean Instrumental Activities of Daily Living.

[f]HRQoL: health-related quality of life.

[g]PCS: physical component summary.

[h]MCS: mental component summary.

[i]CES-D: Center for Epidemiological Studies-Depression.

[j]MSPSS: Multidimensional Scale of Perceived Social Support.

Predictor Variables for Identifying Digital Literacy Profiles

Table 4 represents the results of multinomial logistic regression for predictor variables of digital literacy profiles. The statistically significant predictors that differentiate each group are age, sex, educational level, income-generating activities, dependence on IADL, physical health-related quality of life, physical exercise, in-person social activities, history of digital device usage, and perceived social support.

Among these variables, participants with an educational level below middle school were over 7 times more likely to be categorized in the “low-level” digital literacy group than the “middle-level” group compared to those holding a college degree or higher (OR 7.223, 95% CI 2.314-22.541). In addition, participants who required assistance with 1 of 10 instrumental activities of daily living, such as housework, meal preparation, or laundry, were over 3 times more likely to fall into the “low-level” digital literacy group than those requiring assistance with 2 or more such activities (OR 3.055, 95% CI 1.111-8.399). Furthermore, participants who did not engage in in-person social activities were nearly twice as likely to be in the “low-level” digital literacy group compared to those who did participate (OR 1.810, 95% CI 1.068-3.065). Each 1-month increment in the history of digital device use was associated with a lower likelihood of being in the “low-level” than “middle-level” digital literacy group (OR 0.986, 95% CI 0.981-0.991).

XSL•FO
RenderX

Table 4. Multinomial logistic regression for predicting variables of digital literacy profiles.

Variables	"Low level" vs "middle level" (reference)		"High level" vs "middle level" (reference)	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Age (y)	1.109 (1.066-1.153)	<.001	0.945 (0.902-0.991)	.02
Sex (reference: male)				
Female	0.683 (0.367-1.271)	.23	0.449 (0.253-0.795)	.01
Education level (reference: college and above)				
Below middle school	7.223 (2.314-22.541)	.001	0.166 (0.068-0.405)	<.001
Below high school	2.465 (0.812-7.487)	.11	0.106 (0.050-0.223)	<.001
Below college	1.725 (0.568-5.237)	.34	0.332 (0.179-0.614)	<.001
Living arrangements (reference: multigenerational household)				
Living alone	1.582 (0.880-2.846)	.13	1.669 (0.892-3.125)	.11
Living with a spouse	0.977 (0.613-1.559)	.92	1.080 (0.712-1.638)	.72
Region (reference: metropolitan area[b])				
Nonmetropolitan area[c]	0.797 (0.535-1.188)	.27	0.923 (0.627-1.360)	.69
Engaged in economic activities (reference: yes)				
Never	0.825 (0.298-2.284)	.71	0.234 (0.061-0.902)	.04
Used to	0.894 (0.574-1.392)	.62	1.043 (0.671-1.623)	.85
K-MMSE-2[d] (reference: 22-23)				
≥24	0.703 (0.282-1.748)	.45	0.612 (0.117-3.207)	.56
Number of diagnosed chronic diseases (reference: ≥2)				
0	0.647 (0.388-1.077)	.09	0.949 (0.592-1.522)	.83
1	0.920 (0.584-1.449)	.72	1.255 (0.787-2.002)	.34
Disability (reference: yes)				
No	0.601 (0.236-1.534)	.29	1.020 (0.326-3.187)	.97
Number of dependent items of K-ADL[e] (reference: ≥2)				
0	1.110 (0.292-4.212)	.88	0.558 (0.131-2.380)	.43
1	0.356 (0.065-1.957)	.24	0.886 (0.168-4.669)	.89
Number of dependent items of K-IADL[f] (reference: ≥2)				
0	1.044 (0.494-2.208)	.91	0.501 (0.235-1.066)	.07
1	3.055 (1.111-8.399)	.03	0.513 (0.175-1.501)	.22
HRQoL[g]				
PCS[h]	0.973 (0.943-1.004)	.09	1.047 (1.010-1.085)	.01
MCS[i]	0.986 (0.956-1.017)	.36	1.010 (0.981-1.040)	.49
CES-D[j] (reference: ≥16)				
1-15	1.162 (0.684-1.972)	.58	0.892 (0.516-1.544)	.68
Exercise (reference: yes)				
No	0.953 (0.645-1.410)	.81	0.578 (0.397-0.843)	.004
Alcohol consumption (reference: yes)				
No	0.751 (0.496-1.137)	.18	0.938 (0.629-1.399)	.76
Smoking (reference: current smoker)				
Nonsmoker	1.725 (0.821-3.623)	.15	1.803 (0.938-3.466)	.08

XSL•FO
RenderX

Variables	"Low level" vs "middle level" (reference)		"High level" vs "middle level" (reference)	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Ex-smoker	0.814 (0.408-1.621)	.56	1.082 (0.601-1.946)	.79
In-person social activities (reference: yes)				
No	1.810 (1.068-3.065)	.03	1.058 (0.563-1.987)	.86
History of digital device use	0.986 (0.981-0.991)	<.001	1.009 (1.006-1.013)	<.001
MSPSS[k]	0.813 (0.597-1.109)	.19	1.700 (1.219-2.369)	.002

[a]OR: odds ratio.

[b]Metropolitan area: Seoul, Incheon, and Gyeonggi province.

[c]Nonmetropolitan area: cities and provinces other than the metropolitan area.

[d]K-MMSE-2: Korean-Mini-Mental State Examination, second edition.

[e]K-ADL: Korean-Activities of Daily Living.

[f]K-IADL: Korean-Instrumental Activities of Daily Living.

[g]HRQoL: health-related quality of life.

[h]PCS: physical component summary.

[i]MCS: mental component summary.

[j]CES-D: Center for Epidemiological Studies-Depression.

[k]MSPSS: Multidimensional Scale of Perceived Social Support.

Meanwhile, women were 55.1% less likely to be in the "high-level" digital literacy group than the "middle-level" group than men (OR 0.449, 95% CI 0.253-0.795). Regarding educational attainment, participants with less than a college education were 66.8% less likely to be categorized as "high-level" in digital literacy compared to those with below a college degree (OR 0.332, 95% CI 0.179-0.614). Those with less than a high school education were about 89.4% less likely (OR 0.106, 95% CI 0.050-0.223), and individuals with less than a middle school education were nearly 83.4% less likely (OR 0.166, 95% CI 0.068-0.405). Participants who had never worked were 76.6% less likely to be in the "high-level" digital literacy group compared to those currently engaged in income-generating activities (OR 0.234, 95% CI 0.061-0.902). Participants who did not regularly engage in physical exercise for more than 10 minutes were 42.2% less likely to be in the "high-level" digital literacy group than those who did (OR 0.578, 95% CI 0.397-0.843). Each one-month increment in the history of digital device use was associated with a higher likelihood of being in the "high-level" than "middle-level" digital literacy group (OR 1.009, 95% CI 1.006-1.013). In addition, each 1-point increase in perceived social support nearly doubled the likelihood of being in the "high-level" digital literacy group (OR 1.700, 95% CI 1.219-2.369).

Discussion

Principal Findings

This study identified distinct profiles of digital literacy among community-dwelling older adults using the EDLQ [26], an instrument developed for this demographic in accordance with the European Commission's DigComp framework [11]. We categorized digital literacy into 3 distinct profiles: low level, middle level, and high level. Across all profiles, we observed a common pattern where digital literacy related to online information searching and social networking activities was relatively high. In contrast, skills in content creation and management, such as creating and editing documents, converting document formats, sharing content, and troubleshooting device issues, were notably lower. The low-level group exhibited particularly low skills in the security and safety domain, especially in their ability to block spam or delete search history. Meanwhile, the middle-level group displayed deficiencies in managing device security settings and seeking assistance for resolving technical issues within the security and safety domain.

Comparison to Prior Work

Our results highlight 2 critical areas essential for supporting the digital literacy of older adults. The first relates to the ability to customize digital content according to individual needs. The second underscores the importance of literacy in digital security and safety. Concerns about security and safety, as noted in previous studies [46,47], can present a significant challenge for older adults when adopting new technology. These findings provide valuable insights that have not been adequately addressed in earlier research, which primarily concentrated on information search and online communication as benchmarks for evaluating digital literacy [13,48].

Notably, our findings highlight the presence of a digital divide among community-dwelling older adults in South Korea, emphasizing the need for tailored strategies to address this issue. Consistent with prior research [9,17,49,50], older adults who were female, of greater age, and had lower educational levels were more likely to exhibit low digital literacy. This suggests that despite high smartphone usage among older adults in countries like South Korea [51], a significant digital divide can still exist within this demographic. It further highlights the importance of considering sociodemographic factors when supporting digitally vulnerable populations [17-19]. For instance, in Singapore, a successful initiative known as "Project Wire Up" was implemented to ensure that older adults with lower socioeconomic status are not left behind in the digital

XSL•FO
RenderX

society [52]. Skilled volunteers visited the homes of older adults to deliver a progressively challenging, tiered curriculum designed specifically for their needs. Consequently, the digital literacy scores of the participants in the intervention group improved significantly compared to those in the control group [52].

One approach to mitigating the digital divide among older adults is to increase the accessibility and use of digital devices. In our study, older adults with a longer history of using digital devices were more likely to achieve a middle-level of digital literacy rather than a low-level, and a high-level rather than a middle-level. Project Wire Up in Singapore [52] exemplifies this approach; a major component of the program was assisting older adults in acquiring smartphones and accessing the internet. In addition, the project focused on shifting traditional face-to-face social networking activities to digital platforms [52]. The recreational use of digital devices has been acknowledged as crucial for improving the digital literacy of older adults [23]. Furthermore, both the frequency and duration of internet use have shown a positive correlation with digital health literacy [53].

From a social connectedness perspective, the significance of social support and active participation in social activities is crucial for increasing the digital literacy of older adults. In our study, older adults who perceived greater social support were more likely to belong to the high-level digital literacy group rather than the middle level. Similarly, those engaged in in-person social activities tended to fall into the middle-level digital literacy group rather than the low-level group. A qualitative study investigating how older adults acquire digital literacy found that support from family, professionals, and peers significantly influenced not only the initiation but also the depth and frequency of digital device use [23]. Furthermore, a systematic review of the adoption of digital health technologies indicated that high social connectedness acts as a promoting factor across various cultural groups [54]. Although the exact mechanism of this association requires further exploration, social connectedness may impact digital literacy in older adults in 2 primary ways: first, by aiding them in learning and acquiring the skills to use digital devices [55], and second, by establishing a norm for digital device use [56].

In a similar context, our study showed that older adults currently engaged in economic activities were more likely to belong to the high-level digital literacy group compared to those who had never done so. Since economic activities often fall within the broader spectrum of social activities [57], it is plausible that the mechanism related to social connectedness also played a role here. In addition, previous research has shown that digital literacy in middle-aged and older adults positively affected their participation in managing personal finances [58]. Typically, most older adults may have primarily earned income through their careers until retirement. It is also conceivable that those more proficient in using digital devices and systems had better opportunities to engage in digital-based economic activities, such as using a mobile app for stock investments. However, since engagement in economic activities was measured with a single item, and the involvement of older adults in digital-based economic activities has not been extensively studied, our findings should be interpreted with caution and further research is needed.

Our findings also indicate that health-related factors are closely associated with digital literacy. Interestingly, our analysis revealed that older adults who are more dependent on IADL tend to fall into the middle-level digital literacy group rather than the low-level group. This contrasts with previous studies that focused on older adults with diabetes in South Korea [50,59]. This discrepancy suggests a potential new perspective on technology use among community-dwelling older adults. Since the early 2000s, there has been an increase in the use of smart home technology to assist individuals with diminished capabilities due to aging or disabilities [60-63]. These technologies, which incorporate sensors and robotics, can aid in supporting the IADL of older adults, such as meal preparation and bathing [60,61]. When introducing technology to older adults, the most critical factors to consider are the perceived benefits and ease of use [64,65]. Physical and cognitive aging can significantly hinder the adoption of new technologies by older adults [66]. As aging often requires more time to manage daily tasks like taking medication and undergoing physical therapy [67], older adults may view the adoption of new technologies as an additional burden that necessitates the reallocation of their time and energy [65].

In addition, in our sample, older adults who engaged in regular physical exercise were more likely to be categorized within the high-level digital literacy group. Similarly, a pilot study that used a smartphone app and activity-tracking device for physical training reported that participants with high technological proficiency were more receptive to modern technologies such as fitness trackers or smartphones compared to their less technologically adept counterparts [68]. Further research is required to explore the relationship between the health status or health behaviors of older adults and their use of digital technology.

Limitations

This study has several limitations. First, it relies on cross-sectional data, which precludes making causal inferences about the relationships between variables. Further longitudinal research is necessary to determine the temporality of the variables identified. Second, although the study assessed a wide range of sociodemographic, health-related, and social variables, it is possible that some potential correlates of digital literacy were overlooked. In addition, significant variables measured by a single item, such as economic activity, may not allow for robust interpretation. Third, our assessment of digital literacy covered a broad spectrum of domains using a measurement that reflects the most recent definition of the concept. However, as digital literacy is an evolving field, future studies might find our assessment to be incomplete. Despite these limitations, this study effectively provides a current understanding and characterization of digital literacy among community-dwelling older adults in South Korea.

Conclusions

Our findings reveal a significant disparity in digital literacy among older adults in a rapidly digitalizing country. Specifically,

XSL•FO
RenderX

factors such as social connectedness and health status were closely associated with the levels of digital literacy observed within this population. Older adults with higher digital literacy may have more opportunities in the digital society. To prevent older adults, particularly those who are digitally vulnerable, from being marginalized in the digital society, we propose a tailored approach. This approach focuses on understanding the characteristics of these older adult groups and addressing modifiable factors. Strategies to enhance digital literacy among older adults could include improving access to digital devices and providing step-by-step instructions tailored to the needs and education levels of digitally vulnerable older adults. In addition, efforts are needed to promote social support and encourage participation in social activities.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2020R1A6A1A03041989) and Institute for Innovation in Digital Healthcare, Yonsei University.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JC conceptualized and supervised the study. JC, JS, and SC developed the study and analytical design. SC prepared the dataset, and JS performed data analysis. HK and JS wrote the draft, and JC and SHC reviewed and edited the manuscript. All the authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
2. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: a literature review. Int J Med Inform 2011 Jan;80(1):1-12 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
3. Carnemolla P. Ageing in place and the internet of things – how smart home technologies, the built environment and caregiving intersect. Vis. Eng 2018 Dec 06;6(1):22. [doi: 10.1186/S40327-018-0066-5]
4. Banskota S, Healy M, Goldberg EM. 15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic. West J Emerg Med 2020 Apr 14;21(3):514-525 [FREE Full text] [doi: 10.5811/westjem.2020.4.47372] [Medline: 32302279]
5. Rowe JW, Kahn RL. Successful aging. Gerontologist 1997 Aug;37(4):433-440. [doi: 10.1093/geront/37.4.433] [Medline: 9279031]
6. van Dijk J. The Digital Divide. Hoboken, NJ: John Wiley & Sons; 2020.
7. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
8. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
9. Park SY, Kim JW. A panel analysis (2010-2018) of the digital divide among different age groups among older adults. Korean Data Anal Soc 2019 Oct 31;21(5):2407-2418. [doi: 10.37727/jkdas.2019.21.5.2407]
10. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
11. Ferrari A, Punie Y, Brecko B. DIGCOMP: a framework for developing and understanding digital competence in Europe. European Commission. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC83167 [accessed 2024-04-29]
12. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021 Feb 03;23(2):e26145 [FREE Full text] [doi: 10.2196/26145] [Medline: 33533727]
13. Kang H, Baek J, Chu SH, Choi J. Digital literacy among Korean older adults: a scoping review of quantitative studies. Digit Health 2023;9:20552076231197334 [FREE Full text] [doi: 10.1177/20552076231197334] [Medline: 37654708]
14. Bulfin S, McGraw K. Digital literacy in theory, policy and practice: old concerns, new opportunities. In: Henderon M, Romeo G, editors. Teaching and Digital Technologies: Big Issues and Critical Questions. Cambridge, MA: Cambridge University Press; 2015:266-281.

XSL•FO
RenderX

15. Law NW, Woo DJ, de la Torre J, Wong KW. A global framework of reference on digital literacy skills for indicator 4. United Nations Educational, Scientific and Cultural Organization. URL: https://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf [accessed 2024-02-05]
16. Vuorikari R, Punie Y, Gomez SC, van den Brande G. DigComp 2.0: the digital competence framework for citizens. Update phase 1: the conceptual reference model. European Commission. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC101254 [accessed 2024-04-29]
17. Acilar A, Sæbø Ø. Towards understanding the gender digital divide: a systematic literature review. Glob Knowl Mem Commun 2021 Dec 16;72(3):233-249. [doi: 10.1108/GKMC-09-2021-0147]
18. Chang J, McAllister C, McCaslin R. Correlates of, and barriers to, internet use among older adults. J Gerontol Soc Work 2015;58(1):66-85. [doi: 10.1080/01634372.2014.913754] [Medline: 24941050]
19. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
20. Lee J, Tak SH. The effects of digital literacy and the frequency of contact with children or grandchildren via ICT on the cognitive function of the elderly. Korea Gerontol Soc 2020 Jun 01;40(3):413-428. [doi: 10.31888/JKGS.2020.40.3.413]
21. Gatto SL, Tak SH. Computer, internet, and e-mail use among older adults: benefits and barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
22. Lee MA, Ferraro KF, Kim G. Digital technology use and depressive symptoms among older adults in Korea: beneficial for those who have fewer social interactions? Aging Ment Health 2021 Oct 02;25(10):1839-1847. [doi: 10.1080/13607863.2020.1839863] [Medline: 33131296]
23. Tsai HS, Shillair R, Cotten SR. Social support and "playing around": an examination of how older adults acquire digital literacy with tablet computers. J Appl Gerontol 2017 Jan;36(1):29-55 [FREE Full text] [doi: 10.1177/0733464815609440] [Medline: 26491029]
24. Kim J, Lee HY, Christensen MC, Merighi JR. Technology access and use, and their associations with social engagement among older adults: do women and men differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]
25. Howard MC, Hoffman ME. Variable-centered, person-centered, and person-specific approaches. Organ Res Methods 2017 Dec 10;21(4):846-876. [doi: 10.1177/1094428117744021]
26. Choi J, Choi S, Song K, Baek J, Kim H, Choi M, et al. Everyday digital literacy questionnaire for older adults: instrument development and validation study. J Med Internet Res 2023 Dec 14;25:e51616 [FREE Full text] [doi: 10.2196/51616] [Medline: 38095999]
27. We help you develop and use the most reliable psychological tests. INPSYT Co. 2020. URL: https://inpsyt.co.kr [accessed 2025-02-09]
28. Tsoi KK, Chan JY, Hirai HW, Wong SY, Kwok TC. Cognitive tests to detect dementia: a systematic review and meta-analysis. JAMA Intern Med 2015 Sep;175(9):1450-1458. [doi: 10.1001/jamainternmed.2015.2152] [Medline: 26052687]
29. Enforcement decree of the act on welfare of persons with disabilities. The Korean Law Information Center. URL: https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=141243&chrClsCd=010203&urlMode=engLsInfoR&viewCls=engLsInfoR [accessed 2024-10-26]
30. Won CW, Rho YG, Kim SY, Cho BR, Lee YS. The validity and reliability of Korean Activities of Daily Living (K-ADL) scale. J Korean Soc Geriatr Neurosurg 2002;6(2):98-106 [FREE Full text]
31. Won CW, Rho YG, Sunwoo D, Lee YS. The validity and reliability of Korean Instrumental Activities of Daily Living (K-IADL) scale. J Korean Soc Geriatr Neurosurg 2002;6(4):273-280 [FREE Full text]
32. Ware Jr J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
33. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: results from the IQOLA project. International quality of life assessment. J Clin Epidemiol 1998 Nov;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7] [Medline: 9817135]
34. Jeon G, Choi S, Yang B. Integrated Korean version of CES-D development. J Korean Psychol Assoc Health 2001;6(1):59-76.
35. Vilagut G, Forero CG, Barbaglia G, Alonso J. Screening for depression in the general population with the center for epidemiologic studies depression (CES-D): a systematic review with meta-analysis. PLoS One 2016;11(5):e0155431 [FREE Full text] [doi: 10.1371/journal.pone.0155431] [Medline: 27182821]
36. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
37. Korean longitudinal study of ageing (KLoSA). Korea Employment Information Service. URL: https://survey.keis.or.kr/klosa/klosaque/List.jsp [accessed 2024-02-05]
38. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. J Pers Assess 1988 Mar;52(1):30-41. [doi: 10.1207/S15327752jpa5201_2]
39. Shin JS, Lee YB. The effects of social supports on psychosocial well-being of the unemployed. Korean J Soc Welf 1999;37:241-269 [FREE Full text]

XSL•FO
RenderX

40. Spurk D, Hirschi A, Wang M, Valero D, Kauffeld S. Latent profile analysis: a review and "how to" guide of its application within vocational behavior research. J Vocat Behav 2020 Aug;120:103445. [doi: 10.1016/j.jvb.2020.103445]
41. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Model 2007 Dec 05;14(4):535-569. [doi: 10.1080/10705510701575396]
42. Weller BE, Bowen NK, Faubert SJ. Latent class analysis: a guide to best practice. J Black Psychol 2020 Jun 05;46(4):287-311. [doi: 10.1177/0095798420930932]
43. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif 1996 Sep;13(2):195-212 [FREE Full text] [doi: 10.1007/BF01246098]
44. Williams GA, Kibowski F. Latent class analysis and latent profile analysis. In: Jason LA, Glenwick DS, editors. Handbook of Methodological Approaches to Community-Based Research: Qualitative, Quantitative, and Mixed Methods. Oxford, UK: Oxford Academic Press; 2016:143-152.
45. Asparouhov T, Muthén B. Auxiliary variables in mixture modeling: three-step approaches using Mplus. Struct Equ Modeling 2014 Jun 09;21(3):329-341. [doi: 10.1080/10705511.2014.915181]
46. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
47. Vitak J, Shilton K. Trust, privacy, and security, and accessibility considerations when conducting mobile technologies research with older adults. In: National Academies of Sciences, Engineering, and Medicine , editor. Mobile Technology for Adaptive Aging: Proceedings of a Workshop. Washington, DC: National Academies Press; 2020:1-20.
48. An S, Lim Y, Chung S. The effects of digital literacy and health empowerment on elders' communication with doctors: focusing on moderating effect of health beliefs. J Korean Acad Community Health Nurs 2022;33(1):53. [doi: 10.12799/jkachn.2022.33.1.53]
49. Lee J, Tak SH. Factors associated with eHealth literacy focusing on digital literacy components: a cross-sectional study of middle-aged adults in South Korea. Digit Health 2022;8:20552076221102765 [FREE Full text] [doi: 10.1177/20552076221102765] [Medline: 35615270]
50. Park S, Kim B. Readiness for utilizing digital intervention: patterns of internet use among older adults with diabetes. Prim Care Diabetes 2020 Dec;14(6):692-697. [doi: 10.1016/j.pcd.2020.08.005] [Medline: 32839128]
51. Gallup report: the 2012-2022 report on smartphone use and brand, smart watch, and wireless earphone. Korea Gallup Institute. URL: https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=1309 [accessed 2024-04-29]
52. Ngiam NH, Yee WQ, Teo N, Yow KS, Soundararajan A, Lim JX, et al. Building digital literacy in older adults of low socioeconomic status in Singapore (project wire up): nonrandomized controlled trial. J Med Internet Res 2022 Dec 02;24(12):e40341 [FREE Full text] [doi: 10.2196/40341] [Medline: 36459398]
53. Liu S, Zhao H, Fu J, Kong D, Zhong Z, Hong Y, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in Southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996 [FREE Full text] [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
54. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and facilitators of digital health among culturally and linguistically diverse populations: qualitative systematic review. J Med Internet Res 2023 Feb 28;25:e42719 [FREE Full text] [doi: 10.2196/42719] [Medline: 36853742]
55. Hyman A, Stacy E, Mohsin H, Atkinson K, Stewart K, Novak Lauscher H, et al. Barriers and facilitators to accessing digital health tools faced by South Asian Canadians in surrey, British Columbia: community-based participatory action exploration using photovoice. J Med Internet Res 2022 Jan 13;24(1):e25863 [FREE Full text] [doi: 10.2196/25863] [Medline: 35023842]
56. Zhang B, Ying L, Khan MA, Ali M, Barykin S, Jahanzeb A. Sustainable digital marketing: factors of adoption of m-technologies by older adults in the Chinese market. Sustainability 2023 Jan 19;15(3):1972. [doi: 10.3390/su15031972]
57. Levasseur M, Richard L, Gauvin L, Raymond E. Inventory and analysis of definitions of social participation found in the aging literature: proposed taxonomy of social activities. Soc Sci Med 2010 Dec;71(12):2141-2149 [FREE Full text] [doi: 10.1016/j.socscimed.2010.09.041] [Medline: 21044812]
58. Wang Q, Liu C, Lan S. Digital literacy and financial market participation of middle-aged and elderly adults in China. Econ Polit Stud 2022 Sep 06;11(4):441-468. [doi: 10.1080/20954816.2022.2115191]
59. Park S, Kim B. Predictors of internet use among older adults with diabetes in South Korea: survey study. JMIR Med Inform 2020 Dec 23;8(12):e19061 [FREE Full text] [doi: 10.2196/19061] [Medline: 33277232]
60. Ding D, Cooper RA, Pasquina PF, Fici-Pasquina L. Sensor technology for smart homes. Maturitas 2011 Jun;69(2):131-136. [doi: 10.1016/j.maturitas.2011.03.016] [Medline: 21531517]
61. Wilson G, Pereyda C, Raghunath N, de la Cruz G, Goel S, Nesaei S, et al. Robot-enabled support of daily activities in smart home environments. Cogn Syst Res 2019 May;54:258-272 [FREE Full text] [doi: 10.1016/j.cogsys.2018.10.032] [Medline: 31565029]
62. Jamwal R, Jarman HK, Roseingrave E, Douglas J, Winkler D. Smart home and communication technology for people with disability: a scoping review. Disabil Rehabil Assist Technol 2022 Aug;17(6):624-644. [doi: 10.1080/17483107.2020.1818138] [Medline: 32924660]

XSL•FO
RenderX

63. Perez AJ, Siddiqui F, Zeadally S, Lane D. A review of IoT systems to enable independence for the elderly and disabled individuals. Internet Things 2023 Apr;21:100653. [doi: 10.1016/j.iot.2022.100653]
64. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 01;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
65. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506 [FREE Full text] [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
66. Czaja SJ, Lee CC. The impact of aging on access to technology. Univ Access Inf Soc 2006 Dec 8;5(4):341-349. [doi: 10.1007/S10209-006-0060-X]
67. Owsley C, Sloane M, McGwin Jr G, Ball K. Timed instrumental activities of daily living tasks: relationship to cognitive function and everyday performance assessments in older adults. Gerontology 2002;48(4):254-265 [FREE Full text] [doi: 10.1159/000058360] [Medline: 12053117]
68. Steinert A, Buchem I, Merceron A, Kreutel J, Haesner M. A wearable-enhanced fitness program for older adults, combining fitness trackers and gamification elements: the pilot study fMOOC@Home. Sport Sci Health 2018 Jan 10;14(2):275-282. [doi: 10.1007/S11332-017-0424-Z]

Abbreviations

CES-D: Center for Epidemiological Studies-Depression
DigComp: Digital Competence
EDLQ: Everyday Digital Literacy Questionnaire
K-ADL: Korean-Activities of Daily Living
K-IADL: Korean-Instrumental Activities of Daily Living
LPA: latent profile analysis
OR: odds ratio

Edited by N Cahill; submitted 05.02.24; peer-reviewed by C Urquhart, J Eronen; comments to author 27.09.24; revised version received 10.11.24; accepted 21.01.25; published 19.02.25.

Please cite as:
Shin J, Kang H, Choi S, Chu SH, Choi J
Identifying Profiles of Digital Literacy Among Community-Dwelling Korean Older Adults: Latent Profile Analysis
J Med Internet Res 2025;27:e57122
URL: https://www.jmir.org/2025/1/e57122
doi:10.2196/57122
PMID:39969960

©Jiyoung Shin, Hun Kang, Seongmi Choi, Sang Hui Chu, JiYeon Choi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Health Literacy Questionnaire for Older Adults: Instrument Development and Validation Study

Xinxin Wang[1,2*], MD; Chengrui Zhang[2*], BSc; Yue Qi[3*], BA; Ying Xing[2], BSc; Yawen Liu[4], BSc; Jiayi Sun[4], BSc; Wei Luan[1], MD

[1]Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai, China

[2]Shanghai Jiao Tong University School of Nursing, Shanghai, China

[3]Department of Social Medicine, School of Health Management, Harbin Medical University, Harbin, China

[4]Shuguang Clinical Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Wei Luan, MD
Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine
Number 528, Zhangheng Road, Pudong New Area
Shanghai
China
Phone: 86 15921317982
Email: luanwei_shuguang@126.com

Abstract

Background: The integration of digital technology into older adult health and care has enhanced the intelligence of health and older adult care products and services while also transforming how seniors acquire and share health information. Assessing older adults' digital health literacy (DHL) is crucial for developing targeted interventions.

Objective: This study aims to develop and validate a DHL assessment questionnaire for older adults. It also seeks to evaluate the questionnaire's scientific validity and feasibility among community-dwelling older adults in China, providing a reliable tool for assessing their level of DHL.

Methods: A literature review, focus group discussions, and the Delphi method were used to construct the questionnaire item pool and perform item screening. Item analysis was conducted for comprehensive evaluation, and questionnaire validity was assessed through construct validity (exploratory factor analysis, confirmatory factor analysis, convergent validity, and discriminant validity), content validity, and criterion-related validity. Reliability was analyzed using Cronbach alpha coefficient, split-half reliability, and test-retest reliability.

Results: The study included 710 participants. Item analysis indicated that the questionnaire had strong discriminant validity. Correlation coefficient analysis showed that the item-total correlation coefficients ranged from 0.497 to 0.920 (P<.01). After multiple exploratory factor analyses, 6 common factors were extracted, with a cumulative variance contribution rate of 73.745%. Confirmatory factor analysis demonstrated a good model fit (χ^2/df=2.803, root-mean-square error of approximation=0.071, comparative fit index=0.907, goodness-of-fit index=0.773, incremental fit index=0.908, Tucker-Lewis index=0.901, normed fit index=0.863). The questionnaire demonstrated favorable convergent validity, content validity, and criterion-related validity. The total Cronbach α coefficient was 0.976, with dimension-specific Cronbach α coefficients ranging from 0.819 to 0.952, indicating satisfactory internal consistency. Additionally, the test-retest reliability coefficient for the total questionnaire was 0.925, demonstrating good stability over time.

Conclusions: This study developed a questionnaire specifically designed to assess DHL in older adults through a scientifically rigorous and systematic process. The questionnaire demonstrates strong psychometric properties and can serve as an empirical tool for health professionals to design personalized intervention policies and enhance health service delivery.

(J Med Internet Res 2025;27:e64193) doi:10.2196/64193

KEYWORDS

digital health literacy; digital literacy; older adults; instruments; reliability; psychological measures; questionnaire; China

XSL•FO
RenderX

Introduction

Health literacy (HL) is the ability of an individual to access, understand, evaluate, and apply health information to make appropriate health decisions. It can help people make rational judgments about health care, disease prevention, and health promotion in their daily lives, thereby maintaining or enhancing an individual's quality of life throughout the life span [1]. Digital health literacy (DHL) is an expansion of HL in the digital era, which refers to an individual's ability to access, process, communicate, and understand health information and services in the context of the use of digital technology and information to promote and improve individual and collective health through effective health decision-making [2]. Different from HL, DHL emphasizes the technical skills required in the digital environment, such as using digital devices and the internet to search for health information, operating health-related apps, identifying and assessing the reliability of online health information, safeguarding the privacy and security of personal health information, and communicating effectively on digital platforms. In addition, DHL involves using critical thinking in digital environments, using digital tools for self-management and health monitoring, and adapting learning to the development of digital technology and health information, which are not included in the concept of HL [2,3]. Studies have shown significant associations between DHL and anxiety [4,5], health-promoting behaviors [6], self-efficacy [6,7], and self-care competence [7] in the older adult population. After controlling for sociodemographic variables, DHL in older adults was found to be able to directly and positively influence health-related quality of life in the study by Liu et al [8]. Kim et al [9] stated that there is a positive association between DHL and health-related behaviors. DHL in older adults significantly promotes health behaviors, health support behaviors, and disease management behaviors. Arcury et al [10] found a positive correlation between DHL in older adults and individual health knowledge, attitude, and computer use.

In recent years, digital technology has developed rapidly and has been increasingly applied to the field of older adults' health and aging. This advancement has enhanced the intelligence of health and aging products while also requiring a higher level of DHL for older adults to adapt to society and promote health [11]. To date, 5 DHL evaluation tools have been used for older adults [3]. Among them, the eHealth Literacy Scale (eHEALS) [12] is the most widely used DHL assessment tool, primarily measuring an individual's self-perceived ability to use information technology to process health information. However, it has a relatively limited scope of skills for assessment. By contrast, the electronic Health Literacy Scale [13] and the Digital Health Literacy Instrument [14] focus on evaluating individuals' operational skills in gathering online health information. However, these 2 tools may only partially cover the dimensions of DHL, particularly the practical skills required for operating digital devices and interactive communication. While the eHealth Literacy Questionnaire [15] was designed to keep pace with the growing prevalence of digital technology, it does not sufficiently address the security of personal information, content creation, and information exchange. The Digital Health Literacy Assessment [16] scale primarily focuses on using computers, smartphones, and other devices to access and assess health information. However, it falls short of accommodating the rapidly evolving demands of digital health capabilities. The emergence of new digital technologies may extend beyond the scope of assessment in existing DHL assessment tools. If assessment tools fail to accurately measure older adults' DHL levels, they may struggle to effectively use digital health information resources and fully engage with digital tools for health management. This could lead to deficiencies in understanding health issues, difficulties in making informed health decisions, and an increased risk of health inequalities among older adults.

The Digital Competence Framework (DIGCOMP) was developed by the European Commission's Joint Research Centre in 2013 to identify and define digital competencies relevant to all citizens living and working in Europe. The framework outlines 5 key competence areas: information, communication, content creation, safety, and problem-solving [17]. The framework provides a comprehensive and systematic approach to assessing digital competence, enabling comparisons across regions and groups. It also categorizes different levels of digital competence and encourages individuals to engage in lifelong learning to continuously refine and enhance their digital skills. Developed through extensive research and practice, DIGCOMP encompasses a range of digital competencies at different levels, from basic to advanced, allowing assessment tools to align with older adults' varying skill levels. Additionally, the DIGCOMP framework covers various aspects of digital competence and can be adapted to meet the specific needs and real-life situations of older adults, ensuring better alignment with their characteristics. Studies have been conducted to develop and validate assessment questionnaires for the everyday digital literacy of older adults based on the DIGCOMP framework, demonstrating good psychometric properties. These questionnaires have been successfully used to assess the digital literacy of older adults in Korea [18].

Because of declining cognitive function and changes in health status, older adults have an increasing need for health information and services. As the digital health domain continues to evolve, health management processes and efficiencies are being optimized. However, this also presents the challenge of a digital divide for older adults. Differences in educational background and technology acceptance lead to significant variations in their digital health capabilities [19]. Therefore, accurately assessing the digital health competence of older adults and implementing targeted interventions are crucial for improving their DHL. Currently, no digital health assessment tools specifically designed for older adults have been developed. In this context, our study constructed a questionnaire to assess DHL in older adults, using DIGCOMP as a theoretical framework while considering their cognitive and physiological characteristics. This questionnaire aims to identify the current level of DHL among older adults, providing a scientific basis for health care professionals to develop personalized interventions and health management policies, as well as to optimize health service processes.

XSL•FO
RenderX

Methods

Building DHL Questionnaire Entries for Older Adults

The research group conducted a scoping review to comprehensively gather relevant research findings on DHL evaluation questionnaires for older adults [3]. The initial literature search was performed in January 2022, followed by a supplementary search in June 2022. After the screening process, 16 published articles were identified that included evaluations of DHL in older adults, encompassing 5 assessment tools. Our analysis revealed that previous studies on DHL among older adults have primarily focused on assessing their ability to access, evaluate, and apply health information, while lacking assessments of their ability to communicate, integrate health information, and protect their privacy. Furthermore, no assessment tools specifically designed to address the unique characteristics of the older population have yet emerged. Based on these findings, we developed a new questionnaire with particular emphasis on the above dimensions, considering the cognitive and operational characteristics of the older population. Semistructured interviews were conducted before expert consultation (Multimedia Appendix 1). The corresponding author (WL) recruited experts through her personal network and applied a snowball sampling method to further expand the participant pool. The criteria for selecting experts for the semistructured interviews are listed in Textbox 1.

Textbox 1. Criteria for selecting experts for the semistructured interviews.

1. Health management of older adults
   - Experts were required to have over 10 years of experience in the field, hold at least an intermediate professional title, possess a postgraduate degree or higher, voluntarily participate, and be able to complete consultations in a timely manner.
2. Electronic health and digital health
   - Experts were required to have more than 5 years of experience, hold at least an intermediate professional title, have a bachelor's degree or higher, voluntarily participate, and be able to complete consultations in a timely manner.

A total of 6 experts participated in interviews and group discussions. These included a PhD supervisor specializing in health management, a chief physician in community general medicine, a chief physician in geriatric medicine, a master's supervisor specializing in public health, a master's supervisor in geriatric nursing, and a chief nurse. The interview outline included the following questions:

- What do you think should be included in the evaluation index of DHL for older adults?
- What fields of expertise do you think should be consulted via correspondence, and for what reasons?

The original DHL questionnaire was developed based on expert interviews. Five rounds of group discussions were conducted to refine the questionnaire, ensuring its alignment with China's national context and cultural background. Subsequently, the first version of the DHL Questionnaire was formulated, using DIGCOMP as its theoretical framework.

Validation of Questionnaire Contents

This study used expert consultation to validate the questionnaire's content. Sixteen experts from Shanghai's tertiary hospitals, colleges, universities, and technology companies—who have long been engaged in geriatric care, geriatrics, internet and health care, health policy, and computing—were selected for this study. The criteria for selecting experts were the same as those used for the semistructured interviews; however, the experts themselves were different. The experts' questionnaire consisted of 3 parts (Textbox 2).

Textbox 2. Parts of the experts' questionnaire.

1. Introduction
   - Outlining the purpose of the study, instructions for completing the questionnaire, details on how and when to submit it, contact information, and acknowledgments.
2. Digital health literacy evaluation questionnaire correspondence form for older adults
   - Experts rated the importance and relevance of each dimension and item on a scale of 1 to 5 (ranging from "very unimportant" to "very important") and 1 to 4 (ranging from "very irrelevant" to "very relevant"), with 2 additional columns—"Opinions and Reasons for Indicator Modification or Deletion" and "Items to be Added"—for experts to provide modification suggestions.
3. General information and expert authority questionnaire
   - Collecting basic information about the experts, along with their self-assessments of their authority and familiarity with the questionnaire content.

Three rounds of expert correspondence were conducted, with questionnaires distributed and returned via WeChat (Tencent Holdings Limited) or email. Upon receipt, the research team checked for completeness and verified any ambiguities or omissions with the experts to ensure the validity of the responses. The inclusion criteria for the entries were as follows:

XSL•FO
RenderX

a mean importance score >3.5, a total score rate >20%, and a coefficient of variation <0.25 [20]. Entries were adjusted based on discussions within the research team, incorporating expert modifications. After each round, responses were compiled, and items with significant opinion divergences were identified. In the subsequent round, these contentious items, along with summaries of expert opinions (median and standard deviation), were presented to encourage reflection and adjustment, gradually moving toward consensus. The study used Kendall *W* to measure the consistency of evaluations among experts, with the consultation stopping when Kendall *W* ranged between 0.4 and 0.5, indicating a satisfactory level of consensus [21]. The final DHL questionnaire for older adults included 6 dimensions and 46 items.

Evaluation of the Reliability and Validity of the Questionnaire

Participants and Data Collection

Given the high acceptance and response rate among older adults following advocacy by community health workers, and the fact that those able to attend the physical examination in person generally met the study's inclusion criteria, convenience sampling was used. The inclusion criteria were as follows: (1) signed informed consent and willingness to participate in the research; (2) aged 60 years or older and residing in the community for over 6 months; (3) possession of full communicative language abilities, including listening comprehension, oral expression, reading comprehension, and written expression. The exclusion criteria were (1) diagnosed mental disorders (eg, schizophrenia, delusions) or cognitive disorders (eg, Alzheimer disease, vascular cognitive impairment) and (2) diagnosis of a major or advanced terminal illness, such as a malignant tumor or end-stage renal disease.

Our study utilized an online survey method, with uniformly trained investigators rigorously selecting participants based on the inclusion and exclusion criteria. After eligible older adults completed their physical examination, the investigator explained the purpose and significance of the study. Participants were guided on how to navigate the questionnaire interface, key considerations for completing the questionnaire, and the rules for successful submission. Older adults completed the questionnaire independently on the spot. To prevent duplicate participation, the questionnaire was configured to allow only 1 submission per WeChat user. After the survey concluded, researchers conducted a quality check on all questionnaires. A preexperiment revealed that it was nearly impossible for older adults to complete the questionnaire within 2 minutes while carefully reading each question. Therefore, to ensure data quality, questionnaires completed in under 2 minutes were excluded.

The sample size was determined based on the principle that factor analysis requires at least 5-10 times the number of questionnaire items [22]. Given that the questionnaire contains 46 items and considering a 20% increase for potential invalid responses, the sample size was calculated as $46 \times 5 \times (1 + 20\%) = 276$. As both exploratory and confirmatory factor analyses require 2 independent samples, the minimum required sample size for the study was 552. Study participants were selected from August to November 2023 among 710 older adults undergoing their annual physical examination in Shanghai. The obtained samples were randomly divided into 2 groups for pretesting (n=355) and formal testing (n=355).

Ethics Approval

The study was approved by the Ethics Committee of Renji Hospital, affiliated with Shanghai Jiao Tong University School of Medicine (approval number RA-2021-465). Participants understood the purpose, process, risks, and benefits of the study and voluntarily agreed to participate. The survey adhered to the principles of anonymity and confidentiality, and responses were collected anonymously. Health science popularization books were provided to participants as compensation.

Instruments

In our study, the questionnaire was in Chinese and consisted of 4 parts (Multimedia Appendix 2): (1) informed consent form; (2) general information questionnaire (designed by the researcher), which included information on gender, age, education level, marital status, residential situation, household monthly income, the primary source of income, and medical payment methods; (3) the eHEALS (developed by Norman et al [12]), which consists 8 questions assessing the ability to apply online health information and services, judgment ability, and decision-making ability. It is scored on a 5-point Likert scale, with a total score ranging from 8 to 40, where higher scores indicate better eHealth literacy (eHL). The scale demonstrated high internal consistency (α=.967); (4) Digital Health Literacy Questionnaire for Older Adults (designed by researchers WL and XW), which includes 6 dimensions: Information (14 items), Interaction (11 items), Content (3 items), Safety (7 items), Attitude (4 items), and Behavior (7 items). A 5-point Likert scale was used, with each item scored from 1 to 5, ranging from "very inconsistent" to "very consistent," where higher scores indicate better DHL.

Statistical Analysis

Excel 2019 (Microsoft Corporation) was used for raw data entry, and SPSS 25.0 (IBM Corp.) and AMOS 26.0 (IBM Corp.) statistical software were used for statistical analysis. Sociodemographic information was expressed as the number of cases and percentages. Given that all variables in our study were bivariate normally distributed and exhibited a linear relationship, we used Pearson correlation analysis.

The first pretest sample (N=355) was subjected to item analysis, exploratory factor analysis (EFA), and further screening of entries to develop the formal questionnaire. The questionnaire items were comprehensively evaluated based on the item analysis screening criteria listed in Textbox 3.

XSL•FO
RenderX

Textbox 3. Item analysis screening criteria.

1. Critical ratio method
   - The total questionnaire score was divided into high and low groups, comprising the top 27% and bottom 27%, respectively. The critical value (*t* value) for each item between these groups was compared, with a requirement of $t \geq 3$ and a statistically significant difference.
2. Correlation coefficient method
   - The correlation coefficient between each item and the total score, as well as the corrected item-total correlation coefficient, should be ≥0.4.
3. Homogeneity test
   - The Cronbach α coefficient of the total questionnaire should not increase after deleting any item. The factor loading of each item on the common factors of the questionnaire should be ≥0.4, and the communality should be ≥0.2. An item was deleted if it failed to meet at least three of these 6 indicators [23].

The Kaiser-Meyer-Olkin (KMO) test and Bartlett test of sphericity were used to assess the adequacy of the data. A KMO value close to 1 and $P<.05$ after the Bartlett test indicate that the original variables are suitable for factor analysis. Principal component analysis was used to extract common factors with eigenvalues greater than 1, combined with the scree plot for judgment. The varimax method of orthogonal rotation was applied to enhance factor loadings, with strong associations considered to exist when factor loadings were ≥0.4. If an item had factor loadings ≥0.4 across multiple dimensions, it was assigned to the dimension with the highest factor loading [24]. The deletion criteria were as follows: (1) items with loadings <0.4 on their respective factors; (2) items with communalities <0.2; (3) items with multiple loadings of similar values across dimensions; (4) items with factor loadings inconsistent with the original conceptual framework and inexplicable; and (5) dimensions with fewer than 3 items [25].

Validation factor analysis, convergent and discriminant validity tests, content validity tests, validity scale assessments, and internal consistency reliability tests were conducted on the second round of formal questionnaire test samples (N=355) to verify the reliability and validity of the questionnaire.

To further verify the structural validity of the questionnaire, confirmatory factor analysis was conducted using the maximum likelihood method to estimate model parameters. The chi-square to degrees of freedom ratio (χ^2/df), root-mean-square error of approximation, goodness-of-fit index, adjusted goodness-of-fit index, comparative fit index, incremental fit index, Tucker-Lewis index, and normed fit index were used to evaluate the model [26]. When the average variance extracted (AVE) is greater than 0.5 and the composite reliability exceeds 0.7, the questionnaire demonstrates ideal convergent validity. Additionally, when the correlation between questionnaire dimensions is significant but remains lower than the square root of the corresponding AVE, the questionnaire exhibits desirable discriminant validity [27].

In this study, experts from relevant fields (N=16) were invited to evaluate the content validity of the questionnaire. The item-level content validity index (I-CVI) represents the proportion of experts who rated each item as 3 or 4. The scale-level content validity index (S-CVI/Ave) is the average of all I-CVIs. When the number of experts exceeds 6, an I-CVI>0.78 and an S-CVI/Ave>0.90 are generally considered to indicate good relevance of the items and questionnaire to the conceptual content being measured [28].

The eHEALS was used as the criterion tool in this study. A positive and statistically significant correlation between the criterion tool and the DHL questionnaire for older adults, both in total score and across dimensions, indicates that the criterion-related validity is satisfied [29].

Cronbach α coefficient and Spearman-Brown split-half reliability are the most commonly used methods for testing the internal consistency reliability of questionnaires. Generally, a Cronbach α coefficient greater than 0.500 for each dimension and a total Cronbach α coefficient and Spearman-Brown split-half reliability greater than 0.700 indicate acceptable reliability. Retest reliability analyses were conducted with a retest sample (N=20) [29]. For all tests, the significance level was set at $\alpha=.05$.

Results

Characteristics of Participants

The study included 710 participants, consisting of 354 males (49.9%) and 356 females (50.1%). Participants ranged in age from 65 to 100 years, with a mean age of 70.7 (SD 6.3) years. The results are presented in Table 1.

XSL•FO
RenderX

Table 1. Participant characteristics (N=710).

Variables	Group 1 (N=355), n (%)	Group 2 (N=355), n (%)	Between groups 1 and 2	
			Chi-square (*df*)	*P* value
Gender			0.09 (1)	.77
Male	179 (50.4)	175 (49.3)		
Female	176 (49.6)	180 (50.7)		
Age			3.529 (3)	.32
60-65	210 (59.2)	231 (65.1)		
66-70	83 (23.4)	78 (22.0)		
71-75	30 (8.5)	22 (6.2)		
＞76	32 (9.0)	24 (6.8)		
Educational level			3.772 (3)	.29
Junior high and below	188 (53.0)	164 (46.2)		
High school/secondary vocational	126 (35.5)	138 (38.9)		
College/university	37 (10.4)	47 (13.2)		
Postgraduates	4 (1.1)	6 (1.7)		
Marital status			4.198 (2)	.12
Married	287 (80.8)	304 (85.6)		
Divorce	24 (6.8)	13 (3.7)		
Widow	44 (12.4)	38 (10.7)		
Residential situation			7.786 (4)	.10
Living alone	46 (13.0)	42 (11.8)		
Living with spouse	146 (41.1)	172 (48.5)		
Living with children	73 (20.6)	49 (13.8)		
Living with spouse and children	86 (24.2)	90 (25.4)		
Other	4 (1.1)	2 (0.6)		
Household monthly income			7.662 (3)	.054
<4000 CNY[a]	108 (30.4)	126 (35.5)		
4001-6000 CNY	110 (31.0)	104 (29.3)		
6001-8000 CNY	97 (27.3)	71 (20.0)		
>8000 CNY	40 (11.3)	54 (15.2)		
Primary source of income			5.541 (3)	.14
Pension	306 (86.2)	323 (91.0)		
Child support	27 (7.6)	22 (6.2)		
Reemployment income	12 (3.4)	6 (1.7)		
Other	10 (2.8)	4 (1.1)		
Medical payment methods			6.681 (3)	.08
Urban Medical Insurance	249 (70.1)	275 (77.5)		
Rural Medical Insurance	66 (18.6)	42 (11.8)		
Employee Medical Insurance	36 (10.1)	34 (9.6)		
Other	4 (1.1)	4 (1.1)		
Chronic disease			6.613 (3)	.08
Without	62 (17.5)	58 (16.3)		
1	157 (44.2)	135 (38.0)		

XSL•FO
RenderX

Variables	Group 1 (N=355), n (%)	Group 2 (N=355), n (%)	Between groups 1 and 2	
			Chi-square (*df*)	*P* value
2	88 (24.8)	90 (25.4)		
≥3	48 (13.5)	72 (20.3)		

[a]1 CNY=US $0.14.

Item Analysis

The results of the item analysis are summarized in Multimedia Appendix 3. In the extreme group comparisons, all items showed statistically significant differences between the high- and low-score groups (P<.001). Correlation coefficient analysis indicated that item-total correlation coefficients ranged from 0.468 to 0.744, while corrected item-total correlation coefficients ranged from 0.442 to 0.725 (P<.001). The Cronbach α coefficient for the total questionnaire was 0.981, and no increase was observed after deleting any item. Factor analysis revealed that item factor loadings ranged from 0.365 to 0.697, with communalities ranging from 0.446 to 0.677. Ultimately, 3 items failed to meet 1 criterion; however, they did not meet the deletion threshold. Therefore, all 46 items were retained for further analysis.

Validity Analysis

Construct Validity

Exploratory Factor Analysis

EFA was conducted on the 46 items, yielding a KMO value of 0.955 and a significant Bartlett test of sphericity (χ^2_{1035}=8299.714, P<.001), indicating that the data were suitable for factor analysis. Six common factors with eigenvalues >1 were extracted, accounting for 53.810% of the cumulative variance. The scree plot also confirmed that retaining 6 factors was appropriate. EFA was performed multiple times, setting the number of factors to 6, and noncompliant items were gradually removed. Ultimately, 39 items remained, forming 6 factors with a cumulative variance contribution rate of 73.745%. The factor loadings for the items in each dimension ranged from 0.452 to 0.815, as detailed in Multimedia Appendix 4.

Confirmatory Factor Analysis

The results indicated that the model was identifiable and successfully converged but initially exhibited poor goodness of fit. The modification indices revealed high correlations between 5 pairs of error terms, each with values greater than 10, suggesting potential local dependence issues in the model [26]. Upon further analysis, we found that while these item pairs were designed to measure distinct concepts, they exhibited some conceptual overlap. This finding was consistent with the high modification indices, indicating that the error terms might not be entirely independent as initially assumed. To address this, and after confirming that the model's complexity was not artificially inflated, we sequentially added 5 error covariances to account for these observed relationships. Following these adjustments, the standardized regression coefficients of the revised model ranged from 0.538 to 0.931, indicating a good model fit. The model fit indices are presented in Table 2, and the confirmatory factor analysis model diagram is shown in Figure 1.

XSL•FO
RenderX

Figure 1. The confirmatory factor analysis model diagram.

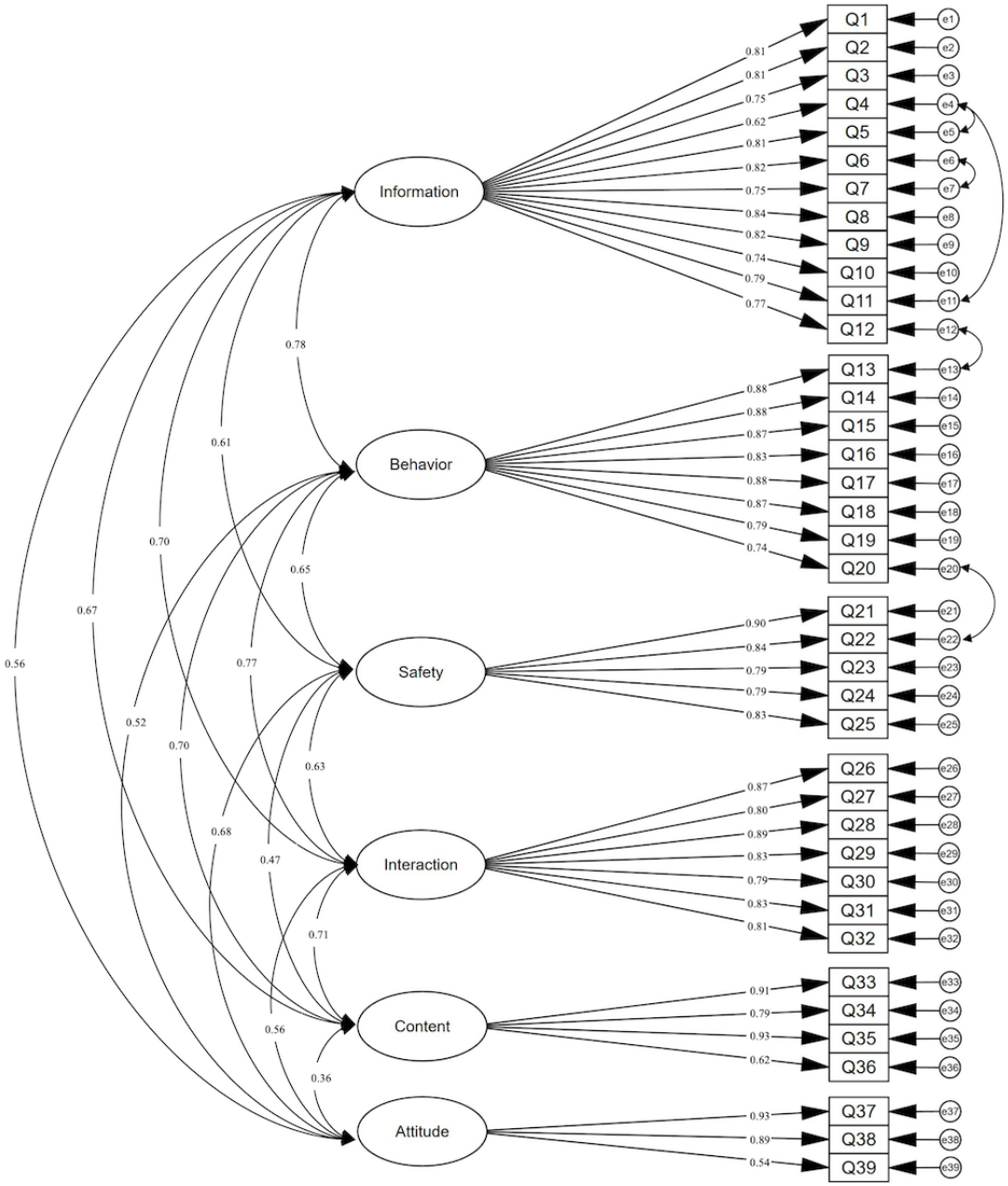

Table 2. Goodness-of-fit indices for confirmatory factor analysis of the digital health literacy questionnaire in older adults.

Item	Chi-square/*df*	Root-mean-square error of approximation	Comparative fit index	Goodness-of-fit index	Incremental fit index	Tucker-Lewis index	Normed fit index
Reference standard	<5	<0.08	>0.9	>0.9	>0.9	>0.9	>0.9
Initial model	3.404	0.082	0.875	0.733	0.876	0.855	0.833
Revised model	2.803	0.071	0.907	0.773	0.908	0.901	0.863

Convergent and Discriminant Validity

Convergent Validity

Based on the well-fitted model, the convergent validity of the questionnaire was tested, with results presented in Table 3. The factor loadings for the items across the 6 dimensions ranged from 0.538 to 0.931, all exceeding 0.500, indicating strong representativeness of the items within their respective dimensions. Additionally, the AVE for each dimension ranged from 0.608 to 0.768, all above the 0.500 threshold, while the composite reliability values ranged from 0.841 to 0.952, all exceeding 0.700, confirming the excellent convergent validity of the questionnaire.

XSL•FO
RenderX

Table 3. Convergent validity of the digital health literacy questionnaire in older adults.

Factors and items	Factor loadings	Average variance extracted	Composite reliability
Information		0.608	0.949
1. I pay attention to whether health information is released and disseminated by official or authoritative institutions	0.814		
2. I compare similar health information	0.809		
3. I verify the correctness of health information from other sources	0.752		
4. I remain vigilant about the health information I obtain and do not easily believe it	0.618		
5. I understand digital health technologies (eg, wearable devices, smart health electronic products)	0.810		
6. I understand that digital technologies can be used for health management or health promotion (such as health and medical mobile apps)	0.817		
7. I do not immediately share health information with others after receiving it but first check the content	0.751		
8. I can browse, search, and obtain health information through digital devices or software	0.844		
9. I pay attention to updates on health information	0.818		
10. I can judge whether health information is related to commercial interests (eg, contains product advertisements)	0.742		
11. I understand that digital health devices or software can be used to store personal health information	0.788		
12. I have used digital health devices or software to record personal health information	0.768		
Behavior		0.714	0.952
13. I know how to use digital health tools to track my health behavior	0.880		
14. I can judge whether digital health tools are trustworthy	0.879		
15. During the use of digital health tools, I can adjust my frequency, intensity, and methods based on the actual situation	0.872		
16. I am used to using digital services to handle health information	0.832		
17. I can use digital devices or electronic health products or software	0.878		
18. I know when, how, and what health information to use	0.874		
19. If necessary, I think I can persist in using digital health tools	0.790		
20. I believe the use of digital technologies is beneficial for my health management	0.743		
Safety		0.688	0.917
21. I believe I have the right to pursue legal responsibility for unauthorized data acquisition or improper data storage that leads to data breaches	0.895		
22. I believe I have ownership of personal data, and others can only obtain my personal health data with my authorization	0.840		
23. I do not click on unsafe web links; I do not visit websites that are flagged as risky	0.787		
24. I can avoid health risks related to the use of digital technologies that threaten physical and mental health	0.792		
25. I know the potential security risks in the online environment	0.829		
Interaction		0.693	0.940

XSL•FO
RenderX

Factors and items	Factor loadings	Average variance extracted	Composite reliability
26. I can share information with others on the internet	0.869		
27. I can share the information I obtained online with others offline	0.803		
28. I can use digital devices or software to communicate health information with others	0.894		
29. I am familiar with the user interface of digital devices or software	0.829		
30. I use information dissemination platforms (eg, Weibo, WeChat Moments) to share information	0.792		
31. I can use digital devices or software to communicate health information with artificial intelligence	0.825		
32. I imitate the health-promoting behaviors or health management methods mentioned in health information	0.811		
Content		0.768	0.908
33. I can edit and improve health content created by myself or others	0.907		
34. I can protect the integrity of original works and cite sources when referencing	0.787		
35. I can integrate health information from multiple sources and rephrase it	0.929		
36. In the past 12 months, I have participated in online health lectures and health care experience sharing activities	0.621		
Attitude		0.650	0.841
37. I adhere to the correct political direction in online behavior	0.931		
38. I do not fabricate or spread false, unverified health information	0.891		
39. I care about health information related to myself	0.538		

Discriminant Validity

Further examination of discriminant validity showed that all questionnaire dimensions were significantly correlated, with correlation coefficients ranging from 0.363 to 0.779 (P<.001). These values were all lower than the square roots of the corresponding AVEs, indicating that while the dimensions are interrelated, they maintain sufficient discriminant validity. This confirms that the questionnaire structure demonstrates ideal discriminant validity. See Table 4 for details.

Table 4. Discriminant validity of the digital health literacy questionnaire in older adults.

Dimensions	Information	Behavior	Safety	Interaction	Content	Attitude
Information	1	—[a]	—	—	—	—
Behavior	0.779[b]	1	—	—	—	—
Safety	0.614[b]	0.650[b]	1	—	—	—
Interaction	0.704[b]	0.773[b]	0.629[b]	1	—	—
Content	0.660[b]	0.704[b]	0.465[b]	0.708[b]	1	—
Attitude	0.555[b]	0.518[b]	0.676[b]	0.560[b]	0.363[b]	1
Square root of average variance extracted	0.779	0.845	0.829	0.832	0.876	0.806

[a]Not applicable.

[b]P<.01.

Content Validity

Based on expert ratings, the I-CVI of the questionnaire ranged from 0.81 to 1.000 and S-CVI/Ave was 0.92, both meeting the required standards, indicating good content validity.

Criterion-Related Validity

The criterion validity analysis showed that the total and dimension scores of the Digital Health Literacy Questionnaire for Older Adults were positively correlated with the total and dimension scores of the eHEALS, with correlation coefficients ranging from 0.363 to 0.980, all statistically significant (P<.01).

XSL•FO
RenderX

This demonstrates good criterion-related validity of the questionnaire, as shown in Multimedia Appendix 5.

Reliability Analysis

Cronbach α Coefficient

The results showed that the total questionnaire's Cronbach α coefficient was 0.976, and each dimension's Cronbach α coefficient ranged from 0.819 to 0.952, demonstrating high internal consistency of the questionnaire.

Split-Half Reliability

The results showed that the total questionnaire's Spearman-Brown split-half reliability coefficient was 0.925, and the Spearman-Brown split-half reliability coefficients for each dimension ranged from 0.739 to 0.956, indicating good internal consistency of the questionnaire.

Test-Retest Reliability

In this study, test-retest reliability was assessed by readministering the questionnaire to 20 individuals 2 weeks after the initial administration. The results showed that the total questionnaire's test-retest reliability coefficient was 0.925, while the coefficients for each dimension ranged from 0.875 to 0.933, all exceeding 0.700. These findings indicate that the questionnaire exhibits good stability over time. See Table 5 for detailed results.

Table 5. Reliability analysis of the digital health literacy questionnaire in older adults.

Dimensions	Item numbers	Cronbach α coefficient	Split-half reliability	Test-retest reliability
Information	12	0.950	0.946	0.903
Behavior	8	0.952	0.956	0.933
Safety	5	0.916	0.837	0.897
Interaction	7	0.939	0.888	0.880
Content	4	0.880	0.888	0.891
Attitude	3	0.819	0.739	0.875
Digital health literacy	39	0.976	0.925	0.925

Discussion

DIGCOMP-Based Framework for Assessing Digital Health Literacy

The study utilized the DIGCOMP framework as a conceptual foundation, integrating expert opinions and empirical research to refine the structure and content of the questionnaire. The final instrument assesses 6 dimensions of DHL and demonstrates strong psychometric properties. This provides a solid empirical basis to support health professionals in developing individualized intervention strategies and optimizing health service delivery.

Questionnaire Formation Is Meaningful

As the global population ages, older adults are becoming a priority in the health care sector [30]. With the rapid advancement of digital technology, they face multiple challenges in managing their health. Cognitive decline and limited digital training can make it difficult for them to use digital tools for health management. This can hinder their access to essential health information and resources, as well as services such as online appointments, telemedicine consultations, and electronic health records. As a result, these barriers may increase the risk of medication errors and exacerbate health inequalities.

From the perspective of enhancing older adults' engagement in health management, improving DHL is a top priority in developing a digital service system for older adults [11]. The concept of DHL is continuously evolving and requires distinct competencies compared with eHL. Although both are related, they differ significantly in their practical application for older adults. eHL primarily focuses on their ability to process and apply health information online, such as searching for disease-related information, comparing data from multiple authoritative sources, and identifying false medical advertisements. It emphasizes critical thinking and the ability to assess the authenticity of health information [14,31]. By contrast, DHL emphasizes older adults' ability to effectively use digital health devices, such as smart blood pressure monitors or health tracking bracelets, to manage their health data. It involves applying digital tools to adjust exercise routines, modify daily diets based on professional advice, and engage in self-care. This requires a certain level of technical competence. Additionally, DHL highlights the importance of exchanging and evaluating health information, as well as ensuring privacy protection in digital interactions [3]. Therefore, if an assessment tool is not designed based on the latest DHL concepts, it may fail to align with the essential skills older adults need to develop and demonstrate. There will be no direct comparison between the results of different studies, which may limit their reproducibility and generalizability. However, most existing assessment tools for older adults have been developed based on eHL, without considering the unique characteristics of this population [12-16]. Currently, there is no specific DHL assessment tool designed for older adults. The questionnaire developed in this study addresses this gap by evaluating the current status of DHL in older adults and identifying areas where digital health competencies need improvement. This provides a foundation for developing appropriate intervention strategies and targeted health education, making it highly relevant in the current digital health landscape.

XSL•FO
RenderX

Questionnaire Construction Process Is Reasonable

This study conducted a comprehensive review of relevant literature and government policy documents. Using the widely accepted DHL concept, the DIGCOMP framework was adopted as the theoretical foundation to ensure the questionnaire's scientific rigor. The initial framework was developed through face-to-face semistructured interviews with 6 experts and 5 rounds of research group discussions. To ensure content validity, 3 rounds of expert consultations were conducted with 16 specialists from hospitals, academic institutions, and technology companies. These experts, with extensive experience in geriatric care, provided valuable insights into the characteristics and needs of older adults. The recommendations provided by these experts were comprehensive and highly instructive. The questionnaire development process emphasized aspects that had been overlooked in previous assessment tools, including interaction (both with others and with devices), content editing (such as information integration and copyright awareness), and security (covering privacy protection, device security, and health-related safety). The questionnaire items were designed to align with the realities of older adults' daily lives and health care experiences, focusing on potential challenges they may face when using digital technologies to access, process, communicate, and understand health information and services. To ensure its relevance and usability, a presurvey was conducted with 10 older adults from diverse backgrounds. Through individual interviews, participants' experiences with completing the questionnaire were observed and recorded. Based on these insights, the questionnaire was simplified in terms of language and optimized in layout to enhance its applicability and improve data collection efficiency. Additionally, complex terminology was replaced with more accessible wording to make it easier for older adults to respond. For essential terms that needed to be retained but might be difficult to understand, clear explanations were added to ensure clarity.

The Contents of the Questionnaire Are Scientific

The study refined the questionnaire entries through empirical research. The research team conducted multiple discussions on the professional significance of each entry, making modifications and deletions based on predefined criteria and expert knowledge.

The "Information" dimension evaluates older adults' ability to access, assess, and utilize health information through digital devices, as well as their understanding of digital health. This dimension is a core component of DHL measurement. In our study, 12 of the original 14 items in this dimension were retained in the final questionnaire. The item "I will check if the health information is of value to me" was removed after discussions with the research team, as it was semantically similar to other items in the same dimension and had a factor loading below 0.4. Additionally, the item "I care about health information related to myself" was reassigned to the "Attitude" dimension.

The "Behavior" dimension assesses older adults' ability to manage their health using digital devices and their understanding of how digital technology can support their well-being [32]. During the revision process, the item "I think digital health information is trustworthy," originally part of the "Attitude" dimension, was removed, while the remaining items were reassigned to the "Behavior" dimension. Additionally, through expert consultations and group discussions, the "Problem-Solving" dimension from the DIGCOMP framework was subdivided into the "Attitude" and "Behavior" dimensions. As a result, the empirical study justified the reclassification of certain items between these 2 dimensions.

The "Safety" dimension assesses older adults' ability to protect their personal information and physical and mental well-being while managing their health through digital devices. This competency, which has not been explicitly addressed in the concept of eHL [3], is crucial for ensuring the safe and effective participation of older adults in digital health activities [17,19]. During the questionnaire evaluation, most items remained within this dimension. However, the item "I believe that personal health data can be accessed by medical staff directly involved in the treatment" was removed due to insufficient factor loading. Additionally, the item "I know how to protect personal digital devices from cyber attacks" exhibited double loading with similar values. Apart from these adjustments, no attributional changes were made to the remaining items.

The "Interaction" dimension in evaluating older adults' DHL emphasizes their ability to communicate health information with individuals, health care platforms, artificial intelligence, and other digital systems. Previous research has primarily focused on older adults' ability to communicate with people in the context of DHL [33], likely due to the traditional perspective that health information exchange occurs mainly between individuals. However, with advancements in technology, the ability of older adults to effectively interact with digital devices is not only crucial for accessing timely and accurate health information but also directly influences the quality of their health management and self-care. During the evaluation of questionnaire entries, 1 item exhibited multiple loadings, and 3 items were reassigned to other dimensions after discussion. No items from other dimensions were added to this category.

The "Content" dimension emphasizes older adults' ability to integrate and enhance digital health information. While it is related to other dimensions, it is distinct in its focus on evaluating overall competencies [17]. In this dimension, we incorporated the item, "I have participated in online health knowledge seminars, health care experience-sharing activities, and similar events in the past 12 months," which was originally part of the "Interaction" dimension. Although each dimension of the DIGCOMP framework has unique characteristics, there are inherent overlaps and interconnections between them [17]. The "Content creation" competency in the DIGCOMP framework includes "Sharing information and content," which aligns with the competencies assessed in the "Content" dimension. As interaction serves as a prerequisite for content creation and addressing health concerns [17], this reallocation is justified.

Additionally, we identified 2 entries from the original "Interaction" dimension—"I can adhere to the correct political direction in my online behavior" and "I can refrain from fabricating or spreading false and unverified health information"—as well as 1 entry from the former "Information"

XSL•FO
RenderX

dimension—"I care about health information related to myself." These 3 entries differ significantly in connotation from the other items in their respective dimensions. After an in-depth group discussion, we consolidated them into a newly established "Attitude" dimension. This dimension was created to assess older adults' sense of responsibility, political sensitivity, and concern for personal health information when managing digital health information.

The Implications of This Study for Future Research

The DHL questionnaire for older adults developed in our study provides future researchers with a standardized measurement tool, facilitating comparative studies across different regions and populations. Our findings also underscore the importance of enhancing DHL in older adults. Future research can build on our study to further refine the questionnaire's structure and content. Additionally, researchers can explore targeted educational interventions in greater detail, such as developing digital skills training programs tailored for older adults and designing user-friendly digital health education resources. Additionally, researchers can use these findings to develop more targeted policies, such as promoting digital inclusion initiatives for older adults and establishing a comprehensive community support system. This could include community health education programs, volunteer service networks, and mutual aid groups. Such measures will effectively enhance the DHL of older adults, ultimately improving their overall health and quality of life and supporting the goal of healthy aging.

Limitations

There are some limitations to our study. The questionnaire developed was a self-assessment tool, which may lack objectivity. Older adults completing self-assessment questionnaires might select responses they believe are more socially acceptable or aligned with the investigators' expectations rather than their true thoughts, feelings, or behaviors regarding digital health. This tendency could introduce social desirability and response biases, potentially reducing data accuracy. Additionally, self-assessments may be affected by recall bias, as older adults might struggle with memory inaccuracies or overestimate their DHL, leading to discrepancies between their reported and actual abilities. Methods such as scenario modeling could be considered for objective evaluation in the future. Additionally, our study used a web-based survey, requiring older adults to complete the questionnaire independently, which may have excluded those with poorer numerical ability and health. These individuals may prefer obtaining health information through traditional channels (such as television, newspapers, and face-to-face communication), have lower familiarity with and less frequent use of the internet and digital technologies, and may face difficulties in accessing and interpreting health information. However, excluding them may result in health needs and issues specific to certain groups of older adults being inadequately explored, potentially affecting the ubiquity and applicability of the findings. Future research could incorporate both online and traditional face-to-face surveys. Additionally, a more concise and comprehensible questionnaire should be designed, and older adults should be provided with the necessary digital skills training to facilitate their participation in research. Furthermore, this questionnaire has only undergone psychometric testing among Chinese older adults. Therefore, cross-linguistic validation is needed to confirm its cultural invariance.

Conclusions

As scientific and technological advancements intertwine with theoretical research developments, the concept of DHL must be continuously explored and updated. Our study developed a DHL questionnaire for older adults based on the latest DHL concepts, assessing 6 dimensions: Information, Behavior, Security, Interaction, Content, and Attitude. The questionnaire closely reflects the real-life experiences of older adults and demonstrates strong differentiation, reliability, and practicality. It provides health-related practitioners with a scientifically sound and feasible tool for accurately and comprehensively assessing the current status of DHL in older adults. Future research should conduct an in-depth assessment of the questionnaire's applicability across different linguistic and cultural contexts and evaluate its validity among older adults with varying health statuses, educational backgrounds, and socioeconomic conditions. Additionally, the effectiveness of the questionnaire in specific intervention programs should be assessed to validate its practical guiding role. Furthermore, it is essential to explore ways to integrate assessment results into targeted health promotion strategies and incorporate theories and methods from multiple disciplines, such as psychology, sociology, and education. Through interdisciplinary collaboration, DHL enhancement initiatives for older adults can be more effectively implemented, providing comprehensive support for their health management.

Acknowledgments

This work was supported by the Shanghai Education Science Research Project (grant C2023136), the Ministry of Education Industry-University Co-operation Collaborative Education Project (grant 230805236311535), and the Shanghai Jiao Tong University Liberal Arts Young Talent Cultivation Project (grant 2022QN00X).

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author (WL) on reasonable request.

Disclaimer

No generative AI tools were used in the writing of this manuscript.

XSL•FO
RenderX

Authors' Contributions

WL contributed to the study design, data collection, supervision, and review. XW was involved in the study design, expert consultation, data collection, data analysis, and article writing. CZ and YQ participated in the data collection, data analysis, and article writing. YX, JS, and YL were responsible for data collection. All authors have read and agreed to the publication of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The implementation and analysis processes of the semistructured interviews.
[DOCX File , 16 KB - jmir_v27i1e64193_app1.docx]

Multimedia Appendix 2
Digital Health Literacy Survey for Older Adults.
[DOCX File , 36 KB - jmir_v27i1e64193_app2.docx]

Multimedia Appendix 3
Item analysis.
[DOCX File , 18 KB - jmir_v27i1e64193_app3.docx]

Multimedia Appendix 4
Exploratory factor analysis factor loadings matrix of the digital health literacy (DHL) Questionnaire for older adults.
[DOCX File , 14 KB - jmir_v27i1e64193_app4.docx]

Multimedia Appendix 5
Criterion-related validity of the digital health literacy (DHL) Questionnaire in older adults.
[DOCX File , 13 KB - jmir_v27i1e64193_app5.docx]

References

1. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
2. van DVR, Drossaert CHC, de HM, Taal E, van DLMAFJ. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
3. Wang X, Luan W. Research progress on digital health literacy of older adults: a scoping review. Front Public Health 2022 Aug 5;10:906089 [FREE Full text] [doi: 10.3389/fpubh.2022.906089] [Medline: 35991040]
4. Leung AYM, Parial LL, Tolabing MC, Sim T, Mo P, Okan O, et al. Sense of coherence mediates the relationship between digital health literacy and anxiety about the future in aging population during the COVID-19 pandemic: a path analysis. Aging Ment Health 2022 Mar;26(3):544-553. [doi: 10.1080/13607863.2020.1870206] [Medline: 33438448]
5. Park S, Chung S. E-health literacy and associated factors among Korean adults during the COVID-19 pandemic: age-group differences. Health Promot Int 2023 Oct 01;38(5):daad099. [doi: 10.1093/heapro/daad099] [Medline: 37665719]
6. Wang Y, Song Y, Zhu Y, Ji H, Wang A. Association of eHealth literacy with health promotion behaviors of community-dwelling older people: the chain mediating role of self-efficacy and self-care ability. Int J Environ Res Public Health 2022 May 17;19(10):6092 [FREE Full text] [doi: 10.3390/ijerph19106092] [Medline: 35627627]
7. Park C. Electronic health literacy as a source of self-efficacy among community-dwelling older adults. Clin Gerontol 2024 Jul 01:1-10. [doi: 10.1080/07317115.2024.2373894] [Medline: 38949203]
8. Liu S, Lu Y, Wang D, He X, Ren W, Kong D, et al. Impact of digital health literacy on health-related quality of life in Chinese community-dwelling older adults: the mediating effect of health-promoting lifestyle. Front Public Health 2023 Jun 21;11:1200722 [FREE Full text] [doi: 10.3389/fpubh.2023.1200722] [Medline: 37415711]
9. Kim K, Shin S, Kim S, Lee E. The relation between eHealth literacy and health-related behaviors: systematic review and meta-analysis. J Med Internet Res 2023 Jan 30;25:e40778 [FREE Full text] [doi: 10.2196/40778] [Medline: 36716080]
10. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
11. Dunn P, Conard S. Digital health literacy in cardiovascular research. Int J Cardiol 2018 Oct 15;269:274-275. [doi: 10.1016/j.ijcard.2018.07.011] [Medline: 29996976]

XSL•FO
RenderX

12. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
13. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an informed consumer of health information and assessment of electronic health literacy in a national sample of internet users: validity and reliability of the e-HLS instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
14. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
15. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the ehealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
16. Liu P, Yeh L, Wang J, Lee S. Relationship between levels of digital health literacy based on the Taiwan digital health literacy assessment and accurate assessment of online health information: cross-sectional questionnaire study. J Med Internet Res 2020 Dec 21;22(12):e19767 [FREE Full text] [doi: 10.2196/19767] [Medline: 33106226]
17. Ferrari A. DIGCOMP: a framework for developing and understanding digital competence in Europe. JRC Publications Repository. 2013. URL: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC83167/lb-na-26035-enn.pdf [accessed 2025-02-17]
18. Choi J, Choi S, Song K, Baek J, Kim H, Choi M, et al. Everyday Digital Literacy Questionnaire for older adults: instrument development and validation study. J Med Internet Res 2023 Dec 14;25:e51616 [FREE Full text] [doi: 10.2196/51616] [Medline: 38095999]
19. Kebede AS, Ozolins L, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 07;24(12):e40192 [FREE Full text] [doi: 10.2196/40192] [Medline: 36477006]
20. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014 Apr;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
21. Hsu C, Sandford B. The Delphi technique: making sense of consensus. Practical Assessment, Research, and Evaluation 2007;12(10):1-8 [FREE Full text] [doi: 10.7275/pdz9-th90]
22. MacCallum RC, Widaman KF, Preacher KJ, Hong S. Sample size in factor analysis: the role of model error. Multivariate Behav Res 2001 Oct 01;36(4):611-637. [doi: 10.1207/S15327906MBR3604_06] [Medline: 26822184]
23. Tsai T, Lee SD. Development and validation of a Weight-Specific Health Literacy Instrument (WSHLI). Obes Res Clin Pract 2018;12(2):214-221. [doi: 10.1016/j.orcp.2017.11.003] [Medline: 29269247]
24. Gaskin CJ, Happell B. On exploratory factor analysis: a review of recent evidence, an assessment of current practice, and recommendations for future use. Int J Nurs Stud 2014 Mar;51(3):511-521. [doi: 10.1016/j.ijnurstu.2013.10.005] [Medline: 24183474]
25. Watkins MW. Exploratory factor analysis: a guide to best practice. Journal of Black Psychology 2018 Apr 27;44(3):219-246. [doi: 10.1177/0095798418771807]
26. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Electr J Bus Res Meth 2008;6(1):53-60 [FREE Full text] [doi: 10.21427/D7CF7R]
27. Rönkkö M, Cho E. An updated guideline for assessing discriminant validity. Organizational Research Methods 2020 Nov 23;25(1):6-14. [doi: 10.1177/1094428120968614]
28. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. Res Social Adm Pharm 2019 Feb;15(2):214-221. [doi: 10.1016/j.sapharm.2018.03.066] [Medline: 29606610]
29. Roberts P, Priest H. Reliability and validity in research. Nurs Stand 2006;20(44):41-45. [doi: 10.7748/ns2006.07.20.44.41.c6560] [Medline: 16872117]
30. Cui G, Li S, Yin Y, Chen L, Li J, Liang F, et al. The relationship among social capital, eHealth literacy and health behaviours in Chinese elderly people: a cross-sectional study. BMC Public Health 2021 Jan 06;21(1):45 [FREE Full text] [doi: 10.1186/s12889-020-10037-4] [Medline: 33407275]
31. Lee J, Lee E, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021 Nov 15;23(11):e30644 [FREE Full text] [doi: 10.2196/30644] [Medline: 34779781]
32. Zhang Y, Lee EWJ, Teo W. Health-seeking behavior and its associated technology use: interview study among community-dwelling older adults. JMIR Aging 2023 May 04;6:e43709 [FREE Full text] [doi: 10.2196/43709] [Medline: 36996003]
33. Elavsky S, Knapova L, Janiš K, Cimler R, Kuhnova J, Cernicky T. Multiple perspectives on the adoption of SMART technologies for improving care of older people: mixed methods study. J Med Internet Res 2024 Feb 07;26:e45492 [FREE Full text] [doi: 10.2196/45492] [Medline: 38324345]

Abbreviations

AVE: average variance extracted

XSL•FO
RenderX

DHL: digital health literacy
DIGCOMP: Digital Competence Framework
EFA: exploratory factor analysis
eHEALS: eHealth Literacy Scale
eHL: eHealth literacy
HL: health literacy
I-CVI: item-level content validity index
KMO: Kaiser-Meyer-Olkin
S-CVI/Ave: scale-level content validity index

Edited by T de Azevedo Cardoso; submitted 11.07.24; peer-reviewed by J Busch-Casler, I Mircheva, Y Wang; comments to author 27.09.24; revised version received 18.11.24; accepted 20.02.25; published 19.03.25.

Please cite as:
Wang X, Zhang C, Qi Y, Xing Y, Liu Y, Sun J, Luan W
Digital Health Literacy Questionnaire for Older Adults: Instrument Development and Validation Study
J Med Internet Res 2025;27:e64193
URL: https://www.jmir.org/2025/1/e64193
doi:10.2196/64193
PMID:

©Xinxin Wang, Chengrui Zhang, Yue Qi, Ying Xing, Yawen Liu, Jiayi Sun, Wei Luan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Decomposition and Comparative Analysis of Urban-Rural Disparities in eHealth Literacy Among Chinese University Students: Cross-Sectional Study

Yao Yu[1*], BA; Zhenning Liang[1,2*], MPA; Qingping Zhou[1*], BA; Yusupujiang Tuersun[1], BA; Siyuan Liu[3], MPH; Chenxi Wang[1], BA; Yuying Xie[4], BA; Xinyu Wang[3]; Zhuotong Wu[3]; Yi Qian[1], PhD

[1]School of Health Management, Southern Medical University, Guangzhou, China

[2]The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China

[3]School of Public Health, Southern Medical University, Guangzhou, China

[4]Shenzhen Longhua Maternity and Child Health Hospital, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Yi Qian, PhD
School of Health Management
Southern Medical University
Number 1023, Shatai South Road
Baiyun District
Guangzhou,
China
Phone: 86 13632101533
Email: qianyi_smu@163.com

Abstract

Background: Mobile health care is rapidly expanding in China, making the enhancement of eHealth literacy a crucial strategy for improving public health. However, the persistent urban-rural divide may contribute to disparities in eHealth literacy between urban and rural university students, potentially affecting their health-related behaviors and outcomes.

Objective: This study aims to examine disparities in eHealth literacy between university students in urban and rural China, identifying key influencing factors and their contributions. The findings will help bridge these gaps, promote social equity, enhance overall health and well-being, and inform future advancements in the digital health era.

Methods: The eHealth Literacy Scale (eHEALS) was used to assess eHealth literacy levels among 7230 university students from diverse schools and majors across 10 regions, including Guangdong Province, Shanghai Municipality, and Jiangsu Province. Descriptive statistics summarized demographic, sociological, and lifestyle characteristics. Chi-square tests examined the distribution of eHealth literacy between urban and rural students. A binary logistic regression model identified key influencing factors, while a Fairlie decomposition model quantified their contributions to the observed disparities.

Results: The average eHealth literacy score among Chinese university students was 29.22 (SD 6.68), with 4135 out of 7230 (57.19%) scoring below the passing mark. Rural students had a significantly higher proportion of inadequate eHealth literacy (2837/4510, 62.90%) compared with urban students (1298/2720, 47.72%; P<.001). The Fairlie decomposition analysis showed that 71.4% of the disparity in eHealth literacy was attributable to urban-rural factors and unobserved variables, while 28.6% resulted from observed factors. The primary contributors were monthly per capita household income (13.4%), exercise habits (11.7%), and 9-item Patient Health Questionnaire (PHQ-9) scores (2.1%).

Conclusions: Rural university students exhibit lower eHealth literacy levels than their urban counterparts, a disparity influenced by differences in socioeconomic status, individual lifestyles, and personal health status. These findings highlight the need for targeted intervention strategies, including (1) improving access to eHealth resources in rural and underserved areas; (2) fostering an environment that encourages physical activity to promote healthy behaviors; (3) expanding school-based mental health services to enhance health information processing capacity; and (4) implementing systematic eHealth literacy training with ongoing evaluation. These strategies will support equitable access to and utilization of eHealth resources for all students, regardless of their geographic location.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63671) doi:10.2196/63671

KEYWORDS

university students; eHealth literacy; urban-rural disparities; Fairlie decomposition model; health equity

Introduction

Health literacy is a midstream determinant of health that can improve outcomes, reduce inequalities, and promote health-related behaviors [1]. Today, the internet is widely used due to its accessibility, broad information coverage, ease of use, affordability, and anonymity [2]. It has become a key platform for disseminating health information, particularly in middle- and high-income countries such as China, where health care resources and face-to-face medical appointments are limited [3]. However, the widespread availability of the internet has also exposed shortcomings in online health information, especially concerning its quality [4]. Individuals frequently encounter false or dubious health information online, a challenge that is not new [5]. eHealth literacy extends the concept of health literacy into the digital domain. It was first introduced by Norman and Skinner [6]. Gilstad [7] later expanded its definition to encompass the ability to address health problems by communicating, locating, understanding, evaluating, and applying health information and digital technologies within specific cultural and social contexts [8]. It should be emphasized that eHealth literacy requires not only the ability to access necessary information but also the ability to assess the quality of eHealth information and distinguish between reliable and unreliable sources [6]. eHealth literacy is essential for maintaining good health and preventing disease. Improving eHealth literacy enables individuals to access health information for disease prevention, treatment, and health promotion [9]. People with strong eHealth literacy can make informed decisions, select appropriate health services, and manage their health effectively. This includes monitoring their health status, following medical advice, and taking preventive measures against disease. Furthermore, eHealth literacy can enhance communication between patients and physicians, improving the effectiveness and overall satisfaction of medical services [10].

As the future builders and pillars of development in their home countries, the health status of young people and their ability to maintain it deserve attention. Their knowledge, attitudes, and behaviors not only impact their own lives but also influence society as a whole [11,12]. For university students transitioning to adulthood, the development of health behaviors during this stage is crucial for maintaining a healthy lifestyle throughout adulthood [13,14]. However, they are exposed to various health risks [15-17]. At the same time, even students in good health often engage in risky health behaviors [18]. This highlights an increased need for reliable health information. Conducting an internet search may align with young people's desire for autonomy and self-sufficiency [19]. Currently, it is estimated that most university students (99%) rely on the internet as their primary source of health-related information [20]. The eHealth Literacy Scale (eHEALS) has emerged as a valid measurement and intervention tool for promoting healthy behaviors among adolescents [21]. Improving eHealth literacy can encourage students to adopt healthier behaviors [22], better manage their health, and maintain both their physical and mental well-being [23,24]. However, despite frequently using digital media to access health information, a significant proportion of university students lack adequate eHealth literacy [25,26], which severely hinders disease prevention and the effectiveness of health services [27]. Stellefson et al [9] demonstrated that while students are confident in their ability to use the internet, their practical eHealth literacy is often insufficient. Moreover, another study found that a considerable proportion of students struggle to critically evaluate the health information they encounter online [26].

According to the latest data from the Chinese Ministry of Education, the total number of students enrolled in various forms of higher education in China in 2023 was 47.63 million [28]. Research on university students' eHealth literacy is both important and urgent, particularly given the lack of a strict verification mechanism for health information on most websites in China [29]. While research on health literacy in China began relatively late, the widespread use of internet technology in health care has gradually brought eHealth literacy into public focus. As university students are a key demographic of internet users, they have become a focal point of eHealth literacy research due to their frequent use of digital tools and ability to access information. One major area of research is the localized adaptation of eHealth literacy measurement tools. In 2013, Guo et al [30] pioneered the translation and adaptation of eHEALS in China, focusing on adolescents, with a reported Cronbach alpha coefficient of 0.913. Subsequent applications of this version among university students have consistently demonstrated high reliability and validity [31,32]. Another important research direction is examining the current state of students' eHealth literacy and analyzing the factors that influence it. Overall, the eHealth literacy of Chinese university students appears to be at a relatively optimistic level. In recent years, studies have shown that the average eHealth literacy scores of university students typically range between 27 and 29 points [33-35], with variations influenced by the time and region of the research. The key factors affecting university students' eHealth literacy primarily include individual characteristics such as gender, major, academic performance, and place of origin (urban or rural). Family-related factors, such as income and number of siblings, as well as the students' own health status, also play a significant role [36]. With the promotion of initiatives such as the "Healthy China Strategy" and the "14th Five-Year Plan for National Health Informatization," the eHealth literacy of university students has become a key issue, highlighting the urgent need for intervention research. However, existing studies are largely limited to single-region samples, lacking large-scale cross-sectional surveys that cover multiple regions. Additionally, research on the determinants of eHealth literacy remains insufficient, with limited detailed data analysis and stratified comparisons among students with different characteristics. Furthermore, there is a lack of systematic

XSL•FO
RenderX

comparisons on how variations in eHealth literacy impact health outcomes, making it difficult to establish a concrete and reliable empirical foundation for intervention research and policy development.

Existing academic research on the role of the internet in health communication and promotion often emphasizes macro-level social inequalities, commonly referred to as the digital divide [37,38]. eHealth literacy plays a crucial role in promoting health equity, improving health outcomes, and reducing health care costs [5], offering significant potential to bridge this divide. This study highlights the widening digital gap between urban and rural students in China, with a particular focus on the challenges faced by rural students as a digitally disadvantaged group. Specifically, it explores the following: (1) Are there differences in eHealth literacy between urban and rural students? (2) What similarities or differences exist in the factors influencing their eHealth literacy? (3) How do these factors contribute to the observed differences? The study's results will serve as a reference for developing intervention methods and improvement strategies for eHealth literacy among students. Additionally, it seeks to recognize and address these social inequalities, ensuring that eHealth can truly benefit people worldwide, including those in marginalized and underserved areas.

Methods

Data Source

This study is a cross-sectional survey. From January to February 2023, the researcher distributed an anonymous electronic questionnaire via the Questionnaire Star platform [39] using a convenience sampling method. The questionnaire was distributed to university students across 10 regions, including Guangdong Province, Shanghai Municipality, and Jiangsu Province, covering majors such as economics, medicine, management, and literature.

To ensure data accuracy and reliability, the questionnaire included 2 verification questions: a repetitive question and a general knowledge question ("What is the capital of China?"). If a respondent answered either question incorrectly, the questionnaire was deemed invalid and excluded from further analysis. A total of 7503 questionnaires were collected. After screening, 253 questionnaires with anomalous response times and 20 with irregular completion patterns were excluded, resulting in 7230 valid questionnaires and an effective response rate of 96.36%. Participants in this study were required to meet the following inclusion criteria: (1) undergraduate students and (2) normal cognitive function with the ability to complete the questionnaire independently. The exclusion criterion was an incorrect response to the verification questions. The exclusion process is illustrated in Figure 1.

Figure 1. Participant screening flowchart.

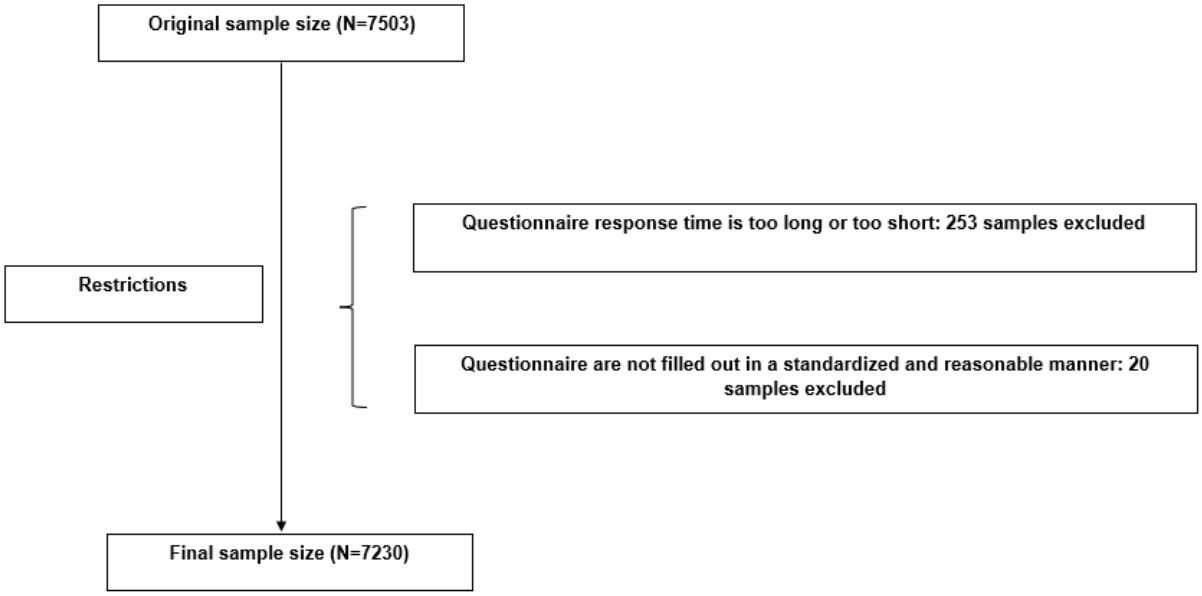

eHealth Literacy

This study used the Chinese version of the eHEALS to assess university students' eHealth literacy levels. The eHEALS was developed by Norman et al [6] in 2006, based on the concept of eHealth literacy and Lily's model for evaluating an individual's ability to search for, screen, understand, assess, and utilize health information [6]. Since its inception, eHEALS has become the most widely used scale for measuring eHealth literacy across different populations in contemporary studies [40]. The Cronbach alpha coefficient for eHEALS is high across numerous studies involving diverse populations, typically ranging from 0.84 to 0.94. This indicates that the scale demonstrates strong internal consistency and reliability. The scale consists of 8 items, categorized into 3 competency areas related to the use of online health information and services: application competencies (items 1-5), judgmental competencies (items 6 and 7), and decision-making competencies (item 8). The scale uses a 5-point Likert scale, with scores ranging from 1 (very poorly met) to 5 (very well met). The total score for the

XSL•FO
RenderX

8 items ranges from 8 to 40, with higher scores indicating a higher level of eHealth literacy. In this study, the widely recognized cut-off score of 32 was used to classify the eHealth literacy of urban and rural university students into 2 categories: "qualified" (≥32 points) and "unqualified" (<32 points) [33,41]. The overall Cronbach α coefficient for this study was 0.974, with 0.973 for the rural university student group and 0.974 for the urban university student group.

Ethics Approval

The study was approved by the Biomedical Ethics Committee of Southern Medical University (Ethics Review Board [2023] No. 46). All participants signed an informed consent form before the survey began. The distribution, collection, and storage of questionnaires were managed via the Questionnaire Star platform. Data access was restricted to the researcher, and the information was tamper-proof to ensure security and integrity. This study strictly adhered to relevant laws and regulations. To protect participant privacy, all public reports will exclude personal identifiers. The final data set was exported in EpiData format for subsequent statistical analysis.

Grouping Variables

Respondents were classified as rural or urban based on the household type they reported in the survey.

Covariates

To obtain more reliable results, a range of potential confounding variables were controlled for, including demographic and sociological characteristics, personal lifestyles, and mental health status—factors previously considered in studies on eHealth literacy levels.

Demographic Characteristics

Participants were classified based on 2 variables: gender (male or female) and personal religious belief (yes or no).

Sociological Characteristics

Respondents were classified into 4 groups based on their academic performance, determined by class ranking: below 25%, 25%-50%, 50%-75%, and above 75%. Additionally, monthly per capita household income was categorized into 4 groups: less than 2500 RMB (1 RMB=US $0.14), 2500-<5000 RMB, 5000-10,000 RMB, and above 10,000 RMB.

Personal Lifestyle

The variables of exercise, smoking, and alcohol use were dichotomized into a binary format, with responses categorized as "yes" or "no." "Exercise" was defined as engaging in physical activity at least three times in the past month, with each session lasting a minimum of 30 minutes. Additionally, the use of tobacco products and alcoholic beverages (including distilled spirits, beer, wine, and rice wine) in the past month was recorded as a dichotomous variable (yes or no).

Mental Health Status

The 7-item Generalized Anxiety Disorder (GAD-7) tool was used to assess the prevalence and severity of anxiety symptoms among study participants. This tool consists of 7 items, each rated on a 4-point scale ranging from 0 to 3: 0="not at all," 1="a few days," 2="more than half the time," and 3="almost every day." A score below 5 indicates the absence of anxiety symptoms, while a score of 5 or higher suggests the presence of anxiety symptoms [42]. In this study, the GAD-7 demonstrated a Cronbach α coefficient of 0.944.

The 9-item Patient Health Questionnaire (PHQ-9) was used to assess depressive symptoms among participants. This tool consists of 9 items and follows the same scoring method as the GAD-7. A score below 5 indicates the absence of depressive symptoms, whereas a score of 5 or higher suggests their presence [43]. In this study, the PHQ-9 demonstrated a Cronbach α coefficient of 0.931.

Statistical Analysis

The demographic characteristics and lifestyles of the study population were analyzed descriptively. Continuous variables were presented as mean (SD), while categorical variables were expressed as percentages. A chi-square test was used to examine the distribution of eHealth literacy levels among urban and rural university students. Subsequently, a binary logistic regression model was constructed to identify key determinants of eHealth literacy levels in these groups. The outcome variable was defined as a binary classification of eHealth literacy (qualified/unqualified), while independent variables were selected from 4 domains: demographic characteristics, sociological characteristics, personal lifestyle, and mental health status. Initially, variables were screened using univariate analysis (retention criterion: $P<.05$) and subsequently included in the multivariate model. Statistically significant predictors ($P<.05$) were retained in the final model. All statistical analyses were performed using SPSS Statistics 23 (IBM Corp.). To examine the factors contributing to discrepancies in eHealth literacy levels between urban and rural university students, the Fairley model was applied using Stata MP 18.0 (StataCorp). The significance level was set at .05.

Fairlie Decomposition Model

In this study, we applied a multivariate Fairlie decomposition analysis (FDA) based on a binary regression model. FDA is a decomposition technique used in multivariate models to quantify the contribution of predicted differences between 2 groups to an outcome variable. It extends the Blinder-Oaxaca decomposition, which has been widely criticized for its inefficiency in handling logit and probit models. FDA was specifically developed for nonlinear regression models, including logit and probit models [44]. The FDA quantifies the contribution of independent variables to differences between groups by calculating the change in the mean predicted probability when substituting 1 independent variable at a time in 1 group (eg, rural university students) while holding other variables constant in the other group (eg, urban university students). The Fairlie decomposition technique ensures that the predicted probability remains within the range of 0-1. Studies have shown that the FDA provides a more precise quantification of variable contributions and significance levels in nonlinear regression models [45-47]. Given that the dependent variable was dichotomous, we applied Fairlie nonlinear decomposition to attribute differences in university students' eHealth literacy to various contributing factors. As outlined by Fairlie [48], the

XSL•FO
RenderX

decomposition of the nonlinear equation can be expressed as follows:



The symbols  represent the mean probabilities of the 2 binary outcomes. The F is used to indicate the cumulative distribution function of the logistic distribution. Y_e-Y_w represents the total difference in the differences between the 2 groups; N^e and N^w represent the sample sizes of the 2 samples. In equation 1, the first term in parentheses represents the portion of the gap attributable to differences in observed characteristics, while the second term denotes the portion of the gap explained by differences in estimated coefficients.

Results

Basic Characteristics of the Study Participants

The study included a total of 7230 participants. The mean eHealth literacy score among university students was 29.22 (SD 6.68; Table 1). Prior research on eHealth literacy among Chinese university students has established a passing threshold of 32 points [33], with scores below this cutoff indicating insufficient eHealth literacy. As shown in Table 2, 4135 out of 7230 (57.19%) students did not meet the eHealth literacy standard, while 3095 out of 7230 (42.81%) students qualified. Notably, a higher proportion of rural university students (2837/4510, 62.90%) failed to meet the standard compared with their urban counterparts (1298/2720, 47.72%; P<.001).

The chi-square test results in Table 2 indicate significant differences in the distribution of 8 covariates between urban and rural university students. These covariates include gender (P<.001), religion (P<.001), class ranking (P=.02), monthly per capita household income (P<.001), exercise (P<.001), drinking (P<.001), GAD-7 scores (P=.002), and PHQ-9 scores (P=.04). However, no significant difference was observed in the distribution of smoking (P=.93).

Table 1. eHealth literacy score for university students.

Item	Overall, mean (SD)	Rural, mean (SD)	Urban, mean (SD)
Q1: I know how to find helpful health resources on the internet	3.69 (0.93)	3.59 (0.91)	3.86 (0.94)
Q2: I know how to use the internet to answer my health questions	3.67 (0.92)	3.57 (0.90)	3.85 (0.93)
Q3: I know what health resources are available on the internet	3.64 (0.91)	3.53 (0.89)	3.82 (0.91)
Q4: I know where to find helpful health resources on the internet	3.65 (0.90)	3.54 (0.88)	3.83 (0.91)
Q5: I know how to use the health information I find on the internet to help me	3.69 (0.89)	3.58 (0.87)	3.86 (0.91)
Q6: I have the skills I need to evaluate the health resources I find on the internet	3.63 (0.90)	3.52 (0.87)	3.81 (0.92)
Q7: I can tell high-quality from low-quality health resources on the internet	3.67 (0.90)	3.55 (0.88)	3.86 (0.90)
Q8: I feel confident in using information from the internet to make health decisions	3.59 (0.90)	3.48 (0.88)	3.76 (0.92)
Total score	29.22 (6.68)	28.36 (6.48)	30.65 (6.75)

XSL•FO
RenderX

Table 2. Distribution of the variables in rural and urban respondents.

Variable	Rural (n=4510), n (%)	Urban (n=2720), n (%)	Chi-square (*df*)	*P* value
eHealth literacy			159.781 (1)	<.001
≥32	1673 (37.10)	1422 (52.28)		
<32	2837 (62.90)	1298 (47.72)		
Gender			58.356 (1)	<.001
Male	1355 (30.04)	1055 (38.79)		
Female	3155 (69.96)	1665 (61.21)		
Religious belief			18.783 (1)	<.001
Yes	291 (6.45)	110 (4.04)		
No	4219 (93.55)	2610 (95.96)		
Class ranking			9.832 (3)	.02
<25.0%	1217 (26.98)	781 (28.71)		
25.0%-<50.0%	1623 (35.99)	1026 (37.72)		
50.0%-<75.0%	1233 (27.34)	658 (24.19)		
≥75.0%	437 (9.69)	255 (9.38)		
Monthly per capita household income (RMB[a])			10002.180 (3)	<.001
<2500	1158 (25.68)	219 (8.05)		
2500-<5000	2163 (47.96)	852 (31.32)		
5000-<10,000	952 (21.11)	1030 (37.87)		
≥10,000	237 (5.25)	619 (22.76)		
Exercise			28.080 (1)	<.001
Yes	3078 (68.25)	2016 (74.12)		
No	1432 (31.75)	704 (25.88)		
Smoking			0.008 (1)	.93
Yes	141 (3.13)	84 (3.09)		
No	4369 (96.87)	2636 (96.91)		
Drinking			49.534 (1)	<.001
Yes	626 (13.88)	549 (20.18)		
No	3884 (86.12)	2171 (79.82)		
7-item Generalized Anxiety Disorder			8.135 (1)	.002
≥5	1812 (40.18)	1001 (36.80)		
<5	2698 (59.82)	1719 (63.20)		
9-item Patient Health Questionnaire			4.420 (1)	.04
≥5	2141 (47.47)	1222 (44.93)		
<5	2369 (52.53)	1498 (55.07)		

[a]1 RMB=US $0.14.

Comparison of Variables' Distribution

Table 3 presents the relationship between covariate distribution and eHealth literacy qualification status among urban and rural university students. The data indicate that certain covariate distributions share similar characteristics among students who passed or failed the eHealth literacy assessment. Gender, religious belief, monthly per capita household income, and drinking were identified as factors associated with eHealth literacy among university students.

XSL•FO
RenderX

Table 3. Distribution of the variables in unqualified eHealth literacy and qualified eHealth literacy respondents.

Variable	Unqualified eHealth literacy				Qualified eHealth literacy			
	Rural (n=2837), n (%)	Urban (n=1298), n (%)	Chi-square (*df*)	*P* value	Rural (n=1673), n (%)	Urban (n=1422), n (%)	Chi-square (*df*)	*P* value
Gender			30.794 (1)	<.001			19.263 (1)	<.001
Male	807 (28.45)	481 (37.06)			548 (32.76)	574 (40.37)		
Female	2030 (71.55)	817 (62.94)			1125 (67.24)	848 (59.63)		
Religious belief			15.514 (1)	<.001			4.876 (1)	.03
Yes	184 (6.49)	45 (3.47)			107 (6.40)	65 (4.57)		
No	2653 (93.51)	1253 (96.53)			1566 (93.60)	1357 (95.43)		
Class ranking			3.912 (3)	.27			5.043 (3)	.17
<25.0%	713 (25.13)	336 (25.89)			504 (30.13)	445 (31.29)		
25.0%-<50.0%	1032 (36.38)	487 (37.52)			591 (35.33)	539 (37.90)		
50.0%-<75.0%	802 (28.27)	330 (25.42)			431 (25.76)	328 (23.07)		
≥75.0%	290 (10.22)	145 (11.17)			147 (8.79)	110 (7.74)		
Monthly per capita household income (RMB[a])			536.789 (3)	<.001			389.931 (3)	<.001
<2500	789 (27.81)	117 (9.01)			369 (22.06)	102 (7.17)		
2500-5000	1406 (49.56)	458 (35.29)			757 (45.25)	394 (27.71)		
5000-10,000	523 (18.43)	465 (35.82)			429 (25.64)	565 (39.73)		
≥10,000	119 (4.19)	258 (19.88)			118 (7.05)	361 (25.39)		
Exercise			2.261 (1)	.13			0.791 (1)	.37
Yes	1652 (58.23)	788 (60.71)			1426 (85.24)	1228 (86.36)		
No	1185 (41.77)	510 (39.29)			247 (14.76)	194 (13.64)		
Smoking			0.092 (1)	.76			0.298 (1)	.59
Yes	97 (3.42)	42 (3.24)			44 (2.63)	42 (2.95)		
No	2740 (96.58)	1256 (96.76)			1629 (97.37)	1380 (97.05)		
Drinking			31.217 (1)	<.001			18.871 (1)	<.001
Yes	395 (13.92)	270 (20.80)			231 (13.81)	279 (19.62)		
No	2442 (86.08)	1028 (79.20)			1442 (86.19)	1143 (80.38)		
7-item Generalized Anxiety Disorder			0.612 (1)	.43			0.572 (1)	.45
≥5	1285 (45.29)	571 (43.99)			527 (31.50)	430 (30.24)		
<5	1552 (54.71)	727 (56.01)			1146 (68.50)	992 (69.76)		
9-item Patient Health Questionnaire			0.002 (1)	.97			0.002 (1)	.97
≥5	1528 (53.86)	700 (53.93)			613 (36.64)	522 (36.71)		
<5	1309 (46.14)	598 (46.07)			1060 (63.36)	900 (63.29)		

[a]1 RMB=US $0.14.

Logistic Model Results

Table 4 presents the results of logistic modeling for eHealth literacy qualification among urban and rural university students. Among rural university students, the following were identified as risk factors for failing to meet the eHealth literacy standard: female gender (odds ratio [OR] 0.847), class ranking in the 25.0%-49.9% and 50.0%-74.9% ranges (both OR 0.788), a PHQ-9 score of ≤5 (OR 0.648), and lack of exercise (OR 0.260). Conversely, higher monthly per capita household income was identified as a protective factor (5000-10,000 RMB: OR 1.607; ≥10,000 RMB: OR 1.944). Among urban university students, risk factors included class ranking in the top 25% (OR 0.708), a PHQ-9 score of <5 (OR 0.603), and lack of exercise (OR 0.267).

XSL•FO
RenderX

In conclusion, the discrepancies in eHealth literacy between urban and rural university students were primarily reflected in 2 key areas. First, gender was a significant factor only in rural settings, where being female was associated with a lower likelihood of meeting the eHealth literacy standard (OR 0.847). Second, higher monthly per capita household income (5000-10,000 RMB: OR 1.607; ≥10,000 RMB: OR 1.944) served as a protective factor in rural contexts but was not significant in urban ones.

Table 4. Logistic regression results for sociodemographic characteristics associated with eHealth literacy levels.

Variable	Overall, odds ratio (95% CI)	Rural, odds ratio (95% CI)	Urban, odds ratio (95% CI)
Gender			
Male	Reference	Reference	Reference
Female	0.843[a] (0.755-0.940)	0.847[a] (0.733-0.980)	0.893 (0.754-1.058)
Religious belief			
Yes	Reference	Reference	Reference
No	0.935 (0.752-1.163)	1.003 (0.772-1.303)	0.688 (0.454-1.042)
Class ranking			
<25.0%	Reference	Reference	Reference
25.0%-<50.0%	0.816[a] (0.720-0.924)	0.788[a] (0.670-0.926)	0.864 (0.708-1.054)
50.0%-<75.0%	0.798[a] (0.696-0.915)	0.788[a] (0.662-0.938)	0.830 (0.664-1.037)
≥75.0%	0.791[a] (0.653-0.958)	0.839 (0.655-1.074)	0.708[a] (0.521-0.962)
Monthly per capita household income (RMB[b])			
<2500	Reference	Reference	Reference
2500-5000	1.081 (0.939-1.245)	1.061 (0.904-1.246)	0.861 (0.626-1.184)
5000-10,000	1.712[c] (1.472-1.992)	1.607[c] (1.331-1.940)	1.161 (0.848-1.589)
≥10,000	2.079[c] (1.727-2.503)	1.944[c] (1.439-2.626)	1.273 (0.914-1.774)
Exercise			
Yes	Reference	Reference	Reference
No	0.262[c] (0.232-0.296)	0.260[c] (0.222-0.304)	0.267[c] (0.220-0.324)
Smoking			
Yes	Reference	Reference	Reference
No	0.828 (0.608-1.128)	0.802 (0.532-1.209)	0.897 (0.554-1.452)
Drinking			
Yes	Reference	Reference	Reference
No	1.041 (0.903-1.200)	1.007 (0.826-1.228)	1.015 (0.826-1.249)
7-item Generalized Anxiety Disorder			
≥5	Reference	Reference	Reference
<5	0.867 (0.748-1.003)	0.839 (0.694-1.013)	0.920 (0.727-1.165)
9-item Patient Health Questionnaire			
≥5	Reference	Reference	Reference
<5	0.638[c] (0.553-0.736)	0.648[c] (0.539-0.780)	0.603[c] (0.480-0.758)

[a]Significant at $P<.05$.

[b]1 RMB=US $0.14.

[c]Significant at $P<.01$.

XSL•FO
RenderX

Decomposition Analysis Results

To ensure the stability of the results, the decomposition model was repeated 100 times using the Stata MP 18.0 (StataCorp). Multimedia Appendix 1 presents the decomposition model results for differences in eHealth literacy levels between urban and rural university students. The findings indicate that observed factors accounted for 28.6% of the disparity, while 71.4% was attributed to urban-rural differences and unobserved factors. Significant contributors to the eHealth literacy gap (P<.001) included monthly per capita household income (13.4%; P=.007), exercise habits (11.7%; P<.001), and PHQ-9 scores (2.1%; P<.001).

Discussion

Principal Findings

This study explores the relationship between eHealth literacy among urban and rural university students in China and factors such as sociodemographic characteristics, personal lifestyle, and psychological health. It quantifies the extent to which these factors contribute to disparities in eHealth literacy between the 2 groups. Our findings confirm significant differences in eHealth literacy levels between urban and rural students, influenced by multiple factors.

The study found that 57.2% of university students had eHealth literacy below the passing threshold, with an average score of 29.22 (6.68). These scores were comparable to those reported for Chinese university students (mean 30.16, SD 6.31) [34], higher than those of students from other Asian countries [49,50], and similar to those of their European and American peers [51,52]. However, all average scores remained below the qualifying standard, highlighting a concerning issue: overall eHealth literacy levels among university students are insufficient. Previous research by Bailey et al [53], Chesser et al [54], Gustafson et al [55], Mengestie et al [56], and others have documented disparities in eHealth literacy between urban and rural populations. Our study found that a significantly higher proportion of rural university students (2837/4510, 62.90%) did not meet eHealth literacy standards compared with their urban counterparts (1298/2720, 47.72%; P<.001). These findings highlight substantial urban-rural disparities in eHealth literacy among Chinese university students, aligning with previous research by Chinese scholars [57]. The study identified a 15.2% point gap between urban and rural students failing to meet eHealth literacy standards, emphasizing the critical role of the upbringing environment in shaping eHealth literacy. Enhancing eHealth literacy among rural students should be a priority to address and potentially mitigate these disparities. Notably, the study by Giger et al [58] found that eHEALS scores were not influenced by rural status or gender. This outcome may stem from uncertainties in the theoretical framework of eHealth literacy, potentially leading to misinterpretations of data measured by the eHEALS scale. Additionally, the lack of variation in scores could be due to a somewhat biased study population.

The logistic regression analysis revealed both similarities and differences in the covariates influencing eHealth literacy among Chinese urban and rural university students. Class ranking, exercise, and PHQ-9 scores were significant factors in both groups. Notably, better academic performance emerged as a protective factor for higher eHealth literacy. To some extent, higher grades indicate better information access and comprehension skills, aligning with findings from previous studies [12,23]. Exercise is a common protective factor, consistent with findings from Tsukahara et al [49], Mitsutake et al [21], and others. Exercise habits reflect an individual's greater health awareness and willingness to seek information. Additionally, many physical activities facilitate social interaction and information sharing. Individuals with high eHealth literacy are also more likely to engage in regular exercise [59], which collectively contributes to the accumulation and enhancement of health literacy. Conversely, PHQ-9 scores were identified as a risk factor for eHealth literacy among both urban and rural students, a finding supported by previous studies [59-61]. Extensive research has confirmed that depressive symptoms negatively impact cognitive function [62], information processing, and health decision-making. Furthermore, depression directly undermines self-efficacy [63], which in turn influences health literacy both directly and indirectly [64], ultimately hindering its development and maintenance.

The covariates influencing eHealth literacy levels differ between urban and rural students. This study found that gender significantly affects rural students, with rural female students exhibiting lower eHealth literacy levels. The existing literature presents mixed findings on this issue. In countries with prevalent gender inequalities, women tend to have lower health literacy, whereas in some Western nations, women demonstrate higher health literacy levels [65,66]. In the context of this study, these results may partially reflect the persistent gender gap in education in rural China, as well as traditional societal views on gender roles. This disparity objectively limits rural female students' access to and use of health information, while also indirectly influencing their health awareness, needs, and confidence in using eHealth resources. Additionally, potential gender differences in health needs, attitudes, and communication patterns underscore the importance of tailored interventions and support. By contrast, urban students do not exhibit significant gender differences, likely due to relatively greater resource availability and a more equitable social environment, where information dissemination is more diverse and access to electronic tools and health information is more convenient. In addition, monthly per capita household income serves as a protective factor for eHealth literacy among rural students but not among urban students. This may be because high-income rural families can better support students in accessing and utilizing eHealth information resources, whereas in urban settings, the impact of income differences is mitigated by greater environmental support and resource availability. The study [67] highlights the importance of tailoring eHealth literacy interventions to specific populations.

The results of the Fairlie decomposition model indicate that 28.6% of the difference in eHealth literacy between urban and rural students can be attributed to observed factors, primarily monthly per capita family income (13.4%), exercise habits (11.7%), and PHQ-9 scores (2.1%). These factors not only demonstrate statistical significance in the empirical data but

XSL•FO
RenderX

also have clear, real-life implications, which is common in logistic regression analyses. Several studies have shown that family income is a key determinant of health literacy, both in the general population and among students [68], as it directly influences their ability to access health information. Students from higher-income families are more likely to afford smart devices and internet services, facilitating easier access to online medical resources [69]. Exercise, by contrast, reflects students' healthy habits, which are not only positively correlated with health literacy but also subconsciously motivate them to seek more health knowledge [70]. Additionally, the PHQ-9 score, as a self-report measure of depressive states, suggests that students with higher scores may be less likely to engage with and comprehend health information due to low mood, a relationship that has been confirmed in studies on mental health and health literacy [71].

In this study, 71.4% of the difference in eHealth literacy between urban and rural students was attributed to unobserved factors, which may be linked to structural disparities between urban and rural areas. Health inequalities between urban and rural populations remain a global issue, with developing countries experiencing more pronounced disparities than developed countries [72]. Strengthening health services in rural communities is widely recognized as a priority. Multiple factors contribute to the observed disparities in health outcomes between rural and urban areas [73]. The pervasive socioeconomic disadvantages endemic to rural areas often result in reduced access to health care services, perpetuating a cycle of disadvantage. Geographic isolation and a shortage of health care providers further exacerbate this issue. Additionally, rural populations face higher risks of injury, encounter significant barriers related to transportation and communication, and are affected by diseconomies of scale due to their sparse distribution. Collectively, these factors negatively impact the health outcomes of individuals living in nonmetropolitan areas. Nevertheless, evidence suggests that even after geographic barriers are removed, health service utilization remains low in some rural communities [74]. Additionally, studies indicate that barriers to self-management—such as limited formal education and poverty—may be further exacerbated in areas with fewer health care professionals and weaker health infrastructure [75]. Patterson et al [76] found that adjusting for socioeconomic status did not fully account for the observed differences in health risk factors between rural and urban populations. Moreover, other factors that differ between these regions may further contribute to disparities in health outcomes. Given the critical role of eHealth literacy in health outcomes, it is plausible that it represents one of these contributing factors.

Existing research has confirmed the crucial role of mobile technology in enhancing eHealth literacy and improving health outcomes, particularly in rural areas and among disadvantaged groups [77]. As health services continue to transition to digital platforms, the need to promote eHealth literacy becomes increasingly urgent. Our findings offer valuable insights for policy makers. First, efforts should be made to improve access to eHealth resources in rural and underserved areas. Additionally, the government should implement measures to enhance residents' incomes and living conditions, thereby addressing structural barriers to eHealth engagement. Given the urban-rural divide, collaboration between the government and educational institutions is essential to developing a multichannel eHealth education platform. This platform should provide tailored eHealth knowledge and telemedicine services, such as live-streamed health lectures and regular dissemination of health information, specifically targeting rural and resource-limited areas to bridge the urban-rural information access gap. Importantly, interventions should be designed with consideration for less educated and lower-income populations, who have the greatest need for such support [78]. Second, fostering a positive environment for physical activity can help encourage the adoption of healthy behaviors, as health literacy and healthy behaviors are mutually reinforcing [79]. To encourage students to develop healthy exercise habits, universities should collaborate with local communities to establish well-equipped sports facilities and organize regular sports events and health-related activities. Additionally, implementing a reward system—where students earn points for participating in sports and health lectures, redeemable for incentives—could further motivate engagement. Third, expanding on-campus mental health services is crucial for enhancing students' ability to process health information. Given the significant impact of mental health on eHealth literacy observed in this study, universities should establish a comprehensive mental health support system. This could include online psychological counseling, crisis intervention services, and routine psychological assessments to ensure early detection and support. Universities should provide high-quality online mental health resources through their health and counseling center websites [80] and establish an integrated framework linking mental health support with eHealth literacy training. This approach would enhance students' understanding and application of health knowledge while effectively reducing barriers to accessing health information. Finally, eHealth literacy training should be systematic, with a structured approach to evaluating its effectiveness. The influence of education on health is well established [81], and numerous studies have shown that elective courses can significantly impact students' health literacy levels [82,83]. Universities can implement targeted eHealth literacy training programs that comprehensively cover all aspects of eHealth literacy. Regular monitoring through surveys and evaluations can help identify educational gaps and optimize the curriculum accordingly. Additionally, fostering supportive environments at home and school can enhance students' self-efficacy and confidence, ultimately contributing to improved health outcomes.

Strengths and Limitations

This study is the first to investigate disparities in eHealth literacy between urban and rural university students in China. Using Fairlie's model for quantitative decomposition, we identified key influencing factors and their specific contributions. However, certain limitations must be acknowledged. First, as this study is based on a cross-sectional questionnaire survey, it does not capture dynamic changes in participants' eHealth literacy over time, limiting our ability to establish causal relationships between eHealth literacy and urban-rural factors. Second, although the eHEALS has been widely validated and

XSL•FO
RenderX

is considered reliable across various cultural contexts and countries, it remains a self-assessment tool and is therefore susceptible to individual subjective biases. Notably, large-scale epidemiological studies worldwide, such as the UK Biobank and China Health and Retirement Longitudinal Study (CHARLS) databases, often rely on self-reported data. These subjective assessments can be supplemented with objective indicators to provide a more comprehensive evaluation of actual eHealth skill levels. Third, our sample size is relatively limited compared with the overall population of university students. Therefore, caution should be exercised when generalizing the findings to a broader or global scale. To enhance the universality of these results, future studies could refine the sampling method and expand the survey scope. Comparative studies in other countries could also provide insights into the similarities and differences between China and other contexts. Finally, while this study considered a broad range of demographic, sociological, and personal lifestyle factors, eHealth literacy is influenced by numerous variables. Future research should explore additional influencing factors. Moreover, if disparities in eHealth literacy between urban and rural students continue to be observed, it will be crucial to develop scientifically sound, context-specific interventions. Researchers should also focus on validating, refining, and addressing potential shortcomings in the implementation of these interventions.

Conclusions

The regression and decomposition analyses in this study revealed that urban university students had higher eHealth literacy levels than their rural counterparts. Key factors contributing to this disparity included monthly per capita household income, exercise habits, and PHQ-9 scores. The Fairlie decomposition model indicated that 28.6% of the difference in eHealth literacy between urban and rural students could be attributed to observed factors, with monthly per capita household income accounting for 13.4%, exercise for 11.7%, and PHQ-9 scores for 2.1%. These findings provide novel insights into urban-rural disparities in eHealth literacy among Chinese university students, offering valuable guidance for developing or refining policies to enhance eHealth literacy.

The policy strategies specifically include the following: (1) improving access to eHealth resources in rural and underserved areas; (2) fostering an environment that encourages physical activity to promote healthy behaviors; (3) expanding school-based mental health services to enhance health information processing capacity; and (4) implementing systematic eHealth literacy training with ongoing evaluation. These strategies and policies will help promote equitable access to and utilization of eHealth resources for all students, regardless of their location.

Acknowledgments

The research is supported by the National Social Science Foundation project of China (project number 21BGL301), entitled “A Multidimensional Evaluation of University Students’ Health and the Development of an Evidence-Based Intervention Model in the Context of the Healthy China Strategy.” We thank the participants and researchers of this survey for making this study possible. Because of ethical constraints and privacy commitments approved by the Institutional Review Board, the storage of raw data sets containing personally identifiable information about undergraduate students in public repositories is not permitted. In accordance with the stipulated data use agreement, applicants are obligated to sign a data use agreement and specify the intended use of the study. Following the acquisition of approval from the Southern Medical University Bioethics Committee, applicants can formally request deidentified data and analysis scripts from the corresponding author (YQ).

Authors' Contributions

YQ, YY, and ZL designed the study. YY, ZL, and QZ checked the data quality and performed the statistical analysis. QZ, YT, CW, and SL managed and checked all the data. YQ, YY, ZL, QZ, YX, XW, and ZW participated in the drafting, editing, reviewing, and revising of the manuscript. All authors read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Fairlie decomposition Model of eHealth literacy Status in Urban and Rural university students.
[DOCX File , 14 KB - jmir_v27i1e63671_app1.docx]

References

1. Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. Annu Rev Public Health 2021 Apr 01;42(Annual Review of Public Health):159-173 [FREE Full text] [doi: 10.1146/annurev-publhealth-090419-102529] [Medline: 33035427]
2. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021 Dec 16;9(12):1740 [FREE Full text] [doi: 10.3390/healthcare9121740] [Medline: 34946466]

XSL•FO
RenderX

3. Xiong Z, Zhang L, Li Z, Xu W, Zhang Y, Ye T. Frequency of online health information seeking and types of information sought among the general Chinese population: cross-sectional study. J Med Internet Res 2021 Dec 02;23(12):e30855 [FREE Full text] [doi: 10.2196/30855] [Medline: 34860676]
4. Maloney EK, D'Agostino TA, Heerdt A, Dickler M, Li Y, Ostroff JS, et al. Sources and types of online information that breast cancer patients read and discuss with their doctors. Palliat Support Care 2015 Apr;13(2):107-114 [FREE Full text] [doi: 10.1017/S1478951513000862] [Medline: 24182945]
5. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
6. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
7. Gilstad H. Toward a comprehensive model of eHealth literacy. In: Proceedings of the 2nd European Workshop on Practical Aspects of Health Information. 2014 May 19 Presented at: The 2nd European Workshop on Practical Aspects of Health Information; May 19, 2014; Trondheim, Norway p. 19-20 URL: http://ceur-ws.org/Vol-1251/paper7.pdf
8. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
9. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011 Dec;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
10. Chang F, Chiu C, Chen P, Miao N, Lee C, Chiang J, et al. Relationship between parental and adolescent eHealth literacy and online health information seeking in Taiwan. Cyberpsychol Behav Soc Netw 2015 Oct;18(10):618-624. [doi: 10.1089/cyber.2015.0110] [Medline: 26375050]
11. Dashti S, Abadibavil D, Roozbeh N. Evaluating e-health literacy, knowledge, attitude and practice regarding COVID-19 prevention and self-protection among Iranian students: a cross-sectional online survey. BMC Med Educ 2022 Mar 05;22(1):148 [FREE Full text] [doi: 10.1186/s12909-022-03210-3] [Medline: 35248025]
12. Kühn L, Bachert P, Hildebrand C, Kunkel J, Reitermayer J, Wäsche H, et al. Health literacy among university students: a systematic review of cross-sectional studies. Front Public Health 2021;9:680999 [FREE Full text] [doi: 10.3389/fpubh.2021.680999] [Medline: 35127605]
13. Daw J, Margolis R, Wright L. Emerging adulthood, emergent health lifestyles: sociodemographic determinants of trajectories of smoking, binge drinking, obesity, and sedentary behavior. J Health Soc Behav 2017 Jun;58(2):181-197 [FREE Full text] [doi: 10.1177/0022146517702421] [Medline: 28661779]
14. Nelson MC, Story M, Larson NI, Neumark-Sztainer D, Lytle LA. Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change. Obesity (Silver Spring) 2008 Oct;16(10):2205-2211. [doi: 10.1038/oby.2008.365] [Medline: 18719665]
15. Davy SR, Benes BA, Driskell JA. Sex differences in dieting trends, eating habits, and nutrition beliefs of a group of midwestern college students. J Am Diet Assoc 2006 Oct;106(10):1673-1677. [doi: 10.1016/j.jada.2006.07.017] [Medline: 17000202]
16. Racette SB, Deusinger SS, Strube MJ, Highstein GR, Deusinger RH. Weight changes, exercise, and dietary patterns during freshman and sophomore years of college. J Am Coll Health 2005;53(6):245-251. [doi: 10.3200/JACH.53.6.245-251] [Medline: 15900988]
17. Yahia N, Wang D, Rapley M, Dey R. Assessment of weight status, dietary habits and beliefs, physical activity, and nutritional knowledge among university students. Perspect Public Health 2016 Jul;136(4):231-244. [doi: 10.1177/1757913915609945] [Medline: 26475773]
18. Musaiger AO, Awadhalla MS, Al-Mannai M, AlSawad M, Asokan GV. Dietary habits and sedentary behaviors among health science university students in Bahrain. Int J Adolesc Med Health 2017 Apr 01;29(2). [doi: 10.1515/ijamh-2015-0038] [Medline: 26251982]
19. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 04;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
20. Wong AKC, Bayuo J, Wang S, Kwan RYC, Lam SC, Wong FKY. Factors associated with the perceptions of eHealth technology of Chinese nurses and nursing students. Nurse Educ Pract 2023 May;69:103605. [doi: 10.1016/j.nepr.2023.103605] [Medline: 37028224]
21. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
22. Kwon M, Oh J. The relationship between depression, anxiety, e-health literacy, and health-promoting behavior in nursing students during COVID-19. Medicine (Baltimore) 2023 Feb 10;102(6):e32809 [FREE Full text] [doi: 10.1097/MD.0000000000032809] [Medline: 36820579]

XSL•FO
RenderX

23. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
24. Nguyen HC, Nguyen MH, Do BN, Tran CQ, Nguyen TTP, Pham KM, et al. People with suspected COVID-19 symptoms were more likely depressed and had lower health-related quality of life: the potential benefit of health literacy. J Clin Med 2020 Mar 31;9(4):965 [FREE Full text] [doi: 10.3390/jcm9040965] [Medline: 32244415]
25. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006 Apr 21;8(2):e6 [FREE Full text] [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
26. Dadaczynski K, Okan O, Messer M, Leung AYM, Rosário R, Darlington E, et al. Digital health literacy and web-based information-seeking behaviors of university students in Germany during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2021 Jan 15;23(1):e24097 [FREE Full text] [doi: 10.2196/24097] [Medline: 33395396]
27. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
28. In 2023, the total number of students in higher education in my country will be 47.6319 million. Ministry of Education of the People's Republic of China. URL: http://www.moe.gov.cn/fbh/live/2024/55831/mtbd/202403/t20240301_1117760.html [accessed 2024-05-02]
29. Lu X, Zhang R. Association between eHealth literacy in online health communities and patient adherence: cross-sectional questionnaire study. J Med Internet Res 2021 Sep 13;23(9):e14908 [FREE Full text] [doi: 10.2196/14908] [Medline: 34515638]
30. Guo S, Yu X, Sun Y, Nie D, Li X, Wang L. Exploration of the localization and applicability of the eHEALS health literacy scale. Chinese Journal of Health Education 2013 Feb 20:106-108. [doi: 10.16168/j.cnki.issn.1002-9982.2013.02.019]
31. Yang T, Chen J, He L, Fu J. Correlation analysis between e-health literacy and health promotion lifestyle and emotion regulation of nursing students. Chinese Journal of Health Education 2024 Sep 20:835-838. [doi: 10.16168/j.cnki.issn.1002-9982.2024.09.013]
32. Mai J, Zhou L, Lin L. A cross-sectional study of students' e-health literacy in colleges and universities in Guangzhou. Health Vocational Education 2021 Jan 15:56-57.
33. He Y, Guo L, Zauszniewski JA, Wei M, Zhang G, Lei X, et al. A reliability and validity study of the electronic health literacy scale among stroke patients in China. Top Stroke Rehabil 2023 Apr;30(3):272-280. [doi: 10.1080/10749357.2021.2016100] [Medline: 34927574]
34. Li S, Cui G, Kaminga AC, Cheng S, Xu H. Associations between health literacy, eHealth literacy, and COVID-19-related health behaviors among Chinese college students: cross-sectional online study. J Med Internet Res 2021 May 06;23(5):e25600 [FREE Full text] [doi: 10.2196/25600] [Medline: 33822734]
35. Kuang H, Li J, Gu Z, Yue F. The mediating effect of e-health literacy on the relationship between college students' mental health and online psychological help-seeking behavior. China Journal of Health Psychology 2023 Nov 24:1876-1880. [doi: 10.13342/j.cnki.cjhp.2023.12.022]
36. Xu S, Li X. Research progress on e-health literacy of college students. Occupation and Health 2024 Feb 01:419-423. [doi: 10.13329/j.cnki.zyyjk.2024.0063]
37. Medero K, Merrill K, Ross MQ. Modeling access across the digital divide for intersectional groups seeking web-based health information: national survey. J Med Internet Res 2022 Mar 15;24(3):e32678 [FREE Full text] [doi: 10.2196/32678] [Medline: 35289761]
38. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
39. Questionnaire Star platform. WJX. URL: https://www.wjx.cn/ [accessed 2025-03-12]
40. Marsall M, Dinse H, Schröder J, Skoda E, Teufel M, Bäuerle A. Assessing electronic health literacy in individuals with the post-COVID-19 condition using the German Revised eHealth Literacy Scale: validation study. JMIR Form Res 2024 Apr 25;8:e52189 [FREE Full text] [doi: 10.2196/52189] [Medline: 38662429]
41. Zhang Y, Xu P, Sun Q, Baral S, Xi L, Wang D. Factors influencing the e-health literacy in cancer patients: a systematic review. J Cancer Surviv 2023 Apr;17(2):425-440 [FREE Full text] [doi: 10.1007/s11764-022-01260-6] [Medline: 36190672]
42. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
43. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
44. Fagbamigbe AF, Morakinyo OM, Balogun FM. Sex inequality in under-five deaths and associated factors in low and middle-income countries: a Fairlie decomposition analysis. BMC Public Health 2022 Feb 16;22(1):334 [FREE Full text] [doi: 10.1186/s12889-022-12679-y] [Medline: 35172780]
45. Zhang J, Li D, Gao J. Health disparities between the rural and urban elderly in China: a cross-sectional study. Int J Environ Res Public Health 2021 Jul 29;18(15):8056 [FREE Full text] [doi: 10.3390/ijerph18158056] [Medline: 34360346]

XSL•FO
RenderX

46. Yuan L, Yu B, Bing Y, Du M, Zhao Z, Sun J. Gender differences in depressive symptoms among Chinese older adults based on fairlie decomposition analysis. Heliyon 2023 Nov;9(11):e21555 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e21555] [Medline: 37954279]
47. Zhao Z, Yu B, Hu F, Zheng C, Gui J, Liu J, et al. Decomposition and comparative analysis of health inequities between the male and female older adults in China: a national cross-sectional study. BMC Public Health 2023 Oct 19;23(1):2045 [FREE Full text] [doi: 10.1186/s12889-023-15814-5] [Medline: 37858126]
48. Fairlie RW. The absence of the African-American owned business: an analysis of the dynamics of self-employment. SSRN. 2014 Jan 12. URL: https://ssrn.com/abstract=2378091 [accessed 2025-03-16]
49. Tsukahara S, Yamaguchi S, Igarashi F, Uruma R, Ikuina N, Iwakura K, et al. Association of eHealth literacy with lifestyle behaviors in university students: questionnaire-based cross-sectional study. J Med Internet Res 2020 Jun 24;22(6):e18155. [doi: 10.2196/18155] [Medline: 32579126]
50. Shudayfat T, Hani SB, Shdaifat E, Al-Mugheed K, Alsenany SA, Farghaly Abdelaliem SM. Electronic health literacy and its association with lifestyle behavior among undergraduate students: a cross-sectional survey. Digit Health 2023;9:20552076231185429 [FREE Full text] [doi: 10.1177/20552076231185429] [Medline: 37434731]
51. Holch P, Marwood JR. EHealth literacy in UK teenagers and young adults: exploration of predictors and factor structure of the eHealth Literacy Scale (eHEALS). JMIR Form Res 2020 Sep 08;4(9):e14450 [FREE Full text] [doi: 10.2196/14450] [Medline: 32897230]
52. Gu Y, Kalibatseva Z, Song X, Prakash S. Effective use of online COVID-19 information and eHealth information literacy among US university students. J Am Coll Health 2024 Jul;72(5):1458-1465. [doi: 10.1080/07448481.2022.2080505] [Medline: 35658100]
53. Bailey CE, Kohler WJ, Makary C, Davis K, Sweet N, Carr M. eHealth literacy in otolaryngology patients. Ann Otol Rhinol Laryngol 2019 Nov;128(11):1013-1018. [doi: 10.1177/0003489419856377] [Medline: 31195809]
54. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016 Feb;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
55. Gustafson DH, McTavish FM, Stengle W, Ballard D, Hawkins R, Shaw BR, et al. Use and impact of eHealth system by low-income women with breast cancer. J Health Commun 2005;10 Suppl 1:195-218. [doi: 10.1080/10810730500263257] [Medline: 16377608]
56. Mengestie ND, Yilma TM, Beshir MA, Paulos GK. eHealth literacy of medical and health science students and factors affecting eHealth literacy in an Ethiopian university: a cross-sectional study. Appl Clin Inform 2021 Mar;12(2):301-309 [FREE Full text] [doi: 10.1055/s-0041-1727154] [Medline: 33827143]
57. Zhang J, Zhong M, Chen Z. A study on the association between e-health literacy and health behaviors among college students in Guangzhou, China. China Public Health Management 2023 Aug 20;39(04):497-501. [doi: 10.19568/j.cnki.23-1318.2023.04.0015]
58. Giger JT, Barnhart S, Feltner F, Slone M, Lawler MJ, Windsor L, et al. Validating the eHealth literacy scale in rural adolescents. J Rural Health 2021 Jun;37(3):504-516. [doi: 10.1111/jrh.12509] [Medline: 32881112]
59. Korkmaz Aslan G, Kartal A, Turan T, Taşdemir Yiğitoğlu G, Kocakabak C. Association of electronic health literacy with health-promoting behaviours in adolescents. Int J Nurs Pract 2021 Apr;27(2):e12921. [doi: 10.1111/ijn.12921] [Medline: 33615625]
60. Wang Y, Zhang J, Huang K. A study on the effect of school and family environments and self-efficacy on health literacy of college students. Front Public Health 2024;12:1449819 [FREE Full text] [doi: 10.3389/fpubh.2024.1449819] [Medline: 39211899]
61. Chaibi LS, Fekih-Romdhane F, Ben Said Saffar C, Cherif W, Cheour M. Mental health information-seeking behaviors and e-health literacy in Tunisian community adults. Eur Psychiatr 2023 Jul 19;66(S1):S858-S858. [doi: 10.1192/j.eurpsy.2023.1818]
62. Peng R, Fan Y, Li J, Zhu F, Tian Q, Zhang X. Abnormalities of electroencephalography microstates in patients with depression and their association with cognitive function. World J Psychiatry 2024 Jan 19;14(1):128-140 [FREE Full text] [doi: 10.5498/wjp.v14.i1.128] [Medline: 38327889]
63. Szota M, Rogowska AM, Kwaśnicka A, Chilicka-Hebel K. The indirect effect of future anxiety on the relationship between self-efficacy and depression in a convenience sample of adults: revisiting social cognitive theory. J Clin Med 2024 Aug 19;13(16):4897 [FREE Full text] [doi: 10.3390/jcm13164897] [Medline: 39201038]
64. Berens E, Pelikan J, Schaeffer D. The effect of self-efficacy on health literacy in the German population. Health Promot Int 2022 Feb 17;37(1):85. [doi: 10.1093/heapro/daab085] [Medline: 34115848]
65. Duong VT, Lin I, Sorensen K, Pelikan JM, Van Den Broucke S, Lin Y, et al. Health literacy in Taiwan: a population-based study. Asia Pac J Public Health 2015 Nov;27(8):871-880. [doi: 10.1177/1010539515607962] [Medline: 26419635]
66. Amoah PA, Phillips DR. Socio-demographic and behavioral correlates of health literacy: a gender perspective in Ghana. Women Health 2020 Feb;60(2):123-139. [doi: 10.1080/03630242.2019.1613471] [Medline: 31092133]
67. Wang J, Shahzad F. A visualized and scientometric analysis of health literacy research. Front Public Health 2021;9:811707 [FREE Full text] [doi: 10.3389/fpubh.2021.811707] [Medline: 35155357]

XSL•FO
RenderX

68. Wu T, Wang X, Zhao S, Xiao Y, Shen M, Han X, et al. Socioeconomic determinants of melanoma-related health literacy and attitudes among college students in China: a population-based cross-sectional study. Front Public Health 2021;9:743368 [FREE Full text] [doi: 10.3389/fpubh.2021.743368] [Medline: 34858925]
69. Cinar S, Boztepe H, Ozcebe H. The effect of sociodemographic characteristics of adolescents on e-health literacy. Eur J Public Health 2020 Jan 1;30(supplement 5):ckaa166.934. [doi: 10.1093/eurpub/ckaa166.934]
70. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
71. Matthews EB, Savoy M, Paranjape A, Washington D, Hackney T, Galis D, et al. Acceptability of health information exchange and patient portal use in depression care among underrepresented patients. J Gen Intern Med 2022 Nov 07;37(15):3947-3955. [doi: 10.1007/s11606-022-07427-2] [Medline: 35132548]
72. Aljassim N, Ostini R. Health literacy in rural and urban populations: a systematic review. Patient Educ Couns 2020 Oct;103(10):2142-2154. [doi: 10.1016/j.pec.2020.06.007] [Medline: 32601042]
73. Dunbar JA, Peach E. The disparity called rural health: what is it, and what needs to be done? Aust J Rural Health 2012 Dec;20(6):290-292. [doi: 10.1111/ajr.12000] [Medline: 23181811]
74. Clark RA, Coffee N, Turner D, Eckert KA, van Gaans D, Wilkinson D, Cardiac AccessibilityRemoteness Index for Australia (Cardiac ARIA) Project Group. Application of geographic modeling techniques to quantify spatial access to health services before and after an acute cardiac event: the Cardiac Accessibility and Remoteness Index for Australia (ARIA) project. Circulation 2012 Apr 24;125(16):2006-2014. [doi: 10.1161/CIRCULATIONAHA.111.083394] [Medline: 22451583]
75. Bell E, Orpin P. Self-management of chronic conditions: implications for rural physicians of a demonstration project Down Under. Can J Rural Med 2006;11(1):33-40 [FREE Full text] [Medline: 16454970]
76. Patterson KA, Cleland V, Venn A, Blizzard L, Gall S. A cross-sectional study of geographic differences in health risk factors among young Australian adults: the role of socioeconomic position. BMC Public Health 2014 Dec 15;14:1278 [FREE Full text] [doi: 10.1186/1471-2458-14-1278] [Medline: 25512127]
77. Shiferaw KB, Tilahun BC, Endehabtu BF, Gullslett MK, Mengiste SA. E-health literacy and associated factors among chronic patients in a low-income country: a cross-sectional survey. BMC Med Inform Decis Mak 2020 Aug 06;20(1):181 [FREE Full text] [doi: 10.1186/s12911-020-01202-1] [Medline: 32762745]
78. El Benny M, Kabakian-Khasholian T, El-Jardali F, Bardus M. Application of the eHealth literacy model in digital health interventions: scoping review. J Med Internet Res 2021 Jun 03;23(6):e23473 [FREE Full text] [doi: 10.2196/23473] [Medline: 34081023]
79. McAnally K, Hagger MS. Health literacy, social cognition constructs, and health behaviors and outcomes: a meta-analysis. Health Psychol 2023 Apr;42(4):213-234. [doi: 10.1037/hea0001266] [Medline: 37023324]
80. Yuen EK, Gangi CE, Barakat K, Harrison F. College students' utilization of the Internet to search for mental health information: effects on mental health literacy, stigma, and help-seeking. J Am Coll Health 2025 Mar;73(3):961-971. [doi: 10.1080/07448481.2024.2404948] [Medline: 39303085]
81. Raghupathi V, Raghupathi W. The influence of education on health: an empirical assessment of OECD countries for the period 1995-2015. Arch Public Health 2020;78:20 [FREE Full text] [doi: 10.1186/s13690-020-00402-5] [Medline: 32280462]
82. Li X. Health literacy and its influencing factors among college students in tianjin in 2021. Chinese Journal of Prevention and Control of Chronic Diseases 2023 Jan 15:66-70. [doi: 10.16386/j.cjpccd.issn.1004-6194.2023.01.015]
83. Zhou W, Luo C, Fan C, Li Y, Ma S. Analysis of the current situation and influencing factors of health literacy of college students in Guangzhou. Chinese Journal of Disease Control and Prevention 2014 Jul 10:654-658 [FREE Full text]

Abbreviations

CHARLS: China Health and Retirement Longitudinal Study
eHEALS: eHealth Literacy Scale
FDA: Fairlie decomposition analysis
GAD-7: 7-item Generalized Anxiety Disorder
OR: odds ratio
PHQ-9: 9-item Patient Health Questionnaire

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 26.06.24; peer-reviewed by A Ahsan, JR Bautista; comments to author 17.09.24; revised version received 09.12.24; accepted 25.02.25; published 26.03.25.

Please cite as:
Yu Y, Liang Z, Zhou Q, Tuersun Y, Liu S, Wang C, Xie Y, Wang X, Wu Z, Qian Y
Decomposition and Comparative Analysis of Urban-Rural Disparities in eHealth Literacy Among Chinese University Students: Cross-Sectional Study
J Med Internet Res 2025;27:e63671
URL: https://www.jmir.org/2025/1/e63671
doi:10.2196/63671
PMID:

©Yao Yu, Zhenning Liang, Qingping Zhou, Yusupujiang Tuersun, Siyuan Liu, Chenxi Wang, Yuying Xie, Xinyu Wang, Zhuotong Wu, Yi Qian. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identification of Intracranial Germ Cell Tumors Based on Facial Photos: Exploratory Study on the Use of Deep Learning for Software Development

Yanong Li[1*], MD; Yixuan He[2*], MD; Yawei Liu[3,4], MSc; Bingchen Wang[5], MD; Bo Li[1], MD, PhD; Xiaoguang Qiu[1,6], MD

[1]Department of Radiation Oncology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

[2]Department of Pediatric Department, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

[3]School of Basic Medical Sciences, Capital Medical University, Beijing, China

[4]Beijing Luhe Hospital Affiliated to Capital Medical University, Beijing, China

[5]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China

[6]Beijing Neurosurgical Institute, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xiaoguang Qiu, MD
Department of Radiation Oncology
Beijing Tiantan Hospital
Capital Medical University
119 West Southern 4th Ring Road
Fengtai District
Beijing, 100070
China
Phone: 86 10 59975581
Email: qiuxiaoguang@bjtth.org

Abstract

Background: Primary intracranial germ cell tumors (iGCTs) are highly malignant brain tumors that predominantly occur in children and adolescents, with an incidence rate ranking third among primary brain tumors in East Asia (8%-15%). Due to their insidious onset and impact on critical functional areas of the brain, these tumors often result in irreversible abnormalities in growth and development, as well as cognitive and motor impairments in affected children. Therefore, early diagnosis through advanced screening techniques is vital for improving patient outcomes and quality of life.

Objective: This study aimed to investigate the application of facial recognition technology in the early detection of iGCTs in children and adolescents. Early diagnosis through advanced screening techniques is vital for improving patient outcomes and quality of life.

Methods: A multicenter, phased approach was adopted for the development and validation of a deep learning model, GVisageNet, dedicated to the screening of midline brain tumors from normal controls (NCs) and iGCTs from other midline brain tumors. The study comprised the collection and division of datasets into training (n=847, iGCTs=358, NCs=300, other midline brain tumors=189) and testing (n=212, iGCTs=79, NCs=70, other midline brain tumors=63), with an additional independent validation dataset (n=336, iGCTs=130, NCs=100, other midline brain tumors=106) sourced from 4 medical institutions. A regression model using clinically relevant, statistically significant data was developed and combined with GVisageNet outputs to create a hybrid model. This integration sought to assess the incremental value of clinical data. The model's predictive mechanisms were explored through correlation analyses with endocrine indicators and stratified evaluations based on the degree of hypothalamic-pituitary-target axis damage. Performance metrics included area under the curve (AUC), accuracy, sensitivity, and specificity.

Results: On the independent validation dataset, GVisageNet achieved an AUC of 0.938 (P<.01) in distinguishing midline brain tumors from NCs. Further, GVisageNet demonstrated significant diagnostic capability in distinguishing iGCTs from the other midline brain tumors, achieving an AUC of 0.739, which is superior to the regression model alone (AUC=0.632, P<.001) but less than the hybrid model (AUC=0.789, P=.04). Significant correlations were found between the GVisageNet's outputs and 7

XSL•FO
RenderX

endocrine indicators. Performance varied with hypothalamic-pituitary-target axis damage, indicating a further understanding of the working mechanism of GVisageNet.

Conclusions: GVisageNet, capable of high accuracy both independently and with clinical data, shows substantial potential for early iGCTs detection, highlighting the importance of combining deep learning with clinical insights for personalized health care.

(J Med Internet Res 2025;27:e58760) doi:10.2196/58760

KEYWORDS

deep learning; facial recognition; intracranial germ cell tumors; endocrine indicators; software development; artificial intelligence; machine learning models; software engineering; neural networks; algorithms; cohort studies

Introduction

Intracranial germ cell tumors (iGCTs) are primary malignant brain tumors of the central nervous system, predominantly affecting children and adolescents [1]. In East Asian countries, iGCTs account for 15% of all pediatric and adolescent brain tumors, making them the third most common primary brain tumor [2]. In other regions, the incidence of iGCTs is only 2%-3%, classifying them as rare tumors [1]. IGCTs typically occur in the midline structures of the brain, with approximately 90% located in the pituitary and pineal regions, and a minority in the fourth ventricle and brain parenchyma [3]. IGCTs can be treated with radiotherapy and chemotherapy, with only about 20% of patients requiring surgical resection [4,5]. The symptoms of iGCTs vary depending on the tumor's location, size, and cerebrospinal fluid dissemination. Common symptoms include headaches, nausea, vomiting, visual changes, endocrine disorders, ataxia, seizures, behavioral and cognitive changes, and cranial nerve dysfunction [6]. The manifestation and severity of these symptoms differ among individuals and according to the specific tumor characteristics. Patients often present with various symptoms and are diagnosed with an intracranial tumor through computed tomography or magnetic resonance imaging scans. However, due to the rarity of iGCTs in most regions, patients are frequently misdiagnosed with craniopharyngiomas, pituitary tumors, gliomas, or pinealocytomas and undergo unnecessary surgical resection [7]. Postoperative pathology often reveals iGCTs, necessitating subsequent radiotherapy and chemotherapy, thus indirectly leading to overtreatment [8].

According to the clinical consensus, iGCTs can be clinically diagnosed when tumor markers (eg, human chorionic gonadotropin [HCG] and alpha-fetoprotein) are elevated [7]. However, due to a lack of diagnostic experience, these tumor markers are often overlooked at initial diagnosis, leading to unwarranted surgical treatment. Therefore, developing a noninvasive, efficient, and universally applicable screening tool is crucial to provide clinicians with indicative information at the first patient visit, guiding targeted examinations. Since iGCTs patients exhibit facial feature changes compared with normal children, we hypothesize that facial recognition algorithms can be developed to alert clinicians at the initial consultation, providing personalized diagnostic approaches [9,10].

The application of facial recognition algorithms is extensive, especially those based on machine learning algorithms, which have been profoundly studied in the medical field. For instance, Bottinelli et al [11] used a binocular rivalry task to assess emotional processing in patients with panic disorder (PD), finding that with PD had a greater initial threat bias toward fearful faces compared with controls, indicating that threat-evoking stimuli more rapidly captured their attention. Yeung et al [12] analyzed 148 studies on facial emotion recognition in autism spectrum disorder, revealing significant impairments in recognizing all basic facial emotions in individuals with autism spectrum disorder compared with typically developing controls. Nigam et al [13] examined facial emotion recognition deficits in remitted patients with bipolar disorder and their first-degree relatives compared with healthy controls, finding significant deficits in recognizing emotions such as fear, anger, surprise, and happiness in patients with bipolar disorder compared with first-degree relatives and healthy controls. iGCTs are a type of intracranial tumor often accompanied by significant endocrine abnormalities, and facial recognition technology holds the potential for uncovering diagnostic information. Nevertheless, current research on this aspect remains unclear

Therefore, this study aims to design a deep learning-driven binary facial recognition model using frontal facial photographs of patients to first distinguish between midline brain tumors and normal controls and then differentiate iGCTs from other midline brain tumors. We will initially screen for the optimal model among commonly used deep learning (DL) models in facial recognition, using development and independent external validation datasets composed of open-source databases (Wild Young Labeled Faces) and multicenter clinical patient facial photographs. Finally, we will develop visualization software to demonstrate the practical clinical application of our model.

Methods

Study Design

The objective of this multicenter retrospective diagnostic study was to develop and validate a 2 stage DL facial recognition classification model, GVisageNet, using frontal facial photographs to distinguish patients with midline brain tumors and normal controls (NCs), and further distinguish iGCTs from other patients with midline brain tumors, thereby facilitating early detection. Data collection occurred across 4 Chinese medical institutes from January 2010 to May 2023. The study design and participant flow are illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. The study design and participant flow. DL: deep learning; iGCT: intracranial germ cell tumor; NC: normal control.

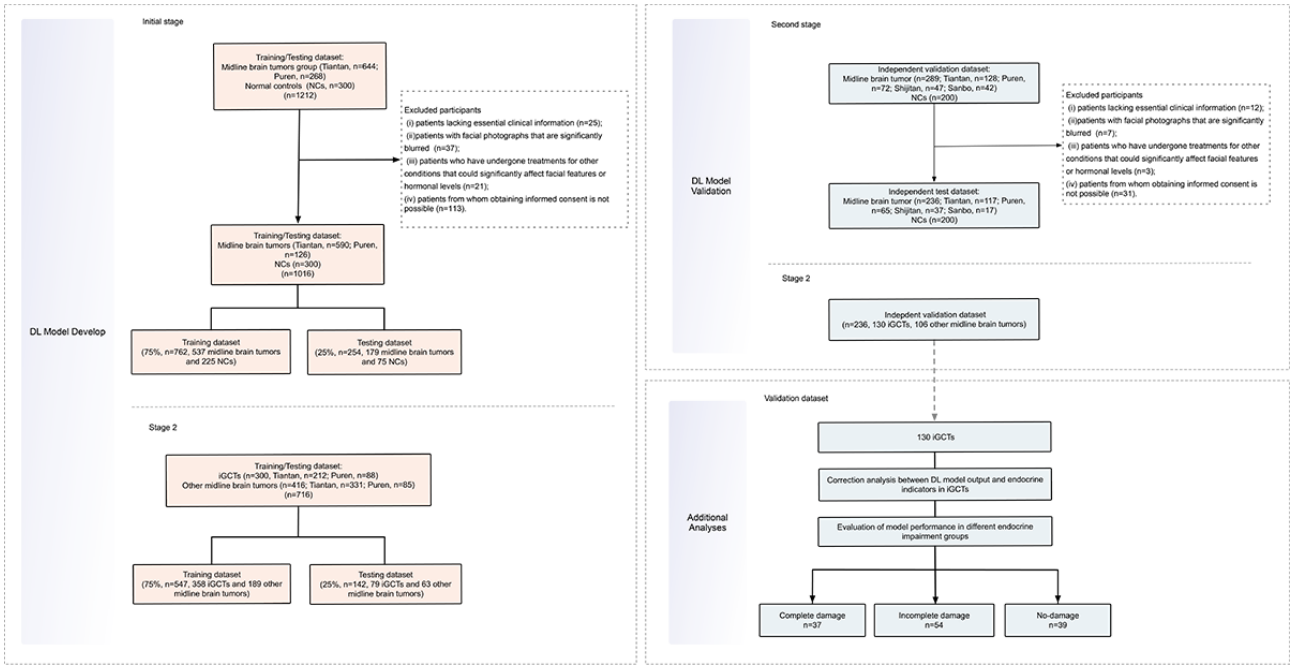

Participants

Inclusion criteria were (1) patients diagnosed pathologically with midline brain tumors (specify tumors originating in the sellar region, pineal region, or the fourth ventricle); (2) availability of frontal facial photographs taken within 2 weeks before the date of diagnosis; (3) gender- and age-matched NCs from the “Wild Young Labeled Faces” dataset with age and gender matching from the GitHub repository managed by the Institute of Systems and Robotics at the University of Coimbra. Exclusion criteria were (1) patients lacking essential clinical information, including but not limited to demographics (eg, age and gender), symptoms, and hormonal levels; (2) patients with facial photographs that are significantly blurred, to the extent that key facial landmarks or features are indistinguishable; (3) patients who have undergone treatments for other conditions that could significantly affect facial features or hormonal levels; and (4) patients from whom obtaining informed consent is not possible. Finally, patients were categorized into the midline brain tumors group and NCs, and further divided into iGCTs and the other midline brain tumor groups.

Study Setting

This investigation was systematically conducted in 3 phases, using a multicenter approach for patient recruitment and data collection. During the initial phase, NCs and patients diagnosed with midline brain tumors (including iGCTs and other midline brain tumors) were recruited from the wild young labeled faces dataset, as well as from Tiantan, Puren Hospital, Capital Medical University. This can be deleted from the text. In this phase, our goal was to construct a binary tandem DL model (GVisageNet). We randomly divided the NCs and patients with midline brain tumors (n=1016) into 2 groups, 75% for training the GvisageNet and 25% for testing its performance (762:254). Furthermore, an independent external dataset was compiled from 4 distinct medical institutions (Tiantan, Puren, Shijitan, and Sanbo Hospital, Capital Medical University) between January 2020 and May 2023, aiming to establish an independent validation dataset (n=436) for validating the DL model. Subsequently, the patients with midline brain tumors (n=716) from the first phase were randomly divided into a training dataset and a testing dataset in a 75:25 ratio to differentiate iGCTs from other midline brain tumor patients (574:142). Finally, the facial photographs of patients with midline brain tumors (n=236) from the independent external validation dataset were used as a separate validation dataset in the second phase.

Since both iGCTs and other midline brain tumors affect the brain's midline functional areas in clinical diagnosis, leading to similar facial characteristics caused by impacts on the patient’s endocrine function, differential diagnosis becomes even more challenging. However, incorrect predictions could result in misdiagnosis, misinterpretation, and inefficient use of resources. To further enhance predictive accuracy, the subsequent phase of the study aimed to develop a regression model using clinical variables marked by statistical disparities between the 2 groups (iGCTs vs other midline brain tumors). This included demographic characteristics, clinical manifestations, and hormonal profiles, where significant differences were observed. This model was then integrated with the outputs generated by GVisageNet (from the second stage), enabling the construction of an integrated hybrid model. The primary goal of this integration was to clarify the contribution of clinical data in improving the model’s predictive performance.

In the final phase of the study, the correlation analysis between the outputs of GVisageNet and 15 endocrine indicators in the independent validation dataset was conducted to explore the potential mechanism of GVisageNet. Finally, stratified studies were carried out in the independent validation dataset based on the degree of hypothalamic-pituitary-target (HPT) axis damage (complete damage, incomplete damage, and no damage), further verifying that endocrine changes are the potential mechanism behind the effectiveness of GVisageNet.

XSL•FO
RenderX

Data Collection

Frontal views of each patient were routinely captured at the initial hospital visit using a standardized protocol. This involved a digital camera with a resolution exceeding 10 million pixels (the criteria and details of photograph taking are presented in Method S1 in Multimedia Appendix 1).

In studying the correlation between facial recognition features and endocrine functions, the selection of specific clinical indicators is designed to comprehensively assess patients' endocrine conditions and their potential impacts on facial characteristics. This includes baseline data such as gender, age, BMI, initial symptoms, disease duration, and the primary site of the tumor, which provide an initial understanding of the patient's overall health and endocrine function. Key endocrine indicators include the adrenal axis (Adrenocorticotropic hormone), thyroid axis (thyroid stimulating hormone [TSH], triiodothyronine, free triiodothyronine [FT3], thyroxine, and free thyroxine [FT4]), growth hormone-insulin-like growth factor-I axis (growth hormone [GH], insulin-like growth factor 1, insulin-like growth factor binding protein 3 [IGFBP-3]), and gonadal axis (follicle-stimulating hormone [FSH], luteinizing hormone [LH], testosterone, estradiol, progesterone, and prolactin). Variations in these hormone levels can lead to significant changes in facial features, such as puffiness, growth retardation, or precocious puberty. In addition, routine monitoring of 24-hour urine volume, specific gravity, and osmolality indirectly reflects the patient's endocrine health status, thereby supporting the potential application of facial recognition technology in clinical diagnostics. This comprehensive data collection is crucial for exploring the link between facial characteristics and endocrine functions, holding significant academic value.

Pathological Diagnosis

The criterion for correct GVisageNet prediction is consistency with pathological results. All patients enrolled in the study were classified according to the 2021 WHO (World Health Organization) pathological classification [14]. Two experienced neuropathologists, XL and YJH, each with 10 years of experience in neuropathology, independently reviewed the histological diagnoses for both the development and independent test datasets. Their reviews were aligned with the 2021 WHO classification of tumors of the central nervous system [15]. Any specimens that did not conform to the 2021 WHO classification underwent a secondary review by other neuropathologists (JD with 30 years of experience). The criteria for immunohistochemical and diagnostic assessments in the independent validation dataset were consistent with those used for the development dataset.

Facial Photo Preprocessing

The quality of facial photographs was evaluated by 2 investigators who were blinded to the study design, as outlined in the protocol (Method S2 in Multimedia Appendix 1.). Those that were deemed qualified were processed by cropping to isolate only the facial area, thus removing any excess background and clothing, and were resized to the dimensions of 256×256 pixels. Figure 2 depicts the ResNet-50 model workflow for facial photo analysis, including preprocessing, configuration, training with loss function and optimizer selection, and evaluation using receiver operating characteristic (ROC) and precision-recall curves on an independent external test dataset.

XSL•FO
RenderX

Figure 2. The figure depicts the GVisageNet model workflow for facial photo analysis, including preprocessing, configuration, training with loss function and optimizer selection, and evaluation using ROC and precision-recall curves on an independent external validation data set. iGCTs: intracranial germ cell tumors; NCs: normal controls.

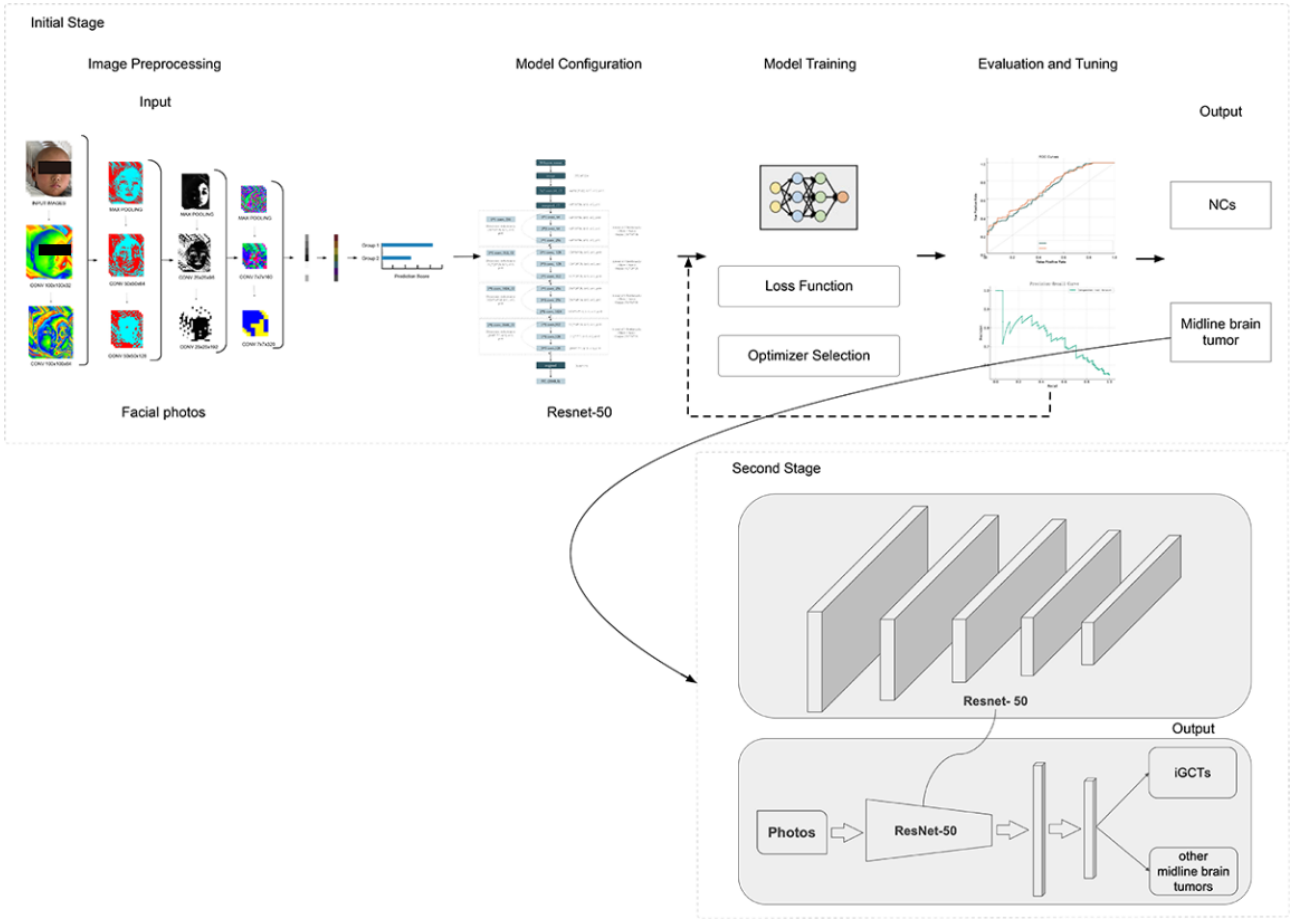

Facial Recognition Deep Learning Model Development

DL Model Architecture

Our facial recognition DL model (GVisageNet) was primarily based on a Convolutional Neural Network (CNN; Figure S1 in Multimedia Appendix 1). We evaluated 3 of the most commonly used candidate models in the field of facial recognition—AlexNet, ResNet-50, and Inception—with ResNet-50 emerging as the most effective through comparative analysis. The development dataset was randomly divided into training and testing datasets in a 75:25 ratio. We used the area under the curve (AUC) to evaluate the classification performance on the testing dataset. To address sample imbalance, resampling techniques were applied to the training dataset, as detailed in Method S3 in Multimedia Appendix 1.

Development and Validation of the Facial Recognition DL Model

We selected ResNet [16], specifically ResNet-50 [17], as our model due to its training efficiency. ResNet-50 is composed of 4 blocks, each containing CNN layers of various depths, culminating in a total of 50 CNN layers (Figure S2 in Multimedia Appendix 1). A fully connected layer was added following ResNet-50 for classification purposes. During training, images were resized to 256×256 pixels. Each patient's facial photos were combined into a 12-channel image stack, ensuring a comprehensive representation of facial features. The supervised model underwent 10 training epochs, using Adam as the optimizer. The initial learning rate was set at $1e^{-4}$ and adjusted through cosine learning rate decay. The loss function used was cross-entropy. Training was conducted on the training dataset and testing was performed on the testing dataset. Upon completion, the DL model produced probabilities for each image belonging to one of 2 categories, with the highest probability determining the category label. The DL framework was based on PyTorch and operated on an Ubuntu 18.04 system equipped with an NVIDIA V100 Tensor Core GPU. Method S4 in Multimedia Appendix 1 shows the details of the development and validation ResNet-50 based DL model.

Evaluating the DL Model

We developed 2 additional models to detect iGCTs, aiming to compare their performance with our primary model. First, we constructed a multivariable logistic regression model using 7 clinical variables, which showed statistical significance in the training dataset as detailed in Table 1. Second, we built a hybrid model that integrates the output of GVisageNet with the results from the logistic regression model, Figure S3 in Multimedia Appendix 1. Both differential diagnostic models underwent training on the training dataset, testing on the testing dataset, and validation on the independent validation dataset.

XSL•FO
RenderX

Table 1. The demographic data, symptoms, and endocrine hormone levels of patients across all datasets.

Variable	Development dataset						Independent external validation dataset		
	Training dataset (n=547)			Testing dataset (n=142)			(n=236)		
	Intracranial germ cell tumors (n=358)	Other midline brain tumors (n=189)	*P* values	Intracranial germ cell tumors (n=79)	Other midline brain tumors (n=63)	*P* values	Intracranial germ cell tumors (n=130)	Other midline brain tumors (n=106)	*P* values
Demographics									
Age (years), mean (SD)	16.95 (8.01)	10.47 (4.72)	.31	15.61 (5.86)	10.61 (4.48)	.42	16.29 (5.49)	14.67 (4.45)	.76
Sex									
Female: Male, n:n	167:191	109:80	.52	51:68	33:30	.67	72:58	61:45	.72
Duration (months)	15.40	19.60	.39	18.20	17.50	.62	14.60	23.40	.15
BMI	18.28	18.59	.88	19.64	16.44	.50	15.82	17.66	.50
Primary tumor site (Sellar: Pineal: Fourth ventricle), n:n	186:160:12	89:59: 41	.40	42:31:6	19:29:15	.87	59:61:10	47:51:8	.37
Symptoms									
Symptoms of cranial hypertension (headache, vomiting, papilledema), %	76.33	67.42	.41	61.43	52.29	.32	47.66	58.72	.66
Vision changes (vision loss, visual field loss), %	43.24	33.21	.56	39.43	48.21	.62	29.56	36.77	.47
Growth retardation, %	27.64	10.21	—[a]	—	—	—	—	—	—
Limb movement disorder, %	19.88	5.31	<.01[b]	25.66	7.62	.018[b]	17.68	10.32	.22
Hypothalamic syndrome, %									
Diabetes insipidus	44.56	21.01	.03[b]	40.98	17.66	<.01[b]	38.21	21.44	.34
Abnormal feeding	9.61	7.77	.24	11.43	16.59	.11	20.44	17.43	.52
Sleep-wake abnormality	11.21	10.33	.11	20.42	9.61	.03[b]	25.33	11.64	.02[b]
Sexual abnormality	8.89	4.23	.04[b]	11.56	6.33	.22	28.42	10.21	.03[b]
Affective disturbance	33.21	11.21	.04[b]	29.56	9.43	<.01[b]	19.08	22.44	.54
Convulsive seizure	21.99	9.82	.05[b]	36.99	6.55	<.01[b]	18.42	10.33	.08
Endocrine hormone levels									
Corticotropic hormone (ACTH), pg/mL	26.63	22.07	.42	22.54	24.67	.78	21.33	19.06	.13
Cortisol (COR), μg/dL	7.44	14.32	.07	6.79	12.03	.05[b]	9.65	13.42	.24
Thyroid stimulating hormone, mIU/L	0.57	2.33	.02[b]	1.02	3.46	.06	0.99	2.77	<.01[b]
Free triiodothyronine, ng/dL	0.77	1.23	.13	0.39	1.45	.08	1.55	2.47	.08
Free thyroxine, ng/dL	1.22	1.79	.452	1.44	1.87	.459	0.99	1.68	.09
Growth hormone, ng/dL	8.55	21.43	.03[b]	10.22	25.64	.02[b]	9.08	22.37	<.01[b]
Insulin-like growth factor-1, ng/mL	189.35	246.88	.48	207.56	255.64	.64	125.42	366.90	.04[b]

XSL•FO
RenderX

Variable	Development dataset						Independent external validation dataset		
	Training dataset (n=547)			Testing dataset (n=142)			(n=236)		
	Intracranial germ cell tumors (n=358)	Other midline brain tumors (n=189)	*P* values	Intracranial germ cell tumors (n=79)	Other midline brain tumors (n=63)	*P* values	Intracranial germ cell tumors (n=130)	Other midline brain tumors (n=106)	*P* values
Insulin-like growth factor binding protein-3, mcg/mL	2.33	5.63	.21	3.42	4.55	.64	1.99	4.43	.07
Follicle-stimulating hormone, IU/mL	3.56	7.58	.42	4.43	5.46	.79	3.31	5.73	.57
Luteinizing hormone, IU/L	0.43	0.68	.43	0.58	0.77	.64	0.46	0.37	.68
Testosterone, ng/dL	78.63	87.43	.88	66.42	69.02	.67	54.37	80.09	.08
Estradiol, pg/mL	99.45	121.44	.67	122.44	109.78	.78	133.21	90.76	.29
Progesterone, ng/mL	3.77	6.42	.23	5.33	5.89	.82	8.98	6.77	.75
Prolactin, µg/L	37.88	33.42	.80	40.66	37.42	.69	29.88	45.53	.35

[a]Not applicable.

[b]Indicate statistical difference.

The *P* values were determined by the one-way analysis of variance test for age and tumor volume, and the chi-square test for gender, T1WI performance, T2WI performance, enhancement appearances, degree of enhancement, hemorrhage, and necrosis.

The Working Mechanism of Interpretation Model Potential

To explore the hypothesis that endocrine changes contribute to alterations in facial characteristics, we conducted a correlation analysis between prediction scores and endocrine indicators. This analysis focused on the independent validation dataset, aimed to uncover the underlying mechanisms. In addition, we segmented the independent validation dataset into 3 subgroups based on the extent of HPT axis damage: complete damage, incomplete damage, and no damage, Method S5 in Multimedia Appendix 1 defines the “complete damage,” “incomplete damage,” and “no-damage” of the HPT axis damage. The performance of GVisageNet was then compared across these subgroups to determine if the degree of endocrine damage influenced the model’s effectiveness.

Statistical Analysis

For our statistical analysis, we used SPSS (version 25.0) software (IBM Corp). Continuous variables were presented as mean (SD), and discrete variables as percentages. We assessed the normality of continuous variables using the Shapiro-Wilk test. Depending on the results, either the independent samples *t* test, Friedman test, or the Mann-Whitney *U* test was used for comparison. Categorical variables, such as gender and endocrine abnormalities in patients with NCs and midline brain tumors (including iGCTs and other midline brain tumors), were analyzed using the chi-square test. We calculated the 95% exact CIs for both proportions and mean differences.

To evaluate the performance of the DL model, we computed its diagnostic accuracy, sensitivity, specificity, and the area under the ROC curve, using pathological diagnosis as the benchmark. The model’s performance was evaluated at 2 specific points on the receiver operating characteristic ROC curve, chosen for their maximum sensitivity and specificity. We used the Delong test [18] to compare the AUCs of the different models. Multinomial logistic regression was used to examine the correlation between the prediction scores and the endocrine indicators. All comparisons were conducted as 2-sided tests, with a *P* value of less than .05 considered statistically significant.

Ethical Considerations

Ethical approval was granted by the institutional review board of Beijing Tiantan Hospital, Capital Medical University (KY2021-142-02). Informed consent was secured from all patients or their guardians. Privacy and Confidentiality: In this study, all participant data, including images, are processed with strict measures to ensure privacy and confidentiality. Each participant’s data, including facial photographs, is anonymized and deidentified before analysis to prevent any direct or indirect identification. For cases where anonymization is not feasible, such as if the integrity of facial features is required for model training or analysis, robust data protection protocols are implemented. These include secure storage solutions, restricted access, and ethical review board oversight, ensuring compliance with data privacy standards.

Compensation details are that the participants involved in the study were compensated fairly for their contributions. All individuals provided informed consent, detailing the nature and amount of compensation received, proportional to the time and effort required for their participation, in accordance with ethical guidelines.

XSL•FO
RenderX

Results

Patient Characteristics

Initially, we obtained frontal facial photographs of 989 subjects (including 300 NCs, 437 patients with iGCTs, and 252 patients with other midline brain tumors). These photos were collected for the training and testing of the DL model with facial recognition, GVisageNet, from January 2010 to December 2020 at 2 medical institutes. In addition, an independent validation dataset, comprising 200 NCs, 130 patients with iGCTs, and 106 patients with other midline brain tumors, was collected from four medical institutes between January 2021 and May 2023 for validation of the DL model. The demographic data, symptoms, and endocrine hormone levels of all patients across all datasets (training, testing, and independent validation) are summarized in Table 1. The training, testing and independent validation datasets included 500 NCs (250 males and 250 females, with mean age of 13.00, SD 5.15 years), 567 patients with iGCTs (256 males and 290 females, with a mean age of 12.81, SD 5.79 years) and 358 patients with other midline brain tumors 155 males and 203 females, with a mean age of 13.61, SD 4.93 years), respectively. We filtered 7 clinical variables (limb movement disorder, diabetes insipidus, sexual abnormality, affective disturbance, convulsive seizure, TSH, and GH) that were statistically different between the iGCTs group and other midline brain tumors group in the training/testing dataset. Figures 3A and 3B illustrate the distribution of iGCT tumor sites across various hospitals in different datasets, as well as the distribution of tumor types within all datasets among other midline brain tumor groups.

Figure 3. (A) Tumor site distribution of iGCTs in different hospitals of different datasets. (B) Distribution of tumor types in all datasets in the other midline brain tumors group. (C) Comparison of performance of different algorithms (accuracy, sensitivity, specificity, and area under the curve). AUC: area under the curve; iGCTs: intracranial germ cell tumors.

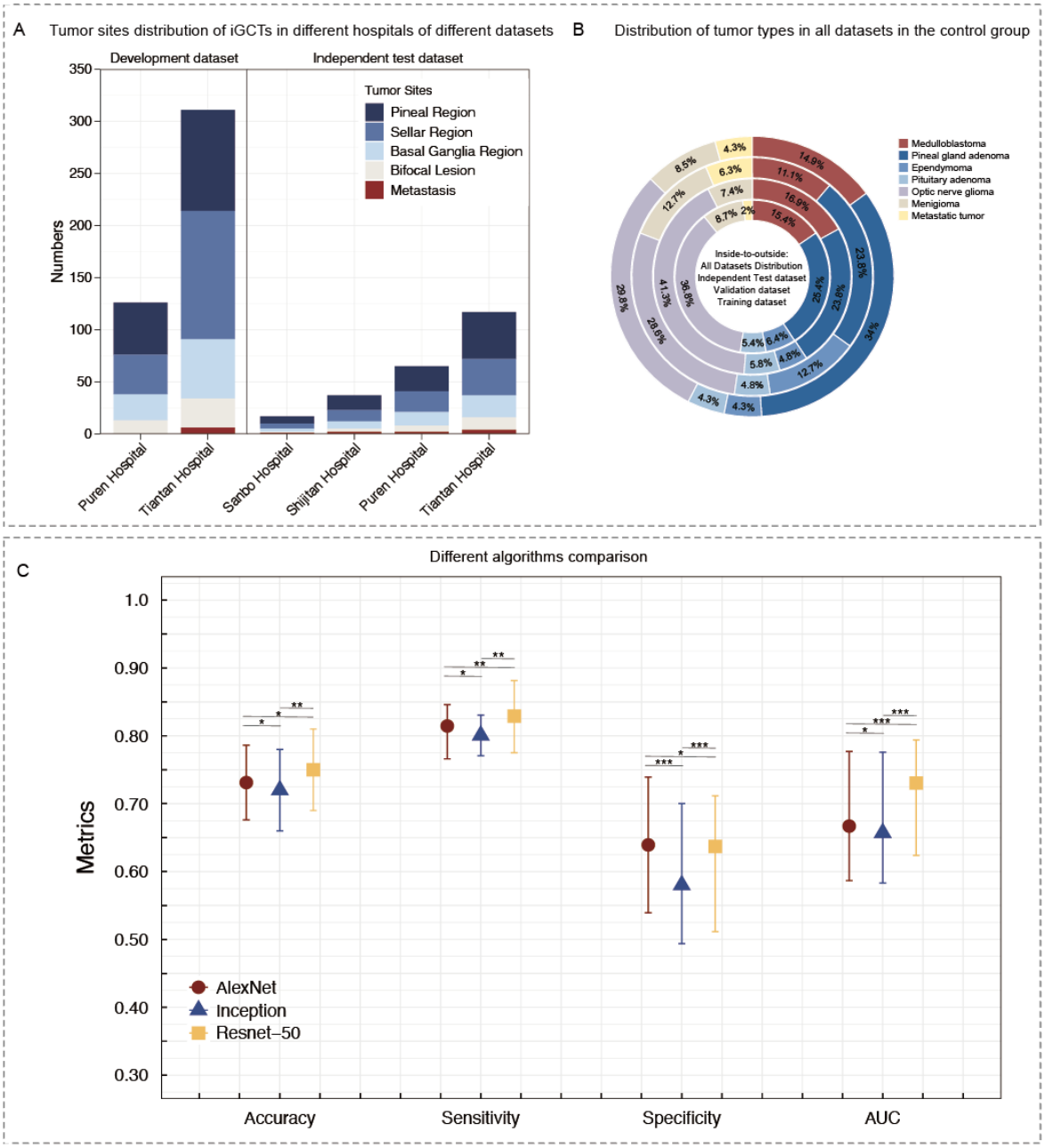

XSL•FO
RenderX

Evaluation of Three Candidate Algorithmic Models

In this investigation, we conducted a comparative analysis of 3 distinct candidate algorithms by training each on the training dataset and then evaluating their respective performances on the testing dataset. The algorithms tested were AlexNet, ResNet-50, and Inception. In the testing dataset, GVisageNet and the other models (AlexNet and Inception) yielded similar results in the first stage (distinguishing patients with midline brain tumors and NCs) with AUCs of 0.972 (95% CI 0.951-0.992), 0.963 (95% CI 0.943-0.984), and 0.983 (95% CI 0.937-1.000), respectively, and $P>.05$. Therefore, we proceeded to compare the results from the second stage. In the second stage, based on the AUC values with 95% CIs, the following results were observed: AlexNet achieved an AUC of 0.667 (95% CI 0.582-0.777), ResNet-50 reached an AUC of 0.705 (95% CI 0.590-0.760), and Inception yielded an AUC of 0.657 (95% CI: 0.583-0.776; Figure 3C and Table 2). Given these outcomes, ResNet-50 emerged as the most proficient model in terms of classification performance and was subsequently chosen as the final model for our classification tasks. Table S1 in Multimedia Appendix 1 and Figure S2 in Multimedia Appendix 1 presented additional details.

Table 2. The AUC, accuracy, sensitivity, and specificity of the different model of testing and independent external validation datasets.

Model	Dataset	AUC[a] (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
GVisageNet	Testing	0.705 (0.590-0.760)	71.71% (57.42%-77.11%)	75.24% (70.89%-77.57%)	73.40% (69.04%-76.80%)
GVisageNet	Independent validation	0.739 (0.728-0.773)	73.10% (69.85%-77.34%)	68.33% (53.65%-75.14%)	63.24% (56.59%-81.20%)
Logistic regression	Testing	0.612 (0.493-0.709)	59.41% (50.85%-64.28%)	65.18% (56.78%-73.54%)	55.93% (48.48%-65.74%)
Logistic regression	Independent validation	0.632 (0.571-0.679)	68.97% (62.68%-75.44%)	57.88% (49.45%-68.74%)	58.59% (49.24%-70.04%)
Hybrid model	Testing	0.711 (0.676-0.798)	70.77% (65.48%-76.65%)	72.94% (66.99%-76.84%)	67.78% (66.31%-72.53%)
Hybrid model	Independent validation	0.789 (0.728-0.793)	71.56% (68.90%-73.84%)	71.77% (56.48%-76.88%)	58.45% (50.73%-66.05%)

[a]AUC: area under the curve.

Performance of GVisageNet

In the testing dataset, GVisageNet achieved an AUC of 0.942 (95% CI 0.921-0.972), an accuracy of 94.33% (95% CI: 91.00%-95.96%), a sensitivity of 93.51% (95% CI: 90.00%-96.50%), and a specificity of 93.39% (95% CI 89.48%-97.00%) in distinguishing NCs from midline brain tumors. In second stage, GVisageNet achieved an AUC of 0.705 (95% CI 0.590-0.760), an accuracy of 71.71% (95% CI 57.42%-77.11%), a sensitivity of 75.24% (95% CI 70.89%-77.57%), and a specificity of 73.40% (95% CI 69.04%-76.80%) in distinguishing iGCTs from other midline brain tumors.

In the independent external validation dataset, GVisageNet demonstrated effectiveness in distinguishing NCs and midline brain tumors. It achieved an AUC of 0.938 (95% CI 0.913-0.967), an accuracy of 83.77% (95% CI: 71.45%-92.33%), a sensitivity of 91.67% (95% CI 87.42%-100.00%), and a specificity of 77.55% (95% CI 66.43%-80.46%; Figure 4A). In distinguishing iGCTs from other midline brain tumors, as shown in Figure 4B, it achieved an AUC of 0.739 (95% CI 0.728-0.773), an accuracy of 73.10% (95% CI 69.85%-77.34%), a sensitivity of 68.33% (95% CI 53.65%-75.14%), and a specificity of 63.24% (95% CI 56.59%-81.20%). The AUC, accuracy, sensitivity, and specificity of GVisageNet in the testing and validation datasets of the second stage are depicted in Figures 4C and 4D.

XSL•FO
RenderX

Figure 4. (A) The receiver operator characteristic curve of GVisageNet in the initial stage of development in the testing and independent external validation datasets. (B) The receiver operator characteristic curve of GVisageNet in the second stage of development in the testing and independent external validation datasets. (C) The area under the curve, accuracy, sensitivity, and specificity of the GVisageNet in the testing dataset with 95 % CI. (D) The AUC, accuracy, sensitivity, and specificity of the GVisageNet in the independent external validation dataset with 95% CI. AUC: area under the curve.

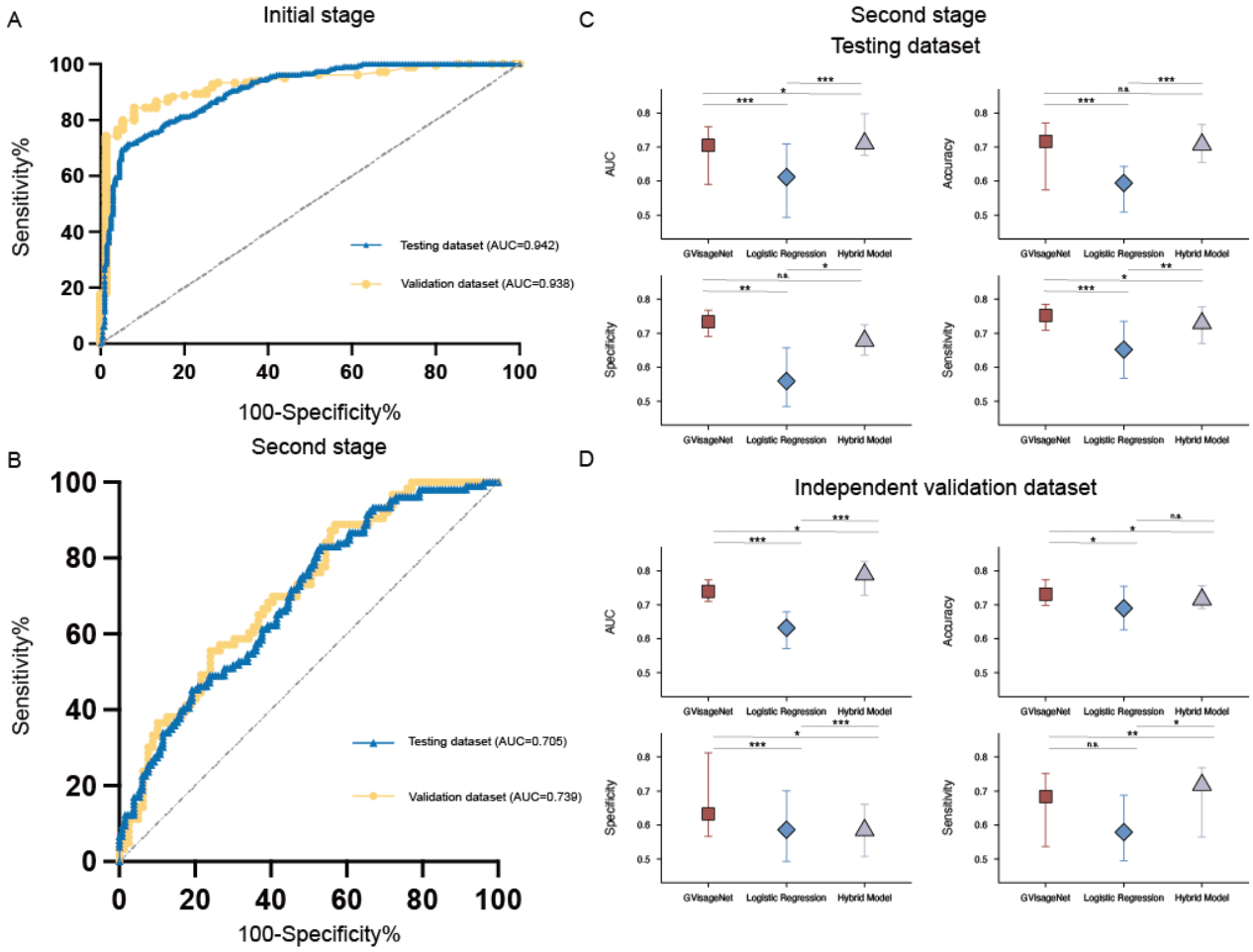

Clinical Variables and Hybrid Model Performance in Differentiating iGCTs From Other Midline Brain Tumors

To enhance GvisageNet's ability to distinguish iGCTs from other midline brain tumors, we compared the performance of a logistic regression model using clinical variables with a hybrid model. In the testing dataset, the logistic regression model, which used clinical variables, achieved an AUC of 0.612 (95% CI 0.493-0.709), an accuracy of 59.41% (95% CI 50.85%-64.28%), a sensitivity of 65.18% (95% CI 56.78%-73.54%), and a specificity of 55.93% (95% CI 48.48%-65.74%). In the independent external validation dataset, this model attained an AUC of 0.632 (95% CI 0.571-0.679), an accuracy of 68.97% (95% CI 62.68%-75.44%), a sensitivity of 57.88% (95% CI 49.45%-68.74%), and a specificity of 58.59% (95% CI: 49.24%-70.04%), as depicted in Figures 4C and 4D, and Table S2 in Multimedia Appendix 1.

The hybrid model in the testing dataset achieved an AUC of 0.711 (95% CI 0.676-0.798), an accuracy of 70.77% (95% CI 65.48%-76.65%), a sensitivity of 72.94% (95% CI 66.99%-76.84%), and a specificity of 67.78% (95% CI 66.31%-72.53%). In the independent external validation dataset, the hybrid prediction model reached an AUC of 0.789 (95% CI 0.728-0.793), an accuracy of 71.56% (95% CI 68.90%-73.84%), a sensitivity of 71.77% (95% CI 56.48%-76.88%), and a specificity of 58.45% (95% CI 50.73%-66.05%), as shown in Figures 4C and 4D.

In a comparative analysis conducted on the independent external validation dataset, GVisageNet was found to significantly outperform the logistic regression model, demonstrating higher accuracy (AUC 0.739 vs 0.632; P<.001). However, its performance was slightly inferior to that of the hybrid model (AUC 0.739 vs 0.789; P=.04).

Correlation Analysis Between Prediction Output and Endocrine Indicators

Correlation analysis revealed a significant positive correlation between the prediction scores and the levels of TSH and IGFBP-3 (P<.05). Conversely, there was a significant negative correlation between the prediction scores and the levels of total thyroxine (TT4), FT3, FT4, GH, prolactin, and estradiol (P<.05). Detailed pairwise concordances of endocrine indicators are presented in Figure 5A.

XSL•FO
RenderX

Figure 5. (A) Correlation analysis between the output results of GVisageNet and endocrine indicators in the independent external validation dataset. (B) The area under the curve, accuracy, sensitivity, and specificity of the GVisageNet for the stratified analysis of the HPT axis damage degree in the independent external validation dataset. (C) The visual integration of GVisageNet into the facial photo collection system (independent validation dataset). ACTH: adrenocorticotrophic hormone; E2: estradiol; FT3: free triiodothyronine; FT4: free thyroxine; GH: growth hormone; iGCTs: intracranial germ cell tumors; IGF-1: insulin-like growth factor 1; IGFBP-3: insulin-like growth factor binding protein 3; LH: luteinizing hormone; P: progesterone; PRL: prolactin; T: testosterone; TSH: thyroid-stimulating hormone; TT3: total tri-iodothyronine; TT4: total thyroxine.

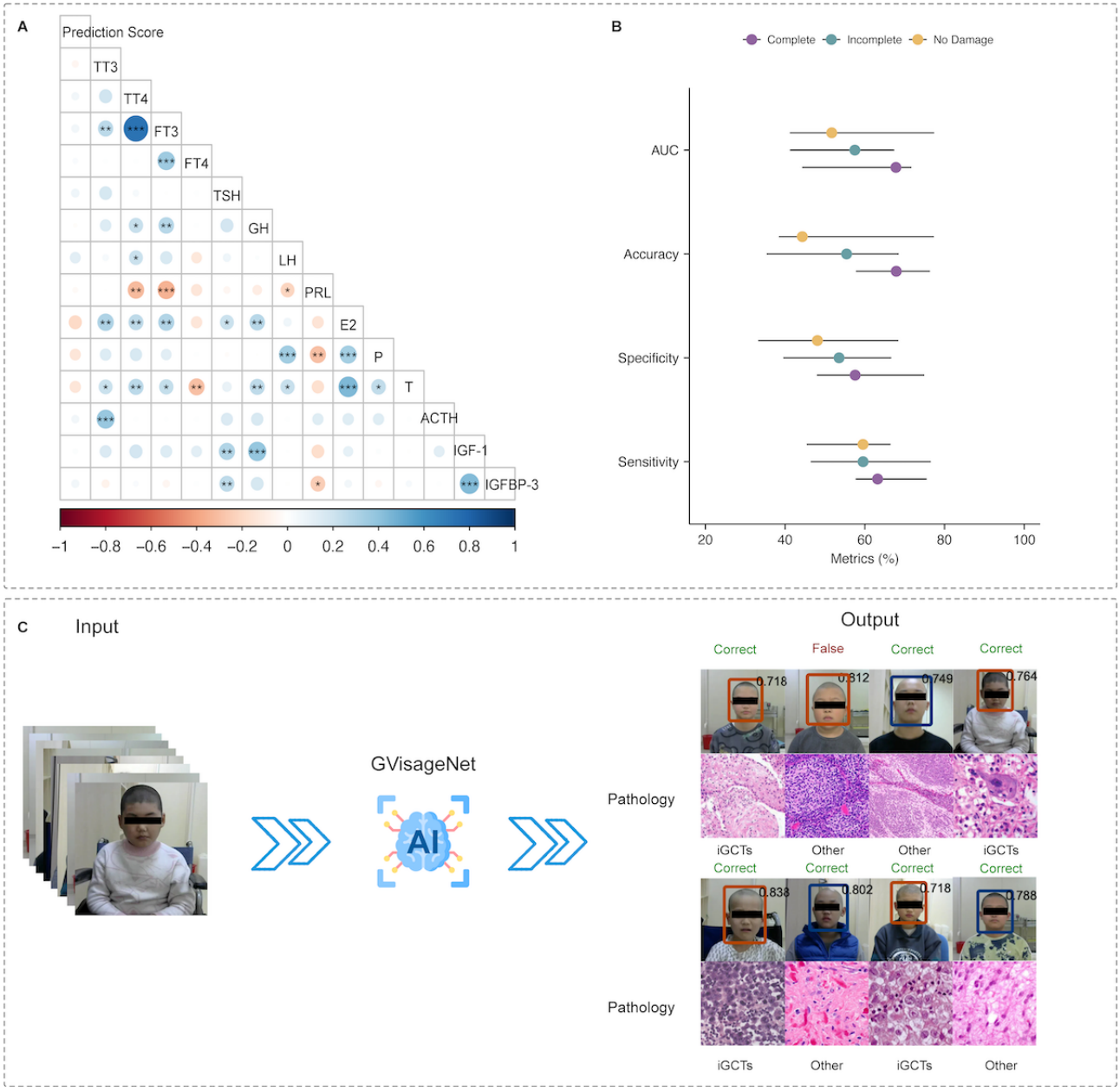

Moreover, the independent validation dataset was divided into 3 groups based on the degree of HPT axis damage: complete damage (n=37), incomplete damage (n=54), and no damage (n=39). In the task of differentiating iGCTs from other midline brain tumors, GVisageNet achieved varying levels of performance across these groups. In the complete damage group, GVisageNet achieved an AUC of 0.678 (95% CI 0.443-0.717), an accuracy of 67.89% (95% CI 57.79%-76.33%), a sensitivity of 63.21% (95% CI 57.67%-75.54%), and a specificity of 57.58% (95% CI 47.99%-74.91%). For the incomplete damage group, the model reached an AUC of 0.575 (95% CI 0.412-0.674), an accuracy of 55.43% (95% CI 35.37%-68.54%), a sensitivity of 59.56% (95% CI 46.43%-76.54%), and a specificity of 53.54% (95% CI 39.54%-66.64%). In the no-damage group, GVisageNet achieved an AUC of 0.517 (95% CI 0.412-0.774), an accuracy of 44.32% (95% CI 38.43%-77.33%), a sensitivity of 59.54% (95% CI 45.43%-66.47%), and a specificity of 48.11% (95% CI 33.23%-68.42%; Table 3). The AUC, accuracy, sensitivity, and specificity for each group are detailed in Figure 5B. An analysis of variance (Friedman test) revealed a significant difference in AUC among the 3 groups (P=.03). The visual integration of GVisageNet into the facial photo collection system is shown in Figure 5C.

XSL•FO
RenderX

Table 3. The AUC, accuracy, sensitivity, and specificity of the GVisageNet in the independent external validation dataset for the stratified analysis of the HPT axis damage degree.

Dataset	AUC[a] (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Complete damage	68.00% (44.30%-71.70%)	67.89% (57.79%-76.33%)	63.21% (57.67%-75.54%)	57.58% (47.99%-74.91%)
Incomplete damage	57.00% (41.20%-67.40%)	55.43% (35.37%-68.54%)	59.56% (46.43%-76.54%)	53.54% (39.54%-66.64%)
No-damage	52.00% (41.20%-77.40%)	44.32% (38.43%-77.33%)	59.54% (45.43%-66.47%)	48.11% (33.23%-68.42%)

[a]AUC: area under the curve.

Discussion

Principal Findings

This study successfully developed and validated an automated binary DL model, GVisageNet, to differentiate midline brain tumors from NCs and further distinguish iGCTs from other midline brain tumors using facial images. We also constructed a logistic regression model and a hybrid model using clinical data to explore whether clinical variables could enhance the diagnostic performance of the model. The results showed that in differentiating iGCTs from other midline brain tumors, the hybrid model slightly outperformed GVisageNet (0.739 vs 0.789; P=.04). Finally, we attempted to explain the mechanism of GVisageNet by analyzing endocrine changes in patients and found that the model's predicted probability values were strongly correlated with seven endocrine indicators (P<.05). In addition, GVisageNet performed best in the group with complete endocrine axis damage, suggesting that endocrine axis impairment may be a reason for the facial texture changes observed in patients. GVisageNet can be integrated into outpatient facial photo collection systems, enabling the routine collection of facial images during a patient's initial visit. This allows for real-time preliminary screening diagnoses, guiding clinicians to conduct targeted examinations, thereby saving medical resources and reducing patient costs.

The enhanced ResNet-50 model demonstrated significant advancements in this field, outperforming both its original version and other conventional CNN models in terms of training and testing accuracy. This improvement was achieved through structural modifications and an optimized training process, including an adaptive learning rate that effectively adjusts weights within the network layers. Moreover, the model effectively addressed the common issue of overfitting in DL, thereby enhancing its reliability and consistency in screening applications. A crucial aspect of this improvement was the model's reduced loss values and fluctuations, contributing to its increased stability. Importantly, visualization tools were used for feature extraction from specific layers of ResNet-50 [19], enabling precise identification of potential tumor indicators from facial photographs [20], which is essential for early screening and intervention. In addition, the introduction of a standard operating procedure for preprocessing facial photos significantly improved the quality of input data, aiding in more accurate tumor screening.

In addition to the analyses conducted, it is essential to consider the potential issue of collinearity among features. Collinearity, or high intercorrelation among predictor variables, can adversely affect the performance and interpretability of regression models and other machine learning techniques. Detecting and addressing collinearity is crucial to ensure the robustness of our models. Methods such as correlation matrices, variance inflation factors, and condition indices can be used to detect collinearity [21]. Mitigation strategies include removing or combining highly correlated features and using regularization techniques like Ridge regression or Lasso [22]. Future work should include a thorough investigation of collinearity and its impact on model performance. By addressing collinearity, we can enhance the stability and interpretability of our models, leading to more reliable diagnostic predictions. Therefore, it is recommended that future studies incorporate collinearity diagnostics and report any steps taken to mitigate its effects. This approach will provide a more comprehensive understanding of the model development process and ensure the robustness of the predictive models.

Furthermore, DL models, especially those designed for facial recognition, are adept at extracting and analyzing complex image patterns that may reflect physiological changes due to endocrine imbalances. Several mechanisms underlie the potential correlations between DL model outputs and endocrine indicators. First, models like CNNs can identify subtle changes in facial features associated with hormonal imbalances, such as those caused by conditions like acromegaly or Cushing's syndrome, which alter fat distribution, skin texture, or facial bone structure. Second, these models use extensive training datasets to recognize patterns associated with specific conditions, such as the protruding eyes or facial puffiness seen in hyperthyroidism, and can correlate these visual cues with hormone levels (eg, TSH, triiodothyronine, and thyroxine) to predict endocrine states from the facial analysis. In addition, hormones significantly affect physical characteristics, where elevated cortisol levels may cause a "moon face" appearance or a buffalo hump, and excess growth hormone can change the jawline and brow prominence, all detectable through DL analysis. Finally, integrating DL outputs with clinical data, such as hormone levels, enhances predictive accuracy. A hybrid model that merges image-based phenotypic features with biochemical markers provides a comprehensive understanding of a patient's endocrine state, enabling nuanced predictions and personalized health assessments. This integrated approach leverages facial recognition technology to noninvasively assess endocrine conditions, thereby enhancing diagnostic precision and patient care.

In addition, exploring the potential broader applications of our DL model in diagnosing other diseases with distinct phenotypic manifestations is imperative. For instance, conditions such as Parkinson disease, characterized by specific facial expressions and muscle rigidity, or genetic disorders like Down syndrome,

XSL•FO
RenderX

which presents with unique facial features, could benefit from analogous DL-based diagnostic approaches [23,24]. By training models to recognize these specific phenotypic patterns, we can expand the use of our DL model to provide early and accurate diagnoses across a variety of medical conditions. This broader application not only enhances the model's relevance in the medical field but also contributes to more comprehensive patient care through the early detection of diseases that manifest through observable physical changes.

Significant correlations were observed between DL model outputs and endocrine markers, including TSH, IGFBP-3, TT4, FT3, FT4, GH, prolactin, and estradiol, suggesting that endocrine variations may play a crucial role in the development of facial features in patents with iGCT. Notably, approximately 40% of iGCTs secrete β-HCG [1,5,25], which also induces endocrine changes at sufficiently high levels. The cross-immunoreactivity of β-HCG with LH triggers peripheral precocious puberty or gonadotropin-releasing hormone–independent precocious puberty, distinct from central precocious puberty and nonprogressive pubertal variants. In males, elevated β-HCG levels activate LH receptors on Leydig cells, enhancing testosterone production [26]. Conversely, in females, β-HCG does not induce precocious puberty, as activation of both FSH and LH receptors is requisite. Consequently, precocious puberty in iGCTs is predominantly observed in male toddlers and school-aged boys, characterized by facial features such as acne, beard growth, increased oil secretion, and skeletal maturity. Excess GH production leads to acromegaly [27], resulting in enlarged facial bones, while hypothyroidism causes facial puffiness [28-32], especially around the eyes [33-35]. When DL model results were integrated with clinical indicators, the detection efficacy was significantly enhanced. It is imperative to clarify that β-HCG was excluded from our study since only 40% of patients exhibit β-HCG secretion, and the DL model serves as a prediagnostic screening tool, hinting at further examination. Premature inclusion of tumor markers could compromise model accuracy. This observation was further underscored in the stratified analysis. Our findings indicate that the predictive efficacy of GVisageNet was least effective in cases without HPT axis impairment, and most effective in instances of complete HPT axis disruption. This underscores a direct correlation between the model's predictive capacity and endocrine function. Notably, when the HPT axis was fully compromised, the resultant endocrine-driven alterations in facial characteristics were most pronounced.

This study had several limitations that should be considered when interpreting the findings. First, the relatively small sample size of each iGCT subtype (germ cell tumors [germinomas] and nongerminomatous germ cell tumors) limited the ability to conduct a more detailed analysis and might have impacted the statistical power of the study. Future research should aim to include larger cohorts to enhance the robustness and generalizability of the findings. Second, the study only collected baseline facial photos of the patients. The lack of longitudinal follow-up data limits the ability to observe changes in facial features and endocrine parameters over time. Longitudinal studies are needed to evaluate the progression and potential predictive value of facial recognition technology in monitoring patients with iGCTs. Third, the study was conducted across multiple centers, which introduced variability in data collection and image quality. Although efforts were made to standardize procedures, some degree of inconsistency is inevitable. Future studies should implement stricter protocols and quality control measures to minimize such variability. Finally, while the study focused on differentiating iGCTs from other midline brain tumors, it did not explore the potential of this technology in distinguishing iGCTs from other types of intracranial tumors or nontumorous conditions. Expanding the scope of future research could provide a more comprehensive understanding of the clinical applications of deep learning-based facial recognition in neuro-oncology.

In addition, "no identifiable personal information" (such as unique facial characteristics) is presented in any figures or supplementary materials of the manuscript. In cases where participant consent is necessary for image inclusion, signed consent forms have been obtained and can be uploaded in the "Other files" field upon resubmission. This ensures transparency and adheres to the highest standards of ethical research practices.

Conclusion

In summary, this study demonstrates the potential application of facial recognition technology in the diagnosis of iGCTs. Facial recognition could not only aid in early detection and screening of iGCTs but also might become a valuable tool in advancing future clinical decision-making.

Acknowledgments

We thank all the patients for their efforts.

The work was supported by the Dominant discipline - Excellent discipline award - Department of Radiotherapy (YXXKJ-2024-01 and YXXKJ-2025-01), the Natural Science Foundation of Beijing (L232079), National Key Research and Development Program of China (2022ZD0210100), National Natural Science Foundation of China (82273343, 82172608, 82101356, and 81902975), National Science Fund of Beijing for Distinguished Young Scholars (JQ24040), Beijing Nova Star Program (20220484058), and Capital Medical University Fund for Excellent Young Scholars (KCB2304).

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Additional methods, figures, and tables for the protocol.
[DOCX File , 579 KB - jmir_v27i1e58760_app1.docx]

References

1. Packer RJ, Cohen BH, Cooney K. Intracranial germ cell tumors. Oncologist 2000;5(4):312-320 [FREE Full text] [Medline: 10964999]
2. Takami H, Fukuoka K, Fukushima S, Nakamura T, Mukasa A, Saito N, et al. Integrated clinical, histopathological, and molecular data analysis of 190 central nervous system germ cell tumors from the iGCT Consortium. Neuro Oncol 2019;21(12):1565-1577 [FREE Full text] [doi: 10.1093/neuonc/noz139] [Medline: 31420671]
3. Bjornsson J, Scheithauer BW, Okazaki H, Leech RW. Intracranial germ cell tumors: pathobiological and immunohistochemical aspects of 70 cases. J Neuropathol Exp Neurol 1985;44(1):32-46. [Medline: 4038412]
4. Rickert CH, Paulus W. Epidemiology of central nervous system tumors in childhood and adolescence based on the new WHO classification. Childs Nerv Syst 2001;17(9):503-511. [doi: 10.1007/s003810100496] [Medline: 11585322]
5. Frappaz D, Dhall G, Murray MJ, Goldman S, Faure Conter C, Allen J, et al. EANO, SNO and Euracan consensus review on the current management and future development of intracranial germ cell tumors in adolescents and young adults. Neuro Oncol 2022;24(4):516-527 [FREE Full text] [doi: 10.1093/neuonc/noab252] [Medline: 34724065]
6. Lo AC, Hodgson D, Dang J, Tyldesley S, Bouffet E, Bartels U, et al. Intracranial germ cell tumors in adolescents and young adults: a 40-year multi-institutional review of outcomes. Int J Radiat Oncol Biol Phys 2020;106(2):269-278. [doi: 10.1016/j.ijrobp.2019.10.020] [Medline: 31654785]
7. Calaminus G, Kortmann R, Worch J, Nicholson JC, Alapetite C, Garrè ML, et al. SIOP CNS GCT 96: final report of outcome of a prospective, multinational nonrandomized trial for children and adults with intracranial germinoma, comparing craniospinal irradiation alone with chemotherapy followed by focal primary site irradiation for patients with localized disease. Neuro Oncol 2013;15(6):788-796 [FREE Full text] [doi: 10.1093/neuonc/not019] [Medline: 23460321]
8. Khatua S, Dhall G, O'Neil S, Jubran R, Villablanca JG, Marachelian A, et al. Treatment of primary CNS germinomatous germ cell tumors with chemotherapy prior to reduced dose whole ventricular and local boost irradiation. Pediatr Blood Cancer 2010;55(1):42-46. [doi: 10.1002/pbc.22468] [Medline: 20222020]
9. McCarthy BJ, Shibui S, Kayama T, Miyaoka E, Narita Y, Murakami M, et al. Primary CNS germ cell tumors in Japan and the United States: an analysis of 4 tumor registries. Neuro Oncol 2012;14(9):1194-1200 [FREE Full text] [doi: 10.1093/neuonc/nos155] [Medline: 22869621]
10. Gittleman H, Cioffi G, Vecchione-Koval T, Ostrom QT, Kruchko C, Osorio DS, et al. Descriptive epidemiology of germ cell tumors of the central nervous system diagnosed in the United States from 2006 to 2015. J Neurooncol 2019;143(2):251-260. [doi: 10.1007/s11060-019-03173-4] [Medline: 31025275]
11. Bottinelli F, Delvecchio G, Moltrasio C, Ferro A, Diwadkar VA, Brambilla P. Facial emotion recognition in panic disorder: a mini-review of behavioural studies. J Affect Disord 2021;282:173-178. [doi: 10.1016/j.jad.2020.12.064] [Medline: 33418364]
12. Yeung MK. A systematic review and meta-analysis of facial emotion recognition in autism spectrum disorder: the specificity of deficits and the role of task characteristics. Neurosci Biobehav Rev 2022;133:104518. [doi: 10.1016/j.neubiorev.2021.104518] [Medline: 34974069]
13. Nigam SG, Shenoy S, Sharma PSVN, Behere RV. Facial emotion recognition and its association with quality of life and socio-occupational functioning in patients with bipolar disorder and their first-degree relatives. Asian J Psychiatr 2021;65:102843. [doi: 10.1016/j.ajp.2021.102843] [Medline: 34547596]
14. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro Oncol 2021;23(8):1231-1251 [FREE Full text] [doi: 10.1093/neuonc/noab106] [Medline: 34185076]
15. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro Oncol 2021;23(8):1231-1251 [FREE Full text] [doi: 10.1093/neuonc/noab106] [Medline: 34185076]
16. Tian T, Wang L, Luo M, Sun Y, Liu X. ResNet-50 based technique for EEG image characterization due to varying environmental stimuli. Comput Methods Programs Biomed 2022;225:107092. [doi: 10.1016/j.cmpb.2022.107092] [Medline: 36058062]
17. Panthakkan A, Anzar SM, Jamal S, Mansoor W. Concatenated Xception-ResNet50 - A novel hybrid approach for accurate skin cancer prediction. Comput Biol Med 2022;150:106170. [doi: 10.1016/j.compbiomed.2022.106170] [Medline: 37859280]
18. Demler OV, Pencina MJ, D'Agostino RB. Misuse of DeLong test to compare AUCs for nested models. Stat Med 2012;31(23):2577-2587 [FREE Full text] [doi: 10.1002/sim.5328] [Medline: 22415937]
19. Nithya VP, Mohanasundaram N, Santhosh R. An early detection and classification of Alzheimer's disease framework based on ResNet-50. Curr Med Imaging 2024;20:e250823220361. [doi: 10.2174/1573405620666230825113344] [Medline: 37622561]

XSL•FO
RenderX

20. Liu ZS, Siu WC, Chan YL. Features guided face Super-Resolution via hybrid model of deep learning and random forests. IEEE Trans Image Process 2021;30:4157-4170. [doi: 10.1109/TIP.2021.3069554] [Medline: 33819156]
21. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 2012;40(7):e49 [FREE Full text] [doi: 10.1093/nar/gkr1293] [Medline: 22217600]
22. Dupré la Tour T, Eickenberg M, Nunez-Elizalde AO, Gallant JL. Feature-space selection with banded ridge regression. Neuroimage 2022;264:119728 [FREE Full text] [doi: 10.1016/j.neuroimage.2022.119728] [Medline: 36334814]
23. Srisraluang W, Rojnueangnit K. Facial recognition accuracy in photographs of thai neonates with down syndrome among physicians and the Face2Gene application. Am J Med Genet A 2021;185(12):3701-3705. [doi: 10.1002/ajmg.a.62432] [Medline: 34288412]
24. Assogna F, Pontieri FE, Caltagirone C, Spalletta G. The recognition of facial emotion expressions in Parkinson's disease. Eur Neuropsychopharmacol 2008;18(11):835-848. [doi: 10.1016/j.euroneuro.2008.07.004] [Medline: 18707851]
25. Kreutz J, Rausin L, Weerts E, Tebache M, Born J, Hoyoux C. Intracranial germ cell tumor. JBR-BTR 2010;93(4):196-197. [doi: 10.5334/jbr-btr.286] [Medline: 20957891]
26. Zhou R, Wu J, Liu B, Jiang Y, Chen W, Li J, et al. The roles and mechanisms of Leydig cells and myoid cells in regulating spermatogenesis. Cell Mol Life Sci 2019;76(14):2681-2695 [FREE Full text] [doi: 10.1007/s00018-019-03101-9] [Medline: 30980107]
27. Melmed S. Acromegaly pathogenesis and treatment. J Clin Invest 2009;119(11):3189-3202 [FREE Full text] [doi: 10.1172/JCI39375] [Medline: 19884662]
28. Kosilek RP, Frohner R, Würtz RP, Berr CM, Schopohl J, Reincke M, et al. Diagnostic use of facial image analysis software in endocrine and genetic disorders: review, current results and future perspectives. Eur J Endocrinol 2015 Oct;173(4):M39-M44. [doi: 10.1530/EJE-15-0429] [Medline: 26162404]
29. Zenker M, Edouard T, Blair JC, Cappa M. Noonan syndrome: improving recognition and diagnosis. Arch Dis Child 2022;107(12):1073-1078 [FREE Full text] [doi: 10.1136/archdischild-2021-322858] [Medline: 35246453]
30. Qiang J, Wu D, Du H, Zhu H, Chen S, Pan H. Review on Facial-Recognition-Based applications in disease diagnosis. Bioengineering (Basel) 2022;9(7):273 [FREE Full text] [doi: 10.3390/bioengineering9070273] [Medline: 35877324]
31. Lin CL, Wu KC. Development of revised ResNet-50 for diabetic retinopathy detection. BMC Bioinformatics 2023;24(1):157 [FREE Full text] [doi: 10.1186/s12859-023-05293-1] [Medline: 37076790]
32. Kim ES, Heo JM, Eun SJ, Lee JY. Development of Early-Stage stroke diagnosis system for the elderly neurogenic bladder prevention. Int Neurourol J 2022;26(Suppl 1):S76-S82 [FREE Full text] [doi: 10.5213/inj.2244030.015] [Medline: 35236050]
33. Rose SR, Wassner AJ, Wintergerst KA, Yayah-Jones N, Hopkin RJ, Chuang J, SECTION ON ENDOCRINOLOGY EXECUTIVE COMMITTEE, COUNCIL ON GENETICS EXECUTIVE COMMITTEE. Congenital hypothyroidism: screening and management. Pediatrics 2023;151(1):e2022060419. [doi: 10.1542/peds.2022-060419] [Medline: 36827523]
34. Bhattacharyya A, Chatterjee S, Sen S, Sinitca A, Kaplun D, Sarkar R. A deep learning model for classifying human facial expressions from infrared thermal images. Sci Rep 2021;11(1):20696 [FREE Full text] [doi: 10.1038/s41598-021-99998-z] [Medline: 34667253]
35. Goceri E. Deep learning based classification of facial dermatological disorders. Comput Biol Med 2021;128:104118. [doi: 10.1016/j.compbiomed.2020.104118] [Medline: 33221639]

Abbreviations

AUC: area under the curve
CNN: convolutional neural network
DL: deep learning
FSH: follicle-stimulating hormone
FT3: free triiodothyronine
FT4: free thyroxine
GH: growth hormone
HCG: human chorionic gonadotropin
HPT: hypothalamic-pituitary-target
iGCT: intracranial germ cell tumor
IGFBP-3: insulin-like growth factor binding protein 3
LH: luteinizing hormone
NC: normal control
PD: panic disorder
ROC: receiver operating characteristic
TSH: thyroid stimulating hormone
TT4: total thyroxine
WHO: World Health Organization

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 24.03.24; peer-reviewed by J Mistry, V Nagesh; comments to author 05.04.24; revised version received 26.05.24; accepted 25.11.24; published 30.01.25.

Please cite as:
Li Y, He Y, Liu Y, Wang B, Li B, Qiu X
Identification of Intracranial Germ Cell Tumors Based on Facial Photos: Exploratory Study on the Use of Deep Learning for Software Development
J Med Internet Res 2025;27:e58760
URL: https://www.jmir.org/2025/1/e58760
doi:10.2196/58760
PMID:

©Yanong Li, Yixuan He, Yawei Liu, Bingchen Wang, Bo Li, Xiaoguang Qiu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Electronic Health Record Data Collection Practices to Advance Standardization and Interoperability of Patient Preferences for Interpretation Services: Qualitative Study

Krysta Heaney-Huls[1*], MPH; Rida Shams[1*], MSPH; Ruth Nwefo[1*], MPH; Rachel Kane[2*], BSPH; Janna Gordon[3*], PhD; Alison M Laffan[1*], PhD; Scott Stare[4*], MBA; Prashila Dullabh[1*], MD

[1]NORC at the University of Chicago, Chicago, IL, United States

[2]Yale School of Public Health, New Haven, CT, United States

[3]Riverbend Health Psychology, PLLC, Salt Lake City, UT, United States

[4]Office of Minority Health, Centers for Medicare & Medicaid Services, Baltimore, MD, United States

[*]all authors contributed equally

Corresponding Author:
Krysta Heaney-Huls, MPH
NORC at the University of Chicago
55 E Monroe St 30th Floor
Chicago, IL, 60603
United States
Phone: 1 7734017110
Email: heaney-huls-krysta@norc.org

Abstract

Background: Poor health outcomes are well documented among patients with a non-English language preference (NELP). The use of interpreters can improve the quality of care for patients with NELP. Despite a growing and unmet need for interpretation services in the US health care system, rates of interpreter use in the care setting are consistently low. Standardized collection and exchange of patient interpretation needs can improve access to appropriate language assistance services.

Objective: This study aims to examine current practices for collecting, documenting, and exchanging information on a patient's self-reported preference for an interpreter in the electronic health record (EHR) and the implementation maturity and adoption level of available data standards. The paper identifies standards implementation; data collection workflows; use cases for collecting, documenting, and exchanging information on a patient's self-reported preference for an interpreter; challenges to data collection and use; and opportunities to advance standardization of the interpreter needed data element to facilitate patient-centered care.

Methods: We conducted a narrative review to describe the availability of terminology standards to facilitate health care organization documentation of a patient's self-reported preference for an interpreter in the EHR. Key informant discussions with EHR developers, health systems, clinicians, a practice-based research organization, a national standards collaborative, a professional health care association, and Federal agency representatives filled in gaps from the narrative review.

Results: The findings indicate that health care organizations value standardized collection and exchange of patient language assistance service needs and preferences. Informants identified three use cases for collecting, documenting, and exchanging information on a patient's self-reported preference for an interpreter, which are (1) person-centered care, (2) transitions of care, and (3) health care administration. The discussions revealed that EHR developers provide a data field for documenting interpreter needed data, which are routinely collected across health care organizations through commonly used data collection workflows. However, this data element is not mapped to standard terminologies, such as Logical Observation Identifiers Names and Codes (LOINC) or Systematized Medical Nomenclature for Medicine–Clinical Terminology (SNOMED-CT), consequently limiting the opportunities to electronically share these data between health systems and community-based organizations. The narrative review and key informant discussions identified three potential challenges to using information on a patient's self-reported preference for an interpreter for person-centered care and quality improvement, which are (1) lack of adoption of available data standards, (2) limited electronic exchange, and (3) patient mistrust.

Conclusions: Collecting and documenting patient's self-reported interpreter preferences can improve the quality of services provided, patient care experiences, and equitable health care delivery without invoking a significant burden on the health care

XSL•FO
RenderX

system. Although there is routine collection and documentation of patient interpretation needs, the lack of standardization limits the exchange of this information among health care and community-based organizations.

(J Med Internet Res 2025;27:e62670) doi:10.2196/62670

KEYWORDS

health information exchange; interoperability; electronic health records; interpreter; limited English proficiency; communication barriers

Introduction

Background

Health disparities are a prevalent issue in health care delivery in the United States today. Among the contributors to care inequities are barriers to language assistance services. Barriers to language assistance services are associated with misdiagnosis and inappropriate treatment by providers [1] and contribute to poor health outcomes that are well documented among patients with a non-English language preference (NELP) [2]. Compared with English-speaking patients, patients with NELP experience worse health outcomes, such as undiagnosed or uncontrolled hypertension, prolonged hospital stays, and poor asthma control [3]. Among Hispanic communities, which represent the fastest growing non-English language population in the United States, stroke is a leading cause of death and has been attributed to factors such as barriers to language assistance services [4]. Studies have also found that patients who experience barriers to language assistance services have a reduced likelihood of physician and mental health provider visits and are more likely to have an unplanned emergency department visit compared with patients who are proficient in English. Furthermore, patients with NELP are at an increased risk for unplanned repeat emergency department visits within 72 hours of discharge [5,6].

Interpretation services represent both a growing and unmet need in the United States health care system [7]. The National Council on Interpreting in Health Care defines a medical interpreter as an individual who “interprets in health care settings of any sort, including doctor’s offices, clinics, hospitals, home health visits, mental health clinics, and public health presentations” [8]. In 2021, the Migration Policy Institute estimated that approximately 26 million individuals in the United States reported a NELP [9,10].

The use of interpreters has been found to improve the quality of care for patients with NELP, with patients reporting overall high satisfaction with care and communication with their care team [11]. However, despite the benefits, rates of interpreter use are consistently low, and additional research is needed to better understand how patients assess the need for and use an interpreter [12]. The findings from Schwei et al [12] inform our current understanding of a patient’s decision-making processes (Figure 1) and could be beneficial to inform how health care organizations engage with patients to gather and use data on a patient’s self-reported preference for an interpreter.

Figure 1. Factors that patients say influence whether an interpreter is needed or used (adapted from Schwei et al [12]).

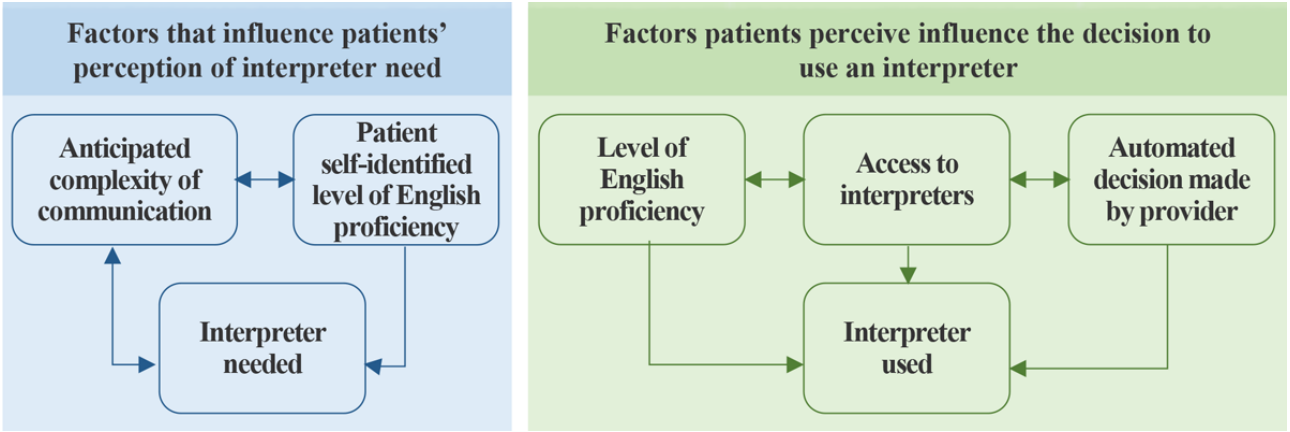

In the absence of medical interpreters, patients with NELP are left to rely on ad hoc interpreters such as family members, hospital staff members, or their own language fluency to obtain needed medical information. This leaves room for a variety of common interpretation errors to occur such as omissions, embellishments, false fluency, paraphrasing, and family members or ad hoc interpreters providing their own opinions [13]. Interpreter use can also be impacted by concerns from health providers or patients regarding the privacy and confidentiality of the communications between the patient and interpreter [5,14,15]. Patients may be reluctant to use an interpreter if they are uncertain of the privacy and confidentiality protections. These issues highlight the importance of both identifying patient language access needs and providing linguistically appropriate interpretation and translation services.

The standardized collection and exchange of patient interpretation needs can help improve access to appropriate language assistance services. Once collected from patients, this data can be stored in the electronic health record (EHR) and accessed by the patient’s care team to support person-centered care or exchanged electronically with other providers (eg, specialists and long-term care facilities) to support transitions of care.

XSL•FO
RenderX

Objectives

This paper examines current practices for collecting, documenting, and exchanging information on a patient's self-reported preference for an interpreter in the EHR. The paper identifies data standards implementation, data collection workflows, use cases for information on a patient's self-reported preference for an interpreter, challenges to data collection and use, and opportunities to advance the standardized collection and use of the interpreter needed data element to facilitate patient-centered care.

Methods

We used 2 methods to assess the collection, documentation, and use of information on a patient's self-reported preference for an interpreter, which are (1) a narrative review and (2) key informant discussions.

Literature Scan

We conducted a narrative review [16] to describe the availability of terminology standards to facilitate health care organization documentation of a patient's self-reported preference for an interpreter in the EHR, and the implementation maturity and adoption level of available data standards.

We searched both the peer-reviewed and gray literature using PubMed and Google (Textbox 1). We also searched the websites of standards development organizations including the Health Level Seven (HL7) website and Confluence pages, the Assistant Secretary for Technology Policy/Office of the National Coordinator for Health Information Technology (ASTP/ONC) website, the Logical Observation Identifiers Names and Codes (LOINC) website, the Gravity Project website, PCORnet (the National Patient Centered Clinical Research Network), and the Centers for Medicare & Medicaid Services (CMS) website. We conducted supplemental searches on barriers to language assistance services and available standards (eg, Systematized Medical Nomenclature for Medicine-Clinical Terminology [SNOMED-CT] and Clinical Document Architecture) based on discussions with key informants. In total, we included a total of 31 sources.

Textbox 1. Example search terms.

PubMed search terms

- "interpreter needed" AND "data element", "interpreter" AND "data element", "interpreter" AND "data" AND "EHR", language translator needed data element, interpretation, communication barriers, limited English proficiency, and "translating*"

Gray literature keywords

- "interpreter*", "interpreter needed", "interpreter service", "interpreter required", "translating*", "language translator", "communication facilitator*", United States Core Data for Interoperability (USCDI), "data element", electronic health record (EHR), language translator needed data element, inclusion of language preference in EHR, inclusion of language preference in EHR, Centers for Medicaid & Medicaid Services (CMS) Data Element Library (DEL) Health Information Technology Workgroup, Health Level 7 (HL7) Workgroup, Office of the National Coordinator (ONC), Gravity Project, and PCORnet Common Data Model.

Key Informant Discussions

We conducted key informant discussions to complement findings from the literature scan and contextual information gaps by identifying emergent topics not yet reflected in the literature. We identified a purposive sample of informants across 3 stakeholder categories to gather diverse perspectives on the collection and use of data on a patient's self-reported preference for an interpreter in the EHR, which are (1) EHR developers (n=3), (2) health systems and clinicians (n=8), and (3) policy-making and advocacy organizations (n=5). The EHR developers included in the study represented over 70% of the market share by bed count in the inpatient setting [17] and nearly half of the ambulatory market share among clinicians that report using a 2015 edition certified EHR product [18]. Health system and clinician informants included representatives from academic medical centers, a Federally Qualified Health Center, a primary care clinic, an integrated health care system, and acute care hospitals. Finally, policy-making and advocacy organization informants were comprised of Federal agency staff, language service professionals, members of a national consensus–based standards collaborative, and a community health center association. The key informants provided perspectives on the experience of a wide variety of health care organizations' processes for collecting, documenting, and using information on a patient's self-reported preference for an interpreter.

We generated transcripts for each key informant discussion. We used qualitative content analysis to identify key themes using the existing United States Core Data for Interoperability (USCDI) ONC New Data Element and Class Submission System, focusing on new data element submission requirements and emergent themes from gaps in the literature. The USCDI ONC New Data Element and Class submission requirements include (1) similar or related data elements in USCDI, (2) why the new data element should be considered separately, (3) main use cases to support the adoption of the data element into the USCDI, (4) estimates of the breadth of applicability of the use cases for the new data element, (5) data element maturity (eg, existing vocabulary, terminology, or content standards), (6) availability of additional technical specifications, (7) level determination representing the use of the new data element, (8) level of electronic exchange with external organizations, and (9) challenges to implementation (eg, restrictions on standardization or use, privacy and security concerns, and implementation burden) [19].

XSL•FO
RenderX

Analytic Approach

The findings from the literature scan were summarized and grouped into two primary categories, which are (1) standards for certified EHR and health information technology systems and (2) health care quality improvement. We used a data abstraction matrix to identify themes from the key informant discussions. We abstracted data from the discussions into five domains: (1) current practices for collecting information on a patient's self-reported preference for an interpreter, (2) use cases for this information, (3) EHR product support for an interpreter needed data element, (4) interoperability standards, and (5) challenges to collecting, documenting, and using information on a patient's self-reported preference for an interpreter. The authors also identified common themes from across the 5 domains to identify key considerations and opportunities to improve data interoperability and expand upon the use of the interpreter needed data element.

Ethical Considerations

This study was reviewed by the NORC at the University of Chicago Institutional Review Board and did not meet the definition of research with human subjects set forth in Federal Regulations at 45 CFR 46.102.

Results

Overview

Health care organizations routinely collect information on a patient's self-reported preference for an interpreter and preferred language data to facilitate interpretation service delivery. Health care organizations recognize the limitation of using either preferred language or English proficiency alone to identify a patient's preference or need for an interpreter and, therefore, collect a patient's self-reported preference for interpretation services in tandem with preferred language to accurately reflect language service access needs.

Current Health Care Organization Data Collection Practices

We did not find evidence of health care organizations using available standard terminologies (eg, LOINC and SNOMED-CT) in the literature. Key informant interviews with health systems, clinicians, and EHR developers affirmed this finding, stating that health care organizations use structured fields to document a patient's self-reported preference for an interpreter in the EHR but do not map this data to available clinical terminologies.

Through the key informant discussions, we identified common practices for the collection, documentation, and use of information on a patient's self-reported preference for an interpreter in both the ambulatory and inpatient settings. We present a workflow diagram for the common data collection and documentation processes in Figure 2 and describe each step below.

XSL•FO
RenderX

Figure 2. Interpretation service requests and delivery workflow. EHR: electronic health record.

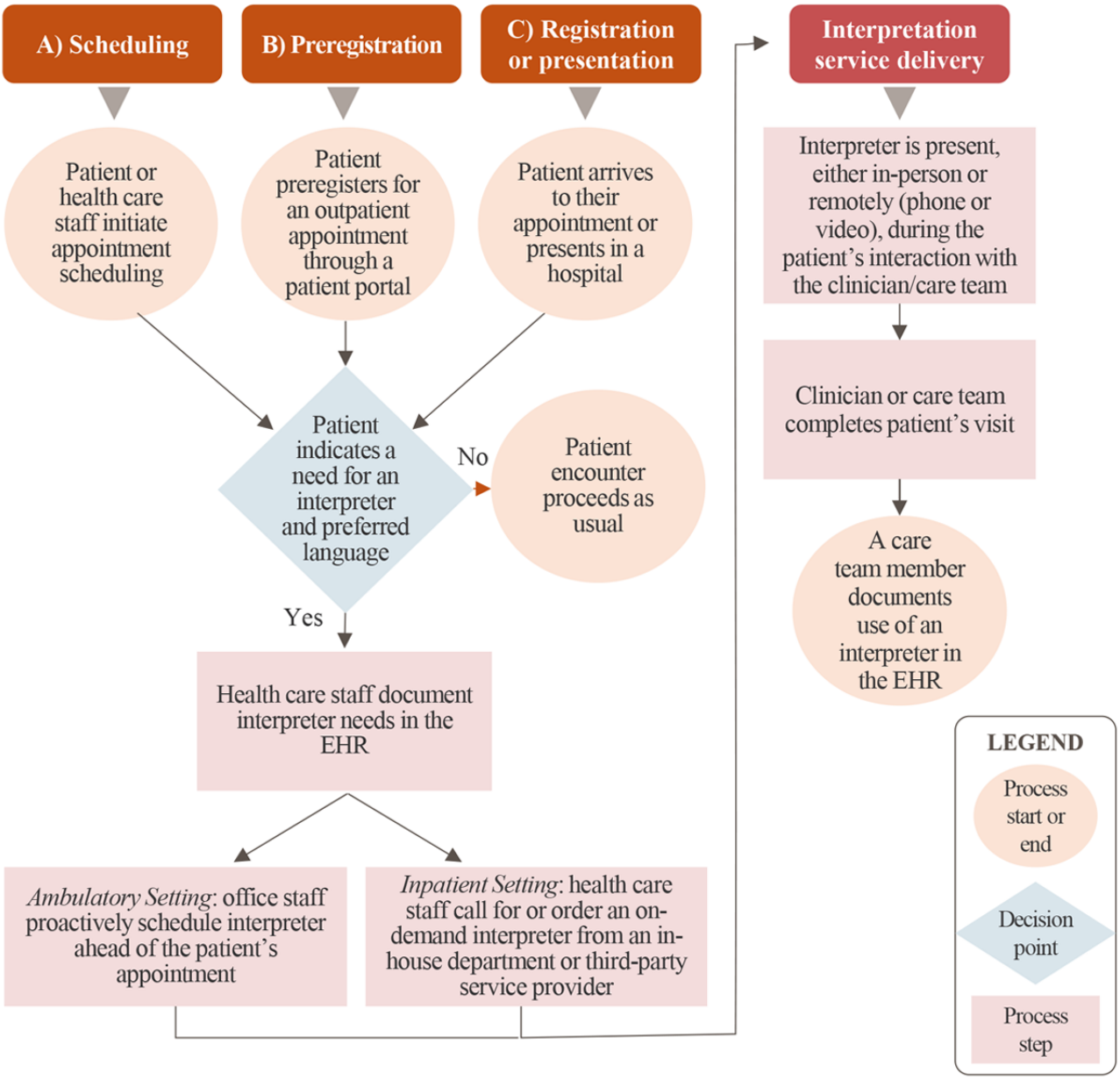

In Figure 2, three events (A-C) begin the process of gathering information from the patient on their self-reported preference for an interpreter and the subsequent provision of interpretation services. However, nuances exist depending on the EHR, organizational practices, care delivery setting, the availability of interpretation services, and interpretation service delivery method. Event A indicates scheduling a patient encounter. When a patient or health care staff initiates appointment scheduling, office staff members document the need for interpretation services in the EHR through an interpreter needed data field. Office staff members may proactively schedule interpretation services for the patient. Event B begins with patient preregistration. During preregistration, the patient may self-report their preference for an interpreter and their preferred language within the patient portal. Office staff members document this information in the patient's record and may proactively schedule interpretation services for the patient. Event C occurs with patient registration or when a patient presents in a hospital. When a patient self-reports their preference for an interpreter at arrival for an outpatient appointment or presents at a hospital, staff members document this information in the EHR and contact an interpreter for on-demand service.

Interpretation services can be delivered in-person or remotely through phone or video connection. Providers may use an interpreter from an internal interpretation services department, or a third-party vendor, or rely upon a clinician or staff member who is medically proficient in the patient's preferred language. Outpatient clinics often proactively schedule in-person or remote interpreters from third-party services because they may have collected a patient's self-reported preference for an interpreter before the patient encounter. If the clinic serves a significant number of patients with NELP, they may have in-house interpreters available or staff members who are medically proficient (ie, someone whose non-English language skills include communicating medical terminology and information) in the patient's preferred language. In the inpatient setting, interpretation services are provided either on-demand from an

XSL•FO
RenderX

in-house interpretation service department or remotely (eg, through a video connection). Upon completion of the patient visit, a member of the care team may document the use of the interpreter within the patient notes section of the EHR.

Use Cases for Data on a Patient's Self-Reported Preference for an Interpreter

Overview

Using the information on a patient's self-reported preference for an interpreter can assist health care organizations in providing more appropriate patient-centered care and eliminate guesswork or assumptions made by health care teams regarding which patients need an interpreter, ultimately improving the patient experience. Key informants described three primary use cases for the collection, documentation, and use of information on a patient's self-reported preference for an interpreter, which are (1) person-centered care, (2) transitions of care, and (3) health care administration (Table 1).

Table 1. Use cases and illustrative quotes for using data on a patient's self-reported preference for an interpreter.

Use case	Illustrative key informant quotes
Person-centered care	• The biggest benefit of collecting interpreters needed is improving the patient care experience for patients with limited English proficiency. [Health system representative] • Accuracy of language is especially important because you need to understand what specific problems the patient is experiencing and ask follow-up questions to make a solid differential diagnosis. [FQHC[a] representative] • It helps drastically with patient satisfaction when we think about being able to communicate with your provider, understand what they're saying, and feeling comfortable that you are understanding medical jargon in your native tongue. [Health system representative]
Transition of care	• If that information is exchanged, then it can be good, because other providers have that information in advance, and can arrange for interpretation services. This could be really good in situations where it's a referral. [Health system representative] • I think one place where exchanging this field would be helpful is when we bring in contracted specialists...So, if a patient was going for a first-time visit with an external provider, that would be valuable for the specialist to know, and to meet with patient needs. [Health system representative] • I could see exchanging Interpreter Needed being useful for patients that are being transferred to long-term acute care hospitals (LTACH). That facility should have the ability to see what language the patient needs to be served in. [FQHC representative]
Health care administration	• Requiring the [interpreter needed] data element might help normalize requesting interpreter services, and/or create a downstream effect of improving the interpreter services payment structure. [Health system representative]

[a]FQHC: Federally Qualified Health Center.

Use Case 1: Person-Centered Care

Improving communication with patients and ensuring person-centered care is the central purpose behind collecting information on a patient's self-reported preference for an interpreter and their preferred language. Key informants and available literature both indicate that interpretation services improve patient comprehension of clinician discussions, patient care quality, and care processes [11,20].

Patients with NELP who receive professional interpretation services are less likely to be readmitted to the hospital and more likely to experience a shorter length of stay than patients with NELP who do not receive interpretation services [11]. Key informants echoed that interpretation services can enhance the delivery of safer patient care and patient decision-making by facilitating better communication between providers and patients regarding patient preferences, medical and nonmedical care needs, clinical information, and care instructions [21].

Finally, collecting information on a patient's self-reported preference for an interpreter also facilitates services for patients with communication needs beyond spoken language, such as indications for an American Sign Language interpreter.

Use Case 2: Transitions of Care

Multiple key informants stated that exchanging data on a patient's interpretation needs facilitates care transitions and referrals to specialty care, social care organizations, and postacute care (PAC) facilities. Exchanging information on a patient's self-reported preference for an interpreter allows specialists to arrange interpretation services in advance of the patient's appointment and may decrease the likelihood that patients are lost to follow-up due to interpretation service scheduling issues. Sharing patient language assistance needs through social care coordination platforms such as Unite Us facilitates patient access to interpreters at community partner organizations.

PAC facilities are required to collect data on whether a patient needs or wants an interpreter using standardized patient assessments under the CMS Quality Reporting Program. The data are collected using the Interpreter Needed LOINC terminology [22], allowing for interoperable data exchange among postacute providers and timely coordination of interpretation services.

Use Case 3: Health Care Administration

A few key informants noted that collecting data on the number of patients requiring interpretation services assists health care

XSL•FO
RenderX

administrators in tailoring such services to adequately serve their patient population. For example, administrators can provide in-person interpretation services for the most requested languages and offer telephonic or remote interpretation services for less commonly requested languages.

Administrators also use information on a patient's self-reported preference for an interpreter to fulfill reporting requirements and support quality improvement efforts. For example, Federally Qualified Health Centers are required to report the number of patients "best served in another language" as part of the Health Resources and Services Administration Uniform Data System. In addition, one of the 2023 quality improvement activities for the CMS Quality Payment Programs is to "Create and Implement a Language Access Plan" that adheres to the National Standards for Culturally and Linguistically Appropriate Services in Health and Health Care [23,24].

EHR Products Support Documentation of Data on a Patient's Self-Reported Preference for an Interpreter

Key informants from the 3 EHR developers stated that functionality to document information on a patient's self-reported preference for an interpreter is available to health care organizations through a structured field in the EHR.

Interpreter Needed Data Element Interoperability Standards

There are well-specified standard terminologies for the interpreter needed data element. The literature and key informants identified both LOINC and SNOMED-CT codes for identifying a patient's self-reported preference for an interpreter: LOINC code 54588-9, "patient/resident's need or want an interpreter to communicate with a doctor or health care staff" and the SNOMED-CT code 315593009, "need for interpreter (finding)" [25,26]. The ASTP/ONC 2024 Interoperability Standards Advisory identifies LOINC code 54588-9 as an available value set to use when communicating information in a patient's preferred language [27].

Mapping information on a patient's self-reported preference for an interpreter to one or both standards can facilitate the exchange of these data between providers, community-based organizations, and others to support patient needs. The HL7 Fast Healthcare Interoperability Resources (FHIR) and Clinical Document Architecture standards define how health care data can be exchanged between health care organizations with different health IT systems [28,29]. In addition, informants identified a FHIR standard (the FHIR Patient Resource-Interpreter Required) for exchanging information on whether a patient "requires an interpreter to communicate health care information to the practitioner" along with patient demographic and administrative data. This FHIR standard offers a practical solution to exchanging these data [30,31]. While our key informants reported that health care organizations do not routinely map information on a patient's self-reported preference for an interpreter to standard terminology, representatives from the 3 EHR developers we spoke to indicate that mapping can be implemented with minimal burden.

Challenges

Our key informants revealed the following 3 potential challenges to using information on a patient's self-reported preference for an interpreter for person-centered care and quality improvement, which are (1) lack of adoption of available data standards, (2) limited electronic exchange, and (3) patient mistrust (Table 2).

XSL•FO
RenderX

Table 2. Challenges to using information on a patient's self-reported preference for an interpreter and illustrative key informant quotes.

Challenge	Illustrative key informant quotes
Lack of adoption of available data standards	• I don't think Interpreter Needed is exchanged right now in a structured field...When I do see data coming through from other systems, I see interpreter information coming through in a free text field. [Health system representative] • If it's a continuation of care document (CCD) that we're sending, it's not a required field. I did check with our [EHR[a] developer lead], and we aren't sharing that Interpreter Needed field. It's currently not technically possible...I imagine if it was something required by regulators, that might spur development. [Health system representative]
Limited electronic exchange of interpreter needed data	• When we refer patients to specialists, the need for an interpreter is often included in free text notes and tends to get lost. I know anecdotally of instances where the specialist's office misses that note, so they don't know to provide an interpreter for the patient. Because it's so challenging for patients to make those appointments in the first place, if the interpreter is not available at the specialist's office the first time, it is common for them to never reschedule. [FQHC[b] representative] • I would guess that Interpreter Needed is a bit difficult to exchange between healthcare systems because it is not consistently labeled with one code" [FQHC representative]
Patient mistrust	• This [interpreter needed] standardized data could also be used in a discriminatory way. On one hand, I want to advocate for patients to have better access to interpreters, but I also don't want it to be used by providers to avoid patients that do need an interpreter. We need to be careful about how we use that information. [Health system representative] • Another issue is that it can feel like there will be an added charge or cost to you if you request a translator, even if it is a free service. Patients don't always feel like they're entitled to ask for those services. Even with Public Charges, when they're told they're entitled to interpretation services, they may not feel like they can access those resources. [Language services professional representative]

[a]EHR: electronic health record.

[b]FQHC: Federally Qualified Health Center.

Lack of Adoption of Available Data Standards

Key informants noted that while most health care organizations currently collect interpreter needed data, most do not map this data to an existing terminology standard. Limited mapping impedes data interoperability and efficient monitoring, tracking, and reporting on the need for interpretation services.

Limited Electronic Exchange of Interpreter Needed Data

Due to the lack of data standardization, there is limited electronic exchange of interpreter needed data among most health care organizations. Key informants stated that a patient's self-reported preference for an interpreter is often exchanged as free text in an administrative or clinical note, or it is shared through phone or fax when notifying a provider that a patient has requested an interpreter. Without standardized electronic exchange, communication of patient needs and preferences is inefficient, creating additional burden on the patient and room for care coordination errors. For instance, key informants shared examples of clinical notes indicating a patient's self-reported preference for an interpreter not being reviewed, causing delayed or missed care if an interpreter is not available at the specialist's office when the patient arrives for the appointment.

Patient Mistrust

The findings from the literature revealed that collecting a patient's self-reported preference for an interpreter from all patients may pose implications for patient trust. For example, a patient with NELP who would benefit from interpretation services may decline due to concerns about privacy and stigmatization [32]. Similarly, negative past experiences with interpreters (eg, misinterpretations or extended wait times for interpretation services) may lead to a patient declining interpretation service [33]. A few key informants also suggested that patients with NELP may decline an interpreter because they believe they will be billed for using those services, even when the patient is informed that they will not be charged. Although these findings transcend standardized collection and use of interpreter needed data, they are important challenges for health care organizations to consider when asking patients for this information.

Discussion

Principal Findings

Overall, while EHR developers and health systems indicated that interpreter needed data are routinely collected and used by most health care organizations, these data are not often mapped to available standards within the EHR. All the health care organizations and EHR developers we spoke with described well-defined administrative and clinical workflows to facilitate data collection of patients' self-reported preference for an interpreter.

Collection of Interpreter Needed Data Can Improve Quality of Care

Despite limited standardization, our findings suggest the collection, documentation, and use of interpreter needed data can improve the quality of care for patients with NELP. The provision of language and communication services for individuals with a NELP or other communication needs (eg, those who are deaf or hard of hearing) can facilitate the accurate

XSL•FO
RenderX

exchange of information regarding prevention, symptoms, diagnosis, treatment, care coordination, discharge planning, and shared decision-making between patients and their care team [34,35]. In short, addressing communication needs may assist in improving patient satisfaction, patient safety, and health outcomes of historically underserved populations.

Use of Interpreter Needed Data Has the Potential to Enhance Health Equity Research

Much of the research on English proficiency–related disparities in health care uses preferred language data abstracted from the EHR, patient experiences of care surveys (eg, Consumer Assessment of Healthcare Providers and Systems), and standardized PAC patient assessments to identify care gaps. Access to more accurate data on interpretation needs could assist researchers in more granularly identifying important health disparities among patients with NELP, understanding experiences of care among NELP individuals, and longitudinally tracking disparities. In addition, the use of standardized interpreter needed data alongside preferred language can support the development of more accurate estimates of language assistance needs that can be used to define quality measures for language service access for patients with NELP and help providers understand ways to better serve patients with NELP.

Greater Adoption of Standardized Interpreter Needed Data Is Necessary

The National Standards for Culturally and Linguistically Appropriate Services in Health and Health Care provide implementation guidance for offering communication and language assistance, which includes developing processes for identifying the languages an individual communicates in and documenting this information in the patient's medical record [34]. Our findings support using available LOINC and SNOMED-CT codes for documenting a patient's self-reported preference for an interpreter. Adoption of these standards will further efforts to use interpreter needed data for quality improvement initiatives aimed at achieving the Quintuple Aim.

Exchange of Interpreter Needed Data Has the Potential to Improve Disease Surveillance

EHR developers and health system key informants noted the potential benefits of exchanging interpreter needed data between health care organizations and public health authorities for the purpose of certain mandated public reporting, such as disease surveillance. For example, contact tracing could be more effective if information on individuals with language service access needs was shared in advance with public health authorities, thus improving communication regarding disease risk and transmission control.

Limitations

The primary objective of this study was to identify health care organization practices for documenting a patient's self-reported preference for an interpreter in the EHR, and the adoption level of available standards. We acknowledge that quantifying the impact of standardized interpreter needed data on health outcomes and patient experiences is important; however, this was out of the scope of our project. Future research on the adoption of the interpreter needed data element will help elucidate the impact of standardized data collection practices to address health disparities and improve patient outcomes.

Conclusion

Improving the quality of services provided to patients with NELP requires health care organizations to systematically collect information from patients regarding their language preferences and to use these data to plan for and offer communication and language assistance services. Despite the presence of common practices for the collection, documentation, and use of data on patient interpretation needs, the lack of standardization limits the exchange of this information among health systems. The use of standard terminologies for interpreter needed data offers one mechanism to facilitate the consistent use of the data within and across health care organizations, resulting in quality improvement [36]. As the US demographics change, there is the potential for unmet patient needs to become more glaring. Standardized data collection and EHR documentation can streamline the use of these data and may help to reduce health disparities.

Acknowledgments

The report was prepared under the contract GS-10F-0033M/75FCMC19F0079 between the Department of Health and Human Services Centers for Medicare & Medicaid Services Office of Minority Health and NORC at the University of Chicago. The views expressed in this article are the authors and do not necessarily represent the views of the US Department of Health and Human Services or the US Government.

Conflicts of Interest

None declared.

References

1. Rajbhandari P, Keith MF, Braidy R, Gunkelman SM, Smith E. Interpreter use for limited English proficiency patients/families: a QI Study. Hosp Pediatr 2021;11(7):718-726. [doi: 10.1542/hpeds.2020-003889] [Medline: 34583318]

XSL•FO
RenderX

2. Ortega P, Shin TM, Martínez GA. Rethinking the term "Limited English Proficiency" to improve language-appropriate healthcare for all. J Immigr Minor Health 2022;24(3):799-805 [FREE Full text] [doi: 10.1007/s10903-021-01257-w] [Medline: 34328602]
3. Fernandez A, Schillinger D, Warton EM, Adler N, Moffet HH, Schenker Y, et al. Language barriers, physician-patient language concordance, and glycemic control among insured Latinos with diabetes: the diabetes study of Northern California (DISTANCE). J Gen Intern Med 2011;26(2):170-176 [FREE Full text] [doi: 10.1007/s11606-010-1507-6] [Medline: 20878497]
4. Salinas M, Elena M, Mendoza J, Gaddy B. Hispanics experience higher risk of heart disease, stroke while facing language barriers. ABC News. 2023. URL: https://abcnews.go.com/GMA/Wellness/hispanics-experience-higher-risk-heart-disease-stroke-facing/story?id=103328285 [accessed 2023-09-15]
5. Flores G. The impact of medical interpreter services on the quality of health care: a systematic review. Med Care Res Rev 2005;62(3):255-299. [doi: 10.1177/1077558705275416] [Medline: 15894705]
6. VanderWielen LM, Enurah AS, Rho HY, Nagarkatti-Gude DR, Michelsen-King P, Crossman SH, et al. Medical interpreters: improvements to address access, equity, and quality of care for limited-English-proficient patients. Acad Med 2014;89(10):1324-1327 [FREE Full text] [doi: 10.1097/ACM.0000000000000296] [Medline: 25054413]
7. Taira BR, Kim K, Mody N. Hospital and health system-level interventions to improve care for limited English proficiency patients: a systematic review. Jt Comm J Qual Patient Saf 2019;45(6):446-458. [doi: 10.1016/j.jcjq.2019.02.005] [Medline: 30910471]
8. FAQ - translators and interpreters. National Council on Interpreting in Health Care. URL: https://www.ncihc.org/faq-translators-and-interpreters [accessed 2024-12-17]
9. Haldar S, Pillai D, Artiga S. Overview of health coverage and care for individuals with limited english proficiency (LEP). KFF. 2023. URL: https://www.kff.org/racial-equity-and-health-policy/issue-brief/overview-of-health-coverage-and-care-for-individuals-with-limited-english-proficiency/ [accessed 2024-12-17]
10. Frequently requested statistics on immigrants and immigration in the United States. Migration Policy Institute. 2023. URL: https://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states#:~:text=In%202021%2C%20approximately%2046%20percent,country's%2025.9%20million%20LEP%20individuals [accessed 2024-12-17]
11. Karliner LS, Jacobs EA, Chen AH, Mutha S. Do professional interpreters improve clinical care for patients with limited English proficiency? A systematic review of the literature. Health Serv Res 2007;42(2):727-754 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00629.x] [Medline: 17362215]
12. Schwei RJ, Schroeder M, Ejebe I, Lor M, Park L, Xiong P, et al. Limited English proficient patients' perceptions of when interpreters are needed and how the decision to utilize interpreters is made. Health Commun 2018;33(12):1503-1508 [FREE Full text] [doi: 10.1080/10410236.2017.1372047] [Medline: 28929813]
13. Flores G, Abreu M, Barone CP, Bachur R, Lin H. Errors of medical interpretation and their potential clinical consequences: a comparison of professional versus ad hoc versus no interpreters. Ann Emerg Med 2012;60(5):545-553. [doi: 10.1016/j.annemergmed.2012.01.025] [Medline: 22424655]
14. Ali PA, Watson R. Language barriers and their impact on provision of care to patients with limited English proficiency: Nurses' perspectives. J Clin Nurs 2018;27(5-6):e1152-e1160 [FREE Full text] [doi: 10.1111/jocn.14204] [Medline: 29193568]
15. Kuo D, Fagan MJ. Satisfaction with methods of Spanish interpretation in an ambulatory care clinic. J Gen Intern Med 1999;14(9):547-550 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.07258.x] [Medline: 10491243]
16. Sukhera J. Narrative reviews: flexible, rigorous, and practical. J Grad Med Educ 2022;14(4):414-417 [FREE Full text] [doi: 10.4300/JGME-D-22-00480.1] [Medline: 35991099]
17. Holmgren AJ, Apathy NC. Trends in US hospital electronic health record vendor market concentration, 2012-2021. J Gen Intern Med 2023;38(7):1765-1767 [FREE Full text] [doi: 10.1007/s11606-022-07917-3] [Medline: 36348217]
18. 2015 Edition market readiness for hospitals and clinicians. Office of the National Coordinator for Health Information Technology. 2019. URL: https://www.healthit.gov/data/quickstats/2015-edition-market-readiness-hospitals-and-clinicians [accessed 2022-12-01]
19. USCDI ONDEC submission prep sheet. HealthIT.gov. URL: https://www.healthit.gov/isp/ONDEC [accessed 2024-10-08]
20. Committee on Accounting for Socioeconomic Status in Medicare Payment Programs, Board on Population Health and Public Health Practice, Board on Health Care Services, Institute of Medicine, Medicine National Academies of Sciences. Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors Kindle Edition. Washington, DC: National Academies Press; 2016.
21. Center for Clinical Standards and Quality and the Office of Minority Health, RTI International, RAND Corporation. Final specifications for SNF QRP quality measures and standardized patient assessment data elements (SPADEs). Centers for Medicare & Medicaid Services. 2019. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/Final-Specifications-for-SNF-QRP-Quality-Measures-and-SPADEs.pdf [accessed 2024-12-17]
22. CMS data element library. Centers for Medicare & Medicaid Services. URL: https://del.cms.gov/DELWeb/pubHome [accessed 2024-09-08]

XSL•FO
RenderX

23. Traditional MIPSxplore measures & activities. U.S. Centers for Medicare & Medicaid Services Quality Payment Program. URL: https://qpp.cms.gov/mips/explore-measures?tab=improvementActivities&py=2023 [accessed 2023-08-28]
24. U.S. Department of Health and Human Services Office of Minority Health. National culturally and linguistically appropriate services standards. 2023. URL: https://thinkculturalhealth.hhs.gov/clas/standards [accessed 2023-08-25]
25. Interpreter needed. LOINC. URL: https://loinc.org/54588-9 [accessed 2023-09-15]
26. Need for interpreter (finding). SNOMED CT Browser. URL: https://browser.ihtsdotools.org/?perspective=full&conceptId1=315593009&edition=MAIN/2023-10-01&release=&languages=en [accessed 2023-09-15]
27. 2024 Interoperability standards advisory reference edition. Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/isp/sites/isp/files/2024-02/2024%20ISA%20Reference%20Edition_508.pdf [accessed 2024-10-23]
28. What is FHIR? Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf [accessed 2024-12-17]
29. International H. section 1a: clinical document architecture (CDA). HL7 International. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7 [accessed 2024-12-17]
30. Office of the National Coordinator for Health Information Technology. Introduction to FHIR resources. URL: https://www.healthit.gov/sites/default/files/page/2021-04/Intro%20to%20FHIR%20Resources%20Fact%20Sheet.pdf [accessed 2024-12-17]
31. Extension: patient interpreter required. HL7 International. URL: https://hl7.org/fhir/extensions/StructureDefinition-patient-interpreterRequired.html [accessed 2024-12-17]
32. Khoong EC, Fernandez A. Addressing gaps in interpreter use: time for implementation science informed multi-level interventions. J Gen Intern Med 2021;36(11):3532-3536 [FREE Full text] [doi: 10.1007/s11606-021-06823-4] [Medline: 33948799]
33. Steinberg EM, Valenzuela-Araujo D, Zickafoose JS, Kieffer E, DeCamp LR. The "Battle" of managing language barriers in health care. Clin Pediatr (Phila) 2016;55(14):1318-1327 [FREE Full text] [doi: 10.1177/0009922816629760] [Medline: 26896341]
34. Office of Minority Health, U.S. Department of Health and Human Services. National standards for culturally and linguistically appropriate services in health and health care: a blueprint for advancing and sustaining CLAS policy and practice. Think Cultural Health. 2013 Apr. URL: https://thinkculturalhealth.hhs.gov/assets/pdfs/enhancedclasstandardsblueprint.pdf [accessed 2025-01-07]
35. IMPACT Act of 2014 Data Standardization and Cross Setting Measures. Centers for Medicare & Medicaid Services. 2023. URL: https://www.cms.gov/medicare/quality/initiatives/pac-quality-initiatives/impact-act-2014-data-standardization-cross-setting-measures [accessed 2023-09-15]
36. HHS invests $8 million to improve health care access by training new physicians to care for individuals with disabilities and individuals with limited english proficiency. U.S. Department of Health and Human Services. 2023. URL: https://www.hhs.gov/about/news/2023/01/25/hhs-invests-8-million-improve-health-care-access-training-new-physicians-care-individuals-disabilities-individuals-limited-english-proficiency.html [accessed 2024-12-17]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
DEL: Data Element Library
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level 7
LOINC: Logical Observation Identifiers Names and Codes
NELP: non-English language preference
PAC: postacute care
SNOMED-CT: Systematized Medical Nomenclature for Medicine-Clinical Terminology
USCDI: United States Core Data for Interoperability

XSL•FO
RenderX

Edited by G Tsafnat; submitted 03.06.24; peer-reviewed by U Sinha, R Schwei; comments to author 21.09.24; revised version received 31.10.24; accepted 22.11.24; published 31.01.25.

Please cite as:
Heaney-Huls K, Shams R, Nwefo R, Kane R, Gordon J, Laffan AM, Stare S, Dullabh P
Electronic Health Record Data Collection Practices to Advance Standardization and Interoperability of Patient Preferences for Interpretation Services: Qualitative Study
J Med Internet Res 2025;27:e62670
URL: https://www.jmir.org/2025/1/e62670
doi:10.2196/62670
PMID:39888652

©Krysta Heaney-Huls, Rida Shams, Ruth Nwefo, Rachel Kane, Janna Gordon, Alison M Laffan, Scott Stare, Prashila Dullabh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Prognostic Significance of Sleep and Circadian Rhythm for Myocardial Infarction Outcomes: Case-Control Study

Wei-Chih Chin[1,2,3*], MD; Pao-Hsien Chu[2,4*], MD; Lung-Sheng Wu[2,4], MD; Kuang-Tso Lee[2,4], MD; Chen Lin[5], PhD; Chien-Te Ho[2,4], MD; Wei-Sheng Yang[6], MD; I-Hang Chung[1,2], MD; Yu-Shu Huang[1,2], MD, PhD

[1]Division of Psychiatry and Sleep Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[2]College of Medicine, Chang Gung University, Taoyuan City, Taiwan

[3]College of Life Sciences and Medicine, National Tsing Hua University, Hsinchu, Taiwan

[4]Department of Cardiology, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[5]Department of Biomedical Sciences and Engineering, National Central University, Taoyuan, Taiwan

[6]Department of Psychiatry, New Taipei City Tucheng Hospital, New Taipei City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Shu Huang, MD, PhD
Division of Psychiatry and Sleep Center
Chang Gung Memorial Hospital
No. 5, Fuxing St., Guishan
Taoyuan, 333423
Taiwan
Phone: 886 3 3281200 ext 2479
Email: yushuhuang1212@gmail.com

Abstract

Background: Myocardial infarction (MI) is a medical emergency resulting from coronary artery occlusion. Patients with acute MI often experience disturbed sleep and circadian rhythm. Most previous studies assessed the premorbid sleep and circadian rhythm of patients with MI and their correlations with cardiovascular disease. However, little is known about post-MI sleep and circadian rhythm and their impacts on prognosis. The use of actigraphy with different algorithms to evaluate sleep and circadian rhythm after acute MI has the potential for predicting outcomes and preventing future disease progression.

Objective: We aimed to evaluate how sleep patterns and disrupted circadian rhythm affect the prognosis of MI, using actigraphy and heart rate variability (HRV). Nonparametric analysis of actigraphy data was performed to examine the circadian rhythm of patients.

Methods: Patients with MI in the intensive care unit (ICU) were enrolled alongside age- and gender-matched healthy controls. Actigraphy was used to evaluate sleep and circadian rhythm, while HRV was monitored for 24 hours to assess autonomic nerve function. Nonparametric indicators were calculated to quantify the active-rest patterns, including interdaily stability, intradaily variability, the most active 10 consecutive hours (M10), the least active 5 consecutive hours (L5), the relative amplitude, and the actigraphic dichotomy index. Follow-ups were conducted at 3 and 6 months after discharge to evaluate prognosis, including the duration of current admission, the number and duration of readmission and ICU admission, and catheterization. Independent sample *t* tests and analysis of covariance were used to compare group differences. Pearson correlation tests were used to explore the correlations of the parameters of actigraphy and HRV with prognosis.

Results: The study included 34 patients with MI (mean age 57.65, SD 9.03 years) and 17 age- and gender-matched controls. MI patients had significantly more wake after sleep onset, an increased number of awakenings, and a lower sleep efficiency than controls. Circadian rhythm analysis revealed significantly lower daytime activity in MI patients. Moreover, these patients had a lower relative amplitude and dichotomy index and a higher intradaily variability and midpoint of M10, suggesting less sleep and wake activity changes, more fragmentation of the rest-activity patterns, and a more delayed circadian rhythm. Furthermore, significant correlations were found between the parameters of circadian rhythm analysis, including nighttime activity, time of M10 and L5, and daytime and nighttime activity$_{SD}$, and patient prognosis.

Conclusions: Patients with acute MI experienced significantly worse sleep and disturbed circadian rhythm compared with healthy controls. Our actigraphy-based analysis revealed a disturbed circadian rhythm, including reduced daytime activities,

XSL•FO
RenderX

greater fluctuation in hourly activities, and a weak rest-activity rhythm, which were correlated with prognosis. The evaluation of sleep and circadian rhythm in patients with acute MI can serve as a valuable indicator for prognosis and should be further studied.

(J Med Internet Res 2025;27:e63897) doi:10.2196/63897

KEYWORDS

myocardial infarction; circadian rhythm; actigraphy; nonparametric analysis; prognosis; sleep; heart rate variability; activity

Introduction

Myocardial infarction (MI) is a medical emergency resulting from the occlusion of coronary arteries, which causes impaired perfusion and myocardial damage. It continues to be a leading cause of mortality globally despite progress in clinical management and revascularization techniques, especially in developed countries [1,2]. Annually, more than 2.4 million deaths occur due to MI in the United States, and over 4 million deaths occur across Europe and northern Asia. Furthermore, the burden of cardiovascular disease and MI has increased in middle- and low-income countries, where more than 80% of cardiovascular deaths now occur [3], highlighting its significant socioeconomic impact.

Accumulating evidence shows that the incidence and prognosis of MI, including mortality, cardiovascular events such as reinfarction, admission duration, and readmission, are influenced by numerous factors. Risk factors identified in the INTERHEART study include smoking status, lipid profile, blood pressure, diabetes, obesity, diet, physical activity, alcohol consumption, and psychosocial factors [4,5]. These risk factors account for over 90% of the risk of acute MI and are similar in all geographic regions, races, and genders. According to a previous study [6], when all modifiable risk factors are optimal, the lifetime risk of coronary artery disease or MI for the middle age group is estimated to be less than 5%, and with more than two major risk factors, it rises to 50% for men and 31% for women. These findings highlight the importance of identifying modifiable risk factors. The correlations among sleep, circadian rhythm, and risk of cardiovascular disease have been studied [7,8]. Insomnia has been found to be associated with cardiovascular and metabolic disease risk and mortality [9]. Both short sleep (less than 5 hours) and long sleep (more than 9 hours) can increase the incidence of cardiovascular events [10,11]. Disturbed circadian rhythm due to shift work can also increase cardiovascular risk [12]. Additionally, the frequency and mortality of MI have been found to increase in the morning [13], and patients with onset during the dark-to-light transition period can have a larger infarct size [14]. This is known as the circadian variation of MI, which is related to variations in clock genes [15].

Nowadays, many studies have shown the efficacy of the use of wearable technology in clinical practice and its potential in predicting outcomes. Actigraphy objectively measures patients' activities and can represent their sleep-wake patterns and sleep quality. Studies have found that monitoring sleep patterns and activity intensity can help in the diagnosis, monitoring, and treatment of depression, anxiety, and narcolepsy [16-18]. Different mathematical algorithm analyses of actigraphy recordings can predict the recurrence of panic attacks in patients with panic disorder and hypersomnia attacks in those with Kleine-Levin syndrome [19,20]. While abundant evidence supports the correlation of heart rate variability (HRV) with the prognosis of MI [21], actigraphy may also provide valuable insights. A study has identified its correlation with the outcomes of patients with heart failure [22], while the study by Peter-Marske et al [23] discovered associations of actigraphy-measured physical activity and sedentary behavior with the incidence of cardiovascular disease, MI, and ischemic stroke. The application of different mathematical algorithms of actigraphy to evaluate sleep and circadian rhythm and their correlations with the prognosis of MI can thus be worthy of investigation.

Currently, most studies have evaluated the premorbid sleep and circadian rhythm of patients with MI and analyzed their correlations. Little is known about sleep and circadian rhythm following the onset of acute MI. Although studies have reported the predictive role of HRV for the prognosis of MI, no study has explored the correlations among sleep, circadian rhythm, and prognosis. Therefore, we conducted a case-control prospective follow-up study to investigate the sleep and circadian rhythm of patients with acute MI, using actigraphy, along with HRV assessment. We followed these patients for 6 months after discharge to evaluate their prognosis. Considering the crucial role of circadian rhythm in cardiovascular function [8], actigraphy data were further analyzed with different mathematical algorithms to assess the circadian rhythm of patients.

Methods

Participants

This case-control prospective follow-up study aimed to evaluate sleep and circadian rhythm in patients with acute MI and their correlations with prognosis. Patients who met the inclusion criteria without the exclusion criteria were recruited from the intensive care unit (ICU) following admission. They received standard MI treatments and light therapy during their ICU stay, alongside assessments by a team of cardiologists and psychiatrists experienced in sleep medicine during hospitalization.

The inclusion criteria were as follows: (1) age between 35 and 85 years, (2) diagnosis of acute MI (Killip I-III), and (3) agreement to participate in this study and provide informed consent. The exclusion criteria were as follows: (1) unclear consciousness, (2) use of sedative or hypnotic drugs, (3) blindness or severe cataracts, (4) severe neurological diseases (eg, seizure, brain injury, or severe organic brain diseases), (5) severe psychiatric disorders (eg, schizophrenia, bipolar disorder,

XSL•FO
RenderX

mental retardation, or substance use disorder), (6) unwillingness to participate, and (7) high mortality risk.

An age- and gender-matched healthy control group was also recruited for comparison.

Study Design

Upon admission to the ICU, patients were screened and invited if they met the inclusion criteria and were not excluded based on the criteria. Physical evaluation and the diagnosis and treatments of acute MI were conducted by experienced cardiologists, while mental status evaluation was performed by psychiatrists to rule out severe psychiatric disorders. One-hour light therapy was administered each morning during ICU stay to prevent delirium. Baseline evaluation included age, gender, BMI, and blood pressure. Participants also underwent blood sampling for tests, including blood routine tests (white blood cell count, red blood cell count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, platelet, platelet distribution width, and mean platelet volume), renal and liver function tests (blood urea nitrogen, creatinine, alanine aminotransferase, and aspartate aminotransferase), electrolyte tests (sodium and potassium), thyroid function tests (thyroid-stimulating hormone and free T4), and vitamin D tests. Actigraphy was used to evaluate sleep and circadian rhythm, while HRV was assessed within 24 hours of ICU admission to evaluate autonomic nerve function (Figure 1). Participants' conditions were monitored throughout their ICU stay and subsequent general ward stay. Two delirium screening scales (the Confusion Assessment Method for the ICU and the Intensive Care Delirium Screening Checklist) were administered by psychiatrists daily to screen delirium during ICU stay and every 2 days in the general ward [24,25], and only 1 participant had delirium during the study period. Prognosis was evaluated with follow-ups at 3 and 6 months after discharge, and data on prognosis, including the duration of current admission, the number and duration of readmission and ICU admission, and catheterization during the 6-month follow-up period, were collected and confirmed via phone contact and by medical records (Figure 1).

Figure 1. Timeline of the study. Patients with myocardial infarction (MI) underwent actigraphy after admission to the intensive care unit (ICU) and kept wearing the device until discharge. They also underwent heart rate variability (HRV) evaluation for the initial 24 hours in the ICU. After discharge, they were followed for 6 months to evaluate the prognosis.

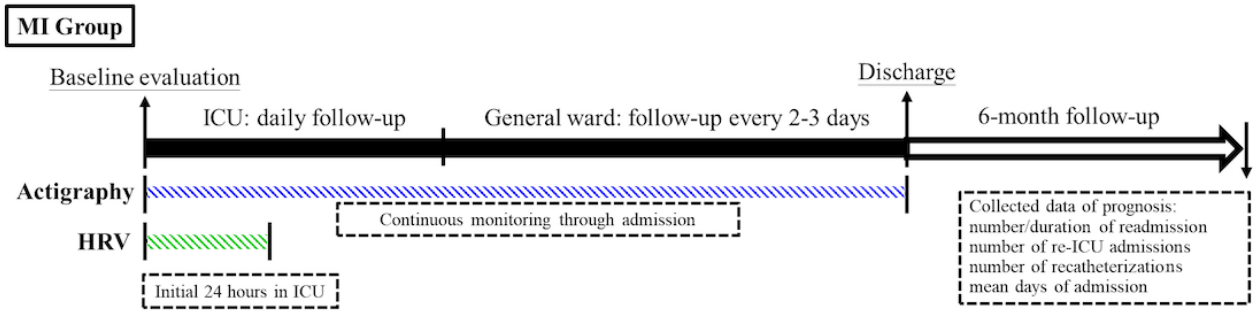

Actigraphy

Actigraphy, a nonintrusive approach involving a wristwatch-like device, can monitor participants' motions and light exposures over extended periods [26]. Data analysis by the widely used parametric method provided sleep parameters, including sleep onset latency, total sleep time, sleep efficiency, wake after sleep onset, and times of awakening, and all 34 participants were included in the analysis. The estimation of these parameters is based on the observation that there is less motion during sleep than during wakefulness [27]. Participants wore the device from enrollment until discharge or withdrawal from the study, which could occur if they met exclusion criteria such as hypnotic use or delirium. Healthy controls wore the device for 2 weeks. To analyze circadian rhythm, participants with actigraphy recording of less than 3 days were excluded. Consequently, 28 of the 34 patients with acute MI and 17 healthy controls were included in the circadian rhythm analysis (Figure 2). Representative actograms for the MI and control groups are shown in Figures 3A and 3B, respectively.

XSL•FO
RenderX

Figure 2. The flowchart outlines the methodological framework for circadian rhythm analysis in the study, from patient recruitment to the processing and analysis of actigraphic data. The chart includes the criteria for patient enrollment based on age and diagnosis, and patients with recording of less than 3 days were excluded from the circadian rhythm analysis. HRV: heart rate variability; ICU: intensive care unit; MI: myocardial infarction.

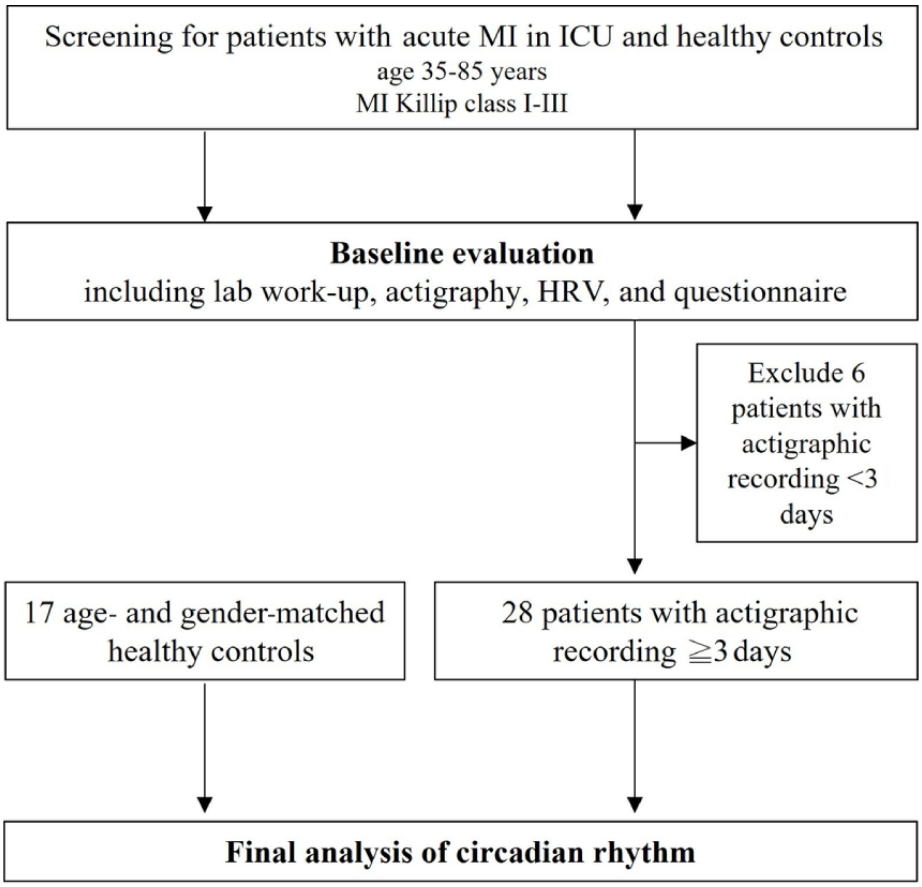

Figure 3. Representative actograms for the myocardial infarction (A) and healthy control (B) groups.

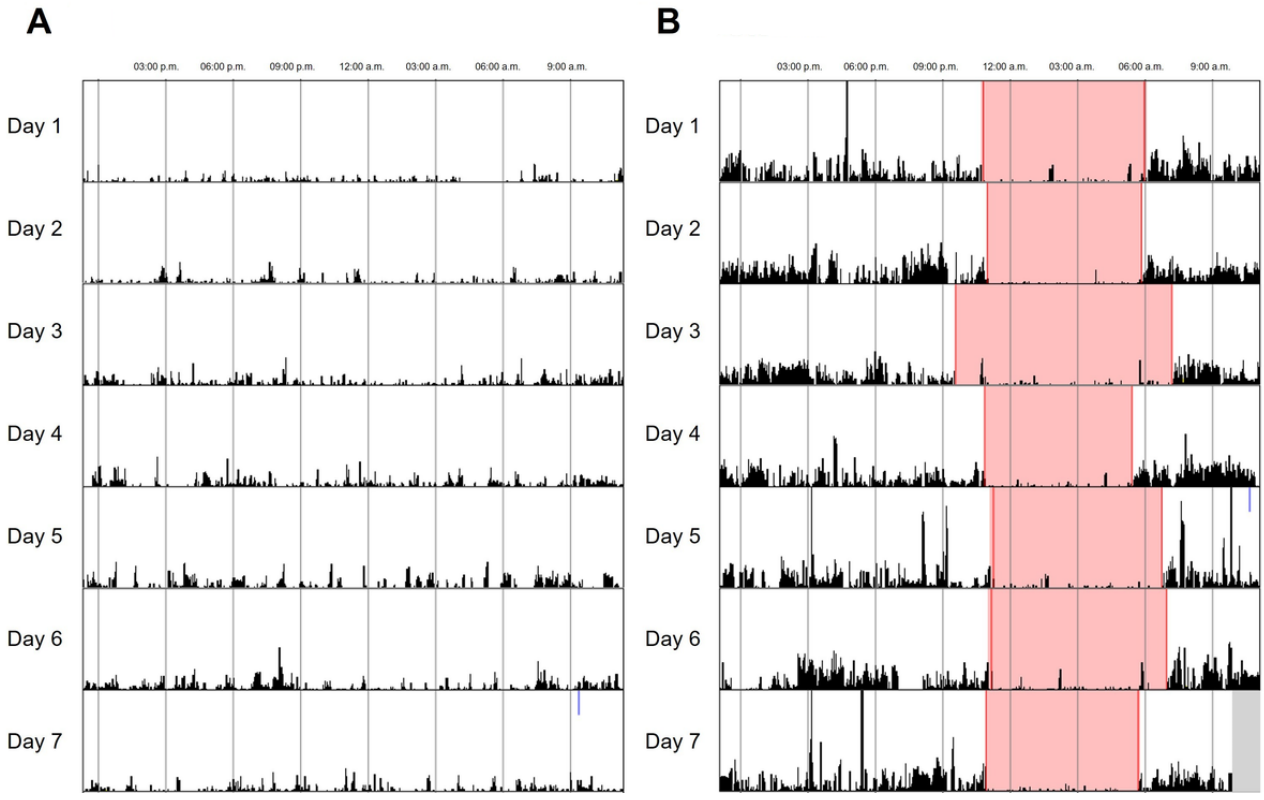

XSL•FO
RenderX

Circadian Rhythm Analysis and Algorithm of Actigraphy

Since the circadian rhythms of patients with MI can be disrupted and are not easily modeled by a 24-hour sinusoidal wave, we calculated several nonparametric indicators based on the hourly averaged activity from minute-by-minute actigraphy epochs. These indicators are designed to quantify the active-rest patterns and include interdaily stability, intradaily variability, the most active 10 consecutive hours (M10), the least active 5 consecutive hours (L5), the relative amplitude, and the actigraphic dichotomy index (I<O). Detailed mathematical formulas can be found elsewhere [20,28-30]. The interdaily stability and intradaily variability values quantify the stability and fragmentation of the rest-active patterns, respectively, and lower interdaily stability and higher intradaily variability indicate a disrupted circadian rhythm. While M10 and L5 identify the peak and trough of daily activity intensities, relative amplitude quantifies the daily circadian activity amplitude by calculating the difference between M10 and L5, normalized to their means. The midpoints of M10 and L5 represent the central times of these periods. The average daily values of M10, L5, and relative amplitude, and the midpoints of M10 and L5 were used to represent the circadian activity profiles of each participant (Multimedia Appendix 1). The actigraphic dichotomy index (I<O) measures the percentage of in-bed activity counts that fall below the median of out-of-bed activity counts. A lower index suggests a weaker circadian rest-activity rhythm [31].

HRV Assessment

HRV, reflecting autonomic nervous system tone, was monitored for 24 hours in patients with acute MI after enrollment. Healthy controls underwent HRV assessment for 15 minutes. It has a well-established role as a marker of cardiovascular risk [32]. Previous studies have found that HRV can predict the mortality rate of patients with acute MI [33-35]. Six patients with a history of arrhythmia were excluded, and 22 patients with acute MI and 16 healthy controls were included in the HRV analysis (Figure 4).

Figure 4. Protocol for heart rate variability (HRV) monitoring within the initial 24 hours of intensive care unit (ICU) admission for patients with acute myocardial infarction (MI). The chart delineates the steps from HRV recording to data analysis.

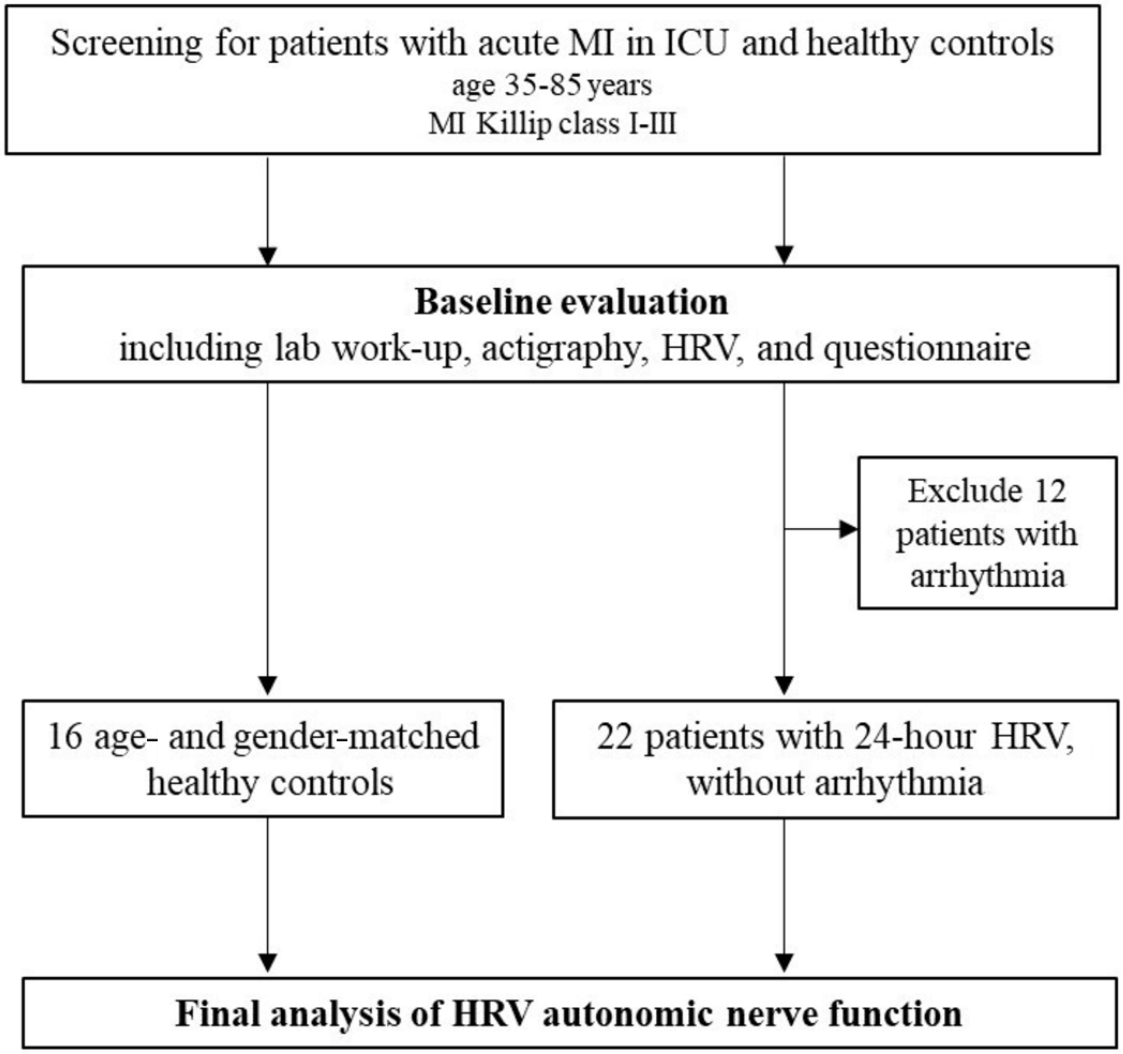

XSL•FO
RenderX

Statistical Analysis

Data were analyzed using SPSS version 20.0 (IBM Corp). Variables were presented by mean and percentage. Independent sample *t* tests and analysis of covariance (ANCOVA) were used to compare differences between the MI group and healthy control group, and age, gender, and BMI were controlled. Pearson correlation tests measured correlations between the parameters of actigraphy and HRV and the prognosis of patients with MI, and partial correlation coefficient (*r*) analysis was used to adjust age and gender. A *P* value of <.05 was considered statistically significant, while a *P* value of <.01 was considered highly statistically significant.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of Chang Gung Memorial Hospital (201902108A3). Before enrollment, participants were fully informed about the purpose, voluntary participation, and confidentiality of the study, and all signed an informed consent form. Data were anonymized and deidentified. Participants were not offered any form of compensation for their involvement.

Results

A total of 34 patients with acute MI (mean age 57.65, SD 9.03 years; range 41-73 years; 29 males and 5 females) were enrolled, and all completed the 6-month follow-up. The healthy control group consisted of 17 age- and gender-matched controls (mean age 57.71, SD 12.56 years; 11 males and 6 females). Table 1 shows the comparison of baseline characteristics between the MI group and the healthy control group. BMI was significantly higher in the MI group than in the control group (mean 26.79, SD 4.11 vs mean 23.09, SD 2.15; *P*=.001). After controlling for BMI, laboratory findings showed that many variables were significantly lower in the MI group than in the control group, including vitamin D (mean 17.43, SD 5.00 vs mean 26.53, SD 5.72 ng/mL; *P*<.001), sodium (mean 137.39, SD 2.75 vs mean 139.47, SD 7.25 mEq/L; *P*=.049), potassium (mean 3.83, SD 0.45 vs mean 4.15, SD 0.27 mEq/L; *P*=.04), and thyroid-stimulating hormone levels (mean 0.91, SD 0.46 vs mean 1.66, SD 0.62 µIU/mL; *P*=.02). On the other hand, some variables were significantly higher in the MI group than in the control group, including white blood cell count (mean 9.87, SD 2.56 vs mean 6.47, SD 1.87 ×1000/µL; *P*<.001), aspartate aminotransferase (mean 37.06, SD 23.77 vs mean 19.53, SD 3.47 U/L; *P*=.02), and alanine aminotransferase (mean 31.88, SD 17.62 vs mean 19.71, SD 5.72 U/L; *P*=.04). Of the 34 participants, 5 (18%) experienced initial symptoms of MI within an hour after awakening, and the onset time for 6 participants could not be confirmed due to atypical clinical presentations.

XSL•FO
RenderX

Table 1. Comparison of baseline characteristics between the myocardial infarction and healthy control groups.

Variable	Myocardial infarction group (n=34)	Healthy control group (n=17)	*P* value[a]
Age (years), mean (SD)	57.65 (9.03)	57.71 (12.56)	.99
Gender, n (%)			.09
Male	29 (85)	11 (65)	
Female	5 (15)	6 (35)	
BMI (kg/m^2), mean (SD)	26.79 (4.11)	23.09 (2.15)	.001
Initial symptoms within 1 hour after awakening, n (%)	5 (18)	—[b]	—
Laboratory finding, mean (SD)			
Total vitamin D (ng/mL)	17.43 (5.00)	26.53 (5.72)	<.001
WBC[c] (1000/μL)	9.87 (2.56)	6.47 (1.87)	<.001
RBC[d] (million/μL)	4.64 (1.01)	4.74 (0.20)	.42
HGB[e] (g/dL)	13.89 (2.87)	14.16 (0.74)	.64
HCT[f] (%)	41.56 (8.33)	43.09 (2.07)	.37
MCV[g] (fL)	90.04 (6.48)	90.91 (2.36)	.79
MCH[h] (pg/cell)	30.11 (2.63)	29.88 (0.86)	.22
MCHC[i] (gHb/dL)	33.40 (1.38)	32.89 (0.90)	.07
RDW[j] (%)	13.26 (1.49)	12.77 (0.64)	.76
Platelet (1000/μL)	237.00 (64.28)	215.82 (45.94)	.90
PDW[k] (fL)	10.44 (1.60)	9.91 (1.58)	.53
MPV[l] (fL)	9.72 (0.73)	9.39 (0.81)	.31
BUN[m] (mg/dL)	31.79 (33.32)	16.25 (4.20)	.19
Creatinine (mg/dL)	2.66 (3.95)	0.81 (0.22)	.19
ALT[n] (U/L)	31.88 (17.62)	19.71 (5.72)	.04
AST[o] (U/L)	37.06 (23.77)	19.53 (3.47)	.02
Na (mEq/L)	137.39 (2.75)	139.47 (7.25)	.049
K (mEq/L)	3.83 (0.45)	4.15 (0.27)	.04
Free T4 (ng/dL)	0.98 (0.22)	0.93 (0.11)	.97
TSH[p] (μIU/mL)	0.91 (0.46)	1.66 (0.62)	.002

[a]*P* values were calculated using independent sample *t* tests; laboratory findings were assessed using analysis of covariance (ANCOVA) controlling for BMI.

[b]Not applicable.

[c]WBC: white blood cell.

[d]RBC: red blood cell.

[e]HGB: hemoglobin.

[f]HCT: hematocrit.

[g]MCV: mean corpuscular volume.

[h]MCH: mean corpuscular hemoglobin.

[i]MCHC: mean corpuscular hemoglobin concentration.

[j]RDW: red cell distribution width.

[k]PDW: platelet distribution width.

[l]MPV: mean platelet volume.

[m]BUN: blood urea nitrogen.

XSL•FO
RenderX

[n]ALT: alanine aminotransferase.

[o]AST: aspartate aminotransferase.

[p]TSH: thyroid-stimulating hormone.

Table 2 shows the comparison of the general parametric sleep parameters of actigraphy between the MI group and the healthy control group after controlling for age, gender, and BMI. In the MI group, the average duration of actigraphy monitoring was 5.75 (SD 2.01) days. The parametric analysis of actigraphy showed that patients with MI had significantly more wake after sleep onset (mean 97.70, SD 59.95 vs mean 50.2, SD 18.17 min; *P*=.01), increased number of awakenings (mean 33.29, SD 12.61 vs mean 22.25, SD 6.14 times; *P*=.007), and lower sleep efficiency (mean 77.96, SD 7.73% vs mean 84.65, SD 4.03%; *P*=.01) than controls (Table 2). Circadian rhythm analysis of actigraphy is shown in Table 3 after controlling for age, gender, and BMI. Compared with the control group, the circadian patterns in the MI group showed less diurnal change and more individual differences (Figures 5 and 6; Multimedia Appendix 2 and 3). Moreover, patients with MI had significantly lower daytime activity (mean 70.65, SD 32.74 vs mean 188.46, SD 84.53; *P*<.001) and averaged M10 (mean 91.00, SD 42.79 vs mean 233.66, SD 106.05; *P*<.001) than controls, suggesting lower activity when awake. Both nighttime activity and averaged L5 were not significantly different between the groups, but the MI group had significantly lower daytime (mean 130.27, SD 39.55 vs mean 283.69, SD 100.78; *P*<.001) and nighttime activity$_{SD}$ (mean 87.13, SD 42.12 vs mean 122.58, SD 43.71; *P*=.03), suggesting less daytime and nighttime activity daily changes. The MI group had a significantly lower relative amplitude (mean 0.69, SD 0.13 vs mean 0.87, SD 0.10; *P*<.001) and dichotomy index (mean 0.85, SD 0.05 vs mean 0.93, SD 0.04; *P*<.001) than the control group, indicating that patients with MI had less sleep and wake activity changes and a weak active-rest rhythm. Moreover, the significantly higher intradaily variability in the MI group than in the control group (mean 1.13, SD 0.27 vs mean 0.94, SD 0.21; *P*=.02) suggested that the MI group had more fragmentations and disturbances during the wake and sleep states. The time of M10 (mean 14.86, SD 1.76 vs mean 13.52, SD 1.07; *P*=.02) was significantly higher in the MI group than in the control group, which supported a more delayed circadian rhythm.

Table 2. Comparison of the results of actigraphy between the myocardial infarction and healthy control groups after controlling for age, gender, and BMI.

Variable	Myocardial infarction group (n=28)	Healthy control group (n=17)	*P* value[a]
SOL[b] (min), mean (SD)	33.99 (34.65)	14.40 (9.98)	.11
TST[c] (min), mean (SD)	501.74 (199.41)	420.59 (88.32)	.15
SE[d] (%), mean (SD)	77.96 (7.73)	84.65 (4.03)	.01
WASO[e] (min), mean (SD)	97.70 (59.95)	50.2 (18.17)	.01
Awake (times), mean (SD)	33.29 (12.61)	22.25 (6.14)	.007

[a]*P* values were calculated using analysis of covariance (ANCOVA) controlling for age, gender, and BMI.

[b]SOL: sleep onset latency.

[c]TST: total sleep time.

[d]SE: sleep efficiency.

[e]WASO: wake after sleep onset.

XSL•FO
RenderX

Table 3. Comparison of the results of circadian rhythm analysis of actigraphy between the myocardial infarction and healthy control groups after controlling for age, gender, and BMI.

Variable	Myocardial infarction group (n=28)	Healthy control group (n=17)	*P* value[a]
Activity count, mean (SD)			
Daytime activity	70.65 (32.74)	188.46 (84.53)	<.001
Daytime activity$_{SD}$	130.27 (39.55)	283.69 (100.78)	<.001
Nighttime activity	33.16 (26.65)	37.45 (14.86)	.53
Nighttime activity$_{SD}$	87.13 (42.12)	122.58 (43.71)	.03
Active-rest rhythm, mean (SD)			
M10[b]	91.00 (42.79)	233.66 (106.05)	<.001
Time of M10	14.86 (1.76)	13.52 (1.07)	.02
L5[c]	16.02 (9.10)	14.84 (8.38)	.92
Time of L5	6.47 (7.73)	2.80 (1.25)	.08
RA[d]	0.69 (0.13)	0.87 (0.10)	.001
IV[e]	1.13 (0.27)	0.94 (0.21)	.02
IS[f]	0.31 (0.12)	0.41 (0.12)	.06
Dichotomy index	0.85 (0.05)	0.93 (0.04)	<.001

[a]*P* values were calculated using analysis of covariance (ANCOVA) controlling for age, gender, and BMI.

[b]M10: most active 10 consecutive hours.

[c]L5: least active 5 consecutive hours.

[d]RA: relative amplitude.

[e]IV: intradaily variability.

[f]IS: interdaily stability.

Figure 5. The average circadian activity template of the myocardial infarction (MI) patient cohort, synthesized from actigraphy data. It depicts the group's collective rest-activity cycles over a standard 24-hour period, highlighting the marked decreased amplitude of daytime and nighttime activity changes observed after acute MI.

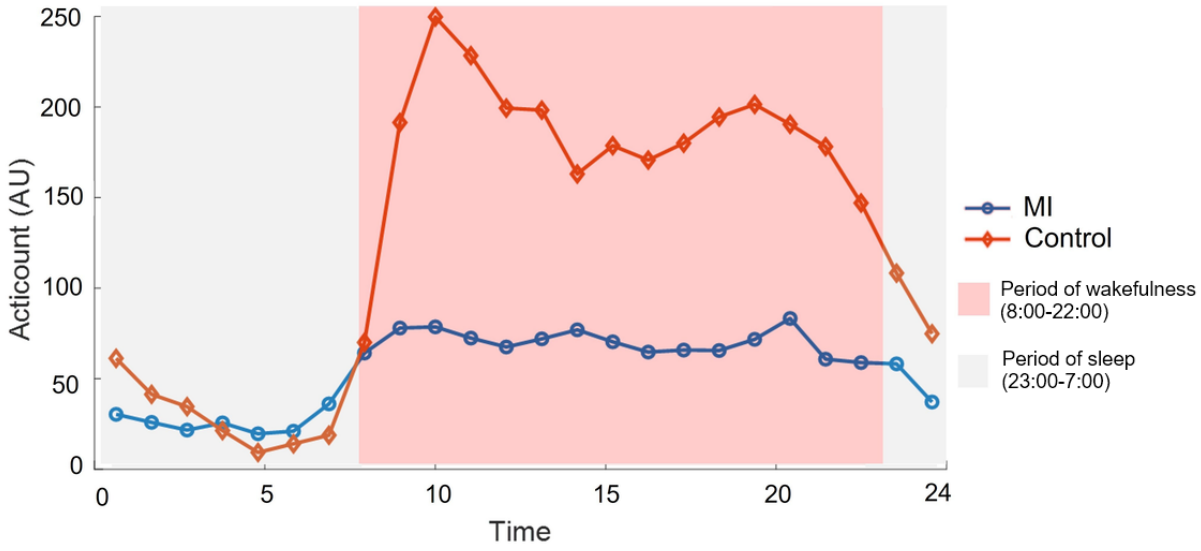

XSL•FO
RenderX

Figure 6. Acrophase map showing the timings of the midpoint of M10 of myocardial infarction (MI) patients and healthy controls. Patients with MI had higher individual variability in the timing of the midpoint of M10 compared to healthy controls. M10: most active 10 consecutive hours.

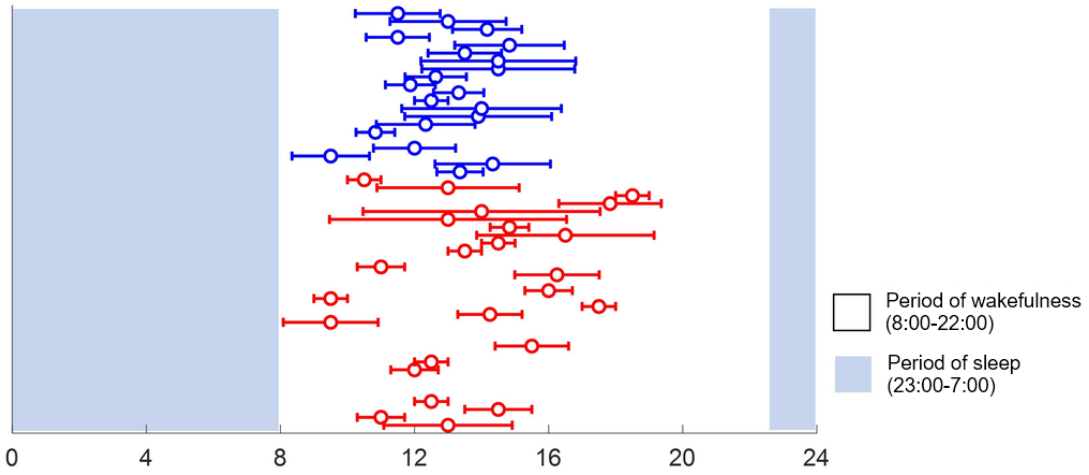

Table 4 shows the comparison of the parameters of HRV between the MI group and the healthy control group after controlling for age, gender, and BMI. The MI group showed a significantly lower root mean square of successive differences (mean 17.27, SD 9.49 vs mean 28.94, SD 17.22; *P*=.04), percentage of successive RR intervals that differ by more than 50 ms (mean 3.19, SD 4.63 vs mean 12.33, SD 16.60; *P*=.04), Poincaré plot SD perpendicular to the line of identity (SD1) (mean 12.20, SD 6.70 vs mean 20.43, SD 12.16; *P*=.04), and SD ratio (mean 0.27, SD 0.09 vs mean 0.44, SD 0.23; *P*=.01) than the control group, suggesting lower parasympathetic nerve function in the MI group.

Table 4. Comparison of the results of heart rate variability between the myocardial infarction and healthy control groups after controlling for age, gender, and BMI.

Variable	Myocardial infarction group (n=22)	Healthy control group (n=16)	*P* value[a]
Overall, mean (SD)	836.35 (129.88)	878.91 (130.74)	.40
HR[b], mean (SD)	73.53 (10.12)	69.65 (9.96)	.32
SDRR[c], mean (SD)	37.89 (19.59)	37.38 (15.39)	.90
RMSSD[d], mean (SD)	17.27 (9.49)	28.94 (17.22)	.04
pNN50[e], mean (SD)	3.19 (4.63)	12.33 (16.60)	.04
pNN20[f], mean (SD)	21.45 (19.14)	38.34 (25.12)	.13
SD1[g], mean (SD)	12.20 (6.70)	20.43 (12.16)	.04
SD2[h], mean (SD)	51.83 (27.08)	47.96 (19.96)	.84
SD ratio[i], mean (SD)	0.27 (0.09)	0.44 (0.23)	.01
DC value[j], mean (SD)	6.25 (3.33)	9.22 (5.40)	.25

[a]*P* values were calculated using analysis of covariance (ANCOVA) controlling for age, gender, and BMI.

[b]HR: heart rate.

[c]SDRR: SD of the RR interval.

[d]RMSSD: root mean square of successive differences.

[e]pNN50: percentage of successive RR intervals that differ by more than 50 ms.

[f]pNN20: percentage of successive RR intervals that differ by more than 20 ms.

[g]SD1: Poincaré plot SD perpendicular to the line of identity.

[h]SD2: Poincaré plot SD along the line of identity.

[i]SD ratio: ratio of SD1 to SD2.

[j]DC value: deceleration capacity of heart rate.

Table 5 shows the correlations between the parameters of actigraphy and the prognosis of patients with MI during the 6-month follow-up. Significantly positive correlations were found between nighttime activity and total days of current

XSL•FO
RenderX

admission (r=0.529; P=.004); between time of M10 and any ICU readmission (r=0.452; P=.02), times of readmission (r=0.496; P=.007), total days of readmission (r=0.571; P=.001), and mean days of readmission (r=0.459; P=.01); and between time of L5 and total days of current admission (r=0.504; P=.006). Significantly negative correlations were found between daytime activity$_{SD}$ and any ICU readmission (r=–0.417; P=.03), times of readmission (r=–0.415; P=.03), and total days of readmission (r=–0.450; P=.02); and between nighttime activity$_{SD}$ and times of readmission (r=–0.421; P=.03).

Table 5. Correlations (r) of the parameters of actigraphy and the prognosis of patients with myocardial infarction (n=28).

Variable	Total days of current admission	Any readmission	Any ICU[a] readmission	Times of readmission	Total days of readmission	Any catheterization	Mean days of readmission
TST[b]	–0.164	0.166	0.092	0.199	0.066	0.309	–0.011
SOL[c]	–0.089	0.068	–0.172	0.044	–0.111	0.283	–0.134
SE[d]	0.141	0.305	0.158	0.310	0.281	0.228	0.303
WASO[e]	–0.075	0.260	–0.039	0.204	–0.065	0.279	–0.068
Awake number	–0.304	0.120	0.104	0.120	–0.122	0.207	–0.206
Activity counts							
Daytime activity	0.347	–0.107	–0.334	–0.269	–0.314	–0.265	–0.152
Daytime activity$_{SD}$	0.171	–0.269	–0.417[f]	–0.415[f]	–0.450[f]	–0.271	–0.330
Nighttime activity	0.529[g]	–0.068	–0.151	–0.171	–0.138	–0.166	0.006
Nighttime activity$_{SD}$	0.314	–0.319	–0.357	–0.421[f]	–0.358	–0.288	–0.246
Active-rest rhythm							
M10[h]	0.369	–0.176	–0.351	–0.326	–0.336	–0.312	–0.180
Time of M10 (pm)	–0.040	0.358	0.452[f]	0.496[g]	0.571[g]	0.125	0.459[f]
L5[i]	0.352	–0.117	–0.256	–0.255	–0.317	–0.100	–0.206
Time of L5 (am)	0.504[g]	0.207	0.088	0.079	–0.002	0.004	0.167
RA[j]	–0.074	–0.130	–0.097	–0.107	–0.028	–0.169	–0.012
IV[k]	–0.124	0.282	0.328	0.343	0.315	0.122	0.250
IS[l]	–0.332	–0.060	–0.227	–0.153	–0.312	–0.050	–0.293
Dichotomy index	–0.354	–0.143	0.029	–0.016	0.101	–0.261	–0.011

[a]ICU: intensive care unit.
[b]TST: total sleep time.
[c]SOL: sleep onset latency.
[d]SE: sleep efficiency.
[e]WASO: wake after sleep onset.
[f]P<.05.
[g]P<.01.
[h]M10: most active 10 consecutive hours.
[i]L5: least active 5 consecutive hours.
[j]RA: relative amplitude.
[k]IV: intradaily variability.
[l]IS: interdaily stability.

After adjusting for age and gender, there were still significant correlations between several parameters of actigraphy and the prognosis of patients with MI (Multimedia Appendix 4). Significantly positive correlations were found between nighttime activity and total days of current admission (r=0.56; P=.003); and between time of M10 and any ICU readmission (r=0.457; P=.02), times of readmission (r=0.527; P=.006), total days of readmission (r=0.579; P=.002), and mean days of readmission (r=0.466; P=.02). Significantly negative correlations were found between IS and recatheterization (r=–0.423; P=.03) and mean days of readmission (r=–0.413; P=.04).

Table 6 shows the correlations between the parameters of HRV and the prognosis of patients with MI during the 6-month

XSL•FO
RenderX

follow-up. Significantly positive correlations were found between the SD ratio and days of readmission (r=0.425; P=.049), recatheterization (r=0.426; P=.048), and mean days of readmission (r=0.425; P=.049). Significantly negative correlations were found between the SD of the RR interval and any readmission (r=–0.447; P=.04) and times of readmission (r=–0.447; P=.04); between Poincaré plot SD along the line of identity (SD2) and any readmission (r=–0.448; P=.04), times of readmission (r=–0.448; P=.04), days of readmission (r=–0.424; P=.049), and mean days of readmission (r=–0.424; P=.049); and between the deceleration capacity of heart rate (DC value) and any readmission (r=–0.457; P=.03) and times of readmission (r=–0.457; P=.03).

Table 6. Correlations (r) of the parameters of heart rate variability and the prognosis of patients with myocardial infarction (n=22).

Variable	Total days of current admission	Any readmission	Any ICU[a] readmission	Times of readmission	Total days of readmission	Any catheterization	Mean days of readmission
Overall mean	0.053	–0.260	–0.069	–0.260	–0.173	–0.121	–0.173
HR[b]	–0.072	0.245	0.041	0.245	0.147	0.096	0.147
SDRR[c]	0.189	–0.447[d]	–0.032	–0.447[d]	–0.421	–0.314	–0.421
RMSSD[e]	–0.038	–0.370	–0.159	–0.370	–0.323	–0.164	–0.323
pNN50[f]	0.062	–0.339	–0.142	–0.339	–0.301	–0.175	–0.301
pNN20[g]	–0.037	–0.351	–0.182	–0.351	–0.317	–0.100	–0.317
SD1[h]	–0.038	–0.370	–0.159	–0.370	–0.323	–0.164	–0.323
SD2[i]	0.199	–0.448[d]	–0.023	–0.448[d]	–0.424[d]	–0.321	–0.424[d]
SD ratio[j]	–0.116	0.317	–0.294	0.317	0.425[d]	0.426[d]	0.425[d]
DC value[k]	–0.095	–0.457[d]	–0.126	–0.457[d]	–0.403	–0.278	–0.403

[a]ICU: intensive care unit.

[b]HR: heart rate.

[c]SDRR: SD of the RR interval.

[d]P<.05.

[e]RMSSD: root mean square of successive differences.

[f]pNN50: percentage of successive RR intervals that differ by more than 50 ms.

[g]pNN20: percentage of successive RR intervals that differ by more than 20 ms.

[h]SD1: Poincaré plot SD perpendicular to the line of identity.

[i]SD2: Poincaré plot SD along the line of identity.

[j]SD ratio: ratio of SD1 to SD2.

[k]DC value: deceleration capacity of heart rate.

After adjusting for age and gender, there were fewer significant correlations between the parameters of HRV and the prognosis of patients with MI (Multimedia Appendix 5). Significantly positive correlations were found only between the SD ratio and days of readmission (r=0.455; P=.04), recatheterization (r=0.456; P=.04), and mean days of readmission (r=0.455; P=.04).

Discussion

Principal Findings

Previous studies have used actigraphy to evaluate activities and sleep in patients with MI [23,36,37]. However, to our knowledge, no study has examined the circadian rhythm of patients after the onset of acute MI using actigraphy and analyzed its correlations with prognosis. Therefore, our findings are significant as they not only reveal sleep and circadian rhythm disturbances after acute MI but also support the application of actigraphy and HRV in evaluating the prognosis.

Unsurprisingly, baseline laboratory work-up assessments showed significant differences between patients with MI and healthy controls. Patients with MI exhibited more obesity and more abnormalities, including a lower vitamin D level, which is known to be negatively associated with the incidence of cardiovascular diseases [38,39]. This study further revealed significant differences in sleep patterns between the 2 groups. After acute MI, patients had significantly worse sleep, characterized by more wake after sleep onset and number of awakenings, and lower sleep efficiency. These findings align with a previous study, which has reported sleep disturbances in patients after acute MI, including more wakefulness and awakening, more stage shifts, less rapid eye movement, decreased sleep efficiency, and increased daytime sleep [40]. These changes can be attributed to MI itself rather than differences in a hospital setting [40] and can be normalized with an improved physical condition. Sleep has been proposed as a vital sign and a marker for disease and recovery [41], and thus,

XSL•FO
RenderX

routine monitoring and evaluation of sleep are recommended in patients with MI.

The circadian rhythm analysis using actigraphy data supports that both sleep quality and circadian rhythm can be disturbed in patients with MI. Compared with controls, the activities of patients with MI were significantly less during the daytime, while nighttime activities remained unchanged. Moreover, there was less variability (SD) in both daytime and nighttime activities, indicating less change in activity levels throughout the day. Although these findings may be linked to the bed rest and activity restrictions commonly advised for patients with acute MI, the higher intradaily variability contradicts this hypothesis and instead suggests greater fluctuations in hourly activities during both the daytime and nighttime. Furthermore, a lower relative amplitude and actigraphic dichotomy index suggested less distinct activity changes between sleep and wakefulness and a weak circadian rest-activity rhythm. Patients with MI also exhibited a significant delay in their circadian rhythm, as indicated by a higher time of M10.

Most studies on circadian rhythm and MI have focused on the circadian variances of MI [42-44], which can be explained by the fact that circadian rhythm–related coronary vasoconstriction and increased sympathetic tone may disrupt atherosclerotic plaque [45]. In this study, 5 out of 34 participants (18%) experienced initial symptoms within 1 hour after awakening, but we did not find any correlation between their onset time and prognosis. The nonsignificant result can be influenced by several factors. We only enrolled patients with Killip I to III and excluded those with severe comorbidities and a high mortality risk. Some participants could have recall bias due to their atypical clinical presentations. Circadian variances of MI and their correlations with prognosis necessitate further investigation.

Delayed and disturbed circadian rhythm has been found to be associated with cardiovascular diseases [46,47]. Consistent with our findings, a cross-sectional study also found a lower 24-hour rest-activity rhythm amplitude and overall rhythmicity in patients with coronary artery disease, using actigraphy [48]. Although little is known about circadian rhythm after the onset of MI, our findings indicate that several parameters derived from the circadian rhythm analysis have significant correlations with prognosis, including the total days of current admission, ICU readmission, and the number and duration of readmissions during the 6-month follow-up period. Patients with a more delayed circadian rhythm (higher times of L5 and M10) were more likely to have a longer duration of current admission, more frequent readmissions, and a longer duration of readmission. Since gender and age can have impacts on both sleep and the prognosis of MI [49-51], we adjusted age and gender in the correlation analysis. Most of the significant correlations could still be found (Multimedia Appendix 4). Although a causal relationship cannot be confirmed, these findings are consistent with those of previous studies that suggest abnormal circadian rhythm as a core vulnerability factor for MI [15,52]. Interventions targeting the circadian rhythm of patients with MI, especially those with a disturbed and delayed sleep phase, may be beneficial. Exercise programs and rehabilitations have already been demonstrated to improve prognosis and reduce mortality in patients with MI [53], and they can also help improve sleep. During the acute phase, light therapy and environmental modifications may aid in the regulation of circadian rhythm and are worthy of further investigation.

We found that patients with MI had lower parasympathetic activities compared to healthy controls, and there were correlations between several HRV parameters and prognosis. HRV is an indirect measure of autonomic nervous activity [29] and has been investigated for its ability to predict prognosis in patients with acute MI. Previous studies have reported associations between HRV and mortality or cardiovascular events [54-57]. In this study, better parasympathetic nerve activity after acute MI could be correlated with a better prognosis. However, we also found positive correlations between the ratio of SD1 to SD2 (SD ratio) and prognosis, including recatheterization and the duration of readmission, contradicting other findings. A previous study found a correlation between a high SD ratio and arrhythmia such as ventricular tachycardia [58]. Thus, although a higher parasympathetic nerve activity can correlate with a better prognosis, the SD ratio may not always be “the higher, the better.”

Comparing correlations with the prognosis of MI, the circadian rhythm analysis using actigraphy showed more significant findings than the HRV analysis, especially in nighttime activity and the time of M10. Moreover, fewer significant correlations between HRV parameters and prognosis were found after adjusting for age and gender (Multimedia Appendix 5). These findings suggest that the evaluation of sleep and circadian rhythm after the onset of MI is a valuable prognosis indicator. Previous studies have demonstrated the potential of actigraphy to predict the outcomes of various diseases [19,20]. The application of actigraphy in the evaluation and treatment of patients with MI is promising, as it has already been suggested to play a role in the prognostication of heart failure [59]. Another study showed that the evaluation of subjective sleep complaints through questionnaires may serve as a prognostic marker in patients after acute MI [60], emphasizing the importance of sleep in the risk stratification of MI. Although our study found fewer HRV correlations with prognostic factors, it differed from past research as there was no morality after excluding patients with Killip IV. Furthermore, instead of analyzing mortality, our study focused on correlations with prognostic factors such as admission duration, readmission, and recatheterization. Given that only 22 cases were included in the HRV analysis, factors, including age, gender, and medications such as beta-blockers [61], might have affected the findings. Further research is required to compare the 2 evaluation tools in terms of their correlations with the prognosis of MI.

Limitations

The study has some limitations. First, the study had a small sample size, which limits detailed subgroup analysis by age or gender. Second, we only evaluated participants after the onset of MI during their admission, and therefore, the observed delayed and disturbed circadian rhythm may not represent their premorbid sleep and circadian rhythm. They did not undergo actigraphy monitoring during follow-up, and thus, subsequent changes in sleep and circadian rhythm were not known. Further

XSL•FO
RenderX

studies following MI patients with actigraphy after discharge can increase the understanding of the long-term changes in sleep and circadian rhythm after MI and their impacts on prognosis. Third, patients in our study received light therapy during their stay in the ICU to prevent delirium. However, the duration of light therapy was short (1 hour per day for only 1-2 days), limiting its potential impacts. Fourth, some patients with MI were excluded, such as those classified as Killip IV. Therefore, caution is required when generalizing our findings, especially to severe cases with high mortality rates. Lastly, comorbidities that can contribute to sleep disturbance were not controlled. Sleep disorders, such as obstructive sleep apnea, cannot be ruled out, as polysomnography is challenging to perform in patients with acute MI in the ICU setting. Nevertheless, our findings not only support but also extend the use of actigraphy beyond its conventional role in sleep medicine to include applications within cardiology.

Conclusions

Patients with acute MI experienced significantly worse sleep and disturbed circadian rhythm compared with healthy controls. Actigraphy-based analysis revealed a disturbed circadian rhythm, including reduced daytime activities, a greater fluctuation in hourly activities, and a weak rest-activity rhythm, which were correlated with prognosis. The evaluation of sleep and circadian rhythm in patients with acute MI can serve as a valuable indicator for prognosis and warrants further investigation.

Acknowledgments

This study was partially supported by a grant from the Chang Gung Memorial Hospital (CMRPG3L0291-G3L0292, CMRPG3N0441) awarded to WCC.

Data Availability

Anonymized data that support the findings of this study are available upon reasonable request from the corresponding author (YSH).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Circadian rhythm analysis.
[DOCX File , 22 KB - jmir_v27i1e63897_app1.docx]

Multimedia Appendix 2
Individual circadian rhythm profiles of patients with myocardial infarction (MI).
[DOCX File , 215 KB - jmir_v27i1e63897_app2.docx]

Multimedia Appendix 3
Individual circadian rhythm profiles of healthy controls.
[DOCX File , 151 KB - jmir_v27i1e63897_app3.docx]

Multimedia Appendix 4
Correlations of the parameters of actigraphy and prognosis of patients with myocardial infarction after adjusting for age and gender.
[DOCX File , 23 KB - jmir_v27i1e63897_app4.docx]

Multimedia Appendix 5
Correlations of the parameters of heart rate variability and prognosis of patients with myocardial infarction after adjusting for age and gender.
[DOCX File , 22 KB - jmir_v27i1e63897_app5.docx]

References

1. Anderson JL, Morrow DA. Acute myocardial infarction. N Engl J Med 2017 May 25;376(21):2053-2064. [doi: 10.1056/nejmra1606915]
2. Bergmark BA, Mathenge N, Merlini PA, Lawrence-Wright MB, Giugliano RP. Acute coronary syndromes. Lancet 2022 Apr 02;399(10332):1347-1358 [FREE Full text] [doi: 10.1016/S0140-6736(21)02391-6] [Medline: 35367005]

XSL•FO
RenderX

3. Chandrashekhar Y, Alexander T, Mullasari A, Kumbhani DJ, Alam S, Alexanderson E, et al. Resource and infrastructure-appropriate management of ST-segment elevation myocardial infarction in low- and middle-income countries. Circulation 2020 Jun 16;141(24):2004-2025. [doi: 10.1161/circulationaha.119.041297]
4. Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, INTERHEART investigators. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):953-962. [doi: 10.1016/S0140-6736(04)17019-0] [Medline: 15364186]
5. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):937-952. [doi: 10.1016/S0140-6736(04)17018-9] [Medline: 15364185]
6. Berry JD, Dyer A, Cai X, Garside DB, Ning H, Thomas A, et al. Lifetime risks of cardiovascular disease. N Engl J Med 2012 Jan 26;366(4):321-329. [doi: 10.1056/nejmoa1012848]
7. Daghlas I, Dashti HS, Lane J, Aragam KG, Rutter MK, Saxena R, et al. Sleep duration and myocardial infarction. J Am Coll Cardiol 2019 Sep 10;74(10):1304-1314 [FREE Full text] [doi: 10.1016/j.jacc.2019.07.022] [Medline: 31488267]
8. Lin J, Kuang H, Jiang J, Zhou H, Peng L, Yan X, et al. Circadian rhythms in cardiovascular function: implications for cardiac diseases and therapeutic opportunities. Med Sci Monit 2023 Nov 21;29:e942215 [FREE Full text] [doi: 10.12659/MSM.942215] [Medline: 37986555]
9. Dean YE, Shebl MA, Rouzan SS, Bamousa BAA, Talat NE, Ansari SA, et al. Association between insomnia and the incidence of myocardial infarction: A systematic review and meta-analysis. Clin Cardiol 2023 Apr 25;46(4):376-385 [FREE Full text] [doi: 10.1002/clc.23984] [Medline: 36841256]
10. Meisinger C, Heier M, Löwel H, Schneider A, Döring A. Sleep duration and sleep complaints and risk of myocardial infarction in middle-aged men and women from the general population: the MONICA/KORA Augsburg cohort study. Sleep 2007 Sep;30(9):1121-1127 [FREE Full text] [doi: 10.1093/sleep/30.9.1121] [Medline: 17910384]
11. Sabanayagam C, Shankar A. Sleep duration and cardiovascular disease: results from the National Health Interview Survey. Sleep 2010 Aug;33(8):1037-1042 [FREE Full text] [doi: 10.1093/sleep/33.8.1037] [Medline: 20815184]
12. Torquati L, Mielke GI, Brown WJ, Kolbe-Alexander T. Shift work and the risk of cardiovascular disease. A systematic review and meta-analysis including dose-response relationship. Scand J Work Environ Health 2018 May 01;44(3):229-238 [FREE Full text] [doi: 10.5271/sjweh.3700] [Medline: 29247501]
13. Manfredini R, Boari B, Bressan S, Gallerani M, Salmi R, Portaluppi F, et al. Influence of circadian rhythm on mortality after myocardial infarction: data from a prospective cohort of emergency calls. Am J Emerg Med 2004 Nov;22(7):555-559. [doi: 10.1016/j.ajem.2004.08.014] [Medline: 15666260]
14. Suárez-Barrientos A, López-Romero P, Vivas D, Castro-Ferreira F, Núñez-Gil I, Franco E, et al. Circadian variations of infarct size in acute myocardial infarction. Heart 2011 Jun 27;97(12):970-976. [doi: 10.1136/hrt.2010.212621] [Medline: 21525526]
15. Škrlec I, Milić J, Heffer M, Wagner J, Peterlin B. Circadian clock genes and circadian phenotypes in patients with myocardial infarction. Adv Med Sci 2019 Sep;64(2):224-229. [doi: 10.1016/j.advms.2018.12.003] [Medline: 30818221]
16. Fedor S, Lewis R, Pedrelli P, Mischoulon D, Curtiss J, Picard RW. Wearable technology in clinical practice for depressive disorder. N Engl J Med 2023 Dec 28;389(26):2457-2466. [doi: 10.1056/nejmra2215898]
17. Pastre M, Lopez-Castroman J. Actigraphy monitoring in anxiety disorders: A mini-review of the literature. Front Psychiatry 2022 Aug 3;13:984878 [FREE Full text] [doi: 10.3389/fpsyt.2022.984878] [Medline: 35990052]
18. Chin W, Huang Y, Trevor Lam NY, Mak KY, Tang I, Wang C, et al. Effects of modafinil on nocturnal sleep patterns in patients with narcolepsy: A cohort study. Sleep Med 2024 Jul;119:95-102. [doi: 10.1016/j.sleep.2024.04.024] [Medline: 38669836]
19. Tsai C, Christian M, Kuo Y, Lu CC, Lai F, Huang W. Sleep, physical activity and panic attacks: A two-year prospective cohort study using smartwatches, deep learning and an explainable artificial intelligence model. Sleep Med 2024 Feb;114:55-63. [doi: 10.1016/j.sleep.2023.12.013] [Medline: 38154150]
20. Lin C, Chin W, Huang Y, Chu K, Paiva T, Chen C, et al. Different circadian rest-active rhythms in Kleine-Levin syndrome: a prospective and case-control study. Sleep 2021 Sep 13;44(9):zsab096. [doi: 10.1093/sleep/zsab096] [Medline: 33851710]
21. Brinza C, Floria M, Covic A, Covic A, Scripcariu D, Burlacu A. The usefulness of assessing heart rate variability in patients with acute myocardial infarction (HeaRt-V-AMI). Sensors (Basel) 2022 May 07;22(9):3571 [FREE Full text] [doi: 10.3390/s22093571] [Medline: 35591260]
22. Anchouche K, Elharram M, Oulousian E, Razaghizad A, Avram R, Marquis-Gravel G, et al. Use of actigraphy (wearable digital sensors to monitor activity) in heart failure randomized clinical trials: a scoping review. Can J Cardiol 2021 Sep;37(9):1438-1449 [FREE Full text] [doi: 10.1016/j.cjca.2021.07.001] [Medline: 34256087]
23. Peter-Marske KM, Evenson KR, Moore CC, Cuthbertson CC, Howard AG, Shiroma EJ, et al. Association of accelerometer-measured physical activity and sedentary behavior with incident cardiovascular disease, myocardial infarction, and ischemic stroke: the women's health study. J Am Heart Assoc 2023 Apr 04;12(7):e028180 [FREE Full text] [doi: 10.1161/JAHA.122.028180] [Medline: 36974744]

XSL•FO
RenderX

24. Ely EW, Margolin R, Francis J, May L, Truman B, Dittus R, et al. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit Care Med 2001 Jul;29(7):1370-1379. [doi: 10.1097/00003246-200107000-00012] [Medline: 11445689]
25. Bergeron N, Dubois MJ, Dumont M, Dial S, Skrobik Y. Intensive Care Delirium Screening Checklist: evaluation of a new screening tool. Intensive Care Med 2001 May;27(5):859-864. [doi: 10.1007/s001340100909] [Medline: 11430542]
26. Huang Y, Hsu J, Paiva T, Chin W, Chen I, Guilleminault C. Sleep-disordered breathing, craniofacial development, and neurodevelopment in premature infants: a 2-year follow-up study. Sleep Med 2019 Aug;60:20-25. [doi: 10.1016/j.sleep.2018.10.015] [Medline: 30466820]
27. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 01;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
28. Witting W, Kwa IH, Eikelenboom P, Mirmiran M, Swaab DF. Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. Biol Psychiatry 1990 Mar 15;27(6):563-572. [doi: 10.1016/0006-3223(90)90523-5] [Medline: 2322616]
29. Van Someren EJW, Swaab DF, Colenda CC, Cohen W, McCall WV, Rosenquist PB. Bright light therapy: improved sensitivity to its effects on rest-activity rhythms in Alzheimer patients by application of nonparametric methods. Chronobiol Int 1999 Jul 07;16(4):505-518. [doi: 10.3109/07420529908998724] [Medline: 10442243]
30. Minors D, Akerstedt T, Atkinson G, Dahlitz M, Folkard S, Levi F, et al. The difference between activity when in bed and out of bed. I. Healthy subjects and selected patients. Chronobiol Int 1996 May;13(1):27-34. [doi: 10.3109/07420529609040839] [Medline: 8761934]
31. Natale V, Innominato PF, Boreggiani M, Tonetti L, Filardi M, Parganiha A, et al. The difference between in bed and out of bed activity as a behavioral marker of cancer patients: A comparative actigraphic study. Chronobiol Int 2015;32(7):925-933. [doi: 10.3109/07420528.2015.1053909] [Medline: 26158757]
32. Cygankiewicz I, Zareba W. Heart rate variability. Handb Clin Neurol 2013;117:379-393. [doi: 10.1016/B978-0-444-53491-0.00031-6] [Medline: 24095141]
33. Kleiger RE, Miller J, Bigger J, Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. Am J Cardiol 1987 Feb 01;59(4):256-262 [FREE Full text] [doi: 10.1016/0002-9149(87)90795-8] [Medline: 3812275]
34. Papaioannou VE, Maglaveras N, Houvarda I, Antoniadou E, Vretzakis G. Investigation of altered heart rate variability, nonlinear properties of heart rate signals, and organ dysfunction longitudinally over time in intensive care unit patients. J Crit Care 2006 Mar;21(1):95-103; discussion 103. [doi: 10.1016/j.jcrc.2005.12.007] [Medline: 16616632]
35. Pomeranz B, Macaulay RJ, Caudill MA, Kutz I, Adam D, Gordon D, et al. Assessment of autonomic function in humans by heart rate spectral analysis. American Journal of Physiology-Heart and Circulatory Physiology 1985 Jan 01;248(1):H151-H153. [doi: 10.1152/ajpheart.1985.248.1.h151]
36. Chin W, Huang Y, Wu L, Lee K, Ho C, Lin C, et al. The prognosis of patients with myocardial infarction after light therapy: a preliminary study. Biomed J 2024 Aug 07:100783 [FREE Full text] [doi: 10.1016/j.bj.2024.100783] [Medline: 39122187]
37. Munir H, Goldfarb M. Sleep quality and characteristics of older adults with acute cardiovascular disease. J Geriatr Cardiol 2024 Mar 28;21(3):369-373 [FREE Full text] [doi: 10.26599/1671-5411.2024.03.004] [Medline: 38665281]
38. Pilz S, März W, Wellnitz B, Seelhorst U, Fahrleitner-Pammer A, Dimai HP, et al. Association of vitamin D deficiency with heart failure and sudden cardiac death in a large cross-sectional study of patients referred for coronary angiography. J Clin Endocrinol Metab 2008 Oct;93(10):3927-3935. [doi: 10.1210/jc.2008-0784] [Medline: 18682515]
39. Wang L, Song Y, Manson JE, Pilz S, März W, Michaëlsson K, et al. Circulating 25-hydroxy-vitamin D and risk of cardiovascular disease. Circ: Cardiovascular Quality and Outcomes 2012 Nov;5(6):819-829. [doi: 10.1161/circoutcomes.112.967604]
40. Broughton R, Baron R. Sleep patterns in the intensive care unit and on the ward after acute myocardial infarction. Electroencephalogr Clin Neurophysiol 1978 Sep;45(3):348-360. [doi: 10.1016/0013-4694(78)90187-6] [Medline: 79474]
41. Grandner MA, Malhotra A. Sleep as a vital sign: why medical practitioners need to routinely ask their patients about sleep. Sleep Health 2015 Mar;1(1):11-12 [FREE Full text] [doi: 10.1016/j.sleh.2014.12.011] [Medline: 27840846]
42. Bulluck H, Nicholas J, Crimi G, White SK, Ludman AJ, Pica S, et al. Circadian variation in acute myocardial infarct size assessed by cardiovascular magnetic resonance in reperfused STEMI patients. Int J Cardiol 2017 Mar 01;230:149-154 [FREE Full text] [doi: 10.1016/j.ijcard.2016.12.030] [Medline: 28038815]
43. Fournier S, Eeckhout E, Mangiacapra F, Trana C, Lauriers N, Beggah AT, et al. Circadian variations of ischemic burden among patients with myocardial infarction undergoing primary percutaneous coronary intervention. Am Heart J 2012 Feb;163(2):208-213 [FREE Full text] [doi: 10.1016/j.ahj.2011.11.006] [Medline: 22305838]
44. Muller JE, Tofler GH, Stone PH. Circadian variation and triggers of onset of acute cardiovascular disease. Circulation 1989 Apr;79(4):733-743. [doi: 10.1161/01.cir.79.4.733] [Medline: 2647318]
45. Muller JE, Stone PH, Turi ZG, Rutherford JD, Czeisler CA, Parker C, et al. Circadian variation in the frequency of onset of acute myocardial infarction. N Engl J Med 1985 Nov 21;313(21):1315-1322. [doi: 10.1056/NEJM198511213132103] [Medline: 2865677]

XSL•FO
RenderX

46. Rüger M, Scheer FAJL. Effects of circadian disruption on the cardiometabolic system. Rev Endocr Metab Disord 2009 Dec;10(4):245-260 [FREE Full text] [doi: 10.1007/s11154-009-9122-8] [Medline: 19784781]
47. Rana S, Prabhu SD, Young ME. Chronobiological influence over cardiovascular function. Circulation Research 2020 Jan 17;126(2):258-279. [doi: 10.1161/circresaha.119.313349]
48. Moon C, Benson CJ, Albashayreh A, Perkhounkova Y, Burgess HJ. Sleep, circadian rhythm characteristics, and melatonin levels in later life adults with and without coronary artery disease. J Clin Sleep Med 2023 Feb 01;19(2):283-292 [FREE Full text] [doi: 10.5664/jcsm.10308] [Medline: 36148612]
49. Walton JC, Bumgarner JR, Nelson RJ. Sex differences in circadian rhythms. Cold Spring Harb Perspect Biol 2022 Jul 01;14(7):a039107 [FREE Full text] [doi: 10.1101/cshperspect.a039107] [Medline: 35101914]
50. Liakos M, Parikh PB. Gender disparities in presentation, management, and outcomes of acute myocardial infarction. Curr Cardiol Rep 2018 Jun 16;20(8):64. [doi: 10.1007/s11886-018-1006-7] [Medline: 29909444]
51. Goldberg RJ, Gore JM, Gurwitz JH, Alpert JS, Brady P, Strohsnitter W, et al. The impact of age on the incidence and prognosis of initial acute myocardial infarction: the Worcester Heart Attack Study. Am Heart J 1989 Mar;117(3):543-549. [doi: 10.1016/0002-8703(89)90727-8] [Medline: 2919534]
52. Holmes DR, Aguirre FV, Aplin R, Lennon RJ, Nestler DM, Bell MR, et al. Circadian rhythms in patients with ST-elevation myocardial infarction. Circ: Cardiovascular Quality and Outcomes 2010 Jul;3(4):382-389. [doi: 10.1161/circoutcomes.109.913343]
53. Lawler PR, Filion KB, Eisenberg MJ. Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: a systematic review and meta-analysis of randomized controlled trials. Am Heart J 2011 Oct;162(4):571-584.e2. [doi: 10.1016/j.ahj.2011.07.017] [Medline: 21982647]
54. Balanescu S, Corlan AD, Dorobantu M, Gherasim L. Prognostic value of heart rate variability after acute myocardial infarction. Med Sci Monit 2004 Jul;10(7):CR307-CR315. [Medline: 15232505]
55. Boskovic A, Belada N, Knezevic B. Prognostic value of heart rate variability in post-infarction patients. VSP 2014;71(10):925-930. [doi: 10.2298/vsp1410925b]
56. Coviello I, Pinnacchio G, Laurito M, Stazi A, Battipaglia I, Barone L, et al. Prognostic role of heart rate variability in patients with ST-segment elevation acute myocardial infarction treated by primary angioplasty. Cardiology 2013 Jan 16;124(1):63-70. [doi: 10.1159/000345779] [Medline: 23328532]
57. Katz A, Liberty IF, Porath A, Ovsyshcher I, Prystowsky EN. A simple bedside test of 1-minute heart rate variability during deep breathing as a prognostic index after myocardial infarction. Am Heart J 1999 Jul;138(1 Pt 1):32-38. [doi: 10.1016/s0002-8703(99)70242-5] [Medline: 10385760]
58. Huikuri HV, Seppänen T, Koistinen MJ, Airaksinen KJ, Ikäheimo MJ, Castellanos A, et al. Abnormalities in beat-to-beat dynamics of heart rate before the spontaneous onset of life-threatening ventricular tachyarrhythmias in patients with prior myocardial infarction. Circulation 1996 May 15;93(10):1836-1844. [doi: 10.1161/01.cir.93.10.1836] [Medline: 8635263]
59. Tan MKH, Wong JKL, Bakrania K, Abdullahi Y, Harling L, Casula R, et al. Can activity monitors predict outcomes in patients with heart failure? A systematic review. Eur Heart J Qual Care Clin Outcomes 2019 Jan 01;5(1):11-21 [FREE Full text] [doi: 10.1093/ehjqcco/qcy038] [Medline: 30215706]
60. Clark A, Lange T, Hallqvist J, Jennum P, Rod NH. Sleep impairment and prognosis of acute myocardial infarction: a prospective cohort study. Sleep 2014 May 01;37(5):851-858 [FREE Full text] [doi: 10.5665/sleep.3646] [Medline: 24790263]
61. Fatisson J, Oswald V, Lalonde F. Influence diagram of physiological and environmental factors affecting heart rate variability: an extended literature overview. Heart Int 2016 Apr 25;11(1):e32-e40 [FREE Full text] [doi: 10.5301/heartint.5000232] [Medline: 27924215]

Abbreviations

HRV: heart rate variability
ICU: intensive care unit
L5: least active 5 consecutive hours
M10: most active 10 consecutive hours
MI: myocardial infarction
SD1: Poincaré plot SD perpendicular to the line of identity
SD2: Poincaré plot SD along the line of identity

XSL•FO
RenderX

Edited by A Coristine; submitted 02.07.24; peer-reviewed by A Parganiha, AK Pati; comments to author 07.10.24; revised version received 19.11.24; accepted 18.12.24; published 04.02.25.

Please cite as:
Chin WC, Chu PH, Wu LS, Lee KT, Lin C, Ho CT, Yang WS, Chung IH, Huang YS
The Prognostic Significance of Sleep and Circadian Rhythm for Myocardial Infarction Outcomes: Case-Control Study
J Med Internet Res 2025;27:e63897
URL: https://www.jmir.org/2025/1/e63897
doi:10.2196/63897
PMID:

©Wei-Chih Chin, Pao-Hsien Chu, Lung-Sheng Wu, Kuang-Tso Lee, Chen Lin, Chien-Te Ho, Wei-Sheng Yang, I-Hang Chung, Yu-Shu Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mortality Risk Prediction in Patients With Antimelanoma Differentiation–Associated, Gene 5 Antibody–Positive, Dermatomyositis–Associated Interstitial Lung Disease: Algorithm Development and Validation

Hui Li[1,2*], MD; Ruyi Zou[1*], MD; Hongxia Xin[3*], MD; Ping He[4*], MD; Bin Xi[5*], MD; Yaqiong Tian[1], MD; Qi Zhao[1], MD, PhD; Xin Yan[1], MD; Xiaohua Qiu[1], MD; Yujuan Gao[1], MD, PhD; Yin Liu[1], MD; Min Cao[1], MD; Bi Chen[5], MD, PhD; Qian Han[6], MD, PhD; Juan Chen[3], MD, PhD; Guochun Wang[7], MD, PhD; Hourong Cai[1], MD

[1]Department of Respiratory and Critical Care Medicine, Affiliated Drum Tower Hospital, Nanjing University Medical School, Nanjing, China

[2]Department of Pulmonary and Critical Care Medicine, Seventh Affiliated Hospital, Sun Yatsen University, Shenzhen, China

[3]Department of Pulmonary and Critical Care Medicine, General Hospital of Ningxia Medical University, Yinchuan, China

[4]Department of Respiratory and Critical Care Medicine, Third People's Hospital of Chengdu, Chengdu, China

[5]Department of Respiratory and Critical Care Medicine, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

[6]National Center for Respiratory Medicine, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[7]China-Japan Friendship Hospital, Key Laboratory of Myositis (Beijing Key Laboratory for Immune Mediated Inflammatory Diseases), Department of Rheumatology, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Hourong Cai, MD
Department of Respiratory and Critical Care Medicine
Affiliated Drum Tower Hospital
Nanjing University Medical School
321 Zhongshan Rd., Gulou District
Nanjing, 210008
China
Phone: 86 025 83106666
Email: caihourong2013@163.com

Abstract

Background: Patients with antimelanoma differentiation–associated gene 5 antibody–positive dermatomyositis–associated interstitial lung disease (anti-MDA5+DM-ILD) are susceptible to rapidly progressive interstitial lung disease (RP-ILD) and have a high risk of mortality. There is an urgent need for a reliable prediction model, accessible via an easy-to-use web-based tool, to evaluate the risk of death.

Objective: This study aimed to develop and validate a risk prediction model of 3-month mortality using machine learning (ML) in a large multicenter cohort of patients with anti-MDA5+DM-ILD in China.

Methods: In total, 609 consecutive patients with anti-MDA5+DM-ILD were retrospectively enrolled from 6 hospitals across China. Patient demographics and laboratory and clinical parameters were collected on admission. The primary endpoint was 3-month mortality due to all causes. Six ML algorithms (Extreme Gradient Boosting [XGBoost], logistic regression (LR), Light Gradient Boosting Machine [LightGBM], random forest [RF], support vector machine [SVM], and k-nearest neighbor [KNN]) were applied to construct and evaluate the model.

Results: After applying inclusion and exclusion criteria, 509 (83.6%) of the 609 patients were included in our study, divided into a training cohort (n=203, 39.9%), an internal validation cohort (n=51, 10%), and 2 external validation cohorts (n=92, 18.1%, and n=163, 32%). ML identified 8 important variables as critical for model construction: RP-ILD, erythrocyte sedimentation rate (ESR), serum albumin (ALB) level, age, C-reactive protein (CRP) level, aspartate aminotransferase (AST) level, lactate dehydrogenase (LDH) level, and the neutrophil-to-lymphocyte ratio (NLR). LR was chosen as the best algorithm for model construction, and the model demonstrated excellent performance, with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.866, a sensitivity of 84.8%, and a specificity of 84.4% on the validation data set and an AUC of 0.90, a sensitivity

XSL•FO
RenderX

of 85.0%, and a specificity of 83.9% on the training data set. Calibration curves and decision curve analysis (DCA) confirmed the model's accuracy and clinical applicability. Moreover, the model showed strong predictive performance in the external validation cohorts (cohort 1: AUC=0.836, 95% CI 0.754-0.916; cohort 2: AUC=0.915, 95% CI 0.871-0.959), indicating good generalizability. This model was integrated into a web-based tool to predict the 3-month mortality for patients with anti-MDA5+DM-ILD.

Conclusions: We successfully developed a robust clinical prediction model and an accompanying web tool to estimate the 3-month mortality risk for patients with anti-MDA5+DM-ILD.

(J Med Internet Res 2025;27:e62836) doi:10.2196/62836

KEYWORDS

antimelanoma differentiation–associated gene 5 antibody; dermatomyositis; interstitial lung disease; 3-month mortality; machine learning; ML; tool; web based; mortality; idiopathic inflammatory myopathy; myopathy; lung disease; melanoma; imaging; clinical outcome

Introduction

Idiopathic inflammatory myopathies (IIMs) encompass a heterogenous group of autoimmune diseases characterized by muscle weakness and inflammation. Dermatomyositis (DM), which predominantly affects the skin and muscles, is a prevalent subtype within the IIM spectrum. Interstitial lung disease (ILD) is a common comorbidity in DM, with an incidence varying widely from 5% to 65% [1]. The clinical presentations, treatment responses, and outcomes of IIMs are highly variable and correlate with the presence of specific myositis-specific antibodies (MSAs). One such antibody, the antimelanoma differentiation–associated gene 5 (MDA5) antibody, targets the MDA5 protein, a member of the retinoic acid inducible gene I (RIG-I) receptor family, which is crucial for the detection of viral RNA and the activation of the interferon pathway in response to viral infections [2]. Approximately 1%-30% of patients with DM test positive for anti-MDA5 antibodies [3]. A recent surge in MDA5+DM cases has been observed in England in the wake of the COVID-19 pandemic, with an associated mortality rate of 32% for antimelanoma differentiation–associated gene 5 antibody–positive dermatomyositis–associated interstitial lung disease (anti-MDA5+DM-ILD) [4]. Patients with anti-MDA5 autoantibody often exhibit minimal muscle involvement. Initially, anti-MDA5 autoantibody was identified in a cohort of East Asian patients with clinically amyopathic myositis (81%) and rapidly progressive ILD (74%) [5], with >90% of rapidly progressive interstitial lung disease (RP-ILD) cases and 84% mortality occurring within the first 6 months of disease onset [6]. Notably, the first 3 months following diagnosis are particularly critical, with 46% of the deaths occurring in this period [6]. However, recent studies have revealed that early and aggressive intervention can significantly improve the 6-month survival rate of patients with anti-MDA5+DM-ILD from 33% to 89% [7-9]. Thus, there is an urgent need to accurately identify these high-risk patients at an early stage of the disease.

In recent years, various mortality risk models incorporating independent predictors have been explored to predict outcomes in patients with anti-MDA5+DM-ILD. Lian et al [10] developed the FLAIR score, which includes 5 key indicators: ferritin levels, lactate dehydrogenase (LDH) levels, the anti-MDA5 antibody grade, the high-resolution computed tomography (HRCT) imaging score, and RP-ILD [10]. However, the FLAIR score was designed to predict mortality in patients with amyopathic DM. Some independent prognostic factors, such as Krebs von den Lungen-6 (KL-6), interleukin (IL)-6, IL-18, and soluble CD206, are challenging to access in clinical practice [11], and HRCT imaging scores are subject to considerable interobserver variability [10]. Most existing prediction models rely on logistic or Cox regression analysis, with limited external validation and only fair overall performance. Recent evidence suggests that machine learning (ML) algorithms, which can handle numerous multidimensional variables with nonlinear relationships, may offer superior predictive accuracy for clinical outcomes [12]. To date, ML approaches for predicting mortality risk and prognosis in patients with anti-MDA5+DM-ILD have been infrequent, and there remains a lack of a concise, practical tool for directly calculating mortality risk in these patients.

Considering the limited sample sizes and inadequate establishment and validation of practical quantitative methods in previous studies, we identified the 3-month mortality risk, established a prediction model, and validated the model using real-world data from multicenter studies in China involving patients with anti-MDA5+DM-ILD. The aim was to create a reliable and easily implemented tool for clinical use.

Methods

Ethical Considerations

The Medical Ethics Committee of the Drum Tower Hospital Affiliated to Nanjing University Medical School approved the study (2020-050-01). The patients included in this study were derived from a secondary analysis of the previously confirmed study. The ethics committee determined that a second round of approval was not required.

Patients and Study Design

From January 2017 to December 2022, our study screened patients with anti-MDA5+DM-ILD across participating institutions. Inclusion criteria were patients with a newly observed ILD, detectable serum anti-MDA5, and DM. The diagnostic criteria for DM included Bohan and Peter criteria and were retrospectively reconfirmed according to the 2017 European League Against Rheumatism (EULAR)/American College of Rheumatology IIM classification criteria or the 2018

XSL•FO
RenderX

European Neuromuscular Centre (ENMC) DM criteria by experienced rheumatologists and pulmonary clinicians [13-15]. Exclusion criteria were as follows: (1) DM-ILD with other positive MSAs; (2) other connective tissue diseases (CTDs); (3) evidence of pulmonary infections at initial admission, as diagnosed clinically and by identifying pathogens; and (4) incomplete data or lost to follow-up.

Following application of these criteria, 509 (83.6%) eligible cases from an initial 609 patients were identified. Patients from the Nanjing Drum Tower Hospital (NJDTH; n=254, 49.9%) were randomly allocated to a training data set (n=203, 79.9%) for model development and internal validation using k-fold cross-validation and a testing set (n=51, 20.1%) for further model validation. External validation was conducted using cohorts from 4 additional institutes, referred to as external cohort 1 (n=92, 18.1%) and external cohort 2 (n=163, 32%).

Clinical Data Collection and Reference Standards

All clinical and laboratory data were extracted from electronic medical records (EMRs) at the beginning of the patients' admission. Demographic characteristics, including age and gender, and clinical signs, such as muscle weakness, the Raynaud phenomenon, mechanic's hands, Gottron sign/papules, heliotrope rash, fever, and arthralgia/arthritis, were documented and compared. Serum levels of muscle enzymes, such as creatine kinase (CK), alanine aminotransferase (ALT), aspartate aminotransferase (AST), and lactate dehydrogenase (LDH), were recorded. Serum antibody abnormalities, including Ro/SSA-52kDa antibody (Ro52) and MDA5 (autoimmune inflammatory myopathies 16 Ag-(IgG); EUROLINE), were analyzed. Disease severity evaluations and inflammatory mediators, such as C-reactive protein (CRP), the erythrocyte sedimentation rate (ESR), the white blood cell (WBC) count, the lymphocyte count, the neutrophil-to-lymphocyte ratio (NLR), and ferritin levels, were recorded and compared. In this study, all these variables were meticulously recorded during hospitalization, with a low incidence of missing data of less than 2%. Missing data were processed with the k-nearest neighbor (KNN) algorithm.

ILD has been defined by the American Thoracic Society/European Respiratory Society [16]. RP-ILD is characterized by the acute and progressive deterioration of ILD, manifesting as worsening dyspnea and interstitial changes on chest HRCT, necessitating hospitalization due to disease-related acute respiratory failure within 3 months of the onset of respiratory symptoms, after exclusion of other potential etiologies (eg, pulmonary infection, heart failure, embolism) [17,18]. The duration of the disease is marked from the initial occurrence of pulmonary, muscle, joint, or skin symptoms/signs.

The primary outcome was the risk of mortality at 3 months. Secondary outcomes included the 1-month mortality risk and the 6-month mortality risk.

Screening of Important Features for Model Construction

We used the BorutaShap algorithm for feature selection, which integrates Boruta and Shapley Additive Explanations (SHAP) algorithms to pinpoint the most influential features within a data set [19]. Consequently, the BorutaShap algorithm was used to sift through 25 parameters, including demographics, clinical features, and laboratory parameters, from the EMR system and identify the most pertinent features for model construction. In this study, only features categorized as "accepted" by the BorutaShap algorithm were incorporated into model development. Features labeled "tentative" or "rejected" were omitted.

Establishment and Validation of Prediction Models

Six ML algorithms were deployed to build prediction models: Extreme Gradient Boosting (XGBoost), logistic regression (LR), Light Gradient Boosting Machine (LightGBM), random forest (RF), support vector machine (SVM), and KNN. To guarantee robustness and reproducibility, all models were provided with identical input features. A thorough search of the hyperparameter space was conducted using both grid and random searches to optimize each model's performance. The reliability of these models was appraised using various evaluation metrics, including the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, the positive predictive value (PPV), the negative predictive value (NPV), accuracy, and the F_1-score. Moreover, 10-fold cross-validations were implemented in the derivation cohort to validate the models.

The optimal model was elucidated using the SHAP methodology, which delineates the importance and contribution of each feature to the model. The model results were interpreted by qualifying the impact of each feature on the predictions. Ultimately, the optimal prediction model was integrated into a user-friendly web-based tool.

Statistical Analysis

Descriptive statistics were applied to summarize continuous variables as mean (SD) values and categorical variables as frequencies and percentage. The Fisher exact test was performed for categorical variables, while the Mann-Whitney *U* test was used for continuous variables to assess differences between the survival and mortality groups within 3 months. Statistical significance was set at $P<.05$. All statistical analyses were performed using SPSS Statistics version 26.0 (IBM Corporation), R Studio version 3.6.3 (R Foundation for Statistical Computing) and Python version 3.7.

Results

Patients' Basic Clinical Characteristics

The study flowchart is shown in Figure 1. The NJDTH cohort had 280 (55%) patients with anti-MDA5+DM-ILD, and after exclusion, the final count was 254 (49.9%) patients. These patients were followed for a median of 6 months, divided into a training cohort of 203 (79.9%) patients and an internal testing cohort of 51 (20.1%) patients. The external validation cohort was further refined to 92 (18.1%) patients in cohort 1 and 163 (32%) patients in external cohort 2 (Figure 1). The entire patient group was analyzed for all-cause mortality. A comparative baseline analysis of 31 clinical variables, encompassing demographics, clinical severity, inflammatory indicators, serum antibody levels, is presented in Table 1. The study evaluated

XSL•FO
RenderX

statistically significant differences in these variables among the groups. Of 254 valid samples, 159 (62.6%) were in the 3-month survival group and 95 (37.4%) in the 3-month death group in the NJDTH cohort. In terms of mortality, 108 (42.5%) patients died during the investigation period, with 71 (27.9%) within 30 days, 95 (37.4%) within 90 days, and 101 (39.78%) within 180 days. The 3-month mortality was up to 37.4%, suggesting the majority of deaths occurred within the first 3 months. Patients who died within 3 months showed more severe injury, as indicated by higher levels of inflammatory markers, such as CRP, ESR, LDH, WBC, ferritin, and NLR, and lower levels of lymphocytes compared to patients who survived beyond 3 months. RP-ILD characteristics were observed in 78 (82.1%) of 95 patients who died within 3 months in contrast to only 30 (18.9%) of 159 survivors. The mean age of patients who died within 3 months was 56.68 (SD 9.97) years, which was higher than that of patients who survived for 3 months (mean 51.50, SD 10.68 years). There were no statistically significant differences in patients' characteristics between the training and testing sets. Other severity indices were comparable between the training and testing sets (Table 2).

Figure 1. Flowchart of patient selection and development of a predictive model. Anti-MDA5+DM-ILD: melanoma differentiation–associated gene 5 antibody–positive dermatomyositis–associated interstitial lung disease; CTD: connective tissue disease.

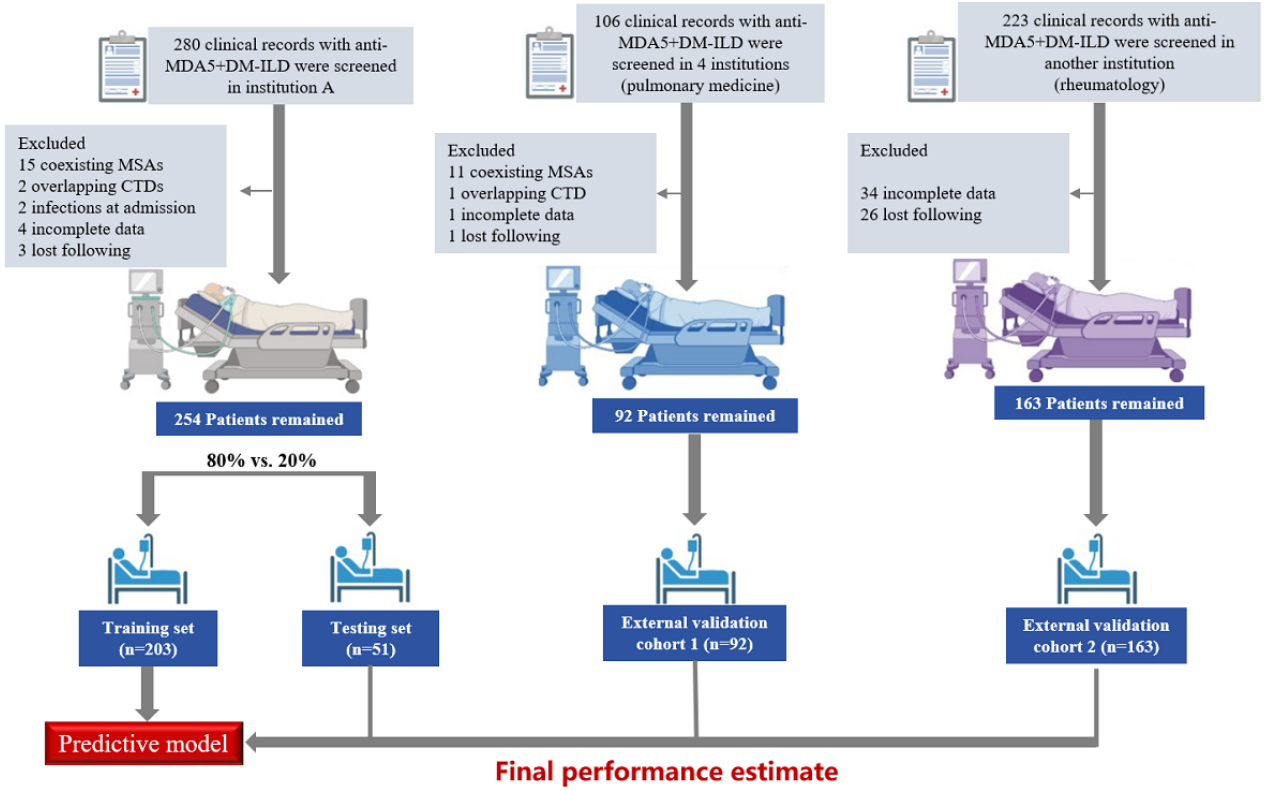

XSL•FO
RenderX

Table 1. Comparison of baseline clinical parameters and all-cause death data of patients from the NJDTH[a] cohort.

Variables	Total NJDTH cohort (N=254)	3-month survival (n=159)	3-month death (n=95)	*P* value
Age (years), mean (SD)	53.44 (10.72)	51.50 (10.68)	56.68 (9.97)	<.001
Gender, n (%)				
Male	77 (30.3)	47 (29.6)	30 (31.6)	.735
Female	177 (69.7)	112 (70.4)	65 (68.4)	N/A[b]
Smoking, n (%)	33 (13.0)	17 (10.7)	16 (16.8)	.158
Clinical features				
Survival time (months)	13.03	28.70	0.50	<.001
Duration (months)	2.00	2.00	1.00	<.001
RP-ILD[c], n (%)	108 (42.5)	30 (18.9)	78 (82.1)	<.001
Muscle weakness, n (%)	63 (24.8)	46 (28.9)	17 (17.9)	.049
Raynaud phenomenon, n (%)	13 (5.1)	9 (5.7)	4 (4.2)	.612
Mechanic's hands, n (%)	74 (29.1)	42 (26.4)	32 (33.7)	.217
Gottron sign/papules, n (%)	167 (65.6)	110 (69.2)	57 (60.0)	.136
Heliotrope rash, n (%)	114 (45.1)	78 (49.4)	36 (37.9)	.076
RF[d], n (%)	19 (7.5)	15 (9.4)	4 (4.2)	.126
ANA[e], n (%)	103 (40.6)	69 (43.4)	34 (35.8)	.232
Fever, n (%)	132 (52.0)	65 (40.9)	67 (70.5)	<.001
Arthralgia/arthritis, n (%)	87 (34.3)	68 (42.8)	19 (20.0)	<.001
Laboratory parameters				
Ro52[f], n (%)	177 (69.7)	105 (66.0)	72 (75.8)	.102
MDA5[g] grade +, n (%)	53 (20.9)	36 (22.6)	17 (17.9)	.665
MDA5 grade ++, n (%)	65 (25.6)	40 (25.2)	25 (26.3)	N/A
MDA5 grade +++, n (%)	136 (53.5)	83 (52.2)	53 (55.8)	N/A
ALB[h] (g/L)	33.57	34.70	31.66	<.001
ALT[i] (U/L)	51.70	44.10	60.80	.021
AST[j] (U/L)	43.30	37.60	56.10	<.001
CK[k] (U/L)	33.00	31.00	41.00	.188
LDH[l] (U/L)	349.00	304.00	439.00	<.001
CRP[m] (mg/L)	6.00	4.70	15.40	<.001
ESR[n] (mm/hour)	39.00	33.00	46.00	<.001
WBC[o] ($\times 10^9$)	6.00	5.50	6.50	.003
Lymphocytes ($\times 10^9$)	0.70	0.80	0.70	.011
NLR[p]	5.50	5.00	8.57	<.001
PaO_2/FiO_2[q]	248.00	328.00	185.00	<.001
Ferritin (ng/mL)	1063.5	698.30	1650.00	<.001

[a]NJDTH: Nanjing Drum Tower Hospital.

[b]N/A: not applicable.

[c]RP-ILD: rapidly progressive interstitial lung disease.

XSL•FO
RenderX

[d]RF: rheumatoid factor.
[e]ANA: antinuclear antibody.
[f]Ro52: Ro/SSA-52kDa antibody.
[g]MDA5: melanoma differentiation–associated gene 5.
[h]ALB: albumin.
[i]ALT: alanine aminotransferase.
[j]AST: aspartate aminotransferase.
[k]CK: creatine kinase.
[l]LDH: lactate dehydrogenase.
[m]CRP: C-reactive protein.
[n]ESR: erythrocyte sedimentation rate.
[o]WBC: white blood cell.
[p]NLR: neutrophil-to-lymphocyte ratio.
[q]PaO_2/FiO_2: ratio of partial pressure of oxygen in arterial blood to fraction of inspired oxygen.

XSL•FO
RenderX

Table 2. Comparison of baseline parameters among training, testing, and validation cohorts.

Variables	Total participants (N=509)	Training cohort (n=203)	Testing cohort (n=51)	External validation cohort 1 (n=92)	External validation cohort 2 (n=163)
Age (years), mean (SD)	53.2 (11.2)	52.9 (11.5)	54.0 (10.2)	52.6 (10.4)	49.3 (11.2)
Gender, n (%)					
Male	170 (33.4)	59 (29.1)	18 (35.3)	40 (43.5)	53 (32.5)
Female	339 (66.6)	144 (70.9)	33 (64.7)	52 (56.5)	110 (67.5)
Smoking, n (%)	87 (17.1)	25 (12.3)	8 (15.7)	17 (18.5)	37 (22.7)
Clinical features					
Duration (months)	2.00	2.00	1.00	2.00	3.00
RP-ILD[a], n (%)	198 (38.9)	88 (43.4)	20 (39.2)	42 (45.7)	48 (29.4)
Heliotrope rash, n (%)	267 (52.6)	90 (44.6)	24 (47.1)	36 (39.1)	117 (71.8)
Gottron sign/papules, n (%)	337 (66.2)	136 (67.0)	31 (60.8)	35 (38.0)	135 (82.8)
Raynaud phenomenon, n (%)	29 (5.7)	13 (6.4)	0	3 (3.3)	13 (8.0)
Mechanic's hands, n (%)	194 (38.1)	61 (30.1)	13 (25.5)	33 (35.9)	87 (53.4)
Muscle weakness, n (%)	165 (32.4)	50 (24.6)	13 (25.5)	26 (28.3)	76 (46.6)
Arthralgia/arthritis, n (%)	183 (36.0)	71 (35.0)	16 (31.4)	39 (42.4)	57 (35.0)
Fever, n (%)	224 (44.0)	108 (53.2)	24 (47.1)	31 (33.7)	61 (37.4)
Skin ulcer, n (%)	90 (19.2)	24 (11.8)	11 (21.6)	7 (13.5)	48 (29.5)
Laboratory parameters					
Ro52[b], n (%)	328 (65.1)	140 (69.0)	37 (72.6)	62 (67.4)	89 (56.3)
RF[c], n (%)	49 (9.8)	14 (6.9)	5 (9.8)	10 (10.9)	20 (13.1)
ANA[d], n (%)	232 (46.2)	80 (39.4)	23 (45.1)	43 (47.3)	86 (54.8)
ALB[e] (g/L)	33.40	33.30	34.10	30.50	34.70
ALT[f] (IU/L)	47.70	53.90	47.70	42.40	44.00
AST[g] (IU/L)	45.50	41.90	50.90	61.00	40.00
CK[h] (IU/L)	40.90	33.00	30.00	87.00	41.00
LDH[i] (IU/L)	335.00	351.00	349.00	392.00	297.00
CRP[j] (mg/L)	5.20	6.00	6.60	6.83	4.01
ESR[k] (mm/hour)	34.00	38.00	42.00	39.00	25.00
WBC[l] ($\times 10^9$)	5.91	5.90	6.20	5.90	5.91
Lymphocytes ($\times 10^9$)	0.71	0.70	0.81	0.69	0.74
NLR[m]	33.40	6.00	5.25	6.94	4.98
PaO_2/FiO_2[n]	309.00	254.00	224.00	248.00	391.91
Ferritin (ng/mL)	804.20	1048.60	1164.30	1351.00	475.10
Primary outcome , n (%)					
Death at 3 months	178 (35.0)	75 (37.0)	20 (39.2)	46 (50.0)	37 (22.7)
Secondary outcomes , n (%)					
Death at 1 month	129 (25.3)	55 (27.1)	16 (31.4)	26 (28.3)	32 (19.6)
Death at 6 months	191 (37.5)	80 (39.4)	21 (41.2)	50 (54.4)	40 (24.5)
All deaths	202 (39.7)	87 (42.9)	21 (41.2)	52 (56.5)	42 (25.8)

XSL•FO
RenderX

[a]RP-ILD: rapidly progressive interstitial lung disease.
[b]Ro52: Ro/SSA-52kDa antibody.
[c]RF: rheumatoid factor.
[d]ANA: antinuclear antibody.
[e]ALB: albumin.
[f]ALT: alanine aminotransferase.
[g]AST: aspartate aminotransferase.
[h]CK: creatine kinase.
[i]LDH: lactate dehydrogenase.
[j]CRP: C-reactive protein.
[k]ESR: erythrocyte sedimentation rate.
[l]WBC: white blood cell.
[m]NLR: neutrophil-to-lymphocyte ratio.
[n]PaO_2/FiO_2: ratio of partial pressure of oxygen in arterial blood to fraction of inspired oxygen.

To account for varying severities of with anti-MDA5+DM-ILD patients, we recruited external cohort 1 mainly from the pulmonary medicine department and external cohort 2 from the rheumatology department. Patients in external cohort 1 tended to have severe illness than those in external cohort 2, as patients with respiratory failure were more likely to be admitted to the pulmonary medicine department. Analysis on a diverse range of severities of patients enhanced our model's practicality and robustness.

Screening of Characteristic Factors for Risk of 3-Month Mortality

In total, 9 parameters (ie, RP-ILD, age, LDH, albumin [ALB], CRP, ESR, AST, NLR, and duration of disease) were selected for analysis using the BorutaShap algorithm (Figure 2). However, due to the overlap between disease duration and RP-ILD, RP-ILD was left for model development to balance model complexity and generalizability. Hence, 8 parameters were selected for further analysis.

XSL•FO
RenderX

Figure 2. By using the BorutaShap algorithm, selected parameters or patients' characteristics were calculated for the importance of evaluation. Parameters with green color were categorized as "accepted" for the next step of analysis ALB: albumin; ALT: alanine aminotransferase; ANA: antinuclear antibody; AST: aspartate aminotransferase; CK: creatine kinase; CRP: C-reactive protein; ESR: erythrocyte sedimentation rate; LDH: lactate dehydrogenase; MDA5: melanoma differentiation–associated gene 5; NLR: neutrophil-to-lymphocyte ratio; RF: rheumatoid factor; RP-ILD: rapidly progressive interstitial lung disease; SHAP: Shapley Additive Explanations; WBC: white blood cell.

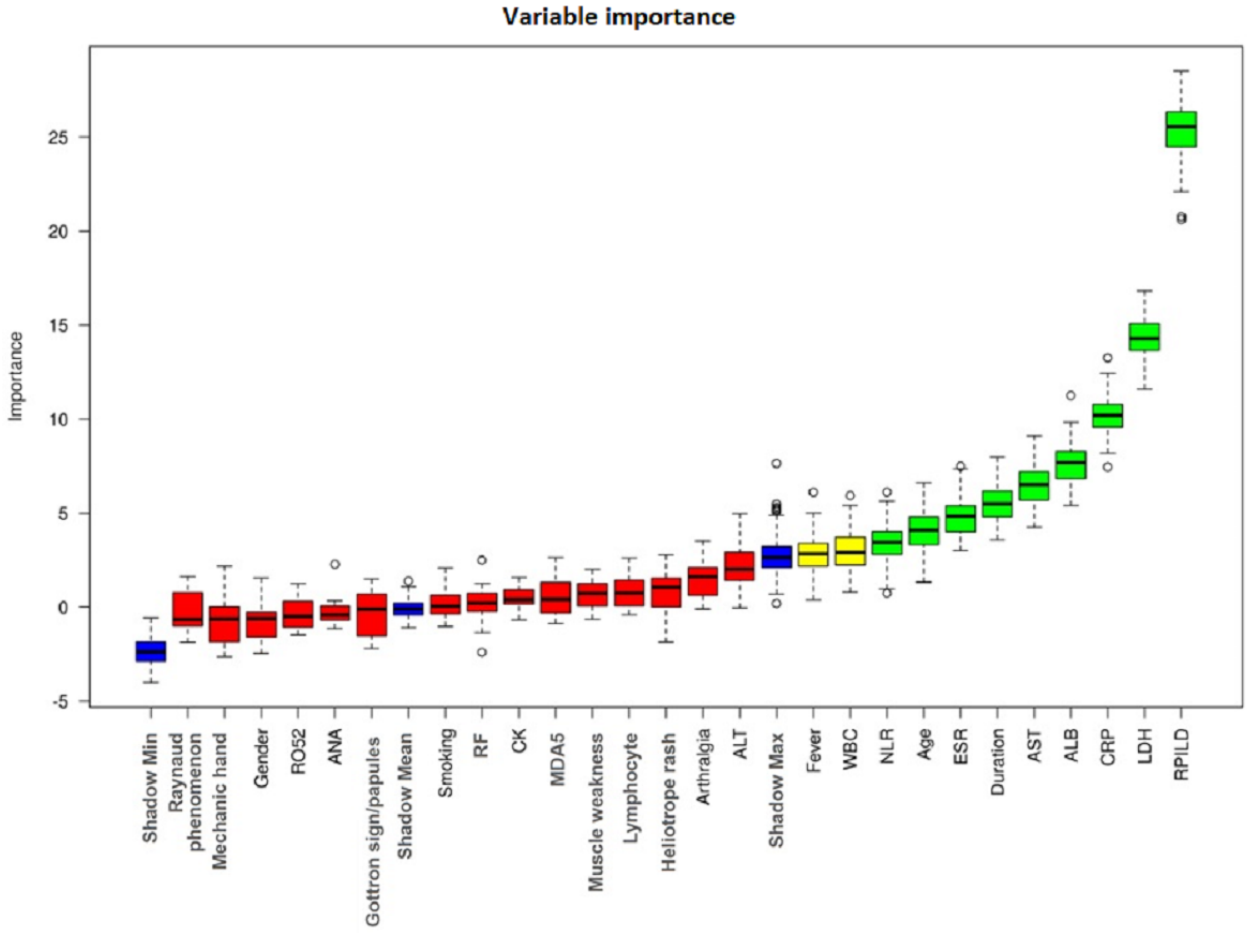

Comprehensive Analysis of the Prediction Model and Selection of the Optimal Model

To identify the best prediction model for the 3-month mortality risk in patients with anti-MDA5+DM-ILD, all models were evaluated using identical input features and assessed via AUC values, decision curve analysis (DCA), calibration curves, and precision-recall (PR) curves. XGBoost, LightGBM, and RF algorithms showed the highest AUCs in the training data set, but LR demonstrated the highest AUC of 0.877 (95% CI 0.744-0.985) in the validation data set (Figure 3A,B). Tables S1 and S2 in Multimedia Appendix 1 provide more details, indicating potential overfitting in the XGBoost, LightGBM, and RF algorithms. Furthermore, the LR model achieved a significantly lower Brier score of 0.145 compared to other models. The DCA curves in Figure 3C and the calibration plots in Figure 3D demonstrate that all 6 models provided a net clinical benefit over the full treat-all or treat-none strategies, with the LR model outperforming the other 5 models in the validation data set. Additionally, the LR model obtained the highest average precision (AP) value of 0.813 on the validation data set (Figure 3E,F). Based on these findings, the LR model was selected as the final model due to its superior performance.

XSL•FO
RenderX

Figure 3. Comparison of prediction models for performance using 6 ML algorithms. ROC results for predicting 3-month mortality obtained by using different models for the (A) training cohort and (B) validation cohort. (C) DCA, (D), calibration curves, and (E, F) PR curves were used to verify which model was preferable. DCA: decision curve analysis; KNN: k-nearest neighbor; LightGBM: Light Gradient Boosting Machine; LR: logistic regression; ML: machine learning; PR: precision-recall; RF: rheumatoid factor; ROC: receiver operating characteristic; SVM: support vector machine; XGBoost: Extreme Gradient Boosting.

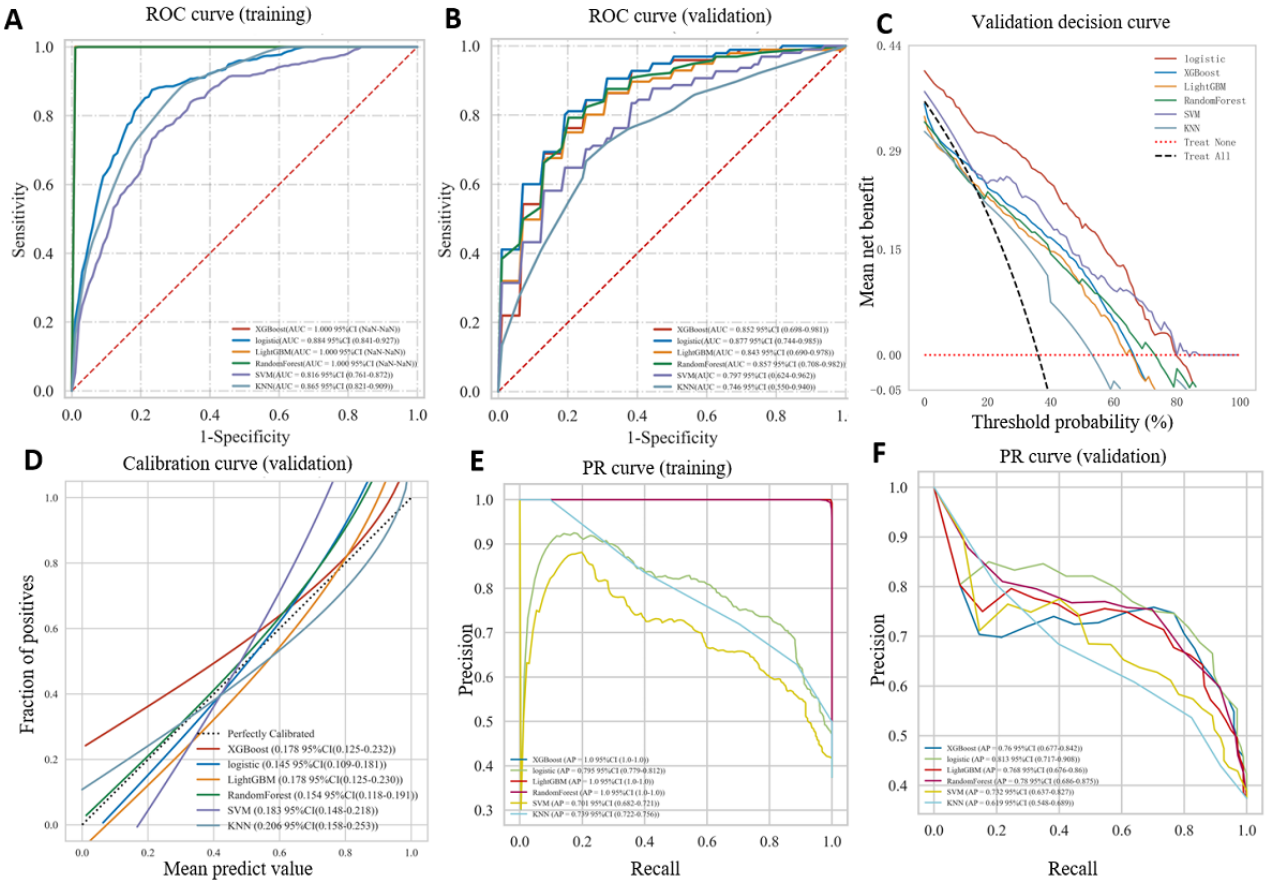

Ultimately, we selected RP-ILD, age, LDH, AST, ALB, NLR, CRP, and ESR as input features for the LR model. As shown in Figure 4 and Table S3 in Multimedia Appendix 1, the LR model demonstrated optimal performance with an AUC, sensitivity, specificity, and accuracy of 0.86, 84.8%, 84.4%, and 79.3%, respectively, on the validation data set and 0.90, 85.0%, 83.9%, and 84.3%, respectively, on the training data set (Figure 4A-D). The calibration of the LR model indicated strong concordance between the predicted probabilities and the actual observed 3-month deaths using Brier scores (Figure 4E). DCA revealed that the model provided greater net benefits over both treat-all and treat-none strategies (Figure 4F).

XSL•FO
RenderX

Figure 4. Performance of the LR model. (A) Training data set ROC and AUC and (B) internal cross-validation set ROC and AUC. The red dashed line represents the training data set, and the blue dashed line represents the validation data set. Values are expressed as averages (95% CI). Solid lines of different colors represent 10 different results. (C) Training data set ROC and AUC. Test results for patients. (D) Learning curve in training and validation data set, (E) calibration curves, and (F) DCA. AUC: area under the receiver operating characteristic curve; DCA: decision curve analysis; LR: logistic regression; ROC: receiver operating characteristic.

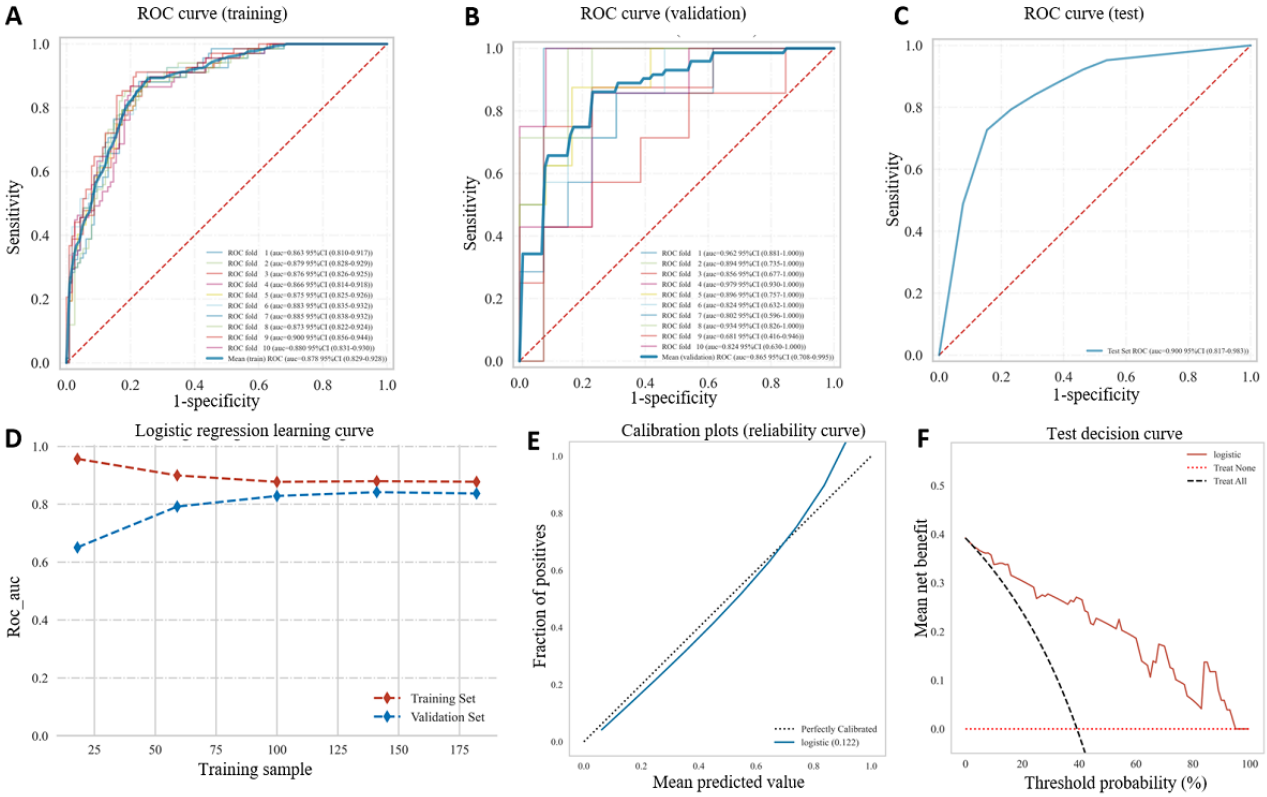

SHAP Model Interpretation and Individual Analysis

To provide a clear explanation of the selected variables, we used SHAP to illustrate how these variables predict the 3-month mortality within the model. Figure 5A displays the 8 most influential features (RP-ILD, ESR, ALB, age, CRP, AST, NLR, and LDH) in our model. Each feature's importance line is represented by colored dots, where red dots indicate high-risk values and blue dots signify low-risk values. Figure 5B ranks these 8 risk factors based on their average absolute SHAP values, with the x axis representing SHAP values that denote the prediction model's importance. Additionally, in this model, the probability threshold for the 3-month mortality risk was established at 0.40473. We then provided 2 exemplary cases to illustrate the model's interpretability. In one instance, a patient presenting with RP-ILD exhibited a notably elevated SHAP-predicted mortality risk of 0.718 (Figure 5C), and unfortunately, this individual succumbed to the disease within a 3-month period postonset. In contrast, a 64-year-old participant demonstrated a substantially lower SHAP-predicted mortality risk of 0.105 (Figure 5D) and remained alive, underscoring the potential utility of the web-based prognostic tool in clinical practice.

XSL•FO
RenderX

Figure 5. SHAP interpretation of the model. (A) Attributes of characteristics in the SHAP chart, (B) mean SHAP value of 8 important variables, (C, D) 2 clinical cases illustrating the usage of a weighing model. ALB: albumin; AST: aspartate aminotransferase; CRP: C-reactive protein; ESR: erythrocyte sedimentation rate; LDH: lactate dehydrogenase; NLR: neutrophil-to-lymphocyte ratio; RP-ILD: rapidly progressive interstitial lung disease; SHAP: Shapley Additive Explanations.

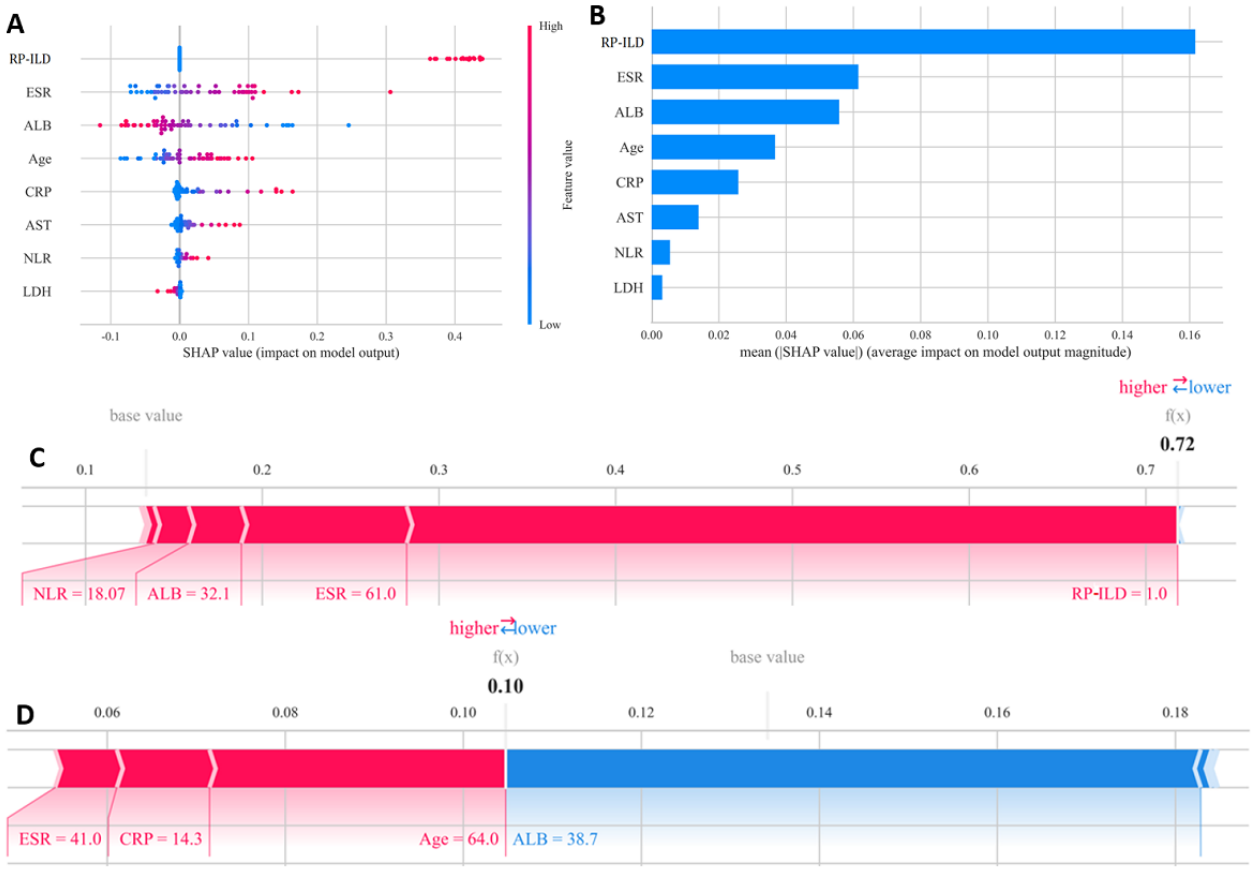

Model External Validation and Establishment of a Web-Based Calculator

The AUC of our proposed model on external cohort 1 from 4 hospitals in China was 0.836 (Figure 6). The sensitivity, specificity, and accuracy for this cohort were 0.87, 0.739, and 0.783, respectively. For external cohort 2, the AUC, sensitivity, specificity, and accuracy were 0.915, 0.892, 0.865, and 0.847, respectively, indicating the model's robust extrapolation capabilities. In addition, secondary outcomes (1-month and 6-month mortality risks) were evaluated in the 2 external validation cohorts. The model's superior discrimination performance, using the AUC, sensitivity, specificity, precision, accuracy, and the F_1-score, is detailed in Tables S4 and S5 in Multimedia Appendix 1, Figure S1 in Multimedia Appendix 2, and Figure S2 in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 6. AUC and ROC of the LR model in the 2 external validation cohorts. AUC: area under the receiver operating characteristic curve; LR: logistic regression; ROC: receiver operating characteristic.

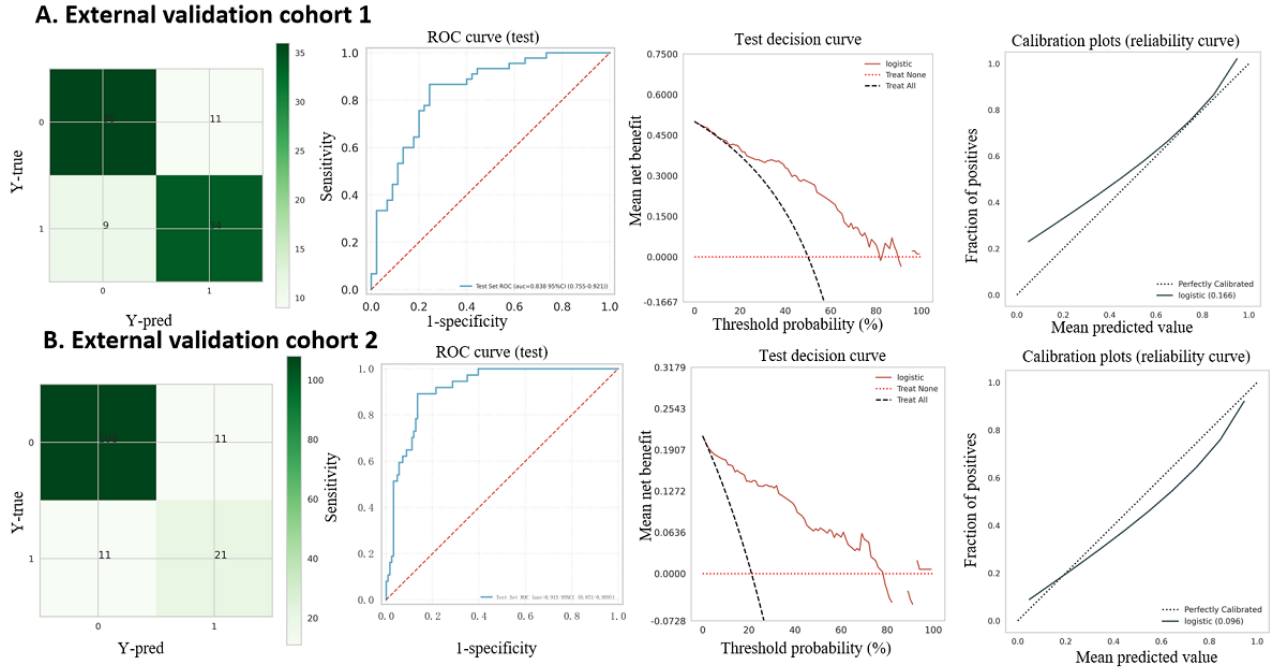

To enhance generalizability and facilitate clinical application, the optimal ML model was adapted to create a web-based calculator [20]. The calculator serves as a practical clinical tool for physicians specializing in respiratory medicine, rheumatology, and critical care medicine. It is particularly useful in the assessment of patients presenting with RP-ILD attributed to anti-MDA5+DM-ILD. A score exceeding 40.47% indicates a heightened risk of mortality, warranting consideration for aggressive management strategies.

Discussion

Principal Findings

To the best of our knowledge, this is the largest retrospective multicenter study to investigate the risk factors for 3-month mortality and to explore 6 ML models for comprehensive predictive analyses on anti-MDA5+DM-ILD cohorts in China. We identified a set of predictive risk factors and constructed a prediction model for patients with anti-MDA5+DM-ILD using ML algorithms in conjunction with clinical and laboratory data easily extracted from the EMR system. In this study, LR, using routinely collected data from real-world clinical practice, demonstrated superior performance across 8 predictive measures. The optimal model exhibited favorable discrimination and calibration, while also emphasizing explainability to enhance trust and transparency. Furthermore, variable importance was investigated at both individual and population levels. Additionally, external validation confirmed the superior predictive performance of this optimal model. To facilitate clinical application, a web-based tool was developed for health care professionals to use the proposed prediction model.

Previous studies have consistently reported high mortality rates exceeding 60% among patients with anti-MDA5+DM-ILD, particularly in East Asian populations, such as China [21]. The majority of deaths occurred within the first 3 months [22]. It is widely recognized that an excessive inflammatory response is associated with a heightened risk of RP-ILD and all-cause mortality [23]. Hyperinflammation plays a crucial role in the development and progression of anti-MDA5+DM-ILD. Uncontrolled hyperinflammation and persistent immune activation contribute to severe lung injury and RP-ILD, which may resemble the phenomenon known as “cytokine storm” observed in COVID-19 [24]. Previous studies have indicated that poor outcomes in patients with DM are correlated with certain factors, including advanced age and elevated levels of serum ferritin, LDH, NLR, and CRP [21]. Consistent with these findings, we established a prediction model incorporating 8 key variables of significant predictive importance: RP-ILD, ESR, ALB, age, CRP, AST, NLR, and LDH. Most of these variables are inflammatory markers that can reflect the hyperinflammatory state and were easy to obtain and observe in previous studies on patients with DM in a clinical setting. Moreover, classical statistical analysis has identified risk factors in these patient populations, including RP-ILD and age [21,25,26]. In line with previous findings, RP-ILD was found to be the most important feature according to the SHAP value and contributed to this prediction model. Age was ranked fourth in terms of its significance as a parameter associated with worsened outcomes. Therefore, this prediction model demonstrated superior performance, yielding AUCs of 0.90 in the internal cohorts and 0.832-0.919 in external cohorts.

There are currently no existing ML-based prediction models for 3-month mortality in individuals with anti-MDA5+DM-ILD, making it impossible to make a direct comparison with prior work. Previous studies have used logistic or Cox proportional hazard models as alternatives [10,27-29]. For instance, a nomogram model incorporating the duration of first-time symptoms, the presence of fever, pleural effusion (PE), total CT scores, and AST has been recommended as a prognostic predictor for anti-MDA5+DM-ILD outcomes [29]. Another

XSL•FO
RenderX

model known as the "FLAIR score," which combines ferritin, LDH, the semiquantitative anti-MDA5 grade, the HRCT imaging score, and RP-ILD/non-RP-ILD classification, was proposed to predict mortality in amyopathic DM-related ILD based on a large-scale Chinese single-center cohort study [10]. In contrast to our study, which specifically targeted patients with anti-MDA5+DM-ILD, Lian et al's [10] investigation encompassed a more expansive patient population, not restricting their analysis solely to anti-MDA5+DM-ILD cases. Furthermore, although their investigation incorporated HRCT findings for scoring purposes, our study deliberately excluded HRCT as a discriminating factor in the evaluation process, attributable to the recognized interobserver and intraobserver variability inherent in HRCT interpretation.

Ouyang et al [30] reported a novel matrix prediction model that includes baseline characteristics, such as fever, ferritin≥1250 μg/L, and positive carcinoma embryonic antigen, which can predict 6-month all-cause mortality in patients with anti-MDA5+DM-ILD. However, these prediction models lack thorough external validation due to their reliance on small and nonrepresentative populations (eg, individuals from a single hospital), primarily due to the rarity of the disease. Moreover, the HRCT imaging score and the radiomics-based prediction model [31,32] are overly complex and impractical for routine clinical use. In contrast, this study leveraged routinely collected data from real-world clinical practices in China to develop an optimal prediction model, which is now accessible as a web-based tool. This tool enables clinicians to calculate individualized risks of 3-month mortality based on easily obtainable clinical data, including RP-ILD, age, AST, CRP, NLR, LDH, ALB, and NLR. It is important to emphasize that in real-world clinical practice, such models should be used holistically, considering all key features, rather than relying on a single feature for risk prediction. Therefore, the models require the input of a comprehensive set of features to accurately assess the risk in patients with anti-MDA5+DM-ILD.

Strengths and Limitations

Our study has several notable strengths. First, we conducted external validation in independent cohorts from 4 hospitals in China, encompassing various regions and settings. This inclusion of external cohorts 1 and 2 in China allows our findings to be representative of real-life clinical practice in our country. Second, we extracted, analyzed, and used data from routine clinical practice and finally established a practicable web-based algorithm for predicting 3-month mortality for patients with anti-MDA5+DM-ILD. Additionally, our study included a substantial number of patients with anti-MDA5+DM-ILD, further enhancing the robustness of our results.

However, it is important to acknowledge some limitations of our study. First, it was a retrospective study, which inherently introduces certain biases. Our patient population was sourced from large tertiary referral hospitals and consisted only of hospitalized patients, potentially leading to the inclusion of individuals with more severe disease. Second, although the predicted forced vital capacity (FVC<50%) was shown to independently predict 6-month all-cause mortality in patients with anti-MDA5-positive DM [29], we were unable to incorporate pulmonary function parameters into our prediction model due to missing data, particularly in patients with severe disease who had no opportunity to finish the pulmonary function test. Lastly, although the model demonstrated good discrimination and calibration in our study, the external validation cohort was limited to a Chinese population. It is necessary to develop a multiregion and multirace worldwide analysis and generate a more comprehensive and classical system to predict 3-month mortality for patients with anti-MDA5+DM-ILD.

Conclusion

In summary, our use of ML techniques and routinely collected clinical features helped us successfully develop an LR model that reliably predicts the risk of 3-month mortality in patients with anti-MDA5+DM-ILD. To facilitate the translation of our findings into clinical practice, we created a web-based tool that allows clinicians to calculate individualized risks of 3-month mortality. This tool can guide risk-stratified approaches to patient care and treatment.

Acknowledgments

We thank Xtreme Smart Analysis for technical support and analytical and visualization support provided for our research. This study was supported by Jiangsu Provincial Key Research and Development (no BE2016611) and the National Natural Science Foundation of China (grant 81970063).

Data Availability

Due to patient privacy, the data related to patients is not available for public access. However, it can be obtained from the corresponding author upon reasonable request, subject to approval by the institutional review board of the enrolled centers.

Authors' Contributions

HL wrote the draft and interpreted the data regarding the anti-melanoma differentiation–associated gene 5–positive dermatomyositis complicated with interstitial lung disease; RYZ, HXX, PH, and BX collected crude data; YQT, QZ, XY, XHQ, YJG, YL, MC, and HYJ managed patients in the capital institution; BC, QH, JC, GCW, and HRC conceived the study; and HRC analyzed the data and strictly revised the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Performance metrics for 6 models and discrimination performance of death at different times.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v27i1e62836_app1.xlsx]

Multimedia Appendix 2
Secondary outcomes for external cohort 1.
[PNG File , 177 KB - jmir_v27i1e62836_app2.png]

Multimedia Appendix 3
Secondary outcomes for external cohort 2.
[PNG File , 199 KB - jmir_v27i1e62836_app3.png]

References

1. Lega J, Reynaud Q, Belot A, Fabien N, Durieu I, Cottin V. Idiopathic inflammatory myopathies and the lung. Eur Respir Rev 2015 Jun 31;24(136):216-238 [FREE Full text] [doi: 10.1183/16000617.00002015] [Medline: 26028634]
2. Palterer B, Vitiello G, Carraresi A, Giudizi MG, Cammelli D, Parronchi P. Bench to bedside review of myositis autoantibodies. Clin Mol Allergy 2018 Mar 7;16(1):5 [FREE Full text] [doi: 10.1186/s12948-018-0084-9] [Medline: 29540998]
3. Lu X, Peng Q, Wang G. Anti-MDA5 antibody-positive dermatomyositis: pathogenesis and clinical progress. Nat Rev Rheumatol 2024 Jan 06;20(1):48-62. [doi: 10.1038/s41584-023-01054-9] [Medline: 38057474]
4. David P, Sinha S, Iqbal K, De Marco G, Taheri S, McLaren E, et al. MDA5-autoimmunity and interstitial pneumonitis contemporaneous with the COVID-19 pandemic (MIP-C). EBioMedicine 2024 Jun;104:105136 [FREE Full text] [doi: 10.1016/j.ebiom.2024.105136] [Medline: 38723554]
5. McHugh NJ, Tansley SL. Autoantibodies in myositis. Nat Rev Rheumatol 2018 Apr 20;14(5):290-302. [doi: 10.1038/nrrheum.2018.56] [Medline: 29674612]
6. You H, Wang L, Wang J, Lv C, Xu L, Yuan F, et al. Time-dependent changes in RPILD and mortality risk in anti-MDA5+ DM patients: a cohort study of 272 cases in China. Rheumatology (Oxford) 2023 Mar 01;62(3):1216-1226. [doi: 10.1093/rheumatology/keac450] [Medline: 35961045]
7. Tsuji H, Nakashima R, Hosono Y, Imura Y, Yagita M, Yoshifuji H, et al. Multicenter prospective study of the efficacy and safety of combined immunosuppressive therapy with high-dose glucocorticoid, tacrolimus, and cyclophosphamide in interstitial lung diseases accompanied by anti-melanoma differentiation-associated gene 5-positive dermatomyositis. Arthritis Rheumatol 2020 Mar 27;72(3):488-498. [doi: 10.1002/art.41105] [Medline: 31524333]
8. Matsuda KM, Yoshizaki A, Kuzumi A, Fukasawa T, Ebata S, Yoshizaki-Ogawa A, et al. Combined immunosuppressive therapy provides favorable prognosis and increased risk of cytomegalovirus reactivation in anti-melanoma differentiation-associated gene 5 antibody-positive dermatomyositis. J Dermatol 2020 May 24;47(5):483-489. [doi: 10.1111/1346-8138.15274] [Medline: 32096271]
9. Matsushita T, Mizumaki K, Kano M, Yagi N, Tennichi M, Takeuchi A, et al. Antimelanoma differentiation-associated protein 5 antibody level is a novel tool for monitoring disease activity in rapidly progressive interstitial lung disease with dermatomyositis. Br J Dermatol 2017 Feb 19;176(2):395-402. [doi: 10.1111/bjd.14882] [Medline: 27452897]
10. Lian X, Zou J, Guo Q, Chen S, Lu L, Wang R, et al. Mortality risk prediction in amyopathic dermatomyositis associated with interstitial lung disease: the FLAIR model. Chest 2020 Oct;158(4):1535-1545 [FREE Full text] [doi: 10.1016/j.chest.2020.04.057] [Medline: 32428508]
11. Horiike Y, Suzuki Y, Fujisawa T, Yasui H, Karayama M, Hozumi H, et al. Successful classification of macrophage-mannose receptor CD206 in severity of anti-MDA5 antibody positive dermatomyositis associated ILD. Rheumatology (Oxford) 2019 Dec 01;58(12):2143-2152. [doi: 10.1093/rheumatology/kez185] [Medline: 31143953]
12. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021 Jul 09;11(7):e048008 [FREE Full text] [doi: 10.1136/bmjopen-2020-048008] [Medline: 34244270]
13. Bohan A, Peter JB. Polymyositis and dermatomyositis. N Engl J Med 1975 Feb 13;292(7):344-347. [doi: 10.1056/nejm197502132920706]
14. Lundberg IE, Bottai M, Tjärnlund A. Response to: 'performance of the 2017 European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies in clinical practice' by Hočevar. Ann Rheum Dis 2018 Dec 09;77(12):e91-e91. [doi: 10.1136/annrheumdis-2017-212786] [Medline: 29317406]

XSL•FO
RenderX

15. Mammen AL, Allenbach Y, Stenzel W, Benveniste O, ENMC 239th Workshop Study Group. 239th ENMC International Workshop: classification of dermatomyositis, Amsterdam, the Netherlands, 14-16 December 2018. Neuromuscul Disord 2020 Jan;30(1):70-92. [doi: 10.1016/j.nmd.2019.10.005] [Medline: 31791867]
16. Travis WD, Costabel U, Hansell DM, King TE, Lynch DA, Nicholson AG, ATS/ERS Committee on Idiopathic Interstitial Pneumonias. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 2013 Sep 15;188(6):733-748 [FREE Full text] [doi: 10.1164/rccm.201308-1483ST] [Medline: 24032382]
17. Moghadam-Kia S, Oddis CV, Sato S, Kuwana M, Aggarwal R. Anti-melanoma differentiation-associated gene 5 is associated with rapidly progressive lung disease and poor survival in US patients with amyopathic and myopathic dermatomyositis. Arthritis Care Res (Hoboken) 2016 May 27;68(5):689-694 [FREE Full text] [doi: 10.1002/acr.22728] [Medline: 26414240]
18. Gono T, Kawaguchi Y, Satoh T, Kuwana M, Katsumata Y, Takagi K, et al. Clinical manifestation and prognostic factor in anti-melanoma differentiation-associated gene 5 antibody-associated interstitial lung disease as a complication of dermatomyositis. Rheumatology (Oxford) 2010 Sep 23;49(9):1713-1719. [doi: 10.1093/rheumatology/keq149] [Medline: 20498012]
19. Keany E. BorutaShap: a wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values. Zenodo. 2020 Nov 5. URL: https://zenodo.org/records/4247618 [accessed 2024-12-12]
20. Predictive model for 3-month mortality risk in anti-MDA5+DM-ILD. Department of Respiratory Medicine, Nanjing Drum Tower Hospital. URL: https://www.xsmartanalysis.com/model/list/predict/model/html?mid=8988&symbol=61nx6gI9797375mB74LX [accessed 2024-12-12]
21. Yang B, Liu S, Qian Z, Tong Z. Predicting the death of patients with anti-melanoma differentiation-associated protein-5-positive dermatomyositis-associated interstitial lung disease: a systematic review and meta-analysis. Mod Rheumatol 2024 Mar 28;34(3):541-550. [doi: 10.1093/mr/road042] [Medline: 37364274]
22. Allenbach Y, Uzunhan Y, Toquet S, Leroux G, Gallay L, Marquet A, et al. Different phenotypes in dermatomyositis associated with anti-MDA5 antibody. Neurology 2020 Jul 07;95(1):e70-e78. [doi: 10.1212/wnl.0000000000009727]
23. Lv C, You H, Xu L, Wang L, Yuan F, Li J, et al. Coexistence of anti-Ro52 antibodies in anti-MDA5 antibody-positive dermatomyositis is highly associated with rapidly progressive interstitial lung disease and mortality risk. J Rheumatol 2023 Feb 15;50(2):219-226. [doi: 10.3899/jrheum.220139] [Medline: 35705235]
24. Yongzhi X. COVID-19-associated cytokine storm syndrome and diagnostic principles: an old and new issue. Emerg Microbes Infect 2021 Dec 18;10(1):266-276 [FREE Full text] [doi: 10.1080/22221751.2021.1884503] [Medline: 33522893]
25. Wang H, Chen X, Du Y, Wang L, Wang Q, Wu H, et al. Mortality risk in patients with anti-MDA5 dermatomyositis is related to rapidly progressive interstitial lung disease and anti-Ro52 antibody. Arthritis Res Ther 2023 Jul 24;25(1):127 [FREE Full text] [doi: 10.1186/s13075-023-03100-z] [Medline: 37488657]
26. Xie H, Zhang D, Wang Y, Shi Y, Yuan Y, Wang L, et al. Risk factors for mortality in patients with anti-MDA5 antibody-positive dermatomyositis: a meta-analysis and systematic review. Semin Arthritis Rheum 2023 Oct;62:152231 [FREE Full text] [doi: 10.1016/j.semarthrit.2023.152231] [Medline: 37348186]
27. Gono T, Masui K, Nishina N, Kawaguchi Y, Kawakami A, Ikeda K, the Multicenter Retrospective Cohort of Japanese Patients with Myositis-Associated ILD (JAMI) Investigators. Risk prediction modeling based on a combination of initial serum biomarker levels in polymyositis/dermatomyositis-associated interstitial lung disease. Arthritis Rheumatol 2021 Apr 22;73(4):677-686. [doi: 10.1002/art.41566] [Medline: 33118321]
28. Li Y, Li Y, Wang Y, Shi L, Lin F, Zhang Z, et al. A clinical risk model to predict rapidly progressive interstitial lung disease incidence in dermatomyositis. Front Med (Lausanne) 2021 Sep 27;8:733599 [FREE Full text] [doi: 10.3389/fmed.2021.733599] [Medline: 34646845]
29. Niu Q, Zhao L, Ma W, Xiong L, Wang X, He X, et al. A new predictive model for the prognosis of MDA5 DM-ILD. Front Med (Lausanne) 2022 Jun 15;9:908365 [FREE Full text] [doi: 10.3389/fmed.2022.908365] [Medline: 35783655]
30. Ouyang Z, Lin J, Tang A, Yang Z, Yang L, Wei X, et al. A matrix prediction model for the 6-month mortality risk in patients with anti-melanoma differentiation-associated protein-5-positive dermatomyositis. Front Med (Lausanne) 2022 Apr 1;9:860798 [FREE Full text] [doi: 10.3389/fmed.2022.860798] [Medline: 35433730]
31. Xu W, Wu W, Zhang D, Chen Z, Tao X, Zhao J, et al. A novel CT scoring method predicts the prognosis of interstitial lung disease associated with anti-MDA5 positive dermatomyositis. Sci Rep 2021 Aug 23;11(1):17070 [FREE Full text] [doi: 10.1038/s41598-021-96292-w] [Medline: 34426622]
32. Li Y, Deng W, Zhou Y, Luo Y, Wu Y, Wen J, et al. A nomogram based on clinical factors and CT radiomics for predicting anti-MDA5+ DM complicated by RP-ILD. Rheumatology (Oxford) 2024 Mar 01;63(3):809-816. [doi: 10.1093/rheumatology/kead263] [Medline: 37267146]

Abbreviations

anti-MDA5+DM-ILD: antimelanoma differentiation–associated gene 5 antibody–positive dermatomyositis–associated interstitial lung disease
ALB: albumin

XSL•FO
RenderX

ALT: alanine aminotransferase
ANA: antinuclear antibody
AST: aspartate aminotransferase
AUC: area under the receiver operating characteristic curve
CK: creatine kinase
CRP: C-reactive protein
CT: computed tomography
CTD: connective tissue disease
DCA: decision curve analysis
DM: dermatomyositis
EMR: electronic medical record
ESR: erythrocyte sedimentation rate
HRCT: high-resolution computed tomography
IIM: idiopathic inflammatory myopathies
IL: interleukin
ILD: interstitial lung disease
KL-6: Krebs von den Lungen-6
KNN: k-nearest neighbor
LDH: lactate dehydrogenase
LightGBM: Light Gradient Boosting Machine
LR: logistic regression
MDA5: melanoma differentiation–associated gene 5
ML: machine learning
MSA: myositis-specific antibody
NJDTH: Nanjing Drum Tower Hospital
NLR: neutrophil-to-lymphocyte ratio
PR: precision-recall
RF: rheumatoid factor
ROC: receiver operating characteristic
Ro52: Ro/SSA-52kDa antibody
RP-ILD: rapidly progressive interstitial lung disease
SHAP: Shapley Additive Explanations
SVM: support vector machine
WBC: white blood cell
XGBoost: Extreme Gradient Boosting

Edited by X Ma; submitted 02.06.24; peer-reviewed by M Wu, S Ye, X Tao; comments to author 06.09.24; revised version received 08.10.24; accepted 22.11.24; published 05.02.25.

Please cite as:
Li H, Zou R, Xin H, He P, Xi B, Tian Y, Zhao Q, Yan X, Qiu X, Gao Y, Liu Y, Cao M, Chen B, Han Q, Chen J, Wang G, Cai H
Mortality Risk Prediction in Patients With Antimelanoma Differentiation–Associated, Gene 5 Antibody–Positive, Dermatomyositis–Associated Interstitial Lung Disease: Algorithm Development and Validation
J Med Internet Res 2025;27:e62836
URL: https://www.jmir.org/2025/1/e62836
doi:10.2196/62836
PMID:

©Hui Li, Ruyi Zou, Hongxia Xin, Ping He, Bin Xi, Yaqiong Tian, Qi Zhao, Xin Yan, Xiaohua Qiu, Yujuan Gao, Yin Liu, Min Cao, Bi Chen, Qian Han, Juan Chen, Guochun Wang, Hourong Cai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Implementation of Medication-Related Technology and Its Impact on Pharmacy Workflow: Real-World Evidence Usability Study

Wei-Ning Yu[1,2*], MS; Yih-Dih Cheng[1,2*], PhD; Yu-Chi Hou[2], PhD; Yow-Wen Hsieh[1,2*], PhD

[1]Department of Pharmacy, China Medical University Hospital, Taichung City, Taiwan

[2]College of Pharmacy, School of Pharmacy, China Medical University, Taichung City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yih-Dih Cheng, PhD
Department of Pharmacy
China Medical University Hospital
No 2, Yude Rd, North Dist
Taichung City, 404327
Taiwan
Phone: 886 422052121 ext 12272
Fax: 886 422070583
Email: tovis168@gmail.com

Abstract

Background: Medication errors constitute a major contributor to patient harm, driving up health care costs and representing a preventable cause of medical incidents. Over the past decade, many hospitals have integrated various medication-related technologies into their pharmacy operations. However, real-world evidence of the impact of these advanced systems on clinical prescription dispensing error rates remains limited.

Objective: This study aims to prospectively detect and record the categories and rates of dispensing errors to illustrate how medication-related technologies, such as automated dispensing cabinet (ADC), barcode medication administration (BCMA), and smart dispensing counter (SDC), can be used to minimize dispensing errors.

Methods: This study used a before-and-after design at a 2202-bed academic medical center in Taiwan to assess the impact of implementing medication-related technologies (ADC, BCMA, and SDC) on patient medication safety. Dispensing error rates were analyzed from January 1, 2017, to December 31, 2023, using data from the China Medical University Hospital Patient Safety Database. The study periods were defined as stage 0 (preintervention, January to November 2017), stage 1 (post-ADC intervention, December 2017 to June 2018), stage 2 (post-BCMA intervention, July 2018 to October 2020), and stage 3 (post-SDC intervention, November 2020 to December 2023). Medication errors were defined according to the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). Chi-square or Fisher exact tests were used to analyze differences between intervention periods, with Bonferroni correction for multiple comparisons. Statistical significance was set at $P<.05$.

Results: Following the introduction of medication-related technologies, the average dispensing error incidence rate significantly decreased by 39.68%, 44.44%, and 77.78%, from 0.0063% in stage 0 to 0.0038%, 0.0035%, and 0.0014% in stages 1, 2, and 3, respectively ($P<.001$). The frequency of “wrong drug” errors, the most common error type in stage 0, significantly decreased by 51.15%, 56.85%, and 81.26% in stages 1, 2, and 3, respectively. All error types, except for “wrong dosage form,” “wrong strength,” “wrong time,” and “others,” demonstrated statistically significant differences ($P<.001$). The majority of harm severities were categorized as “A” (no error; 97%-98.8%) and “B-D” (error, no harm; 1.2%-3%) according to the NCC MERP classification. The severity of “no error” (category A) significantly decreased at each stage ($P<.001$). Statistically significant differences in dispensing error rates were observed between all stages ($P<.001$), except between stages 2 and 1 ($P>.99$).

Conclusions: This study provides significant evidence that the implementation of medication-related technologies, including ADC, BCMA, and SDC, effectively reduces dispensing errors in a hospital pharmacy setting. Specifically, we observed a substantial decrease in the average dispensing error rate across 3 stages of technology implementation. Importantly, this study appears to be the first to investigate the combined impact of these 3 specific technologies on dispensing error rates within a hospital pharmacy.

(J Med Internet Res 2025;27:e59220) doi:10.2196/59220

XSL•FO
RenderX

KEYWORDS

medication error; dispensing error; medication-related technology; pharmacy; smart dispensing counter

Introduction

Medication errors pose a significant risk to patient safety and incur substantial economic losses. Recognizing this, medication safety has been identified as a critical area for improvement in all health care settings globally [1]. In March 2017, the World Health Organization (WHO) launched the third Global Patient Safety Challenge, focusing specifically on medication safety. This initiative aims to address systemic weaknesses within health care systems that contribute to medication errors and their potentially severe consequences. The WHO has set an ambitious goal to globally reduce the level of severe, avoidable medication-related harm by 50% within 5 years. This challenge emphasizes improvements across all stages of the medication process, including prescribing, dispensing, administering, monitoring, and patient use [2,3]. In the United States, the Joint Commission on Accreditation of Healthcare Organizations also prioritizes medication safety by including it among its annual patient safety goals [4]. Previous research indicates that approximately 22%-25% of medication errors occur during the dispensing phase [5]. Dispensing errors often precede administration errors, highlighting the critical role of the pharmacy in preventing medication errors. Estimates of medication errors vary significantly across different regions within the prescription (10%-39%), dispensing (11%-40%), and administration (10.5%-38%) stages of medication management. These variations can be attributed to multiple factors [6-14].

Medication errors can arise when weak medication safety systems and human factors, such as fatigue, poor environmental conditions, or staff shortages, compromise any stage of the medication management process [15]. Dispensing errors can lead to preventable patient harm, including adverse drug events, hospitalization, or even death. A systematic review examining the global prevalence of dispensing errors across various pharmacy settings revealed rates ranging from 0% to 33.3%. The worldwide prevalence of dispensing errors was determined to be 1.6% across community, hospital, and other pharmacy settings [16]. Factors influencing dispensing errors include pharmacists' machine usage, workload, and the length of their monthly vacations. Adequate staffing and the use of automated medication preparation systems may be necessary to minimize dispensing errors. Pharmacists returning to dispensing duties after an absence exceeding 72 hours should gradually reintegrate to regain their proficiency [17]. Near-misses, incidents with the potential for medication errors, often share the same root causes as actual errors [18]. Proactively identifying and addressing near-misses is crucial for preventing future medication errors.

Health information technology is believed to be transformative in addressing inefficiencies, preventing medication errors, and improving overall care quality [19]. The integration of health information technology throughout the medication use process is expected to enhance health care safety and efficiency. Past research demonstrates that automated equipment and systems are generally more effective than manual methods, providing a more standardized and reliable execution of medication use processes and thereby reducing errors [20-22].

The use of medication-related technologies, such as automated dispensing cabinets (ADCs), barcode medication administration (BCMA), and smart dispensing counters (SDCs), is increasingly prevalent in hospitals. These technologies aim to reduce medication errors, safeguard medications from improper use, and improve the efficiency of medication processes. This study aims to evaluate the impact of medication-related technologies on both workflow efficiency and prescription dispensing accuracy following their implementation in a Taiwan academic medical center.

Methods

Definitions of Medication Error

The definition of medication error mainly refers to the definition provided by the US National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). A medication error is defined as "any preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of the health care professional, patient, or consumer" [3]. Medication errors occur at each stage of the medication process, including prescribing, dispensing, administering, monitoring, and use.

Dispensing medication is the core function of pharmaceutical care and involves a complex combination of processes, technologies, and human interactions. Dispensing errors include, but are not limited to, dispensing the medicine for the wrong patient, incorrect medicine name, incorrect strength, and incorrect dosage [7].

Study Design

Our hospital experiences 190,000 outpatient visits; 12,000 emergency visits; 7600 inpatient admissions; and 5100 surgeries monthly. To evaluate the impact of implementing medication-related technologies (ADCs, BCMAs, and SDCs) on patient medication safety, we conducted a before-and-after study at our 2202-bed academic medical center in Taiwan.

We analyzed medication error rates within the medication use process from January 1, 2017, to December 31, 2023, using data from the China Medical University Hospital Patient Safety Database (Multimedia Appendices 1 and 2). Our analysis specifically focused on medication errors that occurred during pharmacist medication dispensing. Power business intelligence was used for visual data analytics to conduct a frequency and cause analysis of these errors.

Intervention-Medication–Related Technologies

To address common errors in pharmacist medication dispensing, we used ADC, BCMA, and SDC as interventions. The study periods were defined as follows: stage 0 (preintervention, January to November 2017), stage 1 (post-ADC intervention, December 2017 to June 2018), stage 2 (post-BCMA

XSL•FO
RenderX

intervention, July 2018 to October 2020), and stage 3 (post-SDC intervention, November 2020 to December 2023).

ADC Intervention

An ADC, also called a unit-based cabinet, automated dispensing device, or automated dispensing machine, is a computerized medicine cabinet for hospitals and health care settings. ADCs allow medications to be stored and dispensed near the point of care while controlling and tracking drug distribution [20].

BCMA Intervention

BCMA is a barcode system designed to prevent medication errors in health care settings and improve the quality and safety of medication administration. The overall goals of BCMA are to improve accuracy, prevent errors, and generate digital records of medication administration [20,22,23].

SDC Intervention

The SDC is also known as the LED-guided picking plus lockable drawer (LED-LD) system. The LED-LD system is linked to the medicine management information system, allowing the extraction of data on the location of the medication. The LED-LD system has adopted LED technology to aid in locating the correct medication bin. When scanning the quick response code on a medication label, the LED corresponding to the medication location lights up, directing the pharmacy staff to the correct medication bin. The LED-LD system uses a remote lockable drawer system integrated with LED technology. Upon scanning the quick response code, only the drawer containing the corresponding medication is unlocked, enabling its retrieval. The drawer locks automatically when it is shut [24] (Multimedia Appendix 3).

Outcome

All medication orders, including those written or cosigned by physicians during the study period, were included in the analysis. The pharmacy, providing continuous services for both outpatients and inpatients, adheres to standardized protocols for medication preparation and dispensing. The primary outcome measure was the monthly rate of prescription dispensing errors before and after the implementation of medication-related technologies (ADC, BCMA, and SDC). This rate was calculated by dividing the number of reported prescription dispensing errors by the total number of prescriptions filled each month. Secondary outcome measures included the types of dispensing errors and their severity. The severity of medication errors was assessed by two pharmacists using the NCC MERP method. In case of disagreement, consensus was reached through discussion with a third senior pharmacist.

Statistical Analysis

Descriptive statistics were used to summarize the dispensing error rates at each stage of the medication use process. The dispensing error rate was calculated by dividing the total number of dispensing errors by the total number of medications dispensed. We compared the incidence of dispensing errors before and after the implementation of medication-related technologies (ADCs, BCMAs, and SDCs). Furthermore, we stratified the analyses by the type and severity of dispensing errors.

Differences among the intervention periods were analyzed using chi-square or Fisher exact tests. The Bonferroni method was applied to adjust for multiple comparisons. All statistical analyses were conducted using R software (version 4.1.0, R Foundation for Statistical Computing). Statistical significance was defined as $P<.05$.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the Research Ethics Committee at China Medical University Hospital, Taichung, Taiwan (approval CMUH114-REC2-041). Following the 'Patient Safety Incident Reporting and Reward Operation Guidelines' of China Medical University Hospital, reporter information is automatically identified and transferred upon acceptance and establishment of a patient safety incident report, facilitating subsequent review and collaborative processes. This study utilized secondary use data, derived from a descriptive qualitative analysis of the patient safety notification database, to quantify error type categories related to healthcare information technologies. Notably, this study did not involve any patient or staff identification, and therefore, participant compensation disclosure is not applicable.

Results

Characteristics of Prescription Dispensing Errors of Prevalence, Type, and Potential Severity

Table 1 presents the proportions of prescription dispensing errors by prevalence, type, and potential severity before and after the implementation of medication-related technologies. A total of 15,410,968 medication orders were dispensed in stage 0 (preintervention), followed by 10,721,238 in stage 1; 44,193,666 in stage 2; and 56,260,136 in stage 3. The total number of dispensing errors observed were 968, 406, 1556, and 773 in stages 0, 1, 2, and 3, respectively. The average number of dispensing errors per month decreased from 88 in stage 0 to 58 in stage 1, 56 in stage 2, and 20 in stage 3. Correspondingly, the dispensing error rates declined from 0.0063% in stage 0 to 0.0038% in stage 1, 0.0035% in stage 2, and 0.0014% in stage 3.

In stage 0 (preintervention), the 3 most frequent types of dispensing errors were wrong drug (459/968, 47.4%), wrong dose (335/968, 34.6%), and wrong technique (55/968, 5.7%). While the specific ranks varied slightly across stages, wrong drug, wrong dose, and wrong patient consistently ranked among the top 3 types of dispensing errors. According to the NCC MERP classification, the severity of harm (97%-98.8%) was categorized as “A” (no error) in all stages. The remaining severity of harm primarily fell into categories “B-D” (error, no harm). In stage 2, two cases were categorized as “C” (dose omission and wrong dose), and one case was categorized as “D” (wrong drug), all with no harm. In stage 3, only one case of “C” (wrong dose) was observed, again with no harm.

XSL•FO
RenderX

Table 1. Characteristics of prescription dispensing error and the severity of harm before and after the implementation of medication-related technology.

Stage	Stage 0	Stage 1	Stage 2	Stage 3
Time periods	Preintervention (January to November 2017)	Post-ADC[a] intervention (December 2017 to June 2018)	Post-BCMA[b] intervention (July 2018 to October 2020)	Post-SDC[c] intervention (November 2020 to December 2023)
Months, n	11	7	28	38
Total medication orders per prescription dispensed, n	15,410,968	10,721,238	44,193,666	56,260,136
Total dispensing errors, n	968	406	1556	773
Average dispensing errors per month, n	88	58	56	20
Dispensing error rate (%)[d]	0.0063	0.0038	0.0035	0.0014
Types of dispensing errors, n (%)[e]				
Wrong drug	459 (47.4)	156 (38.4)	568 (36.5)	314 (40.6)
Wrong dose	335 (34.6)	133 (32.8)	549 (35.3)	216 (27.9)
Wrong dosage form	9 (0.9)	13 (3.2)	63 (4.0)	15 (1.9)
Wrong strength	12 (1.2)	4 (1.0)	58 (3.7)	35 (4.5)
Wrong patient	42 (4.3)	33 (8.1)	149 (9.6)	62 (8.0)
Wrong time	0 (0)	1 (0.2)	2 (0.1)	1 (0.1)
Dose omission	23 (2.4)	10 (2.5)	22 (1.4)	22 (2.8)
Monitoring error	9 (0.9)	4 (1.0)	2 (0.1)	0 (0)
Wrong technique	55 (5.7)	28 (6.9)	52 (3.3)	54 (7.0)
Others	24 (2.5)	24 (5.9)	91 (5.8)	54 (7.0)
Severity of harm[f], n (%)[e]				
Category A (no error)	939 (97)	401 (98.8)	1529 (98.3)	760 (98.3)
Category B (error, no harm)	29 (3)	5 (1.2)	24 (1.5)	12 (1.6)
Category C (error, no harm)	0 (0)	0 (0)	2 (0.1)	1 (0.1)
Category D (error, no harm)	0 (0)	0 (0)	1 (0.1)	0 (0)
Category E (error, harm)	0 (0)	0 (0)	0 (0)	0 (0)
Category F (error, harm)	0 (0)	0 (0)	0 (0)	0 (0)
Category G (error, harm)	0 (0)	0 (0)	0 (0)	0 (0)
Category H (error, harm)	0 (0)	0 (0)	0 (0)	0 (0)
Category I (error, death)	0 (0)	0 (0)	0 (0)	0 (0)

[a]ADC: automated dispensing cabinet.

[b]BCMA: barcode medication administration.

[c]SDC: smart dispensing counter.

[d]The number of dispensing errors divided by the number of medication orders.

[e]Percentage are calculated using the number of dispensing errors as the denominator.

[f]National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) classification [3]: no error (category A); error, no harm (categories B-D); error, harm (categories E-H); and error, death (category I). A: circumstances or events that have the capacity to cause an error; B: an error occurred but it did not reach the patient. C: an error occurred that reached the patient but did not cause harm to the patient; D: an error occurred that reached the patient and required monitoring to confirm that it resulted in no harm to the patient, or if necessary, required intervention to preclude harm; E: an error occurred that may have contributed to or resulted in temporary harm to the patient and required intervention; F: an error occurred that may have contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization; G: an error occurred that may have contributed to or resulted in permanent harm to the patient; H: an error occurred that required intervention necessary to sustain life; and I: an error occurred that may have contributed to or resulted in the patient's death.

XSL•FO
RenderX

Effect of Medication-Related Technology on Reducing Dispensing Errors

Figure 1 shows a significant decrease in the rate of reported prescription dispensing errors across the observation periods. Compared with the preintervention stage 0 (traditional manual medication picking), where the average dispensing error rate was 0.0063%, a notable reduction was observed in subsequent stages: stage 1 (post-ADC intervention) at 0.0038% (39.68% decrease), stage 2 (post-BCMA intervention) at 0.0035% (44.44% decrease), and stage 3 (post-SDC intervention) at 0.0014% (77.78% decrease).

Figure 1. The trend curve of the dispensing error rate before and after the implementation of medication-related technology in each stage from 2017 to 2023. ***P<.001. ADC: automated dispensing cabinet; BCMA: barcode medication administration; SDC: smart dispensing counter.

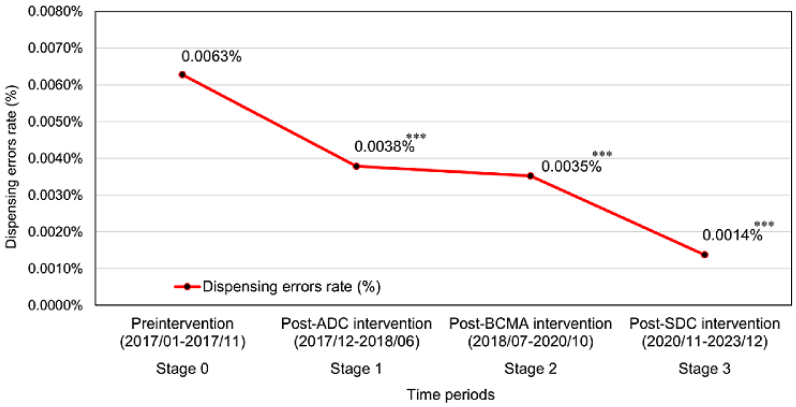

Effect of Introducing ADC on Reducing Dispensing Errors

Table S1 in Multimedia Appendix 4 shows the incidence of dispensing errors by type before and after the introduction of an ADC. Compared with the preintervention stage (stage 0), the post-ADC intervention stage (stage 1) demonstrated significant reductions in the incidence of the wrong drug (0.0030%-0.0015%; P<.001), wrong dose (0.0022%-0.0012%; P<.001), and total dispensing error rates (0.0063%-0.0038%; P<.001). Other error types, including wrong dosage form, wrong strength, wrong patient, wrong time, dose omission, monitoring error, wrong technique, and other errors, showed either minimal changes or slight decreases.

Regarding the severity of harm, a significant reduction was observed in category A (no error) from 0.0061% in stage 0 to 0.0037% in stage 1 (P<.001). Furthermore, the prevalence of category B (error, no harm) decreased from 0.00019% in stage 0 to 0.00005% in stage 1 (P=.02).

Effect of Introducing BCMA on Reducing Dispensing Errors

Table S1 in Multimedia Appendix 4 shows the incidence of dispensing errors by type before and after the introduction of BCMA. Compared with the preintervention stage (stage 0), the post-BCMA intervention stage (stage 2) demonstrated significant reductions in the incidence of wrong drug (0.0030%-0.0013%; P<.001), wrong dose (0.0022%-0.0012%; P<.001), dose omission (0.00015%-0.00005%; P=.001), monitoring errors (0.00006%-0.00001%; P<.001), wrong technique (0.0004%-0.0001%; P<.001), and the total dispensing error rate (0.0063%-0.0035%, P<.001). Compared with the post-ADC intervention stage (stage 1), only the wrong technique showed a significant reduction in stage 2 (P=.005).

Regarding the severity of harm, a significant reduction was observed in category A (no error) from 0.0061% in stage 0 to 0.0035% in stage 2 (P<.001). Similarly, category B (error, no harm) also demonstrated a significant reduction from 0.00019% in stage 0 to 0.00005% in stage 2 (P<.001). No significant differences in the severity of harm were observed between stage 1 and stage 2.

Effect of Introducing SDC on Reducing Dispensing Errors

Table S1 in Multimedia Appendix 4 shows the incidence of dispensing errors by type before and after the introduction of SDC. Compared with the preintervention stage (stage 0), the post-SDC intervention stage (stage 3) demonstrated significant reductions in the incidence of wrong drug (0.0030%-0.0006%; P<.001), wrong dose (0.0022%-0.0004%; P<.001), wrong patient (0.0003%-0.0001%; P<.001), dose omission (0.00015%-0.00004%; P<.001), monitoring errors (0.00006%-0%; P<.001), wrong technique (0.0004%-0.0001%; P<.001), and the total dispensing error rate (0.0063%-0.0014%; P<.001). Significant reductions were also observed between stage 1 and stage 2, as well as between stage 2 and stage 3 for multiple error types, including wrong drug, wrong dose, wrong dosage form, wrong patient, others, and the total dispensing error rate.

Regarding the severity of harm, a significant reduction was observed in category A (no error) from 0.0061% in stage 0 to 0.0014% in stage 3 (P<.001). Similarly, category B (error, no harm) also demonstrated a significant reduction from 0.00019% in stage 0 to 0.00002% in stage 3 (P<.001). Significant reductions in category A were also observed between stage 1 and stage 3 (P<.001), as well as between stage 2 and stage 3 (P<.001).

As given in Table 2, statistically significant reductions in prescription dispensing error rates were observed between each stage (P<.001), except when comparing stage 1 to stage 2.

XSL•FO
RenderX

Table 2. Comparisons of the prescription dispensing error rates at each stage.

Stage	Stage name	Prescribed medications, n	Dispensing errors, n (%)	*P* value	Comparison *P* value			
					Vs stage 0	Vs stage 1	Vs stage 2	Vs stage 3
0	Preintervention	15,410,968	968 (0.0063)	<.001[a]	—[b]	<.001[a]	<.001[a]	<.001[a]
1	Post-ADC[c] intervention	10,721,238	406 (0.0038)	—	<.001[a]	—	>.99	<.001[a]
2	Post-BCMA[d] intervention	44,193,666	1556 (0.0035)	—	<.001[a]	>.99	—	<.001[a]
3	Post-SDC[e] intervention	56,260,136	773 (0.0014)	—	<.001[a]	<.001[a]	<.001[a]	—

[a]Statistically significant (P<.001) after adjustment using the Bonferroni correction.

[b]Not available.

[c]ADC: automated dispensing cabinet.

[d]BCMA: barcode medication administration.

[e]SDC: smart dispensing counter.

Discussion

Main Findings of Medication-Related Technology in Reducing Dispensing Errors

The study revealed that prescription dispensing error rates significantly decreased after the introduction of medication-related technologies, including ADC, BCMA, and SDC. Specifically, error rates reduced by 39.68%, 44.44%, and 77.78% from stage 0 (preintervention) to stage 1 (post-ADC intervention), stage 2 (post-BCMA intervention), and stage 3 (post-SDC intervention), respectively. The average number of dispensing errors per month also significantly decreased from 88 in stage 0 to 58 in stage 1, to 56 in stage 2, and to 20 in stage 3. The most frequent error type in stage 0, wrong drug, was significantly reduced by 51.15%, 56.85%, and 81.26% in stages 1 to 3, respectively. Similarly, the wrong dose, the second most frequent error type, was significantly reduced by 42.93%, 42.85%, and 82.34% across the stages. Compared with stage 0, the severity of harm, as categorized by category A (no error) and category B (error, no harm), was significantly decreased in all subsequent stages. These findings demonstrate the effectiveness of medication-related technologies in reducing dispensing errors in a hospital pharmacy and improving medication safety.

The Actual Effectiveness of Introducing ADC in Reducing Dispensing Errors

In stage 1 (post-ADC intervention), the dispensing error rate decreased significantly from 0.0063% in stage 0 (preintervention) to 0.0038% (a reduction rate of 39.71%). This reduction was observed across all error types, including wrong drug, wrong dose, and the total number of dispensing errors. Notably, the reduction rate of wrong drug was substantial, decreasing from 0.0030% to 0.0015% (a 51.15% reduction), and the reduction rate of wrong dose also decreased significantly from 0.0022% to 0.0012% (a 42.93% reduction). These findings are consistent with previous studies that have demonstrated the effectiveness of ADCs in reducing dispensing errors [22,25-27]. The primary impact of ADC implementation was observed in a significant reduction of wrong drug and wrong dose. Previous research by Shah et al [26] highlighted that while ADCs can eliminate certain error types, they may also introduce new error types from a pharmacist's perspective [26]. Therefore, pharmacists play a crucial role in ensuring the safe and effective use of ADCs by implementing appropriate policies, procedures, and quality assurance programs to address safety, accuracy, security, and patient confidentiality [28].

The Actual Effectiveness of Introducing BCMA in Reducing Dispensing Errors

In stage 2 (post-BCMA intervention), the dispensing error rate decreased significantly by 43.95% compared with stage 0 (0.0063% vs 0.0035%). These findings were consistent across all error types, including wrong drug, wrong dose, monitoring error, wrong technique, dose omission, and the total number of dispensing errors. Notably, the reduction rate of the wrong drug was substantial, decreasing from 0.0030% to 0.0013% (56.85% reduction). The reduction rates of wrong dose, dose omission, and monitoring errors were also significant, decreasing from 0.0022% to 0.0012% (42.85% reduction), from 0.00015% to 0.00005% (66.67% reduction), and from 0.000058% to 0.000005% (92.25% reduction), respectively. The wrong technique error rate also decreased significantly from 0.0004% to 0.0001% (67.03% reduction). Compared with stage 1, only the wrong technique error rate showed a statistically significant difference. These findings are consistent with previous studies [29-31] and demonstrate the effectiveness of BCMA in reducing dispensing errors, particularly in preventing wrong drug, wrong dose, dose omission, monitoring error, and wrong technique.

Consistent with other studies, our analysis showed that BCMA implementation was associated with a reduction in dispensing errors, especially for medications with high error rates or those prone to confusion. However, the effectiveness of BCMA relies heavily on proper implementation, including comprehensive education and training on standard operating procedures. For example, simultaneous scanning of multiple medications can lead to errors, and issues with barcode quality or accuracy can also hinder the system's effectiveness. In addition, damaged and unreadable drug barcodes, as well as medications labeled with incorrect barcodes, can also cause problems.

XSL•FO
RenderX

The Actual Effectiveness of Introducing SDC in Reducing Dispensing Errors

In stage 3 (post-SDC intervention), the dispensing error rate decreased significantly by 78.13% compared with stage 0 (0.0063% vs 0.0014%). These findings were consistent across all error types, including wrong drug, wrong dose, wrong patient, dose omission, monitoring error, wrong technique, and the total number of errors. Notably, the reduction rate of wrong drug errors was substantial, decreasing from 0.0030% to 0.0006% (81.26% reduction). The reduction rates of wrong dose and wrong patient errors were also significant, decreasing from 0.0022% to 0.0004% (82.34% reduction) and from 0.0003% to 0.0001% (59.56% reduction), respectively. Monitoring errors were completely eliminated in stage 3. The wrong technique and dose omission error rates also decreased significantly. Compared with stage 1 and stage 2, significant reductions were observed in multiple error types in stage 3, including wrong drug, wrong dose, wrong patient, and others. The implementation of SDC demonstrated significant reductions in wrong drug, wrong dose, wrong patient, dose omission, monitoring error, and wrong technique. These findings are consistent with a previous study by Teo et al [24], which observed that technology-assisted medication picking systems, such as an LED-guided system with a lockable drawer, significantly reduced near-miss medication errors in a tertiary referral hospital in Singapore.

Our observations highlight the potential of the SDC system, designed to address the unique dispensing needs of Taiwan, to significantly reduce medication errors. By incorporating visual, auditory, and perceptual cues, the SDC system can guide pharmacists through the dispensing process, minimizing errors that may occur due to factors such as infrequent medication use, similar packaging, or distractions. This is particularly important for new pharmacists who are still familiarizing themselves with medication locations and workflow.

Limitations

This study has several limitations. First, variations in medication error definitions across the literature make comparisons challenging. We attempted to standardize error classification according to the NCC MERP guidelines and calculated error rates based on the total number of dispensing errors and medication orders per prescription. Second, the staggered implementation of ADC, BCMA, and SDC (December 2017, July 2018, and November 2020, respectively) may have introduced a time-related bias. To mitigate this, we implemented standardized staff training on system operation and troubleshooting and provided on-site technical support. Furthermore, direct observation of drug dispensing was not conducted, making it difficult to confirm the absence of management errors despite random checks of standard dispensing procedures.

Automated systems may face challenges with irregularly shaped or infrequently used medications, requiring backup systems during downtime and maintenance [32]. Concurrent manual workflows during automation can also introduce potential sources of error. Based on our findings, optimizing the benefits of dispensing and management technologies requires adequate resource allocation for postimplementation support, including education, training, and technical assistance. The occurrence of system-related errors highlights the need for continuous monitoring and proactive problem-solving. For example, the system should track the movement of drugs after delivery, including any necessary withdrawals by pharmacists and nursing staff. Clear procedures for responding to system failures must be communicated to all employees. To maximize the benefits of these technologies, continuous evaluation, system refinement, and process optimization are crucial.

Conclusions

This study demonstrated that the implementation of 3 medication-related technologies (ADC, BCMA, and SDC) significantly reduced dispensing errors compared with traditional manual medication picking. Overall, the study observed a significant reduction in dispensing errors across all stages. Except for wrong dosage form, wrong strength, wrong time, and other minor error types, all major error categories (including wrong drug, wrong dose, and wrong patient) showed statistically significant reductions (P<.001). The severity of harm also decreased significantly at each stage (P<.001). Significant differences in dispensing error rates were observed between all stages (P<.001). These findings underscore the potential of medication-related technologies to enhance medication safety and improve the efficiency of medication dispensing processes in hospital settings. Notably, the study observed significant reductions in critical error types such as wrong drug and wrong dose, which can have serious consequences for patient safety. This study shows that medication-related technology reduces dispensing errors and improves medication safety in hospitals. As hospitals work to enhance efficiency and safety, more targeted research is needed to assess the benefits and risks of these technologies. The findings highlight that technology can reduce errors and enhance medication management by improving dispensing and prescribing, optimizing clinical pharmacist resources, and allowing staff to focus more on patient care.

Acknowledgments

We are grateful to China Medical University Hospital for providing administrative and technical support. We also thank Ying-Shu You (Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan) for her statistical analysis and consultation.

XSL•FO
RenderX

Authors' Contributions

All authors have read and have approved of the submission of this manuscript. WNY and YDC did the primary conceptualization of the study. WNY curated the data, performed the formal analysis, and conducted the methodology. WNY and YDC wrote the original manuscript draft. YDC reviewed and edited the manuscript. YCH and YWH supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Types of dispensing error.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v27i1e59220_app1.xlsx]

Multimedia Appendix 2
Severity of harm.
[XLSX File (Microsoft Excel File), 28 KB - jmir_v27i1e59220_app2.xlsx]

Multimedia Appendix 3
Smart dispensing counter (an LED-guided manual picking system) in a hospital pharmacy. LED-LD: LED-guided picking plus lockable drawer.
[PNG File , 877 KB - jmir_v27i1e59220_app3.png]

Multimedia Appendix 4
Table S1. Incidence of prescription dispensing errors and severity of harm before and after implementation of medication-related technology.
[DOCX File , 18 KB - jmir_v27i1e59220_app4.docx]

References

1. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human: Building a Safer Health System. Washington, DC: The National Academies Press; 2000.
2. Medication without harm. World Health Organization. 2017. URL: https://iris.who.int/bitstream/handle/10665/255263/WHO-HIS-SDS-2017.6-eng.pdf?sequence=1 [accessed 2024-11-26]
3. About medication errors: what is a medication error? National Coordinating Council for Medication Error Reporting and Prevention (NCCMERP). URL: https://www.nccmerp.org/about-medication-errors [accessed 2024-11-26]
4. National patient safety goals. The Joint Commission. URL: https://www.jointcommission.org/standards/national-patient-safety-goals/ [accessed 2024-11-26]
5. Hicks RW, Cousins DD, Williams RL. Selected medication-error data from USP's MEDMARX program for 2002. Am J Health Syst Pharm 2004 May 15;61(10):993-1000. [doi: 10.1093/ajhp/61.10.993] [Medline: 15160775]
6. Aldhwaihi K, Umaru N, Pezzolesi C, Schifano F. A systematic review of the nature of dispensing errors in hospital pharmacies. IPRP 2016 Jan;5:1-10. [doi: 10.2147/iprp.s95733]
7. Cheung KC, Bouvy ML, de Smet PAGM. Medication errors: the importance of safe dispensing. Br J Clin Pharmacol 2009 Jun;67(6):676-680 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03428.x] [Medline: 19594537]
8. Lewis PJ, Dornan T, Taylor D, Tully MP, Wass V, Ashcroft DM. Prevalence, incidence and nature of prescribing errors in hospital inpatients: a systematic review. Drug Saf 2009;32(5):379-389. [doi: 10.2165/00002018-200932050-00002] [Medline: 19419233]
9. Berdot S, Gillaizeau F, Caruba T, Prognon P, Durieux P, Sabatier B. Drug administration errors in hospital inpatients: a systematic review. PLoS One 2013;8(6):e68856 [FREE Full text] [doi: 10.1371/journal.pone.0068856] [Medline: 23818992]
10. Tully MP, Ashcroft DM, Dornan T, Lewis PJ, Taylor D, Wass V. The causes of and factors associated with prescribing errors in hospital inpatients: a systematic review. Drug Saf 2009;32(10):819-836. [doi: 10.2165/11316560-000000000-00000] [Medline: 19722726]
11. Turner E, Kennedy MC, Barrowcliffe A. An investigation into prescribing errors made by independent pharmacist prescribers and medical prescribers at a large acute NHS hospital trust: a cross-sectional study. Eur J Hosp Pharm 2020 Mar 11;28(3):149-153. [doi: 10.1136/ejhpharm-2019-002074]
12. Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. JAMA 1995 Jul 05;274(1):35-43. [Medline: 7791256]
13. Keers RN, Williams SD, Cooke J, Ashcroft DM. Prevalence and nature of medication administration errors in health care settings: a systematic review of direct observational evidence. Ann Pharmacother 2013 Feb;47(2):237-256. [doi: 10.1345/aph.1R147] [Medline: 23386063]

XSL•FO
RenderX

14. McLeod MC, Barber N, Franklin BD. Methodological variations and their effects on reported medication administration error rates. BMJ Qual Saf 2013 Apr;22(4):278-289. [doi: 10.1136/bmjqs-2012-001330] [Medline: 23322752]
15. Reporting and learning systems for medication errors: the role of pharmacovigilance centres. World Health Organization. URL: https://iris.who.int/handle/10665/137036 [accessed 2024-11-26]
16. Um IS, Clough A, Tan ECK. Dispensing error rates in pharmacy: a systematic review and meta-analysis. Res Social Adm Pharm 2024 Jan;20(1):1-9 [FREE Full text] [doi: 10.1016/j.sapharm.2023.10.003] [Medline: 37848350]
17. Chou H, Wang Y, Liao L, Chen J, Chen X, Tang K, et al. Exploring susceptibility factors to medication dispensing errors through a retrospective study of patient-reported dispensing errors over 11 years: are dispensing errors indeed due to personal reasons for pharmacists? Eur J Hosp Pharm 2024 Dec 30:ejhpharm-2023-004064. [doi: 10.1136/ejhpharm-2023-004064] [Medline: 38839267]
18. Sheikhtaheri A. Near misses and their importance for improving patient safety. Iran J Public Health 2014 Jun;43(6):853-854 [FREE Full text] [Medline: 26110160]
19. Identification and prioritization of HIT patient safety measures. National Quality Forum. URL: https://www.qualityforum.org/Publications/2016/02/Identification_and_Prioritization_of_HIT_Patient_Safety_Measures.aspx [accessed 2024-11-24]
20. Zheng WY, Lichtner V, van Dort BA, Baysari MT. The impact of introducing automated dispensing cabinets, barcode medication administration, and closed-loop electronic medication management systems on work processes and safety of controlled medications in hospitals: a systematic review. Res Social Adm Pharm 2021 May;17(5):832-841 [FREE Full text] [doi: 10.1016/j.sapharm.2020.08.001] [Medline: 32891535]
21. James KL, Barlow D, Bithell A, Hiom S, Lord S, Pollard M, et al. The impact of automation on workload and dispensing errors in a hospital pharmacy. Int J Pharm Pract 2013 Apr;21(2):92-104. [doi: 10.1111/j.2042-7174.2012.00238.x] [Medline: 23418851]
22. Takase T, Masumoto N, Shibatani N, Matsuoka Y, Tanaka F, Hirabatake M, et al. Evaluating the safety and efficiency of robotic dispensing systems. J Pharm Health Care Sci 2022 Oct 01;8(1):24 [FREE Full text] [doi: 10.1186/s40780-022-00255-w] [Medline: 36180937]
23. Johnson CL, Carlson RA, Tucker CL, Willette C. Using BCMA software to improve patient safety in Veterans Administration Medical Centers. J Healthc Inf Manag 2002;16(1):46-51. [Medline: 11813523]
24. Teo MEH, Chia BY, Lee YC, Tay PLY, Wong JA, Lee SB, et al. Cost-effectiveness of two technology-assisted manual medication picking systems versus traditional manual picking in a hospital outpatient pharmacy. Eur J Hosp Pharm 2021 Mar;28(2):100-105 [FREE Full text] [doi: 10.1136/ejhpharm-2019-001997] [Medline: 33608438]
25. Jumeau M, Francois O, Bonnabry P. Impact of automated dispensing cabinets on dispensing errors, interruptions and pillbox preparation time. Eur J Hosp Pharm 2023 Jul;30(4):237-241 [FREE Full text] [doi: 10.1136/ejhpharm-2021-002849] [Medline: 34426488]
26. Shah SR, Galt KA, Fuji KT. Error types with use of medication-related technology: a mixed methods research study. Res Social Adm Pharm 2019 Dec;15(12):1480-1483. [doi: 10.1016/j.sapharm.2019.01.010] [Medline: 30665824]
27. Tu HN, Shan TH, Wu YC, Shen PH, Wu TY, Lin WL, et al. Reducing medication errors by adopting automatic dispensing cabinets in critical care units. J Med Syst 2023 Apr 27;47(1):52 [FREE Full text] [doi: 10.1007/s10916-023-01953-0] [Medline: 37103718]
28. Cello R, Conley M, Cooley T, de la Torre C, Dorn M, Ferer DS, et al. ASHP guidelines on the safe use of automated dispensing cabinets. Am J Health Syst Pharm 2022 Jan 01;79(1):e71-e82. [doi: 10.1093/ajhp/zxab325] [Medline: 34453436]
29. Jessurun JG, Hunfeld NGN, van Rosmalen J, van Dijk M, van den Bemt PMLA. Effect of automated unit dose dispensing with barcode scanning on medication administration errors: an uncontrolled before-and-after study. Int J Qual Health Care 2021 Nov 13;33(4):mzab142 [FREE Full text] [doi: 10.1093/intqhc/mzab142] [Medline: 34662396]
30. Cochran GL, Barrett RS, Horn SD. Comparison of medication safety systems in critical access hospitals: combined analysis of two studies. Am J Health Syst Pharm 2016 Aug 01;73(15):1167-1173. [doi: 10.2146/ajhp150760] [Medline: 27440623]
31. Shah K, Lo C, Babich M, Tsao NW, Bansback NJ. Bar code medication administration technology: a systematic review of impact on patient safety when used with computerized prescriber order entry and automated dispensing devices. Can J Hosp Pharm 2016;69(5):394-402 [FREE Full text] [doi: 10.4212/cjhp.v69i5.1594] [Medline: 27826157]
32. Alanazi MF, Shahein MI, Alsharif HM, Alotaibi SM, Alanazi AO, Alanazi AO, et al. Impact of automated drug dispensing system on patient safety. PHARMPRACT 2022 Dec 01;20(4):01-11. [doi: 10.18549/pharmpract.2022.4.2744]

Abbreviations

ADC: automated dispensing cabinet
BCMA: barcode medication administration
LED-LD: LED-guided picking plus lockable drawer
NCC MERP: National Coordinating Council for Medication Error Reporting and Prevention
SDC: smart dispensing counter
WHO: World Health Organization

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 05.04.24; peer-reviewed by J Waterson, E Oluokun; comments to author 18.04.24; revised version received 11.01.25; accepted 26.02.25; published 27.03.25.

Please cite as:
Yu WN, Cheng YD, Hou YC, Hsieh YW
Implementation of Medication-Related Technology and Its Impact on Pharmacy Workflow: Real-World Evidence Usability Study
J Med Internet Res 2025;27:e59220
URL: https://www.jmir.org/2025/1/e59220
doi:10.2196/59220
PMID:40019479

©Wei-Ning Yu, Yih-Dih Cheng, Yu-Chi Hou, Yow-Wen Hsieh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Intelligent System for Classifying Patient Complaints Using Machine Learning and Natural Language Processing: Development and Validation Study

Xiadong Li[1*], MD, PhD; Qiang Shu[1*], MD, PhD; Canhong Kong[2], MD; Jinhu Wang[1], MD, PhD; Gang Li[3], MD; Xin Fang[4], MA; Xiaomin Lou[5*], MD; Gang Yu[1], PhD

[1]Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center For Child Health, Hang Zhou, China

[2]Patient Service Surveillance Office, Medical Information Department, Hangzhou Red Cross Hospital, Hang Zhou, China

[3]Department of Radiation Oncology, Zhe Jiang Xiaoshan hospital, Hangzhou Normal University, Hang Zhou, China

[4]Hospital Management Office, Hangzhou Cancer Hospital, Hang Zhou, China

[5]Patient Service Surveillance Office, Hangzhou Red Cross Hospital, Hang Zhou, China

[*]these authors contributed equally

Corresponding Author:
Gang Yu, PhD
Children's Hospital
Zhejiang University School of Medicine
National Clinical Research Center For Child Health
No. 3333, Binsheng Road, Binjiang District, Hangzhou
Hang Zhou, 310020
China
Phone: 86 13588773370
Email: yugbme@zju.edu.cn

Abstract

Background: Accurate classification of patient complaints is crucial for enhancing patient satisfaction management in health care settings. Traditional manual methods for categorizing complaints often lack efficiency and precision. Thus, there is a growing demand for advanced and automated approaches to streamline the classification process.

Objective: This study aimed to develop and validate an intelligent system for automatically classifying patient complaints using machine learning (ML) and natural language processing (NLP) techniques.

Methods: An ML-based NLP technology was proposed to extract frequently occurring dissatisfactory words related to departments, staff, and key treatment procedures. A dataset containing 1465 complaint records from 2019 to 2023 was used for training and validation, with an additional 376 complaints from Hangzhou Cancer Hospital serving as an external test set. Complaints were categorized into 4 types—communication problems, diagnosis and treatment issues, management problems, and sense of responsibility concerns. The imbalanced data were balanced using the Synthetic Minority Oversampling Technique (SMOTE) algorithm to ensure equal representation across all categories. A total of 3 ML algorithms (Multifactor Logistic Regression, Multinomial Naive Bayes, and Support Vector Machines [SVM]) were used for model training and validation. The best-performing model was tested using a 5-fold cross-validation on external data.

Results: The original dataset consisted of 719, 376, 260, and 86 records for communication problems, diagnosis and treatment issues, management problems, and sense of responsibility concerns, respectively. The Multifactor Logistic Regression and SVM models achieved weighted average accuracies of 0.89 and 0.93 in the training set, and 0.83 and 0.87 in the internal test set, respectively. Ngram-level term frequency–inverse document frequency did not significantly improve classification performance, with only a marginal 1% increase in precision, recall, and F_1-score when implementing Ngram-level term frequency–inverse document frequency (n=2) from 0.91 to 0.92. The SVM algorithm performed best in prediction, achieving an average accuracy of 0.91 on the external test set with a 95% CI of 0.87-0.97.

Conclusions: The NLP-driven SVM algorithm demonstrates effective classification performance in automatically categorizing patient complaint texts. It showed superior performance in both internal and external test sets for communication and management problems. However, caution is advised when using it for classifying sense of responsibility complaints. This approach holds

XSL•FO
RenderX

promises for implementation in medical institutions with high complaint volumes and limited resources for addressing patient feedback.

(J Med Internet Res 2025;27:e55721) doi:10.2196/55721

KEYWORDS

complaint analysis; text classification; natural language processing; NLP; machine learning; ML; patient complaints

Introduction

Background

Patient complaints refer to patients feeling dissatisfied or having opinions regarding medical services, treatment processes, medical staff attitudes, or health care institution management during their participation in medical care [1]. These complaints are expressed through written or verbal feedback or grievances to relevant organizations or individuals. Complaints can involve various aspects, including medical errors, poor treatment outcomes, unfriendly service attitudes, and excessive waiting times among other issues.

Handling patient complaints is an essential part of health care institution management, as it can help improve the quality of medical services and enhance patient satisfaction. Patient complaints also serve as valuable sources for gaining insights into safety-related issues within health care institutions. Patients are often more sensitive to a range of issues within medical services compared with health care professionals within the institution. Some of these issues may not be identified by traditional medical surveillance systems [2], such as adverse event reporting systems or mortality case reviews. Therefore, patient complaints can provide crucial information for health care institutions on how to improve patient safety.

Currently, patient satisfaction surveys are conducted through SMS text messages or online methods (such as through a WeChat [Tencent] official account) in our hospital. If a patient responds negatively, a staff member from the outpatient office follows up with a phone call. Under traditional management methods, satisfaction survey personnel meticulously document specific complaints from patients regarding dissatisfaction with the hospital's medical management. These complaints are then summarized and regularly sent to the hospital's Patient Advocacy Center (PAC). They are manually categorized and reported every 2 weeks during hospital meetings and also distributed to the email addresses of relevant clinical and medical technology department heads for corrective action and improvement. Only if the patient receives feedback from the relevant department, it will be considered as closing the loop on the complaint-handling process.

On average, about 25 new complaints are submitted through various channels each day, while the PAC can only analyze and process approximately 10 complaints loops per day. At this rate, it can be estimated that the hospital's PAC will still have over 5400 patient complaints waiting to be analyzed and addressed within a year. Clearly, without significant changes to the screening and categorization process, the hospital's PAC will be unable to analyze all submitted complaints within a reasonable timeframe to provide adequate feedback and resolution. The crux of the issue lies in the inability to classify patient complaints in a timely manner, which consequently hampers targeted communication between subsequent departments and patients. This delay leads to patients lodging secondary or even multiple complaints due to prolonged waiting times. In addition, departments with the capability to address specific issues remain idle as they do not receive timely patient complaints. Experience with several hundred such interventions at community and academic medical centers shows fewer subsequent complaints associated with most of those receiving timely and effective feedback [3]. Therefore, tardily and ineffectively addressing patient complaints has become a significant bottleneck restricting the enhancement of hospital management quality and medical safety.

Automatic classification tools play a crucial role in streamlining the triage process, enabling health care providers to efficiently prioritize and address patient concerns based on the severity and nature of their complaints. By automating this classification process, clinicians can allocate resources more effectively, leading to improved patient outcomes and satisfaction. In addition, such tools facilitate the organization and analysis of large volumes of patient data, ultimately contributing to evidence-based decision-making and enhanced health care delivery. The problem is that there are hardly any studies, articles, or systems focusing on the automatic classification of patient complaints. The need for automatic classification of patient complaints is crucial, as manual identification of complaints is inefficient, time-consuming, and highly prone to errors.

The main contributions of this study are by automating the patient complaint triage process, PAC can allocate resources more efficiently, thereby improving patient care experience and satisfaction. In addition, such tools help organize and analyze large amounts of patient complaint data, ultimately contributing to evidence-based medical decisions and improved medical services to ensure patient treatment safety.

The structure of the article is as follows. "Existing Approaches to Text Classification" introduces previous studies on existing approaches to text classification and the patients complain auto-classification methodology. "Methods" and "Results" introduce machine learning (ML) models and experimental results. "Discussion" presents an analysis and discussion of selected results and offers some suggestions for future research. Finally, "Conclusion" summarizes our research.

Existing Approaches to Text Classification

The classification of patient complaint texts falls within the realm of text auto-classification. Therefore, many previous studies in text classification provide valuable insights for us (Multimedia Appendix 1) [4-9]. Especially, auto-classification

XSL•FO
RenderX

of texts based on patients' chief complaints (CC) during the diagnostic process is the most relevant to our research on auto-classifying patient complaint texts.

CC text classification refers to the task of automatically categorizing or classifying text data based on the main concerns or symptoms expressed by patients when they seek medical attention [10]. In principle, each patient complaint should be classified into 1 category, but typically 1 patient complaint may be classified into a single or multiple categories because it contains more than 1 category of disease. A recorded free-text CC is assigned to a category either manually by emergency medicine physicians or other emergency department (ED) professionals, or by using an automated mapping algorithm [11-14]. Automated CC categorization, on the other hand, are more suitable for a wide range of (ED) applications. Traditional automated CC classification algorithms typically use a linear mapping algorithm (keyword search and category matching algorithm [15]), or a semantic model based on Bayesian network [12].

The study by Arnaud et al [16] leveraged natural language processing (NLP) solutions to predict medical specialties at hospital admission by integrating structured data with unstructured textual notes. An MLP model and a convolutional neural network (CNN) were independently used to achieve promising accuracy in the analysis of over 260,000 ED records. Although the classifier model could only achieve about 68% accuracy of prediction, this research contributes to the growing use of NLP methods in health care analytics. NLP models achieved high accuracy in predicting the need for admission, triage score, critical illness, and mapping free-text CC to structured fields. Incorporating both structured data and free-text data improved results when compared with models that used only structured data. However, the majority of studies (16/20, 80%) were assessed to have a high risk of bias, and only 1 study reported the deployment of an NLP model into clinical practice [17].

These traditional linear classification methods are all based on the bag-of-words (BOW) representation, which treats text as an unordered collection of words, ignoring the position and grammatical structure of words, and only focusing on the frequency of word occurrences. In the BOW model, text is represented as a vector, where each dimension corresponds to a word, and each element of the vector represents the frequency or occurrence of the corresponding word in the text. However, it is known that BOW technique suffers from issues such as loss of word order information, curse of dimensionality, sparsity, and semantic loss. These limitations restrict the performance of the BOW model in NLP tasks [18,19]. Therefore, researchers have proposed various improvement methods to overcome these issues, such as using word embedding techniques and models that consider contextual information [20]. A representative work was proposed by Alhazzani et al [21], who used pretrained static word embeddings and pretrained vector word embeddings to train deep learning (DL)–based BiLSTM (Bidirectional Long Short-Term Memory) and BiGRU (Bidirectional Gated Recurrent Unit) classifiers. Lee et al [13] developed a recurrent neural network (RNN)–based long short-term memory (LSTM) and gated recurrent unit (GRU) cells for text classification [13], and they reported that in all instances, the RNN models outperformed the BOW classifiers suggesting DL models could substantially improve the automatic classification of unstructured text for syndromic surveillance.

Currently, 1 significant innovation using the transformer architecture [22] and using bidirectional encoding to effectively capture dependency relationships from both directions of context simultaneously, has significantly propelled technological advancements across various CC text classification tasks [23,24]. The performance of those models (BERT [Bidirectional Encoder Representations from Transformers; Google], Bio_BERT, Clinical BERT, KP_BERT, CC_BERT) was compared with that of the term frequency–inverse document frequency (TF-IDF) [25] model serving as the baseline. However, the results revealed that the TF-IDF model outperforms a robust BERT-based model on the test dataset and exhibits statistical comparability in terms of misspelling sets.

Despite the application of DL methods, including CNNs, RNNs, and BERT, in text categorization—particularly in handling large-scale, high-dimensional data and significantly enhancing text classification performance for complex tasks [26,27]—traditional ML approaches continue to offer advantages in terms of computational resource requirements and the interpretability of results compared with DL, CNNs, and RNNs. Modern RNNs can have millions of free parameters and thus require huge datasets [28]. Training models of this size require significant computing resources. In the task of patient complaints classifying, besides pursuing classification accuracy, the interpretability of ML models should be prioritized. This is because our goal was to identify the root causes of patient complaints through a correct interpretation of classification results, and the interpretability of models such as DL, CNNs, and RNNs is notably low. This significantly undermines the credibility of clinical interpretations of the aforementioned analysis results and their subsequent applicability. Therefore, 3 interpretable common ML classification models (Multifactor Logistic Regression [MLR], Multinomial Naive Bayes [MNB], and Support Vector Machines [SVMs]) are used in our research work.

By training ML models on large datasets of labeled complaint data, these models can learn to accurately categorize complaints based on their content and context. This automated classification approach can significantly improve the efficiency and accuracy of complaint management systems, reducing reliance on subjective human intervention and enabling faster processing and resolution of patient complaints. In addition, it can help to identify patterns and trends in complaints so that PAC can proactively address common issues and improve overall patient satisfaction.

In this study, we explore the automatic text classification of patient complaints using 2 types of TF-IDF techniques and trained an ML tool for automatic classification of patient complaint texts written in Chinese based on the existing mature English text automatic classification models.

We posit that an automated classification architecture for the categorization of patient complaints can be created, which consists of 2 modules.

XSL•FO
RenderX

First, the NLP module is based on the TF-IDF technique. The function of this module is to perform word frequency and extract features from the patient complaint texts. By using the TF-IDF technique (Word-level TF-IDF and Ngram-level TF-IDF), the importance of each word and semantics in the text can be calculated and keywords representing the core meaning of the complaint can be extracted. This helps the model to better understand and classify the complaint texts.

Second, the ML module is to classify complaint texts. ML algorithms have been widely adopted for text classification [29-33]. However, existing research in the literature has primarily focused on the English language. Therefore, patient comments written in Chinese are currently difficult to automatically analyze using traditional ML or statistical approaches. While several classification models based on Chinese text have emerged, their efficacy, algorithmic intricacy, and substantial hardware requirements have impeded the broad implementation of this approach in scholarly discourse [34,35]. The challenge of dealing with Chinese text lies in its complex linguistic features, including character-based writing, tonal variations, and the absence of explicit word boundaries. In addition, Chinese exhibits rich contextual meanings and cultural nuances, which can pose difficulties for NLP tasks such as machine translation, sentiment analysis, and text summarization. By using translation tools (such as ChatGPT 3.5 [OpenAI] or other professional translation tools) to translate Chinese text into English, and then having English professionals review and correct the translated results, these reviewed English texts were used as input data for ML models for classification training. This approach can overcome the challenges of directly processing Chinese text and provides an effective way to use Chinese text data in the test classification process.

Methods

Data Acquisition and Preprocessing Flow

A retrospective study of patient complaints was conducted at Hangzhou Red Cross Hospital and Hangzhou Cancer Hospital from 2015 to 2019. A comprehensive compilation of 1817 documented complaints were carefully collated. Complaints involving the same patient and the same incident were grouped into 1 independent complaint case. This study was approved by the ethics committee, which granted a waiver of review and exempted it from the informed consent requirement. The data from Hangzhou Red Cross Hospital were used for model training and internal validation, while the data from Hangzhou Cancer Hospital were used for external testing of the model with the best performance in the ML model. All data collection and analysis methods in this study strictly adhere to the "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research" [36].

Due to the current difficulty of directly applying ML and classification training to Chinese text, coupled with the availability of mature software and tools for translating Chinese into English, along with the review of translation results by English-proficient graduate students, converting the Chinese written text into English for training purposes is a feasible and efficient approach for automatic text classification. All the texts recorded in Chinese were translated into English by ChatGPT 3.5. Furthermore, 2 postgraduates with English language skills were responsible for reviewing the machine-translated texts; if there were obvious translation errors, the texts were corrected after the 2 individuals proposed unanimous modification suggestions.

If 2 individuals had a disagreement regarding the translation of the same Chinese word, we would escalate the matter to native English-speaking professors in the International Education Department of the hospital. They would make the final decision to ensure that each identical Chinese word in all Chinese complaint texts was translated into English with complete consistency. To ensure the efficiency of the expert correction mechanism, we have established a reward system. For every flaw or error found in the machine-translated text, a certain amount of funds would be allocated from our project budget for rewards.

Once sufficient data have been collected, preprocessing is divided into the 3 steps.

In the first step, the complaint texts are tokenized. After comparing jieba (Multimedia Appendices 2 and 3), THULAC [37], and HanLP [38], the more mature tool jieba is used in this study. To achieve the best balance between speed and accuracy in tokenization, the Hidden Markov Model Bigram (HMM-Bigram) [39] algorithm is used. Part-of-speech tagging is performed at the same time as tokenization. This involves assigning labels to words based on their respective parts of speech, such as adjectives, verbs, nouns, and so on. Part-of-speech tagging is based on maximum entropy and maximum likelihood approaches.

The second step is to eliminate duplicate parts in the tokenized results and ensure that each word appears only once. In this way, a total vocabulary is generated.

In the third step, all complaint texts are searched. If a word from the total vocabulary occurs in a particular complaint text, the corresponding position is marked with 1. If it does not appear, it is marked with 0. This creates a matrix with words and tokens as horizontal and vertical axes. Each complaint text can be regarded as a vector of 1s and 0s.

Feature Extraction

After preprocessing, feature extraction was adopted before model training. The features are presented in text form, whereby words and phrases with a strong semantic meaning are included as a feature set. The feature set contains most of the information from the entire text, which is beneficial for document classification. Content that is not included in the feature set may result in the loss of some semantic information but has minimal impact on the classification task. Complaint texts usually have a length of 10-100 words and are therefore relatively short. To achieve a relatively accurate classification, the TF-IDF technique was used for the extraction of text features.

After the preprocessing step, there are still many semantically meaningless particles or signs in the text. For example, "doctor" appears frequently in the text, but its contribution to text classification is relatively small. In addition to the common 841

XSL•FO
RenderX

Chinese stop words, we included an additional 17 frequently occurring but semantically insignificant words from the medical complaint domain into our stop words list based on unanimous recommendations from the staff of our collaborating hospital's PAC. This decision was made based on our previous experience, as detailed in Multimedia Appendix 4. By including them as stop words, we can focus on more specific and meaningful keywords and thus improve the accuracy and relevance of extracting high-frequency words.

Semantic Analysis

Due to the singularity of the environment in which patients complain, the content of the text has a high similarity in terms of scenes and semantics. Therefore, in addition to using Word-level TF-IDF, we also used Ngram-level TF-IDF to examine the impact of consecutive words on the importance of text. Given our dataset's modest size of approximately 2600 cases and the need for higher granularity in data analysis, the decision was made to use n=2 for Ngram-level TF-IDF.

After TF-IDF analysis and extension of existing stop words, the problem of particles such as "doctor" can be alleviated, and the feature text is more suitable for the following ML text classification training. For example, the complaint text "Doctor issued unnecessary examination orders for financial reasons" was changed to "unnecessary examination orders financial reasons." The top 10 high-probability words and word clouds for each type of complaint are calculated and displayed. This step not only allows for an understanding of the basic structure of existing complaint text data but also enables the examination of the correctness of manually labeled complaint texts, preparing necessary groundwork for subsequent model training and verification.

Unbalanced Data Handling

The degree of imbalance of dataset is based on the proportion of a minority class in the whole dataset and could range from mild (20%-40%), moderate (1%-20%) to vigorous (<1%) imbalances [40]. Previous studies showed that the resampling approach is a useful preprocessing step to handle the imbalanced dataset [41]. This method modifies the imbalanced distribution of the majority and minority classes at the data level before training with classifiers. Before training the datasets, we used the Synthetic Minority Oversampling Technique (SMOTE) as an imbalanced adjustment strategy [42]. The decision was made due to a moderate dataset imbalance (sense of responsibility/communication problem=11.9%). The datasets of the minority class, including diagnosis and treatment, management problem, and sense of responsibility, were all up sampled to 719. Unlike regular SMOTE, we matched the data volume of minority classes to the majority class through up sampling, while keeping the actual amount of data in the majority class unchanged. This maneuvering minimized the impact on the raw data distribution. One important point to emphasize is that SMOTE only oversamples the internal data used for training and validation, while the independent external data used for testing remains untouched.

Classification Model Training and Validation

In order to ensure the consistency of the labeling by the 2 experts, the complaint data were independently labeled by the 2 experts. Secondary annotations for the data of different categories were made by PAC staff from our cooperative hospital, comprising 3 people in total. Finally, the classification results were counted from 5 people for this annotation, and the category that received the most votes was determined as the final complaint classification annotation. We implemented a comparison of several well-known ML classifiers to determine if a supervised ML algorithm can achieve a reasonable classification accuracy compared to the expert rules. The complaint types are "communication problem," "diagnosis and treatment," "management problem," and "sense of responsibility." The sample counts for each type are shown in Table 1.

In total, 3 ML methods (MLR, MNB, and SVMs) were used to train and validate the model. In the training process, 75% (539/719) of the data in each category were used for training, the rest of 25% (180/719) was used for verification. Another 376 patient complaint records from Hangzhou Cancer Hospital were used for external testing of the trained model with the best performance in the training and verification section. The composition of the 376 patient complaint records is communication problem (n=122), diagnosis and treatment (n=89), management problem (n=85), and sense of responsibility (n=80). The performance of ML algorithms in the training and verification sets is shown in Table 2. To assess the generalization ability of the trained model in a more objective manner, a 5-fold cross-validation (CV) was applied to evaluate the model performance during external testing. Figure 1 summarizes the entire process of preprocessing and ML method for patient complaint classification.

Table 1. Sample number of complaints for each type in the training and verification sets.

Category	Sample size of raw data	Sample size after SMOTE[a]
Communication problem	719	719
Diagnosis and treatment	376	719
Management problem	260	719
Sense of responsibility	86	719
Total	1441	2876

[a]SMOTE: Synthetic Minority Oversampling Technique.

XSL•FO
RenderX

Table 2. Comparison of the performance of 3 machine learning algorithms.

Machine learning models	Communication problem		Diagnosis and treatment		Management problem		Sense of responsibility	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
MLR[a]								
Precision	0.83	0.71	0.89	0.83	0.99	0.97	1.00	0.99
Recall	0.95	0.91	0.88	0.81	0.94	0.88	0.96	0.89
F_1-score	0.89	0.80	0.88	0.82	0.96	0.92	0.98	0.94
Support	160	140	195	183	209	197	210	200
SVM[b]								
Precision	0.75	0.67	1.00	0.95	1.00	0.99	1.00	1.00
Recall	1.00	0.98	0.96	0.81	0.98	0.92	0.96	0.92
F_1-score	0.86	0.80	0.98	0.87	0.99	0.95	0.98	0.96
Support	167	122	219	210	220	193	209	195
MNB[c]								
Precision	0.73	0.67	0.98	0.95	1.00	0.99	1.00	1.00
Recall	1.00	0.98	0.94	0.81	0.96	0.92	0.97	0.92
F_1-score	0.84	0.80	0.96	0.87	0.98	0.95	0.98	0.96
Support	134	122	219	210	220	193	219	195

[a]MLR: Multifactor Logistic Regression.

[b]SVM: Support Vector Machine.

[c]MNB: Multinomial Naive Bayes.

Figure 1. Illustration of the text preprocessing workflow and patient complaint classification model development and validation process. HMM-Bigram: Hidden Markov Model Bigram; MLR: Multifactor Logistic Regression; MNB: Multinomial Naive Bayes; SMOTE: Synthetic Minority Oversampling Technique; SVM: Support Vector Machine; TF-IDF: term frequency–inverse document frequency.

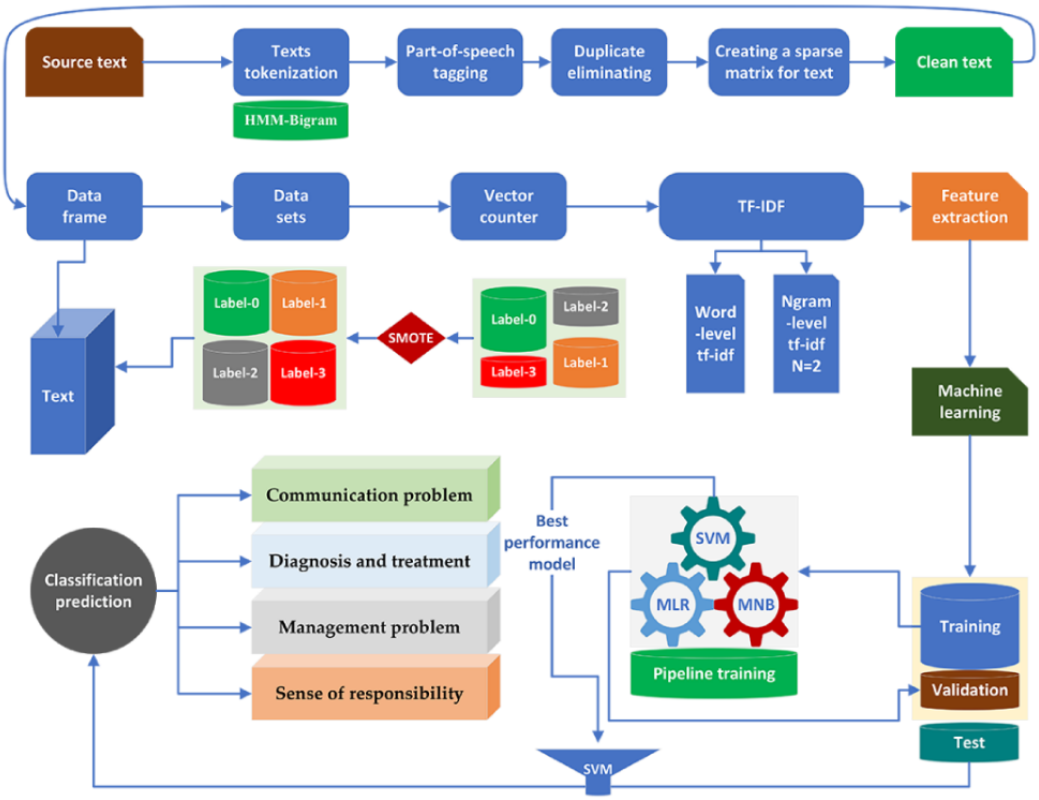

XSL•FO
RenderX

Ethical Considerations

From an ethical standpoint, the present study focused on the examination of public behavior in human subjects. The information collection methods used by the researchers were designed to ensure that subjects could not be directly identified, nor could they be indirectly identified through associated identifiers. To facilitate the effective handling of patient complaints, informed consent was obtained for telephone recordings or written complaints, with access to this information restricted to designated hospital departments. In the context of this research, only the specific content of patient complaints was analyzed, without any extraction or linkage to personal identifiable information. Furthermore, the identities of hospital staff members who were the subjects of complaints were anonymized, with only their departmental affiliation being retained. This approach served to mitigate potential conflicts between complainants and the staff involved. In addition, the potential for harm or discomfort anticipated in the study did not surpass that which individuals might have experienced in daily life or during standard physical and psychological examinations. Consequently, the hospital's ethics review committee granted a waiver of review for this research.

The study was carried out in compliance with the Declaration of Helsinki and received ethical approval as a waiver of review from the Ethics Committee of Hangzhou Red Cross Hospital.

Results

Overview

In our investigation, we discovered that out of 1817 patient complaints, approximately 23 cases, accounting for 1.3% of the total, required modification post translation by ChatGPT 3.5 to avoid potential ambiguity. Among these cases, communication problem accounted for 7 instances, diagnosis and treatment only had 1, management problem had 4, while sense of responsibility constituted 11 cases. Most ambiguities were concentrated within communication problem and sense of responsibility categories, with relatively fewer occurrences in the more objective diagnosis and treatment and Management Problem categories. This observation may be attributed to the inherent difficulty patients encounter in articulating issues related to communication problem and sense of responsibility, compared with the relatively straightforward nature of diagnosis and treatment and management problem categories.

The HMM-Bigram algorithm was used for part-of-speech tagging. For the input text: "General surgery patients complained that unnecessary examination items were applied by the doctor," the tagging results are as follows: [('General', 'JJ-TL'), ('surgery', 'NN'), ('patients', 'NNS'), ('complained', 'VBD'), ('that', 'DT'), ('doctor', 'NN'), ('applied', 'VBN'), ('unnecessary', 'JJ'), ('examination', 'NN'), ('items', 'NNS'), ('performance', 'NN'), ('commissions', 'NNS')]. Each tag consists of 2 parts, the first being the word itself, and the second being the part-of-speech tag for that word. For example, in the first tag of the results ('General', 'JJ-TL'), "General" is the word, and "JJ-TL" is its part-of-speech tag. In this tag, "JJ" indicates an adjective, while "TL" represents the word's tag in the Brown Corpus. Similarly, ('surgery', 'NN') indicates that "surgery" is a noun.

The following image (Figure 2) illustrates a matrix with words and tokens as horizontal and vertical axes with the HMM-Bigram algorithm.

In the methodology section, 2 methods, Word-level TF-IDF and Ngram-level TF-IDF (with n=2), were used for feature extraction from the text. The TF-IDF values and their performance on the training and verification sets are shown in Figure 3 and Table 3, respectively.

Figure 2. Illustration of a matrix with words and tokens as horizontal and vertical axes with the Hidden Markov Model Bigram token algorithm.

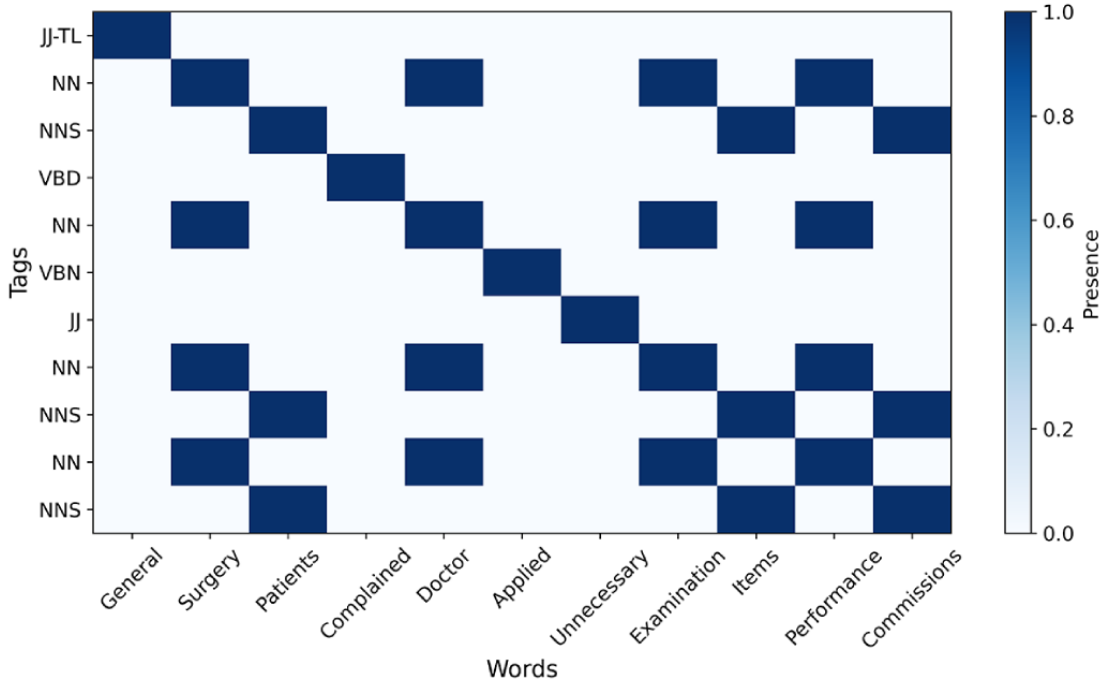

XSL•FO
RenderX

Figure 3. Comparison of the term frequency–inverse document frequency values between Word-level and Ngram-level term frequency–inverse document frequency (with n=2). TF-IDF: term frequency–inverse document frequency.

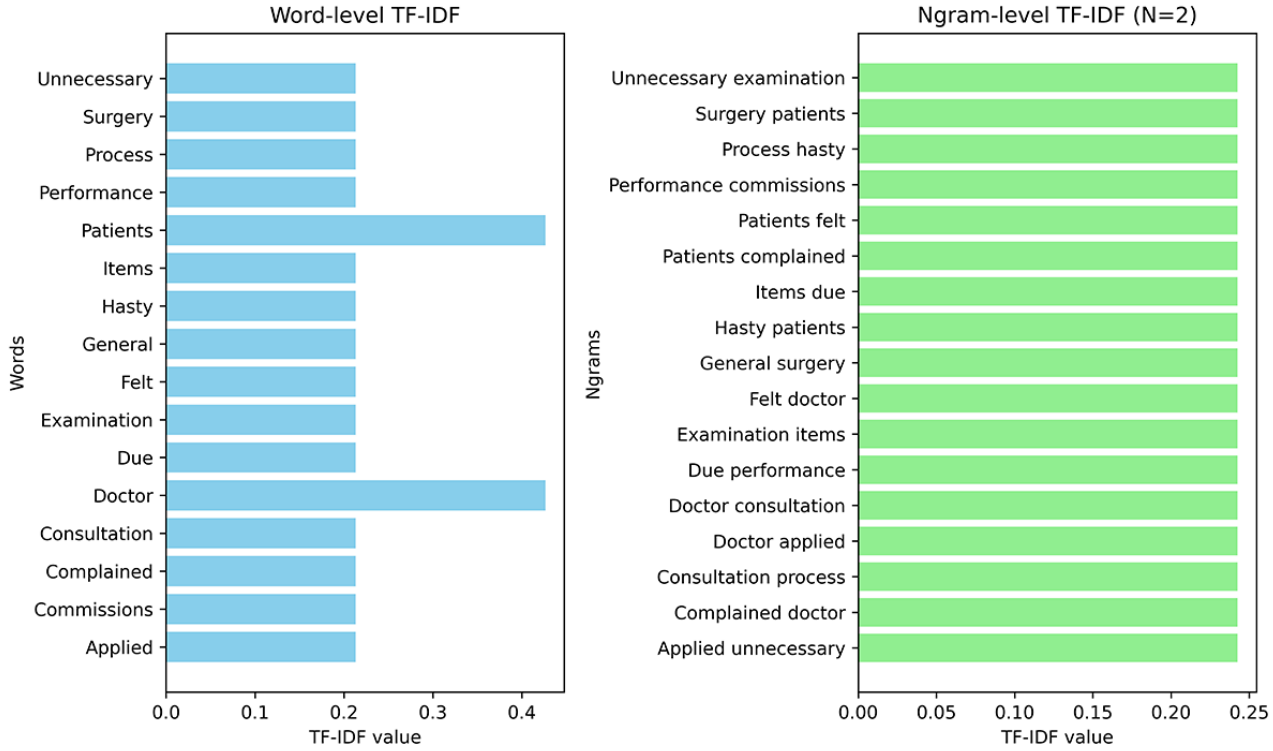

Table 3. Classification results comparison of Word-level term frequency–inverse document frequency and Ngram-level term frequency–inverse document frequency (n=2).

	Precision		Recall		F_1-score		Support	
	WL[a]	NL[b]	WL	NL	WL	NL	WL	NL
Communication problem	0.73	0.74	0.95	0.95	0.82	0.83	138	140
Management problem	0.93	0.93	0.89	0.88	0.91	0.91	189	189
Diagnosis and treatment	0.99	0.99	0.89	0.9	0.94	0.94	201	200
Sense of responsibility	1	1	0.94	0.94	0.97	0.97	192	191
Accuracy	__[c]	—	—	—	0.91	0.92	720	720
Macro average	0.91	0.92	0.92	0.92	0.91	0.91	720	720
Weighted average	0.93	0.93	0.91	0.92	0.92	0.92	720	720

[a]Word-level term frequency–inverse document frequency.

[b]Ngram-level term frequency–inverse document frequency.

[c]Not applicable.

Semantic Analysis

A total of 4 complaint categories underwent preprocessing after incorporating expended stop words to enhance the specificity of medical complaints. This process generated high-frequency word histograms, and the word frequency analysis results for the 4 types of complaints will be uploaded as supplementary files (Multimedia Appendices). The top 3 vocabulary words in the category of "communication problem" are "explanation," "enough," and "believes." In the category of "diagnosis and treatment" complaints, apart from "diagnosis and treatment" itself, the main concerns were "effect," "pediatric," and "emergency." In the category of "management problem" complaints, the primary content revolves around "clinic," "waiting," and "time." As for the category of "sense of responsibility," the prominent content includes "medicine," "back," and "went."

Performance of Patient Complaint Classification Models

The performance of 3 ML models in the training and validation sets was demonstrated in Table 2 (for more detailed data including the 95% CIs for each metric, please refer to Multimedia Appendix 5). It was found that MLR achieved the best performance in classifying the communication problem, as assessed by F_1-score. However, in the classification tasks of

XSL•FO
RenderX

the remaining 3 categories, SVM outperformed MLR and MNB, comprehensively. Furthermore, from the training and validation iteration process, it can be observed that the model began to converge well by the eighth epoch (Figures 4 and 5). In addition, the best model achieved very high receiver operating characteristic (ROC) values (ROC>0.97) in classifying the 4 types on the test set. Hence, the SVM algorithm was adopted for the subsequent external data validation section with a 5-fold CV method applied. A more detailed comparison of the point estimate and 95% CIs for each model's performance metrics has been included in Multimedia Appendix 6.

Figure 4. Demonstration of training and validation performance in machine learning iteration process. Training and validation performance of Support Vector Machine. acc: accuracy.



Figure 5. Training and validation performance of Support Vector Machine (receiver operating characteristic curve). ROC: receiver operating characteristic.

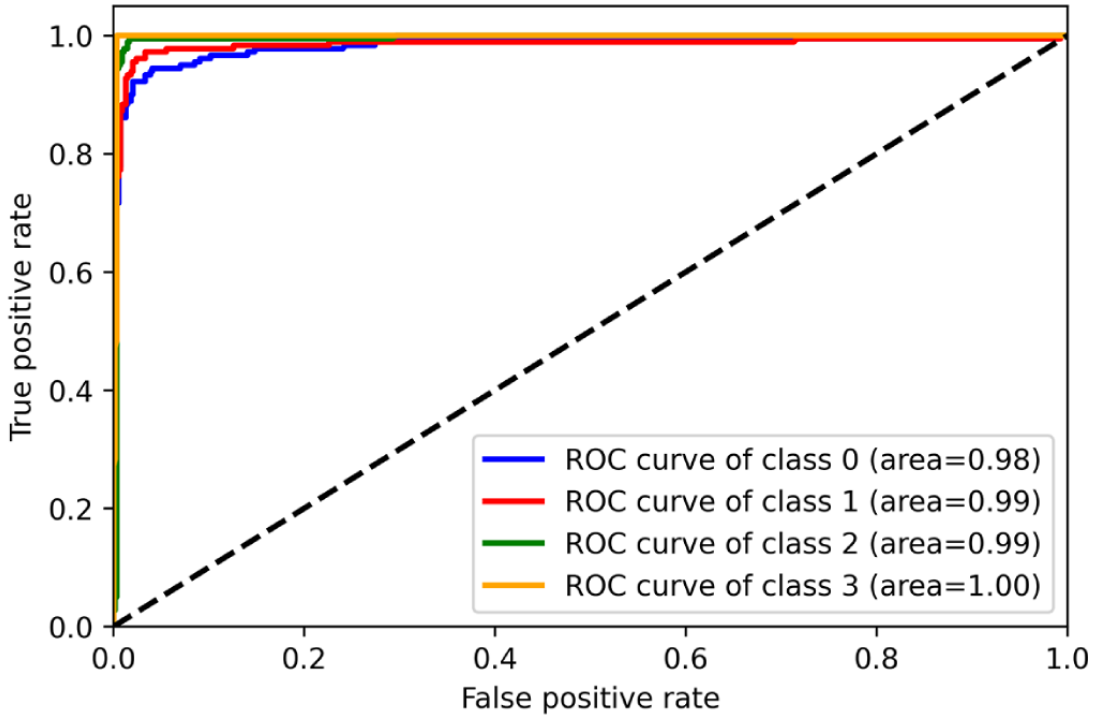

Stability of Model Iteration Process

To monitor the stability of the model's training effects during the machine-learning process, we set a maximum iteration number of 30. It was found that all ML algorithms perfectly converged in the first 20 iterations, and the MLR had the most stable output and the best generalization performance during the training and validation iterations (Figures 6-8). However, due to the use of the SGD (Stochastic Gradient Descent) algorithm, the SVM algorithm tended to output significantly fluctuating results during the training process (Figures 4 and 5). Although the MNB algorithm performed well on the training set, it exhibited overfitting and mediocre generalization ability on the test set (Figures 9 and 10).

XSL•FO
RenderX

Figure 6. Demonstration of training and validation performance in machine learning iteration process. Confusion matrix of Multinomial Naive Bayes, Logistic Regression, and Support Vector Machine in testing set. svm: Support Vector Machine.

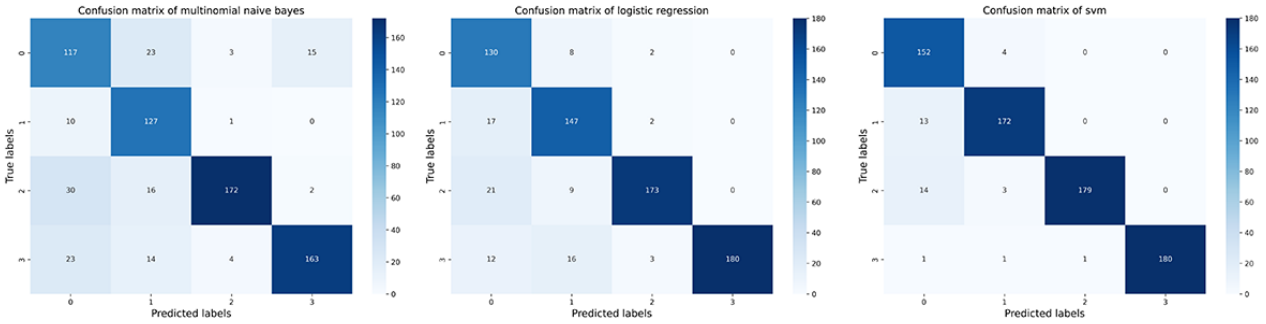

Figure 7. Demonstration of training and validation performance in machine learning iteration process. Training and validation performance of Multinomial Naive Bayes. acc: accuracy.

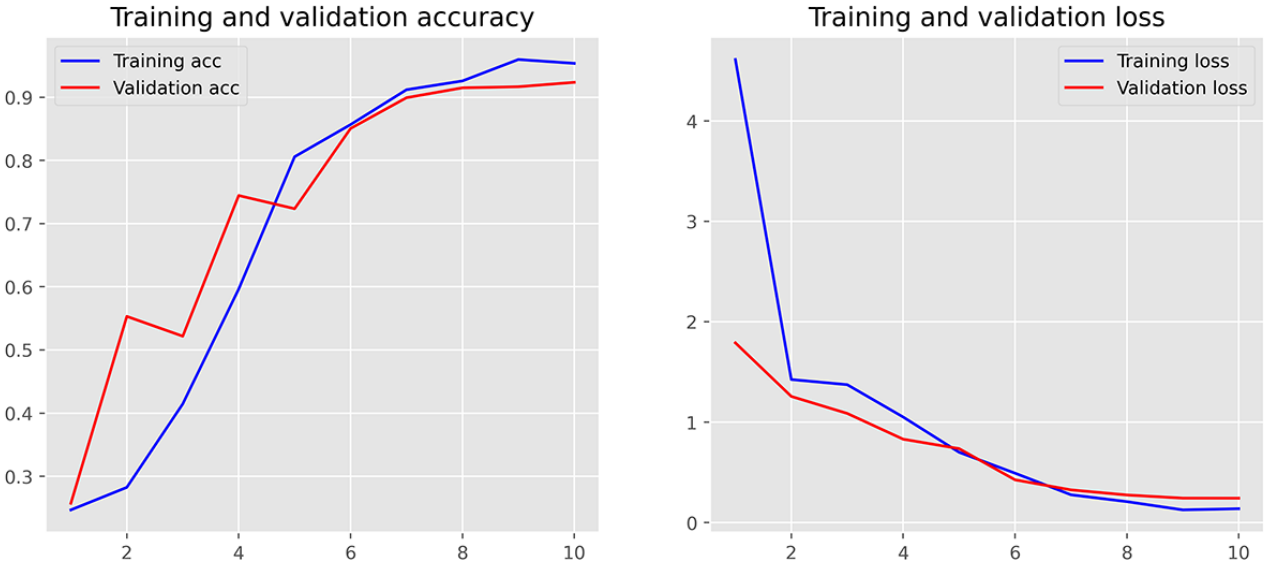

Figure 8. Training and validation performance of Multinomial Naive Bayes (receiver operating characteristic curve). ROC: receiver operating characteristic.

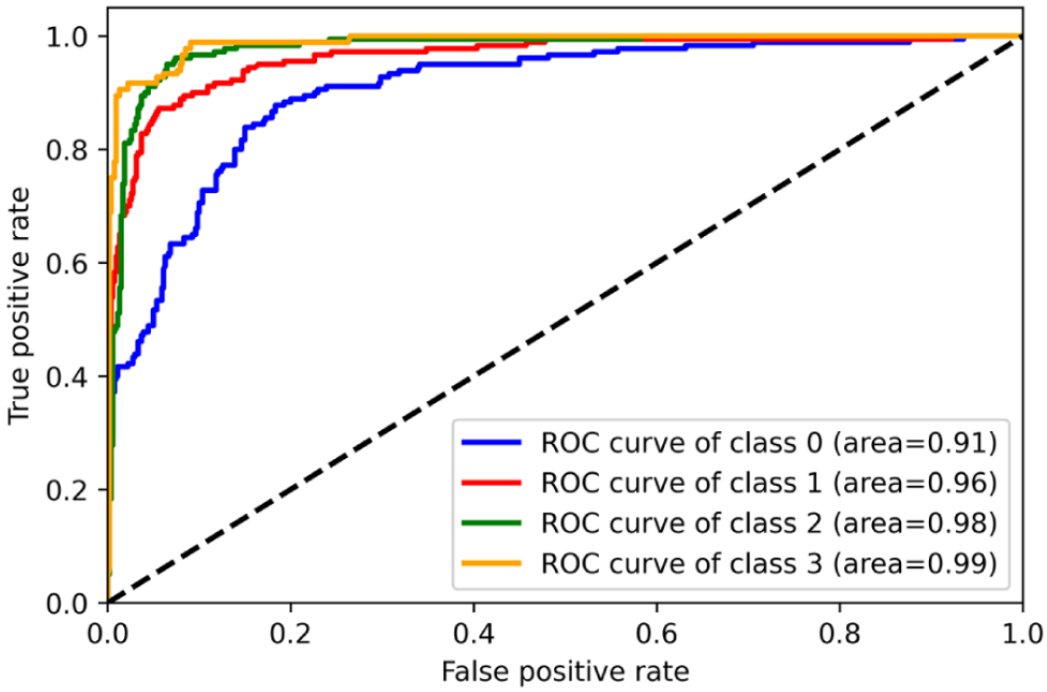

XSL•FO
RenderX

Figure 9. Demonstration of training and validation performance in machine learning iteration process. Training and validation performance of Multifactor Logistic Regression. acc: accuracy.

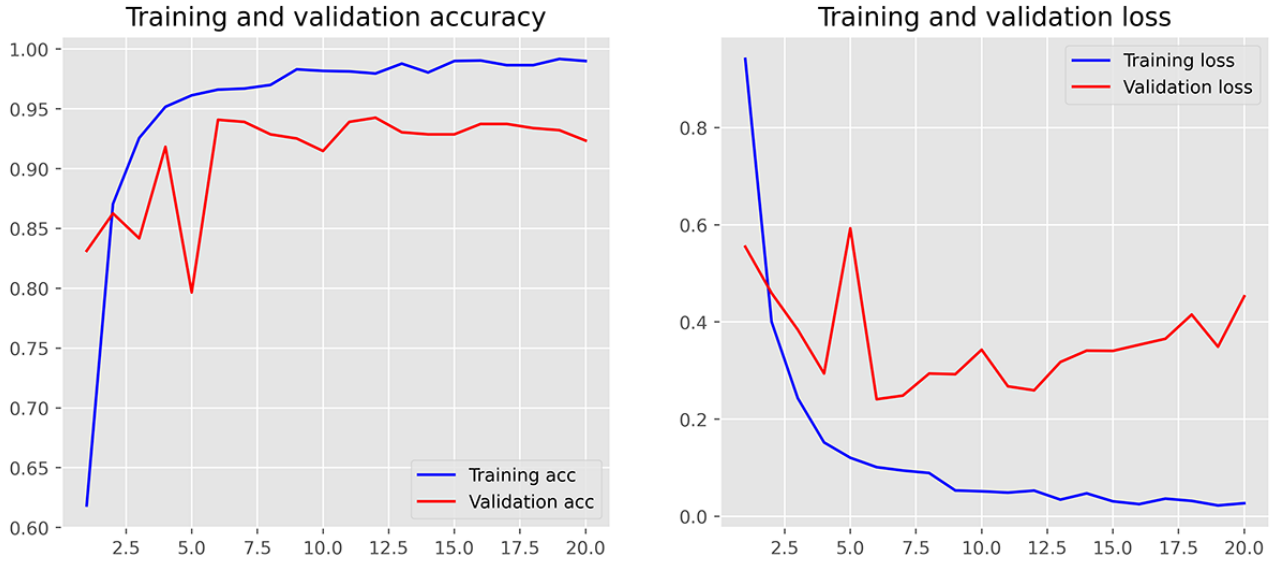

Figure 10. Training and validation performance of Multifactor Logistic Regression (receiver operating characteristic curve). ROC: receiver operating characteristic.

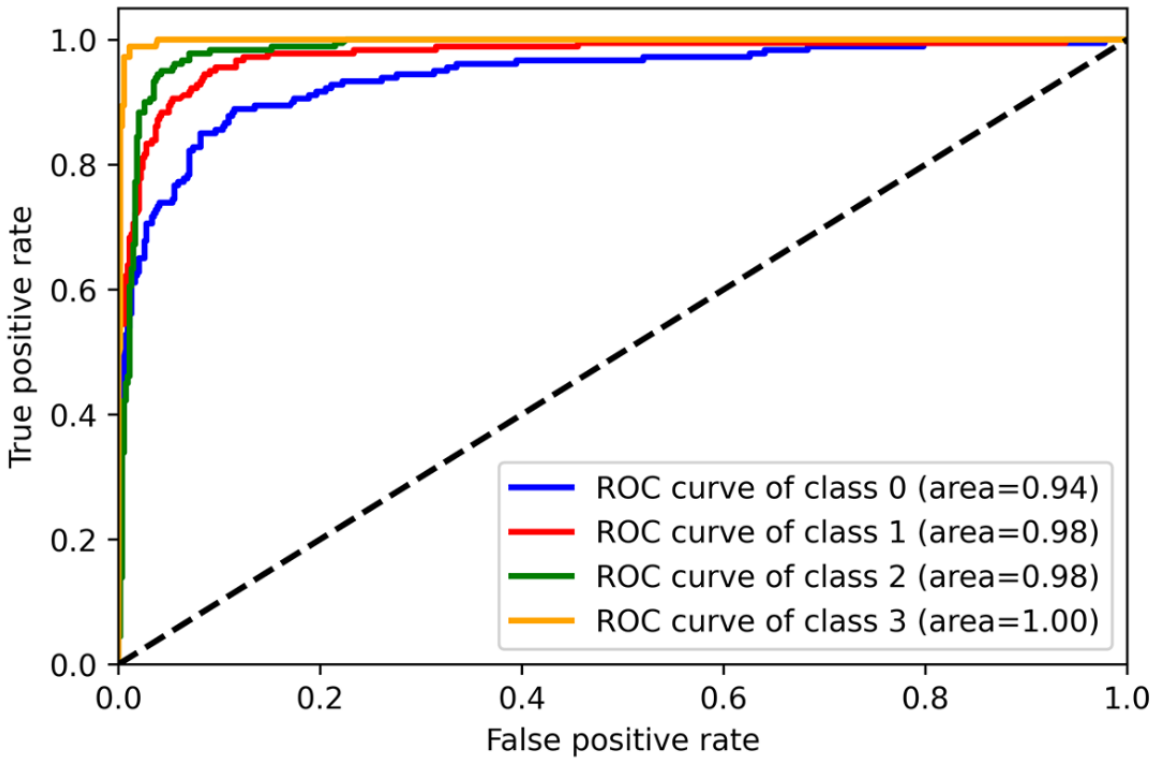

Performance of the 5-Fold SVMs Algorithm on the External Test Dataset

It was inspiring to see that the SVM classification algorithm achieved a commendable result in accurately categorizing 3 distinct types of patient complaints (communication problem, diagnosis, treatment, and management problem) on the external test dataset. The average area under the curve (AUC) for these 3 types was 0.94, 0.86, and 0.81, respectively, with an average SD below 0.1 in all 3 types of patient complaints (Figures 11-13). However, the average AUC for the sense of responsibility category was only 0.7 (in the 5-fold testing, the minimum AUC is 0.48, and the maximum AUC is only 0.82; Figure 14).

XSL•FO
RenderX

Figure 11. A 5-fold cross-validation curve of receiver operating characteristic of Support Vector Machine on the external test data set. Receiver operating characteristic for communication problem classification. AUC: area under the curve; ROC: receiver operating characteristic.

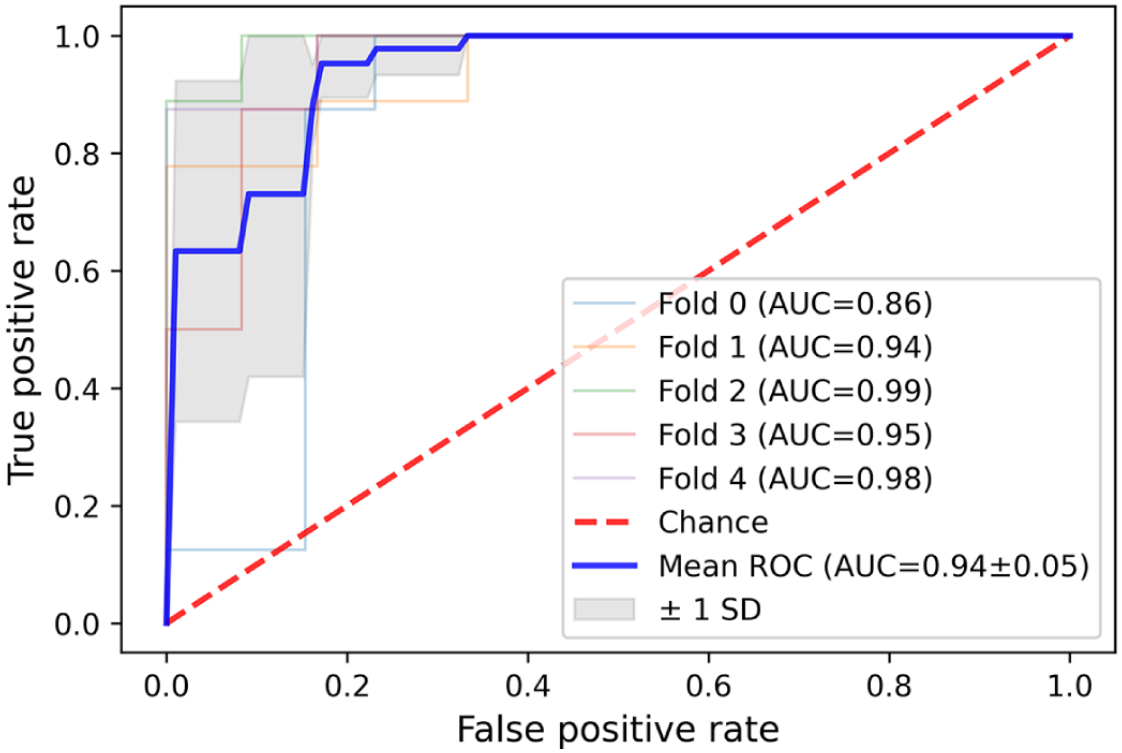

Figure 12. A 5-fold cross-validation curve of receiver operating characteristic of Support Vector Machine on the external test data set. Receiver operating characteristic for diagnosis and treatment problem classification. AUC: area under the curve; ROC: receiver operating characteristic.

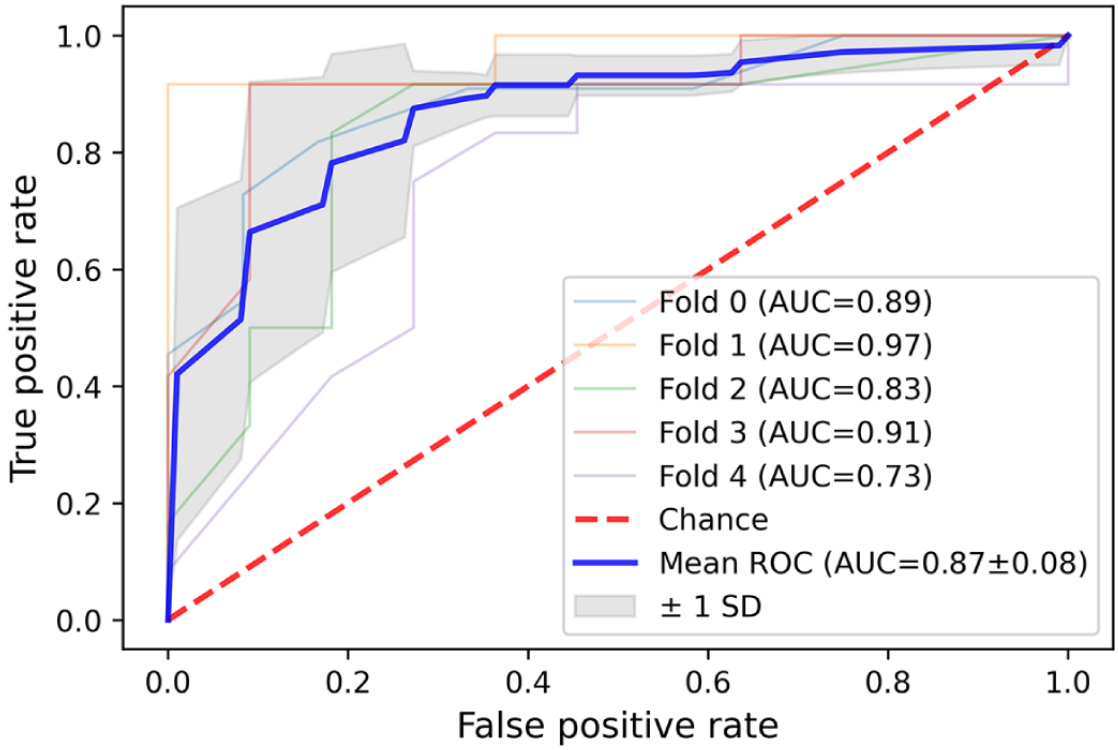

XSL•FO
RenderX

Figure 13. A 5-fold cross-validation curve of receiver operating characteristic of Support Vector Machine on the external test data set. Receiver operating characteristic for management problem classification. AUC: area under the curve; ROC: receiver operating characteristic.

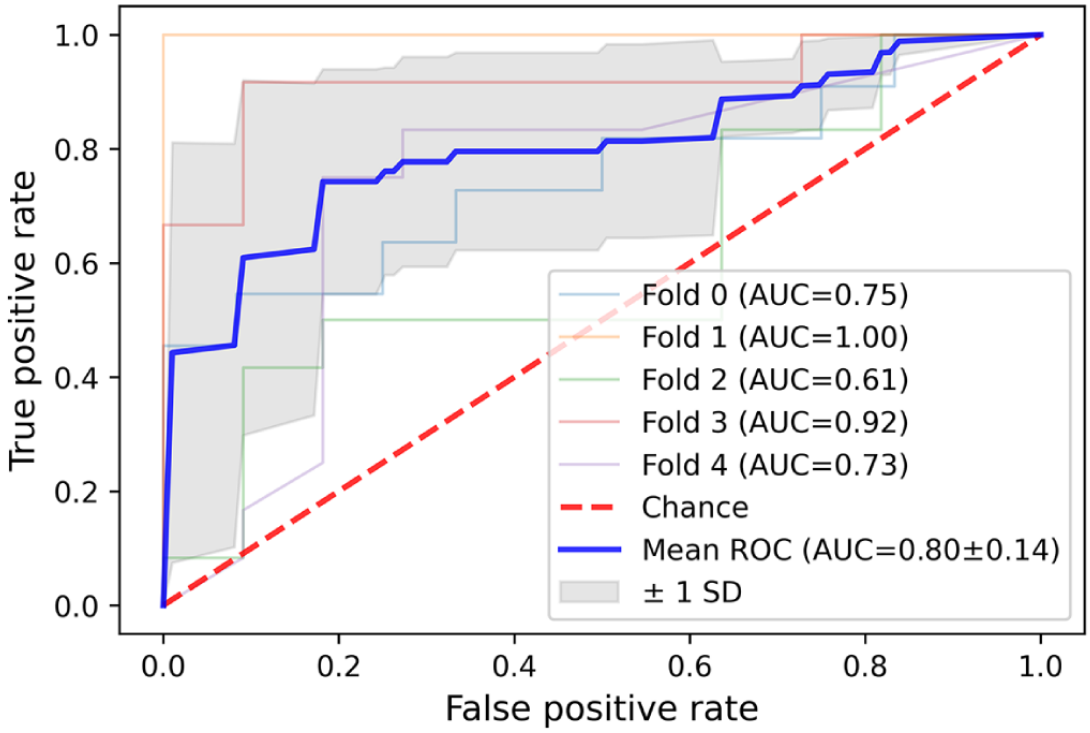

Figure 14. A 5-fold cross-validation curve of receiver operating characteristic of Support Vector Machine on the external test data set. Receiver operating characteristic for sense of responsibility problem classification. AUC: area under the curve; ROC: receiver operating characteristic.

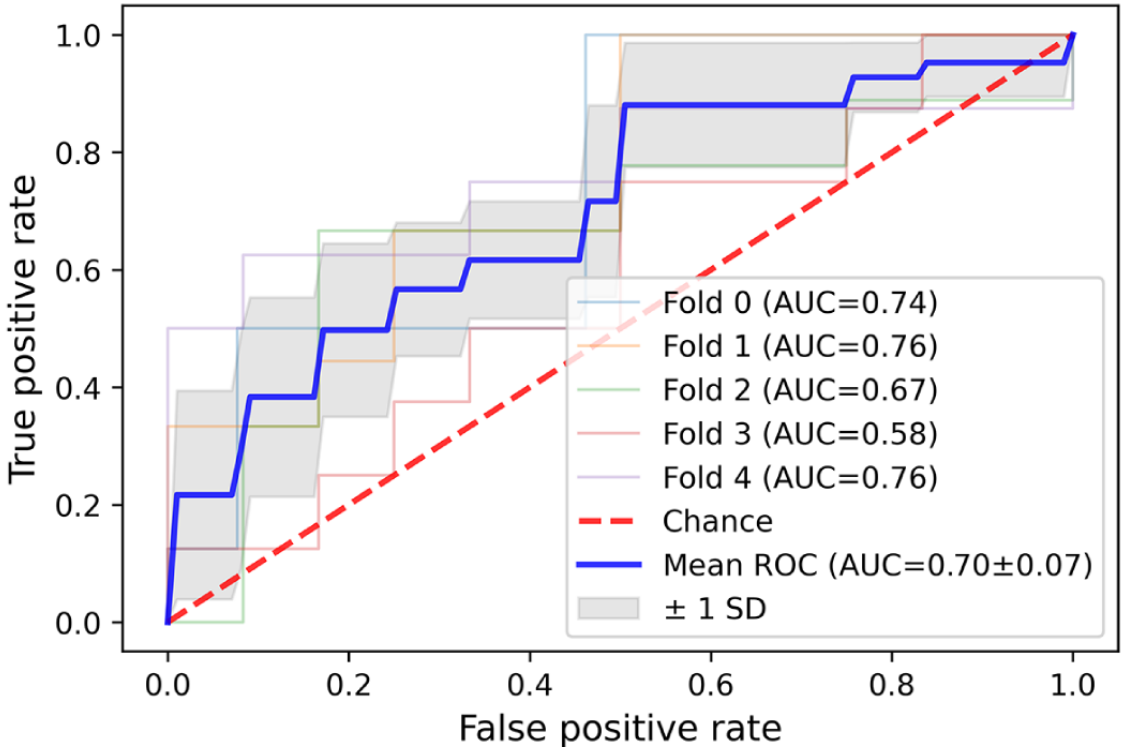

Discussion

Principal Findings

Ambiguities arise more frequently in areas where patients may struggle to clearly articulate their concerns, particularly in subjective categories like communication and responsibility. Word-level TF-IDF proved sufficient for our needs. The result also revealed that Ngram-level methods did not significantly enhance model performance, even after optimal tuning of N. The NLP-driven SVM algorithm effectively classifies patient complaint texts, demonstrating superior performance in both internal and external test sets for communication and management problems.

XSL•FO
RenderX

Related Systems for Patient Complaints Text Classification

Traditional text classification relies on hand-coded keyword-based methods suffer from low efficiency and accuracy. The manual creation of these classifiers (often in the form of expert rules) is a difficult and expensive process [43] that requires the coordinated work of medical experts and knowledge engineers. To the best of our knowledge, there are currently 3 manual approaches to classifying patient complaint texts: The first one is Patient Advocacy Reporting System (PARS), the second mainly focused on Healthcare Complaint Analysis Tool (HCAT [44]), developed by UK researchers Alex Gillespie and Tom Reader, and the last one is based on the use of the qualitative analysis software NVivo [45,46], launched by the Australian company QSR (now Lumivero). However, both HCAT and NVivo require human classification at various stages of the classification process. Therefore, these methods are at best semiautomatic means for text classification. They have not yet fully solved the problem of manual intervention in the classification process [47].

Ngram-Level TF-IDF Failed to Enhance Classification Performance

As gleaned from Table 3, the increment in precision, recall, and F_1-score with Ngram-level TF-IDF (n=2) merely amounted to 1%, rising from 0.91 to 0.92. The reason analysis for why Ngram-level TF-IDF (n=2) did not improve the classification performance of Word-level TF-IDF could be as described further. First, the primary reason might be Ngram-level TF-IDF with n=2 considers pairs of words, capturing some contextual information. However, this may not always lead to significant improvements, especially when text prone to ambiguity undergoes review by language experts, the additional information obtainable from conjunctions becomes relatively limited. The second possible reason was Ngram-level TF-IDF, particularly with smaller datasets (only 2157 cases for training), which could be prone to overfitting. It might capture too much noise or specific patterns from the training data that did not generalize well to unseen data, resulting in worse performance on the test set. For tuning the parameter of Ngram-level TF-IDF, we used the MNB method and set N to values ranging from 1 to 5. The results were depicted in Multimedia Appendix 7. It can be observed that when N=1, the model achieved its maximum values for precision, recall, and F_1-score. Consequently, N=1 for Ngram-level TF-IDF corresponds to using Word-level TF-IDF.

The Feasibility of Using ChatGPT as a Bulk Translation Tool

The process of translating patient complaints into English through ChatGPT could potentially introduce several issues into the classification, especially concerning the accuracy of translation. Common knowledge holds by European and American litterateurs that translating ancient Chinese poetry into English is extremely challenging. However, in a study reported by Gao et al [48], they compared the performances of ChatGPT with Google Translate and DeepL Translator in translating Chinese classical poetry in terms of fidelity, fluency, language style, and machine translation style. The results revealed that ChatGPT outperformed Google Translate and DeepL Translator in all evaluation criteria. Therefore, in our research, the difficulty level of the Chinese text of patient complaints was significantly lower compared with classical poetry. The texts were generally very clear and complete, containing specific elements and structures such as precise time, location, individuals involved, and events. As the complainants were patients and their family members, the language was relatively common, devoid of esoteric medical terminology. Consequently, for ChatGPT 3.5, the translation results were very close to the outcomes of human translations. In addition, 2 postgraduates with English language skills were responsible for reviewing the machine-translated texts. If there were obvious translation errors, the texts were corrected after the 2 individuals proposed unanimous modification suggestions. Therefore, we believe that at least using ChatGPT 3.5 for translating bulk of simple complaint texts into English was acceptable and efficient, as we did not have to struggle with balancing the accuracy and efficiency of the translation.

Appropriate Number and Types of Categories

Reader et al [49] developed a coding taxonomy for analyzing patient complaints. The subcategories were thematically grouped into 7 categories and then 3 conceptually distinct domains. Their "3-domain" classification system is similar to our "4-category" one. For instance, "the safety and quality of clinical care" corresponds to our "diagnosis and treatment," "management of health care organizations" corresponds to our "management problem," and "healthcare staff-patient relationships" corresponds to our "communication problem" and "sense of responsibility."

We subdivided "healthcare staff-patient relationships" into 2 categories because it was considered that communication issues were skill-related problems that can be addressed through targeted training. At the same time, the sense of responsibility is a matter of the staff's character that requires long-term education and may even necessitate stricter rewards and penalties to improve. They also reported that the most common issues complained about were "treatment" (113,738/88,069, 15.6%) and "communication" (12,065/88,069, 13.7%). However, our research shows that complaints mainly stem from "communication" (719/1465, 49.9%) and "diagnosis and treatment" (376/1465, 26.1%), respectively. However, when focusing on categorizing complaints, we must analyze the data sources, as there can be significant differences between patient complaint data from specialized hospitals and data from comprehensive hospitals. Even within the same hospital, there can be significant variations in complaint data between different departments.

Another similar research carried out by HaCohen-Kerner et al [50], showed that their study attempted to reduce the number of categories for automatic text classification from 7 to 4 across 2073 samples. The predictive model's performance reached its peak accuracy (93.8%) on the test set when the number of categories was reduced to 4. Therefore, it was reasonable in our research to limit the predicted categories to 4 in similar sample sizes.

XSL•FO
RenderX

Optimal Number of Folds to Use in a K-Fold Cross-Validation in External Dataset

It is well-accepted that larger k means less bias toward overestimating the true expected error but higher variance and higher running time. In both of these cases, selecting k depends on the same thing. We must ensure that the training set and testing set are drawn from the same distribution and that both sets contain sufficient variation such that the underlining distribution is represented. The problem we encountered involved using training and verification data sourced from a general hospital and external test data from a specialized cancer hospital. Discrepancies in patient treatment expectations and compliance between the 2 hospitals may introduce biases in the distribution of complaint data. Consequently, we opted to use CV testing during external testing to mitigate the potential impact of these factors on data representation.

Selecting k is not an exact science because it is challenging to find a balance in a fold of data that effectively represents the overall dataset while also minimizing data bias. We try to use a 5-fold cross validation, in this case, with approximately 75 (376×0.2) instances per fold, conducting 10-folds CV would result in only about 37 instances per fold. As a result, all 4 types of complaints may not be encompassed within a single fold. However, if our dataset size increases dramatically, like if we have over 10,000 instances, a 10-fold cross validation would lead in folds of 1000 instances. In a related work carried by Oyedele [51], a bigger independent testing set of size 10,000 (independently sampled from the same distribution and using the same parameters with 5000 samples per class) was created in order to estimate the performance (ie, error probability) of the network. Therefore, for a task with 2 classifications and 10,000 samples, they adopt a 10-fold CV, which poses much less risk of bias in the results compared with our 5-fold CV approach.

Advantages of Using TF-IDF and SVM for Patient Complaint Text Classification

Elmessiry et al [52] compared a text classification method using 6 ML classifiers and suggested that the TF-IDF–based random forest classification algorithm has advantages in text classification both in accuracy and F_1-score. However, based on their results, the achieved accuracy, F_1-score, sensitivity, and specificity in correctly classifying patient complaints were relatively low, with values of 82%, 81%, 0.76, and 0.87 respectively.

Previous studies have used ML algorithms such as SVMs [53], Logistic Regression [54], and Naive Bayes classifiers [55] for classification purposes [56]. Compared with other ML algorithms, the SVM algorithm has demonstrated superior performance in text classification, which aligns closely with our research findings. The strong generalization capability of SVM in high-dimensional feature spaces can be credited for their effectiveness in text classification tasks. SVMs eliminate the need for feature selection, simplifying the classification process. Their regularization parameters help prevent overfitting, making them robust in handling noisy data and datasets with many features. In addition, SVMs are efficient in processing small to medium-sized datasets, producing reliable performance even with limited training data. Our research also confirmed that, even with very limited sample sizes, such as in the cases of “management problem” and “sense of responsibility” with raw data sizes of only 260 and 86, the SVM algorithm achieved average AUC scores of 0.81 and 0.7, respectively, in external test sets.

In the text preprocessing and ML section, we used the pipeline training architecture, which can be seen as a tool to connect multiple data processing and model training steps. It allows users to combine multiple operations in a specific order to form a complete flow and ensures data flows between each operation. Through pipeline, data preprocessing, feature engineering, model training, model evaluation, and other steps can be organically combined into a complete ML process. The advantage of doing this is improving code maintainability and reusability, making it easier to perform the same workflow on different datasets and better managing dependencies between each step.

We initialized SVM parameters as listed in Multimedia Appendix 8 and adopted a fine-tuning with genetic algorithm–SVM as proposed by Ali et al [57] and Nair et al [58], which has the best model performance while keeping a reasonable optimization time. The detailed tuning process and prediction results before and after tuning have been illustrated in Multimedia Appendices 9 and 10).

ML Model Selection Strategy

A critical consideration is the balance between model accuracy and model simplicity or interpretability. High accuracy is essential for a model’s effectiveness, especially in scenarios that rely on making correct predictions. While complex models (such as DL or ensemble methods) can achieve higher accuracy due to their ability to capture intricate patterns in the data, they often require large amounts of data and substantial computational resources. Models with high interpretability foster trust and accountability, as stakeholders can see and understand how decisions are made. However, there is a trade-off here; complex models may exhibit lower bias but higher variance, meaning they might fit the training data very well but perform poorly on unseen data due to overfitting. Conversely, simpler models tend to have higher bias, potentially missing significant patterns in the data, but they perform more consistently across different datasets.

In this study, a hybrid strategy was adopted in ML model selection. In the first stage, 3 ML algorithms (MLR, MNB, and SVM) were used for model training and validation. In the second stage, the best-performing model was tested using 5-fold CV on external data. Ultimately, the choice of model should be guided by the specific context of the problem, including the importance of accuracy versus interpretability, the stakes involved, and the audience for the model outputs. In the specific task of patient complaint classification, we prioritize the generalization ability of the model, which means achieving the best prediction accuracy on unseen data. This preference is crucial because if the model’s generalization is insufficient, misclassified complaint cases will require more human effort to resolve, and the resulting delay in complaint resolution might exacerbate patient-doctor conflicts and disputes.

XSL•FO
RenderX

Limitations and Further Research Directions

In this study, the original data were initially in Microsoft Excel format, and each row of complaint content was separately saved as a text file before subsequent data cleaning and analysis were performed. Consequently, similar processing was required during application to ensure compatibility with the model. Failure to align the actual input data format with that of the model could lead to the inability of the model to predict outcomes correctly. Therefore, in subsequent work, a complaint text organization code will be developed to support both individual complaint text prediction through copy-and-paste operations and batch text processing.

It is noteworthy that despite ultimately selecting the SVM algorithm for validating external data, it displayed consistent fluctuations in learning accuracy and verification loss throughout in training iteration. We attempted to improve the situation by reducing the learning rate from 0.5 to 0.1, but this did not yield positive results. After ruling out factors such as randomness and model complexity, we speculate that it is likely due to the SVM algorithm using SGD optimization, which can cause the model to converge to local minimum or exhibit more significant fluctuations in results. Another possibility may be the imbalance in our raw data distribution, with the third class, “management problem,” and the fourth class, “sense of responsibility,” having too few data samples.

The primary challenge addressed in this study pertains to the imbalanced distribution of raw data and the limited dataset size. Although we used SMOTE to adjust for imbalanced data, the synthetic samples generated by SMOTE may introduce noise or distortions into the dataset, potentially leading to information loss and affecting the classifier’s performance on the minority class. To address this issue, our future work will involve collaborative efforts with multiple hospitals to enhance data volume. Furthermore, as is well-known, more advanced NLP models (eg, BERT) provide contextual understanding, enabling them to grasp complex relationships within text and improve analysis accuracy and depth. As a future direction for our work, we plan to develop a user-friendly graphical user software interface powered by BERT to facilitate the classification and management of complaints by PAC personnel.

Another issue that must be stated is that the semantic features extracted using NLP technology are quite different from the common scenarios based on explicit independent and dependent variables. In ML problems with clear relationships between independent and dependent variables, it is possible to rank the importance of independent variables, as these features have stable meanings themselves. However, semantic features are influenced by the linguistic context, meaning that the same word can have different meanings in different language environments and contexts. Therefore, our results cannot pinpoint which specific words significantly influence the classification of complaint text types.

Conclusion

The NLP combined SVM algorithm performs well in automatically categorizing patient complaint texts. Furthermore, it has the best performance in both internal and external test sets for communication and management problems. However, caution is necessary when used for the classification of sense of responsibility. It has excellent prospects for application in medical institutions with many complaints and a shortage of medical-patient processing specialists.

Acknowledgments

This research was funded by the Natural Science Foundation of Zhejiang Province, China (grant LGF22H220007); and Hangzhou Health Science and Technology Project, China (grant A20200746).

Data Availability

Data that support the findings of the study are available upon request from the corresponding author.

In addition, we have provided the important Python code related to this study in Multimedia Appendix 11 for interested readers to download and use.

Authors' Contributions

Conceptualization was contributed by X Li, X Lou, and XF. Methodology was contributed by CK and GL. Validation was contributed by X Lou. Formal analysis was contributed by X Li. Writing—review and editing was contributed by XF. All authors have read and agreed to the published version of the manuscript. Authors GY (yugbme@zju.edu.cn) and XL (lxd2019@zju.edu.cn) are co-corresponding authors for this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Review and Prospect of Multi-Label Text Classification Research.
[PDF File (Adobe PDF File), 2923 KB - jmir_v27i1e55721_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
A method and system for classifying short text of product names based on attention mechanism: Part a.
[PDF File (Adobe PDF File), 2989 KB - jmir_v27i1e55721_app2.pdf]

Multimedia Appendix 3
A method and system for classifying short text of product names based on attention mechanism: Part b.
[PDF File (Adobe PDF File), 2979 KB - jmir_v27i1e55721_app3.pdf]

Multimedia Appendix 4
High frequency but meaningless stop words related to medical complaints.
[DOCX File , 18 KB - jmir_v27i1e55721_app4.docx]

Multimedia Appendix 5
Comparison of the Performance of Three Machine Learning Algorithms: Point Estimate and 95% Confidence Intervals.
[DOCX File , 20 KB - jmir_v27i1e55721_app5.docx]

Multimedia Appendix 6
Point Estimate and 95% CIs of each model's performance metrics in distinguishing the four categories.
[DOCX File , 19 KB - jmir_v27i1e55721_app6.docx]

Multimedia Appendix 7
Precision, Recall, and F1-score on different N in Ngram-level TF-IDF (term frequency–inverse document frequency).
[PNG File , 367 KB - jmir_v27i1e55721_app7.png]

Multimedia Appendix 8
Initialized Support Vector Machine parameters.
[DOCX File , 18 KB - jmir_v27i1e55721_app8.docx]

Multimedia Appendix 9
The fine tuning hyperparameters.
[PNG File , 4420 KB - jmir_v27i1e55721_app9.png]

Multimedia Appendix 10
The fine tuning hyperparameters.
[PNG File , 3965 KB - jmir_v27i1e55721_app10.png]

Multimedia Appendix 11
Core Python code used in this study.
[RAR File , 14 KB - jmir_v27i1e55721_app11.rar]

References

1. Yoder-Wise PS. Patient complaints: not ours to own, but ours to fix. J Contin Educ Nurs 2015;46(12):531-532. [doi: 10.3928/00220124-20151112-01] [Medline: 26641145]
2. Wofford MM, Wofford JL, Bothra J, Kendrick SB, Smith A, Lichstein PR. Patient complaints about physician behaviors: a qualitative study. Acad Med 2004;79(2):134-138. [doi: 10.1097/00001888-200402000-00008] [Medline: 14744713]
3. Hardavella G, Aamli-Gaagnat A, Saad N, Rousalova I, Sreter KB. How to give and receive feedback effectively. Breathe (Sheff) 2017;13(4):327-333 [FREE Full text] [doi: 10.1183/20734735.009917] [Medline: 29209427]
4. 4 W. Zhang, X. Xi, Z. Cui, Y. Zou and J. Review and Prospect of Multi-Label Text Classification Research. 2023. URL: https://caod.oriprobe.com/articles/65876090/Review_and_Prospect_of_Multi_Label_Text_Classifica.htm [accessed 2023-12-29]
5. Rao Y, Yang F. A method for classifying information in education policy texts based on an improved attention mechanism model. Wirel. Commun. Mob. Comput 2022;2(1) [FREE Full text] [doi: 10.1155/2022/5467572]
6. 6 T. Schopf, D. 2023 Jan 18 Presented at: Braun and F. Matthes17th International Conference on Web Information Systems and Technologies (WEBIST) Electr Network 2023 Oct 2021 2023 Pages; Oct 26, 2021 - Oct 28, 2021; Valletta, Malta. [doi: 10.1007/978-3-031-24197-0]

XSL•FO
RenderX

7. Wu H, Liu Y, Wang J. Review of text classification methods on deep learning. Cmc-Computers Materials & Continua 2020;63(3):1309-1321. [doi: 10.32604/cmc.2020.010172]
8. Gao H, Zeng X, Yao C. Application of improved distributed naive Bayesian algorithms in text classification. J Supercomput 2019;75(3):5831-5847. [doi: 10.1007/s11227-019-02862-1]
9. Ma Y, Li Y, Wu X, Zhang X. Chinese text classification review. 2018 Presented at: 9th International Conference on Information Technology in Medicine and Education (ITME); 2018 October 19-21; Hangzhou, China.
10. Branca Vergano L, Monesi M, Vicenti G, Bizzoca D, Solarino G, Moretti B. Posterior approaches in malleolar fracture: when, why and how. J Biol Regul Homeost Agents 2020;34(3 Suppl. 2):89-95. ADVANCES IN MUSCULOSKELETAL DISEASES AND INFECTIONS. [Medline: 32856446]
11. Chapman WW, Dowling JN, Wagner MM. Classification of emergency department chief complaints into 7 syndromes: a retrospective analysis of 527,228 patients. Ann Emerg Med 2005;46(5):445-455. [doi: 10.1016/j.annemergmed.2005.04.012] [Medline: 16271676]
12. Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. Artif Intell Med 2005;33(1):31-40. [doi: 10.1016/j.artmed.2004.04.001] [Medline: 15617980]
13. Lee SH, Levin D, Finley PD, Heilig CM. Chief complaint classification with recurrent neural networks. J Biomed Inform 2019;93:103158 [FREE Full text] [doi: 10.1016/j.jbi.2019.103158] [Medline: 30926471]
14. Karagounis S, Sarkar IN, Chen ES. Coding free-text chief complaints from a health information exchange: a preliminary study. AMIA Annu Symp Proc 2020;2020:638-647 [FREE Full text] [Medline: 33936438]
15. Tootooni MS, Pasupathy KS, Heaton HA, Clements CM, Sir MY. CCMapper: an adaptive NLP-based free-text chief complaint mapping algorithm. Comput Biol Med 2019;113:103398. [doi: 10.1016/j.compbiomed.2019.103398] [Medline: 31454613]
16. Arnaud E, Elbattah M, Gignon M, Dequen G. NLP-based prediction of medical specialties at hospital admission using triage notes. : IEEE; 2021 Presented at: 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI); 2021 August 09-12; Victoria, BC, Canada URL: https://doi.org/10.1109/ICHI52183.2021.00103 [doi: 10.1109/ichi52183.2021.00103]
17. Stewart J, Lu J, Goudie A, Arendts G, Meka SA, Freeman S, et al. Applications of natural language processing at emergency department triage: a narrative review. PLoS One 2023;18(12):e0279953 [FREE Full text] [doi: 10.1371/journal.pone.0279953] [Medline: 38096321]
18. Branca Vergano L, Monesi M, Vicenti G, Bizzoca D, Solarino G, Moretti B. Posterior approaches in malleolar fracture: when, why and how. J Biol Regul Homeost Agents 2020;34(3 Suppl. 2):89-95. ADVANCES IN MUSCULOSKELETAL DISEASES AND INFECTIONS. [Medline: 32856446]
19. Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS One 2021;16(2):e0245909 [FREE Full text] [doi: 10.1371/journal.pone.0245909] [Medline: 33630869]
20. Karakaya O, Kilimci ZH. An efficient consolidation of word embedding and deep learning techniques for classifying anticancer peptides: FastText+BiLSTM. PeerJ Comput Sci 2024;10:e1831 [FREE Full text] [doi: 10.7717/peerj-cs.1831] [Medline: 38435607]
21. Alhazzani NZ, Al-Turaiki IM, Alkhodair SA. Text classification of patient experience comments in Saudi dialect using deep learning techniques. Applied Sciences 2023;13(18):10305 [FREE Full text] [doi: 10.3390/app131810305]
22. Cascella M, Semeraro F, Montomoli J, Bellini V, Piazza O, Bignami E. The breakthrough of large language models release for medical applications: 1-year timeline and perspectives. J Med Syst 2024;48(1):22 [FREE Full text] [doi: 10.1007/s10916-024-02045-3] [Medline: 38366043]
23. Peng C, Yang X, Chen A, Yu Z, Smith KE, Costa AB, et al. Generative large language models are all-purpose text analytics engines: text-to-text learning is all your need. J Am Med Inform Assoc 2024;31(9):1892-1903. [doi: 10.1093/jamia/ocae078] [Medline: 38630580]
24. Valmianski, I, Goodwin C, Finn IM, Khan N, Zisook DS. Evaluating robustness of language models for chief complaint extraction from patient-generated text. arXiv:1911.06915 2019 [FREE Full text]
25. Jones KS. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 1972;28(1):11-21. [doi: 10.1108/eb026526]
26. Lu H, Ehwerhemuepha L, Rakovski C. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. BMC Med Res Methodol 2022;22(1):181 [FREE Full text] [doi: 10.1186/s12874-022-01665-y] [Medline: 35780100]
27. Kim Y. Convolutional neural networks for sentence classification. 2014 Presented at: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); November 28, 2024; Doha, Qatar p. 1746-1751. [doi: 10.3115/v1/d14-1181]
28. Wang R, Li Z, Cao J, Chen T, Wang L. Convolutional recurrent neural networks for text classification. : IEEE; 2019 Presented at: 2019 International Joint Conference on Neural Networks (IJCNN); July 14-19, 2019; Budapest, Hungary p. 1-6. [doi: 10.1109/ijcnn.2019.8852406]

XSL•FO
RenderX

29. Mccallum A, Nigam K. A comparison of event models for naive Bayes text classification. AAAI-98 Workshop on Learning for text Categorization. 1998. URL: https://cdn.aaai.org/Workshops/1998/WS-98-05/WS98-05-007.pdf [accessed 2024-12-04]
30. Lin CY, Hovy E. Automatic evaluation of summaries using N-gram co-occurrence statistics. New York: ACM; 2003 Presented at: NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology; 2003 May 27; Edmonton, Alberta, Canada. [doi: 10.3115/1073445.1073465]
31. Zhang W, Yoshida T, Tang X. TFIDF, LSI and multi-word in information retrieval and text categorization. Piscataway: IEEE; 2009 Presented at: 2008 IEEE International Conference on Systems, Man and Cybernetics; 2008 October 12-15; Singapore p. 108-113. [doi: 10.1109/icsmc.2008.4811259]
32. Joachims T. Text categorization with support vector machines: learning with many relevant features. 1998 Presented at: ECML'98: Proceedings of the 10th European Conference on Machine Learning; 1998 April 21-23; Berlin, Heidelberg p. 137-142. [doi: 10.1007/bfb0026683]
33. Genkin A, Lewis DD, Madigan D. Large-scale Bayesian logistic regression for text categorization. Technometrics 2007;49(3):291-304. [doi: 10.1198/004017007000000245]
34. Xu X, Chang Y, An J, Du Y. Chinese text classification by combining Chinese-BERTology-wwm and GCN. PeerJ Comput Sci 2023;9:e1544 [FREE Full text] [doi: 10.7717/peerj-cs.1544] [Medline: 37705631]
35. Li Y, Wang X, Xu P. Chinese text classification model based on deep learning. Future Internet 2018;10(11):113 [FREE Full text] [doi: 10.3390/fi10110113]
36. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
37. Li Z, Sun M. Punctuation as implicit annotations for Chinese word segmentation. Computational Linguistics 2009;35(4):505-512. [doi: 10.1162/coli.2009.35.4.35403]
38. Beauprez SA, Bidet-Ildei CB, Hiraki K. Does watching Han Solo or C-3PO similarly influence our language processing? Psychol Res 2020;84(6):1572-1585. [doi: 10.1007/s00426-019-01169-3] [Medline: 30931488]
39. Zhou Y, Huang F, Chen H. Combining probability models and web mining models: a framework for proper name transliteration. Inf Technol Manage 2007;9(2):91-103. [doi: 10.1007/s10799-007-0031-9]
40. Bansal A, Saini M, Singh R, Yadav JK. Analysis of SMOTE: modified for diverse imbalanced datasets under the IoT environment. International Journal of Information Retrieval Research 2021;11(2):23. [doi: 10.4018/ijirr.2021040102]
41. Chen Z, Li X, Li J, Zhang S, Zhou P, Yu X, et al. A COVID-19 risk score combining chest CT radiomics and clinical characteristics to differentiate COVID-19 pneumonia from other viral pneumonias. Aging (Albany NY) 2021;13(7):9186-9224 [FREE Full text] [doi: 10.18632/aging.202735] [Medline: 33713401]
42. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. JAIR 2002;16:321-357. [doi: 10.1613/jair.953]
43. Wilcox AB, Hripcsak G. The role of domain knowledge in automating medical text report classification. J Am Med Inform Assoc 2003;10(4):330-338 [FREE Full text] [doi: 10.1197/jamia.M1157] [Medline: 12668687]
44. Jerng JS, Huang SF, Yu HY, Chan YC, Liang HJ, Liang HW, et al. Comparison of complaints to the intensive care units and those to the general wards: an analysis using the Healthcare Complaint Analysis Tool in an academic medical center in Taiwan. Crit Care 2018;22(1):335 [FREE Full text] [doi: 10.1186/s13054-018-2271-y] [Medline: 30522508]
45. Dhakal K. NVivo. J Med Libr Assoc 2022;110(2):270-272 [FREE Full text] [doi: 10.5195/jmla.2022.1271] [Medline: 35440911]
46. Wang B, Lin Q, Wang Y, Tang S. Policy analysis of the integration of sports and medicine against the backdrop of "Healthy China": a qualitative study using NVivo. Int J Environ Res Public Health 2023;20(3):2079 [FREE Full text] [doi: 10.3390/ijerph20032079] [Medline: 36767445]
47. Yamanouchi Y, Nakamura T, Ikeda T, Usuku K. An alternative application of natural language processing to express a characteristic feature of diseases in Japanese medical records. Methods Inf Med 2023;62(3-04):110-118 [FREE Full text] [doi: 10.1055/a-2039-3773] [Medline: 36809794]
48. Gao R, Lin Y, Zhao N, Cai ZG. Machine translation of Chinese classical poetry: a comparison among ChatGPT, Google Translate, and DeepL Translator. Humanit Soc Sci Commun 2024 Jun 26;11(1). [doi: 10.1057/s41599-024-03363-0]
49. Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. BMJ Qual Saf 2014;23(8):678-689 [FREE Full text] [doi: 10.1136/bmjqs-2013-002437] [Medline: 24876289]
50. HaCohen-Kerner Y, Dilmon R, Hone M, Ben-Basan MA. Automatic classification of complaint letters according to service provider categories. Information Processing & Management 2019;56(6):102102 [FREE Full text] [doi: 10.1016/j.ipm.2019.102102]
51. Oyedele O. Determining the optimal number of folds to use in a K-fold cross-validation: a neural network classification experiment. Research in Mathematics 2023;10(1). [doi: 10.1080/27684830.2023.2201015]
52. Elmessiry A, Cooper WO, Catron TF, Karrass J, Zhang Z, Singh MP. Triaging patient complaints: Monte Carlo cross-validation of six machine learning classifiers. JMIR Med Inform 2017;5(3):e19 [FREE Full text] [doi: 10.2196/medinform.7140] [Medline: 28760726]

XSL•FO
RenderX

53. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273-297. [doi: 10.1007/BF00994018]
54. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. JAMA 2016;316(5):533-534. [doi: 10.1001/jama.2016.7653] [Medline: 27483067]
55. McCallum A. Graphical Models, Lecture2: Bayesian Network Representation. 2019. URL: https://people.cs.umass.edu/~mccallum/courses/gm2011/02-bn-rep.pdf [accessed 2022-10-09]
56. Vakharia V, Shah M, Suthar V, Patel VK, Solanki A. Hybrid perovskites thin films morphology identification by adapting multiscale-SinGAN architecture, heat transfer search optimized feature selection and machine learning algorithms. Phys. Scr 2023;98(2):025203. [doi: 10.1088/1402-4896/acae49]
57. Ali YA, Awwad EM, Al-Razgan M, Maarouf A. Hyperparameter search for machine learning algorithms for optimizing the computational complexity. Processes 2023;11(2):349 [FREE Full text] [doi: 10.3390/pr11020349]
58. Nair P, Vakharia V, Shah M, Kumar Y, Woźniak M, Shafi J, et al. AI-driven digital twin model for reliable lithium-ion battery discharge capacity predictions. International Journal of Intelligent Systems 2024;2024:1-18 [FREE Full text] [doi: 10.1155/2024/8185044]

Abbreviations

AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
BiGRU: Bidirectional Gated Recurrent Unit
BiLSTM: Bidirectional Long Short-Term Memory
BOW: bag-of-words
CC: chief complaints
CNN: convolutional neural network
CV: cross-validation
DL: deep learning
ED: emergency department
GRU: gated recurrent unit
HMM-Bigram: Hidden Markov Model Bigram
HCAT: Healthcare Complaint Analysis Tool
LSTM: long short-term memory
ML: machine learning
MLR: Multifactor Logistic Regression
MNB: Multinomial Naive Bayes
NLP: natural language processing
PAC: Patient Advocacy Center
PARS: Patient Advocacy Reporting System
RNN: recurrent neural networks
ROC: receiver operating characteristic
SGD: Stochastic Gradient Descent
SMOTE: Synthetic Minority Oversampling Technique
SVM: Support Vector Machine
TF-IDF: term frequency–inverse document frequency

Edited by A Coristine; submitted 21.12.23; peer-reviewed by J Liang, M Elbattah, TAR Sure, V Vakharia; comments to author 12.03.24; revised version received 28.04.24; accepted 04.11.24; published 08.01.25.

Please cite as:
Li X, Shu Q, Kong C, Wang J, Li G, Fang X, Lou X, Yu G
An Intelligent System for Classifying Patient Complaints Using Machine Learning and Natural Language Processing: Development and Validation Study
J Med Internet Res 2025;27:e55721
URL: https://www.jmir.org/2025/1/e55721
doi:10.2196/55721
PMID:

©Xiadong Li, Qiang Shu, Canhong Kong, Jinhu Wang, Gang Li, Xin Fang, Xiaomin Lou, Gang Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.01.2025. This is an open-access article distributed under the

XSL•FO
RenderX

terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Revealing Patient Dissatisfaction With Health Care Resource Allocation in Multiple Dimensions Using Large Language Models and the International Classification of Diseases 11th Revision: Aspect-Based Sentiment Analysis

Jiaxuan Li[1], MSc; Yunchu Yang[1], MSc; Chao Mao[1], MSc; Patrick Cheong-Iao Pang[1], PhD; Quanjing Zhu[2], MSc; Dejian Xu[1], BSc; Yapeng Wang[1], PhD

[1]Faculty of Applied Sciences, Macao Polytechnic University, Macao, Macao

[2]Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Patrick Cheong-Iao Pang, PhD
Faculty of Applied Sciences
Macao Polytechnic University
Rua de Luís Gonzaga Gomes
Macao, 999078
Macao
Phone: 853 85996886
Email: mail@patrickpang.net

Abstract

Background: Accurately measuring the health care needs of patients with different diseases remains a public health challenge for health care management worldwide. There is a need for new computational methods to be able to assess the health care resources required by patients with different diseases to avoid wasting resources.

Objective: This study aimed to assessing dissatisfaction with allocation of health care resources from the perspective of patients with different diseases that can help optimize resource allocation and better achieve several of the Sustainable Development Goals (SDGs), such as SDG 3 (“Good Health and Well-being”). Our goal was to show the effectiveness and practicality of large language models (LLMs) in assessing the distribution of health care resources.

Methods: We used aspect-based sentiment analysis (ABSA), which can divide textual data into several aspects for sentiment analysis. In this study, we used Chat Generative Pretrained Transformer (ChatGPT) to perform ABSA of patient reviews based on 3 aspects (patient experience, physician skills and efficiency, and infrastructure and administration)00 in which we embedded chain-of-thought (CoT) prompting and compared the performance of Chinese and English LLMs on a Chinese dataset. Additionally, we used the *International Classification of Diseases 11th Revision* (ICD-11) application programming interface (API) to classify the sentiment analysis results into different disease categories.

Results: We evaluated the performance of the models by comparing predicted sentiments (either positive or negative) with the labels judged by human evaluators in terms of the aforementioned 3 aspects. The results showed that ChatGPT 3.5 is superior in a combination of stability, expense, and runtime considerations compared to ChatGPT-4o and Qwen-7b. The weighted total precision of our method based on the ABSA of patient reviews was 0.907, while the average accuracy of all 3 sampling methods was 0.893. Both values suggested that the model was able to achieve our objective. Using our approach, we identified that dissatisfaction is highest for sex-related diseases and lowest for circulatory diseases and that the need for better infrastructure and administration is much higher for blood-related diseases than for other diseases in China.

Conclusions: The results prove that our method with LLMs can use patient reviews and the ICD-11 classification to assess the health care needs of patients with different diseases, which can assist with resource allocation rationally.

(J Med Internet Res 2025;27:e66344) doi:10.2196/66344

XSL•FO
RenderX

KEYWORDS

ICD-11; International Classification of Diseases 11th Revision; disease classification; patient reviews; patient satisfaction; ChatGPT; Sustainable Development Goals; chain of thought; large language model

Introduction

With the imbalance in regional economic development, there are 2 extreme problems in the accurate assessment of medical resources. On the one hand, health care management cannot accurately measure patients' health care needs [1]. On the other hand, the assessment of medical resources cannot be accurate for each clinical department in hospitals, which leads to wastage of health care resources [2]. Research has shown that the improvement of health care resources enhances patient satisfaction [3]. To promote the sustainable development of the world, the United Nations has been advocating 17 Sustainable Development Goals (SDGs) since 2015. SDG 3 ("Good Health and Well-being") and SDG 10 ("Reducing Inequalities") suggest that the realization of universal health care coverage and avoiding lack or wastage of health care resources are urgent tasks [4-6]. SDG 9 ("Industry, Innovation and Infrastructure") is about maintaining scientific research and innovation with the sustainable development of health care infrastructure and resources [7]. Overall, it is crucial to guarantee equitable access to health care services for all [8]. In this context, it is particularly important to assess health care resources based on patient needs within specific clinical departments [9]. For this reason, in this study, we planned to analyze patient dissatisfaction through their reviews after consultations to understand a macro view of their health care needs.

Chat Generative Pretrained Transformer (ChatGPT), a representative model of large language models (LLMs), was introduced in November 2022 [10], and its responsiveness as well as performance make it a possibility for further scientific research [11]. Several studies have pointed out that LLMs show strong performance in medical [12], financial [13], and education [14] fields, especially in the medical field, where ChatGPT shows reliable accuracy in questioning dialogues [15], diagnosing illnesses [16], suggesting treatments [17], and predicting prognosis [18]. Therefore, we wanted introduce LLMs into this health care needs study because with the powerful capability of LLMs, we can accomplish an in-depth assessment of health care resources.

Traditional sentiment analysis can determine whether the sentiment of a text is negative or positive, but it cannot determine sentiment from multiple perspectives [19]. However, some studies have shown that using aspect-based sentiment analysis (ABSA) can determine the affective state of a text from multiple perspectives, which is more suitable for us to determine the different perspectives of a patient's health care needs [20-22], and therefore, we adopted ABSA for this study [20]. In this work, we proposed ABSA prompts that include "patient experience," "physician skills and efficiency," and "infrastructure and administration." ABSA can more accurately indicate which of the patients' health care needs are not being met by refining the direction of patient dissatisfaction with health care services, coupled with modern macro analysis of big data, in order to assess overall health care resources. However, in specific clinical application scenarios, different diseases have different disease characteristics, as well as their own different functional managers [23,24], so it is clinically important to analyze data from patient reviews of different disease types.

The *International Classification of Diseases 11th Revision* (ICD-11) coding data have been widely used in many fields [25], such as health statistics, public health monitoring, health care performance evaluation, and reimbursement [26]. ICD-11 is the latest international health information standard, adopted at the 72nd World Health Assembly in 2019, and was implemented in 2022. ICD-11 is a modern, integrated classification and nomenclature system with a modern digital medical model with application programming interfaces (APIs). One of the critical reasons why the World Health Organization (WHO) wants countries to move to ICD-11 is that a standardized classification based on the latest medical and scientific knowledge provides the best access to valuable data for prevention, resource allocation, or evaluation [2]. Several studies have shown that ICD-11 can be integrated with electronic medical records, and WHO's multiple tools and guidelines can reduce training and implementation costs. Therefore, it is critical to expand ICD-11 research and program implementation worldwide [27].

As such, using ICD-11 to classify patients with different diseases, together with our ABSA prompts, it is possible to find differences in the satisfaction and dissatisfaction related to health care resource allocation among patients with different diseases. However, since the classification and statistical standards of medical diseases vary from one health care institution to another, normalization using ICD-11 as the standard will allow comparison of the same disease types to assess the balance of medical resources, as well as a comparison of different disease types to evaluate the differences in health care needs. By standardizing the classification of diseases, it will be possible to not only collect as much valid data as possible under the same standard but also conduct comparative studies of health care resources more effectively. Some researchers have indicated that they tested the coding accuracy of ICD-11 in a large-scale pilot study in 59 public hospitals in China. The results indicated that ICD-11 is well in line with Chinese disease classification habits, with the accuracy of the total coding domains and the accuracy of the word stem codes reaching 82.9% and 92.2%, respectively [28].

In this study, we analyzed patient comments using a LLM and determined the health care medical needs of patients with different diseases using ICD-11. We also discussed the reasons for these results in the context of our findings and the potential applications in clinical settings.

XSL•FO
RenderX

Methods

Architecture of the Experiment

Figure 1 illustrates the proposed architecture for identifying medical needs from patient comments. In this architecture, we used our ABSA prompts in conjunction with chain-of-thought (CoT) prompting to generate questions fed into an LLM to analyze the content of patient comments. Considering the practicality and cost, we chose ChatGPT 3.5 Turbo as the LLM for our experiments as it is the most cost-effective choice. In the output, the model should be able to provide 3 categories of sentiment results. For example, the sample review presented in Figure 1 shows a positive sentiment for the category "patient experience" but a negative sentiment for the categories "physician skills and efficiency" and "infrastructure and administration." Finally, we preprocessed the patients' disease names through ChatGPT and typed the disease codes through the ICD-11 API. For example, the disease code 1B1Z returned from the API suggested that the disease belongs to Chapter 01 (the first digits of the code) of ICD-11.

Figure 1. Proposed architecture for ABSA for analyzing patient data across different diseases using ChatGPT and the ICD-11 API. ABSA: aspect-based sentiment analysis; API: application programming interface; ChatGPT: Chat Generative Pretrained Transformer; CoT: chain of thought; ICD-11: *International Classification of Diseases 11th Revision*.

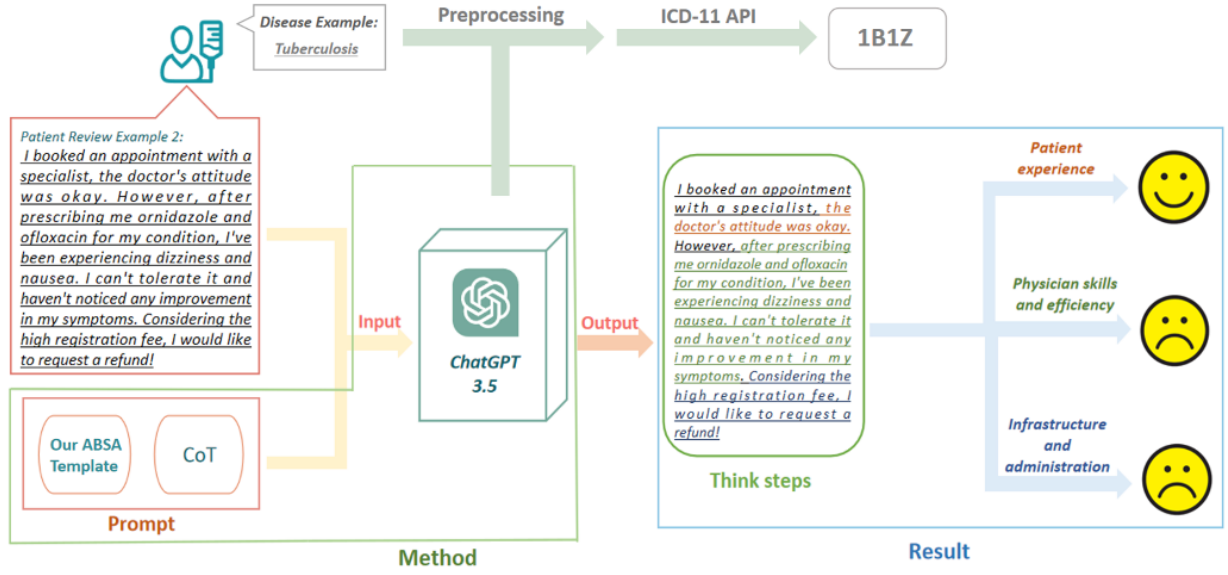

Data

The data used in this study came from Haodf Online [29], which is the most well-known online health care website in mainland China, with more than 6.1 million patients having uploaded their visit experiences [30]. Other studies have pointed out the data from online health care websites as important and valuable for research [31,32]. In addition, a study has shown that ChatGPT can directly process Chinese data without additional translation [33]. The comments contain the name of the disease, the type of treatment, the treatment outcome, the date of the visit, the location, the hospital name, the physician's name, and patient comments. The datasets are in Chinese. The platform validates the comments to ensure that actual patients publish them.

We analyzed 552,764 comments on Haodf Online from November 2022 to June 2023. The data were not manually screened or deleted and spanned from October 2003 to June 2023. A total of 504,198 comments (91.2%) were retained after initial cleaning, including removing incomplete and duplicate data.

Aspect-Based Sentiment Analysis Prompts

Traditional sentiment analysis can only categorize sentiments for overall reviews and can only be performed with the entire content. However, positive and negative emotions may exist in long comments addressing different aspects of the medical process. Based on the existing literature and data, we designed prompts to address this issue [29,34]. Afrashtehfar et al [34] constructed an instrument that successfully assessed patient satisfaction in several ways. We optimized the latest version of their tool. Specifically, as shown in Figure 2, author QZ summarized 3 categories based on Afrashtehfar et al's [34] findings and his own clinical experience and observations.

XSL•FO
RenderX

Figure 2. Illustration of formulating ABSA prompts from the dimensions of patient satisfaction. ABSA: aspect-based sentiment analysis.

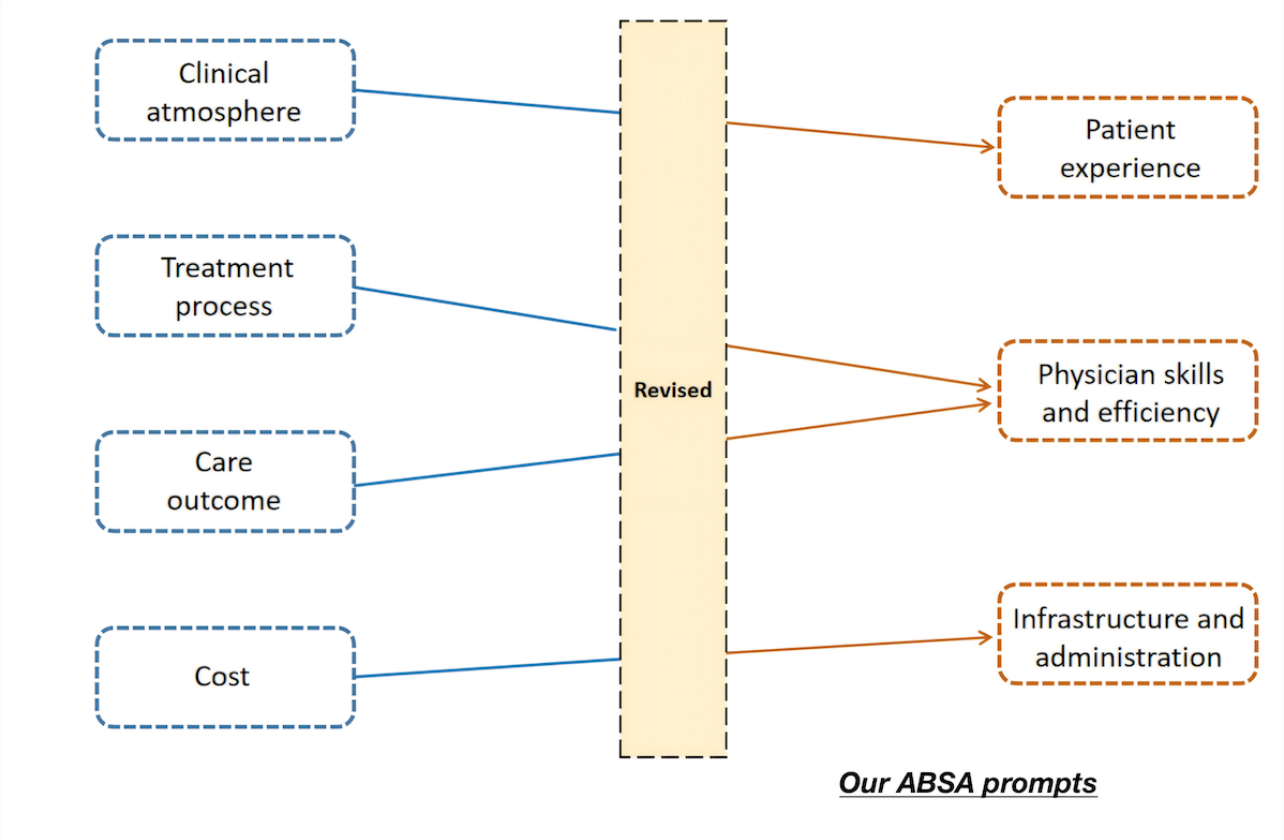

Specifically, "patient experience" includes perceptions of physicians, such as attitudes toward attending the hospital and the overall experience of attending the hospital. Second, "physician skill and efficiency" includes the physician's treatment plan for the patient's disease, the remission of the disease, and the physician's treatment efficiency. Third, "infrastructure and administration" includes medical facility conditions, costs, the environment, hygiene, registration, and other administrative issues of the hospital. In summary, our ABSA prompts consisted of 3 categories (patient experience, physician skills and efficiency, and infrastructure and administration), all of which are clinically relevant and clinically meaningful.

Chain of Thought

Some studies have shown that the use of CoT can alleviate problems such as hallucinations in an LLM to improve its accuracy [35-38]. We used the CoT technique to differentiate between the emotions related to the various topics in the comments. This process required logical reasoning to guide the LLM step by step through the thinking process, ensuring that steps were not skipped or omitted. With CoT, the LLM was guided to analyze the comments, confirm that each topic was been evaluated, and complete all the necessary steps in generating the final answer. Compared to outputs that do not use CoT, our approach can improve accuracy and facilitate checking the reasoning logic of the model.

In addition, as shown in Figure 3, we used a zero-shot technique and added an example to the prompt to allow the model to learn to control its output. Once the LLM received our task, the CoT forced it to complete it step by step. Specifically, when ChatGPT received a patient's comment, the model first extracted all the sentences related to "patient experience" from the comment and then analyzed the sentiment. After completing the sentiment analysis for the first category, the model extracted all sentences about "physician skills and efficiency" and performed sentiment analysis. Finally, the model extracted sentences about "infrastructure and administration" for sentiment analysis and output the results of all sentiment analyses. Our prompts were written in English, which aligns with the recommendations of other studies that suggest that using English in interactions with ChatGPT results in better performance. Other studies have noted that CoT performs better in English.

XSL•FO
RenderX

Figure 3. Prompt and CoT examples used in our work. CoT: chain of thought.

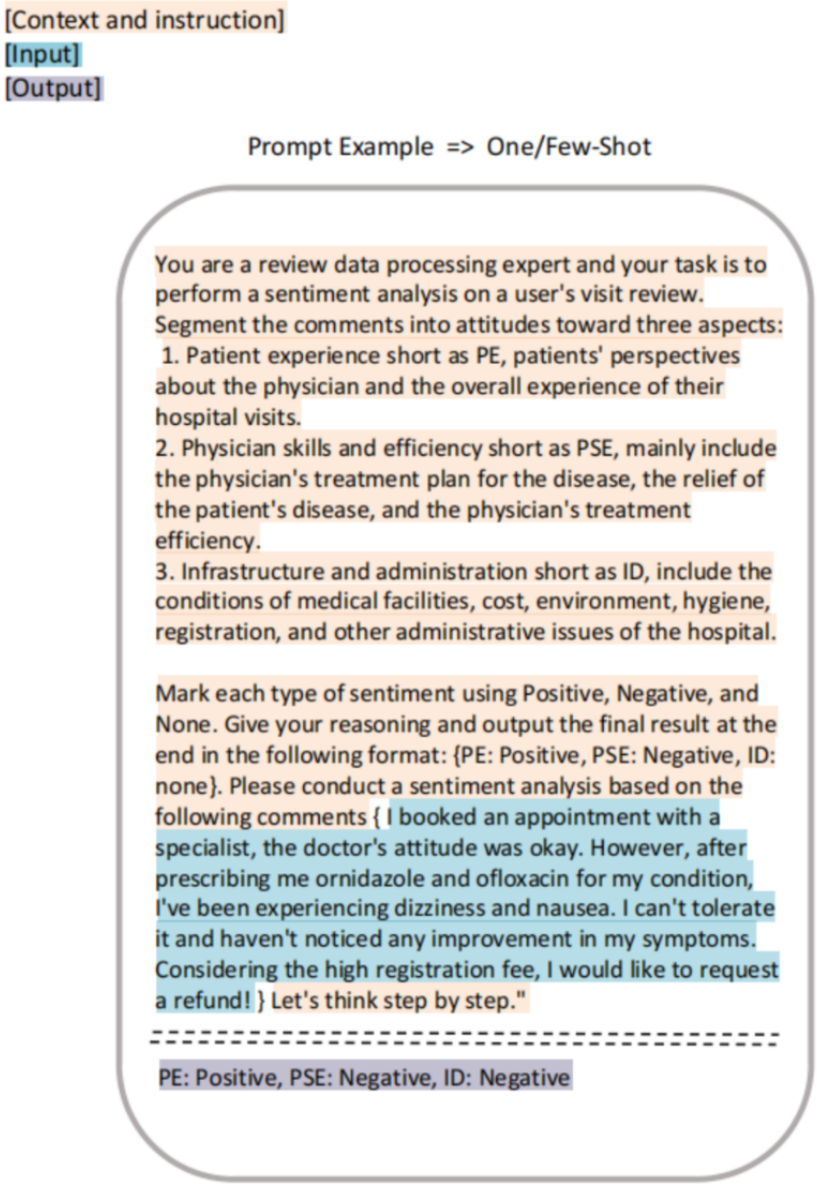

ICD-11 API

It has been pointed out that the study of individual diseases is of little significance but that quantifying the distribution and determinants of different kinds of diseases is of significant importance [39,40]. Because diseases of the same kind tend to be highly interdependent, and similar diseases often share similar pathogenesis, risk factors, etc, it is much more important to perform clustered disease analyses than individual disease analyses [39,40]. Therefore, we intended to explore more meaningful information about patients' health care needs by categorizing dispersed information about individual diseases through the ICD-11 API. It is also for this reason that some countries are piloting the possibility of health care system access to the ICD-11 API [28,41-44].

The ICD API is an HTTP-based REST API that allows programmatic access to the ICD [45]. Before the advent of the ICD-11 API, the only way to obtain a disease code was to search for it one at a time. In the 4-digit disease code, the first digit is the disease chapter code. There are 24 chapters in total. In version 2024, the International Classification of Functioning, Disability, and Health (ICF) was added as the second linearization of the API [46]. All linearization endpoints can access the ICF [46]. WHO provides a set of open source ICD-11 classification tools called API [47], we used Python's Requests library to call the linearlization search method provided by the API remotely.

Although in this version (2023-01), the ICD-11 API claims to have the capability of workin with the Chinese language, in our actual test, we found that the accuracy rate is lower than that for English when the input language is Chinese. One of the reasons for this is that when inputting Chinese, there are more requirements for the text to match. For example, we found that the ICD-11 API cannot recognize the Chinese characters for "disease" or "symptom" if they are missing (similar to issues reported in other work [28,48]). However, when we translated the terms into English and entered them into the ICD-11 API, it accurately recognized them and provided a disease code regardless of whether the Chinese character was or was not missing. In summary, we used ChatGPT 3.5 to translate Chinese

XSL•FO
RenderX

disease names to solve the problem of the Chinese version of the ICD-11 API being too stringent in its search scope. In this way, the ICD-11 API generated correct results (as shown in the validation before).

Evaluation Method

To more fully assess accuracy of our method, we used 2 approaches. First, we tested the performance of our method using manually labeled test sets and calculated precision, recall, and F_1-scores. Evaluating the results using manual random sampling enhances their interpretability and makes them more comprehensive and credible. Therefore, we used a dual criterion to evaluate the output fully. Finally, we evaluated the cost and run time to assess the pertinence of our method in practical applications.

Model Performance

We manually annotated patient comment data using the 3 categories defined in our ABSA prompts. We labeled each category as "positive" or "negative" within the same comment. If the comment did not refer to a particular category, it was labeled "none." Multiple authors with medical backgrounds performed the labeling. Two annotators worked independently (and one of them is a PhD candidate in the artificial intelligence [AI] field, while the other is a specialized physician at West China Hospital of Sichuan University). If the annotation results of the two annotators were inconsistent, we invited a professional doctor from a third-party hospital to resolve the conflict. We selected a test set of 2000 comments to assess precision, recall, and F_1-scores. The F_1-score combines precision and recall and is a widely used assessment metric in multicategory and uneven classification situations. Therefore, we use the F_1-score to evaluate sentiment categorization methods. Its calculation is as follows:



TP (true positive) stands for actual cases, indicating the number of samples that the model correctly predicted as positive cases; FN stands for false-negative cases, indicating the number of samples that the model failed to predict as positive cases correctly; and FP stands for false-positive cases, indicating the number of samples that the model incorrectly predicted as positive cases. To avoid sample imbalance affecting the accuracy of the conclusions, we chose to use the modified weighted F_1-scores to calculate the model's scores using Equation 4, where i is the index of the category, which is used to indicate the category of indicators currently being calculated, and w^i is the weight of category i, defined as the proportion of the sample size of that category to the total sample size.



Reliability and Stability Evaluation

Although we used a manually labeled test set to assess the model's accuracy, several studies have pointed to the instability and unreliability of LLMs. Therefore, we used a mixture of 3 sampling methods, whose merits have been pointed out by multiple studies, to assess the reliability and stability of the LLM model in our study. Specifically, mode 1 was simple random sampling, where 2000 pieces of data were drawn at a time within the total sample; mode 2 was incremental random sampling, where a total of 4 samples were taken, with 200, 400, 600, and 800 pieces of data, from the total sample; and mode 3 was isometric sampling, with a total of 4 draws each with 500 pieces of data taken from the total sample.

Hyperparameterization

There are several adjustable hyperparameters in the GPT model we used, such as temperature, top_p, frequency_penalty, and presence_penalty. We did not change these hyperparameters but followed the default settings to ensure our experiments were broadly applicable. Thus, temperature was set to 1.0, top_p was set to 1.0, and frequency_penalty and presence_penalty were set to 0.

Ethical Considerations

The dataset in this study was anonymized, and ethical clearance was obtained from the Macao Polytechnic University (approval no: HEA005-FCA-2024).

Results

Model Comparisons

Results showed that the accuracy of GPT-4o is slightly higher than that of ChatGPT 3.5, while its running time and costs are substantially higher than those of ChatGPT 3.5, which is not consistent with the research goal of sustainable development. GPT-4o took 104 minutes to process 1000 patient reviews, while ChatGPT 3.5 took only 71 minutes. In addition, GPT-4o cost US $1.7 to run 1000 patient reviews, while ChatGPT 3.5 cost only US $0.8 per 1000 reviews. The open source model Qwen-7B cost US $0 in comparison. In addition, the accuracy of GPT-4o was only 0.004 times higher than that of ChatGPT 3.5, while the accuracy of Qwen-7B was too low at 0.682 to meet the research criteria of this study. Therefore, we selected ChatGPT 3.5 as the experimental base model.

Table 1. Accuracy, run time, and cost of the 3 candidate models.

Model	Accuracy	Run time[a] (minutes)	Cost[a] (US $)
ChatGPT[b] 3.5	0.907	71	0.8
GPT-4o	0.911	104	1.7
Qwen-7B	0.682	73	0

[a]Per 1000 reviews.

[b]ChatGPT: Chat Generative Pretrained Transformer.

XSL•FO
RenderX

Model Performance

We tested the model on a test set of 2000 patient reviews with labels. To demonstrate the variability of the study, we introduced 95% CIs. As shown in Table 2, the model's weighted total results indicated a precision of 0.907 (95% CI 0.865-0.949), an F_1-score of 0.793 (95% CI 0.784-0.802), and a recall of 0.748 (95% CI 0.737-0.759). Thus, the model fulfilled the quality requirements of our experiments.

Table 2. Results for model precision, F_1-score, and recall by category.

Category	F_1-score	Precision	Recall
Patient experience	0.965	0.971	0.970
Physician skills and efficiency	0.899	0.948	0.879
Infrastructure and administration	0.405	0.860	0.394
Weighted total	0.793	0.907	0.748

Model Reliability and Stability

We used 3 different manual test sampling methods in this work. As shown in Table 3, the average accuracies of the 3 different sampling methods were 90.1%, 88.5%, and 89.4%, respectively. This was close to our accuracy of 0.907 described before, again confirming the reliability of both evaluation methodologies in the experiment.

Table 3. Model accuracy results for 3 manual random sampling methods.

Mode and characteristics	Sampling			
	First	Second	Third	Fourth
Mode 1 (entire sample accuracy: 90.1%)				
Sample size, n (%)	2000 (100)	—[a]	—	—
Patient experience accuracy (%)	92.2	—	—	—
Physician skills and efficiency accuracy (%)	89.4	—	—	—
Infrastructure and administration accuracy (%)	88.6	—	—	—
Mode 2 (entire sample accuracy: 88.5%)				
Sample size, n (%)	200 (10)	400 (20)	600 (30)	800 (40)
Patient experience accuracy (%)	94.0	90.3	92.5	91.3
Physician skills and efficiency accuracy (%)	90.5	88.8	89.8	88.6
Infrastructure and administration accuracy (%)	83.0	82.8	86.2	84.3
Mode 3 (entire sample accuracy: 89.4%)				
Sample size, n (%)	500 (25)	500 (25)	500 (25)	500 (25)
Patient experience accuracy (%)	93.8	95.8	89.6	90.8
Physician skills and efficiency accuracy (%)	88.8	92.4	95.4	90.4
Infrastructure and administration accuracy (%)	83.2	84.4	81.8	86.0

[a]Not applicable.

ICD-11 API Accuracy

We used Python to randomly extract 1000 disease names with ICD-11 codes from the datasets with ICD-11 chapter labels. Under the guidance of one of the authors of this paper, from the West China Hospital of Sichuan University, these data were reviewed for accuracy, and the results showed that 954 (95.4%) were correct, while 46 (4.6%) were incorrect; this accuracy of 95.4% met our experimental requirements.

Chapters of ICD-11

We calculated the distribution of the different types of diseases in ICD-11 across the 3 different categories of health care needs in our ABSA prompts. As shown in Figure 4, the diseases in ICD-11 Chapter 3 ("Diseases of the Blood or Blood-Forming Organs") showed a high degree of variability, with the percentage of negative comments on the "infrastructure and administration" category exceeding 15%, which is significantly higher than in all other chapters. However, the percentage of negative comments on "patient experience" and "physician skills and efficiency" was only about 4%, which was only higher than for diseases of the nervous system, sleep-wake disorders, and diseases of the circulatory system (ICD-11 Chapters 8, 7, and 11, respectively).

XSL•FO
RenderX

Figure 4. Percentage of negative comments by chapter in ICD-11. The x-axis was sorted by the average percentage of negative reviews on all aspects. ICD-11: *International Classification of Diseases 11th Revision*.

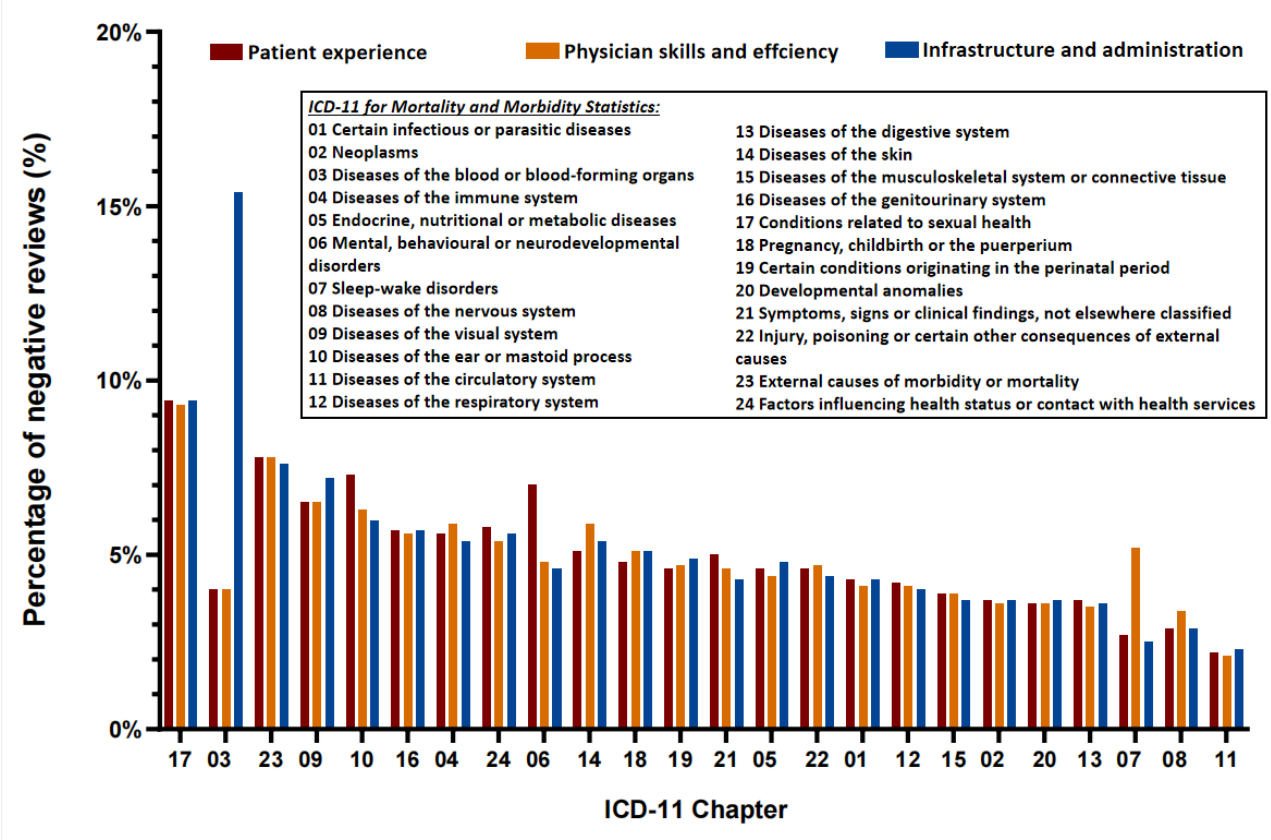

Next, to explore the overall patient demand by disease, we calculated the average percentage of negative ABSA comments and ranked them in descending order. As shown in Table 4, most of the chapters had similar rates of poor ratings for all 3 categories of health care needs in our ABSA prompts. In chapters on diseases of the blood or blood-forming organs; mental, behavioural, or neurodevelopmental disorders; and sleep-wake disorders (Chapters 3, 6, and 7, respectively), there was a significant difference in a single category. In the chapter on mental, behavioural, or neurodevelopmental disorders (Chapter 6), the “patient experience” category had a negative review percentage of about 7%, while the remaining 2 categories were below 5%. In the chapter on sleep-wake disorders (Chapter 7), the percentage of negative comments for “physician skills and efficiency” exceeded 5%, while the remaining 2 categories were below 3%.

XSL•FO
RenderX

Table 4. Percentage of negative comments by patients with different diseases based on ICD-11[a] chapters.

ICD-11 chapter	Patient experience (%)	Physician skills and efficiency (%)	Infrastructure and administration (%)	Average (%)
17	9.40	9.30	9.40	9.37
3	4.00	4.00	15.40	7.80
23	7.80	7.80	7.60	7.73
9	6.50	6.50	7.20	6.73
10	7.30	6.30	6.00	6.53
16	5.70	5.60	5.70	5.67
4	5.60	5.90	5.40	5.63
24	5.80	5.40	5.60	5.60
6	7.00	4.80	4.60	5.47
14	5.10	5.90	5.40	5.47
18	4.80	5.10	5.10	5.00
19	4.60	4.70	4.90	4.73
21	5.00	4.60	4.30	4.63
5	4.60	4.40	4.80	4.60
22	4.60	4.70	4.40	4.57
1	4.30	4.10	4.30	4.23
12	4.20	4.10	4.00	4.10
15	3.90	3.90	3.70	3.83
2	3.70	3.60	3.70	3.67
20	3.60	3.60	3.70	3.63
13	3.70	3.50	3.60	3.60
7	2.70	5.20	2.50	3.47
8	2.90	3.40	2.90	3.07
11	2.20	2.10	2.30	2.20

[a]ICD-11: *International Classification of Diseases 11th Revision*.

As also listed in Table 4, chapters on diseases of the blood or blood-forming organs (Chapter 3), conditions related to sexual health (Chapter 17), external causes of morbidity or mortality (Chapter 23), and diseases of the visual system (Chapter 9) had the top 4 highest percentages of negative comments, exceeding 9.8%, 7.8%, 7.7%, and 6.7%, respectively. The direction of "infrastructure and administration" in diseases of the visual system (Chapter 9) and "patient experience" in diseases of the ear or mastoid process (Chapter 10) reached 7.2% and 7.3%, respectively, significantly higher than the other 2 categories. In addition, chapters on diseases of the digestive system, diseases of the nervous system, sleep-wake disorders, and diseases of the circulatory system (Chapters 13, 8, 7, and 11, respectively) had the lowest poor ratings, which were below 4%, 3.5%, 3%, and 2.5%, respectively, except for the classification of "physician skills and efficiency" in sleep-wake disorders (Chapter 7).

Figure 5 shows a better visualization of the percentage of negative reviews in each category in the ABSA prompts and the differences between the different chapters. In the "patient experience" category, the 3 chapters on conditions related to sexual health, external causes of morbidity or mortality, and diseases of the ear or mastoid process (Chapters 17, 23, and 10, respectively) had the highest percentage of negative comments, which exceeded 9%, 7.5%, and 7%, respectively. In addition, chapters on mental, behavioural, or neurodevelopmental disorders; diseases of the visual system; factors influencing the health status or contact with health services; diseases of the genitourinary system; diseases of the immune system; and diseases of the skin and symptoms not elsewhere classified (Chapters 6, 9, 24, 16, 4, 14, and 21, respectively) had negative comments exceeding 5 %, for a total of 9 chapters. Except for these 9 chapters, the rest were below 5%, and the 3 chapters with the lowest percentage of negative comments were on diseases of the nervous system, sleep-wake disorders, and diseases of the circulatory system (Chapters 8, 7, and 11, respectively), at less than 3%, 3%, and 2.5%, respectively.

XSL•FO
RenderX

Figure 5. Percentage of negative reviews by ICD-11 chapter. ICD-11: *International Classification of Diseases 11th Revision*.

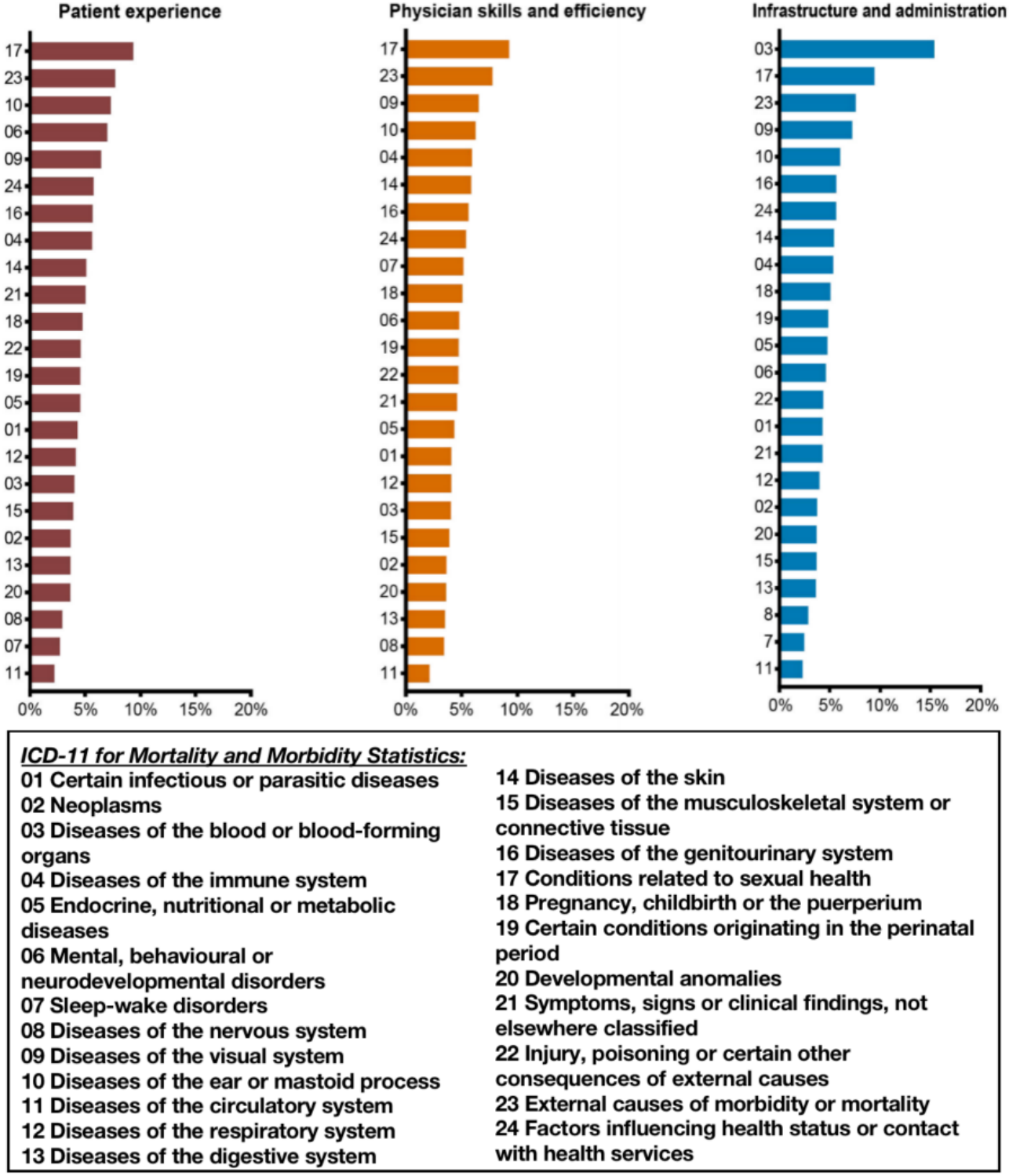

In the “physician skills and efficiency” category, chapters on conditions related to sexual health, external causes of morbidity or mortality, and diseases of the visual system (Chapters 17, 23, and 9, respectively) had the highest percentage of negative comments at 9%, 7.5%, and 6.5%, respectively. In addition, chapters on diseases of the ear or mastoid process , diseases of the immune system, diseases of the skin, diseases of the genitourinary system, factors influencing the health status or contact with health services, sleep-wake disorders, and diseases of the nervous system (Chapters 10, 4, 14, 16, 24, 7, and 8, respectively) had more than 5% of negative comments, for a total of 10 chapters. In addition, the 3 chapters with the lowest negative comment percentage were on diseases of the digestive system, diseases of the nervous system and diseases of the circulatory system (Chapters 13, 8, and 11, respectively) at about 3.5%, 3.5%, and 2%, respectively. The rest of the chapters had a negative comment percentage between 5 and 3.5%.

Finally, in the “infrastructure and administration” category, the 3 chapters with the highest percentage of negative comments were on diseases of the blood or blood-forming organs,

XSL•FO
RenderX

conditions related to sexual health, and external causes of morbidity or mortality (Chapters 3, 17, and 23, respectively), at 15%, 9%, and 7.5%, respectively. Additionally, 15% was the highest percentage of negative comments among all 3 categoriess. The next highest percentage of negative comments was found for chapters on diseases of the visual system, diseases of the ear or mastoid process, diseases of the genitourinary system, factors influencing the health status or contact with health services, diseases of the skin, diseases of the immune system and pregnancy, childbirth or the puerperium (Chapters 9, 10, 16, 24, 14, 4, and 18, respectively), with the following highest percentage of negative comments being found for chapters on diseases of the circulatory system, sleep-wake disorders, and diseases of the nervous system (Chapters 11, 7, and 8, respectively), at less than 3%.

To visually observe the relationship of ABSA results among diseases, we visualized the top 5 ICD-11 chapters with the highest and the lowest percentage of negative patient reviews using a body map. Based on the visualization, we obtained a clearer image of different body parts with the highest and lowest percentages of negative reviews. As shown in Figure 6, most of the top 5 chapters with the lowest percentages focus on diseases near the brain, including neurological disorders, sleep disorders, and developmental disorders.

Figure 6. Visualization of parts of the body systems with the highest and lowest percentages of negative patient reviews by ICD-11 chapter. ICD-11: *International Classification of Diseases 11th Revision*.

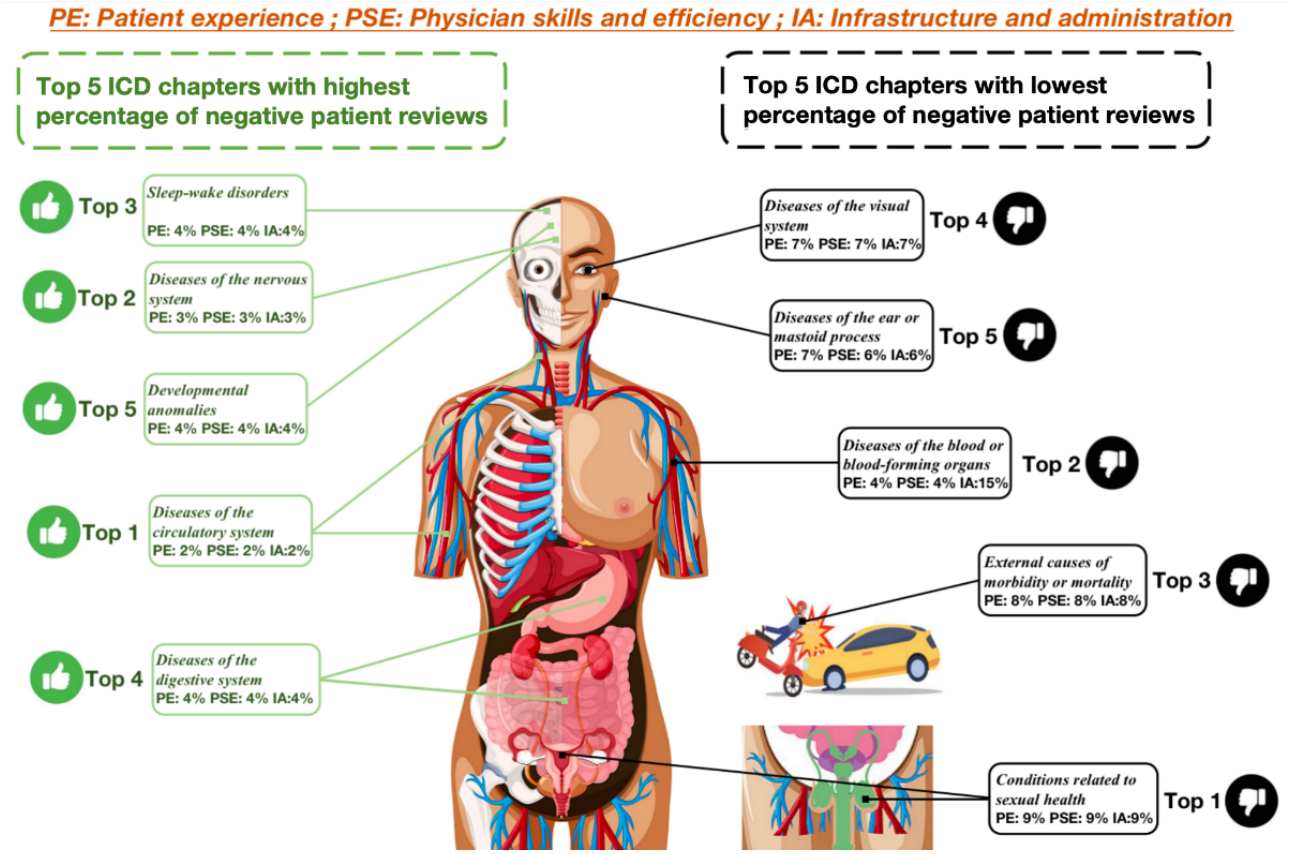

Discussion

Principal Findings

Model Accuracy and Feasibility

The results showed that our model achieves an F_1-score, precision, and recall of 0.793, 0.907, and 0.748, respectively, proving the effectiveness of our combination of ABSA prompts and CoT architecture. In addition, the model achieved 90.1%, 88.5%, and 89.4% accuracy in the stability and reliability tests conducted in random sampling modes 1, 2, and 3, respectively. We used an earlier version of LLMs (ChatGPT-3.5-turbo-0613), which was released in 2023. This proved that the initial version of GPT was already sufficient for our task, and it was expected that newer versions of the model can also be used for this type of task.

Differences in Patients' Health Care Needs Across ICD-11 Chapters

Overall, across all classifications, patients with sexual health–related illnesses had high health care needs and showed high levels of dissatisfaction with all aspects. Based on the clinical setting and research in China [49-51], we hypothesize that this may be related to the extremely private nature of the disease, which leads to shame and low self-esteem, which, in turn, leads to a sensitive and vulnerable mindset and, therefore, extreme concern for the physician's attitude, skills, and the environment in which the consultation takes place. In contrast, patients with circulatory disorders expressed satisfaction with current medical resources, suggesting that China has an advantage in researching and treating this type of disease.

Second, although the overall medical needs of patients with sleep-wake disorders are not high, patients with these disorders

XSL•FO
RenderX

show a higher demand for physician skills and efficiency, probably because it is difficult to see immediate results in the treatment of these patients and because insomnia can lead to neurasthenia, etc, which can have a significant impact on work and life. Therefore, there is a strong demand for efficacy for such types of conditions.

In addition to the aforementioned diseases, patients with ear or mastoid disorders also have a relatively high demand for patient experience. This may be because ear diseases make patients wary and uncomfortable with all the sounds in their environment and, therefore, extremely sensitive to the attitude and tone of the physician's dialogue. Furthermore, immune system disorders have a higher need for physician skill and efficiency, which may be related to the degree of difficulty in treating the diseases, as immune system disorders are generally more difficult to cure. Therefore, there is generally a robust and heartfelt expectation from patients regarding the physician's skill.

Lastly, neurological or hematopoietic organ diseases lead the way regarding the "infrastructure and administration" category. This may be because these diseases depend highly on advanced hospital infrastructure, such as specific hemodialysis equipment. In short, each disease has different medical needs. Therefore, medical management can allocate or solve problems according to the medical needs of various diseases. If one needs to continue to explore the specific causes of a problem, as proposed in a previous work [52], one can do so downstream of the task, such as word frequency analysis and word cloud construction. In this way, we can uncover specific negative reasons in patient comments.

Implications for Sustainable Development Goals

Our work on analyzing patient comments using ICD-11 and LLMs to identify health care needs for different diseases has important implications for the SDGs. First, in SDG 3 ("Good Health and Well-being"), the precise analysis of patient comments can provide a more accurate understanding of the specific needs of different diseases, promote accurate consultation and treatment in health care units, and improve the quality of their health care services. In SDG 9 ("Industry, Innovation and Infrastructure"), our research demonstrates the use of LLMs and ICD-11 for health care innovation to improve diagnosis and treatment efficiency, promote smart health care infrastructure, and enhance the accessibility and efficiency of health care services in various ICD-11 chapters. In addition, for SDG 10 ("Reducing Inequalities"), by revealing the health needs and health care disparities among different groups, we can help develop more targeted health policies to reduce health care inequalities and ensure that all people have access to appropriate health care, as well as emphasizing the importance of the patient's voice and promoting inclusive health care.

Finally, for SDG 17 ("Partnerships for Goal Achievement"), this research highlights the importance of interdisciplinary collaboration, combining medicine, data science, and AI technologies to encourage multiple parties, including public health organizations, academic institutions, and technology companies, to work together to address health challenges worldwide.

In conclusion, our research not only has important application value in the medical field but also provides valuable insights into the achievement of the SDGs and promotes the innovation and development of health care services through the full use of modern scientific and technological tools to enhance health worldwide and promote the sustainable development of society.

Potential Applications of the Assessment of Health Care Services

Several studies have shown that physicians or hospitals in the United States are incentivized by patient scores or ratings [53-55], which helps them provide efficient, high-quality care [56,57]. At the same time, these scores or ratings are made public on medical websites, and people can choose their preferred departments or hospitals through the ratings. In China, there is a need for ratings for hospitals [58], and although some medical websites have ratings for physicians or hospitals, the ratings on websites could be more sketchy and include only a small number of physicians or hospitals [59]. Therefore, governments and health authorities can use our approach to obtain fine-grained ratings of each health care unit to enhance monitoring and management, which would benefit both patients and the administration.

From the patients' perspective, aggregating satisfaction and dissatisfaction scores based on ICD-11 codes can help them select health care providers based on their preferences. For example, if a patient is more concerned about the facilities and customer experience of a hospital, they can choose a hospital with a lower ratio of negative reviews in the "infrastructure and administration" category. In addition, as shown in Multimedia Appendix 1, the health care administration can obtain ABSA results with ICD-11 codes using our method and use it for continuous improvements. The result with ICD-11 codes can be shown as a dashboard, and management can identify areas with high satisfaction and dissatisfaction. Patient reviews in such areas can be further analyzed using natural language processing techniques (eg, word frequency analysis or topic modeling) to identify the root causes.

Taking our datasets as an example, assuming that the they originate from the same hospital, management can prioritize the verification and optimization of health care resources for equipment and environments in different departments with a high percentage of negative comments in the "infrastructure and administration"category (eg, ICD-11 Chapters 1, 17, and 23 in this case). Second, management can organize quality education and psychoeducation for doctors to optimize departments with a high percentage of negative comments in the "patient experience" category (eg, ICD-11 Chapters 10, 17, and 23). Finally, management can provide systematic training to doctors and organize professional meetings to address departments with a high percentage of negative comments in the "physician skills and efficiency" category (eg, ICD-11 Chapters 9, 17, and 23). These evaluation cycles can be iterated over time for checking whether there is any improvement.

Limitations

This study has several limitations. Since this study proposed a new ABSA template, there are no other datasets with annotations

XSL•FO
RenderX

to be compared and cross-validated with this study for now; therefore, it is not possible to validate whether this method can be applied to datasets in other languages. In addition, it is not possible to verify whether it is meaningful to apply the ABSA in this study to regions with different health care environments. In terms of application, we might be able to make in-depth analyses by comparing data from multiple regions in the future. Finally, to prove that the problem of SDGs has been solved requires a comprehensive evaluation framework and evaluation method [32,60]. In the future, we anticipate collaboration with multiple hospitals to conduct a more comprehensive evaluation and practice in a specific clinical setting to determine the positive impact of this study on health care resource allocation and SDGs.

Conclusion

We analyzed the distribution of health care resources by disease through an LLM of patient reviews. In addition, we combined the results with the ICD-11 classification and found that conditions related to sexual health have the lowest overall patient satisfaction. In contrast, patient satisfaction is highest among patients with diseases of the circulatory system. Our approach demonstrates the possibility of using an LLM to understand patients' needs and extract aspect-based information from reviews written by them and health consumers.

Acknowledgments

We thank Rong Chen and the First Affiliated Hospital, Sun Yat-Sen University, for their contributions and guidance in this research. This research was supported by Macao Polytechnic University (submission code: fca.1c2e.6826.4) and the Macao Science and Technology Development Fund (funding ID: 0088/2023/ITP2).

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JL contributed to conceptualization, methodology, formal analysis, writing—original draft, project administration, and data curation. YY contributed to methodology, project administration, and software. CM contributed to project administration, investigation, and methodology. PCIP contributed to supervision, funding acquisition, and writing—review and editing. QZ and DX contributed to visualization, validation, and resources. YW contributed to supervision and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Two potential directions of the application of our method. ABSA: aspect-based sentiment analysis; ICD-11: *International Classification of Diseases 11th Revision*.
[PNG File , 175 KB - jmir_v27i1e66344_app1.png]

References

1. Phillips-Wren G, Sharkey P, Dy SM. Mining lung cancer patient data to assess healthcare resource utilization. Expert Syst Appl 2008 Nov;35(4):1611-1619. [doi: 10.1016/j.eswa.2007.08.076]
2. Pezzella P. The ICD-11 is now officially in effect. World Psychiatry 2022 Jun;21(2):331-332 [FREE Full text] [doi: 10.1002/wps.20982] [Medline: 35524598]
3. Rosenbaum K, Lasater K, McHugh M. Hospital performance on hospital consumer assessment of healthcare providers and system ratings: associations with nursing factors. Med Care 2024;62(5):288-295. [doi: 10.1097/mlr.0000000000001966]
4. Sundewall J, Forsberg BC. Understanding health spending for SDG 3. Lancet 2020 Sep;396(10252):650-651. [doi: 10.1016/s0140-6736(20)30963-6]
5. Elhadi M, Msherghi A, Alkeelani M, Zorgani A, Zaid A, Alsuyihili A, et al. Assessment of healthcare workers' levels of preparedness and awareness regarding COVID-19 infection in low-resource settings. Am J Trop Med Hyg 2020 Aug;103(2):828-833 [FREE Full text] [doi: 10.4269/ajtmh.20-0330] [Medline: 32563273]
6. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of COVID-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/nejmsb2005114]
7. Kynčlová P, Upadhyaya S, Nice T. Composite index as a measure on achieving Sustainable Development Goal 9 (SDG-9) industry-related targets: the SDG-9 index. Applied Energy 2020 May;265:114755. [doi: 10.1016/j.apenergy.2020.114755]
8. Küfeo?lu S. SDG-9: industry, innovation and infrastructure. In: Emerging Technologies: Value Creation for Sustainable Development. Cham: Springer; 2022:349-369.
9. Mallia P. Allocation of resources in medicine – a moral imperative for doctors as well. Medicine 2020 Oct;48(10):671-674. [doi: 10.1016/j.mpmed.2020.07.002]

XSL•FO
RenderX

10. Teubner T, Flath CM, Weinhardt C, van der Aalst W, Hinz O. Welcome to the Era of ChatGPT et al.: the prospects of large language models. Bus Inf Syst Eng 2023 Mar 13;65(2):95-101. [doi: 10.1007/s12599-023-00795-x]
11. Wu T, He S, Liu J, Sun S, Liu K, Han Q, et al. A brief overview of ChatGPT: the history, status quo and potential future development. IEEE/CAA J Autom Sinica 2023 May;10(5):1122-1136. [doi: 10.1109/jas.2023.123618]
12. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. npj Digit Med 2023 Nov 16;6(1):210. [doi: 10.1038/s41746-023-00958-w]
13. Dowling M, Lucey B. ChatGPT for (finance) research: the Bananarama conjecture. Fin Res Lett 2023 May;53:103662. [doi: 10.1016/j.frl.2023.103662]
14. Baidoo-Anu D, Ansah L. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teachinglearning. JAI 2023;7(1):52-62. [doi: 10.61969/jai.1337500]
15. Spiegel BMR, Liran O, Clark A, Samaan JS, Khalil C, Chernoff R, et al. Feasibility of combining spatial computing and AI for mental health support in anxiety and depression. NPJ Digit Med 2024 Jan 26;7(1):22 [FREE Full text] [doi: 10.1038/s41746-024-01011-0] [Medline: 38279034]
16. Afshar M, Gao Y, Gupta D, Croxford E, Demner-Fushman D. On the role of the UMLS in supporting diagnosis generation proposed by large language models. J Biomed Inform 2024 Sep;157:104707. [doi: 10.1016/j.jbi.2024.104707]
17. Yan C, Grabowska ME, Dickson AL, Li B, Wen Z, Roden DM, et al. Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real-world clinical validation. NPJ Digit Med 2024 Feb 26;7(1):46 [FREE Full text] [doi: 10.1038/s41746-024-01038-3] [Medline: 38409350]
18. Gakuba C, Le BC, Sar A. Evaluation of ChatGPT in predicting 6-month outcomes after traumatic brain injury. Crit Care Med 2024;52(6):942-950. [doi: 10.1097/ccm.0000000000006236]
19. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl-Based Syst 2021 Aug;226:107134. [doi: 10.1016/j.knosys.2021.107134]
20. Zhang W, Li X, Deng Y, Bing L, Lam W. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. IEEE Trans Knowl Data Eng 2023 Nov 1;35(11):11019-11038. [doi: 10.1109/tkde.2022.3230975]
21. Nazir A, Rao Y, Wu L, Sun L. Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. IEEE Trans Affect Comput 2022 Apr 1;13(2):845-863. [doi: 10.1109/taffc.2020.2970399]
22. Do HH, Prasad P, Maag A, Alsadoon A. Deep learning for aspect-based sentiment analysis: a comparative review. Expert Syst Appl 2019 Mar;118:272-299. [doi: 10.1016/j.eswa.2018.10.003]
23. Fanti MP, Mangini AM, Dotoli M, Ukovich W. A three-level strategy for the design and performance evaluation of hospital departments. IEEE Trans Syst Man Cybern Syst 2013 Jul;43(4):742-756. [doi: 10.1109/tsmca.2012.2217319]
24. Rowland HS, Rowland BL, Díaz de Santos E. Hospital Management: A Guide to Departments. Philadelphia, PA: Lippincott Williams and Wilkins; 1984.
25. International Classification of Diseases, Eleventh Revision (ICD-11). World Health Organization. 2021. URL: https://icd.who.int/browse11 [accessed 2025-02-25]
26. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak 2021 Nov 09;21(Suppl 6):206 [FREE Full text] [doi: 10.1186/s12911-021-01534-6] [Medline: 34753471]
27. Fenton S, Stanfill M, Giannangelo K. An ICD for the digital world: what does the ICD-11 research show? In: MEDINFO 2023—The Future Is Accessible. Amsterdam, the Netherlands: IOS Press; 2024:58-62.
28. Zhang M, Wang Y, Jakob R, Su S, Bai X, Jing X, et al. Methodologies and key considerations for implementing the International Classification of Diseases-11th Revision morbidity coding: insights from a national pilot study in China. J Am Med Inform Assoc 2024 Apr 19;31(5):1084-1092 [FREE Full text] [doi: 10.1093/jamia/ocae031] [Medline: 38427850]
29. Triola MM, Burk-Rafel J. Precision medical education. Acad Med 2023;98(7):775-781. [doi: 10.1097/acm.0000000000005227]
30. Li J, Pang P. Have patients' attitudes towards doctors changed with COVID-19? Insights from online doctor reviews in Beijing, China. 2022 Presented at: ICEME 2022: 13th International Conference on E-business, Management and Economics; July 16-18, 2022; Beijing, China. [doi: 10.1145/3556089.3556163]
31. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
32. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
33. Li L, Zhang H, Li C. Evaluation on ChatGPT for Chinese language understanding. Data Intell 2023;5(4):885-903. [doi: 10.1162/dint_a_00232]
34. Afrashtehfar KI, Assery MKA, Bryant SR. Patient Satisfaction in medicine and dentistry. Int J Dent 2020;2020:6621848 [FREE Full text] [doi: 10.1155/2020/6621848] [Medline: 33456467]
35. Wei J, Wang X, Schuurmans D. Chain-of-thought prompting elicits reasoning in large language modelsJ. 2022 Presented at: NIPS'22: 36th International Conference on Neural Information Processing Systems; November 28-December 9, 2022; New Orleans, LA p. 24824-24837.

XSL•FO
RenderX

36. Feng G, Zhang B, Gu Y. Towards revealing the mystery behind chain of thought: a theoretical perspectiveJ. 2023 Presented at: NIPS '23: 37th International Conference on Neural Information Processing Systems; December 10-16, 2023; New Orleans, LA p. 70757-70798.
37. Wang X, Wei J, Schuurmans D. Self-consistency improves chain of thought reasoning in language models. arXiv Preprint posted online 2022. [doi: 10.48550/arXiv.2203.11171]
38. Zhang Z, Zhang A, Li M. Automatic chain of thought prompting in large language models. arXiv Preprint posted online 2022. [doi: 10.48550/arXiv.2210.03493]
39. Webster AJ, Gaitskell K, Turnbull I, Cairns BJ, Clarke R. Characterisation, identification, clustering, and classification of disease. Sci Rep 2021 Mar 08;11(1):5405 [FREE Full text] [doi: 10.1038/s41598-021-84860-z] [Medline: 33686097]
40. Smart B. On the classification of diseases. Theor Med Bioeth 2014;35:251-269 [FREE Full text] [doi: 10.1007/s11017-014-9301-9]
41. Mugisha M, Byiringiro JB, Uwase M, Abizeyimana T, Ndikubwimana B, Karema N, et al. Integration of International Classification of Diseases version 11 application program interface (API) in the Rwandan electronic medical records (openMRS): findings from two district hospitals in Rwanda. Stud Health Technol Inform 2020 Jun 26;272:280-283. [doi: 10.3233/SHTI200549] [Medline: 32604656]
42. Ibrahim I, Alrashidi M, Al-Salamin M, Kostanjsek N, Jakob R, Azam S, et al. ICD-11 morbidity pilot in Kuwait: methodology and lessons learned for future implementation. Int J Environ Res Public Health 2022 Mar 05;19(5):3057 [FREE Full text] [doi: 10.3390/ijerph19053057] [Medline: 35270745]
43. Eastwood CA, Southern DA, Khair S, Doktorchik C, Ghali WA, Quan H. The ICD-11 field trial: creating a large dually coded database. ResearchGate. 2021 May 13. URL: https://www.researchgate.net/publication/351551650_The_ICD-11_Field_Trial_Creating_a_Large_Dually_Coded_Database [accessed 2025-02-25]
44. Bach B, Kramer U, Doering S, di Giacomo E, Hutsebaut J, Kaera A, et al. The ICD-11 classification of personality disorders: a European perspective on challenges and opportunities. Borderline Pers Disord Emot Dysregul 2022 Apr 01;9(1):12 [FREE Full text] [doi: 10.1186/s40479-022-00182-0] [Medline: 35361271]
45. Liu H. Service-oriented means of obtaining the name of the disease according to the international code. National Aviation University. 2023 Dec. URL: https://er.nau.edu.ua/items/aefaeebb-73e5-4070-801e-2a499476de4a [accessed 2025-02-25]
46. Della Mea V, Almborg A, Martinuzzi M, Tu SW, Martinuzzi A. Harmonization of ICF body structures and ICD-11 anatomic detail: one foundation for multiple classifications. PLoS One 2023 Jul 27;18(7):e0280106 [FREE Full text] [doi: 10.1371/journal.pone.0280106] [Medline: 37498874]
47. ICD API documentation version 2. World Health Organization. URL: https://icd.who.int/docs/icd-api/APIDoc-Version2/ [accessed 2024-06-05]
48. Xu Y, Zhou J, Li H, Cai D, Zhu H, Pan S. Improvements in neoplasm classification in the International Classification of Diseases, Eleventh Revision: systematic comparative study with the Chinese clinical modification of the International Classification of Diseases, Tenth Revision. Interact J Med Res 2024 Mar 08;13:e52296 [FREE Full text] [doi: 10.2196/52296] [Medline: 38457228]
49. Li W, Li S, Lu P, Chen H, Zhang Y, Cao Y, et al. Sexual dysfunction and health condition in Chinese doctor: prevalence and risk factors. Sci Rep 2020 Sep 16;10(1):15180 [FREE Full text] [doi: 10.1038/s41598-020-72072-w] [Medline: 32938977]
50. Li T, Luo Y, Meng Y, Yue J, Nie M, Fan L, et al. Sexual activity and related factors of older women in Hunan, China: a cross-sectional study. J Sex Med 2022 Feb;19(2):302-310. [doi: 10.1016/j.jsxm.2021.11.020] [Medline: 34973899]
51. Dong Y. "Premature" girls with “illegitimate” diseases: a qualitative exploration of the stigma of gynecological diseases for unmarried young women in China. J Health Commun 2023 Feb 01;28(2):111-120. [doi: 10.1080/10810730.2023.2185322] [Medline: 36855814]
52. Li J, Pang PC, Xiao Y, Wong D. Changes in doctor–patient relationships in China during COVID-19: a text mining analysis. Int J Environ Res Public Health 2022 Oct 18;19(20):13446 [FREE Full text] [doi: 10.3390/ijerph192013446] [Medline: 36294022]
53. Khanna K, Diab M. Physician ratings: determinants, accuracy, and impact. J Bone Joint Surg Am 2021 Apr 07;103(7):e27. [doi: 10.2106/JBJS.20.01048] [Medline: 33787556]
54. Syed U, Acevedo D, Narzikul A, Coomer W, Beredjiklian P, Abboud J. Physician rating websites: an analysis of physician evaluation and physician perception. Arch Bone Jt Surg 2019 Mar;7(2):136-142 [FREE Full text] [Medline: 31211191]
55. Sabin JE. Physician-rating websites. Virtual Mentor 2013 Nov 01;15(11):932-936 [FREE Full text] [doi: 10.1001/virtualmentor.2013.15.11.ecas2-1311] [Medline: 24257083]
56. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-based physician ratings for California physicians on probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
57. Johnson C. Survey finds physicians very wary of doctor ratings. Physician Exec 2013;39(1):6-8,10,12. [Medline: 23437751]
58. Han X, Li B, Zhang T, Qu J. Factors associated with the actual behavior and intention of rating physicians on physician rating websites: cross-sectional study. J Med Internet Res 2020 Jun 04;22(6):e14417 [FREE Full text] [doi: 10.2196/14417] [Medline: 32496198]

XSL•FO
RenderX

59. Yang Y, Zhang X, Lee PK. Improving the effectiveness of online healthcare platforms: an empirical study with multi-period patient-doctor consultation data. Int J Product Econ 2019 Jan;207:70-80. [doi: 10.1016/j.ijpe.2018.11.009]
60. Giles-Corti B, Lowe M, Arundel J. Achieving the SDGs: evaluating indicators to be used to benchmark and monitor progress towards creating healthy and sustainable cities. Health Policy 2020 Jun;124(6):581-590. [doi: 10.1016/j.healthpol.2019.03.001] [Medline: 30935701]

Abbreviations

ABSA: aspect-based sentiment analysis
AI: artificial intelligence
API: application programming interface
ChatGPT: Chat Generative Pretrained Transformer
CoT: chain of thought
ICD-11: *International Classification of Diseases 11th Revision*
ICF: International Classification of Functioning, Disability, and Health
LLM: large language model
SDG: Sustainable Development Goal

Edited by A Coristine; submitted 10.09.24; peer-reviewed by Y Peng, J Zhang, X Du; comments to author 04.11.24; revised version received 19.12.24; accepted 04.02.25; published 17.03.25.

Please cite as:
Li J, Yang Y, Mao C, Pang PCI, Zhu Q, Xu D, Wang Y
Revealing Patient Dissatisfaction With Health Care Resource Allocation in Multiple Dimensions Using Large Language Models and the International Classification of Diseases 11th Revision: Aspect-Based Sentiment Analysis
J Med Internet Res 2025;27:e66344
URL: https://www.jmir.org/2025/1/e66344
doi:10.2196/66344
PMID:

©Jiaxuan Li, Yunchu Yang, Chao Mao, Patrick Cheong-Iao Pang, Quanjing Zhu, Dejian Xu, Yapeng Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Estimating Patient and Family Costs and CO2 Emissions for Telehealth and In-Person Health Care Appointments in British Columbia, Canada: Geospatial Mixed Methods Study

Graham Mainer-Pearson[1], BA; Kurtis Stewart[1], MSc; Kim Williams[2], BSCN, MSCN; John Pawlovich[2,3,4], MD; Scott Graham[5,6], MA; Linda Riches[7], MEd; Sonya Cressman[1], MBA, PhD; Kendall Ho[1], MD

[1]Digital Emergency Medicine, Department of Emergency Medicine, University of British Columbia, Vancouver, BC, Canada

[2]Rural Coordination Centre of British Columbia, Vancouver, BC, Canada

[3]Carrier Sekani Family Services, West Vancouver, BC, Canada

[4]Department of Family Practice, University of British Columbia, Vancouver, BC, Canada

[5]First Nations Health Authority, Vancouver, BC, Canada

[6]School of Community and Regional Planning, University of British Columbia, Vancouver, BC, Canada

[7]Patient Partner Researcher, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Sonya Cressman, MBA, PhD
Digital Emergency Medicine
Department of Emergency Medicine
University of British Columbia
Rm 6400 - 818 West 10th Avenue
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 6048220327
Email: sonya.cressman@ubc.ca

Abstract

Background: Patients inevitably incur some cost for accessing health care, even in universal systems such as Canada. The COVID-19 pandemic dramatically shifted health care delivery from in-person to telehealth services, also shifting the proportion of costs offset by patients and their families by reducing the need to travel to in-person appointments.

Objective: This study aimed to develop a method for estimating the costs patients and their families incur and CO_2 emissions attributed to travel needed for emergency department (ED) visits, hospitalizations, and physician appointments.

Methods: We present a method to evaluate the costs associated with in-person and telehealth care appointments from the perspective of patients, their families, and the environment. We used ED locations, road distances, and duration of appointment to account for costs paid by patients (ie, lost productivity, informal caregiving, and out-of-pocket expenses) attributed to travel to receive medical care. Costs to the environment were evaluated by calculating the amount of CO_2 emitted per medical visit. Using our costs calculated per visit, we apply our method to calculate total patient costs for a simulated population over 1 year.

Results: Our method estimates that patients in British Columbia pay up to $300 (2023 CAD, CAD $1=US $0.86) on average to attend an in-person ED visit, depending on where they live; $166 may be attributed to lost productivity, $83 to informal caregiving, and $50 to out-of-pocket expenses. These estimates are higher than most observed cost estimates. In addition, avoiding in-person care diverts up to 13 kg of CO_2 per medical visit, depending on the distance and frequency of travel to appointments. This translates to up to $0.70 in carbon costs per visit, or cumulatively $44,120 per year in British Columbia, conventionally not included in patient cost estimates.

Conclusions: We present a novel method for estimating patient-incurred costs and CO_2 emissions from accessing health care and apply it to estimate that every year, patients in British Columbia pay upwards of 30 million dollars to access health care services, primarily for medical travel. Our method adds to the economic evaluation literature by providing a more comprehensive and context-modifiable calculation of patient costs that will allow for more informed decision-making regarding health care services.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e56766) doi:10.2196/56766

KEYWORDS

virtual care; economic evaluation; patient costs; lost productivity; informal caregiving; out-of-pocket costs; environmental costs; geospatial; patient; family; CO2; emission costs; health care; Canada; virtual service; emergency department; hospitalization; physician visit; consultation; sensitivity analysis; patient-paid; telehealth

Introduction

Patients and their families incur costs when accessing health care services, even in universal, publicly funded systems such as Canada. However, patient-paid costs are often overlooked or underestimated in evaluations of health services because they are difficult to quantify and often considered out-of-scope for health policy decisions [1,2]. All patients pay to access health services either directly through co-pay cost-sharing arrangements or in other ways, such as needing to take time off work, travel to health care facilities, or attend an appointment as a caregiver. Methods to calculate and analyze patient costs are therefore of interest to account for the degree of cost-sharing that is passed off to patients and to protect them against financial hardship associated with receiving basic medical care. There are concerns about equity in cost-sharing because certain components contributing to patient-paid costs (ie, travel time or visit duration) may vary according to geographic location or service type, resulting in patient-paid costs that depend on where patients live and what type of service they access. Intergenerational inequities may arise from environmental costs to be paid in years to come, including greenhouse gas (GHG) emissions, such as CO_2, associated with the need to travel to receive in-person service visits.

Over the past decade, studies on the patient-paid portion of health care costs have focused primarily on out-of-pocket expenses, such as travel costs [3-7]. Occasionally, some studies also consider time costs by accounting for lost productivity (ie, foregone paid or unpaid work) [8-13] and GHG emissions [13-19]. Studies on time costs and lost productivity have found that the majority of these costs are attributable to the amount of time patients spend attending appointments, estimating that patients pay between $80-$125 (2023 CAD, CAD $1=US $0.86) for their time to attend in-person medical appointments in developed countries [9,10,13]. Of the studies that accounted for GHG emissions and environmental costs, a systematic review found that 0.70-372 kg of CO_2-equivalent emissions were avoided by telemedicine appointments [16]. Many of these studies are based on assumed values for key parameters, such as cost per unit of travel distance, wage, fuel efficiency, and CO_2 emissions, for different patient groups defined by age, sex, or type of service accessed. Costs are typically presented either by determining per-visit costs or analyzing costs over a longer time horizon.

Based on costing methods currently available, we see a significant gap in estimating patient-paid costs. Current methods calculate costs for a specific population subgroup, often including patient data that prevents cost results from being generalizable to other cases. Furthermore, more general studies of patient costs do not include all aspects of costs; instead, they focus on a specific type of cost, resulting in an incomplete picture of the full burden of costs to patients. We propose to construct a multidimensional method for a cost analysis to calculate patient costs and do so in a way that is able to be objectively and consistently applied to different service models, be they in-person or telehealth. In doing so, we present an open-source analytical tool that can be used to calculate these costs generally, using easy-to-obtain or public datasets, and without the need for specific patient-level administration data or costly patient-reported data. Incorporating a geospatial component allows our method to function for both rural and urban areas and can be applied by other health systems to capture patient costs specific to their region. The advantage of this approach is that it provides a standardized way to calculate these costs that will allow for increased opportunities for economic evaluations and comparisons across different service approaches, such as in-person, telehealth, or hybrid care. The aim of this paper, thus, is to present a transferrable method to comprehensively account for patient costs to access health care, inclusive of lost productivity, informal caregiving, out-of-pocket costs, and GHG emissions.

Methods

Overview

The geospatial method was developed to calculate the per-visit costs for health service use from the perspective of patients, their family caregivers, and future generations. The method is inclusive of costs attributed to the receipt of care that are paid at the time of health service use but are not covered by the health care system, in alignment with methods used by the World Health Organization to measure progress on universal health systems' objective to provide financial protection from the high costs of illness [20]. We calculate costs attributed to attending emergency departments (EDs), being admitted to the hospital, and receiving in-person or telehealth physician services. We use publicly available data on distances, appointment durations, and average wage statistics to determine productivity losses for paid or unpaid work with the human capital method of valuation. All data and code used in the method are available in a companion repository on GitHub [21]. Costs to receive care are expressed as service-specific unit costs, comprised of the subunits of lost productivity, informal caregiving, and out-of-pocket expenses, disaggregated by age and region. GHG emissions are expressed as the volume of CO_2-equivalent emissions. All costs are reported in 2023 CAD, approximately equal to US $0.86. Results were simulated for individuals in British Columbia based on the age, region, and service use patterns observed from the real British Columbia population.

XSL•FO
RenderX

Geospatial Evaluation

The geospatial method accounts for geography specific to British Columbia, a province in Canada with approximately 16% rural and 84% urban population (population density ranging between 0.01/km^2 to 18804.48/km^2) [22]. The geographic regions are defined as community health service areas (CHSA) and 5 broader health authority (HA) geographies. At the time of analysis, there were 218 CHSAs defined in British Columbia [23]. Our geospatial method operates primarily at the CHSA level, with results aggregated and presented at the HA level. Travel distances were calculated for hospitalizations, physician visits, and ED visits by determining the geographic center of each CHSA and then calculating the real street distance from each CHSA center to the nearest emergency department, assumed to be a proxy for the nearest location to access care, using road network data obtained through the *OpenStreetMap* package (version 4.0) and calculating with R (version 4.2; R Core Team). Travel duration was also calculated with this method, using speed limits of the specific roads traveled on each trip. The calculated travel distance and duration for each CHSA were aggregated to the HA level using the population-weighted mean of all CHSAs within each HA.

In addition, we also analyzed unit costs for CHSAs using an urban and rural stratification. In the Government of British Columbia record of all health regions, each CHSA is given a classification, numbering 1 through 7, of which 4 and below are considered urban and 5 and above are considered rural [22]. We then aggregated data for all CHSAs considered urban or rural and computed a unit cost for each. This involved calculating travel distance and duration, using our geospatial method, for both urban and rural CHSAs.

Patient and Family Costs

Overview

British Columbia has a publicly funded, single-payer health system, where the provincial government pays for core health services. In our model, costs not paid by the health care system were defined as patient and family costs and were attributed to the following health service types: ED visits, hospital admissions, and telehealth or in-person physician encounters. Costs were calculated for each service type by age group and health authority. For ED visits, costs were calculated separately for Canadian Triage Acuity Scale (CTAS) levels I-III and IV-V [24]. The CTAS is a method for grouping patients according to the severity of their condition upon presentation to an ED, with level I being the most severe and level V being the least severe, resulting in differentials in the amount of time patients spend at ED. We used age groupings of 0-14, 15-64, and ≥65 years, which were chosen to distinguish the working-age population to account for lost productivity and align with Statistics Canada's definitions of working and retirement ages [25,26]. Average hourly wage rate data taken from Statistics Canada begins at age 15, which is consistent with the Employment Standards Act of British Columbia, which allows light work for persons aged 15 without a permit [27,28]. Each calculated cost for each service type and category is comprised of the following subunits: lost productivity, informal caregiving costs, and out-of-pocket expenses. A goal of the method is to create a unit cost database containing all costs, organized by age group, service type, and region.

Lost Productivity

Lost productivity was calculated from the time costs attributed to any foregone paid or unpaid work required to attend appointments or clinic visits. Time costs include time spent on travel, which was obtained by our travel distance calculations, time spent waiting to receive health services, and appointment time (see Multimedia Appendix 1 for details). For physician appointments, ED visits, hospitalizations, and telehealth visits, the total time was multiplied by the average hourly wage rate for British Columbia for 2022. Lost productivity costs were calculated for all working-age patients (15-64 years), and productivity costs were zero for individuals aged 0-14 and ≥65 years. For hospitalizations, due to the increased length of stay, we capped time costs for lost productivity at 8 hours per 24-hour period.

Informal Caregiving

Informal caregiving was evaluated from the time required for caregivers to accompany patients for physician appointments, ED visits, or admissions. For physician, ED, and telehealth visits, we assumed that 50% of all patients aged 15 years and older attend appointments with a caregiver, and 100% of patients in the 0-14 age group attend with a caregiver. We assumed that the caregiver would have the same time commitment for travel time, wait time, and appointment duration as the patient. For hospitalizations for admitted patients, we assumed that 25% of all patients in the 15-64 and ≥65 years age groups and 75% of patients in the 0-14 years age group have a caregiver present for the duration of their visit. For hospital admissions, the caregiver costs were prorated to 8 hours per day to account for lost wages only within a regular workday. Given the lack of clear estimates available in the published literature to draw from, we derived estimates from several Canadian sources and updated them with input from clinical study team members who see patients in clinical practice in order to estimate caregiver attendance percentages [29,30].

Out-of-Pocket Costs

The out-of-pocket subunit includes costs incurred by patients that are a direct result of the appointment but not for the appointment itself. This includes costs such as, but not limited to, vehicle costs, gas, parking, or accommodation. For physician visits, ED visits, and hospital admissions, direct travel costs were calculated based on standard Canadian rates for a compact car (eg, Toyota Corolla) with 80% highway and 20% urban road driving, gas prices of approximately $2 per liter, and average annual mileage of 20,000 km [31]. For physician and ED visits, accommodation costs were assumed to be $0 for all visits as none of the travel distances would require an overnight stay. For hospitalizations, we assumed accommodation would be needed for the caregiver for 2 days if the total travel distance was greater than 50 km to account for extra travel and time necessary on the patient's admission and discharge days. As a result, accommodation costs were only included for hospitalizations in Northern Health Authority and equaled $0 for all other unit costs. When applicable, the accommodation

XSL•FO
RenderX

cost was determined by averaging hotel costs in the area using Google Travel [32]. This accommodation cost was then multiplied by the appropriate caregiver attendance percentage, depending on the age group (75% for age <15 years, 25% for age ≥15 years) to estimate the final caregiver accommodation cost. For ED visits, one meal, at a $15 value, was assumed to be needed for patients in CTAS I-V. For hospitalizations, 3 meals per day for 2 days (admission and discharge days), at $15 value, was assumed to be needed for caregivers. Meal costs were then also multiplied by the appropriate attendance coefficient (75% for age <15 years, 25% for age ≥15 years) to estimate the final caregiver meal costs. Parking costs were determined using rates from a city within the relevant health authority. Average parking rates from the chosen city's core were used for physician visits, and hospital rates were used for ED visits and hospitalizations. Furthermore, for hospitalizations, parking was assumed to be required for the caregiver for 2 days to align with the assumption made for accommodation costs.

Telehealth Visit Costs

Telehealth visit costs were similarly comprised of lost productivity, informal caregiving, and out-of-pocket subunits. However, as many expenses for in-person visits, such as travel-related costs, are not applicable to telehealth visits, the content of the subunits differ. The lost productivity and informal caregiving costs included only appointment wait time and duration, given that travel time is zero. For out-of-pocket expenses, instead of meals, parking, or direct travel costs, the telehealth visit cost included data usage costs. Data usage costs were obtained using estimates of data used for an average telehealth encounter length and using per-unit data costs from the least expensive plans available in British Columbia. Telephone and internet were both considered, based on the current modality of using telehealth care, with 80% weight given to telephone data costs and 20% weight to internet data costs to determine the final data usage rate. Multimedia Appendix 1 shows the details of the full data usage cost calculation. We assumed as a baseline that everyone has access to a phone, so the cost of a device was considered to be zero, and electricity costs were negligible.

Greenhouse Gas Emissions

In-person visits also carry an environmental cost in the form of carbon costs from travel. We included this in our calculations by inputting total travel distance into an emissions calculator [33], using a fuel efficiency rate of 7.06 L/100 km for a 2022 Toyota Corolla. We obtained CO_2-equivalent emissions, in kilograms, per visit for each health authority, based on the distance traveled for a visit in that region.

Sensitivity Analysis

As our method requires assumptions to be made for certain parameter values, we have conducted a sensitivity analysis to explore how the cost output changes based on changes to specific parameters. We conducted the sensitivity analysis on 4 different parameters: wage, travel distance, travel duration, and caregiver attendance percentage. We tested the wage assumption using British Columbia minimum wage ($16.75) [34], $20, $25, $30.54 (base case), and each $5 increment from $35 to $50 (all CAD). The caregiver attendance percentage was tested for 0%, 25%, 50% (base case), 75%, and 100% of appointments attended. Travel distance and duration were varied in 10% increments from 50%-150% of the base case average value. The overall street distance calculation method was also compared with the haversine (straight-line) distance from the center of each CHSA to the nearest in-person visit location. When varying one parameter value, we kept all other parameters constant and applied the method to compare the resulting unit cost output to our base case.

Simulated Population

To illustrate a potential application of this method, we simulated costs and CO_2 footprint for 100,000 individuals distributed across age groups and regions (health authority). We used a Canadian Institute for Health Information (CIHI) report on ED visits to estimate the total number of ED visits per person in British Columbia in one year [35]. We then used NACRS data from British Columbia in 2022 to determine what proportion of ED visits were assigned CTAS I-III and CTAS IV-V in order to apply the appropriate unit cost [35]. For physician visits, we used an international report that found an average of 2.7 provider visits per year in Canada [36]. For telehealth visits, we used a cross-sectional study of telehealth care usage in Ontario, which found an average rate of 48 visits per 1000 people per week, which we then converted to 2.5 visits per person per year [14]. We then distributed these visits between age groups and regions, according to the population distribution in British Columbia, and simulated the impact across British Columbia. Costs were applied to determine the estimated total patient cost by service type for a 1-year period. In addition, we calculated GHG emissions for the simulated population over the 1-year period using the total distance traveled across all ED and physician visits.

Ethical Considerations

All datasets analyzed in this study are publicly available, and their secondary use does not require human consent; therefore, ethics approval was not needed for this study as per the University of British Columbia's Behavioral Research Ethics Board guidelines [37].

Results

Geospatial Evaluation

Travel distances and durations were calculated using our method for each CHSA in British Columbia. The longest distance calculated was 302.37 km for the Mackenzie CHSA, and the longest duration calculated was 429.58 minutes for the Mackenzie CHSA. The shortest distance was 0.59 km for the Fort Nelson Population Centre CHSA, and the shortest duration calculated was 1.21 minutes for the Fort Nelson Population Centre CHSA. Figure 1 shows the resulting map of all CHSA with their calculated route. The street distance calculation method uses a route beginning at the center of each CHSA. In cases where a street is not present, the method may deviate to find the nearest street. Distances and durations calculated from each CHSA were then aggregated to the HA level using a

XSL•FO
RenderX

weighted mean based on the population of each CHSA, the results of which are presented in Table 1.

Figure 1. Map of calculated routes.

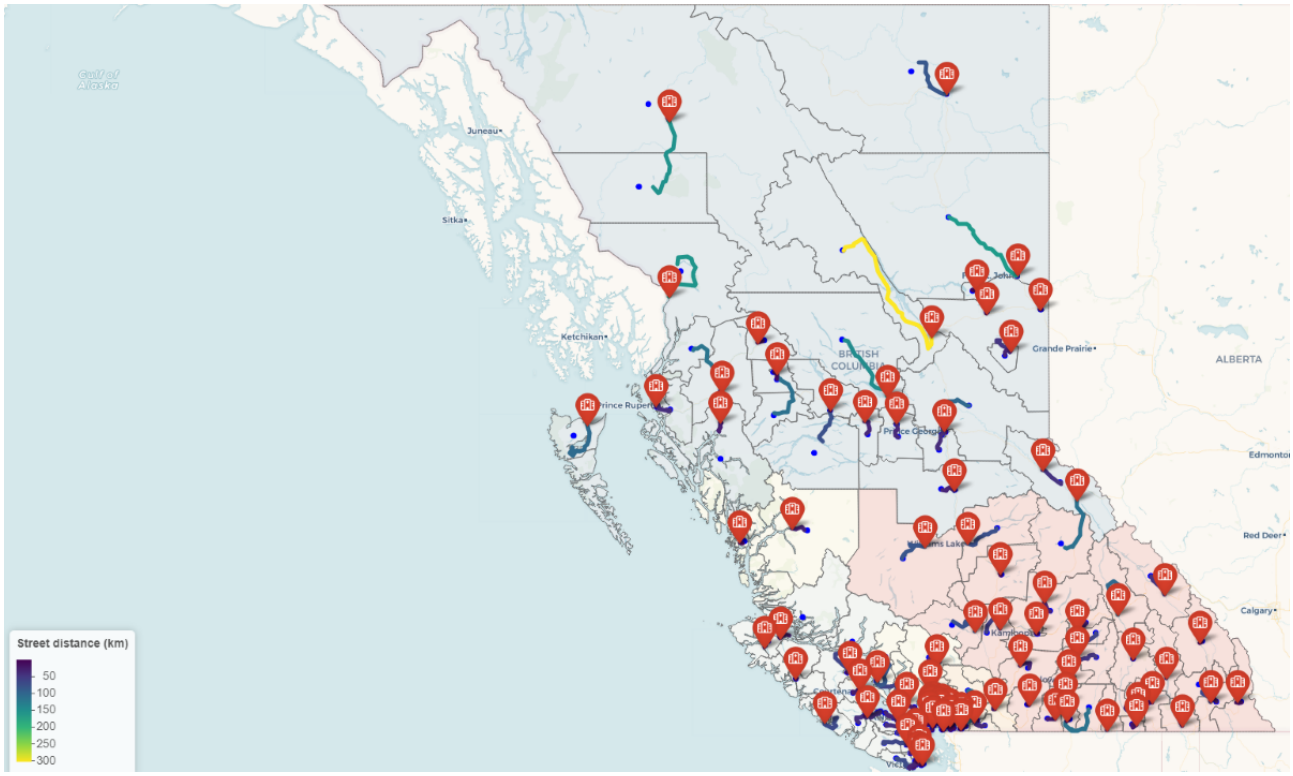

Table 1. Distance and duration by health authority.

Health authority	Weighted mean distance (km)	Weighted mean duration (min)	CO_2-equivalent emissions per visit (kg)
Fraser	30.6	64	4.67
Interior	42.8	65.2	6.53
Northern	69.4	93.8	10.59
Vancouver Coastal	12.06	24.2	1.84
Vancouver Island	24.4	37.8	3.72

Patient and Family Costs

Our method produced individual costs associated with each category (ie, each combination of age group, health authority, and service type), which we have defined as unit costs. Each unit cost, therefore, represents the total cost for one visit of that type. The cost database also includes the cost for each subunit: lost productivity, informal caregiving, and out-of-pocket costs (see Multimedia Appendix 2 for full database). The creation of the cost database allows for the comparison of unit costs across different categories. Costs can be compared between service types, regions, or age groups. From the cost database, we determined the range of costs across the different categories, from a minimum cost of $31 for a physician consult in Vancouver Coastal Health to a maximum cost of $300 for an ED visit in Northern Health.

Visits to EDs with the highest urgency (CTAS IV-V) in the Northern Health region among patients in the 15-64 years age group had the highest costs per visit ($245), considerably higher than the amount for patients accessing the ED in regions covered by Vancouver Coastal Health ($181; Figure 2).

XSL•FO
RenderX

Figure 2. Single visit emergency department (ED) unit costs by region with subunit breakdown. CTAS: Canadian Triage Acuity Scale.

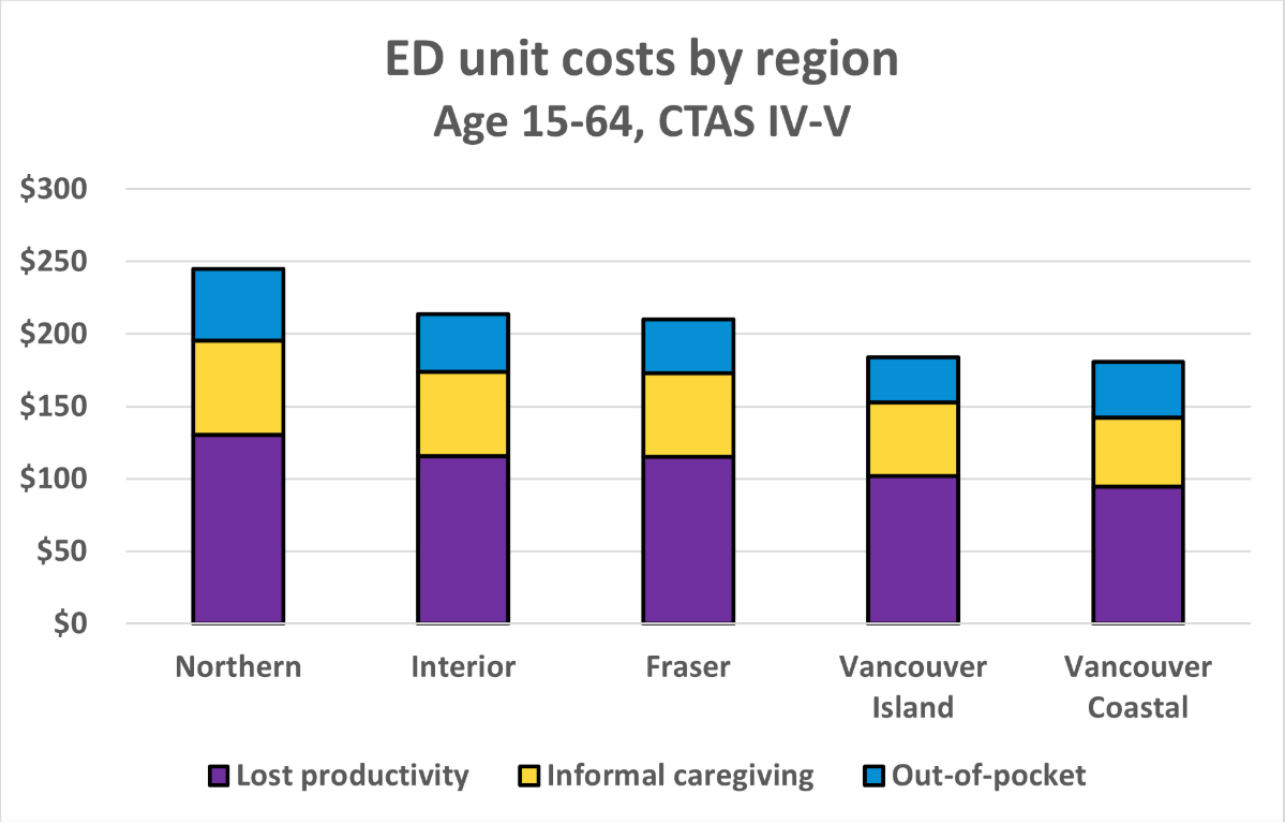

Costs among the 15-64 years age group were driven by lost productivity. This finding relies on the assumption of negligible productivity impacts among the 0-14 and ≥65 years age groups and that the attendance of an informal caregiver is 100% for the 0-14 years group and 50% for the ≥65 years age group. However, when applied to a specific population, it is possible that absolute costs may differ, given that the population may be more heavily weighted toward one or more age groups. We also observe from Figures 3-4 that the geographic trend observed previously continues to hold, with the Northern Health region having the highest costs across all age groups and service types.

Figure 3. Single visit emergency department (ED) unit costs by age group. VIH: Vancouver Island; VCH: Vancouver Coastal.

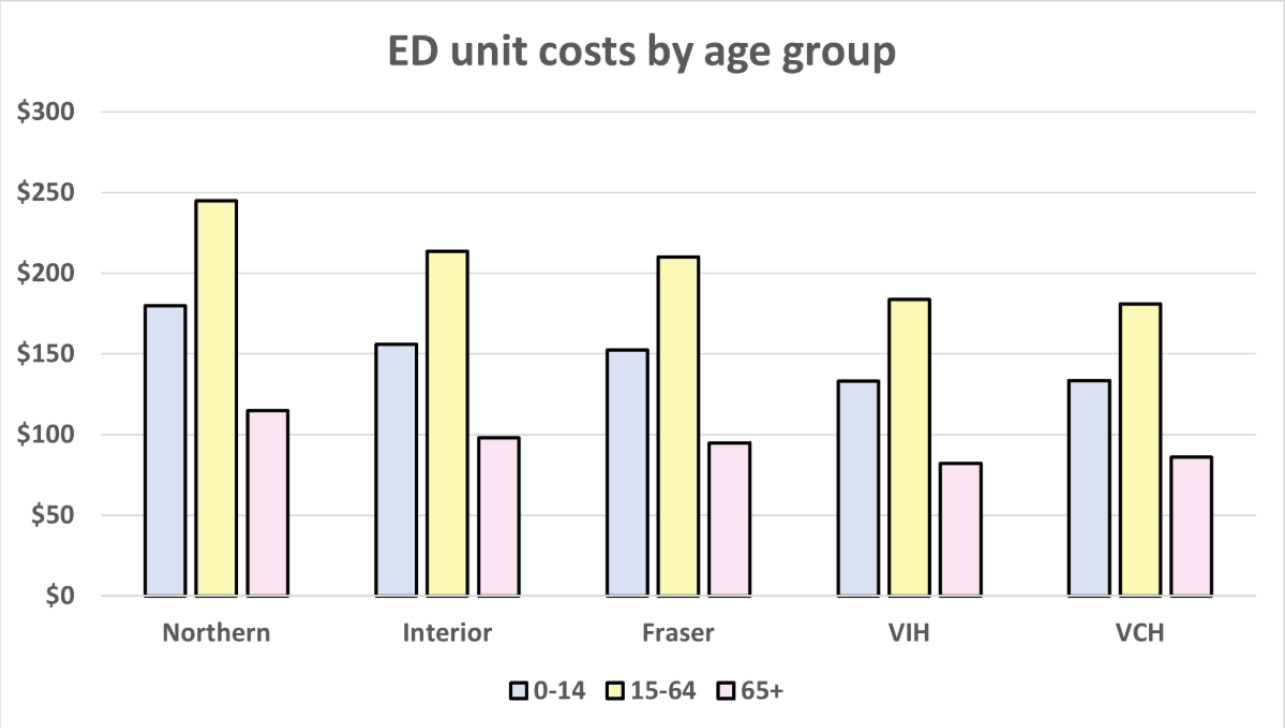

XSL•FO
RenderX

Figure 4. Single visit unit costs by service type and region. CTAS: Canadian Triage Acuity Scale.

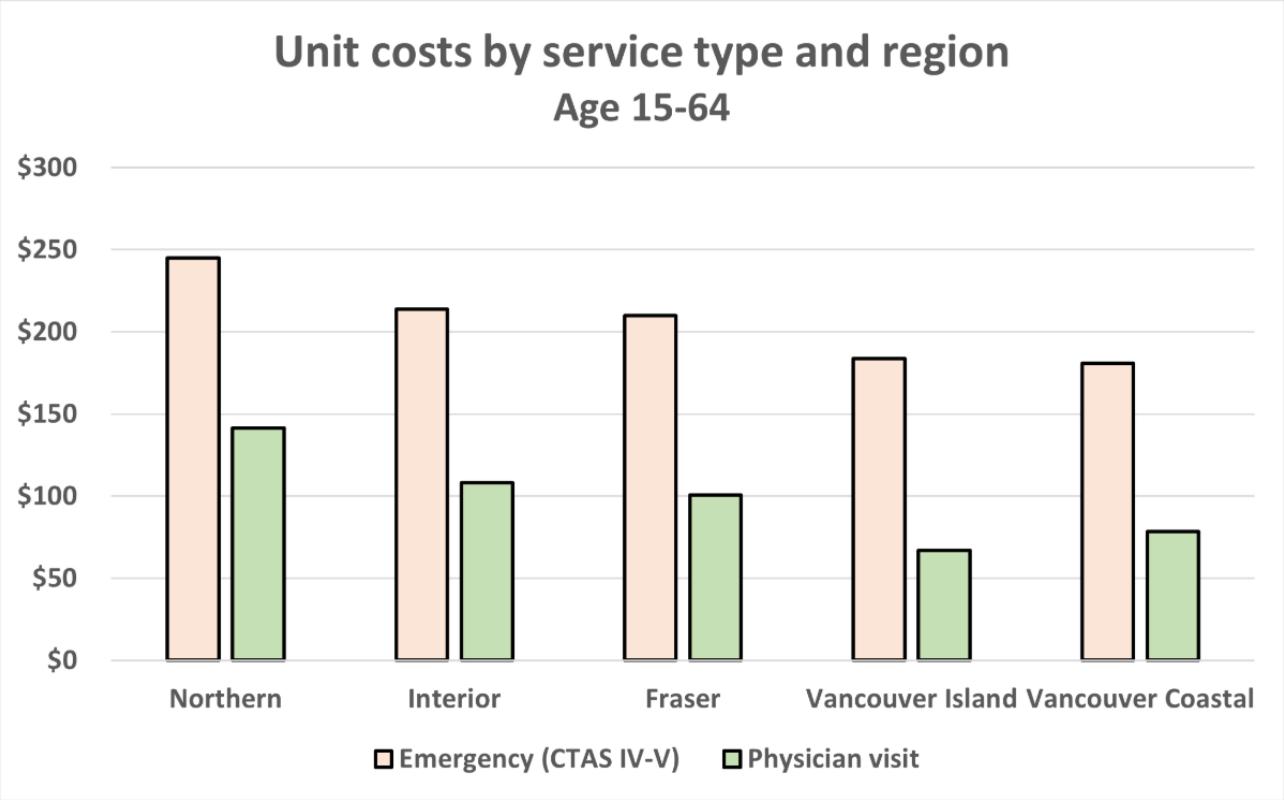

Comparing costs among patients in the age 15-64 group, ED visits at the CTAS IV-V level were higher across all health authorities, ranging from $181 to $245 compared with the estimated $67 to $142 estimated cost for patients to attend primary care appointments.

Unit costs for telehealth visits were found to be $30 for the 15-64 years age group. Given the lack of travel-related costs, telehealth visit unit costs did not vary geographically across the province.

In addition to the aggregation to the health authority level, we calculated unit costs based on a stratification of each CHSA between urban and rural, which are presented in Table 2. This involved aggregating 4 classes of urban CHSAs and 3 classes of rural CHSAs. We then calculated unit costs for physician visits and for ED visits in the CTAS I-III and CTAS IV-V categories. Given the longer travel distances and durations associated with rural locations, we expected to see that rural unit costs are higher than urban for all age groups for both ED and physician service types by a considerable margin.

Table 2. Base per-visit unit costs by urban and rural community health service area (CHSA) classification.

Urban				Rural			
Age group (years)	Physician visit (unit cost), $	Emergency medicine (unit cost), $		Age group (years)	Physician visit (unit cost), $	Emergency medicine (unit cost), $	
		CTAS[a] I-III	CTAS IV-V			CTAS I-III	CTAS IV-V
0-14	52	157	117	0-14	146	250	211
15-64	72	222	162	15-64	196	345	286
≥65	32	92	72	≥65	96	156	136

[a]CTAS: Canadian Triage Acuity Scale.

Greenhouse Gas Emissions

We also determined GHG emissions and costs associated with the travel distance for both the rural and urban CHSAs. We used the same reference vehicle and fuel efficiency (2022 Toyota Corolla, 7.06 L/100 km) and obtained kilograms of CO_2-equivalent emissions per visit, which are shown in Table 3.

XSL•FO
RenderX

Table 3. Base per-visit emissions by urban and rural community health service area (CHSA) classification.

	Distance per visit (km)	CO_2-equivalent emissions per visit (kg)
Rural	91.2	13.92
Urban	18.94	2.89

Sensitivity Analysis

A sensitivity analysis was applied to show how the unit cost outputs change when parameter values are deterministically varied for each of the health services included. The magnitudes of the observed cost variations are similar relative to the base case for each service type and consistently represent a linear 2%-3% change in total unit cost for every $1 change in wage (see Multimedia Appendix 2). Costs in the Northern Health authority had the largest variation and represented the maximum of the range of changes observed. A noteworthy caveat is that although a lower wage value results in lower total costs, this may still represent a larger burden to patients with this income profile. In the case of minimum wage, visit costs are approximately 3.5-4 times the cost of 1 hour of work, whereas in our base case using British Columbia average wage, visit costs are approximately 3-3.5 times the cost of 1 hour of work.

We also varied assumptions about the percentage of appointments attended by a caregiver and found that each 25% change in caregiver attendance resulted in changes of less than $35 in total, except for hospitalizations; the method was sensitive to assumptions about how often a patient has a caregiver attend. The impact changed the estimated patient and family costs by $434-$494, indicating the importance both in the magnitude of this cost overall and also in the significant amount of time that caregivers devote to supporting patients in the hospital. Similar to the analysis of wage changes, costs in Northern Health authority had the largest variation based on a change in caregiver attendance percentage. The observed change in costs for each service type was consistent across all age groups in terms of absolute value. Given that the base case unit costs differ for each service type and age group, the observed change in costs represents a difference in relative change even when the absolute values remain constant. For the 0-14 years age group, a 25% variation in caregiver attendance percentage represents approximately a 17-32% variation in unit costs when compared to the base case. For the 15-64 age group, costs varied by approximately 12.5%-22% compared with the base case. Finally, for the ≥65 years age group, costs varied by 28%-94% compared with the base case. The ≥65 years age group shows significant variation due to 25% caregiver attendance representing a large variation in the absolute value of the unit cost, particularly for hospitalizations, combined with informal caregiving costs being a larger proportion of this age group's total unit cost. Without considering the hospitalization service type, costs varied by approximately 28%-35% for the ≥65 years age group.

Our analysis of the method's travel parameters found that a 10% variation in distance and duration resulted in a change in cost ranging from $2-$8 for patients aged 0-14, $2-$10 for patients aged 15-64 years, and $1-$6 for patients aged ≥65 years. The magnitude of variation remained consistent across all service types. Similar to other parameters included in the sensitivity analysis, the largest variation in costs was observed in the Northern Health authority. While the magnitude of cost variation was consistent for all service types, the relative change in costs did vary due to the difference in the base cost. The observed variation in costs from a 10% change in travel distance and duration represented a change of approximately 0.5%-7.5% compared with the base case. Full sensitivity analysis results for wage, travel, and caregiver parameters are presented in Multimedia Appendix 2.

We also tested the impact of the method used to calculate distance and travel costs within each CHSA by comparing the street distance method with the haversine method. We find that for 167 of 218 (76.6%) CHSAs, the haversine method was within 10 km of the street distance method. The remaining 51 of 218 (23.4%) had a difference greater than 10 km between the two methods; however, for some of the most remote regions, the haversine method underestimated the street distance by up to 137 km.

Simulated Population

We applied our method to our simulated population to determine the total costs of ED, physician visits, and telehealth visits from the patient perspective over 1 year. Demographics and service use for the synthetic population are provided in Multimedia Appendix 1. Total costs are presented in Table 4, organized by service type, acuity, and health authority.

XSL•FO
RenderX

Table 4. Estimated patient-paid costs by service type and acuity for 1 year for a simulated population of 100,000.

Health Authority	Total ED[a] CTAS[b] I-III visit costs ($)	Total ED CTAS IV-V visit costs ($)	Total physician visit costs ($)	Total telehealth visit costs ($)
Fraser	2,292,674	982,692	8,508,221	2,261,474
Interior	1,005,335	431,293	3,987,006	972,891
Northern	424,195	187,685	1,985,431	368,011
Vancouver Coastal	1,350,332	561,400	3,758,996	1,487,933
Vancouver Island	934,499	387,732	3,009,011	1,020,796
Total	6,007,035	2,550,802	21,248,664	6,111,105

[a]ED: emergency department.

[b]CTAS: Canadian Triage Acuity Scale.

We see from this application of the method that substantial costs are likely paid by patients for both ED and physician visits. The simulation showed that ED and physician visits have a cost of just over $30,000,000 over 1 year per 100,000 people. While this total cost is not representative of current absolute costs, it illustrates the type of information that can be obtained by our costing method. Furthermore, we see that while physician visit costs are more than double those of ED visits, ED visits make up just 13.6% (42,573/312,573) of the total visits between the two. This suggests that ED visits are a relatively substantial cost driver despite being smaller in absolute number. In comparison, telehealth visits show considerably less expensive total costs than either ED or physician visits despite a comparable number of total visits to physicians. In addition, GHG emissions were calculated at varying proportions of hybrid care, which assumes both in-person and telehealth visits, the results of which are presented in Table 5.

Table 5. CO_2-equivalent emissions for 1 year for a simulated population of 100,000.

	Estimated number of in-person visits	Total distance (km)	CO_2-equivalent emissions (metric tons)
Physician visits	270,000	3,959,416	586.32
Hybrid care			
25% in-person	67,500	989,854	146.58
50% in-person	135,000	1,979,708	293.16
75% in-person	202,500	2,969,562	439.74
ED[a] visits	42,573	624,311	92.45
60% visits to the ED avoided by telehealth care	25,544	374,586.6	55.47

[a]ED: emergency department.

When applying the method to a population, such as this simulated example, age group definitions can impact total cost output. Labor force participation may vary in complex ways with age or other factors, particularly when youth may start work and when individuals may retire [38,39]. To test the impact of age group definition calculated costs based on age group definitions of 0-19, 20-59, and ≥60 years and compared with the initial results. Total costs for each service type in Table 4 were reduced by 6%-8% with this age group definition.

Discussion

Principal Findings

Our study presents a novel method for estimating the patient-paid portion of costs attributable to receiving medical care with geography as our basis. The core calculation method and the categorization and the types of costs that are included can be applied to any set of parameters, whether they are based on routine data or customized values. Therefore, all parameters are flexible and can be swapped to those relevant to other jurisdictions, allowing the method to be generalizable to many different contexts. Furthermore, when applying the method to populations, labor force participation rates can be applied to change the age group definition or distribution to customized results for the population of interest. Our findings highlight the range and magnitude of cost variations depending on the type of care received, where patients live, and their age. This implies that economic evaluations of telehealth or in-person service from a patient's perspective should consider a range of demographic characteristics of the relevant population to fully capture these costs. Our findings concur with previous research showing that a substantial portion of health care costs are passed off to patients when accessing care, particularly in rural areas [18]. Applying our method to a simulated population estimates that patients in British Columbia may collectively pay up to 30 million dollars per year to access health care services, primarily for medical travel.

XSL•FO
RenderX

Comparison With Previous Work

Our findings add to the literature on patient-paid costs to access health care. These studies, however, are often tied to a specific cohort, using methods to determine travel and other out-of-pocket expenses that consider specific geographic data on patient locations and determine the distance traveled to either the nearest facility or the designated facility appropriate for the relevant treatment. Some methods also use patient location data and then calculate average distances to their facility [8]. Methods used by other researchers to calculate lost productivity—occasionally referred to as opportunity costs—account for total time and multiply by the relevant average wage [9,10,13]. Lost productivity costs using these methods have been estimated at $13.35 (2023 CAD equivalent) for a telehealth visit and $82.35 for an in-person visit [13]. An earlier study estimated the opportunity cost of an in-person visit to be $81.86 [9]. Our method found similar results, with lost productivity costs ranging from $42 to $166 for ED and physician visits. In addition, other studies do not report the per-visit cost directly but estimate cost savings of $124.10 for telehealth care [10]. Depending on the service being analyzed, calculations of patient lost productivity can be done over multiyear time horizons posttreatment and are thus geared toward providing insight on treatment outcomes rather than costs of visits directly. Travel distances, costs, and visit durations contributing to patient cost calculations are also often obtained from patient survey methods [5,7,40]. Survey results have shown that travel-related costs represent a substantial majority of the time spent and the costs that patients pay, with telehealth visits having considerably less cost [5,7,40]. Furthermore, studies of patient travel in Canada show that patients pay an average of $31 of the cost per person for travel to attend primary care appointments [14,41]. By including time costs by lost productivity and informal caregiving, we add the perspective of a comprehensive account of the total costs paid by patients that may be tailored depending on the scope of the economic evaluation. The World Health Organization's methods of quantifying financial hardship currently refer to all costs attributed to out-of-pocket expenses at the time of receiving medical care; however, if there is an opportunity to save patients time during business hours by telehealth care provision, one may argue that accounting for lost productivity would be within the interest of patients, and applicable to providing more patient-centered care.

Strengths and Limitations

Despite a motive for the deliberate embodiment of patient and family costs in evaluation, our method is primarily limited by our stated assumptions and the available evidence about what people pay to access health care. Specifically, our assumptions generalize anticipated differences in wages, labor force participation [38,39], the need for and availability of informal caregiving, fuel, vehicle, and meal costs. In addition, our analysis is limited to an account of the economic and environmental impacts of receiving medical care paid by patients, informal caregivers, and future generations. It does not consider the quality of care, which may result in further avoidable patient costs due to unnecessary services, nor does our method address questions related to the value of care. We rely on local travel maps and assumptions about the day-to-day travel considerations. In doing so, we likely underestimated the impact of seasonal variations, for example, during winter driving conditions or climate-related concerns such as road closures due to flooding or wildfires. The method, in its current form, does not account for constraints on the health care system, such as unanticipated ED closures that will cause patients to travel further. Furthermore, costs that are less easily monetized, such as those attributable to stress and emotional labor involved in emergencies and acute conditions, can create additional costs to patients that are less straightforward to quantify, especially with large distances between many patients and their usual health care centers. Therefore, this may result in the method underestimating the true costs to patients. The inclusion of the sensitivity analysis, while not addressing additional potential costs, is intended to mitigate the impact by showing how costs scale if more weight is given to particular parameters.

Future Directions

This method is applicable to both research and policy contexts for economic evaluations of existing and newly proposed health care services. For example, studies calculating costs for rural health services, especially those with significant travel costs, could be augmented by our method in order to obtain a more comprehensive cost estimate and improve comparability between other studies and evaluations [40,42]. This may be particularly salient in travel reimbursement policies such as Canadian medical transport benefits [43]. Without a comprehensive and accurate calculation of patient costs, the level of burden placed on patients to access care can be unclear and likely underestimate these costs. While it is often indicated that access to care is a greater burden for those living in rural areas, without quantifiable cost information, certain geographic areas and populations may not receive sufficient resources and services. Using the method presented here will allow policymakers to more clearly understand the specific costs of accessing care for all patients in their jurisdiction and thus contribute to a more comprehensive evaluation to inform resource allocation.

Conclusion

In this paper, we present a novel, generalizable method to estimate patient-paid costs and GHG emissions from accessing health care services. Through our geospatial costing method, we are able to show that patients pay substantial costs to access health services. Furthermore, in a simple simulation using historical data, we estimate that patients in British Columbia may be paying millions of dollars to access health care services. By including lost productivity, informal caregiving, and out-of-pocket expenses, the method is able to provide a more comprehensive and accurate calculation of patient-paid costs. Using this method to gain a better understanding of patient-paid costs will allow for more informed evaluations of new and existing services and aid in the comparison between potential alternatives.

XSL•FO
RenderX

Acknowledgments

SC, KS, and GM-P designed the study methods; GM-P and KS analyzed the data. All study authors reviewed the results and contributed to the drafting of the manuscript and its strategy for dissemination.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional explanation of method, values of each parameter, and details of simulated population.
[DOCX File , 26 KB - jmir_v27i1e56766_app1.docx]

Multimedia Appendix 2
Full database of all unit costs obtained from our method. Full details of all sensitivity analyses completed.
[XLSX File (Microsoft Excel File), 134 KB - jmir_v27i1e56766_app2.xlsx]

References

1. Trenaman L, Kaal K, Laba T, Safari A, Aguiar M, Burch T, et al. The financial burden of accessing care for people with scleroderma in Canada: a patient-oriented, cross-sectional survey. CMAJ Open 2023;11(4):E630-E636 [FREE Full text] [doi: 10.9778/cmajo.20220227] [Medline: 37437955]
2. Rai KS, Mann U, Harasemiw O, Tangri N, Eng A, Patel P, et al. A prospective evaluation of patient perspectives and financial considerations during prostate cancer treatment decision-making. Can Urol Assoc J 2023;17(9):E244-E251 [FREE Full text] [doi: 10.5489/cuaj.8228] [Medline: 37458740]
3. Cappitelli A, Wenzinger E, Langa O, Nuzzi L, Ganor O, Rogers-Vizena C, et al. Cost and satisfaction implications of using telehealth for plagiocephaly. Plast Reconstr Surg Glob Open 2022;10(6):e4392 [FREE Full text] [doi: 10.1097/GOX.0000000000004392] [Medline: 35747260]
4. Chan S, O'Riordan A, Appireddy R. Exploring the determinants and experiences of senior stroke patients with virtual care. Can J Neurol Sci 2021;48(1):87-93. [doi: 10.1017/cjn.2020.162] [Medline: 32713397]
5. Vallasciani S, Abdo B, Rauf Z, Anjum A, Ghulman S, Alghammas H, et al. Telehealth for the assessment of patients referred for pediatric urological care: a preliminary cost savings analysis and satisfaction survey. Telemed J E Health 2019;25(8):756-761. [doi: 10.1089/tmj.2018.0159] [Medline: 30359550]
6. Yabroff KR, Guy G, Ekwueme D, McNeel T, Rozjabek H, Dowling E, et al. Annual patient time costs associated with medical care among cancer survivors in the United States. Med Care 2014;52(7):594-601 [FREE Full text] [doi: 10.1097/MLR.0000000000000151] [Medline: 24926706]
7. Finkelstein JB, Cahill D, Young K, Humphrey K, Campbell J, Schumann C, et al. Telemedicine for pediatric urological postoperative care is safe, convenient and economical. J Urol 2020;204(1):144-148. [doi: 10.1097/JU.0000000000000750] [Medline: 31951495]
8. Geerdink TH, Geerdink N, van Dongen JM, Haverlag R, Goslings J, van Veen RN, Virtual Fracture Care Study Collaborative. Cost-effectiveness of direct discharge from the emergency department of patients with simple stable injuries in the Netherlands. Trauma Surg Acute Care Open 2021;6(1):e000763 [FREE Full text] [doi: 10.1136/tsaco-2021-000763] [Medline: 34722930]
9. Ray KN, Chari A, Engberg J, Bertolet M, Mehrotra A. Opportunity costs of ambulatory medical care in the United States. Am J Manag Care 2015;21(8):567-574 [FREE Full text] [Medline: 26295356]
10. Head WT, Garcia D, Mukherjee R, Kahn S, Lesher A. Virtual visits for outpatient burn care during the COVID-19 pandemic. J Burn Care Res 2022;43(2):300-305. [doi: 10.1093/jbcr/irab202] [Medline: 34687201]
11. Jones G, Brennan V, Jacques R, Wood H, Dixon S, Radley S. Evaluating the impact of a 'virtual clinic' on patient experience, personal and provider costs of care in urinary incontinence: a randomised controlled trial. PLoS One 2018;13(1):e0189174 [FREE Full text] [doi: 10.1371/journal.pone.0189174] [Medline: 29346378]
12. White-Means S, Chollet D. Opportunity wages and workforce adjustments: understanding the cost of in-home elder care. J Gerontol B Psychol Sci Soc Sci 1996;51(2):S82-S90. [doi: 10.1093/geronb/51b.2.s82] [Medline: 8785696]
13. Tselapedi-Sekeitto B, Rocha T, Sowerby L, Rotenberg B, Biadsee A. Telemedicine as an environmental ally - the social, financial, and environmental impact of virtual care in the otolaryngology clinic. Am J Otolaryngol 2023;44(2):103791 [FREE Full text] [doi: 10.1016/j.amjoto.2023.103791] [Medline: 36706717]

XSL•FO
RenderX

14. Welk B, McArthur E, Zorzi A. Association of virtual care expansion with environmental sustainability and reduced patient costs during the COVID-19 pandemic in Ontario, Canada. JAMA Netw Open 2022;5(10):e2237545 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.37545] [Medline: 36264577]
15. Ellis I, Cheek C, Jaffray L, Skinner T. Making a case for telehealth: measuring the carbon cost of health-related travel. Rural Remote Health 2013;13(4):2723 [FREE Full text] [Medline: 24299428]
16. Purohit A, Smith J, Hibble A. Does telemedicine reduce the carbon footprint of healthcare? A systematic review. Future Healthc J 2021;8(1):e85-e91 [FREE Full text] [doi: 10.7861/fhj.2020-0080] [Medline: 33791483]
17. Miah S, Dunford C, Edison M, Eldred-Evans D, Gan C, Shah T, et al. A prospective clinical, cost and environmental analysis of a clinician-led virtual urology clinic. Ann R Coll Surg Engl 2019;101(1):30-34 [FREE Full text] [doi: 10.1308/rcsann.2018.0151] [Medline: 30286648]
18. Croghan SM, Rohan P, Considine S, Salloum A, Smyth L, Ahmad I, et al. Time, cost and carbon-efficiency: a silver lining of COVID era virtual urology clinics? Ann R Coll Surg Engl 2021;103(8):599-603 [FREE Full text] [doi: 10.1308/rcsann.2021.0097] [Medline: 34464571]
19. Paquette S, Lin J. Outpatient telemedicine program in vascular surgery reduces patient travel time, cost, and environmental pollutant emissions. Ann Vasc Surg 2019;59:167-172. [doi: 10.1016/j.avsg.2019.01.021] [Medline: 31077768]
20. Households with out-of-pocket payments greater than 40% of capacity to pay for health care (food, housing and utilities approach - developed by WHO/Europe). URL: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4989 [accessed 2023-10-03]
21. Mainer-Pearson G. Geospatial-Patient-Costing-Method-Companion. URL: https://github.com/GrahamMP/Geospatial-Patient-Costing-Method-Companion/tree/main [accessed 2024-02-01]
22. Community Health Service Areas Boundaries - B.C. Health Region Master table - Version 2022. URL: https://catalogue.data.gov.bc.ca/dataset/community-health-service-areas-chsa/resource/874aa151-afe6-400c-876c-aef1ce55102e [accessed 2023-12-14]
23. Community Health Service Areas - CHSA - Datasets - Data Catalogue. Government of British Columbia. 2019. URL: https://catalogue.data.gov.bc.ca/dataset/community-health-service-areas-chsa [accessed 2021-11-17]
24. CTAS introduction. URL: https://ctas-phctas.ca/?page_id=17 [accessed 2023-10-26]
25. Employee wages by industry, annual 2022. Statistics Canada. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410006401 [accessed 2022-12-08]
26. Retirement age by class of worker, annual. Statistics Canada. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410006001 [accessed 2024-10-31]
27. 2016 Census of Population: Age and sex release 2017. Statistics Canada. URL: https://www12.statcan.gc.ca/census-recensement/2016/ref/98-501/98-501-x2016002-eng.cfm [accessed 2022-11-21]
28. Hiring young people. BC Ministry of Labour. URL: https://www2.gov.bc.ca/gov/content/employment-business/employment-standards-advice/employment-standards/hiring/young-people [accessed 2022-11-21]
29. Differences in the characteristics of caregivers and caregiving arrangements of Canadians, 2018. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/220114/dq220114c-eng.htm [accessed 2024-10-31]
30. Stall N. We should care more about caregivers. CMAJ 2019 Mar 04;191(9):E245-E246 [FREE Full text] [doi: 10.1503/cmaj.190204] [Medline: 30833490]
31. CAA Driving Costs Calculator. URL: https://carcosts.caa.ca/ [accessed 2022-12-08]
32. Google Travel. URL: https://www.google.com/travel/ [accessed 2023-12-14]
33. Carbon Footprint Calculator. URL: https://www.carbonfootprint.com/calculator.aspx [accessed 2023-10-05]
34. Minimum wage. British Columbia. URL: https://www2.gov.bc.ca/gov/content/employment-business/employment-standards-advice/employment-standards/wages/minimum-wage [accessed 2023-11-30]
35. NACRS emergency department visits and lengths of stay | CIHI. URL: https://www.cihi.ca/en/nacrs-emergency-department-visits-and-lengths-of-stay [accessed 2023-08-16]
36. Irving G, Neves A, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open 2017;7(10):e017902 [FREE Full text] [doi: 10.1136/bmjopen-2017-017902] [Medline: 29118053]
37. Navigating Ethics Review. The University of British Columbia: Office of Research Ethics. URL: https://researchethics.ubc.ca/behavioural-research-ethics/navigating-ethics-review [accessed 2025-02-14]
38. Krol M, Hosseinnia N, Brouwer W, van Roijen LH. Multiplier effects and compensation mechanisms for inclusion in health economic evaluation: a systematic review. Pharmacoeconomics 2023;41(9):1031-1050 [FREE Full text] [doi: 10.1007/s40273-023-01304-4] [Medline: 37592122]
39. Brouwer W, Verbooy K, Hoefman R, van Exel J. Production losses due to absenteeism and presenteeism: the influence of compensation mechanisms and multiplier effects. Pharmacoeconomics 2023;41(9):1103-1115 [FREE Full text] [doi: 10.1007/s40273-023-01253-y] [Medline: 36856941]
40. Kornelsen J, Khowaja A, Av-Gay G, Sullivan E, Parajulee A, Dunnebacke M, et al. The rural tax: comprehensive out-of-pocket costs associated with patient travel in British Columbia. BMC Health Serv Res 2021;21(1):854 [FREE Full text] [doi: 10.1186/s12913-021-06833-2] [Medline: 34419025]

XSL•FO
RenderX

41. Recommendations for virtual care in primary care practices: a survey of patients across Ontario, Canada. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10549079 [accessed 2024-07-25]
42. Exploration of the costs of accessing health services: data from a longitudinal study of young people in transition from paediatric to adult services. BMC Health Services Research. URL: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-021-06280-z [accessed 2024-10-31]
43. Medical transportation benefits for First Nations and Inuit. URL: https://www.sac-isc.gc.ca/eng/1574177172364/1574177196509 [accessed 2024-10-31]

Abbreviations

CAA: Canadian Automobile Association
CHSA: community health service area
CIHI: Canadian Institute for Health Information
CTAS: Canadian Triage Acuity Scale
ED: emergency department
GHG: greenhouse gas
HA: health authority
NACRS: National Ambulatory Care Reporting System

Edited by A Schwartz; submitted 01.02.24; peer-reviewed by A Ohinmaa, ATG Paulus; comments to author 30.04.24; revised version received 22.08.24; accepted 09.12.24; published 19.02.25.

Please cite as:
Mainer-Pearson G, Stewart K, Williams K, Pawlovich J, Graham S, Riches L, Cressman S, Ho K
Estimating Patient and Family Costs and CO2 Emissions for Telehealth and In-Person Health Care Appointments in British Columbia, Canada: Geospatial Mixed Methods Study
J Med Internet Res 2025;27:e56766
URL: https://www.jmir.org/2025/1/e56766
doi:10.2196/56766
PMID:39969971

©Graham Mainer-Pearson, Kurtis Stewart, Kim Williams, John Pawlovich, Scott Graham, Linda Riches, Sonya Cressman, Kendall Ho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Ubiquitous News Coverage and Its Varied Effects in Communicating Protective Behaviors to American Adults in Infectious Disease Outbreaks: Time-Series and Longitudinal Panel Study

Anqi Shao[1], MS; Kaiping Chen[1], PhD; Branden Johnson[2], PhD; Shaila Miranda[3], PhD; Qidi Xing[4], PhD

[1]Department of Life Sciences Communication, University of Wisconsin - Madison, Madison, WI, United States

[2]Decision Research, Eugene, OR, United States

[3]Department of Information Systems, University of Arkansas, Fayetteville, AR, United States

[4]School of Computing and Analytics, Northern Kentucky University, Highland Heights, KY, United States

Corresponding Author:
Anqi Shao, MS
Department of Life Sciences Communication, University of Wisconsin - Madison
1545 Observatory Drive
Madison, WI, 53706
United States
Phone: 1 6083389939
Email: anqi.shao@wisc.edu

Abstract

Background: Effective communication is essential for promoting preventive behaviors during infectious disease outbreaks like COVID-19. While consistent news can better inform the public about these health behaviors, the public may not adopt them.

Objective: This study aims to explore the role of different media platforms in shaping public discourse on preventive measures to infectious diseases such as quarantine and vaccination, and how media exposure influences individuals' intentions to adopt these behaviors in the United States.

Methods: This study uses data from 3 selected top national newspapers in the United States, Twitter discussions, and a US nationwide longitudinal panel survey from February 2020 to April 2021. We used the Intermedia Agenda-Setting Theory and the Protective Action Decision Model to develop the theoretical framework.

Results: We found a 2-way agenda flow between selected national newspapers and the social media platform Twitter, particularly in controversial topics like vaccination ($F_{1,426}$=16.39; P<.001 for newspapers; $F_{1,426}$=44.46; P<.001 for Twitter). Exposure to media coverage increased individuals' perceived benefits of certain behaviors like vaccination but did not necessarily translate into behavioral adoption. For example, while individuals' media exposure increased perceived benefits of mask-wearing (β=.057; P<.001 for household benefits; β=.049; P<.001 for community benefits), it was not consistently linked to higher intentions to wear masks (β=–.026; P=.04).

Conclusions: This study integrates media flow across platforms with US national panel survey data, offering a comprehensive view of communication dynamics during the early stage of an infectious disease outbreak. The findings caution against a one-size-fits-all approach in communicating different preventive behaviors, especially where individual and community benefits may not always align.

(J Med Internet Res 2025;27:e64307) doi:10.2196/64307

KEYWORDS

risk communication; panel study; computational method; intermedia agenda setting; protective behaviors; infectious disease

XSL•FO
RenderX

Introduction

Background

The shift from traditional to digital media marks a significant change in how the public accesses scientific information, with social media emerging as a central platform [1]. Media coverage may amplify or attenuate risk judgments of its audiences [2], increase risk perception of disease outbreaks [3], and encourage them to take protective actions [4]. Although previous research has explored the relationship between media exposure, risk perception, and protective behaviors during infectious disease outbreaks, the specific impact of distinct media content on the public's adoption of protective behaviors remains unclear. This study aims to bridge this gap by examining how exposure to specific media contents from different sources might alter people's perspectives and willingness to adopt protective measures, building upon but distinct from the broader media agenda-setting literature.

The Role of Different Media in Communicating Preventative Behavior Through the Lens of Intermedia Agenda-Setting Theory

During infectious disease outbreaks, nationwide health messages are warranted for successfully building up healthier collective behavior. It is assumed that such messages, including those to overcome the COVID-19 pandemic, would work if they are clear [5], consistent [6], and come from reliable, nonpolitical sources [7]. These messages typically include descriptions of and recommendations to enact, various protective behaviors for COVID-19. Yet the term "protective behavior" during the pandemic was often used too broadly, covering both simple actions like handwashing and more complex ones like getting vaccinated. It is difficult to effectively address the questions and concerns raised by such varied actions in single messages, or even a single information campaign. This problem is further complicated by other challenges of public health communication, such as engaging diverse communities [8], stigmatization [9], and misinformation, such as antivaccination propaganda [10,11].

Since early 2020, the US Centers for Disease Control and Prevention (CDC) issued guidelines for protective behaviors against COVID-19, such as handwashing and mask-wearing [12]. While some behaviors like mask mandates were communicated top-down from authorities to the public [13], others, like vaccination debates before the widespread availability of COVID-19 vaccines in December 2020, first emerged on social media and then flowed upward (ie, bottom-up) to legacy media [14]. As communication about protective behaviors happens across different media types, it is crucial for scholars to understand the transition of information between traditional media outlets, such as newspapers and television, and social media platforms like Facebook and X (formerly Twitter, during our analysis period).

However, it remains unclear whether these social media discussions are driven by traditional media outlets such as newspapers' coverage or vice versa. Understanding this relationship is crucial because behaviors like mask-wearing and quarantine mandates have different properties that affect their perceived benefits and costs, and may therefore have different origins and dissemination pathways for information about them. Gaining such understanding is key to comprehending how media shapes public health behavior. This leads to our first research focus: investigating the reciprocal influence between legacy newspapers and social media in setting the agenda for COVID-19 protective behaviors. This study specifically looks at the representation of these behaviors, from basic handwashing to complex vaccination debates, in selected national newspapers and on Twitter to discern the pattern and direction of this information flow. The first research question thus aims to dissect this interplay between selected national newspapers, as examples of legacy media, and Twitter, as an example of social media, in setting the agenda for infectious disease protective behaviors, RQ1: in what direction did the agenda for COVID-19 protective behaviors move between the selected top US national newspapers and Twitter?

Intermedia agenda setting serves as a framework for understanding how various news sources influence one another and the consequent breadth of information presented to the public [15]. Originating in political communication, this theory has been applied to health topics to examine the dynamics between traditional and emerging media sources. It highlights the "co-orientation" of narratives within media, influenced by both economic and local social factors [16]. This interaction affects how issues, including health, are framed and perceived, impacting public opinion and decisions. The theory explains how news content transfers between media, illustrating how dominant sources can shape the agendas of others, thereby influencing the entire media landscape [16,17]. For instance, in health communication, significant agenda flows from social media to traditional media have been observed for topics like the Measles, Mumps, and Rubella vaccine-autism debate [18] and climate change [19].

From Communication to Action: How Media Exposure Might Influence People's Protective Behavior

While media agendas on infectious diseases may influence public risk perceptions [20], the evidence of their impact on protective behaviors during pandemics is mixed, particularly in terms of media source and effect. First, there is limited recent evidence on how legacy media use impacts behavioral adaptations. Increased time spent on legacy media can lead to heightened anxiety and stress, which may result in misplaced health-protective behaviors [21]. Official government media use tends to encourage protective actions, while commercial media often results in overprotective behaviors driven by anxiety [22]. Notably, legacy media coverage, including national and local newspapers, television, and radio, has a significant influence on the protective actions of less-educated individuals [23]. Additionally, past research has shown that merely relying on massive media coverage of infectious diseases is not sufficient for promoting protective actions, as other factors like social norms and perceived threats also play crucial roles [24-26].

Second, although social media are key tools for disseminating health messages and can positively influence health-related

XSL•FO
RenderX

behaviors [27-31], their effectiveness in behavior change is not consistent. For instance, increased exposure to media content about Ebola heightened risk perceptions, but a similar increase in exposure to social media content about the Zika virus did not shift risk perceptions among mainland Americans [3,20]. Furthermore, evidence suggests a negative relationship between health-protective behaviors and social media use, where heightened concerns and a lower construal level associated with social media do not encourage preventive actions but instead reinforce negative emotional reactions and mental health problems [32].

In this research, we drew from the Protective Action Decision Model (PADM) for a comprehensive theoretical framework to understand how media agendas of protective behaviors relate to actual individual practices [33]. The PADM framework details how decision-making begins with environmental cues like sight and smell, social cues observed from the behavior of others, and warnings from information sources [33,34]. It specifies prompts for action that are crucial for encouraging engagement in health-related behaviors during infectious disease outbreaks such as H7N9 [35]. The PADM has been applied to the COVID-19 context with a focus on the early stages [36], and in different cultural backgrounds [37,38].

This study collected panel survey data from around 2000 American adults to examine the relationship between media exposure and individuals' behavioral intentions to protect themselves during the COVID-19 pandemic. The dataset spans from early in the US COVID-19 pandemic experience (February 2020) to the period after vaccines became widely available (April 2021). In line with the PADM, which considers intentions as a function of risk perceptions, behavior perceptions, and stakeholder perceptions, this study focuses specifically on behavior perceptions. Since our media content analysis primarily dealt with protective behaviors rather than the overall COVID-19 threat or trust in institutions, we chose to emphasize how respondents perceive the benefits of these behaviors. Based on PADM, we developed our main hypothesis as H1: US individuals' perceived benefit of protective behaviors is positively associated with their media exposure to those behaviors.

We argue that media exposure's impact on behavior is not straightforward but mediated through perceived benefits. According to the original model by Lindell and Perry, before behavior adoption, individuals assess it based on media, evaluating factors such as effectiveness, safety, and relevance. Their perception of these benefits is a crucial factor in deciding whether to adopt certain behaviors, while perceived barriers (such as resource limitations) can hinder the translation of intention into action. During the COVID-19 pandemic, media exposure indirectly influenced health behaviors by promoting protective actions in public spaces through combined perceptions of social initiatives and moderate fear [39], and by positively affecting perceived efficacy and threat awareness [40], leading to improved health-protective behaviors. In other contexts, exposure to health-relevant messages indirectly influences behavioral intentions, such as perceived social norms affecting binge drinking [41] or perceived benefits encouraging regular exercise [42]. This body of evidence, illustrating how media exposure may shape behavior through the lens of perceived benefits, lays the foundation for our second hypothesis H2: the association between media exposure to protective behaviors and the actual practice of these behaviors is mediated by individuals' perceived benefits of these behaviors.

Methods

This section describes our data collection, measurements, and analysis methods for testing our hypotheses.

Data Collection

Example Legacy Media and Social Media Contents

In this study, we used major national newspapers as examples for legacy media, and Twitter as an example for social media. We selected three major US newspapers: the New York Times (NYT), Wall Street Journal (WSJ), and USA Today, due to their high nationwide circulation numbers [43]. This collection of newspapers also represents a range of political ideologies—USA Today is considered centrist, The WSJ leans right of center, and The NYT leans left of center [44]. We collected newspaper coverage of the COVID-19 pandemic via the news data archive Factiva database by searching for the keywords "coronavirus" or "covid-19" between January 2020 and June 2021. Social media data were collected from Twitter using daily searches for the same keywords and within the same time period. While we aimed to focus on US-based conversations, our Twitter dataset may include tweets from English-speaking users outside the United States, as geo-tagging is not always available. In total, we collected 41,235 news articles and 2,398,046 tweets in English mentioning either or both of the keywords "coronavirus" or "covid-19."

We cleaned the media content data by filtering out those irrelevant to protective behaviors. For initial filtering, we used a regular expression search based on a lexicon that covers all protective behaviors suggested by the CDC, such as handwashing and mask-wearing (Multimedia Appendix 1). However, we recognized that keyword-based filtering alone might miss the broader themes in the media content. Therefore, we used structural topic modeling using the "stim" package in R (R Project for Statistical Computing), with the publication date serving as a covariate [45]. This additional step helped us identify 47 topics in news articles and 20 in tweets (the numbers of topics were decided by the searchK diagnosis function in the R "stm" package). Based on several criteria (Multimedia Appendix 2), we selected models of size 47 topics for newspapers and 20 for Twitter posts as optimal topic models). We then excluded content that was not closely related to health, such as international politics or entertainment (for the human validation process, see Multimedia Appendix 2). In the end, 10,528 news articles and 209,261 tweets met our criteria for health-focused protective behaviors. Figure 1 visually represents how these data trends evolved over time.

XSL•FO
RenderX

Figure 1. Standardized coverage from newspapers and Twitter on protective behaviors (the raw volume of daily Twitter coverage is much greater than that from newspapers. Therefore, we conducted a scaling process [ie, calculating z scores] for each set of coverage to make the figure clearer. We also smooth the trendlines with the spline function for better aesthetic quality) over time including (A) mask-wearing, (B) handwashing, (C) avoiding large public gatherings and traveling, (D) isolation at home (self-quarantine), and (E) getting vaccinated. The red lines represent the volumes of legacy media content (specifically from selected national newspapers), and the blue dashed lines represent the volumes of Twitter content.

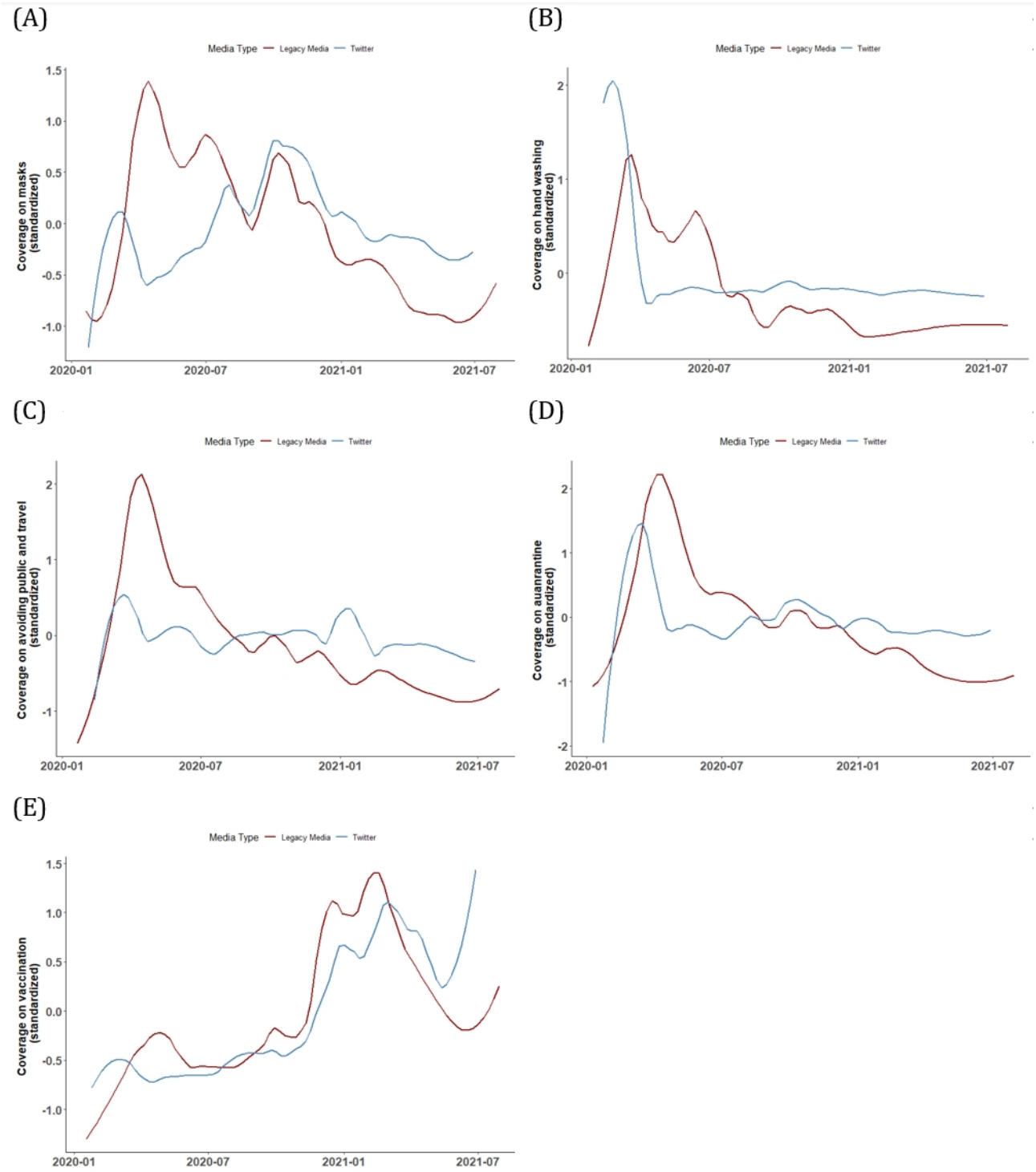

Public Opinion Panel Survey

We recruited 2005 American adults from the Prolific digital survey panel for a longitudinal study spanning 1 year and 2 months, consisting of six waves. In this longitudinal design, respondents from each completed wave were invited to the next wave, with the final wave open to all initial respondents. Each wave lasted about 2 weeks, except for the first. Data collection occurred on February 28-29, 2020 (wave 1), April 27-May 6, 2020 (wave 2), August 5-13, 2020 (wave 3), October 12-22, 2020 (wave 4), January 22-February 11, 2021 (wave 5), and March 25-April 13, 2021 (wave 6). However, not all respondents completed every wave or answered all questions, resulting in some variation in response rates across different waves. We had 2002 responses in wave 1, a total of 1613 responses in wave 2, a total of 1197 responses in wave 3, a total of 1026 responses

XSL•FO
RenderX

in wave 4, a total of 866 responses in wave 5, and a total of 1025 responses in wave 6.

Respondents were asked for their intentions about each protective behavior on a 6-point Likert scale from "My household has never considered taking this action" to "My household has taken this action and will continue to take this action as needed." To assess respondents' media exposure to major national newspapers such as the NYT, WSJ, and USA Today, and for the social media platform Twitter, we asked them about the frequency of using these media sources for COVID-19–related information. Some examples of survey questions can be found in Multimedia Appendix 3.

To measure respondents' perceived benefit of protective behaviors, we asked how beneficial they think it is to practice each protective behavior in reducing risks for their household, and separately for vulnerable people in their community. To assess potential barriers to behavioral intentions, we asked respondents if they had sufficient resources to practice protective behaviors like mask-wearing and handwashing. Perceived barriers were coded as "resources" and were included in our analysis as a control on intentions, as shown in Figure 2. We also recorded demographic information, including age and gender, during the first wave.

Figure 2. Multilevel regression model visualization, with independent variables categorized into time-varying (eg, media exposure, perceived benefits, case updates) and time-unvarying factors (eg, demographics), with fixed and random effects to capture perceptual and behavioral changes over time.

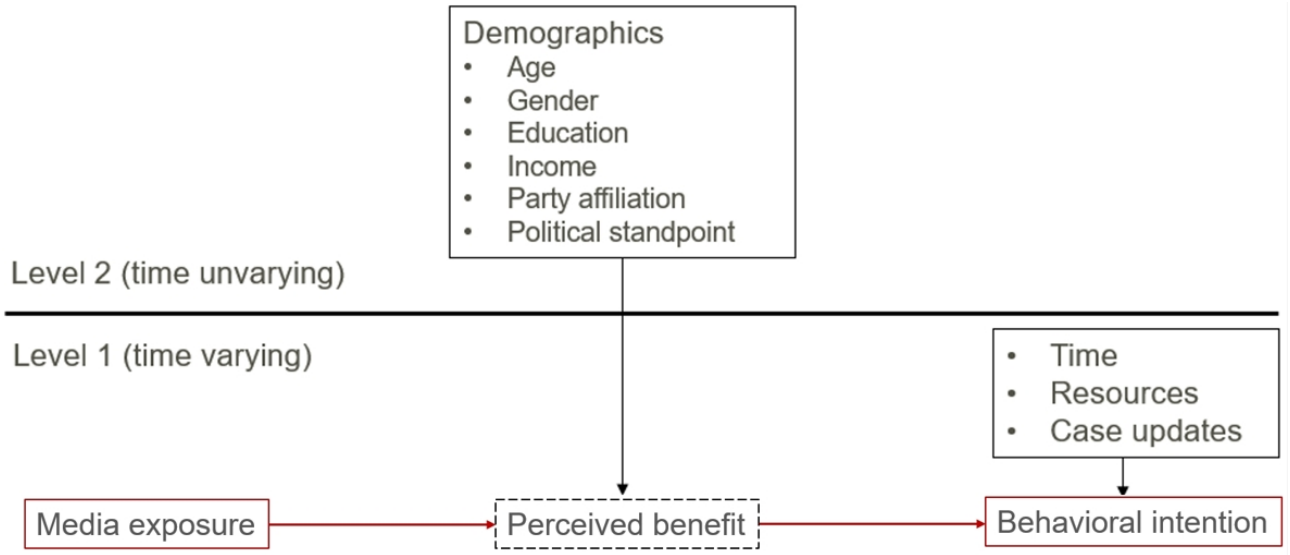

Ethical Considerations

Data collection from human participants was approved as exempt by the Decision Research institutional review board chair. After reading an informed consent form at each wave, outlining survey aims, benefits, risks, and other relevant information, consent was indicated by the respondent checking a box for consent. Anyone who checked the "No" box was not allowed to proceed to the survey itself. The prolific panel operator recruited all participants at wave 1, using a compensation scheme that is proprietary (ie, researchers were not informed of its type or amount), and reported all responses to researchers with a unique identifier, so that researchers were unaware of the identity of any respondent. While the survey was not password-protected, it was a closed survey in practice, as only responses from participants with matching unique identifiers were included in this analysis.

Granger Predictive Relationship in Time Series (RQ1)

The Granger causality test is a statistical tool commonly used to infer whether one time series can predict another. It does not establish true causality but rather identifies predictive relationships between variables. This method is often used in intermedia agenda-setting research to examine the lagged predictive influence of one platform's agenda on another [17,46,47].

The Granger causality test requires data to be stationary, as it would otherwise produce unreliable results [48]. Before running this test, we used the augmented Dickey-Fuller stationarity check, which sets a null hypothesis that the data series is nonstationary. Our tests yielded *P*-values less than .05, leading us to reject the null hypothesis for each data series. Therefore, we used Granger tests to explore the predictive relationship between daily newspaper coverage and Twitter activity for each protective behavior. We used daily coverage summary data for both newspapers and Twitter and conducted Granger tests in both directions, that is, from newspaper coverage to Twitter activity and vice versa. This allowed us to examine the predictive influence of each platform's agenda on the other for various protective behaviors.

In this study, statistical analyses were run on R (version 4.1.2), and the time-series analyses in this session were conducted with open-sourced packages "tseries" [49] for the stationarity check and "lmtest" for Granger causality detection [50].

Modeling Respondents' Media Exposure and Protective Behavior (H1)

Comprehensive media coverage of protective behaviors does not indicate everyone gets the same information. To avoid making arbitrary assumptions about individual media exposure, we chose not to solely rely on either "media coverage" or survey-reported "media exposure frequency." Instead, we

XSL•FO
RenderX

calculated a "weekly media exposure" metric for each respondent by multiplying their self-reported frequency of using specific media outlets, like the NYT or Twitter, with the daily proportion of relevant coverage from those outlets. To estimate this weekly media exposure, we used one-week "exposure windows" prior to each survey wave. For each type of protective behavior, such as mask-wearing or handwashing, we calculated the specific daily proportion of coverage related to that behavior in selected top national newspapers and Twitter. Therefore, the media exposure metrics for each respondent are dynamic across different waves of survey and different types of protective behaviors. Details for calculating the media exposure can be found in Multimedia Appendix 4.

According to the PADM, the behavioral response may also be impacted by situational facilitators [33]. For situational facilitators in this study, we referred to the Center for Systems Science and Engineering at Johns Hopkins University for daily confirmed case updates in the United States from January 2020 to June 2021 [51]. More specifically, we chose to present average weekly and monthly trends of their local case updates (ie, the gradient of daily confirmed cases in the same zip code area) before each response date.

We constructed our model with multilevel modeling. As shown in Figure 2, we categorized the independent variables into two sets. The first set includes time-varying variables like respondents' media exposure, perceived benefits, and situational factors, which change with each survey wave. The second set includes time-unvarying variables like demographics, which remain constant throughout the survey. In total, we created five models, each corresponding to one of the five protective behaviors previously mentioned. To test hypothesis H1, we examined the relationship between media exposure and perceived benefits for each protective behavior. For H2, we also looked at how these perceived benefits influence the intention to practice these behaviors. We chose the "lme4" package [52] in R because it can handle our multilevel data, which includes fixed effects like demographics and random effects showing changes in respondents across surveys. This package is effective for modeling individual changes in behavior and perception over time, thus providing a more unbiased view of how behavior changes due to media exposure and perceived benefits. To ensure comparability and interpretability of coefficients, numeric variables were standardized before analysis.

Results

Intermedia Agenda Setting Between Selected National Newspapers and Twitter

Table 1 presents the results of how newspaper coverage of various protective behaviors might predictably influence Twitter discourse. Table 2 shows the reverse results about how Twitter activity might predictably influence newspaper coverage.

Table 1. Granger test results (F statistics) of predictive causality from newspapers to Twitter posts about protective behavior issue agendas.

Lag (in days)	Wash hands, *F* test (df)	Wear masks, *F* test (df)	Self-quarantine, *F* test (df)	Avoid gathering and travel, *F* test (df)	Take vaccination, *F* test (df)
1	3.04 (1, 163)	0.19 (1, 439)	3.62 (1, 434)	1.51 (1, 368)	16.39 (1, 426)[a]
2	0.52 (2, 131)	0.18 (2, 416)	1.19 (2, 410)	0.28 (2, 343)	14.54 (2, 398)[a]
3	0.45 (3, 110)	0.28 (3, 402)	0.63 (3, 398)	0.19 (3, 331)	8.17 (3, 381)[a]
4	0.85 (4, 92)	1.04 (4, 392)	0.44 (4, 392)	0.25 (4, 321)	4.97 (4, 371)[a]
5	0.69 (5, 77)	0.86 (5, 390)	0.57 (5, 386)	0.26 (5, 312)	3.89 (5, 361)[a]

[a]*P*<.001.

Table 2. Granger test results (F statistics) of predictive causality from Twitter posts to newspapers about protective behavior issue agendas.

Lag (in days)	Wash hands, *F* test (df)	Wear masks, *F* test (df)	Self-quarantine, *F* test (df)	Avoid gathering and travel, *F* test (df)	Take vaccination, *F* test (df)
1	2.75 (1, 163)	0.14 (1, 439)	9.68 (1, 434)[a]	1.57 (1, 368)	44.46 (1, 426)[b]
2	0.99 (2, 131)	0.34 (2, 416)	4.62 (2, 410)[c]	1.54 (2, 343)	19.05 (2, 398)[b]
3	0.57 (3, 110)	0.37 (3, 402)	3.14 (3, 398)[c]	0.42 (3, 331)	11.81 (3, 381)[b]
4	2.14 (4, 92)	0.48 (4, 392)	2.72 (4, 392)[c]	0.41 (4, 321)	9.31 (4, 371)[b]
5	1.45 (5, 77)	0.41 (5, 390)	1.46 (5, 386)	0.70 (5, 312)	7.58 (5, 361)[b]

[a]*P*<.01.

[b]*P*<.001.

[c]*P*<.05.

Our first research question (RQ1) explored the relationship between selected national newspapers and Twitter in shaping agendas around protective behaviors. Contrary to our expectations, the results do not support a clear predictive

XSL•FO
RenderX

relationship in either direction for certain behaviors like handwashing ($F_{1,163}$=3.04; P=.08), mask-wearing ($F_{1,439}$=0.19; P=.66), and avoiding gatherings ($F_{1,368}$=1.51; P=.22). In these cases, no significant patterns were observed, suggesting a lack of agenda influence between newspapers and Twitter.

However, for self-quarantine, Twitter activity was found to predictably influence subsequent newspaper coverage ($F_{1,434}$=9.68; P=.002 at a one-day lag). In the case of the more contentious, high-impact behavior of vaccination, a reciprocal relationship was observed: newspaper coverage appeared to shape Twitter discussions ($F_{1,426}$=16.39; P<.001 at a one-day lag), and Twitter activity, in turn, seemed to influence newspaper coverage ($F_{1,426}$=44.46; P<.001 at a one-day lag).

Predicting Behavioral Response With Media Coverage and Real-World Data

To address the variability in behavioral responses to media coverage of protective behaviors, this section presents a detailed examination of two hypotheses. Our first hypothesis (H1) proposed that individuals' perceived benefit of protective behaviors increases with their media exposure to these behaviors. Additionally, in our second hypothesis (H2) based on PADM, we expected that people's perceived benefit of protective behaviors is positively related to their media exposure. Here, we observed varied patterns from our multilevel regression in the relationship between respondents' media exposure, perceived household benefits, perceived community benefits, and behavioral intention (Bintn; Multimedia Appendix 5).

Access to resources (β=.097; $F_{1,5812}$=64.40; P<.001) and perceived household benefits (β=.294; $F_{1,5812}$=330.55; P<.001) significantly predicted handwashing behavior. Contrarily, perceived community benefits did not significantly influence handwashing (β=–.010; $F_{1,5812}$=0.50; P=.48). Therefore, H1 is only partially supported in the context of handwashing behavior. Besides, the weekly trend of increasing disease cases also positively predicted handwashing (β=.024; $F_{1,5812}$=5.69; P=.02), reflecting behavior change in response to case numbers. However, Twitter exposure had a negative association with both perceived benefits to the household (β=–.058; $F_{1,7589}$=24.70; P<.001) and community (β=–.030; $F_{1,7592}$=5.73; P=.02). Weekly exposure to the selected national newspapers is negatively associated with perceived benefits to the household (β=–.085; $F_{1,7592}$=44.46; P<.001) and to the community (β=–.090; $F_{1,7592}$=45.67; P<.001). These results suggest that increased exposure to Twitter and the newspapers negatively correlates with perceived benefits of handwashing for both the household and community, rejecting H2 in the handwashing scenario.

Regarding mask-wearing behavior, we found that respondents' behavioral intention is significantly predicted by resources (β=.196; $F_{1,5288}$=369.02; P<.001), benefit to the household (β=.164; $F_{1,5064}$=149.13; P<.001), and benefit to the community (β=.101; $F_{1,5465}$=57.09; P<.001), supporting H1. Twitter exposure positively influenced perceived household benefits (β=.057; $F_{1,6403}$=24.70; P<.001) and community benefits (β=.049; $F_{1,6312}$=17.35; P<.001), while news exposure to the three selected newspapers was positively related to household benefits (β=.031; $F_{1,6442}$=6.68; P=.009) but negatively to community benefits (β=–.029; $F_{1,6259}$=5.34; P=.02). These findings were partly in line with H2, which suggests a positive link between media exposure and perceived benefits of mask-wearing behaviors.

For avoiding public gatherings and travel, both perceived benefits for community (β=.044; $F_{1,5698}$=5.13; P=.02) and for household (β=.348; $F_{1,5723}$=280.26; P<.001) significantly predicted respondents' self-reported behavioral intention, supporting H1. Access to resources (β=.157; $F_{1,5688}$=171.46; P<.001) and increasing weekly case trends (β=.035; $F_{1,4468}$=12.78; P<.001) also contributed to this avoidance behavior. Twitter (β=.058; $F_{1,5859}$=28.62; P<.001) and exposure to the three selected national newspapers (β=.031; $F_{1,6458}$=6.70; P=.009) were positively related to perceived benefits for households, while Twitter exposure alone (β=.051; $F_{1,5979}$=20.60; P<.001) enhanced perceived community benefits, partially in line with H2.

Resources (β=.338; $F_{1,5559}$=757.64; P<.001) and perceived household benefits (β=.268; $F_{1,5559}$=246.06; P<.001) were key predictors of quarantine behavior, which aligns with H1. Perceived community benefits (β=.029; $F_{1,5559}$=3.29; P=.07) showed a positive but not significant trend. Exposure to the selected national newspapers negatively impacted perceived benefits for both the household (β=–.081; $F_{1,5617}$=16.16; P<.001) and community (β=–.090; $F_{1,5617}$=12.80; P<.001), which did not support H2 for quarantine behavior.

Resources (β=.176; $F_{1,5808}$=136.69; P<.001) and perceived household benefits (β=.377; $F_{1,5808}$=280.91; P<.001) significantly predicted intentions to vaccinate, suggesting alignment with H1 for household perception. Weekly case trends (β=.026; $F_{1,5808}$=5.40; P=.02) also positively predicted vaccination intentions. Twitter exposure positively affected the perceived benefits for households (β=.073; $F_{1,7590}$=34.03; P<.001) and communities (β=.058; $F_{1,7591}$=18.36; P<.001), indicating support for H2 with respect to social media influence. In contrast, exposure to the selected newspaper sources negatively correlated with household benefits (β=–.026; $F_{1,7590}$=4.24; P=.04) and was not significant for community benefits (β=–.013; $F_{1,7591}$=0.96; P=.33), which presents a mixed outcome for H2 when considering different media sources.

Interestingly, an inspection of the random effects analysis highlights that time-related variance in vaccination behavior (σ^2=0.016) was lower than individual differences (σ^2=0.437). This indicates that differences in vaccination behavior across different time points were relatively small compared to the differences between individuals. The random effects showed substantial variability in baseline vaccination behavior across individuals (σ^2=0.718), suggesting that some individuals were consistently more inclined to vaccinate than others. Additionally, there was a tendency for those with higher initial intentions to experience smaller changes over time (σ^2=0.046, corr.=–0.90), implying that people with strong initial intentions to vaccinate

XSL•FO
RenderX

were less likely to significantly increase or decrease their intentions across the study period.

Discussion

Principal Findings

Our findings advance the literature on health communication in several ways. First, our research contributes to the intermedia agenda-setting theory within the field of health communication that social media platforms, such as Twitter, can set the agenda for some legacy media such as the selected newspaper outlets in our paper, by revealing nuances in patterns across protective behaviors in a pandemic. We noticed no significant agenda flow between the selected newspapers and Twitter in handwashing and mask-wearing, and in avoiding public gatherings. By contrast, there is also a two-way agenda flow—from selected newspapers to Twitter and from Twitter to those newspapers—about vaccination. Contrary to our expectations, the agenda for self-quarantine originated from Twitter and predicted legacy media outlets' content. That is, our results implied that Twitter users' discussion on quarantining causes the coverage of quarantine in these legacy media outlets we chose as examples. One explanation could be the massive mental health concerns reflected on social media platforms early in the pandemic in the United States, considering the widespread policy of quarantine at that time [53-55]. These studies identified Twitter (or other social media) users' mental health concerns at home at an early stage of the pandemic (ie, at or before March 2020). However, it was not until mid-June to July 2020 that there was a boost in these three US top national newspapers' coverage of mental health issues [56].

Second, our research contributes to the PADM by demonstrating that the relationship between media exposure and protective behavior adoption is contingent upon the specific behavior in question. For certain protective behaviors like avoiding public gatherings and travel and taking up vaccination once vaccines become available, increased exposure to the selected top national newspaper content positively influences individuals' perceived benefits and behavioral intentions. This aligns with existing literature on the positive impact of media on health behavior [5]. The controversy about vaccinations—such as the misleading vaccine-autism relationship and vaccination mandate debate—mostly happened on social media platforms such as Twitter [18]. Legacy media outlets, on the other hand, have to shoulder the responsibility for communicating scientific and other authorities' views about appropriate behavior in the pandemic. Therefore, the positive relationship found between exposure to vaccination-related media content and perceived benefit is understandable.

We also see unexpected, but interesting, negative relationships in quarantine and washing hands. People's perception of the benefits of quarantining might decrease or even backfire in the later stages of the pandemic. This could be due to two factors: (1) mental health concerns arising from social isolation, as previously discussed (eg, [53]), and (2) improvements in nationwide pandemic control [57], leading some states to cancel stay-at-home mandates. Whatever the causes, we observed that respondents' self-reported practices of quarantine decreased over time. We advance PADM testing by showing that, while exposure to information sources such as national newspapers and Twitter can indeed affect perceptions about protective behaviors and behavioral intentions as the model posits, we find that this impact varies across both behaviors and media platforms. For some behaviors, such as handwashing and avoiding gatherings, both types of media exposure had the same effect on intentions, whereas for others, such as mask-wearing and vaccination, the media platforms did not contribute equally to the public's behavioral outcomes, with opposite signs for national newspapers and Twitter. There were also differences within a given media type: for instance, while Twitter exposure positively predicts the perceived benefits of mask-wearing and quarantine behaviors, it is negatively associated with the perceived benefits of handwashing in preventing COVID-19. This underscores the need for targeted message strategies that consider platform-specific outcomes to promote public health actions more effectively.

Third, we examined the relationship between media-influenced benefit perception and individuals' behavioral intention. The positive relationship between perceived benefit and behavioral adoption is within our expectations and is consistent with suggestions from multiple health behavior-related theories, such as the Theory of Planned Behavior [58]. We also noticed negative relationships between the perceived benefits of vaccination and handwashing for the community, and behavioral intentions. Currently, we cannot assertively attribute this phenomenon to a definite theory, but it may be related to egoism versus altruism or prosocial intentions. When individual interests conflict with collective interests in considering vaccination, this situation can be seen as a social dilemma and may result in vaccine hesitation. Our results implied that beliefs about the altruistic effects of these two behaviors undermined intentions, while communication findings elsewhere yielded mixed findings. Altruism-centered messages, aiming to increase individuals' perceived benefit of vaccination for the community, showed no significant results in increasing vaccination behavior as well [59]. However, while messages conveying self-protection had no effect on vaccine intentions, altruistic messages emphasizing protecting other individuals, population health, and the economy had substantially stronger effects [60], while individuals are responsive to altruistic messaging about vaccination [61].

Limitations and Future Research

Although we tried to simulate an authentic environment for individuals reading news articles and making decisions, our model cannot replicate the entire decision-making process. The current model suggests that individuals read news articles about protective behaviors, assess their usefulness, consider available resources, check pandemic updates, and then decide whether to adopt these behaviors. However, there are several factors in the PADM missing in this story: for example, it ignores the effects of risk perceptions, or of trust in authorities or the information source, which are also associated with behavioral intentions [36]. There are also limitations in how we calculate the exposure to relevant Twitter content due to the inherent differences in how content is consumed on the platform compared to the national newspaper we selected. In newspaper

XSL•FO
RenderX

outlets like the NYT, the number of studies on a given topic is relatively limited and stable, allowing for a straightforward calculation of exposure based on study frequency and reader habits. Twitter's dynamic, high-volume nature, however, makes a simple multiplication-based metric less meaningful compared to those newspapers.

It is worth noting that the PADM theorists mentioned the factors "should form causal chains" [33], while our statistical test for this model can only convince us of correlation instead of solid causal relationships. The varied patterns in media exposure's relationship with individuals' risk perceptions prompt us to reconsider how we filter media content related to protective behaviors. We focused only on the content discussed, specifically whether keywords were mentioned (first step of dictionary-based filtering) and whether the topic was health-related (second step of topic-based filtering). We did not look into how they discussed these protective behaviors. For example, media coverage about vaccination could vary in tone as the development of effective COVID-19 vaccines evolved from uncertain to clinically proven during our six survey waves [62].

Additionally, this study is US-based, with a US national survey and US national newspapers in analyses, which limits the generalizability of our findings to a global perspective. Although we collected tweets in English and aimed to capture US users' conversations, our dataset may include tweets from English-speaking users outside the United States due to limitations in geo-tagging availability. This introduces a degree of uncertainty in assuming that all social media data reflect US perspectives. Furthermore, there are well-documented differences in public risk perceptions and protective behaviors across countries, influenced by sociocultural factors such as individualistic or prosocial values, trust in authorities, and personal experience with the virus [63]. When applying the US-based findings to other contexts, we cannot ignore the role of cross-cultural variations and how they differently shape people's responses to health risks.

Communication during the COVID-19 pandemic has faced a concern of an "infodemic" with wicked challenges like misinformation, resistance, and fear [64,65]. Exposure to this infodemic might further elicit individuals' worries about uncertainty and information-seeking behavior [66,67]. Therefore, we believe studying the agenda consistency between different media platforms is essential for effective pandemic management. However, scholars have long been concerned that media can have multiple roles in communicating risk-protective behaviors to lay audiences, and different channels (eg, legacy media vs nonlegacy media such as social media) can have different communication effects [4]. The transmission of quarantine agendas from Twitter to the selected top national newspapers, as observed in this study for two of the five behaviors, along with the differing perceived benefits for household versus community protective behaviors, caution practitioners to consider multiple factors that may be salient to audiences when promoting protective behaviors. This study also calls for further steps for building up more profound and solid models in predicting individuals' behavioral intentions in health communication research.

Our findings carry several actionable strategies for health communicators and decision makers to organize and publish media content. First, practitioners can adopt a platform-specific approach. Social media platforms like Twitter and legacy media such as the national newspapers can influence public perceptions in different ways. For Twitter, practitioners should stay aware of trending conversations about key health behaviors (eg, vaccinating) and engage in addressing misinformation or reinforcing protective behaviors, including changes over time (eg, [68]). Such engagement can help in spreading accurate information and serve as a feedback mechanism with metrics such as engagement rates.

Second, although exposure to some legacy media outlets (eg, national newspapers) can raise awareness and perceived benefits of protective behaviors like vaccination, our data show that this awareness does not always translate into self-reported behavior change. For example, high newspaper exposure was associated with lower perceived household and community benefits for quarantine, even though these perceived benefits were a positive indicator of behavioral change. Therefore, practitioners should design campaigns that not only raise awareness but also emphasize actionable steps of specific protective behaviors, such as tailoring clear instructions and feasible solutions to barriers. Public opinion shows there are barriers for individuals to fully follow "safer at home" policies, including emotional strain and practical challenges [69]. By acknowledging these issues, communicators can offer practical solutions and support, enhancing effectiveness, other than merely increasing the behavior's presence in media.

Third, increased media exposure can sometimes have unintended effects by reinforcing negative perceptions, as seen in our findings where both Twitter and newspaper exposure were linked to lower perceived benefits of handwashing. To avoid these unintended consequences, practitioners should pretest messages with diverse audiences to identify potential misinterpretations or negative framing before widespread dissemination. Additionally, focusing on clear, fact-based messaging that emphasizes practical benefits and personal relevance can help reduce skepticism. The overwhelming amount of unclear, ambiguous, and inaccurate information during the early stage of the COVID-19 pandemic contributed to information overload, which heightened health-related anxiety and increased the spread of misinformation [70,71]. Clear communication can help counteract these negative effects by providing trustworthy, straightforward guidance.

Finally, it is important to consider whether the heightened media consumption during the global pandemic translates into healthy communication strategies for more "normal" times, when media use may be more fragmented. Most of our survey respondents (85%) reported in the first wave that they were using their usual media channels for COVID-19 information, with the exceptions primarily COVID-19–specific outlets (eg, US CDC and World Health Organization websites). Furthermore, unpublished analyses of patterns of self-reported media exposure across different media types (eg, the most common was moderate use of national newspapers, digital news, and social media) found that about 10% of people shifted these patterns from one wave to another, again suggesting that there was mainly stability in

XSL•FO
RenderX

media use from before to during the pandemic through April 2021, the last point at which we collected these data. However, media consumption patterns may shift post pandemic, requiring communicators to adjust their strategies.

Conclusions

While news consumption surged during the pandemic [72], postpandemic media engagement has stabilized but remains above prepandemic levels, and audiences are engaging with more new technology [73,74]. Considering this fragmented attention, health communication practitioners should focus on targeting specific platforms, moving away from blanket exposure and toward tailored content that resonates with diverse, targeted audiences [75]. Practitioners should adapt by segmenting audiences and leveraging analytics to track and optimize message delivery across different platforms.

Acknowledgments

The research on which this paper is based was partially supported by a grant from the US National Science Foundation (#2022216).

Data Availability

The dataset analyzed in this study is publicly available in the Interuniversity Consortium for Political and Social Research repository [76].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords for protective behaviors.
[DOCX File , 19 KB - jmir_v27i1e64307_app1.docx]

Multimedia Appendix 2
Topic modeling process for filtering irrelevant news articles and tweets.
[DOCX File , 21 KB - jmir_v27i1e64307_app2.docx]

Multimedia Appendix 3
Key survey constructs used for measuring media exposure, perceived benefit, recourses, and behavioral intention.
[DOCX File , 21 KB - jmir_v27i1e64307_app3.docx]

Multimedia Appendix 4
How media exposure was calculated.
[DOCX File , 25 KB - jmir_v27i1e64307_app4.docx]

Multimedia Appendix 5
Updated table for the regression.
[DOCX File , 29 KB - jmir_v27i1e64307_app5.docx]

References

1. Brossard D, Scheufele DA. The chronic growing pains of communicating science online. Science 2022;375(6581):613-614. [doi: 10.1126/science.abo0668] [Medline: 35143290]
2. Kasperson JX, Kasperson RE, Pidgeon N, Slovic P. The social amplification of risk: assessing 15 years of research and theory. In: Social Amplification of Risk. Cambridge, England: Cambridge University Press; 2012.
3. Wirz CD, Mayorga M, Johnson BB. A longitudinal analysis of Americans' media sources, risk perceptions, and judged need for action during the zika outbreak. Health Commun 2021;36(12):1571-1580. [doi: 10.1080/10410236.2020.1773707] [Medline: 32496934]
4. Neuwirth K, Dunwoody S, Griffin RJ. Protection motivation and risk communication. Risk Anal 2000;20(5):721-734. [doi: 10.1111/0272-4332.205065] [Medline: 11110218]
5. Noar SM, Austin L. (Mis)communicating about COVID-19: insights from health and crisis communication. Health Commun 2020;35(14):1735-1739. [doi: 10.1080/10410236.2020.1838093] [Medline: 33112180]
6. Covello VT. Best practices in public health risk and crisis communication. J Health Commun 2003;8 Suppl 1:5-8; discussion 148. [doi: 10.1080/713851971] [Medline: 14692565]
7. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]

XSL•FO
RenderX

8. Savoia E, Harriman NW, Piltch-Loeb R, Bonetti M, Toffolutti V, Testa MA. Exploring the association between misinformation endorsement, opinions on the government response, risk Perception, and COVID-19 vaccine hesitancy in the US, Canada, and Italy. Vaccines (Basel) 2022;10(5):671 [FREE Full text] [doi: 10.3390/vaccines10050671] [Medline: 35632427]
9. White SJ, Barello S, di San Marco EC, Colombo C, Eeckman E, Gilligan C, et al. Critical observations on and suggested ways forward for healthcare communication during COVID-19: pEACH position paper. Patient Educ Couns 2021;104(2):217-222 [FREE Full text] [doi: 10.1016/j.pec.2020.12.025] [Medline: 33419600]
10. Viswanath K, Bekalu M, Dhawan D, Pinnamaneni R, Lang J, McLoud R. Individual and social determinants of COVID-19 vaccine uptake. BMC Public Health 2021;21(1):818 [FREE Full text] [doi: 10.1186/s12889-021-10862-1] [Medline: 33910558]
11. Fahim C, Cooper J, Theivendrampillai S, Pham B, Straus SE. Exploring Canadian perceptions and experiences of stigma during the COVID-19 pandemic. Front Public Health 2023;11:1068268 [FREE Full text] [doi: 10.3389/fpubh.2023.1068268] [Medline: 36960376]
12. How to protect yourself and others. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/covid/prevention/index.html [accessed 2025-01-21]
13. Honein MA, Christie A, Rose DA, Brooks JT, Meaney-Delman D, Cohn A, et al. Summary of guidance for public health strategies to address high levels of community transmission of SARS-CoV-2 and related deaths, December 2020. MMWR Morb Mortal Wkly Rep 2020;69(49):1860-1867 [FREE Full text] [doi: 10.15585/mmwr.mm6949e2] [Medline: 33301434]
14. Lang J, Erickson WW, Jing-Schmidt Z. #MaskOn! #MaskOff! digital polarization of mask-wearing in the United States during COVID-19. PLoS One 2021;16(4):e0250817 [FREE Full text] [doi: 10.1371/journal.pone.0250817] [Medline: 33909669]
15. McCombs M, Valenzuela S. Setting the Agenda: Mass Media and Public Opinion. Hoboken, New Jersey: John Wiley & Sons; 2020.
16. Harder RA, Sevenans J, Van Aelst P. Intermedia agenda setting in the social media age: how traditional players dominate the news agenda in election times. Int J Press/Polit 2017;22(3):194016121770496. [doi: 10.1177/1940161217704969]
17. Vargo CJ, Guo L. Networks, big data, and intermedia agenda setting: an analysis of traditional, partisan, and emerging online U.S. News. Journalism Mass Commun Q 2016;94(4):1031-1055. [doi: 10.1177/1077699016679976]
18. Jang SM, Mckeever BW, Mckeever R, Kim JK. From social media to mainstream news: the information flow of the vaccine-autism controversy in the US, Canada, and the UK. Health Commun 2019;34(1):110-117. [doi: 10.1080/10410236.2017.1384433] [Medline: 29028371]
19. Jones-Jang SM, Hart PS, Feldman L, Moon W. Diversifying or reinforcing science communication? Examining the flow of frame contagion across media platforms. Journalism Mass Commun Q 2019;97(1):98-117. [doi: 10.1177/1077699019874731]
20. Wirz CD, Mayorga M, Johnson BB. Evaluating the effects of news-following, volume and content of news coverage on Americans' risk perceptions during the 2014-2016 Ebola outbreak. J Health Commun 2021;26(5):328-338. [doi: 10.1080/10810730.2021.1927257] [Medline: 34185622]
21. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
22. Liu Y, Liu C. Protective and overprotective behaviors against COVID-19 outbreak: media impact and mediating roles of institutional trust and anxiety. Int J Environ Res Public Health 2023;20(2):1368 [FREE Full text] [doi: 10.3390/ijerph20021368] [Medline: 36674124]
23. Chan MPS, Winneg K, Hawkins L, Farhadloo M, Jamieson KH, Albarracín D. Legacy and social media respectively influence risk perceptions and protective behaviors during emerging health threats: a multi-wave analysis of communications on Zika virus cases. Soc Sci Med 2018;212:50-59 [FREE Full text] [doi: 10.1016/j.socscimed.2018.07.007] [Medline: 30005224]
24. Ajzen I. From actions to intentions: a theory of planned behavior. In: Kuhl J, Beckman J, editors. Action-Control: From Cognition to Behavior. Cham: Springer; 1985.
25. Rogers RW. A protection motivation theory of fear appeals and attitude change1. J Psychol 1975;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
26. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
27. Farsi D. Social media and health care, part I: literature review of social media use by health care providers. J Med Internet Res 2021;23(4):e23205 [FREE Full text] [doi: 10.2196/23205] [Medline: 33664014]
28. de Vere Hunt I, Linos E. Social media for public health: framework for social media-based public health campaigns. J Med Internet Res 2022;24(12):e42179 [FREE Full text] [doi: 10.2196/42179] [Medline: 36515995]
29. Liu PL. COVID-19 information on social media and preventive behaviors: managing the pandemic through personal responsibility. Soc Sci Med 2021;277:113928 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113928] [Medline: 33865093]

XSL•FO
RenderX

30. Al-Dmour H, Masa'deh R, Salman A, Abuhashesh M, Al-Dmour R. Influence of social media platforms on public health protection against the COVID-19 pandemic via the mediating effects of public health awareness and behavioral changes: integrated model. J Med Internet Res 2020;22(8):e19996 [FREE Full text] [doi: 10.2196/19996] [Medline: 32750004]
31. Li M, Liu L, Yang Y, Wang Y, Yang X, Wu H. Psychological impact of health risk communication and social media on college students during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020;22(11):e20656 [FREE Full text] [doi: 10.2196/20656] [Medline: 33108308]
32. Jung S, Jung S. The impact of the COVID-19 infodemic on depression and sleep disorders: focusing on uncertainty reduction strategies and level of interpretation theory. JMIR Form Res 2022;6(1):e32552 [FREE Full text] [doi: 10.2196/32552] [Medline: 34870609]
33. Lindell MK, Perry RW. The protective action decision model: theoretical modifications and additional evidence. Risk Anal 2012;32(4):616-632. [doi: 10.1111/j.1539-6924.2011.01647.x] [Medline: 21689129]
34. Fitriani WR, Sutanto J, Handayani PW, Hidayanto AN. User compliance with the health emergency and disaster management system: systematic literature review. J Med Internet Res 2023;25:e41168 [FREE Full text] [doi: 10.2196/41168] [Medline: 37145840]
35. Wang F, Wei J, Shi X. Compliance with recommended protective actions during an H7N9 emergency: a risk perception perspective. Disasters 2018;42(2):207-232. [doi: 10.1111/disa.12240] [Medline: 28799670]
36. Johnson BB, Mayorga M. Americans' early behavioral responses to COVID-19. Hum Ecol Risk Assess: Int J 2021;27(7):1733-1746. [doi: 10.1080/10807039.2021.1884842]
37. Duan T, Jiang H, Deng X, Zhang Q, Wang F. Government intervention, risk perception, and the adoption of protective action recommendations: evidence from the COVID-19 prevention and control experience of China. Int J Environ Res Public Health 2020;17(10):3387 [FREE Full text] [doi: 10.3390/ijerph17103387] [Medline: 32414013]
38. Guo Y, An S, Comes T. From warning messages to preparedness behavior: the role of risk perception and information interaction in the COVID-19 pandemic. Int J Disaster Risk Reduct 2022;73:102871 [FREE Full text] [doi: 10.1016/j.ijdrr.2022.102871] [Medline: 35261877]
39. Scopelliti M, Pacilli MG, Aquino A. TV news and COVID-19: media influence on healthy behavior in public spaces. Int J Environ Res Public Health 2021;18(4):1879 [FREE Full text] [doi: 10.3390/ijerph18041879] [Medline: 33671977]
40. Zhao S, Wu X. From information exposure to protective behaviors: investigating the underlying mechanism in COVID-19 outbreak using social amplification theory and extended parallel process model. Front Psychol 2021;12:631116 [FREE Full text] [doi: 10.3389/fpsyg.2021.631116] [Medline: 34113280]
41. Yang B, Zhao X. TV, social media, and college students' binge drinking intentions: moderated mediation models. J Health Commun 2018;23(1):61-71. [doi: 10.1080/10810730.2017.1411995] [Medline: 29265924]
42. Cho J, Chun JW, Lee MJ. Impacts of message framing and social distance in health campaign for promoting regular physical exercise. J Health Commun 2018;23(9):824-835. [doi: 10.1080/10810730.2018.1527872] [Medline: 30325699]
43. Top 10 U.S. daily newspapers. CISION. 2019 Jan 04. URL: https://web.archive.org/web/20210417130324/https://www.cision.com/2019/01/top-ten-us-daily-newspapers/ [accessed 2021-04-17]
44. Media bias chart. Ad Fontes Media. 2020. URL: https://www.adfontesmedia.com/ [accessed 2025-01-21]
45. Roberts ME, Stewart BM, Tingley D. stm: R package for structural topic models. J Stat Softw 2019;91(2):1-40. [doi: 10.18637/jss.v091.i02]
46. Kleinnijenhuis J, Schultz F, Utz S, Oegema D. The mediating role of the news in the BP oil spill crisis 2010. Commun Res 2013;42(3):408-428. [doi: 10.1177/0093650213510940]
47. Meraz S. Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks. Journalism Mass Commun Q 2011;88(1):176-194. [doi: 10.1177/107769901108800110]
48. Papana A, Kyrtsou C, Kugiumtzis D, Diks C. Identification of causal relationships in non-stationary time series with an information measure: evidence for simulated and financial data. Empir Econ 2022;64(3):1399-1420. [doi: 10.1007/s00181-022-02275-9]
49. Trapletti A, Hornik K, LeBaron B, Hornik MK. Package 'tseries'. 2022. URL: https://cran.r-project.org/web/packages/tseries/tseries.pdf [accessed 2024-09-23]
50. Hothorn T, Zeileis A, Farebrother RW, Cummins C, Millo G, Mitchell D. Package 'lmtest'. Testing linear regression models. 2015. URL: https://cran.r-project.org/web/packages/lmtest/lmtest.pdf [accessed 2025-01-21]
51. CSSEGISandData/COVID-19. CSSEGISandData. 2023. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2025-01-17]
52. Bates DM, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J. Stat. Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
53. Koh JX, Liew TM. How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 twitter feeds. J Psychiatr Res 2022;145:317-324 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.015] [Medline: 33190839]
54. Lu W, Yuan L, Xu J, Xue F, Zhao B, Webster C. The psychological effects of quarantine during COVID-19 outbreak: sentiment analysis of social media data. medRxiv. Preprint posted online on June 26, 2020 2020. [doi: 10.1101/2020.06.25.20140426]

XSL•FO
RenderX

55. Yu M, Li Z, Yu Z, He J, Zhou J. Communication related health crisis on social media: a case of COVID-19 outbreak. Curr Issues Tourism 2020;24(19):2699-2705. [doi: 10.1080/13683500.2020.1752632]
56. Wirz CD, Shao A, Bao L, Howell EL, Monroe H, Chen K. Media systems and attention cycles: volume and topics of news coverage on COVID-19 in the United States and China. Journalism Mass Commun Q 2021;99(4):1048-1071. [doi: 10.1177/10776990211049455]
57. Gold JAW, DeCuir J, Coyle JP, Duca LM, Adjemian J, Anderson KN, et al. COVID-19 case surveillance: trends in person-level case data completeness, United States, April 5-September 30, 2020. Public Health Rep 2021;136(4):466-474 [FREE Full text] [doi: 10.1177/00333549211006973] [Medline: 33789540]
58. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Processes 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
59. Gong Z, Tang Z, Li J. What strategy is better for promoting COVID-19 vaccination? A comparison between gain-framed, loss-framed, and altruistic messages. Ann Behav Med 2022;56(4):325-331. [doi: 10.1093/abm/kaab070] [Medline: 34398184]
60. Galasso V, Pons V, Profeta P, McKee M, Stuckler D, Becher M, et al. Addressing vaccine hesitancy: experimental evidence from nine high-income countries during the COVID-19 pandemic. BMJ Global Health 2023;8(9):e012658 [FREE Full text] [doi: 10.1136/bmjgh-2023-012658] [Medline: 37739450]
61. Cucciniello M, Pin P, Imre B, Porumbescu G, Melegaro A. Altruism and vaccination intentions: evidence from behavioral experiments. Soc Sci Med 2022 Jan;292:114195 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114195] [Medline: 34602309]
62. Fauci AS. The story behind COVID-19 vaccines. Science 2021;372(6538):109. [doi: 10.1126/science.abi8397] [Medline: 33833099]
63. Dryhurst S, Schneider CR, Kerr J, Freeman ALJ, Recchia G, van der Bles AM, et al. Risk Perceptions of COVID-19 Around the World. United Kingdom: Routledge; 2022:162-174.
64. Krause NM, Freiling I, Beets B, Brossard D. Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19. J Risk Res 2020;23(7-8):1052-1059. [doi: 10.1080/13669877.2020.1756385]
65. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
66. Fung TKF, Griffin RJ, Dunwoody S. Testing links among uncertainty, affect, and attitude toward a health behavior. Sci Commun 2018;40(1):33-62. [doi: 10.1177/1075547017748947]
67. Griffin RJ, Dunwoody S, Zabala F. Public reliance on risk communication channels in the wake of a cryptosporidium outbreak. Risk Anal 1998;18(4):367-375. [doi: 10.1111/j.1539-6924.1998.tb00350.x] [Medline: 9775446]
68. Johnson BB. Factors in intention to get the COVID-19 vaccine change over time: evidence from a two-wave U.S. study. Health Risk Soc 2023;25(3-4):151-179. [doi: 10.1080/13698575.2023.2173727]
69. Chen K, Bao L, Shao A, Ho P, Yang S, Wirz C, et al. How public perceptions of social distancing evolved over a critical time period: communication lessons learnt from the American state of Wisconsin. JCOM 2020;19(05):A11. [doi: 10.22323/2.19050211]
70. Laato S, Islam AKMN, Farooq A, Dhir A. Unusual purchasing behavior during the early stages of the COVID-19 pandemic: the stimulus-organism-response approach. J Retail Consum Serv 2020;57:102224. [doi: 10.1016/j.jretconser.2020.102224]
71. Freiling I, Krause NM, Scheufele DA, Brossard D. Believing and sharing misinformation, fact-checks, and accurate information on social media: the role of anxiety during COVID-19. New Media Soc 2023;25(1):141-162 [FREE Full text] [doi: 10.1177/14614448211011451] [Medline: 36620434]
72. Nelson JL, Lewis SC. The structures that shape news consumption: evidence from the early period of the COVID-19 pandemic. Journalism (Lond) 2022;23(12):2495-2512 [FREE Full text] [doi: 10.1177/14648849221095335] [Medline: 38603182]
73. Comscore research captures evolving consumer mobile behavior. Comscore, Inc. 2020. URL: https://www.comscore.com/Insights/Press-Releases/2020/11/Comscore-Research-Captures-Evolving-Consumer-Mobile-Behavior [accessed 2025-01-17]
74. The internet and the pandemic. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/ [accessed 2025-01-17]
75. Moorman C, McCarthy T. CMOs: adapt your social media strategy for a post-pandemic world. Harvard Business Review. 2021. URL: https://hbr.org/2021/01/cmos-adapt-your-social-media-strategy-for-a-post-pandemic-world [accessed 2021-01-19]
76. Johnson B, Mayorga M. RAPID: media exposure, objective knowledge, risk perceptions, and risk management preferences of Americans regarding the novel coronavirus outbreak. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research [distributor], 2025-02-01. URL: https://www.openicpsr.org/openicpsr/project/215382/version/V1/view [accessed 2025-02-05]

Abbreviations

CDC: Centers for Disease Control and Prevention
NYT: New York Times
PADM: Protective Action Decision Model

XSL•FO
RenderX

WSJ: Wall Street Journal

Edited by A Mavragani; submitted 15.07.24; peer-reviewed by D Valvi, Y Xu, D Micallef; comments to author 20.09.24; revised version received 18.10.24; accepted 09.12.24; published 10.03.25.

Please cite as:
Shao A, Chen K, Johnson B, Miranda S, Xing Q
Ubiquitous News Coverage and Its Varied Effects in Communicating Protective Behaviors to American Adults in Infectious Disease Outbreaks: Time-Series and Longitudinal Panel Study
J Med Internet Res 2025;27:e64307
URL: https://www.jmir.org/2025/1/e64307
doi:10.2196/64307
PMID:

©Anqi Shao, Kaiping Chen, Branden Johnson, Shaila Miranda, Qidi Xing. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Stakeholder Consensus on an Interdisciplinary Terminology to Enable the Development and Uptake of Medication Adherence Technologies Across Health Systems: Web-Based Real-Time Delphi Study

Alexandra Lelia Dima[1,2,3], PhD; Urska Nabergoj Makovec[4], PhD; Janette Ribaut[5,6], PhD; Frederik Haupenthal[7], MD; Pilar Barnestein-Fonseca[8], PhD; Catherine Goetzinger[9], PhD; Sean Grant[10], PhD; Cristina Jácome[11], PhD; Dins Smits[12], PhD; Ivana Tadic[13], PhD; Job van Boven[14], PhD; Ioanna Tsiligianni[15], PhD; Maria Teresa Herdeiro[16*], PhD; Fátima Roque[17*], PhD; European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE)[18,19*]

[1]Avedis Donabedian Research Institute (FAD), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

[2]Avaluació de tecnologies sanitàries en atenció primària i salut mental (PRISMA), Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain

[3]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública CIBERESP, Madrid, Spain

[4]Department of Social Pharmacy, Faculty of Pharmacy, University of Ljubljana, Ljubljana, Slovenia

[5]Institute of Nursing Science, Department Public Health, Faculty of Medicine, University of Basel, Basel, Switzerland

[6]Department of Theragnostic, Hematology, University Hospital of Basel, Basel, Switzerland

[7]Division of Nephrology and Dialysis, Department of Medicine III, Medical University of Vienna, Vienna, Austria

[8]CUDECA Institute for Training and Research in Palliative Care, CUDECA Hospice Foundation, Málaga, Spain

[9]Observatoire national de la santé, Strassen, Luxembourg

[10]Prevention Science Institute, University of Oregon, Eugene, OR, United States

[11]CINTESIS@RISE, MEDCIDS, Faculty of Medicine, University of Porto, Porto, Portugal

[12]Faculty of Health and Sports Sciences, Institute of Public Health, Riga Stradins University, Riga, Latvia

[13]Clinical Pharmacy, Innsbruck University, Innsbruck, Austria

[14]Department of Clinical Pharmacy and Pharmacology, Medication Adherence Expertise Center of the Northern Netherlands (MAECON), University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[15]Department of Social Medicine, School of Medicine, University of Crete, Crete, Greece

[16]Institute of Biomedicine, Medical Sciences Department, University of Aveiro, Aveiro, Portugal

[17]Research Unit for Inland Development, Polytechnic of Guarda, Guarda, Portugal

[18]University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[19]See Acknowledgements

[*]these authors contributed equally

Corresponding Author:
Alexandra Lelia Dima, PhD
Avedis Donabedian Research Institute (FAD)
Universitat Autònoma de Barcelona (UAB)
Provença, 293
Barcelona, 08037
Spain
Phone: 34 683163092
Email: adima@fadq.org

Abstract

Background: Technology-mediated medication adherence interventions have proven useful, yet implementation in clinical practice is low. The European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE) European Cooperation in Science and Technology Action (CA19132) online repository of medication adherence technologies (MATechs) aims to provide an open access, searchable knowledge management platform to facilitate innovation and support medication

XSL•FO
RenderX

adherence management across health systems. To provide a solid foundation for optimal use and collaboration, the repository requires a shared interdisciplinary terminology.

Objective: We consulted stakeholders on their views and level of agreement with the terminology proposed to inform the ENABLE repository structure.

Methods: A real-time web-based Delphi study was conducted with stakeholders from 39 countries active in research, clinical practice, patient representation, policy making, and technology development. Participants rated terms and definitions of MATech and of 21 attribute clusters on product and provider information, medication adherence descriptors, and evaluation and implementation. Relevance, clarity, and completeness criteria were rated on 9-point scales, and free-text comments were provided interactively. Participants could reconsider their ratings based on real-time aggregated feedback and revisit the survey throughout the study period. We quantified agreement and process indicators for the complete sample and per stakeholder group and performed content analysis on comments. Consensus was considered reached for ratings with a disagreement index of <1. Median ratings guided decisions on whether attributes were considered mandatory, optional, or not relevant. We used the results to improve the terminology and repository structure.

Results: Of 250 stakeholders invited, 117 (46.8%) rated the MATech definition, of whom 83 (70.9%) rated all attributes. Consensus was reached for all items. The definition was considered appropriate and clear (median ratings 7.02, IPR 6.10-7.69, and 7.26, IPR 6.73-7.90, respectively). Most attributes were considered relevant, mandatory, and sufficiently clear to remain unchanged except for *ISO certification* (considered optional; median relevance rating 6.34, IPR 5.50-7.24) and *medication adherence phase*, *medication adherence measurement*, and *medication adherence intervention* (candidates for optional changes; median clarity ratings 6.07, IPR 4.86-7.17; 6.37, IPR 4.80-6.67; and 5.67, IPR 4.66-6.61, respectively). Subgroup analyses found several attribute clusters considered moderately clear by some stakeholder groups. Results were consistent across stakeholder groups and time, yet response variation was found within some stakeholder groups for selected clusters, suggesting targets for further discussion. Comments highlighted issues for further debate and provided suggestions informing modifications to improve comprehensiveness, relevance, and clarity.

Conclusions: By reaching agreement on a comprehensive MATech terminology developed following state-of-the-art methodology, this study represents a key step in the ENABLE initiative to develop an information architecture capable of structuring and facilitating the development and implementation of MATech across Europe. The debates and challenges highlighted in stakeholders' comments outline a potential road map for further development of the terminology and the ENABLE repository.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-059674

(J Med Internet Res 2025;27:e59738) doi:10.2196/59738

KEYWORDS

health technology; medication adherence; Delphi study; stakeholder engagement; digital health; behavioral science; implementation science

Introduction

Background

Technology-supported interventions for enhancing medication adherence have shown promise [1-3]. Addressing the prevalent low adherence to medication is likely to improve the use of limited resources and population health [4]. However, the sustainable implementation of health technologies in routine practice is lagging, in part due to the lack of accessibility to users in real-life settings [5]. Many repositories have been developed to improve accessibility and, thus, facilitate uptake of digital health [6], including digital apps [7] and medication adherence interventions [8]. Beyond accessibility, despite positive expectations from different stakeholders, digital health uptake is slowed down by misalignment of priorities and lack of a shared understanding of successful implementation, requiring dedicated efforts to increase communication and coordinated action from early development to routine use [9]. For medication adherence technologies (MATechs), additional barriers are the lack of awareness among stakeholders of the high prevalence and negative impact of suboptimal adherence and the lack of coordinated action for large-scale adoption of adherence-enhancing solutions across health systems [10].

The European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE) European Cooperation in Science and Technology Action (CA19132) [11] was a 4-year project (2020-2024) funded by the European Commission bringing together >250 members from 40 countries to (1) raise awareness of medication adherence and related solutions, (2) expand multidisciplinary knowledge on medication adherence at multiple levels, (3) accelerate knowledge translation to clinical settings, and (4) encourage collaboration toward the implementation of MATech across health systems. Within ENABLE, a work package has been dedicated to identifying and showcasing MATech developed or available in participating countries by developing and maintaining a web-based repository where different stakeholders would be able to search for technologies that fit their needs, preferences, and context. The ENABLE repository aims to improve the accessibility of technologies though proactive and continued stakeholder involvement in building a community of practice that would sustain collaborative efforts throughout the process of technology development and implementation.

XSL•FO
RenderX

Managing the knowledge necessary for developing and using appropriate MATech for each health care context is challenging due to the complexity and heterogeneity of the relevant information and of its potential use contexts. This challenge is well known in medical informatics, which has a long history of developing standardized terminologies, classifications, and ontologies for knowledge management in health care [12]. These commonly agreed upon collections of "entities" (objects and phenomena in the world of health care) and their "attributes" (or properties) are used to encode specific types of information and, thus, structure knowledge in ways that allow for automated inference and learning. Such fundaments of information architecture are essential not only for health care delivery but also as a common language for multi-stakeholder communication about health care planning or improvement. Known examples are the World Health Organization (WHO) International Classification of Diseases [13]; International Classification of Functioning, Disability, and Health [14]; and International Classification of Health Interventions [15], as well as the Systematized Nomenclature of Medicine [16]. While well-established terminologies are intended to be comprehensive and evolve with the evolution of health sciences and clinical practice, newer terminologies continue to be proposed focusing on specific knowledge areas and aiming for interoperability with established terms. The recently developed Classification of Digital Health Interventions [17] and the Behaviour Change Intervention Ontology [18] are 2 examples relevant for MATech. The ENABLE work continues these efforts and has already established a definition of MATech and 385 attributes that represent the foundation of the ENABLE repository and MATech-related activities [19] within the broader context of an ENABLE terminology for intervention, technology, and practice established across working groups [20].

Objectives

This paper reports on stakeholders' views on the proposed MATech definition and attributes on which the ENABLE repository is built. We aimed to examine (1) the level of agreement on the relevance, clarity, and completeness of the proposed definition of MATech; (2) the level of agreement on the relevance, clarity, and completeness of the proposed framework of attributes; and (3) the stakeholders' needs and requirements regarding MATech description.

Methods

Study Design and Process

An international interdisciplinary steering committee (SC) was formed by 11 ENABLE members to guide the development of the terminology and repository. The SC prepared the study protocol, proposed relevant attributes for MATech characterization, provided the rationale and methodology for the stakeholder consultation, managed the consultation process, and performed the data analysis and interpretation. The real-time Delphi stakeholder consultation was conducted on the eDelphi platform (Metodix Ltd), which allowed for anonymous data collection and management in line with European regulations on data protection. The real-time Delphi method is appropriate for capturing the views of large groups of experts and encouraging debate to work toward consensus on specific topics. The proposed terminology and the study protocol were improved based on feedback requested from 30 ENABLE members on face validity, user experience, and clarity of the instructions. The description of the ENABLE terminology and its development, the real-time Delphi survey design, and supporting materials are available in the published study protocol [19].

Ethical Considerations

Ethics approval for the ENABLE Action, including this study, was granted by the Málaga Regional Research Ethics Committee (Comite de Etica de la Investigacion Provincial de Malaga) (CEI29.04.2021.MS1) on April 29, 2021, and a data protection assessment was performed by the Data Protection Officer at the University of Basel, which assessed the protocol as compliant with data protection and security standards. Participants were provided with study information and gave informed consent electronically on the study entry page. Anonymity was ensured via the eDelphi platform. No compensation was offered for participation. More details are reported in the published study protocol [19].

Sampling and Sample Size

Purposive sampling was performed to identify at least 195 panelists from 39 countries participating in ENABLE at the time, aiming to represent 5 stakeholder groups in each country: adherence and eHealth research, clinical care, patient representation, policy and decision-making (PDM), and technology development. The sampling aimed to ensure variation in expertise, countries, and other characteristics such as years of expertise and gender. ENABLE Management Committee members in each country were requested to recommend at least 5 panelists among persons in their network whom they considered to have experience in medication adherence in at least one of the stakeholder groups mentioned previously; all recommended persons were invited to take part.

Survey and Supporting Materials

The repository proposed to focus on MATech, defined as "devices, procedures or systems developed based on evidence to support patients to take their medications as agreed with health care providers (i.e., to initiate, implement and persist with the medication regimen)" following established definitions of related terms [21-24]. A framework of attributes was developed to describe MATech, covering 3 domains: *product and provider information* (domain 1), *medication adherence descriptors* (domain 2), and *evaluation and implementation* (domain 3). The survey consisted of several sections. First, information on the aim of the consultation, rules of engagement, and instructions for the survey and informed consent were provided. Second, participants were required to provide information on their sociodemographic characteristics and professional background, including gender, age, country, level of education, field of professional expertise (eg, medicine, pharmacy, nursing, psychology, or data science), stakeholder group (research, practice, PDM, patient representation, or technology development), and years of experience related to the field of medication adherence. Third, participants were asked to rate their level of agreement with and perceived clarity of the

XSL•FO
RenderX

MATech definition and provide comments or revision suggestions as free text. Fourth, they were asked to familiarize themselves with the proposed terminology and provide general feedback on its completeness as free text. Next, attributes were presented to participants to rate their level of agreement with and clarity of each cluster and provide comments and suggestions. To facilitate response, attributes were grouped into 21 thematic clusters following the structure of the attribute framework. The questions addressing relevance and clarity were displayed in a Live2D format, a 2D grid where participants could rate relevance (on the x-axis) and clarity (on the y-axis) simultaneously on 9-point Likert scales, where 1 represented extremely irrelevant/unclear and 9 represented extremely relevant/clear. Participants provided their rating by placing a dot on the grid, which triggered the display in real time of all ratings of previous participants. They were encouraged to reconsider their ratings based on this information and rerate if needed. Participants were also able to view each other's comments and were encouraged to interact in the free-text comment area for each cluster.

Data Collection

Participants' email addresses were entered into the survey management system, which generated automatic email invitations with personalized links. Before starting the survey, every participant provided informed consent electronically. Participants were able to access and answer the survey at any time during the study period between October 25, 2021, and January 15, 2022. To reach the full potential of the real-time approach, participants were encouraged to revisit the survey several times to follow and engage in discussions and revisit their ratings if their opinion changed. Participation was moderated by the study team to encourage interaction and discussions. Summaries of the study progress and active debates and encouragements to participate were provided weekly via email through the platform (Multimedia Appendix 1). The comment areas were monitored for any technical questions, which were answered in the same interactive area accessible to all, and the answers were included in the weekly summaries if appropriate. To monitor and analyze the evolution of survey responses in time, data were downloaded daily from day 7 to the end of the study period (78 days). This was necessary to bypass the platform limitation of overwriting previous data of individual participants if they decided to reconsider their answers. Data collection was stopped when three predetermined criteria were met: (1) a total response rate of ≥30%, (2) ≥10 complete responses in each stakeholder group, and (3) representation of two-thirds of the ENABLE countries. A survey response was considered complete if background questions and at least 75% of the questions on the terminology were answered.

Data Analysis

Quantitative analyses were conducted in R (R Foundation for Statistical Computing) [25], and qualitative analyses were conducted in NVivo Pro (version 12; QSR International) [26]. Descriptive statistics were used for the background characteristics of the sample and to summarize responses on the terminology.

Level of Agreement on Relevance and Clarity

The level of agreement with the definition, completeness, and relevance and clarity of every attribute cluster was quantified using the interpercentile range adjusted for symmetry (IPRAS) analysis technique from the RAND/UCLA Appropriateness Method [27]. The disagreement index (DI) was calculated as a ratio between the interpercentile range (IPR) from the 30th to the 70th percentile and the IPRAS, and DI>1 (ie, IPR>IPRAS) was considered to indicate disagreement [27]. Levels of agreement were determined based on median ratings rounded to the nearest integer, and the DI was used to steer decisions about each cluster (Table 1).

Table 1. Criteria for levels of agreement and decisions about the attribute clusters.

Criterion	Relevance	Clarity
Median 7-9 and DI[a]<1	Attributes are relevant and mandatory.	Attributes are sufficiently clear to remain unchanged.
Median 4-6 or DI>1	Attributes are optional.	Proposed changes to the attributes are optional.
Median 1-3 and DI<1	Attributes are not relevant and candidates for exclusion.	Attributes are not clear and candidates for rephrasing.

[a]DI: disagreement index.

Subgroup Analyses

The descriptive analyses were repeated per stakeholder group to examine variation in opinions among stakeholders applying the criteria described previously and considering each participant in all the groups they identified themselves with. The reliability of the ratings per question within the stakeholder group was determined via intraclass correlation coefficients (ICCs) based on a 2-way random model considering type (average measures) and definition of relationship (consistency), with ICC>0.70 indicating moderate to good reliability [19]. To examine differences per item, 2-level models were performed for each item considering participants nested within stakeholder group. For these models, participants identifying with more than one group were allocated to the group least represented to maximize the equal distribution of participants in the groups.

Analysis of Process Indicators

The coefficient of quartile variation (CQV) was used to describe response stability (ie, the consistency of stakeholders' responses) and was calculated at the end of the study for all participants (CQV_{total}) and within the same stakeholder group (CQV_{sub}), with CQV_{total}<30% and CQV_{sub}<15% indicating a stable (consistent) response [28]. Longitudinal changes in ratings for each question from individual participants, color coded by stakeholder group, and the evolution of median values of ratings for each stakeholder group were explored visually to gain deeper insights into the evolution of responses during the survey period.

XSL•FO
RenderX

Qualitative Analysis of Participants' Interactive Feedback

Comments provided in free text were analyzed qualitatively using deductive content analysis based on the structure of the attribute framework [19]. Coding was performed independently by 2 researchers (UNM [pharmacy background and experienced in qualitative analysis] and FH [medicine background and trained new user of qualitative methods]). Both coders had in-depth knowledge of the terminology and had lead roles in managing the consultation process. Discrepancies were discussed and harmonized through an iterative process, initially between the 2 coders and then in discussions with other SC members (AD, JR, TH, and FR). This process generated a revised thematic tree, which was presented in a narrative and table format. Proposed modifications to the terminology were considered if suggested by median ratings or DIs and reflected in participants' free-text comments. The qualitative results guided improvements in the attribute framework (eg, rewording labels and definitions and excluding or adding attributes) and the subsequent development of the ENABLE repository.

Results

Participation Rates and Sample Characteristics

From 250 invited panelists, 174 (69.6%) started the survey, 117 (46.8%) panelists from 29 countries answered the questions on background information and the definition of MATech, and 83 (33.2%) panelists from 27 countries completed the survey (see details in the data analysis report in Multimedia Appendix 2). On average, 3 (SD 2) panelists represented each of the 29 countries, with a minimum of 1 per country (2 countries) and a maximum of 10 per country (1 country). The countries represented in the real-time Delphi study were Albania, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. The panel consisted predominantly of women (74/117, 63.2%), participants with PhD-level education (64/117, 54.7%), health care professionals (HCPs; 85/117, 72.6%), and participants with background in research or health care practice (74/117, 63.2% and 62/117, 53%, respectively); 55.6% (65/117) of the participants identified with 1 background, whereas others identified with up to 5 backgrounds (13/117, 11.1%; Table 2).

XSL•FO
RenderX

Table 2. Characteristics of panelists who completed medication adherence technology (MATech) definition ratings and panelists who completed the survey.

	Completed MATech definition rating (n=117), n (%)	Completed the survey (n=83), n (%)
Sex		
Female	74 (63.2)	54 (65.1)
Male	43 (36.8)	29 (34.9)
Age (years)		
18-30	4 (3.4)	2 (2.4)
31-40	28 (23.9)	22 (26.5)
41-50	43 (36.8)	28 (33.7)
51-60	35 (29.9)	26 (31.3)
61-70	7 (6)	5 (6)
Educational level		
Bachelor's degree	4 (3.4)	3 (3.6)
Doctorate (PhD)	64 (54.7)	42 (50.6)
High school diploma	4 (3.4)	3 (3.6)
Master's degree	28 (23.9)	23 (27.7)
Specialty degree (health care)	17 (14.5)	12 (14.5)
Expertise		
Computer science or software engineering	5 (4.3)	4 (4.8)
Data science or statistics	5 (4.3)	4 (4.8)
Economics or management	11 (9.4)	8 (9.6)
Medicine	26 (22.2)	16 (19.3)
Nursing	11 (9.4)	11 (13.3)
Pharmacy	48 (41)	33 (39.8)
Psychology	3 (2.6)	2 (2.4)
Sociology	1 (0.9)	0 (0)
Other	7 (6)	5 (6)
Background[a]		
Research or education	74 (63.2)	50 (60.2)
HCP[b]	62 (53)	45 (54.2)
PDM[c]	29 (24.8)	19 (22.9)
Patient representative	26 (22.2)	18 (21.7)
eHealth developer	32 (27.4)	24 (28.9)
Number of backgrounds		
1	65 (55.6)	49 (59)
2	27 (23.1)	16 (19.3)
3	9 (7.8)	7 (8.4)
4	3 (2.6)	1 (1.2)
5	13 (11.1)	10 (12)

[a]Multiple answers possible for this question.

[b]HCP: health care professional.

[c]PDM: policy and decision-making.

XSL•FO
RenderX

Level of Agreement on Relevance and Clarity

No disagreements we identified among the 46.8% (117/250) of the participants who contributed to definition and attribute rating. The median level of agreement with the proposed MATech definition was 7.02 (IPR 6.10-7.69), and the median clarity rating was 7.26 (IPR 6.73-7.90). Thus, the definition was considered relevant and sufficiently clear. The median ratings of relevance and clarity for attribute clusters based on 33.2% (83/250) of the participants are presented visually in Figure 1. A total of 95% (20/21) of attribute clusters were considered relevant (Table S1 in Multimedia Appendix 3), whereas 1 attribute (*ISO standard*) received a moderately relevant rating (median 6.34, IPR 5.50-7.24). *Target use scenario* received the highest rating (median 7.66, IPR 6.99-8.01). Thus, most attributes were considered relevant and mandatory for inclusion in the repository except for *ISO standard*, which was considered optional. In total, 86% (18/21) of the attribute clusters were considered clear (Table S2 in Multimedia Appendix 3). A total of 14% (3/21) of the attribute clusters were rated as moderately clear: *medication adherence phase* (median 6.07, IPR 4.86-7.17), *medication adherence measurement* (median 6.37, IPR 4.80-6.67), and *medication adherence intervention* (median 5.67, IPR 4.66-6.61). *Implementation outcomes* received the highest clarity rating (median 7.67, IPR 7.20-8.06). Thus, most attributes were considered sufficiently clear to remain unchanged, whereas for 3 attributes, changes could be considered if appropriate to improve clarity.

Figure 1. Median ratings with interpercentile range for relevance and clarity per attribute cluster (N=83). ISO: International Organization for Standardization.

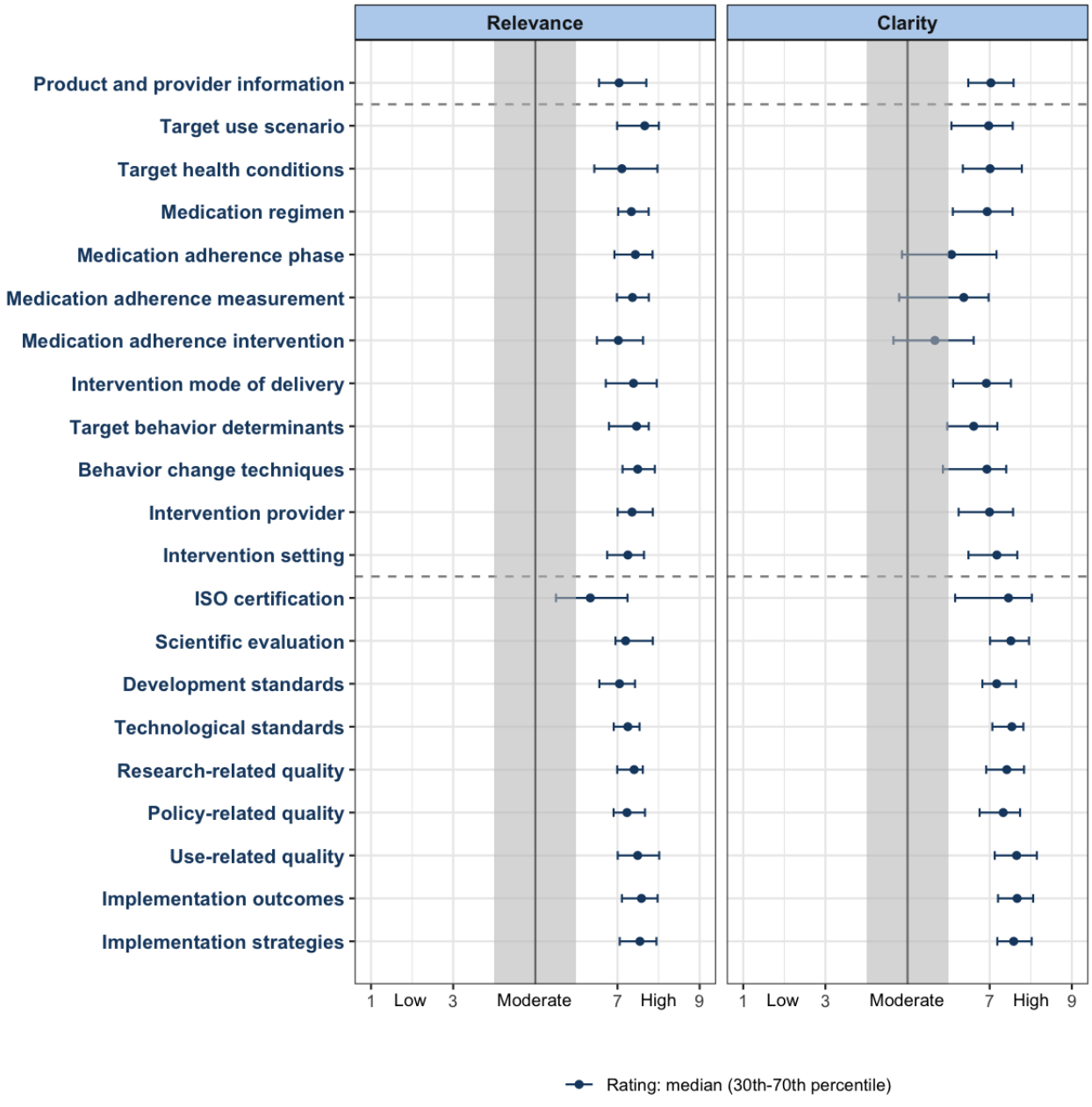

XSL•FO
RenderX

Subgroup Analyses

No disagreements were identified at the subgroup level among researchers, HCPs, PDM stakeholders, patient representatives, and technology developers. For the MATech definition, levels of agreement and clarity ratings within the subgroups were above the established thresholds for sufficient relevance and clarity (Table 3). For attribute clusters (Figure S1 in Multimedia Appendix 3), relevance ratings within all subgroups were above the established thresholds for being considered relevant and sufficiently clear except for *ISO certification* (considered optional by all subgroups) and *intervention setting* (considered optional only by technology developers). Clarity ratings within subgroups confirmed the full sample results for *medication adherence phase* and *medication adherence intervention*, which were perceived as moderately clear by all subgroups. *Medication adherence measurement* ratings were above the clarity threshold for subgroups, suggesting that participants identifying with multiple subgroups rated this attribute higher on clarity. Moreover, *target behavior determinants* was considered moderately clear by patient representatives, and *intervention provider* was considered moderately clear by PDM stakeholders, patient representatives, and technology developers.

Across attribute clusters, the ICC value for the median ratings of the 5 subgroups was 0.95 for relevance and 0.97 for clarity, indicating good interrater reliability between subgroups (considered as "raters"). To describe the differences specific to each attribute cluster, 41% (34/83) of the participants, who identified with more than one group, were allocated to the least represented group. This decision resulted in 33% (27/83) of the participants in the research group, 22% (18/83) in the HCP group, 10% (8/83) in the PDM group, 22% (18/83) in the patient representative group, and 14% (12/83) in the technology developer group. For most attribute clusters, the proportion of variance explained by the stakeholder group was either 0 (model almost or near singular) or <0.1, with only 2 items with an ICC of >0.15 (ie, *target health conditions* [relevance; ICC=0.19] and *intervention setting* [relevance; ICC=0.205]). Thus, ≥80% of the variance was within groups, and therefore, the differences across groups did not justify further investigation per item (Multimedia Appendix 4).

Table 3. Descriptive statistics for level of agreement with and clarity of medication adherence technology definition per stakeholder subgroup.

Subgroup and criterion	Median	30th-70th percentile	Interpercentile range	Interpercentile range adjusted for symmetry	Disagreement index
Researchers (n=50)					
Agreement	7.05	6.10-7.77	1.67	5.25	0.32
Clarity	7.48	6.91-8.01	1.10	6.04	0.18
Health care professionals (n=45)					
Agreement	7.01	6.11-7.45	1.34	5.02	0.27
Clarity	7.38	6.51-7.86	1.35	5.63	0.24
Policy and decision-making(n=19)					
Agreement	6.86	6.19-7.40	1.22	5.05	0.24
Clarity	7.55	7.14-7.97	0.84	6.19	0.14
Patient representatives (n=18)					
Agreement	6.74	5.99-7.20	1.20	4.74	0.25
Clarity	7.09	6.63-7.74	1.12	5.63	0.20
Technology developers (n=24)					
Agreement	6.74	6.01-7.41	1.40	4.92	0.28
Clarity	7.60	7.01-8.01	1.00	6.11	0.16

Process Indicators

For the whole sample, CQV values for the MATech definition were 13.23 for agreement and 12.74 for clarity, indicating consistency (<30%); consistent responses were found for all attribute clusters. Subgroup analyses found CQV values below the established threshold for most subgroups (<15%), with some exceptions. For the MATech definition, response variation was found among HCPs and technology developers. For relevance (Table 4), response variation was found among technology developers regarding *medication adherence intervention* and among researchers, HCPs, PDM stakeholders, and technology developers regarding the *ISO certification.* For clarity (Table 5), response variation was found among one or more stakeholder groups regarding 48% (10/21) of the attribute clusters. Notably, *target health conditions* was below the stability threshold for patient representatives and technology developers; *medication adherence phase* was below the stability threshold for researchers, HCPs, patient representatives, and technology developers; *medication adherence measurement* was below the stability threshold for researchers, PDM stakeholders, and patient representatives; *medication adherence intervention* was below the stability threshold for researchers and patient representatives; *target behavior determinants* was below the stability threshold for technology developers; *intervention setting* was below the stability threshold for researchers, HCPs,

XSL•FO
RenderX

patient representatives, and technology developers; and *ISO certification* was below the stability threshold for researchers and patient representatives. Moreover, *target use scenario*, *medication regimen*, *intervention mode of delivery*, and *behavior change techniques* were below the stability threshold for patient representatives only.

Most participants did not change their ratings throughout the real-time Delphi survey (Multimedia Appendix 4). The visualization of the evolution of median ratings for relevance and clarity for each attribute cluster per stakeholder group (Figure 2) shows a stabilization of the median values in the second half of the real-time Delphi period at levels above the threshold values.

Table 4. Coefficient of quartile variation (CQV) overall and per subgroup for relevance (N=83). Values below the relevance threshold are italicized.

Rating item	CQV total	CQV for researchers	CQV for health care professionals	CQV for policy and decision-making	CQV for patient representatives	CQV for technology developers
Medication adherence technology definition (agreement)	13.23	14.15	16.54	6.05	12.05	17.62
Product and provider information	11.14	10.09	6.99	6.54	13.75	9.77
Target use scenario	7.88	8.69	6.66	6.72	8.63	10.62
Target health conditions	11.35	12.62	8.42	4.29	18.47	13.03
Medication regimen	7.38	6.32	6.57	3.07	11.64	11.77
MA[a] phase	8.47	8.07	6.68	5.44	10.64	8.16
MA measurement	6.73	4.98	4.61	10.37	6.41	8.28
MA intervention	11.23	9.56	5.01	9.56	9.19	*19.95* [b]
Intervention mode of delivery	10.08	6.47	7.34	8.17	10.65	8.96
Target behavior determinants	7.08	6.33	2.88	8.05	7.06	3.42
Behavior change technique	6.71	6.04	6.72	5.29	10.05	11.32
Intervention provider	8.38	4.84	4.92	3.71	6.94	8.74
Intervention setting	10.55	8.97	7.42	7.73	8.84	7.88
ISO certification	18.84	*19.62*	*18.55*	*19.64*	11.64	*31.10*
Scientific evaluation	7.52	9.32	7.55	5.44	12.28	7.53
Development standards	10.99	8.89	6.67	3.03	11.23	9.74
Technological standards	6.17	6.37	4.06	3.61	5.24	6.47
Research-related quality	5.87	6.02	5.43	3.35	8.45	5.01
Policy-related quality	7.00	6.08	5.10	6.89	9.26	5.93
Use-related quality	7.61	7.25	6.86	4.44	9.19	9.40
Implementation outcomes	6.41	6.46	5.31	4.66	6.28	9.54
Implementation strategies	6.78	7.76	4.96	7.37	6.19	6.05

[a]MA: medication adherence.

[b]Values below the relevance threshold are italicized.

XSL•FO
RenderX

Table 5. Coefficient of quartile variation (CQV) overall and per subgroup for clarity (N=83). Values below the clarity threshold are italicized.

Rating item	CQV total	CQV for researchers	CQV for health care professionals	CQV for policy and decision-making	CQV for patient representatives	CQV for technology developers
Medication adherence technology definition (agreement)	12.74	12.01	13.76	6.58	10.43	14.50
Product and provider information	13.65	12.53	8.68	4.06	10.03	11.43
Target use scenario	13.1	13.18	13.83	8.69	*15.62* [a]	10.61
Target health conditions	13.14	8.74	7.69	13.84	*16.37*	*21.60*
Medication regimen	12.77	12.75	9.02	7.95	*18.72*	14.91
MA[b] phase	25.03	*21.49*	*23.47*	6.84	*28.13*	*37.23*
MA measurement	21.26	*20.51*	12.31	*18.74*	*29.43*	14.92
MA intervention	13.12	*19.95*	9.1	9.39	*27.3*	13.34
Intervention mode of delivery	12.61	7.9	11.27	11.65	*19.54*	14.21
Target behavior determinants	14.24	14.34	6.15	11.75	14.64	*20.96*
Behavior change technique	11.97	12.49	9.32	13.19	*19.29*	8.38
Intervention provider	10.75	8.42	7.23	6.57	14.55	14.64
Intervention setting	23.72	*22.19*	*20.55*	14.68	*18.61*	*26.34*
ISO certification	14.91	*18.71*	12.4	6.45	*17.57*	9.73
Scientific evaluation	7.41	6.64	7.6	12.2	12.4	14.3
Development standards	7.77	9.27	8.33	8.17	4.19	8.41
Technological standards	6.33	6.7	3.25	2.58	6.69	12.32
Research-related quality	10.02	8.59	4.35	5.02	14.35	7.54
Policy-related quality	8.28	8.64	4.17	8.59	9.42	8.4
Use-related quality	7.9	8.02	4.22	4.84	10.33	8.04
Implementation outcomes	6.97	6.54	7.66	3.87	7.45	5.35
Implementation strategies	6.87	7.28	6.2	6.57	3.83	7.3

[a]Values below the clarity threshold are italicized.

[b]MA: medication adherence.

XSL•FO
RenderX

Figure 2. Evolution of median ratings of relevance and clarity per stakeholder group. Each line represents the median value of the ratings for 1 attribute cluster and property (relevance or clarity) calculated on each day of the survey based on the responses of the participants to date. HCP: health care professional; PDM: policy and decision-making.

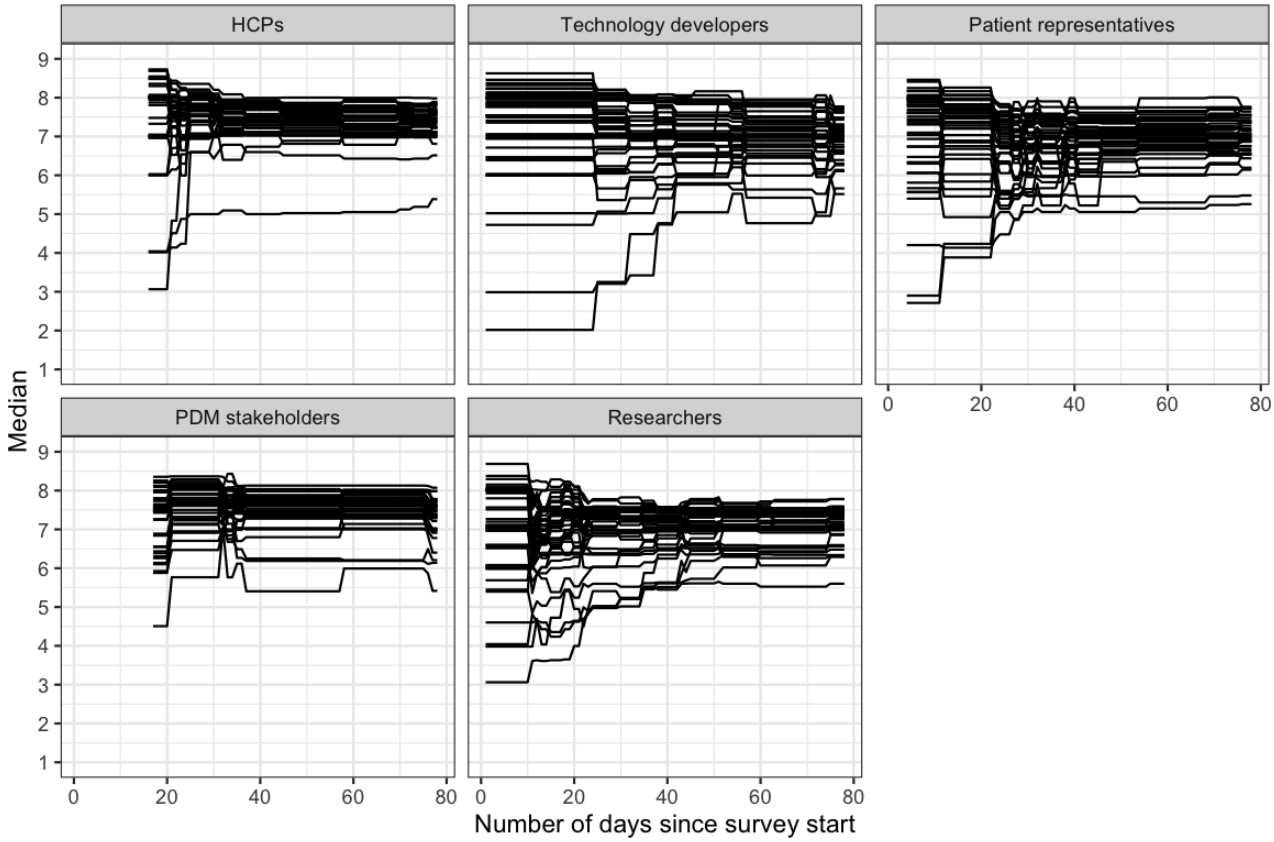

Qualitative Analysis of Participants' Interactive Feedback

Overview

We examined the feedback provided in the interactive comment area of each survey page, pertaining to 6 predefined topics: general feedback on the survey and on the terminology, specific feedback on the MATech definition, and feedback on the 3 attribute domains (*product and provider information*, *medication adherence descriptors*, and *evaluation and implementation*). In addition to identifying suggestions for improving the terminology, we aimed to understand stakeholders' needs and expectations regarding MATech description that would guide further steps in the development of the ENABLE repository and community of practice. Further details and representative quotes are presented in Multimedia Appendix 5.

General Feedback on the Terminology

Several participants noted the complexity and comprehensiveness of the terminology. Some highlighted challenges related to its use for the repository (eg, anticipating how the framework would need to be translated in a data collection tool and expressing concerns on the feasibility of data collection and long-term maintenance of the repository). Some comments pointed to the importance of representing patient perspectives in the terminology (eg, by including patient-related determinants of adherence and technology use such as medication burden, affordability, or inequities of access).

Feedback on the MATech Definition

Many participants commented on the MATech definition, with several aspects generating debate. Referring to the term "technology," some participants considered that MATech should be understood as digital tools, whereas others considered that a broader scope (eg, including also analog tools and services) was appropriate. Different opinions were expressed on whether "medication adherence" stood for taking medicines as *prescribed by* or as *agreed with* an HCP. Debate ensued on the meaning of the term "evidence based," especially concerning the strength of the evidence needed and finding a balance between inclusiveness and evidence standards (ie, between including technologies under development for which limited evidence is available vs limiting the repository to technologies that meet higher thresholds of evidence).

Feedback on Product and Provider Information

In this first domain, 2 general questions were raised. Some participants were concerned about how the product maturity (versioning) would be handled in the repository and how products that comprised a combination of technologies could be described. The implementation of these terms in the repository would need to consider these different scenarios.

Some comments referred to the need to consider the intended use of the product as a first question and adapt the description depending on the response to this question.

XSL•FO
RenderX

Feedback on Medication Adherence Descriptors

In this second domain, there was debate over the distinction between monitoring and support as 2 elements of adherence management. Some comments recommended to follow the Ascertaining Barriers to Compliance medication adherence taxonomy [21] without adaptations. Others questioned whether an MATech could only monitor the patient without also intervening and affecting their behavior. Inclusion of behavior change techniques following the capability, opportunity, motivation, behavior model was viewed positively by some participants, whereas others questioned whether these models are widely known and simple enough for their use in practice, for example, by HCPs. Suggestions were given on clarifying how certain types of MATech could be described according to the proposed attributes (eg, modes of delivery and provider and setting).

Feedback on Evaluation and Implementation

The third domain was the least commented on. The need for an evidence base, especially scientific evidence, was challenged and considered conditional on the intended use of the technology. Participants expressed that the implementation outcomes and strategies should be more visible in the terminology. Some overlap was noted in the proposed attributes. Several comments referred to the lesser relevance of the ISO certification.

Stakeholders' Needs and Expectations

Some comments highlighted contrasting needs and expectations from participants regarding the MATech description. Some expected more detailed descriptions to provide comprehensive information for different potential users, whereas others preferred more succinct descriptions feasible for MATech providers to enter into the platform. Several comments expressed expectations for continued testing and adaptation of the terminology to evolving technologies. Ensuring accurate information, continuously evaluating MATech, and updating MATech descriptions regularly (eg, every year) were highlighted as important expected features for the repository to provide value to potential users in the context of a rapidly changing field.

Discussion

Principal Findings

The ENABLE European Cooperation in Science and Technology Action offered the opportunity to develop a comprehensive interdisciplinary terminology regarding MATech aligned with existing classifications and ontologies and to consult stakeholders from diverse backgrounds and geographical locations on the terminology's relevance and clarity. We found agreement among stakeholders on the MATech definition and attribute clusters, which indicates an interdisciplinary network ready for collaboration on this conceptual basis. Most terms were considered relevant and clear, and several moderate ratings and related comments suggested further improvements, which informed a new version of the terminology. This initial consensus represents a strong foundation for current efforts to advance the development and use of technologies in medication adherence management across health systems, in which the ENABLE network and collaborating organizations have already made important progress [10,11,20]. This conceptual foundation was an important step in the development of the ENABLE repository (Figure 3 [29]). However, the variation in opinions within stakeholder groups and debates that ensued in the interactive feedback areas over several topics point to the need to continue engaging stakeholders in constructive and inclusive discussions to align conceptual representations and expectations on how they can guide concrete action.

XSL•FO
RenderX

Figure 3. Screenshot of the European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE) repository as available at manuscript submission date.

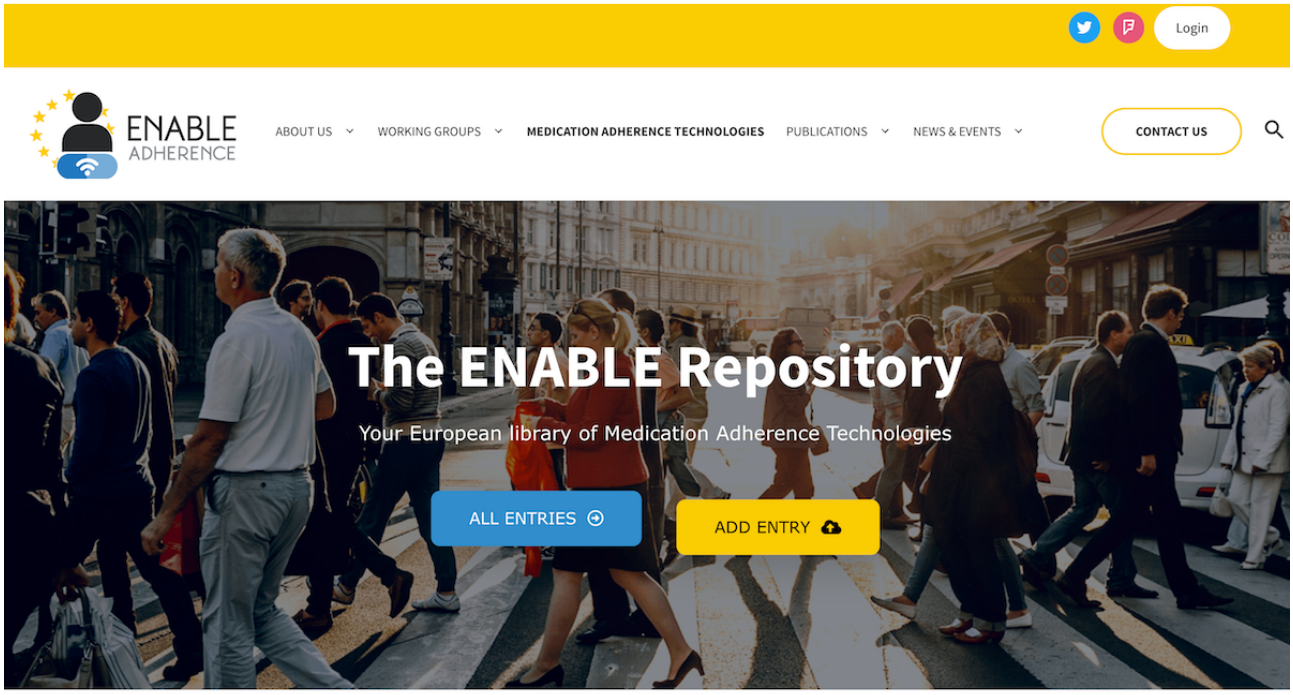

Interpretation and Comparison to Prior Work

It was agreed that the MATech definition was appropriate and clear; however, several debates ensued regarding "technology," "medication adherence," and "evidence." These debates have practical implications for the ENABLE repository and network. First, the term "health technology," which we adopted from the WHO definition as "the application of organized knowledge and skills in the form of devices, medicines, vaccines, procedures and systems developed to solve a health problem and improve quality of lives" [23], was considered by some participants as too broad for the scope of the repository. At the same time, the attribute framework was criticized for not explicitly including types of technologies considered of high potential, such as bio- and nanotechnologies. Considering this feedback, it was agreed that the scope of the ENABLE repository would focus initially on digital technologies, with the possibility of further extension to other types of technologies in the future. Nevertheless, these different perspectives highlight the challenges inherent in structuring and describing health innovation. The discussion that ensued on the term "medication adherence" as taking medicines "as prescribed" or "as agreed" reflects a long-standing challenge of harmonizing descriptive versus prescriptive definitions of medication adherence. From a descriptive viewpoint, the process of medication taking starts once a prescription has been issued; is operationalized as the comparison between 2 time series (ie, of prescribed and actual dosing); and is distinct from its determinants, including the process of shared decision-making leading to an agreement on the treatment regimen between the patient and their health care team. These distinctions are essential for accurate measurement in research and evaluation [30]. From a prescriptive viewpoint, adhering to a medication regimen would require as a precondition an appropriately prescribed treatment resulting from a process of decision-making that involves mutual agreement on a course of treatment for which responsibility is

XSL•FO
RenderX

shared between the patient and health care team [24]. Thus, integrating treatment agreement into the broader process of medical treatment is essential for a person-centered approach to medication taking. Harmonizing these perspectives requires further consensus work with the involvement of different stakeholders. Third, the role of scientific evidence was debated, with some participants underlining the necessity of including only MATech supported by scientific evidence in the repository and others promoting a more inclusive approach to MATech selection. Following this feedback, we adopted the technology readiness level (TRL) as a framework to guide eligibility, including MATech with at least a small-scale prototype validated in a local environment; however, challenges remained regarding mapping candidate MATech onto the TRL framework [31]. On the basis of these considerations and further work within the ENABLE network on a higher-level ENABLE terminology including medication adherence–enhancing interventions, reimbursement, and best practice, the MATech definition has evolved to "evidence-based health technologies used in the management of medication adherence by different stakeholders" [20]. Therefore, the initial version of the repository includes MATech that (1) are presented as aiming to address medication adherence management (by either obtaining information or offering support), (2) have a digital component as per the DigitalHealthEurope definition [32], and (3) meet at least TRL 4. These criteria are likely to evolve depending on the interest in and uses of the repository by different stakeholder groups.

Most attribute clusters were also considered relevant and clear except for *ISO qualification*, which was considered moderately relevant, and the *medication adherence phase*, *medication adherence measurement*, and *medication adherence intervention* attributes, which were considered moderately clear. The comments received guided improvements; for example, the *ISO qualification* attribute was changed to a broader and more flexible attribute of *certification label*, and the definitions of the medication adherence–related attributes were modified to improve clarity and refer to the Ascertaining Barriers to Compliance taxonomy [21], which informed their selection. The comments on each attribute domain were taken into account by modifying labels and definitions of some terms and restructuring others or in the formulation of the MATech data collection form. While these modifications were considered sufficient for the initial development of the repository, it is necessary to continue improving this terminology and allow it to evolve in synchrony with similar efforts in related fields. In particular, the methodology and evaluation criteria for health technology are evolving rapidly. For example, a recent Delphi study on the assessment of patient-facing eHealth tools reached consensus on 46 criteria classified as "foundational" (eg, technical aspects, clinical utility, and safety) and "contextual" (eg, data protection compliance, safety regulatory compliance, interoperability, and data integration) [33]. While this extensive assessment specific to different use scenarios can be applicable to MATech as well, a briefer description specific to medication adherence would be more appropriate for an initial overview of an MATech at a given moment, which could be followed up with more extensive examination. To facilitate the repository development and its function of boosting research collaborations within the ENABLE network, we subsequently chose to include only a selection of attributes in the MATech description and develop support materials for participating countries to search, contact, and involve MATech developers in describing their products on the ENABLE website. To date, 18 countries have started the MATech search process, and 8 have included at least one MATech in the repository. User feedback will continue to be gathered to inform the iterative development of the terminology and repository.

The subgroup analyses and process indicators showed limited variation in responses between and within groups and a stabilization of responses halfway through the survey. This stability and convergence toward above-threshold ratings for relevance and clarity indicate a strong basis for multi-stakeholder action on this topic. However, there were some attribute clusters that showed moderate clarity only for some stakeholder groups (ie, *target behavior determinants* for patient representatives and *intervention provider* for PDM stakeholders, patient representatives, and technology developers). While these ratings suggest only optional modifications, it is important to interpret these findings as a recommendation to reword and explain professional terminology in all stakeholder-facing materials. Participants recommended repeatedly adapting the technical language to diverse audiences, and this advice has been considered in the development of repository-related materials and needs to guide further developments.

Developing a consensus terminology to structure the description and facilitate the search of available solutions has been used in developing repositories in related areas, such as the taxonomy of self-management interventions on which the Comparing Effectiveness of Self-Management Interventions in 4 High-Priority Chronic Diseases in Europe (COMPAR-EU) platform was built [34]. These frameworks prove useful in systematic searches and descriptions of published research evidence and available digital products, as performed in the COMPAR-EU project [35] and in ENABLE [36]. Other repositories may not rely on an extensive conceptual development and rather adopt a limited number of descriptive terms to serve for initial explorations of the available offer in a specific topic or geographical area. This is the case the Radar Digital Health Uptake [6], the different country-specific repositories cataloged on the European mHealth Hub [7], and Interventienet for medication adherence interventions [8]. As the digital health field will continue to develop, it becomes increasingly relevant to work on the integration and continuity of these platforms. Ontological approaches may provide part of the solution for the technical integration, whereas consensus methodologies such as the Delphi method could be used to reach agreement on policy priorities and their implementation.

Strengths and Limitations

This real-time Delphi study succeeded in engaging a diverse stakeholder panel regarding geographical location, professional experience, and sociodemographic characteristics and in reaching a consensus on a comprehensive terminology to inform research on and adoption of MATech. The real-time web-based format facilitated interactions and flexible participation of

XSL•FO
RenderX

panelists at their chosen time and provided valuable feedback for further improvements. The subgroup and process analyses allowed for a more in-depth understanding of the variation in opinions and, thus, potential avenues for future stakeholder engagement. However, 2 main limitations are important to consider. First, some limitations of the tool did not permit a full analysis of the change in opinions and might have overestimated the level of consensus. For example, once a panelist provided an initial response to a question, they were able to see the responses of others and readjust their response in light of this information, in which case the initial response was overwritten. Nevertheless, survey data were downloaded every day to capture temporal changes. Therefore, while we could not capture changes between initial and reconsidered responses if such changes occurred within the same day, changes occurring in consecutive days were identified. Second, participation rates were <50%, and few participants engaged with the survey multiple times during the survey period. While we attempted to encourage participation via pretesting of materials and weekly summaries emailed to all participants, we could not overcome other likely barriers such as the complexity of the participation task and the novelty of the real-time web-based Delphi format. Thus, our findings likely represent a selected population, and future work on the ENABLE terminology would need to involve a wider audience and apply different methods to communicate and ask for feedback about medication adherence–related terms and their different practical applications.

Future Directions

This initial work opens several avenues for potential development of the ENABLE terminology. An important next step is to develop it into an ontology of MATech research and deployment in health systems in line with methodological recommendations [37]. While our terminology closely followed ontology development principles and steps [19], further work remains to be done, such as testing interrater reliability, specifying relationships between entities, and disseminating and maintaining the future ontology. This work would need to be integrated into the sustainability of the ENABLE repository and community of practice. The ENABLE Action has been encouraging the development of national centers and networks of medication adherence in collaboration with stakeholder organizations [10]. These centers and the pan-European community that has developed during the ENABLE project could adopt and further improve this terminology and conduct translations into different languages following recent examples of other terminologies [38,39]. The repository is intended as a common platform for developing and implementing digital adherence technologies in Europe and needs to be supported by policies that acknowledge medication adherence as a priority for coordinated action, inform of available solutions, and incentivize and support their deployment [40].

Beyond its use to inform the repository, the ENABLE terminology can be applied to assess stakeholder needs and develop training, for example, training HCPs in selecting and using MATech in their practice or MATech developers in assessing user needs and health system requirements for technology adoption. Part of the attribute framework was already used in developing content for an interdisciplinary ENABLE training school in 2023, and other materials are under development. In the long term, the more consistent use of MATech-related terms and definitions has the potential to lead to better standardization of research and implementation efforts, as well as of recording practices in real-life settings, by providing the information architecture to collect real-world data for the development, evaluation, and implementation of MATech in line with current developments to use real-world data for agile regulatory approval [41]. As medication adherence is a transversal topic in health services worldwide, we can envisage the possibility of using future iterations of this taxonomy for standardized operationalization of medication adherence measurement and intervention into digital systems, for example, within the innovative Digital Adaptation Kits recently piloted by the WHO in several clinical areas [42]. Given its built-in interoperability, the ENABLE terminology can also be expanded to encompass terms describing other related medication use processes (eg, medication review and prescribing practices) and other self-management and health behaviors (eg, dietary recommendations, physical exercise, sedentariness, sleep, and alcohol and tobacco use). Continuing its development with an active engagement of patients and other stakeholders would allow for the setting of priorities in line with users' expectations and, thus, maximizing its utility. Challenges remain on aligning stakeholders and maintaining a community of practice that would apply and evolve this taxonomy to serve the effective and efficient development and use of MATech in health systems.

Conclusions

This paper reports reaching an agreement among diverse stakeholders from 27 countries on a comprehensive MATech terminology developed following state-of-the-art methodologies. Therefore, this study represents a key contribution toward the ambitious goal of building an information architecture to facilitate the development and implementation of MATech in health systems. Further improvements to this terminology, together with knowledge mobilization activities such as interdisciplinary networking and training across stakeholder groups, would ensure a shared conceptual and practical foundation for efficient innovation in the field of medication adherence.

Acknowledgments

The authors would like to thank the European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE) members for their feedback on the development of the ENABLE repository and early versions of this work and the ENABLE Core Team for their leadership and coordination of ENABLE activities, which this work is part of. The authors are grateful for the participation of 83 experts in the real-time Delphi study, among whom they are honored to name the following: Pavel Panayotov, Roche Bulgaria; Salvatore Pirri, Sant'Anna School of Advanced Studies, Pisa, Italy; Ema Paulino, National

XSL•FO
RenderX

Association of Pharmacies, Lisbon, Portugal; Iva Bužančić, Faculty of Pharmacy and Biochemistry, University of Zagreb, and Gradska ljekarna Zagreb, Zagreb, Croatia; Fatjona Kamberi, Ismail Qemali University of Vlora, Faculty of Health, Research Centre of Public Health, Vlore, Albania; Pawel Lewek, Department of Family Medicine, Medical University of Lodz, Poland; Katarina Fehir Šola, Pharmacy of Bjelovar, Bjelovar, Croatia; Samuel Allemann, University of Basel, Basel, Switzerland; Marie P Schneider, School of Pharmaceutical Sciences and Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Switzerland; Elena Kkolou, the Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; Emma Aarnio, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; Enrica Menditto, Center of Pharmacoeconomics and Drug Utilization Research and Department of Pharmacy, University of Naples Federico II, Naples, Italy; Harm Geers, Utrecht University, Utrecht, the Netherlands; Ulrika Gillespie, Uppsala University Hospital; Edgars Labsvīrs, former director of the National Health Service, Latvia; Katerina Mala-Ladova, Faculty of Pharmacy in Hradec Králové, Charles University, Czech Republic; Tamas Agh, Syreon Research Institute, Budapest, Hungary; Alberto Moreno-Conde, Virgen Macarena University Hospital, Seville, Spain; Dimitris Plexousakis, Institute of Computer Science, Foundation for Research and Technology – Hellas, Greece; Vildan Mevsim, Faculty of Medicine, Department of Family Medicine, Dokuz Eylül University, Izmir, Turkey; Fischer Bigna, University of Basel, Basel, Switzerland; Dermot Ryan, Usher Institute, University of Edinburgh, and Respiratory Effectiveness Group; Yvonne Wengström, Karolinska Institutet, Division of Nursing, Stockholm, Sweden; Tanja Spanic, Europa Donna Slovenia; Ana-Maria Schweitzer, Baylor Black Sea Foundation, Constanta, Romania; Maria Kamusheva, Faculty of Pharmacy, Medical University of Sofia, Sofia, Bulgaria; Elin Dahlén, Karolinska Institutet, Department of Clinical Science and Education, Södersjukhuset, Stockholm, Sweden; Marina Odalović, Faculty of Pharmacy, University of Belgrade, Belgrade, Serbia; Jasna Anđelković, Apoteka Kragujevac, Kragujevac, Serbia; Antonella Celano, Italian National Association of People With Rheumatological and Rare Diseases, Italy; Kevin Dolgin, Observia, Paris, France; Kirsten K Viktil, Diakonhjemmet Hospital Pharmacy, Oslo, Norway; Tasmania del Pino-Sedeño, Canary Islands Foundation Canary Islands Health Research Institute, Santa Cruz de Tenerife, Spain; Ana Margarida Pisco Almeida, Department of Communication and Art and DigiMedia, University of Aveiro, Portugal; Fruzsina Mezei, Center for Health Technology Assessment, Semmelweis University, and Syreon Research Institute, Budapest, Hungary; Bernard Vrijens, AARDEX Group, Liège, Belgium; Lacramioara Ochiuz, Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania; Daniela Scala, Cardarelli Hospital (radiopharmacy), Naples, Italy; Martina Šepetavc, Farmacia community pharmacy chain, Zagreb, Croatia; Nikolaos Tsakountakis, Maria Primary Care Centre, Malia, Crete, Greece; Vesna Vujic-Aleksic, the Republic of Srpska Agency for Certification, Accreditation, and Health Care Quality Improvement, Banja Luka, Bosnia and Herzegovina; Hanne Andresen, Norwegian Pharmacy Association, Norway; Davide Cafiero, Helaglobe, Florence, Italy; Paulo Félix Lamas, Singular Research Centre on Intelligent Technologies, University of Santiago de Compostela, Santiago de Compostela, Spain; Unn Sollid Manskow, Norwegian Centre for E-health Research, Tromsø, Norway; Katharina Blankart, University of Duisburg-Essen, Essen, Germany; Ivett Jakab, Patient Policy Research Unit, Syreon Research Institute, Budapest, Hungary; Ioanna Chouvarda, Aristotle University of Thessaloniki, Thessaloniki, Greece; Joao A Fonseca, Faculty of Medicine of the University of Porto, Porto, Portugal; Hanna Burkhalter, Department of Nursing and Specialist Support, Cantonal Hospital of Graubünden, Switzerland; Sean MacBride-Stewart, Pharmacy Services, National Health Service Greater Glasgow and Clyde, Glasgow, Scotland, United Kingdom; Patrizia Künzler-Heule, Cantonal Hospital St. Gallen, St. Gallen, Switzerland; Simona Zito, Italian Medicines Agency; Manuel Tabelli, Servier, Rome, Italy; Bistra Angelovska, Faculty of Medical Sciences, Goce Delchev University of Stip, Stip, Republic of North Macedonia; Lana Nezic, Faculty of Medicine, University of Banja Luka, Banja Luka, Bosnia and Herzegovina; Przemyslaw Kardas, Medication Adherence Research Center, Department of Family Medicine, Medical University of Lodz, Poland; Francisco Javier Acosta Artiles, Department of Mental Health, General Management of Healthcare Programs, Canary Islands Health Service, Las Palmas de Gran Canaria, Spain; Eric Tousset, AARDEX Group, Seraing, Belgium; Anita Strmljan, Health Insurance Institute of Slovenia, Slovenia; Želimir J Janjić, Pharmaceutical Chamber of the Republic of Srpska, Banjaluka, Republic of Srpska; Elena Drakalska Sersemova, Goce Delcev University of Stip, Republic of North Macedonia; Sonja Beckmann, University of Basel, Basel, and University Hospital Zurich, Zurich, Switzerland; Timo Heikkilä, Popit Ltd, Espoo, Finland; Aurelie Bourmaud, Robert Debré University Hospital, Greater Paris University Hospitals, Paris, France; Tina Tarre Monfrino, Swedish eHealth Agency, Sweden; and Steljana Lapa, Faculty of Public Health, Vlore, Albania. This work was funded by the European Cooperation in Science and Technology Action CA19132 ENABLE supported by the European Cooperation in Science and Technology. AD was supported by a Miguel Servet Fellowship (CP21/00062) during part of this study. The funders had no role in the study design, content work and preparation, or writing of the manuscript.

Authors' Contributions

ALD, UNM, JR, CG, FH, PB-F, SG, CJ, DS, I Tadic, MTH, and FR contributed to conceptualization and investigation. ALD, UNM, JR, FH, MTH, and FR contributed to data curation. ALD, UNM, JR, CG, FH, MTH, and FR contributed to formal analysis. JvB, I Tsiligianni, ALD, PB-F, and DS contributed to funding acquisition. ALD, UNM, and SPG contributed to methodology. ALD and UNM contributed to supervision. ALD and SG contributed to data visualization. UNM and ALD wrote the original draft. UNM, ALD, JR, CG, FH, PB-F, SG, CJ, DS, I Tadic, JvB, I Tsiligianni, MTH, and FR reviewed and edited the drafts and final version.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Delphi survey bulletins.
[DOC File , 925 KB - jmir_v27i1e59738_app1.doc]

Multimedia Appendix 2
Delphi survey data and analysis script.
[7Z File , 2463 KB - jmir_v27i1e59738_app2.7z]

Multimedia Appendix 3
Relevance and clarity ratings for the full sample and stakeholder subgroups.
[DOCX File , 333 KB - jmir_v27i1e59738_app3.docx]

Multimedia Appendix 4
Quantitative data report.
[ZIP File (Zip Archive), 3378 KB - jmir_v27i1e59738_app4.zip]

Multimedia Appendix 5
Qualitative data report.
[DOCX File , 44 KB - jmir_v27i1e59738_app5.docx]

References

1. Chan A, De Simoni A, Wileman V, Holliday L, Newby CJ, Chisari C, et al. Digital interventions to improve adherence to maintenance medication in asthma. Cochrane Database Syst Rev 2022 Jun 13;6(6):CD013030 [FREE Full text] [doi: 10.1002/14651858.CD013030.pub2] [Medline: 35691614]
2. Chan AH, Foot H, Pearce CJ, Horne R, Foster JM, Harrison J. Effect of electronic adherence monitoring on adherence and outcomes in chronic conditions: a systematic review and meta-analysis. PLoS One 2022;17(3):e0265715 [FREE Full text] [doi: 10.1371/journal.pone.0265715] [Medline: 35312704]
3. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
4. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/en/publications/investing-in-medication-adherence-improves-health-outcomes-and-health-system-efficiency_8178962c-en.html [accessed 2024-04-29]
5. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
6. European digital health uptake (DHU) radar. Digital Health Uptake (DHU). URL: https://digitalhealthuptake.eu/radar/ [accessed 2024-02-19]
7. Health apps repositories in Europe. European mHealth Hub. URL: https://mhealth-hub.org/health-apps-repositories-in-europe [accessed 2024-02-19]
8. Therapietrouwe interventies portal. InterventieNet. URL: https://interventienet.nl/ [accessed 2024-02-19]
9. Lyles CR, Adler-Milstein J, Thao C, Lisker S, Nouri S, Sarkar U. Alignment of key stakeholders' priorities for patient-facing tools in digital health: mixed methods study. J Med Internet Res 2021 Aug 26;23(8):e24890 [FREE Full text] [doi: 10.2196/24890] [Medline: 34435966]
10. Kardas P, Ágh T, Dima A, Goetzinger C, Potočnjak I, Wettermark B, et al. Half a century of fragmented research on deviations from advised therapies: is this a good time to call for multidisciplinary medication adherence research centres of excellence? Pharmaceutics 2023 Mar 14;15(3):933 [FREE Full text] [doi: 10.3390/pharmaceutics15030933] [Medline: 36986794]
11. van Boven JF, Tsiligianni I, Potočnjak I, Mihajlović J, Dima AL, Nabergoj Makovec U, et al. European network to advance best practices and technology on medication adherence: mission statement. Front Pharmacol 2021 Oct 11;12:748702 [FREE Full text] [doi: 10.3389/fphar.2021.748702] [Medline: 34707502]
12. Liyanage H, Krause P, De Lusignan S. Using ontologies to improve semantic interoperability in health data. J Innov Health Inform 2015 Jul 10;22(2):309-315 [FREE Full text] [doi: 10.14236/jhi.v22i2.159] [Medline: 26245245]

XSL•FO
RenderX

13. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak 2021 Nov 09;21(Suppl 6):206 [FREE Full text] [doi: 10.1186/s12911-021-01534-6] [Medline: 34753471]
14. Kostanjsek N. Use of the international classification of functioning, disability and health (ICF) as a conceptual framework and common language for disability statistics and health information systems. BMC Public Health 2011 May 31;11 Suppl 4(Suppl 4):S3 [FREE Full text] [doi: 10.1186/1471-2458-11-S4-S3] [Medline: 21624189]
15. Fortune N, Madden R, Almborg AH. Use of a new international classification of health interventions for capturing information on health interventions relevant to people with disabilities. Int J Environ Res Public Health 2018 Jan 17;15(1):145 [FREE Full text] [doi: 10.3390/ijerph15010145] [Medline: 29342077]
16. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak 2008 Oct 27;8 Suppl 1(Suppl 1):S2 [FREE Full text] [doi: 10.1186/1472-6947-8-S1-S2] [Medline: 19007439]
17. Classification of digital health interventions v1.0. World Health Organization. URL: https://www.who.int/publications-detail-redirect/WHO-RHR-18.06 [accessed 2024-02-19]
18. Michie S, West R, Finnerty AN, Norris E, Wright AJ, Marques MM, et al. Representation of behaviour change interventions and their evaluation: development of the Upper Level of the Behaviour Change Intervention Ontology. Wellcome Open Res 2020;5:123. [doi: 10.12688/wellcomeopenres.15902.2] [Medline: 33614976]
19. Makovec UN, Goetzinger C, Ribaut J, Barnestein-Fonseca P, Haupenthal F, Herdeiro MT, European Network to Advance Best practices and technoLogy on medication adherencE (ENABLE), European Network to Advance Best Practices and TechnoLogy on Medication AdherencE (ENABLE). Developing a medication adherence technologies repository: proposed structure and protocol for an online real-time Delphi study. BMJ Open 2022 Apr 22;12(4):e059674 [FREE Full text] [doi: 10.1136/bmjopen-2021-059674] [Medline: 35459677]
20. Kardas P, Aarnio E, Agh T, van Boven JF, Dima AL, Ghiciuc CM, et al. New terminology of medication adherence enabling and supporting activities: ENABLE terminology. Front Pharmacol 2023;14:1254291 [FREE Full text] [doi: 10.3389/fphar.2023.1254291] [Medline: 37900155]
21. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
22. Perleth M, Jakubowski E, Busse R. What is 'best practice' in health care? State of the art and perspectives in improving the effectiveness and efficiency of the European health care systems. Health Policy 2001 Jun;56(3):235-250. [doi: 10.1016/s0168-8510(00)00138-x] [Medline: 11399348]
23. WHO definition of health technologies. World Health Organization. 2007. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA60/A60_R29-en.pdf [accessed 2024-04-13]
24. Sabaté E. Adherence to long-term therapies : evidence for action. World Health Organization. 2003. URL: https://iris.who.int/handle/10665/42682 [accessed 2024-04-29]
25. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2024-04-29]
26. Turn data complexity into clarity. NVivoPro. URL: https://lumivero.com/ [accessed 2024-04-29]
27. Fitch K, Bernstein SJ, Aguilar MD, Burnand B. The RAND/UCLA appropriateness method user's manual. RAND Corporation. 2001. URL: https://www.rand.org/pubs/monograph_reports/MR1269.html [accessed 2024-04-29]
28. Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? Eur J Integr Med 2015 Aug 20;7(4):423-428 [FREE Full text] [doi: 10.1016/j.eujim.2015.07.002]
29. The ENABLE repository. European Network to Advance Best Practices and Technology on Medication Adherence. URL: https://enableadherence.eu/enable-repository-matech/ [accessed 2024-04-21]
30. Dima AL, Allemann SS, Dunbar-Jacob J, Hughes DA, Vrijens B, Wilson IB. Methodological considerations on estimating medication adherence from self-report, electronic monitoring and electronic healthcare databases using the TEOS framework. Br J Clin Pharmacol 2023 Jul;89(7):1918-1927 [FREE Full text] [doi: 10.1111/bcp.15375] [Medline: 35491721]
31. Jansen-Kosterink S, Broekhuis M, van Velsen L. Time to act mature-Gearing eHealth evaluations towards technology readiness levels. Digit Health 2022 Jul 11;8:20552076221113396 [FREE Full text] [doi: 10.1177/20552076221113396] [Medline: 35847525]
32. Digital health. Digital Health Europe. URL: https://digitalhealtheurope.eu/glossary/digital-health/ [accessed 2024-04-13]
33. Jacob C, Lindeque J, Müller R, Klein A, Metcalfe T, Connolly SL, et al. A sociotechnical framework to assess patient-facing eHealth tools: results of a modified Delphi process. NPJ Digit Med 2023 Dec 15;6(1):232 [FREE Full text] [doi: 10.1038/s41746-023-00982-w] [Medline: 38102323]
34. Orrego C, Ballester M, Heymans M, Camus E, Groene O, Niño de Guzman E, COMPAR-EU Group. Talking the same language on patient empowerment: development and content validation of a taxonomy of self-management interventions for chronic conditions. Health Expect 2021 Oct;24(5):1626-1638 [FREE Full text] [doi: 10.1111/hex.13303] [Medline: 34252259]
35. Heijmans M, Poortvliet R, Van der Gaag M, González-González AI, Beltran Puerta J, Canelo-Aybar C, et al. Using a taxonomy to systematically identify and describe self-management interventions components in randomized trials for

XSL•FO
RenderX

COPD. Int J Environ Res Public Health 2022 Oct 04;19(19):12685 [FREE Full text] [doi: 10.3390/ijerph191912685] [Medline: 36231985]
36. Ekenberg M, Landin F, Wettermark B. A cross-sectional study identifying medication adherence technologies (MATech) in Sweden using behavior change techniques. Patient Prefer Adherence 2024;18:2281-2293 [FREE Full text] [doi: 10.2147/PPA.S481152] [Medline: 39559483]
37. Wright AJ, Norris E, Finnerty AN, Marques MM, Johnston M, Kelly MP, et al. Ontologies relevant to behaviour change interventions: a method for their development. Wellcome Open Res 2020;5:126 [FREE Full text] [doi: 10.12688/wellcomeopenres.15908.3] [Medline: 33447665]
38. Haag M, Lehmann A, Hersberger KE, Schneider MP, Gauchet A, Vrijens B, et al. The ABC taxonomy for medication adherence translated into French and German. Br J Clin Pharmacol 2020 Apr;86(4):734-744 [FREE Full text] [doi: 10.1111/bcp.14172] [Medline: 31721271]
39. Félix IB, Silva CC, Guerreiro MP, Rafael Henriques H, Michie S, Henriques MA, et al. Translating the behaviour change technique taxonomy v.1 to other languages: the approach used in European Portuguese (BCTTv1-PT). Wellcome Open Res 2024;9:114 [FREE Full text] [doi: 10.12688/wellcomeopenres.20609.1] [Medline: 39512382]
40. van Boven JF, Dima AL, Wettermark B, Potočnjak I, Ágh T, ENABLE collaborators. Leveraging digital medication adherence technologies to enhance sustainability of European health systems: ENABLE's key recommendations. Lancet Reg Health Eur 2025 Jan;48:101164 [FREE Full text] [doi: 10.1016/j.lanepe.2024.101164] [Medline: 39697213]
41. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health 2022 Mar;4(3):e200-e206 [FREE Full text] [doi: 10.1016/S2589-7500(21)00292-2] [Medline: 35216754]
42. Tamrat T, Ratanaprayul N, Barreix M, Tunçalp Ö, Lowrance D, Thompson J, et al. Transitioning to digital systems: the role of World Health Organization's Digital Adaptation Kits in operationalizing recommendations and interoperability standards. Glob Health Sci Pract 2022 Feb 28;10(1):e510 [FREE Full text] [doi: 10.9745/GHSP-D-21-00320] [Medline: 35294382]

Abbreviations

COMPAR-EU: Comparing Effectiveness of Self-Management Interventions in 4 High-Priority Chronic Diseases in Europe
CQV: coefficient of quartile variation
DI: disagreement index
ENABLE: European Network to Advance Best Practices and Technology on Medication Adherence
HCP: health care professional
ICC: intraclass correlation coefficient
IPR: interpercentile range
IPRAS: interpercentile range adjusted for symmetry
MATech: medication adherence technology
PDM: policy and decision-making
SC: steering committee
TRL: technology readiness level
WHO: World Health Organization

Edited by A Schwartz; submitted 28.04.24; peer-reviewed by U Sinha, S Mucherino; comments to author 28.09.24; revised version received 29.09.24; accepted 19.12.24; published 25.03.25.

Please cite as:
Dima AL, Nabergoj Makovec U, Ribaut J, Haupenthal F, Barnestein-Fonseca P, Goetzinger C, Grant S, Jácome C, Smits D, Tadic I, van Boven J, Tsiligianni I, Herdeiro MT, Roque F, European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE)
Stakeholder Consensus on an Interdisciplinary Terminology to Enable the Development and Uptake of Medication Adherence Technologies Across Health Systems: Web-Based Real-Time Delphi Study
J Med Internet Res 2025;27:e59738
URL: https://www.jmir.org/2025/1/e59738
doi:10.2196/59738
PMID:

©Alexandra Lelia Dima, Urska Nabergoj Makovec, Janette Ribaut, Frederik Haupenthal, Pilar Barnestein-Fonseca, Catherine Goetzinger, Sean Grant, Cristina Jácome, Dins Smits, Ivana Tadic, Job van Boven, Ioanna Tsiligianni, Maria Teresa Herdeiro,

XSL•FO
RenderX

Fátima Roque, European Network to Advance Best Practices and Technology on Medication Adherence (ENABLE). Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

New Perspective on Digital Well-Being by Distinguishing Digital Competency From Dependency: Network Approach

Si Chen[1], BS; Omid V Ebrahimi[2,3], PhD; Cecilia Cheng[1], PhD

[1]Department of Psychology, University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

[3]Department of Psychology, University of Oslo, Oslo, Norway

Corresponding Author:
Cecilia Cheng, PhD
Department of Psychology
University of Hong Kong
Pokfulam
Hong Kong
China (Hong Kong)
Phone: 852 39174224
Email: ceci-cheng@hku.hk

Abstract

Background: In the digital age, there is an emerging area of research focusing on digital well-being (DWB), yet conceptual frameworks of this novel construct are lacking. The current conceptualization either approaches the concept as the absence of digital ill-being, running the risk of pathologizing individual digital use, or follows the general subjective well-being framework, failing to highlight the complex digital nature at play.

Objective: This preregistered study aimed to address this gap by using a network analysis, which examined the strength of the relationships among affective (digital stress and web-based hedonic well-being), cognitive (online intrinsic needs satisfaction), and social (online social connectedness and state empathy) dimensions of DWB and their associations with some major DWB protective and risk factors (ie, emotional regulation, nomophobia, digital literacy, self-control, problematic internet use, coping styles, and online risk exposure).

Methods: The participants were 578 adults (mean age 38.7, SD 13.14 y; 277/578, 47.9% women) recruited from the United Kingdom and the United States who completed an online survey. Two network models were estimated. The first one assessed the relationships among multiple dimensions of DWB, and the second examined the relationships between DWB dimensions and related protective and risk factors.

Results: The 2 resulting network structures demonstrated high stability, with the correlation stability coefficients being 0.67 for the first and 0.75 for the second regularized Gaussian graphical network models. The first network indicated that all DWB variables were positively related, except for digital stress, which was negatively correlated with the most central node—online intrinsic needs satisfaction. The second network revealed 2 distinct communities: *digital competency* and *digital dependency*. Emotional regulation emerged as the most central node with the highest bridge expected influence, positively associated with emotion-focused coping in the *digital competency* cluster and negatively associated with avoidant coping in the *digital dependency* cluster. In addition, some demographic differences were observed. Women scored higher on nomophobia (χ^2_4=10.7; *P*=.03) and emotion-focused coping (χ^2_4=14.9; *P*=.01), while men scored higher on digital literacy (χ^2_4=15.2; *P*=.01). Compared with their older counterparts, younger individuals scored lower on both emotional regulation (Spearman ρ=0.27; *P*<.001) and digital self-control (Spearman ρ=0.35; *P*<.001) and higher on both digital stress (Spearman ρ=−0.14; *P*<.001) and problematic internet use (Spearman ρ=−0.25; *P*<.001).

Conclusions: The network analysis revealed how different aspects of DWB were interconnected, with the cognitive component being the most influential. Emotional regulation and adaptive coping strategies were pivotal in distinguishing digital competency from dependency.

(J Med Internet Res 2025;27:e70483) doi:10.2196/70483

XSL•FO
RenderX

KEYWORDS

digital wellness; affective well-being; emotional regulation; coping; digital competence; digital autonomy; artificial intelligence

Introduction

Background

Digital well-being (DWB), which refers to the subjective experience of optimal balance between the benefits and drawbacks obtained from online connectivity, has emerged as a critical area of concern in the current digital era [1]. As of July 2024, global estimates indicate that the number of active internet users has reached 5.45 billion, representing two-thirds of the world's population, with a significant portion also engaging actively on social media platforms [2]. In this context, understanding and promoting DWB has become essential for bolstering individual well-being in an increasingly digitalized society.

Existing studies adopted 2 distinct approaches for conceptualizing this novel construct. The first approach conceptualizes DWB as the absence of problems resulting from internet and technology use, drawing parallels to frameworks of digital addiction (DA) [3-5]. This perspective encompasses various subtypes of problematic digital use, such as internet gaming disorder, smartphone addiction, and social media addiction, which have attracted growing attention from both the professional community and the general public [6-8]. DWB research that follows the DA framework has provided valuable insights into the risks of unhealthy digital behaviors [9-12]. However, equating well-being with the absence of such behaviors fundamentally differs from the broader conception of well-being as a state of "optimal psychological experience and functioning" [13]. An exclusive focus on ill-being can medicalize individuals' relationships with digital devices, thereby overlooking the factors that foster positive and fulfilling digital experiences [1,14,15].

The second approach to conceptualizing DWB is grounded in subjective well-being (SWB) frameworks, emphasizing the hedonic and eudaimonic benefits of digital use [16,17]. This SWB approach to the conceptualization addresses affective well-being (eg, pleasure), satisfaction across various life domains (eg, interpersonal relations and work productivity), and overall satisfaction with contemporary digital life marked by constant internet connectivity and availability [17]. In contrast to the DA framework, the SWB approach highlights the concept of "controlled pleasure" and the use of digital devices to meet individuals' intrinsic needs [18,19]. Despite the advantages of the SWB framework in promoting healthy digital relationships, its reliance on general SWB measures often fails to account for the domain specificity of digital connectivity [18,20].

Toward a Tripartite Framework: Affective, Cognitive, and Social DWB

To address the shortcomings of both the DA and SWB frameworks, Büchi [17] proposed a new DWB framework that emphasizes how distal well-being outcomes (eg, psychological health and SWB) are related to 3 key constructs, namely, digital practices (eg, social media use and online behaviors), proximal harms (eg, digital stress), and benefits (eg, social connectedness and hedonic experience). These connections are further moderated by individual and situational variables [17]. Building on this theoretical understanding, this study expands the SWB framework by conceptualizing DWB as comprising individual dimensions of affective (ie, positive and negative affect), cognitive (ie, intrinsic needs satisfaction), and social (ie, experiences of connectedness and empathy) well-being, with an emphasis on the inherently digital nature of DWB [13,21].

In line with the conceptualization of affective well-being, the affective dimension of DWB reflects both positive and negative emotional experiences. On the negative side, digital stress encompasses the stress arising from the pressure of mobile connectivity, which can lead to expectations for constant availability (ie, availability stress) and feelings of being overwhelmed due to an excess of information on the web (ie, information overload) [22]. These affective stressors have been shown to have negative influences on mental health [22-24]. Conversely, the positive aspect of affective DWB includes the hedonic pleasure derived from engaging in digital activities [1,25].

A key conceptual component of DWB, aligned with the SWB framework, pertains to the cognitive dimension. Cognitive well-being refers to individuals' evaluative judgments of their overall life satisfaction and satisfaction within specific domains [21]. Drawing from the SWB framework, previous DWB research has highlighted the importance of 3 core needs—autonomy, competence, and relatedness—in determining satisfaction with digital use [16]. This study focuses on intrinsic needs for autonomy and competence, which represent eudaimonic benefits derived from digital connectivity [1,16].

The increasingly pervasive role of social media in daily social interactions, a factor that has fundamentally transformed the processes of socialization, highlights the need to examine the social aspect as a distinct dimension of individual DWB. To unpack how individuals perceive and experience their relations with others and society at large through digital activities, this study operationalizes social DWB as comprising online social connectedness and online state empathy [26-30].

Enhancing the DWB Framework: Role of Protective and Risk Factors

To formulate a comprehensive conceptual framework of DWB that addresses the variability inherent in digital connectivity [17], this research incorporates a cluster of DWB-related factors (ie, protective and risk factors).

One body of studies focusing on DWB enhancement highlighted several protective factors, including the ability to regulate one's emotions (ie, emotional regulation), the ability to resolve conflicts between digital use and real-life goals (ie, digital self-control), skillful use of technologies (ie, digital literacy), and adaptive coping strategies such as emotion-focused coping [31-33]. Digital platforms provide users access to emotionally

XSL•FO
RenderX

charged content and facilitate swift emotional exchanges through media emotional affordances such as likes, comments, and reactions [34,35]. Therefore, effective emotional regulation and adaptive emotion-focused coping are essential to prevent negative impacts from these interactions. Studies have indicated that individuals with stronger emotional regulation report better well-being and fewer problematic digital behaviors [36-38]. In addition to emotional affordances, online connectivity poses another threat to people's DWB. Self-control becomes crucial for maintaining responsible internet use and resisting excessive use [4,39-41]. Studies also revealed the important role of digital literacy in a meaningful digital experience, which contributes to individual cognitive well-being by meeting their needs for competence and autonomy through skillful digital use [42,43].

Another body of DWB studies that follows the DA framework has identified several key affective and behavioral risk factors, including emotional attachment to mobile phones (ie, nomophobia), addictive internet use (ie, problematic internet use), maladaptive coping (ie, avoidant coping), and involuntary exposure to unwanted experiences (ie, online risk exposure) [10,44-46]. Meta-analyses revealed the harmful effect of overreliance on phone and internet use on mental health outcomes [47,48]. In addition, online risks, including unwanted explicit content, cyberbullying, sexual solicitation, and leaking of privacy, have been shown to elicit emotional and psychological harm, especially for younger internet users [46,49]. Avoidant coping refers to attempts to ignore or evade stressors rather than addressing them by attending to distractive digital activities. This maladaptive coping can reduce overall resilience and exacerbate feelings of stress, anxiety, and dissatisfaction [50].

In summary, this study examines the relationships among the affective, social, and cognitive dimensions of DWB as well as the associations between its related factors, which are a selected cluster of affective, cognitive, and behavioral protective and risk factors. This study uses a system-based approach [51] to uncover patterns and interactions that contribute to the overall structure of DWB. Using network analysis [52], it aims to reveal how various psychological and contextual factors (ie, related factors) relate to specific DWB components while accounting for the influence of other variables in the network. In doing so, this approach aims to provide a deeper understanding of the construct, illustrating how dimensions of DWB and its related factors collectively shape digital wellness.

Methods

Recruitment

This online study was accessible to participants using any type of digital device with internet access. Recruitment was facilitated through Prolific Academic, a crowdsourcing platform that offers access to large and diverse samples while ensuring high data quality for academic research [53]. Eligible participants included adults aged ≥18 years from the United Kingdom and the United States with a good track record on the survey platform (ie, an approval rate of ≥90% in previous surveys). Participants were informed about the study's objectives, which involved completing a series of survey questions exploring their digital and online habits and activities. The hourly compensation rate of £2.50 (US $2.30) was provided upfront to ensure transparency and encourage voluntary participation. A total of 578 participants were recruited.

Research Design and Procedure

This study used a cross-sectional design to collect data using self-report questionnaires administered through the Qualtrics software. Interested and eligible individuals were directed to a survey via a URL provided on the Prolific recruitment platform. Participants indicated their informed consent by checking a box before starting the survey. The survey included the measures described in the Overview section, which were presented in English and randomized in order. After completing the survey, participants were thanked and compensated.

Measures

Overview

For each variable, the sum scores of the measures were incorporated as nodes in the network. Further details of the measures administered in the study are shown in Table 1.

XSL•FO
RenderX

Table 1. All measured variables included in the network analysis (N=578)[a].

Label	Variable or node	Definition	Sample items
Affective DWB[b]			
ADWB[c]1	Digital stress	The degree to which people felt stressed and pressured due to mobile communication and information overloads	• "I feel stressed by the amount of contact through my mobile phone." • "I am pressured to respond quickly to all calls or texts." • "I am frequently overwhelmed by the amount of information available online."
ADWB2	Web-based hedonic well-being	Hedonic experience from online activities	• "Being online entertains and stimulates my mind."
Cognitive DWB			
CDWB[d]	Intrinsic needs satisfaction	Individual satisfaction with needs for autonomy and competence resulted from the use of social media	• "When I am on social media, I feel free to be who I am." • "When I am on social media, I feel very capable and effective."
Social DWB			
SDWB[e]	Social digital well-being	The ability to understand, feel, and sympathize with others in digital space (online state empathy) and the degree to which people feel connected and related to others on the web (online social connectedness)	• "I knew what the person I was interacting with felt emotionally." • "I felt the same way as the individual I was interacting through digital interactions." • "I feel close to people online." • "I don't feel related to most people on social media." [Reverse coded]
Affective-related factors			
AC[f]1	Emotional regulation	The ability to recognize and exert control over one's own emotional state	• "I have difficulty making sense out of my feelings." [Reverse coded] • "When I am upset, I have difficulty thinking about anything else." [Reverse coded]
AC2	Nomophobia	The fear of being disconnected from the digital space	• "If I did not have my smartphone, I would be uncomfortable because I could not stay up to date with social media and online networks." • "Running out of battery in my smartphone would scare me."
Cognitive-related factors			
CC[g]1	Digital literacy	Individual ability of skillful digital use	• "I know how to open a new tab in my browser." • "I feel comfortable deciding who to follow online (eg, on services like Twitter or Tumblr)."
CC2	Digital self-control	The ability to successfully manage goal conflicts between digital use and real-life activities	• "How often do you give in to a desire to use social media even though your social media use at that particular moment makes you delay other things you want or need to do?" [Reverse coded]
Behavioral-related factors			
BC[h]1	Problematic internet use	Compulsive and excessive use of internet	• "I lose track of time online." • "I feel lost if I can't go online."
BC2	Problem-focused coping	Coping by directly confronting a stressor in an attempt to decrease or eliminate it	• "I've been thinking hard about what steps to take."

XSL•FO
RenderX

Label	Variable or node	Definition	Sample items
BC3	Emotion-focused coping	Coping using skills for processing and dealing with feelings that arise due to stressful situations	• "I've been making fun of the situation."
BC4	Avoidant coping	Coping by not addressing the problem directly but instead disengaging from the situation	• "I've been turning to work or other activities to take my mind off things."
BC5	Online risk exposure	The frequency at which individuals are exposed to unwanted stressful events on the web	• "Someone made rude or mean comments about you or threatened you in some way online." • "You saw online stories, images, or videos that contained excessive violence that made you feel uncomfortable."

[a]This table outlines all the variables tested in the network analysis, detailing their operational definitions and sample items. All the variables were measured using validated psychological scales in an online self-report survey conducted from June 2024 to July 2024.

[b]DWB: digital well-being.

[c]ADWB: affective digital well-being.

[d]CDWB: cognitive digital well-being.

[e]SDWB: social digital well-being.

[f]AC: affective-related factors.

[g]CC: cognitive-related factors.

[h]BC: behavioral-related factors.

Digital Stress

The Entrapment Scale [54] consists of 6 items that assess stress related to mobile communication and 3 items that evaluate information overload. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), indicating the extent to which they feel pressured by online communication and information available on the web. The sum score ranges from 0 to 36, with higher scores indicating greater levels of digital stress due to mobile entrapment and information overload.

Web-based Hedonic Well-Being

Web-based hedonic well-being was measured using the Online Hedonic Scale [55], which comprises 3 items that assess pleasant subjective experience with digital activities. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), indicating the degree to which they derive pleasure from being on the web. The sum score ranges from 0 to 12, and higher scores indicate greater levels of web-based hedonic well-being.

Intrinsic Needs Satisfaction

The 6-item Autonomy and Competence subscale was adapted from the Intrinsic Need Satisfaction Scale [16], which is a 12-item measure designed to assess individual intrinsic need satisfaction derived from digital use. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), reflecting the extent to which they feel their needs for autonomy and competence were fulfilled through social media engagement. The sum score ranges from 0 to 48, with higher scores indicating greater levels of internal need satisfaction.

Online Social Well-Being

The State Empathy Scale [56] is a measure consisting of 9 items that assess an individual's ability to understand, feel, and empathize with others in the digital space. The online social connectedness measures were adapted from the Social Connectedness Scale [28], which also includes 9 items evaluating the extent to which individuals feel connected and related to others on social media. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), indicating the extent to which they experience empathy during online interactions. The sum score ranges from 0 to 72, with higher scores indicating greater levels of online social well-being.

Emotional Regulation

The Difficulties in Emotion Regulation Scale [57] consists of 16 items that assess individuals' challenges in recognizing and managing their emotional states. Participants rated each item on a Likert scale ranging from 0 (strongly agree) to 4 (strongly disagree), reflecting the degree to which they experience difficulties in emotional regulation. The item scores are reversely coded, and the sum score ranges from 0 to 64, with higher scores reflecting greater levels of emotional regulation.

Nomophobia

The Nomophobia Scale [58] comprises 20 items that measure the fear of being unable to communicate and access information, losing connectedness, and sacrificing convenience due to disconnection from mobile phone connectivity. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), reflecting the extent to which they experience the aforementioned type of fear. The sum score ranges from 0 to 80, with higher scores indicating greater levels of nomophobia.

XSL•FO
RenderX

Digital Literacy

The Digital Literacy Scale [59] consists of 35 items that assess individuals' ability in skillful digital use. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), reflecting their level of digital competence. The sum score ranges from 0 to 140, with higher scores reflecting greater levels of digital literacy.

Digital Self-Control

The Social Media Self-Control Failure Scale [60] consists of 3 items that evaluate self-control failures arising from goal conflicts between digital and real-life activities. Participants rated each item on a Likert scale ranging from 0 (strongly agree) to 4 (strongly disagree), reflecting the extent to which they struggle to manage goal conflicts between online engagement and real-life activity. The sum score ranges from 0 to 12, with higher scores indicating greater levels of digital self-control.

Problematic Internet Use

The Problematic Internet Usage Scale [8] consists of 15 items that measure compulsive and excessive internet use. Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), indicating the extent of their problematic internet use. The sum score ranges from 0 to 60, with higher scores reflecting greater levels of problematic internet use.

Coping Styles

The Brief Coping Orientation to Problems Experienced Inventory [61] consists of 28 items that assess individual preferred coping styles (ie, problem focused, emotion focused, and avoidant). Participants rated each item on a Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree), reflecting the extent to which they adopt a particular coping style. The sum score ranges from 0 to 32 for problem-focused coping, 0 to 48 for emotion-focused coping, and 0 to 32 for avoidant coping. Higher scores indicate greater deployment of the corresponding coping style.

Online Risk Exposure

The Online Risk Exposure Scale [46] consists of 16 items that measure the frequency of involuntary exposure to unwanted stressful events on the web. Participants rated each item on a Likert scale ranging from 0 (never) to 4 (always), indicating the frequency of their risk exposure to unwanted online activities. The sum score ranges from 0 to 64, with higher scores indicating greater frequencies of online risk exposure.

Ethical Considerations

The study was preregistered on the Open Science Framework [62]. Ethics approval for this study was obtained from the Human Research Ethics Committee of the University of Hong Kong before the study began (EA240174). Before participation, individuals were presented with information regarding the study's objectives, their rights to withdraw at any time without consequences, and the requirement for informed consent for research participation. All participants were compensated at a rate of £2.50 (US $2.30) per hour upon completing the study. To protect privacy and confidentiality, all data were anonymized and identified only through a randomly assigned participant ID. Responses were securely stored following data protection guidelines, including encrypting data during both storage and transmission.

Statistical Analysis

All analyses were conducted using R software (version 4.1.3) and RStudio (R Core Team, 2024).

Preliminary Analysis

Shapiro-Wilk normality tests were first used to check normality. Kruskal-Wallis tests were conducted to examine potential differences across gender groups in cases of nonnormally distributed data. If significant differences (ie, $P<.05$) were detected, post hoc pairwise comparisons were performed using the Wilcoxon rank sum test with continuity correction to investigate differences across gender categories. The comparisons were adjusted using the Bonferroni correction method. Spearman rank-order correlation tests were conducted to assess differences based on age distribution.

Network Analysis

The R package *huge* was first used to apply nonparanormal data transformation for skewed data (version 1.3.5; Jiang et al [63]). Following guidelines for network analysis, variables were theoretically selected to avoid conceptual overlap [52,64]. A data-driven method was also applied to confirm this [65]. First, the correlation matrix was checked to ensure that it was positive definite, controlling for linear combinations among the variables. Then, the *goldbricker* function from the R package *networktools* (version 1.5.2; Jones [66]) was used to identify potential redundant variables. No redundancies were found, supporting that the theoretical selection had yielded nonoverlapping variables.

A regularized graphical Gaussian model (GGM) was used to estimate 2 network structures [67,68]. The first network (DWB network) contained the affective, cognitive, and social dimensions of DWB. The second network included DWB factors from the first network and related risk and protective factors selected from previous DWB research. The DWB-related associations were examined in the second network model (DWB-related network).

The network models were constructed using the EBICglasso algorithm in the R package *qgraph* (version 1.9.4; Epskamp et al [69]). This procedure involves using the graphical least absolute shrinkage and selection operator to estimate a GGM through a regularization technique. This method balances model complexity and sample size to select the best-fitting model based on the extended Bayesian information criterion, which includes a penalty term that discourages overfitting. The graphical least absolute shrinkage and selection operator applies L1-regularization, shrinking small edge weights to zero, which results in a simpler network that highlights the most important connections. After constructing a series of networks, the model with the lowest extended Bayesian information criterion is chosen to ensure interpretability and avoid overfitting [70]. Networks were then visualized using the Fruchterman-Reingold algorithm [71]. Thicker edges indicate stronger associations

XSL•FO
RenderX

between nodes, with red edges representing negative and blue edges representing positive associations.

The strength centrality measure for each variable was obtained across the 2 networks. This index reflects the overall strength connectivity of a variable with other variables from the studied system. The node with the highest strength centrality is considered the most central or influential, as it has the most and strongest associations within the network, underscoring its overall importance in the system. In this study, the strength centrality index was used to identify the most influential variable in examining the construct of DWB.

Community Detection for DWB-Related Network

The *spin glass* algorithm was used to detect the community structure of the DWB-related network (Yang et al [72]) using the R package *igraph* (version 7.2.3; Csárdi [73]). The communities represent clusters of nodes, which demonstrate stronger connections with one another compared to nodes from other clusters. In addition to strength centrality measures, the bridge expected influence (BEI) of each node was computed to reflect each node's overall strength of connectivity with other communities detected in the DWB-related network using the R package *networktools* (version 1.5.2 [74]). This BEI index reveals the nodes that act as bridges between the different communities in the network. The node with the highest BEI indicates the crucial role played by the variable in linking the other cluster. In this study, the BEI index allows for the understanding of the mechanisms underlying the relationship between the protective and risk-oriented aspects of digital health.

Network Accuracy and Stability

Nonparametric and case-dropping bootstrapping tests were used to assess the accuracy and stability of the resulting networks using the R package *bootnet* (version 1.6 [75]). The accuracy of edge weights was tested using nonparametric bootstrapping with 1000 iterations. The stability of strength centrality and BEI measures were tested using case-dropping subset bootstrapping with 1000 iterations. In this process, the correlation between the original centrality indices and those derived from smaller subsets was calculated, where up to 75% (434/578) of the participants were dropped. The correlation stability coefficients were then computed, showing the largest proportion of data that can be excluded while still maintaining a correlation ≥0.70 with the original centrality indices, with 95% confidence. A network is considered stable if the correlation stability coefficient is at least 0.25, with a preferred value >0.50 [76].

Results

Preliminary Results

Table 2 summarizes the demographic characteristics of participants of this study.

The descriptive statistics of DWB and its related factors are summarized in Table 3. The Shapiro-Wilk normality results indicated that data for all measurements, except problematic internet use, were nonnormally distributed (Multimedia Appendix 1). Gender was found to be a significant factor in nomophobia (χ^2_4=10.7; *P*=.03), digital literacy (χ^2_4=15.2; *P*=.01), and emotion-focused coping (χ^2_4=14.8; *P*=.01). Specifically, a significant difference was observed, with women (median 48.00, IQR 36.00-60.00) scoring significantly higher than men (median 43.00, IQR 28.35-57.65) on nomophobia (*P*=.02). In addition, women (median 22.00, IQR 19.50-24.50) also scored significantly higher than men (median 21.00, IQR 17.50-24.50) on emotion-focused coping (*P*=.01). Conversely, men (median 113.00, IQR 102.50-123.50) scored significantly higher than women (median 107.00, IQR 96.00-118.00) on digital literacy (*P*=.01).

The Spearman rank-order correlation test results are summarized in Table 4. Statistically significant positive relationships were found between age and intrinsic need satisfaction, emotional regulation, and digital self-control. Conversely, age was found to be negatively related to digital stress, nomophobia, problematic internet use, problem-focused coping, emotion-focused coping, and online risk exposure.

XSL•FO
RenderX

Table 2. Demographic characteristics of the sample (N=578)[a].

Demographic variable	Values
Age (y), mean (SD)	38.70 (13.14)
Gender, n (%)	
Women	277 (47.9)
Men	288 (49.8)
Nonbinary or third gender	10 (1.7)
Other	1 (0.2)
Prefer not to say	2 (0.3)
Race, n (%)	
Asian	33 (5.7)
Black or African American	48 (8.3)
Hispanic or Latino	8 (1.4)
Native American	2 (0.4)
White	467 (80.8)
Other	22 (3.8)
Education, n (%)	
Less than high school diploma	12 (2.1)
High school diploma	126 (21.8)
Higher diploma	91 (15.7)
Associate degree	23 (4)
Bachelor’s degree	214 (37)
Master’s degree	82 (14.2)
Professional degree	13 (2.2)
Doctoral degree	17 (2.9)
Marital status, n (%)	
Single	161 (27.8)
In a relationship	154 (26.6)
Married	235 (40.7)
Divorced	18 (3.1)
Widowed	6 (1)
Other	4 (0.7)

[a]The table provides a detailed breakdown of the major demographic information of the participants who took part in the online self-report survey conducted between June 2024 and July 2024.

XSL•FO
RenderX

Table 3. Descriptive statistics of digital well-being–related measurements (N=578)[a].

	Scores, mean (SD; range)	Scores, median (IQR)
Digital stress	18.70 (7.07; 1.00-40.00)	18.00 (13.50-22.50)
Web-based hedonic well-being	10.80 (2.43; 0.00-16.00)	11.00 (10.00-12.00)
Intrinsic needs satisfaction	14.90 (4.16; 2.00-24.00)	15.00 (12.00-18.00)
Social digital well-being	35.60 (9.06; 9.00-61.00)	35.00 (29.00-41.00)
Emotional regulation	35.80 (11.80; 4.00-64.00)	36.00 (27.00-45.00)
Nomophobia	43.20 (18.00; 0.00-80.00)	45.00 (31.25-58.75)
Digital literacy	108.00 (17.00; 35.00-138.00)	110.00 (98.50-121.50)
Digital self-control	6.20 (2.76; 0.00-12.00)	6.00 (4.00-8.00)
Problematic internet use	26.80 (8.79; 0.00-49.00)	27.00 (21.00-33.00)
Problem-focused coping	13.70 (8.44; 0.00-43.00)	12.00 (6.00-18.00)
Emotion-focused coping	20.70 (4.52; 4.00-31.00)	21.00 (18.00-24.00)
Avoidant coping	25.90 (5.44; 3.00-42.00)	26.00 (22.63-29.38)
Online risk exposure	13.00 (4.45; 1.00-30.00)	13.00 (10.00-16.00)

[a]This table summarizes the descriptive statistics of all variables assessed in the online self-report survey conducted between June 2024 and July 2024.

Table 4. Spearman rank-order correlation test results of age differences (N=578)[a].

	Spearman rank-order correlation	*P* value
Digital stress	−0.138	<.001
Web-based hedonic well-being	−0.022	.59
Intrinsic needs satisfaction	0.088	.03
Social digital well-being	−0.081	.051
Emotional regulation	0.265	<.001
Nomophobia	−0.200	<.001
Digital literacy	−0.077	.07
Digital self-control	0.350	<.001
Problematic internet use	−0.249	<.001
Problem-focused coping	−0.373	<.001
Emotion-focused coping	0.042	.31
Avoidant coping	−0.204	<.001
Online risk exposure	−0.144	<.001

[a]This table shows the correlational test results between age and measured variables, highlighting age-related patterns in digital behaviors. Data were collected through an online survey conducted between June 2024 and July 2024.

Network Structure of DWB Factors

Network Estimation

Figure 1 presents the regularized GGM network structure of the DWB measures. The well-being components were positively associated with each other, except for digital stress, which was negatively linked to intrinsic needs satisfaction (CDWB) and unrelated to web-based hedonic well-being (ADWB2). The strongest positive association was between CDWB and ADWB2. The second and third strongest positive edges were those between ADWB2 and social digital well-being (SDWB) and between intrinsic needs satisfaction (CDWB) and SDWB.

XSL•FO
RenderX

Figure 1. The digital well-being (DWB) network structure (N=578). The figure illustrates the relationships among 4 dimensions of DWB derived from a network analysis performed on survey data collected between June 2024 and July 2024. Blue edges indicate positive associations, while the red edge indicates a negative association between variables. The thickness of each edge reflects the strength of the observed relationships, with thicker edges denoting stronger associations. This DWB network highlights the interconnected nature of the DWB dimensions. ADWB1: digital stress; ADWB2: web-based hedonic well-being; CDWB: intrinsic needs satisfaction; SDWB: social digital well-being.

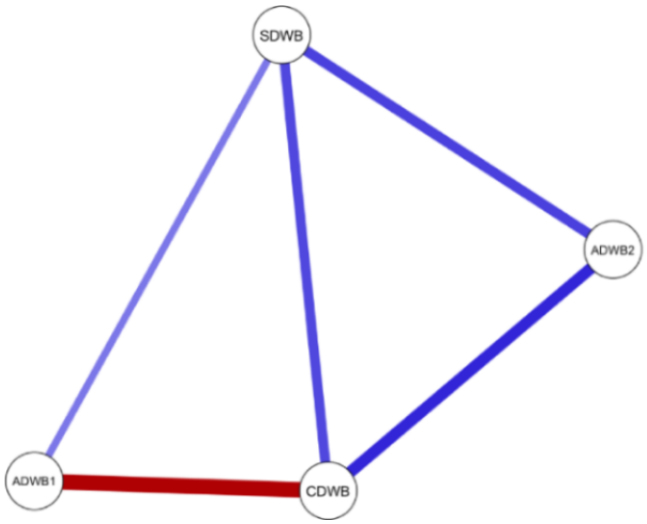

Strength Centrality

To examine the relative importance of each node in the DWB network, Figure 2 presents the strength centrality results. The cognitive component of DWB (ie, intrinsic needs satisfaction) had the highest strength centrality, followed by SDWB, suggesting the important roles played by the cognitive and social dimensions.

Figure 2. The centrality plot of the digital well-being network (N=578). This figure displays the centrality strength of all digital well-being variables derived from a network analysis performed on survey data collected between June 2024 and July 2024, highlighting their relative importance. The x-axis represents the centrality strength values, with higher values indicating more central roles in the network. ADWB1: digital stress; ADWB2: web-based hedonic well-being; CDWB: intrinsic needs satisfaction; SDWB: social digital well-being.

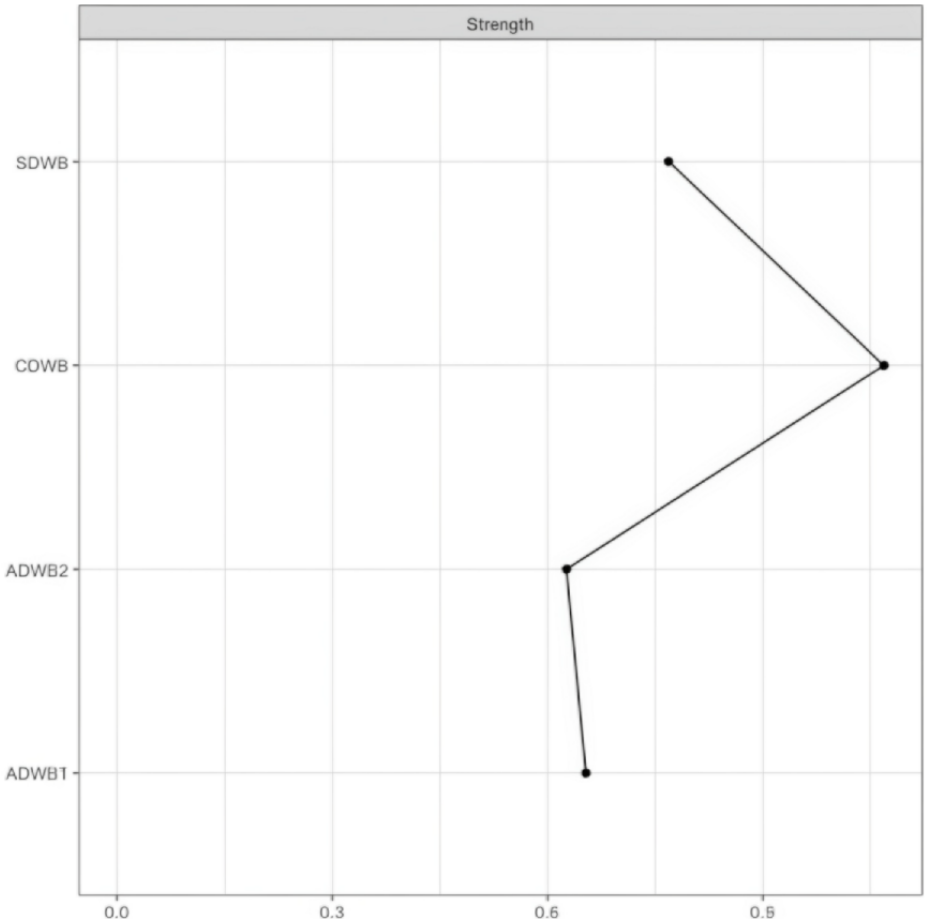

XSL•FO
RenderX

Network Stability

The correlation stability coefficient of strength centrality with the case-dropping subset bootstrapping method was 0.67, indicating that the centrality estimates were stable. Bootstrapping results for the estimated edge weights also provided support for the accuracy of edge weights in the resulting network (Multimedia Appendix 2).

Network Structure of the DWB and Related Factors

Network Estimation

Figure 3 presents the estimated structure of the DWB-related network. The spin glass community analysis detected 2 clusters. Consistent with the DWB network, the DWB factors were split into a *digital competency* (ie, healthy and effective use of technologies) community and a *digital dependency* (ie, harmful reliance on technologies) community. *Digital competency* included all DWB factors except digital stress (ADWB1) and affective and cognitive protective factors, including emotional regulation (AC1), digital literacy (CC1), and digital self-control (CC2). By contrast, negative affective DWB (ie, digital stress) was grouped into *digital dependency* with risk factors, including nomophobia (AC2), problematic internet use (BC1), problem-focused coping (BC2), avoidant coping (BC4), and the measure of participants' past experiences with involuntary online risk exposure (BC5).

Figure 3. The digital well-being–related network structure (N=578). This figure illustrates the network structure connecting dimensions of digital well-being and associated protective and risk variables derived from a network analysis conducted on survey data collected between June 2024 and July 2024. The nodes represent specific variables, with orange nodes indicating the digital competency cluster and blue nodes representing the digital dependency cluster. Edges between nodes represent associations, with blue edges indicating positive relationships and red edges indicating negative relationships. The edges' thickness corresponds to the associations' strength, with thicker edges denoting stronger connections. AC2: nomophobia; ADWB1: digital stress; BC1: problematic internet use; BC2: problem-focused coping; BC3: emotion-focused coping; BC4: avoidant coping; BC5: web-based risk exposure; CC1: digital literacy; CC2: digital self-control.

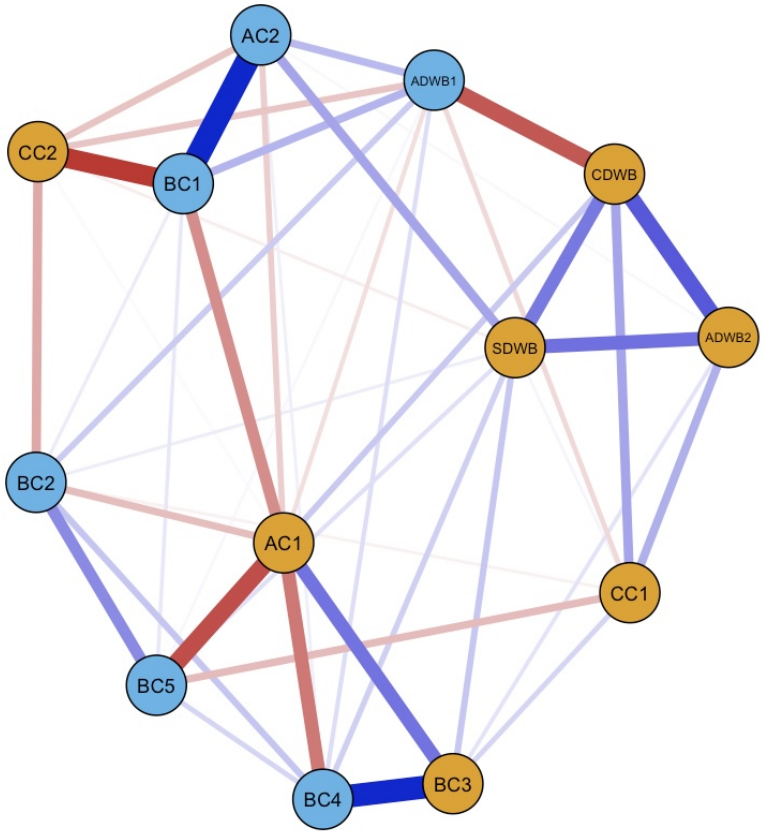

Strength Centrality

Figures 4 and 5 present the strength centrality and the BEI measures for all DWB and related protective and risk factors. The nodes with the highest strength centrality indices were AC1, BC1, and CDWB. The most central nodes for each community were then examined. For *digital competency*, the most central factors were AC1 and CDWB. For *digital dependency*, BC1 and BC4 were identified with the highest centrality index.

XSL•FO
RenderX

Figure 4. The strength centrality plot of the digital well-being–related network (N=578). The figure displays the strength centrality indices of each node derived from a network analysis conducted on survey data collected between June 2024 and July 2024. The x-axis represents the relative centrality strength values, with higher values indicating more central roles of the associated variables in the network. AC1: emotional regulation; AC2: nomophobia; ADWB1: digital stress; ADWB2: web-based hedonic well-being; BC1: problematic internet use; BC2: problem-focused coping; BC3: emotion-focused coping; BC4: avoidant coping; BC5: web-based risk exposure; CC1: digital literacy; CC2: digital self-control; CDWB: intrinsic needs satisfaction; SDWB: social digital well-being.

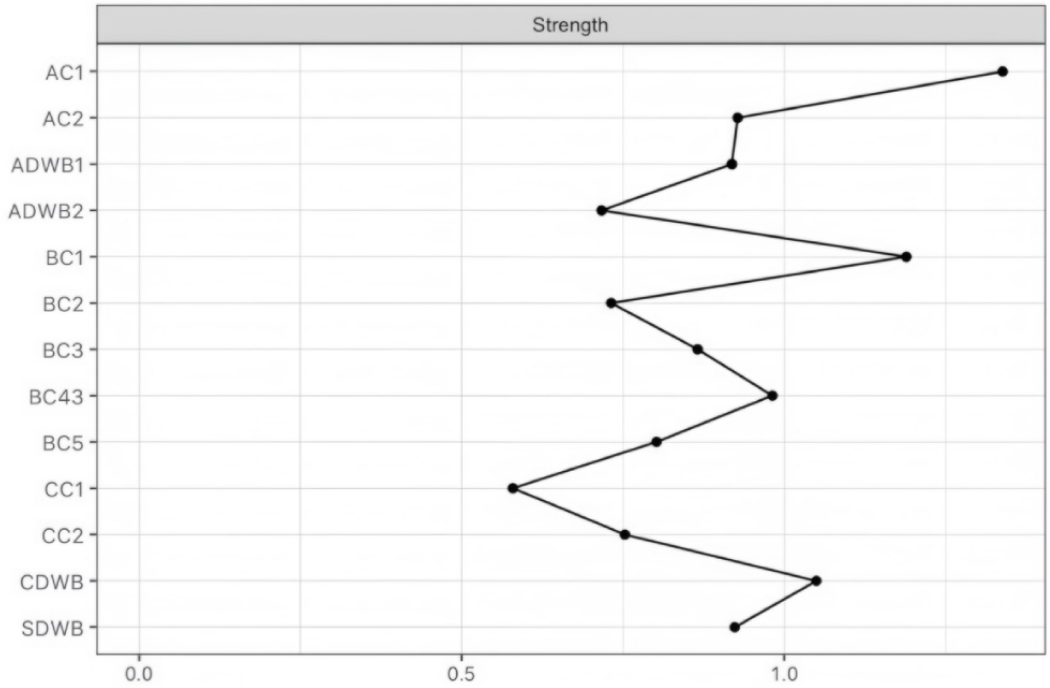

Figure 5. The bridge expected influence of the digital well-being–related network (N=578). The figure illustrates the bridge expected influence (1-step) values for each node in the digital well-being–related network derived from a network analysis conducted on survey data collected between June 2024 and July 2024. The y-axis lists the variables in the network, and the x-axis represents the relative bridge expected influence values. Higher values reflect nodes that act as critical bridging variables, connecting the digital competency and digital dependency clusters. AC1: emotional regulation; AC2: nomophobia; ADWB1: digital stress; ADWB2: web-based hedonic well-being; BC1: problematic internet use; BC2: problem−focused coping; BC3: emotion−focused coping; BC4: avoidant coping; BC5: web-based risk exposure; CC1: digital literacy; CC2: digital self-control; CDWB: intrinsic needs satisfaction; SDWB: social digital well-being.

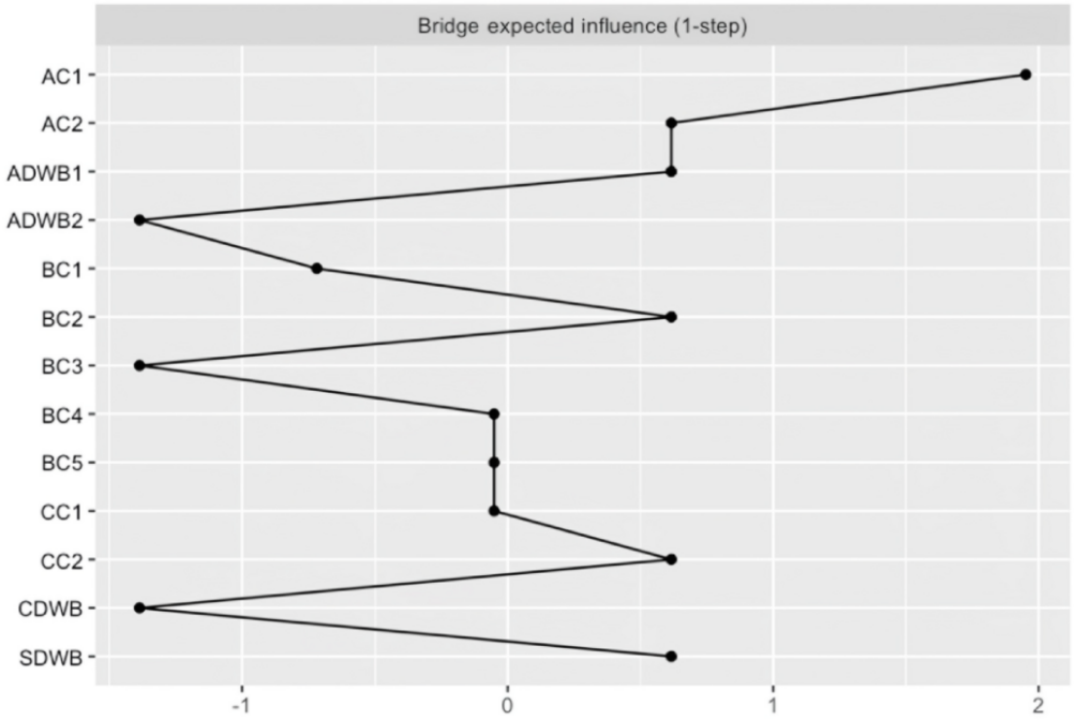

XSL•FO
RenderX

Intercluster Connections

The importance of each node in connecting the identified communities within the network was examined with the BEI, as illustrated in Figure 5. The node with the highest BEI index was AC1, which played an important role in connecting the *digital dependency* and the *digital competency* cluster. Specifically, greater emotional regulation was positively associated with BC3 in the *digital competency* cluster. By contrast, greater emotional regulation was associated with lower BC5 and less BC4 from the *digital dependency* cluster.

The strongest positive links appeared between AC2 and BC1 in the *digital dependency* cluster and between CDWB and ADWB2 in the *digital competency* cluster. The strongest negative links were between CC2 and BC1, between AC1 and BC5, and between CDWB and ADWB1.

Network Stability

The correlation stability coefficients for strength centrality and BEI were both 0.75, implying a high stability of centrality and BEI indices in the generated DWB-related network. Bootstrapping results for the estimated edge weights also indicated accurate edge weights of the resulting network (Multimedia Appendix 3).

Discussion

This study adopted a system-based framework to yield findings for conceptualizing the construct of DWB [51,64].

Dimensions of DWB

The affective dimension comprises 2 relatively independent components: digital stress and web-based hedonic well-being (Figure 1). This finding contributes to the longstanding debates in affective well-being literature regarding whether positive and negative affect represent bipolar ends of a single construct or operate as independent dimensions [77-79]. Recent evidence supports that interactions between seemingly opposite affective dimensions can be context dependent [79]. In some contexts, people might experience both types of affect simultaneously (eg, during bittersweet experiences), while in other contexts, the presence of one might diminish the other [79].

Building on this line of work, our findings support the independent view of the affective well-being debate that digital stress and hedonic well-being appear relatively independent in digital contexts. The concept of a “stress paradox” on social network sites, where users simultaneously experience enjoyment and stress from factors such as social comparison and information overload, further speaks to our finding [80]. Together, these insights highlight the importance of examining the positive and negative affective dimensions independently, as they may coexist and impact DWB in unique ways.

Indeed, the positive and negative affective DWB showed opposite associations with the most central component of DWB, namely, cognitive DWB. Precisely, online intrinsic needs satisfaction was positively correlated with hedonic well-being and negatively correlated with digital stress. These 2 associations also represented the strongest edges within the DWB network. Such observations align with the self-determination theory, which posits that optimal well-being results from the fulfillment of individual psychological needs for competence, autonomy, and relatedness [81-83]. This study further distinguishes the cognitive needs for competence and autonomy, which directly address the eudaimonic benefits of digital connectivity, from the social needs for relatedness. Consequently, the findings resonate with psychological well-being literature, underscoring that digital eudaimonic experiences (ie, using digital devices to feel competent and autonomous) could play a crucial role in cultivating healthy digital relationships and minimizing potential harm from digital use [83,84].

The social dimension of DWB, encompassing digital empathy and online social connectedness, emerged as the second most central node in the network. It showed positive relations with all the other variables, albeit less strongly with digital stress, which was more closely related to the lack of perceived autonomy and competence. Research has shown that digital platforms not only allow people to extend their social relations but also provide avenues to maintain existing social networks [85,86]. Hence, our findings on the positive associations between the social dimension and the positive affective and cognitive components of DWB emphasize the pivotal role played by online experiences of empathy and connectedness in cultivating a satisfying and pleasurable digital relationship.

These findings have practical implications for intervention designs to leverage digital hedonic and eudaimonic experiences, an area of growing interest with artificial intelligence (AI) advancements. For instance, DWB programs can incorporate AI-driven tools to facilitate online intrinsic needs satisfaction with goal-tracking interfaces and adaptive feedback systems. These features can provide users with a tailored online experience and allow for monitoring of one’s progress, thus satisfying individual needs for competence and autonomy [87]. Moreover, AI-powered interventions could strengthen digital empathy and connectedness by analyzing and responding to online interactions to promote interpersonal understanding, which used to be a key challenge in online interactions due to the lack of nonverbal cues [88]. For example, social platforms could benefit from their large database to train AI to identify opportunities for meaningful engagement, such as recommending empathetic responses or fostering inclusive discussions.

Despite the benefits, caution should also be practiced in designing interventions that incorporate AI-driven tools, as the effectiveness of human-AI collaboration depends heavily on the relative strengths of each contributor [89]. While AI can greatly enhance users’ content creation and analytics strength by automating repetitive tasks, relying on AI for decision-making tasks may undermine human judgment due to the lack of contextual understanding. Therefore, interventions must carefully consider the context in which AI tools are deployed, ensuring that they complement rather than replace human agency. The importance of this balance is further discussed in the following section on digital competency and dependency.

XSL•FO
RenderX

Emotional Regulation in Differentiating Competency From Dependency

To further understand the protective and risk factors associated with positive and negative dimensions of DWB identified in the initial network (Figure 1), the DWB-related network (Figure 3) revealed 2 distinct communities: *digital competency* and *digital dependency*. *Digital competency* refers to the effective integration and responsible use of technologies, such as successful self-control and skillful use, that can enhance one's ability to achieve personal and professional goals while fostering optimal health outcomes [90,91]. By contrast, *digital dependency* represents a reliance on digital technologies that leads to undesirable psychological consequences, such as problematic internet use characterized by excessive or compulsive use of digital media [92,93]. These 2 distinct communities encapsulated the protective and risk-oriented dimensions of DWB, highlighting the dual nature of digital engagement, that is, its potential for both benefits and harm.

The affective aspect played a key role in fostering competency. Specifically, emotional regulation emerged as the most central actor in the network, displaying a positive link with emotion-focused coping from the *digital competency* cluster and a negative link with avoidant coping from the *digital dependency* cluster. This pattern aligns with previous coping literature that revealed a positive connection between successful emotional regulation and emotion approach coping [94]. Emotion-focused coping involves the conscious regulation of emotions to alleviate distress without necessarily resolving the underlying issue [94]. Conversely, avoidant coping, which emerged as the second most central node in the *digital dependency* cluster, involves evasion of both the stressor and the emotional response. It often consists of suppression, distraction, and withdrawal from one's negative emotions, which leads to reduced well-being outcomes [95,96].

In the context of digital connectivity, a common motivation for digital use is to escape from boredom or real-life adversities, which is generally considered negative media consumption [97,98]. However, recent research on media escapism demonstrated that emotion approach coping coupled with cognitive control and self-efficacy can result in positive well-being outcomes [99,100]. Our findings further support this finding by emphasizing the role of emotional regulation in bridging the affective and behavioral coping aspects of DWB, thereby differentiating *digital competency* from *digital dependency*.

Within the *digital competency* cluster, the strongest positive association was between the satisfaction of individual needs for autonomy and competence and hedonic well-being. This finding aligns with the definition of DWB by Vanden Abeele [1] as an experience culminating in "controlled pleasure." The issue of user autonomy has gained increasing significance. Amid the evolution of intricate algorithms, user autonomy is constantly shaped, constrained, and enabled by different dimensions of human-algorithm relations [101]. Consequently, individual decision-making, actions, and consequent well-being outcomes are guided by their knowledge, perception, and critical evaluation of omnipresent algorithmic outputs [101]. A recent study examining user well-being in online retail settings showed that perceived individual autonomy can predict hedonic enjoyment when engaging in both personalized purchase experiences and general browsing [102], supporting the observed link between autonomy need satisfaction and hedonic well-being.

In addition to digital autonomy, digital competence is another key conceptual component of DWB (eg, the European Commission's Digital Competence Framework) [103]. Notably, research conducted during the COVID-19 pandemic revealed digital competence as a significant protective factor for general well-being [104]. Recent studies further advocated the benefits of digital competence across diverse age groups [105-108]. These findings underscore the importance of fulfilling individuals' competence needs in both digital and offline spaces to successfully navigate their daily lives with pervasive technology presence.

The strongest positive association in the *digital dependency* cluster was between nomophobia and problematic internet use, both ascribing to an unhealthy attachment to digital connectivity. This finding aligns with DWB studies that use the DA framework [109,110]. One possible explanation for this positive association could be their shared dependence on mobile devices for connectivity, emotional support, social interaction, and information. Both problematic internet use and nomophobia involve the fear of disconnection and an unhealthy reliance on the internet, creating a cycle of compulsive phone and internet use.

Interestingly, this study revealed a stronger negative link between digital self-control and problematic internet use, while nomophobia displayed a positive, albeit weaker, association with social DWB from the digital competency cluster. The former finding is consistent with most DA literature that identifies self-control failures as the primary predictor for problematic internet use [32,40,111]. By contrast, the observed positive relationship between nomophobia and social well-being points to the need for a critical examination of frequent digital use. Research indicates that nomophobia is often linked to increased smartphone use [39,112]. As tools for social connection, smartphones enable individuals to maintain constant contact with their social networks, both on an interpersonal and societal level. This reliance on smartphones for communication may enhance empathy by fostering emotional exchanges and social feedback loops. Consequently, individuals higher in nomophobia also obtain more opportunities to experience and respond to others' emotional states, thus reinforcing their sense of empathy through online engagement.

However, metacognitive evidence suggests that nomophobia may be related to difficulties in recognizing and expressing emotions and reduced face-to-face empathy [113-115]. Thus, while increased smartphone use may contribute to the observed positive association between nomophobia and social DWB, further studies are needed to understand how experiences of empathy and social connectedness differ in online and offline spaces.

Earlier, we discussed fostering DWB by satisfying users' needs for competence and autonomy using AI-driven tools. However,

XSL•FO
RenderX

there is a risk of AI technologies taking over users' agency, thus leading to dependency. Here, the findings of the DWB-related network highlight the roles of emotional regulation and adaptive coping in distinguishing competency from dependency, thus offering several practical avenues for interventions aimed at reducing excessive use of technologies.

Meta-analytical findings support the effectiveness of digital gaming interventions in fostering adaptive emotional regulation [116,117]. Thus, integrating similar gamified features within digital social platforms could encourage users to identify and regulate their emotions during online interactions. For instance, social media platforms could introduce features such as interactive modules consisting of scenarios (eg, receiving negative feedback on social media), where users are guided through emotional regulation steps to interpret the situation and flexibly cope with the digital stressors presented to them.

In addition to general programs designed to cultivate individuals' emotional regulation skills, tailored interventions can be implemented on digital platforms via emotion-sensitive algorithms to further enhance users' abilities to regulate and cope with potential stressors arising from digital engagement. Excessive screen time or patterns of negative interactions, for example, can be identified to trigger break reminders or suggestions for positive content [118]. Such strategies will not only allow individuals to regulate their emotions but can also give rise to optimal DWB based on our findings.

Gender and Age Differences

In addition to network findings on DWB and its related factors, statistically significant differences were found across age groups, suggesting that younger individuals may be more susceptible to unhealthy online behaviors, including problematic network use, difficulties in maintaining digital self-control, and heightened vulnerability to online risks [44,119]. These findings underscore the need for early onset of DWB interventions. School-based digital literacy programs and interactive workshops can teach students how to navigate digital spaces responsibly, develop appropriate self-control strategies, and critically assess online content. Such interventions should also be tailored to address the developmental needs of adolescents and young adults, equipping them with the tools to balance digital engagement with offline activities and develop healthier long-term digital habits.

Consistent with previous literature on digital literacy, men scored significantly higher than women on digital literacy, implying potential barriers that may prevent women from fully capitalizing on the advantages of digital technologies [120]. Conversely, women scored higher than men on nomophobia, which suggested escalating health concerns faced by women [121]. Interventions to bridge this gap could include gender-focused digital education programs and mentorship initiatives aimed at empowering women in digital spaces, emphasizing skill development in areas such as cybersecurity, data literacy, and content creation. In addition, tailored resources can be created to address the specific challenges women face, such as overreliance on digital connectivity and emotional susceptibility to online interactions.

Limitations and Future Directions

Several limitations should be considered when interpreting the results of this study. One limitation is that the network analysis adopted a cross-sectional design, which presents methodological issues regarding the directionality and temporality of the results. Thus, longitudinal studies with multiple time points are needed to examine temporal changes in the variables included in the networks, as well as how the current networks relate to long-term well-being changes.

Another limitation is that the DWB networks mainly investigated psychological factors at the individual level. Given that DWB is a dynamic construct that emerges from the interplay between users and external digital factors, the influence of external factors should also be investigated [1,17]. For instance, algorithms often prioritize content that maximizes engagement with emotionally charged and polarizing posts, which can inadvertently amplify stress, anxiety, and feelings of inadequacy among users [122]. Similarly, interface nudges, such as the infinite scrolling design of TikTok, can prolong user engagement, making it harder for individuals to disengage and manage their screen time effectively [123]. To capture these dynamics, future efforts should broaden the current conceptualization of DWB by scrutinizing the interactions between psychological and external digital factors, such as unveiling the mechanisms underlying media affordances that serve to fulfill diverse motivational needs among users [35].

In addition to the study design, the sample of this study comprising adults from the United States and the United Kingdom raised issues on the generalizability of findings. The 2023 United Nations' report on information society highlighted a growing gap between digital coverage and technological development between low-income (eg, South Africa) and high-income (eg, North America and Northern Europe) countries [124]. This disparity underscores the importance of contextualizing digital inequality, as it can significantly influence individuals' digital health. For example, residents from countries with limited internet infrastructure may lack access to digital literacy initiatives, which are vital for fostering meaningful online experiences. Besides digital inequality, cultural differences in technology adoption can lead to varying effects of digital use on social well-being outcomes [125,126]. For example, higher self-disclosure has been associated with greater perceived social support and connection in Western cultures, where openness and self-promotion are valued [127]. In contrast, in cultures that prioritize modesty, increased online self-disclosure may lead to discomfort or stress in navigating different digital norms [128]. Therefore, our findings consisting mainly of White participants from Western high-income countries may not fully apply to countries with differing digital landscapes. More large-scale multinational studies on DWB should be conducted to increase the global applicability of our findings.

Conclusions

Despite certain limitations, to the best of our knowledge, this is the first study to empirically conceptualize DWB and its protective and risk factors through a system-based approach. This research found that younger individuals face greater risks

XSL•FO
RenderX

of problematic online behaviors, while gender differences in digital literacy and nomophobia indicate potential obstacles and health risks associated with digital use.

The conceptual network findings on various DWB factors suggest that digital stress and hedonic well-being are independent constructs of the affective dimension of DWB. Consistent with SWB and psychological well-being literature, the most central aspect of optimal DWB concerns fulfilling individuals' psychological needs for competence and autonomy through digital eudaimonic benefits. By contrast, the hedonic experience was closely related to the social dimension of DWB. The network results of DWB and its related factors showed that emotional regulation and adaptive coping are key in distinguishing between digital competency and dependency, emphasizing the role of emotional regulation and cognitive control in fostering positive digital experiences. These insights highlight the importance of promoting autonomous and skilled digital use across demographics to ensure healthier digital engagement.

Acknowledgments

This work was funded by the Hong Kong Research Grants Council's General Research Fund (17601420) and the University of Hong Kong's Seed Fund for Basic Research (106200251) to CC, and grants from the Research Council of Norway to OVE (288083, 314843). No generative artificial intelligence was used in the conceptualization, design, data analysis, or writing of this manuscript, except for the use of Grammarly for grammar checking.

Data Availability

The datasets generated and analyzed during this study are available on the Open Science Framework repository [62]. This repository includes the anonymized dataset included in the analysis. For further information regarding the data and statistical analyses, please contact the first author (SC).

Authors' Contributions

SC and CC conceptualized the study and contributed to methodology development. SC and OVE developed the software and curated the data. Validation was conducted by SC and CC, while formal analysis and investigation were carried out by SC. Resources were provided by SC and CC. The original draft was prepared by SC, with CC and OVE contributing to review and editing. Visualization was handled by SC, and supervision was conducted by CC. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Nonparametric transformation for the skewed data for digital well-being and related variables.
[PNG File , 383 KB - jmir_v27i1e70483_app1.png]

Multimedia Appendix 2
Bootstrapping results for the estimated edge weights, edge difference, and node strength difference for the digital well-being network.
[PNG File , 33 KB - jmir_v27i1e70483_app2.png]

Multimedia Appendix 3
Bootstrapping results for the estimated edge weights, edge difference, and node strength difference for the digital well-being–related network.
[PNG File , 125 KB - jmir_v27i1e70483_app3.png]

References

1. Vanden Abeele MM. Digital wellbeing as a dynamic construct. Commun Theory 2021;31(4):932-955. [doi: 10.1093/ct/qtaa024]
2. 2024 global digital report: insights into social media, internet, and mobile usage worldwide. Meltwater. 2024. URL: https://www.meltwater.com/en/global-digital-trends [accessed 2024-04-29]
3. Allcott H, Gentzkow M, Song L. Digital addiction. Am Econ Rev 2022 Jul 01;112(7):2424-2463. [doi: 10.1257/aer.20210867]
4. Błachnio A, Przepiorka A, Cudo A, Angeluci A, Ben-Ezra M, Durak M, et al. Self-control and digital media addiction: the mediating role of media multitasking and time style. Psychol Res Behav Manag 2023 Jun;16:2283-2296 [FREE Full text] [doi: 10.2147/PRBM.S408993] [Medline: 37366482]

XSL•FO
RenderX

5. Chen Y, Li M, Guo F, Wang X. The effect of short-form video addiction on users' attention. Behav Inf Technol 2022 Dec 06;42(16):2893-2910. [doi: 10.1080/0144929X.2022.2151512]
6. Király O, Griffiths MD, Demetrovics Z. Internet gaming disorder and the DSM-5: conceptualization, debates, and controversies. Curr Addict Rep 2015 Sep 01;2(3):254-262. [doi: 10.1007/S40429-015-0066-7]
7. Sun Y, Zhang Y. A review of theories and models applied in studies of social media addiction and implications for future research. Addict Behav 2021 Mar;114:106699. [doi: 10.1016/j.addbeh.2020.106699] [Medline: 33268185]
8. Caplan SE. Problematic Internet use and psychosocial well-being: development of a theory-based cognitive–behavioral measurement instrument. Comput Human Behav 2002 Sep;18(5):553-575. [doi: 10.1016/S0747-5632(02)00004-3]
9. Cemiloglu D, Almourad MB, McAlaney J, Ali R. Combatting digital addiction: current approaches and future directions. Technol Soc 2022 Feb;68:101832. [doi: 10.1016/j.techsoc.2021.101832]
10. Abdoli N, Sadeghi-Bahmani D, Salari N, Khodamoradi M, Farnia V, Jahangiri S, et al. Nomophobia (no mobile phone phobia) and psychological health issues among young adult students. Eur J Investig Health Psychol Educ 2023 Sep 12;13(9):1762-1775 [FREE Full text] [doi: 10.3390/ejihpe13090128] [Medline: 37754467]
11. Rodríguez-García AM, Moreno-Guerrero AJ, López Belmonte J. Nomophobia: an individual's growing fear of being without a smartphone-a systematic literature review. Int J Environ Res Public Health 2020 Jan 16;17(2):580 [FREE Full text] [doi: 10.3390/ijerph17020580] [Medline: 31963208]
12. Tang MC, Ebrahimi OV, Cheng C. Symptomatology, risk, and protective factors of gaming disorder: a network analytical approach. Comput Human Behav 2023 Nov;148:107899. [doi: 10.1016/j.chb.2023.107899]
13. Deci EL, Ryan RM. Hedonia, eudaimonia, and well-being: an introduction. J Happiness Stud 2006 Nov 18;9(1):1-11. [doi: 10.1007/S10902-006-9018-1]
14. Meier A. Studying problems, not problematic usage: do mobile checking habits increase procrastination and decrease well-being? Mob Media Commun 2021 Aug 04;10(2):272-293. [doi: 10.1177/20501579211029326]
15. van Rooij AJ, Ferguson CJ, Colder Carras M, Kardefelt-Winther D, Shi J, Aarseth E, et al. A weak scientific basis for gaming disorder: let us err on the side of caution. J Behav Addict 2018 Mar 01;7(1):1-9 [FREE Full text] [doi: 10.1556/2006.7.2018.19] [Medline: 29529886]
16. Reinecke L, Vorderer P, Knop K. Entertainment 2.0? The role of intrinsic and extrinsic need satisfaction for the enjoyment of Facebook use. J Commun 2014 May 19;64(3):417-438. [doi: 10.1111/jcom.12099]
17. Büchi M. Digital well-being theory and research. New Media Soc 2021 Nov 14;26(1):172-189. [doi: 10.1177/14614448211056851]
18. Lukoff K, Yu C, Kientz J, Hiniker A. What makes smartphone use meaningful or meaningless? Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-26. [doi: 10.1145/3191754]
19. Bauer AA, Loy LS, Masur PK, Schneider FM. Mindful instant messaging. J Media Psychol 2017 Jul;29(3):159-165. [doi: 10.1027/1864-1105/a000225]
20. Nieboer A, Lindenberg S, Boomsma A, Bruggen AC. Dimensions of well-being and their measurement: the SPF-IL scale. Soc Indic Res 2005 Sep;73(3):313-353. [doi: 10.1007/S11205-004-0988-2]
21. Diener E. Subjective well-being. Psychol Bull 1984;95(3):542-575. [doi: 10.1037/0033-2909.95.3.542]
22. Steele RG, Hall JA, Christofferson JL. Conceptualizing digital stress in adolescents and young adults: toward the development of an empirically based model. Clin Child Fam Psychol Rev 2020 Mar 7;23(1):15-26. [doi: 10.1007/s10567-019-00300-5] [Medline: 31392451]
23. Qi C, Yang N. Digital resilience and technological stress in adolescents: a mixed-methods study of factors and interventions. Educ Inf Technol 2024 Mar 18;29(14):19067-19113. [doi: 10.1007/S10639-024-12595-1]
24. Hefner D, Freytag A. Consciously connected: the role of mindfulness for mobile phone connectedness and stress. Media Psychol 2023 Sep 05;27(4):503-532. [doi: 10.1080/15213269.2023.2253732]
25. van Koningsbruggen GM, Hartmann T, Eden A, Veling H. Spontaneous hedonic reactions to social media cues. Cyberpsychol Behav Soc Netw 2017 May;20(5):334-340. [doi: 10.1089/cyber.2016.0530] [Medline: 28422519]
26. James C, Davis K, Charmaraman L, Konrath S, Slovak P, Weinstein E, et al. Digital life and youth well-being, social connectedness, empathy, and narcissism. Pediatrics 2017 Nov 01;140(Supplement 2):S71-S75. [doi: 10.1542/peds.2016-1758f]
27. Soundararajan A, Lim JX, Ngiam NH, Tey AJ, Tang AK, Lim HA, et al. Smartphone ownership, digital literacy, and the mediating role of social connectedness and loneliness in improving the wellbeing of community-dwelling older adults of low socio-economic status in Singapore. PLoS One 2023 Aug 30;18(8):e0290557 [FREE Full text] [doi: 10.1371/journal.pone.0290557] [Medline: 37647263]
28. Riedl C, Köbler F, Goswami S, Krcmar H. Tweeting to feel connected: a model for social connectedness in online social networks. Int J Hum Comput Interact 2013 Oct 03;29(10):670-687. [doi: 10.1080/10447318.2013.768137]
29. Hasell A, Nabi RL. Emotion, information sharing, and social media. In: Nabi RL, Myrick JG, editors. Emotions in the Digital World: Exploring Affective Experience and Expression in Online Interactions. Oxford, UK: Oxford Academic Press; 2023:381-400.
30. Albashrawi M, Yu J, Binsawad M, Asiri Y. Moving to digital-healthy society: empathy, sympathy, and wellbeing in social media. J Assoc Inf Syst 2022 Jan;14:71-89. [doi: 10.17705/1pais.14206]

XSL•FO
RenderX

31. Gainsburg I, Earl A. Trigger warnings as an interpersonal emotion-regulation tool: avoidance, attention, and affect depend on beliefs. J Exp Soc Psychol 2018 Nov;79:252-263. [doi: 10.1016/j.jesp.2018.08.006]
32. Roffarello AM, De Russis L. Achieving digital wellbeing through digital self-control tools: a systematic review and meta-analysis. ACM Trans Comput Hum Interact 2023 Sep 12;30(4):1-66. [doi: 10.1145/3571810]
33. Shin M, Downes C, Hopwood J, Byers M, Kemps E. Media multitasking, negative mood, and avoidance coping. Behav Inf Technol 2024 May 13:1-11. [doi: 10.1080/0144929X.2024.2353276]
34. Wadley G, Smith W, Koval P, Gross JJ. Digital emotion regulation. Curr Dir Psychol Sci 2020 Jun 18;29(4):412-418. [doi: 10.1177/0963721420920592]
35. Steinert S, Dennis MJ. Emotions and digital well-being: on social media's emotional affordances. Philos Technol 2022 Apr 13;35(2):36 [FREE Full text] [doi: 10.1007/s13347-022-00530-6] [Medline: 35450167]
36. Díaz-Moreno A, Bonilla I, Chamarro A. Negative social comparison: the influence of anxiety, emotional regulation and problematic use of social media. Anxiety Stress 2023;29(3):181-186 [FREE Full text] [doi: 10.5093/anyes2023a22]
37. Fortes AB, Broilo PL, Lisboa CS. Smartphone use and psychological well-being: the moderating role of emotion regulation. Trends in Psychol 2020 Nov 11;29(2):189-203. [doi: 10.1007/S43076-020-00051-1]
38. Yue Z, Zhang R, Xiao J. Passive social media use and psychological well-being during the COVID-19 pandemic: the role of social comparison and emotion regulation. Comput Human Behav 2022 Feb;127:107050 [FREE Full text] [doi: 10.1016/j.chb.2021.107050] [Medline: 34646057]
39. Safaria T, Saputra NE, Arini DP. The impact of nomophobia: exploring the interplay between loneliness, smartphone usage, self-control, emotion regulation, and spiritual meaningfulness in an Indonesian context. J Technol Behav Sci 2024 Sep 11:27. [doi: 10.1007/S41347-024-00438-2]
40. Du J, Kerkhof P, van Koningsbruggen GM. The reciprocal relationships between social media self-control failure, mindfulness and wellbeing: a longitudinal study. PLoS One 2021 Aug 4;16(8):e0255648 [FREE Full text] [doi: 10.1371/journal.pone.0255648] [Medline: 34347832]
41. Reinecke L, Gilbert A, Eden A. Self-regulation as a key boundary condition in the relationship between social media use and well-being. Curr Opin Psychol 2022 Jun;45:101296. [doi: 10.1016/j.copsyc.2021.12.008] [Medline: 35085935]
42. Chiu TK, Sun JC, Ismailov M. Investigating the relationship of technology learning support to digital literacy from the perspective of self-determination theory. Educ Psychol 2022 May 24;42(10):1263-1282. [doi: 10.1080/01443410.2022.2074966]
43. Gui M, Fasoli M, Carradore R. Digital well-being: developing a new theoretical tool for media literacy research. Ital J Sociol Educ 2017;9(1):155-173. [doi: 10.14658/pupj-ijse-2017-1-8]
44. Benzi IM, Fontana A, Lingiardi V, Parolin L, Carone N. “Don’t leave me behind!” problematic internet use and fear of missing out through the lens of epistemic trust in emerging adulthood. Curr Psychol 2023 Nov 18;43(15):13775-13784. [doi: 10.1007/S12144-023-05440-0]
45. Agha Z, Badillo-Urquiola K, Wisniewski PJ. "Strike at the root": co-designing real-time social media interventions for adolescent online risk prevention. Proc ACM Hum Comput Interact 2023 Apr 16;7(CSCW1):1-32. [doi: 10.1145/3579625]
46. McHugh BC, Wisniewski P, Rosson MB, Carroll JM. When social media traumatizes teens: the roles of online risk exposure, coping, and post-traumatic stress. Internet Res 2018 Oct 02;28(5):1169-1188. [doi: 10.1108/intr-02-2017-0077]
47. Augner C, Vlasak T, Aichhorn W, Barth A. The association between problematic smartphone use and symptoms of anxiety and depression-a meta-analysis. J Public Health (Oxf) 2023 Mar 14;45(1):193-201. [doi: 10.1093/pubmed/fdab350] [Medline: 34585243]
48. Cai Z, Mao P, Wang Z, Wang D, He J, Fan X. Associations between problematic internet use and mental health outcomes of students: a meta-analytic review. Adolesc Res Rev 2023;8(1):45-62 [FREE Full text] [doi: 10.1007/s40894-022-00201-9] [Medline: 36744121]
49. Gupta M. Understanding social media users' perceptions of trigger and content warnings. Virginia Tech. 2023. URL: http://hdl.handle.net/10919/116505 [accessed 2024-11-22]
50. Melodia F, Canale N, Griffiths MD. The role of avoidance coping and escape motives in problematic online gaming: a systematic literature review. Int J Ment Health Addiction 2020 Nov 30;20(2):996-1022. [doi: 10.1007/s11469-020-00422-w]
51. Ebrahimi OV. Systems-based thinking in psychology and the mental health sciences. Nat Rev Psychol 2023 May 10;2(6):332 [FREE Full text] [doi: 10.1038/s44159-023-00193-w] [Medline: 37304192]
52. Briganti G, Scutari M, Epskamp S, Borsboom D, Hoekstra RH, Golino HF, et al. Network analysis: an overview for mental health research. Int J Methods Psychiatr Res 2024 Dec 14;33(4):e2034. [doi: 10.1002/mpr.2034] [Medline: 39543824]
53. Peer E, Rothschild D, Gordon A, Evernden Z, Damer E. Data quality of platforms and panels for online behavioral research. Behav Res Methods 2022 Aug 29;54(4):1643-1662 [FREE Full text] [doi: 10.3758/s13428-021-01694-3] [Medline: 34590289]
54. Hall JA, Baym NK. Calling and texting (too much): mobile maintenance expectations, (over)dependence, entrapment, and friendship satisfaction. New Media Soc 2011 Aug 24;14(2):316-331. [doi: 10.1177/1461444811415047]
55. Nambisan S, Baron RA. Interactions in virtual customer environments: implications for product support and customer relationship management. J Interact Mark 2007 May 01;21(2):42-62. [doi: 10.1002/dir.20077]

XSL•FO
RenderX

56. Powell PA, Roberts J. Situational determinants of cognitive, affective, and compassionate empathy in naturalistic digital interactions. Comput Human Behav 2017 Mar;68:137-148. [doi: 10.1016/j.chb.2016.11.024]
57. Bjureberg J, Ljótsson B, Tull MT, Hedman E, Sahlin H, Lundh L, et al. Development and validation of a brief version of the difficulties in emotion regulation scale: the DERS-16. J Psychopathol Behav Assess 2016 Jun 14;38(2):284-296 [FREE Full text] [doi: 10.1007/s10862-015-9514-x] [Medline: 27239096]
58. Yildirim C, Correia AP. Exploring the dimensions of nomophobia: development and validation of a self-reported questionnaire. Comput Human Behav 2015 Aug;49:130-137. [doi: 10.1016/j.chb.2015.02.059]
59. van Deursen AJ, van Dijk JA. Toward a multifaceted model of internet access for understanding digital divides: an empirical investigation. Inf Soc 2015 Sep 11;31(5):379-391. [doi: 10.1080/01972243.2015.1069770]
60. Du J, van Koningsbruggen GM, Kerkhof P. A brief measure of social media self-control failure. Comput Human Behav 2018 Jul;84:68-75. [doi: 10.1016/j.chb.2018.02.002]
61. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
62. Unveiling the conceptual framework of digital well-being: a network analysis approach. Open Science Framework. URL: https://osf.io/j8t9z/?view_only=c17c559cfb344989a1aa52f15739e8e3 [accessed 2024-04-29]
63. Jiang H, Fei X, Liu H, Roeder K, Lafferty J, Wasserman L, et al. Package 'huge': high-dimensional undirected graph estimation (R package version 1.3.5). Cran R. URL: https://cran.r-project.org/web/packages/huge/huge.pdf [accessed 2024-04-29]
64. Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. Nat Rev Methods Primers 2021 Aug 19;1(1):46. [doi: 10.1038/s43586-021-00055-w]
65. Skjerdingstad N, Johnson MS, Johnson SU, Hoffart A, Ebrahimi OV. Feelings of worthlessness links depressive symptoms and parental stress: a network analysis during the COVID-19 pandemic. Eur Psychiatry 2021 Jul 27;64(1):e50 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.2223] [Medline: 34311806]
66. Jones P. Package 'networktools': tools for identifying important nodes in networks. Cran R. URL: https://cran.r-project.org/web/packages/networktools/networktools.pdf [accessed 2024-04-29]
67. Isvoranu AM, Epskamp S. Which estimation method to choose in network psychometrics? Deriving guidelines for applied researchers. Psychol Methods 2023 Aug;28(4):925-946. [doi: 10.1037/met0000439] [Medline: 34843277]
68. Epskamp S, Waldorp LJ, Mõttus R, Borsboom D. The Gaussian Graphical Model in cross-sectional and time-series data. Multivariate Behav Res 2018 Apr 16;53(4):453-480 [FREE Full text] [doi: 10.1080/00273171.2018.1454823] [Medline: 29658809]
69. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Package 'qgraph': graph plotting methods, psychometric data visualization and graphical model estimation. Cran R. URL: https://cran.r-project.org/web/packages/qgraph/qgraph.pdf [accessed 2024-04-29]
70. Isvoranu AM, Epskamp S, Waldorp L, Borsboom D. Network Psychometrics with R: A Guide for Behavioral and Social Scientists. New York, NY: Taylor & Francis; 2022.
71. Fruchterman TM, Reingold EM. Graph drawing by force ‐ directed placement. Softw Pract Exp 2006 Oct 30;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
72. Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. Sci Rep 2016 Aug 01;6:30750 [FREE Full text] [doi: 10.1038/srep30750] [Medline: 27476470]
73. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006. URL: https://cran.r-project.org/web/packages/igraph/igraph.pdf [accessed 2024-04-29]
74. Jones P, Jones MP. Package 'networktools'. CRAN-R. 2018. URL: http://cran.nexr.com/web/packages/networktools/networktools.pdf [accessed 2024-04-29]
75. Epskamp S, Fried EI. Package 'bootnet': bootstrap methods for various network estimation routines. CRAN-R. 2015. URL: https://cran.r-project.org/web/packages/bootnet/bootnet.pdf [accessed 2024-04-29]
76. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods 2018 Feb 24;50(1):195-212 [FREE Full text] [doi: 10.3758/s13428-017-0862-1] [Medline: 28342071]
77. Russell JA, Carroll JM. On the bipolarity of positive and negative affect. Psychol Bull 1999 Jan;125(1):3-30. [doi: 10.1037/0033-2909.125.1.3] [Medline: 9990843]
78. Huppert FA, Whittington JE. Evidence for the independence of positive and negative well-being: implications for quality of life assessment. Br J Health Psychol 2003 Feb 16;8(Pt 1):107-122. [doi: 10.1348/135910703762879246] [Medline: 12643820]
79. Robinson MD, Irvin RL, Persich MR, Krishnakumar S. Bipolar or independent? Relations between positive and negative affect vary by emotional intelligence. Affect Sci 2020 Dec 06;1(4):225-236 [FREE Full text] [doi: 10.1007/s42761-020-00018-6] [Medline: 36042817]
80. Cheikh-Ammar M. The bittersweet escape to information technology: an investigation of the stress paradox of social network sites. Inf Manag 2020 Dec;57(8):103368. [doi: 10.1016/j.im.2020.103368]
81. Adeel A, Batool S, Madni ZA. Intrinsic motivation and creativity: the role of digital technology and knowledge integration ability in facilitating creativity. Int J Manag Stud 2023;30(1):1-36. [doi: 10.32890/ijms2023.30.1.1]

XSL•FO
RenderX

82. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
83. Huta V. An overview of hedonic and eudaimonic well-being concepts. In: Reinecke L, Oliver MB, editors. The Routledge Handbook of Media Use and Well-Being: International Perspectives on Theory and Research on Positive Media Effects. New York, NY: Routledge; 2016:14-33.
84. Zambianchi M. Digital technologies and eudaimonic well-being in the emerging adults. Psychol Mag 2022 Dec 15;41(1):147-164. [doi: 10.18800/psico.202301.006]
85. Petersen B, Khalili-Mahani N, Murphy C, Sawchuk K, Phillips N, Li KZ, et al. The association between information and communication technologies, loneliness and social connectedness: a scoping review. Front Psychol 2023 Mar 22;14:1063146 [FREE Full text] [doi: 10.3389/fpsyg.2023.1063146] [Medline: 37034933]
86. Simons M, Reijnders J, Janssens M, Lataster J, Jacobs N. Staying connected in old age: associations between bonding social capital, loneliness and well-being and the value of digital media. Aging Ment Health 2023 Jan 14;27(1):147-155 [FREE Full text] [doi: 10.1080/13607863.2022.2036947] [Medline: 35156860]
87. Kelly CA, Sharot T. Web-browsing patterns reflect and shape mood and mental health. Nat Hum Behav 2025 Jan 21;9(1):133-146. [doi: 10.1038/s41562-024-02065-6] [Medline: 39572688]
88. Venter E. Challenges for meaningful interpersonal communication in a digital era. Theol Stud 2019 Apr 29;75(1):1-6. [doi: 10.4102/hts.v75i1.5339]
89. Vaccaro M, Almaatouq A, Malone T. When combinations of humans and AI are useful: a systematic review and meta-analysis. Nat Hum Behav 2024 Dec 28;8(12):2293-2303. [doi: 10.1038/s41562-024-02024-1] [Medline: 39468277]
90. Nyikes Z. Contemporary digital competency review. Interdiscip Descr Complex Syst 2018;16(1):124-131. [doi: 10.7906/indecs.16.1.9]
91. Karsenti T, Poellhuber B, Parent S, Michelot F. What is the Digital Competency framework? Revue Int Technol Pedagog Univer 2020;17(1):11-14. [doi: 10.18162/ritpu-2020-v17n1-04]
92. Harley D. Digital wellbeing: making sense of digital dependency. In: Harley D, editor. Mindfulness in a Digital World. Cham, Switzerland: Springer; 2022:1-23.
93. Lambert A, Rome A, Fornari F. Stayin' alive? Reflections on navigating digital dependency. Mark. Theory 2023 Nov 30;24(3):359-366. [doi: 10.1177/14705931231218102]
94. Baker JP, Berenbaum H. Emotional approach and problem-focused coping: a comparison of potentially adaptive strategies. Cogn Emot 2007 Jan;21(1):95-118. [doi: 10.1080/02699930600562276]
95. Seiffge-Krenke I, Klessinger N. Long-term effects of avoidant coping on adolescents' depressive symptoms. J Youth Adolesc 2000 Dec;29(6):617-630. [doi: 10.1023/A:1026440304695]
96. Boals A, Vandellen MR, Banks JB. The relationship between self-control and health: the mediating effect of avoidant coping. Psychol Health 2011 Aug;26(8):1049-1062. [doi: 10.1080/08870446.2010.529139] [Medline: 21598180]
97. Kardefelt-Winther D. The moderating role of psychosocial well-being on the relationship between escapism and excessive online gaming. Comput Human Behav 2014 Sep;38:68-74. [doi: 10.1016/j.chb.2014.05.020]
98. Kırcaburun K, Griffiths MD. Problematic Instagram use: the role of perceived feeling of presence and escapism. Int J Ment Health Addiction 2018 Jun 4;17(4):909-921. [doi: 10.1007/S11469-018-9895-7]
99. Stenseng F, Falch-Madsen J, Hygen BW. Are there two types of escapism? Exploring a dualistic model of escapism in digital gaming and online streaming. Psychol Pop Media 2021 Jul;10(3):319-329. [doi: 10.1037/ppm0000339]
100. Lee YH, Chen M. Seeking a sense of control or escapism? The role of video games in coping with unemployment. Games Cult 2022 May 12;18(3):339-361. [doi: 10.1177/15554120221097413]
101. Savolainen L, Ruckenstein M. Dimensions of autonomy in human–algorithm relations. New Media Soc 2022 Jun 14;26(6):3472-3490. [doi: 10.1177/14614448221100802]
102. Shen P, Wan D, Li J. How human–computer interaction perception affects consumer well-being in the context of online retail: from the perspective of autonomy. Nankai Bus Rev Int 2022 Jun 03;14(1):102-127. [doi: 10.1108/nbri-03-2022-0034]
103. Riina V, Stefano K, Yves P. DigComp 2.2: the digital competence framework for citizens - with new examples of knowledge, skills and attitudes. Joint Research Centre. URL: https://ideas.repec.org/p/ipt/iptwpa/jrc128415.html [accessed 2024-12-01]
104. Kumpikaitė-Valiūnienė V, Aslan I, Duobienė J, Glińska E, Anandkumar V. Influence of digital competence on perceived stress, burnout and well-being among students studying online during the COVID-19 lockdown: a 4-country perspective. Psychol Res Behav Manag 2021;14:1483-1498 [FREE Full text] [doi: 10.2147/PRBM.S325092] [Medline: 34588828]
105. Huu PT. Impact of employee digital competence on the relationship between digital autonomy and innovative work behavior: a systematic review. Artif Intell Rev 2023 Apr 29;56(12):1-30 [FREE Full text] [doi: 10.1007/s10462-023-10492-6] [Medline: 37362897]
106. Mungra Y, Srivastava R, Sharma A, Banerji D, Gollapudi N. Impact of digital competence on employees' flourishing through basic psychological needs satisfaction. J Comput Inf Syst 2023 Feb 15;64(1):47-61. [doi: 10.1080/08874417.2023.2176948]
107. Schirmer M, Dalko K, Stoevesandt D, Paulicke D, Jahn P. Educational concepts of digital competence development for older adults-a scoping review. Int J Environ Res Public Health 2023 Jun 30;20(13):6269 [FREE Full text] [doi: 10.3390/ijerph20136269] [Medline: 37444116]

XSL•FO
RenderX

108. Su J, Yang W. Digital competence in early childhood education: a systematic review. Educ Inf Technol 2023 Jul 11;29(4):4885-4933. [doi: 10.1007/S10639-023-11972-6]
109. Liu W, Chen JS, Gan WY, Poon WC, Tung SE, Lee LJ, et al. Associations of problematic internet use, weight-related self-stigma, and nomophobia with physical activity: findings from Mainland China, Taiwan, and Malaysia. Int J Environ Res Public Health 2022 Sep 25;19(19):12135 [FREE Full text] [doi: 10.3390/ijerph191912135] [Medline: 36231435]
110. Tung SE, Gan WY, Chen JS, Ruckwongpatr K, Pramukti I, Nadhiroh S, et al. Internet-related instruments (bergen social media addiction scale, smartphone application-based addiction scale, internet gaming disorder scale-short form, and nomophobia questionnaire) and their associations with distress among Malaysian University Students. Healthcare (Basel) 2022 Aug 02;10(8):1448 [FREE Full text] [doi: 10.3390/healthcare10081448] [Medline: 36011105]
111. Hofmann W, Reinecke L, Meier A. Of sweet temptations and bitter aftertaste: self-control as a moderator of the effects of media use on well-being. In: Reinecke L, Oliver MB, editors. The Routledge Handbook of Media Use and Well-Being: International Perspectives on Theory and Research on Positive Media Effects. New York, NY: Routledge; 2016:211-222.
112. Lin CY, Potenza MN, Ulander M, Broström A, Ohayon MM, Chattu VK, et al. Longitudinal relationships between nomophobia, addictive use of social media, and insomnia in adolescents. Healthcare (Basel) 2021 Sep 11;9(9):1201 [FREE Full text] [doi: 10.3390/healthcare9091201] [Medline: 34574975]
113. Cambier R, Derks D, Vlerick P. Detachment from work: a diary study on telepressure, smartphone use and empathy. Psychol Belg 2019 Jun 27;59(1):227-245 [FREE Full text] [doi: 10.5334/pb.477] [Medline: 31328018]
114. Pera A. The psychology of addictive smartphone behavior in young adults: problematic use, social anxiety, and depressive stress. Front Psychiatry 2020;11:573473 [FREE Full text] [doi: 10.3389/fpsyt.2020.573473] [Medline: 33101087]
115. Yavuz M, Altan B, Bayrak B, Gündüz M, Bolat N. The relationships between nomophobia, alexithymia and metacognitive problems in an adolescent population. Turk J Pediatr 2019 Jun 25;61(3):345-351 [FREE Full text] [doi: 10.24953/turkjped.2019.03.005] [Medline: 31916711]
116. Reynard S, Dias J, Mitic M, Schrank B, Woodcock KA. Digital interventions for emotion regulation in children and early adolescents: systematic review and meta-analysis. JMIR Serious Games 2022 Aug 19;10(3):e31456 [FREE Full text] [doi: 10.2196/31456] [Medline: 35984681]
117. Cheng C, Li S, Chen S. Evaluation of a multicomponent positive psychology program to prevent gaming disorder and enhance mental wellness in primary pupils: a randomized controlled trial. J Behav Addict 2024 Oct 04;13(3):871-883. [doi: 10.1556/2006.2024.00052] [Medline: 39348198]
118. Plackett R, Blyth A, Schartau P. The impact of social media use interventions on mental well-being: systematic review. J Med Internet Res 2023 Aug 11;25:e44922 [FREE Full text] [doi: 10.2196/44922] [Medline: 37565693]
119. Blachnio A, Przepiórka A, Gorbaniuk O, McNeill M, Bendayan R, Ben-Ezra M, et al. Lonely online: a social model of digital media addiction: a study in 21 countries. Soc Sci J 2024 Jan 10:1-12. [doi: 10.1080/03623319.2023.2228439]
120. Kaur M, Sharma J. The role of digital literacy to promote the gender equality. Shodh Sari 2023 Oct 05;02(04):315-327. [doi: 10.59231/SARI7642]
121. Arpaci I. Gender differences in the relationship between problematic internet use and nomophobia. Curr Psychol 2020 Nov 02;41(9):6558-6567. [doi: 10.1007/s12144-020-01160-x]
122. Milli S, Carroll M, Wang Y, Pandey S, Zhao S, Dragan AD. Engagement, user satisfaction, and the amplification of divisive content on social media. arXiv Preprint posted online May 26, 2023 [FREE Full text] [doi: 10.1093/pnasnexus/pgaf062]
123. Virós-Martín C, Montaña-Blasco M, Jiménez-Morales M. Can't stop scrolling! Adolescents' patterns of TikTok use and digital well-being self-perception. Humanit Soc Sci Commun 2024 Oct 30;11(1):1-11. [doi: 10.1057/s41599-024-03984-5]
124. Secretary-General U, Secretariat U. Progress made in the implementation of and follow-up to the outcomes of the World Summit on the Information Society at the regional and international levels. Coherent Digital. URL: https://coilink.org/20.500.12592/xsj3wfd [accessed 2024-04-29]
125. Cheng C, Li AY. Internet addiction prevalence and quality of (real) life: a meta-analysis of 31 nations across seven world regions. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):755-760 [FREE Full text] [doi: 10.1089/cyber.2014.0317] [Medline: 25489876]
126. Cheng C, Cheung MW, Montasem A, 44 members of the International Network of Well-Being Studies. Explaining differences in subjective well-being across 33 nations using multilevel models: universal personality, cultural relativity, and national income. J Pers 2016 Feb;84(1):46-58. [doi: 10.1111/jopy.12136] [Medline: 25234240]
127. Li H, Xu Y. Unraveling the cross-cultural differences in online expression of social anxiety in online support communities. Cyberpsychol Behav Soc Netw 2024 May 01;27(5):328-335. [doi: 10.1089/cyber.2023.0539] [Medline: 38526233]
128. Gao X, Takai J. Psychosocial effects of self-disclosure among Chinese international students in Japan. J Int Stud 2023 Mar 19;14(1):58-77. [doi: 10.32674/jis.v14i1.5098]

Abbreviations

AI: artificial intelligence
DA: digital addiction
GGM: graphical Gaussian model

XSL•FO
RenderX

SWB: subjective well-being

Edited by A Mavragani; submitted 23.12.24; peer-reviewed by C Nnajiofor, M Adnan; comments to author 25.01.25; revised version received 12.02.25; accepted 13.02.25; published 25.03.25.

Please cite as:
Chen S, Ebrahimi OV, Cheng C
New Perspective on Digital Well-Being by Distinguishing Digital Competency From Dependency: Network Approach
J Med Internet Res 2025;27:e70483
URL: https://www.jmir.org/2025/1/e70483
doi:10.2196/70483
PMID:

©Si Chen, Omid V Ebrahimi, Cecilia Cheng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patients' Use of e-Consultations as an Alternative to Other General Practitioner Services: Cross-Sectional Survey Study

Eli Kristiansen[1,2], MSc; Helen Atherton[3], MSc, MPH, PhD; Bjarne Austad[4], MD, PhD; Trine Strand Bergmo[2], MSc, PhD; Børge Lønnebakke Norberg[2,4], MD; Chris Salisbury[5], MSc, MD; Paolo Zanaboni[2], PhD

[1]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[2]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[3]School of Primary Care Population Health and Medical Education, University of Southampton, Southampton, United Kingdom

[4]General Practice Research Unit, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

[5]Centre for Academic Primary Care, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Eli Kristiansen, MSc
Department of Clinical Medicine
Faculty of Health Sciences
UiT The Arctic University of Norway
Hansine Hansens veg 18
Tromsø, 9019
Norway
Phone: 47 97162311
Email: eli.kristiansen@ehealthresearch.no

Abstract

Background: e-Consultations, defined as asynchronous text-based messaging, have transformed how patients interact with their general practitioner (GP). While e-consultations can improve patient access to GP care, concerns about increased workload for GPs are raised.

Objective: This study aimed to address three research questions: (1) For what purpose and with what expectations do patients initiate e-consultations? (2) If e-consultations had not been available, what alternative actions would the patient have taken? and (3) How are the alternative actions associated with patient and e-consultation characteristics?

Methods: A cross-sectional study was conducted through a web-based survey on Helsenorge. Helsenorge is the national citizen portal for digital health services in Norway, including e-consultations with the GP. All users who sent e-consultations through Helsenorge were invited to participate between January and February 2023. The survey addressed questions on users' expectations and experience with e-consultations. The association between patient and e-consultation characteristics and alternative actions to e-consultations were analyzed using multinomial logistic regression.

Results: Overall, 13,011 users answered the survey. The most common reason for initiating an e-consultation was requesting a sick certificate (4940/13,011, 38%). Overall, 68.7% (8802/13,011) of respondents expected an answer within 24 hours, and 17.7% (2310/13,011) anticipated that the GP would ask them to attend a physical examination. If e-consultations had not been available, 45.5% (5917/13,011) of respondents would have booked a GP appointment, and 44.9% (5846/13,011) would have called the front desk. Users who expected a quicker response (odds ratio [OR] 1.64, 95% CI 1.46-1.85) and were less concerned about their health issues (OR 1.29, 95% CI 1.18-1.40) were more likely to call the front desk. Only 2.5% (323/13,011) of respondents would have contacted out-of-hours services. Users with longer travel time to the GP office (OR 6.08, 95% CI 3.46-10.66) and with a new health problem (OR 2.71, 95% CI 2.09-3.51) were more likely to choose this option. In addition, 4.7% (609/13,011) of the users would not have sought help if e-consultations had not been available. Younger patients (OR 2.16, 95% CI 1.38-3.37) and those with a longer travel time to the GP office (OR 2.19, 95% CI 1.27-3.80) or a new health issue (OR 1.74, 95% CI 1.43-2.12) had higher odds for not seeking help.

Conclusions: e-Consultations were often the patients' first choice of access route, and users expected a fast response. e-Consultations were mostly perceived as an alternative to GP appointments or calling the front desk. Patients with lower availability to the GP office had higher odds of using e-consultations as an alternative to out-of-hours service or waiting and not

XSL•FO
RenderX

seeking GP care. Guidance for patient use should be developed to ensure appropriate and safe use. Further research should assess the effect of e-consultations on health outcomes and efficiency.

(J Med Internet Res 2025;27:e55158) doi:10.2196/55158

KEYWORDS

e-consultation; remote consultation; telehealth; primary care; general practitioner; patient experience; cross-sectional; digital health; survey

Introduction

Background

The increased availability of e-consultations with general practitioners (GPs) has raised questions about how they affect patients' health care service use. An e-consultation consists of asynchronous text-based messaging between the patient and GP [1]. Norwegian authorities state that e-consultations must be used to solve clinical problems [2] and expect that e-consultations will be used as an alternative to physical appointments to relieve GPs from some of the workload pressure [3]. Users perceive e-consultations as more accessible and available than physical GP consultations [4-6]. At the same time, GPs have reported increased workload due to patients seeking help for minor issues [5,7] and a need for follow-ups after e-consultations [8].

e-Consultations in Norwegian General Practice

Remote consultations with the GP can be conducted through e-consultations, video consultations, or telephone consultations. Adoption and use of remote consultations among GPs and patients in Norway experienced significant growth in relation to the COVID-19 pandemic [9]. While e-consultations and telephone consultations were offered by most GPs, video consultations were only used to a limited extent in 2023 [10].

According to Norwegian authorities, e-consultations must provide medical advice, and GPs should generally respond within 5 working days. Patients pay the same out-of-pocket fee for e-consultations as for face-to-face GP appointments [2]. In most situations, e-consultations are initiated by patients. However, there are no official guidelines to assist patients in assessing if their issues can be handled through e-consultations.

Theories and Knowledge About the Use of and Access to e-Consultations

The implementation of e-consultations has changed patients' access to GPs and possibly affected the patients' way of seeking health care [11]. Patients' adoption and use of e-consultations are affected by demographic characteristics [5,12] and health needs [13,14], in addition to perceived availability, accessibility, and capability to use the service [4,15]. Perceived availability and accessibility of e-consultations can, in turn, depend on factors such as self-selected use, information about the service, patient expectations and experience, expected waiting time for a GP appointment, and cost of service [4,5,16,17].

Research shows that patients value e-consultations for their high accessibility and efficiency [18-21]. e-Consultations are primarily used for less severe health concerns [18] and already-known problems [22]. Patients are less willing to use e-consultations for serious issues [20] and think that complex issues should not be addressed in e-consultations [21]. However, a study from Denmark showed that the complexity of the content of e-consultations was high in terms of both the number of questions asked and the disease perspective [23]. An established relationship between GP and patient facilitates good communication around clinical issues and mutual trust in e-consultations [4,24]. If e-consultations are to reduce the workload for GPs, they should be used as an alternative to a face-to-face consultation without the need for a follow-up examination at the GP office [7,25]. However, there is a lack of clear evidence regarding the circumstances under which an e-consultation can effectively replace a face-to-face consultation and the potential impacts of such use [5]. Some studies show that e-consultations are not used as a replacement for face-to-face consultations but rather as an additional service [5,6,26]. Research shows that patients can use e-consultations as an alternative to calling the front desk (ie, patients asking the GP for help to triage their health issues) [3,27,28]. Another study suggests that the high availability of e-consultations can reduce visits to out-of-hours primary care services [29].

It has been argued that e-consultations should not replace other general practice services, instead, they should be viewed as a complementary service that adds flexibility to general care [4,7,18,30,31]. The flexibility and availability of e-consultations can generate new demand for GP care. By lowering patients' threshold for contact, e-consultations may lead to requests that would not have otherwise entered the health care system, thereby increasing the GP workload [25,27]. However, it remains unclear whether this new demand arises from previously unmet needs due to limited access or from patients reaching out more readily for minor issues [20,25].

Study Aim

This study aimed to investigate how users perceived their use of e-consultations. Did they consider it an alternative to other GP services, or was a new demand for GP care created? We defined other GP services as GP appointments, phone calls to the front desk, or out-of-hours services. Information about circumstances for use and how patients perceived the role of e-consultations in the general practice setting is important for future service development.

The research questions (RQs) addressed by this study were as follows: (RQ1) For what purpose and with what expectations do patients initiate e-consultations? (RQ2) If e-consultations had not been available, what alternative actions would the patient have taken? and (RQ3) How are the alternative actions associated with patient and e-consultation characteristics?

XSL•FO
RenderX

Methods

Study Setting

General practices in Norway are organized in a scheme where citizens can choose a specific GP. The average patient list per GP was approximately 1040 patients in 2023 [32]. GPs serve as gatekeepers to specialized health care services and provide welfare services such as sick certifications. They have substantial autonomy in organizing their practices, including the option to offer e-consultations. Most GPs in Norway are responsible for their own GP list but share office locations and support personnel with other GPs. On average, the Norwegian citizen conducted 3 consultations (including e-consultations) with the GP in 2023 [33]. In recent years, the GP scheme has had problems with high work pressure [34]. The national health portal, Helsenorge, hosts several digital health services for citizens and is the main platform for digital GP services, including e-consultations. GPs are also responsible for providing out-of-hours services and taking turns staffing clinics outside regular hours. The out-of-hours service addresses urgent health care needs, such as acute illnesses, injuries, or conditions that cannot wait until regular clinic hours. The Norwegian health care system offers free GP care with minimal out-of-pocket expenses for patients. In January 2023, patients paid the same amount for regular consultations and e-consultations, 160 NOK (US $14.36) or 212 NOK (US $19.03) for consultations with GPs without or with specialization in general practice, respectively. In addition to the out-of-pocket fee from patients, GPs receive extra payment per consultation from the reimbursement system. GPs received the same total amount for both regular consultations and e-consultations, 175 NOK (US $15.71) without specialization or 285 NOK (US $25.59) with specialization [2].

Study Design

We conducted a cross-sectional study through a nationwide web-based survey of e-consultation users. The user survey was developed by the study authors, who have expertise in digital health and primary health care. It was based on academic knowledge and previous research on digital health care use and theories of access to health care [4,12,13,15-17,22]. Textbox 1 shows the survey themes. The survey was in Norwegian. The whole survey, and English translations can be found in Multimedia Appendix 1.

Textbox 1. Overview of survey topics.

Patient's background characteristics

1. Gender
2. Age
3. Education
4. Health care use: both general practitioner (GP) appointments and e-consultations
5. Digital maturity: first-time user or not
6. Proxy user

Availability of GP services

7. Travel time to the GP office
8. Ease of getting through on the phone to the doctor's front desk
9. Ease of getting a GP appointment within a reasonable time

Use of and access to e-consultations

10. Information about the service
11. Used as the first choice for issues in the matter
12. Expected response time
13. Cost of service
14. Satisfaction with the use of e-consultations
15. Alternative action if e-consultations had not been available

Characteristics of the problem handled in the e-consultation

16. Level of concern about the problem
17. Perceived suitability of issue to be handled digitally
18. Reason for sending e-consultation

Experts from Norsk Helsenett, the national service provider of the health portal Helsenorge, revised the survey. The survey was piloted on 5 e-consultation users in different age groups. Norsk Helsenett facilitated the survey invitation on the web page Helsenorge.no. All users who sent an e-consultation to their GP through Helsenorge in 3 weeks from Monday, January 30, 2023, to Sunday, February 2, 2023, received an invitation to the survey. The invitation was presented through a pop-up

XSL•FO
RenderX

on the web page and read "Do you want to participate in research about e-consultations?" The e-consultation service is only accessible through a secure log-in. If a user sent more than one e-consultation during the study period, a pop-up would appear for each e-consultation they sent. The respondents were informed that participation in the survey was voluntary and anonymous. They gave consent by clicking "I agree to be a part of the survey" at the bottom of the page and were then redirected to an external web page containing the survey. The web-based data collection solution used was Questback Essential.

The survey had up to 20 questions (depending on the respondent's answers) distributed over 15 pages. The survey was written in Norwegian, and pilot estimations showed that the survey took about 7 minutes to complete. All questions were mandatory to answer. The questions had multiple choice answers, except for 3 questions about availability and use that were answered through a 5-point Likert scale. Only completed questionnaires were analyzed. It was not possible to calculate the completion rate since no information about uncompleted questionnaires was provided. We used the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cross-sectional studies [35] (Multimedia Appendix 2) and the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [36] when reporting the results (Multimedia Appendix 3).

Data Analysis

The response rate was calculated based on routine data about all patient-initiated e-consultations sent through Helsenorge during the 3-week period.

RQ1 was answered using descriptive statistics, including the frequency and percentages of all respondents. Survey topics 10-18 (except 15) were analyzed (Textbox 1).

RQ2 was answered through survey topic 15 ("If you had not had the opportunity to send an e-consultation, what would you have done?"). The answer options to this question were book a GP appointment, call the front desk, wait or seek information on the internet, or take other actions. These are hereafter called alternative actions. Descriptive statistics were reported, including the frequency and percentages of all respondents.

RQ3 was examined through a multinomial logistic regression. Conducting a regression made it possible to investigate the association between the characteristics of the patients and e-consultations and their alternative action to sending the e-consultation. The dependent variable was the users' assessment of alternative actions to using e-consultation (survey topic 15). The reference group of the analysis was the users choosing "booking a GP appointment." This group was chosen since e-consultations aim to be an alternative to physical GP appointments for their most effective use. We compared it with users who would rather have called the front desk, contacted out-of-hours service, or waited or searched for information on the internet. This gave reasonable indications of use patterns and reasons for sending an e-consultation in different circumstances. The group that answered "other actions" was not included in the analysis. The respondents of this group are presumably heterogeneous, making it difficult to analyze their group characteristics meaningfully. We included all other survey topics as independent variables in the regression, except the time and day for answering the survey (Textbox 1). For simplicity, the survey questions with 5 answer options were merged into 3 options ("strongly agree" and "agree" were merged to "agree," "strongly disagree" and "disagree" were merged to "disagree," "very satisfied" and "satisfied" were merged to "satisfied," and "very dissatisfied" and "dissatisfied" were merged to "dissatisfied").

We examined the dataset for multicollinearity before performing the regression, and no indications of multicollinearity were found. As the goal was to investigate associations rather than build a model, we did not present model fit values. The likelihood-ratio test showed that not all variables were statistically significant; however, we included them all to not bias the results. We presented the odds ratio (OR) and 95% CI. The *P* value for statistical significance was set to .05. All analyses were conducted using SPSS (version 25; IBM Corp).

Ethical Considerations

The study and the data handling procedure were approved by the Data Protection Officer of the University Hospital of North Norway (#03057). According to the Norwegian Act on Medical and Health Research §2 and §4, the study did not require approval from the ethics committee. The study involved human participants. Participants were informed that they gave their consent to participate in the study by answering the survey, and no compensation to participants was given. All data were anonymous.

Results

Respondent Statistics

During the period the survey was available, 163,977 patient-initiated e-consultations were sent in the portal. A total of 13,648 e-consultations were assessed by users in the survey, giving a response rate of 8.32%. After two and a half days, we were informed that the question "If you had not had the opportunity to send an e-consultation, what would you have done?" was mistakenly set up as a multiple-choice question. This was corrected to a single-choice format. Respondents who had provided multiple answers (n=637) were removed, leaving a final sample of 13,011 responses.

Table 1 lists the characteristics of the respondents. A total of 78.1% (10,155/13,011) of the respondents had less than 30 minutes of travel time to get to the GP office. Almost half of the respondents (6095/13,011, 46.8%) agreed that it was usually easy to get through on the phone to the front desk, and 64.3% (8366/13,011) agreed that they usually got a GP appointment within a reasonable time.

XSL•FO
RenderX

Table 1. Characteristics of survey respondents (N=13,011).

Topic and options	Value, n (%)
Gender	
Female	9059 (69.6)
Male	3848 (29.6)
Other or I don’t want to answer	104 (0.8)
Age range (years)	
16-25	823 (6.4)
26-40	3595 (27.6)
41-55	4764 (36.6)
56-70	3123 (24)
≥71	706 (5.4)
Highest completed education level	
10 years of primary school or less	795 (6.1)
Upper secondary school	2753 (21.2)
Vocational school	2170 (16.7)
University less than 4 years	3156 (24.2)
University more than 4 years	3875 (29.8)
Other	262 (2)
Number of GP[a] appointments in the last 12 months	
0-3	6236 (47.9)
4 -9	5572 (42.8)
10-19	1029 (7.9)
20 or more	174 (1.4)
Number of e-consultations in the last 12 months	
1-3	6646 (51.1)
4-9	4607 (35.4)
10-19	1358 (10.4)
20 or more	400 (3.1)
Proxy user?	
No, sending it on behalf of myself	11,828 (90.9)
Yes, on behalf of my child	1068 (8.2)
Yes, on behalf of others I have power of attorney for	115 (0.9)
Time of answering the survey	
06:00-09:59 (day)	3199 (24.6)
10:00-13:59 (day)	4532 (34.8)
14:00-17:59 (day)	2483 (19.1)
18:00-21:59 (day)	1825 (14)
22:00-01:59 (night)	803 (6.2)
02:00-05:59 (night)	169 (1.3)
Day of answering the survey	
Monday	3091 (23.8)
Tuesday	2245 (17.3)
Wednesday	2338 (18)

XSL•FO
RenderX

Topic and options	Value, n (%)
Thursday	2364 (18.2)
Friday	1660 (12.8)
Saturday	371 (2.9)
Sunday	942 (7)
User's perception of the availability of GP services	
Travel time to GP office	
0-30 minutes	10,155 (78.1)
30-60 minutes	2178 (16.7)
1-2 hours	468 (3.6)
More than 2 hours	210 (1.6)
It is usually easy to get through on the phone to my GP office front desk	
Strongly agree	2199 (16.9)
Agree	3896 (29.9)
Neither agree nor disagree	3365 (25.9)
Disagree	2272 (17.5)
Strongly disagree	1279 (9.8)
I usually get a GP appointment within a reasonable time	
Strongly agree	3601 (27.7)
Agree	4765 (36.6)
Neither agree nor disagree	2513 (19.3)
Disagree	1489 (11.5)
Strongly disagree	643 (4.9)

[a]GP: general practitioner.

Use of e-Consultations: for What Reasons and With What Expectations Were e-Consultations Initiated?

Most respondents (10,351/13,011, 79.7%) used e-consultations as their first choice to contact the GP that day, and 83.9% (10,913/13,011) were satisfied with using it.

The results indicate a lack of information about the service given to users. More than half (7466/13,011, 57.4%) of the users had not received information about the possibility of sending an e-consultation but found out about the service by themselves. As many as 20.5% (2670/13,011) of the respondents did not know whether or not they had to pay to send an e-consultation. Recommendations for GPs' use of e-consultations state that an answer should be given within 5 days. However, more than two-thirds (8802/13,011, 68.7%) of the respondents expected a response to their e-consultations within 24 hours.

Only a minority of the e-consultations (1824/13,011, 14%) dealt with problems the user was very worried about. A total of 17.8% (2310/13,011) thought that the e-consultation would result in the GP asking them to come to the office for a physical examination (Table 2).

Table 3 shows the reasons why a patient sent an e-consultation. Multiple answers were permitted. Most e-consultations contained a request for a sick certification (4940/13,011, 38%) or were about a known problem (4778/13,011, 36.7%). A new health problem was the reason for 23.2% (3015/13,011) of the consultations. Only 8.5% (1110/13,011) of the e-consultations were initiated to ask the GP whether they should book a GP appointment for their problem.

XSL•FO
RenderX

Table 2. Patient's perception of access and use of e-consultations (N=13,011).

Topic and options	Value, n (%)
Access and use of e-consultations	
Who told you that sending an e-consultation to the GP[a] is possible?	
Found the service myself	7466 (57.4)
The GP or the health receptionist	4553 (35)
Brochures, advertisements, or other written material	101 (0.8)
Friends, family, or colleagues	652 (5)
Other	239 (1.8)
Sending an e-consultation was my first choice to get an answer from my GP about my problem today.	
Strongly agree	7497 (57.6)
Agree	2872 (22.1)
Neither agree nor disagree	1410 (10.8)
Disagree	804 (6.2)
Strongly disagree	428 (3.3)
Expected time to get an answer to this e-consultation	
Within 12 hours	4445 (34.2)
Within 24 hours	4356 (33.5)
Within 48 hours	2340 (18)
Between 48 hours and 5 days	1860 (14.3)
Getting the e-consultation for free?[b]	
Yes	1140 (8.8)
No	9201 (70.7)
Don't know	2670 (20.5)
Would you have sent this e-consultation if you had to pay for it?[c] (n=1150)	
Yes	870 (76.3)
No	141 (12.4)
I don't know	139 (11.3)
Characteristics of the problem handled in the e-consultation	
How concerned are you about the issue you sent an e-consultation for?	
Not worried	4744 (36.5)
Somewhat worried	6443 (49.5)
Very worried	1824 (14)
Do you think the GP will answer this e-consultation by asking you to come to the office for a physical examination?	
Yes	2310 (17.7)
No	5396 (41.5)
Don't know	5305 (40.8)
All in all, how satisfied were you with contacting the GP through an e-consultation today?	
Very satisfied	6495 (49.9)
Satisfied	4418 (34)
Neither satisfied nor dissatisfied	1776 (13.6)
Dissatisfied	221 (1.7)
Very dissatisfied	101 (0.8)

XSL•FO
RenderX

[a]GP: general practitioner.

[b]Some patient groups receive free health care services, including e-consultations (eg, pregnant women, first-time military service members, and patients with high costs of health care services receiving exemption cards).

[c]Question was only asked to the ones who answered that they got the e-consultation for free.

Table 3. Patients' reasons for sending an e-consultation (N=13,011).

Reasons	Value, n (%)
Sick certificate or other certificates	4940 (38)
Known health problems	4778 (36.7)
New health problems	3015 (23.2)
Medication use	1595 (12.3)
Other issues	1537 (11.8)
Asked if I needed to book a GP[a] appointment	1110 (8.5)
Test results	943 (7.2)
Answered an e-consultation from the GP	453 (3.5)

[a]GP: general practitioner.

Handling of the Medical Problem if e-Consultations had not Been Available

Nearly 9 out of 10 respondents would have contacted the GP office if the e-consultation service had not been available. Only 4.7 % (609/13,011) of the patients would not have contacted health care services but rather sought information on the internet or waited Table 4).

Table 4. Patients' assessment of alternative action if the e-consultation service had not been available (N=13,011).

Alternative action	Value, n (%)
Book a GP[a] appointment	5917 (45.5)
Call the front desk	5846 (44.9)
Contact out-of-hours service	323 (2.5)
Wait or seek information on the internet	609 (4.7)
Other	316 (2.4)

[a]GP: general practitioner.

Factors Associated With Using e-Consultation as an Alternative to Other Actions

Overview

Table 5 shows the results from the multinomial regression model. The table presents associations between the characteristics of patients and e-consultations and how the respondent would have handled the medical problem if e-consultations had not been available. The reference category is "Book a GP appointment." Associations with the patient's alternative options (calling the front desk, contacting out-of-hours, and waiting or seeking information) relative to booking a GP appointment are reported with all other variables held constant. Descriptive statistics of all groups are found in Multimedia Appendix 4.

XSL•FO
RenderX

Table 5. Associations between characteristics of patients and e-consultations and their perception of what they alternatively have done if e-consultations had not been available.

	Call front desk (n=5846)	Out-of-hours service (n=323)	Wait or seek information (n=609)[a]
	OR[b] (95% CI)	OR (95% CI)	OR (95% CI)
Background patient characteristics			
Gender (reference: female)			
Male	0.85*** (0.78-0.93)	1.26 (0.98-1.63)	0.94 (0.77-1.14)
Age range (years; reference: ≥71)			
16-25	1.15 (0.91-1.45)	1.44 (0.74-2.81)	2.16*** (1.38-3.37)
26-40	1.03 (0.85-1.24)	1.07 (0.61-1.85)	1.29 (0.89-1.86)
41-55	0.91 (0.76-1.08)	0.88 (0.52-1.51)	0.75 (0.52-1.07)
56-70	0.96 (0.80-1.16)	0.90 (0.52-1.56)	0.69* (0.47-1.0)
Education (reference: university education)[c]			
Nonuniversity education	1.12* (1.0-41.22)	1.24 (0.98-1.58)	0.99 (0.83-1.19)
Number of GP[d] appointments in the last 12 months (reference: 1-3)			
4-9	1.07 (0.98-1.16)	1.13 (0.87-1.47)	1.02 (0.84-1.24)
10-19	1.05 (0.90-1.22)	1.31 (0.84-2.05)	1.07 (.077-1.50)
20 or more	1.09 (0.77-1.55)	2.75* (1.29-5.86)	1.24 (0.59-2.59)
Number of e-consultations in the last 12 months (reference: 1-3)			
4-9	1.06 (0.97-1.16)	1.04 (0.79-1.38)	1.01 (0.82-1.24)
10-19	1.26** (1.10-1.44)	1.33 (0.88-2.02)	1.19 (0.87-1.62)
20 or more	1.30* (1.02-1.65)	1.01 (0.49-2.07)	1.45 (0.89-2.38)
First time sending e-consultation (reference: no)[e]			
Yes	1.00 (0.88-1.13)	0.88 (0.61-1.27)	0.90 (0.68-1.18)
Availability of GP services			
Travel time to GP office (reference: less than 30 minutes)			
30-60 minutes	0.97 (0.87-1.07)	1.55* (1.17-2.05)	1.01 (0.80-1.27)
1-2 hours	1.05 (0.86-1.29)	2.22** (1.36-3.61)	1.48 (0.99-2.22)
>2 hours	1.27 (0.92-1.76)	6.08*** (3.46-10.66)	2.19** (1.27-3.79)
I usually get through on the phone to my doctor's front desk (reference: neither agree nor disagree)			
Agree	1.07 (0.98-1.18)	0.99 (0.73-1.34)	0.84 (0.68-1.03)
Disagree	1.03 (0.92-1.14)	1.11 (0.82-1.51)	0.96 (0.76-1.21)
I usually get a GP appointment within a reasonable time (reference: neither agree nor disagree)			
Agree	1.03 (0.93-1.14)	0.58*** (0.43-0.77)	1.04 (0.83-1.31)
Disagree	0.96 (0.85-1.09)	0.91 (0.65-1.27)	1.08 (0.81-1.43)
Use of and access to e-consultations			
Received information about e-consultations (reference: info from HCP[f])[g]			
No info from HCP	1.11 (1.03-1.20)	0.92 (0.72-1.17)	1.46*** (1.20-1.78)
Sending an e-consultation was my first choice (reference: neither agree nor disagree)			
Agree	0.89 (0.78-1.01)	1.19 (0.81-1.74)	1.15 (0.85-1.54)
Disagree	1.07 (0.90-1.26)	1.45 (0.91-2.30)	0.81 (0.52-1.25)
Expected response time (reference: between 48 hours and 5 days)			
Within 12 hours	1.64*** (1.46-1.85)	2.10*** (1.43-3.08)	0.84 (0.65-1.08)

XSL•FO
RenderX

	Call front desk (n=5846) OR[b] (95% CI)	Out-of-hours service (n=323) OR (95% CI)	Wait or seek information (n=609)[a] OR (95% CI)
Within 24 hours	1.29*** (1.14-1.45)	1.35 (0.91-2.01)	0.77* (0.60-0.99)
Within 48 hours	1.03 (0.90-1.17)	0.66 (0.40-1.09)	0.70** (0.53-0.92)
Getting e-consultation for free (reference: don't know)			
Yes	0.72*** (0.62-0.84)	0.67 (0.40-1.11)	0.73 (0.51-1.04)
No	0.71*** (0.64-0.78)	0.84 (0.62-1.12)	0.73* (0.59-0.90)
Characteristics of the problem handled in e-consultation			
How concerned were you about the issue? (reference: somewhat worried)			
Not worried	1.29*** (1.18-1.40)	0.76 (0.55-1.05)	1.28* (1.06-1.55)
Very worried	1.02 (0.90-1.14)	2.71*** (2.07-3.54)	0.88 (0.67-1.1)
Do you think the GP will ask for a physical examination? (reference: don't know)			
Yes	1.02 (0.92-1.14)	1.06 (0.80-1.41)	0.76* (0.60-0.98)
No	1.37*** (1.26-1.49)	0.70* (0.52-0.95)	1.05 (0.86-1.29)
Reason for sending e-consultations			
Sick certification or other certification	1.07 (0.98-1.16)	0.94 (0.72-1.25)	0.31*** (0.24-0.39)
New problem	0.94 (0.85-1.03)	2.71*** (2.09-3.51)	1.74*** (1.43-2.12)
Former known problem	0.62*** (0.57-0.68)	0.71* (0.54-0.92)	0.69*** (0.57-0.84)
Medicine use	1.04 (0.92-1.17)	0.96 (0.65-1.40)	1.16 (0.91-1.48)
Test results	1.31*** (1.13-1.52)	0.53* (0.29-0.97)	1.27 (0.93-1.74)
Ask if needed to order an appointment	0.97 (0.84-1.11)	1.13 (0.79-1.63)	0.90 (0.67-1.22)
Answer question from GP	1.11 (0.90-1.38)	1.20 (0.64-2.25)	1.54* (1.01-2.33)
Proxy user (reference: no)			
Yes	1.26** (1.09-1.45)	1.59* (1.06-2.37)	0.60* (0.42-0.84)
Satisfaction with e-consultations			
How satisfied were you with sending the e-consultation today? (reference: neither satisfied nor dissatisfied)			
Satisfied	0.74*** (0.66-0.83)	0.49*** (0.36-0.66)	0.79 (0.60-1.02)
Not satisfied	1.22 (0.92-1.61)	0.84 (0.44-1.61)	1.39 (0.77-2.52)

[a]The dependent variable "wait or seek information" consists of two options: "Sought information on the internet for my problem" and "waited a while."

[b]OR: odds ratio.

[c]Variable about education was merged into two categories: with or without university education.

[d]GP: general practitioner.

[e]We constructed a dummy variable for first-time users based on the question "Are you a first-time user?"

[f]HCP: health care personnel.

[g]Variable about receiving information about service was merged into two categories: "Info from HCP" and "No info from HCP" including "found the service by myself," "brochures, advertisements/other material," "friends/family/colleagues," and "other"

*P<.05, **P<.01, ***P<.001

Calling the Front Desk Relative to Booking a GP Appointment

The patients who expected a response within 12 hours (OR 1.64, 95% CI 1.46-1.85) and 24 hours (OR 1.29, 95% CI 1.14-1.45) had higher odds of calling the front desk compared with those who expected to wait 5 days for an answer.

The respondents who were not worried about the issue handled in the e-consultation had higher odds (OR 1.29, 95% CI 1.18-1.40) of calling the front desk compared with those who were somewhat worried.

The patients who thought the GP would not ask them to come in for a physical examination had higher odds of contacting the front desk (OR 1.37, 95% CI 1.26-1.49) compared with the

XSL•FO
RenderX

group who said they did not know if the GP would recommend a physical examination.

The patients who asked about test results had higher odds (OR 1.31, 95% CI 1.13-1.52) of contacting the front desk in comparison to the ones who did not use e-consultations to ask about test results.

Contacting the Out-of-Hours Service Relative to Booking a GP Appointment

Among the respondents who had more than 30 minutes of travel time to the GP office, the odds of contacting out-of-hours service were much higher (30-60 minutes: OR 1.55, 95% CI 1.17-2.05; 1-2 hours: OR 2.22, 95% CI 1.36-3.61; over 2 hours: OR 6.08, 95% CI 3.46-10.66) compared with those who had less than 30 minutes of travel time.

The odds for contacting out-of-hours services were higher (OR 2.10, 95% CI 1.43-3.08) for those who expected an answer within 12 hours compared with those who expected an answer within 5 days.

For patients who usually get a GP appointment within a reasonable time, the odds of contacting out-of-hours service were lower (OR 0.58, 95% CI 0.43-0.77) compared with patients who answered that they thought GP appointments were neither available nor unavailable.

Among the patients who were very worried about their health issues, the odds of contacting out-of-hours service were higher (OR 2.71, 95% CI 2.07-3.54) compared with those who were only somewhat worried about their health issues.

For patients who had a new problem, the odds of contacting out-of-hours service were higher (OR 2.71, 95% CI 2.09-3.51) compared with those who did not have a new problem.

Waiting or Seeking Information on the Internet Relative to Booking a GP Appointment

Among the patients aged 16-25 years, the odds for waiting or seeking information on the internet were higher (OR 2.16, 95% CI 1.38-3.37) than for users over 70 years old.

Also, among patients with a particularly long travel time to the GP office (more than 2 hours), the odds for waiting or seeking information on the internet were higher (OR 2.19, 95% CI 1.27-3.80) than users with a shorter travel time.

The respondents who expected a short response rate had lower odds (OR 0.77, 95% CI 0.60-0.99 and OR 0.70, 95% CI 0.53-0.92) for choosing the waiting or seeking information on the internet alternative compared with those expecting a response rate of over 48 hours.

Among the respondents who had not received information about the service from health care personnel, the odds of waiting or seeking information on the internet were higher (OR 1.15, 95% CI 1.20-1.78) than the ones who had received information.

If the patients' reason for sending an e-consultation was a new problem, the odds of choosing to wait or seek information on the internet were higher (OR 1.74, 95% CI 1.43-2.12) compared with users who did not ask about a new problem.

Discussion

Principal Findings

This web-based survey gathered experiences from 13,011 users of e-consultation in Norway. Most had short distances to travel to the GP office, but still, 79.6% (10,369/13,011) reported that e-consultations were their first choice for contacting the GP. The 2 most common reasons for sending an e-consultation were obtaining a sick certification or asking a question about an already known health problem. In addition, 1 out of 4 e-consultations concerned a new health issue. However, only 14% (1824/13,011) of the requests addressed issues the patient was worried about. Almost 70% (8801/13,011, 67.7%) of the users expected to receive an answer from their GP within 24 hours, despite the Norwegian recommendation to GPs to provide an answer within 5 working days.

If e-consultations had not been available, 45.5% (5917/13,011) of the patients would have booked a physical GP appointment, and 44.9% (5846/13,011) would have contacted the front desk. Patients who alternatively would have called the front desk were associated with not worrying about their health problems and not thinking a physical examination would be necessary. Few patients (323/13,011, 2.5%) answered that they would alternatively have contacted out-of-hours services. These were associated with longer travel distances to the GP office and low availability of GP appointments. Finally, 4.7% (609/13,011) responded that waiting or seeking information on the internet was their potential alternative for sending an e-consultation. These were associated with younger age, a new health problem they did not worry about, and long travel distances to the GP office.

Strengths and Limitations

Despite a low response rate, which is often the case in web-based surveys, the web-based survey design facilitated an efficient collection of a large sample size. Reminders may have increased the response rate. However, as the survey was made available through a pop-up in the web browser and no e-mail addresses or other identifiers were collected, there was no possibility of sending survey reminders. Since the pop-up with the invitation to the survey appeared for each e-consultation sent, it was potentially possible for a user to answer more than once.

The Norwegian official register data for remote consultations do not distinguish between e-consultations, video consultations, and telephone consultations [9]. No specific data on users of e-consultations on Helsenorge are available. However, Norsk Helsenett have provided us with the total number of e-consultations sent in the portal during the survey period. Because no data are available on the characteristics of e-consultation users, it was not possible to assess the overall representativeness of the survey respondents. We could only rely on the large number of respondents.

The respondents were invited to answer the survey immediately after sending an e-consultation. Therefore, the patients' answers were not biased by the GP's response to the e-consultation. Since data were collected after the COVID-19 pandemic, we

XSL•FO
RenderX

can assume that e-consultations were used for all purposes and not mainly to minimize the spread of infections.

We recognize that different appointment booking systems among general practices could lead to inconsistency in answers about alternative actions for e-consultation. Most general practices have a web-based booking system, but some still arrange appointment bookings by telephone at the front desk. In addition, practices can operate with on-the-day acute appointments only available by booking from the front desk. This could have made it hard for some respondents to choose whether e-consultation was an alternative to calling the front desk or booking a GP appointment. The differences we found between the groups should be interpreted with care.

The regression analysis was based on the simplified assumption that e-consultations were used as an alternative or replacement for other GP services. In real life, a patient may intentionally want to use e-consultations for their request without reflecting if it replaces another service. In this setting, e-consultations have their own valuable and independent role in helping patients manage their health. Despite this simplification of actual use, our analysis provides insights into the circumstances of using e-consultations in different settings. It also provides knowledge of patients' perception of using e-consultations compared with the other GP services.

The study did not measure the outcome of the e-consultation (eg, how many e-consultations required an immediate follow-up at the GP office or how many were perceived as solved in writing). Therefore, we cannot conclude whether the e-consultation service adds extra workload to the GP due to the need for follow-up after e-consultations.

Comparison With Previous Work

Use of e-Consultations

This study confirms previous findings demonstrating that users are satisfied with e-consultations and appreciate their high availability [18,31,37]. The expectation of a rapid response time and that most users found out about the service by themselves, probably while navigating the web-based health portal Helsenorge, aligns with findings from a previous study involving older users of e-consultations in Norway [4]. Our study confirms that e-consultations are mostly used for existing medical problems [22]. However, every fourth e-consultation was about a new problem. This is a high share, taking into account that previous research has shown that e-consultations may be less suitable for newly emerged problems [37]. Using e-consultations for sick certificates has been perceived as more efficient and suitable for known health problems [28,38].

e-Consultation as an Alternative to Physical Consultations

In 45.5% (5917/13,011) of the assessed e-consultations, the respondent answered that a GP appointment would have been the alternative action for help. For both the patient and the GP, a medical issue solved by sending an e-consultation is convenient and efficient. However, the potential efficiency gain decreases if a physical examination or many follow-up e-consultations are required after the e-consultation [4,23,25]. Almost 20% (2310/13,011, 17.7%) of the patients in our study believed that the GP would answer the e-consultation by asking them to come to the office for a physical examination. This indicates a potential follow-up through a face-to-face examination. Research has shown ambiguous findings related to follow-up contact after digital contact [25]. These findings range from 34% of all e-visits in the United States to 66% of all e-consultations in the United Kingdom needing follow-up after initial digital contact [39,40]. Another study indicates that digital contacts do not lead to more or sooner follow-up consultations than in-person service [41]. Our study did not measure how often an issue was solved in an e-consultation or how often a follow-up was required. Thus, it is not possible to draw a definitive conclusion regarding the impact of e-consultations on efficiency in the Norwegian context.

e-Consultations as an Alternative to Calling the Front Desk

Previous research has shown that patients use e-consultations to obtain quick answers to minor issues [4,18,20]. Our study found that 44.9% (5846/13,011) of the users would have contacted the front desk if e-consultations had not been available. These users were associated with health concerns they did not worry about or deem necessary for a physical examination. In addition, the patients who had sent many e-consultations in the last 12 months had higher odds of choosing this alternative. We interpret this as e-consultations that meet a need for communication about minor issues when in-person GP appointments are not necessary. Even for minor issues, these patients had high expectations of a fast response from the GP.

Studies of the eConsult service in England show that issues previously handled by the front desk are directly sent to the GP's inbox after implementing the service [26,28]. The eConsult service differs from the Norwegian e-consultations service, as it is a web-based triage platform meant to be the first point of contact for triaging and often requires a follow-up [42]. In Norway, front desk personnel usually triage incoming phone calls and advise patients by giving in-person advice and clarifying requests. We believe some of the e-consultations to the GP could have been efficiently handled by the front desk personnel, thus reducing the inbox for the GP.

e-Consultations as an Alternative to Out-of-Hours Service

Previous studies have shown that patients seek out-of-hours services when GP offices are closed or are perceived as unavailable [42-44]. The users who answered that out-of-hours services would have been the alternative if e-consultations had not been available were associated with longer travel distances to their GPs. Previous research has shown that the high availability of e-consultations can slightly reduce the demand for out-of-hours services [29]. It raises a concern that users with new and worrisome problems who expect a fast response may turn to e-consultations instead of seeking immediate care. The e-consultation service is unsuitable for urgent care as the response time norm is 5 working days. Ensuring that e-consultations do not become an alternative for urgent care is crucial, as this risks patient safety. The risk of patients using

XSL•FO
RenderX

e-consultations for urgent issues has been highlighted as an unintended consequence of e-consultations in another study from England [28].

e-Consultations as an Alternative for Waiting or Seeking Information on the Internet

Our study found that 4.7% (609/13,011) of the users who initiated e-consultations would alternatively have waited to seek GP care. Earlier research has also described how e-consultations can create an extra demand for GP care [25,45]. The group that alternatively would have waited or sought information on the internet had higher odds of being younger and addressing a new issue they were less worried about. Previous research supports that better access to primary care prompts seeking help for minor complaints [46]. When e-consultations lower the threshold for use and lead to requests that could just as well be handled outside health care services, the GPs' health care resources may be misallocated. It can also add to the GP's workload. On the other hand, our study indicates that users of e-consultations with a long way to the GP office had higher odds of waiting and seeking GP care if e-consultations had not been available. This indicates that e-consultations improve access to GP care for patients who would alternatively not have sought help due to unavailability.

Implications and Future Research

Users of e-consultations are mainly satisfied, probably due to the increased convenience of e-consultations. Our study indicates a potential efficiency gain for patients and GPs as many patients initiate e-consultations in circumstances where they alternatively would have booked a GP appointment. However, we do not know how many e-consultations would need a follow-up, such as a physical appointment or requesting more information in writing. In addition, our findings suggest a slight increase in demand for GP care and a shift in tasks from front desk staff to GPs, which challenges the total efficiency gain for the GP. The pressure on GPs from patients expecting a fast response can also add to the GPs' workload. Our study was not designed to investigate efficiency gains or a potential increase or decrease in GP workload, and this question needs further investigation.

Our findings indicate that some of the initiated e-consultations were probably less suitable for the written modality. Examples include e-consultations in which patients deemed a physical examination necessary or where the issue was new or caused the patient significant concern. Knowing that few users had received information about the service from health care personnel and that several were uncertain about the cost of service shows that more information about the service is necessary. Guidance for patients could help ensure appropriate use. This is important for patient safety, keeping GPs' workload down, and helping GPs prioritize the patients with the most need. Keeping the e-consultation manageable for the GP is important to maintain the offer for the patients. As the service is voluntary for GPs to offer, an overwhelming and unmanageable inbox may drive some GPs to stop providing the service.

Future research is needed to assess the overall efficacy, safety, and impact on GP resource allocation. It should evaluate the added value of e-consultations on health outcomes. A focus should be put on evaluating the added outcome of e-consultation when patients experience increased demand for GP care due to higher availability through e-consultations.

Conclusion

This study demonstrated that e-consultations are primarily used for convenience, as most respondents reported having good access to their GP due to short travel distances to the GP office and relatively accessible GP appointments. Patients' expectations for a fast response to the e-consultation were high. Given that the GP norm is a 5-day response time, the proportion of e-consultations addressing new issues and issues the patient was worried about should raise concerns about patient safety. Nearly all patients would have booked a GP appointment or contacted the front desk had e-consultations not been available. This indicates that e-consultations may serve as an alternative to both these parts of the GP service. The study suggests a small increase in demand for GP care through e-consultations for patients who would not have sought health care if e-consultations had not been available. For patients with low availability to the GP office, some e-consultations seemed to be used as an alternative to out-of-hours service or not seeking help. Patients' concerns about their medical problems influenced how they assessed alternative actions to e-consultations. Those highly concerned had higher odds of alternatively choosing out-of-hours service, while those less concerned were more likely to call the front desk or not seek help. Ensuring appropriate use of e-consultations is crucial for patient safety and managing GP workload. Clear guidance could help patients determine when e-consultations are appropriate. Future research should explore the clinical outcomes and efficiency gains of e-consultations.

Acknowledgments

We thank Elin Synnøve Berge and Vilda Breivyte Buhs at Norwegian Health Network (Norsk Helsenett SF) for supporting the development and distribution of the survey. This work was supported by the Research Council of Norway (grant #315404). CS receives support from the UK National Institute for Health Research as a Senior Investigator (NIHR 201314). The views expressed are those of the authors and not necessarily those of the Research Council of Norway, the National Institute for Health and Care Research (NIHR), or the UK Department of Health and Social Care. No generative artificial intelligence was used to write the manuscript.

XSL•FO
RenderX

Data Availability

The dataset analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

EK, PZ, BA, TSB, and BLN developed the survey. EK drafted the results. All authors contributed substantially to the study's design and interpretation of data. All authors reviewed the manuscript for important intellectual content and approved the final version.

Conflicts of Interest

CS has received research grants from the National Institute for Health Research. No conflicts of interest are declared by the other authors.

Multimedia Appendix 1
Survey questionnaire.
[PDF File (Adobe PDF File), 126 KB - jmir_v27i1e55158_app1.pdf]

Multimedia Appendix 2
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[PDF File (Adobe PDF File), 86 KB - jmir_v27i1e55158_app2.pdf]

Multimedia Appendix 3
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[PDF File (Adobe PDF File), 162 KB - jmir_v27i1e55158_app3.pdf]

Multimedia Appendix 4
Descriptive statistics cross-table.
[PDF File (Adobe PDF File), 186 KB - jmir_v27i1e55158_app4.pdf]

References

1. Zanaboni P, Fagerlund AJ. Patients' use and experiences with e-consultation and other digital health services with their general practitioner in Norway: results from an online survey. BMJ Open 2020;10(6):e034773 [FREE Full text] [doi: 10.1136/bmjopen-2019-034773] [Medline: 32554721]
2. Normaltariff for fastleger og legevakt. List of reimbursement codes for the general practitioner services and out-of-hours services. Legeforeningen. Norwegian Medical Association. URL: https://normaltariffen.no [accessed 2024-11-29]
3. Handlingsplan for allmennlegetjenesten 2020-2024. Action plan for the general practitioner services 2020-2024. Helse- og omsorgsdepartementet. Ministry of Health and Care Services. 2024. URL: https://www.regjeringen.no/contentassets/96f6581333ee48559cdabf23c8772294/handlingsplan-for-allmennleger.pdf [accessed 2024-11-29]
4. Kristiansen E, Atherton H, Austad B, Bergmo T, Norberg BL, Zanaboni P. Older patients' experiences of access to and use of e-consultations with the general practitioner in Norway: an interview study. Scand J Prim Health Care 2023;41(1):33-42 [FREE Full text] [doi: 10.1080/02813432.2022.2161307] [Medline: 36592342]
5. Darley S, Coulson T, Peek N, Moschogianis S, van der Veer SN, Wong DC, et al. Understanding how the design and implementation of online consultations affect primary care quality: systematic review of evidence with recommendations for designers, providers, and researchers. J Med Internet Res 2022;24(10):e37436 [FREE Full text] [doi: 10.2196/37436] [Medline: 36279172]
6. Leighton C, Cooper A, Porter A, Edwards A, Joseph-Williams N. Effectiveness and safety of asynchronous telemedicine consultations in general practice: a systematic review. BJGP Open 2024;8(1):BJGPO.2023.0177 [FREE Full text] [doi: 10.3399/BJGPO.2023.0177] [Medline: 37783479]
7. Eriksson P, Hammar T, Lagrosen S, Nilsson E. Digital consultation in primary healthcare: the effects on access, efficiency and patient safety based on provider experience; a qualitative study. Scand J Prim Health Care 2022;40(4):498-506 [FREE Full text] [doi: 10.1080/02813432.2022.2159200] [Medline: 36573065]
8. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
9. Statistikk og rapporter fra Kommunalt pasient-og brukeregister (KPR). Statistics and reports from Norwegian Registry for Primary Health Care (KPR). Folkehelseinstituttet. Norwegian Institute of Public Health. URL: https://www.fhi.no/he/kpr/statistikk-og-rapporter/ [accessed 2024-11-22]

XSL•FO
RenderX

10. Assing Hvidt E, Atherton H, Keuper J, Kristiansen E, Lüchau EC, Lønnebakke Norberg B, et al. Low adoption of video consultations in post-COVID-19 general practice in Northern Europe: barriers to use and potential action points. J Med Internet Res 2023;25:e47173 [FREE Full text] [doi: 10.2196/47173] [Medline: 37213196]
11. Ricketts TC, Goldsmith LJ. Access in health services research: the battle of the frameworks. Nurs Outlook 2005;53(6):274-280. [doi: 10.1016/j.outlook.2005.06.007] [Medline: 16360698]
12. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021;5(3):BJGPO.2021.0040 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
13. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
14. Bavafa H, Hitt LM, Terwiesch C. The impact of E-visits on visit frequencies and patient health: evidence from primary care. Manage Sci 2018;64(12):5461-5480 [FREE Full text] [doi: 10.1287/mnsc.2017.2900] [Medline: 33033417]
15. Levesque JF, Harris MF, Russell G. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. Int J Equity Health 2013;12:18 [FREE Full text] [doi: 10.1186/1475-9276-12-18] [Medline: 23496984]
16. Bryce C, O'Connell MD, Dale J, Underwood M, Atherton H. Online and telephone access to general practice: a cross-sectional patient survey. BJGP Open 2021;5(4) [FREE Full text] [doi: 10.3399/BJGPO.2020.0179] [Medline: 33910917]
17. Holmgren AJ, Byron ME, Grouse CK, Adler-Milstein J. Association between billing patient portal messages as e-visits and patient messaging volume. JAMA 2023;329(4):339-342 [FREE Full text] [doi: 10.1001/jama.2022.24710] [Medline: 36607621]
18. Gabrielsson-Järhult F, Kjellström S, Josefsson KA. Telemedicine consultations with physicians in Swedish primary care: a mixed methods study of users' experiences and care patterns. Scand J Prim Health Care 2021;39(2):204-213 [FREE Full text] [doi: 10.1080/02813432.2021.1913904] [Medline: 33974502]
19. Moschogianis SF, Darley S, Coulson T, Peek N, Cheraghi-Sohi S, Brown B. Patient experiences of an online consultation system: a qualitative study in English primary care post-COVID-19. Br J Gen Pract 2024;74(744):e489-e497 [FREE Full text] [doi: 10.3399/BJGP.2023.0076] [Medline: 38164550]
20. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
21. Nilsson E, Sverker A, Bendtsen P, Eldh AC. A human, organization, and technology perspective on patients' experiences of a chat-based and automated medical history-taking service in primary health care: interview study among primary care patients. J Med Internet Res 2021;23(10):e29868 [FREE Full text] [doi: 10.2196/29868] [Medline: 34661544]
22. Chappell P, Dias A, Bakhai M, Ledger J, Clarke GM. How is primary care access changing? A retrospective, repeated cross-sectional study of patient-initiated demand at general practices in England using a modern access model, 2019-2022. BMJ Open 2023;13(8):e072944 [FREE Full text] [doi: 10.1136/bmjopen-2023-072944] [Medline: 37591638]
23. Møller JE, Fage-Butler A, Brøgger MN. Complexity and simplicity in doctor-patient email consultations. Conjunctions Transdisciplinary Journal of Cultural Participation 2021;8(1) [FREE Full text] [doi: 10.7146/tjcp.v8i1.123039]
24. Bakhai M, Atherton H. How to conduct written online consultations with patients in primary care. BMJ 2021;372:n264. [doi: 10.1136/bmj.n264] [Medline: 33627324]
25. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]
26. Atherton H, Boylan AM, Eccles A, Fleming J, Goyder CR, Morris RL. Email consultations between patients and doctors in primary care: content analysis. J Med Internet Res 2020;22(11):e18218 [FREE Full text] [doi: 10.2196/18218] [Medline: 33164902]
27. Grønning A, Assing Hvidt E, Nisbeth Brøgger M, Fage-Butler A. How do patients and general practitioners in Denmark perceive the communicative advantages and disadvantages of access via email consultations? A media-theoretical qualitative study. BMJ Open 2020;10(10):e039442 [FREE Full text] [doi: 10.1136/bmjopen-2020-039442] [Medline: 33109668]
28. Turner A, Morris R, Rakhra D, Stevenson F, McDonagh L, Hamilton F, et al. Unintended consequences of online consultations: a qualitative study in UK primary care. Br J Gen Pract 2022;72(715):e128-e137 [FREE Full text] [doi: 10.3399/BJGP.2021.0426] [Medline: 34903520]
29. Clarke G, Pariza P, Wolterz A. How are total triage and remote consultations changing the use of emergency care. The Health Foundation. 2020. URL: https://www.health.org.uk/news-and-comment/charts-and-infographics/how-are-total-triage-and-remote-consultation-changing-the-us [accessed 2024-11-22]
30. Mold F, Cooke D, Ip A, Roy P, Denton S, Armes J. COVID-19 and beyond: virtual consultations in primary care-reflecting on the evidence base for implementation and ensuring reach: commentary article. BMJ Health Care Inform 2021 Jan;28(1):e100256 [FREE Full text] [doi: 10.1136/bmjhci-2020-100256] [Medline: 33436372]
31. Clarke GDA, Wolters A. Access to and delivery of general practice services: a study of patients at practices using digital and online tools. The Health Foundation. 2022. URL: https://www.health.org.uk/publications/access-to-and-delivery-of-general-practice-services [accessed 2024-11-29]

XSL•FO
RenderX

32. Fastlegestatistikk. Statistics on general practitioners. HELFO. The Norwegian Health Economics Administration. URL: https://www.helfo.no/fastlegeordninga/fastlegestatistikk [accessed 2024-11-29]
33. Statistikk om allmenlegetjenesten. Statistics on general practitioner services. Statistics Norway. URL: https://www.ssb.no/helse/helsetjenester/statistikk/allmennlegetjenesten [accessed 2024-11-29]
34. Ekspertutvalg. Expert committee. Gjennomgang av allmennlegetjenesten. Review of general practitioner services. Regjeringen. 2023. URL: https://www.regjeringen.no/contentassets/092e9ca0af5e49f39b55c6aded2cf18d/230418_ekspertutvalgets_rapport_allmennlegetjenesten.pdf [accessed 2024-11-09]
35. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
36. Eysenbach G. Improving the quality of Web surveys: the Checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
37. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. Alternatives to the face-to-face consultation in general practice: focused ethnographic case study. Br J Gen Pract 2018;68(669):e293-e300 [FREE Full text] [doi: 10.3399/bjgp18X694853] [Medline: 29378697]
38. Breivik E, Kristiansen E, Zanaboni P, Johansen MA, Øyane N, Bergmo TS. Suitability of issuing sickness certifications in remote consultations during the COVID-19 pandemic. A mixed method study of GPs' experiences. Scand J Prim Health Care 2024;42(1):7-15 [FREE Full text] [doi: 10.1080/02813432.2023.2282587] [Medline: 37982708]
39. Penza KS, Murray MA, Pecina JL, Myers JF, Furst JW. Electronic visits for minor acute illnesses: analysis of patient demographics, prescription rates, and follow-up care within an asynchronous text-based online visit. Telemed J E Health 2018;24(3):210-215. [doi: 10.1089/tmj.2017.0091] [Medline: 28731388]
40. Nijhof D, Ingram A, Ochieng R, Roberts EJ, Poulton B, Ochieng B. Examining GP online consultation in a primary care setting in east midlands, UK. BMC Health Serv Res 2021;21(1):1030 [FREE Full text] [doi: 10.1186/s12913-021-07039-2] [Medline: 34592980]
41. Cao YJ, Chen D, Smith M. Use telehealth as needed: telehealth substitutes in-person primary care and associates with the changes in unplanned events and follow-up visits. BMC Health Serv Res 2023;23(1):426 [FREE Full text] [doi: 10.1186/s12913-023-09445-0] [Medline: 37138327]
42. Eccles A, Hopper M, Turk A, Atherton H. Patient use of an online triage platform: a mixed-methods retrospective exploration in UK primary care. Br J Gen Pract 2019;69(682):e336-e344 [FREE Full text] [doi: 10.3399/bjgp19X702197] [Medline: 30910874]
43. Zhou Y, Abel G, Warren F, Roland M, Campbell J, Lyratzopoulos G. Do difficulties in accessing in-hours primary care predict higher use of out-of-hours GP services? Evidence from an english national patient survey. Emerg Med J 2015;32(5):373-378 [FREE Full text] [doi: 10.1136/emermed-2013-203451] [Medline: 24850778]
44. Kelly SJ, Piercy H, Ibbotson R, Fowler Davis SV. Who attends out-of-hours general practice appointments? Analysis of a patient cohort accessing new out-of-hours units. BMJ Open 2018;8(6):e020308 [FREE Full text] [doi: 10.1136/bmjopen-2017-020308] [Medline: 29886444]
45. Rosen R. Meeting need or fuelling unnecessary demand? Understanding the impact of improved access to primary care. Nuffield Trust. 2014. URL: https://www.nuffieldtrust.org.uk/research/meeting-need-or-fuelling-unnecessary-demand-understanding-the-impact-of-improved-access-to-primary-care#partners [accessed 2024-11-22]
46. van Loenen T, van den Berg MJ, Faber MJ, Westert GP. Propensity to seek healthcare in different healthcare systems: analysis of patient data in 34 countries. BMC Health Serv Res 2015;15:465 [FREE Full text] [doi: 10.1186/s12913-015-1119-2] [Medline: 26453459]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GP: general practitioner
OR: odds ratio
RQ: research question
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by N Cahill, T Leung; submitted 05.12.23; peer-reviewed by F Mold, A Eccles, NS Sood, S Vanja, I Muli; comments to author 22.07.24; revised version received 16.09.24; accepted 28.11.24; published 08.01.25.

Please cite as:
Kristiansen E, Atherton H, Austad B, Bergmo TS, Norberg BL, Salisbury C, Zanaboni P
Patients' Use of e-Consultations as an Alternative to Other General Practitioner Services: Cross-Sectional Survey Study
J Med Internet Res 2025;27:e55158
URL: https://www.jmir.org/2025/1/e55158
doi:10.2196/55158
PMID:

©Eli Kristiansen, Helen Atherton, Bjarne Austad, Trine Strand Bergmo, Børge Lønnebakke Norberg, Chris Salisbury, Paolo Zanaboni. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Role of the Installed Base in Information Exchange Among General Practitioners in Germany: Mixed Methods Study

Tim Holetzek[1], MA; Andreas Häusler[2], DipPsych; Kathrin Gödde[1], Dr rer nat; Michael Rapp[2], Prof Dr Med, PhD; Jacob Spallek[3,4], MSc, Prof Dr; Christine Holmberg[1,5], MPH, Prof Dr

[1]Institute of Social Medicine and Epidemiology, Brandenburg Medical School Theodor Fontane, Brandenburg/Havel, Germany

[2]Social and Preventive Medicine, University of Potsdam, Potsdam, Germany

[3]Department of Public Health, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg, Germany

[4]Lusatian Centre for Digital Public Health, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg, Germany

[5]Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Brandenburg/Havel, Germany

Corresponding Author:
Tim Holetzek, MA
Institute of Social Medicine and Epidemiology
Brandenburg Medical School Theodor Fontane
Hochstrasse 15
Brandenburg/Havel, 14770
Germany
Phone: 49 03381411282
Fax: 49 03381411289
Email: tim.holetzek@mhb-fontane.de

Abstract

Background: Digitalization is steadily advancing on a global scale, exerting a profound influence on health care systems. To facilitate acceptance of the digital transformation, guiding principles emphasize the need for digital health structures to be person-centered and promote high-quality care. This paper examines the implementation challenges within the German health care system, with a particular focus on how change initiatives engage with existing infrastructures and organizational modes of health care delivery. This approach provides a framework for analyzing how established infrastructure determines new developments while also highlighting the procedural dynamics of change and the integration of innovations within existing information infrastructures. These established infrastructures are referred to as the installed base.

Objective: The aim of the study is to examine the installed base encountered by the digital transformation within the German health care system by investigating information exchange practices among general practitioners (GPs) and their communication with other health care actors.

Methods: A mixed methods study including a quantitative survey and semistructured qualitative interviews was conducted. The study sample consisted of all publicly accessible GP practices (N=1348) situated in the state of Brandenburg, Germany. The survey captured demographic data, communication practices, and perceived barriers to digitalization. The interviews explored experiences with digital applications. Quantitative data were analyzed using R (R Foundation for Statistical Computing), and qualitative data were managed and analyzed in MAXQDA (VERBI Software GmbH) through content analysis.

Results: A total of 250 questionnaires (response rate 18.5%) and 10 interviews with GPs were included in the analysis. GPs primarily use the telephone (n=138, 55.2%, SD 24.64), fax (n=109, 43.9%, SD 25.40), or post (n=50, 20.2%, SD 9.46) to exchange information. Newer digital communication channels such as messenger applications (n=2, 0.8%, SD 0.72) and Communication in the Medical Sector (n=1, 0.5%, SD 0.97) play a minor role. We identified three intertwined clusters displaying diverse barriers to the digitalization of GPs' communication practices: (1) incompatibility issues and technical immaturity, (2) lack of knowledge and technical requirements, and (3) additional technical, financial, and time-related burdens. These barriers were perceived as significant deterrents to the adoption of digital tools, with older GPs more reliant on analog systems and more likely to view digitalization as a source of frustration.

Conclusions: Newly established communication channels in the German health care system compete with the existing information infrastructure, which is deeply integrated into GPs' practice routines and care processes. However, this installed base has been largely overlooked in digital transformation initiatives. While newer channels hold potential, they often malfunction and are

XSL•FO
RenderX

incompatible with long-established, individualized GP workflows. Addressing these issues rather than imposing coercive measures is crucial for increasing adoption. Incorporating health care providers' perspectives and aligning new channels with established routines can prevent frustration and facilitate a smoother digital transformation.

(J Med Internet Res 2025;27:e65241) doi:10.2196/65241

KEYWORDS

digitalization; general practitioners; Germany; information and communication technologies; information exchange; primary health care; digital transformation; mixed methods study; digital health; health application; qualitative interview

Introduction

Digitalization is steadily increasing worldwide and is also massively impacting health care systems. The hope associated with digital health care systems is that they are able to mitigate the effects of staff shortages, deliver equal care in rural and deprived areas, or provide incentives for people to increase healthy living behaviors [1]. Furthermore, it is assumed that the integration of new technologies in diagnostic procedures contributes to lowering misdiagnosis rates and improving treatment processes, including the reduction of treatment delays [2]. The World Health Organization global strategy on digital health outlines principles for the digital transformation of health care systems, stating that digital health structures will be accepted when they serve quality health care and are person-centered [3]. This approach deliberately includes the health care workforce and calls for consideration of the local context, including its health care system, health needs, and resources. Simultaneously, the strategy emphasizes health data sharing and system interoperability, broadening the objectives of health care systems to encompass data production, thereby illustrating the integral role of technology in achieving these expanded goals.

This becomes particularly relevant when digitalization is considered in relation to information structures. In the context of digitalizing health care systems, the term information infrastructure describes all those "sociotechnical networks consisting of physical facilities and heterogeneous, interconnected subinfrastructures" [4] that are crucial for the exchange of information between the relevant stakeholders in a given health care system. A common way to examine processes of technological disruption in information infrastructures is to study the acceptance and adoption of new systems. Such approaches assume the presence of "old" and "new" systems, with the new either replacing or complementing the old. In this context, challenges in the transformation process are seen in organizational structures, in the users or patients, or in the technologies used [5-7]. Contrasting this perspective, Aanestad et al [8] approach implementation challenges by studying how change initiatives encounter existing infrastructures and modes of organization. Such a perspective, they argue, allows for the investigation of how existing elements shape, mobilize, and represent resources for new developments. Instead of clearly separating "old" and "new" into distinct entities, this approach allows for a focus on the procedural nature of change and the integration of innovations as part of existing information infrastructures. These information infrastructures are deeply embedded in daily work routines across organizations as well as "in social, organizational, and technical fabrics and coevolve with each other" [4,9]. Important work conducted in these routines often is tacit activities that remain hidden in formal work descriptions. Reconfiguring such work routines through technologies can lead to unanticipated and unintended consequences. Given that medical work is supported by technologies that convey information in specific formats, digitalization represents not merely a change in information transmission but actively reshapes the nature of medical practice. This is demonstrated by Vikkelsø [10] in her analysis of electronic records in Danish hospitals, showing that the introduction of digitalization leads to a reconfigured health care system rather than an improved one. She cautions us to understand the complexities and the coemergence of structure and technology rather than viewing change as a binary of good or bad, working or not working, and encourages us to conceptualize innovation in terms of the distribution of work, responsibilities, attention, and risk [10].

According to Aanestad et al [11], the successful development and implementation of information infrastructures require more than just a description of implementation objectives, necessary technological capabilities, and human skills. Instead, it is imperative to maintain a comprehensive awareness of existing organizational, institutional, regulatory, sociotechnical arrangements, established work routines, and the built environment that together form an installed base of information infrastructure in a given setting [11]. Since the existing installed base of infrastructure significantly influences the design possibilities of new infrastructure, failing to consider the existing infrastructure can lead to technical issues in newly implemented systems, poor interoperability with current systems, increased costs, and an overall lack of efficiency [12]. To accelerate the transformation process, the German federal government has enacted significant legislation in recent years to convert the German health care system into a digital one [13-20]. This includes enhancing communication among ambulatory physician offices, hospitals, and patients and resulted in mandation to implement various applications, such as digital health applications (German: *Digitale Gesundheitsanwendungen* [DHAs]) [21], electronic health records (German: *elektronische Patientenakte*) [22], or electronic medication prescriptions (German: *elektronisches Rezept*) [23]. For information exchange in the German ambulatory sector, the telematics infrastructure (TI; German: *Telematikinfrastruktur*) has been developed. The TI is a secure network specifically designed for the exchange of medical data in the German health care system. It serves as a communication platform for various stakeholders in the health care sector, such as physicians, dentists, pharmacies, hospitals, and health insurance companies. It was designed to enable the

XSL•FO
RenderX

secure exchange of medical data, including patient information, electronic prescriptions, emergency data, and electronic patient records. Key technical components of the TI are so-called connectors in all health care facilities, which identify these as authorized participants using digital certificates on chip cards [23]. These connectors have had to be connected to the TI since June 2019. Relevant elements of the TI structure are electronic prescriptions, electronic certificates of incapacity, electronic patient records, but also DHAs, video communication, or the KIM (Communication in the Medical Sector, German: *Kommunikation im Medizinwesen*) module. KIM is an email-based end-to-end encrypted information transmission procedure intended to enable the secure sending and receiving of electronic physician letters and other patient-related documents between different types of providers. The use of KIM to transmit electronic certificates of incapacity for work and electronic physician's letters has been mandatory since 2021 [16]. In the context of this paper, we define the installed base of the information infrastructure as the established communication channels that predate the current digital transformation efforts in the German health care system, which is evident in the implementation of the TI and associated channels. While a limited number of approaches address the installed base within the German context [24], the significance of specific pre-existing modes of information infrastructures for medical work has been regularly underestimated in current transformation processes to more digital information infrastructures [10]. This, in turn, has resulted in somewhat incomplete explanations for the below-average digital development in Germany compared to other European countries [23,25]. For this reason, this paper identifies practices of information exchange between general practitioners (GPs) and other stakeholders in the German health care system, using the state of Brandenburg as an example. Understanding the role and influence of the installed base of information infrastructure, we argue, will aid in anticipating challenges in the ongoing digitalization of health care services.

Methods

Study Design

To reach the study aim of capturing the installed base of information exchange in the health care system from the perspective of GPs, the study was set up as a mixed methods approach using a parallel design both for data collection and analysis (details on this mixed methods study following the Good Reporting of a Mixed Methods Study framework are provided as Multimedia Appendix 1). The quantitative arm of the study consisted of a survey distributed to all GPs residing in the state of Brandenburg in Germany. The qualitative arm consisted of semistructured, qualitative interviews with a subsample of GPs (10 interviews) and with citizens aged 65 years and older (21 interviews). The qualitative interviews were designed to provide an in-depth understanding of information exchange practices as well as the use and challenges of digital applications in ambulatory care settings. In contrast, the survey aimed to capture the current state of these practices, including the stakeholders involved and the methods of information exchange. Data collection for the qualitative study arm was conducted between September 2021 and October 2022, while the quantitative arm took place from March to June 2022. This paper focuses specifically on the perspective of GPs, incorporating data from both the GP survey and interviews.

Ethical Considerations

Ethics approval for this study was obtained from the Brandenburg Medical School ethics committee (E-02-20210531), ensuring adherence to all applicable regulations governing research involving human participants such as the Declaration of Helsinki. Informed consent was obtained from all participants prior to their inclusion in the study, encompassing both quantitative and qualitative data collection and analysis, with secondary analyses explicitly permitted under the primary consent agreements. To safeguard participant confidentiality, pseudonymization was implemented for all data: IDs were assigned to participating GPs during the survey, and pseudonyms were created for each qualitative interview. Identifying information was replaced with pseudonyms in the interview transcripts. The pseudonymization lists and participants' contact information were securely stored in paper form within a locked cabinet at the Institute of Social Medicine and Epidemiology, accessible only to authorized study personnel. All data were securely managed in compliance with institutional policies and General Data Protection Regulation guidelines. Participants were not compensated monetarily, and no identifying details appear in this manuscript or supplementary materials, ensuring both anonymity and ethical integrity throughout the research and publication process.

Data Collection

Quantitative Study Arm

The questionnaire used in this study was developed by the research team in collaboration with GPs from the state of Brandenburg to align with the specific objectives of the study. The survey themes were shaped by a thorough analysis of relevant research gaps identified in the literature. Before implementation, the questionnaire was pilot-tested with 2 GPs to ensure clarity and appropriate understanding of the questions. The finalized instrument comprises 124 items and was designed to assess the communication challenges faced by GPs in their routine practice using a series of targeted queries: (1) frequency of use of specific communication channels for information exchange with specific health care stakeholders, (2) assessment of quality of exchange, and (3) expected barriers to use of digital channels (all 5-point Likert scales). In addition, sociodemographic data on age, sex, number of years of professional experience, type of practice, and population size at the practice location were collected. The questionnaire was offered as a paper-pencil or digital version. All GPs in the state of Brandenburg were invited to participate. The address data of the GPs to be invited were made available via the website of the Brandenburg Association of Statutory Health Insurance Physicians. The total of all GPs in Brandenburg was divided into 2 random subgroups, of which at baseline (T0), one group received the questionnaire by mail as a paper-pencil version including an invitation letter, study information, and consent form, whereas the other group received a mailed invitation letter to participate in a web-based questionnaire including a link and

XSL•FO
RenderX

QR code. After 3 weeks (T1), a reminder in the form of postcards was sent to all GPs who had not yet participated at that time, with each postcard corresponding to an invitation for a web-based questionnaire. After 3 more weeks (T2), a second reminder was sent, with those physicians who received a paper-pencil invitation in the first wave receiving a digital invitation this time and vice versa. The web-based version was hosted on the professional survey platform SoSci Survey (SoSci Survey GmbH) under a commissioned data processing agreement to ensure secure and compliant data collection. The invitation to participate in the questionnaire was sent by post for both the paper-pencil version and the web-based version, with the invitation for the latter containing a QR code that led to the survey provider's website. The invitation and all communication with participants were managed by the study staff at the Institute of Social Medicine and Epidemiology.

Qualitative Study Arm

The semistructured interviews aimed at identifying practices and routines that GPs exhibit in the exchange of information with other health care actors and how these routines affect health care delivery. From this perspective, the communication practices of GPs from the German state of Brandenburg were to be identified, which in turn should shed light on attitudes and concrete behaviors in the communication practices of GPs. In particular, the role of digitalization in medical practices was to be recorded, including the associated potentials and hurdles. An interview guideline, informed by a review of existing literature on communication practices in health care and the role of digitalization in medical practices, was developed and used for conducting the interviews (Multimedia Appendix 2). All qualitative data were collected by the first author (TH). The interviews were audio-recorded and transcribed following the methodology outlined by Kuckartz [26]. GPs were recruited for the qualitative interviews in a variety of ways. On the one hand, eligible individuals from the extended research network of the Institute of Social Medicine and Epidemiology were contacted. These were selected based on convenience sampling logic. Additionally, the General Practitioners Association Brandenburg (German: *Hausärzteverband Brandenburg e.V.*) was involved as gatekeeper, supporting the recruitment by reaching out to members informing about the study and the possibility to give an interview. Finally, interviewed GPs were asked if they knew other potentially relevant GPs who would be interested in being interviewed for the study, using a snowball sampling approach. Identified GP practices received the study information by study personnel, either via email or postal mail. GPs who provided written informed consent were included in the study.

Analysis

Quantitative Study Arm

The frequency distributions of the sociodemographic data were mapped by category using a table, and relevant location parameters (mean, median, and IQR) were reported in text form. Depending on the presence of normally distributed data, 2-tailed *t* tests for independent samples or Mann-Whitney *U* tests were carried out in order to determine whether significant differences exist in the age groups according to sex. Frequency distributions for the variables were visualized using stacked bar charts. The frequency distributions for the variables were visualized with stacked bar charts, whereby categories 1 and 2 as well as 4 and 5 of the Likert scale items were combined. Spearman correlation was used to identify variations in answering behavior by age, practice site, and population size. Eta-squared statistics were calculated to determine the proportion of the variance of the variables that is explained by the grouping by practice type. Point-biserial correlations were used to examine whether participant's sex influenced answering behavior. CIs for the Spearman correlations were calculated using bootstrap replicates (K=1000), and CIs for the point-biserial correlations of the sex variable were calculated using approximations by Fisher Z-transformations. All tests for correlations were solely exploratory. R (version 2022.07.2; R Foundation for Statistical Computing) was used for the statistical analyses. In addition to the standard R functions available in base R, the following packages were used: *readxl* (version 1.4.3), *psych* (version 2.4.1), and *boot* (1.3-28) [27-29].

Qualitative Study Arm

The interview materials and responses to 2 free-text questions from the survey were included in the qualitative analysis, which focused on reasons for poor exchange quality and general experiences with information exchange in general practice. The analysis was conducted using content-structuring qualitative content analysis according to Kuckartz [26]. This method aims to identify content-related aspects within the material, conceptualize the data concerning these aspects, and systematically describe them [30]. Within the framework of the analysis, deductive categories were generated from the semistructured interview guideline. During the analysis, inductive categories using open coding were added. The deductive and inductive categories were then used to categorize the free-text responses of the survey, and additional inductive categories were added to the analysis. The category construction was not theory-based and took place in several steps: first, the deductive categories for the GP interviews were discussed among the authors in terms of precision and meaningfulness. Subsequently, the entire interviews were coded using the deductive categories, and new inductive categories were determined. The newly identified inductive categories were then discussed and finalized before being applied to the interview data. In the final step, the deductive and inductive category system was used to code the free-text responses. Here, additional inductive categories were formed by the coders, then discussed and finalized, and applied to the entire free-text dataset. The analyses took place at the category level, relating the categories developed and their concrete manifestations in the material. The selection of the categories studied in more detail was informed by their importance to the research question and absolute coding frequency. Each step was carried out independently by 2 researchers. Between the individual steps, the categories developed were compared, discussed, and harmonized. All qualitative data were managed using MAXQDA 2022 software (version 22.0.1; VERBI Software GmbH).

XSL•FO
RenderX

Results

Sociodemographic Data

In total, questionnaires were received from 250 of the 1348 GPs contacted, corresponding to a response rate of approximately 18.5%. Of those 250 responses, 81 were collected via the web-based version, and 169 via the paper-based survey, corresponding to 32.4% and 67.6% of the total responses, respectively. Of the participants, 63.6% (n=159) were female, and the 50- to 59-year age group was particularly well represented (n=88, 35.2%). The mean age was 54.03 (SD 9.82; median 56, IQR 46-61) years. On average, the participating physicians had 19.3 (SD 12.16; median 19, IQR 8-30) years of professional work experience. Most practices were located in regions with a population density of 5000 to 20,000 or 20,000 to 100,000 inhabitants. Most participants worked in solo practices (Table 1). Age showed no significant difference between male and female participants. Thus, the distribution of age is comparable for male and female participants in this sample. A total of 10 GPs were included in the interview study, 3 of whom were male and 7 female. The participants' ages ranged from 45 to 67 years. Of the GPs interviewed, 7 practiced in solo practices, 1 in a group practice, 1 in a joint practice, and 1 in a medical care center. The length of the GP interviews ranged from 9 to 93 minutes (mean 44.86, SD 25.1, median 46.73, IQR 28.38-52.13 minutes).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of survey participants (n=250): general practitioners in Brandenburg, Germany, participating in the survey study on information exchange (2022).

	Values
Age (years), mean (SD)	54.03 (9.82)
Age group (years), n (%)	
30-39	22 (8.8)
40-49	58 (23.2)
50-59	88 (35.2)
60-69	69 (27.6)
70+	10 (4)
Not specified	3 (1.2)
Sex, n (%)	
Female	159 (63.6)
Male	88 (35.2)
Diverse	0 (0)
Not specified	3 (1.2)
Population size at practice location, n (%)	
<1000	7 (2.8)
1001-5000	39 (15.6)
5001-20,000	98 (39.2)
20,001-100,000	77 (30.8)
>100,000	23 (9.2)
Not specified	6 (2.4)
Work experience (years), mean (SD)	19.3 (12.16)
Work experience (years), n (%)	
<5	40 (16)
5-10	31 (12.4)
11-20	67 (26.8)
21-30	56 (22.4)
31-40	40 (16)
>40	10 (4)
Not specified	6 (2.4)
Type of practice, n (%)	
Solo practice	120 (48)
Medical care center	28 (11.2)
Joint practice	51 (20.4)
Group practice	29 (11.6)
Other	17 (6.8)
Not specified	5 (2)

Survey

Frequency of Use of Specific Communication Channels for Information Exchange

The survey determined how often GPs exchange information with other actors in the health care system via specific channels. It was found that the exchange with all the actors surveyed takes place via channels that have been used for decades, namely, by postal mail, by telephone, and by fax. On average, GPs use the telephone (n=138, 55.2%; SD 24.64), fax (n=109, 43.9%; SD 25.40), and post (n=50, 20.2%; SD 9.46) often to always in order to exchange information with other stakeholder groups.

XSL•FO
RenderX

However, the exact use characteristics of the communication channels differ depending on the communication partner. With other physicians, health departments, care actors, and pharmacies or medical supply stores, faxes in particular are sent frequently. Communication with patients is largely done by telephone and sometimes by electronic or postal mail. Email is the most frequently used of all the more digital applications (n=20, 8.2% among all communication partners; SD 4.65) but is clearly behind the older channels mentioned. Messenger applications (n=2, 0.8%; SD 0.72) and KIM (n=1, 0.5%; SD 0.97), on the other hand, are used very rarely on average (Figure 1).

Figure 1. Frequency of use of selected communication channels for information exchange with various health care system actors from the perspective of general practitioners (n=250) in Brandenburg, Germany (2022). KIM: Communication in the Medical Sector.

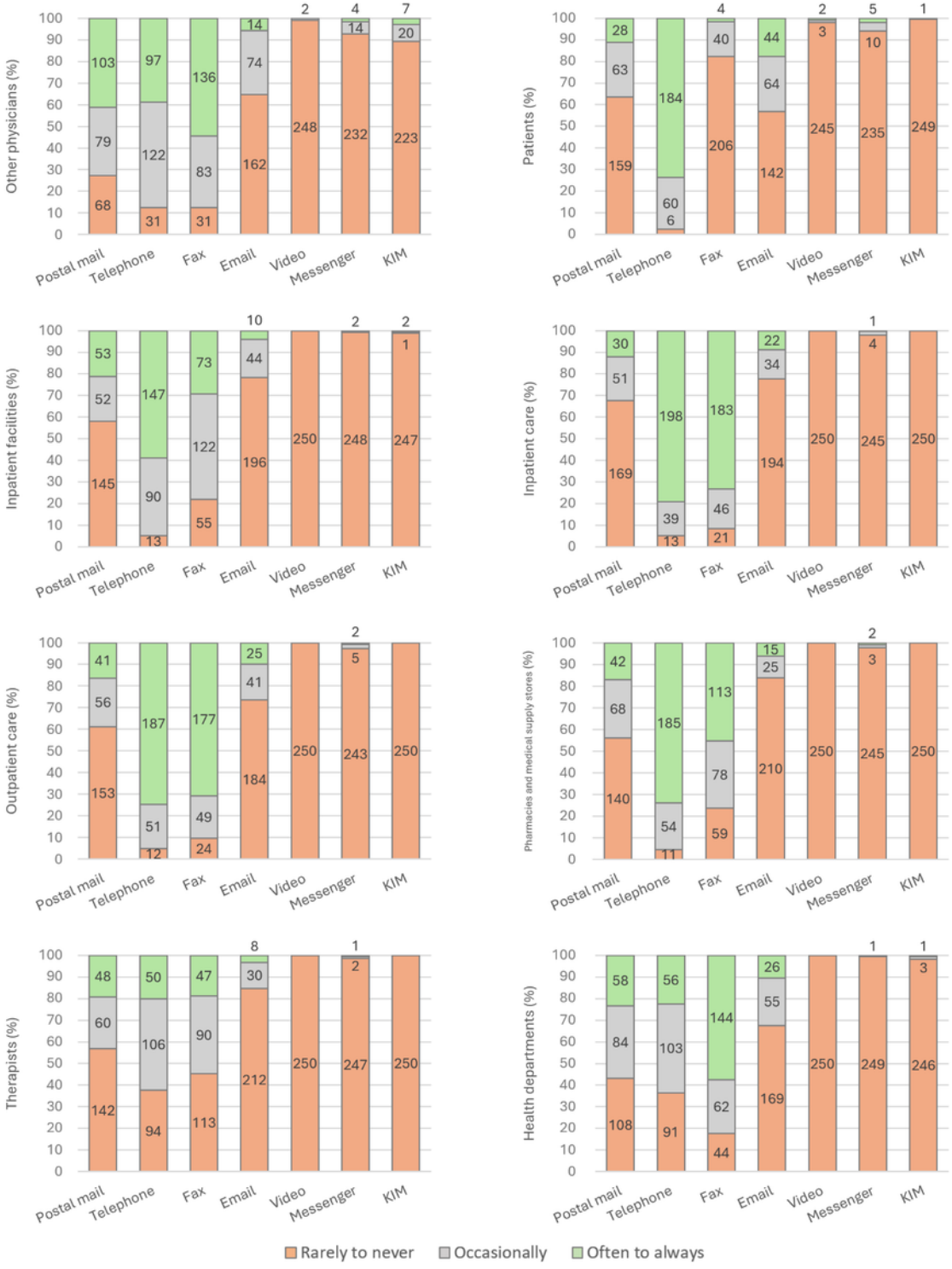

Positive correlations between age or professional experience and the frequency of use of numerous different analog channels indicate that older physicians with more professional experience may be more likely to use older communication channels such

XSL•FO
RenderX

as the telephone and postal mail. Email and fax, on the other hand, show negative correlation coefficients, which indicates that these communication channels are used less with increasing age. However, in all cases, the correlation coefficients are in the low range between –0.3 and 0.3 [31]. In contrast, sex, the size of the population at the practice location, and the type of practice show only minor or no correlation to the use of certain communication channels, which is why an influence on the use of the information channels surveyed is unlikely (Multimedia Appendix 3).

Assessment of Quality of Exchange

Looking at the satisfaction of the physicians surveyed with the exchange of information, it can be seen that GPs are relatively satisfied with the quality of the exchange, especially with pharmacies, patients, and care actors. Health departments, inpatient facilities, and therapists can be seen as exceptions here. Consequently, it might be argued that GPs are generally quite satisfied with the installed base of communication channels (Figure 2).

Figure 2. Satisfaction of general practitioners (n=250) with information exchange across various actors in the health care system in Brandenburg, Germany (2022). N/A: not applicable.

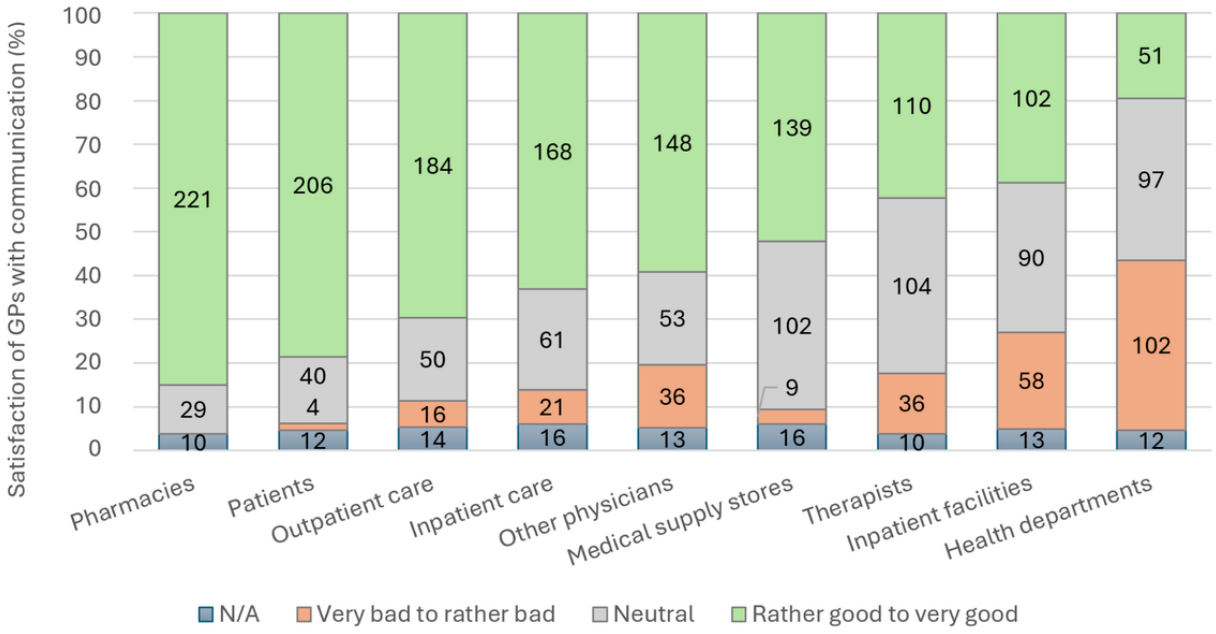

The sociodemographic variables show at best occasional links to satisfaction with the exchange of information with other actors in the German health care system. Age appears to have a certain influence on satisfaction with the exchange of information with certain actors (therapists and medical supply stores). Similarly, an increasing amount of professional years correlates with higher satisfaction with a wide range of actors (patients, health authorities, outpatient care, inpatient care, therapists, and medical supply stores). Thus, increasing age appears to contribute to greater acceptance or satisfaction with the exchange practices. All correlations are weak to moderate. Sex, population size, and practice type show little to no correlation with satisfaction with existing information exchange (Multimedia Appendix 4).

Expected Barriers to the Use of Digital Channels

The survey showed that GPs considered numerous barriers to digitalization to be potentially relevant to their practice (Figure 3). The barriers were assigned to 3 clusters, which are addressed again in the qualitative analysis: cluster 1—"incompatibility issues and technical immaturity" (containing "impracticality," "technical immaturity," "incompatibility problems," "insufficient data transfer rates," and "data protection concerns"), cluster 2—"lack of knowledge and technical requirements" (containing "lack of confidence in dealing with digital technologies [physicians]," "lack of confidence in dealing with digital technologies [patients]," "lack of technical equipment [patients]," and "lack of acceptance [patients]"), and cluster 3—"additional technical, financial, and time-related burdens" (containing "high costs," "additional technical effort," and "additional time effort"). While the additional technical and time expenditure as well as the incompatibilities and technical immaturities are particularly significant, insufficient data transfer rates and GPs' lack of skills in using digital technologies seem less relevant.

XSL•FO
RenderX

Figure 3. Perceived barriers to digitalization among general practitioners (n=250), categorized into 3 clusters: insights from a mixed methods study in Brandenburg, Germany (2022). N/A: not applicable.

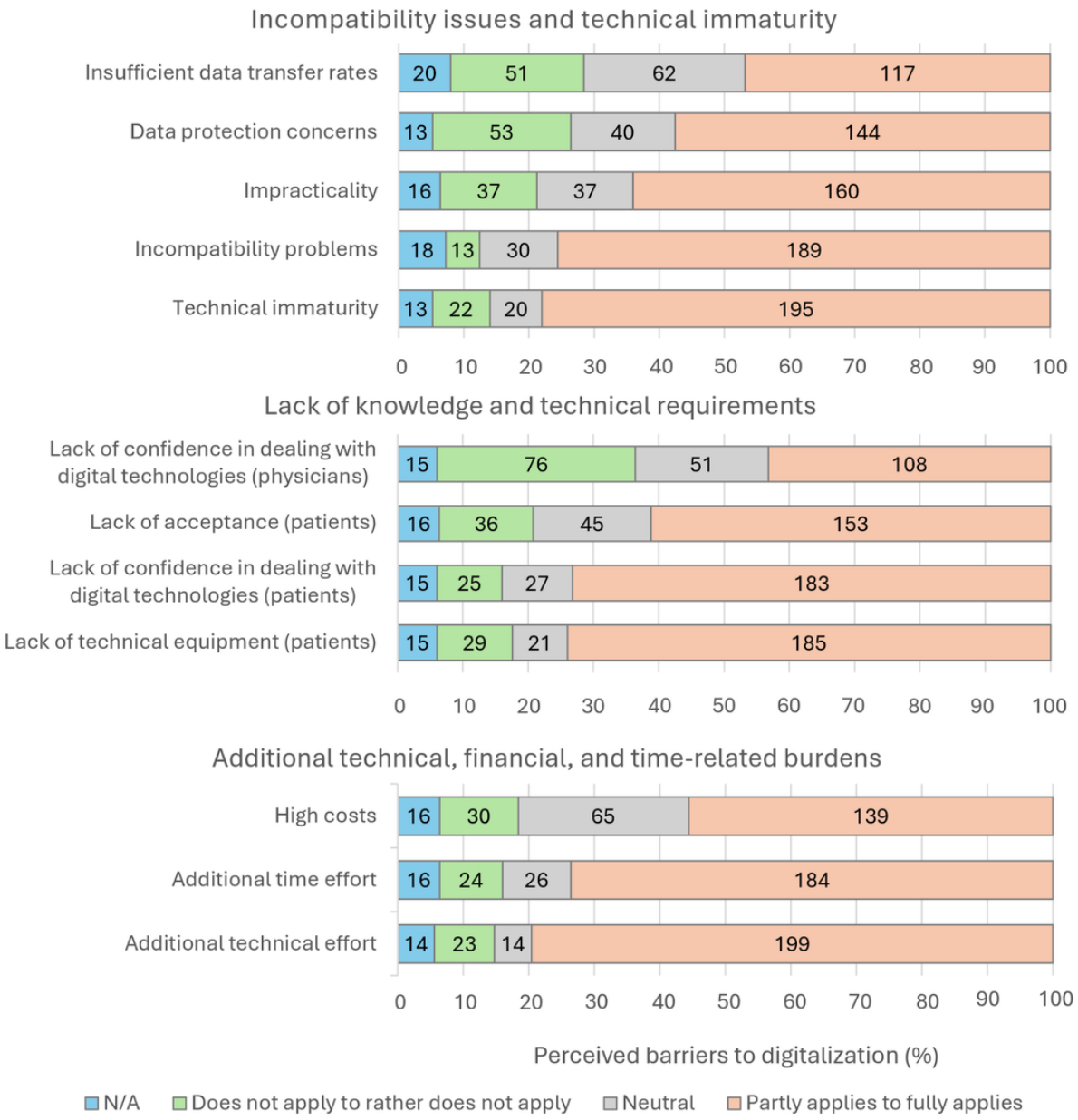

Age appears to be relevant in the perception of numerous barriers to digitalization. With the exception of "high costs," "data protection concerns," and "insufficient data transfer rates," all barriers appear to be seen as significantly more problematic with increasing age. In every case, the effect strength is low to moderate. Similar trends can be seen for the participants' years of experience: with the exception of data protection concerns and patients' lack of confidence using digital communication channels, all other barriers show significant correlations of weak to medium strength. Issues of high costs, data protection, and patients' lack of confidence using digital communication channels become more pronounced, as population size increases. In some cases, there are also minor sex-related differences in the perception of digitalization barriers. Female participants are more likely to perceive data protection and insufficient data transfer rates as problematic compared to male participants. Additionally, female participants are more likely to perceive a lack of confidence in dealing with digital technologies on the side of the physician as a barrier. This might indicate a sex difference in how certain challenges or obstacles are viewed, with female participants potentially seeing these barriers as more significant or problematic. However, the effect sizes are again small to medium (Multimedia Appendix 5).

Qualitative Results

Overview

While the survey was able to show to what degree the established base of communication channels is used by GPs and that certain barriers exist in the use of more recent applications, it was not yet possible to understand how these barriers manifest in everyday practice. To do so, the qualitative analyses can offer valuable insights and identify specific needs

XSL•FO
RenderX

regarding digitalization. The results are arranged according to the 3 clusters mentioned earlier, which were developed as part of the categorization of the qualitative data material. They appear to be strongly interrelated.

Cluster 1—Incompatibility Issues and Technical Immaturity

Both the qualitative interviews as well as the free-text analyses revealed that malfunctioning technology is a particularly relevant problem in the digital exchange of information. This includes practice management software and modules of the TI, which sometimes run very slowly. In addition, there are noticeable incompatibilities between numerous existing practice management software options and the TI, resulting in frequent system crashes. This negatively affects the user experience:

> *I have found many colleagues to be fundamentally open to digitalization—but there is a great deal of dissatisfaction with the implementation. The devices and software are often incompatible, it takes more time [to use them], there are system crashes and registrations are complex. Instead of reducing bureaucracy, the introduction of digital structures results in printouts and digital dispatch—which makes twice as much work.*

In particular, the use of KIM, which is intended to enable the secure sending and receiving of electronic physician letters and other patient-related documents between different types of providers, has been prone to numerous errors in the past and is also relatively unintuitive from a physician's perspective and involves too many clicks on the part of the sender. In addition, faulty health insurance card readers regularly cause practice management systems to fail. The various and regularly occurring technical difficulties result in the need to consult external IT support service providers. Participating GPs had varying levels of experience with such providers, but it was frequently noted that services were unsatisfactory, often caused by the associated high costs for GPs as well as poor availability of services due to high levels of capacity use.

Cluster 2—Lack of Knowledge and Technical Requirements

Another barrier associated with digitalization is the lack of knowledge about technical processes, which limits the use of digital applications. This applies to older physicians or those who have little know-how in terms of digitalization:

> *I belong to a generation that never received an introduction to data technology. As a result, all technical applications are all Greek to me. [...] At almost 63 years old, technology is also a huge burden for me, and I don't feel I have the time to cope with the innovations alongside my medical work.*

This also affects patients, especially older ones:

> *Due to the increasing aging of the population, patients are increasingly experiencing cognitive problems, explanations have to be retrieved more frequently, and additional limitations (such as hearing loss) make this even more difficult. The use of digital media overwhelms almost all patients older than 80, many no longer use cell phones, PCs or tablets. In my experience, the intended extensive digitalization does not reach about 25% of my patients, and about 50% of patients with a permanent medication.*

In addition to a lack of technical affinity, affected individuals, especially older patients, lack technical equipment. This includes hardware such as terminal devices (computers, smartphones, and tablets) but also an inadequate internet connection. Especially some rural medical practices only have a poorly developed internet connection at the practice location, which limits the use of the TI:

> *The internet connection is unfortunately inadequate, in some cases only 0.1 Mbit/s when downloading. This is a major hurdle with regard to digitalization, especially when using electronic certificates of incapacity or e-prescriptions.*

Cluster 3—Additional Technical, Financial, and Time-Related Burdens

As a direct consequence of the previously mentioned deficits in digital information exchange, these cause enormous additional time and administrative expenses. The need for frequent error management and software updates ties up physicians' time and limits patient care:

> *We are witnessing terrible changes with the introduction of the telematics infrastructure! It takes an extreme amount of extra time to personally create all the prerequisites for this. You need time for the installation, the constant updates, there is a constant need for remote maintenance for normal working functions of the practice management system. [...] A lot more "clicks" are required for every work step. I don't see any improvement or simplification through the TI, but an extreme time burden with bureaucratic and technical tasks.*

In addition, the various susceptibilities to failure of existing hardware and software result in a technology dependency, which also manifests itself in continuous error management and thus ties up time that could be used for patient care:

> *And it's also a huge nuisance in everyday practice when things don't work. If the system fails, care is actually no longer possible. You can't do anything anymore. You can talk to the patient, but you can't write a prescription. You can't call up a single form. In fact, nothing is possible if the telematics infrastructure is blocked. Because it is also not possible to read the health insurance card.*

The high speed of innovation, expressed in a constant stream of new applications and updates, also makes it necessary for physicians to regularly spend part of their working time keeping up to date with the latest applications and initiating their implementation in their own practices:

> *It takes 3-9 months between being informed about innovations and ordering the components and delivery or installation. After installation, you have to start*

XSL•FO
RenderX

> *reading up again, as you have forgotten the information in the meantime and there were five other innovations in parallel—it never stops and you can't keep up.*

Another significant negative factor is the financial outlay that digitalization entails for GPs. This includes the comparatively high acquisition and setup costs but also the maintenance and update costs, which are correspondingly high if problems occur regularly. In particular, physicians without a deeper understanding of IT and thus problem-solving skills are at risk of having to pay more money for more frequent IT services. This is likely to include older physicians in particular. Furthermore, switching practice management software providers is very costly, since the data must be transferred by external IT companies.

The perceived additional burden of the aspects described here can become problematic for a health care system, especially if it leads to physicians resigning in the face of the hurdles posed by digitalization. In this context, especially older medical practices are at risk, if owners do not feel up to the demands of digitalization. As a result, physicians may give up their professional activities prematurely, which can lead to supply bottlenecks that may not be able to be closed quickly. This is particularly true in very rural regions with a lower physician density:

> *If my daughter hadn't said, with the pandemic and with all this digitalization, "I'm taking over the practice," then it would have been a reason for me to give up the practice. Yes, it would have been a reason for me. I have also had practice nurses who are also my age, that is, those who are about to retire. And I have to tell you, that was really a big challenge for us to handle. And there I would have said, "No, I'm not going to do that to myself now at my age." You know, first and foremost, I'm a doctor and not somebody who's constantly dealing with technology here. Because I can see it. When I come here, there's always some problem. There's always something wrong with all the software and I don't know what all. And I wouldn't have had the nerve to do it anymore. I would have said, "That's enough. I'm not going to do this to myself anymore." So that's where I'm at.*

However, the barriers and especially the high bureaucratic hurdles can also have a deterrent effect on younger generations of physicians:

> *It's just awful that there is such extensive interference with professional self-employment of physicians. I am not surprised if I do not find a successor for my practice. I would never become self-employed again under these circumstances, and I will close my practice at some point, not just at the age of 67—I am used up, exhausted, at the end of my tether.*

Discussion

Principal Findings

This paper examined the role of the installed base in information exchange between GPs and other health care actors in Brandenburg, as an exemplary more rural region in Germany. The analysis revealed several relevant findings. At the time of data collection, the participating GPs very frequently interacted via communication channels that had existed for several decades, such as telephone, fax, or postal mail. The older age of physicians appears to be associated with a greater use of more traditional communication channels within the installed base and to greater skepticism toward digitalization. This is shown in the positive and significant correlations between age or professional experience and the frequency of use of numerous different older channels, indicating that older physicians with more professional experience may be more likely to use older communication channels such as the telephone and postal mail. These older communication channels work well from the perspective of those GPs' who took part in the survey. In contrast, new communication applications developed and introduced in recent years in the context of digitalization have not been used as frequently. Among those applications, the low rate of use of the KIM service is particularly notable since KIM has been mandatory for GPs to use for information transfer while this survey was conducted. The installed base in the form of postal letters, fax, and other channels appears to compete with the newly developed KIM channel, limiting its use. Although the exact use practices varied depending on the communication partner, this general trend applies equally to all of the surveyed actors, namely, other physicians, patients, inpatient facilities, in- and outpatient care actors, therapists, pharmacies, and medical supply stores as well as health departments. As was shown, the participating GPs perceived hurdles in the use of digital communication channels, which are arranged around the 3 clusters "incompatibility issues and technical immaturity," "lack of knowledge and technical requirements," and "additional technical, financial, and time-related burdens." Taken together, the hurdles perceived by GPs could result in digital channels being seen more as an additional burden than an added value, which could limit their use and ultimately jeopardize health care provision.

The results obtained here coincide to a certain extent with those of similar studies. While the user behavior of GPs toward existing information infrastructure for the German context has scarcely been examined from the point of view of the installed base, findings do exist on the barriers to the use of digital exchange channels. In a previous study, Schendzielorz et al [32] found that in a more rural area in Germany, the high cost of applications, technical problems (slow internet connection and nonfunctioning applications), the time required to process digital and analog documents, and a lack of personal contact in particular were obstacles to the use of digital technologies.

According to the "PraxisBarometer" of the German National Association of Statutory Health Insurance Physicians 2021, GPs in Germany have been noted for their rather negative attitudes toward digitalization, with technical hurdles in the use of

XSL•FO
RenderX

associated applications appearing to be particularly significant, as expressed by a lack of user-friendliness, high conversion costs (financial costs, information and training requirements, and time needed), an unfavorable cost-benefit ratio, and concerns about data protection [33]. A lack of knowledge in the use of digital applications as well as a low tolerance for telemedicine and an advanced age are also not conducive to the willingness to use digital applications [34]. It was further shown that physicians rarely receive diagnostic data, doctor's letters, or radiological findings in digital form, and just under a quarter of medical practices do not receive any digital data from other outpatient facilities. In addition, about two-thirds of medical practices did not send any treatment-relevant data in digital form themselves. Further, the frequency of errors in the use of relevant applications has increased over the past few years [33]. These trends have continued across Germany in the years following the data collection for this paper. Despite increasing use rates of corresponding applications, digitalization has not yet become fully established in practices [35,36]. The barriers seem to be similar across applications. With regard to DHAs, for example, it is regularly noted that GPs lack confidence in their ability to introduce patients to DHAs and support their use [37]. In addition, privacy, security, and legal concerns as well as costs related to reimbursement and fees are among the most stated barriers to a more intense use of DHAs [38]. Overall, various barriers to the implementation of digital applications were identified. However, these are by no means unique to Germany but are relevant in many other countries and specialist areas outside of GP care and in relation to various digital tools [39-42].

Ultimately, it can be concluded that the newly established communication channels and applications associated with the TI compete with the installed base of information infrastructure. Some GPs have been using these established channels for decades, with their practice routines and care processes built around them. These systems function adequately for many providers at a procedural level. Moreover, the newer applications experience various malfunctions that exceed the limitations of the older communication channels. If the technical problems of the newer communication channels are sustainably eliminated, the willingness of GPs to rely more on them in their practices would very likely increase. However, if the technical issues persist, health care will be severely impacted, testing the willingness and patience of the physician community. Under these circumstances, the potential of digitalization for health care can only be guessed at. To this end, adjustments must be made above all at the individual, that is, the practice level. This is all the more relevant, as the Digital Act has declared the electronic patient file as an opt-out version for 2025, which will lead to GPs in particular having to deal with the topic and technology more intensively [19].

Strengths and Limitations

This paper is accompanied by some limitations. For example, the survey's response rate of 18.5% (n=250) among Brandenburg GPs may come along with several limitations that warrant consideration. First, the relatively low response rate raises concerns about the representativeness of the sample, which may not adequately reflect the broader population of GPs in the region. This could limit the generalizability of the findings to all Brandenburg GPs. Additionally, nonresponse bias might be present if the GPs who chose to participate differ systematically from those who did not, potentially affecting the validity of the results. It is also possible that those who responded were more likely to have strong opinions or experiences related to the survey topic, which could skew the results. The same logic applies with regard to the interviews. Accordingly, the extent to which the results are applicable to the rest of Germany beyond the federal state of Brandenburg must be carefully considered. However, the participating physicians adequately reflected the German age structure of GPs (German: *Allgemeinmedizin* and *Interne Medizin mit hausärztlicher Tätigkeit*), according to which 7.5% (n=2561) and 8.6% (n=1478) are younger than 40 years of age, 20.7% (n=7127) and 28.5% (n=4893) between 40 and 49 years, 35.3% (n=12,153) and 37% (n=6361) between 50 and 59 years, 20.6% (n=7097) and 15% (n=2584) between 60 and 65 years, and 15.9% (n=5455) and 10.9% (n=1875) older than 65 years. The national average age is 55.3 years [43]. The sex distribution roughly corresponded to the Brandenburg GP group (n=731, 62.9% of Brandenburg GPs are female) [44]. For data protection reasons, the geographical location of the participating practices was not determined. A notable limitation of the parallel mixed methods design was the inability to iteratively refine data collection tools based on preliminary findings from either method. The concurrent collection of quantitative survey data and qualitative interviews precluded adjustments to the survey instrument to incorporate emerging insights from the qualitative interviews or modifications to the interview guide to explore trends identified in the survey responses. This limitation may have constrained the ability to capture more nuanced interconnections between the 2 datasets and to fully investigate unanticipated findings. Finally, while the results presented here show how GPs in Brandenburg communicate with other actors and what kinds of barriers they experience with regard to digitalization, they only show part of the practices in communication. Thus, the perspectives of GPs only provide a limited view of communication. Including the perspectives of GPs' direct communication partners may do justice to the complexity of the German health care system by identifying cross-actor problems and offering more comprehensive solutions. This explicitly includes the concrete needs of patients and citizens with regard to digital communication and their views with regard to the installed base.

Conclusions

In the exchange of information, it appears that a number of conditions must be met in order to guarantee the added value of digitalization compared to those exchange formats associated with the installed base from the physician's point of view. To successfully establish digital communication in the long term, patience and technical support should be granted by the political leadership instead of maintaining formal coercive liability and sanctions. Innovations in information exchange should be integrated in a planned, careful, and step-by-step manner, rather than simply being implemented, to reduce the risk of undesirable effects and potential failures [45,46]. From our data, we see that the sole legislative implementation of an infrastructure

XSL•FO
RenderX

component (like KIM) does not automatically result in its integration as a functional component of the installed base. Quite the opposite, it may even have counterproductive effects on resources. Considering these aspects, irritation, frustration, and even resignation could be limited. This implies that new applications should not be rolled out until their functionality is guaranteed, for example, through completed test phases with the help of pilot practices. Furthermore, the perspectives of the health care providers who use the infrastructure to be replaced should be more integrated into the planning of the adjustments [47], with particular attention to individual practice processes. Primary care in Germany is typically very individualized, with communication practices often based on decades-old routines. Digitalization is forcing such individuals, first, to adapt their established practices and routines and, second, to deal with technologies that might be completely alien to them. The frustration that arises here is reasonable and, if it translates into resignation and abandonment of practice, could potentially lead to problems and bottlenecks in care. This should be prevented. GPs play a crucial role in the digitalization of the German health care system. This is due to several factors, including their involvement in a significant portion of patient-related communication, their role in prescribing DHAs [37], and their responsibility for filling electronic health records with data [22]. Facilitating the associated processes for them seems expedient to the overall objective. In the necessary parallel existence of long-established communication channels like postal mail or fax and newer channels such as KIM, arguments need to be found in favor of the latter that do not run counter to individual practice routines.

As has been shown in this paper, especially physicians of older age and with long-established practices need to be given mediating or organizing instances to overcome the individual hurdles of digitalization. This can be done by younger, more technology-savvy colleagues, specially trained practice personnel who take on the digitalization, or other actors or bodies. In this context, a guidance concept is conceivable, where external personnel provide in-house support in primary care practices or comparable settings. Comprehensive continuing education and training programs appear to be necessary, as well as the visualization of links to existing offers. In addition, central quality criteria for training courses should be established in order to provide interested parties with guidance in selecting suitable programs. Aspects of digitalization should also be integrated into the curriculum of medical students. It seems important to identify the specific technical knowledge physicians need to effectively perform their duties.

Finally, it is relevant to remember that barriers to the use of digital communication channels can have different causes. On the one hand, technologies may not operate appropriately because the technology is prone to failure, because users do not possess relevant knowledge, or for other reasons. On the other hand, however, new applications can disrupt the general work routines of GPs, some of which may have been in place for decades and which can be assumed to have meaning and to meet the individual needs of GPs. At this point, the significance of the installed base of communication channels for health care becomes evident beyond their purely technical-material presence. Understanding how communication channels and technologies interact with practice routines can help align newer applications with the needs and demands of GPs, making them more attractive to them. In this way, health care areas that are characterized by a low provider density and a high average age of the population and medical practitioners can benefit in particular.

Acknowledgments

The study is funded by the Brandenburg Medical School publication fund supported by the Ministry of Science, Research and Cultural Affairs of the State of Brandenburg. The analysis of the data collected in the cause of the underlying project—DigiKO—was financially supported by the Economic Development Agency Brandenburg (Wirtschaftsförderung Brandenburg). The authors would also like to thank Josephin Schliephacke for her organizational support in conducting the study. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this paper.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to data protection concerns and the written informed consent but are available from the corresponding author on reasonable request.

Authors' Contributions

TH contributed to writing the original draft, data collection and management, qualitative and quantitative data analysis, participant recruitment, methodology, visualization of tables and graphs, as well as reviewing and editing the manuscript. AH contributed to data collection and management, participant recruitment, methodology, and reviewing and editing the manuscript. KG contributed to data analysis and reviewing and editing the manuscript. MR and JS contributed to reviewing and editing the manuscript. CH contributed to writing the original draft, methodology, as well as supervision.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Good Reporting of a Mixed Methods Study.
[PDF File (Adobe PDF File), 199 KB - jmir_v27i1e65241_app1.pdf]

Multimedia Appendix 2
Overview of themes and subthemes explored in qualitative, semistructured interviews conducted with general practitioners in Brandenburg, Germany (2021-2022).
[DOCX File , 15 KB - jmir_v27i1e65241_app2.docx]

Multimedia Appendix 3
Correlations between frequency of communication channel use in information exchange by general practitioners and sociodemographic factors (age, sex, size of population at practice location, type of practice, and years of practice experience): significant survey results from Brandenburg, Germany (2022), including correlations with communication partners such as pharmacies and medical supply stores, sorted by sociodemographic variable and strength of correlation.
[DOCX File , 17 KB - jmir_v27i1e65241_app3.docx]

Multimedia Appendix 4
Correlations between satisfaction with communication with health care system actors and sociodemographic factors (age, sex, size of population at practice location, type of practice, and years of practice experience): significant survey results from general practitioners in Brandenburg, Germany (2022), sorted by sociodemographic variable and strength of correlation.
[DOCX File , 14 KB - jmir_v27i1e65241_app4.docx]

Multimedia Appendix 5
Correlations between perceived barriers to digital communication and sociodemographic factors (age, sex, size of population at practice location, type of practice, and years of practice experience): significant survey results from general practitioners in Brandenburg, Germany (2022), sorted by sociodemographic variable and strength of correlation.
[DOCX File , 17 KB - jmir_v27i1e65241_app5.docx]

References

1. Diel S, Doctor E, Reith R, Buck C, Eymann T. Examining supporting and constraining factors of physicians' acceptance of telemedical online consultations: a survey study. BMC Health Serv Res 2023;23(1):1128 [FREE Full text] [doi: 10.1186/s12913-023-10032-6] [Medline: 37858170]
2. Sadoughi F, Nasiri S, Ahmadi H. The impact of health information exchange on healthcare quality and cost-effectiveness: a systematic literature review. Comput Methods Programs Biomed 2018;161:209-232. [doi: 10.1016/j.cmpb.2018.04.023] [Medline: 29852963]
3. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2025-02-20]
4. Aanestad M, Jensen TB. Building nation-wide information infrastructures in healthcare through modular implementation strategies. J Strateg Inf Syst 2011;20(2):161-176. [doi: 10.1016/j.jsis.2011.03.006]
5. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
6. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-339. [doi: 10.2307/249008]
7. Taylor S, Todd PA. Understanding information technology usage: a test of competing models. Inf Syst Res 1995;6(2):144-176. [doi: 10.1287/isre.6.2.144]
8. Aanestad M, Grisot M, Hanseth O, Vassilakopoulou P. Introduction. In: Information Infrastructures Within European Health Care: Working with the Installed Base. Cham: Springer International Publishing; 2017.
9. Berg M, Winthereik BR. Waiting for godot: episodes from the history of patient records. In: Health Information Management. London: Routledge; 2003:11-44.
10. Vikkelsø S. Subtle redistribution of work, attention and riskslectronic patient records and organisational consequences. Scand J Inf Syst 2005;17(1):3-30 [FREE Full text]
11. Aanestad M, Grisot M, Hanseth O, Vassilakopoulou P. Information infrastructures and the challenges of the installed base. In: Information Infrastructures Within European Health Care: Working with the Installed Base. Cham: Springer International Publishing; 2017:25-34.
12. Hanseth O, Lyytinen K. Theorizing about the design of information infrastructures: design kernel theories and principles. Sprouts: Working Papers on Information Systems. 2005. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1067&context=sprouts_all [accessed 2025-02-28]

XSL•FO
RenderX

13. Gesetz für sichere digitale Kommunikation und Anwendung im Gesundheitswesen sowie zur Änderung weiterer Gesetze. 2015 Dec 21. URL: https://dip.bundestag.de/vorgang/gesetz-f%C3%BCr-sichere-digitale-kommunikation-und-anwendungen-im-gesundheitswesen-sowie/67134 [accessed 2025-02-20]
14. Gesetz für schnellere Termine und bessere Versorgung (Terminservice- und Versorgungsgesetz—TSVG). 2019 May 10. URL: https://dip.bundestag.de/vorgang/gesetz-f%C3%BCr-schnellere-termine-und-bessere-versorgung-terminservice-und-versorgungsgesetz/240321 [accessed 2022-02-20]
15. Gesetz für mehr Sicherheit in der Arzneimittelversorgung. 2019 Aug 09. URL: https://dip.bundestag.de/vorgang/gesetz-f%C3%BCr-mehr-sicherheit-in-der-arzneimittelversorgung/243802 [accessed 2025-02-20]
16. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz—DVG). 2019 Dec 09. URL: https://dip.bundestag.de/vorgang/gesetz-f%C3%BCr-eine-bessere-versorgung-durch-digitalisierung-und-innovation-digitale-versorgung-gesetz/251761?term=Digitale-Versorgung-Gesetz&rows=25&pos=1 [accessed 2025-02-20]
17. Gesetz zum Schutz elektronischer Patientendaten in der Telematikinfrastruktur (Patientendaten-Schutz-Gesetz—PDSG). 2020 Oct 14. URL: https://dip.bundestag.de/vorgang/gesetz-zum-schutz-elektronischer-patientendaten-in-der-telematikinfrastruktur-patientendaten-schutz-gesetz/260962 [accessed 2025-02-20]
18. Gesetz für ein Zukunftsprogramm Krankenhäuser (Krankenhauszukunftsgesetz—KHZG). 2020 Oct 23. URL: https://dip.bundestag.de/vorgang/gesetz-f%C3%BCr-ein-zukunftsprogramm-krankenh%C3%A4user-krankenhauszukunftsgesetz-khzg/266662
19. Gesetz zur Beschleunigung der Digitalisierung des Gesundheitswesens (Digital-Gesetz—DigiG). 2024 Mar 22. URL: https://www.recht.bund.de/bgbl/1/2024/101/VO.html [accessed 2024-03-22]
20. Gesetz zur verbesserten Nutzung von Gesundheitsdaten (Gesundheitsdatennutzungsgesetz—GDNG). 2024 Mar 22. URL: https://www.recht.bund.de/bgbl/1/2024/102/VO.html [accessed 2025-02-28]
21. Radic M, Donner I, Waack M, Brinkmann C, Stein L. Digitale Gesundheitsanwendungen: Die Akzeptanz steigern [Digital health applications: increasing acceptance]. Deutsches Ärzteblatt 2021;118(6):A-286-B-250 [FREE Full text]
22. Pohlmann S, Kunz A, Ose D, Winkler EC, Brandner A, Poss-Doering R, et al. Digitalizing health services by implementing a personal electronic health record in Germany: qualitative analysis of fundamental prerequisites from the perspective of selected experts. J Med Internet Res 2020;22(1):e15102 [FREE Full text] [doi: 10.2196/15102] [Medline: 32012060]
23. Stachwitz P, Debatin JF. Digitalisierung im Gesundheitswesen: Heute und in Zukunft [Digitalization in healthcare: today and in the future]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2023;66(2):105-113. [doi: 10.1007/s00103-022-03642-8] [Medline: 36648499]
24. Klein S, Schellhammer S. Medication infrastructure development in Germany. In: Information Infrastructures Within European Health Care: Working with the Installed Base. Cham: Springer International Publishing; 2017:151-172.
25. European Commission. Digital Economy and Society Index (DESI) 2022: Human capital. URL: https://digital-strategy.ec.europa.eu/en/policies/desi-human-capital [accessed 2023-08-25]
26. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung [Qualitative Content Analysis: Methods, Practice, Computer Support]. Weinheim, Basel: Beltz Juventa; 2022.
27. Boot: bootstrap R (S-Plus) functions. R Package Version 1.3-28. 2024. URL: https://CRAN.R-project.org/package=boot [accessed 2025-02-25]
28. psych: Procedures for Personality and Psychological Research. 2024. URL: https://CRAN.R-project.org/package=psych [accessed 2025-02-20]
29. readxl: Read Excel files. R Package Version 1.4.3. 2024. URL: https://CRAN.R-project.org/package=readxl [accessed 2025-02-20]
30. Schreier M. Ways of doing qualitative content analysis: disentangling terms and terminologies. Forum Qual Soc Res 2014;15(1). [doi: 10.17169/fqs-15.1.2043]
31. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. Mahwah: Lawrence Erlbaum Associates; 1988:1096.
32. Schendzielorz J, Harre K, Tarara M, Oess S, Holmberg C. Digitale technologien in der allgemeinmedizin [Digital technologies in primary care]. Z Allg Med 2023 Nov 22;100(1):21-29. [doi: 10.1007/s44266-023-00145-w]
33. PraxisBarometer Digitalisierung 2021: Stand und Perspektiven der Digitalisierung in der vertragsärztlichen und -psychotherapeutischen Versorgung. Kassenärztliche Bundesvereinigung. 2021. URL: https://www.kbv.de/media/sp/Schlussfolgerungen_Praxisbarometer_2021.pdf
34. Mühlensiepen F, Petit P, Knitza J, Welcker M, Vuillerme N. Factors associated with telemedicine use among German general practitioners and rheumatologists: secondary analysis of data from a nationwide survey. J Med Internet Res 2022;24(11):e40304 [FREE Full text] [doi: 10.2196/40304] [Medline: 36449333]
35. PraxisBarometer Digitalisierung 2023. Kassenärztliche Bundesvereinigung. 2023. URL: https://www.kbv.de/media/sp/KBV_PraxisBarometer_Digitalisierung_2023.pdf [accessed 2024-03-08]

XSL•FO
RenderX

36. Ergebnisse der bundesweiten Online-Befragung von Niedergelassenen zu aktuellen Erfahrungen mit e-AU und e-Rezept. Kassenärztliche Bundesvereinigung. 2022. URL: https://www.kbv.de/media/sp/2022-08-16_Befragung_eAU_eRezept.pdf
37. Wangler J, Jansky M. How can primary care benefit from digital health applications? A quantitative, explorative survey on attitudes and experiences of general practitioners in Germany. BMC Digit Health 2024;2(1):1-14. [doi: 10.1186/s44247-024-00068-x]
38. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]
39. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
40. Rettinger L, Kuhn S. Barriers to video call-based telehealth in allied health professions and nursing: scoping review and mapping process. J Med Internet Res 2023;25:e46715 [FREE Full text] [doi: 10.2196/46715] [Medline: 37526957]
41. Whitelaw S, Pellegrini D, Mamas M, Cowie M, Van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
42. Sony M, Antony J, Tortorella GL. Critical success factors for successful implementation of healthcare 4.0: a literature review and future research agenda. Int J Environ Res Public Health 2023;20(5):4669. [doi: 10.3390/ijerph20054669]
43. Statistische Informationen aus dem Bundesarztregister. Kassenärztliche Bundesvereinigung. 2021. URL: https://www.kbv.de/html/bundesarztregister.php#content6889
44. Land Brandenburg gesamt—Berufstätige Ärzte—nach Anerkennung und Altersgruppen. Landesärztekammer Brandenburg. 2022. URL: https://www.laekb.de/fileadmin/user_upload/Kammer/Statistik/2021/2021_-_Land_Brandenburg_gesamt_-_Berufstaetige_AErzte_-_nach_Anerkennung_und_Altersgruppen.pdf [accessed 2025-02-28]
45. Grisot M, Hanseth O, Thorseng A. Innovation of, in, on infrastructures: articulating the role of architecture in information infrastructure evolution. JAIS 2014;15(4):197-219. [doi: 10.17705/1jais.00357]
46. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure revisited: an ethnographic case study of how health information infrastructure shapes and constrains technological innovation. J Med Internet Res 2019;21(12):e16093 [FREE Full text] [doi: 10.2196/16093] [Medline: 31855184]
47. Berg M, Timmermans S. Orders and their others: on the constitution of universalities in medical work. Configurations 2000;8(1):31-61. [doi: 10.1353/con.2000.0001]

Abbreviations

DHA: digital health application
GP: general practitioner
KIM: Communication in the Medical Sector
TI: telematics infrastructure

Edited by A Mavragani; submitted 12.08.24; peer-reviewed by C Stallmann, M Aanestad; comments to author 15.10.24; revised version received 25.11.24; accepted 26.12.24; published 24.03.25.

Please cite as:
Holetzek T, Häusler A, Gödde K, Rapp M, Spallek J, Holmberg C
The Role of the Installed Base in Information Exchange Among General Practitioners in Germany: Mixed Methods Study
J Med Internet Res 2025;27:e65241
URL: https://www.jmir.org/2025/1/e65241
doi:10.2196/65241
PMID:

©Tim Holetzek, Andreas Häusler, Kathrin Gödde, Michael Rapp, Jacob Spallek, Christine Holmberg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Dynamics of Doctor-Patient Communication During Remote Consultations: Qualitative Study Among Norwegian Contract General Practitioners

Børge Lønnebakke Norberg[1,2], MD; Bjarne Austad[1], MD, PhD; Eli Kristiansen[2], MSc; Paolo Zanaboni[2], PhD; Linn Okkenhaug Getz[1], MD, PhD

[1]General Practice Research Unit, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway
[2]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Corresponding Author:
Børge Lønnebakke Norberg, MD
General Practice Research Unit
Department of Public Health and Nursing
Norwegian University of Science and Technology
PO Box 8905
Trondheim, 7491
Norway
Phone: 47 73598859
Email: borge.norberg@ntnu.no

Abstract

Background: Patient consultations in general practice are undergoing a digital transformation, embracing diverse modalities such as video, text-based, and telephone consultations. The quality of communication in medical consultations is pivotal for successful outcomes, necessitating a comprehensive assessment of the impact of this transformation on doctor-patient communication and interaction.

Objective: This study aims to explore general practitioners' (GPs') perspectives on how the communication between Norwegian contract GPs and patients has been affected by the large-scale implementation of remote consultations following the onset of the COVID-19 pandemic.

Methods: Five focus groups, comprising 18 purposefully recruited GPs from diverse settings and geographical regions in Norway, were carried out in 2022. We applied thematic analysis guided by the framework proposed by Braun and Clarke.

Results: Six themes resulted from the analysis. First, suitability regarding remote communication is context-dependent: knowing the characteristics of the patient as a person and the clinical relationship is more important than the reason for contact or type of health problem—even more so than during ordinary physical consultations. Second, remote consultations favor a demarcated communication style, "keeping things simple—the one-problem approach," which can increase work effectiveness. Third, a downside of such effective minimalism is that the uncritical use of remote consultations may undermine the quality of care. Communication becomes too transactional, limiting the chances of addressing more implicit and complex issues, with the risk of missing vital information. Fourth, remote modalities can help engage hesitant and vulnerable patients. Fifth, GPs make communicative trade-offs in the name of continuity to be able to maintain relationships with patients they see as vulnerable or fugitive. Finally, there are advantages and dilemmas stemming from text-based consultations. Although they offer benefits such as multimedia-enabled patient expression and sharing of digital information, some concerns include the risk of information loss through triage errors, managing informal language, and ending chat-like interactions between patients and doctors.

Conclusions: The implementation of remote consultations has many effects on clinical interaction and communication. Although these modalities can enhance efficiency, there is a discernible risk of compromised retrieval of essential information and unvoiced problems, potentially resulting in unintended consequences. The preservation of continuity of care emerges as a pivotal strategy to mitigate some of these challenges.

(J Med Internet Res 2025;27:e57679) doi:10.2196/57679

XSL•FO
RenderX

KEYWORDS

remote consultations; digital consultations; telemedicine; eHealth; communication; safety; general practice; family medicine; focus groups; telehealth; digital health; relationship; patient-physician; general practitioner; thematic analysis; qualitative analysis

Introduction

Background

Strong clinical relationships and effective communication during general practitioner (GP) encounters are considered crucial to the overall quality of primary care [1-5]. This importance extends to remote consultations [6-11], which have seen rapid and widespread adoption in response to the challenges posed by the COVID-19 pandemic [12-15]. This global shift has altered the dynamics of medical consultations [16,17] and had a profound impact on the substance and efficacy of communication between GPs and their patients [8,11,18-24].

Remote communication introduces both novel opportunities and challenges in clinical practice [6-11,25-27]. Remote consultations inherently elevate accessibility, offering efficient, environmentally friendly, and positively received solutions for patients [12]. However, the "gold standard" patient-centered communicative approach was developed before the era of remote communication. In other words, the extensive implementation of digital modalities disrupts established practices in ways that have been hard to anticipate [28-33]. Recent research indicates that remote consultations may compromise communicative and diagnostic quality, posing suboptimal clinical evaluation and the risk of overlooking critical information [24-29]. Furthermore, remote communication might also curtail the handling of social determinants of health [8,12,31].

Seen in isolation, each type of remote communication is associated with documented strengths and weaknesses [24-28]. Telephone consultations are deemed appropriate for noncomplex, relatively simple contacts, including triage [34-37], but are susceptible to limitations, including simplistic dialogue, challenges in conveying emotions and sensitive messages, and unsatisfactory clinical assessment [29,32,38-41]. Video consultations, while comparable with telephone consultations in terms of duration, content, and quality [38-41], still present challenges in establishing a doctor-patient relationship, with patients reporting feelings of rush and difficulties in turn-taking in the conversation [8]. Recent publications nevertheless reveal strengthened alliances between patients and physicians in video consultations, compared with the pandemic context [20,21,38,39]. Video consultations are found to be superior to telephone consultations for more demanding triage involving visual evaluation, certain psychological issues, discussing test results, and addressing changes in medication [20-22,38-41]. Text-based consultations, often recognized for their brevity, can reproduce crucial components of face-to-face consultations, encompassing history-taking, investigation, and final assessment [42-46]. Despite their streamlined nature, text-based consultations may generate additional workload for GPs, stemming from unfiltered access and heightened demands [28,46,47]. Some practitioners have noted a decline in the doctor-patient relationship linked to text-based consultations [44-46]. However, studies have also shown instances of strengthened therapeutic relationships [24,45] and improved accessibility over time, especially for individuals dealing with long-term health conditions, attributable to enhanced availability [30,31,42-46].

Much of the existing research on remote consultations stems from contexts where services were only piloted or during the COVID-19 pandemic when use was forced upon both clinicians and patients [12-15]. Contextual factors evidently play a decisive role [48], and consensus on communicative changes or challenges associated with remote consultations remains elusive [8-11,25,45,49,50].

Consequently, it seems essential to continuously observe the impacts of the digital transition on communication dynamics in remote consultations across different contexts, including contractual GP schemes. How will the digital transformation settle into the "new normal" after the resolution of the pandemic [7-9,18,27]? This study examined remote consultations as an integral part of the daily routines of contract GPs in Norway in a more or less normalized postpandemic situation without societal restrictions. The study's aim was to explore the impact of remote consultations on interaction and communication between GPs and patients in this context.

Study Context

Remote Consultations

This study focuses on 3 types of remote consultations: video, telephone, and text-based consultations. Video and telephone consultations are synchronous interactions occurring between doctors and patients on various devices, including phones, tablets, or personal computers. Text-based consultations occur as asynchronously written dialogues between the patient and the doctor, which are initiated or prearranged with the patient. Norwegian GPs are expected to respond to patient-initiated text-based consultations within 5 working days. Text-based consultations normally take the form of a freely formulated essay-type request potentially supplemented by a predefined questionnaire filled out by the patient.

Remote consultations in Norway peaked at nearly two-thirds of all consultations during the COVID-19 lockdown and have since stabilized at 20%-25% of all GP consultations in 2024 [51]. The prevalence of video consultations remains relatively low [52], while telephone and text consultations have sustained high utilization postpandemic, influenced partly by the reimbursement system [53].

In Norway, 80% of all patients use the national health portal Helsenorge [54] for electronic booking of appointments, electronic prescription renewals, electronic contact with the GP office for nonmedical inquiries, and text consultations for clinical inquiries with the GP. The use of remote consultations is not mandatory for GPs, so each GP can decide whether to offer text, video, or telephone. Patients can typically book physical, video, and telephone consultations on the web.

XSL•FO
RenderX

Technical equipment for video consultations is available through both private and public external providers and is becoming more integrated with patient record systems.

The Norwegian Contract GP Scheme

Norway's health care system is grounded in the principles of universal access, decentralization, and continuity of care [55]. Since 2001, all Norwegian citizens have the option to enroll (or change enrollment) with a GP, a choice exercised by 99% of the population. This scheme is designed to ensure continuity of care, with GPs serving as coordinators of municipal services and gatekeepers to specialized services. Norway's 5391 (per May 2024) contract GPs are each responsible for a defined patient list (typically around 1000 citizens or patients) from 8 AM to 4 PM on weekdays and often participate in evening and night emergency shifts in the municipality or region.

Being a specialist GP is a unique specialty in Norway, similar to, for example, ophthalmology or psychiatry. It entails a 5-year program of specified clinical services in accredited practices or institutions, combined with defined learning objectives such as practical procedures, courses, and guidance. The specialty certification must be reaccredited every 5 years [56].

Contract GPs' income systems vary to some extent. The majority receive a fixed annual contribution per patient to help cover practice expenses. Each consultation is reimbursed by a set fee, with additional reimbursement for specified procedures, for example, electrocardiogram testing, surgical procedures, and conversational therapy. Furthermore, GPs receive out-of-pocket fees from patients, typically around 230 NOK (US $21) per consultation (whether physical, text, telephone, or video), up to an annual maximum per citizen for essential medical expenses (currently 3165 NOK = US $287). Expenses above this are fully covered. The provision of remote consultation services is not mandatory for contract GPs, and a small fraction disables 1 or more of these options.

Alongside the well-developed publicly regulated system (contract GP scheme), a few fully private actors offer regular medical care at the patient's expense, and some private digital care companies offer video or text consultations. Private medical insurance is limited in Norway but is increasing.

Methods

Study Design

We conducted a qualitative focus group study among Norwegian contract GPs. Focus groups are deemed suitable for exploring attitudes, experiences, and areas with limited prior knowledge, and they allow for the exploration of tentative and conflicting views [57-59]. The Standards for Reporting Qualitative Research Reporting Guidelines [60] have been consulted in the reporting process (checklist in Multimedia Appendix 1).

Our research team comprised 2 experienced contract GPs (BLN and BA) with substantial experience with remote consultations, a former GP who is now a full-time professor in behavioral sciences in medicine (LOG), and 2 experienced digital health researchers (PZ and EK) with backgrounds in engineering and economics, respectively. LOG, PZ, and BA hold PhDs, with LOG and PZ serving as professors. LOG and BA are seniors in a general practice research unit, introducing solid academic perspectives on general practice, including the documented relevance of continuity of care, patient-centered communication, and relational trust. BLN has extensive experience as a lecturer in clinical communication for medical students. All senior team members had previous experience with qualitative research.

Recruitment Procedure

We used purposive sampling to include experienced contract GPs with familiarity with remote consultations to participate in the focus groups. We approached participants or educational groups by phone and mail from diverse geographical regions, including both urban and remote municipalities, as well as participants with varying income systems in the contract GP scheme (fixed salary, subsidized positions, or paid per capita).

Participants were recruited in 2 ways: 3 groups comprised GPs from preexisting postgraduate educational groups associated with the formal recertification system for GPs in Norway. Such groups are compulsory and meet regularly to engage in structured activities and discussions of general practice topics. These GPs were professionally acquainted beforehand, something we considered to be a facilitating factor. The remaining 2 groups included experienced, opinionated, and nuanced GPs, individually recruited through the research team's professional network.

All but 1 invited GP agreed to participate. Some of the participants were known to members of the research team beforehand, but there were no personal bonds likely to impact the integrity of the data. The initial plan was to conduct 4 focus groups, but a fifth group was added after preliminary analysis of groups 1-4 to further enhance information power [57].

Following the initial invitation, participants received written information outlining the project's purpose, data handling, and withdrawal procedures (Multimedia Appendix 2). No exclusions or dropouts occurred.

Data Collection

The focus group interviews were conducted a while after pandemic restrictions had been lifted in Norwegian health care, between January 15 and April 20, 2022. Three interviews were conducted physically in local medical center meeting rooms, while 2 were carried out remotely on the platform teams, involving GPs from rural regions and other parts of Norway. Interviews were conducted in Norwegian.

Prior to the recorded interview sessions, participating GPs were reminded that participation was voluntary before they signed informed consent (Multimedia Appendix 2). Participants were instructed to carefully anonymize all patient-related experiences. The participants provided anonymized information about themselves and their practice.

During the interviews, moderators BLN and BA used a semistructured interview guide, collaboratively designed by BLN, BA, and LOG, featuring flexible, open-ended questions (guide in Multimedia Appendix 3). The team's previous publications on remote consultations [4,20] have significantly contributed to a robust database of physician responses, which

XSL•FO
RenderX

in turn informed the thematic focus of the interview guide. The guide was pilot-tested with 2 GPs and 2 nonmedical PhD students involved in qualitative research.

The focus group interviews lasted 95-120 minutes. They were audio-recorded and transcribed verbatim. The process remained open for corrections and additions. One follow-up interview provided in-depth insights into the lack of emotional exchange during remote consultations. Participants were compensated approximately 1000 NOK (US $91), which is about half of the compensation for 2 hours of medical consultant work.

Data Analysis

Reflexive thematic analysis, following the approach proposed by Braun and Clarke [57-59], was used to establish patterns of meaning across the data. Initially, LOG and BLN conducted individual analyses: LOG manually and BLN using NVivo (QSR International [Lumivero]) for digital processing. Along the way, BLN and LOG compared and discussed results, followed by joint meetings with BA. The main analytic strategy was inductively derived from the data, albeit influenced by the researchers' professional situatedness. Some elements of the analysis were deductively guided by theory related to the phenomenon of trust. Subsequently, preliminary themes and candidates for final themes were shared with the entire research team, supported by 2 external project collaborators (see the Acknowledgments section). At this stage, theme development underwent a major revision. This paper focuses on microperspectives (here defined as doctor-patient communication), leaving meso- and macroperspectives (ie, addressing the impact of remote consultations on GP offices and the health care system at large) to a forthcoming publication.

In the final stages of analysis and writing, exchanges among all authors aided in challenging individual biases and presuppositions, enhancing the validity and relevance of our findings.

Ethical Considerations

All participants signed an informed consent to participate. This study was conducted among health care workers, involving no patients or sensitive health care information. According to the Norwegian Act on Medical and Health Research §2 and §4, the study did not require approval from the regional ethics committee. Interview data were safely secured in accordance with national and institutional regulations. In the transcription and publication process, personal and demographic information was anonymized to prevent indirect identification of the GPs. The procedure for handling the data was approved by the Norwegian Center for Research Data NSD/SIKT (reference: 531672).

Results

Participant Characteristics

A total of 18 GPs participated in 5 different focus groups, each consisting of 3-5 members. They all had substantial experience using all forms of remote consultations, but 3 had limited their access to text consultations due to high demand. An overview of participant characteristics can be found in Table 1.

Analysis of the focus group discussions on the communicative aspects of remote consultations unveiled a nuanced landscape marked by both advantages and disadvantages, giving rise to 6 themes, summarized in Figure 1.

XSL•FO
RenderX

Table 1. Characteristics of participating GPs.

Number[a]	Age category (years)	Sex	Use of video, telephone, or text consultations	Specialist in general practice	Geographical location	List size (approximate)	Interview format
1	30-40	F[b]	Video, text, and telephone	Under specialization	Rural	500	Digital
2	30-40	M[c]	Video, text, and telephone	Yes	Rural	1000	Digital
3	40-50	F	Video, text, and telephone	Yes	Rural	500	Digital
4	50-60	F	Video, text, and telephone	Yes	Rural	600	Digital
5	50-60	F	Video, text, and telephone	Yes	Urban	1100	Digital
6	50-60	F	Video, text, and telephone	Yes	Urban	1300	Digital
7	40-50	F	Video, text, and telephone	Yes	Urban	1200	Digital
8	50-60	F	Video and telephone; text turned off because of high demand	Yes	Urban	1200	Digital
9	40-50	M	Video, text, and telephone	Yes	Urban	1000	Physical
10	60-70	M	Video, text, and telephone	Yes	Urban	900	Physical
11	40-50	F	Video, text, and telephone	Yes	Urban	800	Physical
12	40-50	F	Video, text, and telephone; limits access to text because of high demand	Yes	Rural-urban fringe	900	Physical
13	40-50	M	Video, text, and telephone	Yes	Rural-urban fringe	1200	Physical
14	40-50	F	Video, text, and telephone	Yes	Rural-urban fringe	900	Physical
15	40-50	M	Video, text, and telephone	Yes	Urban	1500	Physical
16	50-60	M	Video, text, and telephone	Yes	Some urbanity	1300	Physical
17	60-70	M	Video, text, and telephone; limits access to text because of high demand	Yes	Rural-urban fringe	1200	Physical
18	30-40	M	Video, text, and telephone; limits access to text because of high demand	Yes	Rural-urban fringe	800	Physical

[a]Numbers 1-4, 5-8, 9-12, 13-15, and 16-18 represent the 5 different focus groups.

[b]F: female.

[c]M: male.

XSL•FO
RenderX

Figure 1. Overview of the themes developed from 5 focus group discussions. GP: general practitioner.

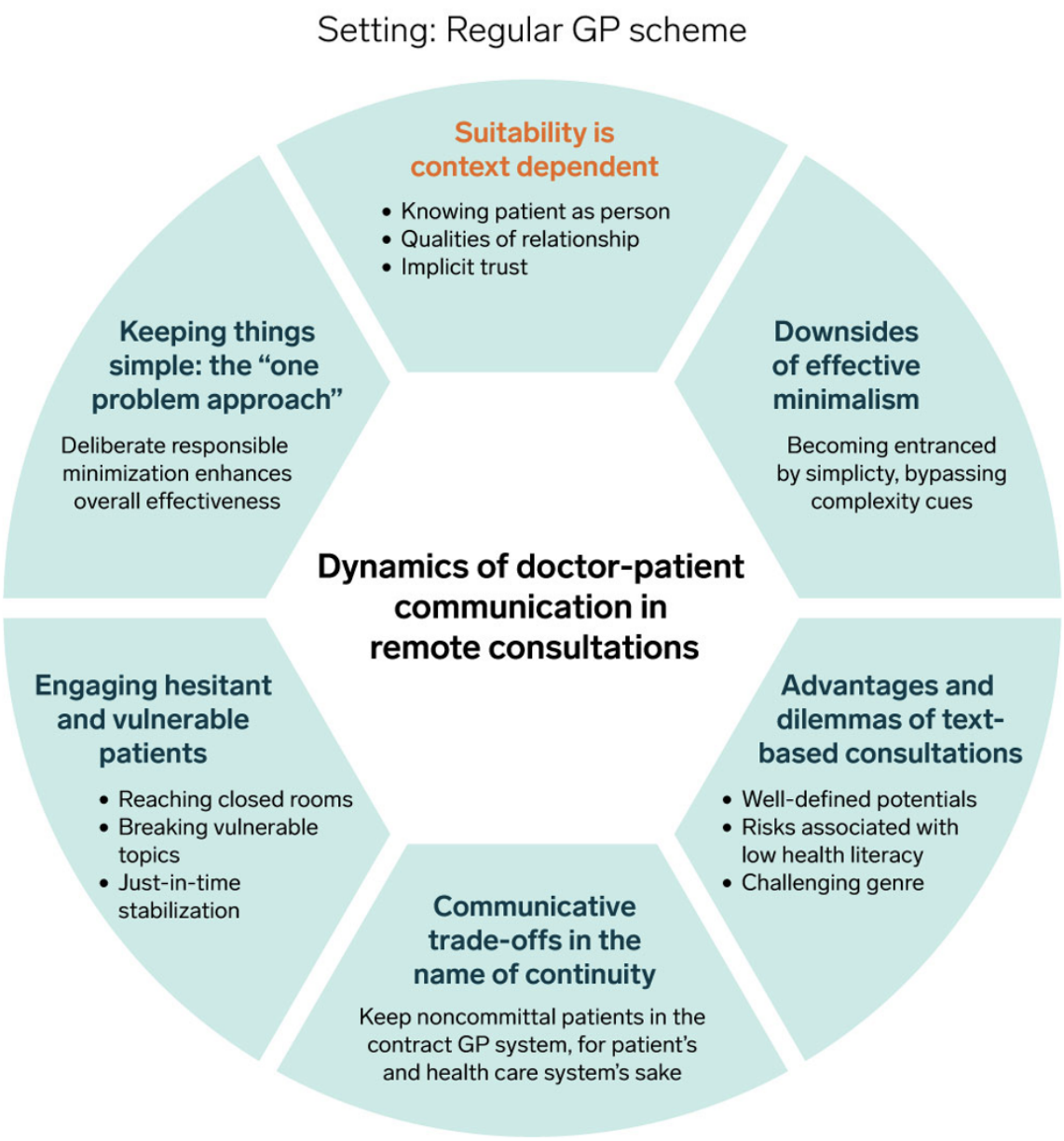

Suitability is Context-Dependent

A dominant view in all focus groups was that the applicability of remote consultations to communication between GPs and patients is largely independent of the reason for contact or specific diagnoses. Suitability largely depends on the doctor-patient relationship, the doctor's knowledge of the patient's medical history, and the patient's ability to formulate his or her concerns concisely. In many ways, this theme has an overarching character, connecting and informing the subsequent themes that manifested from our analysis, as shown in Figure 1.

> *When you know the patient well, it is easier to agree on a plan...Background knowledge means a lot for an assessment.* [ID6]

One GP quoted the influential physician Sir William Osler (1848-1919) [61]: "It is much more important to know what sort of a patient has a disease than what sort of a disease a patient has."

The sentiment among GPs resonated with the idea that:

> *Topics in consultations are often interconnected, mutually influencing each other [...]. A troubled marriage can manifest in several ways, and the need for sleeping pills may conceal significant underlying information.* [ID16]

The participants emphasized that both verbal communication and clinical examination and treatment via video were surprisingly feasible, especially when relational trust had been established between the GP and the patient. For example, a GP who had previously prescribed medication to a young patient

XSL•FO
RenderX

with chronic urticaria recounted how it was easy to calm and instruct her over video as she needed to inject the drug:

> *She posed in shorts on her kitchen chair, and she was able to insert this medication subcutaneously without any problem.* [ID5]

Keeping Things Simple—the "One-Problem Approach"

The GPs seemed to agree that remote communication has some fundamental qualities that differ from physical communication. Without any specific preparations or instructions from the GP's side, the modality appeared to narrow things down to what the GPs called a "one-problem approach." This phenomenon could often be perceived as beneficial:

> *The best thing about remote consultations is that they go quickly and smoothly, and you stick to the problem. Concise...it stops everything outside the original theme.* [ID9]

Furthermore, participants described that remote consultations, in general, appeared suitable for (presumably) demarcated problems with negligible risk of missing clinically relevant issues or risk factors. Therefore, they can be performed more quickly. In such instances, GPs experienced that remote consultations represented "good enough" medical work from a here-and-now perspective.

A benefit of keeping things simple relates to worried patients who have a habit of overseeking regular consultations. GPs have noted that even brief remote consultations can "work miracles" (quote from a participating GP) for these patients and reduce the number of time-consuming physical consultations.

Remote communication tends to have a transactional quality that seems to bypass implicit expectations of relational care. Once a distressed patient has waited for a physical consultation and perhaps traveled a long way, expectations can build up, and the GP can sense an obligation to spend more time with the patient as a sign of care and respect. This phenomenon rarely applied to remote consultations:

> *When you take that video, it is a bit more acceptable to...keep simple things simple.* [ID2]

However, the benefits of this approach were balanced by concerns about potential downsides, such as overlooking clinically significant aspects.

Downsides of Effective Minimalism

Consultations characterized by the typical "one-problem approach" apparently entailed diminished cognitive and emotional demands on the GP in a here-and-now perspective, thereby contributing to welcomed work intensity variation during the day. However, the discussions of new potentials did not last long in any of the focus groups before a certain ambivalence and skepticism started to surface. GPs described a risk of being carried away by apparent simplicity, missing clinically significant aspects of their everyday work as they knew it from years of physical consultations. Uncritical use of remote consultations may undermine the quality of care:

> *It can easily take off in a spiral of efficiency—and then it quickly becomes a dangerous disadvantage.* [ID11]

The discussions revealed that the most evident pitfalls of remote consultations represent the flip side of the reported advantages. Participants described how the experienced efficiency could turn into a spiral of unintended minimalism:

> *If the GP makes too much use of remote aids and don't take consultations physically...one can imagine that the doctor fixes things (way) too quickly….* [ID8]

At a more specific level, several GPs were worried that in the flow of efficiency, relevant and essential information associated with the patient's reason for contact remained out of sight and unaddressed:

> *You might miss these things on remote consultations. It can be central to realize that there is something else …something unsaid or unwritten.* [ID17]

GPs shared experiences pertaining to typical communicative qualities of remote consultations. Compared with physical encounters, communication becomes more linear, literal, and business-like. Information residing between the verbal lines seems to go unnoticed or remain unaddressed:

> *It is like an interpreter call…because it has become such an additional filter in between.* [ID14]

Keywords regarding what might easily be lost during remote consultations were context and cues, referring to relational and emotional factors that affect the symptom picture and have relevance for effective resolution:

> *Have you noticed? No one ever cries on video! ….* [ID16]

If contextual insight and emotional cues go unnoticed, the chance of an effective problem solution diminishes. This hampers the GPs' ability to reveal serious illness. It is, therefore, not just about the number of problems that are lost, they elaborated, but more about the range of dimensions lost in communication:

> *I had a patient on video who had difficulty sleeping and wanted sleeping pills...I did not prescribe them. I rather wanted to find out what his life was really like and invited him to have a proper [physical] conversation […] Remote consultations may also generate 4-5 extra consultations for extending sick leave because the patient does not want to talk about his terrible boss on a remote consultation.* [ID16]

GPs recognized a distinct ritual associated with traditional, in-person consultations—a ritual that, for better or worse, imbues the medical encounter with a sense of uniqueness. Traditional health care requires patients to travel, wait, meet the doctor, and possibly undergo physical tests. Remote consultations eliminate these steps, saving time and seamlessly fitting into the patient's daily routine. Participants elaborated that in remote GP appointments, patients often seem less mentally focused. Furthermore, such consultations are often initiated for minor complaints. These are seen as potential

XSL•FO
RenderX

drawbacks. To minimize these drawbacks, some GPs aimed to meet patients physically at an early stage:

> *I like to start with a physical consultation to sense the non-verbal clues and include the big context.* [ID15]

In relation to video consultations, GPs described patients' casual use of mobile cameras, simultaneously multitasking during all kinds of transportation and errands:

> *It is quite common, we have video consultations while the patient is driving, lies in bed or is all over the place.* [ID14]

GPs exhibited ambivalence toward patients' multitasking while in remote consultations, experiencing a spectrum from frustration to recognizing occasional valuable moments. However, they explicitly appreciated their own multitasking opportunities. The pandemic prompted a shift in how GPs use eye contact during video consultations, transitioning from its prioritization to a more flexible approach amidst multitasking demands. This adaptation mirrored the younger generation's informal use of social media, offering GPs a sense of freedom while acknowledging both the advantages and drawbacks.

The traditional face-to-face consultation was seen to strengthen the therapeutic alliance and give the GP a better understanding of the totality of the patient as a person. Overall, it appeared that most GPs remained alert to the potential side effects of "pleasing the patients" too much. Offering all patients who are reluctant to attend a physical consultation a choice of remote consultations as an easy way out might ultimately undermine these patients' abilities to engage in trustful human relations and face the challenges of everyday life in a physical community. Still, none of the focus groups reached a clear consensus as to when or how the physical encounter is clearly to be preferred for therapeutic reasons.

Engaging Hesitant and Vulnerable Patients

In all focus groups, GPs discussed how remote consultations may facilitate the establishment of new relationships. Examples include socially withdrawn patients and young people with high digital competence who have little experience with defining their own problems and seeking formal health care:

> *I have several patients with autism-like diagnoses or personality disorders with relational issues, who prefer to communicate remotely—by text, telephone, or video—rather than sitting in the waiting room for a physical appointment. They communicate better and seem to profit from these new services.* [ID13]

Another communicative gain was referred to by some participants as an "ice-breaking" effect, facilitating the sharing of sensitive information that, for some patients, could be harder to address in a face-to-face consultation:

> *...I do not necessarily think they would have been able to say these things when they were sitting here, right in front of me in a chair. So, sometimes that distance on video can make it easier to bring things up, eventually leading to a physical consultation.* [ID3]

Specific examples of sensitive subjects that might sometimes emerge due to, and not despite, the digital modality involve histories of sexual abuse:

> *It was difficult regarding trust initially…. But then a nice little dialogue started on text-consultation. A few small questions. So, in a way, I understand that she was trying to say something important [about abuse]. And then we go on…finally arriving at what was really the case. And this is a new unique way of working.* [ID14]

Some of the GPs had experienced that remote consultations were used by patients with suicidal thoughts that might not otherwise contact the health care system:

> *I talked to him in the car a good distance from his home and office...When he talked about his suicidal thoughts and how he might end up taking his own life, he liked to sit alone in his car. So, in a way, he set the conditions for being protected…* [ID9]

Communicative Trade-Offs in the Name of Continuity

Although the GPs agreed that communication during remote consultations entailed a risk of losing vital symptoms and unvoiced problems, they shared experiences of how they would sometimes deliberately stretch their limits and agree to perform remote consultations simply to avoid losing young, immature, or otherwise vulnerable, noncommittal patients to competing commercial digital doctors. The GPs thought that such actors might not provide adequate, affordable, and high-quality services for complex patients in the long term because they would lack continuity of care. As one GP said about such patients, with reference to the contract GP scheme:

> *I work hard to "keep them in the system.*" [ID7]

An additional reason why the GPs tried to avoid the involvement of other digital doctors was that such consultations tended to create double work and even more additional work, as the contract GP might typically be contacted anyway and have to sort out the status of a process initiated by an unknown colleague:

> *I would rather help them myself than get help from someone else in the private sector. If that happened, the patients typically say they did not understand anything, so they still come to talk with me.* [ID6]

Ideally, as already mentioned, many of the GPs prefer to see shy and reluctant patients, many of whom have a psychiatric diagnosis, physically with the aim of fostering relational trust and a better working relationship:

> *[I prefer]...making them leave the house, so that they do not just stay at home and isolate themselves. So, it is better to have psychiatric follow-up consultations at the medical center.* [ID18]

However, the GPs consider "good enough" remote medical care by a contract GP to represent an acceptable compromise, if the patient specifically prefers it, to maintain some sort of relationship.

XSL•FO
RenderX

Advantages and Dilemmas of Text-Based Consultations

On the positive side, GPs discussed how text-based consultations have introduced new possibilities. The patient can freely describe his or her reason for contact and include illustrative photographs or videos. In situations with a risk of disease transmission, this can be helpful:

> *We use that quite a lot; text-based consultations for typical wounds, rashes, and similar things. Most recently I had a chicken pox diagnosis last week. It is nice not to have children with chicken pox in the waiting room.* [ID5]

Attachment of photographs or videos was also described as effective in the follow-up of a healing process and rehabilitation. GPs can, on their side, use text-based consultations to attach files with ready-made patient information and other instructive material at any stage of the clinical trajectory. They can also ask patients to fill in various forms digitally, including mental health diagnostic checklists or scales, symptom diaries, or text for declarations required by the Norwegian Welfare System, as well as adding self-written parts in the patient record:

> *That is such a "game changer" for me. It is an incredibly clever new way of doing it, which also makes the patient quite responsible. "Ok, sit down and think about it and write until next time 'or' if you are having a headache, write down all about it from the beginning till now."* [ID13]

The group dynamics then prompted deliberations on more negative aspects, particularly the dilemma regarding the utilization of predefined questionnaires. These questionnaires encompass a series of clinical inquiries presented in a flowchart format intended to streamline the process, reduce dependence on health secretaries for triage, and aid doctors in promptly addressing the patient's reason for contact. The general tone was, however, marked by skepticism and call for precaution:

> *This development [with fixed questions] is happening quickly and quietly without anyone knowing the long-term consequences.* [ID18]

Participating GPs referred to a well-documented communicative routine in physical, effective patient-centered communication, namely, to invite patients to freely formulate their reason for contact before zooming in on further clinical details:

> *Letting the patient express him/herself freely is preferable [...] and prefixed algorithmic questions should be used with the greatest caution. It takes away the information we are so used to handling; listening to what the patient says.* [ID4]

Losing out on the patient's unfiltered, personal account at the beginning of a consultation might lead to time-consuming diversions and a need to revisit the reason for contact at a later stage. In other words, this created more work for the GP.

In text-based consultations, some patients exaggerate to get attention and skip the line:

> *Some patients write dramatically as if it were an emergency, yet there is not...* [ID15]

Another issue of concern related to text-based consultations was that some patients use rough language characterized by profanity and exaggerations, resembling the comment field jargon in social media. GPs worry because the communication is automatically saved in the patient's medical record, with a potential for unforeseen future implications:

> *There is a lot of dialect, slang, profanity, and things that are not suitable for the medical record...They can write anything. There are no restrictions.* [ID12]

Beyond the fear of missing serious disease in a flow of incoming requests, some GPs had noted that informal chat-like text-based consultation sequences back and forth between patient and doctor can be difficult to terminate:

> *It is difficult with text consultations because it turns out to be a running dialogue that goes back and forth all the time, like a tennis match...which I could have ended in a stricter way, but I find it difficult to do it in a gentle way.* [ID11]

Considering all these factors, participants suggested that remote consultations might augment the GP's workload.

Discussion

Principal Findings

This focus group study among Norwegian contract GPs shows that the rapid implementation of remote consultations elicited diverse effects on clinical communication and interaction for better and for worse. The appropriateness of remote communication appears contingent on context; recognizing the patient's characteristics and the clinical relationship takes precedence over the reason for contact or the nature of the health problem. Although entailing a definite potential to enhance efficiency, remote modalities also pose a discernible risk of compromising the retrieval of crucial information and unarticulated problems with clinical relevance. Nonetheless, a somewhat unexpected advantage of remote consultations lies in their potential role as relational "icebreakers," as some patients find it easier to seek help from their GP and broach sensitive topics from a distance than in face-to-face encounters. Text consultations evoke ambivalence among GPs. Although being of beneficial use in relation to clear reasons for contact, their safety hinges on patients' ability to articulate their complaints. Strategic preservation of continuity of care emerges as crucial in mitigating the identified challenges.

Findings in Light of Existing Literature

Context-Dependent Suitability

A central finding in our study is that the observed changes in communication during remote consultations are not primarily associated with specific reasons for contact. Instead, they are more influenced by the quality of the doctor-patient relationship and the GP's evaluation of the patient's individual ability to express health concerns coherently. The context-dependent qualities of remote consultations resonate with the conclusion of health care researcher Greenhalgh et al [62] that dilemmas about establishing remote consultation services cannot be resolved by standard procedures or algorithms but rather by

XSL•FO
RenderX

attending to here-and-now practicalities applied with contextual judgment. Our findings also fit with a recent large in-depth study of patient experiences of remote consultations, stating that it can be influenced by previously unreported patient characteristics and the conditions they consult about [63].

Optimizing Communication Styles and Workflow

Our research indicates that remote communication can enhance workflow and information gathering, streamlining a portion of the tasks encountered in a doctor's daily routine. This aligns with previous research highlighting convenience and time savings offered, especially when physical examination is unnecessary [20-22,63]. Our participating contract GPs reported how, during normal investigations, they could leverage patient-provided information, symptom-scoring forms, and other digitalized documentation more effectively than previously documented [48,52]. However, we also found that remote consultations could potentially increase the overall workload, in accordance with recent research findings [28,33,55]. In total, this paints a nuanced perspective on workflow dynamics and applicability. Ideally, GPs can blend various modalities, leveraging their respective strengths, integrating the straightforward "one-problem approach" with a more comprehensive exploration of the complexity and wider context as needed.

Communicative Threats to Quality of Care

Alongside the aforementioned benefits, we identified concerns about unintended loss of contextual information, which is crucial for gaining a comprehensive understanding of more complex health problems. Disregarding life stressors and sociocultural factors might result in degraded clinical decision-making, imprecise or missed diagnoses, and underestimation of severity, findings that expand on prior research [63-69]. This is consistent with research suggesting that remote communication might lead to heightened clinical uncertainty [28-30]. In summary, these challenges suggest that remote communication not only tends to address fewer issues but may also address each problem with a narrower focus, overlooking certain communicative cues and contextual details. Interestingly, this aligns with patients' perspectives, which highlights the lack of nonverbal communication [26]. Our research also corresponds with a new "masterclass" [23] on communication skills for remote health care, including openings, active listening, and closing nuances, recognizing the differences from face-to-face consultations.

Previous studies have highlighted the constraints of remote consultations in dealing with social determinants of health [12,31,68]. Our results, in particular GPs' reports on "the one-problem approach," indicate that unless the GP actively counteracts the inclination to "narrow down," less urgent but still significant matters on chronic diseases might go unaddressed. This situation calls to mind a noteworthy 1979 paper by GP academic Nigel Stott, who highlighted the distinctive opportunities for including the management of continuing problems and providing some opportunistic preventive health care during routine consultations initiated for other reasons [69]. Our findings suggest an untapped potential for preventive health care on digital platforms. In particular, text consultations give patients more time to formulate nuanced questions, making them well-suited for meeting the needs of patients with complex and chronic conditions over time, in accordance with previous studies [44,45]. Combining informative materials from reputable health libraries with personalized guidance might enhance understanding and patient engagement. However, effective written consultations rely on patients' ability to articulate their concerns [45].

Simultaneously, GPs in our study describe challenges associated with unfiltered, unfocused, and even verbally profane written patient accounts. Patients may invertedly divert their problems away from the most adequate management, delaying the detection of serious conditions and limiting care planning [25-29]. In addition, our contract GPs reported making concessions to "please the customer" when selecting remote modalities as they endeavor to maintain the clinical relationship with vulnerable patients. Overall, our findings underscore the importance of considering precautionary measures to uphold quality in the evolving landscape of remote consultations.

Notably, we found that contract GPs, after having acquired substantial experience with remote consultations, describe themselves as more observant of the multidimensional aspects of communication in traditional physical consultations. They underscore the importance of safeguarding these qualities in the future.

Facilitating Relationships With Reluctant Patients

Up to this point, face-to-face consultations have been considered essential for establishing workable doctor-patient relationships [40,70]. Our study confirms the importance of physical consultations but provides a more nuanced picture. Under certain circumstances, remote consultations can foster new alliances with reluctant patients in need of health care who might otherwise not consult at all. Remote consultations can also function as icebreakers, giving patients an opportunity to open sensitive topics. Examples mentioned included histories of sexual abuse and suicidal thoughts that the patient finds hard to share with the GP in a physical encounter. Furthermore, remote consultations can offer expedited, just-in-time relief to well-known, vulnerable, and unstable patients. In these cases, remote communication seems to consolidate trust over time [71-74]. This also covers text-based consultations, extending the insights from prior literature [32,44]. The findings confirm emerging patient perspectives on text consultations, where attitudes vary widely on their suitability for handling sensitive topics [64].

It has been argued that youth and young adults, groups potentially underusing health services for psychosocial challenges, can benefit from remote consultations, although such consultations have been found to be of shorter duration than physical consultations [68,74]. Our GPs note how consultations with youth often proceed in a somewhat casual manner, with a handheld camera, sometimes on the move, bringing to mind the jargon term "snapping." Similarly, the discourse in text-based consultations with young individuals may "ping-pong" back and forth, resembling a conversational chat. Despite such informality, the GP-patient relationship may derive advantages. We propose that the rise in informality during remote communication, driven by technological advancements

XSL•FO
RenderX

and societal shifts, attracts new user groups to these modalities. Health care professionals adopt this approach to connect with these specific patient groups.

Power, Trust, and Risk

Norwegian philosopher Grimen [71,72] wrote substantially about trust as a fundamentally important phenomenon in clinical relationships, describing a "nexus" of power, trust, and risk. Despite publishing before the era of remote consultations, he underlined that health care is constantly undergoing change and invited others to reflect upon how changes impact the power-trust-risk nexus. We have already discussed how remote consultations might offer new opportunities in relation to certain reluctant and vulnerable patients. At first glance, remote consultations can be seen as empowering patients in general, offering "consumer" freedom. However, if not deliberately chosen as appropriate, the remote mode may undermine the quality of the GP's work and ultimately damage trust. Contrary to some prior research [9,64], our findings imply caution in granting patients unrestricted choice regarding consultation type.

Strength and Limitations

The research team, composed of seasoned contract GPs and digital health researchers, ensured a robust and multifaceted approach to the study. Another key strength was the recruitment and participation of experienced GPs representing diverse demographics, including gender, age groups, and remuneration systems. Their practices spanned urban and rural regions across Norway, contributing to a comprehensive understanding of remote consultations in different settings. This purposive sampling provided a wide range of perspectives. Furthermore, conducting focus groups proved particularly effective for exploring attitudes, experiences, and areas with limited prior knowledge. This methodology facilitated the exploration of tentative and conflicting views, allowing for nuanced reflections, examples, and insights. The use of semistructured interview guides, detailed transcription, and reflexive thematic analysis contributed to the depth and reliability of the findings. We believe that we have sufficient information power to answer our research questions [75].

However, one limitation is that the study did not address potential challenges related to ethnic and cultural diversity among patients and doctors [64]. This omission may be partly due to demographic realities in Norway and the tendency of dedicated regular GPs to avoid repeated digital consultations with patients if they encounter language problems or other contextual barriers, thereby having acquired few significant experiences. We also have sparse data from young GPs; this could have added nuance to our results. The study might also have benefited from the formal inclusion of citizen or patient user perspectives [41].

Potential bias from a few participants' familiarity with the research team may have influenced responses, yet it facilitated more open dialogue. Also, contrary to expectations that remote interviews would reduce data richness, they appeared to yield engaging interactions. The study was not designed to evaluate each consultation modality (ie, video, text-based, and telephone consultations) in depth. It is important to note that the study was conducted during the aftermath of the pandemic, and it is questionable whether enough time had elapsed to establish "the new normal" [12,13]. During the study period, GPs were still receiving full remuneration for telephone consultations, which might have influenced their choices of remote modality. However, since the telephone has long been a contact medium in general practice, participants appeared particularly interested in discussing novel modalities, such as video and text consultations. Finally, the study focused on microperspectives of doctor-patient communication, leaving broader systemic impacts to be addressed in future research.

Conclusions

The implementation of remote consultations introduces a spectrum of effects on clinical interaction and communication. These modalities can enhance efficiency but also pose risks, such as compromised retrieval of essential information and unvoiced problems, potentially leading to unintended consequences. Preserving continuity of care is crucial to mitigate these challenges.

Knowledge of the patient and the context plays a crucial role in determining when remote consultations are appropriate. For some patients, remote communication may even contribute to building trustful alliances and enhancing the doctor-patient relationship. Future research could delve into the possibilities of proactive health care initiatives during remote consultations and examine the broader ramifications they have on physicians, medical practices, and societal dynamics.

Acknowledgments

The authors want to thank professors Chris Salisbury and Helen Atherton for important advice in the analysis and writing process. The paper is part of a PhD project funded by the Research Council of Norway (315404). Norberg also holds a part-time position at the Norwegian Centre for eHealth Research in Tromsø.

Data Availability

The datasets generated and analyzed during this study are not publicly available, but access to the anonymized material can be granted upon reasonable request.

XSL•FO
RenderX

Disclaimer

ChatGPT was, in some instances, consulted to enhance linguistic or grammatical clarity but was in no instance used to generate original text.

Authors' Contributions

All authors were involved in the design of the study. BLN developed the interview guide together with BA and LOG. BLN and BA moderated the focus group interviews. BLN and LOG developed initial themes, gradually involving BA. PZ and EK contributed substantially to data analysis, interpretation of results, and writing. BLN wrote the first draft of the manuscript, while BA and LOG contributed to all phases of the writing process. All authors contributed actively to the later stages of the writing process and have accepted the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ checklist.
[DOCX File , 427 KB - jmir_v27i1e57679_app1.docx]

Multimedia Appendix 2
Information and consent scheme/application for privacy and anonymity (same text used for both purposes; translated from Norwegian).
[DOCX File , 42 KB - jmir_v27i1e57679_app2.docx]

Multimedia Appendix 3
Interview guide (translated from Norwegian).
[DOCX File , 34 KB - jmir_v27i1e57679_app3.docx]

References

1. Burt J, Abel G, Elmore N, Newbould J, Davey A, Llanwarne N, et al. Rating communication in GP consultations: the association between ratings made by patients and trained clinical raters. Med Care Res Rev 2018;75(2):201-218 [FREE Full text] [doi: 10.1177/1077558716671217] [Medline: 27698072]
2. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
3. Sandvik H, Hetlevik Ø, Blinkenberg J, Hunskaar S. Continuity in general practice as predictor of mortality, acute hospitalisation, and use of out-of-hours care: a registry-based observational study in Norway. Br J Gen Pract 2022;72(715):e84-e90 [FREE Full text] [doi: 10.3399/BJGP.2021.0340] [Medline: 34607797]
4. Norberg BL, Johnsen TM, Kristiansen E, Krogh FH, Getz LO, Austad B. Primary care gatekeeping during the COVID-19 pandemic: a survey of 1234 Norwegian regular GPs. BJGP Open 2024;8(1):BJGPO.2023.0095 [FREE Full text] [doi: 10.3399/BJGPO.2023.0095] [Medline: 37907336]
5. Chipidza FE, Wallwork RS, Stern TA. Impact of the doctor-patient relationship. Prim Care Companion CNS Disord 2015;17(5). [doi: 10.4088/pcc.15m01840]
6. Li E, Tsopra R, Jimenez G, Serafini A, Gusso G, Lingner H, et al. General practitioners' perceptions of using virtual primary care during the COVID-19 pandemic: an international cross-sectional survey study. PLOS Digit Health 2022;1(5):e0000029 [FREE Full text] [doi: 10.1371/journal.pdig.0000029] [Medline: 36812543]
7. Speakman EM, Jarvis H, Whiteley D. Opportunities and risks within the expanding role of general practice. Br J Gen Pract 2021;71(709):344-345 [FREE Full text] [doi: 10.3399/bjgp21X716489] [Medline: 34326075]
8. Mughal F, Atherton H, Awan H, Kingstone T, Poppleton A, Silverwood V, et al. The impact of remote consultations on brief conversations in general practice. BJGP Open 2022;6(2):BJGPO.2021.0199 [FREE Full text] [doi: 10.3399/BJGPO.2021.0199] [Medline: 35217511]
9. Payne R, Clarke A, Swann N, van Dael J, Brenman N, Rosen R, et al. Patient safety in remote primary care encounters: multimethod qualitative study combining safety I and safety II analysis. BMJ Qual Saf 2024;33(9):573-586 [FREE Full text] [doi: 10.1136/bmjqs-2023-016674] [Medline: 38050161]
10. Payne RE, Clarke A. How and why are video consultations used in urgent primary care settings in the UK? A focus group study. BJGP Open 2023;7(3):BJGPO.2023.0025-BJGPO.2023.1-11. [doi: 10.3399/BJGPO.2023.0025] [Medline: 37068795]
11. Gordon HS, Solanki P, Bokhour BG, Gopal RK. 'I'm Not Feeling Like I'm Part of the Conversation.' Patients' perspectives on communicating in clinical video telehealth visits. J Gen Intern Med 2020;35(6):1751-1758 [FREE Full text] [doi: 10.1007/s11606-020-05673-w] [Medline: 32016705]

XSL•FO
RenderX

12. General practice in the post Covid world: challenges and opportunities for general practice. Royal College of General Practitioners. 2020 Jul 11. URL: https://tinyurl.com/4hr83ejj [accessed 2025-03-25]
13. Hughes G, Moore L, Maniatopoulos G, Wherton J, Wood GW, Greenhalgh T, et al. Theorising the shift to video consulting in the UK during the COVID-19 pandemic: analysis of a mixed methods study using practice theory. Soc Sci Med 2022;311:115368 [FREE Full text] [doi: 10.1016/j.socscimed.2022.115368] [Medline: 36152402]
14. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
15. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
16. Alsnes IV, Munkvik M, Flanders WD, Øyane N. How well did Norwegian general practice prepare to address the COVID-19 pandemic? Fam Med Community Health 2020;8(4):e000512 [FREE Full text] [doi: 10.1136/fmch-2020-000512] [Medline: 33277356]
17. Renaa T, Brekke M. Restructuring in a GP practice during the COVID-19 pandemic - a focus-group study. Tidsskr Nor Laegeforen 2021 Feb 02;141(2):1-11 [FREE Full text] [doi: 10.4045/tidsskr.20.0713] [Medline: 33528128]
18. Greenhalgh T, Rosen R. Remote by default general practice: must we, should we, dare we? Br J Gen Pract 2021;71(705):149-150 [FREE Full text] [doi: 10.3399/bjgp21X715313] [Medline: 33771790]
19. Meurs M, Keuper J, Sankatsing V, Batenburg R, van Tuyl L. 'Get Used to the Fact That Some of the Care Is Really Going to Take Place in a Different Way': general practitioners' experiences with e-health during the COVID-19 pandemic. Int J Environ Res Public Health 2022;19(9):5120 [FREE Full text] [doi: 10.3390/ijerph19095120] [Medline: 35564519]
20. Norberg NL, Getz LO, Johnsen TM, Austad B, Zanaboni P. General practitioners' experiences with potentials and pitfalls of video consultations in primary care: qualitative study. JMIR Form Res 2022;25(3):e28379. [doi: 10.2196/28379]
21. Elliott T, Tong I, Sheridan A, Lown BA. Beyond convenience: patients' perceptions of physician interactional skills and compassion via telemedicine. Mayo Clin Proc Innov Qual Outcomes 2020;4(3):305-314 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.04.009] [Medline: 32542222]
22. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
23. Roberts LC, Osborn-Jenkins L. Delivering remote consultations: talking the talk. Musculoskelet Sci Pract 2021;52:102275 [FREE Full text] [doi: 10.1016/j.msksp.2020.102275] [Medline: 33132068]
24. Dixon S, Frost L, Feder G, Ziebland S, Pope C. Challenges of safeguarding via remote consulting during the COVID-19 pandemic: a qualitative interview study. Br J Gen Pract 2022;72(716):e199-e208 [FREE Full text] [doi: 10.3399/BJGP.2021.0396] [Medline: 35074797]
25. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020;4(1):01-08 [FREE Full text] [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
26. Verma P, Kerrison R. Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. BJGP Open 2022;6(2):1-16 [FREE Full text] [doi: 10.3399/BJGPO.2021.0192] [Medline: 35031558]
27. Neves AL, Li E, Serafini A, Jimenez G, Lingner H, Koskela TH, et al. Evaluating the impact of COVID-19 on the adoption of virtual care in general practice in 20 countries (inSIGHT): protocol and rationale study. JMIR Res Protoc 2021;10(8):e30099 [FREE Full text] [doi: 10.2196/30099] [Medline: 34292867]
28. Turner A, Morris R, Rakhra D, Stevenson F, McDonagh L, Hamilton F, et al. Unintended consequences of online consultations: a qualitative study in UK primary care. Br J Gen Pract 2022;72(715):e128-e137 [FREE Full text] [doi: 10.3399/BJGP.2021.0426] [Medline: 34903520]
29. Rosen R, Wieringa S, Greenhalgh T, Leone C, Rybczynska-Bunt S, Hughes G, et al. Clinical risk in remote consultations in general practice: findings from in-COVID-19 pandemic qualitative research. BJGP Open 2022;6(3):1-11 [FREE Full text] [doi: 10.3399/BJGPO.2021.0204] [Medline: 35487581]
30. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
31. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
32. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021;8(2):e28589 [FREE Full text] [doi: 10.2196/28589] [Medline: 33822736]
33. Salisbury H. Helen Salisbury: E-consultations are increasing the GP workload. BMJ 2021;375:n2867. [doi: 10.1136/bmj.n2867] [Medline: 34815221]

XSL•FO
RenderX

34. McKinstry B, Watson P, Pinnock H, Heaney D, Sheikh A. Telephone consulting in primary care: a triangulated qualitative study of patients and providers. Br J Gen Pract 2009;59(563):e209-e218 [FREE Full text] [doi: 10.3399/bjgp09X420941] [Medline: 19520019]
35. Pope C, McKenna G, Turnbull J, Prichard J, Rogers A. Navigating and making sense of urgent and emergency care processes and provision. Health Expect 2019;22(3):435-443 [FREE Full text] [doi: 10.1111/hex.12866] [Medline: 30632242]
36. Huibers L, Smits M, Renaud V, Giesen P, Wensing M. Safety of telephone triage in out-of-hours care: a systematic review. Scand J Prim Health Care 2011;29(4):198-209 [FREE Full text] [doi: 10.3109/02813432.2011.629150] [Medline: 22126218]
37. Bryce C, O'Connell MD, Dale J, Underwood M, Atherton H. Online and telephone access to general practice: a cross-sectional patient survey. BJGP Open 2021;5(4):1-9 [FREE Full text] [doi: 10.3399/BJGPO.2020.0179] [Medline: 33910917]
38. Car J, Koh GCH, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
39. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among norwegian general practitioners. J Med Internet Res 2021;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
40. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021;5(3):1-7 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
41. Wanderås MR, Abildsnes E, Thygesen E, Martinez SG. Video consultation in general practice: a scoping review on use, experiences, and clinical decisions. BMC Health Serv Res 2023;23(1):316 [FREE Full text] [doi: 10.1186/s12913-023-09309-7] [Medline: 36997997]
42. Eccles A, Hopper M, Turk A, Atherton H. Patient use of an online triage platform: a mixed-methods retrospective exploration in UK primary care. Br J Gen Pract 2019;69(682):e336-e344 [FREE Full text] [doi: 10.3399/bjgp19X702197] [Medline: 30910874]
43. Atherton H, Boylan AM, Eccles A, Fleming J, Goyder CR, Morris RL. Email consultations between patients and doctors in primary care: content analysis. J Med Internet Res 2020;22(11):e18218 [FREE Full text] [doi: 10.2196/18218] [Medline: 33164902]
44. Bakhai M, Atherton H. How to conduct written online consultations with patients in primary care. BMJ 2021;372:n264. [doi: 10.1136/bmj.n264] [Medline: 33627324]
45. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. Characteristics of patients who seek care via eVisits instead of office visits. Telemed J E Health 2013;19(7):515-519 [FREE Full text] [doi: 10.1089/tmj.2012.0221] [Medline: 23682589]
46. Bavafa H, Hitt LM, Terwiesch C. The impact of e-visits on visit frequencies and patient health: evidence from primary care. Manage Sci 2018;64(12):5461-5480 [FREE Full text] [doi: 10.1287/mnsc.2017.2900] [Medline: 33033417]
47. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]
48. Paparini S, Papoutsi C, Murdoch J, Green J, Petticrew M, Greenhalgh T, et al. Evaluating complex interventions in context: systematic, meta-narrative review of case study approaches. BMC Med Res Methodol 2021;21(1):225 [FREE Full text] [doi: 10.1186/s12874-021-01418-3] [Medline: 34689742]
49. Hewitt H, Gafaranga J, McKinstry B. Comparison of face-to-face and telephone consultations in primary care: qualitative analysis. Br J Gen Pract 2010;60(574):e201-e212 [FREE Full text] [doi: 10.3399/bjgp10X501831] [Medline: 20423575]
50. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Aug 29;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
51. Helsedirektoratet. E-konsultasjoner hos fastleger. [E-consultations with GPs] [Website in Norwegian]. The Norwegian Directorate of Health. 2020. URL: https://www.helsedirektoratet.no/statistikk/statistikk-om-allmennlegetjenester/e-konsultasjoner-hos-fastleger [accessed 2025-03-25]
52. Wanderås MR, Abildsnes E, Thygesen E, Martinez SG. Hammering nails with a screwdriver: how GPs perceive video consultations. BJGP Open 2024:1-10. [doi: 10.3399/bjgpo.2024.0010]
53. Den Norske Legeforening. Den Norske Legeforening. Normaltariffen. [The Norwegian Medical Association. Normal tariff.] [Website in Norwegian]. URL: https://normaltariffen.legeforeningen.no [accessed 2024-12-11]
54. National Online Health Services in Norway [Website in Norwegian]. 2016 Oct 1. URL: https://www.helsenorge.no [accessed 2025-03-25]
55. Tikkanen R, Osborn R, Mossialos E, Djordjevic A, Wharton GA. International Health Care System Profiles—Norway. 2020. URL: https://www.commonwealthfund.org/international-health-policy-center/countries/norway [accessed 2025-03-25]
56. Helsedirektoratet; autorisasjon og spesialistutdanning. [The Norwegian Directorate of Health; authorization and specialist training.] [Website in Norwegian]. Helsedirektoratet. 2020 Aug 31. URL: https://www.helsedirektoratet.no/tema/autorisasjon-og-spesialistutdanning/spesialistutdanning-for-leger/allmennmedisin [accessed 2025-03-25]
57. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: SAGE; 2021.
58. Terry G, Hayfield N, Clarke V, Braun V. In: Willig C, Stainton Rogers W, editors. The SAGE Handbook of Qualitative Research in Psychology. 2nd ed. London: SAGE Publications; 2017:17-37.

XSL•FO
RenderX

59. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
60. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
61. Casadevall C, Fang FC. Diseased science: 'It is more important to know what sort of person has a disease than to know what sort of disease a person has.'— attributed to Hippocrates. Microbe (Washington, DC) 2014;9(10):390-392. [doi: 10.1128/microbe.9.390.1]
62. Greenhalgh T, Rosen R, Shaw SE, Byng R, Faulkner S, Finlay T, et al. Planning and evaluating remote consultation services: a new conceptual framework incorporating complexity and practical ethics. Front Digit Health 2021;3:726095 [FREE Full text] [doi: 10.3389/fdgth.2021.726095] [Medline: 34713199]
63. Moschogianis S, Darley S, Coulson T, Peek N, Cheraghi-Sohi S, Brown BC. Patient experiences of an online consultation system: a qualitative study in English primary care post-COVID-19. Br J Gen Pract 2023;74(744):BJGP.2023.0076. [doi: 10.3399/bjgp.2023.0076]
64. Kristiansen E, Atherton H, Austad B, Bergmo T, Norberg BL, Zanaboni P. Older patients' experiences of access to and use of e-consultations with the general practitioner in Norway: an interview study. Scand J Prim Health Care 2023 Mar;41(1):33-42. [doi: 10.1080/02813432.2022.2161307] [Medline: 36592342]
65. Yee V, Bajaj SS, Stanford FC. Paradox of telemedicine: building or neglecting trust and equity. Lancet Digit Health 2022;4(7):e480-e481 [FREE Full text] [doi: 10.1016/S2589-7500(22)00100-5] [Medline: 35750399]
66. Baur C. Limiting factors on the transformative powers of e-mail in patient-physician relationships: a critical analysis. Health Commun 2000;12(3):239-259. [doi: 10.1207/s15327027hc1203_02]
67. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016;6(1):e009388 [FREE Full text] [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]
68. Mann C. The impact of remote consultations on personalized care.: University of Bristol; 2021. URL: https://arc-w.nihr.ac.uk/Wordpress/wp-content/uploads/2021/08/Remote-consultation-briefing-website-final.pdf [accessed 2024-12-30]
69. Stott NC, Davis RH. The exceptional potential in each primary care consultation. J R Coll Gen Pract 1979;29(201):201-205. [Medline: 448665]
70. Ladds E, Khan M, Moore L, Kalin A, Greenhalgh T. The impact of remote care approaches on continuity in primary care: a mixed-studies systematic review. Br J Gen Pract 2023;73(730):e374-e383 [FREE Full text] [doi: 10.3399/BJGP.2022.0398] [Medline: 37105731]
71. Grimen H. Tillit og makt—tre samanhengar [Trust and power—three interrelationships] [Article in Norwegian]. Tidsskr Nor Laegeforen 2001;121(30):3617-3619 [FREE Full text] [Medline: 11808029]
72. Grimen H. Power, trust, and risk: some reflections on an absent issue. Med Anthropol Q 2009;23(1):16-33. [doi: 10.1111/j.1548-1387.2009.01035.x] [Medline: 19449710]
73. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
74. Anderson J, Walsh J, Anderson M, Burnley R. Patient satisfaction with remote consultations in a primary care setting. Cureus 2021;13(9):e17814 [FREE Full text] [doi: 10.7759/cureus.17814] [Medline: 34660024]
75. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
GP: general practitioner

Edited by N Cahill; submitted 23.02.24; peer-reviewed by I Gagyor, L van Tuyl, T Koskela; comments to author 17.06.24; revised version received 08.08.24; accepted 29.11.24; published 27.03.25.

Please cite as:
Norberg BL, Austad B, Kristiansen E, Zanaboni P, Getz LO
The Dynamics of Doctor-Patient Communication During Remote Consultations: Qualitative Study Among Norwegian Contract General Practitioners
J Med Internet Res 2025;27:e57679
URL: https://www.jmir.org/2025/1/e57679
doi:10.2196/57679
PMID:

XSL•FO
RenderX

©Børge Lønnebakke Norberg, Bjarne Austad, Eli Kristiansen, Paolo Zanaboni, Linn Okkenhaug Getz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical Question Answering Using Reddit Data: Proof-of-Concept Study

Sudeshna Das[1*], PhD; Yao Ge[1*], PhD; Yuting Guo[1], MS; Swati Rajwal[2], MTech; JaMor Hairston[1], MS; Jeanne Powell[1], PhD; Drew Walker[1], PhD; Snigdha Peddireddy[3], MPH; Sahithi Lakamana[1], MS; Selen Bozkurt[1], PhD; Matthew Reyna[1], PhD; Reza Sameni[1,4], PhD; Yunyu Xiao[5], PhD; Sangmi Kim[6], PhD; Rasheeta Chandler[6], PhD; Natalie Hernandez[7], PhD; Danielle Mowery[8], PhD; Rachel Wightman[9], MD; Jennifer Love[10], MD; Anthony Spadaro[11], MD; Jeanmarie Perrone[12], MD; Abeed Sarker[1,4], PhD

[1]Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States

[2]Department of Computer Science and Informatics, Emory University, Atlanta, GA, United States

[3]Department of Behavioral, Social & Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[4]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, United States

[5]Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, United States

[6]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States

[7]Center for Maternal Health Equity, Morehouse School of Medicine, Atlanta, GA, United States

[8]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[9]Department of Emergency Medicine, Warren Alpert Medical School of Brown University, Providence, RI, United States

[10]Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[11]Department of Emergency Medicine, Rutgers New Jersey Medical School, Newark, NJ, United States

[12]Department of Emergency Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States

[*]these authors contributed equally

Corresponding Author:
Sudeshna Das, PhD
Department of Biomedical Informatics
School of Medicine
Emory University
101 Woodruff Circle
Atlanta, GA, 30322
United States
Phone: 1 4047270229
Email: sudeshna.das@emory.edu

Abstract

Background: The increasing use of social media to share lived and living experiences of substance use presents a unique opportunity to obtain information on side effects, use patterns, and opinions on novel psychoactive substances. However, due to the large volume of data, obtaining useful insights through natural language processing technologies such as large language models is challenging.

Objective: This paper aims to develop a retrieval-augmented generation (RAG) architecture for medical question answering pertaining to clinicians' queries on emerging issues associated with health-related topics, using user-generated medical information on social media.

Methods: We proposed a two-layer RAG framework for query-focused answer generation and evaluated a proof of concept for the framework in the context of query-focused summary generation from social media forums, focusing on emerging drug-related information. Our modular framework generates individual summaries followed by an aggregated summary to answer medical queries from large amounts of user-generated social media data in an efficient manner. We compared the performance of a quantized large language model (Nous-Hermes-2-7B-DPO), deployable in low-resource settings, with GPT-4. For this

XSL•FO
RenderX

proof-of-concept study, we used user-generated data from Reddit to answer clinicians' questions on the use of xylazine and ketamine.

Results: Our framework achieves comparable median scores in terms of relevance, length, hallucination, coverage, and coherence when evaluated using GPT-4 and Nous-Hermes-2-7B-DPO, evaluated for 20 queries with 76 samples. There was no statistically significant difference between GPT-4 and Nous-Hermes-2-7B-DPO for coverage (Mann-Whitney U=733.0; n_1=37; n_2=39; P=.89 two-tailed), coherence (U=670.0; n_1=37; n_2=39; P=.49 two-tailed), relevance (U=662.0; n_1=37; n_2=39; P=.15 two-tailed), length (U=672.0; n_1=37; n_2=39; P=.55 two-tailed), and hallucination (U=859.0; n_1=37; n_2=39; P=.01 two-tailed). A statistically significant difference was noted for the Coleman-Liau Index (U=307.5; n_1=20; n_2=16; P<.001 two-tailed).

Conclusions: Our RAG framework can effectively answer medical questions about targeted topics and can be deployed in resource-constrained settings.

(J Med Internet Res 2025;27:e66220) doi:10.2196/66220

KEYWORDS

retrieval-augmented generation; substance use; social media; large language models; natural language processing; artificial intelligence; GPT; psychoactive substance

Introduction

Large language models (LLMs) present opportunities for solving complex biomedical natural language processing problems, such as medical question answering (MQA). However, operational challenges (eg, high computational resource requirements) hinder their real-life deployment and use. Another issue with LLM-generated text for MQA is "hallucination": generated text that is plausible-sounding but nonsensical or incorrect [1]. Chain-of-thought prompting [2], self-reflection [1], and retrieval-augmented generation (RAG) are forerunners in mitigating hallucination. RAG also aids in constraining generated texts and improves in-context learning [3]. LLMs in RAG frameworks have been used in the biomedical domain owing to the need for timely, accurate, and transparent responses [4]. As LLMs become increasingly integrated into clinical practice [5], it is important to ensure their operability in low-resource settings [6] while generating accurate and coherent texts.

We present a proof-of-concept study for a two-layer RAG framework for MQA that ingests user-generated medical information from Reddit. We used smaller, quantized, open-source LLMs that can run on personal computers without specialized hardware, allowing our framework to be used in low-resource settings, thus ensuring equitable access to timely medical information.

Methods

Study Design

We evaluated our proof-of-concept study in a setting where copious amounts of data are available for a topic but gathering insights and answering questions require substantial manual work—the topic of emerging drugs from Reddit. Reddit has ~52 million daily active users, is commonly used to study emerging medical themes [7], and features numerous discussions on the nonmedical uses of substances. Recently, Reddit data have been leveraged to study novel psychoactive substances since such information is not typically available elsewhere. We chose two substances that have gained attention recently—xylazine (because of its increasing impact and association with the US opioid crisis) and ketamine (because of its recent popularity as a treatment for depression). We collected all available data (~2.5 billion posts) from Reddit until December 31, 2023, and extracted all posts mentioning xylazine (n=177,684) and ketamine (n=7699) for our retrieval engine. Based on clinician-driven interests, we formulated 20 queries (Table 1).

XSL•FO
RenderX

Table 1. Queries used for evaluating the framework.

Query ID	Query
1	What are the side effects of xylazine?
2	What does xylazine do to the skin?
3	How does xylazine impact rehab?
4	What is xylazine withdrawal like?
5	What drugs contain xylazine?
6	What treatments work for xylazine?
7	What drugs are mixed or cut with xylazine?
8	What areas of the United States are impacted by xylazine?
9	How is xylazine different from pure heroin?
10	What is the general sentiment associated with xylazine?
11	Does narcan or naloxone work for xylazine overdose?
12	What are the side effects of ketamine?
13	What is ketamine withdrawal like?
14	What are k cramps like?
15	How do the users describe k hole?
16	Does ketamine work for depression?
17	What drugs are ketamine coused with recreationally?
18	Is ketamine effective for the treatment of suicidal behavior?
19	How can you treat ketamine addiction?
20	Does ketamine use cause cramps?

System Architecture

As depicted in Figure 1, the user submits a query to be parsed by the information retrieval engine, which returns a ranked list of documents. The top n documents are chosen to be sources for answer generation. In the first layer, the LLM is provided with (1) a query, (2) text from the retrieved documents, and (3) a prompt that embeds the text and instructs the LLM to summarize it (Multimedia Appendix 1). Since the prompt context window is finite, feeding the LLM all the retrieved text for answer generation is typically impossible. Even single documents can be too long. Thus, the framework allows for the specification of segment lengths for the retrieved text in each iteration, ensuring the framework is applied to relatively small LLMs with shorter context lengths. The first layer generates short, query-focused summaries (Figure 1). The LLM states if the retrieved text segment does not contain an answer to the question. Examples of this first-layer summarization are provided in Multimedia Appendix 2.

The second layer takes as input the original query and individual short summaries embedded within a second prompt (Multimedia Appendix 1) while ignoring summaries that the LLM states did not contain the answer. Figure 1 depicts an example of the final, synthesized summary.

We used the 8-bit quantized model Nous-Hermes-2-7B-DPO as our LLM, which is tuned on 1,000,000 high-quality instructions [8]. To test the performance of the proposed framework with larger models, we also performed an evaluation using GPT-4 [9].

XSL•FO
RenderX

Figure 1. Overview of the two-layer RAG framework. The first layer generates individual summaries based on retrieved posts relevant to the original query. The second layer generates the final summary based on the individual summaries generated in the first layer. LLM: large language model; RAG: retrieval-augmented generation.

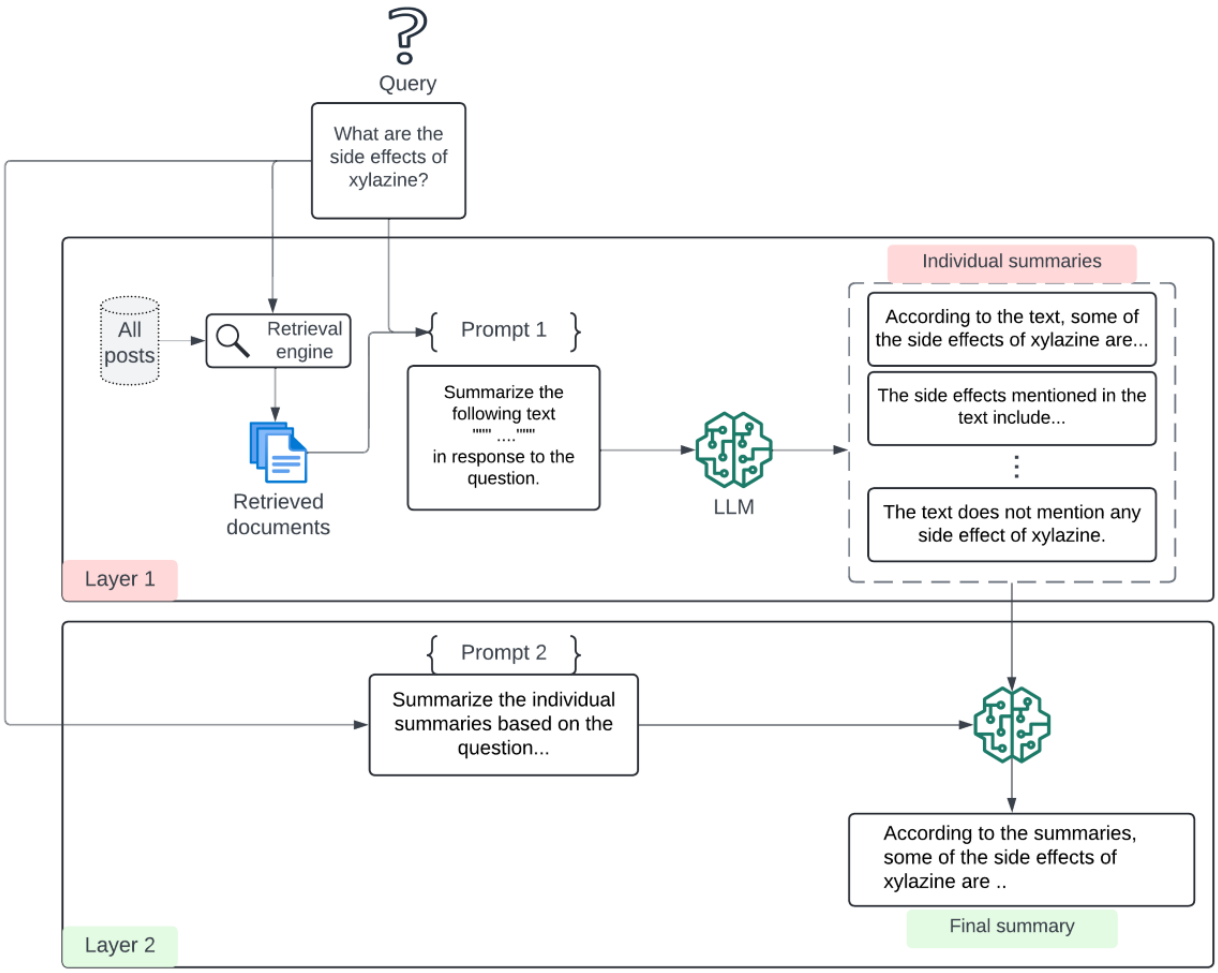

RAG Architecture

Since the retrieval aspect is not our primary focus, we used a simple keyword-based approach using the default search settings provided by our information retrieval package "Whoosh," which uses Okapi BM25F ranking [10]. The top 50 retrieved documents were chosen for generating the first-layer summaries. This number may be adjusted without changes to the architecture. The number of text segments is typically higher when posts do not fit within the context window of the LLM after being embedded within the prompt.

Evaluation

Our evaluation focused on the architecture's summary generation quality, rather than retrieval performance. Commonly used automatic summary evaluation methods, such as Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [11] and bilingual evaluation understudy (BLEU) [12], primarily focus on text overlap between generated and gold-standard summaries. In the absence of gold-standard summaries, subject matter experts manually and qualitatively evaluated the important nuances of generative summaries, which is impossible with ROUGE or BLEU. We used Likert-scale evaluations (Table 2). Each query–individual summary–final summary triplet was evaluated by ≥2 evaluators (at least a master's degree in medicine, public health, informatics, or allied fields). Overall, 21 experts generated 76 evaluations for 20 unique queries.

We also assessed the readability of the final summaries using the Coleman-Liau Readability Index (CLI) [13], which approximates the US grade level required to comprehend text.

We performed nonparametric tests for proportions (Mann-Whitney *U* test) with the null hypothesis (H_0: "The two populations are equal") to determine if the scores assigned to answers generated by GPT-4 and Nous-Hermes-2-7B-DPO vary significantly. All tests were performed using the *SciPy* package [14]. The null hypothesis (H_0) was rejected if $P<.05$ (two-tailed).

XSL•FO
RenderX

Table 2. Evaluation criteria and scales presented to annotators.

Criteria	Question	Evaluation scale
Coverage	Does the final summary accurately represent the information present in the original text?	• 5: Yes; the final summary covers all the important information present in the original text. • 4: Mostly; the final summary covers most, but not all of the important information. • 3: Somewhat; the final summary covers some of the important information, but also misses some of them. • 2: Not really; the final summary misses most of the important information. • 1: No; the final summary does not cover any of the important information present in the original text.
Coherence	Is the final summary coherent?	• 5: Yes; the final summary is easy to read and understand. • 4: Mostly; the final summary is readable, but not straightforward to understand. • 3: Somewhat; the final summary is readable but confusing. • 2: Not really; the final summary has some grammatical errors or nonsequiturs. • 1: No; the final summary is unintelligible or incomprehensible.
Relevance	Does the final summary answer the original question?	• 3: Yes; the summary answers the original question. • 2: Partially; the summary answers the original question, but not fully. • 1: No; the summary does not answer the original question.
Length	Is the length of the final summary appropriate?	• 3: Yes; the summary is appropriate in length. • 2: Somewhat; the summary could be shorter or longer. • 1: No; the summary is long-winded or too short.
Hallucination	Does the summary contain information not present in the original text?	• 0: No; the summary does not contain information not present in the original text. • 1: Yes; the summary contains information not present in the original text.

Ethical Considerations

This study was deemed to be exempt from review per the Emory University Institutional Review Board's guidelines. The data used in this study are anonymous by default. We ensured that self-disclosed, personally identifiable information is not used by only reporting aggregated data. We removed posts that were deleted by the user.

Results

We conducted extensive expert evaluations of the generated answers for coverage, coherence, relevance, length, and hallucination (Figure 2). Annotators were not made aware of which LLM was used to generate the summaries for fair evaluation. On a 5-point Likert scale, median coverage scores were 5 (IQR 4-5) for both models; the distributions did not differ significantly (U=733.0; n_1=37; n_2=39; P=.89). Median coherence scores were 5 (IQR 5-5 for GPT-4; IQR 4-5 for Nous-Hermes-2-7B-DPO) for both; they did not differ significantly (U=670.0; n_1=37; n_2=39; P=.49).

On a 3-point Likert scale, the median relevance scores were 3 (IQR 3-3) for both; they did not differ significantly (U=662.0; n_1=37; n_2=39; P=.15). Median length scores were 3 (IQR 2-3) for both; they did not differ significantly (U=672.0; n_1=37; n_2=39; P=.55). On a binary Likert scale, median hallucination scores were 0 (IQR 0-0) for both; they did not differ significantly (U=859.0; n_1=37; n_2=39; P=.10). The median CLIs were 16.635 (IQR 13.860-17.675) for GPT-4 and 12.125 (IQR 11.02-13.98) for Nous-Hermes-2-7B-DPO; there was a statistically significant difference (U=307.5; n_1=20; n_2=16; P<.001).

Median token counts for queries posed to GPT-4 and Nous-Hermes-2-7B-DPO were 5 (IQR 5-7) and 7 (IQR 5-8), respectively; there was no significant difference (U=165.0; n_1=20; n_2=16; P=.66). Median lengths of responses generated by GPT-4 and Nous-Hermes-2-7B-DPO were 1118 (IQR 709-2986) and 441 (IQR 231-695) for the combined individual summaries and 141.5 (IQR 115-159) and 61 (IQR 28-87) for the final summaries, respectively. Both were significantly different (U=300.0; n_1=20; n_2=16; P<.001, and U=145.5; n_1=20; n_2=16; P<.001).

XSL•FO
RenderX

Figure 2. Box plots illustrating the distribution of scores for the evaluation criteria used. (A) Coverage on a 5-point Likert scale. (B) Coherence on a 5-point Likert scale. (C) Relevance on a 3-point Likert scale. (D) Length on a 3-point Likert scale. (E) Hallucination on a binary scale. (F) Values for the Coleman-Liau Index. (G) Token counts for Questions. (H) Token counts for combined individual summaries. (I) Token counts for the final summary.

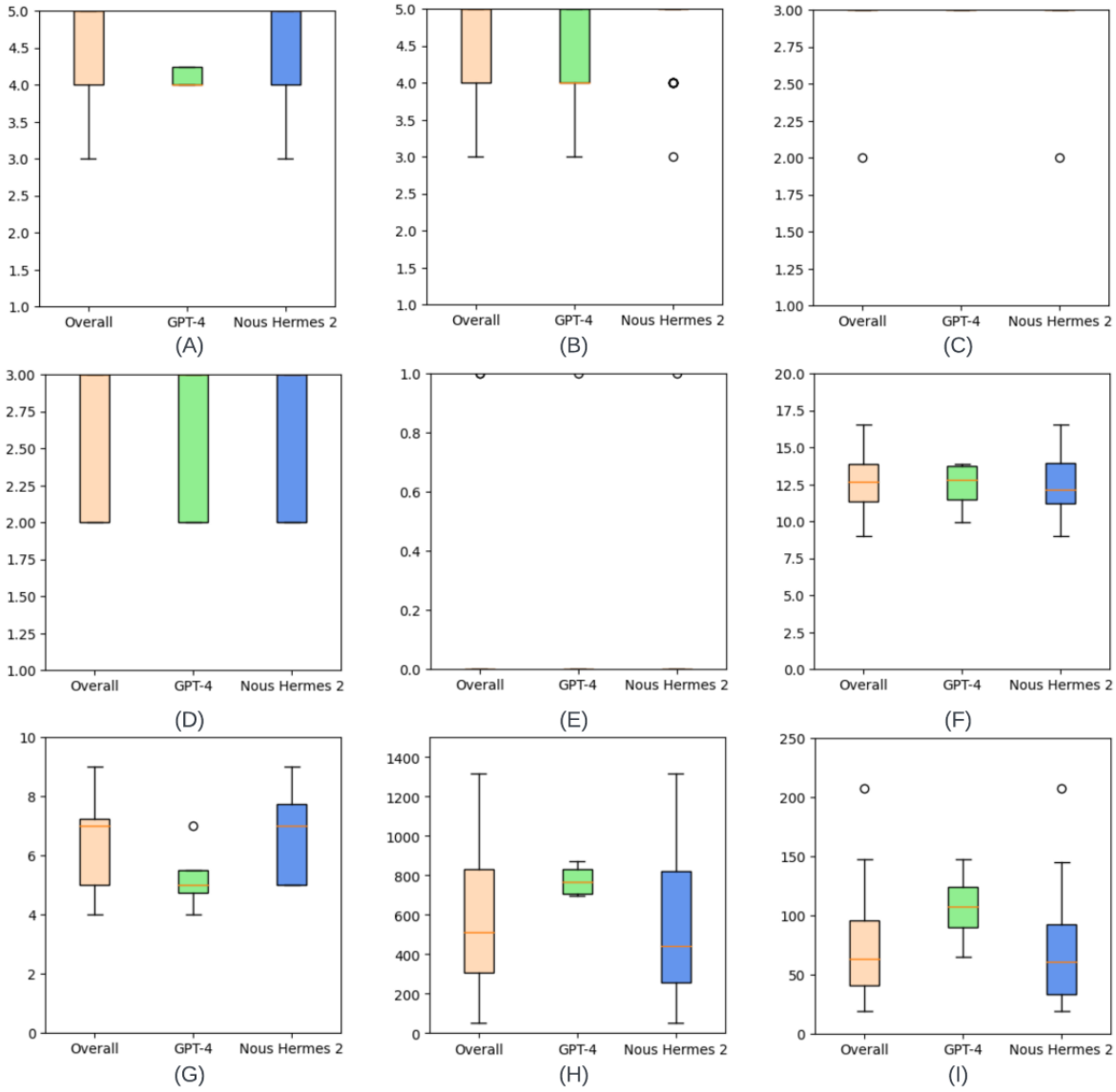

Discussion

Principal Findings

This study presents a novel two-layer RAG framework for MQA that uses user-generated content from Reddit. Our findings demonstrate that the framework effectively synthesizes accurate and contextually relevant answers even in low-resource settings, aligning with our goal to create an accessible, computationally lightweight tool. Focusing on small, quantized, open-source LLMs ensures equitable access to valuable insights about emerging trends, potential side effects, and general perception of substances, as reflected in Reddit posts.

The modular structure of the framework enables good performance without requiring specialized hardware, which is critical in low-resource environments. This modularity also supports using different retrieval engines or LLMs, providing flexibility to adapt to various use cases. The system's ability to answer nuanced queries (eg "What are k cramps like?"—which would require extensive manual curation) illustrates its potential for real-world applications. The framework's ability to specify temporal ranges in queries allows it to track trends over time, offering opportunities for longitudinal studies and misinformation detection.

Unlike previous work [15], where segments of text are generated chronologically, we performed segmentation at the post level without accounting for chronology. Compared with existing literature, which predominantly relies on large LLMs for MQA [4,16], this study underscores the potential of smaller models for tasks requiring domain-specific, contextually accurate outputs. Prior work often focuses on high-resource settings [17,18] with robust computational infrastructure, leaving gaps

XSL•FO
RenderX

in applicability for low-resource environments. We fill this gap by showing that reliable performance can be achieved with computationally efficient architectures, expanding the reach of artificial intelligence tools to underresourced regions. Although smaller LLMs have been used [19], summarization from large volumes of text with aggregated information has not been addressed before.

Limitations

Despite its promising features, the framework has limitations. It relies on the accuracy and representativeness of the Reddit data it ingests. Reddit posts may include biases, inaccuracies, or misinformation that could influence the system's output. While faithfully summarizing misinformation is valuable for transparency, users need to exercise caution in interpreting the results. Additionally, we evaluated the framework using a small set of queries pertaining to substance use; further validation is necessary to assess performance across diverse medical domains.

Conclusions

This study demonstrates that a modular, lightweight RAG framework can effectively address complex MQA using social media data in low-resource settings. By enabling clinicians to rapidly extract insights about substance use trends and potential side effects from Reddit posts, the framework holds significant potential for improving public health.

Acknowledgments

Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health (NIH) under award R01DA057599. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data Availability

All data used in this study were publicly available from Reddit at the time of data collection. The second-level summaries are available in Multimedia Appendix 3. The original posts and social media posts analyzed during this study are available from the corresponding author upon reasonable request and the completion of a data use agreement.

Authors' Contributions

SD and Y Ge led analysis, evaluation, visualization, and original draft preparation. Y Guo, SR, JH, J Powell, DW, SP, and SL contributed to the evaluation and manuscript preparation. SB, MR, RS, YX, SK, RC, NH, DM, RW, JL, A Spadaro, and J Perrone worked on the evaluation. A Sarker conceptualized the study, led model design and implementation, and supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prompts used.
[DOCX File , 12 KB - jmir_v27i1e66220_app1.docx]

Multimedia Appendix 2
Sample first-layer individual summaries.
[DOCX File , 14 KB - jmir_v27i1e66220_app2.docx]

Multimedia Appendix 3
Final summaries generated by the framework for each of the 20 queries used for evaluation.
[DOCX File , 30 KB - jmir_v27i1e66220_app3.docx]

References

1. Ji Z, Yu T, Xu Y, Lee N, Ishii E, Fung P. Towards mitigating LLM hallucination via self reflection. 2023 Presented at: Findings of the Association for Computational Linguistics: EMNLP 2023; December 6-10, 2023; Singapore, Singapore p. 1827-1843 URL: https://aclanthology.org/2023.findings-emnlp.123/ [doi: 10.18653/v1/2023.findings-emnlp.123]
2. Dhuliawala S, Komeili M, Xu J, Raileanu R, Li X, Celikyilmaz A, et al. Chain-of-verification reduces hallucination in large language models. arXiv Preprint posted online on September 20, 2023 [FREE Full text] [doi: 10.18653/v1/2024.findings-acl.212]
3. Ge J, Sun S, Owens J, Galvez V, Gologorskaya O, Lai JC, et al. Development of a liver disease-specific large language model chat interface using retrieval augmented generation. medRxiv Preprint posted online on November 11, 2023 [FREE Full text] [doi: 10.1101/2023.11.10.23298364] [Medline: 37986764]

XSL•FO
RenderX

4. Xiong G, Jin Q, Lu Z, Zhang A. Benchmarking retrieval-augmented generation for medicine. arXiv Preprint posted online on February 20, 2024 [FREE Full text] [doi: 10.18653/v1/2024.findings-acl.372]
5. McNamara SL, Yi PH, Lotter W. The clinician-AI interface: intended use and explainability in FDA-cleared AI devices for medical image interpretation. NPJ Digital Med 2024;7(1):80 [FREE Full text] [doi: 10.1038/s41746-024-01080-1] [Medline: 38531952]
6. Ghosh S, Tyagi U, Kumar S, Manoch D. BioAug: conditional generation based data augmentation for low-resource biomedical NER. 2023 Presented at: SIGIR '23: The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 27, 2023; Taipei, Taiwan. [doi: 10.1145/3539618.3591957]
7. Somani S, Balla S, Peng AW, Dudum R, Jain S, Nasir K, et al. Contemporary attitudes and beliefs on coronary artery calcium from social media using artificial intelligence. NPJ Digital Med 2024;7(1):83 [FREE Full text] [doi: 10.1038/s41746-024-01077-w] [Medline: 38555387]
8. Teknium, theemozilla, karan4d, huemin_art. Nous-Hermes-2-Mistral-7B-DPO. Hugging Face. URL: https://huggingface.co/NousResearch/Nous-Hermes-2-Mistral-7B-DPO [accessed 2024-07-10]
9. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Leoni Aleman F, et al. GPT-4 technical report. arXiv Preprint posted online on March 15, 2023 [FREE Full text] [doi: 10.48550/arXiv.2303.08774]
10. Chaput M. Whoosh 2.7.4 documentation. Whoosh. URL: https://whoosh.readthedocs.io/en/latest/ [accessed 2024-07-10]
11. Lin CY. ROUGE: a package for automatic evaluation of summaries. 2004 Presented at: Text Summarization Branches Out; July 25-26, 2004; Barcelona, Spain p. 74-81 URL: https://aclanthology.org/W04-1013.pdf
12. Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. 2024 Presented at: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics; July 7-12, 2002; Philadelphia, PA URL: https://aclanthology.org/P02-1040.pdf [doi: 10.3115/1073083.1073135]
13. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60(2):283-284. [doi: 10.1037/h0076540]
14. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
15. Subbiah M, Zhang SS, Chilton L, Zhang B, McKeown K. Reading subtext: evaluating large language models on short story summarization with writers. arXiv Preprint posted online on March 02, 2024 [FREE Full text] [doi: 10.1162/tacl_a_00702]
16. Hammane Z, Ben-Bouazza FE, Fennan A. SelfRewardRAG: enhancing medical reasoning with retrieval-augmented generation and self-evaluation in large language models. 2024 Presented at: International Conference on Intelligent Systems and Computer Vision (ISCV); August 12, 2024; Fez, Morocco URL: https://ieeexplore.ieee.org/abstract/document/10620139/ [doi: 10.1109/iscv60512.2024.10620139]
17. Alkhalaf M, Yu P, Yin M, Deng C. Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. J Biomed Inform 2024;156:104662 [FREE Full text] [doi: 10.1016/j.jbi.2024.104662] [Medline: 38880236]
18. Soong D, Sridhar S, Si H, Wagner J, Sá ACC, Yu CY, et al. Improving accuracy of GPT-3/4 results on biomedical data using a retrieval-augmented language model. PLOS Digital Health 2024;3(8):e0000568. [doi: 10.1371/journal.pdig.0000568] [Medline: 39167594]
19. Jeong M, Sohn J, Sung M, Kang J. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. Bioinformatics 2024;40(Suppl 1):i119-i129 [FREE Full text] [doi: 10.1093/bioinformatics/btae238] [Medline: 38940167]

Abbreviations

BLEU: Bilingual Evaluation Understudy
CLI: Coleman-Liau Index
LLM: large language model
MQA: medical question answering
RAG: retrieval-augmented generation
ROUGE: Recall-Oriented Understudy for Gisting Evaluation

XSL•FO
RenderX

Edited by A Mavragani; submitted 06.09.24; peer-reviewed by G Xiong, M Jeong; comments to author 26.09.24; revised version received 30.10.24; accepted 05.12.24; published 06.01.25.

Please cite as:
Das S, Ge Y, Guo Y, Rajwal S, Hairston J, Powell J, Walker D, Peddireddy S, Lakamana S, Bozkurt S, Reyna M, Sameni R, Xiao Y, Kim S, Chandler R, Hernandez N, Mowery D, Wightman R, Love J, Spadaro A, Perrone J, Sarker A
Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical Question Answering Using Reddit Data: Proof-of-Concept Study
J Med Internet Res 2025;27:e66220
URL: https://www.jmir.org/2025/1/e66220
doi:10.2196/66220
PMID:

©Sudeshna Das, Yao Ge, Yuting Guo, Swati Rajwal, JaMor Hairston, Jeanne Powell, Drew Walker, Snigdha Peddireddy, Sahithi Lakamana, Selen Bozkurt, Matthew Reyna, Reza Sameni, Yunyu Xiao, Sangmi Kim, Rasheeta Chandler, Natalie Hernandez, Danielle Mowery, Rachel Wightman, Jennifer Love, Anthony Spadaro, Jeanmarie Perrone, Abeed Sarker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Bias Mitigation in Primary Health Care Artificial Intelligence Models: Scoping Review

Maxime Sasseville[1,2], PhD; Steven Ouellet[1], PhD; Caroline Rhéaume[2,3,4], MD, PhD; Malek Sahlia[5], MSc, MEng; Vincent Couture[1], PhD; Philippe Després[6], PhD; Jean-Sébastien Paquette[2,3], MD; David Darmon[7], PhD; Frédéric Bergeron[8], MSI; Marie-Pierre Gagnon[1,2], PhD

[1]Faculté des sciences infirmières, Université Laval, Québec, QC, Canada

[2]Vitam Research Center on Sustainable Health, Québec, QC, Canada

[3]Département de médecine familiale et de médecine d'urgence de la Faculté de médecine, Université Laval, Québec, QC, Canada

[4]Research Center of Quebec Heart and Lungs Institute, Québec, QC, Canada

[5]École Nationale des Sciences de l'Informatique, Université de La Manouba, La Manouba, Tunisia

[6]Département de physique, de génie physique et d'optique de la Faculté des sciences et de génie, Université Laval, Québec, QC, Canada

[7]Risques, Epidémiologie, Territoires, Informations, Education et Santé. Département d'enseignement et de recherche en médecine générale, Université Côte d'Azur, Nice, France

[8]Direction des services-conseils de la Bibliothèque, Université Laval, Québec, QC, Canada

Corresponding Author:
Maxime Sasseville, PhD
Faculté des sciences infirmières, Université Laval
1050 Av. de la Médecine
Québec, QC, G1V 0A6
Canada
Phone: 1 418 656 3356
Email: maxime.sasseville@fsi.ulaval.ca

Abstract

Background: Artificial intelligence (AI) predictive models in primary health care have the potential to enhance population health by rapidly and accurately identifying individuals who should receive care and health services. However, these models also carry the risk of perpetuating or amplifying existing biases toward diverse groups. We identified a gap in the current understanding of strategies used to assess and mitigate bias in primary health care algorithms related to individuals' personal or protected attributes.

Objective: This study aimed to describe the attempts, strategies, and methods used to mitigate bias in AI models within primary health care, to identify the diverse groups or protected attributes considered, and to evaluate the results of these approaches on both bias reduction and AI model performance.

Methods: We conducted a scoping review following Joanna Briggs Institute (JBI) guidelines, searching Medline (Ovid), CINAHL (EBSCO), PsycINFO (Ovid), and Web of Science databases for studies published between January 1, 2017, and November 15, 2022. Pairs of reviewers independently screened titles and abstracts, applied selection criteria, and performed full-text screening. Discrepancies regarding study inclusion were resolved by consensus. Following reporting standards for AI in health care, we extracted data on study objectives, model features, targeted diverse groups, mitigation strategies used, and results. Using the mixed methods appraisal tool, we appraised the quality of the studies.

Results: After removing 585 duplicates, we screened 1018 titles and abstracts. From the remaining 189 full-text articles, we included 17 studies. The most frequently investigated protected attributes were race (or ethnicity), examined in 12 of the 17 studies, and sex (often identified as gender), typically classified as "male versus female" in 10 of the studies. We categorized bias mitigation approaches into four clusters: (1) modifying existing AI models or datasets, (2) sourcing data from electronic health records, (3) developing tools with a "human-in-the-loop" approach, and (4) identifying ethical principles for informed decision-making. Algorithmic preprocessing methods, such as relabeling and reweighing data, along with natural language processing techniques that extract data from unstructured notes, showed the greatest potential for bias mitigation. Other methods aimed at enhancing model fairness included group recalibration and the application of the equalized odds metric. However, these approaches sometimes exacerbated prediction errors across groups or led to overall model miscalibrations.

XSL•FO
RenderX

Conclusions: The results suggest that biases toward diverse groups are more easily mitigated when data are open-sourced, multiple stakeholders are engaged, and during the algorithm's preprocessing stage. Further empirical studies that include a broader range of groups, such as Indigenous peoples in Canada, are needed to validate and expand upon these findings.

Trial Registration: OSF Registry osf.io/9ngz5/; https://osf.io/9ngz5/

International Registered Report Identifier (IRRID): RR2-10.2196/46684

(J Med Internet Res 2025;27:e60269) doi:10.2196/60269

KEYWORDS

artificial intelligence; AI; algorithms; expert system; decision support; bias; community health services; primary health care; health disparities; social equity; scoping review

Introduction

Developments in computer science have led to artificial intelligence (AI) models that learn from large datasets and can perform independent analysis [1-4]. Significant progress has been made in these tasks with the development of machine learning (ML). This branch of AI focuses on understanding, generating, and reasoning based on data without explicit human instructions [2,3] Such ML algorithms use datasets known as "training datasets" to capture the patterns required for clustering tasks or predictive modeling [3,4]. These models are now used in multiple contexts and industries to predict the likelihood of an event or to support human decision-making [4]. In health care, AI models applied in radiology can potentially detect and predict the progression of cancerous tumors accurately [5]. Algorithms can also be useful in community-based primary health care (CBPHC) for identifying individuals, such as heart failure or diabetes outpatients, who require specific health care services [6]. As defined by the Canadian Institutes of Health Research, CBPHC encompasses a comprehensive array of services aimed at community well-being, including primary prevention (such as public health), health promotion, disease prevention, diagnosis, treatment, and management of chronic and episodic illnesses, rehabilitation support, and end-of-life care [7].

Despite the potential benefits of AI, such as compensating for workforce shortage and maximizing access to CBPHC [6], algorithm biases toward diverse groups can hinder their application in health care settings. These biases may be perpetuated when protected attributes [1], as identified by the place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital (PROGRESS-Plus) framework [8], are underrepresented or misrepresented in the training data of algorithms [1,9]. Strategies aimed at identifying and mitigating bias, defined as a persistent inclination either in favor or toward something [9], in predictive models are in development and beginning to be empirically applied [10,11]. In computer science, attempts to achieve algorithmic fairness can involve which are (1) preprocessing, (2) in-processing, or even, (3) postprocessing strategies, such as those used in "out-of-the-box" commercial AI models [4]. Academic disciplines beyond computer science, such as medicine, management, and ethics, are also closely involved in addressing issues related to identifying potential bias toward diverse groups in AI models [1,3]. However, there remains a knowledge gap regarding which strategies and methods have been empirically applied to mitigate bias toward diverse groups in CBPHC algorithms [10,12].

To address this gap, we conducted a scoping review aimed at identifying and describing (1) the attempts made to mitigate bias in primary health care AI models, (2) which diverse groups or protected attributes have been considered, and (3) the results regarding bias attenuation and the overall performance of the models.

Methods

Search Strategy

We conducted a scoping review informed by the Joanna Briggs Institute (JBI) [13] and used the Population (or Participant), Concept, and Context Framework [14], as shown in Table 1.

XSL•FO
RenderX

Table 1. Population (or Participant), Concept, and Context framework used for the search strategy.

PCC[a] elements [14]	Definition (per JBI[b] Reviewer's Manual)	PCC elements applied in this review
Population	"Important characteristics of participants, including age and other qualifying criteria" (11.2.4)	Any diverse groups [8] based on their personal or protected attributes [1].
Concept	"The core concept examined by the scoping review should be clearly articulated to guide the scope and breadth of the inquiry. This may include details that pertain to elements that would be detailed in a standard systematic review, such as the "interventions" or "phenomena of interest" (11.2.4)	Strategies, attempts, or methods for assessing and mitigating bias in artificial intelligence.
Context	"May include...cultural factors such as geographic location or specific racial or gender-based interests. In some cases, context may also encompass details about the specific setting."	Community-based primary health care [7].

[a]PCC (Population [or Participant], Concept, and Context) framework [14].

[b]JBI: Joanna Briggs Institute.

Bias Mitigation in Primary Health Care Artificial Intelligence Models

Primary review questions are (1) What attempts have been made to mitigate bias in primary health care AI models? (2) Which diverse groups or protected attributes have been considered? and (3) What are the results regarding bias attenuation and model performance?

In November 2022, we developed a search strategy aligned with the main concepts of our primary review questions with an experienced librarian in 4 relevant databases (MEDLINE [Ovid], CINAHL [EBSCO], PsycInfo [Ovid], and Web of Science). The results of the search strategy in Web of Science were limited to the following 2 indexes: Science Citation Index Expanded and Emerging Sources Citation Index. We used 5 relevant articles to test the sensitivity of our search strategy, focusing on peer-reviewed publications from the past 5 years (between January 1, 2017, and November 15, 2022). The search strategies for each database can be found in Multimedia Appendix 1.

Data Collection

We imported all sources (n=1603) into the web-based collaborative tool Covidence (Veritas Health Innovation) [15], which automatically identified and removed 581 duplicates, with an additional 4 removed manually. The inclusion and exclusion criteria are presented in Table 2. During the title and abstract screening phase, 7 reviewers independently assessed the abstracts based on the selection criteria. We piloted the screening process on 50 sources that all reviewers independently assessed. Reviewers included a source if it met our inclusion criteria, such as featuring an AI predictive model in health, targeting primary health care populations, and presenting a strategy or method for reducing bias. All titles and abstracts were screened independently by at least 2 reviewers, with any discrepancies resolved through consensus involving all reviewers, including at least 1 senior researcher.

Table 2. Inclusion and exclusion criteria.

PCC (Population, Concept, and Context) elements [14]	Inclusion criteria	Exclusion criteria
Population	• Any populations targeted by CBPHC[a] interventions.	• Any populations targeted by hospital or specialized care interventions.
Concept	• All methods or strategies deployed to assess and mitigate bias toward diverse groups or protected attributes in AI models. • All mitigation methods or strategies deployed to promote and increase equity, diversity, and inclusion in CBPHC algorithms.	• Methods or strategies deployed to assess and mitigate bias in the AI model itself (eg, biased prediction of treatment effects), rather than bias related to individuals' characteristics or protected attributes. • Strategies, methods, or interventions that are not related to CBPHC. • CBPHC interventions that do not include any algorithm or AI system.
Context	• Include all CBPHC algorithms (AI) applications that can perpetuate or introduce potential biases toward diverse groups based on their characteristics or protected attributes.	• Algorithms used by primary health care providers for support in administrative tasks and operational aspects, rather than for clinical decisions.
Study design, study type, and time frame	• All empirical studies published in English or French between 2017 and 2022.	• Reviews, opinions, commentaries, editorial content, conference papers, communications, protocols, magazine articles, and so on.

[a]CBPHC: Community-based primary health care.

XSL•FO
RenderX

For the remaining articles assessed for eligibility at the full-text review stage, we searched for and obtained any missing full texts of selected references, then imported them into Covidence. Out of 5 reviewers independently applied the same selection criteria, and all reasons for exclusion were recorded in Covidence. All full texts underwent dual screening. As in the previous stage, any discrepancies regarding the included studies were resolved through consensus among all reviewers, including at least one senior researcher.

Data Extraction

One experienced reviewer performed the extraction of the included studies, and 2 senior researchers validated the data for all of them. We also hand-searched [16] and identified 2 relevant articles [17,18] related to 2 included studies [19,20], which were added to Covidence for extraction. Based on reporting standards for AI in health care [21], we extracted the following information (title of the paper, year of publication, lead author, and country), study objective, discipline and study design, AI model features, study population and setting, AI model architecture and evaluation, bias assessment method, strategy for deployment, diverse groups concerned, bias mitigation results, and the impact on AI model performance and accuracy.

Quality Assessment

One senior reviewer appraised the quality of the included studies by applying the Mixed-Methods Appraisal Tool (MMAT) [22,23] and at least one senior researcher validated each of them.

Data Synthesis

In accordance with the JBI recommendations [24], we synthesized data using structured narrative summaries around our review concepts (eg, model data source, model input, model output, diverse groups, or protected attributes), mitigation strategies deployed, and the results on bias mitigation and overall model performance. We reported our findings based on the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [25].

Ethical Considerations

We obtained approval from the ethics board of the “Comité d’éthique de la recherche sectoriel en Santé des Populations et Première Ligne du Centre Intégré Universitaire de Santé et de Services Sociaux de la Capitale-Nationale” for the Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse and Equitable AI (PREMIA) project (#2023-2726).

Results

Out of a total of 1018 titles and abstracts, along with 189 full-text articles that underwent dual screening, 17 studies [19,20,26-40] met our eligibility criteria. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram is shown in Figure 1 [41].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram.

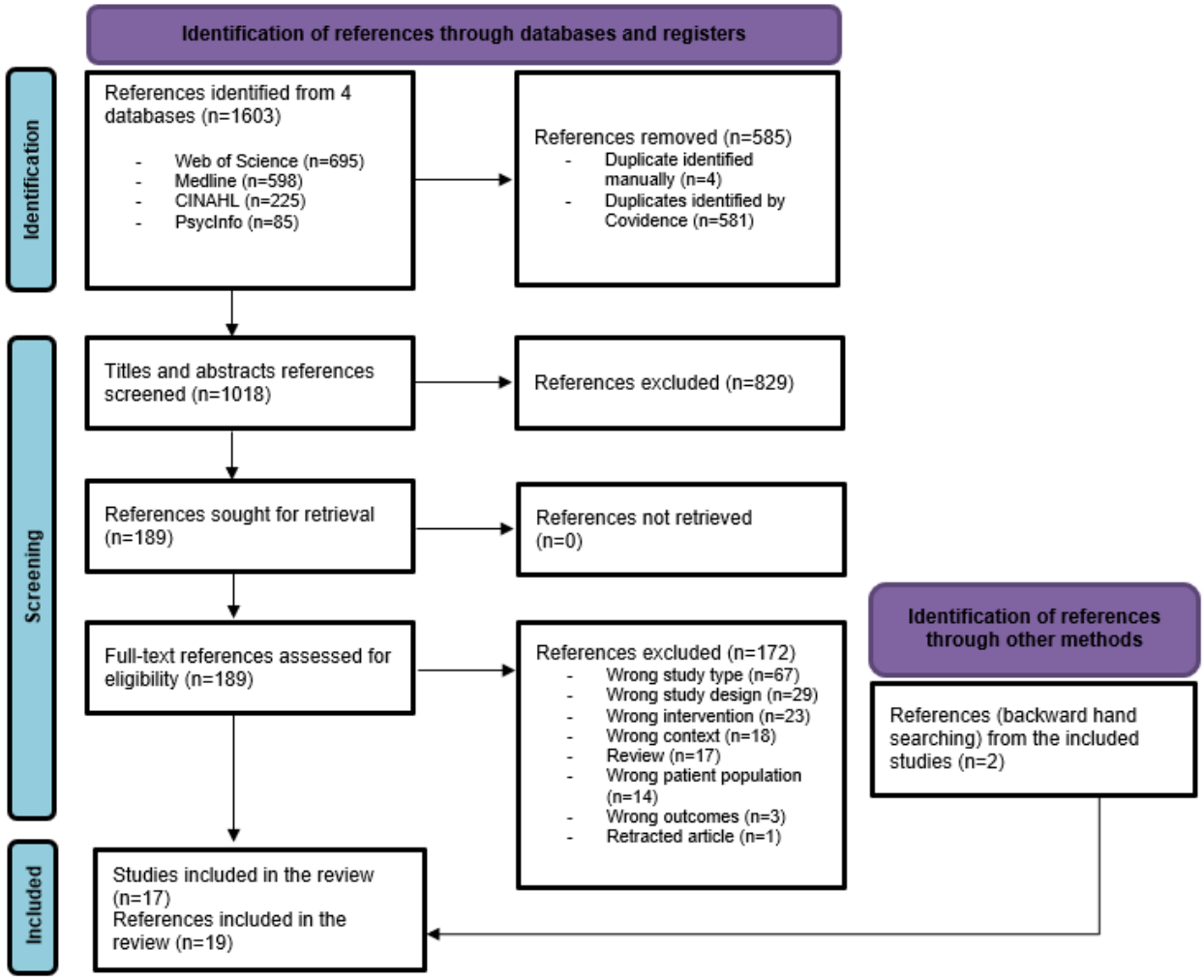

The relatively high number of exclusions at the full-text review stage (172/189, 91%) can be attributed to our inclusive approach in the previous stage. For example, some reviews (17/189, 9%) and incorrect study types (67/189, 35%), such as editorials, commentaries, or conference papers, were excluded at this stage. Other exclusion reasons (88/189, 47%) included models that lacked AI components, models focusing on health care operational processes (eg, workflow modeling), studies targeting populations receiving specialized care (eg, hospitalized or cancer patients), interventions such as imaging research that were outside the scope of CBPHC, and methods for mitigating bias that were applied to the AI model itself (eg, biased predictions of treatment effects) rather than addressing biases related to diverse groups or personal attributes.

Overview of Included Studies

Of the 17 included studies published between 2019 and 2022, we identified 7 studies in the discipline of data science or informatics, 7 in medical informatics, 1 in medical ethics and informatics, 1 in medical ethics using a Delphi method, and 1 in management care ethics using a user-centered design. Most studies have been conducted in the United States (15/17, 88%), 1 in the United Kingdom, and 1 in Italy. The main characteristics of the included studies can be found in Multimedia Appendix 2.

Quality Assessment of Included Studies

Most studies had a quantitative descriptive study design (14/17, 82%), while 2 used a mixed methods design, and 1 used a qualitative design. All studies showed high quality, receiving scores of 3 or 4 stars (on a possibility of 5). All MMAT scores can be found in Multimedia Appendix 3.

Diverse Groups Considered

The most frequently studied protected attributes were race (or ethnicity), examined in 71% (12/17) of studies, and sex (defined as binary male versus female), considered in 59% (10/17) of studies. None of the studies distinguished between biological sex and socially constructed gender, and 5 of them incorrectly identified sex as gender. Race or ethnicity was most often categorized as White or Black, Black or non-Black or, in one study, as Asian, Black, White, and other.

Other protected attributes considered by the studies included age (7/17, 41%), socioeconomic status or its proxies, such as income, work class, education, health care insurance (5/17, 29%), place of residence (2/17, 12%), marital status (1/17, 6%), and disability status (1/17, 6%).

XSL•FO
RenderX

Categorization of Deployed Bias Mitigation Strategies

We identified considerable heterogeneity across the studies, which used various strategies and methods to assess and mitigate bias in algorithms impacting diverse groups. We categorized these efforts into four groups: (1) addressing bias in existing AI models or datasets, (2) mitigating biases from data sources such as electronic health records (EHRs), (3) developing tools that incorporate a "human-in-the-loop" approach, and (4) identifying ethical principles to guide informed decision-making.

Attempts in Existing AI Models or Datasets

We identified 7 studies that attempted to mitigate biases in existing AI models or datasets [19,20,27,28,35,37,39].

A debiasing attempt was made on an insurance coverage algorithm designed to identify individuals who could benefit from health resources according to their health needs [35]. Risk scores were initially calculated based on projected future costs rather than uncontrolled or unmanaged illnesses, disadvantaging Black patients. By changing the data labeling to focus on future illness rather than future costs, the percentage of Black patients who could benefit from health resources increased significantly [35].

Another cohort study [37] using a Medicaid enrollees' dataset showed that reweighing was more effective at reducing bias in postpartum depression risk scores between White and Black individuals compared with training without the race variable for comparison. Initially, it was found that the White individuals had higher rates of postpartum depression and mental health service use. However, after comparing postpartum depression rates between races based on population surveys, it became clear that the higher rates in White women might be due to disparities in the timely assessment, screening, and detection of symptoms in Black women [37].

A total of three other studies include (1) retraining models with data that incorporated health equity measures resulted in a slight decrease in performance for detecting abnormal electrocardiograms but significantly reduced gender, race and age biases [19]; (2) increasing diversity in the training data of a predictive pulmonary disease model improved its performance [27]; and (3) although a mental health assessment model achieved high accuracy, its performance was statistically higher and more accurate for men than for women [18]. The use of an algorithmic disparate remover, by adjusting the modeling data, significantly reduced this disparity while maintaining high accuracy [20].

Another attempt to assess bias involved replicating models predicting liver disease [39]. Importing an existing dataset reproduced predictive models with high accuracy but revealed a previously unobserved bias, with women experiencing a higher false negative rate.

We identified only 1 in-processing debiasing attempt [28]. Out of 2 algorithmic fairness strategies, group recalibration and equalized odds, were used to recalibrate a predictive model of cardiovascular diseases that was not initially adjusted for attributes such as sex or race. This resulted in an exacerbation of false positive and negative rates differences between groups, as well as overall model miscalibration.

Attempts in Data Sourcing

We identified 5 studies that attempted to mitigate biases in data sourcing [26,31,32,38,40].

Based on published synthetic datasets, such as the analysis of the American Time Use Survey dataset, using fairness metrics revealed potential discrepancies in representativeness between real and synthetic data across age, sex, and race [26].

Out of 4 other studies investigated EHRs datasets [31,32,38,40]. A natural language processing model was developed to extract vital sign features from unstructured notes, comparing risk scores with 2 convenience samples. This method reduced the missingness of vital signs by 31%, thereby mitigating possible discrimination toward diverse groups, such as Black men or Black women [32]. Based on data from a previous study, 2 ML models were trained to compare balanced error rates across different socioeconomic status levels and the incompleteness of EHRs data [31]. Asthmatic children with lower socioeconomic status exhibited larger balanced error rates than those with higher socioeconomic status and had more missing information regarding asthma care, severity, or undiagnosed asthma, despite meeting asthma criteria [31].

Potential bias based on place of residence in EHRs was examined by 2 studies [38,40]. Rebalancing class labels by adding zip-code level information to 19,367 EHRs during the preprocessing step showed no significant deviation in performance, indicating that bias can be mitigated through preprocessing [38]. Meanwhile, a simple 30-day readmission prediction model was developed, categorizing each patient as local (nearby) or not (far) [40]. The performance with and without this variable was assessed, revealing no significant differences. Considering that living locally only affects the observability of the outcome (eg, a patient may be readmitted to a different hospital), differential bias assessment cannot rely solely on observed data [40].

Attempts in Developing Tools With a "Human-in-the-Loop" Approach

We identified 3 studies that attempted to mitigate biases by incorporating a "human-in-the-loop" approach [29,30,36].

These studies led to the development of "human-in-the-loop" tools: (1) a visual tool for auditing and mitigating bias from tabular datasets, which was tested through experiments on 3 datasets with user participation and significantly reduced bias compared with another commercial debiasing toolkit [29]; (2) pragmatic tools developed for better use of risk scores with a Medicare members' dataset, allowing users to identify appropriate risk scores for each subgroup to achieve equality of opportunity [30]; and (3) a tool called "FairLens" capable of identifying and explaining biases, which was tested using a fictitious black box model serving as a decision support system [36]. Empirically validated by injecting biases into this fictitious decision support system, this tool outperformed other standard measures and enabled experts to identify problematic groups

XSL•FO
RenderX

or affected patients, thereby allowing for the detection of potential misclassification [36].

Attempts at Identifying Ethical Principles for Informed Decision-Making

We identified 2 empirical studies that attempted to mitigate biases by identifying ethical principles for informed decision-making [33,34].

To assess the potential missingness of EHR data from phenotyping technology, a Delphi study was conducted to address ethical challenges and reach a consensus on the importance of privacy, transparency, consent, accountability, and fairness [33]. In addition, a user-centered design study was conducted to identify user requirements, mainly intended for health managers and clinicians, to support informed decision-making and confidence in using a hepatitis C severity illness predictive model prototype [34].

Discussion

Principal Findings

The reviewed studies illustrate a multifaceted approach to mitigating bias in primary care AI models. Strategies include retraining, reweighing, relabeling, adding more diversity, and attempting to replicate existing modeling data [19,20,27,35,37,39], as well as algorithmic recalibration applied to an existing prediction model [28]. Other strategies involve the development and application of fairness metrics to ensure equitable distributions in previously published databases [26], and the identification of missingness in EHRs datasets by rebalancing class labels or adding information [31,32,38]. Another group of strategies includes the introduction of visual interactive tools for human-in-the-loop bias auditing [29,30,36]. All these attempts cover a broad spectrum of interventions, ranging from data preprocessing and algorithmic modification to post hoc analysis, demonstrating the complexity and variety of approaches needed to address bias in AI models in primary health care.

The studies collectively address a wide range of protected attributes [1,8], including race or ethnicity [19,26,28-37], sex [19,20,26-31,36,39], age [19,26,27,29-31,36], socioeconomic status (SES) [27,29,31,33,36], and other demographic variables such as place of residence [38,40]. This underlines the recognition of the multifaceted nature of bias, which can intersect across various dimensions of identity and social determinants of health [9,42]. However, we have identified disparities in the number of protected attributes studied. Race (White vs Black) and sex (male vs female) are most frequently investigated, whereas other attributes, such as disability and gender, are underresearched or not studied at all.

Bias mitigation efforts reveal a nuanced landscape where attempts to reduce bias across protected attributes can result in complex trade-offs with model performance. For example, a decrease in overall model performance accompanied by significant reductions in bias was observed following the implementation of constrained optimization [19]. Similarly, improvements in calibration for specific groups came at the cost of increased disparities in false positive and false negative rates between groups [28]. Despite these trade-offs, the efforts have largely been successful in reducing bias, as evidenced by a study that achieved fairer distributions in synthetic data [26], and in another study where human-in-the-loop interventions significantly reduced bias while maintaining utility [29].

These empirical findings reinforce theoretical insights that emphasize the importance of health equity between protected and unprotected attributes [1,8]. To mitigate bias in AI health models, distributive justice options for ML have been proposed: (1) equal patient outcomes; (2) equal performance; and (3) equal allocation of resources [1]. Since these different types of fairness options are often incompatible, optimizing all these parameters seems challenging, as demonstrated by an identified study [28]. Trade-offs are essential, and a participatory process involving key stakeholders, including ethicists, clinicians, and marginalized populations, is strongly encouraged [1]. While striving to create ethically robust AI models, selected studies often reveal tension, as efforts to reduce bias can sometimes lead to a decrease in the model's overall performance. This presents a critical challenge: balancing the imperative of fairness with the need to maintain high accuracy and efficiency in algorithmic outputs.

Comparison With Previous Work

Initiatives focused on the fair use of AI in health care and the assessment of bias risk in AI predictive models have been published in recent years. Notable initiatives include Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) and Standard Protocol Items Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) [43], which provide guidelines for the ethical presentation of the results of trials conducted with AI in the health care field. To assess the risk of bias in diagnostic and prognostic prediction model studies, the "Prediction Model Risk of Bias Assessment Tool" (PROBAST) [44] can be used. PROBAST consists of a list of signaling questions grouped into 4 categories: participants, predictors, outcomes, and analysis. This tool was used in a systematic scoping review to assess the quality of primary studies reporting applications of AI in CBPHC [45].

However, the objective of our scoping review differs; it is not to identify biases in the AI prediction models themselves, but rather to examine biases toward groups that are underrepresented or misrepresented in these AI models. An identified review has used and adapted PROBAST to assess related protected attributes, but the AI predictive models studied were hospital-based and not relevant to primary care [11]. We also identified a scoping review protocol that focused on bias toward diverse groups in AI systems in primary care; however, unless we are mistaken, the results of this protocol have never been published [10]. Another identified review aimed to assess age-related bias in AI but did not focus on primary health care [46]. Finally, we identified another systematic review investigating health inequities in primary care, but it adopted a system-wide perspective, focusing on aspects such as patient consultation and effects on health systems [47].

XSL•FO
RenderX

To our knowledge, no other published review has the objectives of identifying (1) the bias mitigation strategies or methods in primary health care, (2) the diverse groups that are underrepresented or misrepresented, and (3) the results of bias mitigation and AI model performance.

Strengths and Limitations

The strengths of this review include results that can be translated into recommendations for various stakeholders, such as AI developers, researchers, and decision makers. However, we acknowledge some limitations. First, we limited our search strategy to the last 5 years before November 2022 and focused on 4 databases, which may have excluded some relevant studies. Second, the extraction of studies and quality assessment were conducted only once, although all of them were validated by at least one senior researcher. Third, due to the heterogeneity of the studies, we were unable to combine results through a quantitative synthesis and remained at a narrative level of reporting. Finally, our review primarily identified research from a North American setting, which reduces its transferability to other continents.

Future Directions and Dissemination Plan

This scoping review serves as the initial phase of the iterative project "Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse, and Equitable AI" (PREMIA).

Following the results of this review, we have developed a framework currently validated by a diverse group of experts, including clinicians, public health managers, primary care researchers, data scientists, and patient and citizen partners. This group is concentrating on existing AI predictive models and the bias mitigation strategies identified in our scoping review. Diverse populations, such as older adults, individuals with disabilities, and people from various racial and ethnic backgrounds, are actively involved in this second phase of PREMIA. We plan to prepare and submit a manuscript based on the findings of this Delphi study.

In addition, in recognition of the rapid advancements in this field, we plan to update this literature review in 2027 using a similar search strategy. This iterative approach will allow us to refine our framework and track the progress of bias mitigation in AI models within primary health care. Indigenous peoples in Canada represent a group historically underrepresented in health research, leading to inequities [3]. Since no other study has addressed bias related to Indigenous status, we collaborate with Indigenous representatives to develop methods for mitigating this bias in CBPHC algorithms.

Conclusion

This review identifies strategies and methods for mitigating bias in primary health care algorithms, considers diverse groups based on their personal or protected attributes, and examines the results of bias attenuation and model performance. The findings suggest that biases toward diverse groups can be more effectively mitigated when data are open-sourced, multiple stakeholders are involved, and during the preprocessing stage of algorithm development. More empirical studies are needed, with a focus on including participants who embrace greater diversity, such as nonbinary gender identities or Indigenous peoples in Canada.

Acknowledgments

The Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse, and Equitable AI project is funded by the International Observatory on the Societal Impacts of AI and Digital Technology. The authors would like to thank Karine Gentelet for her contribution to the study's design and for obtaining the funding.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1-3.

Authors' Contributions

MS, MPG, CR, VC, PD, JSP, and DD designed the study and obtained the funding. MS, MPG, SO, and FB developed the search strategy. MS, SO, MS, CR, MPG, VC, and FB participated in the screening of sources. MS, SO, and MPG conducted the data extraction. SO, MS, and MPG completed the first draft of the manuscript, and all authors participated in the revision and editing of the manuscript versions. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database Search Strategies.
[PDF File (Adobe PDF File), 84 KB - jmir_v27i1e60269_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

Characteristics of Included Studies.
[PDF File (Adobe PDF File), 328 KB - jmir_v27i1e60269_app2.pdf]

Multimedia Appendix 3
Quality assessment: MMAT (Mixed-Methods Appraisal Tool) scores.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v27i1e60269_app3.xlsx]

Multimedia Appendix 4
PRISMA-ScR checklist.
[DOCX File , 109 KB - jmir_v27i1e60269_app4.docx]

References

1. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169(12):866-872 [FREE Full text] [doi: 10.7326/M18-1990] [Medline: 30508424]
2. Qu Y, Wei C, Du P, Che W, Zhang C, Ouyang W, et al. Integration of cognitive tasks into artificial general intelligence test for large models. iScience 2024;27(4):109550 [FREE Full text] [doi: 10.1016/j.isci.2024.109550] [Medline: 38595796]
3. Gurevich E, El Hassan B, El Morr C. Equity within AI systems: what can health leaders expect? Healthc Manage Forum 2023;36(2):119-124 [FREE Full text] [doi: 10.1177/08404704221125368] [Medline: 36226507]
4. Alabdulmohsin I, Lucic M. A near-optimal algorithm for debiasing trained machine learning models. ArXiv. Preprint posted online on August 23, 2022 2022 [FREE Full text]
5. van Leeuwen KG, de Rooij M, Schalekamp S, van Ginneken B, Rutten MJCM. How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr Radiol 2022;52(11):2087-2093 [FREE Full text] [doi: 10.1007/s00247-021-05114-8] [Medline: 34117522]
6. Kang J, Hanif M, Mirza E, Khan MA, Malik M. Machine learning in primary care: potential to improve public health. J Med Eng Technol 2021;45(1):75-80. [doi: 10.1080/03091902.2020.1853839] [Medline: 33283565]
7. Canadian Institutes of Health Research. Community-based primary health care. URL: https://cihr-irsc.gc.ca/e/43626.html [accessed 2024-04-01]
8. Cochrane methods. PROGRESS-Plus. URL: https://methods.cochrane.org/equity/projects/evidence-equity/progress-plus [accessed 2024-04-01]
9. Delgado J, de Manuel A, Parra I, Moyano C, Rueda J, Guersenzvaig A, et al. Bias in algorithms of AI systems developed for COVID-19: a scoping review. J Bioeth Inq 2022;19(3):407-419 [FREE Full text] [doi: 10.1007/s11673-022-10200-z] [Medline: 35857214]
10. Wang JX, Somani S, Chen JH, Murray S, Sarkar U. Health equity in artificial intelligence and primary care research: protocol for a scoping review. JMIR Res Protoc 2021;10(9):e27799 [FREE Full text] [doi: 10.2196/27799] [Medline: 34533458]
11. Wang HE, Landers M, Adams R, Subbaswamy A, Kharrazi H, Gaskin DJ, et al. A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. J Am Med Inform Assoc 2022;29(8):1323-1333 [FREE Full text] [doi: 10.1093/jamia/ocac065] [Medline: 35579328]
12. Sasseville M, Ouellet S, Rhéaume C, Couture V, Després P, Paquette JS, et al. Risk of bias mitigation for vulnerable and diverse groups in community-based primary health care artificial intelligence models: protocol for a rapid review. JMIR Res Protoc 2023;12:e46684 [FREE Full text] [doi: 10.2196/46684] [Medline: 37358896]
13. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
14. Apply PCC. University of South Australia. URL: https://guides.library.unisa.edu.au/ScopingReviews/ApplyPCC [accessed 2024-04-01]
15. Veritas health innovation. Covidence. URL: https://www.covidence.org/ [accessed 2024-04-01]
16. Handsearching. Cochrane. URL: https://training.cochrane.org/resource/tsc-induction-mentoring-training-guide/5-handsearching [accessed 2024-04-01]
17. Reyna M, Sadr N, Gu A, Perez Alday EA, Liu C. Will two do? Varying dimensions in electrocardiography: the physioNet/computing in cardiology challenge 2021 v1.0.3. physionet.org. URL: https://physionet.org/content/challenge-2021/1.0.3/ [accessed 2024-01-11]
18. Singh VK, Long T. Automatic assessment of mental health using phone metadata. Proc. Assoc. Info. Sci. Tech 2019;55(1):450-459. [doi: 10.1002/pra2.2018.14505501049]
19. Perez Alday EA, Rad AB, Reyna MA, Sadr N, Gu A, Li Q, et al. Age, sex and race bias in automated arrhythmia detectors. J Electrocardiol 2022;74:5-9. [doi: 10.1016/j.jelectrocard.2022.07.007] [Medline: 35878534]
20. Park J, Arunachalam R, Silenzio V, Singh VK. Fairness in mobile phone-based mental health assessment algorithms: exploratory study. JMIR Form Res 2022;6(6):e34366 [FREE Full text] [doi: 10.2196/34366] [Medline: 35699997]

XSL•FO
RenderX

21. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020;27(12):2011-2015 [FREE Full text] [doi: 10.1093/jamia/ocaa088] [Medline: 32594179]
22. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. EFI 2018;34(4):285-291. [doi: 10.3233/efi-180221]
23. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. Mixed methods appraisal tool (MMAT) version 2018: user guide. Mixed Methods Appraisal Tool. 2018. URL: https://content.iospress.com/articles/education-for-information/efi180221 [accessed 2024-04-01]
24. Pollock D, Peters MDJ, Khalil H, McInerney P, Alexander L, Tricco AC, de Moraes, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Bhanot K, Qi M, Erickson JS, Guyon I, Bennett KP. The problem of fairness in synthetic healthcare data. Entropy (Basel) 2021;23(9):1165 [FREE Full text] [doi: 10.3390/e23091165] [Medline: 34573790]
27. Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. Front Artif Intell 2020;3:561802 [FREE Full text] [doi: 10.3389/frai.2020.561802] [Medline: 33981989]
28. Foryciarz A, Pfohl SR, Patel B, Shah N. Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation. BMJ Health Care Inform 2022;29(1):e100460 [FREE Full text] [doi: 10.1136/bmjhci-2021-100460] [Medline: 35396247]
29. D-BIAS: a causality-based human-in-the-loop system for tackling algorithmic bias. IEEE Journals & Magazine | IEEE Xplore. URL: https://ieeexplore.ieee.org/document/9903601/authors#authors [accessed 2024-01-11]
30. Hane CA, Wasserman M. Designing equitable health care outreach programs from machine learning patient risk scores. Med Care Res Rev 2023;80(2):216-227 [FREE Full text] [doi: 10.1177/10775587221098831] [Medline: 35685000]
31. Juhn YJ, Ryu E, Wi C, King KS, Malik M, Romero-Brufau S, et al. Assessing socioeconomic bias in machine learning algorithms in health care: a case study of the HOUSES index. J Am Med Inform Assoc 2022;29(7):1142-1151 [FREE Full text] [doi: 10.1093/jamia/ocac052] [Medline: 35396996]
32. Khurshid S, Reeder C, Harrington LX, Singh P, Sarma G, Friedman SF, et al. Cohort design and natural language processing to reduce bias in electronic health records research. NPJ Digit. Med 2022;5(1):47. [doi: 10.1038/s41746-022-00590-0]
33. Martinez-Martin N, Greely HT, Cho MK. Ethical development of digital phenotyping tools for mental health applications: delphi study. JMIR Mhealth Uhealth 2021;9(7):e27343 [FREE Full text] [doi: 10.2196/27343] [Medline: 34319252]
34. Nong P, Raj M, Platt J. Integrating predictive models into care: facilitating informed decision-making and communicating equity issues. Am J Manag Care 2022;28(1):18-24 [FREE Full text] [doi: 10.37765/ajmc.2022.88812] [Medline: 35049257]
35. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
36. Panigutti C, Perotti A, Panisson A, Bajardi P, Pedreschi D. FairLens: auditing black-box clinical decision support systems. Inf Process Manag 2021;58(5):102657. [doi: 10.1016/j.ipm.2021.102657]
37. Park Y, Hu J, Singh M, Sylla I, Dankwa-Mullan I, Koski E, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. JAMA Netw Open 2021;4(4):e213909 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3909] [Medline: 33856478]
38. Seker E, Talburt JR, Greer ML. Preprocessing to address bias in healthcare data. Stud Health Technol Inform 2022;294:327-331. [doi: 10.3233/SHTI220468] [Medline: 35612086]
39. Straw I, Wu H. Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. BMJ Health Care Inform 2022;29(1):e100457 [FREE Full text] [doi: 10.1136/bmjhci-2021-100457] [Medline: 35470133]
40. Yan M, Pencina MJ, Boulware LE, Goldstein BA. Observability and its impact on differential bias for clinical prediction models. J Am Med Inform Assoc 2022;29(5):937-943 [FREE Full text] [doi: 10.1093/jamia/ocac019] [Medline: 35211742]
41. PRISMA Flow diagram. URL: https://www.prisma-statement.org/prisma-2020-flow-diagram [accessed 2024-04-12]
42. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2023;2(6):e0000278 [FREE Full text] [doi: 10.1371/journal.pdig.0000278] [Medline: 37347721]
43. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
44. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]

XSL•FO
RenderX

45. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. J Med Internet Res 2021;23(9):e29839 [FREE Full text] [doi: 10.2196/29839] [Medline: 34477556]
46. Chu C, Donato-Woodger S, Khan SS, Nyrup R, Leslie K, Lyn A, et al. Age-related bias and artificial intelligence: a scoping review. Humanit Soc Sci Commun 2023;10(1) [FREE Full text] [doi: 10.1057/s41599-023-01999-y]
47. d'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Fam Med Community Health 2022;10(Suppl 1):e001670 [FREE Full text] [doi: 10.1136/fmch-2022-001670] [Medline: 36450391]

Abbreviations

AI: artificial intelligence
CBPHC: community-based primary health care
CONSORT-AI: Consolidated Standards of Reporting Trials-Artificial Intelligence
EHR: electronic health record
JBI: Joanna Briggs Institute
ML: machine learning
MMAT: Mixed-Methods Appraisal Tool
PREMIA: Protecting and Engaging Vulnerable Populations in the Development of Predictive Models in Primary Health Care for Inclusive, Diverse and Equitable AI
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
PROBAST: Prediction model Risk Of Bias Assessment Tool
PROGRESS: Place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital
SES: socioeconomic status
SPIRIT-AI: Standard Protocol Items Recommendations for Interventional Trials-Artificial Intelligence

Edited by A Coristine; submitted 06.05.24; peer-reviewed by J Bensemann, L He; comments to author 14.09.24; revised version received 26.09.24; accepted 07.11.24; published 07.01.25.

Please cite as:
Sasseville M, Ouellet S, Rhéaume C, Sahlia M, Couture V, Després P, Paquette JS, Darmon D, Bergeron F, Gagnon MP
Bias Mitigation in Primary Health Care Artificial Intelligence Models: Scoping Review
J Med Internet Res 2025;27:e60269
URL: https://www.jmir.org/2025/1/e60269
doi:10.2196/60269
PMID:

©Maxime Sasseville, Steven Ouellet, Caroline Rhéaume, Malek Sahlia, Vincent Couture, Philippe Després, Jean-Sébastien Paquette, David Darmon, Frédéric Bergeron, Marie-Pierre Gagnon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

AI Interventions to Alleviate Healthcare Shortages and Enhance Work Conditions in Critical Care: Qualitative Analysis

Nadine Bienefeld[1], PD, PhD; Emanuela Keller[2], MD, Prof Dr; Gudela Grote[1], Prof Dr

[1]ETH Zurich, Zurich, Switzerland

[2]University Hospital of Zurich, Zurich, Switzerland

Corresponding Author:
Nadine Bienefeld, PD, PhD
ETH Zurich
Weinbergstrasse 56/58
Zurich, 8093
Switzerland
Phone: 41 44 632 70 78
Email: nbienefeld@ethz.ch

Abstract

Background: The escalating global scarcity of skilled health care professionals is a critical concern, further exacerbated by rising stress levels and clinician burnout rates. Artificial intelligence (AI) has surfaced as a potential resource to alleviate these challenges. Nevertheless, it is not taken for granted that AI will inevitably augment human performance, as ill-designed systems may inadvertently impose new burdens on health care workers, and implementation may be challenging. An in-depth understanding of how AI can effectively enhance rather than impair work conditions is therefore needed.

Objective: This research investigates the efficacy of AI in alleviating stress and enriching work conditions, using intensive care units (ICUs) as a case study. Through a sociotechnical system lens, we delineate how AI systems, tasks, and responsibilities of ICU nurses and physicians can be co-designed to foster motivating, resilient, and health-promoting work.

Methods: We use the sociotechnical system framework COMPASS (Complementary Analysis of Sociotechnical Systems) to assess 5 job characteristics: autonomy, skill diversity, flexibility, problem-solving opportunities, and task variety. The qualitative analysis is underpinned by extensive workplace observation in 6 ICUs (approximately 559 nurses and physicians), structured interviews with work unit leaders (n=12), and a comparative analysis of data science experts' and clinicians' evaluation of the optimal levels of human-AI teaming.

Results: The results indicate that AI holds the potential to positively impact work conditions for ICU nurses and physicians in four key areas. First, autonomy is vital for stress reduction, motivation, and performance improvement. AI systems that ensure transparency, predictability, and human control can reinforce or amplify autonomy. Second, AI can encourage skill diversity and competence development, thus empowering clinicians to broaden their skills, increase the polyvalence of tasks across professional boundaries, and improve interprofessional cooperation. However, careful consideration is required to avoid the deskilling of experienced professionals. Third, AI automation can expand flexibility by relieving clinicians from administrative duties, thereby concentrating their efforts on patient care. Remote monitoring and improved scheduling can help integrate work with other life domains. Fourth, while AI may reduce problem-solving opportunities in certain areas, it can open new pathways, particularly for nurses. Finally, task identity and variety are essential job characteristics for intrinsic motivation and worker engagement but could be compromised depending on how AI tools are designed and implemented.

Conclusions: This study demonstrates AI's capacity to mitigate stress and improve work conditions for ICU nurses and physicians, thereby contributing to resolving health care staffing shortages. AI solutions that are thoughtfully designed in line with the principles for good work design can enhance intrinsic motivation, learning, and worker well-being, thus providing strategic value for hospital management, policy makers, and health care professionals alike.

(J Med Internet Res 2025;27:e50852) doi:10.2196/50852

XSL•FO
RenderX

KEYWORDS

artificial intelligence; AI; work design; sociotechnical system; work; job; occupational health; sociotechnical; new work; future of work; satisfaction; health care professionals; intensive care; ICU; stress mitigation; worker; employee; stress; health care professional; overburdened; burden; burnout; autonomy; competence; flexible; task; workplace; hospital

Introduction

According to projections by the World Health Organization, by 2030, the world will witness a global deficit of 18 million health care workers, more than double the current shortfall, with potentially far-reaching implications for patient care and health care delivery [1,2]. This alarming prediction underscores not only a public health crisis but also a wider social and economic dilemma, as the shortage of health care workers significantly impairs the resilience of health systems, undermines public health outcomes, and exacerbates inequities in health care access [3]. This global concern is attributed to demographic changes such as an aging population, an increase in chronic illnesses, and a scarcity of both educational and financial resources [4,5]. Escalating this issue are high rates of absence and turnover, driven by an unprecedented level of physician and nurse burnout due to psychosocial stressors and occupational hazards such as exposure to infectious diseases, shiftwork, or ergonomic challenges, that place health care workers at risk for long-term mental and physical health consequences [6-9]. These conditions are aggravated by the emotional labor and the high-stakes environment of health care settings, resulting in professionals leaving their careers due to overwork, demotivation, or dissatisfaction [10,11].

Artificial intelligence (AI) and machine learning–based applications have emerged as promising tools to address these challenges by potentially improving work conditions and mitigating personnel shortages [12-14]. Yet, the adoption of AI in real-world health care settings has been notably sluggish, a trend associated, in part, with high acquisition costs, increased training needs, and other implementation challenges such as the "black box" nature of AI systems [15,16]. Due to the opacity of current AI systems, health care practitioners often are uncertain about the AI decision-making process, which has implications for trust and AI acceptance [15,17,18]. Moreover, it is a fallacy to presume that AI will automatically enhance human performance and work conditions. As past evidence shows, the introduction of digital technologies intended to improve the work of health care professionals has, at times, introduced new burdens on workers due to systems being ill-designed, with scant regard for the nuances of organizational and worker needs [19-23].

For technology to have the desired positive impact on performance and to engender motivating working conditions for health care staff, a comprehensive sociotechnical system design approach is required—one that incorporates principles of human-centered design along with work design [24,25]. This approach advocates for technology, organizational processes, and human tasks to be jointly designed and implemented to fully harness the benefits of a new technology such as AI [26,27].

While there is abundant research on the human-centered design of AI in health care—with a primary focus on enhancing the user experience and facilitating the interaction between single users and AI [28]—research often overlooks how multiple clinicians use AI collaboratively as an integral part of their work tasks and professional roles. Therein lies a research gap: understanding the use of AI beyond individual interactions to encompass the multiprofessional dynamic collaborations so vital to today's complex health care delivery systems.

We, therefore, examine the design of human tasks and processes within the framework of sociotechnical systems in the introduction of AI within dynamic interprofessional team collaboration in critical care settings. We apply the well-established and empirically validated work system assessment tool COMPASS (Complementary Analysis of Sociotechnical Systems [29-31]), to analyze structured data from workplace observation and interviews and to assess five job characteristics essential for enhancing intrinsic motivation, learning, and resilience among health care workers [32-34]. As depicted in Figure 1, the five job characteristics of human-centered work design are (1) autonomy in decision-making; (2) skill variety and competence development; (3) flexibility in terms of time, place, and type of work; (4) problem-solving opportunities; and (5) task identity and variety.

XSL•FO
RenderX

Figure 1. Five job characteristics for motivating, resilient, and health-promoting work.

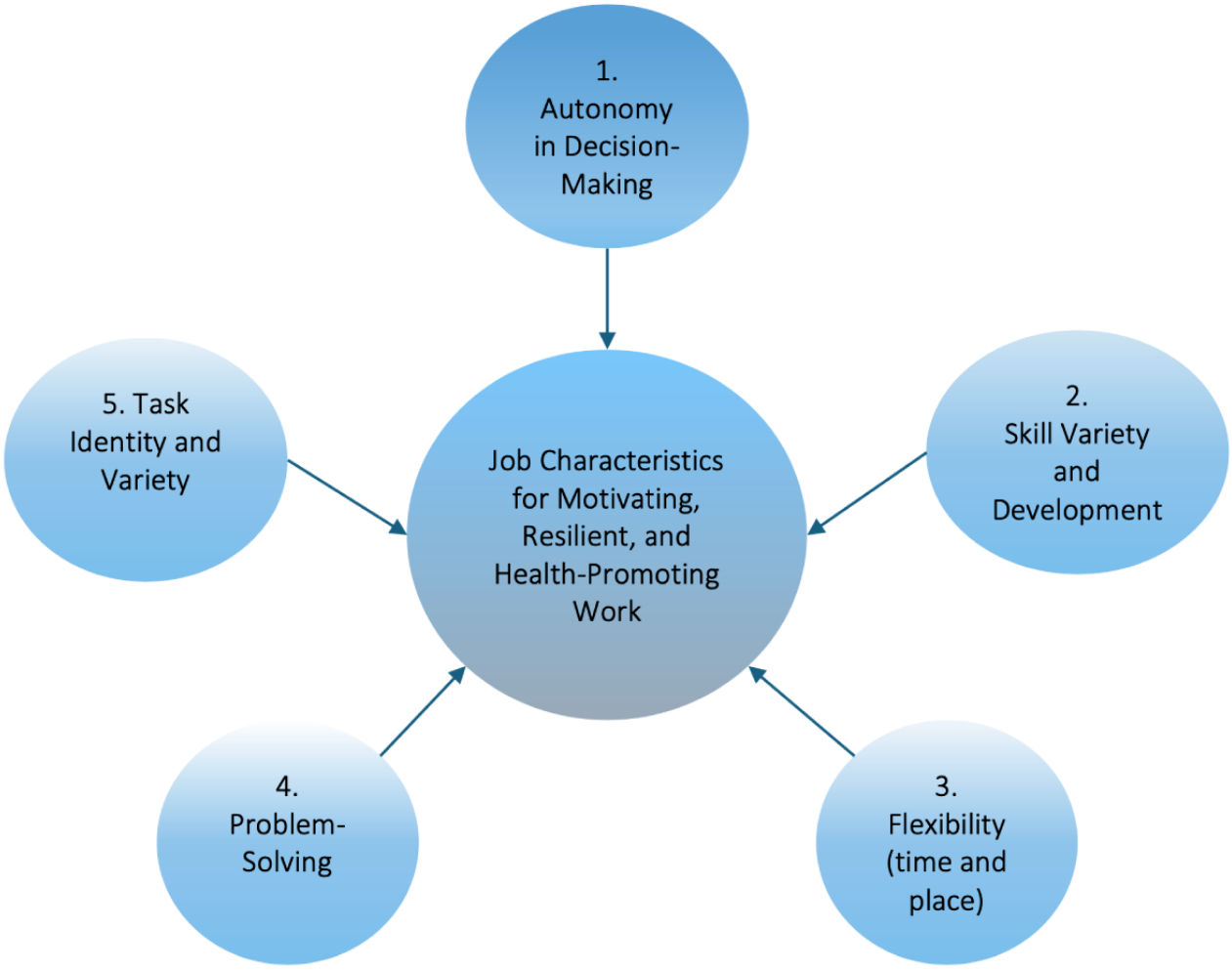

Our study of the sociotechnical integration of AI centers around critical care teams, specifically intensive care unit (ICU) nurses and physicians. The ICU context was selected due to the particularly high job demands and increased rates of burnout, which were exacerbated by the extreme work conditions during the SARS-CoV-2 pandemic [35-37]. Because the joint optimization of social and technical aspects of work systems must be considered before a technology is introduced into the workplace [38], we first assess the current work situation in ICUs to then compare it with future work scenarios after AI implementation. This comparison is based on the optimal levels of human-AI teaming for core ICU tasks as identified in a prior study [39] and key job characteristics for motivating, resilient, and health-promoting work. Our findings lead to actionable recommendations for AI design and implementation that alleviate working conditions for critical care nurses and physicians to ultimately improve the quality of patient care.

Methods

Ethical Considerations

Ethics approval was obtained from the ETH Zurich Ethics Commission (EK 2019-N-51 and EK 2019-N-190). Informed consent was obtained from participants before any form of data collection and the option to opt-out was provided. All data were anonymized. Participants did not receive any form of compensation. No patient data or information were collected.

Study Context

The study was carried out in six ICUs that were selected based on unit size (an average of 10 to 14 beds per unit) and patient scope (a capacity of about 800 patients treated per year). Of the 6 ICUs contacted, all agreed to participate in the study. Working shifts lasted 8.5 hours (morning 7 AM to 3:30 PM, afternoon 3 PM to 11:30 PM, and night 11 PM to 7:30 AM).

Sampling Strategy

In line with the COMPASS guidelines (Multimedia Appendix 1), we observed 5 morning shifts in each ICU, resulting in a total of 30 shifts lasting 8.5 hours each. Contingent upon each ICU's number of beds, the team composition varied slightly, with 6 to 8 physicians and 8 to 12 nurses working together for each shift. This resulted in a total of approximately 559 health care professionals observed in this study. All nurses were registered nurses. Nurses in training and nurse practitioners were excluded from the study as they could have distorted the results based on their lower level of expertise and potentially different experience of work demands. Resident physicians oversaw up to 5 patients and attending physicians were ultimately responsible for all patients in the unit. Due to the relatively small number of attending physicians working in each

XSL•FO
RenderX

unit, we combined residents and attending physicians into one professional group, hereafter referred to as physicians. To complement the worker perspectives, we additionally conduct n=12 structured interviews (6 head physicians, 5 nursing staff heads, and one head of nursing education).

Work System Assessment

Assessment Tool

We used the well-established and empirically validated work system assessment tool COMPASS [29-31,40] to assess and compare the current and future work states across the 6 ICUs. COMPASS analyzes structured data from workplace observations and work unit head interviews using normative criteria based on sociotechnical systems theory (Multimedia Appendix 1).

Workplace Observations

The first author (NB)—a work and organizational psychologist with 15 years of experience in work system analysis and qualitative research methodology—applied the structured observational guidelines provided by COMPASS, to observe the work practices of ICU nurses and physicians in-situ (see the full set of observational guidelines and criteria in Multimedia Appendix 1). As a nonparticipant observer, she took detailed notes on a notepad standing in the background about five feet away, and where suitable, asked clarifying questions from health care professionals.

Interviews With Work Unit Heads

In addition to workplace observations, the first author (NB) also conducted 12 structured COMPASS interviews with each of the 6 ICU head physicians, 5 nursing staff heads, and 1 head of nursing education. Interview questions related to the overall work system structure, professional roles and responsibilities, work process variances, and disturbances (see the full list of interview questions in Multimedia Appendix 2). Interviews were conducted in single-person offices, lasted around 60 minutes, and were audio-recorded and transcribed ad verbatim.

Data Analysis

Interview transcripts and notes from observations were methodically analyzed using the COMPASS guidelines [29,30] and a grounded theory approach [41] to ensure qualitative rigor in qualitative research [42]. The software MAXQDA (version 2020; Verbi) was used to facilitate the analytical process [43]. Qualitative results are reported as per the standard published by the Academy of Medicine [44] (see the completed checklist in the Multimedia Appendix 3). To further enhance the trustworthiness of results, the data from observations and interviews were triangulated with additional data gathered from hospital documentation such as descriptions of job profiles and professional competencies, organizational charts, and technology user manuals. A total of 6-member checking workshops [45] were conducted (one in each unit), and 2 work unit heads validated the accuracy and adequacy of the analysis concerning the correctness of medical knowledge, contextual fit of interpretations, and adequate use of terminology [46].

Mixed Methods Study to Identify Optimal Levels of Human-AI Teaming Across ICU Tasks

To assess AI-induced work system changes, we used COMPASS to evaluate each of the 5 job characteristics based on the optimal levels of human-AI teaming for each ICU task. The optimal levels of human-AI teaming for each ICU task were identified in a mixed methods study [39] and summarized in Figure 2 [39].

Figure 2. Summary of results from Bienefeld et al [39], a mixed-method study identifying optimal levels of human-AI teaming in a sample of n=19 data science experts and n=61 nurses and physicians. AI: artificial intelligence.

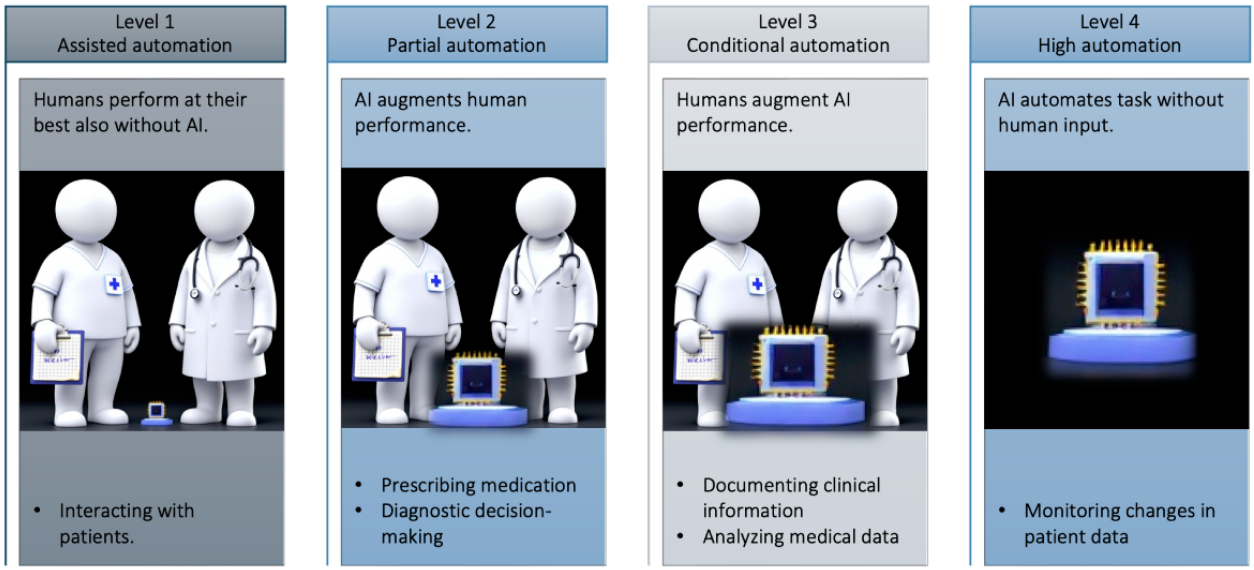

XSL•FO
RenderX

Results

Overview

In the following, we report the results of our analysis of AI-induced changes to key job characteristics for motivating, resilient, and health-promoting work in the example of ICU nurses and physicians. Table 1 summarizes the 5 job characteristics together with definitions and example quotes from interviews.

Table 1. Five job characteristics, their definition, and example quotes from structured interviews.

Job characteristics	Definition	Example quotes from interviews (n=12 ICU[a] head physicians and nurses)
Autonomy in decision-making	The ability to make work-related decisions based on one's own level of professional expertise rather than decisions by supervisors or SOPs[b]	"Of course, we need to follow lots of standards and guidelines here [in the ICU], but it doesn't feel that way [restrictive] because every patient is different, and I have the possibility to adapt my actions to what a specific patient needs. I always need to anticipate, observe, and decide what's next, not one patient reacts in the same way. That makes it interesting. I see the guidelines as some sort of crash barriers, they are there to protect you. But they cannot be too narrow, you need to improvise sometimes, or else it wouldn't work. Like when things go wrong with technology, you need to solve it quickly on your own because your patient's life depends on it. All this makes our jobs challenging but interesting." [Head nurse, #7]
Skill variety and competence development	Opportunities for using a range of skills, applying one's expertise and further developing competencies based on the qualification demands and the level of qualification	"Sometimes it can be tricky for us [nurses] when we are at the patient's bedside and we clearly see that we need to do something [eg, administer some medication], but technically we can't because it's their [physicians'] responsibility. Rather than having to wait, it would be great if we could check with an AI or something right there in the moment and have the competence to act." [Head of nurse education, #12]
Flexibility regarding time, place, and type of work	Influence over working conditions regarding the type and place of work and temporal flexibility regarding when the work is done	"[The lack of flexibility] in scheduling is what makes it difficult for many of us [nurses]. Especially when you have kids, [working in shifts] is just not compatible. We are quite a good team and organize ourselves via WhatsApp to swap around etc but that's not a great solution. Or if we could at least have lunch in the canteen [away from the ICU], to air out our heads or get some air, that would already help. Sometimes, it feels like you are tethered to your patient's bed all day long." [Head nurse, #8].
Problem-solving opportunities	The degree of cognitive complexity required for problem-solving tasks and the possibility to perform problem-solving tasks	"Doing the mentally challenging detective work [to diagnose a patient] that's what makes this profession [of an ICU physician] so rewarding. When you are able to make a very sick patient well again because you figure out how all the pieces of the puzzle fit together. I think that's an important part [of being a physician] for most of us and I sure wouldn't want to miss it." [Head physician, #1]
Task identity and variety	The ability to perform complete tasks or processes and to perform multiple tasks within the work system	"There is never a dull moment here [in the ICU] for sure. So many different things to do, maybe sometimes too many [laughs]. To me, and I think many of my colleagues is to see when a patient comes to us, very sick and we think he [or she] might not survive. And after several weeks here [in the ICU], he [or she] can go home to his [or her] family. Then we know, we [the team of physicians and nurses] did everything we could to make this happen and that's when I know why I like this job." [Head physician, #3].

[a]ICU: intensive care unit.

[b]SOP: standard operating procedures.

Autonomy in Decision-Making

The generally high levels of autonomy in decision-making and the associated ability to solve problems locally at their source were identified as major strengths in the current work system without AI. When implementing AI in safety-critical systems like the ICU, it is important to maintain high levels of human autonomy to enable agility and resilience in terms of quick reactions to fast-changing patient states, varying patient volumes, technological disturbances, and fluctuating personnel resources. If AI were to drastically increase the levels of AI-based standardization and regulation (eg, to enable robust AI performance), work processes would become more rigid, people's decision-making autonomy would be undermined and the equilibrium between standardization and flexible action by human operators could eventually tip out of balance. In addition, people's sense of autonomy in decision-making is essential for intrinsic motivation and job satisfaction. Physicians generally have high levels of autonomy in performing tasks, though within the typical decision-making hierarchy (eg, attending physicians must sign off decisions suggested by resident physicians). With the proposed level 2 (partial automation) autonomy in decision-making can be maintained for diagnostic decision-making and prescribing medication or treatment tasks but could potentially be threatened for the task of analyzing medical data (level 3). Without AI, nurses' autonomy in decision-making is currently limited with a high dependency on decisions made by physicians for any task that is outside of their decision-making scope (eg, physicians must sign off before medication can be administered). The use of AI could eventually increase autonomy in decision-making for nurses, for example, by using AI to perform additional tasks such as analyzing medical data (level 3) or delivering certain types of treatment (level 2). This increase, however, would require significant changes to nurses' education, their roles, and responsibilities, as will be discussed in more detail below.

XSL•FO
RenderX

Skill Variety and Competence Development

A second consideration for improved work design with AI is the currently low level of cross-skilling between nurses and physicians. This hampers interprofessional collaboration and problem-solving abilities as well as their competence development. Even though without AI, nurses and physicians are regularly applying their unique professional skills and can expand their knowledge through academic or on-the-job learning, these learning opportunities are mostly offered separately within each professional group. With the assistance of increasingly reliable and robust AI, for example, to analyze medical data and document clinical information (level 3), especially nurses could broaden their skill set to increase polyvalence across professional groups. AI augmentation of these tasks would thus potentially enable more cohesive and synergetic interprofessional collaboration practices. At the same time, with AI facilitating administrative tasks, physicians could spend more time interacting with patients, which would further strengthen interpersonal skills.

Flexibility Regarding Time, and Place of Work

The work system analysis revealed that without AI, nurses and physicians have very limited control over the type, time, or place of work. This negatively impacts their sense of personal agency at work. Low flexibility is inevitable due to shift work requiring physical presence 24/7. However, low flexibility is a major barrier to intrinsic motivation and job satisfaction because it restricts people's sense of autonomy and mobility, as well as their ability to integrate work with other life domains [47]. Flexibility could be increased, for instance, if AI would automate the task of monitoring patient data (level 4), enabling nurses to temporarily leave the patient's bedside without missing critical events. Moreover, with highly reliable AI to provide early warnings of upcoming complications, clinical staff could intervene long before patients enter a critical state [48,49]. This would increase their ability to plan and coordinate tasks within and across professional boundaries. In addition, if AI could assist in the documentation of clinical data and the analysis of medical data (level 3), not only would there be significant efficiency gains, but also the flexibility regarding when and where to accomplish these tasks would be increased.

Problem-Solving Opportunities

The analysis of work in the ICU without AI revealed high levels of problem-solving opportunities for physicians and medium levels for nurses. Typical activities performed by physicians such as overseeing, planning, and deciding about patients' diagnoses and treatments, require high levels of cognitive engagement. With AI-based task augmentation in diagnostic decision-making and prescribing medication and treatment (level 2), problem-solving opportunities for physicians are likely to remain high because the ability to validate or refute the suggestions made by an AI requires advanced expert knowledge and problem-solving skills. For nurses, a significant amount of attention and cognitive resources are needed for monitoring patients and interacting with them. With AI automation of monitoring tasks (level 4), nurses would likely require fewer cognitive resources in this area, which would free up cognitive capacity for other problem-solving tasks. For example, experienced nurses who have developed implicit knowledge or intuitive hunches about upcoming patient complications but are uncertain whether to communicate this information across interprofessional hierarchies could use AI to analyze medical data (level 3) in order to validate (or refute) these hunches, which would ultimately improve knowledge sharing and problem-solving in interprofessional care teams.

Task Identity and Variety

The completeness of tasks (eg, doing a work task or process from start to finish), is the basis for task identity; which enables local control of variances and disturbances and allows workers to anticipate and swiftly react to upcoming problems. Without AI, task identity and variety are rated as medium for both nurses and physicians. Physicians concentrate on patient diagnosis, goal setting, and planning behaviors, while nurses focus more on preparation, execution, and controlling behaviors, as well as patient care activities. The suggested levels of AI automation for three core tasks (monitoring medical data, documenting medical information, and analyzing medical data), could lead to a significant decrease in task identity and variety. Furthermore, if nurses and physicians were to take on only supervisory control or trouble-shooting roles when collaborating with AI, their ability to solve problems locally and directly would be heavily diminished. However, whether such changes negatively impact job satisfaction and well-being depends on how important professionals perceive each task as part of their identity. For instance, (partly) automating the labor-intensive and tedious task of documenting medical information might be a welcome relief for physicians, whereas an AI-based analysis of medical data might impair their capacity to anticipate and adequately react to patient complications [50,51].

Figure 3 shows the comparative summary of results, sorted along the 5 job characteristics and optimal levels of human-AI teaming [39]. Level 1 automation is not included because AI developers and clinicians saw the use of AI in direct patient interactions as undesirable due to social and ethical concerns.

XSL•FO
RenderX

Figure 3. Summary of AI-induced changes to key job characteristics, based on a comparative analysis of ICU nurses' and physicians' work with versus without AI combining observational data (approximately 559 ICU nurses and physicians) and data from structured interviews (n=12 ICU head nurses and physicians). Arrows pointing to the right signify an improvement; arrows to the left a deterioration. The equal sign signifies no change. Nurse-specific changes are displayed in orange; physician-specific changes are in blue. The category "Nurses" includes only registered nurses. The category "Physicians" combines resident and attending physicians. AI: artificial intelligence; COMPASS: Complementary Analysis of Sociotechnical Systems; ICU: intensive care unit.

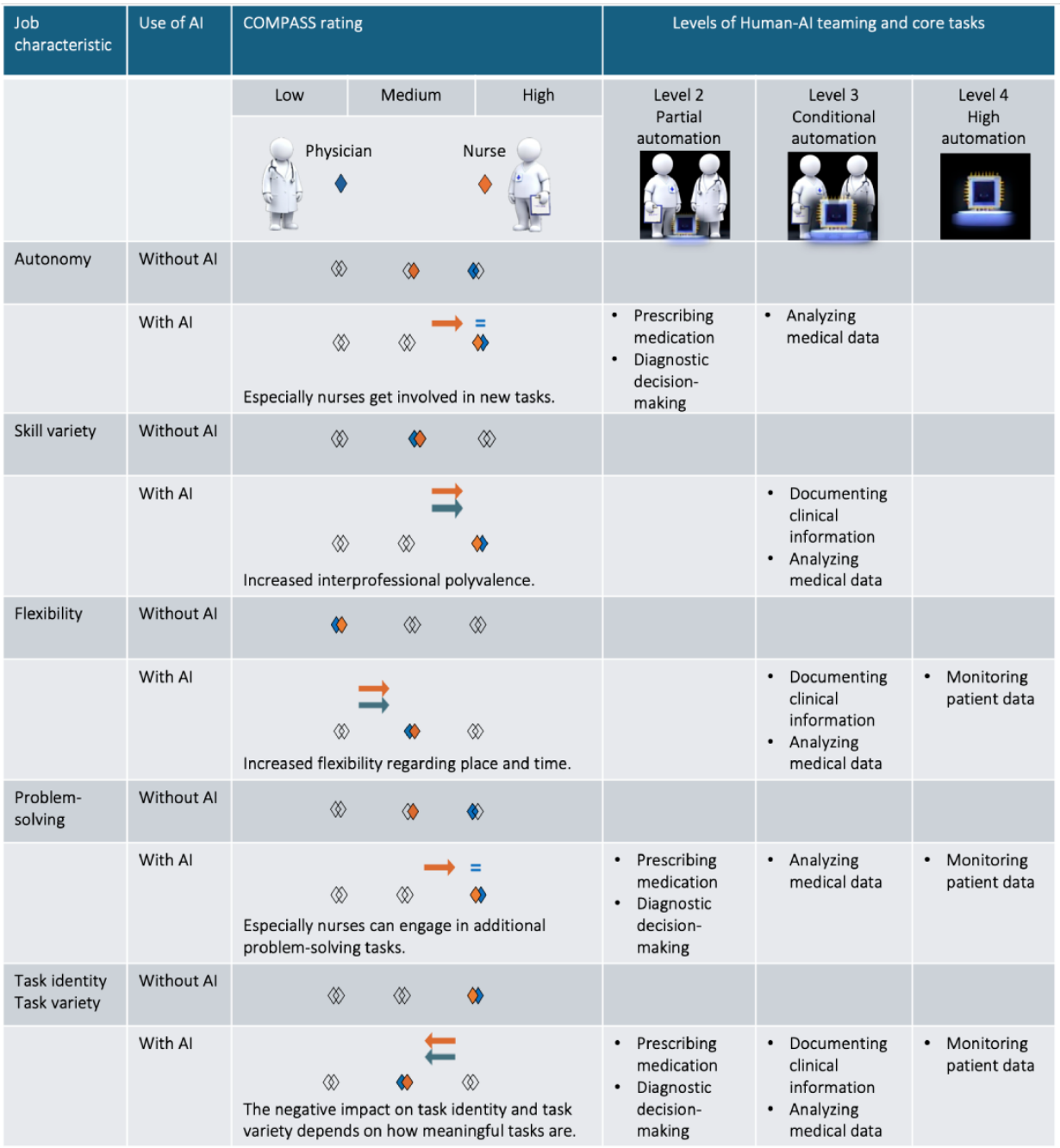

Discussion

Principal Findings

This study has explored how the use of AI can mitigate the increasing rates of stress and dissatisfaction among overworked ICU nurses and physicians, thus improving their work conditions and potentially contributing to reducing the global scarcity of skilled labor in this domain. We applied a sociotechnical system perspective in the analysis of current and future work scenarios to identify how AI impacts key job characteristics for motivating, resilient, and health-promoting work. As discussed in detail below, our findings highlight four areas where AI can enhance work conditions for at least one professional group and one area where the application of AI should be cautiously approached.

First, autonomy in decision-making is a fundamental aspect of work design due to its impact on multiple outcomes such as stress, motivation, learning, and performance. From a stress perspective, autonomy enables individuals to exert control over environmental demands, thereby reducing stress [52]. Moreover,

XSL•FO
RenderX

extensive research suggests that job autonomy contributes to a sense of meaning and purpose in the workplace [53]. This, in turn, leads to lower turnover rates, decreased absenteeism, and increased creativity, proactivity, and, ultimately, performance [54]. Autonomy leads to improved problem-solving because decisions can be made locally, that is, problems are solved directly at the source where they arise [26,55]. A decrease in autonomy in decision-making would not only lead to the loss of these beneficial effects but likely result in limited AI acceptance and adequate use altogether. A recent study demonstrated that when radiologists felt their autonomy in decision-making was reduced by AI (because they could not understand why the system made certain recommendations), they refused to use it or found ways to override the system. Other AI systems that provided higher levels of transparency and allowed radiologists to feel "in charge" had higher acceptance rates and truly augmented their work [56]. Similar results were found in a study on an AI-based system for sepsis prediction [57] and in a study where data scientists and clinicians co-designed an algorithm to predict delayed cerebral ischemia [18]. In line with this literature and work design theory [58], maintaining human autonomy in decision-making is crucial and can be achieved by designing AI systems that provide high levels of transparency, predictability, and controllability [39].

Second, AI can be used to boost skill variety and competence development, two important job characteristics for perceiving work as stimulating, experiencing mastery of one's work, and enabling lifelong learning [59]. Especially nurses would benefit from AI to develop new competencies and gain additional responsibilities. AI-enabled upskilling would not only be an important source of personal growth and motivation for nurses but could create a much-needed interprofessional skill set for managing clinical staff shortages. Furthermore, training nurses and physicians on how to best collaborate with AI—and each other—could help improve interprofessional collaboration overall [60]. On the other hand, failure to consider skill variety and competence development can lead to the deskilling of experienced professionals, that is, the loss of relevant knowledge and skills over time [61]. This would be problematic not only for experienced professionals but especially for younger colleagues who still need to develop a broad set of skills and expertise to be able to evaluate AI-based recommendations and intervene in case of system malfunctions.

Third, providing people with more flexibility regarding time, type, and place of work stimulates personal agency, which is at the core of intrinsic motivation [62]. AI could be used to increase flexibility by (partly) automating the tasks of monitoring patient data and documenting clinical information. By liberating nurses and physicians from such time-consuming monitoring and administrative tasks, they could engage in more meaningful work such as interacting with and more directly caring for their patients. Moving away from the computer and back to the patient's bedside is a wish many health care professionals entertain. Strengthening the uniquely human qualities of empathy, compassion, and interpersonal connection has become increasingly important in a technology-driven world [63]. AI could also increase flexibility regarding the place of work (eg, by predictive analytics and remote patient monitoring [48,64]) and time of work (eg, by designing smart rosters that consider individual needs and preferences). Increased flexibility regarding the time and place of work would ultimately facilitate better work-life integration to accommodate family and other life domains—a known factor associated with high turnover rates in health care professionals [65,66].

Fourth, problem-solving opportunities help create stimulating and meaningful work [67,68]. According to our assessment, problem-solving opportunities are likely to stay high for physicians because they remain in control of tasks such as diagnostic decision-making. For nurses, problem-solving opportunities were associated mainly with monitoring activities. With the automation of monitoring tasks by AI, nurses would have fewer problem-solving needs in this domain, but AI could create new opportunities for problem-solving in other areas. For instance, by developing new skills and using AI to analyze medical data, nurses could engage more deeply in cognitive problem-solving activities, which would deepen their understanding of a patient's medical condition and likely improve interprofessional communication and collaboration. However, given the current hierarchical structure of ICU teams, such changes would require significant adaptations in nurses' educational curricula, roles, and career paths, including new regulations regarding legal accountability and reimbursement strategies [69,70].

Finally, task identity and variety are important factors for intrinsic motivation and worker engagement, and were rated as high in work conditions without AI. To maintain high task identity and variety, it is important to balance different forms of human-AI teaming as suggested by Bienefeld et al [39]. Assigning humans with leftover tasks (ie, tasks that cannot be executed by technology) can otherwise result in decreased motivation and make automated systems brittle and prone to errors [71,72]. Against the common belief that automating complex tasks leaves the easy aspects for humans, automation can make the role of humans more challenging, thus hindering their ability to intervene effectively in the event of system failures. This phenomenon, known as the "out of the loop" performance problem [73], results in human operators becoming progressively less capable of detecting system errors and reacting appropriately when automation malfunctions. This problem has been observed, for example, in the aviation industry, where pilots could no longer foresee or understand the automated actions of the flight management system, prompting inappropriate reactions that led to many tragic accidents [74-76]. Moreover, even as the potential of AI tools in health care increases and more and more tasks can potentially be automated or augmented by AI [77], it is important to allow ICU nurses and physicians to engage in a range of activities to maintain adequate levels of task variety.

Considering today's grand health care challenges such as the high rates of attrition and shortage of skilled labor, we hope that our predictive analysis of how AI will impact the quality of work of nurses and physicians inspires all involved stakeholder groups (eg AI developers, hospital management, and policy makers) to think carefully about whether a given technology really improves the work task or process in question. Our results predict higher worker motivation, satisfaction, and

XSL•FO
RenderX

well-being and fewer AI implementation hurdles if AI tools are developed and implemented in line with the principles for good work and sociotechnical system design [32,58,78]. Future researchers should continue to analyze the short and long-term effects of AI implementation in real-world health care settings, also considering additional factors such as resistance to change, team dynamics, and the role of leadership [79]. Implementation research [80] practices, like technological co-design [18] and stakeholder involvement [81], could be used to facilitate such endeavors.

Limitations and Future Research Directions

As with any study, there are various limitations to consider when interpreting the results. While the COMPASS methodology is a theory-based, empirically validated, and highly structured tool in sociotechnical systems and work design research, we are mindful of the potential biases such as the Hawthorne effect [82] such a tool may introduce. In this study, we have undertaken due diligence to minimize such biases by taking the following measures: (1) observations were conducted as unobtrusively as possible with the observer wearing nurse gowns and standing in the background about five feet away, (2) evenly distributing observations and interviews across the six ICUs to ensure comparability, (3) standardized observer and interview guidelines (Multimedia Appendices 1 and 2), (4) extensive observer training (>15 years of experience using the COMPASS tool), and (5) centering data collection within one experienced qualitative researcher with domain expertise in work and organizational psychology to avoid interrater bias [83]. Furthermore, our findings are likely most applicable to settings with similar boundary conditions regarding workload, tasks, and team dynamics (eg, emergency or acute care teams). Nonetheless, the qualitative nature of our work has allowed us to capture the situated work practices of ICU staff, which we hope provides valuable insights for future research and real-world clinical practice.

Even though we included a relatively large sample of approximately 570 ICU nurses and physicians in this study, we may not have captured the full spectrum of experiences within interprofessional ICU teams.

Future research should examine how additional roles, particularly those with lower levels of experience (eg, nurse students and junior residents), perceive and interact with AI. The use of AI is often seen as particularly beneficial for less experienced professionals [84] but at the same time, is a cause of concern: Less experienced professionals may rely too much on the assistance of AI to perform their work, thus cannot build important skills, and lack the necessary expertise to verify the adequacy of AI outputs. Studies to examine the nuanced effects across different levels of expertise are needed to avoid these dangers and to build AI systems that support all users effectively.

Finally, despite our focus on improving job resources through AI-based technologies, ICU nurses and physicians face multiple job demands such as time pressure, very complex patient cases, and emotionally demanding situations that may continue to exert a toll on them despite AI support. As previously mentioned, a balance between job resources and job demands [33,34] can render these demands tolerable. People-focused strategies such as mentorship or emotional support programs might also help to mitigate the many challenges [85]. In addition, exploring job crafting initiatives [86,87] could enrich our understanding as a complement to the top-down approach used in this study. For instance, job crafting activities, whereby workers themselves initiate changes to their work conditions, have proven highly effective in bypassing potential management or union resistance [58].

Finally, this paper could not delve into broader issues like economic pressures in health care. While worker motivation is multifaceted, fair pay as a fundamental need is, of course, also pivotal [88].

Conclusions

The integration of AI into critical care work presents multifaceted opportunities to enhance job satisfaction and well-being and ease the burden on staff. Specifically, our analysis demonstrates that AI can augment decision-making autonomy, encourage skill diversity and competence development, and increase flexibility—each a critical determinant of job satisfaction, well-being, and professional retention. Moreover, AI's potential to redistribute routine and administrative tasks can allow health care professionals to allocate more time to patient care and collaborative problem-solving, thus amplifying their effectiveness and job fulfillment.

It is imperative to say that AI's contribution extends beyond mere numerical staffing solutions. By strategically leveraging AI's capabilities, health care systems can enhance the quality of work for ICU professionals, fostering environments where job resources are amplified and job demands are diminished. This, in turn, can lead to heightened professional engagement and a corresponding reduction in the attrition rates that exacerbate staffing shortages.

However, it is critical to acknowledge that the deployment of AI must be thoughtfully managed to prevent the deskilling of professionals and to maintain the integrity of task identity and variety, which are essential criteria for continuously safe performance. Thus, while AI holds promise for addressing aspects of the health care staffing crisis, these technologies must be used as part of a broader strategy that includes sociotechnical system design and human-centric principles.

To this end, we conclude that AI—when incorporated in alignment with good work and sociotechnical system design principles—has the potential to significantly alleviate health care staffing shortages by making the work of nurses and physicians more attractive. The study results indicate that by tailoring AI to redistribute work tasks based on the synergistic interplay between human and AI capabilities, intrinsic motivation, resilience, and worker well-being can be elevated, and the retention of professional expertise can be ensured. Future research into the sociotechnical integration of AI in health care should look to include a wider array of team-based settings beyond critical care to encompass a broader spectrum of workers and explore the long-term impacts on the health care workforce and the patients they serve.

XSL•FO
RenderX

Acknowledgments

Our immense gratitude goes to all the clinicians who devoted their time to participate in this study. This work was supported by funding from the Swiss National Science Foundation (grant 407740_187331). The content and conclusions presented in this paper are those of the authors and should neither be construed as the official position or policy of nor should any endorsements be inferred by the Swiss National Science Foundation. We used Grammarly for Microsoft Office (free version) to suggest corrections for grammatical and spelling errors, as well as improve the flow of written sentences [89]. The generative AI tool ChatGPT by OpenAI [90] (free version) was used to suggest the title of this manuscript based on the abstract and in adherence with JMIR author guidelines. The original ChatGPT transcripts are made available as Multimedia Appendix 4.

Data Availability

Datasets generated during or analyzed during this study are available in the ETH Zurich Research Collection repository and accessible by the digital object identifier (DOI; 10.3929/ethz-b-000693893). The DOI issued to datasets in the repository can be included as part of a data citation in publications, allowing the datasets underpinning a publication to be identified and accessed. The full transcripts of the interviews are not publicly available to minimize the risk of participant reidentification. Summaries of the interview contents and related metadata that support these findings are available from the corresponding author upon reasonable request.

Authors' Contributions

NB and GG contributed to the initial study design. EK provided her medical expertise to set up the study and verify findings from the work system analysis. NB conducted all parts of the study and implemented the study protocol. NB and GG contributed to the review and analysis of the results. All authors contributed to the preparation of the manuscript. All authors approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COMPASS Observation Guidelines and Criteria.
[PDF File (Adobe PDF File), 182 KB - jmir_v27i1e50852_app1.pdf]

Multimedia Appendix 2
COMPASS Interview Guidlines.
[PDF File (Adobe PDF File), 196 KB - jmir_v27i1e50852_app2.pdf]

Multimedia Appendix 3
Completed Checklist for Qualitative Research.
[PDF File (Adobe PDF File), 65 KB - jmir_v27i1e50852_app3.pdf]

Multimedia Appendix 4
Declaration of the Use of Generative AI.
[PDF File (Adobe PDF File), 591 KB - jmir_v27i1e50852_app4.pdf]

References

1. Sousa A, Scheffler RM, Nyoni J, Boerma T. A comprehensive health labour market framework for universal health coverage. Bull World Health Organ 2013;91(11):892-894 [FREE Full text] [doi: 10.2471/BLT.13.118927] [Medline: 24347720]
2. WHO guideline on self-care interventions for health and well-being, 2022 revision.: World Health Organization; 2022. URL: https://clacaidigital.info/handle/123456789/1654 [accessed 2023-07-06]
3. Boniol M, Kunjumen T, Nair TS, Siyam A, Campbell J, Diallo K. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage? BMJ Global Health 2022;7(6)::e009316 [FREE Full text] [doi: 10.1136/bmjgh-2022-009316] [Medline: 35760437]
4. Drennan VM, Ross F. Global nurse shortages—the facts, the impact and action for change. Br Med Bull 2019;130(1):25-37. [doi: 10.1093/bmb/ldz014] [Medline: 31086957]
5. Zhang X, Lin D, Pforsich H, Lin VW. Physician workforce in the United States of America: forecasting nationwide shortages. Hum Resour Health 2020;18(1):8 [FREE Full text] [doi: 10.1186/s12960-020-0448-3] [Medline: 32029001]

XSL•FO
RenderX

6. Amanullah S, Shankar RR. The impact of COVID-19 on physician burnout globally: a review. Healthcare (Basel) 2020;8(4):421 [FREE Full text] [doi: 10.3390/healthcare8040421] [Medline: 33105757]
7. Shah MK, Gandrakota N, Cimiotti JP, Ghose N, Moore M, Ali MK. Prevalence of and Factors Associated With Nurse Burnout in the US. JAMA Netw Open 2021 Feb 01;4(2):e2036469 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.36469] [Medline: 33538823]
8. Llop-Gironés A, Vračar A, Llop-Gironés G, Benach J, Angeli-Silva L, Jaimez L, et al. Employment and working conditions of nurses: where and how health inequalities have increased during the COVID-19 pandemic? Hum Resour Health 2021;19(1):112 [FREE Full text] [doi: 10.1186/s12960-021-00651-7] [Medline: 34530844]
9. Søvold LE, Naslund JA, Kousoulis AA, Saxena S, Qoronfleh MW, Grobler C, et al. Prioritizing the mental health and well-being of healthcare workers: an urgent global public health priority. Front Public Health 2021;9:679397 [FREE Full text] [doi: 10.3389/fpubh.2021.679397] [Medline: 34026720]
10. Bees J. The worldwide nurse staffing crisis intensifies. NEJM Catal 2023;4(3). [doi: 10.1056/cat.23.0049]
11. LeClaire M, Poplau S, Linzer M, Brown R, Sinsky C. Compromised integrity, burnout, and intent to leave the job in critical care nurses and physicians. Crit Care Explor 2022;4(2):e0629 [FREE Full text] [doi: 10.1097/CCE.0000000000000629] [Medline: 35156049]
12. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull 2021;139(1):4-15. [doi: 10.1093/bmb/ldab016] [Medline: 34405854]
13. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. Int J Environ Res Public Health 2021;18(1):271 [FREE Full text] [doi: 10.3390/ijerph18010271] [Medline: 33401373]
14. Chenais G, Lagarde E, Gil-Jardiné C. Artificial intelligence in emergency medicine: viewpoint of current applications and foreseeable opportunities and challenges. J Med Internet Res 2023;25:e40031 [FREE Full text] [doi: 10.2196/40031] [Medline: 36972306]
15. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
16. Ossa LA, Starke G, Lorenzini G, Vogt JE, Shaw DM, Elger BS. Re-focusing explainability in medicine. Digital Health 2022;8:20552076221074488 [FREE Full text] [doi: 10.1177/20552076221074488] [Medline: 35173981]
17. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
18. Bienefeld N, Boss JM, Lüthy R, Brodbeck D, Azzati J, Blaser M, et al. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. NPJ Digital Med 2023;6(1):94 [FREE Full text] [doi: 10.1038/s41746-023-00837-4] [Medline: 37217779]
19. Aarts J. A sociotechnical perspective of health information technology. Int J Med Inform 2013;82(12):1133-1135. [doi: 10.1016/j.ijmedinf.2013.10.007] [Medline: 24216291]
20. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
21. Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The influence of electronic health record use on physician burnout: cross-sectional survey. J Med Internet Res 2020;22(7):e19274 [FREE Full text] [doi: 10.2196/19274] [Medline: 32673234]
22. Blijleven V, Koelemeijer K, Jaspers M. SEWA: a framework for sociotechnical analysis of electronic health record system workarounds. Int J Med Inform 2019;125:71-78. [doi: 10.1016/j.ijmedinf.2019.02.012] [Medline: 30914183]
23. Gawande AA. Why doctors hate their computers. The New Yorker. URL: https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers [accessed 2024-11-12]
24. Waterson PE, Gray MTO, Clegg CW. A sociotechnical method for designing work systems. Hum Factors 2002;44(3):376-391. [doi: 10.1518/0018720024497628] [Medline: 12502156]
25. Shields C, Cunningham SG, Wake DJ, Fioratou E, Brodie D, Philip S, et al. User-centered design of a novel risk prediction behavior change tool augmented with an artificial intelligence engine (MyDiabetesIQ): a sociotechnical systems approach. JMIR Hum Factors 2022;9(1):e29973 [FREE Full text] [doi: 10.2196/29973] [Medline: 35133280]
26. Grote G. Management of Uncertainty: Theory and Application in the Design of Systems and Organizations. Germany: Springer Science & Business Media; 2009.
27. Clegg CW. Sociotechnical principles for system design. Appl Ergon 2000;31(5):463-477. [doi: 10.1016/s0003-6870(00)00009-0] [Medline: 11059460]
28. Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders SM. Methods for human-centered eHealth development: narrative scoping review. J Med Internet Res 2022;24(1):e31858 [FREE Full text] [doi: 10.2196/31858] [Medline: 35084359]
29. Grote G, Ryser C, Windischer A, Weik S. KOMPASS: a method for complementary function allocation in automated work systems. Int J Hum-Comput Stud 2000;52(2):267-287. [doi: 10.1006/ijhc.1999.0289]

XSL•FO
RenderX

30. Waefler T, Grote G, Windischer A, Ryser C. KOMPASS: A method for complementary system design. In: Handbook of Cognitive Task Design, in Human Factors and Ergonomics. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2003:477-502.
31. Boos D, Grote G, Guenter H. A toolbox for managing organisational issues in the early stage of the development of a ubiquitous computing application. Pers Ubiquit Comput 2013;17(6):1261-1279. [doi: 10.1007/s00779-012-0634-y]
32. Parker SK, Van den Broeck A, Holman D. Work design influences: a synthesis of multilevel factors that affect the design of jobs. ANNALS 2017;11(1):267-308. [doi: 10.5465/annals.2014.0054]
33. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001;86(3):499-512. [doi: 10.1037/0021-9010.86.3.499]
34. Schaufeli WB, Bakker AB. Job demands, job resources, and their relationship with burnout and engagement: a multi‐sample study. J Organ Behavior 2004;25(3):293-315. [doi: 10.1002/job.248]
35. Heesakkers H, Zegers M, van Mol MMC, van den Boogaard M. The impact of the first COVID-19 surge on the mental well-being of ICU nurses: a nationwide survey study. Intensive Crit Care Nurs 2021;65:103034 [FREE Full text] [doi: 10.1016/j.iccn.2021.103034] [Medline: 33863609]
36. Meynaar IA, Ottens T, Zegers M, van Mol MMC, van der Horst ICC. Burnout, resilience and work engagement among Dutch intensivists in the aftermath of the COVID-19 crisis: a nationwide survey. J Crit Care 2021;62:1-5 [FREE Full text] [doi: 10.1016/j.jcrc.2020.11.010] [Medline: 33232847]
37. Moradi Y, Baghaei R, Hosseingholipour K, Mollazadeh F. Challenges experienced by ICU nurses throughout the provision of care for COVID-19 patients: a qualitative study. J Nurs Manag 2021;29(5):1159-1168 [FREE Full text] [doi: 10.1111/jonm.13254] [Medline: 33480145]
38. Hollnagel E, Woods DD. Joint Cognitive Systems: Foundation of Cognitive Systems Engineering. New York: Taylor and Francis; 2005.
39. Bienefeld N, Keller E, Grote G. Human-AI teaming in critical care: a comparative analysis of data scientists' and clinicians' perspectives on AI augmentation and automation. J Med Internet Res 2024;26:e50130 [FREE Full text] [doi: 10.2196/50130] [Medline: 39038285]
40. Grote G. Design approaches for the complementary interaction of people and technology in Industry 4.0. In: Digitization of Industrial Work. Germany: Nomos Verlagsgesellschaft mbH & Co; 2018:215-232.
41. Glaser BG, Strauss AL. Discovery of Grounded Theory: Strategies for Qualitative Research. UK: Routledge; 2017.
42. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research. Organ Res Methods 2012;16(1):15-31. [doi: 10.1177/1094428112452151]
43. Kuckartz U, Rädiker S. Analyzing qualitative data with MAXQDA. In: Text, Audio, and Video. Cham: Springer International Publishing; 2020.
44. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
45. Pratt MG, Sonenshein S, Feldman MS. Moving beyond templates: a bricolage approach to conducting trustworthy qualitative research. Organ Res Methods 2020;25(2):1094428120927466. [doi: 10.1177/1094428120927466]
46. Seale C. Quality in qualitative research. Qual Inq 1999;5(4):465-478. [doi: 10.1177/107780049900500402]
47. Parker SK, Grote G. Automation, algorithms, and beyond: why work design matters more than ever in a digital world. Appl Psychol 2020;71(4):1171-1204. [doi: 10.1111/apps.12241]
48. Muroi C, Meier S, De Luca V, Mack DJ, Strässle C, Schwab P, et al. Automated false alarm reduction in a real-life intensive care setting using motion detection. Neurocrit Care 2020;32(2):419-426. [doi: 10.1007/s12028-019-00711-w] [Medline: 31290067]
49. Schwab P, Keller E, Muroi C, Mack DJ, Strässle C, Karlen W. Not to cry wolf: distantly supervised multitask learning in critical care. 2018 Presented at: Proceedings of the 35th International Conference on Machine Learning, PMLR; July 10-15, 2018; Stockholm, Sweden p. 4518-4527 URL: https://proceedings.mlr.press/v80/schwab18a.html
50. McKoane A, Sherman DK. Diagnostic uncertainty in patients, parents, and physicians: a compensatory control theory perspective. Health Psychol Rev 2023;17(3):439-455. [doi: 10.1080/17437199.2022.2086899] [Medline: 35672909]
51. Landau MJ, Kay AC, Whitson JA. Compensatory control and the appeal of a structured world. Psychol Bull 2015;141(3):694-722. [doi: 10.1037/a0038703] [Medline: 25688696]
52. Karasek RA. Job demands, job decision latitude, and mental strain: implications for job redesign. Adm Sci Q 1979;24(2):285-308. [doi: 10.2307/2392498]
53. Hackman JR, Oldham GR. Motivation through the design of work: test of a theory. Organ Behav Hum Perform 1976;16(2):250-279. [doi: 10.1016/0030-5073(76)90016-7]
54. Parker SK, Morgeson FP, Johns G. One hundred years of work design research: looking back and looking forward. J Appl Psychol 2017;102(3):403-420. [doi: 10.1037/apl0000106] [Medline: 28182465]
55. Wall TD, Cordery JL, Clegg CW. Empowerment, performance, and operational uncertainty: a theoretical integration. Appl Psychol 2002;51(1):146-169. [doi: 10.1111/1464-0597.00083]

XSL•FO
RenderX

56. Lebovitz S, Lifshitz-Assaf H, Levina N. To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis. Organ Sci 2022;33(1):126-148. [doi: 10.1287/orsc.2021.1549]
57. Henry KE, Kornfield R, Sridharan A, Linton RC, Groh C, Wang T, et al. Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. NPJ Digital Med 2022;5(1):97 [FREE Full text] [doi: 10.1038/s41746-022-00597-7] [Medline: 35864312]
58. Parker SK, Grote G. Automation, algorithms, and beyond: why work design matters more than ever in a digital world. Appl Psychol 2020;71(4):1171-1204. [doi: 10.1111/apps.12241]
59. Parker SK, Ward MK, Fisher GG. Can high-quality jobs help workers learn new tricks? A multidisciplinary review of work design for cognition. ANNALS 2021;15(2):406-454. [doi: 10.5465/annals.2019.0057]
60. Bienefeld N, Kolbe M, Camen G, Huser D, Buehler PK. Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness. Front Psychol 2023 Aug 04;14:1-10 [FREE Full text] [doi: 10.3389/fpsyg.2023.1208019] [Medline: 37599773]
61. Salas E, Maurino D, Curtis M. Chapter 1—human factors in aviation: an overview. In: Human Factors in Aviation (Second Edition). San Diego: Academic Press; 2010:3-19.
62. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008;49(3):182-185. [doi: 10.1037/a0012801]
63. Zulman DM, Verghese A. Virtual care, telemedicine visits, and real connection in the era of COVID-19: unforeseen opportunity in the face of adversity. JAMA 2021;325(5):437-438. [doi: 10.1001/jama.2020.27304] [Medline: 33528520]
64. Jamil F, Ahmad S, Iqbal N, Kim DH. Towards a remote monitoring of patient vital signs based on IoT-based blockchain integrity management platforms in smart hospitals. Sensors (Basel) 2020;20(8):2195 [FREE Full text] [doi: 10.3390/s20082195] [Medline: 32294989]
65. Shanafelt TD, West CP, Dyrbye LN, Trockel M, Tutty M, Wang H, et al. Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic. Mayo Clin Proc 2022;97(12):2248-2258 [FREE Full text] [doi: 10.1016/j.mayocp.2022.09.002] [Medline: 36229269]
66. Dyrbye LN, West P, Johnson PO, Cipriano PF, Beatty DE, Peterson C, et al. Burnout and satisfaction with work-life integration among nurses. J Occup Environ Med 2019;61(8):689-698. [doi: 10.1097/JOM.0000000000001637] [Medline: 31348422]
67. Humphrey SE, Nahrgang JD, Morgeson FP. Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. J Appl Psychol 2007;92(5):1332-1356. [doi: 10.1037/0021-9010.92.5.1332] [Medline: 17845089]
68. Lysova EI, Allan BA, Dik BJ, Duffy RD, Steger MF. Fostering meaningful work in organizations: a multi-level review and integration. J Voc Behav 2019;110:374-389. [doi: 10.1016/j.jvb.2018.07.004]
69. Briscoe F, Gardner H. Richard Susskind and Daniel Susskind: the future of the professions: how technology will transform the work of human experts. Adm Sci Q 2017;62(4):NP42-NP44. [doi: 10.1177/0001839217716083]
70. Susskind RE, Susskind D. The Future of the Professions: How Technology Will Transform the Work of Human Experts. Somerville: Oxford University Press; 2015.
71. Fairbanks RJ, Wears RL, Woods DD, Hollnagel E, Plsek P, Cook RI. Resilience and resilience engineering in health care. Jt Comm J Qual Patient Saf 2014;40(8):376-383. [doi: 10.1016/s1553-7250(14)40049-7] [Medline: 25208443]
72. Bainbridge L. Ironies of automation. In: Analysis, Design and Evaluation of Man-Machine Systems. 1983 Presented at: Proceedings of the IFAC/IFIP/IFORS/IEA Conference, Baden-Baden; September 27-29, 1982; Germany p. 129-135. [doi: 10.1016/b978-0-08-029348-6.50026-9]
73. Endsley MR, Kiris EO. The out-of-the-loop performance problem and level of control in automation. Hum Factors 1995;37(2):381-394. [doi: 10.1518/001872095779064555]
74. Ferris T, Sarter N, Wickens CD. Cockpit automation: still struggling to catch up? In: Salas E, Maurino D, editors. Human Factors in Aviation. New York: Academic Press; 2010:503.
75. Billings, CE. Billings, Aviation Automation: The Search for A Human-centered Approach. United States: CRC Press; 2018.
76. Sheridan TB. Automation. In: Boehm-Davis DA, Durso FT, Lee JD, editors. APA Handbook of Human Systems Integration. Washington: American Psychological Association; 2015:465.
77. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature 2023;616(7956):259-265. [doi: 10.1038/s41586-023-05881-4] [Medline: 37045921]
78. Parker SK. Beyond motivation: job and work design for development, health, ambidexterity, and more. Annu Rev Psychol 2014;65(1):661-691. [doi: 10.1146/annurev-psych-010213-115208] [Medline: 24016276]
79. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Chazerand P, et al. An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. In: Floridi L, editor. Ethics, Governance, and Policies in Artificial Intelligence. Cham: Springer; 2021:39.
80. Brownson RC, Colditz GA, Proctor EK, editors. Proctor, Dissemination and Implementation Research in Health: Translating Science to Practice. United Kingdom: Oxford University Press; 2023.

XSL•FO
RenderX

81. Hogg HDJ, Al-Zubaidy M, Talks J, Denniston AK, Kelly CJ, Malawana J, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. J Med Internet Res 2023;25:e39742 [FREE Full text] [doi: 10.2196/39742] [Medline: 36626192]
82. Wickström G, Bendix T. The Hawthorne effect—what did the original Hawthorne studies actually show? Scand J Work Environ Health 2000;26(4):367 [FREE Full text]
83. Pratt MG. From the editors: for the lack of a boilerplate: tips on writing up (and reviewing) qualitative research. AMJ 2009;52(5):856-862. [doi: 10.5465/amj.2009.44632557]
84. Brynjolfsson E, Li D, Raymond LR. Generative AI at work. Natl Bur Econ Res 2023:31161. [doi: 10.3386/w31161]
85. Bakker AB, de Vries JD. Job demands-resources theory and self-regulation: new explanations and remedies for job burnout. Anxiety Stress Coping 2021;34(1):1-21. [doi: 10.1080/10615806.2020.1797695] [Medline: 32856957]
86. Knight C, Tims M, Gawke J, Parker SK. When do job crafting interventions work? The moderating roles of workload, intervention intensity, and participation. J Voc Behav 2021;124:103522. [doi: 10.1016/j.jvb.2020.103522]
87. Zhang F, Parker SK. Reorienting job crafting research: a hierarchical structure of job crafting concepts and integrative review. J Organ Behav 2018;40(2):126-146. [doi: 10.1002/job.2332]
88. Rynes SL, Gerhart B, Minette KA. The importance of pay in employee motivation: discrepancies between what people say and what they do. Hum Resour Manage 2004;43(4):381-394. [doi: 10.1002/hrm.20031]
89. Grammarly for MS office. 2024. URL: https://www.grammarly.com/terms [accessed 2020-03-26]
90. OpenAI. ChatGPT. 2024. URL: https://chat.openai.com/chat [accessed 2022-11-30]

Abbreviations

AI: artificial intelligence
COMPASS: Complementary Analysis of Sociotechnical Systems
ICU: intensive care unit

Edited by A Mavragani; submitted 14.07.23; peer-reviewed by E Vashishtha, R Bidkar, E Fisher, J Vietas; comments to author 22.11.23; revised version received 07.02.24; accepted 11.10.24; published 13.01.25.

Please cite as:
Bienefeld N, Keller E, Grote G
AI Interventions to Alleviate Healthcare Shortages and Enhance Work Conditions in Critical Care: Qualitative Analysis
J Med Internet Res 2025;27:e50852
URL: https://www.jmir.org/2025/1/e50852
doi:10.2196/50852
PMID:

©Nadine Bienefeld, Emanuela Keller, Gudela Grote. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usefulness of Automatic Speech Recognition Assessment of Children With Speech Sound Disorders: Validation Study

Do Hyung Kim[1], MS; Joo Won Jeong[1], MS; Dayoung Kang[1], MS; Taekyung Ahn[2,3], BEd; Yeonjung Hong[3], PhD; Younggon Im[3], BEng; Jaewon Kim[1], MD; Min Jung Kim[4], PhD; Dae-Hyun Jang[1], MD, PhD

[1]Department of Rehabilitation Medicine, Incheon St Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Department of English Language and Literature, Korea University, Seoul, Republic of Korea

[3]MediaZen, Seongnam-si, Gyeonggi-do, Republic of Korea

[4]Department of Special Education, Dankook University, Youngin-si, Gyeonggi-do, Republic of Korea

Corresponding Author:
Dae-Hyun Jang, MD, PhD
Department of Rehabilitation Medicine
Incheon St Mary's Hospital, College of Medicine
The Catholic University of Korea
22 Banpo-daero
Seocho-gu
Seoul, 06591
Republic of Korea
Phone: 82 0322806601
Email: dhjangmd@naver.com

Abstract

Background: Speech sound disorders (SSDs) are common communication challenges in children, typically assessed by speech-language pathologists (SLPs) using standardized tools. However, traditional evaluation methods are time-intensive and prone to variability, raising concerns about reliability.

Objective: This study aimed to compare the evaluation outcomes of SLPs and an automatic speech recognition (ASR) model using two standardized SSD assessments in South Korea, evaluating the ASR model's performance.

Methods: A fine-tuned wav2vec 2.0 XLS-R model, pretrained on 436,000 hours of adult voice data spanning 128 languages, was used. The model was further trained on 93.6 minutes of children's voices with articulation errors to improve error detection. Participants included children referred to the Department of Rehabilitation Medicine at a general hospital in Incheon, South Korea, from August 19, 2022, to June 14, 2023. Two standardized assessments—the Assessment of Phonology and Articulation for Children (APAC) and the Urimal Test of Articulation and Phonology (U-TAP)—were used, with ASR transcriptions compared to SLP transcriptions.

Results: This study included 30 children aged 3-7 years who were suspected of having SSDs. The phoneme error rates for the APAC and U-TAP were 8.42% (457/5430) and 8.91% (402/4514), respectively, indicating discrepancies between the ASR model and SLP transcriptions across all phonemes. Consonant error rates were 10.58% (327/3090) and 11.86% (331/2790) for the APAC and U-TAP, respectively. On average, there were 2.60 (SD 1.54) and 3.07 (SD 1.39) discrepancies per child for correctly produced phonemes, and 7.87 (SD 3.66) and 7.57 (SD 4.85) discrepancies per child for incorrectly produced phonemes, based on the APAC and U-TAP, respectively. The correlation between SLPs and the ASR model in terms of the percentage of consonants correct was excellent, with an intraclass correlation coefficient of 0.984 (95% CI 0.953-0.994) and 0.978 (95% CI 0.941-0.990) for the APAC and UTAP, respectively. The *z* scores between SLPs and ASR showed more pronounced differences with the APAC than the U-TAP, with 8 individuals showing discrepancies in the APAC compared to 2 in the U-TAP.

Conclusions: The results demonstrate the potential of the ASR model in assessing children with SSDs. However, its performance varied based on phoneme or word characteristics, highlighting areas for refinement. Future research should include more diverse speech samples, clinical settings, and speech data to strengthen the model's refinement and ensure broader clinical applicability.

(J Med Internet Res 2025;27:e60520) doi:10.2196/60520

XSL•FO
RenderX

KEYWORDS

speech sound disorder; speech recognition software; speech articulation tests; speech-language pathology; child

Introduction

Background

Speech sound disorders (SSDs) are common communication disorders among children, wherein they experience difficulty producing speech sounds or using them correctly when compared to children of the same age [1]. In South Korea, approximately 44.1% of children undergoing speech therapy have SSDs [2], and around 9% of 6-year-old children exhibit articulation problems [3].

SSDs among children are evaluated by prompting spontaneous speech using single-word naming tasks, sentence repetition tasks, and connected speech tasks [4]. In South Korea, standardized assessment tools such as the Assessment of Phonology and Articulation for Children (APAC) and Urimal Test of Articulation and Phonology (U-TAP) have been predominantly used to assess children with SSDs in the clinical setting [5,6]. The percentage of correct consonants (PCC, %) serves as the primary criterion for diagnosing SSDs. Aside from PCC, speech error patterns must also be examined and analyzed throughout the entire transcription to guide the direction of intervention for children.

Nonetheless, traditional evaluation methods face two primary issues. First, these methods are time-consuming [4,7]. In fact, the process by which speech-language pathologists (SLPs) evaluate children with SSDs lasts approximately 2-2.5 hours, due to the need for accurate transcription and analysis of individual speech error patterns [7]. The second issue involves reliability, which has been attributed to slight differences in results depending on the tester, potentially arising from variations in the tester's skill level and the test environment [8,9].

Consensus among the SLPs is influenced by patterns of speech errors, with certain patterns being more readily recognizable than others. Studies have shown that misarticulations involving only one feature change are usually ignored, whereas those with two or more feature changes are readily recognized [9,10]. Moreover, evidence suggests that identifying misarticulations of vowels is more challenging than that of consonants [11], with combinations of both further increasing the difficulty [12]. Regarding consonants, listeners can more consistently detect voicing errors than they would changes in the place of articulation [13]. In particular, Klein et al [14] showed diversity in the ratings of /r/ production among SLPs, due to difficulties in auditory-perceptual discrimination. These studies suggest that SLPs find it more challenging to transcribe incorrectly produced phonemes than those produced correctly. Thus, streamlined and standardized evaluations are essential for achieving accurate and consistent assessments of speech sound errors in children with SSDs.

Prior Studies

Automating the process of transcribing children's speech and analyzing patterns of speech errors could enhance the efficiency and reliability of the evaluation process benefiting both patients and SLPs [8]. Investigations on automatic speech recognition (ASR) for assessing children with SSDs have used various approaches to enhance accuracy and reliability [15-18].

Traditional ASR models using the hidden Markov model (HMM) are divided into three parts: Lexicon, Acoustic, and Language models. Each part is modeled independently from each other. Suanpirintr and Thubthong [16] obtained a phoneme error rate (PER; the number of incorrectly recognized phonemes divided by the total number of phonemes) of 51.5% for dysarthria in Thailand using an HMM. Mazenan et al [17] also used an HMM for Malay SSDs, reporting low PER rates (approximately 3%) on various isolated phonemes; however, their results were limited to Malay alveolar sounds. Lee et al [18] evaluated articulation using an HMM targeting patients with dysarthria in South Korea and confirmed that their PCC results were strongly correlated with those of SLPs; however, no information was provided regarding the agreement between individual phonemes in the actual transcription results. Even though HMM has been used in many studies, developing HMM-based models requires considerable time and expertise, involving detailed data analysis and careful model validation.

Building on the strengths and limitations of HMM-based models, recent advancements have led to more advanced approaches like end-to-end models. These models simplify the process by directly linking inputs to outputs, removing the need for separate steps. As a preliminary investigation to this study, Ahn et al [19] confirmed that the end-to-end model (Wav2Vec2-XLS-R-1B) is suitable for evaluating children with SSDs even with limited training datasets for detecting speech errors. Owing to the pretrained framework of the Wav2Vec2-XLS-R-1B, it effectively learns from a wide variety of waveform contexts [19]. This pretraining allows the model to generalize well from smaller datasets and enables the model to better recognize and more accurately process Korean phonemes. They achieved approximately 90% accuracy in recognizing speech from children with SSDs in 73 Korean words using only about 1.5 hours of training data and without a language model. This suggests that sufficient accuracy can be attained in speech recognition models with low-resource languages to identify speech errors by using an end-to-end model.

Objective

This study aimed to (1) compare the evaluation results of the SLPs and the ASR model using South Korea's standardized SSD evaluation tools, specifically the APAC and U-TAP, and (2) analyze the cases of recognition errors occurring frequently in the ASR model.

Methods

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board

XSL•FO
RenderX

of the Catholic University of Korea, Incheon St. Mary's Hospital (protocol code OC22OISI0041; received approval on August 19, 2022). All participants provided informed consent and were compensated with KRW 50,000 (approximately US $40) for their participation. Written informed consent was obtained from all parents. For children aged 7 years and older, additional written assent was collected. Research data were securely stored on hospital servers, separate from personally identifiable information except for study identification numbers. Access to and analysis of data were limited exclusively to preapproved researchers. This paper and any supplementary materials contain no identifiable images or personal information regarding participants.

Recruitment

Participants were recruited between August 19, 2022, and June 14, 2023, from children referred for speech and language evaluations in the Department of Rehabilitation Medicine at Incheon St. Mary's Hospital. Eligible participants were native Korean-speaking children younger than 18 years of age who were referred primarily due to concerns with articulation. Exclusion criteria included (1) intellectual disabilities or autism spectrum disorders that impaired effective engagement in the assessment; and (2) motor speech disorders, fluency disorders, a history of cleft palate, or severely reduced speech intelligibility that would hinder accurate transcription. Only children meeting these criteria, and whose parents provided consent, were included in the study.

We included 30 children with suspected SSD (20 male and 10 female children; aged 3-7 years). The age and sex distribution of the participants were as follows: aged 3 years (2 male children), aged 4 years (10 male and 5 female children), aged 5 years (4 male and 1 female children), aged 6 years (2 male and 3 female children), and aged 7 years (2 male and 1 female children). The sample size was determined based on the median value observed in previous studies [20]. Additional participant information is provided in Multimedia Appendix 1.

ASR Model

The ASR model used in this study is an end-to-end model (Wav2Vec2-XLS-R-1B) originally provided by Facebook (Meta) as an open-source model [21] and modified by Ahn et al [19] to enhance its ability to detect speech error well in Korean children.

It was originally developed through training with 436,000 hours of data from adult voice databases, encompassing 128 languages. To enhance the model's ability to detect articulation errors, Ahn et al [19] used 93.6 minutes of speech data comprising 73 words from the APAC and U-TAP, and an additional 12 words, yielding a total of 6935 words from 95 children. The additional 12 words were selected during the engine fine-tuning stage to include phonemes and contexts not originally covered in the APAC or U-TAP. Detailed information regarding the word list for the model training can be found in Multimedia Appendix 2 [5,6].

Evaluation of Model Performance

This study used two Korean standardized SSD tests, namely the APAC and U-TAP, which comprised 37 and 30 words, respectively, with 6 words overlapping between the two tests. The examinations and recordings were conducted in the speech therapy room at Incheon St. Mary's Hospital by three experienced SLPs (DHK, JWJ, and DK), all of whom were trained in the assessment and treatment of children with SSDs. Although the facility was not completely soundproof, it was conducted in an environment where external noise was controlled.

During the test, the participants were asked to name pictures, and their responses were recorded using an iPhone X (Apple). In the recording process, a sufficient pause of 500-1000 ms was ensured between each target word. The recordings were saved in WAV format and converted to a 16 kHz sampling rate to verify the recognition results. Each target word was attempted at least twice, allowing for the exclusion of words with quality issues, such as noise or insufficient volume. After recording, files were reviewed, and the most suitable file was selected for analysis.

To ensure transcription accuracy, each word spoken by the participants in Korean was independently transcribed by two SLPs, and any disagreements were resolved by a third party for final decision-making. Transcription agreement between the two SLPs across all phonemes was 95% and 93.80% for the APAC and U-TAP, respectively, based on a 10% random sample of the data. The transcriptions created by the model were then compared to those obtained by the SLPs (Figure 1).

XSL•FO
RenderX

Figure 1. Comparison of the transcription results between the ASR model and the SLPs. APAC: Assessment of Phonology and Articulation for Children; ASR: automatic speech recognition; C-PER: consonant phoneme error rate; PER: phoneme error rate; SLP: speech-language pathologist; U-TAP: Urimal Test of Articulation and Phonology.

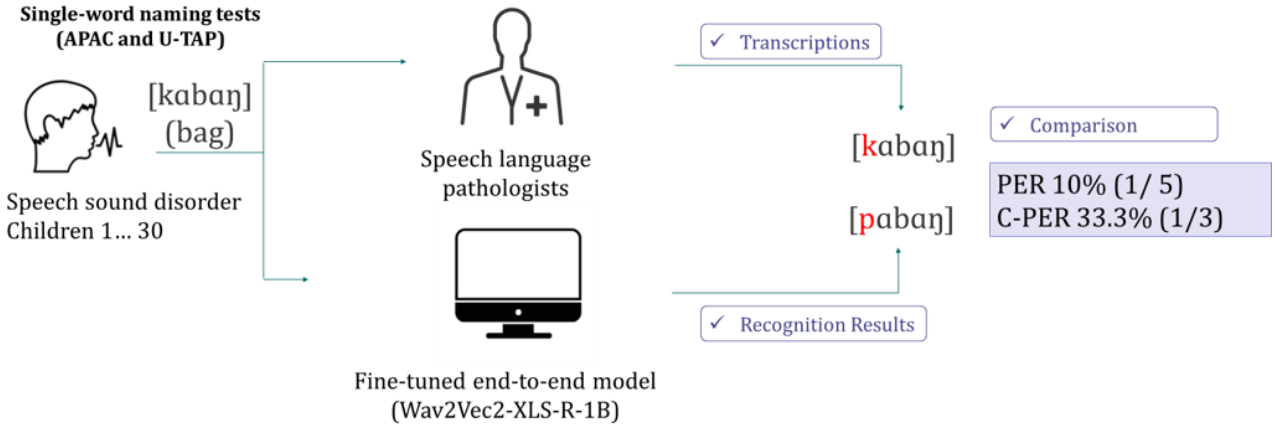

Variables

PER

The PER (%) has been used to determine the difference in the overall phoneme when comparing the transcription results of the SLPs and ASR model. The consonant PER (C-PER, %) was examined separately to analyze consonant errors in the ASR model.





PER and C-PER between the SLPs and the ASR model were conducted across all phonemes, regardless of whether they are target phonemes. This comprehensive approach is crucial given that the accuracy of the entire transcription is essential for assessing the clinical utility of the ASR.

Common ASR Disagreements

This count encompasses discrepancies between the SLPs and the ASR model, including instances where the SLPs flagged correct articulations that were identified as misarticulations by the ASR model, as well as cases where the SLPs identified misarticulations within the set that the ASR model failed to match. This indicator assesses which test words (or phonemes) differed considerably in evaluation between the SLPs and the ASR model during tests.

Target PCC

The PCC (%) is a metric used to assess consonant articulation accuracy. In this study, "target PCC" specifically refers to the percentage of correct consonant sounds produced for the target phonemes, in each word test (ie, the APAC and U-TAP). Additional information about the target phonemes, and the number of production opportunities are provided in Multimedia Appendix 3.



Severity

The *z* scores discussed in this study were based on the target PCC, which is used to evaluate the severity of SSDs. These *z* scores were calculated from the children's assessments conducted by both the ASR model and the SLPs, after which the severity was compared between the two.

Statistical Analysis

All data collected in this study were analyzed using SPSS (version 25.0; IBM Corp) for Windows. Descriptive statistics were used to describe the general characteristics of the evaluation results. The interrater reliability was calculated using the intraclass correlation coefficient (ICC) with a two-way random model, which served as the reliability coefficient for assessing the agreement on target PCC between the SLPs and the ASR model, using the absolute agreement method. Statistical significance was set at .05. The interpretation of the ICC was based on Landis and Koch [22].

Results

PERs

The PER, which indicated that the percentage of phonemes where the transcriptions by the SLPs and the ASR model differed among total phonemes, were 8.42% (457/5430) and 8.91% (402/4514) for the APAC and U-TAP, respectively. Additionally, the C-PER was 10.58% (327/3090) and 11.86% (331/2790) for APAC and U-TAP, respectively. The APAC uses a total of 174 phonemes, whereas the U-TAP consists of a total of 151 phonemes.

Common ASR Disagreements

Disagreements in Correct Articulations

An average of 2.60 (SD 1.54) and 3.07 (SD 1.39) phonemes per child were transcribed as correctly produced by the SLPs but incorrectly produced by the ASR model, totaling 78 and 92 instances among 30 children, in the APAC and U-TAP, respectively. Table 1 presents the percentage of phonemes that were judged by the SLPs as correctly produced but identified as incorrectly produced by the ASR model. Table 2 lists the types of disagreements that occurred three or more times.

XSL•FO
RenderX

Table 1. The number of disagreements between the ASR[a] and the SLPs[b] for phonemes that were judged by the SLPs as correctly produced.

	Plosives, n/N[c] (%)	Nasals, n/N (%)	Fluids, n/N (%)	Fricatives, n/N (%)	Affricatives, n/N (%)
APAC[d]					
Syllable-initial word-initial	11/381 (2.89)	0/115 (0)	—[e]	2/143 (1.40)	4/84 (4.76)
Syllable-initial word-medial	5/407 (1.23)	2/83 (2.41)	2/27 (7.41)	5/96 (5.21)	5/82 (6.10)
Syllable-final word-medial	3/43 (6.98)	9/205 (4.39)	0/39 (0)	—	—
Syllable-final word-final	7/116 (6.03)	12/143 (8.39)	6/66 (9.09)	—	—
U-TAP[g]					
Syllable-initial word-initial	14/349 (3.69)	1/117 (0.85)	1/10 (10)	0/70 (0)	5/78 (6.41)
Syllable-initial word-medial	9/343 (2.62)	3/163 (1.84)	5/34 (14.71)	4/63 (6.35)	5/71 (7.04)
Syllable-final word-medial	6/31 (19.35)	14/226 (6.19)	—	—	—
Syllable-final word-final	6/81 (7.41)	10/181 (5.52)	9/75 (12)	—	—

[a]ASR: automatic speech recognition model.

[b]SLP: speech-language pathologist.

[c]The number of disagreements between the ASR and the SLPs/total number of cases flagged as correct articulations by the SLPs.

[d]APAC: Assessment of Phonology and Articulation for Children.

[e]Not applicable.

[f]Fricatives and affricates are not realized as final consonants in Korean.

[g]U-TAP: Urimal Test of Articulation and Phonology.

Table 2. Common ASR[a] disagreements transcribed as correct articulations by the SLPs[b].

Target words (meaning) [IPA[c]]	ASR disagreement type	Frequency, n/N[d] (%)
APAC[e]		
화장실 (toilet) [hwad□ɑŋ□il]	Deletion of correctly sounded [l]	4/78 (5.13)
눈사람 (snowman) [nuns*ɑram]	Deletion of correctly sounded [n]	3/78 (3.84)
눈사람 (snowman) [nuns*ɑram]	Deletion of correctly sounded [m]	3/78 (3.84)
화장실 (toilet) [hwad□ɑŋ□il]	Substitution of correctly sounded [d□] with [d]	3/78 (3.84)
컵 (cup) [k[h]□]	Deletion of correctly sounded [□]	3/78 (3.84)
U-TAP[f]		
짹짹 (tweet) [t□*□t□*□]	Deletion of correctly sounded [□] (medial sound)	4/92 (4.34)
그림 (drawing) [k rim]	Deletion of correctly sounded [k]	3/92 (3.26)
그림 (drawing) [k rim]	Deletion of correctly sounded [m]	3/92 (3.26)
귀 (ears) [kwi]	Substitution of correctly sounded [k] with [d□]	3/92 (3.26)
바지 (pants) [pɑd□i]	Substitution of correctly sounded [d□] with [d]	3/92 (3.26)

[a]ASR: automatic speech recognition model.

[b]SLP: speech-language pathologist.

[c]IPA: International Phonetic Alphabet.

[d]The number of specific ASR disagreements occurred/total number of ASR disagreements, specifically among cases flagged as correct articulations by the SLPs.

[e]APAC: Assessment of Phonology and Articulation for Children.

[f] U-TAP: Urimal Test of Articulation and Phonology.

XSL•FO
RenderX

Disagreements in Misarticulations

The disagreements for phonemes transcribed as misarticulations by the SLPs averaged 7.87 (SD 3.66) and 7.57 (SD 4.85) occurrences per child, totaling 236 and 227 times among 30 children in the APAC and U-TAP, respectively. Table 3 presents the percentage of sounds that were judged by the SLPs as incorrectly produced but were identified as different by the ASR model. Table 4 lists the types of disagreements that occurred three or more times.

Table 3. The number of disagreements between the ASR[a] and the SLPs[b] for phonemes that were judged by the SLPs as incorrectly produced.

	Plosives, n/N[c] (%)	Nasals, n/N (%)	Fluids, n/N (%)	Fricatives, n/N (%)	Affricatives, n/N (%)
APAC[d]					
Syllable-initial word-initial	24/39 (61.54)	4/5 (80)	—[e]	20/67 (29.85)	11/36 (30.56)
Syllable-initial word-medial	27/73 (36.99)	2/7 (28.57)	5/93 (5.38)	42/114 (36.84)	18/68 (26.47)
Syllable-final word- medial	15/47 (31.91)	24/125 (19.20)	11/21 (52.38)	—	—
Syllable-final word-final	11/34 (32.35)	12/37 (32.43)	10/24 (41.67)	—	—
U-TAP[g]					
Syllable-initial word-initial	22/41 (51.22)	2/3 (66.67)	3/20 (15)	12/50 (24)	12/72 (16.67)
Syllable-initial word-medial	22/47 (46.81)	7/17 (41.18)	19/146 (13.01)	22/87 (25.29)	13/49 (26.53)
Syllable-final word-medial	13/29 (44.83)	34/134 (25.37)	—	—	—
Syllable-final word-final	14/39 (35.90)	26/59 (44.07)	7/45 (15.56)	—	—

[a]ASR: automatic speech recognition model.

[b]SLP: speech-language pathologist.

[c]The number of disagreements between the ASR and the SLPs/total number of cases flagged as misarticulations by the SLPs.

[d]APAC: Assessment of Phonology and Articulation for Children.

[e]Not applicable.

[f]Fricatives and affricates are not realized as final consonants in Korean.

[g]U-TAP: Urimal Test of Articulation and Phonology.

XSL•FO
RenderX

Table 4. Common ASR[a] disagreements transcribed as misarticulations by the SLPs[b].

Target words (meaning) [IPA[c]]	SLPs' judgment	ASR results	Frequency, n/N[d] (%)
APAC[e]			
딸기(strawberry) [t*algi]	[l] → Omitted	Recognizing misarticulation as correct articulation	6/236 (2.54)
이빨 (teeths) [i p*al]	[l] → Omitted	Recognizing misarticulation as correct articulation	5/236 (2.12)
딸기 (strawberry) [t*algi]	[g] → [k*]	Recognizing misarticulation as correct articulation	5/236 (2.12)
단추 (button) [tant□ʰu]	[t□ʰ] → [t□*]	Recognizing misarticulation as correct articulation	4/236 (1.69)
눈사람(snowman) [nuns*aram]	[n] → Omitted	Recognizing misarticulation as correct articulation	4/236 (1.69)
U-TAP[f]			
동물원 (zoo) [doŋmurw n]	[ŋ] → [m]	Recognizing misarticulation as correct articulation	6/227 (2.64)
괴물(monster) [kwemul]	[l] → Omitted	Recognizing misarticulation as correct articulation	5/227 (2.2)
눈썹 (eyebrow) [nuns*□]	[□] → Omitted	Recognizing misarticulation as correct articulation	4/227 (1.76)
단추 (button) [tant□ʰu]	[t□ʰ] → [t□*]	Recognizing misarticulation as correct articulation	4/227 (1.76)
짹짹 (tweet) [t□*□t□*□]	[□(final)] → Omitted	Recognizing misarticulation as correct articulation	4/227 (1.76)

[a]ASR: automatic speech recognition model.

[b]SLP: speech-language pathologist.

[c]IPA: International Phonetic Alphabet.

[d]The number of specific ASR disagreements occurred/total number of ASR disagreements, specifically among cases flagged as misarticulations by the SLPs.

[e]APAC: Assessment of Phonology and Articulation for Children.

[f]U-TAP: Urimal Test of Articulation and Phonology.

Target PCC

For the APAC, the mean target PCC was 74.76% (SD 15.21%) for SLPs and 76.71% (SD 15.20%) for the ASR model. The ICC for APAC demonstrated excellent reliability, with a value of 0.984 (95% CI 0.953-0.994). Similarly, for the U-TAP, the mean target PCC was 73.88% (SD 16.13%) for SLPs and 76.05% (SD 15.26%) for ASR. The ICC for U-TAP also indicated strong reliability, with a value of 0.978 (95% CI 0.941-0.990). The APAC uses 70 target consonants out of a total of 174 phonemes, whereas the U-TAP is based on 43 target consonants out of a total of 151 phonemes.

Severity

Table 5 displays the *z* scores for the target PCC in each test, as evaluated by the SLPs and ASR. The *z* scores between SLPs and ASR showed more pronounced differences with the APAC than with the U-TAP, with 8 individuals showing discrepancies in the APAC compared to only 2 in the U-TAP. The variations in *z* scores for each child are detailed in Multimedia Appendix 1.

XSL•FO
RenderX

Table 5. Comparison of children's z scores assessed by the SLPs[a] and the ASR[b] model.

z score	APAC[c]		U-TAP[d]	
	SLPs[c]	ASR[d]	SLPs[c]	ASR[d]
Over +1	0	0	0	0
0 to 1	1	4	2	3
–1 to 0	3	3	0	0
–2 to –1	6	2	1	1
Under –2	17	18	24	23
Total[e]	27	27	27	27

[a]SLP: speech-language pathologist.

[b]ASR: automatic speech recognition model.

[c]APAC: Assessment of Phonology and Articulation for Children.

[d]U-TAP: Urimal Test of Articulation and Phonology.

[e]The APAC and U-TAP provide *z* scores specifically for children aged 2 to 6. Thus, 3 children who were aged 7 were excluded from the analysis of the *z* score comparisons.

Discussion

Overview

This study highlights the effectiveness of the end-to-end model in accurately capturing children's articulations at the phoneme level. The PER between the SLPs and ASR model for the APAC and U-TAP assessments showed close correspondence (approximately 8%), with C-PER exhibiting similar ranges (ie, 10%-11%). Notably, the higher C-PER compared to the overall PER suggests fewer disagreements for vowels. Although the ASR model's results did not perfectly align with the SLPs' transcriptions, it is important to note that studies using standardized tests for SSDs typically report interrater reliability of 0.90 or higher, though it never reaches 100% [5,6,9,23]. Therefore, inherent variability in transcription, even among the SLPs, may have influenced the ASR evaluation process.

The occurrence of ASR disagreements in cases of correct articulation was relatively low, averaging approximately 2-3 instances per child. However, discrepancies in plosive sounds, particularly in the syllable-final word-medial position in the U-TAP, were notably higher compared to the APAC. This difference may stem from children imitating bird sounds when producing 짹짹 [t□* □t□* □] (tweet, onomatopoeia) in the U-TAP, reducing transcription reliability, as seen in common ASR disagreements cases.

In contrast, disagreements between the SLPs and the ASR model for misarticulations were approximately three times higher than for correct articulations. This finding is consistent with Jing and Grigos [9], who noted that the reliability of evaluation diminishes further when assessing misarticulated phonemes rather than correctly articulated ones. Disagreements between the ASR and the SLPs were higher for nasals and plosives but lower for fricatives and affricates in both the APAC and U-TAP, particularly in the syllable-initial word-initial position. This discrepancy may be attributed to the limited dataset, as fricatives and affricates are developmentally acquired later, resulting in a greater opportunity for the model to learn more error data related to these sounds.

The results of this study suggest several implications for the selection of test words when using the ASR. First, test words with long and complex syllable structures, such as 옥수수 [o□s*usu] (corn), 화장실 [hwɑd□aŋ□il] (bathroom), and 눈사람 [nuns*ɑram] (snowman) can hinder ASR systems in recognizing specific phonemes produced by children, particularly when the target phoneme is located in the word-medial position. Furthermore, accurate evaluation of the consonants ㅈ [t□, d□], ㄷ [t, d], and ㄱ [k, g] was challenging when followed by the vowel /i/, as seen in words like 바지 [pɑd□i] (pants) and 귀 [kwi] (ears), with ASR showing confusion between these phonemes. This difficulty is likely attributed to the proximity of their articulatory positions, which leads to similar auditory perceptual characteristics. In other words, assessing the accurate production of the preceding phoneme becomes particularly difficult when followed by the vowel /i/. In addition, the ASR demonstrated low performance in recognizing incorrect articulations of liquid sounds in syllable-final, word-medial positions. For example, it struggled to accurately detect the omission of the /l/ sound, as seen in words like 딸기 [t*ɑlgi] (strawberry) and 빨대 [p*ɑlt*] (straw).

Moreover, certain speech error patterns may not be accurately assessed using the ASR. First, it was difficult to distinguish whether plosive sounds at the end of a word were omitted or correctly produced (eg, 책 [t□ʰ□] (book)). Since syllable-final plosives are unreleased stops in Korean, the ASR may fail to accurately assess the speech of children who omit plosive consonants in the coda. Therefore, visual observation by SLPs should accompany the ASR evaluation to provide a more accurate assessment. Second, children exhibiting deaspiration errors may not be appropriately evaluated by the ASR, as misarticulation of aspirated consonants as tense consonants is difficult to detect (eg, 단추 [tɑnt□ʰu] vs [tɑnt□*u]; buttons). This highlights the limitations of ASR in capturing certain types

XSL•FO
RenderX

of speech errors, underscoring the need for supplementary evaluation methods, such as human auditory judgment, to ensure accurate diagnosis.

In addition, it is essential to compare the PCC and *z* scores derived from the ASR recognition results, as these provide a more comprehensive analysis of its clinical usefulness. The reliability of the ASR model developed by Ahn et al [19], using both the APAC and U-TAP, was deemed excellent, as indicated by the target PCC, which showed a strong correlation with the SLPs' evaluation results. However, this could be inflated with respect to diagnosing SSDs, as it is limited to results for only the targeted phoneme, rather than considering all phonemes in the test word.

When comparing the *z* scores for the target PCC, more pronounced differences were observed when using the APAC than when using the U-TAP (8 vs 2 individuals, respectively). This finding suggests that differences between the ASR model and SLPs may be more pronounced when using the APAC. These differences likely reflect variations in the normative data used for each test, with the inclusion of children with SSD being a key factor. Normative samples should include children with speech or language issues to avoid overidentification, placing normally developing children at the lower end of the distribution [24]. In this regard, the U-TAP included only "normal children," while the APAC incorporated both "general children" and those suspected of having functional articulation and phonological disorders, though the exact number was unspecified. In the U-TAP, most children in this study were classified as severe cases with *z* scores below –2, which may explain the lack of significant change in *z* scores despite slight differences in ASR and SLP evaluations. This finding is also aligned with Yi and Kim [23], who compared *z* scores across different SSD assessment tools [23].

Limitations and Suggestions

The limitations of this study include the use of a fine-tuned end-to-end model based on Korean phonemes, which may not be applicable for assessing speech errors in other languages or distortions not present in Korean.

Additionally, while machine learning relies on human transcriptions, it is essential to acknowledge that SLPs' transcriptions are not always consistent. These discrepancies must be considered in the ASR's training and evaluation, and future research should compare disagreement patterns among SLPs and between SLPs and the ASR.

In particular, this study conducted internal validation, meaning that both the model development and verification processes were carried out within the same setting. In addition, the small sample size makes it challenging to generalize the results. Therefore, external validation using other clinical settings or recording devices and including a larger number of participants is necessary.

For future studies, we suggest that incorporating connected speech tests may enhance the model's applicability in real-world clinical settings. Additionally, integrating automated error pattern analysis programs into the ASR to evaluate and categorize children's speech sound errors could streamline the evaluation process and save time for SLPs.

Conclusions

This study has been the first to assess the efficacy of ASR at the phoneme level in discerning speech errors among children afflicted with SSDs in South Korea. The model exhibited a commendable reliability when compared to transcriptions provided by the SLPs. Thus, our findings clearly confirm the viability and potential of using such a model in the domain of speech-language pathology. Still, given that assessing children with SSDs is difficult solely through auditory perception, ASR cannot fully replace traditional evaluations.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; RS-2022-NR071928), and by the Bio Industrial Technology Development Program (20017960) funded by the Ministry of Trade, Industry, and Energy (MOTIE, Korea). TA was also affiliated with MediaZen Inc at the time of the trial and is currently affiliated with Enuma Inc. YH was affiliated with MediaZen Inc at the time of the trial and is now affiliated with LG AI Research.

Data Availability

The datasets generated during or analyzed during this study are not publicly available due to the fact that no prior consent was obtained from the participants to share their data but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant demographics and evaluation results.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v27i1e60520_app1.xlsx]

Multimedia Appendix 2

XSL•FO
RenderX

Words list for the model training.
[DOCX File , 20 KB - jmir_v27i1e60520_app2.docx]

Multimedia Appendix 3
Target consonants and number of opportunities for production.
[DOCX File , 24 KB - jmir_v27i1e60520_app3.docx]

References

1. Waring R, Knight R. How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. Int J Lang Commun Disord 2013;48(1):25-40. [doi: 10.1111/j.1460-6984.2012.00195.x] [Medline: 23317382]
2. Kim SJ, Kim MJ, Ha S, Ha J. A survey of speech sound disorders in clinical settings. Commun Sci Disord 2015;20(2):133-144. [doi: 10.12963/csd.15236]
3. Kim SJ, Ko YK, Seo EY, Oh GA. Prevalence of speech sound disorders in 6-year-old children in Korea. Commun Sci Disord 2017;22(2):309-317. [doi: 10.12963/csd.17403]
4. Mcleod S, Baker E. Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. Clin Linguist Phon 2014;28(7-8):508-531. [doi: 10.3109/02699206.2014.926994] [Medline: 25000375]
5. Kim M, Pae SY, Lee SE. The development of the 'Test of Articulation for Children': concurrent validity. Commun Sci Disord 2005;10(1):82-96 [FREE Full text]
6. Kim Y, Shin M, Kim S, Ha J. Urimal Test of Articulation and Phonology (U-TAP). Seoul, South Korea: Hakjisa; 2004.
7. Skahan SM, Watson M, Lof GL. Speech-language pathologists' assessment practices for children with suspected speech sound disorders: results of a national survey. Am J Speech Lang Pathol 2007;16(3):246-259. [doi: 10.1044/1058-0360(2007/029)] [Medline: 17666550]
8. Keshet J. Automatic speech recognition: a primer for speech-language pathology researchers. Int J Speech Lang Pathol 2018;20(6):599-609. [doi: 10.1080/17549507.2018.1510033] [Medline: 31274357]
9. Jing L, Grigos MI. Speech-language pathologists' ratings of speech accuracy in children with speech sound disorders. Am J Speech Lang Pathol 2022;31(1):419-430. [doi: 10.1044/2021_ajslp-20-00381]
10. Cole RA. Listening for mispronunciations: a measure of what we hear during speech. Percept Psychophys 1973;13(1):153-156. [doi: 10.3758/bf03207252]
11. Howard SJ, Heselwood BC. Learning and teaching phonetic transcription for clinical purposes. Clin Linguist Phon 2002;16(5):371-401. [doi: 10.1080/02699200210135893] [Medline: 12185984]
12. Kewley-Port D, Burkle TZ, Lee JH. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. J Acoust Soc Am 2007;122(4):2365-2375. [doi: 10.1121/1.2773986] [Medline: 17902871]
13. Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 1955;27(2):338-352. [doi: 10.1121/1.1907526]
14. Klein HB, Grigos MI, Byun TM, Davidson L. The relationship between inexperienced listeners' perceptions and acoustic correlates of children's /r/ productions. Clin Linguist Phon 2012;26(7):628-645. [doi: 10.3109/02699206.2012.682695] [Medline: 22690718]
15. Dudy S, Bedrick S, Asgari M, Kain A. Automatic analysis of pronunciations for children with speech sound disorders. Comput Speech Lang 2018;50:62-84 [FREE Full text] [doi: 10.1016/j.csl.2017.12.006] [Medline: 29628620]
16. Suanpirintr S, Thubthong N. The effect of pauses in dysarthric speech recognition study on Thai cerebral palsy children. 2007 Presented at: i-CREATe07: International Conference for Rehabilitation Engineering & Assistive Technolog; April 26, 2007; Singapore. [doi: 10.1145/1328491.1328530]
17. Mazenan MN, Swee TT, Soh SS. Recognition test on highly newly robust Malay corpus based on statistical analysis for Malay articulation disorder. In: IEEE. 2014 Presented at: The 7th 2014 Biomedical Engineering International Conference; November 28, 2014; Fukuoka, Japan p. 1-5. [doi: 10.1109/bmeicon.2014.7017394]
18. Lee SH, Kim M, Seo HG, Oh BM, Lee G, Leigh JH. Assessment of dysarthria using one-word speech recognition with hidden Markov models. J Korean Med Sci 2019;34(13):e108 [FREE Full text] [doi: 10.3346/jkms.2019.34.e108] [Medline: 30950253]
19. Ahn T, Hong Y, Im Y, Kim DH, Kang D, Jeong JW, et al. Automatic speech recognition (ASR) for the diagnosis of pronunciation of speech sound disorders in Korean children. Clin Linguist Phon 2024:1-14. [doi: 10.1080/02699206.2024.2387609] [Medline: 39162064]
20. McKechnie J, Ahmed B, Gutierrez-Osuna R, Monroe P, McCabe P, Ballard KJ. Automated speech analysis tools for children's speech production: a systematic literature review. Int J Speech Lang Pathol 2018;20(6):583-598. [doi: 10.1080/17549507.2018.1477991] [Medline: 29996691]

XSL•FO
RenderX

21. Baevski A, Zhou Y, Mohamed A, Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations. Adv NeuralInf Process Syst 2020:33. [doi: 10.48550/arXiv.2006.1147]
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174. [doi: 10.2307/2529310]
23. Yi RD, Kim SJ. Comparisons among single-word tests for Korean speech sounds and results of assessment for children with speech sound disorders. Commun Sci Disord 2019;24(4):1040-1053. [doi: 10.12963/csd.19664]
24. Dodd B, Holm A, Hua Z, Crosbie S. Phonological development: a normative study of British English-speaking children. Clin Linguist Phon 2003;17(8):617-643. [doi: 10.1080/0269920031000111348] [Medline: 14977026]

Abbreviations

APAC: Assessment of Phonology and Articulation for Children
ASR: automatic speech recognition
C-PER: consonants phoneme error rate
HMM: hidden Markov model
ICC: intraclass correlation coefficient
PER: phoneme error rate
PCC: percentage of consonants correct
SLP: speech-language pathologist
SSD: speech sound disorder
U-TAP: Urimal Test of Articulation and Phonology

Edited by A Mavragani; submitted 15.05.24; peer-reviewed by JW Ha, H Goldstein, A Ahmadi; comments to author 01.07.24; revised version received 23.09.24; accepted 17.11.24; published 14.01.25.

Please cite as:
Kim DH, Jeong JW, Kang D, Ahn T, Hong Y, Im Y, Kim J, Kim MJ, Jang DH
Usefulness of Automatic Speech Recognition Assessment of Children With Speech Sound Disorders: Validation Study
J Med Internet Res 2025;27:e60520
URL: https://www.jmir.org/2025/1/e60520
doi:10.2196/60520
PMID:39576242

©Do Hyung Kim, Joo Won Jeong, Dayoung Kang, Taekyung Ahn, Yeonjung Hong, Younggon Im, Jaewon Kim, Min Jung Kim, Dae-Hyun Jang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Supervised Explainable Machine Learning Model for Perioperative Neurocognitive Disorder in Liver-Transplantation Patients and External Validation on the Medical Information Mart for Intensive Care IV Database: Retrospective Study

Zhendong Ding[1], MD; Linan Zhang[1], MD; Yihan Zhang[1], MD; Jing Yang[1], MD, PhD; Yuheng Luo[2], MD; Mian Ge[1], MD, PhD; Weifeng Yao[1], MD, PhD; Ziqing Hei[1], MD, PhD; Chaojin Chen[1], MD, PhD

[1]Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[2]Guangzhou AI & Data Cloud Technology Co., LTD, Guangzhou, China

Corresponding Author:
Chaojin Chen, MD, PhD
Department of Anesthesiology
The Third Affiliated Hospital of Sun Yat-sen University
No. 600 Tianhe Road
Guangzhou, 510630
China
Phone: 86 13430322182
Email: chenchj28@mail.sysu.edu.cn

Abstract

Background: Patients undergoing liver transplantation (LT) are at risk of perioperative neurocognitive dysfunction (PND), which significantly affects the patients' prognosis.

Objective: This study used machine learning (ML) algorithms with an aim to extract critical predictors and develop an ML model to predict PND among LT recipients.

Methods: In this retrospective study, data from 958 patients who underwent LT between January 2015 and January 2020 were extracted from the Third Affiliated Hospital of Sun Yat-sen University. Six ML algorithms were used to predict post-LT PND, and model performance was evaluated using area under the receiver operating curve (AUC), accuracy, sensitivity, specificity, and F_1-scores. The best-performing model was additionally validated using a temporal external dataset including 309 LT cases from February 2020 to August 2022, and an independent external dataset extracted from the Medical Information Mart for Intensive Care Ⅳ (MIMIC-Ⅳ) database including 325 patients.

Results: In the development cohort, 201 out of 751 (33.5%) patients were diagnosed with PND. The logistic regression model achieved the highest AUC (0.799) in the internal validation set, with comparable AUC in the temporal external (0.826) and MIMIC-Ⅳ validation sets (0.72). The top 3 features contributing to post-LT PND diagnosis were the preoperative overt hepatic encephalopathy, platelet level, and postoperative sequential organ failure assessment score, as revealed by the Shapley additive explanations method.

Conclusions: A real-time logistic regression model-based online predictor of post-LT PND was developed, providing a highly interoperable tool for use across medical institutions to support early risk stratification and decision making for the LT recipients.

(J Med Internet Res 2025;27:e55046) doi:10.2196/55046

KEYWORDS

machine learning; risk factors; liver transplantation; perioperative neurocognitive disorders; MIMIC-Ⅳ database; external validation

XSL•FO
RenderX

Introduction

Perioperative neurocognitive disorder (PND), encompassing various postsurgical cognitive impairments identified especially in the postoperative period, was first proposed in 2018 [1]. These cognitive changes are consistent with the clinical diagnostic criteria for neurocognitive disorders outlined in the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) [1-3]. In addition to postoperative delirium (POD) [4,5], other components of PND include emergence delirium, delayed neurocognitive recovery, and postoperative neurocognitive dysfunction [2,6]. POD or PND incidence is 2%-3% after general surgery [5,7] and 50%-70% in high-risk patients [8]. In addition, PND not only contributes to increased mortality rates but also extends hospitalization in patients undergoing liver transplantation (LT) [7,9], escalating health care costs and resource use. Preventative strategies and timely interventions for post-LT PND are crucial for enhancing patient outcomes and easing health care burdens [10].

Existing studies identify risk factors for post-LT PND, such as excessive alcohol consumption, Child-Turcotte-Pugh scores, and model for end-stage liver disease (MELD) scores [11,12]. Potential biomarkers for cognitive impairment prediction have also been proposed, including calcium binding protein β and neuron-specific enolase [13], yet their practical application is hindered by complex clinical scenarios and expense.

Machine learning (ML), a branch of artificial intelligence, offers a solution by distilling extensive clinical data into actionable insights, identifying relative risk factors for PND [14,15]. However, there is a dearth of ML-based models predicting post-LT–related complications [16-22] and postoperative delirium during specific surgeries [4,23]. There are currently no appropriate models for predicting PND in LT recipients, with most current clinical prediction models often failing to maintain accuracy when applied to external datasets, resulting in significant limitations to their generalizability.

This study aimed to extract critical predictors and develop an efficient ML algorithm to predict PND in LT recipients using routinely collected clinical data and to validate its performance using the Medical Information Mart for Intensive Care Ⅳ (MIMIC-Ⅳ) database.

Material and Methods

Study Design and Patients

This retrospective, single-center study was conducted at our institution following the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis guidelines. We enrolled 1267 patients who underwent LT between January 2015 and August 2022. Records were extracted using the perioperative specialist database platform (PSDP) and electronic patient record (EPR) systems. The inclusion and exclusion criteria are shown in Textbox 1.

All included recipients were formalized and registered in the China Organ Transplant Response System.

Textbox 1. Inclusion and exclusion criteria for the study.

Inclusion criteria

- Age >18 years.
- Allogeneic liver transplantation.

Exclusion criteria

- Simultaneous liver and kidney transplantation.
- Preoperative overt hepatic encephalopathy.
- Emergency reoperation.
- Persistent postoperative coma and inability to screen for cognitive function.
- Post–liver transplantation cerebral infarction or hemorrhage.
- Incomplete medical records.

Data Collection

The development and temporal validation cohort datasets were created by extracting original records from the Docare System (Medical system), Hospital Information System, and Laboratory Information System, and integrating them into the PSDP platform and EPR systems. To increase ML model accuracy and applicability, we included the following variables: (1) demographic characteristics; (2) liver donor characteristics; (3) preoperative comorbidities, complications, preoperative treatment, and LT etiology; (4) preoperative laboratory test results; (5) intraoperative surgery characteristics and medications; (6) postoperative MELD scores, sequential organ failure assessment (SOFA) scores, and laboratory test results; and (7) complications and prognosis in LT recipients. All of the original data were made anonymous throughout the study.

Definitions of Outcomes

The primary outcome was postoperative PND occurrence from surgery until discharge from the hospital. A summary of perioperative neurocognitive impairments is shown in Table S1 in Multimedia Appendix 1**.** The initial diagnosis criteria was the retrieval of any of the following terms from the medical records: “Delirium”, “Confusion”, “Confusional arousals”,

XSL•FO
RenderX

"Clouding of consciousness", "Soma", "Drowsiness", "Changes in mental status", "Hallucinations", "Disorientation", "Dyscalculia", "Haziness of spirit-mind", "Irritability", "Agitation", "Inattentiveness", "Reactive confusion", "Somatization disorder", "Irritability", and "Somatoform disorders", or equivalent terms in Chinese [4,24,25]. Next, each patient was evaluated based on the *DSM-5* criteria by a designated neurologist without prior access to the patient's records [3,26].

Variable Selection

A comprehensive set of 137 variables was extracted for the initial analysis (Table S2 in Multimedia Appendix 1). Table S3 in Multimedia Appendix 1 provides a concise explanation of the main complications and relevant term definitions. Postoperative SOFA scores were calculated by intensive care unit (ICU) physicians immediately after surgery according to European Society of Intensive Care Medicine criteria [27] and submitted for statistical analysis.

To account for multicollinearity and confounding variables affecting the overall model fitting performance, variables that were statistically significant (P<.05) in the univariate test were subjected to stability selection (Table S4 in Multimedia Appendix 1) [28]. After 100 iterations of least absolute shrinkage and selection operator (LASSO) regression, the top 10 features with the highest selection frequencies were chosen to train the ML models. For each LASSO regression, 90% of the training set samples were randomly selected as subsamples.

Machine Learning Models

The following 6 ML models were developed, and their performances were further evaluated: logistic regression (LR), multilayer perceptron classifier (MLP), extreme gradient boosting with classification trees (XGB), light gradient boosting machine (LGB), support vector machine (SVM), and random forest classifier (RF). All models were constructed using the XGB, LGB, and Scikit-learn packages.

The primary cohort dataset was randomly divided into 80% development and 20% internal validation sets. The bootstrap method was implemented 1000 times on the internal validation set to determine a 95% CI for the discrimination assessment metrics for each model: the area under the receiver operating curve (AUC), accuracy, sensitivity, specificity, and F_1-scores. Considering that ML models have multiple hyperparameters that are essential for model performance, a 5-fold cross-validation grid search method was used to optimize the parameters and AUCs (Table S5 in Multimedia Appendix 1). The Shapley additive explanations (SHAP) method was used to assess predictive feature importance and explain the ML algorithms' predictions [29].

Model Performance Comparison and MIMIC-Ⅳ Dataset

Because the SOFA and MELD scores have been reported as potential predictors of various post-LT complications [16,30], our study also compared the ML model's performance against SOFA and MELD scores.

An external validation set extracted from the MIMIC-Ⅳ (version 2.2) [31] database was used to evaluate the ML model's performance, which was authorized by the review committee of Massachusetts Institute of Technology (agreement 1.5.0). Patients who underwent LT surgery and were diagnosed with PND according to the International Classification of Diseases (9th and 10th revisions) were enrolled. Data extraction and cleaning were performed using PostgreSQL (version 15.3) and Navicate Premium (version 16) with a Structured Query Language (Figure S1 in Multimedia Appendix 1).

Statistical Analysis

Data cleaning used Python (version 3.9.13) packages Pandas (version 1.4.4) and Numpy (version 1.23.5). Data analysis used the Python Scipy package (version 3.7), and SHAP (0.41.0) was used to visualize and analyze feature importance.

Data distribution was evaluated using the Kolmogorov Smirnov test. Normally distributed continuous variables are presented as mean (SD) and were compared by independent sample t tests. Non-normally distributed continuous data are presented as median (IQR) and were compared using the nonparametric equivalent (Mann Whitney test). Categorical variables are expressed as frequencies and percentages and were tested using the chi-square test or Fisher exact test. Long-term survival rates were estimated using the Kaplan Meier method. Group comparisons were conducted using the Gehan-Breslow Wilcoxon test and log-rank tests.

All tests were 2-tailed, with statistical significance set at 0.05. Before ML model training, continuous variables were normalized, dichotomous variables were coded as binary variables, and multicategory variables were coded as uniform numbers.

Variables with missing values exceeding 20% were excluded, and missing values below 20% were imputed with the median (for numeric variables) or mode (for categorical variables). The overall data distribution after imputation exhibited an acceptable level of variability.

Visualized Online Calculator

An online calculator with a visual interface was developed to facilitate the easy input of clinical variables and to generate clear and meaningful output indicating the absolute risk in percentages.

Ethical Considerations

The study protocol was approved by the Ethics Committee of the Third Affiliated Hospital of Sun Yat-sen University on July 27, 2022 (No. (2019)02-609-04) and was conducted in accordance with the Declaration of Helsinki. The requirement for informed patient consent was waived due to the study's retrospective nature, and all data were anonymized before analysis.

Results

Patient Demographic Characteristics

The flowchart for patient recruitment is shown in Figure 1. Of the 958 patients who underwent LT, 751 patients were enrolled

XSL•FO
RenderX

randomly into the development set (n=600) and internal validation set (n=151). Notably, PND occurred in 201 patients, accounting for 33.5% of the development cohort. Table 1 and Table S6 in Multimedia Appendix 1 summarizes the development set's demographic characteristics, donor features, and perioperative variables of patients with or without post-LT PND.

Figure 1. Diagram of experimental procedure and flowchart, (A) brief diagram of the experimental procedure and (B) flowchart for patient enrollment, development and selection of machine learning model. LGB: light gradient boosting machine; LR: logistic regression; MIMIC-Ⅳ: the Medical Information Mart for Intensive Care Ⅳ; ML: machine learning; MLP: multilayer perceptron classifier; RF: random forest classifier; SVM: support vector machine; XGB: extreme gradient boosting with classification trees.

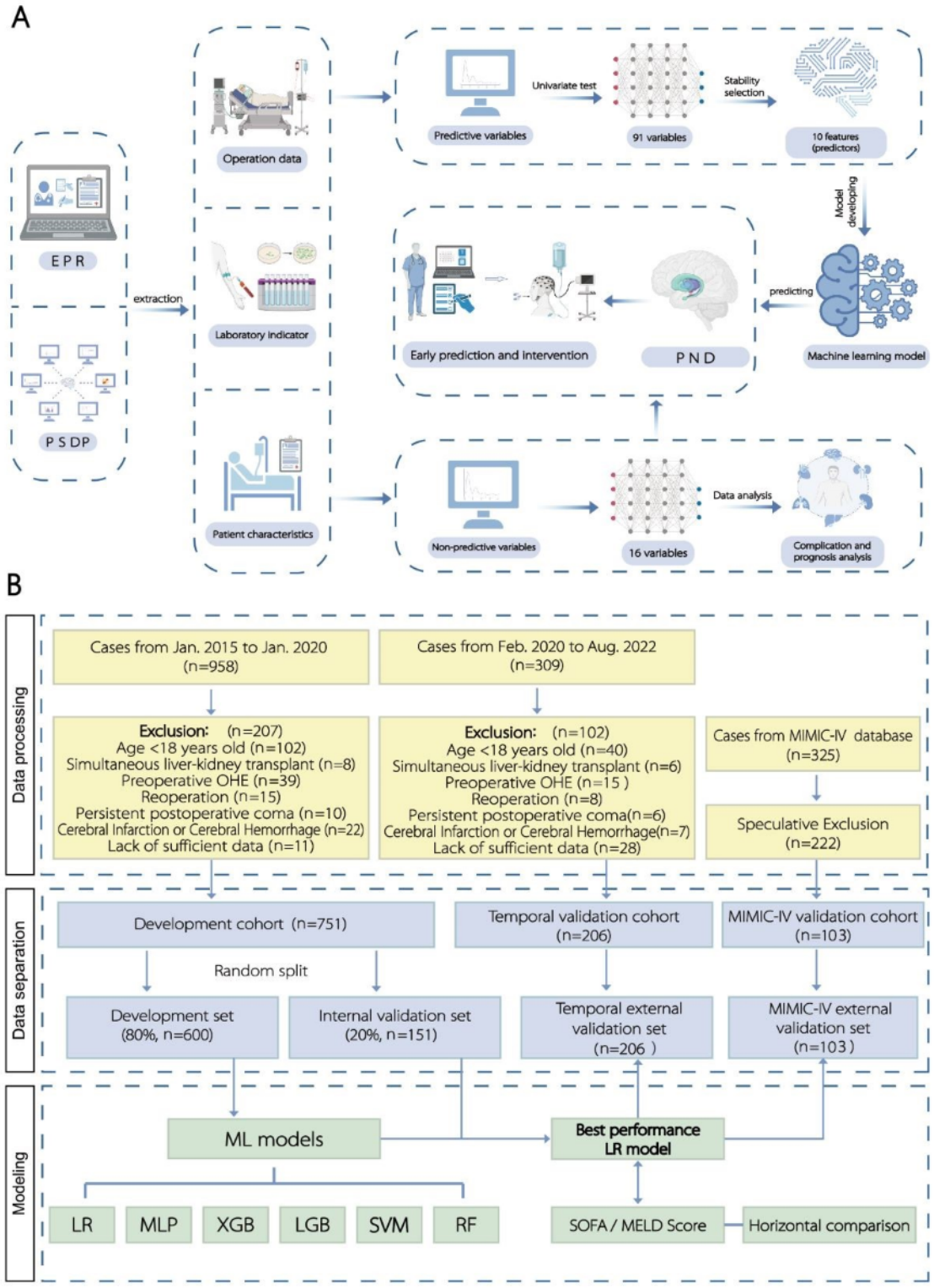

XSL•FO
RenderX

Table 1. Demographic characteristics and donor characteristics variables of patients with stratification by perioperative neurocognitive disorder.

Characteristics	Total (n=600)	NonPND[a] (n=399)	PND[a] (n=201)	*P* value
Demographic characteristics				
Age (years), mean (SD)	49 (10.34)	49.24 (10.16)	48.53 (10.7)	.43
Sex				.06
Female, n (%)	74 (12.33%)	42 (10.63%)	32 (16%)	
Male, n (%)	521 (86.83%)	353 (89.37%)	168 (84%)	
Height (cm), median (IQR)	170 (172-165)	170 (172-165)	169 (170-163)	.17
Weight (kg), median (IQR)	64 (71-58)	64 (72-58.88)	64 (70-58)	.26
BMI, median (IQR)	22.84 (24.85-20.43)	22.78 (24.90-20.45)	22.86 (24.74-20.44)	.79
Blood group, n (%)				.76
A	233 (38.83)	156 (39.49)	77 (38.5)	
B	156 (26)	106 (26.84)	50 (25)	
O	167 (27.83)	110 (27.85)	57 (28.5)	
AB	39 (6.5)	23 (5.82)	16 (8)	
Donor characteristics				
Donor age (years), median (IQR)	40 (49-28)	40 (50-28)	40 (48-29.5)	.64
Donor BMI, median (IQR)	22.49 (24.22-20.7)	22.49 (24.22-20.76)	22.59 (24.52-20.38)	.35
Donor Type				.03
DBD[b], n (%)	321 (53.5)	228 (62.3)	93 (50.54)	
DCD[c], n (%)	225 (37.5)	136 (37.16)	89 (48.37)	
DBCD[d], n (%)	4 (0.67)	2 (0.55)	2 (1.09)	
Steatosis of donor liver				.27
Steatosis grade 0, n (%)	390 (65)	265 (71.05)	125 (65.79)	
Steatosis grade 1, n (%)	147 (24.5)	94 (25.2)	53 (27.89)	
Steatosis grade 2, n (%)	26 (4.33)	14 (3.75)	12 (6.32)	

[a]PND: perioperative neurocognitive dysfunction.

[b]DBD: donation after brain death.

[c]DCD: donation after circulatory death.

[d]DBCD: donation after brain death followed by circulatory death.

Perioperative Characteristics

Among the preoperative characteristics, American Society of Anesthesiologists classification and preoperative comorbidities such as acute respiratory distress syndrome, and laboratory results including hemoglobin, white blood cell (WBC) count, liver function, coagulation function, and serum calcium were significantly different between patients with and without postoperative PND (*P*<.01, Table S6 in Multimedia Appendix 1). Specifically, individuals diagnosed with post-LT PND exhibited a notably elevated prevalence of preoperative cover hepatic encephalopathy (CHE; 45.27% vs 8.77%, *P*<.001) and hypercalcemia (7.57% vs 1.36%, *P*<.001). Furthermore, patients with post-LT PND had higher Child Pugh and MELD scores (*P*<.001), longer preoperative ICU stays, increased continuous blood purification, increased plasma exchange, longer mechanical ventilation, and higher tracheal intubation (all *P*<.001, Table S6 in Multimedia Appendix 1).

Regarding intraoperative characteristics, patients with post-LT PND had longer anesthesia durations; increased sodium bicarbonate levels, red blood cell counts, plasma levels, and levels of cryoprecipitate transfusion; increased estimated blood loss (EBL); and reduced urine output (all *P*<.001, Table S6 in Multimedia Appendix 1). Differences in intraoperative medications between the 2 groups were not significant, except for recombinant activated factor VII (*P*<.001). Interestingly, our results showed no association between day or night surgery and the incidence of PND (*P*=.44, Table S6 in Multimedia Appendix 1).

For the postoperative characteristics, patients with post-LT PND showed significantly higher levels of aspartate aminotransferase (AST), total bilirubin, blood urea nitrogen, prothrombin time

XSL•FO
RenderX

(PT), international normalized ratio, hypersensitive C-reactive protein (hsCRP), procalcitonin, and serum calcium, as well as lower levels of hemoglobin, hematocrit, WBC, platelet (PLT), gamma-glutamyltransferase, albumin, and serum osmolality (all $P<.05$, Table S6 in Multimedia Appendix 1).

Feature Selection

The frequency of LASSO algorithm selection for each variable is shown in detail in Figure S2 in Multimedia Appendix 1. The top 10 features chosen as predictors for ML model development were preoperative CHE, PLT, PT, estimated glomerular filtration rate (eGFR), Ca^{2+}, MELD score, intraoperative EBL, postoperative SOFA score, hsCRP, and AST.

Model Performance and Horizontal Comparison

The performance of the 6 ML models is shown in Figure 2. The LR model achieved the highest AUC (0.799, 95% CI 0.709-0.877) with acceptable accuracy (0.722, 95% CI 0.642-0.795), sensitivity (0.714, 95% CI 0.575-0.833), and specificity (0.73, 95% CI 0.639-0.811) compared with the other 5 models.

The SOFA (AUC=0.459, 95% CI 0.365-0.555), preoperative MELD (AUC=0.672, 95% CI 0.581-0.768), and postoperative MELD scores (AUC=0.679, 95% CI 0.587-0.772) had significantly lower AUCs than the LR model in the internal validation set (Figure 3A).

Figure 2. Performance metrics for six ML models. (A) ROC curves of six ML models. (B) Details of the model performance metrics. Accuracy=(TP+TN)/(TP+TN+FP+FN); AUC, the area under the receiver-operating curve; F1=2*Precision*Recall/ (Precision + Recall); FN: false negative; FP: false positive; LGB: light gradient boosting machine; LR: logistic regression; MLP: multilayer perceptron classifier; RF: random forest classifier; Sensitivity=TP/ (TP + FN); Specificity (Recall)=TN/ (TN + FP); SVM: support vector machine; TN: true negative; TP: true positive; XGB: extreme gradient boosting with classification trees.

A

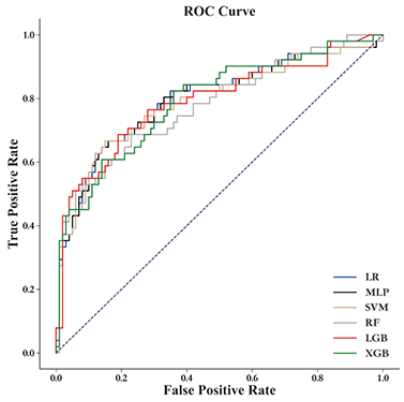

B

Model	AUC	Accuracy	Sensitivity	Specificity	F1
LR	0.799	0.722	0.714	0.73	0.632
	(0.709, 0.877)	(0.642, 0.795)	(0.575, 0.833)	(0.639, 0.811)	(0.513, 0.734)
MLP	0.797	0.788	0.61	0.882	0.66
	(0.708, 0.879)	(0.722, 0.854)	(0.463, 0.733)	(0.812, 0.942)	(0.537, 0.76)
SVM	0.787	0.728	0.725	0.734	0.643
	(0.692, 0.862)	(0.656, 0.788)	(0.592, 0.837)	(0.645, 0.809)	(0.525, 0.735)
RF	0.791	0.722	0.628	0.772	0.603
	(0.705, 0.871)	(0.656, 0.795)	(0.49, 0.767)	(0.686, 0.85)	(0.473, 0.708)
LGB	0.792	0.748	0.708	0.769	0.655
	(0.709, 0.872)	(0.675, 0.821)	(0.583, 0.83)	(0.685, 0.851)	(0.547, 0.755)
XGB	0.795	0.709	0.688	0.72	0.611
	(0.711, 0.863)	(0.636, 0.781)	(0.542, 0.809)	(0.636, 0.806)	(0.495, 0.712)

XSL•FO
RenderX

Figure 3. SHAP analysis of the LR model and model performance in horizontal comparison and external validation. (A) Horizontal comparison of predicting performance between the LR model and MELD/SOFA scores in the internal validation set. (B-C) The SHAP summary plot demonstrated the general importance of each feature in LR model. The color bar on the right indicates the relative value of a feature in each case, with red color representing higher value and blue color representing lower value. (D-E) ROC curves and model performance in the external validation. AST: aspartate aminotransferase; AUC: the area under the receiver-operating curve; CHE: cover hepatic encephalopathy; EBL: estimated blood loss; eGFR: estimated glomerular filtration rate; hsCRP: hypersensitive C-reactive protein; LR: logistic regression; MELD scores: model for end-stage liver disease score; MIMIC-IV: Medical Information Mart for Intensive Care Ⅳ; PLT: platelet; PT: prothrombin time; SHAP: Shapley additive explanations; SOFA scores: sequential organ failure assessment score.

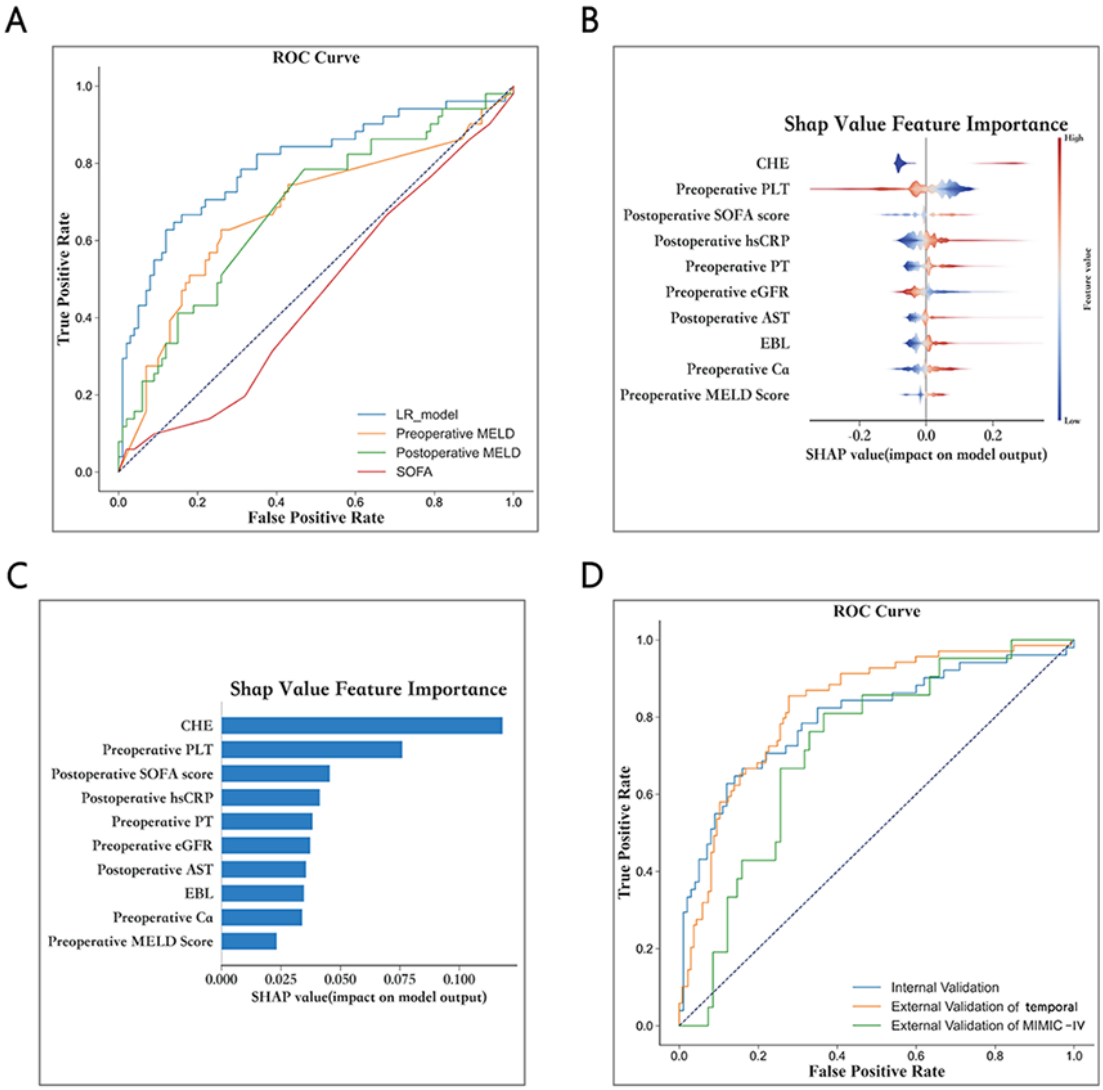

E

Validation set	AUC	Accuracy	Sensitivity	Specificity	F1
Internal	0.799 (0.709, 0.877)	0.722 (0.642, 0.795)	0.714 (0.575, 0.833)	0.73 (0.639, 0.811)	0.632 (0.513, 0.734)
Temporal external	0.826 (0.765, 0.887)	0.748 (0.689, 0.811)	0.766 (0.662, 0.861)	0.743 (0.667, 0.821)	0.671 (0.589, 0.755)
MIMIC-IV external	0.72 (0.606, 0.829)	0.631 (0.534, 0.728)	0.812 (0.64, 0.95)	0.589 (0.481, 0.694)	0.472 (0.312, 0.602)

XSL•FO
RenderX

Feature Importance

The SHAP summary plot (Figures 3B and 3C) illustrates the correlation between the feature value magnitudes in the LR model. Both SHAP plots revealed that the presence of CHE, lower preoperative PLT, higher postoperative SOFA score, higher postoperative hsCRP, and higher preoperative PT were associated with a higher SHAP value output in the LR model, indicating a heightened likelihood of post-LT PND and forming the top 5 effective variables.

Three correctly classified examples (eg, patients 48, 80, and 122) are presented in Figure S3 in Multimedia Appendix 1, showing the SHAP decision and force plots.

Temporal External Validation and MIMIC-Ⅳ Dataset Validation

A comparison of the main demographic characteristics and key predictive variables between the development and validation sets is shown in Table S7 in Multimedia Appendix 1, and the incidence rates of post-LT PND in the temporal and MIMIC-Ⅳ external validation were 27.1%, and 20.3%, respectively. The LR model exhibited a comparable performance in the temporal external validation set (AUC=0.826, 95% CI 0.765-0.887) (Figure 3D). Surprisingly, the LR model also provided acceptable predictions for the MIMIC-Ⅳ dataset (Figure 3D, AUC=0.72, 95% CI 0.606-0.829). Figure 3E summarizes the main performance metrics of the LR model.

Effect of Perioperative Neurocognitive Dysfunction on Patients' Outcomes and Prognosis

Compared with patients without post-LT PND, patients with PND were more likely to experience perioperative complications (Table S8 in Multimedia Appendix 1), including higher incidences of sepsis (51.63% vs 21.55%, *P*<.001), pneumonia (75.56% vs 65.46%, *P*<.05), acute kidney injury (69.5% vs 39.75%, *P*<.001), and hemodialysis (51.35% vs 12.81%, *P*<.001). Furthermore, patients with post-LT PND had higher hospitalization costs (CNY 377,801.69 [US $51,566.83], SD 177,855.53 [US $24,275.82] vs CNY 277,018.95 [US $37,810.82], SD 92,779.91 [US $12,663.70]; *P*<.001), prolonged postoperative stays (25 {18} vs 21 {11} days, *P*<.001), longer postoperative ICU stay (113 {114} vs 65 {48.5} hours, *P*<.001), and a markedly higher in-hospital mortality rate (12.44% vs 2.51%, *P*<.001).

Further survival analysis (Figure 4) was conducted to assess patient prognosis. The PND group exhibited significantly lower survival rates at 30 days (87.1% vs 97.84%, *P*<.001), 3 months (83.99% vs 96.46%, *P*<.001), 6 months (82.78% vs 95.38%, *P*<.001), and 12 months (78.85% vs 88.44%, *P*<.001), and overall survival (*P*=.03).

Figure 4. Post–liver transplantation survival associated with perioperative neurocognitive dysfunction. Patients with post–liver transplantation perioperative neurocognitive dysfunction showed a significantly lower survival rate. LT: liver transplantation; PND: perioperative neurocognitive dysfunction.

A

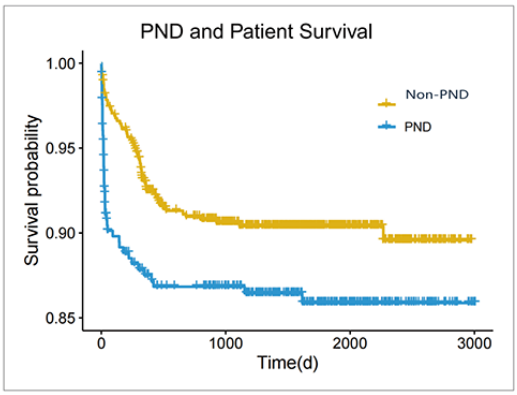

B

Survival duration	Total (N=980)	Non-PND (N=649)	PND (N=331)	*P-value*
30-day*	925(94.39%)	635(97.84%)	290(87.61%)	**<0.001**
3-month*	904(92.24%)	626(96.46%)	278(83.99%)	**<0.001**
6-month*	893(91.12%)	619(95.38%)	274(82.78%)	**<0.001**
12-month*	835(85.20%)	574(88.44%)	261(78.85%)	**<0.001**
3-years	673(68.67%)	448(69.03%)	225(67.98%)	0.737

Clinical Availability of the Logistic Regression Model

Given the accessibility of the 10 predictive features, we constructed a visually oriented online calculator to facilitate clinical decision making. The perioperative information of 2 typical patients was entered into the online calculator: patient 48 had a positive final predicted probability of PND occurrence (probability: 96%), and patient 122 had a negative final predicted probability of PND occurrence (probability: 17%; Figure 5). The online calculator is freely accessible at the hospital website.

XSL•FO
RenderX

Figure 5. Online calculator for the clinical interface of the post–liver transplantation perioperative neurocognitive dysfunction risk prediction logistic regression model. (A) Patient No. 48 post–liver transplantation perioperative neurocognitive dysfunction will occur (probability of perioperative neurocognitive dysfunction: 94%); (B) Patient No. 122 post–liver transplantation perioperative neurocognitive dysfunction will not occur (probability of perioperative neurocognitive dysfunction: 17%).

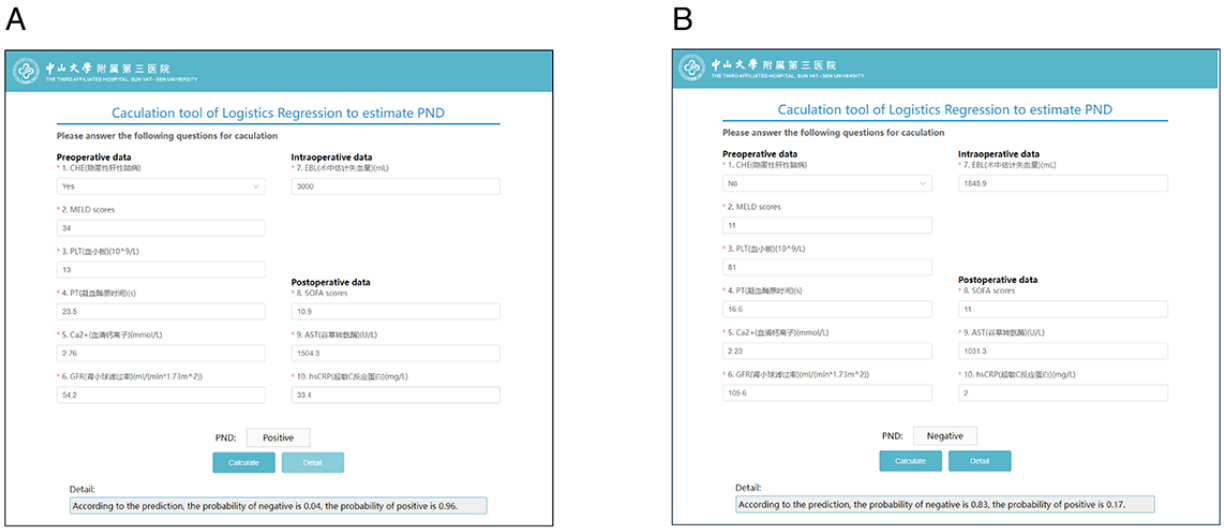

Discussion

Principal Findings

Our retrospective study assessed 6 different ML algorithms to predict post-LT PND, using 10 readily available clinical parameters. We found that post-LT PND incidence was 33.5%. The 10 predictive features significantly associated with PND included preoperative CHE, PLT, PT, eGFR, Ca^{2+}, MELD score, intraoperative EBL and postoperative SOFA score, hsCRP, and AST. The LR model demonstrated superior performance, with high AUC, accuracy, sensitivity, and specificity, surpassing traditional SOFA and MELD scores in predicting post-LT PND and performed acceptably in the rigorous temporal and MIMIC-Ⅳ external validations.

This study aids clinicians in detecting postoperative cognitive changes in LT recipients. Patients with PND typically faced more perioperative complications, higher hospitalization costs, and prolonged hospital and ICU stays, consistent with previous studies [4,23]. Hepatic encephalopathy has been reported as an independent risk factor for postoperative neurocognitive disorders [32]. To ensure cognitive assessment accuracy, we excluded patients with overt hepatic encephalopathy according to the spectrum of neurocognitive impairment in cirrhosis criteria [33]. CHE emerged as a significant predictor in our model analysis. Both oxidative stress and neuroinflammation have been implicated in POD pathophysiology [10,34]. A recent systematic review also links increased perioperative CRP levels to a high delirium risk [35], supporting our inclusion of hsCRP as a predictor. Calcium ions (Ca^{2+}) are important cell signaling molecules, and previous studies reported a positive correlation between Ca^{2+} concentration and neuronal apoptosis extent in vitro [36], consistent with our results. Furthermore, the model identified PLT as an unconventional indicator of PND, showcasing ML's ability to highlight nontraditional risk factors. This discovery is partly supported by Eyer et al [37] suggesting a relationship between lower PLT and delirium tremens.

Our study used preoperative, intraoperative, and postoperative data (SOFA scores, hsCRP, and AST levels) to develop the LR model. Earlier studies have revealed that multiple postoperative factors were also risk factors for PND [11,12]. The postoperative variables included in this study were predominantly assessed upon initial admission to the ICU. Stability selection analysis revealed a positive correlation between elevated postoperative SOFA scores, hsCRP levels, and AST levels, and an increased likelihood of post-LT PND. This highlights the predictive value of these commonly observed postoperative variables for PND.

Our results suggest that LR outperforms other ML models in predicting post-LT PND, which is not surprising. A recent systematic review showed no performance superiority of other ML models over LR in predicting clinical complications [38]. Wiredu et al [35] also found that compared to ML algorithms, LR had the highest AUC when predicting sex-specific hip fractures. Song et al [4] developed an LR model to predict POD in older adult patients, achieving the highest AUC compared with other models. Given the evident linear relationships among the top 10 features, the LR may be more appropriate for capturing distribution patterns. In contrast to other algorithms, LR performs well on nonoversized and high-dimensional datasets, exhibits computational efficiency, and imposes lower dataset requirements.

As demonstrated by the example of prediction cases (Figure 5), we successfully developed a predictive model for post-LT PND, with its primary advantage in its reliable predictive performance, validated using 2 external datasets. The importance of early detection and prevention of PND in patients undergoing cardiac surgery or transplantation is clearly emphasized in current international guidelines [39]. However, the implementation of preventive measures is often challenged by limited resources [40], especially in cases where the shortage of liver donors persists. On accurate identification by the LR model, patients

XSL•FO
RenderX

at high risk for post-LT POD could be referred to enhanced LT perioperative management strategies, such as individualized pharmacological or nonpharmacological comprehensive multicomponent interventions, according to the 10 commonly accessible predictive parameters filtered by the ML algorithm.

Limitations

However, this study had several limitations. First, it was a single-center retrospective study, meaning the Confusion Assessment Method (CAM) or the associated CAM-ICU and 3D-CAM were inappropriate for our database. Instead, patients with PND were identified from medical records according to the *DSM-5* criteria [3,6,26]. Second, as a real-world study, researchers can only infer precise risk factors based on the data available, and inhomogeneous confounding among the datasets could affect the study conclusions [41]. While our online decision tool has the potential to aid surgeons and anesthesiologists in clinical decision making, the causes and underlying mechanisms of PND remain subjects of intense debate, necessitating further research.

Conclusions

This study successfully develops a real-time and easily accessible parameter requiring LR-based PND prediction algorithm for post-LT settings. The LR model outperformed the other five models owing to its enhanced model performance and interpretability. The optimal use of our freely accessible online predictor would enable timely and convenient risk stratification, enhanced perioperative management strategies, and comprehensive multicomponent interventions.

Acknowledgments

We express our gratitude to Ms Li Jing from the Chinese MCC5 University for her valuable assistance in processing the figures. This study was supported partly by the Special Support Project of Guangdong Province (0720240209), the Natural Science Foundation of Guangdong Province (grant 2022A1515012603), the Joint Funds of the National Natural Science Foundation of China (U22A20276), Science and Technology Planning Project of Guangdong Province-Regional Innovation Capacity and Support System Construction (2023B110006), Provincial-enterprise Joint Funds of Guangdong Basic and Applied Basic Research Foundation (2021B1515230012), Science and Technology Program of Guangzhou, China (202201020429) and the “Five and five” project of the Third Affiliated Hospital of Sun Yat-Sen University (2023WW501).

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 1, and the codes used in this study are all common codes in Python packages mentioned in the “Methods” section of the manuscript.

Authors' Contributions

Authors ZD and CC had full access to all of the data in this study and take responsibility for the integrity of the data and the accuracy of the data analysis. ZD, ZH, and CC contributed to conceptualization and design. ZD, LZ, YZ, YL, and MG handled acquisition, analysis, or interpretation of data. ZD, CC, and WY managed drafting of the manuscript. All authors contributed to critical revision of the manuscript for important intellectual content. ZD, YL, and JY performed statistical analysis. ZH and CC obtained funding. JY and YL managed administrative, technical, or material support.

Address correspondence to CC (for access to raw data and statistical analysis): Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, chenchj28@mail.sysu.edu.cn; and ZH (regarding study design and funding acquisition): Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, heiziqing@sina.com; and WY (regarding the managed drafting and revision of the manuscript): Department of Anesthesiology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, yaowf3@mail.sysu.edu.cn.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional online content.
[DOCX File , 1000 KB - jmir_v27i1e55046_app1.docx]

References

1. Evered L, Silbert B, Knopman DS, Scott DA, DeKosky ST, Rasmussen LS, Nomenclature Consensus Working Group. Recommendations for the nomenclature of cognitive change associated with anaesthesia and surgery-2018. Anesthesiology 2018;129(5):872-879 [FREE Full text] [doi: 10.1097/ALN.0000000000002334] [Medline: 30325806]
2. Tasbihgou SR, Absalom AR. Postoperative neurocognitive disorders. Korean J Anesthesiol 2021;74(1):15-22 [FREE Full text] [doi: 10.4097/kja.20294] [Medline: 32623846]
3. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC London, England: American Psychiatric Association; 2013.

XSL•FO
RenderX

4. Song YX, Yang XD, Luo YG, Ouyang CL, Yu Y, Ma YL, et al. Comparison of logistic regression and machine learning methods for predicting postoperative delirium in elderly patients: a retrospective study. CNS Neurosci Ther 2023;29(1):158-167 [FREE Full text] [doi: 10.1111/cns.13991] [Medline: 36217732]
5. Marcantonio ER. Delirium in hospitalized older adults. N Engl J Med 2017;377(15):1456-1466 [FREE Full text] [doi: 10.1056/NEJMcp1605501] [Medline: 29020579]
6. Kong H, Xu LM, Wang DX. Perioperative neurocognitive disorders: a narrative review focusing on diagnosis, prevention, and treatment. CNS Neurosci Ther 2022;28(8):1147-1167 [FREE Full text] [doi: 10.1111/cns.13873] [Medline: 35652170]
7. Gleason LJ, Schmitt EM, Kosar CM, Tabloski P, Saczynski JS, Robinson T, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. JAMA Surg 2015;150(12):1134-1140 [FREE Full text] [doi: 10.1001/jamasurg.2015.2606] [Medline: 26352694]
8. Jin Z, Hu J, Ma D. Postoperative delirium: perioperative assessment, risk reduction, and management. Br J Anaesth 2020;125(4):492-504 [FREE Full text] [doi: 10.1016/j.bja.2020.06.063] [Medline: 32798069]
9. Mottaghi S, Nikoupour H, Firoozifar M, Jalali SS, Jamshidzadeh A, Vazin A, et al. The effect of taurine supplementation on delirium post liver transplantation: a randomized controlled trial. Clin Nutr 2022;41(10):2211-2218. [doi: 10.1016/j.clnu.2022.07.042] [Medline: 36081295]
10. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: advances in diagnosis and treatment. JAMA 2017;318(12):1161-1174 [FREE Full text] [doi: 10.1001/jama.2017.12067] [Medline: 28973626]
11. Zhou S, Deng F, Zhang J, Chen G. Incidence and risk factors for postoperative delirium after liver transplantation: a systematic review and meta-analysis. Eur Rev Med Pharmacol Sci 2021;25(8):3246-3253 [FREE Full text] [doi: 10.26355/eurrev_202104_25733] [Medline: 33928610]
12. Zhou J, Xu X, Liang Y, Zhang X, Tu H, Chu H. Risk factors of postoperative delirium after liver transplantation: a systematic review and meta-analysis. Minerva Anestesiol 2021;87(6):684-694 [FREE Full text] [doi: 10.23736/S0375-9393.21.15163-6] [Medline: 33594873]
13. Wang CM, Chen WC, Zhang Y, Lin S, He HF. Update on the mechanism and treatment of sevoflurane-induced postoperative cognitive dysfunction. Front Aging Neurosci 2021;13:702231 [FREE Full text] [doi: 10.3389/fnagi.2021.702231] [Medline: 34305576]
14. Arora A. Artificial intelligence: a new frontier for anaesthesiology training. Br J Anaesth 2020;125(5):e407-e408 [FREE Full text] [doi: 10.1016/j.bja.2020.06.049] [Medline: 32682557]
15. Mathis MR, Kheterpal S, Najarian K. Artificial intelligence for anesthesia: what the practicing clinician needs to know: more than black magic for the art of the dark. Anesthesiology 2018;129(4):619-622 [FREE Full text] [doi: 10.1097/ALN.0000000000002384] [Medline: 30080689]
16. Chen C, Chen B, Yang J, Li X, Peng X, Feng Y, et al. Development and validation of a practical machine learning model to predict sepsis after liver transplantation. Ann Med 2023;55(1):624-633 [FREE Full text] [doi: 10.1080/07853890.2023.2179104] [Medline: 36790357]
17. Tran J, Sharma D, Gotlieb N, Xu W, Bhat M. Application of machine learning in liver transplantation: a review. Hepatol Int 2022;16(3):495-508. [doi: 10.1007/s12072-021-10291-7] [Medline: 35020154]
18. Zhang Y, Yang D, Liu Z, Chen C, Ge M, Li X, et al. An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation. J Transl Med 2021;19(1):321 [FREE Full text] [doi: 10.1186/s12967-021-02990-4] [Medline: 34321016]
19. Chen C, Yang D, Gao S, Zhang Y, Chen L, Wang B, et al. Development and performance assessment of novel machine learning models to predict pneumonia after liver transplantation. Respir Res 2021;22(1):94 [FREE Full text] [doi: 10.1186/s12931-021-01690-3] [Medline: 33789673]
20. Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, et al. Applying machine learning in liver disease and transplantation: a comprehensive review. Hepatology 2020;71(3):1093-1105. [doi: 10.1002/hep.31103] [Medline: 31907954]
21. Lee BP, Vittinghoff E, Hsu C, Han H, Therapondos G, Fix OK, et al. Predicting low risk for sustained alcohol use after early liver transplant for acute alcoholic hepatitis: the sustained alcohol use post-liver transplant score. Hepatology 2019;69(4):1477-1487 [FREE Full text] [doi: 10.1002/hep.30478] [Medline: 30561766]
22. Ayllón MD, Ciria R, Cruz-Ramírez M, Pérez-Ortiz M, Gómez I, Valente R, et al. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. Liver Transpl 2018;24(2):192-203. [doi: 10.1002/lt.24870] [Medline: 28921876]
23. Zhang Y, Wan DH, Chen M, Li YL, Ying H, Yao GL, et al. Automated machine learning-based model for the prediction of delirium in patients after surgery for degenerative spinal disease. CNS Neurosci Ther 2023;29(1):282-295. [doi: 10.1111/cns.14002] [Medline: 36258311]
24. Zhang LM, Hornor MA, Robinson T, Rosenthal RA, Ko CY, Russell MM. Evaluation of postoperative functional health status decline among older adults. JAMA Surg 2020;155(10):950-958 [FREE Full text] [doi: 10.1001/jamasurg.2020.2853] [Medline: 32822459]
25. Hornor MA, Ma M, Zhou L, Cohen ME, Rosenthal RA, Russell MM, et al. Enhancing the American college of surgeons NSQIP surgical risk calculator to predict geriatric outcomes. J Am Coll Surg 2020;230(1):88-100.e1. [doi: 10.1016/j.jamcollsurg.2019.09.017] [Medline: 31672676]

XSL•FO
RenderX

26. Kuhn E, Du X, McGrath K, Coveney S, O'Regan N, Richardson S, et al. Validation of a consensus method for identifying delirium from hospital records. PLoS One 2014;9(11):e111823 [FREE Full text] [doi: 10.1371/journal.pone.0111823] [Medline: 25369057]
27. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European society of intensive care medicine. Intensive Care Med 1996;22(7):707-710. [doi: 10.1007/BF01709751] [Medline: 8844239]
28. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc 2010;72(4):417-473. [doi: 10.1111/j.1467-9868.2010.00740.x]
29. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
30. Yao L, Li Y, Yin R, Yang L, Ding N, Li B, et al. Incidence and influencing factors of post-intensive care cognitive impairment. Intensive Crit Care Nurs 2021;67:103106. [doi: 10.1016/j.iccn.2021.103106] [Medline: 34246526]
31. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023;10(1):1 [FREE Full text] [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
32. Garcia-Martinez R, Rovira A, Alonso J, Jacas C, Simón-Talero M, Chavarria L, et al. Hepatic encephalopathy is associated with posttransplant cognitive function and brain volume. Liver Transpl 2011;17(1):38-46 [FREE Full text] [doi: 10.1002/lt.22197] [Medline: 21254343]
33. Bajaj JS, Cordoba J, Mullen KD, Amodio P, Shawcross DL, Butterworth RF, International Society for Hepatic Encephalopathy Nitrogen Metabolism (ISHEN). Review article: the design of clinical trials in hepatic encephalopathy--an international society for hepatic encephalopathy and nitrogen metabolism (ISHEN) consensus statement. Aliment Pharmacol Ther 2011;33(7):739-747 [FREE Full text] [doi: 10.1111/j.1365-2036.2011.04590.x] [Medline: 21306407]
34. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet 2014;383(9920):911-922 [FREE Full text] [doi: 10.1016/S0140-6736(13)60688-1] [Medline: 23992774]
35. Wiredu K, Aduse-Poku E, Shaefi S, Gerber SA. Proteomics for the discovery of clinical delirium biomarkers: a systematic review of major studies. Anesth Analg 2023;136(3):422-432 [FREE Full text] [doi: 10.1213/ANE.0000000000006246] [Medline: 36580411]
36. Kahraman S, Zup S, McCarthy M, Fiskum G. GABAergic mechanism of propofol toxicity in immature neurons. J Neurosurg Anesthesiol 2008;20(4):233-240 [FREE Full text] [doi: 10.1097/ANA.0b013e31817ec34d] [Medline: 18812886]
37. Eyer F, Schuster T, Felgenhauer N, Pfab R, Strubel T, Saugel B, et al. Risk assessment of moderate to severe alcohol withdrawal--predictors for seizures and delirium tremens in the course of withdrawal. Alcohol Alcohol 2011;46(4):427-433. [doi: 10.1093/alcalc/agr053] [Medline: 21593124]
38. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22. [doi: 10.1016/j.jclinepi.2019.02.004] [Medline: 30763612]
39. Foley KA, Djaiani G. Update of the European society of anaesthesiology and intensive care medicine evidence-based and consensus-based guideline on postoperative delirium in adult patients. Eur J Anaesthesiol 2025;42(1):86-87. [doi: 10.1097/EJA.0000000000002043] [Medline: 39628423]
40. Hughes CG, Boncyk CS, Culley DJ, Fleisher LA, Leung JM, McDonagh DL, Perioperative Quality Initiative (POQI) 6 Workgroup. American society for enhanced recovery and perioperative quality initiative joint consensus statement on postoperative delirium prevention. Anesth Analg 2020;130(6):1572-1590 [FREE Full text] [doi: 10.1213/ANE.0000000000004641] [Medline: 32022748]
41. Concato J, Corrigan-Curay J. Real-world evidence - where are we now? N Engl J Med 2022;386(18):1680-1682. [doi: 10.1056/NEJMp2200089] [Medline: 35485775]

Abbreviations

AST: aspartate aminotransferase
AUC: area under the receiver operating curve
CAM: Confusion Assessment Method
CHE: cover hepatic encephalopathy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
EBL: estimated blood loss
eGFR: estimated glomerular filtration rate
EPR: electronic patient record
hsCRP: hypersensitive C-reactive protein
ICU: intensive care unit
LASSO: least absolute shrinkage and selection operator
LGB: light gradient boosting machine

XSL•FO
RenderX

LR: logistic regression
LT: liver transplantation
MELD: model for end-stage liver disease
MIMIC-Ⅳ: Medical Information Mart for Intensive Care Ⅳ
ML: machine learning
MLP: multilayer perceptron classifier
PLT: platelet
PND: perioperative neurocognitive disorder
POD: postoperative delirium
PSDP: perioperative specialist database platform
PT: prothrombin time
RF: random forest classifier
SHAP: Shapley additive explanations
SOFA: sequential organ failure assessment
SVM: support vector machine
WBC: white blood cell
XGB: extreme gradient boosting with classification trees

Edited by T de Azevedo Cardoso; submitted 30.11.23; peer-reviewed by K Xie, V Chauhan; comments to author 07.02.24; revised version received 12.04.24; accepted 30.10.24; published 15.01.25.

Please cite as:
Ding Z, Zhang L, Zhang Y, Yang J, Luo Y, Ge M, Yao W, Hei Z, Chen C
A Supervised Explainable Machine Learning Model for Perioperative Neurocognitive Disorder in Liver-Transplantation Patients and External Validation on the Medical Information Mart for Intensive Care IV Database: Retrospective Study
J Med Internet Res 2025;27:e55046
URL: https://www.jmir.org/2025/1/e55046
doi:10.2196/55046
PMID:39813086

©Zhendong Ding, Linan Zhang, Yihan Zhang, Jing Yang, Yuheng Luo, Mian Ge, Weifeng Yao, Ziqing Hei, Chaojin Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Ten Machine Learning Models for Predicting Preoperative and Postoperative Coagulopathy in Patients With Trauma: Multicenter Cohort Study

Xiaojuan Xiong[1*], PhD; Hong Fu[2*], Prof Dr Med; Bo Xu[3*], Prof Dr Med; Wang Wei[4*], Prof Dr Med; Mi Zhou[1], MD; Peng Hu[5], MD; Yunqin Ren[1], MD; Qingxiang Mao[1], Prof Dr Med

[1]Department of Anesthesiology, Daping Hospital, Army Medical University, Chongqing, China

[2]Department of Anesthesiology, Chongqing Emergency Medical Center, Chongqing University Central Hospital, School of Medicine, Chongqing University, Chongqing, China

[3]Department of Anesthesiology, General Hospital of Southern Theater Command of PLA, Guangzhou, China

[4]Department of Anesthesiology, The PLA Rocket Force Characteristic Medical Center, Beijing, China

[5]School of Public Policy and Administration, Chongqing University, Chongqing, China

[*]these authors contributed equally

Corresponding Author:
Qingxiang Mao, Prof Dr Med
Department of Anesthesiology
Daping Hospital
Army Medical University
Yuzhong District
10 Changjiang Zhilu
Chongqing
China
Phone: 86 68729729
Email: qxmao@tmmu.edu.cn

Abstract

Background: Recent research has revealed the potential value of machine learning (ML) models in improving prognostic prediction for patients with trauma. ML can enhance predictions and identify which factors contribute the most to posttraumatic mortality. However, no studies have explored the risk factors, complications, and risk prediction of preoperative and postoperative traumatic coagulopathy (PPTIC) in patients with trauma.

Objective: This study aims to help clinicians implement timely and appropriate interventions to reduce the incidence of PPTIC and related complications, thereby lowering in-hospital mortality and disability rates for patients with trauma.

Methods: We analyzed data from 13,235 patients with trauma from 4 medical centers, including medical histories, laboratory results, and hospitalization complications. We developed 10 ML models in Python (Python Software Foundation) to predict PPTIC based on preoperative indicators. Data from 10,023 Medical Information Mart for Intensive Care patients were divided into training (70%) and test (30%) sets, with 3212 patients from 3 other centers used for external validation. Model performance was assessed with 5-fold cross-validation, bootstrapping, Brier score, and Shapley additive explanation values.

Results: Univariate logistic regression identified PPTIC risk factors as (1) prolonged activated partial thromboplastin time, prothrombin time, and international normalized ratio; (2) decreased levels of hemoglobin, hematocrit, red blood cells, calcium, and sodium; (3) lower admission diastolic blood pressure; (4) elevated alanine aminotransferase and aspartate aminotransferase levels; (5) admission heart rate; and (6) emergency surgery and perioperative transfusion. Multivariate logistic regression revealed that patients with PPTIC faced significantly higher risks of sepsis (1.75-fold), heart failure (1.5-fold), delirium (3.08-fold), abnormal coagulation (3.57-fold), tracheostomy (2.76-fold), mortality (2.19-fold), and urinary tract infection (1.95-fold), along with longer hospital and intensive care unit stays. Random forest was the most effective ML model for predicting PPTIC, achieving an area under the receiver operating characteristic of 0.91, an area under the precision-recall curve of 0.89, accuracy of 0.84, sensitivity of 0.80, specificity of 0.88, precision of 0.88, F_1-score of 0.84, and Brier score of 0.13 in external validation.

XSL•FO
RenderX

Conclusions: Key PPTIC risk factors include (1) prolonged activated partial thromboplastin time, prothrombin time, and international normalized ratio; (2) low levels of hemoglobin, hematocrit, red blood cells, calcium, and sodium; (3) low diastolic blood pressure; (4) elevated alanine aminotransferase and aspartate aminotransferase levels; (5) admission heart rate; and (6) the need for emergency surgery and transfusion. PPTIC is associated with severe complications and extended hospital stays. Among the ML models, the random forest model was the most effective predictor.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300078097; https://www.chictr.org.cn/showproj.html?proj=211051

(J Med Internet Res 2025;27:e66612) doi:10.2196/66612

KEYWORDS

traumatic coagulopathy; preoperative; postoperative; machine learning models; random forest; Medical Information Mart for Intensive Care

Introduction

Severe trauma is a significant global public health burden and a leading cause of reduced life expectancy. The Global Burden of Diseases, Injury, and Risk Factors Study 2017 estimates that trauma accounts for 8% of total deaths annually [1]. Traumatic coagulopathy (TIC) is a systemic inflammatory state accompanied by coagulation dysfunction, acidosis, and hypothermia that occurs after traumatic injury [2]. TIC is observed in one-quarter to one-third of patients with trauma [3,4] and is associated with increased rates of massive transfusion and multiple organ failure (MOF), prolonged intensive care unit (ICU) stay and length of stay (LOS), and a 4-fold increase in mortality [3]. Harhangi et al [5] reported that TIC is linked to a 9-fold increase in mortality and a 36-fold increase in the likelihood of adverse outcomes [5]. Early endogenous TIC is associated with increased bleeding, higher red blood cell (RBC) transfusion rates, and a greater risk of secondary complications like multiple organ dysfunction syndrome (MODS) and thromboembolism [6]. Cohen and Christie [7] also reported that patients with TIC have worse prognoses, including higher rates of infection, thromboembolism, acute lung injury, MOF, and mortality. Without proper and timely diagnosis and treatment, posttraumatic bleeding and associated TIC remain potentially preventable causes of MOF [8]. Therefore, the latest European guidelines indicate that the immediate identification and management of TIC can improve outcomes for severely injured patients [9].

In recent years, artificial intelligence (AI) has received significant attention for its potential use in various aspects of human activities, including health care [10]. Sidey-Gibbons and Sidey-Gibbons [11] and Liu and Salinas [12] suggested that machine learning (ML) excels over traditional methods in handling large, unstructured, nonlinear, or incomplete datasets, improving outcome predictions for patients with trauma. With the advent of ML and improved computational analysis methods, recent approaches have aimed to enhance predictions and identify which factors are most important in contributing to posttraumatic mortality [13]. Rashidi et al [14] emphasized the critical importance of rapid diagnosis and intervention for managing hemostatic disorders, which can lead to life-threatening bleeding or clotting. AI or ML models can significantly enhance turnaround times, accuracy, and the timeliness of diagnoses and interventions, potentially improving patient care and outcomes.

Our data show that despite surgical interventions such as hemostasis, vascular reconstruction, fracture reduction, organ repair, blood transfusion, and the correction of electrolyte imbalances in patients who experienced preoperative TIC, 50% of these patients still had TIC postoperatively [15]. These patients were more prone to develop complications during hospitalization, and preoperative and postoperative traumatic coagulopathy (PPTIC) was identified as an independent risk factor for these complications. Therefore, we plan to explore the risk factors for PPTIC in patients with trauma and its relationship with in-hospital complications. We will use admission indicators from patients with trauma to predict the risk of PPTIC, applying 10 different ML models to estimate the probability of PPTIC occurrence and conducting a comprehensive evaluation of each model. This predictive approach can aid clinicians in implementing timely and appropriate interventions to reduce the incidence of PPTIC and related complications, thereby lowering in-hospital mortality and disability rates for these patients.

Methods

Ethical Considerations

Data from 4 medical centers were anonymized and integrated. Ethical approvals were granted by the respective ethics committees: Daping Hospital, Medical Research Ethics (2023) 261; Chongqing Emergency Medical Center, 2023 ethical review (48); the People's Liberation Army Rocket Force Characteristic Medical Center (KY 2023037); and the General Hospital of Southern Theatre Command of People's Liberation Army (NZLLKZ2024021). As this was a retrospective study, patient informed consent was waived by the ethics committees. It was registered with the Chinese Clinical Trial Registry (ChiCTR2300078097) and adhered to the Declaration of Helsinki. This study adheres to 2 key guidelines to ensure rigorous reporting: the TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence) guidelines for the development and reporting of multivariable ML predictive models in biomedical research (Multimedia Appendix 1) [16] and the STROCSS (Strengthening the Reporting of Cohort Studies in Surgery) criteria, which standardize the reporting of surgical observational studies [17].

XSL•FO
RenderX

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria are as follows:

- Inclusion criteria: Patients who underwent surgical treatment after trauma from January 2017 to February 2023.
- Exclusion criteria: Patients without trauma (snake bites, bee stings, and burns); patients who did not undergo surgery or stopped treatment; patients younger than 18 years of age; patients with hematological diseases; patients with a history of liver cirrhosis; pregnant women; and patients without coagulation-related tests results either preoperatively or postoperatively.

Data Collection

We retrieved patients who underwent trauma or concurrent surgeries using the MIMIC IV (Medical Information Mart for Intensive Care) 2.2 database, with the detailed data extraction process outlined in Multimedia Appendix 2. We then collected the necessary indicators using subject_id and hadm_id, which included the following information: ID number; sex; age; height; weight; type of surgery (elective or emergency); and preoperative comorbidities such as high blood pressure, diabetes, cerebrovascular diseases, chronic obstructive pulmonary disease, chronic bronchitis, asthma, and chronic renal insufficiency. Preoperative indicators included systolic blood pressure, heart rate (HR), saturation of pulse oximetry, and the presence of shock upon admission. The laboratory examination indicators (latest results prior to admission and first results postoperatively) included the following: complete blood count (white blood cell, lymphocyte [%], monocyte [%], neutrophil [%], eosinophil [%], basophil [%], RBC, mean corpuscular hemoglobin concentration, mean corpuscular volume, red cell distribution width, hemoglobin, hematocrit, platelet count [PLT], mean platelet volume, renal function [urea, creatinine, and blood uric acid]), liver function (albumin, alkaline phosphatase, lactate dehydrogenase, gamma-glutamyl transferase, total protein, globulin, direct bilirubin, indirect bilirubin, total bilirubin, alanine aminotransferase [ALT], and aspartate aminotransferase [AST]), electrolytes (sodium, potassium, and calcium), and coagulation function (activated partial thromboplastin time [APTT], prothrombin time [PT], international normalized ratio [INR], and D-dimer). The relevant intraoperative indicators included the need for perioperative blood transfusion. Complications during hospitalization included pulmonary embolism, sepsis, septicemia, heart failure, delirium, pulmonary edema, abnormal coagulation, endotracheal tube intubation, invasive mechanical ventilation, tracheostomy, end-stage renal disease, systemic inflammatory response, acute respiratory distress syndrome, pleural effusion, cardiac arrest, anemia, venous thrombosis, renal failure, acute kidney failure, urinary tract infection, respiratory failure, pneumonia, LOS, in-hospital death, and length of ICU stay. Data from the other 3 medical centers were collected via electronic medical records, consistent with the previously mentioned information. Our definition criteria for TIC were as follows: PLT<100×10^9/L; INR>1.25; PT>14 second; and APTT>36 second [18]. PPTIC was defined as patients who experienced TIC both within 24 hours before surgery (preoperative) and within 24 hours after surgery (postoperative), meeting one or more of the above TIC diagnostic criteria in both timeframes.

Process of ML Modeling

Handling Missing Data

We processed the data using Python (version 3.9; Python Software Foundation), discarding variables with over 40% missing data (Multimedia Appendix 3). For binary and categorical variables, we applied 1-hot encoding. Continuous variables were imputed using k-nearest neighbor (KNN) with k=5, while categorical variables underwent multiple imputations with 5 iterations.

Data Normalization and Feature Selection

Data Normalization

We created a *StandardScaler* object to standardize the data. Both the training and test data were standardized via the same procedure.

Outlier Handling

We calculated z scores for the training data, considering values above 3 as outliers, and replaced these with the mean using *np.where*.

Feature Selection

We assessed continuous variables with the Pearson correlation coefficient and converted categorical variables to numerical ones via label encoding. Then, the Pearson correlation coefficient was calculated for each feature relative to the target variable, and the top 50 features with the highest correlations were selected. These selected features were then refined using least absolute shrinkage and selection operator regression, resulting in a final set of 25 features.

Dataset Splitting

We modeled data from 10,023 patients with trauma in MIMIC IV, splitting it into a training set of 7631 cases and a test set of 1908 cases (7:3 ratio) using the *train_test_split* function. Additionally, we used data from 3212 patients from 3 other medical centers for external validation (Figure 1).

XSL•FO
RenderX

Figure 1. The data processing. MIMIC: Medical Information Mart for Intensive Care; ML: machine learning; PLA: People's Liberation Army.

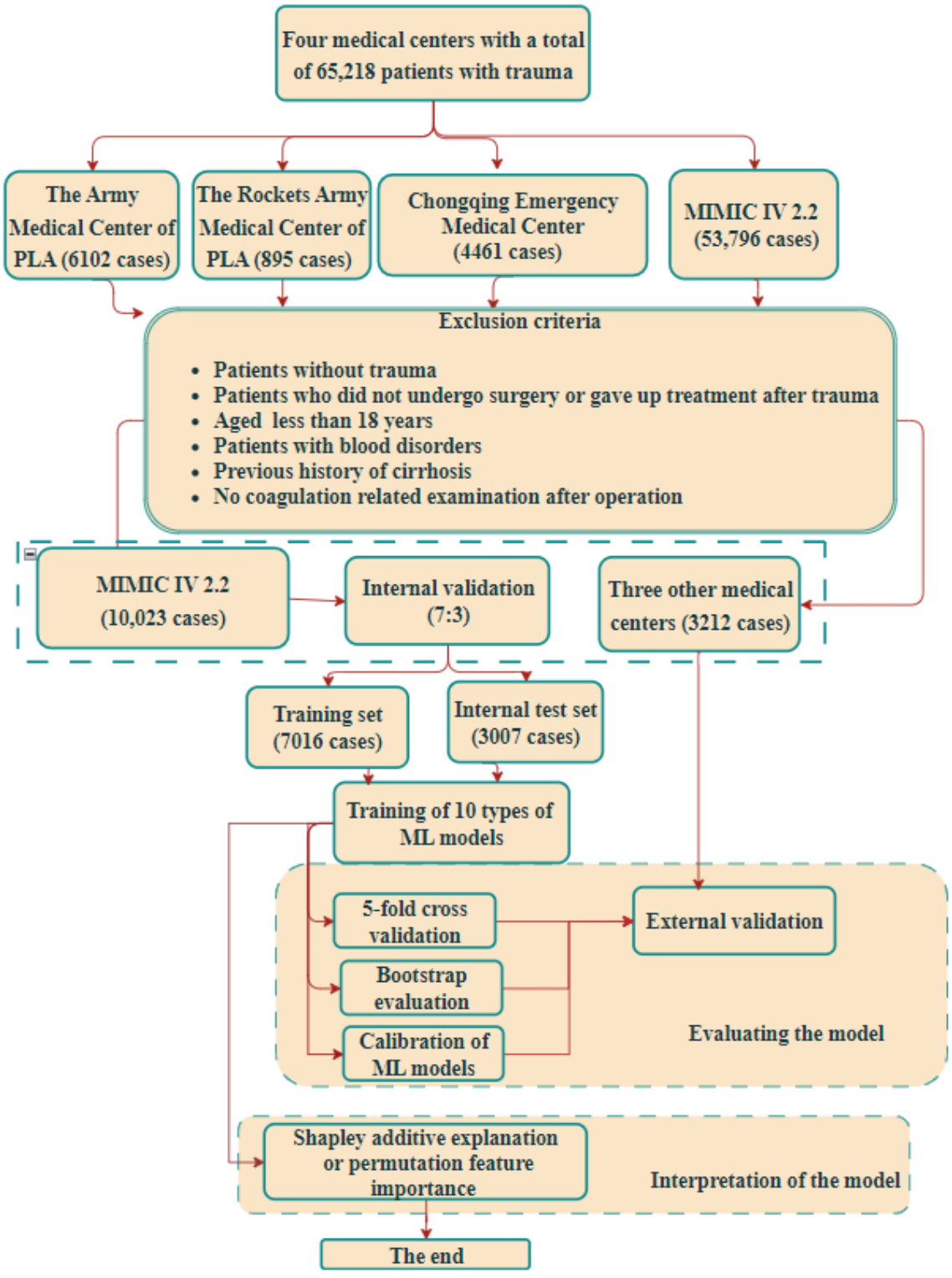

Building the ML Models

We used Python 3.9 to build the ML models, using a total of 10 models (Multimedia Appendix 4): logistic regression (LR), random forest (RF), support vector machine (SVM), decision tree (DT), KNN, gradient boosting (GB), neural networks (NN), naive Bayes (NB), AdaBoost, and extreme gradient boosting (XGBoost). We assessed the models' performances via 7 metrics: area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), accuracy, sensitivity, specificity, precision, and F_1-score. The Python 3.9 packages used for data processing and analysis are shown in Multimedia Appendix 5. Additionally, the full code and data processing steps, along with detailed instructions, have been made publicly available via GitHub [19].

XSL•FO
RenderX

Model Performance and Evaluation

In this study, we used "grid search" for hyperparameter optimization. We conducted 5-fold cross-validation by dividing the dataset into 5 subsets, using each subset once for testing and the remaining four for training. This process, repeated 5 times, helps reduce overfitting, stabilize performance evaluation, and address sample selection bias [20]. We measured accuracy, precision, recall, and F_1-score.

To evaluate the robustness and generalization of 10 ML models, we used bootstrap sampling with 1000 iterations. Each iteration involved retraining the models on a bootstrap sample and calculating accuracy, precision, recall, and F_1-score. We then estimated the distributions and 95% CIs for these metrics, producing 12 key indicators: accuracy mean, accuracy lower or upper 95% CI, precision mean, precision lower or upper 95% CI, recall mean, recall lower or upper 95% CI, F_1-score mean, and F_1-score lower or upper 95% CI.

To assess model calibration, we used the Brier score (BS), which ranges from 0 to 1, with lower scores indicating better accuracy of predicted probabilities. We trained the model, evaluated predictions on the test set, computed the BS, and plotted a calibration curve to visualize the model's performance.

Interpretation of the ML

We use Shapley additive explanation (SHAP) values to assess the significance of variables in the model by evaluating their impact on predictions across all feature subsets and averaging these effects [21]. We used the Python package *SHAP* to visualize and rank the top 15 variables with the highest SHA*P* values related to PPTIC. These variables were integrated into the model, and SHA*P* values ranging from 0 to 1 were used for risk interpretation. We also conducted univariate logistic regression on these top 15 variables.

Statistics

Statistical analysis was performed using SPSS Statistics (version 27.0; IBM Corp). Patients were divided into PPTIC and non-PPTIC groups, with PPTIC as the independent variable and other factors as dependent variables. Categorical data are presented as percentages, and continuous data as mean (SD). Depending on the distribution, continuous variables were assessed via 2-tailed *t* tests or Mann-Whitney *U* tests, and categorical variables via chi-square tests or exact tests. Univariate analysis identified risk factors for persistent trauma-induced coagulopathy, and multivariable logistic regression analyzed complications during hospitalization in the PPTIC group. A *P* value less than .05 was considered statistically significant.

Results

Basic Information of the Patients

We ultimately included 13,237 patients with trauma, among whom 6586 (49.8%) experienced PPTIC. The mean age of our patients was 62.45 (SD 17.85) years. In the PPTIC group, the average age was 63.6 (SD 17.4) years, whereas in the non-PPTIC group, it was 61.3 (SD 18.2) years, with *P*<.001 (Table 1). Among the patients, there were 5590 (42.2%) male individuals and 7646 (57.8%) female individuals, and this difference was significant (*P*<.001). The racial distribution included 3452 (26.1%) Asian individuals, 7402 (55.9%) White individuals, 342 (2.6%) Black individuals, 940 (7.1%) Latino individuals, and 1101(8.3%) individuals of other races.

XSL•FO
RenderX

Table 1. The basic information of the patients.

	PPTIC[a] group	Non-PPTIC group	*P* value[b]
Age (years), mean (SD)	63.6 (17.4)	61.3 (18.2)	<.001
Weight (kg), mean (SD)	74.9 (23.8)	71.8 (20)	.002
BMI (kg/m^2), mean (SD)	27.9 (8)	27.3 (7.4)	.001
Admission HR[c], mean (SD)	86.3 (19.3)	84.6 (17)	<.001
Admission SBP[d] (mmHg), mean (SD)	125 (22.1)	131.3 (131.4)	.001
Pre-hematocrit (%), mean (SD)	31.9 (6.4)	34.5 (6.5)	<.001
Pre-hemoglobin (g/L), mean (SD)	104.1 (22)	113.7 (22.6)	<.001
Pre-RBC[e] (×10×12/L), mean (SD)	3.5 (0.8)	3.8 (0.7)	<.001
Pre-ALT[f] (U/L), mean (SD)	69.6 (258.2)	44.4 (86)	<.001
Pre-AST[g] (U/L), mean (SD)	87.2 (305.3)	48.4 (101.2)	<.001
Pre-GGT[h] (U/L), mean (SD)	54.1 (104)	46.4 (95.7)	.031
Pre-calcium (mmol/L), mean (SD)	2.1 (0.2)	2.2 (0.2)	<.001
Pre-sodium (mmol/L) mean (SD)	0.0552	0.0476	<.001
Pre-PT[i] (s), mean (SD)	18.3 (7.9)	12.8 (1.8)	<.001
Pre-APTT[j] (s), mean (SD)	40.3 (13.5)	30.8 (6.7)	<.001
Pre-PLT[k] (×10×9/L), mean (SD)	212.5 (124.3)	232.8 (103)	<.001
Pre-INR[l], mean (SD)	1.6 (0.7)	1.1 (0.2)	<.001
Post-INR, mean (SD)	1.6 (0.6)	1.1 (0.1)	<.001
Post-PT (seconds), mean (SD)	17.5 (6.7)	12.9 (1.5)	<.001

[a]PPTIC: preoperative and postoperative traumatic coagulopathy.
[b]*P* value less than .05 was considered statistically significant.
[c]HR: heart rate.
[d]SBP: systolic pressure.
[e]RBC: red blood cell
[f]ALT: alanine aminotransferase.
[g]AST: aspartate amino transferase.
[h]GGT: glutamyl transferase
[i]PT: prothrombin time
[j]APTT: activated partial thromboplastin time.
[k]ALT: alanine aminotransferase.
[l]INR: international normalized ratio.

Results of the ML Models

Training Results of the ML Models

We developed 10 ML risk models for predicting PPTIC. In the training set, all 10 models achieved an AUROC of 0.83 or higher, with RF, DT, and XGBoost reaching AUROC values of 1. In the internal testing set, RF achieved an excellent AUROC of 0.92, whereas the AUROC of SVM was 0.9, that of GB was 0.93, that of NN was 0.9, that of AdaBoost was 0.91, and XGBoost performed well (Figure 2). RF, GB, and XGBoost demonstrated superior performance in both the training and testing sets, with AUROC, accuracy, sensitivity, specificity, precision, and F_1-score all above 0.8 (Multimedia Appendix 6).

XSL•FO
RenderX

Figure 2. The AUROC curves of 10 machine learning models in the (A) training, (B) test, and (C) external validation set for predicting PPTIC in patients with trauma. The ROC curve for predicting PPTIC in patients with trauma is defined by the false positive rate on the x-axis, representing the proportion of negative cases misclassified as positive, and the true positive rate on the y-axis, reflecting the proportion of correctly identified positive cases (sensitivity), with both axes ranging from 0 to 1. The AUROC is a key metric for assessing the model's predictive performance, where a higher AUROC signifies superior discriminatory ability in predicting PPTIC and improved generalizability. AUROC: area under the receiver operating characteristic curve; PPTIC: preoperative and postoperative traumatic coagulopathy; ROC: receiver operating characteristic curve; SVM: support vector machine.

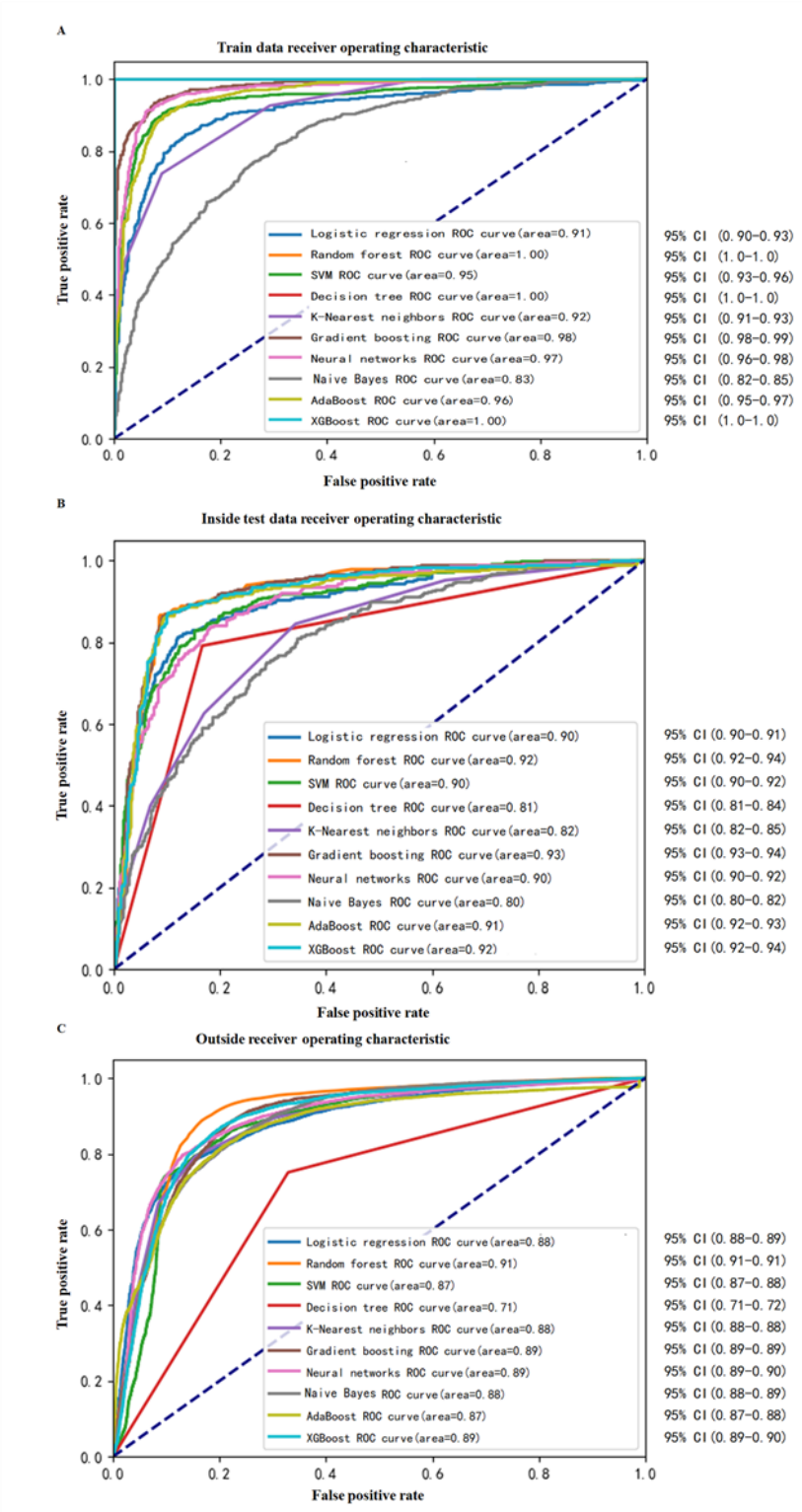

XSL•FO
RenderX

Evaluating the ML Models

Results of the 5-Fold Cross-Validation

The results from the 5-fold cross-validation (Multimedia Appendix 7) showed that our RF, GB, AdaBoost, and XGBoost models consistently performed well across diverse subsets, demonstrating high accuracy and consistency. The SDs of the accuracy, precision, recall, and F_1-scores were minimal, confirming the robustness and reliability of these ML models.

Bootstrap Results

The bootstrap analysis results indicated that RF, DT, GB, AdaBoost, and XGBoost achieved average accuracy, precision, recall, and F_1-score all exceeding 0.91, accompanied by narrow 95% CIs (Table 2). This suggests consistently high accuracy across various sample resampling conditions, underscoring the robustness of these models. Moreover, RF, DT, GB, AdaBoost, and XGBoost demonstrated substantial stability and reliability across their respective performance metrics.

Table 2. The bootstrap results of 10 kinds of machine learning models.

	Mean accuracy	95% CI accuracy	Mean precision	95% CI precision	Mean recall	95% CI recall	Mean F_1-score	95% CI F_1-score
LR[a]	0.766	0.713-0.821	0.766	0.712-0.821	0.766	0.713-0.821	0.763	0.709-0.82
RF[b,c]	*0.961*	*0.956-0.966*	*0.961*	*0.956-0.966*	*0.961*	*0.956-0.966*	*0.961*	*0.956-0.966*
SVM[d]	0.573	0.572-0.573	0.503	0.328-0.755	0.573	0.572-0.573	0.419	0.417-0.421
DT[e]	0.934	0.926-0.941	0.935	0.926-0.941	0.934	0.926-0.941	0.934	0.926-0.941
KNN[f]	0.661	0.651-0.673	0.657	0.647-0.669	0.661	0.651-0.673	0.656	0.645-0.668
GB[g]	*0.918*	*0.914-0.923*	*0.918*	*0.914-0.923*	*0.918*	*0.914-0.923*	*0.918*	*0.914-0.923*
NN[h]	0.633	0.49-0.742	0.697	0.625-0.746	0.633	0.49-0.742	0.594	0.399-0.743
NB[i]	0.728	0.689-0.75	0.74	0.713-0.759	0.728	0.689-0.75	0.724	0.685-0.749
AdaBoost	0.892	0.885-0.899	0.892	0.885-0.898	0.892	0.885-0.899	0.891	0.884-0.898
XGBoost[j]	*0.96*	*0.954-0.964*	*0.96*	*0.954-0.964*	*0.96*	*0.954-0.964*	*0.96*	*0.954-0.964*

[a]LR: logistic regression.

[b]RF: random forest.

[c]Italics are used to highlight the better-performing machine learning models in our study.

[d]SVM: support vector machine

[e]DT: decision tree.

[f]KNN: k-nearest neighbor.

[g]GB: gradient boosting.

[h]NN: neural networks.

[i]NB: naive Bayes.

[j]XGBoost: extreme gradient boosting.

Calibration and Precision-Recall of the ML Models

In the training set, RF, DT, GB, and XGBoost achieved BSs of 0.01, 0.00, 0.05, and 0.00, respectively, with precision-recall values above 0.9, demonstrating robustness. In the internal testing set, RF, GB, and XGBoost had BSs of 0.1, 0.09, and 0.10, and precision-recall values of 0.87, 0.87, and 0.86. During external validation, they recorded BSs of 0.13, 0.13, and 0.15, with precision-recall values around 0.88, 0.84, and 0.87. The BS and precision-recall values of RF, GB, and XGBoost in the external validation set closely mirrored those in the internal testing set (Figure 3), underscoring the consistency and reliability of these models.

XSL•FO
RenderX

Figure 3. The precision-recall curve and train data calibration curve of 10 machine learning models in (A-B) train, (C-D) internal test, and (E-F) external validation sets predicted patients with PPTIC. PPTIC: preoperative and postoperative traumatic coagulopathy; SVM: support vector machine.

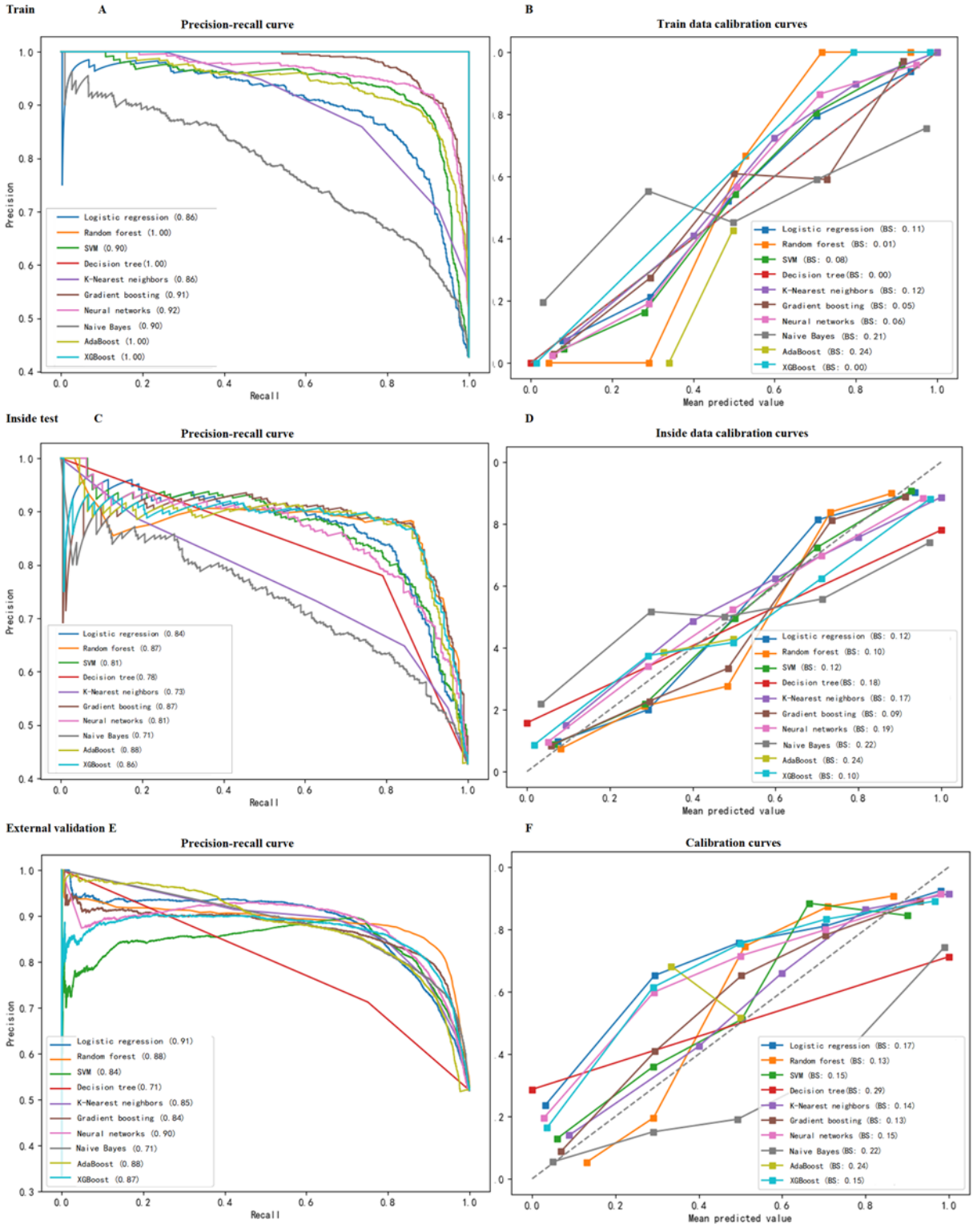

Results of the External Validation of the ML Models

In the external validation set, RF, GB, NN, and XGBoost achieved AUROC values of 0.91, 0.89, 0.89, and 0.89, respectively (Figure 2), while other models attained approximately 0.87. RF showed the highest performance with AUROC 0.91 (95% CI 0.90-0.913), AUPRC 0.89 (95% CI 0.88-0.90), accuracy 0.84 (95% CI 0.83-0.85), sensitivity 0.80 (95% CI 0.79-0.81), specificity 0.88 (95% CI 0.87-0.89), precision 0.88 (95% CI 0.87-0.89), and F_1-score 0.84 (95% CI 0.83-0.84). Despite being slightly lower than in the training set, RF's performance remained stable, demonstrating strong generalizability. Thus, RF was identified as the optimal model based on 5-fold cross-validation, bootstrap, calibration, precision-recall, and external validation results.

XSL•FO
RenderX

Interpretation of the ML Models

Analysis of SHAP values highlighted the key features influencing model predictions. For RF, GB, and XGBoost (Figure 4), as well as the other 7 ML models (Multimedia Appendix 8), we observed the top 15 variables ranked by their average impact. In RF, the primary predictors of PPTIC in patients with trauma were preoperative APTT, PT, INR, AST, ALT, calcium, hemoglobin, sodium, hematocrit, admission SBP, RBC, admission HR, and the need for emergency surgery or perioperative blood transfusion (Table 3). Univariate logistic regression identified prolonged APTT, PT, and INR; low levels of hemoglobin, hematocrit, RBC, calcium, and sodium; elevated ALT and AST; and the need for emergency surgery or blood transfusion were significant risk factors for PPTIC.

Figure 4. SHAP values of (A) GB, (B) XGBoost, and (C) RF for predicting the risk of TIC before and after surgery in patients with trauma. We selected the three best-performing ML models (GB, XGBoost, and RF) from a set of 10 to predict PPTIC outcomes and evaluated their SHAP values. The SHAP plot highlights the top 15 features with the most significant impact on model predictions, ranked in descending order of importance from top to bottom. The x-axis represents the SHAP value, while the y-axis lists the corresponding feature names. Red points represent higher feature values, while blue points represent lower feature values. If red points are skewed toward the positive direction (higher SHAP values), it indicates that higher feature values contribute positively to the prediction outcome (increasing the predicted value). Conversely, if blue points are skewed toward the negative direction (lower SHAP values), it suggests that lower feature values contribute negatively to the prediction outcome (decreasing the predicted value). ALT: alanine aminotransferase; APTT: activated partial thromboplastin time; AST: aspartate aminotransferase; DBP: diastolic blood pressure; GB: gradient boosting; HR: heart rate; PPTIC: preoperative and postoperative traumatic coagulopathy; PT: prothrombin time; RBC: red blood cell; RF: random forest; SBP: systolic pressure; SHAP: Shapley additive explanation; TIC: traumatic coagulopathy.

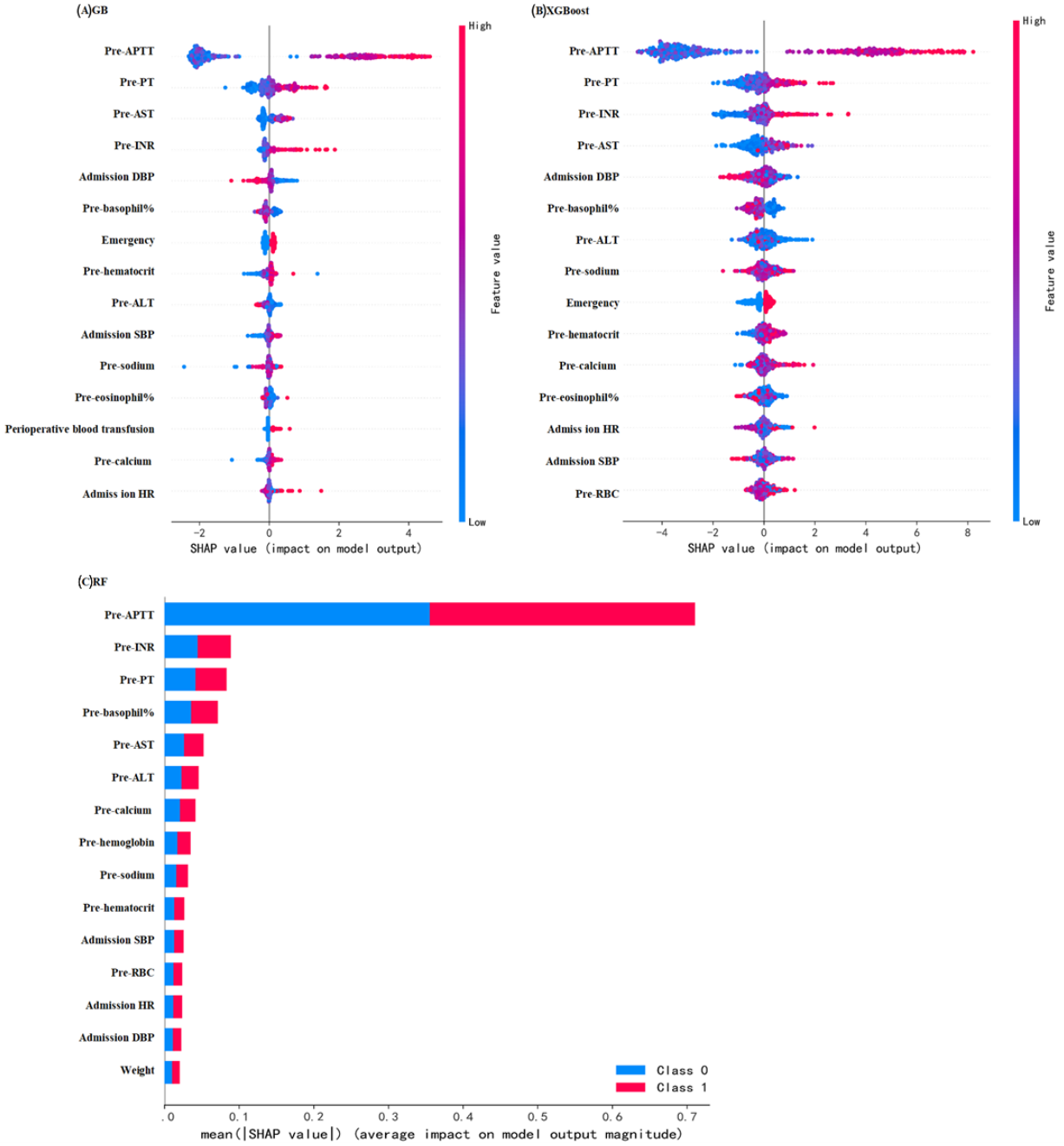

XSL•FO
RenderX

Table 3. Univariate logistic regression of risk factors for preoperative and postoperative TIC[a] in patients with trauma.

	OR[b] (95% CI)	*P* value[c]
Pre-APTT[d] (s)	1.14 (1.13-1.14)	<.001
Pre-PT[e] (s)	1.99 (1.92-2.05)	<.001
Pre-INR[f]	1289.15 (936.81-1773.99)	<.001
Pre-hemoglobin (g/L)	0.98 (0.98-0.98)	<.001
Pre-hematocrit (%)	0.94 (0.94-0.95)	<.001
Pre-RBC[g] (×10×12/L)	0.61 (0.57-0.64)	<.001
Pre-calcium (mmol/L)	0.4 (0.33-0.49)	<.001
Pre-sodium (mmol/L)	0.96 (0.95-0.97)	<.001
Admission SBP[h] (mmHg)	0.99 (0.99-0.99)	<.001
Admission HR[i]	1.004 (1.004-1.006)	.005
Emergency	1.34 (1.24-1.45)	<.001
Pre-ALT[j] (U/L)	1.001 (1.001-1.002)	<.001
Pre-AST[k] (U/L)	1.001 (1.001-1.002)	<.001
Perioperative blood transfusion	1.42 (1.12-1.81)	.004

[a]TIC: traumatic coagulopathy.
[b]OR: odds ratio.
[c]*P* value less than .05 was considered statistically significant.
[d]APTT: activated partial thromboplastin time.
[e]PT: prothrombin time.
[f]INR: international normalized ratio.
[g]RBC: red blood cell.
[h]SBP: systolic pressure.
[i]HR: heart rate.
[j]ALT: alanine aminotransferase.
[k]AST: aspartate amino transferase.

Complications During Hospitalization in Patients With PPTIC

Multivariate logistic regression analysis showed that compared with patients with non-PPTIC, those with PPTIC had significantly higher postoperative risks of the following complications: sepsis (1.75-fold, 95% CI 1.18-2.6; *P*=.006), heart failure (1.5-fold, 95% CI 1.24-1.82; *P*<.001), delirium (3.08-fold, 95% CI 1.62-5.83; *P*=.001), abnormal coagulation (3.57-fold, 95% CI 2.26-5.63; *P*<.001), tracheostomy (2.76-fold, 95% CI 2.02-3.77, *P*=.001), in-hospital mortality (2.19-fold, 95% CI 1.84-2.61; *P*<.001), urinary tract infection (1.95-fold, 95% CI 1.46-2.59; *P*<.001), and longer LOS and ICU stay (Figure 5).

XSL•FO
RenderX

Figure 5. Multivariate logistic regression of postoperative complications in patients with trauma who developed TIC both before and after surgery. LOS: length of stay; ICU: intensive care unit; OR: odds ratio; TIC: traumatic coagulopathy. *P* value .05 was considered statistically significant.

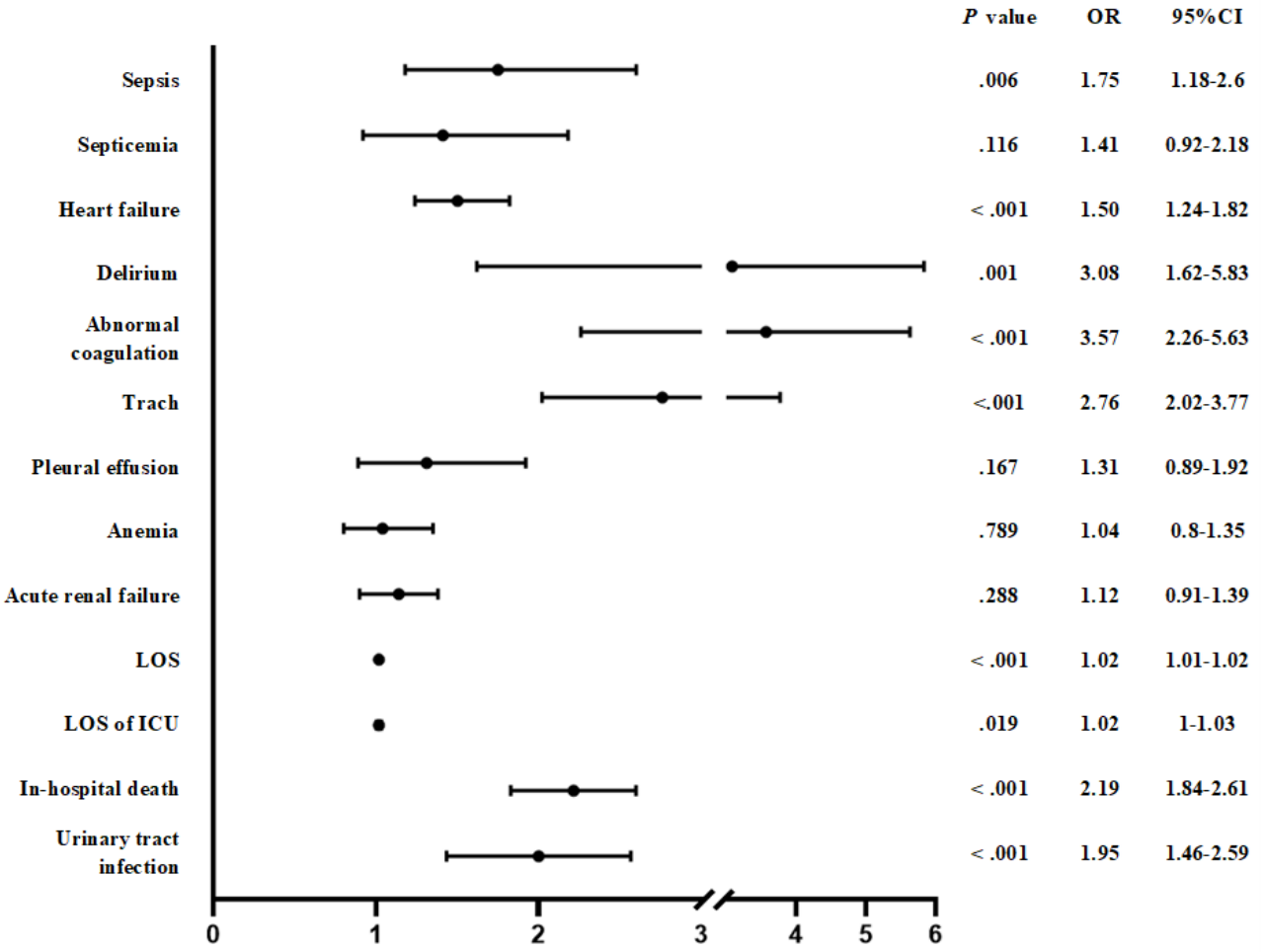

Discussion

ML Models for Predicting PPTIC in Patients With Trauma

Our study found that RF, GB, and XGBoost had AUROC values above 0.89 for predicting PPTIC, with RF performing best at 0.91. RF showed excellent performance across training, testing, and external validation sets. RF is an ML algorithm capable of capturing potential nonlinear relationships between risk factors and outcomes [22]. It operates by creating hundreds or thousands of smaller decision trees and combining their outputs for prediction [23]. This approach helps overcome the tendency of decision trees to overfit [23] while enhancing model generalizability [24]. Therefore, we recommend the use of the RF to predict the occurrence of PPTIC in patients with trauma. The prediction variables were primarily continuous laboratory indicators, eliminating the need for conversion into binary or categorical forms and facilitating practical use. Physicians can directly input these indicators into the RF model to assess the risk of PPTIC in patients with trauma. Recent guidelines highlight the importance of early detection and management of TIC to improve outcomes for severely injured patients [9]. Our study aims to predict PPTIC using preoperative indicators to facilitate early intervention and thereby improve clinical outcomes in such patients.

Most robust PPTIC prediction models in our study are tree-based, such as RF, DT, GB, and XGBoost, a trend also seen in other coagulopathy-related ML models. Perkins et al [25] used BN to predict early TIC, defined as an INR>1.2. Their model, trained on 600 patients with trauma, achieved an AUROC of 0.93 with a BS of 0.06 in the development dataset and an AUROC of 0.95 with a BS of 0.05 in the validation set [25]. In comparison, Our RF model achieved an AUROC of 1 in training, 0.92 in validation, and 0.91 in external validation, with BS values of 0.01, 0.10, and 0.13, respectively. Li et al [26] defined acute traumatic coagulopathy (ATC) as an INR>1.5 and collected data from 818 patients with trauma admitted to the emergency department. They used LR and RF to predict ATC occurrence, with LR achieving an AUROC of 0.858 and RF achieving an AUROC of 0.830 in the internal validation set [27]. Their study used only 2 ML models, with RF and LR achieving moderate predictive performance for ATC. Peltan et al [27] predicted the occurrence of prehospital ATC in 285 severely injured patients, defining coagulopathy as an INR>1.5, with an AUROC value of only 0.73. Previous studies had small sample sizes and moderate predictive power, and were limited to single centers without external validation. In contrast, our multicenter study employed 10 ML models, with 3 demonstrating excellent performance. It included diverse medical centers and populations (White, Asian, Black, and

XSL•FO
RenderX

Latino) and featured external validation, showcasing the models' broad applicability and generalizability.

Preoperative Risk Factors and In-Hospital Complications of PPTIC in Patients With Trauma

The definition of TIC relies primarily on coagulation tests, drawing from extensive literature on the subject. Brohi et al [3] recommended that patients with a PT > 18 s, an INR > 1.5, an APTT>60 s, or a TT>15 s could be diagnosed with the hypocoagulable phenotype of TIC, and this diagnosis was not affected by the use of anticoagulant medications or abnormal blood samples. MacLeod et al [28] predicted early coagulopathy in 20,103 patients with trauma and reported a higher incidence after trauma, which was associated with mortality. They defined early coagulopathy as a PT>14 s, an APTT>35 s, and a PLT<100×10^9/L, highlighting its role as an independent predictor of death despite other risk factors [28]. Currently, the accepted definition of TIC utilizes the INR, with an INR>1.2 being the threshold for detecting TIC [8]. Yuan et al [18] defined TIC for predicting mortality among patients with trauma as a PLT<100×10^9/L, an INR>1.25, a PT>14 seconds, and an APTT>36 seconds [18]. We defined TIC based on coagulation test reference ranges from various centers and literature, using slightly lower standards than Brohi et al [3]. The main reasons for adopting these criteria are as follows: (1) lower diagnostic criteria alert clinicians to take early preventive measures to prevent PPTIC from progressing to disseminated intravascular coagulation; and (2) patients with preexisting TIC have a 50% chance of persisting with TIC postoperatively.

The prolongation of the preoperative APTT, PT, and INR in patients with trauma is understandable as a risk factor for PPTIC. The decreases in hemoglobin, hematocrit, and RBC levels and lower admission DBP are associated mainly with acute anemia due to preoperative bleeding in patients with trauma. Ryan et al [29] noted a close correlation between initial low hematocrit or hemoglobin levels in patients with trauma and hemorrhagic shock. RBC and hematocrit levels are key factors in platelet transport and adhesion to endothelial injury. Platelet adhesion increases 5-fold when hematocrit rises from 10% to 40% [30]. Therefore, the RBC level is closely related to the platelet level. Steele et al [31] reported an independent correlation between hyponatremia at admission and hypocalcemia. Ionic calcium is crucial not only for the formation and stability of fibrin polymerization sites but also for many platelet-related functions [32]. Platelet activation and aggregation are also calcium-dependent [33] and play crucial roles in the pathophysiology of TIC [34]. Elevated ALT and AST levels indicate liver damage, which impacts coagulation factor production and the coagulation process.

Hess et al [35] reported that patients with TIC experience more blood loss, higher transfusion needs, and a greater incidence of MODS compared to those without TIC. Brohi et al [3] highlighted that TIC is associated with a 4-fold increase in mortality, increased rates of massive transfusion and MODS, and prolonged ICU and LOS. Smalls et al [36] reported that patients with TIC due to splenic injury face 2.4 times more complications, with increased mortality (1.3-fold), sepsis (2-fold), acute respiratory distress syndrome (2.6-fold), acute renal failure (1.5-fold), and cardiac arrest (1.5-fold). These findings align with our study, where patients with PPTIC had increased risks of sepsis (1.75-fold), heart failure (1.5-fold), delirium (3.08-fold), abnormal coagulation (3.57-fold), tracheostomy (2.76-fold), mortality (2.19-fold), urinary tract infection (1.95-fold), and longer LOS and ICU stay.

Limitations

This retrospective study focused primarily on predicting PPTIC in patients with trauma via admission laboratory indicators and medical records. In this study, we addressed missing data to varying extents by excluding variables with more than 40% missing values and applying distinct imputation methods for variables with less than 40% missing data (KNN for continuous variables and multiple imputation for categorical variables). While multiple imputation is an effective method to mitigate systematic bias from nonrandom missing data, some data remained missing, which could affect the model's predictive accuracy for specific patient characteristics. Due to missing values exceeding 40% for certain indicators, such as height, admission peripheral capillary oxygen saturation, temperature on admission, pre-lymphocyte, pre-total protein, pre-globulin, pre-direct bilirubin, pre-lactate dehydrogenase, pre-D-dimer, and post-D-dimer (Multimedia Appendix 3), these variables were excluded from model development. As these missing data may impact the predictive performance of the model, future studies could address this by collecting more comprehensive data to improve model robustness. Additionally, large-sample prospective clinical trials are necessary to validate the model's applicability in clinical settings and assess its effectiveness in reducing the incidence of PPTIC.

Conclusions

Our study, using data from 4 medical centers, identified risk factors for PPTIC in patients with trauma. Significant factors included (1) prolonged preoperative APTT, PT, and INR; (2) decreased hemoglobin, hematocrit, RBC, calcium, and sodium levels; (3) lower admission DBP; (4) elevated ALT and AST levels; (5) increased admission HR; and (6) the need for emergency surgery or perioperative transfusion. PPTIC was associated with higher rates of sepsis, heart failure, delirium, abnormal coagulation, tracheostomy, mortality, and extended hospital and ICU stays. Of the 10 ML models evaluated for predicting PPTIC, RF demonstrated the best performance with an AUROC of 0.91, AUPRC of 0.89, accuracy of 0.84, and a BS of 0.13.

Acknowledgments

The authors declare that no generative artificial intelligence tools were used in the writing or preparation of this manuscript. The authors received funding from the Chongqing Talent Program (Innovative leading talents CSTC2024YCJH-BGZXM0011);

XSL•FO
RenderX

General Program of Joint Medical Research of Chongqing Science and Health Commission and Chong-qing Health Commission (2023MSXM127); Chongqing Municipal Studios for Young and Middle-aged Top Medical Talents (ZQNYXGDRCGZS2019006); and Daping Hospital (ZXAIYB020).

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to ongoing research and unpublished results but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence) guidelines.
[DOCX File , 79 KB - jmir_v27i1e66612_app1.docx]

Multimedia Appendix 2
A comprehensive description of the methodology.
[DOCX File , 16 KB - jmir_v27i1e66612_app2.docx]

Multimedia Appendix 3
The amount of data missing and the coding process.
[DOCX File , 25 KB - jmir_v27i1e66612_app3.docx]

Multimedia Appendix 4
Sklearn packages for machine learning models.
[DOCX File , 14 KB - jmir_v27i1e66612_app4.docx]

Multimedia Appendix 5
Python packages used for data processing and analysis.
[DOCX File , 15 KB - jmir_v27i1e66612_app5.docx]

Multimedia Appendix 6
Ten machine learning models predicting patients with trauma who have preoperative and postoperative traumatic coagulopathy and related performance evaluation.
[XLSX File (Microsoft Excel File), 13 KB - jmir_v27i1e66612_app6.xlsx]

Multimedia Appendix 7
Results of 5-fold cross-validation for 10 machine learning models.
[DOCX File , 22 KB - jmir_v27i1e66612_app7.docx]

Multimedia Appendix 8
SHA*P* values of AdaBoost, DT, KNN, LR, NB, NN, and SVM for predicting the risk of TIC before and after surgery in patients with trauma. DT: decision tree; KNN: k-nearest neighbor; LR: logistic regression; NB: naive Bayes; NN: neural network; SVM: support vector machine; SHAP: Shapley additive explanation; TIC: traumatic coagulopathy.
[DOCX File , 1266 KB - jmir_v27i1e66612_app8.docx]

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Kornblith LZ, Moore HB, Cohen MJ. Trauma-induced coagulopathy: the past, present, and future. J Thromb Haemost 2019;17(6):852-862 [FREE Full text] [doi: 10.1111/jth.14450] [Medline: 30985957]
3. Brohi K, Singh J, Heron M, Coats T. Acute traumatic coagulopathy. J Trauma 2003;54(6):1127-1130. [doi: 10.1097/01.TA.0000069184.82147.06] [Medline: 12813333]

XSL•FO
RenderX

4. Maegele M, Lefering R, Yucel N, Tjardes T, Rixen D, Paffrath T, et al. Early coagulopathy in multiple injury: an analysis from the German Trauma Registry on 8724 patients. Injury 2007;38(3):298-304. [doi: 10.1016/j.injury.2006.10.003] [Medline: 17214989]
5. Harhangi BS, Kompanje EJO, Leebeek FWG, Maas AIR. Coagulation disorders after traumatic brain injury. Acta Neurochir 2008;150(2):165-175. [doi: 10.1007/s00701-007-1475-8] [Medline: 18166989]
6. Duque P, Calvo A, Lockie C, Schöchl H. Pathophysiology of trauma-induced coagulopathy. Transfus Med Rev 2021;35(4):80-86. [doi: 10.1016/j.tmrv.2021.07.004] [Medline: 34610877]
7. Cohen MJ, Christie SA. Coagulopathy of trauma. Crit Care Clin 2017;33(1):101-118. [doi: 10.1016/j.ccc.2016.08.003] [Medline: 27894491]
8. Cole E, Weaver A, Gall L, West A, Nevin D, Tallach R, et al. A decade of damage control resuscitation: new transfusion practice, new survivors, new directions. Ann Surg 2021;273(6):1215-1220. [doi: 10.1097/SLA.0000000000003657] [Medline: 31651535]
9. Rossaint R, Afshari A, Bouillon B, Cerny V, Cimpoesu D, Curry N, et al. The European guideline on management of major bleeding and coagulopathy following trauma: sixth edition. Crit Care 2023;27(1):80 [FREE Full text] [doi: 10.1186/s13054-023-04327-7] [Medline: 36859355]
10. National Academy of Medicine. In: Whicher D, Ahmed M, Thadaney Israni S, Matheny M, editors. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: National Academies Press; 2023.
11. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19(1):64 [FREE Full text] [doi: 10.1186/s12874-019-0681-4] [Medline: 30890124]
12. Liu NT, Salinas J. Machine learning for predicting outcomes in trauma. Shock 2017;48(5):504-510. [doi: 10.1097/SHK.0000000000000898] [Medline: 28498299]
13. Seki T, Kawazoe Y, Ohe K. Machine learning-based prediction of in-hospital mortality using admission laboratory data: a retrospective, single-site study using electronic health record data. PLoS One 2021;16(2):e0246640 [FREE Full text] [doi: 10.1371/journal.pone.0246640] [Medline: 33544775]
14. Rashidi HH, Bowers KA, Reyes Gil M. Machine learning in the coagulation and hemostasis arena: an overview and evaluation of methods, review of literature, and future directions. J Thromb Haemost 2023;21(4):728-743 [FREE Full text] [doi: 10.1016/j.jtha.2022.12.019] [Medline: 36696218]
15. Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S, et al. MIMIC-IV. PhysioNet 2021. [doi: 10.13026/s6n6-xd98]
16. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:e078378 [FREE Full text] [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
17. Mathew G, Agha R, Albrecht J, Goel P, Mukherjee I, Pai P, et al. STROCSS 2021: strengthening the reporting of cohort, cross-sectional and case-control studies in surgery. Int J Surg 2021;96:106165 [FREE Full text] [doi: 10.1016/j.ijsu.2021.106165] [Medline: 34774726]
18. Yuan Q, Yu J, Wu X, Sun YR, Li ZQ, Du ZY, et al. Prognostic value of coagulation tests for in-hospital mortality in patients with traumatic brain injury. Scand J Trauma Resusc Emerg Med 2018;26(1):3 [FREE Full text] [doi: 10.1186/s13049-017-0471-0] [Medline: 29304855]
19. PPTIC-ML. GitHub. URL: https://github.com/kk1937/PPTIC-ML [accessed 2024-12-31]
20. Yu M, Wang S, He K, Teng F, Deng J, Guo S, et al. Predicting the complexity and mortality of polytrauma patients with machine learning models. Sci Rep 2024;14(1):8302 [FREE Full text] [doi: 10.1038/s41598-024-58830-0] [Medline: 38594313]
21. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Comput Biol Med 2021;137:104813 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104813] [Medline: 34481185]
22. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019;19(1):281 [FREE Full text] [doi: 10.1186/s12911-019-1004-8] [Medline: 31864346]
23. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, et al. An introduction to machine learning. Clin Pharmacol Ther 2020;107(4):871-885 [FREE Full text] [doi: 10.1002/cpt.1796] [Medline: 32128792]
24. Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, et al. Predicting pressure injury in critical care patients: a machine-learning model. Am J Crit Care 2018;27(6):461-468 [FREE Full text] [doi: 10.4037/ajcc2018525] [Medline: 30385537]
25. Perkins ZB, Yet B, Marsden M, Glasgow S, Marsh W, Davenport R, et al. Early identification of trauma-induced coagulopathy: development and validation of a multivariable risk prediction model. Ann Surg 2021;274(6):e1119-e1128. [doi: 10.1097/SLA.0000000000003771] [Medline: 31972649]
26. Li K, Wu H, Pan F, Chen L, Feng C, Liu Y, et al. A machine learning-based model to predict acute traumatic coagulopathy in trauma patients upon emergency hospitalization. Clin Appl Thromb Hemost 2020;26:1076029619897827 [FREE Full text] [doi: 10.1177/1076029619897827] [Medline: 31908189]

XSL•FO
RenderX

27. Peltan ID, Rowhani-Rahbar A, Vande Vusse LK, Caldwell E, Rea TD, Maier RV, et al. Development and validation of a prehospital prediction model for acute traumatic coagulopathy. Crit Care 2016;20(1):371 [FREE Full text] [doi: 10.1186/s13054-016-1541-9] [Medline: 27846895]
28. MacLeod JBA, Lynn M, McKenney MG, Cohn SM, Murtha M. Early coagulopathy predicts mortality in trauma. J Trauma 2003;55(1):39-44. [doi: 10.1097/01.TA.0000075338.21177.EF] [Medline: 12855879]
29. Ryan ML, Thorson CM, Otero CA, Vu T, Schulman CI, Livingstone AS, et al. Initial hematocrit in trauma: A paradigm shift? J Trauma Acute Care Surg 2012;72(1):54-59. [doi: 10.1097/TA.0b013e31823d0f35] [Medline: 22310116]
30. Turitto VT, Weiss HJ. Red blood cells: their dual role in thrombus formation. Science 1980;207(4430):541-543. [doi: 10.1126/science.7352265] [Medline: 7352265]
31. Steele T, Kolamunnage-Dona R, Downey C, Toh C, Welters I. Assessment and clinical course of hypocalcemia in critical illness. Crit Care 2013;17(3):R106 [FREE Full text] [doi: 10.1186/cc12756] [Medline: 23734769]
32. Lier H, Krep H, Schroeder S, Stuber F. Preconditions of hemostasis in trauma: a review. The influence of acidosis, hypocalcemia, anemia, and hypothermia on functional hemostasis in trauma. J Trauma 2008;65(4):951-960. [doi: 10.1097/TA.0b013e318187e15b] [Medline: 18849817]
33. Matthay ZA, Fields AT, Nunez-Garcia B, Patel MH, Cohen MJ, Callcut RA, et al. Dynamic effects of calcium on in vivo and ex vivo platelet behavior after trauma. J Trauma Acute Care Surg 2020;89(5):871-879 [FREE Full text] [doi: 10.1097/TA.0000000000002820] [Medline: 32852184]
34. Vulliamy P, Gillespie S, Armstrong PC, Allan HE, Warner TD, Brohi K. Histone H4 induces platelet ballooning and microparticle release during trauma hemorrhage. Proc Natl Acad Sci U S A 2019;116(35):17444-17449 [FREE Full text] [doi: 10.1073/pnas.1904978116] [Medline: 31405966]
35. Hess JR, Brohi K, Dutton RP, Hauser CJ, Holcomb JB, Kluger Y, et al. The coagulopathy of trauma: a review of mechanisms. J Trauma 2008;65(4):748-754. [doi: 10.1097/TA.0b013e3181877a9c] [Medline: 18849786]
36. Smalls N, Obirieze A, Ehanire I. The impact of coagulopathy on traumatic splenic injuries. Am J Surg 2015;210(4):724-729. [doi: 10.1016/j.amjsurg.2015.05.035] [Medline: 26384795]

Abbreviations

AI: artificial intelligence
ALT: alanine aminotransferase
APTT: activated partial thromboplastin time
AST: aspartate amino transferase
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
BS: Brier score
DBP: diastolic blood pressure
DT: decision tree
GB: gradient boosting
HR: heart rate
ICU: intensive care unit
INR: international normalized ratio
KNN: k-nearest neighbor
LOS: length of stay
LR: logistic regression
ML: machine learning
MODS: multiple organ dysfunction syndrome
MOF: multiple organ failure
NB: naive Bayes
NN: neural networks
PLT: platelet count
PPTIC: preoperative and postoperative traumatic coagulopathy
PT: prothrombin time
RBC: red blood cell
STROCSS: Strengthening the Reporting of Cohort Studies in Surgery
TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence
RF: random forest
SHAP: Shapley additive explanation
SVM: support vector machine
TIC: traumatic coagulopathy

XSL•FO
RenderX

XGBoost: extreme gradient boosting

Edited by A Mavragani; submitted 18.09.24; peer-reviewed by Z Ehtesham, CC Yen; comments to author 04.11.24; revised version received 19.11.24; accepted 27.11.24; published 22.01.25.

Please cite as:
Xiong X, Fu H, Xu B, Wei W, Zhou M, Hu P, Ren Y, Mao Q
Ten Machine Learning Models for Predicting Preoperative and Postoperative Coagulopathy in Patients With Trauma: Multicenter Cohort Study
J Med Internet Res 2025;27:e66612
URL: https://www.jmir.org/2025/1/e66612
doi:10.2196/66612
PMID:39841523

©Xiaojuan Xiong, Hong Fu, Bo Xu, Wang Wei, Mi Zhou, Peng Hu, Yunqin Ren, Qingxiang Mao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

AI-Driven Innovations for Early Sepsis Detection by Combining Predictive Accuracy With Blood Count Analysis in an Emergency Setting: Retrospective Study

Tai-Han Lin[1*], MD; Hsing-Yi Chung[1*], MS; Ming-Jr Jian[1], MS, PhD; Chih-Kai Chang[1], MS; Hung-Hsin Lin[1], MD; Chiung-Tzu Yen[1], MS; Sheng-Hui Tang[1], MS; Pin-Ching Pan[1], MS; Cherng-Lih Perng[1], MS; Feng-Yee Chang[2], MD, PhD; Chien-Wen Chen[3], MD; Hung-Sheng Shang[1], MD, PhD

[1]Division of Clinical Pathology, Department of Pathology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[2]Division of Infectious Diseases and Tropical Medicine, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[3]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Hung-Sheng Shang, MD, PhD
Division of Clinical Pathology, Department of Pathology
Tri-Service General Hospital, National Defense Medical Center
No. 161, Sec. 6, Minquan E. Rd., Neihu Dist.
Taipei, 11490
Taiwan
Phone: 886 920713130
Fax: 886 87927226
Email: iamkeith001@gmail.com

Abstract

Background: Sepsis, a critical global health challenge, accounted for approximately 20% of worldwide deaths in 2017. Although the Sequential Organ Failure Assessment (SOFA) score standardizes the diagnosis of organ dysfunction, early sepsis detection remains challenging due to its insidious symptoms. Current diagnostic methods, including clinical assessments and laboratory tests, frequently lack the speed and specificity needed for timely intervention, particularly in vulnerable populations such as older adults, intensive care unit (ICU) patients, and those with compromised immune systems. While bacterial cultures remain vital, their time-consuming nature and susceptibility to false negatives limit their effectiveness. Even promising existing machine learning approaches are restricted by reliance on complex clinical factors that could delay results, underscoring the need for faster, simpler, and more reliable diagnostic strategies.

Objective: This study introduces innovative machine learning models using complete blood count with differential (CBC+DIFF) data—a routine, minimally invasive test that assesses immune response through blood cell measurements, critical for sepsis identification. The primary objective was to implement this model within an artificial intelligence–clinical decision support system (AI-CDSS) to enhance early sepsis detection and management in critical care settings.

Methods: This retrospective study at Tri-Service General Hospital (September to December 2023) analyzed 746 ICU patients with suspected pneumonia-induced sepsis (supported by radiographic evidence and a SOFA score increase of ≥2 points), alongside 746 stable outpatients as controls. Sepsis infection sources were confirmed through positive sputum, blood cultures, or FilmArray results. The dataset incorporated both basic hematological factors and advanced neutrophil characteristics (side scatter light intensity, cytoplasmic complexity, and neutrophil-to-lymphocyte ratio), with data from September to November used for training and data from December used for validation. Machine learning models, including light gradient boosting machine (LGBM), random forest classifier, and gradient boosting classifier, were developed using CBC+DIFF data and were assessed using metrics such as area under the curve, sensitivity, and specificity. The best-performing model was integrated into the AI-CDSS, with its implementation supported through workshops and training sessions.

Results: Pathogen identification in ICU patients found 243 FilmArray-positive, 411 culture-positive, and 92 undetected cases, yielding a final dataset of 654 (43.8%) sepsis cases out of 1492 total cases. The machine learning models demonstrated high

XSL•FO
RenderX

predictive accuracy, with LGBM achieving the highest area under the curve (0.90), followed by the random forest classifier (0.89) and gradient boosting classifier (0.88). The best-performing LGBM model was selected and integrated as the core of our AI-CDSS, which was built on a web interface to facilitate rapid sepsis risk assessment using CBC+DIFF data.

Conclusions: This study demonstrates that by providing streamlined predictions using CBC+DIFF data without requiring extensive clinical parameters, the AI-CDSS can be seamlessly integrated into clinical workflows, enhancing rapid, accurate identification of sepsis and improving patient care and treatment timeliness.

(J Med Internet Res 2025;27:e56155) doi:10.2196/56155

KEYWORDS

sepsis; artificial intelligence; critical care; complete blood count analysis; CBC analysis; artificial intelligence clinical decision support systems; AI-CDSS

Introduction

The World Health Organization classifies sepsis as a critical global health issue, responsible for approximately 20% of all deaths worldwide in 2017, with 48.9 million cases and 11 million fatalities [1,2]. Sepsis, a life-threatening condition triggered by the body's response to infection, leads to severe consequences, such as septic shock, organ failure, and even death if not promptly managed [3]. The mortality rate for sepsis is alarmingly high at 32.2%, and it rises significantly to 38.5% for those who develop septic shock, as reported in a meta-analysis [4]. It predominantly affects high-risk groups such as older adults, intensive care unit (ICU) patients, and those with chronic conditions or compromised immune systems [5]. The prevalence of sepsis, which is intensified by drug-resistant infections, highlights the need for more effective prevention and treatment methods [6] and rapid diagnostic strategies [7], particularly in optimizing their implementation in clinical settings [8]. The key to mitigating the heavy health toll of sepsis lies in its efficient management and, crucially, early detection, which significantly enhances the probability of recovery and diminishes its overall detrimental health impact [1,9].

According to the Surviving Sepsis Campaign, sepsis diagnosis relies on the Sequential Organ Failure Assessment (SOFA) score, which evaluates organ dysfunction resulting from infection by assessing 6 components: respiratory function through the ratio of partial pressure of oxygen in arterial blood to the fraction of inspiratory oxygen concentration, coagulation through the platelet count, liver function with bilirubin levels, cardiovascular status through hypotension or the use of vasopressors, central nervous system function using the Glasgow Coma Scale, and renal function measured through creatinine levels or urine output [3]. An increase of 2 or more points in the SOFA score, combined with suspected infection, indicates sepsis [3,10]. For rapid assessments outside of ICUs, the quick SOFA criteria, altered mental state, respiratory rate of ≥22 breaths per minute, or low blood pressure (systolic blood pressure≤100 mm Hg) are used for prompt evaluation, which is critical for reducing sepsis-related health impacts [3,10]. Alongside identifying and addressing the source of infection or removing suspected infected routes or devices, immediate administration of broad-spectrum antimicrobial agents within 1 hour of detecting sepsis is crucial for effective management [3,10,11]. Rapid intravenous crystalloid resuscitation and ongoing fluid adjustments are essential to maintain circulatory volume and ensure organ perfusion [3,10]. Vasopressors, such as norepinephrine, should be applied if a mean arterial pressure of at least 65 mm Hg cannot be maintained [3,10]. Dynamic monitoring of stroke volume, pulse pressure variations, and respiratory status is essential for effective treatment, enabling precise, real-time adjustments to fluid therapy [3,10].

Traditional diagnostic methods such as clinical assessments, laboratory evaluations, and patient histories, while foundational, often lack the immediacy and specificity required to promptly confirm infections [3,5,12]. The progression from infection to sepsis can be insidious, with initial clinical manifestations that are highly variable and influenced by both host and pathogen factors [10]. This variability can be particularly detrimental, as rapid identification within the first crucial hours is vital; delays in diagnosis or the initiation of antibiotic treatment can lead to rapid deterioration in the patient's condition, significantly worsening the prognosis [10]. While existing machine learning models have sought to reduce these delays by expediting the diagnosis of sepsis, many rely on complex clinical factors that can still delay results or affect accuracy [13-15]. In addition, blood cultures, though critical for identifying pathogens, could take days and are prone to false negatives, further complicating the timely management of sepsis [3,5,12]. This highlights a critical need for simpler, more effective diagnostic approaches that can be swiftly applied and have high reliability.

Addressing these challenges, our research introduces a novel machine learning model centered on complete blood count with differential (CBC+DIFF) data, a routine and minimally invasive test that measures the levels of various blood cell types, including white blood cells, red blood cells, and platelets. This approach was selected for its broad clinical use, rapid analytical time, and ability to evaluate immune response through white blood cell differential analysis (such as neutrophils and lymphocytes), which are crucial for early sepsis detection. This makes it an accessible and cost-effective tool for diagnostics, allowing us to simplify the diagnostic process without sacrificing accuracy. The primary objective of this study was to develop and implement an artificial intelligence–clinical decision support system (AI-CDSS) that uses this model to provide rapid and precise indications of sepsis. This system aims to facilitate earlier intervention and potentially improve patient outcomes across various medical settings, from outpatient departments to emergency departments, with a particular focus on high-stakes environments such as ICUs, where timely decisions are crucial.

XSL•FO
RenderX

Methods

Study Focus and Data Collection

In our retrospective study conducted at the ICU of Tri-Service General Hospital from September to December 2023, we focused on cases suspected of pneumonia-induced sepsis, as approved by the hospital's institutional review board. The study site, a major teaching hospital affiliated with the National Defense Medical Center in Taipei, Taiwan, is a facility with approximately 1800 beds, serving as the primary referral center for 3 branch hospitals in northern Taiwan. This comprehensive medical center provides advanced medical facilities and multidisciplinary care to both civilian and military patients, while its branch hospitals focus on chronic condition management and refer complex acute cases, including sepsis, to the main facility. This centralized health care network structure enabled us to collect data from a diverse patient population across various clinical settings. Our data were exclusively derived from the electronic medical records of patients who were either transferred from the general ward owing to deteriorating conditions or directly admitted from the emergency department owing to severe initial clinical presentations. These records include admission notes, progress notes, laboratory results, and radiographic findings. We included only those patients whose records indicated pneumonia as the suspected cause of sepsis, corroborated by radiographic evidence consistent with pneumonia and an increase in the SOFA score of 2 or more points, indicating significant organ dysfunction. Specifically, we extracted and analyzed data on CBC+DIFF, sputum cultures, blood cultures, and FilmArray analyses. These selected data points formed our core dataset used to analyze the diagnostic and management strategies for pneumonia-induced sepsis in the ICU.

Control Group Selection

In addition to the groups directly involved in the sepsis investigation, the study included a control group of 746 randomly selected outpatients from the same hospital, enrolled during the same period. These patients typically visited the hospital for routine follow-up appointments. On their appointment days, they first stopped at our blood draw station, where standard blood tests were performed. The results of these tests were available to their physicians within 1 hour, ensuring that the most current health data would be reviewed during their visit. Data from these patients were obtained from the hospital's electronic medical records. These records confirmed that none of these individuals required subsequent transfer to the emergency department for acute issues, verifying their stable health condition at the time of their visit. CBC+DIFF data were collected specifically for this group.

Ethical Considerations

Our retrospective study was conducted following strict ethical guidelines approved by the Tri-Service General Hospital's Institutional Review Board (approval C202305073). The study adhered to the principles outlined in the Declaration of Helsinki and relevant local ethical regulations. The study involved analyzing anonymized patient data to ensure privacy and confidentiality and was exempt from the requirement for patient consent under local and international ethical regulations. No compensation was provided to the patients as no direct interaction was involved.

Sepsis Identification and Classification Criteria

Criteria for Confirming Sepsis

We initially considered all ICU patients with suspected sepsis. Confirmation depended on positive sputum or blood cultures, or FilmArray test results [16], in conjunction with clinical symptoms and identified infection sources.

Exclusion Parameters

Cases lacking data or with poor sample quality (including blood samples with volumes less than 10 mL, sputum samples less than 1 mL, or showing contamination indicators such as squamous cell count>10/low power field or mixed bacterial flora), impacting the trustworthiness of culture reports, were excluded.

Determination of Nonseptic Status

Patients with negative results in both culture and FilmArray tests were classified as nonseptic.

Data Processing and Model Formulation

Data Organization

We meticulously organized the patient data using Python's *Pandas* framework (Python Software Foundation) to enhance data handling and analysis efficiency.

Selection of Features

Our approach involved a comprehensive analysis of hematological parameters from the CBC+DIFF data using the Sysmex XN-9100 system, focusing on parameters relevant to sepsis pathology. These included white blood cell count (10^3/μL), red blood cell count (10^6/μL), hemoglobin (g/dL), hematocrit (%), mean corpuscular volume (fL), mean corpuscular hemoglobin (pg), mean corpuscular hemoglobin concentration (g/dL), platelet count (10^3/μL), neutrophil percentage (%), lymphocyte percentage (%), monocyte percentage (%), eosinophil percentage (%), basophil percentage (%), immature granulocyte count (10^3/μL), immature granulocyte percentage (%), neutrophil side scatter light intensity, neutrophil cytoplasmic complexity, and neutrophil-to-lymphocyte ratio. Special attention was paid to neutrophil side scatter light intensity and neutrophil cytoplasmic complexity regarding their insights into neutrophil characteristics. Heat map visualization aided in identifying the most significant features of our predictive model and in discarding less relevant features to optimize the dataset for targeted analysis. In addition, we used Shapley additive explanations (SHAP) values to interpret feature importance and contribution to the model's output. SHAP summary plots generated with the *TreeExplainer* illustrated the impact of individual features on sepsis risk predictions, enhancing our understanding of the model's decision-making process and ensuring interpretability and reliability.

XSL•FO
RenderX

Model Development and Evaluation

Data Segmentation

The dataset was divided, with data from September to November forming the training set and data from December used for validation.

Using Machine Learning Algorithms

We used a diverse range of machine learning models for the extensive analysis of predictive capabilities. These included logistic regression (LR), linear discriminant analysis, the random forest classifier (RFC), the gradient boosting classifier (GBC), the AdaBoost classifier, extreme gradient boosting (XGBoost), and the light gradient boosting machine (LGBM). Each model was selected to provide a comprehensive perspective on the individual and collective abilities to accurately predict sepsis.

Assessment of Predictive Accuracy

These models were stringently tested using receiver operating characteristic curves, the area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the F_1-score to measure the accuracy of sepsis prediction. We also evaluated the models using the Brier score and Brier skill score (BSS) to assess the accuracy and calibration of probabilistic predictions. The Brier score measures the mean squared difference between predicted probabilities and actual outcomes, with lower scores indicating better accuracy. The BSS compares the model's performance against a baseline model using LR, which provides a more representative benchmark. Values above 0 for the BSS indicate superior performance.

Clinical Utility Evaluation

To select the best-performing model for accurate sepsis prediction, we validated the models using data collected in December and evaluated key performance metrics, including AUC, sensitivity, specificity, PPV, NPV, and F_1-score. The optimal model was saved using Python's *joblib* package for seamless implementation. For clinical deployment, we created a web-based, AI-CDSS interface that allows clinicians to initiate sepsis risk assessment by entering a patient ID. The system retrieves laboratory values, sends them to a Python Flask server for analysis, and then displays the calculated sepsis probability. In addition, health care professionals received comprehensive support through group workshops and online training sessions. Feedback collected through surveys and interviews was used to continuously refine the tool, enhancing its usability and integration into clinical workflows.

Results

Patient Demographics and Analysis

Our study included 746 ICU patients and a control group of 746 randomly selected healthy outpatients, which served as a normative benchmark. Among the ICU patients, 654 were classified in the sepsis group based on confirmed pathogen detection: 243 were identified through FilmArray tests and 411 through culture tests. The remaining 92 patients, initially suspected of sepsis, were reclassified into the nonsepsis group due to the absence of pathogen detection, suggesting alternative etiologies for their symptoms. This classification ultimately yielded a final dataset consisting of 654 (43.8%) sepsis cases out of a total of 1492 cases (Figure 1). We compiled the demographic and clinical characteristics, including the CBC+DIFF parameters, for each group, as detailed in Table 1.

XSL•FO
RenderX

Figure 1. Patient selection flowchart.

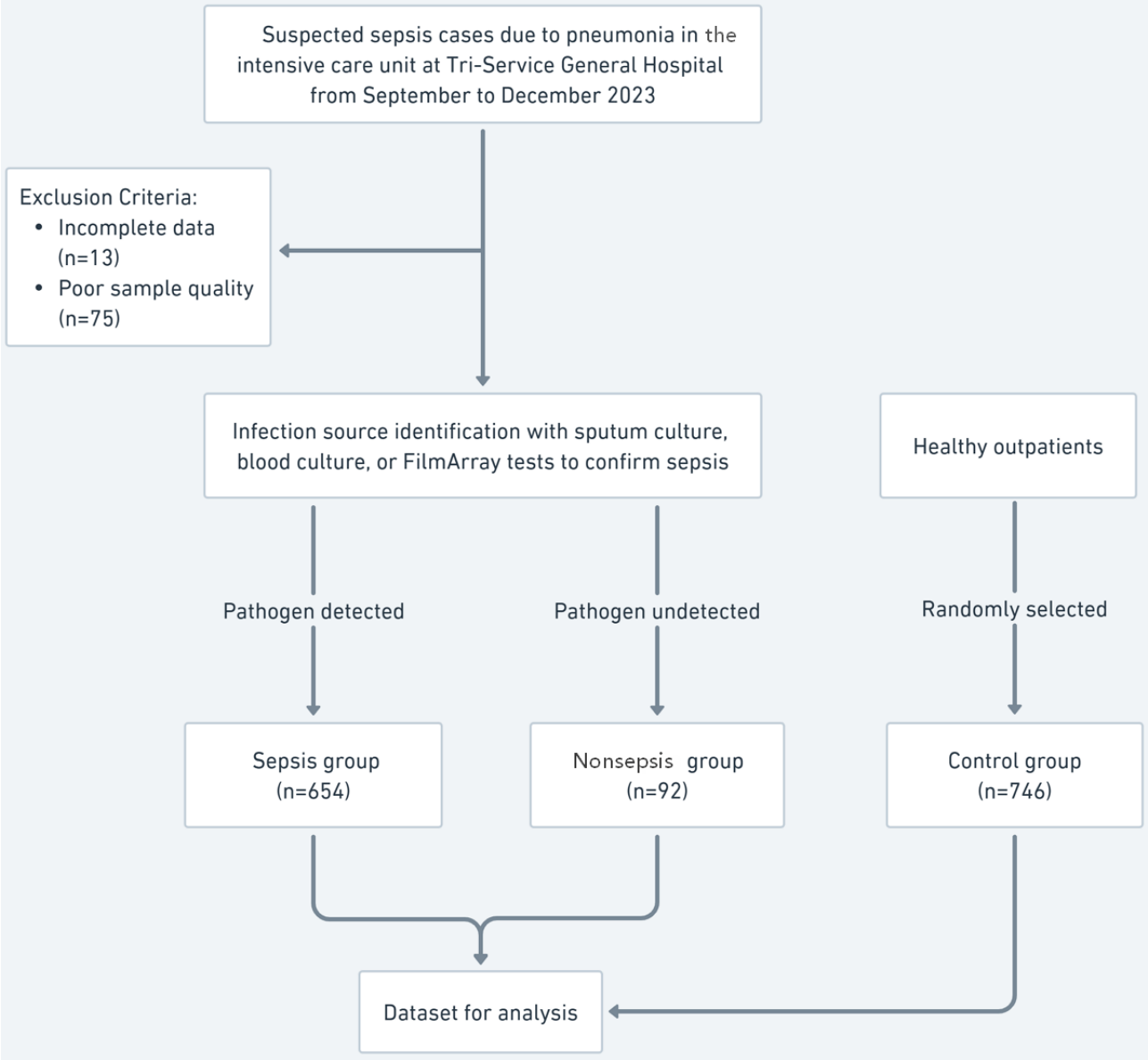

XSL•FO
RenderX

Table 1. Demographic and hematological characteristics of study groups.

Parameter	Sepsis group (n=654)	Nonsepsis group (n=92)	Control group (n=746)
Age (years), mean (SD)	69.04 (16.4)	67.98 (18.17)	55.52 (17.6)
Sex (male), n/N (%)	425/654 (64.98)	57/92 (61.96)	482/746 (64.61)
WBC[a] (10^3/μL), mean (SD)	10.06 (5.84)	11.34 (6.15)	6.88 (3.67)
RBC[b] (10^6/μL), mean (SD)	3.51 (0.82)	3.35 (0.71)	4.47 (0.77)
Hb[c] (g/dL), mean (SD)	10.26 (2.14)	9.89 (2.02)	13.05 (2.14)
HCT[d] (%), mean (SD)	31.62 (6.45)	30.15 (6.18)	40.01 (6.04)
MCV[e] (fL), mean (SD)	90.76 (7.35)	90.45 (6.72)	90.21 (8.33)
MCH[f] (pg), mean (SD)	29.46 (2.7)	29.68 (2.3)	29.42 (3.31)
MCHC[g] (g/dL), mean (SD)	32.45 (1.45)	32.83 (1.39)	32.56 (1.26)
Platelet count (10^3/μL), mean (SD)	187.63 (110.81)	196.45 (106.68)	238.41 (87.26)
Neutrophil (%), mean (SD)	78.45 (13.85)	84.62 (9.99)	60.92 (12.82)
Lymphocyte (%), mean (SD)	13.43 (10.9)	9.31 (7.27)	29.13 (11.86)
Monocyte (%), mean (SD)	6.18 (4.43)	5.08 (3.35)	6.57 (2.52)
Eosinophil (%), mean (SD)	1.65 (4.07)	0.76 (1.58)	2.79 (2.7)
Basophil (%), mean (SD)	0.3 (0.26)	0.23 (0.19)	0.6 (0.39)
Immature granulocytes (10^3/μL), mean (SD)	0.17 (0.34)	0.25 (0.56)	0.06 (0.2)
Immature granulocytes (%), mean (SD)	1.51 (2.41)	1.78 (2.68)	0.64 (1.4)
Neutrophil side scatter light intensity, mean (SD)	52.38 (9.22)	52.7 (5.41)	47.03 (3.34)
Neutrophil cytoplasmic complexity, mean (SD)	688.13 (184.1)	681.86 (110.85)	608.05 (79.31)
Neutrophil to lymphocyte ratio, mean (SD)	13.76 (39.89)	20.81 (51.03)	3.19 (4.35)

[a]WBC: white blood cell count.

[b]RBC: red blood cell count.

[c]Hb: hemoglobin.

[d]HCT: hematocrit.

[e]MCV: mean corpuscular volume.

[f]MCH: mean corpuscular hemoglobin.

[g]MCHC: mean corpuscular hemoglobin concentration.

In Figure 1, a comprehensive flowchart illustrates the patient selection process used in our study. It outlines the initial screening of ICU patients with suspected pneumonia; the criteria used to confirm sepsis (such as positive results from sputum or blood cultures or FilmArray tests); and how patients were categorized into groups of confirmed sepsis, suspected nonsepsis, and outpatient control. In addition, this flowchart delineates the number of patients at each stage and the reasons for exclusion.

Data Visualization and Feature Analysis

For feature selection, we used heat maps to graphically represent the correlation between various CBC+DIFF parameters and sepsis, as shown in Figure 2. This visualization technique enabled us to eliminate parameters, such as the mean corpuscular hemoglobin concentration, mean corpuscular volume, and monocyte percentage, which had minimal impact on our model's predictions. The SHAP summary plot, as shown in Figure S1 in Multimedia Appendix 1, further revealed that lymphocyte percentage was the most influential feature, with high values decreasing the sepsis risk. Neutrophil percentage and neutrophil-to-lymphocyte ratio strongly correlated with an increased sepsis risk. Features such as mean corpuscular hemoglobin concentration, mean corpuscular volume, and monocyte percentage also show the least impact, aligned with the interpretation from Figure 2. This confirms the critical role of specific blood parameters in sepsis detection and enhances our model's interpretability and clinical relevance.

XSL•FO
RenderX

Figure 2. Correlation heat map of complete blood count with differential (CBC+DIFF) parameters and sepsis. BASO: basophil (%); EO: eosinophil (%); HCT: hematocrit (%); HGB: hemoglobin (g/dL); IG#: immature granulocytes (10^3/μL); IG%: immature granulocytes (%); LYMPH: lymphocyte (%); MCH: mean corpuscular hemoglobin (pg); MCHC: mean corpuscular hemoglobin concentration (g/dL); MCV: mean corpuscular volume (fL); MONO: monocyte (%); NE-SFL: neutrophil side scatter light intensity; NEU: neutrophil (%); NE-WY: neutrophil cytoplasmic complexity; NLR: neutrophil-to-lymphocyte ratio; PLT: platelet count (10^3/μL); RBC: red blood cell count (10^6/μL); WBC: white blood cell count (10^3/μL).

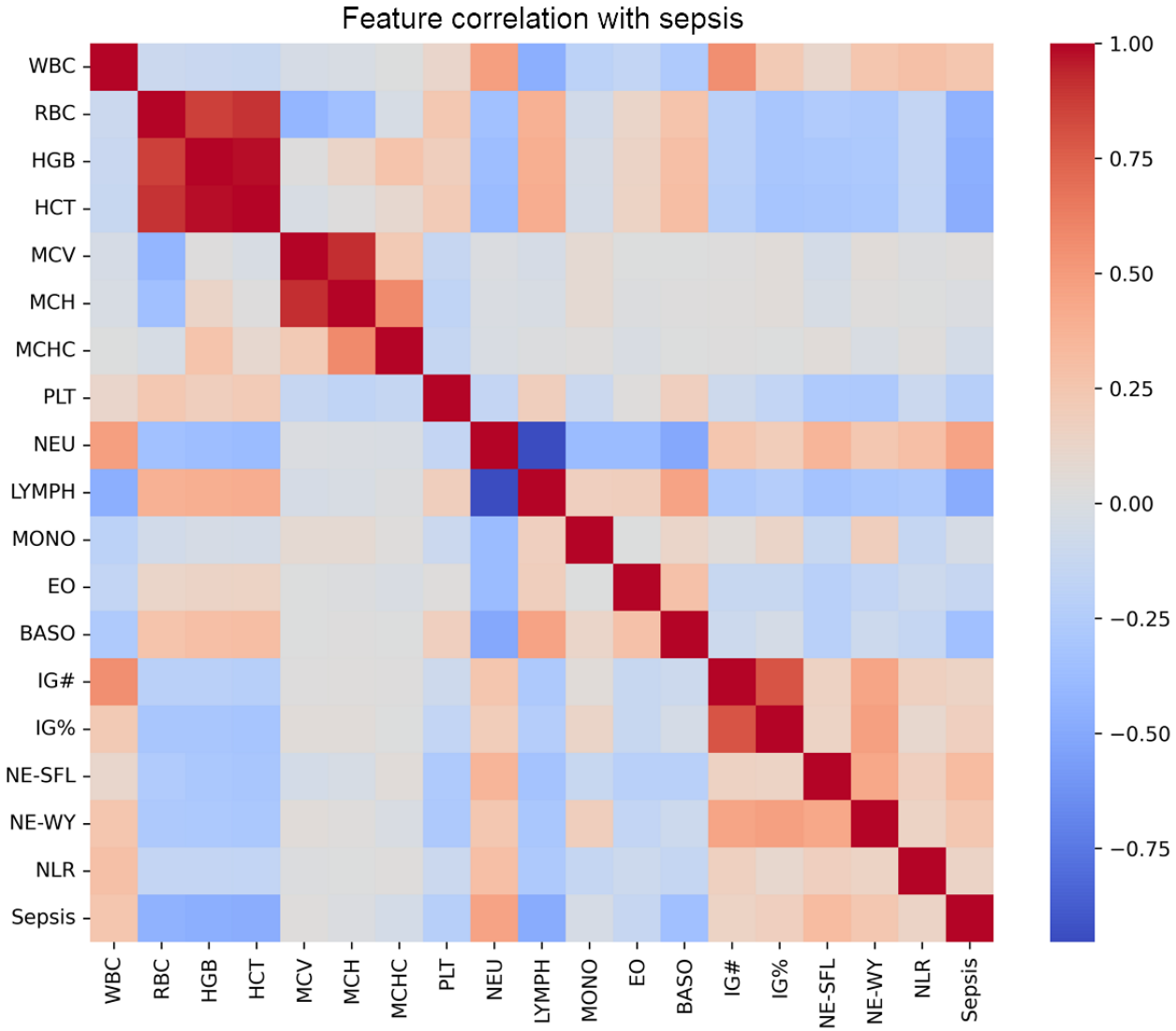

In Figure 2, the heat map depicts the correlations between various CBC+DIFF parameters and the occurrence of sepsis. This effectively shows the relevance of each parameter in predicting sepsis and serves as a crucial tool for identifying and selecting key features for the machine learning model. The color gradients in the heat map correspond to the varying strengths of these correlations.

Machine Learning Model Performance

Our analysis of the machine learning models for detecting sepsis yielded significant results, as detailed in Table 2 and illustrated in Figure 3. The GBC and RFC models exhibited superior predictive performance, with the GBC model achieving a perfect training AUC of 0.99 and a testing AUC of 0.88. The RFC model demonstrated excellent reliability, with a training AUC of 0.94 and a testing AUC of 0.89.

XSL•FO
RenderX

Table 2. Overview of machine learning models' performance.

Model	Training AUC[a]	Testing AUC	Accuracy	Sensitivity	Specificity	PPV[b]	NPV[c]	F_1-score	Brier Score	BSS[d]
LR[e]	0.86	0.85	0.80	0.80	0.80	0.74	0.85	0.77	0.15	Ref
LDA[f]	0.86	0.86	0.80	0.80	0.80	0.73	0.85	0.76	0.15	0.17%
RFC[g]	0.94	0.89	0.81	0.84	0.79	0.73	0.88	0.78	0.13	15.79%
GBC[h]	0.99	0.88	0.83	0.86	0.81	0.75	0.89	0.80	0.13	12.54%
ABC[i]	0.90	0.85	0.79	0.77	0.80	0.72	0.83	0.74	0.22	–47.29%
XGBoost[j]	0.97	0.88	0.80	0.81	0.80	0.73	0.86	0.77	0.13	14.33%
LGBM[k]	0.99	0.90	0.83	0.85	0.82	0.76	0.89	0.80	0.13	10.38%

[a]AUC: area under the curve.
[b]PPV: positive predictive value.
[c]NPV: negative predictive value.
[d]BSS: Brier skill score.
[e]LR: logistic regression.
[f]LDA: linear discriminant analysis.
[g]RFC: random forest classifier.
[h]GBC: gradient boosting classifier.
[i]ABC: AdaBoost classifier.
[j]XGBoost: extreme gradient boosting.
[k]LGBM: light gradient boosting machine.

Figure 3. Receiver operating characteristic curves of the evaluated machine learning models.

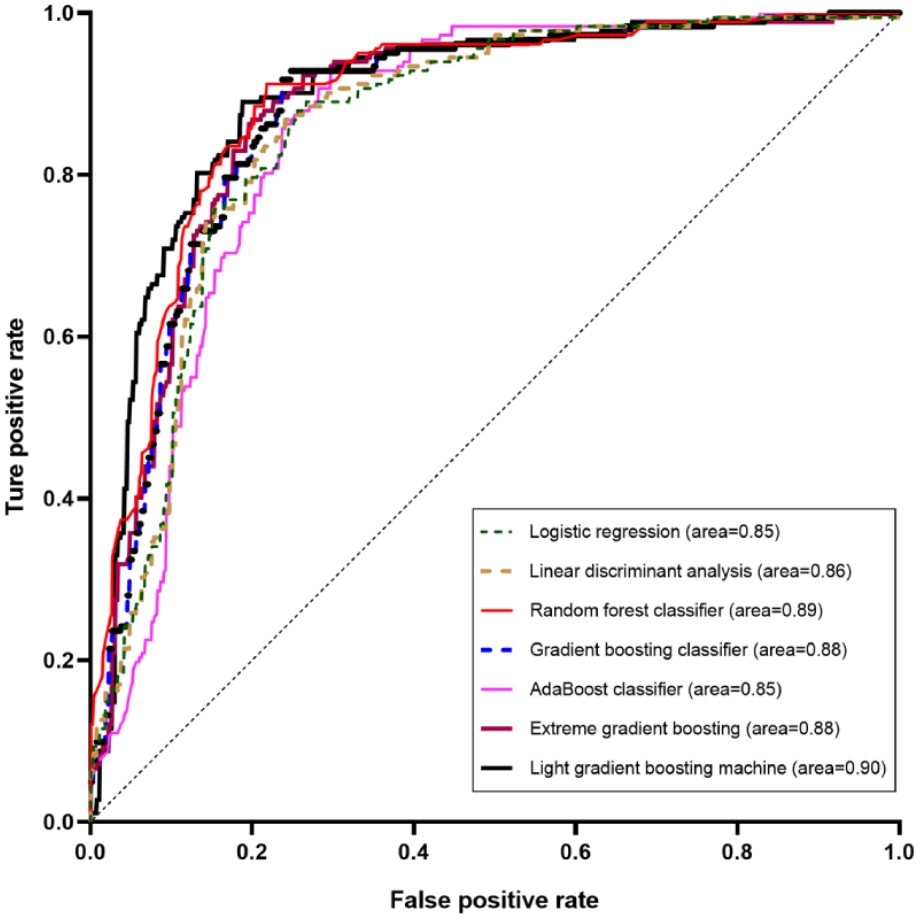

XSL•FO
RenderX

Moreover, the LGBM model surpassed the others in terms of testing accuracy, scoring a testing AUC of 0.90, highlighting its analytical efficiency. The XGBoost model also showed substantial potential for accurately predicting sepsis with an AUC of 0.88.

The GBC model showed a high PPV of 0.75, indicating its effectiveness in correctly identifying septic cases. Furthermore, the RFC and LGBM models also demonstrated strong predictive reliability with PPV of 0.73 and 0.76, respectively, reinforcing their utility for sepsis detection.

Brier score and BSS evaluations reveal that the RFC model achieved the best performance with the lowest Brier Score (0.13) and highest BSS (15.79%), substantially outperforming the baseline LR model with a Brier score of 0.15. The GBC, LGBM, and XGBoost models also showed strong performance, each with a Brier score of 0.13 and BSS values of 12.54%, 10.38%, and 14.33%, respectively. In contrast, the AdaBoost classifier had the highest Brier score (0.22) and a negative BSS (-47.29%), indicating poorer performance. These findings highlight that RFC, GBC, LGBM, and XGBoost models provide both accurate and well-calibrated probabilistic estimates for sepsis, as detailed in Table 2.

In Figure 3, the receiver operating characteristic curves for the various machine learning models are assessed in the study, including the LR, linear discriminant analysis, RFC, GBC, AdaBoost classifier, XGBoost, and the LGBM models. These curves graphically represent the true-positive rate versus the false-positive rate for each model, elucidating their efficacy in differentiating sepsis from nonsepsis cases. The AUC values were also provided to gauge the predictive accuracy of each model.

The receiver operating characteristic curves depicted in Figure 3 provide a direct comparison of the model performances. These curves plot the true-positive rate against the false-positive rate and identify the ensemble models (GBC, RFC, and LGBM) that are particularly effective in differentiating between sepsis and nonsepsis cases, making them valuable tools for ICU diagnostics.

In summary, our findings suggest that the GBC, RFC, and LGBM models, with their high AUC values, balanced performance, and strong PPV scores, are robust tools for early sepsis prediction and can substantially enhance the clinical workflow by enabling timely and accurate sepsis interventions.

Clinical Application of the Predictive Model

Following a comprehensive evaluation, the LGBM model was selected as the backbone of our AI-CDSS owing to its superior performance on key metrics such as the AUC, sensitivity, and specificity. A key advancement in clinical decision-making is the development of a web-based interface for health care professionals. This tool allows for the quick input of CBC data to obtain an immediate assessment of the sepsis risk. A screenshot of this user-friendly functional interface is shown in Figure 4, demonstrating its applicability in real-time clinical settings. This interface marks a significant step forward in applying our machine learning model for swift and reliable sepsis risk evaluation, thereby optimizing the diagnostic process in ICU settings.

Figure 4. Interface demonstration for web-based sepsis risk assessment. A screenshot of the innovative web-based interface developed in this study. This interface was designed to allow health care professionals to load complete blood count data and swiftly assess sepsis risk. AI-CDSS: artificial intelligence–clinical decision support system.

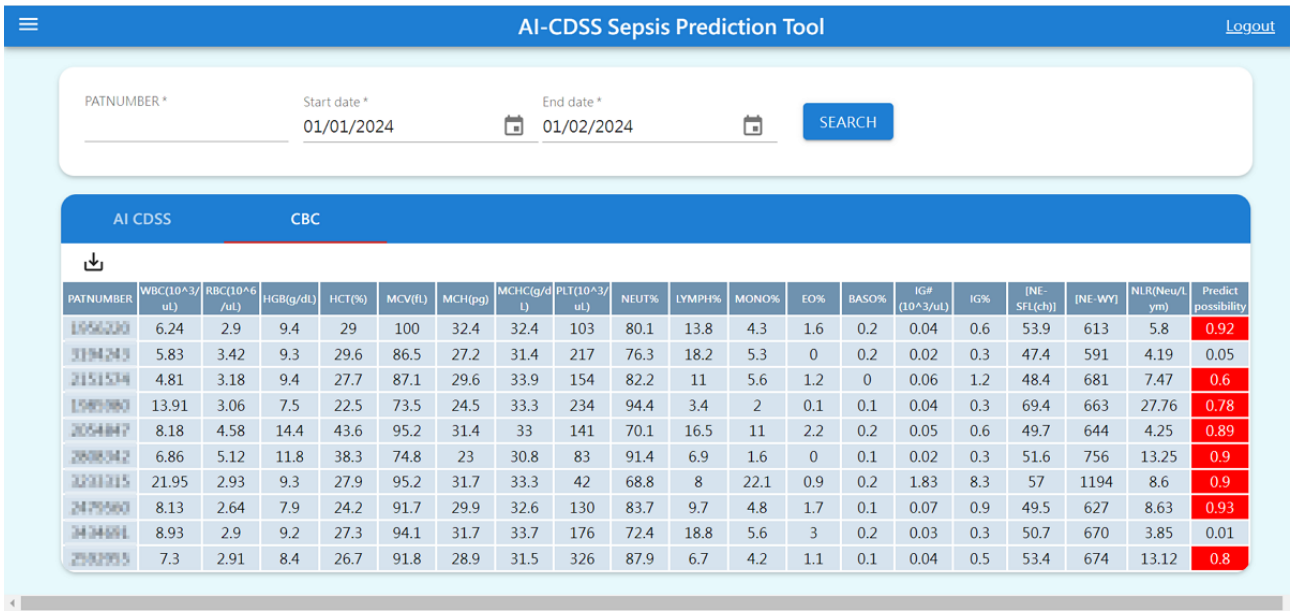

Discussion

Principal Findings

Our research contributes to advancements in rapid sepsis diagnostics through the development and validation of machine learning models that use CBC+DIFF data. The encouraging performance of the LGBM, GBC, and RFC models, as indicated by their AUC values during the testing phases, suggests a potential shift from traditional, more cumbersome diagnostic methods to a swift, data-driven approach using routine blood tests for the early detection of sepsis.

XSL•FO
RenderX

Importance of Early Sepsis Detection and Advanced Diagnostic Technologies

Early detection of sepsis is crucial to prevent progression to severe sepsis and septic shock, improving patient outcomes and reducing mortality. The mortality rate for septic shock (38.5%) is higher than that for sepsis (32.2%) [4], highlighting the need for early intervention and treatment for sepsis. Early intervention can significantly reduce the length of hospitalization; lower health care costs; and most importantly, improve the survival rates of patients [1,9]. Enhanced diagnostic accuracy can be achieved using new parameters that provide information on the morphological and functional characteristics of leukocytes, reflecting their activation and functional activity in response to infections [17]. In our study, we use the Sysmex XN-9100 system, which uses fluorescence flow cytometry along with blood-cell membrane surfactant reagents and fluorescent dyes targeting nucleic acids and proteins. This advanced technology provides detailed insights into the morphological characteristics of neutrophils, lymphocytes, and monocytes, further enhancing our ability to detect and understand sepsis at the cellular level.

Interpreting the Models' Effectiveness

The strong efficacy of LGBM, GBC, and RFC models underscores the potential of regular CBC+DIFF data as a valuable yet underexplored tool for early sepsis detection. The ability of these machine learning models to distinguish between septic and nonseptic states is particularly critical in ICU settings, where prompt diagnosis is of paramount importance [18], as multiple studies endeavor to detect early sepsis using machine learning models [15]. To further understand the contribution of each feature, we used heat map visualizations and SHAP values. The heat map analysis highlighted significant features such as neutrophil percentage, lymphocyte percentage, and neutrophil-to-lymphocyte ratio from the CBC+DIFF data, and SHAP values offered an additional perspective on how each feature influenced the model's predictions, enhancing clinical interpretability. In addition to the AUC, we evaluated the models using other metrics, including accuracy, sensitivity, specificity, PPV, NPV, and the F_1-score. The LGBM model outperformed the RFC and GBC models in these metrics, showcasing its superior overall performance. Furthermore, we assessed the models using the Brier score and BSS. The Brier score for the RFC, GBC, and LGBM models was 0.13, indicating similar accuracy among these models. However, the BSS showed that the RFC model had a slightly higher value of 15.79%, compared to the GBC model's 12.54% and the LGBM model's 10.38%, reflecting better performance relative to the baseline LR model, despite the LGBM model outperforming the RFC model in other metrics. Therefore, the LGBM model was integrated into our AI-CDSS backbone core owing to its overall performance.

Comparison to Previous Work

Our sepsis prediction model based on CBC+DIFF data provides a streamlined and efficient approach compared to existing models that require numerous physiological and clinical parameters from laboratory profiles, electronic health records, vital signs, and monitoring devices [13,14,19]. Unlike these complex models, our approach relies exclusively on readily available hematological data, making it highly accessible across various clinical settings, including non-ICU environments. Despite its simplicity, our model achieves an AUC of 0.90, surpassing the performance of more complex models that report AUCs ranging from 0.83 to 0.87, demonstrating that comprehensive predictive capability can be achieved using basic hematological parameters alone [13,14,19]. The model's simplicity offers significant advantages in terms of speed, cost-efficiency, and ease of implementation, particularly in settings where extensive data collection is challenging. By enabling rapid risk assessment through routine blood work, our model provides relative real-time decision support that facilitates efficient patient management across diverse health care settings, from emergency departments to resource-constrained environments.

Workflow Integration

Currently, our AI-CDSS web system is deployed on our hospital's internal server after validation. Physicians can log in to the system with their account, input the patient's medical record number, and select the relevant CBC+DIFF test values for analysis. The system analyzes the test data and provides an estimated probability of sepsis. This deployment allows for seamless integration into daily clinical workflows, enabling quick and accurate assessments that support timely and informed clinical decisions. Compared to traditional methods, even when physicians use the SOFA score to assess sepsis, they typically have to wait approximately 3 days to confirm a positive bacterial culture from blood or sputum samples. In contrast, our new AI-CDSS system can provide an earlier indication of infection within minutes, allowing physicians to confirm sepsis sooner and intervene earlier, thereby improving patient outcomes.

Clinical Relevance and Implementation

Considering the complex interplay of sepsis-related mortality and underlying diseases [20], the development of more precise diagnostic tools, such as our machine learning models, is essential to improve the early detection and management of sepsis in ICU settings. Our research, centered on the LGBM, GBC, and RFC, aligns with the broader trend of integrating diverse machine learning models in health care [21], reflecting the growing significance of artificial intelligence in enhancing sepsis diagnostics across various hospital settings [15]. Incorporating these models into ICU practices aligns with the emerging trends in health care [7] and could show potential to transform sepsis diagnostics, offering a rapid and precise alternative to traditional diagnostic approaches with only a minimally invasive blood draw required. The models' adoption is expected to enable early therapeutic intervention, thereby enhancing patient outcomes [15]. The creation of a user-friendly, web-based interface for health care professionals exemplifies the practicality of these models, facilitating real-time clinical decision-making and improving patient care.

In resource-limited countries, our study presents a potential innovation for sepsis management using machine learning for CBC+DIFF analysis, providing a rapid, effective alternative to traditional methods, essential where resource constraints cause life-threatening diagnostic delays [22]. Using routine blood tests, health care providers in these regions can rapidly identify sepsis risk, enabling quicker intervention and potentially

XSL•FO
RenderX

reducing the heavy dependence on limited antibiotics [23]. This method could assist in areas with limited access to advanced diagnostic facilities. Our study highlights the potential global impact of artificial intelligence–enhanced blood analysis, offering a promising approach to enhance sepsis diagnosis and improving patient outcomes in under-resourced settings.

Limitations

Although our findings are encouraging, there are some limitations to our study. The retrospective nature of data collection may introduce selection bias, and reliance on a singular data source (CBC+DIFF) could restrict the breadth of the models. In addition, this study was conducted at a single center, although supported by branches that refer to sepsis cases and focused exclusively on pneumonia-induced sepsis, which may limit the generalizability of the findings to other causes of sepsis. Although our results are promising, they are preliminary and require further confirmation through prospective studies in various clinical settings to verify their robustness and applicability.

Future Research

We recommend conducting extensive validation studies to confirm each model's generalizability and performing cost-benefit analyses to evaluate their impact on health care systems. As these models are integrated into clinical practice, it is essential to provide comprehensive training for health care professionals to ensure the optimal utilization of web-based tools. Expanding the model data inputs to encompass more patient-specific factors could further improve the accuracy of sepsis prediction.

Conclusions

In summary, our study contributes to advancements in sepsis diagnostics by exploring the use of machine learning to interpret CBC+DIFF data for enhanced sepsis diagnosis. This innovative approach may represent a significant shift in critical care and has the potential to impact diagnostics in various health care settings, including outpatient and emergency departments. By aiming to enable faster and more precise detection of sepsis, these models could significantly contribute to global health care efforts against sepsis, potentially aiding in improvements in patient care and outcomes. This approach may offer benefits in resource-limited countries, where diagnosing sepsis with only CBC+DIFF data provides a crucial alternative in scenarios where traditional diagnostic means are lacking.

Acknowledgments

Financial support for this research was provided by the Tri-Service General Hospital in Taipei, Taiwan, under grants TSGH_D_114102, TSGH_D_114101, MND-MAB-D-114203, and MND-MAB-D-113168. The sponsors did not participate in the design of the study, the gathering and analysis of data, or the determination to publish this research. Furthermore, the authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this study.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

FYC, CWC, and HSS contributed to conceptualization. CKC, CLP, HHL, CTY, SHT, and PCP assisted with data curation. THL, HYC, and MJJ contributed to formal analysis. HSS was responsible for funding acquisition. CKC, CLP, and HSS performed the investigation. FYC, CWC, and HSS helped with methodology. HSS contributed to project administration. FYC and HSS assisted with resources and writing–review & editing. HSS and CLP were responsible for supervision. THL, HYC, and MJJ performed validation. THL and HSS contributed to visualization. THL and HSS assisted with writing–original draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
SHAP summary plot for feature importance in the sepsis prediction model. SHAP: Shapley additive explanations.
[DOCX File , 176 KB - jmir_v27i1e56155_app1.docx]

References

1. Sepsis. World Health Organization. 2024 May 03. URL: https://www.who.int/news-room/fact-sheets/detail/sepsis [accessed 2025-01-01]
2. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. Lancet 2020;395(10219):200-211 [FREE Full text] [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31954465]

XSL•FO
RenderX

3. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Crit Care Med 2021;49(11):e1063-e1143. [doi: 10.1097/CCM.0000000000005337] [Medline: 34605781]
4. Bauer M, Gerlach H, Vogelmann T, Preissing F, Stiefel J, Adam D. Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019- results from a systematic review and meta-analysis. Crit Care 2020;24(1):239 [FREE Full text] [doi: 10.1186/s13054-020-02950-2] [Medline: 32430052]
5. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
6. Fleischmann C, Scherag A, Adhikari NKJ, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. Am J Respir Crit Care Med 2016;193(3):259-272. [doi: 10.1164/rccm.201504-0781OC] [Medline: 26414292]
7. Ackermann K, Baker J, Green M, Fullick M, Varinli H, Westbrook J, et al. Computerized clinical decision support systems for the early detection of sepsis among adult inpatients: scoping review. J Med Internet Res 2022;24(2):e31083 [FREE Full text] [doi: 10.2196/31083] [Medline: 35195528]
8. Rogers P, Boussina AE, Shashikumar SP, Wardi G, Longhurst CA, Nemati S. Optimizing the implementation of clinical predictive models to minimize national costs: sepsis case study. J Med Internet Res 2023;25:e43486 [FREE Full text] [doi: 10.2196/43486] [Medline: 36780203]
9. Zhu Y, Zhang R, Ye X, Liu H, Wei J. SAPS III is superior to SOFA for predicting 28-day mortality in sepsis patients based on Sepsis 3.0 criteria. Int J Infect Dis 2022;114:135-141 [FREE Full text] [doi: 10.1016/j.ijid.2021.11.015] [Medline: 34775116]
10. Gyawali B, Ramakrishna K, Dhamoon AS. Sepsis: the evolution in definition, pathophysiology, and management. SAGE Open Med 2019;7:2050312119835043 [FREE Full text] [doi: 10.1177/2050312119835043] [Medline: 30915218]
11. Andersson M, Östholm-Balkhed Å, Fredrikson M, Holmbom M, Hällgren A, Berg S, et al. Delay of appropriate antibiotic treatment is associated with high mortality in patients with community-onset sepsis in a Swedish setting. Eur J Clin Microbiol Infect Dis 2019;38(7):1223-1234 [FREE Full text] [doi: 10.1007/s10096-019-03529-8] [Medline: 30911928]
12. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. Crit Care Med 2006;34(6):1589-1596. [doi: 10.1097/01.ccm.0000217961.75225.e9]
13. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015;7(299):122-299. [doi: 10.1126/scitranslmed.aab3719] [Medline: 26246167]
14. Moor M, Bennett N, Plečko D, Horn M, Rieck B, Meinshausen N, et al. Predicting sepsis using deep learning across international sites: a retrospective development and validation study. EClinicalMedicine 2023;62:102124 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102124] [Medline: 37588623]
15. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]
16. Kyriazopoulou E, Karageorgos A, Liaskou-Antoniou L, Koufargyris P, Safarika A, Damoraki G, et al. BioFire FilmArray pneumonia panel for severe lower respiratory tract infections: subgroup analysis of a randomized clinical trial. Infect Dis Ther 2021;10(3):1437-1449 [FREE Full text] [doi: 10.1007/s40121-021-00459-x] [Medline: 34120316]
17. Aguirre U, Urrechaga E. Diagnostic performance of machine learning models using cell population data for the detection of sepsis: a comparative study. Clin Chem Lab Med 2023;61(2):356-365. [doi: 10.1515/cclm-2022-0713] [Medline: 36351434]
18. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng 2016;104(2):444-466 [FREE Full text] [doi: 10.1109/JPROC.2015.2501978] [Medline: 27765959]
19. Zhou L, Shao M, Wang C, Wang Y. An early sepsis prediction model utilizing machine learning and unbalanced data processing in a clinical context. Prev Med Rep 2024;45:102841. [doi: 10.1016/j.pmedr.2024.102841] [Medline: 39188971]
20. Rannikko J, Syrjänen J, Seiskari T, Aittoniemi J, Huttunen R. Sepsis-related mortality in 497 cases with blood culture-positive sepsis in an emergency department. Int J Infect Dis 2017;58:52-57 [FREE Full text] [doi: 10.1016/j.ijid.2017.03.005] [Medline: 28288925]
21. Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, et al. Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med 2020;104:101820 [FREE Full text] [doi: 10.1016/j.artmed.2020.101820] [Medline: 32498999]
22. Milton R, Gillespie D, Dyer C, Taiyari K, Carvalho MJ, Thomson K, BARNARDS Group, et al. Neonatal sepsis and mortality in low-income and middle-income countries from a facility-based birth cohort: an international multisite prospective observational study. Lancet Glob Health 2022;10(5):e661-e672 [FREE Full text] [doi: 10.1016/S2214-109X(22)00043-2] [Medline: 35427523]

XSL•FO
RenderX

23. Becker JU, Theodosis C, Jacob ST, Wira CR, Groce NE. Surviving sepsis in low-income and middle-income countries: new directions for care and research. Lancet Infect Dis 2009;9(9):577-582. [doi: 10.1016/S1473-3099(09)70135-5] [Medline: 19695494]

Abbreviations

AI-CDSS: artificial intelligence–clinical decision support system
AUC: area under the curve
BSS: Brier skill score
CBC+DIFF: complete blood count with differential
GBC: gradient boosting classifier
ICU: intensive care unit
LGBM: light gradient boosting machine
LR: logistic regression
NPV: negative predictive value
PPV: positive predictive value
RFC: random forest classifier
SHAP: Shapley additive explanations
SOFA: Sequential Organ Failure Assessment
XGBoost: extreme gradient boosting

Edited by N Cahill, T Leung; submitted 08.01.24; peer-reviewed by J Waterson, S Liu; comments to author 24.04.24; revised version received 26.05.24; accepted 04.12.24; published 24.01.25.

Please cite as:
Lin TH, Chung HY, Jian MJ, Chang CK, Lin HH, Yen CT, Tang SH, Pan PC, Perng CL, Chang FY, Chen CW, Shang HS
AI-Driven Innovations for Early Sepsis Detection by Combining Predictive Accuracy With Blood Count Analysis in an Emergency Setting: Retrospective Study
J Med Internet Res 2025;27:e56155
URL: https://www.jmir.org/2025/1/e56155
doi:10.2196/56155
PMID:

©Tai-Han Lin, Hsing-Yi Chung, Ming-Jr Jian, Chih-Kai Chang, Hung-Hsin Lin, Chiung-Tzu Yen, Sheng-Hui Tang, Pin-Ching Pan, Cherng-Lih Perng, Feng-Yee Chang, Chien-Wen Chen, Hung-Sheng Shang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Transformers for Neuroimage Segmentation: Scoping Review

Maya Iratni[1], MS; Amira Abdullah[1], MS; Mariam Aldhaheri[1], MS; Omar Elharrouss[1], PhD; Alaa Abd-alrazaq[2], PhD; Zahiriddin Rustamov[1], MS; Nazar Zaki[1], PhD; Rafat Damseh[1], PhD

[1]Department of Computer Science and Software Engineering, United Arab Emirates University, Al Ain, United Arab Emirates
[2]Weill Cornell Medicine-Qatar, Doha, Qatar

Corresponding Author:
Rafat Damseh, PhD
Department of Computer Science and Software Engineering
United Arab Emirates University
Sheik Khalifa Bin Zayed St - 'Asharij - Shiebat Al Oud - Abu Dhabi
Al Ain, 15551
United Arab Emirates
Phone: 971 37135586
Email: rdamseh@uaeu.ac.ae

Abstract

Background: Neuroimaging segmentation is increasingly important for diagnosing and planning treatments for neurological diseases. Manual segmentation is time-consuming, apart from being prone to human error and variability. Transformers are a promising deep learning approach for automated medical image segmentation.

Objective: This scoping review will synthesize current literature and assess the use of various transformer models for neuroimaging segmentation.

Methods: A systematic search in major databases, including Scopus, IEEE Xplore, PubMed, and ACM Digital Library, was carried out for studies applying transformers to neuroimaging segmentation problems from 2019 through 2023. The inclusion criteria allow only for peer-reviewed journal papers and conference papers focused on transformer-based segmentation of human brain imaging data. Excluded are the studies dealing with nonneuroimaging data or raw brain signals and electroencephalogram data. Data extraction was performed to identify key study details, including image modalities, datasets, neurological conditions, transformer models, and evaluation metrics. Results were synthesized using a narrative approach.

Results: Of the 1246 publications identified, 67 (5.38%) met the inclusion criteria. Half of all included studies were published in 2022, and more than two-thirds used transformers for segmenting brain tumors. The most common imaging modality was magnetic resonance imaging (n=59, 88.06%), while the most frequently used dataset was brain tumor segmentation dataset (n=39, 58.21%). 3D transformer models (n=42, 62.69%) were more prevalent than their 2D counterparts. The most developed were those of hybrid convolutional neural network-transformer architectures (n=57, 85.07%), where the vision transformer is the most frequently used type of transformer (n=37, 55.22%). The most frequent evaluation metric was the Dice score (n=63, 94.03%). Studies generally reported increased segmentation accuracy and the ability to model both local and global features in brain images.

Conclusions: This review represents the recent increase in the adoption of transformers for neuroimaging segmentation, particularly for brain tumor detection. Currently, hybrid convolutional neural network-transformer architectures achieve state-of-the-art performances on benchmark datasets over standalone models. Nevertheless, their applicability remains highly limited by high computational costs and potential overfitting on small datasets. The heavy reliance of the field on the brain tumor segmentation dataset hints at the use of a more diverse set of datasets to validate the performances of models on a variety of neurological diseases. Further research is needed to define the optimal transformer architectures and training methods for clinical applications. Continuing development may make transformers the state-of-the-art for fast, accurate, and reliable brain magnetic resonance imaging segmentation, which could lead to improved clinical tools for diagnosing and evaluating neurological disorders.

(J Med Internet Res 2025;27:e57723) doi:10.2196/57723

KEYWORDS

3D segmentation; brain tumor segmentation; deep learning; neuroimaging; transformer

XSL•FO
RenderX

Introduction

Neuroimaging refers to the visualization of the structure and function of the brain. It is one of the most important tools in the understanding of different neurological disorders. Generally, neuroimages can be obtained using 3 principal imaging modalities, where each modality shows the complexities of the brain from a different perspective. Of the 3, magnetic resonance imaging (MRI) is still the most frequently used due to high contrasting ability of brain tissues, high spatial resolution, and no risk of radiation exposure [1-3]. For different brain regions to be viewed, multiple MRI sequences are needed, such as T1, T1ce, T2, and fluid-attenuated inversion recovery, as presented in Figure 1 [4]. The second neuroimaging modality is computed tomography (CT), which can produce high-resolution images. On the other hand, it has limited soft tissue characterization, and its radiation risk makes it unsuitable for repetitive use [3,5]. The third neuroimaging modality is positron emission tomography (PET), which integrates nuclear medicine to visualize metabolic activity [2]. PET has high sensitivity, making it effective in detecting metastases, finding abnormalities, and imaging deep structures. However, it has limited resolution, and repeated use causes radiation risk [3,5]. Finding changes in brain tissue through neuroimaging analysis is critical for detecting and monitoring neurological disorders [6] and brain tumors [7]. Segmentation is a useful process in outlining regions of interest in medical images [8], which enables the quantitative assessment of atrophy, growths, and anatomical differences that depict conditions like Alzheimer disease, schizophrenia, and brain tumors among other neurodegenerative diseases [2]. Because of this, segmentation is applied broadly in different medical applications in diagnosis, tissue classification, radiotherapy treatment, and surgical planning [2,9].

Figure 1. Magnetic resonance imaging modalities of brain tumor. FLAIR: fluid-attenuated inversion recovery; T1: T1-weighted imaging; T2: T2-weighted imaging; T1ce: T1-weighted contrast-enhanced imaging.

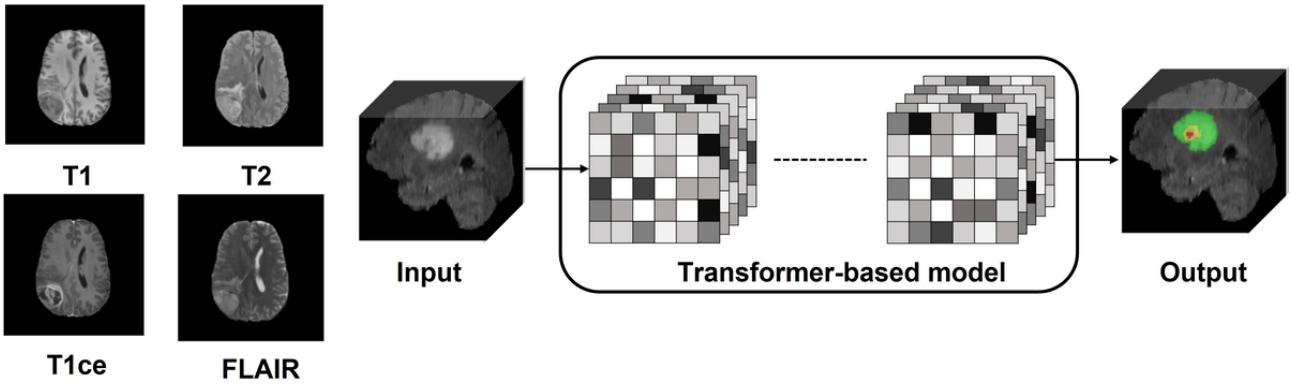

Segmentation techniques can be classified into 3 categories: manual, semiautomated, and fully automated. Manual segmentation is the standard for segmentation because it is believed to be the most accurate [10]. The technique, however, is laborious, time-consuming, and subjective, since it depends on human judgment, and this may result in variation in the results because of the different interpretations. Due to this, there has been a great deal of research into automated segmentation techniques to replicate the results from manual segmentation but with a higher level of efficiency and consistency [2]. To do this, 2 early paradigms were used: intensity-based approaches, which include thresholding, edge-detection, and region-based [8], and traditional machine learning paradigms, including support vector machine, k-nearest neighbor clustering, and random forest [8,11]. Each of these methods has been applied in 1 or more ways, but their applicability and performance within the task of image segmentation remain limited [11,12]. Since then, deep learning (DL) methods have transformed medical imaging applications and became a strong alternative to classical techniques.

DL is a subclass of machine learning that involves artificial neural networks with multiple layers. These networks are designed to progressively learn hierarchical representations and features of data, which both eliminates the need for manual feature engineering [2] and enables the extraction of complicated patterns from large datasets [13]. Different DL architectures have been used for medical image segmentation, but the most widely used and popular one is convolutional neural networks (CNNs), which have achieved state-of-the-art performances in different medical imaging tasks, including segmentation [12,14]. U-Net [15] is another notable model that was specially designed for biomedical image segmentation and has produced very good results in its field [16-18]. Some other notable models include SegNet [19], ResNet [20], DenseNet [21], 3D-ConvNet [22], and DeepLab [23]. These models have served as a solid foundation for the imaging field and have resulted in a plethora of variants, each developed for specific imaging modalities, anatomical structures, and segmentation tasks. Transformers [24] are a type of neural network architectures that mainly rely on self-attention mechanisms. They were first proposed in 2017 and have since yielded state-of-the-art results in the field of natural language processing [25]. More recently, transformers have also shown success when applied to a wide array of computer vision tasks, one of which is segmentation [26]. Although CNNs have achieved impressive performances in image-related tasks, they may not capture global and long-range dependencies well due to the small kernel size [26,27].

Transformers have recently gained popularity in imaging due to their self-attention mechanism, which can model these long-range dependencies—especially useful in brain segmentation [27]. The great success of transformers has motivated the construction of vision transformers (ViT) [28],

XSL•FO
RenderX

which forego the use of convolutional layers and rely instead on a multihead self-attention mechanism [13,29]. This architecture divides an image into fixed-size patches, linearly embeds them, and processes them through a transformer network, thereby allowing it to model long-range dependencies with reduced inductive bias [27,30]. Recently, ViT architectures specifically designed for medical image segmentation have been explored and resulted in models like TransUnet [31] and Swin-UNet [32] for general-purpose use and models like TransBTS [33] and Swin-UNETR [34] with the backbone for brain tumor segmentation [29,30]. Special study deserves transformer use in neuroimaging, as the structures of the brain are complicated. Neural networks based on transformers can model long-range dependencies and spatial relationships of the brain images [27], which is very important in brain segmentation.

Although transformers have shown very promising results in many medical imaging tasks, their use in neuroimaging segmentation remains an evolving field that had not been systematically reviewed. Existing literature reviews have either examined the use of transformers for general medical image segmentation without focusing specifically on brain segmentation [25,29,30] or have reviewed brain segmentation techniques using various DL methods without emphasizing the role of transformers [2,13]. One more difference that exists between this review and others is the focus on applying transformers to neuroimage segmentation, which is a central task in neurological disorder diagnosis and treatment. For example, compared with more general surveys such as those by Shamshad et al [29] and Xiao et al [26], which address a wide range of tasks or organ systems, our work specifically focuses on the unique challenges and developments within brain image segmentation. Thus, this scoping review will seek to fill the gap by focusing solely on transformer applications in neuroimage segmentation, an area of paramount importance for the diagnosis and treatment of neurological disorders. A scoping review on this topic is appropriate because the application of transformers in this area is relatively new and fast-developing; hence, it allows for comprehensively mapping the current research landscape and identifying knowledge gaps.

The main purpose of this scoping review is to synthesize and critically evaluate the existing literature on the use of different transformer models for neuroimaging segmentation. This review aims at summarizing the types of transformer models applied, their performance, applications in various neurological conditions and imaging modalities, limitations of the current literature, and highlighting the existing gaps in research.

Methods

Study Design

The approach of this scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Multimedia Appendix 1) [35]. Our primary research question was “What are the current applications, performance, and limitations of transformer models in neuroimaging segmentation?” The goal was the extraction of key themes within recent literature related to transformer use in neuroimaging segmentation that will guide future research and clinical applications. Only the literature starting from 2019 was considered, since the rapid evolution in the development of transformer models for medical imaging is a key recent development in this field.

We defined transformer models as DL architectures relying on self-attention mechanisms, capable of processing sequential data and capturing long-range dependencies. From the neuroimaging perspective, we considered those studies where these models were applied to different modalities of brain imaging, focusing on MRI due to its prevalence in neurological diagnostics.

Our review process followed a systematic search in 4 major databases: Scopus, IEEE Xplore, PubMed, and ACM Digital Library. We present a comprehensive review of methodologies, results, strengths, and limitations of the included studies to derive useful insights that bridge technical developments with their implications in the clinical domain of neuroimage segmentation.

Search Strategy

Studies were retrieved on May 22, 2023, through searching the following databases: IEEE Xplore, ACM Digital Library, Scopus, and PubMed. The search was limited to 5 years, from 2019 to 2023, to prioritize recent research and consisted of search queries related to transformers such as “transformer,” “deep learning,” and “self-attention”; queries related to neuroimaging such as “brain,” “neuroimaging,” “MRI,” “CT,” and “PET”; and queries related to the medical field such as “health care,” “medical,” “health.”

Study Eligibility Criteria

This review only included papers whose primary purpose was on the use of transformers for the segmentation of neuroimages. Our search included journal papers, conference papers, and dissertations that focused on applying transformer models to imaging scans (eg, MRI and CT) of the human brain. We excluded all studies that (1) used transformers for the segmentation of nonneuroimaging, raw brain signals, or electroencephalogram data; (2) were not in English, review papers, conference abstracts, preprints, protocols, and conference abstracts; (3) focused on neuroimaging tasks other than segmentation (eg, classification and prediction); and (4) were published before 2019.

Study Selection

For this review, we used a 3-step study selection process. First, we used EndNote (Clarivate) to remove duplicate studies returned by our initial search. Next, 3 independent reviewers (MI, AA, and MA) screened the titles and abstracts of the remaining papers to exclude irrelevant studies. We then obtained full texts of the studies that passed the initial screening, and the same 3 reviewers (MI, AA, and MA) examined them against our predefined inclusion criteria. Any disagreements between reviewers during the screening processes were resolved through in-person discussions until a consensus was reached.

XSL•FO
RenderX

Data Collection

The data extraction for this review is done in Microsoft Excel by 2 independent groups of 2 reviewers (MI and AA and MA and OE) to share the workload for extraction and resolve conflicts between the groups. Disagreements in data extraction were resolved through consensus during face-to-face discussions. Data extractions fall into 3 broad categories: study characteristics, neuroimaging acquisition, and transformer features.

Synthesis of included data was done using a narrative approach. Descriptive texts, tables, and figures describe and show the summary and characteristics of the data. Microsoft Excel was used to manage and synthesize the data. First, we depict the characteristics of each included study concerning publication year, type of publication, and country of origin. Then, we describe the neuroimaging acquisition of these studies by including the imaging modality, dataset, dataset accessibility, and neurological condition. Finally, it explains the transformer architecture of the included studies: the number of parameters, transformer type, hybrid component, and the training and evaluation methodology used along with loss function, optimizer, and metrics.

Ethical Considerations

This scoping review synthesized and analyzed publicly available research studies. No direct human participant research was conducted; therefore, approval from an institutional review board or a research ethics committee was unnecessary. There was no collection, use, or dissemination of personal data from human participants in this study. Data extracted and analyzed in this review had been sourced from published studies, previously subjected to ethical review processes as part of their original publication. The review followed ethics in research practices, in that the representation of the included studies was true to form, and the methodology for the selection of studies and extraction was transparent. No individual participant data were accessed or reported in this review; hence, privacy and confidentiality were ensured. Since this study did not involve direct contact with human participants, issues regarding informed consent and protection of privacy and compensation of participants were therefore not relevant in this study. No images or supplementary material showing identifiable information of any individual were used in this review.

Results

Overview

A total of 1246 publications were retrieved from the initial search of the selected databases. In the first round of screening, 261 duplicates were identified and removed through the use of EndNote X9, leaving 985 publications remaining. In the second round, 761 publications were excluded through the analysis of their titles and abstracts against our predefined inclusion or exclusion criteria. The remaining 224 publications continued to the third round of screening, which included a detailed full-text read-through and resulted in the exclusion of 156 publications. Of the 1246 initial publications, only 68 studies met our criteria and were thereby included in this review. Figure 2 depicts the full screening process in more detail.

Figure 2. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart of the study selection process.

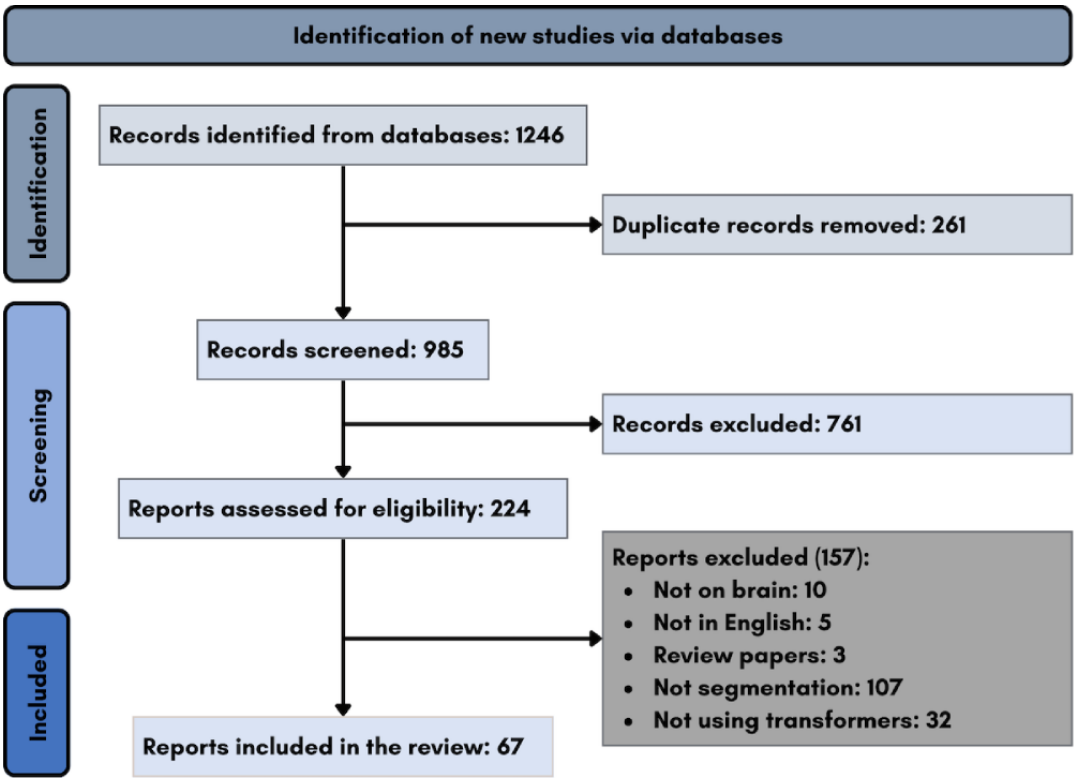

XSL•FO
RenderX

Characteristics of the Included Studies

Table 1 depicts the characteristics and metadata of each included study, including the publication year, country, and type. Included studies ranged between 2019 and 2023. Over half of the included studies were published in 2022, followed by 32.84% (n=22) in 2023. Studies included peer-reviewed papers (n=48, 71.64%) and conference papers (n=19, 29.36%). The included studies spanned a total of 13 countries, with China being by far the largest contributor in this domain, representing 68.66% (n=46) of the total studies. Following China, we can find the United States (n=5, 7.46%), the United Kingdom (n=4, 5.97%), and India (n=3, 4.48%), with other countries contributing 1 paper apiece.

Table 1. Characteristics of the studies used in this review, including the year, type, and country of publication.

Features	Studies, n (%)	References
Year of publication		
2023	22 (32.84)	[36-57]
2022	34 (50.75)	[58-91]
2021	10 (14.93)	[33,34,92-99]
2019	1 (1.49)	[100]
Type of publication		
Journal paper	48 (71.64)	[34,36-51,53-76,80,83,86,87,91,92,96]
Conference paper	19 (29.36)	[33,52,77-79,81,82,84,85,88-90,93-95,97-100]
Country of publication		
China	46 (68.66)	[33,37-42,44,46-48,50,54-67,69,72,74-77,79-84,86,88,90-94,96,99]
United States	5 (7.46)	[34,52,85,92,100]
United Kingdom	4 (5.97)	[53,68,71,97]
India	3 (4.48)	[49,95,98]
Other	9 (13.43)	[36,43,45,51,70,73,78,87,89]

Neuroimaging Acquisition and Neurological Condition

Table 2 depicts the different imaging modalities used, the datasets used, and the different neurological conditions across the included studies. The included studies included a range of 6 different modalities, with MRI being by far the most common with 88% (n=59), followed by CT with 10.45% (n=7), and the remaining modalities with 1 each. Over half of the included studies used only 1 dataset (n=40, 59.70%) for training and evaluation purposes, followed by 23.88% (n=16) using 2 datasets. Of the 44 unique datasets used across the included studies, 70.45% (n=31) are public or open-source datasets, and 29.55% (n=13) are private datasets obtained directly from medical institutions. Regarding the public dataset category, the brain tumor segmentation dataset (BraTS dataset) is by far the most widely used, with 58.21% (n=39) of total studies using its variants (including BraTS 2015, 2017, 2018, 2019, 2020, and 2021), followed by Medical Segmentation Decathlon and low-grade glioma-Kaggle with 5.97% (n=4) each. The main neurological condition of the included studies was the segmentation of brain tumors, with 71.64% (n=48) of studies conducting research specifically in this area.

XSL•FO
RenderX

Table 2. Description of features used in the included studies, including modalities, datasets, dataset types, and neurological conditions.

Features	Studies, n (%)	References
Imaging modality		
MRI[a]	59 (88.06)	[33,34,36-51,54-69,71-85,87,88,90,92,93,95-99]
CT[b]	7 (10.45)	[52,53,71,75,86,89,94]
PET[c]	1 (1.49)	[73]
Interventional ultrasound	1 (1.49)	[70]
Electron microscopy	1 (1.49)	[100]
Digital subtraction angiography	1 (1.49)	[91]
Number of datasets used		
1	40 (59.7)	[34,36,39,40,42-50,53,57,63-65,69,70,74,77,78,80,83-88,90-95,97-100]
2	16 (23.88)	[33,37,38,41,51,56,58,59,61,67,68,73,76,82,89,96]
3+	10 (14.93)	[54,55,60,62,66,71,72,75,79,81]
Not mentioned	1 (1.49)	[52]
Dataset accessibility		
Public	31 (70.45)	__[d]
Private	13 (29.55)	—
Dataset		
Public: BraTS[e]	39 (58.21)	[33,34,37,40-42,44-46,48,50,51,54,56,60-62,64-66,68,69,71,72,74,76,77,79-83,90,92,93,95,97-99]
MSD[f]	4 (5.97)	[39,62,85,88]
iseg-2017	3 (4.48	[58,59,96]
ADNI[g]	2 (2.99)	[43,73]
MRBrainS[h]	2 (2.99)	[59,96]
LGG[i]-Kaggle	4 (5.97)	[49,63,66,84]
ISLES[j]	3 (4.48)	[51,54,75]
WMH[k]	3 (4.48)	[71,78,81]
Other	12 (17.91)	[55,58,68,70,71,73,75,81,87,89,91,100]
Private	9	[36,38,47,53,57,67,82,86,94]
Not stated	1 (1.49)	[52]
Neurological condition		
Brain tumor	48 (71.64)	[33,34,37-42,44-46,48-51,54,56,60-72,74,76,77,79-85,88,90,92,93,95,97-99]
Ischemic stroke	4 (5.97)	[51,52,75,94]
Alzheimer disease	3 (4.48)	[43,73,81]
Parkinson disease	2 (2.99)	[36,57]
Intracerebral hemorrhage	3 (4.48)	[53,86,89]
Intracranial aneurysms	1 (1.49)	[91]
Autism	1 (1.49)	[55]
Brain lesions	1 (1.49)	[78]
Healthy brain	6 (8.96)	[47,58,59,87,96,100]

[a]MRI: magnetic resonance imaging.

[b]CT: computed tomography.

[c]PET: positron emission tomography.

XSL•FO
RenderX

[d]Not available.

[e]BraTS: brain tumor segmentation dataset.

[f]MSD: Medical Segmentation Decathlon.

[g]ADNI: Alzheimer's Disease Neuroimaging Initiative.

[h]MRBrainS: magnetic resonance brain image segmentation.

[i]LGG: low-grade glioma.

[j]ISLES: ischemic stroke lesion segmentation.

[k]WMH: white matter hyperintensities.

Transformer-Based Techniques Types, Training Parameters, and Evaluation

The proposed neuroimage segmentation techniques used various artificial intelligence (AI) techniques. In this review, we focused on the deep transformer–based techniques that have gained more attention recently. From the proposed models, we can find transformer-based, CNN with transformer-based, and generative adversarial network with transformer-based techniques. At the same time, the methods based on TransBTS, TransUNet, SeinUNet, and U-Net with transformer are the most used models for neuroimage segmentation. Figure 3 illustrates these models in terms of architecture. Table 3 depicts the characteristics of transformer models used within the included studies. From Table 3, we can find that 58.21% (n=39) of the included studies did not explicitly report the number of parameters of their proposed models. Of the studies that did, however, the majority of the transformer models proposed had between 20 and 40 million parameters (n=10, 14.93%), followed by 1 and 19 million (n=8, 11.94%). A majority of studies implemented a 3D segmentation network (n=42, 62.69%), with 37.31% (n=25) being 2D. An overwhelming 85.07% (n=57) of included studies proposed transformer models that are hybrid, with only 14.93% (n=10) of them being standalone transformer models. ViT was the most used transformer architecture, with 55.22% (n=37) of studies using it as its main component. Another significant transformer model is the Swin transformer, with 20.89% (n=14), followed by TransUnet, with 5.97% (n=4). Of the 57 hybrid transformer models, 55 (96.49%) studies opted for a combination of CNN with their transformer, and of those 55 CNN-transformer models, 56.36% (n=31) were U-Net based, and 9.09% (n=5) were ResNet based. Both generative adversarial network (n=2, 3.51%) and autoencoders (n=2, 3.51%) were also combined with transformers.

Figure 3. Various transformer-based architectures used for neuroimage segmentation. (A) U-Net+transformer, (B) TransBTS, (C) TransUNet, (D) SwinUnet, (E) UNETR, and (F) transformer.

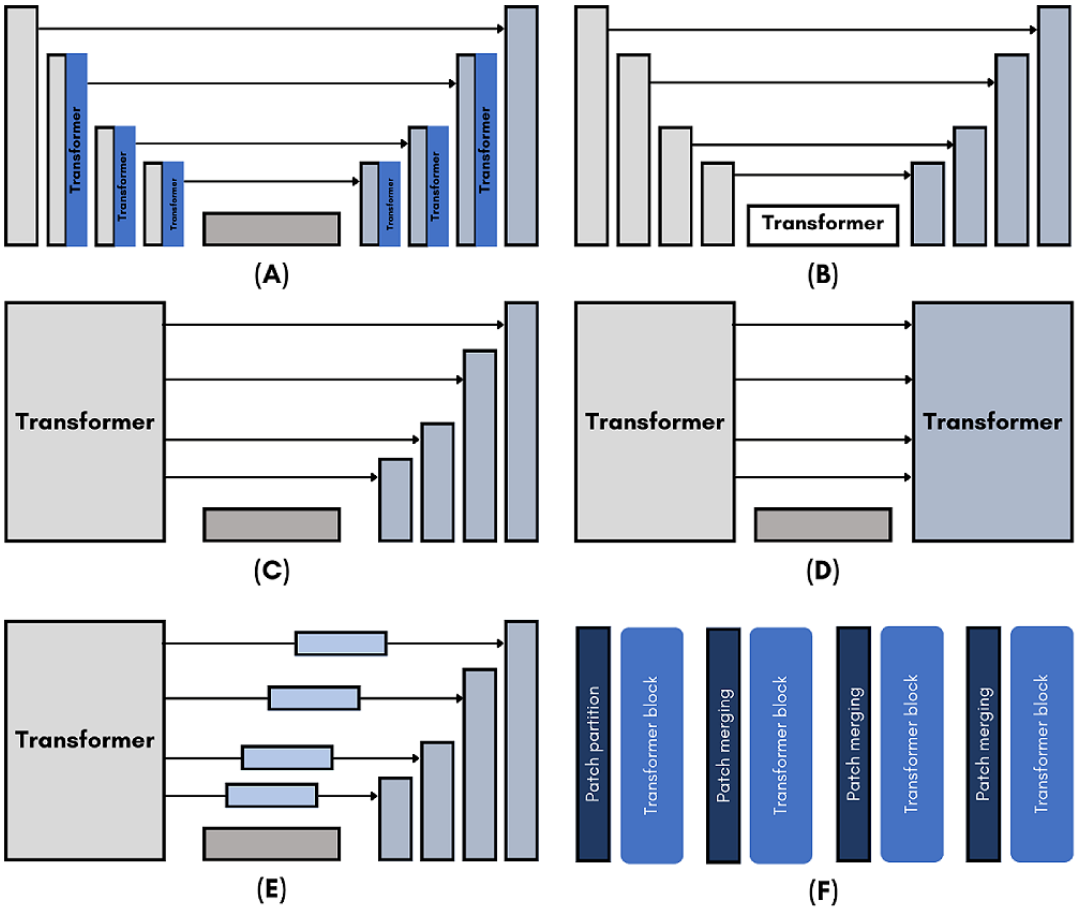

XSL•FO
RenderX

Table 3. Proposed methods based on the weight of the models, type of data used, and type of transformer technique used.

Features	Studies, n (%)	References
Number of parameters (in millions)		
1-19	8 (11.94)	[41,42,45,51,68,76,79,82]
20-39	10 (14.93)	[33,37,39,56,58,62,63,65,69,88]
40-59	4 (5.97)	[42,52,60,86]
60-100	3 (4.48)	[34,47,85]
100-120	3 (4.48)	[53,80,97]
120+	2 (2.99)	[42,54]
Not mentioned	39 (58.21)	[36,38,40,43,44,46,48-50,55,57,59,61,64,66,67,70-75,77,78,81,83,84,87,89-96,98-100]
Dimensionality		
2D	25 (37.31)	[40,43,45,48,49,53,54,58,63,70,75,77,78,84,86,88,89,91,94,99,100]
3D	42 (62.69)	[33,34,36-39,41,42,44,46,47,50-52,55-57,59-62,64,69,71-74,76,79-83,85,87,90,92,93,95-98]
Transformer model		
Standalone	10 (14.93)	[38,42,54,55,61,75,78,82,89,100]
Hybrid	57 (85.07)	[33,34,36,37,39-41,43-53,56-60,62-74,76,77,79-81,83-88,90-99]
Type of transformer		
ViT[a]	37 (55.22)	[33,36,38,40,44,46,47,49-55,57-60,65,66,71,73,75,81-83,85,87,88,90,91,93,95-99]
Swin	14 (20.89)	[34,37,42,43,56,62,67-69,72,74,76,77,89]
SwinUnet	2 (2.99)	[61,78]
TransUnet	4 (5.97)	[45,70,84,94]
TransBTS	2 (2.99)	[64,92]
Other	8 (11.94)	[39,41,48,63,79,80,86,100]
Type of hybrid component		
CNN[b]	55 (96.49)	[33,34,36,37,39-41,43-53,56-60,62-70,72-74,76,77,79,83-88,90-99]
U-Net	31 (56.36)	[34,36,37,40,43,45,46,48-51,53,56,57,60,62-64,70,76,83-86,88,90-94,97]
ResNet	5 (9.09)	[59,60,66,67,94]
GAN[c]	2 (3.51)	[43,60]
Autoencoder	2 (3.51)	[71,81]

[a]ViT: vision transformer.

[b]CNN: convolutional neural network.

[c]GAN: generative adversarial network.

Table 4 depicts the loss function used, the optimizer used, and the different evaluation methods used across each included study. The loss function was not mentioned in 11.94% (n=8) of the included studies. Of the studies that mentioned it, the most popular loss function is a combination of cross-entropy and Dice loss with 40.30% (n=27) of included studies, followed by Dice loss with 19.40% (n=13). Adam is the most used optimizer, with 47.76% (n=32) of included studies using it, followed by AdamW at 14.92% (n=10). However, the optimizer was not mentioned in 22.39% (n=15) of studies. In terms of evaluation, over half of the included studies used at least two evaluation metrics (n=34, 50.75%), followed by 1 metric (n=11, 16.42%). Of these evaluation metrics, the Dice score is by far the most used, with 94.03% (n=63) of all studies using it, followed by HD95, 52.24% (n=35), and sensitivity, 28.36% (n=19).

XSL•FO
RenderX

Table 4. Experimental setups and evaluation for the proposed transformer-based techniques.

Features	Studies, n (%)	References
Loss function1		
Dice loss	13 (19.4)	[34,37,38,41,45-49,59,79,80,87]
Cross-entropy	9 (13.43)	[43,44,50,61,63,65,77,90,96]
Dice cross-entropy	27 (40.3)	[39,40,42,51,54,56,57,60,62,69,72,76,78,81-86,91,93-95,97-99]
Other	10 (14.93)	[33,52,53,55,58,66,68,71,74,89]
Not mentioned	8 (11.94)	[36,67,70,73,75,88,92,100]
Optimizer		
Adam	32 (47.76)	[33,37,40,41,44,49,53,55,60-62,65-69,72,75,76,78,80,86-88,91-93,95,97-99]
AdamW	10 (14.92)	[42,48,51,52,56,74,81,82,85,89]
SGD[a]	7 (10.45)	[47,50,54,57,83,94,97]
Ranger	2 (2.99)	[38,79]
RMSprop[b]	1 (1.49)	[63]
Apollo	1 (1.49)	[64]
Not mentioned	15 (22.39)	[34,36,39,43,45,46,58,59,70,71,73,84,90,96,100]
Evaluation metrics		
Dice score	63 (94.03)	[33,34,36-42,44-66,68-72,74-99]
HD95[c]	35 (52.24)	[33,34,40-42,44-48,50,51,54,56-58,61,62,64,65,72,74,76,78,79,81,82,85,90,92,93,95,97-99]
Recall or sensitivity	19 (28.36)	[43,44,46,47,53-55,57,60,64,69,75,83,86,87,89,90,93,94]
IoU[d]	12 (17.91)	[45,49,52-54,63,66,68,77,87,89,91]
Precision	10 (11.94)	[43,44,47,53,55,60,75,87,89,90]
Accuracy	6 (8.96)	[43,49,55,67,73,86]
Specificity	5 (7.46)	[47,86,89,90,93]
AUC[e]	5 (7.46)	[43,67,71,89,100]
F-measure	3 (4.48)	[43,78,89]
Jaccard index	4 (5.97)	[44,53,57,90]
Other	5 (7.46)	[47,63,66,78,91]

[a]SGD: stochastic gradient descent.

[b]RMSprop: root mean square propagation.

[c]HD95: Hausdorff distance at the 95th percentile.

[d]IoU: Intersection Over Union.

[e]AUC: area under the curve.

Strengths and Limitations of Transformer-Based Techniques

Transformers have revolutionized the area of neuroimage segmentation by offering unparalleled capabilities in modeling complex features in medical imaging. It has the ability to model both local and global information, which substantially improves the accuracy of segmentation and therefore becomes very useful in various neurological applications. As shown in Table 5, the common strengths of transformer-based techniques include a high mean Dice score, effective fusion of multimodal MRI, and robust performance across diverse and complex datasets. However, these models also have substantial limitations in terms of high computational and memory costs, sensitivity to small areas of tumors, and possible overfitting on smaller datasets.

XSL•FO
RenderX

Table 5. Strengths and limitations of common transformer-based techniques.

References	Strengths	Limitations
ViT[a]		
[33,36,38, 40,44,46, 47,49-55, 57-60,65, 66,71,73, 75,81-83, 85,87,88, 90,91,93, 95-99]	• Effectively models local and global features in 3D MRIb data. • High mean Dice score. • Demonstrates consistent improvements in segmentation performance. • Effective in emphasizing informative brain regions. • Uses symmetry of brain structures for improved feature learning. • Outperforms state-of-the-art SSLc methods and medical image segmentation models on benchmarks. • Incorporates gradient-based scoring for attentive reconstruction. • Effective multimodal MRI fusion. • Enhanced long-term dependencies within individual modalities. • Complementary contextual information among modalities.	• Computationally intensive. • High computational and memory cost. • Memory constraints when lowering patch resolution. • Complexity in integrating CNNd and transformer features. • Sensitivity to small tumor areas in LGGe. • Overlap in feature dimensions between CNN and transformer branches. • Misclassification of voxels in LGG. • Imbalance in dataset affecting performance. • Need for extensive validation on more diverse datasets.
Swin		
[34,37,42, 43,56,62, 67-69,72, 74,76,77, 89]	• Capable of learning multiscale contextual information, enhancing performance across various tasks. • Combines advantages of ViT and CNNs, balancing both local and global feature learning. • Maintains high-resolution features, crucial for precise segmentation tasks. • Performs efficient tri-level preprocessing, including noise removal improving input quality for better results. • Incorporates advantages of 3D Swin transformer, improving performance in 3D medical image analysis.	• Slight decrease in performance for specific areas like tumor core segmentation in some instances. • ViTs have many parameters and structures, making them complex and resource-intensive. • Potential risk of overfitting on smaller datasets due to high model complexity. • Existing neural network algorithms may often extract redundant features, reducing overall efficiency. • Limited exploration in preprocessing and postprocessing techniques, which might enhance model performance further.
SwinUnet		
[61,78]	• Improves the efficiency of using limited labeled data. • Competitive performance in Dice score, Hausdorff distance, and other segmentation metrics.	• Requires verification of the improvement with diverse and larger datasets. • Lower recall and F1-scores compared to other CNN-based methods.
TransUnet		
[45,70,84, 94]	• High effectiveness in model design. • Combines the strengths of U-Net and transformer models. • Achieves higher Dice scores compared to U-Net and transformer. • Effectively learns global context features in images.	• Some models within the study showed poor performance. • Cross-application limitations in certain scenarios. • High data requirements for effective training. • Limited dataset size affecting generalization.
TransBTS		
[64,92]	• Residual basis blocks reduce feature loss and enhance feature extraction. • Combines CNN and transformer for improved segmentation performance, leveraging both local and global information. • Attention mechanisms enhance the model's ability to focus on relevant features, improving accuracy.	• Increased network depth leads to higher parameter counts, increasing computational requirements. • May struggle with unseen patterns in the testing phase, affecting performance robustness.
Other transformer types		
[39,41,48, 63,79,80, 86,100]	• Effectively capture both local and global context in medical images. • Enhance segmentation accuracy. • Robust across diverse and complex datasets. • Improve semantic information representation. • Combine multiresolution information effectively.	• High computational complexity. • Significant memory use. • Requires substantial hardware resources. • Extensive data preprocessing is needed. • Limited scalability to very large datasets. • Sensitive to variations in input data quality.

[a]ViT: vision transformer.

[b]MRI: magnetic resonance imaging.

[c]SSL: self-supervised learning.

[d]CNN: convolutional neural network.

[e]LGG: low-grade glioma.

XSL•FO
RenderX

Discussion

Principal Findings

The main purpose of this scoping review is to conduct a thorough investigation into the use of different transformer models in the field of neuroimaging, specifically segmentation. From the gathered data, it is clear that the use of transformers in neuroimaging experienced a great boost in research from 2021 to 2022, with over half of the included studies being published in 2022 compared to only 10 studies in 2021. It is also important to note that for the year 2023, only the studies up to May 22 were included; yet, this constitutes a total percentage of the included studies of 32.84% (n=22) and could very well be even higher when the whole year is considered.

From the studies included in this review, it is clear that MRI is by far the most popular image modality for applying transformer models to neuroimaging segmentation. This can be attributed to how common the use of MRI is in the diagnosis of neurological illnesses, especially for brain tumors [27], wherein it is able to provide functional, structural, and metabolic information [27] through the use of its different modalities (T1, T2, T1ce, and fluid-attenuated inversion recovery). MRI is particularly suitable for neuroimaging segmentation purposes because of the high spatial resolution and soft tissue contrast, both being critical for any form of precise segmentation it exhibits, since it is able to show good detailed visualization of structures in the brain and distinction between different tissues of the brain sizes [1,2].

Another reason for the popularity of MRI in the included studies is the availability of brain MRI scans sourced from the widely used BraTS datasets [101]. This yearly and open-source dataset contains a wide variety of different MRI modalities that are manually annotated, making it a very important resource for developing and benchmarking segmentation methods based on different transformer models. This is why it is no surprise that it is by far the most used dataset in the included studies.

When it comes to neurological conditions, a majority of included studies in this review focused on the use of transformers in brain tumor segmentation. This can be attributed to multiple factors, including the availability of MRI scans from the BraTS dataset that are specifically for brain tumor segmentation. Brain tumors are also highly prevalent among all ages and have a high fatality rate [102], making it a prime area for research into new methods of diagnosis and treatment. In addition, brain tumors are fairly complex and irregular in both location and shape [27], which makes manual segmentation a very tedious and time-consuming process that would benefit greatly from increased research into more automated methods for segmentation. The BraTS dataset is also a factor, as it provides a large variety of MRI scans that are specifically for brain tumor segmentation. Transformers are particularly useful for brain tumor segmentation due to their self-attention mechanism, which allows them to account for different variations in tumor characteristics, such as size and shape, during the segmentation process [26].

Most included studies proposed and developed models with 3D segmentation networks, specifically for 3D imaging data. In terms of neuroimaging, 3D scans are more common in part due to the 3D nature of MRI scans. Since MRI is the most common imaging modality used in neuroimaging, it makes sense that it is preferable to develop models for 3D imaging data in order to avoid the loss of information. Even though 3D models are typically more accurate for 3D imaging segmentation, they are computationally expensive [26], which is why some proposed models in the included studies chose to instead extract 2D slices from 3D imaging data. While this technique is suitable, reducing a 3D scan into 2D slices can lead to a degradation of volume and spatial characteristics native to 3D data [103].

CNN-transformer hybrid models were used far more than standalone transformer models in the included studies, specifically in the form of a U-Net and transformer combination. These combinations capitalize on the strengths of both CNN and transformers while minimizing their weaknesses. CNN is particularly useful in extracting local features and spatial information from the provided scans; however, it often struggles to capture long-range dependencies due to its small kernel size [26,27]. On the other hand, transformers are able to model these long-range dependencies due to their self-attention module, making them very useful for neuroimaging segmentation, especially in the case of brain tumors [27]. This is why most included studies opted to use the use of CNN to capture local features and transformers to capture global features to increase the performance of their models in the task of segmentation [102].

Research and Practical Implications

This scoping review provides an overview of the available research regarding the use of transformers in the context of neuroimaging segmentation. These findings underline important implications for future research and applications in this area.

It is a notable finding of this review that many studies apply transformers, specifically to brain tumor segmentation, which might hint at the potential of transformers in assisting diagnosis and treatment planning in this field. As shown here, transformers are well-suited for this task. However, further research is needed to assess the real-world clinical usefulness of transformers for brain tumor segmentation. While brain tumors are an important challenge, the focus on this single application at this level would seem indicative of the current lack of large and good-quality datasets in many other big neurological diseases and conditions, such as Alzheimer and Parkinson disease, and strokes. Making publicly available manually annotated datasets of different neurological conditions would motivate new research and developments on the application of transformers in this field. On top of this, the heavy reliance of studies on the BraTS dataset shows that there is a need to diversify datasets in order to validate different models correctly. Most of the included studies favored hybrid use by combining CNN and transformers, which illustrates the complementary strengths of these architectures for neuroimaging segmentation. Success in hybrid techniques shows that further exploration of novel integrations between transformers, CNNs, and other modules could become a promising direction to achieve better performances on more complex medical image analysis problems. Improved accuracy in neuroimaging segmentation, through the ability of transformer

XSL•FO
RenderX

models to extract local and global features, allows for more accurate identification of neurological conditions such as brain tumors. This will provide earlier diagnosis and treatment. Moreover, automation with these models will save much time of the clinicians in performing manual segmentation so that they can concentrate on the care of patients and other important tasks. Treatment planning may also be improved with transformer models, where the potential for more accurate and consistent segmentation results helps a lot in this respect. Moreover, these models can also potentially be integrated into clinical workflows without much hassle by developing user-friendly interfaces and collaboration between AI researchers and clinicians to ensure these tools are adopted and effectively used in practice.

Strengths and Limitations

This scoping review has numerous key strengths with regard to the analysis of transformer applications in neuroimaging segmentation. First, it gives a broad overview of the fast-evolving field by capturing recent works from 2019 through 2023. Second, it allows focusing on current research so that the review reflects the state-of-the-art in transformer applications for medical imaging. It is a systematic approach, covering 4 major databases; hence, wide and comprehensive coverage of the literature reviewed. The inclusion of journal papers and conference papers facilitates a wide view of both consolidated and emergent research. Third, this review gives elaborate insights into various aspects of transformer use in neuroimaging: imaging modality, dataset, neurological condition, and metric for performance evaluation. This level of analysis provides rich information relevant to both researchers and practitioners within the field. Finally, the review's focus on brain tumor segmentation, while a limitation in some respects, also serves as a strength by providing an in-depth look at transformer applications in a critical area of medical imaging with significant clinical implications.

While this scoping review offers a number of strengths, its limitations need to be acknowledged so as to strike a balance. First, the review was on transformers in neuroimaging segmentation alone, excluding other medical imaging tasks or organs. This narrow focus allows for an in-depth analysis of transformer applications in brain imaging but may not be representative of the full spectrum of use that transformers have seen in medical imaging. This limitation could be reduced by expanding the scope of future reviews to multiple organ systems or imaging tasks, giving a wider look at transformer applications in medical imaging.

Second, the review was focused on studies published in the English language, published from 2019 up to 2023. This narrowing was necessary, as most current works are favored in this novelty area of transformer use in medical imaging. In so doing, this review criterion may have left out important non-English language publications or early applications of transformers. This is likely a limitation in the representation of research trends worldwide. In this respect, future studies can be designed to include more languages, also extending the date range to capture more diverse sets of publications and track the evolution of transformer use in medical imaging over an extended period.

Third, the fact that 58.21% (n=39) of the works included in this review were based on the BraTS dataset introduces a certain bias in the domain toward the segmentation of brain tumors. Though it is a very critical area, it might not be useful to represent transformers completely for other neurological conditions. Future research needs to give more emphasis to developing and publicly releasing manually annotated datasets about more neurological conditions to address this limitation. This will further encourage diverse applications of transformers in neuroimaging and provide a wider understanding of the capability of transformers across different pathologies.

The review demonstrated a high dominance by studies from China, with 46 (68.66%) studies of the total (see Multimedia Appendix 2 for detailed analysis). This aligns with broader publication patterns in AI research where Chinese institutions contribute approximately 40% of global publications. While this distribution reflects documented trends in international research output, future reviews might benefit from more diversified search strategies to ensure comprehensive coverage of global research activities in this field.

Finally, no formal quality or risk-of-bias assessment of the included studies was performed. Although this represents a common approach when it comes to scoping reviews, this limits the degree to which strong conclusions can be drawn about the relative effectiveness of various approaches to transformers. Future systematic reviews or meta-analyses may involve quality assessments to support more robust evidence in terms of the efficacy of transformer models in neuroimaging segmentation.

Future Directions

These findings point to a variety of promising directions for future research on the application of transformers to neuroimaging segmentation. First, future studies should develop novel integrations between transformers, CNNs, and other advanced modules that will further improve performance for complex medical image analysis tasks. This might be achieved by investigating various hybrid models leveraging the strengths of transformers and more traditional DL methods. Second, the extension of transformer applications to more neurological conditions other than brain tumors, which would allow a wider grasp of the potential capability of transformers across different pathologies. More clinical applications are likely to follow from here. Third, the development of new transformer-based methods or their combination with emerging techniques like diffusion models could further improve efficiency and robustness for both 2D and 3D brain segmentation. Fourth, future studies shall be done to bridge the current limitations in dataset diversity. This may be in creating and publishing manually annotated datasets for a wider range of neurological conditions that can enable transformers to apply to neuroimaging in more diversified ways. Finally, the translation of research findings into clinical practice remains a high unmet need. This transition will require extensive validation of transformer models on diverse, real-world datasets and close collaboration between AI researchers and clinicians. Such collaboration could result in the development of more clinically relevant models and user-friendly interfaces, which

XSL•FO
RenderX

would expedite the translation of these advanced technologies into routine clinical practice.

Conclusions

This scoping review has thoroughly investigated the applications of transformers in neuroimaging segmentation and discovered a highly evolving field with great potential. The results of this paper have shown that transformer models, especially combined with CNNs in hybrid architectures, are also very promising for the task of brain MRI segmentation. Some of the big advantages of transformers include the modeling of long-range dependencies in images through self-attention mechanisms while still being able to perform local feature extraction. Such a combination uniquely allows for more accurate and detailed segmentation in highly complex neurological pathologies, like brain tumors.

There is clearly a trend toward 3D transformer models and hybrid CNN-transformer architectures, dominated by ViT as the variant of transformer used most frequently. These approaches also obtain superior performance on benchmark datasets, such as brain tumor segmentation tasks. However, reliance on the BraTS dataset highlights a requirement for more diverse data sources to ensure that performance could be validated across more multiple neurological conditions.

While this is promising, there are still important issues in the field: high computational costs associated with transformer models, overfitting on smaller datasets, and validation in larger clinical settings. Another issue is the geographical concentration of research output that highlights the need for greater diversity in the origins of studies worldwide to improve the generalizability of findings.

The future prospect of transformer models will unlock the potential that neuroimaging segmentation demands. Refining both architectures and training methods and integration into clinical workflows, transformations may provide state-of-the-art for fast, accurate, and reproducible brain-MRI segmentation, hence advancing clinical diagnosis and evaluation techniques for a better outcome in regard to patients with neurological disorders.

Although transformers have shown great improvement in neuroimaging segmentation, much potential is yet to be realized. Future work will need to be focused on present limitations, the extension of applications across a wider range of neurological conditions, and narrowing the gap between research and clinical practice to ensure that transformers are a valuable and impactful technology in medical imaging analysis.

Acknowledgments

This work was supported in part by the United Arab Emirates University (UAEU grant 12T037) and in part by the Big Data Analytics Center (UAEU grant 12R239).

Data Availability

All data generated or analyzed during this study are included in this published paper and Multimedia Appendix 1.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 84 KB - jmir_v27i1e57723_app1.docx]

Multimedia Appendix 2
Significant prevalence of studies published in China.
[DOCX File , 104 KB - jmir_v27i1e57723_app2.docx]

References

1. Abd-Ellah MK, Awad AI, Khalaf AAM, Hamed HFA. A review on brain tumor diagnosis from MRI images: practical implications, key achievements, and lessons learned. Magn Reson Imaging 2019;61:300-318. [doi: 10.1016/j.mri.2019.05.028] [Medline: 31173851]
2. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. J Digit Imaging 2017;30(4):449-459 [FREE Full text] [doi: 10.1007/s10278-017-9983-4] [Medline: 28577131]
3. Shoeibi A, Khodatars M, Jafari M, Ghassemi N, Moridian P, Alizadehsani R, et al. Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: a review. Inf Fusion 2023;93:85-117. [doi: 10.1016/j.inffus.2022.12.010]
4. Hasan AM, Meziane F, Aspin R, Jalab HA. Segmentation of brain tumors in MRI images using three-dimensional active contour without edge. Symmetry 2016;8(11):132. [doi: 10.3390/sym8110132]
5. Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. Artif Intell Rev 2019;53(6):4093-4160. [doi: 10.1007/s10462-019-09788-3]

XSL•FO
RenderX

6. Noor MBT, Zenia NZ, Kaiser MS, Mamun SA, Mahmud M. Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. Brain Inform 2020;7(1):11 [FREE Full text] [doi: 10.1186/s40708-020-00112-2] [Medline: 33034769]
7. Ghaffari M, Sowmya A, Oliver R. Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012-2018 challenges. IEEE Rev Biomed Eng 2020;13:156-168. [doi: 10.1109/RBME.2019.2946868] [Medline: 31613783]
8. Fawzi A, Achuthan A, Belaton B. Brain image segmentation in recent years: a narrative review. Brain Sci 2021;11(8):1055 [FREE Full text] [doi: 10.3390/brainsci11081055] [Medline: 34439674]
9. Wadhwa A, Bhardwaj A, Singh Verma V. A review on brain tumor segmentation of MRI images. Magn Reson Imaging 2019;61:247-259. [doi: 10.1016/j.mri.2019.05.043] [Medline: 31200024]
10. Gau K, Schmidt CSM, Urbach H, Zentner J, Schulze-Bonhage A, Kaller CP, et al. Accuracy and practical aspects of semi- and fully automatic segmentation methods for resected brain areas. Neuroradiology 2020;62(12):1637-1648 [FREE Full text] [doi: 10.1007/s00234-020-02481-1] [Medline: 32691076]
11. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: a survey. IET Image Process 2022;16(5):1243-1267. [doi: 10.1049/ipr2.12419]
12. Seo H, Badiei Khuzani M, Vasudevan V, Huang C, Ren H, Xiao R, et al. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. Med Phys 2020;47(5):e148-e167 [FREE Full text] [doi: 10.1002/mp.13649] [Medline: 32418337]
13. Zhao Z, Chuah JH, Lai KW, Chow CO, Gochoo M, Dhanalakshmi S, et al. Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: a review. Front Comput Neurosci 2023;17:1038636 [FREE Full text] [doi: 10.3389/fncom.2023.1038636] [Medline: 36814932]
14. Balwant MK. A review on convolutional neural networks for brain tumor segmentation: methods, datasets, libraries, and future directions. IRBM 2022;43(6):521-537. [doi: 10.1016/j.irbm.2022.05.002]
15. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. 2015 Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention, Proceedings, Part III; October 5-9, 2015; Munich, Germany p. 18. [doi: 10.1007/978-3-319-24574-4_28]
16. Mall PK, Singh PK, Srivastav S, Narayan V, Paprzycki M, Jaworska T, et al. A comprehensive review of deep neural networks for medical image processing: recent developments and future opportunities. Healthc Anal 2023;4:100216. [doi: 10.1016/j.health.2023.100216]
17. Gul S, Khan MS, Bibi A, Khandakar A, Ayari MA, Chowdhury MEH. Deep learning techniques for liver and liver tumor segmentation: a review. Comput Biol Med 2022;147:105620. [doi: 10.1016/j.compbiomed.2022.105620] [Medline: 35667155]
18. Yousef R, Khan S, Gupta G, Siddiqui T, Albahlal BM, Alajlan SA, et al. U-net-based models towards optimal MR brain image segmentation. Diagnostics (Basel) 2023;13(9):1624 [FREE Full text] [doi: 10.3390/diagnostics13091624] [Medline: 37175015]
19. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39(12):2481-2495. [doi: 10.1109/TPAMI.2016.2644615] [Medline: 28060704]
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV, United States. [doi: 10.1109/cvpr.2016.90]
21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI, United States. [doi: 10.1109/cvpr.2017.243]
22. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3D convolutional networks. 2015 Presented at: 2015 IEEE International Conference on Computer Vision (ICCV); December 7-13, 2015; Santiago, Chile. [doi: 10.1109/iccv.2015.510]
23. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40(4):834-848. [doi: 10.1109/TPAMI.2017.2699184] [Medline: 28463186]
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA, United States.
25. He K, Gan C, Li Z, Rekik I, Yin Z, Ji W, et al. Transformers in medical image analysis. Intell Med 2023;3(1):59-78. [doi: 10.1016/j.imed.2022.07.002]
26. Xiao H, Li L, Liu Q, Zhu X, Zhang Q. Transformers in medical image segmentation: a review. Biomed Signal Process Control 2023;84:104791. [doi: 10.1016/j.bspc.2023.104791]
27. Akinyelu AA, Zaccagna F, Grist JT, Castelli M, Rundo L. Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to MRI: a survey. J Imaging 2022;8(8):205. [doi: 10.3390/jimaging8080205] [Medline: 35893083]

XSL•FO
RenderX

28. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. ArXiv Preprint posted online on June 3, 2021 [FREE Full text]
29. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: a survey. Med Image Anal 2023;88:102802. [doi: 10.1016/j.media.2023.102802] [Medline: 37315483]
30. Thisanke H, Deshan C, Chamith K, Seneviratne S, Vidanaarachchi R, Herath D. Semantic segmentation using vision transformers: a survey. Eng Appl Artif Intell 2023;126:106669. [doi: 10.1016/j.engappai.2023.106669]
31. Chen J, Lu Y, Yu Q, Luo X, Wang Y, Adeli E, et al. TransUNet: transformers make strong encoders for medical image segmentation. ArXiv Preprint posted online on February 8, 2021 [FREE Full text]
32. Cao H, Wang Y, Chen Y, Jiang D, Zhang X, Tian X, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation. 2022 Presented at: Computer Vision—ECCV 2022 Workshops; October 23-27, 2022; Tel Aviv, Israel. [doi: 10.1007/978-3-031-25066-8_9]
33. Wang W, Chen C, Ding M, Yu M, Zha S, Li J. TransBTS: multimodal brain tumor segmentation using transformer. 2021 Presented at: Medical Image Computing and Computer Assisted Intervention—MICCAI 2021; September 27-October 1, 2021; Strasbourg, France. [doi: 10.1007/978-3-030-87193-2_11]
34. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images. 2021 Presented at: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries; September 27, 2021; Virtual Event. [doi: 10.1007/978-3-031-08999-2_22]
35. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
36. Song J, Hahm J, Lee J, Lim CY, Chung MJ, Youn J, et al. Comparative validation of AI and non-AI methods in MRI volumetry to diagnose Parkinsonian syndromes. Sci Rep 2023;13(1):3439 [FREE Full text] [doi: 10.1038/s41598-023-30381-w] [Medline: 36859498]
37. Cai Y, Long Y, Han Z, Liu M, Zheng Y, Yang W, et al. Swin Unet3D: a three-dimensional medical image segmentation network combining vision transformer and convolution. BMC Med Inform Decis Mak 2023;23(1):33 [FREE Full text] [doi: 10.1186/s12911-023-02129-z] [Medline: 36788560]
38. Wu W, Yan J, Zhao Y, Sun Q, Zhang H, Cheng J, et al. Multi-task learning for concurrent survival prediction and semi-supervised segmentation of gliomas in brain MRI. Displays 2023;78:102402. [doi: 10.1016/j.displa.2023.102402]
39. Yan Q, Liu S, Xu S, Dong C, Li Z, Shi JQ, et al. 3D medical image segmentation using parallel transformers. Pattern Recognit 2023;138:109432. [doi: 10.1016/j.patcog.2023.109432]
40. Liu Z, Ma C, She W, Xuan W. TransMVU: multi‐view 2D U‐Nets with transformer for brain tumour segmentation. IET Image Process 2023;17(6):1874-1882. [doi: 10.1049/ipr2.12762]
41. Lu Y, Chang Y, Zheng Z, Sun Y, Zhao M, Yu B, et al. GMetaNet: multi-scale ghost convolutional neural network with auxiliary MetaFormer decoding path for brain tumor segmentation. Biomed Signal Process Control 2023;83:104694. [doi: 10.1016/j.bspc.2023.104694]
42. Wei C, Ren S, Guo K, Hu H, Liang J. High-resolution Swin transformer for automatic medical image segmentation. Sensors (Basel) 2023;23(7):3420 [FREE Full text] [doi: 10.3390/s23073420] [Medline: 37050479]
43. Gharaibeh N, Abu-Ein AA, Al-hazaimeh OM, Nahar KMO, Abu-Ain WA, Al-Nawashi MM. Swin transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning. Int J Onl Eng 2023;19(04):22-50. [doi: 10.3991/ijoe.v19i04.37677]
44. Wang J, Li S, Yu L, Qu A, Wang Q, Liu J, et al. SDPN: a slight dual-path network with local-global attention guided for medical image segmentation. IEEE J Biomed Health Inform 2023;27(6):2956-2967. [doi: 10.1109/JBHI.2023.3260026] [Medline: 37030687]
45. Anaya-Isaza A, Mera-Jiménez L, Fernandez-Quilez A. CrossTransUnet: a new computationally inexpensive tumor segmentation model for brain MRI. IEEE Access 2023;11:27066-27085. [doi: 10.1109/access.2023.3257767]
46. Lin J, Lin J, Lu C, Chen H, Lin H, Zhao B, et al. CKD-TransBTS: clinical knowledge-driven hybrid transformer with modality-correlated cross-attention for brain tumor segmentation. IEEE Trans Med Imaging 2023;42(8):2451-2461. [doi: 10.1109/TMI.2023.3250474] [Medline: 37027751]
47. Huang X, Liu Y, Li Y, Qi K, Gao A, Zheng B, et al. Deep learning-based multiclass brain tissue segmentation in fetal MRIs. Sensors (Basel) 2023;23(2):655 [FREE Full text] [doi: 10.3390/s23020655] [Medline: 36679449]
48. Hu Z, Li L, Sui A, Wu G, Wang Y, Yu J. An efficient R-Transformer network with dual encoders for brain glioma segmentation in MR images. Biomed Signal Process Control 2023;79:104034. [doi: 10.1016/j.bspc.2022.104034]
49. Dhamija T, Gupta A, Gupta S, Anjum, Katarya R, Singh G. Semantic segmentation in medical images through transfused convolution and transformer networks. Appl Intell (Dordr) 2023;53(1):1132-1148 [FREE Full text] [doi: 10.1007/s10489-022-03642-w] [Medline: 35498554]
50. Gao H, Miao Q, Ma D, Liu R. Deep mutual learning for brain tumor segmentation with the fusion network. Neurocomputing 2023;521:213-220. [doi: 10.1016/j.neucom.2022.11.038]

XSL•FO
RenderX

51. Ashtari P, Sima DM, De Lathauwer L, Sappey-Marinier D, Maes F, Van Huffel S. Factorizer: a scalable interpretable approach to context modeling for medical image segmentation. Med Image Anal 2023;84:102706. [doi: 10.1016/j.media.2022.102706] [Medline: 36516557]
52. Marcus A, Bentley P, Rueckert D. Concurrent ischemic lesion age estimation and segmentation of CT brain using a transformer-based network. IEEE Trans Med Imaging 2023;42(12):3464-3473. [doi: 10.1109/TMI.2023.3287361] [Medline: 37335797]
53. Piao Z, Gu YH, Jin H, Yoo SJ. Intracerebral hemorrhage CT scan image segmentation with HarDNet based transformer. Sci Rep 2023;13(1):7208 [FREE Full text] [doi: 10.1038/s41598-023-33775-y] [Medline: 37137921]
54. Yang H, Zhou T, Zhou Y, Zhang Y, Fu H. Flexible fusion network for multi-modal brain tumor segmentation. IEEE J Biomed Health Inform 2023;27(7):3349-3359. [doi: 10.1109/JBHI.2023.3271808] [Medline: 37126623]
55. Wang Z, He M, Lv Y, Ge E, Zhang S, Qiang N, et al. Accurate corresponding fiber tract segmentation via FiberGeoMap learner with application to autism. Cereb Cortex 2023;33(13):8405-8420. [doi: 10.1093/cercor/bhad125] [Medline: 37083279]
56. Liang J, Yang C, Zhong J, Ye X. BTSwin-unet: 3D U-shaped symmetrical swin transformer-based network for brain tumor segmentation with self-supervised pre-training. Neural Process Lett 2022;55(4):3695-3713. [doi: 10.1007/s11063-022-10919-1]
57. Rui-Qiang L, Xiao-Dong C, Ren-Zhe T, Cai-Zi L, Wei Y, Dou-Dou Z, et al. Automatic localization of target point for subthalamic nucleus-deep brain stimulation via hierarchical attention-UNet based MRI segmentation. Med Phys 2023;50(1):50-60. [doi: 10.1002/mp.15956] [Medline: 36053005]
58. Zhang S, Ren B, Yu Z, Yang H, Han X, Chen X, et al. TW-Net: transformer weighted network for neonatal brain MRI segmentation. IEEE J Biomed Health Inform 2023;27(2):1072-1083. [doi: 10.1109/JBHI.2022.3225475] [Medline: 36446007]
59. Khaled A, Han JJ, Ghaleb TA. Learning to detect boundary information for brain image segmentation. BMC Bioinformatics 2022;23(1):332 [FREE Full text] [doi: 10.1186/s12859-022-04882-w] [Medline: 35953776]
60. Huang L, Zhu E, Chen L, Wang Z, Chai S, Zhang B. A transformer-based generative adversarial network for brain tumor segmentation. Front Neurosci 2022;16:1054948 [FREE Full text] [doi: 10.3389/fnins.2022.1054948] [Medline: 36532274]
61. Chen S, Zhang JX, Zhang T. LETCP: a label-efficient transformer-based contrastive pre-training method for brain tumor segmentation. Appl Sci 2022;12(21):11016. [doi: 10.3390/app122111016]
62. Liang J, Yang C, Zeng L. 3D PSwinBTS: an efficient transformer-based Unet using 3D parallel shifted windows for brain tumor segmentation. Digital Signal Process 2022;131:103784. [doi: 10.1016/j.dsp.2022.103784]
63. Zhang J, Liu Y, Wu Q, Wang Y, Liu Y, Xu X, et al. SWTRU: star-shaped window transformer reinforced U-net for medical image segmentation. Comput Biol Med 2022;150:105954. [doi: 10.1016/j.compbiomed.2022.105954] [Medline: 36122443]
64. Liu X, Cheng X. Segmentation method of magnetoelectric brain image based on the transformer and the CNN. Information 2022;13(10):445. [doi: 10.3390/info13100445]
65. Xu Y, He X, Xu G, Qi G, Yu K, Yin L, et al. A medical image segmentation method based on multi-dimensional statistical features. Front Neurosci 2022;16:1009581 [FREE Full text] [doi: 10.3389/fnins.2022.1009581] [Medline: 36188458]
66. Gai D, Zhang J, Xiao Y, Min W, Zhong Y, Zhong Y. RMTF-Net: residual mix transformer fusion net for 2D brain tumor segmentation. Brain Sci 2022;12(9):1145 [FREE Full text] [doi: 10.3390/brainsci12091145] [Medline: 36138880]
67. Wu J, Xu Q, Shen Y, Chen W, Xu K, Qi XR. Swin transformer improves the IDH mutation status prediction of gliomas free of MRI-based tumor segmentation. J Clin Med 2022;11(15):4625 [FREE Full text] [doi: 10.3390/jcm11154625] [Medline: 35956236]
68. Huang J, Fang Y, Wu Y, Wu H, Gao Z, Li Y, et al. Swin transformer for fast MRI. Neurocomputing 2022;493:281-304. [doi: 10.1016/j.neucom.2022.04.051]
69. Chen Y, Yin M, Li Y, Cai Q. CSU-Net: a CNN-transformer parallel network for multimodal brain tumour segmentation. Electronics 2022;11(14):2226. [doi: 10.3390/electronics11142226]
70. Zeineldin RA, Pollok A, Mangliers T, Karar ME, Mathis-Ullrich F, Burgert O. Deep automatic segmentation of brain tumours in interventional ultrasound data. Curr Dir Biomed Eng 2022;8(1):133-137. [doi: 10.1515/cdbme-2022-0034]
71. Pinaya WHL, Tudosiu PD, Gray R, Rees G, Nachev P, Ourselin S, et al. Unsupervised brain imaging 3D anomaly detection and segmentation with transformers. Med Image Anal 2022;79:102475 [FREE Full text] [doi: 10.1016/j.media.2022.102475] [Medline: 35598520]
72. Jiang Y, Zhang Y, Lin X, Dong J, Cheng T, Liang J. SwinBTS: a method for 3D multimodal brain tumor segmentation using swin transformer. Brain Sci 2022;12(6):797 [FREE Full text] [doi: 10.3390/brainsci12060797] [Medline: 35741682]
73. Kadri R, Bouaziz B, Tmar M, Gargouri F. Multimodal deep learning based on the combination of EfficientNetV2 and ViT for Alzheimer's disease early diagnosis enhanced by SAGAN data augmentation. Int J Comput Inf Syst Ind Manag Appl 2022;14:313-325 [FREE Full text]
74. Liu J, Zheng J, Jiao G. Transition Net: 2D backbone to segment 3D brain tumor. Biomed Signal Process Control 2022;75:103622. [doi: 10.1016/j.bspc.2022.103622]
75. Wang J, Wang S, Liang W. METrans: multi‐encoder transformer for ischemic stroke segmentation. Electron Lett 2022;58(9):340-342. [doi: 10.1049/ell2.12444]

XSL•FO
RenderX

76. Liang J, Yang C, Zeng M, Wang X. TransConver: transformer and convolution parallel network for developing automatic brain tumor segmentation in MRI images. Quant Imaging Med Surg 2022;12(4):2397-2415 [FREE Full text] [doi: 10.21037/qims-21-919] [Medline: 35371952]
77. Wang X, Li Y. STC-Net: fusing Swin transformer and convolution neural network for 2D medical image segmentation. 2022 Presented at: 2022 2nd International Conference on Electronic Information Engineering and Computer Technology (EIECT); October 28-30, 2022; Yan'an, China. [doi: 10.1109/eiect58010.2022.00069]
78. Viteri J, Piguave B, Pelaez CE, Loayza F. Automatic brain white matter hyperintensities segmentation with Swin U-net. 2022 Presented at: 2022 IEEE ANDESCON; November 16-19, 2022; Barranquilla, Colombia. [doi: 10.1109/andescon56260.2022.9989775]
79. Wang P, Liu S, Peng J. AST-Net: lightweight hybrid transformer for multi-modal brain tumor segmentation. 2022 Presented at: 2022 26th International Conference on Pattern Recognition (ICPR); August 21-25, 2022; Montreal, QC, Canada. [doi: 10.1109/icpr56361.2022.9956705]
80. Zhang Y, He N, Yang J, Li Y, Wei D, Huang Y, et al. mmFormer: multi-modal medical transformer for incomplete multimodal learning of brain tumor segmentation. 2022 Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; September 18-22, 2022; Singapore. [doi: 10.1007/978-3-031-16443-9_11]
81. Huang J, Li H, Wan X. Attentive symmetric autoencoder for brain MRI segmentation. 2022 Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; September 18-22, 2022; Singapore. [doi: 10.1007/978-3-031-16443-9_20]
82. Xing Z, Yu L, Wan L, Han T, Zhu L. NestedFormer: nested modality-aware transformer for brain tumor segmentation. 2022 Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; September 18-22, 2022; Singapore. [doi: 10.1007/978-3-031-16443-9_14]
83. Chen Y, Wang J. TSEUnet: A 3D neural network with fused Transformer and SE-Attention for brain tumor segmentation. 2022 Presented at: 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS); July 21-23, 2022; Shenzen, China. [doi: 10.1109/cbms55023.2022.00030]
84. Wang E, Hu Y, Yang X, Tian X. TransUNet with attention mechanism for brain tumor segmentation on MR images. 2022 Presented at: 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); June 24-26, 2022; Dalian, China. [doi: 10.1109/icaica54878.2022.9844551]
85. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B. UNETR: transformers for 3D medical image segmentation. 2022 Presented at: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); January 3-8, 2022; Waikoloa, HI, United States. [doi: 10.1109/wacv51458.2022.00181]
86. Nijiati M, Tuersun A, Zhang Y, Yuan Q, Gong P, Abulizi A, et al. A symmetric prior knowledge based deep learning model for intracerebral hemorrhage lesion segmentation. Front Physiol 2022;13:977427 [FREE Full text] [doi: 10.3389/fphys.2022.977427] [Medline: 36505076]
87. Laiton-Bonadiez C, Sanchez-Torres G, Branch-Bedoya J. Deep 3D neural network for brain structures segmentation using self-attention modules in MRI images. Sensors (Basel) 2022;22(7):2559 [FREE Full text] [doi: 10.3390/s22072559] [Medline: 35408173]
88. Li Y, Cai W, Gao Y, Li C, Hu X. More than encoder: introducing transformer decoder to upsample. 2022 Presented at: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 6-8, 2022; Las Vegas, NV, United States. [doi: 10.1109/bibm55620.2022.9995378]
89. Rasoulian A, Salari S, Xiao Y. Weakly supervised intracranial hemorrhage segmentation using hierarchical combination of attention maps from a Swin transformer. 2022 Presented at: International Workshop on Machine Learning in Clinical Neuroimaging; September 18, 2022; Singapore p. 63-72. [doi: 10.1007/978-3-031-17899-3_7]
90. Ayivi W, Zeng L, Yussif SB, Browne JA, Agbesi VK, Sam F, et al. Segmentation of glioblastoma multiforme via-attention neural network. 2022 Presented at: 33rd Irish Signals and Systems Conference (ISSC); June 9-10, 2022; Cork, Ireland. [doi: 10.1109/issc55427.2022.9826163]
91. Ou C, Qian Y, Chong W, Hou X, Zhang M, Zhang X, et al. A deep learning-based automatic system for intracranial aneurysms diagnosis on three-dimensional digital subtraction angiographic images. Med Phys 2022;49(11):7038-7053. [doi: 10.1002/mp.15846] [Medline: 35792717]
92. Jia Q, Shu H. BiTr-Unet: a CNN-transformer combined network for MRI brain tumor segmentation. 2021 Presented at: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries; September 27, 2021; Virtual Event URL: https://europepmc.org/abstract/MED/36005929
93. Yang Y, Wei S, Zhang D, Yan Q, Zhao S, Han J. Hierarchical and global modality interaction for brain tumor segmentation. 2022 Presented at: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries; September 27, 2021; Quebec City, Canada p. 441-450. [doi: 10.1007/978-3-031-08999-2_38]
94. Luo C, Zhang J, Chen X, Tang Y, Weng X, Xu F. UCATR: based on CNN and transformer encoding and cross-attention decoding for lesion segmentation of acute ischemic stroke in non-contrast computed tomography images. 2021 Presented at: 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); November 1-5, 2021; Mexico. [doi: 10.1109/embc46164.2021.9630336]

XSL•FO
RenderX

95. Sagar A. Vitbis: Vision transformer for biomedical image segmentation. 2021 Presented at: MICCAI Workshop on Distributed and Collaborative Learning; October 1, 2021; Strasbourg, France p. 34-45. [doi: 10.1007/978-3-030-90874-4_4]
96. Sun Q, Fang N, Liu Z, Zhao L, Wen Y, Lin H. HybridCTrm: bridging CNN and transformer for multimodal brain image segmentation. J Healthc Eng 2021;2021:7467261 [FREE Full text] [doi: 10.1155/2021/7467261] [Medline: 34630994]
97. Fidon L, Shit S, Ezhov I, Paetzold JC, Ourselin S, Vercauteren T. Generalized wasserstein dice loss, test-time augmentation, and transformers for the BraTS 2021 challenge. 2022 Presented at: International MICCAI Brainlesion Workshop; September 27, 2021; Singapore p. 187-196. [doi: 10.1007/978-3-031-09002-8_17]
98. Sagar A. Emsvit: Efficient multi scale vision transformer for biomedical image segmentation. 2021 Presented at: International MICCAI Brainlesion Workshop; September 27, 2021; Quebec City, Canada p. 39-51.
99. Yang H, Shen Z, Li Z, Liu J, Xiao J. Combining global information with topological prior for brain tumor segmentation. 2021 Presented at: International MICCAI Brainlesion Workshop; October 8, 2023; Quebec City, Canada. [doi: 10.1007/978-3-031-08999-2_16]
100. Li J, Chen Y, Cai L, Davidson I, Ji S. Dense transformer networks for brain electron microscopy image segmentation. 2019 Presented at: Proceedings of the 28th International Joint Conference on Artificial Intelligence; August 10-16, 2019; Macao, China. [doi: 10.24963/ijcai.2019/401]
101. Menze B, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging 2015;34(10):1993-2024 [FREE Full text] [doi: 10.1109/TMI.2014.2377694] [Medline: 25494501]
102. Wang P, Yang Q, He Z, Yuan Y. Vision transformers in multi-modal brain tumor MRI segmentation: a review. Meta-Radiology 2023;1(1):100004. [doi: 10.1016/j.metrad.2023.100004]
103. Khan RF, Lee B, Lee MS. Transformers in medical image segmentation: a narrative review. Quant Imaging Med Surg 2023;13(12):8747-8767 [FREE Full text] [doi: 10.21037/qims-23-542] [Medline: 38106306]

Abbreviations

AI: artificial intelligence
BraTS dataset: brain tumor segmentation dataset
CNN: convolutional neural network
CT: computed tomography
DL: deep learning
MRI: magnetic resonance imaging
PET: positron emission tomography
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
ViT: vision transformer

Edited by A Coristine; submitted 25.02.24; peer-reviewed by J Mistry, Y Shen, S Mary; comments to author 02.07.24; revised version received 23.08.24; accepted 19.11.24; published 29.01.25.

Please cite as:
Iratni M, Abdullah A, Aldhaheri M, Elharrouss O, Abd-alrazaq A, Rustamov Z, Zaki N, Damseh R
Transformers for Neuroimage Segmentation: Scoping Review
J Med Internet Res 2025;27:e57723
URL: https://www.jmir.org/2025/1/e57723
doi:10.2196/57723
PMID:

©Maya Iratni, Amira Abdullah, Mariam Aldhaheri, Omar Elharrouss, Alaa Abd-alrazaq, Zahiriddin Rustamov, Nazar Zaki, Rafat Damseh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Large Language Models to Detect and Understand Drug Discontinuation Events in Web-Based Forums: Development and Validation Study

William Trevena[1], PhD; Xiang Zhong[1], PhD; Michelle Alvarado[1], PhD; Alexander Semenov[1], PhD; Alp Oktay[2], BSc; Devin Devlin[3], BSc; Aarya Yogesh Gohil[1], MSc; Sai Harsha Chittimouju[1], MSc

[1]Department of Industrial and Systems Engineering, The University of Florida, GAINESVILLE, FL, United States

[2]Department of Industrial and Systems Engineering, The University of San Diego, San Diego, CA, United States

[3]Microsoft, Seattle, WA, United States

Corresponding Author:
Xiang Zhong, PhD
Department of Industrial and Systems Engineering
The University of Florida
PO BOX 115002
GAINESVILLE, FL, 32611-5002
United States
Phone: 1 3523922477
Email: xiang.zhong@ise.ufl.edu

Abstract

Background: The implementation of large language models (LLMs), such as BART (Bidirectional and Auto-Regressive Transformers) and GPT-4, has revolutionized the extraction of insights from unstructured text. These advancements have expanded into health care, allowing analysis of social media for public health insights. However, the detection of drug discontinuation events (DDEs) remains underexplored. Identifying DDEs is crucial for understanding medication adherence and patient outcomes.

Objective: The aim of this study is to provide a flexible framework for investigating various clinical research questions in data-sparse environments. We provide an example of the utility of this framework by identifying DDEs and their root causes in an open-source web-based forum, MedHelp, and by releasing the first open-source DDE datasets to aid further research in this domain.

Methods: We used several LLMs, including GPT-4 Turbo, GPT-4o, DeBERTa (Decoding-Enhanced Bidirectional Encoder Representations from Transformer with Disentangled Attention), and BART, among others, to detect and determine the root causes of DDEs in user comments posted on MedHelp. Our study design included the use of zero-shot classification, which allows these models to make predictions without task-specific training. We split user comments into sentences and applied different classification strategies to assess the performance of these models in identifying DDEs and their root causes.

Results: Among the selected models, GPT-4o performed the best at determining the root causes of DDEs, predicting only 12.9% of root causes incorrectly (hamming loss). Among the open-source models tested, BART demonstrated the best performance in detecting DDEs, achieving an F_1-score of 0.86, a false positive rate of 2.8%, and a false negative rate of 6.5%, all without any fine-tuning. The dataset included 10.7% (107/1000) DDEs, emphasizing the models' robustness in an imbalanced data context.

Conclusions: This study demonstrated the effectiveness of open- and closed-source LLMs, such as GPT-4o and BART, for detecting DDEs and their root causes from publicly accessible data through zero-shot classification. The robust and scalable framework we propose can aid researchers in addressing data-sparse clinical research questions. The launch of open-access DDE datasets has the potential to stimulate further research and novel discoveries in this field.

(J Med Internet Res 2025;27:e54601) doi:10.2196/54601

KEYWORDS

natural language processing; large language models; ChatGPT; drug discontinuation events; zero-shot classification; artificial intelligence; AI

XSL•FO
RenderX

Introduction

The development of the transformer model [1] has significantly advanced natural language processing (NLP), a key area of artificial intelligence (AI). These models, commonly referred to as large language models (LLMs), often contain billions or trillions of parameters and have greatly improved our ability to extract meaningful insights from unstructured data [2], achieving state-of-the-art results in various NLP tasks since the introduction of bidirectional encoder representations from transformers [3]. LLMs have excelled in tasks, such as translation, answering of questions, summarization, sentiment analysis, language generation, and named entity recognition. They are used in diverse fields, including analyzing corporate filings for investment decisions [4], assessing online reviews [5], and performing sentiment analysis of tweets [6].

Recently, advanced LLMs such as DeBERTa (Decoding-Enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention) [7] and OpenAI's GPT-4 [8] have demonstrated substantial performance improvements. These enhancements include improvements in model architecture, such as DeBERTa's disentangled attention mechanism and GPT-4's increased model size and training data, which enhance their prediction accuracy and their capacity to handle complex language understanding tasks. These advancements have opened new possibilities for health care applications of NLP, such as analyzing social media and online forums to assess trust and confidence toward vaccine use [9], adverse drug reactions [10,11], and depression [12], among others [13-16]. However, many potential applications of NLP in health care remain underexplored.

One critical issue is early drug discontinuation and nonadherence to prescribed medication, which negatively affects patient outcomes and increases health care costs [17]. The reasons behind early medication discontinuation are not well understood [18,19], and detecting drug discontinuation events (DDEs) from clinical narratives and social media data is underexplored.

Some studies have explored medication discontinuation using data from social media and NLP techniques. Golder et al [20] conducted a mixed methods study analyzing patient-reported reasons for switching or discontinuing statin therapy using social media data from WebMD. Their work focused specifically on statins and relied on manual annotation for classifying DDEs and their root causes. Similarly, van Buchem et al [21] developed and validated the AI patient-reported experience measure (AI-PREM) by analyzing patient experiences using NLP. Their study primarily used sentiment analysis to gauge patient experiences but did not focus on detecting specific events such as DDEs.

Another study by Tsai et al [22] used NLP and network analysis to examine patients withdrawing from life-sustaining treatments using electronic health records from 119 patients in Taiwan. Their approach involved generating document-term matrices and analyzing word frequencies to identify patterns. However, they did not use pretrained LLMs, and their study was limited to a small dataset.

While these studies and others [23-26] contribute valuable insights into patient experiences and reasons for medication discontinuation, they have limitations such as focusing on specific medications, relying on manual annotation, or not leveraging advanced LLMs for broader applicability. To the best of our knowledge, no prior studies have systematically analyzed the root causes of medication discontinuation using state-of-the-art LLMs such as OpenAI's GPT-4o, especially in a zero-shot classification setting applied to large-scale, unstructured patient-generated data from web-based forums.

Historically in many NLP applications, the models needed to be fine-tuned or adjusted for each specific task. However, state-of-the-art LLMs are capable of zero-shot classification, which are especially valuable in settings where data are scarce, as they do not require modifications or fine-tuning on task-specific data. This feature is particularly beneficial in the medical field, where access to data is often limited due to patient confidentiality concerns. Motivated by challenges associated with limited data availability in health care research, the primary aim of this study is to develop a versatile framework for addressing clinical research questions in data-sparse settings by leveraging pretrained LLMs and zero-shot classification of unstructured data from web-based health forums. The specific objectives of this paper are as follows:

1. To present a methodical approach to detecting DDEs posted on web-based health forums
2. To identify the root causes of DDEs posted on web-based health forums
3. To evaluate the effectiveness of different classification strategies and models in identifying DDEs and their root causes
4. To introduce and validate the first open-source DDE detection dataset and the first open-source DDE root cause classification dataset to support ongoing and future research

In addition, our work is novel in 4 additional aspects. First, studies in the extant work mainly used electronic health records (EHRs) as the data source [17]. Due to the poor documentation of drug discontinuation events (DDEs) in EHRs [17], it is crucial to understand the causes of discontinuation from the patient's perspective to improve medication adherence. Traditional EHR data often lack direct patient input, as they are usually filtered through health care providers. Moreover, privacy concerns surrounding patient health data often restrict accessibility to large open-source datasets for health research. Meanwhile, web-based forums such as MedHelp offer unmediated patient perspectives, provide insights not captured in EHRs, and involve a large and diverse population. Identifying the root causes of DDEs from the patient's viewpoint on the basis of a representative sample can enhance our understanding of patient behavior and the challenges they face.

Second, to the best of our knowledge, this is the first work that compares state-of-the-art text-to-text models with the previous generation of models that receive text as an input and respond with text as an output, such as GPT-4o. Many open-source pretrained large language models (LLMs) fine-tuned for zero-shot classification are available through platforms like Hugging Face [27]. These models can predict whether a given

XSL•FO
RenderX

text supports an arbitrary hypothesis without needing additional training for each new task. This means that if a human reader can infer the truth of a hypothesis from the text, the model can make a prediction to do the same [28]. Typically, these models output a probability between 0 and 1, indicating how likely it is that the hypothesis is true based on the given text. To be able to apply these approaches to the problem, we designed a novel framework that leverages these models to identify DDE and the associated root causes.

Third, we demonstrated the effectiveness of LLMs in addressing clinical research questions without extensive training data. This approach overcomes the limitations of previous methods [17] by using transfer learning to tackle data scarcity. Our study shows that LLMs such as bidirectional and auto-regressive transformers (BART) and GPT-4o, with their generalization capabilities, are versatile for new tasks. This novel methodology could revolutionize the way clinical research questions are approached, opening a new paradigm in health informatics research.

Fourth, this paper introduces 2 open-source datasets, the first of their kind: one focused on DDE detection and another on root cause classification of DDEs. By creating and publishing these datasets, we aimed to motivate and facilitate future research efforts toward investigating drug discontinuation. For example, researchers could use these datasets to investigate why patients discontinue specific medications early. By extracting and analyzing comments that describe DDEs, they can identify the root causes and guide future research to address side effects or other factors associated with drug discontinuation.

Overall, this research provides valuable insights into patient behavior and decision-making regarding medication use, highlighting the real-world challenges faced by the patients. These insights can inform health care providers, policy makers, and pharmaceutical companies, helping them to better understand and address barriers to medication adherence.

Methods

Overview

The effectiveness of research outcomes in NLP relies heavily on the methods and materials used. This section outlines the strategies and resources used in this study, focusing on LLMs. In particular, this study introduces an innovative approach by redefining DDE detection as a natural language inference task. This approach enables LLMs to leverage their existing training to deduce relationships between text elements, achieving the desired outcomes without the need for additional training data or model fine-tuning.

We sourced data from MedHelp [29], a publicly accessible web-based health community where users discuss health experiences and share advice. This section details the methodology used for detecting DDEs from comments posted on MedHelp and analyzes the root causes of these DDEs using LLMs.

Data Scraping and Preprocessing

Our dataset was compiled by scraping 759,872 questions and answers from 131 of the 193 communities on MedHelp. Each question or answer has a unique URL, with information corresponding to each being recorded as a separate row in our dataset.

The dataset was then trimmed down by excluding questions and answers that lacked relevant study keywords. These keywords were derived from the International Drug Dictionary dataset [30], reducing our dataset to 183,565 questions and answers.

Data Labeling

Overview

To create the first open-source DDE datasets and to assess the accuracy of LLMs at detecting DDEs and identifying their root causes, we used graduate students to label the data obtained from MedHelp. For DDE detection, 1000 comments were labeled, resulting in 893 (89.3%) non-DDE comments and 107 (10.7%) DDE comments. Of the 89 (8.9%) comments that the labelers did not unanimously agree upon, the majority vote was taken as the “ground truth.” Here, the “ground truth” refers to the correct classification for each comment, based on the consensus of the labelers.

For DDE root cause analysis, 1000 (100%) comments suspected of containing DDEs were categorized based on predefined criteria. Multiple root causes can apply to a single DDE and a single comment can express multiple DDEs, allowing multiple root causes to be assigned to one comment. When the labelers disagreed, they discussed the comments to reach a consensus.

Classifying Comments as DDE or Non-DDE

Labelers were asked to categorize each comment into either of the following categories:

- 0—non-DDE (the comment does not contain a DDE)
- 1—DDE (the comment contains a DDE)

The comments were classified as DDE if a specific person’s discontinuation of a recurring medication or treatment could be inferred from it. Medication changes were considered as DDEs while one-time treatments were not. Apart from this, labelers were asked to indicate their level of confidence in their labels by selecting one of the three confidence levels: (1) very confident (2890/3000, 96.33%); (2) somewhat confident (102/3000, 3.4%); and (3) not confident (8/3000, 0.27%). The percentage in parentheses alongside each confidence level indicates the rate at which the confidence level was selected by the labelers.

Classifying the Root Causes of DDEs

After identifying comments containing DDEs, the next step involved categorizing the root causes of these discontinuations. Labelers were provided with detailed guidelines and definitions for each category to ensure consistency and accuracy (Textbox 1).

Labelers were instructed to apply these categories to each comment identified as containing a DDE. They could assign

XSL•FO
RenderX

multiple categories to a single comment if it mentioned >1 reason for discontinuation. The comprehensive guidelines, including examples and the decision tree shown in Figure 1, helped labelers accurately and consistently classify the root causes of DDEs. The aim was to capture the complex and multifaceted reasons behind medication discontinuation, providing valuable insights for further analysis.

Textbox 1. Categories for root causes.

- Treatment success: discontinuation due to successful treatment completion or sufficient improvement in health
- Treatment inefficacy: perceived ineffectiveness of the treatment or loss of belief in its efficacy
- Adverse reactions: discontinuation due to adverse side effects, allergic reactions, or negative interactions with other medications
- Accessibility issues: issues related to financial constraints, changes in prescriber decisions, or market availability
- Personal choices: discontinuation based on personal decisions, cultural or religious reasons, or general nonadherence without medical advice
- Alternative medical reasons: other medical reasons not explicitly covered by the other categories
- Indeterminate: unclear or unspecified reasons for discontinuation
- Nondiscontinuation: comments that do not indicate any form of drug discontinuation

XSL•FO
RenderX

Figure 1. Decision tree for labeling the root causes of drug discontinuation events.

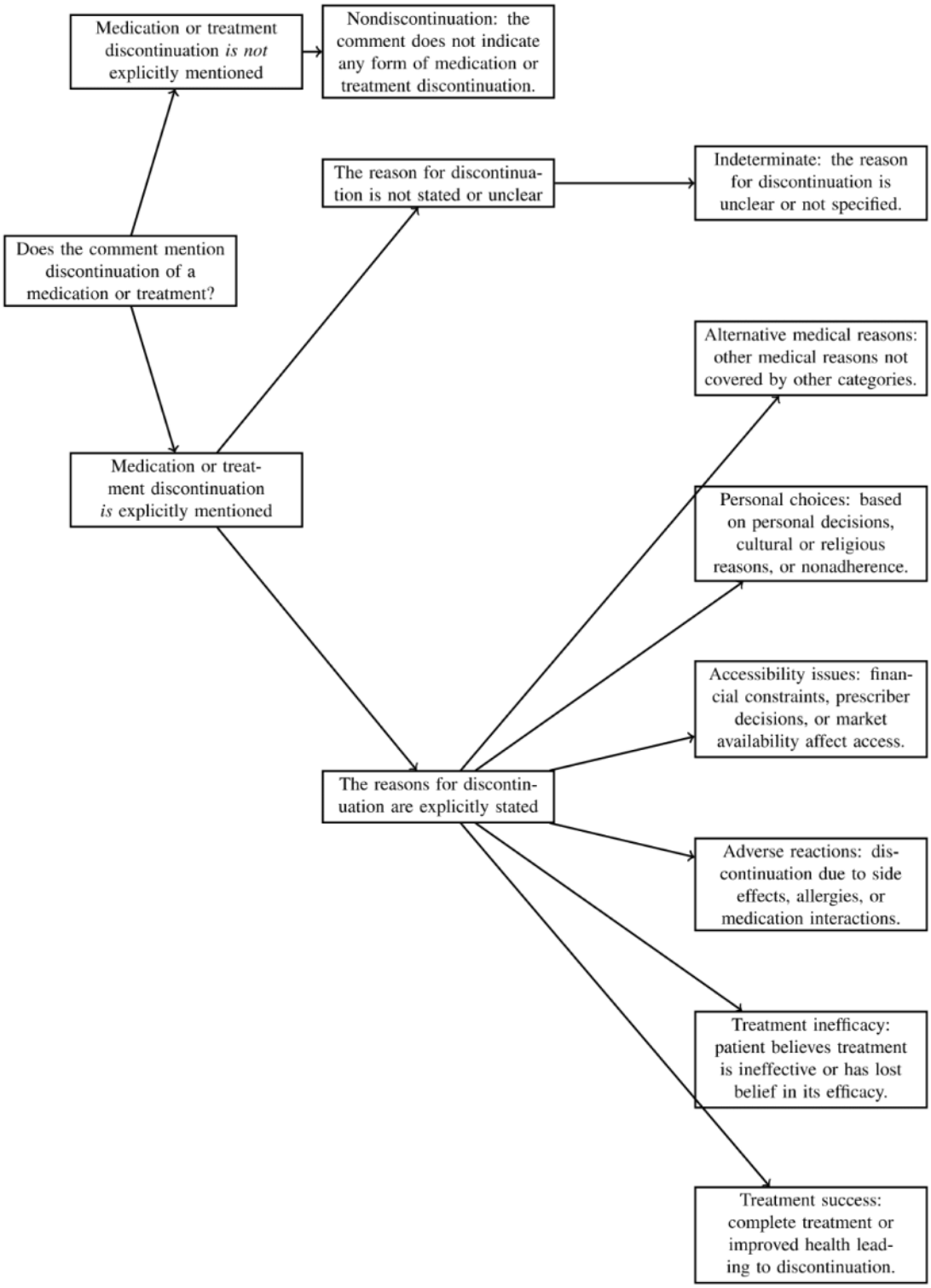

Models

In this study, we used the following models from Hugging Face (Hugging Face, Inc)—all of which are LLMs fine-tuned for zero-shot classification—to detect DDEs in comments posted on web-based forums: nli-deberta-base [31]; bart-large-mnli [32]; roberta-large-mnli [33]; distilbert-base-uncased-mnli [34]; nli-distilroberta-base [35].

In addition, we used OpenAI's GPT-3.5 Turbo and GPT-4 models [36]. These LLMs have been trained on a wide range of internet text and can generate humanlike text, making them useful for our research. GPT-3.5 Turbo is smaller and less costly than GPT-4 because of its fewer parameters [37]. As illustrated in Figure 2, these models can be used to analyze a premise from a web-based forum post (eg, a post from MedHelp) alongside a hypothesis (eg, "person stopped taking medication") and output a probability, such as 0.8, for whether the premise entails the hypothesis. This approach allows us to identify potential discontinuation events in a zero-shot manner without requiring additional training data.

XSL•FO
RenderX

Figure 2. Illustration of how large language models fine-tuned to perform natural language inference can be used to perform zero-shot classification. MNLI: Multi-Genre Natural Language Inference; RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.

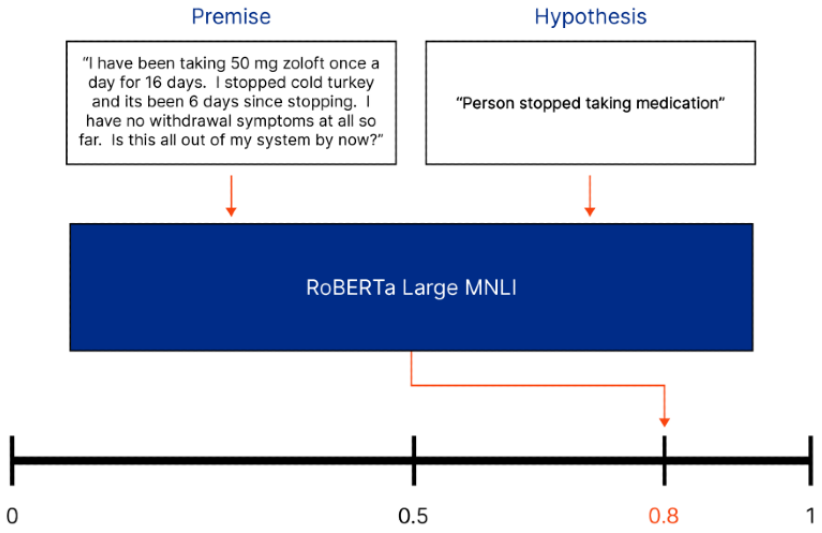

Data Classification Procedure Used to Detect DDEs

Detecting DDEs can be framed as a binary classification task, where the goal is to determine whether a text sequence contains a DDE or not. For example, the statement "patient stopped Lipitor because of cost" would be classified as containing a DDE. Each model, including GPT-3.5 Turbo and GPT-4, performed zero-shot classification as described in the classification strategies later. For models from the Hugging Face library, comments from MedHelp were treated as premises with the hypothesis being the following: "Person stopped taking medication." Note that the results presented in this paper and the performance of the proposed framework are sensitive to the hypothesis used, and using a different hypothesis could result in different results.

For GPT-3.5 Turbo and GPT-4, the models were prompted to generate a response to a query using the comment as context. The model's output was processed to derive a probability of entailment, which was used for classification. The evaluation setting for the GPT models varied with the classification strategy used and was influenced by the model version. It is important to note that the DDE detection experiments were performed before the release of GPT-4 Turbo and the release of the "set seed" and structured JavaScript Object Notation response features by OpenAI. As a result, carefully structured prompts were needed to obtain parseable responses from the GPT models for the task of DDE detection. These prompts and the complete code used in this work are available on GitHub [38]. The subsequent sections detail the procedures used to classify comments into the 2 categories, with practical examples of these classification strategies shown as figures.

Step 1: Split Each Comment Into Sentences and Tokenize

The models used for DDE detection, including GPT-3.5 Turbo and GPT-4, have a maximum token input length. This means each model can only process text that, when tokenized, is within its maximum token limit. Given that questions and answers on MedHelp can exceed these limits, we split longer comments into smaller chunks for classification. A comment was classified as discussing a DDE if any of its segments were classified as such. To handle the prevalence of typos and missing punctuation, we used the spaCy Python library's Sentencizer method [39] to segment each question and answer into sentences. Despite some inaccuracies due to informal language, typos, or incorrect punctuation, we referred to the output of the Sentencizer as the consecutive "sentences" forming each question or answer from MedHelp.

Step 2: Apply a Classification Strategy

Each comment was processed using 3 distinct classification strategies: classification strategy 1 (CS1, see Figure 3 and Textbox 2), classification strategy 2 (CS2; see Figures 4 and 5 and Textbox 3), and classification strategy 3 (CS3; see Figure 5 and Textbox 4).

In CS1 (Figure 3; Textbox 2), each sentence from the text is treated as a premise and passed into the model individually. The maximum probability (ie, model output) returned for any premise is identified. If this maximum probability is above a user-defined cutoff level, the original post is classified as containing a DDE. In the example shown in Figure 3, the post is classified as containing a DDE because the maximum model output exceeds the user-defined cutoff level (95%).

In CS2 (Figures 4 and 5; Textbox 3), consecutive sentences are concatenated and passed into the model as single premises. The maximum model output for any premise is compared to a user-defined cutoff value. In the example shown in Figure 4, the post is classified as containing a DDE because the maximum probability (99.4%) is higher than the cutoff level (95%).

In Figure 5, an example shows a scenario where the total number of tokens in the original post exceeds the maximum input length that the model can accept. Here, sequential sentences are grouped to ensure no group exceeds the model's maximum token limit (Textbox 4).

XSL•FO
RenderX

Figure 3. An example illustrating classification strategy 1. DDE: drug discontinuation event; Max: maximum; MNLI: Multi-Genre Natural Language Inference; RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.

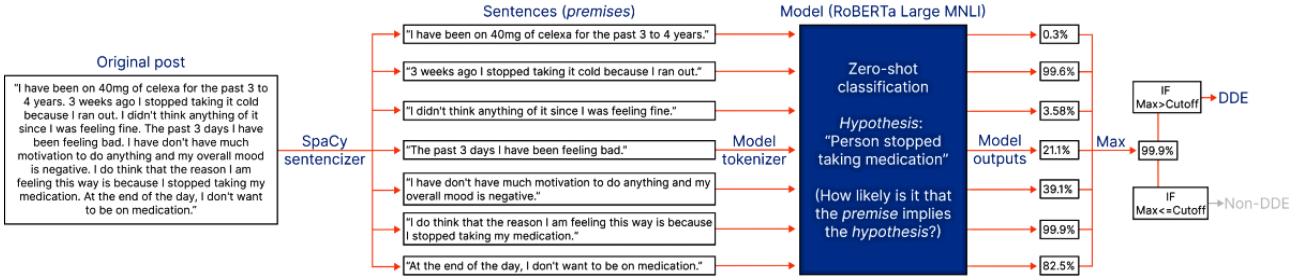

Textbox 2. CS1—classification strategy 1: individual sentences.

1. Each sentence in a comment was classified individually, generating a "model output" for each sentence ranging from 0 (non-DDE) to 1 (DDE).
2. The "model prediction" was set as the maximum of all the "model outputs" for the sentences.
3. If the "model prediction" exceeded a user-defined cutoff value, the comment was classified as a DDE. Otherwise, it was categorized as a non-DDE.

Figure 4. An example illustrating classification strategy 2 when using a model with a large maximum token input length. DDE: drug discontinuation event; Max: maximum; MNLI: Multi-Genre Natural Language Inference; RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.



Figure 5. An example illustrating classification strategy 2 when using a model with a small maximum token input length. DDE: drug discontinuation event; Max: maximum; MNLI: Multi-Genre Natural Language Inference; RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.



Textbox 3. CS2—classification strategy 2: groups of sentences.

1. Groups of sentences were created by concatenating as many consecutive sentences together as possible while staying within the maximum token input length of the model.
2. Each group of sentences was classified as a single block of text, generating a "model output" for each group of sentences ranging from 0 (non-DDE) to 1 (DDE).
3. The "model prediction" was set as the maximum of all the "model outputs" for the different groups of sentences.
4. If the "model prediction" exceeded a user-defined cutoff value, the comment was classified as a DDE. Otherwise, it was classified as a non-DDE.

Textbox 4. CS3—classification strategy 3: entire text (ChatGPT only).

The entire text is passed into the ChatGPT model at once, and the model is prompted to provide a binary "1" (DDE) or "0" (non-DDE) classification for each comment.

XSL•FO
RenderX

Step 3: Compare the "Model Prediction" to a User-Defined "Cutoff" Value

To derive a binary classification for a comment from the set of model outputs for each sentence (CS1) or block of sentences (CS2), we compared the maximum of all the model outputs to a cutoff value. If the maximum model output for any sentence or block of sentences is greater than the cutoff value, the comment is classified as a DDE. Otherwise, it is classified as a non-DDE. Cutoff values were tested from 0.05 to 0.95 in increments of 0.05. Adjusting the cutoff impacts the trade-off between sensitivity and specificity, reflected in the resulting false positive rate (FPR) and false negative rate (FNR) of each classifier. This approach helps identify an optimal cutoff that balances false positives and false negatives. This structured approach ensures a comprehensive analysis and maximizes the accuracy of DDE detection in unstructured web-based forum data.

Data Classification Procedure Used to Detect the Root Causes of DDEs

Once DDEs were identified, we proceeded to classify the root causes of these discontinuations. The root cause classification was treated as a multilabel classification task, allowing each comment to be associated with multiple reasons for discontinuation.

Applying CS1, CS2, and CS3 to DDE Root Cause Classification

The same classification strategies (ie, CS1, CS2, and CS3) used to detect DDEs were applied to identify the root causes of DDEs.

CS1: Individual Sentences

Each sentence in a DDE-identified comment was classified individually to determine the root causes. Each sentence was analyzed to see if it fit any of the predefined root cause categories in the Classifying the Root Causes of DDEs section.

If any sentence indicated a root cause, it was considered a relevant cause for the entire comment. The final set of root causes for the comment was derived from the union of causes identified in individual sentences.

CS2: Groups of Sentences

Groups of sentences were classified together to determine the root causes. This method ensured that contextual information was preserved, which is crucial for accurately identifying root causes that may span multiple sentences. Each group of sentences was evaluated against the predefined categories of root causes, and the maximum model output for each group determined the presence of specific root causes.

CS3: Entire Comment

The entire text of each DDE-identified comment was passed into the ChatGPT model to classify the root causes. The model was prompted to first summarize the reasons for discontinuation and then was asked to generate a list of applicable root causes from the predefined categories. This approach leveraged the model's ability to understand the context of the entire comment, providing a comprehensive assessment of the root causes.

Comparing Model Predictions to Cutoff Values

For both CS1 and CS2, the maximum model output for each root cause category was compared to a user-defined cutoff value. If the model output exceeded the cutoff, the root cause was assigned to the comment.

The complete set of prompts and hypotheses used for both DDE detection and root cause analysis can be found on our GitHub repository [38]. These details are omitted here for brevity, but they are available on the web and in Multimedia Appendix 1 to ensure transparency and facilitate replication of our methods. Researchers and practitioners can access these resources to better understand the specific inputs used in our classification models and apply similar techniques to their own datasets.

Ethical Considerations

This study involves the analysis of user-generated content on MedHelp, a publicly accessible health forum where individuals share personal medical experiences and seek health advice. Given the sensitive nature of the health information discussed, we are committed to protecting user privacy and ensuring ethical handling of all data. While the content on MedHelp is publicly available, users have not explicitly consented to participate in this research. We acknowledged the need to respect users' implied expectations of privacy, especially given the health-focused context of the discussions. No attempt was made to contact individual users nor were they asked to participate directly in this study. The scraped data have been processed in a way that prevents the association of responses or health conditions with specific users. This study complies with all relevant ethical guidelines for internet-based research, and we did not find it necessary to seek an ethics review board assessment based on the guidelines set by the University of Florida institutional review board.

Results

Overview

This section presents the results of our experiments on detecting DDEs and identifying their root causes. We evaluated the performance of various LLMs, using different classification strategies. The evaluation included precision-recall (PR) curves, receiver operating characteristic (ROC) curves, and detailed performance metrics for each classifier and strategy combination.

DDE Detection Task

The effectiveness of the classifiers in detecting DDEs was evaluated using PR and ROC curves. Figures 6 and 7 show the PR and ROC curves for each classifier under each classification strategy, respectively. These figures also include the area under the curve (AUC) for each combination of classifier and strategy. Of all the models and classification strategies, bidirectional and auto-regressive transformers (BART) under CS1 achieved the highest ROC AUC of 0.977 and the highest PR AUC of 0.904. On the other hand, GPT-3.5 Turbo under CS2 had the lowest ROC AUC of 0.611 and the lowest PR AUC of 0.215.

Under CS3, GPT-4 achieved an F_1-score of only 0.56, an FPR of 12.9%, an FNR of 18.7%, and an accuracy of 86.5%. GPT-3.5

XSL•FO
RenderX

Turbo performed even more poorly under CS3, achieving an F_1-score of only 0.27, an FPR of 5.6%, an FNR of 76.6%, and an accuracy of 86.8%, illustrating the imbalance in the dataset. In comparison, the best-performing classifier at the task of DDE detection was BART under CS1, achieving an F_1-score of 0.86, an accuracy of 96.8%, an FPR of 2.8%, and an FNR of 6.5%. A summary of these performance metrics can be found in Tables S1-S9 in Multimedia Appendix 1.

Figure 6. The precision-recall curve of each classifier on the task of detecting drug discontinuation events. BART: Bidirectional and Auto-Regressive Transformers; DeBERTa: Decoding-Enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention; RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.

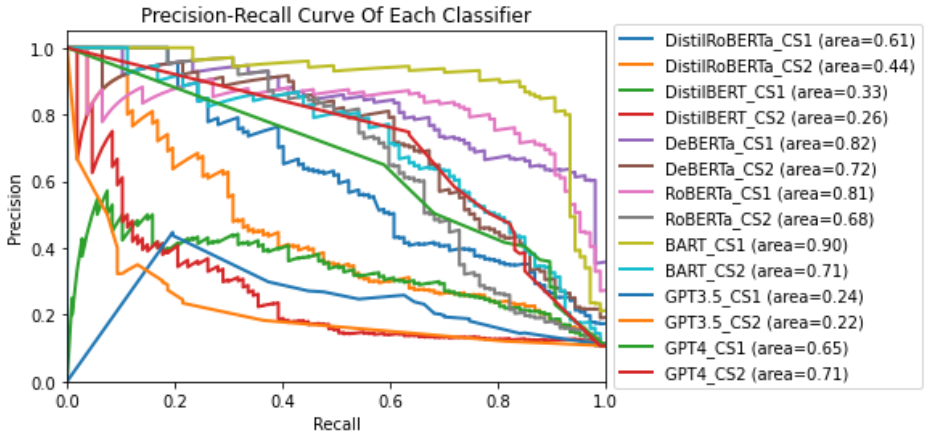

Figure 7. The receiver operating characteristic curve of each classifier on the task of detecting drug discontinuation events. BART: Bidirectional and Auto-Regressive Transformers; DeBERTa: Decoding-Enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention; RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach.

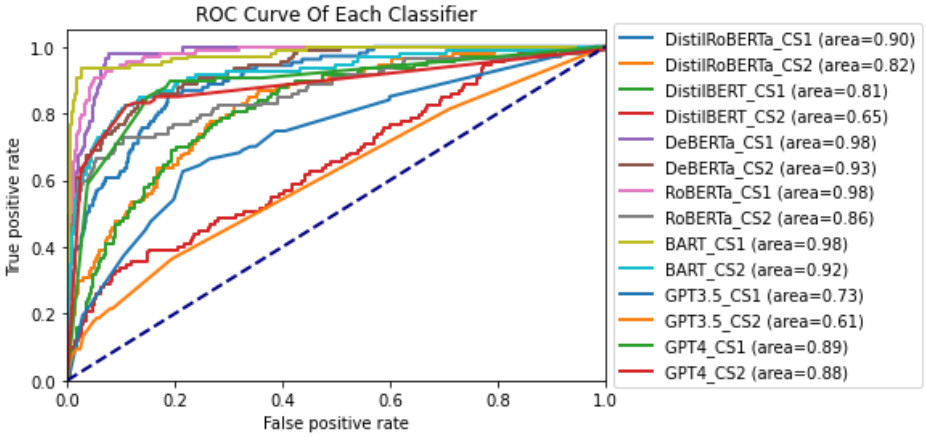

DDE Root Cause Classification Task

To assess the performance of the classifiers in identifying the root causes of DDEs, we used multiple evaluation metrics, including micro-average ROC AUC, macro-average ROC AUC, weighted-average ROC AUC, micro-average PR AUC, macro-average PR AUC, weighted-average. PR AUC, micro-average F_1, macro-average F_1, weighted-average F_1, hamming loss, Jaccard score, and subset accuracy. The results across all DDE root causes are summarized in Tables 1 and 2.

Tables 3 and 4 show the performance of GPT-4 Turbo and GPT-4o at predicting each individual DDE root cause. Overall, GPT-4o performed the best across the board achieving the highest averaged F_1-scores, the lowest hamming loss, and the highest Jaccard score and subset accuracy of all the classifiers. On the other hand, BART and DeBERTa performed very poorly across the board on the task of DDE root cause detection as shown in Tables 1 and 2.

An overview of the frequency of each DDE root cause in our labeled dataset of 1000 comments is provided in Table 5.

The most common DDE root causes in our dataset were *personal choices* with 670 (67%) occurrences and *adverse reactions* with 627 (62.7%) occurrences, and the least common was *treatment success* with 96 (9.6%) occurrences. In addition, only 10% (100/1000) of the comments were labeled as non-DDE by labelers, suggesting the accuracy performance is consistent with the experiments performed regarding DDE detection in a smaller labeled dataset.

XSL•FO
RenderX

Table 1. Evaluation metrics for DDE[a] root cause classification (part 1).

Classifier	Strategy	Micro-average ROC[b] AUC[c]	Macro-average ROC AUC	Weighted-average ROC AUC	Micro-average PR[d] AUC	Macro-average PR AUC	Weighted-average PR AUC
DeBERTa[e]	CS1[f]	0.568669	0.507989	0.502451	0.338243	0.300314	0.472326
DeBERTa	CS2[g]	0.528049	0.508299	0.500792	0.312381	0.300947	0.472003
BART[h]	CS1	0.604359	0.510527	0.507081	0.368773	0.301987	0.474543
BART	CS2	0.581217	0.518657	0.518826	0.344869	0.306675	0.480845

[a]DDE: drug discontinuation event.

[b]ROC: receiver operating characteristic.

[c]AUC: area under the curve.

[d]PR: precision-recall.

[e]DeBERTa: Decoding-Enhanced Bidirectional Encoder Representations from Transformer with Disentangled Attention.

[f]CS1: classification strategy 1.

[g]CS2: classification strategy 2.

[h]BART: Bidirectional and Auto-Regressive Transformers.

Table 2. Evaluation metrics for DDE[a] root cause classification (part 2).

Classifier	Strategy	Micro-average F_1	Macro-average F_1	Weighted-average F_1	Hamming loss	Jaccard score	Subset accuracy
DeBERTa[b]	CS1[c]	0.337991	0.245975	0.359274	0.36594	0.218692	0.014905
DeBERTa	CS2[d]	0.222708	0.184899	0.244992	0.366095	0.119609	0.005095
BART[e]	CS1	0.425243	0.284461	0.441498	0.363548	0.292373	0.029905
BART	CS2	0.445648	0.349929	0.516783	0.469571	0.297445	0.015476
GPT-4 Turbo	CS3[f]	0.729443	0.616037	0.711924	0.153	0.634567	0.344
GPT-4o	CS3	0.783084	0.739096	0.786899	0.128875	0.6823	0.356

[a]DDE: drug discontinuation event.

[b]DEBRTa: Decoding-Enhanced Bidirectional Encoder Representations from Transformer with Disentangled Attention.

[c]CS1: classification strategy 1.

[d]CS2: classification strategy 2.

[e]BART: Bidirectional and Auto-Regressive Transformers.

[f]CS3: classification strategy 3.

XSL•FO
RenderX

Table 3. Detailed performance metrics by DDE[a] root cause for GPT-4 Turbo.

Class	F_1-score	Precision	Recall	Accuracy	FPR[b]	TPR[c]	FNR[d]	TNR[e]
Treatment success	0.6339	0.6667	0.6042	0.933	0.0321	0.6042	0.3958	0.9679
Treatment inefficacy	0.6840	0.6840	0.6840	0.842	0.1053	0.6840	0.3160	0.8947
Adverse reactions	0.8467	0.8267	0.8676	0.803	0.3056	0.8676	0.1324	0.6944
Accessibility issues	0.6524	0.8971	0.5126	0.935	0.0079	0.5126	0.4874	0.9921
Personal choices	0.8138	0.8362	0.7925	0.757	0.3152	0.7925	0.2075	0.6848
Alternative medical reasons	0.5690	0.6918	0.4833	0.694	0.1546	0.4833	0.5167	0.8454
Indeterminate	0.0702	0.3636	0.0388	0.894	0.0078	0.0388	0.9612	0.9922
Non-DDE	0.6583	0.5643	0.7900	0.918	0.0678	0.7900	0.2100	0.9322

[a]DDE: drug discontinuation event.
[b]FPR: false positive rate.
[c]TPR: true positive rate.
[d]FNR: false negative rate.
[e]TNR: true negative rate.

Table 4. Detailed performance metrics by DDE[a] root cause for GPT-4o.

Class	F_1-score	Precision	Recall	Accuracy	FPR[b]	TPR[c]	FNR[d]	TNR[e]
Treatment success	0.7551	0.74	0.7708	0.952	0.028	0.7708	0.2292	0.9712
Treatment inefficacy	0.7968	0.8016	0.792	0.899	0.0653	0.792	0.208	0.9347
Adverse reactions	0.8836	0.9089	0.8596	0.858	0.1448	0.8596	0.1404	0.8552
Accessibility issues	0.8087	0.8378	0.7815	0.956	0.0204	0.7815	0.2185	0.9796
Personal choices	0.8419	0.8576	0.8269	0.792	0.2788	0.8269	0.1731	0.7212
Alternative medical reasons	0.6457	0.6295	0.6627	0.696	0.2801	0.6627	0.3373	0.7199
Indeterminate	0.3810	0.3438	0.4272	0.857	0.0937	0.4272	0.5728	0.9064
Non-DDE	0.8000	0.7810	0.8200	0.959	0.0256	0.8200	0.1800	0.9744

[a]DDE: drug discontinuation event.
[b]FPR: false positive rate.
[c]TPR: true positive rate.
[d]FNR: false negative rate.
[e]TNR: true negative rate.

Table 5. Frequency of labels in the dataset.

Label	Comments (n=1000), n (%)
Personal choices	670 (67)
Adverse reactions	627 (62.7)
Alternative medical reasons	418 (41.8)
Treatment inefficacy	250 (25)
Accessibility issues	119 (11.9)
Indeterminate	103 (10.3)
Nondiscontinuation	100 (10)
Treatment success	96 (9.6)

XSL•FO
RenderX

Discussion

DDE Detection

Overview

The results from the PR and ROC curves highlight significant differences in the performance of the models at detecting DDEs using the 3 different classification strategies: CS1, CS2, and CS3. These metrics are essential for understanding the trade-offs between sensitivity and specificity in identifying DDEs, which is vital in clinical settings where missing a DDE can have serious implications.

The PR curves (Figure 6) demonstrate the effectiveness of each classifier in handling the imbalanced nature of the dataset. BART under CS1 achieved the highest PR AUC, indicating its superior ability to maintain high precision without sacrificing recall. This performance suggests that BART is particularly adept at identifying DDEs in a dataset where DDEs are relatively rare, representing only 107 (10.7%) of the 1000 comments we labeled. In contrast, GPT-3.5 Turbo under CS2 had the lowest F_1-score of only 0.22, showing significant challenges in balancing precision and recall. These results highlight the limitations of some models when applied with the classification strategies we developed and to our imbalanced data set. However, these results highly depend on the prompts used with the models, and better results may be achieved by using different prompts in conjunction with the proposed classification strategies. Further research into this area is needed to better understand this phenomenon.

Similarly, the ROC curves (Figure 7) illustrate the classifiers' ability to distinguish between DDE and non-DDE cases across various thresholds. The area under the ROC curve provides a measure of the overall discriminative power of the models. BART's high area under the ROC curve under CS1 indicates its robustness in differentiating between the 2 classes, even at different decision thresholds. This robustness is crucial in clinical applications where both high sensitivity and specificity are required to minimize false negatives and false positives.

Classification Strategies

CS1: Individual Sentences

This strategy was particularly effective for both DeBERTa and BART in detecting DDEs, maintaining a higher balance between precision and recall as evidenced by their PR AUC scores. By analyzing sentences individually, this approach mitigates the dilution of relevant information, which is particularly beneficial in contexts where critical information might be embedded within longer text.

CS2: Groups of Sentences

While this strategy provided context by combining multiple sentences, it often resulted in a higher FNR due to the dilution of relevant information with surrounding text. This observation aligns with our hypothesis that feeding multiple sentences to a model simultaneously can lead to relevant DDE mentions being overshadowed by unrelated information. However, an intriguing exception was observed for GPT-4, which performed slightly better under CS2 than CS1 at higher cutoff values. This suggests that GPT-4 might better leverage the entire context of comments to make more informed classifications, indicating its potential to handle more complex and detailed contexts.

CS3: Entire Comment (GPT 3.5 Turbo and GPT-4 Models Only)

Interestingly, the performance of GPT-4 under CS3 was comparable to its performance under CS2 with a cutoff of 0.75, contradicting our initial hypothesis that the GPT models would perform better under the binary classification task considered in CS3. This implies that the architecture and training methodology of GPT-4 may not align with the approach implemented in CS3 or that our prompt design in CS3 was not optimally formulated. Further investigation is needed to understand this phenomenon better.

Key Takeaways

Superior Performance of BART

BART's superior performance, particularly under CS1, highlights its robustness and reliability in detecting DDEs from unstructured patient-generated content. This finding suggests that BART can be effectively used in clinical research to identify relevant data from imbalanced datasets without the need for extensive fine-tuning, which is crucial for real-world applications where data imbalance is common.

Because of BART's exceptional performance in DDE detection, we used BART under the classification strategy and cutoff that it achieved the best performance with to curate the dataset we labeled and used for our DDE root cause analysis. BART achieved the highest F_1-score of 0.86 under CS1 with a cutoff of 0.9.

Class Imbalance

The significant class imbalance, with only 107 out of 1000 (10.7%) comments labeled as DDEs, emphasizes the need for careful evaluation of classifier performance beyond overall accuracy. Metrics such as PR AUC provide a more comprehensive understanding of model performance, ensuring that high accuracy does not mask poor performance in detecting less frequent but critical events.

Real-World Applications

High FPRs and FNRs in DDE detection can lead to missed DDEs or unnecessary interventions, respectively. Therefore, achieving a balance in FPR and FNR, as demonstrated by BART under CS1 with a cutoff of 0.9, is essential for practical applications. This balance ensures that models can effectively support health care providers in making informed decisions based on patient-generated data.

DDE Root Cause Analysis

Our multilabel classification approach enabled the identification of multiple reasons for drug discontinuation within a single comment, offering a comprehensive understanding of patient behavior and decision-making. This robust methodology allowed us to leverage the strengths of LLMs to tackle complex clinical research questions related to DDEs.

XSL•FO
RenderX

Performance Highlights

DeBERTa and BART Struggled at Detecting DDE Root Causes

Both models struggled significantly at identifying the root causes of DDEs, as indicated by their much lower F_1-scores and higher hamming loss in Table 2 in comparison to GPT-4 Turbo and GPT-4o. Another interesting observation was that 10% (100/1000) of the comments in the dataset were non-DDEs, meaning that BART under CS1 with a cutoff of 0.9 inaccurately classified them as DDEs. We noticed that BART tended to label non-DDE comments as DDEs if they contained phrases such as "I am not on any medication." This is an important takeaway as it shows that even the models that performed the best at detecting DDEs still have tendencies to make mistakes in certain scenarios.

GPT-4 Turbo and GPT-4o Performed Exceptionally Well at Detecting DDE Root Causes

These models outperformed DeBERTa and BART across most metrics. GPT-4o, in particular, achieved the highest F_1-scores, lowest hamming loss, and highest Jaccard scores, indicating its superior ability to accurately and reliably classify the root causes of DDEs. The advanced capabilities of GPT-4 Turbo and GPT-4o in processing complex textual inputs holistically contributed to their outstanding performance.

Classification Strategies

CS1 (Individual Sentences) and CS2 (Groups of Sentences)

Although these strategies were effective for both DeBERTa and BART at detecting DDEs, the models struggled at detecting the root causes of DDEs under these strategies, as indicated by their extremely poor performance results shown in Tables 1 and 2.

CS3: Entire Comment (GPT-4 Turbo and GPT-4o Models Only)

GPT-4 Turbo and GPT-4o performed exceptionally well under CS3, likely due to their advanced capabilities in understanding and processing complex textual inputs holistically. This strategy allowed the models to capture the full context of the comments, leading to more accurate and reliable DDE root cause identification.

Key Takeaways

Superior Performance of GPT-4 Turbo and GPT-4o

The outstanding performance of GPT-4 Turbo and GPT-4o models highlights the importance of using advanced language models for complex classification tasks in health care. Their ability to accurately identify the root causes of DDEs can provide valuable insights for health care providers, enabling them to address patient concerns more effectively and improve medication adherence.

Challenges With Indeterminate Class

One notable challenge was the performance of GPT-4 Turbo in detecting DDEs belonging to the *indeterminate* class. For instance, a comment such as "My dad was diagnosed with bipolar disorder and stopped taking his meds after being in a rehabilitation center twice" did not clearly state the reason for discontinuation. Instead of classifying this as *indeterminate*, GPT-4 Turbo attempted to infer specific reasons such as *personal choices* or *adverse reactions*. This tendency to infer explicit conditions from ambiguous statements indicates a potential overfitting issue, where the model predicts more common classes instead of recognizing the ambiguity inherent in the input.

Class Imbalance

The significant class imbalance presented a challenge for all classifiers. Classes with higher frequencies, such as *personal choices* and *adverse reactions*, exhibited better performance. In contrast, less frequent classes such as *indeterminate* showed lower performance, highlighting the need for more balanced training datasets or advanced techniques to handle class imbalance effectively.

Clinical Implications

The superior performance of GPT-4 Turbo and GPT-4o in both DDE detection and root cause analysis underscores their potential application in clinical settings. These models can help health care providers identify common reasons for drug discontinuation, enabling the development of targeted interventions to improve medication adherence and patient outcomes. By leveraging these advanced NLP models, clinicians and researchers can gain deeper insights into patient behavior and decision-making processes, facilitating more effective and personalized health care strategies.

This research emphasizes the importance of capturing the patient's perspective directly from web-based health forums, which offers a richer and more authentic understanding of patient behavior and the real-world challenges they face compared to traditional electronic health record data. By identifying the root causes of DDEs from the patient's standpoint, health care providers, policy makers, and pharmaceutical companies can develop more targeted interventions to enhance medication adherence and patient outcomes.

Future Research Directions

To further enhance the applicability and reliability of these models in real-world settings, future research should focus on the opportunities outlined in the subsequent sections.

Real-World Applications

This research provides a foundation for clinical applications, such as improving patient adherence, by identifying common reasons for drug discontinuation. Health care providers can use these insights to develop targeted interventions and support systems to address the specific challenges faced by patients, ultimately enhancing treatment outcomes. Furthermore, applying these findings in real-world settings, such as clinical decision support systems, where accurate and reliable classification of text can significantly impact patient outcomes and health care processes, represents a substantial opportunity. Developing scalable and cost-effective solutions for deploying these models in practice will be crucial.

XSL•FO
RenderX

Prompt Optimization

Future research should explore various prompt optimization techniques to enhance model performance. This includes experimenting with chain-of-thought prompting, which has been shown to improve the performance of LLMs on complex reasoning tasks. In addition, investigating the impact of prompt specificity and structure on DDE root cause classification performance can provide valuable insights into optimal prompt design. For instance, future research could examine how the performance of GPT-4 Turbo and GPT-4o changes when examples of positive instances for low-frequency classes are included in the prompt. One concern with this approach is the potential for "prompt dependency," where providing specific examples might restrict the model's ability to generalize and accurately classify comments that differ in nature from those presented in the prompt.

Exploring Other State-of-the-Art LLMs

Research should be expanded to include state-of-the-art language models from developers other than OpenAI, such as models from Anthropic, Google, and Cohere. Models such as Claude by Anthropic and Gemini by Google have shown promising results in various NLP tasks and could be valuable for this study. In addition, models from research institutions, such as Command R by Cohere, should also be considered for comprehensive performance comparisons. By comparing the performance of these models on DDE root cause classification, we can identify the best tools for specific multilabel classification tasks and potentially reduce dependency on any single provider. This comparative analysis will help in understanding the strengths and limitations of each model in handling complex health care–related data, ultimately leading to more robust and versatile NLP applications.

Cost Minimization Strategies

Developing strategies to minimize the costs associated with using advanced NLP models is essential. This includes exploring open-source alternatives, such as LLaMA, optimizing model use to reduce application programming interface calls, and implementing efficient data processing pipelines. Research directions toward cost minimization could potentially align with advanced prompting techniques, such as classifying multiple comments with a single prompt and application programming interface call. However, preliminary research suggests that longer prompts lead to degraded performance of LLMs, even when the number of tasks in a single prompt does not increase.

Limitations

In summary, this research provides valuable insights into the ability of LLMs to detect DDEs from textual comments. However, the results should be interpreted with some caution due to the small sample size of DDE-labeled comments and the case imbalance inherent in our dataset. It is also worth mentioning that, as our study was restricted to English language comments from MedHelp, the generalizability of our results to other languages and platforms is not guaranteed. Furthermore, because the models used in this work generate stochastic outputs, the reproducibility of these results may be limited.

Despite these limitations, our study provides a foundation for further exploration into the use of LLMs in health informatics, specifically in identifying DDEs. Future studies should focus on improving the performance of transformer models by investigating other strategies, exploring ensemble methods, refining the classification prompt for the ChatGPT models, or incorporating different forms of data augmentation.

We anticipate that such innovative methods will stimulate and support further research into drug discontinuation, aiding health care researchers in identifying factors associated with the discontinuation of specific medications and treatments. These advancements are crucial in spotlighting new research directions with the potential to ameliorate treatment protocols and mitigate issues associated with medication discontinuation and nonadherence.

The ultimate aim is to enhance the capability of these models in real-world applications where accurate, timely detection of DDEs is of paramount importance. Improved detection can contribute to patient safety, optimize drug therapies, and facilitate better health outcomes.

Conclusions

This study has provided a thorough evaluation of advanced LLMs for detecting DDEs and identifying their root causes using data from MedHelp. The research demonstrates the effectiveness of state-of-the-art LLM models, particularly GPT-4 and GPT-4o, in handling complex multilabel classification tasks without the need for extensive task-specific training.

In conclusion, this research contributes significantly to the field of NLP and health care informatics by showcasing the potential of advanced classifiers to derive actionable insights from patient-generated content. The findings underscore the transformative impact these technologies can have on understanding and improving medication adherence, ultimately leading to better health care outcomes.

Acknowledgments

This work was supported by the University of Florida Foundation Grant.

Data Availability

The datasets generated and analyzed during this study are openly available in the GitHub repository [38]. This repository includes the raw data, processed data, and any scripts used in data processing and analysis. Further documentation is provided within the

XSL•FO
RenderX

repository to guide data use and interpretation. For any additional inquiries regarding the dataset, please contact the corresponding author.

Authors' Contributions

WT, AO, AS, and DD conceived and conceptualized the study. WT, XZ, MA, AS, AO, DD, AYG, and SHC participated in the methodology design. WT, AS, AO, DD, AYG, and SHC contributed to the coding. WT, XZ, MA, AS, and AO completed the formal analysis. WT led the investigation and validation, with support from AO, AYG, and SHC. WT, XZ, MA, AS, AO, and DD drafted the manuscript. All authors reviewed and revised the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional details regarding the methods and results presented in this research.
[DOCX File , 62 KB - jmir_v27i1e54601_app1.docx]

References

1. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, CA URL: https://dl.acm.org/doi/10.5555/3295222.3295349
2. Rogers A, Kovaleva O, Rumshisky A. A primer in BERT ology: what we know about how BERT works. Transact Assoc Comput Linguist 2020;8:842-866 [FREE Full text] [doi: 10.1162/tacl_a_00349]
3. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL-HLT 2019; June 2-7, 2019; Minneapolis, MN. [doi: 10.18653/v1/N19-1423]
4. Adosoglou G, Lombardo G, Pardalos PM. Neural network embeddings on corporate annual filings for portfolio selection. Expert Syst Appl 2021 Feb;164:114053. [doi: 10.1016/j.eswa.2020.114053]
5. Fagernäs S, Hamilton W, Espinoza N, Miloff A, Carlbring P, Lindner P. What do users think about Virtual Reality relaxation applications? A mixed methods study of online user reviews using natural language processing. Internet Interv 2021 Apr;24:100370 [FREE Full text] [doi: 10.1016/j.invent.2021.100370] [Medline: 33665136]
6. Bera A, Ghose MK, Pal DK. Sentiment analysis of multilingual tweets based on natural language processing (NLP). Int J Syst Dyn Appl 2021 Jan;10(4). [doi: 10.4018/IJSDA.20211001.oa16]
7. He P, Liu X, Gao J, Chen W. DeBERTa: decoding-enhanced BERT with disentangled attention. arXiv Preprint posted online on June 5, 2020 [FREE Full text]
8. OpenAI. GPT-4 technical report. arXiv. Preprint posted online on March 27, 2023 2024 [FREE Full text]
9. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res 2020;21(140):1-67 [FREE Full text]
10. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. ALBERT: a lite BERT for self-supervised learning of language representations. arXiv Preprint posted online on September 26, 2019 [FREE Full text]
11. Liu Z, Lin W, Shi Y, Zhao J. A robustly optimized BERT pre-training approach with post-training. In: Proceedings of the Chinese Computational Linguistics: 20th China National Conference. 2021 Presented at: CCL 2021; August 13-15, 2021; Hohhot, China URL: https://aclanthology.org/2021.ccl-1.108 [doi: 10.1007/978-3-030-84186-7_31]
12. Wei J, Ren X, Li X, Huang W, Liao Y, Wang Y, et al. NEZHA: neural contextualized representation for chinese language understanding. arXiv Preprint posted online on August 31, 2019 [FREE Full text]
13. Zhang Z, Han X, Liu Z, Jiang X, Sun M, Liu Q. ERNIE: enhanced language representation with informative entities. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: ACL 2019; July 28-August 2, 2019; Florence, Italy URL: https://aclanthology.org/P19-1139 [doi: 10.18653/v1/p19-1139]
14. Zhang Z, Wu Y, Zhao H, Li Z, Zhang S, Zhou X, et al. Semantics-aware BERT for language understanding. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020 Presented at: AAAI-20; February 7-12, 2020; New York, NY. [doi: 10.1609/aaai.v34i05.6510]
15. Liu X, He P, Chen W, Gao J. Multi-task deep neural networks for natural language understanding. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: ACL 2019; July 28-August 2, 2019; Florence, Italy URL: https://www.aclweb.org/anthology/P19-1441 [doi: 10.18653/v1/p19-1441]
16. Wang W, Bi B, Yan M, Wu C, Bao Z, Xia B, et al. StructBERT: incorporating language structures into pre-training for deep language understanding. arXiv Preprint posted online on August 13, 2019 [FREE Full text]

XSL•FO
RenderX

17. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020 Presented at: NIPS'20; December 6-12, 2020; Vancouver, BC.
18. de Wynter A, Perry DJ. Optimal subarchitecture extraction for BERT. arXiv Preprint posted online on October 20, 2020 [FREE Full text]
19. Xue L, Constant N, Roberts A, Kale M, Al-Rfou R, Siddhant A, et al. mT5: a massively multilingual pre-trained text-to-text transformer. arXiv Preprint posted online on October 22, 2020 [FREE Full text] [doi: 10.18653/v1/2021.naacl-main.41]
20. Golder S, Weissenbacher D, O'Connor K, Hennessy S, Gross R, Hernandez GG. Patient-reported reasons for switching or discontinuing statin therapy: a mixed methods study using social media. Drug Saf 2022 Sep 07;45(9):971-981 [FREE Full text] [doi: 10.1007/s40264-022-01212-0] [Medline: 35933649]
21. van Buchem MM, Neve OM, Kant IM, Steyerberg EW, Boosman H, Hensen EF. Analyzing patient experiences using natural language processing: development and validation of the artificial intelligence patient reported experience measure (AI-PREM). BMC Med Inform Decis Mak 2022 Jul 15;22(1):183 [FREE Full text] [doi: 10.1186/s12911-022-01923-5] [Medline: 35840972]
22. Tsai WC, Tsai YC, Kuo KC, Cheng SY, Tsai JS, Chiu TY, et al. Natural language processing and network analysis in patients withdrawing from life-sustaining treatments: a retrospective cohort study. BMC Palliat Care 2022 Dec 22;21(1):225 [FREE Full text] [doi: 10.1186/s12904-022-01119-8] [Medline: 36550430]
23. Breden A, Moore L. Detecting adverse drug reactions from Twitter through domain-specific preprocessing and BERT ensembling. arXiv Preprint posted online on May 11, 2020 [FREE Full text]
24. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8 [FREE Full text] [doi: 10.2105/ajph.2016.303512]
25. Liu F, Pradhan R, Druhl E, Freund E, Liu W, Sauer BC, et al. Learning to detect and understand drug discontinuation events from clinical narratives. J Am Med Inform Assoc 2019 Oct 01;26(10):943-951 [FREE Full text] [doi: 10.1093/jamia/ocz048] [Medline: 31034028]
26. Thill M, Schmidt M. Management of adverse events during cyclin-dependent kinase 4/6 (CDK4/6) inhibitor-based treatment in breast cancer. Ther Adv Med Oncol 2018 Sep 03;10:1758835918793326 [FREE Full text] [doi: 10.1177/1758835918793326] [Medline: 30202447]
27. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020 Presented at: EMNLP 2020; November 16-20, 2020; Online URL: https://aclanthology.org/2020.emnlp-demos.6 [doi: 10.18653/v1/2020.emnlp-main]
28. Dagan I, Glickman O, Magnini B. The PASCAL recognising textual entailment challenge. In: Proceedings of the Evaluating Predictive Uncertainty, Visual Object Classification, and Recognizing Textual Entailment, First Pascal Machine Learning Challenges Workshop. 2005 Presented at: MLCW 2005; April 11-13, 2005; Southampton, UK. [doi: 10.1007/11736790_9]
29. org. All Ask a Doctor Forums And Medical Communities - MedHelp ? medhelp. URL: https://www.medhelp.org/forums/list [accessed 2023-04-26]
30. Khaleel MA, Khan AH, Ghadzi SM, Alshakhshir S. Curation of an international drug proprietary names dataset. Data Brief 2022 Feb;40:107701 [FREE Full text] [doi: 10.1016/j.dib.2021.107701] [Medline: 34988273]
31. cross-encoder / nli-deberta-base. Hugging Face. URL: https://huggingface.co/cross-encoder/nli-deberta-base [accessed 2023-08-24]
32. facebook / bart-large-mnli. Hugging Face. URL: https://huggingface.co/facebook/bart-large-mnli [accessed 2023-08-24]
33. FacebookAI / roberta-large-mnli. Hugging Face. URL: https://huggingface.co/roberta-large-mnli [accessed 2023-08-24]
34. typeform / distilbert-base-uncased-mnli. Hugging Face. URL: https://huggingface.co/typeform/distilbert-base-uncased-mnli [accessed 2023-08-24]
35. cross-encoder / nli-distilroberta-base. Hugging Face. URL: https://huggingface.co/cross-encoder/nli-distilroberta-base [accessed 2023-08-24]
36. Models. OpenAI Platform. URL: https://platform.openai.com/docs/models [accessed 2023-07-23]
37. Pricing. OpenAI. URL: https://openai.com/pricing [accessed 2023-08-01]
38. Trevena W. Utilizing open source language models and ChatGPT for zero-shot identification of drug discontinuation events (DDEs) in online forums. GitHub. URL: https://github.com/wtrevena/Open-Source-Language-Models-ChatGPT-for-Zero-Shot-Identification-of-DDEs-in-Online-Forums [accessed 2025-01-06]
39. Honnibal M, Montani I. spaCy 2: natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Sentometrics Research. 2017. URL: https://sentometrics-research.com/publication/72/ [accessed 2024-12-30]

Abbreviations

AI: artificial intelligence
AUC: area under the curve

XSL•FO
RenderX

BART: Bidirectional and Auto-Regressive Transformers
CS1: classification strategy 1
CS2: classification strategy 2
CS3: classification strategy 3
DDE: drug discontinuation event
DeBERTa: Decoding-Enhanced Bidirectional Encoder Representations from Transformers with Disentangled Attention
FNR: false negative rate
FPR: false positive rate
LLM: large language model
NLP: natural language processing
PR: precision-recall
ROC: receiver operating characteristic

Edited by A Mavragani; submitted 20.11.23; peer-reviewed by J Yang, K Hur, F Uzoaru; comments to author 19.03.24; revised version received 03.07.24; accepted 07.11.24; published 30.01.25.

Please cite as:
Trevena W, Zhong X, Alvarado M, Semenov A, Oktay A, Devlin D, Gohil AY, Chittimouju SH
Using Large Language Models to Detect and Understand Drug Discontinuation Events in Web-Based Forums: Development and Validation Study
J Med Internet Res 2025;27:e54601
URL: https://www.jmir.org/2025/1/e54601
doi:10.2196/54601
PMID:39883487

©William Trevena, Xiang Zhong, Michelle Alvarado, Alexander Semenov, Alp Oktay, Devin Devlin, Aarya Yogesh Gohil, Sai Harsha Chittimouju. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study

Godwin Denk Giebel[1], MSc; Pascal Raszke[1], MA; Hartmuth Nowak[2,3], PD, Dr Med; Lars Palmowski[2], Dr Med; Michael Adamzik[2], Prof Dr; Philipp Heinz[4], Dipl-Kfm; Marianne Tokic[5], MSc; Nina Timmesfeld[5], Prof Dr; Frank Brunkhorst[6], Prof Dr; Jürgen Wasem[1], Prof Dr; Nikola Blase[1], Dr Med

[1]Institute for Healthcare Management and Research, University of Duisburg-Essen, Essen, Germany

[2]Department of Anesthesiology, Intensive Care and Pain Therapy, University Hospital Knappschaftskrankenhaus Bochum, Bochum, Germany

[3]Center for Artificial Intelligence, Medical Informatics and Data Science, University Hospital Knappschaftskrankenhaus Bochum, Bochum, Germany

[4]Knappschaft Kliniken GmbH, Recklinghausen, Germany

[5]Department of Medical Informatics, Biometry and Epidemiology, Ruhr University Bochum, Bochum, Germany

[6]German Sepsis Society, Jena, Germany

Corresponding Author:
Godwin Denk Giebel, MSc
Institute for Healthcare Management and Research
University of Duisburg-Essen
Thea-Leymann-Str. 9
Essen
Germany
Phone: 49 201 18 331
Email: godwin.giebel@medman.uni-due.de

Abstract

Background: Digitalization is currently revolutionizing health care worldwide. A promising technology in this context is artificial intelligence (AI). The application of AI can support health care providers in their daily work in various ways. The integration of AI is particularly promising in clinical decision support systems (CDSSs). While the opportunities of this technology are numerous, the problems should not be overlooked.

Objective: This study aimed to identify challenges and barriers in the context of AI-based CDSSs from the perspectives of experts across various disciplines.

Methods: Semistructured expert interviews were conducted with different stakeholders. These included representatives of patients, physicians and caregivers, developers of AI-based CDSSs, researchers (studying AI in health care and social and health law), quality management and quality assurance representatives, a representative of an ethics committee, a representative of a health insurance fund, and medical product consultants. The interviews took place on the web and were recorded, transcribed, and subsequently subjected to a qualitative content analysis based on the method by Kuckartz. The analysis was conducted using MAXQDA software. Initially, the problems were separated into “general,” “development,” and “clinical use.” Finally, a workshop within the project consortium served to systematize the identified problems.

Results: A total of 15 expert interviews were conducted, and 309 expert statements with reference to problems and barriers in the context of AI-based CDSSs were identified. These emerged in 7 problem categories: technology (46/309, 14.9%), data (59/309, 19.1%), user (102/309, 33%), studies (17/309, 5.5%), ethics (20/309, 6.5%), law (33/309, 10.7%), and general (32/309, 10.4%). The problem categories were further divided into problem areas, which in turn comprised the respective problems.

Conclusions: A large number of problems and barriers were identified in the context of AI-based CDSSs. These can be systematized according to the point at which they occur (“general,” “development,” and “clinical use”) or according to the problem category (“technology,” “data,” “user,” “studies,” “ethics,” “law,” and “general”). The problems identified in this work should be further investigated. They can be used as a basis for deriving solutions to optimize development, acceptance, and use of AI-based CDSSs.

International Registered Report Identifier (IRRID): RR2-10.2196/preprints.62704

(J Med Internet Res 2025;27:e63377) doi:10.2196/63377

XSL•FO
RenderX

KEYWORDS

decision support; artificial intelligence; machine learning; clinical decision support system; digitalization; health care; technology; innovation; semistructured interview; qualitative; quality assurance; web-based; digital health; health informatics

Introduction

Background

AI, and in particular machine learning (ML), is becoming more prevalent in health care systems worldwide. This is illustrated, for example, by the number of AI- or ML-enabled medical devices approved in the United States. Here, the annual increase in 2020, 2021, and 2022 was 39%, 15%, and 14%, respectively, and was forecasted to rise to ≥30% in 2023 [1]. In January 2024, the Food and Drug Administration listed 692 medical devices enabling AI and ML [1]. For Europe, no reliable statement can be made about the number of approved AI- or ML-enabled devices as the corresponding European database, the European Database on Medical Devices, does not provide a corresponding filter function [2]. This means that it is not possible to search explicitly for devices with AI or ML component.

AI-enabled devices cover a wide range of medical specialties, such as dermatology, diabetology, gastroenterology, gynecology, heart surgery, cardiology, nephrology, ophthalmology, orthopedics, pathology, psychiatry, and radiation oncology [3]. In the United States, in the first 7 months of 2023, a total of 79% of newly approved devices covered the field of radiology; 9% covered the cardiovascular field; 5% covered the field of neurology; 4% covered the field of gastroenterology or urology; 2% covered the field of anesthesiology; and 1% each covered the ear, nose, and throat as well as ophthalmic fields [1].

Analogous to the large number of medical specialties using AI, there are also many different types of applications. For example, AI can support the detection of infectious disease outbreaks; identify rare and common diseases by combining clinical, genetic, and many other laboratory test results; and support hospital business operations [4].

In addition to the various possibilities to benefit from AI in health care, there are also some concerns and problems with its use. These can be categorized in different ways. While Farhud and Zokaei [5], for example, identified “privacy and data protection,” “informed consent and autonomy,” “social gaps and justice,” and “medical consultation, empathy, and sympathy” as problem areas, Khan et al [6] identified “data collection and algorithm developing,” “ethical,” “social,” and “clinical implementation” as concerns.

An interesting application of AI in the medical context is its use in CDSSs. In general, CDSSs can be classified into knowledge based or non–knowledge based. While the knowledge-based systems follow clear rules (if-then statements), the non–knowledge-based systems use AI, ML, or statistical pattern recognition to make decisions on a data source [7]. Several approaches exist for ML. These include supervised learning, where algorithms benefit from information provided and labeled by humans (eg, by adding the final diagnosis to a set of clinical data), and unsupervised learning, where algorithms search for patterns in the underlying data on their own [8]. There is also reinforcement learning, where algorithms learn by interacting with an environment and receive feedback in the form of rewards to adapt and improve their actions [9]. AI-based CDSSs can be further distinguished based on the number of characteristics (eg, support on demand or unprompted) and in terms of their functions (eg, supporting diagnosis, outcome prediction, treatment planning, prescribing and managing medications, preventative care, chronic disease management, image interpretation, and many others) [10].

Regardless of AI, CDSSs also have several problems in addition to their many benefits. These problems include their content (“elimination or shifting of human roles,” “difficulty in keeping content current,” and “inappropriate content”) and their presentation (“rigidity of the system,” “alert fatigue,” and “potential for errors”) [11].

AI has the ability to evaluate large amounts of data where humans would fail because they cannot perceive the entire context beyond the data [12]. Thus, clinical decision-making can benefit from the use of AI. It can assist with the prognosis, diagnosis, and treatment of diseases; support the clinical workflow; and expand the availability of medical expertise [13]. However, in addition to the benefits, the problems and barriers of AI applications should be critically and responsibly considered rather than ignored [3].

Objectives

The use of AI-based CDSSs affects a large number of different stakeholders, such as physicians, caregivers, patients, and developers, and raises a multitude of different questions (eg, of legal or ethical nature). Therefore, in this research, various stakeholders were interviewed with the aim to identify potential problems and barriers related to AI-based CDSSs.

Methods

Overview

Qualitative research, concretely, semistructured expert interviews, was conducted to identify problems and barriers in the context of AI-based CDSSs. The decision to conduct qualitative research was made with the aim to approach the question exploratively because the use of AI is still not well established in health care systems worldwide. To ensure transparency in all aspects of our qualitative research, we followed the standards by O’Brien et al [14] and subsequently checked the manuscript by applying the 32-item COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [15].

Theoretical Framework

This study is part of a larger research project called KI@work (User-Oriented Requirements for AI-Based Clinical Decision Support Systems) funded by the German Federal Joint Committee. Building on the results of this interview study, a corresponding scoping review, and focus groups with physicians

XSL•FO
RenderX

and carers, a quantitative survey of physicians will be developed to investigate their preferences as well as problems in the context of AI-based CDSSs.

The aim was to collect the perspectives and opinions of experts. The interviews were conducted with stakeholders playing a pivotal role in the context of AI in medicine. On the basis of a previously conducted scoping review in the context of the project, a guideline was created and used to structure the interviews.

During the interviews, we asked about both problems and barriers as well as solutions to overcome them. This paper focuses especially on the problems and barriers identified in the underlying conversations.

Participant Selection

Before contacting potential participants, the project consortium discussed relevant stakeholders. During the discussion, they came to the consensus that, in addition to representatives of patients, service providers, AI developers, and health insurance funds, further experts with knowledge on regulation of AI as well as experts on ethical questions should be included.

Once the relevant stakeholder groups were determined, the University of Duisburg-Essen recruited corresponding experts with the help of the entire consortium. Before conducting the interviews, the stakeholders received an informative letter about the topic of the interviews—problems in the implementation of AI-based CDSSs in patient care—and further information about the procedure, including information about data protection, via email.

A total of 4 contacted stakeholders did not respond to the request. There was no clear refusal to participate in the interviews. After recruitment, during the interviews, none of the experts included dropped out of the study.

Setting and Data Collection

All the interviews were held online either via the videoconferencing platform integrated into Microsoft Teams or Zoom (Zoom Video Communications). Each expert interview was guided by a moderator and accompanied by at least one additional comoderator. All 3 moderators (GDG, PR, and NB) were from the Institute for Health Care Management and Research, University of Duisburg-Essen. While one moderator (NB) is head of the research team, the 2 others are scientific staff within the team. In total, 2 moderators were male (GDG and PR), and one moderator was female (NB). Of the moderators, 2 have a (health) economic background, and the third moderator holds a medical doctor degree in addition (NB). GDG and NB are experienced in conducting qualitative research, especially interviews. PR gained experience in conducting interviews through his previous work in market research. There were no relationships between the researchers and the interviewees. The researchers did not have a clear stance either for or against the use of AI in health care. No further circumstances or events influencing the conversations are known.

Preliminary results of a scoping review conducted within the project (to be published) before the stakeholder interviews served to develop a noninfluencing interview guideline. This guideline followed a uniform structure, including three main topics: (1) general problems with AI-based CDSSs, (2) problems in the development process of AI-based CDSSs, and (3) problems in the clinical use of AI-based CDSSs. Each of the topics started with open questions followed by more specific questions. Each topic included a core set of fixed questions for all stakeholders and was also slightly extended or adapted to the stakeholders to receive stakeholder-specific information (Multimedia Appendix 1).

Data collection took place between the beginning of June 2023 and the middle of August 2023. If the interviews were conducted using Microsoft Teams, they were automatically transcribed using the integrated function. The machine-generated transcripts were revised and pseudonymized by research assistants. If the interviews were conducted using Zoom, the recordings were transcribed by research assistants. In both cases, the transcripts were quality checked by 1 of the 3 moderators. The quality-assured transcripts were sent to the corresponding stakeholders interviewed to allow for requests for changes within 1 week (if no request has been made to extend the deadline). During the study, no adjustments (except for the stakeholder-specific questions in the guideline) were made to the method of data generation as the interviews worked without problems and the guideline proved to be understandable.

Data Analysis

After incorporating the requested changes by 3 stakeholders, the transcripts were loaded into MAXQDA (VERBI Software GmbH), and video recordings were deleted. Subsequently, data analysis started in September 2023 and was completed in October 2023.

As recommended by Kuckartz [16], deductive codes were defined before data analysis and applied subsequently to the transcripts (GDG). Deductive codes were determined according to the structure of the interview and included the 3 topics of the guideline (“general,” “development,” and “clinical use”).

The statements identified were concisely summarized and printed on individual pieces of paper (GDG). In a subsequent workshop (GDG, NB, and PR), the printed statements that were thematically similar were grouped together into clusters (“problem areas”). Whenever a problem or barrier arose that did not fit into an existing cluster, a new cluster was created. The clusters were named appropriately according to the problems and barriers they contained.

Subsequently, thematically similar clusters were organized into higher-level groups (“problem categories”) and named accordingly. Finally, the stakeholder statements and their respective summaries were sorted into a matrix with deductive codes (“general,” “development,” and “clinical use”) on the horizontal axis and inductive codes (“problem categories” and “problem areas”) on the vertical axis.

Ethical Considerations

As the topic of the interviews was not a sensitive subject and sociodemographic or other potentially identifying characteristics of the interviewees were not collected, there was no need for

XSL•FO
RenderX

approval by an ethics review board. This was also confirmed by the Ethics Committee of the Faculty of Medicine of the University of Duisburg-Essen. During the introduction, before the interviews, the participants were informed about the exact procedure and conditions and asked for their consent to participate. Through pseudonymization and deletion of the recordings, reassignment of statements to participants was no longer possible. On request, the experts received an expense allowance of €120 (US $124.27) for their participation. In the end, 4 of the 15 interview partners did so.

Results

Overview

A total of 15 interviews were conducted with 17 experts. In total, 13% (2/15) of the interviews conducted were double interviews with 2 interview partners in each. Experts included all requested stakeholder groups: a patient and a physician representative, developers of AI systems, consultants, experts in the fields of quality assurance and quality management, caregiver representatives, members of research institutions in the fields of AI and health care as well as social and health law, a representative of a health insurance fund, and an expert from an ethics committee. In general, intrinsic motivation and participation were high. An overview of each interview is provided in Table 1.

Table 1. Conducted interviews.

Number	Stakeholder	Setting	Date
1	Caregiver representative	Individual interview	June 5, 2023
2	Quality management representative	Individual interview	June 5, 2023
3	Researcher—AI[a] in health care	Individual interview	June 6, 2023
4	Medical product consultant	Individual interview	June 16, 2023
5	Caregiver representative	Individual interview	June 27, 2023
6	Caregiver representative	Double interview	June 27, 2023
7	Quality assurance representative	Double interview	July 4, 2023
8	Representative of an ethics committee	Individual interview	July 17, 2023
9	Researcher—social and health law	Individual interview	July 19, 2023
10	Medical product consultant	Individual interview	July 27, 2023
11	Developer of AI-based CDSSs[b]	Individual interview	July 27, 2023
12	Patient representative	Individual interview	July 31, 2023
13	Physician representative (intensive care)	Individual interview	August 8, 2023
14	Developer of AI-based CDSSs	Individual interview	August 11, 2023
15	Representative of a health insurance fund	Individual interview	August 17, 2023

[a]AI: artificial intelligence.

[b]CDSS: clinical decision support system.

The interviews provided 309 statements on problems and barriers in the context of AI-based CDSSs. A total of 7 problem categories were identified: “technology,” “data,” “user,” “studies,” “ethics,” “law,” and “general.” The categories with their corresponding problem areas are shown in Figure 1.

XSL•FO
RenderX

Figure 1. Identified problem categories and corresponding problem areas. AI: artificial intelligence; CDSS: clinical decision support system.

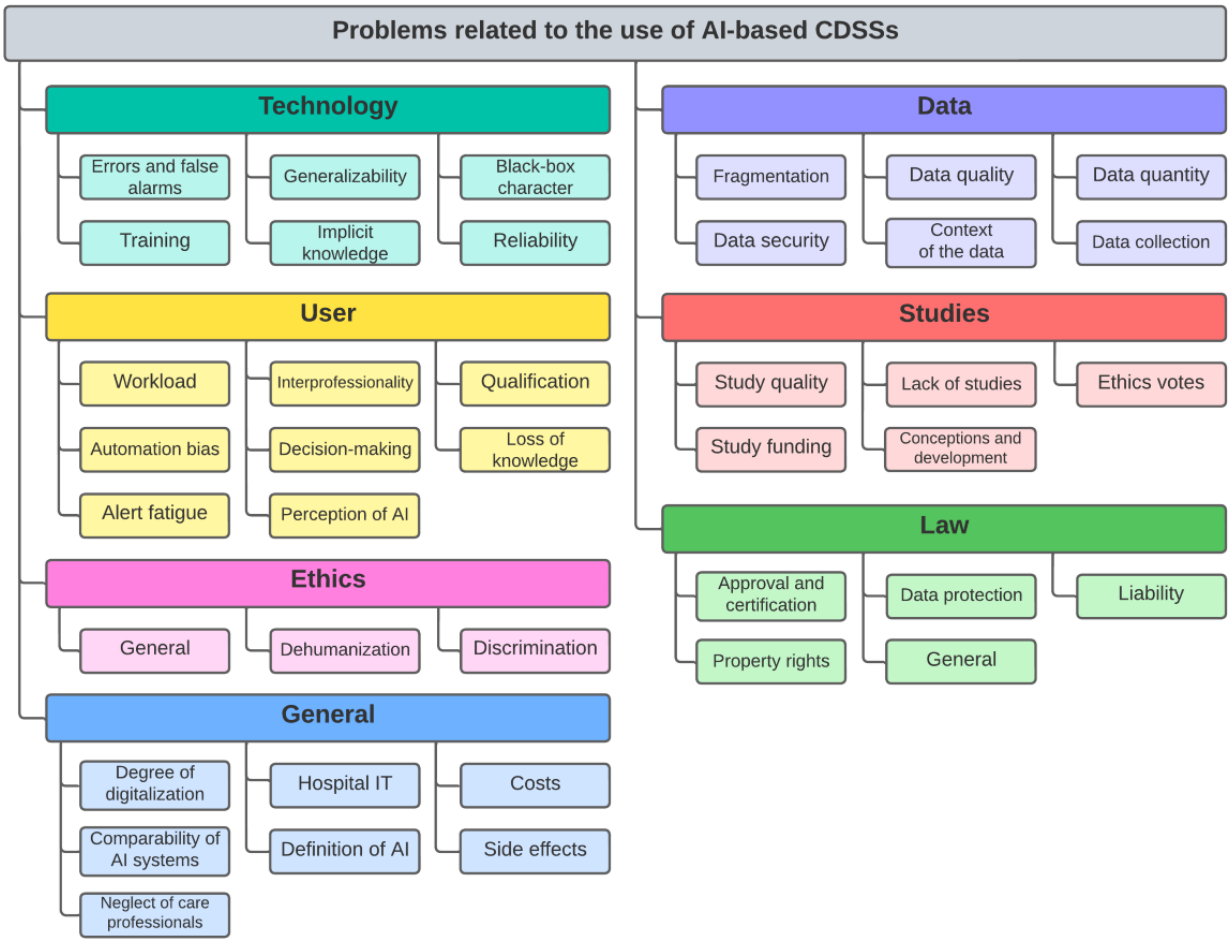

Even if the aim of qualitative research is more explorative than deterministic, it should be mentioned that some problem categories were more often thematized than others (user: 102/309, 33%; data: 59/309, 19.1%; technology: 46/309, 14.9%; law: 33/309, 10.7%; general: 32/309, 10.4%; ethics: 20/309, 6.5%; studies: 17/309, 5.5%). The problem categories are described in the following sections, and relevant expert statements are reproduced to illustrate the problems. The order of the problem categories does not indicate their relevance. A final list with the systematization of all identified problems can be found in Multimedia Appendix 2.

Technology

Problems described in the "technology" category were directly related to the technology underlying AI-based CDSSs. Named problems concerned 6 areas: "errors and false alarms," "generalizability," "black-box-character," "training," "implicit knowledge," and "reliability."

"Errors and false alarms" were observed in clinical use and in general. While incorrect recommendations resulting from use were seen as a threat to patients, a more general problem in this context was the tension between sensitivity and specificity:

> *[W]ith every AI procedure...there exists an area of conflict. So do I tend to develop the system more towards the direction of being more sensible, so to speak. In other words,...if it triggers, there's a high probability that it's actually correct. But then you lose, then suddenly there might be cases that you don't diagnose....* [Developer of AI-based CDSSs]

In addition to weighing up individualization and generalization during the development process and the system adaptation to local data before use, named problems with "generalizability" were of a more general nature. These problems were observed particularly in relation to regional validity, hospital-specific validity (eg, specialized or smaller hospitals), and population-based validity (eg, genetic). As AI is always developed on a distinct database, experts questioned whether the systems were ubiquitously valid. One expert described that the quality of the results in the real setting is not comparable with the quality of the results based on training and test data:

> *Yes, I see problems with generalizability. We hear from several developers that if they have a data set of a hundred thousand patients, take 50,000 for training and test it with the other 50,000, then they get good results. If they take data outside of this cohort, the results are different.* [Medical product consultant]

Furthermore, the "black-box-character" problem was thematized. In general, more complex AI systems tend to be not comprehensible to humans. As a result, users have a lack of understanding of AI-proposed decisions. This is not only a general problem but also a concrete problem during clinical use, as described by a caregiver representative:

XSL•FO
RenderX

> *[T]raceability is a very important factor for such a system; if I, as a user or collaborator, don't understand it, then the collaboration doesn't work well either.* [Caregiver representative]

During the development process, problems can also occur in the "training" of the AI. This is the case if the training is suboptimal or even poor and biased. This might be due to (1) an insufficiently analyzed database, (2) biased training because of too specific datasets, or (3) the system learning from lowly qualified users. Another problem was observed if the development phase was not completed before actual clinical use (eg, due to the prospect of quick profits). If the development phase is not supervised, so-called drifts can emerge, and the validity of systems might change over time.

Further problems were expressed in the context of "implicit knowledge." As this type of knowledge cannot be represented in words or numbers, experts explained that it cannot be integrated into AI-based CDSSs. Problems in this context concerned mainly clinical use because experts stated the need for this knowledge and warned against neglecting it by using AI. Furthermore, it was assumed that the lack of implicit knowledge might lead to a reductionist patient image:

> *We don't have the context that doctors sometimes have when the patient is right in front of them, and they see him as a human being and ask a few questions. So that's why it's very dangerous, because we end up with a very reductionist image of a patient as far as the AI system is concerned.* [Developer of AI-based CDSSs]

"Reliability" was the last problem area thematized within "technology." A general problem leading to lack of reliability was observed in the functionality of AI, which is based on probabilities and statistics. According to one expert, these would not always lead to the right decisions:

> *[T]he problem with AI is that it is not a physical algorithm, but statistics. That means there will always be things where you can be wrong.* [Medical product consultant]

Furthermore, such systems would neither critically question inputs nor provide absolute reliability. Premature deployment in resource-critical care situations was seen as a particular risk.

Data

"Data" as the underlying input of AI-based CDSSs are essential for the development and use of these systems. During the interviews, numerous problems were described in the context of data. These included "fragmentation," "data quality," "data quantity," "data security," "context of the data," and "data collection."

The first problem in this category is the "fragmentation" of the health care system and, subsequently, the fragmentation of data. Problems concerning "fragmentation" usually have a general background but have a strong impact on the development process. In particular, the German health care system is characterized by a great variety of institutions and actors with different interests, making it very complex. Breaks exist at the boundaries between different sectors (eg, between inpatient and outpatient supply), which means that data are not collected uniformly and, therefore, are often not comparable. In some cases, there was a lack of uniform standards. These problems are also reflected in the development process of AI-based CDSSs. Nonharmonized, nonunified, and unprocessed data, as well as a multitude of different actors, lead to major barriers for the development of these systems. One expert described a lack of standardization as a problem that hinders the comparability of data:

> *Well, first of all, I would say that it heavily relies on the training data, that we have different standards there. And the question is to what extent they can be harmonized or standardized, so that you're not comparing apples with oranges....* [Representative of an ethics committee]

Once the data had been collected, their quality was described to be often insufficient. It was even stated that data do not always represent reality. The insufficient "data quality" is especially a problem during the development process. In this context, the "garbage-in, garbage-out" principle was mentioned. It states that, if the data entered are not good, the results will not be good either:

> *Because that's the garbage-in, garbage-out principle; the worse the data you put in, the worse the result will be.* [Developer of AI-based CDSSs]

A threat in the context of low-quality data was observed in the use of suboptimal data if no other data exist. This might lead to the development of an AI-based CDSS that produces wrong results and is subsequently inappropriate for clinical use.

Experts also saw a problem in "data quantity." While, on the one hand, there was a general lack of data, experts such as a researcher in the field of AI in health care mentioned especially a lack of structured and quality-proved data on the other hand:

> *I think there is a quality problem with the data, that's undisputed....there is a massive problem, but our main problem is the quantity of data that is available at all.* [Researcher—AI in health care]

Furthermore, it was mentioned that some data are not available in digitized form. A special case was observed in rare treatment situations as these would not provide a sufficient data basis. In contrast to the lack of data, one developer mentioned an uncontrolled flood of data.

"Data security" was reported as a further problem. In general, the experts themselves questioned data security and mentioned that patients are particularly afraid of data leaks, too. According to the experts, those leaks and problems with the IT and its cybersecurity could emerge during the clinical use of AI-based CDSSs. A threat was observed in hacker attacks and subsequent manipulation or deletion of training data or algorithms, as well as system failures. The risk of being affected by hacker attacks and the absolute danger that this poses were described as being increased in the context of AI:

> *Data protection is certainly a problem, the more AI you have, the more you are likely to be affected by*

XSL•FO
RenderX

> *external hacker attacks and so on. So, because if I have access to the system, the consequences of a hacker attack will be on another level.* [Representative of a health insurance fund]

A problem in clinical use could emerge if the (medical) "context of the data" changes. Therefore, the quality of the information provided by AI-based CDSSs could change. An example given by an expert of such an altering context was a changing understanding of diseases or a change in diagnoses over time:

> *So, you can't just say that this is a technical problem. You always have to ask again about the current state of science. Is it still valid if I take the data from eight or ten years ago, because sometimes the concept of illness changes, e.g. in the case of psychological experiences, but also in the case of other diagnoses.* [Quality assurance representative]

Finally, it was explained that the "data collection" process might be a problem. Experts stated that only a fraction of treatment data is available in digital form and that the extraction of analog data can be very complex.

User

A major problem category was the "user" of AI-based CDSSs. Related problem areas were "workload," "interprofessionality," "qualification," "automation bias," "decision-making," "loss of knowledge," "alarm fatigue," and "perception of AI."

Lack of time is omnipresent among physicians and caregivers. Therefore, experts warned of additional time expenditure, especially in the context of clinical use. Nevertheless, the lack of time and the additional workload were also a problem during development because caregivers were therefore not able to insert data for training purposes. In the context of clinical use, an additional workload was stated, assumed, considered, or feared by different experts:

> *But in cases where someone has to enter something, you must be careful, what does the doctor have to do, are you taking time away from the patient again?* [Quality assurance representative]

Experts stated that some data cannot be collected automatically and, thus, would lead to a need for additional personnel. Potentially necessary manual corrections of incorrect measurements would be required. Furthermore, a problem related to workload was the attempt to misuse AI-based CDSSs to cut staff costs.

A problem area concerning development was "interprofessionality." Generally, experts stated the need for interdisciplinary development of an AI-based CDSS. Otherwise, they foresaw problems in clinical use or even stated that, without interdisciplinary development, no useful systems would be developed at all. In particular, the lack of involvement of medical and care professionals was expressed by the experts. They mentioned that medical professionals are more likely to notice professional errors during development but also that they tend to underestimate the effort to develop an AI-based CDSS. A further problem with the involvement of physicians is that the development of an AI-based CDSS is usually only a secondary activity for them. The caregiver representatives felt that caregivers are insufficiently involved in the development of such systems, and thereby, important caregiver knowledge would not be incorporated. Development problems are also aggravated by the lack of skilled workers with knowledge of AI.

Once started, development can be difficult because of nonfunctioning cooperation between different professions, the time required for interprofessional communication, different specialist terminology, and conflicts of interest between actors involved:

> *Yes, so I would totally agree. So that's clear in such a—especially in the development phase, there are various stakeholders and various interests that simply come together, and to even get into the discourse, to have a conversation and to unite them, I think that's a very high art.* [Caregiver representative]

The "qualification" of users can also be a problem. It was stated by the experts that, in general, the users are poorly qualified for the use of AI-based CDSSs. They sometimes lack digital experience or further training. A caregiver representative mentioned a low intrinsic motivation of care professionals to deal with the topic of AI-based CDSSs. A lack of experience might also lead to more difficult interpretation and communication in the context of the systems. One expert explained that a lack of training was a problem and that it was not enough to train just 1 employee. Otherwise, it was likely that further problems would result:

> *I think one problem is the training. No matter who uses it, I must train these people. And I believe that's always a source of error, so it's shown to one person. The entire staff that has to use it simply isn't thoroughly trained. Then mistakes can of course happen in the application of this AI.* [Representative of a health insurance fund]

Experts described that users tend to prefer suggestions from (AI-based) CDSSs and ignore contradictory information provided without automation even if it is correct. This was described as blind trust, blind replying, uncritical adoption, or "automation bias." It was assumed that automation bias can lead to inadequate action and potentially jeopardize patient health. The problem was reinforced due to time pressure, stress, and a high shortage of skilled workers. This was especially emphasized by a caregiver representative:

> *So, I definitely see the danger that, especially against the background of such a fast clinical routine, the recommendation is simply adopted without being asked.* [Caregiver representative]

In addition to automation bias, further problems exist within the "medical or caregiving decision-making" process. Thus, AI-based CDSSs can be seen as a decision-making barrier in the case of disagreements. They might endanger the autonomy of users or undermine the medical and nursing expertise. An expert described this new knowledge asymmetry in which the patient is ahead of the physician as follows:

XSL•FO
RenderX

> *And I believe that quite a lot of doctors, and certainly a few nursing experts and nurses, see in it a bit of a curtailment of their field, their expertise, their unique selling point. I mean, when AI-based applications are brought into the public, I, as a person who may even be suffering from a rare disease and knows more through these support systems than my treating doctor, then they have to deal with that too, yes.* [Quality management representative]

A further problem area expressed was "loss of knowledge." In particular, the risk that the regular use of the systems would result in users no longer being able to access their own knowledge, which would then diminish, was described:

> *I see the risk of knowledge loss equally significant in all areas, so we're handing something over to AI in all aspects of life. And we are also losing competencies as a result.* [Caregiver representative]

Furthermore, experts explained that too many (false) alarms might lead to "alert fatigue." This means that system users deliberately ignore alerts or no longer manage to follow them up. It was described that the number of alerts produced across all software systems would be overwhelming and lead to alert fatigue:

> *That's a huge problem, especially false alarms, regardless of artificial intelligence. I mean, all of our or many of our software systems produce alerts. But in such a volume that one can't even follow up on them anymore, and then, of course, we're also talking about alert fatigue.* [Quality management representative]

Finally, the "perception of AI" was described as a problem area emerging from the user side. The problem partly originated in negative communication. The perception described ranged from low or lack of trust to doubts about validity and benefit to a partly irrational fear. A negative perception was observed among patients and health care providers but also among authorities and notified bodies. Health care providers described a fear of being replaced by AI. Excessive expectations of AI and a lack of awareness of the limits of AI systems were also reported. One expert described the perceptions in both directions (positive or negative) as extreme.

In clinical use, it was observed that distrust among users might emerge due to an undermining of medical competence or due to measuring physicians' diagnostic quality. A lack of trust was associated with a lack of added value provided by the systems. During clinical use, the perception of patients toward AI-based CDSS might also be impacted. This could happen through perceived false alerts or misleading medical wording. One expert mentioned that a big problem lies in failing to open up to AI and focusing too narrowly on its problems:

> *And I think, that's one of the biggest problems with AI-based solutions, if I don't approach the whole issue decisively or openly. That means, if I only see all the risks, if I'm somehow maybe even fear the AI-powered, self-evolving robot that might wipe out humanity, so I'm exaggerating it to some extent, but there are still some people out there who have this idea.* [Medical product consultant]

Studies

A further problem category was "studies," which are necessary to examine and demonstrate the evidence in terms of clinical effectiveness, reliability, and real-world applicability of AI-based CDSSs. This included the "study quality"; a "lack of studies"; and problems with "ethics votes," "study funding," and "conception and development."

In general, experts questioned the quality of available studies. They criticized that studies are not conducted with conscientiousness and that results may be manipulated. Furthermore, a risk of low-quality evidence in studies was described. An expert described the varying quality of studies that must be considered:

> *That's just—and I believe that often, work is not done properly enough, because people simply think, "Oh, it's just a study, and it'll be fine." If we look closely...we often find that, well, not all studies are created equally right.* [Representative of a health insurance fund]

In addition to the quality, the "lack of studies" was expressed to be a problem. It was mentioned that there are only few study results from Germany available. In particular, a lack of valid and reliable studies, prospective studies, and studies on the additional benefits of the systems was named:

> *The problem with all these AI algorithms so far is that they are based on retrospective data. And what we need is a prospective—or prospective studies with algorithms, to prove that there is actually a prospective advantage compared to standard treatment.* [Physician representative—intensive care]

"Ethic votes" were described as being an obstacle to study conduct. In detail, according to an expert, this means that ethic votes are usually necessary but might be rejected:

> *[A]nd even if we had data, we would need appropriate ethics approvals for the data, unless it is de-identified, which simply takes a very long time. And I have actually heard from colleagues that such ethics approvals have been rejected because they were supposed to be used for technical AI studies.* [Researcher—AI in health care]

A further problem area in the context of studies was "study funding." The experts called for more funding to be able to conduct studies to prove the evidence of AI-based CDSSs. This was described by a physician representative as giving up medical success due to lack of study funding for fear of AI.

Finally, problems were mentioned in the context of "conception and development." A threat was observed if developers were not independent of influence. Experts questioned the objectivity and knowledge base of algorithms and the transparency and trustworthiness of the development process. The latter was partly described as exploratory instead of purposeful, and the lack of a holistic view of the data was criticized. This was illustrated with the help of a metaphor by an expert:

XSL•FO
RenderX

> *So, if you collect data for the sake of data and then run it through a system, with some machine learning algorithm, and then get something nice out of it, it's just like knowing: Somewhere in a pond there are 20,000 different species of fish. I throw my fishing rod in there, catch a carp and afterwards I end up saying, that's exactly what I wanted. So maybe we should also think about what kind of data do I need for what, so what are my outcomes....* [Caregiver representative]

Ethics

The problem category "ethics" included "general," "dehumanization," and "discrimination." While "general" only consisted of general concerns about ethical problems, the other problem areas were more specified.

During clinical use, "dehumanization" of patients was feared. This was due to the dehumanization of the medical treatment and the risk of deteriorating communication with patients. Experts described that patients might be perceived as data points and their individual stories might not be considered:

> *But the question is, how can you link this in a meaningful way with general practitioners who actually see the patient and their history and not just specific data points?* [Representative of an ethics committee]

The most discussed problem area in the context of ethics was "discrimination." These problems occurred within the "development" category, concerned clinical use, and were partly general. General problems were observed in the reidentification of small population groups and a special paradox:

> *I see it as a bit of a paradox that, on the one hand, you of course want to collect data from as many minorities as possible to make the data more representative and, on the other hand, you have to classify them and that this can lead to new forms of stigmatization.* [Representative of an ethics committee]

An inherent problem of the development of an AI-based CDSS was observed in the dependence on the underlying data. The latter were regularly discriminatory or at least imbalanced between different patient groups. One such discrimination was observed in primarily male-oriented data. Once discrimination exists within the data, the AI developed adopts these patterns and reflects them in clinical use. Thus, a threat was described that the use of AI-based CDSSs might result in discrimination, for example, in terms of social inequalities or racist behavior. That this problem is already subject to research was described by a medical product consultant:

> *Basically, I know that there are a lot of AI research activities to see whether AI behaves in a racist way. So, there are definitely some research projects that point in that direction.* [Medical product consultant]

Law

In terms of law-related problems, experts named the problem areas "approval and certification," "data protection," "liability," "property rights," and "general." The latter encompassed that AI-based CDSSs are currently not sufficiently reflected in medical law.

In the context of "approval and certification," 2 problems were described. First, the certification procedure for AI-based CDSSs was described as overly complex and costly. Among other things, additional work due to different requirements for different authorization regions was named. Especially for start-ups with scarce resources, the framework conditions were described as too cost intensive. Second, experts stated that there are regulatory issues in the context of medical device authorization and uncertainties with and an insufficient number of certifying bodies:

> *The regulator hurdle is the certification as a medical device. There are currently only four notified bodies in the EU for the certification of software as a medical device. And that means we have waiting lists.* [Medical product consultant]

"Data protection" was found to be a general problem but also a problem during clinical use. For clinical use, on the one hand, there were experts that stated that data protection was a real problem. On the other hand, experts stated the problem that data protection was used as a pretext against the use of AI-based CDSSs or to avoid renewal. Real data protection problems in clinical use were observed, for example, in insufficient reading and writing permissions in the digital patient file for caregivers due to data protection.

In general, requirements for data protection and data security were described as cumbersome, and their implementation was seen as problematic. In particular, data that cannot be anonymized and automated data processing were seen as special issues. To some extent, experts described that data protection would hamper the exchange of data even within the same hospital and would lead to poorer data. In the case of withdrawal of consent to use data, identification of specific data among vast amounts of data might be a problem. Finally, it was criticized that data protection was prioritized over health protection:

> *[M]y view is that we have a big problem when it comes to data protection. In my opinion, in Germany or at least in my perception, data protection is more important than health protection.* [Quality management representative]

Furthermore, experts saw problems with "liability." There were concerns that physicians were always liable for decisions even if the errors stemmed from AI-based CDSSs. It was stated that the problem of liability is a difficult question and that, in some cases, responsibility remains unclear or is not regulated at all:

> *Is it actually the user or is it the person who developed the system? Who is actually responsible afterwards if there is some damage because the system did not function properly?* [Researcher—social and health law]

Finally, the issue of "property rights" was raised. Given the multidisciplinary development processes of AI-based CDSSs, it is not always clear who has sovereignty over data and

XSL•FO
RenderX

development results. Therefore, conflicts over ownership of data and patents might arise.

General

The last problem category was "general." It included all problems that could not be matched to another category. It consisted of (nonlegal) framework conditions such as "degree of digitalization," "hospital IT," "definition of AI," or "costs" on the one hand and of more direct problem areas such as "comparability of AI systems," "side effects," or "neglect of caring professions" on the other hand.

Different experts described that there was a lack of digitalization or that the "degree of digitalization" in the health care sector was too low. Furthermore, a lack of integration and, subsequently, a lack of data availability between sectors and institutions was described:

> *If you look at the overall process that a patient goes through, it is ultimately still too analogue and too disruptive and not digitized enough. That's a big problem overall as a basis for AI-based support systems. Because ultimately, you need a relatively large amount of data and also from a specific patient in order to compare it accordingly.* [Physician representative]

A problem area in clinical use was the "hospital IT." It was described as inadequate and lacking in technical equipment, its reliability was questioned, and a lack of IT staff was mentioned. A lack of system compatibility; a lack of interfaces between individual departments; and the unavailability of standardized, structured data within the hospital were more concrete problems. Furthermore, it was described that hospitals are not considered attractive places to work for IT specialists and that their pay would be too low. The unattractiveness of hospitals for IT specialists was, for example, described by a researcher in the field of AI in health care:

> *The costs are exploding of course. The IT sins have of course piled up everywhere and you also know that hospitals are not seen as desirable workplaces for IT specialists.* [Researcher—AI in health care]

Furthermore, problems with "costs" were observed during development in the lack of financial resources for projects, as well as the implementation costs being a barrier for clinical use. A problem that prevents cost savings was observed in AI-based CDSS manufacturers aiming to maximize profits. Within the remuneration system, the lack of specific billing codes for algorithm-based laboratory test results and the absence of separate billing options for AI systems were criticized:

> *Then we certainly have another issue, that the costs for these systems are currently not separately billable. Instead, they are part of the medical service, and you can take the position, well, that's just how it is, because the reimbursement code exists for the image and the diagnosis, and if the doctor buys additional software to save time in diagnosis, then that's an economic effect on his side, which the system doesn't have to worry about.* [Medical product consultant]

A challenge in deciding between different AI-based CDSSs, or whether to adopt such a system at all, could arise from the insufficient "comparability of AI systems." In addition, the lack of reasonable data to make health economic decisions was criticized:

> *I'm not sure whether those who are supposed to use it have sufficient criteria to decide which system to buy and how to compare the different systems with each other. So how do the different systems on the market actually compete with each other in terms of quality?* [Medical product consultant]

A problem expressed by caregiver representatives was a "neglect of caring professions" within the context of AI-based CDSSs. A lack of funding for the development of AI in the care sector, as well as a strong focus on medical applications resulting in a deficit in nursing applications, was described.

Furthermore, the occurrence of "side effects" in the clinical use of AI-based CDSSs was named as a problem requiring evaluation and follow-up to guarantee patient safety.

Finally, experts thematized different "definitions of AI" as a problem:

> *So, if you tell someone in practice that this is artificial intelligence, then there are an incredible number of opinions and ideas about what it could actually be.* [Caregiver representative]

Discussion

Principal Findings

We identified 7 different stakeholder-relevant problem categories ("technology," "data," "user," "studies," "ethics," "law," and "general"). While "technology" included problems that are directly related to the technology of AI-based CDSSs themselves, the other problem categories concerned the environment and the circumstances in which the systems are conceptualized, developed, studied, and used.

While our systematization is fundamentally new, many of the problems identified were in accordance with the literature on CDSSs. For example, Sutton et al [7] described the following problem areas in the context of CDSSs: "fragmented workflows," "alert fatigue and inappropriate alerts," "impact on user skill," "dependence on computer literacy," "system and content maintenance," "operational impact of poor data quality and incorrect content," "lack of transportability and interoperability," and "financial challenges." Similar results were obtained in a review of medication-related CDSSs. Most reported factors influencing the acceptance of such systems were (a lack of) "usefulness," "relevance" of information, "ease of use," and "efficiency" [17].

While these problems were valid for CDSSs in general, our study showed that stakeholders also saw problems that are especially attributable to the integration of AI into these systems (eg, "generalizability," "black-box character," and "perception of AI").

XSL•FO
RenderX

Challenges of AI in health and medicine have been described, for example, by Rajpurkar et al [18]. They found problems similar to those identified in our research. Challenges were identified in "implementation," "accountability," and "fairness." Implementation challenges included model trust and data limitations; accountability challenges stemmed from regulations and responsibilities; and, finally, fairness challenges comprised ethical data use as well as equity and bias.

In the German context in which the study took place, the Ethics Committee of the German Medical Association provides a profound overview on ethical and legal problem areas in the field of AI. Ethical aspects are "trust and trustworthiness," "responsibility," "autonomy," "communication and empathy," "medical expertise," "risk of discrimination," and "data sovereignty and privacy." Legal aspects include "regulation of medical devices," "obligations in the context of examination or treatment methods," "duties of care of physicians," "patient information and consent," "liability," and "data protection and confidentiality" [19].

In addition to individual problems of CDSSs as well as AI, there is some research on factors influencing the acceptance of AI-based CDSSs. Factors described in this context include concern of system failure, overtrusting of the systems, locality of the systems, and an intuitive user interface [20]. Knop et al [21] listed further characteristics and human factors that influence and shape the relationship and collaboration between AI-based CDSSs and humans. Technological characteristics include "training data quality," "performance," "explainability or transparency," and "adapted output or adaptability." Human factors comprise "medical expertise," "technological expertise," "personality," "cognitive biases," and "trust" [21]. While technological characteristics included a part of the problems that we identified and categorized under "technology" and "data," we included human factors under the category "user."

The combination of AI and CDSSs can increase the relevance of individual problems. While traditional CDSSs typically provide transparent and understandable reasoning, the black-box nature of AI-based CDSSs can make it difficult to build trust in these systems. If, for example, the causes of alarms are not clear, this could lead to them being ignored. On the other hand, the need for scrutinizing proposed decisions is even more important in AI-based CDSSs. As the reason for a decision is not comprehensible and relies on statistics, physicians should be even more sensitized to automation bias to uncover wrong decisions.

In the context of AI-based CDSSs, a special focus during development should be placed on the creation of the data basis. As the systems learn from these data, there is a threat that wrong decisions will emerge if the data are inadequate.

A problem that arises specifically with AI-based CDSSs is liability for incorrect decisions. As the systems themselves cannot be held liable, it is conceivable that either the physicians using them or the vendors or health care organizations could be held liable [22]. However, it should be noted that there is no easy way of understanding black-box decisions, and therefore, the responsibility of physicians is not clear.

While the problems identified in our research are in accordance with the literature, it should be emphasized that some stakeholders saw certain topics as problems, whereas others did not. One example is "loss of knowledge." While one caregiver representative mentioned that it would be a problem, another stated that the systems are an upgrade for human decision-making instead of leading to loss of knowledge. Another such example was the problem of executing untested recommendations (automation bias). While most of the stakeholders saw a problem here, one stakeholder emphasized that, if the reliability of the systems was close to 100% in the future, it would no longer be a problem. One interviewee emphasized that, in general, he would rather refer to them as "to-dos" than as problems.

To increase physicians' acceptance of CDSSs, Khairat et al [23] proposed 2 different models. On the one hand, physicians should be included in the design process by examining user needs and expectations as well as the usability of prototypic designs. On the other hand, it is important for physicians to include the CDSS as a component of their decisions while maintaining professional autonomy. These approaches should also be pursued in the context of AI-based CDSSs.

Even if most problems of AI-based CDSSs were already known from CDSSs or AI as individual technologies, our investigation shows that they are still relevant. Our work in particular makes a contribution by compiling and systematizing the problems in the context of AI-based CDSSs. This provides an ideal starting point for further investigation of individual problem categories or manifestations.

Implications

AI-based CDSSs can lead to an improved physician performance and better medical outcomes [24]. Nevertheless, we identified several problems that are in accordance with the literature [7,18,19]. These should be solved as, by addressing errors, the potential of AI to improve the future of medicine can be realized [18].

Even if there are regulations for CDSSs, such as the 21st Century Cures Act in the United States and the Medical Device Regulation in the European Union, not all problems regarding this technology are covered [25]. McKee and Wouters [8] identified five challenges for regulators in the context of AI: (1) the AI application is only one part of a complex clinical system, (2) the training process may incorporate existing values and biases, (3) the performance will change over time using ML, (4) there might be conflicts with data protection legislation, and (5) gathering vast quantities of data might raise issues of privacy.

Given the heterogeneity of the problems in the context of AI-based CDSSs, one cannot assume a definite solution to be available. While many problems emerge from the technology behind AI-based CDSSs, there are other problems emerging from the environment in which the systems should be used. Therefore, existing risks cannot be solved only on the technical level but require interdisciplinary effort. To conceptualize valuable systems, clear communication between IT professionals

XSL•FO
RenderX

(developers) and health care practitioners (users) during development is needed [26].

Even if future systems can reliably derive decisions from databases, the treating physician should not be left out. Some problems, such as lack of implicit knowledge, dehumanization, and most of the ethical problems or the (medical) context of the data, cannot be sufficiently considered by the systems. Therefore, to achieve the best outcome for all actors within the health care system, the goal should be to combine the capabilities and strengths of AI and human intelligence [12,27]. To do so, it is necessary to identify the attitudes of relevant stakeholders toward the technology and its use. Such an approach was adopted by Laï et al [28]. In an interview study with 40 French stakeholders, they found that views on AI were very heterogeneous and that only cooperation between the involved parties would lead to the development of satisfactory AI tools.

As mentioned in the Methods section, this work is part of a research project—KI@work. The results of this interview study together with the results of a scoping review and focus groups with physicians and carers will be used to conceptualize a quantitative survey for intensive care physicians. Thereby, the relevance of individual problems from the point of view of physicians might be determined more precisely.

Limitations

Given the novelty of the technology of AI-based CDSSs, we conducted interviews to explore problems and barriers from the point of view of German stakeholders. Due to the qualitative methodology, we interviewed only representatives of each stakeholder group. Therefore, the results are not necessarily representative and should be interpreted with caution. Thus, the findings of this study should be seen as preliminary evidence to guide further qualitative or quantitative research rather than as definitive conclusions.

There are 2 further limitations of this study that should be mentioned. First, we did not calculate any agreement rate or other measurements regarding coding discrepancies. However, quality assurance was conducted via the internal workshop, which aimed to systematize the identified problems and barriers. Second, as the expert interviews took place in Germany, the interviews were conducted in German. Thus, the statements included in this paper were translated. After a first translation, it was checked within the consortium whether the content of the statements had been changed. Upon reasonable request, the original anonymized citations are available from the authors.

In this study, we aimed to obtain a comprehensive overview of problems and barriers from the point of view of different stakeholders. Nevertheless, we had a slight overrepresentation of care professionals. As the views of patients compared to clinicians differ regarding certain aspects, such as trust [29], further studies should explicitly focus on physicians and patients as key stakeholders of AI-based CDSSs.

As we interviewed German individuals and the German system has specific problems, such as low digitalization in health care [30-32], strict regulation regarding the handling of data [33], and a severe shortage of physicians and caregivers [34], the relevance and transferability of the problems to other health care systems should always be critically scrutinized.

Conclusions

Experts see several problems within the context of AI-based CDSSs. Problems concerned the technology itself as well as the context in which the systems are developed and used. In total, we identified 7 different problem categories, namely, “technology,” “data,” “user,” “studies,” “ethics,” “law,” and “general.” To guarantee a sustainable, safe, and effective integration of AI-based CDSSs, the problems identified in this study should be considered when developing and using these systems. Furthermore, problems that focus on regulation should also be taken into account by policy makers. As this study followed an explorative qualitative design, the results should be further investigated in additional qualitative and quantitative studies.

Acknowledgments

This study is part of a larger research project (User-Oriented Requirements for Artificial Intelligence–Based Clinical Decision Support Systems [KI@work]; project 01VSF22050). The project is funded by the German Federal Joint Committee. The funders had no influence on the study design, the conduct of the study, or the decision to publish or prepare the manuscript. The authors acknowledge support from the Open Access Publication Fund of the University of Duisburg-Essen.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guidelines.
[DOCX File , 17 KB - jmir_v27i1e63377_app1.docx]

Multimedia Appendix 2
Systematization of all identified problems.
[XLSX File (Microsoft Excel File), 99 KB - jmir_v27i1e63377_app2.xlsx]

XSL•FO
RenderX

References

1. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. U.S. Food and Drug Administration. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices [accessed 2024-06-11]
2. EUDAMED - European database on medical devices. European Commission. 2024. URL: https://ec.europa.eu/tools/eudamed/#/screen/search-device [accessed 2024-02-06]
3. Wehkamp K, Krawczak M, Schreiber S. The quality and utility of artificial intelligence in patient care. Dtsch Arztebl Int 2023 Jul 10;120(27-28):463-469 [FREE Full text] [doi: 10.3238/arztebl.m2023.0124] [Medline: 37218054]
4. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023 Mar 30;388(13):1201-1208. [doi: 10.1056/NEJMra2302038] [Medline: 36988595]
5. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. Iran J Public Health 2021 Nov;50(11):i-v [FREE Full text] [doi: 10.18502/ijph.v50i11.7600] [Medline: 35223619]
6. Khan B, Fatima H, Qureshi A, Kumar S, Hanan A, Hussain J, et al. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. Biomed Mater Devices 2023 Feb 08:1-8 [FREE Full text] [doi: 10.1007/s44174-023-00063-2] [Medline: 36785697]
7. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 06;3(1):17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
8. McKee M, Wouters OJ. The challenges of regulating artificial intelligence in healthcare comment on "clinical decision support and new regulatory frameworks for medical devices: are we ready for it?: a viewpoint paper". Int J Health Policy Manag 2023 Sep 19;12:7261 [FREE Full text] [doi: 10.34172/ijhpm.2022.7261] [Medline: 36243948]
9. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. Boston, MA: MIT Press; 2018.
10. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. Med Phys 2020 Jun 17;47(5):e228-e235 [FREE Full text] [doi: 10.1002/mp.13562] [Medline: 32418341]
11. Ash JS, Sittig DF, Campbell EM, Guappone KP, Dykstra RH. Some unintended consequences of clinical decision support systems. AMIA Annu Symp Proc 2007 Oct 11;2007:26-30 [FREE Full text] [Medline: 18693791]
12. Lee J. Is artificial intelligence better than human clinicians in predicting patient outcomes? J Med Internet Res 2020 Aug 26;22(8):e19918 [FREE Full text] [doi: 10.2196/19918] [Medline: 32845249]
13. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
14. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
16. Kuckartz U. Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. 4th edition. Berlin, Germany: Kiligry; 2018.
17. Westerbeek L, Ploegmakers KJ, de Bruijn GJ, Linn AJ, van Weert JC, Daams JG, et al. Barriers and facilitators influencing medication-related CDSS acceptance according to clinicians: a systematic review. Int J Med Inform 2021 Aug;152:104506 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104506] [Medline: 34091146]
18. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022 Jan;28(1):31-38 [FREE Full text] [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
19. Decision support for medical work through artificial intelligence. Central Ethics Committee at the German Medical Association. 2021. URL: https://www.zentrale-ethikkommission.de/stellungnahmen/kuenstliche-intelligenz-2021 [accessed 2024-02-06]
20. Fujimori R, Liu K, Soeno S, Naraba H, Ogura K, Hara K, et al. Acceptance, barriers, and facilitators to implementing artificial intelligence-based decision support systems in emergency departments: quantitative and qualitative evaluation. JMIR Form Res 2022 Jun 13;6(6):e36501 [FREE Full text] [doi: 10.2196/36501] [Medline: 35699995]
21. Knop M, Weber S, Mueller M, Niehaves B. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence-enabled clinical decision support systems: literature review. JMIR Hum Factors 2022 Mar 24;9(1):e28639 [FREE Full text] [doi: 10.2196/28639] [Medline: 35323118]
22. Khullar D, Casalino LP, Qian Y, Lu Y, Chang E, Aneja S. Public vs physician views of liability for artificial intelligence in health care. J Am Med Inform Assoc 2021 Jul 14;28(7):1574-1577 [FREE Full text] [doi: 10.1093/jamia/ocab055] [Medline: 33871009]
23. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]

XSL•FO
RenderX

24. Kruse CS, Ehrbar N. Effects of computerized decision support systems on practitioner performance and patient outcomes: systematic review. JMIR Med Inform 2020 Aug 11;8(8):e17283 [FREE Full text] [doi: 10.2196/17283] [Medline: 32780714]
25. Van Laere S, Muylle KM, Cornu P. Clinical decision support and new regulatory frameworks for medical devices: are we ready for it?: a viewpoint paper. Int J Health Policy Manag 2022 Dec 19;11(12):3159-3163 [FREE Full text] [doi: 10.34172/ijhpm.2021.144] [Medline: 34814678]
26. Wolfien M, Ahmadi N, Fitzer K, Grummt S, Heine KL, Jung IC, et al. Ten topics to get started in medical informatics research. J Med Internet Res 2023 Jul 24;25:e45948 [FREE Full text] [doi: 10.2196/45948] [Medline: 37486754]
27. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019 Jul;8(7):2328-2331 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_440_19] [Medline: 31463251]
28. Laï MC, Brian M, Mamzer MF. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
29. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
30. Thiel R, Deimel L, Schmidtmann D, Kart L. #SmartHealthSystems: Digitalisierungsstrategien im internationalen Vergleich. Bertelsmann Stiftung. 2018. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/smarthealthsystems [accessed 2024-04-29]
31. Blase N, Diehl A, Wasem J. Digitalisierung im Krankenhaus. In: Klauber J, Wasem J, Beivers A, Mostert C, editors. Krankenhaus-Report 2023: Schwerpunkt: Personal. Cham, Switzerland: Springer; 2023:165-178.
32. Stachwitz P, Debatin JF. [Digitalization in healthcare: today and in the future]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2023 Feb 17;66(2):105-113 [FREE Full text] [doi: 10.1007/s00103-022-03642-8] [Medline: 36648499]
33. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. European Parliament & Council of the European Union. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2024-04-29]
34. Köppen J, Busse R. Die Personalsituation im Krankenhaus im internationalen Vergleich. In: Klauber J, Wasem J, Beivers A, Mostert C, editors. Krankenhaus-Report 2023: Schwerpunkt: Personal. Cham, Switzerland: Springer; 2023:19-23.

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
COREQ: Consolidated Criteria for Reporting Qualitative Research
ML: machine learning

Edited by G Tsafnat; submitted 18.06.24; peer-reviewed by P Cimiano, L Cossu; comments to author 24.08.24; revised version received 17.10.24; accepted 13.11.24; published 03.02.25.

Please cite as:
Giebel GD, Raszke P, Nowak H, Palmowski L, Adamzik M, Heinz P, Tokic M, Timmesfeld N, Brunkhorst F, Wasem J, Blase N
Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study
J Med Internet Res 2025;27:e63377
URL: https://www.jmir.org/2025/1/e63377
doi:10.2196/63377
PMID:39899342

©Godwin Denk Giebel, Pascal Raszke, Hartmuth Nowak, Lars Palmowski, Michael Adamzik, Philipp Heinz, Marianne Tokic, Nina Timmesfeld, Frank Brunkhorst, Jürgen Wasem, Nikola Blase. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

AI for IMPACTS Framework for Evaluating the Long-Term Real-World Impacts of AI-Powered Clinician Tools: Systematic Review and Narrative Synthesis

Christine Jacob[1], PhD; Noé Brasier[2], MD; Emanuele Laurenzi[1], PhD; Sabina Heuss[1*], PhD, Prof Dr; Stavroula-Georgia Mougiakakou[3,4*], PhD, Prof Dr; Arzu Cöltekin[1*], PhD, Prof Dr; Marc K Peter[1*], Prof Dr

[1]FHNW, University of Applied Sciences and Arts Northwestern Switzerland, Windisch, Switzerland

[2]Institute of Translational Medicine, Department of Health Science and Technology, ETH Zurich, Zurich, Switzerland

[3]ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

[4]University of Nicosia, Nicosia, Cyprus

[*]these authors contributed equally

Corresponding Author:
Christine Jacob, PhD
FHNW, University of Applied Sciences and Arts Northwestern Switzerland
Bahnhofstrasse 6
Windisch, 5210
Switzerland
Phone: 41 62 957 29 78
Email: christine.k.jacob@gmail.com

Abstract

Background: Artificial intelligence (AI) has the potential to revolutionize health care by enhancing both clinical outcomes and operational efficiency. However, its clinical adoption has been slower than anticipated, largely due to the absence of comprehensive evaluation frameworks. Existing frameworks remain insufficient and tend to emphasize technical metrics such as accuracy and validation, while overlooking critical real-world factors such as clinical impact, integration, and economic sustainability. This narrow focus prevents AI tools from being effectively implemented, limiting their broader impact and long-term viability in clinical practice.

Objective: This study aimed to create a framework for assessing AI in health care, extending beyond technical metrics to incorporate social and organizational dimensions. The framework was developed by systematically reviewing, analyzing, and synthesizing the evaluation criteria necessary for successful implementation, focusing on the long-term real-world impact of AI in clinical practice.

Methods: A search was performed in July 2024 across the PubMed, Cochrane, Scopus, and IEEE Xplore databases to identify relevant studies published in English between January 2019 and mid-July 2024, yielding 3528 results, among which 44 studies met the inclusion criteria. The systematic review followed PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines and the Cochrane Handbook for Systematic Reviews. Data were analyzed using NVivo through thematic analysis and narrative synthesis to identify key emergent themes in the studies.

Results: By synthesizing the included studies, we developed a framework that goes beyond the traditional focus on technical metrics or study-level methodologies. It integrates clinical context and real-world implementation factors, offering a more comprehensive approach to evaluating AI tools. With our focus on assessing the long-term real-world impact of AI technologies in health care, we named the framework AI for IMPACTS. The criteria are organized into seven key clusters, each corresponding to a letter in the acronym: (1) I—integration, interoperability, and workflow; (2) M—monitoring, governance, and accountability; (3) P—performance and quality metrics; (4) A—acceptability, trust, and training; (5) C—cost and economic evaluation; (6) T—technological safety and transparency; and (7) S—scalability and impact. These are further broken down into 28 specific subcriteria.

Conclusions: The AI for IMPACTS framework offers a holistic approach to evaluate the long-term real-world impact of AI tools in the heterogeneous and challenging health care context and lays the groundwork for further validation through expert consensus and testing of the framework in real-world health care settings. It is important to emphasize that multidisciplinary

XSL•FO
RenderX

expertise is essential for assessment, yet many assessors lack the necessary training. In addition, traditional evaluation methods struggle to keep pace with AI's rapid development. To ensure successful AI integration, flexible, fast-tracked assessment processes and proper assessor training are needed to maintain rigorous standards while adapting to AI's dynamic evolution.

Trial Registration: reviewregistry1859; https://tinyurl.com/ysn2d7sh

(J Med Internet Res 2025;27:e67485) doi:10.2196/67485

KEYWORDS

eHealth; assessment; adoption; implementation; artificial intelligence; clinician; efficiency; health technology assessment; clinical practice

Introduction

Background

Artificial intelligence (AI) is profoundly transforming health care across a range of applications, enhancing both clinical outcomes and operational efficiency. In medical imaging, AI algorithms improve diagnostic accuracy by analyzing complex imaging data, such as from magnetic resonance imaging and computed tomography scans, for highly precise and rapid clinical diagnostics [1]. Decision support systems powered by AI assist clinicians in making evidence-based decisions by providing real-time data-driven insights and predictive analytics [2]. Large language models are increasingly used for generating detailed medical reports and streamlining triage processes by analyzing and summarizing patient data quickly and accurately [3]. In addition, innovative digital health technologies such as electronic skins use wearable sensor technologies and AI to offer continuous, real-time monitoring of various health indicators, further enhancing personalized care [4]. These advancements have the potential to contribute to a more efficient, accurate, responsive, and holistic health care, reshaping how patient care is delivered and managed.

Despite the growing body of literature on AI in health care, its implementation has lagged behind other industries [5,6]. Previous studies have highlighted substantial barriers to the successful adoption of AI in health care, including issues related to trust; potential risks of harm; accuracy and perceived usefulness; reproducibility; evidentiary standards; and ethical, legal, and societal concerns [7,8]. In addition, uncertainty surrounding postadoption outcomes further complicates the implementation process [7].

A significant barrier identified by health care leaders worldwide is that despite the emergence of various new frameworks for assessing AI in health care, most focus primarily on the quality of study methodologies or technical aspects [9,10]. There remains a lack of a comprehensive, systematic framework that assesses the real-world impact of AI and offers guidance on clinical implementation, monitoring, procurement, and evaluation [9,11]. Most research overlooks the complex, multistep process required for successful AI integration, leaving critical gaps in understanding how to effectively implement and sustain AI tools in clinical practice [9,11]. As a result, the adoption of AI in clinical practice has fallen short of expectations, with only a few algorithms showing sustained clinical impact [12]. This gap is often due to inadequate or incomplete evaluation and the lack of universally recognized standards for AI assessment. The limited understanding of AI's true added value in health care highlights the need for a more comprehensive evaluation framework [13-15]. To ensure confidence in the added clinical value and successful integration of AI into health care workflows, a practical, comprehensive tool is needed so that the translational readiness of AI systems can be evaluated. Current approaches assessing AI in health care often focus on foundational technical metrics such as sensitivity and specificity, which fail to capture the full clinical impact [13,16]. A robust valuation should encompass factors such as patient outcomes, effects on clinical decision-making, workflow efficiency, and the tangible benefits for patients to fully determine AI's true contribution to and impact on health care [10,17,18].

In the context outlined earlier, regulatory approval is an important milestone for demonstrating overall performance, although the scientific evidence supporting AI tools in health care remains limited compared to traditional medical standards [9,19]. In addition, new regulations are being introduced to keep pace with rapidly evolving AI technologies, such as the European Union (EU) AI Act, which aims to ensure the trustworthiness of high-risk AI tools including those used in health care [20]. Despite the potentially positive impact of regulatory frameworks on AI-related developments, a recent study revealed that nearly half of Food and Drug Administration (FDA)–authorized AI devices lacked clinical validation data, raising concerns about their safety and effectiveness [21]. Without robust clinical validation, these technologies could pose significant risks to patient care. Despite efforts to create reporting guidelines for AI in health care, such as Standard Protocol Items Recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI) [13], CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) [14], Standards for Reporting of Diagnostic Accuracy Studies–Artificial Intelligence [22], Checklist for Artificial Intelligence in Medical Imaging [23], Prediction Model Risk of Bias Assessment Tool–Artificial Intelligence [24], and others, a unified international consensus on the evaluation of AI-based tools has yet to be established. While these guidelines address key methodological issues and share significant overlap, indicating the importance of certain assessment criteria, the absence of a standardized, universally accepted framework remains a significant challenge [4]. This lack of consensus complicates the consistent evaluation and implementation of AI technologies in clinical practice.

XSL•FO
RenderX

Objectives

The goal of this study was to develop a comprehensive framework for assessing the impact of AI tools in health care. This involved synthesizing and consolidating the various evaluation criteria found in existing literature regarding the quality and impact of AI tools. On the basis of the outcomes of this study, we plan on validating the framework through expert consensus using the Delphi process. However, this validation effort will be addressed in the subsequent phase of the project and is beyond the scope of this foundational paper. This approach aims to create a rigorous, evidence-based structure for AI evaluation, ensuring its relevance and applicability in health care settings.

In doing so, we adopted the perspective of the World Health Organization (WHO) on AI in health care, defining it as "the ability of algorithms and software to analyze complex medical data and support health care providers by improving decision-making, predicting outcomes, and enhancing clinical efficiency" [25]. AI tools in health care span a broad spectrum of applications, such as (1) diagnostic support, (2) prognosis of diseases course, (3) personalized treatment recommendations, (4) patient monitoring, and (5) overall health management, driving innovation across the health care landscape [25].

To address this, a systematic review was conducted to offer a comprehensive and current analysis of the criteria used in existing research to evaluate the quality and impact of AI in health care, from technological, social, and organizational perspectives. The review also explores the potential implications of AI implementation for key stakeholders and offers recommendations on how to effectively assess AI-powered clinical tools under consideration for clinical impact. This study builds upon and extends the findings of a prior research project, which examined the sociotechnical assessment criteria for patient-facing eHealth tools, that is already published [26,27].

We believe the results of this review will provide valuable insights for clinicians, pharmaceutical leaders, insurance professionals, technology providers, and policy makers by presenting an up-to-date, thorough overview of the criteria used to assess AI-powered clinical tools. These insights will help stakeholders make informed decisions about which tools to implement, recommend to patients, invest in, partner with, or provide reimbursement for, based on their assessed quality and potential impact.

Methods

Overview

The methodology for this review was based on established best practices, specifically following the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines [28] and the Cochrane Handbook for Systematic Reviews of Interventions [29]. These frameworks were chosen to ensure a rigorous and methodologically sound approach to the systematic literature review process. All review methods were predetermined and documented in advance, with the protocol being publicly registered in the research registry (reviewregistry1859) to enhance transparency and accountability [30]. The primary research question guiding this systematic review was the following: What technical, social, and organizational criteria should be considered when assessing the quality and impact of AI-powered clinical tools? This question served as the foundation for the analysis and exploration of the criteria relevant to AI's evaluation in clinical settings. The study remained highly consistent with the initial protocol from a methodological standpoint, adhering to the predefined review question; search strategy; databases; inclusion and exclusion criteria; participants, intervention, comparators, and outcomes (PICO) framework elements; data extraction strategy; quality assessment; and data synthesis approach as originally outlined. The only variation from the protocol was in the presentation of the findings: rather than merely listing the results as an inventory of criteria, we organized them into a cohesive framework. This structured approach enhances both the memorability and practical applicability of the results in real-world settings.

Search Strategy

A comprehensive search of the PubMed, Cochrane, Scopus, and IEEE Xplore databases was conducted in July 2024 to identify relevant studies. The review was limited to peer-reviewed papers published in English between January 2019 and mid-July 2024. We focused on this specific time frame and limited the search to the last 5 years to ensure the findings reflect the most recent advancements and challenges, particularly with the emergence of new generative AI technologies. Going back further would have added limited value, as older studies may not capture the rapid technological shifts and evolving complexities that are relevant today. Only fully published research articles were included, while other formats, such as editorials and study protocols, were excluded from the analysis. In accordance with the Cochrane Handbook for Systematic Reviews of Interventions, we chose not to include articles sourced through manual reference list searches, as "positive studies are more likely to be cited," which could introduce bias [29].

This systematic review focused on AI-powered tools designed specifically for clinicians, excluding tools meant solely for patients or medical students as these will most likely not reflect the implementation aspects in real-world health care organizations. The search strategy targeted manuscripts with titles including the terms "AI" or "Artificial Intelligence," reflecting the intervention focus on AI technologies. Outcomes of interest were assessment criteria, captured through titles containing the terms "assessment," "assess," "evaluation," "evaluating," "effectiveness," "efficacy," "quality," "efficiency," "usability," or "usefulness," as well as abstracts mentioning "criteria," "framework," "method," "methodology," "methodologies," "measurement," "toolkit," "tool," "tools," "approach," or "scorecard." No condition-based restrictions were applied, aligning with a broad approach to capture all relevant studies on assessment methodologies for clinician-targeted AI tools.

Textbox 1 illustrates the search string designed using the PICO framework. To ensure the relevance of the retrieved papers, the search was mostly restricted to manuscript titles, focusing on studies that addressed AI assessment criteria comprehensively

XSL•FO
RenderX

rather than those evaluating specific tools or pilot studies. Because comparators were not relevant to this review, they were excluded from the search parameters.

Textbox 1. The search string according to the participants, intervention, comparators, and outcomes framework.

Participants: clinicians

- Focus on artificial intelligence (AI)–powered tools for clinicians, excluding those designed solely for patients or medical students.

Intervention: AI-powered clinician tools

- Focus is on AI-powered clinician tools: the search targeted manuscript titles containing the terms (AI OR "Artificial Intelligence").

Comparator: not applicable

- There were no restrictions on eligible conditions for inclusion.

Outcome: assessment criteria

- The search targeted manuscript titles also containing AND (assessment OR assess OR evaluation OR evaluating OR effectiveness OR efficacy OR quality OR efficiency OR usability OR usefulness). As well as manuscript tiles and abstracts containing AND (criteria OR framework OR method OR methodology OR methodologies OR measurement OR toolkit OR tool OR tools OR approach OR scorecard).

Study Selection

Two researchers (CJ and EL) participated in the screening, eligibility, and inclusion phases of the study. Any discrepancies during these stages were resolved through discussion among them. If consensus could not be reached, a third coauthor was consulted to make the final decision. The team used the open-source Rayyan app (Qatar Computing Research Institute) to streamline collaborative screening efforts [31]. The screening process took place between July and August 2024.

The inclusion and exclusion criteria, outlined in Textbox 2, were developed following the PICO framework. Included studies centered on AI-powered tools in clinical settings, addressing criteria to assess the quality and impact of these tools. Eligible studies were peer-reviewed, published between January 2019 and mid-July 2024, and written in English. Exclusions were made for studies involving only patients or medical students as they were not likely to reflect implementation factors, AI technologies outside clinical settings (eg, patient use chatbots), studies assessing specific tools in isolation, or frameworks solely evaluating AI research methodology or clinical trials rather than the implementation of the tools in real-world settings. Editorials, study protocols, and non-English publications were also excluded.

Following the completion of the screening process and resolution of any conflicting views among the researchers, CJ and EL proceeded to assess the full texts of the selected studies for eligibility. Any remaining disagreements were addressed through consultation with a third coauthor. CJ evaluated the risk of bias using the Critical Appraisal Skills Program (CASP) checklist [32], which assesses key quality criteria in the included studies. These criteria include the following: the presence of a clear statement of the research aims, the appropriateness of the methodology for the research objectives, the suitability of the research design in addressing those aims, the relevance of the recruitment strategy, the adequacy of data collection methods in relation to the research question, the consideration given to the researchers' roles, the evaluation of ethical issues, the rigor of data analysis, the clarity of the study's findings, and whether the researchers discussed the study's contribution to existing knowledge, such as its implications for current practice, policy, or relevant literature. The results of this appraisal are available in Multimedia Appendix 1.

XSL•FO
RenderX

Textbox 2. Inclusion and exclusion criteria according to the participants, intervention, comparators, and outcomes framework.

Inclusion criteria

- Participants: focused on clinicians
- Intervention: focused on artificial intelligence (AI)–powered clinician tools
- Comparators: does not apply
- Outcomes: addresses the different criteria used to assess the quality and impact of AI-powered clinician tools regardless of the condition
- Publication type: peer-reviewed and published papers
- Time frame: studies published between January 2019 and mid-July 2024
- Language: studies published in English

Exclusion criteria

- Participants: focused solely on patients or medical students
- Intervention: technologies used outside of clinical environments, such as chatbots used by patients to obtain health care information
- Comparators: does not apply
- Outcomes: individual assessments of pilot studies singling out specific tools, and assessment frameworks that focus on the reporting and methodological quality of AI research and clinical studies rather than evaluating the AI tool itself
- Publication type: editorials and study protocols
- Time frame: studies published before January 2019 or after mid-July 2024
- Language: studies published in languages other than English

Data Collection and Synthesis

The procedures and outcomes across the included studies were too diverse to support a quantitative analysis. As a result, a narrative synthesis was used following the sociotechnical approach, organized around the social, organizational, and technical criteria used to evaluate the quality and impact of AI-powered tools for clinicians. The authors were influenced by the sociotechnical theory, which emphasizes that the design and performance of innovations can only be fully understood when both social and technical aspects are considered as interdependent components of a larger system [33]. This approach aligns with recommendations from several scholars who advocate for moving beyond purely technology-focused frameworks to incorporate the broader context, including societal and implementation factors [34-36]. To facilitate this process, NVivo (version 1.7.2; Lumivero), a qualitative data analysis software, was used.

Data coding began with a preliminary extraction grid, which was structured around themes derived from previous research and established technology acceptance frameworks. The initial codebook was informed by our prior work on factors influencing eHealth evaluation and adoption [26,27,36-38], with additional codes being incorporated as new themes emerged during the review. Thematic analysis, as outlined by Braun and Clarke [39], was conducted to identify and extract themes based on the social, technical, and organizational assessment criteria relevant to the research question. This analysis followed 7 key phases: familiarizing with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, linking themes to explanatory frameworks, and producing the final report.

In line with the approach of Braun and Clarke [40], we opted not to use interrater reliability as it aligns more closely with quantitative methods and standardized interpretation. Thematic analysis in a qualitative context prioritizes depth, subjectivity, and the unique insights each researcher brings to the data. Rather than using numerical reliability measures such as interrater reliability, reliability in this approach is often ensured through collaborative discussions that allow for consensus and a nuanced understanding of the themes. Accordingly, the first author, CJ, conducted the initial analysis and coding and NB reviewed the coding. Any cases of disagreement were discussed and mutually agreed upon in conjunction with a third author. Using the sociotechnical framework as our guide, we developed our initial codebook and grouped the criteria accordingly. This approach ensures a holistic evaluation of each tool, capturing the complex interdependencies between technical capabilities, social contexts, and organizational fit and readiness. By doing so, we moved beyond a narrow technical focus or methodological evaluation at the study level, ensuring that the social and organizational dimensions are fully integrated into the analysis. As a result, this work prioritizes the often-overlooked social and organizational dimensions that are critical for the successful implementation of AI technologies. Unlike frameworks that focus solely on clinical study quality, our analysis and synthesis specifically emphasize social and organizational factors such as user trust, support and training, interoperability, and integration.

However, we intentionally did not apply any hierarchy or prioritization within this foundational framework, as the purpose here is to treat all criteria as equally significant. Prioritization and potential gap identification will occur in the next phase (beyond the scope of this paper), where the Delphi process will engage an expert panel to further refine and prioritize these

XSL•FO
RenderX

criteria. The coding and analysis process was carried out from August to October 2024.

Results

Study Selection Flow and Characteristics of the Included Studies

Figure 1 presents the PRISMA flow diagram, illustrating the progression of study selection during the systematic review. It details the number of records identified, screened, included, and excluded, along with reasons for exclusion. After applying these criteria, 44 articles were selected for the qualitative synthesis.

Figure 1. Study selection flow diagram based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. AI: artificial intelligence.

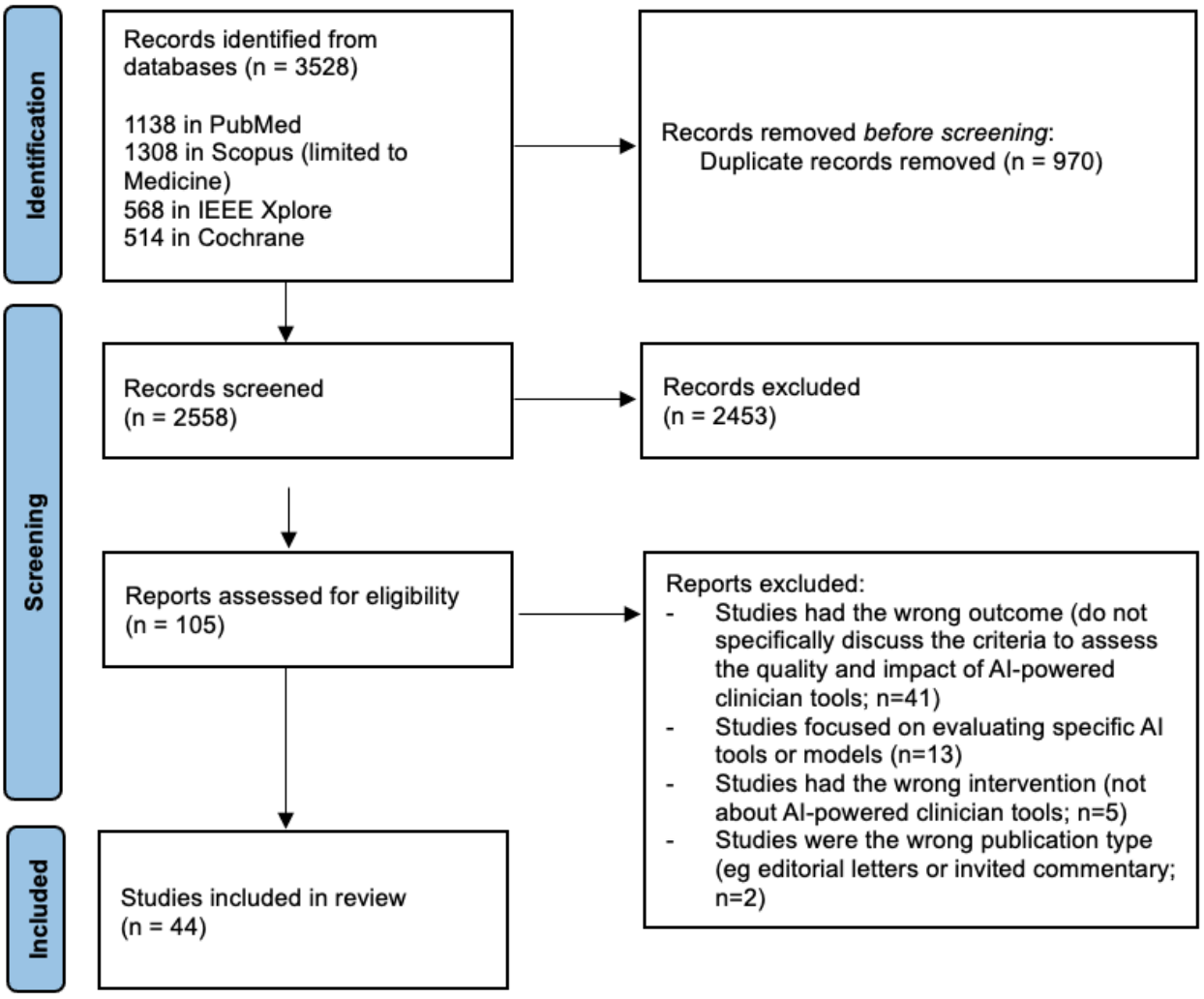

Table 1 outlines the characteristics of these studies, offering insights into their research methodologies, geographic distributions, and clinical focuses. This comprehensive overview highlights the diversity of approaches and topics addressed within the included studies.

XSL•FO
RenderX

Table 1. Characteristics of the included studies (N=44).

Study characteristics	Studies, n (%)	References
Country of authors		
Multiple	21 (48)	[41-61]
United States	5 (11)	[62-66]
France	3 (7)	[67-69]
Netherlands	3 (7)	[70-72]
Australia	2 (5)	[73,74]
Canada	2 (5)	[75,76]
Others		
China	1 (2)	[77]
Denmark	1 (2)	[78]
Germany	1 (2)	[79]
Greece	1 (2)	[80]
India	1 (2)	[81]
Saudi Arabia	1 (2)	[82]
Sweden	1 (2)	[83]
United Kingdom	1 (2)	[84]
Focus (some papers encompassed multiple areas of focus)		
No specific focus	9 (21)	[49,54,56,63,65,67,68,70,75]
Clinical focus		
Cardiovascular	3 (7)	[43,58,59]
Dermatology	2 (5)	[42,66]
ENT[a]	1 (2)	[82]
Medical imaging	12 (27)	[41,45,46,55,64,69,74,76,78,81,83,84]
Nuclear medicine	1 (2)	[48]
Radiation oncology	1 (2)	[47]
Technology focus		
ANN[b]	2 (5)	[48]
CDSSs[c]	3 (7)	[52,53,77]
DQMs[d]	1 (2)	[51]
LLMs[e]	3 (7)	[57,62,73]
ML[f]	2 (5)	[46,53]
Prediction models	2 (5)	[58,71]
Thematic focus		
EEs[g]	4 (9)	[44,61,72,79]
Ethics and equity	3 (7)	[60,63,66]
Explainability	1 (2)	[80]
Regulatory and trust	2 (5)	[50,70]
Paper type		
Original research		
Delphi process	3 (7)	[44,75,77]
Survey or questionnaire	2 (5)	[80,82]

XSL•FO
RenderX

Study characteristics	Studies, n (%)	References
Expert consensus	3 (7)	[42,57,69]
Expert perspective or comment	9 (21)	[51,52,54,55,62,64-66,76]
Guidelines or statements	6 (14)	[41,47,48,58,60,63]
Policy brief	1 (2)	[70]
Review	10 (23)	[45,46,50,53,56,59,74,81,83,84]
Scoping review	6 (14)	[43,68,71,73,78,79]
Systematic review	4 (9)	[49,61,67,72]
Publication year		
2019 (from January)	2 (5)	[46,52]
2020	2 (5)	[53,60]
2021	5 (11)	[50,55,56,70,74,77]
2022	10 (23)	[42,45,48,59,71,72,78,79,83]
2023	12 (27)	[43,47,51,58,61,63,65,67,69,73,75,76,80,81]
2024 (until mid-July)	13 (30)	[41,44,49,54,57,62,64,66,68,82,84]
Frameworks resulting from the included studies		
ABCDS[h]	1 (2)	[63]
CHEERS-AI[i]	1 (2)	[44]
CLEAR[j]	1 (2)	[42]
DQM[k]	1 (2)	[51]
DRIM France AI grid[l]	1 (2)	[69]
ECLAIR[m]	1 (2)	[55]
HEAL[n]	1 (2)	[66]
MAS-AI[o]	1 (2)	[78]
RADAR[p]	1 (2)	[41]
RELAINCE guidelines[q]	1 (2)	[48]
R-AI-DIOLOGY checklist[r]	1 (2)	[45]
TEHAI[s]	1 (2)	[56]
TREE[t]	1 (2)	[60]
Frameworks used in or referred to in the included studies		
CHEERS[u]	3 (7)	[44,61,72]
CLAIM[v]	4 (9)	[67,76,78,81]
CONSORT-AI[w]	6 (14)	[42,44,61,67,68,74]
DECIDE-AI[x]	2 (5)	[42,84]
FUTURE-AI[y]	1 (2)	[41]
GEP-HI[z]	1 (2)	[52]
HTA[aa]	6 (14)	[43,59,61,67,68,78]
MAST[ab]	2 (5)	[43,78]
PROBAST-AI[ac]	4 (9)	[42,44,67,74]
QAMAI[ad]	1 (2)	[57]

XSL•FO
RenderX

Study characteristics	Studies, n (%)	References
QMS[ae]	1 (2)	[65]
RQS[af]	1 (2)	[76]
SPIRIT-AI[ag]	6 (14)	[42,44,61,67,68,74]
STARD-AI[ah]	3 (7)	[42,67,74]
STARE-HI[ai]	1 (2)	[52]
TRIPOD-AI[aj]	7 (16)	[42,44,58,60,63,68,74]

[a]ENT: ear, nose, and throat.
[b]ANN: artificial neural network.
[c]CDSS: clinical decision support system.
[d]DQM: diagnostic quality model.
[e]LLM: large language model.
[f]ML: machine learning.
[g]EE: economic evaluation.
[h]ABCDS: Algorithm-Based Clinical Decision Support.
[i]CHEERS-AI: Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence.
[j]CLEAR: Derm Consensus Guidelines from the International Skin Imaging Collaboration Artificial Intelligence Working Group.
[k]DQM: Diagnostic Quality Model.
[l]DRIM France AI grid: French community grid for the evaluation of radiological artificial intelligence solutions.
[m]ECLAIR: Evaluating Commercial Artificial Intelligence Solutions in Radiology.
[n]HEAL: Health Equity Assessment of Machine Learning Performance.
[o]MAS-AI: Model for Assessing the Value of Artificial Intelligence in Medical Imaging.
[p]RADAR: Radiology Artificial Intelligence Deployment and Assessment Rubric.
[q]RELAINCE guidelines: Recommendations for Evaluation of Artificial Intelligence for Nuclear Medicine.
[r]R-AI-DIOLOGY checklist: a practical checklist for evaluation of artificial intelligence tools in clinical neuroradiology.
[s]TEHAI: Translational Evaluation of Healthcare Artificial Intelligence.
[t]TREE: transparency, reproducibility, ethics, and effectiveness.
[u]CHEERS: Consolidated Health Economic Evaluation Reporting Standards.
[v]CLAIM: Checklist for Artificial Intelligence in Medical Imaging.
[w]CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence.
[x]DECIDE-AI: Reporting Guideline for the Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence.
[y]FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in health care.
[z]GEP-HI: Good Evaluation Practice in Health Informatics.
[aa]HTA: Health Technology Assessment.
[ab]MAST: Model for Assessment of Telemedicine.
[ac]PROBAST-AI: Prediction Model Risk of Bias Assessment Tool–Artificial Intelligence.
[ad]QAMAI: Quality Analysis of Medical Artificial Intelligence.
[ae]QMS: Quality Management System.
[af]RQS: Radiomics Quality Score.
[ag]SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence.
[ah]STARD-AI: Standards for Reporting of Diagnostic Accuracy Studies–Artificial Intelligence.
[ai]STARE-HI: Statement on Reporting of Evaluation Studies in Health Informatics.
[aj]TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence.

Critical Appraisal

We evaluated the quality of the included studies using the CASP checklist [32]. This tool was selected due to the variety of methodologies used in the studies and the narrative approach of our synthesis, which differed from meta-analyses and other quantitative methods. The CASP is widely recognized as the most frequently used tool for appraising the quality of qualitative evidence in health research, with endorsement from the Cochrane Qualitative and Implementation Methods Group [85]. The studies included in our review used a range of methodologies (quantitative, qualitative, mixed methods, and systematic literature reviews), which meant that some questions on the checklist were not applicable to all study types. As per

XSL•FO
RenderX

the checklist's recommendations, we did not assign scores to the studies.

Following the critical appraisal of the 44 studies, several issues were identified. While all studies clearly stated their aims, presented well-defined findings, and provided valuable insights for health care stakeholders, 21 (N=44, 48%) studies lacked a dedicated methods section, making it difficult to assess the appropriateness and suitability of their approach. Similarly, the absence of clear methods in these studies hindered the evaluation of the research design and data collection techniques.

In addition, out of 44 studies, 25 (57%) studies did not detail their analysis methods, making it challenging to gauge the rigor and reliability of their approach. Furthermore, 28 (64%) studies lacked validation of their findings, while 8 (18%) offered only partial validation (eg, expert consensus), highlighting the need for empirical validation in real-world clinical applications to ensure the findings' robustness. The comprehensive quality assessment of the included studies can be found in Multimedia Appendix 1.

Studies were not excluded based on the results of the quality assessment, as this was unlikely to significantly impact the definition of the assessment criteria or the development of the aggregated framework. However, the quality assessment offered valuable insight into the overall robustness of the development processes behind the existing frameworks, helping to gauge the strength and reliability of the evidence presented [85]. An in-depth exploration of this topic can be found in the Discussion section, where the challenges associated with current initiatives and frameworks are examined.

Synthesized Assessment Criteria

We synthesized comparable measures from various papers, frameworks, and initiatives, ultimately identifying a set of unique criteria that reflected all relevant assessment methods referenced in the included studies. Notably, several criteria are closely interrelated and could fit into multiple categories; however, they were placed in the most appropriate category based on their significance and impact. For instance, while "user trust" and "model explainability" are inherently linked, because trust often correlates with the level of explainability provided by an AI system, we categorized trust under the cluster "acceptability, trust, and training," which focuses on user-centric aspects, whereas "explainability" was assigned to the cluster evaluating model performance metrics, given its technical focus. In addition, we intentionally included assessment criteria applicable to high-risk tools, enabling us to compile a more comprehensive list. We recognized that not all criteria would apply to lower-risk AI-powered health care tools, such as patient safety assessments, which are more relevant to high-risk tools that pose potential safety concerns. We are guided by National Institute for Health and Care Excellence's Evidence Standards Framework for Digital Health Technologies to assess and understand the risk levels of health care technologies [86].

Figure 2 provides a visual overview of the aggregated criteria, organized into clusters and subclusters, while Table 2 presents these criteria grouped into 7 primary clusters and their respective subcriteria, outlining their occurrences across the included studies, along with their definitions and corresponding references. A detailed exploration of each criteria cluster and its corresponding subcriteria is provided in the Discussion section.

XSL•FO
RenderX

Figure 2. Visual overview of the aggregated assessment criteria, organized into clusters and subcriteria. AI: artificial intelligence.

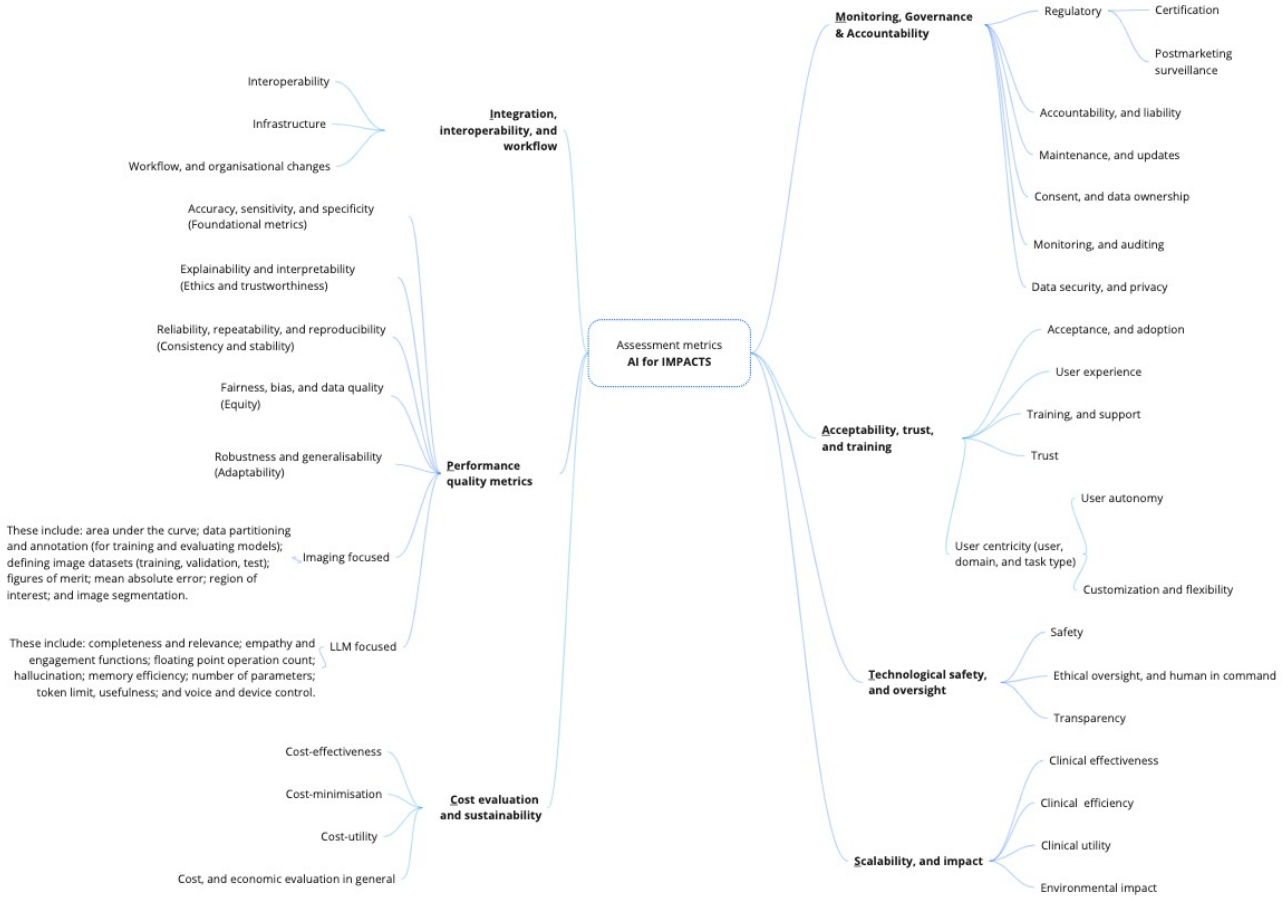

XSL•FO
RenderX

Table 2. Assessment criteria, their definitions, occurrence, and respective references (N=44).

Criteria	Definition	Studies, n (%)	Studies in which the criteria occurred
Integration			
Infrastructure	• The underlying technological, hardware, and software systems required to support the deployment and scalability of the AI[a] tool	15 (34)	[41,44,49,53-55,60,65,68,69,71,75,76,78,84]
Interoperability	• The AI tool's ability to seamlessly exchange and integrate data with different health care platforms and devices	19 (43)	[41,45,51,53-56,59,64,67-69,71,74-78,82]
Workflow and organizational changes	• The degree to which the AI tool impacts existing clinical workflows and health care operations, ensuring minimal disruption while enhancing efficiency, communication, and overall care delivery	22 (50)	[41,45,47,49,51-56,58,64,65,67-69,71,76,78,79,82,84]
Monitoring, governance, and accountability			
Accountability and liability	• The clear attribution of responsibility for errors or outcomes and the establishment of legal and ethical frameworks to address potential issues and ensure proper recourse	13 (30)	[49-51,54,55,63-65,70,75,78,81,84]
Consent and data ownership	• Evaluates the processes for obtaining informed consent from patients regarding the use of their data and ensuring clear policies on data ownership, privacy, and control	5 (11)	[49,54,68,78,81]
Maintenance and updates	• Evaluates the processes for ongoing support, including regular updates and bug fixes, to ensure the AI tool remains effective, secure, and aligned with evolving medical standards and practices	13 (30)	[44,45,49,50,52,53,55,59,65,69,71,76,77]
Monitoring and governance	• Evaluates the systems in place for overseeing the AI's performance, including regular assessments and audits to ensure ethical use and effectiveness	22 (50)	[44,48-56,59,60,64,65,68,69,71,74,76,81,83,84]
Regulatory compliance	• Evaluates adherence to established regulations throughout the AI tool's life cycle, including ongoing monitoring and reporting after deployment to ensure continued safety, efficacy, and adherence to legal requirements	23 (52)	[45,49-55,59,60,63-65,67-69,72,74,76,78,81,83,84]
Security and privacy	• Evaluates the measures implemented to protect sensitive patient data from unauthorized access and breaches while ensuring compliance with privacy regulations	26 (59)	[45,51-56,60,62,64-71,73-75,77,78,80-83]
Performance quality metrics			
Accuracy, sensitivity, and specificity (foundational metrics)	• Accuracy: the proportion of correct predictions (both true positives and true negatives) out of all predictions. It gives an overall measure of performance but may be misleading if the dataset is imbalanced (ie, when one class dominates). • Sensitivity (recall): the ability of the model to correctly identify true positives (ie, people with the condition). In health care, this often refers to how well the model detects cases like diseases. High sensitivity ensures that most cases of the disease are caught, reducing the chance of missing sick patients. • Specificity: the ability to correctly identify true negatives (ie, people without the condition). High specificity means the model avoids false positives, reducing unnecessary interventions for healthy people	26 (59)	[41,43,45-53,57,58,60,62,67,69,70,73,74,76,80-84]

XSL•FO
RenderX

Criteria	Definition	Studies, n (%)	Studies in which the criteria occurred
Explainability and interpretability (ethics and trustworthiness)	• Explainability: refers to the degree to which the model's predictions and decisions can be understood by humans. In health care, explainability is crucial because clinicians need to trust AI recommendations and understand why the AI made a particular decision. • Interpretability: closely related to explainability, it is about how easily a human can comprehend the internal workings of the model. For example, an interpretable model may allow clinicians to track how specific features (like patient age or laboratory results) influenced the AI's prediction	19 (43)	[45,47-49,51,52,55,60,62,67,68,71,73-75,78,80,81,83]
Fairness (equity)	• Fairness: ensures that the AI model does not systematically discriminate against any specific group of people (eg, based on race, gender, or socioeconomic status). Fairness in health care is key to avoid bias in diagnoses or treatments.	32 (73)	[42-47,49,51-56,58,60,62-69,71,73-75,77,78,81,83,84]
Reliability, repeatability, and reproducibility (consistency and stability)	• Reliability: refers to the consistency of the model over time. Can the AI be trusted to perform in the same way under similar conditions in the future? • Repeatability: the ability of the model to provide consistent results when the same input is given multiple times in the same environment. In health care, this ensures that if a patient is reevaluated using the same AI tool, it will give the same outcome. • Reproducibility: refers to how well the model performs when applied to different datasets or by different teams. This is critical in health care, where models trained on one population must still perform well when tested on different populations or data collected in different hospitals.	24 (55)	[42,43,45,48,50-56,58-60,64,67,68,71,74,76,77,81,83,84]
Robustness and generalizability (adaptability)	• Robustness: the model's ability to maintain performance despite slight variations or noise in the input data. In a health care setting, this might mean the model works well even with slightly lower-quality images or laboratory results from different equipment. • Generalizability: the ability of the model to perform well on new, unseen data that may differ from the training data. In health care, it is crucial that an AI model trained in one hospital or region can generalize to others.	23 (52)	[41-46,48,49,52,54-56,58,60,64,67,68,70,71,75,76,83,84]
Imaging-focused	• These may include area under the curve; data partitioning and annotation (for training and evaluating models); defining image datasets (training, validation, and testing); figures of merit; mean absolute error; region of interest and image segmentation	10 (23)	[42,45-48,55,69,76,81,83]
Large language model-focused	• These may include completeness and relevance; empathy and engagement functions; floating point operation count; hallucination; memory efficiency; number of parameters; token limit and usefulness; voice and device control	3 (7)	[57,62,73]
Acceptability, trust, and training			

XSL•FO
RenderX

Criteria	Definition	Studies, n (%)	Studies in which the criteria occurred
Acceptance and adoption	• Evaluates how well the AI tool is embraced by health care professionals and patients, including their willingness to integrate it into routine practice	18 (41)	[47,49,52-54,56,60,63,65,67-69,72,73,75,77,78,82]
Training and support	• Evaluates the effectiveness and availability of resources provided to users for learning and using the AI tool, ensuring they have the necessary guidance and assistance for successful implementation and operation	17 (39)	[45,49,52-55,64,65,67,69,71,76,78,80-82,84]
Trust	• Evaluates the degree to which health care professionals and patients believe in the reliability, accuracy, and ethical considerations of the AI tool, influencing their willingness to use it	11 (25)	[45,47,49,52,60,62,68,73,75,82,84]
Usability	• Evaluates how easily and effectively health care professionals and patients can interact with and use the AI tool, ensuring it enhances rather than hinders the user experience and clinical workflows	18 (41)	[47,49-53,55,56,62,63,68,71,73,75-77,82,84]
User centricity (user, domain, and task type)	• Evaluates how well the AI tool is designed to meet the specific needs, preferences, and contexts of its users, domain-specific requirements, and task types it is intended to support	19 (43)	[42,45,49,52,53,55,56,58,60,62,64,65,71,73,75-77]
Cost and economic evaluation			
Costs and economic evaluation in general	• Evaluates the financial implications of implementing the AI tool, ensuring it provides value without imposing excessive financial burdens on health care systems or patients	18 (41)	[44,49,51,55,59-61,64,67-69,72,73,75,76,79,82,84]
Cost-effectiveness analysis	• Compares the relative costs and outcomes of different interventions. The outcomes are typically measured in natural units like life years saved, cases prevented, or symptom-free days	12 (27)	[41,53,55,59-61,68,71-74,84]
Cost-minimization analysis	• Used when 2 or more interventions or treatments are assumed to produce identical outcomes or equivalent effectiveness. Given that the outcomes are considered the same, the focus is entirely on minimizing costs.	5 (11)	[41,61,72,78,82]
Cost-utility analysis	• Measures outcomes in terms of both quantity (life expectancy) and quality of life. It uses a metric called quality-adjusted life years or disability-adjusted life years to quantify health benefits	3 (7)	[41,61,72]
Technological safety and transparency			
Safety	• Evaluation of an AI tool's ability to avoid causing harm to patients by ensuring that it operates reliably, adheres to clinical standards, and mitigates potential risks	26 (59)	[42,47,49-56,60,62-68,71,73,75,78,80,81,83,84]
Transparency	• Refers to the extent to which an AI tool's processes, decision-making logic, and data sources are made understandable and accessible to stakeholders	27 (61)	[42,44,46,49,50,52-54,56,58-61,63-65,67,68,70,71,73-76,78,80,81]
Ethical oversight, human in command	• Assesses whether the AI tool is designed to support human decision-making, allowing clinicians to maintain control and override AI decisions when necessary, ensuring AI complements rather than replaces human judgment	14 (32)	[41,42,45,49,51,64,65,67,68,70,74,75,78,80]
Scalability and impact			

XSL•FO
RenderX

Criteria	Definition	Studies, n (%)	Studies in which the criteria occurred
Clinical effectiveness	• Assesses how well the AI tool works in real-world practice, including its ability to achieve desired clinical outcomes across diverse populations and settings	26 (59)	[41-44,47-49,51-54,60,63-65,67-69,71-73,75,77,78,83,84]
Clinical efficiency	• Focuses on the optimal use of resources (time, staff, and cost) to deliver care	8 (18)	[47,52,53,55,56,68,76,79]
Clinical utility	• Refers to the practical benefits of a treatment or intervention in improving patient care, such as guiding clinical decision-making or reducing risks	14 (32)	[41-43,47,48,50,52,56,58,60,62,69,77,82]
Environmental impact	• Evaluates how the development, deployment, and operation of AI tools affect environmental sustainability, such as energy consumption and carbon footprint	1 (2)	[75]

[a]AI: artificial intelligence.

With our focus on assessing the long-term real-world impact of AI technologies in health care, we named the framework AI for IMPACTS. The criteria were organized into seven key clusters, each corresponding to a letter in the acronym: (1) I — integration, interoperability, and workflow; (2) M — monitoring, governance, and accountability; (3) P — performance and quality metrics; (4) A — acceptability, trust, and training; (5) C — cost and economic evaluation; (6) T — technological safety and transparency; and (7) S — scalability and impact.

Discussion

Principal Results

Through our systematic review of the literature, which culminated in the inclusion of 44 relevant papers, we conducted a narrative synthesis guided by the sociotechnical framework. This synthesis identified and categorized the key technical, social, and organizational criteria critical for the practical and effective implementation of AI technologies in health care. The results are organized into 7 main clusters, further divided into 28 specific subcriteria, providing a structured framework to address the multifaceted considerations highlighted in the reviewed literature.

By synthesizing and aggregating the assessment criteria from all included studies, we developed the AI for IMPACTS framework. This framework goes beyond focusing solely on technical metrics or methodological guidance at the study level. It integrates the clinical context and real-world implementation factors to ensure AI tools are evaluated holistically. Most criteria in our proposed framework can be aligned with existing frameworks, but none covers all relevant categories without extensions. For successful AI implementation in health care, it is essential to integrate these tools within the broader organizational context. Frameworks should account for the complexities of the sociotechnical environment, recognizing the interplay between technical, social, and organizational dimensions. Our consolidated framework achieves this by synthesizing and expanding existing frameworks for AI assessment in health care. It uses a sociotechnical approach to consider all contextual factors, their interactions, and the long-term real-world impact of these technologies in clinical practice.

The sociotechnical theory, which emphasizes the dynamic interplay between social, organizational, and technical aspects, provides a holistic approach to evaluating novel technologies [33]. This is critical in health care, where the successful implementation of novel technologies requires a balance of these factors to optimize both technology adoption and clinical outcomes [36]. Each component of the AI for IMPACTS framework reflects this sociotechnical foundation, as described below.

- I: integration, interoperability, and workflow — sociotechnical theory stresses the need for alignment between technology and workflow. This criteria cluster ensures that AI tools integrate seamlessly within existing systems and workflows, minimizing disruptions and supporting health care professionals in their work.
- M: monitoring, governance, and accountability — governance structures are vital for ensuring AI applications adhere to clinical standards and ethical norms. The sociotechnical theory supports the need for oversight that considers not just technical capabilities but also social and organizational responsibilities, promoting accountability in decision-making.
- P: performance and quality metrics — effective AI assessment requires robust performance metrics that span technical and clinical outcomes. By applying sociotechnical principles, this criteria cluster ensures that quality standards are met in ways that resonate with both technical requirements and patient care priorities.
- A: acceptability, trust, and training — for AI to be widely adopted, it must be trusted and understood by users. The sociotechnical theory emphasizes the role of social factors such as trust and user training, which are essential for fostering acceptance among health care providers and patients.
- C: cost and economic evaluation — costs are a key concern in health care. The sociotechnical approach underscores the importance of evaluating not just technical implementation costs but also the economic implications

XSL•FO
RenderX

for patients and health care systems, ensuring that AI tools are financially sustainable and valuable.

- T: technological safety and transparency — safety and transparency are core to AI in health care, as they directly affect user trust and patient safety. The sociotechnical theory highlights that these technical attributes must be coupled with transparent communication and organizational processes that make AI's functioning understandable and dependable.
- S: scalability and impact — sociotechnical principles stress adaptability within complex systems. This criteria cluster considers how AI can be scaled effectively across diverse health care settings, evaluating both technical scalability and the social and organizational impact for expansion.

By leveraging the sociotechnical theory, the AI for IMPACTS framework ensures that each criterion is evaluated in a way that respects the complex interdependencies between technical capabilities, social context, and organizational readiness, providing a balanced and comprehensive approach to AI assessment in health care. We selected the acronym IMPACTS to underscore our emphasis on real-world outcomes over isolated, study-level evaluations. This highlights our commitment to assessing the broader, practical effects in health care settings.

Figure 3 depicts the 7 assessment clusters of the AI for IMPACTS framework. Each cluster contains multiple subcriteria, all of which are summarized in a comprehensive checklist presented in Table 3. The framework provides a systematic approach for evaluating AI's holistic role and potential in health care applications. The following subsections provide a detailed analysis of each criteria cluster and their respective subcriteria, offering a comprehensive breakdown of how each factor contributes to the overall assessment.

Figure 3. AI for IMPACTS: a comprehensive framework for evaluating the long-term real-world impacts of artificial intelligence (AI)–powered clinician tools.

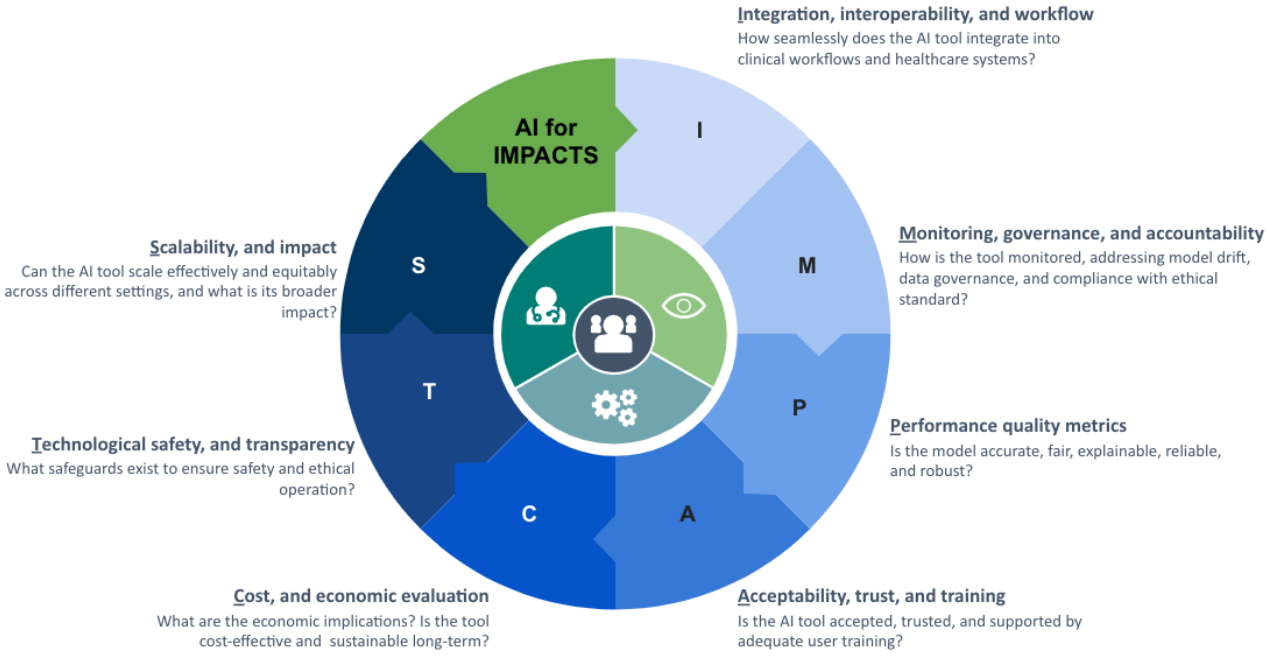

XSL•FO
RenderX

Table 3. The AI for IMPACTS framework assessment criteria for evaluating the long-term real-world impacts of artificial intelligence–powered clinician tools.

Criteria	Assessment
Integration	
Infrastructure	Does the deployment and scalability of the AI[a] tool require additional technological, hardware, or software infrastructure beyond what is already available in the current clinical setting?
Interoperability	Does the AI tool seamlessly integrate and exchange data with various health care platforms and devices, ensuring interoperability across different systems without requiring significant modifications?
Workflow and organizational changes	Does the AI tool integrate smoothly into existing clinical workflows and health care operations, minimizing disruption while enhancing efficiency, communication, and the overall delivery of care?
Monitoring, governance, and accountability	
Accountability and liability	Is there clear attribution of responsibility for errors or outcomes, supported by well-defined legal and ethical frameworks that ensure accountability and proper recourse in the event of any issues?
Consent and data ownership	Does the AI tool have clear and robust processes for obtaining informed consent from patients, including transparent policies on data ownership, privacy, and control, ensuring patients fully understand how their data will be used?
Maintenance and updates	Does the AI tool have established processes for ongoing support, including regular updates and bug fixes, to ensure it remains effective, secure, and compliant with evolving medical standards and practices?
Monitoring and governance	Does the AI tool have systems in place for ongoing oversight of its performance, including regular assessments and audits to ensure ethical use, effectiveness, and adherence to relevant standards?
Regulatory compliance	Does the AI tool demonstrate adherence to established regulations throughout its entire life cycle, with systems in place for ongoing monitoring and reporting postdeployment to ensure continued safety, efficacy, and compliance with legal requirements?
Security and privacy	Does the AI tool have robust measures in place to protect sensitive patient data from unauthorized access and breaches, while ensuring full compliance with relevant privacy regulations?
Performance quality metrics	
Foundational metrics	These are application-specific metrics to ensure each tool is assessed appropriately based on its function: • Diagnosis and prediction applications: use classification metrics (eg, accuracy, sensitivity, specificity, and area under the curve) for diagnosis tasks and regression metrics (eg, mean absolute error and root mean square error) for predicting continuous outcomes (classification, anomaly detection, and recommendation systems) • Image and pattern analysis: focus on segmentation accuracy and reinforcement learning's long-term performance optimization (eg, Dice coefficient, Jaccard index, and cumulative reward) • Text and language processing applications: evaluate the accuracy and quality of AI-extracted or generated text (eg, completeness and relevance, empathy and engagement, floating-point operation count, and hallucination)
Explainability (ethics and trustworthiness)	Is the AI tool able to clearly show how it reached a specific decision or prediction in a way that clinicians can understand?
Interpretability	Is it easy for clinicians to understand the relationship between the input data and the AI tool's outputs, without needing detailed technical explanations?
Fairness (equity)	Does the AI tool ensure fairness by avoiding systematic discrimination against any specific group, such as race, gender, or socioeconomic status, and promoting equitable outcomes in diagnoses and treatments?
Reliability, repeatability, and reproducibility (consistency and stability)	Does the AI tool demonstrate reliability, repeatability, and reproducibility by consistently delivering the same results over time, under similar conditions, and when applied to different data sets or used by different teams?
Robustness and generalizability (adaptability)	Does the AI tool demonstrate both robustness and generalizability by maintaining strong performance despite variations or noise in input data, and by performing well on new, unseen data from different hospitals or regions compared to its training data?
Acceptability, trust, and training	
Acceptance and adoption	Does the AI tool demonstrate strong acceptance by health care professionals and patients, including their willingness to adopt and integrate it into routine clinical practice?
Training and support	Does the AI tool provide comprehensive and readily available resources for users, ensuring they have the necessary guidance, training, and assistance to successfully implement and operate it in clinical practice?
Trust	Does the AI tool inspire trust among health care professionals and patients in terms of its reliability, accuracy, and ethical considerations, thereby positively influencing their willingness to use it?

XSL•FO
RenderX

Criteria	Assessment
Usability	Does the AI tool offer an intuitive and user-friendly interface that allows health care professionals and patients to interact with it easily and effectively, ensuring it enhances the user experience and integrates smoothly into clinical workflows?
User centricity (user, domain, and task type)	Does the AI tool effectively meet the specific needs, preferences, and contexts of its users, while addressing domain-specific requirements and supporting the relevant tasks for which it is intended?
Cost and economic evaluation	
Costs and economic evaluation	Does the AI tool provide financial value by enhancing care without imposing excessive costs on health care systems or patients, ensuring that its implementation is economically sustainable? This can be measured using one or more of the following methods*:* • Does the AI tool demonstrate cost-effectiveness by offering a favorable balance between its costs and the health outcomes it achieves, such as life years saved, cases prevented, or symptom-free days, when compared to alternative interventions? • Does the AI tool demonstrate cost-utility by providing measurable improvements in both life expectancy and quality of life, quantified through metrics such as quality-adjusted life years or disability-adjusted life years? • Does the AI tool demonstrate cost-minimization by achieving equivalent outcomes or effectiveness compared to alternative interventions, while focusing on minimizing overall costs?
Technological safety and transparency	
Safety	Does the tool reliably adhere to clinical standards, consistently mitigate potential risks, and demonstrate the ability to avoid causing harm to patients through reliable operation and risk management?
Transparency	Does the AI tool provider ensure transparency by making its processes, decision-making logic, and data sources understandable and accessible to all relevant stakeholders?
Ethical oversight, human in command	Does the AI tool incorporate ethical oversight by ensuring that it supports human decision-making, allowing clinicians to maintain control and override AI-generated decisions, when necessary, thereby complementing rather than replacing human judgment?
Scalability and impact	
Clinical effectiveness	Does the AI tool demonstrate clinical effectiveness by consistently achieving the desired clinical outcomes in real-world practice, across diverse patient populations and health care settings?
Clinical efficiency	Does the AI tool demonstrate clinical efficiency by optimizing the use of resources, including time, staff, and costs, to effectively deliver care without compromising quality?
Clinical utility	Does the AI tool demonstrate clinical utility by offering practical benefits that improve patient care, such as guiding clinical decision-making or reducing risks during treatment?
Environmental impact	Does the AI tool minimize its environmental impact by considering sustainability in its development, deployment, and operation, including factors such as energy consumption and carbon footprint?

[a]AI: artificial intelligence.

Integration

This criteria cluster focuses on evaluating how effectively the AI tool integrates into existing clinical workflows and health care systems.

Infrastructure plays a crucial role in the successful implementation of AI tools in health care settings. Adequate computational power, specialized hardware, and robust IT infrastructure are often necessary to support the processing of large datasets and the operational demands of AI technologies [49,71]. This may include advanced components such as graphics processing units, which are not always standard in health care systems [55]. In addition, integrating these tools might require significant investment in new hardware or upgrades [60,69]. For cloud-based AI solutions, attention must be paid to network security and performance [55]. Ensuring infrastructure compatibility is essential for the smooth deployment and optimal functionality of AI in health care [41,54].

Interoperability ensures seamless integration with existing systems, such as electronic health records and imaging software. It allows AI tools to operate within current workflows without disrupting established clinical processes, enhancing data exchange across platforms [41,68]. It also ensures that AI tools adhere to industry standards, facilitating communication between different health care technologies and minimizing issues such as data misinterpretation or workflow inefficiencies [71]. Proper integration can reduce the resource burden on health care facilities and improve the overall usability and effectiveness of AI systems in diverse clinical settings [64].

Understanding the impact on *clinical workflows and organizational structures* is essential. AI tools must be seamlessly integrated into workflows to avoid disrupting clinical processes [49,82]. Evaluating how AI affects the redistribution of tasks among health care professionals and identifying necessary organizational changes are essential [64,67]. Poor integration or failure to align with clinical routines can

XSL•FO
RenderX

negatively impact efficiency, increase cognitive burdens, and require significant resources to adapt systems [45,58].

Monitoring, Governance, and Accountability

This criteria cluster focuses on evaluating how effectively the AI tool is monitored throughout its life cycle, addressing critical aspects such as model drift, data governance, and adherence to ethical standards.

Clarity on *accountability and liability* is essential when assessing AI tools in health care due to the potential risks involved in their implementation [49,81]. AI systems can make errors or offer recommendations that may not be followed by clinicians, raising complex questions about who is responsible when mistakes occur [54,78]. The lack of clear guidelines on whether liability lies with the developer, the health care institution, or the clinician using the tool poses significant legal and ethical concerns [55,84]. Proper assessment frameworks must ensure that accountability is well-defined, including clear roles for all stakeholders involved (eg, clinicians, developers, and institutions) particularly in cases of adverse events or errors [50,64,70].

Data security, privacy, informed consent, and data ownership are vital criteria for assessing AI tools in health care. These tools often require large amounts of sensitive patient data, which must be protected from unauthorized access, breaches, or misuse [75,83]. Ensuring compliance with relevant regulations, such as General Data Protection Regulation or Health Insurance Portability and Accountability Act, is essential to safeguard patient privacy [55,60,71]. In addition, clear processes for obtaining informed consent are critical, ensuring that patients understand how their data will be used [68,81]. Proper data ownership policies must also be in place, ensuring transparency around who controls the data and how it can be accessed or shared [49,78]. These measures are crucial for building trust and ensuring ethical AI deployment in health care settings [54,68].

Regulatory compliance and certification are essential but insufficient assessment criteria for AI tools in health care [21]. Although regulatory bodies like the FDA in the United States and CE marking in the EU set minimum safety and efficacy standards, there are significant gaps between legal certification and real-world clinical validation, workflow integration, and ongoing use [21,45]. For instance, FDA clearance does not always assure users that an AI tool will meet their expectations for effective performance in all clinical settings, leading to skepticism among health care professionals [21,84]. Similarly, in the EU, AI tools with CE marking are often assumed to be clinically validated, but many lack sufficient validation for real-world clinical use, such as in dementia diagnosis via magnetic resonance imaging [45,83]. These gaps highlight the need for stronger regulatory frameworks and postmarket surveillance to ensure AI tools are not only certified but also thoroughly validated and integrated into health care workflows for effective and safe use [21,72,76].

Monitoring and governance mechanisms, including feedback loops, are critical for ensuring the continued safety, effectiveness, performance, and reliability of AI tools in health care [71]. It is essential that the responsibility for monitoring these tools is shared between the developer, regulator, and the health care organization deploying the tool [84]. Developers are responsible for ongoing performance evaluations, including regular updates to address issues such as data drift or algorithmic failure [48,74]. Regulators must ensure compliance with postmarket surveillance requirements and set clear guidelines for monitoring practices [60,84]. Health care organizations must implement local oversight systems, ensuring that the AI tool continues to meet clinical needs without causing disruption or harm [49,64,65,71,83]. By assigning responsibility to all 3 entities, health care systems can ensure comprehensive, multi-layered oversight that addresses technical, clinical, and regulatory concerns [84].

The *maintenance and updating* of AI tools are critical to ensuring their continued effectiveness and safety in health care [71]. Regular updates, including adjustments to algorithms and reference datasets, are essential to avoid performance degradation and ensure accurate results [53,71]. Without proper maintenance, different software versions could introduce biases or inconsistencies, which might affect clinical outcomes [45,76]. Establishing clear protocols for updates, including version control and procedures for managing software changes, ensures that AI tools remain reliable and aligned with current medical standards, safeguarding patient care [50].

Performance Quality Metrics

This criteria cluster focuses on evaluating the performance and quality of the AI tool by assessing key metrics such as foundational performance metrics, fairness, explainability, reliability, and robustness.

Foundational performance metrics play a crucial role in assessing the effectiveness of AI tools. The systematic review revealed that 59% (26/44) of studies primarily focused on accuracy, sensitivity, and specificity as key metrics. However, it is essential to consider application-specific metrics when evaluating AI performance, as different AI tools require tailored measures depending on their intended use. For example, diagnosis and prediction tools encompass applications like classification (eg, disease diagnosis), regression (eg, predicting disease progression), anomaly detection, and recommendation systems. These tools can be assessed through metrics such as accuracy, sensitivity, specificity, and the area under the curve for classification tasks [41,43,69] and mean absolute error and root mean square error for regression tasks [46,81]. Image and pattern analysis covers tasks such as image segmentation and reinforcement learning, using metrics like the Dice coefficient and Jaccard index for segmentation accuracy [87,88], and cumulative reward for evaluating reinforcement learning performance [89]. On the other hand, text and language processing applications, such as natural language processing and large language models, are assessed using metrics like relevance, engagement, empathy, token limits, hallucination rates, memory efficiency, and floating-point operation count [57,62,73]. These metrics ensure the AI tool is properly evaluated based on its intended use and technology type.

Explainability and interpretability are essential for ensuring the ethical and trustworthy use of AI tools in health care. These

XSL•FO
RenderX

criteria allow health care professionals to understand how AI models arrive at their conclusions, fostering trust in their recommendations [47,83]. Explainability helps to demystify the AI's decision-making process, making it transparent and accessible to users [67,68]. This, in turn, improves adoption, as clinicians are more likely to trust and rely on AI tools that are interpretable [49,60]. Ultimately, clear explainability supports ethical deployment, reducing risks associated with "black box" systems [73,78].

Fairness or equity ensures that AI models provide unbiased, consistent performance across diverse demographic groups, including those defined by race, gender, age, or socioeconomic status [62,66,83]. This criterion addresses the risk of bias in training data, including sample size and representativeness, which can lead to unequal treatment or outcomes for underrepresented populations [42,43,71]. By focusing on fairness, AI tools can avoid perpetuating disparities and contribute to more equitable health care delivery for all patients [55,63,71].

Reliability, repeatability, and reproducibility ensure that the AI tool can produce consistent outputs when presented with similar inputs, is repeatable under identical conditions, and is reproducible in diverse environments, including different institutions or patient populations [52,55,56,64]. Maintaining consistency and stability is essential for the tool's trustworthiness and its broader applicability in real-world health care scenarios [54,84].

Robustness and generalizability are essential criteria for assessing the adaptability of AI tools in health care [42,83]. Robustness ensures the tool can maintain high performance even when exposed to slight variations in input data or operational environments [70,83]. Generalizability, on the other hand, evaluates whether the AI tool can effectively perform across different populations, clinical settings, or geographic regions beyond the environment in which it was trained [48,83]. These criteria ensure that AI tools remain reliable and effective when scaled or applied to diverse health care contexts [49,54].

Acceptability, Trust, and Training

This criteria cluster evaluates user-centric aspects of the AI tool, focusing on its acceptance, trustworthiness, and the adequacy of user training and support.

User *acceptance and adoption* are crucial for the successful implementation and translation of AI-powered health tools in real-life settings [56,65,82]. Key challenges include fostering trust and confidence among health care professionals, ensuring ease of use, and integrating these tools seamlessly into clinical workflows [77]. User acceptance depends significantly on the perceived benefits, transparency, and safety of the AI systems [47,49,52]. Moreover, ethical concerns, the potential for bias, and the need for comprehensive testing also impact adoption [67]. Clinicians are more likely to embrace these tools when they complement human expertise and are introduced with adequate training and support, ensuring they enhance patient outcomes without compromising safety [68]. User acceptance and adoption of technology are typically measured through surveys (eg, Technology Acceptance Model and Unified Theory of Acceptance and Use of Technology) assessing factors like perceived usefulness and ease of use, as well as use metrics such as adoption rates, frequency, and retention.

Trust is built through factors such as validation, transparency, safety, privacy, and interpretability of the AI tool [62]. Both health care professionals and patients must trust that the AI tool is reliable, safe, and effective in clinical practice [60,68]. Validating AI performance using local data is essential to build clinician confidence, while demonstrating that the tool adheres to rigorous standards helps address concerns about its real-world application [49,84]. Trust also influences adoption, making it vital for the successful implementation of AI tools in health care [82]. User trust in technology is commonly assessed through surveys and trust scales, such as the Technology Trust Index, which evaluate key dimensions like reliability, competence, transparency, and security. Behavioral metrics, including use patterns and reliance during critical tasks, offer additional insights into how trust manifests in practice.

User centricity emphasizes the need for a clear understanding of the intended users, domain, and specific tasks the AI tool is designed to support [42,58,71]. AI tools must be tailored to meet the unique requirements of their end users, whether clinicians, nurses, or patients, and address the particular medical conditions they aim to diagnose, monitor, or treat [42,56]. Clarity in defining the tool's intended use, the health care domain it serves, and the tasks it performs ensures that it delivers meaningful value in its practical application [64,73].

Usability ensures that the tool is user-friendly and intuitive for both health care professionals and patients [49,53]. An AI tool's ease of use and minimal training requirements are essential for successful adoption [47,53]. Usability also impacts user satisfaction, influencing acceptance and trust in the system [52,77]. Proper design should minimize cognitive load, provide relevant information in context, and allow customization by users [71]. Evaluating usability ensures that AI tools can be effectively deployed in real-world clinical environments, enhancing rather than hindering care delivery [55,56].

Adequate *training* ensures that clinicians and other end users can effectively use AI tools, minimizing user error and maximizing the tool's potential to improve patient outcomes [55,65,71]. Training programs should cover how to interact with the AI interface, interpret its outputs, and understand the tool's limitations [49,54]. Continuous education is also crucial, and end users should not only be trained on interpreting the algorithm's output but also be made aware of the factors that can affect its performance [64]. Moreover, accessible and responsive technical *support* is necessary to address user concerns, provide ongoing assistance, and maintain confidence in the AI tool's reliability and safety over time [52]. Without proper *training and support*, the integration of AI tools into clinical practice may face significant barriers, limiting their overall effectiveness [45,65].

Cost and Economic Evaluation

This criteria cluster evaluates the economic implications of the AI tool to determine its financial viability and long-term sustainability.

XSL•FO
RenderX

Economic evaluation and cost considerations are crucial in assessing AI tools in health care. AI interventions must demonstrate not only clinical value but also health economic impact to ensure their long-term sustainability [76,84]. This includes evaluating both direct costs, such as acquisition, maintenance, and implementation, as well as indirect costs like staff training or workflow disruptions [49,61]. Transparent and comprehensive economic evaluations help health care organizations determine the financial viability of AI tools, guiding decision-making on investments, reimbursement, and long-term sustainability [59,64,67,79]. Incomplete or unclear cost assessments can hinder AI adoption and create financial risks [72,79].

The choice of an economic evaluation method for an AI tool in health care depends on its intended use and desired outcomes. *Cost-effectiveness* analysis is useful when comparing costs with health outcomes like life years saved [41,59,68,71]. *Cost-utility* analysis is ideal when focusing on both life expectancy and quality of life improvements, measured in quality-adjusted life years or disability-adjusted life years [41,61,72]. *Cost-minimization* analysis is appropriate when the AI tool achieves similar outcomes as alternatives but aims to reduce costs [61,72,78]. The method chosen should align with the tool's specific goals and intended health care impact.

Technological Safety and Transparency

This criteria cluster focuses on evaluating the technological safety and transparency of the AI tool by assessing the safeguards in place to ensure safe and ethical operation.

Safety ensures that AI systems operate reliably and securely in clinical environments beyond laboratory settings and clinical trials [73,78]. This includes compliance with safety regulations, minimizing the risks of harmful outcomes, and maintaining high standards for long-term safety and patient protection [53,67,68]. Safety also encompasses the reliability of the AI model after its implementation, ensuring it consistently avoids errors and unintended consequences [47,65,78]. Ongoing monitoring, risk management, and thorough clinical validation are necessary to ensure that AI tools remain safe and effective in diverse health care settings and the long-term safety of constant updates [49,67,68,83].

Transparency is a critical assessment criterion for AI tools in health care, ensuring clarity in data processing, coding standards, and the overall functioning of AI systems [71,78,80]. Transparent models allow health care professionals to understand how decisions are made, promoting trust and enabling accurate assessments of the AI's performance [63,67,68,73]. Clear documentation and disclosure of data processing methods, coding protocols, and the AI's decision-making processes ensures accountability and reproducibility [42,54,64]. A recent review of 692 FDA-approved AI enabled medical devices highlighted major gaps in transparency and safety reporting [90]. Key data such as ethnicity (reported in only 3.6% of approvals), socioeconomic information (absent in 99.1%), and study participants' age (missing in 81.6%) were often underreported [90]. In addition, only 46.1% of devices provided detailed performance results and only 1.9% were linked to scientific publications on safety and efficacy [90]. These findings underscore the urgent need for improved transparency and more comprehensive safety reporting to reduce algorithmic bias and ensure equitable health care outcomes.

Ethical oversight and human in command ensure human control and responsibility in the AI decision-making processes [64,70]. This criterion emphasizes that humans must retain ultimate authority over AI-generated decisions, particularly in critical health care scenarios [68,70]. Human in command ensures that clinicians can review, intervene, or override AI decisions, maintaining ethical standards and safeguarding patient outcomes [42,70]. This oversight protects against overreliance on automated systems and ensures that AI tools support, rather than replace, human judgment in clinical practice [45,68,80].

Scalability and Impact

This criteria cluster focuses on evaluating scalability and impact by determining the AI tool's clinical utility and effectiveness and examining its broader impact.

Clinical effectiveness focuses on the tool's ability to positively impact patient outcomes [68,71,78]. This involves evaluating whether the AI tool contributes to better therapeutic results or patient-reported outcomes [43,71]. The assessment examines how well the AI tool integrates into real-world clinical settings and measures its tangible benefits in terms of patient health and health care quality [49,69]. Clinical effectiveness ensures that AI tools do more than function technically; they must provide meaningful improvements in patient care [47,51].

Clinical utility focuses on how effectively the tool supports clinical tasks and decision-making, including its ability to assist with diagnoses, treatment recommendations, and overall health care delivery [51,82]. Ensuring clinical utility means the AI tool must provide tangible benefits that align with clinical needs and enhance health care practices [43,79]. *Clinical efficiency* focuses on the tool's ability to optimize resource use while maintaining or improving care quality [68]. This includes evaluating how well it improves productivity, reduces time spent on routine tasks, and streamlines workflows for health care professionals [47,53,55].

Environmental impact is an important, yet often overlooked, criterion for assessing AI tools in health care; only 1 out of 44 studies addressed this criterion. The energy consumption and resource use associated with developing, deploying, and maintaining AI systems, such as data centers, computational power, and device infrastructures, can lead to significant environmental harm, including e-waste and greenhouse gas emissions [75]. Implementing eco-responsible practices, such as energy-efficient computing and sustainable data storage, is essential to minimizing the ecologic footprint of AI tools [75].

Practical Implications and Persisting Challenges

The wide array of frameworks and initiatives focused on AI assessment in health care shown in this systematic review highlights the significant lack of standardization in this field, creating additional challenges for stakeholders [43,70,71]. Faced with a growing number of assessment tools, they often struggle to determine which approach is most appropriate or how to

XSL•FO
RenderX

apply it effectively [63]. This diversity in assessment methods can lead to confusion and hinder comparability [43,68,69,79]. Variations in data collection and evaluation methods, ranging from self-reported to objective measures, and from qualitative to quantitative assessments, only add to the complexity, further complicating the establishment of clear, universal guidelines for AI evaluation in health care [62].

Most frameworks included in this analysis were driven by the recognition that many existing methods for assessing AI tools in health care were not specifically tailored to AI-based medical devices or health care applications [62,65,67]. Traditional technology assessments often lack a critical focus on the unique, dynamic challenges and opportunities AI presents [57]. This underscores the need for health care–specific frameworks that account for the evolving nature and complexities of AI systems in clinical environments [60]. Moreover, existing frameworks tend to prioritize technical metrics such as algorithm accuracy, precision, and validation [62,73]. While these factors are undeniably important, this narrow focus often overlooks broader considerations, including clinical relevance, practical application, and long-term impact on patient outcomes [41,56,82]. Consequently, these frameworks can fall short in delivering a holistic evaluation of AI tools, which is essential for ensuring their safe, effective, and seamless integration into real-world health care settings [47,83].

This study builds upon and advances the ongoing discussion on AI assessment in health care, aiming to address the recognized gaps by developing the AI for IMPACTS framework. This proposed framework integrates technical, social, and organizational dimensions, ensuring that the adaptive nature of AI and the complexity of the health care ecosystem are fully considered. By encompassing these critical aspects, the framework provides a more comprehensive and nuanced approach to evaluating AI tools, helping shape the field and offering a robust method for assessing AI's real-world impact in health care settings.

However, numerous challenges still remain. These challenges extend beyond just setting the assessment criteria, to include practical difficulties in implementing, validating, and standardizing these criteria across diverse health care environments. A key challenge in assessing AI tools in health care is the variation across different contexts and settings [64,71]. Most available evidence focuses on high-income countries, limiting the generalizability of findings to diverse health care environments, particularly in low- and middle-income countries [43,49]. Recent studies underscore the importance of collaborative efforts and context-sensitive solutions to effectively address the unique health care challenges faced in these regions [91]. Another challenge is the need for a multidisciplinary team of assessors. Effective evaluation requires collaboration among professionals from various fields, such as medicine, IT, and social sciences to ensure a comprehensive assessment [77,83]. This diversity of expertise is necessary to address the complexities of AI, from technical and ethical considerations to clinical relevance and real-world impact [55,65,84].

It is crucial to emphasize the importance of adequate training in assessment methods [69,74]. Many assessors may lack the specific expertise required to thoroughly evaluate AI-based tools [46]. Proper training in the complexities of AI technology and appropriate evaluation techniques is essential for conducting accurate and meaningful assessments [55]. Without this, the assessment process may be compromised, potentially leading to inaccurate or incomplete evaluations of an AI tool's safety and effectiveness, which could undermine its implementation in health care settings [68]. Furthermore, the rapid pace of AI development, with AI-based medical devices having shorter product life cycles compared to traditional medical devices, underscores the need for more adaptive and fast-tracked health technology assessment processes [49,68]. Conventional health technology assessments are often too time-consuming, taking about a year to complete, which is incompatible with the fast-evolving nature of AI technologies [59]. Balancing the need for robust evidence with the dynamic nature of AI development is essential to ensure timely, informed decision-making and avoid delays in implementation and potential reimbursement [59,68].

Limitations and Future Research

This study enhances the understanding of various criteria for assessing the quality and impact of AI tools in health care, but several limitations must be acknowledged. Relevant studies may have been missed due to language restrictions or limited database searches, and the exclusion of gray literature may have omitted valuable insights. In addition, no follow-up was conducted with the study authors to validate the findings, and manual reference searches were avoided to minimize citation bias. As a result, some relevant frameworks or assessment criteria may not have been captured in this review. Future research could expand to include studies in other languages, offering a more comprehensive understanding of potential interregional or intercultural differences in the assessment of AI tools in health care.

The critical appraisal of the frameworks included in this review highlighted that many papers discussing AI tool assessment in health care lacked rigorous validation, with some omitting the methods section entirely. To address this gap, we propose rigorously validating the AI for IMPACTS framework proposed in this work through a Delphi process. The Delphi method was selected as a means to validate the framework as it is specifically designed to achieve reliable expert consensus, particularly in addressing complex issues [92,93]. This method is widely recognized across various fields of medicine, especially for developing best practice guidance and clinical guidelines, where expert agreement is critical [94,95]. This approach will involve key stakeholders to critically apply, reflect on, and refine the framework, ensuring it is relevant, comprehensive, and user-friendly. The goal is to cocreate practical, accessible tools with industry experts that can support the effective evaluation of AI tools in real-world health care settings.

It is also important to highlight that new frameworks were published after the cutoff date of this systematic review, including the Organizational Perspective Checklist for Artificial Intelligence Adoption [96], Stanford's framework for evaluating

XSL•FO
RenderX

Fair, Useful, and Reliable Artificial Intelligence Models in Health Care Systems [97], and the Transparent Reporting of Ethics for Generative Artificial Intelligence checklist [98]. While an initial review shows that their assessment dimensions align with this work, a deeper integration will be undertaken before the validation study. This will ensure that the foundation for the Delphi process is as comprehensive and up-to-date as possible.

Conclusions

AI has the potential to transform health care by improving clinical outcomes and operational efficiency. However, its adoption has progressed more slowly than anticipated, partly due to the absence of robust and comprehensive evaluation frameworks. Existing frameworks often focus too narrowly on technical metrics, such as accuracy and validation, neglecting real-world factors like clinical impact, workflow integration, and economic viability. Furthermore, the variety of frameworks and initiatives focused on AI assessment in health care, as highlighted in this systematic review, underscores a significant lack of standardization in the field, creating additional challenges for stakeholders and making it difficult to compare and implement AI tools effectively.

This study builds on and advances the ongoing discussion surrounding AI assessment in health care by developing the AI for IMPACTS framework. It aims to address key gaps identified in existing evaluation approaches, offering a comprehensive model that incorporates technical, social, and organizational dimensions. It is organized around 7 key criteria clusters: I—integration, interoperability, and workflow; M—monitoring, governance, and accountability; P—performance and quality metrics; A—acceptability, trust, and training; C—cost and economic evaluation; T—technological safety and transparency; S—scalability and impact.

While the framework provides a more holistic approach, significant challenges persist. The diverse contexts and settings in health care make it difficult to apply a one-size-fits-all framework. Multidisciplinary teams are necessary to evaluate AI tools thoroughly, as expertise from fields such as medicine, IT, and social sciences is required to address the complexities of AI. In addition, many assessors lack the specific training needed to evaluate these tools accurately. The rapid pace of AI development further complicates the assessment process, as conventional evaluation methods are often too slow to keep up with AI's short product life cycles. To ensure successful AI integration in health care, adaptive and fast-tracked assessment processes are essential, allowing for timely decision-making and implementation while maintaining the necessary rigor.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

CJ is an editorial board member of JMIR Human Factors at the time of this publication.

Multimedia Appendix 1
Critical Appraisal Skills Program appraisal of the included studies.
[XLSX File (Microsoft Excel File), 70 KB - jmir_v27i1e67485_app1.xlsx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 33 KB - jmir_v27i1e67485_app2.docx]

References

1. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan 7;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
2. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
3. Pressman SM, Borna S, Gomez-Cabello CA, Haider SA, Haider CR, Forte AJ. Clinical and surgical applications of large language models: a systematic review. J Clin Med 2024 May 22;13(11):3041 [FREE Full text] [doi: 10.3390/jcm13113041] [Medline: 38892752]
4. Xu C, Solomon SA, Gao W. Artificial intelligence-powered electronic skin. Nat Mach Intell 2023 Dec 18;5(12):1344-1355 [FREE Full text] [doi: 10.1038/s42256-023-00760-z] [Medline: 38370145]
5. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health 2021 Mar;3(3):e195-e203 [FREE Full text] [doi: 10.1016/S2589-7500(20)30292-2] [Medline: 33478929]

XSL•FO
RenderX

6. Zhang J, Whebell S, Gallifant J, Budhdeo S, Mattie H, Lertvittayakumjorn P, et al. An interactive dashboard to track themes, development maturity, and global equity in clinical artificial intelligence research. Lancet Digit Health 2022 Apr;4(4):e212-e213 [FREE Full text] [doi: 10.1016/S2589-7500(22)00032-2] [Medline: 35337638]
7. Goh S, Goh RS, Chong B, Ng QX, Koh GC, Ngiam KY, et al. Challenges in implementing artificial intelligence in breast cancer screening programs: a systematic review and framework for safe adoption. J Med Internet Res (Forthcoming) 2022 [FREE Full text] [doi: 10.2196/62941]
8. Hassan M, Kushniruk A, Borycki E. Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. JMIR Hum Factors 2024 Aug 29;11:e48633 [FREE Full text] [doi: 10.2196/48633] [Medline: 39207831]
9. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021 Apr;27(4):582-584. [doi: 10.1038/s41591-021-01312-x] [Medline: 33820998]
10. European Society of Radiology (ESR). Value-based radiology: what is the ESR doing, and what should we do in the future? Insights Imaging 2021 Jul 27;12(1):108-152 [FREE Full text] [doi: 10.1186/s13244-021-01056-9] [Medline: 34318401]
11. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022 Jul 01;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
12. Artificial intelligence in health care: benefits and challenges of technologies to augment patient care. U.S. Government Accountability Office. URL: https://www.gao.gov/products/gao-21-7sp [accessed 2024-04-29]
13. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ 2020 Sep 09;370:m3210 [FREE Full text] [doi: 10.1136/bmj.m3210] [Medline: 32907797]
14. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health 2020 Oct 09;2(10):e537-e548 [FREE Full text] [doi: 10.1016/S2589-7500(20)30218-1] [Medline: 33328048]
15. Nsoesie EO. Evaluating artificial intelligence applications in clinical settings. JAMA Netw Open 2018 Sep 07;1(5):e182658 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2658] [Medline: 30646173]
16. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015 Jan 06;13(1):1 [FREE Full text] [doi: 10.1186/s12916-014-0241-z] [Medline: 25563062]
17. Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the joint European and north American multisociety statement. Can Assoc Radiol J 2019 Nov 29;70(4):329-334. [doi: 10.1016/j.carj.2019.08.010] [Medline: 31585825]
18. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the editorial board. Radiology 2020 Mar;294(3):487-489. [doi: 10.1148/radiol.2019192515] [Medline: 31891322]
19. van Leeuwen KG, Schalekamp S, Rutten MJ, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 2021 Jun 15;31(6):3797-3804 [FREE Full text] [doi: 10.1007/s00330-021-07892-z] [Medline: 33856519]
20. Stettinger G, Weissensteiner P, Khastgir S. Trustworthiness assurance assessment for high-risk AI-based systems. IEEE Access 2024;12:22718-22745. [doi: 10.1109/access.2024.3364387]
21. Chouffani El Fassi S, Abdullah A, Fang Y, Natarajan S, Masroor AB, Kayali N, et al. Not all AI health tools with regulatory authorization are clinically validated. Nat Med 2024 Oct 26;30(10):2718-2720. [doi: 10.1038/s41591-024-03203-3] [Medline: 39187696]
22. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, STARD-AI Steering Committee. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021 Jun 28;11(6):e047709 [FREE Full text] [doi: 10.1136/bmjopen-2020-047709] [Medline: 34183345]
23. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020 Mar 01;2(2):e200029 [FREE Full text] [doi: 10.1148/ryai.2020200029] [Medline: 33937821]
24. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021 Jul 09;11(7):e048008. [doi: 10.1136/bmjopen-2020-048008] [Medline: 34244270]
25. Ethics and governance of artificial intelligence for health. World Health Organization. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2024-10-02]
26. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023 Mar 23;10:e45143 [FREE Full text] [doi: 10.2196/45143] [Medline: 36843321]

XSL•FO
RenderX

27. Jacob C, Lindeque J, Müller R, Klein A, Metcalfe T, Connolly SL, et al. A sociotechnical framework to assess patient-facing eHealth tools: results of a modified Delphi process. NPJ Digit Med 2023 Dec 15;6(1):232 [FREE Full text] [doi: 10.1038/s41746-023-00982-w] [Medline: 38102323]
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
29. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd edition. Hoboken, NJ: Wiley-Blackwell; 2019.
30. Jacob C. AI-powered clinician tools assessment criteria: protocol of a systematic review of the literature. Research Registry. URL: https://www.researchregistry.com/browse-the-registry#registryofsystematicreviewsmeta-analyses/registryofsystematicreviewsmeta-analysesdetails/669754107bdc5b002704adbe/ [accessed 2024-04-29]
31. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
32. Critical appraisal skills programme checklists. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2024-04-29]
33. Leonardi PM. Methodological guidelines for the study of materiality and affordances. In: Mir R, Jain S, editors. The Routledge Companion to Qualitative Research in Organization Studies. New York, NY: Routledge; 2017:279-290.
34. Ammenwerth E. Technology acceptance models in health informatics: TAM and UTAUT. Stud Health Technol Inform 2019 Jul 30;263:64-71. [doi: 10.3233/SHTI190111] [Medline: 31411153]
35. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315 [FREE Full text] [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
36. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
37. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
38. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 05;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
39. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA: Sage Publications; 2013.
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Boverhof BJ, Redekop WK, Bos D, Starmans MP, Birch J, Rockall A, et al. Radiology AI deployment and assessment rubric (RADAR) to bring value-based AI into radiological practice. Insights Imaging 2024 Feb 05;15(1):34 [FREE Full text] [doi: 10.1186/s13244-023-01599-z] [Medline: 38315288]
42. Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, et al. Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR derm consensus guidelines from the international skin imaging collaboration artificial intelligence working group. JAMA Dermatol 2022 Jan 01;158(1):90-96 [FREE Full text] [doi: 10.1001/jamadermatol.2021.4915] [Medline: 34851366]
43. Di Bidino R, Piaggio D, Andellini M, Merino-Barbancho B, Lopez-Perez L, Zhu T, et al. Scoping meta-review of methods used to assess artificial intelligence-based medical devices for heart failure. Bioengineering (Basel) 2023 Sep 22;10(10):1109 [FREE Full text] [doi: 10.3390/bioengineering10101109] [Medline: 37892839]
44. Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, CHEERS-AI Steering Group. Consolidated health economic evaluation reporting standards for interventions that use artificial intelligence (CHEERS-AI). Value Health 2024 Sep;27(9):1196-1205 [FREE Full text] [doi: 10.1016/j.jval.2024.05.006] [Medline: 38795956]
45. Haller S, Van Cauter S, Federau C, Hedderich DM, Edjlali M. The R-AI-DIOLOGY checklist: a practical checklist for evaluation of artificial intelligence tools in clinical neuroradiology. Neuroradiology 2022 May 31;64(5):851-864. [doi: 10.1007/s00234-021-02890-w] [Medline: 35098343]
46. Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. AJR Am J Roentgenol 2019 Jan;212(1):38-43. [doi: 10.2214/AJR.18.20224] [Medline: 30332290]
47. Jackson GP, Vergis R. Evaluation of artificial intelligence in radiation oncology. In: Mun SK, Dieterich S, editors. Artificial Intelligence in Radiation Oncology. New York, NY: World Scientific Publishing; 2023:359-368.
48. Jha AK, Bradshaw TJ, Buvat I, Hatt M, Kc P, Liu C, et al. Nuclear medicine and artificial intelligence: best practices for evaluation (the RELAINCE guidelines). J Nucl Med 2022 Sep 26;63(9):1288-1299 [FREE Full text] [doi: 10.2967/jnumed.121.263239] [Medline: 35618476]

XSL•FO
RenderX

49. Khan SD, Hoodbhoy Z, Raja MH, Kim JY, Hogg HD, Manji AA, et al. Frameworks for procurement, integration, monitoring, and evaluation of artificial intelligence tools in clinical settings: a systematic review. PLOS Digit Health 2024 May 29;3(5):e0000514 [FREE Full text] [doi: 10.1371/journal.pdig.0000514] [Medline: 38809946]
50. Larson DB, Harvey H, Rubin DL, Irani N, Tse JR, Langlotz CP. Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: summary and recommendations. J Am Coll Radiol 2021 Mar;18(3 Pt A):413-424 [FREE Full text] [doi: 10.1016/j.jacr.2020.09.060] [Medline: 33096088]
51. Lennerz JK, Salgado R, Kim GE, Sirintrapun SJ, Thierauf JC, Singh A, et al. Diagnostic quality model (DQM): an integrated framework for the assessment of diagnostic quality when using AI/ML. Clin Chem Lab Med 2023 Mar 28;61(4):544-557 [FREE Full text] [doi: 10.1515/cclm-2022-1151] [Medline: 36696602]
52. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform 2019 Aug 25;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
53. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. Med Phys 2020 Jun 17;47(5):e228-e235 [FREE Full text] [doi: 10.1002/mp.13562] [Medline: 32418341]
54. Mahmood U, Shukla-Dave A, Chan HP, Drukker K, Samala RK, Chen Q, et al. Artificial intelligence in medicine: mitigating risks and maximizing benefits via quality assurance, quality control, and acceptance testing. BJR Artif Intell 2024 Jan;1(1):ubae003 [FREE Full text] [doi: 10.1093/bjrai/ubae003] [Medline: 38476957]
55. Omoumi P, Ducarouge A, Tournier A, Harvey H, Kahn CE, Louvet-de Verchère F, et al. To buy or not to buy-evaluating commercial AI solutions in radiology (the ECLAIR guidelines). Eur Radiol 2021 Jun 05;31(6):3786-3796 [FREE Full text] [doi: 10.1007/s00330-020-07684-x] [Medline: 33666696]
56. Reddy S, Rogers W, Makinen VP, Coiera E, Brown P, Wenzel M, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ Health Care Inform 2021 Oct 12;28(1):e100444 [FREE Full text] [doi: 10.1136/bmjhci-2021-100444] [Medline: 34642177]
57. Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, et al. Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool: a new tool to assess the quality of health information provided by AI platforms. Eur Arch Otorhinolaryngol 2024 Nov 04;281(11):6123-6131 [FREE Full text] [doi: 10.1007/s00405-024-08710-0] [Medline: 38703195]
58. van Royen FS, Asselbergs FW, Alfonso F, Vardas P, van Smeden M. Five critical quality criteria for artificial intelligence-based prediction models. Eur Heart J 2023 Dec 07;44(46):4831-4834 [FREE Full text] [doi: 10.1093/eurheartj/ehad727] [Medline: 37897346]
59. Vervoort D, Tam DY, Wijeysundera HC. Health technology assessment for cardiovascular digital health technologies and artificial intelligence: why is it different? Can J Cardiol 2022 Feb;38(2):259-266. [doi: 10.1016/j.cjca.2021.08.015] [Medline: 34461229]
60. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020 Mar 20;368:l6927 [FREE Full text] [doi: 10.1136/bmj.l6927] [Medline: 32198138]
61. Vithlani J, Hawksworth C, Elvidge J, Ayiku L, Dawoud D. Economic evaluations of artificial intelligence-based healthcare interventions: a systematic literature review of best practices in their conduct and reporting. Front Pharmacol 2023 Aug 8;14:1220950 [FREE Full text] [doi: 10.3389/fphar.2023.1220950] [Medline: 37693892]
62. Abbasian M, Khatibi E, Azimi I, Oniani D, Shakeri Hossein Abad Z, Thieme A, et al. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. NPJ Digit Med 2024 Mar 29;7(1):82 [FREE Full text] [doi: 10.1038/s41746-024-01074-z] [Medline: 38553625]
63. Economou-Zavlanos NJ, Bessias S, Cary MP, Bedoya AD, Goldstein BA, Jelovsek JE, et al. Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. J Am Med Inform Assoc 2024 Feb 16;31(3):705-713. [doi: 10.1093/jamia/ocad221] [Medline: 38031481]
64. Larson DB, Doo FX, Allen B, Mongan J, Flanders AE, Wald C. Proceedings from the 2022 ACR-RSNA workshop on safety, effectiveness, reliability, and transparency in AI. J Am Coll Radiol 2024 Jul;21(7):1119-1129. [doi: 10.1016/j.jacr.2024.01.024] [Medline: 38354844]
65. Overgaard SM, Graham MG, Brereton T, Pencina MJ, Halamka JD, Vidal DE, et al. Implementing quality management systems to close the AI translation gap and facilitate safe, ethical, and effective health AI solutions. NPJ Digit Med 2023 Nov 25;6(1):218 [FREE Full text] [doi: 10.1038/s41746-023-00968-8] [Medline: 38007604]
66. Schaekermann M, Spitz T, Pyles M, Cole-Lewis H, Wulczyn E, Pfohl SR, et al. Health equity assessment of machine learning performance (HEAL): a framework and dermatology AI model case study. EClinicalMedicine 2024 Apr;70:102479 [FREE Full text] [doi: 10.1016/j.eclinm.2024.102479] [Medline: 38685924]
67. Farah L, Davaze-Schneider J, Martin T, Nguyen P, Borget I, Martelli N. Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A systematic review. Artif Intell Med 2023 Jun;140:102547 [FREE Full text] [doi: 10.1016/j.artmed.2023.102547] [Medline: 37210155]

XSL•FO
RenderX

68. Farah L, Borget I, Martelli N, Vallee A. Suitability of the current health technology assessment of innovative artificial intelligence-based medical devices: scoping literature review. J Med Internet Res 2024 May 13;26:e51514 [FREE Full text] [doi: 10.2196/51514] [Medline: 38739911]
69. Guenoun D, Zins M, Champsaur P, Thomassin-Naggara I, DRIM France AI Study Group. French community grid for the evaluation of radiological artificial intelligence solutions (DRIM France Artificial Intelligence initiative). Diagn Interv Imaging 2024 Feb;105(2):74-81. [doi: 10.1016/j.diii.2023.09.002] [Medline: 37749026]
70. Bimczok SP, Godynyuk EA, Pierey J, Roppel MS, Scholz ML. How are excellence and trust for using artificial intelligence ensured? Evaluation of its current use in EU healthcare. South East Eur J Public Health 2023 Jan 24:239 [FREE Full text] [doi: 10.56801/seejph.vi.239]
71. de Hond AA, Leeuwenberg AM, Hooft L, Kant IM, Nijman SW, van Os HJ, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022 Jan 10;5(1):2 [FREE Full text] [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
72. Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic review of health economic evaluations focused on artificial intelligence in healthcare: the tortoise and the cheetah. Value Health 2022 Mar;25(3):340-349 [FREE Full text] [doi: 10.1016/j.jval.2021.11.1362] [Medline: 35227444]
73. Ding H, Simmich J, Vaezipour A, Andrews N, Russell T. Evaluation framework for conversational agents with artificial intelligence in health interventions: a systematic scoping review. J Am Med Inform Assoc 2024 Feb 16;31(3):746-761. [doi: 10.1093/jamia/ocad222] [Medline: 38070173]
74. Goergen SK, Frazer HM, Reddy S. Quality use of artificial intelligence in medical imaging: what do radiologists need to know? J Med Imaging Radiat Oncol 2022 Mar 03;66(2):225-232. [doi: 10.1111/1754-9485.13379] [Medline: 35243782]
75. Lehoux P, Rocha de Oliveira R, Rivard L, Silva HP, Alami H, Mörch CM, et al. A comprehensive, valid, and reliable tool to assess the degree of responsibility of digital health solutions that operate with or without artificial intelligence: 3-phase mixed methods study. J Med Internet Res 2023 Aug 28;25:e48496 [FREE Full text] [doi: 10.2196/48496] [Medline: 37639297]
76. Tanguay W, Acar P, Fine B, Abdolell M, Gong B, Cadrin-Chênevert A, et al. Assessment of radiology artificial intelligence software: a validation and evaluation framework. Can Assoc Radiol J 2023 May 06;74(2):326-333. [doi: 10.1177/08465371221135760] [Medline: 36341574]
77. Ji M, Genchev GZ, Huang H, Xu T, Lu H, Yu G. Evaluation framework for successful artificial intelligence–enabled clinical decision support systems: mixed methods study. J Med Internet Res 2021 Jun 2;23(6):e25929. [doi: 10.2196/25929]
78. Fasterholdt I, Naghavi-Behzad M, Rasmussen BS, Kjølhede T, Skjøth MM, Hildebrandt MG, et al. Value assessment of artificial intelligence in medical imaging: a scoping review. BMC Med Imaging 2022 Oct 31;22(1):187 [FREE Full text] [doi: 10.1186/s12880-022-00918-y] [Medline: 36316665]
79. Gomez Rossi J, Feldberg B, Krois J, Schwendicke F. Evaluation of the clinical, technical, and financial aspects of cost-effectiveness analysis of artificial intelligence in medicine: scoping review and framework of analysis. JMIR Med Inform 2022 Aug 12;10(8):e33703 [FREE Full text] [doi: 10.2196/33703] [Medline: 35969458]
80. Panagoulias DP, Virvou M, Tsihrintzis GA. Applying DOI theory to assess the required level of explainability in artificial intelligence-empowered medical applications. In: Proceedings of the 14th International Conference on Information, Intelligence, Systems & Applications. 2023 Presented at: IISA '23; July 10-12, 2023; Volos, Greece p. 1-7 URL: https://ieeexplore.ieee.org/document/10345846 [doi: 10.1109/iisa59645.2023.10345846]
81. Bhatnagar S. Checklist for medical imaging using artificial intelligence by evaluation of machine learning models. In: Proceedings of the 5th International Conference on Inventive Research in Computing Applications. 2023 Presented at: ICIRCA '23; August 3-5, 2023; Coimbatore, India p. 865-871 URL: https://ieeexplore.ieee.org/document/10220939 [doi: 10.1109/icirca57980.2023.10220939]
82. Alshehri S, Alahmari KA, Alasiry A. A comprehensive evaluation of AI-assisted diagnostic tools in ENT medicine: insights and perspectives from healthcare professionals. J Pers Med 2024 Mar 28;14(4):354 [FREE Full text] [doi: 10.3390/jpm14040354] [Medline: 38672981]
83. Lundström C, Lindvall M. Mapping the landscape of care providers' quality assurance approaches for AI in diagnostic imaging. J Digit Imaging 2023 Apr 09;36(2):379-387 [FREE Full text] [doi: 10.1007/s10278-022-00731-7] [Medline: 36352164]
84. Ross J, Hammouche S, Chen Y, Rockall A, Royal College of Radiologists AI Working Group. Beyond regulatory compliance: evaluating radiology artificial intelligence applications in deployment. Clin Radiol 2024 May;79(5):338-345 [FREE Full text] [doi: 10.1016/j.crad.2024.01.026] [Medline: 38360516]
85. Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. Res Meth Med Health Sci 2020 Aug 06;1(1):31-42. [doi: 10.1177/2632084320947559]
86. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE evidence standards framework for digital health and care technologies - developing and maintaining an innovative evidence framework with global impact. Digit Health 2021 Jun 24;7:20552076211018617 [FREE Full text] [doi: 10.1177/20552076211018617] [Medline: 34249371]

XSL•FO
RenderX

87. Sarwar N, Irshad A, Naith QH, D Alsufiani K, Almalki FA. Skin lesion segmentation using deep learning algorithm with ant colony optimization. BMC Med Inform Decis Mak 2024 Sep 27;24(1):265 [FREE Full text] [doi: 10.1186/s12911-024-02686-x] [Medline: 39334181]
88. Zubair M, Owais M, Mahmood T, Iqbal S, Usman SM, Hussain I. Enhanced gastric cancer classification and quantification interpretable framework using digital histopathology images. Sci Rep 2024 Sep 28;14(1):22533 [FREE Full text] [doi: 10.1038/s41598-024-73823-9] [Medline: 39342030]
89. Bourdillon AT, Garg A, Wang H, Woo YJ, Pavone M, Boyd J. Integration of reinforcement learning in a virtual robotic surgical simulation. Surg Innov 2023 Feb 03;30(1):94-102. [doi: 10.1177/15533506221095298] [Medline: 35503302]
90. Muralidharan V, Adewale BA, Huang CJ, Nta MT, Ademiju PO, Pathmarajah P, et al. A scoping review of reporting gaps in FDA-approved AI medical devices. NPJ Digit Med 2024 Oct 03;7(1):273 [FREE Full text] [doi: 10.1038/s41746-024-01270-x] [Medline: 39362934]
91. Yang J, Dung NT, Thach PN, Phong NT, Phu VD, Phu KD, et al. Generalizability assessment of AI models across hospitals in a low-middle and high income country. Nat Commun 2024 Sep 27;15(1):8270 [FREE Full text] [doi: 10.1038/s41467-024-52618-6] [Medline: 39333515]
92. Barrett D, Heale R. What are Delphi studies? Evid Based Nurs 2020 Jul 19;23(3):68-69. [doi: 10.1136/ebnurs-2020-103303] [Medline: 32430290]
93. Grime MM, Wright G. Delphi method. In: Grime MM, editor. Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; 2016:1-6.
94. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021 Jul 20;11(4):116-129 [FREE Full text] [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]
95. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and REporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017 Sep 13;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
96. Dagan N, Devons-Sberro S, Paz Z, Zoller L, Sommer A, Shaham G, et al. Evaluation of AI solutions in health care organizations — the OPTICA tool. NEJM AI 2024 Aug 22;1(9):65. [doi: 10.1056/aics2300269]
97. Callahan A, McElfresh D, Banda JM, Bunney G, Char D, Chen J, et al. Standing on FURM ground: a framework for evaluating fair, useful, and reliable AI models in health care systems. NEJM Catal 2024;5(10):131. [doi: 10.1056/cat.24.0131]
98. Ning Y, Teixayavong S, Shang Y, Savulescu J, Nagaraj V, Miao D, et al. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. Lancet Digit Health 2024 Nov;6(11):e848-e856 [FREE Full text] [doi: 10.1016/S2589-7500(24)00143-2] [Medline: 39294061]

Abbreviations

AI: artificial intelligence
CASP: Critical Appraisal Skills Program
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
EU: European Union
PICO: participants, intervention, comparators, and outcomes
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

Edited by A Mavragani; submitted 14.10.24; peer-reviewed by D Vogel, Q Ng; comments to author 04.11.24; revised version received 14.11.24; accepted 30.12.24; published 05.02.25.

Please cite as:
Jacob C, Brasier N, Laurenzi E, Heuss S, Mougiakakou SG, Cöltekin A, Peter MK
AI for IMPACTS Framework for Evaluating the Long-Term Real-World Impacts of AI-Powered Clinician Tools: Systematic Review and Narrative Synthesis
J Med Internet Res 2025;27:e67485
URL: https://www.jmir.org/2025/1/e67485
doi:10.2196/67485
PMID:

©Christine Jacob, Noé Brasier, Emanuele Laurenzi, Sabina Heuss, Stavroula-Georgia Mougiakakou, Arzu Cöltekin, Marc K Peter. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the

XSL•FO
RenderX

Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Citizens' Response to Social Activities on Twitter in US Metropolises During the COVID-19 Recovery Phase Using a Fine-Tuned Large Language Model: Application of AI

Ryuichi Saito[1]; Sho Tsugawa[1], PhD

Institute of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan

Corresponding Author:
Ryuichi Saito
Institute of Systems and Information Engineering
University of Tsukuba
1-1-1 Tennodai
Tsukuba, 305-8577
Japan
Phone: 81 08055751714
Email: saito.ryuichi.tkb_gw@u.tsukuba.ac.jp

Abstract

Background: The COVID-19 pandemic continues to hold an important place in the collective memory as of 2024. As of March 2024, >676 million cases, 6 million deaths, and 13 billion vaccine doses have been reported. It is crucial to evaluate sociopsychological impacts as well as public health indicators such as these to understand the effects of the COVID-19 pandemic.

Objective: This study aimed to explore the sentiments of residents of major US cities toward restrictions on social activities in 2022 during the transitional phase of the COVID-19 pandemic, from the peak of the pandemic to its gradual decline. By illuminating people's susceptibility to COVID-19, we provide insights into the general sentiment trends during the recovery phase of the pandemic.

Methods: To analyze these trends, we collected posts (N=119,437) on the social media platform Twitter (now X) created by people living in New York City, Los Angeles, and Chicago from December 2021 to December 2022, which were impacted by the COVID-19 pandemic in similar ways. A total of 47,111 unique users authored these posts. In addition, for privacy considerations, any identifiable information, such as author IDs and usernames, was excluded, retaining only the text for analysis. Then, we developed a sentiment estimation model by fine-tuning a large language model on the collected data and used it to analyze how citizens' sentiments evolved throughout the pandemic.

Results: In the evaluation of models, GPT-3.5 Turbo with fine-tuning outperformed GPT-3.5 Turbo without fine-tuning and Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach (RoBERTa)–large with fine-tuning, demonstrating significant accuracy (0.80), recall (0.79), precision (0.79), and F_1-score (0.79). The findings using GPT-3.5 Turbo with fine-tuning reveal a significant relationship between sentiment levels and actual cases in all 3 cities. Specifically, the correlation coefficient for New York City is 0.89 (95% CI 0.81-0.93), for Los Angeles is 0.39 (95% CI 0.14-0.60), and for Chicago is 0.65 (95% CI 0.47-0.78). Furthermore, feature words analysis showed that COVID-19–related keywords were replaced with non–COVID-19-related keywords in New York City and Los Angeles from January 2022 onward and Chicago from March 2022 onward.

Conclusions: The results show a gradual decline in sentiment and interest in restrictions across all 3 cities as the pandemic approached its conclusion. These results are also ensured by a sentiment estimation model fine-tuned on actual Twitter posts. This study represents the first attempt from a macro perspective to depict sentiment using a classification model created with actual data from the period when COVID-19 was prevalent. This approach can be applied to the spread of other infectious diseases by adjusting search keywords for observational data.

(J Med Internet Res 2025;27:e63824) doi:10.2196/63824

KEYWORDS

COVID-19; restriction; United States; X; Twitter; sentiment analysis; large language model; LLM; GPT-3.5; fine-tuning

XSL•FO
RenderX

Introduction

Background

The global SARS-CoV-2 outbreak, which began in 2020, remains vivid in the collective memory of humanity as of 2024. It is beyond dispute that it was an unprecedented pandemic in human history. For >3 years since the World Health Organization declared the novel COVID-19 a Public Health Emergency of International Concern on January 30, 2020 [1,2], humanity has grappled with COVID-19 as a major social challenge. As of March 2024, the total number of COVID-19 cases has surpassed 676 million, >6 million people have died, and >13 billion vaccine doses have been administered [3]. To assess the scale and severity of the COVID-19 pandemic, it is crucial to confirm public health indicators such as these. While numerous studies have been undertaken to address the pandemic, from artificial intelligence and medical imaging to printing technology [4], it is also important to conduct sociopsychological observations to understand how citizens have perceived the COVID-19 pandemic.

We assess public sentiment during the pandemic by analyzing social media posts using natural language processing (NLP) techniques, which have significantly advanced since the 2010s. An NLP-based approach can address the limitations of traditional social science research methods, which are often constrained by limited observational data and the potential for indirect bias from nonrespondents [5]. Previous studies that analyzed sentiment during the COVID-19 pandemic can be categorized into 2 main groups. One group focused its investigations on the initial outbreak that occurred in March 2020 [6-10], while the other focused on specific thematic areas, such as vaccination efforts [11-14]. However, no study to date has captured the cyclic fluctuations in social sentiment from macro and long-term perspectives because these studies are based on data collected using keywords related to COVID-19, and these words usually reflect negative feelings. Furthermore, despite variations in the spread of COVID-19 between urban and rural areas [15], many studies have focused on sentiment analysis at the level of language or country, thereby limiting the extraction of insights considering local infection status, which is crucial from a public health perspective.

Our previous study [16] analyzed social media posts about citizens' activities that were constrained by the pandemic, and the degree of sentiment changed depending on the context of the source observational data instead of COVID-19–related texts. In addition, we focused on major American cities such as New York City, Los Angeles, and Chicago, which have similar social conditions. Using this approach, we addressed the problems of previous studies and revealed the general trends of citizens' sentiments from December 2019 to December 2021 during the COVID-19 pandemic.

This study further develops this research and examines the period from December 2021, when the Omicron variants surged, to December 2022, a period when public interest in COVID-19 generally waned. In the United States in 2021, outbreaks caused by the Omicron variant strain continued intermittently, starting with BA.1 at the end of 2021, followed by BA.2 and BA.3 [17], prompting state governments to issue repeated warnings and subsequently rescind them. The World Health Organization's Public Health Emergency of International Concern ended in May 2023 [2], and thus, 2022 can be considered a transition period from the pandemic period to the postpandemic period. We believe it is important to observe citizens' sentiments in the United States during this period. In addition, because previous research involving sentiment classification models relied on lexicon dictionaries or training datasets before 2020 when SARS-CoV-2 emerged, the domain adaption between the lexicon dictionary or training data (source domain) and the observational data (target domain) [18] was not sufficiently considered in NLP. To address this problem, we constructed a sentiment classification model fine-tuned using data extracted under conditions identical to those of the actual observational dataset. Using these methodologies, we propose a definitive approach for estimating social sentiment during the COVID-19 pandemic.

Objectives

Sentiment classification methods based on algorithms can be broadly divided as follows: lexicon-based approaches that infer text sentiment from tokens that have been assigned sentiment scores in advance and machine learning approaches that infer text sentiment from models trained on datasets [19]. There are several studies on lexicon-based sentiment analysis during a crisis. A study in Wuhan, China, tracked the change in public sentiment during the first 12 weeks after the identification of COVID-19 by examining posts on Weibo. The study revealed the pattern of trajectories—from confusion and fear, through disappointment, frustration, depression, and anxiety, and finally to happiness and gratitude—using the Emotion Vocabulary of the Dalian University of Technology [7]. A study in Europe and the United States measured the sentiment toward immigration by examining related tweets in Germany, Italy, Spain, the United Kingdom, and the United States during the early stages of the COVID-19 pandemic. The study estimated the sentiment scores of tweets, using Valence Aware Dictionary and Sentiment Reasoner, and revealed various themes regarding immigration for each country through topic modeling [20].

Some studies on sentiment analysis during crises use machine learning approaches. An analysis in the United States investigated the communication patterns of provaccine and antivaccine users on Twitter (now called X) by visualizing a retweet network related to the measles, mumps, and rubella vaccine during the 2015 California Disneyland measles outbreak. The study classified the users into provaccination, antivaccination, and neutral groups using a support vector machine. The results showed that most of the users were overwhelmingly provaccination, while antivaccine users resided in their own enclosed structural community [21]. A study in the United States also analyzed public perceptions of gig work on Twitter for 2 weeks before and after the COVID-19 emergency declaration. The study trained a machine learning model based on 10 different labels. The results showed that tweets reflected an increased sense of community and concern toward gig workers during the pandemic [22]. An investigation on Instagram demonstrated the changes in hate and misinformation against East Asians that occurred through Meta's content

XSL•FO
RenderX

moderation early in the COVID-19 pandemic. The study used supervised machine learning methods to estimate text emotions associated with human faces in posts with the hashtag #coronavirus [8]. A study on Twitter and Weibo revealed a pattern in which negative sentiment peaked before the 2020 lockdowns in many countries, followed by a gradual recovery in sentiment [9]. A study in China examined the psychological impacts of the COVID-19 epidemic declaration on individuals using Weibo data. Findings revealed an increase in negative emotions, such as anxiety and depression, alongside a decrease in life satisfaction and positive emotions using the pretrained psychological prediction model [10].

The disadvantage of lexicon-based approaches is that they are highly domain oriented [19]. Then, to predict sentiment in a specific domain, such as the discourse space of Twitter during the COVID-19 pandemic, a machine learning–based approach using training data, especially a neural network approach, is reasonable in terms of accuracy. This approach involves Transformer-based models with attention mechanisms [23], which assign weights to tokens based on their relationships, without processing time-series data sequentially, as is done in natural language. Language models such as GPT-1 [24] and Bidirectional Encoder Representations from Transformers (BERT) [25], which leverage the Transformer architecture and have undergone extensive pretraining on large datasets, have become readily available tools for NLP tasks.

A study on Twitter examined how users with different ideological views and follower bases expressed vaccine favorability and specific vaccine-related concerns. Users' perception of the vaccine was classified by a fine-tuned BERT model using training data that were coded by the authors. The results from linear mixed-effects models showed a contrast in vaccines between conservative and liberal users, and users with large numbers of followers tended to be more favorable toward vaccines, while those with an average number of followers were prone to be more concerned about vaccines [11]. A study from January 2021 to January 2022 on Twitter used a change-point detection algorithm to identify significant shifts in public sentiment regarding the pandemic. Validation of these findings was accomplished by cross-referencing with contemporaneous news reports. Furthermore, the estimation used BERT fine-tuned with labeled tweets from Kaggle to gauge public attitudes [26]. A study on Twitter and Reddit investigates public sentiment regarding COVID-19 vaccines from January 2020 to March 2022. Using a fine-tuned DistilRoBERTa model, augmented with back-translation, it reveals that Twitter sentiment was predominantly negative, whereas Reddit sentiment was mostly positive [13]. A study in the United Kingdom used a hybrid model, combining Valence Aware Dictionary and Sentiment Reasoner, TextBlob, and BERT, to analyze public perceptions of COVID-19 contact tracing apps on Twitter and Facebook. Results indicated varying sentiments influenced by debates on centralized versus decentralized data handling in app-based contact tracing [27].

In a language model based on the Transformer architecture, there is a pretraining stage in which initial parameters are set and a fine-tuning stage in which parameters are adapted for a specific target task or domain. Given that language models may not possess adequate pretraining on social discourse after major transformative events such as the COVID-19 pandemic, it is appropriate to fine-tune the model using training data with conditions similar to the observed data. Therefore, the following research question (RQ) initially guided our study: *To what extent can the performance of classification tasks be enhanced through the use of posts on Twitter during the pandemic as training data? (RQ1)*

In addition, since 2020, there have been advancements in large language models (LLMs) such as GPT-3, making them capable of not only understanding but also generating sentences [28]. In this paper, we use GPT-3.5 Turbo [29], which is the latest version that allows fine-tuning, and we ask the following question: *To what extent does GPT-3.5 Turbo enhance the performance of sentiment classification tasks in comparison with conventional Transformer-based models? (RQ2)*

Previous studies have focused mainly on the several months of lockdown starting in April 2020 and the rollout of vaccinations from late 2020 onward, but no study has captured the long-term trend in social sentiment for specific regions during the COVID-19 pandemic period from 2020 to 2023. These tendencies simultaneously indicate the difficulty of analyzing public sentiment during the COVID-19 pandemic from a macro and diachronic perspective. We intend to undertake a long-term and macroscopic observation by targeting text that discusses the social activities of citizens restricted by lockdowns and similar measures implemented in response to the spread of COVID-19, rather than focusing solely on text explicitly containing keywords such as "COVID-19," "virus," and "lockdown," as was the case in the previous studies. According to a study about adaptation to the "new normal," the term "the new normal" first emerged during the 2008 financial crisis and was used again to point out how the COVID-19 pandemic had changed essential aspects of human life [30]. A study compared risk perception in a sample in April 2021 and January 2022, based on the assumption that the "new normal" involved ongoing restrictions starting with widespread vaccine access and the vaccination of populations considered vulnerable. The study found that people tend to overestimate COVID-19 risks, particularly for children and healthy individuals [31]. This study considers this "new normal" period in the time frame from December 2021 to December 2022 and asks the following question: *To what extent will the cyclical sentiment of citizens toward restricted activities in major US cities during the "new normal" period weaken over time following its peak in December 2021? (RQ3)*

Previous studies focused on observations from countries and language areas such as the English-speaking world. However, a geospatial study showed that the spread of COVID-19 in the United States differed between urban and rural areas even within the same state and that infection status was correlated with each county's social vulnerability and community resilience [15]. Moreover, a study that surveyed rural and urban areas during the pandemic showed that rural residents were less sensitive to preventive health behaviors compared with urban residents due to the influence of political ideology or demographic factors [32]. On the basis of these findings, it is imperative to distinguish and analyze urban and rural areas separately,

XSL•FO
RenderX

particularly within the US context. Therefore, in this study, we intentionally focused on major urban centers within the United States to estimate social sentiment and asked the following question: *In comparing citizens' sentiments across major metropolitan areas in the United States, namely, New York City, Los Angeles, and Chicago, between December 2021 and December 2022, what differences or similarities in sentiment were there? (RQ4)*

Methods

Overview

We collected posts on Twitter related to restricted activities during the COVID-19 pandemic in New York City, Los Angeles, and Chicago from October 2021 to December 2022. Posts from October 2021 to November 2021 were used to train the sentiment classification model, while posts from December 2021 to December 2022 were used to estimate sentiment and extract feature words. Then, the correlation coefficient between estimated sentiment and actual infected cases was determined. In addition, the complete code for this study is available on the GitHub repository [33].

Data Collection

In this study, we chose the 3 largest cities in the United States according to the US Census Bureau [34]: New York City, Los Angeles, and Chicago. Specifically, New York County, Los Angeles County, and Cook County were selected as observation targets. In the data collection process, the Full-archive Search Application Programming Interface (API) of Twitter API v2 [35] was used to retrieve messages posted within a 25-mile radius of the latitude and longitude of the city hall for each city. This radius is the upper limit of the Full-archive Search API, ensuring comprehensive coverage of the urban core. Our observation period was from December 2021, when the emergence of the Omicron variant led to record-high number of cases in the United States, to December 2022, when interest in COVID-19 had waned; this builds on the author's previous study [16] that estimated sentiment in the same 3 cities from December 2019 to January 2022. Posts from October 2021 to November 2021 were also used as training data for neural network models.

The search string used for retrieving posts on Twitter uses keywords associated with citizens' activities that were constrained by lockdowns and similar measures implemented in response to the COVID-19 pandemic, as described in the study by Saito and Haruyama [16]. These keywords are shown in Multimedia Appendix 1. The use of these specific keywords enables a coherent overview of the citizens' psychological landscape from a macro perspective. These keywords are categorized by restriction type (ie, stay-at-home order, restrictions on gatherings, and travel restrictions) and are specified as arguments for the Twitter API version 2's Search API, using "OR" conditions. In addition, retweets were excluded from the search, and only English posts were retrieved. The retrieved posts were processed to remove noise, including URLs, mentions, and hashtags, followed by text normalization and deduplication, before being used for training and classification with neural networks.

Observational Data

We obtained posts from December 2021 to December 2022 using the Search API of Twitter API v2. The total number of posts was 119,437, and the total file size was 34.4 MB. Table 1 shows the number of posts by restriction type, and the number of unique users. Moreover, this data collection was conducted during the period from June 15, 2023, to June 22, 2023. For privacy protection, identifying information such as author IDs and usernames was removed, retaining only the text for analysis.

Table 1. Number of posts and unique users collected from Twitter between December 2021 and December 2022.

	New York City	Los Angeles	Chicago
Posts, n (%)			
Stay-at-home order	12,969 (23.97)	9622 (21.98)	5308 (24.64)
Restrictions on gatherings	14,095 (26.05)	13,448 (30.71)	5803 (26.94)
Travel restrictions	27,045 (49.98)	20,715 (47.31)	10,432 (48.42)
Total	54,109 (100.00)	43,785 (100.00)	21,543 (100.00)
Unique users, n	20,575[a]	17,588	8948

[a]For 113 tweets in New York City, unique users could not be identified because the author ID could not be obtained. Therefore, they were excluded from this table.

Neural Network Model

To create a sentiment classifier, we used a Transformer-based neural network model using training data that are searched on the same data source, keywords, geolocation information, and language except for a period as actual observed data.

Training Data

As training data, we used Twitter posts obtained using Search API of Twitter API v2.0 under the same search keywords, geolocation, and language as observed data during the period from October 31, 2021, to November 27, 2021. This period was chosen to prevent any overlap with the observation data. The total number of results obtained was 3149, and the retrieved posts were preprocessed, such as noise removal, text normalization, and deduplication. For these pieces of data, an author and 2 Amazon Mechanical Turk (MTurk) workers labeled them as positive, negative, or neutral based on the instructions shown in Textbox 1, and the final label was determined by a majority vote. Tweets that could not be decided by a majority

XSL•FO
RenderX

vote because each of the 3 people indicated different opinions were labeled as neutral. Concerning the instructions in Textbox 1, positive and negative sentiments were defined based on the available sentiment by [36], and neutral sentiments were defined by the authors according to the characteristics of the Twitter posts obtained. We also selected workers with master's qualifications who have consistently been recognized for their high performance on Amazon MTurk. In addition, only texts were extracted for individual privacy measures, and due to Amazon MTurk's inability to display 4-byte Unicode Transformation Format-8 emojis, we converted them to HTML spans in the task instructions, ensuring that the workers considered emojis when evaluating emotions. After labeling 3149 posts, the unanimous agreement rate among all 3 evaluators was 55.22% (1739/3149), and the 2-out-of-3 agreement rate was 28.74% (905/3149). The remaining 16.04% (505/3149) of posts were marked as neutral because they were of different values.

Labeled posts were shuffled and then filtered to extract 2400 instances for training the neural network model. These were then partitioned into training (1800/2400, 75%) and validation (600/2400, 25%) datasets. The remaining 749 instances were reserved for testing purposes. Table 2 shows the number of divided posts for each sentiment.

Textbox 1. Instructions for 3 evaluators to create training data and the prompt for GPT-3.5 Turbo for classifying sentiment. In the prompt in GPT-3.5 Turbo, this instruction was modified only to return numerical values such as positive to 0, neutral to 1, and negative to 2.

You are requested to perform sentiment classification (positive, negative, or neutral) for each tweet based on the following rules. These tweets have been collected for the purpose of scientific research.

- Positive sentiments: admire, amazing, assure, celebration, charm, eager, enthusiastic, excellent, fancy, fantastic, frolic, graceful, happy, joy, luck, majesty, mercy, nice, patience, perfect, proud, rejoice, relief, respect, satisfactorily, sensational, super, terrific, thank, vivid, wise, wonderful, zest, expectations, etc.
- Negative sentiments: abominable, anger, anxious, bad, catastrophe, cheap, complaint, condescending, deceit, defective, disappointment, embarrass, fake, fear, filthy, fool, guilt, hate, idiot, inflict, lazy, miserable, mourn, nervous, objection, pest, plot, reject, scream, silly, terrible, unfriendly, vile, wicked, etc.
- Neutral sentiments: neither positive nor negative, such as text without sentiment, stating a fact, question, news article, advertisement, solicitation, request, quote, unintelligible text, etc.
- When the sentiment is mixed, such as expressing both joy and sadness, use your judgment and choose the more strongly expressed emotion.

Table 2. Training, validation, and test data for creating the neural network models.

Data types	Positive, n (%)	Neutral, n (%)	Negative, n (%)
Training data (n=1800)	648 (36)	515 (28.61)	637 (35.39)
Validation data (n=600)	225 (37.5)	159 (26.5)	216 (36)
Test data (n=749)	270 (36.05)	214 (28.57)	265 (35.38)

Training of the Neural Network Model

For sentiment classification in this research, we adopted GPT-3.5 Turbo, the latest version of GPT-3 that is trained using huge parameters and can be fine-tuned using real data. GPT-3.5 was trained on data up to September 2021, marking a distinction from GPT-3, which used training data through 2019. This difference suggests a substantial potential contribution to the contextual analysis of the globally prevalent COVID-19 pandemic since 2020. In the development of our model, we used gpt-3.5-turbo-1106 as the base model, subsequently fine-tuning it with the training data described in the Training Data section. The model was created using the following hyperparameters: a learning rate multiplier of 2, a batch size of 3, and a total of 3 epochs. These values were automatically optimized by the fine-tuning API based on the size of the training dataset. Figure 1 presents the learning curve, where the x-axis represents the number of steps and the y-axis represents the loss, which indicates how closely the model's predictions align with the actual outcomes. The validation loss, which reflects the model's performance on unseen data, decreases until approximately step 500 but then fluctuates intermittently up to step 1500, suggesting potential overfitting to the training data. Toward the final step, however, the validation loss decreases to 0.39, indicating that the model stabilizes and improves its generalization performance.

XSL•FO
RenderX

Figure 1. The learning curve of fine-tuning GPT-3.5 Turbo.



Ethical Considerations

The study was conducted using publicly posted large volumes of Twitter data collected via the Twitter API, which is permitted only for academic research but does not involve any human subject research. Furthermore, these posts are not produced within a virtual community on Twitter. Data collection, creation of training and test datasets, and analysis were carried out following Twitter's terms and conditions. In addition, all identifying information, such as author IDs, tweet IDs, display names, and usernames, was excluded, with only the tweet content used for the research. This study also ensures adherence to the guidelines outlined in the study by Luo et al [37].

Results

Evaluation of Sentiment Estimation Models

To evaluate the developed classification model, we compared it with the model Robustly Optimized BERT Pretraining Approach (RoBERTa)–large [38], which was fine-tuned using the same training data as the developed model, as well as GPT-3.5 Turbo, which was deliberately not fine-tuned. In addition, the following hyperparameters were used in fine-tuning a RoBERTa-large model: a learning rate of 5e-5, a batch size of 16 for training and 64 for validation, and a total of 3 epochs. In both GPT-3.5 Turbo and GPT-3.5 Turbo with fine-tuning, the same prompt (the role's system parameter of the API) was set based on Textbox 1, with the only change being the modification to return numerical values, such as 0 for positive, 1 for neutral, and 2 for negative. Both models also implemented exception handling to skip the output as a value error if the generated text is not numeric. (There was only 1 value error in this evaluation when using GPT-3.5 Turbo with fine-tuning.)

Table 3 shows the performance evaluation of each model. Accuracy is the proportion of correctly predicted instances out of all the cases in the test data. Recall is the proportion of correctly predicted instances out of all actual instances for each class (positive, neutral, and negative). Precision is the proportion of correctly predicted instances out of all the cases predicted as each class. Although the neutral class recall was low for all models, with precision remaining relatively stable across all models, GPT-3.5 Turbo with fine-tuning still significantly outperformed the other models in the recall. This difference might have a significant impact on the accuracy difference. F_1-score is a metric generally used to evaluate the performance of neural network models, and it provides a balanced assessment by considering both recall and precision. In both models, the F_1-score value for the neutral label was much lower than the values for the other labels, which means that the subtle nuances of the boundaries between negative and neutral and between neutral and positive were not sufficiently trained. However, in both models, the F_1-scores for the positive and negative labels exceeded 0.80, so we assume that there was no fatal effect on the results. Comparing the 2 models, RoBERTa-large and GPT-3.5 Turbo showed small differences in accuracy and average F_1-score, but it can be said that GPT-3.5 Turbo without fine-tuning showed the same performance as RoBERTa-large with fine-tuning.

RQ1 asks to what extent the performance of classification tasks can be enhanced through the use of Twitter's posts during the pandemic as training data. As shown in Table 3, GPT-3.5 Turbo with fine-tuning had a 5% higher accuracy and a 6% higher F_1-score compared with GPT-3.5 Turbo without fine-tuning, which clearly demonstrates the improvement in performance because of fine-tuning. RQ2 asks to what extent GPT-3.5 Turbo enhances the performance of sentiment classification tasks compared with conventional Transformer-based models. GPT-3.5 Turbo with fine-tuning had a 4% higher accuracy and a 7% higher F_1-score compared with RoBERTa-large with fine-tuning, which distinctly demonstrates the performance improvement because of large-scale learning. On the basis of the abovementioned results, this study used GPT-3.5 Turbo with fine-tuning.

Table 4 shows examples of sentiments classified and indexed by GPT-3.5 Turbo with fine-tuning. Each tweet is classified as –1 for positive, 0 for neutral, or 1 for negative. The higher the value of the index, the more pessimistic the sentiment is throughout the week, and the lower the value, the more optimistic the sentiment.

XSL•FO
RenderX

Table 3. Performance evaluation of the neural network models for sentiment classification.

Model and class	Accuracy	Recall	Precision	F_1-score
RoBERTa[a]-large with fine-tuning	0.76			
Positive		0.87	0.76	0.82
Neutral		0.37	0.85	0.52
Negative		0.94	0.72	0.82
Average		0.73	0.78	0.72
GPT-3.5 Turbo without fine-tuning	0.75			
Positive		0.83	0.80	0.81
Neutral		0.48	0.73	0.58
Negative		0.90	0.73	0.81
Average		0.74	0.75	0.73
GPT-3.5 Turbo with fine-tuning	0.80			
Positive		0.89	0.81	0.85
Neutral		0.57	0.74	0.64
Negative		0.91	0.83	0.86
Average		0.79	0.79	0.79

[a]RoBERTa: Robustly Optimized BERT Pretraining Approach.

Table 4. Sample of tweets classified by GPT-3.5 Turbo with fine-tuning.

Tweets	Score	Sentiment
"I get you. I feel like masks block my turkey gobbler neck, so, sometimes I don't mind wearing them. That, and they keep my face warm in the winter when I go out. .	–1	Positive
"teaches you how to be an effective, empathetic leader in 0 lessons that you can find at Shot at our Level-0 complex of stages in Sunset Park! . Tô learn more about our stages, please visit ."	0	Neutral
"I have both vaccines and just got the booster (which takes a while to kick in). Unfortunately, there are a lot of people who go out unmasked when they're sick and are spreading it like crazy,"	1	Negative

Result of Sentiment Estimation

Overview

To answer RQ3 and RQ4, the sentiment was extracted from posts made between November 28, 2021, and December 31, 2022, in New York City, Los Angeles, and Chicago. Similarly, the number of new cases in New York City, Los Angeles, and Chicago between November 28, 2021, and December 31, 2022, was taken from the New York Times COVID-19 data [39]. The research about the cyclicity of COVID-19 cases reported that in the United States in 2020, the number of new cases was lower on weekends than on weekdays [40]. Therefore, the sentiment was expressed as the weekly arithmetic mean of the sentiment score in each period, which was classified as –1 for positive, 0 for neutral, and 1 for negative. In addition, a 4-week moving average was applied to the sentiment data to smooth out random fluctuations or noise and to account for sentiment trends preceding the number of infected cases [16]. This value is defined as the sentiment index. At the same time, new cases were aggregated weekly to offset the weekend's effects. The cases were described as a logarithm of 10 to normalize extreme increases due to specific mutant strains. Moreover, these time-series data were plotted on the graph to understand the patterns and trends and to confirm the relevance of each city's timeline. To quantify the degree and confirm the direction of the linear relationship between the sentiment index and new cases, the correlation coefficient was used to extract the main topics and themes that characterize each period when the number of cases increased or decreased.

New York City

The left axis of Figure 2 shows the number of cases represented by a logarithmic scale in New York County, New York, and the right axis shows the average sentiment over 4 weeks extracted by GPT-3.5 Turbo fine-tuned with training data. In December 2021, citizens in New York were exposed to a surge in Omicron variant [41] cases and subsequently experienced cyclical waves of infections from subvariants such as Omicron BA.2 [42] and Omicron XBB.1.5 [43], which became predominant, particularly in the Northeast, by January 2023. Unlike in 2020, when COVID-19 had a high death rate, measures such as stay-at-home orders, bans on large gatherings, and travel restrictions were not enforced by the state government during this period, although mandates such as vaccinations [44]

XSL•FO
RenderX

for specific workers and indoor mask wearing [45] in clinics were implemented or recommended.

Here, we performed a statistical test on the sentiment waveform. As shown in Table 5, the correlation coefficient between sentiment and cases is 0.89, indicating a strong positive relationship. This suggests that in New York City, the sentiment of posts related to pandemic restrictions was consistently associated with COVID-19 infection status throughout the period. The correlation between the sentiment of posts related to restrictions on gatherings and cases was relatively weaker at 0.40 than for other types. There is a possibility that certain anxiety about commuting to work, going to school, and traveling may have persisted due to the spread of COVID-19, but it can be assumed that citizens exercised less caution when it came to leisure activities such as going to the movies, going on dates, and shopping.

Table 6 shows the feature words extracted using term frequency–inverse document frequency (TF-IDF) for each period according to the infection status. During December 2021, when infections caused by the Omicron strain spiked, themes related to polymerase chain reaction tests and the Omicron strain were dominant, whereas from January 2022 onward, concerns about COVID-19 ceased to be the main topic. Although Table 5 confirms a strong relationship between sentiment and infection status, the number of posts using keywords related to COVID-19 decreased from the beginning of 2022.

Figure 2. New COVID-19 cases and sentiment index extracted by GPT-3.5 Turbo with fine-tuning in New York City. The closer the sentiment index is to 1, the higher the degree of negativity, while the closer it is to −1, the higher the degree of positivity.

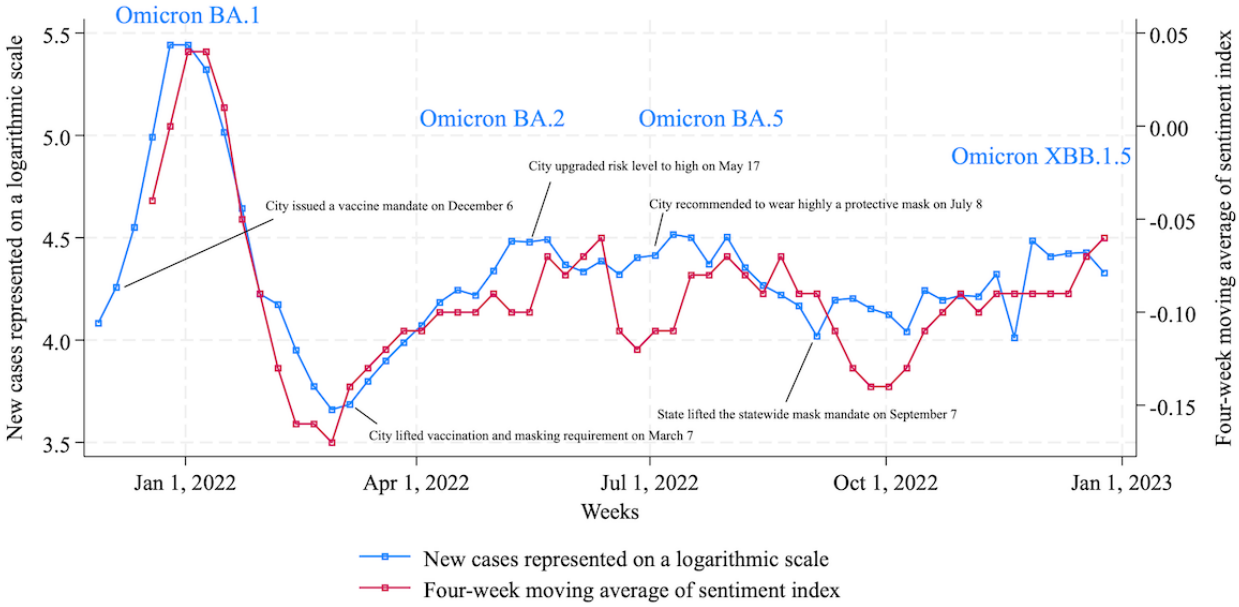

XSL•FO
RenderX

Table 5. The correlation coefficient between new COVID-19 cases and sentiment index extracted by GPT-3.5 Turbo with fine-tuning from December 2021 to December 2022.

	df[a] (cases and index pairs)	Average log cases[b]	Average index[c]	*r*[d] (95% CI[e])	*P* value
New York City					
Total	52	4.3	–0.1	0.89 (0.81 to 0.93)	<.001
Stay-at-home order	52	4.3	0.2	0.69 (0.41 to 0.76)	<.001
Restrictions on gatherings	52	4.3	–0.1	0.40 (0.15 to 0.61)	<.001
Travel restrictions	52	4.3	–0.2	0.73 (0.58 to 0.84)	<.001
Los Angeles					
Total	52	4.3	–0.1	0.39 (0.14 to 0.60)	<.001
Stay-at-home order	52	4.3	0.2	0.41 (0.17 to 0.61)	<.001
Restrictions on gatherings	52	4.3	−0.1	0.13 (−0.15 to 0.38)	<.001
Travel restrictions	52	4.3	–0.3	0.52 (0.29 to 0.69)	<.001
Chicago					
Total	52	4.0	–0.1	0.65 (0.47 to 0.78)	<.001
Stay-at-home order	52	4.0	0.1	0.63 (0.43 to 0.77)	<.001
Restrictions on gatherings	52	4.0	–0.1	0.42 (0.18 to 0.62)	<.001
Travel restrictions	52	4.0	–0.3	0.48 (0.24 to 0.66)	<.001

[a]*df* calculated as n – 2, where n is the number of paired observations.

[b]Avg log cases: the average of the logarithmically transformed number of new COVID-19 cases that were summarized on a weekly basis.

[c]Avg index: the average sentiment index derived from GPT-3.5 Turbo with fine-tuning that was summarized on a weekly basis.

[d]The correlation coefficient between average log cases and average index.

[e]95% CI for the correlation coefficient.

XSL•FO
RenderX

Table 6. Feature words extracted by term frequency–inverse document frequency in New York City, Los Angeles, and Chicago.

Period	Trend[a]	Feature words
New York City		
December 2021 to January 2022	Rise[b]	"nye", "santa", "*tests*", "*cases*", "*pcr*", "xmas", "eve", "playmates", "*omicron*", "citymd"
January 2022 to March 2022	Fall[c]	"adon", "magazine", "partnership", "snow", "playmates", "jorge", "cric", "holly", "ostine", "har"
March 2022	Rise	"adon", "magazine", "partnership", "mfa", "thesis", "easter", "ukraine", "ecea", "abortion", "ladies"
May 2022 to June 2022	Fall	"gelato", "abortion", "juneteenth", "britney", "Imran", "kele", "sonny", "ceremony", "conference"
June 2022 to August 2022	Rise	"taiwan", "pelosi", "gelato", "abortion", "gyamfua", "teresa", "dj", "speaker", "supplies", "nancy"
August 2022 to October 2022	Fall	"supplies", "hurricane", "marcos", "dj", "princess", "backpacks", "zuccotti", "sp", "sept", "depot"
October 2022 to December 2022	Rise	"thanksgiving", "exile", "election", "pt", "santa", "comics", "thankful", "profile", "mets", "feet"
Los Angeles		
December 2021 to January 2022	Rise	"shooky", "*omicron*", "mang", "xmas", "rain", "nye", "*tested*", "eve", "*vaccinated*", "pickup"
January 2022 to February 2022	Fall	"meye", "spotify", "launch", "vibration", "return", "colors", "reach", "train", "katelyn"
February 2022 to March 2022	Rise	"drumz", "ikeboy", "corey", "chorus", "braves", "platforms", "speed", "putin", "pickup"
March 2022 to April 2022	Fall	"anaheim", "katani", "axie", "mace", "wednesdays", "mombasa", "syokimau", "rd", "gig"
April 2022 to June 2022	Rise	"gun", "train", "demand", "mass", "lots", "easter", "stand", "anaheim", "responders", "dodger"
June 2022 to July 2022	Fall	"kbla", "abortion", "kele", "tumi", "calhope", "inquiries", "usc", "alex", "folks", "lots"
July 2022 to November 2022	Rise	"halloween", "return", "lots", "train", "sinners", "folks", "stand", "sharing", "bday", "sept"
November 2022 to December 2022	Rise	"thanksgiving", "xmas", "return", "folks", "rain", "lots", "picture", "code", "oregon", "stub"
Chicago		
December 2021 to January 2022	Rise	"christmas", "*testing*", "*vaccinated*", "*vaxxed*", "winter", "merry", "*tested*", "teachers", "*tests*", "*masks*"
January 2022 to February 2022	Fall	"snow", "polio", "valentines", "n", "mae", "quite", "*vaccine*", "lens", "strap", "rethink"
February 2022 to March 2022	Rise	"ukraine", "creature", "ukrainian", "women", "session", "uppf", "group", "required", "*vaccine*", "*masks*"
March 2022 to May 2022	Fall	"orvieto", "pauline", "uppf", "pagbalik", "ni", "fralaine", "writers", "wines", "nevada", "celebrate"
May 2022 to May 2022	Rise	"nra", "kardashian", "kourtney", "send", "lee", "government", "tragedy", "travis", "robb", "families"
May 2022 to June 2022	Fall	"national", "women", "final", "logical", "bush", "solution", "event", "tourists", "crops", "alexander"
June 2022 to July 2022	Rise	"women", "abortion", "highland", "harris", "tamil", "celebrate", "tornado", "session", "send"
July 2022 to August 2022	Fall	"bash", "supplies", "finkl", "saudi", "group", "biden", "bag", "faculty", "lee", "jaz"
August 2022 to October 2022	Rise	"bark", "tickets", "birthday", "dance", "dress", "event", "n", "hc", "downtown", "celebrate"
October 2022 to November 2022	Fall	"halloween", "dailey", "juniors", "sign", "visiting", "encouraged", "hermanos", "fascism", "attend"
November 2022 to December 2022	Rise	"christmas", "thanksgiving", "santa", "winter", "merry", "snow", "tickets", "gift", "storm"

[a]Trend denotes the current infection status of COVID-19. "Rise" signifies the period of infection expansion, and "Fall" indicates the decline or conclusion of the infection period.

[b]Rise signifies the period of infection expansion.

[c]Fall indicates the decline or conclusion of the infection period. Italicized texts in the Feature words column represent words relevant to COVID-19.

Los Angeles

The left axis of Figure 3 shows the number of cases represented by a logarithmic scale in Los Angeles County, California, and the right axis shows the average sentiment over 4 weeks extracted by GPT-3.5 Turbo fine-tuned with training data. Similar to New York City, Los Angeles experienced an increase in Omicron infections in December 2021 [46], followed by subvariant BA.2 infections in spring 2022 [47] and subvariant BA.5 infections in summer 2022 [48]. Unlike in New York City, cases caused by the Omicron subvariant XBB.1.5 were limited in Los Angeles in December 2022 [49]. During this

XSL•FO
RenderX

period, as in New York City, the state government did not order any restrictions on public movements, mainly requiring masks in public indoor settings [50] and recommending vaccinations for health care workers [51].

Here, we examine the correlation between sentiment and cases from Table 5. For the total sentiment index and cases pairs, *r* is 0.39, which is lower than that in New York City (*r*=0.89). However, by setting a 2-week lag in cases in the correlation coefficient, it was confirmed that *r* was 0.61 (*P*<.001). From this, it can be confirmed that in Los Angeles, sentiment preceded infection status. Similar to New York City, there was a relatively high correlation (0.52) between posts related to travel restriction and infection status and a low correlation (0.13) between posts related to restrictions on gatherings and infection status. In addition, as shown in Figure 3, it is notable that although the number of cases decreased since fall 2022, sentiment did not move positively as it did in spring 2022.

Table 6 shows the feature words extracted for each period according to infection status. In late December 2021 when infections caused by the Omicron strain rose, words related to Omicron and vaccination became common. However, from January 2, 2022, words related to COVID-19 dropped from the top of the TF-IDF scores, following the same trend observed in New York City. This means that although subvariants such as BA.2 and BA.5 appeared after January 2022, public interest in COVID-19 had waned.

Figure 3. New COVID-19 cases and sentiment index extracted by GPT-3.5 Turbo with fine-tuning in Los Angeles.

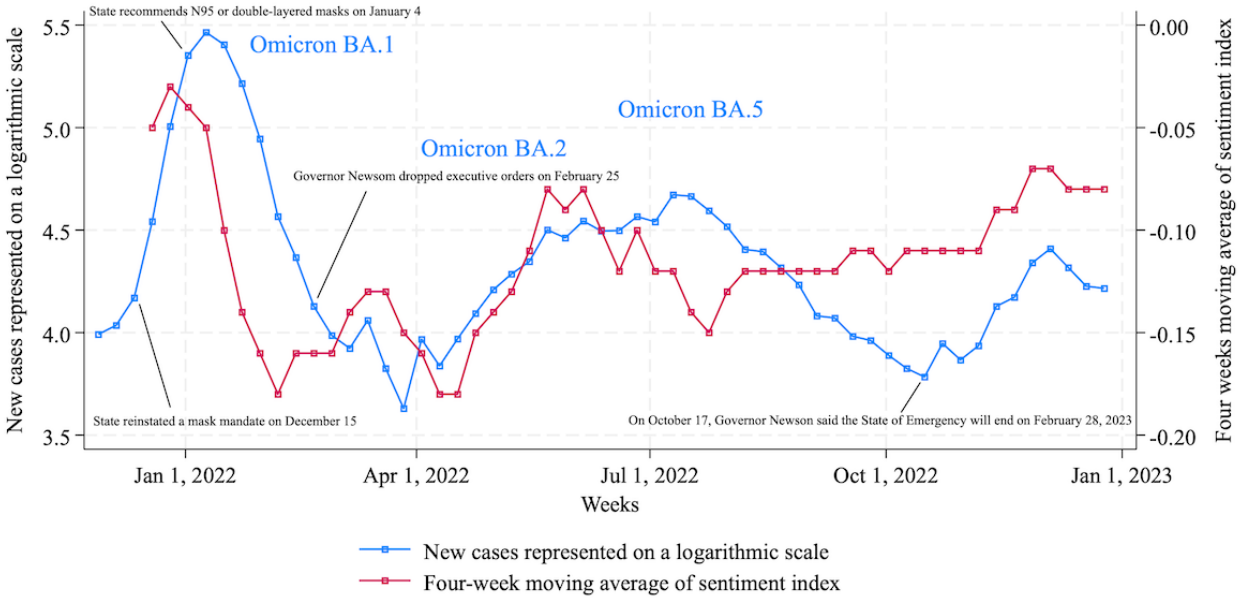

Chicago

The left axis of Figure 4 shows the number of cases represented by a logarithmic scale in Cook County, Illinois, and the right axis shows the average sentiment over 4 weeks extracted by GPT-3.5 Turbo fine-tuned with training data. In Chicago as well, after the surge of Omicron infection at the end of 2021 [52], subvariant BA.2 became dominant in spring [53], and BA.5 replaced it in summer [54]. In Illinois, from December 2021 to 2022, vaccination and mask mandates were intermittently issued for specific situations and workers, but as in New York City and Los Angeles, restrictions on movements were also imposed.

As shown in Table 5, the correlation coefficient between cases and sentiment in Chicago was 0.65 throughout the period, which is lower than that in New York City (0.89) and higher than that in Los Angeles (0.39). Regarding the correlation between cases and sentiment, unlike New York City and Los Angeles, there was a relatively higher correlation (0.63) between sentiment related to stay-at-home orders and cases. In Chicago, concerns about going to school or work may have been higher than concerns about social activities and travel.

Table 6 shows the feature words extracted for each period according to infection status. Unlike New York City and Los Angeles, we can see that keywords related to COVID-19 were common until March 2022. In Chicago, the indoor mask mandate was lifted [55], and proof of vaccination in public places was repealed on February 28 [56], but Chicagoans may have had a more sustained interest in wearing masks and getting vaccinated compared with residents in the other 2 cities.

XSL•FO
RenderX

Figure 4. New COVID-19 cases and sentiment index extracted by GPT-3.5 Turbo with fine-tuning in Chicago.

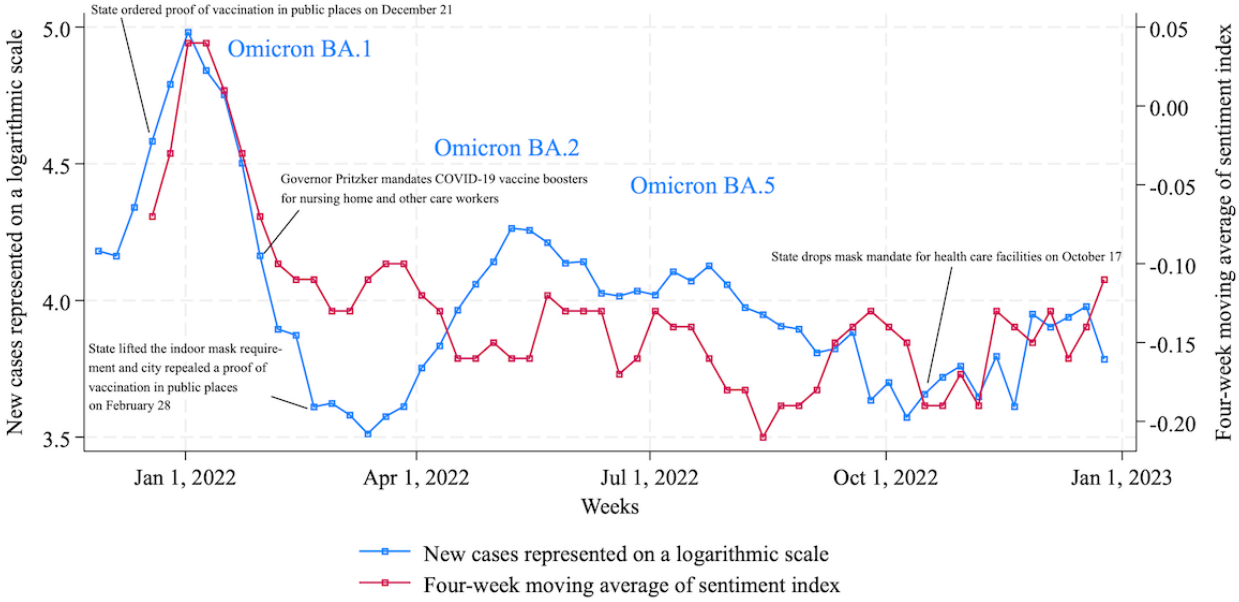

Discussion

Principal Findings

Considering the findings presented in the Results section, here we address RQ3 and RQ4. First, we undertake RQ3: "To what extent will the cyclical sentiment of citizens toward restricted activities in major US cities during the 'new normal' period weaken over time following its peak in December 2021?." Our results indicate a decrease in negative sentiment starting in April 2022, as evidenced by the peak of negative sentiment being below –0.05 in New York City (Figure 2) and below –0.1 in Chicago (Figure 4), despite the emergence of additional variants such as Omicron BA.2 and BA.5. In contrast to New York City and Chicago, sentiment in Los Angeles remained consistently within 75% of the initial peak observed during the week of December 26, 2021, maintaining this level from August 2022 onward (Figure 3). This phenomenon can be attributed to the relatively modest peak value (–0.03) for Los Angeles compared with the peak values of 0.03 in New York City and 0.04 in Chicago during the weeks of January 2 and January 9, 2022. Overall, we confirmed that the tone of the sentiment toward restricted social activities changed from negative to moderate throughout 2022, a trend observed consistently across these 3 US cities. According to the investigation about prolonged pandemic symptoms, significantly lower levels of limiting one's activities were reported among adults aged 18 to 39 years in the United States in June 2022 compared with other age groups [57]. This age group overlapped with the Twitter user group [58], partially supporting the sentiment estimation results.

Second, we address RQ4: "In comparing citizens' sentiments across major metropolitan areas in the United States, namely, New York City, Los Angeles, and Chicago, between December 2021 and December 2022, what differences or similarities in sentiment were there?" In terms of sentiment, the correlation coefficient between New York City and Los Angeles was 0.64, between Los Angeles and Chicago was 0.32, and between Chicago and New York City was 0.60, indicating an overall positive correlation. In feature words extracted by TF-IDF, during the peak of the initial Omicron wave in New York City and Los Angeles, feature words associated with COVID-19 were common; however, they were notably absent in subsequent periods. Meanwhile, in Chicago, feature words related to COVID-19 remained prominent in Twitter's message space until March 2022, indicating sustained interest in COVID-19. From then on, COVID-19–related keywords in each city were replaced by words from political contexts such as "Ukraine" and "abortion," weather-related words such as "hurricane" and "rain," and other words with unknown contexts. The abovementioned results are consistent with the trend of weakening sentiment throughout 2022.

Third, we consider the methodology. A novel aspect of our study is the accuracy of the sentiment estimation model, which was developed using Twitter posts extracted under the same conditions as the observed message group focused on COVID-19 awareness among people in 2021. Previous studies relied on data from before 2020, when COVID-19 emerged in the sentiment classification model. In such cases, a domain shift problem is likely to occur between the registered lexicon or training data and the observed data, which would affect the accuracy of the results. This problem remains an issue to be addressed, from conventional neural networks to LLMs, and technical solutions are being discussed [59,60]. We overcame these issues by training the model explicitly on data from the COVID-19 period.

Comparison With Prior Work

We compare our results with those of previous studies. The novelty of this study is that it showed a positive correlation over time between sentiment and cases during the "new normal" period. In a similar study, a survey in the Greater London area delved into the recovery of sentiment at the city level during the COVID-19 pandemic and found a gradual recovery in sentiment over time after reopening, similar to the findings of

XSL•FO
RenderX

this study [61]. However, this study merely confined the visualization of sentiment estimation results to their association with reopening events and did not reveal a correlation with any variables related to COVID-19. In the prior study, sentiment transitioned positively with a delay after reopening events, whereas in our study, sentiment shifted positively concurrently with the issuance of reopening orders. The difference in these results is thought to stem from whether the observed messages associated with keywords related mainly to COVID-19 or to social activities restricted by the pandemic.

Next, we refer to the previous study conducted by the author [16], which explored sentiment in New York City, Los Angeles, and Chicago from December 2019 to December 2021, using data collected under the same conditions as in this study. The sentiment waveforms correlated with infection status show the same trend, but the crucial difference is the tendency of feature words related to COVID-19 to appear. Both papers show that feature words related to COVID-19 had high TF-IDF values in New York City and Los Angeles from March 2020 to January 2022 and in Chicago from March 2020 to March 2022. In contrast, after the aforementioned period, feature words associated with COVID-19 did not surface among the top rankings in these 3 cities. In the context of COVID-19, it is plausible that Twitter's message space reverted to a state akin to that of February 2020 after the spring of 2022.

Limitations

It is important to acknowledge certain limitations of our study. First, Twitter users in the United States have tended to skew toward the 18- to 39-year age group, which constitutes 62% of the user base [58]. Moreover, there was a notable gender bias, with males accounting for 63% of the total users [62]. It is essential to recognize that these demographics may not perfectly align with the overall population distribution in the United States. Consequently, our study's observed data may exhibit a slight bias toward younger male demographics. Second, in this study, we estimated citizen sentiment by focusing on activities that were restricted during the COVID-19 pandemic, but these may not necessarily be messages posted in the context of the pandemic. As the goal of this study was to capture the societal atmosphere during the "new normal" period, we did not intend to limit our observations to the context of the COVID-19 pandemic. However, it is important to keep this context in mind as a premise.

Conclusions

This study estimated citizens' sentiment in major cities in the United States, spanning the period from the exponential surge in COVID-19 infections to the gradual abatement of the pandemic as it approached its conclusion. This research makes 3 primary contributions. First, it enhances the performance of the sentiment classifier by fine-tuning the LLM using data under the same conditions as the observed target. The efficacy of this improvement was validated through a comparison with previous models using F_1-score metrics. Second, our findings revealed a positive correlation between the estimated sentiment of citizens and actual cases of COVID-19 in New York City, Los Angeles, and Chicago. Although variations exist among these cities, a consistent trend emerged showing a gradual decline in sentiment, which corresponded to a reduction in the number of infections. Third, across these cities in 2022, references to COVID-19 gradually disappeared in the context of restricted social activities as the pandemic neared its end. This phenomenon is evidenced by the disappearance of COVID-19–related words in feature words after the spring of 2022.

The observational data design and sentiment classifier creation method used in this study have potential applications beyond the context of the COVID-19 pandemic. They can be adapted to address various other crises that humanity encounters intermittently, including infectious diseases, natural disasters, terrorism, and warfare. In the future, this field of study is expected to evolve and be applied more broadly to society, supported by stronger collaboration between social science and computer science.

Acknowledgments

The authors received no financial support for this study's research, authorship, and publication.

Authors' Contributions

RS contributed to the study's conception, the development of the neural network model, the execution of experiments, and the writing of the manuscript. ST guided the study's direction and conducted a thorough review of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords related to citizens' activities that were constrained by lockdowns and similar measures implemented in response to the COVID-19 pandemic.
[PDF File (Adobe PDF File), 98 KB - jmir_v27i1e63824_app1.pdf]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/europe/emergencies/situations/covid-19 [accessed 2024-06-30]

XSL•FO
RenderX

2. Statement on the fifteenth meeting of the IHR (2005) emergency committee on the COVID-19 pandemic. World Health Organization. URL: https://tinyurl.com/mr3c9854 [accessed 2024-06-30]
3. COVID-19 map. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2024-06-30]
4. Sood SK, Rawat KS, Kumar D. Scientometric analysis of ICT-assisted intelligent control systems response to COVID-19 pandemic. Neural Comput Appl 2023 Jun 27;35(26):18829-18849. [doi: 10.1007/S00521-023-08788-3]
5. Groves RM. Nonresponse rates and nonresponse bias in household surveys. Public Opin Q 2006 Jan 01;70(5):646-675. [doi: 10.1093/poq/nfl033]
6. Yi J, Gina Qu J, Zhang WJ. Depicting the emotion flow: super-spreaders of emotional messages on Weibo during the COVID-19 pandemic. Soc Media Soc 2022 Mar 12;8(1):205630512210849. [doi: 10.1177/20563051221084950]
7. Zheng P, Adams PC, Wang J. Shifting moods on Sina Weibo: the first 12 weeks of COVID-19 in Wuhan. New Media Soc 2021 Nov 27;26(1):346-367. [doi: 10.1177/14614448211058850]
8. Hong T, Tang Z, Lu M, Wang Y, Wu J, Wijaya D. Effects of #coronavirus content moderation on misinformation and anti-Asian hate on Instagram. New Media Soc 2023 Aug 04 [FREE Full text] [doi: 10.1177/14614448231187529]
9. Wang J, Fan Y, Palacios J, Chai Y, Guetta-Jeanrenaud N, Obradovich N, et al. Global evidence of expressed sentiment alterations during the COVID-19 pandemic. Nat Hum Behav 2022 Mar;6(3):349-358. [doi: 10.1038/s41562-022-01312-y] [Medline: 35301467]
10. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
11. Jiang X, Su MH, Hwang J, Lian R, Brauer M, Kim S, et al. Polarization over vaccination: ideological differences in Twitter expression about COVID-19 vaccine favorability and specific hesitancy concerns. Soc Media Soc 2021 Sep 30;7(3):205630512110484. [doi: 10.1177/20563051211048413]
12. Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, et al. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. J Med Internet Res 2021 Sep 10;23(9):e30854 [FREE Full text] [doi: 10.2196/30854] [Medline: 34346888]
13. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: comparative study. J Med Internet Res 2022 Oct 17;24(10):e40408 [FREE Full text] [doi: 10.2196/40408] [Medline: 36174192]
14. Niu Q, Liu J, Kato M, Nagai-Tanima M, Aoyama T. The effect of fear of infection and sufficient vaccine reservation information on rapid COVID-19 vaccination in Japan: evidence from a retrospective twitter analysis. J Med Internet Res 2022 Jun 09;24(6):e37466 [FREE Full text] [doi: 10.2196/37466] [Medline: 35649182]
15. Huang Q, Jackson S, Derakhshan S, Lee L, Pham E, Jackson A, et al. Urban-rural differences in COVID-19 exposures and outcomes in the South: a preliminary analysis of South Carolina. PLoS One 2021;16(2):e0246548 [FREE Full text] [doi: 10.1371/journal.pone.0246548] [Medline: 33534870]
16. Saito R, Haruyama S. Estimating time-series changes in social sentiment @Twitter in U.S. metropolises during the COVID-19 pandemic. J Comput Soc Sci 2023;6(1):359-388 [FREE Full text] [doi: 10.1007/s42001-022-00186-4] [Medline: 36405087]
17. Anthes E. A C.D.C. airport surveillance program found the earliest known U.S. cases of Omicron subvariants. The New York Times. URL: https://www.nytimes.com/2022/03/24/health/cdc-us-ba2.html [accessed 2024-04-29]
18. Ben-David S, Blitzer J, Crammer K, Pereira FC. Analysis of representations for domain adaptation. Adv Neural Inf Process Syst 2006:137-144 [FREE Full text] [doi: 10.7551/mitpress/7503.003.0022]
19. Wankhade M, Rao AC, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 2022 Feb 07;55(7):5731-5780. [doi: 10.1007/S10462-022-10144-1]
20. Rowe F, Mahony M, Graells-Garrido E, Rango M, Sievers N. Using Twitter to track immigration sentiment during early stages of the COVID-19 pandemic. Data Policy 2021 Dec 28;3:e36. [doi: 10.1017/dap.2021.38]
21. Yuan X, Schuchard RJ, Crooks AT. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. Soc Media Soc 2019 Sep 04;5(3):205630511986546. [doi: 10.1177/2056305119865465]
22. Agrawal S, Schuster AM, Britt N, Liberman J, Cotten SR. Expendable to essential? Changing perceptions of gig workers on Twitter in the onset of COVID-19. Inf Commun Soc 2021 Dec 31;25(5):634-653. [doi: 10.1080/1369118X.2021.2020323]
23. Vaswani A, Shazeer NM, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst 2017 Jun 12;30:5998-6008 [FREE Full text]
24. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. OpenAI CDN. 2018. URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf [accessed 2024-04-29]
25. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL-HLT '19; June 2-7, 2019; Minneapolis, MN p. 4171-4186 URL: https://aclanthology.org/N19-1423.pdf [doi: 10.18653/v1/2024.naacl-short]

XSL•FO
RenderX

26. Theocharopoulos PC, Tsoukala A, Georgakopoulos SV, Tasoulis SK, Plagianakos VP. Analysing sentiment change detection of COVID-19 tweets. Neural Comput Appl 2023 May 31:1-11 [FREE Full text] [doi: 10.1007/s00521-023-08662-2] [Medline: 37362564]
27. Cresswell K, Tahir A, Sheikh Z, Hussain Z, Domínguez Hernández A, Harrison E, et al. Understanding public perceptions of COVID-19 contact tracing apps: artificial intelligence-enabled social media analysis. J Med Internet Res 2021 May 17;23(5):e26618 [FREE Full text] [doi: 10.2196/26618] [Medline: 33939622]
28. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv Preprint posted online May 28, 2020 [FREE Full text]
29. OpenAI platform. OpenAI. URL: https://platform.openai.com/docs/models/gpt-3-5-turbo [accessed 2024-06-30]
30. Corpuz JC. Adapting to the culture of 'new normal': an emerging response to COVID-19. J Public Health (Oxf) 2021 Jun 07;43(2):e344-e345 [FREE Full text] [doi: 10.1093/pubmed/fdab057] [Medline: 33683346]
31. Graso M. The new normal: COVID-19 risk perceptions and support for continuing restrictions past vaccinations. PLoS One 2022;17(4):e0266602 [FREE Full text] [doi: 10.1371/journal.pone.0266602] [Medline: 35395026]
32. Callaghan T, Lueck JA, Trujillo KL, Ferdinand AO. Rural and urban differences in COVID-19 prevention behaviors. J Rural Health 2021 Mar;37(2):287-295 [FREE Full text] [doi: 10.1111/jrh.12556] [Medline: 33619836]
33. RyuichiSaito1 /COVID 19-Twitter-USA-restoring. GitHub Repository. URL: https://github.com/RyuichiSaito1/covid19-twitter-usa-restoring/ [accessed 2024-04-29]
34. City and town population totals: 2020-2023. US Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html [accessed 2024-06-30]
35. Twitter API. X Developer Platform. URL: https://developer.x.com/en/docs/x-api [accessed 2024-04-29]
36. Jurafsky D, Martin JH. Speech and language processing. Stanford University. URL: https://web.stanford.edu/~jurafsky/slp3/ [accessed 2024-06-30]
37. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
38. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized Bert pretraining approach. arXiv Preprint posted online July 26, 2019 [FREE Full text]
39. COVID-19-data: a repository of data on coronavirus cases and deaths in the U.S. GitHub. URL: https://github.com/nytimes/covid-19-data [accessed 2024-06-30]
40. Soukhovolsky V, Kovalev A, Pitt A, Shulman K, Tarasova O, Kessel B. The Cyclicity of coronavirus cases: "waves" and the "weekend effect". Chaos Solitons Fractals 2021 Mar;144:110718 [FREE Full text] [doi: 10.1016/j.chaos.2021.110718] [Medline: 33531739]
41. Omicron is spreading fast. Can New York do more to slow it down? The New York Times. URL: https://www.nytimes.com/2021/12/22/nyregion/omicron-nyc-spread.html [accessed 2024-02-11]
42. As yet another wave of COVID looms, New Yorkers ask: should i worry? The New York Times. URL: https://www.nytimes.com/2022/04/07/nyregion/omicron-variant-ba2.html [accessed 2024-02-11]
43. Health experts warily eye XBB.1.5, the latest Omicron subvariant. The New York Times. URL: https://www.nytimes.com/2023/01/07/science/covid-omicron-variants-xbb.html [accessed 2024-02-11]
44. Mayor de Blasio announces vaccine mandate for private-sector workers, and major expansions to nation-leading “key to NYC” program. The City of New York. URL: https://www.nyc.gov/office-of-the-mayor/news/807-21/mayor-de-blasio-vaccine-mandate-private-sector-workers-major-expansions-to [accessed 2024-02-11]
45. Health officials urge New Yorkers to wear masks again and visit test to treat clinics as COVID cases rise. CBS Broadcasting. URL: https://www.cbsnews.com/newyork/news/health-officials-urge-new-yorkers-to-wear-masks-again-as-covid-cases-rise-test-to-treat/ [accessed 2024-02-11]
46. Coronavirus cases soaring in L.A. county. Los Angeles Times. URL: https://www.latimes.com/california/story/2021-12-23/coronavirus-cases-soaring-in-l-a-county [accessed 2024-02-11]
47. California could see coronavirus uptick with BA.2 subvariant. Los Angeles Times. URL: https://www.latimes.com/california/story/2022-03-29/spring-omicron-ba-2-wave-is-likely-but-how-big-will-it-be [accessed 2024-02-11]
48. Stunning spread of BA.5 shows why this California COVID wave is so different. Los Angeles Times. URL: https://www.latimes.com/california/story/2022-07-12/omicron-ba5-coronaviru [accessed 2024-02-12]
49. The XBB.1.5 variant is taking over on the East Coast. Will it happen in California too? Los Angeles Times. URL: https://www.latimes.com/science/story/2023-01-05/xbb-1-5-variant-is-taking-over-the-east-coast-will-it-happen-in-california [accessed 2024-02-12]
50. CDPH requires masking for all public indoor settings to slow the spread of COVID-19 in response to increasing case rates and hospitalization. California Department of Public Health. URL: https://www.cdph.ca.gov/Programs/OPA/pages/nr21-352.aspx [accessed 2024-02-12]
51. State public health officer order of March 3, 2023. California Department of Public Health. URL: https://www.cdph.ca.gov/Programs/CID/DCDC/pages/covid-19/order-of-the-state-public-health-officer-health-care-worker-vaccine-requirement.aspx [accessed 2024-02-12]

XSL•FO
RenderX

52. Omicron creates ‘fifth wave’ Of COVID-19 in Chicago, with officials urging people to get tested before holidays. Block Club Chicago. URL: https://blockclubchicago.org/2021/12/22/omicron-creates-fifth-wave-of-covid-19-in-chicago-with-officials-urging-people-to-get-tested-before-holidays/ [accessed 2024-02-12]
53. Sub-lineages of Omicron in Chicago. Chicago Department of Public Health. URL: https://www.chicago.gov/content/dam/city/sites/covid/reports/2022/Omicron_sub-lineages_Data_Brief_May-2022.pdf [accessed 2024-02-12]
54. ‘The virus not done with us yet': health officials warn of reinfection as BA.5 variant spreads. ABC Eyewitness News. URL: https://abc7chicago.com/covid-19-ba5-variant/12049925/ [accessed 2024-02-12]
55. Illinois indoor masking requirement to end Monday, February 28, 2022. State of Illinois Government. URL: https://www.illinois.gov/news/press-release.24545.html [accessed 2024-02-12]
56. Public health order no. 2021-2 - proof of vaccination in public places (fourth amended and re-issued). City of Chicago Government. URL: https://www.chicago.gov/city/en/sites/covid-19/home/health-orders.html [accessed 2024-02-12]
57. Ford ND, Slaughter D, Edwards D, Dalton A, Perrine C, Vahratian A, et al. Long COVID and significant activity limitation among adults, by age - United States, June 1-13, 2022, to June 7-19, 2023. MMWR Morb Mortal Wkly Rep 2023 Aug 11;72(32):866-870 [FREE Full text] [doi: 10.15585/mmwr.mm7232a3] [Medline: 37561665]
58. U.S. social network users 2023, by age group. Statista. URL: https://www.statista.com/statistics/1337525/us-distribution-leading-social-media-platforms-by-age-group/ [accessed 2024-06-30]
59. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. J Mach Learn Res 2016;17:1-35 [FREE Full text] [doi: 10.1007/978-3-319-58347-1_10]
60. Eisenschlos JM, Cole JR, Liu F, Cohen WW. WinoDict: probing language models for in-context word acquisition. In: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. 2023 Presented at: EACL '23; May 2-6, 2023; Dubrovnik, Croatia p. 94-102 URL: https://aclanthology.org/2023.eacl-main.7.pdf [doi: 10.18653/v1/2023.eacl-main.7]
61. Chen Y, Niu H, Silva EA. The road to recovery: sensing public opinion towards reopening measures with social media data in post-lockdown cities. Cities 2023 Jan;132:104054 [FREE Full text] [doi: 10.1016/j.cities.2022.104054] [Medline: 36345535]
62. X/Twitter user distribution by gender in the U.S. Statista. URL: https://www.statista.com/statistics/678794/united-states-twitter-gender-distribution/ [accessed 2024-06-30]

Abbreviations

API: Application Programming Interface
BERT: Bidirectional Encoder Representations from Transformers
LLM: large language model
MTurk: Mechanical Turk
NLP: natural language processing
RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach
RQ: research question
TF-IDF: term frequency–inverse document frequency

Edited by A Mavragani; submitted 01.07.24; peer-reviewed by H Wang, KS Rawat; comments to author 18.11.24; revised version received 08.12.24; accepted 24.12.24; published 11.02.25.

Please cite as:
Saito R, Tsugawa S
Understanding Citizens’ Response to Social Activities on Twitter in US Metropolises During the COVID-19 Recovery Phase Using a Fine-Tuned Large Language Model: Application of AI
J Med Internet Res 2025;27:e63824
URL: https://www.jmir.org/2025/1/e63824
doi:10.2196/63824
PMID:39932775

©Ryuichi Saito, Sho Tsugawa. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Large Language Models–Supported Thrombectomy Decision-Making in Acute Ischemic Stroke Based on Radiology Reports: Feasibility Qualitative Study

Jonathan Kottlors[1], Dr med; Robert Hahnfeldt[1], MD; Lukas Görtz[1], Dr med; Andra-Iza Iuga[1], Dr med; Philipp Fervers[1], Dr med; Johannes Bremm[1], Dr med; David Zopfs[1], Dr med; Kai R Laukamp[1], Dr med; Oezguer A Onur[2], Prof Dr; Simon Lennartz[1], Dr med; Michael Schönfeld[1], Dr med; David Maintz[1], Prof Dr; Christoph Kabbasch[1], Prof Dr; Thorsten Persigehl[1*], Prof Dr; Marc Schlamann[1*], Prof Dr

[1]Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[2]Department of Neurology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[*]these authors contributed equally

Corresponding Author:
Jonathan Kottlors, Dr med
Institute for Diagnostic and Interventional Radiology
Faculty of Medicine and University Hospital Cologne
University of Cologne
Kerpener Straße 62
Cologne, 50937
Germany
Phone: 49 221 47896063
Email: jonathan.kottlors@uk-koeln.de

Abstract

Background: The latest advancement of artificial intelligence (AI) is generative pretrained transformer large language models (LLMs). They have been trained on massive amounts of text, enabling humanlike and semantical responses to text-based inputs and requests. Foreshadowing numerous possible applications in various fields, the potential of such tools for medical data integration and clinical decision-making is not yet clear.

Objective: In this study, we investigate the potential of LLMs in report-based medical decision-making on the example of acute ischemic stroke (AIS), where clinical and image-based information may indicate an immediate need for mechanical thrombectomy (MT). The purpose was to elucidate the feasibility of integrating radiology report data and other clinical information in the context of therapy decision-making using LLMs.

Methods: A hundred patients with AIS were retrospectively included, for which 50% (50/100) was indicated for MT, whereas the other 50% (50/100) was not. The LLM was provided with the computed tomography report, information on neurological symptoms and onset, and patients' age. The performance of the AI decision-making model was compared with an expert consensus regarding the binary determination of MT indication, for which sensitivity, specificity, and accuracy were calculated.

Results: The AI model had an overall accuracy of 88%, with a specificity of 96% and a sensitivity of 80%. The area under the curve for the report-based MT decision was 0.92.

Conclusions: The LLM achieved promising accuracy in determining the eligibility of patients with AIS for MT based on radiology reports and clinical information. Our results underscore the potential of LLMs for radiological and medical data integration. This investigation should serve as a stimulus for further clinical applications of LLMs, in which this AI should be used as an augmented supporting system for human decision-making.

(J Med Internet Res 2025;27:e48328) doi:10.2196/48328

KEYWORDS

artificial intelligence; radiology; report; large language model; text-based augmented supporting system; mechanical thrombectomy; GPT; stroke; decision-making; thrombectomy; imaging; model; machine learning; ischemia

XSL•FO
RenderX

Introduction

Acute ischemic stroke (AIS) is a major cause of death and disability, requiring effective and efficient therapeutic strategies. To prevent the progression of ischemic damage and improve patient outcomes, timely and accurate diagnosis and treatment are crucial [1-3]. The standard of care and diagnostic workflow for acute AIS involves a combination of clinical information, clinical assessment, imaging, and decision-making tools, such as the National Institutes of Health Stroke Scale (NIHSS) [2,4-6]. Current treatment approaches include the administration of intravenous thrombolytic therapy, as well as mechanical thrombectomy (MT), either as a combined or sole treatment option [5,7,8]. While MT is an invasive procedure aiming for the removal of an occluding thrombus, only patients fulfilling certain requirements are eligible for this specific treatment [9,10]. The indication for MT is established with a combination of clinical assessment, image-based results, and guidelines such as the European Society of Neuroradiology or European Stroke Organization guidelines, taking into account factors such as patient's age, onset time of symptoms, and location and extent of the occlusion [8,10,11].

However, the subjective nature of the clinical assessment and decision-making tools, combined with the growing workload of health care professionals, highlights the need for a supporting, objective, and efficient approach to decision-making [12]. The task of deciding a patient's suitability for MT can be particularly challenging for inexperienced physicians, as it requires a thorough understanding of the latest medical evidence and guidelines, as well as a high degree of clinical judgment [13-15].

Studies have demonstrated the feasibility and potential of using artificial intelligence (AI) in medical decision-making, particularly in the radiology field [16,17]. For example, recent studies show the application of AI in cancer imaging analysis or in detecting acute intracranial hemorrhage on computed tomography (CT) or magnetic resonance imaging scans [18,19]. Similar studies have also shown promising results in the use of AI in the diagnosis of ischemic stroke (IS), with potential improvement of diagnostic accuracy and reduction of variability in the decision-making process. However, these studies all focus on the AI-based processing of visual information [18,20].

As nearly all information can be expressed in human language, recorded in text form in the radiology report, recent advancements in the development of large language models (LLMs) have opened new opportunities for medical information processing [21-24]. One of these autoregressive LLMs is GPT-3 used within the pretrained chatbot application version, named ChatGPT, developed by OpenAI [25]. GPT-3 is a third-generation deep learning model that has been trained on a massive amount of text data on 175 billion parameters, allowing it to generate humanlike logical and semantical responses to text-based questions and input information [26,27]. In an experimental setting, a GPT-3–based LMM model shows a significant improvement in medical question-answering tasks and showed a passing score in a medical licensing examination, due to its ability to provide logical and informational context across most answers [28]. To our knowledge, no large language processing model has been evaluated for immediate clinical decision-making based on radiology report texts. This could carry several benefits, including increasing the objectivity, speed, and standardization of decisions, as well as assisting physicians and medical staff by highlighting or reminding them of potential courses of action based on the latest medical evidence and guidelines.

In the context of AIS diagnosis and treatment, the application of such an LLM to a radiology report with a resulting decision-making component related to treatment using MT represents an exemplary use case. In this way, possible AI-assisted decisions related to MT could be made faster and more standardized based on up-to-date medical evidence and guidelines, which can be checked and correlated to the radiology report. Additionally, the application of an LLM as a monitoring, AI-augmented background system, supporting inexperienced or younger radiologists by highlighting the potential MT indication based on the written report, is conceivable. Thus, this study aims to evaluate the accuracy and efficiency of LLMs in predicting the suitability of patients for MT and to compare its decisions with those made in consensus by experienced health care professionals.

Methods

Ethical Considerations

This retrospective study received human subject ethics review approvals (23-1061-retro). Informed consent was waived due to the retrospective design of this investigation including the ability of participants to opt out. Privacy and confidentiality were ensured—beyond the patient's age and sex, no identifiable features about the patient were transmitted to the LLM model; especially, no patient-identifying information was provided to the AI.

Patient Population

The inclusion criteria for this study were (1) clinical symptoms consistent with AIS; (2) evaluable CT scan according to in-house stroke protocol and board-certified radiology report; (3) availability of the NIHSS score; (4) presentation to the hospital within a potential time of treatment options including MT; (5) for group affiliation, the decision to perform MT was evaluated based on an interdisciplinary expert consensus immediately before MT in each case and classified into (a) indication of MT and (b) no indication of MT; and (6) aged 18 years or older.

The exclusion criteria for this study were (1) patients in whom the juristic decision-making authority or written intention potentially influences the decision to undergo MT and (2) significant comorbidities that would limit the ability to undergo MT or the ability to provide informed consent.

Workflow

The Institute for Diagnostic and Interventional Radiology University Hospital Cologne performs about 250 intracranial MT per year. We retrospectively screened the database according to the abovementioned inclusion criteria in the period between January 2020 to December 2022 and randomly selected a total of 100 patients with AIS, whereby 50% (50/100) patients

XSL•FO
RenderX

have an MT indication and 50% (50/100) patients have no MT indication. All patients were eligible for a priori consideration of MT, the NIHSS was assessed before CT imaging, and neurological symptoms were recorded. The onset of neurological symptoms was also recorded based on case history, including unknown time frames. The image information, in the form of the radiology report, was transmitted to the AI, together with additional information on neurological symptoms, symptoms onset, and patients' age, but the final decision regarding MT was not transmitted to the AI (Figure 1). Based on 20% (20/100) of the cases, we performed a test re-evaluation by the LLM.

Figure 1. A hundred patients with AIS were analyzed. The patients were divided into two groups: 50% (50/100) had an indication for MT and the other 50% (50/100) had no MT indication. The image information, in the form of the radiology report, was transmitted to the AI, together with additional information on NIHSS, neurological symptoms, symptoms onset, and patients' age. However, the final decision regarding MT was not transmitted to the AI. The text-based AI was asked whether there was an indication for MT. The AI decision was compared with the decision of the health care professional's. AI: artificial intelligence; AIS: acute ischemic stroke; IS: ischemic stroke; MT: mechanical thrombectomy; NIHSS: National Institutes of Health Stroke Scale.

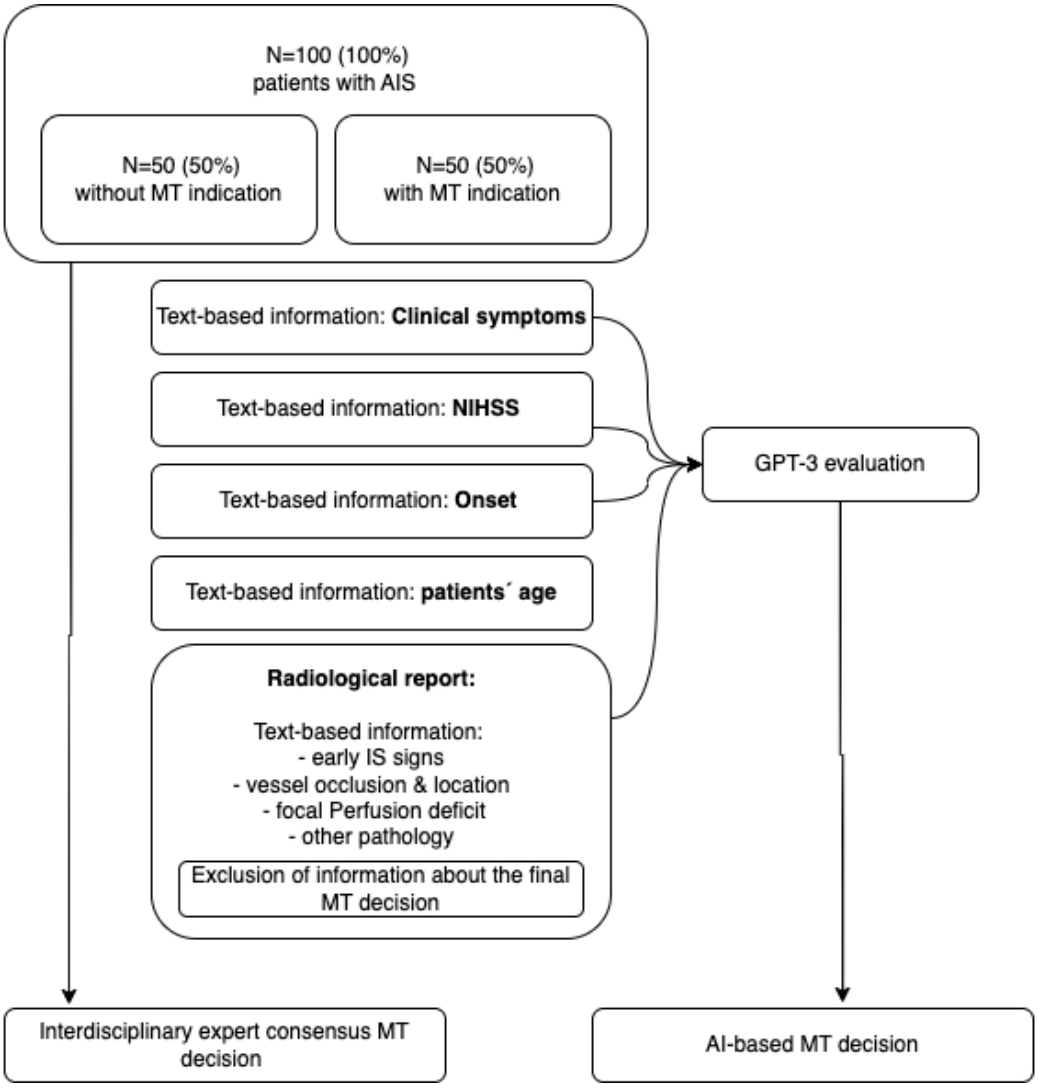

Neurological Assessment

A comprehensive neurological assessment was performed on all patients by a trained neurologist before the radiological examination. This assessment included the documentation of the NIHSS score including neurological symptoms and their onset, which was then communicated as part of the patient's medical history included in the request for radiologic imaging.

Imaging Protocol

All examinations used 128-254 slice CT scanners. Patients were scanned when supine from head to toe. The protocol included an unenhanced neurocranium CT and arterial phase imaging of cervical and intracranial vessels after injecting 60 mL of contrast media (Accupaque 350 mg/mL) at 4 mL/s, followed by 20 mL saline. Bolus tracking activated a 3-second scan delay after reaching 150 Hounsfield units in the descending thoracic aorta. The settings were pitch 1, 120 kVp, and 512 matrix with the DoseRight 3D-DOM (Philips Healthcare). Images were reconstructed axially with a 0.67-mm slice thickness. In some cases, CT perfusion was carried out with an additional 40 mL contrast agent and 20 mL saline using an 80-kVp, 512-matrix DoseRight 3D-DOM and processed with Philips IntelliSpace software (Philips Healthcare).

XSL•FO
RenderX

Image Analysis

In the study, 12 assistant radiologists (2-4 years' experience in neuroradiological imaging) were primarily responsible for conducting the primary image analysis. The radiologic findings were then secondarily reviewed and evaluated in an interdisciplinary consensus with neuroradiologists and neurologists, who had an average experience of over 5 years in neuroradiological imaging. The final decision regarding MT was reached based on this interdisciplinary consensus.

Radiology Report

The text information in the form of a narrative, text-based radiology report in the German language included various imaging findings related to the presence of an AIS. This information was recorded as a prose text. The radiological findings were transmitted to the LLM together with the patient's age and respective NIHSS score, neurologic symptomatology, and the corresponding time windows (time between onset of symptomatology and time of CT scan). Any information on the final MT decision was not included. No other personal data were transmitted to the LLM during the process. The average word count per report was 186 (SD 14; range 113-278).

LLM-Based Decision-Making

GPT-3 is a commercial-sector, pretrained, state-of-the-art, language-processing AI model developed by OpenAI, which is known for its ability to assess text information and has been seen as a significant advancement in natural language processing technology [27]. It has been trained on a massive amount of text data obtained from the internet and has been fine-tuned for various language tasks such as text completion, translation, and question answering. The used LLM model is specifically designed for conversational AI applications, such as virtual assistants [27].

In this study, GPT-3 in the form of ChatGPT was used for information processing and decision-making tasks. The text-based information received was used to generate a decision response. As described above, the information from the radiology report as well as clinical information and patients' ages were provided to the LLM, and the following request was made to it to provide a decision [29]:

> *Please decide whether mechanical thrombectomy is indicated or not, based on the following data. Please answer yes or no.*

Statistical Analysis

Statistical data analysis was performed using R (version 3.6.2; R Foundation for Statistical Computing) on RStudio (version 1.2.5033.35; R Foundation). Figures were plotted using the *ggplot2* package. Continuous variables were reported as mean and SD. Demographic characteristics were compared using the chi-square test for categorical variables and 2-tailed *t* test for continuous variables. The diagnosis' overall accuracy, specificity, and sensitivity were assessed using contingency tables.

Results

Baseline Characteristics

The demographic data were collected and analyzed for a total of 100 patients with AIS, who were divided into two groups: group 1 (n=50, 50%) received MT treatment, while group 2 (n=50, 50%) did not receive MT treatment. The average age of the participants was 68.34 (SD 6.95) years, with no significant difference between the two groups (*P*=.32). The distribution of sex was 54% (27/50) male in group 1 and 46% (23/50) in group 2 (*P*=.82). There were no significant differences between the two groups according to the demographic data (Table 1).

Table 1. Baseline characteristics of this study's population, presented by treatment group. The table displays the mean age (SD), sex distribution (male and female), and NIHSS[a] score and time from clinical onset to imaging (SD) for two treatment groups. Group 1 received MT[b], while group 2 did not receive MT.

Group	Age (years), mean (SD)	Sex, n (%)		NIHSS score, mean (SD)	Onset to imaging (min)
		Male	Female		
Group 1 (MT; n=50)	67.44 (8.34)	27 (54)	23 (46)	14.28 (2.21)	93 (50[c])
Group 2 (no MT; n=50)	69.23 (5.56)	23 (46)	27 (54)	9.22 (2.71)	129 (113[d])

[a]NIHSS: National Institutes of Health Stroke Scale.

[b]MT: mechanical thrombectomy.

[c]Unknown onset of symptoms in 8 cases.

[d]Unknown onset of symptoms in 14 cases.

Regarding the NIHSS scores, the average score was found to be higher in group 1 (14.28, SD 2.21) than in group 2 (9.22, SD 2.71), suggesting that the patients in group 1 had more severe IS symptoms than those in group 2. The difference between the two groups was statistically significant (*P*=.02; Table 1). In the MT group, the time from the clinical onset of symptoms to CT imaging was on average 93 (SD 50) minutes (unknown onset of symptoms in n=8 cases), while in the no-MT indication group, the time was 129 (SD 113) minutes (unknown onset of symptoms in n=14 cases). There was no significant difference between the two groups (*P*=.051; Table 1).

In both groups, the most common symptoms were paresis and aphasia. In the group with MT indication, 72% (36/50) of patients presented with paresis and 46% (23/50) of patients presented with aphasia. In the group without MT indication, 76% (38/50) of patients presented with paresis and 26% (13/50) of patients presented with aphasia. Anopsia was less frequent,

XSL•FO
RenderX

with only 2% (1/50) of patients in the MT group and 6% (3/50) of patients in the group without MT indication. Loss of consciousness was also less common, with 18% (9/50) of patients in the MT group and 4% (2/50) of patients in the group without MT indication (Table 2).

Table 2. Frequency and percentage of the most common clinical symptoms observed before imaging in two groups.

Group	Paresis, n (%)	Aphasia, n (%)	Anopsia, n (%)	Loss of consciousness, n (%)
Group 1 (MT[a]; n=50)	36 (72)	23 (46)	1 (2)	9 (18)
Group 2 (no MT; n=50)	38 (76)	13 (26)	3 (6)	2 (4)

[a]MT: mechanical thrombectomy.

Performance of AI Decision-Making

The decision-making process of the AI-based model was compared to the interdisciplinary consensus regarding MT. The specificity, sensitivity, and accuracy of the AI-based model were calculated based on the agreement with the expert consensus. The results showed that the AI had a specificity of 0.96, a sensitivity of 0.8, and an accuracy of 0.88 in comparison to the expert consensus (Table 2). The area under the curve for report-based MT decision was calculated to be 0.92, indicating good discrimination between the two treatment options (Table 3). The re-evaluation of 20% (20/50) of the cases showed a 100% (20/20) concordance between the results of the first and second LLM evaluations.

Table 3. Number of true positive, false negative, true negative, and false positive cases for two groups: group 1 (MT[a]) and group 2 (no MT).

Group	True positive, n	False negative, n	True negative, n	False positive, n
Group 1 (MT)	40	10	—[b]	—
Group 2 (no MT)	—	—	48	2

[a]MT: mechanical thrombectomy.

[b]Not applicable.

Discussion

This investigation assesses a new approach applying a text-processing AI model to the decision-making process for MT indication in IS, primarily based on the information from the radiology report. Descriptive information on the radiological findings, along with information on patients' neurological symptoms, symptom onset, and demographic data, were provided to the LLM to evaluate the possibility of MT.

This study compared the AI-based decision-making process with the clinical interdisciplinary consensus for MT. Although the accuracy was good, the specificity of the AI model was lower than the sensitivity, with a specificity of 96% and a sensitivity of 80%. The overall accuracy of the AI-based model was 88%, indicating a good discrimination between the two treatment options, with an area under the curve of 0.92 for the AI report–based MT decision.

One of the strengths of this study is the robust gold truth: the well-defined interdisciplinary performed expert consensus for or against MT. Additionally, all cases were evaluated based on radiographic findings, neurological diagnostic information, as well as demographic data on the patient [10]. The AI model used in this study was not specifically trained for this task of making a clinical decision, being rather a language processing algorithm that considers a large amount of human declarative and conceptual knowledge up until 2021.

However, there are several limitations to this study. First, only text-based information was included in the decision-making process, and descriptive statements can be inaccurate and lack quantitative precision. For example, the adjective "large" can have varying interpretations between individuals, which potentially can lead to misinterpretation of the decision and result in AI decision consequences. Second, as the LLM model only has knowledge up until 2021, newer guidelines and developments in the field are not considered [27]. Third, the basis of a decision made by a model is not deterministically understandable and post hoc traceable, which represents a fundamental "black-box" problem of decisions based on structures of deep neural networks [12]. Moreover, within LLM, there is no possibility to show sources and retrieve information about the source of a decision. Further, and this is a fact that may be attributed to the commercial structure of OpenAI's GPT-3, although the algorithm has been trained on vast amounts of text data, it remains unclear which data have been included or excluded [12,26,27,30]. In particular, the possibility that potentially older outdated information that would not have resulted in an MT indication and has been made available to the algorithm at an unspecified location for training purposes could lead to problems. Another limitation pertains to the fact that the algorithm can only consider general information for the MT decision, without considering individual factors about the individual health care facility or medical personnel. Factors such as (1) availability of material for MT, (2) personnel or time capacities, and (3) personal skills of the interventionist are not part of the MT decision [14]. Experimental settings and individual treatment attempts are not part of a potential LLM-based decision. Another limitation is that only text-based reports from one center in one language were included; further studies should include radiological findings from other centers with diverse linguistic styles and also other languages.

XSL•FO
RenderX

In particular, the described facts that current data and information, as well as the individual skills of the interventionalist, could be considered in the decision-making process may explain the lower sensitivity of the AI-based MT decision in relation to its specificity.

In summary, there are several reasons why AI should not make important medical decisions such as MT decisions on its own. Nonetheless, a decision-making algorithm based on the written radiology report can potentially be used to run in the background and alert to a potential MT indication. A potential application would be in the context of an augmenting system that "monitors" radiology reports of younger or inexperienced physicians in the background and alerts in specific cases so that an experienced colleague can be consulted for the final decision. This integration of augmented text-based AI could be especially useful in night shifts, as it has the potential to shorten the time from diagnosis to MT (door-to-needle time). In summary, this could lead to a sustainable improvement in patient care and apply to other areas beyond interventional neuroradiology.

Future studies should consider incorporating additionally other AI networks, such as deep learning–based image recognition networks, to make decisions based on more quantitative information, rather than just on declarative text information. This could provide more objective information, such as the amount of brain tissue that can be saved, which would help to further refine the decision-making process. In the context of further research, it would be preferable that the algorithm specifies the exact source or guideline on which the decision is based, highlighting the need for stable, finalized, and specifically tuned systems for particular tasks in clinical settings. For clinical use of text-based advanced assist for radiologists, quantitative and qualitative information about the AI training data used should be available. Additionally, future studies should also consider incorporating more recent information and developments in the field, as well as available MT devices and the experience of the treating interventionalist.

In conclusion, this study highlights within a first application the potential of LLMs for integrating text-based radiological information with other clinical data, which may be useful for augmented clinical decision-making. Further research should elucidate the limitations of such systems, particularly regarding the traceability of related training and reference data.

Acknowledgments

This study was partly funded by NUM 2 (funding reference number: 01KX2121) and Else Kröner-Fresenius-Stiftung (grant 2023_EKEA.77).

Conflicts of Interest

DZ is on the speaker's bureau of Philips and a lecturer for Amboss GmbH. SL received authorship and speaker fees from Amboss.

References

1. Saini V, Guada L, Yavagal DR. Global epidemiology of stroke and access to acute ischemic stroke interventions. Neurology 2021;97(20 Suppl 2):S6-S16. [doi: 10.1212/WNL.0000000000012781] [Medline: 34785599]
2. Herpich F, Rincon F. Management of acute ischemic stroke. Crit Care Med 2020;48(11):1654-1663 [FREE Full text] [doi: 10.1097/CCM.0000000000004597] [Medline: 32947473]
3. Feske SK. Ischemic stroke. Am J Med 2021;134(12):1457-1464. [doi: 10.1016/j.amjmed.2021.07.027] [Medline: 34454905]
4. Sandset EC, Anderson CS, Bath PM, Christensen H, Fischer U, Gąsecki D, et al. European Stroke Organisation (ESO) guidelines on blood pressure management in acute ischaemic stroke and intracerebral haemorrhage. Eur Stroke J 2021;6(2):XLVIII-LXXXIX [FREE Full text] [doi: 10.1177/23969873211012133] [Medline: 34780578]
5. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019;50(12):e344-e418 [FREE Full text] [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
6. Brott T, Adams HP, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: a clinical examination scale. Stroke 1989;20(7):864-870. [doi: 10.1161/01.str.20.7.864] [Medline: 2749846]
7. Ganesh A, Fraser JF, Gordon Perue GL, Amin-Hanjani S, Leslie-Mazwi TM, Greenberg SM, et al. Endovascular treatment and thrombolysis for acute ischemic stroke in patients with premorbid disability or dementia: a scientific statement from the American Heart Association/American Stroke Association. Stroke 2022;53(5):e204-e217 [FREE Full text] [doi: 10.1161/STR.0000000000000406] [Medline: 35343235]
8. Berge E, Whiteley W, Audebert H, de Marchis G, Fonseca AC, Padiglioni C, et al. European Stroke Organisation (ESO) guidelines on intravenous thrombolysis for acute ischaemic stroke. Eur Stroke J 2021;6(1):I-LXII [FREE Full text] [doi: 10.1177/2396987321989865] [Medline: 33817340]
9. Goyal M, Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. Lancet 2016;387(10029):1723-1731. [doi: 10.1016/S0140-6736(16)00163-X] [Medline: 26898852]
10. Jadhav AP, Desai SM, Jovin TG. Indications for mechanical thrombectomy for acute ischemic stroke: current guidelines and beyond. Neurology 2021;97(20 Suppl 2):S126-S136. [doi: 10.1212/WNL.0000000000012801] [Medline: 34785611]

XSL•FO
RenderX

11. From the American Association of Neurological Surgeons (AANS), American Society of Neuroradiology (ASNR), CardiovascularInterventional Radiology Society of Europe (CIRSE), Canadian Interventional Radiology Association (CIRA), Congress of Neurological Surgeons (CNS), European Society of Minimally Invasive Neurological Therapy (ESMINT), European Society of Neuroradiology (ESNR), European Stroke Organization (ESO), Society for Cardiovascular AngiographyInterventions (SCAI), Society of Interventional Radiology (SIR), Society of NeuroInterventional Surgery (SNIS),World Stroke Organization (WSO), Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, et al. Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. AJNR Am J Neuroradiol 2018;39(6):E61-E76 [FREE Full text] [doi: 10.3174/ajnr.A5638] [Medline: 29773566]
12. Čartolovni A, Tomičić A, Lazić Mosler E. Ethical, legal, and social considerations of AI-based medical decision-support tools: a scoping review. Int J Med Inform 2022;161:104738. [doi: 10.1016/j.ijmedinf.2022.104738] [Medline: 35299098]
13. Cimflova P, Kappelhof M, Singh N, Kashani N, Ospel JM, McDonough R, et al. Factors influencing thrombectomy decision making for primary medium vessel occlusion stroke. J Neurointerv Surg 2022;14(4):350-355. [doi: 10.1136/neurintsurg-2021-017472] [Medline: 33947769]
14. Kashani N, Ospel J, Saposnik G, Goyal M, Menon BK, Almekhlafi MA. Geographical variations in the decision to proceed with endovascular treatment under "Current" and "Ideal" practice settings across regions of the world: results from an international multidisciplinary study. 2019 Presented at: Society of NeuroInterventional Surgery Meeting; July 22-26, 2024; Miami, FL p. 3. [doi: 10.26226/morressier.5cbf25c0469f6e000f9affad]
15. Ospel JM, Singh R, Kashani N, Almekhlafi M, Wilson A, Fischer U, et al. Endovascular treatment decision making in patients with low baseline ASPECTS: insights from UNMASK EVT, an international multidisciplinary study. J Stroke Cerebrovasc Dis 2020;29(12):105411. [doi: 10.1016/j.jstrokecerebrovasdis.2020.105411] [Medline: 33254375]
16. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 2021;31(6):3797-3804 [FREE Full text] [doi: 10.1007/s00330-021-07892-z] [Medline: 33856519]
17. Brady AP, Neri E. Artificial intelligence in radiology-ethical considerations. Diagnostics (Basel) 2020;10(4):231 [FREE Full text] [doi: 10.3390/diagnostics10040231] [Medline: 32316503]
18. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol 2022;19(2):132-146 [FREE Full text] [doi: 10.1038/s41571-021-00560-7] [Medline: 34663898]
19. Matsoukas S, Scaggiante J, Schuldt BR, Smith CJ, Chennareddy S, Kalagara R, et al. Accuracy of artificial intelligence for the detection of intracranial hemorrhage and chronic cerebral microbleeds: a systematic review and pooled analysis. Radiol Med 2022;127(10):1106-1123. [doi: 10.1007/s11547-022-01530-4] [Medline: 35962888]
20. Clèrigues A, Valverde S, Bernal J, Freixenet J, Oliver A, Lladó X. Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. Comput Biol Med 2019;115:103487 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.103487] [Medline: 31629272]
21. Biswas S. ChatGPT and the future of medical writing. Radiology 2023;307(2):e223312. [doi: 10.1148/radiol.223312] [Medline: 36728748]
22. The Lancet Digital Health. ChatGPT: friend or foe? Lancet Digit Health 2023;5(3):e102 [FREE Full text] [doi: 10.1016/S2589-7500(23)00023-7] [Medline: 36754723]
23. Cornacchia S, Errico R, Balzano RF, Fusco V, Maldera A, Pierpaoli E, et al. Medical radiological procedures: which information would be chosen for the report? Radiol Med 2019;124(8):783-793. [doi: 10.1007/s11547-019-01032-w] [Medline: 30972532]
24. Mañas-García A, González-Valverde I, Camacho-Ramos E, Alberich-Bayarri A, Maldonado JA, Marcos M, et al. Radiological structured report integrated with quantitative imaging biomarkers and qualitative scoring systems. J Digit Imaging 2022;35(3):396-407 [FREE Full text] [doi: 10.1007/s10278-022-00589-9] [Medline: 35106674]
25. Schubert L, Petrov M, Carter S, Cammarata N, Goh G, Olah C. OpenAI Microscope. OpenAI. URL: https://openai.com/index/microscope/ [accessed 2024-12-27]
26. Floridi L, Chiriatti M. GPT-3: its nature, scope, limits, and consequences. Minds Mach 2020;30(2):681-694. [doi: 10.1007/s11023-020-09548-1]
27. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. 2020 Presented at: NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems; December 6-12, 2020; Vancouver, BC p. 1877-1901 URL: https://dl.acm.org/doi/abs/10.5555/3495724.3495883
28. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
29. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25:e50638 [FREE Full text] [doi: 10.2196/50638] [Medline: 37792434]
30. Ji J, Hou Y, Chen X, Pan Y, Xiang Y. Vision-language model for generating textual descriptions from clinical images: model development and validation study. JMIR Form Res 2024;8:e32690 [FREE Full text] [doi: 10.2196/32690] [Medline: 38329788]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
AIS: acute ischemic stroke
CT: computed tomography
IS: ischemic stroke
LLM: large language model
MT: mechanical thrombectomy
NIHSS: National Institutes of Health Stroke Scale

Edited by Q Jin; submitted 19.04.23; peer-reviewed by BS Kang, C Ehrengut; comments to author 26.02.24; revised version received 30.06.24; accepted 18.07.24; published 13.02.25.

Please cite as:
Kottlors J, Hahnfeldt R, Görtz L, Iuga AI, Fervers P, Bremm J, Zopfs D, Laukamp KR, Onur OA, Lennartz S, Schönfeld M, Maintz D, Kabbasch C, Persigehl T, Schlamann M
Large Language Models–Supported Thrombectomy Decision-Making in Acute Ischemic Stroke Based on Radiology Reports: Feasibility Qualitative Study
J Med Internet Res 2025;27:e48328
URL: https://www.jmir.org/2025/1/e48328
doi:10.2196/48328
PMID:39946168

©Jonathan Kottlors, Robert Hahnfeldt, Lukas Görtz, Andra-Iza Iuga, Philipp Fervers, Johannes Bremm, David Zopfs, Kai R Laukamp, Oezguer A Onur, Simon Lennartz, Michael Schönfeld, David Maintz, Christoph Kabbasch, Thorsten Persigehl, Marc Schlamann. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Artificial Intelligence in Lymphoma Histopathology: Systematic Review

Yao Fu[1], MMedSc; Zongyao Huang[2], MD; Xudong Deng[3], BSc; Linna Xu[2], MMedSc; Yang Liu[2], MD; Mingxing Zhang[2], BMed; Jinyi Liu[4], BMed; Bin Huang[2], MMedSc

[1]Sichuan Tianfu New Area People's Hospital, Chengdu, China

[2]Department of Pathology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu, China

[3]Wonders Information Co., Ltd, Shanghai, China

[4]Phase I Clinical Trial Unit, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu, China

Corresponding Author:
Bin Huang, MMedSc
Department of Pathology
Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute
Sichuan Cancer Center, University of Electronic Science and Technology of China
South Renmin Road
Chengdu, 610041
China
Phone: 86 18236170185
Email: 18236170185@163.com

Abstract

Background: Artificial intelligence (AI) shows considerable promise in the areas of lymphoma diagnosis, prognosis, and gene prediction. However, a comprehensive assessment of potential biases and the clinical utility of AI models is still needed.

Objective: Our goal was to evaluate the biases of published studies using AI models for lymphoma histopathology and assess the clinical utility of comprehensive AI models for diagnosis or prognosis.

Methods: This study adhered to the Systematic Review Reporting Standards. A comprehensive literature search was conducted across PubMed, Cochrane Library, and Web of Science from their inception until August 30, 2024. The search criteria included the use of AI for prognosis involving human lymphoma tissue pathology images, diagnosis, gene mutation prediction, etc. The risk of bias was evaluated using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Information for each AI model was systematically tabulated, and summary statistics were reported. The study is registered with PROSPERO (CRD42024537394) and follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 reporting guidelines.

Results: The search identified 3565 records, with 41 articles ultimately meeting the inclusion criteria. A total of 41 AI models were included in the analysis, comprising 17 diagnostic models, 10 prognostic models, 2 models for detecting ectopic gene expression, and 12 additional models related to diagnosis. All studies exhibited a high or unclear risk of bias, primarily due to limited analysis and incomplete reporting of participant recruitment. Most high-risk models (10/41) predominantly assigned high-risk classifications to participants. Almost all the articles presented an unclear risk of bias in at least one domain, with the most frequent being participant selection (16/41) and statistical analysis (37/41). The primary reasons for this were insufficient analysis of participant recruitment and a lack of interpretability in outcome analyses. In the diagnostic models, the most frequently studied lymphoma subtypes were diffuse large B-cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia, and mantle cell lymphoma, while in the prognostic models, the most common subtypes were diffuse large B-cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia, and Hodgkin lymphoma. In the internal validation results of all models, the area under the receiver operating characteristic curve (AUC) ranged from 0.75 to 0.99 and accuracy ranged from 68.3% to 100%. In models with external validation results, the AUC ranged from 0.93 to 0.99.

Conclusions: From a methodological perspective, all models exhibited biases. The enhancement of the accuracy of AI models and the acceleration of their clinical translation hinge on several critical aspects. These include the comprehensive reporting of data sources, the diversity of datasets, the study design, the transparency and interpretability of AI models, the use of cross-validation and external validation, and adherence to regulatory guidance and standardized processes in the field of medical AI.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e62851) doi:10.2196/62851

KEYWORDS

lymphoma; artificial intelligence; bias; histopathology; tumor; hematological; lymphatic disease; public health; pathologists; pathology; immunohistochemistry; diagnosis; prognosis

Introduction

Lymphoma, a malignancy that originates from the lymphohematopoietic system, is recognized as one of the most prevalent hematological cancers globally. Epidemiological studies indicate that Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL) are prevalent malignant lymphatic disorders that pose significant public health challenges. Data from GLOBOCAN 2020 reveal that the projected global incidence and mortality for HL are 83,087 and 23,376 cases, respectively, while for NHL, these figures stand at 544,352 and 259,793, respectively [1,2]. In China, lymphoma is a considerable public health challenge. Data from the Global Burden of Diseases (GBD), Injuries, and Risk Factors Study for 2019 reveal that the age-standardized incidence rate for HL is 0.57 cases per 100,000 individuals, with an age-standardized mortality rate of 0.15 per 100,000 individuals. For NHL, the age-standardized incidence rate is significantly higher at 4.99 per 100,000 individuals, and the age-standardized mortality rate is 2.32 per 100,000 individuals [3].

Histopathology, which involves examining tissue specimens at the cellular level, is the gold standard for the diagnosis of lymphomas [4]. The conventional diagnostic process typically involves pathologists using hematoxylin-eosin (HE) staining of tissues and immunophenotyping for diagnosis. For the diagnosis of high-grade B-cell lymphomas, fluorescence in situ hybridization (FISH) is also commonly employed in conjunction [5]. However, this method has several drawbacks, including subjectivity, time-consuming procedures, and high costs [6].

Traditionally, pathologists have relied on optical microscopes to analyze pathological tissue sections. However, the advent of digital pathology has seen a shift toward the use of computers for reviewing and analyzing scanned whole slide images (WSIs). This transition is not only driven by the potential for increased efficiency but also opens new avenues for the development of automated diagnostic tools [7]. These tools have the potential to enhance the accuracy, efficiency, objectivity, and consistency of diagnoses, which is crucial in addressing the global shortage of pathologists. They can also increase diagnostic throughput and reduce the reliance on referrals and additional tests [8]. This field of research is burgeoning, and for certain types of malignant tumors, these systems are beginning to demonstrate clinical utility [9]. However, although many research models of artificial intelligence (AI) applied to lymphoma histopathology have been published, it is unclear whether there are methodological biases in these models, and the clinical utility of AI applied to lymphoma histopathology has not been summarized.

In this extensive study, we systematically reviewed the literature exploring the use of AI technologies, including traditional machine learning (ML) and deep learning (DL) methods, to assess digital pathology images for lymphoma diagnosis, prognosis, and other pertinent applications. Our review encompasses research that focuses on individual diagnostic factors, such as histological subtypes, as well as studies that perform computer-assisted tasks like tumor segmentation. We also assessed the clinical utility of these AI methods with consideration of potential biases. The objective of this review is to provide valuable insights and actionable recommendations based on the existing body of literature. Thus, this review aims to provide insights and recommendations based on published literature to improve the clinical utility of future research, including reducing the risk of bias, improving reproducibility, and increasing generalizability.

Methods

Literature Search

A thorough search was conducted across 3 prominent research databases: PubMed, Cochrane, and Web of Science. The search was restricted to peer-reviewed journals and conference proceedings to ensure the quality and credibility of the studies included. The search timeline extended from the inception of each database up to August 30, 2024. We employed MeSH terms for more precise retrieval.

Given the multitude of terms related to AI, our search strategy incorporated keywords such as “artificial intelligence,” “machine learning,” “neural network,” and “network, neural (computer),” along with “lymphoma.” We combined multiple relevant terms for each concept using the OR operator (eg, “artificial intelligence” OR “machine learning”) and then merged “lymphoma” with “artificial intelligence” using the AND operator. This approach ensured that the retrieved studies met both criteria.

Subsequently, we screened the articles based on their relevance to histopathological AI, focusing on the title and abstract. The review protocol was registered on PROSPERO (CRD42024537394) prior to the screening of search results for inclusion. Detailed search strategies and methods are provided in Multimedia Appendix 1.

Literature Selection

A researcher (YF) manually removed duplicate papers with the assistance of the reference management software EndNote 20. Subsequently, another researcher (ZH) independently screened the articles for inclusion in 2 stages: the first based on titles and abstracts, and the second based on the full text. Disagreements were discussed and arbitrated by a third researcher (JL).

The inclusion criteria required the research to evaluate the use of at least one AI approach to make diagnostic or prognostic inferences on human histopathology images from suspected or confirmed cases of lymphoma. Studies were only included if AI methods were applied directly to the digital pathology images

XSL•FO
RenderX

or to features that were automatically extracted from the images. Fundamental tasks, such as segmentation and cell counting, were included as these could be used by pathologists for computer-aided diagnosis. Only conventional light microscopy images were considered, with other imaging modalities, such as fluorescence and hyperspectral imaging, excluded. Publications that did not include primary research, such as review papers, were excluded. Non-English language articles and research where a full version of the manuscript was not accessible were excluded.

In the studies included, models were deemed of interest if they adhered to the same inclusion criteria. When several models were compared against the same outcome, the model of interest was typically the newly proposed one. If this was ambiguous, the model with the best performance during the validation phase was selected. When multiple models from a single study exhibited similar modeling techniques, the one with superior validation performance was included in the assessment. Results from the same model at varying levels of precision (eg, patch level, slice level, and patient level) were not treated as distinct outcomes.

Risk of Bias Assessment

The risk of bias in the models of interest was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST) [10]. The tool evaluates the likelihood that the reported results are distorted due to limitations in study design, conduct, and analysis. PROBAST includes 20 guiding questions categorized into 4 domains: Participants, Predictors, Outcomes, and Analysis. These questions are summarized to indicate a high risk or low risk of bias or are marked as unclear when insufficient information is available for a comprehensive assessment and no information is available to suggest a high risk of bias. It is important to note that an unclear risk of bias does not imply a methodological flaw but rather indicates incomplete reporting.

The Participants domain involves the recruitment and selection of participants to ensure the consistency and representativeness of the study population targeting the intended demographic. Relevant details include the recruitment strategy, inclusion criteria, and number of participants enrolled.

The Predictors domain addresses the consistent definition and measurement of predictive variables, which in this context often refers to the generation of digital pathology images. This encompasses methods for the fixation, staining, scanning, and digital processing of tissues prior to modeling.

The Outcomes domain involves the appropriate definition and consistent determination of ground truth labels. This includes the criteria used to ascertain diagnoses or prognoses, the expertise of those determining these labels, and whether the labels are independent of any model outputs.

The Analysis domain encompasses statistical considerations in the evaluation of model performance to ensure valid and not overly optimistic results. It includes various factors, such as the number of participants for each outcome in the test set, the validation methods used (cross-validation, external validation, internal validation, etc), the metrics for assessing performance, and the methods to address the impact of censoring, confounding, and missing data. Some of these factors are interrelated. For example, the risk of bias due to a small dataset is somewhat mitigated by cross-validation, which increases the effective size of the test set and can be used to assess variability, reducing the optimism of the results. Additionally, the risk associated with using a small dataset depends on the type of outcome being predicted; robust analysis for a 5-class classification requires more data than a binary classification. There must also be sufficient data across all relevant patient subgroups; for instance, if multiple subtypes of lymphomas are included, it is not acceptable for 1 subtype to be represented by only a few patients. Due to these interrelated factors, there are no rigid standards for determining the appropriate size of a dataset.

Inconsistencies in methodology often lead to bias risk. For example, inconsistencies in HE staining from different research centers can lead to heterogeneity in the visual characteristics of digital pathology slides, potentially causing spurious correlations through random or systematic differences within or between subgroups in the dataset. Using a large dataset during training may enhance the model's generalizability, but this must be tightly controlled to avoid introducing systematic confounding. Inconsistencies in the determination of outcomes may mean that the results of a study are unreliable due to spurious correlations in the underlying factual labels or invalid due to misjudgment of the labels.

While PROBAST provides a framework for assessing the risk of bias, there is a degree of subjectivity in interpreting the signal questions. Therefore, each model was analyzed by 2 independent researchers (YF and ZH), with at least one computer scientist and one pathologist involved in the bias risk assessment of each model.

The Quality Assessment of Diagnostic Accuracy Studies-AI (QUADAS-AI) tool was used to evaluate the sensitivity of the included studies. QUADAS-AI is the AI-specific extension of QUADAS-2 [11] and QUADAS-C [12], and includes 4 domains for determining the risk of bias (patient selection, reference standard, index test, and flow and timing) and 3 domains for applicability issues (index test, patient selection, and reference standard) (Multimedia Appendix 2).

Data Synthesis

Data extraction was independently performed by 2 researchers (YF and MZ), using a form containing 67 fields within the categories Overview, Data, Methods, Results, and Miscellaneous. A summary of this process is provided in Multimedia Appendix 3.

Information was sought from full-text articles, as well as references and supplementary materials where appropriate. Inferences were made only when both researchers were confident that this gave the correct information, with disagreements resolved through discussion. Fields that could not be confidently completed were labeled as being unclear.

All extracted data were summarized in 2 tables, 1 each for study-level and model-level characteristics. Only models of interest were included in these tables. The term model outcome

XSL•FO
RenderX

refers to the model output (whether this was a clinical outcome [diagnosis or prognosis] or a diagnostically relevant outcome that could be used for computer-aided diagnosis, such as tumor segmentation). The data synthesis did not include any meta-analysis due to the diversity of the included methods and model outcomes. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines for reporting systematic reviews were followed, with checklists provided in Multimedia Appendix 3.

Results

Results of Literature Screening

As shown in Figure 1, the initial literature search identified a total of 3565 records, with 1375 being duplicates. After screening titles and abstracts, 2052 records were excluded, leaving 41 studies for inclusion in the review. All studies that met the inclusion criteria were identified through research databases, with no eligible records found in trial registries. Although the search was performed from 1949, all included studies were published since 2010, and more than 80% were published since 2020. The characteristics of these studies are summarized in Table 1. The model construction parameters of the 41 included studies are detailed in Table 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart. AI: artificial intelligence.

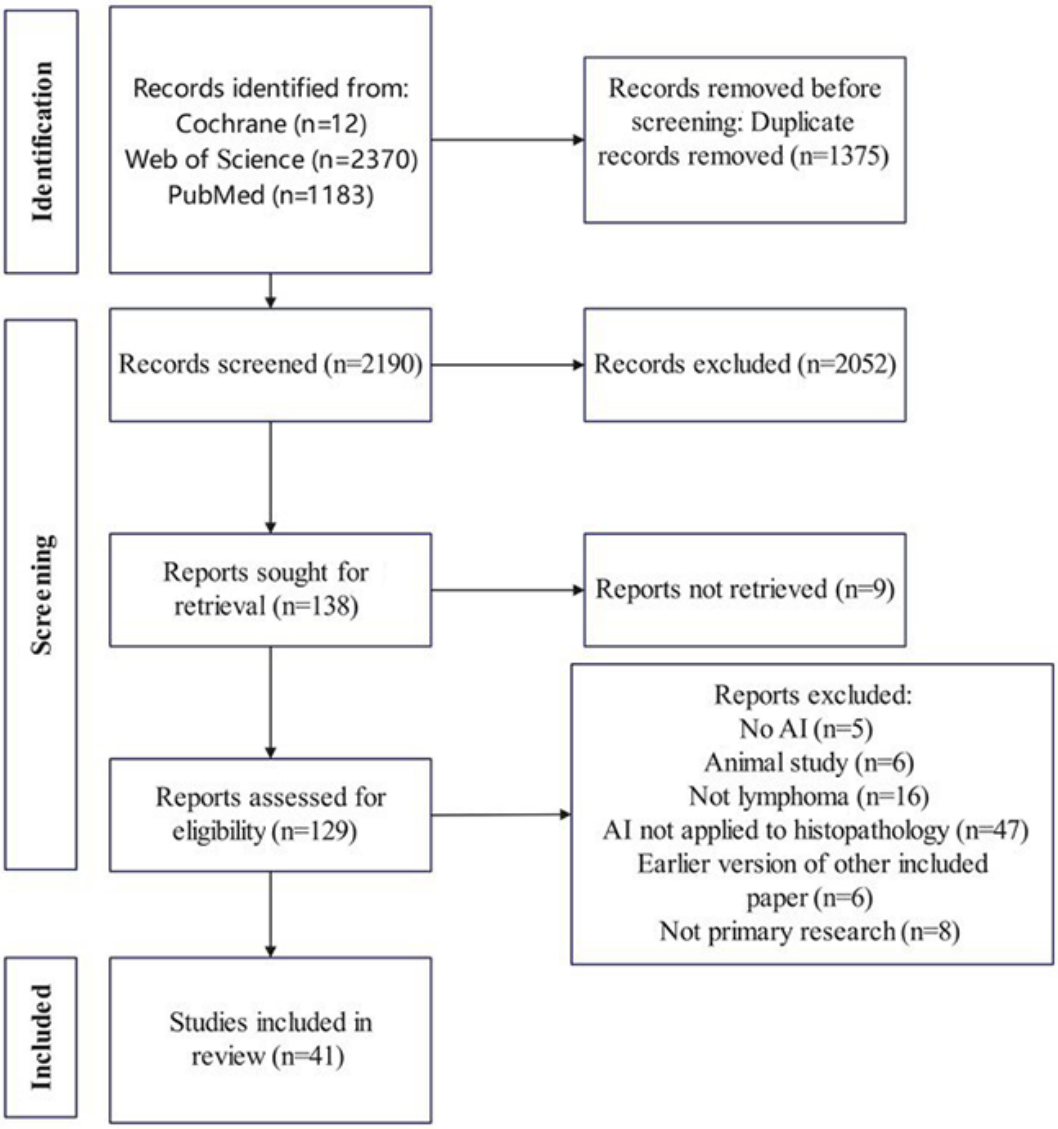

XSL•FO
RenderX

Table 1. Characteristics of the 41 studies included in this systematic review.

Publication	Data source	Geographical distribution	Internal participants, n	Internal pathology images, n	Subtypes	Outcome type	Model task	Code available
Achi et al [13], 2019	Single center	America	128	2560	Burkitt lymphoma, DLBCL[a], and small lymphocytic lymphoma	Diagnosis	Identification	None
Steinbuss et al [14], 2021	Single center	Germany	629	Unclear	Nodal small lymphocytic lymphoma/CLL[b] and DLBCL	Diagnosis	Classification	[15]
Miyoshi et al [16], 2020	Single center	Japan	388	Unclear	FL[c] and DLBCL	Diagnosis	Classification	None
Syrykh et al [17], 2020	Single center	France	378	Unclear	FL	Diagnosis	Identification	[18]
Sereda et al [19], 2023	Single center	Germany	53	Unclear	Pediatric nodular lymphocyte–predominant Hodgkin lymphoma	Prognosis	Identification, semantic segmentation	None
Zhang et al [20], 2020	Single center	China	Unclear	374	CLL, MCL[d], and FL	Diagnosis	Classification	None
Li et al [21], 2022	Multicenter	America	1005	3123	DLBCL	Diagnosis	Classification	[22]
Zhang et al [23], 2021	Single center	China	Unclear	374	CLL, MCL, and FL	Other	Classification	None
Yu et al [24], 2021	Multicenter	China	40	40	Monomorphic epitheliotropic intestinal T-cell lymphoma	Other	Classification, cell nuclei segmentation	None
Takagi et al [25], 2023	Single center	Japan	842	Unclear	DLBCL and FL	Other	Classification	None
Hamdi et al [26], 2023	Single center	Saudi Arabia	Unclear	15,000	CLL, MCL, and FL	Diagnosis	Classification	[27]
Mohlman et al [28], 2020	Single center	America	70	10,818	DLBCL and Burkitt lymphoma	Other	Classification	None
Karabulut et al [29], 2023	Single center	Turkey	11	60	MF[e]	Diagnosis	Identification, classification, cell nuclei segmentation	None
do Nascimento et al [30], 2018	Single center	Brazil	347	15,353	CLL, MCL, and FL	Diagnosis	Classification	None
Zhu et al [31], 2019	Multicenter	China	Unclear	374	CLL, MCL, and FL	Other	Classification	None
Hashimoto et al [32], 2022	Single center	Japan	262	Unclear	DLBCL, Burkitt lymphoma, and FL	Other	Classification	None
Michail et al [33], 2014	Single center	Greece	Unclear	300	FL	Other	Identification, semantic segmentation	None
Kornaropoulos et al [34], 2014	Single center	Greece	17	500	FL	Other	Classification	None
Chuang et al [35], 2022	Single center	China	103	309	MCL	Prognosis	Semantic segmentation	None

XSL•FO
RenderX

Publication	Data source	Geographical distribution	Internal participants, n	Internal pathology images, n	Subtypes	Outcome type	Model task	Code available
Vrabac et al [36], 2021	Multicenter	America	209	Unclear	DLBCL	Prognosis	Semantic segmentation	[37]
Motmaen et al [38], 2023	Single center	Germany	83	Unclear	HL[f]	Prognosis	Classification	None
Swiderska-Chadaj et al [39], 2021	Multicenter	Holland	287	354	DLBCL	MYC translocation detection	Identification	None
Tagami et al [40], 2023	Single center	Japan	129	1290	Ocular adnexal mucosa–associated lymphoid tissue lymphoma	Other	Classification	None
Irshaid et al [41], 2022	Single center	America	61	Unclear	CLL and FL	Prognosis	Cell nuclei segmentation, classification	None
El Hussein et al [42], 2022	Multicenter	America	125	213	CLL	Prognosis	Cell nuclei segmentation, classification	None
Chen et al [43], 2022	Multicenter	America	135	213	CLL	Prognosis	Cell nuclei segmentation, classification	None
Zhang et al [44], 2024	Multicenter	China	Unclear	Unclear	Primary central nervous system lymphoma	Diagnosis	Classification	[45]
Tavolara et al [46], 2024	Single center	America	172	376	DLBCL	Other	Identification	[47]
Perry et al [5], 2023	Single center	Israel	57	Unclear	High-grade B-cell lymphoma	Diagnosis	Classification	None
Yan et al [48], 2024	Single center	China	220	220	DLBCL	Prognosis	Cell nuclei segmentation	[49]
Lee et al [50], 2024	Single center	America	216	251	DLBCL	Prognosis	Cell nuclei segmentation	None
Duan et al [51], 2024	Single center	China	114	132	Primary central nervous system lymphoma	Prognosis	Cell nuclei segmentation	None
Quan et al [52], 2024	Single center	China	Unclear	350	Gastric MALT[g] lymphoma	Diagnosis	Cell nuclei segmentation	None
Al-Mekhlafi et al [53], 2022	Single center	Saudi Arabia	Unclear	1500	CLL, FL, and MCL	Diagnosis	Classification	None
Somaratne et al [54], 2019	Single center	Australia	Unclear	374	FL	Diagnosis	Classification	None
Codella et al [55], 2016	Multicenter	America	Unclear	374	CLL, FL, and MCL	Diagnosis	Classification	None
Swiderska-Chadaj et al [56], 2020	Multicenter	America	91	157	DLBCL	MYC translocation detection	Identification	None
Basu et al [57], 2022	Unclear	India	Unclear	700	MF	Diagnosis	Classification	None
Shankar et al [58], 2023	Single center	America	Unclear	Unclear	cHL, DLBCL, and MCL	Diagnosis	Classification	[59]

XSL•FO
RenderX

Publication	Data source	Geographical distribution	Internal participants, n	Internal pathology images, n	Subtypes	Outcome type	Model task	Code available
Soltane et al [60], 2022	Unclear	Saudi Arabia	Unclear	323	cHL, nodular lymphoma predominant, Burkitt lymphoma, FL, MCL, large B-cell lymphoma, and T-cell lymphoma	Diagnosis	Classification	None
Tagami et al [61], 2024	Single center	Japan	127	1270	Orbital MALT lymphoma	Diagnosis	Classification	None

[a]DLBCL: diffuse large B-cell lymphoma.

[b]CLL: chronic lymphocytic leukemia.

[c]FL: follicular lymphoma.

[d]MCL: mantle cell lymphoma.

[e]MF: mycosis fungoides.

[f]HL: Hodgkin lymphoma.

[g]MALT: mucosa-associated lymphoid tissue.

XSL•FO
RenderX

Table 2. Model construction parameters for the 41 included studies.

Publication	Stain type	Original image size	Patch size	Magnification	Feature extraction	Final model	Final model prediction precision	Validation type	External validation data	Metric	Internal results	External results
Achi et al [13]	HE[a]	WSI[b]	40×40	40×	Learned	CNN[c]	Patch	Internal validation	None	Accuracy	95%	None
Steinbuss et al [14]	HE	WSI	395×395	40×	Handcrafted	Efficient Net	Patch	Internal validation	None	Accuracy	95%	None
Miyoshi et al [16]	HE	WSI	64×64	5×	Handcrafted	CNN	Patch	5-fold cross-validation	None	Accuracy	94%	None
Syrykh et al [17]	HE	WSI	299×299	20×	Handcrafted	BNN[d]	Patch	Internal validation	None	AUC[e]	0.92-0.99	None
Zhang et al [20]	HE	WSI	224×224	Unclear	Learned	ResNet-50	Patch	5-fold cross-validation	None	Accuracy	95.4%	None
Li et al [21]	HE	WSI	945×945	40×	Learned	17CNN+transform	Patch	External validation	402	Accuracy	100%	100%
Hamdi et al [26]	HE	WSI	Unclear	Unclear	Handcrafted	MobileNet-VGG-16, decision tree–based machine learning	WSI	Internal validation	None	AUC, accuracy	0.99, 99.8%	None
Karabulut et al [29]	HE	Microscopic images	600×600	200×	Learned	DL[f]	Patch	Internal validation	None	Accuracy	94.2%	None
do Nascimento et al [30]	HE	Microscopic images	Unclear	1000×	Unclear	Classification using the polynomial	Unclear	Internal validation	None	Accuracy	96%-100%	None
Sereda et al [19]	IHC[g]	Patch	256×256	Unclear	Handcrafted	YOLOv4-tiny CNN	Patch	Internal validation	None	Accuracy	95.43%	None
Chuang et al [35]	HE	WSI	132×132	40×	Handcrafted	CNN	Patch	Internal validation	None	AUC	0.94	None
Vrabac et al [36]	HE	WSI	224×224	40×	Handcrafted	Hover-Net	Patch	External validation	179	CI	None	95%
Motmaen et al [38]	Picrosirius Red	WSI	320×320	20×	Handcrafted	YOLOv4	Patch	Internal validation	None	AUC	0.79	None
Irshaid et al [41]	HE	WSI	128×128	40×	Handcrafted	CNN	Patch	Internal validation	None	AUC	0.85	None
El Hussein et al [42]	HE	WSI	256×256	20×	Handcrafted	Hover-Net	WSI	External validation	28	AUC	None	0.93
Chen et al [43]	HE	WSI	256×256	20×	Handcrafted	Hover-Net	WSI	External validation	68	Accuracy	None	92.5%

XSL•FO
RenderX

Publication	Stain type	Original image size	Patch size	Magnification	Feature extraction	Final model	Final model prediction precision	Validation type	External validation data	Metric	Internal results	External results
Swiderska-Chadaj et al [39]	HE	WSI	Unclear	20×	Learned	U-Net	WSI	External validation	49	Sensitivity	None	93%
Zhang et al [23]	HE	WSI	Unclear	Unclear	Handcrafted	ResNet-50	Patch	Internal validation	None	Accuracy	98.63%	None
Yu et al [24]	HE	WSI	115×115	40×	Handcrafted	Decision tree–based machine learning	Patch	Internal validation	None	AUC	0.96	None
Takagi et al [25]	HE	WSI	224×224	40×	Learned	CNN	Patch	5-fold cross-validation	None	Accuracy	0.83	None
Mohlman et al [28]	HE	WSI	224×224	200×	Learned	CNN	Patch	Internal validation	None	AUC	0.92	None
Zhu et al [31]	HE	WSI	64×64	Unclear	Learned	VGG-16, LSTM[h]	Patch	10-fold cross-validation	None	Overall grading accuracy	0.98	None
Hashimoto et al [32]	HE	WSI	224×224	20×	Learned	CNN	Patch	5-fold cross-validation	None	Accuracy	68.3%	None
Michail et al [33]	HE	Microscopic images	Unclear	40×	Handcrafted	SVM[i]	Microscopic images	Internal validation	None	Accuracy	97.4%	None
Kornaropoulos et al [34]	HE	WSI	71×71	Unclear	Handcrafted	Laplacian Eigenmaps	WSI	Hold-out K-folds	None	Accuracy	99.22%	None
Tagami et al [40]	HE	WSI	2048×2048	20×	Handcrafted	SVM	Patch	10-fold cross-validation	None	AUC	0.86	None
Zhang et al [44]	HE	WSI	256×256	40×	Learned	DL	Patch	External validation	None	AUC	0.96	None
Tavolara et al [46]	IHC	WSI	224×224	40×	Handcrafted	ResNet-50	Patch	External validation	108	Sensitivity, specificity	None	0.857, 0.991
Perry et al [5]	HE	WSI	384×384	40×	Handcrafted	DL	Patch	Internal validation	None	AUC	0.95	None
Yan et al [48]	IHC	WSI	256×256	40×	Handcrafted	CNN	Patch	External validation	61	ICC[j]	None	96%
Lee et al [50]	HE	WSI	224×224	40×	Learned	ViT-S/8	Patch	External validation	48	Sensitivity, specificity	None	90.2%, 70.0%
Duan et al [51]	HE	WSI	512×512	20×	Handcrafted	KNN[k]	Patch	Internal validation	46	AUC	None	0.92

XSL•FO
RenderX

Publication	Stain type	Original image size	Patch size	Magnification	Feature extraction	Final model	Final model prediction precision	Validation type	External validation data	Metric	Internal results	External results
Quan et al [52]	HE	WSI	512×512	40×	Handcrafted	ResNet 50	Patch	5-fold cross-validation	None	Sensitivity, specificity	96.79%, SD 1.50%; 99.38%, SD 0.15%	None
Al-Mekhlafi et al [53]	HE	WSI	512×512	40×	Handcrafted	ResNet 50	Patch	Internal validation	None	AUC	0.99	None
Somaratne et al [54]	HE	WSI	227×227	Unclear	Handcrafted	AlexNet	Patch	External validation	213	AUC	None	0.99
Codella et al [55]	HE	WSI	Unclear	Unclear	Unclear	Unclear	Patch	3-fold cross-validation	None	Accuracy	92.3%	None
Swiderska-Chadaj et al [56]	HE	WSI	512×512	20×	Handcrafted	CNN	Patch	External validation	66	AUC	None	0.83
Basu et al [57]	HE	WSI	224×224	40×	Handcrafted	CNN	Patch	Internal validation	None	Sensitivity, specificity	94.67%, 97.3%	None
Shankar et al [58]	HE	WSI	Unclear	40×	Handcrafted	ResNet 50	Patch	Internal validation	None	AUC	0.95	None
Soltane et al [60]	HE	WSI	224×224	Unclear	Learned	ResNet 50	Patch	5-fold cross-validation	None	Accuracy	91.6%	None
Tagami et al [61]	HE	WSI	2048×2048	20×	Handcrafted	DL	Patch	5-fold cross-validation	None	AUC	0.8	None

[a]HE: hematoxylin-eosin.

[b]WSI: whole slide image.

[c]CNN: convolutional neural network.

[d]BNN: Bayesian neural network.

[e]AUC: area under the receiver operating characteristic curve.

[f]DL: deep learning.

[g]IHC: immunohistochemical.

[h]LSTM: long short-term memory.

[i]SVM: support vector machine.

[j]ICC: intraclass correlation coefficient.

[k]KNN: k-nearest neighbors.

Risk of Bias Assessment

The PROBAST assessment findings are detailed in Table 3. Despite some studies encompassing multiple models of interest, each paper highlighted 1 model with superior predictive value for bias risk analysis. Most models exhibited either a high overall bias risk (13/41) or an unclear overall bias risk (28/41), with none of the models presenting a low overall bias risk (0/41). Most high-risk models predominantly allocated their high-risk scores in the Participants domain (10/41). Conversely, most low-risk scores were concentrated in the Predictors (26/41) and Outcomes (26/41) domains. Almost all studies reported an unclear risk of bias in at least one domain, with the Participants (16/41) and Statistical Analysis (37/41) domains being the most frequently affected. Qualitative summaries are presented in Figure 2.

XSL•FO
RenderX

Table 3. Prediction Model Risk of Bias Assessment Tool (PROBAST) risk of bias assessment results for the 41 papers included in this review.

Publication	Participants	Predictors	Outcomes	Analysis	Overall judgement
Achi et al [13]	Unclear	Low	Unclear	Unclear	Unclear concerns
Steinbuss et al [14]	High	Unclear	Low	Unclear	High concerns
Miyoshi et al [16]	High	Low	Low	Unclear	High concerns
Syrykh et al [17]	High	Low	Low	Low	High concerns
Sereda et al [19]	Low	Low	Low	Unclear	Unclear concerns
Zhang et al [20]	Unclear	Low	Unclear	Unclear	Unclear concerns
Li et al [21]	High	Low	Low	Unclear	High concerns
Zhang et al [23]	Unclear	Unclear	Low	Unclear	Unclear concerns
Yu et al [24]	Unclear	Low	Low	Unclear	Unclear concerns
Takagi et al [25]	Unclear	Low	Low	Unclear	Unclear concerns
Hamdi et al [26]	Unclear	Low	Unclear	Unclear	Unclear concerns
Mohlman et al [28]	High	Low	Low	Low	High concerns
Karabulut et al [29]	Unclear	Low	Low	Unclear	Unclear concerns
do Nascimento et al [30]	Unclear	Unclear	Unclear	Unclear	Unclear concerns
Zhu et al [31]	Unclear	Low	Unclear	Unclear	Unclear concerns
Hashimoto et al [32]	Unclear	Low	Unclear	Unclear	Unclear concerns
Michail et al [33]	Unclear	Unclear	Unclear	Unclear	Unclear concerns
Kornaropoulos et al [34]	Low	Low	Low	Unclear	Unclear concerns
Chuang et al [35]	Unclear	Low	Unclear	Unclear	Unclear concerns
Vrabac et al [36]	Low	Low	Low	Unclear	Unclear concerns
Motmaen et al [38]	High	Unclear	Low	Unclear	High concerns
Swiderska-Chadaj et al [39]	Unclear	Unclear	Unclear	Unclear	Unclear concerns
Tagami et al [40]	High	Unclear	Low	Unclear	High concerns
Irshaid et al [41]	High	Low	Low	Unclear	High concerns
El Hussein et al [42]	High	Low	Low	Unclear	High concerns
Chen et al [43]	High	Low	Low	Unclear	High concerns
Zhang et al [44]	Unclear	Unclear	Low	Unclear	Unclear concerns
Tavolara et al [46]	Low	Low	Unclear	Unclear	Unclear concerns
Perry et al [5]	Low	High	Low	High	High concerns
Yan et al [48]	Low	Low	Low	Unclear	Unclear concerns
Lee et al [50]	Low	Low	Low	Unclear	Unclear concerns
Duan et al [51]	Low	Low	High	High	High concerns
Quan et al [52]	Low	Unclear	Unclear	Unclear	High concerns
Al-Mekhlafi et al [53]	Low	Unclear	Low	Unclear	High concerns
Somaratne et al [54]	Low	Unclear	Low	Unclear	Unclear concerns
Codella et al [55]	Low	Unclear	Unclear	High	High concerns
Swiderska-Chadaj et al [56]	Low	Unclear	Low	Unclear	Unclear concerns
Basu et al [57]	Unclear	Unclear	Unclear	Unclear	Unclear concerns
Shankar et al [58]	Low	Low	Low	Unclear	Unclear concerns
Soltane et al [60]	Unclear	Low	Unclear	Unclear	Unclear concerns
Tagami et al [61]	Low	Low	Low	Unclear	Unclear concerns

XSL•FO
RenderX

Figure 2. Prediction Model Risk of Bias Assessment Tool (PROBAST) risk of bias.

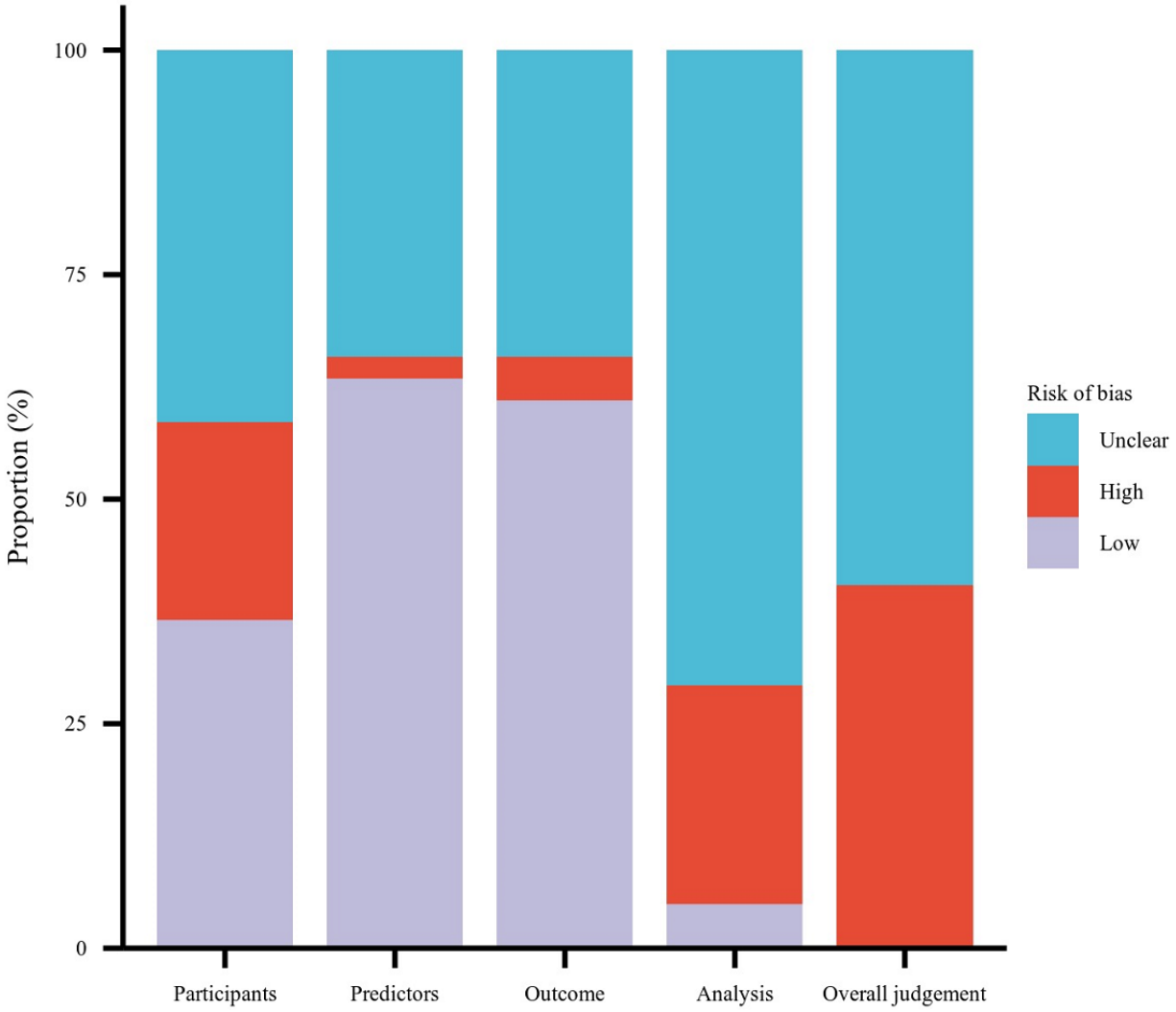

Data Synthesis Results

Data in the Included Literature

The number of participants across internal datasets varied significantly, with studies recruiting anywhere from 10 to 1005 patients diagnosed with lymphoma. Concurrently, the model development used a broad range of histopathological slides (count: 10-15,353). In most studies, the samples for model development were WSIs involving excised or biopsied tissues (38/41), with other samples using microscopic images (3/41). Most studies used HE-stained tissues (37/41), while others employed various immunohistochemical (IHC) staining methods (4/41). Some studies employed a multimodal analysis method that integrated pathological images with clinical information [25,35,50].

Among the included studies, most (29/41) used single-center data and few (10/41) used multicenter data. A small number of studies had unclear data sources, and the United States was the most common (12/41) source country.

Models in the Included Literature

The studies encompassed a variety of models, with the most prevalent being convolutional neural networks (CNNs), accounting for 30 out of 41 studies. A minority of studies employed support vector machines (SVMs) and random forests (2 studies each). The CNN architectures that were explored included Mobile Net, VGG-16, Hover-Net, U-Net, and ResNet-50. These newer CNNs generally incorporated multiple standardized blocks that featured layers for convolution, normalization, activation, and pooling [23]. One study stood out by leveraging transfer learning and integrating 17 distinct DL models to create a highly accurate platform, achieving a diagnostic accuracy rate of 100% [21]. This approach significantly bolstered the model's ability to generalize. Another innovative study developed a novel architecture by applying a topological optimization method to the conventional VGG-16 model [26]. Some studies opted for a hybrid approach, combining traditional ML techniques with DL. They used CNNs for feature extraction, followed by decision tree–based methods for quantification and classification [24,26]. Notably, 1 study implemented a CNN framework grounded in Multiple Instance Learning (MIL), which autonomously concentrated on image patches from regions of interest within tumors, showcasing an advanced method for image analysis [32].

In the analysis encompassing various models, most studies predominantly used patches (33/41), with a subset operating at

XSL•FO
RenderX

the WSI level (6/41). Two distinct aggregation approaches were implemented: premodeling and postmodeling aggregation. The premodeling method necessitated the creation of slide-level features prior to modeling, whereas the postmodeling approach entailed consolidating patch-level model outputs to formulate slide-level predictions. For models that used patch images as the basis for final modeling, it was essential to segment the original images into individual patches before proceeding with modeling. The patch sizes varied from 40×40 to 2048×2048 pixels, with the most frequently employed dimensions being 224×224 pixels (9/41) and 256×256 pixels (5/41). Subsequently, a variety of feature extraction techniques were applied, encompassing both handcrafted or predefined features (27/41) and features that were automatically learned by the models (12/41).

The handcrafted features encompassed a diverse spectrum of attributes, including texture, color, cellular, and nuclear morphological characteristics. These meticulously crafted features were predominantly employed as inputs for traditional ML algorithms, such as SVM and random forest models.

In contrast, learned features were predominantly extracted through the application of CNNs, which also frequently served as the classifier of choice. Ultimately, the outputs from the patch-level models were synthesized to develop predictive models. This aggregation was achieved through various methods, such as attention-based weighted averaging, concatenation, and more sophisticated embedding techniques. These included Fisher vector encoding and k-means clustering, with the process often culminating in the selection of the maximum value to enhance the predictive accuracy of the models [36,62].

Among the papers that specified magnification levels, the most prevalent were 20× (10/41) and 40× (17/41). A handful of studies employed varying magnifications strategically to pinpoint informative tissue regions and to enhance their modeling accuracy [16,17,40].

A limited number of models integrated histopathological data with other data modalities [25,35,50]. The multimodal approaches observed in the literature included the premodeling integration of unimodal features extracted independently, as well as the amalgamation of unimodal predictions from distinct models [63]. Additionally, transformer-based methods were frequently employed for encoding the intricate relationships between different modalities [64,65]. While attention-based methods have been used in the study of other malignancies for several years [66], their application in lymphoma research is a relatively recent development. Among the studies reviewed, 1 study stood out by using a variant of the transformer architecture to encode the interplay between medical imaging data and clinical records. This study introduced a novel personalized attention mechanism (PersAM) for the classification of lymphoma subtypes, marking a significant advancement in the field [25].

Analysis in the Included Literature

Most studies relied on internal validation (30/41), while external validation using independent lymphoma datasets was seldom conducted (11/41). In terms of internal validation, partial validation was typically executed through a 5-10–fold cross-validation approach. Some papers detailed the hyperparameter selection process using the training dataset, yet only reported evaluations on a test set derived from the same data source [24,33,67]. For external validation, models were trained on WSIs and subsequently validated on either WSIs (10/11) or tissue microarrays (TMAs) (1/11) from separate independent sources. Notably, the model of 1 study was externally validated against data from normal lymph node tissues [14]. In a particular instance, a model that achieved perfect validation accuracy (area under the receiver operating characteristic curve [AUC]=1.0) with internal validation underperformed on external cases, with an AUC ranging from 0.63 to 0.69. This discrepancy may stem from the sensitivity of ML algorithms to preprocessing steps, and neural networks, in accordance with statistical principles, necessitate a representative sample to ensure reliable inductive reasoning [12]. In another study, a comprehensive evaluation was conducted using polynomial, SVM, random forest, and decision tree classifiers to assess the efficacy of the proposed method [30].

Most models were assessed using accuracy or AUC, as well as other metrics, including sensitivity, specificity, hazard ratio, and the C-index. Even when studies reported sensitivity and specificity, CIs were not reported. In the internal validation results of all models, the AUC ranged from 0.75 to 0.99 and accuracy ranged from 68.3% to 100%. In models with external validation results, the AUC ranged from 0.93 to 0.99.

The burgeoning demand for AI methods in health care is undeniable, yet the lack of interpretability remains a significant impediment to their clinical adoption [68,69]. Enhancing the interpretability of AI models is crucial for fostering trust among medical professionals in the future AI systems they will rely upon. In a thorough analysis of the studies included, it was observed that the majority of studies (20/41) undertook efforts to analyze the interpretability of their models, with a notable minority (8/20) delving into visual interpretability analysis of the histopathological images that significantly influenced the model's prognostic assessments. Several studies meticulously characterized the spatial distribution and interrelationships of typical cells, their nuclei, and the microenvironments within the regions of interest [24,32,36,42], thereby showcasing the interpretability of their AI systems. One study presented graphical features that were correlated with clinical prognostic information [38]. A handful of studies opted for traditional ML models, such as decision trees [26], which are inherently more transparent and align closely with human reasoning processes, thus facilitating a more intuitive understanding of the decision-making process.

Clinical Utility

Among the 41 models included, 17 were diagnostic models, 10 were prognostic models, 2 were models to detect gene translocation, and 12 were other prediction- and diagnosis-related information models. The tasks of these models included identification (8/41), classification (24/41), and segmentation (9/41).

XSL•FO
RenderX

In the field of AI-based diagnostic models for lymphoma histopathology, the most common subtypes included diffuse large B-cell lymphoma (DLBCL) (5/41), follicular lymphoma (FL) (8/41), chronic lymphocytic leukemia (CLL) (5/41), and mantle cell lymphoma (MCL) (8/41). Additionally, a small number of studies developed diagnostic models for Burkitt lymphoma, central nervous system lymphoma, high-grade lymphoma, HL, and T-cell lymphoma. A study using CNNs for cell segmentation of WSIs of NHL achieved an average diagnostic accuracy of 100%, 99.73%, and 99.20% for CLL, FL, and MCL, respectively [20]. Researchers have successfully harnessed the power of both DL and traditional ML to develop a diagnostic tool with remarkable accuracy rates ranging from 95% to 100% for identifying MCL, FL, and CLL. This cutting-edge approach involves the precise segmentation of cell nuclei and the meticulous measurement of key morphological features such as area, perimeter, eccentricity, and diameter [30].

In the realm of AI-based prognostic models for lymphoma histopathology, the most common subtypes include DLBCL (3/4), CLL (2/42), HL (2/41), and FL (2/41). Sereda et al [19] used DL-based cell detection on digital slides from patients with nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL) to quantitatively assess the histological patterns of lymphocyte-predominant cells. They identified 6 key features of lymphocyte-predominant cell spatial patterns and achieved a high average precision in cell detection (mean 95.24%, SD 0.17%). Furthermore, they found a strong correlation between treatment response and the density and number of lymphocyte-predominant cells (P<.05). Several studies have identified independent prognostic factors for DLBCL and MCL by segmenting cell nuclei and calculating the geometric characteristics of each segmented nucleus [35,36]. Several studies [41-43] addressed the clinical challenge of large cell transformation in indolent B-cell lymphomas, such as FL and CLL. They trained a CNN to predict large cell transformation based on tumor cell morphology, including the small cell proportion, chromatin pattern, presence of distinct nucleoli, and proliferation index. The machine-generated quantifications demonstrated superior reproducibility compared to estimates made by pathologists and showed a stronger correlation with the outcome data. The precise assessment and evaluation of PD-L1 biomarkers are crucial for the targeted immunotherapy triage of cancer patients. Notably, Yan et al [48] developed an AI-based image analysis method that encompasses the detection, segmentation, and classification of PD-L1+ cells for the evaluation of PD-L1 expression in patients with DLBCL. This method has produced highly correlated quantitative results compared to the subjective assessments of pathologists. However, none of the prognostic models included T-cell lymphoma subtypes in the studies.

Regarding AI-based histopathological models for detecting gene translocations in lymphoma, 2 studies focused on DLBCL as the tumor type [39,56]. Their results showed that it is possible to predict MYC translocation based on morphology alone. This would allow simple and fast prescreening, saving about 34% of genetic testing using the current algorithm.

Overall, methodologically, all studies exhibited a high or unclear risk of bias, primarily due to limited analysis and incomplete reporting on participant recruitment. Most high-risk models (10/41) predominantly assigned high-risk classifications to participants. Almost all studies presented an unclear risk of bias in at least one domain, with the most frequent being participant selection (16/41) and statistical analysis (37/41). The primary reasons for this were insufficient analysis of participant recruitment and a lack of interpretability in outcome analyses. In the diagnostic models, the most frequently studied lymphoma subtypes were DLBCL, FL, CLL, and MCL, while in the prognostic models, the most common subtypes were DLBCL, FL, CLL, and HL. None of the prognostic models included T-cell lymphoma subtypes in the studies. In the internal validation results of all models, the AUC ranged from 0.75 to 0.99 and accuracy ranged from 68.3% to 100%. In models with external validation results, the AUC ranged from 0.93 to 0.99.

Sensitivity Analysis

To further evaluate the sensitivity of our conclusions, we conducted a sensitivity analysis by selecting diagnostic models from the literature we included and performing a QUADAS-AI diagnostic evaluation. The results are presented in Multimedia Appendix 4.

Our findings revealed that out of the 17 diagnostic models considered, 15 were rated as high-risk models and 2 were deemed unclear. Most high-risk models were classified as such primarily due to their use of nonpublic datasets (13/17). A minority of studies (2/17) were rated as high risk because they failed to provide a clear description of their data sources. Therefore, we used QUADAS-AI to evaluate the diagnostic models and concluded that all the methodologies included in the diagnostic models were biased.

To further assess the sensitivity of our conclusions, we considered that the distribution of lymphoma subtypes varies by region, which could potentially bias the results. Therefore, we only included studies from the United States, as it was the most common source country (12/41). We found that in the United States, the most common subtypes in AI-based diagnostic models for lymphoma histopathology were DLBCL (5/12), FL (2/12), and MCL (2/12). Regarding AI-based prognostic models for lymphoma histopathology, the most common subtypes were DLBCL (2/12), CLL (3/12), and FL (1/12). In the models for detecting gene translocations in lymphoma histopathology, 1 study focused on DLBCL. These findings are largely consistent with our previous conclusions.

Discussion

Current Status of AI in Lymphoma Assessment

AI has significantly enhanced the precision of lymphoma diagnostics by eliminating the subjectivity often associated with human observation [16,26]. For instance, a study by Achi et al [13] demonstrated the power of CNN in accurately distinguishing between various lymphoma subtypes. Their diagnostic model, designed for 4 distinct lymphoma categories (benign lymph nodes, DLBCL, Burkitt lymphoma, and small lymphocytic lymphoma), achieved an impressive 95% accuracy rate in image prediction. Notably, in a multicenter study [21], the diagnostic accuracy of AI for DLBCL reached 100%.

XSL•FO
RenderX

Moreover, AI provides invaluable insights into the tumor microenvironment, enabling the identification and quantification of image features that surpass simple density assessments. It delves into higher-order relationships and offers a quantitative evaluation of lymphocyte aggregation patterns and the complex interplay between tumor regions. Such capabilities are pivotal for advancing cancer clinical research and the development of new therapeutics [36,70].

However, methodologically, the risk assessments of the included studies in this review were all rated as high or unclear, primarily due to incomplete reporting, absence of detailed patient source information, and inadequate explanation of the predictors used. This highlights the need to expedite the clinical translation of AI in lymphoma diagnosis, ensuring that these advanced tools are rigorously validated and seamlessly integrated into clinical practice.

Frequently omitted details include the precise origin of patient data, the total number of patients involved, the quantity of samples or images used, and the techniques employed for tissue processing and digitization. Most studies reported data from single centers, and this scarcity may stem from AI researchers not dedicating sufficient effort to comprehend these images, whether for training purposes or external validation. Information about the predictors (histopathology images and their features) was generally better reported; however, there remains an absence or inadequacy of the detailing of certain critical aspects. For example, it is often unclear whether the investigators assessed the predictors without knowing the outcomes or whether all histopathology images were processed uniformly, which could have introduced bias. Moreover, some researchers rely on a limited dataset and analyze a single test data split without implementing methods to mitigate overfitting and model optimism, such as cross-validation or external validation. These limitations are prevalent in lymphoma AI research, resulting in weak validation and an elevated risk of bias within the models.

Code sharing is crucial for enhancing the reproducibility of research findings and mitigating the effects of incomplete reporting. However, in the review of 41 papers, only 9 included codes, and the data in the other studies were either incomplete or difficult to access. To foster better reproducibility, code repositories should provide comprehensive documentation. This should include instructions for setting up the environment, an overview of the code's functionality, guidance on how to produce results, and, of course, the code itself [14,17,21,26,38,44,46,48,58].

Several studies are dedicated to enhancing the interpretability of DL tools by employing existing methods. These include post hoc techniques and supervised ML models that interpret the outcomes after DL models have generated predictions [71,72]. In the realm of AI research focused on lymphoma, most current studies offer personalized interpretability for analysis. This includes visual attention heatmaps and traditional ML highlighting the spatial locations of key feature areas and their interrelationships. However, traditional ML, often crafted in partnership with domain experts, can provide greater interpretability as it relies on manually engineered features. Despite this advantage, the process of handcrafting features is inherently challenging and complex. It demands a substantial time commitment from pathologists or oncologists, who are responsible for developing these methods.

In recent years, there has been a notable increase in hybrid approaches that combine DL with handcrafted strategies. These methods might involve using DL algorithms for the preliminary detection of cells or elements, followed by the application of easily interpretable traditional ML techniques for making predictions. By doing so, they harness domain knowledge to ensure the biological interpretability of the approach [73].

Development of the Field

The domain of AI in lymphoma histopathology diagnosis and prognosis is experiencing rapid growth, with a notable surge in scholarly publications since 2019. Most of these studies have leveraged deep neural networks for automated feature extraction and classification. In contrast, a smaller subset of research has employed conventional ML algorithms [24,30,33,40]. Recent investigations have expanded their scope to encompass a wider array of diagnostic outcomes, such as identifying specific lymphoma subtypes [17,21], predicting prognosis [35,36], and detecting genetic translocations [39,56].

Despite advancements regarding the role of AI in lymphoma research, there has not been a noticeable trend toward larger datasets, either in the number of slides analyzed or the number of participants involved. While there is no indication that more recent studies have adopted stricter internal validation methods, there is a positive shift toward increased external validation. Prior to 2019, no studies incorporated external validation on lymphoma data; however, 10 recent publications have done this. Although these validations often involve limited data, their inclusion signifies a step forward in research methodology.

These external validations are essential for the practical application of AI models in clinical settings, as they must be robust enough to handle the visual diversity present in data from various sources. This diversity can fluctuate both between and within different data centers over time. As the field evolves, we expect to see an increase in studies that rigorously validate their models against larger, high-quality, independent datasets. This includes transparent reporting on patient recruitment and selection processes, histopathology slide preparation, and digitization techniques. Such practices will be instrumental in mitigating the biases, limited reproducibility, and restricted generalizability that currently plague much of the research in this domain.

In the realm of oncology, there has been a marked surge in the number of published multimodal research studies since 2019. Notably, this growth includes only a handful of studies focused on multimodal research in lymphoma. Histopathological examination of tissue sections continues to be the cornerstone for cancer diagnosis, yet even seasoned pathologists often seek support from biomarker assays to enhance diagnostic accuracy. Multimodal research, which amalgamates diverse data types, such as genomics, proteomics, transcriptomics, and clinical data, has been pivotal in steering the trajectory of cancer research. This approach not only consolidates information from various sources but also transforms it, offering novel insights

XSL•FO
RenderX

[74,75]. For instance, a multimodal framework can facilitate a dual-modality analysis where a pathological image is processed to yield outputs from another domain, such as genetic sequencing or different imaging formats [63]. Such a model, when adeptly trained, can be leveraged to analyze pathological images from patients without overt medical conditions, extracting valuable indicators pertinent to precision medicine, including genetic sequences. As we anticipate advancements in high-throughput technologies, alongside the fields of transcriptomics, metabolomics, and proteomics, the future holds promise for an increase in the integration of multidimensional omics data with histopathological images in multimodal analyses.

This integration is poised to significantly bolster the clinical utility of AI, enabling more precise and personalized treatment strategies in oncology. By harnessing the power of multimodal research, we can expect a future where AI plays an even more integral role in clinical decision-making, thereby enhancing patient outcomes.

Current Limitations and Future Recommendations

A considerable amount of published research lacks the necessary clinical and pathological details to evaluate potential biases effectively. As a result, it is imperative for AI researchers to meticulously document the origins of their data. This transparency is crucial for understanding the variability within the dataset and for determining whether this diversity has been adequately addressed in the research methodology. Additionally, the modeling and analytical techniques employed must be thoroughly described to ensure the reliability and reproducibility of the findings.

To further improve reproducibility, we recommend that researchers provide code and data whenever possible. Digital pathology studies on lymphoma are currently constrained by the lack of publicly available data. Furthermore, WSIs from different centers can lead to significant heterogeneity in image data due to differences in scanning equipment across various centers. This variability can introduce confounding factors that complicate the task of developing robust AI models and assessing their generalizability. Such limitations can increase the risk of bias and confusion in research findings.

To mitigate these issues, during the image preprocessing stage, it is essential to diligently address and eliminate confounding factors arising from variations in staining, the presence of bubbles, and other artifacts. This meticulous attention to detail will enhance the accuracy and reliability of AI-driven diagnostic and analytical tools.

For AI to be clinically valuable, rigorous validation is paramount, particularly considering the constraints inherent in existing datasets. We recommend that researchers employ comprehensive analytical methods, such as cross-validation and external validation, to substantiate the robustness of their findings and the capacity of their models to extend to new datasets. Moreover, it is essential to report CIs for results, with a focus on the 95% CI, especially when comparing various models. This practice aids in discerning whether observed differences in model performance are genuinely significant or merely a product of random variation. By doing so, researchers can make more informed decisions about the efficacy and reliability of different AI models in clinical settings.

Researchers are recommended to follow regulatory guidance and standardized processes in the field of medical AI, such as reporting guidelines and quality assessment tools, such as QUADAS-AI, which provide a specific framework for assessing the risk of bias and the applicability of studies of diagnostic test accuracy in AI centers.

Moreover, a lack of interpretability is a barrier to the clinical adoption of AI. Therefore, we recommend that researchers strive to demonstrate the interpretability of their models to enhance the understanding and trust of clinical and pathological professionals.

Lymphoma is a diverse type of blood cancer that includes a variety of subtypes. From the literature we have reviewed, the current application of AI in lymphoma histopathology has primarily been focused on the diagnosis and prognosis of B-cell lymphoma. There are relatively fewer AI models for HL or T-cell lymphoma and other rare subtypes. This could be attributed to the higher incidence of B-cell lymphomas compared to other subtypes [1,2]. However, given the high aggressiveness and heterogeneity of T-cell lymphoma and HL [76], we hope that in future model development, more researchers will develop AI models for T-cell lymphoma and HL from a multidimensional clinical perspective.

Conclusion

Methodologically, the diagnostic and prognostic models of AI applied to lymphoma histopathology were evaluated, and the models were found to be biased. The enhancement of the accuracy of AI models and the acceleration of their clinical translation hinge on several critical aspects. These include the comprehensive reporting of data sources, the diversity of datasets, the study design, the transparency and interpretability of AI models, the use of cross-validation and external validation, and adherence to regulatory guidance and standardized processes in the field of medical AI.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant number: 82300216), the Fellowship of China Postdoctoral Science Foundation (grant number: 2024M760362), and the Sichuan Provincial Foundation of Science and Technology (grant number: 2025NSFSC0425).

XSL•FO
RenderX

Authors' Contributions

ZH, JL, and BH provided ideas for literature writing. YF drafted and developed literature search strategies, conducted the literature search, and wrote the manuscript. MZ, LX, and XD extracted field information and created the graphs. YL and JL discussed and arbitrated differences. ZH and BH directed the writing of the manuscript. All authors read and approved the final manuscript.
Author JL is the co-corresponding author of this paper and can be reached at: Phase 1 Clinical Trial Unit, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu, China; liujinyi@scszlyy.org.cn.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategies in PubMed, Cochrane Library, and Web of Science.
[DOCX File , 18 KB - jmir_v27i1e62851_app1.docx]

Multimedia Appendix 2
Description of quality assessment based on Quality Assessment of Diagnostic Accuracy Studies-AI (QUADAS-AI) domains used to evaluate the methodological quality of the studies included.
[DOCX File , 16 KB - jmir_v27i1e62851_app2.docx]

Multimedia Appendix 3
Prediction Model Risk of Bias Assessment Tool (PROBAST) table.
[XLS File (Microsoft Excel File), 152 KB - jmir_v27i1e62851_app3.xls]

Multimedia Appendix 4
Quality Assessment of Diagnostic Accuracy Studies-AI (QUADAS-AI) table.
[XLS File (Microsoft Excel File), 93 KB - jmir_v27i1e62851_app4.xls]

Multimedia Appendix 5
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 32 KB - jmir_v27i1e62851_app5.docx]

References

1. Huang J, Pang WS, Lok V, Zhang L, Lucero-Prisno DE, Xu W, NCD Global Health Research Group, Association of Pacific Rim Universities (APRU). Incidence, mortality, risk factors, and trends for Hodgkin lymphoma: a global data analysis. J Hematol Oncol 2022 May 11;15(1):57 [FREE Full text] [doi: 10.1186/s13045-022-01281-9] [Medline: 35546241]
2. Chu Y, Liu Y, Fang X, Jiang Y, Ding M, Ge X, et al. The epidemiological patterns of non-Hodgkin lymphoma: global estimates of disease burden, risk factors, and temporal trends. Front Oncol 2023;13:1059914 [FREE Full text] [doi: 10.3389/fonc.2023.1059914] [Medline: 37333805]
3. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2133-2161 [FREE Full text] [doi: 10.1016/S0140-6736(24)00757-8] [Medline: 38642570]
4. Pathological Technology Expert Group of Medical Technician Committee of Chinese Medical Doctor Association, Standardization Department of Pathological Equipment Branch of China Medical Equipment Association, Pathological Technology Group, Pathology Professional Committee. [Chinese expert consensus on detection techniques and interpretation of lymphoma gene rearrangement (2023 version)]. Zhonghua Bing Li Xue Za Zhi 2023 Jun 08;52(6):558-565. [doi: 10.3760/cma.j.cn112151-20230106-00010] [Medline: 37263919]
5. Perry C, Greenberg O, Haberman S, Herskovitz N, Gazy I, Avinoam A, et al. Image-based deep learning detection of high-grade b-cell lymphomas directly from hematoxylin and eosin images. Cancers (Basel) 2023 Oct 29;15(21):5205 [FREE Full text] [doi: 10.3390/cancers15215205] [Medline: 37958379]
6. Ferry JA. Scientific advances and the evolution of diagnosis, subclassification and treatment of lymphoma. Arch Med Res 2020 Nov;51(8):749-764. [doi: 10.1016/j.arcmed.2020.05.022] [Medline: 32553461]
7. Baidoshvili A, Bucur A, van Leeuwen J, van der Laak J, Kluin P, van Diest PJ. Evaluating the benefits of digital pathology implementation: time savings in laboratory logistics. Histopathology 2018 Nov;73(5):784-794. [doi: 10.1111/his.13691] [Medline: 29924891]

XSL•FO
RenderX

8. Stenzinger A, Alber M, Allgäuer M, Jurmeister P, Bockmayr M, Budczies J, et al. Artificial intelligence and pathology: From principles to practice and future applications in histomorphology and molecular profiling. Semin Cancer Biol 2022 Sep;84:129-143. [doi: 10.1016/j.semcancer.2021.02.011] [Medline: 33631297]
9. Raciti P, Sue J, Retamero JA, Ceballos R, Godrich R, Kunz JD, et al. Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. Arch Pathol Lab Med 2023 Oct 01;147(10):1178-1185 [FREE Full text] [doi: 10.5858/arpa.2022-0066-OA] [Medline: 36538386]
10. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan 01;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]
11. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
12. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, QUADAS-C Group, et al. QUADAS-C: A tool for assessing risk of bias in comparative diagnostic accuracy studies. Ann Intern Med 2021 Nov;174(11):1592-1599 [FREE Full text] [doi: 10.7326/M21-2234] [Medline: 34698503]
13. Achi HE, Belousova T, Chen L, Wahed A, Wang I, Hu Z, et al. Automated diagnosis of lymphoma with digital pathology images using deep learning. Ann Clin Lab Sci 2019 Mar;49(2):153-160. [Medline: 31028058]
14. Steinbuss G, Kriegsmann M, Zgorzelski C, Brobeil A, Goeppert B, Dietrich S, et al. Deep learning for the classification of non-Hodgkin lymphoma on histopathological images. Cancers (Basel) 2021 May 17;13(10):2419 [FREE Full text] [doi: 10.3390/cancers13102419] [Medline: 34067726]
15. Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images. GitHub. URL: https://github.com/AG-Computational-Diagnostic/pacltune-pup [accessed 2025-01-24]
16. Miyoshi H, Sato K, Kabeya Y, Yonezawa S, Nakano H, Takeuchi Y, et al. Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma. Lab Invest 2020 Oct 29;100(10):1300-1310 [FREE Full text] [doi: 10.1038/s41374-020-0442-3] [Medline: 32472096]
17. Syrykh C, Abreu A, Amara N, Siegfried A, Maisongrosse V, Frenois FX, et al. Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning. NPJ Digit Med 2020;3:63 [FREE Full text] [doi: 10.1038/s41746-020-0272-0] [Medline: 32377574]
18. Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning. GitHub. URL: https://github.com/ArnaudAbreu/DiagFLFH [accessed 2025-01-24]
19. Sereda S, Shankar A, Weber L, Ramsay AD, Hall GW, Hayward J, et al. Digital pathology in pediatric nodular lymphocyte-predominant Hodgkin lymphoma: correlation with treatment response. Blood Adv 2023 Oct 24;7(20):6285-6289 [FREE Full text] [doi: 10.1182/bloodadvances.2023010652] [Medline: 37611165]
20. Zhang J, Cui W, Guo X, Wang B, Wang Z. Classification of digital pathological images of non-Hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis. Med Phys 2020 Sep;47(9):4241-4253. [doi: 10.1002/mp.14357] [Medline: 32593219]
21. Li D, Bledsoe JR, Zeng Y, Liu W, Hu Y, Bi K, et al. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. Nat Commun 2020 Nov 26;11(1):6004 [FREE Full text] [doi: 10.1038/s41467-020-19817-3] [Medline: 33244018]
22. Li D, Bledsoe J, Zeng Y, Liu W, Hu Y, Bi K, et al. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. UMass Chan Medical School. URL: https://fts.umassmed.edu/ [accessed 2025-01-24]
23. Zhang X, Zhang K, Jiang M, Yang L. Research on the classification of lymphoma pathological images based on deep residual neural network. Technol Health Care 2021;29(S1):335-344 [FREE Full text] [doi: 10.3233/THC-218031] [Medline: 33682770]
24. Yu W, Li C, Wang R, Yeh C, Chuang S. Machine learning based on morphological features enables classification of primary intestinal t-cell lymphomas. Cancers (Basel) 2021 Oct 30;13(21):5463 [FREE Full text] [doi: 10.3390/cancers13215463] [Medline: 34771625]
25. Takagi Y, Hashimoto N, Masuda H, Miyoshi H, Ohshima K, Hontani H, et al. Transformer-based personalized attention mechanism for medical images with clinical records. J Pathol Inform 2023;14:100185 [FREE Full text] [doi: 10.1016/j.jpi.2022.100185] [Medline: 36691660]
26. Hamdi M, Senan EM, Jadhav ME, Olayah F, Awaji B, Alalayah KM. Hybrid models based on fusion features of a CNN and handcrafted features for accurate histopathological image analysis for diagnosing malignant lymphomas. Diagnostics (Basel) 2023 Jul 04;13(13):2258 [FREE Full text] [doi: 10.3390/diagnostics13132258] [Medline: 37443652]
27. Hamdi M, Senan EM, Jadhav ME, Olayah F, Awaji B, Alalayah KM. Hybrid models based on fusion features of a CNN and handcrafted features for accurate histopathological image analysis for diagnosing malignant lymphomas (Dataset). kaggle. URL: https://www.kaggle.com/datasets/obulisainaren/multi-cancer [accessed 2025-01-24]

XSL•FO
RenderX

28. Mohlman JS, Leventhal SD, Hansen T, Kohan J, Pascucci V, Salama ME. Improving augmented human intelligence to distinguish Burkitt lymphoma from diffuse large B-cell lymphoma cases. Am J Clin Pathol 2020 May 05;153(6):743-759. [doi: 10.1093/ajcp/aqaa001] [Medline: 32067039]
29. Karabulut YY, Dinç U, Köse E, Türsen Ü. Deep learning as a new tool in the diagnosis of mycosis fungoides. Arch Dermatol Res 2023 Jul;315(5):1315-1322. [doi: 10.1007/s00403-022-02521-1] [Medline: 36571610]
30. do Nascimento MZ, Martins AS, Azevedo Tosta TA, Neves LA. Lymphoma images analysis using morphological and non-morphological descriptors for classification. Comput Methods Programs Biomed 2018 Sep;163:65-77. [doi: 10.1016/j.cmpb.2018.05.035] [Medline: 30119858]
31. Zhu H, Jiang H, Li S, Li H, Pei Y. A novel multispace image reconstruction method for pathological image classification based on structural information. Biomed Res Int 2019;2019:3530903 [FREE Full text] [doi: 10.1155/2019/3530903] [Medline: 31111048]
32. Hashimoto N, Ko K, Yokota T, Kohno K, Nakaguro M, Nakamura S, et al. Subtype classification of malignant lymphoma using immunohistochemical staining pattern. Int J Comput Assist Radiol Surg 2022 Jul;17(7):1379-1389 [FREE Full text] [doi: 10.1007/s11548-021-02549-0] [Medline: 35147848]
33. Michail E, Dimitropoulos K, Koletsa T, Kostopoulos I, Grammalidis N. Morphological and textural analysis of centroblasts in low-thickness sliced tissue biopsies of follicular lymphoma. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:3374-3377. [doi: 10.1109/EMBC.2014.6944346] [Medline: 25570714]
34. Kornaropoulos EN, Niazi MKK, Lozanski G, Gurcan MN. Histopathological image analysis for centroblasts classification through dimensionality reduction approaches. Cytometry A 2014 Mar;85(3):242-255 [FREE Full text] [doi: 10.1002/cyto.a.22432] [Medline: 24376080]
35. Chuang W, Yu W, Lee Y, Zhang Q, Chang H, Shih L, et al. Deep learning-based nuclear morphometry reveals an independent prognostic factor in mantle cell lymphoma. Am J Pathol 2022 Dec;192(12):1763-1778 [FREE Full text] [doi: 10.1016/j.ajpath.2022.08.006] [Medline: 36150505]
36. Vrabac D, Smit A, Rojansky R, Natkunam Y, Advani RH, Ng AY, et al. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. Sci Data 2021 May 20;8(1):135 [FREE Full text] [doi: 10.1038/s41597-021-00915-w] [Medline: 34017010]
37. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. GitHub. URL: https://github.com/stanfordmlgroup/DLBCL-Morph [accessed 2025-01-24]
38. Motmaen I, Sereda S, Brobeil A, Shankar A, Braeuninger A, Hasenclever D, et al. Deep-learning based classification of a tumor marker for prognosis on Hodgkin's disease. Eur J Haematol 2023 Nov;111(5):722-728. [doi: 10.1111/ejh.14066] [Medline: 37549921]
39. Swiderska-Chadaj Z, Hebeda KM, van den Brand M, Litjens G. Artificial intelligence to detect MYC translocation in slides of diffuse large B-cell lymphoma. Virchows Arch 2021 Sep;479(3):617-621 [FREE Full text] [doi: 10.1007/s00428-020-02931-4] [Medline: 32979109]
40. Tagami M, Nishio M, Katsuyama-Yoshikawa A, Misawa N, Sakai A, Haruna Y, et al. Machine learning model with texture analysis for automatic classification of histopathological images of ocular adnexal mucosa-associated lymphoid tissue lymphoma of two different origins. Curr Eye Res 2023 Dec;48(12):1195-1202. [doi: 10.1080/02713683.2023.2246696] [Medline: 37566457]
41. Irshaid L, Bleiberg J, Weinberger E, Garritano J, Shallis RM, Patsenker J, et al. Histopathologic and machine deep learning criteria to predict lymphoma transformation in bone marrow biopsies. Arch Pathol Lab Med 2022 Jan 02;146(2):182-193 [FREE Full text] [doi: 10.5858/arpa.2020-0510-OA] [Medline: 34086849]
42. El Hussein S, Chen P, Medeiros LJ, Wistuba II, Jaffray D, Wu J, et al. Artificial intelligence strategy integrating morphologic and architectural biomarkers provides robust diagnostic accuracy for disease progression in chronic lymphocytic leukemia. J Pathol 2022 Jan;256(1):4-14 [FREE Full text] [doi: 10.1002/path.5795] [Medline: 34505705]
43. Chen P, El Hussein S, Xing F, Aminu M, Kannapiran A, Hazle JD, et al. Chronic lymphocytic leukemia progression diagnosis with intrinsic cellular patterns via unsupervised clustering. Cancers (Basel) 2022 May 13;14(10):2398 [FREE Full text] [doi: 10.3390/cancers14102398] [Medline: 35626003]
44. Zhang X, Zhao Z, Wang R, Chen H, Zheng X, Liu L, et al. A multicenter proof-of-concept study on deep learning-based intraoperative discrimination of primary central nervous system lymphoma. Nat Commun 2024 May 04;15(1):3768 [FREE Full text] [doi: 10.1038/s41467-024-48171-x] [Medline: 38704409]
45. A multicenter proof-of-concept study on deep learning-based intraoperative discrimination of primary central nervous system lymphoma. GitHub. URL: https://github.com/Kepler1647b/LGNet/tree/main [accessed 2025-01-24]
46. Tavolara TE, Niazi MKK, Feldman AL, Jaye DL, Flowers C, Cooper LAD, et al. Translating prognostic quantification of c-MYC and BCL2 from tissue microarrays to whole slide images in diffuse large B-cell lymphoma using deep learning. Diagn Pathol 2024 Jan 19;19(1):17 [FREE Full text] [doi: 10.1186/s13000-023-01425-6] [Medline: 38243330]
47. Translating prognostic quantification of c-MYC and BCL2 from tissue microarrays to whole slide images in diffuse large B-cell lymphoma using deep learning. GitHub. URL: https://github.com/cialab/tma_to_wsi [accessed 2025-01-24]

XSL•FO
RenderX

48. Yan F, Da Q, Yi H, Deng S, Zhu L, Zhou M, et al. Artificial intelligence-based assessment of PD-L1 expression in diffuse large B cell lymphoma. NPJ Precis Oncol 2024 Mar 27;8(1):76 [FREE Full text] [doi: 10.1038/s41698-024-00577-y] [Medline: 38538739]
49. Artificial intelligence-based assessment of PD-L1 expression in diffuse large B cell lymphoma. GitHub. URL: https://github.com/yanfang-research/Digital-PD-L1-Scoring [accessed 2025-01-24]
50. Lee JH, Song G, Lee J, Kang S, Moon KM, Choi Y, et al. Prediction of immunochemotherapy response for diffuse large B-cell lymphoma using artificial intelligence digital pathology. J Pathol Clin Res 2024 May;10(3):e12370 [FREE Full text] [doi: 10.1002/2056-4538.12370] [Medline: 38584594]
51. Duan L, He Y, Guo W, Du Y, Yin S, Yang S, et al. Machine learning-based pathomics signature of histology slides as a novel prognostic indicator in primary central nervous system lymphoma. J Neurooncol 2024 Jun;168(2):283-298 [FREE Full text] [doi: 10.1007/s11060-024-04665-8] [Medline: 38557926]
52. Quan J, Ye J, Lan J, Wang J, Hu Z, Guo Z, et al. A deep learning model fusion algorithm for the diagnosis of gastric Mucosa-associated lymphoid tissue lymphoma. Biomedical Signal Processing and Control 2024 Jun;92:106064. [doi: 10.1016/j.bspc.2024.106064]
53. Al-Mekhlafi ZG, Senan EM, Mohammed BA, Alazmi M, Alayba AM, Alreshidi A, et al. Diagnosis of histopathological images to distinguish types of malignant lymphomas using hybrid techniques based on fusion features. Electronics 2022 Sep 10;11(18):2865. [doi: 10.3390/electronics11182865]
54. Somaratne U, Wong K, Parry J, Sohel F, Wang X, Laga H. Improving Follicular Lymphoma Identification using the Class of Interest for Transfer Learning. 2019 Presented at: 2019 Digital Image Computing: Techniques and Applications (DICTA); December 02-04, 2019; Perth, WA, Australia. [doi: 10.1109/DICTA47822.2019.8946075]
55. Codella N, Moradi M, Matasar M, Sveda-Mahmood T, Smith J. Lymphoma diagnosis in histopathology using a multi-stage visual learning approach. In: Proceedings Volume 9791, Medical Imaging 2016: Digital Pathology. 2016 Presented at: SPIE Medical Imaging; February 27-March 3, 2016; San Diego, CA. [doi: 10.1117/12.2217158]
56. Swiderska-Chadaj Z, Hebeda K, van den Brand M, Litjens G. Predicting MYC translocation in HE specimens of diffuse large B-cell lymphoma through deep learning. In: Proceedings Volume 11320, Medical Imaging 2020: Digital Pathology. 2020 Presented at: SPIE Medical Imaging; February 15-20, 2020; Houston, TX. [doi: 10.1117/12.2549650]
57. Basu S, Agarwal R, Srivastava V. Deep discriminative learning model with calibrated attention map for the automated diagnosis of diffuse large B-cell lymphoma. Biomedical Signal Processing and Control 2022 Jul;76:103728. [doi: 10.1016/j.bspc.2022.103728]
58. Shankar V, Yang X, Krishna V, Tan B, Silva O, Rojansky R, et al. LymphoML: An interpretable artificial intelligence-based method identifies morphologic features that correlate with lymphoma subtype. Proceedings of Machine Learning Research 2023;225:528-558 [FREE Full text]
59. LymphoML: An interpretable artificial intelligence-based method identifies morphologic features that correlate with lymphoma subtype. GitHub. URL: https://github.com/rajpurkarlab/LymphoML [accessed 2025-01-24]
60. Soltane S, Alshreef S, Eldin SM. Classification and diagnosis of lymphoma's histopathological images using transfer learning. Computer Systems Science and Engineering 2022;40(2):629-644. [doi: 10.32604/csse.2022.019333]
61. Tagami M, Nishio M, Yoshikawa A, Misawa N, Sakai A, Haruna Y, et al. Artificial intelligence-based differential diagnosis of orbital MALT lymphoma and IgG4 related ophthalmic disease using hematoxylin-eosin images. Graefes Arch Clin Exp Ophthalmol 2024 Oct 03;262(10):3355-3366. [doi: 10.1007/s00417-024-06501-1] [Medline: 38700592]
62. Samsi S, Krishnamurthy AK, Gurcan MN. An efficient computational framework for the analysis of whole slide images: application to follicular lymphoma immunohistochemistry. J Comput Sci 2012 Sep 01;3(5):269-279 [FREE Full text] [doi: 10.1016/j.jocs.2012.01.009] [Medline: 22962572]
63. Qiao Y, Zhao L, Luo C, Luo Y, Wu Y, Li S, et al. Multi-modality artificial intelligence in digital pathology. Brief Bioinform 2022 Nov 19;23(6):bbac367 [FREE Full text] [doi: 10.1093/bib/bbac367] [Medline: 36124675]
64. Ruchti A, Neuwirth A, Lowman AK, Duenweg SR, LaViolette PS, Bukowy JD. Homologous point transformer for multi-modality prostate image registration. PeerJ Comput Sci 2022;8:e1155 [FREE Full text] [doi: 10.7717/peerj-cs.1155] [Medline: 36532813]
65. Dong W, Yang Q, Wang J, Xu L, Li X, Luo G, et al. Multi-modality attribute learning-based method for drug-protein interaction prediction based on deep neural network. Brief Bioinform 2023 May 19;24(3):bbad161. [doi: 10.1093/bib/bbad161] [Medline: 37114624]
66. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 2021 Jun 01;5(6):555-570 [FREE Full text] [doi: 10.1038/s41551-020-00682-w] [Medline: 33649564]
67. Frankenstein Z, Uraoka N, Aypar U, Aryeequaye R, Rao M, Hameed M, et al. Automated 3D scoring of fluorescence in situ hybridization (FISH) using a confocal whole slide imaging scanner. Appl Microsc 2021 Apr 09;51(1):4 [FREE Full text] [doi: 10.1186/s42649-021-00053-y] [Medline: 33835321]
68. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal 2016 Oct;33:170-175 [FREE Full text] [doi: 10.1016/j.media.2016.06.037] [Medline: 27423409]

XSL•FO
RenderX

69. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
70. Xu Z, Li Y, Wang Y, Zhang S, Huang Y, Yao S, et al. A deep learning quantified stroma-immune score to predict survival of patients with stage II-III colorectal cancer. Cancer Cell Int 2021 Oct 30;21(1):585 [FREE Full text] [doi: 10.1186/s12935-021-02297-w] [Medline: 34717647]
71. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. Am J Surg Pathol 2018 Dec;42(12):1636-1646 [FREE Full text] [doi: 10.1097/PAS.0000000000001151] [Medline: 30312179]
72. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. Arch Pathol Lab Med 2019 Jul;143(7):859-868 [FREE Full text] [doi: 10.5858/arpa.2018-0147-OA] [Medline: 30295070]
73. Wang X, Janowczyk A, Zhou Y, Thawani R, Fu P, Schalper K, et al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H-E images. Sci Rep 2017 Oct 19;7(1):13543 [FREE Full text] [doi: 10.1038/s41598-017-13773-7] [Medline: 29051570]
74. Zhao L, Dong Q, Luo C, Wu Y, Bu D, Qi X, et al. DeepOmix: A scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis. Comput Struct Biotechnol J 2021;19:2719-2725 [FREE Full text] [doi: 10.1016/j.csbj.2021.04.067] [Medline: 34093987]
75. d'Este SH, Nielsen MB, Hansen AE. Visualizing glioma infiltration by the combination of multimodality imaging and artificial intelligence, a systematic review of the literature. Diagnostics (Basel) 2021 Mar 25;11(4):592 [FREE Full text] [doi: 10.3390/diagnostics11040592] [Medline: 33806195]
76. Horwitz SM, Ansell S, Ai WZ, Barnes J, Barta SK, Brammer J, et al. T-Cell lymphomas, version 2.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2022 Mar;20(3):285-308. [doi: 10.6004/jnccn.2022.0015] [Medline: 35276674]

Abbreviations

AI: artificial intelligence
AUC: area under the receiver operating characteristic curve
CLL: chronic lymphocytic leukemia
CNN: convolutional neural network
DL: deep learning
DLBCL: diffuse large B-cell lymphoma
FL: follicular lymphoma
HE: hematoxylin-eosin
HL: Hodgkin lymphoma
IHC: immunohistochemical
MCL: mantle cell lymphoma
ML: machine learning
NHL: non-Hodgkin lymphoma
PROBAST: Prediction Model Risk of Bias Assessment Tool
QUADAS-AI: Quality Assessment of Diagnostic Accuracy Studies-AI
SVM: support vector machine
WSI: whole slide image

Edited by A Schwartz; submitted 03.06.24; peer-reviewed by M Zhang, S Marletta; comments to author 11.09.24; revised version received 03.11.24; accepted 07.01.25; published 14.02.25.

Please cite as:
Fu Y, Huang Z, Deng X, Xu L, Liu Y, Zhang M, Liu J, Huang B
Artificial Intelligence in Lymphoma Histopathology: Systematic Review
J Med Internet Res 2025;27:e62851
URL: https://www.jmir.org/2025/1/e62851
doi:10.2196/62851
PMID:

©Yao Fu, Zongyao Huang, Xudong Deng, Linna Xu, Yang Liu, Mingxing Zhang, Jinyi Liu, Bin Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.02.2025. This is an open-access article distributed under

XSL•FO
RenderX

the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Future Use of AI in Diagnostic Medicine: 2-Wave Cross-Sectional Survey Study

Bernardo Pereira Cabral[1,2], PhD; Luiza Amara Maciel Braga[1], PhD; Carlos Gilbert Conte Filho[3], PhD; Bruno Penteado[4], PhD; Sandro Luis Freire de Castro Silva[5,6], PhD; Leonardo Castro[4,7], PhD; Marcelo Fornazin[4,7], PhD; Fabio Mota[1], PhD

[1]Cellular Communication Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

[2]Department of Economics, Faculty of Economics, Federal University of Bahia, Salvador, Brazil

[3]Department of Economics, Federal University of Santa Maria, Santa Maria, Brazil

[4]Fiocruz Strategy for the 2030 Agenda, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

[5]National Cancer Institute, Rio de Janeiro, Brazil

[6]Graduate Program in Management and Strategy, Federal Rural University of Rio de Janeiro, Seropedica, Brazil

[7]National School of Public Health, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

Corresponding Author:
Fabio Mota, PhD
Cellular Communication Laboratory
Oswaldo Cruz Institute
Oswaldo Cruz Foundation
Avenida Brasil, 4365
Manguinhos
Rio de Janeiro, 21040-900
Brazil
Phone: 55 2125984220
Email: fabio.mota@fiocruz.br

Abstract

Background: The rapid evolution of artificial intelligence (AI) presents transformative potential for diagnostic medicine, offering opportunities to enhance diagnostic accuracy, reduce costs, and improve patient outcomes.

Objective: This study aimed to assess the expected future impact of AI on diagnostic medicine by comparing global researchers' expectations using 2 cross-sectional surveys.

Methods: The surveys were conducted in September 2020 and February 2023. Each survey captured a 10-year projection horizon, gathering insights from >3700 researchers with expertise in AI and diagnostic medicine from all over the world. The survey sought to understand the perceived benefits, integration challenges, and evolving attitudes toward AI use in diagnostic settings.

Results: Results indicated a strong expectation among researchers that AI will substantially influence diagnostic medicine within the next decade. Key anticipated benefits include enhanced diagnostic reliability, reduced screening costs, improved patient care, and decreased physician workload, addressing the growing demand for diagnostic services outpacing the supply of medical professionals. Specifically, x-ray diagnosis, heart rhythm interpretation, and skin malignancy detection were identified as the diagnostic tools most likely to be integrated with AI technologies due to their maturity and existing AI applications. The surveys highlighted the growing optimism regarding AI's ability to transform traditional diagnostic pathways and enhance clinical decision-making processes. Furthermore, the study identified barriers to the integration of AI in diagnostic medicine. The primary challenges cited were the difficulties of embedding AI within existing clinical workflows, ethical and regulatory concerns, and data privacy issues. Respondents emphasized uncertainties around legal responsibility and accountability for AI-supported clinical decisions, data protection challenges, and the need for robust regulatory frameworks to ensure safe AI deployment. Ethical concerns, particularly those related to algorithmic transparency and bias, were noted as increasingly critical, reflecting a heightened awareness of the potential risks associated with AI adoption in clinical settings. Differences between the 2 survey waves indicated a growing focus on ethical and regulatory issues, suggesting an evolving recognition of these challenges over time.

XSL•FO
RenderX

Conclusions: Despite these barriers, there was notable consistency in researchers' expectations across the 2 survey periods, indicating a stable and sustained outlook on AI's transformative potential in diagnostic medicine. The findings show the need for interdisciplinary collaboration among clinicians, AI developers, and regulators to address ethical and practical challenges while maximizing AI's benefits. This study offers insights into the projected trajectory of AI in diagnostic medicine, guiding stakeholders, including health care providers, policy makers, and technology developers, on navigating the opportunities and challenges of AI integration.

(J Med Internet Res 2025;27:e53892) doi:10.2196/53892

KEYWORDS

artificial intelligence; AI; diagnostic medicine; survey research; researcher opinion; future

Introduction

Background

Artificial intelligence (AI) is a broad term that covers various technologies such as machine learning, natural language processing, neural networks, and specialized rule-based systems [1,2]. It has emerged as a new paradigm in health care [3-7], with applications in disease risk assessment, treatment outcomes prediction, medical complication prevention or mitigation, clinical research, drug development, and patient care [8]. The AI market in health care is expanding rapidly, with a market size of US $26.69 billion in 2024 and projected to grow to approximately US $613.81 billion by 2034 [9]. In the United States alone, the Food and Drug Administration received <500 submissions between 2016 and 2022 [10].

Specifically, AI is expected to promote major changes in diagnostic medicine [4,11-13]. This branch of medicine focuses on identifying diseases and medical conditions by evaluating symptoms, physical signs, and test results [14,15]. Expected benefits of adopting AI include more reliable diagnoses, reduced screening costs, improved health care access, and reduced physician workload [13,16-18]. AI's ability to analyze large datasets quickly and accurately can also facilitate early detection of diseases [19-21] and support personalized treatment decisions [22], offering opportunities to improve patient outcomes significantly [23]. In addition, AI-driven tools can assist in managing diagnostic workflows by automating routine tasks [23,24], allowing physicians to focus on more complex cases. These benefits are especially desirable as the demand for imaging diagnoses has been rising faster than the supply of physicians for decades [3,11,17].

However, several factors can influence the direction and speed of changes in diagnostic medicine. Among these, one such factor is ethical and regulatory issues, which include difficulties in accessing and sharing databases with patient information [1,19], data validation and auditing [20-22], uncertainties about legal liability for the use of algorithms [23-25], and bias of algorithmic underrepresentation [20,26]. Technical challenges, such as the integration of AI systems into existing health care infrastructure [27] and the need for ongoing training of AI models to adapt to new medical knowledge [28], also pose significant hurdles. In addition, the lack of standardization in AI development and deployment can lead to inconsistent performance across different clinical settings [29,30]. All these factors mean that the future of AI use in diagnostic medicine is still quite uncertain, despite its potential.

Objectives

What, then, is the future of AI use in diagnostic medicine? This study aimed to anticipate future possibilities of AI use in diagnostic medicine, comparing researchers' expectations over time. For this purpose, we conducted 2 global cross-sectional surveys with authors of recent scientific publications on AI and diagnostic medicine, indexed in the Web of Science (WoS) Core Collection database. Some previous studies provided literature reviews to anticipate expected changes in health care using AI [1,3,8,12,18,27,28,31-36]. More specific studies on AI applications in diagnostic medicine focused on specific diagnostic tools, such as medical imaging [29], electrocardiograms [30], or cancer classification [37]. However, none of them comprehensively addressed aspects related to the future use of AI in diagnostic medicine and attempted to follow up on possible changes in the expected future. Our study addresses this gap by surveying the opinions of >3000 researchers worldwide at 2 different moments, in September 2020 and February 2023, with 28 months in between. A similar approach has been adopted in other studies to assess changes in attitudes or perceptions over time [38-41]. The survey participants shared their expectations on aspects related to (1) the probability of occurrence of events pointed out by the scientific literature, (2) the integration between AI and various types of diagnostic instruments, and (3) the main barriers to the use of AI in diagnostic medicine.

This study provides a unique perspective on the expected changes in diagnostic medicine due to the integration of AI and diagnostic medicine. Our findings may help health care providers and policy makers make informed decisions on integrating AI into clinical practice. In addition, our study may interest those investing in research and development and those expected to apply AI technologies in diagnostic medicine.

Methods

Literature Review

A literature review was conducted to explore the challenges and opportunities of AI applications in diagnostic medicine and build the questionnaire. We searched for recent articles, editorials, and reviews on this topic in WoS-indexed journals. The publications were identified using the search strategy provided in Textbox 1.

XSL•FO
RenderX

Textbox 1. The search strategy used to identify publications in Web of Science (WoS)–indexed journals.

TS=("Artificial intelligence" OR "Computational Intelligence" OR "Machine Intelligence" OR "Computer Reasoning" OR AI OR "Computer Vision System*") AND TS=(future* OR foresight* OR forthcoming* OR prospective* OR imminent*) AND TS=(diagnostics OR medicine OR "clinical practice")

AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Editorial Material OR Review)

Indexes: SCI-Expanded; timespan: 2015 to 2020

The search strategy combined thesaurus terms related to AI, medical diagnosis obtained from the Medical Subject Headings (MeSH) of the US National Library of Medicine, as well as free-text words for terms related to the future. In the WoS advanced search mode, we used the tag topic to search for those terms in the titles, abstracts, and keywords of articles, editorial materials, and reviews published between 2015 and May 2020 and indexed in the Science Citation Index (SCI)-Expanded. The SCI-Expanded was used to retrieve documents published in journals of sciences, and the period was set to obtain recent information on applications of AI in diagnostic medicine.

The search was conducted in May 2020 and recovered 536 publication records, imported in plain text format to the data- and text-mining software VantagePoint (version 11.0; Search Technology Inc). After reading their titles and abstracts, we selected 65 publication records for further analysis. These records were then imported into the software Citavi (version 6.1; Swiss Academic Software GmbH), where we read the texts fully and managed the references. Of these 65 publications, we selected 27 (41%) that formed the basis of the literature review and the survey questionnaire [1-3,5,8,11,16-18,20,23,29,33-35,42-52]. As the questionnaire's content could not be changed to allow for comparing the results from both waves, we did not update the literature review.

Survey Design

We followed the Checklist for Reporting Results of Internet E-Surveys [53], and the detailed checklist is provided in Multimedia Appendix 1. The questionnaire considered a horizon of 10 years and was divided into 6 parts. The first part introduced the study's aim, data collection and treatment procedures, anonymity and confidentiality guarantees, voluntary participation conditions, and informed consent. The second part asked about the respondents' knowledge regarding diagnostic medicine. Only those who reported having good or some knowledge qualified for the survey. Those who reported no knowledge were disqualified and excluded from the survey; hence, they were prevented from proceeding further with the main questionnaire. The third part asked about the respondents' expectations of AI's impact on diagnostic medicine in terms of (1) radical changes, (2) future outcomes (eg, reducing physicians' workload), and (3) integration with diagnostic instruments (eg, interpreting brain magnetic resonance imaging). The fourth part asked about the general barriers to using AI in diagnostic medicine. The respondents had to choose 1 of the 4 options as the most important barrier (eg, the difficulty of incorporation into clinical practice). On the basis of their choice, they were directed to a question about specific barriers within that category, where they had to select one as the most important barrier (eg, conflicts between AI and other clinical strategies). The bibliographic references for each question in these parts of the questionnaire are listed in Multimedia Appendix 2.

The fifth and sixth parts were optional and complementary to the questionnaire as they were not part of the survey's core. Thus, the data collected were not considered when calculating the number of fully completed questionnaires. The fifth part was an open-ended question where respondents could provide new information about the use of AI in diagnostic medicine or submit comments on the survey. The last part consisted of 5 demographic questions. As the results of this type of study were not influenced by the respondents' demographics [54-57], these questions were included to provide an overview of respondents' academic degrees, professional occupations, institutional affiliation, professional experience, and the region where they live. The questionnaire consisted of 10 pages.

Respondent Recruitment

Respondent recruitment followed a detailed and systematic process. The first wave (W1) and second wave (W2) respondents were authors of articles or review articles related to AI and diagnostic medicine published in peer-reviewed journals indexed in WoS's SCI-Expanded between January 1, 2015, and September 20, 2020 (W1), and between September 21, 2020, and December 30, 2022 (W2). To identify the participants, we used the query provided in Textbox 2 using the WoS advanced search mode.

Textbox 2. The query used to identify participants using Web of Science advanced search mode.

TS=("Artificial intelligence" OR "Computational Intelligence" OR "Machine Intelligence" OR "Computer Reasoning" OR AI OR "Computer Vision System*") AND TS=(Diagnos* OR Medicine OR Medical OR Ultrasound* OR Ultrasonograph* OR "Ultrasonic Diagnos*" OR Echotomograph* OR Echograph* OR "Radionuclide Imaging" OR "Radioisotope Scanning" OR radiograph* OR Roentgenograph* OR "X-ray")

Refined by DOCUMENT TYPES: (ARTICLE OR REVIEW)

Indexes: SCI-Expanded

The query was designed to capture a broad range of AI applications in diagnostic medicine. The thesaurus terms related to AI and diagnostic medicine were collected in MeSH. The query was set to retrieve recently published peer-reviewed articles and review articles. In W1, conducted in September 2020, we retrieved 15,084 publication records. These records were processed using the VantagePoint software to extract 23,053 authors' emails. The extracted emails were compiled

XSL•FO
RenderX

into a CSV file containing the authors' names, emails, and titles of their respective articles. An in-house Python (Python Software Foundation) script was used to cross-check and link 80.11% (18,469/23,053) of the extracted emails to their respective authors, enabling a more personalized approach to sending the survey invitations.

Similarly, in W2, conducted in December 2022, we retrieved 35,146 publication records, leading to 65,149 author emails. We linked 80.03% (52,138/65,149) of the emails to their respective authors using the same process. This consistent approach ensured the recruitment of a wide-reaching and diverse sample of respondents. The use of personalized invitations, sent directly to authors, aimed to enhance the engagement and response rate. The recruitment process was systematic and designed to ensure the survey's reach across key stakeholders in the AI and diagnostic medicine fields. The procedures used were similar to those of other studies using the same method [39,54,56-60], especially those with 2 survey waves [38]. A summarized and visual representation of this process of recruiting respondents is provided in Multimedia Appendix 3.

In choosing this methodological approach, we aimed to capture the rapidly evolving landscape of AI in diagnostic medicine. The decision to conduct 2 waves of surveys with different cohorts was driven by the need to reflect the dynamic nature of this field. AI technologies and their applications in medicine are advancing at a fast pace, and by expanding the cohort in W2, we sought to provide a broader and more current perspective on these developments. This approach allowed us to present a comprehensive understanding of the trends, challenges, and expectations within the AI and diagnostic medicine community over time.

The inclusion of respondents from engineering and computer science backgrounds was intentional, reflecting the interdisciplinary nature of AI development and its application in diagnostic medicine. The integration of AI into clinical practice is not solely a clinical challenge but also a technological one [34]. Engineers and computer scientists play a crucial role in advancing the algorithms and systems that underpin AI applications in diagnostics. Their insights into technological feasibility, innovation potential, and future directions are invaluable for understanding how these tools might evolve and impact clinical practice. Thus, their contributions are essential to forming a well-rounded perspective on the future of AI use in diagnostic medicine.

Data Collection Procedures

The list of respondents was imported into the web-based survey platform SurveyMonkey (SurveyMonkey Inc), where the questionnaire was designed and the survey was conducted. After uploading the list of respondents, the number of emails was reduced due to bounced emails and opted-out respondents (people who previously chose not to participate in surveys conducted via SurveyMonkey). In W1, the final number was 20,952, whereas in W2, it was 56,480. The personalized email invitations included a direct link to the questionnaire, inviting participants to take part in the survey. The email outlined the purpose of the study, its importance, and instructions for accessing and completing the survey. To encourage participation and improve response rates, reminder emails were sent to nonrespondents every 2 days for a period of 1 week, totaling a maximum of 3 reminders in addition to the invitation to nonrespondents.

In W1, before the formal study, the questionnaire was validated through a pilot study with a random sample of 2000 researchers (2000/20,952, 9.55% of the total emails). In this phase, we evaluated the questionnaire (ie, application routine, consistency, internal logic, completion rate, and response time) and collected feedback from respondents. Of 2000 researchers, 91 (4.55%) who participated in the pilot study did not suggest any changes in the questionnaire. Thus, neither the questionnaire nor the application routine was modified, and the data collected were added to the study results. Following this validation, the full survey was launched to the remaining sample (18,952/20,000, 94.76%), using the same recruitment process for both waves. In W1, the pilot and the formal study were conducted in September 2020. The formal study of W2 was conducted between January and February 2023. To maintain data integrity, each email address was restricted to submitting only 1 response. The SurveyMonkey platform automatically prevented duplicate entries by tracking responses based on individual email addresses. No randomization of questionnaire items was used.

Ethical Considerations

The questionnaire, invitation, and reminder emails informed the respondents about the survey. Before answering the questionnaire, they were informed that the survey was being conducted for research purposes only, personal or sensitive data would not be collected, answers would not be identified in the results, participation would be voluntary, and informed consent would be provided by answering the questionnaire. Thus, all respondents who participated in this survey gave us their informed consent to use the data collected. Given the voluntary participation, anonymity in the results, and absence of sensitive or personal questions, an examination by an ethics committee was not necessary. This study followed the guidelines set forth by Brazilian Resolution 510 of April 7, 2016 (Official Federal Gazette [61]), which provides an exemption from ethics committee registration and evaluation for public opinion research with unidentified participants. In addition, the study data were anonymous, and there was no compensation for the respondents. Again, these procedures followed previous studies that used the same method [39,54,56-60].

Statistical Analysis of the Sample

We used the nonparametric marginal homogeneity test at a 5% significance level to compare the responses obtained in both waves and assess any changes in respondents' expectations over time. This test is typically used to compare nominal data from 2 related samples and determine whether there is a significant difference between their proportions, particularly when the data are not normally distributed [62,63]. It is commonly applied in before-and-after studies [64,65].

In addition to analyzing the differences between W1 and W2 responses, we evaluated the differences between the responses of the 2 knowledge groups (ie, good and some knowledge) for each wave. We used the Mann-Whitney *U* test to identify

XSL•FO
RenderX

whether the level of knowledge interfered with the results. The Mann-Whitney *U* test is a nonparametric statistical test applied when data are not normally distributed [63]. It is commonly used to compare 2 independent groups and determine whether there is a statistically significant difference between them [66]. The test is often applied in cross-sectional research studies [67-69].

Out of the 25 questions presented to the respondents, only 1 (4%) revealed a statistical difference between W1 and W2 by the marginal homogeneity test results. In 2022, the respondents anticipated that AI would be used to interpret heart rhythm and provide improved outcomes in a shorter period than expected in 2020. The results of the tests are reported in Multimedia Appendix 4.

The results of the Mann-Whitney *U* test indicated that there was no statistical difference between most of the responses from the good and some knowledge groups. Only 30% (9/30) of the W1 questions and 10% (3/30) of the W2 questions displayed significant variations between the 2 groups. Hence, we opted to present the aggregated results. The responses categorized by knowledge level are provided in Multimedia Appendix 4.

Results

Table 1 presents an overview of the survey details, including response rates, completed questionnaires, confidence level, margin of error, and the demographics of respondents in W1 and W2. The respondents had a high level of education, with >80% (1039/1245) having a doctoral degree, and the most common occupation was professor or researcher (W1: 901/1039, 72.1% and W2: 1395/2097, 66.5%). Most respondents worked at universities or research organizations (W1: 1007/1251, 80.3% and W2: 1607/2095, 76.7%) and had a varied range of experience, with approximately 30% (383/1249) having 10 to 20 years and 25% (345/1249) to almost 30% (664/2097) having >20 years of experience in both waves. The most common regions of residence were Europe (W1: 458/1251, 36.6% and W2: 849/2099, 40.5%), Asia (W1: 367/1251, 29.3% and W2: 640/2099, 30.5%), and North America (W1: 258/1251, 20.6% and W2: 350/2099, 16.7%).

In addition to the demographic questions, we mapped the publication profile of both participants and nonparticipants of the survey through the metadata of their article records collected in WoS. The results are provided in Multimedia Appendix 5, and it shows that the survey participants and nonparticipants are similar concerning the research areas whose publications have been indexed.

We first asked the respondents about their general expectations of AI's potential to change diagnostic medicine radically in the future. Most respondents (W1: 965/1410, 68.5% and W2: 1555/2324, 66.9%) anticipated that this would happen within 10 years, while a smaller proportion (W1: 420/1410, 29.7% and W2: 706/2324, 30.1%) expected this to take longer. A minor percentage (W1: 25/1410, 1.8% and W2: 63/2324, 2.7%) believed that such a radical change was unlikely.

The respondents' expectations of 7 future events resulting from using AI in diagnostic medicine are depicted in Figure 1. Reduction of patients' hospitalization time (<50% in both waves) and improvement in treatment compliance (>50% in both waves) received the lowest percentages of likely before 10 years in both waves. Most respondents expected a lower screening cost (W1: 942/1360, 69.3% and W2: 1514/2243, 67.6%) and a higher diagnosis reliability (W1: 926/1362, 68% and W2: 1479/2244, 66%) within 10 years. The latter was also the most agreed-upon event, with only approximately 3% (37/1362) of respondents considering its occurrence unlikely. Improvements in treatment compliance and reduction in patients' hospitalization received the highest percentages of unlikely in both waves.

XSL•FO
RenderX

Table 1. Overview of respondent demographics, knowledge level, and survey details for 2 waves of a global cross-sectional survey on artificial intelligence use in diagnostic medicine. Data include number of invitations, response rates, knowledge levels, educational degrees, occupations, institutional affiliations, years of experience, and regional distribution of respondents in wave 1 (WI; September 2020) and wave 2 (W2; February 2023).

	W1, n (%)	W2, n (%)
Survey summary (W1: n=20,952; W2: n=56,480)[a]		
Response rate	1622 (7.7)	2606 (4.6)
Fully completed questionnaires	1208[b] (5.7)	1724[c] (3)
Knowledge level (W1: n=1622; W2: n=2606)		
Good knowledge	724 (44.6)	1192 (45.7)
Some knowledge	706 (43.5)	1160 (44.5)
No knowledge[d]	192 (11.8)	254 (9.7)
Demographic information		
Degree of education		
Doctoral degree	1039 (83.5)	1717 (82.3)
Master's degree	162 (13)	292 (14)
Bachelor's degree	19 (1.5)	48 (2.3)
Associate's degree	25 (2)	30 (1.4)
Occupation		
Professor or researcher	901 (72.1)	1395 (66.5)
Physician or clinician	133 (10.6)	276 (13.2)
Master's degree student or PhD[e] student	137 (11)	238 (11.3)
Other	40 (3.2)	62 (3)
Public health or health care professional	14 (1.1)	61 (2.9)
Manager or executive	23 (1.8)	60 (2.9)
Policy maker	1 (0.1)	5 (0.2)
Work institution		
University or research organization	1007 (80.5)	1607 (76.7)
Hospital or similar organizations	148 (11.8)	315 (15)
Industry	69 (5.5)	120 (5.7)
Government	27 (2.2)	53 (2.5)
Experience (years)		
10 to 20	383 (30.7)	676 (32.2)
>20	345 (27.6)	664 (31.7)
5 to 10	343 (27.5)	510 (24.3)
<5	178 (14.3)	247 (11.8)
Region		
Europe	458 (36.6)	849 (40.4)
Asia (including the Middle East)	367 (29.3)	640 (30.5)
North America (including Central America and the Caribbean)	258 (20.6)	350 (16.7)
South America	102 (8.2)	173 (8.2)
Africa	34 (2.7)	50 (2.4)
Australasia or Pacific Islands	32 (2.6)	37 (1.8)

[a]Number of participants to whom the survey invitation was sent.

[b]Confidence level of 95% and a margin of error of 2.7%.

XSL•FO
RenderX

[c]Confidence level of 95% and a margin of error of 2.3%.

[d]Respondents with no knowledge were disqualified and prevented from proceeding to the main questionnaire.

[e]PhD: Doctor of Philosophy.

Figure 1. The likelihood of expected future events from artificial intelligence use in diagnostic medicine based on 2 global surveys: wave 1 (WI) in September 2020 and wave 2 (W2) in February 2023. Responses showed the expected impact on screening costs, diagnostic reliability, disease prediction, physician workload, access to care, treatment compliance, and hospitalization time, categorized as likely before 10 years, after 10 years, or unlikely. *P* value refers to the marginal homogeneity test.

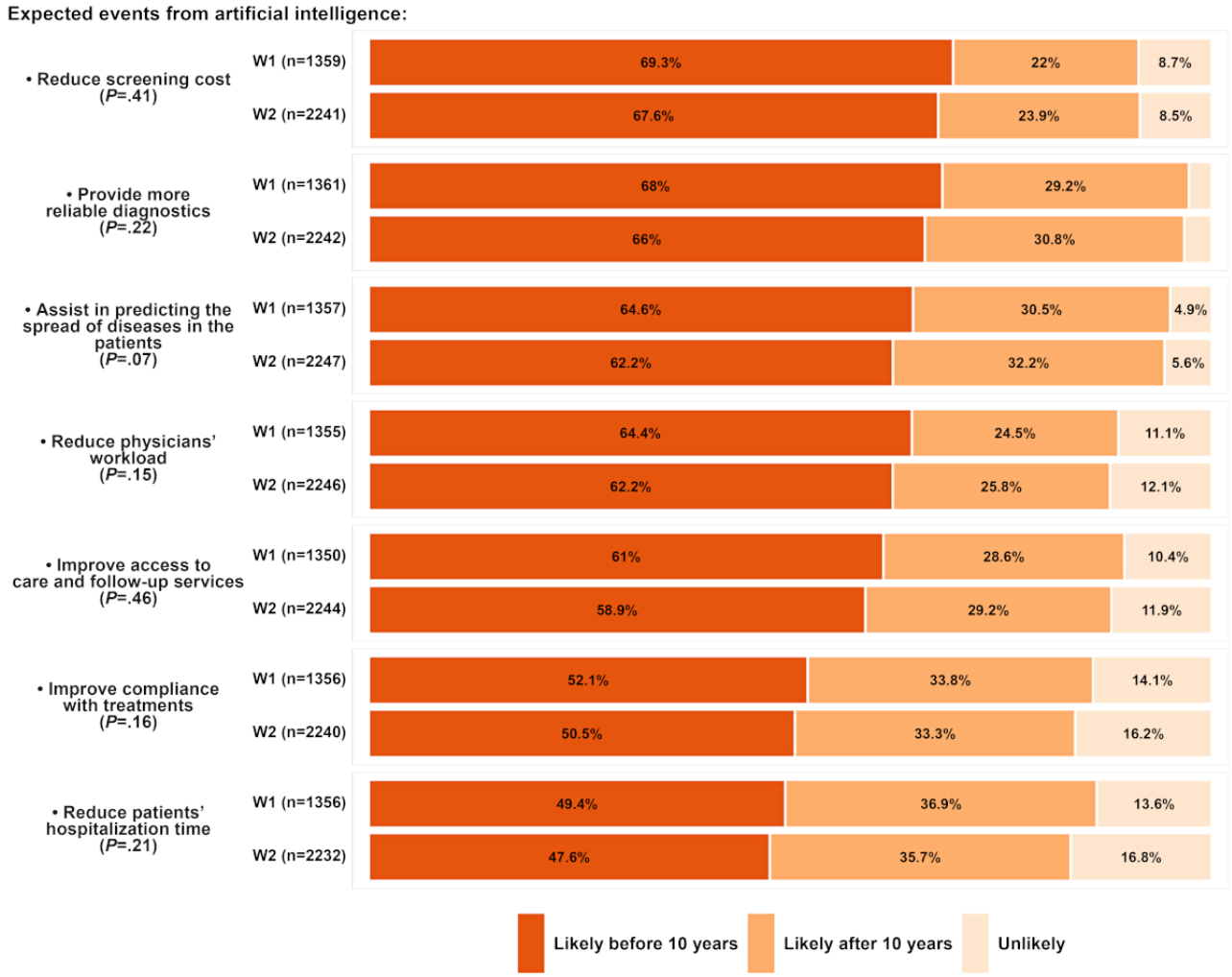

The respondents' expectations of integrating AI and 12 selected diagnostic instruments to enhance diagnostic outcomes are illustrated in Figure 2. The 3 most expected integrations were x-ray diagnosis (W1: 1071/1322, 81% and W2: 1787/2183, 81.6%), heart rhythm interpretation (W1: 1062/1313, 80.9% and W2: 1578/2181, 72.4%), and skin malignancy diagnosis (W1: 1024/1320, 77.6% and W2: 1680/2176, 77.2%). Intrapartum monitoring (W1: 577/1235, 46.7% and W2: 1045/2070, 50.5%) and identification of sepsis symptoms (W1: 770/1257, 61.3% and W2: 1290/2095, 61.6%) were the events with the highest percentages of unlikely.

The respondents' perceptions of the main barriers to using AI in diagnostic medicine are shown in Figure 3. The most common barriers in both waves were the difficulty of incorporating AI into clinical practice and ethical or regulatory issues. In W1, these barriers were selected by 41% (531/1296) and 37.6% (487/1296) of the respondents who answered this question, respectively. In W2, ethical or regulatory issues were slightly more prevalent (941/2164, 43.5%) than the difficulty of incorporating AI into clinical practice (788/2164, 36.4%). The impact of AI on the workforce and the lack of improvement in medical diagnostics were less frequently chosen as barriers, with a combined percentage of approximately 15% (W1: 167/1296 and W2: 333/2164) in both waves. The respondents who identified the difficulty of incorporating AI into clinical practice as a barrier also reported its specific related challenges. The most common challenge in both waves was aligning AI to the specific context of clinical practice (W1: 255/535, 47.6% and W2: 399/789, 50.4%).

Similarly, the respondents who identified ethical or regulatory issues as barriers reported their related challenges. The most common challenge in both waves was the uncertainty about legal responsibility and accountability for AI-supported clinical decisions (W1: 235/489, 48.4% and W2: 449/940, 47.8%). When comparing other specific challenges, the one where waves differed the most was the unsuitability of AI to a real-world

XSL•FO
RenderX

context of care and services (W1: 22/87, 25.3% and W2: 57/158, 36.1%).

Figure 2. The likelihood of integrating artificial intelligence with various diagnostic instruments to deliver better results, based on global survey responses from wave 1 (W1; September 2020) and wave 2 (W2; February 2023). Instruments included heart rhythm interpretation, x-ray diagnosis, skin malignancy diagnosis, and others. *Statistical significance was at the 5% level. MRI: magnetic resonance imaging.

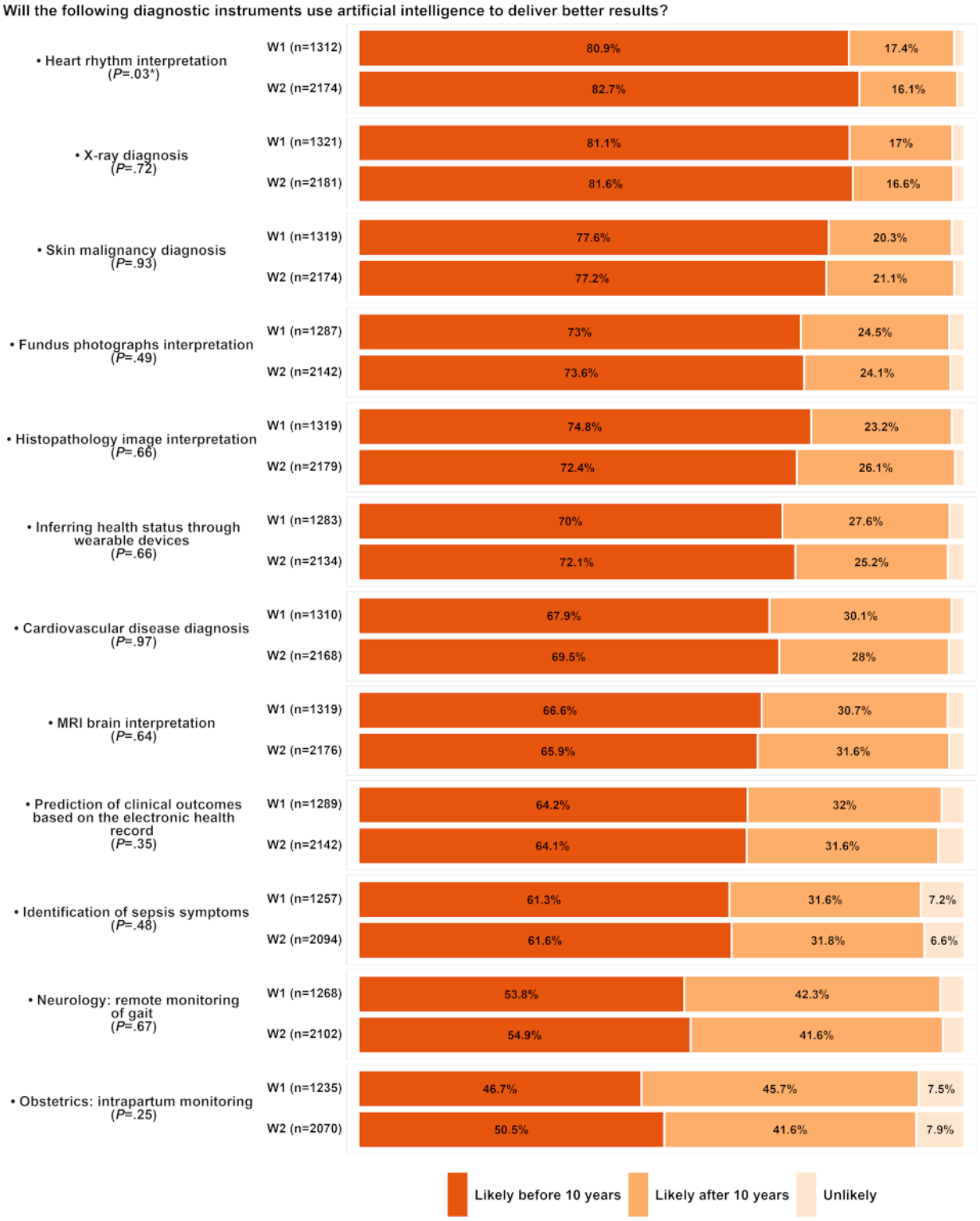

XSL•FO
RenderX

Figure 3. Barriers to artificial intelligence (AI) adoption in diagnostic medicine, comparing wave 1 (WI; September 2020) and wave 2 (W2; February 2023) survey results. Responses addressed clinical, ethical, regulatory, and workforce-related challenges. *P* value refers to the marginal homogeneity test.

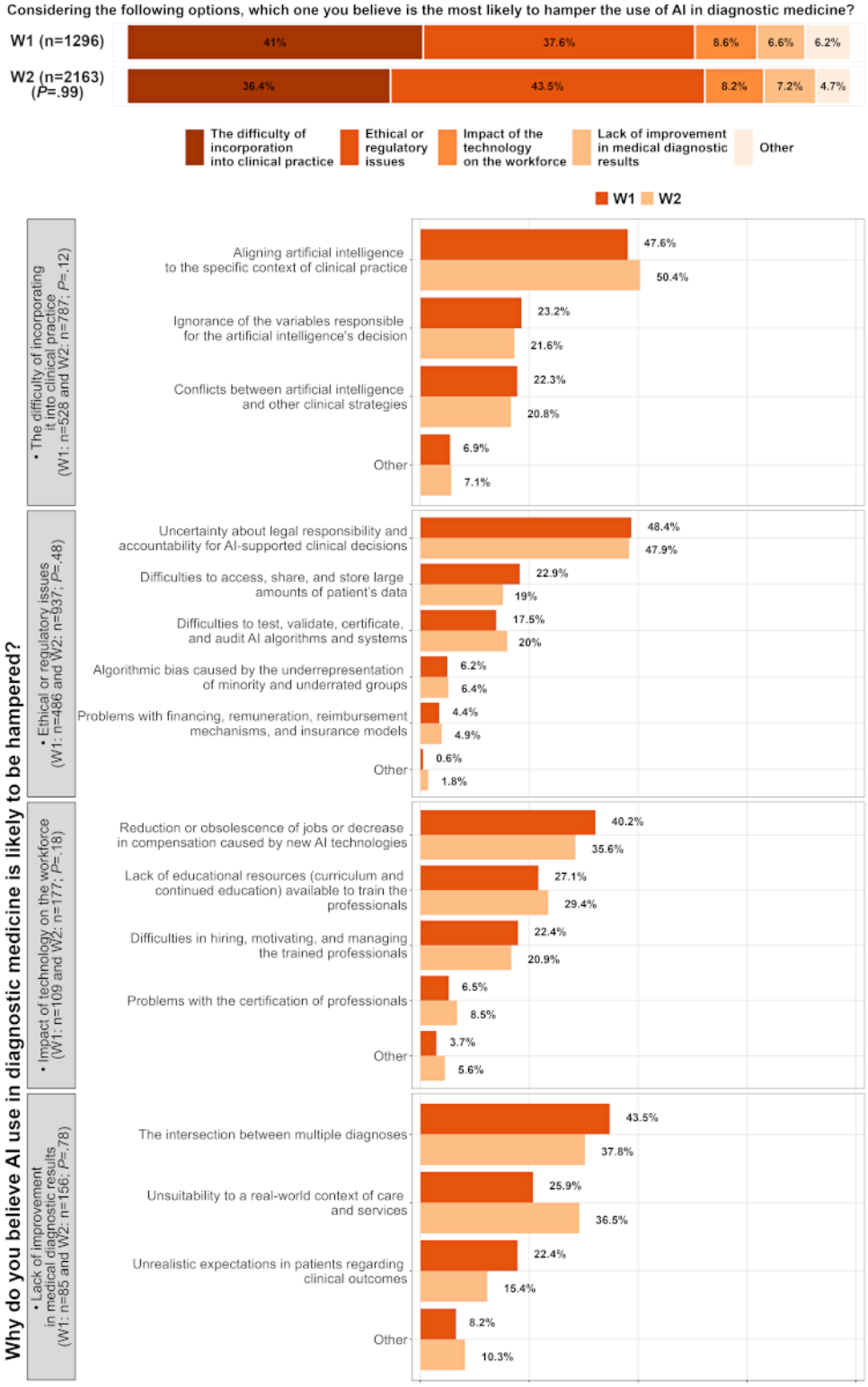

Discussion

Principal Findings

Although the span of 28 months may seem a short period to assess changes in researchers' expectations on a given topic, the case of AI in diagnostic medicine is peculiar. This is an area where transformations are quick, and innovations are the norm [51]. One indicator of this rapid transformation is the surge in publications on this topic. For instance, a quick search on WoS between January 1, 2015, and September 20, 2020, showed >6000 articles on AI in diagnostic medicine. However, between September 21, 2020, and December 30, 2022, this number rose to >12,000, an increase of 113.5% (the search strategies used

XSL•FO
RenderX

to identify these publications are the same as those described in Textbox 2).

AI is expected to improve the quality of health care for patients, mainly by addressing patient safety and reducing medical errors [18,70,71]. A study compared deep learning algorithms and physicians in diagnosing skin conditions from photos and dermoscopic images to evaluate how AI can support or automate health care decision-making. The results showed that the algorithms performed better than the average physician [18]. Image reconstruction using deep learning algorithms is another example of how AI can benefit patients. This technique can potentially reduce radiation exposure and the time needed for image acquisition and segmentation in preparation for radiotherapy, which may reduce side effects and material costs [72].

Survey respondents in both waves agreed that AI could positively impact quality over a 10-year time horizon by (1) aiding the prediction of disease progression, (2) increasing diagnostic reliability, and (3) improving treatment compliance. Furthermore, they believed that AI might improve quality in diagnostic medicine in <10 years by producing more reliable, compliable, and predictable outcomes. The answers to the questionnaire (Figure 2) show that AI-based quality improvements are associated with patient outcomes, but the technology may also benefit health care practitioners by reducing human error and fatigue [18,73-75].

Information accuracy is fundamental for achieving better-quality health care outcomes. However, technology and information accuracy may become a barrier to adopting AI [20,76]. For example, most of the researchers (W1: 255/535, 47.7% and W2: 399/789, 50.8%) who selected the difficulty of incorporating AI into clinical practice as the main barrier expressed their concern that the use of AI might be hampered by its misalignment with the specific context of clinical practice. This issue is associated with the generalizability and reproducibility of AI algorithms [20], which are often trained on clean datasets with no poor-quality information but are applied in a real clinical setting where data may be incomplete or erroneous. When comparing responses in both waves, this topic was more relevant for researchers in W2 than those in W1, which may indicate a growing concern about this misalignment. Furthermore, data accuracy issues may influence the quality of AI algorithm outputs by introducing biases [20,71,77,78] due to discrepancies in patient population (eg, individuals from low socioeconomic backgrounds and underrepresented minority groups) and organizational settings (eg, hospitals and primary care) [79,80].

Cost reduction is another potential benefit of AI in diagnostic medicine, as it relates to the optimization of resource use [72], processes, and services [71], as well as financial management [18,70]. As a result, the same therapy could be offered at a lower cost, or more services could be provided at the same cost. The use of AI to process large image datasets, for example, could offer workflow and productivity gains and reduce the workforce in many back-office activities, such as billing, clinical appointments, and staffing [72]. AI could have a variable impact on patient-, organizational-, and system-level tasks, potentially leading insurers to revise their reimbursement policy to reduce health care costs and improve quality [18]. Moreover, AI may make costly and time-consuming screening programs more affordable in resource-limited countries [23,81].

A total of 3 questions assessed the respondents' beliefs about the changes in costs driven by AI. Most respondents (W1: 982/1360, 69.3% and W2: 1514/2243, 67.5%) in both waves anticipated that screening costs and physician workloads would be reduced in <10 years and patients' hospitalization time would be shortened. These changes could potentially lower the cost of health care services and improve their quality, as physicians could perform more tasks concurrently and focus on more skilled work by delegating repetitive tasks to AI algorithms [18,72]. These benefits could be offset by the increased costs and risks for patients and the health system due to unnecessary testing and treatments prompted by AI [20]. Furthermore, there are concerns about the initial investments required for state-of-the-art AI products, the potential for long-term cost savings, and the need for a balanced evaluation of costs and benefits when considering AI integration [82,83].

Conversely, when appropriately implemented, AI can enhance diagnostic accuracy and reduce the need for redundant or follow-up tests. This efficiency is already leading to cost savings and a reduced burden on the health care system, and there is potential for further reduction of redundant and follow-up tests [84]. Therefore, AI value analyses should consider the technical aspects, short-term returns, and the overall value of automating processes [20].

The survey results in both waves have highlighted the anticipated transformation of AI in diagnostic instruments within a decade. In particular, image analysis and interpretation applications, such as x-ray, skin malignancy, and histopathologic diagnosis, have received optimistic perspectives. This may relate to advances in deep learning techniques for image analysis, such as convolutional neural networks and transformer architectures [85], which have significantly improved tasks such as image classification and object detection [3,29,46]. Deep learning is already the most used AI technique in health-related applications [71], and combining computing power, large datasets, and convolutional neural networks is also responsible for the subsequent revolution in medical imaging [18]. Since deep learning was introduced in 2012 for image recognition, it surpassed the human accuracy rate in specific, large data-labeled datasets after 5 years [72].

Furthermore, technology's maturity may relate to perceptions of the use of AI in medical imaging. Digital medical images have been in use since the 1960s, and mature technologies have been built around them. Most radiology departments maintain picture archiving and communication systems containing historical images [18]. Cardiology applications highlighted by the respondents in both waves (eg, heart rhythm interpretation and disease diagnostics) have been in use for many decades, but recent developments are seen as paradigm shifts in clinical practice [86,87]. Because AI algorithms rely on the design of distinctive features to learn from data, more mature technologies are expected to provide a better set of such features, as is the case for these diagnostic instruments [79,88].

XSL•FO
RenderX

Conversely, the least optimistic results in both waves were related to applications that use other data formats, such as electronic health record texts (eg, identification of sepsis symptoms and prediction of clinical outcomes) or 1D signal data (eg, intrapartum monitoring and remote monitoring of gait). Multiple reasons may account for that, for example, data heterogeneity in electronic health records, including mixed data types, such as free text clinical notes, radiological reports, medication dosages, and medical codes [70,89]. Moreover, privacy policies and data interoperability impact information transfer across institutions that construct large datasets for deep learning representations of electronic health records [89]. For gait and intrapartum monitoring, some issues, such as the lack of large datasets and the use of clinical context as a complement for decision-making, also affect its performance [90].

Another factor that may help explain these results is regulation. In 2018, the Food and Drug Administration published a fast-track approval plan for AI medical algorithms for different applications [72]. Since then, the number of algorithms approved has soared, and there are now almost 400 algorithms associated with radiology and 58 associated with cardiology, many of which are dedicated to x-ray diagnostics and heart rhythm interpretation, respectively [10]. Thus, technological innovation and maturity combined with adequate regulation seem correlated to the respondents' perceptions of AI's impact on diagnostics in the coming years.

According to the respondents' answers in both waves, the difficulty of incorporating AI into clinical practice, despite the reported benefits in terms of health outcomes, may hamper these innovations. These issues may be related to resistance to change in adopting AI in diagnostic medicine [91-93]. Another relevant challenge to AI adoption was ethical and regulatory issues. Accountability and legal responsibility for AI clinical decisions were marked as the most relevant issues in this topic in both waves. Accountability issues arise when AI algorithms make autonomous decisions about diagnoses and treatments beyond their role as support tools [20,23]. As respondents in W2 (941/2164, 43.5%) rated ethical and regulatory issues as more relevant than those in W1 (487/1296, 37.6%; in fact, as the most relevant), this seems to be a growing concern. As more AI applications become available to the public and cause social changes shortly, it is unsurprising that this issue increased in relevance from one wave to another. Large language models, such as ChatGPT, are examples of how AI has become more accessible to people and how this may raise more concerns over time [94,95]. For example, the United Kingdom's Medicines and Healthcare Products Regulatory Agency points out some issues with incorporating large language models in medical devices. While it anticipates their use in health care soon, it emphasizes the need to uphold safety, effectiveness, and ethical standards [96].

In short, there are still doubts about who should be held responsible if a patient experiences an adverse event due to AI-based technology [34]. Insurers and regulators will have to be able to distinguish algorithmic errors from those resulting from misuse by the clinician, the organization, or even the patient. This issue is exacerbated by the black box nature of AI systems [20]. The increased use of AI in medicine will probably lead to legal challenges regarding medical negligence attributed to complex decision support systems [18]. Radiologist associations globally have been evaluating the ethical integration of AI in medical imaging. The Canadian Association of Radiologists underscores the potential of AI in medical imaging, emphasizing the importance of addressing its ethical and legal challenges [97]. Furthermore, the Royal Australian and New Zealand College of Radiologists has formulated ethical guidelines and deployment standards in response to AI's rapid growth in radiology [98]. Simultaneously, a Joint European and North American Multisociety Statement, representing several radiological entities, emphasizes AI's potential to enhance radiology efficiency and introduce systemic errors, advocating for ethical AI use that prioritizes patient well-being and transparent practices [99].

According to the respondents in both waves, other relevant issues are accessing and sharing large amounts of patient data. The use of big data raises concerns that may hamper the potential of AI in diagnostic medicine. These relate to data protection and confidentiality and the need to access the large volumes of data needed to train AI algorithms. Regarding data protection and confidentiality, key concerns include the data's origin, obtaining patient consent, authorization for reuse, data ownership and responsibility, who can access and reuse the data, and the conditions under which the data can be used [20]. For example, a recent report from the European Union Agency for Cybersecurity on the pivotal role of AI in medical imaging diagnosis shows the importance of cybersecurity and privacy controls in properly addressing these many concerns [100]. Other relevant issues raised in the literature are the risk of the inappropriate use of databases and imprecise consent terms [34]. Technology-related obstacles include limited deidentification techniques and the need to integrate and standardize large amounts of data [34].

A shortage of well-annotated datasets for training AI algorithms is a key obstacle to the large-scale introduction of these systems. Furthermore, the absence of clear, standardized regulations could lead to the illicit collection of data from unknown sources [23]. These concerns relate to the very nature of machine learning technologies, as they need large amounts of training materials to be taught. On the other hand, as clinical data are gathered from multiple and diverse sources, it becomes easier to trace them to patients, threatening their privacy [49]. As ubiquitous data collection becomes commonplace, consensus must be reached for a consent framework to guide health-related data sharing [18]. In addition, there is growing concern about the lack of clinician involvement in the development of AI applications. A recent systematic review showed that developers typically consulted clinicians late in the design process, and many applications lacked validation against clinical expertise or detailed algorithm descriptions [101].

A particularly sensitive problem is iatrogenic risks related to a lack of transparency in the algorithm training processes. Before an AI algorithm can be unleashed in clinical practice, it has to be debugged, audited, simulated, and validated, along with prospective scrutiny [72]. Respondents identified these problems as major issues in both waves, with higher percentages in W2 (170/789, 21.5%). Besides the need for certification of AI

XSL•FO
RenderX

systems, professionals, and teams, the rapid evolution of machine learning–based models presents an enormous challenge to regulation as more data are collected and used in algorithm development. A critical problem is how these updates should be evaluated and audited [18]. Quality control instruments for AI algorithms are needed to prevent abuse by AI system developers, as clinical decision support systems could be programmed to favor certain drugs, tests, or devices without users being aware of this manipulation. Transparency will be difficult to achieve if companies make their algorithms purposefully opaque for proprietary or financial reasons [34].

Another issue, which is quite present in the reviewed literature but received relatively little attention from the respondents (W1: 32/489, 6.5% and W2: 60/940, 6.4%) who answered this question, relates to embedded bias present in AI algorithms due to a lack of inclusion of minority individuals in datasets [72]. Such biases, resulting from underrepresenting minority groups and those considered vulnerable in the datasets used to develop AI systems, could reinforce discriminatory practices based on race, sex, or other features [34]. Machine learning datasets must be large, but the often-used clinical trial research databases are derived mainly from majority populations. The resulting algorithms may be more likely to fail when applied to underserved and possibly underrepresented patient groups [49]. Algorithms trained on health care datasets, which inherently mirror biases in health care spending, have been observed to exacerbate racial disparities in access to care within the United States and can affect medical diagnosis [102]. This problem will require careful attention from regulatory agencies [103].

Limitations

This study relies on previous studies that have explored future scenarios of science and technology by surveying researchers [39,54-58,60]. However, it also inherits its limitations. The first one is the lack of diversity among respondents, who were identified and selected based on their publications in scientific journals. Therefore, most respondents (W1: 1007/1251, 80.5% and W2: 1607/2095, 76.7%) are researchers and professors affiliated with universities and research organizations. While their insights offer valuable academic and technological perspectives on AI use in diagnostic medicine, the relatively lower representation of clinicians might not capture the complete range of insights from the frontline of clinical practice. Considering that our respondents were sourced from scientific publications, a domain where professors and researchers are more commonly the authors, this was somewhat expected given our methodology.

Another limitation is the potential optimism bias of respondents. As researchers involved with AI and diagnostic medicine, they may have more positive expectations about the future of this technology than other groups (eg, patients, managers, businesspeople, and politicians). Nevertheless, they are among the most qualified to inform about future possibilities of AI use in diagnostic medicine as they are helping to advance the scientific and technological knowledge in the field. The self-attribution of knowledge level by the respondents in the questionnaire is another limitation. We cannot verify the accuracy of the respondents' self-assessments or assign knowledge levels to them. Therefore, the self-assigned knowledge level reflects how the respondents perceive their own knowledge in the area. However, this study only includes participants who are authors of peer-reviewed scientific articles related to AI and diagnostic medicine indexed in WoS, which minimizes the risk of incorporating opinions from people with no knowledge of the topic.

A potential limitation of our study is the varied interpretations of the term "artificial intelligence" among respondents. AI is a broad and multifaceted concept, encompassing everything from machine learning and neural networks to natural language processing and specialized rule-based systems [1,2]. The use of thesaurus terms to identify relevant respondents and survey questions may have introduced biases based on how individuals conceptualize AI. This variation in understanding could influence responses, particularly regarding the anticipated impact and integration of AI into diagnostic medicine. Hence, future studies may benefit from providing a clearer, more uniform definition of AI to respondents, ensuring a more consistent interpretation across the survey population.

We included radiology-specific terms in the search for respondents because this field is one of the most likely to be affected by the advances of AI in diagnostics [104-106]. We acknowledge that there is a potential selection bias in extending the search to radiology experts. However, it is important to highlight that the query also included general terms, such as diagnostics, medicine, and medical, which mitigates this potential bias by engaging a more diverse range of diagnostic experts.

Furthermore, there is the possibility of regional demographic bias. Collecting respondents' emails from scientific publications may introduce a bias toward specific demographic groups, possibly leading to an underrepresentation of experts from certain regions. Our methodology for identifying experts does not allow us to have demographic information on nonrespondents. The lack of this demographic information prevents an extensive analysis of nonresponse bias and demographic balance. However, we have included a regional breakdown of the responses in Multimedia Appendix 3 to address this concern and provide additional information. While slight percentage variations exist in the responses from each region, the overarching response patterns remain consistent across them.

Conclusions

We reported the results of a global, 2-wave, cross-sectional survey of >3000 researchers with knowledge in diagnostic medicine. Most respondents (W1: 965/1410, 68.4% and W2: 1555/2324, 66.9%) believed that AI would transform diagnostic medicine radically in 10 years. Furthermore, they expected that AI would reduce screening costs and increase diagnostic reliability in this period and that x-ray diagnosis and heart rhythm interpretation are the most likely diagnostic tools to be integrated with AI. According to the respondents, the main general barriers to using AI in diagnostic medicine were the difficulty of incorporating AI into clinical practice and ethical or regulatory issues. While there were little changes in expectations when comparing the 2 waves, ethical and regulatory

XSL•FO
RenderX

issues were more relevant to respondents in the second survey (941/2164, 43.5%), which indicates that this issue is becoming more urgent.

Moreover, adapting AI to the specific context of clinical practice and uncertainties about legal responsibility and accountability for AI-supported clinical decisions were the main specific barriers in both surveys. While our results provide a broad overview of the future of AI use in diagnostic medicine, they also pave the way for more detailed, focused research into specific aspects, enriching our understanding of this evolving domain. This may include an in-depth analysis of clinicians' trust in AI applications in diagnostic medicine; further analysis of the integration of AI in specific diagnostic tools, such as x-ray diagnosis, heart rhythm interpretation, and skin malignancy diagnosis; or a cost analysis of AI implementation.

The use of AI in diagnostic medicine is rapidly evolving and can potentially transform the quality and cost of health care for patients. As AI continues to advance, it is essential to address these challenges and ensure that AI-based technologies are developed and deployed responsibly and ethically. The success of AI in diagnostic medicine will ultimately depend on the willingness of stakeholders to collaborate, innovate, and work together to ensure that these technologies are used to improve patient outcomes and enhance the overall quality of health care.

Acknowledgments

The authors are grateful to all those who completed the questionnaire. They also thank the Fiocruz Strategy for Agenda 2030/Oswaldo Cruz Foundation and the National Health Fund/Ministry of Health for paying the article processing charge.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to its size but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys.
[DOCX File , 29 KB - jmir_v27i1e53892_app1.docx]

Multimedia Appendix 2
Survey references.
[DOCX File , 122 KB - jmir_v27i1e53892_app2.docx]

Multimedia Appendix 3
Summarized and visual representation of the process of recruiting respondents and data collection.
[DOCX File , 376 KB - jmir_v27i1e53892_app3.docx]

Multimedia Appendix 4
Data by knowledge level with Mann-Whitney U test and marginal homogeneity test results.
[DOCX File , 85 KB - jmir_v27i1e53892_app4.docx]

Multimedia Appendix 5
Respondents' publication analysis.
[DOCX File , 102 KB - jmir_v27i1e53892_app5.docx]

References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun 13;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
2. Lu Y. Artificial intelligence: a survey on evolution, models, applications and future trends. J Manag Anal 2019 Feb 06;6(1):1-29. [doi: 10.1080/23270012.2019.1570365]
3. Alexander A, Jiang A, Ferreira C, Zurkiya D. An intelligent future for medical imaging: a market outlook on artificial intelligence for medical imaging. J Am Coll Radiol 2020 Jan;17(1 Pt B):165-170 [FREE Full text] [doi: 10.1016/j.jacr.2019.07.019] [Medline: 31918875]
4. Chiu HY, Chao HS, Chen YM. Application of artificial intelligence in lung cancer. Cancers (Basel) 2022 Mar 08;14(6):1370 [FREE Full text] [doi: 10.3390/cancers14061370] [Medline: 35326521]

XSL•FO
RenderX

5. Meskó B. The real era of the art of medicine begins with artificial intelligence. J Med Internet Res 2019 Nov 18;21(11):e16295 [FREE Full text] [doi: 10.2196/16295] [Medline: 31738169]
6. Sezgin E. Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. Digit Health 2023 Jul 02;9:20552076231186520 [FREE Full text] [doi: 10.1177/20552076231186520] [Medline: 37426593]
7. Al Kuwaiti A, Nazer K, Al-Reedy A, Al-Shehri S, Al-Muhanna A, Subbarayalu AV, et al. A review of the role of artificial intelligence in healthcare. J Pers Med 2023 Jun 05;13(6):951 [FREE Full text] [doi: 10.3390/jpm13060951] [Medline: 37373940]
8. Becker A. Artificial intelligence in medicine: what is it doing for us today? Health Policy Technol 2019 Jun;8(2):198-205. [doi: 10.1016/j.hlpt.2019.03.004]
9. Artificial intelligence in healthcare market. Precedence Research. URL: https://www.precedenceresearch.com/artificial-intelligence-in-healthcare-market [accessed 2024-04-29]
10. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. Food and Drug Administration. URL: https://tinyurl.com/3s6ka23j [accessed 2023-08-14]
11. Biot J. How will clinical practice be impacted by artificial intelligence? Eur J Dermatol 2019 Apr 01;29(S1):8-10. [doi: 10.1684/ejd.2019.3536] [Medline: 31017581]
12. El Alaoui Y, Elomri A, Qaraqe M, Padmanabhan R, Yasin Taha R, El Omri H, et al. A review of artificial intelligence applications in hematology management: current practices and future prospects. J Med Internet Res 2022 Jul 12;24(7):e36490 [FREE Full text] [doi: 10.2196/36490] [Medline: 35819826]
13. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med 2023 May 25;388(21):1981-1990. [doi: 10.1056/nejmra2301725]
14. Stedman T. Stedman's Medical Dictionary. 28th edition. Philadelphia, PA: Lippincott Williams & Wilkins; 2006.
15. Yang D, Fineberg HV, Cosby K. Diagnostic excellence. JAMA 2021 Nov 16;326(19):1905-1906. [doi: 10.1001/jama.2021.19493] [Medline: 34709367]
16. Blum A, Zins M. Radiology: is its future bright? Diagn Interv Imaging 2017 May;98(5):369-371 [FREE Full text] [doi: 10.1016/j.diii.2017.04.002] [Medline: 28477770]
17. Powell AC. Impact of the Artificial Nudge. Acad Radiol 2020 Jan;27(1):143-146. [doi: 10.1016/j.acra.2019.09.010] [Medline: 31818383]
18. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct 10;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
19. Grzybowski A, Jin K, Wu H. Challenges of artificial intelligence in medicine and dermatology. Clin Dermatol 2024 May;42(3):210-215. [doi: 10.1016/j.clindermatol.2023.12.013] [Medline: 38184124]
20. Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, et al. Artificial intelligence and health technology assessment: anticipating a new level of complexity. J Med Internet Res 2020 Jul 07;22(7):e17707 [FREE Full text] [doi: 10.2196/17707] [Medline: 32406850]
21. Oala L, Murchison AG, Balachandran P, Choudhary S, Fehr J, Leite AW, et al. Machine learning for health: algorithm auditing and quality control. J Med Syst 2021 Nov 02;45(12):105 [FREE Full text] [doi: 10.1007/s10916-021-01783-y] [Medline: 34729675]
22. Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. Lancet Digit Health 2022 May;4(5):e384-e397. [doi: 10.1016/s2589-7500(22)00003-6]
23. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging 2018 Oct;9(5):745-753 [FREE Full text] [doi: 10.1007/s13244-018-0645-y] [Medline: 30112675]
24. Jassar S, Adams SJ, Zarzeczny A, Burbridge BE. The future of artificial intelligence in medicine: medical-legal considerations for health leaders. Healthc Manage Forum 2022 May 31;35(3):185-189 [FREE Full text] [doi: 10.1177/08404704221082069] [Medline: 35354409]
25. Mello MM, Guha N. Understanding liability risk from using health care artificial intelligence tools. N Engl J Med 2024 Jan 18;390(3):271-278. [doi: 10.1056/nejmhle2308901]
26. Nazer LH, Zatarah R, Waldrip S, Ke JX, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2023 Jun 22;2(6):e0000278 [FREE Full text] [doi: 10.1371/journal.pdig.0000278] [Medline: 37347721]
27. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689 [FREE Full text] [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
28. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MA, Dwivedi YK. A systematic literature review of artificial intelligence in the healthcare sector: benefits, challenges, methodologies, and functionalities. J Innov Knowl 2023 Jan;8(1):100333. [doi: 10.1016/j.jik.2023.100333]
29. Alexander A, McGill M, Tarasova A, Ferreira C, Zurkiya D. Scanning the future of medical imaging. J Am Coll Radiol 2019 Apr;16(4 Pt A):501-507 [FREE Full text] [doi: 10.1016/j.jacr.2018.09.050] [Medline: 30522844]

XSL•FO
RenderX

30. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019 Sep 07;394(10201):861-867. [doi: 10.1016/S0140-6736(19)31721-0] [Medline: 31378392]
31. d'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Fam Med Community Health 2022 Nov 30;10(Suppl 1):e001670 [FREE Full text] [doi: 10.1136/fmch-2022-001670] [Medline: 36450391]
32. Derevianko A, Pizzoli SF, Pesapane F, Rotili A, Monzani D, Grasso R, et al. The use of artificial intelligence (AI) in the radiology field: what is the state of doctor-patient communication in cancer diagnosis? Cancers (Basel) 2023 Jan 12;15(2):470 [FREE Full text] [doi: 10.3390/cancers15020470] [Medline: 36672417]
33. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017 Apr;69S:S36-S40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
34. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
35. Kulkarni S, Seneviratne N, Baig MS, Khan AH. Artificial intelligence in medicine: where are we now? Acad Radiol 2020 Jan;27(1):62-70. [doi: 10.1016/j.acra.2019.10.001] [Medline: 31636002]
36. Karimian G, Petelos E, Evers SM. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. AI Ethics 2022 Mar 28;2(4):539-551. [doi: 10.1007/S43681-021-00131-7]
37. Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, et al. Superior skin cancer classification by the combination of human and artificial intelligence. Eur J Cancer 2019 Oct;120:114-121 [FREE Full text] [doi: 10.1016/j.ejca.2019.07.019] [Medline: 31518967]
38. Mota F, Braga LA, Cabral BP, Conte Filho CG. What is the future of lab-on-a-chip diagnostic devices? Assessing changes in experts' expectations over time. Foresight 2021 Aug 10;23(6):640-654. [doi: 10.1108/fs-05-2021-0101]
39. Mota FB, Braga LA, Cabral BP, Conte Filho C. Future of bioprinted tissues ànd organs: a two-wave global survey. Foresight STI Gov 2022:20. [doi: 10.17323/2500-2597.2022.1.6.20]
40. Vissers S, Stolle D. Spill-over effects between Facebook and on/offline political participation? Evidence from a two-wave panel study. J Inf Technol Politics 2014 Aug;11(3):259-275. [doi: 10.1080/19331681.2014.888383]
41. Wang D, Ross B, Zhou X, Meng D, Zhu Z, Zhao J, et al. Sleep disturbance predicts suicidal ideation during COVID-19 pandemic: a two-wave longitudinal survey. J Psychiatr Res 2021 Nov;143:350-356 [FREE Full text] [doi: 10.1016/j.jpsychires.2021.09.038] [Medline: 34571319]
42. Comaniciu D, Engel K, Georgescu B, Mansi T. Shaping the future through innovations: from medical imaging to precision medicine. Med Image Anal 2016 Oct;33:19-26. [doi: 10.1016/j.media.2016.06.016] [Medline: 27349829]
43. Makridakis S. The forthcoming artificial intelligence (AI) revolution: its impact on society and firms. Futures 2017 Jun;90:46-60. [doi: 10.1016/j.futures.2017.03.006]
44. Martín Noguerol T, Paulano-Godino F, Martín-Valdivia MT, Menias CO, Luna A. Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning applications in radiology. J Am Coll Radiol 2019 Sep;16(9 Pt B):1239-1247. [doi: 10.1016/j.jacr.2019.05.047] [Medline: 31492401]
45. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? Am J Med 2018 Feb;131(2):129-133. [doi: 10.1016/j.amjmed.2017.10.035] [Medline: 29126825]
46. Nakata N. Recent technical development of artificial intelligence for diagnostic medical imaging. Jpn J Radiol 2019 Feb 31;37(2):103-108. [doi: 10.1007/s11604-018-0804-6] [Medline: 30706381]
47. Recht M, Bryan RN. Artificial intelligence: threat or boon to radiologists? J Am Coll Radiol 2017 Nov;14(11):1476-1480. [doi: 10.1016/j.jacr.2017.07.007] [Medline: 28826960]
48. Recht MP, Dewey M, Dreyer K, Langlotz C, Niessen W, Prainsack B, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. Eur Radiol 2020 Jun;30(6):3576-3584. [doi: 10.1007/s00330-020-06672-5] [Medline: 32064565]
49. Safdar NM, Banja JD, Meltzer CC. Ethical considerations in artificial intelligence. Eur J Radiol 2020 Jan;122:108768. [doi: 10.1016/j.ejrad.2019.108768] [Medline: 31786504]
50. Savadjiev P, Chong J, Dohan A, Vakalopoulou M, Reinhold C, Paragios N, et al. Demystification of AI-driven medical image interpretation: past, present and future. Eur Radiol 2019 Mar 13;29(3):1616-1624. [doi: 10.1007/s00330-018-5674-x] [Medline: 30105410]
51. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022 Jan 20;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
52. van Hoek J, Huber A, Leichtle A, Härmä K, Hilt D, von Tengg-Kobligk H, et al. A survey on the future of radiology among radiologists, medical students and surgeons: students and surgeons tend to be more skeptical about artificial intelligence and radiologists may fear that other disciplines take over. Eur J Radiol 2019 Dec;121:108742. [doi: 10.1016/j.ejrad.2019.108742] [Medline: 31734640]
53. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

XSL•FO
RenderX

54. Pereira Cabral B, da Graça Derengowski Fonseca M, Batista Mota F. What is the future of cancer care? A technology foresight assessment of experts' expectations. Econ Innov New Technol 2018 Nov 26;28(6):635-652. [doi: 10.1080/10438599.2018.1549788]
55. Cabral BP, Bonventre JV, Wieringa F, Mota FB. Probing expert opinions on the future of kidney replacement therapies. Artif Organs 2021 Jan 15;45(1):79-87. [doi: 10.1111/aor.13784] [Medline: 32697860]
56. Mota F, Braga L, Rocha L, Cabral B. 3D and 4D bioprinted human model patenting and the future of drug development. Nat Biotechnol 2020 Jun 09;38(6):689-694. [doi: 10.1038/s41587-020-0540-1] [Medline: 32518405]
57. Rocha LF, Braga LA, Mota FB. Gene editing for treatment and prevention of human diseases: a global survey of gene editing-related researchers. Hum Gene Ther 2020 Aug 01;31(15-16):852-862. [doi: 10.1089/hum.2020.136] [Medline: 32718240]
58. Cabral BP, da Graça Derengowski Fonseca M, Mota FB. Long term prevention and vector control of arboviral diseases: what does the future hold? Int J Infect Dis 2019 Dec;89:169-174 [FREE Full text] [doi: 10.1016/j.ijid.2019.10.002] [Medline: 31606414]
59. Cabral BP, Braga LA, Syed-Abdul S, Mota FB. Future of artificial intelligence applications in cancer care: a global cross-sectional survey of researchers. Curr Oncol 2023 Mar 16;30(3):3432-3446 [FREE Full text] [doi: 10.3390/curroncol30030260] [Medline: 36975473]
60. Cabral B, Terlizzi V, Laselva O, Conte Filho C, Mota F. Anticipating new treatments for cystic fibrosis: a global survey of researchers. J Clin Med 2022 Feb 26;11(5):1283 [FREE Full text] [doi: 10.3390/jcm11051283] [Medline: 35268374]
61. Resolution 510 of 7 April 2016. Minister of State for Health, Brazil. URL: https://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510_07_04_2016.html [accessed 2024-04-29]
62. Dmitrienko A. Modelling paired categorical outcomes in clinical trials. Pharm Stat 2003 Dec 05;2(4):279-289. [doi: 10.1002/pst.48]
63. Mehta CP, Patel NR. IBM SPSS exact tests. IBM Corp. URL: https://www.ibm.com/docs/en/SSLVMB_27.0.0/pdf/en/IBM_SPSS_Exact_Tests.pdf [accessed 2024-04-29]
64. Devaraj SM, Rockette-Wagner B, Miller RG, Arena VC, Napoleone JM, Conroy MB, et al. The impact of a yearlong diabetes prevention program-based lifestyle intervention on cardiovascular health metrics. J Prim Care Community Health 2021 Jul 08;12:21501327211029816 [FREE Full text] [doi: 10.1177/21501327211029816] [Medline: 34236004]
65. Vismara L, Rollè L, Agostini F, Sechi C, Fenaroli V, Molgora S, et al. Perinatal parenting stress, anxiety, and depression outcomes in first-time mothers and fathers: a 3- to 6-months postpartum follow-up study. Front Psychol 2016 Jun 24;7:938 [FREE Full text] [doi: 10.3389/fpsyg.2016.00938] [Medline: 27445906]
66. Wall ER. Mann-Whitney test and t-test. J Vis Impair Blind 2023;117(1):100. [doi: 10.1007/springerreference_100245]
67. Herpel C, Drusko A, Schwindling FS, Rammelsberg P, Tesarz J. Head and neck pain drawing area correlates with higher psychosocial burden but not with joint dysfunction in temporomandibular disorders: a cross-sectional study. J Pain 2023 Jun;24(6):970-979 [FREE Full text] [doi: 10.1016/j.jpain.2023.01.010] [Medline: 36682594]
68. Hinds Z, Lockwood P. A cross-sectional student survey of the impact of the COVID-19 lockdowns on clinical placement in England. Radiography (Lond) 2023 Jan;29(1):190-199 [FREE Full text] [doi: 10.1016/j.radi.2022.11.006] [Medline: 36476513]
69. Torres-Puebla G, Poblete P, Rodríguez-Núñez I, Báez C, Muñoz-Pareja M, Zenteno D. The cardiorespiratory response and physiological determinants of the 6-min handbike cycle test in healthy children: a cross-sectional study. Pediatr Pulmonol 2023 Apr 16;58(4):1152-1159. [doi: 10.1002/ppul.26309] [Medline: 36617493]
70. Lorkowski J, Pokorski M. Medical records: a historical narrative. Biomedicines 2022 Oct 17;10(10):2594 [FREE Full text] [doi: 10.3390/biomedicines10102594] [Medline: 36289856]
71. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. J Biomed Inform 2019 Dec;100:103311 [FREE Full text] [doi: 10.1016/j.jbi.2019.103311] [Medline: 31629922]
72. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
73. Li J, Zhou L, Zhan Y, Xu H, Zhang C, Shan F, et al. How does the artificial intelligence-based image-assisted technique help physicians in diagnosis of pulmonary adenocarcinoma? A randomized controlled experiment of multicenter physicians in China. J Am Med Inform Assoc 2022 Nov 14;29(12):2041-2049 [FREE Full text] [doi: 10.1093/jamia/ocac179] [Medline: 36228127]
74. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. World J Gastroenterol 2022 Jan 07;28(1):108-122 [FREE Full text] [doi: 10.3748/wjg.v28.i1.108] [Medline: 35125822]
75. Wani SU, Khan NA, Thakur G, Gautam SP, Ali M, Alam P, et al. Utilization of artificial intelligence in disease prevention: diagnosis, treatment, and implications for the healthcare workforce. Healthcare (Basel) 2022 Mar 24;10(4):608 [FREE Full text] [doi: 10.3390/healthcare10040608] [Medline: 35455786]

XSL•FO
RenderX

76. Alami H, Rivard L, Lehoux P, Hoffman SJ, Cadeddu SB, Savoldelli M, et al. Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. Global Health 2020 Jun 24;16(1):52 [FREE Full text] [doi: 10.1186/s12992-020-00584-1] [Medline: 32580741]
77. Harmon DM, Attia ZI, Friedman PA. Current and future implications of the artificial intelligence electrocardiogram: the transformation of healthcare and attendant research opportunities. Cardiovasc Res 2022 Feb 21;118(3):e23-e25 [FREE Full text] [doi: 10.1093/cvr/cvac006] [Medline: 35187568]
78. Heising L, Angelopoulos S. Operationalising fairness in medical AI adoption: detection of early Alzheimer's disease with 2D CNN. BMJ Health Care Inform 2022 Apr;29(1):e100485 [FREE Full text] [doi: 10.1136/bmjhci-2021-100485] [Medline: 35477689]
79. Jha N, Lee KS, Kim YJ. Diagnosis of temporomandibular disorders using artificial intelligence technologies: a systematic review and meta-analysis. PLoS One 2022;17(8):e0272715 [FREE Full text] [doi: 10.1371/journal.pone.0272715] [Medline: 35980894]
80. Jones DT, Kerber KA. Artificial intelligence and the practice of neurology in 2035: the neurology future forecasting series. Neurology 2022 Feb 08;98(6):238-245. [doi: 10.1212/WNL.0000000000013200] [Medline: 35131918]
81. Hao S, Liu C, Li N, Wu Y, Li D, Gao Q, et al. Clinical evaluation of AI-assisted screening for diabetic retinopathy in rural areas of midwest China. PLoS One 2022 Oct 13;17(10):e0275983 [FREE Full text] [doi: 10.1371/journal.pone.0275983] [Medline: 36227905]
82. Gomez Rossi J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. JAMA Netw Open 2022 Mar 01;5(3):e220269 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.0269] [Medline: 35289862]
83. Parikh RB, Helmchen LA. Paying for artificial intelligence in medicine. NPJ Digit Med 2022 May 20;5(1):63 [FREE Full text] [doi: 10.1038/s41746-022-00609-6] [Medline: 35595986]
84. Khanna NN, Maindarkar MA, Viswanathan V, Fernandes JF, Paul S, Bhagawati M, et al. Economics of artificial intelligence in healthcare: diagnosis vs. treatment. Healthcare (Basel) 2022 Dec 09;10(12):2493 [FREE Full text] [doi: 10.3390/healthcare10122493] [Medline: 36554017]
85. Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Cuadrado-Godia E, et al. State-of-the-art review on deep learning in medical imaging. Front Biosci (Landmark Ed) 2019 Jan 01;24(3):392-426 [FREE Full text] [doi: 10.2741/4725] [Medline: 30468663]
86. Bridge J, Fu L, Lin W, Xue Y, Lip GY, Zheng Y. Artificial intelligence to detect abnormal heart rhythm from scanned electrocardiogram tracings. J Arrhythm 2022 Jun 29;38(3):425-431 [FREE Full text] [doi: 10.1002/joa3.12707] [Medline: 35785392]
87. Laad M, Kotecha K, Patil K, Pise R. Cardiac diagnosis with machine learning: a paradigm shift in cardiac care. Appl Artif Intell 2022 Jan 26;36(1):1-24. [doi: 10.1080/08839514.2022.2031816]
88. Żurek M, Jasak K, Niemczyk K, Rzepakowska A. Artificial intelligence in laryngeal endoscopy: systematic review and meta-analysis. J Clin Med 2022 May 12;11(10):2752 [FREE Full text] [doi: 10.3390/jcm11102752] [Medline: 35628878]
89. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604 [FREE Full text] [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
90. O'Sullivan ME, Considine EC, O'Riordan M, Marnane WP, Rennie JM, Boylan GB. Challenges of developing robust AI for intrapartum fetal heart rate monitoring. Front Artif Intell 2021 Oct 26;4:765210 [FREE Full text] [doi: 10.3389/frai.2021.765210] [Medline: 34765970]
91. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019 Mar 05;69(2):127-157 [FREE Full text] [doi: 10.3322/caac.21552] [Medline: 30720861]
92. Chen M, Zhang B, Cai Z, Seery S, Gonzalez MJ, Ali NM, et al. Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. Front Med (Lausanne) 2022 Aug 31;9:990604 [FREE Full text] [doi: 10.3389/fmed.2022.990604] [Medline: 36117979]
93. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
94. Biswas S. ChatGPT and the future of medical writing. Radiology 2023 Apr 01;307(2):e223312. [doi: 10.1148/radiol.223312] [Medline: 36728748]
95. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. Lancet Digit Health 2023 Mar;5(3):e105-e106 [FREE Full text] [doi: 10.1016/S2589-7500(23)00019-5] [Medline: 36754725]
96. Ordish J. Large language models and software as a medical device. Government of UK. URL: https://medregs.blog.gov.uk/2023/03/03/large-language-models-and-software-as-a-medical-device/ [accessed 2024-09-26]
97. Jaremko JL, Azar M, Bromwich R, Lum A, Alicia Cheong LH, Gibert M, et al. Canadian association of radiologists white paper on ethical and legal issues related to artificial intelligence in radiology. Can Assoc Radiol J 2019 May 01;70(2):107-118 [FREE Full text] [doi: 10.1016/j.carj.2019.03.001] [Medline: 30962048]

XSL•FO
RenderX

98. Kenny LM, Nevin M, Fitzpatrick K. Ethics and standards in the use of artificial intelligence in medicine on behalf of the Royal Australian and New Zealand College of Radiologists. J Med Imaging Radiat Oncol 2021 Aug 02;65(5):486-494 [FREE Full text] [doi: 10.1111/1754-9485.13289] [Medline: 34342139]
99. Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the joint European and north American multisociety statement. Radiology 2019 Nov;293(2):436-440. [doi: 10.1148/radiol.2019191586] [Medline: 31573399]
100. Cybersecurity and privacy in AI - medical imaging diagnosis. European Union Agency for Cybersecurity (ENISA). URL: https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research [accessed 2024-04-29]
101. Tulk Jesso S, Kelliher A, Sanghavi H, Martin T, Henrickson Parker S. Inclusion of clinicians in the development and evaluation of clinical artificial intelligence tools: a systematic literature review. Front Psychol 2022 Apr 7;13:830345 [FREE Full text] [doi: 10.3389/fpsyg.2022.830345] [Medline: 35465567]
102. The Lancet. AI in medicine: creating a safe and equitable future. Lancet 2023 Aug 12;402(10401):503. [doi: 10.1016/S0140-6736(23)01668-9] [Medline: 37573071]
103. Roche C, Wall PJ, Lewis D. Ethics and diversity in artificial intelligence policies, strategies and initiatives. AI Ethics 2022 Oct 06;3(4):1-21 [FREE Full text] [doi: 10.1007/s43681-022-00218-9] [Medline: 36246014]
104. Gampala S, Vankeshwaram V, Gadula SS. Is artificial intelligence the new friend for radiologists? A review article. Cureus 2020 Oct 24;12(10):e11137 [FREE Full text] [doi: 10.7759/cureus.11137] [Medline: 33240726]
105. Moawad AW, Fuentes DT, ElBanan MG, Shalaby AS, Guccione J, Kamel S, et al. Artificial intelligence in diagnostic radiology: where do we stand, challenges, and opportunities. J Comput Assist Tomogr 2022;46(1):78-90. [doi: 10.1097/RCT.0000000000001247] [Medline: 35027520]
106. Ursin F, Timmermann C, Steger F. Explicability of artificial intelligence in radiology: is a fifth bioethical principle conceptually necessary? Bioethics 2022 Feb 12;36(2):143-153. [doi: 10.1111/bioe.12918] [Medline: 34251687]

Abbreviations

AI: artificial intelligence
MeSH: Medical Subject Headings
SCI: Science Citation Index
W1: wave 1
W2: wave 2
WoS: Web of Science

Edited by A Mavragani; submitted 23.10.23; peer-reviewed by A Hidki, R Bidkar, L He, B Mesko; comments to author 21.03.24; revised version received 06.05.24; accepted 18.10.24; published 27.02.25.

Please cite as:
Cabral BP, Braga LAM, Conte Filho CG, Penteado B, Freire de Castro Silva SL, Castro L, Fornazin M, Mota F
Future Use of AI in Diagnostic Medicine: 2-Wave Cross-Sectional Survey Study
J Med Internet Res 2025;27:e53892
URL: https://www.jmir.org/2025/1/e53892
doi:10.2196/53892
PMID:40053779

©Bernardo Pereira Cabral, Luiza Amara Maciel Braga, Carlos Gilbert Conte Filho, Bruno Penteado, Sandro Luis Freire de Castro Silva, Leonardo Castro, Marcelo Fornazin, Fabio Mota. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of AI for Enhancing Computed Tomography Image Quality and Radiation Protection in Radiology: Systematic Review and Meta-Analysis

Subo Zhang[1,2,3*], BS; Zhitao Zhu[1,2,3*], BS; Zhenfei Yu[1,2,3], BS; Haifeng Sun[1,2,3], BS; Yi Sun[1,2,3], BS; Hai Huang[1,2,3], BS; Lei Xu[1,2,3], BS; Jinxin Wan[1,2,3], BS

[1]Department of Medical Imaging, The Second People's Hospital of Lianyungang, Lianyungang, China

[2]Department of Medical Imaging, Cancer Hospital of Lianyungang, Lianyungang, China

[3]Lianyungang Clinical College, Jiangsu University, Lianyungang, China

[*]these authors contributed equally

Corresponding Author:
Jinxin Wan, BS
Department of Medical Imaging
The Second People's Hospital of Lianyungang
No 41 Hailian East Road
Lianyungang, 222000
China
Phone: 86 051885775003
Email: jxwlyg@126.com

Abstract

Background: Artificial intelligence (AI) presents a promising approach to balancing high image quality with reduced radiation exposure in computed tomography (CT) imaging.

Objective: This meta-analysis evaluates the effectiveness of AI in enhancing CT image quality and lowering radiation doses.

Methods: A thorough literature search was performed across several databases, including PubMed, Embase, Web of Science, Science Direct, and Cochrane Library, with the final update in 2024. We included studies that compared AI-based interventions to conventional CT techniques. The quality of these studies was assessed using the Newcastle-Ottawa Scale. Random effect models were used to pool results, and heterogeneity was measured using the I^2 statistic. Primary outcomes included image quality, CT dose index, and diagnostic accuracy.

Results: This meta-analysis incorporated 5 clinical validation studies published between 2022 and 2024, totaling 929 participants. Results indicated that AI-based interventions significantly improved image quality (mean difference 0.70, 95% CI 0.43-0.96; P<.001) and showed a positive trend in reducing the CT dose index, though not statistically significant (mean difference 0.47, 95% CI –0.21 to 1.15; P=.18). AI also enhanced image analysis efficiency (odds ratio 1.57, 95% CI 1.08-2.27; P=.02) and demonstrated high accuracy and sensitivity in detecting intracranial aneurysms, with low-dose CT using AI reconstruction showing noninferiority for liver lesion detection.

Conclusions: The findings suggest that AI-based interventions can significantly enhance CT imaging practices by improving image quality and potentially reducing radiation doses, which may lead to better diagnostic accuracy and patient safety. However, these results should be interpreted with caution due to the limited number of studies and the variability in AI algorithms. Further research is needed to clarify AI's impact on radiation reduction and to establish clinical standards.

(J Med Internet Res 2025;27:e66622) doi:10.2196/66622

KEYWORDS

artificial intelligence; computed tomography; image quality; radiation protection; meta-analysis

XSL•FO
RenderX

Introduction

Computed tomography (CT) has revolutionized medical imaging since its introduction, providing high-resolution, 3D visualizations of anatomical structures. Its widespread adoption has significantly enhanced diagnostic capabilities across various medical specialties [1-3]. However, the increasing use of CT scans has raised concerns about radiation exposure and its potential long-term health effects on patients [4,5].

Simultaneously, maintaining optimal image quality is crucial for accurate diagnosis and treatment planning. This creates a challenging balance between minimizing radiation dose and preserving diagnostic image quality [6-8]. In recent years, artificial intelligence (AI) has emerged as a promising solution to address the balance between maintaining high image quality and minimizing radiation dose in CT imaging [9-11].

AI, particularly deep learning algorithms, has shown potential in various aspects of CT imaging, including image reconstruction, noise reduction, and automated image analysis [12]. These AI-driven approaches aim to enhance image quality while allowing for lower radiation doses, potentially improving both diagnostic accuracy and patient safety [13].

Several studies have investigated the application of AI in CT imaging, focusing on different anatomical regions and clinical scenarios. For instance, Hu et al [14] explored the use of deep learning models for intracranial aneurysm detection on CT angiography images. Zhang et al [15] investigated AI reconstruction algorithms in CT imaging of sports injuries. Other researchers have examined AI's role in breast cancer radiotherapy planning [16] and liver CT imaging [17].

Despite the growing body of research, the overall effectiveness of AI in improving CT image quality and reducing radiation dose remains unclear [14,17]. The variability in study designs, AI algorithms, and outcome measures makes it challenging to draw definitive conclusions about the broader impact of AI on CT imaging practices [18]. This meta-analysis aims to synthesize the available evidence on the role of AI in CT image quality control and radiation protection, providing a comprehensive evaluation of AI's role across various clinical scenarios and anatomical regions. Our study contributes to the literature by offering a systematic overview of the current state of AI in CT imaging, which can inform future research directions and clinical applications.

Methods

Search Strategy and Study Selection

A comprehensive literature search was conducted across multiple electronic databases, including PubMed, Embase, Web of Science, Science Direct, and Cochrane Library. The search strategy used a combination of Medical Subject Headings terms and keywords related to AI, deep learning, and CT imaging. The specific search string used was as follows: (Intelligence, Artificial [Title/Abstract]) OR (Computer Reasoning [Title/Abstract]) OR (Reasoning, Computer [Title/Abstract]) OR (AI (Artificial Intelligence [Title/Abstract]) OR (Machine Intelligence [Title/Abstract])) and so on (Multimedia Appendix 1).

The selected databases are renowned for their extensive coverage of medical and scientific literature, ensuring a thorough retrieval of studies pertinent to our research question. The choice of these databases was also influenced by their indexing of a wide array of medical journals and conference proceedings, which are critical for capturing the latest advancements in the field of AI and CT imaging.

The search was conducted without language restrictions and conducted from the inception of the databases until the search date in 2024, to ensure inclusion of the most current and relevant studies.

Inclusion and Exclusion Criteria

Studies were eligible for inclusion if they met the following criteria: (1) clinical validation studies comparing AI-based interventions with conventional CT imaging techniques; (2) reported outcomes related to image quality, radiation dose, or diagnostic performance; and (3) provided sufficient data for quantitative analysis. Exclusion criteria included (1) nonclinical studies or (2) those without a control group, (3) studies focusing solely on AI development without clinical validation, (4) review articles, (5) case reports, (6) conference abstracts, and (7) studies with insufficient data for meta-analysis.

Data Extraction and Quality Assessment

Two independent reviewers extracted data from the included studies using a standardized form. The extracted information included study characteristics (author, year, and study design), sample size and patient demographics, AI intervention details, and outcome measures (image quality metrics, radiation dose indicators, and diagnostic performance).

The quality of the included studies was assessed using the Newcastle-Ottawa Scale for nonrandomized studies. This scale evaluates studies based on selection, comparability, and outcome domains, with a maximum score of 9 indicating the highest quality. The risk of bias for each included study was assessed using the Cochrane risk of bias tool, which provides a standardized evaluation of study quality.

Statistical Analysis

Meta-analyses were performed using Review Manager (version 5.4; The Cochrane Collaboration). For continuous outcomes (eg, image quality scores and CT dose index), mean differences with 95% CIs were calculated. For binary outcomes (eg, diagnostic accuracy), odds ratios with 95% CIs were computed.

Random effect models were used to account for potential heterogeneity among studies. Heterogeneity was assessed using the I^2 statistic, with values of 25%, 50%, and 75% considered as low, moderate, and high heterogeneity, respectively.

Publication bias was assessed visually using funnel plots. Sensitivity analyses were conducted by sequentially removing each study to evaluate its impact on the overall effect size.

Results were considered statistically significant at $P<.05$. Forest plots were generated to visually represent the meta-analysis

XSL•FO
RenderX

results, displaying individual study effects and the pooled effect size.

Ethical Considerations

An ethical statement is not applicable because this study is based exclusively on published literature.

Results

Study Selection and Characteristics

The initial database search yielded 835 records (PubMed: n=38, Embase: n=191, Web of Science: n=3, Science Direct: n=283, and Cochrane Library: n=320). After removing duplicates, 371 unique records remained. Following title and abstract screening, 11 full-text articles were assessed for eligibility. Ultimately, 5 studies met the inclusion criteria and were included in the meta-analysis (Figure 1).

The 5 included studies were all clinical validation studies published between 2022 and 2024. They encompassed a total of 929 participants (464 in the experimental AI group and 465 in the control group). The studies focused on various anatomical regions and clinical applications, including intracranial aneurysm detection [14], knee anterior cruciate ligament injury [15], breast cancer radiotherapy planning [16], head CT angiography [9], and liver CT imaging [17].

All included studies used deep learning–based AI algorithms for image reconstruction or analysis. The sample sizes ranged from 60 to 14,715 participants. The Newcastle-Ottawa Scale scores for the included studies ranged from 7 to 8, indicating good methodological quality (Table 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for study selection, detailing the process from initial records to final inclusion across various databases and time frames.

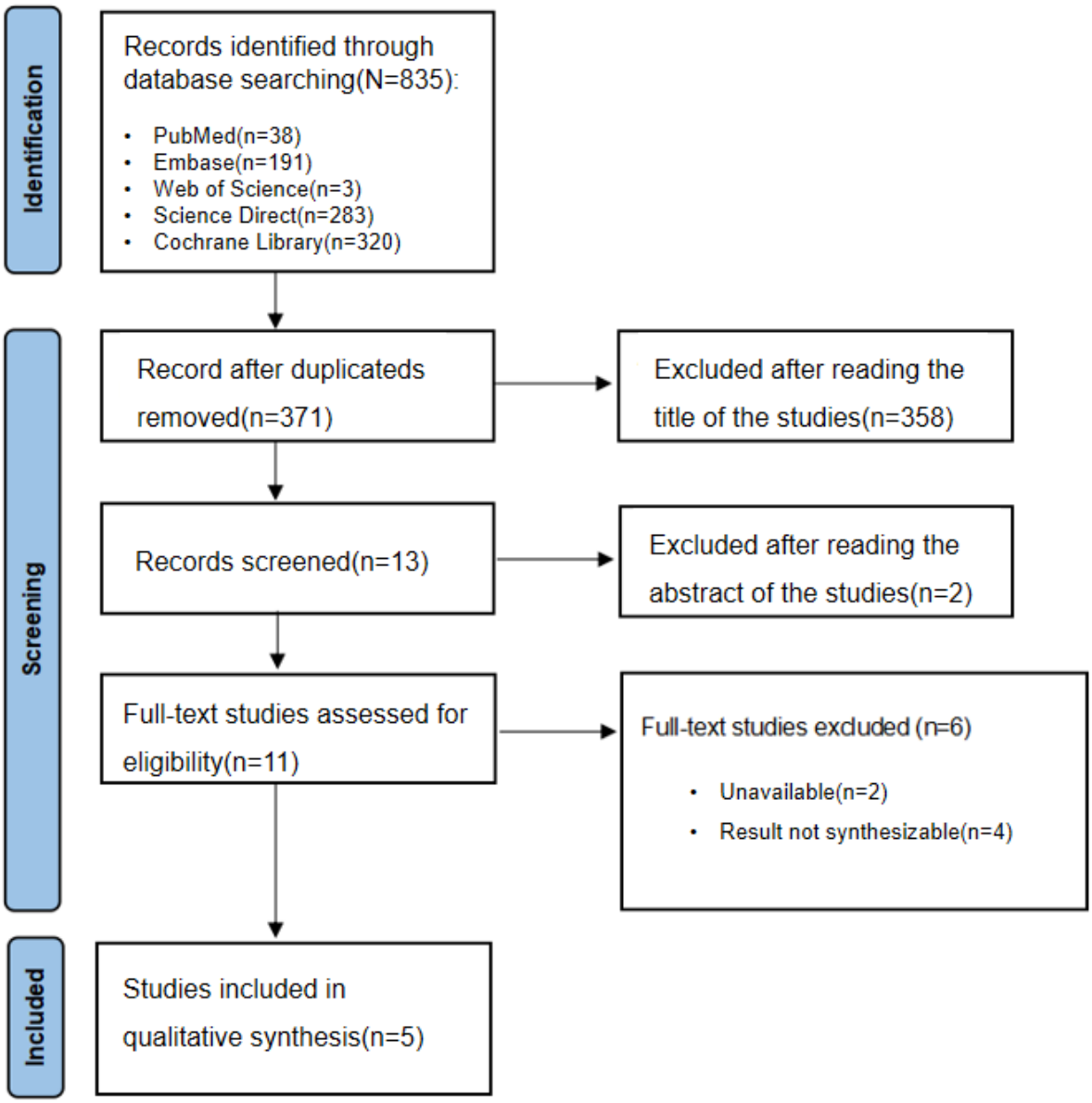

XSL•FO
RenderX

Table 1. Basic characteristics of included literature.

Author (year)	Type of invention group	Type of study	Training dataset, n	Internal validation dataset, n	External validation dataset, n	Exposure factors	NOS[a] score
Hu et al [14] (2024)	AI[b]	Clinical validation study	12,817	1700	1198	Accuracy, sensitivity, positive predictive value, and negative predictive value	7
Zhang et al [15] (2022)	AI	Clinical validation study	30	30	30	Image quality, CTDI[c] volume, and dose length generation (DLP[d])	8
Ma et al [16] (2023)	AI	Clinical validation study	230	280	22	Image quality	8
Huang et al [9] (2023)	AI	Clinical validation study	63	38	25	CTDI, CM[e] dose, and image quality	7
Lee et al [17] (2024)	AI	Clinical validation study	296	246	90	CT image noise level and diagnostic performance	8

[a]NOS: Newcastle-Ottawa Scale.

[b]AI: artificial intelligence.

[c]CTDI: computed tomography dose index.

[d]DLP: dose length product.

[e]CM: contrast medium.

Risk-of-Bias Assessment

Overview

The risk of bias for each included study was assessed using the Cochrane risk of bias tool. Figure 2 presents the risk of bias graph, showing the distribution of judgments across all included studies for each domain. Random sequence generation (selection bias) was assessed as having a low risk in 40% (2/5) of studies and unclear risk in 60% (3/5). Allocation concealment was judged as low risk in 80% (4/5) of studies and unclear in 20% (1/5). The blinding of participants and personnel (performance bias) was unanimously assessed as an unclear risk across all studies. Blinding outcome assessment (detection bias) was determined to be low risk in 80% (4/5) of studies and unclear in 20% (1/5). Incomplete outcome data (attrition bias) were consistently judged as low risk across all studies. Selective reporting (reporting bias) was assessed as low risk in 40% (2/5) of studies and unclear in 60% (3/5). Other bias was unanimously judged as unclear across all studies.

Figure 3 [9,14-17] presents a comprehensive risk-of-bias summary for each study included in the analysis. Several key observations can be made from this summary. Notably, all 5 studies demonstrated a low risk for incomplete outcome data, indicating robust data collection and reporting practices. However, the blinding of participants and personnel was unclear across all studies, which is a common limitation in imaging studies where complete blinding can be challenging to achieve. Random sequence generation was assessed as low risk in 3 (60%) out of the 5 studies, suggesting generally sound randomization practices were used. Interestingly, the “other bias” category was consistently rated as unclear for all studies, pointing to potential unidentified sources of bias that may warrant further investigation. Finally, there was considerable variability in the assessment of selective reporting and blinding of outcome assessment across the studies, highlighting areas where methodological consistency could be improved in future research.

XSL•FO
RenderX

Figure 2. Risk-of-bias graph: review authors' judgments about each risk-of-bias item presented as percentages across all included studies. This graph provides a visual representation of the risk of bias across different domains, indicating the methodological quality of the studies included in the meta-analysis.

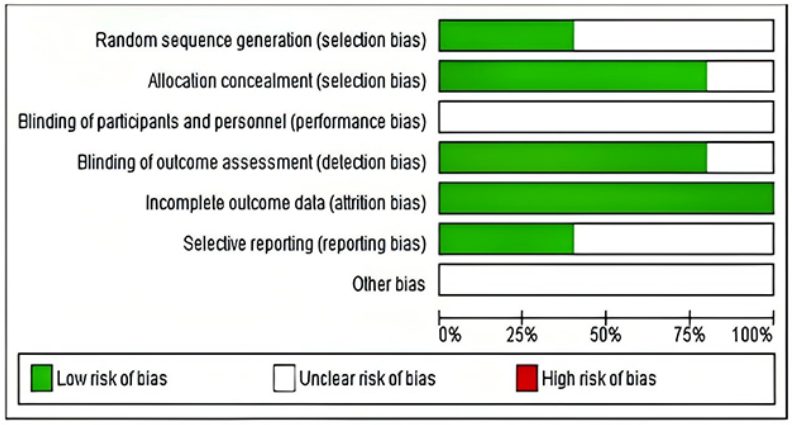

Figure 3. Risk-of-bias summary: review authors' judgments about each risk-of-bias item for each included study. This summary offers a comprehensive overview of the risk of bias for each study, highlighting areas of strength and potential concerns within the meta-analysis.

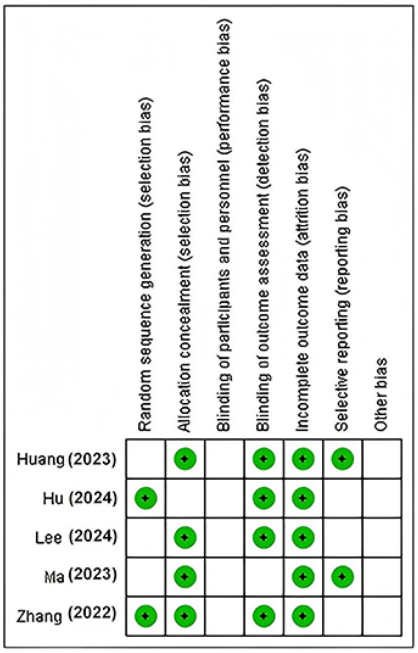

Image Quality

All 5 studies reported on image quality outcomes. The meta-analysis showed a significant improvement in image quality with AI-based interventions compared to conventional methods (mean difference 0.70, 95% CI 0.43-0.96; *P*<.001; Figure 4 [9,14-17]). There was moderate heterogeneity among the studies (I^2=39%), suggesting some variability in the effect of AI on image quality across different clinical contexts.

The largest effects were observed in the studies by Hu et al [14] and Lee et al [17], which focused on intracranial aneurysm detection and liver CT imaging, respectively. These studies carried the highest weights in the analysis (27.6% and 33.8%, respectively).

XSL•FO
RenderX

Figure 4. Forest plot comparing image quality scores between AI-based interventions and conventional CT imaging techniques across 5 clinical validation studies conducted from 2022 to 2024, focusing on various anatomical regions and clinical applications. This plot visually represents the mean differences in image quality, with confidence intervals, favoring either the experimental or control groups. AI: artificial intelligence; CT: computed tomography; IV: inverse variance.

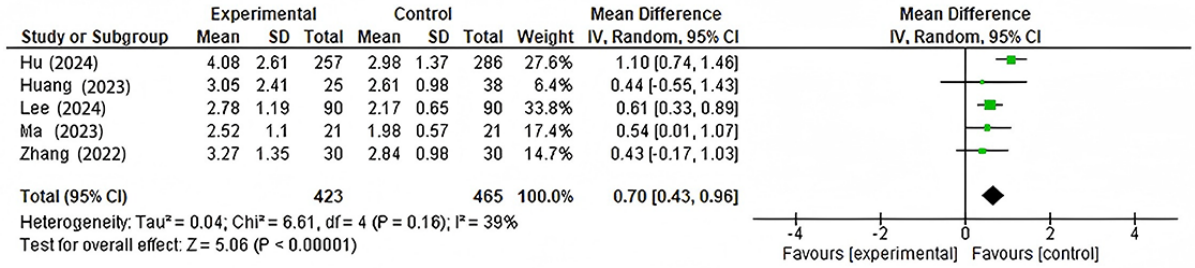

CT Dose Index

Three studies [9,15,17] reported data on the CT dose index. The meta-analysis showed a trend toward reduction in CT dose index with AI-based interventions, but the result was not statistically significant (mean difference 0.47, 95% CI –0.21 to 1.15; *P*=.18; Figure 5 [9,14-17]). There was high heterogeneity among these studies (I^2=87%), indicating substantial variability in the effect of AI on radiation dose across different applications.

The study by Zhang et al [15] showed the largest effect size in favor of AI, while Lee et al [17] reported almost no difference between the AI and conventional groups.

Figure 5. Forest plots of changes in CT dose index. This figure displays the impact of AI-based interventions on the CT dose index, with mean differences and CIs, across 3 studies that reported this outcome. AI: artificial intelligence; CT: computer tomography; IV: inverse variance.

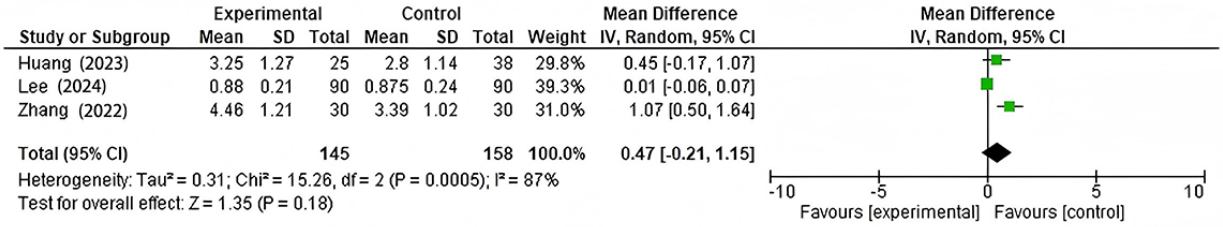

Efficiency in Image Analysis

All 5 studies provided data on the efficiency of AI in image analysis. The meta-analysis demonstrated a significant improvement in efficiency with AI-based methods compared to conventional approaches (odds ratio 1.57, 95% CI 1.08-2.27; *P*=.02; Figure 6 [9,14-17]). There was no significant heterogeneity among the studies for this outcome (I^2=0%).

The study by Lee et al [17] had the largest weight (39.7%) in this analysis, likely due to its larger sample size. All studies showed an odds ratio favoring the AI group, although the precision of the estimates varied, with wider confidence intervals observed in smaller studies.

Figure 6. Forest plots of efficiency in image analysis. This forest plot demonstrates the comparison of image analysis efficiency between AI-based methods and conventional approaches, showing a significant improvement with AI across all 5 studies. AI: artificial intelligence; M-H: Mantel-Haenszel analysis.

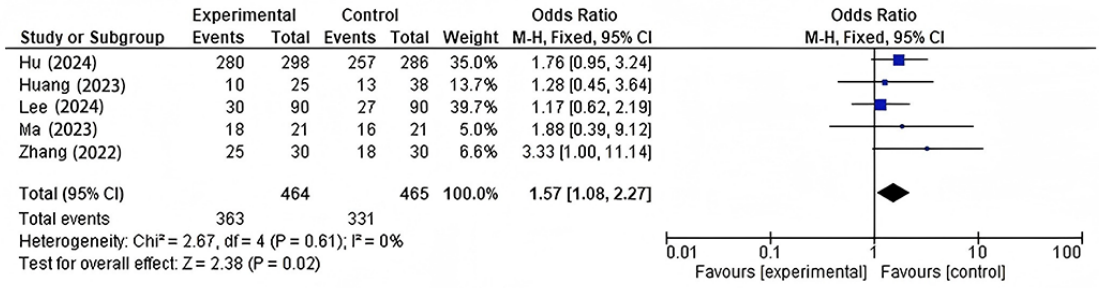

Diagnostic Performance

While not all studies reported on diagnostic performance metrics, those that did showed promising results. Hu et al [14] reported high accuracy (0.951), sensitivity (0.974), and specificity (0.928) for their AI model in detecting intracranial aneurysms. Lee et al [17] found that the diagnostic performance of low-dose CT with AI reconstruction was noninferior to standard-dose CT for liver lesion detection.

XSL•FO
RenderX

Publication Bias

Visual inspection of the funnel plot (Figure 7 [9,14-17]) did not reveal substantial asymmetry, suggesting a low risk of publication bias. However, the small number of included studies limits the reliability of this assessment.

Figure 7. Funnel plot of efficiency in image analysis by AI. The funnel plot is used to assess publication bias, with the distribution of effect sizes visually represented against their precision. AI: artificial intelligence; OR: odds ratio.

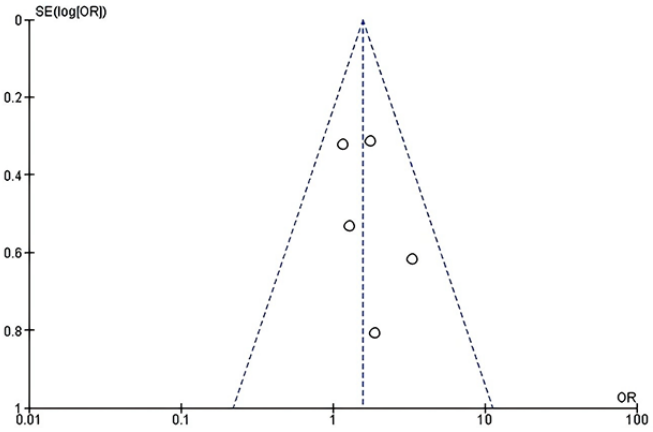

Discussion

Principal Findings

This meta-analysis synthesized evidence from 5 clinical validation studies to evaluate the role of AI in CT image quality control and radiation protection. The findings suggest that AI-based interventions can significantly improve CT image quality and efficiency in image analysis, with a potential trend toward radiation dose reduction.

The significant improvement in image quality observed with AI-based interventions (mean difference 0.70, 95% CI 0.43-0.96; P<.001) is a key finding of this meta-analysis. This improvement was consistent across different anatomical regions and clinical applications, suggesting that AI algorithms have broad applicability in enhancing CT image quality. The enhanced image quality achieved through AI could have substantial clinical implications. Improved image clarity and detail may lead to more accurate diagnoses, potentially reducing the need for repeat scans or additional imaging studies [19]. This, in turn, could contribute to reduced overall radiation exposure for patients and improved workflow efficiency in radiology departments.

While the meta-analysis showed a trend toward CT dose index reduction with AI-based interventions, the result was not statistically significant (mean difference 0.47, 95% CI –0.21 to 1.15; P=.18). The lack of statistical significance does not necessarily imply that AI has no effect on radiation dose reduction. Rather, it highlights the need for further research to elucidate the factors influencing AI's impact on radiation dose across different CT applications. It is possible that some AI algorithms are more effective at dose reduction than others, or that dose reduction capabilities may be more pronounced in certain anatomical regions or clinical scenarios [20].

Moreover, the trend toward dose reduction, even if not statistically significant, is encouraging. When combined with the significant improvements in image quality, these findings suggest that AI may enable a favorable shift in the balance between image quality and radiation dose in CT imaging [21].

The significant improvement in the efficiency of image analysis with AI-based methods (odds ratio 1.57, 95% CI 1.08-2.27; P=.02) is another important finding of this meta-analysis. This improvement in efficiency could have far-reaching implications for radiology workflows and patient care. Faster image analysis times could lead to reduced reporting turnaround times, potentially enabling quicker clinical decision-making and treatment initiation [22]. Additionally, improved efficiency could help address the growing demand for CT imaging services and the increasing workload on radiologists [23]. The lack of heterogeneity (I^2=0%) in this analysis suggests that the efficiency gains with AI are consistent across different applications. This consistency is particularly noteworthy given the diversity of clinical scenarios represented in the included studies.

Comparison With Prior Work

While not all studies reported comprehensive diagnostic performance metrics, those that did showed promising results. The high accuracy, sensitivity, and specificity reported by Hu et al [14] for intracranial aneurysm detection, as well as the noninferiority of low-dose CT with AI reconstruction for liver lesion detection reported by Lee et al [17], suggest that AI can maintain or even enhance diagnostic performance while potentially reducing radiation dose. These findings align with previous studies that have demonstrated the potential of AI to improve diagnostic accuracy in various imaging modalities [24]. The combination of improved image quality, potential dose reduction, and maintained or enhanced diagnostic performance positions AI as a powerful tool for advancing CT imaging practices.

The principle of action by which AI, particularly deep learning algorithms, enhances CT image quality is multifaceted. First, AI-driven approaches use advanced computational techniques to reconstruct images from raw data, optimizing the balance between image sharpness and noise reduction. This is achieved

XSL•FO
RenderX

through the use of convolutional neural networks that learn to identify and enhance features of interest while suppressing background noise, leading to clearer and more detailed images [24]. Second, AI algorithms can adaptively adjust image reconstruction parameters based on the specific characteristics of the scanned tissue, resulting in improved contrast resolution and better visualization of anatomical structures. Third, AI-based techniques offer the capability for automated image analysis, which can standardize the interpretation process and reduce interobserver variability, thereby contributing to more accurate and consistent diagnostic outcomes [22]. These mechanisms collectively contribute to the significant improvement in image quality observed in our meta-analysis, underscoring the potential of AI to revolutionize CT imaging practices by enhancing diagnostic accuracy and patient safety.

Strengths and Limitations

Several limitations of this meta-analysis should be noted. First, the small number of included studies limits the generalizability of the findings and precludes more detailed subgroup analyses. Second, the heterogeneity in AI algorithms, CT protocols, and outcome measures across studies made direct comparisons challenging. Third, the lack of long-term, follow-up data in the included studies means that the potential long-term impacts of AI-enhanced CT imaging on patient outcomes remain unclear. Future research should focus on larger, multicenter studies with standardized protocols and outcome measures to more definitively establish the impact of AI on CT image quality and radiation dose [25]. Long-term studies are needed to assess the clinical impact of AI-enhanced CT imaging on patient outcomes and radiation-induced cancer risk. Additionally, research into the cost-effectiveness of implementing AI in CT imaging workflows would be valuable for health care systems considering the adoption of these technologies.

Conclusions

This meta-analysis provides evidence that AI-based interventions can significantly improve CT image quality and efficiency in image analysis, with the potential for radiation dose reduction. These findings suggest a promising role for AI in enhancing CT imaging practices, potentially leading to improved diagnostic accuracy, reduced radiation exposure, and enhanced patient care. As AI technologies continue to evolve, their integration into CT imaging workflows may become increasingly important in addressing the ongoing challenges of balancing image quality, radiation safety, and diagnostic accuracy.

Acknowledgments

The fundings for this study were provided by The Health Science and Technology Project of the Lianyungang Health Commission（202325）and The Scientific Research Project of the Lianyungang Anti-Cancer Association (MS202312)

Data Availability

All data generated or analyzed during this study are included in this paper. Further inquiries can be directed to the corresponding author.

Authors' Contributions

JW and YS conceived the study, and SZ, ZZ, ZY, HS, HH, and LX participated in its design and data analysis and statistics. SZ and ZZ helped to draft the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed Search Strategy for "The Role of Artificial Intelligence in CT Image Quality Control and Radiation Protection: A Meta-Analysis".
[DOCX File , 16 KB - jmir_v27i1e66622_app1.docx]

Multimedia Appendix 2
PRISMA checklist.
[PDF File (Adobe PDF File), 97 KB - jmir_v27i1e66622_app2.pdf]

References

1. Kalra MK, Maher MM, Toth TL, Hamberg LM, Blake MA, Shepard J, et al. Strategies for CT radiation dose optimization. Radiology 2004;230(3):619-628. [doi: 10.1148/radiol.2303021726] [Medline: 14739312]
2. Kleber CEJ, Karius R, Naessens LE, van Toledo CO, van Osch JAC, Boomsma MF, et al. Advancements in supervised deep learning for metal artifact reduction in computed tomography: a systematic review. Eur J Radiol 2024;181:111732 [FREE Full text] [doi: 10.1016/j.ejrad.2024.111732] [Medline: 39265203]

XSL•FO
RenderX

3. Jeong G, Li F, Mitcham TM, Villa U, Duric N, Anastasio MA. Investigating the use of traveltime and reflection tomography for deep learning-based sound-speed estimation in ultrasound computed tomography. IEEE Trans Ultrason Ferroelectr Freq Control 2024:1358-1376. [doi: 10.1109/TUFFC.2024.3459391] [Medline: 39264782]
4. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. N Engl J Med 2007;357(22):2277-2284. [doi: 10.1056/NEJMra072149] [Medline: 18046031]
5. Lohiya N, Hussein M, Sahu AK, Aggarwal B, Maheshwari J, Iyengar KP, et al. Assessing the current role of AP and Bernageau view radiographs in measurement of glenoid bone loss in patients with recurrent shoulder dislocation: correlation with computed tomography, magnetic resonance imaging, and arthroscopy. Skeletal Radiol 2024. [doi: 10.1007/s00256-024-04797-y] [Medline: 39264417]
6. McCollough CH, Primak AN, Braun N, Kofler J, Yu L, Christner J. Strategies for reducing radiation dose in CT. Radiol Clin North Am 2009;47(1):27-40 [FREE Full text] [doi: 10.1016/j.rcl.2008.10.006] [Medline: 19195532]
7. Tunissen SAM, Smit EJ, Mikerov M, Prokop M, Sechopoulos I. Performance evaluation of a 4D similarity filter for dynamic CT angiography imaging of the liver. Med Phys 2024;51(12):8814-8827. [doi: 10.1002/mp.17394] [Medline: 39264288]
8. Li W, Gong Y, Zhang J, Liu J, Li J, Fu S, et al. Construction of CXCR4 receptor-targeted CuFeSe nano theranostic platform and its application in MR/CT dual model imaging and photothermal therapy. Int J Nanomedicine 2024;19:9213-9226 [FREE Full text] [doi: 10.2147/IJN.S470367] [Medline: 39263631]
9. Huang X, Zhao W, Wang G, Wang Y, Li J, Li Y, et al. Improving image quality with deep learning image reconstruction in double-low-dose head CT angiography compared with standard dose and adaptive statistical iterative reconstruction. Br J Radiol 2023;96(1143):20220625 [FREE Full text] [doi: 10.1259/bjr.20220625] [Medline: 36606518]
10. Yang Y, Zhou Z, Zhang N, Wang R, Gao Y, Ran X, et al. Performance of artificial intelligence in detecting the chronic total occlusive lesions of coronary artery based on coronary computed tomographic angiography. Cardiovasc Diagn Ther 2024;14(4):655-667 [FREE Full text] [doi: 10.21037/cdt-23-407] [Medline: 39263478]
11. Huang H, Yan Z, Li B, Lu W, He P, Fan L, et al. LungPath: artificial intelligence–driven histologic pattern recognition for improved diagnosis of early-stage invasive lung adenocarcinoma. Transl Lung Cancer Res 2024;13(8):1816-1827 [FREE Full text] [doi: 10.21037/tlcr-24-258] [Medline: 39263012]
12. Willemink MJ, Noël PB. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. Eur Radiol 2019;29(5):2185-2195 [FREE Full text] [doi: 10.1007/s00330-018-5810-7] [Medline: 30377791]
13. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. Med Phys 2019;46(1):e1-e36 [FREE Full text] [doi: 10.1002/mp.13264] [Medline: 30367497]
14. Hu B, Shi Z, Lu L, Miao Z, Wang H, Zhou Z, et al. A deep-learning model for intracranial aneurysm detection on CT angiography images in China: a stepwise, multicentre, early-stage clinical validation study. Lancet Digit Health 2024;6(4):e261-e271 [FREE Full text] [doi: 10.1016/S2589-7500(23)00268-6] [Medline: 38519154]
15. Zhang H, Zheng H, Deng R, Luo K, Duan S. Computed tomography imaging under artificial intelligence reconstruction algorithm used in recovery of sports injury of the knee anterior cruciate ligament. Contrast Media Mol Imaging 2022;2022:1199841 [FREE Full text] [doi: 10.1155/2022/1199841] [Medline: 35685654]
16. Ma L, Yang Y, Ma J, Mao L, Li X, Feng L, et al. Correlation between AI-based CT organ features and normal lung dose in adjuvant radiotherapy following breast-conserving surgery: a multicenter prospective study. BMC Cancer 2023;23(1):1085 [FREE Full text] [doi: 10.1186/s12885-023-11554-2] [Medline: 37946125]
17. Lee DH, Lee JM, Lee CH, Afat S, Othman A. Image quality and diagnostic performance of low-dose liver CT with deep learning reconstruction versus standard-dose CT. Radiol Artif Intell 2024;6(2):e230192 [FREE Full text] [doi: 10.1148/ryai.230192] [Medline: 38231025]
18. Shetty S, Mubarak AS, R David L, Al Jouhari MO, Talaat W, Al-Rawi N, et al. The application of mask region–based convolutional neural networks in the detection of nasal septal deviation using cone beam computed tomography images: proof-of-concept study. JMIR Form Res 2024;8:e57335 [FREE Full text] [doi: 10.2196/57335] [Medline: 39226096]
19. Yasaka K, Abe O. Deep learning and artificial intelligence in radiology: current applications and future directions. PLoS Med 2018;15(11):e1002707 [FREE Full text] [doi: 10.1371/journal.pmed.1002707] [Medline: 30500815]
20. Greffier J, Hamard A, Pereira F, Barrau C, Pasquier H, Beregi JP, et al. Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. Eur Radiol 2020;30(7):3951-3959. [doi: 10.1007/s00330-020-06724-w] [Medline: 32100091]
21. Yao Y, Guo B, Li J, Yang Q, Li X, Deng L. The influence of a deep learning image reconstruction algorithm on the image quality and auto-analysis of pulmonary nodules at ultra-low dose chest CT: a phantom study. Quant Imaging Med Surg 2022;12(5):2777-2791 [FREE Full text] [doi: 10.21037/qims-21-815] [Medline: 35502370]
22. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
23. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence—an ESR white paper. Insights Imaging 2019;10(1):44. [doi: 10.1186/s13244-019-0738-2] [Medline: 30949865]
24. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88. [doi: 10.1016/j.media.2017.07.005] [Medline: 28778026]

XSL•FO
RenderX

25. Gillan C, Hodges B, Wiljer D, Dobrow M. Health care professional association agency in preparing for artificial intelligence: protocol for a multi-case study. JMIR Res Protoc 2021;10(5):e27340 [FREE Full text] [doi: 10.2196/27340] [Medline: 34009136]

Abbreviations

AI: artificial intelligence
CT: computed tomography
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Coristine; submitted 20.09.24; peer-reviewed by X Gu, LT Qian; comments to author 22.10.24; revised version received 27.11.24; accepted 16.12.24; published 27.02.25.

Please cite as:
Zhang S, Zhu Z, Yu Z, Sun H, Sun Y, Huang H, Xu L, Wan J
Effectiveness of AI for Enhancing Computed Tomography Image Quality and Radiation Protection in Radiology: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e66622
URL: https://www.jmir.org/2025/1/e66622
doi:10.2196/66622
PMID:40053787

©Subo Zhang, Zhitao Zhu, Zhenfei Yu, Haifeng Sun, Yi Sun, Hai Huang, Lei Xu, Jinxin Wan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Deep Learning–Based Electrocardiogram Model (EIANet) to Predict Emergency Department Cardiac Arrest: Development and External Validation Study

Shao-Chi Lu[1], MS; Guang-Yuan Chen[1], BS; An-Sheng Liu[1], MS; Jen-Tang Sun[2], MD; Jun-Wan Gao[3], BS; Chien-Hua Huang[3], MD, PhD; Chu-Lin Tsai[3*], MD, SCD; Li-Chen Fu[1*], PhD

[1]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

[2]Department of Emergency Medicine, Far Eastern Memorial Hospital, Taipei, Taiwan

[3]Department of Emergency Medicine, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chu-Lin Tsai, MD, SCD
Department of Emergency Medicine
National Taiwan University Hospital and National Taiwan University College of Medicine
No.1 Jen-Ai Road
Taipei, 100
Taiwan
Phone: 886 23123456 ext 267684
Email: chulintsai@ntu.edu.tw

Abstract

Background: In-hospital cardiac arrest (IHCA) is a severe and sudden medical emergency that is characterized by the abrupt cessation of circulatory function, leading to death or irreversible organ damage if not addressed immediately. Emergency department (ED)–based IHCA (EDCA) accounts for 10% to 20% of all IHCA cases. Early detection of EDCA is crucial, yet identifying subtle signs of cardiac deterioration is challenging. Traditional EDCA prediction methods primarily rely on structured vital signs or electrocardiogram (ECG) signals, which require additional preprocessing or specialized devices. This study introduces a novel approach using image-based 12-lead ECG data obtained at ED triage, leveraging the inherent richness of visual ECG patterns to enhance prediction and integration into clinical workflows.

Objective: This study aims to address the challenge of early detection of EDCA by developing an innovative deep learning model, the ECG-Image-Aware Network (EIANet), which uses 12-lead ECG images for early prediction of EDCA. By focusing on readily available triage ECG images, this research seeks to create a practical and accessible solution that seamlessly integrates into real-world ED workflows.

Methods: For adult patients with EDCA (cases), 12-lead ECG images at ED triage were obtained from 2 independent data sets: National Taiwan University Hospital (NTUH) and Far Eastern Memorial Hospital (FEMH). Control ECGs were randomly selected from adult ED patients without cardiac arrest during the same study period. In EIANet, ECG images were first converted to binary form, followed by noise reduction, connected component analysis, and morphological opening. A spatial attention module was incorporated into the ResNet50 architecture to enhance feature extraction, and a custom binary recall loss (BRLoss) was used to balance precision and recall, addressing slight data set imbalance. The model was developed and internally validated on the NTUH-ECG data set and was externally validated on an independent FEMH-ECG data set. The model performance was evaluated using the F_1-score, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC).

Results: There were 571 case ECGs and 826 control ECGs in the NTUH data set and 378 case ECGs and 713 control ECGs in the FEMH data set. The novel EIANet model achieved an F_1-score of 0.805, AUROC of 0.896, and AUPRC of 0.842 on the NTUH-ECG data set with a 40% positive sample ratio. It achieved an F_1-score of 0.650, AUROC of 0.803, and AUPRC of 0.678 on the FEMH-ECG data set with a 34.6% positive sample ratio. The feature map showed that the region of interest in the ECG was the ST segment.

XSL•FO
RenderX

Conclusions: EIANet demonstrates promising potential for accurately predicting EDCA using triage ECG images, offering an effective solution for early detection of high-risk cases in emergency settings. This approach may enhance the ability of health care professionals to make timely decisions, with the potential to improve patient outcomes by enabling earlier interventions for EDCA.

(J Med Internet Res 2025;27:e67576) doi:10.2196/67576

KEYWORDS

cardiac arrest; emergency department; deep learning; computer vision; electrocardiogram

Introduction

In-hospital cardiac arrest (IHCA) differs significantly from out-of-hospital cardiac arrest (OHCA), with the former often occurring in patients with significant comorbidities rather than as sudden cardiac events [1-3]. Despite its clinical significance, IHCA has received less research attention. Recent studies using data from the American Heart Association's Get With the Guidelines-Resuscitation registry revealed an IHCA incidence of approximately 10 per 1000 bed-days (290,000 cases annually), with a survival rate of 15% to 20% at hospital discharge [4,5]. Emergency department (ED)–based IHCA (EDCA), accounting for 10% to 20% of all IHCA cases [6], is associated with higher mortality and less well-defined causes than inpatient cardiac arrest (IPCA) [7]. ED crowding, coupled with infrequent monitoring and unstable conditions, further increases EDCA risk [8].

Most patients experiencing IHCA show clinical deterioration hours before arrest, leading to the adoption of rule-based early warning scores to detect high-risk individuals [9-11]. With advancements in information technology, the use of machine learning for predicting IHCA has become increasingly prominent. Recent studies have demonstrated the potential of machine learning and statistical approaches in disease prediction, including heart disease analysis [12]. Although some deep learning models using structured data, such as vital signs [13-15], exist for predicting IPCA, they are not tailored for ED-specific conditions, highlighting the need for a dedicated ED-focused predictive model to address unique challenges in this setting. Using both triage and vital sign time-series data, we recently developed and validated a deep learning–based prediction tool for EDCA [16].

Electrocardiogram (ECG) is an old tool in clinical medicine but has re-emerged for the prediction of low left ventricular ejection fraction [17], arrhythmia [18], dyskalemia [19], or even longer-term mortality [20]. A recent randomized controlled trial using artificial intelligence–enabled ECG to identify hospitalized patients with a high risk of mortality found that the implementation of the artificial intelligence–enabled ECG alert was associated with a significant reduction in all-cause mortality within 90 days [21]. This finding is somewhat surprising, as a recent Cochrane review did not find strong evidence supporting the utility of early warning systems and rapid response systems for the prevention of patient deterioration in acute adult hospital wards [22]. To our knowledge, only 1 study has analyzed ECGs solely from inpatients to predict cardiac arrest within 24 hours of ECG acquisition [23]. This study analyzed signal-based ECGs, which require complex signal processing before analysis. In addition, XML-based protocols, commonly used by several ECG device manufacturers, store structured ECG data and enable integration with systems like General Electric's MUSE server for advanced diagnostics. However, such reliance on proprietary software and infrastructure limits their applicability in the ED, which has diverse device environments. In contrast, image-based 12-lead ECG models leverage universally available PDF reports generated at ED triage. This approach eliminates the need for additional preprocessing or proprietary systems, providing a more accessible, flexible, and scalable solution that seamlessly integrates into real-world clinical workflows.

Accordingly, this study explored an innovative approach by using triage 12-lead ECG images as a more accessible alternative for predicting imminent cardiac arrest in the ED setting. We leveraged ECG reports, commonly available in PDF, and converted them into image-based data inputs to develop a deep learning–based predictive system for IHCA. We also used 2 independent ED data sets for development and validation of our model.

Methods

System Overview

We present a novel cardiac arrest prediction model called the ECG-Image-Aware Network (EIANet) that was designed to predict IHCA with high accuracy. Our study adhered to the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [24]. EIANet operated in 3 key stages, as outlined in Figure 1. In the first stage, ECG waveform images were extracted from triage ECG reports obtained from the National Taiwan University Hospital (NTUH). The NTUH is a tertiary academic medical center with approximately 2400 beds and 100,000 ED visits per year. We obtained 12-lead ECG images at ED triage for adult patients with EDCA (cases) from 2011 to 2019. We defined EDCA as patients who arrived in the ED with vital signs and later developed cardiac arrest in the ED. OHCAs with or without return of spontaneous circulation on ED arrival were excluded. We focused on treated EDCA, so patients with a do-not-resuscitate order were also excluded. The EDCA cases were from the same pool as in our previous studies on EDCA [16,25]. Control ECGs were randomly selected from adult ED patients without cardiac arrest during the same study period.

XSL•FO
RenderX

Figure 1. Overview of the proposed ECG-Image-Aware Network (EIANet) model. SA: spatial attention.

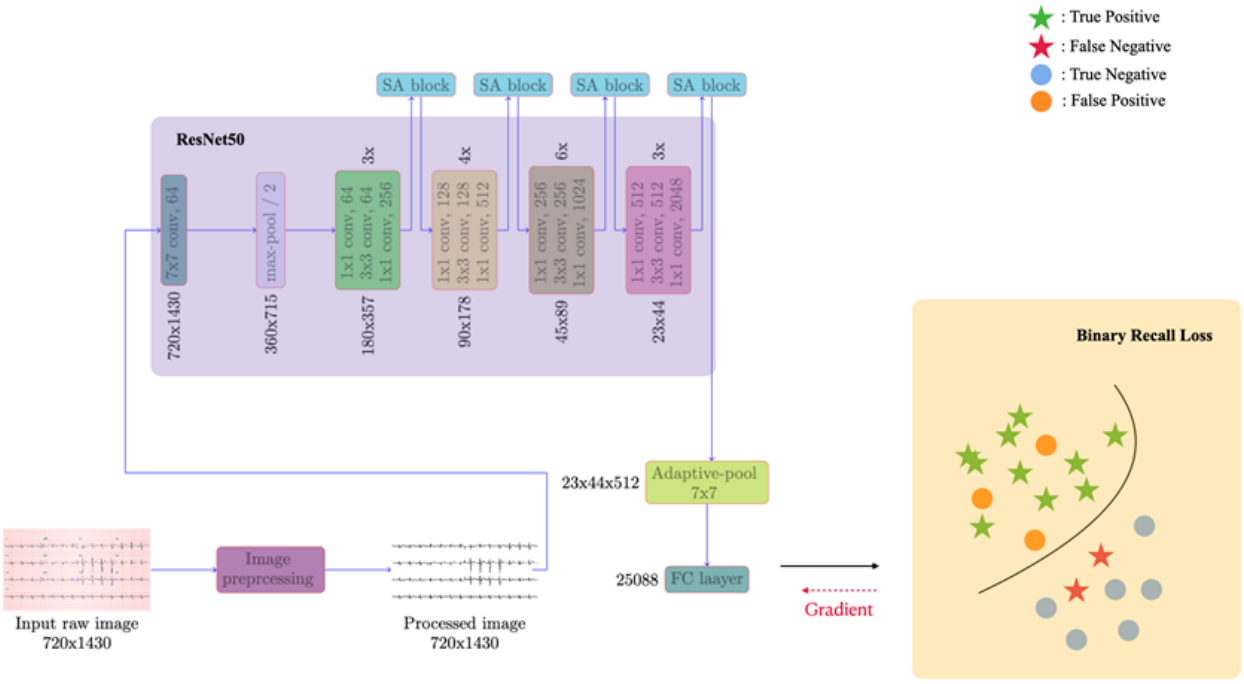

The ECG images were subjected to a thorough preprocessing step, where extraneous information, artifacts, and noise were removed, ensuring that the data fed into the model were clean and of high quality for more accurate analysis. In the second stage, the preprocessed ECG images were passed through a deep learning model, which was enhanced using a spatial attention module. This module allowed the model to focus on more relevant areas of the ECG image, identifying subtle patterns and features critical for predicting cardiac arrest. Finally, in the third stage, the model's output was combined with the recall metric to calculate the binary recall loss (BRLoss). This custom loss function helps address slight class imbalance by prioritizing recall during training, leading to more balanced predictions, particularly in rare IHCA cases.

ECG Data Set for IHCA Prediction

NTUH-ECG Data Set

The NTUH-ECG data set is a comprehensive collection of ECG data sourced from ED patients at NTUH, covering a significant period from 2011 to 2019. During this period, the data set was meticulously curated to include as many positive instances of EDCA as possible, representing critical cases that pose serious risks to patient safety. To balance the data set, random samples of ECGs from patients who did not experience EDCA were collected to serve as negative examples. Distinct from conventional ECG data sets that primarily focus on raw signal data, the NTUH-ECG data set features detailed ECG PDF reports containing 12-lead waveforms (I, II, III, aVR, aVL, aVF, and V1-V6) captured over 10-second intervals at ED triage. The primary objective of this data set is to predict the likelihood of a patient experiencing cardiac arrest during the ED stay, a task of immense clinical significance. The data set comprised a total of 571 cardiac arrest events (positive samples) and 826 noncardiac arrest events (negative samples), resulting in a positive sample ratio of 40.1%. For model training and evaluation, the data were randomly split into training and testing sets, with 80% allocated for training and 20% allocated for testing. This division yielded a test set containing 280 data points, which included 114 positive cases and 166 negative cases. The key characteristics of the NTUH-ECG data set are summarized in Table 1, providing valuable insights into its structure and composition for researchers and clinicians alike. The median time to cardiopulmonary resuscitation was approximately 5 to 7 hours for the cases and controls.

Table 1. Basic characteristics of the training and testing data sets for the National Taiwan University Hospital electrocardiogram (NTUH-ECG) data set.

Variable	NTUH-ECG training data set (n=1117)	NTUH-ECG testing data set (n=280)
Age (years), mean (SD)	65.9 (15.9)	63.7 (17.2)
Male, n (%)	608 (54.4)	141 (50.4)
Annotated positive, n (%)	457 (40.9)	114 (40.7)
Time to CPR[a] (hours), median (IQR)	7.4 (2.5-23.5)	5.1 (1.9-17.9)
Annotated negative, n (%)	660 (59.1)	166 (59.3)

[a]CPR: cardiopulmonary resuscitation.

XSL•FO
RenderX

Far Eastern Memorial Hospital ECG Data Set

The Far Eastern Memorial Hospital (FEMH)-ECG data set consisted of ECG images collected from ED patients presenting to the FEMH from 2016 to 2023. The FEMH is a tertiary academic medical center with approximately 1200 beds and 130,000 ED visits per year. We obtained 12-lead ECG images at ED triage for adult patients with ED cardiac arrest (cases) from 2016 to 2023. Control ECGs were randomly selected from adult ED patients without cardiac arrest during the same study period. Each image was carefully annotated to indicate whether the patient experienced an EDCA during their ED stay. The data format closely mirrors that of the NTUH-ECG data set, with the images being directly extracted from ECG PDF reports. However, the distribution of positive samples in FEMH-ECG was significantly lower than in NTUH-ECG, at 378 positive cases and 713 negative cases, which equates to a positive sample proportion of approximately 34.6%. Table 2 shows the basic characteristics of the FEMH-ECG data set. The median time to cardiopulmonary resuscitation was approximately 2.5 hours.

Table 2. Basic characteristics of the positive and negative samples for the Far Eastern Memorial Hospital electrocardiogram (FEMH-ECG) data set.

Variable	FEMH-ECG positive samples (n=378)	FEMH-ECG negative samples (n=713)
Age (years), mean (SD)	68.3 (14.9)	64.4 (17.9)
Male, n (%)	252 (66.7)	389 (54.6)
Time to CPR[a] (hours), median (IQR)	2.5 (0.7-8.9)	—[b]

[a]CPR: cardiopulmonary resuscitation.

[b]Not applicable.

Ethical Considerations

The NTUH study was approved by the NTUH Institutional Review Board (reference number: 202304129RINC), which waived the requirement for patient informed consent. The FEMH study was approved by the FEMH Institutional Review Board (reference number: 112193-F), which also waived the requirement for patient informed consent.

Preprocessing

Although deep learning models can automatically extract relevant features from input data, their performance can be significantly improved when noisy or unclean data are addressed through careful preprocessing. In our approach, we began by converting the RGB ECG image into a binary format using a thresholding technique. Specifically, we first used Gaussian blur and thresholding to remove unnecessary elements such as thin grid lines. By applying a threshold, pixels exceeding the threshold value were turned black (0), while those below the threshold were converted to white (255). This step ensured that distracting elements were removed, enhancing the clarity of the ECG waveform for further analysis.

Next, we used a connected component algorithm to identify and remove specific unwanted components such as lead labels and other extraneous noise in the image. These elements, if left unaddressed, can interfere with the accuracy of the model's predictions. To further refine the image, we applied a morphological opening operation, which consists of erosion followed by dilation, using a 3 × 3 structuring element. This process smoothed the ECG lines and reduced minor noise, ensuring that the primary features of the ECG waveform were preserved while removing small, irrelevant artifacts. This enhanced the overall quality of the image, making it more suitable for deep learning analysis.

To complete the preprocessing, we applied a Gaussian blur to the image, which helped minimize any remaining noise that may have been left after previous steps. The ECG image was then transformed to gray scale, and we used histogram equalization to improve the contrast between the ECG waveform and any residual background noise. This step ensured that the critical features of the waveform were more distinct, allowing the deep learning model to focus on the most relevant data points. In the final stage of preprocessing, the image was reconverted to binary format, accentuating the essential features necessary for more accurate and effective analysis by the model, ultimately improving its ability to predict IHCA.

Spatial Attention

In our model, we used the ResNet50 [26] architecture as the foundational framework due to its robust performance in handling complex image data. ResNet50's use of residual connections enhances the learning capacity of deep networks by facilitating the flow of information and mitigating issues such as vanishing gradients. We integrated an attention mechanism into the network. This integration ensured that the model can better identify subtle, critical signals in the ECG data, ultimately enhancing its predictive performance and improving patient outcomes.

Following the concept of the Convolutional Block Attention Module [27], we strategically positioned spatial attention blocks after each of the first 4 layers, as Figure 2. This design choice was motivated by 2 primary considerations. First, the nature of ECG waveforms, which display both fine and coarse-grained patterns, necessitates a multiscale approach to feature extraction. By placing spatial attention blocks at the front, middle, and end of the network, we enhanced the model's ability to capture and optimize features across varying scales. This is crucial for accurately interpreting the diverse characteristics present in ECG signals, which can vary significantly in amplitude and duration due to different physiological conditions. Second, we intentionally limited the number of attention blocks to mitigate the risk of overfitting during the training process. By judiciously inserting these blocks after each layer, we balanced the complexity of the model with its generalization capabilities.

XSL•FO
RenderX

Figure 2. We inserted a spatial attention (SA) block after each layer and before the next layer.

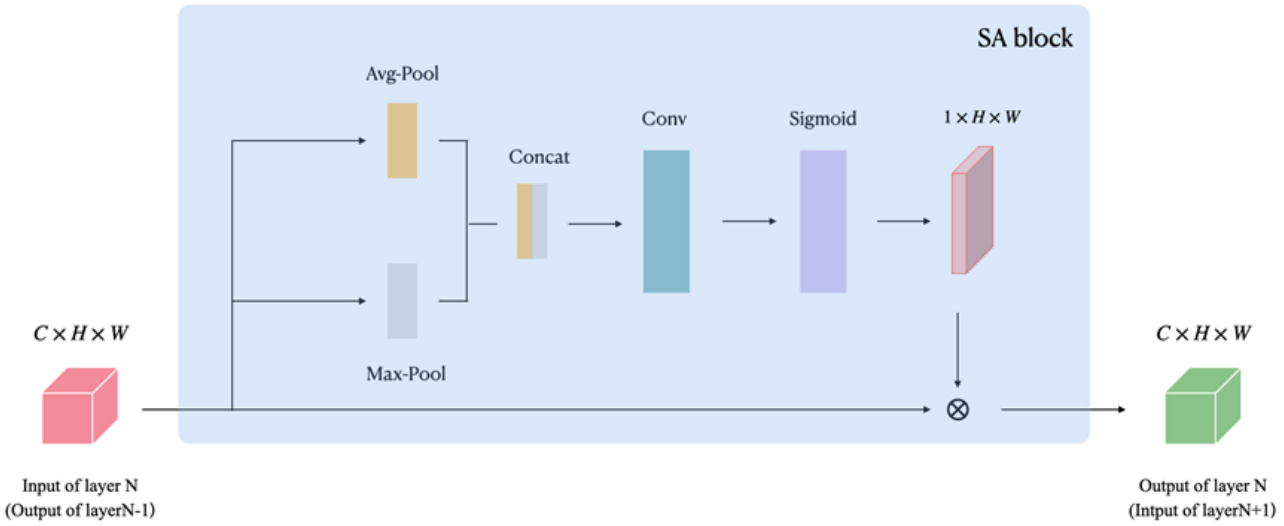

The incorporation of the spatial attention mechanism allowed the model to learn more representative features, leading to improved performance. Visualization results further demonstrated the model's ability to concentrate on more relevant patterns. Detailed findings are discussed in the following section.

Binary Recall Loss

Our training data set exhibited a slight imbalance of positive and negative samples, which posed a challenge for accurately predicting IHCA. Given the serious threat that IHCA presents to patient safety, maximizing the model's recall is of paramount importance, as it measures the system's ability to correctly identify true positive cases. A high recall ensures that most actual cardiac arrest occurrences are detected, allowing health care professionals to respond swiftly and effectively. However, this focus on recall can inadvertently lead to an increase in false positives, overwhelming the health care system with unnecessary alarms and interventions. Such a scenario can strain resources and divert attention from genuine emergencies, potentially leading to alarm fatigue among medical staff. Therefore, it is equally critical to maintain a reasonable level of precision alongside recall. Striking a balance between these 2 metrics is essential to ensure that we effectively identify critical cases of IHCA while also minimizing the burden on health care providers, ultimately safeguarding patient safety and enhancing the efficiency of emergency care.

To address this, rather than using the traditional binary cross-entropy loss, we used BRLoss by drawing on the concept of performance-based loss from [28]. BRLoss used recall-related weights for each class, as shown in Equation 1:



where N is the number of samples in a batch, y_i is the ground truth label of i-th data in the batch, $y_i \in [0, 1]$, p_i is the predicted probability, $0 \leq p_i \leq 1$ and $Recall_c$ means recall of class c, $Recall_c$ [0, 1], c is 0 or 1.

Unlike many other statistics-based weighted loss [29-31] functions that may cause unnecessary false positives, sacrificing precision to gain recall, BRLoss effectively balanced recall and precision. We used the same analytical method as [26] to perform a partial derivative of the loss function. Assuming the $i - th$ data sample in the batch has a final output z_i before applying the sigmoid function, the gradient for BRLoss can be calculated as shown in Equation 2:



In the case where the ground truth is 1, an increase in false negatives leads to a larger negative gradient in the term related to $Recall_1$ in Equation 2. This drives z to increase during gradient descent, reducing false negatives and improving $Recall_1$. Similarly, when the ground truth is 0, more false positives result in a larger positive gradient in the term related to $Recall_0$ (specificity), causing z to decrease and reduce false positives. BRLoss thus balanced recall and precision, leading to more reliable model performance.

Results

We performed experiments using the NTUH-ECG and FEMH-ECG data sets. Given the increasing significance of explainable artificial intelligence, particularly in the medical domain, interpretability had become essential [32-34]. Therefore, we present various qualitative results and conducted ablation studies to clarify model behavior under different conditions.

Implementation Details

We implemented EIANet using the Python and Pytorch [35] framework. The EIANet training was accelerated using an RTX 3090 GPU. The AdamW [36] optimizer was used to train the model over 200 epochs, with a learning rate initialized at 2×10^{-5}.

Evaluation Metrics for IHCA Prediction

In our binary classification task aimed at predicting IHCA using ECG images, we used a comprehensive evaluation strategy to assess model performance. This involved using several metrics derived from the confusion matrix, such as accuracy, precision,

XSL•FO
RenderX

recall (sensitivity), and specificity. These metrics provided a detailed breakdown of how well the model distinguished between positive (IHCA) and negative (non-IHCA) cases, helping us understand not only the overall accuracy but also how well the model balanced false positives and false negatives. We also calculated the area under the receiver operating characteristic curve (AUROC), which measured the trade-off between true positive rates and false positive rates across different decision thresholds. A higher AUROC value indicates that the model is more capable of distinguishing between the 2 classes, even in the presence of class imbalance. Moreover, we incorporated the area under the precision-recall curve (AUPRC) to evaluate performance, particularly in handling imbalanced data sets like ours, where the positive class is underrepresented. The AUPRC focuses on precision (positive predictive value) and recall, offering a more sensitive measure of the model's performance when identifying IHCA cases without being skewed by the larger number of negative cases. Together, these metrics provided a robust, multifaceted assessment of our model's effectiveness at predicting IHCA, ensuring that it performs well across a range of critical evaluation criteria.

Experimental Results With ECG Data Sets

External Comparison

Table 3 presents a comprehensive comparison between our study and the work of Kwon et al [23], which used raw ECG signals to predict IPCA. In addition to differences in patient populations (ED vs inpatients), a key distinction between the 2 studies lies in the format of the input data: Although Kwon et al [23] used raw ECG signals, our approach was built around the use of ECG images. This shift from signal-based input to image-based input enables our model to leverage advanced image processing techniques, such as spatial attention, to enhance feature extraction. In contrast, raw signals can be directly analyzed using the temporal relationships between successive values.

Table 3. Comparison of our study with Kwon et al [23], highlighting differences in input data formats (electrocardiogram images vs raw signals) and data set balance.

Method	Positive rate	Accuracy	Recall	Precision	Specificity	AUROC[a]
Kwon et al [23]	0.0093	0.918	0.778	0.076	0.920	0.948
Ours	0.401	0.843	0.800	0.813	0.884	0.896

[a]AUROC: area under the receiver operating characteristic curve.

Another significant difference is the distribution of positive samples between the 2 data sets. The work by Kwon et al [23] involved a highly imbalanced data set with a larger proportion of negative samples, which is reflected in their higher specificity. In contrast, our data set was more balanced, with a more equitable distribution of positive and negative samples, which had a direct influence on the performance metrics. This difference in the proportion of positive samples played a crucial role in shaping outcomes like precision, accuracy, and AUROC, making direct comparisons between the 2 models less straightforward.

In summary, our approach, with spatial attention, BRLoss, and more balanced sample distribution, can enhance the accurate detection of positive cases, leading to a slightly higher recall. The model by Kwon et al [23], having a predominantly negative sample set, excels in specificity but may sacrifice sensitivity to positive cases. The distinct methodologies and differing data set characteristics highlight the complementary strengths of the 2 approaches, offering insights into how input formats and sample distributions can impact the prediction performance of cardiac arrest models.

Comparison With the Benchmark

To the best of our knowledge, the NTUH-ECG data set stands as the only data set we can get to date that uses image-based ECG data specifically for the prediction of EDCA. Our proposed model, EIANet, represents a pioneering approach in this domain, being the first to tackle this predictive task using ECG images as input. We used the well-established vision model ResNet50 as a baseline for comparative analysis, leveraging its widespread recognition in image-based tasks as a point of reference.

Table 4 shows a side-by-side comparison of the performance of EIANet and ResNet50. Compared with solely using the ResNet50 module, EIANet incorporates spatial attention, enabling the model to focus more on spatially crucial features. With stronger feature representations, combined with BRLoss, the model's ability to differentiate between positive and negative samples is further enhanced. IEIANet significantly outperforms ResNet50 across multiple evaluation metrics, underscoring the efficacy of our tailored approach. We attribute this marked improvement to 2 key innovations in our model design. First, the incorporation of the spatial attention module substantially enhanced the model's feature extraction capabilities. By using spatial attention, EIANet can dynamically focus on the most relevant regions within the ECG images, allowing it to capture critical visual patterns that are often indicative of impending cardiac events. This focused attention provides the model with a more nuanced understanding of the subtle yet vital features present in ECG images. In addition to spatial attention, BRLoss helps to mitigate the risk of too many false positives and false negatives, leading to a slight improvement in recall without affecting precision too much.

XSL•FO
RenderX

Table 4. Comparison of our ECG-Image-Aware Network (EIANet) module with the baseline (ResNet50) module.

Method	Accuracy	Recall	Precision	F_1-score	AUROC[a]	AUPRC[b]
Baseline (ResNet50)	0.786	0.763	0.725	0.744	0.846	0.789
Ours (EIANet)	0.843	0.800	0.813	0.805	0.896	0.842

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

Qualitative Results

Figure 3 and Figure 4 provide feature map [37] visualizations of positive samples from the NTUH-ECG data set. The visualization shows the Grad-CAM output, highlighting regions with Grad-CAM values greater than 0.4. Figure 3 shows the output of a model incorporating spatial attention, while Figure 4 presents the results of a model devoid of spatial attention. Alterations in the ST segment serve as pivotal indicators of myocardial ischemia. Notwithstanding the presence of a bundle branch block in this ECG, EIANet, as depicted in Figure 3, could still accurately pinpoint the ST segment. This underscores EIANet's capacity to discern patterns associated with EDCA within ECG images. Furthermore, tachycardia is a prevalent ECG manifestation preceding EDCA. The regular bright spots in the figures suggest that the algorithm is adept at counting the heart rate. In contrast, although the model without spatial attention, as illustrated in Figure 4, also counted the heart rate, its focus on the ST segment was less precise and more dispersed.

Figure 3. Feature map of a National Taiwan University Hospital electrocardiogram (NTUH-ECG) sample with a spatial attention block.

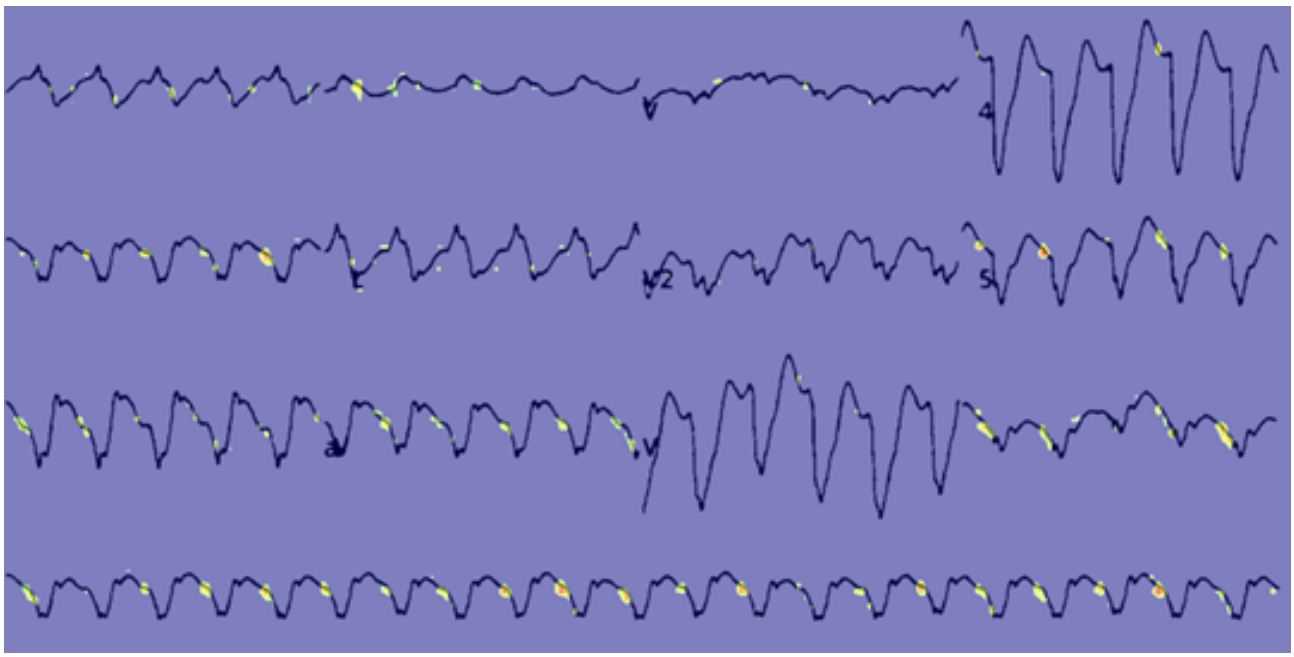

XSL•FO
RenderX

Figure 4. Feature map of a National Taiwan University Hospital electrocardiogram (NTUH-ECG) sample without a spatial attention block.

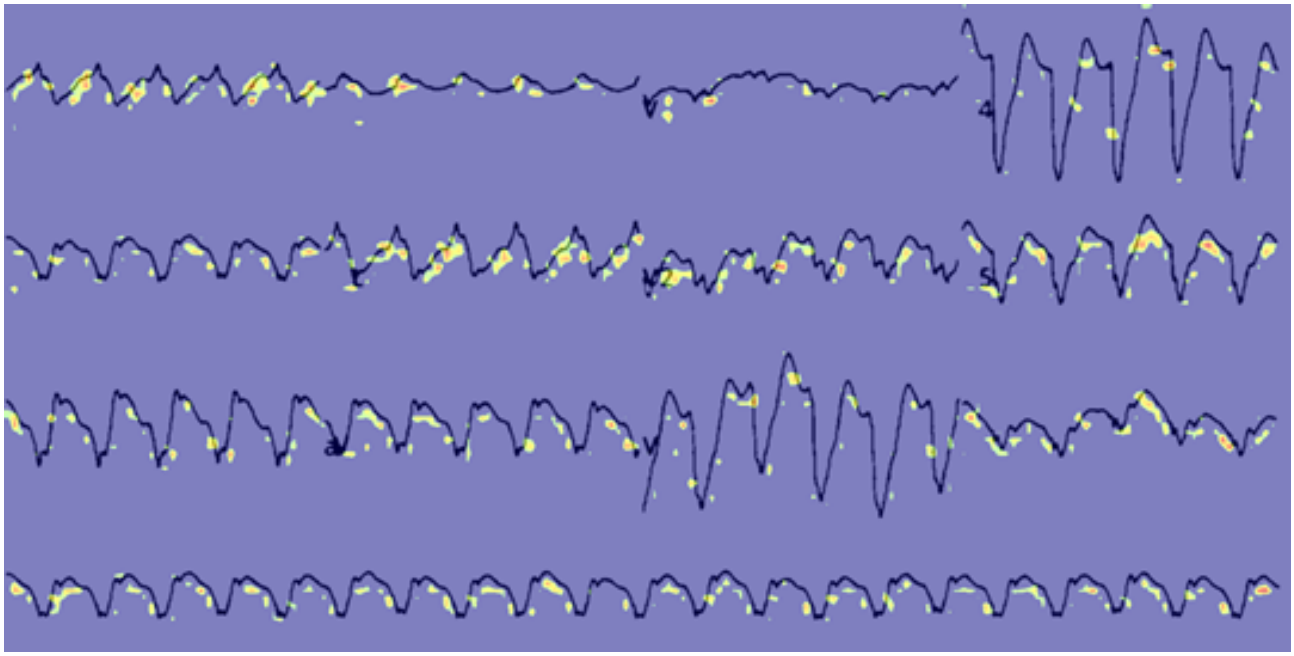

Figure 5 and Figure 6 present visualizations from the FEMH-ECG data set, which further illustrate the impact of spatial attention by focusing on heart rate–related features. In Figure 5 (with spatial attention), the model effectively identifies regular bright spots corresponding to heartbeats, indicating its ability to count the heart rate accurately. Conversely, Figure 6 (without spatial attention) shows a less distinct attention pattern, reducing interpretability.

Figure 5. Feature map of a Far Eastern Memorial Hospital electrocardiogram (FEMH-ECG) sample with a spatial attention block.

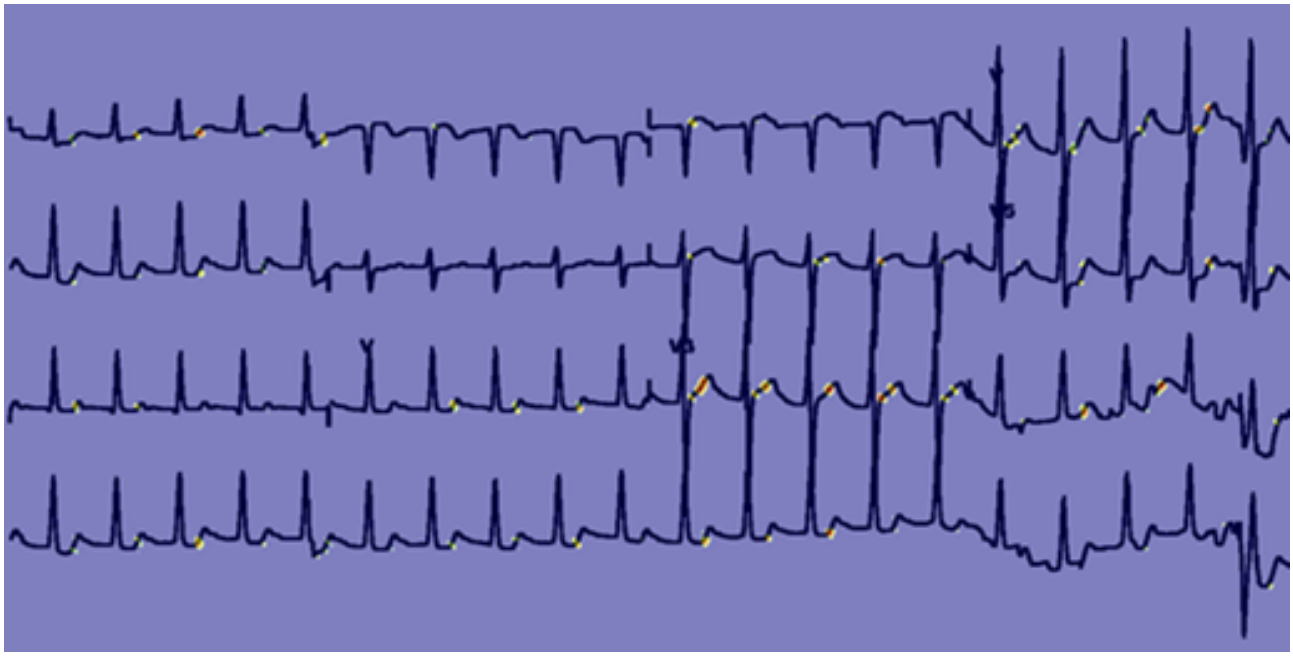

XSL•FO
RenderX

Figure 6. Feature map of a Far Eastern Memorial Hospital electrocardiogram (FEMH-ECG) sample without a spatial attention block.

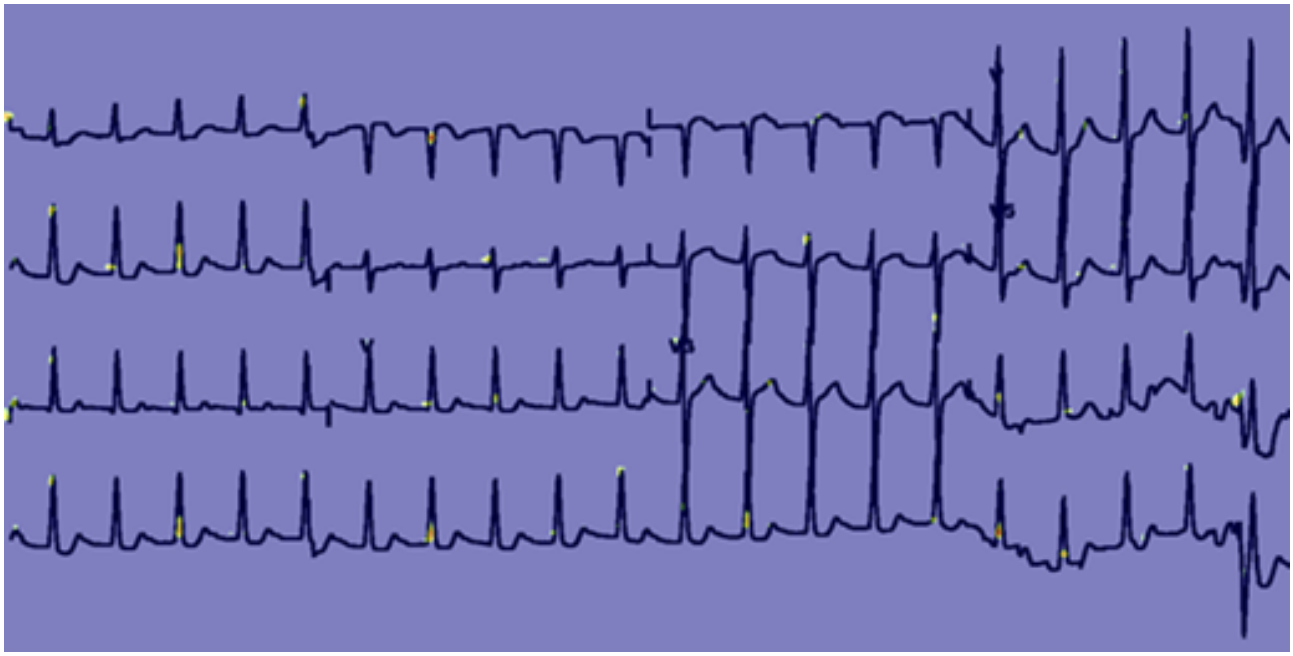

Ablation Study

The first ablation study focused on evaluating the impact of different types of input data on model performance. As detailed in Table 5, using raw ECG images directly extracted from PDF reports for training demonstrated a certain level of predictive capability, indicating that even unprocessed images contain valuable information that the model can learn from. However, the study revealed that, when these raw images undergo a series of image preprocessing steps to remove irrelevant or distracting elements, the model's performance saw marked improvement because the model could focus on crucial visual patterns and structures essential for predicting EDCA. This performance boost underscores the importance of preprocessing for enhancing the quality of the input data.

Table 5. Comparison of raw images and processed images on the National Taiwan University Hospital electrocardiogram (NTUH-ECG) data set.

Input	Accuracy	Recall	Precision	F_1-score	AUROC[a]	AUPRC[b]
Raw image	0.782	0.737	0.718	0.727	0.832	0.785
Processed image	0.843	0.800	0.813	0.805	0.896	0.842

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

By applying preprocessing techniques, such as noise reduction, removal of grid lines, and extraction of key features, the model was better able to focus on the most relevant aspects of the ECG images, allowing it to learn more effectively. Preprocessing not only reduced the amount of noise and irrelevant information present in the images but also helped to clarify the visual patterns and structures that are crucial for predicting EDCA. This led to more accurate predictions and overall model performance.

Table 6 presents an ablation study evaluating the influence of spatial attention and BRLoss on the model's performance and demonstrates the impact of incorporating the spatial attention mechanism and BRLoss on model performance. We trained our model on the NTUH-ECG training set and validated it using both the NTUH-ECG testing set and the FEMH-ECG data set. As expected, due to differences in patient populations, there was a slight decrease in model performance in the FEMH-ECG data set. The results indicate that the spatial attention module and BRLoss elevated the model's performance. The spatial attention mechanism empowered the model to extract more effective ECG features, such as the ST segment and heart rate patterns, thereby enhancing feature representation and overall model accuracy. This mechanism ensured that the model concentrated on critical segments of the ECG image, leading to improved interpretability and reliability. Moreover, BRLoss ameliorated the model's recall without compromising precision to a significant degree. BRLoss focuses on reducing false predictions and improving F_1-score and AUROC. However, its effectiveness also depends on the model's ability to sufficiently represent the data. When combined with spatial attention, the performance of the model can be further enhanced by leveraging the strengths of both modules. Although BRLoss prioritizes recall, decision thresholds can be adjusted in clinical applications to achieve a suitable balance between recall and precision, addressing specific requirements in emergency scenarios.

XSL•FO
RenderX

Table 6. Ablation study results of modules in ECG-Image-Aware Network (EIANet).

Modules	Accuracy	Recall	Precision	F_1-score	AUROC[a]	AUPRC[b]
NTUH-ECG [c] data set						
None	0.742	0.728	0.670	0.700	0.841	0.792
SA[d] only	0.825	0.789	0.783	0.786	0.874	0.825
BRLoss[e] only	0.796	0.781	0.736	0.757	0.869	0.816
SA + BRLoss	0.843	0.800	0.813	0.805	0.896	0.842
FEMH-ECG [f] data set						
None	0.707	0.626	0.571	0.597	0.781	0.621
SA only	0.734	0.645	0.610	0.627	0.814	0.697
BRLoss only	0.726	0.613	0.602	0.607	0.797	0.676
SA + BRLoss	0.743	0.688	0.616	0.650	0.803	0.678

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]NTUH-ECG: National Taiwan University Hospital electrocardiogram.

[d]SA: spatial attention.

[e]BRLoss: binary recall loss.

[f]FEMH-ECG: Far Eastern Memorial Hospital electrocardiogram.

Ablation Study With Varying Class Imbalances

To evaluate the contributions of the spatial attention module and BRLoss under varying class imbalances, we performed random resampling of the NTUH-ECG and FEMH-ECG data sets to adjust the proportion of positive samples to 0.1 and 0.2. Each resampling experiment was repeated 30 times, and the results were averaged to ensure robustness. The ablation study results under positive sample ratios of 0.1 and 0.2 are presented in Table 7. These results clearly demonstrate the contributions of spatial attention and BRLoss to the model's performance. In addition, for metrics that are not dependent on the prevalence of positive cases (eg, AUROC), the model performance remained.

XSL•FO
RenderX

Table 7. Ablation study results of modules in ECG-Image-Aware Network (EIANet).

Modules	Accuracy	Recall	Precision	F_1-score	AUROC[a]	AUPRC[b]
NTUH-ECG [c] data set: positive ratio of 0.1						
None	0.749	0.714	0.238	0.357	0.837	0.436
SA[d] only	0.843	0.783	0.359	0.492	0.871	0.491
BRLoss[e] only	0.805	0.780	0.304	0.438	0.872	0.495
SA + BRLoss	0.865	0.788	0.402	0.532	0.894	0.493
FEMH-ECG [f] data set: positive ratio of 0.1						
None	0.737	0.621	0.215	0.319	0.787	0.265
SA only	0.767	0.637	0.243	0.351	0.811	0.368
BRLoss only	0.769	0.605	0.237	0.340	0.794	0.335
SA + BRLoss	0.764	0.685	0.250	0.366	0.803	0.339
NTUH-ECG data set: positive ratio of 0.2						
None	0.747	0.728	0.421	0.533	0.842	0.612
SA only	0.840	0.800	0.567	0.664	0.875	0.660
BRLoss only	0.802	0.780	0.499	0.609	0.869	0.646
SA + BRLoss	0.858	0.798	0.608	0.690	0.895	0.674
FEMH-ECG data set: positive ratio of 0.2						
None	0.726	0.629	0.386	0.319	0.780	0.444
SA only	0.754	0.647	0.425	0.351	0.814	0.644
BRLoss only	0.750	0.608	0.414	0.340	0.797	0.515
SA + BRLoss	0.755	0.686	0.429	0.366	0.803	0.515

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]NTUH-ECG: National Taiwan University Hospital electrocardiogram.

[d]SA: spatial attention.

[e]BRLoss: binary recall loss.

[f]FEMH-ECG: Far Eastern Memorial Hospital electrocardiogram.

Discussion

Principal Findings

To our knowledge, this is the first study to develop and validate a deep learning algorithm (EIANet) specifically for predicting EDCA using ECG image data. The findings demonstrate that deep learning, a powerful artificial intelligence tool, can detect subtle ECG changes hours before cardiac arrest in the ED. Deployment of this tool at ED triage has the potential to gain lead time to identify high-risk patients for timely interventions and reduce EDCAs.

Comparison With Prior Work

Traditionally, ECG prediction of cardiac arrest has mainly focused on OHCA (ie, using ECG to predict sudden cardiac death at the population level) [38]. For example, an electrical risk score that comprises 6 features (heart rate, prolonged corrected QT interval, Tpeak-Tend interval, QRS-T angle, left ventricular hypertrophy, and delayed QRS transition zone) can accurately predict sudden cardiac death within a community-based cohort [39]. A refined deep learning–based ECG model can further improve its predictive ability using the same 6 features [40]. Notably, tachycardia has been included as one of the features in the electrical risk score and was also documented in our feature map analysis and in a previous IPCA ECG study [23], suggesting the importance of this feature in predicting OHCA, IPCA, and EDCA.

One of our previous studies on EDCA focused on the development and validation of a simple 8-item, rule-based prediction tool using structural data at ED triage [25]. Another study of ours further advanced EDCA prediction by using deep learning with static triage data and dynamic vital signs measured during the ED stay [16]. None of these studies used ECG as an input for deep learning, as we did in this study. Given the promising results of using triage ECG to predict EDCA, the next logical step would be to combine ECG and structural data as a multimodal model for EDCA prediction.

Strengths

The EIANet has several technical innovations. Before model training, we used image processing techniques to eliminate

XSL•FO
RenderX

manual labels and noise from the ECG images. To enhance model performance, we integrated an attention mechanism and BRLoss. These components had a positive impact on EIANet's predictive capabilities. Our research used 2 ECG image data sets: NTUH-ECG and FEMH-ECG. The NTUH-ECG data set, with a positive sample ratio of 40%, yielded impressive results, with EIANet achieving an F_1-score of 0.805 and AUPRC of 0.842. The FEMH-ECG data set, characterized by a lower positive sample ratio of 34.6%, still demonstrated strong performance, with EIANet attaining an F_1-score of 0.650 and AUPRC of 0.678. To validate the effectiveness of the proposed modules, we conducted ablation studies, further confirming their importance. Additionally, Grad-CAM visualization was used to provide interpretability for EIANet's predictions using ECG images.

In the past, ECG prediction models were typically designed for diseases with relatively obvious features, such as myocardial infarction or hyperkalemia [41,42]. However, precardiac arrest ECGs lack absolutely distinct characteristics, making EDCA significantly more challenging to predict. Additionally, most existing image-based ECG models have remained in the proof-of-concept stage, often relying on general purpose framework models like ResNet without substantial customization for this specific task. These approaches, although promising, have struggled with critical limitations, such as high false-negative rates and limited interpretability. To address these challenges, we enhanced ResNet50 by incorporating a spatial attention mechanism, enabling the model to focus on subtle but clinically significant ECG patterns. Furthermore, we introduced a BRLoss function to prioritize recall and reduce false negatives, a critical improvement for predicting rare but life-threatening events like EDCA. These innovations aimed to overcome the inherent limitations of prior approaches and establish a practical, interpretable, and effective prediction framework.

Limitations and Future Directions

Although our results demonstrate a certain level of predictive capability on both the NTUH-ECG and FEMH-ECG data sets, the performance on the FEMH-ECG data set declined. The observed drop in recall when evaluating the EIANet model on the FEMH-ECG data set can be attributed to multiple factors. First, patient population differences between the data sets likely played a significant role. Variations in patient populations could influence the manifestations of ECGs, leading to differing model performance. Second, differences in instrumentation between the 2 data sets may have introduced some biases. These discrepancies, combined with slight differences in image sizes formats and preprocessing process, might have impacted the model's ability to generalize effectively. Addressing these challenges in future work by incorporating more diverse data sets could enhance the model's robustness and applicability across different clinical settings. In the future, we hope to use EIANet as an encoder for ECG data, integrating it with other modalities of patient information to establish a more comprehensive multimodal prediction method for EDCA.

This study has some potential limitations. First, the model's performance was dependent on the quality and variety of the training data. Although we included ECGs from 2 independent medical centers, further external validation studies are needed to evaluate the model's robustness in different patient populations. Second, this was a retrospective study. Even with external validation, as in our retrospective study, models could have poor performance in the prospective setting, such as the Rothman Index by PeraHealth [43,44] and the Epic Sepsis Model [45,46]. Future implementation of the EIANet model in the ED would be warranted to test its real-world effectiveness.

Conclusions

In conclusion, we presented a comprehensive study aimed at understanding the relationship between ECG images and EDCA. To the best of our knowledge, this is the first in the literature that uses readily available 12-lead ECG images at ED triage to predict imminent ED cardiac arrest. Using 2 independent ED data sets, we developed and externally validated the novel deep learning EIANet model that can be used to identify high-risk patients and potentially reduce devastating ED cardiac arrest events. Future implementation of the EIANet in real time is warranted to test whether this tool could flag high-risk patients at ED triage for timely interventions, thereby improving patient outcomes.

Acknowledgments

We thank the staff of the National Taiwan University Hospital-Integrative Medical Data Center (NTUH-iMDC) for providing clinical data for this study.

This research was supported by the Ministry of Science and Technology of Taiwan and Center for Artificial Intelligence & Advanced Robotics, National Taiwan University, under the grant numbers MOST 110-2634-F-002-049 and MOST 110-2221-E-002-166-MY3. This project was also supported by grants from the National Health Research Institutes (NHRI-EX114-11332PI), National Science and Technology Council (NSTC 112-2314-B-002-264), and National Taiwan University Hospital (112-UN-0027 and 113-UN0017).

Generative artificial intelligence was not used in any portion of the manuscript generation.

Data Availability

The artificial intelligence algorithm cannot be made publicly available because it is proprietary intellectual property. The computer code and data for this study are available from the corresponding authors upon reasonable request for research studies.

XSL•FO
RenderX

Authors' Contributions

All authors contributed to the study conception and design. Data collection was performed by JWG and JTS. Data preprocessing and analysis were performed by SCL and GYC. The first draft of the manuscript was written by SCL, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. CLT and LCF are the corresponding authors and supervised the project.

Conflicts of Interest

None declared.

References

1. Morrison LJ, Neumar RW, Zimmerman JL, Link MS, Newby LK, McMullan PW, et al. Strategies for Improving Survival After In-Hospital Cardiac Arrest in the United States: 2013 Consensus Recommendations. Circulation 2013 Apr 09;127(14):1538-1563. [doi: 10.1161/cir.0b013e31828b2770]
2. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-hospital cardiac arrest: a review. JAMA 2019 Mar 26;321(12):1200-1210 [FREE Full text] [doi: 10.1001/jama.2019.1696] [Medline: 30912843]
3. Nolan JP, Berg RA, Andersen LW, Bhanji F, Chan PS, Donnino MW, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the Utstein Resuscitation Registry Template for In-Hospital Cardiac Arrest: a consensus report from a task force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian and New Zealand Council on Resuscitation, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa, Resuscitation Council of Asia). Circulation 2019 Oct 29;140(18):1. [doi: 10.1161/cir.0000000000000710]
4. Merchant RM, Yang L, Becker LB, Berg RA, Nadkarni V, Nichol G, et al. Incidence of treated cardiac arrest in hospitalized patients in the United States*. Critical Care Medicine 2011;39(11):2401-2406. [doi: 10.1097/ccm.0b013e3182257459]
5. Girotra S, Nallamothu BK, Spertus JA, Li Y, Krumholz HM, Chan PS. Trends in survival after in-hospital cardiac arrest. N Engl J Med 2012 Nov 15;367(20):1912-1920. [doi: 10.1056/nejmoa1109148]
6. Kayser RG, Ornato JP, Peberdy MA, American Heart Association National Registry of Cardiopulmonary Resuscitation. Cardiac arrest in the emergency department: a report from the National Registry of Cardiopulmonary Resuscitation. Resuscitation 2008 Aug;78(2):151-160. [doi: 10.1016/j.resuscitation.2008.03.007] [Medline: 18508184]
7. Mir T, Qureshi WT, Uddin M, Soubani A, Saydain G, Rab T, et al. Predictors and outcomes of cardiac arrest in the emergency department and in-patient settings in the United States (2016-2018). Resuscitation 2022 Jan;170:100-106. [doi: 10.1016/j.resuscitation.2021.11.009] [Medline: 34801637]
8. Camargo CA, Tsai C, Sullivan AF, Cleary PD, Gordon JA, Guadagnoli E, et al. Safety climate and medical errors in 62 US emergency departments. Ann Emerg Med 2012 Nov;60(5):555-563.e20. [doi: 10.1016/j.annemergmed.2012.02.018] [Medline: 23089089]
9. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. Chest 2013 Jun;143(6):1758-1765 [FREE Full text] [doi: 10.1378/chest.12-1605] [Medline: 23732586]
10. Churpek MM, Yuen TC, Huber MT, Park SY, Hall JB, Edelson DP. Predicting cardiac arrest on the wards: a nested case-control study. Chest 2012 May;141(5):1170-1176 [FREE Full text] [doi: 10.1378/chest.11-1301] [Medline: 22052772]
11. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
12. Rasool A, Tao R, Kashif K, Khan W, Agbedanu P, Choudhry N. Statistic solution for machine learning to analyze heart disease data. 2020 Presented at: 12th International Conference on Machine Learning and Computi; February 15-17, 2020; Shenzhen, China. [doi: 10.1145/3383972.3384061]
13. Cho K, Kwon O, Kwon J, Lee Y, Park H, Jeon KH, et al. Detecting patient deterioration using artificial intelligence in a rapid response system. Crit Care Med 2020 Apr;48(4):e285-e289. [doi: 10.1097/CCM.0000000000004236] [Medline: 32205618]
14. Kwon J, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in‐hospital cardiac arrest. JAHA 2018 Jul 03;7(13):1. [doi: 10.1161/jaha.118.008678]
15. Lee YJ, Cho K, Kwon O, Park H, Lee Y, Kwon J, et al. A multicentre validation study of the deep learning-based early warning score for predicting in-hospital cardiac arrest in patients admitted to general wards. Resuscitation 2021 Apr 22;163:78-85 [FREE Full text] [doi: 10.1016/j.resuscitation.2021.04.013] [Medline: 33895236]
16. Deng Y, Wang J, Ko C, Huang C, Tsai C, Fu L. Deep learning-based Emergency Department In-hospital Cardiac Arrest Score (Deep EDICAS) for early prediction of cardiac arrest and cardiopulmonary resuscitation in the emergency department. BioData Min 2024 Nov 23;17(1):52 [FREE Full text] [doi: 10.1186/s13040-024-00407-8] [Medline: 39580434]
17. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med 2019 Jan 7;25(1):70-74. [doi: 10.1038/s41591-018-0240-2] [Medline: 30617318]

XSL•FO
RenderX

18. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, Nemani A, Carbonati T, Jing L, et al. Deep neural networks can predict new-onset atrial fibrillation from the 12-lead ECG and help identify those at risk of atrial fibrillation–related stroke. Circulation 2021 Mar 30;143(13):1287-1298. [doi: 10.1161/circulationaha.120.047829]
19. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. JAMA Cardiol 2019 May 01;4(5):428-436 [FREE Full text] [doi: 10.1001/jamacardio.2019.0640] [Medline: 30942845]
20. Popescu DM, Shade JK, Lai C, Aronis KN, Ouyang D, Moorthy MV, et al. Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. Nat Cardiovasc Res 2022 Apr;1(4):334-343 [FREE Full text] [doi: 10.1038/s44161-022-00041-9] [Medline: 35464150]
21. Lin C, Liu W, Tsai D, Lou Y, Chang C, Lee C, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med 2024 Apr 29;30(5):1461-1470. [doi: 10.1038/s41591-024-02961-4]
22. McGaughey J, Fergusson D, Van Bogaert P, Rose L. Early warning systems and rapid response systems for the prevention of patient deterioration on acute adult hospital wards. Cochrane Database Syst Rev 2021 Nov 22;11(11):CD005529 [FREE Full text] [doi: 10.1002/14651858.CD005529.pub3] [Medline: 34808700]
23. Kwon J, Kim K, Jeon K, Lee SY, Park J, Oh B. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. Scand J Trauma Resusc Emerg Med 2020 Oct 06;28(1):98 [FREE Full text] [doi: 10.1186/s13049-020-00791-0] [Medline: 33023615]
24. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
25. Tsai C, Lu T, Fang C, Wang C, Lin J, Chen W, et al. Development and validation of a novel triage tool for predicting cardiac arrest in the emergency department. West J Emerg Med 2022 Feb 23;23(2):258-267 [FREE Full text] [doi: 10.5811/westjem.2021.8.53063] [Medline: 35302462]
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.90]
27. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. In: Ferrari V, Sminchisescu C, Hebert M, Weiss Y, editors. Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Cham, Switzerland: Springer; 2018:3-19.
28. Tian J, Mithun NC, Seymour Z, Chiu HP, Kira Z. Striking the right balance: recall loss for semantic segmentation. 2022 Presented at: International Conference on Robotics and Automation (ICRA); May 23-27, 2022; Philadelphia, PA. [doi: 10.1109/icra46639.2022.9811702]
29. Huang C, Li Y, Loy CC, Tang X. Learning deep representation for imbalanced classification. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.580]
30. Cui Y, Jia M, Lin TY, Song Y, Belongie S. Class-balanced loss based on effective number of samples. 2019 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 15-20, 2019; Long Beach, CA. [doi: 10.1109/cvpr.2019.00949]
31. Wang YX, Ramanan D, Hebert M. Learning to model the tail. In: In Neural Information Processing Systems. 2017 Presented at: 31st Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA.
32. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans. Neural Netw. Learning Syst 2021 Nov;32(11):4793-4813. [doi: 10.1109/tnnls.2020.3027314]
33. Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. J Imaging 2020 Jun 20;6(6):52 [FREE Full text] [doi: 10.3390/jimaging6060052] [Medline: 34460598]
34. Gu R, Wang G, Song T, Huang R, Aertsen M, Deprest J, et al. CA-Net: comprehensive attention convolutional neural networks for explainable medical image segmentation. IEEE Trans. Med. Imaging 2021 Feb;40(2):699-711. [doi: 10.1109/tmi.2020.3035253]
35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. 2019 Presented at: Advances in Neural Information Processing Systems 32 (NeurIPS 2019); December 8-14, 2019; Vancouver, Canada. [doi: 10.7551/mitpress/11474.003.0014]
36. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint (January 4, 2019).
37. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep net- works via gradient-based localization. 2017 Presented at: IEEE International Conference on Computer Vision (ICCV); October 22-29, 2017; Venice, Italy. [doi: 10.1109/iccv.2017.74]
38. Kolk MZ, Ruipérez-Campillo S, Wilde AA, Knops RE, Narayan SM, Tjong FV. Prediction of sudden cardiac death using artificial intelligence: current status and future directions. Heart Rhythm 2024 Sep 06:1 [FREE Full text] [doi: 10.1016/j.hrthm.2024.09.003] [Medline: 39245250]
39. Aro A, Reinier K, Rusinaru C, Uy-Evanado A, Darouian N, Phan D, et al. Electrical risk score beyond the left ventricular ejection fraction: prediction of sudden cardiac death in the Oregon Sudden Unexpected Death Study and the Atherosclerosis

XSL•FO
RenderX

Risk in Communities Study. Eur Heart J 2017 Oct 21;38(40):3017-3025 [FREE Full text] [doi: 10.1093/eurheartj/ehx331] [Medline: 28662567]
40. Holmstrom L, Chugh H, Nakamura K, Bhanji Z, Seifer M, Uy-Evanado A, et al. An ECG-based artificial intelligence model for assessment of sudden cardiac death risk. Commun Med (Lond) 2024 Feb 27;4(1):17 [FREE Full text] [doi: 10.1038/s43856-024-00451-9] [Medline: 38413711]
41. Makimoto H, Höckmann M, Lin T, Glöckner D, Gerguri S, Clasen L, et al. Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction. Sci Rep 2020 May 21;10(1):8445 [FREE Full text] [doi: 10.1038/s41598-020-65105-x] [Medline: 32439873]
42. Kim D, Jeong J, Kim J, Cho Y, Park I, Lee S, et al. Hyperkalemia detection in emergency departments using initial ECGs: a smartphone AI ECG analyzer vs. board-certified physicians. J Korean Med Sci 2023 Nov 20;38(45):e322 [FREE Full text] [doi: 10.3346/jkms.2023.38.e322] [Medline: 37987103]
43. Arnold J, Davis A, Fischhoff B, Yecies E, Grace J, Klobuka A, et al. Comparing the predictive ability of a commercial artificial intelligence early warning system with physician judgement for clinical deterioration in hospitalised general internal medicine patients: a prospective observational study. BMJ Open 2019 Oct 10;9(10):e032187 [FREE Full text] [doi: 10.1136/bmjopen-2019-032187] [Medline: 31601602]
44. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. J Biomed Inform 2013 Oct;46(5):837-848 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.011] [Medline: 23831554]
45. Habib AR, Lin AL, Grant RW. The Epic Sepsis Model falls short-the importance of external validation. JAMA Intern Med 2021 Aug 01;181(8):1040-1041. [doi: 10.1001/jamainternmed.2021.3333] [Medline: 34152360]
46. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 2021 Aug 01;181(8):1065-1070 [FREE Full text] [doi: 10.1001/jamainternmed.2021.2626] [Medline: 34152373]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
BRLoss: binary recall loss
ECG: electrocardiogram
ED: emergency department
EDCA: emergency department–based in-hospital cardiac arrest
EIANet: ECG-Image-Aware Network
FEMH: Far Eastern Memorial Hospital
IHCA: in-hospital cardiac arrest
IPCA: inpatient cardiac arrest
NTUH: National Taiwan University Hospital
OHCA: out-of-hospital cardiac arrest

Edited by A Coristine; submitted 17.10.24; peer-reviewed by D Yoo, A Rasool; comments to author 29.11.24; revised version received 03.01.25; accepted 04.02.25; published 28.02.25.

Please cite as:
Lu SC, Chen GY, Liu AS, Sun JT, Gao JW, Huang CH, Tsai CL, Fu LC
Deep Learning–Based Electrocardiogram Model (EIANet) to Predict Emergency Department Cardiac Arrest: Development and External Validation Study
J Med Internet Res 2025;27:e67576
URL: https://www.jmir.org/2025/1/e67576
doi:10.2196/67576
PMID:40053733

©Shao-Chi Lu, Guang-Yuan Chen, An-Sheng Liu, Jen-Tang Sun, Jun-Wan Gao, Chien-Hua Huang, Chu-Lin Tsai, Li-Chen Fu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Large Language Models' Accuracy in Emulating Human Experts' Evaluation of Public Sentiments about Heated Tobacco Products on Social Media: Evaluation Study

Kwanho Kim[1], PhD; Soojong Kim[2], PhD

[1]Department of Media, College of Politics and Economics, Kyung Hee University, Seoul, Republic of Korea

[2]Department of Communication, University of California Davis, Davis, CA, United States

Corresponding Author:
Soojong Kim, PhD
Department of Communication
University of California Davis
1 Shields Ave, Kerr Hall #361
Davis, CA, 95616
United States
Phone: 1 530 752 0966
Email: sjokim@ucdavis.edu

Abstract

Background: Sentiment analysis of alternative tobacco products discussed on social media is crucial in tobacco control research. Large language models (LLMs) are artificial intelligence models that were trained on extensive text data to emulate the linguistic patterns of humans. LLMs may hold the potential to streamline the time-consuming and labor-intensive process of human sentiment analysis.

Objective: This study aimed to examine the accuracy of LLMs in replicating human sentiment evaluation of social media messages relevant to heated tobacco products (HTPs).

Methods: GPT-3.5 and GPT-4 Turbo (OpenAI) were used to classify 500 Facebook (Meta Platforms) and 500 Twitter (subsequently rebranded X) messages. Each set consisted of 200 human-labeled anti-HTPs, 200 pro-HTPs, and 100 neutral messages. The models evaluated each message up to 20 times to generate multiple response instances reporting its classification decisions. The majority of the labels from these responses were assigned as a model's decision for the message. The models' classification decisions were then compared with those of human evaluators.

Results: GPT-3.5 accurately replicated human sentiment evaluation in 61.2% of Facebook messages and 57% of Twitter messages. GPT-4 Turbo demonstrated higher accuracies overall, with 81.7% for Facebook messages and 77% for Twitter messages. GPT-4 Turbo's accuracy with 3 response instances reached 99% of the accuracy achieved with 20 response instances. GPT-4 Turbo's accuracy was higher for human-labeled anti- and pro-HTP messages compared with neutral messages. Most of the GPT-3.5 misclassifications occurred when anti- or pro-HTP messages were incorrectly classified as neutral or irrelevant by the model, whereas GPT-4 Turbo showed improvements across all sentiment categories and reduced misclassifications, especially in incorrectly categorized messages as irrelevant.

Conclusions: LLMs can be used to analyze sentiment in social media messages about HTPs. Results from GPT-4 Turbo suggest that accuracy can reach approximately 80% compared with the results of human experts, even with a small number of labeling decisions generated by the model. A potential risk of using LLMs is the misrepresentation of the overall sentiment due to the differences in accuracy across sentiment categories. Although this issue could be reduced with the newer language model, future efforts should explore the mechanisms underlying the discrepancies and how to address them systematically.

(J Med Internet Res 2025;27:e63631) doi:10.2196/63631

KEYWORDS

heated tobacco products; artificial intelligence; large language models; social media; sentiment analysis; ChatGPT; generative pre-trained transformer; GPT; LLM; NLP; natural language processing; machine learning; language model; sentiment; evaluation; social media; tobacco; alternative; prevention; nicotine; OpenAI

XSL•FO
RenderX

Introduction

Heated tobacco products (HTPs) are emerging tobacco products that heat processed tobacco leaves, enabling users to breathe nicotine into their lungs [1]. As these products gain global market share at a rapid pace, their potential impacts on tobacco prevention and cessation initiatives are becoming an important topic of public debate [2].

Social media platforms are where a wide range of stakeholders of tobacco regulations distribute their messages, such as policy announcements, product advertisements, and product user feedback [3-5]. Analyses of social media discourses on HTPs provide opportunities to observe and identify the dynamics of these messages, which could affect the public's perception of these products and relevant regulatory issues [3-7].

Sentiment analysis is a widely adopted method to understand the attitudes of the public toward tobacco-related issues by evaluating social media messages [8-11]. Prior research has specifically focused on positive and negative sentiments due to their possible associations with tobacco cessation and prevention outcomes, such as the use of tobacco products and support for tobacco regulations [11-13].

In past sentiment analyses of large-scale social media content data, human evaluators often examined a subset of the dataset rather than analyzing the entire dataset. The subset was then used as a representative sample to inform the sentiment of the whole dataset [8,9] or as a reference for machine learning classifiers tasked with analyzing the entire dataset [14]. This approach stems from the time-consuming and labor-intensive nature of human sentiment evaluation, which involves recruiting, training, and coordinating multiple evaluators. This complexity arises from the fact that latent coding, including sentiment analysis, requires understanding the underlying meanings and subtleties in the text, which can substantially vary depending on the context and across coders [15].

Large language models (LLMs), such as OpenAI's Generative Pre-trained Transformer and Google's Gemini, may be able to alleviate the burdens of human sentiment evaluation. LLMs are artificial intelligence (AI) models that were trained on extensive text data to emulate the linguistic patterns of humans [16,17]. Recent LLMs are known to achieve precision at the level of human decisions on several intellectual tasks [18]. As LLMs become increasingly accessible and available, expectations are growing about the feasibility of using these technologies in public health and social science research [19]. Several examples include analyzing health and medical information [20,21], pretesting the effect of health campaign messages [22], predicting psychological experimental results [23], and simulating sociodemographic groups and their reactions to social issues [24].

We investigate the accuracy of LLMs in analyzing sentiment in social media messages about HTPs. The current research focuses on OpenAI's GPT, given its high accessibility, availability, and popularity. GPT models are easily accessible through chatbot services such as ChatGPT, Microsoft Copilot, and Apple Intelligence and are estimated to have the largest user base worldwide. For instance, ChatGPT has more than 200 million weekly active users as of August 2024 [25]. These aspects contribute to the attractiveness of GPT as an analytic tool for tobacco researchers, especially those with limited budgets and resources.

This study examines the accuracy of GPT-3.5 and GPT-4 Turbo in emulating human sentiment evaluations of social media messages related to HTPs. GPT-3.5 is a milestone model that powered ChatGPT when the service was launched [26]; GPT-4 Turbo is one of the most recent GPT-4 models as of 2024, with a particular development focus on improvements in processing text prompts [27]. This study conducted direct comparisons of the sentiment evaluations made by human coders and these language models based on social media messages gathered from multiple platforms. This investigation could ultimately contribute to assessing the ability of LLMs in examining how the public views alternative tobacco products.

Methods

Data Collection

Messages relevant to HTPs were collected from 2 social media platforms that provide distinct message formats: Facebook (long format) and Twitter (short format). Facebook posts were collected using CrowdTangle (CT)'s keyword search feature. CT was a social media analytic tool provided by Facebook's parent company, Meta. It allowed researchers to access the historical data of Facebook [28]. Tweets were gathered using Twitter's application programming interface 2.0, which could access the historical Twitter dataset through the company's academic research access program [29]. This study focused on messages written in English.

In April 2022, a keyword search was conducted using the following search query: "heat not burn" OR "heat-not-burn" OR "heated tobacco" OR "tobacco heating" OR ((htp OR hnb) AND (smoking OR smoke OR vaping OR vape OR tobacco OR cig OR nicotine)) OR iqos. This query was designed to find tweets and Facebook posts meeting at least one of the following conditions: (1) Containing "heat not burn" in its entirety; (2) containing "heat-not-burn" in its entirety; (3) containing "heated tobacco" in its entirety; (4) containing "tobacco heating" in its entirety; (5) containing at least one of "htp" and "hnb," only when it also contains one of "smoking," "smoke," "vaping," "vape," "tobacco," "cig," and "nicotine;" and (6) containing "iqos." This search yielded 16,284 Facebook posts that were published between January 2014 and December 2021 and 60,031 tweets published in the same period.

Human Evaluation

The procedures for preparing samples for human sentiment evaluations were adapted from sentiment analyses of tobacco-related mass and social media discussions [8-11]. A team of 3 human coders evaluated the sentiment of 1250 Facebook posts and 1200 tweets sampled from the entire pool of keyword-searched social media messages (ie, 16,284 Facebook posts and 60,031 tweets). Those messages were human-labeled as one of the following 5 categories: ANTI (anti-HTP messages), PRO (pro-HTP messages), NEU (neutral

XSL•FO
RenderX

messages), MIX (messages containing a mixture of positive and negative attitudes on HTPs), and IR (messages irrelevant to HTPs).

The messages were sampled through a multistep process designed to increase the likelihood of including both potentially negative and positive messages, ensuring their inclusion in the selected messages for human annotation. The details of the sampling and coding procedures are reported in Multimedia Appendix 1.

GPT Evaluation

From each of the human-evaluated 1250 long-form and 1200 short-form messages, we randomly selected 200 PRO, 200 ANTI, and 100 NEU messages, totaling 1000 messages. All these selected messages for GPT-3.5 and GPT-4 Turbo sentiment classification were relevant to HTPs. A LLM prompt was created for each message, including coding instructions, the message, and the coding scheme. Since this study aims to conduct direct comparisons between the sentiment evaluations of human coders and the language models, the instructions and the coding scheme for the language models were kept consistent with those for human coders, aside from minor formatting adjustments. The instructions included in the prompt directed a language model to categorize a given message based on the coding scheme and to format its response based on formatting rules. The coding scheme, largely identical to the one given to human evaluators, included the definitions and explanations of HTPs and 5 sentiment categories (ANTI, PRO, NEU, MIX, and IR).

LLMs generate a sequence of words by selecting each word based on its preceding words, and the selection is done by sampling a word from a large distribution of possible words [16,17]. Because of the inherent randomness in this sampling process, LLMs may produce different responses to the same prompt. This potential variability can be accounted for by generating multiple responses from an LLM using the same prompt [30]. To be specific, we collected 20 instances of responses for each message, referred to as "response instances." Each instance was obtained by initiating a new chat with a language model, sending the prompt, and saving the response from the model.

A language model's decision for each message was determined by randomly selecting m instances from a pool of 20 response instances with replacement. Then, the majority within the selected instances was identified. This majority outcome was termed the "machine decision." We assessed the machine decision by varying the number of response instances (m=1, 3, 5, 7, 9, and 11). For example, m=5 simulates a scenario where a user generates 5 response instances and identifies the majority among them. m=1 corresponds to a "one-shot" determination, where a single instance was generated and considered as the machine decision. In the case of a tie, an extra response instance was randomly selected until the tie was broken.

For each message and each value of m, the process of determining a machine decision was iterated 1000 times. After each iteration, a variable that we refer to as "human-machine concurrence" was recorded as 1 if the machine decision aligned with the human evaluation of the message. Otherwise, it was recorded as 0. This variable was then averaged across all iterations, yielding a value referred to as "accuracy." Thus, the accuracy in this study indicates how accurately the language models classify the sentiment of a message based on the m number of responses. Alternatively, the accuracy can be interpreted as the proportion of messages classified by the language models that match the human sentiment evaluation of the same messages. For instance, if the accuracy of a language model is 90% for ANTI message classification, this suggests that nine-tenths of human-labeled ANTI messages are categorized as ANTI by the model. To assess the overall accuracy of the model in evaluating the sentiment of a specific set of messages (eg, Facebook messages classified as ANTI by human evaluators), we calculated the average accuracy across messages in each set, denoted as K_m. For example, K_{11} for ANTI messages refers to the proportion of human-labeled ANTI messages that were also classified as ANTI by a language model based on 1000 iterations of the majority of randomly selected 11 responses out of the total of 20 responses. Examples of human-labeled ANTI, PRO, and NEU messages, along with the language models' sentiment classification decisions on the same messages, are provided in Table S16 in Multimedia Appendix 1.

Results

GPT-3.5 Evaluation

The average accuracy, based on 20 response instances (K_{20}), was 0.612 (SE 0.02) for long-form and 0.570 (SE 0.02) for short-form messages. However, the accuracy varied across categories. For messages categorized as ANTI by human evaluators, K_{20} was 0.755 (SE 0.028) for long-form and 0.696 (SE 0.03) for short-form messages. For messages categorized as PRO by human evaluators, the accuracy was 0.544 (SE 0.033) for long-form and 0.476 (SE 0.033) for short-form messages. The language model's average accuracy for messages classified as NEU by human evaluators was 0.461 (SE 0.04) for long-form and 0.507 (SE 0.042) for short-form messages. Table 1 presents the accuracy of different sentiment labels with varying m.

XSL•FO
RenderX

Table 1. Accuracy varying the number of response instances (GPT-3.5).

Platform (format) and m[a]	ANTI[b] (n=200)		PRO (n=200)[c]		NEU (n=100)[d]		All (N=500)	
	K_m[e] (SE)	K_m/K_{20}, %	K_m (SE)	K_m/K_{20}, %	K_m (SE)	K_m/K_{20}, %	K_m (SE)	K_m/K_{20}, %
Facebook (Long format)								
1	0.657 (0.023)	87.1	0.452 (0.025)	83	0.428 (0.024)	92.7	0.529 (0.015)	86.5
3	0.715 (0.025)	94.7	0.506 (0.029)	93.1	0.446 (0.030)	96.7	0.578 (0.017)	94.4
5	0.734 (0.026)	97.2	0.524 (0.030)	96.2	0.451 (0.033)	97.7	0.593 (0.018)	96.9
7	0.740 (0.027)	98	0.530 (0.031)	97.4	0.455 (0.034)	98.6	0.599 (0.018)	97.9
9	0.744 (0.027)	98.6	0.535 (0.031)	98.2	0.456 (0.036)	98.7	0.603 (0.019)	98.5
11	0.749 (0.027)	99.2	0.537 (0.032)	98.7	0.459 (0.037)	99.5	0.606 (0.019)	99.1
20	0.755 (0.028)	—[f]	0.544 (0.033)	—	0.461 (0.040)	—	0.612 (0.020)	—
Twitter (Short format)								
1	0.614 (0.025)	88.3	0.411 (0.025)	86.4	0.444 (0.026)	87.6	0.499 (0.016)	87.5
3	0.662 (0.027)	95.1	0.449 (0.029)	94.3	0.481 (0.033)	94.8	0.540 (0.018)	94.8
5	0.675 (0.028)	97.0	0.459 (0.031)	96.5	0.492 (0.036)	97.1	0.552 (0.019)	96.9
7	0.681 (0.029)	97.9	0.465 (0.031)	97.6	0.496 (0.038)	97.9	0.557 (0.019)	97.8
9	0.685 (0.029)	98.4	0.469 (0.032)	98.6	0.502 (0.039)	98.9	0.562 (0.019)	98.6
11	0.688 (0.029)	99	0.470 (0.032)	98.8	0.502 (0.040)	99	0.564 (0.020)	98.9
20	0.696 (0.030)	—	0.476 (0.033)	—	0.507 (0.042)	—	0.570 (0.020)	—

[a]m indicates the number of response instances used for majority determination. When m is greater than 1, machine decision is the majority among the response instances. When m equals 1, the machine decision is equal to the response instance (1-shot determination).

[b]anti-HTP messages.

[c]pro-HTP messages.

[d]neutral messages.

[e]K_m indicates the average of the machine accuracy of n messages when the machine decision for each message was determined based on m response instances.

[f]Not applicable.

Most discrepancies arose when the language model classified messages as NEU or IR, whereas human evaluators identified positive or negative sentiments in these messages. For instance, the model misclassified 24.5% (49/200) of human-evaluated long-form ANTI messages. Among these misclassified messages, the language model classified 61.2% (30/49), 26.5% (13/49), and 12.2% (6/49) as NEU, IR, and PRO. Tables S2, S4-S7, and Figures S1 and S2 in Multimedia Appendix 1 provide more detailed comparative descriptions of decisions made by human evaluators and GPT-3.5.

The significance of differences in accuracy between sentiment categories was examined. The results indicated that K_{20} for human-evaluated ANTI messages was significantly higher than that of human-labeled PRO (U=26483.5; P<.001) and NEU (U=14675.5; P<.001) messages in long form. This was also true for human-labeled short-form PRO (U=25876.5; P<.001) and NEU (U=13312.5, P<.001) messages. These gaps in accuracy were consistent across all m values and formats (see Tables S12 and S13 in Multimedia Appendix 1).

XSL•FO
RenderX

The accuracy improved as *m* increased, as visualized in Figure 1. In Figure 1, error bars represent the mean (SE) of the mean, and "m" refers to the number of response instances. However, even with a few response instances, the accuracy was comparable to the accuracy based on 20 response instances. For example, even the average accuracy of 1-shot determination (K_1) for human-labeled ANTI, PRO, and NEU messages also reached 87.1%, 83%, and 92.7% of K_{20} for long-form messages and 88.3%, 86.4%, and 87.6% of K_{20} for short-form messages.

Figure 1. Accuracy across response instances and message formats (GPT-3.5). HTP: heated tobacco product. ANTI: anti-HTP; PRO: pro-HTP; NEU: neutral.

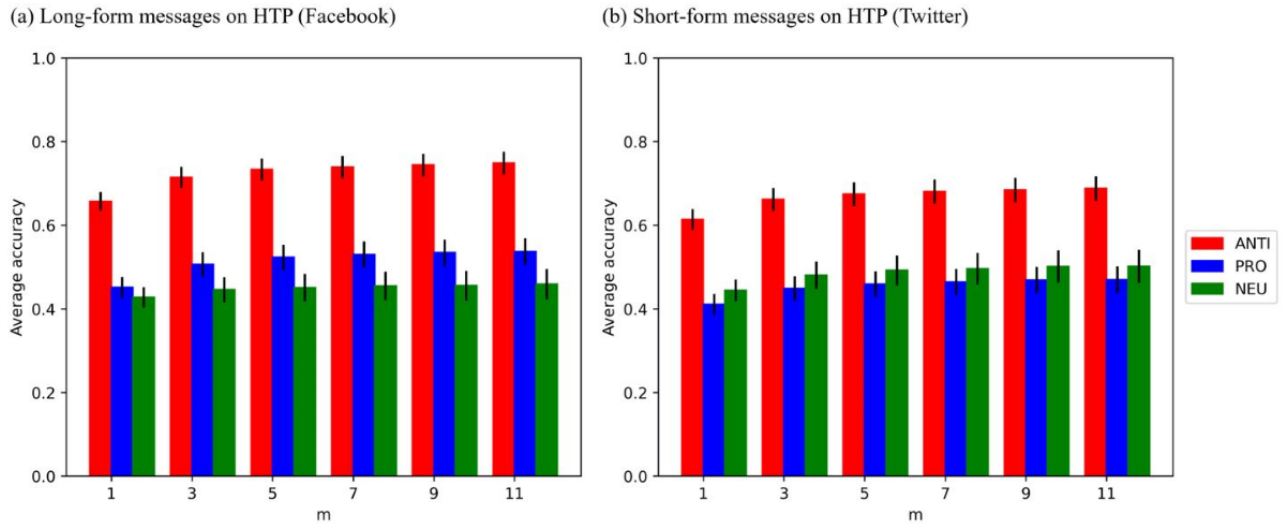

GPT-4 Turbo Evaluation

GPT-4 Turbo demonstrated higher accuracy than GPT-3.5 across all sentiment categories. The language model's overall average accuracy was 0.817 (SE 0.017) for long-form messages and 0.770 (SE 0.019) for short-form messages. Although accuracy varied across categories, the gap between the accuracy in ANTI and PRO sentiment classification decreased compared with GPT-3.5. For human-labeled ANTI messages, K_{20} was 0.861 (SE 0.024) for long-form messages and 0.79 (SE 0.028) for short-form messages. For human-labeled PRO messages, K_{20} was 0.84 (SE 0.025) for long-form messages and 0.783 (SE 0.029) for short-form messages. The accuracy for human-labeled NEU message categorization also increased compared to GPT-3.5. For NEU messages, K_{20} was 0.685 (SE 0.044) for long-form messages and 0.703 (SE 0.045) for short-form messages. The accuracy of sentiment categories with varying *m* is reported in Table 2.

XSL•FO
RenderX

Table 2. Accuracy varying the number of response instances (GPT-4 Turbo).

Platform (format) and m[a]	ANTI[b] (n=200)		PRO[c] (n=200)		NEU[d] (n=100)		All (N=500)	
	K_m[e] (SE)	K_m/K_{20}, %	K_m (SE)	K_m/K_{20}, %	K_m (SE)	K_m/K_{20}, %	K_m (SE)	K_m/K_{20}, %
Facebook (Long format)								
1	0.856 (0.023)	99.4	0.828 (0.024)	98.5	0.661 (0.038)	96.5	0.806 (0.016)	98.6
3	0.859 (0.024)	99.7	0.834 (0.025)	99.3	0.681 (0.041)	99.3	0.813 (0.016)	99.5
5	0.860 (0.024)	99.9	0.837 (0.025)	99.6	0.685 (0.042)	100	0.816 (0.016)	99.8
7	0.860 (0.024)	99.9	0.837 (0.025)	99.6	0.687 (0.042)	100	0.816 (0.017)	99.8
9	0.860 (0.024)	99.9	0.838 (0.025)	99.7	0.687 (0.043)	100	0.816 (0.017)	99.9
11	0.860 (0.024)	99.9	0.838 (0.025)	99.8	0.688 (0.043)	100	0.817 (0.017)	100
20	0.861 (0.024)	__[f]	0.840 (0.025)	—	0.685 (0.044)	—	0.817 (0.017)	—
Twitter (Short format)								
1	0.789 (0.027)	99.9	0.773 (0.028)	98.7	0.704 (0.040)	100	0.765 (0.018)	99.4
3	0.789 (0.028)	99.9	0.776 (0.028)	99.2	0.708 (0.042)	100	0.768 (0.018)	99.8
5	0.789 (0.028)	99.9	0.778 (0.028)	99.4	0.707 (0.043)	100	0.768 (0.018)	99.8
7	0.790 (0.028)	100	0.779 (0.028)	99.6	0.706 (0.043)	100	0.769 (0.018)	99.9
9	0.789 (0.028)	99.9	0.780 (0.029)	99.7	0.705 (0.044)	100	0.769 (0.018)	99.9
11	0.789 (0.028)	99.9	0.781 (0.029)	99.7	0.705 (0.044)	100	0.769 (0.018)	99.9
20	0.790 (0.028)	—	0.783 (0.029)	—	0.703 (0.045)	—	0.770 (0.019)	—

[a]m indicates the number of response instances used for majority determination. When m is greater than 1, machine decision is the majority among the response instances. When m equals 1, the machine decision is equal to the response instance (1-shot determination).

[b]anti-HTP messages.

[c]pro-HTP messages.

[d]neutral messages.

[e]K_m indicates the average of the machine accuracy of n messages when the machine decision for each message was determined based on m response instances.

[f]Not applicable.

GPT-4 Turbo showed fewer false selections of the IR label across all sentiment categories. When examined with 3 randomly selected response instances (m=3), the model misclassified only 10 out of the entire 1000 sample messages as IR. This impacted the pattern of mismatches between human and language model sentiment classification. For instance, the model categorized 14% (28/200) of human-labeled ANTI messages as one of the other labels. Of these mismatches, 93% (26/28) were classified as NEU. Multimedia Appendix 1 includes contingency tables (Tables S3 and S8-S11) as well as flow charts (Figure S2) providing more detailed comparative descriptions of human and language model sentiment labeling.

For long-form messages, the model's accuracy of ANTI classification based on 20 response instances was significantly higher than NEU classification (U=12,561.5; P<.001) but not significantly different from the accuracy of PRO classification (U=20,810.5; P=.30). PRO classification also showed significantly higher accuracy compared with NEU classification (U=12,176; P<.001). This pattern was observed across all m values. For short-form messages with lower m values, the

XSL•FO
RenderX

accuracy of ANTI and PRO classification was significantly greater than NEU labeling. For example, the 1-shot determination (K_1) for ANTI and PRO classification was significantly higher than for NEU classification (ANTI vs NEU: U=11,829; P<.001 and PRO vs NEU: U=11,625; P<.001). However, these differences diminished as m increased; K_{20} was not significantly different across the 3 sentiment categories. The difference test results across all m values are provided in Multimedia Appendix 1 (Table S14 and S15).

The model's accuracy in 1-shot cases was already comparable with that of 20 instances, as shown in Figure 2. Error bars represent the mean (SE) of the mean, and "m" refers to the number of response instances. Table 2 above presented that the accuracy of the 1-shot determination (K_1) reached at least 96.5% accuracy based on 20 response instances (K_{20}).

Figure 2. Accuracy across response instances and message formats (GPT-4 Turbo). HTP: heated tobacco product; ANTI: anti-HTP; PRO: pro-HTP; NEU: neutral.

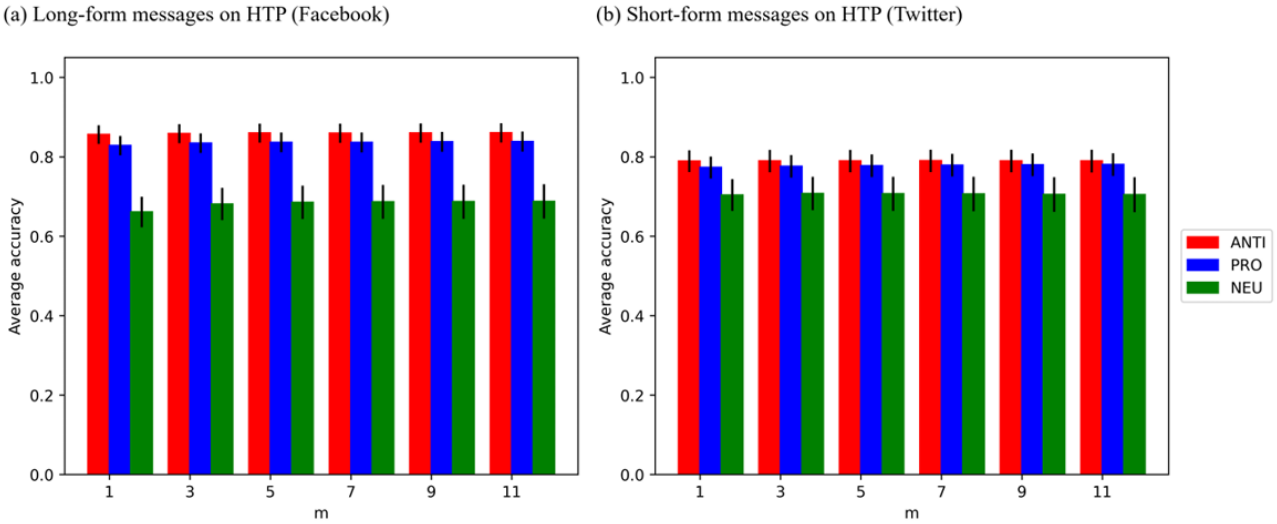

Discussion

GPT-4 Turbo accurately replicated 81.7% of human sentiment evaluations for long-form messages and 77% for short-form messages, based on 20 AI responses. In comparison, GPT-3.5's K_{20} indicated that the model's labeling decisions matched human coders' evaluations with 61.2% accuracy for long-form messages and 57% accuracy for short-form messages. In sum, GPT-4 Turbo showed improvements in accuracy compared to GPT-3.5 due to increased accuracy across all sentiment categories.

Focusing on GPT-4 Turbo, which showed better overall accuracy, the LLM demonstrated already high accuracy with a small number of responses. The difference in accuracy between a small number of responses (eg, m=1, 3) and a high number of responses (eg, m=20) was not statistically significant. For all sentiment categories, the language model's K_3 reached at least 99.2% of K_{20}. The model demonstrated similar levels of accuracy for ANTI and PRO labels for both long-form and short-form messages. While the accuracy of NEU classification was lower than that of ANTI and PRO classification, it increased by approximately 20 percentage points compared with GPT-3.5. These findings suggest that GPT-4 Turbo can yield more accurate sentiment classification decisions, even with a small number of response instances, such as 3.

GPT-3.5 showed discrepancies in accuracy across sentiment categories. Specifically, the accuracy of the ANTI classification was better than the PRO classification. This finding suggests the possibility of a relative underrepresentation of messages with positive sentiment compared with negative sentiment when using LLMs for sentiment analysis of tobacco-related social media discourses. This issue calls for further exploration of approaches, techniques, and procedures to assess, mitigate, or compensate for LLMs' potential inconsistencies across different sentiment categories, as well as the reasons underlying these discrepancies.

Employing a newer model may be the most straightforward solution, as shown by the comparison between older (GPT-3.5) and newer (GPT-4 Turbo) models. The newer model not only improved accuracy across all sentiment categories but also showed no significant difference in accuracy between ANTI and PRO classifications. This is particularly important for tobacco prevention researchers, as the detection of ANTI and PRO sentiments is important due to their possible associations with tobacco prevention outcomes [11,12,31]. However, a more recent model may not always perform better than its predecessors. For example, GPT-4 Turbo experienced a "laziness" issue, where the model does not complete user requests [32]. Therefore, the performance of new language models on specific tasks should also be rigorously tested.

Using language models specialized for health and medical information analyses, such as Google's Med-PaLM [33] and Stanford University's BioMedLM (previously PubMedGPT) [34], may influence accuracy. Research in this area is still emerging, with limited evidence on the application of these specialized LLMs for sentiment analysis. Previous studies have primarily used standard GPT models [35-38]. In addition, using these specialized models could be more difficult than widely used platforms like ChatGPT. Still, their performance in

XSL•FO
RenderX

analyzing the sentiment of public health-related social media messages warrants future investigations, considering their capabilities in handling content from the general public, not just academic researchers and professionals.

Prompt engineering could be another strategy for improving the accuracy of LLMs in sentiment analysis and reducing discrepancies across categories. In line with the objective of this study to facilitate a straightforward comparison between human coders' and language models' labeling decisions, we used a prompt that closely mirrored the coding scheme for human evaluators. However, different prompting techniques can lead to different results for similar requests [39,40]. Techniques such as few-shot prompting, which involves including task-related examples within the prompts, may enhance accuracy. For instance, rather than only defining sentiment labels, the coding scheme can provide several example messages for each label. Although these techniques complicate direct comparisons between human and machine classification, they possibly offer potential accuracy gains worth exploring.

Establishing and adopting coding procedures for LLM-involved coding is also worthwhile. A study investigating LLMs as substitutes for human coders in labeling texts on political topics serves as a good example [30]. The authors proposed a "hybrid" model where disagreements between the "GPT-4 first run" and the "GPT-4 second run" are resolved by a human coder. Their findings demonstrated that the hybrid approach can be optimized with minimal additional human effort and boosted the accuracy of GPT-4's annotations. This hybrid approach can potentially be adapted for the analysis of sentiment about health topics, and other coding procedures should be explored to further enhance the accuracy and efficiency.

The implications and future applications of our findings should be discussed with caution. First, this study is a focused case study on OpenAI's GPT, examining sentiment analysis on Twitter and Facebook messages related to HTPs. Future research can extend beyond this specific focus to evaluate the accuracy of LLMs in analyzing the sentiment of health-related information across a broader range of topics and platforms. Second, while the differences in accuracy for ANTI and PRO sentiment classifications that were present in GPT-3.5 disappeared in GPT-4 Turbo, and the accuracy of NEU classification increased by 20%, the NEU classification accuracy remained lower. A simple explanation might be the underperformance of the models. For instance, GPT-4 Turbo misclassified a human-labeled NEU message, which used HTPs as examples to explain an economic principle, as irrelevant. Alternatively, this difference might stem from the inherent complexity of evaluating neutrality [15,41]. For example, human coders classified a message as NEU, describing IQOS (Philip Morris International) as a device that uses a "patented heat-control technology." In contrast, GPT-4 Turbo classified the same message as PRO, interpreting "patented heat-control technology" to have positive connotations. It may be worthwhile for future research to explore the patterns of misclassifications by LLMs. Third, this study did not address the potential ethical issues of using LLMs for sentiment analysis of social media content. Ethical considerations, such as security, privacy protection, and data ownership, are important when using LLMs to analyze social media messages [42,43]. These issues require careful attention when analyzing health-related information using LLMs, too [44,45]. Future research should use LLMs while carefully considering the potential ethical issues surrounding their content analyses.

Acknowledgments

We thank Kyungmin Kang for her contribution to the message evaluation.

Data Availability

The datasets generated during and/or analyzed during this study are available in the Open Science Framework (OSF) repository [46].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online supplements (Table S1-S16, Figure S1 and Figure S2).
[PDF File (Adobe PDF File), 818 KB - jmir_v27i1e63631_app1.pdf]

References

1. Heated tobacco products. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/tobacco/other-tobacco-products/heated-tobacco-products.html [accessed 2025-01-21]
2. Abroms L, Levine H, Romm K, Wysota C, Broniatowski D, Bar-Zeev Y, et al. Anticipating IQOS market expansion in the United States. Tob Prev Cessat 2022;8:04 [FREE Full text] [doi: 10.18332/tpc/144650] [Medline: 35174298]
3. Jun J, Zhang N, Zain A, Mohammadi E. Social media discussions on the FDA's modified risk tobacco product authorization of IQOS. Subst Use Misuse 2022;57(3):472-480. [doi: 10.1080/10826084.2021.2019777] [Medline: 34986075]

XSL•FO
RenderX

4. Abroms LC, Wysota CN, Tosakoon S, Khayat A, Duan Z, Bar-Zeev Y, et al. Industry marketing of tobacco products on social media: case study of philip morris international's IQOS. Tob Control 2024;33(4):518-524. [doi: 10.1136/tc-2022-057833] [Medline: 36927516]
5. O'Brien EK, Hoffman L, Navarro MA, Ganz O. Social media use by leading US e-cigarette, cigarette, smokeless tobacco, cigar and hookah brands. Tob Control 2020;29(e1):e87-e97. [doi: 10.1136/tobaccocontrol-2019-055406] [Medline: 32217772]
6. Berg CJ, Romm KF, Bar-Zeev Y, Abroms LC, Klinkhammer K, Wysota CN, et al. IQOS marketing strategies in the USA before and after US FDA modified risk tobacco product authorisation. Tob Control 2023;32(4):418-427 [FREE Full text] [doi: 10.1136/tobaccocontrol-2021-056819] [Medline: 34667105]
7. Barker JO, Vassey J, Chen-Sankey JC, Allem J, Cruz TB, Unger JB. Categorizing IQOS-related twitter discussions. Int J Environ Res Public Health 2021;18(9):4836 [FREE Full text] [doi: 10.3390/ijerph18094836] [Medline: 33946546]
8. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of E-cigarette discussions on twitter. J Med Internet Res 2015;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]
9. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among twitter users: content and sentiment analysis. J Med Internet Res 2017;19(2):e56 [FREE Full text] [doi: 10.2196/jmir.5694] [Medline: 28242592]
10. Kim K, Gibson LA, Williams S, Kim Y, Binns S, Emery SL, et al. Valence of media coverage about electronic cigarettes and other tobacco products from 2014 to 2017: evidence from automated content analysis. Nicotine Tob Res 2020;22(10):1891-1900 [FREE Full text] [doi: 10.1093/ntr/ntaa090] [Medline: 32428214]
11. Hornik R, Binns S, Emery S, Epstein V, Jeong M, Kim K, et al. The effects of tobacco coverage in the public communication environment on young people's decisions to smoke combustible cigarettes. J Commun 2022;72(2):187-213 [FREE Full text] [doi: 10.1093/joc/jqab052] [Medline: 35386823]
12. Unger JB, Urman R, Cruz TB, Majmundar A, Barrington-Trimis J, Pentz MA, et al. Talking about tobacco on Twitter is associated with tobacco product use. Prev Med 2018;114:54-56 [FREE Full text] [doi: 10.1016/j.ypmed.2018.06.006] [Medline: 29898418]
13. Siegel LN, Levin AV, Kranzler EC, Gibson LA. Do longitudinal trends in tobacco 21-related media coverage correlate with policy support? An Exploratory analysis using supervised and unsupervised machine learning methods. Health Commun 2022;37(1):29-38 [FREE Full text] [doi: 10.1080/10410236.2020.1816282] [Medline: 32900231]
14. Fu R, Kundu A, Mitsakakis N, Elton-Marshall T, Wang W, Hill S, et al. Machine learning applications in tobacco research: a scoping review. Tob Control 2023;32(1):99-109. [doi: 10.1136/tobaccocontrol-2020-056438] [Medline: 34452986]
15. Riffe D, Lacy S, Watson BR, Lovejoy J. Analyzing Media Messages: Using Quantitative Content Analysis in Research. United Kingdom: Routledge; 2023.
16. Jurafsky D, Martin JH. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2023. URL: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf [accessed 2023-10-05]
17. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. Cambridge, MA, USA: The MIT press; 2018.
18. OpenAI. GPT-4 technical report. ArXiv. Preprint posted online on March 15, 2023 2021 [FREE Full text] [doi: 10.48550/arXiv.2303.08774]
19. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
20. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt J, Laleh NG, et al. The future landscape of large language models in medicine. Commun Med (Lond) 2023;3(1):141 [FREE Full text] [doi: 10.1038/s43856-023-00370-1] [Medline: 37816837]
21. Yu H, Fan L, Li L, Zhou J, Ma Z, Xian L, et al. Large language models in biomedical and health informatics: a review with bibliometric analysis. J Healthc Inform Res 2024;8(4):658-711. [doi: 10.1007/s41666-024-00171-8] [Medline: 39463859]
22. Sheeran P, Kenny A, Bermudez A, Gray K, Galper EF, Boynton M, et al. Artificial intelligence simulation of adolescents' responses to vaping-prevention messages. JAMA Pediatr 2024;178(5):504-506. [doi: 10.1001/jamapediatrics.2024.0274] [Medline: 38526479]
23. Ashokkumar A, Hewitt L, Ghezae I, Willer R. Predicting results of social science experiments using large language models. URL: https://docsend.com/view/ity6yf2dansesucf [accessed 2025-02-12]
24. Argyle LP, Busby EC, Fulda N, Gubler JR, Rytting C, Wingate D. Out of one, many: using language models to simulate human samples. Polit. Anal 2023;31(3):1-15. [doi: 10.1017/pan.2023.2]
25. OpenAI says ChatGPT's weekly users have grown to 200 million. Reuters. 2024. URL: https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/ [accessed 2024-11-29]
26. Introducing ChatGPT. OpenAI. URL: https://openai.com/index/chatgpt/ [accessed 2024-11-29]
27. New models and developer products announced at DevDay. Open AI. URL: https://openai.com/index/new-models-and-developer-products-announced-at-devday/ [accessed 2024-11-29]
28. What data is CrowdTangle tracking? CrowdTangle. 2023. URL: http://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking [accessed 2023-11-01]
29. X data for academic research. X. URL: https://developer.x.com/en/use-cases/do-research/academic-research [accessed 2023-11-01]
30. Heseltine M, Clemm VHB. Large language models as a substitute for human experts in annotating political text. Res Polit 2024;11(1):1-10. [doi: 10.31219/osf.io/cx752]

XSL•FO
RenderX

31. Kim K. Scanned information exposure and support for tobacco regulations among US youth and young adult tobacco product users and non-users. Health Educ Res 2023;38(5):426-444. [doi: 10.1093/her/cyad033] [Medline: 37565566]
32. New embedding models and API updates. OpenAI. URL: https://openai.com/index/new-embedding-models-and-api-updates/ [accessed 2024-11-30]
33. Med-PaLM: A large language model from Google Research, designed for the medical domain. Google Research. URL: https://sites.research.google/med-palm/ [accessed 2024-11-30]
34. Stanford CRFM introduces PubMedGPT 2.7B. Stanford University. URL: https://hai.stanford.edu/news/stanford-crfm-introduces-pubmedgpt-27b [accessed 2024-11-30]
35. Liyanage CR, Gokani R, Mago V. GPT-4 as an X data annotator: Unraveling its performance on a stance classification task. PLoS One 2024;19(8):e0307741 [FREE Full text] [doi: 10.1371/journal.pone.0307741] [Medline: 39146280]
36. Rathje S, Mirea D, Sucholutsky I, Marjieh R, Robertson CE, Van Bavel JJ. GPT is an effective tool for multilingual psychological text analysis. Proc Natl Acad Sci USA 2024;121(34):e2308950121. [doi: 10.1073/pnas.2308950121]
37. Lossio-Ventura JA, Weger R, Lee AY, Guinee EP, Chung J, Atlas L, et al. A comparison of ChatGPT and fine-tuned open pre-trained transformers (OPT) against widely used sentiment analysis tools: sentiment analysis of COVID-19 survey data. JMIR Ment Health 2024;11:e50150 [FREE Full text] [doi: 10.2196/50150] [Medline: 38271138]
38. Fu Z, Hsu YC, Chan CS, Lau CM, Liu J, Yip PSF. Efficacy of ChatGPT in cantonese sentiment analysis: comparative study. J Med Internet Res 2024;26:e51069 [FREE Full text] [doi: 10.2196/51069] [Medline: 38289662]
39. Schulhoff S, Ilie M, Balepur N, Kahadze K, Liu A, Si C, et al. The prompt report: a systematic survey of prompting techniques. ArXiv. Preprint posted online on June 6, 2024 2024 [FREE Full text]
40. Sahoo P, Singh A, Saha S, Jain V, Mondal S, Chadha A. A systematic survey of prompt engineering in large language models: techniques and applications. ArXiv. Preprint posted online on February 5, 2024 2024 [FREE Full text]
41. Liu B. Sentiment Analysis and Opinion Mining. New York City: Springer Nature; 2022.
42. Wu X, Duan R, Ni J. Unveiling security, privacy, and ethical concerns of ChatGPT. Journal of Information and Intelligence 2024;2(2):102-115. [doi: 10.1016/j.jiixd.2023.10.007]
43. Peris C, Dupuy C, Majmudar J, Parikh R, Smaili S, Zemel R, et al. Privacy in the time of language models. : ACM; 2023 Presented at: Proceedings of the WSDM '23: The Sixteenth ACM International Conference on Web Search and Data Mining; February 27-March 3, 2023; Singapore p. 1291-1292. [doi: 10.1145/3539597.3575792]
44. Liu S, Guo LR. Data ownership in the AI-powered integrative health care landscape. JMIR Med Inform 2024;12:e57754 [FREE Full text] [doi: 10.2196/57754] [Medline: 39560980]
45. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digit Health 2023;5(6):e333-e335 [FREE Full text] [doi: 10.1016/S2589-7500(23)00083-3] [Medline: 37120418]
46. Online supplements for "large language models' accuracy in emulating human experts' evaluation of public sentiments about heated tobacco products on social media: evaluation study". OSF. URL: https://doi.org/10.17605/osf.io/6teuz [accessed 2025-02-13]

Abbreviations

AI: artificial intelligence
CT: CrowdTangle
HTP: heated tobacco product
LLM: large language model

Edited by Q Jin; submitted 29.06.24; peer-reviewed by Z Xie, M Navarro, F Chen; comments to author 08.09.24; revised version received 27.12.24; accepted 19.01.25; published 04.03.25.

Please cite as:
Kim K, Kim S
Large Language Models' Accuracy in Emulating Human Experts' Evaluation of Public Sentiments about Heated Tobacco Products on Social Media: Evaluation Study
J Med Internet Res 2025;27:e63631
URL: https://www.jmir.org/2025/1/e63631
doi:10.2196/63631
PMID:40053746

©Kwanho Kim, Soojong Kim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigating Whether AI Will Replace Human Physicians and Understanding the Interplay of the Source of Consultation, Health-Related Stigma, and Explanations of Diagnoses on Patients' Evaluations of Medical Consultations: Randomized Factorial Experiment

Weiqi Guo[1], PhD; Yang Chen[2], PhD

[1]School of Foreign Languages, Renmin University of China, Beijing, China

[2]School of Journalism and Communication, Renmin University of China, Beijing, China

Corresponding Author:
Yang Chen, PhD
School of Journalism and Communication
Renmin University of China
Room 720, Mingde Xinwen Building
No.59 Zhongguancun St., Haidian
Beijing, 100872
China
Phone: 86 1062514835
Fax: 86 1062514835
Email: 20050022@ruc.edu.cn

Abstract

Background: The increasing use of artificial intelligence (AI) in medical diagnosis and consultation promises benefits such as greater accuracy and efficiency. However, there is little evidence to systematically test whether the ideal technological promises translate into an improved evaluation of the medical consultation from the patient's perspective. This perspective is significant because AI as a technological solution does not necessarily improve patient confidence in diagnosis and adherence to treatment at the functional level, create meaningful interactions between the medical agent and the patient at the relational level, evoke positive emotions, or reduce the patient's pessimism at the emotional level.

Objective: This study aims to investigate, from a patient-centered perspective, whether AI or human-involved AI can replace the role of human physicians in diagnosis at the functional, relational, and emotional levels as well as how some health-related differences between human-AI and human-human interactions affect patients' evaluations of the medical consultation.

Methods: A 3 (consultation source: AI vs human-involved AI vs human) × 2 (health-related stigma: low vs high) × 2 (diagnosis explanation: without vs with explanation) factorial experiment was conducted with 249 participants. The main effects and interaction effects of the variables were examined on individuals' functional, relational, and emotional evaluations of the medical consultation.

Results: Functionally, people trusted the diagnosis of the human physician (mean 4.78-4.85, SD 0.06-0.07) more than medical AI (mean 4.34-4.55, SD 0.06-0.07) or human-involved AI (mean 4.39-4.56, SD 0.06-0.07; $P<.001$), but at the relational and emotional levels, there was no significant difference between human-AI and human-human interactions ($P>.05$). Health-related stigma had no significant effect on how people evaluated the medical consultation or contributed to preferring AI-powered systems over humans ($P>.05$); however, providing explanations of the diagnosis significantly improved the functional ($P<.001$), relational ($P<.05$), and emotional ($P<.05$) evaluations of the consultation for all 3 medical agents.

Conclusions: The findings imply that at the current stage of AI development, people trust human expertise more than accurate AI, especially for decisions traditionally made by humans, such as medical diagnosis, supporting the algorithm aversion theory. Surprisingly, even for highly stigmatized diseases such as AIDS, where we assume anonymity and privacy are preferred in medical consultations, the dehumanization of AI does not contribute significantly to the preference for AI-powered medical agents over humans, suggesting that instrumental needs of diagnosis override patient privacy concerns. Furthermore, explaining the diagnosis

XSL•FO
RenderX

effectively improves treatment adherence, strengthens the physician-patient relationship, and fosters positive emotions during the consultation. This provides insights for the design of AI medical agents, which have long been criticized for lacking transparency while making highly consequential decisions. This study concludes by outlining theoretical contributions to research on health communication and human-AI interaction and discusses the implications for the design and application of medical AI.

(J Med Internet Res 2025;27:e66760) doi:10.2196/66760

KEYWORDS

artificial intelligence; AI; medical artificial intelligence; medical AI; human–artificial intelligence interaction; human-AI interaction; medical consultation; health-related stigma; diagnosis explanation; health communication

Introduction

Background

Artificial intelligence (AI) is considered a revolutionary technology for the health care sector. Analysts have forecasted massive growth in the global AI health care market over the next decade [1]. AI is now used in medical scenarios in 2 forms: either to make independent medical decisions (ie, autonomous AI) or to collaborate with human physicians and make decisions with human involvement (ie, human-involved AI) [2]. While ideally, AI diffusion promises to improve diagnostic accuracy, save human effort, and increase patient convenience, empirical research is needed to assess the merits and limitations of medical AI [3], especially from a patient-centered perspective [4]. This perspective is significant because AI as a technological solution does not necessarily improve the patient's confidence in the diagnosis and adherence to treatment at the functional level, establish meaningful interactions between the medical agent and the patient at the relational level, and stimulate positive emotions in the patient and reduce pessimism at the emotional level. This leads to the guiding research question (RQ) of this study as follows: From the patient's perspective, can autonomous AI or human-involved AI replace the role of human physicians at the functional, relational, and emotional levels?

Previous studies and practical necessities motivate this study in several ways. First, although previous studies have explored human-AI interactions guided by some established frameworks, such as the computers are social actors (CASA) paradigm or algorithm aversion theory, by analyzing scenarios of web-based content production and moderation, consumer behavior, and so on [5,6], there is still a lack of empirical evidence as to whether these conclusions can adequately address the complexities of the medical scenario. Little research has systematically examined, from a user-centered approach, the functional, relational, and emotional evaluations of human-AI interactions in medical scenarios, which may yield different conclusions from other scenarios when contextual factors, such as the severe consequences of medical interventions and the attributes of the disease, are taken into account. In addition, few studies have tested user evaluations of human-involved AI, especially vis-à-vis AI and humans; however, this model of human-AI collaboration is increasingly prevalent in medical scenarios.

Second, a crucial distinction between human-AI interaction and human-human interaction in the health domain is that of the human touch, a quality traditionally associated with human physicians, which may affect the patient's evaluation of the medical consultation. Therefore, we incorporated health-related stigma into the analysis of human-AI interaction. Stigma is often associated with human-human interactions, and in the medical scenario, it may lead to patients not seeking a diagnosis or withholding information from physicians. However, this also presents opportunities to introduce AI in a manner that eliminates the human element from medical services. Earlier studies have primarily examined health-related stigma and human-AI interaction separately [7,8]; however, by linking these two variables, this study is able to address the broader question of how disease attributes and communicator characteristics influence the preference for machines over humans in medical scenarios and whether dehumanization is sometimes beneficial in health care.

Third, previous theories have consistently emphasized the accuracy and convenience of AI-powered systems as reasons for preferring AI over humans [9]. In the medical scenario, however, is it enough for AI to merely be accurate and efficient to replace human practitioners? In particular, given the long-standing issue of accurate diagnoses being overlooked due to a lack of explanation and low patient comprehension in traditional patient-physician communication [10] and the black-box nature of AI that systematically ignores explanations for users [11], it is imperative to investigate whether providing explanations of the diagnostic decision-making process enhances effective communication between the medical agent and the patient. This could shed light on how AI-powered medical agents should be designed from a user-centered approach.

To answer these questions, an experiment was conducted based on the theories of human-computer interaction (HCI), human-machine communication (HMC), and health communication. We constructed different scenarios where the source of medical consultation (ie, AI vs human-involved AI vs human), the stigma of the health issue (ie, high vs low), and the explanation of diagnosis (without vs with explanation) were manipulated and examined their individual and interaction effects on people's functional, relational, and emotional evaluations of the medical consultation. On the basis of the findings, we outlined the theoretical contributions to the understanding of human-AI interaction and health communication and practical implications for the principles of human-AI collaboration in health care and the design of AI-powered medical agents.

XSL•FO
RenderX

Literature Review

AI, Human-Involved AI, and Humans as Sources of Medical Consultation

AI is considered promising for enhancing the efficiency, accuracy, and quality of medical services. People are increasingly turning to AI-powered instruments to understand their symptoms, seek diagnoses, and explore treatment options, which previously occurred between patients and human physicians in offline clinical environments [12]. Currently, there are 2 paradigms for the application of AI in medical consultation and diagnosis. Patients may encounter autonomous AI that replaces human physicians and makes medical decisions independently [2]. They can also opt for human-involved AI, where human physicians collaborate with AI, participate in the medical decision-making loop, and monitor AI's diagnostic results and treatment recommendations [13,14]. Human-involved AI is an intermediate paradigm between autonomous AI and human practitioners, synthesizing the functional and relational characteristics of both. Despite the growing importance of AI, few studies have systematically examined people's functional, relational, and emotional evaluations of AI and human-involved AI versus humans in medical settings, along with the merits and shortcomings of the 3 medical agents in disease diagnosis, which are the focus of this study.

Existing HCI and HMC theories offer conflicting predictions regarding how users perceive AI versus humans. The CASA paradigm suggests that people apply the same social rules and expectations to machines as to humans despite knowing that they are inanimate agents. The implication is that there should be no difference in users' perceptions of AI and humans [15]. However, as human-AI interactions become more common and technological affordances continue to advance, some recent evidence does not support the CASA paradigm [16]. Another framework based on the concept of machine heuristic suggests that people should favor AI as a source of communication over humans. The argument is that when people interact with machines, they tend to use mental shortcuts and apply general stereotypes of machines, such as being objective, accurate, unbiased, and predictive, thus viewing AI and its decisions as more favorable compared to human decisions [17]. Evidence of such AI preferences has been reported in health, media, and organizational settings, among others, in both low-risk and high-risk contexts [18-21]. In contrast, the algorithm aversion literature argues that people are suspicious of AI decision makers for several reasons, including the lack of transparency of algorithms in the decision-making process and the absence of human expertise in certain domains [22]. Accordingly, people prefer human decision makers to AI, especially for decisions that are traditionally made by humans, despite acknowledging the fact that machines' overall performance is superior to that of humans [13].

Whether these theories could capture the complexities of the medical scenario remains unclear, as many studies approach this question by analyzing scenarios of web-based content moderation, media production, or consumer behaviors. In addition, little evidence exists to test user perceptions of human-involved AI, especially vis-à-vis AI and humans; however, this model of human-AI cooperation is very common in medical scenarios. Inspired by Guzman and Lewis [23], this study regards medical AI and human-involved AI as communicators similar to human physicians, rather than simple technologies mediating human interactions and explores users' functional, relational, and emotional evaluations of the 3 agents in a medical setting.

Regarding users' functional evaluations of the agents, the machine heuristic argument posits that medical AI exhibits a superior level of information processing ability compared to humans. AI trains and refines its decision-making models based on vast quantities of medical data and adheres to a consistent decision-making process that yields rational solutions, which could culminate in higher functional evaluations of AI from users [9]. Conversely, the algorithmic aversion literature suggests that because disease diagnosis and treatment are traditionally performed by humans, users may perceive human physicians as more authoritative and reliable than medical AI [24]. When AI and human physicians collaborate in diagnosis (ie, human-involved AI), it is expected that the benefits of both approaches will be integrated to provide users with fast, accurate, and authoritative medical decisions and thus generate the highest functional evaluations.

Regarding relational evaluations, machine heuristic may make users less likely to consider relational attributes of AI during interactions, making it more difficult for medical AI to build relationships with users in the way that human physicians do. However, the CASA paradigm offers opposing predictions in this regard [25]. When the communicator is a human-involved AI, the medical consultation involves both human attributes of interpersonal communication and elements of dehumanized HMC. Therefore, the user's relational evaluation of the agent is hypothesized to be intermediate between that of an autonomous AI and a human physician. In addition to functional and relational evaluations, it is essential to understand how different medical agents induce emotional responses in patients. Are different types of medical agents equally effective in helping people comprehend their health conditions (ie, functionally helpful), providing empathetic support (ie, relationally supportive), and thus eliciting positive emotional responses (eg, optimism) and reducing negative emotions (eg, guilt about their conditions)?

Taken together, we proposed RQ1: Are there differences in people's functional, relational, and emotional evaluations of medical consultations with different medical agents (AI vs human-involved AI vs human)?

Health-Related Stigma and Its Influence on Medical Consultations

Stigma is a mark or condition that is related to social dishonor, stereotyping, and discriminatory beliefs [26,27]. One health issue that is typically stigmatized is HIV infection or AIDS. Patients with stigmatized diseases, such as AIDS, are reportedly hesitant to seek treatment due to the social shame attached to these diseases and the unavoidable disclosure of private information, such as risky sexual behaviors, during medical consultations [28]. Face-to-face consultations can be particularly uncomfortable, causing patients to withhold relevant information

XSL•FO
RenderX

or lie to physicians [29]. Consequently, as digital health care applications proliferate, patients with stigmatized health conditions may turn to AI-powered medical agents rather than human physicians for medical services [12]. This means that health-related stigma may influence health care preferences, thereby altering the functional, relational, and emotional evaluations of consultations with different medical agents.

The technological affordances of AI facilitate new experiences in medical consultations for people with stigmatized diseases, compared to interactions with human physicians. Medical AI promises greater anonymity and privacy. If patients want to consult without revealing their identity, AI has an advantage over clinical visits or phone consultations [30]. Furthermore, people tend to view AI as a nonjudgmental tool and thus feel more comfortable disclosing personal health information, such as high-risk sexual conduct, to a medical AI than to human physicians [27,31]. This indicates the potential necessity for dehumanization in medical consultations, particularly when stigma is present [29].

This study examined the effect of stigma on preferences for medical agents in the context of 2 conditions, AIDS and heart disease, the former being more stigmatized and the latter less. Previous research has demonstrated the influence of disease severity on patient preferences for medical agents [32]. This study aimed to control for the effect of disease severity and focus on the role of stigma. While AIDS and heart disease are similar in terms of disease severity, as both have serious health consequences and require accurate treatment, they diverge in terms of stigma. AIDS is typically stigmatized, whereas heart disease is not subject to cultural judgments about one's private life and is not infective. On the basis of the aforementioned theoretical discussions, we expected that in stigma-laden conditions (eg, AIDS), where patients need instrumental guidance and prefer that the medical agent refrains from judging their lifestyles, reducing the human element in medical consultations would enhance patients' evaluations of the medical experience. When stigma was not involved (eg, heart disease), we expected that patients would still prefer human expertise and human touch during medical consultations and would give higher functional, relational, and emotional ratings to human physicians than to autonomous AI or human-involved AI.

RQ2 is as follows: How does health-related stigma (AIDS vs heart disease) affect the functional, relational, and emotional evaluations of the medical consultation?

Explanations of Diagnoses and Evaluation of Medical Consultation

The effectiveness of physician-patient communication is critical to patients' evaluations of medical consultations and the achievement of positive health outcomes. The physician's ability to explain the condition and diagnosis to the patient is of particular significance [33]. Explanation is defined as the details or reasons given to facilitate comprehension. For health care professionals, effective diagnostic explanations require interpreting symptoms and justifying diagnostic and treatment decisions. Operationally, the diagnostic explanation in this study encompassed 4 aspects. First, patients should be informed of the symptom descriptions required to confirm their conditions. Second, the rationale for the diagnosis formulated must be explained, which includes clarifying the range of possible diagnoses based on the symptom descriptions and the reasons why a specific diagnosis is made. Third, different treatment options should be provided and explained, and a treatment recommendation should be offered. Finally, the efficacy of the relevant medication and the suggested medication should be explained.

In traditional clinical visits, patients have complained about the lack of explanation of diagnosis and treatment [10]. Sometimes, patients must turn to web-based communities to decipher the medical knowledge and reasoning embedded in medical disclosure documents [34]. Following this line of reasoning, as AI medical agents become more prevalent, the following questions arise. Do people have the same expectations when using medical AI and human-involved AI? If providing an explanation of the diagnosis enhances people's evaluations of medical consultations, does this approach to enhancing physician-patient communication differ between different types of medical agents (ie, medical AI, human-involved AI, and human physician) and between different types of diseases (ie, highly stigmatized and less stigmatized diseases)? This study aimed to answer these questions.

Owing to the black-box nature of AI, decisions made by AI-involved systems are often criticized for lacking transparency and accountability [11]. When making high-impact decisions such as medical diagnosis, human understanding of the reasoning behind the decision-making process without the need to comprehend the internal structure or underlying algorithms, which is defined as the principle of intelligibility of explainable AI, should be helpful and satisfying to content recipients [35], thereby enhancing the functional, relational, and emotional evaluations of the human-AI interaction. For medical AI to replace human practitioners, being accurate is not enough. Empirical evidence indicates that an accurate diagnosis, without explanation or justification, may be ignored by patients, even if made by a highly competent physician, let alone by AI systems, which have long been criticized as unsafe due to their lack of explicability [4,36,37]. Therefore, we hypothesized that providing explanations of the diagnosis, whether by a human physician, medical AI, or human-involved AI, would improve patients' functional, relational, and emotional evaluations of the medical consultation. Furthermore, we predicted that the greater the involvement of AI in the diagnosis, the more patients would require diagnostic explanations due to the increased black-box problem associated with AI. This, in turn, may affect their evaluations of the medical consultation on all 3 dimensions.

Regarding the interaction effect of health-related stigma and diagnosis explanation on patients' evaluations of medical consultations, as previously discussed, regardless of the type of the medical agent, when diagnosis explanations are provided, patients with stigmatized and nonstigmatized conditions will similarly have a better understanding of their conditions and higher ratings of medical consultations [33]. If the medical agent does not explain the diagnosis, patients with a nonstigmatized condition may be less satisfied with the diagnosis, especially if the consultation is with an AI-involved system rather than a human physician. In contrast, people with stigmatized diseases

XSL•FO
RenderX

are more likely to avoid further discussion with the medical agent, which potentially results in the revelation of more personal and stigmatized health information. Nevertheless, they are compelled to seek instrumental medical services due to the high risk of their health conditions [29]. Therefore, even if the medical agent only offers a diagnosis and treatment options but does not provide any explanation of the decision-making process, patients with stigmatized health issues are more likely to accept the diagnosis without requiring transparency.

Taken together, we proposed RQ3: Does providing explanations of the diagnosis improve the functional, relational, and emotional evaluations of the medical consultation? Do the 3 types of medical agents (ie, medical AI, human-involved AI, and human physician) and the 2 types of diseases (ie, highly stigmatized and less stigmatized diseases) differ in this path to improving the effectiveness of physician-patient communication?

Methods

Research Design

A 3 (consultation source: AI vs human-involved AI vs human) × 2 (health-related stigma: high vs low) × 2 (diagnosis explanation: without explanation vs with explanation) factorial experiment was conducted in November 2023 to test the effects of medical consultation source, health-related stigma, and diagnosis explanation on the functional, relational and emotional evaluations of human-AI interaction in a diagnostic setting. Consultation sources and health-related stigma are between-subject factors. Diagnosis explanation is a within-subject factor. Data were analyzed using SPSS Statistics (IBM Corp), consisting of descriptive statistics, the 2-tailed *t* test, and ANOVA.

Ethical Considerations

Institutional review board approval was obtained from Renmin University of China. Undergraduate students participated in the study for partial course credits. Before collecting data, we asked for participants' consent, explained the experiment's purpose and the use of data, and assured them that participation was completely voluntary and anonymous. The data were anonymized, containing no identifiable information about the participants.

Procedure and Stimuli

Figure 1 illustrates the flow of the experiment. A total of 249 participants (mean age 18.67, SD 0.972 y; n=65, 26.1% male) were randomly assigned to 1 of the 6 experimental groups, which were conducted offline in 6 laboratories. Each participant was presented with condition-specific stimulus stories and given adequate reading time. Stimulus stories are conversations between an information seeker (concerned about AIDS or heart disease) and 1 of the 3 medical agents (an autonomous medical AI, a human-involved medical AI, or a human physician), presented in text form and typeset in a common one-to-one web-based chat interface (refer to Multimedia Appendix 1 for an example). Participants were first required to read an introduction in which they were informed that they had recently been unwell and wished to confirm their condition and obtain advice on treatment through a web-based medical consultation.

Figure 1. Experiment flow. AI: artificial intelligence.

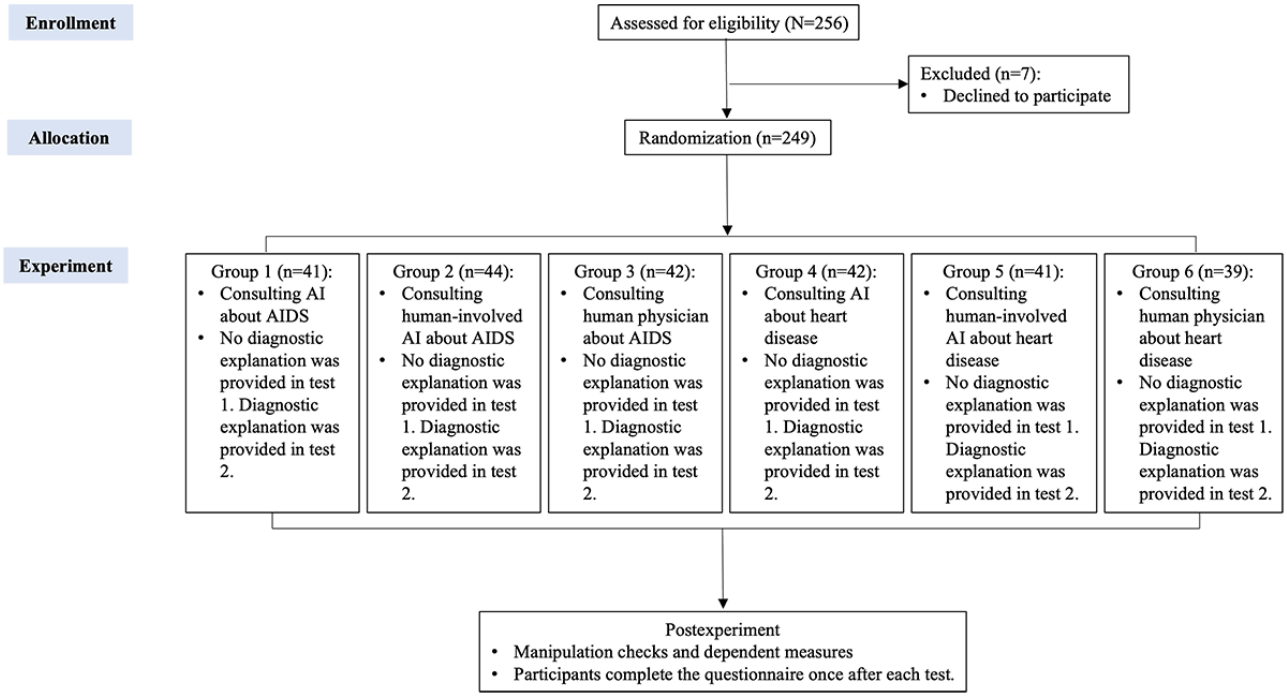

Then, to manipulate the consultation source, participants read prompts introducing the medical agent, such as "You are having a consultation with a medical AI. It learns from past medical data of patients with the same condition, makes a diagnosis based on the symptoms you've described, and offers treatment options. Human physicians are not involved in diagnosis, medication recommendations or monitoring of the consultation process" (the autonomous AI scenario). We also emphasized

XSL•FO
RenderX

the manipulation of the source in the conversation scripts by referring to the medical agent as autonomous AI, human-involved AI, or human physician.

We manipulated health-related stigma through the conversation between the medical agent and the information seeker, with the high-stigma scenario involving symptoms of AIDS and the low-stigma scenario involving symptoms of heart disease. In a question-and-answer format, the patient described the symptoms and provided responses to the queries posed by the medical agent about his or her medical history. The medical agent proceeded to offer a diagnosis of the condition and treatment options. Participants consulting for the same condition received the same diagnosis and treatment options, either from an autonomous AI, a human-involved AI, or a human physician. We adapted the conversations based on the output of several leading medical AI models regarding AIDS and heart disease, also incorporating authentic records of web-based consultations provided by human physicians concerning the 2 diseases. The materials for the AIDS and heart disease groups differed in the patients' symptom descriptions, medical history, and the medical agent's diagnosis and treatment. The structure and content of the rest of the conversation were the same.

Each participant read 2 stimulus stories. The first reading was a brief conversation in which the medical agent made a diagnosis and offered treatment options but did not explain the rationale for the diagnosis. The second reading was an extended conversation in which the medical agent explained the diagnostic decision-making process in more detail, including the analysis of the symptoms, the clarification of potential diagnoses, the reasons why a specific diagnosis was made, and the analysis of treatment options and relevant medications. At the end of each reading, participants completed a questionnaire including manipulation checks and dependent variables.

Dependent Measures

Functional measures of the medical consultation address how participants perceive the medical agent within its intended role [23]. Informed by the findings of Lew and Walther [38] and Appelman and Sundar [39], we explored 3 functional aspects of communication between the medical agent and the patient. *Source credibility* (mean 4.66, SD 0.88; Cronbach α=0.73) was measured with 4 items. Participants rated whether the source was trustworthy, competent, knew a lot, and had goodwill [38,40]. *Message credibility* (mean 4.61, SD 0.75; Cronbach α=0.81) was assessed using the 3-item message credibility scale [39], examining whether the messages were accurate, authentic, and believable. *Persuasive effect* (mean 4.45, SD 1.16; Cronbach α=0.83) was measured with 5 items [41], examining the extent to which participants would accept the diagnosis, follow the treatment advice, take the medication, seek more information from the medical agent in the future, and schedule another appointment with the agent if necessary. All items were rated on a 7-point scale (1=strongly disagree and 7=strongly agree).

Relational measures of the diagnostic experience pertain to how participants perceive the medical agent in relation to themselves [23]. Consistent with the findings of Croes and Antheunis [42] and Ahn et al [43], this study examined 3 relational aspects of medical agent–patient communication. *Perceived empathy* (mean 4.07, SD 1.34; Cronbach α=0.81) was measured with 5 statements (eg, "The medical agent came across as empathic"; 1=strongly disagree and 7=strongly agree) [42]. *Self-disclosure* (mean 4.35, SD 1.78; Cronbach α=0.87) was assessed using 4 statements [44], such as "I feel comfortable disclosing personal information when interacting with the medical agent" (1=strongly disagree and 7=strongly agree). *Psychological distance* (mean 2.99, SD 1.59) was measured with the single item of Inclusion of Other in the Self Scale [45]. Participants chose the picture that best represented their relationship with the medical agent (1=farthest and 7=closest).

Participants' emotional responses to the medical consultation were examined by asking, "Do the following statements describe your feelings after the conversation with the medical agent?" We focused on 2 emotional responses that had opposing valences and were most relevant in the context of medical consultations: feeling optimistic and feeling guilty. Each response was measured with a single item rated on 7 points (1=no such emotion and 7=the emotion is extremely strong).

Participants reported demographic information at the end of the experiment. Sex was entered as a covariate in data analysis.

Results

Manipulation Check

As a manipulation check, participants answered three questions about their agreement with the statements after exposure to the assigned stimulus: (1) the medical agent is an autonomous AI, (2) the disease is socially stigmatized, and (3) the diagnostic explanation is provided. Autonomous AI (mean 2.71, SD 1.33) was rated as more autonomous than human-involved AI (mean 3.20, SD 1.06) and the human physician (mean 5.52, SD 1.75; $F_{2,240}$=91.37; P<.001). AIDS (mean 5.17, SD 0.79) was rated as more stigmatized than heart disease (mean 2.68, SD 1.08; t_{220}=20.69; P<.001). The second scenario (with diagnosis explanation; mean 6.92, SD 0.70) was rated as having more explanation than the first scenario (no explanation; mean 1.65, SD 1.87; t_{219}=–37.99; P<.001). The results indicated effective manipulation of consultation sources, health-related stigma, and diagnosis explanation.

Testing the Influence of the Source of Medical Consultation

RQ1 asked about the influence of consultation sources on the functional, relational, and emotional evaluations of the medical consultation. Regarding the functional evaluations, the main effect of consultation source emerged for source credibility ($F_{2,240}$=7.75; P=.001; partial η^2=0.06), message credibility ($F_{2,240}$=9.52; P<.001; partial η^2=0.07), and persuasive effect ($F_{2,240}$=10.78; P<.001; partial η^2=0.08). Post hoc tests (Bonferroni) indicated that diagnoses made by AI and by human-involved AI had significantly lower source credibility, message credibility, and persuasive effect as compared to those made by humans, but no significant differences were noted between the ratings of AI and human-involved AI (Table 1). Concerning the relational ratings, there were no differences in

XSL•FO
RenderX

perceived empathy ($F_{2,240}$=1.75; P=.18), self-disclosure ($F_{2,240}$=0.63; P=.54), or psychological distance ($F_{2,240}$=2.65; P=.07) between the 3 medical agents. The main effect of the consultation source did not exist for the 2 emotional responses: feeling optimistic ($F_{2,240}$=0.34; P=.71) and feeling guilty ($F_{2,240}$=1.04; P=.35).

Table 1. The main effect of consultation sources on the functional, relational, and emotional evaluations of the medical consultation.

	AI[a], mean (SD)	Human-involved AI, mean (SD)	Human, mean (SD)	*F* test (*df*)	*P* value	Bonferroni test[b]
Functional evaluations						
Source credibility	4.55 (0.06)	4.56 (0.06)	4.85 (0.06)	7.75 (2, 240)	.001	1<3, 2<3, and 1<2[c]
Message credibility	4.47 (0.07)	4.52 (0.07)	4.85 (0.07)	9.52 (2, 240)	<.001	1<3, 2<3, and 1<2[d]
Persuasive effect	4.34 (0.07)	4.39 (0.07)	4.78 (0.07)	10.78 (2, 240)	<.001	1<3, 2<3, and 1<2[e]
Relational evaluations						
Perceived empathy	4.06 (0.08)	3.97 (0.08)	4.17 (0.08)	1.75 (2, 240)	.18	NS[f]
Self-disclosure	4.28 (0.11)	4.30 (0.11)	4.44 (0.11)	0.63 (2, 240)	.54	NS
Psychological distance	2.82 (0.12)	2.94 (0.12)	3.20 (0.12)	2.65 (2, 240)	.07	NS
Emotional responses						
Optimistic	3.61 (0.1)	3.64 (0.1)	3.72 (0.1)	0.34 (2, 240)	.71	NS
Guilty	3.21 (0.1)	3.06 (0.1)	3.26 (0.1)	1.04 (2, 240)	.35	NS

[a]AI: artificial intelligence.

[b]1 indicates mean value of AI group, 2 indicates mean value of human-involved AI group, and 3 indicates mean value of human group.

[c]1<3 (P=.002), 2<3 (P=.003), and 1<2 (P=.99).

[d]1<3 (P<.001), 2<3 (P=.002), and 1<2 (P=.99).

[e]1<3 (P<.001), 2<3 (P=.001), and 1<2 (P=.99).

[f]NS: not significant.

Testing the Influence of Health-Related Stigma

RQ2 explored whether the stigma of the health issue (ie, AIDS vs heart disease) would directly influence or interact with the effect of consultation source on participants' evaluations of the diagnosis. Results demonstrated that there were no significant differences in source credibility ($F_{1,240}$=3.22; P=.07), message credibility ($F_{1,240}$=0.01; P=.94), persuasive effect ($F_{1,240}$=1.46; P=.23), perceived empathy ($F_{1,240}$=0.27; P=.6), self-disclosure ($F_{1,240}$=2.87; P=.09), or psychological distance ($F_{1,240}$=0.02; P=.89) between the 2 diseases, indicating that the functional and relational evaluations were not significantly influenced by health-related stigma. Regarding the emotional responses, the main effect of health-related stigma emerged for feeling guilty ($F_{1,240}$=44.55; P<.001; partial η^2=0.16) but not for feeling optimistic ($F_{1,240}$=1.42; P=.24). Post hoc tests (Bonferroni) revealed that the level of guilt was lower in the heart disease scenario (mean 2.79, SD 0.08) than in the AIDS scenario (mean 3.57, SD 0.08). The interaction effect of consultation source and health-related stigma did not appear for any of the dependent variables.

Testing the Influence of Diagnosis Explanation

RQ3 investigated whether providing an explanation of the diagnosis would enhance the evaluation of the medical consultation and whether there were differences among the 3 medical agents and among the 2 types of diseases. First, the main effect of diagnosis explanation emerged for source credibility ($F_{1,240}$=21.48; P<.001; partial η^2=0.08), message credibility ($F_{1,240}$=16.35; P<.001; partial η^2=0.06), persuasive effect ($F_{1,240}$=16.29; P<.001; partial η^2=0.06), perceived empathy ($F_{1,240}$=20.13; P<.001; partial η^2=0.08), self-disclosure ($F_{1,240}$=4.01; P=.04; partial η^2=0.02), and psychological distance ($F_{1,240}$=16.68; P<.001; partial η^2=0.07). Post hoc tests (Bonferroni) revealed that providing an explanation of the diagnosis significantly increased all the functional and relational evaluations of the medical consultation (Table 2). Regarding emotional responses, the main effect of diagnosis explanation emerged for feeling optimistic ($F_{1,240}$=15.27; P<.001; partial η^2=0.06) and feeling guilty ($F_{1,240}$=4.87; P=.03; partial η^2=0.02). Participants had significantly higher levels of optimism and lower levels of guilt when the diagnosis was explained (Table 2).

XSL•FO
RenderX

Table 2. The main effect of diagnosis explanation on the functional, relational, and emotional evaluations of the medical consultation.

	No explanation, mean (SD)	With explanation, mean (SD)	*F* test (*df*)	*P* value	Bonferroni test[a]
Functional evaluations					
Source credibility	4.2 (0.05)	5.12 (0.04)	21.48 (1, 240)	<.001	1<2
Message credibility	4.26 (0.05)	4.97 (0.04)	16.35 (1, 240)	<.001	1<2
Persuasive effect	4.07 (0.05)	4.94 (0.05)	16.29 (1, 240)	<.001	1<2
Relational evaluations					
Perceived empathy	3.36 (0.06)	4.77 (0.05)	20.13 (1, 240)	<.001	1<2
Self-disclosure	4.15 (0.07)	4.53 (0.07)	4.01 (1, 240)	.04	1<2
Psychological distance	2.3 (0.07)	3.67 (0.09)	16.68 (1, 240)	<.001	1<2
Emotional responses					
Optimistic	3.07 (0.06)	4.24 (0.07)	15.27 (1, 240)	<.001	1<2
Guilty	3.37 (0.07)	2.99 (0.06)	4.87 (1, 240)	.03	2<1

[a]1 indicates the mean value of no diagnostic explanation and 2 indicates the mean value of providing a diagnostic explanation.

Furthermore, a significant interaction effect appeared between consultation source and diagnosis explanation on participants' perceived empathy ($F_{2,240}$=8.71; P<.001; partial η^2=0.07) during the medical consultation. Participants did not differ in their perceived empathy for the 3 medical agents when no diagnosis explanation was provided (P=.15). However, when the diagnosis was explained, participants perceived significantly more empathy from the human physician (mean 5.08, SD 0.09) than from the AI (mean 4.59, SD 0.09; P=.001) and the human-involved AI (mean 4.65, SD 0.09; P=.003); however, there was no significant difference between the AI and the human-involved AI (P=.99).

Another interaction effect emerged between health-related stigma and diagnosis explanation on the evaluation of source credibility ($F_{1,240}$=6.02; P=.02; partial η^2=0.03). When no explanation of the diagnosis was provided, the AIDS group (mean 4.32, SD 0.07) rated the source as significantly more credible than the heart disease group (mean 4.07, SD 0.07; P=.01). However, when the diagnosis was explained, no mean difference was found between the 2 types of diseases (P=.97).

Discussion

Principal Findings

AI is considered a revolutionary technology in the health care sector and is used either independently or in collaboration with humans in medical diagnoses [2]. Nonetheless, the technological expectations of AI from the expert's perspective do not necessarily translate into an improved evaluation of medical services from the patient's perspective. This context led to the formulation of the RQs of the study. From the patient's perspective, can autonomous or human-involved AI replace the role of human physicians at the functional, relational, and emotional levels? How does health-related stigma and its associated need for dehumanization influence patients' evaluations of medical consultations and preferences for different medical agents? Does providing explanations of the diagnosis equally improve patients' evaluations of consultations with different medical agents?

The results showed that at the functional level, people were more likely to trust a human physician's diagnosis than either AI or human-involved AI. However, at the relational and emotional levels, the physician-patient interactions constructed by the 3 medical agents did not differ significantly. Stigma did not significantly affect the evaluation of the medical consultation or the preference for the medical agent, suggesting that in high-stakes medical scenarios, patients' instrumental need for diagnosis is prioritized over privacy concerns. Furthermore, explaining the diagnostic decision-making process significantly improved patients' functional, relational, and emotional evaluations of the consultations, implying that AI, human-involved AI, and human physicians should all prioritize providing understandable explanations to patients, which could enhance treatment adherence, strengthen physician-patient relations, and generate positive emotional responses.

Theoretical Contributions

First, this study contributes to the understanding of HCI and HMC in a medical scenario through the empirical evaluation of 3 influential theories in this field that are divergent in nature, namely, the CASA paradigm, the machine heuristic theory, and the algorithm aversion theory. Previous studies have explored HCI and HMC mostly by analyzing scenarios of web-based content production and moderation, consumer behavior, and so on [5,6]. Which theory best captures the intricacies of the medical scenario remains unclear, especially when considering contextual factors, such as the severe consequences of medical interventions and the attributes of the disease. Moreover, little evidence exists to test, from a user-centered perspective, the perception of human-involved AI, especially vis-à-vis AI and humans; however, this model of human-AI collaboration is increasingly prevalent in medical scenarios. This study found that functionally, people trusted the diagnosis of a human physician more than AI or human-involved AI, and there was no difference between autonomous AI and human-involved AI,

XSL•FO
RenderX

supporting the algorithm aversion theory. This may be due to the opacity of AI systems, as evidenced in RQ3 through the observation that explaining the decision-making process significantly enhanced the perceived credibility and persuasiveness of the diagnosis. Another potential explanation is that, at the current stage of AI development, people place greater trust in human expertise than in accurate AI, particularly in decisions that are traditionally made by humans and that carry severe consequences, such as disease diagnosis. Regarding the relational and emotional aspects, no significant difference was found between human-computer (ie, medical consultation with AI or human-involved AI) and human-human (ie, medical consultation with human) interactions. This may be due to a general lack of care on the part of human physicians during medical consultations and AI's increasing emphasis on creating anthropomorphic relationships.

Second, this study is among the early research endeavors to incorporate the issue of health-related stigma into the evaluation of human-AI interaction [7,8]. It explored whether dehumanized medical consultations were preferred for patients with stigmatized conditions, who might require more anonymity and privacy, or whether human expertise and human touch were preferred in consultations for all diseases. We found that stigma did not significantly influence how users evaluated the consultations or preferred a particular medical agent. This implies that in the highly consequential medical context, the instrumental need for diagnosis outweighs privacy concerns. In the context of illness, patients are keen to receive treatment and less likely to prefer a medical agent simply because of the different levels of privacy exposure that exist between human-AI interaction and human-human interaction. However, including the diagnosis explanation in the analysis (RQ3) revealed an interaction effect between health-related stigma and diagnosis explanation on people's source credibility ratings. We found that stigma had no significant effect on perceived source credibility when the diagnosis was explained in detail, suggesting that patients are likely to judge source credibility based on the diagnostic explanation when it is provided. However, when patients were only given the name of the disease and treatment options but not an explanation of why this decision was made, more stigma associated with the disease led to higher perceived source credibility from the patient. The aforementioned findings collectively indicate that while overall stigma does not significantly influence individual judgment, when patients are provided solely with the name of the disease and treatment options without an explanation of the diagnostic decision-making process, patients experiencing stigmatized diseases are susceptible to the psychology of shame and the eagerness to find solutions without demanding transparency from the medical agent. This is regardless of whether the agent is an AI, a human-involved AI, or a human physician.

Third, we advanced health communication literature by exploring, from a patient-centered perspective, a potential approach to improving patient-physician communication, which is providing explanations for the diagnostic decision-making process. We found that including a diagnosis explanation significantly improved patients' evaluations of the consultation at functional, relational, and emotional levels and for all 3 medical agents. This suggests that regardless of the identity of the communicator (AI, human-involved AI, or human), it is not enough for a medical agent to have the expertise (eg, human physicians) or to be highly accurate (eg, AI and human-involved AI). The explanation for the diagnostic decision-making process is critical for patients to improve treatment adherence, strengthen the physician-patient relationship, and build positive emotions throughout the consultation. On the one hand, this sheds light on how to facilitate traditional physician-patient communication, as evidence showed that patients sometimes had no choice but to turn to web-based communities to decipher the medical knowledge and reasoning embedded in the medical disclosure documents [34]. More importantly, it offers insights into the ways to design an AI-powered medical system, which has long been criticized for lacking transparency in making highly consequential decisions. In addition, an interaction effect appeared between diagnosis explanation and consultation source on perceived empathy. When there was no explanation, participants did not differ in their perceived empathy for the 3 medical agents, but when an explanation was provided, participants perceived significantly more empathy from the human physician than from AI or human-involved AI. This suggests that although unraveling the decision-making process is helpful in building effective medical agent–patient communication, it is still easier for human physicians to establish closer and more empathetic interactions with patients than AI-based systems when the same efforts of diagnosis explanation were made.

Practical Implications

This study contributes to the understanding of the roles of humans and AI in medical diagnosis and identifies key principles for the future application and design of AI medical agents. First, we found that people generally viewed human physicians as more credible and easier to relate to than AI or human-involved AI when the same effort was made (ie, providing the same diagnosis and the same explanation). Even for conditions where we would expect privacy and anonymity to be required (eg, stigma-related diseases), no significant preference for AI-based systems over humans emerged. This suggests that people still prefer human expertise and are more easily moved by human-human interactions in medical consultations. Thus, it is vital to ensure that human physicians assume a pivotal position in medical diagnosis to foster trust and empathy between physicians and patients. As the health care industry progresses toward embracing AI in medical diagnosis, it is advisable to consider incorporating more human elements in the diagnostic process to gain greater user acceptance.

In addition, we propose that the design of AI-based medical agents should be user centered, with the objective of satisfying the user's need to understand their own conditions and treatment options rather than merely providing conclusions that are accurate but difficult to comprehend. Medical diagnosis has historically prioritized the functional objective of furnishing accurate diagnoses and treatments. Consequently, the design of numerous AI-powered diagnostic applications has centered on enhancing diagnostic precision through processes such as data learning and algorithm iteration. However, it is insufficient for AI to simply be accurate and efficient to be trusted and embraced

XSL•FO
RenderX

by patients. From the patient's or user's perspective, an inability to comprehend the diagnosis and link their symptoms to the treatment plan may result in reluctance to adhere strictly to the treatment to achieve desired health outcomes. In examining RQ3, we found that opening the black box of the diagnostic decision-making process and providing explanations to help patients understand their health conditions effectively improved their functional, relational, and emotional evaluations of medical consultations. Therefore, designers of AI-powered systems should place greater focus on taking users' perspectives, opening the black box of decision-making processes, and ensuring that the diagnostic process is understandable to patients. This will not only enhance patient satisfaction and treatment outcomes at the individual level and in health care scenarios, but it also has practical implications for AI-based product design in a broader sense.

Limitations and Future Research

First, experiment research has its limitations. Because the scenarios are simulated, some participants may react differently in real-life situations. For instance, a distinction may emerge between an individual genuinely at risk of AIDS and a participant simulating such a condition. People who are genuinely at risk may be more likely than normal participants to prefer dehumanization in medical assistance because of their tendency to withhold real private information such as risky sexual behaviors [27]. However, another possibility is that those facing imminent peril may prioritize their survival over privacy concerns, opting to seek help from authoritative human physicians [46], whereas, in role-playing situations, participants are only imagining the stigma but not experiencing actual life threats. Although these 2 opposing situations may counteract the effect of role-playing, future research could use qualitative methods, such as ethnography and interviews, to further elucidate the intricacies of patient psychology and interpret how they may influence the perception of different medical agents.

Second, this research focused on the effects of consultation sources, health-related stigma, and explanation of diagnosis on patients' evaluations of medical consultations. We believe that these are highly important variables in the evaluation of medical consultations, but other factors, such as disease severity, may also interact with the aforementioned variables and shape patients' evaluations of medical consultations. The inclusion of more variables, such as disease severity, in the analyses in future studies will provide more in-depth and detailed insights.

Furthermore, while the experimental scenario in this study is medical diagnosis, AI has many other applications in health care, such as health risk analysis and disease screening. Future studies could investigate whether attitudes toward AI vary across different application scenarios.

Conclusions

With the increasing application of AI in the medical field to provide diagnosis, either independently or in collaboration with humans, this study aims to examine, from the patient's perspective, whether AI or human-involved AI can replace the role of human physicians in functional, relational, and emotional dimensions. In addition, we aimed to investigate how health-related stigma and explanations of diagnoses affect patients' evaluation of the medical consultation and their preferences for the medical agent. Findings indicated that individuals exhibited greater trust in diagnoses rendered by a human physician than by AI or human-involved AI, but at the relational and emotional levels, there was no significant difference between the human-AI and human-human interactions. Stigma did not have a significant impact on how people evaluated the medical consultation or contribute to preferring AI-powered systems over humans. However, providing explanations of the diagnosis significantly improved patients' functional, relational, and emotional evaluations of the consultation for all 3 medical agents. By comparing people's perceptions of AI, human-involved AI, and human physicians in medical diagnosis, this study contributed to the understanding of human-AI interaction in the medical scenario and empirically tested 3 influential but divergent theories in the field (CASA, machine heuristic theory, and algorithm aversion theory). Furthermore, we advanced health communication studies by incorporating health-related stigma into the evaluation of human-AI interaction, as well as by demonstrating that explaining the diagnostic decision-making process is effective in improving patient-physician communication.

Acknowledgments

This research was financially supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (project number 23XNL016).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

WG was involved in the conceptualization, methodology, writing the original draft, and reviewing and editing the manuscript. YC was involved in the conceptualization, methodology, project administration, and supervision of the study.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
The stimulus materials and the questionnaire used for the experiment.
[DOCX File , 401 KB - jmir_v27i1e66760_app1.docx]

References

1. Rawal J. Digital health market. Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/industry-reports/digital-health-market-100227 [accessed 2024-09-10]
2. Formosa P, Rogers W, Griep Y, Bankins S, Richards D. Medical AI and human dignity: contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. Comput Human Behav 2022 Aug;133:107296. [doi: 10.1016/j.chb.2022.107296]
3. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021 Apr 22;23(4):e25759 [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
4. Alam L, Mueller S. Examining the effect of explanation on satisfaction and trust in AI diagnostic systems. BMC Med Inform Decis Mak 2021 Jun 03;21(1):178 [FREE Full text] [doi: 10.1186/s12911-021-01542-6] [Medline: 34082719]
5. Cheng X, Zhang X, Cohen J, Mou J. Human vs. AI: understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. Inf Process Manag 2022 May;59(3):102940. [doi: 10.1016/j.ipm.2022.102940]
6. Molina MD, Sundar SS. Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. New Media Soc 2022 Jun 23;26(6):3638-3656. [doi: 10.1177/14614448221103534]
7. Duggan J. Using TikTok to teach about abortion: combatting stigma and miseducation in the United States and beyond. Sex Educ 2022 Jan 21;23(1):81-95. [doi: 10.1080/14681811.2022.2028614]
8. Van Brakel WH. Measuring health-related stigma--a literature review. Psychol Health Med 2006 Aug;11(3):307-334. [doi: 10.1080/13548500600595160] [Medline: 17130068]
9. Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. Artif Intell Med 2013 Jan;57(1):9-19. [doi: 10.1016/j.artmed.2012.12.003] [Medline: 23287490]
10. Weetman K, Dale J, Scott E, Schnurr S. Adult patient perspectives on receiving hospital discharge letters: a corpus analysis of patient interviews. BMC Health Serv Res 2020 Jun 15;20(1):537 [FREE Full text] [doi: 10.1186/s12913-020-05250-1] [Medline: 32539716]
11. Milano S, Taddeo M, Floridi L. Recommender systems and their ethical challenges. AI Soc 2020 Feb 27;35(4):957-967 [FREE Full text] [doi: 10.1007/s00146-020-00950-y]
12. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of self-diagnosis health chatbots in real-world settings: case study. J Med Internet Res 2021 Jan 06;23(1):e19928 [FREE Full text] [doi: 10.2196/19928] [Medline: 33404508]
13. Aoki N. The importance of the assurance that "humans are still in the decision loop" for public trust in artificial intelligence: Evidence from an online experiment. Comput Human Behav 2021 Jan;114:106572. [doi: 10.1016/j.chb.2020.106572]
14. Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' perceptions toward human-artificial intelligence interaction in health care: experimental study. J Med Internet Res 2021 Nov 25;23(11):e25856 [FREE Full text] [doi: 10.2196/25856] [Medline: 34842535]
15. Nass C, Moon Y. Machines and mindlessness: social responses to computers. J Soc Issue 2002 Dec 17;56(1):81-103. [doi: 10.1111/0022-4537.00153]
16. Gambino A, Fox J, Ratan R. Building a stronger CASA: extending the computers are social actors paradigm. Hum Mach Commun 2020 Feb 1;1:71-86. [doi: 10.30658/hmc.1.5]
17. Sundar SS, Kim J. Machine heuristic: when we trust computers more than humans with our personal information. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Scotland, UK p. 1-9 URL: https://dl.acm.org/doi/abs/10.1145/3290605.3300768 [doi: 10.1145/3290605.3300768]
18. Araujo T, Helberger N, Kruikemeier S, de Vreese CH. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 2020 Jan 01;35(3):611-623. [doi: 10.1007/s00146-019-00931-w]
19. Chew HS, Achananuparp P. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. J Med Internet Res 2022 Jan 14;24(1):e32939 [FREE Full text] [doi: 10.2196/32939] [Medline: 35029538]
20. Logg JM, Minson JA, Moore DA. Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process 2019 Mar;151:90-103. [doi: 10.1016/j.obhdp.2018.12.005]
21. Thurman N, Moeller J, Helberger N, Trilling D. My friends, editors, algorithms, and I: examining audience attitudes to news selection. Digit Journal 2018 Oct 16;7(4):447-469. [doi: 10.1080/21670811.2018.1493936]
22. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 2015 Feb;144(1):114-126. [doi: 10.1037/xge0000033] [Medline: 25401381]
23. Guzman AL, Lewis SC. Artificial intelligence and communication: a human–machine communication research agenda. New Media Soc 2019 Jul 04;22(1):70-86. [doi: 10.1177/1461444819858691]
24. Yokoi R, Eguchi Y, Fujita T, Nakayachi K. Artificial intelligence is trusted less than a doctor in medical treatment decisions: influence of perceived care and value similarity. Int J Hum Comput Interact 2020 Dec 26;37(10):981-990. [doi: 10.1080/10447318.2020.1861763]

XSL•FO
RenderX

25. De Togni G, Erikainen S, Chan S, Cunningham-Burley S. What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care. Soc Sci Med 2021 May;277:113874 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113874] [Medline: 33901725]
26. Pescosolido BA, Martin JK. The stigma complex. Annu Rev Sociol 2015 Aug 14;41(1):87-116 [FREE Full text] [doi: 10.1146/annurev-soc-071312-145702] [Medline: 26855471]
27. Branley-Bell D, Brown R, Coventry L, Sillence E. Chatbots for embarrassing and stigmatizing conditions: could chatbots encourage users to seek medical advice? Front Commun 2023 Sep 27;8:1275127. [doi: 10.3389/fcomm.2023.1275127]
28. Saki M, Mohammad Khan Kermanshahi S, Mohammadi E, Mohraz M. Perception of patients with HIV/AIDS from stigma and discrimination. Iran Red Crescent Med J 2015 Jun 23;17(6):e23638. [doi: 10.5812/ircmj.236381v2]
29. Palmer A, Schwan D. Beneficent dehumanization: employing artificial intelligence and carebots to mitigate shame-induced barriers to medical care. Bioethics 2022 Feb 23;36(2):187-193. [doi: 10.1111/bioe.12986] [Medline: 34942057]
30. Nadarzynski T, Puentes V, Pawlak I, Mendes T, Montgomery I, Bayley J, et al. Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: a qualitative analysis. Sex Health 2021 Nov;18(5):385-393. [doi: 10.1071/SH21123] [Medline: 34782055]
31. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
32. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]
33. Larasati R. Trust and explanation in artificial intelligence systems: a healthcare application in disease detection and preliminary diagnosis. The Open University. 2023. URL: https://oro.open.ac.uk/88778/ [accessed 2025-01-20]
34. Kiełkiewicz-Janowiak A, Zabielska M. Please explain what this diagnosis means because I don't quite understand: patients seeking medical expertise and support in online dialogue to deconstruct hospital discharge documents. Lang Dialogue 2024 Jun 13;14(2):214-238. [doi: 10.1075/ld.00170.kie]
35. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020 Jun;58:82-115. [doi: 10.1016/j.inffus.2019.12.012]
36. Bussone A, Stumpf S, O'Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. In: Proceedings of the 2015 International Conference on Healthcare Informatics. 2015 Presented at: ICHI '15; October 21-23, 2015; Dallas, TX p. 160-169 URL: https://ieeexplore.ieee.org/document/7349687 [doi: 10.1109/ichi.2015.26]
37. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare. New York, NY: Academic Press; 2020:295-336.
38. Lew Z, Walther JB. Social scripts and expectancy violations: evaluating communication with human or AI chatbot interactants. Media Psychol 2022 Jun 10;26(1):1-16. [doi: 10.1080/15213269.2022.2084111]
39. Appelman A, Sundar SS. Measuring message credibility: construction and validation of an exclusive scale. Journal Mass Commun Q 2015 Oct 05;93(1):59-79. [doi: 10.1177/1077699015606057]
40. Shi W, Rothman AJ, Yzer MC, Nagler RH. Effects of exposure to conflicting information about mammography on cancer information overload, perceived scientists' credibility, and perceived journalists' credibility. Health Commun 2023 Oct 23;38(11):2481-2490. [doi: 10.1080/10410236.2022.2077163] [Medline: 35607276]
41. Weber R, MacKenzie WI, Lanius C. The impact of source credibility and risk perception attitudes on Americans' willingness to participate in contact tracing applications. J Appl Commun Res 2022 Nov 13;51(3):283-301. [doi: 10.1080/00909882.2022.2143274]
42. Croes EA, Antheunis ML. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. J Soc Pers Relatsh 2020 Sep 25;38(1):279-300. [doi: 10.1177/0265407520959463]
43. Ahn J, Kim J, Sung Y. AI-powered recommendations: the roles of perceived similarity and psychological distance on persuasion. Int J Advert 2021 Oct 26;40(8):1366-1384. [doi: 10.1080/02650487.2021.1982529]
44. Ledbetter AM. Measuring online communication attitude: instrument development and validation. Commun Monogr 2009 Dec;76(4):463-486. [doi: 10.1080/03637750903300262]
45. Aron A, Aron EN, Smollan D. Inclusion of other in the self scale and the structure of interpersonal closeness. J Pers Soc Psychol 1992 Oct;63(4):596-612. [doi: 10.1037/0022-3514.63.4.596]
46. Miles O, West R, Nadarzynski T. Health chatbots acceptability moderated by perceived stigma and severity: a cross-sectional survey. Digit Health 2021 Dec 08;7:20552076211063012 [FREE Full text] [doi: 10.1177/20552076211063012] [Medline: 34917391]

Abbreviations

AI: artificial intelligence
CASA: computers are social actors
HCI: human-computer interaction

XSL•FO
RenderX

HMC: human-machine communication
RQ: research question

Edited by A Coristine; submitted 22.09.24; peer-reviewed by S Pan, JH Kim; comments to author 17.12.24; revised version received 25.01.25; accepted 31.01.25; published 05.03.25.

Please cite as:
Guo W, Chen Y
Investigating Whether AI Will Replace Human Physicians and Understanding the Interplay of the Source of Consultation, Health-Related Stigma, and Explanations of Diagnoses on Patients' Evaluations of Medical Consultations: Randomized Factorial Experiment
J Med Internet Res 2025;27:e66760
URL: https://www.jmir.org/2025/1/e66760
doi:10.2196/66760
PMID:40053785

©Weiqi Guo, Yang Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Augmenting Insufficiently Accruing Oncology Clinical Trials Using Generative Models: Validation Study

Samer El Kababji[1,2], PhD; Nicholas Mitsakakis[2], MSc, PhD; Elizabeth Jonker[2], BA; Ana-Alicia Beltran-Bless[3], MD; Gregory Pond[4], PhD; Lisa Vandermeer[3], MSc; Dhenuka Radhakrishnan[2], MD; Lucy Mosquera[5], MSc; Alexander Paterson[6], MD; Lois Shepherd[7], MDCM, FRCP; Bingshu Chen[7], PhD; William Barlow[8], PhD; Julie Gralow[9], MD; Marie-France Savard[3], MD; Christian Fesl[10], PhD; Dominik Hlauschek[10], MSc; Marija Balic[11], MD, PhD; Gabriel Rinnerthaler[11], MD; Richard Greil[12], MD; Michael Gnant[13], MD; Mark Clemons[3], MD; Khaled El Emam[1,2], BEng, PhD

[1]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[2]CHEO Research Institute, Ottawa, ON, Canada
[3]Division of Medical Oncology, University of Ottawa, Ottawa, ON, Canada
[4]Department of Oncology, McMaster University, Hamilton, ON, Canada
[5]Generate Ops & Data Science, Aetion, Ottawa, ON, Canada
[6]Oncology, Alberta Health Services, Edmonton, AB, Canada
[7]Public Health Sciences, Queens University, Kingston, ON, Canada
[8]Biostatistics, University of Washington, Seattle, WA, United States
[9]Medical Oncology, University of Washington, Seattle, WA, United States
[10]Clinical Statistics, Austrian Breast & Colorectal Cancer Study Group (ABCSG), Vienna, Austria
[11]Division of Clinical Oncology, Medical University Graz, Graz, Austria
[12]Paracelsus Medical University Salzburg, Salzburg, Austria
[13]Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria

Corresponding Author:
Khaled El Emam, BEng, PhD
School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa
75 Laurier Ave E
Ottawa, ON, K1N 6N5
Canada
Phone: 1 6137975412
Email: kelemam@ehealthinformation.ca

Abstract

Background: Insufficient patient accrual is a major challenge in clinical trials and can result in underpowered studies, as well as exposing study participants to toxicity and additional costs, with limited scientific benefit. Real-world data can provide external controls, but insufficient accrual affects all arms of a study, not just controls. Studies that used generative models to simulate more patients were limited in the accrual scenarios considered, replicability criteria, number of generative models, and number of clinical trials evaluated.

Objective: This study aimed to perform a comprehensive evaluation on the extent generative models can be used to simulate additional patients to compensate for insufficient accrual in clinical trials.

Methods: We performed a retrospective analysis using 10 datasets from 9 fully accrued, completed, and published cancer trials. For each trial, we removed the latest recruited patients (from 10% to 50%), trained a generative model on the remaining patients, and simulated additional patients to replace the removed ones using the generative model to augment the available data. We then replicated the published analysis on this augmented dataset to determine if the findings remained the same. Four different generative models were evaluated: sequential synthesis with decision trees, Bayesian network, generative adversarial network, and a variational autoencoder. These generative models were compared to sampling with replacement (ie, bootstrap) as a simple alternative.

XSL•FO
RenderX

Replication of the published analyses used 4 metrics: decision agreement, estimate agreement, standardized difference, and CI overlap.

Results: Sequential synthesis performed well on the 4 replication metrics for the removal of up to 40% of the last recruited patients (decision agreement: 88% to 100% across datasets, estimate agreement: 100%, cannot reject standardized difference null hypothesis: 100%, and CI overlap: 0.8-0.92). Sampling with replacement was the next most effective approach, with decision agreement varying from 78% to 89% across all datasets. There was no evidence of a monotonic relationship in the estimated effect size with recruitment order across these studies. This suggests that patients recruited earlier in a trial were not systematically different than those recruited later, at least partially explaining why generative models trained on early data can effectively simulate patients recruited later in a trial. The fidelity of the generated data relative to the training data on the Hellinger distance was high in all cases.

Conclusions: For an oncology study with insufficient accrual with as few as 60% of target recruitment, sequential synthesis can enable the simulation of the full dataset had the study continued accruing patients and can be an alternative to drawing conclusions from an underpowered study. These results provide evidence demonstrating the potential for generative models to rescue poorly accruing clinical trials, but additional studies are needed to confirm these findings and to generalize them for other diseases.

(J Med Internet Res 2025;27:e66821) doi:10.2196/66821

KEYWORDS

generative models; study accrual; recruitment; clinical trial replication; oncology; validation; simulated patient; simulation; retrospective; dataset; patient; artificial intelligence; machine learning

Introduction

Background

Recruiting a sufficient number of patients for clinical trials is challenging [1], and the inability to recruit participants is the cause of failure for many clinical trials [2]. Approximately, 25% of clinical trials are discontinued before completion [3], with insufficient recruitment being the most frequent reason in 31% of the cases [4]. For adult cancer trials, between 20% and 50% fail to complete or were unable to reach recruitment goals [5-9]. This has been exacerbated by the recent pandemic where many trials experienced a considerable reduction in recruitment rates [10-13], which has continued after the pandemic [12]. While poor accrual is a problem in all trials, it is a greater problem in government (ie, academic) sponsored trials [14,15].

When a study is unable to recruit a sufficient number of patients, the study can be stopped, and the relevant analyses are performed on the available data. However, not reaching accrual targets results in underpowered analyses, and the smaller sample sizes increase the risk of unstable parameter estimates.

Patients have an expectation that their trial participation will lead to some advancement in knowledge that can be beneficial to the community [16], but many enroll in trials that do not answer the primary question adequately [15]. They are, therefore, enrolled in a study and exposed to toxicity and additional costs, with limited scientific benefit, which is considered unethical [17]. In addition to the wasted resources, it also means that those resources were not used for other studies that could have produced useful results.

Data augmentation is one approach to address insufficient accrual by either using real-world data (RWD) or by simulating additional observations.

RWD can be used for matched controls [18] where patient data from external sources are used instead of recruiting patients to the trial itself. In such a case, previous similar trials, registries, or eHealth record datasets on patients under the standard of care are matched to the treatment arm patients, and the matched patients' data are used as the control arm [19,20]. Such an approach with external controls has been used for running single-arm oncology trials [21,22]. However, external controls are challenging for a number of reasons [18]. First, the patients from RWD may have different observed and unobserved characteristics than the treatment arm patients, despite the use of matching. Second, unaccounted for environmental factors, such as seasonal effects, may lead to outcome differences. Third, changes in medical practice may have occurred over time and since the external control data were collected. Fourth, there may be measurement differences between the treatment and external controls resulting in the pooling of incompatible datasets, even for objective metrics. Fifth, if there are no adequate matches in the external data for some of the treatment arm patients, then these treatment arm patients may need to be dropped resulting in loss of valuable data. Sixth, the outcome variables need to be available in the external control dataset to allow a comparison, which can be challenging for surrogate end points or patient-reported outcomes. Finally, and specific to our context, insufficient accrual would occur in all arms in a study and not only in the controls; therefore, external controls would not address our problem.

Augmentation through simulation can be a potential solution when there are insufficient data and is a common practice for imaging data [23-25] and time series data [26,27]. In the case of cross-sectional RWD, augmentation methods, such as sampling with replacement (henceforth referred to as *bootstrap*), and generative models, such as sequential synthesis using decision trees, generative adversarial networks (GANs), and variational autoencoders (VAEs), have been evaluated with encouraging results [28-32]. Augmentation methods have also been applied to small clinical trial datasets as a first step in synthetic data generation [33,34]. One study used a VAE

XSL•FO
RenderX

generative model to simulate additional patients as a mechanism to design smaller clinical trials [35].

Objective

In this study, we therefore adopt augmentation through simulation and expand on this body of work by more comprehensively evaluating multiple types of generative models and on a larger number of clinical trials. We make 2 hypotheses that we evaluate in this work as follows: (1) patients recruited early in a clinical trial are similar on the treatment effect to patients recruited later in a clinical trial and (2) because of hypothesis 1, we can train a generative model on early patients to simulate the remaining patients in insufficiently accruing trials to reach target recruitment and replicate the results of the original study that reached target recruitment.

Specifically, generative models are used to augment breast cancer clinical trials that do not reach target recruitment. We start with datasets from 9 completed breast cancer clinical trials and simulate different levels of insufficient accrual, and in each case, use generative machine learning models to simulate patients to compensate for the insufficient accrual. We then replicate the analyses of the published studies using the augmented datasets to determine if they produce similar findings as if the target number of patients were actually recruited.

Methods

Overview

Data augmentation methods using generative models were applied on 9 breast cancer clinical trial datasets. Insufficient accrual was simulated and augmentation then applied to compensate for that. The question was whether this can produce similar findings to the published analyses with the full data.

Datasets

The clinical trials that were included are summarized in Table 1, with further details described in Table S1 in Multimedia Appendix 1 [36-38]. The Rethinking Clinical Trials (REaCT) were supported by the REaCT program at the Ottawa Hospital [39]. The remaining datasets included a larger number of patients and multiple sites.

Table 2 shows the countries that the patients were recruited from. The studies spanned multiple jurisdictions in North America, Europe, Australia, and South Africa.

XSL•FO
RenderX

Table 1. Key features of the clinical trial datasets used in this study.

Dataset	NCT identifier	Participants, N	Variables in dataset[a], n	Strata info?	Control arm	Patients in the control arm, n (%)	Treatment arms considered	Patients in the treatment arm, n (%)	Enrollment period
REaCT[b]-ILI-AD	NCT02861859	218	8	True	Placebo	105 (48)	Active olanzapine	113 (52)	December 2016 to June 2019
REaCT-BTA[c]	NCT02721433	230	8	True	4 weekly BTA	118 (51)	12 weekly BTA	112 (49)	August 3, 2016, to June 5, 2018
CCTG[d] MA27	NCT00066573	7576	25	True	Anastrozole	3787 (50)	Exemestane	3789 (50)	June 2, 2003, to July 31, 2008
NSABP[e] B34	NCT00009945	3310	49	True	Placebo	1656 (50)	Clodronate	1654 (50)	January 22, 2001, to March 31, 2004
REaCT-G/G2	NCT02428114 and NCT02816164	401	10	True	7 or 10 d of granulocyte colony stimulating factor	248 (62)	5 d of granulocyte colony stimulating factor	153 (38)	May 2015 to September 2018
REaCT-HER2+[f]	NCT02632435	50	47	True	Peripherally inserted central catheter	26 (52)	PORT; totally implanted vascular access device	24 (48)	March 2016 to March 2018
ABCSG[g]-12	NCT00295646	1803	35	False	Tamoxifen[h]	900 (50)	Anastrozole[h]	903 (50)	1999 to 2006
REaCT-ZOL[i]	NCT03664687	211	11	False	—[j]	—	—	—	November 1, 2018, to April 2, 2020
SWOG[k] 0307[l]	NCT00127205	6018	23	False	Clodronate	2268 (38)	Zoledronic acid	2262 (38)	January 2006 to February 2010

[a]This is the total number of variables that were included in the generative models or bootstrap.

[b]REaCT: Rethinking Clinical Trials.

[c]BTA: bone-targeted agents.

[d]CCTG: Canadian Cancer Trials Group.

[e]NSABP: National Surgical Adjuvant Breast and Bowel Project.

[f]HER2+: human epidermal growth factor receptor-2 positive.

[g]ABCSG: Austrian Breast and Colorectal Cancer Study Group.

[h]Tamoxifen represents the arms of Nolvadex/control and Nolvadex/zoledronate in the clinical trial while anastrozole represents the arms of Arimidex/control and Arimidex/zoledronate.

[i]ZOL: zoledronate.

[j]Not applicable.

[k]SWOG: Southwest Oncology Group.

[l]Initially, the trial included 3 arms; however, only the indicated 2 arms had patients assigned to them throughout the duration of the study. Moreover, the available data for our study did not include any randomization codes, and as such, the original primary analysis comparing the outcomes between the 2 arms could not be replicated. Instead, we compared the 5-year survival probabilities between those with negative and positive or equivocal HER2 status. The estimate of the difference of survival probabilities and SE were produced.

XSL•FO
RenderX

Table 2. Countries of recruitment for the clinical trials used in this study.

Dataset	Countries of recruitment
REaCT[a]-ILIAD	Canada
REaCT-BTA[b]	Canada
CCTG[c] MA27	Australia, Canada, Hungary, Italy, Puerto Rico, South Africa, Switzerland, and the United States
NSABP[d] B34	United States
REaCT-G/G2	Canada
REaCT-HER2+[e]	Canada
ABCSG[f]-12	Austria and Germany
REaCT-ZOL[g]	Canada
SWOG[h]0307	Canada and the United States

[a]REaCT: Rethinking Clinical Trials.

[b]BTA: bone-targeted agents.

[c]CCTG: Canadian Cancer Trials Group.

[d]NSABP: National Surgical Adjuvant Breast and Bowel Project.

[e]HER2+: human epidermal growth factor receptor-2 positive.

[f]ABCSG: Austrian Breast and Colorectal Cancer Study Group.

[g]ZOL: zoledronate.

[h]SWOG: Southwest Oncology Group.

Ethical Considerations

This study was a secondary analysis of datasets from already completed clinical trials. The secondary analysis was approved by the Children's Hospital of Eastern Ontario Research Ethics Board (protocol: 23/47X) and the Ontario Cancer Research Ethics Board (project ID: 3749).

Nonmonotonic Treatment Effect Size Hypothesis

We will use the terms "early participants" to indicate participants who were recruited in the earlier stages of a study and "late participants" to indicate those who were recruited in later stages of a study. For early participants to be good candidates for training a generative model that can be used to simulate late participants, there should not be a systematic difference in the estimated effect size between these 2 groups.

Estimated effect sizes tend to vary as patients are recruited and converge to the true value as more information is collected [40]. Instability of estimates at small sample sizes is a contributing factor. This means that training a generative model on earlier patients may enable the simulation of realistic patients that are representative of those that would be recruited later in the trial if the effect over time is not systematic.

However, existing sites gain experience with conducting a trial and this may result in process adjustments along that learning curve that may have an impact on the outcome. This can also happen, for example, when there is treatment effect heterogeneity whereby some patient characteristic (eg, disease severity or age) interacts with the intervention and more patients at one end of the severity or age scale are recruited earlier in the study [41]. One simulation demonstrated a monotonic change in effect size as more patients are recruited [42]. In that example, earlier patients had a meaningfully different estimated effect size compared to later patients with a trend over time. In such a case, training a generative model on earlier patients may not produce simulated patients that are representative of the later patients.

Therefore, it is an empirical question whether such a monotonic effect can be observed in practice.

As can be seen in Table 1, our 9 studies were conducted over an extended period. If the studies were very short, then there is a higher likelihood that early and late participants would be similar. However, the extended enrollment periods suggest that there was ample opportunity for the characteristics of the participants to change over time, as well as adjustments to the trial processes to occur as study staff gain more experience with time.

To test the hypothesis that early and late participants are similar on the estimated effect size, the monotonic relationship of the effect size and order of participant recruitment was examined with a regression model of treatment and recruitment order main effects and their interaction. Point estimates and 95% CIs of the interaction term were obtained, indicating statistical significance at $P<.05$ if they do not include 0. This investigation was not conducted for the trials where the main analysis did not use a statistical model with treatment as predictor (ie, REaCT-zoledronate [ZOL], Southwest Oncology Group 0307, and REaCT–human epidermal growth factor receptor-2 positive [HER2+]). If the effect size was not monotonic, then the interaction term would not be statistically significant.

XSL•FO
RenderX

Generative Modeling Methods

Overview

We used 4 common machine learning–based generative modeling methods for structured tabular data to synthesize the analysis datasets from the clinical trials under investigation. These methods include sequential synthesis using decision trees, Bayesian networks, GAN, and VAE. The last 3 methods were implemented as an adaptation of an open-sourced Python package *Synthcity* [43]. Our implementation, which is publicly available (*Data Availability* section), provides further preprocessing and postprocessing on top of *Synthcity*. No further hyperparameter tuning was performed beyond what was available in the *Synthcity* implementation. For each generative model, the number of variables indicated in Table 1 were synthesized.

In addition, we used the "bootstrap" technique for comparison as a baseline. Bootstrap is simply sampling with replacement from the training data to add the missing patients.

Sequential Decision Trees

Similar to using a chaining method for multilabel classification problems, sequential decision trees generate synthetic data using conditional trees in a sequential fashion [44-46]. It has been commonly used in the health care and social-science domains for data synthesis [28,47-54]. The details of the implementation procedures are described elsewhere [44].

Bayesian Networks

Bayesian networks are models based on directed acyclic graphs that consist of nodes representing the random variables and arcs representing the dependencies among these variables. To construct the Bayesian networks model, the first step is to find the optimal network topology and then to estimate the optimal parameters [55]. Starting with a random initial network structure, the Hill Climb heuristic search is used to find the optimal structure. Then, the conditional probability distributions are estimated using the maximum a posteriori estimator [56]. Once the network structure and the parameters are estimated, we can initialize the nodes with no incoming arcs by sampling from their marginal distributions and predict the rest of the connected variables using the estimated parameters.

Conditional GAN

A basic GAN consists of 2 artificial neural networks, a generator and a discriminator [57]. The generator and the discriminator play a min-max game. The input to the generator is noise, while its output is synthetic data. The discriminator has 2 inputs: the real training data and the synthetic data generated by the generator. The output of the discriminator indicates whether its input is real or synthetic. The generator is trained to "trick" the discriminator by generating samples that look real. In contrast, the discriminator is trained to maximize its discriminatory capability.

Among all the variations of GAN architectures, the conditional tabular GAN (CTGAN) is often used in tabular data synthesis [58]. CTGAN builds on conditional GANs by addressing the multimodal distributions of continuous variables and the highly imbalanced categorical variables [59]. CTGAN solves the first problem by proposing a per-mode normalization technique. For the second problem, each category of a categorical variable serves as the condition passed to the GAN.

Variational Autoencoder

VAEs use artificial neural networks and involve 2 steps (ie, encoding and decoding) to generate new samples [60]. First, an encoder is generated to compress input data into a lower-dimensional latent space, in which the data points are represented by distributions. The second step is a decoding process, in which new data samples are reconstructed as output from the latent space. The neural network is optimized by minimizing the reconstruction loss between the output and the input. VAEs are known to generate complex data of various types due to their ability to learn more complex distributions and relationships [61]. Many variants have been proposed as an extension of VAE, such as triplet-based VAE [62], conditional VAE [63], and Gaussian VAE [64]. In particular, the tabular VAE was proposed as an adaptation of standard VAE to model and generate mixed-type tabular data with a modified loss function [59].

Augmentation

The augmentation procedure ensured that the augmented data had the same number of patients as the original clinical trial. Therefore, there is no difference in size between the published study and the augmented datasets used in our analyses.

Figure 1 illustrates the main steps of generating the augmented datasets for the trials given in Table 1. A trial's *original* dataset, with N number of patients, was first reduced by *r*. The variable *r* signifies the fraction of the *last* patients who were deliberately removed from the input dataset. This results in a *reduced* dataset with "(1-r) N" patients. In practice, the reduced dataset represents a poorly accruing clinical trial that needs to be rescued. The shaded area outlines the typical steps taken by a practitioner during the implementation process.

In our study, the *r* value was varied incrementally from 0.1 to 0.5 in steps of 0.1. For instance, a 0.2 *r* value indicates that the last 20% of patients were deliberately removed from the original trial dataset. The reduced dataset was then used to train a generative model. In case of bootstrapping, the reduced dataset was used for sampling with replacement. After training, the necessary samples for augmentation were generated from the trained model and concatenated onto the reduced dataset. To account for stochasticity, and as depicted by the dotted lines in the Figure 1, ten versions of the synthetic data were generated leading to multiple versions of the augmented datasets. Subsequently, each of the augmented datasets were analyzed in the same way as the published analysis. The 10 analyses results were then combined to obtain a single augmented dataset result, and that was compared with the results obtained from the original dataset (as described in the subsequent section).

XSL•FO
RenderX

Figure 1. Augmentation of clinical trial datasets using the generative model.

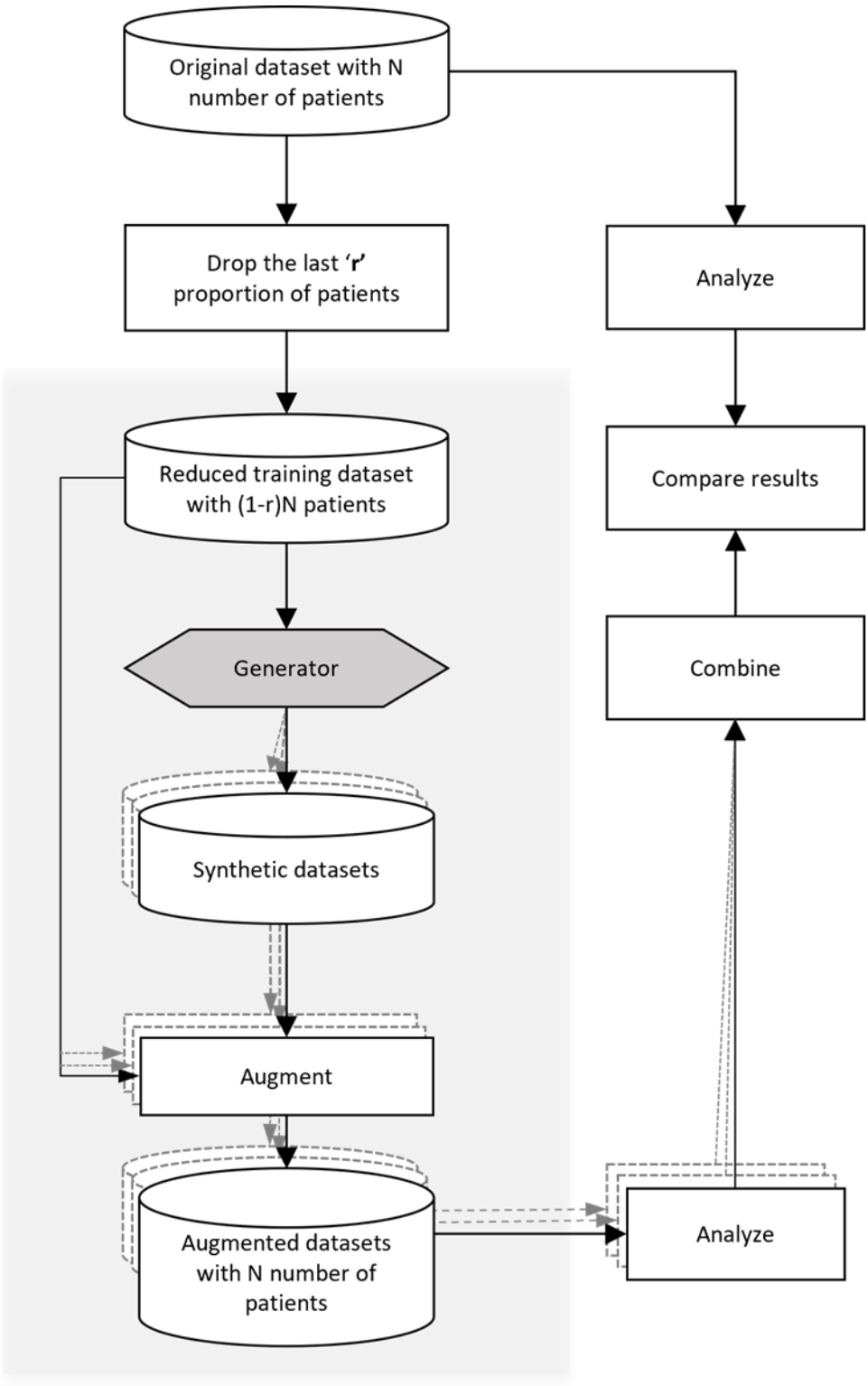

Some of the trials had implemented stratified randomization, and in those cases, we modified the basic process to accommodate stratification, for example, consider that as per the predefined protocol, the original dataset is stratified by 2 variables, the "Cancer Type" and the "Site Number." Rejection sampling was used to draw from the generated datasets to achieve the desired strata proportions along these 2 dimensions, in addition to the appropriate numbers of patients in each arm of the study.

Combining Rules

The original proposal for synthetic data generation treated it as a form of multiple imputation [65]. Under the multiple imputation model, multiple datasets, say *m*, are synthesized and analyzed. The combining rules are used to compute the parameter estimates and variances across the analysis results from the *m* synthetic datasets [66-68]. Such corrections for the parameter estimates and variances ensured that variability introduced by the generative process are accounted for when

XSL•FO
RenderX

estimating parameters and making population inferences from synthetic datasets.

Once the generative model was trained on the underlying distribution of the input data, it was used to create *m* synthetic datasets of size $r \times N$, and then each was added to that training dataset to augment it. This resulted in *m* versions of augmented datasets, each containing N records.

The *m* augmented datasets were analyzed using the same methodology applied to the original data in the relevant publications. This analysis yielded estimated parameters for each of the augmented versions. Subsequently, these estimated parameters were combined in accordance with the following partial synthesis rules.

For a particular model parameter q_i with variance v_i using synthetic dataset *i* where *i=1...m*. The adjustment for the model parameters and variances is as follows [52,68,69]. The combined model parameter  . is the mean across the m model parameters from the synthetic datasets  , and  is the mean variance across the *m* model parameters from the synthetic datasets where  . The between imputation variance is given by  , the adjusted variance is computed as  , and the adjusted large sample 95% CI of the model parameter is computed as  . For this study, we set *m*=10, which is consistent with current practice for the analysis of synthetic data [52,68-70] and has been recommended based on a recent simulation [71].

Augmentation Fidelity

For the generated datasets, we evaluated the fidelity relative to the training dataset. Fidelity indicates the extent to which the distributions of generated data deviate from the training data. For example, if *r*=0.1, then 90% of the dataset is used for training the generative model, and 10% is generated. We assessed the fidelity using the Hellinger distance [72], which has the advantage of being interpretable as it varies from 0 to 1. The Hellinger distance is averaged across the 10 generated datasets.

Evaluation of Study Replicability

We evaluated the replicability of the analysis results using the augmented datasets. Replicability is the reliability of findings when an existing study is repeated using the same analytical methods but different data [73]. We assessed it by comparing the published analysis results using the real datasets for these clinical trials with the results of the same analysis performed on the partially synthetic (ie, augmented) data. The details of the published analyses that were replicated are summarized in Multimedia Appendix 1.

For each clinical trial, a model was fitted to obtain a parameter estimate and its SE. For instance, if generalized estimating equations [36] were used for modeling the real data, an estimate would be the coefficient associated with the selected predictor as per the published study. We obtained a value for the estimate and its 95% CI. The same was applied to each of the *m* versions of the augmented data. We combined the results from all the augmented versions using the combining rules discussed above. Subsequently, the estimates and CIs of the original and augmented data were compared in terms of the *estimate agreement*, the *decision agreement*, *standardized difference*, and the *CI overlap*. These criteria have been used in the literature to assess the replicability of analyses using synthetic data [71,74]. The criteria are defined in Textbox 1.

Textbox 1. Criteria to assess the replicability of analyses using synthetic data.

- *Estimate agreement*: It is a Boolean indicator of whether the estimate produced by the augmented data is within the 95% CI produced by the real data. This requires that an augmented data effect estimate be within the range of plausible values for the true effect based on evidence from the real data. Under the assumption that the parameter variances are equal between the real and augmented datasets, estimate agreement is expected 83% of the time under no bias [75].
- *Decision agreement*: It is a Boolean indicator of whether the same conclusion is drawn from the real and augmented data estimates. This means that the augmented data estimates have the same direction and statistical significance as the real data. The decision agreement does not apply if the analysis is descriptive. We would expect decision agreement to occur at a rate equal to power, which would be at least 80% of the time (ie, assuming the 9 trials are powered by design for at least 80%) [75].
- *Standardized difference*: It is a Boolean indicator of whether the difference in the parameter estimate between real and augmented data is consistent with the null hypothesis of no difference [75]. The Z value is computed and compared with the standard normal ($|Z| \leq 1.96$).
- *CI overlap*: It is a proportion of the real and augmented data parameter CIs overlap [76], which is a commonly used synthetic data utility metric. We would want this to be as close to 100% as possible but set 80% as a minimal value.

The 2 agreement metrics are consistent with previous measures of replicability [77-79], have been used to compare RWD analysis results against a clinical trial reference [75,80-83], and have been used to assess the replicability of psychological studies [77].

Results

The first set of results are shown in Table 3 where the monotonic relationship of the effect size over time is investigated. In none of the trials used for the study was the interaction term (ie, recruitment order by treatment) found to be statistically significant, indicating lack of evidence of monotonically varying treatment effect with respect to the order of participant recruitment.

XSL•FO
RenderX

Table 3. Evaluation of monotonic relationships over recruitment order in the estimated effect size.

Trial	Main effect (95% CI)	Interaction effect (95% CI)
REaCT[a]-ILIAD	0.52 (0.023 to 1.01)	−0.00000371 (−0.0000116 to 4.16×10^{-6})
REaCT-BTA[b]	−1.85 (−4.75 to 1.04)	0.011 (−0.032 to 0.0542)
CCTG[c] MA27	0.033 (−0.117 to 0.18)	−0.0000135 (−0.0000854 to 5.84×10^{-5})
NSABP[d] B34	−0.036 (−0.18 to 0.11)	8.28×10^{-5} (−0.000072 to 0.00024)
REaCT-G/G2	0.015 (−0.0019 to 0.032)	6.6×10^{-6} (-0.000005 to 1.85×10^{-5})
ABCSG[e]-12 (tamoxifen vs anastrozole)	0.11 (−0.13 to 0.35)	7.36×10^{-5} (−0.0004 to 0.00054)
ABCSG-12 (zoledronic acid vs no zoledronic acid)	−0.24 (0.48 to 0.002)	6.69×10^{-5} (−0.00041 to 0.00054)

[a]REaCT: Rethinking Clinical Trials.

[b]BTA: bone-targeted agents.

[c]CCTG: Canadian Cancer Trials Group.

[d]NSABP: National Surgical Adjuvant Breast and Bowel Project.

[e]ABCSG: Austrian Breast and Colorectal Cancer Study Group.

The results on the data with insufficient accrual (ie, the "reduced" datasets with no augmentation) for 2 values of *r* at 0.2 and 0.5 are shown in Table 4.

We can make 4 observations from Table 4:

1. For the REaCT-HER2+ study, which had the smallest sample size, there is no decision agreement as *r* increases. This is not surprising as the low sample sizes would mean unstable parameter estimates and larger CIs. The small sample size also explains the lack of difference in the standardized difference comparison.
2. The REaCT-ILIAD study had a statistically significant result with the full data. Decision agreement is no longer attained as *r* increases due to the smaller sample sizes, lower power, and hence, wider CIs.
3. The Austrian Breast and Colorectal Cancer Study Group (ABCSG)-ZOL analysis had a marginal nonsignificant outcome in the full trial. At small *r* values, small changes can have an impact on the statistical significance of the results, but that becomes less of an issue for large *r* values as the results will no longer be marginal with wider CIs. Although, one would expect that if the original results were marginally statistically significant, there would not be decision agreement even for larger values of *r*.
4. For the remaining trial analyses, the results were not significant in the original data. Therefore, the decision agreement would not be affected with the lower power as *r* increases, and the parameter estimates retained the same direction as the full trial. The standardized difference comparison indicates that the parameters were not different even as *r* increases, which is consistent with the inability to detect a monotonic effect with recruitment order presented above, and detecting a difference becomes more difficult as the sample size decreases.

These results indicate that drawing conclusions from reduced datasets can be misleading and can produce incorrect findings relative to those that would be obtained if target recruitment was achieved. In addition, the nature of any error would not be known a priori.

The fidelity results for the augmented datasets are shown in Figure 2. At low values of *r*, the generated datasets were small, making fidelity comparisons unstable. For the smaller datasets, such as REaCT-HER2+, REaCT–bone-targeted agents (BTA), REaCT-ZOL, and REaCT-ILIAD, the Hellinger distance values were highest but still relatively low on an absolute scale. For the other datasets, the Hellinger values were quite small with variation in a very narrow range, demonstrating high fidelity.

XSL•FO
RenderX

Table 4. The baseline primary results for the original datasets, and for the reduced datasets at 2 different values of *r* (ie, 0.2 and 0.5)

Trial name	Trial short name	Full dataset			*r*=0.2						*r*=0.5					
		Effect size (SE)	Variables used in the analysis	Analysis method	Sample size, N	Effect size (SE)	Estimate agreement	Decision agreement	Standardized difference	CI overlap	Sample size, N	Effect size (SE)	Estimate agreement	Decision agreement	Standardized difference	CI overlap
REaCT[a]-ILIAD	ILIAD [84]	0.52 (0.25)	8	GEE[b]	174	0.64 (0.28)	1	1	1	0.88	109	0.39 (0.36)	1	0	1	0.85
REaCT-BTA[c]	BTA [85]	−1.85 (1.48)	8	LM[d]	184	−2.54 (1.63)	1	1	1	0.89	115	−3.46 (2.03)	1	1	1	0.79
CCTG[e] MA27	CCTG [86]	0.033 (0.076)	6	Cox[f]	6060	0.045 (0.084)	1	1	1	0.96	3788	−0.006 (0.102)	1	1	1	0.87
NSABP[g] B34	NSABP [87]	−0.036 (0.076)	6	Cox	2648	−0.034 (0.085)	1	1	1	0.95	1655	−0.124 (0.108)	1	1	1	0.78
REaCT-G/G2	G/G2 [88]	0.015 (0.009)	10	GEE	320	0.012 (0.010)	1	1	1	0.91	200	0.0005 (0.016)	1	1	1	0.76
REaCT-HER2+[h]	HER2+ [89]	0.0058 (0.0062)	3	GLM[i]	40	0.0075 (0.006)	1	1	1	0.93	25	0.017 (0.009)	1	0	1	0.62
ABCSG[j]-12 (tamoxifen vs anastrozole)	ABCSG [90]	0.11 (0.12)	3	Cox	1442	0.13 (0.14)	1	1	1	0.96	901	0.07 (0.17)	1	1	1	0.87
ABCSG-12 (ZOL[k] vs no ZOL)	ABCSG-ZOL [90]	−0.24 (0.1237)	3	Cox	1442	−0.27 (0.14)	1	0	1	0.95	901	−0.26 (0.17)	1	1	1	0.87
REaCT-ZOL	ZOL [91]	58.98[l] (0.76)	2	*t* test	168	58.07 (0.89)	1	—	1	0.72	105	58.20 (1.10)	1	—	1	0.81
SWOG[m] 0307	SWOG [92]	−0.0087 (0.0145)	3	Surv[n]	4814	4.03×10^{-5} (0.016)	1	1	1	0.86	3009	−0.0096 (0.02)	1	1	1	0.86

[a]REaCT: Rethinking Clinical Trials.

[b]GEE: general estimating equations.

[c]BTA: bone-targeted agents.

[d]LM: linear model.

[e]CCTG: Canadian Cancer Trials Group.

[f]Cox: Cox regression

[g]NSABP: National Surgical Adjuvant Breast and Bowel Project.

[h]HER2+: human epidermal growth factor receptor-2 positive

[i]GLM; general linear model.

[j]ABCSG: Austrian Breast and Colorectal Cancer Study Group.

[k]ZOL: zoledronate.

[l]No effect was estimated for REaCT-ZOL—the figures shown correspond to descriptive analysis.

[m]SWOG: Southwest Oncology Group.

[n]Surv: difference in survival probabilities.

XSL•FO
RenderX

Figure 2. Hellinger distance results by comparing the training dataset with the generated dataset. The values were averaged across all generated 10 datasets. A value of 0 means that the 2 datasets are the same and of 1 indicates maximum difference. Note that the y-axis scales are not the same across all plots to provide better readability. ABCSG: Austrian Breast and Colorectal Cancer Study Group; BTA: bone-targeted agents; CCTG: Canadian Cancer Trials Group; CTGAN: conditional tabular generative adversarial network; HER2+: human epidermal growth factor receptor-2 positive; NSABP: National Surgical Adjuvant Breast and Bowel Project; REaCT: Rethinking Clinical Trials; SWOG: Southwest Oncology Group; TVAE: tabular variational auto encoder; ZOL: zoledronate.

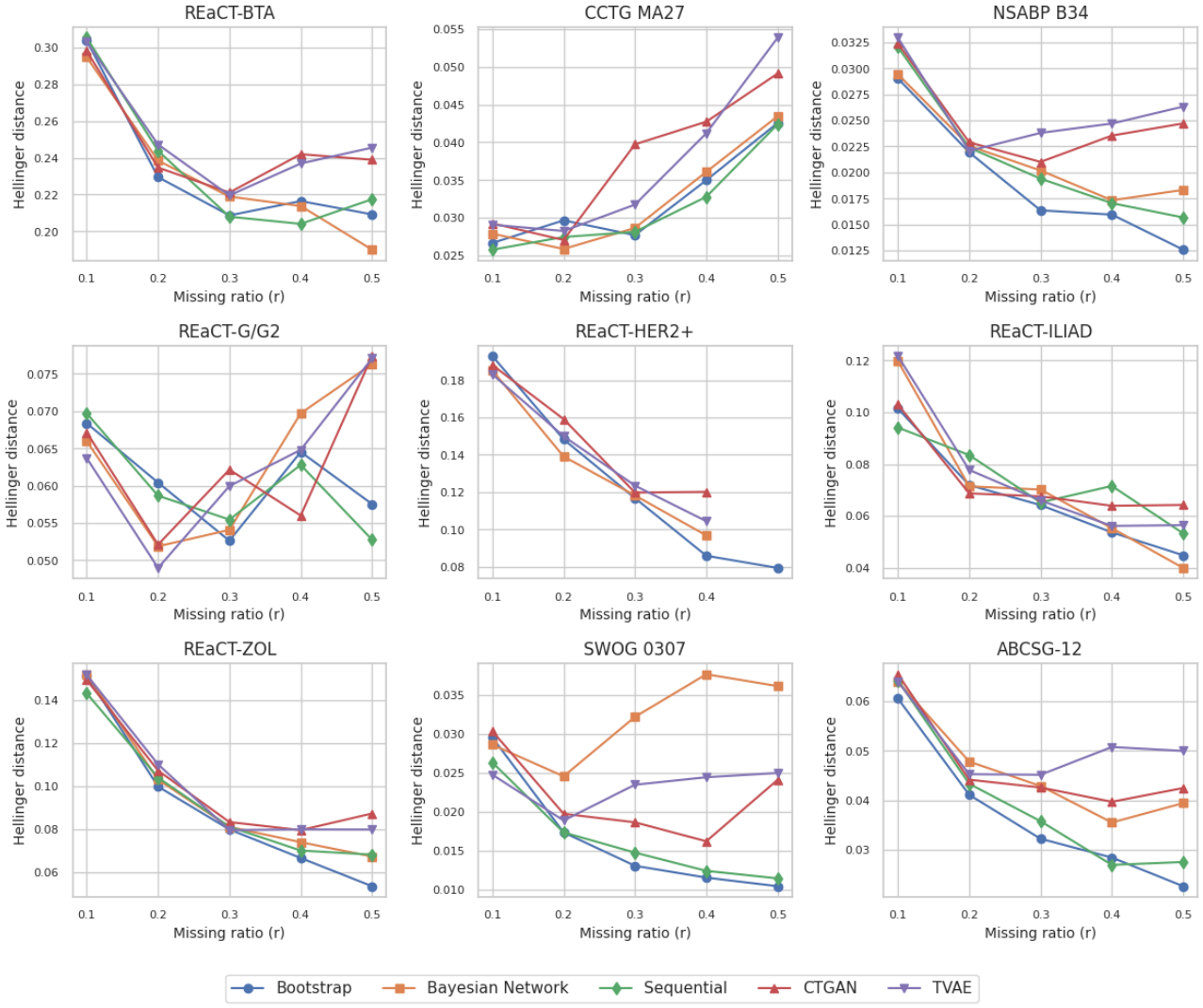

Another representation of fidelity is shown in Figure 3, where we compared the training dataset with a generated dataset of the same size. Here, we can see high fidelity values for all datasets as these results were less affected by small dataset sizes. As the training dataset size decreased with higher *r*, the fidelity decreased, but the changes in fidelity were modest.

All the detailed replicability results for all values of *r* are provided in Table S2 in Multimedia Appendix 1.

The first general observation is that the bootstrap and sequential synthesis tend to perform best among all the generative models. Therefore, in Figure 4 we only show their results across all datasets that were examined for the 4 measures of replicability. Both types of generative models achieved high estimate and decision agreement and maintained a CI overlap above 80% even as *r* values approached 0.5.

XSL•FO
RenderX

Figure 3. Hellinger distance results by comparing the training dataset with a generated dataset of the same size. The values were averaged across all generated 10 datasets. A value of 0 means that the 2 datasets are the same and of 1 indicates maximum difference. Note that the y-axis scales are not the same across all plots to provide better readability. ABCSG: Austrian Breast and Colorectal Cancer Study Group; BTA: bone-targeted agents; CCTG: Canadian Cancer Trials Group; CTGAN: conditional tabular generative adversarial network; HER2+: human epidermal growth factor receptor-2 positive; NSABP: National Surgical Adjuvant Breast and Bowel Project; REaCT: Rethinking Clinical Trials; SWOG: Southwest Oncology Group; TVAE: tabular variational auto encoder; ZOL: zoledronate.

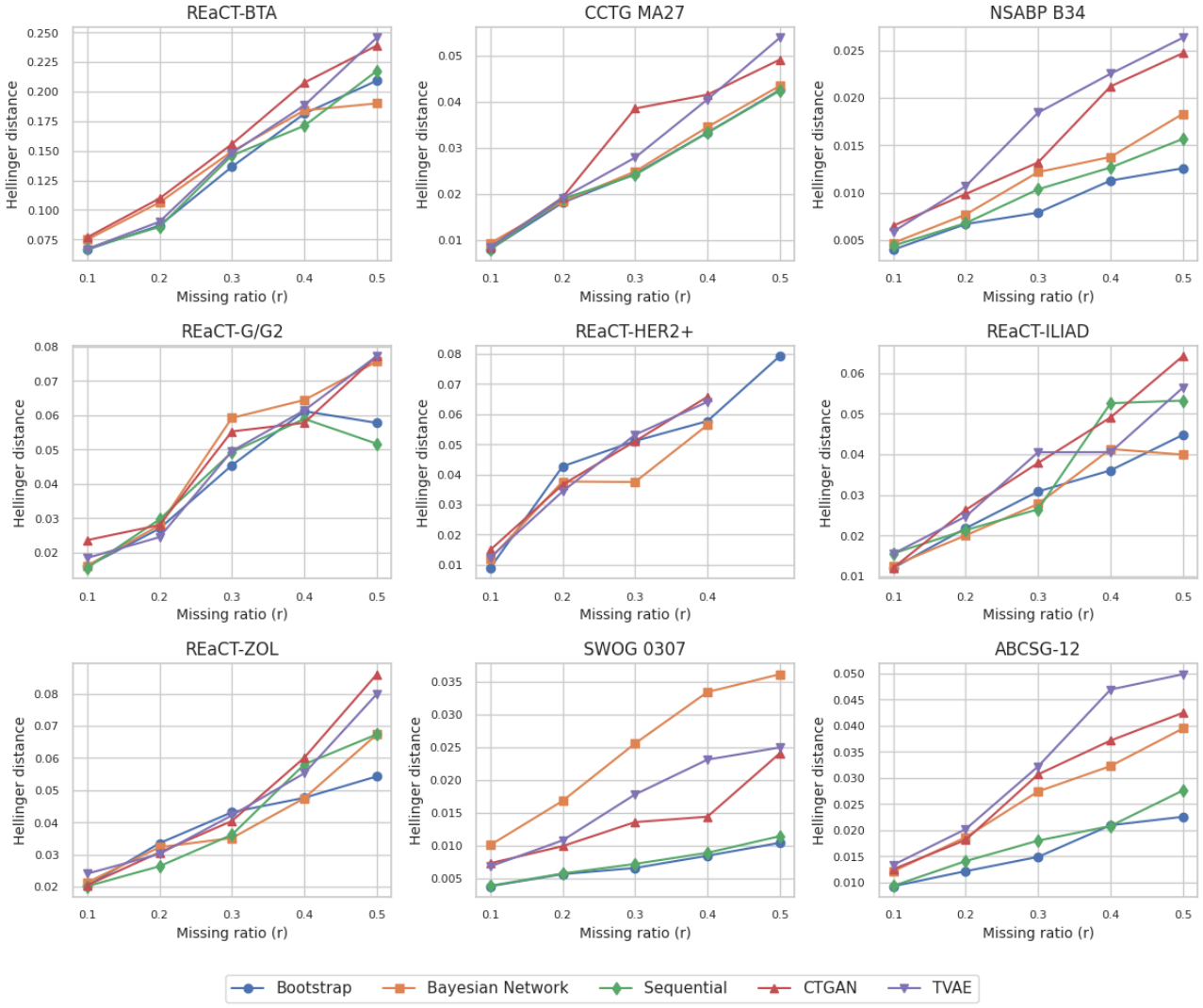

It should be noted that sequential synthesis fails for the smallest clinical trial and that was not included in the denominator for Figure 4. This failure was a design decision in the implementation that we used to not train a model with less than 50 observations. Also note that in the plots model failures are not counted in the denominator.

For the results by dataset, we focus on 4 trials that exemplify all the scenarios in our dataset across all 4 generative models, and are shown in Figure 5 for estimate agreement, Figure 6 for decision agreement, Figure 7 for the standardized difference, and Figure 8 for the CI overlap. The values were consistently high for estimate agreement, standardized difference, and CI overlap across all the scenarios, even at values of *r* approaching 0.5. The results for decision agreement in Figure 6 were more varied and can be characterized as follows:

1. For studies where there was a statistically significant result, for example, REaCT-ILIAD, the ability to maintain decision agreement deteriorates with higher values of *r*. The best performing generative model was sequential synthesis in that decision agreement was maintained for *r* as high as 0.4. Next was the bootstrap, which had a decision agreement for *r* up to 0.3.
2. For studies where the results were not statistically significant (eg, Canadian Cancer Trials Group MA27) or a marginal nonsignificant result (eg, ABCSG–ZOL), the value of *r* had no impact on the replicability of the results in that all the results were successfully replicated.
3. For smaller studies (eg, REaCT-ZOL), some generative models were able to maintain decision agreement for *r* as high as 0.3, although most models failed for *r* greater than that.

XSL•FO
RenderX

Figure 4. All 4 metrics calculated across all the datasets for the bootstrap and sequential generators. For estimate agreement, decision agreement, and standardized difference, the y-axis is the proportion across all datasets. For CI overlap, the y-axis is the average across all data sets. Modeling failures are not considered.

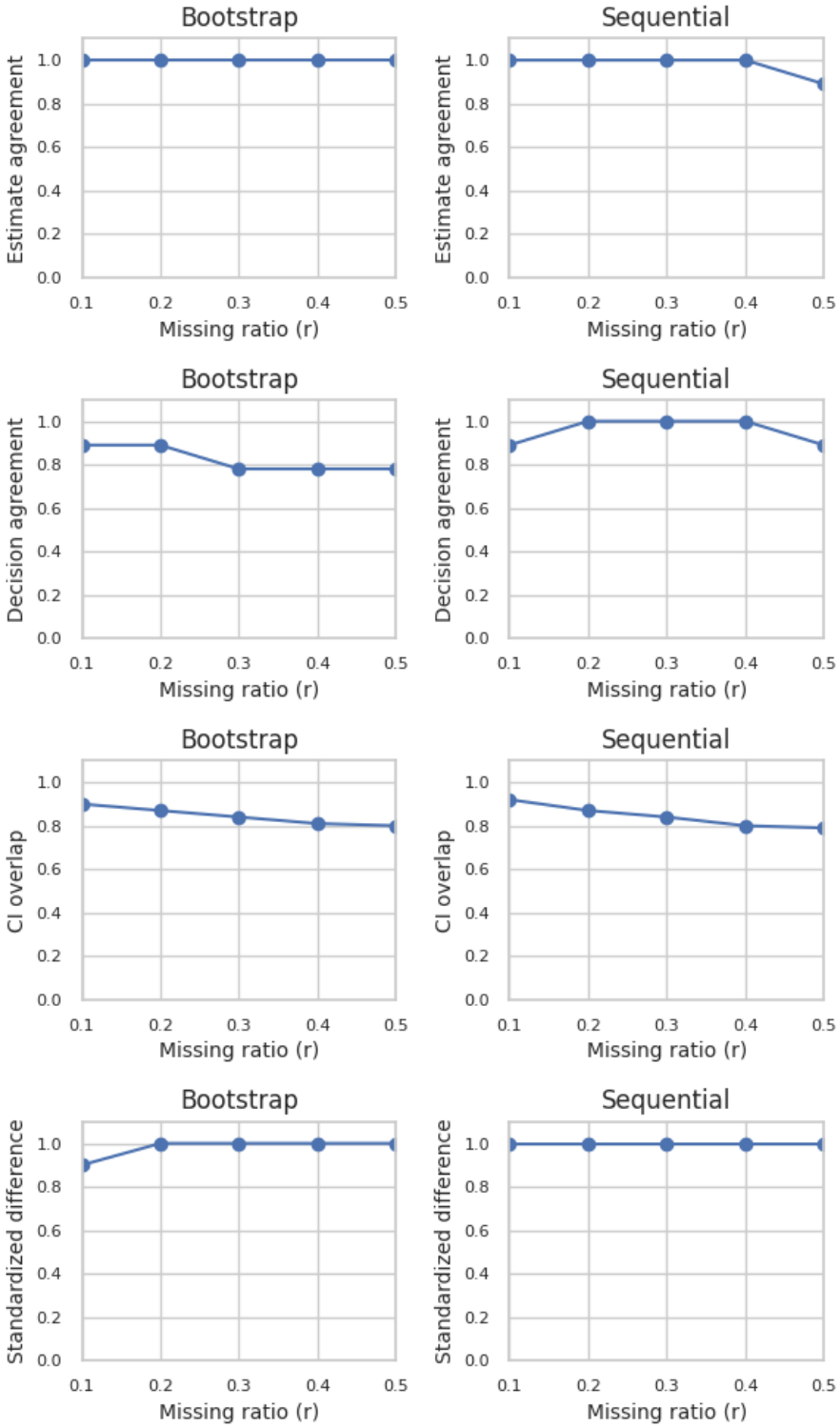

XSL•FO
RenderX

Figure 5. Estimate agreement for selected datasets—proportion across all generators. ABCSG: Austrian Breast and Colorectal Cancer Study Group; CCTG: Canadian Cancer Trials Group; HER2+: human epidermal growth factor receptor-2 positive; REaCT: Rethinking Clinical Trials; ZOL: zoledronate.

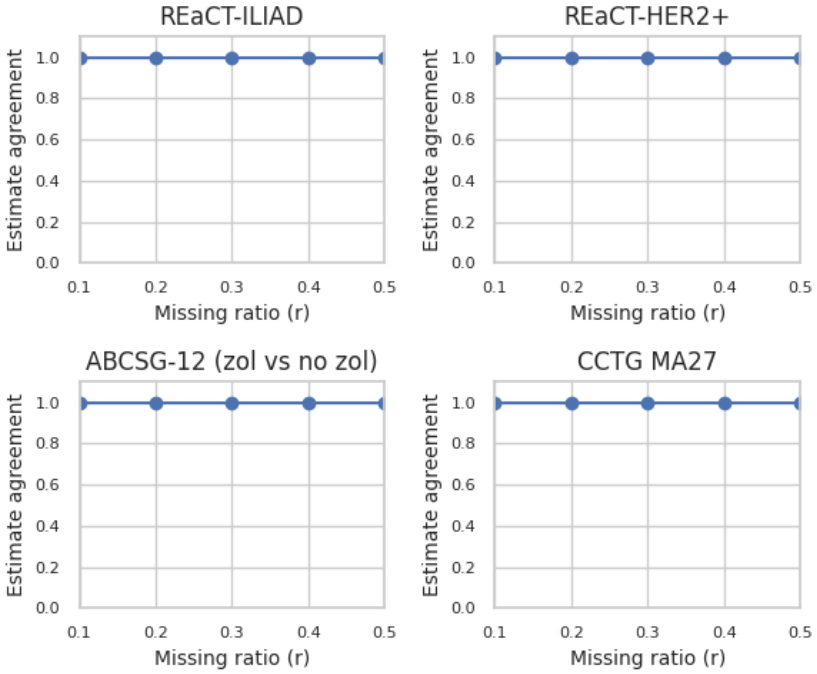

Figure 6. Decision agreement for selected datasets—proportion across all generators. ABCSG: Austrian Breast and Colorectal Cancer Study Group; CCTG: Canadian Cancer Trials Group; HER2+: human epidermal growth factor receptor-2 positive; REaCT: Rethinking Clinical Trials; ZOL: zoledronate.

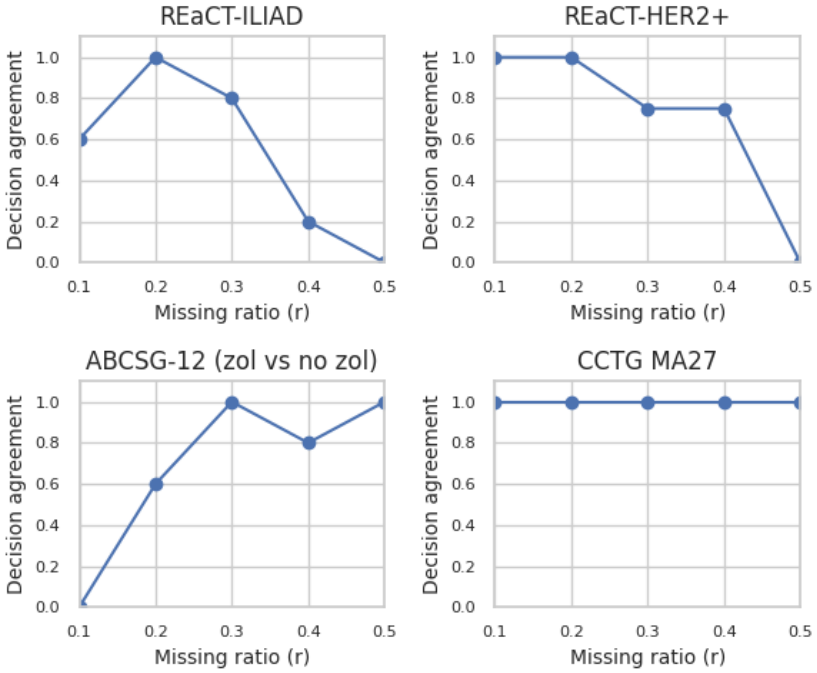

XSL•FO
RenderX

Figure 7. Standardized difference indicator for selected datasets—proportion across all generators. ABCSG: Austrian Breast and Colorectal Cancer Study Group; CCTG: Canadian Cancer Trials Group; HER2+: human epidermal growth factor receptor-2 positive; REaCT: Rethinking Clinical Trials; ZOL: zoledronate.

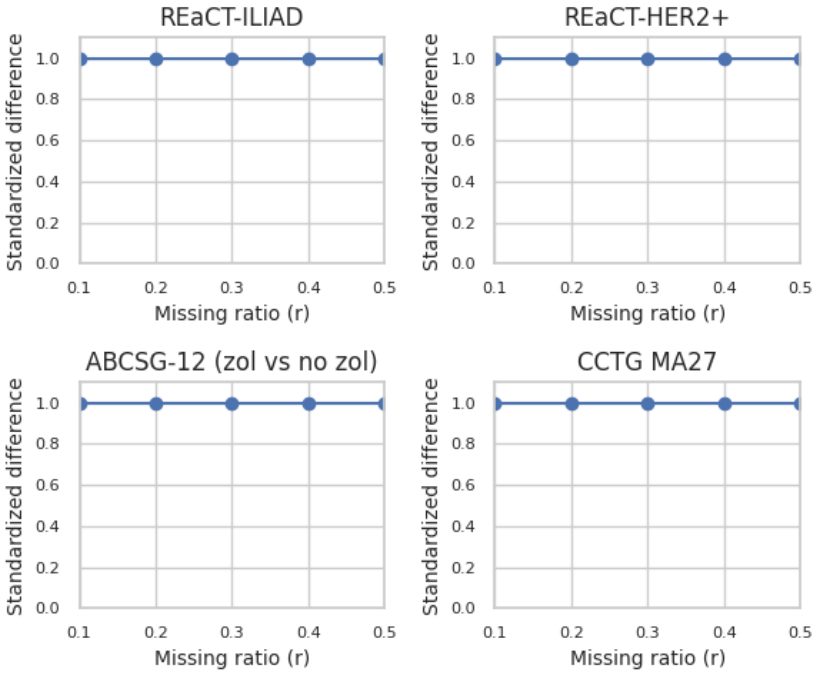

Figure 8. CI overlap for selected datasets—average across all generators. ABCSG: Austrian Breast and Colorectal Cancer Study Group; CCTG: Canadian Cancer Trials Group; HER2+: human epidermal growth factor receptor-2 positive; REaCT: Rethinking Clinical Trials; ZOL: zoledronate.

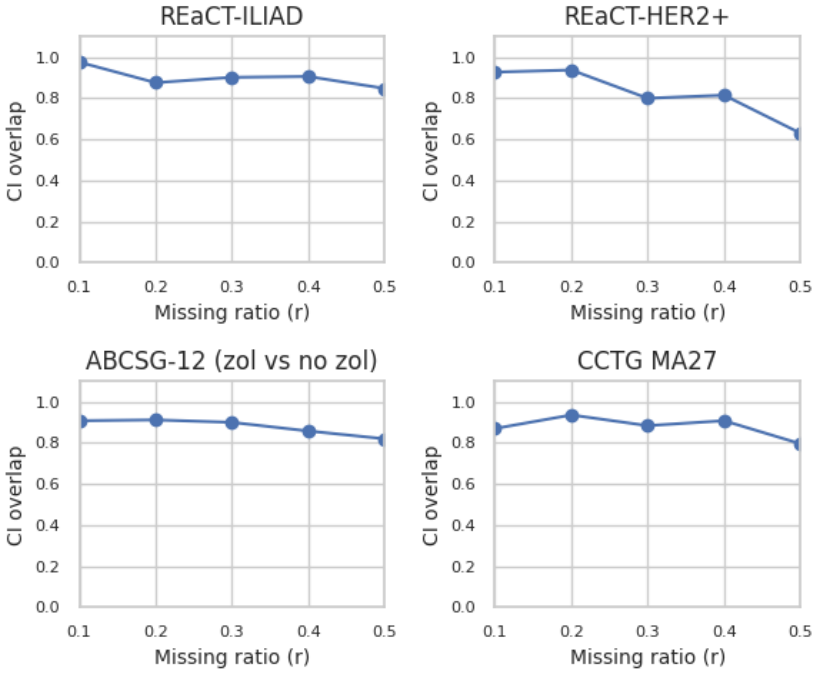

Discussion

Summary

Many clinical trials face accrual problems, resulting in an inability to reach target recruitment and resulting in analysts drawing conclusions from potentially underpowered studies. This exposes patients to toxicity and additional costs, with potentially no scientific benefit. Accrual problems may be due to genuine difficulty with recruitment of patients or with execution quality challenges during the trial itself.

When a study is unable to recruit more patients, the study can be stopped, and the relevant analyses is performed on the available data. For small trials, analyzing the data with insufficient accrual results in even smaller sample sizes, which

XSL•FO
RenderX

can produce unstable parameter estimates and direction. For larger trials, when the complete trial results are statistically significant, an analysis with insufficient accrual can be underpowered and result in nonsignificant findings. For marginal results with the full data, insufficient accrual can reverse their statistical significance. When the full study results are not significant, then insufficient accrual would have less of an impact. A priori, it would not be known which one of these situations pertains to a particular study, making it difficult to interpret the results when analysis is performed with unplanned accrual deficiencies.

The objective of this study was to determine whether generative models can be a useful tool to rescue clinical studies that have insufficient accrual, through augmentation. Generative models have been applied to simulate participants [93-96] and counterfactuals [95,96] in the context of in silico clinical trials. While there have been concerns about generative models overfitting for the small datasets typically encountered in clinical trials [96,97], recent studies have been able to generate synthetic variants of full clinical trial datasets with high utility [74,95,97-102].

To test the ability of generative models to augment clinical trial datasets, we evaluated 4 different types of commonly used generative models (ie, sequential synthesis using decision trees, Bayesian network, GAN, and VAE) on 9 different breast cancer clinical trials and 10 different analyses. The study simulated different degrees of insufficient accrual and the generative models simulated replacement patients to compensate. The last fraction of recruited patients were replaced with simulated ones ranging from r=0.1 to r=0.5. In addition to generative models, we evaluated a bootstrap approach. These augmented datasets were then used to replicate the published analyses (ie, using the complete datasets) on these 9 trials.

An important assumption for these augmentation methods to work was that participants recruited early were not systematically different than late participants in their estimated effect size. It has been argued that estimated effect sizes tend to vary as patients are recruited and converge to the true value with more information [40,42]. In contrast, as sites gain experience and adjust their processes, there could be a monotonic treatment effect over recruitment time. Also, consider, for example, if there was treatment effect heterogeneity on disease severity (ie, the impact of the treatment on the outcome depends on disease severity), with high severity patients having a bigger response to the intervention, and fewer high severity patients were recruited early in the study compared with late in the study. This would manifest itself as a monotonic relationship between the estimated effect size and patients recruited over time.

We tested that monotonic relationship hypothesis by examining the interaction effect between recruitment order and treatment on the outcome. We did not find evidence across the trials that this relationship was monotonic (ie, all interaction effects were small and not statistically significant), meaning that we could not detect an increasing or decreasing effect size as more patients were recruited. This is despite these trials having long enrollment periods, in some cases lasting years. Therefore, if the early and late participants were, on average, similar in the estimated effect size, then generative models and a simple bootstrap would also be expected to work well.

These results are supportive of hypothesis 1, which stated that patients recruited early in a trial are similar to those recruited later in the trial.

Furthermore, the fidelity of the generated datasets was quite high relative to the training datasets from the different generative models. In all cases, the patterns in datasets that were synthesized had a high similarity on the Hellinger distance to the datasets of the participants already recruited.

Several observations can be made from our results:

1. A bootstrap would be attractive due to its simplicity and low computational burden. However, this method tended to lack consistent decision agreement when a trial result was statistically significant and when the trial was small, while in other scenarios, sampling with replacement performed well. However, the former are nontrivial failure modes.
2. All approaches struggled with marginal results (eg, marginally nonsignificant results) when the r value was low. This indicates a general sensitivity to that particular scenario.
3. For r values as high as 0.4, sequential synthesis performed well across all datasets. This means that both decision agreement and estimate agreement were achieved, CI overlap was at or above 0.8, and it was either equivalent to or better than the other methods evaluated.
4. Bayesian networks and the GAN had the next best performance after sequential synthesis; however, the Bayesian network had slightly better CI overlap overall up to an r=0.4.

To ensure reasonable performance across multiple scenarios, the results suggest that sequential synthesis can be used to address insufficient accrual up to r=0.4 (ie, only 60% of the target is recruited).

These results are supportive of hypothesis 2, which stated that generative models can simulate the remaining patients in a clinical trial with insufficient recruitment, and the augmented dataset would replicate the results if the trial did reach target recruitment.

It should be noted that using generative models to simulate additional patients as described in this study would not be preplanned, as opposed to a planned interim analysis. Insufficient accrual becomes a problem when there are no budgeted resources available to continue recruiting patients, for example, by adding sites, extending recruitment time, or changing the inclusion and exclusion criteria. Although, if the results from the augmentation show positive findings, the case may be made to allocate more resources to continue recruitment. Additionally, if the results from augmentation show negative findings, then that would provide a stronger case for terminating the study.

The countries of recruitment for the clinical trials used in this study cover multiple regions around the world as shown in Table

XSL•FO
RenderX

2. Our results were consistent across the different jurisdictions. Therefore, it would be reasonable to have confidence that the findings are generalizable across multiple jurisdictions and not specific to a particular region.

Comparison With Prior Work

Previous studies have demonstrated that sequential synthesis performs well (ie, in terms of replicability of published studies) on oncology clinical trial datasets [74,98], and therefore, our findings are consistent with that evidence. More generally, sequential synthesis has been found to have superior utility across different types of datasets relative to other types of generative models [103-105]. Also, it should be noted that most of the published analyses that were replicated across all clinical trials used datasets that were low dimensional, which imposes lower sample size requirements for the generative models.

An earlier study found that a VAE generative model trained on early patients could augment a clinical trial with simulated patients [35]. In that study, the authors argued that generative models can also enable the design of smaller studies to start off with. This means that studies would be designed to be smaller, with augmentation used to reach target recruitment for the final analysis. However, that analysis only considered a single simulated clinical trial (ie, not real data).

The argument for using augmentation to prospectively design smaller studies is appealing. The largest factor driving up the cost of trials is the number of participants required to achieve sufficient statistical power [106,107]. The median cost per participant in drug trials (ie, in general) was estimated to be US $41,413 [107], the median cost per participant specifically in oncology drug trials was US $100,271 [107], and an earlier study found the 1-year cost to be US $17,003 per patient in the treatment arm and US $15,516 per control participant [108]. Designing studies that require fewer patients to be recruited can improve the cost-effectiveness of studies.

Our current analysis generalizes this work to other types of generative models, including a VAE, but we did find that a VAE did not perform well on our criteria. Plus, we performed the evaluation on 9 real clinical trial datasets of different sizes and durations and compared the generative models to a simple bootstrap. Furthermore, an important difference is that rescuing a study due to insufficient accrual using augmentation is not planned, whereas designing a small study is planned.

Nevertheless, using augmentation to design smaller studies deserves further investigation, and it remains a necessity that researchers aim to recruit the target sample size wherever possible.

Clinical trials are known to underrepresent certain groups, and hence, there is the potential for introducing bias in the results. For instance, a recent study in Canada found that the underrepresentation of Black patients in cancer research remains a significant concern, with 15 out of the 20 most common types of cancer not being studied in Black communities [109,110]; studies on underrepresented populations in clinical trials show racial and ethnic disparities worldwide [111-118]; and there is a consistent underrepresentation of various other groups, such as older adults [111,112,119-122], women [111,113,119,120,123], and individuals of lower socioeconomic status and educational level [111,112,124]. Furthermore, synthetic data generation has been shown to introduce bias in the generated data relative to the training data [125], and these biases are propagated across multiple generations of generative models, where the output of one is used as training for the next one [126].

Our analysis did not explicitly evaluate representation bias as we were replicating the published analyses rather than identifying and correcting any weaknesses and did not explicitly attempt to mitigate such underrepresentation to the extent that it existed in the original datasets. Nevertheless, analysts can also apply augmentation methods to compensate for any biases that may exist in the training datasets [127-129].

Another risk of bias is if there is a relationship between a particular characteristic and the order of recruitment. For example, if the first 60% of patients recruited were aged mostly >70 years and the last 40% were aged mostly <70 years, then the trained generative model and their simulated patients would not include sufficient younger patients. However, such an age bias would only impact the randomized study outcomes if there was an interaction between recruitment order, or factors correlated to it, such as age in this example, and the treatment. We explicitly tested for such an interaction effect, and as seen in the results, there were none detected.

Generative models can simulate a larger number of patients than what is needed to reach target recruitment (ie, data amplification). On the surface, this may seem to be a mechanism to amplify the statistical power of the study and solve the problem of drawing conclusions from small studies. However, with the necessary adjustments using the combining rules described in our methodology, it has been shown that amplification does not increase statistical power for fully synthetic data as the adjusted SEs of parameter estimates are also increased [71]. The same study shows that population inferences and replicability diminish markedly without using the combining rules during the analysis of synthetic data. Further examination is needed to determine whether the same conclusions would hold for hybrid data that is only part synthetic.

The incorporation of generative models and simulated patients in industry-sponsored clinical trials would necessitate collaboration with sponsors to apply and evaluate these methods in their studies across multiple therapeutic areas to determine consistency in performance. However, recent surveys indicate that sponsors see uncertainty about regulators' expectations and requirements for evidence as a critical barrier for the adoption of computer modeling and simulation methods in clinical trials [130]. Some efforts have identified high level principles that can be applied for the quality assurance evaluation of in silico trials [131], as well as a general good practice guidance for the application of simulations [132]. Regulators have noted the potential of simulated patients in clinical trials [133] and have further suggested adopting the Government Accountability Office accountability framework [134] for the application of machine learning models to in silico trials, which includes addressing challenges related to governance, accountability,

XSL•FO
RenderX

and transparency; data considerations; and model development, performance, and validation [135]. Furthermore, synthetic data, because of their privacy-protective properties, can serve as a more readily available proxy for RWD when used in regulatory submissions [136], which can solve the data access challenge and accelerate the generation of real-world evidence. Some experts at regulatory agencies have expressed cautious optimism that synthetic data can be used by manufacturers [137], with replicability of results on real data and the need for further experimental exemplars being emphasized specifically for synthetic data generation [94]. Our study contributes to that evidence base.

Recruitment is a challenging issue not only in clinical trials but in other clinical studies as well [138,139]. These challenges are exacerbated with studies on rare diseases [140] and pediatric studies. Pediatric datasets are typically small due to a scarcity of potential study participants: there are fewer children in the population with severe disease [141]. Additional challenges include the complex ethical issues surrounding research involving children and an extra layer of consent required for pediatric participants (ie, parental consent is required in addition to patients' assent) [142]. Many trials recruited a very small number of children [143], and studies with placebo arms are also at a disadvantage as patients are less likely to participate in case they do not get randomized to the treatment arm [144]. Therefore, the results from our study would have broader applicability beyond oncology trials in adults.

On a methodological point, this study was possible because of the ability to obtain access to the original datasets across 9 different clinical trials. While there has been strong interest in making more clinical trial data available for secondary analysis by journals, funders, the pharmaceutical industry, and regulators [145-153], data access for secondary analysis remains a challenge [154], sometimes taking many months to get data [155,156]; for example, an analysis of the success of getting individual-level data for meta-analysis projects from authors found that the percentage of the time these efforts were successful ranged from 0% to 58% [156-161]. Specifically, recent reports highlight the difficulties in accessing data for health research and machine learning analytics [162-164]. In our case, the process of getting access to all the datasets used in this study took approximately 2 years, including executing the necessary data sharing agreements and establishing collaborations with the original investigators.

One of the reasons that access to individual-level clinical trial data faces friction is concern over patient privacy by the patients and regulators [165,166]. However, the general assumption has been that synthetic data produced through generative models have low identity disclosure vulnerability because there is no unique or one-to-one mapping between the records in the synthetic data with the records in the original (ie, real) data [167-174]. While there are other types of disclosure vulnerabilities that are also relevant [175,176], some authors have argued that synthetic data can be considered nonpersonal information under statuary definitions in North America and Europe [177-179]. This would arguably be the case if disclosure vulnerability measurements demonstrate vulnerability values that are below acceptable thresholds. To that end, it is encouraging that recently study authors have been making synthetic variants of data used in their research papers publicly available to enable open science [180-183]. However, given that generating fully synthetic variants of the clinical trial datasets used in this study were not part of the original protocol, readers interested in data access can make a request to the individual data custodians, with the necessary contacts in the *Data Availability* section.

Limitations

Although our datasets covered single site as well as multisite studies, and trials performed across multiple regions of the world with a wide range of sizes and durations, our results were obtained only from oncology trials (mostly breast cancer). There is no a priori reason for these methods not to work for other diseases and conditions, and different populations. To generalize the findings, it is necessary to replicate the findings in this work for other diseases, particularly those that impose high societal costs and where the acceleration of clinical trial evidence can be most impactful.

It is not known whether there would be evidence of a monotonic effect size over the enrollment period in other types of clinical trials and with different populations. For example, one can argue that surgery trials may exhibit a monotonic effect because surgeons become more experienced with new procedures over time. None of our studies were surgery trials. Therefore, this monotonicity relationship would need to be investigated further in these other contexts before drawing broader conclusions.

This study did not consider safety data, which tend to have fewer observations in a clinical trial. Because of the relatively smaller number of adverse events, it is more challenging to train a generative model on that kind of information. Therefore, studies comparing existing interventions or new indications would be more suited for the application of augmentation methods.

Furthermore, our retrospective analysis was limited to 9 clinical trials that were actually completed. Studies that do not reach accrual targets may have different characteristics, which could lead to different results. While our results are encouraging based on a retrospective analysis, future work should evaluate this approach through a prospective design.

When the *r* value increases, the available dataset size to train a generative model decreases. For the smallest clinical trials, in some rare instances, this resulted in generative model failure. The generative models that did not fail under those conditions had a high risk of overfitting to the training data, although they still performed better than a simple bootstrap on decision agreement. For most of our small studies (REaCT-ILIAD, REaCT–BTA, and REaCT-ZOL), the number of variables that were used for training was also quite small (8, 8, and 11, respectively; Table 1). To the extent that low dimensionality has a diluting impact on the rate of overfitting for a fixed sample size, the small number of variables would reduce that risk. Furthermore, most of the variables in some of the small trials were categorical with very few categories (ie, mostly binary), suggesting quite simple datasets were being modeled (eg, REaCT-HER2+).

XSL•FO
RenderX

We did not find that the deep learning generative models performed as well as the other approaches that were considered. One possible explanation is that additional hyperparameter tuning for these models was not performed and the default implementation model size and characteristics were used. The tuning could have improved the performance of these models. Although the default hyperparameters were suitable for low dimensional data, which was the case for the analysis datasets in many of these clinical trials.

In drawing our conclusions, we used a 0.8 value as a threshold of acceptability on CI overlap. Should one apply a more stringent threshold, then the values of *r* where the results would be acceptable would decrease further.

Future Work

Further research is needed to provide evidence-based parameters for the number of variables that can be simulated with different types of generative models and the appropriate number of patients in a trial for training these models. This will further ensure that augmentation is not applied in inappropriate contexts.

Studies on rare diseases would have few observations, making it more challenging to train a generative model using the methods described here. Evaluations using pretrained generative models to simulate clinical trial patients should be investigated as these may be more applicable to rare disease trials.

Given that our focus was to address the insufficient accrual problem, the augmentation that was performed covered all arms of the study. However, the methods described here can be applied to individual arms as well. For example, it is possible to augment only the control arm of a trial if there were challenges accruing patients in the control arm or it could be by design to only recruit a subset of control patients and simulate the rest. Future work can evaluate such alternative augmentation strategies.

An alternative approach for augmentation to address insufficient accrual is to use pretrained generative models. For example, these models can be trained on historical clinical trial datasets, and then the pretrained models are used to simulate additional patients. This is different from the approach presented in this paper, whereby already collected data from the trial are used to train the generative models. It would be informative to compare both approaches to determine which would work better in practice, and to understand the types of historical datasets that give the best results for training these generative models.

Acknowledgments

This research is funded by a data transformation grant of the Canadian Cancer Society grant 707600, the Canada Research Chairs program through the Canadian Institutes of Health Research, and Discovery Grant RGPIN-2022-04811 from the Natural Sciences and Engineering Research Council of Canada and by the Government of Ontario (Ontario Center of Innovation).

Data Availability

The datasets used in this study are not publicly available. Access to the datasets can be requested from the contacts listed in Multimedia Appendix 1 for each of the datasets used in this study. The synthetic data generation code is available in the *pysdg* package [184].

Authors' Contributions

SEK, NM, GP, MC, and KEE designed the study. All authors participated in collecting and curating the datasets and writing and reviewing the paper. SEK, NM, GP, MC, CF, DH, and KEE performed the data analysis.

Conflicts of Interest

KEE owns shares in Aetion. AABB received a travel grant from Janssen. GP has received consulting fees from Traferox Technologies and Merck, honorariums from Astra-Zeneca and Takeda, and has a close family member who is employed by Roche Canada and who owns stock in Roche Ltd. MFS has received honoraria from GILEAD, AstraZeneca, Merck, Seagen, Novartis, Pfizer, Roche, and Lilly. RG has advisory roles and provided expert testimony to: Celgene, Novartis, Roche, BMS, Takeda, Abbvie, Astra Zeneca, Janssen, MSD, Merck, Gilead, Daiichi Sankyo; Honoria: Celgene, Roche, Merck, Takeda, AstraZeneca, Novartis, Agmen, BMS, MSD, Sandoz, Abbvie, Gilead, Daiichi Sankyo; financing of scientific research: Celgene, Merck, Takeda, AstraZeneca, Novartis, Agmen, BMS, MSD, Sandoz, Gilead, Roche; travel support: Roche, Agmen, Janssen, Astra Zeneca, Novatris, MSD Gilead, Abbvie, Daiichi Sankyo. M. G. reports personal fees / travel support from Amgen, AstraZeneca, Bayer, DaiichiSankyo, EliLilly, EPG Health (IQVIA), Menarini-Stemline, MSD, Novartis, PierreFabre, Veracyte.
KEE is the Editor-in-Chief of JMIR AI at the time of this publication but played no role in the editorial handling or peer review process of this manuscript.

Multimedia Appendix 1
Methodology details and analysis results.
[PDF File (Adobe PDF File), 300 KB - jmir_v27i1e66821_app1.pdf]

References

XSL•FO
RenderX

1. Prescott RJ, Counsell CE, Gillespie AJ, Grant AM, Russell IT, Kiauka S, et al. Factors that limit the quality, number and progress of randomised controlled trials: a review. Health Technol Assess 1999;3(20). [doi: 10.3310/hta3200]
2. Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs 2010 Jan 17;19(1-2):227-233. [doi: 10.1111/j.1365-2702.2009.03041.x] [Medline: 20500260]
3. Kasenda B, von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA 2014 Mar 12;311(10):1045-1051 [FREE Full text] [doi: 10.1001/jama.2014.1361] [Medline: 24618966]
4. Kitterman DR, Cheng SK, Dilts DM, Orwoll ES. The prevalence and economic impact of low-enrolling clinical studies at an academic medical center. Acad Med 2011 Nov;86(11):1360-1366 [FREE Full text] [doi: 10.1097/ACM.0b013e3182306440] [Medline: 21952064]
5. Stensland KD, McBride RB, Latif A, Wisnivesky J, Hendricks R, Roper N, et al. Adult cancer clinical trials that fail to complete: an epidemic? J Natl Cancer Inst 2014 Sep;106(9):dju229. [doi: 10.1093/jnci/dju229] [Medline: 25190726]
6. Feller S. One in four cancer trials fails to enroll enough participants. United Press International. 2015 Dec 30. URL: https://www.upi.com/Health_News/2015/12/30/One-in-four-cancer-trials-fails-to-enroll-enough-participants/2611451485504/ [accessed 2021-03-25]
7. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 2006 Apr 07;7:9 [FREE Full text] [doi: 10.1186/1745-6215-7-9] [Medline: 16603070]
8. Institute of Medicine. Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary. Washington, DC: The National Academies Press; 2010.
9. Sully BG, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013 Jun 09;14:166 [FREE Full text] [doi: 10.1186/1745-6215-14-166] [Medline: 23758961]
10. Mirza M, Siebert S, Pratt A, Insch E, McIntosh F, Paton J, et al. Impact of the COVID-19 pandemic on recruitment to clinical research studies in rheumatology. Musculoskeletal Care 2022 Mar;20(1):209-213 [FREE Full text] [doi: 10.1002/msc.1561] [Medline: 33938621]
11. Mitchell EJ, Ahmed K, Breeman S, Cotton S, Constable L, Ferry G, et al. It is unprecedented: trial management during the COVID-19 pandemic and beyond. Trials 2020 Sep 11;21(1):784 [FREE Full text] [doi: 10.1186/s13063-020-04711-6] [Medline: 32917258]
12. Slow recruitment due to Covid-19 disruptions continues to climb in 2022. Clinical Trials Arena. 2022 Jan 31. URL: https://www.clinicaltrialsarena.com/analyst-comment/slow-recruitment-covid-19-disruptions/ [accessed 2023-10-05]
13. McDonald K, Seltzer E, Lu M, Gaisenband SD, Fletcher C, McLeroth P, et al. Quantifying the impact of the COVID-19 pandemic on clinical trial screening rates over time in 37 countries. Trials 2023 Apr 04;24(1):254 [FREE Full text] [doi: 10.1186/s13063-023-07277-1] [Medline: 37013558]
14. Hauck CL, Kelechi TJ, Cartmell KB, Mueller M. Trial-level factors affecting accrual and completion of oncology clinical trials: a systematic review. Contemp Clin Trials Commun 2021 Dec;24:100843 [FREE Full text] [doi: 10.1016/j.conctc.2021.100843] [Medline: 34765799]
15. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015 Feb;12(1):77-83 [FREE Full text] [doi: 10.1177/1740774514558307] [Medline: 25475878]
16. Lièvre M, Ménard J, Bruckert E, Cogneau J, Delahaye F, Giral P, et al. Premature discontinuation of clinical trial for reasons not related to efficacy, safety, or feasibility. BMJ 2001 Mar 10;322(7286):603-605 [FREE Full text] [doi: 10.1136/bmj.322.7286.603] [Medline: 11238162]
17. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002 Jul 17;288(3):358-362. [doi: 10.1001/jama.288.3.358] [Medline: 12117401]
18. Schmidli H, Häring DA, Thomas M, Cassidy A, Weber S, Bretz F. Beyond randomized clinical trials: use of external controls. Clin Pharmacol Ther 2020 Apr 17;107(4):806-816. [doi: 10.1002/cpt.1723] [Medline: 31725899]
19. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res 2011 May;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
20. Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: The changing landscape of the regulatory approval process. Pharmacoepidemiol Drug Saf 2020 Oct 01;29(10):1201-1212 [FREE Full text] [doi: 10.1002/pds.4932] [Medline: 31823482]
21. Gökbuget N, Kelsh M, Chia V, Advani A, Bassan R, Dombret H, et al. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. Blood Cancer J 2016 Sep 23;6(9):e473 [FREE Full text] [doi: 10.1038/bcj.2016.84] [Medline: 27662202]
22. Davi R, Mahendraratnam N, Chatterjee A, Dawson CJ, Sherman R. Informing single-arm clinical trials with external controls. Nat Rev Drug Discov 2020 Dec 18;19(12):821-822. [doi: 10.1038/d41573-020-00146-5] [Medline: 32811986]

XSL•FO
RenderX

23. Mumuni A, Mumuni F. Data augmentation: a comprehensive survey of modern approaches. Array 2022 Dec;16:100258. [doi: 10.1016/j.array.2022.100258]
24. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019 Jul 6;6:60. [doi: 10.1186/s40537-019-0197-0]
25. Goceri E. Medical image data augmentation: techniques, comparisons and interpretations. Artif Intell Rev 2023 Mar 20:1-45 [FREE Full text] [doi: 10.1007/s10462-023-10453-z] [Medline: 37362888]
26. Wen Q, Sun L, Yang F, Song X, Gao J, Wang X, et al. Time series data augmentation for deep learning: a survey. arXiv Preprint posted online on Feb 27, 2020 [FREE Full text] [doi: 10.24963/ijcai.2021/631]
27. Iwana BK, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. PLoS One 2021 Jul 15;16(7):e0254841 [FREE Full text] [doi: 10.1371/journal.pone.0254841] [Medline: 34264999]
28. Sabay A, Harris L, Bejugama V, Jaceldo-Siegl K. Overcoming small data limitations in heart disease prediction by using surrogate data. SMU Data Sci Rev 2018;1(3) [FREE Full text]
29. Nakhwan M, Duangsoithong R. Comparison analysis of data augmentation using bootstrap, GANs and autoencoder. In: Proceedings of the 14th International Conference on Knowledge and Smart Technology. 2022 Presented at: KST 2022; January 26-29, 2022; Chon Buri, Thailand. [doi: 10.1109/kst53302.2022.9729065]
30. Zhao Y, Duangsoithong R. Empirical analysis using feature selection and bootstrap data for small sample size problems. In: Proceedings of the 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. 2019 Presented at: ECTI-CON 2019; July 10-13, 2019; Pattaya, Thailand. [doi: 10.1109/ecti-con47248.2019.8955366]
31. Pedersen PM. Exploring the value of GANs for synthetic tabular data generation in healthcare with a focus on data quality, augmentation, and privacy. Oslo Metropolitan University. 2023. URL: https://oda.oslomet.no/oda-xmlui/handle/11250/3101091#:~:text=While%20various%20methods%20exist%20for,is%20organized%20in%20tabular%20format [accessed 2025-01-01]
32. Ahmadian M, Bodalal Z, van der Hulst HJ, Vens C, Karssemakers LH, Bogveradze N, et al. Overcoming data scarcity in radiomics/radiogenomics using synthetic radiomic features. Comput Biol Med 2024 May;174:108389. [doi: 10.1016/j.compbiomed.2024.108389] [Medline: 38593640]
33. Wang W, Pai TW. Enhancing small tabular clinical trial dataset through hybrid data augmentation: combining SMOTE and WCGAN-GP. Data 2023 Aug 23;8(9):135. [doi: 10.3390/data8090135]
34. Shafquat A, Beigi M, Gao C, Mezey J, Sun J, Aptekar J. An interpretable data augmentation framework for improving generative modeling of synthetic clinical trial data. In: Proceedings of the 3rd Workshop on Interpretable Machine Learning in Healthcare. 2023 Presented at: IMLH 2023; July 28, 2023; Virtual event URL: https://icml.cc/virtual/2023/27750
35. Papadopoulos D, Karalis VD. Variational autoencoders for data augmentation in clinical studies. Appl Sci 2023 Jul 30;13(15):8793. [doi: 10.3390/app13158793]
36. Carey VJ, Lumley TS, Moler C, Ripley B. gee: generalized estimation equation solver. The Comprehensive R Archive Network. 2024 Dec 11. URL: https://cran.r-project.org/web/packages/gee/index.html [accessed 2024-12-12]
37. Oehlert GW. A note on the delta method. Am Stat 1992 Feb;46(1):27. [doi: 10.2307/2684406]
38. Ver Hoef JM. Who invented the delta method? Am Stat 2012 May;66(2):124-127. [doi: 10.1080/00031305.2012.687494]
39. Beltran-Bless AA, Clemons M, Vandermeer L, El Emam K, Ng TL, McGee S, et al. The Rethinking Clinical Trials Program Retreat 2023: creating partnerships to optimize quality cancer care. Curr Oncol 2024 Mar 06;31(3):1376-1388 [FREE Full text] [doi: 10.3390/curroncol31030104] [Medline: 38534937]
40. The Coronary Drug Project Research Group. Practical aspects of decision making in clinical trials: the coronary drug project as a case study. Control Clin Trials 1981 May;1(4):363-376. [doi: 10.1016/0197-2456(81)90041-6] [Medline: 7261627]
41. Degtiar I, Rose S. A review of generalizability and transportability. Annu Rev Stat Appl 2023 Mar 10;10(1):501-524. [doi: 10.1146/annurev-statistics-042522-103837]
42. Ciolino JD, Kaizer AM, Bonner LB. Guidance on interim analysis methods in clinical trials. J Clin Transl Sci 2023 May 15;7(1):e124 [FREE Full text] [doi: 10.1017/cts.2023.552] [Medline: 37313374]
43. Qian Z, Cebere BC, van der Schaar M. Synthcity: facilitating innovative use cases of synthetic data in different data modalities. arXiv Preprint posted online on January 18, 2023 [FREE Full text]
44. El Emam K, Mosquera L, Zheng C. Optimizing the synthesis of clinical trial data using sequential trees. J Am Med Inform Assoc 2021 Jan 15;28(1):3-13 [FREE Full text] [doi: 10.1093/jamia/ocaa249] [Medline: 33186440]
45. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 2006 Sep;15(3):651-674. [doi: 10.1198/106186006X133933]
46. Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. In: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases. 2009 Presented at: ECML PKDD 2009; September 7-11, 2009; Bled, Slovenia. [doi: 10.1007/978-3-642-04174-7_17]
47. Drechsler J, Reiter JP. An empirical evaluation of easily implemented, nonparametric methods for generating synthetic datasets. Comput Stat Data Anal 2011 Dec;55(12):3232-3243. [doi: 10.1016/j.csda.2011.06.006]
48. Arslan RC, Schilling KM, Gerlach TM, Penke L. Using 26,000 diary entries to show ovulatory changes in sexual desire and behavior. J Pers Soc Psychol 2021 Aug;121(2):410-431. [doi: 10.1037/pspp0000208] [Medline: 30148371]

XSL•FO
RenderX

49. Bonnéry D, Feng Y, Henneberger AK, Johnson TL, Lachowicz M, Rose BA, et al. The promise and limitations of synthetic data as a strategy to expand access to state-level multi-agency longitudinal data. J Res Educ Effect 2019 Aug 02;12(4):616-647. [doi: 10.1080/19345747.2019.1631421]
50. Lauger A, Freiman M, Reiter J. Data synthesis and perturbation for the American community survey at the U.S. Census Bureau. United States Census Bureau. 2016. URL: https://www.census.gov/content/dam/Census/newsroom/press-kits/2016/20160803_lauger_data_synthesis_and_perturbation_for_acs.pdf [accessed 2025-02-20]
51. Nowok B. Utility of synthetic microdata generated using tree-based methods. synthpop. 2016. URL: https://synthpop.org.uk/assets/35_nowok_psd2016.pdf [accessed 2025-01-07]
52. Raab GM, Nowok B, Dibben C. Practical data synthesis for large samples. J Priv Confidentiality 2018 Feb 02;7(3):67-97. [doi: 10.29012/jpc.v7i3.407]
53. Nowok B, Raab GM, Dibben C. Providing bespoke synthetic data for the UK longitudinal studies and other sensitive data with the synthpop package for R. Stat J IAOS 2017 Aug 21;33(3):785-796. [doi: 10.3233/sji-150153]
54. Quintana DS. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. Elife 2020 Mar 11;9:9 [FREE Full text] [doi: 10.7554/eLife.53275] [Medline: 32159513]
55. Kaur D, Sobiesk M, Patil S, Liu J, Bhagat P, Gupta A, et al. Application of Bayesian networks to generate synthetic health data. J Am Med Inform Assoc 2021 Mar 18;28(4):801-811 [FREE Full text] [doi: 10.1093/jamia/ocaa303] [Medline: 33367620]
56. Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA: MIT Press; 2012.
57. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 28th International Conference on Neural Information Processing Systems. 2014 Presented at: NIPS'14; December 8-13, 2014; Montreal, QC URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
58. Bourou S, El Saer A, Velivassaki TH, Voulkidis A, Zahariadis T. A review of tabular data synthesis using GANs on an IDS dataset. Information 2021 Sep 14;12(9):375. [doi: 10.3390/info12090375]
59. Xu L, Skoularidou M, Cuesta-Infante A. Modeling tabular data using conditional GAN. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019 Presented at: NIPS'19; December 8-14, 2019; Vancouver, BC.
60. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv Preprint posted online on December 20, 2013 [FREE Full text]
61. Wan Z, Zhang Y, He H. Variational autoencoder based synthetic data generation for imbalanced learning. In: Proceedings of the IEEE Symposium Series on Computational Intelligence. 2017 Presented at: SSCI; November 27-December 1, 2017; Honolulu, HI. [doi: 10.1109/ssci.2017.8285168]
62. Ishfaq H, Hoogi A, Rubin D. TVAE: triplet-based variational autoencoder using metric learning. arXiv Preprint posted online on February 13, 2018 [FREE Full text]
63. Sohn K, Yan X, Lee H. Learning structured output representation using deep conditional generative models. In: Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 2. 2015 Presented at: NIPS'15; December 7-12, 2015; Montreal, QC.
64. Salim AJ. Synthetic patient generation: a deep learning approach using variational autoencoders. arXiv Preprint posted online on August 20, 2018 [FREE Full text] [doi: 10.1017/9781108891530.014]
65. Rubin DB. Discussion: statistical disclosure limitation. J Off Stat 1993;9:461-468.
66. Raghunathan TE, Reiter JP, Rubin DB. Multiple imputation for statistical disclosure limitation. J Off Stat 2003;19(1):1-16 [FREE Full text]
67. Reiter JP. Satisfying disclosure restrictions with synthetic data sets. J Off Stat 2002 Jan;18(4).
68. Reiter JP. Inference for partially synthetic, public use microdata sets. Survey Methodol 2003;29(2):181-188.
69. Loong B, Zaslavsky AM, He Y, Harrington DP. Disclosure control using partially synthetic data for large-scale health surveys, with applications to CanCORS. Stat Med 2013 Oct 30;32(24):4139-4161 [FREE Full text] [doi: 10.1002/sim.5841] [Medline: 23670983]
70. Taub J, Elliot MJ, Sakshaug JW. The impact of synthetic data generation on data utility with application to the 1991 UK samples of anonymised records. Transact Data Privacy 2020 Jan;13(1):1-23 [FREE Full text] [doi: 10.1201/9781315370910-16]
71. El Emam K, Mosquera L, Fang X, El-Hussuna A. An evaluation of the replicability of analyses using synthetic health data. Sci Rep 2024 Mar 24;14(1):6978 [FREE Full text] [doi: 10.1038/s41598-024-57207-7] [Medline: 38521806]
72. El Emam K, Mosquera L, Fang X, El-Hussuna A. Utility metrics for evaluating synthetic health data generation methods: validation study. JMIR Med Inform 2022 Apr 07;10(4):e35734 [FREE Full text] [doi: 10.2196/35734] [Medline: 35389366]
73. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press; 2019.
74. El Kababji S, Mitsakakis N, Fang X, Beltran-Bless AA, Pond G, Vandermeer L, et al. Evaluating the utility and privacy of synthetic breast cancer clinical trial data sets. JCO Clin Cancer Inform 2023 Nov 27;7. [doi: 10.1200/cci.23.00116]

XSL•FO
RenderX

75. Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. Clin Pharmacol Ther 2020 Apr;107(4):817-826 [FREE Full text] [doi: 10.1002/cpt.1633] [Medline: 31541454]
76. Karr AF, Kohnen CN, Oganian A, Reiter JP, Sanil AP. A framework for evaluating the utility of data altered to protect confidentiality. Am Stat 2006 Aug;60(3):224-232. [doi: 10.1198/000313006x124640]
77. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science 2015 Aug 28;349(6251):aac4716. [doi: 10.1126/science.aac4716] [Medline: 26315443]
78. Patil P, Peng RD, Leek JT. What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. Perspect Psychol Sci 2016 Jul 29;11(4):539-544 [FREE Full text] [doi: 10.1177/1745691616646366] [Medline: 27474140]
79. Camerer CF, Dreber A, Holzmeister F, Ho TH, Huber J, Johannesson M, et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nat Hum Behav 2018 Sep 27;2(9):637-644. [doi: 10.1038/s41562-018-0399-z] [Medline: 31346273]
80. Crown W, Dahabreh IJ, Li X, Toh S, Bierer B. Can observational analyses of routinely collected data emulate randomized trials? Design and feasibility of the observational patient evidence for regulatory approval science and understanding disease project. Value Health 2023 Feb;26(2):176-184 [FREE Full text] [doi: 10.1016/j.jval.2022.07.003] [Medline: 35970705]
81. Yoon D, Jeong HE, Park S, You SC, Bang SM, Shin JY. Real-world data emulating randomized controlled trials of non-vitamin K antagonist oral anticoagulants in patients with venous thromboembolism. BMC Med 2023 Sep 29;21(1):375 [FREE Full text] [doi: 10.1186/s12916-023-03069-1] [Medline: 37775786]
82. Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative, Franklin JM, Desai RJ, Feldman W, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. JAMA 2023 Apr 25;329(16):1376-1385 [FREE Full text] [doi: 10.1001/jama.2023.4221] [Medline: 37097356]
83. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies. Circulation 2021 Mar 09;143(10):1002-1013. [doi: 10.1161/circulationaha.120.051718]
84. Clemons M, Dranitsaris G, Sienkiewicz M, Sehdev S, Ng T, Robinson A, et al. A randomized trial of individualized versus standard of care antiemetic therapy for breast cancer patients at high risk for chemotherapy-induced nausea and vomiting. Breast 2020 Dec;54:278-285 [FREE Full text] [doi: 10.1016/j.breast.2020.11.002] [Medline: 33242754]
85. Clemons M, Ong M, Stober C, Ernst S, Booth C, Canil C, et al. A randomised trial of 4- versus 12-weekly administration of bone-targeted agents in patients with bone metastases from breast or castration-resistant prostate cancer. Eur J Cancer 2021 Jan;142:132-140 [FREE Full text] [doi: 10.1016/j.ejca.2020.08.019] [Medline: 33023785]
86. Goss PE, Ingle JN, Pritchard KI, Ellis MJ, Sledge GW, Budd GT, et al. Exemestane versus anastrozole in postmenopausal women with early breast cancer: NCIC CTG MA.27--a randomized controlled phase III trial. J Clin Oncol 2013 Apr 10;31(11):1398-1404 [FREE Full text] [doi: 10.1200/JCO.2012.44.7805] [Medline: 23358971]
87. Paterson AH, Anderson SJ, Lembersky BC, Fehrenbacher L, Falkson CI, King KM, et al. Oral clodronate for adjuvant treatment of operable breast cancer (National Surgical Adjuvant Breast and Bowel Project protocol B-34): a multicentre, placebo-controlled, randomised trial. Lancet Oncol 2012 Jul;13(7):734-742. [doi: 10.1016/s1470-2045(12)70226-7]
88. Clemons M, Fergusson D, Simos D, Mates M, Robinson A, Califaretti N, et al. A multicentre, randomised trial comparing schedules of G-CSF (filgrastim) administration for primary prophylaxis of chemotherapy-induced febrile neutropenia in early stage breast cancer. Ann Oncol 2020 Jul;31(7):951-957 [FREE Full text] [doi: 10.1016/j.annonc.2020.04.005] [Medline: 32325257]
89. Clemons M, Stober C, Kehoe A, Bedard D, MacDonald F, Brunet MC, et al. A randomized trial comparing vascular access strategies for patients receiving chemotherapy with trastuzumab for early-stage breast cancer. Support Care Cancer 2020 Oct 30;28(10):4891-4899. [doi: 10.1007/s00520-020-05326-y] [Medline: 32002617]
90. Gnant M, Mlineritsch B, Schippinger W, Luschin-Ebengreuth G, Pöstlberger S, Menzel C, et al. Endocrine therapy plus zoledronic acid in premenopausal breast cancer. N Engl J Med 2009 Feb 12;360(7):679-691. [doi: 10.1056/nejmoa0806285]
91. Awan A, Ng T, Conter H, Raskin W, Stober C, Simos D, et al. Feasibility outcomes of a randomised, multicentre, pilot trial comparing standard 6-monthly dosing of adjuvant zoledronate with a single one-time dose in patients with early stage breast cancer. J Bone Oncol 2021 Feb;26:100343 [FREE Full text] [doi: 10.1016/j.jbo.2020.100343] [Medline: 33425673]
92. Gralow JR, Barlow WE, Paterson AH, M'iao JL, Lew DL, Stopeck AT, et al. Phase III randomized trial of bisphosphonates as adjuvant therapy in breast cancer: s0307. J Natl Cancer Inst 2020 Jul 01;112(7):698-707 [FREE Full text] [doi: 10.1093/jnci/djz215] [Medline: 31693129]
93. Wang Z, Draghi B, Rotalinti Y, Lunn D, Myles P. High-fidelity synthetic data applications for data augmentation. In: Domínguez-Morales MJ, Civit-Masot J, Damaševičius R, Muñoz-Saavedra L, Damaševičius R, Engelbrecht A, editors. Deep Learning - Recent Findings and Research. London, UK: IntechOpen; 2024.
94. Myles P, Ordish J, Tucker A. The potential synergies between synthetic data and in silico trials in relation to generating representative virtual population cohorts. Prog Biomed Eng 2023 Jan 24;5:013001. [doi: 10.1088/2516-1091/acafbf]

XSL•FO
RenderX

95. Das T, Wang Z, Sun J. TWIN: personalized clinical trial digital twin generation. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023 Presented at: KDD '23; August 6-10, 2023; Long Beach, CA. [doi: 10.1145/3580305.3599534]
96. Gootjes-Dreesbach L, Sood M, Sahay A, Hofmann-Apitius M, Fröhlich H. Variational autoencoder modular Bayesian networks for simulation of heterogeneous clinical study data. Front Big Data 2020 May 28;3:16 [FREE Full text] [doi: 10.3389/fdata.2020.00016] [Medline: 33693390]
97. Akiya I, Ishihara T, Yamamoto K. Comparison of synthetic data generation techniques for control group survival data in oncology clinical trials: simulation study. JMIR Med Inform 2024 Jun 18;12:e55118 [FREE Full text] [doi: 10.2196/55118] [Medline: 38889082]
98. Azizi Z, Zheng C, Mosquera L, Pilote L, El Emam K, GOING-FWD Collaborators. Can synthetic data be a proxy for real clinical trial data? A validation study. BMJ Open 2021 Apr 16;11(4):e043497 [FREE Full text] [doi: 10.1136/bmjopen-2020-043497] [Medline: 33863713]
99. Krenmayr L, Frank R, Drobig C, Braungart M, Seidel J, Schaudt D, et al. GANerAid: realistic synthetic patient data for clinical trials. Inform Med Unlock 2022;35:101118. [doi: 10.1016/j.imu.2022.101118]
100. Beaulieu-Jones BK, Wu ZS, Williams C, Lee R, Bhavnani SP, Byrd JB, et al. Privacy-preserving generative deep neural networks support clinical data sharing. Circ Cardiovasc Qual Outcomes 2019 Jul;12(7):e005122 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.118.005122] [Medline: 31284738]
101. Beigi M, Shafquat A, Mezey J, Aptekar J. Synthetic Clinical Trial Data while Preserving Subject-Level Privacy. In: Proceedings of the Synthetic Data for Empowering ML Research. 2022 Presented at: NeurIPS 2022 SyntheticData4ML; December 2, 2022; New Orleans, LA URL: https://neurips.cc/virtual/2022/58683
102. Eckardt JN, Hahn W, Röllig C, Stasik S, Platzbecker U, Müller-Tidow C, et al. Mimicking clinical trials with synthetic acute myeloid leukemia patients using generative artificial intelligence. NPJ Digit Med 2024 Mar 20;7(1):76 [FREE Full text] [doi: 10.1038/s41746-024-01076-x] [Medline: 38509224]
103. Latner J, Neunhoeffer M, Drechsler J. Generating synthetic data is complicated: know your data and know your generator. In: Proceedings of the International Conference on Privacy in Statistical Databases. 2024 Presented at: PSD 2024; September 25-27, 2024; Antibes Juan-les-Pins, France. [doi: 10.1007/978-3-031-69651-0_8]
104. Fössing E, Drechsler J. An evaluation of synthetic data generators implemented in the python library synthcity. In: Proceedings of the Privacy in Statistical Databases: International Conference. 2024 Presented at: PSD 2024; September 25-27, 2024; Antibes Juan-les-Pins, France. [doi: 10.1007/978-3-031-69651-0_12]
105. Slokom M, Agrawal S, Krol NC, de Wolf PP. Relational or single: a comparative analysis of data synthesis approaches for privacy and utility on a use case from statistical office. In: Proceedings of the International Conference on Privacy in Statistical Databases. 2024 Presented at: PSD 2024; September 25-27, 2024; Antibes Juan-les-Pins, France. [doi: 10.1007/978-3-031-69651-0_27]
106. Huynh L, Johns B, Liu SH, Vedula SS, Li T, Puhan MA. Cost-effectiveness of health research study participant recruitment strategies: a systematic review. Clin Trials 2014 Oct 30;11(5):576-583 [FREE Full text] [doi: 10.1177/1740774514540371] [Medline: 24980278]
107. Moore TJ, Heyward J, Anderson G, Alexander GC. Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015-2017: a cross-sectional study. BMJ Open 2020 Jun 11;10(6):e038863 [FREE Full text] [doi: 10.1136/bmjopen-2020-038863] [Medline: 32532786]
108. Fireman BH, Fehrenbacher L, Gruskin EP, Ray GT. Cost of care for patients in cancer clinical trials. J Natl Cancer Inst 2000 Jan 19;92(2):136-142. [doi: 10.1093/jnci/92.2.136] [Medline: 10639515]
109. Anand A. Study reveals serious cancer research gaps for Black Canadians. CBC News. 2023 Feb 7. URL: https://www.cbc.ca/news/canada/ottawa/black-canadian-cancer-research-gap-ottawa-problem-1.6737797 [accessed 2024-02-07]
110. Cénat JM, Dromer É, Darius WP, Dalexis RD, Furyk SE, Poisson H, et al. Incidence, factors, and disparities related to cancer among Black individuals in Canada: a scoping review. Cancer 2023 Feb 01;129(3):335-355 [FREE Full text] [doi: 10.1002/cncr.34551] [Medline: 36436148]
111. Gross AS, Harry AC, Clifton CS, Della Pasqua O. Clinical trial diversity: an opportunity for improved insight into the determinants of variability in drug response. Br J Clin Pharmacol 2022 Jun;88(6):2700-2717 [FREE Full text] [doi: 10.1111/bcp.15242] [Medline: 35088432]
112. Ford JG, Howerton MW, Lai GY, Gary TL, Bolen S, Gibbons MC, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer 2008 Jan 15;112(2):228-242 [FREE Full text] [doi: 10.1002/cncr.23157] [Medline: 18008363]
113. Khan MS, Shahid I, Siddiqi TJ, Khan SU, Warraich HJ, Greene SJ, et al. Ten-year trends in enrollment of women and minorities in pivotal trials supporting recent US Food and Drug Administration approval of novel cardiometabolic drugs. J Am Heart Assoc 2020 Jun 02;9(11):e015594 [FREE Full text] [doi: 10.1161/JAHA.119.015594] [Medline: 32427023]
114. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. Health Soc Care Community 2004 Sep;12(5):382-388. [doi: 10.1111/j.1365-2524.2004.00507.x] [Medline: 15373816]

XSL•FO
RenderX

115. Evelyn B, Toigo T, Banks D, Pohl D, Gray K, Robins B, et al. Participation of racial/ethnic groups in clinical trials and race-related labeling: a review of new molecular entities approved 1995-1999. J Natl Med Assoc 2001 Dec;93(12 Suppl):18S-24S. [Medline: 11798060]
116. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/ajph.2013.301706]
117. UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. J Gen Intern Med 2007 Jun 21;22(6):852-863 [FREE Full text] [doi: 10.1007/s11606-007-0126-3] [Medline: 17375358]
118. Ramamoorthy A, Pacanowski MA, Bull J, Zhang L. Racial/ethnic differences in drug disposition and response: review of recently approved drugs. Clin Pharmacol Ther 2015 Mar 20;97(3):263-273. [doi: 10.1002/cpt.61] [Medline: 25669658]
119. Tannenbaum C, Day D, Matera Alliance. Age and sex in drug development and testing for adults. Pharmacol Res 2017 Jul;121:83-93 [FREE Full text] [doi: 10.1016/j.phrs.2017.04.027] [Medline: 28455265]
120. Vitale C, Fini M, Spoletini I, Lainscak M, Seferovic P, Rosano GM. Under-representation of elderly and women in clinical trials. Int J Cardiol 2017 Apr 01;232:216-221. [doi: 10.1016/j.ijcard.2017.01.018] [Medline: 28111054]
121. Ruiter R, Burggraaf J, Rissmann R. Under-representation of elderly in clinical trials: an analysis of the initial approval documents in the Food and Drug Administration database. Br J Clin Pharmacol 2019 Apr 23;85(4):838-844 [FREE Full text] [doi: 10.1111/bcp.13876] [Medline: 30681181]
122. Crome P, Cherubini A, Oristrell J. The PREDICT (increasing the participation of the elderly in clinical trials) study: the charter and beyond. Expert Rev Clin Pharmacol 2014 Jul;7(4):457-468. [doi: 10.1586/17512433.2014.922864] [Medline: 24939469]
123. Heidari S, Babor TF, de Castro P, Tort S, Curno M. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev 2016;1:2 [FREE Full text] [doi: 10.1186/s41073-016-0007-6] [Medline: 29451543]
124. Acharya KP, Pathak S. Applied research in low-income countries: why and how? Front Res Metr Anal 2019 Nov 14;4:3 [FREE Full text] [doi: 10.3389/frma.2019.00003] [Medline: 33870035]
125. Bhanot K, Qi M, Erickson JS, Guyon I, Bennett KP. The problem of fairness in synthetic healthcare data. Entropy (Basel) 2021 Sep 04;23(9):1165 [FREE Full text] [doi: 10.3390/e23091165] [Medline: 34573790]
126. Wyllie S, Shumailov I, Papernot N. Fairness feedback loops: training on synthetic data amplifies bias. arXiv Preprint posted online on March 12, 2024 [FREE Full text] [doi: 10.1145/3630106.3659029]
127. Juwara L, El-Hussuna A, El Emam K. An evaluation of synthetic data augmentation for mitigating covariate bias in health data. Patterns (N Y) 2024 Apr 12;5(4):100946 [FREE Full text] [doi: 10.1016/j.patter.2024.100946] [Medline: 38645766]
128. Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. J Am Med Inform Assoc 2024 Apr 19;31(5):1172-1183. [doi: 10.1093/jamia/ocae060] [Medline: 38520723]
129. Theodorou B, Danek B, Tummala V, Kumar SP, Malin B, Sun J. FairPlay: improving medical machine learning models with generative balancing for equity and excellence. Research Square Preprint posted online on December 13, 2024 [FREE Full text] [doi: 10.21203/rs.3.rs-5252769/v1]
130. Landscape report and industry survey on the use of computational modeling and simulation in medical device development. Medical Device Innovation Consortium. 2024 Feb 16. URL: https://mdic.org/resources/landscape-report-industry-survey-on-the-use-of-computational-modeling-simulation-in-medical-device-development/ [accessed 2025-01-07]
131. Redrup Hill E, Mitchell C, Myles P, Branson R, Frangi AF. Cross-regulator workshop: journeys, experiences and best practices on computer modelled and simulated regulatory evidence— workshop report. InSilicoUK Pro-Innovation Regulations Network. 2023. URL: https://www.pankhurst.manchester.ac.uk/wp-content/uploads/sites/278/2024/01/Cross_Regulator_Workshop_on_CMS_Report_December_2023_final.pdf [accessed 2025-01-07]
132. Viceconti M, Emili L. Toward Good Simulation Practice: Best Practices for the Use of Computational Modelling and Simulation in the Regulatory Process of Biomedical Products. Cham, Switzerland: Springer; 2024.
133. Warraich HJ, Tazbaz T, Califf RM. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. JAMA 2025 Jan 21;333(3):241-247. [doi: 10.1001/jama.2024.21451] [Medline: 39405330]
134. Artificial intelligence: an accountability framework for federal agencies and other entities. General Accountability Office. 2021 Jun 30. URL: https://www.gao.gov/products/gao-21-519sp [accessed 2024-02-07]
135. Using artificial intelligence and machine learning in the development of drug and biological products. U.S. Food & Drug Administration. URL: https://www.fda.gov/media/167973/download [accessed 2024-02-07]
136. Alloza C, Knox B, Raad H, Aguilà M, Coakley C, Mohrova Z, et al. A case for synthetic data in regulatory decision-making in Europe. Clin Pharmacol Ther 2023 Oct 24;114(4):795-801. [doi: 10.1002/cpt.3001] [Medline: 37441734]
137. Myles P, Ordish J, Branson R. Synthetic data and the innovation, assessment, and regulation of AI medical devices. Clinical Practice Research Datalink. 2022. URL: https://www.cprd.com/sites/default/files/2022-02/Synthetic%20data_RA%20proof.pdf [accessed 2024-02-07]

XSL•FO
RenderX

138. Newington L, Metcalfe A. Factors influencing recruitment to research: qualitative study of the experiences and perceptions of research teams. BMC Med Res Methodol 2014 Jan 23;14(1):10 [FREE Full text] [doi: 10.1186/1471-2288-14-10] [Medline: 24456229]
139. Price D, Edwards M, Carson-Stevens A, Cooper A, Davies F, Evans B, et al. Challenges of recruiting emergency department patients to a qualitative study: a thematic analysis of researchers' experiences. BMC Med Res Methodol 2020 Jun 11;20(1):151 [FREE Full text] [doi: 10.1186/s12874-020-01039-2] [Medline: 32522265]
140. Friede T, Posch M, Zohar S, Alberti C, Benda N, Comets E, et al. Recent advances in methodology for clinical trials in small populations: the InSPiRe project. Orphanet J Rare Dis 2018 Oct 25;13(1):186 [FREE Full text] [doi: 10.1186/s13023-018-0919-y] [Medline: 30359266]
141. Reilly K. Sharing data for the benefit of children with cancer. BioMed Central. 2019 Sep 9. URL: https://blogs.biomedcentral.com/on-medicine/2019/09/09/sharing-data-for-the-benefit-of-children-with-cancer/ [accessed 2020-10-02]
142. Bavdekar SB. Pediatric clinical trials. Perspect Clin Res 2013 Jan;4(1):89-99 [FREE Full text] [doi: 10.4103/2229-3485.106403] [Medline: 23533990]
143. Campbell H, Surry SA, Royle EM. A review of randomised controlled trials published in Archives of Disease in Childhood from 1982-96. Arch Dis Child 1998 Aug 01;79(2):192-197 [FREE Full text] [doi: 10.1136/adc.79.2.192] [Medline: 9797608]
144. Silverman WA, Altman DG. Patients' preferences and randomised trials. Lancet 1996 Jan 20;347(8995):171-174 [FREE Full text] [doi: 10.1016/s0140-6736(96)90347-5] [Medline: 8544555]
145. Principles for responsible clinical trial data sharing: our commitment to patients and researchers. European Federation of Pharmaceutical Industries and Associations. 2023. URL: https://www.efpia.eu/media/qndlfduy/phrmaefpiaprinciplesforresponsibledatasharing2023.pdf [accessed 2024-02-07]
146. European Medicines Agency policy on publication of clinical data for medicinal products for human use. European Medicines Agency. 2019 Mar 21. URL: https://www.ema.europa.eu/en/documents/other/policy-70-european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf [accessed 2024-01-11]
147. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. Ann Intern Med 2016 Jan 26;164(7):505-506. [doi: 10.7326/m15-2928]
148. Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: The National Academies Press; 2015.
149. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors. 2015 Jan. URL: https://www.icmje.org/icmje-recommendations.pdf [accessed 2024-02-07]
150. Data, software and materials management and sharing policy. Wellcome. URL: https://wellcome.org/grant-funding/guidance/policies-grant-conditions/data-software-materials-management-and-sharing-policy [accessed 2024-02-07]
151. Final NIH statement on sharing research data. National Institutes of Health. 2003 Feb 26. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html [accessed 2020-06-29]
152. Protection of personal data in clinical documents – a model approach. TransCelerate Biopharma Inc. 2016. URL: http://www.transceleratebiopharmainc.com/wp-content/uploads/2017/02/Protection-of-Personal-Data-in-Clinical-Documents.pdf [accessed 2024-02-07]
153. Data de-identification and anonymization of individual patient data in clinical studies — a model approach. TransCelerate Biopharma Inc. 2016. URL: http://www.transceleratebiopharmainc.com/wp-content/uploads/2015/04/TransCelerate-De-identification-and-Anonymization-of-Individual-Patient-Data-in-Clinical-Studies-V2.0.pdf [accessed 2024-02-07]
154. Doshi P. Data too important to share: do those who control the data control the message? BMJ 2016 Mar 02;352:i1027. [doi: 10.1136/bmj.i1027] [Medline: 26936297]
155. Rankin D, Black M, Bond R, Wallace J, Mulvenna M, Epelde G. Reliability of supervised machine learning using synthetic data in health care: model to preserve privacy for data sharing. JMIR Med Inform 2020 Jul 20;8(7):e18910 [FREE Full text] [doi: 10.2196/18910] [Medline: 32501278]
156. Ventresca M, Schünemann HJ, Macbeth F, Clarke M, Thabane L, Griffiths G, et al. Obtaining and managing data sets for individual participant data meta-analysis: scoping review and practical guide. BMC Med Res Methodol 2020 May 12;20(1):113 [FREE Full text] [doi: 10.1186/s12874-020-00964-6] [Medline: 32398016]
157. Polanin JR. Efforts to retrieve individual participant data sets for use in a meta-analysis result in moderate data sharing but many data sets remain missing. J Clin Epidemiol 2018 Jun;98:157-159. [doi: 10.1016/j.jclinepi.2017.12.014] [Medline: 29288135]
158. Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in The BMJ and PLOS Medicine. BMJ 2018 Feb 13;360:k400 [FREE Full text] [doi: 10.1136/bmj.k400] [Medline: 29440066]

XSL•FO
RenderX

159. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. BMJ 2017 Apr 05;357:j1390 [FREE Full text] [doi: 10.1136/bmj.j1390] [Medline: 28381561]
160. Villain B, Dechartres A, Boyer P, Ravaud P. Feasibility of individual patient data meta-analyses in orthopaedic surgery. BMC Med 2015 Jun 03;13(1):131 [FREE Full text] [doi: 10.1186/s12916-015-0376-6] [Medline: 26040278]
161. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible research practices and transparency across the biomedical literature. PLoS Biol 2016 Jan 4;14(1):e1002333 [FREE Full text] [doi: 10.1371/journal.pbio.1002333] [Medline: 26726926]
162. Expert advisory group report 2: building Canada's health data foundation. Government of Canada. 2021 Nov. URL: https://tinyurl.com/zcsakfv9 [accessed 2024-02-07]
163. Read KB, Ganshorn H, Rutley S, Scott DR. Data-sharing practices in publications funded by the Canadian Institutes of Health Research: a descriptive analysis. CMAJ Open 2021 Nov 09;9(4):E980-E987 [FREE Full text] [doi: 10.9778/cmajo.20200303] [Medline: 34753787]
164. National Academy of Medicine. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: The National Academies Press; 2022.
165. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014 Nov 05;14:1144 [FREE Full text] [doi: 10.1186/1471-2458-14-1144] [Medline: 25377061]
166. Kalkman S, Mostert M, Gerlinger C, van Delden JJ, van Thiel GJ. Responsible data sharing in international health research: a systematic review of principles and norms. BMC Med Ethics 2019 Mar 28;20(1):21 [FREE Full text] [doi: 10.1186/s12910-019-0359-9] [Medline: 30922290]
167. Reiter JP. New approaches to data dissemination: a glimpse into the future (?). CHANCE 2012 Sep 20;17(3):11-15. [doi: 10.1080/09332480.2004.10554907]
168. Park N, Mohammadi M, Gorde K, Jajodia S, Park H, Kim Y. Data synthesis based on generative adversarial networks. Proc VLDB Endow 2018 Jun 01;11(10):1071-1083. [doi: 10.14778/3231751.3231757]
169. Hu J. Bayesian estimation of attribute and identification disclosure risks in synthetic data. arXiv Preprint posted online on April 9, 2018 [FREE Full text]
170. Taub J, Elliot M, Pampaka M, Smith D. Differential correct attribution probability for synthetic data: an exploration. In: Proceedings of the Privacy in Statistical Databases. 2018 Presented at: PSD 2018; September 26-28, 2018; Valencia, Spain. [doi: 10.1007/978-3-319-99771-1_9]
171. Hu J, Reiter JP, Wang Q. Disclosure risk evaluation for fully synthetic categorical data. In: Proceedings of the Privacy in Statistical Databases. 2014 Presented at: PSD 2014; September 17-19, 2014; Ibiza, Spain. [doi: 10.1007/978-3-319-11257-2_15]
172. Wei L, Reiter JP. Releasing synthetic magnitude microdata constrained to fixed marginal totals. Stat J IAOS 2016 Feb 27;32(1):93-108. [doi: 10.3233/sji-160959]
173. Ruiz N, Muralidhar K, Domingo-Ferrer J. On the privacy guarantees of synthetic data: a reassessment from the maximum-knowledge attacker perspective. In: Proceedings of the Privacy in Statistical Databases. 2018 Presented at: PSD 2018; September 26-28, 2018; Valencia, Spain. [doi: 10.1007/978-3-319-99771-1_5]
174. Reiter JP. Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study. J R Stat Soc Ser A Stat Soc 2005 Jan;168(1):185-205. [doi: 10.1111/j.1467-985X.2004.00343.x]
175. El Emam K, Mosquera L, Fang X. Validating a membership disclosure metric for synthetic health data. JAMIA Open 2022 Dec;5(4):ooac083 [FREE Full text] [doi: 10.1093/jamiaopen/ooac083] [Medline: 36238080]
176. El Emam K, Mosquera L, Bass J. Evaluating identity disclosure risk in fully synthetic health data: model development and validation. J Med Internet Res 2020 Nov 16;22(11):e23139 [FREE Full text] [doi: 10.2196/23139] [Medline: 33196453]
177. Arora A, Wagner SK, Carpenter R, Jena R, Keane PA. The urgent need to accelerate synthetic data privacy frameworks for medical research. Lancet Digit Health 2025 Feb;7(2):e157-e160. [doi: 10.1016/s2589-7500(24)00196-1]
178. El Emam K, Mosquera L, Hoptroff R. Practical Synthetic Data Generation: Balancing Privacy and the Broad Availability of Data. Sebastopol, CA: O'Reilly Media; 2020.
179. Beduschi A. Synthetic data protection: towards a paradigm change in data regulation? Big Data Soc 2024 Feb 14;11(1). [doi: 10.1177/20539517241231277]
180. Lun R, Siegal D, Ramsay T, Stotts G, Dowlatshahi D. Synthetic data in cancer and cerebrovascular disease research: a novel approach to big data. PLoS One 2024 Feb 7;19(2):e0295921 [FREE Full text] [doi: 10.1371/journal.pone.0295921] [Medline: 38324588]
181. Rousseau O, Karakachoff M, Gaignard A, Bellanger L, Bijlenga P, Constant Dit Beaufils P, et al. Location of intracranial aneurysms is the main factor associated with rupture in the ICAN population. J Neurol Neurosurg Psychiatry 2021 Feb 23;92(2):122-128. [doi: 10.1136/jnnp-2020-324371] [Medline: 33097563]
182. Guillaudeux M, Rousseau O, Petot J, Bennis Z, Dein CA, Goronflot T, et al. Patient-centric synthetic data generation, no reason to risk re-identification in biomedical data analysis. NPJ Digit Med 2023 Mar 10;6(1):37 [FREE Full text] [doi: 10.1038/s41746-023-00771-5] [Medline: 36899082]

XSL•FO
RenderX

183. Thomas A, Jaffré S, Guardiolle V, Perennec T, Gagnadoux F, Goupil F, et al. Does PaCO correction have an impact on survival of patients with chronic respiratory failure and long-term non-invasive ventilation? Heliyon 2024 Feb 29;10(4):e26437 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e26437] [Medline: 38420381]
184. pysdg. Open Science Framework Home. 2024. URL: https://osf.io/xj9pr/ [accessed 2025-02-20]

Abbreviations

ABCSG: Austrian Breast and Colorectal Cancer Study Group
BTA: bone-targeted agents
CTGAN: conditional tabular generative adversarial network
GAN: generative adversarial network
HER2+: human epidermal growth factor receptor-2 positive
REaCT: Rethinking Clinical Trials
RWD: real-world data
VAE: variational autoencoder
ZOL: zoledronate

Edited by A Schwartz; submitted 24.09.24; peer-reviewed by PA Gourraud, AK Vadathya, A Bhattacharya; comments to author 23.10.24; revised version received 27.12.24; accepted 31.01.25; published 05.03.25.

Please cite as:
El Kababji S, Mitsakakis N, Jonker E, Beltran-Bless AA, Pond G, Vandermeer L, Radhakrishnan D, Mosquera L, Paterson A, Shepherd L, Chen B, Barlow W, Gralow J, Savard MF, Fesl C, Hlauschek D, Balic M, Rinnerthaler G, Greil R, Gnant M, Clemons M, El Emam K
Augmenting Insufficiently Accruing Oncology Clinical Trials Using Generative Models: Validation Study
J Med Internet Res 2025;27:e66821
URL: https://www.jmir.org/2025/1/e66821
doi:10.2196/66821
PMID:40053790

©Samer El Kababji, Nicholas Mitsakakis, Elizabeth Jonker, Ana-Alicia Beltran-Bless, Gregory Pond, Lisa Vandermeer, Dhenuka Radhakrishnan, Lucy Mosquera, Alexander Paterson, Lois Shepherd, Bingshu Chen, William Barlow, Julie Gralow, Marie-France Savard, Christian Fesl, Dominik Hlauschek, Marija Balic, Gabriel Rinnerthaler, Richard Greil, Michael Gnant, Mark Clemons, Khaled El Emam. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perspectives on Using Artificial Intelligence to Derive Social Determinants of Health Data From Medical Records in Canada: Large Multijurisdictional Qualitative Study

Victoria H Davis[1], MSc; Jinfan Rose Qiang[2], BSc; Itunuoluwa Adekoya MacCarthy[2], MSc; Dana Howse[3], PhD; Abigail Zita Seshie[2], PhD; Leanne Kosowan[4], MSc; Alannah Delahunty-Pike[5], MSc; Eunice Abaga[2], MEd, MPA; Jane Cooney[2], MLS; Marjeiry Robinson[2]; Dorothy Senior[2], LPN; Alexander Zsager[2], BA; Kris Aubrey-Bassler[3], CCFP, MD, MSc; Mandi Irwin[5], MD, CCFP; Lois A Jackson[6], MA, PhD; Alan Katz[4], CCFP, MBChB, MSc; Emily Gard Marshall[5], MSc, PhD; Nazeem Muhajarine[7], PhD; Cory Neudorf[7], MD, MHSc; Stephanie Garies[8], MPH, PhD; Andrew D Pinto[2,9,10,11], CCFP, MSc, MD

[1]Department of Health Behavior and Health Equity, School of Public Health, University of Michigan–Ann Arbor, Ann Arbor, MI, United States

[2]Upstream Lab, MAP Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, Unity Health Toronto, Toronto, ON, Canada

[3]Primary Healthcare Research Unit, Memorial University of Newfoundland and Labrador, St. John's, NL, Canada

[4]Department of Family Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

[5]Department of Family Medicine, Dalhousie University, Halifax, NS, Canada

[6]School of Health and Human Performance, Dalhousie University, Halifax, NS, Canada

[7]Department of Community Health & Epidemiology, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada

[8]Department of Family Medicine, University of Calgary, Calgary, Canada

[9]Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[10]Department of Family and Community Medicine, St. Michael’s Hospital, Toronto, ON, Canada

[11]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Andrew D Pinto, CCFP, MSc, MD
Upstream Lab, MAP Centre for Urban Health Solutions
Li Ka Shing Knowledge Institute
Unity Health Toronto
30 Bond Street
Toronto, ON, M5B 1W8
Canada
Phone: 1 416 864 6060 ext 76148
Email: andrew.pinto@utoronto.ca

Abstract

Background: Data on the social determinants of health could be used to improve care, support quality improvement initiatives, and track progress toward health equity. However, this data collection is not widespread. Artificial intelligence (AI), specifically natural language processing and machine learning, could be used to derive social determinants of health data from electronic medical records. This could reduce the time and resources required to obtain social determinants of health data.

Objective: This study aimed to understand perspectives of a diverse sample of Canadians on the use of AI to derive social determinants of health information from electronic medical record data, including benefits and concerns.

Methods: Using a qualitative description approach, in-depth interviews were conducted with 195 participants purposefully recruited from Ontario, Newfoundland and Labrador, Manitoba, and Saskatchewan. Transcripts were analyzed using an inductive and deductive content analysis.

Results: A total of 4 themes were identified. First, AI was described as the inevitable future, facilitating more efficient, accessible social determinants of health information and use in primary care. Second, participants expressed concerns about potential health care harms and a distrust in AI and public systems. Third, some participants indicated that AI could lead to a loss of the human

XSL•FO
RenderX

touch in health care, emphasizing a preference for strong relationships with providers and individualized care. Fourth, participants described the critical importance of consent and the need for strong safeguards to protect patient data and trust.

Conclusions: These findings provide important considerations for the use of AI in health care, and particularly when health care administrators and decision makers seek to derive social determinants of health data.

(J Med Internet Res 2025;27:e52244) doi:10.2196/52244

KEYWORDS

artificial intelligence; social determinants of health; sociodemographic data; social needs; social care; primary care; machine learning; qualitative study

Introduction

The social determinants of health (SDoH), people's daily living and working conditions that are influenced by policies and structures (eg, racism and housing) [1], contribute to systemic and avoidable health inequities across groups [2,3]. These resource and power-related determinants contribute to access to high-quality health care services and delivery and widen the gap in health outcomes across sociocultural groups [4-6]. For example, people with lower-incomes in the United States often experience barriers to access to care due in part to gaps in health insurance, and many Black individuals have experienced health care discrimination and worse postoperative care outcomes than White individuals due to structural racism [7-10]. Income inequities persist in access to primary and specialist care in Canada despite universal health care [11,12].

Primary health care is at the nexus of medical care, public health, and community and social services [13-15], and the collection of SDoH data in primary care is essential to identify and tackle inequities, which contribute to poor health outcomes [4,16,17]. These data could be used to improve care, help patients with their social and financial situations through coordination to local services [17-20], and guide health care changes and public policy [17,19-21]. However, it is challenging to operationalize SDoH data in real-world clinical settings. Practical and technological challenges including a lack of a unified and standardized measurement of SDoH [22], electronic medical record (EMR) system variabilities [23-25], and finite health care system capacity, can lead to limited data interoperability and reduced power to inform health care systems and public health policies [23-25]. In addition, it is important to consider the health care team's capacity with additional workload associated with administrating, collecting, documenting, and responding to needs, which lead to increasing risk of burnout [21,26-29].

Artificial intelligence (AI) can leverage the potential benefits of SDoH data to identify patient needs [30,31] and respond to health care overcapacity and disease complexity [32,33]. Machine learning is a common type of AI used to detect, predict, and categorize outcomes by looking for patterns in the data that are associated with known observations or "ground truth" cases [34,35]. Although this work is still in exploratory stages, machine learning and natural language processing have demonstrated the feasibility of detecting a number of SDoH from EMR data, such as childhood experiences [36], social connections [37], living situation [31,36], employment [31], and predicting health outcomes based on the SDoH [25].

While there is strong potential for adopting AI to identify SDoH data, there are multiple potential harms. Concerns with using AI technology in health care include the exacerbation of biases and inequities, discrimination [32], data security and privacy, and lack of infrastructure [38-40]. For example, the datasets themselves may reflect discriminative or biased practices, in which the algorithms are trained to learn [32,41]. Unstructured physician notes may contain biases which are similarly replicated to determine patterns and impact patient outcomes [32,41].

The perspectives of patients and the general public (who have been or may become patients) are critical to shaping decisions surrounding the implementation of AI in health care. Patient data are used to develop AI algorithms and patients are affected by having AI inform their care [42]. Despite its importance, few articles focus on public or patient perspectives on AI in health care [38-40,42]. A scoping review of 37 articles (grey literature and peer-reviewed) found that many patients reported positive views on AI, although views may differ based on patients' experiences, concerns, and trust [42]. However, there is limited knowledge of participants' perspectives on using AI to derive their SDoH information using existing EMR data, which may elicit different views due to the sensitive nature of this information. Given the widespread adoption of EMRs in primary care settings, the sensitivity of SDoH data, and the potential harms associated with AI in health care, it is essential to learn viewpoints on whether and how AI could be implemented in an equitable, acceptable, and safe manner. This qualitative study aimed to understand the perspectives of a diverse sample of Canadians on using AI to derive SDoH information from existing EMR data in primary care, including potential benefits and concerns.

Methods

Study Background

This study was conducted as part of a multicomponent project that developed and refined a standardized SDoH questionnaire for primary care settings, known as the Screening for Poverty And Related Social Determinants and Intervening to Improve Knowledge of and Links to Resources (SPARK) Tool (Multimedia Appendix 1) [43]. Details on the SPARK Tool and the broader project are provided elsewhere [43]. For this paper, we report on data gathered from in-depth interviews with participants related to their perspectives on having AI derive SDoH data from the EMR. Participants became aware of the

XSL•FO
RenderX

SDoH before in-depth interviews by completing the SPARK Tool and were provided with examples of determinants.

Study Design, Setting, and Sampling Approach

A qualitative description approach was chosen to provide a comprehensive summary of participants' experiences and preferences on using AI to derive SDoH data in primary care settings [44]. One-on-one, semistructured interviews were conducted using video teleconference software across 4 Canadian provinces: Ontario, Newfoundland and Labrador, Saskatchewan, and Manitoba.

Maximum variation sampling helped to ensure that a diverse sample of participants were included across ages, races and ethnicities, location (rural or city), languages, and gender identities where possible [45]. After indicating interest in the study, potential participants were asked about these characteristics, which enabled the purposive selection of individuals who were not initially well-represented in the sample. Adults (aged 18 years and older) were recruited through social media advertisements, email distribution lists, and posters in community and health centers [43]. Advertisements were translated based on the 3 most commonly spoken languages in each province other than English and French, although interviews were only conducted in English due to a lack of non-English speakers contacting the study team. The study aimed to recruit a sample of approximately 200 individuals, with a greater number of participants from Ontario due to the size of the province, to ensure adequate diversity across the multiple domains of the SPARK Tool and to inform the multicomponent study objectives and sampling approach.

Data Collection

The interviewers were female research staff [DH, AD-P, AZS, LK, and IAM] or research assistants, and all received training or had previous experience with qualitative interviewing. The study was also informed regularly through a national advisory group composed of patient partners, scientists, and other collaborators. Data collection occurred from April 2021 to January 2022. The interview guide was established and iteratively revised through meetings with the research team and participant feedback to ensure that the questions were comprehensible for participants without a background in AI. Participants provided verbal informed consent before participating in the study, in accordance with research ethics board requirements. Participants were asked their perspectives on having secure software access patients' records to derive their SDoH information, as opposed to asking patients for this information directly (refer to Multimedia Appendix 2 for interview questions). Follow-up questions included examining possible benefits or concerns with the use of this software.

Data Analysis

Interviews were audio-recorded and transcribed verbatim using a professional service. NVivo (version 12; Lumivero) software was used for data management. The codebook was developed collaboratively by the interviewers and the analysis team (VHD, JRQ, IAM, DH, AZS, LK, AD-P, and ADP). Two study team members (VHD and JRQ) with qualitative research experience led the analysis. An initial codebook outline was created by 2 of the interviewers (AD-P and IAM) during independent review of 2 transcripts. This outline helped to guide the development of a preliminary codebook, which was developed by the lead analysts after examining 10 randomly selected transcripts, including the 2 previously examined transcripts (4 from Ontario, 2 each from Newfoundland and Labrador, Saskatchewan, and Manitoba). The remaining analysis team members reviewed the same 10 transcripts and provided detailed feedback on the preliminary codebook. After the codebook was revised and agreed upon, 2 study team members continued to refine it by reviewing 2 additional transcripts each. Codes were compared for consistency and alignment, and extensive documentation was created to outline decisions made during multiple meetings between the coders. The codebook continued to undergo minor edits as more transcripts were reviewed. After 12 transcripts were thoroughly reviewed and examined between the 2 coders, the remaining interviews were randomly split within each province between the coders.

A qualitative content analysis was conducted to focus on theme development across participant interviews [46-48]. A combined inductive and deductive approach was used to enable flexibility to incorporate emerging codes and findings from transcripts, while also focusing on codes pertaining to the research and interview questions.

The researchers familiarized themselves with the transcripts before analysis by thoroughly reading the contents. For the inductive coding approach, the transcripts were read and meaning units (chunks of data, such as sentences or a paragraph) were selected and coded without a predetermined plan [46,47]. For the deductive coding approach, codes specific to potential benefits, challenges or concerns with using AI to derive SDoH information were created before analyzing the data based on the research aims and interview questions. Meaning units were assigned to these predetermined codes, and codes were grouped into categories and themes in an iterative process. Throughout the data analysis, regular meetings were scheduled with the analysis team to discuss updates, preliminary findings, and assist with the context and interpretation of codes into themes. Figure 1 provides a simplified description of the analysis workflow based on the methodology by Elo and Kyngäs [48].

XSL•FO
RenderX

Figure 1. Qualitative content analysis based on the methodology by Elo and Kyngäs [48].

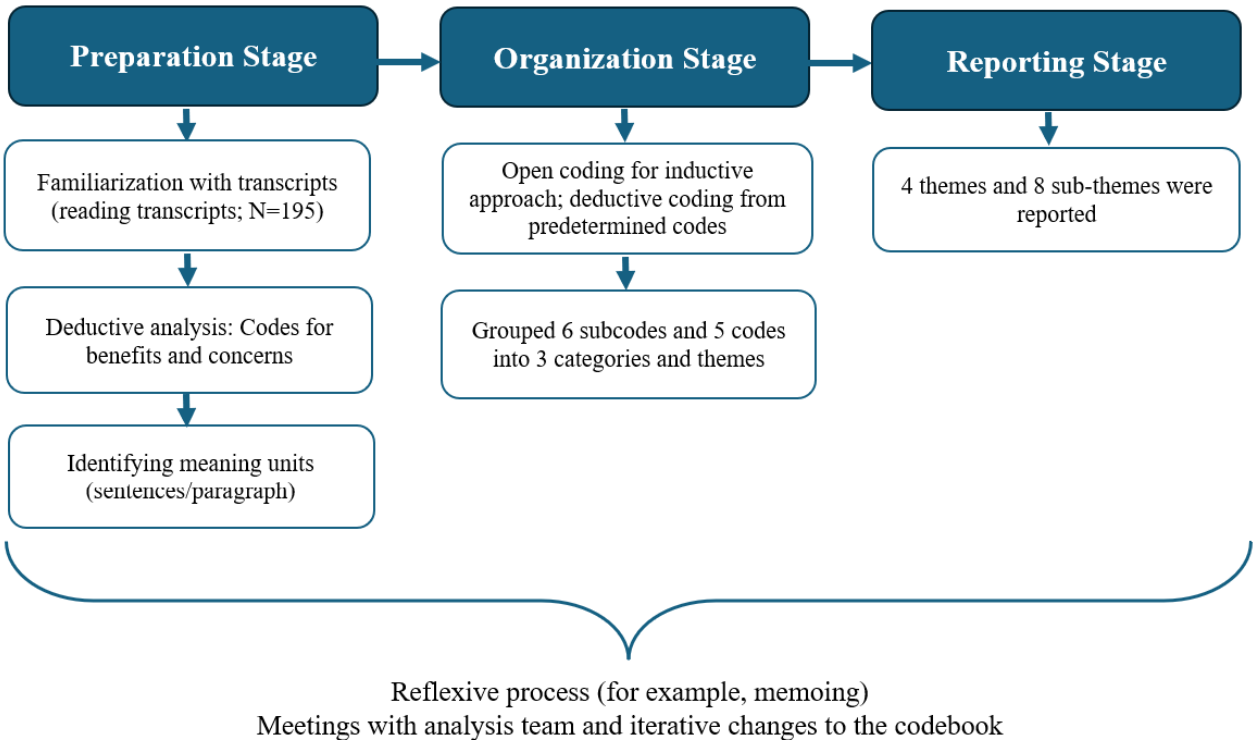

Positionality, Reflexivity, and Trustworthiness

The 5 main interviewers (IA, DH, AZS, LK, and AD-P) reside in different provinces and have different backgrounds and perspectives. The team regularly met to discuss the interviews and the project both while the interviews were being conducted and when the analysis was occurring, alongside VHD, JRQ, ADP, and at times EA. This team approach helped to mitigate any 1 perspective from dominating the interviews and analysis, as a form of data triangulation to enhance the criteria of trustworthiness [49].

The 2 main analysts (VHD and JRQ) engaged in reflexivity to understand how their own biases and preconceptions could influence how the data were approached and analyzed, through regular journaling and discussions before and throughout the analysis [49,50]. They also took memos upon analyzing interviews to document early findings [49,50]. For example, they documented their own beliefs about whether AI should be used to derive SDoH information, which became more nuanced based on the benefits and concerns expressed by participants. Both analysts created extensive documentation outlining the processes and decisions for the coding approach and analysis [49].

Ethical Considerations

This study was approved by the Unity Health Toronto Research Ethics Board (#20-241) Saskatchewan Behavioural Research Ethics Board (#2373), Newfoundland and Labrador Health Research Ethics Board (#2020.259), and University of Manitoba Health Research Ethics Board (#HS24204).

Results

Overview

There were 195 interviews conducted across the 4 provinces, lasting approximately 30-45 minutes each. Most participants lived in Ontario (124/195, 64%), identified as women (126/195, 65%), and non-White (122/169, 72% among those who disclosed their race and ethnicity), while 38% (75/195) of participants reported at least 1 unmet social need (Table 1).

XSL•FO
RenderX

Table 1. Participant demographics (N=195; adapted from Adekoya et al [43], which is published under Creative Commons Attribution 4.0 International License [51]).

Characteristics	Participants, n (%)[a]
Canadian province	
Ontario	125 (64)
Saskatchewan	25 (13)
Manitoba	24 (12)
Newfoundland and Labrador	21 (11)
Race and ethnicity	
Asian	71 (36)
Black	17 (9)
Indigenous	8 (4)
White	47 (24)
Other	10 (5)
Multiracial	16 (8)
Data not collected (Manitoba) or no response[b]	26 (12)
Gender	
Man	58 (30)
Woman	126 (65)
Transgender, gender fluid, or nonbinary	9 (5)
No response	≤5
Sex at birth	
Male	58 (30)
Female	133 (68)
Intersex	≤5
No response	≤5
One or more unmet social needs[c]	
Yes	75 (38)
No	120 (62)
Difficulty making ends meet	
Yes	43 (22)
No	151 (77)
No response	≤5
Highest level of educational attainment	
Less than a high school diploma	≤5
High school diploma or some postsecondary education	48 (25)
Trades certificate or diploma	14 (7)
College or university degree	94 (48)
Postgraduate degree	34 (17)
No response	≤ 5

[a]Percentage values are not provided for values that are ≤5.

[b]Race and ethnicity data were not collected in Manitoba.

[c]Unmet social needs were based on the following categories: precarious employment (presence of all of the following: short term, casual, or temporary employment; fear of being fired if raised employment concerns; and varying pay); living in social housing; missed rent or utility bill payments; missed appointment due to transportation cost; avoided filling prescription or made it last longer due to cost; difficulty making ends meet; and lack of social

XSL•FO
RenderX

support.

Themes

The results are presented as main themes and subthemes. The 4 main themes include AI as the inevitable future; potential health care harms; loss of the human touch; and consent is critical. Additional quotes that provide evidence of themes are provided in Multimedia Appendix 3.

AI as the Inevitable Future: Facilitating More Efficient, Accessible SDoH Information and Use

Participants described common benefits to leveraging the use of AI in medicine to derive their SDoH information, as it represented the future of technological advancements, and "it's much better that way" (05_20, Newfoundland and Labrador). Some participants spoke of AI as inevitable regardless of their perspectives for or against its use and seemed to acquiesce to its use in medicine.

> *Yeah, I'm for it… I mean it's the future of medicine and it's the future of the world so… Whether I have concerns or not, it's going to take over, but personally I don't have concerns, I'm okay with it, yep.* [01_135 (Ontario)]

Efficient, Streamlined Social Determinants of Health Data

Most participants expressed that the benefits of using AI in primary care were to streamline SDoH data collection in a more efficient and timely manner for patients, staff, and the health care system.

> *Well, I see benefits cause it streamlines the process and makes your information more accessible to people so you're not the one who has to like repeat and remind [healthcare staff] all the time.* [04_16 (Manitoba)]

One participant mentioned that, if used correctly and accurately, it could reduce the burden on the health care system to free-up resources for providers to speak with patients.

> *People think that artificial intelligence and technology will be replacing jobs but like that's not true because it only is there to help us move faster in life…Once you give the job [of deriving patients' SDoH data through AI] …the people who were doing that [previously] could have one-on-one…interpersonal communication with patients and you know speak to them while they're in the waiting room.* [01_06 (Ontario)]

AI Could Overcome Barriers to Disclosure

A small number of participants described how using AI to derive SDoH information based on existing EMR data could overcome barriers to verbally disclosure, in a "more accessible" (01_124, Ontario) manner.

> *So, I do [see benefits of AI], I think like folks that have challenges expressing themselves like…people that are vulnerable…the homeless, folks that are coming out of traumatic situations…people that have language barriers, speech barriers; I think this technology would be very beneficial to them…And people that have disabilities.* [01_99 (Ontario)]

Data Use to Improve Health

Participants expressed ways that using AI to derive SDoH information could be useful to improving health or health care. For example, 1 participant described how it would help to

> *…move towards a system of greater continuity of care and a system where you are not forced to tell each new professional you see the same old story*. [01_134 (Ontario)]

Participants expressed that AI could help generate automated alerts of potential conditions for individuals based on group memberships from the SDoH data, in addition to other health information in the EMR.

> *I think it's good for certain things like I think if you're… from a particular country or race or something and you're prone to very like high health risk diseases for example…I think that's important and sure a computer can fill that information out so that way they red flag like check her heart every time she comes in.* [01_68 (Ontario)]

Participants indicated that AI could be useful to help automate local social and community resources to assist with their social situation.

> *That's a way to screen and actually find the best help for you within your area…say for example the person has addiction or…mental health challenges or…are in an abusive situation…Sometimes these programs are so filled there is not enough space available…this program can provide all this information accessible to you and show the patient immediately if tomorrow they can get help… they're put on some waiting list or [indicate where] there are shelters available or…programs from the government.* [01_38 (Ontario)]

Some participants expressed that disaggregated SDoH data could be used at a larger health care or population level, as opposed to the individual level. This included assistance with policy-making decisions to reduce health inequities, disease surveillance, understanding the causes of diseases, and prevention. COVID-19 disparities were mentioned as an example of the use of SDoH data collection.

> *I think the benefits is if we're going to be using to track down diseases like not [individual] people [and determine]…What are the cause?... That would be beneficial and [it could be used for prevention] … [You could] do more campaigns…[but] knowing the stats and releasing that to the public in general I think that would help everybody…[It would help] to make them aware of what's going on… Especially now with COVID.* [03_16 (Saskatchewan)]

Similarly, 1 participant described how the information generated could be used to inform "culturally competent" care through staff training practices.

XSL•FO
RenderX

> *You [can use AI for demographic purposes, such as you] find that you have 50% African, 30% Asian, 10% White, 20% mixed [racial identities] … Then it would therefore affect training because clearly the majority of your clientele will be African so you have to be culturally competent to be able to treat [them] with whatever they are coming with, the possibility so I think in that sense it could be beneficial.* [03_19 (Saskatchewan)]

Potential Health Care Harms: Distrust in AI Used in Public Systems

Many participants held strong views about the potential harms of implementing AI in primary care and expressed an overall distrust in AI.

AI Inaccuracies and Impact on Care

The primary concerns with using AI to derive SDoH information in health care were inaccurate SDoH predictions and the subsequent consequences on their care provision. Many participants were particularly concerned about their race and ethnicity, gender identity, and sexual orientation being incorrectly identified, including "misclassifications" (01_28, Ontario).

> *I think that could actually kind of be dangerous… I'm not sure I would trust the computer…I mean I work in [information technology] but [Laugh] I actually wonder how would a computer know for sure what my racial identity is?... I'm just afraid of the computer making the wrong choice and then it somehow impacting my healthcare negatively… If the computer [was] in the background [and] was only doing it for statistics that is not for an individual's healthcare then I mean I don't really care. I feel like there could be room for error.* [01_32 (Ontario)]

However, a few participants felt that more inaccuracies would result from a human compared with AI and they would trust the computer over humans with their SDoH information:

> *I think maybe a computer would probably be better than an individual doing it because you know [for an] individual there is always I would think a larger chance for a margin of human error rather than a computer.* [01_56 (Ontario)]

With regards to updating SDoH information, some participants expressed that social situations are malleable and may change, but those changes may not be accurately captured in their EMR by their physician unless they are explicitly asked about it through a survey. Participants mentioned this often in the context of financial situations, sexual orientation, or gender identity.

> *I think asking people one-by-one even though it is a bit more time consuming is worth it because you know like even for sexuality and all like it always changes so it's not like they can, it can be generated once because it should come from the main source…because like let's say you're in a court and someone says like you know where did you get this information from? It's like I didn't even provide that information.* [01_39 (Ontario)]

It was described how AI would not pick up on the subtle human cues that are intrinsic to conversation.

> *The only problem, when it comes to like using that type of technology is there's no emotion behind it so when you're speaking with someone one-on-one you might say something in a certain way but you don't mean it that way. So, your facial expression, the way you speak it communicates your idea across differently.* [01_27 (Ontario)]

In addition, some participants had different perspectives on using AI based on the SDoH variable that was predicted. They mentioned that perhaps certain SDoH should be derived using EMR data while others should be asked directly by patients.

> *So, I think for things that are like you know pretty like core defining questions like your race, your gender, your age, your citizenship like I don't want the thing guessing that I'm Chinese based on my last name for example…So, I think those things are, you know like more like the legal aspects of things… Other things like you know your income like okay, it can predict it and probably it's like pretty accurate…I would just want to be assured that whatever predictions it makes like it would be to try to benefit me and not to try to like characterize myself and like sell me some new drugs.* [01_72 (Ontario)]

Privacy and Security

Privacy and security were important concerns of participants:

> *I think that privacy…has to be the number one issue that guides every aspect of this.* [04_21 (Manitoba)]

Participants expressed concerns with the management of SDoH information in the EMR, including that "*…the infrastructure and the environment in which you are operating is not equipped to protect the data you are collecting*." (01_13, Ontario). Another participant mentioned "*It's frightening what computers are being programmed to do without any accountability.*" [04_21 (Manitoba)].

Some participants also mentioned that the data being retrieved using EMR was in itself a "huge invasion of privacy" [01_80 (Ontario)].

Several participants described how they did not want AI to derive their SDoH information as it removed their ability to control and share what they are comfortable with sharing to their doctor, given the sensitive nature of the SDoH questions. For example, some participants were themselves not willing to have their SDoH data on the EMR, such as their race and ethnicity or sexual orientation.

> *Importantly I think it would make me very leery or weary of what I share every time I talk to a healthcare provider cause I might want to share a sensitive piece of information but I don't want that information to be widely available on my record …Your program to retroactively go back and gather this information*

XSL•FO
RenderX

> *from all of my visits just makes me feel very like not in control of my own information.* [01_50 (Ontario)]

The interviews occurred at a time when there was media attention surrounding hacking and privacy breaches of government across provinces, and COVID-19 vaccine policies. Individuals also reported having distrust in government and overall technology.

> *My thoughts on this whole COVID-19, yeah, it's a long story and stuff but I don't trust the government so I wouldn't want it to be done... I don't have a lot of trust in computers... You can say everything is secure but like I said I've seen it over time and especially in the last 20 months there's been a lot of breaches…You want to know? Phone me. You want to question me or ask me? Bring me in person-to-person. Don't do it over the computer... because you can be hacked at any time.* [03_22 (Saskatchewan)]

Data Misuse and Discrimination

Few participants mentioned their concerns about data misuse and discrimination. For example, a participant mentioned how the use of AI would be stigmatizing and could negatively stereotype people, particularly those living with disabilities. Discrimination due to race and ethnicity and income were also reported by participants.

Some participants extrapolated the harms beyond the health care system. Participants compared the potential for AI to discriminate or be used for discrimination, to their concerns regarding the criminal system and policing. For example, 1 participant indicated that it would facilitate "[racial] profiling in general and not only healthcare but the government is going to use it, so I am concerned" and provided an example of it being used in fascist governments for "genocide against communities" [01_07 (Ontario)].

There were also concerns regarding the data being misused if it were to be leaked.

> *The companies are trying to surveil us or they're trying to… use our data for their own means…I'm thinking about like police officers... I think it was in British Columbia where like…the municipal police department…was using AI to surveil folks and… that is like very problematic and could really harm a lot of people and I think that this could just be linked to that and like I feel like police would maybe somehow have access to this if the government does.* [01_41 (Ontario)]

One participant described how the algorithm would essentially be stereotyping individuals in order to predict and identify their SDoH information.

> *Well, and it's based on a whole lot of basically stereotyping groups so saying oh, well I am racialized and low-income therefore these things must be true and I'm not at all comfortable with, cause it's, the whole system relies on those stereotypes and assumptions.* [04_17 (Manitoba)]

However, 1 participant believed that AI "is not racist. [Laugh] They don't see race" [03_24 (Saskatchewan)] and is better suited to SDoH data retrieval as opposed to human data collection.

Loss of the Human Touch: Preference for Provider Relationships and Individualized Care

A few participants expressed an overall preference for having strong patient-provider relationships and interactions and emphasized the need for individualized care. Participants believed AI could pose a barrier to patient-centered care and the patient-provider relationship, by removing the "human factor" with providers which could otherwise be nurtured by having a conversation about their SDoH situation. For example, they expressed that it would be "de-personalizing encounters with the doctor" [04_20 (Manitoba)] or would be "dehumanizing," particularly for people experiencing a mental health condition and social isolation, by "just being dealt with by a machine" [01_125 (Ontario)].

> *I think having the human touch to our view like these surveys is always good because well people will give more personal opinions if you allow them to whereas if it was just AI then you know we're just numbers.* [01_36 (Ontario)]

> *Healthcare should be about individuals and it should be about connection. It should be about you and me face-to-face and to put artificial intelligence in there to pigeon hole me into a certain group I think would be very dangerous. I think we want to get away from that and start looking at individual situations and everybody's you know life and how we can interact with them best.* [03_03 (Saskatchewan)]

Similarly, one participant provided a unique perspective as someone who identifies as living with disabilities.

> *I'd just like to see that be individualized and person centered... That requirement for person-centered care for diversity of options, a diversity of supports, and for that holistic care to be present and connected and supportive. Mostly I think it's a human rights actual systemic problem to use AI.* [03_21 (Saskatchewan)]

Consent Is Critical: Strong Safeguards Are Needed to Protect Patients' Data and Trust

Overview

Many participants advocated for strong safeguards and fully-informed consent before having AI used to derive their SDoH information. Most of these participants described opt-in consent, however some described opt-out consent. There is a need for "full transparency" [01_78 (Ontario)] regarding why the data is being collected by AI, what it will be used for, how it will be stored, who has access, how it will work, and oversight before implementation. Without these safeguards, participants mentioned that it would be a "breach of trust" [01_78 (Ontario)] and they would feel upset and "caught off guard" [01_38 (Ontario)].

> *And I think verbal consent or written consent is very important…as opposed to the passive consent you know it's like when you download something how it's*

XSL•FO
RenderX

> *all agree, agree, agree…We really do need to read the fine print…And because I think that we, our private information is more valuable than we actually realize and we've gotten so accustomed to passing it out at times that we're failing to realize how valuable and how sensitive it can be so….* [05_12 (Newfoundland and Labrador)]
>
> *So it is a black box in the way it uses that [information] but then is there a team which is always correcting it if it's out of whack?* [01_13 (Ontario)]

Data Access and Verification

Multiple participants wanted the opportunity to check what the AI algorithms derived for their SDoH information and have the chance to rectify inaccuracies.

> *Maybe you know it would be okay if the AI [derives the SDoH information] first and then the patient…is asked to confirm like say here is what we think it is, do you think this is correct - and then the patient can say yes or no but I think not to just do it and like keep it there behind closed doors.* [03_05 (Saskatchewan)]

Discussion

Principal Findings

Overall, there were varied perspectives on whether AI should be used in primary care to derive patients' SDoH information. Some participants expressed that the benefits of more efficient data collection or improved care did not outweigh the potential harms from inaccurate predictions, privacy and security concerns, reduced patient-provider interactions, and data misuse and discrimination. Many participants emphasized the need for strong safeguards if AI is used, including fully informed consent, transparency, oversight, and the ability to check and verify predictions.

Benefits

Participants in other studies have similarly identified that the use of AI in health care is efficient and could promote improved health and health care [52]. Integrating SDoH information through AI can help to predict risks of negative health outcomes (eg, suicide [53] and HIV [54]) and health care use [25,55-57]. Other studies have found that the identification of unmet social and financial needs using EMR data could assist with predicting the need for community resources, and therefore facilitate community resource connections and personalized interventions to address patients' social situations [57,58]. It could also help to inform future program design and community planning to build capacity and outreach to encourage access to necessary health services and reduce related inequities [57,58].

Navigating Concerns and Establishing Strong Patient Protections

Many of the potential harms of AI were focused on issues surrounding participants' personal care and treatment at the individual-level, as opposed to the use of aggregated data for system- or organization-level change. This may partly explain the strong concerns that surfaced about AI. Some participants highlighted the impracticality of continuously verifying AI-generated outcomes over time and expressed willingness to contribute their data anonymously for research and quality improvement. Health services may consider the feasibility of adopting alternative models of patient consent [59]. For example, tiered consent enables patients to opt-in or -out of sharing their data under various conditions (eg, data used at the aggregate level for algorithm development or individual-level) [59,60]. It has been recommended for organizations to implement patient education on consent regarding AI and to work alongside patients in creating a consent that is coherent and straightforward [60]. Similar to the participants in this study, others have discussed consent, transparency, verification, control, and oversight as part of recommendations for the ethical use of AI [55,61-64].

AI is a broad term to describe many types of complex statistical methods, which are not often transparent or well understood by the public. This can lead to mistrust in AI and medicine, and given the novelty of AI in health care settings, there are still many unknowns. A major concern is the potential for AI to misclassify one's racial or ethnic background, as well as fluctuating social circumstances. Even with highly accurate AI models, misclassification is possible, as well as algorithmic or data bias and societal bias [32]. Without careful oversight, inaccurate or biased AI models could be used to significantly impact patient care or propagate discrimination or health inequities [32,41]. This may contribute to overcriminalization and systemic discrimination experienced by racialized and low-income communities [65,66]. Thus, although resource-intensive, it is imperative that patients have the ability to verify the AI-generated SDoH outcomes and modify it. It is also incumbent on AI teams to produce algorithms that go further than just being interpretable, but ones that provide justification for the outcome [67] and proactively address health equity [68].

Other studies examining perspectives on AI in health care have reported similar concerns of privacy and data breaches, infrastructure and oversight, and lack of choice or control of the data [24,38,39,52,62,63,69]. During the time of the interviews, the Canadian news media frequently discussed ransomware attacks and data breaches in medical data systems [70], including those described as the "worst in Canadian history" [71]. This context could have impacted participant concerns of using AI to derive SDoH data in this study.

Health care, particularly primary care, is a highly personal "social enterprise, powered by committed, caring, and collaborative connections between the humans involved" [72]. It is not surprising that a major concern in this study was the potential for AI to threaten patient-centered and individualized care, harming the patient-provider relationship. These findings have been reported in the literature [69,73] and may reflect concerns about AI in medicine in general, as opposed to extraction of SDoH data on the backend. Some articles have suggested restricting AI from decision-making [74,75] such as automatically deriving patient SDoH data and recording it into patient records. Instead, AI could be used as a tool [74] as part of patient-centered care that helps providers prioritize discussions about SDoH with their patient based on their risk of having unmet needs. Based on these discussions, team-based

XSL•FO
RenderX

care approaches could support personalized actions following identification of a need.

Implementation of AI for SDoH data extraction should incorporate a strong focus on equity and meaningful partnerships with underrepresented communities in every stage, to promote safety and minimize potential harms [32,38,62]. This should include community governance, particularly for Indigenous and Black communities [61], and would inform whether the SDoH data is collected, the type of data to collect, the frequency for such data collection, and its use. In combination, using an equity framework to guide implementation could help to mitigate medical mistrust and associated health disparities [76,77], particularly among individuals who experience structural racism and discrimination in health care [76,78]. Overall, while the necessary safeguards may diminish some of the returns of more timely, efficient SDoH information retrieval using AI, they must be an inextricable part of using AI to derive SDoH data.

Strengths and Limitations

This study has multiple strengths. It involved a large sample of 195 participants across 4 Canadian provinces and used a maximum variation approach for a diverse sample. For example, the majority of participants who disclosed their race and ethnicity identified as non-White, and most had at least 1 unmet social need. There are few multijurisdictional studies that examine perspectives on using AI to derive SDoH data; thus, this study contributes to reducing knowledge gaps regarding implementation of AI to derive SDoH data in primary care. Various techniques were also used to increase the trustworthiness and quality of the study. Furthermore, the SPARK project is composed of a dedicated team of researchers and patient partners who are passionate about health and social inequities and meet regularly to discuss the study and decision making.

Several limitations were also identified. This study reports the findings from questions on AI that were asked as part of a larger interview on general SDoH data collection and use in primary care. Participants' perspectives may have been impacted by the other questions that preceded the AI conversation. In addition, despite having study recruitment materials available in multiple languages for each province, the team relied entirely on potential participants to make initial contact and were not approached by non-English speakers. While the sampling strategy captured a range of formal educational experiences, approximately one-third had a college, university, or postgraduate degree. Thus, the study sample may not include individuals who experience language barriers and may be less transferable to individuals who did not complete high or grade school. Furthermore, race-based data in Manitoba were not collected due to the need for greater community engagement. This study did not receive Research Ethics Board approval across all provinces to report participants' age; however, each province considered age for a diverse recruitment sample. Finally, given project timelines and resource constraints, qualitative data analysis was not formally initiated until the interviews were completed. Despite this, the close communication, regular meetings, and feedback enabled the interview guide to undergo changes for greater clarity; the interview questions were also piloted by the study team beforehand.

Conclusions

This large qualitative study examined perspectives on using AI to derive SDoH data in primary care using existing EMR data. Participants described the benefits of efficiency, access and improvement of care, and concerns focused on inaccuracies, negative consequences to care, privacy and security breaches, reduced patient-provider interactions, data misuse, and discrimination. Strong safeguards, including fully informed consent, verification and oversight can help to alleviate harm. There is a need to engage communities in the development, implementation, and evaluation of future AI initiatives, to determine opportunities and safety measures for using AI. Interviews with providers, health care administrators, and decision-makers are also necessary to understand the feasibility of integrating AI in primary care for SDoH information.

Acknowledgments

We would like to sincerely thank the study participants from across Canada who kindly offered their time and perspectives. We are grateful for the entire SPARK team, including the many patient partners. We are also appreciative of Ellah San Antonio and Nada Dali for helping with the interviews.

This project was supported in part by Canadian Institutes for Health Research (FRN 156885). Author [ADP] is supported as a Clinician-Scientist by the Department of Family and Community Medicine, Faculty of Medicine at the University of Toronto and at St. Michael's Hospital, the Li Ka Shing Knowledge Institute, St. Michael's Hospital, and a CIHR Applied Public Health Chair in Upstream Prevention. The opinions, results and conclusions reported in this article are those of the authors and are independent from any funding sources.

Authors' Contributions

KA-B, MI, LAJ, AK, EGM, NM, CN and ADP contributed to the study conceptualization and design. IAM, DH, AZS, LK, and AD-P collected the data. The data analysis and interpretation were led by VHD and JRQ, and supported by DH, AZS, LK, AD-P, and ADP. Interpretation of the data was supported by EA, JC, MR, DS, AZ, and SG. VHD wrote the paper, and was assisted by JRQ and SG. All authors revised and approved the final draft.

XSL•FO
RenderX

Conflicts of Interest

None to declare.

Multimedia Appendix 1
SPARK (Screening for Poverty And Related Social Determinants and Intervening to Improve Knowledge of and Links to Resources) tool.
[DOCX File , 27 KB - jmir_v27i1e52244_app1.docx]

Multimedia Appendix 2
Interview questions assessing sociodemographic data Collection using artificial intelligence.
[DOCX File , 13 KB - jmir_v27i1e52244_app2.docx]

Multimedia Appendix 3
Additional representative quotes.
[DOCX File , 28 KB - jmir_v27i1e52244_app3.docx]

References

1. Social determinants of health. World Health Organization. URL: https://www.who.int/health-topics/social-determinants-of-health [accessed 2022-01-20]
2. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2(Suppl 2):19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
3. Marmot M, Friel S, Bell R, Houweling TAJ, Taylor S, Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 2008;372(9650):1661-1669. [doi: 10.1016/S0140-6736(08)61690-6] [Medline: 18994664]
4. Pinto AD, Glattstein-Young G, Mohamed A, Bloch G, Leung FH, Glazier RH. Building a foundation to reduce health inequities: routine collection of sociodemographic data in primary care. J Am Board Fam Med 2016;29(3):348-355 [FREE Full text] [doi: 10.3122/jabfm.2016.03.150280] [Medline: 27170792]
5. Peiris D, Brown A, Cass A. Addressing inequities in access to quality health care for indigenous people. CMAJ 2008;179(10):985-986 [FREE Full text] [doi: 10.1503/cmaj.081445] [Medline: 18981431]
6. Baumann AA, Cabassa LJ. Reframing implementation science to address inequities in healthcare delivery. BMC Health Serv Res 2020;20(1):190 [FREE Full text] [doi: 10.1186/s12913-020-4975-3] [Medline: 32164706]
7. Dickman SL, Himmelstein DU, Woolhandler S. Inequality and the health-care system in the USA. Lancet 2017;389(10077):1431-1441. [doi: 10.1016/S0140-6736(17)30398-7] [Medline: 28402825]
8. Nafiu OO, Mpody C, Kim SS, Uffman JC, Tobias JD. Race, postoperative complications, and death in apparently healthy children. Pediatrics 2020;146(2):e20194113. [doi: 10.1542/peds.2019-4113] [Medline: 32690804]
9. Kraemer K, Cohen ME, Liu Y, Barnhart DC, Rangel SJ, Saito JM, et al. Development and evaluation of the American college of surgeons NSQIP pediatric surgical risk calculator. J Am Coll Surg 2016;223(5):685-693. [doi: 10.1016/j.jamcollsurg.2016.08.542] [Medline: 27666656]
10. Johnson TJ. Intersection of bias, structural racism, and social determinants with health care inequities. Pediatrics 2020;146(2):e2020003657. [doi: 10.1542/peds.2020-003657] [Medline: 32690807]
11. Haggerty J, Levesque JF, Harris M, Scott C, Dahrouge S, Lewis V, et al. Does healthcare inequity reflect variations in peoples' abilities to access healthcare? Results from a multi-jurisdictional interventional study in two high-income countries. Int J Equity Health 2020;19(1):167 [FREE Full text] [doi: 10.1186/s12939-020-01281-6] [Medline: 32977813]
12. Hirello L, Pulok MH, Hajizadeh M. Equity in healthcare utilization in Canada's publicly funded health system: 2000-2014. Eur J Health Econ 2022;23(9):1519-1533. [doi: 10.1007/s10198-022-01441-1] [Medline: 35182272]
13. DeVoe JE, Bazemore AW, Cottrell EK, Likumahuwa-Ackman S, Grandmont J, Spach N, et al. Perspectives in primary care: a conceptual framework and path for integrating social determinants of health into primary care practice. Ann Fam Med 2016;14(2):104-108 [FREE Full text] [doi: 10.1370/afm.1903] [Medline: 26951584]
14. Rasanathan K, Montesinos EV, Matheson D, Etienne C, Evans T. Primary health care and the social determinants of health: essential and complementary approaches for reducing inequities in health. J Epidemiol Community Health 2011;65(8):656-660. [doi: 10.1136/jech.2009.093914] [Medline: 19933684]
15. Exworthy M, Morcillo V. Primary care doctors' understandings of and strategies to tackle health inequalities: a qualitative study. Prim Health Care Res Dev 2019;20:e20 [FREE Full text] [doi: 10.1017/S146342361800052X] [Medline: 32800013]
16. Penman-Aguilar A, Talih M, Huang D, Moonesinghe R, Bouye K, Beckles G. Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. J Public Health Manag Pract 2016;22 Suppl 1(Suppl 1):S33-S42 [FREE Full text] [doi: 10.1097/PHH.0000000000000373] [Medline: 26599027]

XSL•FO
RenderX

17. Andermann A, CLEAR Collaboration. Taking action on the social determinants of health in clinical practice: a framework for health professionals. CMAJ 2016;188(17-18):E474-E483 [FREE Full text] [doi: 10.1503/cmaj.160177] [Medline: 27503870]
18. Fichtenberg CM, Alley DE, Mistry KB. Improving social needs intervention research: key questions for advancing the field. Am J Prev Med 2019;57(6 Suppl 1):S47-S54 [FREE Full text] [doi: 10.1016/j.amepre.2019.07.018] [Medline: 31753279]
19. Gottlieb L, Tobey R, Cantor J, Hessler D, Adler NE. Integrating social and medical data to improve population health: opportunities and barriers. Health Aff (Millwood) 2016;35(11):2116-2123. [doi: 10.1377/hlthaff.2016.0723] [Medline: 27834254]
20. Pinto AD, Bloch G. Framework for building primary care capacity to address the social determinants of health. Can Fam Physician 2017;63(11):e476-e482 [FREE Full text] [Medline: 29138172]
21. Davis VH, Dainty KN, Dhalla IA, Sheehan KA, Wong BM, Pinto AD. "Addressing the bigger picture": a qualitative study of internal medicine patients' perspectives on social needs data collection and use. PLoS One 2023;18(6):e0285795 [FREE Full text] [doi: 10.1371/journal.pone.0285795] [Medline: 37285324]
22. Davis VH, Rodger L, Pinto AD. Collection and use of social determinants of health data in inpatient general internal medicine wards: a scoping review. J Gen Intern Med 2023;38(2):480-489 [FREE Full text] [doi: 10.1007/s11606-022-07937-z] [Medline: 36471193]
23. Hatef E, Rouhizadeh M, Tia I, Lasser E, Hill-Briggs F, Marsteller J, et al. Assessing the availability of data on social and behavioral determinants in structured and unstructured electronic health records: a retrospective analysis of a multilevel health care system. JMIR Med Inform 2019;7(3):e13802 [FREE Full text] [doi: 10.2196/13802] [Medline: 31376277]
24. Alpert J, Kim HJ, McDonnell C, Guo Y, George TJ, Bian J, et al. Barriers and facilitators of obtaining social determinants of health of patients with cancer through the electronic health record using natural language processing technology: qualitative feasibility study With stakeholder interviews. JMIR Form Res 2022;6(12):e43059 [FREE Full text] [doi: 10.2196/43059] [Medline: 36574288]
25. Chen M, Tan X, Padman R. Social determinants of health in electronic health records and their impact on analysis and risk prediction: a systematic review. J Am Med Inform Assoc 2020;27(11):1764-1773 [FREE Full text] [doi: 10.1093/jamia/ocaa143] [Medline: 33202021]
26. Williams-Roberts H, Neudorf C, Abonyi S, Cushon J, Muhajarine N. Facilitators and barriers of sociodemographic data collection in Canadian health care settings: a multisite case study evaluation. Int J Equity Health 2018;17(1):186 [FREE Full text] [doi: 10.1186/s12939-018-0903-0] [Medline: 30591045]
27. Bakken S. Can informatics innovation help mitigate clinician burnout? J Am Med Inform Assoc 2019;26(2):93-94 [FREE Full text] [doi: 10.1093/jamia/ocy186] [Medline: 30668747]
28. Lofters AK, Shankardass K, Kirst M, Quiñonez C. Sociodemographic data collection in healthcare settings: an examination of public opinions. Med Care 2011;49(2):193-199. [doi: 10.1097/MLR.0b013e3181f81edb] [Medline: 21150797]
29. Kirst M, Shankardass K, Bomze S, Lofters A, Quiñonez C. Sociodemographic data collection for health equity measurement: a mixed methods study examining public opinions. Int J Equity Health 2013;12:75 [FREE Full text] [doi: 10.1186/1475-9276-12-75] [Medline: 24119260]
30. Patra BG, Sharma MM, Vekaria V, Adekkanattu P, Patterson OV, Glicksberg B, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. J Am Med Inform Assoc 2021;28(12):2716-2727 [FREE Full text] [doi: 10.1093/jamia/ocab170] [Medline: 34613399]
31. Lybarger K, Dobbins NJ, Long R, Singh A, Wedgeworth P, Uzuner, et al. Leveraging natural language processing to augment structured social determinants of health data in the electronic health record. J Am Med Inform Assoc 2023;30(8):1389-1397 [FREE Full text] [doi: 10.1093/jamia/ocad073] [Medline: 37130345]
32. d'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Fam Med Community Health 2022;10(Suppl 1):e001670 [FREE Full text] [doi: 10.1136/fmch-2022-001670] [Medline: 36450391]
33. Balicer RD, Cohen-Stavi C. Advancing healthcare through data-driven medicine and artificial intelligence. In: Nordlinger B, Villani C, Rus D, editors. Healthcare and Artificial Intelligence. Cham: Springer International Publishing; 2020:9-15.
34. Callahan A, Shah NH. Chapter 19 - machine learning in healthcare. In: Sheikh A, Cresswell KM, Wright A, Bates DW, editors. Key Advances in Clinical Informatics. United States: Academic Press; 2017:279-291.
35. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
36. Bejan CA, Angiolillo J, Conway D, Nash R, Shirey-Rice JK, Lipworth L, et al. Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. J Am Med Inform Assoc 2018;25(1):61-71 [FREE Full text] [doi: 10.1093/jamia/ocx059] [Medline: 29016793]
37. Zhu VJ, Lenert LA, Bunnell BE, Obeid JS, Jefferson M, Halbert CH. Automatically identifying social isolation from clinical narratives for patients with prostate Cancer. BMC Med Inform Decis Mak 2019;19(1):43 [FREE Full text] [doi: 10.1186/s12911-019-0795-y] [Medline: 30871518]

XSL•FO
RenderX

38. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
39. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey. J Med Internet Res 2021;23(8):e26162 [FREE Full text] [doi: 10.2196/26162] [Medline: 34236994]
40. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020;20(1):170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
41. Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? BMJ 2021;372:n304 [FREE Full text] [doi: 10.1136/bmj.n304] [Medline: 33722847]
42. Moy S, Irannejad M, Manning SJ, Farahani M, Ahmed Y, Gao E, et al. Patient perspectives on the use of artificial intelligence in health care: a scoping review. J Patient Cent Res Rev 2024;11(1):51-62 [FREE Full text] [doi: 10.17294/2330-0698.2029] [Medline: 38596349]
43. Adekoya I, Delahunty-Pike A, Howse D, Kosowan L, Seshie Z, Abaga E, et al. Screening for poverty and related social determinants to improve knowledge of and links to resources (SPARK): development and cognitive testing of a tool for primary care. BMC Prim Care 2023;24(1):247 [FREE Full text] [doi: 10.1186/s12875-023-02173-8] [Medline: 38007462]
44. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
45. Patton M. Qualitative Research & Evaluation Methods. Thousand Oaks: Sage; 2002.
46. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
47. Lindgren BM, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. Int J Nurs Stud 2020;108:103632. [doi: 10.1016/j.ijnurstu.2020.103632] [Medline: 32505813]
48. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
49. Guba EG, Lincoln YS. Epistemological and methodological bases of naturalistic inquiry. ECTJ 1982;30(4):233-252. [doi: 10.1007/bf02765185]
50. Guillemin M, Gillam L. Ethics, reflexivity, and “Ethically Important Moments” in research. Qualitative Inquiry 2004;10(2):261-280. [doi: 10.1177/1077800403262360]
51. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2025-03-05]
52. Wu C, Xu H, Bai D, Chen X, Gao J, Jiang X. Public perceptions on the application of artificial intelligence in healthcare: a qualitative meta-synthesis. BMJ Open 2023;13(1):e066322 [FREE Full text] [doi: 10.1136/bmjopen-2022-066322] [Medline: 36599634]
53. Zheng L, Wang O, Hao S, Ye C, Liu M, Xia M, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. Transl Psychiatry 2020;10(1):72. [doi: 10.1038/s41398-020-0684-2] [Medline: 32080165]
54. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using clinical notes and natural language processing for automated HIV risk assessment. J Acquir Immune Defic Syndr 2018;77(2):160-166 [FREE Full text] [doi: 10.1097/QAI.0000000000001580] [Medline: 29084046]
55. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. 2021. URL: https://www.who.int/publications-detail-redirect/9789240029200 [accessed 2023-07-30]
56. Nijhawan AE, Metsch LR, Zhang S, Feaster DJ, Gooden L, Jain MK, et al. Clinical and sociobehavioral prediction model of 30-day hospital readmissions among people with HIV and substance use disorder: beyond electronic health record data. J Acquir Immune Defic Syndr 2019;80(3):330-341 [FREE Full text] [doi: 10.1097/QAI.0000000000001925] [Medline: 30763292]
57. Soy Chen MS, Danielle B, Kelly M, Allison K, John F, John S. Using applied machine learning to predict healthcare utilization based on socioeconomic determinants of care. The American Journal of Managed Care 2020;26(1):26-31 [FREE Full text] [doi: 10.37765/ajmc.2020.42142]
58. Kasthurirathne SN, Vest JR, Menachemi N, Halverson PK, Grannis SJ. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. J Am Med Inform Assoc 2018;25(1):47-53 [FREE Full text] [doi: 10.1093/jamia/ocx130] [Medline: 29177457]
59. Wiertz S, Boldt J. Evaluating models of consent in changing health research environments. Med Health Care Philos 2022;25(2):269-280 [FREE Full text] [doi: 10.1007/s11019-022-10074-3] [Medline: 35286521]
60. Kotsenas AL, Balthazar P, Andrews D, Geis JR, Cook TS. Rethinking patient consent in the era of artificial intelligence and big data. J Am Coll Radiol 2021;18(1 Pt B):180-184. [doi: 10.1016/j.jacr.2020.09.022] [Medline: 33413897]
61. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. BMC Public Health 2022;22(1):2146 [FREE Full text] [doi: 10.1186/s12889-022-14422-z] [Medline: 36419010]

XSL•FO
RenderX

62. Hartzler AL, Xie SJ, Wedgeworth P, Spice C, Lybarger K, Wood BR, SDoH Community Champion Advisory Board. Integrating patient voices into the extraction of social determinants of health from clinical notes: ethical considerations and recommendations. J Am Med Inform Assoc 2023;30(8):1456-1462 [FREE Full text] [doi: 10.1093/jamia/ocad043] [Medline: 36944091]
63. Beets B, Newman TP, Howell EL, Bao L, Yang S. Surveying public perceptions of artificial intelligence in health care in the United States: systematic review. J Med Internet Res 2023;25:e40337 [FREE Full text] [doi: 10.2196/40337] [Medline: 37014676]
64. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019;1(9):389-399. [doi: 10.1038/s42256-019-0088-2]
65. Bingley WJ, Haslam SA, Steffens NK, Gillespie N, Worthy P, Curtis C, et al. Enlarging the model of the human at the heart of human-centered AI: a social self-determination model of AI system impact. New Ideas in Psychology 2023;70:101025. [doi: 10.1016/j.newideapsych.2023.101025]
66. Richardson R, Schultz JM, Crawford K. Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. NYU L Rev Online 2019;94:15.
67. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020;2020:191-200 [FREE Full text] [Medline: 32477638]
68. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169(12):866-872 [FREE Full text] [doi: 10.7326/M18-1990] [Medline: 30508424]
69. Haan M, Ongena YP, Hommes S, Kwee TC, Yakar D. A qualitative study to understand patient perspective on the use of artificial intelligence in radiology. J Am Coll Radiol 2019;16(10):1416-1419. [doi: 10.1016/j.jacr.2018.12.043] [Medline: 30878311]
70. Wilner AS, Luce H, Ouellet E, Williams O, Costa N. From public health to cyber hygiene: cybersecurity and Canada's healthcare sector. International Journal 2022;76(4):002070202110679 [FREE Full text] [doi: 10.1177/00207020211067946]
71. N.L. health-care cyberattack is worst in Canadian history, says cybersecurity expert. CBC News. 2021. URL: https://www.cbc.ca/news/canada/newfoundland-labrador/nl-cyber-attack-worst-canada-1.6236210 [accessed 2023-08-22]
72. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. J Gen Intern Med 2019;34(8):1626-1630 [FREE Full text] [doi: 10.1007/s11606-019-05035-1] [Medline: 31090027]
73. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021;3(9):e599-e611 [FREE Full text] [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
74. Witkowski K, Okhai R, Neely SR. Public perceptions of artificial intelligence in healthcare: ethical concerns and opportunities for patient-centered care. BMC Med Ethics 2024;25(1):74 [FREE Full text] [doi: 10.1186/s12910-024-01066-4] [Medline: 38909180]
75. Bjerring JC, Busch J. Artificial intelligence and patient-centered decision-making. Philos. Technol 2020;34(2):349-371. [doi: 10.1007/s13347-019-00391-6]
76. Jaiswal J, Halkitis PN. Towards a more inclusive and dynamic understanding of medical mistrust informed by science. Behav Med 2019;45(2):79-85 [FREE Full text] [doi: 10.1080/08964289.2019.1619511] [Medline: 31343962]
77. Pellowski JA, Price DM, Allen AM, Eaton LA, Kalichman SC. The differences between medical trust and mistrust and their respective influences on medication beliefs and ART adherence among African-Americans living with HIV. Psychol Health 2017;32(9):1127-1139 [FREE Full text] [doi: 10.1080/08870446.2017.1324969] [Medline: 28475365]
78. Benkert R, Peters RM, Clark R, Keves-Foster K. Effects of perceived racism, cultural mistrust and trust in providers on satisfaction with care. J Natl Med Assoc 2006;98(9):1532-1540. [Medline: 17019925]

Abbreviations

AI: artificial intelligence
EMR: electronic medical record
SDoH: social determinants of health
SPARK: Screening for Poverty And Related Social Determinants and Intervening to Improve Knowledge of and Links to Resources

XSL•FO
RenderX

Edited by T de Azevedo Cardoso, KJ Craig; submitted 28.08.23; peer-reviewed by BG Patra, Y Khan, J Wu, A Frier, K Fuji; comments to author 30.06.24; revised version received 31.10.24; accepted 29.11.24; published 06.03.25.

Please cite as:
Davis VH, Qiang JR, Adekoya MacCarthy I, Howse D, Seshie AZ, Kosowan L, Delahunty-Pike A, Abaga E, Cooney J, Robinson M, Senior D, Zsager A, Aubrey-Bassler K, Irwin M, Jackson LA, Katz A, Marshall EG, Muhajarine N, Neudorf C, Garies S, Pinto AD Perspectives on Using Artificial Intelligence to Derive Social Determinants of Health Data From Medical Records in Canada: Large Multijurisdictional Qualitative Study
J Med Internet Res 2025;27:e52244
URL: https://www.jmir.org/2025/1/e52244
doi:10.2196/52244
PMID:40053728

©Victoria H Davis, Jinfan Rose Qiang, Itunuoluwa Adekoya MacCarthy, Dana Howse, Abigail Zita Seshie, Leanne Kosowan, Alannah Delahunty-Pike, Eunice Abaga, Jane Cooney, Marjeiry Robinson, Dorothy Senior, Alexander Zsager, Kris Aubrey-Bassler, Mandi Irwin, Lois A Jackson, Alan Katz, Emily Gard Marshall, Nazeem Muhajarine, Cory Neudorf, Stephanie Garies, Andrew D Pinto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Generative AI–Enabled Therapy Support Tool for Improved Clinical Outcomes and Patient Engagement in Group Therapy: Real-World Observational Study

Johanna Habicht[1*], MSci; Larisa-Maria Dina[1,2*], BSc, MSc; Jessica McFadyen[1], PhD; Mona Stylianou[3]; Ross Harper[1], PhD; Tobias U Hauser[1,4,5,6], PhD; Max Rollwage[1], PhD

[1]Limbic Ltd, London, United Kingdom

[2]Department of Psychology, King's College London, London, United Kingdom

[3]Everyturn Mental Health, Gosforth, United Kingdom

[4]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom

[5]Department of Psychiatry and Psychotherapy, Medical School and University Hospital, Eberhard Karls University of Tubingen, Tübingen, Germany

[6]German Center for Mental Health, Tübingen, Germany

[*]these authors contributed equally

Corresponding Author:
Max Rollwage, PhD
Limbic Ltd
Kemp House, 128 City Road
London, EC1V 2NX
United Kingdom
Phone: 44 020 3818 3240
Email: max@limbic.ai

Abstract

Background: Cognitive behavioral therapy (CBT) is a highly effective treatment for depression and anxiety disorders. Nonetheless, a substantial proportion of patients do not respond to treatment. The lack of engagement with therapeutic materials and exercises between sessions, a necessary component of CBT, is a key determinant of unsuccessful treatment.

Objective: The objective of this study was to test whether the deployment of a generative artificial intelligence (AI)–enabled therapy support tool, which helps patients to engage with therapeutic materials and exercises in between sessions, leads to improved treatment success and patient treatment adherence compared with the standard delivery of CBT exercises through static workbooks.

Methods: We conducted a real-world observational study of 244 patients receiving group-based CBT in 5 of the United Kingdom's National Health Service Talking Therapies services, comparing 150 (61.5%) patients who used the AI-enabled therapy support tool to 94 (38.5%) patients who used the standard delivery of CBT exercises. The groups were equivalent with respect to the content of the CBT materials and the human-led therapy sessions; however, the intervention group received support from the AI-enabled therapy support tool in conducting CBT exercises.

Results: Patients using the AI-enabled therapy support tool exhibited greater attendance at therapy sessions and fewer dropouts from treatment. Furthermore, these patients demonstrated higher reliable improvement, recovery, and reliable recovery rates when compared to the control group, which was related to the degree of use of the AI-enabled therapy support tool. Moreover, we found that engagement with AI-supported CBT interventions, relative to psychoeducational materials, predicted better treatment adherence and treatment success, highlighting the role of personalization in the intervention's effectiveness. To investigate the mechanisms of these effects further, we conducted a separate qualitative experiment in a nonclinical sample of users (n=113). Results indicated that users perceived the AI-enabled therapy support tool as most useful for discussing their problems to gain awareness and clarity of their situation as well as learning how to apply coping skills and CBT techniques in their daily lives.

Conclusions: Our results show that an AI-enabled, personalized therapy support tool in combination with human-led group therapy is a promising avenue to improve the efficacy of and adherence to mental health care.

(J Med Internet Res 2025;27:e60435) doi:10.2196/60435

XSL•FO
RenderX

KEYWORDS

artificial intelligence; National Health Service; NHS Talking Therapies; mental health; therapy support tool; cognitive behavioral therapy; CBT; chatbot; conversational agent; clinical; patient engagement; therapist; treatment; medication; depression; anxiety disorder; exercise; observational study; control group; patient adherence

Introduction

Background

Mental health conditions are a leading cause of health-related burden across the world, with anxiety and depression being ranked among the top 25 causes of burden in 2019 [1]. Both disorders are very common, affecting >29% of the global population in their lifetime [2]. The recent COVID-19 pandemic has accelerated this even further, raising the global prevalence by 27.6% between 2020 and 2021 for major depressive disorder and 25.6% for anxiety disorders [3], highlighting the need for effective and accessible mental health support.

Cognitive behavioral therapy (CBT) has been shown to be a highly effective treatment for depression and anxiety disorders [4-6] and is the first line of treatment for these conditions in many therapy guidelines worldwide [7,8]. Nonetheless, CBT does not benefit everyone. In fact, a recent meta-analysis of 409 trials reported that 58% of patients do not respond to treatment [5]. Several reasons for unsuccessful therapy have been identified. Among them, the most prominent is a failure to engage with therapeutic exercises and materials (ie, homework assignments) between therapy sessions [9].

Between-session support, such as completion of therapeutic materials and exercises, is a necessary component of CBT [10], allowing patients to maintain the progress made in the therapy and transfer it to real life to achieve short-term and long-term improvements [11]. Broadly, between-session work can be divided into psychoeducational (eg, reading materials about symptomatology), self-assessment (eg, monitoring feelings, thoughts, or behaviors), and modality-specific tasks (eg, exposure to images for specific phobias). Meta-analyses have consistently demonstrated that both the quantity and the quality of such task completion significantly predict treatment outcomes [12,13].

Thus, enhancing high-quality engagement with CBT materials and exercises between therapy sessions is a key objective for improving CBT treatment outcomes [14]. This is particularly critical in group therapy settings, where dropout rates are higher and recovery rates are lower than in individual therapy settings [15-17]. Traditionally, therapeutic materials and exercises in CBT have been distributed as worksheets in pen and paper format, downloadable PDF format, or delivered as digital applications, requiring the patient to engage with these static materials without any intelligent supervision. However, this method has shown limitations, often resulting in delayed completion or incomplete materials, highlighting the need for innovative solutions to enhance patient compliance and treatment effectiveness. One promising avenue for enhancing between-session support has been the deployment of digital tools [18]. Most notably, recent developments in generative artificial intelligence (AI) and large language models (LLMs), such as ChatGPT, have showcased that AI can enable a truly tailored user experience with interactions nearly indistinguishable from humans [19] and lead to increased engagement and treatment adherence [20,21]. Thus, the use of this novel technology holds great potential to support clinical practice, especially for supporting patients between therapy sessions to meaningfully engage with their course of treatment.

Objectives

We developed a generative AI–enabled therapy support tool, Limbic Care, to support patients between sessions during clinician-led group-based CBT. The AI-enabled therapy support tool includes chat-based therapeutic materials to allow for more natural engagement with these materials than traditional, static CBT materials. Recognizing the important role of personalization in overcoming the barrier to the uptake of digital solutions [22], our solution uses generative AI to deliver a user-centric experience. Personalization addresses the diverse needs and preferences of individuals undergoing CBT rather than using the one-size-fits-all approach, which often falls short in addressing unique needs. It may also help cultivate a relationship between users and the AI-enabled therapy support tool. This relationship, also known as therapeutic alliance, has been shown to increase engagement and predict positive clinical change [23]. Thus, we hypothesized that the use of the AI-enabled therapy support tool increases engagement with therapy and thus leads to better outcomes for patients. This support was tailored and personalized, however, treatment was still provided by the clinician in line with medical device regulations [24]. In this observational study, we evaluated a novel solution for patient engagement and clinical outcomes. Assessing 244 patients undergoing CBT treatment in 5 of the United Kingdom's National Health Service (NHS) Talking Therapies services, we found that patients who chose to use the AI-enabled therapy support tool during their treatment achieved better treatment success. Moreover, we showed that engagement with the AI-enabled therapy support tool was linked to increased reliable improvement, recovery, and reliable recovery, as well as more attended treatment sessions and reduced dropout from treatment. Therefore, our results suggest that the AI-enabled therapy support tool may improve individual patients' outcomes and support clinical services by increasing treatment adherence and treatment effectiveness.

Methods

Setting

We evaluated the efficacy of an AI-enabled therapy support tool (Limbic Care; Limbic Ltd [25]) in a real-world clinical context in 5 NHS Talking Therapies for anxiety and depression services provided by Everyturn Mental Health in the United Kingdom. Specifically, the tool was implemented in the context of completing therapeutic materials and exercises between the sessions of a synchronous web-based group-based CBT program

XSL•FO
RenderX

according to the NHS Talking Therapies for anxiety and depression manual [26].

AI-Enabled Therapy Support Tool

The AI-enabled therapy support tool is a mobile app using AI to offer support between therapy sessions to patients in psychological talking therapy who are aged ≥18 years, under the supervision of a trained clinician.

The app features a conversational chatbot (Figure 1) to assist with the completion of therapist-assigned therapeutic materials and exercises, which are a component of CBT. The chatbot was powered by an LLM (GPT-4 [OpenAI] at the time of the study) to deliver clinically validated materials, that is, therapeutic materials and exercises developed by clinical experts and assigned by the treating clinician. Specifically, the clinical materials in the AI-enabled therapy support tool were interventions (interactive exercises helping the patient work through a presenting problem, eg, reframing negative thoughts) and psychoeducational materials (educational materials, eg, about CBT). Importantly, there were multiple safety layers and guardrails implemented to ensure that the LLM strictly adhered to the task at hand [27].

The therapeutic materials primarily consist of standard materials for CBT as they are also commonly used in NHS Talking Therapies. These are delivered in a conversational way to the patient using several custom-tailored LLM modules and supervisory AI tools.

All conversations were constantly monitored using several machine learning safety modules to ensure appropriateness, prevent harmful responses, monitor risks, and ensure regulatory compliance [27]. The conversations as well as these machine learning models were monitored and continuously improved by the company's research team.

In contrast to traditional homework worksheets, the patient can interact conversationally to navigate through the materials and receive notifications regarding assignments. The app encourages and empowers patients to engage in the therapeutic process outside group therapy sessions and continue the progress made in therapy in their real lives as well as provides empathetic support for the user.

Figure 1. Example user interface of the artificial intelligence–enabled therapy support tool, displaying the conversational interface (left screen) and the assignments provided by the treating clinician (right screen).

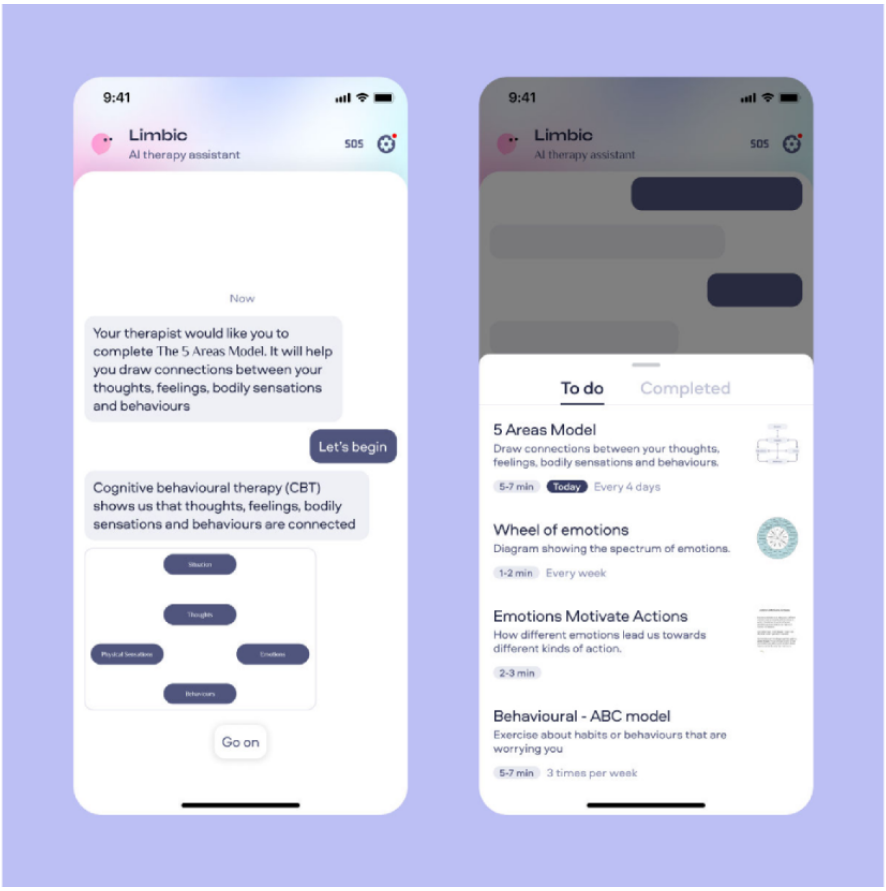

Study Design

This was a multisite, real-world, observational study to investigate the impact of the AI-enabled therapy support tool on clinical outcomes. The AI-enabled therapy support tool was offered to all patients eligible for 11 group-based CBT groups starting between October 2023 and January 2024. All (N=244, 100%) patients had access to the same group sessions irrespective of their decision to use the AI-enabled therapy support tool between the sessions. The clinicians facilitating the group sessions were trained to offer the use of the AI-enabled therapy support tool in the first session of CBT. All (N=244, 100%) patients were encouraged to complete therapeutic

XSL•FO
RenderX

materials and exercises between sessions regardless of which delivery method they chose.

We compared the clinical outcomes of individuals who signed up to use the AI-enabled therapy support tool (the intervention group) with those of individuals who did not (the control group). It is important to note that individuals who chose not to use the AI-enabled therapy support tool had access to CBT worksheets, which represent the standard way of completing homework between therapy sessions and which were also assigned by the treating clinician. Therefore, aside from the introduction of the AI-enabled therapy support tool, no other changes were made to the content or the delivery method of the sessions, ensuring that any observed effects can be attributed to patients choosing to use the tool. In addition, we investigated whether engagement with the AI-enabled therapy support tool, specifically the completion of therapeutic materials and exercises on the app, was linked to the outcome measures.

Ethical Considerations

Clinical audit studies within the NHS Talking Therapies do not require additional patient consent or ethics approval, as determined by the NHS and in line with National Institute for Health and Care Excellence guidelines [28]. This exemption from additional ethical review was further confirmed by the UK Health Research Authority. In this case, a clinical audit was used to determine whether the introduction of a new technology would improve the standard of care. Patients voluntarily opted in to use the AI-enabled therapy support tool. Those who were offered access to the AI-enabled therapy support tool agreed to the privacy and data policy agreement in the app and consented to the use of their anonymized data for audits and research studies. Because this was a real-world observational study in group therapy settings, patients were not provided any financial compensation. Regarding the identification of individuals, treatment and engagement outcome data in this study were anonymous; therefore, it was not possible to identify individual patients.

The qualitative user experience study in a nonclinical sample was separate from the clinical audit and was approved by the University College London Research Ethics Committee (6218/003). Participants in this study were paid £4.20 (US $5.17), and this was not contingent on their use of the AI-enabled therapy support tool. Participation was voluntary, and all users provided informed consent. Free-text answers were anonymized for coding and analysis.

Outcome Measures

Overview

Clinical and app use outcomes were collected as part of the study. Clinical outcomes are routinely assessed throughout treatment provision by NHS Talking Therapies. The measures of treatment engagement include attended sessions, proportion of dropouts, and did not attend (DNA) treatment sessions and the measures of treatment success include reliable improvement, recovery, and reliable recovery [29]. Use outcomes were collected through the mobile app as patients used the AI-enabled therapy support tool and were used as measures of engagement with the intervention. The information collected included the number of sessions in the app and the number of therapeutic materials and exercises completed.

Treatment Engagement

Number of Therapy Sessions

The number of therapy sessions was used to quantify the total interactions that the patient had with the therapist during their treatment journey. The interactions included assessment and therapy sessions. In NHS Talking Therapies, patients usually have 1 assessment session before the start of treatment and 6 synchronous group therapy sessions. Due to how data are reported by the services, it is not possible to differentiate between assessment and treatment sessions.

Proportion of DNA Sessions

DNA sessions refer to sessions that are canceled last minute or not attended by the patient. This was measured as the proportion of DNA sessions out of the total number of sessions that the patient could attend.

Dropouts

Dropouts are measured as the percentage of patients who dropped out during the course of treatment.

Treatment Success

Reliable Improvement

Reliable improvement refers to a clinically significant improvement in symptoms following a course of treatment and is calculated as the score difference between the first and the last validated clinical questionnaire completed. The types of questionnaires patients complete are tailored to their specific condition. For example, the Patient Health Questionnaire-9 (PHQ-9) [30] is used to measure depression symptom severity, and the Generalized Anxiety Disorder-7 (GAD-7) [31] is used to measure anxiety symptom severity. A clinically significant improvement in symptoms is considered a change score ≥6 for PHQ-9 or ≥4 for GAD-7 [26].

Recovery

Recovery is achieved in instances where a patient is defined as a clinical case at the start of treatment and not as a clinical case at the end of the treatment. Clinical caseness refers to a patient with severe enough clinical symptoms to be regarded as a clinical case at the start of treatment. Caseness is measured by the clinical questionnaires and is met when patients score ≥10 on PHQ-9 or ≥8 on GAD-7 [26].

Reliable Recovery

Reliable recovery is calculated based on reliable improvement and recovery and, as such, is the most stringent outcome. It is achieved if a patient meets the criteria for reliable improvement, referring to clinically significant improvements in their symptoms, and if the patient moves to recovery. Moving to recovery refers to a patient meeting the criteria for clinical caseness at the start of treatment to not being a clinical case at the end of it [26].

XSL•FO
RenderX

Use Outcomes

Number of Sessions in the App

A session in the app starts when the AI-enabled therapy support tool app is opened and ends when no activity has been detected for ≥30 minutes. This is measured as the number of sessions in the app and is only available for patients using the AI-enabled therapy support tool between sessions.

Number of Completed Therapeutic Materials and Exercises

Completed therapeutic materials and exercises refer to the tasks available in the AI-enabled therapy support tool that were initiated and finalized by the patients. This is measured as the number of completed homework assignments and is only available for patients using the AI-enabled therapy support tool between sessions.

User Experience

In addition to the impact of the AI-enabled therapy support tool on clinical outcomes, we were further interested in the perceived benefits of the tool and potential areas for improvement. Because it can be challenging to contact patients undergoing treatment and minimize the burden on them, we conducted a separate user study to assess user experience. Therefore, we recruited 113 individuals experiencing depression and anxiety symptoms (PHQ-9 score of ≥5 or GAD-7 score of ≥5) from the web-based platform Prolific [32] and instructed them to install the Limbic Care app. Participation was voluntary, and all (113/113, 100%) participants provided informed consent. Participants were not instructed on how to engage with the AI-enabled therapy support tool but rather were told they could use it as much or as little as they would like for the 4-week period of the study. After 4 weeks, participants were invited to answer some questions about their experience using the app, and only users who had used the app were included in the analysis. To probe the perceived benefits of the AI-enabled therapy support tool, participants were asked if they felt their mental health had benefited from using the app ("What benefits to your mental health (if any) have you gained from using the app?"). To understand potential areas for improvement, participants were asked if they had any feedback on how the app could be improved ("Do you have any feedback or suggestions on how we could improve the Limbic app?"). Participants were reimbursed £4.20 (US $5.24) for completing the study, and this was independent of their use of the AI-enabled therapy support tool.

Qualitative Analysis

Data were collected asynchronously, using free-text answers from a nonclinical, convenience sample on Prolific (n=113). Participants did not have a mental health diagnosis, though they experienced elevated symptoms of anxiety and depression (as indicated by scores of ≥5 for the PHQ-9 and GAD-7). Data were analyzed using reflexive thematic analysis, taking an inductive approach to coding [33], and a realist epistemological position, assuming that meaning and experience are reflected in language [34]. Two researchers (MR and LMD) independently coded a subset of responses and iteratively cross-checked interpretations and codes. Codes were refined using a comparative method, with new interpretations evaluated against existing ones. Inductive analysis of similar codes produced content themes highlighting potential benefits and areas for improvement in the AI-enabled therapy support tool.

Although the semiautomated nature of data collection precluded purposive sampling for saturation, no new codes emerged in later responses that altered the meaning of earlier findings, indicating data saturation was reached [35]. The themes and supporting participant quotes are presented in the Results section.

Coder Qualifications

The coding team consisted of MR, a male researcher, and LMD, a female researcher. Both researchers, while predominantly quantitative, were trained in thematic analysis and had previous experience in qualitative research [36].

Trustworthiness

Credibility was ensured through reflexivity and investigator triangulation [37,38]. Data were collected anonymously and asynchronously, minimizing the influence of researcher characteristics on participant responses. Open-ended prompts encouraged both positive and negative reflections to capture a balanced perspective. While the nature of data collection did not allow methodological triangulation, investigator triangulation was achieved through the involvement of multiple researchers (MR, LMD, JH, and JM). To ensure transferability, we described the study's context and the participants (thick description) to allow the readers to evaluate whether the findings are transferable to other care contexts [37,38]. Dependability and confirmability were ensured by thoroughly documenting the research steps from the start of the study to the interpretation and reporting of the findings [37,38].

Statistical Methods

We compared pretreatment data for patients in both groups to ensure that they did not differ based on demographic characteristics (age, gender, ethnicity, and sexual orientation), wait time to treatment, and depression and anxiety baseline symptom scores. In addition, recognizing that different referral methods can influence clinical outcomes, as illustrated by the AI-enabled self-referral tool increasing treatment success [21,39], we also compared the referral methods between groups. A 2-tailed chi-square test was applied for categorical outcomes, and a 2-tailed independent student *t* test was used for continuous outcomes.

Next, clinical outcomes were compared between the intervention and the control group. Linear regression models were run for continuous outcomes (number of attended appointments and proportion of DNA sessions), and logistic regressions were run for categorical outcomes (dropouts, reliable improvement, recovery, and reliable recovery). In both cases, the outcomes of interest were entered as independent variables and the intervention group (intervention group=1; control group=0) as a dependent variable. In addition, we controlled for baseline anxiety and depression symptom scores and the proportion of women and heterosexual participants. For linear regression, we reported regression coefficients, along with the 95% CIs, and for logistic regressions, we reported odds ratios (ORs) along with 95% CIs.

XSL•FO
RenderX

Use outcomes were investigated within the group with access to the AI-enabled therapy support tool. Specifically, we were interested in investigating if there would be any association between patients' level of engagement with the app and their clinical outcomes, in line with previous work suggesting that the completion of therapeutic materials and exercises and engagement between sessions may lead to an improvement in treatment outcomes [14]. Pearson correlation was used for continuous variables, and point-biserial correlation was used for binary variables.

Results

Participants

Patients (N=244) attending regular group therapy in Talking Therapies in the United Kingdom were offered to use our AI-enabled therapy support tool, of which 150 (61.5%) downloaded and used the AI-enabled therapy support tool (intervention group). The remaining 94 (38.5%) patients used static therapeutic materials, which were either in pen and paper format or downloadable PDF format (control group). We followed both groups from enrolling in the study until the end of their therapy (typically 8-10 weeks long). To ensure that the patients used the AI-enabled therapy support tool as intended, we calculated the percentage of individuals who engaged with the therapeutic materials at a certain number of weeks from the start of their therapy (known as app retention). We found that 79.3% (119/150) of the individuals were engaged at week 1, 64% (96/150) were engaged at week 2, 51.3% (77/150) were engaged at week 3, 42% (63/150) were engaged at week 4, 35.3% (53/150) were engaged at week 5, and 19.3% (29/150) were engaged at week 6.

To check for baseline differences, we compared pretreatment clinical and demographic characteristics. The mean ages of the participants in the intervention group (mean 40.3, SD 13.2 years) and the control group (mean 40.5, SD 15.5 years) were not significantly different (t_{183}=0.11; *P*=.91) nor was the ethnicity of the participants (intervention: 132/150, 88.7% White participants; control: 78/94, 83% White participants; χ^2_1=0.4; *P*=.53). However, the groups differed significantly in the proportion of heterosexual individuals (intervention: 129/150, 86% heterosexual individuals; control: 69/94, 73% heterosexual individuals; χ^2_1=5.4; *P*=.02) and women in the group (intervention: 111/150, 74% women; control: 58/94, 62% women; χ^2_1=4.2; *P*=.04). Therefore, we controlled for the proportion of heterosexual individuals and women in all subsequent analyses. The 2 groups did not differ in anxiety (b=–0.11; *P*=.81) symptom scores nor in depression scores at baseline (b=−0.56; *P*=.36; Table 1). Nonetheless, we decided to control for baseline depression (PHQ-9) and anxiety (GAD-7) scores in all further analyses, as they are known as prognostic factors for changes in symptom severity [40,41]. Finally, the groups did not differ in the time they waited to start treatment (t_{122}=–0.32; *P*=.75) nor the proportion of individuals who referred using the AI-enabled self-referral tool (χ^2_1=1.2; *P*=.28).

Table 1. Baseline characteristics of the therapy support and control groups.

	Intervention group (n=150)	Control group (n=94)	*P* value
Age (y), mean (SD)	40.3 (13.24)	40.5 (15.46)	.91
Gender (women), n (%)	111 (74)	58 (61.7)	.04
Sexuality (heterosexual), n (%)	129 (86)	69 (73.4)	.02
Ethnicity (White), n (%)	132 (88.7)	78 (83)	.53
PHQ-9[a] baseline, mean (SD)	12.6 (4.62)	12.0 (4.55)	.36
GAD-7[b] baseline, mean (SD)	12.0 (3.30)	12.1 (3.74)	.81
Wait time to treatment (d), mean (SD)	35.4 (19.6)	34.4 (20.4)	.75
Referral method (AI[c]-enabled self-referrals; %)	73.3	66	.28

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder-7.

[c]AI: artificial intelligence.

Outcome Data

AI-Enabled Therapy Support Improves Treatment Engagement

To assess the impact of the AI-enabled therapy support tool on treatment adherence, we compared whether the intervention group attended more therapy sessions, missed fewer sessions, and showed a lower therapy dropout rate than the control group.

Increase in the Number of Therapy Sessions

We evaluated whether the AI-enabled therapy support tool facilitated an overall increase in attended therapy sessions. We found that the intervention group attended significantly more sessions, with an average of 2 more sessions attended compared to the control group (b=1.65; 95% CI 1.051-2.200; *P*<.001), illustrating the AI-enabled therapy support tool's potential to improve engagement in therapeutic processes.

XSL•FO
RenderX

Reduction in the Sessions That Were Not Attended

Further analysis focused on the proportion of sessions not attended by the patients (referred to as DNA sessions) to gain a deeper understanding of treatment engagement. We found that the intervention group showed a significant reduction in the rate of DNA sessions (b=–14.97; 95% CI –21.464 to –8.400; *P*<.001), equating to a 15 percentage point reduction in DNA sessions (Table 2). This indicates that the intervention group was more consistent in attending sessions; therefore, it was more engaged in treatment.

Table 2. The percentage point difference between the intervention and control groups for treatment engagement and treatment success measures.

Outcome	Percentage point difference (%)	Group effect (b)	*P* value
DNA[a] session proportion	–15	–14.97	<.001
Dropout rate	–23	–1.14	<.001
Reliable improvement	21	80	.005
Recovery	25	1.03	.001
Reliable recovery	21	86	.004

[a]DNA: did not attend.

Reduction in Dropout From Therapy

More critical than DNA sessions were participants who did not finish their therapy and dropped out completely. Therefore, we investigated whether the AI-enabled therapy support tool was also linked to reduced dropout rates. We found that the intervention group experienced a significant 23 percentage point reduction in dropout rates (OR 0.32, 95% CI 0.171-0.595; *P*<.001; Table 2) compared to individuals in the control group. This suggests that the AI-enabled therapy support tool was associated with an increased likelihood of finishing a course of treatment.

Tables S1 and S2 in Multimedia Appendix 1 further evaluate the impact of the AI-enabled therapy support tool on treatment engagement separately for patients with a primary diagnosis of depression or an anxiety disorder at baseline. Briefly, we found that the tool is equally effective for both conditions.

AI-Enabled Therapy Support Improves Treatment Success

Next, to assess the impact of the AI-enabled therapy support tool, we investigated the 3 key metrics for treatment success that are routinely assessed in NHS Talking Therapies [26] (Figure 2).

Figure 2. Correlations between completed therapeutic materials, exercise assignments, and treatment engagement (A-C) and treatment success measures (D-F).

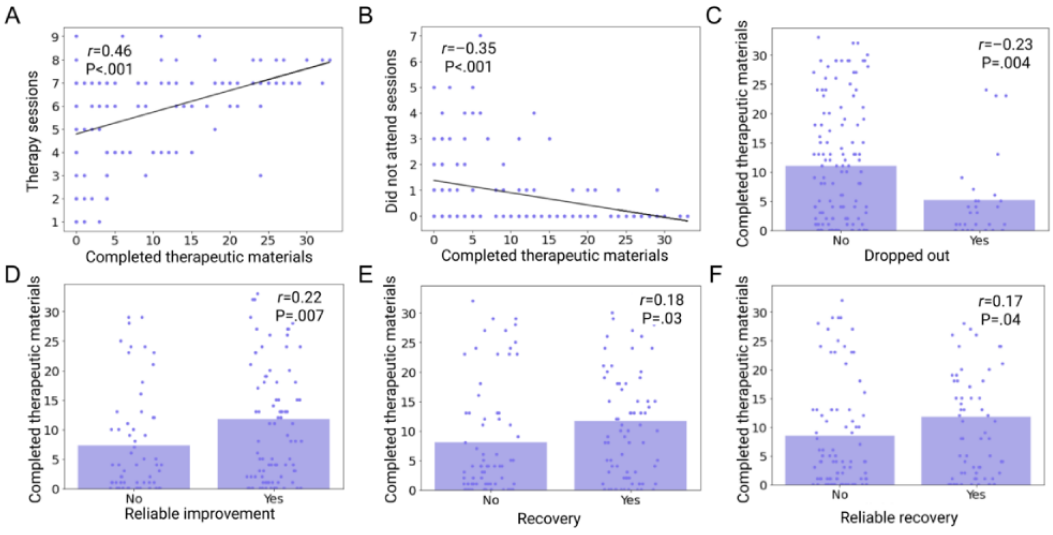

Increased Reliable Improvement

First, we compared the reliable improvement rate between the groups. Reliable improvement is defined as a significant change in patients' symptom scores from the beginning to the end of treatment (reduction of ≥6 points in the PHQ-9 or ≥4 points in the GAD-7 total scores). This measure is considered to reflect meaningful clinical progress. We found that the intervention group experienced a 21 percentage point higher rate of reliable improvement compared to the control group (Table 2). This difference was statistically significant (OR 2.21, 95% CI 1.279-3.834; *P*=.005), indicating that using the AI-enabled therapy support tool was linked to a higher likelihood that the individual saw a clinical improvement in their symptoms.

XSL•FO
RenderX

Improved Recovery

Next, we were interested in recovery, going beyond mere symptom improvement. Therefore, we looked at how many patients moved from caseness (meaning that the patient has severe enough symptoms to be considered a clinical case for depression or anxiety disorders) to below caseness during their treatment, indicating recovery [26]. We found that the intervention group experienced a 25 percentage point higher recovery rate compared to the control group (Table 2). This difference was highly significant between groups (OR 2.81, 95% CI 1.561-5.069; *P*=.001), showing that patients in the intervention group were more likely to recover during the administered therapy.

Better Reliable Recovery

Finally, we investigated the most rigorous criteria for treatment success, namely reliable recovery [26]. Reliable recovery is achieved when a patient shows both of the aforementioned criteria, that is, reliable improvement as well as recovery. In other words, the patient has moved from being considered a clinical case to being a nonclinical case and there has also been a significant improvement in their condition.

We found that the intervention group had 21 percentage points higher reliable recovery rates compared to the control group (Table 2). This difference was significant between the 2 groups (OR 2.37, 95% CI 1.311-4.290; *P*=.004), showing that patients in the intervention group were more likely to achieve reliable recovery. In summary, these results show that the people who used the AI-enabled therapy support tool achieved substantially and consistently better treatment outcomes than those who did their homework using traditional means.

Tables S1 and S2 in Multimedia Appendix 1 further evaluate the impact of the AI-enabled therapy support tool on treatment success separately for patients with a primary diagnosis of depression or an anxiety disorder at baseline. Briefly, we found that the tool is equally effective for both conditions.

Use of AI-Enabled Therapy Support Tool Linked to Better Treatment Adherence and Success

A crucial aspect of the usefulness of AI-enabled therapy support tools is whether these are being used by the patients. During the course of treatment, patients in the intervention group used the AI-enabled therapy support tool, on average, for 23 sessions. However, there was some variability in use within the intervention group. Therefore, we investigated if higher engagement with the AI-enabled therapy support tool was associated with therapy adherence in a dose-dependent manner.

First, we found a statistically significant link between the number of completed therapeutic exercises in the AI-enabled therapy support tool and the number of therapy sessions (*r*=0.46; *P*<.001; Figure 2A), meaning those who engaged more with the AI-enabled therapy support tool were more likely to attend more therapy sessions. In the same spirit, individuals using the AI-enabled therapy support tool between sessions had a lower proportion of recorded DNA sessions (*r*=–0.35; *P*<.001; Figure 2B), indicating that they were more consistent in attending scheduled sessions. Individuals who completed more tasks on the AI-enabled therapy support tool were also found to be less likely to drop out from the treatment program (r_{pb}=–0.23; *P*=.004; Figure 2C), suggesting that the tool not only enhanced session attendance but also contributed to sustained treatment participation.

In addition, we found that the higher use of the AI-enabled therapy support tool was associated with better treatment success, such as reliable improvement (r_{pb}=0.22; *P*=.007; Figure 2D), recovery (r_{pb}=0.18; *P*=.02; Figure 2E), and reliable recovery (r_{pb}=0.17; *P*=.04; Figure 2F). Therefore, these results indicated that the tool was effective in promoting both treatment adherence and clinical treatment efficacy.

Beyond the number of completed therapeutic materials, we were also interested in other measures of engagement with the AI-enabled therapy support tool to gain a more in-depth understanding of the link between app activity and treatment outcomes. As additional engagement measures, we investigated the overall app sessions and total time spent using the app.

We found statistically significant associations between the total number of sessions in the app and dropout rate (*r*=–0.25; *P*=.007) and increased reliable improvement (*r*=0.23; *P*=.01), recovery (*r*=0.20; *P*=.02), as well as reliable recovery (*r*=0.17; *P*=.046). The total number of sessions spent in the app was not correlated with attended appointments (*r*=0.13; *P*=.20) nor DNA sessions (*r*=–0.14; *P*=.18).

Finally, we were also interested in assessing potential links between time spent in the app with clinical and engagement outcomes. Similar to the results mentioned earlier, we found statistically significant associations between time spent in the app and dropout rate (*r*=–0.25; *P*=.007), increased reliable improvement (*r*=0.23; *P*=.01), recovery (*r*=0.20; *P*=.02), as well as the more stringent metric of reliable recovery (*r*=0.17; *P*=.046). Time spent in the app was not associated with attended appointments (*r*=0.14; *P*=.18) nor DNA sessions (*r*=–0.12; *P*=.21).

Higher Engagement with CBT Exercises Relative to Psychoeducation Predicts Better Treatment Adherence and Success

Of the 150 patients in the intervention condition, 118 (78.7%) completed at least 1 AI-supported CBT intervention (an interactive CBT exercise helping patients work through a problem) or psychoeducational material (educational material about CBT). Of these, 4 (2.7%) patients only completed interventions, and 14 (9.3%) patients only engaged with psychoeducational materials. The remaining 100 (66.7%) patients engaged with at least 1 CBT intervention and 1 psychoeducational material (number of interventions: mean 6.7, SD 4.78; number of psychoeducational materials: mean 6.5, SD 4.22). In this subgroup of 100 patients, we evaluated whether engagement with a specific type of AI-supported therapeutic material was predictive of treatment outcomes. To this end, we computed a use ratio of the number of CBT exercises to the number of psychoeducational materials completed (Figure 3A). We found that completing more CBT exercises, relative to psychoeducational materials, predicted better treatment adherence, for example, increased the number of attended

XSL•FO
RenderX

appointments (r=0.30; P=.001) and reduced the proportion of DNA sessions (r=–0.27; P=.003) as well as the dropout rate (r=–0.25; P=.002; Figures 3B-3D). Furthermore, the increased use of CBT interventions was also associated with an improvement in treatment success, for example, increased reliable improvement (r=0.23; P=.005), recovery (r=0.20; P=.01), and reliable recovery (r=0.17; P=.04; Figures 3E-3G). This suggests that especially the engagement with the personalized CBT exercises delivered through generative AI was associated with treatment adherence and treatment success compared to engagement with psychoeducational CBT materials.

Figure 3. Use ratio of cognitive behavioral therapy interventions relative to psychoeducational materials (A), correlations between use ratio and treatment engagement outcomes (B-D), and correlations between use ratio and treatment success outcomes (E-G). Error bars indicate standard error of the mean. CBT: cognitive behavioral therapy; DNA: did not attend.

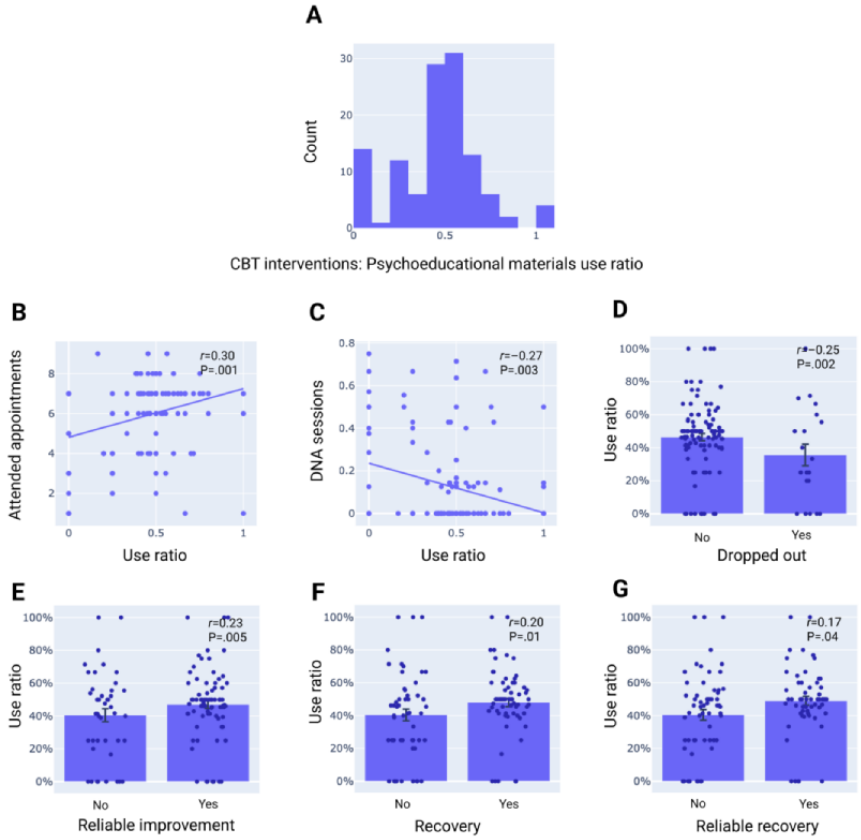

Other Analyses

User Experience

Beyond the investigation of the effects of the AI-enabled therapy support tool on treatment success and treatment adherence outcomes, we were also interested in conducting a qualitative investigation into the perceived benefits of using this AI-enabled therapy support tool as well as potential areas for improvement. Investigating potential benefits in more depth provides the opportunity to understand the mechanism of why the AI-enabled therapy support tool might have advantages over standard care.

Therefore, in a separate user experience study, 113 participants downloaded and used the Limbic Care app for 4 weeks and provided free-text answers on perceived benefits related to using the app as well as on potential areas for improvement. The sample had a mean age of 31.32 (SD 8.72) years and consisted of 53.1% (60/113) women. A total of 47 (41.6%) participants identified as White. Participants in our nonclinical sample experienced some symptoms of depression and anxiety (PHQ-9 score of ≥5 or GAD-7 score of ≥5); their scores were notably lower than the patient sample, as expected (PHQ-9: mean 10.15, SD 6.14; GAD-7: mean 8.74, SD 4.64).

Perceived Benefits

Of the 113 users surveyed, 92 (81.4%) mentioned a specific benefit of the AI-enabled therapy support tool to their mental health. Here, the most mentioned benefit (30/113, 26.5%) was improved awareness and clarity, indicating that the conversations with the AI-enabled therapy tool were especially helpful for users to think through their issues and gain better perspectives on their situation. The second most frequently reported benefit (23/113, 20.4%) consisted of coping strategies and CBT techniques, suggesting that AI-enabled support helped patients understand and apply CBT exercises to their day-to-day lives. A total of 14 (12.4%) users reported the positive benefit of mindfulness and relaxation, indicating that immediate support for dealing with stressful situations and helping to reduce emotional responses were important aspects of the tool. In addition, 9 (8%) users mentioned empathetic and nonjudgmental support as the main benefit, suggesting that a place to vent and be heard when no human is available was an important component of the tool. Finally, 15 (13.3%) users reported a broad positive impact on their mental health without focusing on any of these more specific categories.

XSL•FO
RenderX

Opportunities for Refinement

The themes and supporting participant quotes are presented in Table 3.

Of the 113 users, only 38 (33.6%) had suggestions for areas of potential improvement. The rest of the users had either no suggestions for improvement (n=63, 55.8%) or mentioned positive feedback even when queried about areas of improvement (n=12, 10.6%). Of the users who had suggestions for improvements, the most common comment (17/113, 15%) focused on the fact that while the chatbot was already very personalized, it was still possible to detect that it was not a human. In addition, 11 (9.7%) users had suggestions for specific additional features that they would either like to see or did not enjoy in the app. There were 4 (3.5%) users asking for more tailored notifications, and 3 (2.7%) users found that the additional guard rails to ensure patient safety restricted their user experience.

This indicated that, overall, users suggested few areas for improvement, with many of these dimensions being highly idiosyncratic to individual users.

Table 3. Overview of themes relating to the perceived benefits of using the artificial intelligence (AI)–enabled therapy support tool.

Theme	Definition	Example quotes
Emotional support and empathetic listening	The AI-enabled therapy support tool provided nonjudgmental and empathetic support, which enabled participants to open up about their problems even in the absence of another human being available to discuss their problems.	• "Limbic has helped me take the time to be grateful for things in my life, and to take the time to appreciate things about myself—even if they're small things.... I have done a couple of the exercises in the app, and they are helpful as well. I feel like I have learned new ways to cope with my mental health. It's also just really nice just openly talking to an AI who doesn't judge your feelings or anything." • "I feel more empowered to talk to my doctor and partner about my low mood and lack of energy. I like that I can talk to the chatbot when I feel like I can't control my emotions, and it helped having someone to be completely honest with."
Awareness, clarity, and new perspectives	The AI-enabled therapy support tool supported participants to think through their problems, identify their patterns, and come up with new perspectives on certain issues, resulting in improved clarity about feelings or problems.	• "I feel like I have gained some self-awareness and also that I now have tools and strategies to deal more effectively with my negative emotional states." • "The act of putting my thoughts and feelings into words... helps bring clarity to them for me." • "Thinking of how I should ground myself when I feel really overwhelmed. Trying to put things in a different perspective outside of my own head."
Coping strategies and CBT[a] techniques	The AI-enabled therapy support tool supported participants by providing information about CBT techniques and exercises, which were helpful to apply to the users' problems, helping them to better cope with their mental health.	• "I've been able to see the positive changes that come from opening up to someone and seeking advice on the issues that I'm dealing with. Each time that I have opened up in the Limbic app, it has provided me with a safe space, reassurance, it acknowledges me and validates me, and always provides me with options for strategies that might be helpful to me. I am very grateful to have an app like this on my phone!" • "I learned some new methods on how to deal with some of the problems I have."
Mindfulness and relaxation	The AI-enabled therapy support tool supported participants to be more mindful of their emotions and supported them in dealing with stressful situations, resulting in feeling more relaxed and balanced.	• "More self-awareness and mindfulness. Reminders to check in on myself and be kind to myself."
Gratitude	The AI-enabled therapy support tool supported participants to be more mindful and grateful for the positive experiences in their lives.	• "Setting and keeping track of my goals and being thankful for small things, thanks to the journaling activity." • "It has helped me acknowledge that there are good things in my life to be grateful for. It helped me realise that I do not need to only focus on what is going wrong."
General positive impact	The AI-enabled therapy support tool had generally broad positive impact on participants' mental health	• "It has improved my mental health." • "I think it helped to reduce my stress and my anxiety and enhanced my overall well-being."

[a]CBT: cognitive behavioral therapy.

XSL•FO
RenderX

Discussion

Principal Findings

We investigated the effects of a generative AI–enabled therapy support tool on treatment success and treatment adherence for patients undergoing group therapy in 5 of the UK NHS Talking Therapies provided by Everyturn Mental Health. A key part of therapy is the engagement with therapeutic materials that are assigned to the patient by the therapist. This engagement is crucial for applying tools and techniques learned in the therapy to real-life contexts [10]. In our study, the intervention group received in-between session support through an AI-enabled therapy support tool, while the control group completed the assignments using standard, static CBT worksheets.

Patients who had access to the AI-enabled therapy support tool showed significant improvement in clinical outcomes. Specifically, they achieved higher rates of reliable improvement, recovery, and reliable recovery compared to the control group. Similarly, patients using the AI-enabled therapy support tool attended more therapy sessions, were less likely to not show up to scheduled therapy sessions (that is, fewer DNAs), and were less likely to drop out from therapy. Importantly, the tool's use level was significantly associated with treatment success. Particularly, higher engagement with the AI-enabled therapy support tool was associated with increased reliable improvement and better recovery and reliable recovery. Furthermore, the use of the AI-enabled therapy support tool was significantly associated with treatment adherence, that is, app use was correlated with a higher number of attended therapy sessions and with a lower proportion of missed appointments and treatment dropouts. Therefore, our findings suggest that integrating the AI-enabled therapy support tool into standard therapy may lead to positive outcomes at the individual level by improving clinical outcomes. Moreover, the tool may save precious clinical time by reducing missed appointments and through this may reduce costs for the services.

Interpretation

Our findings indicate that using the AI-enabled therapy support tool to support patients between sessions can improve treatment outcomes. In line with previous research indicating that more meaningful engagement leads to increased intervention effectiveness [42] and hence better treatment success [43], we found that the use of the AI-enabled therapy support tool was associated with increased reliable improvement, recovery, and reliable recovery–the 3 core metrics for treatment success that are routinely assessed in NHS Talking Therapies [26]. One key aspect of our AI-enabled therapy support tool is personalization, a crucial component of CBT [44], which means that patients receive tailored information and interactions to support them in understanding how the CBT exercises relate to their situations and daily problems. Cultivating a therapeutic relationship or therapeutic alliance has been shown to increase engagement and predict positive clinical change [23]. In the case of our AI-enabled therapy support tool, we encouraged this using an interactive chat with the patients, which provided personalized and individualized assistance. It is noteworthy that these effects seem to be specifically linked to the delivery of CBT exercises through generative AI rather than only by displaying CBT psychoeducation materials in a digitized format per se. Moreover, the qualitative analysis of user feedback suggests that these generative conversations were especially useful for helping users explore new perspectives and views on their problems, resulting in increased clarity. These findings suggest a unique role of generative AI in the delivery of CBT exercises.

It is also worth noting that higher engagement with the app was linked to more attended sessions and a lower proportion of reported DNA sessions. This indicates that increased engagement between sessions is not only beneficial to the individual patients but also to the services that experience approximately 12% fewer missed appointments and 42% fewer treatment dropouts [45]. Furthermore, considering the substantial cost per reliable recovery in low-intensity treatment in NHS Talking Therapies, estimated at £1087 (US $1355.91) [46], the increase of 21 percentage points observed in the AI-enabled therapy support tool group translates to substantial cost savings. Assuming that the AI-enabled therapy tool is solely driving this 21% improvement, it would translate to £228 (US $284.40) of generated value per patient based on improved recovery alone, making it highly cost-effective. While future studies should assess the cost-effectiveness of such tools in detail, our data indicate that enhancing treatment effectiveness within NHS Talking Therapies through AI-enabled therapy support could lead to substantial cost savings and improved quality of care.

There has been a large societal debate about the use of AI, and especially LLMs, in clinical settings [47], emphasizing the importance of caution when introducing novel technologies into health care settings. Importantly, the AI-enabled therapy support tool has undergone rigorous testing as part of an International Organization for Standardization 13485–compliant quality management system, ensuring the product has gone through gold-standard risk management processes. Moreover, the AI-enabled therapy support tool was only complementing standard care provided by trained clinicians. Thus, the AI-enabled therapy support tool is only populated with therapeutic materials chosen by the clinician rather than making any clinical decisions itself. While this is a low-risk task, the positive impact on treatment outcomes was nevertheless substantial. These findings exemplify the importance of critically evaluating how AI can be used safely and effectively in clinical settings.

Generalizability

Importantly, our study uses data acquired from real patients receiving treatment from the UK NHS Talking Therapies, which means that the presented data provide high ecological validity for real-world clinical settings. Another advantage of our study is that the AI-enabled therapy support tool offered personalization through generative AI, and a lack of personalization has been identified as a key barrier to the uptake of digital solutions in previous research [22]. Lastly, the novel AI-enabled therapy support tool targeted support between therapy sessions, which is one of the most prominent reasons for disengagement [9].

XSL•FO
RenderX

Limitations

First, it is important to consider the observational design of this study. Following this promising observational data, a randomized controlled trial should be conducted. Second, while we were statistically controlling for several potential confounding factors (eg, patients' symptom severity), we did not have access to and were therefore unable to control for certain aspects, such as socioeconomic status. Nonetheless, given that the participants in both groups were from the same services, and thus similar geographic areas, we are not expecting major differences in their socioeconomic background. Studies with randomized designs may help to control for such confounds and would be a valuable future research direction. Finally, while we incorporated qualitative feedback about the perceived benefits of the AI-enabled therapy support tool, as well as potential areas where the tool could be improved, these data were collected from a nonclinical convenience sample of users. While the users showed elevated levels of depression and anxiety symptoms, it is possible that their views do not align with those of the patient sample. To this end, we would aim to incorporate a mixed methods process evaluation as part of a future randomized controlled trial evaluating the effectiveness of the tool.

Conclusions

In conclusion, this study showed that integrating an AI-enabled therapy support tool can lead to positive outcomes for patients and clinical services alike, by improving therapeutic outcomes and reducing costs associated with missed or dropout sessions. These are important findings considering that AI-enabled digital mental health tools are an emerging field and that they have the potential to increase access to treatment for patients and lessen the burden on mental health services.

Acknowledgments

This work was supported by Limbic Limited.

Data Availability

All analyses were based on sensitive clinical data, which we are not permitted to share with third parties. Summary statistics, as displayed in figures and text, are available upon request from the corresponding author.

Authors' Contributions

JH contributed to the conceptualization, methodology, investigation, project administration, formal analysis, visualization, and writing the original draft. LMD contributed to conceptualization, methodology, investigation, project administration, formal analysis, and writing the original draft. JM was responsible for formal analysis; visualization; and writing, reviewing, and editing the manuscript. MS contributed to the conceptualization; methodology; investigation; and writing, reviewing, and editing the manuscript. RH contributed to r conceptualization; supervision; and writing, reviewing, and editing the manuscript. TUH contributed to the conceptualization; methodology; supervision; and writing, reviewing, and editing the manuscript. MR was responsible for conceptualization; methodology; formal analysis; supervision; and writing, reviewing, and editing the manuscript.

Conflicts of Interest

JH, JM, RH, and MR are employed by Limbic Limited and hold shares in the company. LMD and TUH work as paid consultants for Limbic Limited. Moreover, TUH holds shares in the company.

Multimedia Appendix 1
App user interface, correlations between app engagement and treatment outcomes, and use ratio results.
[DOCX File , 16 KB - jmir_v27i1e60435_app1.docx]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
3. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
4. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
5. Cuijpers P, Miguel C, Harrer M, Plessen CY, Ciharova M, Ebert D, et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including

XSL•FO
RenderX

409 trials with 52,702 patients. World Psychiatry 2023 Feb 14;22(1):105-115 [FREE Full text] [doi: 10.1002/wps.21069] [Medline: 36640411]
6. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 01;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
7. Depression in adults: treatment and management. National Institute for Clinical Excellence. URL: https://www.nice.org.uk/guidance/ng222 [accessed 2024-02-08]
8. Generalised anxiety disorder and panic disorder in adults: management. National Institute for Clinical Excellence. 2011. URL: https://www.nice.org.uk/guidance/cg113 [accessed 2024-02-08]
9. Glenn D, Golinelli D, Rose RD, Roy-Byrne P, Stein MB, Sullivan G, et al. Who gets the most out of cognitive behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. J Consult Clin Psychol 2013 Aug;81(4):639-649 [FREE Full text] [doi: 10.1037/a0033403] [Medline: 23750465]
10. Kazantzis N, Deane FP, Ronan KR. Assessing compliance with homework assignments: review and recommendations for clinical practice. J Clin Psychol 2004 Jun 08;60(6):627-641. [doi: 10.1002/jclp.10239] [Medline: 15141396]
11. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The relationship between homework compliance and therapy outcomes: an updated meta-analysis. Cognit Ther Res 2010 Oct;34(5):429-438 [FREE Full text] [doi: 10.1007/s10608-010-9297-z] [Medline: 20930925]
12. Kazantzis N, Whittington C, Dattilio F. Meta-analysis of homework effects in cognitive and behavioral therapy: a replication and extension. Clin Psychol Sci Pract 2010 Jun;17(2):144-156. [doi: 10.1111/J.1468-2850.2010.01204.X]
13. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behav Ther 2016 Sep;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002] [Medline: 27816086]
14. Lebeau RT, Davies CD, Culver NC, Craske MG. Homework compliance counts in cognitive-behavioral therapy. Cogn Behav Ther 2013;42(3):171-179 [FREE Full text] [doi: 10.1080/16506073.2013.763286] [Medline: 23419077]
15. Fanous M, Daniels J. What are the recovery and attrition outcomes for group CBT and individual CBT for generalised anxiety disorder in an IAPT service? An exploratory study. Cogn Behav Ther 2020 Jun 09;13:e521. [doi: 10.1017/S1754470X20000045]
16. Fernandez E, Salem D, Swift JK, Ramtahal N. Meta-analysis of dropout from cognitive behavioral therapy: magnitude, timing, and moderators. J Consult Clin Psychol 2015 Dec;83(6):1108-1122. [doi: 10.1037/ccp0000044] [Medline: 26302248]
17. Hans E, Hiller W. Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: a meta-analysis of nonrandomized effectiveness studies. J Consult Clin Psychol 2013 Feb;81(1):75-88. [doi: 10.1037/a0031080] [Medline: 23379264]
18. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. JMIR Ment Health 2017 Jun 08;4(2):e20 [FREE Full text] [doi: 10.2196/mental.5283] [Medline: 28596145]
19. Demszky D, Yang D, Yeager DS, Bryan CJ, Clapper M, Chandhok S, et al. Using large language models in psychology. Nat Rev Psychol 2023 Oct 13;54:547-577. [doi: 10.1038/S44159-023-00241-5]
20. McFadyen J, Habicht J, Dina LM, Harper R, Hauser TU, Rollwage M. AI-enabled conversational agent increases engagement with cognitive-behavioral therapy: a randomized controlled trial. medRxiv Preprint posted online November 2, 2024. [doi: 10.1101/2024.11.01.24316565]
21. Rollwage M, Habicht J, Juechems K, Carrington B, Viswanathan S, Stylianou M, et al. Using conversational AI to facilitate mental health assessments and improve clinical efficiency within psychotherapy services: real-world observational study. JMIR AI 2023 Dec 13;2:e44358 [FREE Full text] [doi: 10.2196/44358] [Medline: 38875569]
22. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
23. Valentine L, D'Alfonso S, Lederman R. Recommender systems for mental health apps: advantages and ethical challenges. AI Soc 2022 Jan 17:1-12 [FREE Full text] [doi: 10.1007/s00146-021-01322-w] [Medline: 35068708]
24. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med 2023 Oct 30;29(10):2396-2398. [doi: 10.1038/s41591-023-02412-6] [Medline: 37391665]
25. The clinical AI patient companion for on-demand support. Limbic. URL: https://www.limbic.ai/care [accessed 2024-04-29]
26. NHS Talking Therapies for anxiety and depression manual. The National Collaborating Centre for Mental Health. URL: https://www.england.nhs.uk/wp-content/uploads/2018/06/nhs-talking-therapies-for-anxiety-and-depression-manual-v7.pdf [accessed 2024-04-29]
27. Rollwage M, Juchems K, Pisupati S, Prichard G, Balogh A, McFadyen J, et al. The Limbic Layer: transforming large language models (llms) into clinical mental health experts. PsyArXiv Preprint posted online November 27, 2024 [FREE Full text] [doi: 10.31234/osf.io/9d7tp]
28. Scrivener R, National Institute for Clinical Excellence, Royal College of Nursing, Royal College of Nursing, Commission for Health Improvement. Principles for Best Practice in Clinical Audit. Abingdon, UK: Radcliffe Medical; 2002.

XSL•FO
RenderX

29. Gyani A, Shafran R, Layard R, Clark DM. Enhancing recovery rates: lessons from year one of IAPT. Behav Res Ther 2013 Sep;51(9):597-606 [FREE Full text] [doi: 10.1016/j.brat.2013.06.004] [Medline: 23872702]
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
31. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
32. Our proprietary pool is thoroughly verified - and 100% human. Prolific. URL: https://www.prolific.com/ [accessed 2024-04-29]
33. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
36. Oldham M, Dina LM, Loebenberg G, Perski O, Brown J, Angus C, et al. Evaluating the acceptability of the drink less app and the national health service alcohol advice web page: qualitative interview process evaluation. J Med Internet Res 2024 Jul 18;26:e42319 [FREE Full text] [doi: 10.2196/42319] [Medline: 39024575]
37. Korstjens I, Moser A. Series: practical guidance to qualitative research. Part 4: trustworthiness and publishing. Eur J Gen Pract 2018 Dec;24(1):120-124 [FREE Full text] [doi: 10.1080/13814788.2017.1375092] [Medline: 29202616]
38. Ahmed SK. The pillars of trustworthiness in qualitative research. J Med Surg Public Health 2024 Apr;2:100051. [doi: 10.1016/j.glmedi.2024.100051]
39. Rollwage M, Habicht J, Juchems K, Carrington B, Hauser TU, Harper R. Conversational AI facilitates mental health assessments and is associated with improved recovery rates. BMJ Innov 2024 Mar 01;10(1-2):4-12. [doi: 10.1136/bmjinnov-2023-001110]
40. Toyomoto R, Sakata M, Yoshida K, Luo Y, Nakagami Y, Uwatoko T, et al. Prognostic factors and effect modifiers for personalisation of internet-based cognitive behavioural therapy among university students with subthreshold depression: a secondary analysis of a factorial trial. J Affect Disord 2023 Feb 01;322:156-162 [FREE Full text] [doi: 10.1016/j.jad.2022.11.024] [Medline: 36379323]
41. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006 Jan;163(1):28-40. [doi: 10.1176/appi.ajp.163.1.28] [Medline: 16390886]
42. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
43. Gan DZ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
44. Cronin TJ, Lawrence KA, Taylor K, Norton PJ, Kazantzis N. Integrating between-session interventions (homework) in therapy: the importance of the therapeutic relationship and cognitive case conceptualization. J Clin Psychol 2015 May 21;71(5):439-450. [doi: 10.1002/jclp.22180] [Medline: 25809713]
45. Davis A, Smith T, Talbot J, Eldridge C, Betts D. Predicting patient engagement in IAPT services: a statistical analysis of electronic health records. Evid Based Ment Health 2020 Feb;23(1):8-14 [FREE Full text] [doi: 10.1136/ebmental-2019-300133] [Medline: 32046987]
46. Steen S. A cost-benefit analysis of the improving access to psychological therapies programme using its key defining outcomes. J Health Psychol 2020;25(13-14):2487-2498. [doi: 10.1177/1359105318803751] [Medline: 30284915]
47. AI Chat used by mental health tech company in experiment on real users. NBC News. URL: https://www.nbcnews.com/tech/internet/chatgpt-ai-experiment-mental-health-tech-app-koko-rcna65110 [accessed 2024-03-19]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
DNA: did not attend
GAD-7: Generalized Anxiety Disorder-7
LLM: large language model
NHS: National Health Service
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9

XSL•FO
RenderX

Edited by A Coristine; submitted 10.05.24; peer-reviewed by S Joshi, T Tillmann; comments to author 06.08.24; revised version received 22.10.24; accepted 21.12.24; published 10.03.25.

Please cite as:
Habicht J, Dina LM, McFadyen J, Stylianou M, Harper R, Hauser TU, Rollwage M
Generative AI–Enabled Therapy Support Tool for Improved Clinical Outcomes and Patient Engagement in Group Therapy: Real-World Observational Study
J Med Internet Res 2025;27:e60435
URL: https://www.jmir.org/2025/1/e60435
doi:10.2196/60435
PMID:40063074

©Johanna Habicht, Larisa-Maria Dina, Jessica McFadyen, Mona Stylianou, Ross Harper, Tobias U Hauser, Max Rollwage. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Caregiving Artificial Intelligence Chatbot for Older Adults and Their Preferences, Well-Being, and Social Connectivity: Mixed-Method Study

Brooke H Wolfe[1], PhD; Yoo Jung Oh[1], PhD; Hyesun Choung[2], PhD; Xiaoran Cui[1], BA; Joshua Weinzapfel[1], ABA; R Amanda Cooper[3], PhD; Hae-Na Lee[4], PhD; Rebecca Lehto[5], RN, PhD

[1]Department of Communication, Michigan State University, East Lansing, MI, United States

[2]Brian Lamb School of Communication, Purdue University, West Lafayette, IN, United States

[3]Department of Communication, University of Connecticut, Storrs, CT, United States

[4]College of Engineering, Michigan State University, East Lansing, MI, United States

[5]College of Nursing, Michigan State University, East Lansing, MI, United States

Corresponding Author:
Brooke H Wolfe, PhD
Department of Communication
Michigan State University
404 Wilson Road, Room 473
East Lansing, MI, 48824
United States
Phone: 1 517 355 3470
Email: wolfebro@msu.edu

Abstract

Background: The increasing number of older adults who are living alone poses challenges for maintaining their well-being, as they often need support with daily tasks, health care services, and social connections. However, advancements in artificial intelligence (AI) technologies have revolutionized health care and caregiving through their capacity to monitor health, provide medication and appointment reminders, and provide companionship to older adults. Nevertheless, the adaptability of these technologies for older adults is stymied by usability issues. This study explores how older adults use and adapt to AI technologies, highlighting both the persistent barriers and opportunities for potential enhancements.

Objective: This study aimed to provide deeper insights into older adults' engagement with technology and AI. The technologies currently used, potential technologies desired for daily life integration, personal technology concerns faced, and overall attitudes toward technology and AI are explored.

Methods: Using mixed methods, participants (N=28) completed both a semistructured interview and surveys consisting of health and well-being measures. Participants then participated in a research team–facilitated interaction with an AI chatbot, Amazon Alexa. Interview transcripts were analyzed using thematic analysis, and surveys were evaluated using descriptive statistics.

Results: Participants' average age was 71 years (ranged from 65 years to 84 years). Most participants were familiar with technology use, especially using smartphones (26/28, 93%) and desktops and laptops (21/28, 75%). Participants rated appointment reminders (25/28, 89%), emergency assistance (22/28, 79%), and health monitoring (21/28, 75%). Participants rated appointment reminders (25/28, 89.3%), emergency assistance (22/28, 78.6%), and health monitoring (21/28, 75%) as the most desirable features of AI chatbots for adoption. Digital devices were commonly used for entertainment, health management, professional productivity, and social connectivity. Participants were most interested in integrating technology into their personal lives for scheduling reminders, chore assistance, and providing care to others. Challenges in using new technology included a commitment to learning new technologies, concerns about lack of privacy, and worries about future technology dependence. Overall, older adults' attitudes coalesced into 3 orientations, which we label as technology adapters, technologically wary, and technology resisters. These results illustrate that not all older adults were resistant to technology and AI. Instead, older adults are aligned with categories on a spectrum between willing, hesitant but willing, and unwilling to use technology and AI. Researchers can use these findings by asking older adults about their orientation toward technology to facilitate the integration of new technologies with each person's comfortability and preferences.

XSL•FO
RenderX

Conclusions: To ensure that AI technologies effectively support older adults, it is essential to foster an ongoing dialogue among developers, older adults, families, and their caregivers, focusing on inclusive designs to meet older adults' needs.

(J Med Internet Res 2025;27:e65776) doi:10.2196/65776

KEYWORDS

older adults; technology use; AI chatbots; artificial intelligence; well-being; social connectedness; mobile phone

Introduction

Background

Both global life expectancy and the aging population have continued to increase in recent years [1]. As of 2023, 1 billion people in the world were aged 60 years or older and numbers are expected to rise to 1.4 billion by 2030 [2]. This trend presents challenges for maintaining the well-being of older adults, many of whom choose independent living while requiring assistance with daily tasks and health care routines, as well as support to stay socially connected [2,3].

Advancements in artificial intelligence (AI) technologies have transformed numerous sectors, including health care and caregiving (eg, supporting people with dementia) [4]. AI tools are now being used to monitor health, provide reminders for medication and appointments, and even offer companionship through conversational agents [5,6].

Despite the growing interest and potential in leveraging AI for older adult populations, challenges remain, particularly regarding the suitability of these technologies for older adults. One major issue is the limited accommodation for the unique needs of older adults, such as user-friendliness for those with limited experience using digital resources [7]. Security and privacy concerns are also significant barriers that can hinder older adults' willingness to adopt AI agents for personal health care [8]. Finally, limited research on their long-term effectiveness underscores the need for more studies to ensure these technologies are sustainably introduced and adopted by older people [4,5].

Acknowledging the need for a personalized, older adult-friendly system that addresses the challenges and needs of older adults in their daily lives, this study aims to explore older adults' current use, challenges, and desired needs associated with AI technology, specifically focusing on caregiving chatbots. As a result, this study seeks to provide a groundwork that will inform the development of AI technologies tailored to improve the well-being of older adults.

AI Technologies in Health Care and Caregiving for Older Adults

While older adults use technology less frequently and in fewer ways than younger individuals, they are still capable of engaging with specific technologies that meet their needs [9,10]. Voice-based chatbots, for example, have been shown to help with medication adherence and reducing loneliness and social isolation, provided they are designed with older users in mind [8,11,12].

The rapid advancement of multimodal large language models like ChatGPT (OpenAI) suggests that many existing barriers for conversational agents in health and well-being may soon be overcome [13,14]. These advancements are expected to enhance conversational competence, domain-specific information, and personalization. Such progress can potentially address older adults' concerns with existing technologies and improve their experience using these technologies.

Recent developments have focused on creating AI technologies specifically for older adults, aiming to address their needs for companionship, support, and health management. SeniorTalk [15], for example, offers an AI-powered chat companion that can be personalized through the selection of different personas. ElliQ (Intuition Robotics) is a social robot that uses conversational AI to engage older adults in daily activities, provide reminders, and facilitate social interaction to help reduce loneliness [16]. These technologies reflect the ongoing efforts to adapt AI applications to the specific needs of older adults. However, the deployment of AI in social care must be approached cautiously, ensuring that these technologies are implemented in ways that uphold the rights and dignity of older individuals [17].

While previous studies have explored older adults' needs and concerns regarding technology use and the requirements for technology to assist aging through input from stakeholders like care professionals, technology designers, and policymakers, these insights have not been fully reflected in the development of technology [18,19]. Additional insight into older adults' preferences for using technology is needed to tailor technological tools to benefit older users.

Social Support and Older Adults' Health and Well-Being

AI technology carries the potential as a source of social support that is greatly needed among the most vulnerable older adult population. Approximately a quarter of older adults live in social isolation and approximately one-third of older adults experience loneliness [20,21]. Loneliness is associated with many negative consequences including poor mental and physical health, lower self-rated well-being, and disrupted sleep quality [19]. Among older adults specifically, loneliness is also associated with reduced physical activity, impaired cognition, dementia progression, nursing home placement, and higher mortality rates [22]. Older adults with limited social networks and low social engagement are at greatest risk [23]. AI chatbots may provide a valuable source of companionship and social support, potentially addressing loneliness. The current study explores the technologies older adults currently use, what they are interested in integrating into their daily lives, the challenges they face, and their overall attitudes toward technology and AI. To examine these topics, 4 research questions are addressed:

XSL•FO
RenderX

RQ1: In what ways do older adults report using technology and AI?

RQ2: What technology and AI are older adults interested in integrating into their daily lives?

RQ3: What concerns and challenges do older adults report about technology and AI use?

RQ4: What are older adults' attitudes toward technology and AI?

Methods

Participant Recruitment and Data Collection

A mixed methods cross-sectional descriptive design was used. Participants were recruited using the website of a community research pool, a research participation system, managed by a large midwestern university. Participants needed to meet these criteria: (1) at least 65 years old and (2) able to complete the interview in person at the university site. The community research pool system posted a short synopsis of the study on their website including the inclusion criteria, compensation for participants for completing the study qualification survey and a 45-minute interview, and researcher contact information. Participants selected an interview date and time through the community research pool system. Once a time was selected, a member of the research team reached out to confirm the interview schedule, provide instructions for the interview location (conference room at a university) and parking, and answer any questions through email. A total of 33 participants signed up for the study, with 5 dropping out before the interview due to scheduling conflicts, resulting in 28 participants who participated and completed the study.

After arriving for their interview, participants completed informed consent followed by a survey that included demographics, health, and well-being measures on either an iPad (Apple Inc) or paper, depending on personal preference. The participants then completed the interview with 2 members of the research team. One member led the interview by asking protocol questions. The interview was divided into questions about daily schedules (eg, What are routines that you do every day?), knowledge of technology and AI (eg, What technology do you use daily?), and social support desired for technology and AI (eg, What, if any, kind of challenges do you face when learning new technologies?). The second team member facilitated the participant's interaction with an AI chatbot, Amazon Alexa. The participant was prompted to acknowledge the chatbot and then use multiple features, including asking for a recipe, playing a trivia game, and requesting the news from Alexa. Pseudonyms were assigned to each participant following the interview. The research team consisted of 5 women faculty members (PhD), 1 woman who is currently a graduate student, and 1 man who is currently an undergraduate student. The lead researcher had previously conducted multiple qualitative studies involving interviews. No previous relationship was established with participants before the interview. Researchers informed participants that the study was a part of a research project and disclosed their professional interests. Audio recordings were collected and transcribed for data analysis. They were not returned to participants for comments.

Statistical Analysis

Descriptive statistics were used for analyzing the surveys. A thematic analysis of the interview data was conducted following phases of reflexive thematic coding, which included data familiarization, systematic coding, generating initial themes, developing and reviewing themes, refining and naming themes, and writing the manuscript [24]. To ensure the robustness of our analysis, we used 5 verification procedures: referential adequacy, peer debriefing, negative case analysis, audit trail, and exemplar identification [25,26]. Initially, we split the data to achieve referential adequacy and, after reaching saturation in the first half, we analyzed the second half of the data and did not find any new themes [27]. The first and fourth authors held 2 peer debriefing meetings to resolve differences and reach consensus on all themes. Continuous peer debriefing occurred throughout the analysis, and labels for the findings were determined collaboratively. We refined our analysis to account for all data, meeting the standard for negative case analysis [25]. Detailed notes were kept (ie, an audit trail), informing both the analysis and the selection of exemplars [26].

Survey Measures

The study collected data on the following measures: demographics, general health status, Behavioral Risk Factor Surveillance System (BRFSS) Disability Questions, loneliness (3-item Loneliness Scale), life satisfaction (The Satisfaction with Life Scale) listed in Table 1 and digital literacy (Digital Health Technology Literacy Assessment Questionnaire), AI literacy, and interests in AI chatbot features for caregiving assistance, reported in Table 2 [28-32].

Table 1. Quantitative summary of older adults' health and assistive technology use reported in study survey (N=28).

Health factors	Values
Health status, mean (SD)	3.6 (0.9)
Limited in activities, n (%)	9 (32)
Use Alexa as an assistive technology, mean (SD)	3.6 (1)
Loneliness, mean (SD; range)	4.1 (1.6; 3-8)
Life satisfaction, mean (SD; range)	3.9 (0.7; 2-5)

XSL•FO
RenderX

Table 2. Quantitative summary of older adults' current technology use reported in the study survey (N=28).

Item	Responses, n (%)
AI[a] chatbot use	
Amazon Alexa	10 (36)
Google Assistant	6 (21)
Apple Siri	9 (32)
Facebook Messenger Bot	3 (11)
ChatGPT	2 (7)
Copilot	1 (4)
Challenges using technology devices	
Never	2 (7)
Rarely	6 (21)
Sometimes	16 (57)
Often	3 (11)
Always	1 (4)
Knowledge about AI	
Don’t have much knowledge	9 (32)
Have knowledge	19 (68)
AI chatbots’ caregiving features interested in adopting	
Appointment reminder	25 (89)
Information access (eg, news, weather)	22 (79)
Emergency assistance	22 (79)
Health monitoring	21 (75)
Daily medication reminders	21 (75)
Fall detection and alert notifications	18 (64)
Interactive entertainment	17 (61)
Social connection	17 (61)
Shopping management	17 (61)
Physical activity encouragement	13 (46)
Home automation control	13 (46)
Mental health support	11 (39)
Customized health tips and reminders	11 (39)
Daily routine assistance	11 (39)
Memory aids	11 (39)
Dietary assistance	9 (32)
Language and communication	2 (7)

[a]AI: artificial intelligence.

Ethical Considerations

The institutional review board at Michigan State University determined that the study was exempt before study recruitment was initiated (STUDY00010214). Each person consented to participate in the study at the beginning of the survey and again at the beginning of their interview. Participants were informed that they could end the interview at any time and could skip any questions they did not feel comfortable answering. Each interview was recorded through Zoom (Zoom Communications Inc). The transcripts were reviewed, and all identifiable information was removed during the transcription process to protect participant information. Participants chose between US $50 cash or an Amazon gift card of the same amount as compensation for their time.

XSL•FO
RenderX

Results

Survey Findings

Among 28 participants, the average age was 71 years (SD 5.5), ranging from 65 to 84. More than half identified themselves as women (57%,16/28) and married or living with partners (61%, 17/28). The remaining participants were widowed, divorced, or separated (29%, 8/28) or never married or single (11%, 3/28). Most participants were living with their partner (61%, 17/28) or their partner and adult children 7% (2/28). Some participants reported that they were living alone (29%, 8/28). Almost all participants were white (89%, 25/28), followed by Asian (7%, 2/28), and black orAfrican American (4%, 1/28). Most participants were currently unemployed or retired (75%, 21/28). The highest level of education obtained by participants was, 43% (12/28) had completed high school, general educational development, or some college, while 57% (18/28) were college graduates or had attained a higher level of education. Participants' household income levels were under US $50,000 (25%, 7/28), between US $50,000 and US $100,000 (36%, 10/28), between US $100,000 and US $150,000 (29%, 8/28), and above US $150,000 (4%, 1/28).

Tables 1 and 2 present a detailed report of participant profiles, including older adults' health factors, technology and AI use, challenges, knowledge, and their intention to adopt AI features to support daily health routines. Participants' average health status was 4 (SD 0.9) which was deemed to be good (from 1 being poor to 5 being excellent), and 32% (9/28) reported that they are limited in activities due to physical conditions. The average loneliness felt by the participants was 4 (SD 1.6) [Range 3-8] which indicated that participants felt lonely some of the time. Life satisfaction was 4 (SD 0.7) [Range 2-5] meaning that participants generally were satisfied with their lives.

Most participants were generally familiar with using technology. For chatbot use, 64% (18/28) reported that they have previously used AI chatbots. Participants were generally comfortable with interacting with a chatbot (mean 4, [SD 1.0], range 2-5). Regular use of technology devices was mostly related to smartphone (93%, 26/28) and desktop use (75%, 21/28), followed by iPad or tablet (46%, 13/28). Other types of technology included smartwatch (14%, 4/28), robot vacuum cleaner (18%, 5/28), and robot lawnmower (4%, 1/28). For regular application use, 75% (21/28) used social media, 71% (20/28) used online shopping, 64% (18/28) used banking and streaming services, 57% (16/28) health application, and 39% (11/28) used customer service chatbots. Participants reported that they face challenges using technology devices never (7%, 2/28), rarely (21%, 6/28), sometimes (57%, 16/28), often (11%, 3/28), and always (4%, 1/28). 32% (9/28) reported that they do not have much knowledge about AI and 68% (19/28) reported to have knowledge about AI.

When participants were asked if there were any AI chatbots' caregiving features they may be interested in adopting, appointment reminder was rated the highest (89%, 25/28), followed by information access such as news or weather (79%, 22/28), emergency assistance (79%, 22/28), health monitoring (75%, 21/28), daily medication reminders (75%, 21/28), fall detection and alert notifications (64%, 18/28), interactive entertainment (61%, 17/28), social connection (61%, 17/28), shopping management (61%, 17/28), physical activity encouragement (46%, 13/28), home automation control (46%, 13/28), mental health support (39%, 11/28), customized health tips and reminders (39%, 11/28), daily routine assistance (39%, 11/28), memory aids (39%, 11/28), dietary assistance (32%, 9/28), and language and communication (7%, 2/28).

Qualitative Findings

Overview

Older adults discussed their current and desired use of technology and AI in their daily lives. These discussions were depicted alongside key concerns and challenges that impacted the degree to which participants were willing to accept and integrate these technologies. These considerations informed the construction of 3 distinct attitudes toward technology and AI. These findings respond to 4 research questions outlined.

RQ1: Current Technology and AI Use

The most common technologies participants mentioned are digital devices such as smartphones, smartwatches, computers, and radios (Table 2). These technologies were used for a range of purposes, including entertainment, health management, professional productivity, and social connectivity, illustrating that older adults used these technologies to enhance their quality of life, manage health, and mitigate social isolation.

Many participants reported using a combination of digital devices smartphones, smartwatches, and computers that are often integrated. For instance, Everett mentioned, "I have an iPhone, an iMac, and an iPad." These interconnected ecosystems facilitated a convenient and useable community of digital devices that allowed older adults to engage online. In addition, older adults commonly use radio applications and devices to listen to stories, news, and music. Dennis expressed, "You know, I hear all these things on the radio. I listen to the radio every day." The digital devices used by older adults facilitated their existing implementation of these technologies for social interaction, entertainment, health management, and aiding productivity. Each of these applications is overviewed in Table 2.

RQ2: Interested Integration

Older adults also reflected on areas of their lives where they were not currently using technology but would consider the addition of technology as helpful. Participants communicated that technology would be beneficial for scheduling reminders, chore assistance, and providing care to others. These categories illustrate how older adults desire to use technology to maintain their independence and manage daily tasks more effectively. These findings offer insights into the type of tasks proposed as helpful, outlined in Tables 3 and 4.

XSL•FO
RenderX

Table 3. Examples of current technology use that older adults described in the study interview.

Current technology use	Examples
Social interaction	
Technology helps older adults maintain social connections through social media, email, and video calls, combating feelings of isolation.	• "I guess, I'm in a LGBT and also friendship group…. it's a group of gay people, primarily men who meet once a month at a different restaurant, just a friendship group to explore a different restaurant[s]...very excited." (Raymond) • "it's just kind of our normal. Sunday, you know, church stream. I would say I went more in person before the pandemic, more so before the pandemic. Yeah, and we've all changed our habits I think since then." (Christine)
Entertainment	
Older adults engage with technology content, like streaming movies, listening to music, and playing games to stay mentally active.	• "...it's radio. Sometimes you can, it just gives out the full like the visual kind of like a video version of the news as well. So, it just depends on what you want, you know, like, you can ask ABC news." (Cheryl) • "Well, most of my downtime I probably spend on YouTube because there's terrific history and documentary channels that I follow….I [also] read all my news and everything on online with. I go to a bunch of the different websites so I can get a wider perspective on what people are thinking." (Ingrid)
Health management	
Technologies can assist older adults in health-relates tasks, such as scheduling doctor's appointments, tracking fitness, and managing medication efficiently.	• "I wear hearing aids, and they've got Bluetooth on them, so, you know, I can listen to music through my phone and stuff like that." (Sandra) • "I set my phone as soon as I've eaten, and I reset it for two hours intervals. And then it has a reminder on it to take my vitamins." (Gwen)
Assisted productivity	
Older adults use devices professional activities like remote work, scheduling, and task management.	• She's [Alexa] a good reminder-er. I mean, I've got to tell her to, feel free to tell me to hurry up and sent out Easter cards, because Easter cards really soon….you know can I put egg shells down the garbage disposal and she's No no no no no no, don't do that. I didn't really remember that I couldn't do that. (Devorah) • "So, there's a lot of communication on my phone to them [church group]. I'm in a group of about 500 people, and there's usually things going on with them. Like, right now, we have a friend who's in the hospital dying, so we're making arrangements for him. Just a lot of arranging. You know, I'm a connector person….I'm the head of a couple groups, so sometimes people contact me through email." (Donna)

XSL•FO
RenderX

Table 4. Example of desired technology use that older adults described in the study interview.

Desired technology use	Example
Scheduling reminders	
Older adults want some tools to help them stay organized by reminding them of tasks and events, thus supporting their daily routines and commitments	• ["I use] auto bill pay I never go to the bank. Everything is auto paid. It's, I mean prescriptions. I only take one, but, you know, it just pops into your thing [phone]." (Mary) • "….at work and I do have a medication [reminder], and some reminders on my phone that remind me to turn in my time sheet or a medication reminder. Yeah, reminders that I need to take my medication, a second medication." (Joyce)
Chore assistance	
Older adults are interested in technologies that simplify and automated daily chores to enhances convenience and reduce effort, such as vacuuming, mowing driving cleaning, and cooking.	• "I would love to not have to mow my lawn. If that was automated, that would be awesome. See, I would love not to have to cook. That would be great. Basically, I'd like to live in the jetsons…it would be great to have somebody tale your dishes and not have to do that." (Ingrid) • "So, you know, if something would just automatically cook my food for me, just what I want for dinner without it being Grubhub or Uber Eats or somethings. Yeah, just, you know, my own food because I like my own cooking." (Christine)
Caregiving monitoring	
Older adults expressed interest in technology/AI that would assist them in observing, supervising, and managing the care of others in their life. Participants clarified that they might want caregiving assistance in the future, but at this point, they were focused on completing care tasks for others including their own parents or partner.	• "More support? Not yet…I mean, I drove here…[but] a device for who could tell if I if I had fallen down without me having to hit a button or anything…but that that's something we would. Yeah. I'd be interested in." (Timothy) • "It has to do with my caregiving…I'm always trying to find things that might help. My mom and I found online I heard a story on the radio about this company called LEC, and it's this little box and it's got this head on it, like with a camera. And you talk to it and it, like, lights up, it points at you has a screen, and you can ask it questions and you can have what I was hoping it would do is while I go, I'm gone, like running errands or I'm on the road or whatever I can call in, seen my mom and check on her and talk back and forth." (Dennis)

RQ3: Concerns and Challenges

Older adults specified several concerns and challenges that influenced their use of technology and AI. These barriers impacted their adaptation to new devices and their ongoing worries about the technology already associated with their routines.

Commitment to Learning

Most older adults articulated that the time and energy to learn new technology was not a worthwhile investment. Patricia remarked, "I'm sort of like this with technology, I use what I have to use...Now, if I wasn't retired, I'd be embracing it a thousand percent because it would be necessary. But yeah, I don't need to be." Patricia differentiated the time period of retirement as 1 in which people no longer needed to keep up with innovations. Instead, she reiterated that her current use was at a level that she was comfortable continuing at. Other older adults shared this view, like Sandra who limited her activity on social media, stating:

> *I use Instagram and Facebook, but I just review it. It turns out every time I try to say anything on it, I don't say what I mean. I said happy anniversary to somebody once, but I sent them a picture of a pizza. If that gives you an idea why I don't type on Instagram.*

Sandra focused on completing online tasks that matched her skill level with the tool. She wanted to review the content to stay connected on interpersonal updates but following miscommunication with the addition of a meme, she limited her participation. Older adults like Patricia and Sandra were motivated to use technology with their current abilities but did not prioritize allocating time and effort to using new technologies or improving their capacity with existing offerings.

Lack of Privacy

Another concern that older adults expressed was how AI technologies could threaten their personal privacy. Given the growth of technology throughout their lifetime and the rapid change of the AI landscape in the last few years, many stated concerns about insufficient regulation regarding these digital tools. Anthony conveyed this uneasiness, saying:

> *Part of the problem, I think, is people don't realize how much of their rights they're signing away when they just click through or the acknowledge or whatever...>With digital data, some of its okay, I'm not worried about the fact that I've got, you know, this Google doc or that Google Sheet or whatever...[B]ut what do you think about what's going on in your household? And you know, and think about some things, some artificial intelligence in this instance that's listening and maybe sending information all the time about everything.*

Anthony compared content that had a minimal need for privacy with parts of his life he considered deserving of more extensive restrictions. Anthony was 1 of multiple older adults who questioned if bringing technology into their home was worth the risk of sharing information and communication about their

XSL•FO
RenderX

household with technology companies. Often older adults were not only worried about their own privacy but also articulated that these considerations were likely not considered by other, less technologically knowledgeable older adults.

Dependence on Technology

Older adults also discussed their fear of technology replacing their performance of daily tasks which would ultimately lead to cognitive decline. This concern was shared among older adults who were concerned that technology was an assistant for laziness and that relying on these tools reflected a negative moral character. Exemplifying this, Gwen explained:

> *Probably as I get older, I might could use some more support. But I also feel that I'm a big believer if you don't use what you have, you lose it, so if I don't use the knowledge I have, or the memory I have, or some of that, then it just goes off because you're not helping that muscle to keep developing.*

Gwen offers an analogy of exercising her brain with tasks that keep her mind active. This analogy leaves technology as a cheat tool that limits this exercise from taking place. Older adults integrated the language of exercise while expressing their worries that technology would replace the routines they had in place to support their brain health.

Overall, older adults expressed their concerns with learning new technologies, the lack of privacy related to the implementation of these technologies, and the impact that these devices would have on their health. These concerns informed how older adults were oriented toward technology and AI regarding their willingness to use these digital devices.

RQ4: Attitudes Toward Technology

Overview

Older adults reported having diverse attitudes toward technology and AI. These orientations influence the degree to which older adults were willing to engage with technology and the attributes they described technology with. The orientations ranged from *adapters* who were willing and able to use a range of technology to support their life to *resisters,* who were adults who articulated their aversion to integrating technology into their daily lives. The adaptor and resister orientations represented the opposing positions and incorporating aspects of both orientations, some older adults described a *wary* disposition.

Adapters

Adapters to technology and AI encompassed participants who were most willing and excited to merge digital devices into their everyday routines. Many older adults embraced the positive attributes and the assistance that technology brought to their lives. For instance, Christine remarked:

> *When you stop and think about what my parents had to do to get through a day, I'm almost embarrassed that my life is so easy. It's like Thursday's the day to turn on the robotic vacuum cleaner. And I think, okay, I dont even have to remember to put on the dishwasher. You know, you're just embarrassed that people had to physically do this work and take up all day. And it's so easy for us...Our last two homes, we did have thermostats we could do with our phone and, you know, all of that. [W]e use Ring. You gotta love Ring. You don't really want to talk to strangers at the door, so we use a lot of different tools.*

Christine's remark highlights the extent to which she has integrated technology into her life, including the use of automation and smart home devices. Many older adults emphasized how the use of technology decreased the burden of once time-consuming tasks which was a reward for integrating new digital tools. Technologies were so integral to some older adults' lives that 1 participant reflected, "Sometimes I think she's [Alexa] my best friend" (Deborah). Deborah was 1 of multiple older adults using generative AI systems like Alexa for interpersonal connection. This connection fostered regular connection such that a friendship connection grew. Adapters of technology integrated technology and AI into their daily routines with such enthusiasm that the benefits of such adaptation were clear to older adults.

Wary

Some older adults were hesitant to use technology or were unsure of the advantages that the integration of digital devices would bring to their lives. Most older adults with this orientation were using new(er) technologies but were less confident about their usage of digital devices. For instance, Raymond compared his technology knowledge with his wife's, stating:

> *I think she's [wife] less afraid of making a mistake. She's just not as afraid as me. I'm always a little wary of it, I go by the rule that if it's not broken, don't fix it, you know. But that doesn't get you any further ahead. But it keeps you at that level rather than falling back, but I don't move ahead.*

Gregory emphasized the caution with which approaches new technologies. Gregory concluded that he felt more comfortable continuing to use technology in the same way as before instead of learning or incorporating new interfaces. This was a sentiment shared by other older adults like Clarence, who shared,

> *I tend to view it [artificial intelligence] as a negative because you know, if you click on a couple articles, then all of a sudden you're being fed all these articles and it can support a view that maybe you were just interested in, but then all of a sudden it's giving you, you know stuff that supports maybe someone else's point of view, it wasn't really yours. But the more you're inundated with this information the more you tend to believe it.*

Clarence pointed out how AI could facilitate access to information consistent with his worldview yet also overwhelm him with content inconsistent with his perspective. For Clarence, AI was central to targeted algorithm manipulation that impacted what he saw online. This belief left Clarence, and many others, feeling skeptical about AI and the application of the tool in his life. When introduced to other benefits of using AI, Clarence pondered,

> *Oh, like reminders for medications and workout schedules and meal suggestions. I didn't realize that*

XSL•FO
RenderX

> *that stuff was out there, but I pretty much eat salad and chicken and that's, you know, fruits and vegetables, healthy diet.*

This participant was among a group of older adults who were interested in learning about previously unknown aspects of technology and AI that might support their daily routines, while also acknowledging some concerns about adopting new technologies. These older adults were hesitant to adopt new technologies but were not opposed to learning about new features or watching important others use digital devices.

Resistant

The final orientation toward technology and AI was older adults who were resistant to the adoption and integration of digital devices in their lives. Participants with this orientation reported that they avoided learning new technologies and often wished that they could return to a time in their life before these advancements. Speaking on this, Donna articulated:

> *"I basically don't like it...I also think it takes away from personal interaction, which is harmful to humanity...I don't like it. I kind of rather go back to, you know, people farming and taking care of themselves. I'm a homesteader, so I really don't like technology very much. I know we use it every day and it is helpful in that way. But I basically don't like it...I kind of basically ignore it and don't use it if I don't have to. If I have to, then I have to. But if I don't need to, then I don't.*

Donna articulated a sense of nostalgia and skepticism toward technology, positioning these advancements as contrary to traditional ways of life. Ultimately, she expressed an aversion to technology and AI as she concluded that she would only use these devices when necessary. This orientation was shared by other older adults who were unwilling or unable to incorporate technology and AI. This sentiment was central to 1 participant's perspective because this resistant orientation had led to him losing his job. Raymond recounted:

> *And when I turned 62, they said, we really can't keep you. You can get Social Security if you want, or you can get a job, but we can't keep you. And part of the criticism I heard about me was I wasn't technologically savvy enough and I didn't have any interest.*

Raymond reflected on the termination of his position related to his unwillingness to integrate new technologies. Raymond recounted that he remained focused on completing his job with the tools he learned when training as an accountant and was indifferent to studying new adaptations of software that he was able to apply to his work. Resisters encompassed older adults who were satisfied with their current knowledge and who were unwilling to consider future integrations regardless of the possibilities that technology and AI offered to their lives.

Discussion

Principal Results and Comparison With Previous Work

The first research question aimed at understanding the current use of technology and AI by older adults. All older adults were using some form of technology, and most were using a combination of smartphones, smartwatches, and computers in their daily routines. Older adults specified that their current use of technology aided them in social interaction, entertainment, and health management, and assisted them in productivity tasks. This is consistent with findings from past research, which has documented similar uses of technology among older adults. For example, the use of digital devices for maintaining social connections, such as through social media and video calls [33]. Similarly, the use of technology for health management, including fitness trackers and health apps has been widely reported [34].

The second research question was focused on what technology older adults would be interested in learning and integrating into their lives. These results were crowdsourced from older adults to ensure that any future AI technology tools could incorporate aspects of technology that were interesting and helpful. Participants articulated that helpful technology would provide schedule reminders, help with laborious chores, and reduce their caregiving load by monitoring loved ones. These findings are consistent with previous research, which has highlighted older adults' interest in practical, supportive technologies. For example, previous studies have shown that older adults value technologies that enhance their independence, such as devices that provide medication reminders or manage daily schedules [35,36]. Interestingly, although we aimed to examine the use of technology to benefit older adults' health and well-being, many participants expressed interest in technology to monitor others' health and lighten the load of informal caregiving. This reflects both the diversity of health and independence in older age and the need to develop technologies that can be applied by older adults both for their own well-being and to monitor the health of others [37,38]. To better meet the needs of older adults, developers should prioritize implementing user-centered designs to tailor to older adults' preferences [39]. For instance, participatory design approaches involve older adults early in the design process as "co-designers" [40]. This approach ensures their insights and feedback are incorporated, leading to technological solutions that accurately reflect their preferences and needs.

Related to the third research question, our results revealed that the notable barriers to technology and AI use were related to older adults' lack of commitment to learning new technology, privacy concerns, and fear of becoming too dependent on technology. Older adults have expressed hesitancy to learn new technology, mentioning that the time and energy required do not seem worthwhile. Such hesitancy is often linked to their perception of high learning effort expectancy [41]. This suggests that some older adults view mastering a new system as overly demanding. Thus, to address the perception that technology is too demanding for older adults, it is essential to enhance the

XSL•FO
RenderX

perceived ease of use. Providing clear, step-by-step guides on how to use these technologies can play a crucial role in making them more accessible and less intimidating for this demographic. Another barrier found in our study, privacy concerns, has been brought up in previous studies as a significant barrier to technology adoption [8,41,42]. Especially for voice-based systems, older adults reported that they are not comfortable with devices that may have access to and store their personal conversations [8,38,41]. Finally, some of our participants expressed fear that reliance on technology may replace their performance of daily tasks, leading to cognitive decline. Similar concerns were reported in previous research where older adults reported skepticism about using technology potentially deterring their memory [43]. For developers, it is crucial to recognize the cognitive and physical changes that come with aging to design inclusive and user-friendly systems [44]. Previous work suggests that aging-driven designs may effectively address challenges such as isolation and physical changes for older adults [45]. This approach focuses on tailoring technological solutions to the specific problems faced by older adult users.

Finally, our findings of the fourth research question reveal that not all older adults are resistant to technology and AI. Rather, the spectrum of technology adoption among older adults can be broad, encompassing adapters, weary, and resistant individuals. Previous studies have found that these variations could be influenced by individual factors such as age, education, and other sociodemographic factors. For instance, familiarity with technology was not a significant barrier for relatively young older adults (aged 65-75 years) who are already active users of smartphones, whereas older adults aged >75 years were more likely to face barriers toward adopting technology [8]. In addition, evidence shows that higher education and living with a spouse or partner were positively associated with increased use of information communication technology [46]. Therefore, to encourage greater adoption of new technologies, such as AI-based communication systems, it is crucial to consider individual factors and tailor solutions to meet the diverse circumstances of older adult users.

Limitations

Limitations for all studies should be considered. First, this study did not distinguish between age groups of older adults including young-old (65-74 years), old-old (75-84 years), and oldest-old (over 85 years old) [38]. The age of older adults likely impacts their orientation and willingness to accept new technologies. For instance, some participants in the sample worked in offices before the widespread integration of computers. Learning technology as a requirement of one's employment might impact the way that older adults view technological advancement compared with others who only voluntarily interacted with technologies throughout their adulthood. Future researchers might differentiate between age groups to examine how the orientations outlined are applicable across sections of older adults. Another limitation of this study pertains to self-selection bias. Recruiting older adults through a community research pool may attract participants who are more technologically adept and familiar with technology, potentially underrepresenting those who face technological barriers or hold negative attitudes toward technology [47]. To address this limitation, future research should adopt recruitment strategies, such as using multiple recruitment channels and partnering with organizations (eg, local senior centers), to be more inclusive toward individuals with limited access to or resistance to technology [36,48].

In addition, older adults routinely experience stereotypes and stigma related to agism [46,49]. This ageism might limit an individual's willingness to adopt a learning mindset with new technology given the vulnerability of labeling a knowledge deficient. Examining how stereotypes serve as barriers for older adults integrating technology and seeking support would inform researchers on how to decrease the spread of this harmful communication and combat the negative implications of the messages through public health campaigns targeted at empowering technology use for members of this community.

Conclusions

Despite the increasing interest and potential benefits of using AI technology for older adult populations, significant challenges persist in the current technology landscape. The current study aimed to deepen our understanding of how older adults perceive technology and AI while also identifying challenges and opportunities for improvement. Our findings suggest a critical need for developing more personalized and aging-friendly systems that can be successfully integrated into older adults' daily lives. Moving forward, it is essential to keep the dialogue open between developers, older adults, their families, and their care team to ensure that the design of AI technologies is inclusive and supportive to older adults.

Acknowledgments

This work was supported by the Michigan State University College of Communication Arts and Sciences, Engineering, and Nursing Trifecta Initiative.

Data Availability

The data generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. WHO methods and data sources for life tables 1990-2019. World Health Organization. 2020. URL: https://cdn.who.int/media/docs/default-source/gho-documents/global-health-estimates/ghe2019_life-table-methods.pdf?sfvrsn=c433c229_5 [accessed 2024-06-29]
2. Mental health of older adults. World Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2023-11-15]
3. Teo RH, Cheng WH, Cheng LJ, Lau Y, Lau ST. Global prevalence of social isolation among community-dwelling older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr 2023;107:104904. [doi: 10.1016/j.archger.2022.104904] [Medline: 36563614]
4. Ruggiano N, Brown EL, Roberts L, Framil Suarez CV, Luo Y, Hao Z, et al. Chatbots to support people with dementia and their caregivers: systematic review of functions and quality. J Med Internet Res 2021;23(6):e25006 [FREE Full text] [doi: 10.2196/25006] [Medline: 34081019]
5. Sapci AH, Sapci HA. Innovative assisted living tools, remote monitoring technologies, artificial intelligence-driven solutions, and robotic systems for aging societies: systematic review. JMIR Aging 2019;2(2):e15429 [FREE Full text] [doi: 10.2196/15429] [Medline: 31782740]
6. Rodríguez-Martínez A, Amezcua-Aguilar T, Cortés-Moreno J, Jiménez-Delgado JJ. Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health. Healthcare (Basel) 2023;12(1):62 [FREE Full text] [doi: 10.3390/healthcare12010062] [Medline: 38200967]
7. Chou YH, Lin C, Lee SH, Chang Chien YW, Cheng LC. Potential mobile health applications for improving the mental health of the elderly: a systematic review. Clin Interv Aging 2023;18:1523-1534 [FREE Full text] [doi: 10.2147/CIA.S410396] [Medline: 37727447]
8. Gudala M, Ross MET, Mogalla S, Lyons M, Ramaswamy P, Roberts K. Benefits of, barriers to, and needs for an artificial intelligence-powered medication information voice chatbot for older adults: interview study with geriatrics experts. JMIR Aging 2022;5(2):e32169 [FREE Full text] [doi: 10.2196/32169] [Medline: 35482367]
9. Hülür G, Macdonald B. Rethinking social relationships in old age: digitalization and the social lives of older adults. Am Psychol 2020;75(4):554-566. [doi: 10.1037/amp0000604] [Medline: 32378949]
10. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
11. Latikka R, Rubio-Hernández R, Lohan ES, Rantala J, Nieto Fernández F, Laitinen A, et al. Older adults' loneliness, social isolation, and physical information and communication technology in the era of ambient assisted living: a systematic literature review. J Med Internet Res 2021;23(12):e28022 [FREE Full text] [doi: 10.2196/28022] [Medline: 34967760]
12. Marziali RA, Franceschetti C, Dinculescu A, Nistorescu A, Kristály DM, Mo oi AA, et al. Reducing loneliness and social isolation of older adults through voice assistants: literature review and bibliometric analysis. J Med Internet Res 2024;26:e50534 [FREE Full text] [doi: 10.2196/50534] [Medline: 38498039]
13. GPT-3.5-Turbo-0125 Model. OpenAI. URL: https://platform.openai.com/docs/models/gpt-3-5 [accessed 2023-03-01]
14. Kocaballi AB, Sezgin E, Clark L, Carroll JM, Huang Y, Huh-Yoo J, et al. Design and evaluation challenges of conversational agents in health care and well-being: selective review study. J Med Internet Res 2022;24(11):e38525 [FREE Full text] [doi: 10.2196/38525] [Medline: 36378515]
15. SeniorTalk - Chatbot for Elderly People. URL: https://www.senior-talk.com/ [accessed 2024-09-15]
16. Broadbent E, Loveys K, Ilan G, Chen G, Chilukuri MM, Boardman SG, et al. ElliQ, an AI-driven social robot to alleviate loneliness: progress and lessons learned. JAR Life 2024;13:22-28 [FREE Full text] [doi: 10.14283/jarlife.2024.2] [Medline: 38449726]
17. Emmer De Albuquerque Green C. Defining responsible use of AI chatbots in social care for older adults. Nat Aging 2024;4(6):745-747 [FREE Full text] [doi: 10.1038/s43587-024-00636-w] [Medline: 38740938]
18. Elavsky S, Knapova L, Janiš K, Cimler R, Kuhnova J, Cernicky T. Multiple perspectives on the adoption of SMART technologies for improving care of older people: mixed methods study. J Med Internet Res 2024;26:e45492 [FREE Full text] [doi: 10.2196/45492] [Medline: 38324345]
19. Peek STM, Wouters EJ, Luijkx KG, Vrijhoef HJ. What it takes to successfully implement technology for aging in place: focus groups with stakeholders. J Med Internet Res 2016;18(5):e98 [FREE Full text] [doi: 10.2196/jmir.5253] [Medline: 27143097]
20. National Academies of Sciences. Engineering, and Medicine. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington, DC: Academies Press; 2020.
21. Chawla K, Kunonga TP, Stow D, Barker R, Craig D, Hanratty B. Prevalence of loneliness amongst older people in high-income countries: a systematic review and meta-analysis. PLoS One 2021;16(7):e0255088 [FREE Full text] [doi: 10.1371/journal.pone.0255088] [Medline: 34310643]
22. Park C, Majeed A, Gill H, Tamura J, Ho RC, Mansur RB, et al. The effect of loneliness on distinct health outcomes: a comprehensive review and meta-analysis. Psychiatry Res 2020;294:113514. [doi: 10.1016/j.psychres.2020.113514] [Medline: 33130511]

XSL•FO
RenderX

23. Ong AD, Uchino BN, Wethington E. Loneliness and health in older adults: a mini-review and synthesis. Gerontology 2016;62(4):443-449 [FREE Full text] [doi: 10.1159/000441651] [Medline: 26539997]
24. Dahlberg L, McKee KJ, Frank A, Naseer M. A systematic review of longitudinal risk factors for loneliness in older adults. Aging Ment Health 2022;26(2):225-249. [doi: 10.1080/13607863.2021.1876638] [Medline: 33563024]
25. Jensen V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101.
26. Kidder LH. Research Methods in Social Relations. New York: Holt, Rinehart, and Winston; 1981.
27. Lincoln YS, Guba EG, Pilotta JJ. Naturalistic Inquiry. Thousand Oaks: Sage; 1985.
28. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Washington DC: SAGE Publications, Inc; 2008.
29. Centers for Disease Control and Prevention. A data users' guide to the disability questions included in the behavioral risk factor surveillance system. 2020. URL: https://www.cdc.gov/brfss/data_documentation/pdf/BRFSS_Data_Users_Guide_on_Disability_Questions_2018-508.pdf [accessed 2024-11-19]
30. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
31. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
32. Yoon J, Lee M, Ahn JS, Oh D, Shin S, Chang YJ, et al. Development and validation of digital health technology literacy assessment questionnaire. J Med Syst 2022;46(2):13 [FREE Full text] [doi: 10.1007/s10916-022-01800-8] [Medline: 35072816]
33. Kaur A, Chen W. Exploring AI literacy among older adults. Stud Health Technol Inform 2023;306:9-16. [doi: 10.3233/SHTI230589] [Medline: 37638893]
34. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
35. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
36. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
37. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
38. Dodge HH, Du Y, Saxton JA, Ganguli M. Cognitive domains and trajectories of functional independence in nondemented elderly persons. J Gerontol A Biol Sci Med Sci 2006;61(12):1330-1337. [doi: 10.1093/gerona/61.12.1330] [Medline: 17234830]
39. Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. Transl Behav Med 2019;9(6):1057-1064. [doi: 10.1093/tbm/iby119] [Medline: 30535343]
40. Lindsay S, Jackson D, Schofield G, Olivier P. Engaging older people using participatory design. 2012 Presented at: Proceedings of the SIGCHI conference on human factors in computing systems; 2012 May 05; USA p. 1199-1208. [doi: 10.1145/2207676.2208570]
41. Min SH, Topaz M, Lee C, Schnall R. Understanding changes in mental health symptoms from young-old to old-old adults by sex using multiple-group latent transition analysis. Geroscience 2023;45(3):1791-1801 [FREE Full text] [doi: 10.1007/s11357-023-00729-1] [Medline: 36626018]
42. Barnard Y, Bradley MD, Hodgson F, Lloyd AD. Learning to use new technologies by older adults: perceived difficulties, experimentation behaviour and usability. Computers in Human Behavior 2013;29(4):1715-1724. [doi: 10.1016/j.chb.2013.02.006]
43. Bonilla K, Martin-Hammond A. Older adults' perceptions of intelligent voice assistant privacy, transparency, and online privacy guidelines. URL: https://scholarworks.indianapolis.iu.edu/server/api/core/bitstreams/60e837c4-8292-4987-9559-3e7bf3dc892e/content [accessed 2020-08-11]
44. Vines J, Pritchard G, Wright P, Olivier P, Brittain K. An age-old problem examining the discourses of ageing in HCI and strategies for future research. ACM Trans Comput-Hum Interact 2015;22(1):1-27. [doi: 10.1145/2696867]
45. Carroll JM, Convertino G, Farooq U, Rosson MB. The firekeepers: aging considered as a resource. Univ Access Inf Soc 2011;11(1):7-15. [doi: 10.1007/s10209-011-0229-9]
46. Wang S, Bolling K, Mao W, Reichstadt J, Jeste D, Kim HC, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) 2019;7(2):60 [FREE Full text] [doi: 10.3390/healthcare7020060] [Medline: 30974780]
47. Poli A, Kelfve S, Berg K, Motel-Klingebiel A. Old-age diversity is underrepresented in digital health research: findings from the evaluation of a mobile phone system for post-operative progress monitoring in Sweden. Ageing and Society 2021;43(10):2264-2286. [doi: 10.1017/s0144686x21001641]
48. Peek ST, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237. [doi: 10.1159/000430949] [Medline: 26044243]

XSL•FO
RenderX

49. Vroman KG, Arthanat S, Lysack C. 'Who over 65 is online?' Older adults' dispositions toward information communication technology. Computers in Human Behavior 2015;43:156-166. [doi: 10.1016/j.chb.2014.10.018]

Abbreviations

AI: artificial intelligence
BRFSS: Behavioral Risk Factor Surveillance System

Edited by A Mavragani; submitted 25.08.24; peer-reviewed by B Davis, P Nohl-Deryk, K Elechi; comments to author 11.09.24; revised version received 18.09.24; accepted 29.10.24; published 13.03.25.

Please cite as:
Wolfe BH, Oh YJ, Choung H, Cui X, Weinzapfel J, Cooper RA, Lee HN, Lehto R
Caregiving Artificial Intelligence Chatbot for Older Adults and Their Preferences, Well-Being, and Social Connectivity: Mixed-Method Study
J Med Internet Res 2025;27:e65776
URL: https://www.jmir.org/2025/1/e65776
doi:10.2196/65776
PMID:

©Brooke H Wolfe, Yoo Jung Oh, Hyesun Choung, Xiaoran Cui, Joshua Weinzapfel, R Amanda Cooper, Hae-Na Lee, Rebecca Lehto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning–Based Explainable Automated Nonlinear Computation Scoring System for Health Score and an Application for Prediction of Perioperative Stroke: Retrospective Study

Mi-Young Oh[1], MD, PhD; Hee-Soo Kim[2], MS; Young Mi Jung[3], MD; Hyung-Chul Lee[4,5], MD, PhD; Seung-Bo Lee[2], PhD; Seung Mi Lee[6,7,8,9], MD, PhD

[1]Department of Neurology, Sejong General Hospital, Sejong General Hospital, Bucheon-si, Republic of Korea
[2]Department of Medical Informatics, School of Medicine, Keimyung University, Daegu, Republic of Korea
[3]Department of Obstetrics and Gynecology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[4]Department of Anesthesiology and Pain Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
[5]Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[6]Department of Obstetrics and Gynecology, College of Medicine, Seoul National University, Seoul, Republic of Korea
[7]Department of Obstetrics and Gynecology, Seoul National University Hospital, Seoul, Republic of Korea
[8]Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, Republic of Korea
[9]Institute of Reproductive Medicine and Population & Medical Big Data Research Center, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Seung Mi Lee, MD, PhD
Department of Obstetrics and Gynecology
College of Medicine
Seoul National University
101 Daehak - ro, Jongno - gu
Seoul, 03080
Republic of Korea
Phone: 82 2 2072 4857
Fax: 82 2 762 3599
Email: lbsm@snu.ac.kr

Abstract

Background: Machine learning (ML) has the potential to enhance performance by capturing nonlinear interactions. However, ML-based models have some limitations in terms of interpretability.

Objective: This study aimed to develop and validate a more comprehensible and efficient ML-based scoring system using SHapley Additive exPlanations (SHAP) values.

Methods: We developed and validated the Explainable Automated nonlinear Computation scoring system for Health (EACH) framework score. We developed a CatBoost-based prediction model, identified key features, and automatically detected the top 5 steepest slope change points based on SHAP plots. Subsequently, we developed a scoring system (EACH) and normalized the score. Finally, the EACH score was used to predict perioperative stroke. We developed the EACH score using data from the Seoul National University Hospital cohort and validated it using data from the Boramae Medical Center, which was geographically and temporally different from the development set.

Results: When applied for perioperative stroke prediction among 38,737 patients undergoing noncardiac surgery, the EACH score achieved an area under the curve (AUC) of 0.829 (95% CI 0.753-0.892). In the external validation, the EACH score demonstrated superior predictive performance with an AUC of 0.784 (95% CI 0.694-0.871) compared with a traditional score (AUC=0.528, 95% CI 0.457-0.619) and another ML-based scoring generator (AUC=0.564, 95% CI 0.516-0.612).

Conclusions: The EACH score is a more precise, explainable ML-based risk tool, proven effective in real-world data. The EACH score outperformed traditional scoring system and other prediction models based on different ML techniques in predicting perioperative stroke.

(J Med Internet Res 2025;27:e58021) doi:10.2196/58021

XSL•FO
RenderX

KEYWORDS

machine learning; explainability; score; computation scoring system; Nonlinear computation; application; perioperative stroke; perioperative; stroke; efficiency; ML-based models; patient; noncardiac surgery; noncardiac; surgery; effectiveness; risk tool; risk; tool; real-world data

Introduction

The risk scoring systems have been proposed to prognosticate critical medical conditions, aiming to identify high-risk patients who are likely to experience adverse outcomes [1-6]. Traditionally, risk-scoring systems have been developed using conventional statistical approaches, based on the assumption of linearity between variables and outcomes [7,8]. This method has provided a comprehensive understanding of patient risk profiles, but it may not fully capture complex, nonlinear interactions, potentially leading to less accurate risk assessments. Furthermore, the clinical variables incorporated into these systems were typically ascertained through univariate analysis, enriched by insights from expert opinions, or selected from a range of risk factors established in previous literature [8]. Consequently, these systems faced limitations in rapidly integrating state-of-the-art medical knowledge alongside medical advancements [9-11]. With the rapid expansion and increasing diversity of medical data, these limitations have become more pronounced [10].

To address these limitations, machine learning (ML) techniques have emerged as promising avenues for creating new and diverse risk models by leveraging extensive electronic medical records [12,13]. Despite their exceptional predictive performance, the lack of interpretability has limited their adoption in real-world medical practice [14]. Recent efforts in ML have addressed the "black box" issue of existing models by presenting data in a more understandable fashion [15]. For instance, an ML-based automated scoring model was developed by integrating the Random Forest (RF) algorithm and logistic regression methods [16]. This model was designed to be easy to understand and apply to various clinical situations, such as in-hospital mortality and out-of-hospital cardiac arrest [16,17]. However, these approaches still rely on scoring methods rooted in traditional statistics, using variables selected by the ML models [16]. Furthermore, the generalizability and scalability of ML models are hampered by limited external validation. Therefore, there is a compelling need to develop and validate a fully automated ML-based scoring system that can address nonlinearity assumptions.

In response to these challenges, this research aimed to develop and validate an Explainable Automated nonlinear Computation scoring system for Health (EACH) score. This system addresses nonlinearity assumptions and enhances explainability. In addition, we applied the EACH score to predict perioperative stroke to assess its performance in real-world clinical practice and examined its performance compared with traditional scores and other ML-based scoring systems.

Perioperative stroke significantly impacts postoperative morbidity and mortality. Patients who experience perioperative stroke are more likely to develop postoperative delirium, have prolonged hospitalization, increased in-hospital mortality, and be discharged to nursing facilities [18-20]. However, the detection of perioperative stroke is challenging due to the effects of anesthetic and analgesic medications, which contribute to a low rate of hyperacute thrombolytic treatment [21]. Therefore, to reduce the risk of perioperative stroke, it is essential to use more effective tools for identifying high-risk patients and to implement close monitoring of these patients. To apply the EACH score in predicting perioperative stroke, we aimed to develop and validate a more accurate and explainable prediction model that can be easily adapted to real-world practice.

Methods

Overview

We developed and validated the EACH, designed to automate the development of clinical scoring models for predefined health outcomes. The development and application of EACH were segmented into several steps, as illustrated in Figure 1. We tried to follow the guideline for developing reporting machine learning prediction models in biomedical research [22].

XSL•FO
RenderX

Figure 1. Visual guide to the sequential steps and their detailed execution. SHAP: SHapley Additive exPlanations.

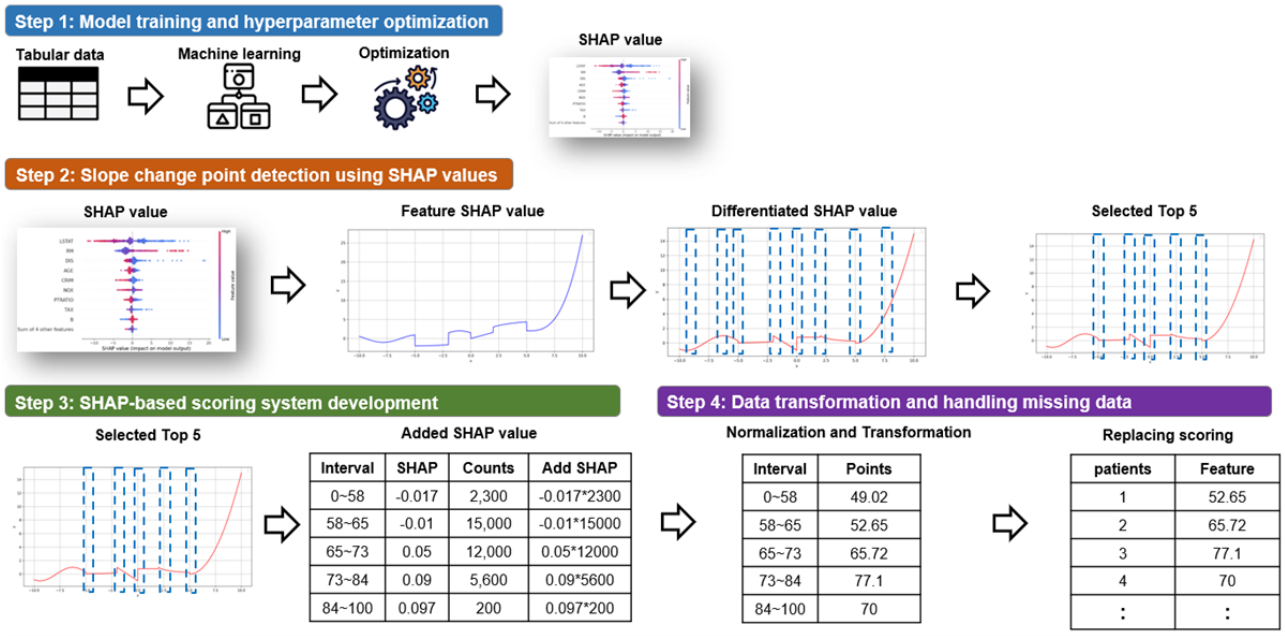

Step 1: Model Training and Hyperparameter Optimization

When implementing the EACH score, the initial step involved training the CatBoost model with hyperparameter optimization for each set of clinical data, followed by determination of important features using SHapley Additive exPlanations (SHAP) values [23].

Step 2: The Detection of Slope Change Point in Plots of SHAP Values and Selection of the Top 5 Steepest Slope Change Points

In this step, we harnessed the power of automation to identify the critical features that influence the accuracy of the prediction model. The process began with the generation of SHAP value plots for each feature, which visually represented the impact of these features on the model predictions. The key advantage of our approach is the automated detection of the slope change points within these plots. The slope change points are significant shifts in the importance of each variable, indicating where the influence of the variable undergoes a notable change.

The algorithm identified the 5 most pronounced slope change points for each feature. These critical points indicate where the model's sensitivity to feature values is dramatically altered, thus playing a pivotal role in subsequent feature scoring. Figure 2 illustrates this method by contrasting linear and nonlinear cases, where the latter demonstrates the effectiveness of the method in recognizing complex nonlinear relationships often present in clinical data.

XSL•FO
RenderX

Figure 2. Differential slope change comparison: linear versus nonlinear cases.

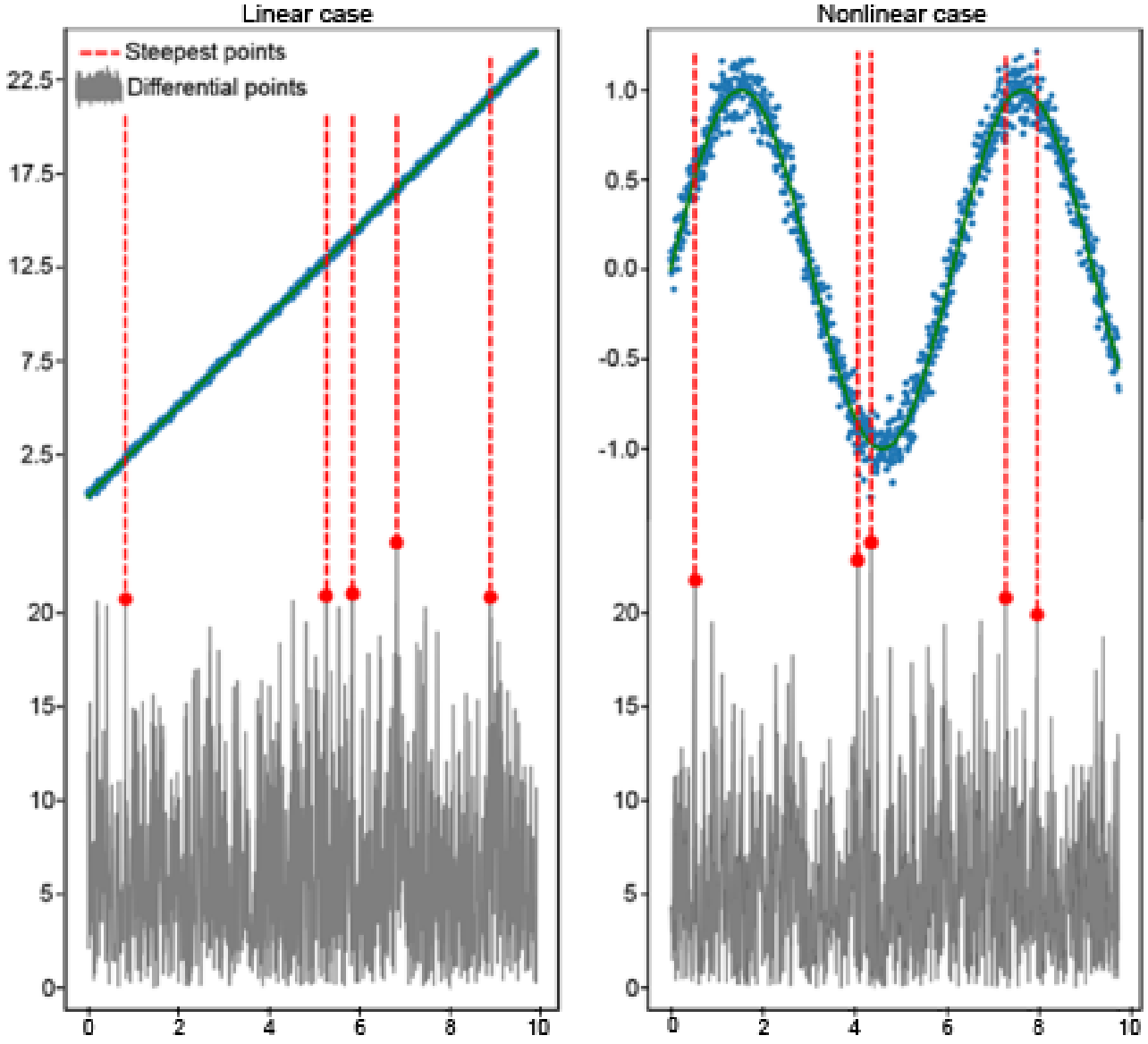

By isolating these key intervals, we delineated the feature value ranges that were most influential on the model's predictions, thus allowing for a more accurate SHAP-based scoring system.

Step 3: SHAP-Based Scoring System Development

The third step shows how to generate scores using the SHAP values for each variable. The score was calculated by adding the SHAP values of the data points obtained from each section to accurately quantify the impact of individual intervals for each feature. This was for the model's sensitive interpretation of the different ranges within the variable and was reflected in the final scoring system.

Step 4: Data Transformation and Handling Missing Data

Normalization of Scores Across Features

In this step, we focused on standardizing the scoring system across all features. This normalization process ensured that the scores from different features were comparable and appropriately weighted within the overall scoring model. Normalization is based on the total range of scores across all features, thereby aligning them on a unified scale.

Handling Missing Data With Interval Averaging

If the actual values for a particular interval were unavailable, our model adopted a fallback strategy that uses the average of values in adjacent intervals. If there is no data in the interval immediately next to the missing value, the entire range of the interval was used to preserve the integrity of the model's risk assessment in the face of data sparsity.

Clinical Study Design

In applying the EACH score to assess stroke risk during the perioperative period in noncardiac surgeries, we approached our clinical study design with a focus on real-world applicability. This study was structured using 3 datasets: training, internal validation, and external validation. The training and internal validation datasets were derived from patient records at Seoul National University Hospital (SNUH) from 2016 to 2019. For a broader perspective and to assess the performance of the model in different settings, we included a geotemporal external

XSL•FO
RenderX

validation using data from surgeries performed at the Boramae Medical Center (BMC) between 2020 and 2021.

A key aspect of our data preparation involved addressing missing values in preoperative variables. Recognizing the potential impact of incomplete data on the model's accuracy, we used a methodological imputation approach. For continuous variables, missing values were imputed using their mean, whereas categorical variables were imputed using the mode [24].

Patients were excluded from the study if the surgery lasted less than 20 minutes, if they were younger than 18 years or weighed less than 30 kg or more than 140 kg, if their height was outside the 135 cm to 200 cm range, or if they had a previous history of stroke. The developmental set consisted of 36,502 patients from Seoul National University Hospital. After applying the exclusion criteria, 404 patients from Boramae Medical Center were included in the external validation set (Figure 3). Perioperative stroke was defined as an ischemic brain infarction occurring within 30 days postoperatively and was identified through new ischemic lesions on diffusion-weighted imaging [18,21]. An experienced neurologist [M-YO] confirmed the stroke diagnosis by reviewing the imaging outcomes.

Figure 3. Flowchart of study population.

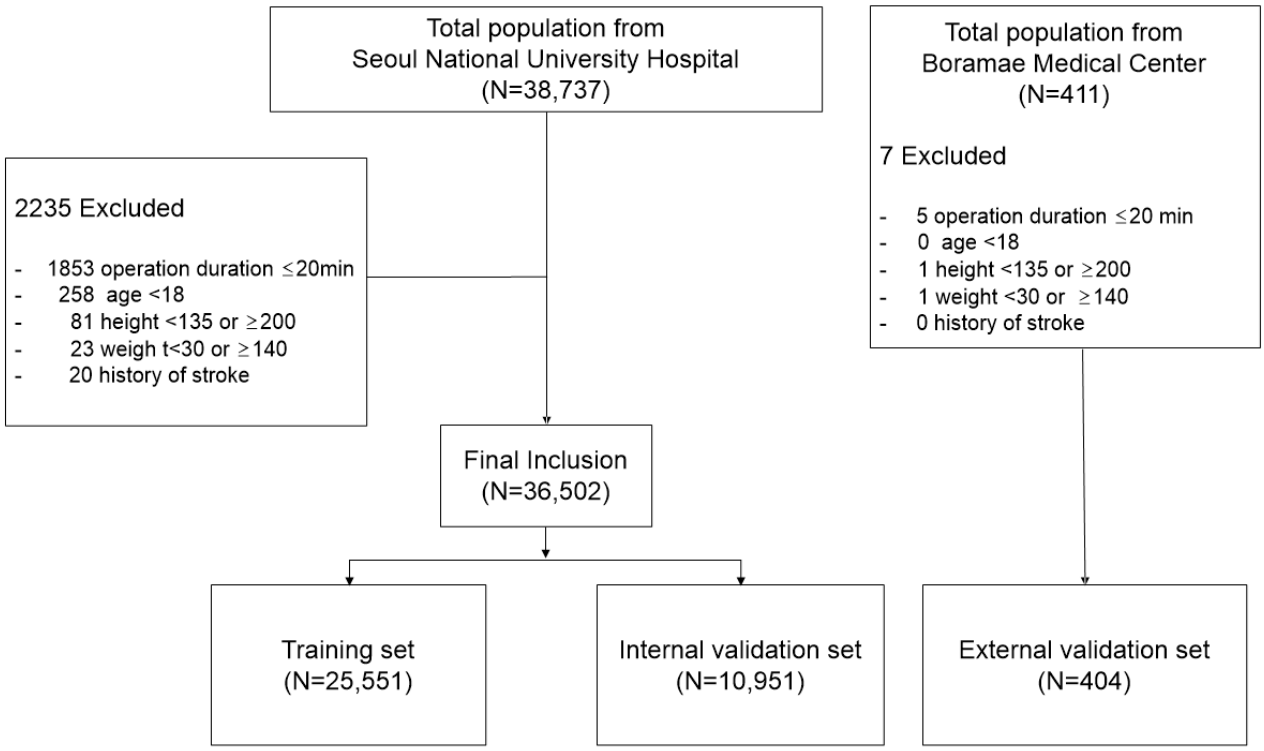

Data Collection

Demographics and Comorbidities

We recorded patients' age, sex, physical metrics (height, weight, and body mass index [BMI]), and a range of preexisting conditions, including hypertension, diabetes (with or without insulin medication), previous cardiovascular events, asthma, chronic obstructive pulmonary disease, liver and kidney diseases, and tuberculosis.

Surgical Information

Surgical risk was classified according to the American Society of Anesthesiologists (ASA) standards. Information on whether the surgery was emergent and the type was collected.

Preoperative Laboratory Findings

A comprehensive set of laboratory results were collected, including hemoglobin levels, renal function indicators (blood urea nitrogen, creatinine, and estimated glomerular filtration rate), nutritional markers (albumin), electrolytes (sodium and potassium), glucose levels, liver enzymes (aspartate aminotransferase and alanine aminotransferase), platelet count, and coagulation status assessed by partial thromboplastin time.

Revised Cardiac Risk Index

This traditional scoring system is used to assess the risk of major adverse cardiac events including stroke in noncardiac surgery. The Revised Cardiac Risk Index (RCRI) is composed of the factors mentioned above, such as type of surgery, history of ischemic heart disease, congestive heart failure, cerebrovascular disease, preoperative treatment with insulin, and a preoperative creatinine level greater than 2 mg/dL [13]. The RCRI was used as the comparative score.

Prediction Model Based on Other ML Techniques

We developed prediction model based on several ML algorithms, including Support Vector Machine (SVM) [25], Decision Tree Classifier (DTC) [26], RF [27], and CatBoost [28] to compare the performance with the EACH score. The developmental dataset was systematically divided into a training set comprising 70% of the data for model development, and the

XSL•FO
RenderX

remaining 30% formed the test set used to assess model performance. Hyperparameter optimization for each model was performed using grid-search cross-validation to identify the optimal settings that maximize the discriminative power of the model [29].

External Validation

External validation was conducted using the BMC cohort to assess the prediction performance of the EACH score, which was developed to predict perioperative stroke based on the SNUH cohort. The data used for validation were entirely separate from the data used during the development of the EACH score. BMC cohort was geographically, and temporally different cohort from the development set, SNUH cohort.

Statistical Analysis

Continuous variables were analyzed using the Student *t* test and the Mann-Whitney *U* test to determine the significance of differences between the 2 groups.

Categorical variables in both datasets were evaluated using the chi-square test to investigate the presence of significant associations or discrepancies between the categories. All tests were 2-sided, and statistical significance was set at $P<.05$. ML modeling was performed in python 3.8 using the Scikit-Learn package [30]. The area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to evaluate the performance of the prediction model [31].

Ethical Considerations

This study was approved by the institutional review board (IRB) of Seoul National University Hospital on April 7, 2020 (IRB Number, 2003-067-1108), and Boramae Medical Center on January 7, 2021 (IRB Number, 30-2021-4). The IRB determined that participant consent was waived in this retrospective study. We tried to follow the tripod guideline [32].

Results

Baseline Characteristics of the Study Cohorts

Our retrospective observational cohort study used data from SNUH comprising the training set (25,551 cases) and the internal validation set (10,951 cases) and the data of 404 patients from BMC to externally validate the prediction model in a geographically and temporally different population. The baseline characteristics of patients in the SNUH and BMC cohorts are summarized in Table 1. Patients in the SNUH cohort were younger than those in the BMC cohort. The prevalence of diabetes was slightly higher in the BMC cohort. Hemoglobin, creatinine, and albumin levels were higher in patients in the SNUH cohort, whereas glucose levels were slightly higher in patients in the BMC cohort.

XSL•FO
RenderX

Table 1. Baseline characteristics.

Characteristics	SNUH[a] cohort	BMC[b] cohort	*P* value
Number of strokes, n (%)	139 (0.38)	10 (2.48)	<.001
Age (years), mean (SD)	56.91 (15.23)	62.44 (14.57)	<.001
Sex, male, n (%)	16,265 (55.44)	204 (50.50)	.48
Height (cm), mean (SD)	161.97 (8.7)	160.7 (8.61)	.01
Weight (kg), mean (SD)	63.46 (12.01)	63.36 (11.91)	.10
BMI (kg/m^2), mean (SD)	24.12 (3.73)	24.49 (3.93)	.04
Preoperative ASA[c], n (%)			.71
1	10,778 (29.53)	25 (6.19)	—[d]
2	21,283 (58.31)	298 (73.76)	—
3	4172 (11.43)	74 (18.32)	—
4	252 (0.69)	7 (1.73)	—
5	17 (0.05)	0 (0)	—
Emergency surgery, n (%)	2132 (5.84)	19 (4.7)	.53
Preoperative hypertension, n (%)	11,395 (31.22)	168 (41.58)	.06
Preoperative diabetes, n (%)	5543 (15.19)	92 (22.77)	.90
Preoperative cardiovascular accident, n (%)	729 (2.00)	26 (6.44)	>.99
Preoperative asthma, n (%)	230 (0.63)	8 (1.98)	>.99
Preoperative COPD[e], n (%)	234 (0.64)	7 (1.73)	.15
Preoperative liver disease, n (%)	1437 (3.94)	42 (10.40)	.69
Preoperative kidney disease, n (%)	1179 (3.23)	39 (9.65)	.97
Preoperative tuberculosis, n (%)	366 (1.00)	26 (6.44)	.91
Surgery type[f], n (%)	19,787 (54.21)	212 (52.48)	.51
Hemoglobin (g/dL), mean (SD)	12.90 (1.86)	12.08 (1.93)	<.001
Platelet (×10^3/μL), mean (SD)	245.18 (79.48)	247.54 (87.81)	.86
Blood urea nitrogen (mg/dL), mean (SD)	15.82 (9.18)	16.88 (11.59)	.34
Creatinine (mg/dL), mean (SD)	1.01 (1.34)	0.43 (1.55)	<.001
Albumin (g/dL), mean (SD)	4.05 (0.47)	3.75 (0.6)	<.001
Sodium (mmol/L), mean (SD)	140.17 (2.66)	138.85 (3.07)	<.001
Potassium (mmol/L), mean (SD)	4.24 (0.41)	3.81 (0.54)	<.001
Glucose (mg/dL), mean (SD)	114.88 (39.92)	121.52 (45.85)	.01
Prothrombin time, n (%)	103.58 (14.86)	11.96 (1.12)	<.001
Partial thromboplastin time (sec), mean (SD)	31.45 (5.56)	27.66 (3.9)	<.001
Aspartate aminotransferase (IU/L), mean (SD)	25.33 (69.93)	25.32 (52.62)	.97
Alanine aminotransferase (IU/L), mean (SD)	25.44 (52.62)	28.68 (32.74)	.01
Estimated glomerular filtration rate (mL/min/1.73 m²), mean (SD)	83.79 (27.25)	83.83 (24.82)	.84

[a]SNUH: Seoul National University Hospital.

[b]BMC: Boramae Medical Center.

[c]ASA: American Society of Anesthesiologists classification.

[d]Not applicable.

[e]COPD: chronic obstructive pulmonary disease.

[f]Surgery types included intrathoracic, intra-abdominal, and supra-inguinal vascular surgery.

XSL•FO
RenderX

Scoring System Based on Clinical Feature Intervals

The entire scoring system is presented in Multimedia Appendix 1. Scores ranging from 0 to 100 were assigned to the clinical features based on specific intervals determined by the slope change points. For example, albumin levels less than 4.1 g/dL were associated with a higher risk of perioperative stroke. Specifically, the highest score of 55.4 was assigned to the 2.4-3.5 g/dL range, indicating a particularly elevated risk for some low albumin levels. Similarly, hemoglobin levels were segmented, with the highest score of 51.5 assigned to the 5.0-11.0 g/dL range. In addition, the 18.6-20.7 kg/m² interval of BMI received a score of 51.8, suggesting a higher risk associated with specific BMI levels. However, age was divided into intervals, with scores increasing incrementally from 49.0 for the 18-49 years age group to 51.8 for those aged 64-97 years old. Creatinine levels did not differ significantly between intervals. These results suggest that continuous variables could impact predictive performance in various nonlinear patterns and that this pattern is well reflected in the EACH scoring system. Categorical variables were scored on the basis of their relative importance in predicting perioperative stroke.

Comparative Analysis of the Performance of EACH Scores Relative to Other ML Models

The EACH score demonstrated superior performance compared with other ML-based models, such as RF and CatBoost, as detailed in Table 2. The EACH score had the highest AUC (0.829) and sensitivity (0.881). The SVM, DTC, RF, and CatBoost models exhibited competitive yet slightly lower AUCs of 0.500, 0.501, 0.802, and 0.822, respectively. In the external validation, the EACH model consistently maintained high performance. It achieved an AUC of 0.784 and attained the highest sensitivity at 0.900 (Table 2).

Table 2. Comparison between the performance of Explainable Automated nonlinear Computation scoring system for Health (EACH) score and that of other machine learning models.

Models	AUC[a]	Accuracy	Sensitivity	Specificity	PPV[b]	NPV[c]
Internal validation						
Support Vector Machine	0.500	0.954	0.888	0.268	0.003	0.999
Decision Tree Classifier	0.501	0.961	0.820	0.566	0.007	0.999
Random Forest	0.802	0.640	0.856	0.639	0.009	0.999
CatBoost	0.822	0.686	0.770	0.686	0.010	0.999
EACH[d] score	0.829	0.655	0.881	0.654	0.010	0.999
External validation						
Support Vector Machine	0.500	0.955	0.700	0.655	0.049	0.989
Decision Tree Classifier	0.503	0.964	0.821	0.621	0.005	0.999
Random Forest	0.744	0.797	0.600	0.802	0.071	0.988
CatBoost	0.782	0.693	0.900	0.688	0.068	0.996
EACH score	0.784	0.688	0.900	0.683	0.067	0.996

[a]AUC: area under the receiver operating characteristic curve.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]EACH score: Explainable Automated nonlinear Computation scoring system for Health.

Comparative Analysis of the Performance of the EACH Explainable Score Relative to Traditional Scores

We compared the effectiveness of the EACH score with that of a traditional scoring system based on classical statistical methods and ML algorithms (Figure 4). The EACH score demonstrated superior performance compared with the RCRI, a scoring system used to predict perioperative stroke based on classical logistic regression analysis (AUC=0.528 vs AUC=0.784).

XSL•FO
RenderX

Figure 4. Comparison of receiver operating characteristic curves.

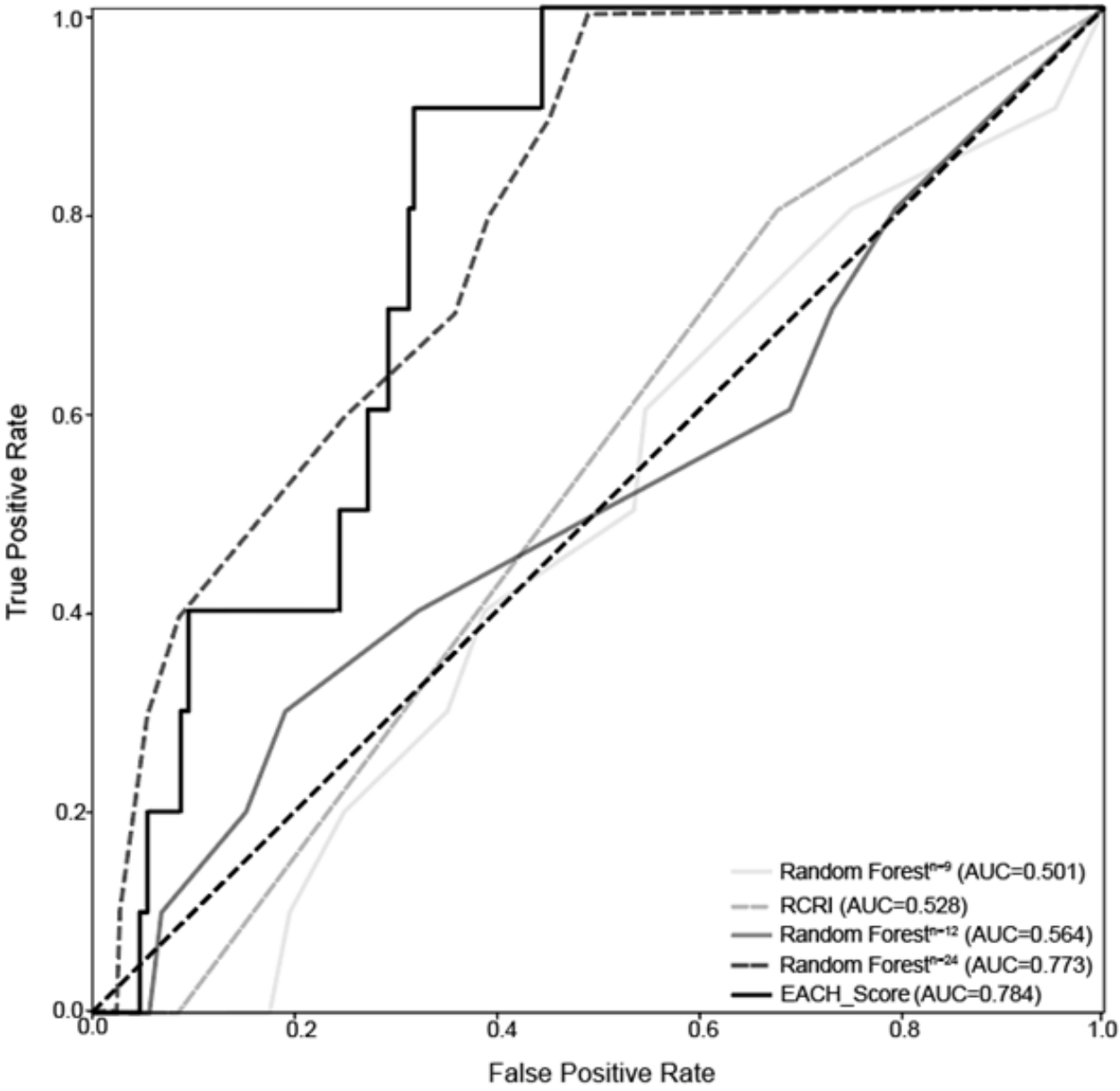

Comparative Analysis of the Performance of the EACH Score Relative to Other ML-Based Scoring Systems

In addition, we compared the performance of the EACH score with that of the ML-based score generation system, AutoScore (eg, $RF^{n=12, n=24}$). The performance of AutoScore increased with the number of features: with 9 features, the AUC was 0.501; with 12 features, the AUC was 0.564; and with 24 features, the AUC was 0.773. However, the EACH score surpassed the performance of AutoScore, demonstrating a superior AUC of 0.784 (Figure 4).

Risk Stratification Using the EACH Score: Low- Versus High-Risk Patients for Perioperative Stroke

The 2 patients representing low and high perioperative stroke risk based on the EACH score are shown in Table 3. The table illustrates how the EACH score model assigned risk scores to individual patient characteristics, leading to a cumulative risk assessment. In assigning scores to patients, the low-risk case, characterized by younger age (42.77 years) and moderate BMI (24.02 kg/m^2), tended to have lower ASA scores, no emergency operation, and no premorbidities. The laboratory results included slightly higher hemoglobin levels (13 g/dL) and higher platelet counts (293 ×10^3/μL), resulting in a lower total score (1379.16) indicative of a lower risk profile. Conversely, the high-risk patient was characterized by older age (71.07 years), lower BMI (22.84 kg/m^2), higher ASA scores, emergency operation, and surgery with a high risk for cardiovascular complications. The patient had lower hemoglobin levels (12 g/dL), lower platelet counts (92 ×10^3/μL), lower albumin levels (3 g/dL), and higher glucose (115 mg/dL) levels. These factors contributed to the higher total score (1517.29) of this patient, indicating a higher-risk profile. The key distinguishing factors in risk assessment included age, ASA score, surgery type, and laboratory values. The model integrated a wide range of clinical

XSL•FO
RenderX

variables into a single risk assessment metric. The ability of the EACH score to distill complex clinical data into a quantifiable stroke risk assessment provides a clear and actionable tool for perioperative stroke risk stratification.

Table 3. Comparative risk stratification using Explainable Automated nonlinear Computation scoring system for Health score: low versus high-risk patients for perioperative stroke.

Characteristic	Actual value	Low risk score	Actual value	High risk score
Age	42.77	49.02	71.07	51.82
Sex, male	1	48.52	1	48.52
Height	160.4	49.78	156.3	49.91
Weight	61.8	48.60	55.8	53.02
BMI	24.02	48.85	22.84	48.85
Preoperative ASA[a]	1	46.17	4	53.76
Emergency operation	0	47.10	1	52.95
Preoperative hypertension	0	49.30	0	49.30
Preoperative diabetes	0	49.83	0	49.83
Preoperative cardiovascular accident	0	47.39	0	47.39
Preoperative asthma	0	50.00	0	50.00
Preoperative COPD[b]	0	49.59	0	49.59
Preoperative liver disease	0	49.81	1	50.17
Preoperative kidney disease	0	49.80	0	49.80
Preoperative tuberculosis	0	49.80	0	49.80
Surgery type[c]	0	0	1	100
Hemoglobin	13	49.66	12	49.49
Platelet	293	49.80	92	50.66
Blood urea nitrogen	14	49.27	8	49.27
Creatinine	0	50.40	0	50.40
Albumin	4	50.54	3	55.35
Sodium	137	50.41	137	50.41
Potassium	4	53.46	3	50.75
Glucose	81	47.39	115	54.08
Prothrombin time	11.9	52.13	11.6	52.13
Partial thromboplastin time	28	46.54	28	46.54
Aspartate aminotransferase	14	47.30	64	52.37
Alanine aminotransferase	22	49.64	78	50.30
Estimated glomerular filtration rate	120	49.06	90	50.83
Total scoring	—[d]	1379.16	—[d]	1517.29

[a]ASA: American Society of Anesthesiologists.

[b]COPD: chronic obstructive pulmonary disease.

[c]Surgery type included several type of surgeries Intrathoracic, intraabdominal, and supra-inguinal vascular surgery.

[d]Not applicable.

Discussion

Principal Findings

In this study, we developed and validated a comprehensible automated scoring system, the EACH score, using CatBoost. This system, when applied to predict perioperative stroke, showed superior performance compared with other ML models. It also outperformed the traditional risk score, the RCRI score, and other ML-based risk-scoring systems, such as AutoScore. Furthermore, the EACH score effectively discriminated between patients with high- and low-risk perioperative stroke.

XSL•FO
RenderX

Recent advancements in ML-based models have shown promising performance by integrating diverse datasets, irrespective of their linearity or nonlinearity. These models have the potential to overcome the limitations inherent in traditional statistical methods. However, broader application of these models is often restricted because of their lack of explainability and insufficient external validation [33,34].

Comparison With Previous Work

To address these issues, Xie et al [16] developed AutoScore, a scoring generator that combines RF-based approaches with traditional logistic regression. However, AutoScore may not fully capture complex clinical relationships or accurately represent biological contexts owing to methodological limitations and reliance on logistic regression [16,17].

The uniqueness of the present study lies in the development of an entirely ML-based scoring system. The EACH score is adaptable to a variety of data types without relying on predefined assumptions. In this system, the range of each continuous variable was determined by the inflection point on the plot using SHAP values. SHAP values demonstrate the key features and the absolute and relative predictive values of each feature within the model and are widely used to enhance the explainability of ML models [14,23]. The inflection point on the plot of SHAP values could be assumed to be a significant change of its predictive impact on outcome [35]. Using the above characteristics of SHAP values, we revealed that the importance of each feature did not increase linearly with its value, but rather showed various relationships. Consequently, the EACH score uniquely reflects the varying significance of identical features, thereby influencing the performance of the prediction model in a distinct manner. Another strength of the EACH score is its robust handling of missing data. If the data for an interval were missing, they were replaced with the average of the surrounding intervals' values, maintaining score completeness and accuracy [24,36].

When applied to the real-world clinical data, EACH outperformed the traditional score and other ML-based systems. This demonstrates that this approach could contribute to the construction of a novel risk assessment tool that achieves high predictive accuracy and is intuitive for clinicians to interpret and apply in real-world practice.

Using the EACH score, clinicians can better differentiate between high-risk and low-risk patients. This allows for the allocation of limited medical resources to high-risk groups, enabling closer monitoring of perioperative stroke occurrence and improving prognosis through timely treatment.

Furthermore, the EACH score can indirectly provide optimal targets for minimizing perioperative stroke risk for key correctable laboratory indicators such as albumin, electrolyte, and glucose by highlighting detailed trends in blood tests that increase risk nonlinearly. For instance, the EACH score identifies an albumin level of 4.1 to 4.4 mg/dL as being associated with the lowest perioperative stroke risk. Adjusting albumin levels to this range could potentially reduce the incidence of perioperative stroke.

To date, there has not been a prediction model that provides treatment strategies for optimal thresholds of preoperative laboratory indicators. The EACH score would be helpful for clinicians in decision-making to reduce perioperative stroke in complex preoperative environments by providing the transparent and straightforward explanation of how it works. Nevertheless, it is crucial to interpret these findings cautiously within the context of each patient's individual situation and clinical context.

Limitations

In addition, we addressed the second common limitation in the application of ML-based prediction models by validating the EACH score in external cohorts from a secondary medical center with different geographical locations, patient volumes, and characteristics from the developmental cohorts. The superior performance of the EACH score in the external validation cohort added robustness to its reliability. Despite these advancements, it is important to acknowledge that the efficacy of the EACH score is contingent on the quality and comprehensiveness of the input data. Incomplete or biased data can result in skewed results. However, the EACH score showed excellent performance even with imbalanced data, such as perioperative stroke, which is a rare disease [19,21]. Nevertheless, there is a need for ongoing updates and validation of the model across diverse clinical settings. Furthermore, because the current dataset comprises only Asian patients, it is imperative to prospectively validate the EACH score in multiracial cohorts to ensure its generalizability.

Conclusion

The EACH score represents a substantial advancement in the development of efficient, explainable, and automated ML-based risk assessment tools. The universal format of the EACH score is easily applicable to various medical scenarios, thereby reducing the need for labor-intensive data collection. Once integrated with electronic medical records, it generates specific predictions of different adverse outcomes [11,37]. Future developments should aim to enhance its user-friendliness for clinicians, potentially through the creation of intuitive interfaces or decision support tools that simplify the interpretation of results [38-40].

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant HI22C1295). This research was also supported by the Seoul National University Hospital research fund (grant 0320222110).

XSL•FO
RenderX

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. In addition, the corresponding code for model development is accessible on GitHub in the repository.

Authors' Contributions

M-YO and H-SK contributed equally as the first authors of this study. S-BL and SML contributed equally as co-corresponding authors of this study.
M-YO, H-SK, S-BL, and SML contributed to conceptualization, data curation, formal analysis, methodology and investigation of study. S-BL and SML handled funding acquisition of study. YMJ and H-CL managed data collection and curation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental Table. Complete SHapley Additive exPlanations value scoring by feature interval.
[DOCX File , 29 KB - jmir_v27i1e58021_app1.docx]

References

1. Foote C, Woodward M, Jardine MJ. Scoring Risk Scores: Considerations Before Incorporating Clinical Risk Prediction Tools Into Your Practice. Am J Kidney Dis 2017 May;69(5):555-557. [doi: 10.1053/j.ajkd.2017.02.005] [Medline: 28434521]
2. Jones HJ, de Cossart L. Risk scoring in surgical patients. Br J Surg 1999 Feb;86(2):149-157. [doi: 10.1046/j.1365-2168.1999.01006.x] [Medline: 10100780]
3. Six AJ, Backus BE, Kelder JC. Chest pain in the emergency room: value of the HEART score. Neth Heart J 2008;16(6):191-196 [FREE Full text] [doi: 10.1007/BF03086144] [Medline: 18665203]
4. Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest 2010 Nov;138(5):1093-1100. [doi: 10.1378/chest.10-0134] [Medline: 20299623]
5. Wells PS, Hirsh J, Anderson DR, Lensing AW, Foster G, Kearon C, et al. Accuracy of clinical assessment of deep-vein thrombosis. Lancet 1995 May 27;345(8961):1326-1330. [doi: 10.1016/s0140-6736(95)92535-x] [Medline: 7752753]
6. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 2010 Feb;137(2):263-272. [doi: 10.1378/chest.09-1584] [Medline: 19762550]
7. Royston P, Altman DG. Visualizing and assessing discrimination in the logistic regression model. Stat Med 2010 Oct 30;29(24):2508-2520. [doi: 10.1002/sim.3994] [Medline: 20641144]
8. Steyerberg EW, Vergouwe YJ. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014 Aug 01;35(29):1925-1931 [FREE Full text] [doi: 10.1093/eurheartj/ehu207] [Medline: 24898551]
9. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
10. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. Sci Rep 2022 Oct 26;12(1):17981 [FREE Full text] [doi: 10.1038/s41598-022-22514-4] [Medline: 36289266]
11. Carrasco-Ribelles LA, Llanes-Jurado J, Gallego-Moll C, Cabrera-Bean M, Monteagudo-Zaragoza M, Violán C, et al. Prediction models using artificial intelligence and longitudinal data from electronic health records: a systematic methodological review. J Am Med Inform Assoc 2023 Nov 17;30(12):2072-2082 [FREE Full text] [doi: 10.1093/jamia/ocad168] [Medline: 37659105]
12. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208 [FREE Full text] [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
13. Goldman L, Caldera DL, Nussbaum SR, Southwick FS, Krogstad D, Murray B, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 1977;297(16):845-850. [doi: 10.1056/nejm197710202971601]
14. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
15. Rudin CJ. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell 2019 May;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
16. Xie F, Chakraborty B, Ong MEH, Goldstein BA, Liu N. AutoScore: a machine learning-based automatic clinical score generator and its application to mortality prediction using electronic health records. JMIR Med Inform 2020;8(10):e21798 [FREE Full text] [doi: 10.2196/21798] [Medline: 33084589]

XSL•FO
RenderX

17. Yuan H, Xie F, Ong MEH, Ning Y, Chee ML, Saffari SE, et al. AutoScore-Imbalance: An interpretable machine learning tool for development of clinical scores with rare events data. J Biomed Inform 2022 May;129:104072 [FREE Full text] [doi: 10.1016/j.jbi.2022.104072] [Medline: 35421602]
18. Leary MC, Varade P. Perioperative Stroke. Curr Neurol Neurosci Rep 2020 Apr 27;20(5):12. [doi: 10.1007/s11910-020-01033-7] [Medline: 32342230]
19. Lindberg A, Flexman AJ. Perioperative stroke after non-cardiac, non-neurological surgery. BJA Educ 2021 Feb;21(2):59-65 [FREE Full text] [doi: 10.1016/j.bjae.2020.09.003] [Medline: 33889431]
20. Mashour GA, Shanks AM, Kheterpal S. Perioperative stroke and associated mortality after noncardiac, nonneurologic surgery. Anesthesiology 2011 Jun;114(6):1289-1296 [FREE Full text] [doi: 10.1097/ALN.0b013e318216e7f4] [Medline: 21478735]
21. Ko SB. Perioperative stroke: pathophysiology and management. Korean J Anesthesiol 2018 Feb;71(1):3-11 [FREE Full text] [doi: 10.4097/kjae.2018.71.1.3] [Medline: 29441169]
22. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
23. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct;2(10):749-760 [FREE Full text] [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
24. Little RJ, Rubin DB. Statistical Analysis with Missing Data, Third Edition. United States: Wiley; 2019.
25. Cortes C, Vapnik VJ. Support-vector networks. Machine Learning 1995 Sep;20(3):273-297. [doi: 10.1007/bf00994018]
26. Safavian SR, Landgrebe DJ. A survey of decision tree classifier methodology. IEEE Trans. Syst., Man, Cybern 1991;21(3):660-674. [doi: 10.1109/21.97458]
27. Breiman L. Random Forests. Machine Learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
28. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV. CatBoost: Unbiased Boosting With Categorical Features. Advances in Neural Information Processing Systems 2018;31:6639-6649 [FREE Full text]
29. Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. 2013 Presented at: Proceedings of the 30th International Conference on Machine Learning; 2023 Feb 08; USA p. 115-123 URL: https://proceedings.mlr.press/v28/bergstra13.html
30. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp. and Comm 2015;19(1):29-33. [doi: 10.1145/2786984.2786995]
31. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29-36. [doi: 10.1148/radiology.143.1.7063747] [Medline: 7063747]
32. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Circulation 2015;131(2):211-219. [doi: 10.1161/circulationaha.114.014508]
33. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021;23(4):e25759 [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
34. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning‐based prediction models in healthcare. WIREs Data Min & Knowl 2020;10(5):e1379. [doi: 10.1002/widm.1379]
35. Arin P, Minniti M, Murtinu S, Spagnolo N. Inflection points, kinks, and jumps: a statistical approach to detecting nonlinearities. Organizational Research Methods 2021;25(4):786-814. [doi: 10.1177/10944281211058466]
36. Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O. A survey on missing data in machine learning. J Big Data 2021;8(1):140 [FREE Full text] [doi: 10.1186/s40537-021-00516-9] [Medline: 34722113]
37. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform for collaborative management of cataracts. Br J Ophthalmol 2019;103(11):1553-1560 [FREE Full text] [doi: 10.1136/bjophthalmol-2019-314729] [Medline: 31481392]
38. Fonseca A, Ferreira A, Ribeiro L, Moreira S, Duque C. Embracing the future-is artificial intelligence already better? A comparative study of artificial intelligence performance in diagnostic accuracy and decision-making. Eur J Neurol 2024;31(4):e16195. [doi: 10.1111/ene.16195] [Medline: 38235841]
39. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial intelligence and surgical decision-making. JAMA Surg 2020;155(2):148-158 [FREE Full text] [doi: 10.1001/jamasurg.2019.4917] [Medline: 31825465]
40. Loftus TJ, Upchurch GR, Bihorac A. Use of artificial intelligence to represent emergent systems and augment surgical decision-making. JAMA Surg 2019;154(9):791-792 [FREE Full text] [doi: 10.1001/jamasurg.2019.1510] [Medline: 31215974]

Abbreviations

ASA: American Society of Anesthesiologists
AUC: area under the receiver operating characteristic curve

XSL•FO
RenderX

BMC: Boramae Medical Center
EACH: Explainable Automated Non-linear Computation scoring system for Health
IRB: institutional review board
NPV: negative predictive value:
ML: machine learning
PPV: positive predictive value
RCRI: Revised Cardiac Risk Index
RF: Random Forest
SHAP: SHapley Additive exPlanations
SNUH: Seoul National University Hospital
SVM: Support Vector Machine

Edited by T de Azevedo Cardoso, S Ma; submitted 03.03.24; peer-reviewed by E Vashishtha, K Singh; comments to author 15.03.24; revised version received 24.03.24; accepted 30.10.24; published 19.03.25.

Please cite as:
Oh MY, Kim HS, Jung YM, Lee HC, Lee SB, Lee SM
Machine Learning–Based Explainable Automated Nonlinear Computation Scoring System for Health Score and an Application for Prediction of Perioperative Stroke: Retrospective Study
J Med Internet Res 2025;27:e58021
URL: https://www.jmir.org/2025/1/e58021
doi:10.2196/58021
PMID:

©Mi-Young Oh, Hee-Soo Kim, Young Mi Jung, Hyung-Chul Lee, Seung-Bo Lee, Seung Mi Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Disclosure of Results From Artificial Intelligence/Machine Learning Research in Health Care: Comprehensive Analysis of ClinicalTrials.gov, PubMed, and Scopus Data (2010-2023)

Shoko Maru[1], PhD; Ryohei Kuwatsuru[1,2], MD, PhD; Michael D Matthias[3], BEng; Ross J Simpson Jr[4], MD, PhD

[1]Real-World Evidence and Data Assessment (READS), Graduate School of Medicine, Juntendo University, Tokyo, Japan

[2]Department of Radiology, School of Medicine, Juntendo University, Tokyo, Japan

[3]Matthias IT Pty Ltd, Brisbane, Australia

[4]Division of Cardiology, School of Medicine, University of North Carolina Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Shoko Maru, PhD
Real-World Evidence and Data Assessment (READS)
Graduate School of Medicine
Juntendo University
2-1-1 Hongo, Bunkyo-ku
Tokyo, 113-8421
Japan
Phone: 81 338133111
Email: shoko.maru@alumni.griffithuni.edu.au

Abstract

Background: Despite the rapid growth of research in artificial intelligence/machine learning (AI/ML), little is known about how often study results are disclosed years after study completion.

Objective: We aimed to estimate the proportion of AI/ML research that reported results through ClinicalTrials.gov or peer-reviewed publications indexed in PubMed or Scopus.

Methods: Using data from the Clinical Trials Transformation Initiative Aggregate Analysis of ClinicalTrials.gov, we identified studies initiated and completed between January 2010 and December 2023 that contained AI/ML-specific terms in the official title, brief summary, interventions, conditions, detailed descriptions, primary outcomes, or keywords. For 842 completed studies, we searched PubMed and Scopus for publications containing study identifiers and AI/ML-specific terms in relevant fields, such as the title, abstract, and keywords. We calculated disclosure rates within 3 years of study completion and median times to disclosure—from the "primary completion date" to the "results first posted date" on ClinicalTrials.gov or the earliest date of journal publication.

Results: Of 842 completed studies (n=357 interventional; n=485 observational), 5.5% (46/842) disclosed results on ClinicalTrials.gov, 13.9% (117/842) in journal publications, and 17.7% (149/842) through either route within 3 years of completion. Higher disclosure rates were observed for trials: 10.4% (37/357) on ClinicalTrials.gov, 19.3% (69/357) in journal publications, and 26.1% (93/357) through either route. Randomized controlled trials had even higher disclosure rates: 11.3% (23/203) on ClinicalTrials.gov, 24.6% (50/203) in journal publications, and 32% (65/203) through either route. Nevertheless, most study findings (82.3%; 693/842) remained undisclosed 3 years after study completion. Trials using randomization (vs nonrandomized) or masking (vs open label) had higher disclosure rates and shorter times to disclosure. Most trials (85%; 305/357) had sample sizes of ≤1000, yet larger trials (n>1000) had higher publication rates (30.8%; 16/52) than smaller trials (n≤1000) (17.4%; 53/305). Hospitals (12.4%; 42/340), academia (15.1%; 39/259), and industry (13.7%; 20/146) published the most. High-income countries accounted for 82.4% (89/108) of all published studies. Of studies with disclosed results, the median times to report through ClinicalTrials.gov and in journal publications were 505 days (IQR 399-676) and 407 days (IQR 257-674), respectively. Open-label trials were common (60%; 214/357). Single-center designs were prevalent in both trials (83.3%; 290/348) and observational studies (82.3%; 377/458).

Conclusions: For over 80% of AI/ML studies completed during 2010-2023, study findings remained undisclosed even 3 years after study completion, raising questions about the representativeness of publicly available evidence. While methodological rigor

XSL•FO
RenderX

was generally associated with higher publication rates, the predominance of single-center designs and high-income countries may limit the generalizability of the results currently accessible.

(J Med Internet Res 2025;27:e60148) doi:10.2196/60148

KEYWORDS

machine learning; ML; artificial intelligence; AI; algorithm; model; analytics; deep learning; health care; health disparities; disparity; social disparity; social inequality; social inequity; data-source disparities; ClinicalTrials.gov; clinical trial; database; PubMed; Scopus; public disclosure of results; public disclosure; dissemination

Introduction

The number of studies on artificial intelligence/machine learning (AI/ML) has surged in recent years, exceeding prior expectations [1]. The growth of AI/ML research in health care continues to gain momentum, driven by its potential to enable early detection of serious conditions in resource-constrained settings or facilitate timely identification of patient deterioration that might otherwise go unnoticed, to name a few. However, publicly available data are highly heterogeneous, including promotional claims, forward-looking statements, white papers, and preprints. It is often unclear what data underpin the claims made for AI/ML tools, whether regulated, nonregulated, or nonproprietary.

Weak publication records in AI/ML research have been highlighted [2,3], and even when AI/ML research results are available in peer-reviewed journals, they were found to be poorly reported [4,5]; lacking transparency, hindering replicability [6]; involving "spin" and hype by overinflating models' predictive abilities [7]; or being at high risk of bias [8,9]. These challenges complicate the interpretation of published results and make it difficult to determine whether new findings ultimately provide tangible benefits to patients.

As AI/ML research continues to proliferate, so do systematic reviews and meta-analyses, which presuppose "all relevant evidence" as their foundation. However, this principle is undermined when underreporting is pervasive. Underreporting can lead to biased effect estimates and compromise the diagnostic or prognostic performance of AI/ML, on which clinical decision-making depends. After all, systematic reviews can only reflect the results accessible—a portion of the total research conducted. Lock and Wells [10] have deemed underreporting of research a form of both scientific and ethical misconduct, as it not only leads to biased and imprecise effect estimates but also breaches implied contracts with study participants, who contribute with the expectation of advancing knowledge [11].

The underreporting of clinical trial results has been well documented, long before the emergence of AI/ML. For example, a 2003 study in Spain revealed that fewer than one-third (31%; 38/123) of clinical trials approved by an ethics committee at a major hospital had results published in peer-reviewed journals within 3 years of study completion [12]. Numerous studies using the ClinicalTrials.gov database have provided further evidence since [13-19]. This enables us to compare the disclosure rates of AI/ML trials with those previously reported for non-AI/ML trials registered on ClinicalTrials.gov.

Finally, AI's impact on health equity is also extensively debated. Some argue that access to the very factors driving AI in health care, such as electronic health records and computing power, may exacerbate existing health care disparities and perpetuate inequities in who benefits most from AI [20]. Certain population groups—demographic, geographic, or economic —can be disproportionately overrepresented in clinical AI/ML research.

This study examines (1) the proportion and patterns of public disclosure of AI/ML research results through ClinicalTrials.gov and peer-reviewed publications indexed in PubMed or Scopus, and (2) whether disclosure rates vary according to key study characteristics, including those related to population representation, such as geographic region and gross national income.

Methods

Data Source

We used ClinicalTrials.gov, a trial registry and results database, and sourced data from the Clinical Trials Transformation Initiative Aggregate Analysis of ClinicalTrials.gov (CTTI AACT) [21], which allows open access to the complete set of studies registered in ClinicalTrials.gov, including additional fields that are not readily available in direct exports from ClinicalTrials.gov. The CTTI AACT data dictionary is publicly accessible [22]. A static version of the CTTI AACT database was downloaded for analysis on February 6, 2024, via PostgreSQL, as previously described [23]. We identified studies that initiated and completed between January 2010 and December 2023 and contained AI/ML-specific terms in the official title, detailed description, brief summary, interventions, conditions, primary outcomes, and keywords. These terms could have appeared in studies where AI/ML was used either as an intervention or a method. A total of 842 AI/ML studies were completed by the end of 2023. The search strategies are provided in Multimedia Appendix 1, Table S1. The study flow diagram is provided in Multimedia Appendix 2, Figure S1. Detailed methods on data extraction, including SQL codes, are available in Multimedia Appendix 3.

AI/ML Publications Linked to ClinicalTrials.gov Identifiers (NCT Numbers)

NCT numbers are unique study identifiers assigned by ClinicalTrials.gov. To identify studies with corresponding journal publications, we searched PubMed and Scopus for these identifiers using each database's application programming interface (API). In PubMed, we searched for NCT numbers in the titles, abstracts, and "trial registration" fields, while in Scopus, we searched the titles and abstracts. Additionally, we

XSL•FO
RenderX

accessed "publication" data on ClinicalTrials.gov using the "reference_type" field in the AACT database. Details on the APIs used for PubMed and Scopus are provided in Multimedia Appendix 3.

Since multiple publications may exist under the same NCT, and not all are related to AI/ML, we included only those that contained AI/ML-specific terms in the title, abstract, keywords, or Medical Subject Headings (MeSH) fields in PubMed. In Scopus, AI/ML terms were searched in the titles and abstracts.

To capture articles likely to be the direct study output, we excluded nonrelevant article types (eg, protocols, reviews, meta-analyses) and those published before primary completion dates. ClinicalTrials.gov defines the primary completion date as the date on which data collection is completed for all primary outcomes [24].

Main Outcomes and Measures

After identifying eligible articles, we determined the proportion of AI/ML studies that disclosed results within 3 years of the primary completion date, either on ClinicalTrials.gov or in journal publications (the number of studies with disclosed results divided by the number of completed studies). We focused on disclosures via ClinicalTrials.gov or peer-reviewed publications (not preprints), as these are typically searched for systematic reviews and meta-analyses.

For studies with disclosed results, we also calculated the median time from the primary completion date to the disclosure date along with the IQR. The disclosure date was defined as either the "results first posted date" recorded on ClinicalTrials.gov or the date of the earliest journal publication linked to the NCT. If a study had multiple publications, only the earliest publication date was used. Each NCT with at least one publication was counted as disclosed.

All variables were derived from ClinicalTrials.gov. The variable on countries (study location) was categorized according to the World Bank's gross national income–based classification: low income (≤US $1135), lower-middle income (US $1136 to $4465), upper-middle income (US $4466 to $13,845), or high income (≥US $13,846) [25]. On lead sponsor, those labeled "other" were reassigned into hospital/clinic, academia, industry, government, and nonprofit organization based on lead sponsor names.

Details on the data extraction are provided in Multimedia Appendix 3.

Statistical Analysis

Data were summarized descriptively: count, percentage, and the median with the IQR. Categorical variables are presented as counts and percentages within 3 years of the primary completion date. Tabulations exclude missing values (ie, n=842 if a variable had no missing values).

Ethics Approval and Reporting Guidelines

Ethics approval was not required for this study because only publicly available data were analyzed. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines [26].

Results

Overview of Completed Studies on AI/ML

Of 842 completed studies (n=357 interventional; n=485 observational), only 5.5% (46/842) disclosed results on ClinicalTrials.gov, 13.9% (117/842) in journal publications, and 17.7% (149/842) through either route within 3 years of completion (Multimedia Appendix 2, Figure S1).

The disclosure rates were higher among trials only: 10.4% (37/357) on ClinicalTrials.gov, 19.3% (69/357) in journal publications, and 26.1% (93/357) through either route. Rates among randomized controlled trials (RCTs) were even higher: 11.3% (23/203) on ClinicalTrials.gov, 24.6% (50/203) in journal publications, and 32% (65/203) through either route.

Among studies with disclosed results, the median reporting times were 505 days (IQR 399-676) on ClinicalTrials.gov and 407 days (IQR 257-674) in journal publications.

Figure 1 shows the trend in studies started each year, completed studies, and studies that reported results within 3 years of study completion. Despite a surge in new studies each year, studies that disclosed results through either route remained scarce.

XSL•FO
RenderX

Figure 1. Studies initiated ("started"), studies completed ("completed"), and completed studies with results reported either on ClinicalTrials.gov or in journal publications within 3 years of completion ("disclosed within 3 years"). Despite a surge in new studies each year, the number of studies reporting results through either route remained scarce.

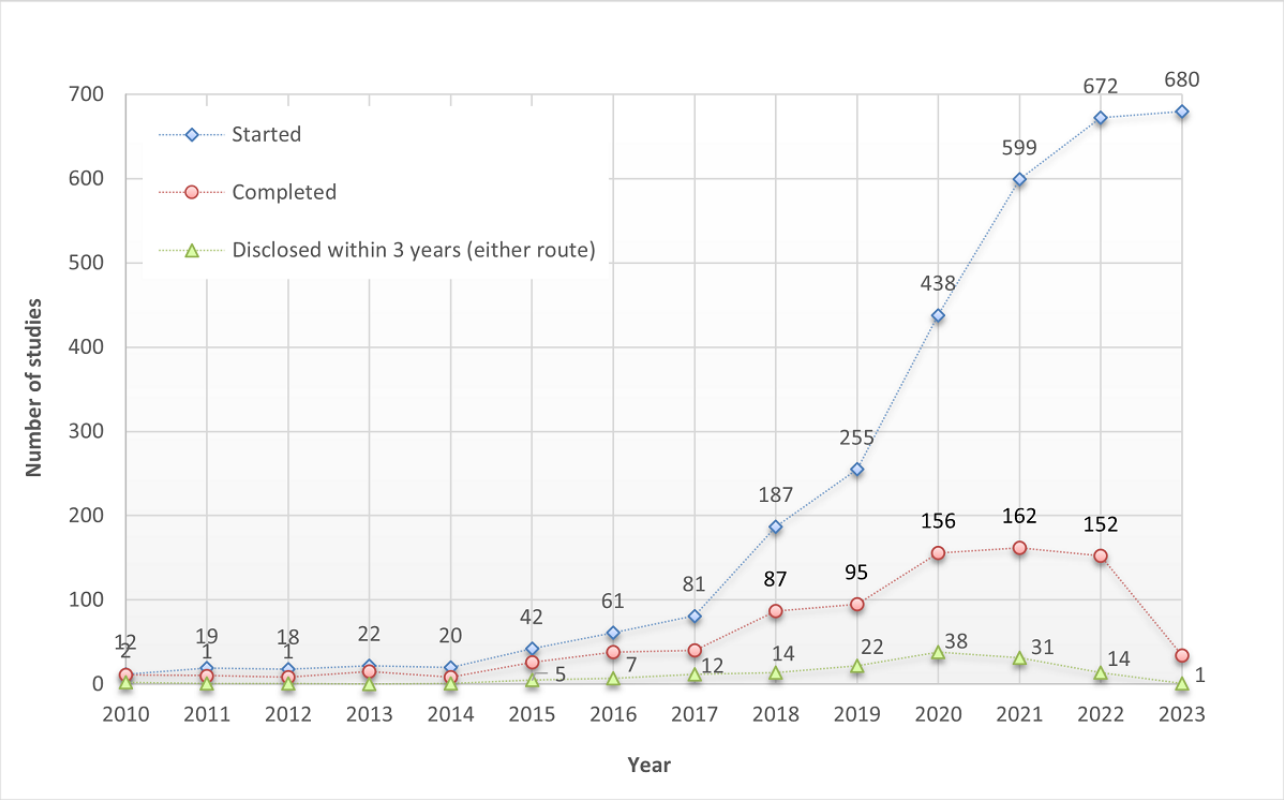

Result Disclosure Rates and Time to Disclosure by Study Characteristics

Table 1 shows reporting rates stratified by the primary completion year. The reporting rates on ClinicalTrials.gov remained low over time. However, from 2018 to 2020, there was a modest increase in publication rates within 3 years of completion, and the median time to publication decreased from approximately 2 years (723 days) to 1 year (385 days). The IQR also decreased from 1166 days in 2018 to 279 in 2020. Before 2018, there were too few AI/ML publications to allow meaningful interpretation, and 2021-2023 was not assessable because the 3-year grace period had not passed as of our data cutoff on December 31, 2023.

Table 2 presents disclosure rates and times stratified by various study characteristics.

Disclosure rates and their timing varied by study design and setting (Table 2), as summarized below.

XSL•FO
RenderX

Table 1. Results disclosure rates and the time to disclosure, stratified by primary completion year.

Primary completion year	Completed studies, n	CTG[a] posting ≤3 years, n (%)	Journal publication ≤3 years, n (%)	Time to CTG posting[b] (days), median (IQR)	Time to journal publication[b] (days), median (IQR)
2010	1	0 (0)	1 (100)	—[c]	—
2011	2	0 (0)	0 (0)	—	—
2012	3	0 (0)	1 (33)	—	—
2013	3	0 (0)	0 (0)	—	—
2014	7	0 (0)	0 (0)	—	—
2015	9	0 (0)	1 (11)	—	1611 (1100-2122)
2016	20	1 (5)	2 (10)	1026 (938-1113)	487 (415-1049)
2017	25	4 (16)	4 (16)	917 (806-953)	820 (586-1077)
2018	55	4 (7)	5 (9)	538 (512-586)	723 (366-1532)
2019	82	6 (7)	14 (17)	611 (433-821)	681 (339-954)
2020	109	8 (7)	24 (22)	427 (288-526)	385 (274-553)

[a]CTG: ClinicalTrials.gov.

[b]Reflects all studies that reported results (including those disclosed after 3 years of completion).

[c]Not applicable.

XSL•FO
RenderX

Table 2. Disclosure rates and the time to disclosure by study characteristics. Tabulations exclude missing values (ie, n=842 if no missing values).

Study characteristics	Completed studies, n	CTG[a] posting ≤3 years, n (%)	Journal publication ≤3 years, n (%)	Time to CTG posting[b] (days), median (IQR)	Time to journal publication[b] (days), median (IQR)
All completed	842	46 (5.5)	117 (13.9)	505 (399-676)	407 (257-674)
Study type (n=842)					
Inverventional	357	37 (10.4)	69 (19.3)	485 (397-574)	368 (272-624)
Observational	485	9 (1.9)	48 (9.9)	722 (526-871)	429 (253-701)
Randomization (interventional; n=357)					
Randomized	203	23 (11.3)	50 (24.6)	485 (399-565)	357 (273-565)
Nonrandomized	154	37 (24)	19 (12.3)	466 (398-593)	549 (262-780)
Masking (interventional; n=357)					
Open label	214	23 (10.7)	37 (17.3)	485 (397-575)	474 (297-672)
Single blind	76	4 (5.3)	16 (21.1)	562 (455-734)	303 (261-405)
Double, triple, quadruple blind	67	10 (14.9)	16 (23.9)	486 (363-548)	335 (215-658)
Study center (interventional; n=348)					
Single center	290	28 (9.7)	55 (19)	495 (399-641)	373 (274-616)
Multicenter	58	8 (13.8)	10 (17.2)	443 (404-494)	308 (218-698)
Study center (observational; n=458)					
Single center	377	8 (2.1)	35 (9.3)	690 (506-834)	427 (252-681)
Multicenter	81	0 (0)	10 (12.3)	__[c]	426 (327-529)
Enrollment (interventional; n=357)					
≤100	181	19 (10.5)	28 (15.5)	427 (397-648)	343 (272-624)
101-1000	124	11 (8.9)	25 (20.2)	490 (418-583)	373 (252-560)
1001-5000	31	3 (9.7)	10 (32.3)	462 (442-465)	594 (334-743)
>5000	21	4 (19)	6 (28.6)	479 (358-537)	413 (279-597)
Enrollment (observational; n=485)					
≤100	145	4 (2.8)	17 (11.7)	794 (627-863)	289 (231-423)
101-1000	206	5 (2.4)	21 (10.2)	638 (526-835)	542 (386-777)
1001-5000	82	0 (0)	9 (11)	—	392 (246-459)
>5000	52	0 (0)	1 (1.9)	—	1282 (944-1745)
Time perspective (observational[d]; n=468)					
Prospective	233	6 (2.6)	25 (10.7)	754 (546-892)	386 (246-628)
Retrospective	164	3 (1.8)	14 (8.5)	690 (527-787)	457 (362-714)
Cross-sectional	71	0 (0)	9 (12.7)	—	427 (333-532)
Primary purpose: top 4 (n=221)					
Treatment	69	9 (13)	11 (15.9)	505 (467-634)	349 (309-521)
Diagnostic	95	7 (7.4)	24 (25.3)	555 (454-671)	430 (282-747)
Prevention	29	7 (24.1)	5 (17.2)	433 (322-550)	188 (94-411)
Health services research	28	5 (17.9)	8 (28.6)	400 (189-424)	293 (215-565)
Lead sponsor: top 5 (n=807)					
Hospital	340	4 (1.2)	42 (12.4)	463 (323-540)	407 (274-674)
Academia	259	22 (8.5)	39 (15.1)	485 (412-618)	402 (273-624)
Industry	146	10 (6.8)	20 (13.7)	651 (318-808)	398 (224-681)

XSL•FO
RenderX

Study characteristics	Completed studies, n	CTG[a] posting ≤3 years, n (%)	Journal publication ≤3 years, n (%)	Time to CTG posting[b] (days), median (IQR)	Time to journal publication[b] (days), median (IQR)
Government	18	2 (11.1)	3 (16.7)	720 (605-835)	280 (233-555)
Nonprofit organization	44	6 (13.6)	8 (18.2)	536 (458-551)	435 (338-522)
Regulation status (n=842)					
FDA[e] regulated	59	14 (23.7)	14 (23.7)	547 (403-707)	467 (233-608)
Not FDA regulated	783	32 (4.1)	103 (13.2)	490 (400-662)	397 (272-678)
US study site (n=842)					
US site	228	35 (15.4)	34 (14.9)	490 (400-611)	440 (296-624)
No US site	614	13 (2.1)	83 (13.5)	571 (374-814)	375 (248-677)
Study location: region (n=806)					
Europe	299	3 (1)	42 (14)	821 (666-966)	375 (251-515)
Asia and Pacific	192	4 (2.1)	24 (12.5)	571 (507-612)	411 (264-732)
North America	235	33 (14)	37 (15.7)	490 (400-611)	430 (282-638)
Middle East	33	3 (9.1)	4 (12.1)	246 (170-540)	200 (158-338)
Africa	10	1 (10)	1 (10)	424 (424-424)	577 (577-577)
Central and South America	6	0 (0)	0 (0)	—	—
Study location: gross national income[f] (n=775)					
High-income countries	613	42 (6.9)	89 (14.5)	505 (399-676)	389 (255-616)
Upper-middle-income countries	139	0 (0)	16 (11.5)	—	391 (236-718)
Lower-middle-income countries	23	2 (8.7)	3 (13)	490 (457-522)	430 (375-504)
Low-income countries	0	0 (0)	0 (0)	—	—

[a]CTG: ClinicalTrials.gov.

[b]Time to results reporting (posting on CTG or in journal publications) also includes those reported after 3 years.

[c]Not applicable.

[d]Time perspective: data available only for observational studies.

[e]FDA: Food and Drug Administration.

[f]Gross national income–based classification as per the World Bank.

Interventional Studies (vs Observational)

Trials accounted for 42.4% (357/842) of completed studies; they had higher publication rates (19.3% vs 9.9%) and posted results on ClinicalTrials.gov more often (10.4% vs 1.9%), with shorter times to disclosure than observational studies.

Randomization

RCTs accounted for 56.9% (203/357) of completed trials and had higher publication rates than non-RCTs (24.6% vs 12.3%). However, non-RCTs posted results on ClinicalTrials.gov more often than RCTs (24% vs 11.3%). RCTs reached publication sooner than non-RCTs (357 vs 549 days).

Masking

Open-label trials were the most common (59.9%; 214/357), followed by single-blind (21.3%; 76/357) and double-blind or higher (18.8%; 67/357) designs. Single-blind trials (21.1%; 16/76), as well as double-, triple-, or quadruple-blinded trials combined (23.9%; 16/67) had higher publication rates and reached publication sooner than open-label trials (17.3%; 37/214).

Single Center (vs Multicenter)

Single-center designs dominated both trials (83.3%; 290/348) and observational studies (82.3%; 377/458). The disclosure rates did not vary substantially between single-center and multicenter studies.

Enrollment

Most trials (85%; 305/357) had sample sizes of ≤1000, yet larger trials (n>1000) had higher publication rates (30.8%; 16/52) than smaller trials (n≤1000) (17.4%; 53/305). However, among observational studies, larger studies (n>1000) had slightly lower publication rates (7.5%; 10/134) than smaller studies (10.8%; 38/351).

Time Perspective (Observational Studies Only)

Nearly half of observational studies used prospective designs (49.8%; 233/468), followed by retrospective designs (35%;

XSL•FO
RenderX

164/468). Prospective studies had slightly higher publication rates (10.7%; 25/233) and reached publication sooner (386 days) than retrospective studies (8.5%; 14/164; 457 days).

Lead Sponsor

The majority of studies were sponsored by hospitals (40.4%; 340/842) and academia (30.8%; 259/842). The most published sponsors were hospitals (12.4%; 42/340), academia (15.1%; 39/259), and industry (13.7%; 20/146), but the percentages were slightly higher for studies sponsored by governments (16.7%; 3/18) or nonprofit organizations (18.2%; 8/44), albeit with small event numbers.

Primary Study Purpose

Diagnostic (31.9%; 95/298) and treatment (23.2%; 69/298) were the most common study purposes. Publication rates were highest for diagnostic (25.3%; 24/95) and health services research (28.6%; 8/28). Prevention studies were more frequently reported on ClinicalTrials.gov (24.1%; 7/29) than in journal publications (17.2%; 5/29). Prevention studies reached publication sooner (188 days) than health services research (293 days), treatment research (349 days), and diagnostic studies (430 days).

Study Locations (Geographic and Economic)

The majority of completed studies were conducted in Europe (37.1%; 299/806); North America (29.2%; 235/806); or the Asia-Pacific (23.8%; 192/806). When categorized by national income level, most studies were conducted in high-income countries (79%; 613/775), followed by upper-middle-income (18%; 139/775) and lower-middle-income (3%; 23/775) countries, and none in low-income countries. Publication rates were also highest in Europe, North America, Asia, and high-income countries, followed by upper-middle-income countries. Finally, 82.4% (89/108) of studies with corresponding publications were conducted in high-income countries.

Regulatory Status

Only 7% (59/842) of completed studies were “FDA-regulated,” while 72.9% (614/842) did not involve US sites. FDA-regulated studies had higher publication rates than non-RCTs (23.7% vs 13.2%). FDA-regulated studies (vs nonregulated) and those with a US site (vs no US site) were more likely to post results on ClinicalTrials.gov.

Discussion

The proliferation of AI/ML research continues, accompanied by a rise in systematic reviews. However, such reviews can capture only the published portion of all relevant evidence. In this cross-sectional analysis of 842 AI/ML studies completed during 2010-2023, we quantified the extent of results that remain undisclosed years after completion.

Principal Findings

First, despite a surge in new AI/ML studies each year, most study findings remained undisclosed even 3 years after study completion, which is more lenient than the 2-year grace period used elsewhere [13,14]. Of 842 completed studies, only 17.7% (149/842) disclosed results through either route within 3 years, although rates were somewhat higher among trials (26.1%; 93/357) and RCTs (32%; 65/203).

Second, study features of greater methodological or logistical rigor—such as interventional (vs observational), randomization (vs nonrandomized), masking (vs open label), and larger sample sizes (>1000 vs ≤1000) *in trials*—had higher disclosure rates with shorter median reporting times through either dissemination route. However, multicenter status had no discernible impact. Less rigorous features, such as single-center designs (82.8%; 667/806) and open-label trials (60%; 214/357), were prevalent.

Finally, most published studies were from high-income countries (82.4%; 89/108).

Comparison With Previous Work

AI/ML Trials Versus Non-AI/ML Trials Registered on ClinicalTrials.gov

To compare results with those of previous research on trials registered on ClinicalTrials.gov, a trial-only subgroup was used. Among the 357 trials, the disclosure rate through either route within 3 years of completion was 26.1% (93/357). This rate was notably lower than that for non-AI/ML settings: oncology trials (60.7%; 7425/12,240 within 2 years) [13], pharmaceutical and biopharmaceutical phase II-IV efficacy trials (25.2%; 3822/15,084 within 1 year) [18], National Institutes of Health–funded trials (46.3%; 294/635 within 30 months) [15], completed trials (45.9%; 311/677 within 2 years) [19], trials by US-based academic medical centers (35.9%; 1560/4347 within 2 years) [14], phase III-IV RCTs of drug interventions (50%; 297/594) [16], and trials approved by a hospital ethics committee in Spain (31%; 38/123 within 3 years). Our 3-year disclosure rate (26.1%; 93/357), equivalent to roughly 8.7% per year, was also lower than the 1-year rate for mobile health trials (18.5%; 25/135) [27].

Several factors could explain the lower publication rates. First, for the non-AI/ML trials cited above, registration and result disclosure may have been mandatory (eg, regulated, publicly funded, efficacy trials), unlike much of digital health research [28]. Second, the rarity of efficacy studies in AI/ML research to date may play a role, as exploratory or formative research may be deemed less publishable. Even so, however, summary results (objective data) can still be posted on ClinicalTrials.gov, making otherwise unpublishable results accessible. Third, the lack of reporting standards for AI/ML studies is another factor. However, a range of AI/ML-specific tools now exist, such as the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) 2024 Update [29] and the Consolidated Reporting Guidelines for Machine Learning Modeling Studies [30]. Fourth, preprint servers are widely used in computer science, reflecting the field’s emphasis on speed, openness, and adaptability. If preprints are considered sufficient—particularly when validated through reputable conferences—the urgency for peer-reviewed publication may be lower within the AI community. As a form of sensitivity analysis, we assessed the use of preprints in AI/ML studies registered on ClinicalTrials.gov by searching arXiv, bioRxiv, Research Square, medRxiv, SSRN, and PsyArXiv. We identified only 14 preprints linked to NCT numbers; of these, 6 were subsequently published in journals (already

XSL•FO
RenderX

captured in our findings), while the remaining 8 were available only as preprints. That is, among the 693 completed studies with no disclosed results, just 1.2% (8/693) had NCT-matched preprints. Thus, even if preprints were considered as an additional dissemination route, the overall disclosure rate would increase only marginally to 18.6% (157/842), compared to the base case of 17.7% (149/842), which does little to explain the low publication rates.

Trial Designs

Previous research in non-AI/ML settings (mostly drugs or biomedical) indicated that trials using randomization, masking, larger sample sizes, or multicenter designs were more likely to be published [13,31,32]. Our findings align with these patterns—except for multicenter status. Although methodological rigor was generally associated with higher publishability, less rigorous features—such as single-center (83.3%; 290/348) or open-label designs (60%; 214/357)—were common among the AI/ML trials. Notably, single-center designs predominated in both trials (83.3%; 290/348) and observational studies (82.3%; 377/458). This pattern was consistent with recent systematic reviews of RCTs on AI, where single-center designs accounted for 62.8% (54/86) [33,34] and 59% (23/39) [34].

The Distribution of Studies by Economic Status

According to scoping reviews on AI publications in health care, high-income countries accounted for 73.3% (33/45) of articles published in 2011-2022 [35] as well as 94.5% (240/254) and 93.3% (235/251) of those published in 2019 [20]. In our data, 82.4% (89/108) of the studies with publications from 2010-2023 were from high-income countries, and 14.8% (16/108) were from upper-middle-income countries, most of which were from China (88%; 14/16). That is, high-income countries and China accounted for 95.4% (103/108) of studies with published results. Despite differences in data sources and methodologies, high-income countries were consistently overrepresented across studies.

Broader Implications of the Study

AI/ML is increasingly being evaluated both as an intervention and as a decision-support tool for clinicians. Yet, in our sample, findings from over 80% of completed studies remained undisclosed even after 3 years, raising concerns about representativeness in systematic reviews or meta-analyses. Reviews that rely on skewed samples of favorable results—excluding null or negative findings, mostly from single-center data—could risk overestimating the effects attributed to AI/ML. This issue may be addressed, in part, by leveraging public registries to improve access to unpublished results. Searching trial registries, such as ClinicalTrials.gov, has been strongly recommended for comprehensive systematic reviews and is mandatory for best-practice Cochrane reviews [36]. This may help reduce publication bias and research waste. To this end, we advocate for broader use of the "results database" feature of public registries and the posting of summary results, even when submission is not mandatory.

The predominance of single-center data and studies from high-income countries has further implications, potentially contributing to existing health inequities. This underscores the need for external validation in diverse populations, especially for models built with single-center data [35]. However, the predominance of single-center designs, along with the shortage or inadequacy of external validation [23,37-39], could be a reflection of what is known as the reproducibility crisis [40-45]. AI researchers face intense pressure to publish quickly, with numerous papers posted daily on arXiv without peer review, and many are reluctant to report failed replications [43]. Unpublished code and sensitivity to training conditions make it difficult, if not impossible, to verify claimed performance [42,46-48]. Notably, papers that fail to replicate are cited at the same rate as those that are successfully replicated, leading to future work being built upon irreproducible results [44,49]. A 2018 study found that only 6% of research presented at top AI conferences explicitly identified research questions being addressed, and just 5% specified hypotheses being tested [40]. Clearly, significant challenges exist even before introducing the added complexities of multicenter studies, which involve substantial clustering (eg, across multiple centers, regions, or countries) and require more rigorous design, analysis, and reporting methods compared to standard prediction model studies [50].

Finally, publishability could improve through the wider adoption of AI/ML-specific reporting guidelines and checklists [51], including those that emphasize critical yet often undocumented details, such as computational reproducibility, data preprocessing, and mitigation of data leakage [41]. Although improved transparency through standardized reporting is the responsibility of authors, guidance such as the REFORMS: Consensus-based Recommendations for Machine-learning-based Science [41] is designed to inform all stakeholders, including readers, about best practices. All the transparency in the field will not suffice unless informed community members critically engage with the key dimensions intrinsic to AI/ML that must be addressed. As preprints gain wider visibility, the role of community literacy will grow increasingly important.

Limitations

First, to identify relevant publications, we searched for study identifiers (NCT numbers) in the title, abstract, and "trial registration" fields in PubMed, while in Scopus, we searched the title and abstract fields. Some publications may have been missed if the NCT was mentioned outside these fields. However, the International Committee of Medical Journal Editors [52] recommends listing the trial registration number at the end of the abstract, and we assumed compliance with this. We think that this was a reasonable assumption, given that publications mentioning NCTs outside abstracts tended to be cited by others or related in some way rather than being direct research outputs.

Second, we searched for publications only for studies marked "completed." We may have missed studies with relevant publications if there were delays by sponsors in updating the study status on ClinicalTrials.gov. Nevertheless, this approach also ensured that we captured publications that were likely to be the direct study output. To mitigate this potential misclassification, however, we delayed the data export until February 6, 2024, despite the data cutoff being December 31,

XSL•FO
RenderX

2023. As for publications, we searched PubMed and Scopus on February 12, 2024, allowing 6 weeks after the data cutoff to account for indexing lag time.

Third, information on the study phase may influence the extent of public disclosure of results. However, it was mostly missing (98%; 826/842), and we were unable to use it as a stratifying variable.

Conclusion

For over 80% of AI/ML studies completed during 2010-2023, study findings remained undisclosed even 3 years after study completion, raising questions about the representativeness of publicly available evidence.

While methodological rigor was generally associated with higher publication rates, the predominance of single-center designs and high-income countries may limit the generalizability of the results that are currently accessible.

Acknowledgments

The authors thank Professor Ray Rogers at the University of Adelaide and Professor Rei Goto at Keio University for their helpful suggestions. This work was in part supported by the Graduate School of Medicine, Juntendo University.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: SM
Methodology: SM
Data analysis: SM
Validation: SM
Manuscript writing: SM
Funding acquisition: RK
Project administration: RK
Data curation: MDM
Software: MDM
Data management: MDM
Supervision: RJS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 68 KB - jmir_v27i1e60148_app1.docx]

Multimedia Appendix 2
Study flow diagram.
[DOCX File , 83 KB - jmir_v27i1e60148_app2.docx]

Multimedia Appendix 3
Technical document.
[PDF File (Adobe PDF File), 568 KB - jmir_v27i1e60148_app3.pdf]

References

1. Collins GS. Making the black box more transparent: improving the reporting of artificial intelligence studies in healthcare. BMJ 2024 Apr 16;385:q832. [doi: 10.1136/bmj.q832] [Medline: 38626954]
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
3. Bruijns B, Folkertsma L, Tiggelaar R. FDA authorized molecular point-of-care SARS-CoV-2 tests: a critical review on principles, systems and clinical performances. Biosens Bioelectron X 2022;11:100158 [FREE Full text] [doi: 10.1016/j.biosx.2022.100158] [Medline: 35619623]

XSL•FO
RenderX

4. Dhiman P, Ma J, Navarro CA, Speich B, Bullock G, Damen JA, et al. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. J Clin Epidemiol 2021;138:60-72 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.06.024] [Medline: 34214626]
5. Logullo P, MacCarthy A, Dhiman P, Kirtley S, Ma J, Bullock G, et al. Artificial intelligence in lung cancer diagnostic imaging: a review of the reporting and conduct of research published 2018-2019. BJR Open 2023;5(1):20220033 [FREE Full text] [doi: 10.1259/bjro.20220033] [Medline: 37389003]
6. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020;368:l6927 [FREE Full text] [doi: 10.1136/bmj.l6927] [Medline: 32198138]
7. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JA, et al. Overinterpretation of findings in machine learning prediction model studies in oncology: a systematic review. J Clin Epidemiol 2023;157:120-133 [FREE Full text] [doi: 10.1016/j.jclinepi.2023.03.012] [Medline: 36935090]
8. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JAA, et al. Risk of bias of prognostic models developed using machine learning: a systematic review in oncology. Diagn Progn Res 2022;6(1):13 [FREE Full text] [doi: 10.1186/s41512-022-00126-w] [Medline: 35794668]
9. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ 2021;375:n2281 [FREE Full text] [doi: 10.1136/bmj.n2281] [Medline: 34670780]
10. Lock S, Wells F. Preface to the second edition. In: Lock S, Wells F, editors. Fraud and Misconduct in Medical Research. London, UK: BMJ Publishing Group; 1996:xi-xii.
11. Antes G, Chalmers I. Under-reporting of clinical trials is unethical. Lancet 2003 Mar 22;361(9362):978-979. [doi: 10.1016/S0140-6736(03)12838-3] [Medline: 12660049]
12. Pich J, Carné X, Arnaiz JA, Gómez B, Trilla A, Rodés J. Role of a research ethics committee in follow-up and publication of results. Lancet 2003 Mar 22;361(9362):1015-1016. [doi: 10.1016/S0140-6736(03)12799-7] [Medline: 12660062]
13. Liu X, Zhang Y, Li WF, Vokes E, Sun Y, Le QT, et al. Evaluation of oncology trial results reporting over a 10-year period. JAMA Netw Open 2021;4(5):e2110438 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.10438] [Medline: 34028549]
14. Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. BMJ 2016;352:i637 [FREE Full text] [doi: 10.1136/bmj.i637] [Medline: 26888209]
15. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. BMJ 2012;344:d7292 [FREE Full text] [doi: 10.1136/bmj.d7292] [Medline: 22214755]
16. Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. PLoS Med 2013;10(12):e1001566; discussion e1001566 [FREE Full text] [doi: 10.1371/journal.pmed.1001566] [Medline: 24311990]
17. Saito H, Gill CJ. How frequently do the results from completed US clinical trials enter the public domain?--a statistical analysis of the clinical Trials.gov database. PLoS One 2014;9(7):e101826 [FREE Full text] [doi: 10.1371/journal.pone.0101826] [Medline: 25025477]
18. Zwierzyna M, Davies M, Hingorani AD, Hunter J. Clinical trial design and dissemination: comprehensive analysis of clinicaltrials.gov and PubMed data since 2005. BMJ 2018;361:k2130 [FREE Full text] [doi: 10.1136/bmj.k2130] [Medline: 29875212]
19. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.Gov: a cross-sectional analysis. PLoS Med 2009 Sep;6(9):e1000144 [FREE Full text] [doi: 10.1371/journal.pmed.1000144] [Medline: 19901971]
20. Celi LA, Cellini J, Charpignon M, Dee EC, Dernoncourt F, Eber R, for MIT Critical Data. Sources of bias in artificial intelligence that perpetuate healthcare disparities-a global review. PLOS Digit Health 2022;1(3):e0000022 [FREE Full text] [doi: 10.1371/journal.pdig.0000022] [Medline: 36812532]
21. AACT Database | Clinical Trials Transformation Initiative. CTTI. URL: https://aact.ctti-clinicaltrials.org/ [accessed 2025-02-08]
22. AACT data dictionary. CTTI. URL: https://aact.ctti-clinicaltrials.org/data_dictionary [accessed 2024-03-08]
23. Maru S, Matthias MD, Kuwatsuru R, Simpson RJ. Studies of artificial intelligence/machine learning registered on ClinicalTrials.gov: cross-sectional study with temporal trends, 2010-2023. J Med Internet Res 2024;26:e57750 [FREE Full text] [doi: 10.2196/57750] [Medline: 39454187]
24. Protocol registration data element definitions for interventional and observational studies. National Library of Medicine. URL: https://clinicaltrials.gov/policy/protocol-definitions [accessed 2024-01-09]
25. How does the World Bank classify countries? World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/378834-how-does-the-world-bank-classify-countries [accessed 2024-03-01]
26. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]

XSL•FO
RenderX

27. Minen MT, Reichel JF, Pemmireddy P, Loder E, Torous J. Characteristics of neuropsychiatric mobile health trials: cross-sectional analysis of studies registered on ClinicalTrials.gov. JMIR Mhealth Uhealth 2020;8(8):e16180 [FREE Full text] [doi: 10.2196/16180] [Medline: 32749230]
28. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of digital health studies registered in ClinicalTrials.gov. JAMA Intern Med 2019;179(6):838-840. [doi: 10.1001/jamainternmed.2018.7235] [Medline: 30801617]
29. Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, CLAIM 2024 Update Panel. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. Radiol Artif Intell 2024;6(4):e240300. [doi: 10.1148/ryai.240300] [Medline: 38809149]
30. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. J Med Internet Res 2023;25:e48763 [FREE Full text] [doi: 10.2196/48763] [Medline: 37651179]
31. Lu J, Xu BB, Shen LL, Wu D, Xue Z, Zheng H, et al. Characteristics and research waste among randomized clinical trials in gastric cancer. JAMA Netw Open 2021;4(9):e2124760 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.24760] [Medline: 34533573]
32. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. PLoS Med 2008;5(9):e191 [FREE Full text] [doi: 10.1371/journal.pmed.0050191] [Medline: 18816163]
33. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health 2024 May;6(5):e367-e373 [FREE Full text] [doi: 10.1016/S2589-7500(24)00047-5] [Medline: 38670745]
34. Lam TYT, Cheung MFK, Munro YL, Lim KM, Shung D, Sung JJY. Randomized controlled trials of artificial intelligence in clinical practice: systematic review. J Med Internet Res 2022;24(8):e37188 [FREE Full text] [doi: 10.2196/37188] [Medline: 35904087]
35. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022;24(10):e40238 [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
36. Hunter KE, Webster AC, Page MJ, Willson M, McDonald S, Berber S, et al. Searching clinical trials registers: guide for systematic reviewers. BMJ 2022;377:e068791. [doi: 10.1136/bmj-2021-068791] [Medline: 35473822]
37. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297 [FREE Full text] [doi: 10.1016/S2589-7500(19)30123-2] [Medline: 33323251]
38. Vedula SS, Tsou BC, Sikder S. Artificial intelligence in clinical practice is here-now what? JAMA Ophthalmol 2022;140(4):306-307. [doi: 10.1001/jamaophthalmol.2022.0040] [Medline: 35238882]
39. d'Elia A, Gabbay M, Rodgers S, Kierans C, Jones E, Durrani I, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. Fam Med Community Health 2022;10(Suppl 1):e001670 [FREE Full text] [doi: 10.1136/fmch-2022-001670] [Medline: 36450391]
40. Gundersen OE. Improving reproducibility of artificial intelligence research to increase trust and productivity. In: Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research. Paris, France: OECD Publishing; 2023.
41. Kapoor S, Cantrell EM, Peng K, Pham TH, Bail CA, Gundersen OE, et al. REFORMS: consensus-based recommendations for machine-learning-based science. Sci Adv 2024;10(18):eadk3452 [FREE Full text] [doi: 10.1126/sciadv.adk3452] [Medline: 38691601]
42. Hutson M. Artificial intelligence faces reproducibility crisis. Science 2018;359(6377):725-726. [doi: 10.1126/science.359.6377.725] [Medline: 29449469]
43. Ball P. Is AI leading to a reproducibility crisis in science? Nature 2023;624(7990):22-25. [doi: 10.1038/d41586-023-03817-6] [Medline: 38052897]
44. Straiton J. Artificial intelligence: help or hindrance in solving the reproducibility crisis? Biotechniques 2024;76(7):291-294 [FREE Full text] [doi: 10.1080/07366205.2024.2355776] [Medline: 38899492]
45. Nickson D, Singmann H, Meyer C, Toro C, Walasek L. Replicability and reproducibility of predictive models for diagnosis of depression among young adults using electronic health records. Diagn Progn Res 2023;7(1):25 [FREE Full text] [doi: 10.1186/s41512-023-00160-2] [Medline: 38049919]
46. Rahrooh A, Garlid AO, Bartlett K, Coons W, Petousis P, Hsu W, et al. Towards a framework for interoperability and reproducibility of predictive models. J Biomed Inform 2024;149:104551 [FREE Full text] [doi: 10.1016/j.jbi.2023.104551] [Medline: 38000765]
47. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Waldron L, Wang B, et al. Transparency and reproducibility in artificial intelligence. Nature 2020;586(7829):E14-E16 [FREE Full text] [doi: 10.1038/s41586-020-2766-y] [Medline: 33057217]
48. Moassefi M, Rouzrokh P, Conte GM, Vahdati S, Fu T, Tahmasebi A, et al. Reproducibility of deep learning algorithms developed for medical imaging analysis: a systematic review. J Digit Imaging 2023;36(5):2306-2312. [doi: 10.1007/s10278-023-00870-5] [Medline: 37407841]

XSL•FO
RenderX

49. Yang Y, Youyou W, Uzzi B. Estimating the deep replicability of scientific findings using human and artificial intelligence. Proc Natl Acad Sci U S A 2020;117(20):10762-10768 [FREE Full text] [doi: 10.1073/pnas.1909046117] [Medline: 32366645]
50. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. BMJ 2023;380:e071018 [FREE Full text] [doi: 10.1136/bmj-2022-071018] [Medline: 36750242]
51. Reporting guidelines - artificial intelligence/machine learning studies. EQUATOR Network. URL: https://www.equator-network.org/reporting-guidelines-study-design/artificial-intelligence-machine-learning-studies/ [accessed 2025-01-05]
52. Clinical Trials. International Committee of Medical Journal Editors. URL: https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html [accessed 2024-06-01]

Abbreviations

AI/ML: artificial intelligence/machine learning
API: application programming interface
CLAIM: Checklist for Artificial Intelligence in Medical Imaging
CTTI AACT: Clinical Trials Transformation Initiative Aggregate Analysis of ClinicalTrials.gov
FDA: Food and Drug Administration
GNI: gross national income
MeSH: Medical Subject Headings
NCT: National Clinical Trial
RCT: randomized controlled trial
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by G Tsafnat; submitted 30.07.24; peer-reviewed by CA Brunette, G Chennekkattu Markose, V Gejjegondanahalli Yogeshappa; comments to author 13.12.24; revised version received 13.01.25; accepted 18.01.25; published 21.03.25.

Please cite as:
Maru S, Kuwatsuru R, Matthias MD, Simpson Jr RJ
Public Disclosure of Results From Artificial Intelligence/Machine Learning Research in Health Care: Comprehensive Analysis of ClinicalTrials.gov, PubMed, and Scopus Data (2010-2023)
J Med Internet Res 2025;27:e60148
URL: https://www.jmir.org/2025/1/e60148
doi:10.2196/60148
PMID:

©Shoko Maru, Ryohei Kuwatsuru, Michael D Matthias, Ross J Simpson Jr. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Explainable AI for Intraoperative Motor-Evoked Potential Muscle Classification in Neurosurgery: Bicentric Retrospective Study

Qendresa Parduzi[1,2,3*], MSc; Jonathan Wermelinger[2*], PhD; Simon Domingo Koller[4], BSc; Murat Sariyar[4], PhD; Ulf Schneider[3], Prof Dr Med; Andreas Raabe[2], Prof Dr Med; Kathleen Seidel[2], Prof Dr Med

[1]Graduate School for Health Sciences, University of Bern, Bern, Switzerland

[2]Department of Neurosurgery, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[3]Department of Neurosurgery, Lucerne Cantonal Hospital, Lucerne, Switzerland

[4]School of Engineering and Computer Science, Bern University of Applied Sciences, Biel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Qendresa Parduzi, MSc
Department of Neurosurgery
Lucerne Cantonal Hospital
Spitalstrasse
Lucerne, 6000
Switzerland
Phone: 41 412056631
Email: qendresa.parduzi@students.unibe.ch

Abstract

Background: Intraoperative neurophysiological monitoring (IONM) guides the surgeon in ensuring motor pathway integrity during high-risk neurosurgical and orthopedic procedures. Although motor-evoked potentials (MEPs) are valuable for predicting motor outcomes, the key features of predictive signals are not well understood, and standardized warning criteria are lacking. Developing a muscle identification prediction model could increase patient safety while allowing the exploration of relevant features for the task.

Objective: The aim of this study is to expand the development of machine learning (ML) methods for muscle classification and evaluate them in a bicentric setup. Further, we aim to identify key features of MEP signals that contribute to accurate muscle classification using explainable artificial intelligence (XAI) techniques.

Methods: This study used ML and deep learning models, specifically random forest (RF) classifiers and convolutional neural networks (CNNs), to classify MEP signals from routine supratentorial neurosurgical procedures from two medical centers according to muscle identity of four muscles (extensor digitorum, abductor pollicis brevis, tibialis anterior, and abductor hallucis). The algorithms were trained and validated on a total of 36,992 MEPs from 151 surgeries in one center, and they were tested on 24,298 MEPs from 58 surgeries from the other center. Depending on the algorithm, time-series, feature-engineered, and time-frequency representations of the MEP data were used. XAI techniques, specifically Shapley Additive Explanation (SHAP) values and gradient class activation maps (Grad-CAM), were implemented to identify important signal features.

Results: High classification accuracy was achieved with the RF classifier, reaching 87.9% accuracy on the validation set and 80% accuracy on the test set. The 1D- and 2D-CNNs demonstrated comparably strong performance. Our XAI findings indicate that frequency components and peak latencies are crucial for accurate MEP classification, providing insights that could inform intraoperative warning criteria.

Conclusions: This study demonstrates the effectiveness of ML techniques and the importance of XAI in enhancing trust in and reliability of artificial intelligence–driven IONM applications. Further, it may help to identify new intrinsic features of MEP signals so far overlooked in conventional warning criteria. By reducing the risk of muscle mislabeling and by providing the basis for possible new warning criteria, this study may help to increase patient safety during surgical procedures.

(J Med Internet Res 2025;27:e63937) doi:10.2196/63937

XSL•FO
RenderX

KEYWORDS

intraoperative neuromonitoring; motor evoked potential; artificial intelligence; machine learning; deep learning; random forest; convolutional neural network; explainability; medical informatics; personalized medicine; neurophysiological; monitoring; orthopedic; motor; neurosurgery

Introduction

The importance of intraoperative neurophysiological monitoring (IONM) during high-risk neurosurgical and orthopedic procedures has been established in recent decades [1]. In particular, the monitoring of motor evoked potentials (MEPs) helps to assess the functional integrity of motor pathways during surgeries and allows postoperative motor outcomes to be predicted [2-9]. However, the features of this complex signal that contribute to its predictive potential are still poorly understood and there are few standardized warning criteria to alert the surgeon. Currently, the best-established and most reliable MEP warning criterion during IONM is a 50% drop in amplitude [10].

Because changes in MEP amplitude are predictive of postoperative motor outcome, it is natural to ask whether other properties of the signals could be important for decision-making. The emergence of machine learning (ML) methods has led to an interest in leveraging these techniques to classify MEPs in the hope of improving intraoperative decision-making [11,12]. However, most of the studies so far have focused on identifying the most accurate and robust ML algorithms rather than on uncovering the underlying patterns leading to the decisions.

In our previous work, we established prediction algorithms for muscle identification to provide a proof of principle within a solid ground truth framework before translating them to outcome predictions [12]. Meanwhile, Boaro et al [13] implemented a similar classification task with additional ML models and muscles. The robustness of ML algorithms on clinical data needs to be established using independent data sources, which is why we have expanded our data set to include signals from an external validation center.

To gain a deeper understanding of our signals, we investigated them in both the time-series and time-frequency domains, which have been shown to be useful in the quantification of disease-related MEP changes [14]. In addition to the standard ML models, we used deep learning methods to leverage their power of internal feature representation. Although these algorithms can accurately predict the identity of muscles based on MEP signals, the specific criteria that these algorithms use to make their predictions are not well understood [15-17]. For research in the field of IONM, this explanatory information is probably at least as important as the predictions themselves, as it can provide new insights into the mechanisms of neurophysiological changes. For this reason, we used methods from the emerging field of explainable artificial intelligence (XAI) [18,19]. The aim was to combine methods to ensure comprehensive interpretability of our different models' decisions.

Our study provides a robust framework for classifying MEPs recorded in routine neurosurgical procedures according to their muscle identity with high accuracy and we validated the methods using data from two independent study centers. Importantly, we elucidate the decision-making processes of our ML models through post hoc analyses, thereby enabling their effective application to previously unseen data and novel situations. These algorithms could act as a safety mechanism in the operating room by detecting mislabeling of muscles and by focusing on new intrinsic features of MEPs. Thus, they may enhance the usage and acceptance of artificial intelligence (AI) in medical decision-making through their interpretability.

Methods

Ethical Considerations

This retrospective study was approved by the cantonal ethics committee of Bern, Switzerland (BASEC-ID 2023–00277). All included patients gave their informed written consent for further use and publication of their anonymized data. The routinely collected data are dated from 2018 to 2022 and coded for the analysis of this paper. Only patients over the age of 18 were included in the study, all received neurosurgical interventions and were not stratified according to their clinical outcome since the outcome prediction of the study was muscle identification. The muscle recordings are set routinely independent of sociodemographic or clinical factors.

In the following, we describe the datasets collected, the preselection approach applied to the data, the methods used for signal data representation, and the analytical techniques used, including XAI approaches to elucidate MEP feature importance.

MEP Data and Signal Recordings

The MEP data used in this study were obtained during routine neurosurgical procedures and were retrospectively collected and analyzed. This study was exploratory in nature, and no formal protocol was prepared. Recordings from 151 surgeries on 144 patients at one center (Inselspital, University Hospital Bern, Switzerland), "center T_1" were used for training and validation, and recordings from 58 surgeries on 57 patients at an independent center (Cantonal Hospital in Lucerne, Switzerland), "center T_2" were used for testing (see Figure 1A). In total, there were 94 females and 107 males and the median age at surgery was 61 years (see Table 1 for distribution between centers). Overall, 182 patients underwent surgery for a brain tumor and 19 for vascular pathologies. The total number of MEP signals was 36,992 for center T_1 and 24,298 for center T_2, with at least 3000 samples for each predicted class (see Table 1 for detailed information on classes). This sample size was determined by the number of routine interventions and ensures that any complexity of the ML models used can appropriately be trained, tested, and validated.

XSL•FO
RenderX

Figure 1. Data analysis pipeline. (A) Bicentric training, validation, and testing setup. (B) Data representation and the algorithms used on each representation. CNN: convolutional neural network.

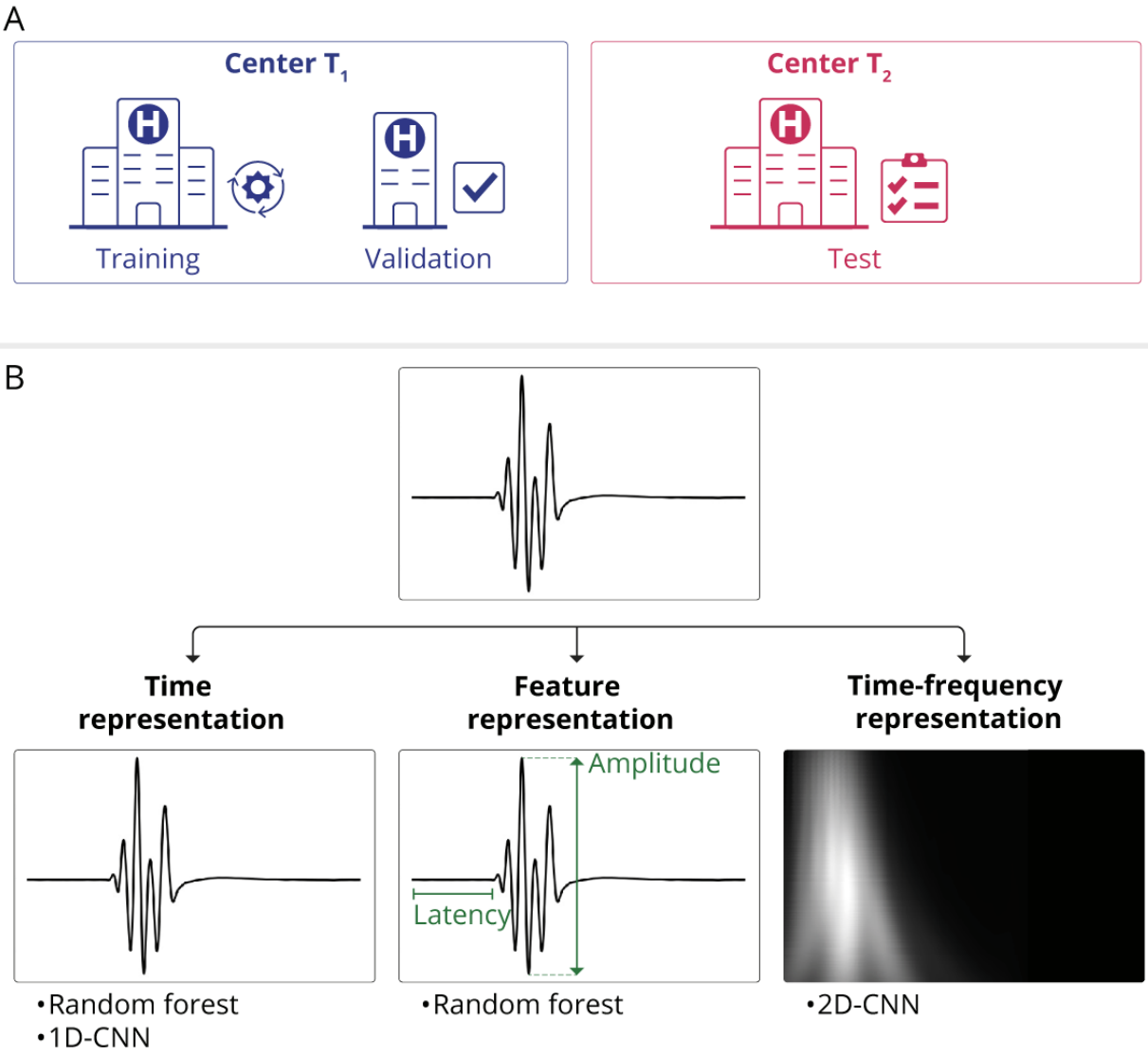

IONM was performed according to a standardized protocol, as previously described [8,20]. The MEPs were elicited either through transcranial electric stimulation (TES) via corkscrew electrodes or direct cortical stimulation (DCS) via strip electrodes placed directly on the cortex and recorded via needle electrodes in the muscle belly. Stimulation to elicit MEPs was conducted under general anesthesia using a train of 5 anodal stimuli with a pulse duration of 0.5 milliseconds, and an interstimulus interval of 4 milliseconds, known as the short train method. Both centers used ISIS Systems to record the MEPs. The sampling frequency was set at 20 kHz, with hardware high- and low-pass filters at 30 Hz and 5 kHz, respectively. TES was used in all the surgeries performed in center T_1 and in 95% (55 of 58) of those performed in center T_2 to elicit MEPs. In addition, DCS was used to elicit MEPs in 134 of the 151 (89%) surgeries performed in center T_1 and 17 of 58 (30%) surgeries performed in center T_2. The recordings consist of 2000 data points, corresponding to 100-millisecond windows for each signal. We selected MEP signals from the following 4 muscles: Extensor digitorum (EXT), abductor pollicis brevis (APB), tibialis anterior (TA), and abductor hallucis (AH). These muscles are routinely monitored during supratentorial surgeries, and the corresponding signals were available for most of the included patients from both sides. As the neurophysiologist labels the recording channels at the start of the surgery, the MEP data are automatically labeled upon saving.

We used custom-made Python 3.0 scripts for all the data analysis and classification tasks.

XSL•FO
RenderX

Table 1. Surgery, patient, and recording characteristics. The percentage of clinical outcome is calculated in relation to the total number of patients, while the percentage of TES and DCS stimulations is calculated with respect to the total number of surgeries (including redo operations).

Categories	Center T_1	Center T_2
Demography, n		
Patients	144	57
M[a]	78	29
F[b]	66	28
Age[c]	62	58
Pathology, n		
Meningioma	0	8
Schwannoma	0	2
Oligodendroglioma	7	2
Astrocytoma	16	10
Glioblastoma	73	10
Metastasis	44	4
Aneurysm	0	13
AVM[d]	1	4
Cavernoma	2	1
Trigeminal neuralgia	0	3
Radio necrosis	1	0
Clinical outcome, n (%)		
Deficits at discharge	29 (19)	9 (16)
Deficits at follow-up	9 (6)	4 (7)
Neurophysiology		
Surgeries, n	151	58
TES[e] stimulation, n (%)	151 (100)	55 (95)
DCS[f] stimulation, n (%)	134 (89)	17 (30)
MEP[g] signals, n	36,992	24,298
EXT[h] signals, n (%)	11,958 (31.8)	3628 (14.5)
APB[i] signals, n (%)	15,800 (42.9)	10,670 (42.6)
TA[j] signals, n (%)	5773 (15.7)	4970 (21.4)
AH[k] signals, n (%)	3461 (9.6)	5030 (21.6)

[a]M: male.
[b]F: female.
[c]Age is the median age of all patients.
[d]AVM: arteriovenous malformation.
[e]TES: transcranial electric stimulation.
[f]DCS: direct cortical stimulation.
[g]MEP: motor-evoked potential.
[h]EXT: extensor digitorum.
[i]APB: abductor pollicis brevis.
[j]TA: tibialis anterior.
[k]AH: abductor hallucis.

XSL•FO
RenderX

Preprocessing and Data Representation

Preselection

An automatic MEP selection algorithm was written to determine whether a given recording contained an MEP [12]. To remove stimulation artifacts from the train of 5, we excluded the first 400 data points (corresponding to 20 milliseconds). Then, two features were computed: the onset latency and duration of the signal. Onset latency is defined as 1 millisecond (empirically determined) before the trace crosses the mean of the baseline (the last 5 milliseconds of the recording) plus or minus the SD of the entire recording. The end of the signal was calculated similarly, by starting from the end of the recording. We defined the duration as the end of signal latency minus onset latency. In addition, the time interval between the first and last peak was determined using the scipy.signal function find_peaks. A recording was considered to contain an MEP if at least one peak was detected, the duration was less than 40 milliseconds, and the interval from the first to last peak was less than 35 milliseconds (in accordance with clinical experience).

In our analysis pipeline, we used three distinct representations of MEP data (see Figure 1B): time, feature, and time-frequency, each tailored to optimize the performance of our ML classifiers.

Time Representation

A finite-impulse response bandpass filter with 30- and 1000-Hz cutoff frequencies was applied to the 1600-dimensional signal vector. These frequency settings align with MEP visualization practices using the monitoring machine at center T_1 (ie, the software filters). The signal vectors are then normalized with respect to the absolute maximum MEP value in each patient. This filtered and normalized time representation of the data was used to train, validate, and test a random forest (RF) classifier, as well as a 1-dimensional convolutional neural network (1D-CNN).

Feature Representation

We used a customized feature extraction algorithm to condense each filtered and normalized MEP signal into characteristic features. The initial choice of predictors is a combination of clinically used predictors (eg, latency, amplitude, minimum, and maximum) and routinely used features in general neurophysiological literature (spectral entropy, frequency, etc). After a correlation analysis (see Multimedia Appendix 1), we chose five features that showed no correlation describing relevant domains of the signal: peak latency, maximum signal value, number of peaks, main frequency, and slope.

Peak values were extracted with the scipy find_peaks function with peak prominence defined as twice the SD of the signal. The main frequency was calculated as the frequency at which the Fourier transform of the filtered data attained its maximum absolute value. Finally, the slope of the signal was computed as the mean of the first derivative of the signal (using the NumPy function gradient). The resulting 5-dimensional feature representation of the data was then used to train, validate, and test a RF classifier.

Time-Frequency Representation

Finally, employing the Python library PyWavelets (pywt), a continuous wavelet transform with a Mexican hat mother wavelet was applied to the data to transform them into 2D time-frequency representations. Scales ranging from 2 to 30 were logarithmically spaced, while the time dimension was undersampled to yield an array of dimensions 224×224. These 2D time-frequency representations were then used to train, validate, and test a 2D-CNN.

Statistical Testing

Using custom Python scripts, Student *t* tests were applied to compare the mean values of two different features. Specifically, for each muscle, the means of the different features were compared across the two centers. Statistical significance was set at P=.05.

Machine and Deep Learning Pipeline

The Python library scikit-learn [21] was used for the RF classifier, while tensorflow [22] and keras [23] were used to obtain the 1D- and 2D-CNN models. Hyperparameter tuning was carried out for the RF and 1D-CNN, while we used fixed parameters for the 2D-CNN (see below). The hyperparameters used in the grid search of the RF are the same as described in Multimedia Appendix 1. The architecture of our 1D-CNN is inspired by the model of Ahmed et al [24], and consists of two consecutive 1D convolutional layers, followed by a MaxPooling layer, a dropout layer, and a BatchNormalization layer. The model then gets flattened, before adding a dense layer and another dropout layer and finally ending in a dense output layer. The structure of our 2D-CNN is inspired by the model of Wang et al [25]. It consists of 4 blocks of 2D convolutional layers followed by BatchNormalization layers, with a MaxPooling layer after the second and third blocks and a GlobalMaxPooling layer after the fourth block. The model is capped off by a dense output layer. The specifics of these models are shown in Multimedia Appendix 1.

The dataset from Center T_1 was split into 70% for training and 30% for validation (stratified according to patients), while the whole of the dataset from Center T_2 was used for testing. In all cases, we used class weighting [26] to deal with the class imbalance problem (ie, the number of leg muscle MEPs is lower than the number of arm muscle MEPs, see Wermelinger et al [12] for a discussion of this issue).

Model Output and Outcome

All prediction models output probabilities of belonging to the predicted class. The decision thresholds were set at a chance level of 0.25 (likelihood of belonging to 1 of 4 classes of muscles). They were systematically explored and reported (see Figure 2B) for decision thresholds up to 0.9.

XSL•FO
RenderX

Figure 2. Classification results and confidence. (A) Validation accuracy (center T1, white) and test accuracy (center T2, colored) of all models. (B) Decision confidence. Solid lines are the accuracies (left y-axis) of the various models for different confidence thresholds. The dashed lines show the proportion of data with these confidences (right y-axis). RFs have a higher increase in accuracy compared with CNNs but at a higher data cost. (C) Bicentric confusion matrices: lower triangle (center T2), upper triangle (center T1) for both RF on feature representation (top) and 2D-CNN (bottom). The RF is slightly more congruent across centers than the 2D-CNN. CNN: convolutional neural network; RF: random forest.

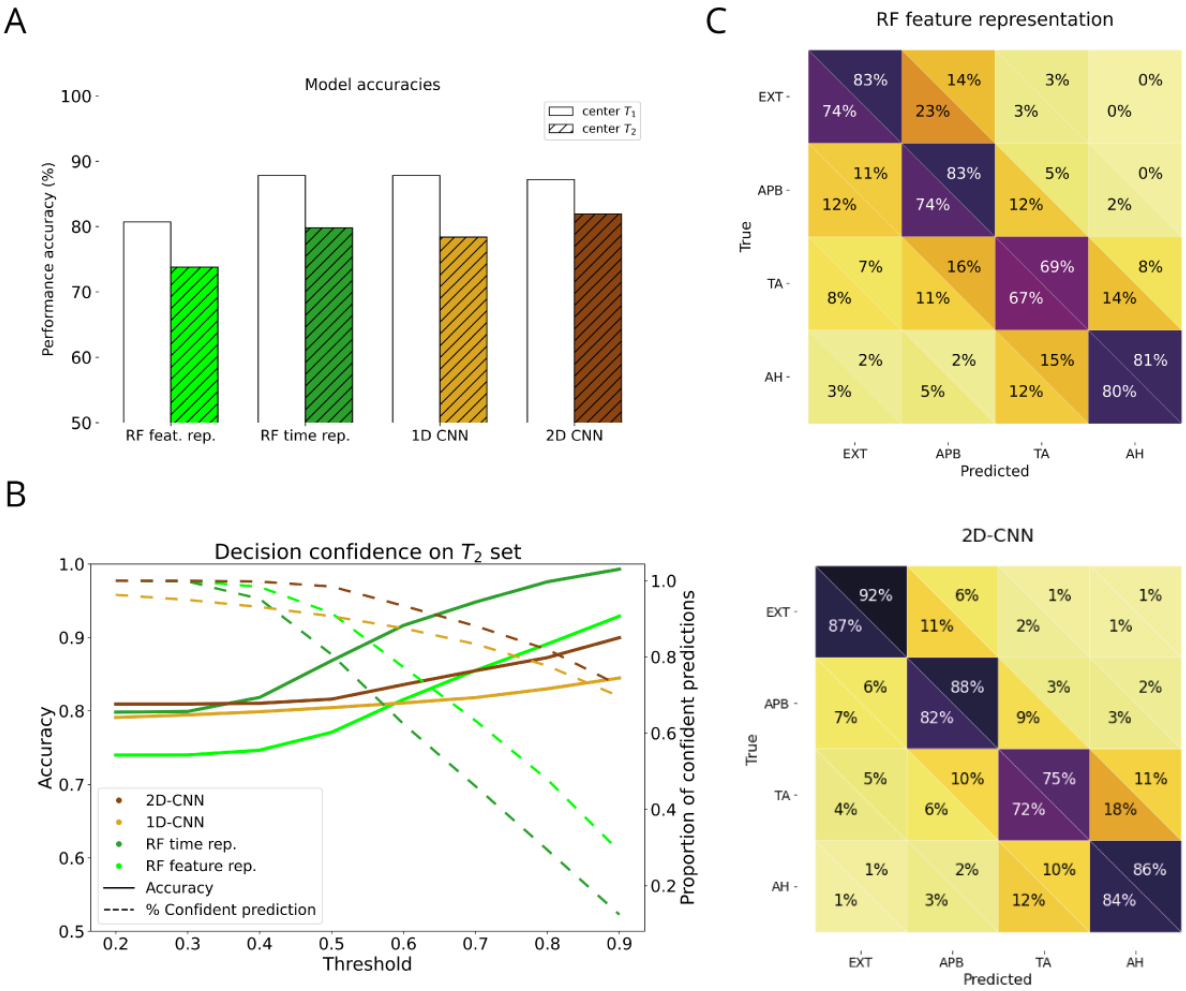

Accuracy was used as the primary performance metric, and the confusion matrix was used to evaluate the performance of the classification algorithm. The outcome assessment does not require subjective interpretation, since the muscle identity is objectively assessable, independent of sociodemographic background and clinical outcome.

No model updating or recalibration was performed during the model evaluation. While some variability in model performance was observed across different centers (Figure 2), we opted to retain the original model without adjustments. Future work may explore model updating to enhance performance in these areas.

Explainability

To elucidate how the RF classifiers made their decisions, we used feature importance and Shapley Additive Explanation (SHAP) values. Feature importance values are provided by the feature_importances_ attribute of the scikit-learn RandomForestClassifier class, while the SHAP values are calculated with the SHAP library [27]. The feature importance values are determined by aggregating (mean and SD) the impurity decrease within each decision tree. SHAP values quantify the impact of each feature on prediction outcomes. Positive values signify a positive influence, while negative values indicate the opposite, with magnitude representing the strength of the effect. SHAP values of the RF on feature representation of the MEP data were computed on a random sample containing 20% (5919 MEPs) of the training data set. In the case of CNNs, we used gradient-weighted class activation mapping (Grad-CAM). This is a type of attention map, a visualization tool highlighting regions within an image considered by the neural network to be pivotal for specific predictions [28]. We adapted preexisting code to generate Grad-CAMs for both 1D- and 2D-CNNs [29,30]. The Grad-CAMs of all signals in the training data set were averaged to obtain the corresponding plots for the 1D-CNN (overall) and 2D-CNN (for each muscle).

Results

Differences and Similarities of MEP Properties Between Centers

Data from 151 surgeries from the training and validation center T_1 yielded a total of 36,992 MEPs (11,958 EXT, 15,800 APB, 5773 TA, and 3461 AH), and the data from 58 surgeries from

XSL•FO
RenderX

the test center T_2 yielded a total of 24,298 MEPs (3628 EXT, 10,670 APB, 4970 TA, and 5030 AH; Table 1).

The distribution of muscle recordings is illustrated in Table 1. A notable discrepancy was observed in the proportion of MEPs recorded from the upper extremities between centers T_1 (75%, 27,758/36,992) and T_2 (58.8%, 14,298/24,298). This variation stems from differences in surgical procedures, montage standards, and stimulation techniques. As shown in Table 1, the proportion of MEPs elicited via DCS was significantly higher in center T_1 (89%, 134/151) than in center T_2 (30%, 17/58). One striking difference in the MEP data was the significantly shorter onset latencies across all muscles for center T_1 (P<.001; Figure 3B). It is crucial to consider these disparities when interpreting the classification results. Furthermore, Figure 3C reveals differences in main frequency patterns between proximal muscles (EXT and TA) and distal muscles (APB and AH) for both centers, adding another layer of complexity to the MEP data analysis.

Figure 3. MEP properties across the 2 centers. (A) Latency distribution for both centers and for each muscle. The latencies of all muscles are significantly shorter at center T1. (B) Main frequency distribution for both centers and each muscle. At both centers, the distal muscles (APB and AH) exhibit a higher main frequency than the proximal muscles (EXT and TA). AH: abductor hallucis; APB: abductor pollicis brevis; EXT: extensor digitorum; TA: tibialis anterior.

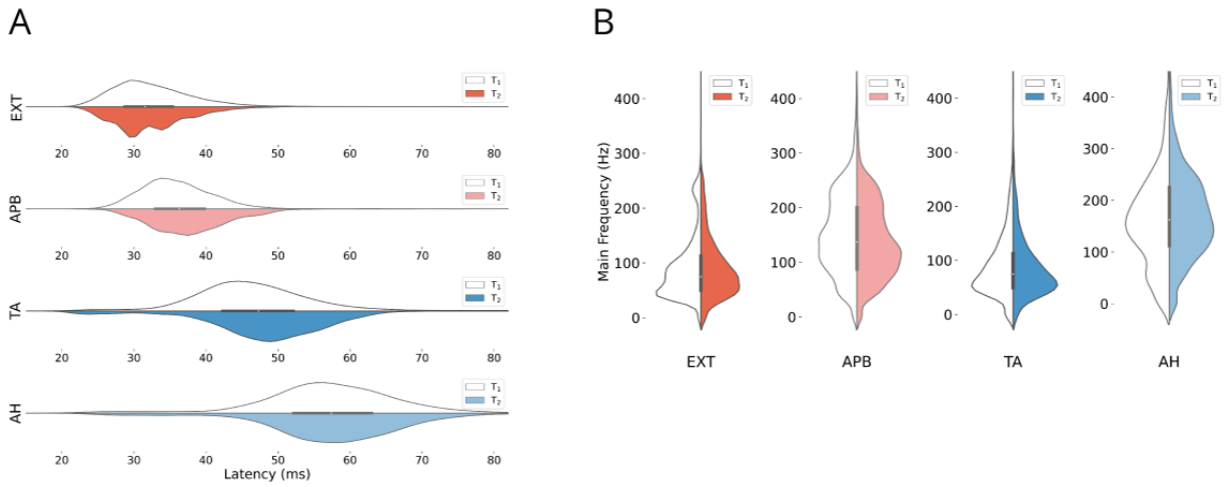

With time representation of the MEPs, the RF classifier achieved 87.9% accuracy on the validation dataset from center T_1 and 80% on the test set from center T_2. The 1D-CNN achieved a validation accuracy of 87.8% and a test accuracy of 78.4%. On the feature representation, the RF achieved 80.3% validation accuracy and 74.5% test accuracy. Finally, the 2D-CNN achieved 87.2% validation accuracy and 81.9% test accuracy on the time-frequency representation of the MEPs (see Figure 2A). Examination of the confusion matrices (Figure 2C and Multimedia Appendix 2) revealed subtle variations in decision-making patterns across muscles for different data representations and models. Generally, the more high-dimensional data input (time and time-frequency representation) performed better at the classification task than the feature representation of the data. Comparing the performances overall muscles between the RF on the feature representation and the 2D-CNN on the time-frequency representation, the former has a lower overall accuracy than the latter. However, when evaluating performance consistency across muscles, the feature representation demonstrated slightly less variability, evidenced by a lower SD (6.44% for RF on feature representation vs 7.3% for 2D-CNN) and coefficient of variation (8.15 vs 8.56, respectively), of performances.

Evaluating Decision Confidence

In intraoperative clinical settings, the confidence with which decisions are made is important. To assess this aspect in our classification task, we examined the confidence of our algorithms in categorizing muscles. Figure 2B illustrates the relationship between the proportion of confident predictions (those meeting a specific confidence threshold) and the corresponding test accuracy (see also Multimedia Appendix 2). For a four-class problem, the chance level is 25%. Notably, our models consistently outperform this baseline, showing robust performance. However, as confidence thresholds increase, the proportion of data that meet these criteria diminishes, albeit resulting in enhanced accuracy. The RF models incur a proportionally higher data cost for achieving this accuracy enhancement compared with CNNs.

Insights Into Model Decision-Making

Explicit Feature Representation

Although algorithms using explicit feature representations showed poorer performance (80% test accuracy), they provided key insights into the decision-making process. Feature importance analysis (Figure 4A) revealed that peak latency, a primary factor in clinical decision-making, was the main driver of classification. Main frequency was the next most important, despite typically not being used by clinicians. This was followed in order of importance by maximum signal value, while slope and number of peaks were the least important features.

XSL•FO
RenderX

Figure 4. Model decisions according to muscle classification. (A) SHAP feature importances of the RF on feature representation. (B) Beeswarm plot of SHAP values for each muscle classification of the RF on data with feature representation. The features are ordered by importance (top to bottom). Feature values are color-coded (black: high value, yellow: low value). Being on the right (positive SHAP values) means that the feature contributes to the (not necessarily correct) prediction of that particular class. Latency is the most important feature in this situation, with short latencies indicating upper extremity muscles, and long latencies lower extremity muscles. The second most important feature is the main frequency, with high frequencies leading to a decision toward distal muscles (APB and AH), whereas low frequencies push the decision toward proximal muscles (EXT and TA). AH: abductor hallucis; APB: abductor pollicis brevis; EXT: extensor digitorum; SHAP: Shapley Additive Explanation; TA: tibialis anterior.

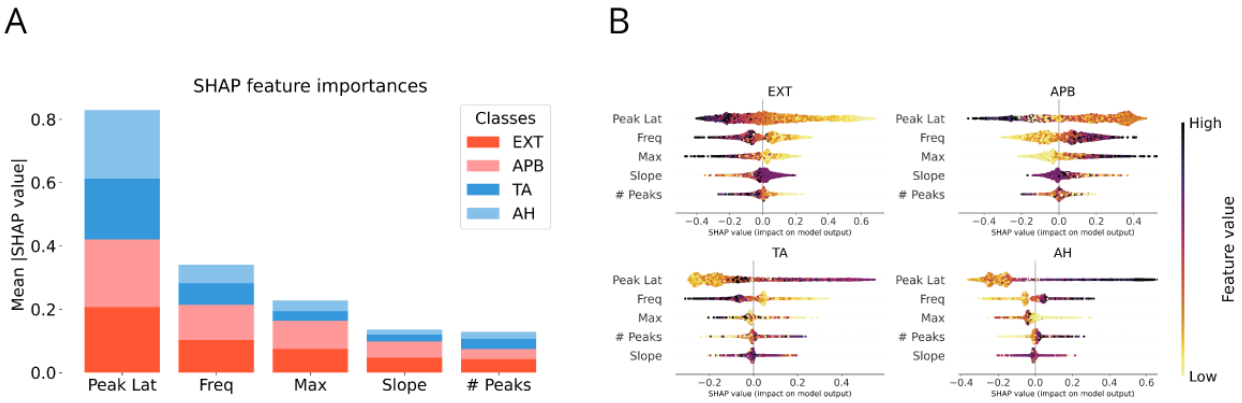

SHAP values from the RF model using feature representation elucidated how different parameters influenced the model's decisions for each class (Figure 4B). In all four muscles, peak latencies were crucial. For the upper extremity, short latencies favored correct predictions, whereas long latencies were accurate indicators for the lower extremity classes. Interestingly, the main frequencies did not exhibit this pattern. High main frequency values favored predictions for distal muscles (APB and AH), whereas low frequencies were associated with proximal muscles (EXT and TA).

Implicit Feature Representation in CNNs

When presented with complex data, such as time-series and wavelet transforms, ML algorithms use internal feature representations to guide their decision-making. For CNNs, these features can be visualized using attention maps, which highlight the areas of the input data that are most salient and decisive for classification. In the depicted Grad-CAMs, these areas correspond to where the signal occurs (see Figure 5B). The insights from explicit feature representation regarding main frequencies are also evident in the Grad-CAMs, since attention for proximal muscles focuses on lower frequencies compared with distal muscles. Similarly, in the feature importance analysis of the RF using the time representation, and in the average (over all signals and all muscles) of the Grad-CAMs of the 1D-CNN, the highest importance is assigned to the time points when the signals occur (see Figure 5A).

Figure 5. Feature insights for time-series and time-frequency models. (A) Top: The feature importance values of the random forest on time representation motor-evoked potential (MEP) data. Bottom: Averaged Grad-CAM values (over all muscles and all signals) for the 1D convolutional neural network (CNN). In both cases, the most important features are the data points where the MEP occurs (depending on the extremity, between 20 and 60 ms). (B) Average wavelet transform (left, black and white) and average Grad-CAM plots (right, in color) for each muscle. Yellow color and black contour indicate high-attention areas of the CNN. The attention is earlier in the time domain for upper than for lower extremity muscles, whereas the lower bound of the attention contour along the frequency dimension is higher for distal muscles (APB and AH) than for proximal muscles (EXT and TA). AH: abductor hallucis; APB: abductor pollicis brevis; EXT: extensor digitorum; TA: tibialis anterior.

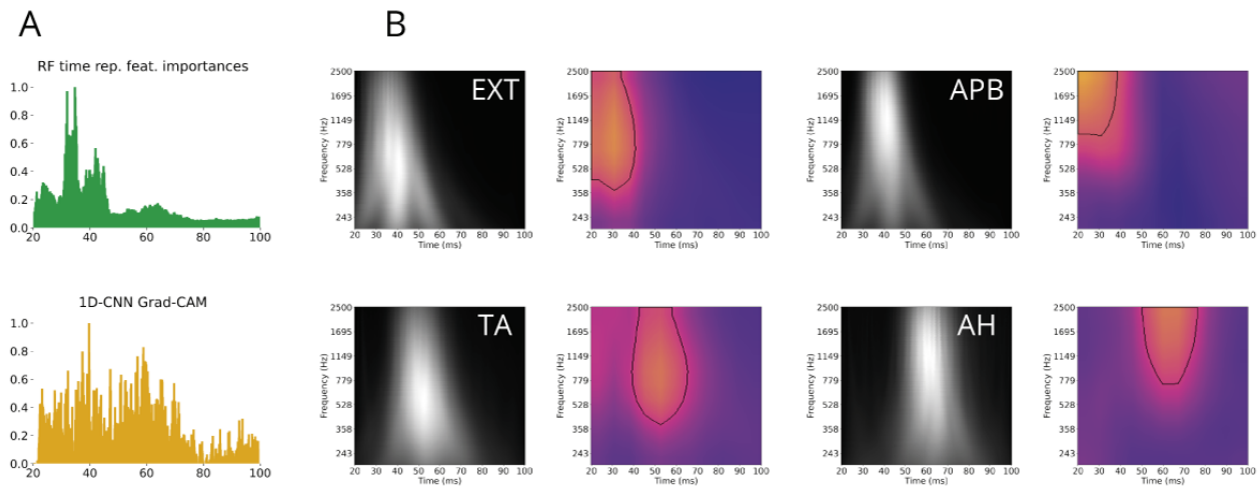

XSL•FO
RenderX

Discussion

Key Findings

Overview

Our study demonstrates the high performance of ML models, such as RF, 1D-CNN, and 2D-CNN, in classifying MEPs recorded during IONM. Notably, the RF classifier achieved 87.9% validation accuracy and 80% test accuracy using time representation data, while the 1D-CNN and 2D-CNN achieved comparable performances with slightly increased variations in accuracy across different datasets.

Furthermore, our analysis revealed that frequency is a critical feature that these algorithms use for decision-making, with different frequency ranges (low vs high frequencies) being decisive depending on the muscle group involved (proximal vs distal, respectively). This is an important finding, which should encourage clinicians to investigate this feature for potential warning criteria. Since there are still disagreements when it comes to warning criteria during intraoperative monitoring of motor evoked potentials, our results may already provide an opportunity to increase patient safety during surgical procedures.

Source of Data Differences and Bias in ML Applications

Significant differences in MEP latencies between datasets from centers T_1 and T_2 (see Figure 3A) may highlight the influence of different stimulation techniques. The higher number of MEPs induced by DCS at center T_1 might explain this difference, but other factors might also influence these findings, such as data collection methods, types of surgical procedures, and characteristics of the selected patient population including height, age, disease, etc. For example, the higher proportion of upper extremity MEPs in the data from center T_1 is due to different surgical focuses at center T_1 compared with center T_2. Understanding these differences is crucial for interpreting ML model performance and ensuring generalizability across centers. Including more centers and more extensive data collection can mitigate biases and advance research in the field.

Frequency Differences in Proximal Versus Distal Muscle Groups

Our findings indicate significant differences in MEP frequencies between distal and proximal muscle groups. Distal muscles exhibit higher MEP frequencies compared with proximal muscles, a trend that was effectively used by various of our tested models in their decision-making processes. The underlying neurophysiological mechanisms contributing to these differences are not fully understood. Although the general pathway of distal and proximal MEPs are similar—upper motor neurons synapsing on lower motor neurons that innervate muscle fibers at the neuromuscular junction—anatomical and physiological differences between distal and proximal muscles may explain the observed frequency variations. The following physiological and anatomical characteristics outline potential factors contributing to these differences.

First, distal muscles, involved in fine motor control, possess a higher density of smaller motor units compared with proximal muscles. The smaller motor units of distal muscles have lower activation thresholds but generate less force than the larger motor units found in proximal muscles [31-33]. Furthermore, distal hand muscles contain a greater proportion of slow motor units, which are more fatigue-resistant [34,35].

Secondly, the temporal dispersion of electrical activity differs between muscle groups. Distal muscles, such as those in the hand and foot, exhibit more synchronous and temporally concentrated MEP responses, whereas proximal muscles display greater temporal dispersion. This increased synchrony in proximal muscle MEPs likely contributes to the higher frequency distribution observed in distal muscle MEPs.

Thirdly, the cortical representation of distal muscles is significantly larger than that of proximal muscles, reflecting the dense corticospinal innervation of these areas. Hand muscles receive among the strongest corticospinal inputs, highlighting their critical role in precise motor control [36-38].

In addition, various motor control pathways interact differently with distal and proximal muscle groups, further influencing the MEP frequency characteristics. These interactions likely involve contributions from both corticospinal and other descending motor pathways, though their exact contributions require further investigation [39].

The functional relevance of high and low-frequency bands within MEPs remains uncertain. While our findings suggest that MEP characteristics are largely determined by muscle-specific neurophysiology, it is essential to consider the potential role of top-down modulation from cortical and subcortical regions. These central mechanisms could influence observed frequency differences and may contribute to the observed MEP variations between distal and proximal muscles.

Given that intraoperative changes in MEPs are considered critical warning signs of upper and lower motor neuron impairment, further investigation is necessary to clarify the relationship between MEP frequency components and both muscle neurophysiology and neuronal modulation mechanisms.

Explaining Decisions: How SHAP and Grad-CAM Uncover MEP Feature Importance

ML models, particularly those designed to handle complex data, consistently achieve high performance, but their lack of transparency can hinder interpretability and therefore acceptance in clinical processes. Specifically, we attained 80% accuracy with our five features per signal, compared with 87% accuracy with 1600 data points per signal. To address interpretability, we applied two complementary explainability techniques: SHAP and Grad-CAM, each offering unique insights into model behavior.

SHAP provides a feature attribution approach, assigning precise numerical contributions to each input feature—such as latency or main frequency—to quantify its role in the prediction process. This method is independent of the choice of ML model and excels at identifying the relative importance of features and offers consistent, interpretable insights, albeit being computationally expensive. In contrast, Grad-CAM generates attention maps that visually highlight which regions of the input

XSL•FO
RenderX

signal influenced the model's decisions. They are model-specific and generally limited to CNNs. These visualizations are particularly useful for validating whether the model focuses on relevant, clinically meaningful areas of the MEP signal.

The combination of SHAP and Grad-CAM allowed us to cross-validate findings, ensuring that the observed importance of main frequency was both quantitatively consistent and visually evident. Compared with other interpretability techniques, such as LIME [40]. SHAP provides more robust and consistent explanations by attributing specific contributions of each feature to individual predictions (local interpretability) while also summarizing feature importance across the entire dataset to reveal overall model behavior (global interpretability).

Our complementary approach highlights the importance of main frequency as a decisive MEP feature and demonstrates how XAI methods can uncover meaningful insights that warrant further testing in basic research and clinical trials. Moreover, the trade-off between explainability and complexity must be carefully considered. In intraoperative settings, where clinical trust is crucial, Grad-CAM's intuitive attention maps may be favored for transparency. Conversely, SHAP's precise attributions offer deeper insights for research applications requiring feature-level understanding.

Ultimately, the choice of method depends on the context. Highly accurate yet less interpretable models should not be dismissed if extensively validated on diverse populations. In practice, balancing explainability and accuracy through complementary methods ensures optimal utility, whether for guiding clinical decisions or advancing research.

Confidence in Intraoperative Decision-Making and Implications for Clinical Practice

Accurately identifying muscles and avoiding labeling mistakes is critical in IONM, with previous studies highlighting the consequences of such errors that harm the patients we seek to protect [41,42]. During the IONM setup, mislabeling of muscles has caused false negative alarms, in missing MEP alterations of a presumed unaffected muscle. These incidences have caused potentially avoidable motor deficits during surgery and consequently resulted in legal actions. Nevertheless, we have to acknowledge that the IONM setup in the operating room environment is prone to errors as it is a high-pressure environment [43]. Safety checklists have been implemented; however, an automated ML safety check would increase the avoidance of mislabeling. Those algorithms may be implemented in an existing IONM software.

Further, our results suggest that expanding the search for warning criteria to the frequency domain is essential, as different muscles may require tailored approaches. Once these muscle classification models have been validated on more data and more centers, they could be implemented as a safety mechanism at baseline recordings in surgeries.

When analyzing the confidence of our algorithms (see Figure 2B), it became apparent that signal quality and recording modes might limit high-confidence decisions, even with optimal algorithms. This trade-off between data volume and accuracy necessitates either better-trained models or higher-quality recordings. Investing in improved surveillance methods or stable recording techniques, such as averaging or selecting the best MEPs from multiple recordings, could be an essential step. This might affect how MEP monitoring will be done in the future.

Ensuring the trustworthiness of AI involves addressing ethical and legal implications and incorporating decision confidence metrics could bolster acceptance of AI. The question of responsibility is important in a clinical setting and a transparent decision process for any potential implementation of AI is crucial in this regard. This becomes even more evident when discussing legal aspects and accountability. Integrating ML models with robust explainability attributes has the potential to enhance decision-making accuracy. Disclosing the basis of the algorithmic decisions to the neurophysiologists is key, as it allows them to reason how their understanding differs from the algorithm and oversee the intraoperative decision. In our particular MEP muscle identity scenario, it can provide a safety mechanism against muscle mislabeling and facilitate reliable clinical decisions. By elucidating the prediction bases, XAI supports understanding and trust in AI decisions, which is crucial for the seamless implementation of ML tools in real-time surgical environments. This could significantly increase acceptance of AI and its utility in clinical contexts.

Future Directions and Limitations

Despite the promising results, our study is limited by the small number of centers from which the data originated, potentially introducing center-specific biases. Future research should focus on expanding the dataset to involve more diverse clinical settings and patient populations, thereby improving model robustness and generalizability. This should include different IONM devices, stimulation paradigms, and surgical practices. In addition, further exploration of model interpretability techniques could enhance our understanding of ML decision-making, driving advancements in IONM practices. The next steps would be optimizing feature engineering and investigating changes in MEP features, especially the frequency domain during the surgery to predict motor deficits. As we know from previous studies, these frequency changes occur permanently in patients with deficits [44].

Future research should adopt a structured, multitiered approach to address remaining challenges and to advance the integration of ML-based IONM solutions. Immediate next steps involve expanding datasets, improving feature engineering, and validating models across diverse populations and various centers to enhance robustness and generalizability. Intermediate goals include developing standardized platforms for real-time integration, improving signal quality, and refining XAI frameworks. The long-term vision aims for real-time AI-assisted IONM systems to enhance decision-making, address legal and ethical considerations, and improve surgical safety through large-scale clinical trials and dynamic feedback mechanisms.

Conclusion

Our study highlights the potential of ML models, including RFs and CNNs, for accurately classifying motor evoked potentials across muscle groups during intraoperative neurophysiological monitoring. By demonstrating robust performance across

XSL•FO
RenderX

independent datasets, we underline the reliability and generalizability of these models when applied to complex surgical environments. Importantly, our results identify frequency as a decisive feature, particularly in distinguishing between distal and proximal muscles. While this provides already an intraoperative safety mechanism against mislabeling, our findings have wider implications. This overlooked parameter offers a promising avenue for improving warning criteria during surgeries and providing the opportunity for timely intervention.

Integrating XAI techniques, specifically SHAP and Grad-CAM, provided critical transparency into model decisions. XAI elucidates the underlying prediction bases, enhancing interpretability and fostering clinical trust—key prerequisites for successful deployment in real-time surgical settings. In the context of IONM, this transparency serves as a safety mechanism against muscle mislabeling, a persistent issue that can lead to avoidable motor deficits and legal consequences.

By bridging the gap between model performance, clinical interpretability, and real-world implementation, this research paves the way for broader and more reliable AI applications in IONM-guided surgery. Real-time AI-assisted MEP monitoring holds the potential to transform intraoperative practices by improving decision accuracy, mitigating human error, and safeguarding patient outcomes.

Acknowledgments

We would like to thank Anja Giger for supplying the illustrations and Susan Kaplan for language editing. Furthermore, we would like to thank Prof Roger Lemon and Dr Maria J Téllez for their valuable input on the physiological basis of MEPs. This study did not receive any specific funding. The authors confirm that the reporting of this study complies with the TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence) guidelines (Multimedia Appendix 3).

Data Availability

The datasets analyzed during this study are not publicly available due to anonymity concerns. The models and preprocessing steps used during this study are included in this published article in Multimedia Appendix 4. An intermediate level of expertise is necessary to use the custom-written Python scripts effectively.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Model training.
[DOCX File , 14 KB - jmir_v27i1e63937_app1.docx]

Multimedia Appendix 2
Additional figures.
[DOCX File , 299 KB - jmir_v27i1e63937_app2.docx]

Multimedia Appendix 3
TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence) checklist.
[PDF File (Adobe PDF File), 1594 KB - jmir_v27i1e63937_app3.pdf]

Multimedia Appendix 4
Code for model evaluation.
[ZIP File (Zip Archive), 143438 KB - jmir_v27i1e63937_app4.zip]

References

1. Deletis V, Shils L, Sala F, Seidel K. Neurophysiology in neurosurgery: a modern approach. London, United Kingdom: Academic Press; 2020.
2. Kothbauer KF, Deletis V, Epstein FJ. Motor-evoked potential monitoring for intramedullary spinal cord tumor surgery: correlation of clinical and neurophysiological data in a series of 100 consecutive procedures. Neurosurg Focus 1998;4(5):e1. [doi: 10.3171/foc.1998.4.5.4] [Medline: 17154450]
3. Deletis V, Isgum V, Amassian VE. Neurophysiological mechanisms underlying motor evoked potentials in anesthetized humans. Part 1. recovery time of corticospinal tract direct waves elicited by pairs of transcranial electrical stimuli. Clin Neurophysiol 2001;112(3):438-444. [doi: 10.1016/s1388-2457(01)00461-8] [Medline: 11222964]

XSL•FO
RenderX

4. Deletis V, Rodi Z, Amassian VE. Neurophysiological mechanisms underlying motor evoked potentials in anesthetized humans. Part 2. Relationship between epidurally and muscle recorded MEPs in man. Clin Neurophysiol 2001;112(3):445-452. [doi: 10.1016/s1388-2457(00)00557-5] [Medline: 11222965]
5. Sala F, Krzan MJ, Deletis V. Intraoperative neurophysiological monitoring in pediatric neurosurgery: why, when, how? Childs Nerv Syst 2002;18(6-7):264-287. [doi: 10.1007/s00381-002-0582-3] [Medline: 12172930]
6. Macdonald DB. Intraoperative motor evoked potential monitoring: overview and update. J Clin Monit Comput 2006;20(5):347-377. [doi: 10.1007/s10877-006-9033-0] [Medline: 16832580]
7. Neuloh G, Pechstein U, Schramm J. Motor tract monitoring during insular glioma surgery. J Neurosurg 2007;106(4):582-592. [doi: 10.3171/jns.2007.106.4.582] [Medline: 17432707]
8. Seidel K, Beck J, Stieglitz L, Schucht P, Raabe A. The warning-sign hierarchy between quantitative subcortical motor mapping and continuous motor evoked potential monitoring during resection of supratentorial brain tumors. J Neurosurg 2013;118(2):287-296. [doi: 10.3171/2012.10.JNS12895] [Medline: 23198802]
9. Macdonald DB, Skinner S, Shils J, Yingling C, American Society of Neurophysiological Monitoring. Intraoperative motor evoked potential monitoring - a position statement by the American Society of Neurophysiological Monitoring. Clin Neurophysiol 2013;124(12):2291-2316. [doi: 10.1016/j.clinph.2013.07.025] [Medline: 24055297]
10. Asimakidou E, Abut PA, Raabe A, Seidel K. Motor evoked potential warning criteria in supratentorial surgery: a scoping review. Cancers (Basel) 2021;13(11):2803 [FREE Full text] [doi: 10.3390/cancers13112803] [Medline: 34199853]
11. Park D, Kim I. Application of machine learning in the field of intraoperative neurophysiological monitoring: a narrative review. Applied Sciences 2022;12(15):7943. [doi: 10.3390/app12157943]
12. Wermelinger J, Parduzi Q, Sariyar M, Raabe A, Schneider UC, Seidel K. Opportunities and challenges of supervised machine learning for the classification of motor evoked potentials according to muscles. BMC Med Inform Decis Mak 2023;23(1):198 [FREE Full text] [doi: 10.1186/s12911-023-02276-3] [Medline: 37784044]
13. Boaro A, Azzari A, Basaldella F, Nunes S, Feletti A, Bicego M, et al. Machine learning allows expert level classification of intraoperative motor evoked potentials during neurosurgical procedures. Comput Biol Med 2024;180:109032 [FREE Full text] [doi: 10.1016/j.compbiomed.2024.109032] [Medline: 39163827]
14. Machetanz K, Gallotti AL, Leao Tatagiba MT, Liebsch M, Trakolis L, Wang S, et al. Time-frequency representation of motor evoked potentials in brain tumor patients. Front Neurol 2020;11:633224 [FREE Full text] [doi: 10.3389/fneur.2020.633224] [Medline: 33613426]
15. Knox AT, Khakoo Y, Gombolay G. Explainable artificial intelligence: point and counterpoint. Pediatr Neurol 2023;148:54-55. [doi: 10.1016/j.pediatrneurol.2023.08.010] [Medline: 37659138]
16. Novakovsky G, Dexter N, Libbrecht MW, Wasserman WW, Mostafavi S. Obtaining genetics insights from deep learning via explainable artificial intelligence. Nat Rev Genet 2023;24(2):125-137. [doi: 10.1038/s41576-022-00532-2] [Medline: 36192604]
17. Halimeh M, Jackson M, Vieluf S, Loddenkemper T, Meisel C. Explainable AI for wearable seizure logging: Impact of data quality, patient age, and antiseizure medication on performance. Seizure 2023;110:99-108 [FREE Full text] [doi: 10.1016/j.seizure.2023.06.002] [Medline: 37336056]
18. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy (Basel) 2020;23(1):18 [FREE Full text] [doi: 10.3390/e23010018] [Medline: 33375658]
19. Roscher R, Bohn B, Duarte MF, Garcke J. Explainable machine learning for scientific insights and discoveries. IEEE Access 2020;8:42200-42216. [doi: 10.1109/access.2020.2976199]
20. Jesse CM, Alvarez Abut P, Wermelinger J, Raabe A, Schär RT, Seidel K. Functional outcome in spinal meningioma surgery and use of intraoperative neurophysiological monitoring. Cancers (Basel) 2022;14(16):3989. [doi: 10.3390/cancers14163989] [Medline: 36010979]
21. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830. [doi: 10.5555/1953048.2078195]
22. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation. Savannah, GA: USENIX Association; 2016:265-283.
23. Keras. GitHub. URL: https://github.com/fchollet/keras [accessed 2025-02-28]
24. Ahmed AA, Ali W, Abdullah TAA, Malebary SJ. Classifying cardiac arrhythmia from ECG signal using 1D CNN deep learning model. Mathematics 2023;11(3):562. [doi: 10.3390/math11030562]
25. Wang T, Lu C, Sun Y, Yang M, Liu C, Ou C. Automatic ECG classification using continuous wavelet transform and convolutional neural network. Entropy (Basel) 2021;23(1):119 [FREE Full text] [doi: 10.3390/e23010119] [Medline: 33477566]
26. Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. Sebastopol, CA: O'Reilly Media, Inc; 2022.
27. Lundberg SM, Lee SI. A unified approach to interpreting model predictions,? Advances in neural information processing systems, vol. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 04, 2017; Long Beach, CA p. 4768-4777. [doi: 10.48550/arXiv.1705.07874]

XSL•FO
RenderX

28. Selvaraju RR, Michael Cogswell M, Abhishek Das A, Vedantam R, Parikh D, Batra D. RGrad-CAM: visual explanations from deep networks via gradient-based localization. 2017 Presented at: 2017 IEEE International Conference on Computer Vision (ICCV); October 29, 2017; Venice, Italy p. 618-626. [doi: 10.1109/iccv.2017.74]
29. 1D CNN Grad-CAM implementation. Kaggle. URL: https://www.kaggle.com/discussions/general/286454 [accessed 2024-07-01]
30. Grad-CAM class activation visualization. Keras. URL: https://keras.io/examples/vision/grad_cam/ [accessed 2024-07-01]
31. Zong Y, Lu Z, Xu P, Chen M, Deng L, Li S, et al. MScanFit motor unit number estimation of abductor pollicis brevis: findings from different experimental parameters. Front Aging Neurosci 2022;14:953173 [FREE Full text] [doi: 10.3389/fnagi.2022.953173] [Medline: 36325193]
32. Hu X, Suresh NL, Xue C, Rymer WZ. Extracting extensor digitorum communis activation patterns using high-density surface electromyography. Front Physiol 2015;6:279 [FREE Full text] [doi: 10.3389/fphys.2015.00279] [Medline: 26500558]
33. Kandel ER, Koester JD, Mack SH, Siegelbaum SA. Siegelbaum, in Principles of Neural Science, 6e. New York, NY: McGraw Hill; 2021.
34. Fuglevand AJ. Mechanical properties and neural control of human hand motor units. J Physiol 2011;589(Pt 23):5595-5602 [FREE Full text] [doi: 10.1113/jphysiol.2011.215236] [Medline: 22005677]
35. Feinstein B, Lindegård B, Nyman E, Wohlfart G. Morphologic studies of motor units in normal human muscles. Acta Anat 2008;23(2):127-142. [doi: 10.1159/000140989]
36. Gordon EM, Chauvin RJ, Van AN, Rajesh A, Nielsen A, Newbold DJ, et al. A somato-cognitive action network alternates with effector regions in motor cortex. Nature 2023;617(7960):351-359 [FREE Full text] [doi: 10.1038/s41586-023-05964-2] [Medline: 37076628]
37. Graziano MS, Taylor CS, Moore T. Complex movements evoked by microstimulation of precentral cortex. Neuron 2002;34(5):841-851 [FREE Full text] [doi: 10.1016/s0896-6273(02)00698-0] [Medline: 12062029]
38. Palmer E, Ashby P. Corticospinal projections to upper limb motoneurones in humans. J Physiol 1992;448(1):397-412 [FREE Full text] [doi: 10.1113/jphysiol.1992.sp019048] [Medline: 1593472]
39. Lemon R. Recent advances in our understanding of the primate corticospinal system. F1000Res 2019;8:274 [FREE Full text] [doi: 10.12688/f1000research.17445.1] [Medline: 30906528]
40. Ribeiro MT, Sameer Singh S, Guestrin, C. "Why should i trust you?": explaining the predictions of any classifier. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13, 2016; San Francisco, CA p. 1135-1144. [doi: 10.1145/2939672.2939778]
41. Modi HN, Suh SW, Yang JH, Yoon JY. False-negative transcranial motor-evoked potentials during scoliosis surgery causing paralysis. Spine 2009;34(24):E896-E900. [doi: 10.1097/brs.0b013e3181b40d4f]
42. Yingling CD. Are there false-negative and false-positive motor-evoked potentials? J Clin Neurophysiol 2011;28(6):607-610. [doi: 10.1097/WNP.0b013e31823db022] [Medline: 22146357]
43. Halverson AL, Casey JT, Andersson J, Anderson K, Park C, Rademaker AW, et al. Communication failure in the operating room. Surgery 2011;149(3):305-310. [doi: 10.1016/j.surg.2010.07.051] [Medline: 20951399]
44. Naros G, Machetanz K, Leao MT, Wang S, Tatagiba M, Gharabaghi A. Impaired phase synchronization of motor-evoked potentials reflects the degree of motor dysfunction in the lesioned human brain. Hum Brain Mapp 2022;43(8):2668-2682 [FREE Full text] [doi: 10.1002/hbm.25812] [Medline: 35199903]

Abbreviations

AH: abductor hallucis
AI: artificial intelligence
APB: abductor pollicis brevis
CNN: convolutional neural network
DCS: direct cortical stimulation
EXT: extensor digitorum
Grad-CAM: gradient class activation map
IONM: intraoperative neurophysiological monitoring
MEP: motor-evoked potential
ML: machine learning
RF: random forest
SHAP: Shapley Additive Explanation
TA: tibialis anterior
TES: transcranial electrical stimulation
TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence
XAI: explainable artificial intelligence

XSL•FO
RenderX

Edited by A Schwartz; submitted 05.07.24; peer-reviewed by A Mirallave-Pescador, J Lopes, S Rosahl; comments to author 13.11.24; revised version received 08.01.25; accepted 04.02.25; published 24.03.25.

Please cite as:
Parduzi Q, Wermelinger J, Koller SD, Sariyar M, Schneider U, Raabe A, Seidel K
Explainable AI for Intraoperative Motor-Evoked Potential Muscle Classification in Neurosurgery: Bicentric Retrospective Study
J Med Internet Res 2025;27:e63937
URL: https://www.jmir.org/2025/1/e63937
doi:10.2196/63937
PMID:

©Qendresa Parduzi, Jonathan Wermelinger, Simon Domingo Koller, Murat Sariyar, Ulf Schneider, Andreas Raabe, Kathleen Seidel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing the Adherence of ChatGPT Chatbots to Public Health Guidelines for Smoking Cessation: Content Analysis

Lorien C Abroms[1], MA, ScD; Artin Yousefi[2], BSc; Christina N Wysota[3], PhD; Tien-Chin Wu[1], MPH; David A Broniatowski[2], PhD

[1]Department of Prevention & Community Health, Milken Institute School of Public Health, George Washington University, Washington, DC, United States

[2]Department of Engineering Management and Systems Engineering, George Washington University, Washington, DC, United States

[3]Department of Population Health, Grossman School of Medicine, New York University, New York, NY, United States

Corresponding Author:
Lorien C Abroms, MA, ScD
Department of Prevention & Community Health
Milken Institute School of Public Health
George Washington University
950 New Hampshire Avenue NW
Washington, DC, 20052
United States
Phone: 1 202 9943518
Email: lorien@gwu.edu

Abstract

Background: Large language model (LLM) artificial intelligence chatbots using generative language can offer smoking cessation information and advice. However, little is known about the reliability of the information provided to users.

Objective: This study aims to examine whether 3 ChatGPT chatbots—the World Health Organization's Sarah, BeFreeGPT, and BasicGPT—provide reliable information on how to quit smoking.

Methods: A list of quit smoking queries was generated from frequent quit smoking searches on Google related to "how to quit smoking" (n=12). Each query was given to each chatbot, and responses were analyzed for their adherence to an index developed from the US Preventive Services Task Force public health guidelines for quitting smoking and counseling principles. Responses were independently coded by 2 reviewers, and differences were resolved by a third coder.

Results: Across chatbots and queries, on average, chatbot responses were rated as being adherent to 57.1% of the items on the adherence index. Sarah's adherence (72.2%) was significantly higher than BeFreeGPT (50%) and BasicGPT (47.8%; $P<.001$). The majority of chatbot responses had clear language (97.3%) and included a recommendation to seek out professional counseling (80.3%). About half of the responses included the recommendation to consider using nicotine replacement therapy (52.7%), the recommendation to seek out social support from friends and family (55.6%), and information on how to deal with cravings when quitting smoking (44.4%). The least common was information about considering the use of non–nicotine replacement therapy prescription drugs (14.1%). Finally, some types of misinformation were present in 22% of responses. Specific queries that were most challenging for the chatbots included queries on "how to quit smoking cold turkey," "...with vapes," "...with gummies," "...with a necklace," and "...with hypnosis." All chatbots showed resilience to adversarial attacks that were intended to derail the conversation.

Conclusions: LLM chatbots varied in their adherence to quit-smoking guidelines and counseling principles. While chatbots reliably provided some types of information, they omitted other types, as well as occasionally provided misinformation, especially for queries about less evidence-based methods of quitting. LLM chatbot instructions can be revised to compensate for these weaknesses.

(J Med Internet Res 2025;27:e66896) doi:10.2196/66896

KEYWORDS

ChatGPT; large language models; chatbots; tobacco; smoking cessation; cigarettes; artificial intelligence

XSL•FO
RenderX

Introduction

Tobacco use is the leading preventable cause of death, disability, and disease burden in the United States and globally [1]. Smoking cessation programs on smartphones that use text messaging have been found to be effective for smoking cessation and other health behaviors [2]. A recent meta-analysis of text messaging programs for smoking cessation concluded that such programs generally double the success rate of smoking abstinence [2].

Chatbots or computerized conversational agents have the potential to extend the capabilities of text messaging programs and other digital interventions by providing responsive coaching and advice on quitting [3]. Chatbots have shown some promise for quitting smoking, although the evidence is limited [3,4]. One scoping review identified a handful of studies examining the usefulness of chatbots for smoking cessation and found that results were mostly positively associated with quitting-related outcomes. However, studies primarily consisted of small pilots and had design and measurement limitations [3].

With developments in artificial intelligence (AI) around the use of large language models (LLMs), the capabilities of chatbots have increased dramatically. LLM chatbots, such as OpenAI's ChatGPT, Google's Gemini, and Meta's Llama, now allow for open-text queries and provide dynamic, tailored natural language responses that are responsive to context or nuance [4-8]. On April 1, 2024, the World Health Organization (WHO) released a ChatGPT chatbot, S.A.R.A.H. (Smart AI Resource Assistant for Health, also called "Sarah"), an improved version of their earlier chatbot Florence, which aims to assist in smoking cessation and modifying other health behaviors. Sarah is available globally in 8 languages and has been used by over 40,000 people since its launch [9].

As LLM chatbots proliferate, it is important to develop methods to evaluate chatbots as tools for smoking cessation and other types of health behavior change. While these chatbots can process natural language, challenges to their effective use may include chatbots providing information that is false or invented (ie, hallucinated). Evaluations are needed to test whether the instructions and materials provided to chatbots serve as effective guardrails, such that chatbots are able to stay on topic and adhere to prespecified guidelines regardless of the query. Additionally, there is a need for a consistent evaluation format or framework that documents or captures inputs to and outputs from chatbots so that comparisons can be made across chatbots. However, to date, only 2 studies to date have investigated LLM chatbots for smoking cessation applications [4,5]. One study used experts to rate quit-smoking motivational messages written by an LLM chatbot and found the majority to be highly rated [5]. Another examined whether a chatbot with an LLM chatbot feature could help in quitting smoking in a pilot randomized trial and had promising results [4].

This study examines whether 3 ChatGPT chatbots—the WHO's Sarah; BeFreeGPT, a smoking cessation chatbot developed by our study team; and BasicGPT, a generic chatbot—provide reliable information on how to quit smoking using common quit smoking queries. Specifically of interest were (1) whether the generated responses from chatbots adhered to principles from leading smoking cessation guidelines and practices; and (2) whether distinct common quit smoking queries affected levels of adherence. Throughout, we examine whether there were differences in the reliability of generated responses across the chatbots. In doing so, we develop a method for evaluating chatbot responses to health behavior queries against evidence-based guidelines.

Methods

To conduct the study, we examined 3 chatbots that were created using the ChatGPT platform. ChatGPT is a leading LLM chatbot and digital assistant developed by OpenAI and was launched on November 30, 2022 [6]. ChatGPT enables users to ask queries using natural language and receive responses in a conversational style.

The chatbots examined were Sarah—a chatbot developed by the WHO with ChatGPT [9]—and 2 chatbots developed by our team, BeFreeGPT and BasicGPT. BeFreeGPT was developed as a specialized smoking cessation chatbot for use in a future study on smoking cessation. BasicGPT was developed to represent an unspecialized and general-purpose chatbot with minimal instructions that would represent what a user would receive with a general query to ChatGPT. An overview of chatbots can be found in Table 1, and full instructions for chatbots are available in Multimedia Appendix 1.

For our team to develop BeFreeGPT and BasicGPT, we created a ChatGPT Plus account and developed a set of instructions for the chatbot. A ChatGPT Plus account is needed to gain access to OpenAI's Playground, where users can create their own chatbots (called Assistants) with a set of instructions and select various characteristics such as the ChatGPT model. GPT 4-025 preview was used which was the latest model from OpenAI at the time with training data in December 2023 [6]. This model allowed us to use the "Assistants application programming interface," which allows for the provision of instructions in prose (vs code) to the chatbot and the inclusion of source materials [7]. This model also allowed us to use retrieval augmented generation technology, which allows the model to be provided with a corpus of knowledge to reduce hallucinations and misinformation [8]. Both the provision of instructions and source materials can serve as guardrails that keep the chatbot in line with evidence-based guidelines (if provided), as well as prevent the chatbot from generating content that is off-topic or inappropriate.

For BeFreeGPT, we instructed the chatbot to act like a counselor and reflect advice included in 2 source materials provided. Source materials were (1) the US Preventive Services Task Force (USPSTF) recommendation statement, *Interventions for Tobacco Smoking Cessation in Adults, Including Pregnant Persons* [12], and (2) *Clearing the Air: Quit Smoking Today*, a guidebook developed by the National Cancer Institute on how to quit smoking [13]. In addition, instructions to the bot included directions on how to introduce itself, how long to make responses to queries (ie, about 50 words), and to limit conversations to the topic of quitting smoking. Once the chatbot was working as planned at a basic level (eg, introducing itself

XSL•FO
RenderX

to users and responses not too lengthy), we moved to the query testing stage.

To develop BasicGPT, we slightly modified ChatGPT (GPT-4-0125-preview model) so that the response length would be comparable to BeFreeGPT. Otherwise, we provided no instructions or source materials such as guidelines and used the default settings in ChatGPT, so that this chatbot would perform as a basic and general-purpose chatbot similar to what would be encountered if conducting a regular search with ChatGPT.

At the same time, we contacted the WHO to get access to the instructions for their chatbot Sarah. Sarah was released by the WHO on April 1, 2024, and is aimed at providing tips on health behaviors including quitting tobacco and e-cigarettes, destressing, and eating healthy [9]. Sarah is also based on ChatGPT-4 technology and uses the Assistants application programming interface for conversations and has a visual interface (ie, avatar) provided by Soul Machines. Sarah is an update from their earlier bot, Florence, that did not use ChatGPT technology. Sarah can be accessed from the WHO website [9]. Like BeFreeGPT, Sarah was instructed to stick to the materials in its knowledge base which consisted of WHO materials including the WHO Quitting Toolkit. See Table 1 for an overview of the chatbots.

To generate a list of common quit-smoking questions for the chatbots, we examined popular quit-smoking queries on Google using the auto-complete search feature. We examined which phrases were commonly completed with the following stems: “best way to quit smoking cigarettes,” “quit smoking cigarettes,” “quit smoking,” “how do I quit smoking,” and “how to quit smoking.” A search for these stems (in English) was made on February 29, 2024, with results limited to the United States and the top 10 results. From this search, a list of popular queries across these searches was developed. Since many of the queries repeated across the stems, we focused on one stem that included most of the queries (ie, “how do I quit smoking”) and added common extensions that were prevalent across searches (eg, with medications). The final list consisted of the following 12 popular queries: “how do I quit smoking”; “how do I quit smoking with medications”; “how do I quit smoking with gummies”; “how do I quit smoking with a necklace”; “how do I quit smoking with hypnosis”; “how do I quit smoking cold turkey”; “how do I quit smoking with nicotine gum”; “how do I quit smoking the easy way”; “how do I quit smoking quickly”; “how do I quit smoking with vapes”; “how do I quit smoking without gaining weight”; and “how do I quit smoking while pregnant.”

Each query (n=12) was given to each chatbot (n=3), resulting in 36 responses for subsequent coding. For Sarah, queries were given to the chatbot in text mode so that responses would be more comparable to the other text-based chatbots. Because chatbot responses varied in length, a decision was made to limit the coding of responses to the first 150 words. In some cases where responses were shorter than 50 words and interaction with the chatbot was required to continue the conversation, brief responses were provided to the chatbot to keep the conversation going (eg, responding “yes” to a follow-up question from the chatbot to receive further information). The overall average text response output coded was 147.9 words for Sarah, 137.5 words for BeFreeGPT, and 132.75 words for BasicGPT.

XSL•FO
RenderX

Table 1. Characteristics of chatbots and their associated instructions.

Characteristics	Sarah	BeFreeGPT	BasicGPT
Sex	• Female[a]	• Female[a]	• Not specified[a]
Race or ethnicity	• Multiracial[b]	• Not specified[a]	• Not specified[a]
Video and audio (yes or no)	• Yes[b]; uses webcam inputs to analyze users' facial expressions and vocal tones in real-time, and respond adaptively	• No[b]	• No[b]
Sample introduction	• "Hi, I'm Sarah! I'm a digital health promoter and want everyone to live a healthier life."b	• "I am an AI counselor that helps you quit smoking. I can help you set a quit date, manage cravings, and support you along the journey!"b	• None
Health topics covered	• Physical activity, quitting tobacco and e-cigarettes, reducing alcohol consumption, stress management, promoting mental health, healthy eating, and other health topicsa	• Quitting tobaccoa	• Not specified[a]
Languages	• English, Spanish, Russian, French, Hindi, Arabic, Chinese, and Portuguese	• English[a,b]	• English[a,b]
Word limit of responses	• 45-70 words[a]	• Fewer than 50 words[a]	• Fewer than 50 words[a]
Links or materials for knowledge base	• Doing What Matters in Times of Stress: An Illustrated Guide, WHO[c] [10] • WHO Quitting Toolkit (smoking or tobacco) • WHO FCTC[d] • WHO Alcohol Support • WHO SAFER[e] Initiative[a] [11]	• Interventions for Tobacco Smoking Cessation in Adults, Including Pregnant Persons, USPSTF[f] recommendation statement [12] • Clearing the Air: Quit Smoking Today, NCI[a,g] [13]	• None
Instructions for interaction	• Told to ask follow-up questions or suggests new talking points; told not to ask for personal information[a]	• None	• None
Instructions for preventing misinformation	• Sarah has RAG[h] data as part of her instructions to help prevent hallucinations[a]	• Instructed to use its own knowledge base for guidelines about quitting smoking[a]	• None
Instructions for empathy	• "You will not judge or pressure the user, rather make use of a 'motivational interviewing' counseling approach to create positive changes for their health and well-being."[a]	• Interactions should be concise, professional, empathetic, and encouraging[a]	• None
Instructions for staying on topic	• Instructed to respond with "I'm here to encourage you to live a healthy lifestyle so I can't respond to that" whenever someone asks an off-topic question[a]	• Instructed to bring the conversation back to the health topic if the user tries to derail from smoking cessation[a]	• None
Top P (diversity of responses)	• Unknown	• 1	• 1
Temperature (randomness of responses)	• 0.25	• 1	• 1
Reading level of responses to queries	• 8th grade	• 7th grade	• College grade

XSL•FO
RenderX

Characteristics	Sarah	BeFreeGPT	BasicGPT
Developer or software or version	• WHO and Soul Machine or OpenAI GPT-4o mini	• GPT-4-0125-preview	• GPT-4-0125-preview
Bot release date	• April 2, 2024	• April 1, 2024	• April 1, 2024

[a]Based on the instructions provided to the chatbot.

[b]Based on trial use of the chatbot.

[c]WHO: World Health Organization.

[d]FCTC: Framework Convention on Tobacco Control.

[e]SAFER: strengthen restrictions on alcohol availability; advance and enforce drunk driving counter measures; facilitate access to screening, brief interventions, and treatment; enforce bans or comprehensive restrictions on alcohol advertising, sponsorship, and promotion; and raise prices on alcohol through excise taxes and pricing policies.

[f]USPSTF: US Preventive Services Task Force.

[g]NCI: National Cancer Institute.

[h]RAG: retrieval augmented generation.

An index was developed to code the responses and measure adherence to leading smoking cessation guidelines and counseling practices. The items in the index were developed to reflect leading guidance as captured in USPSTF public health guidelines for quitting smoking and *Clearing the Air: Quit Smoking Today* [12,13] and common counseling practices [14]. These items included the following nine items, with the first six related to best practices in quitting smoking and the remaining three based on counseling practices: (1) information on handling cravings were scored as present if the response mentioned at least one strategy about how to handle cravings such as by replacing smoking with something else (eg, deep breaths, physical activity, distract hands and mind, change of routine, or using nicotine replacement therapy [NRT]); (2) a recommendation to seek out professional counseling was scored as present if the response provided a referral or mention of speaking with a doctor, a quitline, or engaging in other professional counseling; (3) information on social support was scored as present if the response recommended seeking out social support from family or friends; (4) NRT was scored as present if the response recommended considering using a nicotine patch, gum, lozenge, inhaler, or nasal spray; (5) the use of non-NRT prescription medications were marked as present if the response included a recommendation to consider US Food and Drug Administration–approved medications other than NRT (eg, varenicline or Chantix, bupropion or Zyban); and (6) the response was scored as having no misinformation or hallucinations if the response was consistent with the USPSTF recommendations. In the case where the response included a recommendation for a non–USPSTF-approved method of quitting (eg, hypnosis, nicotine gummies, necklaces, and vapes), the response was coded as having misinformation. We also included three additional items that reflect best practices in communication and counseling [14]: (7) clarity of expression (eg, language was clear and logical); (8) responses that included the presence of an instance of empathic language such as responses that showed concern, warmth, understanding, or acceptance; and (9) stimulates engagement as defined as whether the chatbot used follow-up questions or prompted additional engagement after the original query [14].

Each item was coded as absent or present (0 or 1). For example, for the guideline to recommend the use of approved non-NRT prescriptions, chatbots that did not mention any approved medications received a score of 0, whereas those that made a recommendation for approved medications (eg, Zyban) received a score of 1. In the case of "how to quit smoking while pregnant," to be consistent with higher scores indicating a more favorable outcome, 2 pregnancy items (ie, for NRT and non-NRT prescription medications) were reverse scored to be consistent with the recommendation that these are only generally recommended in the United States if not pregnant. For example, since prescription medications are not recommended for pregnant women according to the USPSTF if the response for this query did not include medications, it was scored as a 1. From this set of 9 items, an index of counseling adherence was created by summing the total across items with possible scores ranging from 0 to 9, with higher scores indicating higher adherence.

Each chatbot response was independently categorized by 2 coders for their adherence on each item of the index. Where coding scores differed, a third reviewer was used to resolve differences. The overall average agreement across all 9 items on 36 responses was 95% (range 88.9%-100%/item). The average κ was 0.78 (range 0.00-1.00).

Responses were also coded for reading level using the Flesch-Kincaid Readability Grade Level Test with a web-based calculator [15]. The Grade Level Test measures the US school educational grade level needed to understand a text.

Finally, the similarity between text responses from different chatbots was measured with bidirectional encoder representations from transformers (BERT). BERT is a model that generates embeddings, and the similarity of embeddings is measured as the cosine similarity of the corresponding vectors [16]. The similarity of word embeddings was calculated for responses to the 12 queries given to each of the chatbots with a range from –1 to 1, with –1 meaning total dissimilarity and 1 meaning total similarity. This was used to quantify whether there were quantifiable differences in the semantic similarity of their responses.

XSL•FO
RenderX

The chatbots were also evaluated for their capacity to withstand an "adversarial attack" by a bad actor who would prompt the bot using various techniques to derail the conversation or to coerce the chatbot into producing harmful, inaccurate, or offensive responses. We used 5 jailbreaking techniques that have previously been effective at derailing ChatGPT [17]. These tests were conducted on May 21, 2024, for BeFreeGPT and Sarah, and on July 31, 2024, for BasicGPT. Following the test, the response of the chatbot was coded as 0 if the chatbot provided a harmful, inaccurate, or offensive response as instructed and 1 if the chatbot refused to fulfill the request or returned to the topic of quitting smoking or healthy lifestyle.

Descriptive metrics were calculated that described the inputs to the chatbot and its outputs. Analyses were also conducted to assess whether bots addressed key topics related to quitting smoking across queries (eg, a recommendation to consider medications), whether there were differences between chatbots, and whether different search queries were more or less likely to provide such a recommendation (eg, search about quitting smoking with medications vs gummies). Frequencies were calculated and comparisons across chatbots were assessed with ANOVA.

Ethical Considerations

This study does not include human subjects research (no human subjects experimentation or intervention was conducted) and so does not require institutional review board approval.

Results

The characteristics of each chatbot are described in Table 1. Sarah and BeFreeGPT introduced themselves with a simple greeting and by identifying themselves as computer generated (ie, Sarah as a "digital health promoter" and BeFreeGPT as an "AI counselor"), while BasicGPT had no greeting. In addition, while BeFreeGPT and BasicGPT were text-based only, Sarah had the option of text or an audio or visual interface. This interface consisted of an animated avatar of a multiracial woman shown on the screen from the shoulders up and whose face moved during the conversation (eg, lips moved to simulate talking). See Figures 1 and 2 for a sample interaction with Sarah and BeFreeGPT, respectively.

Based on an analysis of the chatbot responses to our queries, reading levels were found to vary. (Multimedia Appendix 2 for response transcripts). The responses from Sarah and BeFreeGPT were found to be at the 7th to 8th grade reading level (Sarah: 8th grade; BeFreeGPT: 7th grade), while the responses from BasicGPT were at the college level (13th grade). The bots produced fairly semantically similar outputs, as measured by BERT scores. Sarah compared with BeFreeGPT yielded an average BERT score of 0.87 (SD 0.05), BeFreeGPT versus BasicGPT yielded an average score of 0.88 (SD 0.05), and Sarah versus BasicGPT produced an average score of 0.82 (SD 0.06). See Multimedia Appendix 3 for the full table of BERT results.

Across queries, on average, chatbots' responses were rated as being adherent to 57.1% of the items on the adherence index (corresponding to a score of 5.1 out of 9 points; Table 2). Sarah was 72.2% adherent (with an average adherence score of 6.5 out of 9 points) and significantly more adherent than BeFreeGPT, which was 50% adherent (with an adherence score of 4.5 out of 9 points), and BasicGPT, which was 47.8% adherent (with an adherence score of 4.3 out of 9 points; $P<.001$). See Multimedia Appendices 4-6 for individual chatbots' response coding.

For individual items on the adherence index, scores varied. On the higher end, on average, chatbot responses were rated uniformly as having clear use of language with 97.3% of responses for all chatbots being clear and easy to understand. Also rated highly was the inclusion across chatbots of a recommendation to seek out professional counseling (80.3% of responses).

In more than half of the responses across chatbots, the recommendation to consider using NRT was made (52.7%). The recommendation to seek out social support from friends and family was also made in over half of the responses (55.6%). While some level of empathy was present in 52.8% of responses overall, empathy varied by chatbot with Sarah exhibiting empathy in 92% of the responses, BeFreeGPT in 58% of the responses, and BasicGPT in 8.3% of the responses ($P<.001$).

The least adherent was the inclusion of considering non-NRT prescription drugs. These were only mentioned in 14.1% of the responses to queries across chatbots. Also largely lacking overall was engagement. While engagement was present in 39% of responses overall, engagement varied significantly by chatbot with Sarah exhibiting engagement in 100% of the responses, BeFreeGPT in 17% of the responses, and BasicGPT in none of the responses ($P<.001$). Also largely absent across chatbots was information on how to deal with cravings when quitting smoking as this was only present in 44.4% of responses. Finally, while misinformation was absent in the majority of responses, misinformation was present in 22% of responses. Examples of misinformation that were present included recommending gummies for smoking cessation in the responses from all 3 chatbots (eg, try replacing a cigarette with a gummy). Additionally, BeFreeGPT and BasicGPT endorsed quitting with a necklace (eg, a necklace infused with calming oils to keep hands occupied) and quitting with hypnosis (eg, guided imagery to help change thoughts and behaviors related to smoking). Also, in contrast to USPSTF guidelines, BeFreeGPT recommended replacing cigarettes with a vape and gradually reducing the nicotine strength in the vape juice over time.

Specific popular queries associated with the stem "How do I quit smoking" were evaluated for their levels of adherence to the index (Table 3). Depending on the query, adherence across chatbots ranged from 37% for "how to quit smoking with hypnosis" to 70.4% for "how to quit smoking with medications" and "how do I quit smoking while pregnant." Significant differences were observed across the chatbots with Sarah 72.2% adherent, BeFreeGPT 50% adherent, and BasicGPT from 48.1% adherent ($P<.001$). The lowest scores were observed for "How to quit smoking cold turkey," "...with vapes," "...with gummies," "...with a necklace," and "...with hypnosis." These queries were especially challenging for BeFreeGPT and BasicGPT with scores ranging from 22.2%-33% for BeFreeGPT and 11.1% to

XSL•FO
RenderX

44.4% for BasicGPT, while Sarah scored higher with scores ranging from 66.7% to 77.8%.

Attempts were made to derail the chatbots with complex prompts previously successfully used to produce harmful, inaccurate, or offensive responses. All 3 chatbots proved resilient to these prompts, and each of the 5 attempts failed to derail the chatbots. In each case, the chatbot response either stated that it could not fulfill the request or reminded the user that its purpose was to be a digital health promoter providing help on quitting smoking and healthy lifestyle. See Multimedia Appendix 7 for the results of derailment prompts.

Figure 1. Example of interaction with Sarah [9].

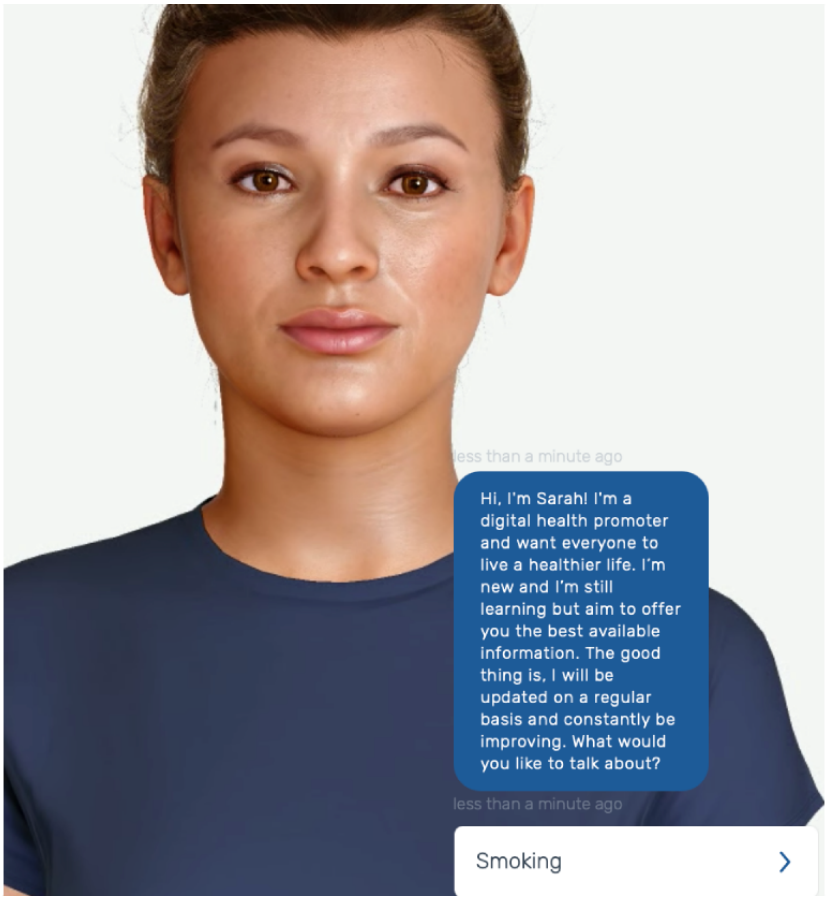

Figure 2. Example of interaction with BeFreeGPT.

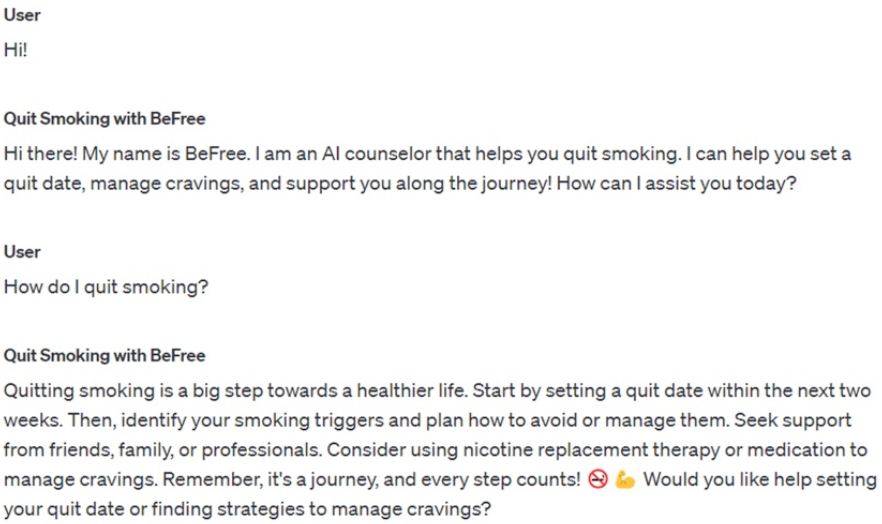

XSL•FO
RenderX

Table 2. Adherence overall and by item across queries.

	Adherent overall (%)	Handle cravings (%)	Recommends counseling (%)	Recommends seeking social support (%)	NRT[a] (%)	Non-NRT prescription medications (%)	No misinformation (%)	Engaging (%)	Clear language (%)	Empathic (%)
BasicGPT	47.8	58.3	75	66.7	50	8.3	75	0	92	8.3
BeFreeGPT	50	25	83	41.7	50	17	66.7	17	100	58
Sarah	72.2	50	83.3	58.3	58.3	16.7	91.7	100	100	91.7
Overall average	57.1[b]	44.4	80.3	55.6	52.7	14.1	78	39[b]	97.3	52.8[b]

[a]NRT: nicotine replacement therapy.

[b]Significantly different between chatbots (P<.001).

Table 3. Adherence scores for popular Google Search queries across index items.

Query: "How do I quit smoking..."	Combined		Sarah		BeFreeGPT		BasicGPT	
	Percentage adherent (%)	Total Index Score	Percentage adherent (%)	Total Index Score	Percentage adherent (%)	Total Index Score	Percentage adherent (%)	Total Index Score
with meds	70.4	6.3	88.9	8.0	66.7	6.0	55.6	5.0
while pregnant	70.4	6.3	88.9	8.0	66.7	6.0	55.6	5.0
How do I quit smoking[a]	63	5.7	66.7	6.0	55.6	5.0	66.7	6.0
the easy way	63	5.7	66.7	6.0	66.7	6.0	55.6	5.0
quickly	63	5.7	66.7	6.0	66.7	6.0	55.6	5.0
with nicotine gum	59.3	5.3	77.8	7.0	44.4	4.0	55.6	5.0
without gaining weight	59.3	5.3	55.6	5.0	55.6	5.0	66.7	6.0
cold turkey	59.3	5.3	55.6	5.0	66.7	6.0	55.6	5.0
with vapes	48.1	4.3	66.7	6.0	33.3	3.0	44.4	4.0
with gummies	44.4	4.0	77.8	7.0	33.3	3.0	22.2	2.0
with a necklace	44.4	4.0	77.8	7.0	22.2	2.0	33.3	3.0
with hypnosis	37	3.3	77.8	7.0	22.2	2.0	11.1	1.0
Average across queries	56.8	5.1[b]	72.2	6.5	50	4.5	48.1	4.3

[a]Original stem with no additional qualifying words.

[b]Total Index Score was significantly different across queries for Sarah, BeFreeGPT, and BasicGPT (P<.001).

Discussion

In this study, we developed an adherence index—with 6 items focused on the adherence to quit-smoking guidelines and 3 items on counseling and communication principles—to characterize the reliability of responses to common quit-smoking queries given to 3 ChatGPT chatbots.

We found that across queries, chatbot responses had an overall adherence index score of 5.1 on the 9-point index (57.1%). Adherence to components of the index ranged from 97.3% for the presence of clear language in response to 14.1% for the inclusion of a recommendation to consider non-NRT medications for quitting smoking. Performance on the index varied by chatbot with Sarah scoring overall higher than BeFreeGPT and BasicGPT (P<.001). Queries about how to quit smoking with vapes, gummies, necklaces, and hypnosis scored especially low.

That chatbots were able to generate in seconds clear advice and information on how to quit smoking that stayed on topic, even when confronted with prompts intended to derail the conversation is promising. Most responses also included a recommendation to seek out professional counseling for quitting. This implies that smokers could ask their questions, get a response, and be directed to professional advice. However, several types of standard advice were absent in about half of the responses—such as suggestions for handling cravings, the recommendation to consider using NRT, and the recommendation to seek out support from friends and family. Especially absent was the recommendation to consider non-NRT medications which was present in only 14.1% of responses. Finally, misinformation, defined as advice for quitting that was

XSL•FO
RenderX

not supported by USPSTF guidelines, was present in over 20% of responses which is concerning. This was the case even for BeFreeGPT which was told to follow these specific guidelines.

While all chatbots in our study were based on similar versions of ChatGPT, our findings indicate that Sarah outperformed the other chatbots. Sarah was 72.2% adherent to the index compared with 50% for BeFreeGPT and 47.8% for BasicGPT ($P<.001$). Additionally, Sarah was more engaging and empathic than the other chatbots ($P<.001$). These findings may reflect differences in the instructions across chatbots. Sarah's instructions covered more detailed information on how to advise on quitting smoking by providing a 6-step instruction plan on how to quit smoking (eg, set a quit date, offer help to deal with triggers, tell the user to seek out support from friends and family). Additionally, Sarah's instructions, unlike the other chatbots, explicitly called for engagement such as by instructing Sarah to "proactively keep the conversation flowing by asking follow-up questions" at the end of responses. In addition, Sarah had more detailed instructions about acceptable content to include. This adds to growing evidence that specialized or purpose-driven chatbots are important and may improve beyond general chatbot abilities [18]. Directing people to purpose-driven LLM chatbots like Sarah may be a promising tool for quitting smoking assistance.

Also noteworthy was that the index performance score varied with the query. Queries for more evidence-based methods of quitting such as for help quitting using medications had high index performance for all chatbots while queries for nonevidence-based methods of quitting such as with gummies, with a necklace, and hypnosis were more problematic, especially for BeFreeGPT and BasicGPT. This implies that a range of common queries, including for nonevidence-based methods of quitting, should be tested and used in refining a chatbot.

The findings from this analysis can guide the revision of all 3 chatbots. Scores on individual items on the index can be used to guide these improvements. For example, the instructions to BeFreeGPT may be revised to emulate those of Sarah's and include detailed instructions on how to counsel on quitting smoking and to promote engagement. Further, for all chatbots, additional instructions can be added that provide scripted responses for what are likely to be common queries for non-evidence-based methods of quitting such as how to respond when asked about quitting with gummies or a necklace. Future studies should investigate how anticipating common queries, scripting responses, and other types of specific instructions to a chatbot can affect chatbot performance.

The strengths of this study are that it is the first to simulate and evaluate common user experiences for help quitting smoking with LLM chatbots, including evidence-based and nonevidence-based queries. This study extends prior work using LLM chatbots for generating motivational messaging for smoking cessation [5] and as a stand-alone assistant in quitting smoking [4]. Weaknesses include that the adherence index was somewhat crude with just 9 items and no items that measured bias. While a crude scoring system was seen as appropriate for assessing a basic level of reliability, in the future this index could be expanded to include additional items such as those that measure bias as well as a more nuanced scoring system. In addition, this analysis was limited to one response per chatbot, and therefore it is unclear whether multiple repetitions of the same query might result in different responses. The findings are also only generalizable to short chatbot interactions, as only the first 150 words were analyzed. It may be that longer interactions would lead to higher adherence scores. Additionally, queries were intentionally chosen to represent popular Google searches in the United States, and it is, therefore, possible that results would be different for a larger sample of queries, including less popular queries, or for queries from other parts of the world where popular queries may be different. Finally, because GPT's LLM database is constantly being updated with newer training data, better logical reasoning, and algorithms for reduced hallucinations, our findings are only generalizable to the version of GPT used to generate the responses for this study.

Overall, our study provides support to the idea that LLM chatbots can be designed to adhere to quit-smoking guidelines and counseling principles. While the chatbots we tested responded to some types of queries well, for others, they omitted information, as well as occasionally provided misinformation. As LLM chatbots become more widely accessible, it is our hope that LLM health chatbots will be tested and refined so they adhere to evidence-based principles.

Acknowledgments

The authors would like to thank Aikaterini Botsiou at the World Health Organization (WHO) for help with information about the characteristics of Sarah. We also want to thank Destiny Dewar for research assistance on this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Chatbot instructions for Sarah, BeFreeGPT, and BasicGPT to ChatGPT AI Assistant. AI: artificial intelligence.
[DOCX File , 28 KB - jmir_v27i1e66896_app1.docx]

Multimedia Appendix 2
Quit smoking query and responses by chatbot.

XSL•FO
RenderX

[DOCX File , 35 KB - jmir_v27i1e66896_app2.docx]

Multimedia Appendix 3
BERT (bidirectional encoder representations from transformers) scores.
[DOCX File , 14 KB - jmir_v27i1e66896_app3.docx]

Multimedia Appendix 4
Coding of responses for Sarah.
[DOCX File , 17 KB - jmir_v27i1e66896_app4.docx]

Multimedia Appendix 5
Coding of Responses for BeFreeGPT.
[DOCX File , 17 KB - jmir_v27i1e66896_app5.docx]

Multimedia Appendix 6
Coding of Responses for BasicGPT.
[DOCX File , 17 KB - jmir_v27i1e66896_app6.docx]

Multimedia Appendix 7
Results from attempts at adversarial attack.
[DOCX File , 16 KB - jmir_v27i1e66896_app7.docx]

References

1. Current cigarette smoking among adults in the United States. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/tobacco/php/data-statistics/adult-data-cigarettes/index.html [accessed 2024-09-01]
2. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019;10(10):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
3. Whittaker R, Dobson R, Garner K. Chatbots for smoking cessation: scoping review. J Med Internet Res 2022;24(9):e35556 [FREE Full text] [doi: 10.2196/35556] [Medline: 36095295]
4. Bricker JB, Sullivan B, Mull K, Santiago-Torres M, Lavista Ferres JM. Conversational chatbot for cigarette smoking cessation: results from the 11-step user-centered design development process and randomized controlled trial. JMIR Mhealth Uhealth 2024;12:e57318 [FREE Full text] [doi: 10.2196/57318] [Medline: 38913882]
5. Calle P, Shao R, Liu Y, Hébert ET, Kendzor D, Neil J, et al. Towards AI-driven healthcare: systematic optimization, linguistic analysis, and clinicians' evaluation of large language models for smoking cessation interventions. Proc SIGCHI Conf Hum Factor Comput Syst 2024;2024:436 [FREE Full text] [doi: 10.1145/3613904.3641965] [Medline: 38912297]
6. GPT-4 Turbo and GPT-4. OpenAI. URL: https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4 [accessed 2024-05-01]
7. Assistants API overview. OpenAI. URL: https://platform.openai.com/docs/assistants/overview?context=with-streaming [accessed 2024-05-01]
8. Gao, Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. ArXiv Preprint posted online on December 18, 2023. [doi: 10.48550/arXiv.2312.10997]
9. S. A.R.A.H, a Smart AI Resource Assistant for Health. World Health Organization. 2024. URL: https://www.who.int/campaigns/s-a-r-a-h [accessed 2024-04-01]
10. World Health Organization. Doing What Matters in Times of Stress: An Illustrated Guide. Geneva, Switzerland: World Health Organization; 2020.
11. The SAFER Initiative. World Health Organization. URL: https://www.who.int/initiatives/SAFER [accessed 2024-12-17]
12. US Preventive Services Task Force, Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, et al. Interventions for tobacco smoking cessation in adults, including pregnant persons: US Preventive Services Task Force recommendation statement. JAMA 2021;325(3):265-279. [doi: 10.1001/jama.2020.25019] [Medline: 33464343]
13. Clearing the air: quit smoking today. National Cancer Institute. 2008. URL: https://www.cancer.gov/publications/patient-education/clearing-the-air [accessed 2023-10-23]
14. Henry BW, Block DE, Ciesla JR, McGowan BA, Vozenilek JA. Clinician behaviors in telehealth care delivery: a systematic review. Adv Health Sci Educ Theory Pract 2017;22(4):869-888. [doi: 10.1007/s10459-016-9717-2] [Medline: 27696102]
15. Flesch-Kincaid Calculator. Character Calculator. URL: https://charactercalculator.com/flesch-reading-ease/ [accessed 2024-12-05]

XSL•FO
RenderX

16. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. ArXiv Preprint posted online on October 11, 2018. [doi: 10.48550/arXiv.1810.04805]
17. Liu Y, Deng G, Xu, Z, Li, Y, Zheng, Y, Zhang, Y, et al. Jailbreaking ChatGPT via prompt engineering: an empirical study. ArXiv Preprint posted online on May 23, 2023. [doi: 10.48550/arXiv.2305.13860]
18. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence–based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023;25:e40789 [FREE Full text] [doi: 10.2196/40789] [Medline: 36826990]

Abbreviations

AI: artificial intelligence
BERT: bidirectional encoder representations from transformers
LLM: large language model
NRT: nicotine replacement therapy
USPSTF: US Preventive Services Task Force
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.09.24; peer-reviewed by F Temidayo, D Patel; comments to author 24.10.24; revised version received 13.11.24; accepted 30.11.24; published 30.01.25.

Please cite as:
Abroms LC, Yousefi A, Wysota CN, Wu TC, Broniatowski DA
Assessing the Adherence of ChatGPT Chatbots to Public Health Guidelines for Smoking Cessation: Content Analysis
J Med Internet Res 2025;27:e66896
URL: https://www.jmir.org/2025/1/e66896
doi:10.2196/66896
PMID:

©Lorien C Abroms, Artin Yousefi, Christina N Wysota, Tien-Chin Wu, David A Broniatowski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

ChatGPT for Univariate Statistics: Validation of AI-Assisted Data Analysis in Healthcare Research

Michael R Ruta[1], BS; Tony Gaidici[1], BS; Chase Irwin[1], MS; Jonathan Lifshitz[1], PhD

University of Arizona College of Medicine – Phoenix, Phoenix, AZ, United States

Corresponding Author:
Michael R Ruta, BS
University of Arizona College of Medicine – Phoenix
475 N 5th St
Phoenix, AZ, 85004
United States
Phone: 1 602 827 2002
Email: mruta@arizona.edu

Abstract

Background: ChatGPT, a conversational artificial intelligence developed by OpenAI, has rapidly become an invaluable tool for researchers. With the recent integration of Python code interpretation into the ChatGPT environment, there has been a significant increase in the potential utility of ChatGPT as a research tool, particularly in terms of data analysis applications.

Objective: This study aimed to assess ChatGPT as a data analysis tool and provide researchers with a framework for applying ChatGPT to data management tasks, descriptive statistics, and inferential statistics.

Methods: A subset of the National Inpatient Sample was extracted. Data analysis trials were divided into data processing, categorization, and tabulation, as well as descriptive and inferential statistics. For data processing, categorization, and tabulation assessments, ChatGPT was prompted to reclassify variables, subset variables, and present data, respectively. Descriptive statistics assessments included mean, SD, median, and IQR calculations. Inferential statistics assessments were conducted at varying levels of prompt specificity ("Basic," "Intermediate," and "Advanced"). Specific tests included chi-square, Pearson correlation, independent 2-sample *t* test, 1-way ANOVA, Fisher exact, Spearman correlation, Mann-Whitney *U* test, and Kruskal-Wallis *H* test. Outcomes from consecutive prompt-based trials were assessed against expected statistical values calculated in Python (Python Software Foundation), SAS (SAS Institute), and RStudio (Posit PBC).

Results: ChatGPT accurately performed data processing, categorization, and tabulation across all trials. For descriptive statistics, it provided accurate means, SDs, medians, and IQRs across all trials. Inferential statistics accuracy against expected statistical values varied with prompt specificity: 32.5% accuracy for "Basic" prompts, 81.3% for "Intermediate" prompts, and 92.5% for "Advanced" prompts.

Conclusions: ChatGPT shows promise as a tool for exploratory data analysis, particularly for researchers with some statistical knowledge and limited programming expertise. However, its application requires careful prompt construction and human oversight to ensure accuracy. As a supplementary tool, ChatGPT can enhance data analysis efficiency and broaden research accessibility.

(J Med Internet Res 2025;27:e63550) doi:10.2196/63550

KEYWORDS

ChatGPT; data analysis; statistics; chatbot; artificial intelligence; biomedical research; programmers; bioinformatics; data processing

Introduction

ChatGPT is a conversational artificial intelligence (AI) created by OpenAI. It has quickly become an invaluable resource for researchers with capabilities that include reviewing literature, identifying gaps in research, and drafting papers [1-5]. With the recent addition of Python code interpretation to the ChatGPT environment, a surge of new research applications has emerged, particularly in data analysis [6]. Accompanied by innovations such as data upload and download, this new feature marks a considerable advancement toward individualized AI-assisted data analysis.

The accessibility of ChatGPT has the potential to democratize data analysis for nonspecialists and serve as a bridge between

XSL•FO
RenderX

programming knowledge and the growing demands of biomedical research [7]. Previously, ChatGPT had proven to be an asset to programmers, with capabilities ranging from debugging and annotating code to translating between coding languages [8-10]. However, for individuals without coding experience, deploying a local programming environment posed a significant challenge [7]. ChatGPT's ability to interpret code bypasses this barrier of entry, making language-to-code translation more accessible [6].

ChatGPT has excelled in some recent data analysis applications by completing bioinformatics exercises with high accuracy [6]. A preliminary analysis of ChatGPT's utility as a data analysis tool found that it provided results consistent with traditional biostatistical software [11]. However, the extent to which ChatGPT can assist with data analysis when the research question is cross-disciplinary remains unclear. For example, ChatGPT's utility was limited in addressing complex bioinformatics tasks, citing restricted file size and exclusive support for Python as key obstacles [7]. Other studies have recommended against using ChatGPT for statistical analysis due to incorrect answers, mislabeled data, and speculative results [12-14]. Continual improvements to the GPT models and refined prompts may improve the value proposition of ChatGPT in biostatistics.

As the capabilities of AI technology expand, the next step in democratizing research is to develop ChatGPT's applications for basic statistical analysis. Our research aims to provide a framework for performing preliminary inferential and descriptive statistics, as well as data management within ChatGPT. The ultimate goal is to avail a more equitable research landscape without the barriers presented by coding knowledge, thereby broadening access and innovation within the research community.

Methods

Dataset and Analyses

The 2019 National Inpatient Sample data from the National Healthcare Cost and Utilization Project was selected to assess ChatGPT's capabilities. This dataset was chosen to represent real-world observational data, similar to typical clinical research studies. Also, these data require processing before testing an array of hypotheses with various statistical tests. This is not a public-use dataset. For researchers interested in replicating these methods, this dataset is often available through institutional access.

Arbitrary inclusion criteria were selected to make the data both manageable and relevant to the aims of this study. ChatGPT's data and processing limits required uploaded files to be less than 100 megabytes. The inclusion criteria were individuals aged 41-70 years old who endured both a cerebral infarct and a myocardial infarction, with total hospital charges less than US $400,000. The resulting dataset included 2740 observations each with 5 attributes: age, gender, race, length of stay, and total charges. The 5 attributes included continuous, categorical, and binary data elements.

Common statistical methods for univariate analysis were used to evaluate ChatGPT's capabilities, including both descriptive and inferential statistics. Data processing, categorization, and tabulation were included as assessments of data management. Each assessment was conducted in a new GPT-4 window with the memory feature disabled. For each iteration, the Microsoft Excel data file and prompt were entered sequentially into the same ChatGPT window.

Ethical Considerations

The National Healthcare Cost and Utilization Project is a limited dataset, and it was deemed exempt by the institutional review board at the University of Arizona.

Data Management and Descriptive Statistics

For data processing, ChatGPT was prompted to create a new variable, reclassifying "race" from 4 dimensions (White, Black, Hispanic, and Other) to 2 dimensions (White and non-White). ChatGPT was also prompted to create a subset of patients aged 45 years to facilitate nonparametric testing later on. The decision to select 45-year-olds within the original sample was arbitrary, intended solely to create a smaller subpopulation suitable for nonparametric tests. For categorization, ChatGPT was prompted to categorize patients into 3 age cohorts (41-50, 51-60, and 61-70 years). Finally, data tabulation involved creating a series table to display the frequencies for the variables used in the categorization and processing steps. The prompts used for each element of data processing and categorization can be found in Multimedia Appendix 1.

For descriptive statistics, the methods tested were mean, SD, median, and IQR. ChatGPT was prompted to compute these values for male and female cohorts, as well as the entire dataset for the following variables: age, length of stay, and total charges. All descriptive statistics prompts were posed as directed questions without opportunity for interpretation. Descriptive statistics prompts can be found in Multimedia Appendix 2.

Inferential Statistics

Inferential statistics were performed as both parametric tests (chi-square, Pearson correlation, independent 2-sample *t* test, and 1-way ANOVA) and equivalent nonparametric tests (Fisher exact, Spearman rank order correlation, Mann-Whitney *U* test, and Kruskal-Wallis *H* test). The pairings of statistical tests are displayed in Table 1. A series of research questions were developed to guide statistical method selection (MS) toward a specific test, as shown in Table 2. Once refined, each research question was posed to ensure that a 2-tailed analysis was conducted, rather than a 1-tailed. The research questions and their proposed analytical tests were selected by one of the authors, a biostatistician, within the medical school.

XSL•FO
RenderX

Table 1. Parametric and nonparametric statistical analysis pairs.

Parametric test	Nonparametric test
Chi-square[a]	Fisher exact
Pearson correlation	Spearman rank order correlation
Independent 2-sample *t* test	Mann-Whitney *U*
1-way ANOVA	Kruskal-Wallis *H*

[a]Chi-square is a nonparametric test, as it does not require the assumption of normality. However, like the other tests listed in the "Parametric test" column, chi-square is better suited for larger sample sizes compared with its counterpart in the "Nonparametric test" column.

Table 2. Research questions and expected results.

Question	Method selection	Statistical assumptions	Statistical values
Is the distribution of White and non-White patients the same across genders?	Chi-square	• Independence • Expected values ≥5	• χ^2 (*df*) 1.385 (1) • *P*=.24
Is the distribution of White and non-White patients the same across genders for individuals who are 45 years old?	Fisher exact	• Independence	• Odds ratio=2.045 • *P*=.66
Is there a significant correlation between total charges and length of stay?[a]	Pearson correlation	• Linearity • Normality • Homoscedasticity • No extreme outliers • Paired data	• *r*=0.842 • *P*<.001
Is there a significant correlation between total charges and length of stay for individuals who are 45 years old?	Spearman rank order correlation	• Monotonicity • Nonnominal data	• ρ=0.447 • *P*=.02
Is there a significant difference in total charges between men and women?	Independent 2-sample *t* test	• Independence • Normality • Homoscedasticity • No extreme outliers • Continuous data	• *t*-stat (*df*) 0.185 (2707) • *P*=.85
Is there a significant difference in total charges between men and women who are 45 years old?	Mann-Whitney *U*	• Independence	• *U*-stat=91.000[b] • *P*=.17
Are there significant differences in length of stay across race categories?	1-way ANOVA	• Independence • Normality • Homoscedasticity • No extreme outliers • Continuous data	• *F*-stat=4.092 • *P*=.007
Are there significant differences in length of stay across race categories[c] for individuals who are 45 years old?	Kruskal-Wallis *H*	• Independence	• *H*-stat=1.984 • *P*=.37

[a]The variables were transformed for this test to improve normality and homoscedasticity.

[b]The *U*-stat was computed in Python and RStudio only, as SAS does not produce an equivalent value.

[c]The Hispanic group was excluded from this test due to insufficient observations.

The prompts used for inferential statistics were formulated at 3 levels of increasing specificity, reflecting a user's familiarity with statistics, ChatGPT, and Python. The lowest level was "Basic," where ChatGPT was provided with the research question, relevant variables, and necessary tasks. Textbox 1 shows a template used for "Basic" prompts. The "Intermediate" level used prompts that were more specific than "Basic" and included additional guidelines, including data cleaning steps and strategies for assessing the statistical assumptions (SA). The "Advanced" level prompted ChatGPT with the same information as "Intermediate" and a recommended statistical test. An overview of the components for each level of prompt is provided in Table 3. The individual prompts used for each test can be found in Multimedia Appendix 3.

XSL•FO
RenderX

Textbox 1. A "Basic" template prompt for inferential statistics.

I am analyzing variables related to hospital visits and demographics to answer the question, "_____?" The relevant variables are as follows:

- x (variable_type): description
- y (variable_type): description

Embody the role of an experienced biostatistician and complete the following tasks:

- Suggest the most relevant statistical method for analyzing this dataset.
- List and verify all of the critical assumptions that must be met to perform this statistical method.
- If any critical assumptions of the primary test are not met, identify a more appropriate alternative analysis. When suggesting a new test, list and verify the assumptions for this new test.
- Perform the most appropriate test and provide test statistics and *P* values to 3 decimal places.

Table 3. Components of inferential statistics prompts at varying knowledge levels.

	Basic	Intermediate	Advanced
Research question	✓	✓	✓
Variables and variable types	✓	✓	✓
Data clean step		✓	✓
Strategies to assess assumptions		✓	✓
Suggested method			✓
Tasks to perform (MS[a], SA[b], SV[c])	✓	✓	✓

[a]MS: method selection.

[b]SA: statistical assumptions.

[c]SV: statistical values.

Response Grading

The data analysis tasks were also performed in Python (Python Software Foundation), SAS (SAS Institute), and RStudio (Posit PBC) to generate an expected statistical output for comparison. For data processing, data categorization, data tabulation, and descriptive statistics, the expected statistical values (SV) can be found in Multimedia Appendix 4.

For inferential statistics, the expected statistical results are in the SV column of Table 2. The results were consistent among Python, SAS, and RStudio. If there were any discrepancies occurring due to algorithmic variations among the platforms, the Python output would have been designated as the "gold standard," as ChatGPT uses this language for its calculations.

The measured outcomes were data processing, data categorization, data tabulation, descriptive statistics, and inferential statistics. A trial consisted of a single interaction with ChatGPT, including one user input (prompt) and one ChatGPT output. Data processing and categorization were assessed through a direct comparison of the frequencies generated by ChatGPT to those in Python, SAS, and RStudio, each across 10 trials. Data tabulation was assessed through ChatGPT's ability to provide the correct columns, rows, and row totals during the processing and categorization trials (20 trials total).

Descriptive statistics were assessed through a direct comparison of the mean, SD, median, and IQR generated by ChatGPT to the expected values across 10 trials. Inferential statistics tests were evaluated on 3 criteria: MS, SA, and SV across 10 trials per statistical method (80 trials total). MS involved selecting the expected method, SA involved reporting the underlying assumptions critical to the statistical test, and SV involved calculating the test statistics and *P* values for the selected statistical method.

For data processing, data categorization, data tabulation, and descriptive statistics, the trial was coded as correct only if all values from ChatGPT's output were within 1% of the expected values. For inferential statistics, MS was coded as correct only if the final test selected by ChatGPT matched the expected test, SA was coded as correct only if the expected assumptions were assessed implicitly or explicitly, and SV was coded as correct only if both the test statistic and *P* value were within 1% of the expected output. A flowchart detailing how all trials were assessed is available in Multimedia Appendix 5.

All measured outcomes were coded as either correct or incorrect. Only a single prompt was provided to ChatGPT in each trial, with no further interaction. If ChatGPT requested further user input or provided multiple answers without indicating the best choice, that trial was coded as incorrect. If ChatGPT timed out during code interpretation, that trial was discarded. During inferential statistics trials, if MS was incorrect, then SA and SV were coded as incorrect as well. The choice of 10 trials per outcome was informed by previous studies on ChatGPT's performance in data analysis tasks, most of which assessed ChatGPT's performance in fewer trials [6,7,11]. In the absence

XSL•FO
RenderX

of an established standard, 10 trials were selected to capture variability in ChatGPT's responses.

Results

Data Management and Descriptive Statistics

ChatGPT performed data processing, categorization, and tabulation correctly across all trials. Variations in table aesthetics were inconsequential to the measured outcomes. For descriptive statistics, ChatGPT provided accurate means, SDs, medians, and IQRs for the 3 continuous variables across all attempts. The compiled tables and descriptive statistics for each variable can be found in Multimedia Appendix 4.

Inferential Statistics

When provided a "Basic" prompt, ChatGPT achieved 47.5% accuracy (38/80 attempts) on MS, 43.8% accuracy (35/80 attempts) on SA, and 32.5% accuracy (26/80 attempts) on SV. With "Intermediate" prompts, ChatGPT achieved 85.0% accuracy (68/80 attempts) on both MS and SA and 81.3% accuracy (65/80 attempts) on SV. With "Advanced" prompts, ChatGPT achieved 92.5% accuracy (74/80 attempts) across all MS, SA, and SV assessments. These results are displayed in Table 4 and Figure 1.

Table 4. Results for 10 attempts at inferential statistics with prompts at varying levels of specificity.

Test	Basic			Intermediate			Advanced		
	MS[a]	SA[b]	SV[c]	MS	SA	SV	MS	SA	SV
Chi-square	10	10	0	10	10	7	10	10	10
Fisher	9	9	9	10	10	10	10	10	10
Pearson	2	2	2	8	8	8	8	8	8
Spearman	9	6	8	7	7	7	10	10	10
t test	0	0	0	10	10	10	9	9	9
Mann-Whitney	7	7	6	8	8	8	9	9	9
ANOVA	1	1	1	7	7	7	8	8	8
Kruskal-Wallis	0	0	0	8	8	8	10	10	10

[a]MS: method selection.

[b]SA: statistical assumptions.

[c]SV: statistical values.

Figure 1. ChatGPT accuracy by prompt specificity.

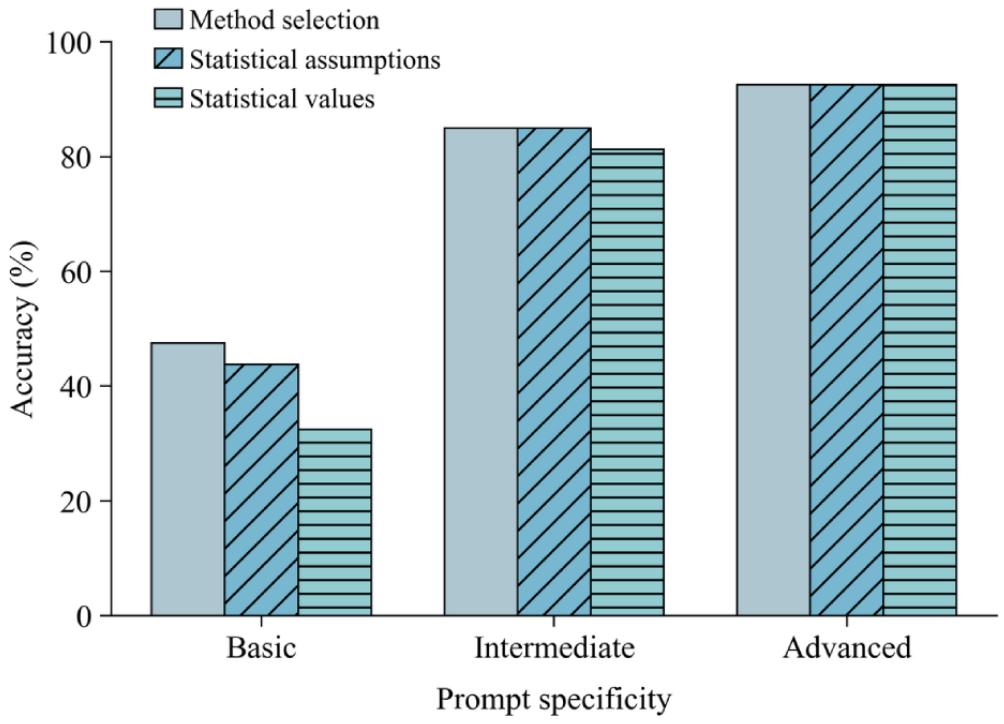

XSL•FO
RenderX

The errors encountered by ChatGPT across all inferential statistics trials included incorrect MS (n=51), coding error (n=13), incomplete response (n=9), incomplete assumptions (n=2), median imputed for missing value (n=1), and outliers incorrectly removed (n=1). A complete list of errors is listed in Table 5. Coding errors were limited to chi-square trials only and occurred due to a feature of the Python function. This function, chi2_contigency, contains a parameter for performing Yates correction, a method to prevent overestimation with small samples. As the parameter defaults to true, ChatGPT performed this correction despite the large sample size.

Table 5. Frequency of errors encountered by ChatGPT during inferential statistics trials.

Test	Basic	Intermediate	Advanced
Chi-square	• Coding error: 10	• Coding error: 3	—[a]
Fisher	• Incorrect method: 1	—	—
Pearson	• Incorrect method: 7 • Incomplete response: 1	• Incorrect method: 2	• Incorrect method: 2
Spearman	• Incomplete assumptions: 2 • Incomplete response: 1 • Imputed value: 1	• Incorrect method: 2 • Incomplete response: 1	—
t test	• Incorrect method: 10	—	• Incorrect method: 1
Mann-Whitney	• Incomplete response: 3 • Outliers removed: 1	• Incorrect method: 2	• Incomplete response: 1
ANOVA	• Incorrect method: 8 • Incomplete response: 1	• Incorrect method: 3	• Incorrect method: 2
Kruskal-Wallis	• Incorrect method: 10	• Incorrect method: 1 • Incomplete response: 1	—

[a]Not applicable.

Discussion

Principal Findings

Our investigation into the analytical capabilities of ChatGPT revealed a nuanced understanding of its implementation that, while promising, exhibits limitations that warrant careful consideration. ChatGPT's performance demonstrated proficiency in data processing, data categorization, data tabulation, and descriptive statistics. The most specific prompts improved response accuracy across inferential statistics trials. Based on our data, "Basic" prompts provided little value, with a low overall accuracy across all trials. Meanwhile, "Intermediate" prompts resulted in a similar overall accuracy compared with "Advanced" prompts despite the addition of a suggested statistical test in the "Advanced" prompts.

The variation observed across trials reflects the probabilistic nature and inherent unpredictability of ChatGPT [15]. Evaluating its accuracy as a tool requires repeated testing [16]. Given that GPT-4 is trained on over 1 trillion parameters [17], a definitive assessment of its performance is nearly impossible, as it would require countless trials to capture all possible response variations. This could explain the anomalies in the *t* test and Spearman trials, where less specific prompts sometimes outperformed more detailed ones.

The calculation errors from ChatGPT likely stem from challenges in implementing user instructions and interpreting intermediary outputs, rather than inherent issues with the Python-based analytical frameworks. With intentional instructions, most inaccuracies can be avoided. Although MS was the most common source of error, ChatGPT's responses included sufficient information for users to verify the suitability of the statistical method. Furthermore, in the "Intermediate" and "Advanced" trials, if ChatGPT selected the correct method, the remaining tasks were frequently completed accurately. Precise and specific prompts generally enhance accuracy; however, erroneous instructions may introduce further inaccuracies, as ChatGPT heavily relies on user input. Successful use of ChatGPT as a statistical doula requires balancing clear analysis goals with the flexibility needed for ChatGPT to present accurate information, without inadvertently introducing bias.

The power and democratization of ChatGPT allow for collaboration to further mitigate concerns about the analytical process. Although these trials were based on single prompts, a series of follow-up prompts by a user can identify determinations with regard to the data, the proposed analysis, or the code. For example, when asked about the Yates correction, ChatGPT acknowledged its use for small samples. To prevent errors from defaulted values like the Yates parameter, prompts should instruct ChatGPT to specify a function's parameters and provide the rationale for setting each parameter. This method was incorporated into the "Intermediate" and "Advanced" trials.

XSL•FO
RenderX

ChatGPT will inevitably produce inaccurate results that, if not carefully verified, may be detrimental to the research community. Therefore, all ChatGPT-generated analyses should be approached with caution. Based on this 10-trial approach, researchers are encouraged to incorporate multiple trials of the requested analysis in the statistical workflow using as specific instructions as possible. ChatGPT is not intended to be a standalone tool for data analysis; consultation with a biostatistician is essential for validating statistical approaches and ensuring reliable results.

Our dataset and assessments are limitations. Although we selected a moderately sized sample, it is still small relative to datasets used in complex analyses and may not assess the extent of ChatGPT's capabilities. In this report, the statistical analyses focused on univariate comparisons, where the capabilities for bivariate data analysis remain to be validated. Furthermore, due to the probabilistic nature of ChatGPT, a definitive performance assessment would require thousands of trials. Finally, we did not assess the validity of ChatGPT's statements outside of the chosen measures. When ChatGPT provided correct, incorrect, or extraneous information in the output, we did not count that against the assessment of accuracy to replicate the expected analysis.

Conclusion

ChatGPT has significant potential as a tool for exploratory data analysis, particularly for researchers who have some statistical knowledge but limited programming expertise. This paper is intended for individuals with a foundational understanding of statistics who could generate "Intermediate" prompts on their own. We hope that these individuals may use ChatGPT to perform preliminary analyses, helping them to understand their data, draw initial insights, and begin writing their papers while waiting for consultation with a statistician. These preliminary analyses are not meant to replace expert review but to accelerate the research process. Furthermore, ChatGPT may serve as an educational resource, helping researchers better understand statistical analyses and tackle unique problems [18].

Further advancements to ChatGPT will undoubtedly enhance its applicability and accuracy in statistical analysis. Data visualization is a crucial component of data analysis, and although still limited, ChatGPT's ability to process visual inputs through its "Vision" feature has already shown promise in interpreting statistical figures [19]. Furthermore, OpenAI is rapidly advancing its GPT models, and GPT-o1, the latest version, is already available for preview. This new model is reportedly better suited to "reason through complex tasks and solve more difficult problems in science, coding, and math" [20]. Improvements to both "Vision" and the GPT model will solidify ChatGPT's role as an asset for researchers.

We encourage researchers to leverage ChatGPT for the programming aspects of their statistical work while leaving the critical decisions to human expertise. By doing so, they can harness the full potential of ChatGPT as a supplementary tool in their research arsenal, ensuring that its application is both productive and scientifically sound. For researchers interested in applying ChatGPT to data analysis, we recommend following the best practices outlined in Textbox 2.

Textbox 2. Best practices for performing data analysis with ChatGPT.

Prompting

- Prompts should have specific and comprehensive goals. Researchers can benefit from creating a flexible template that can be adapted across different data analysis tasks. Additional details relevant to the current query should be included as needed. For those using ChatGPT for inferential statistics, prompts should, at a minimum, resemble the level of detail used in our "Intermediate" trials.

Refinement

- Approach data analysis like a discussion. Engage ChatGPT by asking it to identify any potential gaps in its knowledge or the inputs provided. Request feedback on how prompts can be improved for future trials. Ask ChatGPT to explain any decisions or necessary parameters.

Consistency

- At a minimum, attempt each trial 3 times to verify its consistency.

Verification

- ChatGPT-generated data analysis can be used to start drafting a paper before consulting a statistician. However, ensure that all statistical outputs and interpretations are reviewed by a statistician before publication.

Transparency

- Be transparent about using ChatGPT to assist with data analysis. Consider sharing the prompts used in your analysis to allow readers to replicate or refine the process.

Authors' Contributions

MRR and CI contributed to the study conception. MRR and TG handled the literature search, drafting, and data collection. CI and JL performed critical revisions.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data processing and categorization prompts.
[DOCX File , 8 KB - jmir_v27i1e63550_app1.docx]

Multimedia Appendix 2
Descriptive statistics prompts.
[DOCX File , 7 KB - jmir_v27i1e63550_app2.docx]

Multimedia Appendix 3
Inferential statistics prompts at varying levels of specificity.
[DOCX File , 16 KB - jmir_v27i1e63550_app3.docx]

Multimedia Appendix 4
Data tabulation and descriptive statistics results.
[DOCX File , 17 KB - jmir_v27i1e63550_app4.docx]

Multimedia Appendix 5
Trial design flowchart.
[DOCX File , 475 KB - jmir_v27i1e63550_app5.docx]

References

1. Bhargava DC, Jadav D, Meshram VP, Kanchan T. ChatGPT in medical research: challenging time ahead. Med Leg J 2023;91(4):223-225. [doi: 10.1177/00258172231184548] [Medline: 37802491]
2. Gödde D, Nöhl S, Wolf C, Rupert Y, Rimkus L, Ehlers J, et al. A SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis of ChatGPT in the medical literature: concise review. J Med Internet Res 2023;25:e49368 [FREE Full text] [doi: 10.2196/49368] [Medline: 37865883]
3. Khlaif ZN, Mousa A, Hattab MK, Itmazi J, Hassan AA, Sanmugam M, et al. The potential and concerns of using AI in scientific research: ChatGPT performance evaluation. JMIR Med Educ 2023;9:e47049 [FREE Full text] [doi: 10.2196/47049] [Medline: 37707884]
4. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
5. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Nature 2023;614(7947):224-226. [doi: 10.1038/d41586-023-00288-7] [Medline: 36737653]
6. Piccolo SR, Denny P, Luxton-Reilly A, Payne SH, Ridge PG. Evaluating a large language model's ability to solve programming exercises from an introductory bioinformatics course. PLoS Comput Biol 2023;19(9):e1011511 [FREE Full text] [doi: 10.1371/journal.pcbi.1011511] [Medline: 37769024]
7. Wang L, Ge X, Liu L, Hu G. Code interpreter for bioinformatics: are we there yet? Ann Biomed Eng 2024;52(4):754-756. [doi: 10.1007/s10439-023-03324-9] [Medline: 37482573]
8. Perkel JM. Six tips for better coding with ChatGPT. Nature 2023;618(7964):422-423. [doi: 10.1038/d41586-023-01833-0] [Medline: 37277596]
9. Shue E, Liu L, Li B, Feng Z, Li X, Hu G. Empowering beginners in bioinformatics with ChatGPT. Quant Biol 2023;11(2):105-108 [FREE Full text] [doi: 10.15302/j-qb-023-0327] [Medline: 37378043]
10. Xu D. ChatGPT opens a new door for bioinformatics. Quant Biol 2023;11(2):204-206 [FREE Full text] [doi: 10.15302/j-qb-023-0328] [Medline: 37900935]
11. Huang Y, Wu R, He J, Xiang Y. Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: a comparative analysis with SAS, SPSS, and R. J Glob Health 2024;14:04070 [FREE Full text] [doi: 10.7189/jogh.14.04070] [Medline: 38547497]
12. Ignjatović A, Stevanović L. Efficacy and limitations of ChatGPT as a biostatistical problem-solving tool in medical education in serbia: a descriptive study. J Educ Eval Health Prof 2023;20:28 [FREE Full text] [doi: 10.3352/jeehp.2023.20.28] [Medline: 37840252]
13. Jahangiri Y. Can Chat Generative Pretraining Transformer (ChatGPT) be used for statistical analysis of research data? J Vasc Interv Radiol 2023;34(12):2242-2246.e2. [doi: 10.1016/j.jvir.2023.09.010] [Medline: 37717655]

XSL•FO
RenderX

14. Ordak M. ChatGPT's skills in statistical analysis using the example of allergology: do we have reason for concern? Healthcare (Basel) 2023;11(18):2554 [FREE Full text] [doi: 10.3390/healthcare11182554] [Medline: 37761751]
15. Ruksakulpiwat S, Phianhasin L, Benjasirisan C, Ding K, Ajibade A, Kumar A, et al. Assessing the efficacy of ChatGPT versus human researchers in identifying relevant studies on mHealth interventions for improving medication adherence in patients with ischemic stroke when conducting systematic reviews: comparative analysis. JMIR mHealth uHealth 2024;12:e51526. [doi: 10.2196/51526] [Medline: 38710069]
16. Zhu L, Mou W, Hong C, Yang T, Lai Y, Qi C, et al. The evaluation of generative AI should include repetition to assess stability. JMIR mHealth uHealth 2024;12:e57978 [FREE Full text] [doi: 10.2196/57978] [Medline: 38688841]
17. Miao H, Li C, Wang J. A future of smarter digital health empowered by generative pretrained transformer. J Med Internet Res 2023;25:e49963 [FREE Full text] [doi: 10.2196/49963] [Medline: 37751243]
18. Magalhaes Araujo S, Cruz-Correia R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals. JMIR Med Educ 2024;10:e51151 [FREE Full text] [doi: 10.2196/51151] [Medline: 38506920]
19. Wang J, Ye Q, Liu L, Guo NL, Hu G. Scientific figures interpreted by ChatGPT: strengths in plot recognition and limits in color perception. NPJ Precis Oncol 2024;8(1):84 [FREE Full text] [doi: 10.1038/s41698-024-00576-z] [Medline: 38580746]
20. Introducing OpenAI o1-preview. OpenAI. 2024. URL: https://openai.com/index/introducing-openai-o1-preview [accessed 2024-12-19]

Abbreviations

AI: artificial intelligence
MS: method selection
SA: statistical assumptions
SV: statistical values

Edited by Q Jin; submitted 23.06.24; peer-reviewed by Y Huang, S Piccolo, EM Borges, F Chen; comments to author 31.08.24; revised version received 19.10.24; accepted 24.11.24; published 07.02.25.

Please cite as:
Ruta MR, Gaidici T, Irwin C, Lifshitz J
ChatGPT for Univariate Statistics: Validation of AI-Assisted Data Analysis in Healthcare Research
J Med Internet Res 2025;27:e63550
URL: https://www.jmir.org/2025/1/e63550
doi:10.2196/63550
PMID:

©Michael R Ruta, Tony Gaidici, Chase Irwin, Jonathan Lifshitz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Laypeople's Use of and Attitudes Toward Large Language Models and Search Engines for Health Queries: Survey Study

Tamir Mendel[1], PhD; Nina Singh[2], MD; Devin M Mann[3], MS, MD; Batia Wiesenfeld[4], PhD; Oded Nov[1], PhD

[1]Department of Technology Management and Innovation, Tandon School of Engineering, New York University, New York, NY, United States

[2]Department of Medicine, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

[3]Department of Population Health, Grossman School of Medicine, New York University, New York, NY, United States

[4]Department of Management and Organizations, Stern School of Business, New York University, New York, NY, United States

Corresponding Author:
Tamir Mendel, PhD
Department of Technology Management and Innovation
Tandon School of Engineering
New York University
2 Metrotech Center, Brooklyn
New York, NY, 11201
United States
Phone: 1 8287348968
Email: tamir.mendel@nyu.edu

Abstract

Background: Laypeople have easy access to health information through large language models (LLMs), such as ChatGPT, and search engines, such as Google. Search engines transformed health information access, and LLMs offer a new avenue for answering laypeople's questions.

Objective: We aimed to compare the frequency of use and attitudes toward LLMs and search engines as well as their comparative relevance, usefulness, ease of use, and trustworthiness in responding to health queries.

Methods: We conducted a screening survey to compare the demographics of LLM users and nonusers seeking health information, analyzing results with logistic regression. LLM users from the screening survey were invited to a follow-up survey to report the types of health information they sought. We compared the frequency of use of LLMs and search engines using ANOVA and Tukey post hoc tests. Lastly, paired-sample Wilcoxon tests compared LLMs and search engines on perceived usefulness, ease of use, trustworthiness, feelings, bias, and anthropomorphism.

Results: In total, 2002 US participants recruited on Prolific participated in the screening survey about the use of LLMs and search engines. Of them, 52% (n=1045) of the participants were female, with a mean age of 39 (SD 13) years. Participants were 9.7% (n=194) Asian, 12.1% (n=242) Black, 73.3% (n=1467) White, 1.1% (n=22) Hispanic, and 3.8% (n=77) were of other races and ethnicities. Further, 1913 (95.6%) used search engines to look up health queries versus 642 (32.6%) for LLMs. Men had higher odds (odds ratio [OR] 1.63, 95% CI 1.34-1.99; P<.001) of using LLMs for health questions than women. Black (OR 1.90, 95% CI 1.42-2.54; P<.001) and Asian (OR 1.66, 95% CI 1.19-2.30; P<.01) individuals had higher odds than White individuals. Those with excellent perceived health (OR 1.46, 95% CI 1.1-1.93; P=.01) were more likely to use LLMs than those with good health. Higher technical proficiency increased the likelihood of LLM use (OR 1.26, 95% CI 1.14-1.39; P<.001). In a follow-up survey of 281 LLM users for health, most participants used search engines first (n=174, 62%) to answer health questions, but the second most common first source consulted was LLMs (n=39, 14%). LLMs were perceived as less useful (P<.01) and less relevant (P=.07), but elicited fewer negative feelings (P<.001), appeared more human (LLM: n=160, vs search: n=32), and were seen as less biased (P<.001). Trust (P=.56) and ease of use (P=.27) showed no differences.

Conclusions: Search engines are the primary source of health information; yet, positive perceptions of LLMs suggest growing use. Future work could explore whether LLM trust and usefulness are enhanced by supplementing answers with external references and limiting persuasive language to curb overreliance. Collaboration with health organizations can help improve the quality of LLMs' health output.

(J Med Internet Res 2025;27:e64290) doi:10.2196/64290

XSL•FO
RenderX

KEYWORDS

large language model; artificial intelligence; LLMs; search engine; Google; internet; online health information; United States; survey; mobile phone

Introduction

Search engines, such as Google, democratized access to health information and changed the dynamics of the patient-provider relationship, with 72% of internet users in the United States looking on the web for health information [1-4]. Clinicians and health organizations initially had mixed reactions, including some early efforts to discourage patients from searching for health information on the web because of concerns that web-based information could make them misinformed or anxious [5-7]. Patients' use of internet-based health information enhances their understanding and their ability to manage their health conditions [8,9]. Over time, clinicians and health organizations started to partner with search engine companies to improve content reliability. For example, the World Health Organization has worked with Google to ensure that COVID-19 searches yield evidence-based information [10].

Large language models (LLMs) have the potential to replace internet searches for clinicians and patients. LLMs, such as ChatGPT, have demonstrated promising performance in clinical decision-making [11] and diagnosis [12]. Numerous studies have explored doctors' evaluation of LLMs as an information resource and diagnostic aid [13-18] and some research has begun to consider laypeople's attitudes toward and use of these tools, such as concerning ethical considerations, whether they are differentiable from doctor's responses, and the accuracy of symptom-checkers [19-21]. LLMs can potentially make information accessible to patients in more specific and personalized ways, but previous studies have yet to consider patient use of LLMs in comparison to search engines for health-related questions.

We surveyed laypeople in the United States to compare the use of LLMs and search engines for health queries, the types of health queries posed, and attitudes toward interactions with these tools.

Methods

Recruitment Procedures

US participants aged older than 18 years were recruited from Prolific, a web-based research participant platform, in February 2024. We programmed the survey questions in Qualtrics. The invitation in Prolific included the purpose and a brief description of this study, a link to the Qualtrics survey, and it informed participants that survey completion would take approximately 5 minutes for study 1 and 5-10 minutes for study 2. Participants completed study 1 in 1.21 (SD 1.04) minutes on average. Participants completed study 2 in 10.01 (SD 7.02) minutes on average. All questionnaires are presented in Multimedia Appendix 1.

Ethical Considerations

This study was reviewed and approved by New York University's institutional review board (IRB-FY2024-8278). Participants were presented with an informed consent form including a description of their role as participants in a research study to learn more about patterns of using search engines and LLMs for health queries. If they agreed to participate in this study, they were asked to complete a questionnaire about their experience using a search engine and an LLM for health questions. The research involves no more than minimal risk of harm to participants. At the end of the survey, participants were asked to provide a limited amount of personal data (Prolific ID) to provide compensation and asked about their willingness to participate in the second study if they qualified. Data were stored in password-protected computers of the faculty investigators and their assistants. Participants received compensation of US $0.25 in study 1 and US $2.50 in study 2.

Survey Questionnaire

A screening survey (study 1) identified the prevalence of participants' LLM and search engine use for health queries. Next, study 1 participants who reported using both LLMs and search engines to answer health questions were invited to participate in study 2, in which their use of these tools was compared. Study 2 participants were asked which types of health information they sought, and how the results affected their relationships with their providers. Likert scales from 1 (strongly disagree) to 7 (strongly agree) were used to evaluate the effect of LLMs and search engines on participants' relationships with health care providers.

Participants were asked about LLMs' versus search engines' perceived usefulness and ease of use. These two measures represent people's perception of whether a system would help them perform their tasks and whether the system is easy or difficult to use [22]. Trustworthiness is the extent to which people trust and rely on LLMs or search engines to answer health questions [23]. Relevance is the extent to which participants consider the source as offering output relevant to their health needs. We also asked whether the source's results were perceived to be biased or benefit advertisers. Finally, we asked about participants' reactions to LLM and search engine query results [24] and their perceptions of anthropomorphism [25]. Likert scales from 1 (strongly disagree) to 7 (strongly agree) were used to evaluate attitudes toward LLMs and search engines. A separate Likert scale ranging from 1 (not at all) to 7 (extremely) was used for items assessing participants' feelings about using LLMs and search engines. Finally, we asked about demographics: age, gender, education, income, use of LLMs for health, health status, and perceived tech skills. Perceived technical skills assessed participants' perceptions of their smartphone skills (1: not at all skilled to 7: expert; Multimedia Appendix 1).

Statistical Analysis

We compared the demographics of LLM users and nonusers using logistic regression. We used ANOVAs followed by a Tukey post hoc test to compare the frequency of use of LLMs and search engines for each of the types of health information

XSL•FO
RenderX

they sought. Using the Shapiro-Wilk test, we observed a nonnormal distribution for perceived usefulness, ease of use, relevance, bias, trustworthiness, feelings, perceptions of benefit to advertisers, and perceptions of anthropomorphism. Therefore, paired samples of Wilcoxon tests were used to compare the effects of LLMs and search engines on these perceptions.

Results

Overview

In total, 2002 participants responded to study 1 (see demographic information in Table 1). Furthermore, 300 randomly selected study 1 participants who reported using both LLMs and search engines for health queries completed study 2. Further, 281 participants were retained after indicating the use of LLMs for health-related queries and passing attention checks (see Table 1 for demographic information).

XSL•FO
RenderX

Table 1. Demographic characteristics of respondents in study 1 and study 2, with study 2 participants being a subset of those from study 1.

Variable	Study 1	Study 2
Respondents, n	2002	281
Age (years), mean (SD)	39.4 (13)	40.43 (13.17)
Gender, n (%)		
Male	904 (45.2)	161 (57.3)
Female	1045 (52.2)	112 (39.9)
Other	53 (1.6)	8 (2.8)
Race, n (%)		
Asian	194 (9.7)	35 (12.5)
Black	242 (12.1)	46 (16.4)
Hispanic	22 (1.1)	3 (1.1)
White	1467 (73.3)	187 (66.5)
Other	77 (3.8)	10 (3.6)
Education		
High school or less	578 (28.9)	72 (25.6)
Associate degree	249 (12.4)	39 (13.9)
Bachelor's degree	847 (42.3)	125 (44.5)
Graduate degree	328 (16.4)	45 (16)
Income (US $)		
Less than 50,000	715 (35.7)	106 (37.7)
50,000-99,999	694 (34.7)	103 (36.7)
More than 100,000	518 (25.9)	72 (25.6)
Unknown	75 (3.7)	0 (0)
Insurance, n (%)		
Private insurance	1166 (58.2)	159 (56.6)
Public insurance	592 (29.6)	92 (32.7)
No insurance	192 (9.6)	25 (8.9)
Other insurance	52 (2.6)	5 (1.8)
Using LLMs[a] for health, n (%)		
Yes	642 (32.1)	281 (100)
No	1360 (67.9)	0 (0)
Using search engines for health, n (%)		
Yes	1913 (95.6)	281 (100)
No	89 (4.4)	0 (0)
Health status, n (%)		
Excellent	268 (13.4)	44 (15.7)
Good	1231 (61.5)	158 (56.2)
Only fair	426 (21.3)	66 (23.5)
Poor	67 (3.3)	12 (4.3)
Prefer not to say	10 (0.5)	1 (0.4)
Perceived tech skills, mean (SD)	5.60 (1.05)	5.85 (1.04)

[a]LLM: large language model.

XSL•FO
RenderX

Among the participants from study 2, a total of 218 (77.6%) participants reported using ChatGPT [OpenAI] as their only LLM tool to answer health questions, 48 (17.1%) reported using both ChatGPT and other LLM tools (such as Bard [Google AI] or Copilot [Microsoft]), and only 15 (5.3%) indicated they used other LLM tools without using ChatGPT. For health-related searches, 228 (81.1%) participants used Google, 51 used both Google and other search engines (such as Bing [Microsoft]), and only 2 (<1%) used Bing alone for finding health answers. Therefore, most participants used ChatGPT as their primary LLM tool and Google search to answer health questions.

Demographic Comparison: LLM Users Versus Nonusers for Health-Related Questions

A logistic regression analysis revealed that men had higher odds (odds ratio [OR] 1.63, 95% CI 1.34-1.99; *P*<.001) of using LLMs for health-related questions compared to women. Individuals who reported their race as Black (OR 1.9, 95% CI 1.42-2.54; *P*<.001) and Asian (OR 1.66, 95% CI 1.19-2.30; *P*<.01) used LLMs for health-related questions more than participants who identified as White. Participants who perceived their health status as excellent (OR 1.46, 95% CI 1.1-1.93; *P*=.01) reported using LLMs for health-related questions more than those reporting good health status. Individuals with higher perceived technical proficiency have a higher likelihood of using LLMs for health-related questions (OR 1.26, 95% CI 1.14-1.39, *P*<.001; see Table S1 in Multimedia Appendix 1).

LLMs Versus Search Engines

In the follow-up study (study 2), most participants used search engines first (n=174, 61.9%) to answer health questions, but the second most common first source consulted was LLMs (n=39, 13.9% of participants), followed by doctor's appointment (n=19, 6.8%), calling their doctor (n=17, 6%), messaging their doctor (n=14, 5%), asking friends (n=7, 2.5%), going to the emergency department (n=6, 2.1%), and other options to answer health questions (n=5, 1.8%). They reported using search engines 6.91 (SD 2.67) times over the past year for health queries, compared to using LLMs 4.7 (SD 2.79) times, and other services fewer times (Figure S1 in Multimedia Appendix 1, *P*<.05). Finally, as shown in Figure 1, there was no difference in the likelihood that participants would use LLMs versus search engines to ask about symptoms (LLM: n=227; search engines: n=239), treatment (LLM: n=161; search engines: n=184), routine preventive care (LLM: n=117; search engines: n=145), diagnoses (LLM: n=95; search engines: n=133), and interpretation of test results (LLM: n=85; search engines: n=107), but they were more likely to use search engines than LLMs for administrative queries (LLM: n=60; search engines: n=112).

Figure 1. Distribution of health queries between LLMs and search engines among 281 participants who reported using both for health queries. The question asked was "When consulting [Google Search/ChatGPT] for health questions, which questions do you use it to answer?" LLM: large language model.

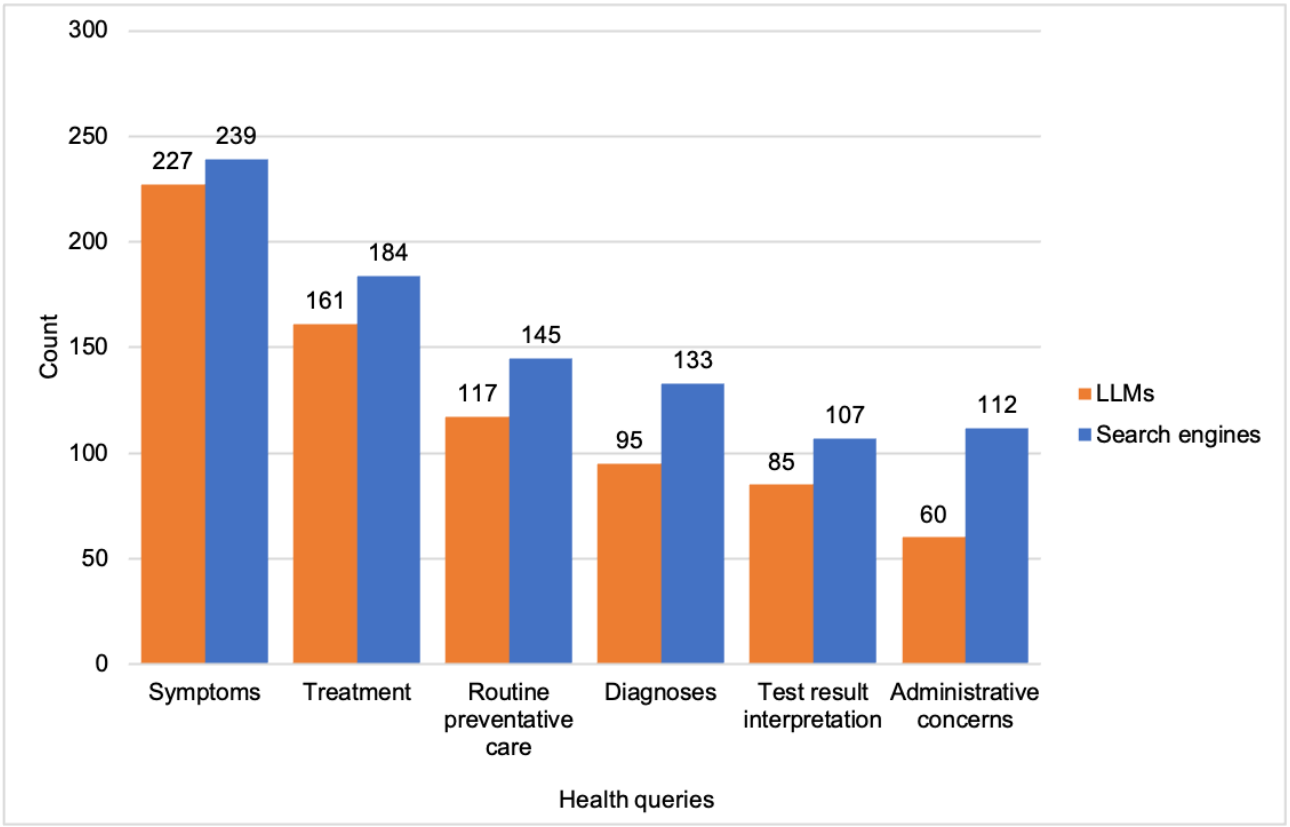

As can be seen in Figure 2, LLMs were perceived to be less useful than search engines for answering health-related questions (W=7214.5, *P*<.01). However, LLMs were also perceived as less biased (W=9533.5, *P*<.001) and less beneficial to advertisers than search engine results (W=20,708, *P*≤.001). Moreover, participants were less likely to view LLMs as being able to replace doctors in answering health care questions than search engines (W=9223, *P*=.003).

XSL•FO
RenderX

Figure 2. Comparison of LLMs and search engines among 281 participants who reported using both for health queries, using the Wilcoxon signed rank test for replacing "my doctor" (W=9223, *P*=.003), usefulness (W=7214.5, *P*<.01), relevance (W=7117, *P*=.07), result bias (W=9533.5, *P*<.001), and benefits to advertisers (W=20708, *P*<.001). LLM: large language model.

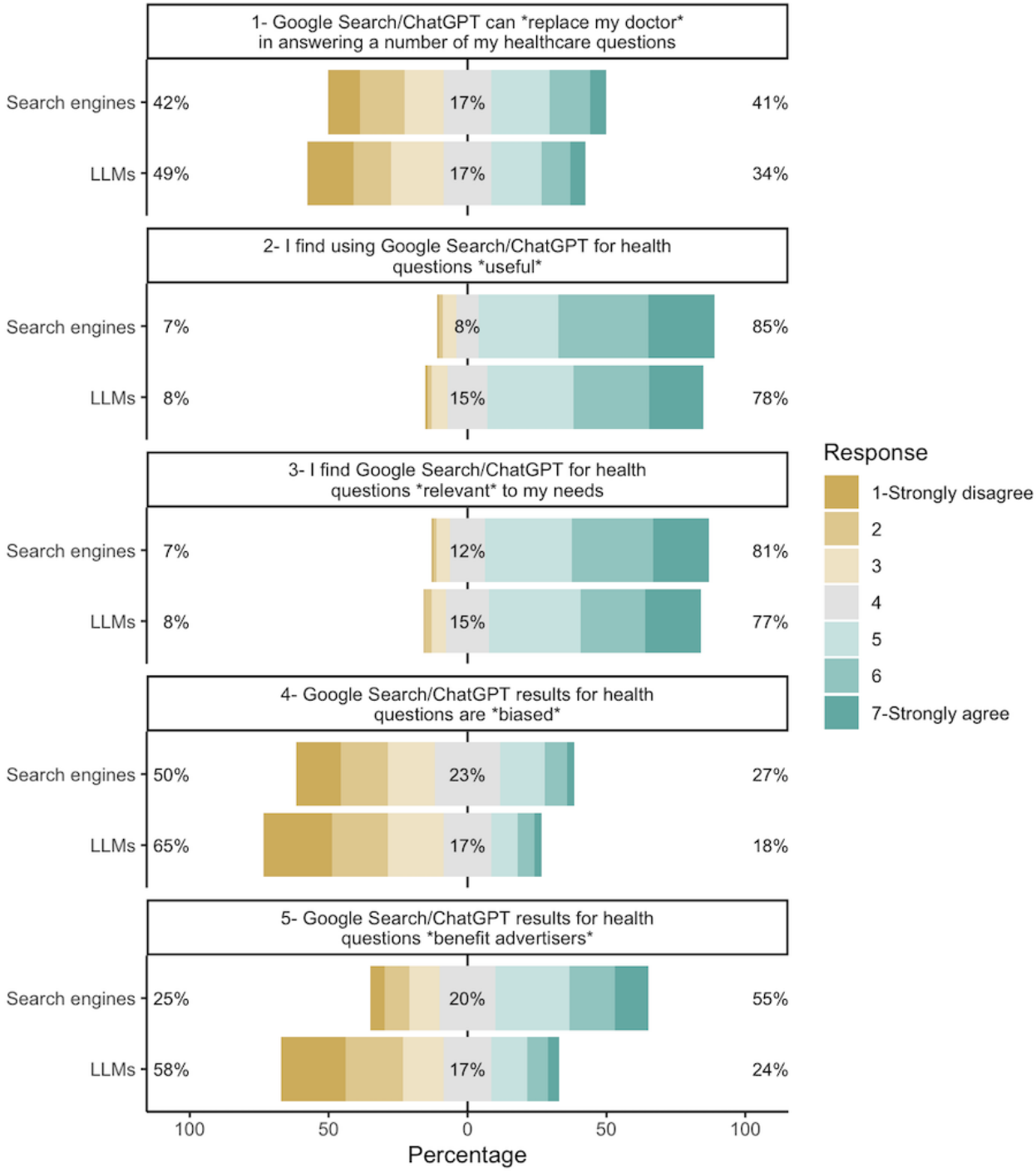

When searching for health information, results can reassure users but they can also make users more concerned [26]. As shown in Figure 3, LLMs elicited fewer negative feelings in response to the results of their queries than did search engines (W=19,496, *P*<.001). The LLMs also elicited more positive feelings in response to results than search engines did (W=9773.5, *P*=.01), but the median was lower than the midpoint of 4 (W=977.5, *P*<.001). On the other hand, while falling short of statistical significance, search engines were perceived to offer responses more relevant to users' needs than LLMs (W=7117, *P*=.07). There were no significant differences between ratings of LLMs and search engines in ease of use or trustworthiness of answers (*P*'s>.05). All paired samples of Wilcoxon tests are available in Table S2 in Multimedia Appendix 1. Finally, most participants (n=160, 57%) perceived LLMs as more human than search engines, but a minority perceived search engines as either more human (n=32, 11%) or equivalent to LLMs (n=32, 12%).

XSL•FO
RenderX

Figure 3. Comparison between LLMs and search engines among 281 participants who reported using both for health queries, based on the Wilcoxon signed rank test for negative feelings (W=19,496, *P*<.001) and positive feelings (W=9773.5, *P*=.01). LLM: large language model.

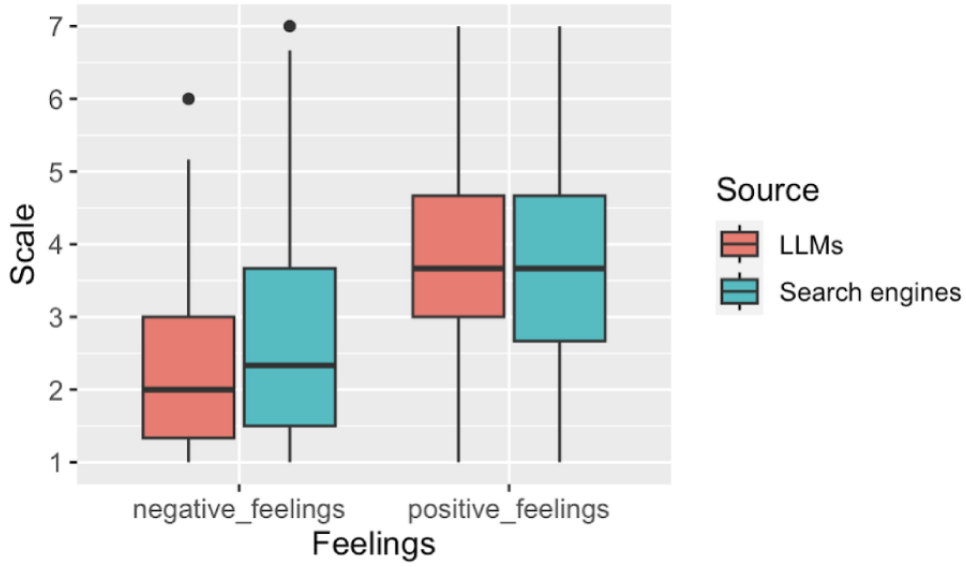

Discussion

Principal Findings

We compared the use of LLMs, search engines, and other sources for health queries and found that while the most common first source of information participants consulted was search engines, 13.9% of respondents consulted LLMs before any other source. Moreover, they were as likely to use LLMs as search engines for the most common health queries (symptoms, treatment, routine preventive care, diagnoses, and interpretation of test results), though they were more likely to use search engines than LLMs for less common administrative queries. In addition, LLMs elicited fewer negative feelings, such as less distress in response to the information provided, and were perceived as less useful and relevant than search engines. LLMs appeared more human-like, less biased, and less favorable toward advertisers. No significant differences were found between trust in LLMs and trust in search engines.

At this early stage of LLM adoption, participants indicated that LLMs were less useful and relevant than search engines for answering health-related questions. One key reason that search engines provide more useful and relevant health information compared to LLMs could include their greater transparency, such as providing external references to information sources that convey credibility and trustworthiness to users—features that are absent or inaccurate in many popular LLMs. The low perceived usefulness and relevance of LLMs can be attributed to issues such as incorrect statements, hallucinations, and ethical concerns [21,27,28], which can be mitigated through the involvement of domain experts, refining input prompts, and enhancing the fine-tuning of LLMs for specific tasks or domains [29].

Strengths of LLMs relative to search engines that emerged in our study include that LLMs elicited fewer negative feelings, such as less distress in response to the information provided. They also appeared more human-like, less biased, and less favorable toward advertisers. LLMs compared to search engines provide people with a feeling of human-like and personalized attention and understanding, enhancing digital communication that makes information delivery more intuitive. Evidence outside of health care suggests that these features make LLMs as persuasive as humans [30]. LLMs have advanced state-of-the-art performance in generating human-like text based on user health questions [19].

With respect to ease of use, there was no significant difference between LLMs and search engines for health queries. This is notable because search engines have been in regular use for decades while LLMs are a relatively new lay user-facing technology. The fact that respondents perceived that new and unfamiliar technology is as easy to use as an extensively used and highly familiar technology suggests that users may shift their use to LLMs further and relatively quickly.

Prior research has found that trust evaluations shape web-based health information-seeking behavior and compliance with health advice [31-33]. No significant differences were found between trust in LLMs and trust in search engines. A possible explanation for the lack of difference in trustworthiness between search engines and LLMs is the counterbalance between the two dimensions of trust: cognitive trust based on perceived accuracy and competence, and affective trust which depends on believing one's interests are protected [32]. People perceived that search engines, compared to LLMs, provide more helpful and relevant health information, reinforcing perceived accuracy and competence in health care information. However, search engines were also perceived to prioritize advertisers' interests more than LLMs did, likely diminishing users' trust in them. While LLMs excel at engaging with users, their output often lacks interpretation within a medical context [34] and they may sometimes fabricate facts or present incorrect information more convincingly and believably [35]. Future work can further analyze the effects of each trust dimension to better understand perceptions of LLMs.

XSL•FO
RenderX

Limitations

Our study has some key limitations. All of the participants were recruited from Prolific. Prolific participants tend to be technically proficient and may be more likely to use new technologies such as LLMs than the general population. Additionally, we screened for US residency so our findings may not generalize to other countries and languages. Future studies are needed to evaluate whether these findings generalize to broader populations. Finally, the current inferential analysis aggregates attitudinal reactions across varied types of health queries. A valuable future study would separately inquire about perceptions of technology (eg, usefulness, ease of use, or trust) for different types of health queries to enable more nuanced assessments.

Conclusion

Overall, most individuals still turn to search engines as their primary source for health-related questions, considering them more useful than LLMs. In the future, people may incorporate LLMs more routinely in their search for answers to health questions. In the case of LLMs, incorporating references and external sources of information would be a beneficial practice to increase trustworthiness. Just as clinicians and health organizations have partnered with search engine companies to enhance the reliability of health-related content [12], similar collaborations could enhance the quality of the health-related information provided by LLMs.

Acknowledgments

The authors received financial support from the US National Science Foundation (grants 1928614 and 2129076) for the submitted work.

Data Availability

The anonymized data analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional materials.
[DOCX File , 106 KB - jmir_v27i1e64290_app1.docx]

References

1. Fox S, Duggan M, Purcell K. Family caregivers are wired for health. Pew Research Center. Washington, D.C: Pew Research Center's Internet & American Life Project; 2013 Jun 20. URL: https://www.pewresearch.org/internet/2013/06/20/family-caregivers-are-wired-for-health/ [accessed 2025-01-15]
2. Luo A, Qin L, Yuan Y, Yang Z, Liu F, Huang P, et al. The effect of online health information seeking on physician-patient relationships: systematic review. J Med Internet Res 2022;24(2):e23354 [FREE Full text] [doi: 10.2196/23354] [Medline: 35142620]
3. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
4. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
5. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with internet-based health information? J Med Internet Res 2006;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]
6. Ahluwalia S, Murray E, Stevenson F, Kerr C, Burns J. 'A heartbeat moment': qualitative study of GP views of patients bringing health information from the internet to a consultation. Br J Gen Pract 2010;60(571):88-94 [FREE Full text] [doi: 10.3399/bjgp10X483120] [Medline: 20132702]
7. DDB. Don't google it. Ads of the World by CLIOS. 2014. URL: https://www.adsoftheworld.com/campaigns/don-t-google-it [accessed 2024-04-22]
8. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
9. Kassirer JP. Patients, physicians, and the Internet. Health Aff (Millwood) 2000;19(6):115-123. [doi: 10.1377/hlthaff.19.6.115] [Medline: 11192394]

XSL•FO
RenderX

10. Germani F, Pattison AB, Reinfelde M. WHO and digital agencies: how to effectively tackle COVID-19 misinformation online. BMJ Glob Health 2022;7(8):e009483 [FREE Full text] [doi: 10.1136/bmjgh-2022-009483] [Medline: 35918074]
11. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J Am Med Inform Assoc 2023;30(7):1237-1245 [FREE Full text] [doi: 10.1093/jamia/ocad072] [Medline: 37087108]
12. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
13. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad A, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. medRxiv Preprint posted online on February 26, 2023 [FREE Full text] [doi: 10.1101/2023.02.21.23285886] [Medline: 36865204]
14. Takita H, Walston SL, Tatekawa H, Saito K, Tsujimoto Y, Miki Y, et al. Diagnostic performance of generative AI and physicians: a systematic review and meta-analysis. medrxiv Preprint posted online on March 18, 2024. [doi: 10.1101/2024.01.20.24301563]
15. Sorin V, Barash Y, Konen E, Klang E. Large language models for oncological applications. J Cancer Res Clin Oncol 2023;149(11):9505-9508. [doi: 10.1007/s00432-023-04824-w] [Medline: 37160626]
16. Benary M, Wang XD, Schmidt M, Soll D, Hilfenhaus G, Nassir M, et al. Leveraging large language models for decision support in personalized oncology. JAMA Netw Open 2023;6(11):e2343689 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.43689] [Medline: 37976064]
17. Kao HJ, Chien TW, Wang WC, Chou W, Chow JC. Assessing chatGPT's capacity for clinical decision support in pediatrics: a comparative study with pediatricians using KIDMAP of Rasch analysis. Medicine (Baltimore) 2023;102(25):e34068 [FREE Full text] [doi: 10.1097/MD.0000000000034068] [Medline: 37352054]
18. Small WR, Wiesenfeld B, Brandfield-Harvey B, Jonassen Z, Mandal S, Stevens ER, et al. Large language model-based responses to patients' in-basket messages. JAMA Netw Open 2024;7(7):e2422399 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.22399] [Medline: 39012633]
19. Nov O, Singh N, Mann D. Putting chatGPT's medical advice to the (Turing) test: survey study. JMIR Med Educ 2023;9:e46939 [FREE Full text] [doi: 10.2196/46939] [Medline: 37428540]
20. Kopka M, von KN, Feufel M. Accuracy of online symptom-assessment applications, large language models, and laypeople for self-triage decisions: a systematic review. medRxiv Preprint posted online on September 14, 2024. [doi: 10.1101/2024.09.13.24313657]
21. Haltaufderheide J, Ranisch R. The ethics of chatGPT in medicine and healthcare: a systematic review on large language models (LLMs). NPJ Digit Med 2024;7(1):183 [FREE Full text] [doi: 10.1038/s41746-024-01157-x] [Medline: 38977771]
22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
23. Gulati S, Sousa S, Lamas D. Design, development and evaluation of a human-computer trust scale. Behav Inf Technol 2019;38(10):1004-1015. [doi: 10.1080/0144929x.2019.1656779]
24. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
25. Li X, Sung Y. Anthropomorphism brings us closer: the mediating role of psychological distance in user–AI assistant interactions. Comput Hum Behav 2021;118:106680. [doi: 10.1016/j.chb.2021.106680]
26. Brown RJ, Skelly N, Chew-Graham CA. Online health research and health anxiety: a systematic review and conceptual integration. Clin Psychol: Sci Pract 2020;27(2):e12299. [doi: 10.1111/cpsp.12299]
27. He Z, Bhasuran B, Jin Q, Tian S, Hanna K, Shavor C, et al. Quality of answers of generative large language models versus peer users for interpreting laboratory test results for lay patients: evaluation study. J Med Internet Res 2024;26:e56655 [FREE Full text] [doi: 10.2196/56655] [Medline: 38630520]
28. Guo Z, Lai A, Thygesen JH, Farrington J, Keen T, Li K. Large language models for mental health applications: systematic review. JMIR Ment Health 2024;11:e57400 [FREE Full text] [doi: 10.2196/57400] [Medline: 39423368]
29. Ahmad MA, Yaramis I, Roy TD. Creating trustworthy LLMs: dealing with hallucinations in healthcare AI. arXiv:2311.01463 Preprint posted online on September 26, 2023. [doi: 10.48550/arXiv.2311.01463]
30. Carrasco-Farré C. Large language models are as persuasive as humans, but how. About the cognitive effort and moral-emotional language of LLM arguments. arXiv:2404.09329. Preprint posted online on April 14, 2024 2024. [doi: 10.48550/arXiv.2404.09329]
31. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
32. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. J Med Internet Res 2018;20(8):e253 [FREE Full text] [doi: 10.2196/jmir.9364] [Medline: 30120087]
33. Mendel T, Nov O, Wiesenfeld B. Advice from a doctor or AI? Understanding willingness to disclose information through remote patient monitoring to receive health advice. Proc ACM Hum-Comput Interact 2024;8(CSCW2):1-34. [doi: 10.1145/3686925]

XSL•FO
RenderX

34. GPT-4 technical report. OpenAI. 2023. URL: https://cdn.openai.com/papers/gpt-4.pdf [accessed 2025-01-10]
35. Fogg B, Marshall J, Laraki O, Osipovich A, Varma C, Fang N, et al. What makes web sites credible?: a report on a large quantitative study. 2001 Presented at: CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2001 March 01; Seattle Washington USA p. 61-68. [doi: 10.1145/365024.365037]

Abbreviations

LLM: large language model
OR: odds ratio

Edited by A Mavragani; submitted 14.07.24; peer-reviewed by K You, R Keshava Murthy; comments to author 27.09.24; revised version received 11.11.24; accepted 25.12.24; published 13.02.25.

Please cite as:
Mendel T, Singh N, Mann DM, Wiesenfeld B, Nov O
Laypeople's Use of and Attitudes Toward Large Language Models and Search Engines for Health Queries: Survey Study
J Med Internet Res 2025;27:e64290
URL: https://www.jmir.org/2025/1/e64290
doi:10.2196/64290
PMID:39946180

©Tamir Mendel, Nina Singh, Devin M Mann, Batia Wiesenfeld, Oded Nov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessment of the Efficiency of a ChatGPT-Based Tool, MyGenAssist, in an Industry Pharmacovigilance Department for Case Documentation: Cross-Over Study

Alexandre Benaïche[1], PharmD; Ingrid Billaut-Laden[1], PhD; Herivelo Randriamihaja[1], PharmD; Jean-Philippe Bertocchio[2,3], MD, PhD

[1]Bayer Healthcare SAS France, Lille, France

[2]SKEZI, Les Papèteries—Image Factory, Annecy, France

[3]Assistance Publique—Hôpitaux de Paris, Pitié-Salpêtrière Hospital, Paris, France

Corresponding Author:
Alexandre Benaïche, PharmD
Bayer Healthcare SAS France
1 Rue Claude Bernard
Lille, 59000
France
Phone: 33 320445962
Email: benaichealexandre@gmail.com

Abstract

Background: At the end of 2023, Bayer AG launched its own internal large language model (LLM), MyGenAssist, based on ChatGPT technology to overcome data privacy concerns. It may offer the possibility to decrease their harshness and save time spent on repetitive and recurrent tasks that could then be dedicated to activities with higher added value. Although there is a current worldwide reflection on whether artificial intelligence should be integrated into pharmacovigilance, medical literature does not provide enough data concerning LLMs and their daily applications in such a setting. Here, we studied how this tool could improve the case documentation process, which is a duty for authorization holders as per European and French good vigilance practices.

Objective: The aim of the study is to test whether the use of an LLM could improve the pharmacovigilance documentation process.

Methods: MyGenAssist was trained to draft templates for case documentation letters meant to be sent to the reporters. Information provided within the template changes depending on the case: such data come from a table sent to the LLM. We then measured the time spent on each case for a period of 4 months (2 months before using the tool and 2 months after its implementation). A multiple linear regression model was created with the time spent on each case as the explained variable, and all parameters that could influence this time were included as explanatory variables (use of MyGenAssist, type of recipient, number of questions, and user). To test if the use of this tool impacts the process, we compared the recipients' response rates with and without the use of MyGenAssist.

Results: An average of 23.3% (95% CI 13.8%-32.8%) of time saving was made thanks to MyGenAssist (P<.001; adjusted R^2=0.286) on each case, which could represent an average of 10.7 (SD 3.6) working days saved each year. The answer rate was not modified by the use of MyGenAssist (20/48, 42% vs 27/74, 36%; P=.57) whether the recipient was a physician or a patient. No significant difference was found regarding the time spent by the recipient to answer (mean 2.20, SD 3.27 days vs mean 2.65, SD 3.30 days after the last attempt of contact; P=.64). The implementation of MyGenAssist for this activity only required a 2-hour training session for the pharmacovigilance team.

Conclusions: Our study is the first to show that a ChatGPT-based tool can improve the efficiency of a good practice activity without needing a long training session for the affected workforce. These first encouraging results could be an incentive for the implementation of LLMs in other processes.

(J Med Internet Res 2025;27:e65651) doi:10.2196/65651

XSL•FO
RenderX

KEYWORDS

MyGenAssist; large language model; artificial intelligence; ChatGPT; pharmacovigilance; efficiency

Introduction

Artificial intelligence (AI) is currently the center of attention in a lot of disciplines. Health is no exception. For instance, AI could be useful for the analysis of pictures for detecting local signs of catheter-associated infections [1] or for reducing both the time and costs required for drug discovery [2]. It could also play a role in the education of future health care professionals [3].

Studies from previous years already provided evidence that the use of AI would benefit pharmaceutical companies [4-6]. However, most of those studies used complex applications of AI to extract or synthesize data from documents like health records [7,8] or social media content [9-11].

Although not yet reachable to a wide audience, large language models (LLMs) offer possibilities to integrate AI in several domains without requiring a long training of the workforce, thanks to its accessibility and ease of use. While everyone can use some LLMs that are open access—including ChatGPT since November 2022 [12]—getting an adequate answer sometimes requires gradually improving the prompts. Hence, the integration of LLMs in the work environment can provide new soft skills at a low cost and promote the empowerment of workers. Several studies already showed that pharmacists and other health care professionals are willing to use LLMs as a help in their work [13-15].

However, the use of LLMs raises concerns about data privacy. Recent news showed that the information provided by some users in their prompts could be transferred by ChatGPT to others [16]. This threat can be answered by companies with the opportunity to develop their own LLM for internal use.

Hence, Bayer AG launched on September 21, 2023, its own internal LLM called MyGenAssist, based on ChatGPT technology. Bayer, as a market authorization holder in Europe, has the duty to set a system to collect, register, and analyze adverse events related to its products according to European [17] and French good pharmacovigilance practices (GVPs) [18] and the European Directive 2001/83/CE [19]. Tasks related to this imperative follow internal procedures, which make some of them repetitive and time-consuming. A semiautomatization of these tasks thanks to an LLM could make them less arduous for workers and procure more time for other activities at higher added value.

In the pharmacovigilance field, the Giens Workshop [20] in 2022 aimed to initiate a reflection on the integration of AI in this area and highlighted the actions implying the writing of a letter to contact a patient or a health care professional as a good opportunity for this. To the best of our knowledge, no study described LLMs' use cases in pharmacovigilance from a daily and practical perspective.

In this study, we aim to determine whether the use of MyGenAssist in the pharmacovigilance case documentation process can provide an improvement in efficiency (ie, save time without leading to a decrease in the answer rate).

Methods

Description of the Activity and Materials Used

Description of the Case Documentation Process and Potential Contribution of MyGenAssist

The current internal process for collecting, recording, and documenting spontaneous case reports of adverse drug events at Bayer France is described here (Figure 1): reporters can notify Bayer about cases by phone, mail, or electronic means. The report is compiled into a source document, which is then added to an internal pharmacovigilance database by a local pharmacovigilance officer. The source document is analyzed by a case evaluator who is part of the company's global pharmacovigilance team. Based on the global analysis, the pharmacovigilance officer is in charge of defining the final list of relevant questions to complete the case. The GVPs, written by the French Agence Nationale de Sécurité du Médicament et des produits de santé, add a duty to contact the reporter a second time if the first attempt of contact was unsuccessful [18].

XSL•FO
RenderX

Figure 1. Description of the process for documenting spontaneous notifications of adverse drug events.

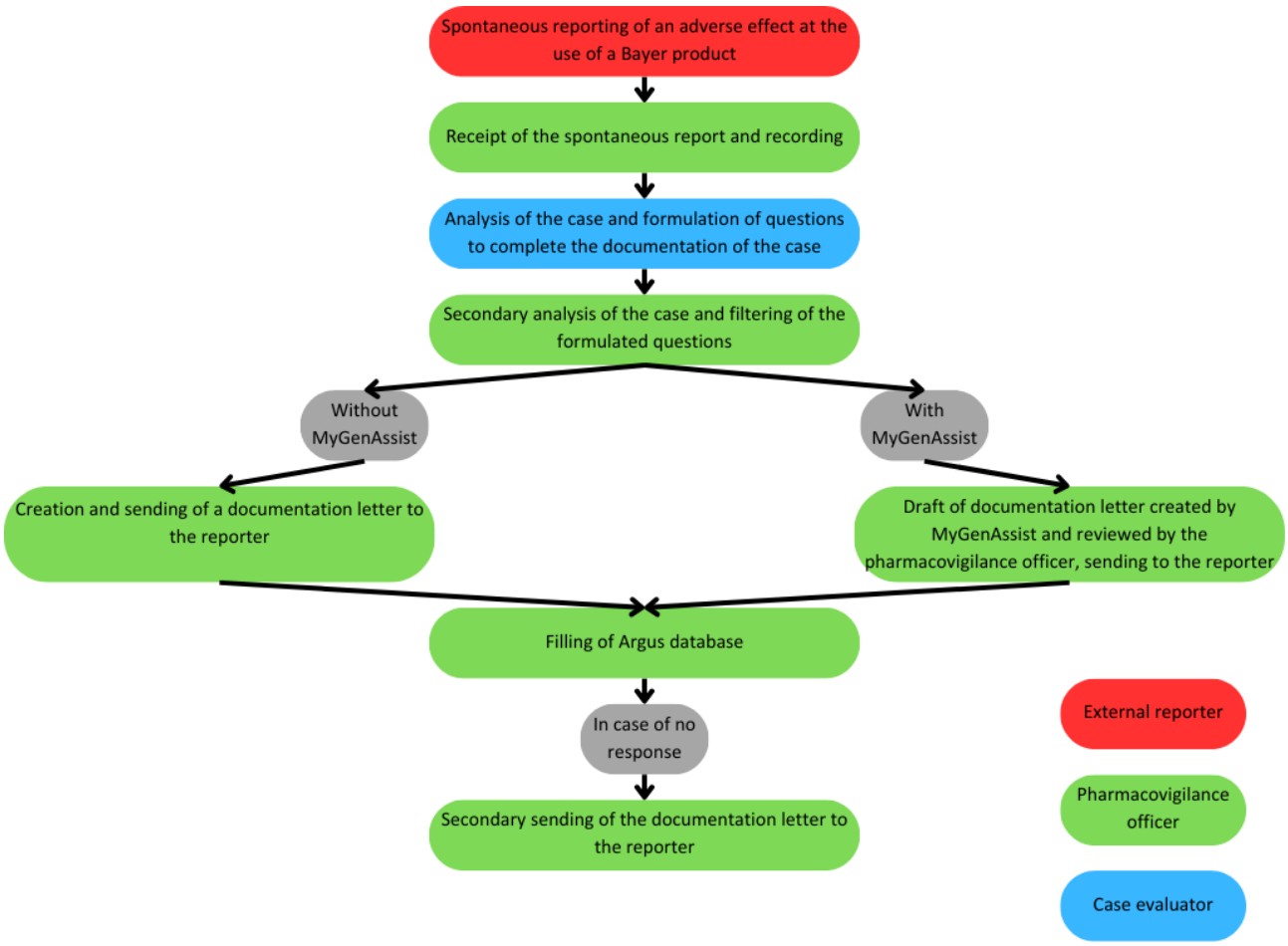

MyGenAssist was added to this process at the stage of drafting documentation letters. This LLM was used to write a first draft of the letters, which was then reviewed and modified if needed by the pharmacovigilance officer.

Software and Applications Used

MyGenAssist is an LLM internal to Bayer based on the technology of ChatGPT-4 Turbo (OpenAI), a conversational agent based on generative AI [21]. The model used here has no differences compared to the one available to the public, apart from the fact that MyGenAssist is securely hosted by Bayer. All manipulations performed on this internal LLM can be reproduced on ChatGPT, which is accessible to the public on the web. Here, MyGenAssist was used by ourselves as final users; the prompt was created without any fine-tuning or modification of its algorithm.

Pharmacovigilance cases reported to Bayer are compiled on an internal pharmacovigilance database supported by the Argus pharmacovigilance case management software developed by Oracle [22]. This software assigns a specific reference to each case. It contains all documents related to the case as well as information regarding the actions taken to analyze the case, such as attempts to contact the reporter.

The questions formulated by the case evaluators following their analysis of the case are listed on the FAST application (Bayer internal application). The pharmacovigilance officer can select the relevant questions and add others if needed.

Use of MyGenAssist for the Activity

Pharmacovigilance case documentation letters were written according to templates validated by an internal process. Two templates exist depending on the type of recipient (patients or health care professionals). The pharmacovigilance officer could adapt the template in function of the pharmacovigilance case. To make the LLM predraft the pharmacovigilance case documentation letters, the 2 templates were used. The elements of the templates corresponding to information specific to each case were replaced by titles placed into brackets in the letter templates. These 2 letter templates were then provided via the following prompt as shown in Multimedia Appendix 1 (original version in French; the translation corresponds to Multimedia Appendix 2, an example of a query is shown in Multimedia Appendix 3, and its translation is shown in Multimedia Appendix 4).

Staff Training

To train staff to use MyGenAssist specifically for this activity, a training session was planned before initiating its use. The earlier-described operating mode was given to the pharmacovigilance officers. A training phase was also planned to verify their good understanding of the use of the tool. The operating procedure was also formalized for the use of potential

XSL•FO
RenderX

new arrivals. To allow pharmacovigilance officers to report any difficulties or suggestions for improvement regarding the operating procedure, meetings were scheduled every other week until the termination of the study.

Data Collection

The study was conducted between January 2 and May 3, 2024. The time spent on each case was measured from the moment the pharmacovigilance case was acknowledged until the letter was sent and the Argus database was filled out.

Over an initial period of 9 weeks, the time spent on drafting each documentation letter, without the use of MyGenAssist, was measured. Over a second equivalent period, the time spent on the same task, but using the LLM as a drafting aid, was also measured. During the 2 periods, apart from the addition of MyGenAssist, no significant changes in the management of the activity took place: 3 workers, including 2 experimented participants and a newcomer were present throughout the 4 months of the study, and the process did not undergo any notable modifications.

Statistical Analysis

Calculation of Time Savings

To determine the average time saved for each case thanks to MyGenAssist, a multiple linear regression was performed, in which the explained variable or outcome was the time spent per case. To highlight the impact of MyGenAssist, the parameters that could potentially influence the time spent per case were researched and selected on the basis of a review of the pharmacovigilance documentation process with experienced pharmacovigilance officers who defined, based on their daily practice, which elements could weigh on the time spent. Because of the company's specificities in the pharmacovigilance documentation process, the judgment of pharmacovigilance officers was then the unique criteria to add these parameters to the model. Among them, all parameters that could be modelized were included in the regression model as independent or explanatory variables (Figure 2):

$$Y=a_1X_1+a_2X_2+a_3X_3+a_4X_4+a_5X_5+b$$

Figure 2. Causal diagram of factors potentially impacting the time spent on the activity, built with Dagitty software (version 3.1).

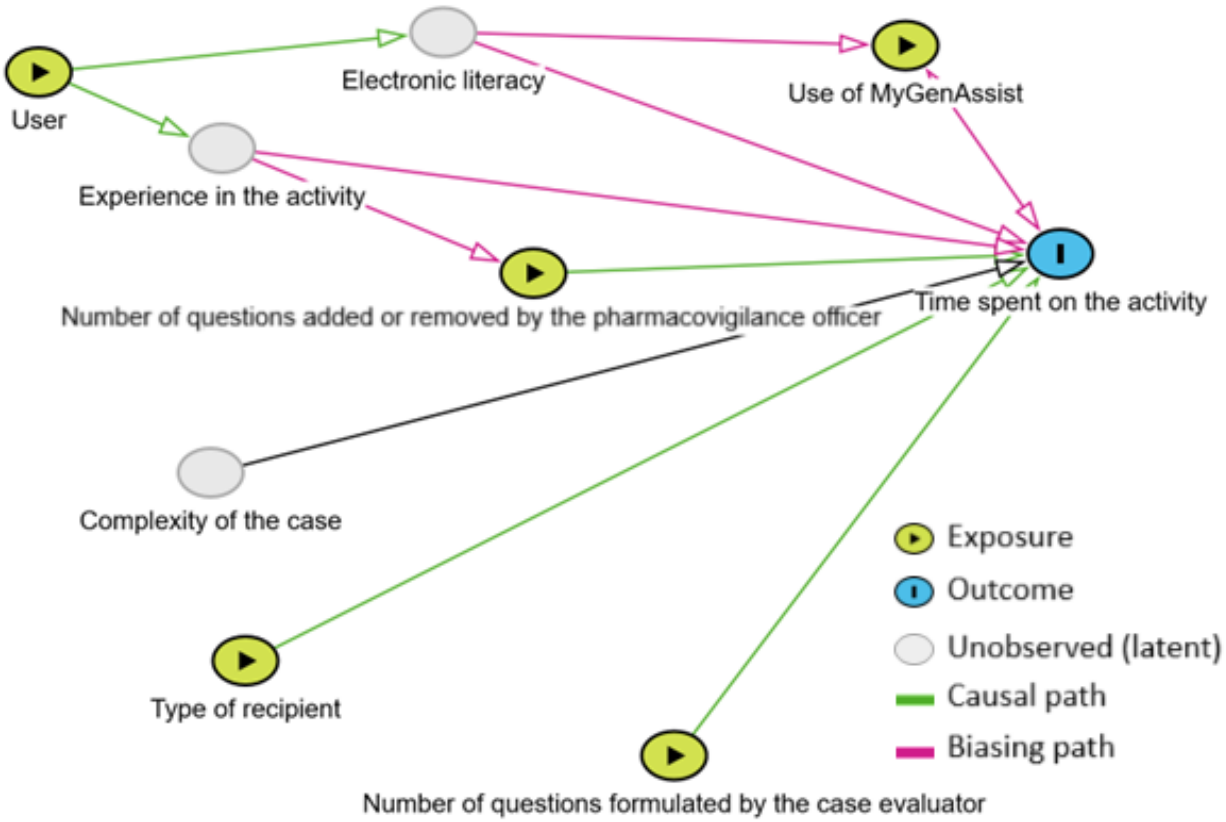

Y is the time spent per case (continuous quantitative variable, expressed in minutes). X_1 is the use of MyGenAssist (dichotomous categorical variable: use or no use). X_2 is the type of recipient (dichotomous categorical variable: patient or health care professional). The letter template differs significantly depending on this parameter. It may also be necessary to simplify the questions contained in the letter for patients, leading to more time spent on the task. X_3 is the number of questions formulated by the case evaluator (discrete quantitative variable). This parameter can also impact the time spent. Each question is asked in English. Without the use of MyGenAssist, a French translation is automatically provided by the internal software FAST, but the translations need to be reviewed and the questions sometimes need to be adapted according to the recipient. This

XSL•FO
RenderX

adaptation is a step in the workflow whose duration could be correlated to the number of questions. When using MyGenAssist, this latter also provides a translation of the questions, but some adaptation work is still necessary. X_4 is the number of questions added or removed by the pharmacovigilance officer (discrete quantitative variable). The pharmacovigilance officer has the option to remove questions prepared by the case evaluator considering them irrelevant in the context of the case. Conversely, this latter can also add questions to obtain information considered as necessary. Adding questions or considering the removal of questions from the case processor leads to additional reflection time that can impact the overall duration of the task. X_5 is the user (categorical variable: user 1, 2, or 3). The user was taken into account. The personal characteristics of the user will influence the time spent on the task. a_1, a_2, a_3, a_4, and a_5 are, respectively, the coefficients of the variables X_1, X_2, X_3, X_4, and X_5, and b is the intercept. Three parameters in this causal diagram were not included in the model because they cannot be measured: the experience in the activity, the electronic literacy of the user, and the complexity of the case.

The multiple linear regression was performed using the R software (R Foundation for Statistical Computing) with the "lm" function, which can be found in the *stats* package (all R equations used in this study are described in Multimedia Appendix 5). The outliers were researched thanks to the function "check_outliers" from the *performance* package [23].

The linear regression was completed with a hierarchical regression to study the relevance of the independent variables in the model. It was performed with the function "stepAIC" from the package *mass* [24] on R. The hierarchical regression is a backward stepwise Akaike information criterion.

Cases were randomly distributed among the pharmacovigilance officers without considering the context of the case, the type of recipient, and the number of questions a priori. The number of questions formulated was not related to the type of recipient.

Measurement of Effectiveness

The reliability of the LLM was not studied because the use of MyGenAssist was intended only as an aid in drafting documentation letters. However, a measure of the effectiveness of the task with or without the tool was carried out. For this, a comparison of the proportion of cases for which a response from the recipient was obtained after the first and the second contact attempt was made, depending on whether MyGenAssist was used or not. A subgroup analysis was also realized in function of the type of recipient (physician or patient). The comparison was made on R thanks to the "chisq.test" function that can be found in the *stats* package. In the subgroup analysis, the same comparison was done but specifically for cases whose recipient was a physician on one hand, and on the other hand, for cases whose recipient was a patient.

Ethical Considerations

This study was not considered as being related to human research in health, as it clearly focused on evaluating work performance with or without an AI-based tool. Regulations in France require such studies to be in compliance with the General Data Protection Regulation (ie, European Regulation 2016/679) [25]. The data used in this study were retrieved first during the normal course of the pharmacovigilance activities and reused for the study as authorized by General Data Protection Regulation [25] and after having received oral consent from pharmacovigilance workers who were informed of the objective of the study and how the data will be used.

Results

Statistical Study

The study included 122 cases (48 without the use of MyGenAssist and 74 with its use). The average time spent on each case was 19.05 (SD 6.05; 95% CI 17.97-20.12) minutes. The average time for cases handled was 22.25 (SD 7.62; 95% CI 20.09-24.41) minutes without the use of MyGenAssist and 16.97 (SD 3.50; 95% CI 16.17-17.77) minutes when the LLM was used (Figure 3). When all factors that potentially impact the time spent on the activity are taken into account in a multiple linear regression model, an average of 23.3% (95% CI 13.8%-32.8%; P<.001) of time savings were realized for each case thanks to MyGenAssist (adjusted R^2=0.286; df=115; Figure 4). All explanatory variables of the linear regression model got a P value inferior to .05 apart from the number of questions added or removed by the pharmacovigilance officer (P=.05).

The hierarchical regression performed found that all the parameters used as explanatory variables were useful to get the best model (the lowest Akaike information criterion is obtained with all the parameters included in the model; Table 1).

XSL•FO
RenderX

Figure 3. Time spent on each case in function of the use of MyGenAssist.

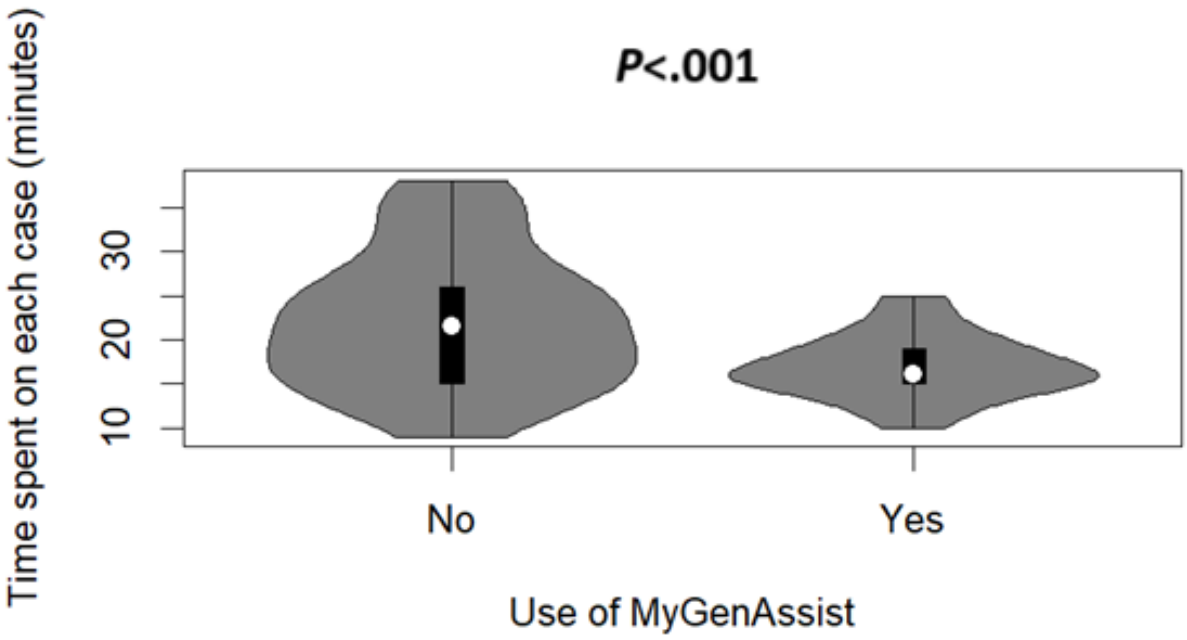

Figure 4. Influence of explanatory variables on the time spent on a case. A logarithmic transformation was performed to fulfill the conditions of the linear regression model (the actual model is log *t*=a1X1+a2X2+a3X3+a4X4+a5X5+b). "log T" is the logarithmic transformation of the variable T; hence, the result obtained is a percentage of difference of time in function of each parameter.

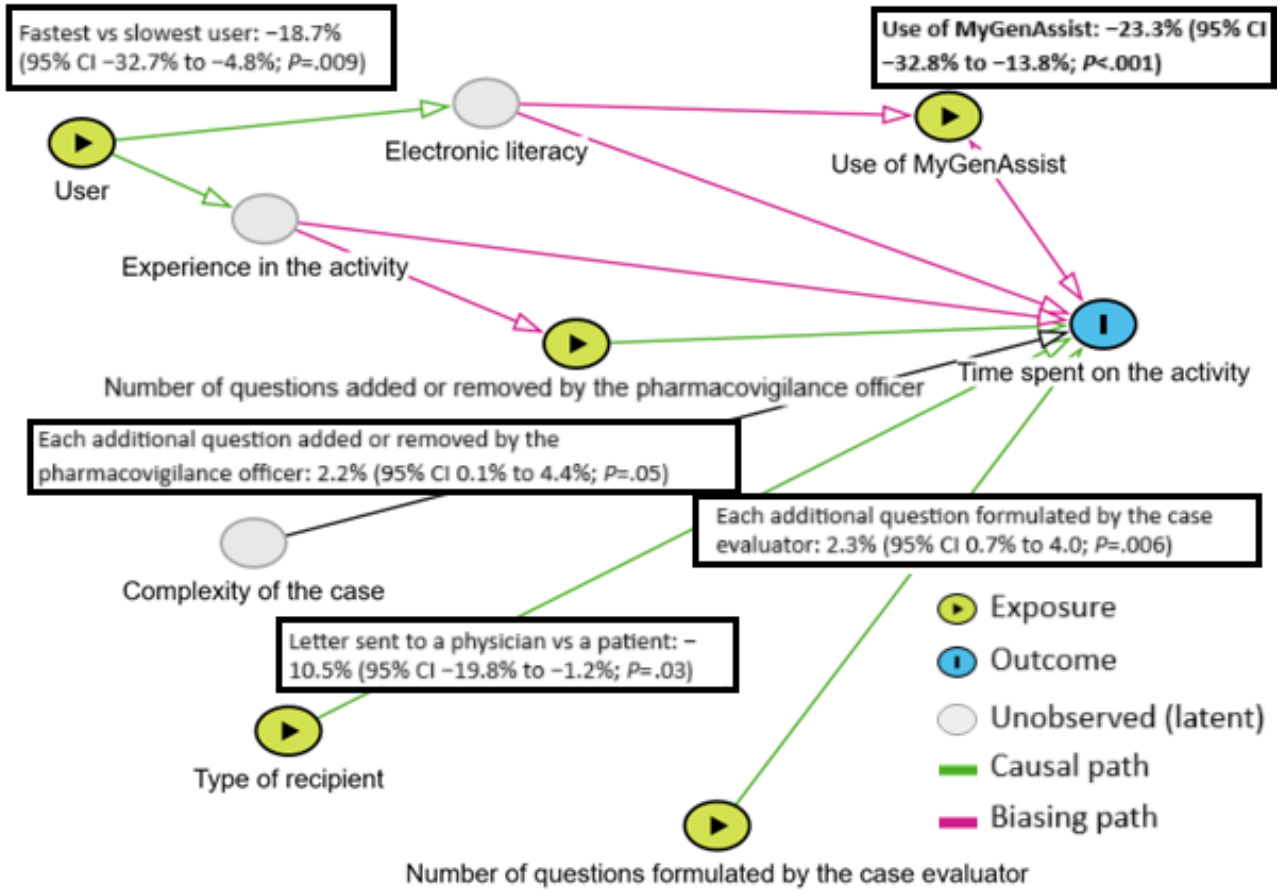

XSL•FO
RenderX

Table 1. Akaike information criterion (AIC) of the multilinear regression model in function of the parameters used.

Parameters included	*df*	Sum of squares	Residual sum of squares	AIC
All	—[a]	—	7.0929	–333.08
Variable "number of questions added or removed by the pharmacovigilance officer" removed	1	0.23076	7.3237	–331.17
Variable "type of recipient" removed	1	0.30945	7.4024	–329.87
Variable "user" removed	2	0.44116	7.5341	–329.72
Variable "number of questions formulated by the case evaluator" removed	1	0.48239	7.5753	–327.05
Variable "use of MyGenAssist" removed	1	1.45757	8.5505	–312.28

[a]Not applicable.

According to the number of spontaneous pharmacovigilance reports for which a documentation request was necessary during the last 3 years, it can be estimated that the mean time saved on the activity with the use of the LLM could be between 66.27 (SD 22.79) and 108.70 (SD 37.23) hours per year, on the basis of an average 23.3% time saved on the average 22.25 (SD 7.62) minutes spent on a case without the use of MyGenAssist. Based on a working day of 8 hours, the use of MyGenAssist for this activity in the 3 previous years could have saved an average of 10.7 (SD 3.6) days of work per year (Table 2).

Table 2. Potential time saved on the pharmacovigilance case documentation process each year with the use of MyGenAssist.

Year	Pharmacovigilance cases requiring documentation request, n	Potential time saved (hours)		Potential time saved (working days)[a]	
		Mean (SD)	95% CI	Mean (SD)	95% CI
2021	1258	108.70 (37.23)	98.16-119.23	13.6 (4.7)	12.3-14.9
2022	947	81.82 (28.02)	73.89-89.75	10.2 (3.5)	9.2-11.2
2023	767	66.27 (22.79)	60.12-72.72	8.3 (2.8)	7.5-9.1

[a]A working day is considered to last 8 hours.

No statistical difference has been found regarding the answers' rates after 1 attempt of contact, whether MyGenAssist was used or not: 31% (15/48) of letters produced without MyGenAssist were provided with an answer and 24% (18/74) when the generative AI was used for the writing (P=.40; χ^2_1=0.7). The same situation occurred after 2 attempts: 42% (20/48) without the tool vs 36% (27/74; P=.57; χ^2_1=0.3). In a subgroup analysis in function of the type of recipient, these results were consistent (Table 3).

Table 3. Rates of answers to letters sent for pharmacovigilance case documentation in function of the use of MyGenAssist and the type of recipient.

	Without the use of MyGenAssist, n/N (%)	With the use of MyGenAssist, n/N (%)	*P* value	Chi-square (*df*=1)
Total	20/48 (42)	27/74 (36)	.57	0.3
Physicians	14/29 (48)	15/45 (33)	.20	1.6
Patients	6/19 (32)	12/29 (41)	.49	0.5

In cases for which an answer is provided, the average time between the answer and the last attempt of contact does not differ significantly (Figure 5): 2.20 (SD 3.27; 95% CI 1.27-3.13) days when the letter was written without MyGenAssist and 2.65 (SD 3.30; 95% CI 1.90-3.41) days with its use (P=.64; t_{44}=–0.46).

XSL•FO
RenderX

Figure 5. Comparison of the duration time between the answer and the last attempt of contact in function of the use of MyGenAssist.

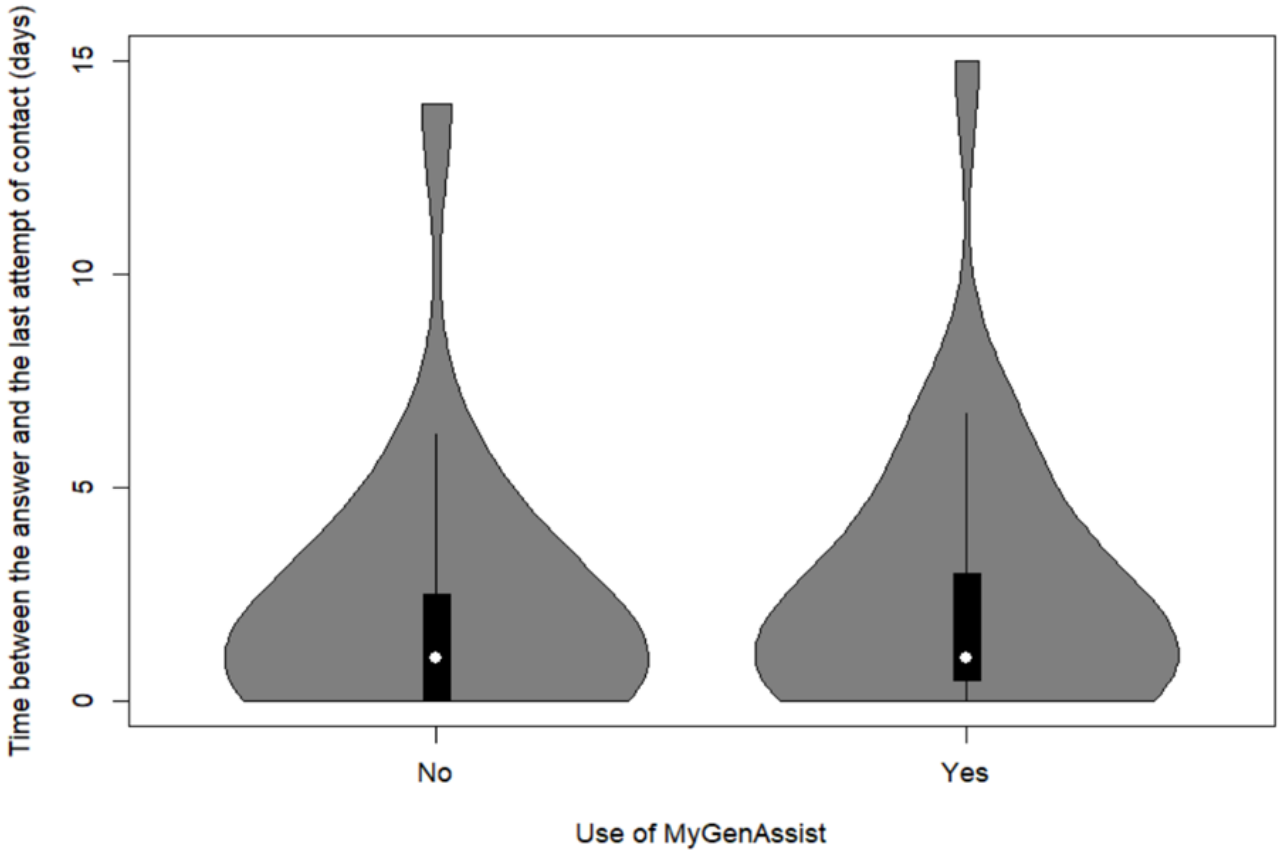

Staff Meetings

Before the study, a 2-hour training was planned to present the use of MyGenAssist in the pharmacovigilance case documentation process to the pharmacovigilance officers. During the study, meetings were planned in the team on a 2-week basis. No difficulties in the use of MyGenAssist were identified, and users only suggested slight improvements in the prompt used to consider specific cases for which additional elements are required in the letter. An operating mode, detailing all the steps required to use MyGenAssist for the activity, was written to offer a reminder for the users if necessary.

Discussion

Main Results

In this study, we demonstrated that the implementation of MyGenAssist to the pharmacovigilance case documentation process provided an average of 23.3% time saving on the task without any modification in the rate of answers. Whatever the experience of the workers, the use of MyGenAssist induces time savings, although this effect is higher in nonexperimented ones. Moreover, while using MyGenAssist, it seems that the time spent on the different pharmacovigilance cases becomes less variable: when the use of the tool is added to the process, the range of time is decreased from 9-38 minutes to 10-25 minutes. MyGenAssist may have a more beneficial impact in cases for which more time is required. The easiest cases needing a slight amount of time to be handled would benefit less (or not) from its use, but the ones that are the more time-consuming would be realized significantly quicker than without it.

Therefore, the use of Bayer's LLM improved the efficiency of this activity. That must save dozens of hours yearly for our local pharmacovigilance team, which could be dedicated to other activities with higher added value. A unique, short, and reachable training was sufficient to implement the use of MyGenAssist. This consisted of introducing the changes in the tool's operating mode to the users. No additional training content was required to enable them to use the LLM in the activity, though they did not have any specific qualifications or skills in informatics. Hence, we showed in this study that making workers use an LLM could be easy, as far as they are assisted at the beginning, while other AI tools require important training to be used daily. However, the daily use of an LLM can make the user continually improve their abilities with this tool. Therefore, LLMs seem easy to integrate into several work environments and a way for the worker to get new skills without needing a large amount of time. Thanks to its ease, all users kept using the generative AI for their activity after the end of the study.

Although we included in the linear regression model all parameters that could have an influence on the time spent on the activity, in our point of view, the adjusted R^2 is low (0.286). Therefore, the most part of the time spent on each case could not be fully explained by the considered parameters. It seems that some aspects that are specific to each and difficult to transcribe in the model are important factors too. Moreover, a "human factor" could play a role here, as the user could spend

XSL•FO
RenderX

a variable duration of time for the same task, in function of a lot of exterior parameters. Whatever explanatory variable was not included here, the average difference created by MyGenAssist (more than 5 minutes on a task requiring about 22 minutes without its use) seems too high to be questioned by these potential "hidden" parameters.

Nonetheless, 3 factors could not be included in the model. First, the experience of the user in the activity may directly influence the time spent on each case. It could also have an effect on the number of questions added or removed by the pharmacovigilance officer: while a newcomer could have difficulties selecting the required information to request from the recipient, an experienced one would more accurately determine which question is important enough to assess an adverse event. Second, electronic literacy, which is the capacity to understand or create messages using electronic means, may have an influence on the time spent on the activity. This impact exists for all cases because they are handled by using informatics, but must be stronger in the cases for which MyGenAssist would be used, as the ability of the user to handle technology could also impact their ease of familiarizing themselves with this new tool. By the time we started this experiment, we were not able to measure either user experience or user electronic literacy [26]. Even if we could not include them in the model, the variable "user" integrates both their experience and their electronic literacy: so, without deciphering which parameter has more impact, we still take them into account. Third, the complexity of the case, which may have an influence on the time spent on the case, is a subjective parameter that could not be measured accurately. The parameters we chose to include in this study were selected during a meeting with experienced pharmacovigilance officers, who were asked to give all factors that could influence the time spent on each case. It seems that no study so far, to the best of our knowledge, has been published reporting the parameters influencing the time spent to document a pharmacovigilance case. Moreover, no comparison of our internal process and the ones of other pharmaceutical companies could be done, as none of them seems to be available publicly. The European GVPs [17] do not define a strict process for the documentation of pharmacovigilance cases and let the market authorization holders organize the activity. Pharmaceutical companies have to create a Pharmacovigilance System Master File that describes the handling of all pharmacovigilance activities. It has to be sent to the national and European health authorities but is not publicly disclosed [27]. Our process could then have specificities (eg, the implication of both a case evaluator and a pharmacovigilance officer) that are not possible to detect and could create a lack of reproducibility of our study.

Other ways could have been explored to determine the effectiveness of generative AI. In this study, we chose to use the fulfillment of the objective of our task as the main criteria of effectiveness: to get the necessary information to analyze the pharmacovigilance case. From our perspective, other criteria did not seem relevant in this context. Moreover, we made the choice not to analyze the quality of the drafts written by MyGenAssist because it was clear in our operating mode and in the training that a human assessment of the letter was mandatory before sending it out. To get a relevant analysis of the effectiveness, recipients were not informed that the letter they received was first written with the help of an LLM: on the one hand, the recipient could have had the impression that its report was "automatically" handled, which could have encouraged this latter not to answer. On the other hand, the potential curiosity created by an eventual mention of generative AI, in a current era in which all eyes are turned to such tools, could have created a bias by making the recipient more willing to answer.

Limitations

While our study was realized over a 4-month period, this latter only includes 122 cases. Even if it was enough to notice a difference in the time spent in function of the use of the tool, this has to be taken into account, particularly while analyzing the answers' rate. However, this study is a first step before the extension of the use of the LLM in the pharmacovigilance case documentation process to all Bayer local pharmacovigilance teams worldwide. The efficiency improvement noticed by the French pharmacovigilance team could give fresh impetus to make other teams adopt the same principle. This will give us the opportunity to retrieve more data concerning our question.

Comparison to Prior Work

Other publications regard the use of LLMs in pharmacovigilance: the use of ChatGPT without any optimization or fine-tuning of the public does not seem to fulfill the objective fixed, whatever this latter is providing answers to pharmacovigilance-related questions [28,29] or assess the risk of mortality of patients with toxic epidermal necrolysis [30]. On the other hand, another LLM, Bidirectional Encoder Representations from Transformers, was used with success for the extraction of adverse events from notes [31,32] or social media [33], but it always implied a step of fine-tuning that could not be handled by a simple user. Therefore, we could not find in the literature any study describing the successful use of ChatGPT or a strict equivalent in pharmacovigilance without the necessity to fine-tune the LLM after its use.

Moreover, the need to explore the possibilities that AI could offer to pharmacovigilance was highlighted by several publications before, particularly with the idea that the implementation of such tools could make both time and cost savings and provide an efficiency improvement, especially as the number of declarations of adverse events has been sharply increasing in the last few years [34]. Gholap et al [35] and Danysz et al [36] evoke the possibility for pharmacovigilance workers to focus on high-level tasks thanks to these time savings on administrative or repetitive tasks. However, in all these papers, these statements remain at the step of hypothesis without any experimental assessment.

Regarding the potential benefits that AI could represent for pharmacovigilance, the TransCelerate initiative, which includes 19 pharmaceutical companies, aimed to ask pharmacovigilance representatives of these companies which steps of the pharmacovigilance case processing might be the best suited for a pharmacovigilance implementation. In 2 surveys set between 2018 [37] and 2021 [38], the pharmacovigilance workers

XSL•FO
RenderX

considered the pharmacovigilance case documentation process as one of the more appropriate steps for an AI implementation. The automation opportunity score, which combines the effort realized by companies to include AI in the task and the benefit or risk assessment regarding this inclusion, was 4.5/5 in 2021. The percentage of companies that set an AI-based process in this task rose from 19% to 35% between 2020 and 2021. Nonetheless, no information was given in these publications about a possible efficiency increase after an AI implementation.

Hence, our study may fulfill what seems to be, to the best of our knowledge, a lack in the medical literature. Descriptions of LLMs' use cases for pharmacovigilance exist, but none of them include a measure of time saving realized thanks to this tool. A study showed that the use of ChatGPT for writing tasks enhanced the productivity of the workers by decreasing the time required by 40%, while the quality rose by 18% [39], but did not concern health. Hence, our objective was to determine if the integration of the LLM in the adverse event case documentation process can improve efficiency by reducing the time required for this GVP-related task without decreasing the response rate from the reporters.

These first encouraging results could be an incentive to implement MyGenAssist in other processes. In the pharmacovigilance field, some tasks fit well with the integration of MyGenAssist. For example, the comparison of different versions of the same procedure with this tool can provide a rapid insight into the modifications in order in a second step to assess their impact on other quality documents. Other use cases could be found in other departments, as the task given to the LLM, which is writing letters based on a template, seems pretty reproducible in other contexts for different objectives. The results of this study show that using LLMs in pharmaceutical activities, whatever the field, is relevant and can create improvements without losing quality concerning the regulations of good practice activities.

Conclusions

In this study, we showed the first example of a use case for a ChatGPT-based tool, MyGenAssist, in a pharmacovigilance industry department and assessed its efficiency over a 4-month period. An average of 23.3% of time savings was achieved thanks to this LLM, while its implementation did not modify the effectiveness of the task. It only required a short training to be set up. These results could be the first step to the largest use of LLMs in pharmaceutical activities.

Acknowledgments

The authors thank all members of the Bayer French local pharmacovigilance team for their kind help and their implication in the study, French Bayer IA champions for the mutual sharing of tips, and Bayer AG for its funding support. No specific funding was necessary or used for this study. Bayer AG provided the authors, who are part of the company, with the usual support provided to their employees. MyGenAssist, ChatGPT, or any other large language models were not used for the writing of our paper.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to our internal data privacy policies, but specific elements are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization and elaboration of the methodology were realized by all authors collectively. Data curation, formal analysis, investigation, project administration, creation of the prompt, and the writing of the original draft were made by AB. The review and editing of the paper were carried out by IB-L, HR, and JPB. IB-L and JPB were in charge of the supervision and validation. Visualization was done by AB, IB-L, and JPB.

Conflicts of Interest

AB, IB-L, and HR were employees of the company (Bayer) at the time this study was performed. Bayer was informed of this study but was not involved in the design of the study or the writing of the manuscript. Bayer gave authorization to the authors to use MyGenAssist for this study.

Multimedia Appendix 1
Original prompt used in French.
[PNG File , 156 KB - jmir_v27i1e65651_app1.png]

Multimedia Appendix 2
Translation of the prompt used in English.
[PNG File , 195 KB - jmir_v27i1e65651_app2.png]

Multimedia Appendix 3

XSL•FO
RenderX

Example of a query by the user and output generated by MyGenAssist in French.
[PNG File , 325 KB - jmir_v27i1e65651_app3.png]

Multimedia Appendix 4
Translation in English of an example of a query by the user and output generated by MyGenAssist.
[PNG File , 417 KB - jmir_v27i1e65651_app4.png]

Multimedia Appendix 5
Description of the R equations used in the paper.
[DOCX File , 15 KB - jmir_v27i1e65651_app5.docx]

References

1. ICUREsearch. DEEPCATH: artificial intelligence for the detection of local signs of infections around catheters. URL: https://icrs.fr/en/ai/deepcath/ [accessed 2024-06-19]
2. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today 2021 Jan;26(1):80-93 [FREE Full text] [doi: 10.1016/j.drudis.2020.10.010] [Medline: 33099022]
3. Mondal H, Panigrahi M, Mishra B, Behera JK, Mondal S. A pilot study on the capability of artificial intelligence in preparation of patients' educational materials for Indian public health issues. J Family Med Prim Care 2023 Aug;12(8):1659-1662 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_262_23] [Medline: 37767452]
4. Cova T, Vitorino C, Ferreira M, Nunes S, Rondon-Villarreal P, Pais A. Artificial intelligence and quantum computing as the next pharma disruptors. Methods Mol Biol 2022;2390:321-347. [doi: 10.1007/978-1-0716-1787-8_14] [Medline: 34731476]
5. Lamberti MJ, Wilkinson M, Donzanti BA, Wohlhieter GE, Parikh S, Wilkins RG, et al. A study on the application and use of artificial intelligence to support drug development. Clin Ther 2019 Aug;41(8):1414-1426. [doi: 10.1016/j.clinthera.2019.05.018] [Medline: 31248680]
6. Singh S, Kumar R, Payra S, Singh SK. Artificial intelligence and machine learning in pharmacological research: bridging the gap between data and drug discovery. Cureus 2023 Aug;15(8):e44359 [FREE Full text] [doi: 10.7759/cureus.44359] [Medline: 37779744]
7. Adamson B, Waskom M, Blarre A, Kelly J, Krismer K, Nemeth S, et al. Approach to machine learning for extraction of real-world data variables from electronic health records. Front Pharmacol 2023 Sep 15;14:1180962 [FREE Full text] [doi: 10.3389/fphar.2023.1180962] [Medline: 37781703]
8. Banerji A, Lai KH, Li Y, Saff RR, Camargo CA, Blumenthal KG, et al. Natural language processing combined with ICD-9-CM codes as a novel method to study the epidemiology of allergic drug reactions. J Allergy Clin Immunol Pract 2020 Mar;8(3):1032-1038.e1 [FREE Full text] [doi: 10.1016/j.jaip.2019.12.007] [Medline: 31857264]
9. Gupta A, Katarya R. PAN-LDA: A latent Dirichlet allocation based novel feature extraction model for COVID-19 data using machine learning. Comput Biol Med 2021 Nov;138:104920 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104920] [Medline: 34655902]
10. Chen LS, Lin ZC, Chang JR. FIR: an effective scheme for extracting useful metadata from social media. J Med Syst 2015 Nov;39(11):139. [doi: 10.1007/s10916-015-0333-0] [Medline: 26330225]
11. Hasan A, Levene M, Weston D. Learning structured medical information from social media. J Biomed Inform 2020 Oct;110:103568 [FREE Full text] [doi: 10.1016/j.jbi.2020.103568] [Medline: 32942027]
12. OpenAI Platform. URL: https://platform.openai.com [accessed 2024-03-13]
13. Abu Hammour K, Alhamad H, Al-Ashwal FY, Halboup A, Abu Farha R, Abu Hammour K. ChatGPT in pharmacy practice: a cross-sectional exploration of Jordanian pharmacists' perception, practice, and concerns. J Pharm Policy Pract 2023;16(1):129 [FREE Full text] [doi: 10.1186/s40545-023-00640-2] [Medline: 37789443]
14. Abu-Farha R, Fino L, Al-Ashwal FY, Zawiah M, Gharaibeh L, Harahsheh MM, et al. Evaluation of community pharmacists' perceptions and willingness to integrate ChatGPT into their pharmacy practice: a study from Jordan. J Am Pharm Assoc 2023 Nov;63(6):1761-1767.e2. [doi: 10.1016/j.japh.2023.08.020] [Medline: 37648157]
15. Temsah MH, Aljamaan F, Malki KH, Alhasan K, Altamimi I, Aljarbou R, et al. ChatGPT and the future of digital health: a study on healthcare workers' perceptions and expectations. Healthcare (Basel) 2023 Jun 21;11(13):1812 [FREE Full text] [doi: 10.3390/healthcare11131812] [Medline: 37444647]
16. Ray S. Samsung bans ChatGPT among employees after sensitive code leak. Forbes. 2023 May 02. URL: https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/ [accessed 2024-04-29]
17. Guideline on good pharmacovigilance practices (GVP): Module VI—collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). European Medicines Agency. URL: https://tinyurl.com/48zjc968 [accessed 2024-03-13]

XSL•FO
RenderX

18. Bonnes pratiques de pharmacovigilance. Agence Nationale de Sécurité du Médicament et des produits de santé. URL: https://ansm.sante.fr/documents/reference/bonnes-pratiques-de-pharmacovigilance [accessed 2024-03-15]
19. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use, vol 311. European Union. 2001. URL: http://data.europa.eu/eli/dir/2001/83/oj/eng [accessed 2024-06-19]
20. Pariente A, Micallef J, Lahouegue A, Molimard M, Auffret M, Chouchana L, et al. What place for intelligent automation and artificial intelligence to preserve and strengthen vigilance expertise in the face of increasing declarations. Therapies 2023 Jan;78(1):131-143. [doi: 10.1016/j.therap.2022.11.004] [Medline: 36572627]
21. Best practices for prompt engineering with the OpenAI API. OpenAI. URL: https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api [accessed 2024-04-01]
22. Argus database. Oracle. URL: https://www.oracle.com/life-sciences/pharmacovigilance/argus-safety-case-management/ [accessed 2024-03-13]
23. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. performance: an R package for assessment, comparison and testing of statistical models. JOSS 2021 Apr;6(60):3139. [doi: 10.21105/joss.03139]
24. Venables WN, Ripley BD. Modern applied statistics with S. In: Statistics and Computing. 4th Edition. New York, NY: Springer; 2002.
25. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (text with EEA relevance), vol 119. European Union. 2016. URL: http://data.europa.eu/eli/reg/2016/679/oj/eng [accessed 2024-12-25]
26. Marsall M, Engelmann G, Skoda EM, Teufel M, Bäuerle A. Measuring electronic health literacy: development, validation, and test of measurement invariance of a revised German version of the eHealth literacy scale. J Med Internet Res 2022 Feb 02;24(2):e28252 [FREE Full text] [doi: 10.2196/28252] [Medline: 35107437]
27. Lavery C, Emmott J, Jeck-Thole S, Rouben P, Usher D, van der Spuij W, et al. An industry survey on managing the Pharmacovigilance System Master File in a global environment: the need for a pragmatic approach. Pharm Med 2022 Jun 20;36(4):233-245 [FREE Full text] [doi: 10.1007/s40290-022-00422-2] [Medline: 35723862]
28. Morath B, Chiriac U, Jaszkowski E, Deiß C, Nürnberg H, Hörth K, et al. Performance and risks of ChatGPT used in drug information: an exploratory real-world analysis. Eur J Hosp Pharm 2023 Jun 01;31(6):491-497. [doi: 10.1136/ejhpharm-2023-003750] [Medline: 37263772]
29. Montastruc F, Storck W, de Canecaude C, Victor L, Li J, Cesbron C, et al. Will artificial intelligence chatbots replace clinical pharmacologists? An exploratory study in clinical practice. Eur J Clin Pharmacol 2023 Oct;79(10):1375-1384. [doi: 10.1007/s00228-023-03547-8] [Medline: 37566133]
30. Pandya S, Patel C, Sojitra B, Karamata H. Causality assessment of adverse drug reaction toxic epidermal necrolysis with the aid of ChatGPT: a case report. Cureus 2024 May;16(5):e60638 [FREE Full text] [doi: 10.7759/cureus.60638] [Medline: 38903274]
31. Silverman AL, Sushil M, Bhasuran B, Ludwig D, Buchanan J, Racz R, et al. Algorithmic identification of treatment‐emergent adverse events from clinical notes using large language models: a pilot study in inflammatory bowel disease. Clin Pharmacol Ther 2024 Mar 08;115(6):1391-1399. [doi: 10.1002/cpt.3226] [Medline: 38459719]
32. Wu L, Gray M, Dang O, Xu J, Fang H, Tong W. RxBERT: Enhancing drug labeling text mining and analysis with AI language modeling. Exp Biol Med (Maywood) 2023 Nov;248(21):1937-1943 [FREE Full text] [doi: 10.1177/15353702231220669] [Medline: 38166420]
33. Dong F, Guo W, Liu J, Patterson TA, Hong H. BERT-based language model for accurate drug adverse event extraction from social media: implementation, evaluation, and contributions to pharmacovigilance practices. Front Public Health 2024;12:1392180 [FREE Full text] [doi: 10.3389/fpubh.2024.1392180] [Medline: 38716250]
34. Desai MK. Artificial intelligence in pharmacovigilance—opportunities and challenges. Perspect Clin Res 2024;15(3):116-121. [doi: 10.4103/picr.picr_290_23] [Medline: 39140015]
35. Gholap AD, Uddin MJ, Faiyazuddin M, Omri A, Gowri S, Khalid M. Advances in artificial intelligence for drug delivery and development: a comprehensive review. Comput Biol Med 2024 Aug;178:108702. [doi: 10.1016/j.compbiomed.2024.108702] [Medline: 38878397]
36. Danysz K, Cicirello S, Mingle E, Assuncao B, Tetarenko N, Mockute R, et al. Artificial intelligence and the future of the drug safety professional. Drug Saf 2018 Oct 20;42(4):491-497 [FREE Full text] [doi: 10.1007/s40264-018-0746-z] [Medline: 30343417]
37. Ghosh R, Kempf D, Pufko A, Barrios Martinez LF, Davis CM, Sethi S. Automation opportunities in pharmacovigilance: an industry survey. Pharm Med 2020 Feb 08;34(1):7-18. [doi: 10.1007/s40290-019-00320-0] [Medline: 32036574]
38. Kassekert R, Grabowski N, Lorenz D, Schaffer C, Kempf D, Roy P, et al. Industry perspective on artificial intelligence/machine learning in pharmacovigilance. Drug Saf 2022 May 17;45(5):439-448 [FREE Full text] [doi: 10.1007/s40264-022-01164-5] [Medline: 35579809]
39. Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. Science 2023 Jul 14;381(6654):187-192. [doi: 10.1126/science.adh2586] [Medline: 37440646]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
GVP: good pharmacovigilance practices
LLM: large language model

Edited by T de Azevedo Cardoso; submitted 21.08.24; peer-reviewed by S Matsuda, D Chrimes; comments to author 26.09.24; revised version received 17.11.24; accepted 13.01.25; published 10.03.25.

Please cite as:
Benaïche A, Billaut-Laden I, Randriamihaja H, Bertocchio JP
Assessment of the Efficiency of a ChatGPT-Based Tool, MyGenAssist, in an Industry Pharmacovigilance Department for Case Documentation: Cross-Over Study
J Med Internet Res 2025;27:e65651
URL: https://www.jmir.org/2025/1/e65651
doi:10.2196/65651
PMID:

©Alexandre Benaïche, Ingrid Billaut-Laden, Herivelo Randriamihaja, Jean-Philippe Bertocchio. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patient Perspectives on Conversational Artificial Intelligence for Atrial Fibrillation Self-Management: Qualitative Analysis

Ritu Trivedi[1], PhD; Tim Shaw[1,2], PhD; Brodie Sheahen[1], MD; Clara K Chow[1,3], PhD; Liliana Laranjo[1], PhD

[1]Westmead Applied Research Centre, Faculty of Medicine and Health, The University of Sydney, Westmead, Australia

[2]Charles Perkins Centre, School of Medical Sciences, The University of Sydney, Sydney, Australia

[3]Cardiology Department, Westmead Hospital, Westmead, Australia

Corresponding Author:
Liliana Laranjo, PhD
Westmead Applied Research Centre
Faculty of Medicine and Health
The University of Sydney
Level 5, Block K, Westmead Hospital
Hawkesbury Road
Westmead, 2145
Australia
Phone: 61 2 8890 3125
Email: liliana.laranjo@sydney.edu.au

Abstract

Background: Conversational artificial intelligence (AI) allows for engaging interactions, however, its acceptability, barriers, and enablers to support patients with atrial fibrillation (AF) are unknown.

Objective: This work stems from the Coordinating Health care with AI–supported Technology for patients with AF (CHAT-AF) trial and aims to explore patient perspectives on receiving support from a conversational AI support program.

Methods: Patients with AF recruited for a randomized controlled trial who received the intervention were approached for semistructured interviews using purposive sampling. The 6-month intervention consisted of fully automated conversational AI phone calls (with speech recognition and natural language processing) that assessed patient health and provided self-management support and education. Interviews were recorded, transcribed, and thematically analyzed.

Results: We conducted 30 interviews (mean age 65.4, SD 11.9 years; 21/30, 70% male). Four themes were identified: (1) interaction with a voice-based conversational AI program (human-like interactions, restriction to prespecified responses, trustworthiness of hospital-delivered conversational AI); (2) engagement is influenced by the personalization of content, delivery mode, and frequency (tailoring to own health context, interest in novel information regarding health, overwhelmed with large volumes of information, flexibility provided by multichannel delivery); (3) improving access to AF care and information (continuity in support, enhancing access to health-related information); (4) empowering patients to better self-manage their AF (encouraging healthy habits through frequent reminders, reassurance from rhythm-monitoring devices).

Conclusions: Although conversational AI was described as an engaging way to receive education and self-management support, improvements such as enhanced dialogue flexibility to allow for more naturally flowing conversations and tailoring to patient health context were also mentioned.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12621000174886; https://tinyurl.com/3nn7tk72

International Registered Report Identifier (IRRID): RR2-10.2196/34470

(J Med Internet Res 2025;27:e64325) doi:10.2196/64325

KEYWORDS

atrial fibrillation; conversational agents; qualitative research; self-management; digital health; patient perspective; conversational artificial intelligence; speech recognition

XSL•FO
RenderX

Introduction

Atrial fibrillation (AF) prevalence is increasing rapidly [1], resulting in substantial costs to health systems [2]. AF care is multifaceted, and appropriate management is complex but key to preventing stroke, heart failure, and all-cause mortality [3]. Novel solutions to address these challenges and deliver integrated AF care are warranted [3]. Digital tools could be used to deliver self-management support and education to promote healthy lifestyle behaviors, medication adherence, symptom management, and overall ability to manage AF and prevent hospitalization [4]. However, existing digital interventions for patients with AF have shown limited engagement [5-7].

Conversational AI is a novel technology that can converse with users interactively by mimicking human-like conversations [8], but its feasibility in supporting AF patients is unknown. The evaluation of conversational AI has been limited, with the majority of studies being quasi-experimental and small-scale feasibility pilots exploring conversational AI for the prevention, diagnosis, and management of chronic conditions, but mostly for psychological conditions (depression and anxiety) and very few in cardiovascular diseases (heart failure, diabetes, and hypertension) [9-11]. Only a few randomized controlled trials (RCTs) have been conducted evaluating the effectiveness of conversational AI, revealing positive outcomes of improvements in symptoms of anxiety and depression in college students [12,13], insulin adherence and glycemic control in patients with type 2 diabetes [10], and medication adherence in patients with both or either hypertension and diabetes [11]. Notably, high engagement with users was common in these trials [10-13] and in other studies [9]. The application of conversational AI may provide a more engaging experience for patients with AF self-managing their condition.

Understanding user perspectives on digital health technologies is essential, especially in AF, as the condition manifests differently for all patients [14]. There is a paucity of qualitative research assessing the perspectives and experiences of patients with AF in using digital self-management interventions, with existing studies focusing on mobile apps [15-18] and none exploring conversational AI. Given the novelty of the technology, qualitative research is required to understand the user experience and acceptability of conversational AI to support AF self-management. Although an existing study has explored the feasibility of a conversational agent (non–AI-based) in patients with AF and has found high acceptance among users, it did not involve a formal qualitative analysis [15]. Qualitative research can help in understanding the barriers and enablers that influence AF patient engagement, which is important in narrowing the gap between proof-of-concept and implementation [19].

The Coordinating Health care with AI–supported Technology for patients with AF (CHAT-AF) [20] was a 6-month RCT evaluating a postdischarge digital program that comprised voice-based conversational AI (automated phone calls with speech recognition and natural language processing), text messages, emails, and an educational website to support patients with AF in their management journey. In this paper, we report on the qualitative findings from semistructured interviews postintervention assessing user perspectives, enablers, and barriers regarding the intervention.

Methods

Study Design

This qualitative study stems from a 6-month RCT, CHAT-AF, aimed at evaluating the feasibility of a conversational AI support program for patients with AF in the community [20]. This study adheres to the Consolidated criteria for Reporting Qualitative (COREQ) research checklist for reporting qualitative research.

Ethical Considerations

The ethics committee at Western Sydney Local Health District (2020/ETH02546) granted approval and the study was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12621000174886). All participants provided informed consent to take part in both the RCT and qualitative interviews. Data collected from this study was only accessible by approved study personnel and was stored on secure servers. Deidentification of data was done before analysis and publication.

Setting and Participants

Participants were recruited for the RCT after discharge from inpatient or outpatient cardiology services at Westmead Hospital (Sydney, Australia). Adult AF patients competent with the English language who had access to a mobile phone were eligible. AF patients randomized to the intervention group were contacted via phone postintervention by study investigator RT and asked whether they would like to participate in a phone interview. These participants were chosen using maximum variation sampling, such that, people of different sex, age, ethnicity, education, type of AF, and engagement with the intervention were invited to take part in the interview.

Intervention

The intervention was designed to monitor patient health status and symptoms and provide self-management and education to AF patients in the community [20] (Figure 1). Briefly, it consisted of 7 outreaches via fully automated conversational AI phone calls, text messages, or email, supplemented with weekly nudges (text message or email) to visit the semipersonalized educational website (personalized according to smoking status, alcohol consumption, hypertension diagnosis, prescription of anticoagulants or warfarin specifically). During the phone calls patients received AF education and were required to verbally respond to risk assessment queries, with certain patient responses triggering an alert to the study team to be actioned and escalated if needed. If the participant was unable to be reached after 3 failed call attempts (including voicemails directed to an inbound phone number), a text message or email was sent directing them to a survey replicating the call content.

XSL•FO
RenderX

Figure 1. Coordinating Health care with Artificial intelligence–supported Technology for patients with Atrial Fibrillation (CHAT-AF) intervention. AF: atrial fibrillation; AI: artificial intelligence.

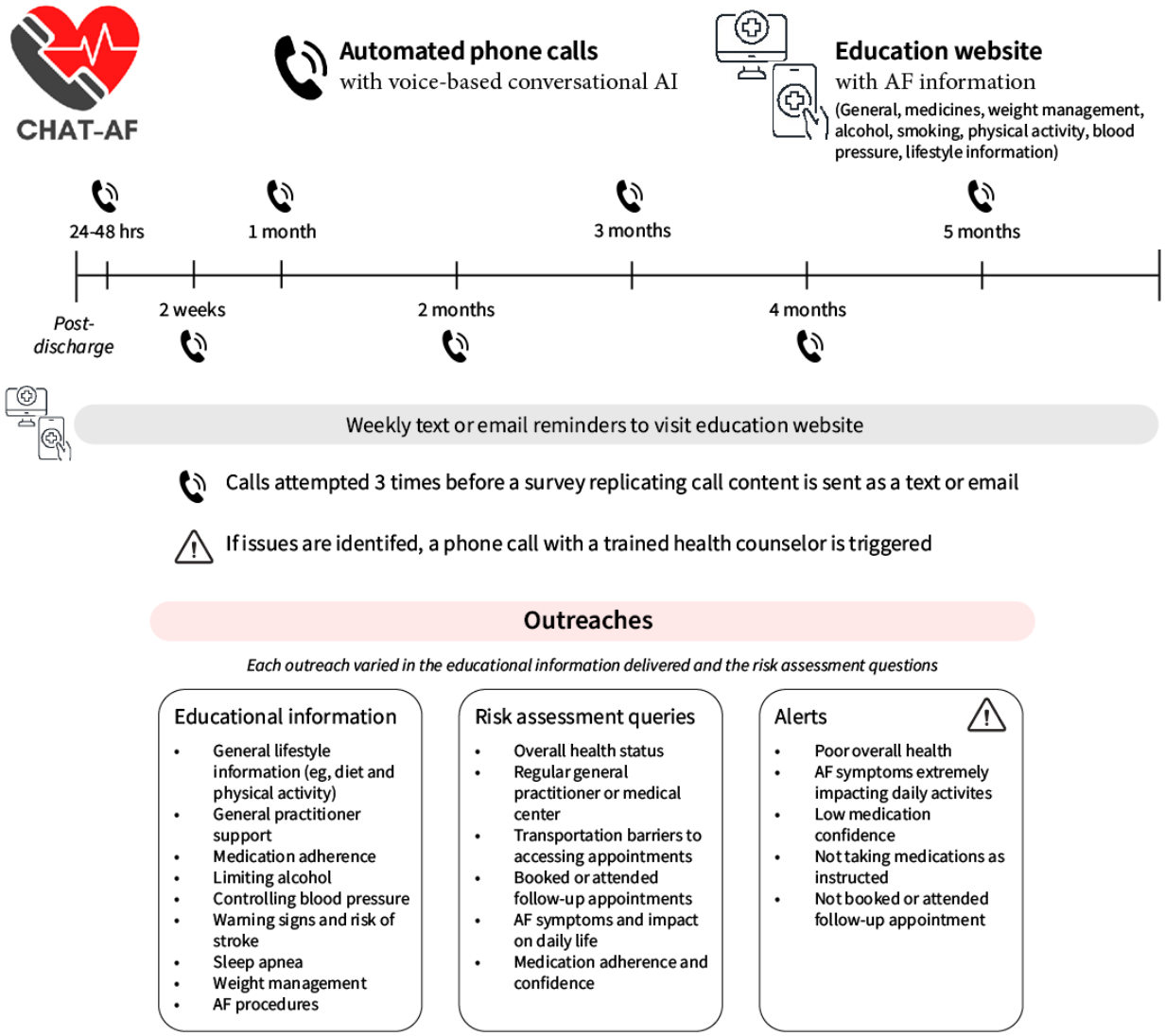

The AI aspect of the intervention included two main components: (1) speech recognition, which allowed the participant's verbal responses to be recognized and then translated into text and (2) natural language processing, which analyzed the semantic and syntactic components of the user's input. The tool was culturally adapted to include a voice with an Australian accent. If the participant's response was unclear, the system would first repeat the question and then prompt the participant to press a button that corresponds to their answer (eg, "press 1 for always, 2 for often, and 3 for sometimes"). The system addressed participants by their first name and was programmed to refer to previous content areas, for example, if a participant had reported they had already visited their GP post their hospital discharge, it would not ask again.

Modified Delivery of the Intervention

We had to prematurely stop recruitment in August 2021 due to the acquisition of the technology partner by another organization. At this point, there were 82 participants enrolled in the intervention arm, of these 20 had completed the intervention and 62 were at various stages of the program. All participants had received at least 2 of 7 outreaches via the conversational AI calls. The delivery of the intervention was then modified, and the call content was replicated and provided via survey only for the remaining outreaches.

Data Collection Techniques

A semistructured interview approach was chosen to explore individual perspectives on conversational AI calls and other intervention components, along with exploring barriers and enablers to engagement with the intervention. The interview guide included open-ended questions and was developed based on existing literature reporting on the self-management needs of patients with AF and their perspectives on digital health interventions to support self-management [14-17] (Multimedia Appendix 1). The interview guide was pilot-tested within the research team. All interviews were conducted by the study investigator (RT) via phone and transcribed verbatim. Comprehensive analysis of interview content resulted in various iterations of themes and concepts. Interviews were conducted until sufficient depth was reached and no new themes, concepts, or relationships were emerging [21].

XSL•FO
RenderX

Analysis

Interview transcripts were uploaded onto NVivo (QSR International) and thematically analyzed by 2 investigators (RT and LL). Coding was done inductively (first-cycle coding), followed by a grouping of codes based on thematic similarities and axial coding (ie, to find patterns between codes) [22]. The initial codebook was developed by RT, with discussions between investigators (RT and LL) occurring every 5 interviews to revise the codebook by analyzing emerging codes and themes. Refined versions of the codebook were reviewed by another investigator (BS) to ensure the appropriateness of focused and axial coding. Coauthor input (LL, BS, CKC, and TS) regarding themes and subthemes was gathered midway and at the completion of interviews.

Results

We invited 33 intervention participants to this study and 3 were not interested in taking part in an interview. A total of 30 semistructured interviews (average duration of 20 minutes and 19 seconds) were conducted between August 2021 to May 2022. The mean participant age was 65.4 (11.91 SD) years, 70% (21/30) were male, 17% (5/30) were of non-Caucasian ethnicity, and 70% (21/30) were nonuniversity graduates (Table 1). Detailed characteristics of both the interviewed participants and the entire intervention cohort are available in Multimedia Appendix 2. Interviewed participants (n=30) received the full 6-month intervention, which included a total of 210 intervention outreaches: 154 were delivered via conversational AI and the remaining 54 were delivered only via the survey tool (modified intervention delivery mode). On average, 5.2 of 7 outreaches were delivered via conversational AI calls (original intervention delivery mode). Themes and subthemes are described below and summarized in Figure 2. Illustrative quotes for each theme are available in Multimedia Appendix 3.

XSL•FO
RenderX

Table 1. Participant characteristics and engagement metrics of the interviewed cohort (N=30 unless otherwise stated).

Characteristics	Values, n (%)
Sex	
Male	21 (70)
Female	9 (30)
Age	
Mean (SD), years	65.4 (12)
Younger than 64 years	13 (43)
65-75 years	8 (27)
Older than 75 years	9 (30)
Ethnicity	
Caucasian	25 (83)
Other Asian	2 (7)
Chinese	1 (3)
Arab or Persian	1 (3)
Other	1 (3)
Most recent atrial fibrillation type	
Paroxysmal	25 (83)
Persistent	5 (17)
Initial atrial fibrillation diagnosis (n=29)	
Less than 5 years ago	21 (70)
5 years or longer	8 (27)
Education	
Primary school	2 (7)
Year 10 school certificate	5 (17)
Year 12 higher school certificate	6 (20)
Diploma/technical	8 (27)
University graduate	9 (30)
Annual household income, AUD[a] (n=20)	
Less than $31,199	4 (13)
$31,200-$77,999	9 (30)
$78,000-$104,000	3 (10)
More than $104,000	4 (13)
Engagement with outreaches	
High (≥4 completed[b])	25 (83)
Low	5 (17)
Engagement with website	
Visited once or less than once	19 (63)
Outreaches delivered via conversational artificial intelligence phone calls, of 7 total, mean (SD)	5.20 (1.81)

[a]The average conversion rate during the study was 1 AUD=US $0.72646.

[b]Participants answered ≥50% of questions asked in the outreach.

XSL•FO
RenderX

Figure 2. Themes and subthemes. AI: artificial intelligence.

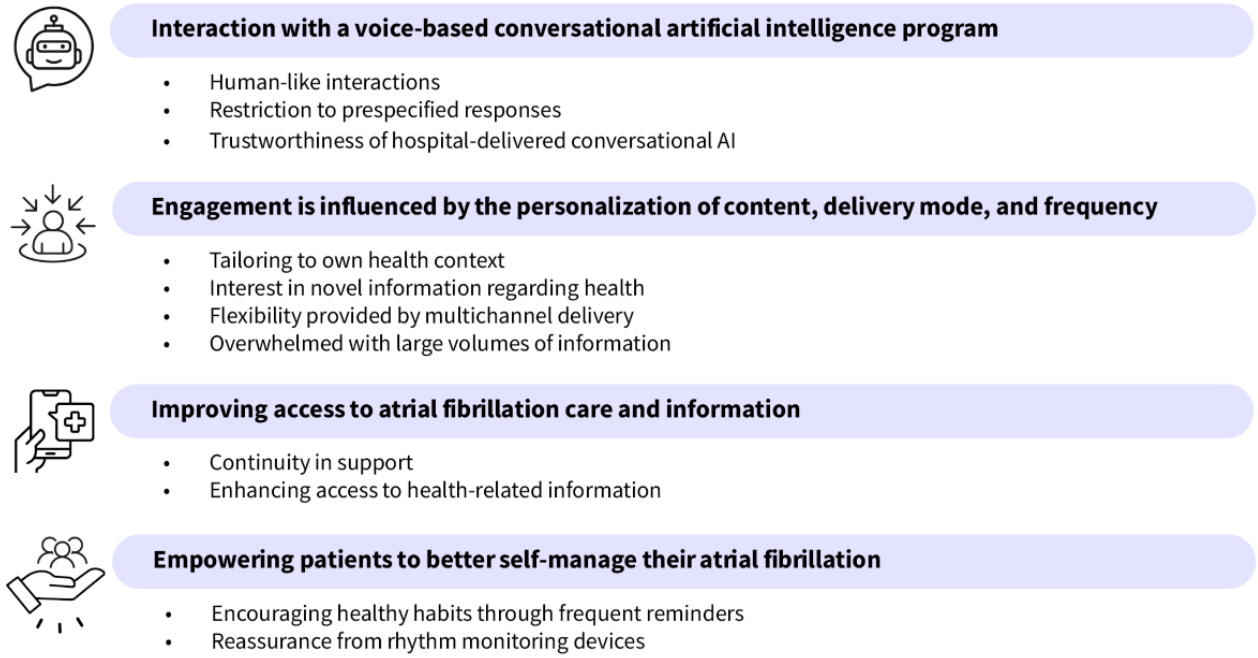

Interaction With a Voice-Based Conversational AI Program

Human-Like Interactions

Participants reported that engaging with the phone calls felt "more human" and "personal" like "talking to a friend" (Multimedia Appendix 3, Table S1, quotes 1-5). Participants mentioned that the phone calls were interactive in the sense they provided a "two-way" channel of communication where they were being asked questions or given information and they were able to respond "easily" (Multimedia Appendix 3, Table S1, quote 5).

Restriction to Prespecified Responses

Participants mentioned the need for more dialogue flexibility, as they felt restricted by the predefined response options in the automated phone calls and wished they had an opportunity to elaborate on their answers (Multimedia Appendix 3, Table S1, quotes 6-8). They also mentioned that some questions did not apply to themselves and due to the lack of appropriate choices in response options, they were at times forced to choose an inappropriate answer, whereas a "does not apply" option would have prevented this issue (Multimedia Appendix 3, Table S1, quote 8).

Trustworthiness of Hospital-Delivered Conversational AI

The importance of the trustworthiness of conversational AI was evident as participants mentioned that bad experiences with marketing-based promotional calls could be a potential barrier to engagement (Multimedia Appendix 3, Table S1, quote 9). However, they highlighted that trusting the provider (ie, the hospital) created a sense of comfort when engaging with automated phone calls (Multimedia Appendix 3, Table S1, quote 10).

Engagement Is Influenced by the Personalization of Content, Delivery Mode, and Frequency

Tailoring to Own Health Context

Participants mentioned that they would have found the program more useful and relevant if they were currently experiencing AF symptoms (Multimedia Appendix 3, Table S2, quote 1) and that they wanted targeted information that was relevant to their disease status (eg, time since diagnosis) and type of AF (eg, paroxysmal AF; Multimedia Appendix 3, Table S2, quotes 2 and 3). Participants noted that because their AF symptoms were not frequent, some questions did not apply to them (Multimedia Appendix 3, Table S2, quote 3).

Participants mentioned the program should be tailored based on their existing knowledge about the disease and their ability to self-manage (Multimedia Appendix 3, Table S2, quotes 4-9**).** Some thought the current content was probably more useful for recently diagnosed patients in greater need of information and support (Multimedia Appendix 3, Table S2, quotes 4 and 5) rather than people who had AF for a while (Multimedia Appendix 3, Table S2, quotes 6 and 7). The need for additional information about pacemakers (Multimedia Appendix 3, Table S2, quote 8) and the impact of COVID-19 on AF (Multimedia Appendix 3, Table S2, quote 9) was mentioned.

The need for tailored information based on an individual's lifestyle behaviors was a recurring theme as participants felt that some of the lifestyle tips did not apply to them, for example, information about quitting smoking and limiting alcohol and caffeine (Multimedia Appendix 3, Table S2, quotes 10 and 11). It was also suggested that we include more comprehensive information about diet (Multimedia Appendix 3, Table S2, quote 12).

XSL•FO
RenderX

Interest in Novel Information Regarding Health

Participants mentioned they would have preferred more novelty in the content, with some of the information being similar to what is available from other reputable sources, such as the Heart Foundation (Multimedia Appendix 3, Table S2, quote 13 and 14). They also indicated the need for progressive content delivery to maintain engagement, for example, the weekly reminders to visit the education website were engaging but led to disappointment when participants realized there was no new content available (Multimedia Appendix 3, Table S2, quote 15).

Flexibility Provided by Multichannel Delivery

Participants differed in their preferences for communication. Some participants noted that it was easier to engage in phone calls and respond using their voice rather than scrolling on a survey, as they could complete the phone calls while multitasking, for example, going for a walk around the park (Multimedia Appendix 3, Table S2, quote 16). Some people also mentioned preferring to listen to information over a phone call, as they thought it was a more interactive and engaging process and they found it easier to interpret the information and visualize what had been said (Multimedia Appendix 3, Table S2, quotes 17-19). Others preferred to respond to the surveys instead of the phone calls, as surveys allowed for adequate time to think about questions before responding, could be completed at a convenient time, and were not interrupted by signal drops (Multimedia Appendix 3, Table S2, quotes 20-23).

Overwhelmed With Large Volumes of Information

Participants mentioned that the frequency of contact should also be tailored based on their needs and preferences, as some felt there were too many phone calls and reminders and would prefer to contact us on an "as needed" basis (Multimedia Appendix 3, Table S2, quotes 24 and 25). When asked about reasons for low intervention engagement, participants mentioned that they were dealing with other medical conditions that took precedence over their AF (Multimedia Appendix 3, Table S2, quote 26), and some participants did not want to be reminded of the problem as they were trying to stay positive (Multimedia Appendix 3, Table S2, quotes 26 and 27).

Improving Access to Atrial Fibrillation Care and Information

Continuity in Support

Overall, participants saw the program as valuable in connecting them with care and providing important information and support (Multimedia Appendix 3, Table S3, quotes 1-8). Participants mentioned that emergency department visits were a "whirlwind experience" and "daunting," and that there was typically no real support and education provided postdischarge, making them feel "helpless" (Multimedia Appendix 3, Table S3, quotes 2 and 3). Participants said that the program was able to address this gap by providing them with information and continuity in follow-up (Multimedia Appendix 3, Table S3, quotes 2-8)—"someone cares enough to contact you"—which was particularly important during COVID-19 when it was difficult to access doctors and for people needing social support (Multimedia Appendix 3, Table S3, quotes 5 and 6). Participants also highlighted the unpredictable and sporadic nature of AF and the need for continuous care as situations change quickly (Multimedia Appendix 3, Table S3, quote 7).

Enhancing Access to Health-Related Information

Participants mentioned that the program provided education and information about AF management that was difficult to obtain from clinicians due to multiple barriers, such as financial, geographical, and mobility constraints along with lack of time during consultations (Multimedia Appendix 3, Table S3, quotes 8-10). They also highlighted that information presented in clinical settings was often hard to absorb due to the high-stress environment and the use of complex terminology, and they appreciated receiving education in lay terms and in a simple format (Multimedia Appendix 3, Table S3, quotes 11 and 12). They also mentioned that the intervention was able to address existing knowledge gaps about AF and improve awareness (Multimedia Appendix 3, Table S3, quotes 13 and 14).

Empowering Patients to Better Self-Manage Their AF

Encouraging Healthy Habits Through Health Reminders

Participants reported that the program improved their health behaviors as it served as a constant reminder and reinforcement of the need to follow healthy lifestyle habits and re-evaluate harmful ones (Multimedia Appendix 3, Table S4, quotes 1 and 2). Some participants mentioned that the program was an influencer in reducing their alcohol intake and promoting weight loss (Multimedia Appendix 3, Table S4, quotes 3 and 4). It was mentioned that the reminders were encouraging and motivational in prompting people to reflect on their current AF management, for example, remember to take medications, exercise, and focus on diet (Multimedia Appendix 3, Table S4, quotes 4-6).

Reassurance From Rhythm Monitoring Devices

The positives and opportunities for improvement of the conversational AI self-management intervention have been outlined in Textbox 1. Participants mentioned that having the ability to monitor their heart rhythm using consumer wireless devices or through an implanted device (loop recorder or pacemaker) can be reassuring in case they suspect the onset of AF and to better manage their symptoms and anxiety related to the disease (Multimedia Appendix 3, Table S4, quotes 7-10). Patients with implantable devices also mentioned that they would prefer being contacted more frequently to ensure that the device was working properly or having access to their own data to understand when things were normal and abnormal, as well as their individual triggers for AF (Multimedia Appendix 3, Table S4, quotes 11 and 12).

XSL•FO
RenderX

Textbox 1. Positives and opportunities for improvement regarding the use of a conversational AI self-management intervention for patients with atrial fibrillation.

Positive elements

- Trustworthiness from receiving information from a hospital-linked provider.
- Engaging human-like interactions.
- Improved continuity in support.
- Improved access to health-related information.
- Frequent reminders and information.

Areas for improvement

- Improving dialogue flexibility of conversational AI phone calls to allow for more natural conversations.
- Enhancing personalization to one's health context.
- Tailoring the amount of information to individual preferences.
- Reducing the repetitiveness of content.

Discussion

Principal Results

Our evaluation of a 6-month AF support program comprising voice-based conversational AI showed that participants valued being supported with their care needs outside of the clinical setting while indicating different barriers and enablers to engagement. Participants' experiences with the conversational AI phone calls were generally positive, with some participants highly valuing that engagement method, and others being less keen on verbal communication for this purpose and preferring a text-based format. Overall, participants indicated the need for more dialogue flexibility, particularly in terms of their response options, to enable more engaging conversations. Personalization of the intervention to patients' needs and preferences was also mentioned as key to engagement, with AF being an especially challenging disease given the high variability in patterns, symptoms, and clinical manifestations.

Comparison With Previous Work

The current intervention was able to address the education and self-management needs of patients with AF, while also providing continuous care outside of a clinical setting. CHAT-AF participants appreciated receiving education and frequent self-care reminders, indicating the intervention addressed a gap in the provision of information and behavior change support. This gap had been previously described in other studies where AF patients expressed their desire for more information, assistance in creating healthy habits, and frequent motivational reminders [23,24]. In addition, it was appreciated that CHAT-AF enabled a greater ability to engage with content easily, when, and how the participant desired outside of a clinical environment. Knowledge gaps are prevalent in patients with AF [25] and patients have indicated insufficient opportunities for education during clinical visits [24]. Managing the ups and downs of the disease is a contributor to the poor quality of life experienced by patients [14] and could be addressed by providing additional support across the patient journey. Participants in our study appreciated the continuity in contact provided by the frequent digital "check-ups," suggesting that interventions like CHAT-AF could enhance the transition between clinical care and community by providing a layer of interaction and support beyond traditional appointments.

Engagement with digital self-management programs by patients with AF has been suboptimal [5-7] but seems to be improved with conversational agents that can facilitate a human-like interaction[15]. A 1-month trial evaluating a mobile application with an embedded non-AI conversational agent showed high acceptability and engagement by patients (18 interactions over one month), even though the interactions required participants to touch prespecified options on the app, instead of using natural language [15]. The interactivity and convenience of voice-based conversational AI were mentioned by participants in CHAT-AF, and engagement was high over the 6-month intervention period, suggesting the value of such technology in supporting patients between clinical visits.

AF has a varied course, and patients appreciate the ability to freely express details regarding their AF rather than being confined to prespecified options, however, at present AI technologies do not allow for safe unconstrained interactions to deliver health information. Users of other AI-based conversational agents have also noted limitations in having naturally flowing conversations [12,13]. However, research has identified various safety concerns in unconstrained speech recognition of voice assistants in health care [26,27]. A study posing medical problems to popular conversational agents (eg, Alexa [Amazon], Google Assistant, and Siri [Apple Inc]), found that only 43% completed their given task, and of these completed tasks, 29% reported potentially harmful actions and 16% may have resulted in death [27]. Future developments in AI, specifically speech recognition and natural language processing, may be able to address the current dialogue shortcomings in our intervention, however, careful consideration is required for the safety implications of their use in health care.

Personalization was highlighted as a key factor to engagement, with challenges mentioned by AF patients given its episodic nature and sporadic symptoms leading to high variability in individual needs regarding education and self-management between different people and across time. AF patients often

XSL•FO
RenderX

experience changes in their symptoms, triggers, and treatments, and their care should be reflective of these varying needs. Newly diagnosed AF patients will often require more education about AF [14], whereas those who have had AF for a prolonged time may have developed the skills to be able to self-manage but still require reinforcement of key concepts [24]. This highlights the need for adaptive health interventions that account for a change in patient needs throughout their disease and adapt accordingly, with the goal of maintaining engagement [19].

Trustworthiness was highlighted by participants as key to engagement with the AI intervention. With the rising popularity of chatbot-based search engines like the large language model ChatGPT, it is increasingly important to consider the factors that influence user trust in health information delivered via digital media [28]. Despite recent studies showing ChatGPT performs remarkably well in answering cardiovascular-related queries [29], its potential use as a health education tool is uncertain, given the lack of transparency in its sources of information. In our study, the credibility of the institution delivering the AI intervention was mentioned by participants as key in trusting the automated phone calls, given the proliferation of unsolicited and fraudulent "robocalls." Patients' acceptability of AI technology is still not well understood but seems to be higher when the intervention is recommended by clinicians [30,31].

Implications

Programs like CHAT-AF could be used to deliver frequent education and self-management support between clinical visits to patients with AF at scale. In addition, conversational AI automated phone calls can improve access to care for traditionally disadvantaged AF patients in rural and remote communities [32], as phone calls are not reliant on internet connectivity, phone models, or operating systems. Engagement of digital interventions in less digitally literate populations remains challenging, however, the simplistic navigation of conversational AI interfaces (ie, requiring users to listen to information and respond using their voice) could provide utility in engaging with older adults who are most susceptible to AF alongside those with low technology literacy and should be the focus of future research. The findings of this work can also guide the implementation of other digital solutions for patients with AF, particularly regarding the optimization of interactivity and personalization, as well as the trustworthiness of providers and technology.

Strengths and Limitations

This is the first qualitative study to explore the user perspectives of using conversational AI in supporting patients with AF. The premature study completion due to the acquisition of the technology partner by another organization meant that some participants had experienced fewer outreaches via conversational AI calls, which may have impacted how they reported on the AI component, however, all participants did receive a minimum of 2 outreaches (of 7) via the calls. As a single-site study, with a predominantly Caucasian and English-speaking population, generalizability is limited, and the results should be interpreted within the limitations of sample characteristics. For instance, the study population was younger (mean 65.4, SD 11.9 years) and had a lower proportion of women (30%) than previously reported literature of the AF patient population in Australia (mean age of 75 years and approximately 45% female) [33]. Selection bias may have occurred where those who agreed to the interviews may have been more engaged, interested, and motivated to provide feedback. We did not interview all trial participants and may have missed some unique experiences, although to address this, interviews were conducted until data saturation was reached. We did not involve patients in the design of the interview guide, and this may have resulted in not capturing experiences that are important to people with AF, however, we did prompt participants to share any additional comments regarding the research and intervention during the interview. Moreover, interview duration varied among participants and was on average around 20 minutes, it is possible that we may not have captured complex concepts in this time, however, the researcher used several prompts in the interview guide to facilitate further discussion during the interviews.

Conclusions

This qualitative study of user experiences with a conversational AI program found that AF patients perceived it could address gaps in current models of AF care by improving access to information and providing continuity in support. Interactions with the voice-based conversational AI were perceived positively by participants, although the need for additional dialogue flexibility and response options was mentioned to better capture their AF experience. Future studies should explore whether delivering more personalized content and providing different interaction options can improve the engagement and effectiveness of such interventions.

Acknowledgments

The authors would like to acknowledge Emma Charlston who was involved in the project management of this trial and Ayesha Thaliph and Alison Robinson for transcribing the interviews. The authors would like to thank the technology partner (HMS) and team members for their efforts in the development of the intervention. This work is supported by Digital Health Cooperative Research Centers Limited. Digital Health Cooperative Research Centers Limited is funded under the Commonwealth's Cooperative Research Centers Program. The funder played no role in the study design, data collection, analysis, and interpretation of data, or the writing of this manuscript.

XSL•FO
RenderX

Authors' Contributions

RT and LL contributed to the conception and design of the work. RT, LL, and BS contributed to the acquisition, analysis, or interpretation of data for the work. RT drafted the manuscript. LL, TS, and CKC critically revised the manuscript. All gave final approval and agreed to be accountable for all aspects of work, ensuring integrity and accuracy.

Conflicts of Interest

None declared

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 131 KB - jmir_v27i1e64325_app1.pdf]

Multimedia Appendix 2
Characteristics of intervention participants and interview cohort.
[PDF File (Adobe PDF File), 158 KB - jmir_v27i1e64325_app2.pdf]

Multimedia Appendix 3
Illustrative quotes for themes and subthemes.
[DOCX File , 44 KB - jmir_v27i1e64325_app3.docx]

References

1. Krijthe BP, Kunst A, Benjamin EJ, Lip GYH, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J 2013;34(35):2746-2751 [FREE Full text] [doi: 10.1093/eurheartj/eht280] [Medline: 23900699]
2. Stewart S, Murphy NF, Walker A, McGuire A, McMurray JJV. Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the UK. Heart 2004;90(3):286-292. [doi: 10.1136/hrt.2002.008748] [Medline: 14966048]
3. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax J, Blomström-Lundqvist C, et al. Corrigendum to: 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021;42(40):4194. [doi: 10.1093/eurheartj/ehab648] [Medline: 34520521]
4. Kirchhof P. The future of atrial fibrillation management: integrated care and stratified therapy. Lancet 2017;390(10105):1873-1887. [doi: 10.1016/S0140-6736(17)31072-3] [Medline: 28460828]
5. Goldenthal IL, Sciacca RR, Riga T, Bakken S, Baumeister M, Biviano AB, et al. Recurrent atrial fibrillation/flutter detection after ablation or cardioversion using the AliveCor KardiaMobile device: iHEART results. Cardiovasc electrophysiol 2019;30(11):2220-2228. [doi: 10.1111/jce.14160] [Medline: 31507001]
6. Desteghe L, Germeys J, Vijgen J, Koopman P, Dilling-Boer D, Schurmans J, et al. Effectiveness and usability of an online tailored education platform for atrial fibrillation patients undergoing a direct current cardioversion or pulmonary vein isolation. Int J Cardiol 2018;272:123-129. [doi: 10.1016/j.ijcard.2018.07.065] [Medline: 30049498]
7. Chen YC, Roebuck AE, Sami A, Ersin OH, Mirro MJ. The use of electronic personal health records to improve medication adherence and patient engagement: a randomized study of non-valvular atrial fibrillation patients. J Innov Card Rhythm Manag 2017;8(8):2804-2813 [FREE Full text] [doi: 10.19102/icrm.2017.080803] [Medline: 32494465]
8. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
9. Schachner T, Keller R, V Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
10. Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of voice-based conversational artificial intelligence for basal insulin prescription management among patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2023;6(12):e2340232 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.40232] [Medline: 38039007]
11. Kassavou A, Mirzaei V, Brimicombe J, Edwards S, Massou E, Prevost AT, et al. A highly tailored text and voice messaging intervention to improve medication adherence in patients with either or both hypertension and type 2 diabetes in a UK primary care setting: feasibility randomized controlled trial of clinical effectiveness. J Med Internet Res 2020;22(5):e16629 [FREE Full text] [doi: 10.2196/16629] [Medline: 32427113]
12. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]

XSL•FO
RenderX

13. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
14. McCabe P, Schumacher K, Barnason S. Living with atrial fibrillation: a qualitative study. J Cardiovasc Nurs 2011;26(4):336-344. [doi: 10.1097/JCN.0b013e31820019b9] [Medline: 21263348]
15. Guhl E, Althouse AD, Pusateri AM, Kimani E, Paasche-Orlow MK, Bickmore TW, et al. The atrial fibrillation health literacy information technology trial: pilot trial of a mobile health app for atrial fibrillation. JMIR Cardio 2020;4(1):e17162 [FREE Full text] [doi: 10.2196/17162] [Medline: 32886070]
16. Magnani JW, Schlusser CL, Kimani E, Rollman BL, Paasche-Orlow MK, Bickmore TW. The atrial fibrillation health literacy information technology system: pilot assessment. JMIR Cardio 2017;1(2):e7 [FREE Full text] [doi: 10.2196/cardio.8543] [Medline: 29473644]
17. Hirschey J, Bane S, Mansour M, Sperber J, Agboola S, Kvedar J, et al. Evaluating the usability and usefulness of a mobile app for atrial fibrillation using qualitative methods: exploratory pilot study. JMIR Hum Factors 2018;5(1):e13. [doi: 10.2196/humanfactors.8004] [Medline: 29549073]
18. Mihas P, Rosman L, Armbruster T, Walker J, Deyo Z, Gehi A. Patient perspectives on performance of a smartphone app for atrial fibrillation self-management. Patient Prefer Adherence 2022;16:2799-2810 [FREE Full text] [doi: 10.2147/PPA.S366963] [Medline: 36281351]
19. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
20. Laranjo L, Shaw T, Trivedi R, Thomas S, Charlston E, Klimis H, et al. Coordinating health care with artificial intelligence-supported technology for patients with atrial fibrillation: protocol for a randomized controlled trial. JMIR Res Protoc 2022;11(4):e34470 [FREE Full text] [doi: 10.2196/34470] [Medline: 35416784]
21. Nelson J. Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research. Qualitative Research 2016;17(5):554-570. [doi: 10.1177/1468794116679873]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Klavebäck S, Svennberg E, Nymark C, Braunschweig F, Lidin M. Management of modifiable risk factors and comorbidities in atrial fibrillation: suggestions for improvement from a patient perspective. Eur J Cardiovasc Nurs 2024;23(2):169-175. [doi: 10.1093/eurjcn/zvad055] [Medline: 37249076]
24. Ferguson C, Hickman LD, Lombardo L, Downie A, Bajorek B, Ivynian S, et al. Educational needs of people living with atrial fibrillation: a qualitative study. J Am Heart Assoc 2022;11(15):e025293 [FREE Full text] [doi: 10.1161/JAHA.122.025293] [Medline: 35876410]
25. Salmasi S, De Vera MA, Barry A, Bansback N, Harrison M, Lynd LD, et al. Assessment of condition and medication knowledge gaps among atrial fibrillation patients: a systematic review and meta-analysis. Ann Pharmacother 2019;53(8):773-785. [doi: 10.1177/1060028019835845] [Medline: 30829052]
26. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
27. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
28. Stokel-Walker C. AI chatbots are coming to search engines - can you trust the results? Nature 2023. [doi: 10.1038/d41586-023-00423-4] [Medline: 36781968]
29. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. JAMA 2023;329(10):842-844 [FREE Full text] [doi: 10.1001/jama.2023.1044] [Medline: 36735264]
30. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021;3(9):e599-e611 [FREE Full text] [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
31. Jermutus E, Kneale D, Thomas J, Michie S. Influences on user trust in healthcare artificial intelligence: a systematic review. Wellcome Open Res 2022;7:65. [doi: 10.12688/wellcomeopenres.17550.1]
32. Essien UR, Kornej J, Johnson AE, Schulson LB, Benjamin EJ, Magnani JW. Social determinants of atrial fibrillation. Nat Rev Cardiol 2021;18(11):763-773 [FREE Full text] [doi: 10.1038/s41569-021-00561-0] [Medline: 34079095]
33. Wong CX, Brooks AG, Cheng Y, Lau DH, Rangnekar G, Roberts-Thomson KC, et al. Atrial fibrillation in Indigenous and non-Indigenous Australians: a cross-sectional study. BMJ Open 2014;4(10):e006242 [FREE Full text] [doi: 10.1136/bmjopen-2014-006242] [Medline: 25344486]

XSL•FO
RenderX

Abbreviations

AF: atrial fibrillation
AI: artificial intelligence
CHAT-AF: Coordinating Health care with AI–supported Technology for patients with AF
COREQ: Consolidated criteria for Reporting Qualitative
RCT: randomized controlled trial

Edited by T de Azevedo Cardoso; submitted 21.07.24; peer-reviewed by N Isakadze, P Mihas, M Chatzimina; comments to author 30.09.24; revised version received 28.10.24; accepted 06.12.24; published 12.03.25.

Please cite as:
Trivedi R, Shaw T, Sheahen B, Chow CK, Laranjo L
Patient Perspectives on Conversational Artificial Intelligence for Atrial Fibrillation Self-Management: Qualitative Analysis
J Med Internet Res 2025;27:e64325
URL: https://www.jmir.org/2025/1/e64325
doi:10.2196/64325
PMID:

©Ritu Trivedi, Tim Shaw, Brodie Sheahen, Clara K Chow, Liliana Laranjo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Influencing the Intentions of Patients With Inflammatory Rheumatic Diseases to Use a Digital Human for Medication Information: Qualitative Study

Lex L Haegens[1], MSc; Victor J B Huiskes[2,3], PhD; Bart J F van den Bemt[1,2,3], Prof Dr; Charlotte L Bekker[3], PhD

[1]Department of Research, Sint Maartenskliniek, Ubbergen, The Netherlands

[2]Department of Pharmacy, Sint Maartenskliniek, Ubbergen, The Netherlands

[3]Department of Pharmacy, Radboudumc, Nijmegen, The Netherlands

Corresponding Author:
Victor J B Huiskes, PhD
Department of Pharmacy
Sint Maartenskliniek
Hengstdal 3
Ubbergen, 6574 NA
The Netherlands
Phone: 31 24 365 9512
Email: v.huiskes@maartenskliniek.nl

Abstract

Background: Introduction: Patients with inflammatory rheumatic diseases (IRDs) frequently experience drug-related problems (DRPs). DRPs can have negative health consequences and should be addressed promptly to prevent complications. A digital human, which is an embodied conversational agent, could provide medication-related information in a time- and place-independent manner to support patients in preventing and decreasing DRPs.

Objective: This study aims to identify factors that influence the intention of patients with IRDs to use a digital human to retrieve medication-related information.

Methods: A qualitative study with 3 in-person focus groups was conducted among adult patients diagnosed with an IRD in the Netherlands. The prototype of a digital human is an innovative tool that provides spoken answers to medication-related questions and provides information linked to the topic, such as (instructional) videos, drug leaflets, and other relevant sources. Before the focus group, participants completed a preparatory exercise at home to become familiar with the digital human. A semistructured interview guide based on the Proctor framework for implementation determinants was used to interview participants about the acceptability, adoption, appropriateness, costs, feasibility, fidelity, penetration, and sustainability of the digital human. Focus groups were recorded, transcribed, and analyzed thematically.

Results: The participants included 22 patients, with a median age of 68 (IQR 52-75) years, of whom 64% (n=22) were female. In total, 6 themes describing factors influencing patients' intention to use a digital human were identified: (1) the degree to which individual needs for medication-related information are met; (2) confidence in one's ability to use the digital human; (3) the degree to which using the digital human resembles interacting with a human; (4) technical functioning of the digital human; (5) privacy and security; and (6) expected benefit of using the digital human.

Conclusions: The intention of patients with IRDs to use a novel digital human to retrieve medication-related information was influenced by factors related to each patient's information needs and confidence in their ability to use the digital human, features of the digital human, and the expected benefits of using the digital human. These identified themes should be considered during the further development of the digital human and during implementation to increase intention to use and future adoption. Thereafter, the effect of applying a digital human as an instrument to improve patients' self-management regarding DRPs could be researched.

(J Med Internet Res 2025;27:e57697) doi:10.2196/57697

XSL•FO
RenderX

KEYWORDS

digital human; information provision; intention to use; qualitative study; focus groups; drug-related problems; medication safety; safety information; information seeking; Netherlands; Pharmacotherapy; medication; telehealth; communication technologies; medication information; rheumatic diseases; rheumatology

Introduction

Pharmacotherapy is a cornerstone treatment for patients with inflammatory rheumatic diseases (IRDs) [1]. Although disease-modifying antirheumatic drugs (DMARDs) are highly effective in reducing disease activity and increasing daily functioning, patients with IRDs frequently experience drug-related problems (DRPs). Previous research demonstrated that patients with IRDs report on average 5 DRPs within 8 weeks, such as side effects, difficulties with medication management, and information needs [2]. DRPs can potentially lead to negative outcomes of medication use, which, if not addressed promptly, can lead to decreased treatment effects and adverse drug events [3]. Because adverse drug events can increase unplanned hospital admissions and morbidity, and thus health care utilization, DRPs must be addressed quickly to prevent these negative consequences [4-7].

Currently, contact between health care providers (HCPs) and patients is infrequent, with generally up to 4 routine consultations annually. However, DRPs occur throughout the year outside these consultations and could thus be overlooked. Furthermore, patients do not always report DRPs during consultations, and if they do, HCPs do not always act upon them [8]. These factors could delay the identification and resolution of DRPs, potentially putting patients at unnecessary risk. To deliver high-quality chronic disease care according to the Chronic Care Model, patients should be informed and activated, HCPs should be prepared and proactive, and interactions between patients and HCPs should be productive [9].

This study will focus on the informed and activated patient by supporting patient self-management, which can ensure that the limited time HCPs have for contact can be spent on issues that truly require in-person interaction. Furthermore, research shows that patient involvement is essential to identify clinically relevant DRPs promptly [10,11]. This can be achieved by supporting patient self-management by providing medication-related information [12]. It is conceivable that the more patients are informed regarding DRPs, the better they can identify and deal with these DRPs themselves. Thus, supporting patients in becoming more informed could potentially help resolve DRPs on time.

Telehealth, which entails the use of communication technologies to deliver health care at a distance [13], has the potential to support patient self-management through information provision. One such telehealth channel to provide medication-related information is a digital human, which is an embodied conversational agent. Conversational agents are software programs designed to interact with users through natural language dialogue (ie, speech or text), and embodied conversational agents have an additional virtual form represented by an avatar to provide information via face-to-face conversations. Digital humans have successfully been applied in health care, such as for educating patients about care after hospital discharge [14], promoting walking in older patients [15], and facilitating accessibility to web-based health information about breastfeeding [16]. Furthermore, numerous digital humans designed for self-management in chronic diseases have been studied [17].

A digital human offers various advantages for supporting patient self-management, such as time- and place-independent access to information; the ability to process large amounts of information quickly; and intuitive, easily accessible information by mimicking human-to-human interaction [18] (eg, allowing speech as input and output for populations with low literacy levels). It can also provide visual support, such as instructional videos, or nonverbal output like medication leaflets, without requiring involvement from HCPs.

Patients with IRDs reported these advantages of digital humans as influencing their preference for telehealth channels when seeking support for medication use, particularly when offered options with or without HCP contact [19]. Additionally, patients believe telehealth can effectively provide informational support about medication use [12]. Thus, a digital human could support patient self-management by providing easily accessible medication-related information and potentially resolving DRPs more efficiently. However, because using a digital human for providing medication-related information is a novel approach, insights into what factors determine patients' intention to use a digital human are lacking. Therefore, this study aims to identify factors shaping the intention of patients with IRDs to use a digital human for accessing medication-related information.

Methods

Overview

This qualitative study is reported according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) [20].

Study Design and Setting

An explorative, qualitative study using focus groups was performed. In total, 3 focus groups were held with the aim of reaching data saturation (ie, the point where additional data collection does not produce new relevant themes). For each focus group, between 8 and 12 participants were included. Focus groups were held in the Sint Maartenskliniek Hospital in the Netherlands from February to March 2023.

Participants

Adult patients (≥18 years of age) were eligible for inclusion if they were prescribed at least 1 DMARD for an IRD from the outpatient pharmacy of the Sint Maartenskliniek Hospital, were proficient in Dutch, and owned an electronic device capable of accessing the digital human (ie, smartphone, tablet, laptop, or desktop) at the time of the study. Participants were purposively

XSL•FO
RenderX

sampled based on age, sex, diagnosis, and digital inclination (defined as patients using or not using the web-based personal health record used by the Sint Maartenskliniek Pharmacy). Eligible patients were recruited via email, which contained study information and a form to state availability for the selected dates for focus groups. Written informed consent was obtained in person from participants prior to the focus groups. Incentives for participation were offered in the form of a 20 Euro (US $20.84) gift card, and travel expenses were fully reimbursed.

Digital Human

Pharmi [21], the digital human used during this study, was developed by Pharmi BV and Deloitte in collaboration with the Sint Maartenskliniek Hospital. Pharmi provides spoken answers to drug-related questions and offers additional information, such as (instructional) videos, drug leaflets, and other relevant sources, for approximately 80 different DMARDs. During the study, a prototype of Pharmi was used to provide information and offer predefined answers to common questions, similar to those found on a “Frequently Asked Questions” page. Topics covered included the drug’s mechanism of action, common side effects, instructional videos, storing instructions for antirheumatic drugs, information on specific IRDs, and logistical details such as how to obtain repeat prescriptions (Figure 1).

Figure 1. The digital human "Pharmi." Sourced from Pharmi BV.



Preparation

Participants were asked to complete a preparatory exercise at home prior to the focus group to familiarize themselves with the digital human. This exercise consisted of a set of predetermined frequently asked questions personalized to each participant’s current DMARD and diagnosis (Multimedia Appendix 1). In addition, participants were asked the following questions: (1) “How would you rate the digital human on a scale from 1 to 10?” (2) “What advantages do you see with the digital human? Name a maximum of 2” (3) “What disadvantages do you see with the digital human? Name a maximum of 2.”

Focus Groups

Focus groups were held at the Sint Maartenskliniek Hospital and lasted 2 hours. Each focus group was led by an experienced moderator (author BJFB or CLB), and an assistant moderator (author LLH) was present to facilitate assignments and take additional field notes. All moderators were researchers at the time of the focus groups. A topic guide was used to standardize and structure each focus group (see Multimedia Appendix 1). This topic guide was based on the implementation outcomes outlined by Proctor et al [22] (ie, acceptability, adoption, appropriateness, costs, feasibility, fidelity, penetration, and sustainability), which are a frequently applied set of outcome measures that describe the concept of implementation outcomes. Service and client outcomes were not included in the topic guide, as these are intended to assess interventions after implementation. All focus groups were audio recorded, and recordings were transcribed verbatim. After 2 focus groups, data saturation was discussed to determine if more than 3 focus groups were necessary.

Analyses of Transcripts

Transcripts were systematically analyzed using inductive thematic content analysis according to the steps outlined by Braun and Clarke [23] using Atlas.ti 9 (ATLAS.ti Scientific Software Development GmbH) and Microsoft Excel (Microsoft Corp). Transcripts were coded by 3 authors (LLH, VJBH, and CLB). First, open codes were generated for one transcript by LLH and checked by VJBH, after which discrepancies were discussed until consensus was achieved. Second, axial codes were generated by LLH and then discussed by LLH, VJBH, and CLB until consensus was achieved. Third, preliminary themes were constructed by LLH, VJBH, and CLB, after which open and axial coding for the remaining 2 focus groups was

XSL•FO
RenderX

performed and themes were revised accordingly. Relevant quotes from participants were selected to support the themes, which were translated into English.

Ethical Considerations

The Medical Ethics Research Committee of Arnhem-Nijmegen waived official ethical approval (case number 2023-16146) as this study was deemed exempt from the Medical Research Involving Humans Act.

Results

Overview

A total of 540 eligible patients were approached for participation, among whom 214 (40%) responded and 31 (6%) agreed to participate. Of these, 9 patients did not participate in a focus group due to sickness (n=4, 44%), sudden unavailability (n=1, 11%) or unspecified reasons (n=4, 44%). Table 1 shows the participants' characteristics.

A total of 6 themes reflecting factors influencing participants' intention to use a digital human to retrieve medication-related information were identified relating to the patient, digital human, and the expected benefit from using the digital human (Figure 2).

Table 1. Characteristics of the study population (N=22).

Age (years), median (IQR)	68 (52-75)
Female sex, n (%)	14 (64)
Diagnosis, n (%)	
Rheumatoid arthritis	14 (64)
Psoriatic arthritis	6 (27)
Other	2 (9)
Digital inclination (yes), n (%)	11 (50)

Figure 2. Schematic representation of overarching themes reflecting factors influencing participants' intentions to use a digital human for retrieving medication-related information.

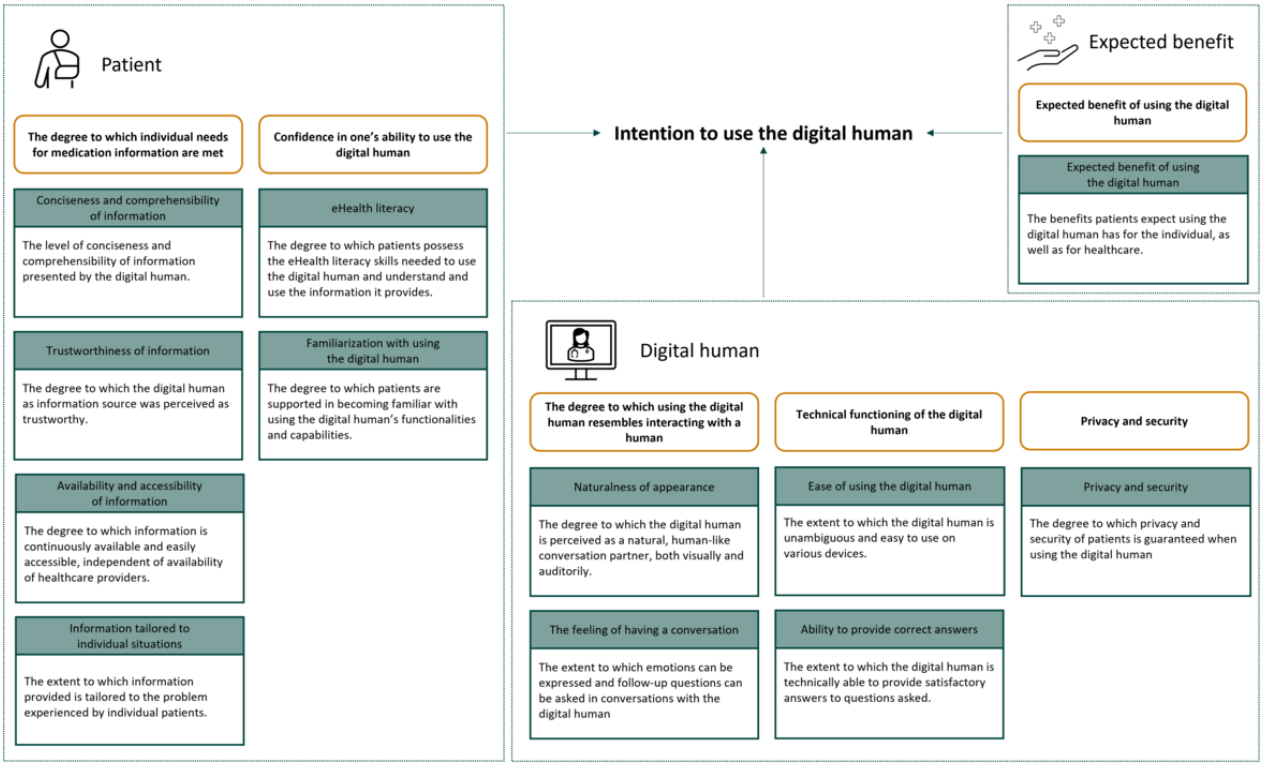

XSL•FO
RenderX

Theme 1: Degree to Which Individual Needs for Medication-Related Information Are Met

Explanation

The intention to use a digital human to retrieve medication-related information was dependent on patients' individual information needs. These needs varied between participants and related to different aspects of the information, and in some cases, the problem experienced.

Conciseness and Comprehensibility of Information

Most participants preferred the concise information presented by the digital human compared with the relatively abundant information found in medication information leaflets, general brochures about medical treatment, and other web-based information sources. On the other hand, some participants felt that the digital human presented limited information compared with other medication information sources.

> *For some people, listing five main side effects is sufficient, but I also know that some people do not dare to take the medication because it has so many side effects.* [Female, 75 years]

Participants reported comprehensibility of information as a factor that influenced their intention to use the digital human. Although some participants found the information presented by the digital human rather vague, others appreciated the clear information it provided. The degree to which the information provided by the digital human was experienced as comprehensible varied. Additionally, the digital human did not verify that the patient understood the information provided.

Trustworthiness of Information

The participants highlighted the importance of the trustworthiness of information provided by the digital human. They perceived information provided by the digital human as trustworthy, particularly when compared to web-based sources, as the digital human was deployed by the hospital. Participants suggested that trustworthiness would increase if they were informed about the availability of the digital human by reliable authorities, such as patient associations or HCPs from the hospital where they were treated. Furthermore, participants trusted that the digital human would provide the most up-to-date information available.

> *If you look at the digital human you get the feeling the information is provided by the pharmacist, and if you just search on Google, you get the feeling it is provided by someone random. So, I see that as an advantage, I would have more trust in that [the digital human].* [Female, 78 years]

Availability and Accessibility of Information

The participants expressed a desire for continuously available information. They sometimes felt that they had to wait for the next HCP consultation to retrieve information and questioned this relatively low accessibility of information. They mentioned that the digital human fulfilled the need for easily accessible information by being available 24/7 without having to wait for an HCP to become available.

> *The benefit of this [the digital human] is that I do not have to wait until I see my treating physician again. I do not have to travel to the hospital to get answers.* [Male, 76 years]

The digital human was regarded as a fast information source that could directly answer questions, whereas, with other information sources, patients had to search for relevant information more intensively.

Information Tailored to Individual Situations

Another factor influencing patients' intentions was to what extent the information was tailored to their concerns. Participants intended to use the digital human for general questions, such as medication side effects or potential medication interactions. For more specific, personal questions, such as regarding the choice of treatment, participants' intention to use the digital human was lower. The main reason was that the digital human provided predefined answers to predefined questions and could not generate tailored information.

> *Well, it is all programmed. If you have really detailed questions, I wouldn't even bother trying, I think.* [Female, 67 years]

Theme 2: Confidence in One's Ability to Use the Digital Human

Explanation

The participants found it important to feel confident in using the digital human, as this would increase their chances of using it more often.

eHealth Literacy

The participants felt that adequate eHealth literacy was needed for using the digital human, including digital skills and health literacy. Regarding digital skills, patients indicated that proficiency in using an electronic device on which the digital human was available was necessary. Some participants believed that older people could be less digitally skilled and thought the digital human was less suitable for this age group.

> *Of course, there are also many people, especially in the older population, who experience difficulties with all digital tools.* [Male, 77 years]

Furthermore, participants felt that health literacy was important for adequately using the digital human. It would enable patients to formulate and articulate questions clearly and understand and apply the health-related information provided. Since the digital human focuses on verbal communication, participants believed it could be especially beneficial for individuals with low literacy (ie, those who struggle with reading and writing), enabling them to ask questions verbally and receive spoken responses.

Familiarization With Using the Digital Human

Participants noted that familiarity with using the digital human would increase their intention to use it. They suggested that HCPs provide clear instructions and support during initial use. Additionally, participants emphasized the importance of having sufficient time to explore and understand the digital human's functionalities and capabilities, as this would increase their

XSL•FO
RenderX

confidence and intention to use it. However, they also highlighted that the quality of their early experiences (whether positive or negative) would influence their use intention.

> *…The fact that you get used to talking with someone, communicating with someone who does not exist...* [Male, 76 years]

Theme 3: Degree to Which Using the Digital Human Resembles Interacting With a Human

Explanation

The extent to which interaction with the digital human resembled human-to-human interaction when retrieving medication-related information influenced participants' intention to use it. Some participants believed that a human-like interaction would positively influence their intention to use the digital human, while others indicated that this did not impact their intention.

Naturalness of Appearance

The perceived naturalness and human-like qualities of the digital human influenced the participants' intention to use it. They felt that the digital human appeared artificial and lacked sufficient human traits. Specifically, they felt that the digital human did not speak like a real person, as it misplaced emphases in sentences, lacked proper pronunciation, and had a tinny tone, leading to nonfluent communication. While the presence of a digital human image increased some participants' attention to the provided answers, others found the unnatural facial expressions and constant movement of the image distracting, negatively influencing their use intention.

The Feeling of Having a Conversation

Participants found interactions with the digital human more impersonal, citing the inability to express emotions compared to conversations with an HCP.

> *When interacting with the digital human, I don't feel that I can genuinely express certain emotions in that interaction.* [Male, 76 years]

Furthermore, participants viewed the lack of follow-up question capabilities negatively. They noted that the digital human did not allow users to ask follow-up questions after receiving an answer, nor did it initiate follow-up questions itself, further diminishing the sense of a meaningful dialogue with the digital human.

> *The digital human, who, by the way, looks very pleasant, only provides preprogrammed responses. So, there is no dialogue where you can ask a question yourself.* [Male, 77 years]

Theme 4: Technical Functionalities of the Digital Human

Explanation

The technical performance of the digital human influenced the participants' intention to use it. Participants expressed frustration when retrieving information was difficult, which negatively influenced their intention to use the digital human.

Ease of Using the Digital Human

Participants emphasized that using the digital human should feel straightforward and easy. While some found it easy to use, others did not due to the prototype's technical flaws. Several issues were reported regarding the synchronicity between the digital human's image and sound. This was especially problematic for patients who rely on lip-reading.

> *What I found very difficult, as I have a hearing impairment, I have a hearing aid and heavily rely on lip reading, the digital human does not synchronize. I think... Well, that's a turn-off for me.* [Female, 50 years]

Additionally, participants found using the digital human on a mobile phone more challenging compared to a laptop or computer. They stressed that the digital human should function smoothly on various types of devices and be easy to read.

Participants had different opinions on the importance of a safe login procedure. Most felt that requiring a login was too cumbersome, while others valued the security, especially if a login could lead to more tailored answers based on the patient's current medication extracted from a personal health record.

Ability to Provide Correct Answers

Many participants experienced difficulties receiving satisfactory answers, mainly due to the digital human not understanding their questions. In several cases, participants did not receive an answer at all, while in other cases, the digital human answered a different question than the one asked. Participants reported that they became impatient and frustrated when having to repeat the question, negatively influencing their intention to use the digital human. Moreover, participants noted that some medication names are difficult to pronounce, which can vary across individuals. They suggested improving speech recognition so the digital human could better handle different pronunciations and accents.

> *If "Have you not heard it correctly," then you have to repeat [your question]. You know, you become impatient.* [Female, 50 years]

> *It mainly has to do with how you pronounce the medication, if she [the digital human] understands it. And I also think it depends on the dialect. My wife tried it as well, there she [the digital human] worked and for me, it did not.* [Male, 79 years]

Theme 5: Privacy and Security

Participants emphasized the importance of privacy and security in their intention to use the digital human. Regarding security, they expressed worries about the possibility of microphones being tapped and the risks associated with cookies from other websites, which could negatively influence their intention to use the digital human. Some participants felt privacy was guaranteed, as questions could be asked anonymously.

However, others had concerns about privacy when using the digital human in public spaces, as verbal questions and answers could be heard by others.

XSL•FO
RenderX

> *I thought that with anonymity, perhaps some people might feel embarrassed to ask questions because they don't know (…)which they think everyone should know. That's why anonymity may be nice.* [Female, 75 years]

Theme 6: Expected Benefits of Using the Digital Human

The anticipated benefits of using the digital human, both at the individual level and societal level, influenced participants' intentions to use it. On an individual level, the digital human was perceived as a tool to enhance self-management by giving patients more control over their own care.

> *I think it is beneficial for health care providers, the better informed patients are about their medication, the better they are doing. As I just said, with medication adherence, but also with trust patients have in their medication. I think that is very important.* [Male, 77 years]

The participants viewed the digital human as a way to increase their knowledge about medication and thus their involvement in their treatment. They felt that the digital human enabled them to address issues independently without having to reach out to HCPs, thereby avoiding additional burdens on HCPs.

> *As has been said before, you don't have to bother anyone, you can look it up yourself.* [Male, 79 years]

Furthermore, participants believed that using the digital human could improve their digital skills and preparedness for the digitization of health care.

Discussion

Principal Findings

This qualitative study showed that the intention of patients with IRDs to use a digital human for retrieving medication-related information was influenced by several factors. These factors included individual patients' information needs and their confidence in using the digital human; the digital human's features, including human-like interaction and technical functionalities; and the expected benefits of using it.

Over the past decade, digital humans have been developed for various chronic conditions to facilitate screening, diagnosis, and education [17]. However, using a digital human specifically for retrieving medication-related information in rheumatic diseases is a novel approach. As a result, direct comparisons with earlier studies investigating digital humans in the same context are difficult.

In this study, the intention to use the digital human largely depended on patients' individual information needs, including their need for concise, trustworthy, accessible, and tailored information. These factors have also been described in other studies regarding the information needs of patients with rheumatic diseases [12,24,25]. Additionally, patients' confidence in their ability to use the digital human and their eHealth literacy skills influenced their use intentions. This aligns with previous research showing that health literacy is associated with the ability to retrieve online health information [26,27] and evaluate its quality [28].

Another theme influencing intention was the ease of use of the digital human. This relation is widely documented in frameworks such as the Technology Acceptance Model, which describes "perceived ease of use" as a factor that indirectly influences the intention to use [29]. Given that the digital human used in this study was a prototype, participants' remarks regarding ease of use should be considered during further development to maximize future intentions to use and thus adoption.

Additionally, participants emphasized the importance of the digital human's technical functionalities, particularly its ability to provide correct answers. A mixed methods study on the use of digital humans for patients with dementia similarly found that technical problems and issues with speech recognition negatively impacted patient's intention to use [30]. In this study, participants also indicated that a negative first experience with the digital human (ie, an experience in which a correct answer was not provided by the digital human) could deter future use intention. Therefore, ensuring an optimal first experience is critical for successful implementation and adoption by patients.

Research on digital humans applied in different contexts suggests that design features, such as facial expressions and tone of voice, influence use intentions [31]. For example, a previous study on a digital human for type 2 diabetes self-management found that its acceptability partly depended on the congruence between its verbal and nonverbal communication [32]. Similarly, in this study, participants indicated that the degree to which the digital human resembled a natural, human-like conversation partner influenced their intention to use it.

Various communication and information channels are available in health care, including frequently asked questions, phone or video calls, and 2-way text messaging with HCPs. These channels vary in interaction levels, user experiences, accessibility, and monetary or personnel costs associated with using the channel. It is conceivable that the digital human is not always the most suitable channel in every situation and for every individual (eg, a patient can experience multiple problems according to which information should be tailored). However, offering the digital human alongside existing channels allows patients to choose the option that best matches their needs and preferences. This could increase the actual usage of such channels to retrieve medication information, leading to patients gaining more medication knowledge, which is recognized as an important element of patients' self-management [33].

Strengths and Limitations

A notable strength of this study is its theory-based approach to constructing the topic guide, which was informed by the Proctor framework for successful implementation [22]. The factors identified in this study are valuable for integrating the digital human into routine care. For example, integrating the digital human into digital patient health records could enhance the trustworthiness and accessibility of medication information. These factors are also relevant for further development of the

XSL•FO
RenderX

digital human to increase future adoption. For example, speech recognition could be improved to increase the digital human's ability to provide correct answers. This is especially valuable because the uptake of telehealth is low among patients with IRDs despite their willingness to use it [34].

A few limitations should be acknowledged. First, the digital human used in this study was a prototype and had several technical flaws, such as suboptimal processing of participants' questions participants and difficulties in controlling the digital human. These issues may have influenced the range and type of technical factors participants identified as impacting their intention to use the digital human.

Second, the participant group was slightly older than the nonparticipant group. The nonparticipants' median age was 59 (IQR 49-68) years. This might be explained by the fact that 2 out of 3 focus groups focus groups were held face-to-face during office hours, potentially making it easier for older adults with more spare time to attend. Additionally, this study had a low participation rate (6%). Most patients did not respond to the online invitation, and those who declined to participate were not obliged to provide reasons, leaving the reasons for nonparticipation unknown. Third, the participants were recruited from a single, specialized rheumatology clinic and might differ from the general IRD population in in terms of their experiences with the level of quality of care and medication information provided. This may limit the generalizability of the study's findings.

Conclusion

This qualitative study highlights that the intention of patients with IRDs to use a novel digital human for medication-related information is influenced by their individual information needs, confidence in their ability to use the digital human, its features, and the expected benefits of using the digital human. These themes should guide the development and implementation of a digital human to maximize patients' intention to use and adoption. Future research should examine the effect of applying the digital human as a tool to improve patients' self-management of DRPs.

Acknowledgments

The authors would like to thank UneeQ for providing access to the digital human during research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions for focus group participants to prepare for focus groups and topic guide used during focus groups.
[DOCX File , 15 KB - jmir_v27i1e57697_app1.docx]

References

1. Abbasi M, Mousavi MJ, Jamalzehi S, Alimohammadi R, Bezvan MH, Mohammadi H, et al. Strategies toward rheumatoid arthritis therapy; the old and the new. J Cell Physiol 2019 Jul;234(7):10018-10031. [doi: 10.1002/jcp.27860] [Medline: 30536757]
2. Haegens LL, Huiskes VJB, Smale EM, Bekker CL, van den Bemt BJF. Drug-related problems reported by patients with rheumatic diseases: an observational study. BMC Rheumatol 2023 Apr 18;7(1):7 [FREE Full text] [doi: 10.1186/s41927-023-00326-x] [Medline: 37069634]
3. van den Bemt PM, Egberts TC, de Jong-van den Berg LT, Brouwers JR. Drug-related problems in hospitalised patients. Drug Saf 2000 Apr;22(4):321-333. [doi: 10.2165/00002018-200022040-00005] [Medline: 10789826]
4. Leendertse AJ, Egberts ACG, Stoker LJ, van den Bemt PMLA, HARM Study Group. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. Arch Intern Med 2008 Sep 22;168(17):1890-1896. [doi: 10.1001/archinternmed.2008.3] [Medline: 18809816]
5. Leendertse AJ, Van Den Bemt PMLA, Poolman JB, Stoker LJ, Egberts ACG, Postma MJ. Preventable hospital admissions related to medication (HARM): cost analysis of the HARM study. Value Health 2011 Jan;14(1):34-40 [FREE Full text] [doi: 10.1016/j.jval.2010.10.024] [Medline: 21211484]
6. Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes. National Institute for Health and Care Excellence. 2015 Mar 04. URL: https://www.nice.org.uk/guidance/ng5 [accessed 2023-11-29]
7. Magdelijns FJH, Stassen PM, Stehouwer CDA, Pijpers E. Direct health care costs of hospital admissions due to adverse events in The Netherlands. Eur J Public Health 2014 Dec;24(6):1028-1033. [doi: 10.1093/eurpub/cku037] [Medline: 24699427]
8. Huiskes VJB, Cramer-van der Welle CM, van den Ende CHM, van Dijk L, Bos JM, Kramers C, et al. Communication about drug-related problems (DRPs) during patients' visits to Dutch physicians and pharmacies. Health Commun 2020 Feb;35(2):201-208. [doi: 10.1080/10410236.2018.1551301] [Medline: 30526088]
9. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]

XSL•FO
RenderX

10. Kwint HF, Faber A, Gussekloo J, Bouvy ML. The contribution of patient interviews to the identification of drug-related problems in home medication review. J Clin Pharm Ther 2012 Dec;37(6):674-680. [doi: 10.1111/j.1365-2710.2012.01370.x] [Medline: 22861493]
11. Kari H, Kortejärvi H, Airaksinen M, Laaksonen R. Patient involvement is essential in identifying drug-related problems. Br J Clin Pharmacol 2018 Sep;84(9):2048-2058 [FREE Full text] [doi: 10.1111/bcp.13640] [Medline: 29774588]
12. Mathijssen EG, Vriezekolk JE, Eijsbouts AM, van den Hoogen FH, van den Bemt BJ. Support needs for medication use and the suitability of eHealth technologies to address these needs: a focus group study of older patients with rheumatoid arthritis. Patient Prefer Adherence 2018;12:349-358 [FREE Full text] [doi: 10.2147/PPA.S152759] [Medline: 29563778]
13. What is telehealth? NEJM Catalyst. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0268 [accessed 2022-10-17]
14. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2:197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
15. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
16. Edwards RA, Bickmore T, Jenkins L, Foley M, Manjourides J. Use of an interactive computer agent to support breastfeeding. Matern Child Health J 2013 Dec;17(10):1961-1968. [doi: 10.1007/s10995-013-1222-0] [Medline: 23329167]
17. Jiang Z, Huang X, Wang Z, Liu Y, Huang L, Luo X. Embodied conversational agents for chronic diseases: scoping review. J Med Internet Res 2024 Jan 09;26:e47134 [FREE Full text] [doi: 10.2196/47134] [Medline: 38194260]
18. Cassell J, Sullivan J, Prevost S, Churchill E. Embodied Conversational Agents. Cambridge, MA: MIT Press; 2018:1-438.
19. Haegens LL, Huiskes VJB, van der Ven J, van den Bemt BJF, Bekker CL. Factors influencing preferences of patients with rheumatic diseases regarding telehealth channels for support with medication use: qualitative study. JMIR Form Res 2023 Jul 20;7:e45086 [FREE Full text] [doi: 10.2196/45086] [Medline: 37471137]
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
21. Pharmi, the digital pharmacist in your pocket. Pharmi. URL: https://www.pharmi.info/en/news/pharmi-digital-pharmacist-your-pocket [accessed 2022-11-08]
22. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
24. Becker C, Diener M, Hueber AJ, Henes J, Krusche M, Ignatyev Y, et al. Unmet information needs of patients with rheumatic diseases: results of a cross-sectional online survey study in Germany. Int J Environ Res Public Health 2022 Jun 09;19(12) [FREE Full text] [doi: 10.3390/ijerph19127071] [Medline: 35742318]
25. Bekker CL, Mohsenian Naghani S, Natsch S, Wartenberg NS, van den Bemt BJF. Information needs and patient perceptions of the quality of medication information available in hospitals: a mixed method study. Int J Clin Pharm 2020 Dec;42(6):1396-1404 [FREE Full text] [doi: 10.1007/s11096-020-01125-x] [Medline: 32857256]
26. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DLG. "It's got to be on this page": age and cognitive style in a study of online health information seeking. J Med Internet Res 2015;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
27. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: implications for public health practice. J Public Health Manag Pract 2016 Dec 15. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
28. Diviani N, van DPB, Giani S, van WJC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
29. Davis F. A technology acceptance model for empirically testing new end-user information systems: Theory and results [dissertation]. University of Arkansas at Fayetteville. URL: https://www.researchgate.net/publication/35465050_A_Technology_Acceptance_Model_for_Empirically_Testing_New_End-User_Information_Systems [accessed 2024-01-28]
30. Stara V, Vera B, Bolliger D, Rossi L, Felici E, Di Rosa M, et al. Usability and acceptance of the embodied conversational agent Anne by people with dementia and their caregivers: exploratory study in home environment settings. JMIR Mhealth Uhealth 2021 Jun 25;9(6):e25891 [FREE Full text] [doi: 10.2196/25891] [Medline: 34170256]
31. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. Int J Human-Computer Stud 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
32. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. Acceptability of an embodied conversational agent for type 2 diabetes self-management education and support via a smartphone app: mixed methods study. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17038 [FREE Full text] [doi: 10.2196/17038] [Medline: 32706734]

XSL•FO
RenderX

33. Been-Dahmen JM, Walter MJ, Dwarswaard J, Hazes JM, van Staa A, Ista E. What support is needed to self-manage a rheumatic disorder: a qualitative study. BMC Musculoskelet Disord 2017 Feb 16;18(1):84 [FREE Full text] [doi: 10.1186/s12891-017-1440-5] [Medline: 28209161]
34. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DMARD: disease-modifying antirheumatic drug
DRP: drug-related problem
HCP: health care provider
IRD: inflammatory rheumatic disease

Edited by A Mavragani; submitted 23.02.24; peer-reviewed by M Scholten, LR Blanco; comments to author 17.09.24; revised version received 02.10.24; accepted 06.11.24; published 13.03.25.

Please cite as:
Haegens LL, Huiskes VJB, van den Bemt BJF, Bekker CL
Factors Influencing the Intentions of Patients With Inflammatory Rheumatic Diseases to Use a Digital Human for Medication Information: Qualitative Study
J Med Internet Res 2025;27:e57697
URL: https://www.jmir.org/2025/1/e57697
doi:10.2196/57697
PMID:

©Lex L Haegens, Victor J B Huiskes, Bart J F van den Bemt, Charlotte L Bekker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinicians' Perceptions and Potential Applications of Robotics for Task Automation in Critical Care: Qualitative Study

Jiafeng Song[1,2,3], MSc; Rishika Iytha Sridhar[2], MSc; Darlene Marie Rogers[4], PhD; Cheryl Hiddleson[5], MSN; Carolyn Davis[6], MD; Tina Lynn Holden[7], PhD; Shanna Ramsey-Haynes[8], DNP; Lisa Reif[9], DNP; Julie Swann[10], MBA, MHA; Craig S Jabaley[7,11], MD; Mary Gullatte[12], PhD; Rishikesan Kamaleswaran[1,13,14,15], PhD

[1]Department of Biomedical Engineering, Duke University, Durham, NC, United States
[2]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States
[3]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States
[4]Emory Hillandale Hospital, Lithonia, GA, United States
[5]eICU Operations, Emory Healthcare, Atlanta, GA, United States
[6]Department of Surgery, Emory University School of Medicine, Atlanta, GA, United States
[7]Emory Critical Care Center, Emory Healthcare, Atlanta, GA, United States
[8]Emory University Hospital Midtown, Atlanta, GA, United States
[9]Emory University Hospital, Atlanta, GA, United States
[10]Emory Saint Joseph's Hospital, Atlanta, GA, United States
[11]Department of Anesthesiology, Emory University, Atlanta, GA, United States
[12]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States
[13]Department of Anesthesiology, Duke University, Durham, NC, United States
[14]Department of Electrical and Computer Engineering, Duke University, Durham, NC, United States
[15]Department of Surgery, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Jiafeng Song, MSc
Department of Biomedical Engineering
Duke University
2301 Erwin Road
Durham, NC, 27710
United States
Phone: 1 7163523686
Email: sjfsjf2010@gmail.com

Abstract

Background: Interest in integrating robotics within intensive care units (ICUs) has been propelled by technological advancements, workforce challenges, and heightened clinical demands, including during the COVID-19 pandemic. The integration of robotics in ICUs could potentially enhance patient care and operational efficiency amid existing challenges faced by health care professionals, including high workload and decision-making complexities.

Objective: This qualitative study aimed to explore ICU clinicians' perceptions of robotic technology and to identify the types of tasks that might benefit from robotic assistance. We focused on the degree of acceptance, perceived challenges, and potential applications for improving patient care in 5 Southeastern US hospitals between January and August 2023.

Methods: A qualitative study through semistructured interviews and questionnaires was conducted with 15 ICU clinicians (7 nurses, 6 physicians, and 2 advanced practice providers) from 5 hospitals in the Southeast United States. Directed content analysis was used to categorize and interpret participants' statements, with statistical tests used to examine any role-based differences in how they viewed robotic integration.

Results: Among the 15 participants, 73% (11/15) were female, with an average of 6.4 (SD 6.3) years of ICU experience. We identified 78 distinct tasks potentially suitable for robotic assistance, of which 50 (64%) involved direct patient care (eg, repositioning patients and assisting with simple procedures), 19 (24%) concerned indirect patient care (eg, delivering supplies and cleaning), 6 (8%) addressed administrative tasks (eg, answering call lights), and 3 (4%) were classified as mixed direct and

XSL•FO
RenderX

indirect (eg, sitting with a patient to keep them calm). Most participants supported the automation of routine, noncritical tasks (eg, responding to nurse calls and measuring glucose levels), viewing this strategy as a way to alleviate workload and enhance efficiency. Conversely, high-complexity tasks requiring nuanced clinical judgment (eg, ventilator settings) were deemed unsuitable for full automation. Statistical analysis revealed no significant difference in how nurses, physicians, and advanced practice providers perceived these tasks (P=.22).

Conclusions: Our findings indicate a significant opportunity to use robotic systems to perform noncomplex tasks in ICUs, thereby potentially improving efficiency and reducing staff burden. Clinicians largely view robots as supportive tools rather than substitutes for human expertise. However, concerns persist regarding privacy, patient safety, and the loss of human touch, particularly for tasks requiring high-level clinical decision-making. Future research should involve broader, more diverse clinician samples and investigate the long-term impact of robotic assistance on patient outcomes while also incorporating patient perspectives to ensure ethical, patient-centered adoption of robotic technology.

(J Med Internet Res 2025;27:e62957) doi:10.2196/62957

KEYWORDS

robotics; intensive care units; critical care; health care technology; qualitative study

Introduction

The intensive care unit (ICU) cohorts patients who are critically ill and who require sophisticated care to manage severe, life-threatening conditions. ICUs are both labor-intensive and technophilic, offering an opportunity to evaluate the potential role of robotics to facilitate critical care despite potential technical challenges [1]. The ICU environment is characterized by an assortment of medical devices, including ventilators, physiologic monitors, intravenous pumps and lines, feeding tubes, various drainage and catheter systems, and dialysis equipment [2]. Each device requires meticulous management, and the urgency and complexity of critical care mean that timely responses can be a matter of life and death. Health care professionals in these settings face immense challenges, including making critical decisions under pressure, caring for patients who are nonresponsive, managing equipment, and providing essential emotional support during patients' most vulnerable moments. Furthermore, the increasing demand for critical care services exacerbates the strain on the limited pool of skilled health care workers, who operate under significant stress [3]. Despite these hurdles, the dedication and efforts of these professionals in overcoming such challenges are noteworthy.

In the health care industry, and particularly in the ICU, the impact of technological advancement has been significant. The adoption of electronic health records, for example, has streamlined access to patient histories and clinical data, facilitating more timely and more effective treatment decisions [4-7]. Given that critical care clinicians work around the clock, innovations that offload common tasks or enhance patient safety have been instrumental, such as automated intravenous pumps [8]. These devices can integrate with the electronic health record to reduce medication errors, ensure the precise delivery of medications, and automatically document medication administration, significantly alleviating the workload of nursing staff while minimizing the potential for errors.

Robotic technology in ICUs, although not widespread, shows potential for enhancing patient care management. Telepresence robots, for example, have dramatically decreased response times to patients, facilitating timelier interventions and contributing to a reduction in mortality [1,9,10]. In the field of physical therapy and rehabilitation, particularly for survivors of stroke, rehabilitation robots have demonstrated efficacy, improved quality of care, and effectively customized rehabilitation activities for motor function recovery [11-16]. Additionally, robots dedicated to medication and equipment delivery and remote patient monitoring have streamlined care processes, enhancing efficiency in patient management [17-19]. These robotic solutions have consistently yielded positive outcomes, underscoring their value in improving patient care. However, it is crucial to acknowledge their limitations and potential concerns regarding cost-effectiveness. Despite these challenges, the contribution of robotics to health care, including ICUs, remains seemingly promising [20].

Understanding clinician perceptions of robotics within the ICU is paramount, as it directly influences the adoption, use, and effectiveness of these technologies. This study aims to explore the multifaceted views of health care professionals on the deployment of robotics in the ICU, identifying perceived benefits, challenges, and areas for improvement. By elucidating these objectives, including assessing clinician readiness, identifying potential tasks for robotic assistance, and evaluating the impact on patient care, we seek to bridge the translational gap between technological capability and clinical application at the bedside.

Methods

Study Design and Participants

The study was cross-sectional with a qualitative, exploratory design. Directed content analysis was used to identify prominent themes emerging from the data. Participant recruitment occurred at 2 academic medical centers with 850 and 500 staffed beds, respectively, and 3 regional or community hospitals with 100, 400, and 450 inpatient beds, respectively, within a large health care system in the Southeastern United States. Additionally, 2 of the hospitals held Magnet designation during the study period. The inclusion criteria for this study were registered nurses, doctors, physicians, and therapists working in ICUs in these hospitals, with experience in treating or caring for patients. Participants were required to be adults aged 18 to 80 years,

XSL•FO
RenderX

capable of understanding the study's purpose and risks, and able to provide informed consent. Additionally, participants needed to be available for a maximum 1-hour Zoom interview (Zoom Video Communications) and actively engaged in providing direct patient care in an ICU setting. Exclusion criteria included individuals with no experience in patient care or treatment, pregnant individuals, prisoners, and those who did not work directly in an ICU setting. These criteria were designed to ensure that participants had relevant clinical experience and could contribute meaningful insights into the study objectives.

Study End Points

Our primary end point was to explore the perceptions of ICU clinicians regarding the integration of robotic technology, focusing on identifying tasks that could reduce staff workload and enhance operational efficiency. The secondary end point was to characterize the ethical and practical considerations—such as privacy, safety, and clinical judgment—surrounding the implementation of robotics in critical care environments.

Sampling Method and Sample Size Calculation

Convenience sampling was used to recruit participants through institutional review board (IRB)–approved flyers. Given the exploratory and qualitative nature of this study, we applied a data saturation approach to inform the sample size. To align with best practices in qualitative research, we referenced the guidelines of Guest et al [21-23], which indicate that 80% to 92% of all concepts within a dataset are typically identified within the first 10 interviews, with 12 to 15 interviews generally sufficient to achieve thematic saturation. Based on this approach, we determined to enroll a minimum of 10 participants to begin the determination of data saturation.

Recruitment Process

For recruitment, a local site contact at each hospital initially posted the flyers with the sign-up QR code on the participating units. The site contacts also introduced the study in ICU huddles and meetings and emailed the recruitment flyers to unit staff on participating units. Once potential participants expressed interest, the research team emailed them directly to confirm their availability for an interview. Follow-up communication included proposed dates and times for the interview, with flexibility to accommodate participants' schedules. Site contacts repeated emails and huddle and meeting attendance until multiple professions were represented and data saturations were reached.

Data Collection

Data collection began in January 2023 and concluded in August 2023 upon achieving the planned 15 interviews. Each study participant provided informed consent to be interviewed and recorded prior to the interview sessions and could withdraw at any point prior to or during the interview. For consistency, 1 research team member (JS) conducted all interviews using a structured interview guide. In addition to asking the questions, the interviewer displayed the questions in Zoom. A detailed interview script is listed in Multimedia Appendix 1.

After the interview, participants filled out a Google Forms questionnaire (Multimedia Appendix 2), which was used to collect relevant information concerning participation demographics and experience while also assessing participant perceptions using a mix of rank order, Likert scale, and yes or no questions. This included reviewing a provided list of tasks derived from existing literature (Multimedia Appendix 3) [24], with the option to suggest additional tasks not listed. All interviews were recorded and transcribed into text files via Zoom. The 3 coders (JS, RIS, and DMR) listened to the recordings and reviewed the text files to correct any obvious transcription errors (eg, "I see you" to ICU).

Analysis

The 3 coders (JS, RIS, and DMR) applied directed content analysis to analyze the transcripts. In content analysis, coders systematically read, listen to, or view communications and assign codes to recurring patterns of communication [25]. In the first cycle of coding, the 3 coders (JS, RIS, and DMR) read each transcript independently to identify analysis categories. Initial categories were correlated to the directed questions in the interview script for descriptive, process, and value coding [26]. Each statement documented from a participant was a unit of analysis for the coders to assign to a category. If a single statement matched multiple categories such as a process and a value, it could be double-coded. Coding categories were finalized after 3 cycles, and the detailed primary codes were abstracted to secondary codes. Interviews and recruitment continued until no new secondary codes emerged. Near data saturation was observed after 11 participants, as comments began to repeat without introducing new insights. By participant 13, true data saturation was achieved, ensuring comprehensive coverage of the interview scope, which is reasonable given the narrow focus of the interview script.

To promote the trustworthiness of the analysis process, the 3 coders (JS, RIS, and DMR) debriefed after each coding cycle. The coders had different professional backgrounds: 2 (JS and RIS) were graduate students in engineering programs and 1 (DMR) was a registered nurse. Coders generated audit trail documents throughout all coding cycles and debrief sessions. All 3 coders (JS, RIS, and DMR) reviewed every unit of analysis. If there was disagreement or uncertainty among the coders about any participant statement, the coders reviewed the transcripts and relistened to the recordings until a consensus was reached. Once final coding was reached, a Python (Python Software Foundation) script was written to tally frequencies within each category.

Statistical Plan

To examine differences in the distribution of thematic codes across participant groups, a chi-square test of independence was conducted. A contingency table was constructed using the frequency of statements categorized under each code by the 3 participant groups. The test assessed whether the observed frequencies of codes were independent of participant roles, with a significance level set at α=.05. This analysis was used to determine if there were significant differences in how each group contributed to various themes, helping to identify potential biases or group-specific priorities in the data.

XSL•FO
RenderX

Ethical Considerations

The Emory University IRB approved this study (IRB STUDY00003902). Consent was obtained from participants both in signature form and prior to the interviews using an IRB-approved oral consent script. Interview recordings, transcripts, and participant data from the surveys were stored on a secured server in the Department of Biomedical Informatics. Participants were assigned ID codes, and transcripts were deidentified prior to analysis to keep responses anonymous. Participants will not be individually identified in any publications, presentations, or supplemental materials. Participants received a US $10 electronic Amazon gift card after each interview.

Results

Participants

After successfully concluding the recruitment phase, our study enlisted 15 participants, comprising 7 registered nurses, 6 physicians, and 2 advanced practice providers (APPs). The demographic profile of the study participants is summarized (Table 1). The sex distribution was predominantly female, with 11 (73%) participants, compared to 4 (27%) male participants. The average age of the participants was 35.9 (SD 8.5) years, and the ages ranged from younger physicians with an average age of 31.8 (SD 2.0) years to APPs with an average age of 42.0 (SD 5.7) years.

Experience in the ICU varied, with nurses and APPs having an average of 8.9 (SD 8.2) and 8.5 (SD 2.1) years, respectively, while physicians had a notably shorter average duration of 2.9 (SD 2.0) years. Exposure to health care robots ranged from 6 (40%) of the participants having encountered nonsurgical health care robots and 4 (27%) having experience with surgical robots.

Table 1. Demographic characteristics of 15 intensive care unit (ICU) clinicians from 6 hospitals in the Southeastern United States (January to August 2023), describing participant roles, age, sex, years of ICU experience, and prior exposure to robotic systems.

	Overall (N=15)	Nurse (n=7)	Physician (n=6)	APP[a] (n=2)
Sex, n (%)				
Female	11 (73)	7 (100)	3 (50)	1 (50)
Male	4 (27)	0 (0)	3 (50)	1 (50)
Age (years), mean (SD)	35.9 (8.5)	38.0 (11.7)	31.8 (2.0)	42.0 (5.7)
Working in the ICU (years), mean (SD)	6.4 (6.3)	8.9 (8.2)	2.9 (2.0)	8.5 (2.1)
Previous encounter with a robot in health care (except for surgical robot), n (%)	6 (40)	1 (14)	4 (67)	1 (50)
Previous encounter with a surgical robot, n (%)	4 (27)	2 (29)	2 (33)	0 (0)

[a]APP: advanced practice provider.

Coding Analysis

After the completion of the first and second cycles of coding analysis, primary codes and 7 secondary codes have been discerned. The secondary codes are administrative, comment or question about the interview, concerns, direct patient care, indirect patient care, minimum capacity, and perceptions (Table 2). Some statements were deemed not codable due to their ambiguous nature or because they were off topic.

XSL•FO
RenderX

Table 2. Secondary codes and associated primary codes derived from a directed content analysis of 15 intensive care unit clinicians showing the framework of perspectives on robotic integration.

Secondary codes	Primary codes
Administrative	• Administrative
Comment or questions about the interview	• Enquiry or comment about study
Concerns	• Cost • Human connection • Job insecurity • Lack of familiarity • Legal or HIPAA[a] • Privacy—patient • Privacy—clinician • Quality patient care • Safety—clinical judgment • Safety—malfunctioning • Safety—need evidence
Direct patient care	• Direct—any procedure • Direct—communicating • Direct—food • Direct—monitoring • Direct—performing procedure • Direct—supporting procedure • Direct—transporting • No physiotherapy • No ventilator
Indirect patient care	• Indirect—communicating • Indirect—custodial • Indirect—monitoring • Indirect—delivering supplies • Indirect—picking supplies • Indirect—supporting procedure • Indirect—transporting
Minimum capacity	• Audio • Device compatibility • Dexterity • Movement • Simultaneous localization and mapping • Storage • Touch screen • Video
Perceptions	• “Complex solution for a simple problem” • “Tasky things” • Reduce burden • Task delegation comfort differ by clinician • User perception—fearful (patients and clinicians) • User perception—need education (patients and clinicians) • User perception—negative (clinicians)

[a]HIPAA: Health Insurance Portability and Accountability Act.

1. Administrative: This category pertains to the logistical and management-related facets of incorporating robots into the ICU. It addresses the administrative duties and potential ramifications therein. Statements within this category, such as “They would function akin to ICU techs and secretaries,” exemplify the envisioned roles robots could undertake to support the administrative backbone of ICU operations.
2. Comment or question about the interview: This category captures any feedback or queries participants had regarding the interview process itself, helping to clarify the interview structure or content. Some statements in this category are “What exactly is your study looking at?” or “What kind of robots you are building?”
3. Concerns: Clinicians shared a broad range of apprehensions about introducing robots into the ICU environment. Their concerns span financial aspects, workforce implications, the impact on patient privacy, the quality of health care delivery, and the loss of “human touch.”

XSL•FO
RenderX

4. Direct patient care: This category refers to tasks and activities where robots would be involved in the immediate clinical management and treatment of patients. Clinicians showed particular interest in the potential of robotics for routine tasks and procedures that require precision and can be standardized, such as turning and repositioning the patient, assisting with physical therapy, assisting bedside procedures, or checking cognition.
5. Indirect patient care: The category includes activities where robots assist with nonclinical tasks that indirectly contribute to patient well-being and the efficient functioning of the ICU. Many participants recognized the utility of robots in roles such as transporting supplies, managing inventory, and performing cleaning tasks, which, while not directly related to patient care, are vital for maintaining a safe and operational environment.
6. Minimum capacity: In addressing the potential role of robots in the ICU, our interview protocol specifically included questions aimed at understanding the essential capabilities such technologies would need. This category collects the discussions focusing on the fundamental requirements that robots must satisfy to function effectively within the intensive care framework. Participants delineated baseline functionalities that robots should possess, such as the ability to navigate complex environments, respond to dynamic situations, and interface with existing hospital systems.
7. Perceptions: Clinicians' perceptions of robotics in the ICU were complex and dualistic. While there was excitement about the innovation and the potential efficiency gains, there was also skepticism about the practicality and readiness of such technologies for the sensitive ICU milieu. The prevailing belief among clinicians is that, while numerous tasks in the ICU could potentially be automated, the present state of robotic technology is best suited for delegating routine tasks to alleviate the heavy workload burden from the clinical staff.

Participant Insights and Perceptions

Guided by insights from clinicians, our coding analysis revealed a total of 78 tasks, where robotic systems might be introduced into ICU practice (Multimedia Appendix 4). Of these tasks, 50 (64%) were primarily related to direct patient care (eg, repositioning patients and assisting with simple procedures), whereas 19 (24%) focused on indirect patient care activities such as delivering supplies or cleaning. An additional 6 (8%) tasks addressed administrative roles (eg, handling patient call lights), and 3 (4%) were classified as mixed, encompassing both direct and indirect elements (eg, sitting with a patient to help keep them calm). These tasks are systematically categorized according to primary codes from our coding analysis. Recognizing the intricate nature of ICU operations, we acknowledge that some tasks may be multifaceted, overlapping across multiple categories. For clarity and coherence, tasks that serve similar purposes have been consolidated and are detailed in Multimedia Appendix 5.

A notable finding from our analysis is the significant number of tasks identified under the category of "direct—performing procedure," where 30 specific tasks were listed as shown in Multimedia Appendix 5. This category encapsulates tasks where robots could directly engage in patient care procedures, highlighting a substantial potential for robotic assistance in hands-on care activities. Additionally, the category "direct—supporting" includes 9 tasks, reflecting the supportive roles that robots could play in enhancing the efficiency and effectiveness of direct patient care.

The distribution of statements across various secondary codes, segmented by participant group (nurses, physicians, and APPs), is illustrated (Table 3). The table provides a quantitative overview of the engagement and contributions of each group in different thematic areas. The "nurse count," "physician count," and "APP count" columns represent the number of participants from each respective group who contributed to the discussions on each secondary code, while the "statements" columns reflect the total number of comments made by each group. In Figure 1, the stacked bar chart shows the proportion of statements contributed by nurses, physicians, and APPs for each primary code, providing a visual overview of how each group engaged with different thematic areas. Primary codes in "minimum capacity" were removed to maintain clarity and ensure emphasis on the most relevant thematic categories.

XSL•FO
RenderX

Table 3. Distribution of discussion topics across secondary codes among 15 intensive care unit clinicians.

Secondary codes	Participant count				Statements count			
	Total, n	Nurse, n (%)	Physician, n (%)	APP[a], n (%)	Total, n	Nurse, n (%)	Physician, n (%)	APP, n (%)
Administrative	10	5 (50)	4 (40)	1 (10)	16	10 (63)	5 (31)	1 (6)
Comment or questions about the interview	10	6 (60)	4 (40)	0 (0)	23	15 (65)	8 (35)	0 (0)
Concerns	15	7 (47)	6 (40)	2 (13)	109	49 (45)	48 (44)	12 (11)
Demographic	15	7 (47)	6 (40)	2 (13)	30	15 (50)	12 (40)	3 (10)
Direct patient care	14	6 (43)	6 (43)	2 (14)	125	64 (51.2)	48 (38.4)	13 (10.4)
Indirect patient care	14	7 (50)	5 (36)	2 (14)	71	38 (54)	22 (31)	11 (15)
Minimum capacity	15	7 (47)	6 (40)	2 (13)	61	26 (43)	25 (41)	10 (16)
Not codable	3	2 (67)	0 (0)	1 (33)	5	2 (40)	0 (0)	3 (60)
Perceptions	11	5 (45)	5 (45)	1 (10)	32	19 (59)	12 (38)	1 (3)

[a]APP: advanced practice provider.

Figure 1. Distribution of thematic contributions by clinician role in a qualitative intensive care unit robotics study. This stacked bar chart illustrates the percentage of statements contributed by 7 nurses (blue), 6 physicians (orange), and 2 advanced practice providers (green) across distinct primary codes derived from interviews at 5 Southeastern United States hospitals between January and August 2023. Technical codes are removed. Each code on the horizontal axis represents a thematic category identified through directed content analysis, and the vertical axis shows the proportion of total statements. APP: advanced practice provider; HIPAA: Health Insurance Portability and Accountability Act.

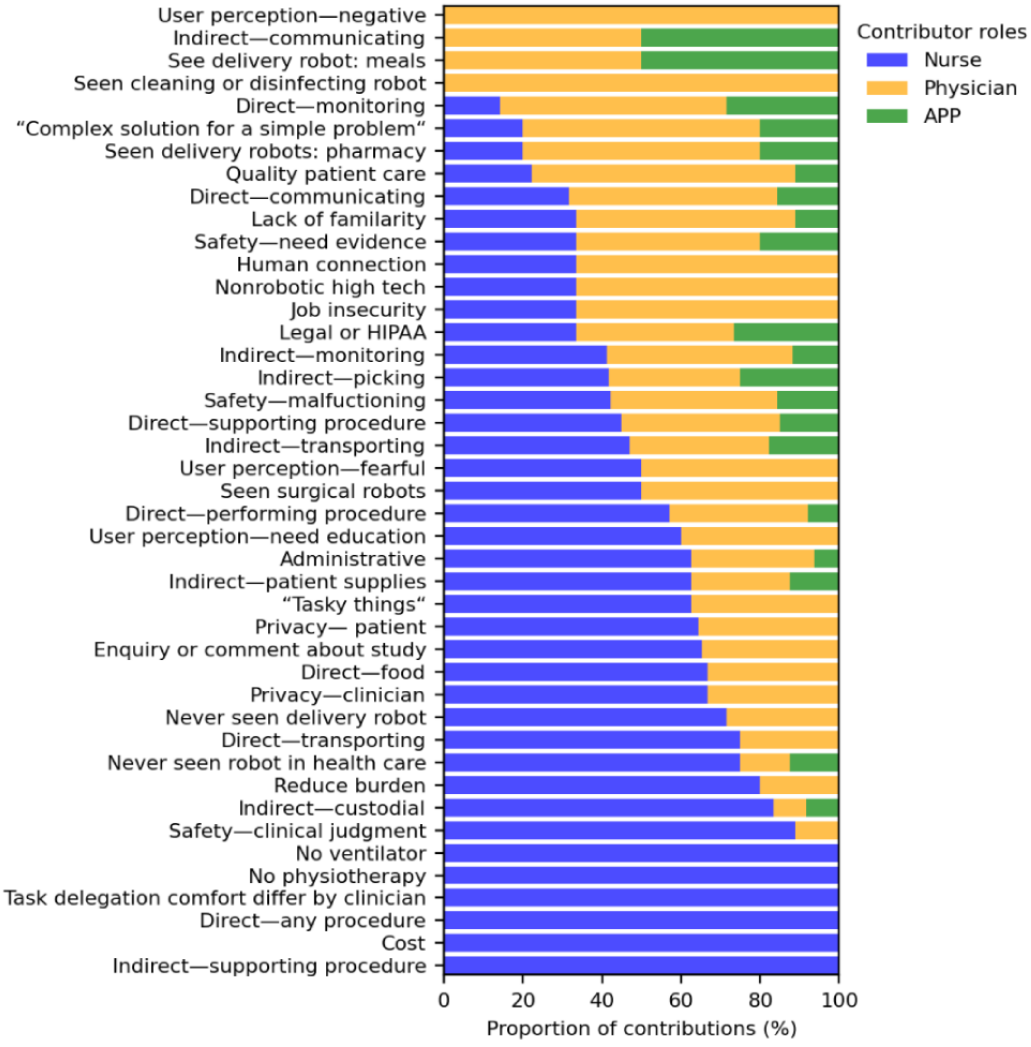

XSL•FO
RenderX

The “concerns” and “direct patient care” categories were the most frequently discussed topics, with nurses and physicians contributing the most statements, indicating shared apprehensions and interests. The “indirect patient care” and “minimum capacity” categories also received considerable attention from all participant groups, with a notably consistent number of statements from physicians and nurses. This suggests a common understanding of the importance of support tasks and the fundamental capabilities that are deemed necessary for robotic systems in the ICU.

Detailed distribution of statements across various primary codes is listed in Multimedia Appendix 6. The chi-square test of independence yielded a chi-square statistic of 112.87 with 102 degrees of freedom and a *P* value of .22. These results indicate no statistically significant differences in the distribution of thematic codes across nurses, physicians, and APPs, suggesting a relatively uniform contribution to the identified themes.

Participant Evaluation of Proposed Tasks

Following the completion of the interviews, clinicians were presented with our preprepared list from the literature of 19 health care tasks performed by robots. There was some overlap with the tasks discussed during the interviews as well as previously unmentioned tasks. Participants were asked to rank the top 7 tasks they believe could be performed by a robot in the ICU. They could also add tasks not on the preprepared list. This yielded a total of 22 tasks for consideration. A radar and heatmap charts visualizing the rankings (Figures 2 and 3) demonstrate the variance in priority assigned to each task by the different clinician groups. The tasks are denoted by short codes for brevity, with detailed descriptions provided in Multimedia Appendix 3. These 2 figures succinctly illustrate the areas where clinicians feel robotic assistance would be most beneficial in the ICU, reflecting their assessment of where robots could best support the clinical team and patient care.

Figure 2. Radar chart of clinician-prioritized robotic tasks in intensive care unit settings. This radar chart illustrates the frequency with which 15 intensive care unit clinicians ranked each task within their top 7. Tasks involve a range of direct, indirect, and administrative care activities, with * highlighting newly suggested tasks. IV: intravenous.

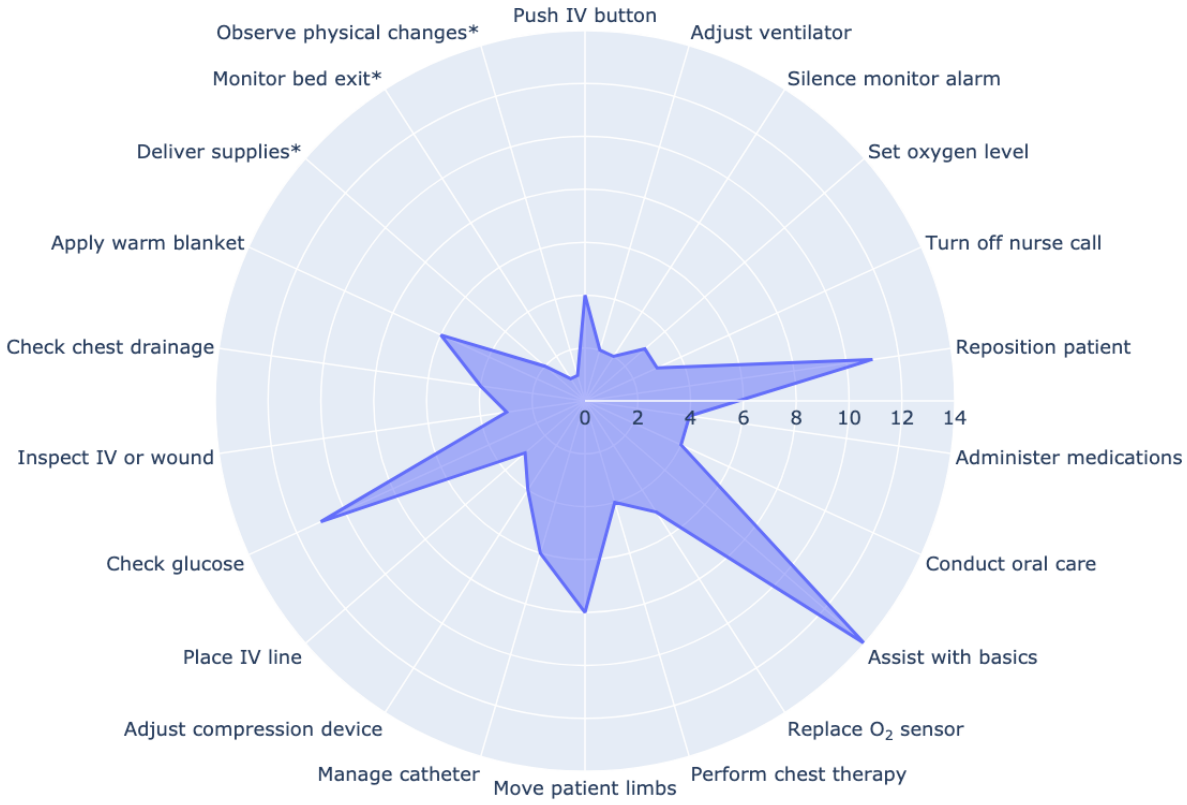

XSL•FO
RenderX

Figure 3. Heatmap of clinician-prioritized robotic tasks in intensive care unit settings. This heatmap shows task rankings by 15 intensive care unit clinicians (7 nurses, 6 physicians, and 2 advanced practice providers) from 5 Southeastern United States hospitals between January and August 2023. Darker cells indicate higher priority for robotic assistance, and tasks marked with * were newly proposed by participants. IV: intravenous.

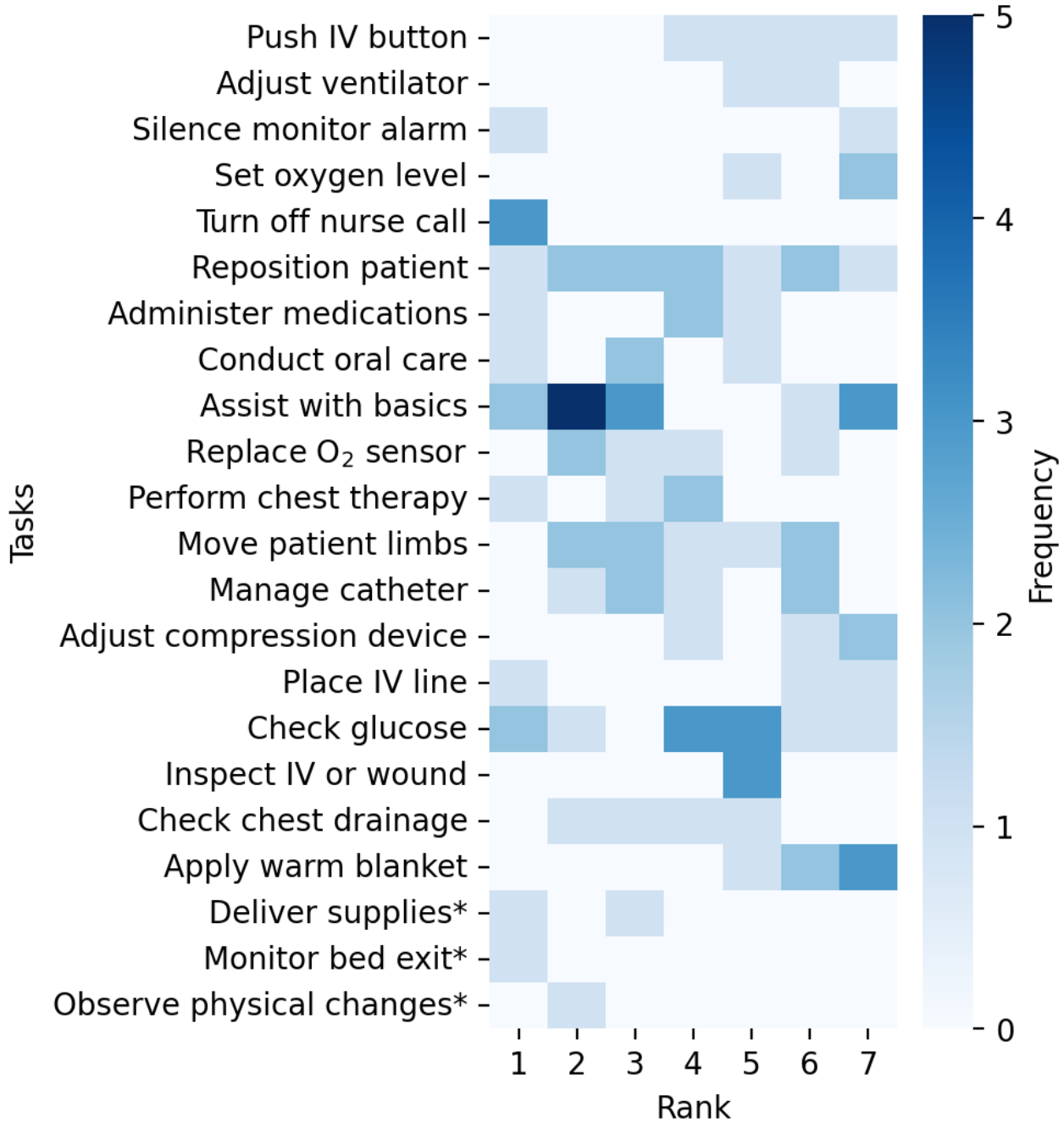

Interestingly, tasks perceived as less critical to direct clinical judgment were ranked higher. These tasks included controlling the nurse call button to mitigate interruptions, acting as a personal assistant for patient-related nonclinical tasks, and performing routine glucose checks. Conversely, tasks that require more nuanced clinical judgment, such as push button on the intravenous machine, adjusting ventilator, and adjust oxygen level from wall fixtures, while ranked, were not as highly prioritized, suggesting clinicians' preference for retaining direct control over complex decision-making processes.

The radar chart reveals a pattern of prioritization among clinicians, with tasks such as acting as the personal assistant, performing glucose checks, and repositioning or boosting a patient in bed consistently appearing as top-ranked tasks. This indicates clinicians' clear preference for assigning routine, recurring tasks requiring less clinical judgment to robotics. However, while these tasks are generally seen as routine, they may not always be suitable for autonomous robotic execution without human oversight. Tasks such as repositioning a patient can vary significantly in complexity depending on the patient's condition. For patients with chest tubes, drains, pressure wounds, or those on ventilators, such movements require heightened situational awareness and clinical judgment.

XSL•FO
RenderX

Survey Responses

Figure 4 presents box plots of responses collected from the survey. There is a similar level of median comfort for both movable and stationary robots, indicating no significant preference for robot mobility. Additionally, the survey responses indicate a division in opinion regarding the simultaneous operation of multiple robots in the ICU, with a majority of 9 (60%) participants comfortable with the idea. In contrast, the remaining 6 (40%) participants expressed discomfort with multiple robots performing tasks concurrently.

Figure 4. Box plot that compares clinicians' comfort levels with different types of robots in the intensive care unit. This box plot compares self-reported comfort levels (1-5 scale) among 15 intensive care unit clinicians (7 nurses, 6 physicians, and 2 advanced practice providers) across 4 categories of robots—movable, stationary, a robot expressing emotions, and a humanoid robot with emotions. Data were collected between January and August 2023 from 5 Southeastern US hospitals. The vertical axis indicates the degree of comfort, and the horizontal axis lists each robot type, revealing varying levels of acceptance for different robotic designs and capabilities in high-acuity health care environments.

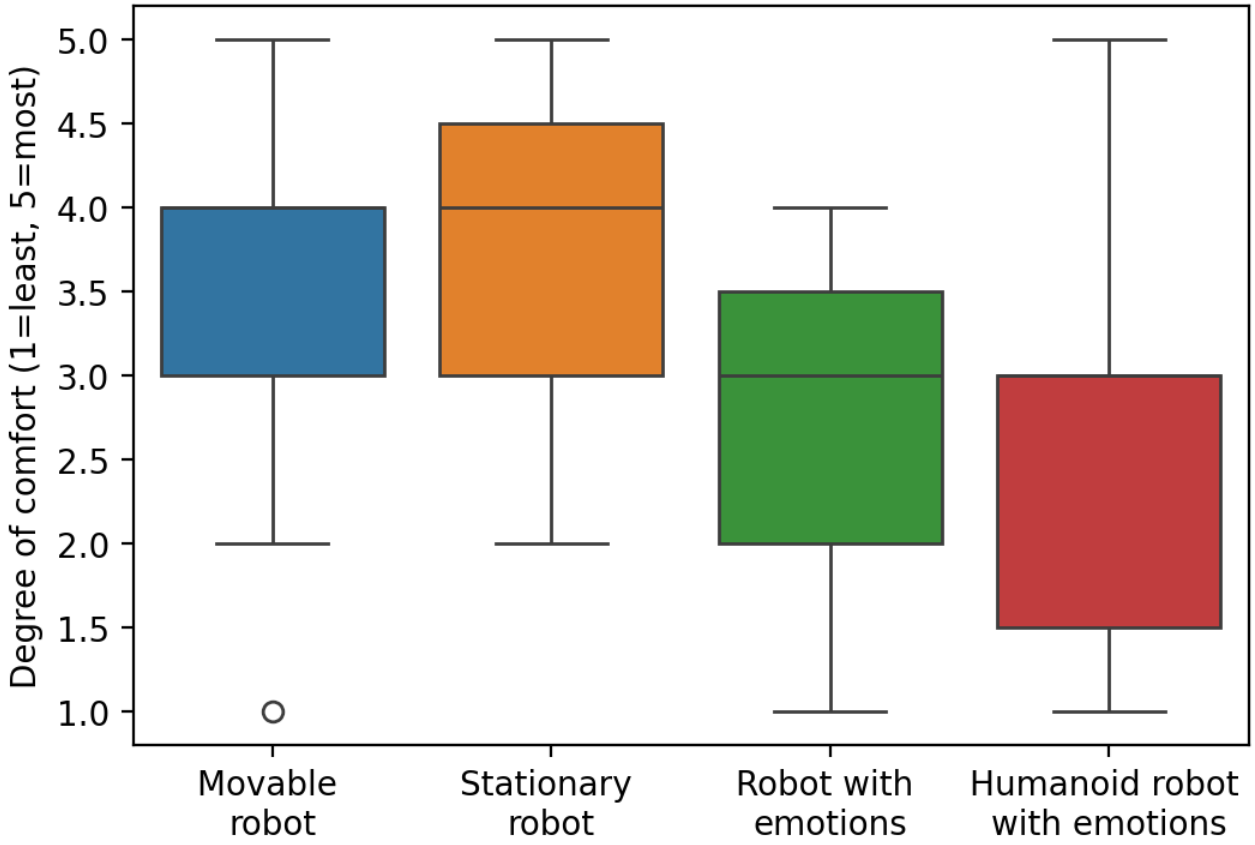

Robots that can express emotions and humanoid robots are met with a lower median comfort level compared to the other types, which may reflect clinicians' uncertainties or concerns about the unpredictability and implications of such features in a clinical setting.

Discussion

Principal Findings

Our study reveals a broad receptiveness among ICU clinicians to the integration of robotic technology, building on mounting evidence that such innovations can reduce staff burden and enhance care delivery in critical settings. While prior investigations have noted that telepresence and task-oriented robots can enhance workflow and reduce clinical workload, our qualitative analysis contributes a more nuanced view by detailing a diverse range of potential tasks—78 in total—where robots might assist without compromising patient safety or clinical judgment. This list offers valuable insights for future research and development, guiding the creation of technologies that align with the real-world needs and preferences of health care professionals.

Clinician Preferences for Robotic Assistance

Clinicians prioritized tasks for robotic assistance that are routine, recurrent, and do not necessarily require clinical judgment. High-frequency, low-complexity tasks such as responding to the nurse call button and performing glucose checks were favored for robotic automation. This suggests that clinicians are interested in leveraging robotic technology to reduce the burden of basic tasks, potentially increasing efficiency and allowing them to focus on more patient-centric care.

On the other hand, tasks requiring higher clinical judgment, such as adjusting ventilator settings, were not as highly prioritized for robotic assistance. An illustrative comment from a nurse highlighted this: "One of the things nurses always learn is that you're not treating the monitor. You're treating a patient so there's plenty of time that the monitor is wrong, and we have to use our nursing judgment." Although such tasks are recognized as potentially automatable, clinicians expressed a

XSL•FO
RenderX

need to maintain human oversight for these critical interventions, underlining an essential preference for direct human involvement in high-stakes decision-making and the delivery of compassionate care. This highlights a key insight that clinicians may prefer to retain control over critical tasks, possibly due to concerns about the consequences of errors and the current limitations of robotic technology in making nuanced decisions.

These insights have significant implications for the design and integration of robotic systems in the ICU. There is an opportunity for technology developers to focus on creating robotic systems that are seen as helpers rather than replacements for clinicians. Systems that can seamlessly integrate into the ICU environment and perform tasks that are perceived as helpful by clinicians are likely to be more readily accepted. Furthermore, the mixed comfort levels with humanoid robots and robots that express emotions suggest that while some clinicians might find these features enhance the patient experience, others might view them as unnecessary or potentially unsettling. This indicates a need for careful consideration of the human-robot interaction design and for the provision of adequate training and familiarization for clinicians.

Ethical and Safety Considerations

Indirect references to the foundational biomedical ethical principles of autonomy, beneficence, nonmaleficence, and justice were reflected in many of the views shared by participants. With respect to autonomy, beneficence, and nonmaleficence, 10 participants explicitly stated that robots in patient rooms must comply with HIPAA (Health Insurance Portability and Accountability Act) protections regarding the privacy and security of patients' individually identifiable personal health information. This is a requirement throughout the health care setting regardless of the introduction of robots. However, 5 participants believed that robots with audiovisual features would pose increased risks for breaches of patient privacy and confidentiality. In contrast, 3 participants noted that while patient personal health information must be protected, patients currently have decreased personal privacy in ICU by design (eg, glass doors and walls) to help clinicians monitor them for safety. Two participants asserted that monitoring for safety (regardless of machine or human) is reasonable if it improves response times and quality of patient care, provided that data are kept secure. One participant was concerned with the possibility of recordings "being used against staff." Three other participants noted that if clinicians are practicing within scope and protocol, they should not worry about being recorded by the hospital. Two other participant comments around privacy and confidentiality were the use of robots in direct patient care should be stated on the consent to receive treatment form, and there should be indicators (eg, a light) when robots or other devices are using audiovisual features.

Specific to beneficence and nonmaleficence, all participants expressed concern about patient safety to varying degrees if robots perform direct patient care. Three participants described how ICU nurses use clinical judgment when treating patients who are critically ill. They described this as "treating the patient" and not responding to clinical device readings in isolation. They did not believe robots would be capable of applying this often experiential clinical decision-making. However, 3 other participants noted that errors and injuries currently occur in hospitals. Akin to safety concerns, 2 nurses explicitly stated that they would want 24×7 support in troubleshooting any robot malfunctions.

Related to justice, one participant noted their belief, "patients who might have limited health literacy or are older ... might be more likely to go to an ICU ... more likely to have complications for all sorts of reasons ... the people who would benefit most from technology tend to already have high health literacy [which leads to the concern of social and identity] inequity." Further patients with a lower understanding about technology or who are in a compromised psychological state such as delirium might be frightened or confused by the presence of robots.

The focus on patient and clinician privacy and safety as well as the concern about missing the "human touch" underscore the need for patient, family, and staff education when deploying robots into the clinical setting to address these concerns. Resources or support would be needed for clinicians to detect and possibly troubleshoot issues with the robots. Three participants noted that their confidence in robots performing direct patient care would increase when they observed this occurring in their clinical setting without any complications. One further commented that robot functions in the ICU should be limited to functions posing less risk to patients and with patients themselves who are less medically fragile. Once the use of robots in ICUs becomes more proven and standardized, then functionality can be increased.

An ethics-related concern emerged about the loss of human connection when robots are included in direct care, regardless of whether process efficiencies are gained. Six clinicians made statements such as patients in ICU are "already prone to feeling isolation ... feeling like there wasn't a person in there with them enough." Would patients miss the human touch or feel "less than" with a robot caregiver? What impact would this have on patients and clinicians? A patient-centered approach to incorporating robots into the ICU setting would be to identify tasks, interactions, and possibly specific patients requiring the expression of human compassion and caring and preserve the human connection in those contexts [27,28].

Strengths and Limitations

This study has several limitations. In addition to the small sample size, which restricts generalizability, there is also a possibility of selection bias, as clinicians who chose to participate may have had greater interest in or familiarity with robotics. Moreover, participants came from a limited geographic region, which may not reflect the diversity of ICU settings across different health care systems and patient populations. Since this study is cross-sectional, it only reflects clinicians' perceptions at a single point in time, without accounting for how attitudes toward robotics may evolve as technology advances. Future investigations should include larger, more diverse samples from multiple institutions and regions, allowing researchers to draw broader conclusions. It would also be beneficial to conduct longitudinal studies to observe how the comfort levels and task preferences of clinicians shift over time with increasing exposure to robotic systems. Finally, collecting input from patients and

XSL•FO
RenderX

their representatives is critical to fully understanding the ethical and practical ramifications of integrating robotics into patient care.

Conclusions

This study aimed to explore the perceptions of ICU clinicians regarding the integration of robotics in critical care, focusing on identifying tasks where robots can effectively reduce clinician workload. Our findings confirm that clinicians view robotics as a supportive tool rather than a replacement for human expertise. By aiding with routine and physically demanding tasks, robots can free providers and nurses to focus on more complex, human-centric aspects of patient care. However, the ethical dimension—encompassing privacy, patient autonomy, and the preservation of the clinician-patient relationship—remains paramount. Ensuring robust oversight of robotics in the ICU, coupled with appropriate safeguards for patient data and clinician well-being, is essential for successful and responsible adoption.

Acknowledgments

This work is funded by the National Institutes of Health (award R01GM139967). The authors thank the study participants and hospital site principal investigators for their time commitment to participating in the study. During the preparation of this work, the authors used ChatGPT-4o to enhance language and readability. All content generated by this tool was subsequently reviewed and edited by the authors, who take full responsibility for the final version of the publication.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional privacy regulations and the sensitive nature of the data but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview script for intensive care unit clinicians.
[DOCX File , 26 KB - jmir_v27i1e62957_app1.docx]

Multimedia Appendix 2
Google Forms questionnaire.
[DOCX File , 27 KB - jmir_v27i1e62957_app2.docx]

Multimedia Appendix 3
Collected list of robotic tasks from literature.
[DOCX File , 28 KB - jmir_v27i1e62957_app3.docx]

Multimedia Appendix 4
Unique tasks for robotic assistance in intensive care unit.
[DOCX File , 32 KB - jmir_v27i1e62957_app4.docx]

Multimedia Appendix 5
Categorized tasks for robotic assistance in intensive care unit.
[DOCX File , 31 KB - jmir_v27i1e62957_app5.docx]

Multimedia Appendix 6
Distribution of discussion topics among health care providers.
[DOCX File , 47 KB - jmir_v27i1e62957_app6.docx]

References

1. Teng R, Ding Y, See KC. Use of robots in critical care: systematic review. J Med Internet Res 2022;24(5):e33380. [doi: 10.2196/33380] [Medline: 35576567]
2. Alsohime F, Temsah MH, Al-Eyadhy A, Ghulman S, Mosleh H, Alsohime O. Technical aspects of intensive care unit management: a single-center experience at a tertiary academic hospital. J Multidiscip Healthc 2021;14:869-875 [FREE Full text] [doi: 10.2147/JMDH.S294905] [Medline: 33907413]

XSL•FO
RenderX

3. Alameddine M, Dainty KN, Deber R, Sibbald WJ. The intensive care unit work environment: current challenges and recommendations for the future. J Crit Care 2009;24(2):243-248 [FREE Full text] [doi: 10.1016/j.jcrc.2008.03.038] [Medline: 19327295]
4. Seymour T, Frantsvog D, Graeber T. Electronic health records (EHR). Am J Health Sci 2012;3(3):201-210. [doi: 10.19030/ajhs.v3i3.7139]
5. Song J, McNeany J, Wang Y, Daley T, Stecenko A, Kamaleswaran R. Riemannian manifold-based geometric clustering of continuous glucose monitoring to improve personalized diabetes management. Comput Biol Med 2024;183:109255. [doi: 10.1016/j.compbiomed.2024.109255] [Medline: 39405732]
6. Song J, Daley T, McNeany J, Kamaleswaran R, Stecenko A. 682 A machine learning approach with silhouette scoring of continuous glucose monitoring enables repeat measure assessment of changes in the glycemic profile in cystic fibrosis. J Cyst Fibros 2024;23:S381-S382. [doi: 10.1016/s1569-1993(24)01520-0]
7. Song J, Daley T, Alvarez J, Harris R, McNeany J, Wang Y, et al. 566 Machine learning analysis of continuous glucose monitoring identifies greater degree of dysglycemia than previously suggested byoral glucose tolerance testing. J Cyst Fibros 2023;22:S298-S299. [doi: 10.1016/s1569-1993(23)01488-1]
8. Nuckols TK, Bower AG, Paddock SM, Hilborne LH, Wallace P, Rothschild JM, et al. Programmable infusion pumps in ICUs: an analysis of corresponding adverse drug events. J Gen Intern Med 2008;23(Suppl 1):41-45 [FREE Full text] [doi: 10.1007/s11606-007-0414-y] [Medline: 18095043]
9. Becevic M, Clarke MA, Alnijoumi MM, Sohal HS, Boren SA, Kim MS, et al. Robotic telepresence in a medical intensive care unit—clinicians' perceptions. Perspect Health Inf Manag 2015;12(Summer):1c [FREE Full text] [Medline: 26396554]
10. Vagvolgyi BP, Khrenov M, Cope J, Deguet A, Kazanzides P, Manzoor S, et al. Telerobotic operation of intensive care unit ventilators. Front Robot AI 2021;8:612964 [FREE Full text] [doi: 10.3389/frobt.2021.612964] [Medline: 34250025]
11. Nieto Agraz C, Pfingsthorn M, Gliesche P, Eichelberg M, Hein A. A survey of robotic systems for nursing care. Front Robot AI 2022;9:832248 [FREE Full text] [doi: 10.3389/frobt.2022.832248] [Medline: 35462781]
12. Costa A, Martinez-Martin E, Cazorla M, Julian V. PHAROS—PHysical Assistant RObot System. Sensors (Basel) 2018;18(8):2633. [doi: 10.3390/s18082633] [Medline: 30103492]
13. Eiammanussakul T, Sangveraphunsiri V. A lower limb rehabilitation robot in sitting position with a review of training activities. J Healthc Eng 2018;2018:e1927807 [FREE Full text] [doi: 10.1155/2018/1927807] [Medline: 29808109]
14. Lo HS, Xie SQ. Exoskeleton robots for upper-limb rehabilitation: state of the art and future prospects. Med Eng Phys 2012;34(3):261-268. [doi: 10.1016/j.medengphy.2011.10.004]
15. Hobbs B, Artemiadis P. A review of robot-assisted lower-limb stroke therapy: unexplored paths and future directions in gait rehabilitation. Front Neurorobot 2020;14:19 [FREE Full text] [doi: 10.3389/fnbot.2020.00019] [Medline: 32351377]
16. Song J, Hardin EC. Monitoring walking asymmetries and endpoint control in persons living with chronic stroke: implications for remote diagnosis and telerehabilitation. Digit Health 2024;10:20552076231220450 [FREE Full text] [doi: 10.1177/20552076231220450] [Medline: 38188863]
17. Miseikis J, Caroni P, Duchamp P, Gasser A, Marko R, Miseikiene N, et al. Lio—a personal robot assistant for human-robot interaction and care applications. IEEE Robot Autom Lett 2020;5(4):5339-5346 [FREE Full text] [doi: 10.1109/LRA.2020.3007462] [Medline: 34192136]
18. Fischinger D, Einramhof P, Papoutsakis K, Wohlkinger W, Mayer P, Panek P, et al. Hobbit, a care robot supporting independent living at home: first prototype and lessons learned. Robot Auton Syst 2016;75:60-78. [doi: 10.1016/j.robot.2014.09.029]
19. Goldau FF, Shastha TK, Kyrarini M, Graser A. Autonomous multi-sensory robotic assistant for a drinking task. 2019 Presented at: 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR); June 24-28, 2019; Toronto, Ontario, Canada. [doi: 10.1109/icorr.2019.8779521]
20. Sale P, Franceschini M, Waldner A, Hesse S. Use of the robot assisted gait therapy in rehabilitation of patients with stroke and spinal cord injury. Eur J Phys Rehabil Med 2012;48(1):111-121. [Medline: 22543557]
21. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
22. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82 [FREE Full text] [doi: 10.1177/1525822x05279903]
23. Morgan MG. Risk Communication: A Mental Models Approach. Cambridge: Cambridge University Press; 2002.
24. Freeman WD, Sanghavi DK, Sarab MS, Kindred MS, Dieck EM, Brown SM, et al. Robotics in simulated COVID-19 patient room for health care worker effector tasks: preliminary, feasibility experiments. Mayo Clin Proc Innov Qual Outcomes 2021;5(1):161-170 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.12.005] [Medline: 33521585]
25. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
26. Miles MB, Huberman AM, Saldana J. Qualitative Data Analysis: A Methods Sourcebook. Washington, DC: SAGE Publications; 2013.

XSL•FO
RenderX

27. Hernandez JPT. Compassionate care with autonomous AI humanoid robots in future healthcare delivery: a multisensory simulation of next-generation models. Biomimetics (Basel) 2024;9(11):687 [FREE Full text] [doi: 10.3390/biomimetics9110687] [Medline: 39590259]
28. Morrow E, Zidaru T, Ross F, Mason C, Patel KD, Ream M, et al. Artificial intelligence technologies and compassion in healthcare: a systematic scoping review. Front Psychol 2022;13:971044 [FREE Full text] [doi: 10.3389/fpsyg.2022.971044] [Medline: 36733854]

Abbreviations

APP: advanced practice provider
HIPAA: Health Insurance Portability and Accountability Act
ICU: intensive care unit
IRB: institutional review board

Edited by A Mavragani; submitted 05.06.24; peer-reviewed by L Hung, E Bignami; comments to author 12.12.24; revised version received 02.01.25; accepted 07.01.25; published 28.03.25.

Please cite as:
Song J, Sridhar RI, Rogers DM, Hiddleson C, Davis C, Holden TL, Ramsey-Haynes S, Reif L, Swann J, Jabaley CS, Gullatte M, Kamaleswaran R
Clinicians' Perceptions and Potential Applications of Robotics for Task Automation in Critical Care: Qualitative Study
J Med Internet Res 2025;27:e62957
URL: https://www.jmir.org/2025/1/e62957
doi:10.2196/62957
PMID:

©Jiafeng Song, Rishika Iytha Sridhar, Darlene Marie Rogers, Cheryl Hiddleson, Carolyn Davis, Tina Lynn Holden, Shanna Ramsey-Haynes, Lisa Reif, Julie Swann, Craig S Jabaley, Mary Gullatte, Rishikesan Kamaleswaran. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of the Smarter Safer Homes Solution on Quality of Life and Health Outcomes in Older People Living in Their Own Homes: Randomized Controlled Trial

Wei Lu[1], PhD; David Silvera-Tawil[1], PhD; Hwan-Jin Yoon[1], PhD; Liesel Higgins[1], BMUS, MBA; Qing Zhang[2], PhD; Mohanraj Karunanithi[3], PhD; Julia Bomke[1], BSN; Joshua Byrnes[4], PhD; Jennifer Hewitt[5], PhD; Vanessa Smallbon[1], MEng; Jill Freyne[1], PhD; Deepa Prabhu[1], PhD; Marlien Varnfield[1], PhD

[1]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, QLD, Australia

[2]Research Center for Frontier Fundamental Studies, Zhejiang Lab, Hangzhou, ZheJiang, China

[3]Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia

[4]Centre for Applied Health Economics, Griffith University, Brisbane, QLD, Australia

[5]Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

Corresponding Author:
Marlien Varnfield, PhD
Australian e-Health Research Centre
Commonwealth Scientific and Industrial Research Organisation
Level 7, Surgical Treatment and Rehabilitation Services, 296 Herston Road
Brisbane, QLD, 4029
Australia
Phone: 61 732533603
Email: marlien.varnfield@csiro.au

Abstract

Background: An increasingly aging population, accompanied by a shortage of residential aged care homes and workforce and consumer feedback, has driven a growing interest in enabling older people to age in place through home-based care. In this context, smart home technologies for remote health monitoring have gained popularity for supporting older people to live in their own homes.

Objective: This study aims to investigate the impact of smart home monitoring on multiple outcomes, including quality of life, activities of daily living, and depressive symptoms among older people living in their own homes over a 12-month period.

Methods: We conducted an open-label, parallel-group randomized controlled trial. The control group continued to receive their existing care from aged care service providers. Meanwhile, the intervention group, in addition to receiving their usual aged care services, had their activities of daily living monitored using a smart home platform. Surveys including the Adult Social Care Outcomes Toolkit (ASCOT), EuroQol-5 Dimensions-5 Levels (EQ-5D-5L), Katz Index of Independence in Activities of Daily Living (Katz ADL), Lawton Instrumental Activities of Daily Living Scale (IADL), and Geriatric Depression Scale (GDS) were conducted at baseline and 6 and 12 months from baseline. Linear mixed-effects models were used to compare the difference between the intervention and control groups, with the ASCOT as the primary outcome measure.

Results: Data from 130 participants were used in the analysis, with no significant differences in baseline characteristics between the control group (n=61) and the intervention group (n=69). In comparison to the control group, the intervention group had a higher ASCOT score at the 6-month assessment (mean difference 0.045, 95% CI 0.001 to 0.089; Cohen *d*=0.377). However, this difference did not persist at the 12-month assessment (mean difference 0.031, 95% CI –0.014 to 0.076; Cohen *d*=0.259). There were no significant differences in EQ-5D-5L, Katz ADL, IADL, and GDS observed between the intervention and control groups at the 6-month and 12-month assessments.

Conclusions: The study demonstrates that smart home monitoring can improve social care–related quality of life for older people living in their own homes. However, the improvement was not sustained over the long term. The lack of statistically significant findings and diminished long-term improvements may be attributed to the influence of the COVID-19 pandemic during the later stage of the trial. Further research with a larger sample size is needed to evaluate the effect of smart home monitoring on broader quality-of-life measures.

XSL•FO
RenderX

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618000829213; https://tinyurl.com/2n6a75em
International Registered Report Identifier (IRRID): RR2-10.2196/31970

(J Med Internet Res 2025;27:e59921) doi:10.2196/59921

KEYWORDS

randomized controlled trial; digital health; eHealth; smart home; sensor; health monitoring; home monitoring; aged care; aging in place; older adult; quality of life

Introduction

Globally, the aging population is steadily increasing, with 703 million individuals aged 65 years or over as of 2019, and this number is projected to reach 1.5 billion by the year 2050, with a global average life expectancy of around 77.1 years [1]. By 2041, the population of Australia aged 65 years and older is expected to see a 54% increase, rising from 4.31 million in 2021 to 6.66 million [2]. This demographic trend presents unique challenges in providing adequate care and support for older individuals, particularly those who choose to age in their own homes. Residential aged care facilities have traditionally been the primary setting for delivering aged care services. However, an aging population, coupled with a global shortage of aged care facilities and workforce, poses significant challenges to residential aged care. This leads to compromised access to essential care services, overburdened existing facilities, and increased strain on caregivers, impacting the quality of care [3-5]. To address these challenges, there has been a growing policy emphasis on aging in place, the concept of older people living longer in their own homes and communities as they age [6,7]. This allows older people to remain connected to their communities, friends, and family networks, promoting overall well-being and social support. Aside from older people's preference, the lack of residential care placements and the rising costs associated with institutional care make aging in place a more feasible and cost-effective option for many individuals.

In this context, smart home technologies have emerged as a promising solution to support the well-being and independence of older individuals living in their own homes. These technologies encompass a wide range of devices and systems, including sensor networks, remote monitoring, and assistive devices, which can enhance safety, health monitoring, and overall quality of life for older adults [8,9]. By integrating these technologies into the home environment, smart home solutions have the potential to address various challenges associated with aging in place, such as falls, medication adherence, and social isolation [10].

While the potential benefits of smart home solutions in supporting aging in place have been widely acknowledged, the evidence base, particularly from randomized controlled trials (RCTs), remains limited [11,12]. The evaluation of smart home technologies through RCTs is important to establish their efficacy, effectiveness, and cost-effectiveness in real-world settings. Moreover, previous studies investigating the influence of smart home technologies on the well-being of older individuals largely focused on health-related quality of life [13-16]. Smart home technologies can detect emergencies, monitor health, and assist in daily tasks, which can lead to increased autonomy and security, not only promoting physical well-being but also addressing critical social and emotional needs [11,17]. However, an increased emphasis on health-related quality of life and less attention to the social care–related quality of life has created a gap in our understanding of how smart home technologies can affect broader aspects of older people's quality of life, beyond health functioning [18-20].

Following our previously published protocol [21], we conducted an RCT, with the primary aim of evaluating how the use of the Smarter Safer Home (SSH) platform impacts the social care–related quality of life in older people living in their own homes. As a secondary aim, we examined whether this technology impacts the health-related quality of life, ability to perform activities of daily living (ADLs), and severity of depression. The findings of this study enhance our understanding of how smart home technologies can impact older people who live in their own homes. Moreover, they have the potential to inform policy and practice in the field of aged care, providing valuable insights into the effectiveness and feasibility of implementing smart home solutions to support aging in place.

Methods

Study Design

This study is a mixed-methods, open-label, parallel-group RCT. Participants were allocated randomly to a control group and an intervention group. The control group received regular aged care services subsidized by the Australian government, which can be categorized into 2 types: the Commonwealth Home Support Programme (CHSP) and the Home Care Package (HCP). CHSP provides basic support to help older Australians live independently, while HCP offers more comprehensive and personalized care for those with higher or more complex needs. The intervention group, in addition to usual aged care services, had the SSH platform installed at their residence. The SSH platform is an in-home monitoring system that uses distributed ambient sensors to track the physical environment and human activities within a household. The sensor data is transmitted to a cloud computing server that assesses residents' ADLs and sleep patterns. During the trial, the assessment was accessible to participants, their families, and aged care service providers through a tablet app or web portals. Survey questionnaires completed by the participants were used to measure primary and secondary outcomes at baseline, 6 months, and 12 months. Additionally, the trial collected various other types of data, including raw sensor data, evaluations from both informal and formal carers, participants' health services utilization, and log spreadsheets kept by aged care service providers detailing their usage of the SSH system [21]. This paper is specifically focused

XSL•FO
RenderX

on the survey questionnaires completed by the participants. Analyses of the additional data types collected during this RCT will be presented in separate publications.

Ethical Considerations

Ethics approval for this study was obtained from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Human and Medical Research Ethics Committee (CHMREC; HREC 4/2018). The trial was registered at the Australian New Zealand Clinical Trials Registry (ACTRN12618000829213). All participants consented to join the study. They are free to withdraw or suspend their participation at any time with no need to provide a reason.

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) older adults (65 years and older), (2) living at home, in the care of a designated aged care service provider, and (3) English-speaking with proficiency in written English. The exclusion criteria were as follows: (1) residing in long-term residential care, (2) unable to give informed consent due to reasons such as severe cognitive impairment, and (3) unwilling to leave their electricity on overnight.

SSH Platform

This study used the SSH platform presented in the protocol paper [21]. The SSH platform included an in-home network of distributed sensors, a cloud server, and a client module with a tablet app, a family web portal, and a service provider web portal (Figure 1). The sensor network consisted of ambient sensors that measured the physical environment (eg, room temperature, humidity) as well as human activities (eg, motion, door opening or closing) within the home. The sensors wirelessly sent data to a gateway hub, which relayed the data to a secure cloud server through a local WiFi network. A cloud computing service was implemented to process the sensor data, assessing the resident's ADLs across 5 domains, including mobility, transfer, hygiene, dressing, and meal preparation. The ADL scores were then compared with the participant's individual baseline profile established with the data collected over the initial 21 days of the trial period. Deviations from the baseline in each ADL domain were categorized as expected, unexpected, or very unexpected. Apart from ADL performance, the resident's sleep measures were also collected with sleep sensors placed under the bed mattress. The analytic results were made accessible to the participants, their families, and associated aged care service providers through a tablet app, a family web portal, and a service provider web portal, respectively.

The setting was home-based, with participants living in their own residences in the community. Participants were recruited from the customer bases of the aged care service providers in metropolitan and regional areas in Queensland, Australia, as defined by the Rural Remote Area and Metropolitan Area classification criteria [22].

Figure 1. The user interface of (A) the SSH tablet app and (B) the SSH web portal. SSH: Smarter Safer Home.



XSL•FO
RenderX

Trial Setting

The RCT was conducted between April 2019 and November 2020, in partnership with 3 aged care service providers (Textbox 1).

Textbox 1. Overview of aged care service providers involved in the study.

- Anglicare Southern Queensland is a member of the Anglicare Australia Network providing support to aged Australians in partnership with government and other support organizations, in response to identified care needs throughout South East Queensland.
- integratedLiving Australia receives funding from the Australian and state governments to provide care services. They provide in-home support services to older people in regional, rural, and remote Australia (including North and Eastern Queensland) and have been focusing on ensuring equitable access for health support to these communities.
- All About Living partners with federal, state, and local government departments and a number of community organizations to deliver a range of high-quality services. All About Living has developed governance, management, and service delivery expertise and excellence that enable consistent service delivery.

Recruitment

The aged care service providers identified eligible clients from their services and provided details to a team of project officers independent from the aged care service providers, who recruited and consented participants to the trial. The project officers contacted potential participants after an introduction by the aged care service provider. The project officer, via phone call, briefed the individual on the trial and assessed their interest in participating. If the individual agreed to participate, they were asked to nominate their informal carer to act as a witness in a future face-to-face meeting, where the project officer confirmed that the prospective participants met the inclusion criteria. Potential participants were provided with the Participant Information and Consent Form and a verbal description of the trial requirements by the project officer. Individuals who agreed to participate were asked to sign the Participant Information and Consent Form.

Procedures

Once participants had been consented, baseline surveys were administered to them by the project officer. After completing the baseline surveys, participants were allocated randomly to either the intervention or control groups. Stratified randomization was used with metropolitan and regional strata to preserve the ratio of the 2 geographical areas in both intervention and control groups. An independent researcher conducted the randomization using a computer program and anonymized identifiers. All participants (intervention and control) continued to receive their existing care and social services in line with local aged care service provider protocols for the 12 months of the trial. In addition, participants allocated to the intervention group received an SSH kit, including sensors, a tablet, and a gateway hub. The installation of the SSH kit was coordinated by the project officer, typically within 2 weeks of group allocation. All participants were contacted by a project officer to complete the follow-up surveys at 6 months and 12 months from baseline. The surveys were conducted either over the phone or in person and were managed using the REDCap (Vanderbilt University) hosted by CSIRO. During the trial period, the project officer was the point of contact for any problems or questions. If an intervention group participant wished to withdraw from the trial, the project officer arranged for the SSH kit to be uninstalled from their home as quickly as possible. Upon completion of the project, all homes had their SSH kits removed by the project officer.

Intervention

The intervention involved monitoring ADLs of participants through the SSH platform. During the trial, the aged care service providers aimed to review the trends in the participants' ADL performance through the SSH service provider portal 5 days a week, specifically on business days. When the aged care service providers identified a participant's ADL performance to be in an unexpected or very unexpected range for 2 consecutive days, they would contact the participant via phone or send an email to the participant's case manager. The aged care service providers created their own intervention logs to record information about actions taken for the participants. Given the workload on aged care service providers, maintaining the intervention log was not a compulsory task to form a complete record of the interventions. However, it offered some insights into how aged care service providers used the SSH platform to trigger interventions. Although no specific tasks were mandated, participants and their families were encouraged to monitor the participants' ADL performance and take any actions they deemed appropriate.

Outcome Measures

The primary outcome measure was social care–related quality of life, measured with a self-completion version of the Adult Social Care Outcomes Toolkit (ASCOT [23]). Secondary outcomes included EuroQol-5 Dimensions-5 Levels (EQ-5D-5L [24,25]), Katz Index of Independence in Activities of Daily Living (Katz ADL [26]), Lawton Instrumental Activities of Daily Living Scale (IADL [27]), and Geriatric Depression Scale (GDS [28]).

Sample Size Calculation

The power calculation for the randomized trial design was based on a clinically significant difference of 10% in ASCOT, with a distribution of 0.80 (SD 0.16) as indicated in prior research [29]. Given a null hypothesis of $\alpha \geq .05$, an effect size of 0.5 (0.08/0.16), identical allocation ratio for the 2 groups in our trial, and a 30% attrition rate, the sample size of this trial was computed to be 100 participants per group to achieve 80% statistical power [21].

XSL•FO
RenderX

Statistical Analysis

Descriptive statistics summarized the baseline characteristics and outcome measures by groups. For continuous variables and categorical variables, means and SD or numbers and percentages were reported, respectively. The primary and secondary outcome measures were analyzed using a linear mixed-effects model to compare the difference between the intervention and control groups. Fixed effects included group allocation, time (6 months and 12 months from baseline), group-by-time interaction, and baseline scores. Additionally, the analysis considered Participant ID as a random effect to account for within-individual correlations. Cohen *d* was used to describe the standardized mean difference of an effect. We conducted sensitivity analyses to examine whether the impact of the SSH platform varied across subgroups defined by baseline demographics, such as gender, care package, and living status. Marriage status and aged care service provider were not included in the sensitivity analyses due to insufficient sample sizes in certain subcategories.

Data Exclusion

COVID-19 Impact

In 2020, while the trial was ongoing, Australia was struck by the COVID-19 pandemic, presenting the country's older community with unexpected and significant challenges. Older adults faced considerable health risks due to their vulnerability to severe illness from COVID-19. Additionally, public health restrictions such as lockdowns and social distancing imposed by the Australian government could have both direct and indirect effects on the physical and mental health of older adults [30-32]. While quantifying the exact impact of the pandemic is challenging, we performed data exclusion to ensure consistency in the pandemic's impact on outcome measures within the 6-month and 12-month assessments.

There were 2 distinct peaks in COVID-19 cases in Australia during the trial period. The first wave, characterized by infections imported from overseas, peaked in March and April 2020. The second wave, largely resulting from increased community transmission, peaked in August 2020 and subsided by October 2020 [33]. Therefore, we identified 2 key dates, April 26, 2020, and October 26, 2020, marking the progression of COVID-19 transmission and the implementation of Australian government public health measures.

April 26, 2020: The Queensland government applied mandatory social distancing, following a series of measures, including restrictions on home confinement, movement, and gathering [34].

October 26, 2020: The end of Australia's second wave of COVID-19 infections in 2020 [35,36].

During the period from April 26 to October 26, 2020, the participants presumably experienced a heightened impact of COVID-19 due to more strict government measures and increased local transmission compared to periods outside this timeframe. Notably, most of the 6-month surveys were administered prior to April 26, 2020, while most 12-month surveys were conducted between April 26 and October 26, 2020. Taking these factors into account, we excluded 6-month surveys conducted after April 26, 2020, and 12-month surveys conducted either before April 26, 2020, or after October 26, 2020. As a result, the 6-month and 12-month assessments were separated into different scenarios, reflecting periods of lower and higher COVID-19 impact, respectively.

Temporal Precision

The 6-month and 12-month surveys were designed to assess participants' outcome measures at 6 months and 12 months from the baseline, respectively. However, the actual timing of these surveys varied depending on participant availability. To enhance the temporal precision of the survey results, we excluded data collected outside a 1-month window around the intended survey time.

Results

Participants

A total of 1086 clients of our partnering aged care service providers were assessed. From those, 195 provided consent to participate in the trial and completed the baseline surveys (Figure 2). Randomization resulted in 97 participants in the control group and 98 participants in the intervention group. The final analysis included the surveys from 61 participants in the control group and 69 participants in the intervention group. The absence of follow-up surveys can be attributed to participants' withdrawals (n=56) or loss of contact (n=9). Additionally, 29 surveys were excluded to account for the impact of COVID-19, and 26 surveys were excluded to ensure the temporal precision of the data.

In total, across the 73 participants included in the aged care service provider's intervention logs, on 284 occasions a decision was made to contact the participant. The contact was made either directly with the participant (144 instances) or through an email sent to the participant's case manager, requesting they contact the participant (140 instances).

XSL•FO
RenderX

Figure 2. Flow diagram of inclusion process and number of included participants.

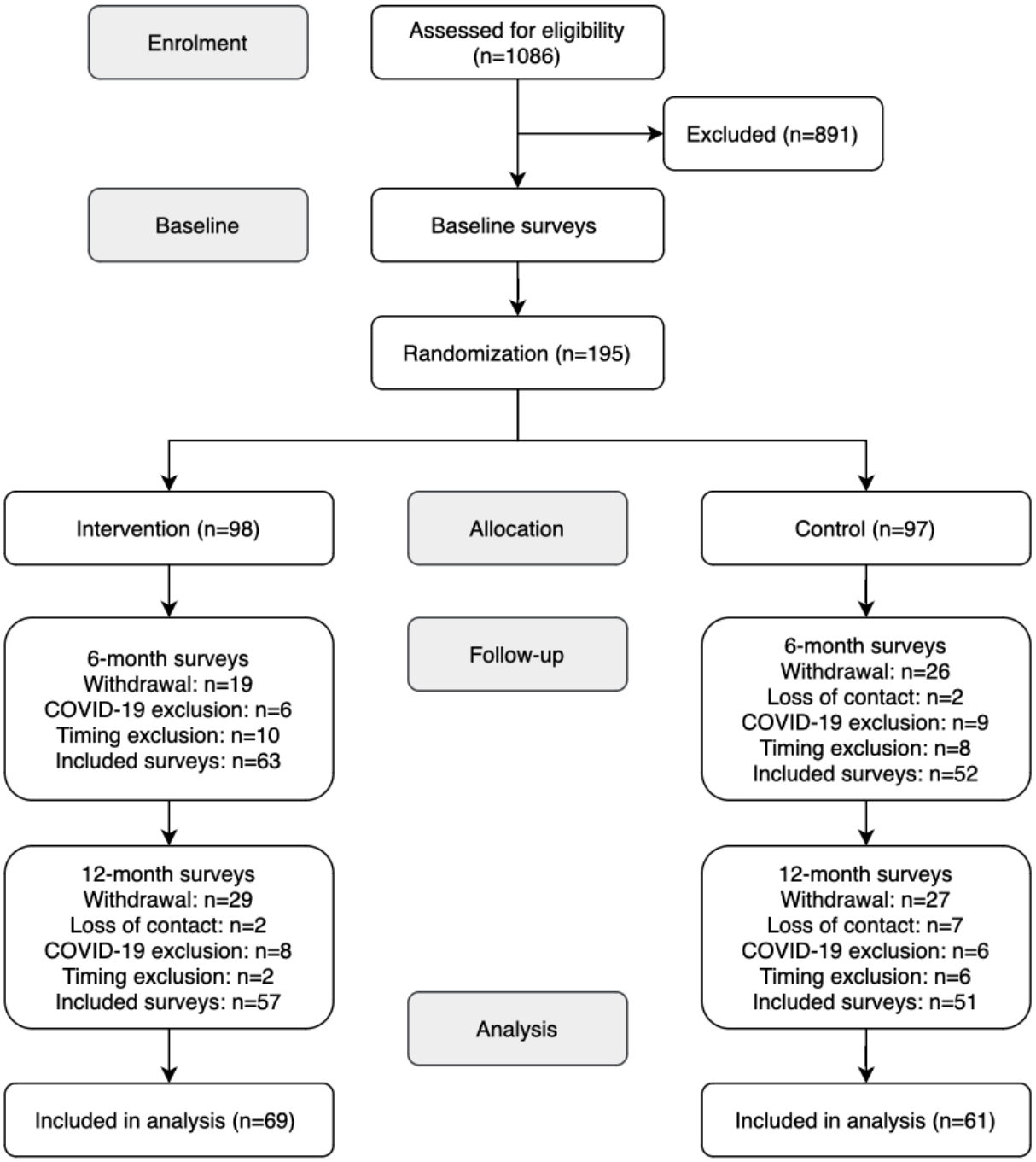

Baseline Characteristics

The sample included 130 participants, with 69 in the intervention group and 61 in the control group. The average age of all participants at the baseline was 82.4 (SD 7.3) years.

Most participants were male (70% overall), with a slightly higher percentage of males in the control group (73.8%) compared to the intervention group (66.7%). Both groups received their regular aged care services, either CHSP (45.4%) or HCP (54.6%). In terms of baseline survey results, there were slight variations between the intervention and control groups in ASCOT, EQ-5D-5L, Katz ADL, IADL, and GDS scores. Additional details are available in Table 1.

XSL•FO
RenderX

Table 1. Baseline characteristics.

Characteristics	Overall (n=130)	Intervention (n=69)	Control (n=61)
Age (years), mean (SD)	82.4 (7.3)	82.7 (7.0)	82.1 (7.7)
Gender, n (%)			
Female	39 (30)	23 (33.3)	16 (26.2)
Male	91 (70)	46 (66.7)	45 (73.8)
Care package, n (%)			
CHSP[a]	59 (45.4)	31 (44.9)	28 (45.9)
HCP[b]	71 (54.6)	38 (55.1)	33 (54.1)
Marriage status, n (%)			
Not married	6 (4.7)	3 (4.3)	3 (4.9)
Married	49 (38)	26 (37.7)	23 (37.7)
Widowed	57 (44.2)	31 (44.9)	26 (42.6)
Divorced	13 (10.1)	7 (10.1)	6 (9.8)
Separated	5 (3.1)	2 (2.9)	3 (4.9)
Living status, n (%)			
Alone	72 (55.4)	35 (50.7)	37 (60.7)
With other	58 (44.6)	34 (49.3)	24 (39.3)
Aged care service provider, n (%)			
Anglicare Southern Queensland	85 (65.4)	45 (65.2)	40 (65.6)
integratedLiving Australia	28 (21.5)	12 (17.4)	16 (26.2)
All About Living	17 (13)	12 (17.4)	5 (8.2)
Baseline surveys, mean (SD)			
ASCOT[c]	0.859 (0.147)	0.866 (0.148)	0.852 (0.148)
EQ-5D-5L[d]	0.724 (0.258)	0.726 (0.244)	0.721 (0.274)
Katz ADL[e]	5.256 (1.201)	5.412 (0.918)	5.082 (1.441)
IADL[f]	4.969 (2.162)	4.910 (2.058)	5.033 (2.287)
GDS[g]	3.146 (2.512)	3.319 (2.518)	2.951 (2.513)

[a]CHSP: Commonwealth Home Support Programme.

[b]HCP: Home Care Package.

[c]ASCOT: Adult Social Care Outcomes Toolkit ranges from –0.17 to 1, with higher scores indicating better social care–related quality of life.

[d]EQ-5D-5L: EuroQol-5 Dimensions-5 Levels ranges from –0.301 to 1, with higher scores indicating better health-related quality of life.

[e]Katz ADL: Katz Index of Independence in Activities of Daily Living ranges from 0 to 6, with higher scores indicating greater independence in daily activities.

[f]IADL: Lawton Instrumental Activities of Daily Living Scale ranges from 0 to 8, with higher scores indicating greater ability to perform instrumental activities of daily living.

[g]GDS: Geriatric Depression Scale ranges from 0 to 15, with higher scores indicating more severe depression symptoms.

Primary and Secondary Outcomes

Table 2 presents the effects of the SSH platform for the intervention and control groups on primary and secondary outcomes. At 6 months, there was a statistically significant difference (P=.046) in ASCOT scores between the intervention and control groups. The mean difference indicated that the intervention group had a 0.045 (95% CI 0.001 to 0.089) higher ASCOT score than the control group. The Cohen d effect size of 0.377 indicated a moderate effect in favor of the intervention group. At 12 months, the mean difference between the intervention and the control group reduced to 0.031 (95% CI –0.014 to 0.076; P=.18). The Cohen d effect size of 0.259 also suggested a smaller effect size compared to the 6-month assessment. The overall test of the group variable suggests there was a statistically significant difference in ASCOT scores between the intervention and control groups (P=.04). However, the overall test of group × time interaction was not significant (P=.59).

XSL•FO
RenderX

As shown in Table 2, there were no statistically significant differences between the 2 groups in the secondary outcomes (EQ-5D-5L, Katz ADL, IADL, and GDS) at 6 months or 12 months.

Table 2. Effects of the Smarter Safer Home platform for the intervention and control groups on primary and secondary outcomes.

Outcome	Intervention, mean (SD)	Control, mean (SD)	Mean difference (95% CI)	*P* value	Cohen *d* (95% CI)	Group, *P* value	Group × time, *P* value
ASCOT[a]						.04	.59
6 months	0.882 (0.119)	0.837 (0.118)	0.045 (0.001 to 0.089)	.046	0.377 (0.008 to 0.749)		
12 months	0.855 (0.118)	0.824 (0.118)	0.031 (–0.014 to 0.076)	.18	0.259 (–0.119 to 0.640)		
EQ-5D-5L[b]						.15	.26
6 months	0.745 (0.217)	0.669 (0.216)	0.076 (–0.004, 0.157)	.06	0.349 (–0.019 to 0.721)		
12 months	0.654 (0.216)	0.634 (0.216)	0.020 (–0.063, 0.102)	.64	0.090 (–0.288 to 0.468)		
Katz ADL[c]						.65	.90
6 months	5.123 (0.992)	5.038 (0.986)	0.085 (–0.284 to 0.454)	.65	0.085 (–0.282 to 0.453)		
12 months	4.930 (0.984)	4.875 (0.985)	0.054 (–0.325 to 0.434)	.78	0.055 (–0.325 to 0.435)		
IADL[d]						.70	.10
6 months	4.623 (1.046)	4.498 (1.043)	0.125 (–0.264 to 0.514)	.53	0.119 (–0.248 to 0.487)		
12 months	4.357 (1.043)	4.610 (1.040)	–0.253 (–0.652 to 0.146)	.21	–0.241 (–0.622 to 0.137)		
GDS[e]						.26	.21
6 months	3.479 (1.976)	3.575 (0.273)	–0.096 (–0.831 to 0.639)	.80	–0.048 (–0.416 to 0.319)		
12 months	3.673 (1.958)	4.289 (0.275)	–0.617 (–1.368 to 0.134)	.11	–0.312 (–0.694 to 0.066)		

[a]ASCOT: Adult Social Care Outcomes Toolkit ranges from –0.17 to 1, with higher scores indicating better social care–related quality of life.

[b]EQ-5D-5L: EuroQol-5 Dimensions-5 Levels ranges from –0.301 to 1, with higher scores indicating better health-related quality of life.

[c]Katz ADL: Katz Index of Independence in Activities of Daily Living ranges from 0 to 6, with higher scores indicating greater independence in daily activities.

[d]IADL: Lawton Instrumental Activities of Daily Living Scale ranges from 0 to 8, with higher scores indicating greater ability to perform instrumental activities of daily living.

[e]GDS: Geriatric Depression Scale ranges from 0 to 15, with higher scores indicating more severe depression symptoms.

Sensitivity Analyses

Significant improvements in ASCOT scores were observed, particularly among female participants (mean difference 0.089, 95% CI 0.016 to 0.162; *P*=.02) and HCP recipients (mean difference 0.079, 95% CI 0.015 to 0.142; *P*=.02) at 6 months (Tables S1 and S2 in Multimedia Appendix 1). Participants living with others showed a marginal improvement in ASCOT (mean difference 0.057, 95% CI –0.003 to 0.117; *P*=.06) at 6 months (Table S3 in Multimedia Appendix 1). The impact of the SSH platform on ASCOT for male participants, CHSP recipients, and participants living alone was not statistically significant. The impact on EQ-5D-5L, Katz ADL, IADL, and GDS was generally consistent with the main analysis (Tables S1, S2, and S3 in Multimedia Appendix 1).

Discussion

Principal Findings

In this RCT, 195 older people living in place were randomly assigned to intervention (using the SSH platform) and control groups. The analysis included 130 participants following withdrawals, loss of contact, and data exclusion. The effects of the SSH platform were assessed through the ASCOT, EQ-5D-5L, Katz ADL, IADL, and GDS collected at baseline, 6 months, and 12 months. The results demonstrated significant improvements in the primary outcome, social care–related quality of life, with the intervention group displaying higher ASCOT scores compared to the control group at 6 months. However, the long-term effect of the smart home solution on

XSL•FO
RenderX

social care–related quality of life, as observed in the 12-month assessment, did not align with the initial findings.

The improvement in ASCOT can be interpreted through several factors related to the characteristics of the SSH platform. ASCOT measures different aspects of quality of life that can be affected by social care, such as safety, social participation, and dignity. The SSH platform's features, including daily updates on activities and carer involvement, can improve social care–related quality of life by providing a sense of security and connectivity. On the other hand, EQ-5D-5L, Katz ADL, and IADL are associated with functional capacities and daily activities that may require specific interventions that the SSH platform does not directly address. For GDS scores, while the SSH platform can provide a sense of security and support, it may not be sufficient to significantly impact depressive symptoms without targeted interventions.

The diminished impact of the SSH platform on ASCOT observed at 12 months might be explained by an unprecedented peak of local COVID-19 transmission and strict public health measures. This is consistent with the studies indicating that the COVID-19 pandemic, characterized by concerns about virus transmission, limitations on social interactions, and disruption of daily activities, can lead to increased loneliness and reduced quality of life [37,38]. Therefore, the ability of smart home solutions to enhance the social care–related quality of life among older individuals needs continued adaptation to external factors, such as the COVID-19 pandemic, which can influence these outcomes.

Although the mean difference in secondary outcomes was not statistically significant, the intervention group demonstrated better point estimates (higher scores in EQ-5D-5L, Katz ADL, and IADL, and lower scores in GDS) compared to the control group at 6 months. Additionally, the estimates of mean difference in the GDS scores showed an increase by the 12-month assessment. However, this study's sample size might have limited the power to detect the effect of the SSH on the secondary outcomes.

Limitations

When interpreting the results of this study, some limitations must be taken into account. The primary limitation stemmed from the unforeseen impact of COVID-19 on trial operations and data analysis. First, implementing blinding for group allocation in this trial was not feasible due to the nature of the intervention, which required installation of SSH in participants' houses. The COVID-19 pandemic had adverse effects on the mental and physical health of older individuals, leading to increased rates of depression, anxiety, and cognitive decline, ultimately affecting their overall quality of life [30-32]. We excluded some data to maintain the consistency of COVID-19's impact on outcome measures at the same time points. This effectively separated the 6-month assessment and 12-month assessment into different scenarios with lower and higher COVID-19 impact, respectively. The adverse impact of COVID-19 might have, to some degree, overshadowed the effects of the SSH platform during the 12-month assessment, influenced by the heightened transmission of COVID-19 and the implementation of public health measures in Australia. In addition, the reduction in sample size from the initial sample size calculation may have reduced the statistical power of the study to detect a difference between groups.

While this study aimed to investigate how the SSH impacts older adults living in their own homes, the original protocol [21] excluded those residing with others, placing an additional emphasis on independent living. However, during the initial phase of recruitment, it became apparent that excluding older adults living with others significantly limited the number of participants available for recruitment and compromised the final sample size. Consequently, we removed this exclusion criterion. This adjustment allowed us to enhance our recruitment efforts and more effectively pursue the main objective of this study.

In addition, among the 195 participants recruited in this trial, 47 participants did not complete the 6-month assessment, and 65 participants did not complete the 12-month assessment. We noted that the participants without 6-month or 12-month follow-up data had lower EQ-5D-5L or IADL scores at baseline (Table S4 in Multimedia Appendix 1).

Comparison With Prior Work

Quality of life has been extensively investigated as one of the health outcomes associated with the use of smart home technologies [9,12,39,40]. However, most studies have primarily examined the impact of these technologies on health-related quality of life, with limited focus on social care–related quality of life within the smart home context. An RCT found that telemonitoring of biometric data along with administering symptom questionnaires did not significantly improve the mental health component of health-related quality of life. However, it may have led to a worsened physical component of health-related quality of life in the telemonitored group compared to the usual care group over the study period [16]. In another RCT of an in-home weight monitoring system in patients with advanced heart failure, the intervention group displayed an increasing trend in quality of life, but the difference was not statistically significant [15]. A systematic review reported that there was no evidence supporting that smart home and home health monitoring technologies improved the health-related quality of life for older adults [9]. Similarly, a meta-analysis of 5 RCTs investigating the effects of smart homes on older patients with chronic conditions indicated that the telemonitoring group did not yield a statistically significant impact on quality of life [12]. A scoping review examining the effectiveness of smart home technologies in supporting community-dwelling older adults with dementia found no significant effects on health-related quality of life in the 5 included studies [40]. In sum, the existing body of literature on RCTs related to smart home technologies has not conclusively demonstrated their positive influence on health-related quality of life. Moreover, evidence regarding the impact of smart home solutions on social care–related quality of life remains considerably limited.

In this study, a significant improvement in social care–related quality of life has been observed among the intervention group as compared to the control group after the first 6 months of using the SSH platform. Despite the heightened COVID-19 impact on the 12-month assessment, the intervention group displayed a possible improvement (mean difference 0.031, 95%

XSL•FO
RenderX

CI –0.014 to 0.076) in social care–related quality of life at 12 months. The ASCOT employed in this study has been proven to be a promising measure to evaluate care services for older people by broadening the evaluative space beyond the health domain to well-being [41,42]. Several studies have recommended the use of the ASCOT in combination with the widely adopted EuroQol-5D (EQ-5D) to measure broader aspects of quality of life beyond physical functioning and health dimensions [18-20]. Even though statistical significance was not achieved in secondary outcome measures, including health-related quality of life, this study filled in a gap in existing literature by assessing the effect of smart home technologies on social care–related quality in community-dwelling older people. This study presented, for the first time, evidence from an RCT, highlighting the possibility of smart home technologies in enhancing the social care–related quality of life for older people living in their own homes.

Future Directions

This RCT has yielded an extensive collection of data, including survey responses from older adult participants, data collected from smart home sensors, and evaluations from both informal and formal caregivers. Beyond this paper, there are several ongoing or planned future research projects that will use various data sources derived from this RCT. For instance, in an exploratory study, we analyzed the correlation between GDS and motion sensor data from participants' homes, aiming to determine if signs of depression could be detected through observed changes in in-home movement patterns [43]. This approach is currently being expanded through machine learning techniques to discover digital biomarkers for various health risks. In addition, a cost-utility analysis will be carried out to evaluate the cost-effectiveness of the intervention. Future research will also include a qualitative analysis to integrate the log spreadsheet maintained by aged care service providers, documenting the usage of SSH throughout the trial, with feedback from both participants and caregivers. The analysis of this data has the potential to open new avenues for predictive health analytics and proactive healthcare interventions. Besides, additional trials should be conducted to further evaluate the effect of the SSH platform on the secondary outcomes or in specific subgroups.

Conclusions

This RCT is a pioneering step in the direction of assessing the impact of smart home technology on older adults' well-being while living in their own homes. It uniquely focused on social care–related quality of life using the ASCOT, along with EQ-5D-5L, Katz ADL, IADL, and GDS measures. The study found significant improvements in ASCOT among participants using the SSH platform during the first 6 months. However, after 12 months, the improvement did not maintain statistical significance. The increased impact of the COVID-19 pandemic on the 12-month assessment was a noteworthy factor, as the pandemic introduced unforeseen challenges and restrictions that might have obfuscated the long-term effects of the SSH platform. While no significant changes were observed in the secondary outcome measures, this study contributes to understanding the potential of smart home technologies in enhancing social care–related quality of life, underscoring the need for further research to investigate the long-term effects and broader implications of such technologies in the care of older adults.

Acknowledgments

Funding was provided by the Department of Health of the Australian Commonwealth Government, under the Dementia and Aged Care Services Research and Innovation Funding, and the CSIRO eHealth Research Program. In-kind support from participating universities and aged care service organizations was also provided.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensitivity analyses and baseline characteristics of the participants with and without follow-up data.
[PDF File (Adobe PDF File), 180 KB - jmir_v27i1e59921_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1285 KB - jmir_v27i1e59921_app2.pdf]

References

1. World population prospects 2019: highlights. United Nations Department of Economic and Social Affairs.: United Nations Department of Economic and Social Affairs; 2019. URL: https://www.ined.fr/fichier/s_rubrique/29368/wpp2019.highlights_embargoed.version_07june2019_vf.fr.pdf [accessed 2019-06-17]
2. Wilson T, Temple J. New population projections for Australia and the States and Territories, with a particular focus on population ageing. 2022. URL: https://cepar.edu.au/publications/working-papers/new-population-projections-australia-and-states-and-territories-particular-focus-population-ageing [accessed 2024-12-25]

XSL•FO
RenderX

3. Xiao LD, Harrington A, Mavromaras K, Ratcliffe J, Mahuteau S, Isherwood L, et al. Care workers' perspectives of factors affecting a sustainable aged care workforce. Int Nurs Rev 2021;68(1):49-58. [doi: 10.1111/inr.12635] [Medline: 33058198]
4. Global Strategy and Action Plan on Ageing and Health. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/329960 [accessed 2024-12-25]
5. Final Report: Care, Dignity and Respect. Royal Commission into Aged Care Quality and Safety. 2021. URL: https://www.royalcommission.gov.au/aged-care/final-report [accessed 2024-03-01]
6. Vasunilashorn S, Steinman BA, Liebig PS, Pynoos J. Aging in place: evolution of a research topic whose time has come. J Aging Res 2012;2012:120952-120956 [FREE Full text] [doi: 10.1155/2012/120952] [Medline: 22175020]
7. An ageing Australia: preparing for the future. Productivity Commission. URL: https://www.pc.gov.au/research/completed/ageing-australia#:~:text=A%20report%20by%20the%20Productivity,budgets%20of%20all%20Australian%20governments [accessed 2013-11-22]
8. Demiris G, Rantz M, Aud M, Marek K, Tyrer H, Skubic M, et al. Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study. Med Inform Internet Med 2004;29(2):87-94. [doi: 10.1080/14639230410001684387] [Medline: 15370989]
9. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rios Rincon A. Smart homes and home health monitoring technologies for older adults: a systematic review. Int J Med Inform 2016;91:44-59. [doi: 10.1016/j.ijmedinf.2016.04.007] [Medline: 27185508]
10. Demiris G, Thompson HJ, Reeder B, Wilamowska K, Zaslavsky O. Using informatics to capture older adults' wellness. Int J Med Inform 2013;82(11):e232-e241 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.03.004] [Medline: 21482182]
11. Morris ME, Adair B, Miller K, Ozanne E, Hampson R, Pearce AJ, et al. Smart-home technologies to assist older people to live well at home. Aging Sci 2013;01(01):101. [doi: 10.4172/2329-8847.1000101]
12. Liu P, Li G, Jiang S, Liu Y, Leng M, Zhao J, et al. The effect of smart homes on older adults with chronic conditions: a systematic review and meta-analysis. Geriatr Nurs 2019;40(5):522-530. [doi: 10.1016/j.gerinurse.2019.03.016] [Medline: 31029481]
13. Pal D, Triyason T, Funikul S. Smart homes and quality of life for the elderly: a systematic review. Taichung: IEEE; 2017 Presented at: IEEE International Symposium on Multimedia (ISM); 2017 December 11-13; Taichung, Taiwan. [doi: 10.1109/ism.2017.83]
14. Pal D, Triyason T, Funilkul S, Chutimaskul W. Smart homes and quality of life for the elderly: perspective of competing models. IEEE Access 2018;6:8109-8122. [doi: 10.1109/access.2018.2798614]
15. Goldberg LR, Piette JD, Walsh MN, Frank TA, Jaski BE, Smith AL, et al. Randomized trial of a daily electronic home monitoring system in patients with advanced heart failure: the weight monitoring in heart failure (WHARF) trial. Am Heart J 2003;146(4):705-712. [doi: 10.1016/s0002-8703(03)00393-4]
16. Pecina JL, Hanson GJ, Van Houten H, Takahashi PY. Impact of telemonitoring on older adults health-related quality of life: the Tele-ERA study. Qual Life Res 2013;22(9):2315-2321. [doi: 10.1007/s11136-013-0361-5] [Medline: 23408299]
17. Morris ME, Adair B, Ozanne E, Kurowski W, Miller KJ, Pearce AJ, et al. Smart technologies to enhance social connectedness in older people who live at home. Australas J Ageing 2014;33(3):142-152. [doi: 10.1111/ajag.12154] [Medline: 24730370]
18. Kaambwa B, Gill L, McCaffrey N, Lancsar E, Cameron ID, Crotty M, et al. An empirical comparison of the OPQoL-Brief, EQ-5D-3 L and ASCOT in a community dwelling population of older people. Health Qual Life Outcomes 2015;13(1):164 [FREE Full text] [doi: 10.1186/s12955-015-0357-7] [Medline: 26420314]
19. Singh J, Pokhrel S, Longworth L. Can social care needs and well-being be explained by the EQ-5D? Analysis of the health survey for England. Value Health 2018;21(7):830-838 [FREE Full text] [doi: 10.1016/j.jval.2018.01.002] [Medline: 30005755]
20. Bulamu NB, Kaambwa B, Ratcliffe J. A systematic review of instruments for measuring outcomes in economic evaluation within aged care. Health Qual Life Outcomes 2015;13(1):179 [FREE Full text] [doi: 10.1186/s12955-015-0372-8] [Medline: 26553129]
21. Zhang Q, Varnfield M, Higgins L, Smallbon V, Bomke J, O'Dwyer J, et al. The smarter safer homes solution to support older people living in their own homes through enhanced care models: protocol for a stratified randomized controlled trial. JMIR Res Protoc 2022;11(1):e31970 [FREE Full text] [doi: 10.2196/31970] [Medline: 35072640]
22. Rural, remote and metropolitan area. Australian Government, Department of Health. 1994. URL: https://www.health.gov.au/%20health-topics/health-workforce/health-workforce-classifications/rural-remote-and-metropolitan-area [accessed 2024-01-25]
23. Netten A, Burge P, Malley J, Potoglou D, Towers A, Brazier J, et al. Outcomes of social care for adults: developing a preference-weighted measure. Health Technol Assess 2012;16(16):1-166 [FREE Full text] [doi: 10.3310/hta16160] [Medline: 22459668]
24. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]

XSL•FO
RenderX

25. Norman R, Mulhern B, Lancsar E, Lorgelly P, Ratcliffe J, Street D, et al. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. Pharmacoeconomics 2023;41(4):427-438 [FREE Full text] [doi: 10.1007/s40273-023-01243-0] [Medline: 36720793]
26. Shelkey M, Wallace M. Katz index of independence in activities of daily living. J Gerontol Nurs 1999;25(3):8-9. [doi: 10.3928/0098-9134-19990301-05] [Medline: 10362969]
27. Coyne R, Kluwer WK. AGACNP-BC, The Lawton Instrumental Activities of Daily Living (IADL) scale. Gerontologist 2019;9:179-186. [doi: 10.4324/9781003525288-7]
28. Greenberg SA. The Geriatric Depression Scale (GDS). Best Pract Nurs Care Older Adults 2012;4:1-2. [doi: 10.1097/01.NAJ.0000292204.52313.f3]
29. Malley JN, Towers AM, Netten AP, Brazier JE, Forder JE, Flynn T. An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people. Health Qual Life Outcomes 2012;10(1):21 [FREE Full text] [doi: 10.1186/1477-7525-10-21] [Medline: 22325334]
30. Bu F, Steptoe A, Fancourt D. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. Public Health 2020;186:31-34 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.036] [Medline: 32768621]
31. Lebrasseur A, Fortin-Bédard N, Lettre J, Raymond E, Bussières EL, Lapierre N, et al. Impact of the COVID-19 pandemic on older adults: rapid review. JMIR Aging 2021;4(2):e26474 [FREE Full text] [doi: 10.2196/26474] [Medline: 33720839]
32. De Pue S, Gillebert C, Dierckx E, Vanderhasselt MA, De Raedt R, Van den Bussche E. The impact of the COVID-19 pandemic on wellbeing and cognitive functioning of older adults. Sci Rep 2021;11(1):4636 [FREE Full text] [doi: 10.1038/s41598-021-84127-7] [Medline: 33633303]
33. The first year of COVID-19 in Australia: direct and indirect health effects. Australian Institute of Health and Welfare. 2021. URL: https://www.aihw.gov.au/reports/burden-of-disease/the-first-year-of-covid-19-in-australia [accessed 2021-09-10]
34. Superseded - Home Confinement, Movement and Gathering Direction (No. 2). Queensland Health.: Queensland Health URL: https://tinyurl.com/28hre3vy [accessed 2024-02-13]
35. Thompson J, McClure R, Blakely T, Wilson N, Baker MG, Wijnands JS, et al. Modelling SARS-CoV-2 disease progression in Australia and New Zealand: an account of an agent-based approach to support public health decision-making. Aust N Z J Public Health 2022;46(3):292-303 [FREE Full text] [doi: 10.1111/1753-6405.13221] [Medline: 35238437]
36. Statement From the Premier. Premier of Victoria. URL: http://www.premier.vic.gov.au/statement-premier-79 [accessed 2024-02-13]
37. Losada-Baltar A, Jiménez-Gonzalo L, Gallego-Alberto L, Pedroso-Chaparro MDS, Fernandes-Pires J, Márquez-González M. 'We Are Staying at Home.' Association of self-perceptions of aging, personal and family resources, and loneliness with psychological distress during the lock-down period of COVID-19. J Gerontol B Psychol Sci Soc Sci 2021;76(2):e10-e16 [FREE Full text] [doi: 10.1093/geronb/gbaa048] [Medline: 32282920]
38. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020;5(5):e256. [doi: 10.1016/s2468-2667(20)30061-x]
39. Facchinetti G, Petrucci G, Albanesi B, De Marinis MG, Piredda M. Can smart home technologies help older adults manage their chronic condition? A systematic literature review. Int J Environ Res Public Health 2023;20(2):1205 [FREE Full text] [doi: 10.3390/ijerph20021205] [Medline: 36673957]
40. Moyle W, Murfield J, Lion K. The effectiveness of smart home technologies to support the health outcomes of community-dwelling older adults living with dementia: a scoping review. Int J Med Inform 2021;153:104513. [doi: 10.1016/j.ijmedinf.2021.104513] [Medline: 34116363]
41. Hackert MQN, Exel JV, Brouwer WBF. Valid outcome measures in care for older people: comparing the ASCOT and the ICECAP-O. Value Health 2017;20(7):936-944 [FREE Full text] [doi: 10.1016/j.jval.2017.03.012] [Medline: 28712623]
42. van Leeuwen KM, Bosmans JE, Jansen AP, Hoogendijk EO, van Tulder MW, van der Horst HE, et al. Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. Value Health 2015;18(1):35-43 [FREE Full text] [doi: 10.1016/j.jval.2014.09.006] [Medline: 25595232]
43. Prabhu D, Dennis M, Kholghi M, Lu W, Sandhu M, Packer K. Detecting depression-related movement changes in older adults using smart home motion sensors - a feasibility study. 2023 Presented at: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 24-27, 2023; Sydney, Australia. [doi: 10.1109/embc40787.2023.10340431]

Abbreviations

ADL: activities of daily living
ASCOT: Adult Social Care Outcomes Toolkit
CHSP: Commonwealth Home Support Programme
CSIRO: Commonwealth Scientific and Industrial Research Organisation
EQ-5D: EuroQol-5 Dimensions
EQ-5D-5L: EuroQol-5 Dimensions-5 Levels

XSL•FO
RenderX

GDS: Geriatric Depression Scale
HCP: Home Care Package
IADL: Lawton Instrumental Activities of Daily Living Scale
Katz ADL: Katz Index of Independence in Activities of Daily Living
RCT: randomized controlled trial
SSH: Smarter Safer Home

Edited by A Schwartz; submitted 29.04.24; peer-reviewed by C Wang, A Andrade; comments to author 20.05.24; revised version received 01.07.24; accepted 11.11.24; published 22.01.25.

Please cite as:
Lu W, Silvera-Tawil D, Yoon HJ, Higgins L, Zhang Q, Karunanithi M, Bomke J, Byrnes J, Hewitt J, Smallbon V, Freyne J, Prabhu D, Varnfield M
Impact of the Smarter Safer Homes Solution on Quality of Life and Health Outcomes in Older People Living in Their Own Homes: Randomized Controlled Trial
J Med Internet Res 2025;27:e59921
URL: https://www.jmir.org/2025/1/e59921
doi:10.2196/59921
PMID:

©Wei Lu, David Silvera-Tawil, Hwan-Jin Yoon, Liesel Higgins, Qing Zhang, Mohanraj Karunanithi, Julia Bomke, Joshua Byrnes, Jennifer Hewitt, Vanessa Smallbon, Jill Freyne, Deepa Prabhu, Marlien Varnfield. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of Artificial Intelligence, Internet of Things, and Edge Intelligence in Long-Term Care for Older People: Comprehensive Analysis Through Bibliometric, Google Trends, and Content Analysis

Shuo-Chen Chien[1], PhD; Chia-Ming Yen[2,3], PhD; Yu-Hung Chang[1], PhD; Ying-Erh Chen[4], PhD; Chia-Chun Liu[1], BS; Yu-Ping Hsiao[1], MSc; Ping-Yen Yang[1], MSc; Hong-Ming Lin[1], BS; Tsung-En Yang[1], BS; Xing-Hua Lu[1], MSc; I-Chien Wu[1], MD, PhD; Chih-Cheng Hsu[1,2], MD, PhD; Hung-Yi Chiou[1,5], PhD; Ren-Hua Chung[1], PhD

[1]Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan

[2]National Center for Geriatrics and Welfare Research, National Health Research Institutes, Yunlin County, Taiwan

[3]Graduate Institute of Biomedical Sciences, China Medical University, Taichung City, Taiwan

[4]Department of Risk Management and Insurance, Tamkang University, New Taipei City, Taiwan

[5]School of Public Health, College of Public Health, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Ren-Hua Chung, PhD
Institute of Population Health Sciences
National Health Research Institutes
35, Keyan Road, Zhunan Town
Miaoli County, 350
Taiwan
Phone: 886 37 246 166 ext 36105
Email: rchung@nhri.edu.tw

Abstract

Background: The global aging population poses critical challenges for long-term care (LTC), including workforce shortages, escalating health care costs, and increasing demand for high-quality care. Integrating artificial intelligence (AI), the Internet of Things (IoT), and edge intelligence (EI) offers transformative potential to enhance care quality, improve safety, and streamline operations. However, existing research lacks a comprehensive analysis that synthesizes academic trends, public interest, and deeper insights regarding these technologies.

Objective: This study aims to provide a holistic overview of AI, IoT, and EI applications in LTC for older adults through a comprehensive bibliometric analysis, public interest insights from Google Trends, and content analysis of the top-cited research papers.

Methods: Bibliometric analysis was conducted using data from Web of Science, PubMed, and Scopus to identify key themes and trends in the field, while Google Trends was used to assess public interest. A content analysis of the top 1% of most-cited papers provided deeper insights into practical applications.

Results: A total of 6378 papers published between 2014 and 2023 were analyzed. The bibliometric analysis revealed that the United States, China, and Canada are leading contributors, with strong thematic overlaps in areas such as dementia care, machine learning, and wearable health monitoring technologies. High correlations were found between academic and public interest, in key topics such as “long-term care” (τ=0.89, P<.001) and “caregiver” (τ=0.72, P=.004). The content analysis demonstrated that social robots, particularly PARO, significantly improved mood and reduced agitation in patients with dementia. However, limitations, including small sample sizes, short study durations, and a narrow focus on dementia care, were noted.

Conclusions: AI, IoT, and EI collectively form a powerful ecosystem in LTC settings, addressing different aspects of care for older adults. Our study suggests that increased international collaboration and the integration of emerging themes such as “rehabilitation,” “stroke,” and “mHealth” are necessary to meet the evolving care needs of this population. Additionally, incorporating high-interest keywords such as “machine learning,” “smart home,” and “caregiver” can enhance discoverability

XSL•FO
RenderX

and relevance for both academic and public audiences. Future research should focus on expanding sample sizes, conducting long-term multicenter trials, and exploring broader health conditions beyond dementia, such as frailty and depression.

(J Med Internet Res 2025;27:e56692) doi:10.2196/56692

KEYWORDS

bibliometric analysis; Google Trends; content analysis; long-term care; older adults; artificial intelligence; Internet of Things; edge intelligence

Introduction

The global long-term care (LTC) sector faces significant challenges, primarily driven by an aging population and increasing demand for services [1]. This surge strains existing systems, exposing critical issues such as workforce shortages and the physically and emotionally demanding nature of LTC work [2,3]. Additionally, the rising costs associated with delivering high-quality care challenge the sustainability of current LTC models, emphasizing the urgent need for innovative and efficient solutions [4,5]. In this context, integrating technologies such as artificial intelligence (AI), the Internet of Things (IoT), and edge intelligence (EI) into LTC presents substantial advantages [6]. AI facilitates predictive health care by enabling timely interventions and personalized care strategies, while IoT devices such as sensors and wearables provide continuous health monitoring, enhancing resident safety and autonomy [7,8]. EI complements these technologies by enabling real-time decision-making through local data processing, reducing latency, and improving responsiveness in care settings [9,10]. Together, these not only improve care quality but also alleviate the workload of health care staff by automating routine tasks.

Numerous studies have investigated the adoption and understanding of these technologies through methods such as bibliometric analysis, Google Trends, and content analysis. Bibliometric analysis provides a quantitative evaluation of academic literature, highlighting key contributors, research trends, and emerging topics within a field [11]. Google Trends offers valuable insights into public interest and search behaviors, tracking the popularity of specific topics across regions over time [12]. Meanwhile, content analysis allows for a qualitative examination of thematic patterns and sentiment across textual data sources [13]. Representative examples are as follows: Maugeri et al [14] analyzed global public and research interest in telemedicine from 2017 to 2022 using Google Trends and bibliometric data, highlighting the impact of COVID-19 on interest and publication growth; Wang et al [15] mapped the global research landscape on AI applications in geriatrics using bibliometric analysis, identifying major contributors and key topics such as dementia and Alzheimer's care. Additionally, Voleti and Bhat [16] conducted a bibliometric analysis focused on IoT and edge computing in health care, revealing the growing application of these technologies in smart health care, with significant challenges in data management and scalability. Similarly, Ziwei et al [17] performed a deep bibliometric study on IoT in smart health care, highlighting key trends such as the increasing integration of blockchain-based security and network connectivity in health care systems. However, while these studies provide valuable insights into specific aspects of these technologies, they often remain limited in scope by focusing solely on either bibliometric analysis, Google Trends, or content analysis. This fragmented approach may hinder a comprehensive understanding of the broader landscape. Therefore, there is a need for a more integrative overview that explores the combined impact of these advanced digital tools on LTC. By synthesizing bibliometric analysis, Google Trends, and content analysis, this study aims to bridge this gap and offer a more holistic perspective on how these smart systems are applied in LTC.

This paper seeks to address the identified gaps by offering a more comprehensive and integrated analysis of AI, IoT, and EI applications in LTC for older adults. By combining bibliometric analysis, Google Trends, and content analysis, the study provides both broader and deeper insights into the usage of these technologies. The bibliometric analysis approach maps the intellectual structure of the field, identifying key contributors and thematic areas, while Google Trends analysis reveals public interest in the same topics, expanding the scope of understanding. Content analysis of the top 1% of the most-cited, high-impact studies offers in-depth insights into the practical applications and challenges of these technologies in real-world LTC settings. This interdisciplinary approach not only enriches the academic discourse but also provides a clearer road map for future research and development in care technologies for older adults.

This study aims to comprehensively explore the integration of AI, the IoT, and EI in LTC for older adults, addressing diverse perspectives, including academic frameworks, public interest, and in-depth insights from high-impact research. Specifically, bibliometric analysis is used to map the intellectual structure of the field, identifying leading contributors, influential publications, and key research trends. Public interest is examined through Google Trends, offering insights into societal attention to relevant topics and their alignment with academic focus. Additionally, a detailed content analysis of the top 1% most-cited research papers is conducted to uncover practical applications and challenges associated with the deployment of these technologies in real-world LTC environments. By synthesizing these diverse methodological approaches, our study results could construct a comprehensive framework that enhances the understanding of current developments and informs future innovations in AI, IoT, and EI for improving the quality and efficiency of care for older adults.

XSL•FO
RenderX

Methods

Overview of the Research Process and Analytical Methods

Figure 1 illustrates the complete research process undertaken in this study. The process begins with defining the research question, which focuses on investigating the application of AI, IoT, and EI among older adults in LTC. After establishing the research focus, relevant bibliometric records were extracted from the chosen databases. A comprehensive bibliometric analysis followed, encompassing the generation of a 3-field plot, performance analysis (based on publication and citation metrics), co-occurrence network analysis, thematic analysis, and keyword analysis to identify key research themes. To enhance the findings from the bibliometric analysis, a Google Trends analysis was conducted, where the top 16 keywords from the bibliometric analysis were expanded with synonyms. Kendall tau (τ) correlation analysis was performed to explore the relationship between keyword occurrences in bibliometric analysis and Google Trends data. Finally, a content analysis of the top 1% of the most-cited papers was conducted, offering deeper insights into practical applications. The results of these analyses inform the discussion and provide recommendations for future research directions.

Figure 1. The overview of workflow in this study.

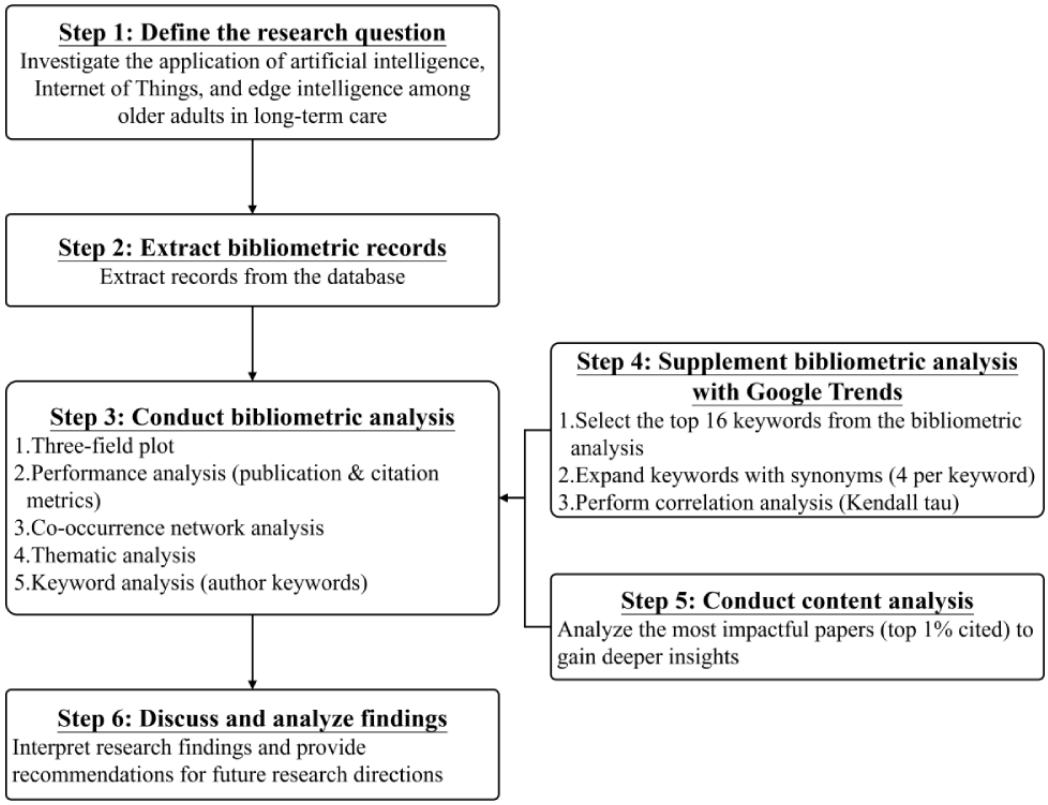

For our analysis, we used the R package "Bibliometrix" (version 4.1.4; R Foundation) for bibliometric analysis, and Python 3.8 (Python Foundation) was used to conduct statistical analyses and create figures [18].

Ethics Considerations

The data utilized in this research were sourced from publicly accessible repositories, qualifying as secondary data and negating the necessity for direct human participant engagement. Consequently, this study does not present any ethical concerns warranting approval from an institutional review board.

The Extraction and Cleaning Process of Bibliometrics Data

We collected data through bibliographic searches across 3 major databases: Web of Science, PubMed, and Scopus. These searches, conducted in early September 2024, used a combination of keywords categorized into 3 groups: (1) terms about older adults; (2) terms linked to AI, IoT, or EI; and (3) terms associated with LTC, with the specific keywords detailed in Table S2 in Multimedia Appendix 1. Our inclusion criteria encompassed proceedings and early access papers, while we excluded papers published before 2013 or after 2023, non-English papers, and those classified under journal categories such as reviews, editorials, letters, books, corrections, and notes. The final search strategy used in each database is listed in Table S3 in Multimedia Appendix 1. Following recommendations from existing literature for ensuring robust bibliometric analysis, we set a minimum sample size of 200 papers for our qualitative review [19].

After the initial data extraction, we used Biblioshiny, a web-based application within the R Bibliometrix package designed for bibliometric analysis, to perform an automated data cleaning process aimed at improving data quality and reliability [18]. Duplicates arising from entries across multiple databases were identified and removed. To ensure consistency,

XSL•FO
RenderX

we also standardized author affiliations, retaining only the most relevant affiliation for each author at the time of publication. This process minimized potential errors, such as incorrect or multiple affiliations, which could have affected the accuracy of the analysis. By implementing this cleaning procedure, we ensured that the final data set accurately represented the scholarly landscape.

Bibliometric Analysis

This study used bibliometric analysis, a quantitative approach that leverages bibliographic data—such as authorship, citations, references, and keywords—to explore relationships and impacts within the field [11]. We focused on 5 distinct analysis methods to gain a comprehensive understanding of the intellectual structure of our research domain: 3-field plot, performance analysis (publication and citation metrics), co-occurrence network analysis, thematic analysis, and keyword analysis (author keywords). These methods are described in detail in the following sections.

Overview Using the Three-Field Plot

To gain a comprehensive understanding of the intellectual landscape of our research field, we used a 3-field plot, a robust method for providing an insightful overview of the research cohort [20]. The 3-field plot, also known as a Sankey diagram, visually represents the flow and connections between different bibliometric dimensions. In this study, we integrated 3 critical dimensions: the geographic distribution of research (country), the dissemination mediums (journal), and the thematic concentration (keywords). This multidimensional approach is instrumental in mapping the intellectual territory of the research field, allowing us to visualize the relationships and transitions between where research is conducted, how it is disseminated, and the thematic focus of the studies.

Performance Analysis Using Publication and Citation Numbers

To measure output performance, we used both the number of publications and citation counts [21]. The number of publications serves as a fundamental and credible metric for quantifying research output, systematically categorized by specific journals, institutions, or countries, providing a reliable measure of scholarly productivity and impact. Citation counts are a robust indicator of a paper's influence within the scientific community, with higher citation counts suggesting that the paper's findings have significantly supported subsequent research, reflecting increased scholarly interest and relevance in the field. Citation counts can be further categorized into 2 distinct metrics: total global citations (TGCs) and total local citations (TLCs). TGCs measure a paper's citations across external publications, highlighting its global influence, while TLCs track citations within our curated data set, providing a focused view of its impact within a specific context. Additionally, the H-index was used to evaluate both the quantity and impact of scientific output, combining aspects of productivity and citation impact. This multifaceted approach, encompassing publication counts, global and local citations, and the H-index, offers a comprehensive view of scientific contributions and their influence within our research domain.

Co-Occurrence Network Analysis

We used a co-occurrence network analysis, a methodological tool designed to visually map the interrelationships among various elements in scholarly literature [22]. This technique is particularly effective for illustrating connections formed through the simultaneous occurrence of elements such as keywords, authors, references, or terms in academic papers. In this study, we focused on analyzing the relationships among countries, institutions, and author's keywords. In the co-occurrence network, each element is represented as a node, with connections (edges) drawn between nodes that frequently appear together in the same articles or contexts. The strength of these connections is determined by the frequency of co-occurrence; more frequent appearances result in thicker, more pronounced links. By examining these networks, researchers can identify clusters of interrelated works or concepts, revealing patterns and intellectual currents that shape academic research. Specifically, the clusters identified through the co-occurrence network of the author's keywords provided the foundation for subsequent thematic mapping, where these clusters were further analyzed to explore thematic concentration and development within the research domain.

Thematic Analysis

Thematic analysis is a method used to identify and interpret patterns (themes) within data, particularly the research themes emerging from bibliometric data. It categorizes and maps research themes based on 2 dimensions: Relevance degree (Centrality), which measures a theme's importance within the research field, and Development degree (Density), which reflects how well-developed a theme is. The results are visually represented in Figure S5 in Multimedia Appendix 1, with themes plotted across 4 quadrants:

- Quadrant I (Motor Themes): Highly relevant and well-developed themes, central to the field and influential in driving research.
- Quadrant II (Niche Themes): Well-developed but less central themes, often specialized within narrower areas of study.
- Quadrant III (Emerging or Declining Themes): Less relevant and underdeveloped themes, either emerging trends or topics in decline.
- Quadrant IV (Basic Themes): Highly relevant but less developed themes that provide a foundation for broader research.

We divided our analysis into 3 phases, 2014-2016, 2017-2020, and 2021-2023, to observe how research themes have evolved. This approach allowed us to track shifts in focus, identify emerging trends, and highlight themes that have remained consistently relevant throughout the periods.

Keyword Analysis Using Author Keywords

In this study, we performed a keyword analysis using author keywords, which are terms selected by authors to encapsulate the core themes of their research papers. These keywords are essential for identifying major research trends and understanding focus areas within a field, as they reflect the concepts scholars consider most important in their studies. We selected the 16

XSL•FO
RenderX

most frequently occurring author keywords from our data set and analyzed their yearly trends to track their prominence in academic research over time. This analysis provided insights into which topics gained or lost attention, highlighting shifts in research priorities. To further explore the relevance of these keywords beyond academia, we used Google Trends to examine public interest in the same 16 keywords, allowing us to compare academic trends with public search behavior and gain a broader understanding of the societal impact of these research themes.

Enhancing Bibliometric Analysis With Public Interest Insights From Google Trends

We incorporated Google Trends as an analytical tool to capture public interest in the research topics identified through our keyword analysis [12]. Google Trends provides search volume scores ranging from 0 to 100 for each keyword, with 100 representing peak popularity, 50 indicating half of the peak popularity, and 0 reflecting insufficient data. The keywords selected for the Google Trends analysis were based on the most frequently occurring keywords from the bibliometric analysis, with the top 16 most frequent ones being chosen. To capture broader variations in public interest, we consulted with experts and selected 4 synonyms for each of these keywords, thereby expanding the search scope. The searches were conducted globally, covering the period from January 1, 2014, to December 31, 2023, with data provided on a monthly basis. We then calculated the annual search interest by averaging the monthly data for each year. Additionally, a weighted average of the search volumes for each keyword and its synonyms was computed to represent the overall Google Trends search interest for each term.

To align the units of comparison between academic keyword trends (from bibliometric analysis) and public interest (from Google Trends), we applied min-max scaling. This scaling was done globally across the 16 keywords and the period from 2014 to 2023, with the maximum and minimum values for both data sets used to normalize the data. After scaling, we multiplied each result by 100, ensuring that every value falls within the range of 0-100, enabling direct comparison between academic research frequency and public search volume. Finally, we applied the Kendall tau correlation coefficient, a nonparametric statistic, to measure the ordinal association between academic focus and public interest by comparing concordant and discordant pairs [23]. The correlation coefficient ranges from –1 (perfect inverse correlation) to +1 (perfect direct correlation), with $P<.05$ considered statistically significant. This method allowed us to quantitatively assess how well academic keyword trends aligned with public search interest over time.

Content Analysis and Quality Assessment

To gain a deeper understanding of the research focus in this study, we conducted a content analysis on the most impactful studies, specifically those with higher TLCs. Content analysis is a qualitative research method used to systematically classify and interpret text data by identifying themes and patterns [13]. This approach allowed us to explore key terms, concepts, and themes related to AI and IoT applications in LTC, helping to uncover trends, patterns, and gaps in the existing research.

Before undertaking the content analysis, a quality assessment was performed to evaluate the methodological rigor of the papers included in our review. We used the Mixed Methods Appraisal Tool to assess the quality of the selected studies (Table S4 in Multimedia Appendix 1). This tool is designed to evaluate a wide range of study types, including qualitative studies, randomized controlled trials, nonrandomized quantitative studies, quantitative descriptive studies, and mixed methods research. Each study was assessed using a nominal scale (yes/no/cannot tell), ensuring that only high-quality papers were included in the content analysis. Two independent reviewers (SCC and CCL) carried out the quality assessment, with discrepancies ultimately resolved by the corresponding author (RHC).

For the content analysis, we selected the top 1% of papers from the overall cohort, ranked primarily by TLC and secondarily by TGC. This selection process ensured that the studies analyzed not only had substantial local impact within the field but also possessed significant global relevance. From these papers, we extracted key information, including the country or region of the corresponding author, whether the study was conducted at a single center or multiple centers, the total number of participants, the percentage of female participants, the illness type of participants, whether assistive robots were used, the study's primary objective, main findings, and stated limitations. Based on this analysis, we gained deeper insights and provided recommendations for future research directions to address the identified gaps and challenges.

Results

Data Collection and Screening Process for Bibliometric Analysis

The keyword search across Web of Science, PubMed, and Scopus yielded 1987, 2934, and 10,369 records, respectively (Figure 2). Our screening process excluded papers published before 2014 or after 2023, resulting in 4569 eliminations. Additionally, we removed 890 review papers, 218 book chapters, 36 letters, 32 notes, 22 editorial materials, and 1993 other types of documents. Furthermore, 179 non-English papers were excluded. An additional 973 papers were discarded due to duplication. Ultimately, 6378 papers were selected for bibliometric analysis in the subsequent step. We also present the distribution of yearly publications and the average TGC per paper in Figure S6 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 2. The procedure involved in choosing appropriate papers.

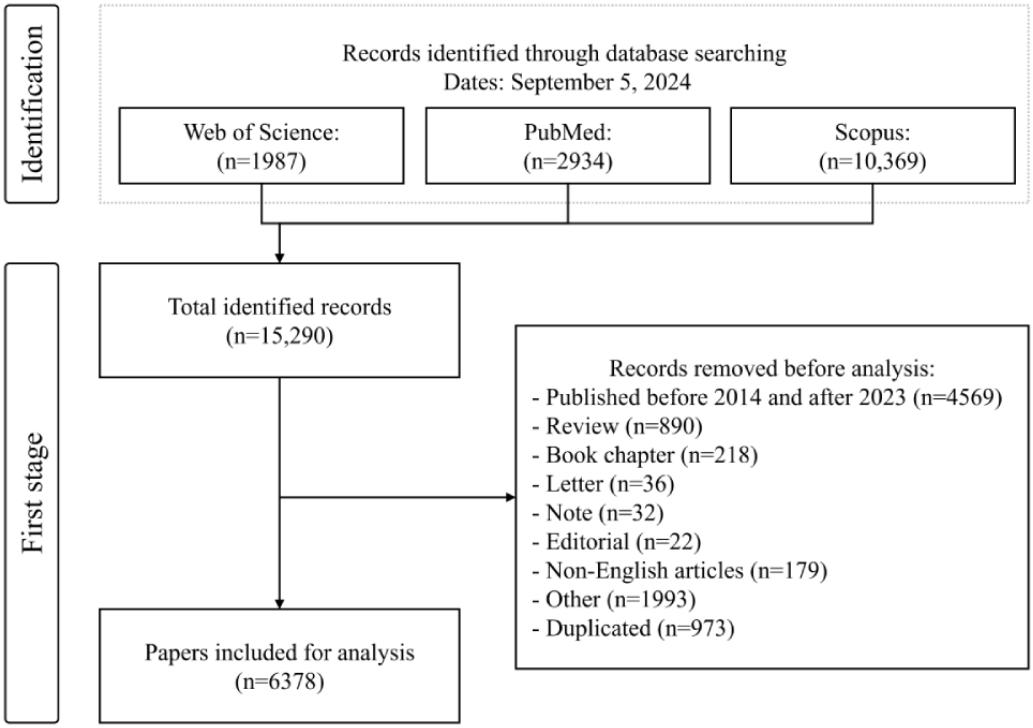

An Overview of This Study Field

We used a Sankey diagram to analyze and illustrate the distribution and interplay among countries, journals, and keywords in the articles. This method effectively highlights the usage of AI, IoT, and EI among older adults in LTC across various countries (Figure 3).

Figure 3. Three-field plot (country, journal, and keyword).

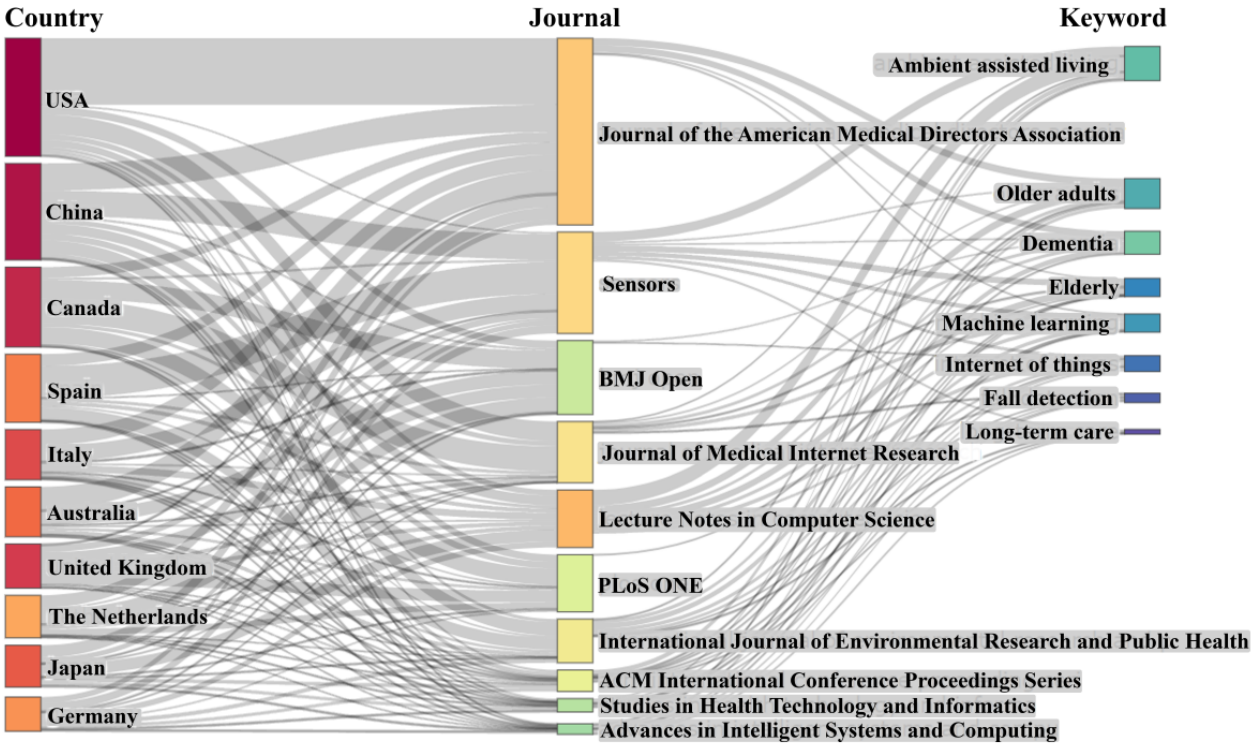

In terms of country contributions, the United States leads with 291 articles, followed by China (n=238), Canada (n=197), Spain (n=168), and Italy (n=124). Other notable contributors include Australia (n=123), the United Kingdom (n=109), the Netherlands (n=105), Japan (n=103), and Germany (n=84). As for journals, the Journal of the American Medical Directors Association

XSL•FO
RenderX

stands out with 460 articles, followed by Sensors (n=250), BMJ Open (n=182), and the Journal of Medical Internet Research (n=151). Other key journals include Lecture Notes in Computer Science (n=142), PLoS ONE (n=141), and the International Journal of Environmental Research and Public Health (n=108). These journals play a critical role in disseminating research on AI, IoT, and EI applications in LTC for older adults. Regarding keywords, prominent terms include “ambient assisted living” (n=84), “older adults” (n=74), “dementia” (n=57), and “elderly” (n=45). Other frequently used keywords are “machine learning” (n=45), “internet of things” (n=39), “fall detection” (n=23), and “long-term care” (n=12), reflecting the research focus on technological solutions aimed at improving the care and well-being of older adults.

The relationships between countries and journals reveal that major contributors such as the United States, China, and Canada publish across a wide range of journals, demonstrating a broad scope of research dissemination. Regarding journal-keyword relationships, distinct focus areas become apparent. For instance, the Journal of the American Medical Directors Association predominantly emphasizes studies on older adults and dementia, while Sensors focuses on topics such as ambient assisted living, machine learning, and care for older adults. By contrast, BMJ Open addresses general studies related to dementia and the IoT without a distinct keyword focus on other specific areas.

Most Productive Journals and Institutions

Table 1 presents a detailed overview of the most productive journals publishing research related to AI, IoT, and EI in LTC for older adults, focusing on key metrics such as publication counts, TGC, and H-Index. Leading the list, Lecture Notes in Computer Science plays an important role in this domain, with 223 publications (N=6378, 3.50%) and a TGC of 1217. The Journal of the American Medical Directors Association, which specializes in geriatrics, aging, and LTC, follows closely with 113 publications (N=6378, 1.77%) and an impressive TGC of 5812. Meanwhile, Sensors, a journal primarily dedicated to technology and sensor applications in health care and IoT, has contributed 108 publications and a TGC of 2560. The Journal of Medical Internet Research, focused on digital health and eHealth technologies, adds 89 publications with a TGC of 1822. Notably, the International Journal of Social Robotics, which focuses on robotics and its applications in social settings, has a smaller output of 36 publications, but its TGC of 1471 highlights its substantial impact per article.

Table 1. Most influential journals (sorted by the number of publications; N=6378).

Ranking	Journals	Item, n (%)	Total global citations	Total global citations per item	H-Index
1	Lecture Notes in Computer Science	223 (3.50)	1217	5.46	15
2	Journal of the American Medical Directors Association	113 (1.77)	5812	51.43	32
3	Sensors	108 (1.69)	2560	23.70	25
4	Journal of Medical Internet Research	89 (1.40)	1822	20.47	24
5	International Journal of Environmental Research and Public Health	85 (1.33)	773	9.09	15
6	ACM International Conference Proceedings Series	80 (1.25)	339	4.24	11
7	PLoS ONE	77 (1.21)	1447	18.79	20
8	BMJ Open	68 (1.07)	912	13.41	15
9	Studies in Health Technology and Informatics	62 (0.97)	336	5.42	10
10	Advances in Intelligent Systems and Computing	51 (0.80)	216	4.24	6
11	BMC Geriatrics	51 (0.80)	911	17.86	15
12	Gerontechnology	51 (0.80)	133	2.61	5
13	CEUR Workshop Proceedings	45 (0.71)	101	2.24	6
14	IEEE Access	41 (0.64)	624	15.22	12
15	JMIR mHealth and uHealth	40 (0.63)	883	22.08	18
16	Lecture Notes in Electrical Engineering	39 (0.61)	60	1.54	5
17	JMIR Aging	38 (0.60)	413	10.87	12
18	International Journal of Social Robotics	36 (0.56)	1471	40.86	20
19	Journal of the American Geriatrics Society	34 (0.53)	755	22.21	18
20	Communications in Computer and Information Science	32 (0.50)	54	1.69	4

The H-Index further underscores the influence of these journals in the field. The Journal of the American Medical Directors Association stands out with a robust H-Index of 32, reflecting its high citation impact. Sensors and Lecture Notes in Computer Science follow with H-Indices of 25 and 15, respectively, illustrating their diverse influence within the field. The Journal of Medical Internet Research also demonstrates substantial scholarly impact, with an H-Index of 24. Finally, despite its

XSL•FO
RenderX

smaller publication count, the International Journal of Social Robotics achieves an impressive H-Index of 20, underscoring the considerable influence of its limited number of publications.

We also identified the 20 most productive institutions worldwide based on their publication output, as detailed in Table S5 in Multimedia Appendix 1. Leading the list are the University of Toronto with 246 publications, the University of California with 169 publications, and the University of Michigan with 136 publications. Following closely are the University of Washington (n=120), Griffith University (n=106), and the University of Pittsburgh (n=102). Our analysis of the top 20 institutions reveals a diverse range of geographic locations, including institutions from Canada, the United States, Australia, the United Kingdom, Sweden, and Denmark. This global distribution underscores the widespread interest and active involvement in this field. The broad geographic representation not only highlights the substantial contributions of these institutions but also offers insights into potential collaboration opportunities across different regions.

Co-Occurrence Network and Thematic Analysis

In the bibliometric analysis, we constructed a co-occurrence network to explore the relationships and thematic intersections across 3 distinct fields: countries, institutions, and author's keywords, covering the period from 2014 to 2023. As shown in Figure 4, the network highlights the complex interactions within the research landscape, delineating patterns of collaboration and thematic focus. The figure visually represents the prominence and centrality of specific countries, institutions, and keywords, revealing key players and dominant themes in the field. This network provides valuable insights into how different entities contribute to and shape research on AI, IoT, and EI in LTC for older adults.

XSL•FO
RenderX

Figure 4. Co-occurrence network with three different fields.

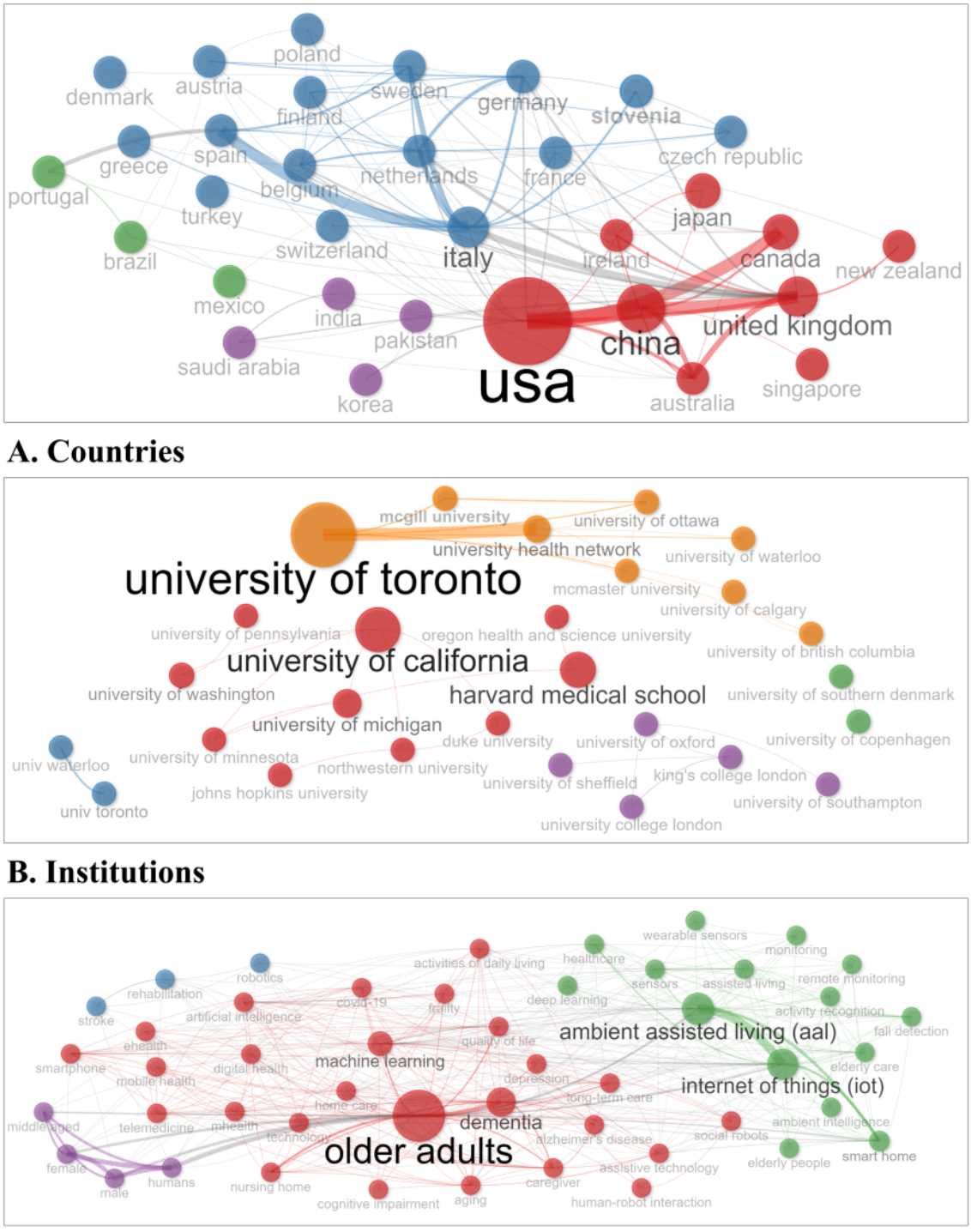

Regarding the countries, the bibliometric network analysis reveals a robust international collaboration framework within the research landscape, as depicted in Figure 4A. The United States is prominently positioned at the core of this network, underscoring its pivotal role in spearheading research initiatives on a global scale. Notable are the substantial bilateral links connecting the United States with other research-heavy nations such as China, the United Kingdom, and Canada, which suggest a high degree of cooperative research activity. European countries, represented by the cohesive cluster of Germany, Italy, and France, demonstrate strong intracontinental connections, indicative of their collective research endeavors. Meanwhile, countries such as Portugal and Brazil, though less central in the visualization, reveal their targeted and specialized contributions to the research community.

The visualization of institutional collaborations (Figure 4B) positions the University of Toronto as a central node within the network, highlighting its significant research output and

XSL•FO
RenderX

influence across disciplines. Other institutions, such as the University of California and Harvard Medical School, also command prominent places within the network, signifying their crucial roles in fostering extensive academic partnerships and knowledge dissemination. The depiction of a wide array of institutions from various corners of the globe points to a richly diverse and interdisciplinary academic milieu, signifying the breadth and depth of research collaborations and intellectual exchange.

The network of the author's keywords (Figure 4C) is dominated by the term "older adults," which acts as a thematic epicenter linking diverse research topics. Surrounding this pivotal term, we find noticeable nodes such as "dementia," "health care," and "assistive technology," reflecting concentrated areas of research focus and societal impact. The prevalence of cutting-edge technological terms such as "artificial intelligence," "internet of things," and "machine learning" adjacent to care topics for older adults underscores the integration of technology in addressing the complexities of aging. Additionally, emergent terms such as "COVID-19," "digital health," and "telemedicine" are a testament to the field's responsiveness to evolving global health issues and the increasing relevance of digital innovation in health care solutions.

To further understand the evolving research landscape, we conducted a thematic analysis of author keywords across 3 distinct periods: 2014-2016, 2017-2020, and 2021-2023. This analysis allowed us to track the shifts in research focus over the years and identify emerging or declining trends. The results, detailed and thoroughly documented in Multimedia Appendix 1 (see the section titled "Using Thematic Analysis for a Meticulous Examination Across Three Periods") show noticeable changes in central research themes such as "older adults," "dementia," and "nursing home," alongside the rise of newer themes such as "mHealth," "digital health," and "artificial intelligence" in recent years (see Figure S1 in Multimedia Appendix 1). These findings provide insights into how the integration of advanced technologies has increasingly shaped the research landscape in care for older adults, reflecting broader societal and health care trends.

Keyword Analysis: Enhancing Bibliometric Analysis With Google Trends for the Top 16 Keywords

We also adopted the keywords analysis to explore the evolving trends and key focus areas in AI, IoT, and EI applications for older adults in LTC. In our study, we first identified the top 16 most frequently occurring keywords from the bibliometric analysis between 2014 and 2023. To further enhance the insights gained, we assessed public interest in these keywords using Google Trends. The detailed results of the keyword analysis, both from bibliometric analysis and Google Trends, are presented in Multimedia Appendix 1 under the section titled "Top 16 Frequent Keywords From Bibliometric Analysis and Their Google Trends Search Interest" (also see Table S1 and Figures S2-S4 in Multimedia Appendix 1). Among the most notable findings, "older adults" ranked as the most frequent keyword in the bibliometric analysis with 733 occurrences, and it consistently showed high search interest in Google Trends, peaking at 65 in 2020. Similarly, "machine learning" appeared 287 times in the bibliometric analysis and saw a notable rise in public interest, with its search score increasing from 15 in 2015 to 95 in 2023. The term "internet of things," which occurred 353 times in the bibliometric analysis, also showed a modest growth in public interest, with its score rising from 39 in 2014 to 46 in 2023. By contrast, "ambient assisted living" peaked at 59 occurrences in the bibliometric analysis in 2017, but its search interest declined from 40 in 2017 to 18 in 2023. Additionally, "fall detection" exhibited lower public interest in the first 5 years, beginning to grow slowly in 2019 but remaining relatively steady over the last 5 years, despite being mentioned 152 times in the bibliometric analysis.

Next, we analyzed the relationship between the keyword results from bibliometric analysis and Google Trends by calculating the correlation using the Kendall rank correlation coefficient. To address differences in scale between the 2 data sets, we applied min-max scaling to normalize the values into a range of 0-100 before conducting the analysis. The results are shown in Figure 5, and the statistical outcomes of this comparison are detailed in Table S6 in Multimedia Appendix 1. In the figure, the red trend lines represent keywords related to the older adults category. Among these, "long-term care" showed the strongest positive correlation (τ=0.89, P<.001), followed by "older adults" (τ=0.87, P<.001), both of which had significant positive correlations (P<.05), indicating a strong alignment between academic and public interest. The terms "dementia" (τ=0.75, P=.003) and "caregiver" (τ=0.72, P=.004) also exhibited significant positive correlations, reflecting growing public attention in these areas.

XSL•FO
RenderX

Figure 5. Correlation between bibliometric analysis and Google Trends volume.

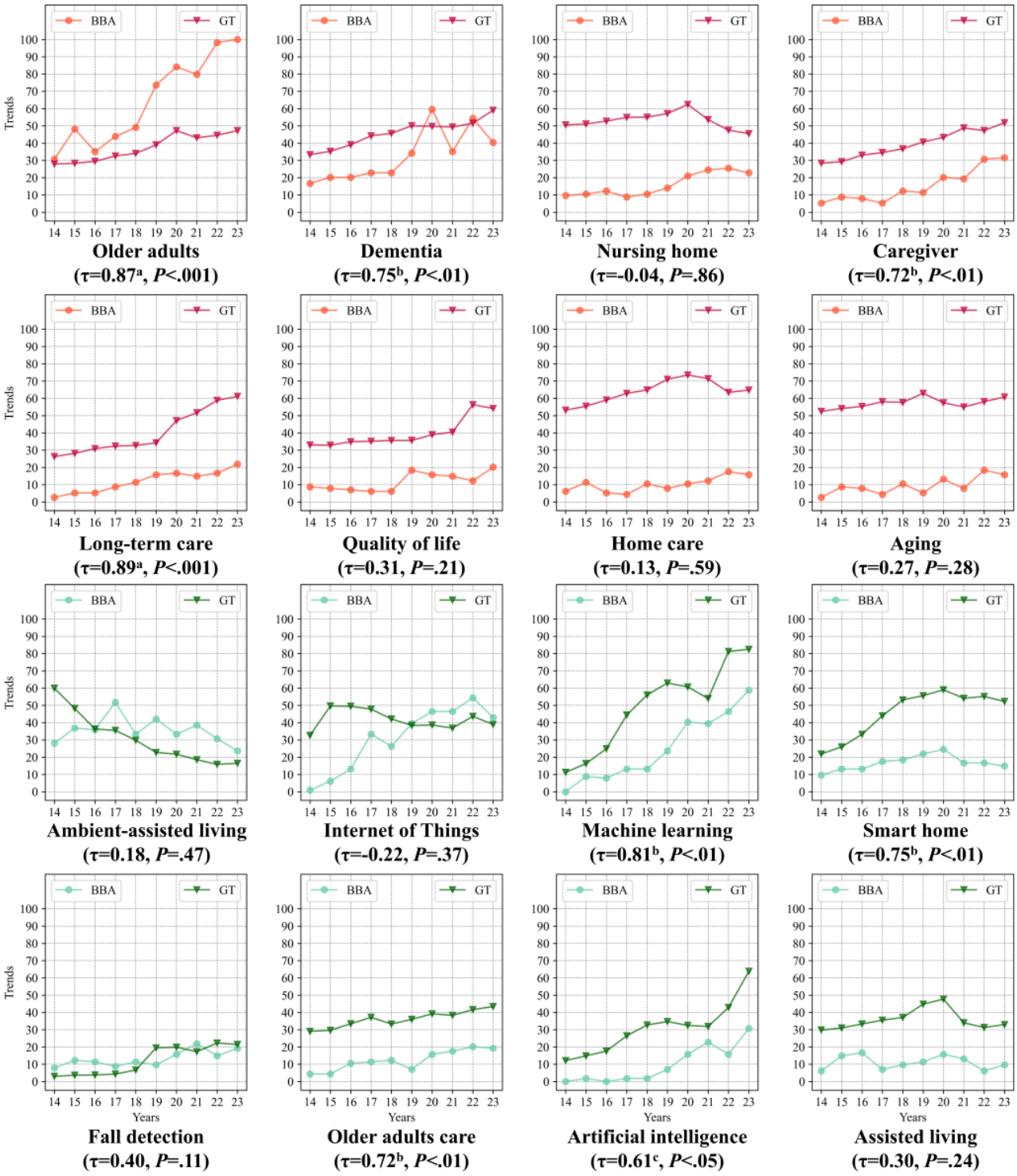

For the green trend lines, which represent keywords related to the ambient assisted living category, the strongest significant correlation was observed for "machine learning" (τ=0.81, *P*=.001), indicating increasing public and research interest. This was followed by "smart home" with a significant positive correlation (τ=0.75, *P*=.003). Similarly, "older adults care" showed a significant positive correlation (τ=0.72, *P*=.004), while "artificial intelligence" exhibited a moderate correlation (τ=0.61, *P*=.02). However, terms such as "fall detection" (τ=0.40, *P*=.11) and "ambient assisted living" (τ=0.18, *P*=.47) did not demonstrate statistically significant correlations, indicating weaker public interest compared with academic research in these areas.

Content Analysis

To gain a deep understanding of the papers, we conducted a content analysis of the top 1% of most impactful papers, ranked primarily by TLC and secondarily by TGC, totaling 64 papers [24-87]. The quality assessment outcomes using the Mixed Methods Appraisal Tool are listed in Table S7 in Multimedia Appendix 1, along with the reasons. For the content analysis, we extracted key information such as the country or region of

XSL•FO
RenderX

the corresponding author, whether the study was conducted at a single center or multiple centers, the total number of participants, the percentage of female participants, the illness type of participants, whether assistive robots were used, the study's primary objective, main findings, and stated limitations in Table S8 in Multimedia Appendix 1.

The content analysis of the top 1% of impactful papers reveals key trends in the use of assistive technologies, particularly social robots, in care for older adults, especially for individuals with dementia. Many studies, such as those involving the robot PARO, consistently highlight its positive impact on reducing agitation and depression, and improving mood among patients with dementia. Long-term benefits were observed in several cases, with the psychosocial environment playing a critical role in the success of these interventions. Despite the overall positive outcomes, limitations such as small sample sizes, short study durations, and difficulties in maintaining participant engagement across various care settings were frequently noted.

A recurring theme is the exploration of the human-robot interaction and how it shapes care outcomes. For instance, multiple studies compared the effectiveness of robotic pets such as PARO against plush toys or live animals. These studies found that robots generally outperformed nonrobotic alternatives in enhancing engagement and managing symptoms. However, some studies, such as the one by Thodberg et al [53], noted a diminishing novelty effect over time, with live animals often prompting more sustained interaction. This highlights the need for future work to focus on maintaining long-term engagement with robotic interventions.

There is a significant focus on addressing the emotional and cognitive needs of older adults, especially those with mild cognitive impairment or more advanced dementia. Robots such as RobuLAB and CuDDler demonstrated potential in supporting cognitive and social functions, though customization and usability were critical factors influencing their acceptance. For example, Pino et al [39] highlighted how older adults and informal caregivers appreciated socially assistive robots for their adaptability, emphasizing the importance of designing robots that can be personalized to meet individual needs. Nevertheless, limitations such as small sample sizes and lack of direct interaction in some studies restricted generalizability.

Table 2 outlines key research gaps and future directions. A major limitation across studies is the absence of long-term follow-up, which constrains understanding of the sustained impacts of these technologies. Furthermore, the small sample sizes limit the generalizability of results. The narrow focus on dementia care, with limited exploration of other conditions such as frailty or depression, represents another critical gap. Lastly, there is insufficient research on the real-world implementation and scalability of these technologies. Future studies should prioritize long-term investigations, expand sample sizes through multicenter collaborations, broaden the scope to include other health conditions, and focus on real-world applicability, particularly scalability and cost-effectiveness.

Table 2. The summary of current research gaps and suggestions.

Current research gap	Suggestion
Studies focus on short-term outcomes.	Conduct longitudinal studies to assess long-term impact.
Small sample sizes limit generalizability.	Increase sample sizes through multicenter studies.
Research mostly targets dementia care.	Expand research to include conditions such as frailty and depression.
Limited real-world implementation and scalability data.	Prioritize studies in real-world settings and assess scalability and cost-effectiveness.

Discussion

Principal Findings

In this study, we utilized bibliometric analysis, Google Trends, and content analysis to gain a comprehensive understanding of the research interest in AI, IoT, and EI applications for older adults in LTC over the last decade. Our bibliometric analysis results included 6378 papers that revealed geographic diversity, with leading contributions from countries like the United States, China, and Canada and institutions such as the University of Toronto and the University of California. The increasing focus on keywords such as "older adults," "machine learning," and "internet of things" highlights their growing importance in enhancing care delivery. Additionally, the convergence of societal and research interests is reflected in the strong correlations between "long-term care," "dementia," and "caregiver" in both bibliometric analysis and Google Trends. Content analysis of the most impactful papers demonstrated that assistive technologies, particularly social robots such as PARO, consistently benefit patients with dementia by improving mood and reducing agitation. Overall, AI, IoT, and EI show significant potential in transforming personalized care, safety, and decision-making in care settings for older adults.

Implications

Artificial Intelligence: The Usage in LTC

AI's potential in the care of older adults is profoundly impactful, particularly in enhancing personalized care and engagement for older adults. Social robots, such as PARO, exemplify how AI can address the emotional and psychological needs of older adults, especially those with dementia. In Norway, a study involving 60 participants found that the use of PARO significantly reduced agitation and depression among patients with dementia, with positive effects lasting up to 3 months after intervention [34]. Similarly, in Australia, a larger study with 415 participants demonstrated that PARO fostered higher engagement and improved mood when compared with a plush toy used in control settings [61]. These results suggest that AI-powered robots can simulate companionship and therapeutic interactions, providing psychological comfort and enhancing the quality of life in LTC facilities. Furthermore, AI's adaptability enables tailored interventions, where robots such

XSL•FO
RenderX

as PARO adjust their responses based on real-time emotional cues, offering a more personalized and human-like experience for older residents. As such, AI is not merely a tool for automation but a transformative technology in care for older adults, contributing to improved mental health, reduced reliance on pharmacological interventions, and enhanced social interaction.

Internet of Things: The Impact on Health Monitoring and Safety in LTC

IoT technologies are revolutionizing safety and continuous health monitoring in older adult care settings by providing real-time data and immediate interventions. Wearable devices and sensors are increasingly used to monitor vital signs such as heart rate, movement, and sleep patterns, significantly contributing to the well-being of older adults. For example, a Hong Kong study developed an Internet of Medical Things system that monitored real-time biometric data, enabling health care providers to proactively modify care plans based on continuous monitoring [83]. This system demonstrated improved operational efficiency in care facilities, reducing the likelihood of health crises such as falls or cardiovascular events. Additionally, IoT solutions such as fall detection systems have been widely adopted in care facilities to enhance resident safety. Studies have shown that IoT-enabled wearable sensors effectively track motor activity and detect falls in real time, thus preventing serious injuries by alerting staff immediately [27,28,57]. These technologies not only improve patient outcomes but also alleviate the burden on caregivers by providing constant, automated oversight. By promoting greater independence among older residents, IoT devices allow them to age in place, contributing to both their safety and quality of life.

Edge Intelligence: Enhancing Real-Time Decision-Making in LTC

EI plays a pivotal role in enhancing care for older adults by enabling the processing of data directly at the source, thus reducing latency and ensuring real-time decision-making. EI allows care facilities to make immediate adjustments to care plans based on real-time data, optimizing both AI and IoT applications in care for older adults. For example, in an Italian study, a radar smart sensor utilizing edge computing successfully detected falls and monitored vital signs in older adults, providing immediate feedback to caregivers [57]. This approach minimizes the delay associated with cloud-based systems, ensuring that critical health decisions are made without the need for external data processing. Another study integrated EI into an IoT-based monitoring system, which facilitated the detection of behavioral changes in older adult residents, enabling care providers to make real-time interventions to prevent health deterioration [77]. By processing data locally, EI enhances the reliability and responsiveness of care systems, reducing the risk of system failures or connectivity issues. Moreover, EI ensures the security of sensitive health data by limiting the amount of personal information transmitted over networks, thus addressing privacy concerns that are prevalent in older adult care technologies.

Summary and Suggestions for Future Research

AI, IoT, and EI collectively form a powerful ecosystem in LTC settings, addressing different aspects of care for older adults. Based on our results, we propose the following suggestions for future research directions. **First**, the co-occurrence network shows that this research topic is globally recognized, with various countries leading relevant studies. We recommend continuing to enhance international collaboration. The thematic map reveals emerging trends in “rehabilitation,” “stroke,” “robotics,” “caregiver,” “mHealth,” and “digital health.” Future research should focus on integrating these themes into broader studies to address the evolving needs in care for older adults. In addition, keyword analysis shows increasing interest in areas such as “older adults,” “machine learning,” and “internet of things.” The correlation of keywords between bibliometric analysis and Google Trends, including “long-term care,” “machine learning,” “dementia,” and “smart home,” reflects alignment between academic and public interest. Thus, we suggest that future research papers could enhance their discoverability by incorporating these keywords, ensuring they resonate with both scholarly communities and the general public. Finally, content analysis shows the potential of assistive technologies, particularly social robots such as PARO, in improving mood and reducing agitation among patients with dementia. Future studies should address limitations such as small sample sizes and short study durations by conducting larger, multicenter trials that examine long-term impacts. Expanding research to cover a wider range of health conditions, such as frailty or depression, would provide a more holistic understanding of how AI, IoT, and EI technologies can transform older adult care.

Comparison With Previous Studies

Previous studies have consistently demonstrated the role of AI and IoT in care for older adults, especially in addressing cognitive impairments such as dementia. Wang et al’s [15] bibliometric analysis identified the United States as a leading contributor to AI applications in care for older adults, particularly in areas such as Alzheimer disease and dementia. Our study corroborates Wang et al’s findings by also highlighting a strong focus on AI and IoT for these conditions; however, we provide a more comprehensive analysis, evaluating over 6000 publications compared with Wang et al’s 230. Similarly, the study by Loveys et al [88] explored the acceptance of AI-augmented social robots and sensors, noting their varied effectiveness across care settings. Our findings align with Loveys et al’s observations, especially in high-income facilities where technologies such as PARO show greater efficacy. Additionally, Lukkien et al’s [89] study emphasized the need for user-centered AI and the importance of contextual sensitivity in older adult care applications—themes that resonate with our focus on personalized interventions. Similar to Maugeri et al [14], we use both bibliometric and Google Trends data to analyze trends in health technologies. However, our study differs by incorporating content analysis, which offers deeper insights by examining the most-cited papers to explore study designs, popular topics, and potential limitations for future research.

XSL•FO
RenderX

Regarding EI, Zhou et al [10] explored EI's potential to enhance real-time decision-making through edge computing, particularly in environments that generate vast amounts of data, such as those supported by IoT devices. Reis [90] emphasized EI's role in optimizing operational efficiency in health care, particularly by processing data closer to the source to ensure timely care interventions. However, our study reveals that EI is not a standalone concept in LTC for older people. Instead, it functions as part of an integrated system, combining AI and IoT to support health care providers in making real-time decisions and offering personalized, technology-assisted interventions for care recipients. This fusion of technologies enables a more efficient and responsive older adult care environment, allowing health care providers to deliver tailored interventions based on immediate data, while providing older adults with customized assistive devices and services. By demonstrating how EI complements AI and IoT in this context, our study contributes to a deeper understanding of how these technologies work together to enhance care quality and personalization in LTC settings.

Limitations

This study has several limitations. First, our analysis was confined to research from the past 10 years, even though the topic has been explored for over 3 decades. This choice was intentional to ensure the inclusion of the most recent and relevant data reflecting current trends. However, it may have excluded some foundational studies that provide historical context. Second, we only included English-language publications, which could introduce language bias and affect the global representativeness of our findings. Additionally, while content analysis was used to address some limitations of quantitative methods, it was restricted to the top 1% (n=64) of highly cited papers, potentially overlooking other important research. Future studies could benefit from adopting methods such as the Delphi study, as suggested by one of the reviewers, to gain deeper insights from expert opinions. This would further strengthen the qualitative depth of the research and make the study even more robust.

Conclusions

The integration of AI, IoT, and EI is driving the evolution of an interconnected ecosystem in LTC for older adults. Our study underscores the significance of this field across multiple regions, emphasizing the need to promote diversity and international collaboration to enhance regional and institutional cooperation. Key themes such as "dementia care," "machine learning," and "wearable health monitoring" are emerging as focal areas of growing research interest. Keywords such as "long-term care," "older adults," "machine learning," "dementia," "smart home," and "caregiver" demonstrate a strong correlation with the public interest, suggesting that incorporating these terms can increase research visibility. Despite the valuable insights into technology acceptance, care quality improvements, and outcomes from interventions, frequently cited studies often face limitations such as small sample sizes, short study durations, and a narrow focus on dementia care, restricting the generalizability of their findings. Future research should broaden the scope to include a wider range of health conditions, conduct long-term trials, and strengthen international cooperation. This approach will significantly enhance the scalability and real-world impact of these technologies globally.

Acknowledgments

This research was funded by the National Health Research Institutes in Taiwan (grant PH-111-GP-07).

Data Availability

All data generated or analyzed during this study are included in this published paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional analysis and results.
[DOCX File , 2622 KB - jmir_v27i1e56692_app1.docx]

References

1. Rahman MM, Rosenberg M, Flores G, Parsell N, Akter S, Alam MA, et al. A systematic review and meta-analysis of unmet needs for healthcare and long-term care among older people. Health Econ Rev 2022 Dec 09;12(1):60 [FREE Full text] [doi: 10.1186/s13561-022-00398-4] [Medline: 36482044]
2. Titley HK, Young S, Savage A, Thorne T, Spiers J, Estabrooks CA. Cracks in the foundation: the experience of care aides in long-term care homes during the COVID-19 pandemic. J Am Geriatr Soc 2023 Jan;71(1):198-205 [FREE Full text] [doi: 10.1111/jgs.18024] [Medline: 36082802]
3. Thwaites C, McKercher J, Fetherstonhaugh D, Blackberry I, Gilmartin-Thomas J, Taylor N, et al. Factors impacting retention of aged care workers: a systematic review. Healthcare (Basel) 2023 Nov 21;11(23):3008 [FREE Full text] [doi: 10.3390/healthcare11233008] [Medline: 38063576]

XSL•FO
RenderX

4. Hacking C, Verbeek H, Hamers JPH, Aarts S. Comparing text mining and manual coding methods: analysing interview data on quality of care in long-term care for older adults. PLoS One 2023;18(11):e0292578 [FREE Full text] [doi: 10.1371/journal.pone.0292578] [Medline: 37939098]
5. Rostad HM, Skinner MS, Wentzel-Larsen T, Hellesø R, Sogstad MKR. Modes and models of care delivery in municipal long-term care services: a cross-sectional study from Norway. BMC Health Serv Res 2023 Jul 31;23(1):813 [FREE Full text] [doi: 10.1186/s12913-023-09750-8] [Medline: 37525166]
6. Sapci AH, Sapci HA. Innovative assisted living tools, remote monitoring technologies, artificial intelligence-driven solutions, and robotic systems for aging societies: systematic review. JMIR Aging 2019 Nov 29;2(2):e15429 [FREE Full text] [doi: 10.2196/15429] [Medline: 31782740]
7. Baig MM, Afifi S, GholamHosseini H, Mirza F. A systematic review of wearable sensors and IoT-based monitoring applications for older adults - a focus on ageing population and independent living. J Med Syst 2019 Jun 15;43(8):233. [doi: 10.1007/s10916-019-1365-7] [Medline: 31203472]
8. Rangraze I, Khan S. A systematic review on the efficacy of artificial intelligence in geriatric healthcare: a critical analysis of current literature. Research Square 2023 Jun 06:e1 [FREE Full text] [doi: 10.21203/rs.3.rs-2878974/v1]
9. Amin SU, Hossain MS. Edge Intelligence and Internet of Things in Healthcare: A Survey. IEEE Access 2021;9:45-59. [doi: 10.1109/access.2020.3045115]
10. Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J. Edge intelligence: paving the last mile of artificial intelligence with edge computing. Proc IEEE 2019 Aug;107(8):1738-1762. [doi: 10.1109/jproc.2019.2918951]
11. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. Journal of Business Research 2021 Sep;133:285-296. [doi: 10.1016/j.jbusres.2021.04.070]
12. Jun S, Yoo HS, Choi S. Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. Technological Forecasting and Social Change 2018 May;130:69-87. [doi: 10.1016/j.techfore.2017.11.009]
13. Harwood TG, Garry T. An overview of content analysis. Mark Rev 2003;3(4):479-498. [doi: 10.1362/146934703771910080]
14. Maugeri A, Barchitta M, Basile G, Agodi A. Public and research interest in telemedicine from 2017 to 2022: infodemiology study of Google Trends data and bibliometric analysis of scientific literature. J Med Internet Res 2024 May 16;26:e50088 [FREE Full text] [doi: 10.2196/50088] [Medline: 38753427]
15. Wang J, Liang Y, Cao S, Cai P, Fan Y. Application of artificial intelligence in geriatric care: bibliometric analysis. J Med Internet Res 2023 Jun 23;25:e46014 [FREE Full text] [doi: 10.2196/46014] [Medline: 37351923]
16. Voleti M, Bhat P. IoT and edge computing in health care: a bibliometric analysis. IJIRT 2022;9(6):582-587 [FREE Full text]
17. Ziwei H, Dongni Z, Man Z, Yixin D, Shuanghui Z, Chao Y, et al. The applications of internet of things in smart healthcare sectors: a bibliometric and deep study. Heliyon 2024 Feb 15;10(3):e25392 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e25392] [Medline: 38356528]
18. Aria M, Cuccurullo C. bibliometrix : An R-tool for comprehensive science mapping analysisa. Journal of Informetrics 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
19. Rogers G, Szomszor M, Adams J. Sample size in bibliometric analysis. Scientometrics 2020 Jul 31;125(1):777-794. [doi: 10.1007/s11192-020-03647-7]
20. Janik A, Ryszko A, Szafraniec M. Scientific Landscape of Smart and Sustainable Cities Literature: A Bibliometric Analysis. Sustainability 2020 Jan 21;12(3):779. [doi: 10.3390/su12030779]
21. Ioannidis JP, Klavans R, Boyack KW. Correction: Multiple citation indicators and their composite across scientific disciplines. PLoS Biol 2016 Aug 22;14(8):e1002548. [doi: 10.1371/journal.pbio.1002548] [Medline: 27548059]
22. Brunson JC, Laubenbacher RC. Applications of network analysis to routinely collected health care data: a systematic review. J Am Med Inform Assoc 2018 Feb 01;25(2):210-221 [FREE Full text] [doi: 10.1093/jamia/ocx052] [Medline: 29025116]
23. Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association 1968 Dec;63(324):1379-1389. [doi: 10.1080/01621459.1968.10480934]
24. Esposito A, Fortunati L, Lugano G. Modeling emotion, behavior and context in socially believable robots and ICT interfaces. Cogn Comput 2014 Oct 14;6(4):623-627. [doi: 10.1007/s12559-014-9309-5]
25. Lattanzio F, Abbatecola AM, Bevilacqua R, Chiatti C, Corsonello A, Rossi L, et al. Advanced technology care innovation for older people in Italy: necessity and opportunity to promote health and wellbeing. J Am Med Dir Assoc 2014 Jul;15(7):457-466. [doi: 10.1016/j.jamda.2014.04.003] [Medline: 24836715]
26. Moyle W, Jones C, Cooke M, O'Dwyer S, Sung B, Drummond S. Connecting the person with dementia and family: a feasibility study of a telepresence robot. BMC Geriatr 2014 Jan 24;14:7 [FREE Full text] [doi: 10.1186/1471-2318-14-7] [Medline: 24456417]
27. Schwenk M, Hauer K, Zieschang T, Englert S, Mohler J, Najafi B. Sensor-derived physical activity parameters can predict future falls in people with dementia. Gerontology 2014;60(6):483-492 [FREE Full text] [doi: 10.1159/000363136] [Medline: 25171300]
28. Yamada Y, Vlachova M, Richter T, Finne-Soveri H, Gindin J, van der Roest H, et al. Prevalence and correlates of hearing and visual impairments in European nursing homes: results from the SHELTER study. J Am Med Dir Assoc 2014 Oct;15(10):738-743. [doi: 10.1016/j.jamda.2014.05.012] [Medline: 24984787]

XSL•FO
RenderX

29. Yu R, Leung J, Woo J. Incremental predictive value of sarcopenia for incident fracture in an elderly Chinese cohort: results from the Osteoporotic Fractures in Men (MrOs) Study. J Am Med Dir Assoc 2014 Aug;15(8):551-558. [doi: 10.1016/j.jamda.2014.02.005] [Medline: 24703927]
30. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Effectiveness of robot Paro in intramural psychogeriatric care: a multicenter quasi-experimental study. J Am Med Dir Assoc 2015 Nov 01;16(11):946-950. [doi: 10.1016/j.jamda.2015.05.007] [Medline: 26115817]
31. Chetty G, White M, Akther F. Smart phone based data mining for human activity recognition. Procedia Computer Science 2015;46:1181-1187. [doi: 10.1016/j.procs.2015.01.031]
32. Claes V, Devriendt E, Tournoy J, Milisen K. Attitudes and perceptions of adults of 60 years and older towards in-home monitoring of the activities of daily living with contactless sensors: an explorative study. Int J Nurs Stud 2015 Jan;52(1):134-148. [doi: 10.1016/j.ijnurstu.2014.05.010] [Medline: 24951084]
33. Gustafsson C, Svanberg C, Müllersdorf M. Using a robotic cat in dementia care: a pilot study. J Gerontol Nurs 2015 Oct;41(10):46-56. [doi: 10.3928/00989134-20150806-44] [Medline: 26488255]
34. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
35. Lloret J, Canovas A, Sendra S, Parra L. A smart communication architecture for ambient assisted living. IEEE Commun Mag 2015 Jan;53(1):26-33. [doi: 10.1109/mcom.2015.7010512]
36. Moyle W, Beattie E, Draper B, Shum D, Thalib L, Jones C, et al. Effect of an interactive therapeutic robotic animal on engagement, mood states, agitation and psychotropic drug use in people with dementia: a cluster-randomised controlled trial protocol. BMJ Open 2015 Aug 12;5(8):e009097 [FREE Full text] [doi: 10.1136/bmjopen-2015-009097] [Medline: 26270953]
37. Moyle W, Jones C, Sung B, Bramble M, O'Dwyer S, Blumenstein M, et al. What effect does an animal robot called CuDDler have on the engagement and emotional response of older people with dementia? A pilot feasibility study. Int J of Soc Robotics 2015 Oct 24;8(1):145-156. [doi: 10.1007/s12369-015-0326-7]
38. Pfadenhauer M, Dukat C. Robot caregiver or robot-supported caregiving? Int J of Soc Robotics 2015 Jan 30;7(3):393-406. [doi: 10.1007/s12369-015-0284-0]
39. Pino M, Boulay M, Jouen F, Rigaud A. "Are we ready for robots that care for us?" Attitudes and opinions of older adults toward socially assistive robots. Front Aging Neurosci 2015;7:141 [FREE Full text] [doi: 10.3389/fnagi.2015.00141] [Medline: 26257646]
40. Pouyet V, Cuvelier G, Benattar L, Giboreau A. Influence of flavour enhancement on food liking and consumption in older adults with poor, moderate or high cognitive status. Food Quality and Preference 2015 Sep;44:119-129. [doi: 10.1016/j.foodqual.2015.04.014]
41. Robinson H, MacDonald B, Broadbent E. Physiological effects of a companion robot on blood pressure of older people in residential care facility: a pilot study. Australas J Ageing 2015 Mar;34(1):27-32. [doi: 10.1111/ajag.12099] [Medline: 24373064]
42. Shiomi M, Iio T, Kamei K, Sharma C, Hagita N. Effectiveness of social behaviors for autonomous wheelchair robot to support elderly people in Japan. PLoS One 2015;10(5):e0128031 [FREE Full text] [doi: 10.1371/journal.pone.0128031] [Medline: 25993038]
43. Yamada Y, Denkinger MD, Onder G, Finne-Soveri H, van der Roest H, Vlachova M, et al. Impact of dual sensory impairment on onset of behavioral symptoms in European nursing homes: results from the Services and Health for Elderly in Long-Term Care study. J Am Med Dir Assoc 2015 Apr;16(4):329-333. [doi: 10.1016/j.jamda.2014.11.006] [Medline: 25523284]
44. Berridge C. Breathing room in monitored space: the impact of passive monitoring technology on privacy in independent living. Gerontologist 2016 Oct;56(5):807-816. [doi: 10.1093/geront/gnv034] [Medline: 26035900]
45. Broadbent E, Kerse N, Peri K, Robinson H, Jayawardena C, Kuo T, et al. Benefits and problems of health-care robots in aged care settings: a comparison trial. Australas J Ageing 2016 Mar 13;35(1):23-29. [doi: 10.1111/ajag.12190] [Medline: 26364706]
46. Jøranson N, Pedersen I, Rokstad AMM, Ihlebaek C. Change in quality of life in older people with dementia participating in Paro-activity: a cluster-randomized controlled trial. J Adv Nurs 2016 Dec;72(12):3020-3033. [doi: 10.1111/jan.13076] [Medline: 27434512]
47. Koceski S, Koceska N. Evaluation of an assistive telepresence robot for elderly healthcare. J Med Syst 2016 May;40(5):121. [doi: 10.1007/s10916-016-0481-x] [Medline: 27037685]
48. Lazarou I, Karakostas A, Stavropoulos TG, Tsompanidis T, Meditskos G, Kompatsiaris I, et al. A novel and intelligent home monitoring system for care support of elders with cognitive impairment. J Alzheimers Dis 2016 Oct 18;54(4):1561-1591. [doi: 10.3233/JAD-160348] [Medline: 27636843]
49. Martinikorena I, Martínez-Ramírez A, Gómez M, Lecumberri P, Casas-Herrero A, Cadore EL, et al. Gait variability related to muscle quality and muscle power output in frail nonagenarian older adults. J Am Med Dir Assoc 2016 Feb;17(2):162-167. [doi: 10.1016/j.jamda.2015.09.015] [Medline: 26577625]

XSL•FO
RenderX

50. Mitoku K, Masaki N, Ogata Y, Okamoto K. Vision and hearing impairments, cognitive impairment and mortality among long-term care recipients: a population-based cohort study. BMC Geriatr 2016 May 27;16:112 [FREE Full text] [doi: 10.1186/s12877-016-0286-2] [Medline: 27233777]
51. Nikoloudakis Y, Panagiotakis S, Markakis E, Pallis E, Mastorakis G, Mavromoustakis CX, et al. A fog-based emergency system for smart enhanced living environments. IEEE Cloud Comput 2016 Nov;3(6):54-62. [doi: 10.1109/mcc.2016.118]
52. Thodberg K, Sørensen LU, Christensen JW, Poulsen PH, Houbak B, Damgaard V, et al. Therapeutic effects of dog visits in nursing homes for the elderly. Psychogeriatrics 2016 Sep;16(5):289-297. [doi: 10.1111/psyg.12159] [Medline: 26510632]
53. Thodberg K, Sørensen LU, Videbech PB, Poulsen PH, Houbak B, Damgaard V, et al. Behavioral responses of nursing home residents to visits from a person with a dog, a robot seal or a toy cat. Anthrozoös 2016 Mar 08;29(1):107-121. [doi: 10.1080/08927936.2015.1089011]
54. Atee M, Hoti K, Parsons R, Hughes JD. Pain assessment in dementia: evaluation of a point-of-care technological solution. J Alzheimers Dis 2017;60(1):137-150. [doi: 10.3233/JAD-170375] [Medline: 28800333]
55. Berridge C. Active subjects of passive monitoring: responses to a passive monitoring system in low-income independent living. Ageing Soc 2017 Mar;37(3):537-560 [FREE Full text] [doi: 10.1017/S0144686X15001269] [Medline: 28239211]
56. Darragh M, Ahn HS, MacDonald B, Liang A, Peri K, Kerse N, et al. Homecare robots to improve health and well-being in mild cognitive impairment and early stage dementia: results from a scoping study. J Am Med Dir Assoc 2017 Dec 01;18(12):1099.e1-1099.e4. [doi: 10.1016/j.jamda.2017.08.019] [Medline: 28974463]
57. Diraco G, Leone A, Siciliano P. A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications. Biosensors (Basel) 2017 Nov 24;7(4):55 [FREE Full text] [doi: 10.3390/bios7040055] [Medline: 29186786]
58. Korchut A, Szklener S, Abdelnour C, Tantinya N, Hernández-Farigola J, Ribes JC, et al. Challenges for service robots-requirements of elderly adults with cognitive impairments. Front Neurol 2017;8:228 [FREE Full text] [doi: 10.3389/fneur.2017.00228] [Medline: 28620342]
59. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc 2017 Oct 01;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]
60. Machado A, Maran V, Augustin I, Wives LK, de Oliveira JPM. Reactive, proactive, and extensible situation-awareness in ambient assisted living. Expert Systems with Applications 2017 Jun;76:21-35. [doi: 10.1016/j.eswa.2017.01.033]
61. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ERA, Shum DKH, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Dir Assoc 2017 Sep 01;18(9):766-773 [FREE Full text] [doi: 10.1016/j.jamda.2017.03.018] [Medline: 28780395]
62. Pigini L, Bovi G, Panzarino C, Gower V, Ferratini M, Andreoni G, et al. Pilot test of a new personal health system integrating environmental and wearable sensors for telemonitoring and care of elderly people at home (SMARTA Project). Gerontology 2017;63(3):281-286. [doi: 10.1159/000455168] [Medline: 28099965]
63. Rafferty J, Nugent CD, Liu J, Chen L. From activity recognition to intention recognition for assisted living within smart homes. IEEE Trans Human-Mach Syst 2017 Jun;47(3):368-379. [doi: 10.1109/thms.2016.2641388]
64. Arends J, Thijs RD, Gutter T, Ungureanu C, Cluitmans P, Van Dijk J, et al. Multimodal nocturnal seizure detection in a residential care setting. Neurology 2018 Nov 20;91(21):e2010-e2019. [doi: 10.1212/wnl.0000000000006545]
65. Cortellessa G, Fracasso F, Sorrentino A, Orlandini A, Bernardi G, Coraci L, et al. ROBIN, a telepresence robot to support older users monitoring and social inclusion: development and evaluation. Telemed J E Health 2018 Feb;24(2):145-154. [doi: 10.1089/tmj.2016.0258] [Medline: 28771398]
66. Erebak S, Turgut T. Caregivers' attitudes toward potential robot coworkers in elder care. Cogn Tech Work 2018 Jul 24;21(2):327-336. [doi: 10.1007/s10111-018-0512-0]
67. Jones C, Moyle W, Murfield J, Draper B, Shum D, Beattie E, et al. Does cognitive impairment and agitation in dementia influence intervention effectiveness? Findings from a cluster-randomized-controlled trial with the therapeutic robot, PARO. J Am Med Dir Assoc 2018 Jul;19(7):623-626. [doi: 10.1016/j.jamda.2018.02.014] [Medline: 29656838]
68. Jones C, Sung B, Moyle W. Engagement of a person with Dementia Scale: establishing content validity and psychometric properties. J Adv Nurs 2018 May 17;74(9):2227-2240. [doi: 10.1111/jan.13717] [Medline: 29772602]
69. Lehoux P, Grimard D. When robots care: public deliberations on how technology and humans may support independent living for older adults. Soc Sci Med 2018 Aug;211:330-337. [doi: 10.1016/j.socscimed.2018.06.038] [Medline: 30015242]
70. Meritam P, Ryvlin P, Beniczky S. User-based evaluation of applicability and usability of a wearable accelerometer device for detecting bilateral tonic-clonic seizures: a field study. Epilepsia 2018 Jun;59 Suppl 1:48-52. [doi: 10.1111/epi.14051] [Medline: 29873828]
71. Moyle W, Bramble M, Jones C, Murfield J. Care staff perceptions of a social robot called Paro and a look-alike Plush Toy: a descriptive qualitative approach. Aging Ment Health 2018 Mar;22(3):330-335. [doi: 10.1080/13607863.2016.1262820] [Medline: 27967207]
72. Moyle W, Jones C, Murfield J, Thalib L, Beattie E, Shum D, et al. Effect of a robotic seal on the motor activity and sleep patterns of older people with dementia, as measured by wearable technology: a cluster-randomised controlled trial. Maturitas 2018 Apr;110:10-17. [doi: 10.1016/j.maturitas.2018.01.007]

XSL•FO
RenderX

73. Mulero R, Almeida A, Azkune G, Abril-Jimenez P, Arredondo Waldmeyer MT, Paramo Castrillo M, et al. An IoT-aware approach for elderly-friendly cities. IEEE Access 2018;6:7941-7957. [doi: 10.1109/access.2018.2800161]
74. Rantanen T, Lehto P, Vuorinen P, Coco K. The adoption of care robots in home care-a survey on the attitudes of Finnish home care personnel. J Clin Nurs 2018 May;27(9-10):1846-1859. [doi: 10.1111/jocn.14355] [Medline: 29575204]
75. Rantanen T, Lehto P, Vuorinen P, Coco K. Attitudes towards care robots among Finnish home care personnel - a comparison of two approaches. Scand J Caring Sci 2018 Jun;32(2):772-782. [doi: 10.1111/scs.12508] [Medline: 28833309]
76. Wittich W, Höbler F, Jarry J, McGilton KS. Recommendations for successful sensory screening in older adults with dementia in long-term care: a qualitative environmental scan of Canadian specialists. BMJ Open 2018 Jan 26;8(1):e019451 [FREE Full text] [doi: 10.1136/bmjopen-2017-019451] [Medline: 29374673]
77. Al-khafajiy M, Baker T, Chalmers C, Asim M, Kolivand H, Fahim M, et al. Remote health monitoring of elderly through wearable sensors. Multimed Tools Appl 2019 Jan 24;78(17):24681-24706. [doi: 10.1007/s11042-018-7134-7]
78. D'Onofrio G, Fiorini L, Hoshino H, Matsumori A, Okabe Y, Tsukamoto M, et al. Assistive robots for socialization in elderly people: results pertaining to the needs of the users. Aging Clin Exp Res 2018 Dec 17;31(9):1313-1329. [doi: 10.1007/s40520-018-1073-z]
79. Daniele K, Marcucci M, Cattaneo C, Borghese NA, Zannini L. How prefrail older people living alone perceive information and communications technology and what they would ask a robot for: qualitative study. J Med Internet Res 2019 Aug 06;21(8):e13228 [FREE Full text] [doi: 10.2196/13228] [Medline: 31389341]
80. Decerle J, Grunder O, Hajjam El Hassani A, Barakat O. A memetic algorithm for multi-objective optimization of the home health care problem. Swarm and Evolutionary Computation 2019 Feb;44:712-727. [doi: 10.1016/j.swevo.2018.08.014]
81. Moyle W, Bramble M, Jones CJ, Murfield JE. "She Had a Smile on Her Face as Wide as the Great Australian Bite": A Qualitative Examination of Family Perceptions of a Therapeutic Robot and a Plush Toy. Gerontologist 2019 Dec 09;59(1):177-185. [doi: 10.1093/geront/gnx180] [Medline: 29165558]
82. Moyle W, Jones C, Murfield J, Thalib L, Beattie E, Shum D, et al. Using a therapeutic companion robot for dementia symptoms in long-term care: reflections from a cluster-RCT. Aging Ment Health 2019 Mar;23(3):329-336. [doi: 10.1080/13607863.2017.1421617] [Medline: 29282989]
83. Tang V, Choy K, Ho G, Lam H, Tsang Y. An IoMT-based geriatric care management system for achieving smart health in nursing homes. IMDS 2019 Sep 09;119(8):1819-1840. [doi: 10.1108/imds-01-2019-0024]
84. Wright J. Robots vs migrants? Reconfiguring the future of Japanese institutional eldercare. Critical Asian Studies 2019 May 10;51(3):331-354. [doi: 10.1080/14672715.2019.1612765]
85. Ho A. Are we ready for artificial intelligence health monitoring in elder care? BMC Geriatr 2020 Sep 21;20(1):358 [FREE Full text] [doi: 10.1186/s12877-020-01764-9] [Medline: 32957946]
86. Łukasik S, Tobis S, Kropińska S, Suwalska A. Role of assistive robots in the care of older people: survey study among medical and nursing students. J Med Internet Res 2020 Aug 12;22(8):e18003 [FREE Full text] [doi: 10.2196/18003] [Medline: 32784187]
87. Khosla R, Chu M, Khaksar SMS, Nguyen K, Nishida T. Engagement and experience of older people with socially assistive robots in home care. Assist Technol 2021 Mar 04;33(2):57-71. [doi: 10.1080/10400435.2019.1588805] [Medline: 31063044]
88. Loveys K, Prina M, Axford C, Domènec, Weng W, Broadbent E, et al. Artificial intelligence for older people receiving long-term care: a systematic review of acceptability and effectiveness studies. Lancet Healthy Longev 2022 Apr;3(4):e286-e297 [FREE Full text] [doi: 10.1016/S2666-7568(22)00034-4] [Medline: 35515814]
89. Lukkien DRM, Nap HH, Buimer HP, Peine A, Boon WPC, Ket JCF, et al. Toward responsible artificial intelligence in long-term care: a scoping review on practical approaches. Gerontologist 2023 Jan 24;63(1):155-168 [FREE Full text] [doi: 10.1093/geront/gnab180] [Medline: 34871399]
90. Reis J. Edge intelligence in enhancing last-mile delivery logistics. PrePrints. Preprint posted online July 23, 2024 [FREE Full text] [doi: 10.20944/preprints202407.1777.v1]

Abbreviations

AI: artificial intelligence
EI: edge intelligence
IoT: Internet of Things
LTC: long-term care
TGC: total global citation
TLC: total local citation

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 24.01.24; peer-reviewed by S Groeneveld, M Vinayak, M Rizvi, J Reis; comments to author 12.08.24; revised version received 21.09.24; accepted 20.01.25; published 04.03.25.

Please cite as:
Chien SC, Yen CM, Chang YH, Chen YE, Liu CC, Hsiao YP, Yang PY, Lin HM, Yang TE, Lu XH, Wu IC, Hsu CC, Chiou HY, Chung RH
Use of Artificial Intelligence, Internet of Things, and Edge Intelligence in Long-Term Care for Older People: Comprehensive Analysis Through Bibliometric, Google Trends, and Content Analysis
J Med Internet Res 2025;27:e56692
URL: https://www.jmir.org/2025/1/e56692
doi:10.2196/56692
PMID:40053718

©Shuo-Chen Chien, Chia-Ming Yen, Yu-Hung Chang, Ying-Erh Chen, Chia-Chun Liu, Yu-Ping Hsiao, Ping-Yen Yang, Hong-Ming Lin, Tsung-En Yang, Xing-Hua Lu, I-Chien Wu, Chih-Cheng Hsu, Hung-Yi Chiou, Ren-Hua Chung. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Metaverse Clinic for Pregnant Women With Subclinical Hypothyroidism: Prospective Randomized Study

Yuanyuan Zheng[1,2*], MSc; Yizhen Chen[1*], MSc; Yan Chen[1,3*], MSc; Liang Lin[1,4], MD, PhD; Ting Xue[1,5], MD, PhD; Chuhui Chen[1,3], MSc; Junping Wen[1,3], MD, PhD; Wei Lin[1,3], MD, PhD; Gang Chen[1,3], MD, PhD

[1]Shengli Clinical Medical College, Fujian Medical University, Fuzhou, China

[2]Department of Geriatric Medicine, Fujian Key Laboratory of Geriatrics Diseases, Fujian Provincial Center for Geriatrics, Fuzhou University Affiliated Provincial Hospital, Fuzhou, China

[3]Department of Endocrinology, Fuzhou University Affiliated Provincial Hospital, Fuzhou, China

[4]Department of Obstetrics and Gynaecology, Fuzhou University Affiliated Provincial Hospital, Fuzhou, China

[5]Department of Health Management, Fuzhou University Affiliated Provincial Hospital, Fuzhou, China

[*]these authors contributed equally

Corresponding Author:
Gang Chen, MD, PhD
Department of Endocrinology
Fuzhou University Affiliated Provincial Hospital
No134 East Road
Fuzhou, 350001
China
Phone: 86 059188217521
Fax: 86 059177788954
Email: chengangfj@163.com

Abstract

Background: Health care is experiencing new opportunities in the emerging digital landscape. The metaverse, a shared virtual space, integrates technologies such as augmented reality, virtual reality, blockchain, and artificial intelligence. It allows users to interact with immersive digital worlds, connect with others, and explore unknowns. While the metaverse is gaining traction across various medical disciplines, its application in thyroid diseases remains unexplored. Subclinical hypothyroidism (SCH) is the most common thyroid disorder during pregnancy and is frequently associated with adverse pregnancy outcomes.

Objective: This study aims to evaluate the safety and effectiveness of a metaverse platform in managing SCH during pregnancy.

Methods: A randomized controlled trial was conducted at Fujian Provincial Hospital, China, from July 2022 to December 2023. A total of 60 pregnant women diagnosed with SCH were randomly assigned into two groups: the standard group (n=30) and the metaverse group (n=30). Both groups received levothyroxine sodium tablets. Additionally, participants in the metaverse group had access to the metaverse virtual medical consultations and metaverse-based medical games. The primary outcomes were adverse maternal and offspring outcomes, and the secondary outcomes included the neurobehavioral development of offspring and maternal psychological assessments.

Results: Of the 30 participants in each group, adverse maternal outcomes were observed in 43% (n=13) of the standard group and 37% (n=11) of the metaverse group (P=.60). The incidence of adverse offspring outcomes was 33% (n=10) in the standard group, compared to 7% (n=2) in the metaverse group (P=.01). The Gesell Development Scale did not show significant differences between the two groups. Notably, the metaverse group demonstrated significantly improved scores on the Self-Rating Depression Scale and the Self-Rating Anxiety Scale scores compared to the standard group (P<.001 and P=.001, respectively).

Conclusions: The use of metaverse technology significantly reduced the incidence of adverse offspring outcomes and positively impacted maternal mental health. Maternal adverse outcomes and offspring neurobehavioral development were comparable between the two groups.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300076803; https://www.chictr.org.cn/showproj.html?proj=205905

(J Med Internet Res 2025;27:e64634) doi:10.2196/64634

XSL•FO
RenderX

KEYWORDS

metaverse; subclinical hypothyroidism; adverse pregnancy outcomes; psychological health; neurobehavioral development

Introduction

The metaverse is a shared virtual space that seamlessly blends the physical and digital realities. It offers an immersive environment where users can interact in a computer-generated universe while gaining real-world benefits [1]. Built upon the foundations of mixed reality, artificial intelligence, and blockchain technology, the metaverse leverages these cutting-edge innovations to deliver dynamic, interactive, and secure user experiences [1-3]. Mixed reality, which includes both augmented reality and virtual reality (VR), bridges the continuum between physical and virtual experiences. Artificial intelligence enables adaptive and intelligent environments, while blockchain technology ensures secure transactions and ownership of digital assets, thereby supporting a resilient digital economy within the metaverse. In health care, the metaverse holds transformative possibilities [4,5]. Compared to traditional health care models, it enables secure and customizable therapeutic environments tailored to the unique needs and treatment goals of individual patients. By integrating the virtual and physical realms, it also allows for real-time monitoring and feedback [1,6]. By 2030, the global metaverse health care market is projected to reach US $98.3 billion, with the United States expected to lead this growth, followed by Europe and the Asia-Pacific region [2]. Early applications of metaverse technologies in health care have primarily focused on surgery, particularly in urology, neurosurgery, and surgical training [7-9]. Recently, its applications have expanded to areas such as cognitive rehabilitation, mental health, and chronic disease management [1,2,10]. Despite these promising developments, its potential role in the management of thyroid disorders remains unexplored.

Subclinical hypothyroidism (SCH), characterized by elevated thyrotropin levels with normal free thyroxine (FT4), is increasingly prevalent during pregnancy [11,12]. SCH is associated with adverse pregnancy outcomes, including spontaneous abortion, preeclampsia, and placental abruption [11,13]. Timely treatment with levothyroxine (LT4) sodium tablets is critical but is often hindered by limited awareness and delayed diagnosis [14].

Aligned with the Institute of Medicine's framework for high-quality health care—which prioritizes patient-centeredness, equity, safety, effectiveness, timeliness, and efficiency—our team has developed an innovative metaverse-based platform tailored for pregnant women with SCH [15]. This digital solution provides personalized care through comprehensive follow-up, starting from the antenatal period and extending to 3 months post partum. The platform aims to enhance SCH management by enabling timely and effective interventions, optimizing resource use, and promoting long-term maternal health. Our primary goal is to assess the safety and effectiveness of this metaverse-based platform in managing SCH during pregnancy. By addressing unmet clinical needs, the initiative seeks to reduce complications and improve the overall quality of prenatal and postnatal care.

Methods

Ethics Approval

This single-center, randomized, open-label trial was approved by the Fujian Provincial Hospital Institutional Review Board (K2022-06-016) and the research ethics board. The trial was registered under ChiCTR2300076803. Recruitment, treatment, and follow-up were conducted from July 2022 to December 2023.

Participants in this clinical trial received compensation for their time and involvement. Compensation was provided to recognize the participants' contributions and to help offset any expenses associated with their participation. Participants who attend the screening visit will receive compensation for their time and travel expenses. Participants will be compensated for each study visit attended. This includes time spent on medical assessments, laboratory tests, and any other required procedures (¥500-600, US $70-$80). Any out-of-pocket expenses incurred during participation, such as transportation or childcare, may be reimbursed upon submission of receipts. Upon successful completion of all study visits, participants will receive a completion bonus (¥1000, US $150).

Participants

Study participants were conveniently recruited from the Fujian Provincial Hospital outpatient practice and prescreened during appointments scheduled for SCH evaluations. Eligibility was confirmed through in-person assessments to ensure compliance with inclusion criteria. Written informed consent was obtained from all participants, and all data were anonymized or deidentified to protect confidentiality.

The inclusion criteria were as follows: pregnant women diagnosed with SCH, as defined by the 2017 American Thyroid Association Guidelines for the Diagnosis and Management of Thyroid Disease During Pregnancy and Postpartum (serum thyrotropin levels above the pregnancy-specific upper reference limit, or ≥4.0 mIU/L, with normal FT4 levels [16]), diagnosis and initiation of LT4 treatment before 8 weeks of gestation, singleton pregnancy, and age ≥18 years. The exclusion criteria were a history of thyroid disease; history of residence in areas with endemic goiter; use of LT4 before pregnancy or medications that affect thyroid function; history of nonthyroidal autoimmune diseases; preexisting cardiovascular, liver, or kidney diseases; or any prepregnancy condition that could interfere with the study. A total of 60 participants were enrolled in the trial.

Randomization and Masking

Participants were randomly assigned in a 1:1 ratio to either the standard or metaverse group using computer-generated randomization. Both groups were managed by the same medical team. Investigators and participants were not masked to treatment allocation. The research team, data manager, and statistician were aware of group assignments during the process of data collection. Data were securely stored in the Fujian

XSL•FO
RenderX

Provincial Hospital database, with the first author overseeing data accuracy and integrity. No placebo was used in the study. As predefined guidelines exist for LT4 use, the potential bias introduced by the nonblinded design on the primary outcome was considered minimal.

Procedures

The research team comprised a chief endocrinologist, an attending endocrinologist, an obstetrician-gynecologist, and a physician specializing in health management, all with extensive experience in thyroid disorder management. Eligible patients were prescribed oral LT4 (Euthyrox, Berlin-Chemie AG, Germany; H20110178, 50 μg × 100 tablets) based on thyroid function test results [16]. Initial dosages ranged from 25 to 50 μg and were adjusted according to pregnancy status [16]. Patients took one tablet daily on an empty stomach in the morning, with thyrotropin and FT4 levels monitored every 14 days for dose adjustments. Once thyrotropin levels were within 0.1-2.5 mIU/L, monitoring was conducted every 4 weeks, with doses adjusted to maintain trimester-specific target ranges [16]. Treatment continued consistently until delivery.

Participants in the standard group attended regular in-person follow-ups to monitor blood pressure, heart rate, weight, thyroid function, medication dosage, and other pregnancy-related indicators. Participants in the metaverse group only visited the hospital for enrollment, blood tests, and required pregnancy assessments. Follow-ups for the metaverse group were conducted through the platform, with medications delivered to their homes. The metaverse platform enabled secure communication between participants and investigators at any time, enhancing adherence to treatment protocols. Technical support was provided, with hardware promptly replaced if issues could not be resolved remotely. Participants used home monitoring devices under investigator guidance, with training provided for platform use. Weekly assessments ensured proficiency. Participants identified as “unfamiliar” received additional training until they could fully engage in medical activities on the platform. The platform supported uploading and recording thyroid function test results and other relevant health data. It also offered lifestyle guidance, psychological counseling, and educational games focused on endocrinology. Immersive experiences using Pico VR glasses (Fujian Huayu Education Technology Co Ltd, China) provided gamified learning and live discussions with physicians. The standard group participated in similar educational activities during weekly in-person sessions.

Adverse pregnancy outcomes were tracked through follow-ups and patient self-reports. Offspring neurobehavioral development was assessed using Gesell Development Scale (GDS) scores at 1 and 3 months of age, evaluating gross motor, fine motor, adaptability, language, and personal-social skills. Maternal mental health was assessed using the Self-Rating Anxiety Scale (SAS) and Self-Rating Depression Scale (SDS) at baseline, during the last trimester, and 6 weeks post partum. Psychological support was offered based on the SAS/SDS results, with interventions delivered via the metaverse platform for the metaverse group and through outpatient sessions for the standard group.

The intervention spanned from pregnancy through 3 months post partum, with participation continuing unless interrupted by withdrawal or loss to follow-up.

Outcomes

The primary outcomes of the study focused on assessing adverse maternal and offspring outcomes. Maternal adverse outcomes included gestational diabetes, gestational hypertension, premature rupture of membranes, amniotic fluid abnormalities, postpartum hemorrhage, and related complications [17]. Offspring adverse outcomes included fetal distress, abortion, stillbirth, preterm birth, neonatal asphyxia, and related complications [13,17]. Secondary outcomes focused on offspring neurobehavioral development and maternal mental health. Neurobehavioral development was considered normal if the developmental quotient on the GDS exceeded 85 [18]. Anxiety was defined as SAS scores greater than 50, and depression was defined as SDS scores greater than 53 [19]. Investigators conducted GDS assessments, while participants completed SAS and SDS questionnaires independently, which were collected immediately after completion.

Statistical Analysis

Based on a significance level of 5%, statistical power of 90%, and an anticipated 5% loss to follow-up, a total of 60 participants (30 per group) was calculated as sufficient for this randomized controlled trial.

Quantitative data were expressed as means and SDs, while categorical data were presented as percentages. Between-group comparisons of normally distributed variables used the independent samples 2-tailed *t* test or Mann-Whitney *U* test, while categorical variables were analyzed using Pearson χ^2 test or Fisher exact test. Analysis of covariance, with baseline thyroid function and initial psychological scores as covariates, was performed to assess postintervention differences in thyroid function and psychological outcomes between groups. Statistical analyses were conducted using SPSS 25.0 (IBM Corp), with a *P* value <.05 considered statistically significant.

Results

Overview

Between July 2022 and July 2023, 87 individuals were screened for eligibility, with 60 participants randomized equally into the standard group (n=30) and the metaverse group (n=30; Figure 1). By December 2023, all participants had completed the trial, including follow-up through 3 months post partum (Figure 2).

The median age of participants was 30 years in both groups, with age ranges of 20-40 years in the standard group and 23-38 years in the metaverse group. The median gestational age at delivery was 39 weeks, ranging from 23 to 40 weeks in the standard group and 36 to 40 weeks in the metaverse group. The median prepregnancy BMI was 22.8 (IQR 15.8-32.7) kg/m^2 in the standard group and 21.5 (IQR 17.6-30) in the metaverse group. There were 19 (63%) nulliparous participants in the standard group and 18 (60%) in the metaverse group. Three participants in each group had a history of obstetric abnormalities (Table 1).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram shows the number of participants. LT4: levothyroxine; SCH: subclinical hypothyroidism.

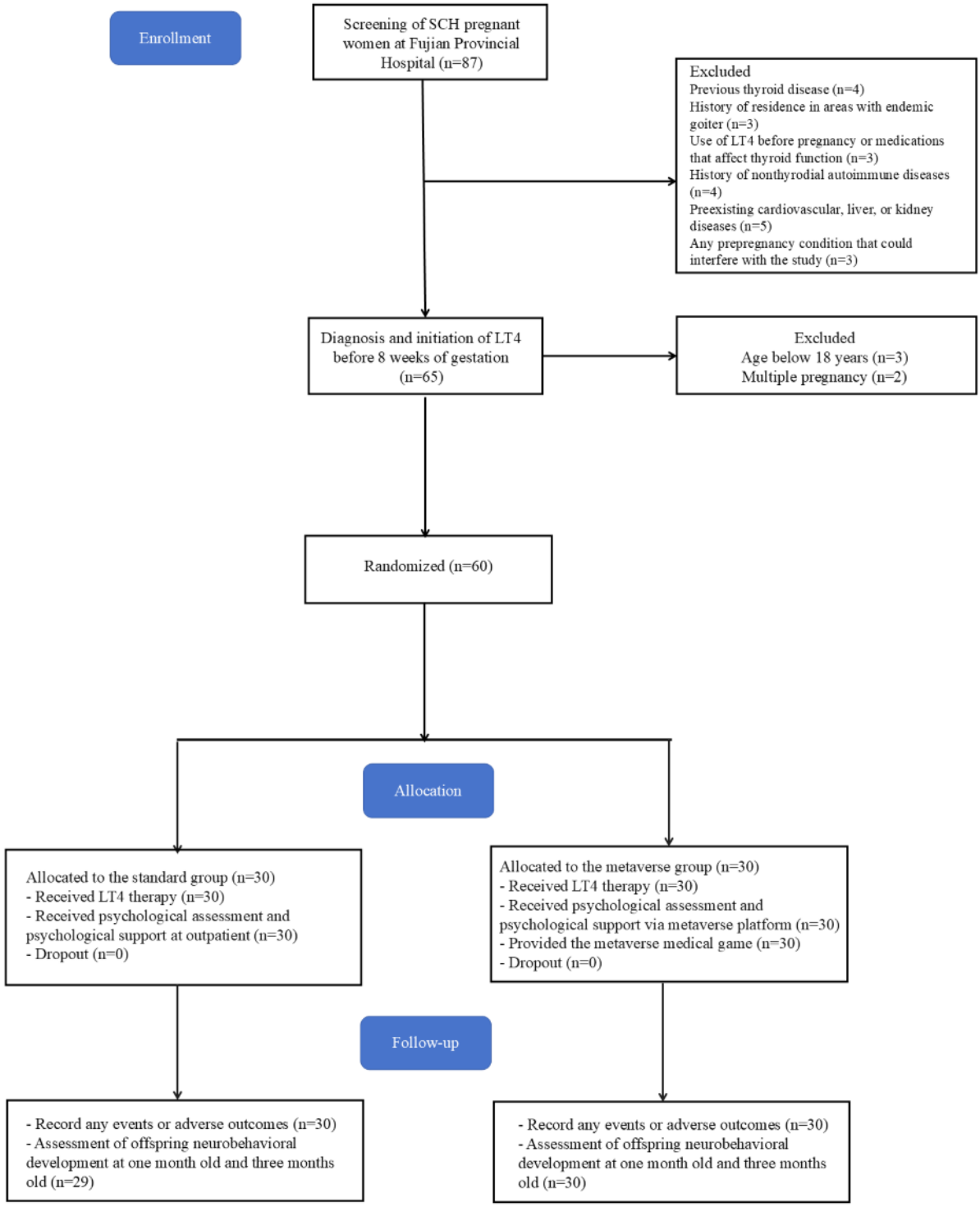

XSL•FO
RenderX

Figure 2. The timeline of the study and assessments. FT4: free thyroxine; GDS: Gesell Development Scale; LT4: levothyroxine; SAS: Self-Rating Anxiety Scale; SDS: Self-Rating Depression Scale; TPOAb: thyroid peroxidase antibodies; TSH: thyrotropin.

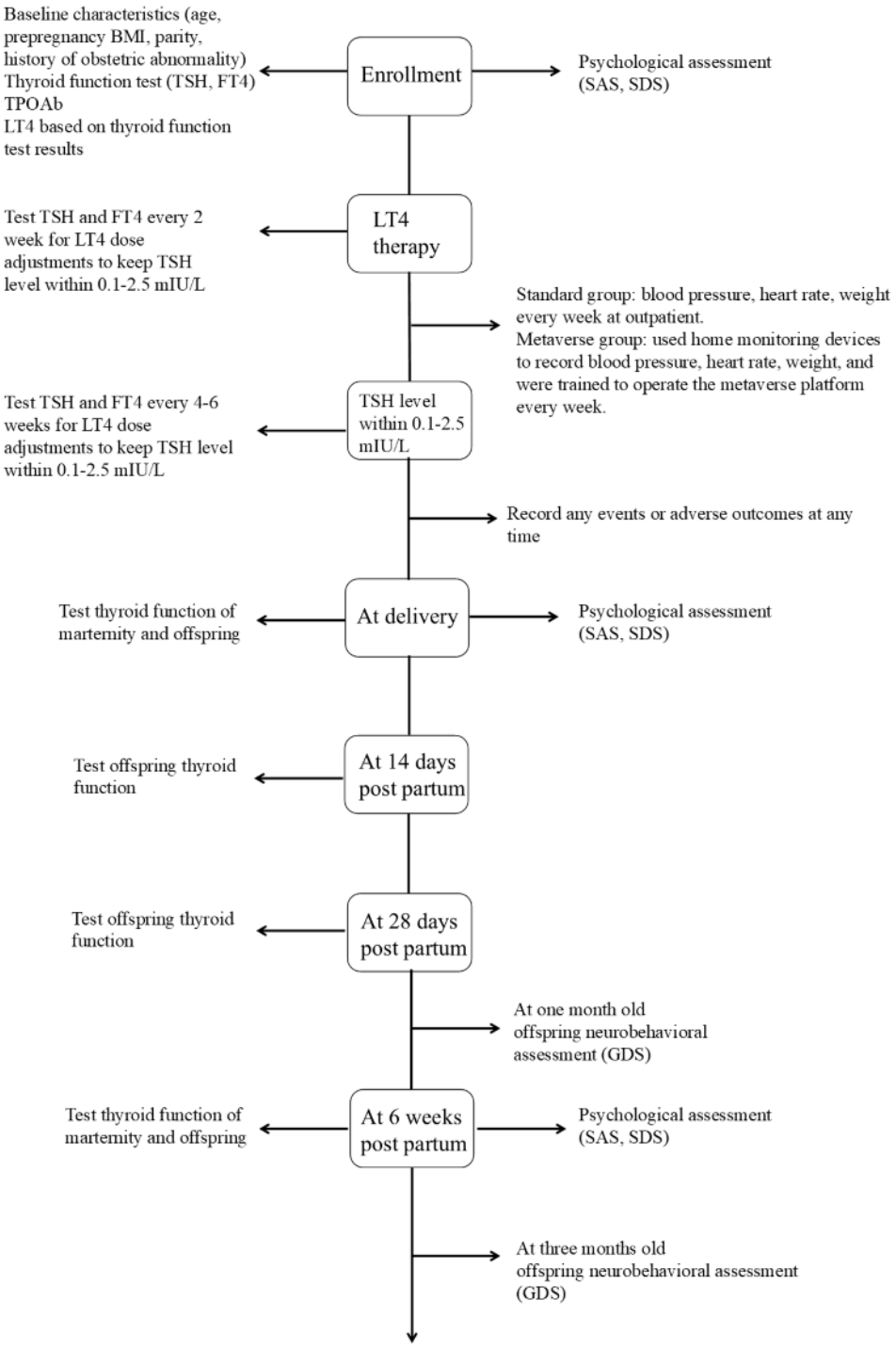

XSL•FO
RenderX

Table 1. Baseline characteristics of study participants (n=60).

	Standard group (n=30)	Metaverse group (n=30)
Age (years), median (IQR)	30 (20-40)	30 (23-38)
Gestational period (weeks), median (IQR)	39 (23-40)	39 (36-40)
Prepregnancy BMI (kg/m^2), median (IQR)	22.8 (15.8-32.7)	21.5 (17.6-30)
Parity		
Unipara, n (%)	19 (63)	18 (60)
Multipara, n (%)	11 (37)	12 (40)
History of obstetric abnormality, n (%)	3 (10)	3 (10)
TPOAb[a] (IU/mL), n (%)		
Positive	21 (70)	18 (60)
Negative	9 (30)	12 (40)
Levothyroxine (μg/d), median (IQR)	50 (50-75)	50 (50-75)

[a]TPOAb: thyroid peroxidase antibodies.

Primary Outcome

At baseline, the mean thyrotropin levels were 4.6 (SD 1.4) mIU/L in the standard group and 4.4 (SD 0.9) mIU/L in the metaverse group. After LT4 treatment, the mean thyrotropin level at delivery was 2.0 (SD 0.1) mIU/L in the standard group and 1.9 (SD 0.1) mIU/L in the metaverse group (*P*=.83). At 6 weeks post partum, the mean thyrotropin levels were 2.7 (SD 0.1) mIU/L in both groups (*P*=.78).

Maternal adverse outcomes occurred in 43% (13/30) of the standard group and 37% (11/30) of the metaverse group (*P*=.60). These included gestational diabetes, preeclampsia, premature rupture of membranes, and amniotic fluid abnormalities (Table 2). Offspring baseline characteristics showed no significant differences between the two groups (Table S1 in Multimedia Appendix 1). The median birthweight was 3290 (IQR 1830-3940) g in the standard group and 3315 (IQR 2870-3660) g in the metaverse group (*P*=.91). Two newborns in the standard group were diagnosed as small for gestational age. Adverse offspring outcomes were significantly lower in the metaverse group (2/30, 7%) compared to the standard group (10/30, 33%; *P*=.01). These outcomes included preterm birth, fetal distress, neonatal intensive care unit admissions, and neonatal hypoxemia (Table 3).

Table 2. Adverse outcomes of maternity.

	Standard group (n=30), n (%)	Metaverse group (n=30), n (%)	*P* value
Gestational diabetes mellitus	6 (20)	4 (13)	.49[a]
Hypertensive disorder complicating pregnancy	0 (0)	0 (0)	>.99[b]
Severe preeclampsia	1 (3)	0 (0)	>.99[b]
Threatened abortion	0 (0)	1 (3)	>.99[b]
Premature rupture of fetal membranes	5 (17)	5 (17)	.25[a]
Abnormal amniotic fluid	1 (3)	1 (3)	>.99[b]
Postpartum hemorrhage	0 (0)	0 (0)	>.99[b]
Total incidence	13 (43)	11 (37)	.60[a]

[a]Pearson χ^2 test.

[b]Continuity correction test.

XSL•FO
RenderX

Table 3. Adverse outcomes of the offspring.

	Standard group (n=30)[a], n (%)	Metaverse group (n=30), n (%)	*P* value
Abnormal perinatal ultrasound and fetal heart monitoring	0 (0)	0 (0)	>.99[b]
Low perinatal weight	1 (3)	0 (0)	>.99[b]
Preterm birth	3 (10)	1 (3)	.61[b]
Stillbirth	0 (0)	0 (0)	>.99[b]
Miscarriage	1 (3)	0 (0)	>.99[b]
Neonatal death	0 (0)	0 (0)	>.99[b]
Admission to neonatal intensive care unit	1 (3)	0 (0)	>.99[b]
Antepartum fetal distress	2 (7)	1 (3)	>.99[b]
Fetal asphyxia[c]	1 (3)	0 (0)	>.99[b]
Fetal malformation	0 (0)	0 (0)	>.99[b]
Neonatal hypoxemia	1 (3)	0 (0)	>.99[b]
Neonatal hypoxic-ischemic encephalopathy	0 (0)	0 (0)	>.99[b]
Neonatal pneumonia or wet lung	0 (0)	0 (0)	>.99[b]
Neonatal jaundice	0 (0)	0 (0)	>.99[b]
Neonatal hypothyroidism	0 (0)	0 (0)	>.99[b]
Total incidence	10 (33)	2 (7)	.01[d]

[a]One pregnant woman in the standard group had a termination due to fetal growth restriction.

[b]Continuity correction test.

[c]Fetal asphyxia refers to an Apgar score <4 at 1 minute or an Apgar score <7 at 5 minutes.

[d]Pearson χ^2 test.

Neonatal thyroid function showed no significant differences between the groups at birth, 14 days post partum, and 28 days post partum (Tables S2-S4 in Multimedia Appendix 1). The median thyrotropin levels at birth were 7.0 (IQR 3.2-16.2) mIU/L in the standard group and 6.2 (IQR 3.0-16.4) mIU/L in the metaverse group (*P*=.31). By 14 days post partum, the median thyrotropin level had decreased to 5.3 (IQR 3.5-7.8) mIU/L in the standard group and 5.0 (IQR 3.2-5.7) mIU/L in the metaverse group (*P*=.26). At 28 days post partum, median thyrotropin levels were 4.8 (IQR 3.2-6.1) mIU/L in the standard group and 4.7 (IQR 3.2-5.4) mIU/L in the metaverse group (*P*=.68). FT4 levels also showed no significant differences between the groups at any time point (all *P*>.05).

Secondary Outcomes

At 1 month of age, neurobehavioral scores showed no significant differences between the groups across all domains (Table 4). The median scores for the standard group in gross motor, fine motor, adaptability, language, and social-emotional domains were 99 (IQR 85.5-109.8), 100 (IQR 83.3-109), 99 (IQR 86-107.3), 101.25 (IQR 85-106.8), and 101 (IQR 87.8-107), respectively. The corresponding scores for the metaverse group were 100 (IQR 90.8-109), 100.9 (IQR 93-108), 100 (IQR 95-109.0), 102 (IQR 94-107.8), and 100.5 (IQR 92-109), respectively. At 3 months of age, similar results were observed, with neurobehavioral scores remaining comparable between groups (Table 4). The median scores for the standard group were 100 (IQR 88-111) for gross motor, 101 (IQR 89-113.3) for fine motor, 102.8 (IQR 88-112.8) for adaptability, 103.7 (IQR 88.8-111.3) for language, and 104.0 (IQR 90.8-111) for social-emotional domains. Scores for the metaverse group were 102 (IQR 92.8-109.8), 103.0 (IQR 95-112.8), 104 (IQR 96-114), 104.4 (IQR 96-111.0), and 104.9 (IQR 92.8-110), respectively.

Maternal mental health outcomes were assessed via the SAS and SDS. The baseline of the mean SAS scores was 44.1 (SD 8.3) in the standard group and 43.8 (SD 7.9) in the metaverse group. It showed significantly lower scores in the metaverse group compared to the standard group at delivery (SAS: mean 44.9, SD 1.0 vs mean 50.0, SD 1.0; *P*=.001; SDS: mean 44.7, SD 0.9 vs mean 49.7, SD 0.9; *P*<.001). This trend persisted at 6 weeks post partum, with the metaverse group continuing to exhibit lower scores (SAS: mean 46.9, SD 1.0 vs mean 51.3, SD 1.0; *P*=.002; SDS: mean 47.9, SD 1.0 vs mean 51.6, SD 1.0; *P*=.009).

XSL•FO
RenderX

Table 4. Comparison of neuropsychological development by the Gesell Development Scale.

	Standard group scores (n=30), median (IQR)[a]	Metaverse group scores (n=30), median (IQR)	*P* value[b]
At 1 month old			
Gross motor	99 (85.5-109.8)	100 (90.8-109)	.24
Fine motor	100 (83.3-109)	100.9 (93-108)	.61
Adaptability	99 (86-107.3)	100 (95-109.0)	.06
Language	101.3 (85-106.8)	102 (94-107.8)	.17
Social-emotional response	101 (87.8-107)	100.5 (92-109)	.64
At 3 months old			
Gross motor	100 (88-111)	102 (92.8-109.8)	.48
Fine motor	101 (89-113.3)	103.0 (95-112.8)	.61
Adaptability	102.8 (88-112.8)	104 (96-114)	.17
Language	103.7 (88.8-111.3)	104.4 (96-111.0)	.31
Social-emotional response	104.0 (90.8-111)	104.9 (92.8-110)	.16

[a]One pregnant woman in the standard group had a termination due to fetal growth restriction.

[b]Wilcoxon rank sum test.

Discussion

The post–COVID-19 era has underscored the transformative potential of Health 4.0, which integrates advanced technologies such as the metaverse, blockchain, and intelligent algorithms into health care systems [20]. The metaverse presents expansive opportunities, including avatar-based consultations for remote diagnoses, immersive simulations for medical training, and social interactions to enhance patient engagement [21]. Despite its growing application in various medical fields, its use in managing thyroid disorders, particularly during pregnancy, remains underexplored [22]. Thyroid dysfunction during pregnancy, especially SCH, has gained increasing attention due to its often overlooked association with adverse pregnancy outcomes [23]. The adoption of innovative technologies like the metaverse to provide personalized care for pregnant women with SCH is crucial for improving maternal and neonatal health.

Recent studies have focused on the etiology, diagnosis, and management of adverse pregnancy outcomes [23,24]. Mounting evidence links SCH during pregnancy to elevated risks of complications. For instance, a meta-analysis reported a higher incidence of preterm birth in women with SCH (odds ratio 2.9, 95% CI 1.0-1.6) [25]. Similarly, a retrospective study found that SCH was associated with increased risks of preterm birth (relative risk 2.2, 95% CI 1.1-4.0) and neonatal respiratory distress syndrome (relative risk 2.8, 95% CI 1.0-7.8), particularly in early pregnancy [26]. These findings emphasize the importance of timely and effective interventions to improve outcomes [17]. Inspired by the establishment of virtual home hospitals in the United States in 2019, our study used virtual simulation technologies for remote monitoring, enabling timely evaluations and interventions [27]. Notably, the incidence of adverse offspring outcomes was significantly lower in the metaverse group compared to the standard group. Although maternal adverse outcomes did not differ significantly between groups (13/30, 43% vs 11/30, 37%), a favorable trend was observed. Neurodevelopmental outcomes, assessed using GDS scores, showed no significant differences between groups. To our knowledge, this is the first study to explore the application of a metaverse-based platform for the management of pregnant women with SCH. Regular treatment and continuous monitoring are essential for this population, and the dynamic follow-up capabilities of the metaverse platform are particularly well suited to their needs. The platform facilitates remote consultations, real-time monitoring, and comprehensive patient education through individualized treatment plans [24-26,28]. Compared to conventional health care services, the metaverse platform offers a more immersive and accessible experience, which may improve engagement and adherence to care protocols.

As illustrated in Figure 3, a virtual clinic was developed to simulate a traditional outpatient environment. Both investigators and participants created digital avatars, enabling seamless access to the virtual clinic from any location. This setup allowed participants to communicate directly with investigators, sharing critical information on medication dosages, test results, diet, lifestyle, and emotional well-being. By eliminating the need for frequent hospital visits, this model fostered greater engagement in the treatment process. Moreover, the platform addressed health care disparities by delivering specialized care to patients in remote or underserved areas.

Maternal mental health is a vital component of perinatal care. Studies indicate that approximately one-quarter of pregnant women with SCH are diagnosed with depression [29]. Impaired mental health during the perinatal period has been linked to adverse outcomes for both mothers and their offspring, including preterm birth, miscarriage, impaired fetal development, and an increased risk of autism spectrum disorders [30-33]. Improving the emotional well-being of perinatal mothers has therefore become a focus of recent research aimed at enhancing pregnancy outcomes [34,35]. Our study explored the use of a metaverse

XSL•FO
RenderX

platform to address emotional disorders in pregnant women with SCH. Evidence supports the efficacy of VR in managing anxiety, depression, and fear [36]. Additionally, metaverse platforms have demonstrated success in providing nonpharmacological interventions for conditions such as posttraumatic stress disorder, depression, and cognitive decline [37]. These platforms enable the creation of highly customized therapeutic environments tailored to individual needs and treatment goals. The incorporation of gamification and interactive scenarios further enhances patient engagement, improving adherence to treatment plans [38,39] (Figure 4). Compared to the standard group, patients in the metaverse group showed significant improvements in emotional well-being and a lower incidence of adverse offspring outcomes, consistent with previous studies. These findings underscore the positive impact of addressing maternal emotional health on offspring outcomes. However, pregnancy is a complex process, and further mechanistic studies are necessary to elucidate the pathways underlying these effects. Our study suggests that the metaverse platform has the potential to benefit all pregnant women, not just those with SCH. By alleviating anxiety and providing foundational medical knowledge, such platforms could enhance maternal well-being and perinatal outcomes on a broader scale.

Figure 3. Screenshots from metaverse clinic, a virtual clinic. (A) Clinic scene. (B) Clinic chatting room. (C) Clinic conference room.



XSL•FO
RenderX

Figure 4. Screenshots from the metaverse medical education games. (A) Metaverse medical games option. (B) Metaverse hospital receptionist. (C) Stage mode of metaverse medical games. (D) Inpatient ward scene in metaverse medical game.

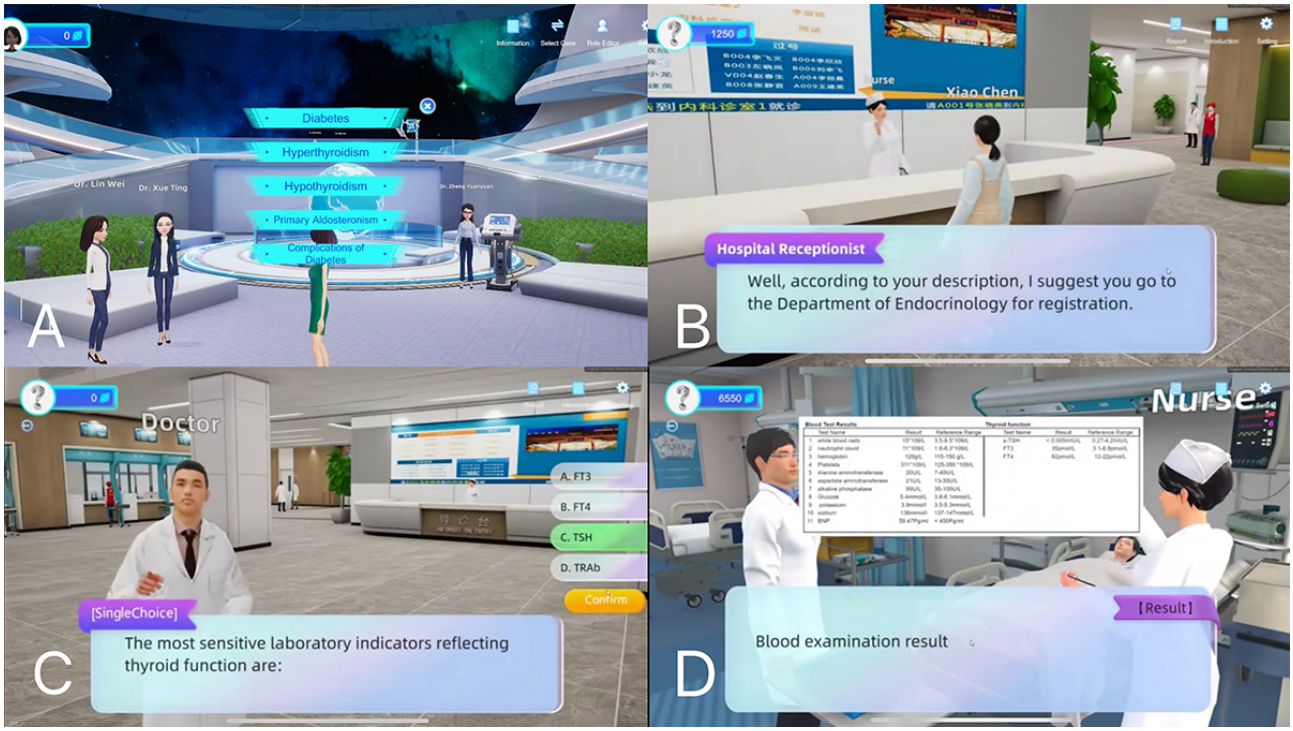

This study has several notable strengths. To our knowledge, it is the first to use a metaverse platform for managing pregnant women with SCH. Compared to traditional telemedicine services, the metaverse offers advancements, particularly in safeguarding patient privacy. The secure virtual environment minimizes data leakage risks, alleviating psychological burdens and reducing the embarrassment often associated with medical consultations. The anonymity afforded by virtual interactions encourages patients to communicate more openly and confidently with physicians, enhancing both diagnostic and therapeutic efficacy. Furthermore, the study extends beyond pregnancy-related adverse outcomes to include assessments of psychological well-being. The metaverse platform also demonstrates the potential to address global disparities in medical resources by enabling equitable access to high-quality health care services, regardless of geographic or economic constraints. This approach represents a step forward in fostering a more balanced distribution of medical resources and advancing global health systems.

Despite these strengths, the study has limitations. Its single-center design and relatively small sample size may restrict the generalizability of the findings. Efforts are underway to expand the application of the metaverse platform to more diverse populations and other endocrine disorders to evaluate its broader effectiveness. Furthermore, the current platform remains in its developmental stage. Dependence on high-quality interactive devices and associated costs limited participant recruitment, potentially introducing statistical bias.

The study highlights the efficacy of an innovative metaverse-based follow-up management system for pregnant women with SCH. Compared to conventional care, this platform significantly reduced adverse offspring outcomes, achieved comparable maternal adverse outcomes, and markedly alleviated psychological distress during pregnancy. By enabling remote access to specialized and personalized care, the metaverse platform addresses disparities in medical resource distribution. In the digital era, this technology offers a transformative approach to managing endocrine disorders, particularly thyroid diseases, with the potential to enhance patient outcomes and advance equitable health care delivery.

Acknowledgments

This research received the following grants: Fujian Provincial Department of Science and Technology, Joint Fund for Science and Technology Innovation (2023Y9290); Department of Endocrinology (National Clinical Priority Specialties; 2023001); and Fujian Clinical Medical Research Center for Type 2 Diabetes (2022Y2013).

We also thank the Fujian Huayu Education Technology Co Ltd for developing and constructing the metaverse platform.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

All authors made a significant contribution to the work. GC and WL conceived and designed the study. YZ, Yizhen Chen, and Yan Chen wrote the manuscript. YZ and Yizhen Chen conducted the data analysis. YZ, Yan Chen, LL, TX, and CC performed the study. JW and GC revised the manuscript and provided some recommendations. All authors read and approved the final manuscript. All authors agree to be accountable for all aspects of the work and the final approval of the version to be published. The above authors all participated in the construction of the metaverse platform.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 23 KB - jmir_v27i1e64634_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2903 KB - jmir_v27i1e64634_app2.pdf]

References

1. Cerasa A, Gaggioli A, Pioggia G, Riva G. Metaverse in mental health: the beginning of a long history. Curr Psychiatry Rep 2024 Jun;26(6):294-303 [FREE Full text] [doi: 10.1007/s11920-024-01501-8] [Medline: 38602624]
2. Girijakumari Sreekantan Nair S, Chandrasekaran B. MetaHealth: unlocking metaverse technologies in digital healthcare. Expert Rev Med Devices 2024 Aug;21(8):727-739. [doi: 10.1080/17434440.2024.2382233] [Medline: 39039648]
3. Checcucci E, Veccia A, Puliatti S, De Backer P, Piazza P, Kowalewski K, et al. Metaverse in surgery - origins and future potential. Nat Rev Urol 2024 Oct 30:10.1038/s41585-024-00941-4. [doi: 10.1038/s41585-024-00941-4] [Medline: 39349948]
4. Sadoughi F, Behmanesh A, Sayfouri N. Internet of things in medicine: a systematic mapping study. J Biomed Inform 2020 Mar;103:103383 [FREE Full text] [doi: 10.1016/j.jbi.2020.103383] [Medline: 32044417]
5. Plechatá A, Makransky G, Böhm R. Can extended reality in the metaverse revolutionise health communication? NPJ Digit Med 2022 Sep 02;5(1):132. [doi: 10.1038/s41746-022-00682-x] [Medline: 36056245]
6. Ali S, Abdullah, Armand TPT, Athar A, Hussain A, Ali M, et al. Metaverse in healthcare integrated with explainable AI and blockchain: enabling immersiveness, ensuring trust, and providing patient data security. Sensors (Basel) 2023 Jan 04;23(2):565 [FREE Full text] [doi: 10.3390/s23020565] [Medline: 36679361]
7. Checcucci E, Amparore D, Volpi G, De Cillis S, Piramide F, Verri P, et al. Metaverse surgical planning with three-dimensional virtual models for minimally invasive partial nephrectomy. Eur Urol 2024 Apr;85(4):320-325. [doi: 10.1016/j.eururo.2023.07.015] [Medline: 37673751]
8. Mishra R, Narayanan MDK, Umana GE, Montemurro N, Chaurasia B, Deora H. Virtual reality in neurosurgery: beyond neurosurgical planning. Int J Environ Res Public Health 2022 Feb 02;19(3):1719 [FREE Full text] [doi: 10.3390/ijerph19031719] [Medline: 35162742]
9. Anwer A, Jamil Y, Bilal M. Provision of surgical pre-operative patient counseling services through the Metaverse technology. Int J Surg 2022 Aug;104:106792 [FREE Full text] [doi: 10.1016/j.ijsu.2022.106792] [Medline: 35918004]
10. Sarasso E, Gardoni A, Tettamanti A, Agosta F, Filippi M, Corbetta D. Virtual reality balance training to improve balance and mobility in Parkinson's disease: a systematic review and meta-analysis. J Neurol 2022 Apr;269(4):1873-1888. [doi: 10.1007/s00415-021-10857-3] [Medline: 34713324]
11. Lee SY, Pearce EN. Testing, monitoring, and treatment of thyroid dysfunction in pregnancy. J Clin Endocrinol Metab 2021 Mar 08;106(3):883-892 [FREE Full text] [doi: 10.1210/clinem/dgaa945] [Medline: 33349844]
12. Dong AC, Stagnaro-Green A. Differences in diagnostic criteria mask the true prevalence of thyroid disease in pregnancy: a systematic review and meta-analysis. Thyroid 2019 Feb;29(2):278-289. [doi: 10.1089/thy.2018.0475] [Medline: 30444186]
13. Li J, Xu Y, Sun Z, Cai Y, Wang B, Zhang M, et al. Differential lipids in pregnant women with subclinical hypothyroidism and their correlation to the pregnancy outcomes. Sci Rep 2021 Oct 04;11(1):19689. [doi: 10.1038/s41598-021-99252-6] [Medline: 34608242]
14. Bein M, Yu OHY, Grandi SM, Frati FYE, Kandil I, Filion KB. Levothyroxine and the risk of adverse pregnancy outcomes in women with subclinical hypothyroidism: a systematic review and meta-analysis. BMC Endocr Disord 2021 Feb 27;21(1):34 [FREE Full text] [doi: 10.1186/s12902-021-00699-5] [Medline: 33639909]
15. Vimalananda VG, Brito JP, Eiland LA, Lal RA, Maraka S, McDonnell ME, et al. Appropriate use of telehealth visits in endocrinology: policy perspective of the endocrine society. J Clin Endocrinol Metab 2022 Nov 23;107(11):2953-2962 [FREE Full text] [doi: 10.1210/clinem/dgac494] [Medline: 36194041]

XSL•FO
RenderX

16. Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, et al. 2017 guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. Thyroid 2017 Mar;27(3):315-389. [doi: 10.1089/thy.2016.0457] [Medline: 28056690]
17. Ge GM, Leung MTY, Man KKC, Leung WC, Ip P, Li GHY, et al. Maternal thyroid dysfunction during pregnancy and the risk of adverse outcomes in the offspring: a systematic review and meta-analysis. J Clin Endocrinol Metab 2020 Dec 01;105(12):dgaa555. [doi: 10.1210/clinem/dgaa555] [Medline: 32810262]
18. Chen R, Li B, Wen J, Zhong C, Ji M. Clinical and electroencephalogram characteristics and treatment outcomes in children with benign epilepsy and centrotemporal spikes. World J Clin Cases 2021 Nov 26;9(33):10116-10125 [FREE Full text] [doi: 10.12998/wjcc.v9.i33.10116] [Medline: 34904081]
19. He L, Wang T, Xu H, Chen C, Liu Z, Kang X, et al. Prevalence of depression and anxiety in women with recurrent pregnancy loss and the associated risk factors. Arch Gynecol Obstet 2019 Oct;300(4):1061-1066. [doi: 10.1007/s00404-019-05264-z] [Medline: 31485778]
20. Liu Z, Ren L, Xiao C, Zhang K, Demian P. Virtual reality aided therapy towards Health 4.0: a two-decade bibliometric analysis. Int J Environ Res Public Health 2022 Jan 28;19(3):1525 [FREE Full text] [doi: 10.3390/ijerph19031525] [Medline: 35162546]
21. Li Y, Gunasekeran DV, RaviChandran N, Tan TF, Ong JCL, Thirunavukarasu AJ, et al. The next generation of healthcare ecosystem in the metaverse. Biomed J 2024 Jun;47(3):100679 [FREE Full text] [doi: 10.1016/j.bj.2023.100679] [Medline: 38048990]
22. Lee CW. Application of metaverse service to healthcare industry: a strategic perspective. Int J Environ Res Public Health 2022 Oct 11;19(20):13038 [FREE Full text] [doi: 10.3390/ijerph192013038] [Medline: 36293609]
23. Lee SY, Pearce EN. Assessment and treatment of thyroid disorders in pregnancy and the postpartum period. Nat Rev Endocrinol 2022 Mar;18(3):158-171 [FREE Full text] [doi: 10.1038/s41574-021-00604-z] [Medline: 34983968]
24. Dong AC, Morgan J, Kane M, Stagnaro-Green A, Stephenson MD. Subclinical hypothyroidism and thyroid autoimmunity in recurrent pregnancy loss: a systematic review and meta-analysis. Fertil Steril 2020 Mar;113(3):587-600.e1 [FREE Full text] [doi: 10.1016/j.fertnstert.2019.11.003] [Medline: 32192591]
25. Consortium on Thyroid and Pregnancy—Study Group on Preterm Birth, Korevaar TIM, Derakhshan A, Taylor PN, Meima M, Chen L, et al. Association of thyroid function test abnormalities and thyroid autoimmunity with preterm birth: a systematic review and meta-analysis. JAMA 2019 Aug 20;322(7):632-641 [FREE Full text] [doi: 10.1001/jama.2019.10931] [Medline: 31429897]
26. Lee SY, Cabral HJ, Aschengrau A, Pearce EN. Associations between maternal thyroid function in pregnancy and obstetric and perinatal outcomes. J Clin Endocrinol Metab 2020 May 01;105(5):e2015-e2023 [FREE Full text] [doi: 10.1210/clinem/dgz275] [Medline: 31838502]
27. Ryan PP, Hawkins KL, Altman S, Granatowski L, Shy BD, Long J, et al. A novel virtual hospital at home model during the coronavirus disease 2019 (COVID-19) pandemic. Infect Control Hosp Epidemiol 2021 Sep;42(9):1140-1142 [FREE Full text] [doi: 10.1017/ice.2020.435] [Medline: 32829735]
28. Uscher-Pines L, Thompson J, Taylor P, Dean K, Yuan T, Tong I, et al. Where virtual care was already a reality: experiences of a nationwide telehealth service provider during the COVID-19 pandemic. J Med Internet Res 2020 Dec 15;22(12):e22727 [FREE Full text] [doi: 10.2196/22727] [Medline: 33112761]
29. Costantine MM, Smith K, Thom EA, Casey BM, Peaceman AM, Varner MW, et al. Effect of thyroxine therapy on depressive symptoms among women with subclinical hypothyroidism. Obstet Gynecol 2020 Apr;135(4):812-820 [FREE Full text] [doi: 10.1097/AOG.0000000000003724] [Medline: 32168208]
30. Toloza FJK, Theriot SE, Singh Ospina NM, Nooruddin S, Keathley B, Johnson SM, et al. Knowledge, attitudes, beliefs, and treatment burden related to the use of levothyroxine in hypothyroid pregnant women in the United States. Thyroid 2021 Apr;31(4):669-677 [FREE Full text] [doi: 10.1089/thy.2020.0629] [Medline: 33176609]
31. Casey S, Carter M, Looney AM, Livingstone V, Moloney G, O'Keeffe GW, SCOPE Consortium. Maternal mid-gestation cytokine dysregulation in mothers of children with autism spectrum disorder. J Autism Dev Disord 2022 Sep;52(9):3919-3932 [FREE Full text] [doi: 10.1007/s10803-021-05271-7] [Medline: 34505185]
32. Osborne LM, Yenokyan G, Fei K, Kraus T, Moran T, Monk C, et al. Innate immune activation and depressive and anxious symptoms across the peripartum: an exploratory study. Psychoneuroendocrinology 2019 Jan;99:80-86 [FREE Full text] [doi: 10.1016/j.psyneuen.2018.08.038] [Medline: 30195110]
33. Mancuso RA, Ross KM, Accortt E, Coussons-Read M, Okun ML, Irwin J, et al. Prenatal mood and anxiety disorders and associated cytokine changes. J Affect Disord 2024 Feb 15;347:635-644. [doi: 10.1016/j.jad.2023.12.014] [Medline: 38070749]
34. Hassdenteufel K, Müller M, Abele H, Brucker SY, Graf J, Zipfel S, et al. Using an electronic mindfulness-based intervention (eMBI) to improve maternal mental health during pregnancy: results from a randomized controlled trial. Psychiatry Res 2023 Dec;330:115599 [FREE Full text] [doi: 10.1016/j.psychres.2023.115599] [Medline: 37988816]
35. Boran P, Dönmez M, Barış E, Us MC, Altaş ZM, Nisar A, et al. Delivering the Thinking Healthy Programme as a universal group intervention integrated into routine antenatal care: a randomized-controlled pilot study. BMC Psychiatry 2023 Jan 06;23(1):14 [FREE Full text] [doi: 10.1186/s12888-022-04499-6] [Medline: 36604685]

XSL•FO
RenderX

36. Usmani SS, Sharath M, Mehendale M. Future of mental health in the metaverse. Gen Psychiatr 2022;35(4):e100825 [FREE Full text] [doi: 10.1136/gpsych-2022-100825] [Medline: 36189180]
37. Zhou H, Gao J, Chen Y. The paradigm and future value of the metaverse for the intervention of cognitive decline. Front Public Health 2022;10:1016680 [FREE Full text] [doi: 10.3389/fpubh.2022.1016680] [Medline: 36339131]
38. Flavián C, Ibáñez-Sánchez S, Orús C, Barta S. The dark side of the metaverse: the role of gamification in event virtualization. Int J Inf Manage 2024 Apr;75:102726. [doi: 10.1016/j.ijinfomgt.2023.102726]
39. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]

Abbreviations

FT4: free thyroxine
GDS: Gesell Development Scale
LT4: levothyroxine
SAS: Self-Rating Anxiety Scale
SCH: subclinical hypothyroidism
SDS: Self-Rating Depression Scale
VR: virtual reality

Edited by A Schwartz; submitted 06.08.24; peer-reviewed by B Dincgez, L Min, H Chen; comments to author 10.12.24; revised version received 15.12.24; accepted 14.01.25; published 05.02.25.

Please cite as:
Zheng Y, Chen Y, Chen Y, Lin L, Xue T, Chen C, Wen J, Lin W, Chen G
Metaverse Clinic for Pregnant Women With Subclinical Hypothyroidism: Prospective Randomized Study
J Med Internet Res 2025;27:e64634
URL: https://www.jmir.org/2025/1/e64634
doi:10.2196/64634
PMID:

©Yuanyuan Zheng, Yizhen Chen, Yan Chen, Liang Lin, Ting Xue, Chuhui Chen, Junping Wen, Wei Lin, Gang Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Pornography-Watching Disorder and Its Risk Factors Among Young Adults: Cross-Sectional Survey

Csaba Erdős[1*], MSc; Oguz Kelemen[2*], MD, PhD; Dávid Pócs[3], MD, PhD; Edit Paulik[1], MD, PhD, Prof Dr Med; András Papp[1], PhD; Edina Horváth[4], MD, PhD; Arbel Golan[1], MD; Krisztián Széll[5], PhD

[1]Department of Public Health, University of Szeged, Albert Szent-Györgyi Medical School, Szeged, Hungary

[2]Department of Behavioral Sciences, University of Szeged, Albert Szent-Györgyi Medical School, Szeged, Hungary

[3]Cardiology Center, Department of Internal Medicine, University of Szeged, Szeged, Hungary

[4]Family Medicine Department, University of Szeged, Szeged, Hungary

[5]Institute of Education, Eötvös Loránd University, Budapest, Hungary

[*]these authors contributed equally

Corresponding Author:
Csaba Erdős, MSc
Department of Public Health
University of Szeged
Albert Szent-Györgyi Medical School
Dóm tér 10
Szeged, H-6720
Hungary
Phone: 36 62545119
Fax: 36 62545120
Email: erdos.csaba@med.u-szeged.hu

Abstract

Background: The widespread availability of internet-based pornography has led to growing concerns about its impact on mental health, particularly among young adults. Despite increasing recognition of problematic pornography use, standardized diagnostic criteria for pornography addiction are lacking.

Objective: This study aimed to address this gap by applying adapted *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) criteria to evaluate "pornography-watching disorder" (PWD) in a large sample of young adults in Hungary. The primary objective was to assess the prevalence of PWD among young adults and identify key risk factors associated with its development using *DSM-5* criteria adapted for pornography use. It also aimed to advance the understanding of PWD as a potential behavioral addiction.

Methods: A cross-sectional web-based survey was conducted between September and December 2018, targeting young adults aged 18-35 years in Hungary. Participants were recruited through social media and the University of Szeged Albert Szent-Györgyi Medical School's web page. Of the 9397 respondents, 7187 (76.5%) had previously consumed pornography and were included in the analysis. PWD was measured using 10 statements adapted from the *DSM-5* substance use disorder criteria. Multivariable binary logistic regression was used to identify significant predictors of PWD.

Results: The prevalence of PWD in the sample was 4.4% (n=315). Frequent pornography consumption was a significant risk factor, with weekly users (odds ratio [OR] 0.45, 95% CI 0.33-0.62, *P*<.001), monthly users (OR 0.18, 95% CI 0.11-0.28, *P*<.001), and less than monthly users (OR 0.05, 95% CI 0.03-0.10, *P*<.001) showing significantly lower odds compared with daily users as a reference category. Male sex was associated with a higher risk (OR 0.53, 95% CI 0.39-0.72, *P*<.001), as were early exposure to pornography (OR 0.94, 95% CI 0.90-0.98, *P*=.006), paraphilia (OR 3.95, 95% CI 2.37-6.56, *P*<.001), dissatisfaction with sexual life (OR 0.94, 95% CI 0.90-0.98, *P*=.006), difficulty forming personal relationships (OR 0.93, 95% CI 0.88-0.98, *P*=.005), and strong adherence to religious norms (OR 1.12, 95% CI 1.06-1.19, *P*<.001). Protective factors included adequate sexual education (OR 0.67, 95% CI 0.53-0.87, *P*=.02) and residing in the capital (OR 0.52, 95% CI 0.30-0.91, *P*=.02). The use of an anonymous web-based questionnaire likely reduced the influence of stigma, resulting in more accurate self-reporting of sensitive behaviors.

XSL•FO
RenderX

Conclusions: This study is among the first to apply *DSM-5* criteria to evaluate PWD, providing important insights into its prevalence and associated risk factors in young adults. The findings highlight the need for standardized diagnostic tools for PWD and suggest targeted interventions, particularly for high-risk groups. These results contribute to the ongoing discussion about whether pornography addiction should be recognized as a distinct behavioral disorder.

(J Med Internet Res 2025;27:e49860) doi:10.2196/49860

KEYWORDS

pornography; Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; DSM-5; paraphilia; satisfaction; sexual disorder; sexual education; online survey; young adults; online; risk factor; cross sectional; online survey; pornography-watching disorder; social media; web-based questionnaire; internet use; public health; prevention programs; sexual well-being

Introduction

Background

The use of pornography has increased significantly in the last few decades, in parallel with the development of the technological means of its transmission, so that access is much easier nowadays than before. Internet pornography shows an increasing trend worldwide, according to Pornhub statistics [1]. To obtain exact data is almost impossible since the number of websites providing erotic content intended for adults is untraceable, and there are no exact data available regarding their use. The massive and ongoing increase in internet use, and internet porn use in particular, has had a profound effect on human sexual behavior over the last decades and related social dynamics [2]. It is estimated that 46%-74% of men and 16%-41% of women in developed countries are active pornography users [3]. Consumption on a weekly basis was found in approximately 50% of young adults under the age of 25, with a general decrease in pornography use with increasing age. According to a study performed in Australia, 84% (8442/10,049) of men from a representative sample population watched porn in their lifetime, while weekly porn consumption was reported in 76% (6131/10,049) in the sample [4]. With the global spread of COVID-19, the use of pornography increased by 4%-24% in several countries in 2020 [5]. Additional features that contribute to the high popularity of internet pornography are the novelty and diversity of the provided contents, which might become a driving force for problematic consumption, as the abundance of sexual imagery can entice further use and lead to tolerance and habituation, well-known in addiction disorders such as alcoholism [6].

Several studies highlight that pornography consumption is often presented as a symptom of sex addiction [7-9]. Individuals with compulsive sexual behavior may misuse the internet in the forms of problematic internet porn use and sex chats as a novel source of sexual gratification, imposing several hazards for themselves and, indirectly, a risk to society [10]. Reid et al [11] found in a field trial on hypersexual disorders that excessive use of pornography was one of the significant behavioral traits of individuals seeking a cure for their disorder, in addition to forced masturbation and frequent casual sex.

Excessive use of internet pornography has emerged as a new type of behavioral addiction, which is in contrast to compulsive sexual behavior disorder being classified as an impulse control disorder rather than an addictive condition in the *ICD-11* (*International Classification of Diseases for Mortality and Morbidity Statistics Eleventh Revision*) [12]. However, the entity pornography addiction is not officially recognized as a disease due to continuous debate among researchers regarding its definition and nature [13]. There is ongoing debate among researchers and clinicians about whether pornography addiction meets the criteria to be considered a distinct mental health disorder. Some researchers contend that compulsive consumption of internet pornography could represent a distinct manifestation of sex addiction or hypersexuality [14], while alternative perspectives advocate for its categorization as a unique subtype of internet addiction [15].

Neither *ICD-11* nor the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders*, [Fifth Edition]) provided accepted criteria to identify and define pornography addiction as a type of behavioral addiction so far [16-18].

Reflecting the urgent need for measuring excessive and problematic consumption of pornography characterized by impulsive and compulsive behavior, an alternative definition of self-perceived pornography addiction (SPPA) was suggested [13]. Based on evidence from neurobiological studies demonstrating an overlap between the mechanisms observed in excessive pornography use and those seen in substance addictions, the existing diagnostic criteria for addiction disorders may also be applicable to pornography addiction [19]. Diagnosing addiction to pornography presents challenges due to the subjective nature of defining problematic use, the diversity of individual experiences with pornography, and the lack of clear diagnostic criteria. Without consensus on standardized diagnostic criteria, it is difficult to establish a clear definition of pornography addiction for inclusion in diagnostic manuals.

Numerous procedures are available to measure problematic (compulsive or addictive) use of pornography. Bőthe et al [20] published the Problematic Pornography Consumption Scale (PPCS), which is based on previous addiction models. Nevertheless, research on problematic internet pornography consumption might be limited or lacking due to the absence of formally accepted scales that are compatible with widely accepted diagnostic criteria. Furthermore, 2 scales that aim to assess the addiction characteristics of internet pornography are the Internet Sex Screening Test (ISST) [21] and the Cyber-Pornography Use Inventory (CPUI) [22], both of which work accurately but with limited clinical relevance. The use of compulsive sexual behavior is not suitable for measuring addiction to pornography, as it is an umbrella term that includes many compulsive behaviors and cannot be reduced to specific aspects of pornography [23].

XSL•FO
RenderX

On the other hand, a functional MRI (magnetic resonance imaging)-based study focused on internet pornography addiction revealed similar brain activity patterns in porn addicts as in drug or alcohol addicts. The observation of more robust activation in compulsive sexual behavior than in healthy volunteers was similar to that seen in drug addiction, suggesting neurobiological similarities between the disorders [24].

Defining sexuality-related addictions is challenging because, although sexual addiction and substance dependence share similarities in brain function and behavior, effectively addressing sexuality-related addictions requires a nuanced understanding of the complex interplay among biological, psychological, and social factors that influence human sexuality.

Objectives of the Study

In the present study, the first aim was to transform the substance use disorder criteria of DSM-5 for measuring porn use based on the above finding. This set of criteria was then used to measure and characterize what we call here "pornography-watching disorder" (PWD) in a young adult population. Such a new measurement method seems to be necessary because the previous methods examined excessive porn use as a behavioral problem, while in the present study, the clinically more relevant criteria of disorder could be assessed. Our second aim was to find relationships between personal features and PWD, which can help future prevention.

Methods

Study Design and Recruitment

The cross-sectional exploratory study was conducted using a web-based questionnaire between September and December 2018 in Hungary. A convenience sampling technique was applied. The questionnaire link was made available on the Hungarian-language website of the Albert Szent-Györgyi Medical School. The survey was conducted in Hungarian. Participants were recruited through internet channels, including advertisements on social media platforms such as Facebook (Meta) and Instagram (Meta), as well as through announcements at university lectures. Given the widespread use of social media among internet users, it was chosen as a key platform for participant recruitment. Opt-in sampling in this research was essential to detect and define the problem. Because of the stigmatized nature of the problem, the opt-in sample was best suited to enable the research participants to disclose their problems more easily. While specific measures to confirm residency in Hungary were not used, recruitment was primarily targeted at Hungarian residents through the University of Szeged Albert Szent-Györgyi Medical School's platforms and Hungarian-language content. Our exploratory research could form the basis for a subsequent representative sample study. The target population was young adults (18-35 years old). We focused on this age group because participants are at a similar stage of life, reducing confounding variables. This age range is relevant because it is when individuals are discovering their sexuality and forming lasting habits. It targets a demographic group that is vulnerable to the effects of pornography because of its accessibility and developmental stage. Narrowing the focus will facilitate subsequent comparative analysis with other age groups, providing insights into developmental trajectories and generational trends.

Survey Development

The study, titled "The Great Sex Test" (in Hungarian: "A nagy szexteszt"), used a comprehensive questionnaire that addressed various topics, including general sexual health, sexual disorders, sexual satisfaction, contraception, sexually transmitted diseases, pornography addiction, and sexual abuse. This paper specifically focuses on the analysis related to pornography. The questionnaire was developed by a multidisciplinary team, including psychiatrists, sociologists, epidemiologists, medical students, as well as a sexologist and an information technology specialist, whose combined expertise ensured the robustness of the tool.

Pilot testing, which included in-depth interviews, was conducted to ensure the clarity and consistency of the questions. Clear instructions were provided to respondents to facilitate accurate responses. The Cantril scale was frequently used to measure attitudes, and validated variables from previous research were incorporated into the question design, ensuring the questionnaire remained concise without compromising quality.

The final questionnaire was created using Google Forms, a widely accessible platform that adheres to Google's public data protection policies. The responses were collected in an aggregated, anonymized form and downloaded as a Microsoft Excel spreadsheet.

The questionnaire, which was administered on a voluntary basis, included 218 questions covering all topics, including both male and female sexual disorders. The use of adaptive questioning significantly reduced the number of questions presented to each respondent based on their relevance. Before beginning the survey, participants were required to read and agree to a consent form that detailed the study's objectives, assured data security, and guaranteed anonymity. The consent form also emphasized the importance of providing accurate information.

An email address was required for questionnaire completion to minimize the risk of multiple submissions. The system handled email addresses separately from the questionnaires. After providing their email, participants were directed to the questionnaire interface through a link. The provided email addresses could not be linked to the completed questionnaires. Although IP address tracking or cookies were not used due to platform limitations, respondents were allowed to review and correct their responses before submission using the back button. Participants could also choose to skip questions by selecting "I don't know" or "I prefer not to answer." Irrelevant questions were filtered out through adaptive questioning, ensuring that participants only encountered questions pertinent to them (eg, those who had never viewed pornographic content were not asked related questions).

The number of questions per page was thematically determined, with most pages containing 5-6 questions. The total number of pages varied based on the relevance of the questions to the respondent, with the average questionnaire completion time being 24 minutes. Only fully completed and submitted questionnaires were included in the final analysis. A multistep

XSL•FO
RenderX

filtering process was used to exclude questionnaires with unrealistic, frivolous, or inappropriate answers, as well as those completed significantly faster than the average time to maintain the integrity of the results.

Measurements

Dependent Variable: Use of Pornography

An adapted version of DSM-5 criteria for substance use disorder was used in this study, omitting the question on withdrawal. The criteria applied to pornography can be classified as non–substance-related addictions. In these disorders, somatic withdrawal symptoms are not observed, and psychological withdrawal symptoms are difficult to differentiate. Withdrawal is typically a consequence of the sudden stop of some substance. A possible alternative, the criteria for gambling disorder, was considered but abandoned as financial gains and losses are irrelevant here. The criteria of *DSM-5* provide standardized, comprehensive, and clinically valuable guidelines. They are integrated into health care systems and have strong research validity. Furthermore, these criteria have been developed through a rigorous process that involves expert consensus and empirical evidence. It is important to note that using these criteria is not intended to create a new health condition but only to better define an existing problem.

The questionnaire used to identify people with PWD consisted of 10 statements. The first 5 statements reflect the loss of control, and the following 5 questions reflect the negative effects associated with pornography use. The respondents had to decide whether the 10 statements were typical of them (0=false and 1=true) in the past year. The answers were summarized by simple addition, resulting in a 0-10 scale (Cronbach α=0.70), where the cutoff point was set at 3. Accordingly, the group of people with PWD included those who stated that 4 or more statements were typical of them in the past year. The cutoff value was the same as for alcohol use disorder in the *DSM-5* criteria [17]. We did not differentiate between mild, moderate, and vigorous disorders. It is important to emphasize that *DSM-5* does not use the term addiction because of its vague definition and negative connotations. Furthermore, addiction is a broader term used to describe physical dependence and psychological compulsion without specific criteria. By using more precise terminology like “substance use disorder,” the *DSM-5* aims to provide more explicit diagnostic criteria and improve clinical communication. Since the PWD criteria were developed using the *DSM-5* criteria, we have avoided the term addiction.

Independent Personal Variables

The questionnaire assessed the demographic characteristics of the participants, including biological sex (0=male and 1=female), age (0=18-24 years and 1=25-35 years), place of residence (0=not in the capital and 1=in the capital), the highest educational attainment of the mother or father (corresponding to International Standard Classification of Education [ISCED] 2011: 0-2, 3 without a general certificate of education, 4 with general certificate of education, 5-8 tertiary education), employment status (working, studying, studying and working, and not studying and not working), parental status (1=have children or 0=not), and relationship status (0=single and 1=in a relationship). We also studied family structure in childhood (1=grew up in a 2-parent family or 0=not) and whether the parents in childhood family are divorced (0=no and 1=yes). A 10-point scale was used to examine the respondent's ideological orientation (between 1=conservative and 10=liberal) and degree of religiosity (by responses to the question, “How much do you adhere to religious norms?”; between 1=not respecting religious norms at all and 10=fully respecting religious norms).

Social relationships and attitudes toward others were captured by a 10-point scale variable that measures the difficulty of establishing personal relationships (“How easily do you make a personal relationship with others?”; from 1=with great difficulty to 10=very easily).

Further questions were aimed at the sexual orientation of the respondents (0=nonheterosexual and 1=heterosexual), the age when they had their first sexual intercourse, the experience in their first sexual intercourse (between 1=very bad and 10=very good), the number of sexual intercourses they had in the last year, the use of sexual performance enhancers (0=not used yet and 1=already used), their satisfaction with their sexual life, body, and genitals (between 1=not satisfied at all and 10=completely satisfied), and how sexually attractive they consider themselves (between 1=not at all and 10=completely). Subjective evaluation areas were measured similarly to the Cantril ladder scale [25].

The questionnaire also covered various sexual disorders. For men, delayed ejaculation, premature (early) ejaculation, hypoactive sexual desire disorder, and erectile disorder, while for women, female sexual interest or arousal disorder, genito-pelvic pain or penetration disorder, and female orgasmic disorder were explored. The answers were summarized into a sexual disorder variable describing the participants being affected by some sexual disorder according to the criteria of *DSM-5* (0=no and 1=yes) [26]. Involvement in paraphilias was measured using the 2 criteria found in *DSM-5*. Criterion “A” explores the qualitative nature of paraphilia, while criterion “B” focuses on the negative consequences. According to *DSM-5*, paraphilic disorders can be diagnosed if both diagnostic criteria are met, and this is also how this disorder was defined in the present study (0=no paraphilia and 1=have paraphilia).

We further measured the frequency of masturbation (1=no masturbation, 2=less than monthly, 3=monthly, 4=weekly, and 5=daily), if the respondents used to talk about their sexual life with others (0=no and 1=yes), and their sexual education, that is, if they received adequate and sufficient sexual information before their first sexual intercourse (0=no and 1=yes), and whether they were ever sexually abused (0=no and 1=yes).

The survey examined the age when the respondents first encountered pornographic content, what kind of experience this first time was (between 1=very bad and 10=very good), and how often they watched porn in the last year (1=never in the last 1 year, 2=less than monthly, 3=monthly, 4=weekly, and 5=daily).

Statistical Analysis

In addition to descriptive statistics and correlation tests (chi-square test and independent samples *t* test), multivariable

XSL•FO
RenderX

binary logistic regression was performed using the Enter method (whereby all variables are added to the model simultaneously) to identify characteristics predicting the risk of PWD. Our model-building strategy was guided by finding the most straightforward, yet most robust model. Since the distribution of those with PWD and those not affected in the sample is very unequal, in order to avoid overfitting and find the simplest model, we built our model based on suggestions by Hosmer et al [27]. The final model includes only those variables that were found to be significant at the $P<.05$ significance level. The final model did not include any variables with interaction effects that were professionally relevant (ie, well-founded, informative, interpretable, significant, and improving model fit). The relative importance of the variables included in the final model was determined using Dominance Analysis [28]. All statistical analyses were conducted with IBM SPSS Statistics (version 28.0) for Windows.

Ethical Considerations

The study protocol was reviewed and approved by the Regional and Institutional Human Medical Biological Research Ethics Committee of the University of Szeged (ethical approval 161/2017-SZTE). Participation in the study was entirely voluntary, and informed written consent was obtained from all participants before their involvement. The study was conducted in accordance with the principles of the Declaration of Helsinki. All data collected were anonymized to ensure the privacy and confidentiality of the participants.

Informed consent was obtained from all participants before their participation in the study. Participants were provided with detailed information about the purpose of the research, the procedures involved, and the nature of the data being collected. They were informed that their participation was voluntary, and that they could withdraw from the study at any time without any consequences. The consent form also assured participants of the confidentiality of their data, which was anonymized during analysis to protect their privacy. The original informed consent explicitly included provisions for the use of the collected data in subsequent analyses, including secondary research. Therefore, no additional consent was required for the secondary analysis performed in this study, as it was covered by the initial consent obtained from the participants.

The privacy and confidentiality of all participants were rigorously protected throughout the study. All data collected through the web-based survey were anonymized, ensuring that no personally identifiable information was linked to the responses. The survey platform used, Google Forms, provided data in a deidentified format, further ensuring that individual participants could not be identified. The anonymized data were securely stored on the internal servers of the University of Szeged Albert Szent-Györgyi Medical School's Department of Public Health and Department of Behavioral Sciences with access restricted to authorized research personnel only. In compliance with ethical guidelines, no personal information, such as names or IP addresses, was collected or stored, thereby maintaining the highest level of confidentiality for the study participants.

To encourage participation, tangible incentives such as theater tickets, movie tickets, and spa passes were offered. These prizes were raffled among participants after the data collection was completed. To ensure transparency and fairness in the compensation process, participants were informed about the raffle at the beginning of the survey. Winners were selected randomly and notified through the email addresses they provided.

Results

A total of 9397 individuals completed the questionnaire. Of these, 7187 had previously consumed pornographic content, and only these participants were included in the present analysis.

A detailed description of the study sample is presented in Tables 1 and 2. The sample consisted of 2398 men and 4789 women. Most of the respondents lived in rural areas and had a relationship. More than one-tenth of the participants (782/7187, 10.9%) have used sexual performance enhancers, while some form of paraphilia was detected in 1.5% (108/7187). Most of them felt that they received adequate sexual information before their first sexual intercourse (5030/7187, 70%). Almost a quarter of them (1631/7187, 22.7%) had not watched porn in the past year, while a quarter of them (1748/7187, 24.3%) did that less often than monthly, less than one-fifth of them (1234/7187, 17.2%), monthly, three-tenths of them (2223/7187, 30.9%), weekly. Every 20th respondent (351/7187, 4.9%) watched porn every day (Table 1).

XSL•FO
RenderX

Table 1. Distribution of categorical variables in the affected and nonaffected subgroup of the study sample and univariate associations of these variables with pornography-watching disorder.

Variables	Pornography-watching disorder absent (n=6872), n (%)	Pornography-watching disorder present (n=315), n (%)	Total sample (N=7187), n (%)	Chi-square (*df*)	*P* value	Φ[a]
Sociodemographic background						
Biological sex				170.6 (1)	<.001	0.15
Male	2186 (91.2)	212 (8.8)	2398 (33.4)			
Female	4686 (97.8)	103 (2.2)	4789 (66.6)			
Age range				0.02 (1)	.90	—[b]
18-24 years	5237 (95.6)	241 (4.4)	5478 (76.2)			
25-35 years	1635 (95.7)	74 (4.3)	1709 (23.2)			
Place of residence				7.9 (1)	.005	0.03
Not in the capital	6251 (95.4)	301 (4.6)	6552 (91.2)			
In the capital	621 (97.8)	14 (2.2)	635 (8.8)			
Father's highest educational attainment				0.6 (3)	.89	—
Primary education (ISCED[c] 0-2)	690 (96)	29 (4)	719 (10)			
Lower secondary education (ISCED 3)	2836 (95.7)	127 (4.3)	2963 (41.2)			
Upper secondary education (ISCED 4)	1927 (95.3)	94 (4.7)	2021 (28.1)			
Tertiary education (ISCED 5-8)	1419 (95.6)	65 (4.4)	1484 (20.6)			
Mother's highest educational attainment				3.2 (3)	.37	—
Primary education (ISCED 0-2)	799 (96.4)	30 (3.6)	829 (11.5)			
Lower secondary education (ISCED 3)	1727 (95.8)	75 (4.2)	1802 (25.1)			
Upper secondary education (ISCED 4)	2483 (95.1)	128 (4.9)	2611 (36.3)			
Tertiary education (ISCED 5-8)	1863 (95.8)	82 (4.2)	1945 (27.1)			
Employment status				1.3 (3)	.72	—
Working	2638 (95.8)	116 (4.2)	2754 (38.3)			
Studying	2392 (95.8)	106 (4.2)	2498 (34.8)			
Studying and working	1554 (95.1)	80 (4.9)	1634 (22.7)			
Not studying and not working	288 (95.7)	13 (4.3)	301 (4.2)			
Family structure in childhood				0.01 (1)	.96	—
Single-parent or other family	1886 (95.6)	86 (4.4)	1972 (27.4)			
2-parent family	4986 (95.6)	229 (4.4)	5215 (72.6)			
Parental divorce in childhood				0.1 (1)	.82	—
Parents are not divorced	2834 (95.5)	132 (4.5)	2966 (41.3)			
Parents are divorced	4038 (95.7)	183 (4.3)	4221 (58.7)			
Parental status				0.4 (1)	.54	—
Have no children	1159 (95.9)	49 (4.1)	1208 (16.8)			
Have at least one child	5713 (95.6)	266 (4.4)	5979 (83.2)			
Relationship status				21.3 (1)	<.001	0.05
Single	1562 (93.6)	107 (6.4)	1669 (23.2)			

XSL•FO
RenderX

Variables	Pornography-watching disorder absent (n=6872), n (%)	Pornography-watching disorder present (n=315), n (%)	Total sample (N=7187), n (%)	Chi-square (*df*)	*P* value	Φ[a]
In a relationship	5310 (96.2)	208 (3.8)	5518 (76.8)			
Sexuality						
Sexual orientation				2.4 (1)	.12	—
Nonheterosexual	551 (94.4)	33 (5.6)	584 (8.1)			
Heterosexual	6321 (95.7)	282 (4.3)	6603 (91.9)			
Use of sexual performance enhancer				8.5 (1)	.004	0.03
Not used yet	6140 (95.9)	265 (4.1)	6405 (89.1)			
Already used	732 (93.6)	50 (6.4)	782 (10.9)			
Sexual disorder				2.0 (1)	.16	—
No	5564 (95.8)	245 (4.2)	5809 (80.8)			
Yes	1308 (94.9)	70 (5.1)	1378 (19.2)			
Paraphilia				111.2 (1)	<.001	0.12
No	6791 (95.9)	288 (4.1)	7079 (98.5)			
Yes	81 (75)	27 (25)	108 (1.5)			
Conversation about the sex life				2.3 (1)	.13	—
No	613 (94.5)	36 (5.5)	649 (9)			
Yes	6259 (95.7)	279 (4.3)	6538 (91)			
Adequate and sufficient sexual education				7.3 (1)	.007	0.03
No	2041 (94.6)	116 (5.4)	2157 (30)			
Yes	4831 (96)	199 (4)	5030 (70)			
Sexual abuse				0.7 (1)	.39	—
No	6135 (95.5)	286 (4.5)	6,421 (89.3)			
Yes	737 (96.2)	29 (3.8)	766 (10.7)			
Pornography						
Frequency of pornography watching				438.9 (4)	<.001	0.25
Never in the last 1 year	1626 (99.7)	5 (0.3)	1631 (22.7)			
Less than monthly	1734 (99.2)	14 (0.8)	1748 (24.3)			
Monthly	1199 (97.2)	35 (2.8)	1234 (17.2)			
Weekly	2035 (91.5)	188 (8.5)	2223 (30.9)			
Daily	278 (79.2)	73 (20.8)	351 (4.9)			

[a]Φ: effect size (phi coefficient or Cramer *V*).

[b]Not applicable.

[c]ISCED: International Standard Classification of Education.

XSL•FO
RenderX

Table 2. Means of scales in the affected and non-affected subgroup of the study sample, and independent-samples t tests by variables in the study sample.

Variables	Pornography-watching disorder absent (n=6872), mean (SD)	Pornography-watching disorder present (n=315), mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Sociodemographic background					
Ideological orientation[a]	6.04 (2.08)	5.79 (2.23)	2.09 (7185)	.04	0.12
Religiosity[b]	1.94 (1.91)	2.21 (2.20)	–2.11 (7185)	.04	0.14
Social relationships					
Building personal relationships[c]	7.19 (2.11)	6.71 (2.37)	3.55 (7185)	<.001	0.23
Sexuality					
Age at first sexual intercourse	16.43 (1.91)	16.74 (2.18)	–2.54 (7185)	.01	0.17
Experience of first sexual intercourse[d]	6.22 (2.67)	6.61 (2.72)	–2.55 (7185)	.01	0.15
Number of sexual partnerships in the last year	1.87 (3.03)	2.07 (2.89)	–1.15 (7185)	.26	__[e]
Satisfaction with sex life[f]	6.75 (2.76)	5.58 (2.95)	6.89 (7185)	<.001	0.42
Satisfaction with genitals[f]	7.48 (2.36)	6.93 (2.49)	4.07 (7185)	<.001	0.23
Being sexually attractive[f]	6.54 (2.02)	6.25 (2.16)	2.32 (7185)	.02	0.14
Pornography					
Age at first encounter with pornographic content	13.98 (3.02)	12.56 (2.90)	8.22 (7185)	<.001	0.47
Experience of the first encounter with pornographic content[d]	6.20 (2.27)	7.09 (2.29)	–6.77 (7185)	<.001	0.39

[a]1=conservative; 10=liberal.

[b]1=not respecting religious norms at all; 10=fully respecting religious norms.

[c]1=with great difficulty; 10=very easily.

[d]1=very bad; 10=very good.

[e]Not applicable.

[f]1=not satisfied at all; 10=completely satisfied.

Responses to questions on a 10-point scale presented in Table 2 were not normally distributed. Nevertheless, after careful consideration, the means with SD and results of the 2-sample *t* tests are presented. The following were our rationale for this decision. First, a large literature argues that parametric tests can be robust to deviations from the normal distribution, that is, the distribution of the variable is not necessarily and not exclusively the relevant criterion [29-31]. One such criterion may be sample size, that is, for large samples (in our case N=7187), parametric tests are reliable even if the distribution is not perfectly normal. Second, for 5 scales (religiosity, building personal relationships, age at first sexual intercourse, satisfaction with sex life, and being sexually attractive), the variances can be said to differ according to the Levene test. This also means that the results of the Mann-Whitney test may not be robust and should be interpreted cautiously because the test looks at the difference in distributions and not just the medians. Third, the results of parametric tests are easier to interpret and, especially for large samples, parametric tests tend to have greater statistical power than nonparametric tests. The latter means that a nonparametric test is less likely to detect true deviations if they really exist. Fourth, we repeated our analysis using nonparametric tests (Mann-Whitney test) and reached the same conclusions as with the parametric tests presented in the manuscript (independent-sample *t* test). Furthermore, the same conclusions can be drawn from the effect sizes (Cohen *d* and r=rank-biserial correlation coefficient). Most of the respondents considered themselves rather liberal (mean 6.03, SD 2.08, 95% CI 5.98-6.08) and nonreligious, that is, they did not follow religious norms (mean 1.96, SD 1.92, 95% CI 1.91-2.00). They found it relatively easy to establish personal relationships with others (mean 7.17, SD 2.12, 95% CI 7.12-7.22). Their first sexual intercourse usually took place between the age of 16 and 17 years (mean 16.44, SD 1.92, 95% CI 16.40-16.48), which they look back on as a moderately good experience (6.23, SD 2.67, 95% CI 6.17-6.30). Their satisfaction with sexual life was also moderately good (mean 6.70, SD 2.78, 95% CI 6.64-6.76). Satisfaction with the own genital organs showed the highest average value (mean 7.46, SD 2.37, 95% CI 7.40-7.51). The respondents tended to consider themselves sexually attractive but not extraordinarily (mean 6.53, SD 2.03, 95% CI 6.48-6.58). On average, the respondents saw pornographic content first at the age of 14 (mean 13.92, SD 3.03, 95% CI 13.85-13.99),

XSL•FO
RenderX

which they had a moderately good experience with (mean 6.24, SD 2.28, 95% CI 6.19-6.29; Table 2).

Among the characteristics listed above, a significant difference was observed between those with PWD and those unaffected by this problem. Concerning all other examined characteristics, no significant differences were detected.

Table 3 displays the number and percentages of individuals who identified the statements related to PWD as true. The regression model's outcome variable is the presence (1) or absence (0) of PWD. The prevalence of PWD was 4.4% (315/7187). Due to the presence of multicollinearity tested using linear regression (tolerance index<0.20), we excluded 3 variables from the initial set of 32 variables (building partnerships, satisfaction with the own body, and frequency of masturbation) from the model. Thus, a total of 29 independent variables were included in the model. Of the 29 independent variables, 12 examine the sociodemographic background of the respondents, 1 examine their attitudes related to the formation of social relationships, 13 examine their sexual characteristics, and 3 are related to pornography.

Table 3. The number and percentage of "true" answers for the pornography-watching disorder-related statements among the respondents (N=7187).

Statement	Answers, n (%)
Pornography is often watched more times or over a longer period than was intended.	698 (9.7)
There is an unsuccessful effort to cut down or control pornography watching.	490 (6.8)
A great deal of time is spent watching pornography.	378 (5.3)
Craving, or a strong desire or urge to watch pornography.	1210 (16.8)
Recurrent pornography watching resulting in a failure to fulfill major role obligations at work, school, or home.	167 (2.3)
Continued pornography watching despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of pornography.	162 (2.3)
Important social, occupational, or recreational activities are given up or reduced because of pornography watching.	62 (0.9)
Recurrent pornography watching in situations in which it is hazardous (eg, in public places, workplaces, and classrooms).	409 (5.7)
Pornography watching is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by pornography.	912 (12.7)
Needs to watch an increasing amount or more extreme types of pornography in order to achieve the desired excitement.	185 (2.6)

Based on the likelihood ratio chi-square test, the final binary logistic regression model run to further analyze the relationship between PWD and the background variables examined is significant (for n=7187, χ^2_{13}=510.6, P<.001); the Hosmer and Lemeshow tests also indicate a suitable model fit (for n=7187, χ^2_8=7.8, P=.46). Based on the Nagelkerke R^2 indicator (0.23), the variable set in the final model explains 22.7% of the variance. Based on the receiver operating characteristics curve analysis, which indicates the probability of true positive classification (sensitivity) as a function of true negative classification (specificity), the area under the curve is 0.84 (cutoff value=0.50; P<.001; 95% CI 0.82-0.86), indicating excellent discrimination [27].

In the final model after variable selection, 10 variables were found to be significant (Table 4).

The most powerful factor in the model was the frequency of watching porn. Compared with those who watched porn daily, the odds of having PWD were significantly lower among those who did not watch porn daily (never in the last 1 year, odds ratio [OR] 0.02, 95% CI 0.01-0.06; less than monthly, OR 0.05, 95% CI 0.03-0.10; monthly, OR 0.18, 95% CI 0.11-0.28; and weekly, OR 0.45, 95% 0.33-0.62). A less powerful but still strongly influential factor was biological sex: females were half as likely as males to have the PWD (OR 0.53, 95% CI 0.39-0.72). The other significant variables are of relatively minor importance. The chance of occurrence of PWD is reduced if the participants encountered pornographic content later (OR 0.94, 95% CI 0.90-0.98), were more satisfied with their sexual life (OR 0.94, 95% CI 0.90-0.98), established personal relationships with others more easily (OR 0.93, 95% CI 0.88-0.98), received adequate sexual information before their first sexual intercourse (OR 0.67, 95% CI 0.53-0.87), or lived in the capital (OR 0.52, 95% CI 0.30-0.91). On the other hand, the chance of developing PWD increased if the respondents were affected by some paraphilia (OR 3.95, 95% CI 2.37-6.56) or sexual disorder (OR 2.01, 95% CI 1.45-2.80) and if they followed the norms of their religion as much as possible (OR 1.12, 95% CI 1.06-1.19).

In summary, the predisposing factors for PWD were higher frequency of pornography watching, male sex, occurrence of paraphilia, an early encounter with pornographic content, lower level of satisfaction with sexual life, difficulties in establishing personal relationships, occurrence of a sexual disorder, stronger adherence to religious norms, inappropriate sexual education, and living in the capital.

XSL•FO
RenderX

Table 4. Results of binary logistic regression for predicting the likelihood of pornography-watching disorder.

Variables	Odds ratio[a] (95% CI)	*P* value	Relative importance (%)
Frequency of pornography watching (Reference: daily)			53.94
Never in the last 1 year	0.02 (0.01-0.06)	<.001	
Less than monthly	0.05 (0.03-0.10)	<.001	
Monthly	0.18 (0.11-0.28)	<.001	
Weekly	0.45 (0.33-0.62)	<.001	
Biological sex (Reference: male)	0.53 (0.39-0.72)	<.001	17.41
Paraphilia (Reference: absent)	3.95 (2.37-6.56)	<.001	7.48
Age at first encounter with pornographic content	0.94 (0.90-0.98)	.006	6.84
Satisfaction with sex life[b]	0.94 (0.90-0.98)	.006	5.59
Building personal relationships[c]	0.93 (0.88-0.98)	.005	2.13
Sexual disorder (Reference: absent)	2.01 (1.45-2.80)	<.001	1.90
Religiosity[d]	1.12 (1.06-1.19)	<.001	1.70
Adequate and sufficient sexual education (Reference: absent)	0.67 (0.53-0.87)	.002	1.51
Place of residence (Reference: Capital)	0.52 (0.30-0.91)	.02	1.50

[a]Values of odds ratio above 1 mean a higher likelihood of inclusion, and values below 1 mean a lower likelihood of inclusion compared with the given reference group. In the table, the variables are listed in descending order of relative importance. Relative importance was determined using dominance analysis.

[b]1=not satisfied at all, 10=completely satisfied.

[c]1=with great difficulty, 10=very easily.

[d]1=not respecting religious norms at all, 10=fully respecting religious norms.

Discussion

Principal Findings

This study found that 4.4% of young adults in Hungary exhibit signs of PWD. Key factors contributing to PWD include frequent pornography use, being male, having paraphilic interests, early exposure to pornography, and dissatisfaction with sexual life. Furthermore, individuals who struggle to form personal relationships, experience sexual disorders, strongly adhere to religious norms, received inadequate sexual education, or live in the capital were also more prone to developing PWD.

Our study identified several important risk factors for PWD. The most significant factor was the frequency of pornography consumption, with daily viewers being much more likely to develop PWD than those who watched less often. Sex was also a critical factor, as males were at higher risk. Furthermore, early exposure to pornography and the presence of paraphilia were strongly linked to PWD. In contrast, higher satisfaction with sexual life and greater ease in forming personal relationships were associated with a lower likelihood of developing PWD, indicating these factors might offer some protection against the disorder.

Comparison With Previous Work

A large-scale survey study of 18,034 participants on gender, sexual orientation, and hypersexuality found a significant association between problematic pornography use and sexual identity because the prevalence was higher among nonheterosexuals than in heterosexuals. The results suggested that GBTQ+ (gay, bisexual, trans, queer, questioning, non-binary or otherwise gender or sexuality-diverse) males have a higher risk of engaging in hypersexual behavior (including high-frequency porn watching) compared with their heterosexual counterparts [32]. The results of our study highlight the role of biological sex, which, in our case, is more relevant than sexual orientation, although the dual association of gender with sexual identity and PWD was not examined.

In connection to religion and ideology, our results suggested a higher prevalence of PWD associated with being religious and having more conservative than liberal views. Both connections were statistically significant. An American study from 2016, on a sample of 713 adult participants, found a higher prevalence of SPPA in religious participants versus nonreligious ones [33], which is in accordance with the current study's findings.

Regarding relationship status (single or in a relationship), a higher prevalence of PWD was found among singles than among those living in a relationship. Kraus et al [34] examined different characteristics of men interested in finding a treatment for their pornography use within a sample of 1298 male participants and found that more single or unmarried individuals were seeking treatment.

Also related to relationship factors was that people affected with PWD found it, on average, more difficult to form any personal relationship than the control group. Those dissatisfied with sexual education, that is, information obtained before the first sexual intercourse, were also more likely to develop the PWD.

XSL•FO
RenderX

These 2 phenomena have apparently not been described in the literature.

Regarding satisfaction with sexual life, the results indicated a significant connection, namely, problematic pornography use was associated with lower levels of satisfaction with sexual life. Results from a study performed on a sample of 832 adults on SPPA, sexual function, and sexual satisfaction indicated that SPPA was associated with sexual dysfunctions and lower sexual satisfaction [35].

Strengths and Limitations

This study is among the first to apply *DSM-5* criteria to assess PWD in a large sample of young adults, providing a clinically relevant approach to understanding this condition. The large sample size and the inclusion of diverse sociodemographic and psychological variables contribute to the robustness of our findings.

A key strength of the study is the use of a self-report web-based questionnaire with ensured anonymity, which likely reduces the influence of stigma and taboo associated with discussing sensitive topics such as sexuality and pornography. This method also allows for the inclusion of a larger and more diverse sample, enhancing the generalizability of the findings and enabling the application of the modified *DSM-5* criteria to a broader population.

However, the study has several limitations. First, its cross-sectional design limits the ability to draw causal inferences; while associations between certain factors and PWD were identified, cause-and-effect relationships cannot be established.

Second, the study was limited to young adults in Hungary, which may affect the generalizability of the findings to other populations. Cultural differences in attitudes toward pornography and sexual behavior could influence both the prevalence and the risk factors for PWD, suggesting a need for future studies that include participants from diverse geographic locations.

Third, the reliance on self-reported data introduces potential bias, as participants may underreport or overreport their behavior due to social desirability or stigma. Although the anonymity of the survey helps mitigate these issues, future research should consider incorporating more objective measures of pornography consumption and related behaviors.

Fourth, the study did not distinguish between mild, moderate, and vigorous forms of PWD, limiting the granularity of the findings. Future research could benefit from examining the full spectrum of PWD severity to provide a more nuanced understanding of the disorder.

Finally, the use of the Cantril ladder scale for measuring subjective attitudes instead of validated question blocks was chosen to prevent respondent fatigue and maintain the quality of responses. However, this choice, along with differences in data-collection methods compared with other studies, may contribute to variability in the results. Despite these challenges, the study's methodological approach, particularly the anonymity provided to respondents, remains a valuable tool in exploring sensitive topics like PWD.

Future Directions

Future research should investigate the longitudinal relationships between early exposure to pornography and the development of PWD to gain a clearer understanding of causality. Furthermore, studies involving more geographically and culturally diverse samples are essential to determine if the risk factors identified in this study are applicable to broader populations. Finally, the development and validation of standardized diagnostic tools for PWD are crucial for advancing our understanding and treatment of this disorder.

Conclusions

This study characterizes PWD and its prevalence in a large sample population using adapted *DSM-5* criteria based on a web-based survey. The information obtained from this study can be used as a base for further research on pornography consumption and can be used as part of public health prevention programs to advance sexual well-being. In addition, the methods and results of the study can be used to standardize other pornography scales and possibly to supply some evidence required to officially recognize pornography addiction as a disorder. The questionnaire used in this study might enable to development of a web-based interface (website or app) that measures and evaluates the user's porn consumption and offers them a quiet assistance program in case of need.

Acknowledgments

This study would not have been possible without the support of the University Szeged Albert Szent-Györgyi Medical School.

The manuscript publication was funded by University of Szeged Open Access Fund (grant 6322).

This research did not involve any costs or funding. All authors met the conditions described in the recommendations by the International Committee of Medical Journal Editors Recommendations and gave final approval to the submitted paper.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to the fact that this study was conducted using a portion of the database, and ongoing analyses and publications are based on the database, which is in continuous use, but data are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization (study concept and design) was contributed by CE, DP, and OK. Methodology was contributed by CE and DP. Software (developing online questionnaire) was handled by CE and DP. Formal Analysis (statistical analyses) was contributed by KS and CE. Resources was managed by CE and AG. Writing—original draft preparation was contributed by CE, KS, and AG. Writing—Review and editing was contributed by EP, KS, OK, and AP. Supervision was managed by EP, OK, AP, and EH. All authors have read and agreed on the final version of the manuscript for submission.

Conflicts of Interest

None declared.

References

1. Pornhub. (2023). The 2022 Year in review. URL: https://www.pornhub.com/insights/2022-year-in-review#top-20-countries [accessed 2023-01-10]
2. Cooper A, Delmonico DL, Burg R. Cybersex users, abusers, and compulsives: new findings and implications. Sexual Addiction & Compulsivity 2000;7(1-2):5-29. [doi: 10.1080/10720160008400205]
3. Zattoni F, Gül M, Soligo M, Morlacco A, Motterle G, Collavino J, et al. The impact of COVID-19 pandemic on pornography habits: a global analysis of Google trends. Int J Impot Res 2020;33(8):824-831 [FREE Full text] [doi: 10.1038/s41443-020-00380-w] [Medline: 33249423]
4. Miller DJ, Raggatt PTF, McBain K. A literature review of studies into the prevalence and frequency of men's pornography Use. American Journal of Sexuality Education 2020;15(4):502-529. [doi: 10.1080/15546128.2020.1831676]
5. Mestre-Bach G, Blycker G, Potenza M. Pornography use in the setting of the COVID-19 pandemic. J Behav Addict 2020;9(2):181-183 [FREE Full text] [doi: 10.1556/2006.2020.00015] [Medline: 32663384]
6. Harper C, Hodgins DC. Examining correlates of problematic internet pornography use among university students. J Behav Addict 2016;5(2):179-191 [FREE Full text] [doi: 10.1556/2006.5.2016.022] [Medline: 27156383]
7. Levi G, Cohen C, Kaliche S, Sharaabi S, Cohen K, Tzur-Bitan D, et al. Sexual addiction, compulsivity, and impulsivity among a predominantly female sample of adults who use the internet for sex. J Behav Addict 2020;9(1):83-92 [FREE Full text] [doi: 10.1556/2006.2020.00007] [Medline: 32359233]
8. Toates F. A motivation model of sex addiction - relevance to the controversy over the concept. Neurosci Biobehav Rev 2022;142:104872. [doi: 10.1016/j.neubiorev.2022.104872] [Medline: 36113783]
9. Shimoni L, Dayan M, Cohen K, Weinstein A. The contribution of personality factors and gender to ratings of sex addiction among men and women who use the Internet for sex purpose. J Behav Addict 2018;7(4):1015-1021 [FREE Full text] [doi: 10.1556/2006.7.2018.101] [Medline: 30378460]
10. Montag C, Reuter M. Internet addiction: Neuroscientific Approaches and Therapeutical Interventions. Cham: Springer; 2017.
11. Reid RC, Carpenter BN, Hook JN, Garos S, Manning JC, Gilliland R, et al. Report of findings in a DSM-5 field trial for hypersexual disorder. J Sex Med 2012;9(11):2868-2877. [doi: 10.1111/j.1743-6109.2012.02936.x] [Medline: 23035810]
12. Shrivastava T, Agarwal P, Vora V, Sethi Y. Aggravation of obsessive-compulsive disorder due to excessive porn consumption: a case report. Cureus 2022;14(12):e33018 [FREE Full text] [doi: 10.7759/cureus.33018] [Medline: 36721539]
13. Duffy A, Dawson DL, das Nair R. Pornography addiction in adults: a systematic review of definitions and reported impact. J Sex Med 2016;13(5):760-777. [doi: 10.1016/j.jsxm.2016.03.002] [Medline: 27114191]
14. Kafka MP. Hypersexual disorder: a proposed diagnosis for DSM-V. Arch Sex Behav 2010;39(2):377-400. [doi: 10.1007/s10508-009-9574-7] [Medline: 19937105]
15. Love T, Laier C, Brand M, Hatch L, Hajela R. Neuroscience of internet pornography addiction: a review and update. Behav Sci (Basel) 2015;5(3):388-433 [FREE Full text] [doi: 10.3390/bs5030388] [Medline: 26393658]
16. World Health Organization.. ICD-11: International Classification of Diseases. Geneva: World Health Organization; 2022.
17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders DSM-5. Arlington, Va: American Psychiatric Publishing; 2013.
18. Darshan MS, Sathyanarayana Rao TS, Manickam S, Tandon A, Ram D. A case report of pornography addiction with dhat syndrome. Indian J Psychiatry 2014;56(4):385-387 [FREE Full text] [doi: 10.4103/0019-5545.146536] [Medline: 25568482]
19. Dwulit AD, Rzymski P. Prevalence, patterns and self-perceived effects of pornography consumption in polish university students: a cross-sectional study. Int J Environ Res Public Health 2019;16(10):1861 [FREE Full text] [doi: 10.3390/ijerph16101861] [Medline: 31137778]
20. Bőthe B, Tóth-Király I, Zsila ?, Griffiths MD, Demetrovics Z, Orosz G. The development of the problematic pornography consumption scale (PPCS). J Sex Res 2018;55(3):395-406. [doi: 10.1080/00224499.2017.1291798] [Medline: 28276929]
21. Delmonico DL, Miller JA. The internet sex screening test: a comparison of sexual compulsives versus non-sexual compulsives. Sexual and Relationship Therapy 2003;18(3):261-276. [doi: 10.1080/1468199031000153900]
22. Grubbs JB, Sessoms J, Wheeler DM, Volk F. The cyber-pornography use inventory: the development of a new assessment instrument. Sexual Addiction & Compulsivity 2010;17(2):106-126. [doi: 10.1080/10720161003776166]

XSL•FO
RenderX

23. Grubbs JB, Volk F, Exline JJ, Pargament KI. Internet pornography use: perceived addiction, psychological distress, and the validation of a brief measure. J Sex Marital Ther 2015;41(1):83-106. [doi: 10.1080/0092623X.2013.842192] [Medline: 24341869]
24. Voon V, Mole TB, Banca P, Porter L, Morris L, Mitchell S, et al. Neural correlates of sexual cue reactivity in individuals with and without compulsive sexual behaviours. PLoS One 2014;9(7):e102419 [FREE Full text] [doi: 10.1371/journal.pone.0102419] [Medline: 25013940]
25. Cantril H. The Pattern of Human Concerns. New Brunswick: Rutgers University Press; 1965.
26. Erdős C, Kelemen O, Pócs D, Horváth E, Dudás N, Papp A, et al. Female sexual dysfunction in association with sexual history, sexual abuse and satisfaction: a cross-sectional study in hungary. J Clin Med 2023;12(3):1112 [FREE Full text] [doi: 10.3390/jcm12031112] [Medline: 36769759]
27. Hosmer DW, Lemeshow S, Sturdivant RX. Model-building strategies and methods for logistic regression. In: Applied logistic regression. Hoboken, New Jersey: Wiley; 2013:89-151.
28. Azen R, Traxel N. Using dominance analysis to determine predictor importance in logistic regression. Journal of Educational and Behavioral Statistics 2009;34(3):319-347. [doi: 10.3102/1076998609332754]
29. Field A. Discovering Statistics Using IBM SPSS Statistics. London: SAGE Publications; 2018.
30. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151-169. [doi: 10.1146/annurev.publhealth.23.100901.140546] [Medline: 11910059]
31. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract 2010;15(5):625-632. [doi: 10.1007/s10459-010-9222-y] [Medline: 20146096]
32. Bőthe B, Bartók R, Tóth-Király I, Reid RC, Griffiths MD, Demetrovics Z, et al. Hypersexuality, gender, and sexual orientation: a large-scale psychometric survey study. Arch Sex Behav 2018;47(8):2265-2276. [doi: 10.1007/s10508-018-1201-z] [Medline: 29926261]
33. Bradley DF, Grubbs JB, Uzdavines A, Exline JJ, Pargament KI. Perceived addiction to internet pornography among religious believers and nonbelievers. Sexual Addiction & Compulsivity 2016;23(2-3):225-243. [doi: 10.1080/10720162.2016.1162237]
34. Kraus SW, Voon V, Potenza MN. Should compulsive sexual behavior be considered an addiction? Addiction 2016;111(12):2097-2106 [FREE Full text] [doi: 10.1111/add.13297] [Medline: 26893127]
35. Blais-Lecours S, Vaillancourt-Morel MP, Sabourin S, Godbout N. Cyberpornography: time use, perceived addiction, sexual functioning, and sexual satisfaction. Cyberpsychol Behav Soc Netw 2016;19(11):649-655. [doi: 10.1089/cyber.2016.0364] [Medline: 27831753]

Abbreviations

CPUI: Cyber-Pornography Use Inventory
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
GBTQ+: gay, bisexual, trans, queer, questioning, non-binary or otherwise gender or sexuality-diverse
ICD-11: International Classification of Diseases for Mortality and Morbidity Statistics Eleventh Revision Eleventh Revision
ISCED: International Standard Classification of Education
ISST: Internet Sex Screening Test
MRI: magnetic resonance imaging
OR: odds ratio
PPCS: Problematic Pornography Consumption Scale
PWD: pornography-watching disorder
SPPA: self-perceived pornography addiction

Edited by N Cahill; submitted 12.06.23; peer-reviewed by I Hanes, N Muhammad, A Csinady; comments to author 08.03.24; revised version received 06.05.24; accepted 15.10.24; published 08.01.25.

Please cite as:
Erdős C, Kelemen O, Pócs D, Paulik E, Papp A, Horváth E, Golan A, Széll K
Pornography-Watching Disorder and Its Risk Factors Among Young Adults: Cross-Sectional Survey
J Med Internet Res 2025;27:e49860
URL: https://www.jmir.org/2025/1/e49860
doi:10.2196/49860
PMID:

XSL•FO
RenderX

©Csaba Erdős, Oguz Kelemen, Dávid Pócs, Edit Paulik, András Papp, Edina Horváth, Arbel Golan, Krisztián Széll. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Comparison of Mobile Social Media Promotion and Volunteer-Driven Strategies for Community Organizations Recruiting Men Who Have Sex with Men for HIV Testing in Zhejiang Province, China: Cross-Sectional Study Based on a Large-Scale Survey

Lin He[1], MA; Shaoqiang Jiang[2], BA; Tingting Jiang[1], MA; Wanjun Chen[1], MA; Jinlei Zheng[1], MA; Hui Wang[1], PhD; Chengliang Chai[1], MA

[1]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[2]Coastal Service Center, Hangzhou, China

Corresponding Author:
Chengliang Chai, MA
Zhejiang Provincial Center for Disease Control and Prevention
3399 Bin Sheng Road, Binjiang District
Hangzhou, 310051
China
Phone: 86 87115193
Fax: 86 87115194
Email: chlchai@cdc.zj.cn

Abstract

Background: China has recently implemented a strategy to promote and facilitate community organization involvement in HIV prevention among men who have sex with men (MSM). Although community-based strategies have been shown to increase HIV testing uptake, the relative effectiveness of mobile social media promotion compared with volunteer-driven recruitment remains underexplored. Limited research has investigated how these strategies differentially affect MSM who have not undergone previous HIV testing.

Objective: This study aimed to compare the differences between a mobile social media promotion strategy and a volunteer-driven strategy for community organizations to recruit MSM for HIV testing.

Methods: A cross-sectional study was conducted from July to December 2023 among MSM in Zhejiang Province, China. Participants aged 16 years with an HIV-negative or unknown status were recruited either through a mobile social media promotion strategy or through a volunteer-driven strategy by a community organization. They completed a questionnaire that collected information on demographics, sexual behavior, and HIV testing history. All participants were tested for HIV after completing the questionnaire. A multivariate logistic regression model was used to identify factors associated with recruitment through mobile social media promotion.

Results: The study included 4600 MSM, of whom 3035 (66%) were recruited through the mobile social media strategy. Overall, 1.4% (66/4600) of participants tested positive for HIV, and 18.8% (865/4600) underwent HIV testing for the first time. Recruitment via the mobile social media promotion strategy was significantly associated with several factors: having only gay sexual partners (adjusted OR [aOR] 1.23, 95% CI 1.05-1.45), having more than 2 sexual partners in the past 3 months (aOR 1.74, 95% CI 1.42-2.11), frequently using rush poppers during sex (aOR 1.39, 95% CI 1.14-1.99), having a history of sexually transmitted infections (aOR 1.56, 95% CI 1.02-2.39), having awareness of pre-exposure prophylaxis (aOR 1.42, 95% CI 1.19-1.71), having awareness of postexposure prophylaxis (PEP; aOR 1.49, 95% CI 1.24-1.79), using mail-in HIV self-testing kits (aOR 2.02, 95% CI 1.77-2.31), testing HIV-positive (aOR 2.02, 95% CI 1.10-3.72), and first-time HIV testing (aOR 1.28, 95% CI 1.09-1.52).

Conclusions: Community organizations play a critical role in expanding HIV testing and identifying undiagnosed individuals infected with HIV. Compared to the volunteer-driven outreach, mobile social media promotion strategies had a higher proportion of first-time testers and a higher rate of HIV positivity. We recommend prioritizing mobile social media strategies in regions with limited LGBTQ+ organizations or HIV health services to increase HIV testing coverage and interventions among MSM.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e66702) doi:10.2196/66702

KEYWORDS

men who have sex with men; MSM; internet; recruit; HIV testing; community organization; strategy; China; mobile phone

Introduction

Men who have sex with men (MSM) represent a significant proportion of individuals living with HIV in China. According to recent estimates, the national prevalence of HIV among MSM was 5.7% in 2018 [1] and rose to 7% in 2022 [2]. Moreover, the percentage of newly diagnosed HIV infections attributed to homosexual transmission increased from 2.5% in 2006 to 25.6% in 2022 [2]. To achieve the United Nations Program on HIV/AIDS 95-95-95 targets (95% diagnosis, 95% treatment, and 95% viral suppression) by 2030, a renewed focus on MSM is essential [3]. While China is on track to meet the second and third 95% targets (treatment and viral suppression) [4], achieving the goal of diagnosing 95% of individuals infected with HIV by 2030 remains a significant challenge. Our previous study demonstrated that 12.9% of MSM will have never undergone HIV testing by 2023 [5]. It is, therefore, crucial to identify MSM who have not been tested, implement targeted HIV education and testing, and expand testing efforts to facilitate early diagnosis. This approach is pivotal for China to meet its 2030 diagnostic target.

Historically, MSM in China has faced significant barriers to HIV awareness, intervention, and counseling services. Traditional social norms and cultural values make it difficult for MSM to disclose their sexual orientation [6], leading many to conceal their identity. This hidden population poses a challenge for the current HIV service network, which is managed by the Centers for Disease Control and Prevention (CDC) and various medical institutions [7,8]. In recent years, China has implemented a series of comprehensive measures aimed at preventing and controlling HIV infections among MSM. One of the primary strategies is to promote and facilitate community organizational involvement in HIV prevention and control initiatives [2]. These encompass the implementation of public education initiatives, disease prevention and intervention strategies, and the provision of testing and counseling services. The community organization is composed of MSM volunteers, which allows for forming trusting relationships and facilitates contact with the MSM population, thus bringing attention to this previously hidden demographic [9,10]. A previous study revealed that 567 MSM community-based organizations in China participated in initiatives funded by the AIDS fund for nongovernmental organizations, facilitating HIV testing for over 250,000 people by 2022 [11]. Of these, 7963 were newly diagnosed, representing 28.9% of all MSM HIV cases nationwide. The involvement of community organizations has increased the reach of MSM communities, facilitating greater access to HIV prevention services and the identification of new infections [12].

HIV testing is a highly effective intervention strategy [13]. The implementation of HIV testing can facilitate early diagnosis, prompt treatment, and subsequent transmission reduction [14]. Studies have shown that MSM who have been tested for HIV are less likely to engage in high-risk sexual behaviors compared to those who have never been tested [15]. Consequently, a significant proportion of funding allocated to community organizations in China is directed toward HIV testing services [9]. Community organizations primarily use 3 methods to recruit MSM for HIV testing. The initial approach was a volunteer-driven strategy that involved active recruitment. Community-organized volunteers and peer educators were recruited from venues frequented by MSM, including parks, bars, and bathhouses, for in-person HIV testing and interventions. In addition, MSM who have not been tested for HIV for over 6 months are also mobilized for testing by these volunteers. Second, a mobile social media promotion or recruitment strategy is also used. Information on HIV promotion, prevention, and testing is disseminated through MSM websites, dating apps, and communication groups, which are then used to recruit MSM to offline service outlets. Alternatively, HIV self-testing kits may be ordered mobile and mailed home. Third, a regular testing strategy is implemented. Some MSM who were initially recruited through the aforementioned strategies continue to test regularly after receiving HIV education without the need for further mobilization by community organizations.

With the rapid growth of China's internet economy and the widespread use of smartphones, mobile dating apps for MSM have surged in popularity [10]. Many MSM now use the internet and dating apps to find sexual partners. This shift has changed HIV testing recruitment from in-person volunteer outreach to internet-based strategies. Studies have shown that the internet is an effective tool for recruiting MSM, especially those who have never been tested, thus expanding the reach of HIV testing and reducing the number of undiagnosed infections [16,17]. Zhejiang Province, an economically developed and thriving internet economy in China, relies heavily on community organizations to implement HIV prevention and provide testing for MSM. In 2023, approximately 66,000 HIV tests have been conducted among MSM in the province, accounting for about 10% of the national total [4]. Notably, over 90% of these tests are performed by community organizations. To improve testing coverage among MSM, these organizations must use a dual recruitment strategy that uses both mobile social promotion and volunteer-driven initiatives. This approach is essential for identifying and engaging MSM who have not yet been tested for HIV. This study, based on the findings of a large-scale survey, aimed to analyze the effectiveness of 2 recruitment strategies, which are mobile social media promotion and volunteer-driven outreach. The results of this analysis will provide a foundation for promoting HIV testing within communities and contribute to recruiting MSM who have never been tested for HIV in China.

XSL•FO
RenderX

Methods

Study Design

A cross-sectional study based on a large-scale survey was conducted among MSM between July and December 2023.

Study Participants

MSM were eligible to participate if they met certain criteria, which include being (1) 16 years or older, (2) reported having sex with men in the past year, (3) were HIV-negative or had an unknown HIV status, (4) received HIV testing as a direct result of mobile social media promotion or volunteer-driven recruitment, and (5) resided in Zhejiang Province, China.

Participant Recruitment and Data Collection

Sunshine Coast Public Welfare is a social service agency that employs 13 full-time and 22 part-time social workers, with over 400 registered volunteers. The organization established a digital HIV prevention service using location-based services. MSM in Zhejiang Province are invited to visit the official "Sunshine Test" account on WeChat, a popular Chinese communication platform, to request free HIV testing and prevention counseling. MSM can choose between mailed self-testing kits or visiting offline services for HIV testing. Full-time and part-time social workers are responsible for regularly recruiting a specified number of MSM for HIV testing. Meanwhile, registered volunteers assist in promoting HIV testing within the MSM community periodically.

Questionnaires

All individuals who engaged with Sunshine Test were required to complete a routine surveillance questionnaire consisting of 20 questions. These questions covered demographic information such as age, marital status, and education level, as well as sexual behavior including sexual roles (receptive, insertive, or both), number of sexual partners, use of rush poppers, sexual history, awareness of pre-exposure prophylaxis (PrEP) and PEP (postexposure prophylaxis), and HIV testing history.

Recruitment through direct engagement by social workers or registered volunteers was classified as "volunteer-driven recruitment." Conversely, if MSM were recruited after encountering HIV prevention information posted by the community organization on internet platforms (such as dating apps, communication groups, or websites) and subsequently applied for testing, this was categorized as "mobile social media promotion recruitment." If MSM had undergone an HIV test at the organization of their own volition, without being prompted by volunteers or mobile social media promotion, their case was classified as a "regular test." During the study, the participating MSM underwent repeated HIV testing on more than 2 occasions, but only the first record was included in the analysis. Details of the Sunshine Coast Public Welfare Program have been described in previous literature [10].

Statistical Analysis

Descriptive analyses were performed, with categorical variables presented as frequencies and proportions and continuous variables presented as median with IQR or mean with SD. The chi-square test (X^2) was used to assess significant differences in demographic characteristics between MSM recruited through mobile social media promotion and those recruited through volunteer-driven recruitment. Factors associated with differences in the sociodemographic characteristics of mobile social media promotion recruitment were included in multivariate logistic regression models, and adjusted odds ratios (aORs) with 95% CIs were calculated. Statistical significance was set at $P<.05$, with $\beta=.1$. All statistical analyses were conducted using SPSS (version 19.0; IBM Corp).

Ethical Considerations

This study was approved by the Zhejiang Provincial Centre for Disease Control and Prevention (approval number 2022-011-01). Written informed consent was obtained from all participants before survey completion, and participants had the ability to drop out anytime during the survey. All data were anonymous; participants received free HIV testing, health counseling, and no compensation; and all study procedures were conducted in accordance with approved guidelines and regulations.

Results

A total of 4600 MSM were enrolled in the study (Figure 1), of whom 3035 (66%) were recruited via mobile social media promotion and 1565 (34%) through volunteer-driven recruitment. The mean age of the participants was 30.0 (SD 9) years. Overall, 78.4% (3606/4600) were single, and 74.5% (3425/4600) had a college degree or higher education. Most participants (3455/4600, 75.1%) reported having male sexual partners. Over the past 3 months, 30.8% (1419/4600) reported having more than 2 sexual partners, and 30.7% (1410/4600) reported using rush poppers during sexual activity. In addition, 43.6% (2007/4600) were aware of their sexual partner's HIV status, 2.7% (126/4600) had a history of sexually transmitted infections (STIs), 47.2% (2169/4600) had used a mailed self-testing kit for HIV testing, 1.4% (66/4600) tested positive for HIV, and 18.8% (865/4600) were undergoing HIV testing for the first time (Table 1).

XSL•FO
RenderX

Figure 1. Study design. MSM: men who have sex with men.

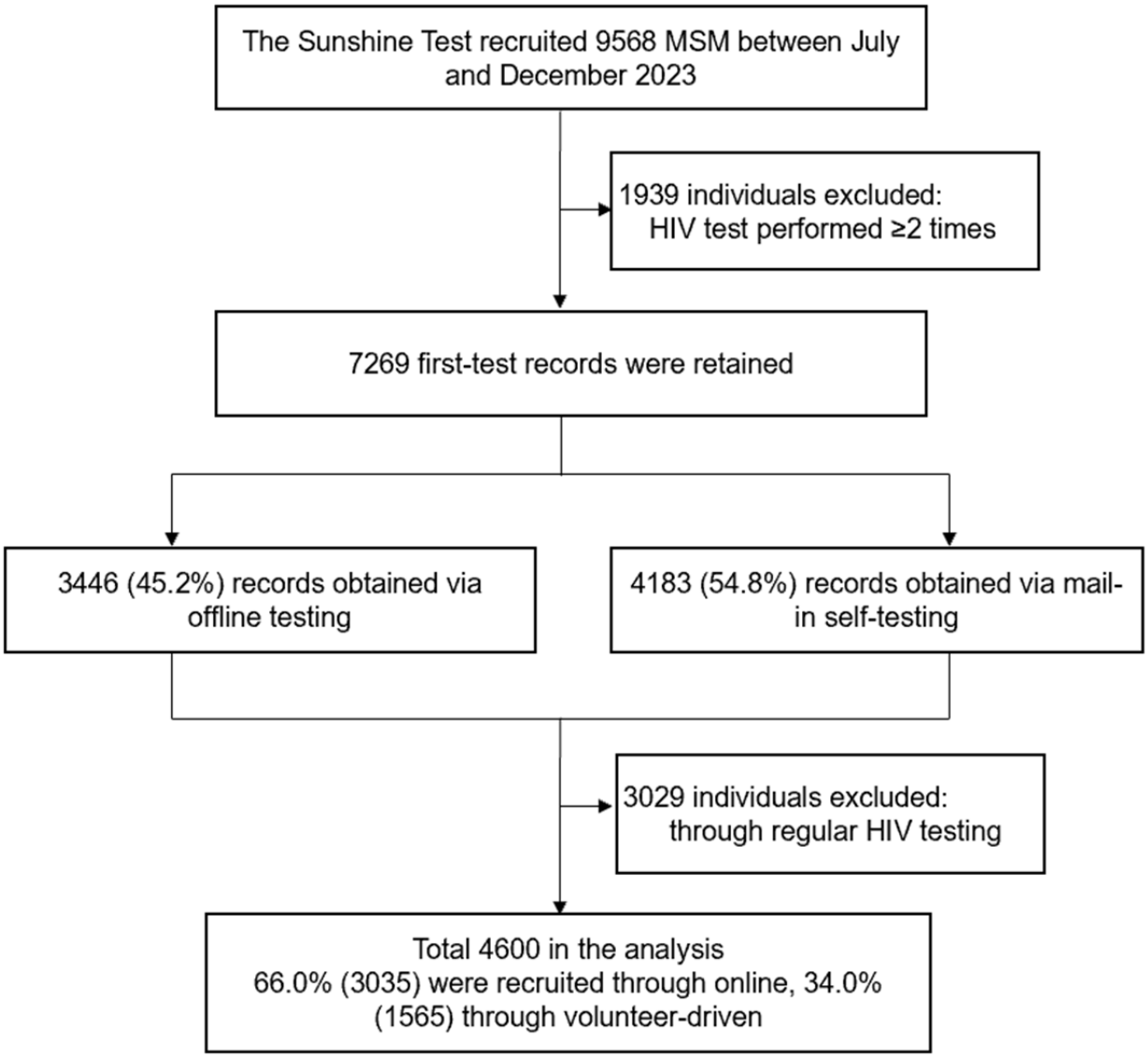

Overall, 66% (3035/4600) of MSM recruited through mobile social media promotion were aged 20-29 years (1842/2719, 67.7%), single (2404/3606, 66.7%), had a master's degree or higher education (283/424, 66.7%), reported insertive anal sex (1037/1512, 68.6%), had exclusively gay sexual partners (2394/3455, 69.3%), had more than 2 sexual partners in the past 3 months (1001/1419, 70.5%), frequently used rush poppers during sex (529/698, 75.8%), were consistently aware of their sexual partner's HIV status (1371/2007, 68.3%), had a history of STIs (96/126, 76.2%), had HIV testing through mail reagent self-testing (1592/2169, 73.4%), tested HIV-positive (52/66, 78.8%), and had been previously tested for HIV (592/865, 68.4%) (Table 1).

XSL•FO
RenderX

Table 1. Social demographic and behavioral characteristics among men who have sex with men who were recruited through mobile social media promotion and volunteer-driven recruitment strategy

Variables	Participants, n (%)	Number of mobile recruitments, n (%)	Number of volunteer-driven recruitment, n (%)	Chi-square (*df*)	*P* value
Overall	4600 (100)	3035 (66.0)	1565 (34.0)		
Age (years), IQR 27 (23-32)				9.5 (2)	.008
<20	306 (6.7)	190 (62.1)	116 (37.9)		
20-29	2719 (59.1)	1842 (67.7)	877 (32.3)		
30 years or older	1575 (34.2)	1003 (63.7)	572 (36.3)		
Marital status				3.5 (2)	.172
Single	3606 (78.4)	2404 (66.7)	1202 (33.3)		
Married	857 (18.6)	544 (63.5)	313 (36.5)		
Divorced or separated	137 (3.0)	87 (63.5)	50 (36.5)		
Education level				1.2 (3)	.750
High school and below	1175 (25.5)	760 (64.7)	415 (35.3)		
College	1175 (25.5)	779 (66.3)	396 (33.7)		
Bachelor’s degree	1826 (39.7)	1213 (66.4)	613 (33.6)		
Master's degree or above	424 (9.2)	283 (66.7)	141 (33.3)		
Occupation				15.9 (3)	.001
Company employee	1605 (34.9)	1111 (69.2)	494 (30.8)		
Student	836 (18.2)	555 (66.4)	281 (33.6)		
Freelance	612 (13.3)	374 (61.1)	238 (38.9)		
Others	1547 (33.6)	995 (64.3)	552 (35.7)		
Sex roles				11.3 (2)	.004
Receptive anal sex	1003 (21.8)	675 (67.3)	328 (32.7)		
Both	2085 (45.3)	1323 (63.5)	762 (36.5)		
Insertive anal sex	1512 (32.9)	1037 (68.6)	475 (31.4)		
Sexuality of partner				67.9 (1)	<.001
Gay	3455 (75.1)	2394 (69.3)	1061 (30.7)		
Bisexual	1145 (24.9)	641 (56.0)	504 (44.0)		
Number of sexual partners in the past 3 months					
0	903 (19.6)	463 (51.3)	440 (48.7)	109.2 (2)	<.001
1	2278 (49.5)	1571 (69.0)	707 (31.0)		
2	1419 (30.8)	1001 (70.5)	418 (29.5)		
Condom use in the past 3 months				75.3 (3)	<.001
Every time	2364 (51.4)	1606 (67.9)	758 (32.1)		
Sometimes use	1122 (24.4)	777 (69.3)	345 (30.7)		
No sex	822 (17.9)	438 (53.3)	384 (46.7)		
Never use	292 (6.3)	214 (73.3)	78 (26.7)		
Rush poppers usage frequency during sex activity				46.2 (2)	<.001
Frequently use	698 (15.2)	529 (75.8)	169 (24.2)		
Occasional use	712 (15.5)	495 (69.5)	217 (30.5)		
Never	3190 (69.3)	2011 (63.0)	1179 (37.0)		
Awareness of partner’s HIV status				36.8 (2)	<.001
Yes	2007 (43.6)	1371 (68.3)	636 (31.7)		

XSL•FO
RenderX

Variables	Participants, n (%)	Number of mobile recruitments, n (%)	Number of volunteer-driven recruitment, n (%)	Chi-square (*df*)	*P* value
Part	1516 (33.0)	1036 (68.3)	480 (31.7)		
No	1077 (23.4)	628 (58.3)	449 (41.7)		
History of STIs[a]				6.0 (1)	.014
No	4474 (97.3)	2939 (65.7)	1535 (34.3)		
Yes	126 (2.7)	96 (76.2)	30 (23.8)		
Awareness of PrEP[b]				108.9 (1)	<.001
Yes	3298 (71.7)	2327 (70.6)	971 (29.4)		
No	1302 (28.3)	708 (54.4)	594 (45.6)		
Awareness of PEP[c]				118.4 (1)	<.001
Yes	3402 (74.0)	2398 (70.5)	1004 (29.5)		
No	1198 (26.0)	637 (53.2)	561 (46.8)		
HIV testing pathway				100.7 (1)	<.001
Mailed self-testing	2169 (47.2)	1592 (73.4)	577 (26.6)		
Offline testing	2431 (52.8)	1443 (59.4)	988 (40.6)		
HIV status				4.9 (1)	.027
Positive	66 (1.4)	52 (78.8)	14 (21.2)		
Negative	4534 (98.6)	2983 (65.8)	1551 (34.2)		
Ever first-time HIV testing				2.9 (1)	.090
No	3735 (81.2)	2443 (65.4)	1292 (34.6)		
Yes	865 (18.8)	592 (68.4)	273 (31.6)		

[a]STIs: sexually transmitted infections.

[b]PrEP: pre-exposure prophylaxis.

[c]PEP: postexposure prophylaxis.

Overall, among the MSM recruited through social media, 78.9% (2394/3035) had gay sexual partners, 33% (1001/3035) had more than 2 sexual partners, 17.4% (529/3035) frequently used rush poppers during sex, 3.2% (96/3035) had a history of STIs, 76.7% (2327/3035) had awareness of PrEP, 79% (2398/3035) had awareness of PEP, 1.7% (52/3035) tested HIV-positive, and 19.5% (592/3035) had been tested for HIV for the first time, all of which were higher than those of the volunteer recruitment (Table 2).

XSL•FO
RenderX

Table 2. Factors associated with mobile social media promotion compared with volunteer-driven recruitment strategy among men who have sex with men.

Variables	Number of mobile social media recruitments, n (%)	Number of volunteer-driven recruitments, n (%)	aOR[a] (95% CI)	*P* value
Overall	3035 (100)	1565 (100)		
Sexuality of partner				
Gay	2394 (78.9)	1061 (67.8)	1.23 (1.05-1.45)	.013
Bisexual	641 (21.1)	504 (32.2)	1	
Number of sexual partners in the past 3 months				
0	463 (15.3)	440 (28.1)	1	
1	1571 (51.8)	707 (45.2)	1.66 (1.38-1.00)	<.001
2	1001 (33.0)	418 (26.7)	1.74 (1.42-2.11)	<.001
Rush poppers usage frequency during sex activity				
Frequently use	529 (17.4)	169 (10.8)	1.39 (1.14-1.99)	.001
Occasional use	495 (16.3)	217 (13.9)	1.12 (0.93-1.35)	.225
Never	2011 (66.3)	1179 (75.3)	1	
History of STIs[b]				
No	2939 (96.8)	1535 (98.1)	1	
Yes	96 (3.2)	96 (1.9)	1.56 (1.02-2.39)	.040
Awareness of PrEP[c]				
Yes	2327 (76.7)	971 (62.0)	1.42 (1.19-1.71)	<.001
No	708 (23.3)	594 (38.0)	1	
Awareness of PEP[d]				
Yes	2398 (79.0)	1004 (64.2)	1.49 (1.24-1.79)	<.001
No	637 (21.0)	561 (35.8)	1	
HIV status				
Positive	52 (1.7)	14 (0.9)	2.02 (1.10-3.72)	.024
Negative	2983 (98.3)	1551 (99.1)	1	
Ever first-time HIV testing				
No	2443 (80.5)	1292 (82.6)	1	
Yes	592 (19.5)	273 (17.4)	1.28 (1.09-1.52)	.003

[a]aOR: adjusted odds ratio.

[b]STI: sexually transmitted infections.

[c]PrEP: pre-exposure prophylaxis.

[d]PEP: postexposure prophylaxis.

In multivariate logistic regression analysis, the following factors were significantly associated with mobile social media recruitment: having only gay sexual partners (aOR 1.23, 95% CI 1.05-1.45), having more than 2 sexual partners in the past 3 months (aOR 1.74, 95% CI 1.42-2.11), frequently using rush poppers during sex (aOR 1.39, 95% CI 1.14-1.99), having a history of STIs (aOR 1.56, 95% CI 1.02-2.39), having awareness of PrEP (aOR 1.42, 95% CI 1.19-1.71), having awareness of PEP (aOR 1.49, 95% CI 1.24-1.79), testing positive for HIV (aOR 2.02, 95% CI 1.10-3.72), and undergoing HIV testing for the first time (aOR 1.28, 95% CI 1.09-1.52) (Table 2). MSM who were recruited by the mobile platform were more likely (aOR 2.02, 95% CI 1.77-2.31) to take the mailed HIV self-test (Table 3).

XSL•FO
RenderX

Table 3. Factors associated with mailed self-testing compared with offline testing among men who have sex with men.

Variables	Participants, n (%)	Number of mailed self-testing, n (%)	Number of offline testings, n (%)	Chi-square (*df*)	*P* value	aOR[a] (95% CI)
Overall	4600 (100)	2169 (47.2)	2431 (52.8)			
Age (years), IQR 27 (23-32)				97.7 (2)	<.001	
<20	306 (6.7)	132 (43.1)	174 (56.9)		.092	1.30 (0.96-1.76)
20-29	2719 (59.1)	1444 (53.1)	1275 (46.9)		<.001	1.43 (1.22-1.68)
≥30	1575 (34.2)	593 (37.7)	982 (62.3)			1
Marital status				99.8 (2)	<.001	
Single	3606 (78.4)	1836 (50.9)	1770 (49.1)			1
Married	857 (18.6)	275 (32.1)	582 (67.9)		<.001	0.59 (0.49-0.72)
Divorced or separated	137 (3.0)	58 (42.3)	79 (57.7)		.876	1.03 (0.71-1.50)
Education level				122.8 (3)	<.001	
High school and below	1175 (25.5)	402 (34.2)	773 (65.8)			1
College	1175 (25.5)	550 (46.8)	625 (53.2)		<.002	1.34 (1.11-1.61)
Bachelor’s degree	1826 (39.7)	993 (54.4)	833 (45.6)		<.001	1.62 (1.35-1.93)
Master’s degree or above	424 (9.2)	224 (52.8)	200 (47.2)		<.001	1.67 (1.29-2.15)
Occupation				59.8 (3)	<.001	
Company employee	1605 (34.9)	865 (53.9)	740 (46.1)		.006	1.25 (1.07-1.46)
Student	836 (18.2)	411 (49.2)	425 (50.8)		.685	0.96 (0.78-1.18)
Freelance	612 (13.3)	250 (40.8)	362 (59.2)		.485	0.93 (0.76-1.14)
Others	1547 (33.6)	643 (41.6)	904 (58.4)			1
Sex roles				29.5 (2)	<.001	
Receptive anal sex	1003 (21.8)	528 (52.6)	475 (47.4)			1
Both	2085 (45.3)	896 (43.0)	1189 (57.0)		<.001	0.64 (0.54-0.76)
Insertive anal sex	1512 (32.9)	745 (49.3)	767 (50.7)		.258	0.91 (0.76-1.08)
Sexuality of partner				0.2 (1)	.638	
Gay	3455 (75.1)	1636 (47.4)	1819 (52.6)			1
Bisexual	1145 (24.9)	533 (46.6)	612 (53.4)		<.001	1.45 (1.22-1.72)
Number of sexual partners in the past 3 months				24.3 (2)	<.001	
0	903 (19.6)	402 (44.5)	501 (55.5)			1
1	2278 (49.5)	1021 (44.8)	1257 (55.2)		.822	1.03 (0.82-1.29)
2	1419 (30.8)	746 (52.6)	673 (47.4)		.013	1.34 (1.05-1.72)
Condom use in the past 3 months				24.6 (3)	<.001	
Every time	2364 (51.4)	1050 (44.4)	1314 (55.6)			1
Sometimes use	1122 (24.4)	594 (52.9)	528 (47.1)		<.001	1.49 (1.28-1.74)
No sex	822 (17.9)	376 (45.7)	446 (54.3)		.095	1.22 (0.97-1.53)
Never use	292 (6.3)	149 (51.0)	143 (49.0)		.013	1.39 (1.07-1.80)
Rush poppers usage frequency during sex activity				36.3 (2)	<.001	
Frequently use	698 (15.2)	278 (39.8)	420 (60.2)			1
Occasional use	712 (15.5)	397 (55.8)	315 (44.2)		<.001	1.93 (1.54-2.41)
Never	3190 (69.3)	1494 (46.8)	1696 (53.2)		<.001	1.57 (1.31-1.88)
Awareness of partner’s HIV status				8.5 (2)	.014	
Yes	2007 (43.6)	956 (47.6)	1051 (52.4)			—[e]

XSL•FO
RenderX

Variables	Participants, n (%)	Number of mailed self-testing, n (%)	Number of offline testings, n (%)	Chi-square (*df*)	*P* value	aOR[a] (95% CI)
Part	1516 (33.0)	745 (49.1)	771 (50.9)			—
No	1077 (23.4)	468 (43.5)	609 (56.5)			—
History of STIs[b]				6.8 (1)	.009	
No	4474 (97.3)	2124 (47.5)	2350 (52.5)		<.003	1.80 (1.22-2.65)
Yes	126 (2.7)	45 (35.7)	81 (64.3)			1
Awareness of PrEP[c]				3.6 (1)	.058	
Yes	3298 (71.7)	1584 (48.0)	1714 (52.0)			—
No	1302 (28.3)	585 (44.9)	717 (55.1)			—
Awareness of PEP[d]				5.8 (1)	.016	
Yes	3402 (74.0)	1640 (48.2)	1762 (51.8)			—
No	1198 (26.0)	529 (44.2)	669 (55.8)			—
HIV status				25.0 (1)	<.001	
Positive	66 (1.4)	11 (16.7)	55 (83.3)			1
Negative	4534 (98.6)	2158 (47.6)	2376 (52.4)		<.001	5.04 (2.55-9.96)
Ever first-time HIV testing				51.4 (1)	<.001	
No	3735 (81.2)	1856 (49.7)	1879 (50.3)		<.001	1.75 (1.48-2.06)
Yes	865 (18.8)	313 (36.2)	552 (63.8)			1
Recruitment				100.7 (1)	<.001	
Volunteer-driven	1565 (34.0)	577 (36.9)	988 (63.1)			1
Mobile recruitment	3035 (66.0)	1592 (52.5)	1443 (47.5)		<.001	2.02 (1.77-2.31)

[a]aOR: adjusted odds ratio.

[b]STI: sexually transmitted infections.

[c]PrEP: pre-exposure prophylaxis.

[d]PEP: postexposure prophylaxis.

[e]Not included in multivariate logistic regression analysis.

Discussion

Principal Results

This study compared the sociodemographic and behavioral characteristics of MSM who were recruited through 2 different recruitment strategies by community organizations to engage in HIV testing in Zhejiang Province, China. The strategies examined were a mobile social media promotion strategy and a volunteer-driven recruitment strategy. The findings indicated that 66% (3035/4600) of MSM were recruited through mobile social media promotion, while both strategies together reached 18.8% (865/4600) of MSM who had never been tested for HIV. Consistent with previous studies [18], community organizations proved effective in increasing HIV testing rates among MSM, particularly those who had not been tested before. The implementation of both strategies successfully expanded intervention coverage and recruited MSM who had not been previously tested for HIV. Furthermore, ongoing support for community organizations is recommended to facilitate the implementation of MSM prevention and control initiatives.

Comparison With Previous Work

Our findings revealed a higher proportion of individuals undergoing their first HIV test (592/3035, 19.5% vs 273/1565, 17.4%) and a higher prevalence of HIV-positive results (52/3035, 1.7% vs 14/1565, 0.9%) among those engaged via mobile social media promotions compared to those recruited through volunteer-driven initiatives. These observations are consistent with previous studies and empirical data. A meta-analysis demonstrated a 50% increase in HIV testing uptake following social media interventions [19], and another study showed a 22% increase in testing through mobile platforms [20], particularly among MSM engaging in high-risk behaviors. In these populations, the HIV-positive rate was 3.5% [21]. These findings suggest that mobile social media promotion may be an effective approach for recruiting MSM, improving HIV-testing coverage, and identifying individuals infected with HIV, particularly those who have not previously undergone HIV testing. As the proportion of MSM who use the internet and dating software continues to grow, it may be beneficial to expand the use of mobile social media promotion strategies to enhance HIV testing among MSM and facilitate the identification of more individuals infected with HIV. This could

XSL•FO
RenderX

be a crucial step toward achieving the goal of diagnosing 95% of individuals infected with HIV by 2030.

Similar to previous studies [22,23], our findings indicate that MSM recruited through mobile social media were more likely to use mail-based HIV self-testing kits. Mobile promotion strategies can facilitate the recruitment of MSM who have not accessed HIV testing services by providing convenient mail-in self-testing kits, thus increasing overall testing and identifying more individuals infected with HIV. Another study demonstrated that HIV self-testing services have the potential to enhance test uptake and reduce undiagnosed infections [24]. It would be beneficial to provide targeted outreach to MSM, such as offering HIV self-testing [25]. The utilization of social media platforms may prove effective in enhancing HIV self-testing rates among MSM aged ≤35 years [26]. Nevertheless, research has indicated that mobile social media may not necessarily increase the rate of HIV testing [27]. This may be attributed to the fact that the testing rate may not be enhanced by repeated HIV publicity and intervention in large cities but rather in areas where HIV testing is not easily accessible and where supportive gay social networks and health professionals are scarce. Therefore, priority should be given to implementing mobile social media promotion in areas lacking LGBTQ+ organizations or HIV health services. This approach can facilitate MSM recruitment, thereby increasing intervention coverage and improving HIV testing. Concurrently, it is also recommended to continue increasing publicity about HIV self-testing and to provide convenient self-testing services.

Consistent with other studies [28,29], we found that MSM recruited via mobile social media promotion were more aware of PrEP and PEP. Dating apps and social media have been identified as underutilized tools for increasing PrEP awareness, uptake, and knowledge among MSM [28,29]. PrEP and PEP represent effective prevention tools for individuals at high risk of HIV exposure, effectively reducing HIV infection and transmission [30]. Since 2021, PrEP and PEP have been widely used in China [31], with much of the advertising and promotion happening mobile. This has substantially increased awareness among MSM using the Internet. This study demonstrated that most MSM initially learned about PrEP through mobile sources [32]. Efforts should be made to further promote PrEP and PEP through mobile platforms while also disseminating preventive information through offline volunteer-driven strategies.

Our findings also revealed that MSM recruited mobile were more likely to report using rush poppers, a recreational drug that has gained considerable popularity among MSM [33], and their rate of use has continued to increase in China [34]. The use of rush poppers has been linked to abnormal arousal and pleasure, and prolonged sexual activity; these factors may increase the likelihood of engaging in unprotected sexual activity with consequent adverse outcomes such as HIV infection [30] and transmission [35]. Furthermore, a greater proportion of MSM seeking sexual activity through social media have reported recreational drug use compared to those using traditional venues [21]. Therefore, it is recommended that internet-based HIV interventions include information on the risks associated with rush poppers and their link to HIV transmission to raise awareness and reduce high-risk behaviors.

The study also revealed that MSM recruited mobile had more sexual partners and a higher prevalence of STIs compared to those recruited through volunteer-driven strategies. A meta-analysis conducted in China revealed that MSM who engage in substance abuse are more likely to seek sexual partners through the internet or social media and engage in unprotected anal intercourse [36]. One potential explanation for this is that the volunteer-driven strategy for MSM tends to focus on offline recruitment in venues such as bars, parks, and bathhouses, where social circles are relatively limited. The advent of the Internet has provided MSM with more opportunities to search for sexual partners. Mobile social media have demonstrated efficacy as a tool for increasing PrEP use [29] and HIV-testing uptake [32]. While PrEP reduces the risk of HIV infection, it does not prevent STIs [37] and may even increase the risk due to reduced condom use and increased high-risk sexual behaviors [38]. Therefore, alongside PrEP promotion, other targeted interventions promoting condom use and awareness of STI risks should be integrated into internet-based HIV prevention efforts.

Limitations

First, the participants recruited by the community organization were primarily younger MSM who use the internet, meaning older MSM (particularly those over 50) were underrepresented. Future research should explore recruitment strategies targeting older MSM who do not engage mobile. Second, the costs associated with the human and financial resources invested in these recruitment strategies were not assessed. Further research on the cost-effectiveness of these strategies is necessary. Finally, the study relied on routine surveillance questionnaires, which may not have captured all relevant factors influencing the effectiveness of the 2 investigated recruitment strategies.

Conclusions

Community organizations effectively use mobile social media promotion and volunteer-driven strategies to recruit MSM, playing an important role in improving HIV testing and identifying undiagnosed individuals infected with HIV. There are notable differences in recruitment outcomes between the 2 strategies, with mobile social media promotion leading to a higher proportion of first-time HIV testing and HIV-positive individual identification. In areas where LGBTQ+ organizations or HIV health services are scarce, mobile social media promotion should be prioritized to increase MSM recruitment and expand the coverage of HIV interventions.

XSL•FO
RenderX

Acknowledgments

We express our gratitude to the participants and Sunshine Coast Public Welfare for their contributions to the study. The study was funded by a grant from the Health Science and Technology Project in Zhejiang Province Social (2023KY632, 2021RC048), Zhejiang Science and Technology Plan for Disease Prevention and Control (project number 2025JK006).

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LH conceived of the study design and coordinated the conduct of this research in the field. LH performed the statistical analysis and drafted the manuscript. CC reviewed and revised the manuscript. SJ, TJ, WC, JZ, and WC designed the research and played a major role in the survey. All of the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Dong MJ, Peng B, Liu ZF, Ye QN, Liu H, Lu XL, et al. The prevalence of HIV among MSM in China: a large-scale systematic analysis. BMC Infectious Diseases 2019 Nov 27;19(1):1-20 [FREE Full text] [doi: 10.1186/s12879-019-4559-1] [Medline: 31775654]
2. Han M. Analysis of the epidemic situation of AIDS in China and prospects for its prevention and control. J AIDS STD 2023 Mar 15;29(03):247-250.
3. Stannah J, Soni N, Lam JKS, Giguère K, Mitchell KM, Kronfli N, et al. Trends in HIV testing, the treatment cascade, and HIV incidence among men who have sex with men in Africa: a systematic review and meta-analysis. Lancet HIV 2023 Aug 10;10(8):e528-e542 [FREE Full text] [doi: 10.1016/S2352-3018(23)00111-X] [Medline: 37453439]
4. CCFD: HIV prevention and control progress report. National Center For AIDS STD Control And Prevention. URL: https://www.cdc.gov/hiv/pdf/policies/progressreports/cdc-hiv-preventionprogressreport.pdf [accessed 2025-01-18]
5. He L, Jiang T, Chen W, Jiang S, Zheng J, Chen W, et al. Examining HIV testing coverage and factors influencing first-time testing among men who have sex with men in Zhejiang province, China: cross-sectional study based on a large internet survey. JMIR Public Health Surveill 2024 Jun 14;10:1-20 [FREE Full text] [doi: 10.2196/56906] [Medline: 38875001]
6. He L, Pan X, Yang J, Ma Q, Jiang J, Wang W, et al. HIV risk behavior and HIV testing among rural and urban men who have sex with men in Zhejiang Province, China: a respondent-driven sampling study. PLoS One 2020 Apr 2;15(4):e0231026 [FREE Full text] [doi: 10.1371/journal.pone.0231026] [Medline: 32240244]
7. Campbell CK, Lippman SA, Moss N, Lightfoot M. Strategies to increase HIV testing among MSM: a synthesis of the literature. AIDS Behav 2018 Aug 22;22(8):2387-2412 [FREE Full text] [doi: 10.1007/s10461-018-2083-8] [Medline: 29550941]
8. Solomon SS, Solomon S, McFall AM, Srikrishnan AK, Anand S, Verma V, Indian National Collaboration on AIDS Study. Integrated HIV testing, prevention, and treatment intervention for key populations in India: a cluster-randomised trial. Lancet HIV 2019 May 1;6(5):e283-e296 [FREE Full text] [doi: 10.1016/S2352-3018(19)30034-7] [Medline: 30952565]
9. Fan EL. HIV testing as prevention among MSM in China: the business of scaling-up. Glob Public Health 2014 Feb 5;9(1-2):85-97 [FREE Full text] [doi: 10.1080/17441692.2014.881520] [Medline: 24498955]
10. Cao B, Liu C, Durvasula M, Tang W, Pan S, Saffer AJ, et al. Social media engagement and HIV testing among men who have sex with men in China: a nationwide cross-sectional survey. J Med Internet Res 2017 Jul 19;19(7):e251 [FREE Full text] [doi: 10.2196/jmir.7251] [Medline: 28724510]
11. Jin J, Li J, Dai S. Eectiveness of social organizations on intervention for detecting MSM in China. Chin J AIDS STD 2024 2024 May 15;30(05):503-507.
12. Zhang W, Hu Q, Tang W, Jin X, Mao X, Lu T, et al. HIV self-testing programs to men who have sex with men delivered by social media key opinion leaders and community-based organizations are both effective and complementary: a national pragmatic study in China. J Acquir Immune Defic Syndr 2020 Aug 15;84(5):453-462. [doi: 10.1097/QAI.0000000000002375] [Medline: 32692103]
13. Booton RD, Ong JJ, Lee A, Liu A, Huang W, Wei C, et al. Modelling the impact of an HIV testing intervention on HIV transmission among men who have sex with men in China. HIV Med 2021 Jan 28;22(6):467-477 [FREE Full text] [doi: 10.1111/hiv.13063] [Medline: 33511687]
14. Cohen MS, Gamble T, McCauley M. Prevention of HIV transmission and the HPTN 052 study. Annu Rev Med 2020 Jan 27;71:347-360. [doi: 10.1146/annurev-med-110918-034551] [Medline: 31652410]

XSL•FO
RenderX

15. Reitsema M, Heijne J, Visser M, van Sighem A, Schim van der Loeff M, Op de Coul ELM, et al. Impact of frequent testing on the transmission of HIV and among men who have sex with men: a mathematical modelling study. Sex Transm Infect 2020 Jul 16;96(5):361-367. [doi: 10.1136/sextrans-2018-053943] [Medline: 31801895]
16. Young S, Cumberland W, Singh P, Coates T. A peer-led online community to increase HIV self-testing among African American and Latinx MSM: a randomized controlled trial. J Acquir Immune Defic Syndr 2022 May 1;90(1):20-26 [FREE Full text] [doi: 10.1097/QAI.0000000000002919] [Medline: 35044989]
17. LeGrand S, Muessig K, Horvath K, Rosengren A, Hightow-Weidman LB. Using technology to support HIV self-testing among MSM. Curr Opin HIV AIDS 2017 Sep 1;12(5):425-431 [FREE Full text] [doi: 10.1097/COH.0000000000000400] [Medline: 28617712]
18. Holland CE, Papworth E, Billong SC, Kassegne S, Petitbon F, Mondoleba V, et al. Access to HIV services at non-governmental and community-based organizations among men who have sex with men (MSM) in Cameroon: an integrated biological and behavioral surveillance analysis. PLoS One 2015 Apr 23;10(4):e0122881 [FREE Full text] [doi: 10.1371/journal.pone.0122881] [Medline: 25906046]
19. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
20. Biello KB, Daddario SR, Hill-Rorie J, Futterman D, Sullivan PS, Hightow-Weidman L, et al. Uptake and acceptability of MyChoices: results of a pilot RCT of a mobile app designed to increase HIV testing and PrEP uptake among young American MSM. AIDS Behav 2022 Jul 13;26(12):3981-3990. [doi: 10.1007/s10461-022-03724-3] [Medline: 35829971]
21. Chiou PY, Ko NY, Chien CY. Mobile HIV testing through social networking platforms: comparative study. J Med Internet Res 2021 Mar 26;23(3):e25031 [FREE Full text] [doi: 10.2196/25031] [Medline: 33769298]
22. Wang Z, Chan PSF, Xin M, Fang YY, Chidgey A, Yu F, et al. An online intervention promoting HIV testing service utilization among Chinese men who have sex with men during the COVID-19 pandemic: a quasi-experimental study. AIDS Behav 2024 Jun 7;28(Suppl 1):77-89 [FREE Full text] [doi: 10.1007/s10461-023-04100-5] [Medline: 37284925]
23. Zhou H, Zhu YY, Gao YY, Chu ZX, Chen S, Liu M, et al. Online distribution of HIV self-testing kits to promote HIV testing among men who have sex with men discontinuing pre-exposure prophylaxis after demonstration project completion in China: a multicentre open-label randomized controlled trial. Lancet Reg Health West Pac 2023 Oct 14;41:100922 [FREE Full text] [doi: 10.1016/j.lanwpc.2023.100922] [Medline: 37867621]
24. Lightfoot MA, Campbell CK, Moss N, Treves-Kagan S, Agnew E, Kang Dufour MS, et al. Using a social network strategy to distribute HIV self-test kits to African American and latino MSM. J Acquir Immune Defic Syndr 2018 Sep 1;79(1):38-45 [FREE Full text] [doi: 10.1097/QAI.0000000000001726] [Medline: 29771792]
25. Noble M, Jones AM, Bowles K, DiNenno EA, Tregear SJ. HIV testing among internet-using MSM in the United States: systematic review. AIDS Behav 2017 Feb 1;21(2):561-575. [doi: 10.1007/s10461-016-1506-7] [Medline: 27498198]
26. Wang L, Podson D, Chen Z, Lu H, Wang V, Shepard C, et al. Using social media to increase HIV testing among men who have sex with men - Beijing, China, 2013-2017. MMWR Morb Mortal Wkly Rep 2019 May 1;68(21):478-482 [FREE Full text] [doi: 10.15585/mmwr.mm6821a3] [Medline: 31145721]
27. Del Río-González AM, Zea MC, Calabrese SK, Betancourt F, Pacheco-Cabrales J, Estrada-Santiago Y, et al. Development and evaluation of an online education-entertainment intervention to increase knowledge of HIV and uptake of HIV testing among Colombian men who have sex with men (MSM). Int J Environ Res Public Health 2021 Feb 1;18(4):1811 [FREE Full text] [doi: 10.3390/ijerph18041811] [Medline: 33673321]
28. Adan MA, Psaros C, Chamberlin G, Zionts D, Iyer S, Platt L, et al. Swiping right on PrEP: a qualitative study of MSM preferences for PrEP public health messaging on dating apps. AIDS Care 2024 Apr 22;36(8):1135-1147 [FREE Full text] [doi: 10.1080/09540121.2024.2343761] [Medline: 38648528]
29. Kudrati SZ, Hayashi K, Taggart T. AIDS Behav 2021 May 3;25(12):4225-4234 [FREE Full text] [doi: 10.1007/s10461-021-03287-9] [Medline: 33939035]
30. Guo W, Li Y, Zhou N, Wu GH, Chang WH, Huan XP, et al. [Risk factors related to HIV new infections among men who have sex with men in a cohort study] [Article in Chinese]. Zhonghua Liu Xing Bing Xue Za Zhi 2018 Jan 15;39(1):16-20. [doi: 10.3760/cma.j.issn.0254-6450.2018.01.003] [Medline: 29374888]
31. Xu J, Huang X, Liu X. Expert consensus on pre exposure prophylaxis for HIV in China. Chinese Journal of AIDS & STD 2020 Nov 15;26(11):7.
32. Gebru NM, Benvenuti MC, Rowland BHP, Kalkat M, Chauca PG, Leeman RF. Relationships among substance use, sociodemographics, pre-exposure prophylaxis (PrEP) awareness and related attitudes among young adult men who have sex with men. Subst Use Misuse 2022 Feb 21;57(5):786-798 [FREE Full text] [doi: 10.1080/10826084.2022.2040030] [Medline: 35188880]
33. He L, Pan X, Wang N, Yang J, Jiang J, Luo Y, et al. New types of drug use and risks of drug use among men who have sex with men: a cross-sectional study in Hangzhou, China. BMC Infect Dis 2018 Apr 17;18(1):182 [FREE Full text] [doi: 10.1186/s12879-018-3091-z] [Medline: 29665785]

XSL•FO
RenderX

34. Zhu Z, Yan H, Wu S, Xu Y, Xu W, Liu L, et al. Trends in HIV prevalence and risk behaviours among men who have sex with men from 2013 to 2017 in Nanjing, China: a consecutive cross-sectional survey. BMJ Open 2019 Jan 3;9(1):e021955 [FREE Full text] [doi: 10.1136/bmjopen-2018-021955] [Medline: 30705234]
35. Zhu ZP, Zhang M, Xu YY, Xu WJ, Liu L, Wu SS, et al. [Cross-sectional surveys on the use of recreational drug nitrous-acid-ester rush-poppers in men who have sex with men, Nanjing] [Article in Chinese]. Zhonghua Liu Xing Bing Xue Za Zhi 2017 Feb 15;38(2):189-193. [doi: 10.3760/cma.j.issn.0254-6450.2017.02.011] [Medline: 28231664]
36. Zhao T, Chen G, Sun C, Gong X, Li H, Fu G. The epidemic of HIV and syphilis and the correlation with substance abuse among men who have sex with men in China: a systematic review and meta-analysis. Front Public Health 2023 Feb 17;11:1082637 [FREE Full text] [doi: 10.3389/fpubh.2023.1082637] [Medline: 36875380]
37. Traeger MW, Guy R, Asselin J, Patel P, Carter A, Wright EJ, Australian Collaboration for Coordinated Enhanced Sentinel Surveillance of Sexually Transmissible Infections and Blood Borne Viruses (ACCESS) Study Group. Real-world trends in incidence of bacterial sexually transmissible infections among gay and bisexual men using HIV pre-exposure prophylaxis (PrEP) in Australia following nationwide PrEP implementation: an analysis of sentinel surveillance data. Lancet Infect Dis 2022 May 25;22(8):1231-1241. [doi: 10.1016/S1473-3099(22)00175-X] [Medline: 35643090]
38. Hoenigl M, Jain S, Moore D, Collins D, Sun X, Anderson PL, California Collaborative Treatment Group 595 Team. Substance use and adherence to HIV preexposure prophylaxis for men who have sex with men. Emerg Infect Dis 2018 Dec 1;24(12):2292-2302 [FREE Full text] [doi: 10.3201/eid2412.180400] [Medline: 30457536]

Abbreviations

aOR: adjusted odds ratio
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
PEP: postexposure prophylaxis
STIs: sexually transmitted infections

Edited by A Mavragani; submitted 20.09.24; peer-reviewed by S Luo, E Miyagi; comments to author 29.11.24; revised version received 29.12.24; accepted 09.01.25; published 13.02.25.

Please cite as:
He L, Jiang S, Jiang T, Chen W, Zheng J, Wang H, Chai C
A Comparison of Mobile Social Media Promotion and Volunteer-Driven Strategies for Community Organizations Recruiting Men Who Have Sex with Men for HIV Testing in Zhejiang Province, China: Cross-Sectional Study Based on a Large-Scale Survey
J Med Internet Res 2025;27:e66702
URL: https://www.jmir.org/2025/1/e66702
doi:10.2196/66702
PMID:

©Lin He, Shaoqiang Jiang, Tingting Jiang, Wanjun Chen, Jinlei Zheng, Hui Wang, Chengliang Chai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Role of the Organization in Promoting Information Security–Related Behavior Among Resident Physicians in Hospitals in Germany: Cross-Sectional Questionnaire Study

Judith Kraushaar[1], Dr rer pol; Sabine Bohnet-Joschko[1], Prof Dr

Chair of Healthcare Management and Innovation, Faculty of Management, Economics and Society, Witten/Herdecke University, Witten, Germany

Corresponding Author:
Judith Kraushaar, Dr rer pol
Chair of Healthcare Management and Innovation
Faculty of Management, Economics and Society
Witten/Herdecke University
Alfred-Herrhausen-Straße 50
Witten, 58455
Germany
Phone: 49 2302 926 47
Email: judith.kraushaar@uni-wh.de

Abstract

Background: Nowadays, optimal patient care should be based on data-driven decisions. In the course of digitization, hospitals, in particular, are becoming complex organizations with an enormously high density of digital information. Ensuring information security is, therefore, essential and has become a major challenge. Researchers have shown that—in addition to technological and regulatory measures—it is also necessary for all employees to follow security policies and consciously use information technology (compliance), because noncompliance can lead to security breaches with far-reaching consequences for the organization. There is little empirical research on information security–related behavior in hospitals and its organizational antecedents.

Objective: This study aimed to explore the impact of specific job demands and resources on resident physicians' information security–related compliance in hospitals through the mediating role of work engagement and information security–related awareness.

Methods: We used a cross-sectional, survey-based study design to collect relevant data from our target population, namely resident physicians in hospitals. For data analysis, we applied structural equation modeling. Our research model consisted of a total of 7 job demands and resources as exogenous variables, 2 mediators, and information security–related compliance as the endogenous variable.

Results: Overall, data from 281 participating physicians were included in the analyses. Both mediators—work engagement and awareness—had a significant positive effect on information security–related compliance (β=.208, *P*=.001 vs β=.552, *P*<.001). Quality of leadership was found to be the only resource with a significant indirect effect on physicians' compliance, mediated by work engagement (β=.086, *P*=.03). Furthermore, awareness mediated the relationships between information security–related communication and information security–related compliance (β=.192, *P*<.001), as well as between further education and training and the endogenous variable (β=.096, *P*=.02). Contrary to our hypothesis, IT resources had a negative effect on compliance, mediated by awareness (β=–.114, *P*=.02).

Conclusions: This study provides new insights into how a high standard of information security compliance among resident physicians could be achieved through strengthening physicians' security work engagement and awareness. Hospital management is required to establish an information security culture that is informative and motivating and that raises awareness. Particular attention should be paid to the quality of leadership, further education and training, as well as clear communication.

(J Med Internet Res 2025;27:e46257) doi:10.2196/46257

KEYWORDS

information security; compliance; work engagement; awareness; leadership; communication; education and training; security; privacy; structural equation modeling; resident; fellow; medical education; continuing education; professional development

XSL•FO
RenderX

Introduction

Information Security–Related Issues in Hospitals

Nowadays, direct, easy, and quick access to and transmission of information are essential factors for health professionals to provide optimal patient care based on data-driven diagnosis and treatment. In the course of digitization, hospitals, in particular, are becoming complex organizations with an enormously high density of such information, which comes from a wide variety of sources and is available in different information systems. In order to make the decision-making process in hospitals more efficient, increasingly more (medical) devices are being digitally connected to hospital networks, providing health care professionals with data from multiple sources at a central access point. In a recently published study, the average number of active medical devices in large university hospitals in Germany was 25,150, with 4500 (17.9%) of them being digitally connected, a rate that is expected to increase rapidly in the near future [1]. Despite their major benefits in terms of the enhanced quality and efficiency of patient care, the use and high variety of connected devices poses potential information security risks to hospitals. Many of the devices used in hospitals are not only connected to local networks but also directly or indirectly connected to the internet or other external systems. For example, patient data can be transitioned to other stakeholders outside the local network or a cloud where they will be further processed [2,3]. Adding to this, with the increasing usage of personal mobile devices for work purposes by health care professionals, the issue becomes even more complex and difficult to handle for hospital management [4-6].

Information security in hospitals refers to the state of full functionality of all information systems, processes, and components and, therefore, the protection of all information required for optimal patient care. It must be guaranteed at all times. Continuous monitoring, as well as rapid responses to breaches and attacks, are essential. Due to the immense impact of poor information security on patient care, as well as the critical standard of security due to years of underinvestment and neglect of this issue, hospitals around the world have recently become targets of cyberattacks, with massive increases each year [7-10]. A study from 2020 shows how vulnerable the German hospital landscape is to ransomware attacks: the authors analyzed the publicly visible system attack surface of hospitals located in Germany and found that more than 1 in 3 hospitals analyzed had vulnerabilities, including those that were part of the critical infrastructure at the time [11]. In times of scarce financial and human resources, ensuring information security is an enormous challenge [4]. In addition to technological and regulatory measures, research literature also highlights and discusses organizational measures to enhance information security in hospitals [2,8,9]. The reason for this is that a large proportion of information security breaches still occur internally as a result of employee misconduct [2,9,12]. German politicians have already recognized that a high level of information security in hospitals can only be achieved with holistic measures at a human, technological, and organizational level. In addition to providing the technological basis, employees must be made aware of security-relevant issues and obliged and motivated to comply with policies. The hospital management must revise information and communication processes and create appropriate support structures (eg, responsibilities, IT, and training) [13].

Empirical research on information security–related behavior of employees at work and its supporting organizational and individual factors is a relatively new field. Several researchers have found that in addition to the technological equipment of the organization to ensure information security, it is also necessary for the employees to follow security policies and consciously use information technology (compliance), because noncompliance can lead to security breaches with far-reaching consequences for the organization.

To explain information security–related behavior within the organization, various behavioral theories and frameworks have been used, including the Theory of Planned Behavior (TPB), Organizational Behavior (OB), and the Job Demands-Resources Model (JDRM) [14]. Hu et al [15] used the TPB to investigate the role of top management, organizational culture, and individual cognitive beliefs on information security–related behavior among alumni of Management Information Systems and Master of Business Administration programs, finding significant effects. D'Arcy and Greene [16] combined different theories for their research model and examined the influence of security culture, job satisfaction, and perceived organizational support on security compliance intentions among computer-using professionals, finding positive effects on security culture and job satisfaction. Solomon and Brown [17] used the TPB and showed relationships between organizational culture, information security culture—as an organizational subculture—and compliance. They further argued that goal orientation among employees has a stronger influence on compliance than rule orientation.

There is little empirical research on information security behavior in hospitals. Yeng et al [18] recently conducted a study in a paperless hospital in Ghana and assessed the security behavior of health care staff. In their study, the authors covered different individual, psychosocial, cultural, and work factors and found several significant correlations with participants' security behavior [18]. In their case studies, Hedström et al [19] showed that employees in health care organizations are exposed to different value conflicts, for example, health care values versus information security values, which they have to resolve quickly for each situation during their practice. This poses security risks that need to be considered by hospital management [19]. In our first article on information security published in 2022, we examined specific security-related practices of physicians when using smartphones and medical apps in everyday clinical practice. Among the apps examined, some were found to pose a certain risk to data protection, information security, and patient safety. A large majority of the participating physicians who communicated by their smartphone did not use General Data Protection Regulation (GDPR)–compliant messenger services and paid little attention to some security-relevant criteria (such as manufacturer information, information on data protection, and information security) when selecting medical apps for professional use [20]. Both the papers from 2022 described above and the following paper are based on the same dataset. While our first paper focused on specific

XSL•FO
RenderX

security-related practices, this paper builds on this and examines the underlying constructs of this behavior.

Due to the particular threat of cyberattacks, the ongoing digitization, increasing regulatory requirements and the existing research gap on information security–related behavior in hospitals, the aim of this study was to explore the impact of job demands and job resources on resident physicians' information security–related compliance.

Theoretical Framework

To develop our research model, we used elements from the Extended JDRM as well as the OB research field.

JDRM Framework

The JDRM by Demerouti et al [21] postulates that working conditions can be divided into 2 categories—demands and resources—and are associated with different job-related outcomes. Demerouti et al [21] originally intended to use the JDRM to explain the development of burnout: in one of their studies, they confirmed the links between high demands and exhaustion and between a lack of resources and disengagement [21]. Three years later, a revised version was presented by Schaufeli and Bakker [22]. Their model included burnout and engagement as mediators between different predictors and different possible consequences. Since then, the JDRM has served as a theoretical basis in many studies in industrial and organizational psychology [23]. The JDRM has proven to be flexible, with researchers adapting it to their specific study context regarding the outcomes, demands, and resources to be considered, as well as the presumed mediators and moderators [24]. Pham et al [25] used the JDRM for their qualitative research on the information security–related behavior of employees in Vietnam, finding 3 security resources, 3 security demands, and 2 personal resources that affected security engagement and security compliance burnout as the mediators of security compliance [25].

OB Framework

OB research has been around for about 100 years now [26]. There are 3 levels at which OB can be studied: the individual, the group, and the organizational level [27]. By analyzing information security–related compliance, our study focuses on how resident physicians behave as individuals in the organization and which job-related antecedents promote or inhibit compliance. Amankwa et al [28] chose—among others—concepts from OB research to explain information security–related compliance, showing that supportive organizational culture and end-user involvement had a significant effect on employees' attitudes toward compliance, which in turn had a positive effect on their behavioral intentions.

Research Model and Hypotheses

Our initial research model is presented in Figure 1. The relevant constructs, definitions, and corresponding items of the questionnaire can be found in Multimedia Appendix 1.

Figure 1. Research model.

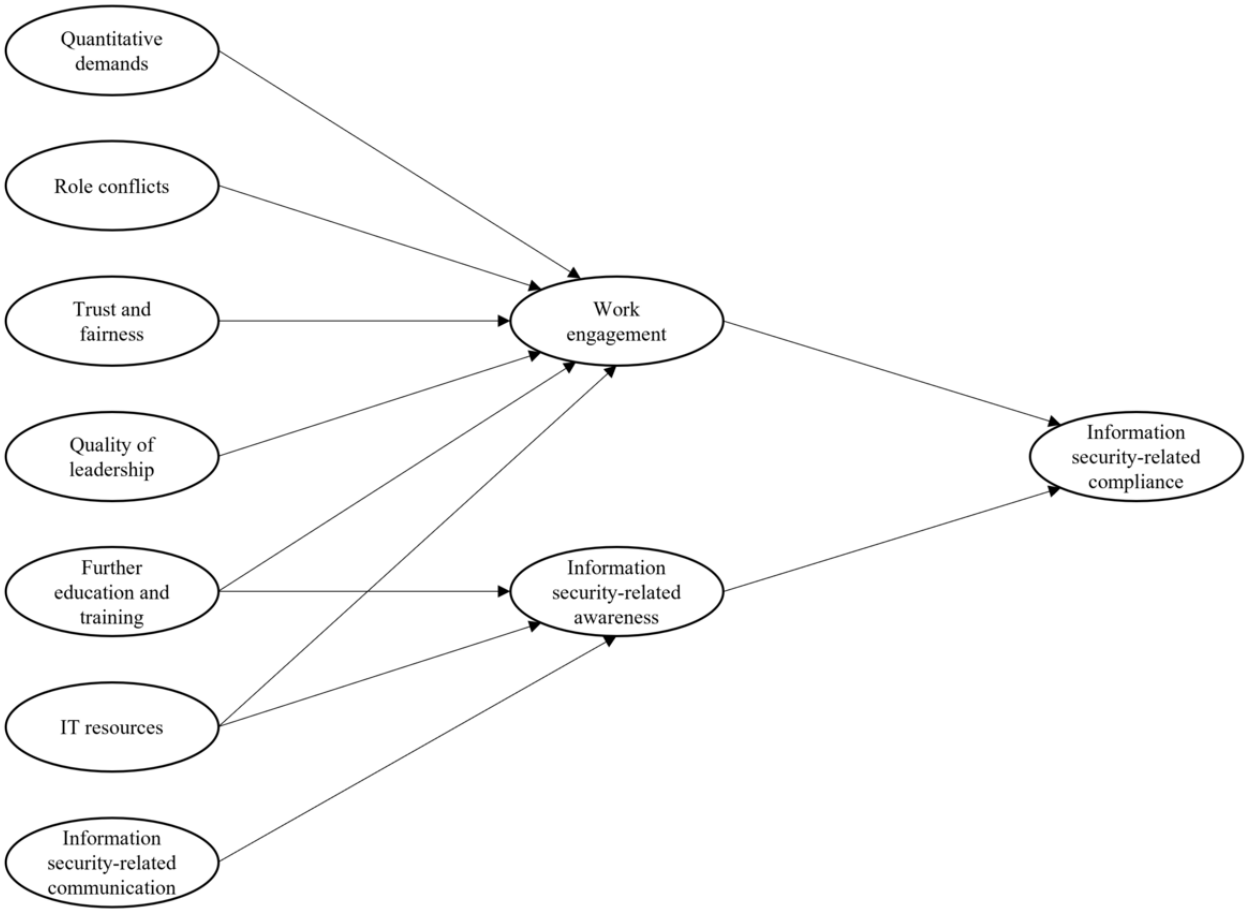

XSL•FO
RenderX

As the first mediator, we have chosen work engagement ("WENG"). In a revised version of the JDRM presented by Schaufeli and Bakker [22], work engagement is considered to be a mediator of the relationship between job resources and job-related outcomes. The authors explained their findings through an underlying motivational process. Next to quality of leadership ("LEAD"), education and training ("EDUC"), and IT resources ("ITRE"), we chose trust and fairness ("TRFA") as the fourth potential resource. Unlike Schaufeli and Bakker [22], we also assumed a negative effect of certain job demands—quantitative demands ("QUAD") and role conflicts ("ROLC")—on work engagement. Here, we wanted to refer to the specific working conditions of physicians in hospitals, which are in general associated with high quantitative demands, but also—as Hedström et al [19] showed—different value conflicts.

Awareness ("AWAR"), as the second mediator, is one of the most commonly investigated variables related to information security–related behavior [29-32]. It is mostly considered to play a mediating role between the exogenous variables and the endogenous compliance variable ("COMP"). In our research model, it also acts as a mediator and reports on physicians' awareness of the risks, threats, policies and responsibilities related to information security in their hospital. In addition to information security–related communication ("COMM"), for which a relationship with information security–related behavior has already been shown in several studies [16,17,28,32], we added EDUC as well as ITRE as potential exogenous variables in our model. Here, the assumption was that both job resources the provision of adequate IT resources and good education and training could lead to greater information security–related awareness. A summary of our research hypotheses is provided in Table 1.

Table 1. Research hypotheses.

#	Hypotheses
H1a	Work engagement is positively related to information security–related compliance.
H1b	Information security–related awareness is positively related to information security–related compliance.
H2a	There is a negative relationship between quantitative demands and information security–related compliance mediated by work engagement.
H2b	There is a negative relationship between role conflicts and information security–related compliance mediated by work engagement.
H2c	There is a positive relationship between trust and fairness and information security–related compliance mediated by work engagement.
H2d	There is a positive relationship between quality of leadership and information security–related compliance mediated by work engagement.
H2e	There is a positive relationship between further education and training and information security–related compliance mediated by work engagement.
H2f	There is a positive relationship between IT resources and information security–related compliance mediated by work engagement.
H3a	There is a positive relationship between further education and training and information security–related compliance mediated by awareness.
H3b	There is a positive relationship between IT resources and information security–related compliance mediated by awareness.
H3c	There is a positive relationship between information security–related communication and information security–related compliance mediated by awareness.

Methods

Study Design

We used a cross-sectional, survey-based study design. A structured questionnaire in German was developed and designed with the free online survey tool LimeSurvey, whereby the first page contained information on the target group, the research project, the content of the questionnaire, and the estimated processing time. After accepting the privacy policy, participants were taken to the second page with demographic questions. These were followed by the main part including seven sections: (1) working conditions; (2) resilience; (3) job satisfaction and work engagement; (4) IT resources, information security, and data protection; (5) information security–related awareness and compliance; (6) technical affinity and innovative work behavior; and (7) mobile device usage. At the end of the survey, participants had the opportunity to share their comments with us. The sections relevant to this report are explained below.

- Working conditions: To assess working conditions, we used the following scales from the Copenhagen Psychosocial Questionnaire (COPSOQ): Quantitative Demands, Predictability, Role Conflicts, Quality of Leadership, Social Support, Feedback, Sense of Community, Trust and Fairness, and Appreciation. The COPSOQ is an internationally established instrument to measure psychosocial work factors, with good to very good validity and reliability for most of its scales [33]. In Germany, the third version of the questionnaire, which we used in our study, was published in 2019 [34]. In contrast to the original questionnaire, we divided the Social Support and Feedback scales into 2 subscales each (supervisors versus colleagues) to separate social support and feedback from supervisors from social support and feedback from colleagues, which could be rated differently, especially in hierarchical organizations. In addition, we used 2 scales (Uncertainty, and Further Education and Training) and 2 individual questions on working hours and shifts from the German instrument for stress-related job analysis for hospital

XSL•FO
RenderX

physicians (ISAK) [35,36]. We also included 4 self-developed items regarding IT resources in the hospital because we could not find a suitable scale in the research literature.

- Information security–related awareness and compliance: In addition to state-of-the-art technical information security solutions, employees of an organization should be aware of the importance of information security and trained accordingly to behave in a compliant manner [31,37]. The items that we used to assess information security–related communication, awareness, self-efficacy, top management commitment, and compliance are based on the works of Hu et al [15], D'Arcy and Greene [16], Karlsson et al [30], and Solomon and Brown [17]. We adapted the items to the clinical situation. Overall, this resulted in 4 items on communication and awareness, 2 on self-efficacy, 1 on top management commitment, and 4 on compliance.

The Work Engagement scale is also part of the COPSOQ and comprises 3 items. All English scales and single items were translated into German, checked independently by 2 bilinguals, and then adapted based on their comments. We used 5-point Likert scales, spanning from "To a very high degree" to "To a very low degree," from "Not true" to "Completely true," from "Always" to "Never/almost never," and from "Strongly disagree" to "Strongly agree," respectively.

We followed the Checklist for Reporting Results of Internet E-Surveys (CHERRIES; Multimedia Appendix 2) [38]. To ensure content validity, the survey was reviewed by faculty members and statisticians and modified accordingly. It was then piloted with a group of residents, who highlighted and took notes on any remaining ambiguities, which we corrected in the final questionnaire. We also tested the usability and technical functionality of the questionnaire before fielding it.

Determining the sample size for structural equation modeling (SEM) is difficult due to its flexibility. There are no generalized guidelines regarding sample size requirements [39]. Therefore, we used the rule of thumb of N≥200 [40].

Data Collection

Data were collected between March and June 2022. Our target group comprised physicians who are currently undergoing medical specialist training/residency training in hospitals in Germany (henceforth referred to as "residents" or "resident physicians"). We focused on residents because they predominantly belong to Generation Y and, compared with older generations, have been used to using digital technologies and especially mobile devices since childhood or youth; therefore, they represent a homogeneous target group in terms of this characteristic. An invitation with a link to the questionnaire was sent directly to the residents by email or social media channels, or indirectly by our contacts in the medical field. Important contacts were chief physicians, senior physicians, university professors, hospital managers, alumni networks, as well as presidents of the German medical societies. In addition, we asked medical experts with significant influence on social media platforms to share the link. The Hartmannbund, an important association of physicians in Germany, forwarded the link to its resident members.

Data Analysis

In total, we examined 2 job demands (QUAD and ROLC) and 5 job resources (TRFA, LEAD, EDUC, ITRE, and COMM) as exposure variables, 2 mediators (WENG and AWAR), and 1 outcome variable (COMP). All variables were measured using Likert scales and are interval scaled. A summary of our research hypotheses is provided in Table 1. To be statistically precise, we would like to highlight that the hypotheses are always intended to analyze the unique effects of the variables in the model and not the bivariate correlations.

For the analysis of our collected data, we applied SEM using the free software environment for statistical computing and graphics R (R Foundation for Statistical Computing) and the *lavaan* package. SEM refers to a family of data analysis techniques for complex relationships between multiple variables in a single research model and has become popular in the social and behavioral sciences [41,42]. Based on our research model and hypotheses, we followed the widely adopted 2-step approach to SEM by Anderson and Gerbing [43], in which the assessment of the measurement model is followed by the analysis of the structural model including tests of the research hypotheses. While the measurement model describes how the latent variables are measured by the indicator variables, the structural model shows how the latent variables are related. In the first step, a confirmatory factor analysis was performed to determine whether the initially proposed measurement model fits the data or needs to be respecified accordingly. Various fit indices exist for evaluating the global model fit. In accordance with Hu and Bentler [44], we used the standardized root mean squared residual (SRMR), the comparative fit index (CFI), and the root mean squared error of approximation (RMSEA). The authors recommended the following thresholds for fit indices that characterize a good global fit: SRMR<.08, CFI≥.95, and RMSEA<.06 [44]. Modification indices were also analyzed to determine the local fit. Here, we only made changes that were theoretically sensible in terms of our model. In addition, the results obtained were assessed regarding reliability, validity, and plausibility. If respecification was indicated, the entire evaluation process was repeated until a good model fit was achieved. In the second step, the relationships between the latent variables are analyzed by again evaluating the model fit and afterward interpreting the path coefficients β and the coefficient of determination R^2. To calculate the results, we used the robust WLSMV (weighted least squares means and variance adjusted) estimator. To ensure model identification, we only used factors with 3 or more items in the initial measurement model.

Ethical Considerations

Ethical review and approval were waived for this study. We are thus following the guidelines of the German Research Foundation (DFG) and the German Data Forum (RatSWD). The RatSWD is an independent body of empirically working scientists and representatives of the most important public institutions for data collection in Germany. It recommends a careful self-examination of research ethics in order to decide whether an ethical review by a committee is necessary for the planned research project. After careful self-examination, we were able to classify our research project as unobjectionable in

XSL•FO
RenderX

terms of research ethics: neither patients nor other vulnerable groups took part in the study. Also, our study was not likely to trigger powerful emotions or cause severe psychological stress or traumatic experiences. It did not involve physical risks to the participants or result in physical pain. Furthermore, the participants were informed of the study. Potential risks of participation (such as social risks, risks of criminal or civil liability, financial loss, professional disadvantages, or damage to reputation; risk due to a difficult security situation in the study room) did not exist. Participation in the study did not involve any type of deception [45,46]. No compensation was provided.

Results

Sample Characteristics

A total of 611 people entered the survey, of whom 349 completed it. An exact statement on the response rate cannot be made because we do not have information on the number of residents who received the questionnaire indirectly through our contacts. Data from 6 participants had to be excluded due to conspicuous response patterns (n=4), work in a hospital abroad (n=1), and specific information in the comment section (n=1). Furthermore, only data from participating physicians who had a clear understanding of information security as distinct from data protection were included. This was checked with the help of a single comprehension item. Overall, data from 281 participants could be used for our analyses.

Table 2 shows the sociodemographic characteristics of the study participants. A large majority of the participants were female (173/281, 61.6%). The two age groups with the highest frequency were 31-35 years (114/281, 40.6%) and 26-30 years (108/281, 38.4%). There was a total of 16 specialties represented by at least two participants. Most participants were part of a residency program for internal medicine (62/281, 22.1%), followed by anesthesiology (47/281, 16.7%), surgery (43/281, 15.3%), and pediatrics and adolescent medicine (30/281, 10.7%). Participants were distributed across all residency levels, with the fewest residents in their fourth year (43/281, 15.3%) and most residents in their fifth year or above (82/281, 29.2%). The majority of the participants worked in a public hospital (174/281, 61.9%), while approximately one-fifth (62/281, 22.1%) worked in a nonprofit hospital and 14.2% (40/281) in a private hospital. Most of the participants were employed in a university/teaching hospital (230/281, 81.9%). The size of the hospitals (measured by the number of beds) in which the residents underwent their training varied, whereby most of them worked in a hospital with more than 800 beds (109/281, 38.8%), followed by hospitals with 300-800 beds (104/281, 37%). Fewer than one-fifth (52/281, 18.5%) of the residents had a job in a hospital with less than 300 beds.

XSL•FO
RenderX

Table 2. Sample characteristics (n=281).

Variable	Value, n (%)
Gender	
Female	173 (61.6)
Male	108 (38.4)
Other	0 (0)
Age group (years)	
21-25	7 (2.5)
26-30	108 (38.4)
31-35	114 (40.6)
36-40	30 (10.7)
40 or older	22 (7.8)
Specialty	
Internal medicine	62 (22.1)
Anesthesiology	47 (16.7)
Surgery	43 (15.3)
Pediatrics and adolescent medicine	30 (10.7)
Other	99 (35.2)
Residency level	
First year	59 (21)
Second year	47 (16.7)
Third year	50 (17.8)
Fourth year	43 (15.3)
Fifth year or higher	82 (29.2)
Hospital sponsorship	
Private	40 (14.2)
Public	174 (61.9)
Nonprofit	62 (22.1)
I don’t know	5 (1.8)
University or teaching hospital	
Yes	230 (81.9)
No	48 (17.1)
I don’t know	3 (1.1)
Hospital size (beds)	
Fewer than 300	52 (18.5)
300-800	104 (37)
More than 800	109 (38.8)
I don’t know	16 (5.7)

Evaluation of the Measurement Model

We analyzed the model fit, the reliability, and the validity of the measurement model and respecified it accordingly until we received sufficient results. We began with an assessment of the global fit of our original measurement model. Table 3 shows the measurement model fit indices for the initial and the final models. It revealed that the CFI was below the recommended threshold, whereas the RMSEA and SRMR already showed acceptable values.

XSL•FO
RenderX

Table 3. Fit indices of the measurement model (N=281).

Model	Chi-square (*df*)	*P* value	CFI[a]	RMSEA[b]	SRMR[c]
Initial measurement model	1217.7 (657)	<.001	.933	.055	.070
Respecified final measurement model	794.3 (514)	<.001	.965	.044	.055
Recommended thresholds	__[d]	≥.05	≥.95	<.06	<.08

[a]CFI: comparative fit index.

[b]RMSEA: root mean squared error of approximation.

[c]SRMR: standardized root mean squared residual.

[d]Not applicable.

Looking at Cronbach α of the latent variables of the initial measurement model, we received a poor value (<0.70) for the latent variable AWAR. Here, we decided to drop the first two items (AWAR1 and AWAR2) because, compared to the third and fourth items, they seem to be too special and an own latent variable, asking about concrete information security–related knowledge. One reason could be that at the time of the survey, information security policies and officers were not yet implemented in most hospitals and could, therefore, not be known to the physicians. After removing those 2 items, the Spearman-Brown coefficient was acceptable.

We further respecified the model according to the modification indices and added an error covariance between QUAD1 and QUAD2. We then dropped items with loadings <.50 (TRFA1 and ITRE1). Although the χ^2 goodness-of-fit statistic was still significant (*P*<.001), these respecifications resulted in global fit indices above the thresholds, as shown in Table 3, meaning that the measurement model can be considered to be an acceptable approximation of the empirical data. According to the Fornell-Larcker [47] criterion, we also compared the square root of the average variance extracted (AVE) for each variable with the corresponding correlations of all other constructs [47]. Since the square root of the AVE was higher than the correlations on the respective horizontal and vertical lines, as shown in Table 4, the discriminant validity of the variables can be confirmed. In addition, the AVE of each variable should be higher than .50 to achieve acceptable convergent validity. All AVEs were above this threshold. Afterward, we looked again at Cronbach α and the composite reliability of each latent variable. All values were higher than the recommended threshold of .70, indicating acceptable reliability of the latent variables. In summary, we obtained acceptable model fit, reliability, and validity for the measurement model. A summary of the quality indicators of the measurement model is presented in Table 5.

XSL•FO
RenderX

Table 4. Mean, SD, and correlations after item exclusion (N=281).

Latent variable	Mean (SD)	1	2	3	4	5	6	7	8	9	10
1. QUAD	3.672 (0.688)										
r		**.778**[a]									
P value											
2. ROLC	3.172 (0.781)										
r		.422	**.746**								
P value		<.001									
3. TRFA	2.961 (0.688)										
r		-.225	-.393	**.732**							
P value		<.001	<.001								
4. LEAD	2.897 (0.857)										
r		-.341	-.391	.348	**.822**						
P value		<.001	<.001	<.001							
5. EDUC	2.898 (0.897)										
r		-.389	-.463	.300	.627	**.820**					
P value		<.001	<.001	<.001	<.001						
6. ITRE	2.367 (0.910)										
r		-.166	-.231	.398	.138	.141	**.813**				
P value		.002	<.001	<.001	.013	.005					
7. COMM	2.686 (0.901)										
r		-.155	-.199	.313	.173	.254	.350	**.792**			
P value		.005	<.001	<.001	.001	<.001	<.001				
8. WENG	3.471 (0.674)										
r		-.135	-.261	.184	.379	.277	.086	.118	**.808**		
P value		.033	<.001	.001	<.001	<.001	.226	.027			
9. AWAR	3.835 (0.971)										
r		-.087	-.076	.076	.188	.182	-.004	.256	.121	**.904**	
P value		.141	.244	.197	.002	.001	.797	<.001	.036		
10. COMP	3.563 (0.731)										
r		-.080	-.196	.175	.201	.239	.084	.337	.219	.502	**.776**
P value		.234	<.001	.002	<.001	<.001	.129	<.001	<.001	<.001	

[a]Bold values on the diagonal are the square root of the corresponding average variance extracted (AVE).

XSL•FO
RenderX

Table 5. Measurement model quality indicators (n=281).

Latent variable	Loadings	Cronbach α	Composite reliability	AVE[a]
Quantitative demands		0.854	.883	.605
QUAD1	.644	__[b]	—	—
QUAD2	.710	—	—	—
QUAD3	.899	—	—	—
QUAD4	.888	—	—	—
QUAD5	.715	—	—	—
Role conflicts		0.743	.789	.557
ROLC1	.827	—	—	—
ROLC2	.711	—	—	—
ROLC3	.693	—	—	—
Trust and fairness		0.720	.774	.536
TRFA1	Dropped	—	—	—
TRFA2	.689	—	—	—
TRFA3	.835	—	—	—
TRFA4	.661	—	—	—
Quality of leadership		0.861	.892	.676
LEAD1	.884	—	—	—
LEAD2	.872	—	—	—
LEAD3	.806	—	—	—
LEAD4	.716	—	—	—
Further education and training		0.853	.891	.672
EDUC1	.784	—	—	—
EDUC2	.910	—	—	—
EDUC3	.721	—	—	—
EDUC4	.852	—	—	—
IT resources		0.776	.851	.661
ITRE1	Dropped	—	—	—
ITRE2	.831	—	—	—
ITRE3	.934	—	—	—
ITRE4	.647	—	—	—
Information security–related communication		0.826	.870	.627
COMM1	.761	—	—	—
COMM2	.731	—	—	—
COMM3	.788	—	—	—
COMM4	.880	—	—	—
Work engagement		0.777	.846	.652
WENG1	.658	—	—	—
WENG2	.969	—	—	—
WENG3	.765	—	—	—
Information security–related awareness		0.855[c]	.899	.817
AWAR1	Dropped	—	—	—

XSL•FO
RenderX

Latent variable	Loadings	Cronbach α	Composite reliability	AVE[a]
AWAR2	Dropped	—	—	—
AWAR3	.866	—	—	—
AWAR4	.940	—	—	—
Information security–related compliance		0.789	.853	.602
COMP1	.896	—	—	—
COMP2	.801	—	—	—
COMP3	.842	—	—	—
COMP4	.504	—	—	—

[a]AVE: average variance extracted.

[b]Not applicable.

[c]Spearman-Brown coefficient.

Analysis of the Structural Model and Hypotheses Testing

The initial structural model already showed a good global fit (Table 6). Nevertheless, we followed a proposed modification index and added a direct effect of COMM on COMP, resulting in the final structural model. The R^2 value for the endogenous variable COMP was .512, indicating that 51.2% of the variance of this variable can be explained by our structural model (Figure 2). The R^2 values for the mediators were .229 (WENG) and .154 (AWAR).

Figure 2 also presents the significant direct effects and standardized path coefficients (β). A summary of the hypotheses tested is provided in Multimedia Appendix 3. The results show a significant relationship between LEAD and WENG (β=.414, *P*=.001). There were no significant unique effects of the other exogenous variables tested (QUAD, ROLC, TRFA, EDUC, and ITRE) on WENG (all *P*>.05). Furthermore, we found a significant positive relationship between EDUC and AWAR (β=.173, *P*=.01), as well as between COMM and AWAR (β=.349, *P*<.001). In addition, there was a significant negative relationship between ITRE and AWAR (β=–.206, *P*=.01). Both mediators WENG (β=.208, *P*=.001) and AWAR (β=.552, *P*<.001) were significantly associated with the endogenous variable COMP, which supports H1a and H1b.

In order to test the mediation hypotheses, we calculated the partial indirect effects. A summary of the results is also presented in Multimedia Appendix 3. The results showed that WENG significantly mediated the relationship between LEAD and COMP (*P*=.03), supporting hypothesis H2d. The positive indirect effect resulted from the positive direct effects between LEAD and WENG and between WENG and COMP. AWAR mediated the relationships between the exogenous variables EDUC (*P*=.02), ITRE (*P*= .02), and COMM (*P*<.001) and the endogenous variable COMP, which supports H3a and H3c. Since we expected a positive indirect effect of ITRE on COMP, H3b could not be confirmed. The positive indirect effects of EDUC and COMM resulted from the positive direct effects involved.

Table 6. Fit indices of the structural model (n=281).

Model	Chi-square (*df*)	*P* value	CFI[a]	RMSEA[b]	SRMR[c]
Initial structural model	799.1 (527)	<.001	.966	.043	.059
Respecified final structural model	769.5 (526)	<.001	.970	.041	.056
Recommended thresholds	—[d]	≥.05	≥.95	<.06	<.08

[a]CFI: comparative fit index.

[b]RMSEA: root mean squared error of approximation.

[c]SRMR: standardized root mean squared residual.

[d]Not applicable.

XSL•FO
RenderX

Figure 2. Structural model and standardized direct effects. *P<.05, **P<.01. Dashed gray arrows indicate statistically nonsignificant effects.

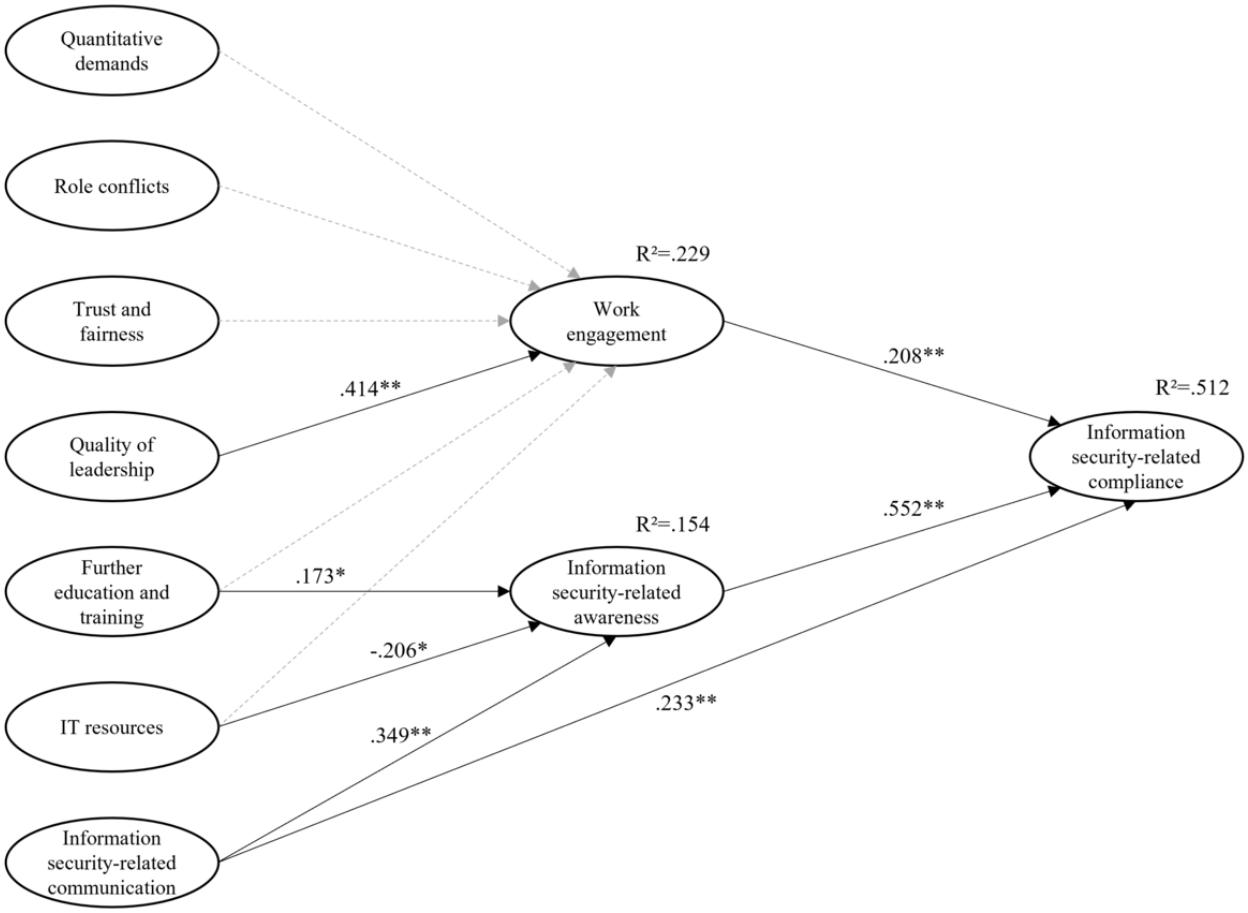

Discussion

Principal Findings and Comparison With Previous Work

With this study, we could statistically prove the influence of resident physicians' job resources and job demands on their information security–related compliance through the mediating roles of work engagement and awareness. This is one of the first studies to empirically analyze organizational antecedents of physicians' information security–related behavior in hospitals. Intensified research activities in this field are of great importance given that increasing digitization in hospitals and the associated risks from cybersecurity attacks require—in addition to technical solutions—well-considered organizational measures to enhance information security.

It was shown from the results that the information security–related compliance of resident physicians is significantly related to their work engagement. The assumed positive effect from our hypothesis was, therefore, confirmed. Pham et al [25] already described this relationship in their explorative study using interview analyses, which could be now supported by our results. In line with the findings of the study by Schaufeli and Bakker [22] and their revised version of the JDRM, in which they stated that "engagement is exclusively predicted by available job resources," we did not find unique relationships between the job demands involved (quantitative demands and role conflicts) and work engagement. Both demands also had no significant indirect effect on compliance. It can be, therefore, assumed that decreasing job demands would not be the most effective strategy in enhancing physicians' information security–related compliance, which would need to be further investigated by subsequent studies. Instead, our study highlights the importance of the job resource "quality of leadership," which had a significant indirect effect on information security–related compliance mediated by work engagement. Based on that, we can conclude that supervisors have the power to increase resident physicians' compliance through the motivational path: a good relationship with their direct supervisors, who are mostly senior physicians, increases resident physicians' work engagement, which leads to better compliance with the hospital's information security measures. According to our study, this relationship can be supported by the supervisor's efforts in the areas of staff development, job satisfaction, work planning, and conflict-solving. In light of the increasing digitization, this should be accompanied by digital and change management competencies [48].

Besides work engagement, awareness was the second mediator in our model, and based on our results, it is also significantly related to information security compliance. Here, our findings are in line with the results of several other studies, analyzing the role of awareness as a mediator in various industries [22,31,32]. We found 2 job resources with a unique positive effect on awareness and also a significant indirect effect on compliance: information security–related communication and

XSL•FO
RenderX

further education and training. Communication is a highly effective factor in improving information security–related compliance, as has been shown in other studies [16,17]. Regular, comprehensible information security–related communication increases physicians' awareness of potential information security risks and threats. Here, we see a particular responsibility on the part of the hospital management. A communication strategy should be developed that makes all employees aware of the risks and threats and their particular information security responsibilities. In addition, information security–related communication had a direct effect on compliance. It could be assumed here that good communication also triggers automatic mechanisms for resolving critical situations. Furthermore, our results show that awareness can be created not only by communication but also by good education and training. Modern training concepts take a holistic approach and use various methods to raise awareness, for example, through personalized phishing simulations for training purposes and classic learning sessions with subsequent skills tests and personalized follow-up training based on individual results. Contrary to our hypothesis, IT resources are negatively related to awareness. One possible explanation could be that digitization and technologization increase confidence in the security of the tools used. On the other hand, if only outdated IT resources are available, physicians might be more sensitive to possible security risks.

Overall, our study provides initial insights into how the hospital as an organization could influence the individual information security–related behavior of resident physicians during clinical practice. In our opinion, the core element is the creation of an information security culture as a subculture of the organizational culture, which increases both the work engagement and the awareness of resident physicians through good leadership, information, and training.

Limitations

A few limitations should be taken into account when considering and evaluating our results, one of which pertains to the study design. First, since our study was conducted in Germany, it is not possible to directly apply the results to other countries. However, our study is intended to create incentives to conduct studies in hospitals in other countries to explore organizational factors for improving information security–related behavior. Second, when using a survey, comprehension problems on the part of the participants cannot be identified and addressed. However, we believe that the approach was the most suitable for the aim of the study and the selected target group as—for example—interviews would have resulted in a much smaller sample, which might have reduced the validity of the results. Third, the cross-sectional design does not allow us to form causal relationships between organizational antecedents, work engagement, awareness, and compliance, but only correlations. Any alterations in physicians' behavior due to changes in organizational factors can only be assessed with a longitudinal design. Nevertheless, we assume causal relationships based on logical considerations.

Even though it was an anonymous questionnaire, the risk of social desirability bias remains, with participants trying to be much more positive about their job-related attitudes and behavior. Furthermore, the fact that we cannot precisely determine the response rate represents another limitation. Since we do not have information on the number of residents who received the questionnaire indirectly through our contacts, for example, through chief physicians or the hospital management, we are not able to make an exact statement on the number of residents being invited, which is the basis for calculating the response rate. However, based on the information available to us, we estimate that fewer than 10% of those who received the questionnaire actually responded. If there are systematic differences between the responders and nonresponders, the results of our survey may not be representative of the target population. This so-called nonresponse bias may threaten the external validity of our study by reducing the representativeness of the results. However, previous research suggests that physician surveys are less susceptible to nonresponse bias than general population studies because they are a more homogeneous study population [49]. Furthermore, it is not always the case that a low response rate automatically reduces the representativeness, which is why the response rate should not be considered in isolation [50]. We believe the main reason for the low response rate was the heavy workload, which did not allow physicians the time to participate. Another indicator of the representativeness of a study is the sampling method [51]. The approach chosen for the data collection, in which residents were invited through different channels, may have led to a selection bias. For example, it is possible that digitally active and networked physicians were primarily addressed. The high proportion of physicians who received the questionnaire through the Hartmannbund may also contribute to the selection bias. Our study may have appealed to physicians with a higher average digital affinity than in the target population. To sum up, statistical conclusions on the entire target population should, therefore, always be drawn, taking into account the supposedly limited representativeness.

Future studies could use a longitudinal design to investigate the specific introduction of information security–related measures and the associated change in behavior. Furthermore, objective criteria, such as the number of breaches, could also be used as a measure of compliance instead of subjective self-reports. In addition, qualitative studies in the form of interviews are suitable for recording specific activities, opinions, perceptions, and concerns of physicians. It would also be interesting to conduct a study among senior physicians who have been working in hospitals for significantly longer and may need other measures to ensure that they behave in accordance with information security. For example, existing habits may be more difficult to change by adjusting resources and demands, and the resource "quality of leadership" may play a weaker role than for resident physicians.

Conclusions

A high standard of information security can only be achieved holistically through coordinated technical and organizational measures. In hospitals, this includes achieving a great level of work engagement among physicians through good leadership, but also creating awareness of the risks and threats relating to information security through communication and training. Hospital management is required to establish an information

XSL•FO
RenderX

security culture that is informative and motivating and that raises awareness.

Acknowledgments

This study was actively supported by the Hartmannbund. The APF is paid for under the ATLAS project “Innovation and digital transformation in healthcare” funded by the State of North Rhine-Westphalia, Germany (grant ITG-1-1). The authors declare that they did not use generative artificial intelligence (AI) in any portion of the manuscript writing.

Data Availability

The datasets used and/or analyzed in the context of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JK designed the study, and SB-J made important contributions during this phase. With the kind support of the Hartmannbund, JK recruited most of the participants. SB-J helped distribute the questionnaire through her professional network. JK analyzed and interpreted the data. JK wrote the manuscript. SB-J was a major contributor to the revision of the manuscript. Both authors have approved the manuscript before submission. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Constructs, definitions, and corresponding items of the questionnaire.
[DOCX File , 38 KB - jmir_v27i1e46257_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 24 KB - jmir_v27i1e46257_app2.docx]

Multimedia Appendix 3
Summary of the hypotheses tested.
[DOCX File , 16 KB - jmir_v27i1e46257_app3.docx]

References

1. Willing M, Dresen C, Haverkamp U, Schinzel S. Analyzing medical device connectivity and its effect on cyber security in german hospitals. BMC Med Inform Decis Mak 2020;20(1):246 [FREE Full text] [doi: 10.1186/s12911-020-01259-y] [Medline: 32993623]
2. Argaw ST, Bempong NE, Eshaya-Chauvin B, Flahault A. The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review. BMC Med Inform Decis Mak 2019;19(1):10 [FREE Full text] [doi: 10.1186/s12911-018-0724-5] [Medline: 30634962]
3. Yaqoob T, Abbas H, Atiquzzaman M. Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices—a review. IEEE Commun. Surv. Tutorials 2019;21(4):3723-3768. [doi: 10.1109/COMST.2019.2914094]
4. Jalali MS, Kaiser JP. Cybersecurity in hospitals: a systematic, organizational perspective. J Med Internet Res 2018;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
5. Wani TA, Mendoza A, Gray K, Smolenaers F. Status of bring-your-own-device (BYOD) security practices in Australian hospitals – a national survey. Health Policy and Technology 2022;11(3):100627. [doi: 10.1016/j.hlpt.2022.100627]
6. Kraushaar J, Bohnet-Joschko S. Prevalence and patterns of mobile device usage among physicians in clinical practice: a systematic review. Health Informatics J 2023;29(2):14604582231169296 [FREE Full text] [doi: 10.1177/14604582231169296] [Medline: 37063054]
7. Millard WB. Where bits and bytes meet flesh and blood. Annals of Emergency Medicine 2017;70(3):A17-A21. [doi: 10.1016/j.annemergmed.2017.07.008]
8. Argaw ST, Troncoso-Pastoriza JR, Lacey D, Florin MV, Calcavecchia F, Anderson D, et al. Cybersecurity of hospitals: discussing the challenges and working towards mitigating the risks. BMC Med Inform Decis Mak 2020;20(1):146 [FREE Full text] [doi: 10.1186/s12911-020-01161-7] [Medline: 32620167]
9. Kruse CS, Frederick B, Jacobson T, Monticone DK. Cybersecurity in healthcare: a systematic review of modern threats and trends. Technol Health Care 2017;25(1):1-10. [doi: 10.3233/THC-161263] [Medline: 27689562]

XSL•FO
RenderX

10. Tully J, Selzer J, Phillips JP, O'Connor P, Dameff C. Healthcare challenges in the era of cybersecurity. Health Secur 2020;18(3):228-231. [doi: 10.1089/hs.2019.0123] [Medline: 32559153]
11. Klick J, Koch R, Brandstetter T. Epidemic? The attack surface of German hospitals during the COVID-19 pandemic. : IEEE; 2021 Presented at: Proceedings of the 13th International Conference on Cyber Conflict (CyCon); May 25-28, 2021; Tallin, Estonia p. 73-94. [doi: 10.23919/cycon51939.2021.9468304]
12. Pollini A, Callari TC, Tedeschi A, Ruscio D, Save L, Chiarugi F, et al. Leveraging human factors in cybersecurity: an integrated methodological approach. Cogn Technol Work 2022;24(2):371-390 [FREE Full text] [doi: 10.1007/s10111-021-00683-y] [Medline: 34149309]
13. Branchenspezifischer Sicherheitsstandard „Medizinische Versorgung“. Deutsche Krankenhaus Gesellschaft. 2022 Dec 8. URL: https://www.dkgev.de/fileadmin/default/Mediapool/2_Themen/2.1_Digitalisierung_Daten/2.1.4._IT-Sicherheit_und_technischer_Datenschutz/2.1.4.1._IT-Sicherheit_im_Krankenhaus/Branchenspezifischer_Sicherheitsstandard_Medizinische_Versorgung_v1.2_Stand_2022-12-08.pdf [accessed 2024-11-10]
14. Kuppusamy P, Samy GN, Maarop N, Magalingam P, Kamaruddin N, Shanmugam B, et al. Systematic literature review of information security compliance behaviour theories. J. Phys.: Conf. Ser 2020;1551(1):012005. [doi: 10.1088/1742-6596/1551/1/012005]
15. Hu Q, Dinev T, Hart P, Cooke D. Managing employee compliance with information security policies: the critical role of top management and organizational culture*. Decision Sciences 2012;43(4):615-660. [doi: 10.1111/j.1540-5915.2012.00361.x]
16. D'Arcy J, Greene G. Security culture and the employment relationship as drivers of employees' security compliance. Info Mngmnt & Comp Security 2014;22(5):474-489. [doi: 10.1108/imcs-08-2013-0057]
17. Solomon G, Brown ITJ. The influence of organisational culture and information security culture on employee compliance behaviour. JEIM 2020;34(4):1203-1228. [doi: 10.1108/jeim-08-2019-0217]
18. Yeng PK, Fauzi MA, Yang B. A comprehensive assessment of human factors in cyber security compliance toward enhancing the security practice of healthcare staff in paperless hospitals. Information 2022;13(7):335. [doi: 10.3390/info13070335]
19. Hedström K, Kolkowska E, Karlsson F, Allen JP. Value conflicts for information security management. The Journal of Strategic Information Systems 2011;20(4):373-384. [doi: 10.1016/j.jsis.2011.06.001]
20. Kraushaar J, Bohnet-Joschko S. Smartphone use and security challenges in hospitals: a survey among resident physicians in Germany. Int J Environ Res Public Health 2022;19(24):16546 [FREE Full text] [doi: 10.3390/ijerph192416546] [Medline: 36554426]
21. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001;86(3):499-512. [Medline: 11419809]
22. Schaufeli WB, Bakker AB. Job demands, job resources, and their relationship with burnout and engagement: a multi‐sample study. J Organ Behavior 2004;25(3):293-315. [doi: 10.1002/job.248]
23. Schaufeli WB, Taris TW. A critical review of the job demands-resources model: implications for improving work and health. In: Bauer GF, Hämmig O, editors. Bridging Occupational, Organizational and Public Health. Dordrecht, the Netherlands: Springer; 2014:43-68.
24. Taris TW, Schaufeli WB. The job demands-resources model. In: Clarke S, Probst TM, Guldenmund F, Passmore J, editors. The Wiley Blackwell Handbook of the Psychology of Occupational Safety and Workplace Health. Hoboken, NJ: Wiley; 2015:155-180.
25. Pham HC, El-Den J, Richardson J. Stress-based security compliance model – an exploratory study. ICS 2016;24(4):326-347. [doi: 10.1108/ics-10-2014-0067]
26. Borkowski N, Meese KA. Organizational Behavior in Health Care. Fourth Edition. Burlington, MA: Jones et Bartlett Learning; 2021.
27. Nahavandi A, Denhardt RB, Denhardt JV, Aristigueta MP. Organizational Behavior (Sage Texts). New Delhi, India: SAGE Publications; 2015.
28. Amankwa E, Loock M, Kritzinger E. Establishing information security policy compliance culture in organizations. ICS 2018;26(4):420-436. [doi: 10.1108/ics-09-2017-0063]
29. Li L, Xu L, He W, Chen Y, Chen H. Research and practical issues of enterprise information systems. In: Tjoa AM, Xu L DA, Raffai M, Novak NM, editors. Cyber Security AwarenessIts Impact on Employee's Behavior. Cham, Switzerland: Springer International Publishing; 2016:103-111.
30. Karlsson F, Karlsson M, Åström J. Measuring employees’ compliance – the importance of value pluralism. ICS 2017;25(3):279-299. [doi: 10.1108/ics-11-2016-0084]
31. Hwang I, Wakefield R, Kim S, Kim T. Security awareness: the first step in information security compliance behavior. Journal of Computer Information Systems 2019;61(4):345-356. [doi: 10.1080/08874417.2019.1650676]
32. Koohang A, Anderson J, Nord JH, Paliszkiewicz J. Building an awareness-centered information security policy compliance model. IMDS 2019;120(1):231-247. [doi: 10.1108/imds-07-2019-0412]
33. Burr H, Berthelsen H, Moncada S, Nübling M, Dupret E, Demiral Y, international COPSOQ Network. The third version of the Copenhagen psychosocial questionnaire. Saf Health Work 2019;10(4):482-503 [FREE Full text] [doi: 10.1016/j.shaw.2019.10.002] [Medline: 31890332]

XSL•FO
RenderX

34. Lincke HJ, Vomstein M, Lindner A, Nolle I, Häberle N, Haug A, et al. COPSOQ III in Germany: validation of a standard instrument to measure psychosocial factors at work. J Occup Med Toxicol 2021;16(1):50 [FREE Full text] [doi: 10.1186/s12995-021-00331-1] [Medline: 34784940]
35. Keller M, Bamberg E, Kersten M, Nienhaus A. Entwicklung eines instruments zur stressbezogenen arbeitsanalyse für klinikärztinnen und -ärzte (ISAK). Z Arb Wiss 2010;64:337-353. [doi: 10.1037/t70200-000]
36. Keller M, Bamberg E, Kersten M, Nienhaus A. Validierung des Instruments zur stressbezogenen Arbeitsanalyse für Klinikärztinnen und -ärzte (ISAK). Zeitschrift für Arbeits- und Organisationspsychologie A&O 2013;57(1):3-21. [doi: 10.1026/0932-4089/a000096]
37. Bulgurcu B, Cavusoglu H, Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Quarterly 2010;34(3):523-548. [doi: 10.2307/25750690]
38. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
39. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;76(6):913-934 [FREE Full text] [doi: 10.1177/0013164413495237] [Medline: 25705052]
40. Kyriazos TA. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. PSYCH 2018;09(08):2207-2230. [doi: 10.4236/psych.2018.98126]
41. Bollen KA, Noble MD. Structural equation models and the quantification of behavior. Proc Natl Acad Sci U S A 2011;108 Suppl 3(Suppl 3):15639-15646 [FREE Full text] [doi: 10.1073/pnas.1010661108] [Medline: 21730136]
42. Kline RB, St C. Principles and Practice of Structural Equation Modeling. New York City, NY: Guilford Press; 2022.
43. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 1988;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
44. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
45. When do I need a statement by an ethics committee? Deutsche Forschungsgemeinschaft. 2023. URL: https://www.konsortswd.de/wp-content/uploads/RatSWD_Output9_Forschungsethik.pdf [accessed 2024-09-21]
46. Rat für Sozial- und Wirtschaftsdaten (RatSWD). Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. Output Series 2017;9(5) [FREE Full text]
47. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981;18(1):39-50. [doi: 10.2307/3151312]
48. Balling S, Bohnet-Joschko S. Ärztliche führung im krankenhaus der zukunft. Gesundheitsökonomie & Qualitätsmanagement 2022;28(02):89-94. [doi: 10.1055/a-1920-3609]
49. Kellerman SE, Herold J. Physician response to surveys. A review of the literature. Am J Prev Med 2001;20(1):61-67. [doi: 10.1016/s0749-3797(00)00258-0] [Medline: 11137777]
50. Schouten B, Cobben F, Bethlehem J. Indicators for the representativeness of survey response. Survey Methodol 2008;35(1):101-113 [FREE Full text]
51. Ortmanns V, Schneider S. Can we assess survey representativeness of cross-national surveys using the education variable? Survey Research Methods 2016;10(3):189-210. [doi: 10.18148/srm/2016.v10i3.6608]

Abbreviations

AVE: average variance extracted
BYOD: bring your own device
CFI: comparative fit index
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COPSOQ: Copenhagen Psychosocial Questionnaire
GDPR: General Data Protection Regulation
JDRM: Job Demands-Resources Model
OB: Organizational Behavior
RMSEA: root mean squared error of approximation
SEM: structural equation modeling
SRMR: standardized root mean squared residual
TPB: Theory of Planned Behavior
WLSMV: weighted least squares means and variance adjusted

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 03.02.23; peer-reviewed by B von Solms, I Schiering, J Klein; comments to author 29.01.24; revised version received 30.03.24; accepted 11.10.24; published 07.01.25.

Please cite as:
Kraushaar J, Bohnet-Joschko S
The Role of the Organization in Promoting Information Security–Related Behavior Among Resident Physicians in Hospitals in Germany: Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e46257
URL: https://www.jmir.org/2025/1/e46257
doi:10.2196/46257
PMID:

©Judith Kraushaar, Sabine Bohnet-Joschko. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Challenges and Opportunities for Data Sharing Related to Artificial Intelligence Tools in Health Care in Low- and Middle-Income Countries: Systematic Review and Case Study From Thailand

Aprajita Kaushik[1*], BSc, MSc; Capucine Barcellona[1*], BA; Nikita Kanumoory Mandyam[1], BSc; Si Ying Tan[1†], PhD; Jasper Tromp[1,2], MD, MPH, PhD

[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[2]Duke-NUS Medical School, Singapore, Singapore

[†]deceased

[*]these authors contributed equally

Corresponding Author:
Jasper Tromp, MD, MPH, PhD
Saw Swee Hock School of Public Health
National University of Singapore
Tahir Foundation Building
12 Science Drive 2, #10-01
Singapore, 117549
Singapore
Phone: 65 6516 4988
Email: jasper_tromp@nus.edu.sg

Abstract

Background: Health care systems in low- and middle-income countries (LMICs) can greatly benefit from artificial intelligence (AI) interventions in various use cases such as diagnostics, treatment, and public health monitoring but face significant challenges in sharing data for developing and deploying AI in health care.

Objective: This study aimed to identify barriers and enablers to data sharing for AI in health care in LMICs and to test the relevance of these in a local context.

Methods: First, we conducted a systematic literature search using PubMed, SCOPUS, Embase, Web of Science, and ACM using controlled vocabulary. Primary research studies, perspectives, policy landscape analyses, and commentaries performed in or involving an LMIC context were included. Studies that lacked a clear connection to health information exchange systems or were not reported in English were excluded from the review. Two reviewers independently screened titles and abstracts of the included articles and critically appraised each study. All identified barriers and enablers were classified according to 7 categories as per the predefined framework—*technical*, *motivational*, *economic*, *political*, *legal and policy*, *ethical*, *social*, *organisational*, and *managerial*. Second, we tested the local relevance of barriers and enablers in Thailand through stakeholder interviews with 15 academic experts, technology developers, regulators, policy makers, and health care providers. The interviewers took notes and analyzed data using framework analysis. Coding procedures were standardized to enhance the reliability of our approach. Coded data were reverified and themes were readjusted where necessary to avoid researcher bias.

Results: We identified 22 studies, the majority of which were conducted across Africa (n=12, 55%) and Asia (n=6, 27%). The most important data-sharing challenges were unreliable internet connectivity, lack of equipment, poor staff and management motivation, uneven resource distribution, and ethical concerns. Possible solutions included improving IT infrastructure, enhancing funding, introducing user-friendly software, and incentivizing health care organizations and personnel to share data for AI-related tools. In Thailand, inconsistent data systems, limited staff time, low health data literacy, complex and unclear policies, and cybersecurity issues were important data-sharing challenges. Key solutions included building a conducive digital ecosystem—having shared data input platforms for health facilities to ensure data uniformity and to develop easy-to-understand consent forms, having standardized guidelines for data sharing, and having compensation policies for data breach victims.

Conclusions: Although AI in LMICs has the potential to overcome health inequalities, these countries face technical, political, legal, policy, and organizational barriers to sharing data, which impede effective AI development and deployment. When tested in a local context, most of these barriers were relevant. Although our findings might not be generalizable to other contexts, this

XSL•FO
RenderX

study can be used by LMICs as a framework to identify barriers and strengths within their health care systems and devise localized solutions for enhanced data sharing.

Trial Registration: PROSPERO CRD42022360644; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=360644

(J Med Internet Res 2025;27:e58338) doi:10.2196/58338

KEYWORDS

artificial intelligence; data sharing; health care; low- and middle-income countries; AI tools; systematic review; case study; Thailand; computing machinery; academic experts; technology developers; health care providers; internet connectivity; data systems; low health data literacy; cybersecurity; standardized data formats; AI development; PRISMA

Introduction

Overview

Artificial intelligence (AI) has the potential to revolutionize health care by empowering health care practitioners in supporting clinical decision-making, enhancing diagnostic accuracy, and refining administrative processes for health care providers, payers, and pharmaceutical organizations [1-5]. Applications such as AI-driven predictive analysis and personalized treatments can improve health care delivery, even in low- and middle-income countries (LMICs). However, despite the promising aspects of AI in health care, there are significant challenges related to limited data sharing and accessibility that impede the development and localization of AI applications in health care, particularly in LMICs [1-3,6-9].

Rationale

Previous studies have identified challenges to health information exchange (HIE) and data sharing, more generally [10-14]. Akhlaq et al [10] reviewed barriers and facilitators to implementing and adopting HIE. This systematic review identified 7 themes influencing HIE—sociopolitical, financial, infrastructure, organizational, technical, individual, and data management [10]. Similarly, Figueiredo [11] and van Panhuis et al [13] identified 20 potential barriers to data sharing across 6 domains—technical, motivational, economic, political, legal, and ethical. However, gaps exist in current research. First, these studies did not focus specifically on the granular and often sensitive data required for AI models in health care, such as high-quality clinical datasets and high-resolution medical images. The complexity of these data types introduces additional challenges, such as ensuring data interoperability across different systems and institutions, addressing heightened privacy concerns, and maintaining data quality for AI algorithms. Furthermore, current studies do not fully test these AI-specific barriers. Second, the absence of empirical testing of frameworks using in-depth formative research is also a pressing gap across current literature. Engaging stakeholders (eg, policy makers, health care providers, and data scientists) in LMICs is critical for identifying practical solutions and for refining theoretical frameworks to fit real-world contexts. Defining clear metrics for the success of data sharing, for example, improved interoperability, increased data accessibility, and reduced privacy risks within national and organizational frameworks, not only informs scalable, adaptable solutions for AI in LMICs but also allows for their evaluation [15].

Thailand

Thailand is developing its AI capabilities and promoting AI adoption, but significant obstacles remain, particularly regarding data sharing. The fragmented nature of its health care service landscape and unclear data-sharing guidelines restrict the effective use of AI in health care. The Thailand National AI Strategy and Action Plan (2022-2027) attempts to close this gap by establishing a data-sharing guideline to enable AI deployment [16]. According to the Government AI Readiness Index 2021, Thailand ranked 59th globally and 9th in East Asia, demonstrating moderate progress in AI readiness but highlighting areas that require improvement, especially in health data governance [17,18].

Aims

This study aimed to address the research gaps by conducting a systematic review to identify the barriers and enablers to data sharing for AI in LMICs. The study further builds on existing literature by testing these findings through stakeholder interviews in Thailand—collecting empirical evidence on local challenges and enablers for AI-related data sharing and future recommendations for sharing data related to AI tools in Thailand. This mixed methods approach allows us to capture both the broader trends from LMICs and context-specific insights from Thailand, contributing to a more nuanced understanding of the intersection between AI and health care data sharing. Our findings will inform policy makers, health care organizations, and technology developers on building effective AI data-sharing frameworks, closing the gap between policy and implementation.

Methods

Overview

We conducted the study in 2 phases. First, we performed a systematic search to identify the literature assessing barriers and facilitators to sharing data for AI in health care. We synthesized and categorized findings using a predefined framework. Second, we conducted semistructured qualitative stakeholder interviews, which we analyzed thematically using the framework's domains. The qualitative interviews' findings validate, complement, and reiterate findings from the systematic review to a localized context (ie, Thailand) and supplement possible gaps in the framework.

Framework

We devised a predefined framework to categorize barriers and enablers on the basis of 2 previously published frameworks for

XSL•FO
RenderX

data sharing [13,19]. The framework included 7 domains categorizing barriers and enablers to data sharing: technical; motivational; economic; political, legal, and policy; ethical; social; organizational, and managerial. These categories are defined in detail in Textbox 1.

Textbox 1. Classification of factors of data sharing and their definitions.

- Technical: barriers and enablers relating to health information system capacity and other technological tools
- Motivational: barriers and enablers relating to personal or institutional motivations to share data
- Economic: barriers and enablers relating to the financial sustainability and costs of data sharing
- Political, legal, and policy: barriers and enablers relating to political principles, regulations, and public policy
- Ethical: barriers and enablers relating to moral principles, values, and ethical guidelines
- Social: barriers and enablers relating to societal norms and attitudes toward data sharing and health technology
- Organizational and managerial: barriers and enablers relating to organizations' strategy, leadership, and management practices

Systematic Review

Search Strategy and Inclusion and Exclusion Criteria

We used the SPICE (settings, perspective, intervention, comparison, and evaluation) framework (Textbox 2) to inform our search strategy and inclusion and exclusion criteria. The selection process focused on studies that explored AI tools, machine learning (ML), electronic decision support systems, or digital interventions, specifically in the context of HIE. We included primary research studies (empirical studies investigating the development, implementation, or evaluation of digital health tools related to HIE); perspectives (opinion or viewpoint articles offering expert insights into AI, ML, electronic decision support systems, or HIE systems); policy landscape analyses (reviews examining the policy or regulatory environment around digital health and HIE); and commentaries (discussions of challenges, advancements, or ethical considerations related to the use of digital interventions in HIE). Only studies published until October 2022 were considered. The inclusion criteria required studies to have a specific focus on the technical and practical aspects of HIE, such as how AI or ML tools facilitate data exchange, the role of interoperability, or the establishment of secure access protocols. Articles that did not address these core components or lacked a clear connection to HIE systems were excluded from the review. All included studies had to be performed in or involve an LMIC context, according to the World Bank country classification (fiscal year 2019) [20]. We excluded studies not reported in English. We defined AI in health care as any software or algorithm made to automate intelligent behavior in a health care setting for diagnosis or treatment directed toward patients, caregivers, health care providers, or a combination [21]. We registered the review protocol with PROSPERO before commencing data extraction (CRD42022343604).

We systematically searched PubMed, SCOPUS, Embase, Web of Science, and ACM. In the search strategy, we incorporated keywords and controlled vocabulary terms for the key concepts of data sharing, AI tools, health care, and barriers and enablers (Multimedia Appendix 1). A librarian at the National University of Singapore vetted the strategy.

Textbox 2. The SPICE (settings, perspective, intervention, comparison, and evaluation) framework for conducting the systematic review.

S (settings)

- Health care domain

P (perspective)

- Experiences, perspectives related to barriers and enablers to data sharing

I (intervention)

- Artificial intelligence tools, machine learning, digital health, electronic decision support systems

C (comparison)

- Traditional health setting or not applicable

E (evaluation)

- Challenges and barriers to data sharing
- Facilitators and enablers to data sharing
- Identification of groups or stakeholders and their perspectives

XSL•FO
RenderX

Data Synthesis

The retrieved search results were imported to Clarivate Endnote (version 20.3) and deduplicated. The remaining articles were imported to Rayyan. Two reviewers independently screened the titles and abstracts of the included articles. All articles that potentially met the inclusion criteria and were agreed upon by both reviewers were included in the review for full-text screening. A third reviewer resolved conflicts.

Data extraction tools were developed, and the data were synthesized. These included article title, first author, year of publication, study characteristics—including country of origin, study design (qualitative and semistructured), description of AI tool, and a summary of barriers and enablers to data sharing. Data synthesis was on the basis of The Joanna Briggs Institute's (JBI) meta-aggregative approach for qualitative systematic reviews [22]. The qualitative data were pooled and represented as a narrative summary for each study per domain.

Critical Appraisal

Two reviewers (AK and CB) critically appraised each study independently using the Critical Appraisal Skills Programme checklist for qualitative studies (Multimedia Appendix 2) [1-3,6-9,23-36] [37]. The Critical Appraisal Skills Programme checklist was used because it is appropriate for all types of qualitative research (including landscape analysis and perspectives).

Stakeholder Interviews

We tested the relevance of barriers and enablers using Thailand as a case study. Thailand was chosen because of the publication of its recent National AI Strategy and Action Plan (2022-2027) and because coauthors of this study had access to local stakeholders [16]. We conducted an exploratory qualitative study to better understand the various challenges faced by the stakeholders in data sharing for AI and the facilitators that help overcome this at a local level.

Study Design

We identified and mapped potentially relevant stakeholders through social media (LinkedIn), search engines, personal, company (university and ministry), and official websites. We included stakeholders aged ≥21 years who could communicate verbally in English, academic experts, technology developers, regulators, or policy makers in governance, regulation, and implementation of digital health technology and novel technology in Thai health care. We used purposive and snowball sampling techniques for stakeholders until saturation was reached.

All mapped stakeholders were contacted via an email titled "Invitation to discuss data sharing for artificial intelligence in Thailand" that included the semistructured interview guide (Multimedia Appendix 3) and the standard letter of invite entailing a short summary of the project, information for the participants regarding the type of data collected, and how data will be used. This aimed to ensure participants could make informed decisions about which interview questions they preferred to answer.

All 5 authors conducted stakeholder interviews on Zoom from December 20, 2022, to February 10, 2023. Interviews followed a semistructured interview guide. The interview guide was vetted by all authors and included relevant comments as per each author's expertise and was further revetted by our partners at the Office of National Higher Education Science Research and Innovation Policy Council, Thailand, to test the local relevance of the topics included and to avoid any potential bias.

Data Collection

Interviewers took detailed written notes, which were subsequently reviewed for accuracy and completeness. The notes were then coded based on predefined domains of barriers and facilitators aligned with the framework. Two researchers independently coded the data using MS Excel, systematically comparing coded segments and themes across interviewees. Illustrative, nonattributable quotations supported the findings, providing a rich, contextual understanding of the data. The coded data were reverified during the analysis, and themes were readjusted where necessary with consensus to avoid researcher bias.

Data Analysis

We conducted a thematic analysis using framework analysis to qualitatively analyze the interview data. This approach is well-suited for applied policy research, as it systematically explores stakeholders' views, knowledge, and perspectives while retaining the contextual integrity of individual narratives. The framework analysis method involves 5 distinct stages: familiarization, identifying a thematic framework, indexing, charting, and mapping and interpretation, which provides a structured approach to organizing and interpreting the data [38].

Throughout the analysis, reflexivity was emphasized, as the researchers continuously examined their own perspectives and potential biases by engaging in discussions to ensure that personal assumptions did not influence the interpretation of the data. We addressed the 3 key criteria for ensuring trustworthiness in qualitative research. (1) Confirmability: we documented the coding and thematic development processes and conducted consensus meetings between coders to ensure that the findings were grounded in the data rather than influenced by researcher bias. (2) Dependability: to enhance the reliability of our approach, we standardized the coding procedures. Consistent data management practices were followed, and any discrepancies between coders were resolved through discussion. (3) Credibility: illustrative, nonattributable quotations supported the themes and ensured that the findings accurately reflected the participants' views.

Ethical Considerations

The National University of Singapore Institutional Review Board approved the qualitative component of this study (NUS-IRB-2022-550). No interviews were recorded, and consent was obtained via email and at the start of the interview. The interview data were not anonymized and were retained solely for potential future contact regarding the study. Participants were informed and consented that their personal information (limited to name, email address, and designation) would be securely stored according to the National University

XSL•FO
RenderX

of Singapore's data storage principles and policies, accessible only to the project team. Consent was obtained to use select quotes from the interviews, ensuring that no identification of individual participants is possible. No compensation in any form was provided to participants.

Results

Systematic Review

Study Characteristics

We retrieved 3512 records, with 2471 articles left after removing duplicates (Figure 1). After excluding titles and abstracts, 58 articles were eligible for full-text screening. Finally, the data extraction and qualitative synthesis included 22 articles published between January 2010 and October 2022 that fulfilled the inclusion criteria.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for included studies. HIE: health information exchange; RQ: research question.

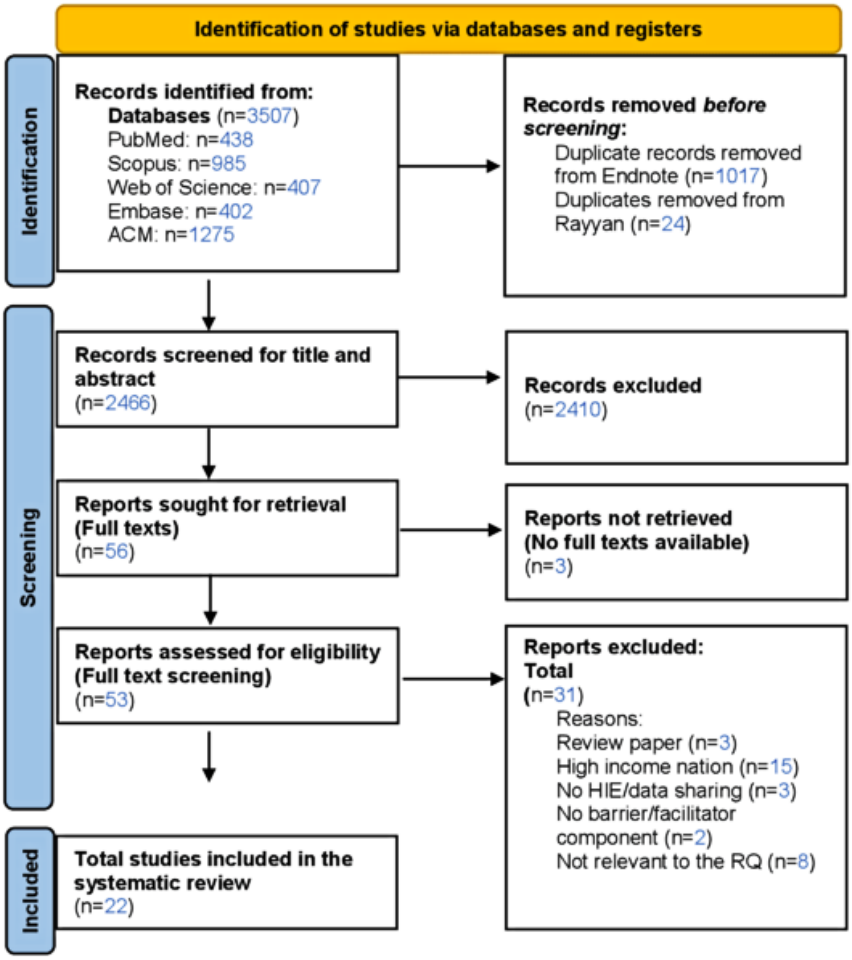

Quality Appraisal

Table S1 in Multimedia Appendix 2 presents the quality appraisal results. All studies reported a clear statement of the aims and findings of the research. The qualitative methodology was appropriate to all studies, while the research design was appropriate to all except one [3]. The appropriateness of recruitment strategy and data collection was unclear or inappropriate for 7 [2,3,23-26,39] and 6 studies [2,3,23,25-27], respectively. Whether the relationship between researchers and participants or ethical issues was adequately considered was unclear or not appropriate for 11 [1,2,6,23,24,26-31] and 12 studies [1,6,23,24,26-28,30,32-34], respectively. All studies were valuable to the local context.

Table 1 shows the characteristics of the included studies. Of all the studies, 8 (36%) were landscape analyses, 7 (32%) had a qualitative design, 2 (9%) were mixed methods, 3 (14%) were case studies, and 2 (9%) were perspective articles. Of all the studies, 12 (55%) studies were from Africa, 6 (27%) from Asia, and 2 (9%) from Latin America and the Caribbean.

XSL•FO
RenderX

Table 1. Characteristics and description of included studies.

Study	Country	Study design	AI[a] tool
Akhlaq et al [35], 2020	Pakistan	Exploratory mixed methods case study	HIE[b] in Pakistan
Bhattacharya and Pandey [23], 2021	Global	Analysis	Internet of things
Botha et al [7], 2015	South Africa	Qual-quant mixed methods analysis with stakeholder interviews and a survey	eHealth systems in South Africa
Chali et al [2], 2017	Tanzania	Case study	Mobile DEC[c] for Tanzanian EHRs[d]
Chukwu et al [1], 2022	Sierra Leone	Semiqualitative primary study	Digital health solutions eIDSR[e] and DHIS2[f], data sharing practices, and current data use in Sierra Leone
de Fatima Marin et al [28], 2022	Brazil	Digital health landscape analysis	National Health Data Network (RNDS[g])
deRiel et al [31], 2018	Haiti	Analysis	EMR[h] program
Fakhkhari et al [6], 2019	Morocco	Analysis	EHR in public hospitals inside and outside Morocco
Guidry et al [8], 2010	Kenya	Information systems landscape analysis	Data models and information systems in Kenya for HIV care
Keny et al [33], 2015	Kenya	Primary qualitative case study	EHR in Kenya
Lei et al [24], 2017	China	Primary qualitative case study	HIE in Xinjin
Muinga et al [36], 2020	Kenya	Semistructured qualitative study	Digital health systems in Kenyan public hospitals
Ndlovu et al [25], 2016	Botswana	Landscape analysis	ICTs[i] in Botswana
Ntlhakana et al [29], 2022	South Africa	Semiqualitative cross-sectional study	South African mines' HCPs[j] data
Nutley et al [34], 2013	Kenya	Semistructured qualitative study	DHP[k] tool in Kenya
Pradhan et al [3], 2021	India	Digital health landscape analysis	Overview of AI in health care
Shrivastava and Srikanth [26], 2020	India	Case study	Federated digital health care ecosystem in India
Simbini et al [39], 2018	Africa	Perspective	African eHealth strategies and ICT
Troncoso [27], 2020	LMICs[l]	Perspective	AI and ML[m] in primary health care
Walcott-Bryant et al [32], 2021	Kenya	Primary qualitative case study	Digital health solution to support management of hypertension in Kenyan private health sector
Wang et al [30], 2019	China	Case study without stakeholder interviews	EMR data
Yaqoob et al [9], 2017	Pakistan	Feasibility analysis	EHR and NHIS[n] in Pakistan and other countries

[a]AI: artificial intelligence.
[b]HIE: health information exchange.
[c]DEC: Data Exchange Component.
[d]EHR: electronic health record.
[e]eIDSR: electronic Integrated Disease Surveillance.
[f]DHIS2: District Health Information Software.
[g]RNDS: National Health Data Network.
[h]EMR: electronic medical record.
[i]ICT: information and communications technology.
[j]HCP: hearing conservation program.
[k]DHP: District Health Profile.
[l]LMIC: low- and middle-income country.
[m]ML: machine learning.
[n]NHIS: National Healthcare Information System.

XSL•FO
RenderX

Factors Associated With Data Sharing

Figure 2 summarizes the factors associated with data sharing according to the framework.

Figure 2. Factors associated with sharing of health care data using artificial intelligence tools.



Technical Factors

We identified 7 technical factors associated with data sharing for AI in LMICs; 5 were related to infrastructure, and 2 were related to data management practices. Limited access to information and communications technology and poor infrastructure lead to poor internet connectivity, power shortages, and software crashes, causing difficulty in backing up and exchanging data [7,23,28,35,36]. Inappropriate and inconsistent data management software and software versions can lead to data fragmentation and format issues [2,29]. Health systems in LMICs often neither use consolidated data management systems (ie, a system that seamlessly brings together various subsystems into one for easier management and coordination [40]) nor use central repositories (ie, a singular storage location for data within an organization, providing visibility, collaboration, and consistency within data management [41,42]). HIEs are often standalone [8,9,35] with limited communication across different databases and various health systems and electronic medical record (EMR) systems implemented in different districts or provinces, leading to duplication [25,32]. This can lead to inconsistent use of data definitions and labels, reducing interoperability [2,8,33]. Investing in a multisourced data triangulation system, storing data in a single database and central repository, and using standardized electronic health records (EHRs) can promote data collection and reduce fragmentation [1,6,9,28-30,35].

Regarding data management, many health care systems in LMICs use handwritten records, often in native (ie, regional) languages [26,34], limiting shareability. Ensuring that the platforms and software used for data handling are ergonomic, user friendly, and tailored to local and regional contexts will increase their uptake and use among health care staff [6].

Motivational Factors

We identified 3 motivational factors. First, health care staff often lack the motivation to systematically collect data and adopt new standards for uniform data collection [8,33]. Education and awareness about the value of health data exchange can address staff resistance to digital data collection [24,27]. Second, managers and organizations might not perceive the value of collecting health care data or emphasize the importance of data quality to their staff [8,43]. Third, incentives for adopting and using AI can improve the digitization of medical data [24,27].

Economic Factors

We identified 3 economic factors associated with data sharing. First, the uneven distribution of financial, human, and technical

XSL•FO
RenderX

resources in LMICs [24,35] is a significant barrier for small clinics or hospitals to collect high-quality data [7,9]. Limited resources constrain the use of data audits for confidentiality, privacy, and standardized data field terminologies [25,31,33]. High equipment costs are a significant barrier for small clinics or hospitals to collect. Financial support enables the adoption of EMR systems and promotes the implementation of regional HIEs by addressing infrastructural barriers, lack of human resources, adequate training, data quality, and security challenges [7,24,35].

Political, Legal, and Policy Factors

We identified 4 political, legal, and policy factors. First, transparent regulations, defined standards for intersystem communication, anonymization of personal data, and security frameworks promote data sharing [3,9,23,27,30,39]. Second, national policy and legal facilitators for data sharing include a national strategy [24,33,35] and data-sharing guidelines to clarify national personal data protection acts [2,3,32]. However, even when national data policies are in place, ensuring regulatory compliance can be challenging [7,24,25]. Data auditing mechanisms can promote compliance [30,39]. Finally, clear institutional-level data-sharing guidelines promote information transfer and increase data-sharing motivation. Similarly, having confidentiality agreements helps uphold data accuracy and privacy protection requirements [25,30,31]. Aligning the description of data elements with current clinical practices and workflows will facilitate work for health care staff and enhance collaboration among entities involved in health data sharing [39].

Ethical Factors

We identified 3 ethical factors. The absence of explicit informed consent and patient confidentiality concerns are significant data-sharing barriers, while unauthorized access caused by bypassing security mechanisms can cause ethical concerns, further putting the patient's data in jeopardy and thus creating a reluctance to share data among patients and practitioners [3,9,26]. Frameworks focusing on ethics and privacy at the national and institutional level facilitate trust among data users and patients [27]. Creating secure channels for data sharing and state-of-the-art security standards will also be important to securing patient privacy [23,31]. Novel technologies such as decentralized consent management blockchain systems might be a potential solution, enabling patients to share consented data through data requests [32].

Social Factors

We identified 3 social factors associated with data sharing. First, the complexity of the types of stakeholders involved in data sharing limits collaboration and interoperability [2,3]. Second, a lack of public awareness regarding patient rights and data laws can also be a barrier that may hinder consent processes [26]. Third, national protocols, certification systems, and increased awareness and information about the benefits and risks associated with data sharing will help build trust between health care practitioners and patients [3,27,30].

Organizational and Managerial Factors

There were 4 organizational and management factors. Workforce limitations [12,13], limited provider digital literacy [24,25,32,35,36,43], and the absence of expert IT teams [6] reduce data capture and quality. Overburdened health care workers' conflicting priorities limit the time to enter high-quality data into digital systems [36,43]. A shortage of staff and leaders can lead to ill-defined responsibilities and ambiguity of data-related roles, further enhancing the ineffectiveness of organizational procedures [7,29,33]. Poor organizational clarity on data rights and permissions, central coordination, and terminology are barriers to data sharing [9]. This can be addressed using standard operating procedures to facilitate health data exchange across facilities [1,29]. Health personnel training can enable the staff to use EMR systems, manage data dictionaries, address data quality challenges, and address reluctance to use digital technology [7,24,31,33,35,43]. The complexity of stakeholders and multiple system contractors and organizations involved in digital health in developing nations is a significant barrier to interoperability [11].

Stakeholder Interviews

We mapped 50 relevant stakeholders for interviews, of which 15 confirmed participation (response rate: 15/50, 30%). Most interviewees were academic experts (6/15, 40%) and technology developers (5/15, 33%). The remaining were policy makers, regulators, and health care experts. Table 2 summarizes the major themes from the interviews. We identified multiple barriers and enablers relevant to the context of Thailand, which were classified under the 7 main domains. Most stakeholders reiterated the technical, economic, political, and social factors of data sharing identified during the systematic review.

Thematic synthesis of the interviews yielded multiple barriers and enablers relevant to Thailand's context, which were classified under 7 main categories of our framework.

XSL•FO
RenderX

Table 2. Summary of major themes identified through in-depth stakeholder interviews.

Framework	Subthemes identified by stakeholders
Technical factors	• Inconsistent data architecture, formats, and standards, eg, different data formats between different EHR_a platforms of the same organization • Misalignment in terminology and structure and data inaccuracy • Different data collection practices between private and public health care sectors • Manual data management and paper-based EHRs
Motivational factors	• Lack of staff time • General unawareness of the value of sharing data, creating awareness and empowerment • Limited incentives to sharing data beyond the organization; therefore, providing incentives such as tax rebates and funding to private sectors
Economic factors	• Lack of financial support and resources (mainly manpower and trained staff) • High equipment costs • Financial disparity between private and public sectors increases the gap in digital skills proficiency
Political, legal, and policy factors	• Complex or unclear data-sharing regulations and guidelines (eg, what data can be shared and what needs to be protected) between organizations or hospitals cause fear of sharing data wrongly • Too many laws and types of laws cause confusion, making data integration cumbersome • Low applicability of international standards • Having clear guidelines mandating standardized practicing on data entry, data protection, and data storage • Identifying a single government organization responsible for collaborating and strengthening all branches of data sharing
Ethical factors	• Confidentiality and consent issues • Frequent breaches of hospital information systems • Improving adherence to practices that mitigate data breaches, such as separating personal and professional use of computers, using 2-factor authentication or a password-protected log-in
Social factors	• Lack of public awareness of data laws and low digital health literacy • Limited stakeholder collaboration
Organizational and managerial factors	• Lack of clarity on what data can be shared • Issue of data ownership, ie, considering data as private and profitable assets • Lack of an organizational protocol, ie, standard operating procedures with necessary steps to approve data-sharing, data protection, and staff training guidelines

[a]EHR: electronic health record.

Factors Associated With Data Sharing: Thematic Analysis

Technical Factors

Stakeholders reported that current data architecture and inconsistent data standards (multiple or uneven systems) did not allow hospitals to share data throughout the health care system. There can be a lack of software standardization even within the same company. In addition, manual data management and paper-based EHRs are common. In lower-income regions or smaller clinics, there is a lack of equipment and infrastructure to collect data (and share data safely), leading to uneven and fragmented data. One participant mentioned that different data collection and health information systems in the clinical workflow are barriers to efficiently sharing data. Most participants also noted the lack of uniformity in gathering data across private and public health sectors:

> *In some clinics, their systems are not fully electronic yet (e.g., patient hand writes, then the document is scanned into a PDF form and is uploaded).* [Policy expert]

> *Previously, governments had no standards for sharing data, which led to each hospital having their own system, format, and structure.* [Health care expert]

> *There is neither a centralised system nor initiative to streamline all the healthcare data into a single platform like a National EMR system.* [Academic expert]

To overcome these technical barriers, stakeholders emphasized the need for a decentralized data infrastructure system to enable data sharing. This can be achieved by making routine data collection electronic, keeping in mind its user friendliness, which would not exacerbate staff workload:

> *All healthcare facilities need to use one single platform for data input in the future. This is important to ensure comparability of the data.* [Academic expert]

Economic Factors

Stakeholders reported a lack of funding and resources (mainly manpower and staff time) as a barrier to standardization, scaling, and data sharing. They cited high equipment and software costs as a reason for still using paper-based data collection. Funding

XSL•FO
RenderX

inequity caused the disparity in digital skills between low versus high-income facilities. Stakeholders mentioned that regular top official turnover limited consistent IT staff training. All these barriers led to a lack of technical capacity to digitize patients' records, especially in most rural hospitals:

> *There is a lack of equity (in funding)... In low-income health facilities, the manpower may not have the technical capabilities to update data and link the data to a central database.* [Policy expert]

> *...The public sector in Thailand should work on security, but it is beyond their budget to buy large tools such as firewalls.* [Academic expert]

> *... unclear who would bear the cost of operating a unifying health data platform in the future.* [Academic expert]

None of the stakeholders mentioned any solutions to the aforementioned economic barriers that would help facilitate data sharing.

Political Factors

Most reported political- and policy-level barriers revolved around the lack of data-sharing regulations and guidelines (eg, who can share what data and under what circumstances), complex and unclear policies, and the low applicability of international data standards. Participants mentioned that there were no clear regulations and national policies on data sharing between hospitals, which may cause the doctors to reject patient requests to share their information with another hospital. Declining requests to share data was also attributed to a fear caused by the lack of clear laws: no laws would protect and support hospitals if they wrongly released a patient's health data. Therefore, too many laws or types of laws were reported to cause confusion and make data integration even more difficult. Other reported barriers were the lack of cross-ministry collaboration and the absence of a central regulatory authority that modulates data-sharing practices among health care providers:

> *The Ministry of Health should work with other ministries to develop these (data sharing guidelines).* [Technology developer]

> *Current law (law that governs digital platforms) is not enough to cover digital health services; an extension of the law is required.* [Regulator]

Possible solutions to these political-level barriers ranged from government decisions to standardize data inputs and sharing to using digital health sandboxes. It was suggested that data sharing between data collectors should be mandated by law, that regulations should be implemented, that there should be an intermediary between all ministries to facilitate collaboration, and that there should be a national committee on AI in health care with a high priority for data sharing:

> *It is suggested that both the public and private sector agree on a set of standards so that they are willing to share the data collected nationally.* [Technology developer and policy expert]

> *Steer the development of regulatory sandboxes as testbeds for regulatory innovation ... it is also important to provide clear costs and benefits for start-ups to join the sandboxes right at the beginning.* [Technology developer]

Social Factors

Important social barriers included a lack of awareness of data laws, data sharing, and collaboration. Hospitals often feel reluctant to share data with private companies because of limited awareness of laws on data sharing (eg, Personal Data Protection Act [PDPA]). Perceived gray areas and limited regulatory clarity exacerbate unclarity on who can share what data under which circumstances. Another significant barrier included the absence of a central body that overlooks communication between different ministries to facilitate interministerial collaboration and data transfer:

> *The implementation of PDPA was launched in Thailand and people stopped sharing data altogether. This is clear evidence that the people have misunderstood the meaning of data sharing.* [Policy maker]

> *There is poor collaboration between the public and private sector; limited evidence and experience related to data sharing.* [Academic expert]

In response to these barriers, interviewees suggested collecting input from various stakeholders involved in data sharing. These should include individuals in the medical field, companies, ministries, and patient organizations. Identifying a single government organization responsible for collaborating and strengthening all branches of data sharing is an unmet need. Another critical consideration was enhancing the Thai population's awareness and knowledge of data sharing:

> *Various stakeholders will need to come together to work together in building a conducive digital ecosystem that facilitates responsible data sharing.* [Academic expert]

Motivational Factors

One participant reported a lack of staff time as a barrier. Other motivational barriers included people's (eg, patients, health care providers, and health staff) unawareness of the value of sharing data and limited incentives for organizations to share data:

> *Doctors often lack the motivation to share data as they dislike sharing data amongst themselves, as they do not like to share their logic and treatment process for a patient.* [Health care expert]

> *Public and private healthcare facilities are unlikely to pull resources and share their data across the public and private spheres.* [Academic expert]

Creating incentive structures for hospitals to share data was among the most widely suggested enablers. One participant suggested that the government can consider providing funding and tax incentives to the private sector if they share data for the public good. Regarding creating awareness about data sharing and data laws, interviewees suggested that executive-level players in organizations, entrepreneurs, and policy makers must

XSL•FO
RenderX

be empowered and educated. Patients should receive information on the benefits of sharing their data. Furthermore, data owners should have an understandable and unambiguous process to share their data securely:

> *Doctors currently do not have the mindset to share the data that they collect. Therefore, it is suggested that there should be incentives from sharing data, such that the one that collects the data (medical doctors) as well as the individual that analyses the data both benefits.* [Policy expert]

> *For instance, an application to use safe drugs would encourage hospitals and drug stores to provide information directly to patients. The inventive for these pharmacies would be patient engagement; there is a greater demand and therefore greater profit which acts as an incentive to the pharmacy. This would allow the information of the individual and the history of purchases to be recorded—whether in the form of websites or memberships—which would encourage customers to come back and buy from the same drug store.* [Technology developer and policy expert]

> *Increase the awareness of data owners; create simpler consent form that is clear, concise and explicit.* [Technology developer]

Ethical Factors

Ethical barriers included a lack of awareness of data laws (eg, PDPA) and data privacy and security issues:

> *Patients are afraid of sharing their data because of the increased rate of cyberattacks on hospitals.* [Policy expert]

Data security discipline among users can mitigate data breaches and help overcome ethical barriers. Possible solutions include separating personal and professional computer use, using 2-factor authentication, and using password-protected log-in and log out:

> *To enable responsible data sharing in digital health, cybersecurity of the health information systems will have to be strengthened. The country will also need to come up with a clear structure to compensate victims whose data are compromised.* [Academic expert]

Organizational and Managerial Factors

Interviewees reported that there are regulations whereby management teams are unable to ensure whether data are secure and safe enough to share, which acts as a barrier at the organizational level (ie, there are certain workflows and policies that fall within a gray area, as mentioned previously). Bureaucracy was also mentioned as a barrier if considering the top-down approach, as changes in high-level leadership in the ministries and organizations caused delayed processing and changed regulations related to data sharing. Data ownership was reported to be a disputed area where health care providers and private hospitals were unwilling to share the collected data, considering it an asset for business profitability. The managerial mindset of organizations regarding data ownership, that is, the notion that if they have generated the data, the data belongs to them, causes unwillingness to share data with other organizations:

> *Mindset of care providers that just because the patients are willing to share their data, the data is now owned by the hospital. However, this is untrue as the hospitals are simply data collectors, and fundamentally, the Personal Data Protection Act (PDPA) states that the personal health information belongs to the patient.* [Technology developer and policy expert]

> *Different organisations (and even departments in the same organisation) do not want to share data as they believe that it is valuable and profitable.* [Academic expert]

Only a few interviewees provided solutions to some organizational barriers. One suggested that having an organizational protocol with the necessary steps to approve data sharing would enhance structure and increase trust:

> *Hospital administrators need to be trained on cybersecurity safeguards, data security, and be given access to resources to improve their digital infrastructures to enable digital transformation (i.e., transforming paper-based data to machine readable data).* [Academic expert]

Discussion

Principal Findings

Our systematic review identified 29 unique factors across the 7 domains associated with data sharing for AI in LMICs. These factors can significantly influence the ability of health care organizations to share data with AI developers and limit the implementation of AI solutions in health systems. Our case study in Thailand demonstrated the relevance of the domains, serving as an example of how the factors identified in the systematic review can be contextualized locally. Strategies to address barriers to data sharing in countries with nascent AI industries can consider the factors identified in this systematic review for policy development.

Two previous studies identified barriers to HIE adoption in LMICs [10] and sharing research and scientific data [11]. Our review is distinct from this previous work by focusing on factors associated with data sharing for AI in health care. Data used to develop AI require more granularity (eg, in the form of text or images). This influences the factors associated with data sharing. For example, data used to deploy AI often involve more sensitive information, such as biometric data, which can be used to identify individuals [44]. This means that data used in AI are often associated with more significant privacy concerns than administrative data, such as birth and death registries. Our study is the first to address specific factors related to data sharing for AI in LMICs, which is of critical importance given AI's potential to enable developing health systems to leapfrog global health inequalities [45].

XSL•FO
RenderX

The most relevant factors we identified were technical and infrastructural, policy based, or motivational. Many of the factors we identified in our review were also seen in 2 previous studies focusing on HIE adoption and barriers to sharing research and scientific data [10,11]. For example, van Panhuis et al [13] found that similar technical factors, such as lack of data standards, guidelines, and policies, were important barriers to sharing data in public health. In our study, infrastructural factors, such as limited access to information and communications technology, unreliable internet connectivity, frequent power shortages, equipment deficiencies, and software instability, were important factors hindering data storage and management. The unavailability of appropriate software and the persistence of handwritten records (often in regional languages) were particular challenges to data integration and sharing efforts. These findings were consistent with previous literature [46-49] and are also seen when data are shared for reasons other than AI, such as global health emergencies and disease outbreaks [50,51]. We found that limited data digitization is an important upstream barrier to downstream data sharing in LMICs. Furthermore, barriers at the political and policy level were significant limiting factors to data sharing for AI in health care. Previous studies have emphasized the need for data-sharing governance mechanisms to protect vulnerable populations in the global south [52-54]. Our results confirm earlier studies and suggest that developing clear policies and closing regulatory gaps can promote data sharing [10,46,48,55]. The motivational barriers we identified, including poor staff and management motivation, confirm previous results [56]. Public and staff education, a clear data-sharing incentive structure, and user-friendly software could improve staff motivation [48,57].

Many barriers identified in the systematic review were relevant to the Thai context. For example, limited staff time to collect data, a limited understanding of data's value, and low health literacy were important barriers in Thailand and the systematic review. Aljunid et al [47] previously reported that Thailand's fragmented financing schemes and service provision limit stakeholder cooperation in data sharing. Our findings concur that this health system fragmentation may hinder clinical and management staff from sharing data. Furthermore, stakeholders pointed to a lack of software standardization within the same organization and continued reliance on manual data management and paper-based EHRs. These findings differ from a previous qualitative study [58], where Thai stakeholders did not see technical challenges as a primary barrier to health care data sharing and instead placed greater weight on ethical concerns, such as informed consent and data control. However, Thai-specific literature [47,54] supports our findings that fragmented databases and software unavailability limit local health data sharing. Finally, we found that unclear data-sharing regulations and guidelines in Thailand, that is, complex and unclear policies and low applicability of international data standards—prevent health care stakeholders from sharing data. These findings are supported by earlier studies, emphasizing health data sharing in Thailand echoes the importance of establishing clear data management and sharing policies [54,55].

During interviews, stakeholders suggested several improvements to overcome barriers to sharing data. First, building a conducive digital ecosystem and having shared data input platforms for health facilities can ensure data uniformity and compatibility. Currently, similar initiatives are being undertaken by the Thai National Health Information Standards Development Centre under the Health System Research Institute of the Ministry of Public Health, Thailand. This organization develops and promotes national standards for health information systems, including data elements, codes, and terminologies, to ensure interoperability across health care organizations [59]. Singapore's National Electronic Health Record initiative is also an example of creating a unified health information system by standardizing data formats and terminologies for seamless data exchange across health care providers [60]. Second, developing predefined templates, easy-to-understand consent forms, and standardized guidelines for data sharing, intellectual property rights, and compensation for data breach victims can help Thailand overcome social-, ethical-, political-, and policy-level barriers at various levels. These measures will increase data owners' and health care professionals' awareness of data sharing and prevent misunderstandings of data sharing guidelines such as the PDPA. In addition, these steps will help developers create and validate transparent and mutually trustworthy data-sharing models across health facilities. Singapore's Advisory Guidelines for Key Concepts on PDPA can serve as a framework for Thailand, offering detailed guidance and examples to understand and interpret key terms and obligations in the PDPA [61]. The document provides illustrative examples of important PDPA terms, including what qualifies as "personal data," the concept of "controlling and owning one's personal data," and the "roles and obligations of data intermediaries." Third, training health care professionals such as hospital administrators, doctors, and care providers can improve the effectiveness of the aforementioned measures. Access to improved digital infrastructures and adequate training on cybersecurity safeguards will enable efficient and standardized data sharing.

Limitations

Although the study focused on a specific context, we have provided detailed descriptions of the settings and participants to allow readers to judge the applicability of the findings to other contexts. Studies included in our systematic review output were mostly from Africa and Asia. Therefore, results might not be generalizable to other regions. LMICs can use this study as a framework and leverage their individual health systems' barriers and strengths to devise local solutions for enhanced data sharing. While the classification of barriers and enablers across domains was done objectively and cross-checked by multiple authors, some factors identified were relevant to various framework categories. This is inevitable due to the complexity of health information systems, the various stakeholders involved, and their interrelationships. Furthermore, the response rate for our stakeholder interviews was low, reducing the sample's representativeness.

Conclusions

AI has the potential for LMICs to leapfrog health inequalities and deficiencies in health systems. Our review identified

XSL•FO
RenderX

significant barriers to sharing data for AI according to a comprehensive framework. In a localized case study, most barriers from the systematic review were relevant to the Thai context. In Thailand, data architecture, inconsistent data standards, complex and unclear policies, uneven distribution of financial resources across institutions, and confidentiality breaches were reported as important barriers to data sharing for AI. Together, our results provide insight into the challenges that LMICs' health systems face regarding data sharing for AI.

Acknowledgments

The authors want to thank the key stakeholder interviewees for their time and wisdom. They also want to thank their partners at the Office of National Higher Education Science Research and Innovation Policy Council for their careful review of the findings and helpful discussions. Finally, the authors thank the Association of Pacific Rim Universities and the United Nations Economic and Social Commission for Asia and the Pacific for their support of the project and Google for the generous funding.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AK, CB, and JT conceived the design of the study. AK and CB conducted the data extraction, data collection, data analysis, interpretation, and initial drafting of the manuscript. JT contributed to the verification of results. All authors contributed to the critical review and final approval of this manuscript.

Conflicts of Interest

JT is supported by the National University of Singapore Start-up grant, the tier 1 grant from the ministry of education, and the CS-IRG New Investigator Grant from the National Medical Research Council; has received consulting or speaker fees from Daiichi-Sankyo, Boehringer Ingelheim, Roche diagnostics, and Us2.ai; and owns patent US-10702247-B2 unrelated to this work. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Search strategy for different databases.
[PDF File (Adobe PDF File), 88 KB - jmir_v27i1e58338_app1.pdf]

Multimedia Appendix 2
Critical appraisal of qualitative studies using Critical Appraisal Skills Programme checklist.
[DOCX File , 23 KB - jmir_v27i1e58338_app2.docx]

Multimedia Appendix 3
Interview guide—barriers and enablers to data sharing in digital health in Thailand.
[PDF File (Adobe PDF File), 95 KB - jmir_v27i1e58338_app3.pdf]

Multimedia Appendix 4
PRISMA checklist.
[PDF File (Adobe PDF File), 61 KB - jmir_v27i1e58338_app4.pdf]

References

1. Chukwu E, Garg L, Foday E, Konomanyi A, Wright R, Smart F. Digital health solutions and state of interoperability: landscape analysis of Sierra Leone. JMIR Form Res 2022 Jun 10;6(6):e29930 [FREE Full text] [doi: 10.2196/29930] [Medline: 35687406]
2. Chali F, Yonah ZO, Kalegele K. Data exchange architecture for the development of mobile applications that support eHealth systems interoperability: a case of Tanzania. Int J Adv Comput Res 2018 Jan 05;8(34):1-10. [doi: 10.19101/ijacr.2017.733025]
3. Pradhan K, John P, Sandhu N. Use of artificial intelligence in healthcare delivery in India. J Hosp Manag Health Policy 2021 Sep;5:28. [doi: 10.21037/jhmhp-20-126]
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
5. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

XSL•FO
RenderX

6. Fakhkhari H, Bounabat B, Bennani M, Bekkali R. Moroccan patient-centered hospital information system: global architecture. In: Proceedings of the ArabWIC 6th Annual International Conference Research Track. 2019 Presented at: ArabWIC '19; March 7-9, 2019; Rabat, Morocco p. 1-6 URL: https://dl.acm.org/doi/10.1145/3333165.3333175 [doi: 10.1145/3333165.3333175]
7. Botha M, Botha A, Herselman M. On the prioritization of data quality challenges in e-health systems in South Africa. In: Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists. 2015 Presented at: SAICSIT '15; September 28-30, 2015; Stellenbosch, South Africa p. 1-10 URL: https://dl.acm.org/doi/10.1145/2815782.2815791 [doi: 10.1145/2815782.2815791]
8. Guidry AF, Walson JL, Abernethy NF. Linking information systems for HIV care and research in Kenya. In: Proceedings of the 1st ACM International Health Informatics Symposium. 2010 Presented at: IHI '10; November 11-12, 2010; Arlington, VA p. 531-535. [doi: 10.1145/1882992.1883078]
9. Yaqoob T, Mir F, Abbas H, Shahid WB, Shafqat N, Amjad MF. Feasibility analysis for deploying national healthcare information system (NHIS) for Pakistan. In: Proceedings of the 19th International Conference on e-Health Networking, Applications and Services. 2017 Presented at: Healthcom '17; October 13-15, 2017; Dalian, China p. 1-6. [doi: 10.1109/healthcom.2017.8210836]
10. Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review. Health Policy Plan 2016 Nov;31(9):1310-1325. [doi: 10.1093/heapol/czw056] [Medline: 27185528]
11. Figueiredo AS. Data sharing: convert challenges into opportunities. Front Public Health 2017;5:327 [FREE Full text] [doi: 10.3389/fpubh.2017.00327] [Medline: 29270401]
12. Bentzen HB, Castro R, Fears R, Griffin G, Ter Meulen V, Ursin G. Remove obstacles to sharing health data with researchers outside of the European Union. Nat Med 2021 Aug;27(8):1329-1333 [FREE Full text] [doi: 10.1038/s41591-021-01460-0] [Medline: 34345050]
13. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014 Nov 05;14:1144 [FREE Full text] [doi: 10.1186/1471-2458-14-1144] [Medline: 25377061]
14. Eden KB, Totten AM, Kassakian SZ, Gorman PN, McDonagh MS, Devine B, et al. Barriers and facilitators to exchanging health information: a systematic review. Int J Med Inform 2016 Apr;88:44-51 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.01.004] [Medline: 26878761]
15. Trusted data sharing framework. Infocomm Media Development Authority. URL: https://www.imda.gov.sg/how-we-can-help/data-innovation/trusted-data-sharing-framework [accessed 2024-04-29]
16. Thailand national AI strategy and action plan. AI Thailand. 2022. URL: https://ai.in.th/wp-content/uploads/2022/12/2022-NAIS-Presentation-eng.pdf [accessed 2024-04-29]
17. Government AI readiness index. Oxford Insights. 2021. URL: https://oxfordinsights.com/wp-content/uploads/2023/11/Government_AI_Readiness_21.pdf [accessed 2024-04-29]
18. Kaewkungwal J, Adams P, Sattabongkot J, Lie RK, Wendler D. Issues and challenges associated with data-sharing in LMICs: perspectives of researchers in Thailand. Am J Trop Med Hyg 2020 Jul;103(1):528-536 [FREE Full text] [doi: 10.4269/ajtmh.19-0651] [Medline: 32394875]
19. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov Inf Q 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
20. Low-and middle-income countries classification. The World Bank Group. 2019. URL: https://data.worldbank.org/income-level/middle-income [accessed 2024-04-29]
21. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021 Sep;3(9):e599-e611 [FREE Full text] [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
22. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2021.
23. Bhattacharya S, Pandey M. Issues and challenges in incorporating the internet of things with the healthcare sector. In: Reddy KA, Devi BR, George B, Raju KS, editors. Data Engineering and Communication Technology. Cham, Switzerland: Springer; 2021:639-651.
24. Lei J, Wen D, Zhang X, Li J, Lan H, Meng Q, et al. Enabling health reform through regional health information exchange: a model study from China. J Healthc Eng 2017;2017(1):22 [FREE Full text] [doi: 10.1155/2017/1053403] [Medline: 29068625]
25. Ndlovu K, Mogotlhwane T, Scott RE, Mars M. E-health interoperability landscape: Botswana. In: Proceedings of the 4th IASTED International Conference on Health Informatics. 2016 Presented at: AfricaHI '16; September 2-6, 2016; Gaborone, Botswana p. 23-55 URL: https://www.actapress.com/Abstract.aspx?paperId=456304 [doi: 10.2316/p.2016.837-010]
26. Shrivastava S, Srikanth TK. e-Governance for healthcare service delivery in India: challenges and opportunities in security and privacy. In: Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance. 2020 Presented at: TPEG '20; September 23-25, 2020; Athens, Greece p. 180-183. [doi: 10.1145/3428502.3428527]

XSL•FO
RenderX

27. Troncoso EL. The greatest challenge to using AI/ML for primary health care: mindset or datasets? Front Artif Intell 2020 Aug 21;3:53 [FREE Full text] [doi: 10.3389/frai.2020.00053] [Medline: 33733170]
28. de Fatima Marin H, de Souza Zinader JP, da Silva Pires FJ, de Barros JV. The Brazilian digital health system: building the digital transformation to engage country citizens. In: Hovenga E, Grain H, editors. Roadmap to Successful Digital Health Ecosystems: A Global Perspective. New York, NY: Academic Press; 2022:e2022-e2087.
29. Ntlhakana L, Nelson G, Khoza-Shangase K, Dorkin E. Occupational hearing loss for platinum miners in South Africa: a case study of data sharing practices and ethical challenges in the mining industry. Int J Environ Res Public Health 2021 Dec 21;19(1):1 [FREE Full text] [doi: 10.3390/ijerph19010001] [Medline: 35010261]
30. Wang Z. Data integration of electronic medical record under administrative decentralization of medical insurance and healthcare in China: a case study. Isr J Health Policy Res 2019 Apr 01;8(1):24 [FREE Full text] [doi: 10.1186/s13584-019-0293-9] [Medline: 30929644]
31. deRiel E, Puttkammer N, Hyppolite N, Diallo J, Wagner S, Honoré JG, et al. Success factors for implementing and sustaining a mature electronic medical record in a low-resource setting: a case study of iSanté in Haiti. Health Policy Plan 2018 Mar 01;33(2):237-246. [doi: 10.1093/heapol/czx171] [Medline: 29253138]
32. Walcott-Bryant A, Ogallo W, Remy SL, Tryon K, Shena W, Bosker-Kibacha M. Addressing care continuity and quality challenges in the management of hypertension: case study of the private health care sector in Kenya. J Med Internet Res 2021 Feb 17;23(2):e18899 [FREE Full text] [doi: 10.2196/18899] [Medline: 33595446]
33. Keny A, Wanyee S, Kwaro D, Mulwa E, Were MC. Developing a national-level concept dictionary for EHR implementations in Kenya. Stud Health Technol Inform 2015;216:780-784. [Medline: 26262158]
34. Nutley T, McNabb S, Salentine S. Impact of a decision-support tool on decision making at the district level in Kenya. Health Res Policy Syst 2013 Sep 08;11(1):34 [FREE Full text] [doi: 10.1186/1478-4505-11-34] [Medline: 24011028]
35. Akhlaq A, McKinstry B, Sheikh A. Stakeholders perspectives and deployment strategies of health information exchange illustrated through an in-depth case study of Pakistan. Inform Health Soc Care 2020 Mar 28;45(2):130-150. [doi: 10.1080/17538157.2019.1582053] [Medline: 30922142]
36. Muinga N, Magare S, Monda J, English M, Fraser H, Powell J, et al. Digital health systems in Kenyan public hospitals: a mixed-methods survey. BMC Med Inform Decis Mak 2020 Jan 06;20(1):2 [FREE Full text] [doi: 10.1186/s12911-019-1005-7] [Medline: 31906932]
37. CASP qualitative studies checklist. CASP. URL: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist/ [accessed 2024-04-29]
38. Jane RL, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. Analyzing Qualitative Data. New York, NY: Routledge; 2018:22.
39. Simbini T, Foster R, Nesara P, Cossio CH. Monitoring diseases across borders: African regional integrative information systems. In: Proceedings of the 13th World Congress on Medical and Health Informatics of International-Medical-Informatics-Association. 2010 Presented at: Medinfo '10; September 12-15, 2010; Cape Town, South Africa p. 12-15.
40. What is an IMS (integrated management system) and how it can help your business? Compliancehelp. URL: https://www.quality-assurance.com/blog/what-is-an-ims-integrated-management-system-how-it-can-help-your-business.html [accessed 2024-12-20]
41. What is a central repository? Hewlett Packard Enterprise Development. URL: https://www.hpe.com/us/en/what-is/central-repository.html [accessed 2024-12-20]
42. Why you need a central repository for your policies. PowerDMS. URL: https://www.powerdms.com/policy-learning-center/why-you-need-a-central-repository-for-your-policies [accessed 2024-12-20]
43. Nutley T, McNabb S, Salentine S. Impact of a decision-support tool on decision making at the district level in Kenya. Health Res Policy Syst 2013 Sep 08;11(1):34 [FREE Full text] [doi: 10.1186/1478-4505-11-34] [Medline: 24011028]
44. Frank FN, Chandra N, Juan T, Eli M. Data-sharing made easy. Deloitte insights. 2018. URL: https://www2.deloitte.com/us/en/insights/focus/tech-trends/2022/data-sharing-technologies.html [accessed 2024-04-29]
45. Shaffer J, Alenichev A, Faure MC. The Gates Foundation's new AI initiative: attempting to leapfrog global health inequalities? BMJ Glob Health 2023 Nov;8(11):23 [FREE Full text] [doi: 10.1136/bmjgh-2023-013874] [Medline: 37923320]
46. Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low- and middle-income countries. Bull World Health Organ 2015 Mar 01;93(3):203-208 [FREE Full text] [doi: 10.2471/BLT.14.139022] [Medline: 25767300]
47. Aljunid SM, Srithamrongsawat S, Chen W, Bae SJ, Pwu RF, Ikeda S, et al. Health-care data collecting, sharing, and using in Thailand, China mainland, South Korea, Taiwan, Japan, and Malaysia. Value Health 2012;15(1 Suppl):S132-S138 [FREE Full text] [doi: 10.1016/j.jval.2011.11.004] [Medline: 22265060]
48. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
49. Sahay S. Big data and public health: challenges and opportunities for low and middle income countries. Commun Assoc Inf Syst 2016;39:419-438. [doi: 10.17705/1cais.03920]

XSL•FO
RenderX

50. Pisani E, Ghataure A, Merson L. Data sharing in public health emergencies: a study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats. Wellcome Trust. 2018 Jan. URL: https://tinyurl.com/3hkcn7wv [accessed 2025-01-10]
51. van Reisen M, Oladipo F, Stokmans M, Mpezamihgo M, Folorunso S, Schultes E, et al. Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. Adv Genet (Hoboken) 2021 Jun 11;2(2):e10050 [FREE Full text] [doi: 10.1002/ggn2.10050] [Medline: 34514430]
52. Evertsz N, Bull S, Pratt B. What constitutes equitable data sharing in global health research? A scoping review of the literature on low-income and middle-income country stakeholders' perspectives. BMJ Glob Health 2023 Mar 28;8(3):e010157 [FREE Full text] [doi: 10.1136/bmjgh-2022-010157] [Medline: 36977523]
53. Tiffin N, George A, LeFevre AE. How to use relevant data for maximal benefit with minimal risk: digital health data governance to protect vulnerable populations in low-income and middle-income countries. BMJ Glob Health 2019 Apr 11;4(2):e001395 [FREE Full text] [doi: 10.1136/bmjgh-2019-001395] [Medline: 31139457]
54. Cheah PY, Tangseefa D, Somsaman A, Chunsuttiwat T, Nosten F, Day NP, et al. Perceived benefits, harms, and views about how to share data responsibly: a qualitative study of experiences with and attitudes toward data sharing among research staff and community representatives in Thailand. J Empir Res Hum Res Ethics 2015 Jul;10(3):278-289 [FREE Full text] [doi: 10.1177/1556264615592388] [Medline: 26297749]
55. Waithira N, Mutinda B, Cheah PY. Data management and sharing policy: the first step towards promoting data sharing. BMC Med 2019 Apr 17;17(1):80 [FREE Full text] [doi: 10.1186/s12916-019-1315-8] [Medline: 30992010]
56. Immunization data: evidence for action. A realist review of what works to improve data use for immunization. PATH PAHO. URL: https://tinyurl.com/526cm4nm [accessed 2024-04-29]
57. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. Bull World Health Organ 2010 Jun;88(6):462-466 [FREE Full text] [doi: 10.2471/BLT.09.074393] [Medline: 20539861]
58. Kaewkungwal J, Adams P, Sattabongkot J, Lie RK, Wendler D. Issues and challenges associated with data-sharing in LMICs: perspectives of researchers in Thailand. Am J Trop Med Hyg 2020 Jul;103(1):528-536 [FREE Full text] [doi: 10.4269/ajtmh.19-0651] [Medline: 32394875]
59. Thai health information standards development center (THIS). Government of Thailand. URL: https://www.this.or.th/en/this/ [accessed 2024-04-29]
60. National electronic health record Singapore. Synapxe. URL: https://www.synapxe.sg/healthtech/national-programmes/national-electronic-health-record-nehr/ [accessed 2024-04-29]
61. Advisory guidelines on key concepts in the personal data protection act. Personal Data Protection Commission. URL: https://tinyurl.com/4xfthjw5 [accessed 2024-04-29]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
EMR: electronic medical record
HIE: health information exchange
LMICs: low- and middle-income countries
ML: machine learning
PDPA: Personal Data Protection Act
SPICE: settings, perspective, intervention, comparison, and evaluation

Edited by A Schwartz; submitted 13.03.24; peer-reviewed by B Thies, D Mmereki, FA Rathore; comments to author 14.05.24; revised version received 06.08.24; accepted 28.11.24; published 04.02.25.

Please cite as:
Kaushik A, Barcellona C, Mandyam NK, Tan SY, Tromp J
Challenges and Opportunities for Data Sharing Related to Artificial Intelligence Tools in Health Care in Low- and Middle-Income Countries: Systematic Review and Case Study From Thailand
J Med Internet Res 2025;27:e58338
URL: https://www.jmir.org/2025/1/e58338
doi:10.2196/58338
PMID:

©Aprajita Kaushik, Capucine Barcellona, Nikita Kanumoory Mandyam, Si Ying Tan, Jasper Tromp. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.02.2025. This is an open-access article distributed under the

XSL•FO
RenderX

terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Health Care Professionals and Data Scientists' Perspectives on a Machine Learning System to Anticipate and Manage the Risk of Decompensation From Patients With Heart Failure: Qualitative Interview Study

Joana Seringa[1], MSc; Anna Hirata[2], BA; Ana Rita Pedro[1], PhD; Rui Santana[1], PhD; Teresa Magalhães[1], PhD

[1]NOVA National School of Public Health, Public Health Research Centre, Comprehensive Health Research Center, NOVA University Lisbon, Lisbon, Portugal

[2]NOVA National School of Public Health, NOVA University Lisbon, Lisbon, Portugal

Corresponding Author:
Joana Seringa, MSc
NOVA National School of Public Health, Public Health Research Centre, Comprehensive Health Research Center, NOVA University Lisbon
Avenida Padre Cruz
Lisbon, 1600-407
Portugal
Phone: 351 910417628
Email: jm.seringa@ensp.unl.pt

Abstract

Background: Heart failure (HF) is a significant global health problem, affecting approximately 64.34 million people worldwide. The worsening of HF, also known as HF decompensation, is a major factor behind hospitalizations, contributing to substantial health care costs related to this condition.

Objective: This study aimed to explore the perspectives of health care professionals and data scientists regarding the relevance, challenges, and potential benefits of using machine learning (ML) models to predict decompensation from patients with HF.

Methods: A total of 13 individual, semistructured, qualitative interviews were conducted in Portugal between October 31, 2022, and June 23, 2023. Participants represented different health care specialties and were selected from different contexts and regions of the country to ensure a comprehensive understanding of the topic. Data saturation was determined as the point at which no new themes emerged from participants' perspectives, ensuring a sufficient sample size for analysis. The interviews were audio recorded, transcribed, and analyzed using MAXQDA (VERBI Software GmbH) through a reflexive thematic analysis. Two researchers (JS and AH) coded the interviews to ensure the consistency of the codes. Ethical approval was granted by the NOVA National School of Public Health ethics committee (CEENSP 14/2022), and informed consent was obtained from all participants.

Results: The participants recognized the potential benefits of ML models for early detection, risk stratification, and personalized care of patients with HF. The importance of selecting appropriate variables for model development, such as rapid weight gain and symptoms, was emphasized. The use of wearables for recording vital signs was considered necessary, although challenges related to adoption among older patients were identified. Risk stratification emerged as a crucial aspect, with the model needing to identify patients at high-, medium-, and low-risk levels. Participants emphasized the need for a response model involving health care professionals to validate ML-generated alerts and determine appropriate interventions.

Conclusions: The study's findings highlight ML models' potential benefits and challenges for predicting HF decompensation. The relevance of ML models for improving patient outcomes, reducing health care costs, and promoting patient engagement in disease management is highlighted. Adequate variable selection, risk stratification, and response models were identified as essential components for the effective implementation of ML models in health care. In addition, the study identified technical, regulatory and ethical, and adoption and acceptance challenges that need to be overcome for the successful integration of ML models into clinical workflows. Interpretation of the findings suggests that future research should focus on more extensive and diverse samples, incorporate the patient perspective, and explore the impact of ML models on patient outcomes and personalized care in HF management. Incorporation of this study's findings into practice is expected to contribute to developing and implementing ML-based predictive models that positively impact HF management.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e54990) doi:10.2196/54990

KEYWORDS

heart failure; machine learning system; decompensation; qualitative research; cardiovascular diseases; heart failure management; interview

Introduction

Heart failure (HF) is a complex clinical syndrome characterized by frequent and rapid decompensation; even long-term, stable patients with HF might rapidly deteriorate in days or hours [1]. HF decompensation is a major cause of hospitalization in patients with HF, negatively influencing prognosis [1], representing the highest share of health care costs for this disease [2].

Several factors can precipitate HF decompensation, including respiratory infections, noncompliance with dietary recommendations or pharmacological treatment, and atrial fibrillation [2]. Early detection of HF decompensation is crucial to allow timely intervention and to prevent hospitalizations and increased morbidity [3].

Studies suggest that telemonitoring systems and predictive models for clinical support and patient empowerment may improve HF management [4]. Based on telemonitored data, artificial intelligence (AI) techniques have been increasingly used to predict and analyze HF decompensation events. AI and machine learning (ML) algorithms have the power to assimilate and integrate multidimensional, multimodal data and create accurate prediction models. The application of AI/ML algorithms could potentially improve workflow and outcomes for patients with HF due to the ever-growing data, particularly time series data gathered via remote monitoring [5]. Furthermore, there have been other health conditions in which AI models have been successfully developed for early detection and screening. A study published in the scientific journal *Nature* showcased a model that surpassed radiologists' performance in interpreting mammograms [6], while another study introduced an ML model that reliably predicts the onset of diabetes [7].

However, there is a limited understanding of health care professionals' perspectives on using such systems and the specific considerations necessary for designing an ML model to be integrated into clinical practice. To bridge this knowledge gap, we conducted a semistructured interview study involving health care professionals and data scientists with profound expertise in health care. The primary objective of this study was to explore their viewpoints concerning an ML system tailored for the early detection and management of HF decompensation. Additionally, this study sought to provide the experts' perceptions on developing crucial features in the ML system, identify potential implementation challenges, and establish design principles to enhance overall system efficiency and effectiveness.

Methods

Study Design

A qualitative research design was used, using individual semistructured interviews to explore the perspectives of health care professionals and data scientists regarding the relevance, challenges, and potential benefits of using ML models to predict decompensation in patients with HF.

Eligibility Criteria

Health care professionals working with patients with HF and data scientists with expertise in the health care field were considered eligible to participate in this study.

Recruitment Strategies

Participants were selected through intentional sampling to ensure the representation of diverse perspectives and expertise in HF and health care data management.

Potential participants were identified based on their expertise in the field. With prior authorization from the experts, key contacts from the researchers' network shared their email addresses with the primary researcher, facilitating direct communication.

The primary researcher emailed potential participants and invited them to a semistructured interview. This email included attachments to the interview guide and a consent form. The participants were required to provide written consent to participate in this study by signing the consent form.

Participants

A total of 13 participants, representing a diverse range of health care professionals and data science experts, were interviewed for this study. The participants recruited were primary care physicians (n=1), cardiologists (n=3), internal medicine physicians (n=3), nurses (n=3), and data scientists (n=3). The composition of this poll of experts aimed to capture the perspective of the health care professionals who usually provide care to patients with HF and the technical perspective of the data scientists. Geographically, participants were in different areas, with 2 participants based in the north of Portugal, 8 participants in the Lisbon area and its surroundings, 2 participants in the south, and 1 participant working outside of Portugal. The diversity of locations is especially relevant to guarantee the results are not biased by the local context.

The sample size of 13 participants was determined based on data saturation, where no new themes emerged from participants' perspectives, aligning with a recent systematic review's findings indicating that saturation can be achieved within a narrow range of interviews (9-17) in qualitative research [8].

XSL•FO
RenderX

Data Collection

The primary researcher conducted face-to-face and online semistructured interviews in Portuguese, using an interview guide (Multimedia Appendix 1) developed by the authors after the participants provided written informed consent to participate. All interviews were conducted in private rooms by a single researcher and were audio recorded to ensure accurate data capture.

The interview guide explored the participants' perspectives on an ML model for predicting HF decompensation. Open-ended questions were included in the interview guide, allowing participants to express their thoughts and opinions freely.

Before the formal data collection, an initial interview was conducted with an internal physician to test and adjust the interview guide, ensuring its relevance and appropriateness; this first interview was not included in the subsequent analysis.

Data Analysis

Each interview lasted, on average, 31 (range 15-48) minutes. The interviews were audio recorded, transcribed verbatim, and assessed for accuracy.

The participant's data were analyzed using interactive and inductive processes following a thematic analysis approach as proposed by Braun and Clarke [9], since the purpose of the research was explanatory and intended to generate new knowledge. The data analysis was conducted in the following phases: (1) familiarizing with the dataset, (2) generating initial codes, (3) searching for themes, (4) reviewing the themes, (5) defining and naming the themes, and (6) producing the report [9].

In the first phase, two authors engaged with the dataset by reviewing the interviews multiple times to understand the content's depth, identifying key expressions, and formulating and validating meanings through discussions. In the second phase, the MAXQDA (VERBI Software GmbH) was used to generate initial codes, allowing for simplified data analysis and management. Subsequently, codes were categorized into overarching themes.

During the third phase, the two authors (JS and AH) systematically identified and organized themes into clusters and categories. In the fourth phase, a third author (ARP) with expertise in qualitative research reviewed the findings to enhance the credibility of the analysis. The author provided feedback on the analytical process and the developing themes, facilitating their refinement.

In the fifth phase, the authors refined and defined the themes by re-examining the dataset and developing a comprehensive description of the themes. In the sixth phase, excerpts of the interview, derived from coding, were translated from Portuguese to English and further edited for clarity, resulting in the report of the analyzed dataset.

Researchers' Reflexivity

In reflexive thematic analysis, the researchers incorporate their knowledge and experiences into their analytical practice [10]. It is important to highlight that all the authors, as health care researchers, have engaged with patients diagnosed with chronic diseases, including HF, and are particularly focused on understanding how digital health can contribute to better management of these conditions. Therefore, our values and assumptions might have influenced the analysis produced. However, efforts were made to minimize this influence through rigorous data analysis and researchers' discussions.

Ethical Considerations

The research was conducted per the Helsinki Declaration (World Medical Association, 2018), and written informed consent was obtained from every participant. All data collected during the interviews were anonymized to ensure participant confidentiality. Only the primary researcher has access to the original recordings, which are securely stored and protected. To ensure anonymity, all participants were assigned unique codes, and personal identifiers were removed during data transcription to safeguard the privacy of participants. No participant received any type of compensation for taking part in this study. The study was approved by the NOVA National School of Public Health research ethics committee (CEENSP no 14/2022).

Results

Overview

Saturation was achieved with the seventh interview, and 6 participants were then interviewed to confirm thematic saturation. Four themes or categories emerged from the analysis: "Machine learning model relevance," "Variables collection and system features," "Risk levels and their management," and "Challenges and potential solutions for implementation."

Thematic Content

Subgroups and themes or categories identified in the analysis of health care professionals and data scientists' perspectives of an ML model to predict the risk of decompensation from patients diagnosed with HF are illustrated in Figure 1. Excerpts of the interview presented in the results were extracted from the coding analysis, and the themes and subgroups resulting from the qualitative coding are reported below.

XSL•FO
RenderX

Figure 1. Subgroups and themes or categories identified in the analysis. ML: machine learning.

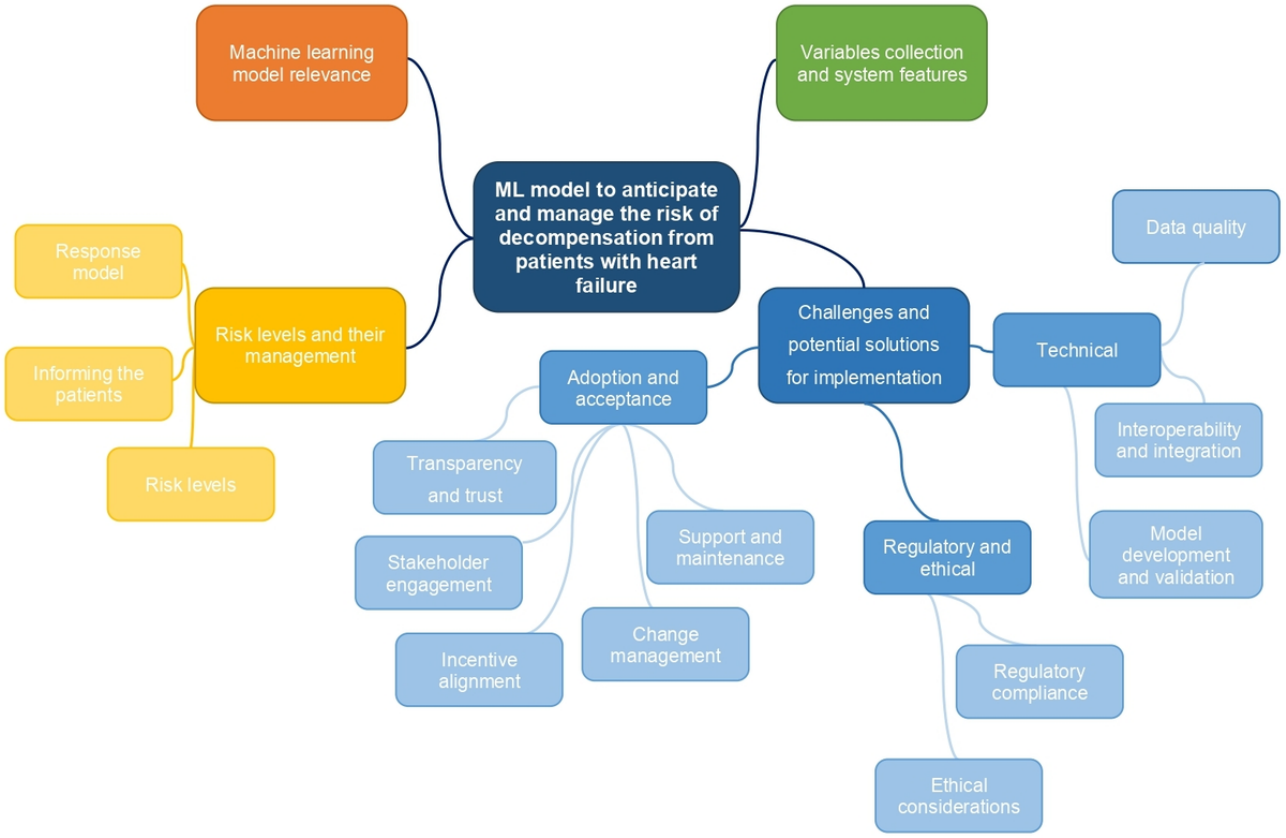

ML Model Relevance

Most participants highlighted the potential benefits of using an ML model to predict decompensation from patients with HF. These benefits included identifying patients at risk of decompensation earlier, allowing for more personalized care, improving patient outcomes, and reducing health care costs.

> *Is a very relevant thing, and early detection of decompensation is something that allows for reducing the number of hospitalizations, which is fundamental. Not only for the patient's quality of life but also for the burden on the National Health Service.* [Cardiologist 2]

> *Regarding models in general, predictive models, I think they are well received by health professionals...There is a clear notion that we are evolving toward an increasingly personalized medicine. We also know that today, waiting for these changes that start as subclinical to become clinical so that professionals or patients then feel that they exist; we are losing, in quotes, time, where intervention could be done earlier and more deleterious outcomes could be avoided, which in the case of heart failure in particular, the one always intended to be achieved is to avoid hospitalizations. Since any additional hospitalization in these patients has a worse prognosis in the natural history of the disease, the earlier we can adjust therapy so that no decompensation of heart failure leads to hospitalization, not only from the individual's perspective but also from the institution's perspective, avoiding pressure on emergency services and hospitals in general.* [Internal medicine physician 2]

Participants also considered the model could contribute to better management of the disease by the patient.

> *For me, the great advantage of digital technology in heart failure, for health professionals, is that it allows us to start preventing hospitalizations earlier. I think that for the patient, even those who are more reluctant to adhere to therapy because in heart failure, therapeutic adherence is also a problem; it gives them a certain empowerment over the disease.....So, I think it's a win-win for both sides.* [Nurse 2]

Variables Collection and System Features

Participants discussed the importance of identifying the correct variables for the ML model. Some variables mentioned include rapid weight gain and symptoms such as tiredness, shortness of breath, and swelling of the ankles (see Textbox 1 for a complete list of variables). Participants noted that selecting variables should be evidence-based and consider the complexity of HF management.

> *Weight gain of more than 2kg in 3 days; increase the number of pillows, the patient who sleeps with one pillow to two or three, or to sleep sitting up; increased edema of the lower limbs; increased coughing at night; finally, a feeling of worsening tiredness for minor efforts.* [Internal medicine physician 1]

XSL•FO
RenderX

Textbox 1. List of variables mentioned by the participants.

- Patient characterization: sex, age, comorbidities, and previous hospitalizations
- Heart failure characterization: New York Heart Association Classification and left ventricular ejection fraction
- Vital signs: weight, height, systolic blood pressure, diastolic blood pressure, heart rate, and blood oxygen saturation
- Laboratory results: creatinine, glomerular filtration rate based on the chronic kidney disease epidemiology collaboration equation, hemoglobin, hematocrit, urea, N-terminal pro b-type natriuretic peptide (NT-Pro BNP), ferritin, sodium, potassium, iron (fix capacity, iron), saturation (from transferrin), natriuretic peptide (BNP), and iron
- Prescribed medication: furosemide, bisoprolol, dapagliflozin, spironolactone, sacubitril, valsartan, carvedilol, lisinopril, nebivolol, ramipril, perindopril, torsemide, enalapril, empagliflozin, captopril, eplerenone, propranolol, atenolol, timolol, trandolapril, fosinopril, metoprolol, and sotalol

The collection of vital signs through wearables was deemed relevant by the participants.

> *I think asking someone every day, and even if it's every week, to record their weight, blood pressure, and frequency is tough. I don't think that's feasible in the long run. So the simpler, the better, and there, the devices are clearly in the way.* [Internal medicine physician 3]

However, as HF is prevalent among older people, using such devices was considered a challenge by some interviewees.

> *Most people with HF are elderly, and the likelihood of using wearables will be low.* [Primary care physician 1]

Interviewers also highlighted the need to select the population to use these tools properly.

> *Telemonitoring has two secrets. One of them is, to be cost-effective, the population you choose; I don't think it's a technology applied to all patients.* [Cardiologist 1]

Risk Levels and Their Management

Risk Levels

Participants emphasized the importance of developing an ML model capable of stratifying patients into distinct risk levels (high, medium, and low risk of decompensation), considering the patient's HF journey stage. This approach would facilitate health care professionals to provide more targeted interventions.

> *The literature in cardiovascular patients structures patients into three levels, possibly four. Roughly speaking, therefore, that is low risk, moderate risk, high risk, and, in some situations, there is a very high or very high-risk level. In the case of heart failure, risk stratification will also depend on the type of patient, namely the stage of their heart failure journey.....three levels, roughly speaking, sounds good, I think it's interesting in your work that you can adjust this risk stratification also to the stage of the heart failure patient journey.* [Internal medicine physician 2]

Response Model

Several participants discussed the importance of having a response model to validate the alerts generated by the ML model. They noted that the validation process should involve a health care professional, such as a nurse or a physician, who could assess the patient's condition and determine the appropriate response.

> *On the first level, the ideal situation would be to activate a call from a specialist heart failure nurse who would help the patient identify what they are doing right and what they are doing wrong. At an intermediate level...the structure would be the equivalent of an open consultation, a day hospital...a medical team. They may go through a nursing assessment first while consulting with the heart failure medical team. In the red level, I think the medical team must be activated to help decide what is going on and how the patient will be assessed.* [Internal medicine physician 3]

Informing the Patients

Participants discussed the potential benefits and challenges of providing alerts to patients identified as high-risk by the ML model. They noted that alerts could help patients take proactive steps to manage their condition, but there were concerns about patient anxiety.

> *When the information is to be given to the patient, either because it is considered that it has to be given, or because the patient has asked for this information, it cannot just be "you have an alert," there has to be an explanation. There has to be an explanation and especially a recommendation of what to do next, because otherwise, it creates a level of anxiety, "you have a change in your oxygen (levels)" (alert) and that's it. It would be a bit distressing for anyone.* [Internal medicine physician 2]

Challenges and Potential Solutions for Implementation

The participants identified several challenges regarding implementing the ML model, along with suggestions to address the identified challenges. They can be grouped into the following categories.

Technical Challenges

Data Quality

Participants emphasized the importance of obtaining high-quality and reliable health care data for training accurate ML models. They noted that health care data can often be messy,

XSL•FO
RenderX

incomplete, or contain errors, posing challenges in data preprocessing and cleaning.

> *There are other issues, not only of having rich data set but also types of quality of data sets, trying to understand how that type of data was collected, how that data was treated, how they were coded. For example, whether ICD 10 or 11 (is used), how are the coding practices like, the standard practice in terms of coding...all of this is necessary to understand the rules inherent to the type of data we have.* [Data scientist 1]

Model Development and Validation

Participants highlighted that designing and developing ML models suitable for health care applications requires domain expertise. It involves selecting appropriate algorithms, fine-tuning models for accuracy, interpretability, and generalizability, and conducting robust testing for model validation.

Interoperability and Integration

The participants mentioned that integrating ML models into existing health care systems and workflows can be challenging. They pointed out issues related to interoperability, compatibility with different data formats and systems, and the need for seamless integration with electronic health records and other health care software.

> *All of this (defining which data to feed into an ML model, EHRs, prescription drugs) also influences the cardiac care...and of course...take into account that many of these systems do not talk to each other, so there is no interoperability of systems, and this makes it difficult from the outset. There is, therefore, work to do beyond quality (assurance) to make all the data talk to each other, trying to see how possible it is to go to the individual level, that is, what data do you have and how can you have as much information as possible at the individual level...But in any case, interoperability here will be very crucial for how fast you can have the data ready for the model, even before deciding which model will be the best for this type of function.* [Data scientist 2]

Regulatory and Ethical Challenges

Regulatory Compliance

Compliance with regulatory frameworks, such as the General Data Protection Regulation in the European Union, was identified as a significant challenge in implementing ML models in health care. Participants stressed the importance of adhering to legal and ethical guidelines regarding data use, consent, and transparency, mainly due to the sensitive nature of health care data, such as patient records. Ensuring patient privacy and data security while using ML models posed notable challenges.

> *Without a doubt, there is a need to ensure that there is some guarantee here in terms of privacy and data security; this becomes immediate when you are joining many sources of data.* [Data scientist 2]

Ethical Considerations

The participants emphasized the need to develop and deploy ML models with ethical considerations. They underscored the importance of avoiding bias, ensuring fairness and equity in health care outcomes, and maintaining transparency in decision-making processes.

> *The real bias is systemic and must be controlled. Other studies (demonstrate the impact of) minorities who were underrepresented; then obviously when calculating the scores, they were underestimated, and the system continued as it always has, leaving those people on the sidelines.* [Data scientist 1]

> *There is another important issue, which is the representativeness and inclusivity within the data set. Many problems have happened because of the type of data and the conclusions reached that sometimes are not fair, (known as) algorithm justice.* [Data scientist 1]

These aspects were identified as crucial for building trust among health care professionals and patients.

Adoption and Acceptance Challenges

Transparency and Trust

As referred before, participants highlighted the importance of transparency in the ML model to gain the trust of health care professionals and patients, favoring the choice of models that have higher explainability, which refers to a characteristic of an AI-driven system that allows understanding and interpretation of the decisions and predictions made by an AI system [11].

Ensuring transparency in the decision-making process was perceived as crucial to fostering trust among stakeholders, as it allows them to understand the underlying factors and reasoning behind the ML model's recommendations or predictions.

> *As you build a model, there is a need to keep those metrics (sensitivity, specificity, therefore, recall, precision, F-score, accuracy) in mind and to perform a primary assessment of those metrics. Another important part that has been talked about more in the last few years and will involve both patients and clinicians is the technical issue of explainability—the idea that AI models can't just be a black box. If we want some form of engagement, some participation from both doctors and patients, and especially trust at the health care system level, as patients and clinical decisions are involved, the system should be something that we have to be able to explain to both stakeholders and patients and doctors.* [Data scientist 1]

Stakeholder Engagement

Engaging health care professionals, including physicians, nurses, and administrators, in designing and implementing ML models was highlighted as essential for successful adoption. Participants stressed the significance of addressing their concerns, providing training, and demonstrating the benefits of ML models to gain their support and involvement.

XSL•FO
RenderX

> *I think our role has to be to encourage through the constant involvement of the professional in the development of the model and implementation, as well as the validation process and everything, but also to demonstrate how those models are not the panacea.* [Data scientist 2]

Change Management

Participants acknowledged that implementing ML models might change health care organizations' workflows, processes, and roles. Additionally, participants expressed concerns that ML models might create an increased workload for health care professionals. They emphasized the need for careful planning and allocation of resources to ensure that the integration of ML models into clinical workflows does not burden the health care workforce.

> *And the machine learning model was a part that, according to the set of information, whether it is from the electronic medical record, such as comorbidities, etc., whether it is collected daily, we have a model here that basically gives us alert levels, which can be, for example, three alert levels, low, medium, high, etc. and act only on those that are medium or high, so that we do not have too many false positives.* [Cardiologist 3]

Incentive Alignment

Participants raised the issue of financial incentives in health care, particularly the existing payment structure that often rewards hospitals based on the activity (number of admissions) rather than patient outcomes. They acknowledged that highly effective ML models capable of accurately predicting admissions might be challenging to implement due to misaligned incentives.

> *Today, the incentive system in virtually all health care systems is based on procedures, it is based on records made, compliance with a process defined as correct, or the number of surgeries or consultations carried out. So, that's how productivity is measured, and compensation is defined, and that is problematic because the system is conditioned to have a short-term perspective. For example, there is a development of a model that manages to reduce hospitalizations by 50%. But the hospital, like everywhere else in Portugal, only gets reimbursed for the number of hospitalized patients. And the more seriously ill the patients are, the more they may be paid. You run the risk that the hospital thinks your model is very good, but in terms of incentives gained, it doesn't make sense to be applied.* [Data scientist 3]

Support and Maintenance

Participants stressed the importance of having dedicated resources and personnel to provide ongoing support and address the doubts and concerns of health care professionals regarding the ML model.

In addition, the need for a designated team responsible for the maintenance and upkeep of the ML model was identified as crucial for ensuring its continuous performance, addressing any technical issues, and incorporating updates or improvements as needed.

> *We don't have medical information officers...for software. I published an article with colleagues about mental health professionals' perceptions of digital health tools, whether apps, websites, etc. One of the biggest problems was that the professionals did not know when to use these apps; they did not have guidelines or recognition from the societies and professional groups they belonged to. It was also noted that there was no contact and no use of those applications—when questioned about the support given by the tool developers, many participants either responded that they did not know or did not answer.* [Data scientist 3]

Discussion

Principal Findings

The findings of this study have several implications for the practical application of ML models in the early detection and management of HF decompensation. The identified themes and subgroups provide insights into critical considerations that health care professionals and data scientists consider relevant when developing and implementing ML models for predicting the risk of decompensation in patients with HF.

The participants recognized the potential benefits of using ML models for the early identification of patients at risk of decompensation, leading to improved patient outcomes and reduced health care costs, a notion that aligns with the findings from previous research [4,12-14].

The emphasis on the importance of early detection is particularly noteworthy, as it aligns with the evolving trend toward personalized care. The ability to intervene before subclinical changes become clinical can significantly impact patient outcomes, emphasizing the critical role of ML models in leading to timely intervention and avoiding hospitalizations [13,15]. This perception was echoed by health care professionals from various domains, that is, cardiologists, internal medicine physicians, and nurses.

The selection of appropriate variables is crucial for accurately predicting HF decompensation using ML models. The participants emphasized the importance of evidence-based variable selection and the consideration of the complexity of HF management. Variables such as rapid weight gain and reported symptoms such as orthopnea and swelling of lower limbs were mentioned as potentially relevant factors for inclusion in the ML models, which aligns with previous studies that identified these as significant predictors of cardiac decompensation events in patients with HF [13].

Furthermore, the use of wearables for collecting vital signs was seen as relevant, but the challenges related to adoption among older patients must be addressed. Additionally, the importance of adequate population selection to ensure cost-effectiveness was highlighted. These findings align with the current literature that acknowledges the potential benefits of the use of wearables but identifies a set of barriers that need to be addressed for

XSL•FO
RenderX

further implementation of these devices into clinical practice [16].

Participants highlighted the importance of ML models identifying different risk levels for patients with HF, considering the patient's HF journey stage [17,18]. By accurate risk stratification, health care professionals can prioritize their interventions and provide more targeted care to those at higher risk [19,20].

It was also identified the relevance of ML models being designed to provide risk stratification outputs that are easily interpretable and actionable by health care professionals. This may involve the development of user-friendly interfaces or decision support systems that show risk levels clearly and intuitively. This perspective aligns with the recognized need for interpretable ML models, as interpretable models enable users to examine, comprehend, troubleshoot, and potentially enhance ML systems. By clarifying the logic behind predictions, interpretable ML systems provide users with justifications for accepting or rejecting predictions and recommendations [21].

Additionally, the participants emphasized the need for a response model to validate the alerts generated by the ML models. The involvement of health care professionals, such as nurses or physicians, in assessing patients' conditions and determining appropriate responses is crucial for effectively managing decompensation.

The study identified several challenges in implementing ML models for predicting HF decompensation and potential solutions. These can be grouped into three categories: technical, regulatory and ethical, and adoption and acceptance challenges, which are aligned with this literature [22,23].

The importance of high-quality and reliable health care data was recognized, with an acknowledgment of the complexities involved in data preprocessing and cleaning, as already addressed in the literature [24]. The need for domain expertise in designing and developing ML models was emphasized, along with challenges related to interoperability and integration with existing health care systems. The existence of proposed frameworks to guide the development and integration of ML models in health care, such as the one developed by Assadi et al [25], frames the findings of this study within this literature.

Compliance with regulatory frameworks, particularly regarding data protection, was identified as a significant challenge. Ethical considerations, including avoiding bias and ensuring fairness and transparency, were highlighted as crucial for building trust among health care professionals and patients, which is in line with what is stated by the European Parliament [26] and recent literature [27].

Transparency and trust were emphasized as critical factors for the successful adoption of ML models [28]. Stakeholder engagement, particularly involving health care professionals, was recognized as essential, along with considerations for change management to avoid disrupting existing workflows, which had already been identified in a similar study [29].

The presence of misaligned financial incentives within the health care system was recognized as a potential obstacle to the integration of ML models. This underscores the importance of establishing a successful reimbursement strategy to align interests among stakeholders, thereby guiding the application of value-based AI in health care, as payment strategies impact utilization, cost, and quality of care across the clinical use cases for AI [30]. A recent paper in *NPJ Digital Medicine* expands into the discussion on the appropriate method of reimbursement for AI systems in health care [31].

The importance of ongoing support, addressing the concerns of health care professionals, and having a dedicated team for the maintenance of the ML model was also recognized as relevant for successfully implementing these models into practice [32].

In general, this study's findings emphasize the potential of ML models in improving the early detection and management of HF decompensation. Considering the implications for practice outlined above, this study is expected to contribute to developing and implementing ML models that positively impact patient outcomes, health care costs [33], and the delivery of personalized care in HF management.

In addition, this study highlighted the synergy between the perspectives of health care professionals and data scientists not only enhances the accuracy and effectiveness of ML models but also fosters a holistic approach that considers both the clinical and technical facets, promoting the successful integration of these models into health care practices. From the health care professionals' perspective, there is a strong emphasis on evidence-based variable selection, drawing on clinical expertise to identify relevant indicators for predicting decompensation in patients with HF. Their insights into the nuanced aspects of patient care, risk stratification, and the practical implications of ML models contribute crucially to the model's alignment with real-world clinical workflows.

On the other side, data scientists bring technical expertise to the table, highlighting challenges related to data quality, model development, and interoperability. They underscore the importance of domain knowledge in ML model development, recognizing the complexities involved in designing accurate, interpretable, and generalizable models for health care applications. Data scientists also stress ethical considerations, algorithmic fairness, and the need for transparent decision-making processes, emphasizing the broader societal impact of ML models in health care. In short, close collaboration between these two professional domains is paramount [25].

Future research should address these concerns and evaluate the effectiveness of ML models in improving patient outcomes.

Strengths and Limitations

The findings of this study provide relevant insights into health care professionals' and data scientists' perspectives on using ML models for predicting HF decompensation. However, it is essential to consider the strengths and limitations of the study methodology and findings.

Regarding the strengths, using a qualitative research design allowed for an in-depth exploration of health care professionals' and data scientists' perspectives on ML models for HF decompensation prediction. The semistructured interviews

XSL•FO
RenderX

provided rich and detailed data, enabling a comprehensive analysis of themes and subgroups. In addition, the thematic analysis approach used in this study, following the guidelines proposed by Braun and Clarke [9], provided a systematic and rigorous process for analyzing the data. Two researchers were involved in the analysis process, enhancing the credibility and reliability of the findings.

The diverse range of health care professionals, including a primary care physician, cardiologists, internal medicine physicians, nurses, and data scientists, and the high level of experience from the participants are also strengths of the study, enhancing the validity and generalizability of the findings. In addition, the fact that participants were from different contexts and regions across the country is also a positive point, reducing the influence of the local context on the overall results.

The determination of data saturation, where no new themes emerged from participants' perspectives, ensured sufficient sample size was achieved to capture the breadth and depth of the participants' experiences and opinions.

Considering the study's limitations, we emphasize that, although data saturation was reached, the sample size of 13 participants may be considered relatively small. The findings may not fully represent the perspectives of all health care professionals involved in HF management and data scientists' considerations on this specific issue.

In addition, data were collected solely through semistructured interviews. While interviews provide rich qualitative data, multiple data collection methods, such as surveys or observation, could have provided a more comprehensive understanding of the topic.

As with any qualitative study, there is a potential for researcher bias during data collection, analysis, and interpretation. The researchers' perspective may have influenced the findings. However, efforts were made to minimize this influence through rigorous data analysis and researchers' discussions.

In addition, another significant limitation of this study is the potential loss of information during the translation process of the interviews from Portuguese to English. Language translation can be complex, and nuances, cultural references, and contextual information may not always be accurately expressed in the translated version. This could lead to the loss of relevant insights and subtle meanings expressed by the participants in their original language. To mitigate this limitation, cross-validation of translations and back-translation were used to ensure the accuracy of the translated data.

Furthermore, the patients' perspectives were not included in this study. Incorporating patient viewpoints would have provided a more holistic understanding of the potential benefits and challenges associated with ML models for HF decompensation prediction.

Future research should address these limitations by including more extensive and diverse samples, incorporating multiple data collection methods, and considering patients' perspectives.

Conclusions

This study explored health care professionals and data scientists' perspectives on using ML models to predict HF decompensation. The findings provide important insights into the potential benefits, challenges, and considerations associated with implementing ML models in HF management.

The study highlighted several key themes, including the relevance of ML models in identifying patients at risk of decompensation earlier, the importance of selecting appropriate variables for model development, the need for risk stratification and response models, and the challenges and potential solutions for implementing ML models in health care settings. These findings have implications for practice in HF management and the development of effective ML-based predictive models.

Future research should address the limitations of this study, such as expanding the sample size and including patients' viewpoints. Additional research areas include evaluating ML models' impact on patient outcomes, health care costs, and the delivery of personalized care.

The incorporation of this study's findings into practice is expected to contribute to the development and implementation of ML models that can potentially improve the early detection and management of HF decompensation, enhancing patient outcomes and reducing health care costs.

Acknowledgments

The authors are grateful to Vanessa Nicolau and Isabel Andrade for their support and suggestions. Special thanks to all the health care professionals and data scientists who participated in this study. This publication was funded by Fundação Ciência e Tecnologia, IP national support through the Comprehensive Health Research Center (UIDP/04923/2020).

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JS collected the data. JS and AH analyzed and interpreted the data. JS wrote the original draft of the manuscript. ARP provided comprehensive feedback during an extensive review of the manuscript and actively participated in discussions about the findings. RS and TM revised the manuscript. All authors contributed to reviewing and editing the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Guide for semistructured interviews.
[PDF File (Adobe PDF File), 20 KB - jmir_v27i1e54990_app1.pdf]

References

1. Krzesiński P, Siebert J, Jankowska EA, Galas A, Piotrowicz K, Stańczyk A, et al. Nurse-led ambulatory care supported by non-invasive haemodynamic assessment after acute heart failure decompensation. ESC Heart Fail 2021;8(2):1018-1026 [FREE Full text] [doi: 10.1002/ehf2.13207] [Medline: 33463072]
2. Verdu-Rotellar JM, Vaillant-Roussel H, Abellana R, Jevsek LG, Assenova R, Lazic DK, et al. Precipitating factors of heart failure decompensation, short-term morbidity and mortality in patients attended in primary care. Scand J Prim Health Care 2020;38(4):473-480 [FREE Full text] [doi: 10.1080/02813432.2020.1844387] [Medline: 33201746]
3. Saner H, Schuetz N, Buluschek P, Du Pasquier G, Ribaudo G, Urwyler P, et al. Case report: ambient sensor signals as digital biomarkers for early signs of heart failure decompensation. Front Cardiovasc Med 2021;8:617682 [FREE Full text] [doi: 10.3389/fcvm.2021.617682] [Medline: 33604357]
4. Larburu N, Artetxe A, Escolar V, Lozano A, Kerexeta J. Artificial intelligence to prevent mobile heart failure patients decompensation in real time: monitoring-based predictive model. Mobile Inf Syst 2018;2018:1-11. [doi: 10.1155/2018/1546210]
5. Gautam N, Ghanta SN, Mueller J, Mansour M, Chen Z, Puente C, et al. Artificial intelligence, wearables and remote monitoring for heart failure: current and future applications. Diagnostics (Basel) 2022;12(12):2964 [FREE Full text] [doi: 10.3390/diagnostics12122964] [Medline: 36552971]
6. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577(7788):89-94. [doi: 10.1038/s41586-019-1799-6] [Medline: 31894144]
7. Dutta A, Hasan MK, Ahmad M, Awal MA, Islam MA, Masud M, et al. Early prediction of diabetes using an ensemble of machine learning models. Int J Environ Res Public Health 2022;19(19):12378 [FREE Full text] [doi: 10.3390/ijerph191912378] [Medline: 36231678]
8. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
9. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
10. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
11. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
12. Kerexeta J, Larburu N, Escolar V, Lozano-Bahamonde A, Macía I, Beristain Iraola A, et al. Prediction and analysis of heart failure decompensation events based on telemonitored data and artificial intelligence methods. J Cardiovasc Dev Dis 2023;10(2):48 [FREE Full text] [doi: 10.3390/jcdd10020048] [Medline: 36826544]
13. Kerexeta J, Larburu N, Escolar V, Lozano-Bahamonde A, Macía I, Beristain Iraola A, et al. Prediction and analysis of heart failure decompensation events based on telemonitored data and artificial intelligence methods. J Cardiovasc Dev Dis 2023;10(2):48 [FREE Full text] [doi: 10.3390/jcdd10020048] [Medline: 36826544]
14. Croon PM, Selder JL, Allaart CP, Bleijendaal H, Chamuleau SAJ, Hofstra L, et al. Current state of artificial intelligence-based algorithms for hospital admission prediction in patients with heart failure: a scoping review. Eur Heart J Digital Health 2022;3(3):415-425 [FREE Full text] [doi: 10.1093/ehjdh/ztac035] [Medline: 36712159]
15. Susič D, Poglajen G, Gradišek A. Identification of decompensation episodes in chronic heart failure patients based solely on heart sounds. Front Cardiovasc Med 2022;9:1009821 [FREE Full text] [doi: 10.3389/fcvm.2022.1009821] [Medline: 36457810]
16. Singhal A, Cowie MR. The role of wearables in heart failure. Curr Heart Fail Rep 2020;17(4):125-132 [FREE Full text] [doi: 10.1007/s11897-020-00467-x] [Medline: 32494944]
17. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2023 focused update of the 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2023;44(37):3627-3639. [doi: 10.1093/eurheartj/ehad195] [Medline: 37622666]
18. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, Writing Committee Members, ACC/AHA Joint Committee Members. 2022 AHA/ACC/HFSA guideline for the management of heart failure. J Card Fail 2022;28(5):e1-e167. [doi: 10.1016/j.cardfail.2022.02.010] [Medline: 35378257]

XSL•FO
RenderX

19. Jing L, Cerna AEU, Good CW, Sauers NM, Schneider G, Hartzel DN, et al. A machine learning approach to management of heart failure populations. JACC Heart Fail 2020;8(7):578-587 [FREE Full text] [doi: 10.1016/j.jchf.2020.01.012] [Medline: 32387064]
20. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. JAMA Netw Open 2020;3(1):e1918962 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18962] [Medline: 31922560]
21. Mckelvey TG, Ahmad MA, Eckert C, Teredesai A, Mckelvey G. Interpretable machine learning in healthcare. In: IEEE Computer Society. 2018 Presented at: IEEE Computer Society; 2018 June 07; USA. [doi: 10.1109/ICHI.2018.00095]
22. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare J 2019;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
23. Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. Nat Biomed Eng 2022;6(12):1330-1345. [doi: 10.1038/s41551-022-00898-y] [Medline: 35788685]
24. Roy R, Mukherjee S, Baral MM, Badhan AK, Ravindra M. Challenges encountered in the implementation of machine learning in the healthcare industry. New York: Springer; 2023.
25. Assadi A, Laussen PC, Goodwin AJ, Goodfellow S, Dixon W, Greer RW, et al. An integration engineering framework for machine learning in healthcare. Front Digital Health 2022;4:932411 [FREE Full text] [doi: 10.3389/fdgth.2022.932411] [Medline: 35990013]
26. Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts. European Parliament. URL: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512 [accessed 2022-06-01]
27. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
28. Wanner J, Herm LV, Heinrich K, Janiesch C. The effect of transparency and trust on intelligent system acceptance: evidence from a user-based study. Electron Markets 2022;32(4):2079-2102. [doi: 10.1007/s12525-022-00593-5]
29. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in sweden. BMC Health Serv Res 2022;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
30. Venkatesh KP, Raza MM, Diao JA, Kvedar JC. Leveraging reimbursement strategies to guide value-based adoption and utilization of medical AI. NPJ Digital Med 2022;5(1):112 [FREE Full text] [doi: 10.1038/s41746-022-00662-1] [Medline: 35948612]
31. Parikh RB, Helmchen LA. Paying for artificial intelligence in medicine. NPJ Digital Med 2022;5(1):63 [FREE Full text] [doi: 10.1038/s41746-022-00609-6] [Medline: 35595986]
32. Hogg HDJ, Al-Zubaidy M, Talks J, Denniston AK, Kelly CJ, Malawana J, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. J Med Internet Res 2023;25:e39742 [FREE Full text] [doi: 10.2196/39742] [Medline: 36626192]
33. Candelieri A, Conforti D, Perticone F, Sciacqua A, Kawecka-Jaszcz K, Styczkiewicz K. Early detection of decompensation conditions in heart failure patients by knowledge discovery: the HEARTFAID approaches. 2008 Presented at: Computers in Cardiology; September 14-17, 2008; Bologna, Italy. [doi: 10.1109/cic.2008.4749186]

Abbreviations

AI: artificial intelligence
HF: heart failure
ML: machine learning

Edited by T de Azevedo Cardoso; submitted 29.11.23; peer-reviewed by H Chen, GB Lee; comments to author 22.05.24; revised version received 30.07.24; accepted 26.10.24; published 20.01.25.

Please cite as:
Seringa J, Hirata A, Pedro AR, Santana R, Magalhães T
Health Care Professionals and Data Scientists' Perspectives on a Machine Learning System to Anticipate and Manage the Risk of Decompensation From Patients With Heart Failure: Qualitative Interview Study
J Med Internet Res 2025;27:e54990
URL: https://www.jmir.org/2025/1/e54990
doi:10.2196/54990
PMID:

XSL•FO
RenderX

©Joana Seringa, Anna Hirata, Ana Rita Pedro, Rui Santana, Teresa Magalhães. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Transformation in Patient Organizations: Interview and Focus Group Study

Simon Wallraf[1], MSc; Sara Köthemann[2], MA; Claudia Wiesemann[2], Prof Dr; Sabine Wöhlke[3], Prof Dr; Marie-Luise Dierks[1], Prof Dr; Marion Andrea Schmidt[2], PhD; Henk Jasper van Gils-Schmidt[3], PhD; Jonas Lander[1], PhD

[1]Institute for Epidemiology, Social Medicine and Health Systems Research, Hannover Medical School, Hanover, Germany

[2]Department of Medical Ethics and History of Medicine, University Medical Center Göttingen, Göttingen, Germany

[3]Department of Health Sciences, Hamburg University of Applied Sciences, Hamburg, Germany

Corresponding Author:
Jonas Lander, PhD
Institute for Epidemiology, Social Medicine and Health Systems Research
Hannover Medical School
Carl-Neuberg-Street 1
Hanover, 30625
Germany
Phone: 49 0511 532 4450
Email: lander.jonas@mh-hannover.de

Abstract

Background: Patient organizations (POs) are an integral part of the health care landscape, serving as advocates and support systems for patients and their families. As the digitalization of health care accelerates, POs are challenged to adapt their diverse roles to digital formats. However, the extent and form of POs' digital adaptation and the challenges POs encounter in their digital transformation remain unexplored.

Objective: This study aims to investigate the digital transformation processes within POs. We examined the types of digital activities and processes implemented, people involved in respective tasks, challenges encountered, and attitudes toward the digitalization of POs.

Methods: The study was carried out by the multicenter interdisciplinary research network Pandora. We adopted a qualitative exploratory approach by conducting 37 semistructured interviews and 2 focus groups with representatives and members of POs in Germany. Results were obtained using a deductive-inductive approach based on a qualitative content analysis. Methods and results were reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.

Results: POs primarily apply basic digital tools to engage in communication, health education, and information dissemination. Some also develop specific mobile apps and collect health data through patient registries. Volunteers cover a considerable part of the workload. Sometimes, POs collaborate with external partners, such as health professionals or other nonprofit organizations. Furthermore, many (13/46, 28%) interviewees referred to the importance of involving members in digitalization efforts to better meet their needs. However, they described the actual practices used to involve members in, for example, developing digital services as limited, passive, or implicit. When evaluating digital transformation processes, representatives and members of POs expressed generally positive attitudes and acknowledged their potential to improve the accessibility of support services, management efficiency, and outreach. Still, resource constraints; the complexity of digital initiatives; and accessibility issues for certain demographic groups, especially older persons, were frequently mentioned as challenges. Several (15/46, 33%) interviewees highlighted POs' increasing responsibility to support their members' digital competencies and digital health literacy.

Conclusions: POs are actively involved in the digital transformation of health services. To navigate challenges and further shape and sustain digital activities and processes, POs may benefit from governance frameworks, that is, a clear plan outlining with whom, how, and with what objectives digital projects are being realized. Support from public, scientific, and policy institutions to enhance the process through training, mentorship, and fostering collaborative networks seems warranted.

(J Med Internet Res 2025;27:e62750) doi:10.2196/62750

KEYWORDS

patient organization; patient support; digitalization; digital transformation; health research

XSL•FO
RenderX

Introduction

Background

The digital transformation of health care is an ongoing process in which health care–related services, such as types of care, diagnostic methods, and health information, are being digitalized [1-3]. The degree of digitalization varies considerably between different health care organizations and health sectors, and this gap has widened during the COVID-19 pandemic. This is due to, inter alia, the available resources and organizational structures, the degree of digital literacy of those responsible for implementing digitalization, and attitudes toward digital services [4,5]. Some institutional actors within the health care system, such as health insurance companies, ask individuals to contribute to the digital transformation and develop their own digital literacy.

It has been argued that the development of individual skills and attitudes toward digital transformation requires health care institutions to foster comprehensive governance frameworks that specify necessary goals, needs, and methods [3]. Furthermore, an individual's health is typically not managed by a single health care professional but rather involves multiple entities. To make this possible, these institutions need to provide the necessary resources and define the appropriate levels of contribution and literacy [3]. Overall, digital transformation requires more collective efforts across the health care sector to counter the uneven degrees of digital transformation [5].

Alongside primary care services, patient organizations (POs) are now recognized as key health advocates, providing vital support to patients, people with chronic conditions, and their families. While varying in size and scope, all POs aim to empower individuals to better manage their condition, for example, by providing health information and self-management resources [6,7]. Previous research shows that POs represent patient voices in health care and health care policy making by developing new forms of collaboration [8-10] and research facilitators and partners [11,12]. In fact, for countries such as the United Kingdom, the Netherlands, Sweden, and Germany, POs are now one of the central contributors to health care research, for example, when their members contribute to the planning and implementation of study projects. The main domain in which this takes place is rare disease research [13,14].

Today, POs face the need to join in the digital transformation of society and the medical field specifically [15]. Furthermore, POs face the challenges associated with shifting to digitalized work processes, for example, the search for appropriate resources. PO-specific challenges, such as the proportion of susceptible and older people who seek help from POs [16-18], add to these demands.

Objectives

To assist POs in shaping their own digital transformation, a better understanding of current practices, opportunities, and potential barriers is needed, which is, to the best of our understanding, currently largely lacking [12]. Hence, this study aimed to explore the current state of digital transformation in German POs. We aimed to do so by answering four specific research questions (RQs):

1. Which digital projects and services do POs currently implement? (RQ 1)
2. Who are the individuals and institutional actors involved in the process of planning and implementing digital projects and activities? (RQ 2)
3. To what extent and how are PO members involved in digitalization efforts? (RQ 3)
4. Which experiences with and attitudes vis-à-vis digitalization in POs do representatives and members express? (RQ 4)

Methods

Study Design and Context

This study was conducted as part of the multicenter research network Pandora [19]. Pandora investigates how POs contribute to the digital health transformation, shape their own digital practices, and address the challenges they face therein [20]. For this study, the 3 study sites pursued distinct, although related, objectives within the same overarching RQs defined by the aim of Pandora. Digitalization practices in POs were explored via semistructured interviews and focus groups. An exploratory design was chosen, as these practices have rarely been investigated so far. We used the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist to report our methodology (Multimedia Appendix 1) [21].

Theoretical Foundation

Overview

To determine the scope and contents of this study, we narratively summarized current insights related to the roles and tasks of POs nationally and internationally. We used this understanding for our own investigation of the digital transformation within these organizations, that is, to derive broad content-wise foci. We chose this strategy as there has been no qualitative study on the digitalization practices in PO. The available evidence on the *ways of working* of PO centers on 3 main pillars.

Tasks and Governance Processes

POs' routine daily tasks and responsibilities have been described as centering on 4 main aspects: policy development and advocacy (eg, POs represent patient interests vis-à-vis political decision makers), empowerment and education (eg, POs develop and deliver health information on all kinds of chronic and acute health conditions), peer support (eg, POs conduct self-management courses to foster individuals' coping with chronic diseases), and research (eg, POs collect and provide individual health data from their members for use by clinical researchers) [15]. More generally, previous studies have concluded that, with the evolving existence, role, and scope of work domains of POs, they increasingly prioritize professionalization, such as creating professional management and organizational structures [8].

Actor and Member Involvement

Given the breadth of tasks and responsibilities of POs, their efforts to professionalize services and structures and the

XSL•FO
RenderX

continuous need for sufficient personnel, technical, administrative, and financial resources, they need to engage in collaboration with other stakeholders [10,22]. Herein, POs not only engage with institutions and organizations but also seek to promote exchange with and involvement of their direct members. To date, this has been most evident in relation to research: POs not only act as intermediaries between research and patients or target groups but also, for example, train their members without a scientific educational background, to become active research partners [12,23].

Principles and Values Underlying the Digital Work of POs

So far, the roles and activities attributed to POs, such as competing with other health care stakeholders for resources, using support from industry to develop and maintain their own services, and advocating for patients' interests, have been described in terms of being directly or indirectly related to ethical aspects [18,24]. For example, POs need to be transparent regarding their partners; they need to balance individual protection and autonomy with collective responsibilities, such as in research advances; and they need to handle rapidly increasing amounts of patient data. However, the distinct challenges and ways in which POs go about their digital transformation have not been studied.

Target Group

To explore the digital transformation in POs, we invited representatives and members of POs in Germany to participate in this study. Except for the recruitment, no previous relationships with the target group existed. We defined representatives as those persons who work full time, part time, or voluntarily for a PO and are involved in leadership roles. Being affected by a chronic illness or disability was not a requirement to be recognized as a PO representative. PO members are individuals affected by the disease or disability that the PO they are registered with advocates for. It was inconsequential whether they were involved in the PO's work as long as they were not in a leadership role. We did not include professionals, such as medical staff or clinical researchers. Regarding the sampling procedure, we aimed to include PO representatives and PO members in equal shares. Furthermore, we aimed to approach numerous POs (refer to the Recruitment Strategy section) to increase diversity in terms of geographic location, size, and the specific health conditions or issues they represent. Each study site recruited participants until these criteria were sufficiently met, and no new topics were mentioned by the interviewees.

Recruitment Strategy

We first conducted a manual web-based search and a comprehensive review of German POs' umbrella organizations and their member organizations. These findings were then merged into a single list of German POs. As part of this process, we screened the POs' websites for clear indications of involvement in digital activities beyond merely having a website. POs that did not meet our criteria were excluded, as we focused on organizations where digital transformation processes were already underway, allowing us to explore ongoing practices and experiences in more detail. The final list contained 96 organizations, which were included in a purposive recruitment process at each study center. This included sending invitations by email, making follow-up calls if there was no response to the initial approach, forwarding the invitation to PO members via the POs' channels of communication, and approaching PO members directly via the Pandora advisory board.

Data Collection

Qualitative data collection was conducted from October 2022 to April 2023 mainly through individual telephone interviews and complemented by 2 web-based focus groups with PO representatives, the latter to increase practicability for the interviewees (S Wallraf, SK, and HJVGS). All junior and senior research fellows had a background in public health or medical ethics. To develop the semistructured interview guideline, we drafted headlines, ie, main topics according to the literature-based conceptualization of POs outlined earlier in the first step. Then, we added concrete questions for each topic and added new aspects where necessary, that is, to establish the focus on digitalization. While the interview guidelines varied slightly per study site, each included three common sections: (1) general aspects related to the level and relevance of digitalization in the respective PO (eg, digital governance, digital activities, and involved actors); (2) perceived opportunities and challenges of digital technologies (ie, appraisal of principles and values); and (3) the digital transformation of the German health care system (eg, digital governance and involved actors; Multimedia Appendix 2).

During the interviews, we focused on understanding how and in which areas such processes take place. This approach allowed us to capture a range of digitalization practices and to identify patterns of digital activity, ensuring a comprehensive understanding of digitalization. Specific technologies were discussed when mentioned by the interviewees. As the third step in the development of the interview guideline, we gathered feedback on the contents and wording of the draft from the Pandora patient advisory board (n=4). Fourth, we pretested the preliminary interview guidelines with representatives of the target group (study center 1: n=2, 50%; study center 2: n=1, 25%; and study center 3: n=1, 25%). Adjustments were made to the number of questions asked, the interview length, the wording of questions to further adapt them to the target group, and the sequencing of topics. At the beginning of the interviews, for which only the researcher and the participants were present, we provided general information on the researchers' role in the study without mentioning individual characteristics. We did not conduct repeat interviews and used field notes to structure them.

Data Analysis

Each audio file was transcribed verbatim to ensure the accuracy of the statements, pseudonymized, and then exported for analysis in MAXQDA 22 (VERBI Software GmbH). To generate our findings, we conducted a qualitative content analysis based on the steps described by Kuckartz [25].

First, we deductively derived the initial main categories for analysis based on our interview guidelines and agreed on these categories among the coauthors. Second, we coded

XSL•FO
RenderX

approximately 10% (4/39) of the manuscripts to test the suitability of the main categories. Third, we compared the results of this test phase and further developed and refined the coding system until the main authors agreed (S Wallraf, SK, HJVGS, and JL); refer to the Results section for an overview of the main categories and subcategories. Fourth, we analyzed all remaining manuscripts with the final version of the codebook (S Wallraf, SK, and HJVGS). For this analysis, we used consensual coding; that is, the main person coding discussed any uncertainties with one of the coauthors based on the definitions established for each category in the codebook.

Fifth, once the process of applying the main codes to all manuscripts was complete, we inductively derived subcategories within each main category by screening each code within the main categories to further concretize the content.

Sixth, we added frequencies to each subcategory based on the individual mentions of each aspect to gain an overview of the weight of each aspect mentioned by our participants.

After completing the first full version of the analysis, we presented summaries of the preliminary findings for each main category to the senior Pandora project staff not involved in the data analysis for feedback and interpretation of the findings. Furthermore, the coded segments in the transcripts were peer checked for correct interpretation by those authors who performed the analysis (HJVGS, JL, SK, and S Wallraf) when refining the coding system and during the actual analysis.

Ethical Considerations

Our study adheres to the Declaration of Helsinki, and ethics approval was obtained from the ethics committees of each study center (project 1: 11_6_22_University Medical Center Göttingen and 2022_20_University of Applied Sciences Hamburg; project 2: 10395_BO_K_2022; and project 3: 11_6_22_University Medical Center Göttingen). All participants consented to their participation before data collection and were given the option to opt out. Participants received no financial compensation. All data were pseudonymized. We did not collect patient data or information.

Results

Participant Characteristics

We interviewed 46 participants (n=26, 57% female individuals and n=20, 43% male individuals) who were either a member (n=26, 57%) or a representative, that is, staff (n=20, 43%) of a PO (Table 1).

The mean duration of interviews was 71 (SD 13.7) minutes, and focus groups lasted for 124 minutes. Most (17/46, 37%) interviewees were aged 45 to 59 years, followed by 15 (33%) participants aged ≥60 years. Interviewees often reported having completed secondary (7/46, 15%) or tertiary (32/46, 70%) education, equivalent to a high school or university degree, respectively. The POs, most (18/19, 95%) of them acting at a national level, involved were diverse in size, with 21% (4/19) acting as umbrella organizations (Table 2).

The main themes that were addressed included digital activities (RQ 1), actors involved in POs' digital transformation (RQ 2), participatory approaches to digital transformation (RQ 3), and current attitudes toward POs' digital efforts (RQ 4). Overall, POs engaged in a range of digital activities with distinct aims; relied on and engaged in collaboration with volunteers and other institutions to cover the workload; and faced considerable technical, conceptual, and motivational hurdles along the way.

Table 1. Characteristics of participants (N=46).

Characteristic	Participants, n (%)
Age group (y)	
18-29	4 (9)
30-44	10 (22)
45-59	17 (37)
≥60	15 (33)
Sex	
Male	20 (43)
Female	26 (57)
Educational level	
Primary education	1 (2)
Secondary education	7 (15)
Vocational training	6 (13)
Tertiary education	32 (70)
Status in patient organization	
Member	26 (57)
Staff or representative	20 (43)

XSL•FO
RenderX

Table 2. Characteristics of patient organizations (n=19).

Characteristic	Patient organizations, n (%)
Size (members), n	
<100	1 (5)
100-500	3 (16)
501-1000	4 (21)
1001-10,000	3 (16)
10,001-100,000	3 (16)
>100,000	1 (5)
N/A[a] (umbrella organization)	4 (21)
Geographic scope	
Regional	1 (5)
National	18 (95)
Head office location (federal states)	
Baden-Württemberg	1 (5)
Bavaria	1 (5)
Berlin	5 (26)
Hesse	2 (11)
Lower Saxony	1 (5)
North Rhine-Westphalia	7 (37)
Rhineland-Palatinate	2 (11)

[a]N/A: not available.

RQ 1: Digital Activities of POs

The analysis of our interviews and focus groups resulted in 4 primary areas where POs engaged in digital activities: communication, administration, health education, and health research.

Subtheme 1.1: Digital Communication

POs mostly used basic communication tools (38/46, 83%), including email, websites, videoconferencing, instant messaging services, and social media, to internally communicate with their members, facilitate communication among members, and externally engage with the wider public (refer to subtheme 1.1 in Table 3).

XSL•FO
RenderX

Table 3. Main themes and subthemes mentioned by research participants.

Main theme and subtheme	Example quotes
Digital activities of POs[a]	
1.1. Digital communication	"Internally,... we started with Zoom and now use Teams for communication. We also use WhatsApp for quick coordination and brief information flows, for example with our self-help group leaders or within the board.... Last year, we started using Office 365 with the goal of working centrally through Teams in the future and using Teams groups, so that, for instance, our treasurer has everything she needs available via Teams. The board can also do its work, and when we collaborate with our partners, we can organize joint projects there."
1.2. Digital administration	"We now have a membership management system, which is also a program where you can fill in all the member data, including account data and so on. We can now also see online who is affected."
1.3. Digital health education	"I know from the PO that they had many offers, especially during the pandemic, where doctors or therapists gave a short lecture on some topic, and where you had the opportunity to submit questions in advance, which were then discussed via the computer."
1.4. Digital health research	"I think we are opening up many things with the register that can go more in the direction of digitalization of care.... So we actually developed this as a hybrid of outpatient clinic, software and register software. This means that the doctor can see everything you enter directly. And also graphically... it looks a bit as if it had at least once swum past Apple."
PO's digital (collaboration) actors	
2.1. POs as independent actors	"This means that I drive forward all digitalization projects within the organization, for example the connection of a new CM system, which was a very large project, or the relaunch of a new website.... I work full-time in our organization and collaborate with many volunteers in the digitalization department."
2.2. Support from volunteers and external partners	"Our homepage is now maintained voluntarily by a member. Whether that's good or bad, I'll leave that for now. Our Instagram is also managed by a member. Neither of them are on the board. The fact that [it] exists at all was also [their own initiative]."
2.3. Public and private financial assistance	"The pharmaceutical company even approached us and offered us this sponsoring membership, and for us it is simply a blessing that we can cover our fixed costs... with it."
Involvement of PO members	
3.1. Motivations to establish and sustain involvement	"It is actually the case that digitalization processes should actually be driven from the bottom up."
3.2. Types and intensity of involvement	"... It's not always easy, there's a lot of tokenism at both national and international level that you involve patients because there may be funding criteria that require it or because it looks good. And that they are not actually taken seriously.... Of course, we are also involved from time to time when it comes to consultation or consultative processes...."
3.3. Requirements	"So what can ultimately diminish the motivation to get involved is if you keep making suggestions that are ultimately not taken into account. I mean, there can certainly be good reasons for this, but if, yes, if you have the feeling that nothing is being taken on board, then I think that would probably have a negative effect."
Attitudes toward digitalization	
4.1. Accelerating management, outreach, and communication	"Young families have also joined us in the meantime.... Of course, that is also a great advantage of digitalization. You can simply put information that is very important online immediately."
4.2. Simplifying access to PO support services	"I thought that was great because it's also a good opportunity for people who don't have the opportunity to travel that far to go somewhere to take part. People who were previously left out somewhere."
4.3. Supporting health research	"Yes, the main reason is because I find it useful for myself and for my treatment, but of course I would like to contribute to making this kind of research possible."
4.4. Ineffective and resource-intensive activities	"[If you introduce digital applications] in healthcare, then... they have to generate added value or have a higher benefit than what has been realized in the previous analog processes.... And this has been almost nonexistent up to now, apart from perhaps the possibility of booking an appointment with a doctor from your computer at home...."
4.5. Challenging personal interaction and access	"And of course there's always the point that we exclude some people who don't have access to the app. Because they simply can't cope with it or are somehow very skeptical about digital. Or because they don't have the means. Yes, those are all things like that."
4.6. Requiring additional efforts	"Financing is a big issue. So where do I get the financing, who takes care of it and so on, I'm busy with all my hands keeping my organization running, I can't write to any foundations, so I say there is money but you have to... know where to get it and... we don't have the manpower at the moment."

XSL•FO
RenderX

Main theme and subtheme	Example quotes
4.7. Encouraging ambivalence	"Well, I think it's a development that probably can't be stopped, because it's happening in many areas now. So as I said, it depends. Sometimes it's helpful. Sometimes, I think, you can do without it. And my big concern is data security."

[a]PO: patient organization.

The aforementioned tools are used by almost all POs, regardless of their size and other digital activities, although not every PO uses all of them combined. In addition, some POs have developed or are developing their own apps or platforms to further support these communication activities. Social media platforms, such as Facebook and Instagram (Meta Platforms, Inc), as well as YouTube (Google LLC) allow POs, for example, to reach a younger and wider audience and disseminate information about the various health care or disease topics they address. While Facebook and Instagram appeared to be relatively established ways of external communication, our interviewees mentioned TikTok (ByteDance) almost never (1/46, 2%), despite the platform's growing reach. PO members used social media platforms to communicate among themselves. However, 1 (2%) interviewee indicated that their organization does not allow WhatsApp (Meta Platforms, Inc) as a formal channel of communication due to considerations regarding data security. Rather, it is used by members on their own initiative and responsibility. Furthermore, POs have established digital or hybrid meetings to supplement or even replace in-person meetings, overcoming contact restrictions as a consequence of the COVID-19 pandemic, accommodating the disabilities of PO members, and appealing to a wider audience from distinct, that is, more distant places (13/46, 28%).

Subtheme 1.2: Digital Administration

Some (9/46, 20%) POs used digital technologies to adapt their administrative and management work (subtheme 1.2 in Table 3). This included managing membership databases digitally, handling bookkeeping via digital systems, generating digital annual financial reports, providing members with digital versions of annual reports, and using collaborative project management software to organize and coordinate their work. Our interview data did not indicate a specific relationship between the use of digital technologies for administrative purposes and the size of the PO. In that sense, digital administration efforts, such as digital membership management, were also found in smaller POs, even if only in a few cases.

Subtheme 1.3: Digital Health Education

The increasingly digitalized world entails a new responsibility of POs as digital health educators (22/46, 48%; subtheme 1.3 in Table 3). As repeatedly stated by our interviewees, many PO members had difficulties in effectively searching for and applying health information or using available digital tools to their advantage. Thus, POs followed at least 4 distinct ways to educate their members in the use of digital technologies as part of their goal to support disease management, improve their digital health literacy, and empower them to take part in the digitalized world. First, POs provided digital health information, primarily via their website (11/46, 24%) but also via an app (4/46, 9%), podcast (2/46, 4%), or in an audiovisual format (2/46, 4%), for instance, via YouTube. Second, they organized digital meetings, events, and conferences to inform members about and engage them in the discussions of recent health care topics, particularly those specific to the disease that a respective PO dealt with (7/46, 15%). Third, POs sometimes offered digital training programs and support groups for their members to help them cope with a disease, for instance, through physical exercises and self-management (5/46, 11%). Fourth, they provided education and training on the use of various digital tools as such to enable them to benefit from such services (8/46, 17%). The overall variety of educational services offered to members tended to vary with the size of the PO, with larger POs generally offering a greater range of services.

Subtheme 1.4: Digital Health Research

Besides providing digital communication and health education as well as digitalizing their administrative work, some POs have started to develop digital technologies, such as mobile health apps (11/46, 24%) and data registries (12/46, 26%; subtheme 1.4 in Table 3). POs did so either on their own or in cooperation with other stakeholders, such as public research institutes and private companies. Through data registries, POs collected health data from their members to facilitate medical and health care services research. Here, they either forwarded such individual health data to research institutions for further investigation or established mutual research cooperations to work together on the data. Remarkably, some of them also applied these registries to conduct their own research projects, that is, without including an external research partner. The management of a data registry appeared to be related to whether the POs had a specific interest in engaging in research rather than their size and resources.

RQ 2: Actors Involved in Planning and Implementing POs' Digital Activities

Subtheme 2.1: POs as Independent Actors

When sufficient personnel and financial resources were available, POs advanced their digital transformation independently by permanently employed staff, that is, individuals with either professional, technical, or media expertise (9/46, 20%). These individuals employed by POs were primarily responsible for planning and implementing activities such as digital projects, services, and strategies (subtheme 2.1 in Table 3). Furthermore, it happened that individual organizations relied on regional or national associations of POs (ie, umbrella organizations) that developed technologies for them (9/46, 20%). For example, the German Rheumatism Association developed an app that each partner of that umbrella organization on the regional level could make use of.

Subtheme 2.2: Support From Volunteers and External Partners

Owing to a lack of expertise within the POs to develop digital services themselves, interviewees frequently mentioned that

XSL•FO
RenderX

they established collaborations with volunteers among their members (13/46, 28%) and external partners (10/46, 22%). Within their own ranks, POs relied on members who, based on their professional background or private interest, had expertise in digital service development and delivery (subtheme 2.2). This support ranged from assisting with short-term tasks to taking full responsibility for certain activities, such as managing the PO's social media channels. Involving members in this direct way rather seemed to be done by smaller POs, with larger POs relying on their regional or national associations or external partners. External partners are needed to develop digital services that are specific to the PO (refer to the Subtheme 2.3: Public and Private Financial Assistance section and Table 3). Examples include the development of an app by which the members can communicate with each other, the design of a new PO website, and the provision of technical infrastructure, for instance, to establish patient registries. Furthermore, external partners are involved in developing health information for the POs' digital formats. Health care professionals, mainly physicians, are asked to provide feedback as an advisory panel.

Subtheme 2.3: Public and Private Financial Assistance

According to some (6/46, 13%) of our interviewees, POs struggled to seek and maintain appropriate financial resources for planning and implementing various activities, such as digital services and projects. In their search for financial support, they applied for external funding from health care insurance and political actors (4/46, 9%). Furthermore, some (4/46, 9%) POs sought funding from pharmaceutical companies (subtheme 2.3 in Table 3).

RQ 3: Involvement of PO Members in Digital Projects

Subtheme 3.1: Motivations to Establish and Sustain Involvement

In general, PO members repeatedly indicated a willingness to be involved in their POs' digital initiatives (10/46, 22%; subtheme 3.1 in Table 3). They cited the benefits of the digital technologies being developed and the feeling that their involvement is making a difference as the main reasons for their involvement. PO representatives and members (13/46, 28%) stated that member involvement is necessary to consider users' (ie, members') needs and preferences regarding digital products. Despite the many positive perceptions and willingness vis-à-vis involvement, interviewees rarely referred to concrete examples of established participatory approaches, indicating a gap between aims or willingness and practice. For instance, only 3 (7%) out of 46 PO members mentioned participatory activities within their POs' digital initiatives, such as members surveys or involving members in the design of a PO website. Regarding a PO's size, interviewees from larger POs more often referred to the importance of and interest in involvement.

Subtheme 3.2: Types and Intensity of Involvement

PO members and representatives indicated rather unanimously that active involvement should take place early in the planning and implementation of the products. However, in contrast, they highlighted several times that currently, member involvement stays rather passive (13/46, 28%), for example, by participating in a member survey (4/46, 9%) or user testing (2/46, 4%). Some (2/46, 4%) representatives mentioned that they do offer members the opportunity to contribute more actively, for example, by serving on an advisory board for the development of a patient registry or on a research committee (subtheme 3.2 in Table 3). Unfortunately, members (7/46, 15%) indicated that they were not aware of such opportunities to become involved. This meshes with representatives, who said that members involved in such digitalization efforts tend to be those who already volunteer within their PO. Moreover, we found more examples of actual involvement practices from interviewees belonging to larger POs.

Subtheme 3.3: Requirements for Involvement

Because of this struggle to achieve "true," that is, active and ongoing involvement, some (5/46, 11%) interviewees pointed out the need to improve the prerequisites and conditions for this. To achieve this, PO members repeatedly emphasized that digital projects should be set up in such ways that involvement is neither too time consuming nor too demanding, for example, in terms of specific professional or technical knowledge (13/46, 28%). This was considered crucial for the motivation of members to get involved (subtheme 3.3 in Table 3). The observation that members can bring specific expertise to digitalization projects, given their professional backgrounds, adds to the importance of involving them.

RQ 4: Attitudes Toward Digitalization in POs

Overview

The impact of the digital activities undertaken by POs can be multiple. These impacts can occur at the individual level by enabling a broader group of members to participate in POs' activities and services (eg, hybrid activities), at the organizational level by streamlining their operations (eg, digital membership system), at the research level by providing better integration into research (eg, digital patient registry), or at the political level by offering better representation in the health care system. In this section, we describe how our participants generally viewed the shift toward more digital communication and the increasing use of digital tools positively. Positive impacts due to digitalization seemed to be expressed more often by interviewees from larger POs, while negative aspects, that is, challenges and disadvantages, were equally expressed by interviewees from smaller and larger POs. However, certain aspects elicited mixed responses, including criticisms regarding how POs are digitally transforming their work.

Subtheme 4.1: Accelerating Management, Outreach, and Communication

Participants (14/46, 30%) mentioned several examples of successful digital initiatives. For instance, POs' websites and, sometimes, social media channels have become key outreach tools to reach a wider audience and attract new members, especially younger people. Moreover, patient registries have been established to support research efforts, and during the COVID-19 pandemic, digital formats for face-to-face interaction and exchange were introduced, further enhancing communication. Furthermore, several (7/46, 15%) representatives highlighted how digital tools help to handle daily or regular tasks related to administration and management.

XSL•FO
RenderX

Participants also mentioned the ability to network at a national or even European level through digital tools, for example by connecting with foreign sister organizations and research institutes that the PO had not previously worked with. Moreover, some (2/46, 4%) mentioned that costs can be reduced by moving from analog to digital formats.

Subtheme 4.2: Simplifying Access to PO Support Services

Many (30/46, 65%) participants described digital tools as helpful and convenient. PO representatives saw videoconferencing software as particularly valuable for simplifying communication or facilitating events. Representatives and members alike appreciated the opportunity to improve access for those who cannot attend PO meetings or events in person, for example, due to a physical impairment.

Subtheme 4.3: Supporting Health Research for People With Chronic and Rare Diseases

Patient registries operated by POs were seen as particularly beneficial and promising by several (11/46, 24%) participants. Registries were seen to advance health research and the development of new treatments. Valuing the potential benefits for themselves and others affected, several (10/46, 22%) PO members indicated that they would generally be willing to provide their data for use in such a patient registry.

Subtheme 4.4: Ineffective, Costly, and Time-Consuming Activities

Despite these various positive aspects, especially PO representatives (8/46, 17%) reported it as costly, requiring significant personnel and financial resources and expertise to implement the projects. This included, but was not limited to, large-scale projects, such as patient registries and mobile apps, for which these resources were not readily available. These challenges were seen as particularly significant for smaller, resource-limited POs. Some PO representatives referred to unsuccessful or failed digital projects (3/46, 7%), although failures were less frequently reported compared to successes (14/46, 30%). For example, a participant noted that the PO's social media channels were not effective enough in reaching a wider audience. Another mentioned a chat forum that was discontinued after a while due to lack of use. The last example is the discontinuance of a patient registry at a smaller PO due to a lack of resources and expertise.

Subtheme 4.5: Challenging Personal Interaction and Access

Representatives and members alike stated that the value of in-person meetings cannot be fully captured by digital alternatives (14/46, 30%). Despite perceived advantages, digital technologies were mainly seen as supplements rather than replacements. In addition, several participants (18/46, 39%) stated that digital formats are less accessible to older PO members or those with medical conditions impairing cognition. Participants stated that these persons may lack access to important information. They lack the literacy to work with digital tools, have physical difficulties using digital equipment, or have concerns regarding the use of digital technologies in general.

Subtheme 4.6: Requiring Additional Efforts for Planning, Training, and Risk Management

Both representatives and members emphasized that POs need to make additional efforts to succeed in their digital transformation (7/46, 15%), which is particularly difficult to achieve for smaller POs (5/46, 11%). Several (15/46, 33%) participants stressed the importance of training members to use digital technologies effectively. Furthermore, they emphasized the importance of digital services meeting members' preferences, providing clear benefits, and being user-friendly. Another key issue our participants brought up is data privacy, for example, of the data stored within patient registries or gathered by the use of digital tools, such as mobile apps. In this study, the participants considered comprehensive privacy policies important and necessary.

Subtheme 4.7: Encouraging Ambivalence or Fostering Undecidedness

Several (11/46, 24%) participants expressed difficulties evaluating the consequences of digital technologies. They were generally uncertain, weighing benefits against potential risks. This was exacerbated, as the digital tools had often only recently been implemented. Others acknowledged that while digital workflows and communication might be feasible and useful in POs, they involved a significant amount of work and financial resources, which were scarce for POs. Furthermore, some (7/46, 15%) representatives pointed out that while digital transformation within the organization was generally positive, this might not be the case for everyone involved in the process. Specifically, they perceived older PO members as being more reluctant to embrace digitalization in the PO, resulting in low motivation levels, skepticism, or even resistance. Moreover, some (5/46, 11%) indicated that the benefits would eventually outweigh the potential drawbacks.

Discussion

Principal Findings

This interdisciplinary interview study explored how German POs enact their digital transformation (RQ 1), who is involved in this process (RQ 2 and RQ 3), and how PO representatives and members evaluate past and current efforts (RQ 4). This study aimed to provide empirical evidence on the *ways of working* of one of the most important institutional actors in health care and research. While digital transformation is seen as promising and potentially transformative for the delivery of care and research, practical insights into this process, its requirements, and the challenges for POs are largely lacking.

First, POs are driving digitalization primarily in 4 key areas: communication, health education, health research, and administration. Communication efforts are the most widespread, with many POs having established *basic* digital formats, such as websites, social media channels, and videoconferencing. Digital health education also plays a central role as POs work to improve members' digital health literacy and provide accessible information. Furthermore, some POs are engaged in health research through digital data collection and have begun to digitalize administrative tasks, although this is comparatively

XSL•FO
RenderX

less common. Second, POs rely heavily on internal support and external collaboration to develop and sustain these digital activities. The participants in our study often referred to volunteer members and collaborations with external public and private partners to manage the necessary workload and resources. Third, while our interviewees considered the active involvement of members, for example, in the design of digital tools, desirable, it is not a *routine* in POs and depends much on the sufficient motivation of those involved. Fourth, the digital transformation in health care comes with a new task for POs: improving the digital health literacy of their members. Finally, PO representatives and members highlighted that digital transformation brings hope to reach a wider audience and attract younger people. However, its advantages may only be realized if several barriers can be overcome. These include, most obviously, a lack of financial resources, the need for digital training for staff and volunteers, and appropriately handling the reduction of face-to-face communication.

Roles and Tasks of POs

Until now, there has been limited research on the roles and tasks of POs in general and especially in relation to digital transformation. Our findings contribute to the evidence base in several ways. For example, according to van de Bovenkamp et al [10], POs seek and depend on collaboration with external stakeholders for a variety of reasons, not least to secure material and immaterial resources, which are often limited, especially in smaller organizations [15]. However, our PO representatives and members were somewhat ambivalent about this fact. On the one hand, POs collaborate in the planning and implementation of digital activities relatively intensively, involving many volunteers (ie, PO members) and, less frequently, sister organizations, such as the national and regional associations. On the other hand, collaboration with external partners, such as pharmaceutical companies or health insurance companies, was often described as hampered. It was perceived that it does not serve *true* collaboration purposes, such as jointly developing a new digital service. Rather, collaboration is meant to secure needed resources or outsource work to external partners given their limited capabilities and skills. However, according to van de Bovenkamp et al [10], POs need to foster actual collaboration to promote their role in the health care system.

While Claus et al [11] state that POs could help to recruit patients for research studies, we found little evidence that POs are indeed using their digital resources for this purpose [12]. Our findings show that POs are using a wide variety of more basic digital tools, mainly for communication purposes. Those tools include websites, social media, and digital events resulting from the POs' internal digitalization efforts. In comparison, the development and use of more elaborate digital tools, such as mobile apps or patient data registries, is less common. A lack of resources, both in terms of finances and skill set, is an important reason for this. Considering the relative novelty of the digital tools and the lack of experience and expertise within the POs to use these tools, a robust digital governance framework is necessary to further advance the digital transformation in POs [3]. Such a governance framework would serve the POs as a guideline in the development and maintenance of their digital activities and services to ensure compliance with their core aims and values.

Although we did not directly assess the impact of these digital activities, participants' experiences and attitudes suggest that digital transformation has a range of effects. For instance, digital tools were perceived as essential for outreach and communication; helpful in facilitating daily administrative tasks; and, in some cases, even reducing costs. Some (11/46, 24%) participants also mentioned that tools, such as videoconferencing, provide better access to support services for members with physical disabilities. In contrast, digital activities were seen as requiring significant financial and human resources, and certain individuals, especially older individuals or those with lower digital literacy, were reported to face barriers in accessing digital tools, potentially leading to exclusion. Hence, the digital transformation in POs, as reported by participants, has a broad impact, including both positive and negative aspects.

Involvement of PO Members

Our findings highlight a consensus among many PO members and representatives regarding the value of participatory approaches to enhance the usability of digital services. This is consistent with the reported general benefits of stakeholder engagement in the development of digital health tools, highlighting the critical role of this approach [26-28].

The interviews revealed a strong willingness among members to engage in digital initiatives, while relatively few (3/46, 7%) reported actual participatory approaches. Some studies, although not specific to the PO context, point to organizational capacity constraints, including time and financial resources, which hinder a more thorough use of participatory approaches [28,29]. Further examining such challenges in the context of POs can provide valuable insights into bridging the gap between intentions to get involved and actual involvement.

To tap into the full range of perspectives and skills of PO's members, organizations need to design involvement methods, as our interview data point to the design of such approaches as a potential pivot point for increasing involvement. For example, participants emphasized the importance of clear benefits, manageable requirements, and early engagement as key factors in their decisions to engage. POs may consider these aspects when planning participatory approaches, as they have also been identified as key participatory design principles for stakeholder engagement in other digital health projects [30,31].

Furthermore, participatory approaches used in POs' digital initiatives seem to focus predominantly on the initial stages of engagement, such as surveys and user testing. The prevalence of these methods in the PO landscape may be due to their relative organizational simplicity and lower demands on participants. More active, collaborative forms, such as advisory boards, were less frequently mentioned. This echoes a trend found in 2 reviews, which identified such collaborations in the participatory development of digital health tools but comparatively more frequently reported forms of involvement that tended toward more passive engagement [29,31]. Regarding the PO context, this may prompt further consideration of the

XSL•FO
RenderX

applicability of active involvement or the potential reluctance of members to take on more control and responsibility.

Building on our findings, future research should explore the willingness of PO members to engage in their PO projects, their specific needs and expectations, and the organizational capacity to support and sustain active involvement. Such studies could guide the design and implementation of participatory approaches in POs' digital initiatives or even establish frameworks to align member expectations with organizational realities.

Ethical Aspects Related to Digital Transformation Within POs

Our findings highlight the ethical complexities that come with the digital transformation. POs are confronted with a wide range of ethical considerations regarding their digital projects and services as well as the processes of developing and maintaining them. Topics such as digital literacy, accessibility of digital tools, and data security [32-34] are highly relevant to the POs' digital transformation processes.

Our findings show the importance of accessibility to the digital services POs provide for their members. Interestingly, digitalizing services can lead both to enhanced and diminished accessibility of the services in question. For an inclusionary approach, POs need to identify the respective groups that will and will not most likely benefit from the digital transformation. One group that was seen as potentially disadvantaged was older members because they might lose the connection to the activities of the PO. This may be due to a lack of resources, insufficient digital literacy, or preferences for analog formats. However, the use of digital tools may also increase the accessibility of PO services. For example, one group for which digital tools increase accessibility is persons unable to attend in-person meetings due to their ailment or disability. Therefore, there is ambivalence concerning accessibility, meaning that POs have to carefully weigh their options and try to forge a path that includes all their members.

Furthermore, a just allocation of scarce resources and collaboration with external partners were considered important. Previous research has found that POs rely heavily on third-party funding; however, there is a lack of transparent information about potential conflicts of interest [18]. POs are challenged with the task of securing enough resources to allow them to develop and maintain their digital activities while remaining independent of undue influences. Their political and economic independence is especially important to maintain integrity and trustworthiness. In this regard, careful consideration is needed when choosing an external partner to safeguard the PO's independence. However, such considerations might be hampered by a lack of choices.

Persons in POs contributing to digitalization projects may easily find themselves in attitudinal and judgmental conflicts when weighing the positive and negative sides of those projects. This stresses the need for an ethical governance framework for POs' digital transformation.

Strengths and Limitations

We were able to include participants from 19 individual POs, which represent a range of target groups, address various aspects in the field of health care and medical issues, and are based in different regions of Germany. By including POs from distinct places, we aimed to reduce potential bias from a concentration of organizations in certain regions and provide a more balanced view regarding digitalization. We believe that the diversity of the included POs may have contributed to a broader range of perspectives on POs' digital transformation while allowing us to clearly identify the most prominent practices, actors, and attitudes. However, as digitalization activities vary considerably between POs, especially regarding their digital educational services, the development of their own technologies, and research activities, there could be a greater differentiation of these digitalization activities and the actors involved depending on the size of a PO, the topics dealt with, and the level of digitalization within each organization. In addition, the observed variations in digital activities are based on subjective reports and may not capture all initiatives. Furthermore, because we included only POs that demonstrated some level of digital activity, the perspectives of those still at the very beginning of their digital transformation or those that have chosen not to engage in this process are not represented in our study; therefore, it may be assumed that for such POs, the extent of digitalization is lower. Given our limited personnel resources, we did not return interview transcripts and the eventual findings to the participants for comments. However, additional steps to increase the validity of the findings were performed, such as peer checks of the coded text segments and peer interpretation of the findings (refer to the Methods section for more validating steps). Our study can be a starting point for further such analyses.

Conclusion

Our study provides insights into the state of digital transformation in German POs. We found that digitalization efforts are particularly evident in 4 core areas: communication, administration, health education, and health research. These processes involve multiple professional and nonprofessional actors, including permanent staff, volunteers, and external partners. Member involvement varies in scope and form, with a high level of willingness to engage. While representatives and members generally view digitalization positively, they also identify barriers, such as resource limitations and accessibility challenges, for certain groups.

In conclusion, German POs are currently investing considerable resources to engage in digital transformation processes, ranging from smaller projects to more sophisticated initiatives, and involving a wide range of stakeholders and individuals. These efforts and collaborations highlight the need for POs to establish more comprehensive ethical governance frameworks to clarify the goals of digitalization, determine what is needed to achieve them, and define how to engage with different stakeholders, including setting clear roles and motivations for their members to become involved. Public institutions could assist POs with these multiple tasks and requirements, for example, by providing training, mentorship, and facilitating networking and sharing among POs.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all of the interviewees for their time and dedication to participating in this study. The authors are grateful for the feedback during the planning and revision of this study provided by their patient advisory board: Miriam Schlangen, Thomas Duda, Wiebke Papenthin, Thorsten Freikamp, and Heidemarie Haase. They would also like to thank the German Federal Ministry of Education and Research for funding the Pandora research project (grant 01GP2115A-C).

Authors' Contributions

S Wallraf, SK, HJVGS, and JL planned the study and collected and analyzed the data. S Wallraf and JL wrote the first draft. SK and HJVGS provided extensive feedback and revised the draft. CW, S Wohlke, MLD, and MAS provided critical input for the draft. All authors read the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 310 KB - jmir_v27i1e62750_app1.pdf]

Multimedia Appendix 2
Overview of interview and focus group questions.
[PDF File (Adobe PDF File), 117 KB - jmir_v27i1e62750_app2.pdf]

References

1. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407 [FREE Full text] [doi: 10.3390/ijerph20043407] [Medline: 36834105]
2. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
3. Digital transformation: shaping the future of European healthcare. Deloitte Centre for Health Solutions. 2020 Sep. URL: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-shaping-the-future-of-european-healthcare.pdf [accessed 2025-01-20]
4. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
5. Burmann A, Tischler M, Faßbach M, Schneitler S, Meister S. The role of physicians in digitalizing health care provision: web-based survey study. JMIR Med Inform 2021 Nov 11;9(11):e31527 [FREE Full text] [doi: 10.2196/31527] [Medline: 34545813]
6. Brown EE. Assessing the quality and reliability of COVID-19 information on patient organization websites. Front Commun 2021 Aug 2;6. [doi: 10.3389/fcomm.2021.716683]
7. Pauer F, Litzkendorf S, Göbel J, Storf H, Zeidler J, Graf von der Schulenburg JM. Rare diseases on the internet: an assessment of the quality of online information. J Med Internet Res 2017 Jan 18;19(1):e23 [FREE Full text] [doi: 10.2196/jmir.7056] [Medline: 28100442]
8. van de Bovenkamp HM, Trappenburg MJ, Grit KJ. Patient participation in collective healthcare decision making: the Dutch model. Health Expect 2010 Mar 17;13(1):73-85 [FREE Full text] [doi: 10.1111/j.1369-7625.2009.00567.x] [Medline: 19719537]
9. Souliotis K, Agapidaki E, Peppou LE, Tzavara C, Varvaras D, Buonomo OC, et al. Assessing patient organization participation in health policy: a comparative study in France and Italy. Int J Health Policy Manag 2018 Jan 01;7(1):48-58 [FREE Full text] [doi: 10.15171/ijhpm.2017.44] [Medline: 29325402]
10. van de Bovenkamp H, de Graaff B, Kalthoff K, Bal R. The patient representation struggle during the COVID-19 pandemic: missed opportunities for resilient healthcare systems. Health Expect 2024 Feb;27(1):e13877 [FREE Full text] [doi: 10.1111/hex.13877] [Medline: 37814486]
11. Claus EB, Feliciano J, Benz LS, Calvocoressi L. Social media partnerships with patient organizations for neuro-oncology patient recruitment. Neurooncol Pract 2020 Mar;7(2):143-151 [FREE Full text] [doi: 10.1093/nop/npz049] [Medline: 32626583]
12. Gentilini A, Miraldo M. The role of patient organisations in research and development: evidence from rare diseases. Soc Sci Med 2023 Dec;338:116332 [FREE Full text] [doi: 10.1016/j.socscimed.2023.116332] [Medline: 37866173]

XSL•FO
RenderX

13. Mitchell D, Geissler J, Parry-Jones A, Keulen H, Schmitt DC, Vavassori R, et al. Biobanking from the patient perspective. Res Involv Engagem 2015;1:4 [FREE Full text] [doi: 10.1186/s40900-015-0001-z] [Medline: 29062493]
14. Rabeharisoa V. The struggle against neuromuscular diseases in France and the emergence of the "partnership model" of patient organisation. Soc Sci Med 2003 Dec;57(11):2127-2136. [doi: 10.1016/s0277-9536(03)00084-4] [Medline: 14512243]
15. Sienkiewicz D, van Lingen C. The added value of patient organisations. European Patients Forum. 2017 Nov. URL: https://www.eu-patient.eu/globalassets/library/publications/epf_added_value_report_final.pdf [accessed 2025-01-20]
16. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023 Dec;2:100171 [FREE Full text] [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
17. Nguyen CQ, Kariyawasam D, Alba-Concepcion K, Grattan S, Hetherington K, Wakefield CE, et al. 'Advocacy groups are the connectors': experiences and contributions of rare disease patient organization leaders in advanced neurotherapeutics. Health Expect 2022 Dec;25(6):3175-3191 [FREE Full text] [doi: 10.1111/hex.13625] [Medline: 36307981]
18. Mikami K, Sturdy S. Patient organization involvement and the challenge of securing access to treatments for rare diseases: report of a policy engagement workshop. Res Involv Engagem 2017;3:14 [FREE Full text] [doi: 10.1186/s40900-017-0065-z] [Medline: 29062539]
19. Patient-centered digitalization: an ethical analysis of the role of patient organizations as actors in the context of digitalization in health-related research and care (PANDORA). Pandora Research Consortium. URL: https://pandora-forscht.de/en/ [accessed 2025-01-22]
20. Patient-centered digitalization: an ethical analysis of the role of patient organizations as actors in the context of digitalization in health-related research and care (PANDORA). PANDORA forscht. URL: https://pandora-forscht.de/en/ [accessed 2024-03-28]
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
22. Rach C, Lukas J, Müller R, Sendler M, Simon P, Salloch S. Involving patient groups in drug research: a systematic review of reasons. Patient Prefer Adherence 2020;14:587-597 [FREE Full text] [doi: 10.2147/PPA.S232499] [Medline: 32210544]
23. Patterson AM, O'Boyle M, VanNoy GE, Dies KA. Emerging roles and opportunities for rare disease patient advocacy groups. Ther Adv Rare Dis 2023;4:26330040231164425 [FREE Full text] [doi: 10.1177/26330040231164425] [Medline: 37197559]
24. Rauter CM, Wöhlke S, Schicktanz S. My data, my choice? - German patient organizations' attitudes towards big data-driven approaches in personalized medicine. An empirical-ethical study. J Med Syst 2021 Feb 22;45(4):43 [FREE Full text] [doi: 10.1007/s10916-020-01702-7] [Medline: 33616768]
25. Kuckartz U. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim, Germany: Beltz Juventa; 2016.
26. Jahnel T, Schüz B. [Participatory development of digital public health: tension between user perspectives and evidence]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb;63(2):153-159. [doi: 10.1007/s00103-019-03082-x] [Medline: 31915866]
27. Hochmuth A, Exner AK, Dockweiler C. [Implementation and participatory design of digital health interventions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb;63(2):145-152. [doi: 10.1007/s00103-019-03079-6] [Medline: 31938837]
28. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
29. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb 22;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
30. Voorheis P, Petch J, Pham Q, Kuluski K. Maximizing the value of patient and public involvement in the digital health co-design process: a qualitative descriptive study with design leaders and patient-public partners. PLOS Digit Health 2023 Oct;2(10):e0000213 [FREE Full text] [doi: 10.1371/journal.pdig.0000213] [Medline: 37878566]
31. Sanz MF, Acha BV, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021 Apr 30;21(2):16 [FREE Full text] [doi: 10.5334/ijic.5573] [Medline: 33981193]
32. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly 2018;148:w14571 [FREE Full text] [doi: 10.4414/smw.2018.14571] [Medline: 29376547]
33. Maeckelberghe E, Zdunek K, Marceglia S, Farsides B, Rigby M. The ethical challenges of personalized digital health. Front Med (Lausanne) 2023;10:1123863 [FREE Full text] [doi: 10.3389/fmed.2023.1123863] [Medline: 37404804]
34. Brall C, Schröder-Bäck P, Maeckelberghe E. Ethical aspects of digital health from a justice point of view. Eur J Public Health 2019 Oct 01;29(Supplement_3):18-22 [FREE Full text] [doi: 10.1093/eurpub/ckz167] [Medline: 31738439]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research

XSL•FO
RenderX

PO: patient organization
RQ: research question

Edited by A Schwartz; submitted 30.05.24; peer-reviewed by V Vyas, U Sinha, V Astha, DA Mertes; comments to author 26.08.24; revised version received 05.11.24; accepted 06.12.24; published 13.02.25.

Please cite as:
Wallraf S, Köthemann S, Wiesemann C, Wöhlke S, Dierks ML, Schmidt MA, van Gils-Schmidt HJ, Lander J
Digital Transformation in Patient Organizations: Interview and Focus Group Study
J Med Internet Res 2025;27:e62750
URL: https://www.jmir.org/2025/1/e62750
doi:10.2196/62750
PMID:39946181

©Simon Wallraf, Sara Köthemann, Claudia Wiesemann, Sabine Wöhlke, Marie-Luise Dierks, Marion Andrea Schmidt, Henk Jasper van Gils-Schmidt, Jonas Lander. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

User Personas for eHealth Regarding the Self-Management of Depressive Symptoms in People Living With HIV: Mixed Methods Study

Ting Zhao[1], MSc; Chulei Tang[2], PhD; Jun Ma[1], MSc; Huang Yan[3], MSc; Xinyi Su[1], MSc; Xueyuan Zhong[1], MSc; Honghong Wang[1], PhD

[1]Xiangya School of Nursing, Central South University, Changsha, China

[2]School of Nursing, Nanjing Medical University, Nanjing, China

[3]Nursing Department, The Third Xiangya Hospital, Central South University, Changsha, China

Corresponding Author:
Honghong Wang, PhD
Xiangya School of Nursing
Central South University
Yuelu District
Changsha, 410013
China
Phone: 86 731 89665663
Email: honghong_wang@hotmail.com

Abstract

Background: eHealth has enormous potential to support the self-management of depressive symptoms in people living with HIV. However, a lack of personalization is an important barrier to user engagement with eHealth. According to goal-directed design, personalized eHealth requires the identification of user personas before concrete design to understand the goals and needs of different users.

Objective: This study aimed to identify user personas for eHealth regarding the self-management of depressive symptoms in people living with HIV and explore the goals and needs of different user personas for future eHealth.

Methods: We used an explanatory sequential mixed methods design at the First Hospital of Changsha City, Hunan Province, China, from April to October 2022. In the quantitative phase, 572 people living with HIV completed validated questionnaires with questions related to demographics, self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. Latent profile analysis was performed to identify different user personas. In the qualitative phase, 43 one-to-one semistructured interviews across different user personas were conducted, transcribed verbatim, and analyzed using conventional content analysis. The findings from both phases were integrated during the interpretation phase.

Results: Three types of user personas could be identified, including “high-level self-managers” (254/572, 44.4%), “medium-level self-managers” (283/572, 49.5%), and “low-level self-managers” (35/572, 6.1%). High-level self-managers had relatively high levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. High-level self-managers had a positive attitude toward using eHealth for the self-management of depressive symptoms and desired access to self-management support for depressive symptoms from eHealth with high usability. Medium-level self-managers had relatively medium levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. Medium-level self-managers felt burdened by using eHealth for the self-management of depressive symptoms and preferred to access self-management support for HIV from eHealth with privacy. Low-level self-managers had relatively low levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. Low-level self-managers had an acceptable attitude toward using eHealth for the self-management of depressive symptoms and desired access to professional guidance from eHealth with privacy and no cost (“free of charge”).

Conclusions: The 3 user personas shed light on the possibility of personalized eHealth to support the self-management of depressive symptoms in different people living with HIV. Further research is needed to examine the generalizability of the user personas across study sites.

(J Med Internet Res 2025;27:e56289) doi:10.2196/56289

XSL•FO
RenderX

KEYWORDS

HIV; depressive symptoms; depression; self-management; eHealth; personas

Introduction

Background

People living with HIV disproportionately experience depressive symptoms. A previous evidence-based study estimated that the global prevalence rate of depression in people living with HIV was approximately 22% to 44% [1], which usually predicts increased sexual risk behavior, poor adherence to antiretroviral therapy, accelerated disease progression, and increased suicidal ideation [2-5].

Self-management plays a crucial role in preventing and reducing depressive symptoms in people living with HIV, and the processes include learning skills, assessing resources, and coping with the illness [6]. However, although a large number of effective resources are available for the self-management of depressive symptoms in people living with HIV, not many seek or receive these [7,8], mainly due to limited access to care, extra financial or commute burdens, and perceived stigma [9-12].

Fortunately, eHealth offers an easily accessible, flexible, potentially cost-effective, and highly anonymous option for the self-management of depressive symptoms in people living with HIV [12-14]. Since the COVID-19 pandemic, eHealth has been greatly recognized for providing better access to mental health services and is being increasingly used as a critical component of normal mental health practices [15-17]. Several existing systematic reviews have demonstrated the effectiveness of eHealth for depressive symptoms in people living with HIV [6,18,19]. Taken together, eHealth has enormous potential to support the self-management of depressive symptoms in people living with HIV [13,20].

It is worth noting, however, that users generally have various goals and needs regarding the use of eHealth [21]. Users may prefer to have an opportunity to select personalized eHealth that can be tailored to their goals and needs, and yet in practice, they always feel “depersonalized” by eHealth and perceive that developers often fail to understand their needs, which can demotivate them when engaging with such resources or even cause them to give up [22-25]. An emerging systematic review found that a lack of personalization is an important barrier to user engagement with eHealth in the mental health domain [26]. Future eHealth needs to encourage user involvement in the development phase to better understand and meet the goals and needs of various people living with HIV for the self-management of depressive symptoms, thereby promoting the adoption and sustained use of future eHealth [27].

Goal-directed design is a user research–based eHealth design method, which requires identifying user personas before concrete design to understand potential or actual user goals and needs (ie, expectations of an end condition) [28]. In this regard, user personas are detailed composite user archetypes that represent different behavior patterns and the goals and needs associated with them, providing a powerful communication tool for developers to better understand different types of users [28]. Previous evidence has indicated that user persona–tailored eHealth could increase user engagement and satisfaction with online user experience [29,30], which holds promise for the adoption and sustained use of eHealth.

According to goal-directed design, user personas can be hypothesized based on demographic variables, domain expertise, and technical expertise. Demographic variables are the most basic factors to help differentiate different types of users. We included age, gender, education, employment status, monthly household income, and health condition as the specific elements of demographic variables based on a literature review [26,31,32]. Domain expertise refers to user proficiency in a specialized subject area relevant to a product. In the area of the self-management of depressive symptoms, patients are empowered to take responsibility for the day-to-day management of their depressive symptoms [33]. However, effective self-management of depressive symptoms requires an individual to have the ability to monitor depressive symptoms and take actions necessary to maintain a satisfactory quality of life [34]. Meanwhile, self-efficacy, which refers to the belief in one’s ability to take specific actions to achieve desired outcomes [35], is closely related to effective self-management of depressive symptoms because self-efficacy, as an important motivational construct, can affect one’s choice of activities, effort, persistence, and achievement in the self-management of depressive symptoms [36]. In other words, both the self-management abilities of depressive symptoms and self-efficacy are indispensable for the desired performance in domain expertise [36,37]. We thus proposed these 2 elements as the specific elements of domain expertise. Lastly, in the eHealth area, technical expertise refers to user proficiency in eHealth technology. We thus defined technical expertise as the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem, namely eHealth literacy [38].

Objectives

To inform the design of future eHealth, this study aimed to identify user personas for eHealth regarding the self-management of depressive symptoms in people living with HIV and explore the goals and needs of different user personas for future eHealth. We propose the following hypotheses:

- Hypothesis 1: There will be different user personas for eHealth regarding the self-management of depressive symptoms in people living with HIV.
- Hypothesis 2: Different user personas will reflect different people living with HIV, their characteristics, and their goals and needs for future eHealth.

Methods

Study Design

We undertook an explanatory sequential mixed methods study [39]. First, we conducted a quantitative survey to provide a preliminary understanding of user personas for eHealth

XSL•FO
RenderX

regarding the self-management of depressive symptoms in people living with HIV. Then, we collected qualitative data through semistructured interviews to explain these user personas in more depth and explore the goals and needs of different user personas for future eHealth. Lastly, the quantitative and qualitative findings were integrated during the interpretation phase.

This study has been reported in line with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cross-sectional studies [40], the SRQR (Standards for Reporting Qualitative Research) checklist [41], and the GRAMMS (Good Reporting of a Mixed Methods Study) checklist [42].

Quantitative Phase

Setting and Sampling

We recruited a consecutive sample of people living with HIV on arrival at the HIV clinic of the First Hospital of Changsha City, Hunan Province, China, from April 2022 to October 2022. Hunan Province is located in south-central China, with a population of 53,030 people living with HIV at the end of October 2022, similar to the national-level prevalence [43]. The First Hospital of Changsha City is a large designated tertiary hospital for HIV diagnosis and treatment in Hunan Province, which follows about 9000 HIV/AIDS patients, with an annual outpatient volume of about 30,000 visits [44]. Individuals were eligible if they met the following criteria: (1) were aged 18 years or older, (2) had been diagnosed with HIV infection (as confirmed by a diagnosis report), and (3) voluntarily participated in this survey after providing informed consent. Individuals were excluded if they had cognitive or audiovisual impairment assessed by clinicians. We consecutively screened people living with HIV for eligibility to participate in this survey until a sufficient sample size was achieved. Meanwhile, some potential participants were referred by medical social workers. Overall, a total of 600 participants were involved in the survey, and 592 returned completed surveys. A final sample of 572 participants was included for analysis after removing responses with over 10% missing data.

Variables and Measures

Demographic variables included age (years), gender (male=0, female=1), educational level (high school or below=0, higher education or above=1), employment status (unemployed=0, employed=1), monthly household income (<10,000 RMB=0, ≥10,000 RMB=1; a currency exchange rate of RMB 1=US $0.1375 is applicable), and health condition (including comorbidities [no=0, yes=1], latest CD4 count [≥200 cells/mm^3=0, <200 cells/mm^3=1], viral load [not "target not detected (TND)" status=0, TND status=1], and severity of depressive symptoms), with these categories determined by reference to previous literature [26,45-48]. We used the 9-item Patient Health Questionnaire (PHQ-9) to assess the severity of depressive symptoms in the past 2 weeks [49], with a total score range from 0 to 27. Scores of 5, 10, 15, and 20 represent mild, moderate, moderately severe, and severe depression, respectively.

The self-management abilities of depressive symptoms were assessed using the 9-item Chinese version of the Depression-Specific Self-Management Questionnaire (DSSM; a 5-point Likert scale), which has a total score ranging from 9 to 45, with higher scores indicating higher levels of self-management abilities of depressive symptoms [50]. Self-efficacy was assessed using the Self-efficacy for Managing Chronic Disease Scale (SEMCDS; a 10-point scale) [51]. The scale has 6 items, and the total score can range from 6 to 60, with higher scores indicating higher levels of self-efficacy. eHealth literacy was assessed using the 8-item Chinese version of the eHealth Literacy Scale (eHEALS; a 5-point Likert scale), which has a total score ranging from 8 to 40, with higher scores indicating higher levels of eHealth literacy [52,53]. All Cronbach α values were above the cutoff value of .70, indicating good internal consistency [54].

Data Collection

Before the formal survey, we pretested the questionnaire and adjusted the questionnaire wording as necessary to ensure comprehensibility and readability. We trained 4 investigators (graduate nursing students) to distribute and collect questionnaires using a traditional pen-and-paper method. Each participant completed the questionnaire themselves after the investigator demonstrated how to fill in responses. For those who could not read or write, the investigator explained the meaning of each item and completed the questionnaire according to the participant's responses. At the end of the survey, participants were invited to participate in a follow-up qualitative interview. Those who expressed an interest were asked to leave their WeChat contact details.

Data Analysis

All returned questionnaires were double-entered in EpiData software (version 3.1; EpiData Association) to ensure accuracy and integrity. Statistical analysis consisted of 3 parts. First, we conducted descriptive statistics to understand the characteristics of the participants.

Second, we used latent profile analysis (LPA) to identify the latent profiles of user personas among people living with HIV based on responses to the following indicators: the self-management abilities of depressive symptoms, self-efficacy, and eHealth literacy. Participants with similar response patterns were classified into the same profile. To ensure adequate accuracy of the classification, the recommended minimum sample size for LPA studies is at least 500 [55,56], and therefore, the current size of 572 is sufficient.

We used a maximum likelihood estimation with robust standard errors (MLR) to identify the participants' latent profiles [57]. To avoid local maxima, we used 7000 random sets of starting values for the initial stage, 500 iterations for each random start, and the 200 best solutions retained for final-stage optimizations. To identify the optimal number of latent profiles, we evaluated 1- to 5-profile models based on the statistical fit, parsimony, and substantive interpretability [58]. We considered a combination of fit indices, including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), sample size–adjusted BIC (saBIC), and entropy [57]. The model with

XSL•FO
RenderX

lower AIC, BIC, and saBIC values had a better fit, while the model with an entropy value closer to 1 had a higher classification accuracy. We also used the adjusted Lo-Mendel-Rubin likelihood ratio test (aLMRT) and bootstrap likelihood ratio test (BLRT) to evaluate model fit [57]. A statistically significant aLMRT or BLRT value indicates that the model with *k* profiles outperforms that with *k-1* profiles. In addition, the percentage of individuals in the smallest profile was set at 5% of the total sample to include at least sufficient individuals (n=30-60) to support the generalizability of the LPA [59,60].

Third, we used the automatic 3-step approach to model the covariates after identifying the optimal number of latent profiles. Specifically, we used the R3STEP command, which assigns individuals to their most likely profiles, to conduct a series of multinomial logistic regressions to examine the effects of covariates on the profile membership [61,62]. This procedure is the recommended method as it accounts for the measurement error in profile classification while modeling the effects of covariates on profile membership [60,61]. All demographic variables were introduced as covariates. In the R3STEP procedure, we used multiple imputation to impute missing values for covariates to minimize bias. A total of 27 imputations were used according to a rule of thumb (ie, the number of imputations should be at least equal to the percentage of incomplete cases) [63]. All statistical analyses were conducted in Mplus software (version 8.4) [64], and a *P* value of <.05 was considered statistically significant.

Qualitative Phase

Sampling

The qualitative phase was conducted using a qualitative descriptive design based on the naturalistic inquiry [65], which is especially appropriate for mixed methods research to provide a rich straight description of the facts from the participants' points of view [66]. From August to October 2022, we recruited 43 participants for the qualitative interviews via a WeChat invitation. We used stratified purposeful sampling to recruit heterogeneous participants with different ages, genders, educational levels, employment statuses, monthly household incomes, or health conditions for each user persona from the quantitative sample [67]. Individuals were eligible if they had completed the quantitative survey of this study and agreed to participate in the qualitative interview. The sample size was determined by the principle of data saturation (ie, no new analytical information arose anymore) [68].

Procedure and Analysis

Three qualitatively trained investigators (female graduate nursing students) used a semistructured interview guide to conduct in-depth interviews. The interview guide consisted of open-ended questions informed by the quantitative findings, a group discussion, and a preliminary pilot study with 5 patients, including the following key topics: (1) how do they self-manage depressive symptoms? (2) what are their attitudes toward using eHealth for the self-management of depressive symptoms? (3) what are their goals and needs toward using eHealth for the self-management of depressive symptoms? Interviews took place in a private room at the HIV clinic, at the participant's home, or in another quiet location, either face-to-face or by telephone, whichever was convenient for participants. The length of interviews ranged from 20 to 82 minutes. No one else was present except the participant and the interviewer. The interviewers had no dependency relationship with any participants.

Each interview was audio-recorded, immediately transcribed verbatim, and member-checked to ensure accuracy. Field notes (eg, verbal and nonverbal cues, "off-the-record" texts) were also used to enrich our results further. Two trained investigators repeatedly reviewed the interview transcripts, immersed themselves in these to obtain a sense of the whole interview, and then independently conducted conventional content analysis using NVivo software (version 12; Lumivero) to identify patterns and categories in an inductive approach. An initial codebook was developed after the first 5 interviews and was iteratively refined in the following interviews until a consensus was reached. Disagreements were resolved through discussion among a 5-member review team having multi-disciplinary backgrounds in nursing, HIV, and mental health.

Mixed Methods Data Integration

We used a mixed methods approach to identify user personas to provide a deeper insight and understanding than either approach [39]. As noted, the quantitative findings informed the sampling frame and interview guide of the following qualitative phase (ie, integration at the methods level) [69]. Second, we used a statistics-by-themes joint display as a visual means to integrate the main quantitative and qualitative findings at the analytic and interpretation levels [70], with the following three possible outcomes: (1) confirmation, where the quantitative and qualitative findings confirm each other; (2) expansion, where the quantitative and qualitative findings diverge and expand insights into user personas; and (3) discordance, where the qualitative and quantitative findings are inconsistent, incongruous, contradictory, or conflicting, or disagree with each other [69]. The integration of quantitative and qualitative findings identified key targets for the design of future eHealth. Multimedia Appendix 1 shows the flow of the overall study.

Ethical Considerations

The study was performed in accordance with the Declaration of Helsinki and received ethics approval from the Institutional Review Board of Xiangya School of Nursing, Central South University (number: E202210). All participants provided verbal informed consent before formal participation in the quantitative or qualitative phase. Pseudonyms were used during data collection, storage, and presentation to maintain anonymization and deidentification (ie, linked only to the study ID). No compensation was provided to participants during the quantitative phase. Upon completing the qualitative phase, participants received US $6.85 for their participation.

XSL•FO
RenderX

Results

Quantitative Findings

Participant Characteristics

A total of 572 valid questionnaires were included in this study (see Multimedia Appendix 2). Participants were between 18 and 72 years old, with a median age of 29 (IQR 25-34.8) years. The majority of participants were male (541/572, 94.6%), had higher education or above (452/572, 79.0%), were employed (472/572, 82.5%), and had a monthly household income of ≥10,000 RMB (≥10,000 Chinese Yuan; 390/572, 68.2%). In addition, 22.2% (127/572) of the participants had comorbidities, but most had a latest CD4 count of ≥200 cells/mm^3 (463/572, 80.9%) and an undetectable HIV viral load (368/572, 64.3%). The PHQ-9 scores of the participants ranged from 0 to 27, with a median score of 6 (IQR 3-11), and 64.3% (368/572) had mild or more depressive symptoms. Detailed descriptive statistics are shown in Table 1.

XSL•FO
RenderX

Table 1. Participant characteristics in the quantitative phase (N=572).

Variable	Total (N=572)	High-level self-managers (n=254)	Medium-level self-managers (n=283)	Low-level self-managers (n=35)
Age (years), median (IQR)	29 (25-34.8)	29 (25-35)	28 (25-34)	31 (25-37)
Gender, n (%)				
Male	541 (94.6)	240 (94.5)	270 (95.4)	31 (88.6)
Female	31 (5.4)	14 (5.5)	13 (4.6)	4 (11.4)
Education, n (%)				
High school or below	120 (21.0)	44 (17.3)	68 (24.0)	8 (22.9)
Higher education or above	452 (79.0)	210 (82.7)	215 (76.0)	27 (77.1)
Employment status, n (%)				
Unemployed	100 (17.5)	31 (12.2)	57 (20.1)	12 (34.3)
Employed	472 (82.5)	223 (87.8)	226 (79.9)	23 (65.7)
Monthly household income (RMB[a]), n (%)				
<10,000[b]	157 (27.4)	55 (21.7)	85 (30.0)	17 (48.6)
≥10,000	390 (68.2)	189 (74.4)	184 (65.0)	17 (48.6)
Missing	25 (4.4)	10 (3.9)	14 (4.9)	1 (2.9)
Presence of comorbidities, n (%)				
No	443 (77.4)	212 (83.5)	208 (73.5)	23 (65.7)
Yes	127 (22.2)	41 (16.1)	74 (26.1)	12 (34.3)
Missing	2 (0.3)	1 (0.4)	1 (0.4)	0 (0.00)
Latest CD4 count (cells/mm^3), n (%)				
≥200	463 (80.9)	203 (79.9)	234 (82.7)	26 (74.3)
<200[c]	31 (5.4)	17 (6.7)	9 (3.2)	5 (14.3)
Missing	78 (13.6)	34 (13.4)	40 (14.1)	4 (11.4)
Viral load, n (%)				
Not TND[d] status	93 (16.3)	39 (15.4)	47 (16.6)	7 (20.0)
TND status	368 (64.3)	167 (65.7)	180 (63.6)	21 (60.0)
Missing	111 (19.4)	48 (18.9)	56 (19.8)	7 (20.0)
Severity of depressive symptoms, median (IQR)	6 (3-11)	3 (1-6)	8 (6-12)	19 (12-25)
Self-efficacy, mean (SE)	41.4 (0.5)	51.4 (0.7)	35.9 (1.1)	16.9 (2.3)
DSSM[e], mean (SE)	31.4 (0.2)	34.7 (0.4)	29.4 (0.3)	24.7 (1.0)
eHealth literacy, mean (SE)	29.5 (0.3)	31.9 (0.5)	27.9 (0.4)	24.8 (1.3)

[a]RMB: Renminbi (Chinese yuan).

[b]RMB 1=US $0.1375.

[c]Latest CD4 count <200 cells/mm^3 indicates advanced HIV disease [46].

[d]TND: target not detected.

[e]DSSM: self-management abilities of depressive symptoms.

Model Selection

Table 2 presents fit statistics for 1- to 5-profile structures. As the number of profiles increased, the solutions provided lower AIC, BIC, and saBIC values and statistically significant BLRT values. However, the smallest profile of the 4- or 5-profile solutions contained less than 5% of the total sample. Considering the principle of substantive interpretability, we did not examine these 2 profile solutions further, as these profiles may be spurious [60]. We finally retained the 3-profile solution as the optimal solution because it exhibited lower AIC, BIC, and saBIC values and a higher entropy value than the 2-profile solution, with statistically significant aLMRT and BLRT values, indicating a significant improvement in fit statistics compared to the 2-profile solution. Meanwhile, the smallest profile of the

XSL•FO
RenderX

3-profile solution contained more than 5% of the total sample (35/572, 6.1%), which is the minimum reference standard to support the generalizability of the LPA [59,60]. Average posterior probabilities for individuals assigned to each profile in this solution were 0.898, 0.879, and 0.928, respectively, with all values greater than 0.70, indicating well-separated profiles [71].

Table 2. Fit statistics for 1- to 5-profile structures (N=572).

Number of profiles	AIC[a]	BIC[b]	saBIC[c]	Entropy	aLMRT[d] (*P*)	BLRT[e] (*P*)	Latent profile proportion
1	4881.797	4907.892	4888.844	—[f]	—	—	1.000
2	4667.982	4711.473	4679.728	0.627	<.001	<.001	0.579/0.421
3	4615.057	4675.944	4631.501	0.762	.003	<.001	0.495/0.444/0.061
4	4591.966	4670.251	4613.109	0.736	.84	<.001	0.521/0.364/0.096/0.019
5	4529.099	4624.780	4554.940	0.790	.008	<.001	0.540/0.308/0.108/0.033/0.010

[a]AIC: Akaike Information Criterion.

[b]BIC: Bayesian Information Criterion.

[c]saBIC: sample-size adjusted BIC.

[d]aLMRT: adjusted Lo-Mendel-Rubin likelihood ratio test.

[e]BLRT: bootstrap likelihood ratio test.

[f]Not applicable.

Profile Characteristics

Table 1 displays the estimated means and SEs for the 3 indicators in each profile. Profile 1 (254/572, 44.4%; *high-level self-managers*) was characterized by relatively higher self-efficacy (mean 51.4, SE 0.7), self-management abilities of depressive symptoms (mean 34.7, SE 0.4), and eHealth literacy (mean 31.9, SE 0.5). Profile 2 (283/572, 49.5%; *medium-level self-managers*) was characterized by relatively medium self-efficacy (mean 35.9, SE 1.1), self-management abilities of depressive symptoms (mean 29.4, SE 0.3), and eHealth literacy (mean 27.9, SE 0.4). Profile 3 (35/572, 6.1%; *low-level self-managers*) was characterized by relatively lower self-efficacy (mean 16.9, SE 2.3), self-management abilities of depressive symptoms (mean 24.7, SE 1.0), and eHealth literacy (mean 24.8, SE 1.3). Figure 1 presents a graphical representation of the 3 profiles based on the z-scores.

Figure 1. Three-profile structure based on z-scores (N=572). DSSM: self-management abilities of depressive symptoms.

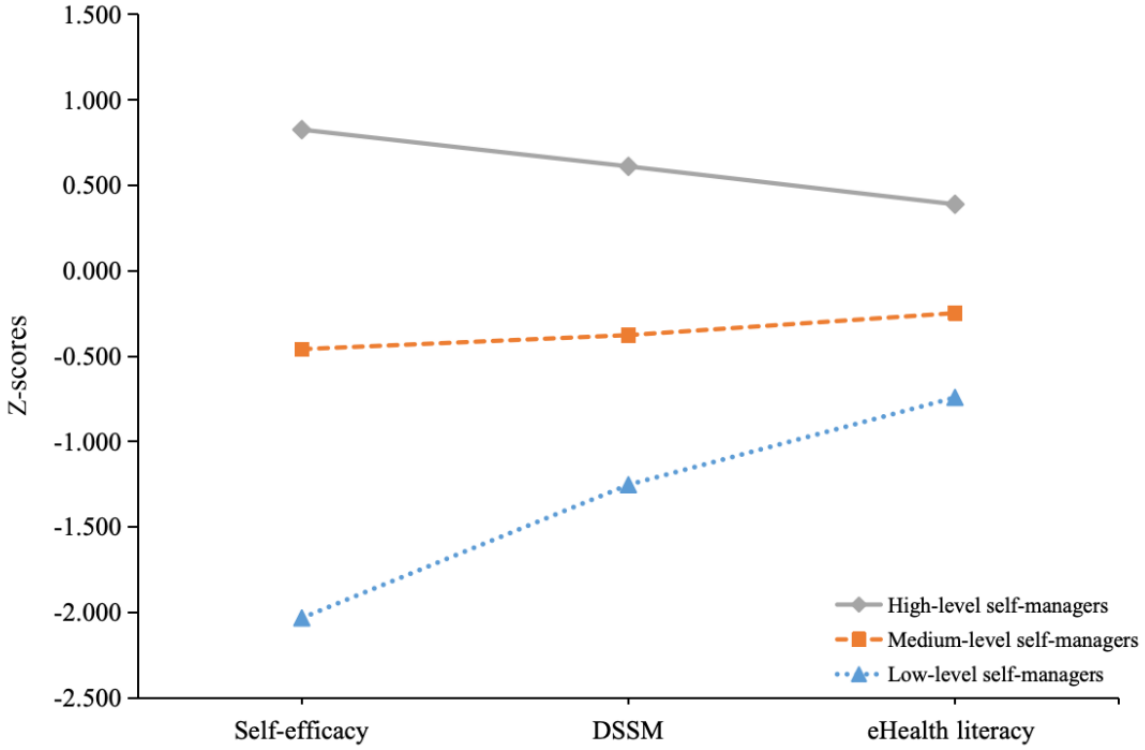

XSL•FO
RenderX

Three-Step Approach

The 3-step results for the covariates are shown in Multimedia Appendix 3. Compared with *high-level self-managers*, we found that *medium-level self-managers* and *low-level self-managers* were more likely to have comorbidities and more severe depressive symptoms. Additionally, we found that *low-level self-managers* were more likely to have more severe depressive symptoms than *medium-level self-managers*.

Qualitative Findings

Interview Sample Characteristics

In total, 43 of 572 individuals who completed quantitative surveys volunteered to participate in semistructured interviews, including 23 *high-level self-managers*, 17 *medium-level self-managers*, and 3 *low-level self-managers*. The sample consisted of 37 males and 6 females, with a median age of 29 years (IQR 25.5-37; range 19-70 years). A total of 30 participants had used eHealth for the self-management of depressive symptoms. See Multimedia Appendix 4 for more details.

The findings are summarized by characteristics for the self-management of depressive symptoms, attitudes toward using eHealth, and goals and needs toward using eHealth. See Multimedia Appendix 5 for representative quotes.

Characteristics for the Self-Management of Depressive Symptoms

High-level self-managers were *particularly proactive* in the self-management of depressive symptoms, such as proactively seeking relevant information and social support. High-level self-managers believed that their emotional management plays a critical role in their health conditions and can, in the worst case, be devastating or fatal.

Medium-level self-managers were *proactive* in the self-management of depressive symptoms, but most used distraction techniques, such as playing on their phones. Medium-level self-managers believed that health conditions were the leading cause of their depressive symptoms and thus emphasized the moderating effects of health conditions on their depressive symptoms.

Low-level self-managers were *negative* about the self-management of depressive symptoms. Low-level self-managers often did not know how to make themselves better.

Attitudes Toward Using eHealth

High-level self-managers had a *positive* attitude toward using eHealth for the self-management of depressive symptoms. High-level self-managers often proactively used eHealth to manage their depressive symptoms, including searching for health information, following up on the latest news about HIV, and seeking peer advice.

Medium-level self-managers felt *burdened* toward using eHealth for the self-management of depressive symptoms. Medium-level self-managers usually did not use eHealth to manage their depressive symptoms to avoid potential burdens caused by frequent pushes, such as anxiety, feeling down, or feeling different from others.

Low-level self-managers felt it was *acceptable* to use eHealth for the self-management of depressive symptoms, especially for online guidance from professionals.

Goals and Needs Toward Using eHealth

High-level self-managers desired *access to self-management support for depressive symptoms* from eHealth, including (1) psychological support: psychological counseling, psychological information, peer support, and referral, and (2) HIV-related support: HIV-related information and counseling. Meanwhile, high-level self-managers desired eHealth to have *high usability*, including privacy, a user-friendly interface, personalized content, a positive and healthy atmosphere, and increased fun.

Medium-level self-managers focused on *access to self-management support for HIV* from eHealth, including HIV-related information and counseling, and emphasized the *privacy* of eHealth.

Low-level self-managers desired *access to guidance from professionals*, either psychological or HIV-related. Meanwhile, low-level self-managers desired eHealth with *privacy* and *no cost (free of charge)*.

Integrated Findings

The main quantitative and qualitative findings were integrated into a joint display. See Multimedia Appendix 5 for more details.

High-level self-managers had relatively high levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. High-level self-managers had a positive attitude toward using eHealth for the self-management of depressive symptoms and desired access to self-management support for depressive symptoms from eHealth with high usability.

Medium-level self-managers had relatively medium levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. Medium-level self-managers emphasized the moderating effects of health conditions on their depressive symptoms and felt burdened by using eHealth for the self-management of depressive symptoms. As a result, medium-level self-managers preferred to access self-management support for HIV from eHealth with privacy.

Low-level self-managers had relatively low levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy. Low-level self-managers had an acceptable attitude toward using eHealth for the self-management of depressive symptoms and desired access to professional guidance from eHealth with privacy and no cost (free of charge).

Discussion

Principal Findings and Comparison With Prior Work

This mixed methods study identified 3 types of user personas for eHealth regarding the self-management of depressive symptoms in people living with HIV, including high-level self-managers, medium-level self-managers, and low-level

XSL•FO
RenderX

self-managers. The different user personas reflect different people living with HIV, their characteristics, and their goals and needs for future eHealth. Therefore, developers may need to consider a tailored perspective in the design of eHealth for different user personas.

High-level self-managers were characterized by relatively high levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy, with minimal depressive symptoms. People living with HIV in this user persona attached great importance to the impact of emotional management on their health condition, and thus, they were particularly proactive in the self-management of depressive symptoms. This phenomenon may be explained by the fact that people with high self-efficacy and self-management abilities tend to see difficulties as challenges to overcome rather than to avoid, so they act with more proactive coping efforts in threatening situations [37]. In addition, people with fewer depressive symptoms tend to perceive more benefits and fewer barriers to taking effective measures to manage their depressive symptoms, so they are more likely to engage in self-management proactively [72]. Moreover, considering their positive attitudes toward eHealth, it is unsurprising that high-level self-managers desired access to more self-management support for depressive symptoms from eHealth, which could make them feel more empowered to control their health [25,26]. Notably, high-level self-managers desired eHealth with high usability, including privacy, a user-friendly interface, personalized content, a positive and healthy atmosphere, and increased fun, similar to the findings in previous research [25,26,73]. These findings may be explained by their relatively high levels of eHealth literacy [25], suggesting that future eHealth for high-level self-managers could be designed from a multi-dimensional perspective to provide an ideal interactive feel.

Medium-level self-managers were characterized by relatively medium levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy, with comorbidities and more severe depressive symptoms. People living with HIV in this user persona were proactive in the self-management of depressive symptoms, whereas most tended to use distraction techniques and placed more emphasis on the moderating effects of health conditions on their depressive symptoms. Accordingly, they preferred to access self-management support for HIV from eHealth. This phenomenon may be partly because people with comorbidities may experience more complexity in self-management, such as additional physical or psychological strain [74,75]. As a result, they tend to struggle with self-management and feel less confident to do so [76-78]. Therefore, it makes sense that medium-level self-managers desired access to self-management support for HIV from eHealth. In addition, this phenomenon may be explained by the traditional cultural beliefs and the stigma against depression in Chinese society [79-81], where people tend to keep their mental discomfort to themselves or refer to mental discomfort as physical symptoms to maintain their “face” (dignity, reputation, and public image), which may prevent them from seeking help for mental health [82,83]. Previous research also confirmed that many people living with HIV in China hardly ever use mental health services, including those with persistent depression [8]. Similar phenomena have been found in other geographical areas, such as Europe, South Asia, and South-East Asia [84]. These findings suggest that future eHealth should not only meet the goals and needs of medium-level self-managers but also improve their recognition in the self-management of depressive symptoms to promote behavioral change. Finally, our findings indicated that medium-level self-managers tended to feel burdened by eHealth resources with frequent push notifications, and thus, they used such resources less often, which is in agreement with previous meta-ethnographic findings [25]. These findings suggest that future eHealth for medium-level self-managers should be designed to promote their use of eHealth in the right amount according to the interactive principle of “Not too much. Not too little. Just right.” [85].

Low-level self-managers were characterized by relatively low levels of self-efficacy, self-management abilities of depressive symptoms, and eHealth literacy, with comorbidities and the most severe depressive symptoms. People living with HIV in this user persona were negative about the self-management of depressive symptoms. One possible explanation is that people with low self-efficacy tend to regard their efforts as futile in the face of difficulties and thus quickly give up trying, which, combined with their low self-management abilities, makes it difficult for them to achieve the desired performance in the self-management of depressive symptoms [37]. Another possible explanation is that people with severe depressive symptoms usually experience diminished interest in doing things, loss of energy, reduced ability to think or concentrate, or indecisiveness, and as a result, they often have difficulty in proactive self-management [72,86,87]. However, our findings indicated that low-level self-managers had an acceptable attitude toward eHealth resources and desired access to professional guidance from eHealth with privacy and no cost (free of charge). This finding suggests that future eHealth for low-level self-managers could be designed as a connection point for 2-way communication with an interactive feeling of “minimizing privacy risks and maximizing public benefits.”

Notably, all 3 types of user personas emphasized the privacy of eHealth, such as invitation codes and non-HIV–related names, which may be attributed to HIV-related stigma and discrimination. Although antiretroviral therapy has transformed HIV infection into a manageable chronic health condition [88], people living with HIV worldwide continue to face significant stigma and discrimination at the individual, interpersonal, community, and societal levels, including avoidance, gossip, verbal or physical abuse, social rejection, denial of health or social services, denial or loss of employment or education opportunities, or even arrest [89,90]. As a result, people living with HIV often value privacy to avoid unwanted disclosures and try to live a “normal” life [91,92]. Previous research has also emphasized the privacy concerns regarding eHealth platforms for HIV management, even in developed countries [92-94]. These findings suggest that future eHealth should be designed with adequate consideration of regulatory, privacy, and security protocols, regardless of user personas.

XSL•FO
RenderX

Strengths and Limitations

To our knowledge, this is the first study to construct user personas for eHealth regarding the self-management of depressive symptoms in people living with HIV. We constructed this powerful communication tool based on goal-directed design, providing an effective way for user involvement in the development phase of eHealth [28,95]. In addition, this study used an explanatory sequential mixed methods design, which integrated quantitative and qualitative findings to provide a more in-depth exploration of user personas. Our findings may be of use to gain a better understanding of intended end users and inform the design of future eHealth.

However, this study has several limitations. First, participants were recruited from a single site in the metropolitan area, and thus, our findings might not be generalizable to other areas of China. However, the site is a large designated tertiary hospital for HIV diagnosis and treatment, with a good representation of the diverse patient population in the survey area. Therefore, it can still be assumed that our findings reflect a social trend. Second, most participants were young males, which might be explained by the difficult access to vulnerable people living with HIV (eg, females) [96]. Therefore, the findings should be interpreted with caution. It would be helpful to include more vulnerable people living with HIV in future research to reflect their unique needs. Third, we recruited only 3 low-level self-managers for qualitative interviews, which might have contributed to potential bias. This may be due to the small total number of low-level self-managers and the fact that low-level self-managers usually experienced the most severe depressive symptoms, with diminished interest in interviews. However, we endeavored to ensure the heterogeneity of the participants by stratified purposeful sampling as much as possible, and the qualitative data revealed no new analytical information, so we presumed data saturation. Finally, as the study did not collect data over time, we did not investigate the dynamics of user personas. Future research can examine this vital area to gain more information to guide the design of future eHealth.

Implications for Practice

Our findings provide a practical communication aid, namely user personas, for interdisciplinary collaborations between health researchers and developers, which contributes to the integration of the “art of caring” and the “art of design” in future eHealth to provide personalized support for the self-management of depressive symptoms in people living with HIV [28,97]. Specifically, user personas could help frontline health researchers better play the role of requirement analysts during interdisciplinary collaborations to bring the user voice to the forefront in a clear and easily understandable way, thereby helping developers to keep different types of users and their goals and needs in mind during the early stages of development and continuously throughout the evaluation and implementation phases [95,97]. Furthermore, user personas could help elicit specific advice from health care providers, regulators, and investors by making the intended end users more tangible, which can contribute to addressing broader system and policy constraints [98].

Conclusions

According to goal-directed design, we identified 3 types of user personas for eHealth regarding the self-management of depressive symptoms in people living with HIV: high-level self-managers, medium-level self-managers, and low-level self-managers. The 3 user personas shed light on the possibility of personalized eHealth to support the self-management of depressive symptoms in different people living with HIV. Further research is needed to examine the generalizability of the user personas across study sites.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (82273746) and the Hainan Provincial Department of Science and Technology, China (ZDYF2024SHFZ042). The funders were not involved in the design of the study, data collection, data analysis, or manuscript preparation. The authors would like to thank all the reviewers and participants for their assistance and support. No generative artificial intelligence was used in any portion of manuscript writing.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart for the explanatory sequential mixed methods design.
[DOC File , 74 KB - jmir_v27i1e56289_app1.doc]

Multimedia Appendix 2
Participant flow throughout the study.
[DOC File , 74 KB - jmir_v27i1e56289_app2.doc]

XSL•FO
RenderX

Multimedia Appendix 3
Three-step results for covariates in the quantitative phase (n=572).
[DOC File , 21 KB - jmir_v27i1e56289_app3.doc]

Multimedia Appendix 4
Participant characteristics in the qualitative phase (n=43).
[DOC File , 104 KB - jmir_v27i1e56289_app4.doc]

Multimedia Appendix 5
Joint display of the main quantitative and qualitative findings.
[DOC File , 122 KB - jmir_v27i1e56289_app5.doc]

Multimedia Appendix 6
GRAMMS (Good Reporting of A Mixed Methods Study) checklist.
[PDF File (Adobe PDF File), 63 KB - jmir_v27i1e56289_app6.pdf]

Multimedia Appendix 7
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cross-sectional studies.
[PDF File (Adobe PDF File), 141 KB - jmir_v27i1e56289_app7.pdf]

Multimedia Appendix 8
SRQR (Standards for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 202 KB - jmir_v27i1e56289_app8.pdf]

References

1. Rezaei S, Ahmadi S, Rahmati J, Hosseinifard H, Dehnad A, Aryankhesal A, et al. Global prevalence of depression in HIV/AIDS: a systematic review and meta-analysis. BMJ Support Palliat Care 2019 Dec 19;9(4):404-412. [doi: 10.1136/bmjspcare-2019-001952] [Medline: 31537580]
2. Musisi S, Wagner GJ, Ghosh-Dastidar B, Nakasujja N, Dickens A, Okello E. Depression and sexual risk behaviour among clients about to start HIV antiretroviral therapy in Uganda. Int J STD AIDS 2014 Feb;25(2):130-137 [FREE Full text] [doi: 10.1177/0956462413495186] [Medline: 23970636]
3. Schuster R, Bornovalova M, Hunt E. The influence of depression on the progression of HIV: direct and indirect effects. Behav Modif 2012 Mar;36(2):123-145. [doi: 10.1177/0145445511425231] [Medline: 22089635]
4. Chakraborty A, Hershow RC, Qato DM, Stayner L, Dworkin MS. Adherence to antiretroviral therapy among HIV patients in India: a systematic review and meta-analysis. AIDS Behav 2020 Jul;24(7):2130-2148. [doi: 10.1007/s10461-020-02779-4] [Medline: 31933019]
5. Necho M, Tsehay M, Zenebe Y. Suicidal ideation, attempt, and its associated factors among HIV/AIDS patients in Africa: a systematic review and meta-analysis study. Int J Ment Health Syst 2021 Jan 23;15(1):13 [FREE Full text] [doi: 10.1186/s13033-021-00437-3] [Medline: 33485362]
6. Yoo-Jeong M, Alvarez G, Khawly G, Voss J, Wang T, Barroso J, et al. A systematic review of self-management interventions conducted across global settings for depressive symptoms in persons with HIV. AIDS Behav 2023 May;27(5):1486-1501. [doi: 10.1007/s10461-022-03945-6] [Medline: 36520336]
7. Pence BW, O'Donnell JK, Gaynes BN. Falling through the cracks: the gaps between depression prevalence, diagnosis, treatment, and response in HIV care. AIDS 2012 Mar 13;26(5):656-658 [FREE Full text] [doi: 10.1097/QAD.0b013e3283519aae] [Medline: 22398574]
8. Niu L, Luo D, Chen X, Wang M, Zhou W, Zhang D, et al. Longitudinal trajectories of emotional problems and unmet mental health needs among people newly diagnosed with HIV in China. J Int AIDS Soc 2019 Aug;22(8):e25332 [FREE Full text] [doi: 10.1002/jia2.25332] [Medline: 31424617]
9. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep 08;370(9590):878-889. [doi: 10.1016/S0140-6736(07)61239-2] [Medline: 17804062]
10. Filiatreau LM, Ebasone PV, Dzudie A, Ajeh R, Pence B, Wainberg M, et al. Correlates of self-reported history of mental health help-seeking: a cross-sectional study among individuals with symptoms of a mental or substance use disorder initiating care for HIV in Cameroon. BMC Psychiatry 2021 Jun 05;21(1):293 [FREE Full text] [doi: 10.1186/s12888-021-03306-y] [Medline: 34090367]
11. Kinsler JJ, Wong MD, Sayles JN, Davis C, Cunningham WE. The effect of perceived stigma from a health care provider on access to care among a low-income HIV-positive population. AIDS Patient Care STDS 2007 Aug;21(8):584-592. [doi: 10.1089/apc.2006.0202] [Medline: 17711383]

XSL•FO
RenderX

12. Karyotaki E. Internet-based interventions for people with HIV and depression. Lancet HIV 2018 Sep;5(9):e474-e475. [doi: 10.1016/S2352-3018(18)30172-3] [Medline: 30135044]
13. Digitally enabled therapies for adults with depression: early value assessment. National Institute for Health and Care Excellence. 2023 May 16. URL: https://www.nice.org.uk/guidance/hte8 [accessed 2025-01-15]
14. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
15. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
16. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
17. Xiang Y, Zhao N, Zhao Y, Liu Z, Zhang Q, Feng Y, et al. An overview of the expert consensus on the mental health treatment and services for major psychiatric disorders during COVID-19 outbreak: China's experiences. Int J Biol Sci 2020;16(13):2265-2270 [FREE Full text] [doi: 10.7150/ijbs.47419] [Medline: 32760195]
18. Cheng LJ, Kumar PA, Wong SN, Lau Y. Technology-delivered psychotherapeutic interventions in improving depressive symptoms among people with HIV/AIDS: A systematic review and meta-analysis of randomised controlled trials. AIDS Behav 2020 Jun;24(6):1663-1675. [doi: 10.1007/s10461-019-02691-6] [Medline: 31587115]
19. Xiao Y, Shao Y, Na Z, Zhao W, Wang R, Fang S, et al. A systematic review and meta-analysis of telephone-based therapy targeting depressive symptoms among low-income people living with HIV. AIDS Behav 2021 Feb;25(2):414-426. [doi: 10.1007/s10461-020-02999-8] [Medline: 32809074]
20. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
21. Powell J, Deetjen U. Characterizing the digital health citizen: mixed-methods study deriving a new typology. J Med Internet Res 2019 Mar 05;21(3):e11279 [FREE Full text] [doi: 10.2196/11279] [Medline: 30835238]
22. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: A qualitative study. Int J Med Inform 2019 Nov;131:103949 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.009] [Medline: 31561192]
23. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
24. Martikainen S, Viitanen J, Korpela M, Lääveri T. Physicians' experiences of participation in healthcare IT development in Finland: willing but not able. Int J Med Inform 2012 Feb;81(2):98-113. [doi: 10.1016/j.ijmedinf.2011.08.014] [Medline: 21956004]
25. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
26. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
27. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
28. Cooper A, Reimann R, Cronin D. About Face 3: The Essentials of Interaction Design. Indianapolis, IN: Wiley Publishing, Inc; 2007.
29. Serio CD, Hessing J, Reed B, Hess C, Reis J. The effect of online chronic disease personas on activation: within-subjects and between-groups analyses. JMIR Res Protoc 2015 Feb 25;4(1):e20 [FREE Full text] [doi: 10.2196/resprot.3392] [Medline: 25720676]
30. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
31. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? results from a web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
32. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
33. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
34. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]

XSL•FO
RenderX

35. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
36. Schunk D, DiBenedetto M. Self-efficacy and human motivation. In: Elliot AJ, editor. Advances in Motivation Science. San Diego, CA: Elsevier Academic Press; 2021:153-179.
37. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
38. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
39. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
40. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
41. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
42. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
43. The 35th World AIDS Day in Hunan Province. Hunan Provincial Center for Disease Control and Prevention. 2022 Dec 1. URL: https://wjw.hunan.gov.cn/wjw/xxgk/gzdt/szdt/202212/t20221213_29159995.html [accessed 2025-01-15]
44. Department of Infection and Immunity. The First Hospital of Changsha. URL: http://www.cssdyyy.com/list/302_4650.html [accessed 2025-01-15]
45. About HIV. Centers for Disease Control and Prevention. 2025 Jan 14. URL: https://www.cdc.gov/hiv/about/index.html [accessed 2025-01-15]
46. HIV. World Health Organization. URL: https://www.who.int/health-topics/hiv-aids [accessed 2025-01-15]
47. Peel NM, Russell TG, Gray LC. Feasibility of using an in-home video conferencing system in geriatric rehabilitation. J Rehabil Med 2011 Mar;43(4):364-366 [FREE Full text] [doi: 10.2340/16501977-0675] [Medline: 21305228]
48. Tao Y, Xiao X, Ma J, Wang H. The relationship between HIV-related stigma and HIV self-management among men who have sex with men: The chain mediating role of social support and self-efficacy. Front Psychol 2022;13:1094575 [FREE Full text] [doi: 10.3389/fpsyg.2022.1094575] [Medline: 36600713]
49. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
50. Yang F, Liu S, Yang B, Fang S, Huang R, Chen P. Chinese version of depression specific self management questionnaire and its reliability and validity test. Chinese Nursing Research 2020;34(19):3413-3417 [FREE Full text]
51. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
52. Norman CD, Skinner HA. eHEALS: The eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
53. Guo S, Yu X, Sun Y, Nie D, Li X, Wang L. Adaptation and evaluation of Chinese version of eHEALS and its usage among senior high school students. Chinese Journal of Health Education 2013;29(02):106-8+23. [doi: 10.16168/j.cnki.issn.1002-9982.2013.02.019]
54. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
55. Wang Y, Kim E, Yi Z. Robustness of latent profile analysis to measurement noninvariance between profiles. Educ Psychol Meas 2022 Feb;82(1):5-28 [FREE Full text] [doi: 10.1177/0013164421997896] [Medline: 34992305]
56. Spurk D, Hirschi A, Wang M, Valero D, Kauffeld S. Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. Journal of Vocational Behavior 2020 Aug;120:103445. [doi: 10.1016/j.jvb.2020.103445]
57. Berlin KS, Williams NA, Parra GR. An introduction to latent variable mixture modeling (part 1): overview and cross-sectional latent class and latent profile analyses. J Pediatr Psychol 2014 Mar;39(2):174-187. [doi: 10.1093/jpepsy/jst084] [Medline: 24277769]
58. Morgan GB, Hodge KJ, Baggett AR. Latent profile analysis with nonnormal mixtures: A Monte Carlo examination of model selection using fit indices. Computational Statistics & Data Analysis 2016 Jan;93:146-161. [doi: 10.1016/j.csda.2015.02.019]
59. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling: A Multidisciplinary Journal 2007 Dec 05;14(4):535-569. [doi: 10.1080/10705510701575396]
60. Ferguson S, G. Moore E, Hull D. Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. International Journal of Behavioral Development 2019 Nov 22;44(5):458-468. [doi: 10.1177/0165025419881721]

XSL•FO
RenderX

61. Asparouhov T, Muthén B. Auxiliary variables in mixture modeling: three-step approaches using Mplus. Structural Equation Modeling: A Multidisciplinary Journal 2014 Jun 09;21(3):329-341. [doi: 10.1080/10705511.2014.915181]
62. Vermunt JK. Latent class modeling with covariates: two improved three-step approaches. Polit. anal 2010 Sep 23;18(4):450-469. [doi: 10.1093/pan/mpq025]
63. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
64. Muthén LK, Muthén BO. Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 2017.
65. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
66. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
67. Sandelowski M. Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. Res Nurs Health 2000 Jun;23(3):246-255. [doi: 10.1002/1098-240x(200006)23:3<246::aid-nur9>3.0.co;2-h] [Medline: 10871540]
68. Moser A, Korstjens I. Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18 [FREE Full text] [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
69. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
70. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
71. Nylund-Gibson K, Choi AY. Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science 2018 Dec;4(4):440-461. [doi: 10.1037/tps0000176]
72. Huang R, Wang XQ, Yang BX, Liu Z, Chen WC, Jiao SF, et al. Self-management of depression among Chinese community individuals: A cross-sectional study using the transtheoretical model. Perspect Psychiatr Care 2022 Jan;58(1):256-265. [doi: 10.1111/ppc.12779] [Medline: 33792031]
73. Ho TQA, Le LK, Engel L, Le N, Melvin G, Le HND, et al. Barriers to and facilitators of user engagement with web-based mental health interventions in young people: a systematic review. Eur Child Adolesc Psychiatry 2024 Feb 14 (forthcoming). [doi: 10.1007/s00787-024-02386-x] [Medline: 38356043]
74. Kenning C, Fisher L, Bee P, Bower P, Coventry P. Primary care practitioner and patient understanding of the concepts of multimorbidity and self-management: A qualitative study. SAGE Open Med 2013;1:2050312113510001 [FREE Full text] [doi: 10.1177/2050312113510001] [Medline: 26770690]
75. Heer E, Kaida A, O'Brien N, Kleiner B, Pierre A, Rouleau D, et al. Prevalence of physical health, mental health, and disability comorbidities among women living with HIV in Canada. J Pers Med 2022 Aug 06;12(8):1294 [FREE Full text] [doi: 10.3390/jpm12081294] [Medline: 36013243]
76. Dickson V, Buck H, Riegel B. Multiple comorbid conditions challenge heart failure self-care by decreasing self-efficacy. Nurs Res 2013;62(1):2-9. [doi: 10.1097/NNR.0b013e31827337b3] [Medline: 23052421]
77. Buck HG, Dickson VV, Fida R, Riegel B, D'Agostino F, Alvaro R, et al. Predictors of hospitalization and quality of life in heart failure: A model of comorbidity, self-efficacy and self-care. Int J Nurs Stud 2015 Nov;52(11):1714-1722. [doi: 10.1016/j.ijnurstu.2015.06.018] [Medline: 26234935]
78. Gobeil-Lavoie A, Chouinard M, Danish A, Hudon C. Characteristics of self-management among patients with complex health needs: a thematic analysis review. BMJ Open 2019 May 24;9(5):e028344 [FREE Full text] [doi: 10.1136/bmjopen-2018-028344] [Medline: 31129599]
79. Ran M, Zhang T, Wong IY, Yang X, Liu C, Liu B, CMHP Study Group. Internalized stigma in people with severe mental illness in rural China. Int J Soc Psychiatry 2018 Feb;64(1):9-16. [doi: 10.1177/0020764017743999] [Medline: 29183250]
80. Lam C, Tsang H, Corrigan P, Lee Y, Angell B, Shi K, et al. Chinese lay theory and mental illness stigma: implications for research and practices. Journal of Rehabilitation 2010;76(1):35-40 [FREE Full text]
81. Chen SX, Mak WWS. Seeking professional help: Etiology beliefs about mental illness across cultures. J Couns Psychol 2008 Oct;55(4):442-450. [doi: 10.1037/a0012898] [Medline: 22017551]
82. Sun KS, Lam TP, Wu D. Chinese perspectives on primary care for common mental disorders: Barriers and policy implications. Int J Soc Psychiatry 2018 Aug;64(5):417-426 [FREE Full text] [doi: 10.1177/0020764018776347] [Medline: 29781372]
83. Yang F, Yang BX, Stone TE, Wang XQ, Zhou Y, Zhang J, et al. Stigma towards depression in a community-based sample in China. Compr Psychiatry 2020 Feb;97:152152 [FREE Full text] [doi: 10.1016/j.comppsych.2019.152152] [Medline: 31838297]
84. Dubreucq J, Plasse J, Franck N. Self-stigma in serious mental illness: a systematic review of frequency, correlates, and consequences. Schizophr Bull 2021 Aug 21;47(5):1261-1287 [FREE Full text] [doi: 10.1093/schbul/sbaa181] [Medline: 33459793]

XSL•FO
RenderX

85. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
86. Liu S, Yang BX, Gong X, Chen J, Liu Z, Zhang J, et al. Prevalence and influencing factors of depression self-management among Chinese community residents: a cross-sectional study. Front Psychiatry 2021;12:559844 [FREE Full text] [doi: 10.3389/fpsyt.2021.559844] [Medline: 34025464]
87. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th edition. Washington, DC: American Psychiatric Publishing, Inc; 2013.
88. HIV and AIDS. World Health Organization. 2024 Jul 22. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids [accessed 2025-01-16]
89. HIV and stigma and discrimination. Joint United Nations Programme on HIV/AIDS. URL: https://www.unaids.org/sites/default/files/media_asset/07-hiv-human-rights-factsheet-stigma-discrmination_en.pdf [accessed 2025-01-16]
90. Goodenow MM, Rausch DM. Recent key efforts to improve HIV-related intersectional stigma and discrimination research. Am J Public Health 2022 Jun;112(S4):S393-S394. [doi: 10.2105/AJPH.2022.306712] [Medline: 35763741]
91. McDonald K, Slavin S, Pitts MK, Elliott JH, HealthMap Project Team. Chronic disease self-management by people with HIV. Qual Health Res 2016 May;26(6):863-870. [doi: 10.1177/1049732315600415] [Medline: 26290540]
92. Hewitt C, Lloyd KC, Tariq S, Durrant A, Claisse C, Kasadha B, et al. Patient-generated data in the management of HIV: a scoping review. BMJ Open 2021 May 19;11(5):e046393 [FREE Full text] [doi: 10.1136/bmjopen-2020-046393] [Medline: 34011598]
93. Marent B, Henwood F, Darking M, EmERGE Consortium. Development of an mHealth platform for HIV care: Gathering user perspectives through co-design workshops and interviews. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e184 [FREE Full text] [doi: 10.2196/mhealth.9856] [Medline: 30339132]
94. Saberi P, Siedle-Khan R, Sheon N, Lightfoot M. The use of mobile health applications among youth and young adults living with HIV: focus group findings. AIDS Patient Care STDS 2016 Jun;30(6):254-260 [FREE Full text] [doi: 10.1089/apc.2016.0044] [Medline: 27214751]
95. Bartels SL, Taygar AS, Johnsson SI, Petersson S, Flink I, Boersma K, et al. Using Personas in the development of eHealth interventions for chronic pain: A scoping review and narrative synthesis. Internet Interv 2023 Apr;32:100619 [FREE Full text] [doi: 10.1016/j.invent.2023.100619] [Medline: 37273935]
96. Wang K, Chen W, Zhang L, Bao M, Zhao H, Lu H. Facilitators of and barriers to HIV self-management: Perspectives of HIV-positive women in China. Appl Nurs Res 2016 Nov;32:91-97 [FREE Full text] [doi: 10.1016/j.apnr.2016.06.004]
97. Zhou Y, Li Z, Li Y. Interdisciplinary collaboration between nursing and engineering in health care: A scoping review. Int J Nurs Stud 2021 May;117:103900. [doi: 10.1016/j.ijnurstu.2021.103900] [Medline: 33677250]
98. Bhattacharyya O, Mossman K, Gustafsson L, Schneider EC. Using human-centered design to build a digital health advisor for patients with complex needs: persona and prototype development. J Med Internet Res 2019 May 09;21(5):e10318 [FREE Full text] [doi: 10.2196/10318] [Medline: 31094334]

Abbreviations

AIC: Akaike Information Criterion
aLMRT: adjusted Lo-Mendel-Rubin likelihood ratio test
BIC: Bayesian Information Criterion
BLRT: bootstrap likelihood ratio test
LPA: latent profile analysis
PHQ-9: 9-item Patient Health Questionnaire
saBIC: sample size–adjusted Bayesian Information Criterion
TND: target not detected

Edited by A Mavragani; submitted 12.01.24; peer-reviewed by O Ezeigwe, C Grov, Y Sun; comments to author 07.10.24; revised version received 30.11.24; accepted 22.12.24; published 17.02.25.

Please cite as:
Zhao T, Tang C, Ma J, Yan H, Su X, Zhong X, Wang H
User Personas for eHealth Regarding the Self-Management of Depressive Symptoms in People Living With HIV: Mixed Methods Study
J Med Internet Res 2025;27:e56289
URL: https://www.jmir.org/2025/1/e56289
doi:10.2196/56289
PMID:

XSL•FO
RenderX

©Ting Zhao, Chulei Tang, Jun Ma, Huang Yan, Xinyi Su, Xueyuan Zhong, Honghong Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Digital Cancer Care for Older Black Adults: Qualitative Study

Paul Wankah[1,2], MD, PhD; Shivani Chandra[2], MSc; Aisha Lofters[2,3], MD, PhD; Nebila Mohamednur[2], BScPharm, MSc; Beverley Osei[2], BHSc; Tutsirai Makuwaza[2], MSc; Ambreen Sayani[2,3], MD, PhD

[1]Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, QC, Canada

[2]Women's College Hospital, Toronto, ON, Canada

[3]University of Toronto, Toronto, ON, Canada

Corresponding Author:
Paul Wankah, MD, PhD
Faculty of Dental Medicine and Oral Health Sciences
McGill University
2001 McGill Avenue
Montreal, QC, H3A 1G1
Canada
Phone: 1 5143962402
Email: paul.wankahnji@mcgill.ca

Abstract

Background: Health systems are rapidly promoting digital cancer care models to improve cancer care of their populations. However, there is growing evidence that digital cancer care can exacerbate inequities in cancer care for communities experiencing social disadvantage, such as Black communities. Despite the increasing recognition that older Black adults face significant challenges in accessing and using health care services due to multiple socioeconomic and systemic factors, there is still limited evidence regarding how older Black adults' access and use digital cancer care.

Objective: This study aims to better understand the digital cancer care experience of older Black adults, their caregivers, and health care providers to identify strategies that can better support patient-centered digital cancer care.

Methods: A total of 6 focus group interviews were conducted with older Black adults living with cancer, caregivers, and health care providers (N=55 participants) across 10 Canadian provinces. Focus group interviews were recorded and transcribed. Through a theory-informed thematic analysis approach, experienced qualitative researchers used the Patient Centered Care model and the synergies of oppression conceptual lens to inductively and deductively code interview transcripts in order to develop key themes that captured the digital cancer care experiences of older Black adults.

Results: In total, 5 overarching themes describe the experience of older Black adults, caregivers, and health care providers in accessing and using digital cancer care: (1) barriers to access and participation in digital care services, (2) shifting caregivers' dynamics, (3) autonomy of choice and choosing based on the purpose of care, (4) digital accessibility, and (5) effective digital communication. We identify 8 barriers and 6 facilitators to optimal digital cancer for older Black adults. Barriers include limited digital literacy, linguistic barriers in traditional African or Caribbean languages, and patient concerns of shifting power dynamics when supported by their children for digital cancer care; and facilitators include community-based cancer support groups, caregiver support, and key features of digital technologies.

Conclusions: These findings revealed a multifaceted range of barriers and facilitators to digital cancer care for older Black adults. This means that a multipronged approach that simultaneously focuses on addressing barriers and leveraging community strengths can improve access and usage of digital cancer care. A redesign of digital cancer care programs, tailored to the needs of most structurally marginalized groups like older Black adults, can enhance the digital care experience for all population groups. Public policies and organizational practices that address issues like availability of internet in remote areas, resources to support linguistic barriers, or culturally sensitive training are important in responding to the complexity of access to digital l cancer care. These findings have implications for other structurally marginalized and underresourced communities that have suboptimal access and usage of digital care.

(J Med Internet Res 2025;27:e63324) doi:10.2196/63324

XSL•FO
RenderX

KEYWORDS

digital care; cancer care; older Black adults; health equity; social determinants of health; access to care; health quality

Introduction

Digital care is defined as "the secure use of information and communication technologies in support of health and health-related fields, including health care services, health surveillance, health literature, and health education, knowledge, and research" [1]. Used in complement with or in lieu of in-person encounters, digital care encompasses a wide array of modalities, including phone calls, videoconferencing, remote monitoring, asynchronous messaging (eg, email and texting), and the use of a patient portal [1]. Digital cancer care has been shown to be effective for pretreatment discussions, the monitoring of adverse events, for counselling purposes, and for long-term cancer care follow-up [2]. However, there is a growing body of evidence that demonstrates inequitable access to digital care services, resulting in a widening health gap between population groups that are able to use digital care versus those that cannot. Reasons for this include an inability to afford the necessary technology, lack of language accessibility within certain applications, lack of infrastructure to support digital connectivity in rural areas, lack of cultural relevance of care provision, and lower levels of comfort in using certain technologies [3-6].

Structurally underresourced [7] and marginalized population subgroups, such as older Black adults, experience barriers specific to their identity, such as physical, cognitive, and sensory limitations or concerns [8]; lack of social or peer support to help with the use of technology [9]; and preference for in-person visits due to limited trust in technology and concerns about privacy [10]. When services are available, acceptability of care [11] is determined by a lack of cultural tailoring of language used in service and technology, such as not adapting the linguistic content of eHealth platforms to the cultural preferences of diverse populations [12], varying views or beliefs for disease management, such as cultural barriers when shifting from traditional in-person care to technology-based intervention platforms [13], and existing historic and societal barriers, such as general distrust of health care providers [14].

While other countries report that cancer disproportionately affects Black people [15], the lack of comprehensive race-based data collection in Canada means that the extent of the care gap affecting Black Canadians is unknown. However, a recent study that used census data to explore patterns of mortality inequities of Black adults in Canada indicates that Black men had an increased risk of dying from prostate cancer (hazard ratio: 1.33) than their White counterparts, and Black women were at an increased risk of dying from stomach cancer (hazard ratio 1.76), corpus uteri cancer (hazard ratio 1.78), and lymphomas and multiple myelomas (hazard ratio 1.26) than their White counterparts [16]. With limited race-based data collection, it is also unclear how race and racism affect digital cancer care use in Canada. It is essential that digital cancer care is delivered in a manner that meets the needs of all communities and does not exacerbate inequities between the most privileged and the most underresourced communities. Knowledge generated in partnership with these communities is essential to developing health care policy and planning that focuses on enhancing equity in access to and quality of digital care. Hence, the purpose of this study was to better understand the digital cancer care experiences of older Black cancer patients, their caregivers, and health care providers. Through this understanding, we sought to identify strategies that can support the delivery of better patient-centered digital cancer care.

Methods

Study Design

Qualitative research methods are appropriate for exploring the lived experiences and perspectives of individuals who are exposed to a social phenomenon like the usage of digital cancer care [17,18]. We used theoretical thematic analysis [17] as a methodology in order to conceptualize, collect, organize, and interpret data for this study. Specifically, we used a codebook approach to thematic analysis [18] that combines structure and reflexivity in methodological choices. In the following subsections of the methods, we describe how we structured our study approach by using 2 theoretical frameworks—the synergies of oppression lens [19] and the patient-centered care model [20]—to create a codebook and using multiple coders to ensure the accuracy of the coding process. The research team's interest in this study stems from a commitment to addressing systemic inequities and amplifying the voices of Black communities. We recognize our positionality influences every stage of the research process, from framing the research question to interpreting findings. Researcher subjectivity was addressed by regular team meetings to discuss and interpret data.

Theoretical Approach

The synergies of the oppression lens [19] is an analytic lens that can be used to illuminate the intersection of multiple systems of oppression that shape the lived or living experiences of older Black patients with cancer as they navigate the health system. Characteristics such as race, age, level of disability, sex at birth and gender, and income levels of individual and social groups are powerful indicators of their access to material and social resources that are necessary to achieve health and wellness. These characteristics (or multiple oppressions) operate in complex synergistic patterns that lead to poor health and have been used previously to unpack and understand inequities in cancer care [21-25]. In this study, we have used the synergies of oppression analytical lens to theorize a combination of oppressions impacting the digital care experiences of older Black patients with cancer. We were also guided by the 7 dimensions of Patient-Centred Care [20] model that outlines the essential components of a patient-responsive health care as (1) aligning the health system's vision, mission, values and quality improvement drivers to patient-centered goals; (2) providing collaborative, coordinated and accessible care at the right time and the right place; (3) providing care that considers physical comfort and emotional well-being; (4) respecting patient and family preferences, values, and cultural traditions

XSL•FO
RenderX

in the organization and delivery of care; (5) integrating family and patients in care teams, and having them fully partake in decision-making at the clinical and system level; (6) encouraging and facilitating the presence of family members in care settings; and (7) supporting informed decision-making by patients and their family by providing complete information in a timely manner. Accordingly, our study questions were as follows: What is the experience of older Black patients who have used digital cancer care? How can we improve the digital cancer care experience of older Black patients with cancer?

We hypothesize that strategies and interventions that will improve the digital cancer care experience of populations that are experiencing the most structural marginalization will improve the digital cancer care experience across all population groups in alignment with Bell Hooks's concept of "centering the margins" [26,27].

Participant Recruitment and Setting

This study was carried out between May 2022 and March 2023. A total of 3 groups of study participants were the focus of this study (Textbox 1).

Textbox 1. Three groups of study participants.

1. Patients: comprised of older Black adults, aged 55 years or older, given the growing health and social challenges of this age group, which includes chronic disease management and barriers in access to high-quality and timely care [28], who had been diagnosed with any form of cancer, including those no longer in active treatment, and had used digital cancer care services across the 10 Canadian provinces.
2. Caregivers: comprised of any individual who provided informal support in activities of daily living to older Black patients with cancer—this included family or paid caregivers.
3. Providers: comprised of health care professionals who regularly delivered clinical care to older Black patients with cancer.

Sampling was carried out through a nonprobabilistic convenience approach. This sampling strategy allowed us to cover various contexts (urban, rural, and remote) and diverse informants who could provide rich data on their experiences in receiving and delivering digital cancer care to older Black patients. The recruitment of potential study participants was supported by the Canadian Cancer Society (CCS), which is the largest cancer charity in Canada. Potential participant recruitment started in August 2022 and ended in January 2023. The CCS posted invitations to participate in the study using their database of patients with cancer, caregivers, and health care professionals. Those who were interested in participating in the study contacted the project coordinator by email—the research team received about 600 responses from potential participants. We sent detailed information letters on the study to all potential participants. In total, 6 groups of participants were created for each focus group, considering time zones and provincial linguistic diversity: (1) patients and caregivers in the Eastern provinces (Newfoundland and Labrador, Nova Scotia, and Prince Edward Island); (2) patients and caregivers in Central or Prairies provinces (Ontario and Manitoba); (3) patients and caregivers in Western or Prairies provinces (Saskatchewan, Alberta, and British Columbia); (4) patients and caregivers in Francophone/Bilingual provinces (Quebec and New Brunswick); (5) health care providers in 8 Canadian provinces; and (6) health care providers in Francophone or bilingual provinces (Quebec and New Brunswick); see Table 1. The final study comprised 55 participants across the 6 groups (Table 1). The final sample of 55 participants was sufficient to achieve code saturation in the qualitative data analysis, ensuring no new themes or codes emerged from the data [29,30].

Table 1. Number of participants in the focus groups.

Focus group	Participants, n
Patients with lived experience and caregivers	
Eastern Provinces (NF, NB, NS, and PEI)	12
Central & Prairie Provinces (ON and MB)	10
Western & Prairie Provinces (BC, AB, and SK)	10
Francophone or bilingual provinces (QC and NB)	8
Health care providers	
English-speaking provinces	10
Francophone or bilingual provinces	5
Total	55

Data Collection

Data were collected through 6 focus group interviews (about 120 minutes duration for each focus group) with the aforementioned groups of participants (Table 1). Focus group interviews were led by 2 Black peer researchers (PW and TM). This was an intentional and important strategy to reduce mistrust of the research process and to increase study participation by leveraging existing relationships between Black communities in Canada. During the focus group interviews, we used 2 distinct interview guides—a patient or caregiver interview guide (Multimedia Appendix 1) and a health care provider interview

XSL•FO
RenderX

guide (Multimedia Appendix 2)—to guide data collection. Both interview guides were developed from the research objectives and targeted literature review. Questions in the interview guides were structured to unpack issues of access and usage of digital cancer care and included probes relating to the roles of various actors in the usage of digital care, preference for various digital care modalities, frequency of digital care use, experiences, and challenges during the usage of digital care. All focus groups were carried out using the Zoom video teleconferencing software, and sessions were recorded with permission. The audio files of the recordings were transcribed externally by CMBusiness & Transcription Services in preparation for data analysis.

Following the focus group interview, a self-reported sociodemographic survey was sent to each focus group participant to better understand the diversity of age, income, and other sociodemographic characteristics that were present in the focus groups. Participation in the sociodemographic survey was optional.

Data Analysis

Qualitative data analysis followed Braun and Clarke's 6-phase thematic analysis approach, as outlined in their original [17] and updated [18] methodological work on codebook approaches to thematic analysis. NVivo 11 software facilitated the organization and coding of data. The analysis began in February 2023 and concluded in May 2023. To ensure familiarity with the data, all authors read and re-read the transcripts. A reflexive stance was maintained throughout the analysis, with the research team acknowledging their positionality, including assumptions and experiences that may influence interpretation. A codebook was initially developed based on the key elements of the patient-centered care model [20]. Line-by-line inductive and deductive coding was conducted independently by 2 researchers (PW and AS), with additional codes iteratively developed and refined during group discussions. Disagreements in coding or categorization were resolved through collaborative reflection and consensus, ensuring confirmability of the coding process. The final coding framework was applied to all transcripts, and excerpts were attributed to relevant codes. Detailed descriptions of each code were developed to deepen understanding of participants lived and living experiences. Themes and subthemes were generated through an iterative process of synthesis and review, prioritizing coherence, internal consistency, and nuanced representation of the data.

For methodological rigor and trustworthiness of our findings, we addressed key dimensions of qualitative research. *Credibility* was ensured through investigator triangulation, as multiple team members were involved in coding and the analysis process, with group discussions validating the consistency and interpretation of the data. To support *transferability*, we provide rich, thick descriptions of participants' experiences and contextual details, enabling readers to assess the applicability of our findings to other settings. *Dependability* was established by systematically documenting each stage of the coding and analysis process, including decisions made during data synthesis, to ensure transparency and reproducibility. Finally, *confirmability* was enhanced through reflexivity, where the research team acknowledged and reflected on their positionality and potential influences on the data interpretation, alongside a collaborative approach to consensus-building during coding and theme development. These steps reinforced the rigor of our analysis and ensured a nuanced, trustworthy representation of the lived and living experiences shared by participants.

Ethical Considerations

Institutional review board approval was provided for this research by the Toronto Women's College Hospital Research Ethics Board (reference 2022-0111-E). The research assistant sent a detailed informed consent form by email to all potential research participants. The researchers and research assistant responded to questions from potential participants by email or by phone conversations before they signed and returned their consent forms prior to the focus group interviews. Privacy and confidentiality of research participants were ensured by removing all identification information during data transcription and analysis; nonidentificatory codes were assigned to focus group interviews that were used in this manuscript. Focus group participants were provided the option to put on or off their camera during the conversation, and we reiterated the need to keep information shared during the conversation confidential at the onset of the focus group. All patients and caregivers who participated in the focus groups were given compensation of CAD $50 (US $35.16) for their time; health care providers were not provided any compensation for participating in the research. Participants had the right to withdraw from the study at any time, without any reasons.

Results

Sociodemographic Results

Of the 55 sociodemographic surveys that were sent to research participants, 24 were completed by patients and caregivers (Table 2). No health care providers completed the survey. All research participants who responded to the survey and whom we identified during the focus groups self-identified as Black people.

About two-thirds of patients and caregivers who participated in the focus group identified as male, and the same proportion of participants were between 50 and 64 years old. Most participants (20/21) had college-level education or higher. More than half of the participants had part-time jobs. There was great diversity in the household income level of participants, with more than half of them earning less than CAD $60,000 (CAD $1=US $0.7) per year.

XSL•FO
RenderX

Table 2. Sociodemographic profile of participants.

Characteristics	Number of participants[a]
Gender	
Male	16
Female	7
Age (years)	
50-64	15
65+	7
Highest level of education	
Primary or middle school	1
Undergraduate degree	7
College diploma or degree or certificate	6
Master's degree	7
Employment	
Full-time	6
Part-time	11
Self-employed	1
Casual	1
Household income (CAD $[b])	
0-29,999	6
30,000-59,999	4
60,000-89,999	3
90,000-119,999	4
120,000-150,000	2

[a]Numbers do not always add up to 24 because of incomplete surveys or missing data.

[b]CAD$1=US $0.70.

Thematic Results

The experiences of patients, caregivers, and health care providers on the use of digital cancer care for older Black adults were similar across the 4 patient or caregiver groups and the 2 provider groups. In total, 5 main themes emerged from this qualitative dataset: (1) barriers to access and participation in digital care services, (2) shifting caregiver dynamics, (3) autonomy of choice and choosing based on the purpose of care, (4) digital accessibility, and (5) effective digital communication.

Barriers to Access and Participation in Digital Care Services

Our findings reflected that older Black patients with cancer are not a homogenous group with respect to their lived experience of digital care. They are a heterogenous group composed of individuals with unique needs and lived experiences. Additionally, individuals within this group will have experienced systems of oppression to various degrees, which has shaped the way they experienced digital cancer care. Our data show 4 main challenges that were experienced by older Black patients:

(1) *Multimorbidity* seemed to be the most salient challenge that influenced the capacity of older Black patients with cancer to participate in virtual encounters. Across the 6 focus groups, study respondents often talked about 2 main features of multimorbidity that limited the capacity of patients to effectively engage in virtual encounters. First, the severity of the patient's cancer, marked by clinical symptoms such as intense chronic pain, loss of autonomy in activities of daily living, and persistent mental fatigue, meant that older Black patients were not always capable of fully participating in virtual encounters. Second, another aspect of multimorbidity was related to the adverse effects of chemotherapy, as explained by a caregiver.

> *I think sometimes we have to consider the patient's state of mind, maybe how long the person just finished taking a medication; the amount of time before you could engage them in a virtual meeting […] you have to look at the possibility of the patient being stable before you could engage into any important communication so that you might not end up getting misinformed by the patient just because of the reactions of medications.* [FG 1—Caregiver]

Several respondents pointed out that chemotherapy is exhausting for most patients with cancer. Older Black patients lost appetite, had nausea and vomiting as well as fatigue after rounds of

XSL•FO
RenderX

chemotherapy. These patients could not effectively participate in virtual meetings without assistance.

(2) The *geographic location* of patients emerged as another issue of concern for digital cancer care as explained by a patient:

> *And I will say maybe most times people in the rural areas tend not to have a good stable communication network system.* [FG 4—Patient]

Participants highlighted issues related to internet connection in rural and remote areas. Internet was often unreliable—they may spend days without any connections so they cannot join the virtual meeting; unstable—they may have an intermittent internet connection, so they are often disconnected from the virtual meeting; or low bandwidth—the poor quality of internet meant the connection is very slow.

(3) *Socially isolated* older Black patients were reported as having difficulties in accessing digital care. One respondent talked about a community-based initiative called the African Cancer Support Group that was launched to support socially isolated individuals of African and Caribbean descent in Alberta.

> *I say that because we have these groups called the African Cancer Support Group to help the friends of Africans or Caribbean and Black Canadians going through cancer. […] We're able to provide some patients with computers.* [FG 4—Caregiver]

This support group identified lack of access to digital technologies such as computers, smartphones, and tablets as an obstacle to digital l care for socially isolated Black patients. One way in which they supported this disadvantaged group was by providing computers to enhance access to digital care.

(4) *Linguistic barriers* emerged as a unique challenge to digital care, as explained by a health care provider:

> *Language barrier is kind of something that is digging deep to the Blacks because we have different kinds of languages.* [FG 5—Health care provider]

Our study participants pointed out that some older Black patients with cancer were not fluent in English or French, the official languages of Canada. Health care providers were also concerned that Black seniors seemed to be more proficient in traditional African or Caribbean languages such as Swahili or Patois. These languages are not easily understood by health care professionals, and it was difficult to find trained interpreters that could facilitate communication with these patients.

Shifting Caregiver Dynamics

Across the 6 focus groups, there was increasing recognition of the central role of caregivers in enabling and supporting digital cancer care for older Black patients. Respondents pointed out 4 main aspects of the shifting roles of caregivers as a dynamic process:

(1) Caregivers were critical in supporting patients with limited digital literacy before, during, and after the virtual encounter:

> *But one of the things that was a concern for me was because my mom is elderly and she's not really tech savvy, so I had to take charge of all portals and mail and all of that.* [FG 2—Caregiver]

(2) Caregivers were more involved in direct caregiving to their patient—a role that was traditionally reserved for health care providers. While family caregivers with no prior experience or health care training seemed to be less comfortable in providing direct care, paid caregivers seemed to be more comfortable in providing direct care.

(3) As caregivers were more involved in supporting digital care and providing direct care to Black seniors, there were privacy concerns related to the roles of caregivers. Specifically, some autonomous patients preferred privacy during their virtual meetings:

> *Like I said earlier, he has prostate cancer, and I think most times there are moments he wants to say things that are personal. So, at the moment, I like to understand his privacy and I'll give him that privacy.* [FG 4—Caregiver]

Other patients with cognitive disorders who needed support from caregivers were not concerned about privacy:

> *So, I think in terms of privacy, because already my patient has a little bit of memory loss, so I think without me being there, I think most things would be left untold.* [FG 4—Caregiver]

And family caregivers, like a spouse supporting the partner in digital care, did not consider privacy an issue

> *I'm not worried about privacy, it's my husband. If the patient is the wife, what kind of privacy? And mostly when you are filling the form, you put all the information they're on the form, so they already know everything about you.* [FG 4—Caregiver]

(4) Some older Black patients expressed anxieties about their children who became caregivers. Specifically, the emerging role of caregivers gave rise to a shift in power or influence where caregivers are more involved in decision-making. Some older Black patients were confronted with their children having more influence in their health care—a role that children do not traditionally hold in these cultures and societies.

Autonomy of Choice and Choosing Based on the Purpose of Care

Focus group participants often positioned digital care dualities and in-person care as alternative and complementary means of delivering care for older Black patients with cancer.

(1) By using digital care, patients particularly appreciated the convenience and comfort of receiving care in their homes, as explained by a patient:

> *You can stay at the comfort of your home. And when you have a virtual conversation with your specialist, you feel this level of comfort. And you can tell your specialist how you're actually feeling.* [FG 4—Patient]

Three other benefits of digital care for patients that emerged from our data included easy access to their specialist, feeling safe by not being exposed to the general population during the

XSL•FO
RenderX

COVID-19 pandemic, and older Black adults felt more empowered and confident to participate in their care.

(2) Caregivers and health care providers often talked about and greatly appreciated the flexibility of digital care in adjusting their work schedules.

(3) Despite these benefits, our respondents recognized that digital care was not always realistic or possible in the continuum of cancer care. They explicitly pointed out areas of their cancer care where digital care was impractical. One patient pointed out the need for a physician examination:

> *I'll say that I prefer in person because while I'm communicating in person, there are some physical examinations and some things I may want to explain myself, which I wouldn't be able to do virtually so I prefer in person.* [FG 1—Patient]

(4) Some focus group members argued for more hybrid models of care delivery that integrated digital care and other care modalities like home care and ambulatory care.

Digital Accessibility

Study participants recognized that an important aspect of digital care meant that they had to frequently use information and communication technologies for their routine care needs. While most of our study participants were relatively comfortable with technology, certain areas of concern emerged from the interviews.

Digital literacy was identified as a problem for all participant groups—patients, caregivers, and health care providers. Older Black adults with limited digital literacy had difficulties in understanding and using computers, tablets, or applications. They needed support from caregivers and health care professionals. Interestingly, some caregivers and health care professionals also faced digital literacy issues, as explained by a health care provider:

> *Yes, actually, it was more time intensive, and it was more stressful because I wouldn't say I'm that good at tech myself.* [FG 6—Health care provider]

The limited capacities of participants to use digital technologies could be frustrating in preparing and operating virtual encounters.

Some caregivers and health care providers mentioned that they had to invest time and efforts to improve their digital literacy skills and competencies in order to be ready to effectively support their patients.

> *During the pandemic, I had to learn a lot of technology skills and at the time I was involved with a virtual meeting with my patient, helping him set up virtual care with his GP and with his oncologist and every kind of health professionals involved with my patient.* [FG 2—Caregiver]

Certain health care organizations offered training resources to help health care providers to improve tech skills.

Finally, we noticed that older Black patients often said that they could not easily build their digital literacy skills due to their age and the burden of cancer on their daily lives. This means that the older Black patients were more reliant on their caregivers for their digital care needs. In this situation, patients seemed to value or prefer caregivers that are compassionate and patient in the way they provided support and assistance during virtual encounters:

> *I think for me, say one of the important things that I benefited from my caregiver, she's very patient with me. That is one thing I benefit from her. Because I think at the time, it took me almost a month to be able to navigate using virtual care. She was there for me, and her patience was something that ... I mean, I can't comprehend on the time, because she took the time to put me through, stage by stage, for me to understand what I was doing, and how to do it.* [FG 3—Patient]

Effective Digital Communication

As focus group participants frequently talked about the importance of digital technologies as a medium for good communication and information exchange with their health care team, there were some concerns about confidentiality and privacy. Specifically, we noticed an aspect of privacy where older Black patients were reluctant to share their health information with their children, as explained by a patient:

> *Yeah, when I had my kids around there were times when I was uncomfortable because there were some things I couldn't answer out loud, so I just did mute it when the question was asked and sometimes I would just tell my doctor, not now (inaudible) when we have a physical appointment or something, or maybe later when it's just me and my wife. If it's just my wife, I can say anything I want to say, but when my kids are around I have to limit what I say because I don't want them having this idea that something terrible is going to happen to dad and all that.* [FG 2—Patient]

These older Black patients seemed to be less concerned about sharing information with their spouses.

Our focus group participants also identified other areas where digital technologies could enhance communication, such as (1) having transcripts of conversations after the virtual encounter, (2) live translations for people that speak different languages, or (3) including or using features of technology to improve care for people with visual or hearing impairments:

> *And when it comes to hearing impairment, I think there's something you use in your ear that could help improve your hearing aid. So, when you're looking at things, you have to look at these tools equally so that you don't know whether you're dealing with a patient that has a hearing problem or an eye problem.* [FG 2—Caregiver]

A summary of barriers and facilitators to digital cancer care based on interview findings is shown in Textbox 2.

XSL•FO
RenderX

Textbox 2. Barriers and facilitators to digital cancer care for older Black adults.

Barriers to digital cancer care

- Limited engagement in virtual encounters of patients fatigued by cancer and chemotherapy or radiotherapy
- Difficulties connecting to virtual encounters due to limited and unstable internet connections in rural and remote areas
- Socially isolated older Black patients had limited access to digital technologies
- Difficult communication between patients and providers during virtual encounters due to linguistic barriers in traditional African or Caribbean languages
- Privacy concerns when caregivers supported competent patients during virtual encounters
- Difficulties in using digital care technologies due to limited digital literacy of patients, caregivers and providers
- Concerns of shifting power dynamics when children support their parents for digital cancer care
- Digital care was impractical in emergency situations

Facilitators to digital cancer care

- Community-based cancer support groups provided computers to socially isolated older Black patients
- Caregivers supported patients with limited digital literacy before, during, and after the virtual encounter
- Patients felt comfortable receiving digital cancer care in their homes
- Digital cancer care offered great flexibility to the work schedules of caregivers and health care providers
- Building digital literacy skills of caregivers and providers to provide better support to older Black patients
- Key features of digital technologies, such as visual and hearing aids or the capacity to have transcripts of virtual meetings, motivated actors to uptake digital cancer care

Discussion

Principal Findings

Inequities in cancer outcomes can be linked to higher cancer risk, delayed diagnosis, and unequal opportunities to access timely treatment. This is more significant for certain population groups that experience marginalized social conditions that are shaped by stigma, discrimination, and intergenerational trauma [15,31]. Inequities in cancer care can be further exacerbated at the intersection of social identities with other marginalizing conditions such as unequal access to the social determinants of health, resulting in health inequities that are structured and reinforced by the health care system [23]. One example of this can be the rapid shift of cancer care services from traditional in-person care towards digital cancer care models that arose in response to social distancing measures during the COVID-19 pandemic [32,33]. The implications of this shift—and the resulting impact in terms of access and optimal digital cancer care for populations that experience marginalizing social conditions—are still unknown. In our study, we sought to unpack these issues so that we could inform the redesign of digital cancer care to meet the needs and priorities of underresourced communities—thereby enabling better digital cancer care for everyone.

Prior systematic reviews reveal increasing evidence of the effectiveness of digital cancer care, which is comparable to in-person care, in counseling, and supportive medical care for people with cancer [2] and the increasing adherence of vulnerable populations, including older people and Black populations, to digital care technologies [9,34]. People living with cancer and health care providers were grateful for the convenience of videoconferencing because of reduced travel time to clinics and less travel expenses [2]. There was increasing recognition of digital health literacy challenges for older adults with investments in training older adults on the use of digital technologies and adapting these technologies to the unique needs of this population [2,9,34]. Our findings concur with previous studies on the barriers and facilitators of digital cancer care. Specifically, we identified 8 barriers to an optimal digital cancer encounter, such as linguistic barriers and limited digital literacy, and 6 facilitators that can enhance the digital cancer care experiences of older Black adults, such as community-based cancer support groups and flexible work schedules. Efforts to improve digital cancer care will require a multipronged approach that targets these specific barriers to digital care while creating system-wide policies that encourage these facilitators and influence the distribution of social determinants of health. For instance, policy interventions may be required to improve the stability of internet connections in rural and remote areas, while interpretation resource guides may help health care providers to know what translation and interpretation resources are available to them at the local or provincial and national level to overcome linguistic barriers. While these multipronged interventions will improve the digital cancer care experience for older Black adults, they are also equity-promoting for other underresourced communities and can enhance care for all.

We elaborate on 2 key findings of our study. First, our findings revealed an increasing recognition of the critical role of caregivers in supporting older Black adults in accessing and using digital cancer care. Older Black adults who are fatigued due to cancer multimorbidity, who had limited digital literacy, or who had cognitive disorders due to cancer relied on the support of caregivers before, during, and after virtual encounters.

XSL•FO
RenderX

However, we also noticed that autonomous patients had privacy concerns when supported by caregivers for their digital cancer care needs. This implies that there is an urgent need for guidance at the policy and organizational level on the roles and scope of practice of caregivers in supporting digital cancer care that is tailored to the unique needs of patients with cancer.

Second, previous studies suggest that the cultural background of racialized patients influences their views on symptoms, diagnosis, and treatments that may hinder access to cancer care [31]. Our findings unraveled concerns of older Black patients with cancer who were confronted with their children having more influence in their health care during virtual encounters and who were reluctant to share their health information with their children. These dynamics between older Black patients with cancer and their children have to be taken into consideration during digital cancer care. Framing virtual cancer care in the context of a person's cultural understanding of health requires the health care provider to have an awareness of the community that they serve [31]. This implies that efforts to develop culturally sensitive guides to digital care may be an important step to supporting health care providers to provide culturally appropriate digital cancer care to older Black patients.

Strengths and Limitations

This study has several strengths. The use of the Synergies of Oppression framework [19] provided a robust lens to examine how intersecting systems of oppression influence access to digital cancer care for older Black patients. Including participants from 10 Canadian provinces enhanced the transferability of findings by reflecting diverse regional contexts. The focus on an understudied population—older Black adults—alongside perspectives from caregivers and health care providers offered a multistakeholder view that highlights barriers, facilitators, and opportunities for improving digital care systems. Applying Braun and Clarke's thematic analysis [17,18] ensured a systematic and reflexive analysis, while the findings hold strong policy and practice relevance for advancing equitable, patient-centered care. Reflexivity throughout the research process further strengthened analytical transparency.

However, the study also has limitations. Conducting focus groups virtually may have introduced selection bias, as individuals with limited digital access were unable to participate. While race concordance between facilitators and participants likely fostered trust, discordance in gender, class, and affiliations may have influenced participation dynamics, and the presence of patient-caregiver dyads in the sessions may have constrained the expression of dissenting views. Additionally, regional differences across provinces were not analyzed, and findings may not reflect the full diversity within Black communities given variations in intersecting identities such as gender, class, and migration status.

Data limitations included the lack of completed socio-demographic surveys from health care providers, which restricted exploration of social factors shaping their experiences. Member checking was not conducted, though reflexivity mitigated this by ensuring transparency and neutrality. Lastly, while focus groups facilitated shared insights, they may have limited the discussion of deeply personal topics. Future research using individual interviews and exploring regional and nonusers' experiences would provide further depth.

Despite these limitations, this study provides timely, actionable evidence to enhance digital cancer care for older Black adults. It highlights key barriers and facilitators, offers multistakeholder insights, and lays the groundwork for equitable improvements in digital health care delivery.

Conclusions

Older Black patients face multiple barriers, such as linguistic, privacy concerns, and limited digital literacy, to accessing and using digital cancer care. A multipronged approach that focuses on addressing barriers, encouraging facilitators, and creating culturally sensitive guides to digital care can form the basis of health system efforts to improve access to digital cancer care. A redesign of digital cancer care programs, tailored to the needs of marginalized social groups like older Black patients, can enhance the digital care experience for all population groups. Public policies and organizational practices that address issues like availability of internet in remote areas, resources to support linguistic barriers, or culturally sensitive training are important in responding to the complexity of access to digital cancer care.

Acknowledgments

The authors would like to acknowledge and thank all the participants who agreed to participate in this study for their valuable time and contributions. The authors would like to thank the Canadian Cancer Society for assisting in the recruitment of potential study participants. This study was commissioned by the Canadian Cancer Society, through funding they received from Merck Inc.

Data Availability

The datasets generated and analyzed during this study are not publicly available as they include confidential data, but are available upon request from the corresponding author, including the need of Women's College Hospital ethical vetting.

Authors' Contributions

AS, AL, PW, and SC contributed to conceptualization; PW and SC contributed to project administration. PW, TM, AS, SC, NM, and BO contributed to data curation and formal analysis; AS, AL, and SC contributed to funding acquisition; AL, AS, and PW

XSL•FO
RenderX

contributed to methodology; PW and AS contributed to writing the original draft; PW, AS, AL, SC, NM, BO, and TM contributed in reviewing and editing the paper; all authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

PW is a recipient of the Provost's Postdoctoral Fellowship from the University of Toronto. AS is a recipient of the Transition to Leadership Stream Award in Patient-Oriented Research from the Canadian Institutes for Health Research and is a Health Equity Expert Advisor to the Canadian Partnership Against Cancer (CPAC). AL is supported as Clinician Scientist by the Department of Family Medicine at the University of Toronto and as Chair of Implementation Science at the Peter Gilgan Centre for Women's Cancers at Women's College Hospital in partnership with the Canadian Cancer Society. All the other authors do not declare any competing interests.

Multimedia Appendix 1
Combined caregiver and patient discussion guide.
[DOCX File , 47 KB - jmir_v27i1e63324_app1.docx]

Multimedia Appendix 2
Virtual cancer care provider discussion guide.
[DOCX File , 27 KB - jmir_v27i1e63324_app2.docx]

References

1. Demaerschalk BM, Hollander JE, Krupinski E, Scott J, Albert D, Bobokalonova Z, et al. Quality frameworks for virtual care: expert panel recommendations. Mayo Clin Proc Innov Qual Outcomes 2023;7(1):31-44 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2022.12.001] [Medline: 36619179]
2. Singh S, Fletcher GG, Yao X, Sussman J. Virtual care in patients with cancer: a systematic review. Curr Oncol 2021;28(5):3488-3506 [FREE Full text] [doi: 10.3390/curroncol28050301] [Medline: 34590602]
3. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
4. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]
5. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
6. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
7. Ho K, Adams O, Sayani A, Cheema G. Defining "Essential Digital Health for the Underserved". Healthc Pap 2024;21(4):5-14. [doi: 10.12927/hcpap.2024.27276] [Medline: 38482653]
8. Batsis JA, DiMilia PR, Seo LM, Fortuna KL, Kennedy MA, Blunt HB, et al. Effectiveness of ambulatory telemedicine care in older adults: a systematic review. J Am Geriatr Soc 2019;67(8):1737-1749. [doi: 10.1111/jgs.15959] [Medline: 31066916]
9. Arsenijevic J, Tummers L, Bosma N. Adherence to electronic health tools among vulnerable groups: systematic literature review and meta-analysis. J Med Internet Res 2020;22(2):e11613 [FREE Full text] [doi: 10.2196/11613] [Medline: 32027311]
10. Fang M, Siden E, Korol A, Demestihas M, Sixsmith J, Sixsmith A. A scoping review exploration of the intended and unintended consequences of eHealth on older people: a health equity impact assessment. Hum Technol 2018:297-323 [FREE Full text] [doi: 10.17011/ht/urn.201811224835]
11. Sayani A, Ali MA, Dey P, Corrado AM, Ziegler C, Nicholson E, et al. Interventions designed to increase the uptake of lung cancer screening: an equity-oriented scoping review. JTO Clin Res Rep 2023;4(3):100469 [FREE Full text] [doi: 10.1016/j.jtocrr.2023.100469] [Medline: 36938372]
12. Bennett GG, Steinberg DM, Stoute C, Lanpher M, Lane I, Askew S, et al. Electronic health (eHealth) interventions for weight management among racial/ethnic minority adults: a systematic review. Obes Rev 2014;15 Suppl 4:146-158. [doi: 10.1111/obr.12218] [Medline: 25196411]
13. Jang M, Johnson CM, D'Eramo-Melkus G, Vorderstrasse AA. Participation of racial and ethnic minorities in technology-based interventions to self-manage type 2 diabetes: a scoping review. J Transcult Nurs 2018;29(3):292-307. [doi: 10.1177/1043659617723074] [Medline: 28826353]
14. Ashfaq A, Esmaili S, Najjar M, Batool F, Mukatash T, Al-Ani H, et al. Utilization of mobile mental health services among syrian refugees and other vulnerable arab populations-a systematic review. Int J Environ Res Public Health 2020;17(4):1295 [FREE Full text] [doi: 10.3390/ijerph17041295] [Medline: 32085422]

XSL•FO
RenderX

15. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer statistics for African American/Black people 2022. CA Cancer J Clin 2022;72(3):202-229 [FREE Full text] [doi: 10.3322/caac.21718] [Medline: 35143040]
16. Tjepkema M, Christidis T, Olaniyan T, Hwee J. Mortality inequalities of Black adults in Canada. Health Rep 2023;34(2):3-16 [FREE Full text] [doi: 10.25318/82-003-x202300200001-eng] [Medline: 36791269]
17. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text]
18. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
19. McGibbon E, McPherson C. Sociology Med 2011;10(1):59-86 [FREE Full text]
20. Gerteis M, Edgman-Levitan S, Daley D, Delbanco T. Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care. 1st ed. San Francisco: Jossey-Bass; 1993.
21. Sayani A. Socially based inequities in breast cancer care: intersections of the social determinants of health and the cancer care continuum. Crit Stud 2017;13(1):24-36. [doi: 10.51357/cs.v13i1.129]
22. Sayani A. Inequities in genetic testing for hereditary breast cancer: implications for public health practice. J Community Genet 2019;10(1):35-39 [FREE Full text] [doi: 10.1007/s12687-018-0370-8] [Medline: 29781042]
23. Sayani A. Health equity in national cancer control plans: an analysis of the Ontario cancer plan. Int J Health Policy Manag 2019;8(9):550-556 [FREE Full text] [doi: 10.15171/ijhpm.2019.40] [Medline: 31657177]
24. Sayani A, Dilney J, Kuhnke JL, McNeil T. "My Cancer Is Worth Only Fifteen Weeks"? A critical analysis of the lived experiences of financial toxicity and cancer in Canada. Int J Health Policy Manag 2022;11(9):1814-1822 [FREE Full text] [doi: 10.34172/ijhpm.2021.83] [Medline: 34634872]
25. Sayani A, Vahabi M, O'Brien MA, Liu G, Hwang SW, Selby P, et al. Perspectives of family physicians towards access to lung cancer screening for individuals living with low income - a qualitative study. BMC Fam Pract 2021;22(1):10 [FREE Full text] [doi: 10.1186/s12875-020-01354-z] [Medline: 33413135]
26. Hooks B. Feminist Theory? From Margin to Center. 3rd ed. New York: Routledge; 2015.
27. Sayani A, Cordeaux E, Wu K, Awil F, Garcia V, Hinds R, et al. Using the power wheel as a transformative tool to promote equity through spaces and places of patient engagement. BMJ Open 2024;14(3):e074277 [FREE Full text] [doi: 10.1136/bmjopen-2023-074277] [Medline: 38514144]
28. Shen J, Tong H, Fuller-Thomson E. Penalty versus premium: social disposition differentiates life satisfaction among living-alone immigrant and native-born older adults-findings from the Canadian Longitudinal Study on Aging (CLSA). Int J Aging Hum Dev 2024:1-26. [doi: 10.1177/00914150241268089] [Medline: 39136136]
29. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
30. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
31. Ezeife DA, Padmore G, Vaska M, Truong TH. Ensuring equitable access to cancer care for black patients in Canada. CMAJ 2022;194(41):E1416-E1419 [FREE Full text] [doi: 10.1503/cmaj.212076] [Medline: 36280246]
32. Berlin A, Lovas M, Truong T, Melwani S, Liu J, Liu ZA, et al. Implementation and outcomes of virtual care across a tertiary cancer center during COVID-19. JAMA Oncol 2021;7(4):597-602 [FREE Full text] [doi: 10.1001/jamaoncol.2020.6982] [Medline: 33410867]
33. Watson L, Qi S, Delure A, Link C, Photitai E, Chmielewski L, et al. Virtual cancer care during the COVID-19 pandemic in alberta: evidence from a mixed methods evaluation and key learnings. JCO Oncol Pract 2021;17(9):e1354-e1361 [FREE Full text] [doi: 10.1200/OP.21.00144] [Medline: 34351822]
34. Budhwani S, Fujioka J, Thomas-Jacques T, De Vera K, Challa P, De Silva R, et al. Challenges and strategies for promoting health equity in virtual care: findings and policy directions from a scoping review of reviews. J Am Med Inform Assoc 2022;29(5):990-999 [FREE Full text] [doi: 10.1093/jamia/ocac022] [Medline: 35187571]

Abbreviations

CCS: Canadian Cancer Society

Edited by A Coristine; submitted 17.06.24; peer-reviewed by RS Bergmans, C Zeigler-Johnson; comments to author 22.11.24; revised version received 23.12.24; accepted 21.01.25; published 19.02.25.

Please cite as:
Wankah P, Chandra S, Lofters A, Mohamednur N, Osei B, Makuwaza T, Sayani A
Improving Digital Cancer Care for Older Black Adults: Qualitative Study
J Med Internet Res 2025;27:e63324
URL: https://www.jmir.org/2025/1/e63324
doi:10.2196/63324
PMID:

XSL•FO
RenderX

©Paul Wankah, Shivani Chandra, Aisha Lofters, Nebila Mohamednur, Beverley Osei, Tutsirai Makuwaza, Ambreen Sayani. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effectiveness of Digital Animation–Based Multistage Education for Patients With Atrial Fibrillation Catheter Ablation: Randomized Clinical Trial

Xiaoyu Shi[1,2,3,4,5*], MD; Yijun Wang[1,2,3,4,5*], MD; Yuhong Wang[1,2,3,4,5*], MD, PhD; Jun Wang[6*], MD; Chen Peng[1,2,3,4,5*], MD; Siyi Cheng[1,2,3,4,5], MD; Lingpeng Song[1,2,3,4,5], MD; Rui Li[1,2,3,4,5], MD; Fuding Guo[7], MD; Zeyan Li[1,2,3,4,5], MD; Shoupeng Duan[1,2,3,4,5], MD; Xiaomeng Yang[1,2,3,4,5], MD; Liping Zhou[1,2,3,4,5], MD, PhD; Hong Jiang[1,2,3,4,5], MD, PhD; Lilei Yu[1,2,3,4,5], MD, PhD

[1]Department of Cardiology, Renmin Hospital of Wuhan University, Wuhan, China

[2]Institute of Molecular Medicine, Renmin Hospital of Wuhan University, Wuhan, China

[3]Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan, China

[4]Hubei Key Laboratory of Autonomic Nervous System Modulation, Wuhan University, Wuhan, China

[5]Hubei Key Laboratory of Cardiology, Cardiovascular Research Institute, Wuhan University, Wuhan, China

[6]Department of Cardiology, The First Affiliated Hospital of Bengbu Medical University, Bengbu, China

[7]Department of Cardiology, Yan'an Affiliated Hospital of Kunming Medical University, Kunming, China

[*]these authors contributed equally

Corresponding Author:
Lilei Yu, MD, PhD
Department of Cardiology
Renmin Hospital of Wuhan University
No. 238 Jiefang Road
Wuhan, 430060
China
Phone: 86 02788041911
Email: lileiyu@whu.edu.cn

Abstract

Background: Digital education for outpatient patients with atrial fibrillation (AF) has gradually increased. However, research on digital education for patients undergoing atrial fibrillation catheter ablation (AFCA) is limited.

Objective: This study aimed to develop a novel digital animation-based multistage education system and evaluate its quality-of-life benefits for patients with AFCA.

Methods: This randomized controlled clinical trial included 208 patients with AF who underwent catheter ablation in the Department of Cardiology at Renmin Hospital of Wuhan University between January 2022 and August 2023. The patients were randomly assigned to the digital animation intervention (n=104) and standard treatment (n=104) groups. The primary outcome was the difference in the quality of life of patients with atrial fibrillation (AF-QoL-18) scores at 3 months. Secondary outcomes included differences in scores on the 5-item Medication Adherence Report Scale (MARS-5), Self-rating Anxiety Scale (SAS), and Self-Rating Depression Scale (SDS) at 3 months.

Results: In the digital animation intervention group, the AF-QoL-18 score increased from 38.02 (SD 6.52) to 47.77 (SD 5.74), the MARS-5 score increased from 17.04 (SD 3.03) to 20.13 (SD 2.12), the SAS score decreased from 52.82 (SD 8.08) to 45.39 (SD 6.13), and the SDS score decreased from 54.12 (SD 6.13) to 45.47 (SD 5.94), 3 months post discharge from the hospital. In the conventional treatment group, the AF-QoL-18 score increased from 36.97 (SD 7.00) to 45.31 (SD 5.71), the MARS-5 score increased from 17.14 (SD 3.01) to 18.47 (SD 2.79), the SAS score decreased from 51.83 (SD 7.74) to 47.31 (SD 5.87), and the SDS score decreased from 52.78 (SD 5.21) to 45.37 (SD 6.18). The mean difference in AF-QoL-18 score change between the 2 groups was 1.41 (95% CI 2.42-0.40, *P*=.006) at 3 months. The mean difference in MARS-5 score change was 1.76 (95% CI 2.42-1.10, *P*<.001). The mean difference in SAS score was –2.91 (95% CI –3.88 to –1.95, *P*<.001). Additionally, the mean difference in SDS score was –1.23 (95% CI –0.02 to –2.44, *P*=.047).

XSL•FO
RenderX

Conclusions: Our study introduces a novel digital animation educational approach that provides multidimensional, easily understandable, and multistage education for patients with AF undergoing catheter ablation. This educational model effectively improves postoperative anxiety, depression, medication adherence, and quality of life in patients at 3 months post discharge.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400081673; https://www.chictr.org.cn/showproj.html?proj=201059

(J Med Internet Res 2025;27:e65685) doi:10.2196/65685

KEYWORDS

animation education; digital health care; atrial fibrillation; catheter ablation; video; mHealth; digital care; digital health; digital animation; randomized clinical trial; RCT; digital education; outpatient; AFCA; atrial fibrillation catheter ablation; therapeutic; cardiac arrhythmia; Asian; animations; comics

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia and significantly increases the risk of mortality, stroke, heart failure, and dementia, thereby severely impacting patient quality of life (QoL) [1]. It is estimated that by 2050, at least 72 million Asian individuals will be diagnosed with AF [1,2]. The primary goals to treat AF include symptom improvement, heart rate and rhythm control, and stroke risk reduction [3,4]. Currently, catheter ablation has progressively emerged as the primary therapeutic approach for rhythm control in AF, owing to its potential to mitigate AF episodes and enhance patients' QoL [3,5]. However, the efficacy of catheter ablation in reducing AF remains uncertain with regard to its impact on patients' QoL.

Studies have indicated that negative emotions, alcohol consumption, smoking, obesity, and sleep disorders, among other high-risk factors, can expedite AF progression and prognosis, potentially directly impacting the efficacy of atrial fibrillation catheter ablation (AFCA) and postoperative recovery [6]. Additionally, anticoagulation therapy significantly reduces the stroke risk in patients with AF and is emerging as a crucial factor affecting treatment efficacy [7,8]. Therefore, alongside enhancing the safety and QoL benefits of AFCA, strategies focusing on risk factor management and medication adherence to improve ablation outcomes are essential in patients with AF [9,10].

Popular science interventions, such as animations and comics, are a noninvasive, entertaining, emerging approach to disease intervention used to ameliorate patient emotions, sleep, lifestyle, and disease prognosis [11-13]. Patients can vividly comprehend various aspects of AF, including etiology, pathology, surgical procedures, and complications, using vivid and engaging animation content, thereby enhancing their interest and engagement in health management. A recent study suggested that using animation education can enhance patients' understanding of AF and increase their satisfaction with clinical care [14].

With the advancement of internet technology, an increasing number of patients obtain information regarding AF via web-based videos. However, most videos vary in the direction and quality of their focus and offer limited benefits to patients undergoing AFCA [15]. Hence, new and more accessible digital animations are needed, and patients should be provided with digital education during the hospitalization and postdischarge phases of catheter ablation to facilitate management at different stages of AF. In this prospective, randomized, controlled clinical trial, we aimed to evaluate the benefits of multistage education based on digital animation for patients with AF undergoing catheter ablation.

Methods

Study Design

The trial protocol followed the guidelines outlined in the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1) for clinical trial reporting. Written informed consent was obtained from all patients. The complete protocol and relevant materials are presented in Multimedia Appendix 2.

Study Population

This randomized controlled trial enrolled 208 patients with AF who underwent AFCA in the Department of Cardiology at Renmin Hospital of Wuhan University between January 2022 and August 2023. The primary inclusion criteria were as follows: (1) aged between 18 and 90 years, (2) confirmed diagnosis of AF via surface electrocardiogram (ECG) or ambulatory ECG examination, and (3) patients with AF who received AFCA for the first time and were successfully treated during their stay at Renmin Hospital of Wuhan University. The procedures for both groups were conducted by the same AFCA team, ensuring consistency in the ablation techniques used. All the AF diagnosis and treatment criteria strictly adhered to the AF management [16]. Participants failing to meet the study criteria, as determined by the clinical investigator, were excluded if they scored below the designated threshold on the Mini-Mental State Examination or had a medical history of percutaneous coronary intervention or coronary artery bypass grafting. The detailed inclusion and exclusion criteria are provided in Multimedia Appendix 2.

Randomization and Masking

A web-based randomization tool was used to allocate 208 patients undergoing radiofrequency ablation to either the digital animation intervention group or the standard treatment group, with equal probability. Stratification based on gender and age (>60 or ≤60 years) was conducted to ensure a balance in baseline characteristics between the 2 groups. The blinding of participants was unfeasible given the inherent nature of the intervention.

XSL•FO
RenderX

Study Procedures

Participants were assessed at baseline (upon admission) and 3 months post discharge. During the recruitment process, baseline datasets containing sociodemographic characteristics and medical histories were obtained from the digital animation intervention and control groups. Standardized questionnaire surveys were conducted, including the quality of life of patients with atrial fibrillation (AF-QoL-18) [17,18], Self-rating Anxiety Scale (SAS) [19], Self-Rating Depression Scale (SDS) [20] and 5-item Medication Adherence Report Scale (MARS-5) [21,22]. Information regarding at-home medication use was collected from the patients in both groups at the time of discharge. Patients completed the standardized questionnaires at the 3-month follow-up. Patient characteristics and outcomes were collected by reviewing medical records of trained clinical physicians and participant self-reports. Multimedia Appendix 2 contains details of data acquisition in the study protocol.

Study Interventions

Based on the recommendations of the guidelines for treating AF, doctors provided all patients with standardized treatment and knowledge regarding AFCA [16]. In the digital animation intervention group, doctors conducted regular knowledge lectures to patients and used customized digital animations during treatment to educate them, ensuring that patients effectively received relevant information regarding AF treatment at appropriate stages. Figure 1 provides a visual summary of digital animation education for AF.

Figure 1. This study presents a visual summary of a digital animation-based educational framework for AF management, structured into 4 distinct phases. The first phase, admission, involves the initial education of patients about AF. The second phase, preoperative, employs digital animation to elucidate the surgical procedure. The third phase, postoperative, offers comprehensive guidance on medication and lifestyle modifications. The final phase, discharge, focuses on developing a structured follow-up plan. The digital animation comprehensively covers essential topics such as AF Knowledge, Check Description, Surgery Explained, Drug Guidance, Lifestyle Advice, Follow-Up Plan, and Things to Notice. The outcomes of the study demonstrate significant improvements, including AF-QoL-18 (P=.006), MARS-5 (P<.001), SAS (P<.001), and SDS (P=.047). AF: atrial fibrillation; AF-Qol-18: atrial fibrillation effect on quality-of-life scale; MARS-5: 5-item Medication Adherence Rating Scale; SAS: Self-rating Anxiety Scale; SDS: Self-Rating Depression Scale.

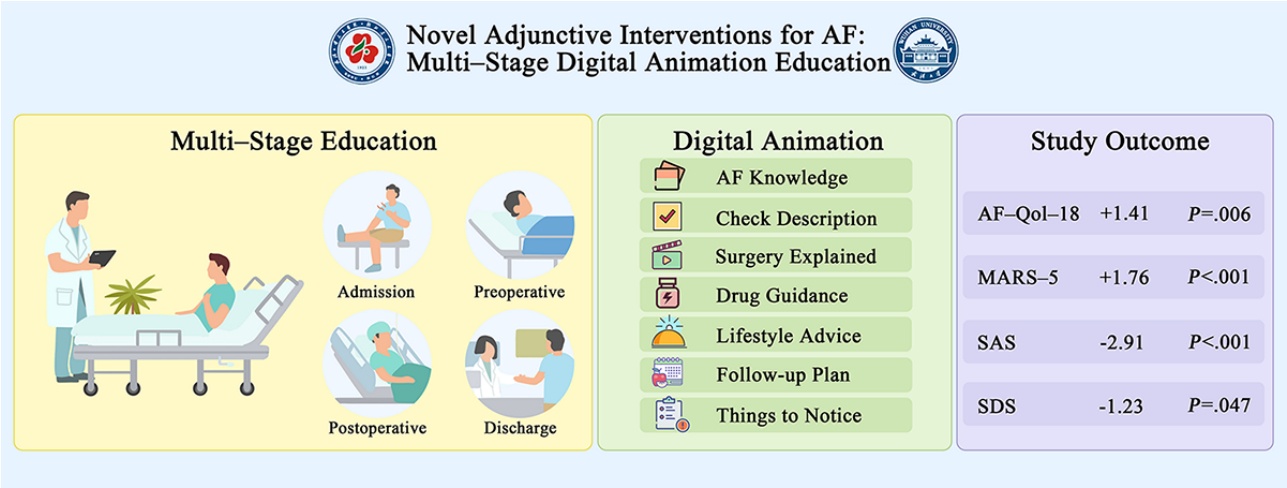

In the first phase, physicians provided standardized iPad devices (Apple) to facilitate playback and guided patients in the digital animation intervention group to watch 4 AF-related animations during admission, preoperative, postoperative, and discharge periods. These animations encompassed various aspects as follows: (1) imparting knowledge on cardiac anatomy, mechanisms underlying AF occurrence, and associated risk factors, as well as explaining the necessary examinations before catheter ablation and their significance; (2) discussing the purpose, process, and precautions of catheter ablation surgery; (3) outlining postoperative care requirements during hospital stay following catheter ablation; (4) emphasizing medication adherence upon discharge along with lifestyle recommendations and plans for regular follow-up visits. Furthermore, physicians addressed any queries raised by patients after viewing the animations to enhance comprehension. Patients in the control group solely received routine preoperative discussions without access to digitally animated educational videos at different stages. All patients completed a comprehensive study scale upon admission.

In the second phase, patients in the intervention and control groups were followed up 3 months after discharge by telephone to complete the study questionnaire. Multimedia Appendix 2 provides a detailed description of the intervention activities and their implementations.

All animation contents were directed and produced under the guidance of multiple cardiologists with extensive expertise. This ensured that the animations included important information about the diagnosis and treatment of diseases, surgeries, and crucial aspects of home-based rehabilitation, including improvements in sleep, emotional care, dietary adjustments, medication adherence, and activity recommendations. To minimize barriers to health education dissemination and language obstacles, animations were created based on elementary literacy levels, with audio and subtitles provided in standard Mandarin Chinese. However, each animation was strictly controlled for <5 min. Additionally, regular quality control and feedback mechanisms were implemented to ensure intervention effectiveness.

XSL•FO
RenderX

Regardless of the study group, all patients received standard hospital education from health care professionals and could discuss any questions or concerns regarding AF treatment with their primary physicians. All patients underwent AFCA and were followed for 3 months post discharge.

Measurements

BMI, calculated as weight (kg) divided by height (m^2), was obtained from the self-reported height and weight.

Measures related to psychometrics and psychological aspects, including the AF-QoL-18, SAS, SDS, and MARS-5, were used. The AF-QoL-18 was used to assess QoL in patients with AF. The SAS was used to assess individual levels of anxiety. The SDS was used to measure depression severity. The MARS-5 was used to evaluate patient medication adherence. The details of the QoL scores and related quality controls are listed in Multimedia Appendix 2.

Outcomes

Clinical outcome measurements were obtained at the 3-month follow-up. The primary outcome was the difference in AF-QoL-18 scores at 3 months, while the secondary outcomes included differences in MARS-5, SAS, and SDS scores. The definitions of the clinical outcomes and details regarding quality control are presented in Multimedia Appendix 2.

Ethical Considerations

This randomized controlled clinical study was approved by the Research Ethics Committee of Renmin Hospital of Wuhan University (approval number WDRY2021-K161) and registered with the Chinese Clinical Trial Registry (registration number ChiCTR2400081673). The trial was conducted in accordance with good clinical practice guidelines, the Declaration of Helsinki, and all relevant laws and regulations in China.

Written informed consent was obtained from all participants prior to enrollment. The consent forms detailed the study objectives, procedures, potential risks, and benefits, as well as data use policies. The informed consent documents have been uploaded as part of this submission (Multimedia Appendix 2).

No identifiable images of participants are included in the manuscript or supplementary materials. Identifiable information was removed from the datasets prior to analysis. All study data were anonymized and deidentified to protect participant privacy. All consent forms explicitly authorized the use of anonymized data for research publication.

The study involved only standard clinical care for AFCA and a digital animation-based multistage education program, which is categorized as routine and noninvasive patient education. In accordance with local regulations for low-risk clinical trials and the ethical approval guidelines of Renmin Hospital of Wuhan University, no financial compensation was provided to participants due to the minimal-risk nature of the study.

Statistical Analysis

The sample size was estimated by comparing 2 independent proportions with a type I error rate (α) set of 2.5%, an intracluster correlation coefficient of 0.01, and an anticipated retention rate of 80%. A significant absolute difference of 20% in clinical outcomes was detected between the intervention and standard treatment groups, with a statistical power of 90% at a follow-up period of 3 months. Based on our calculations, a minimum sample size of 80 assessable patients per group was required to achieve the desired statistical power.

Statistical analyses were conducted using SPSS (version 23.0; IBM Corp) and R (version 4.2.2; R Foundation for Statistical Computing) software. Data with normal distribution are presented as mean (SD). The 2 groups were compared using the independent sample *t* test. Data with nonnormal distributions are presented as medians and interquartile ranges. The Mann-Whitney *U* test was used to analyze the differences between the groups. Categorical data are expressed as numbers and percentages and compared using the chi-square test or Fisher exact test, as appropriate. To assess the effectiveness of the intervention, the differences between the 2 groups were calculated by subtracting the change from baseline to 3 months in the control group from that in the intervention group. Then, an independent *t* test was used to compare differences between groups before and after the intervention. A $P<.05$ was considered statistically significant.

Results

Figure 2 illustrates participant recruitment and retention. Among 208 patients undergoing catheter ablation (average age of 60.9 years), 68 (32.7%) were female and 140 (67.3%) were male. At baseline, the participants were randomly assigned to the digital animation intervention (n=104) and standard treatment (n=104) groups, both of which had similar characteristics. Table 1 presents the baseline characteristics of the digital animation intervention and the standard treatment groups. Table S1 in Multimedia Appendix 3 reveals significant statistical improvements in the AF-QoL-18, MARS-5, SAS, and SDS scores (all $P<.001$) at 3 months compared with baseline. In the digital animation intervention group, the AF-QoL-18 score increased from 38.02 (SD 6.52) to 47.77 (SD 5.74; $P<.001$; Figure 3A), the MARS-5 score increased from 17.04 (SD 3.03) to 20.13 (SD 2.12; $P<.001$; Figure 3B), the SAS score decreased from 52.82 (SD 8.08) to 45.39 (SD 6.13; $P<.001$; Figure 4A), and the SDS score decreased from 54.12 (SD 6.13) to 45.47 (SD 5.94; $P<.001$; Figure 4B). In the conventional treatment group, the AF-QoL-18 score increased from 36.97 (SD 7.00) to 45.31 (SD 5.71; $P<.001$), the MARS-5 score increased from 17.14 (SD 3.01) to 18.47 (SD 2.79; $P<.001$), the SAS score decreased from 51.83 (SD 7.74) to 47.31 (SD 5.87; $P<.001$), and the SDS score decreased from 52.78 (SD 5.21) to 45.37 (SD 6.18; $P<.001$; Table S1 in Multimedia Appendix 3).

The AF-QoL-18 ($P=.002$), MARS-5 ($P<.001$), and SAS ($P=.02$) scores of the digital animation intervention and conventional treatment groups were significantly different at 3 months. However, the SDS score ($P=.90$) did not exhibit a statistically significant difference (Table S2 in Multimedia Appendix 3). Additionally, the mean difference in AF-QoL-18 score change between the 2 groups was 1.41 (95% CI 2.42-0.40) at 3 months ($P=.006$; Table S3 in Multimedia Appendix 3). The mean difference in MARS-5 score change was 1.76 (95% CI

XSL•FO
RenderX

2.42-1.10, *P*<.001; Table S3 in Multimedia Appendix 3). The mean difference in SAS score was –2.91 (95% CI –3.88 to –1.95, *P*<.001; Table S3 in Multimedia Appendix 3). Furthermore, the mean difference in SDS score was –1.23 (95% CI –0.02 to –2.44), indicating an estimated difference between the digital animation intervention and conventional treatment groups at 3 months (*P*=.047; Table S3 in Multimedia Appendix 3).

At the 3-month follow-up, all patients underwent 12-lead ECGs and 24-hour ambulatory ECG monitoring to assess rhythm control. Our findings indicated that 9 patients in the digital animation intervention group experienced atrial arrhythmic events, including AF, atrial flutter, and atrial tachycardia, compared with 13 patients in the control group. However, the difference in the recurrence rates of these arrhythmic events between the 2 groups was not statistically significant (*P*=.50; Table S4 in Multimedia Appendix 3). Additionally, the mean heart rate was 70.08 bpm in the digital animation intervention group and 70.85 bpm in the control group. The difference in mean heart rate between the 2 groups was also not statistically significant (*P*=.25; Table S4 in Multimedia Appendix 3). Neither group experienced serious cardiovascular events, and no adverse reactions related to the medications were reported in either group.

Figure 2. Screening, randomization, and follow-up flowchart. Our study recruited patients diagnosed with atrial fibrillation who were scheduled for catheter ablation at the Department of Cardiology, Renmin Hospital of Wuhan University, between January 2022 and August 2023. Following a rigorous screening process, eligible participants were randomly assigned to either the digital animation intervention, which received digital animation-based education, or the standard treatment groups, which received standard care. The follow-up period was 3 months. A total of 208 patients with complete data were included in the final analysis. MMSE: Mini-Mental State Examination.

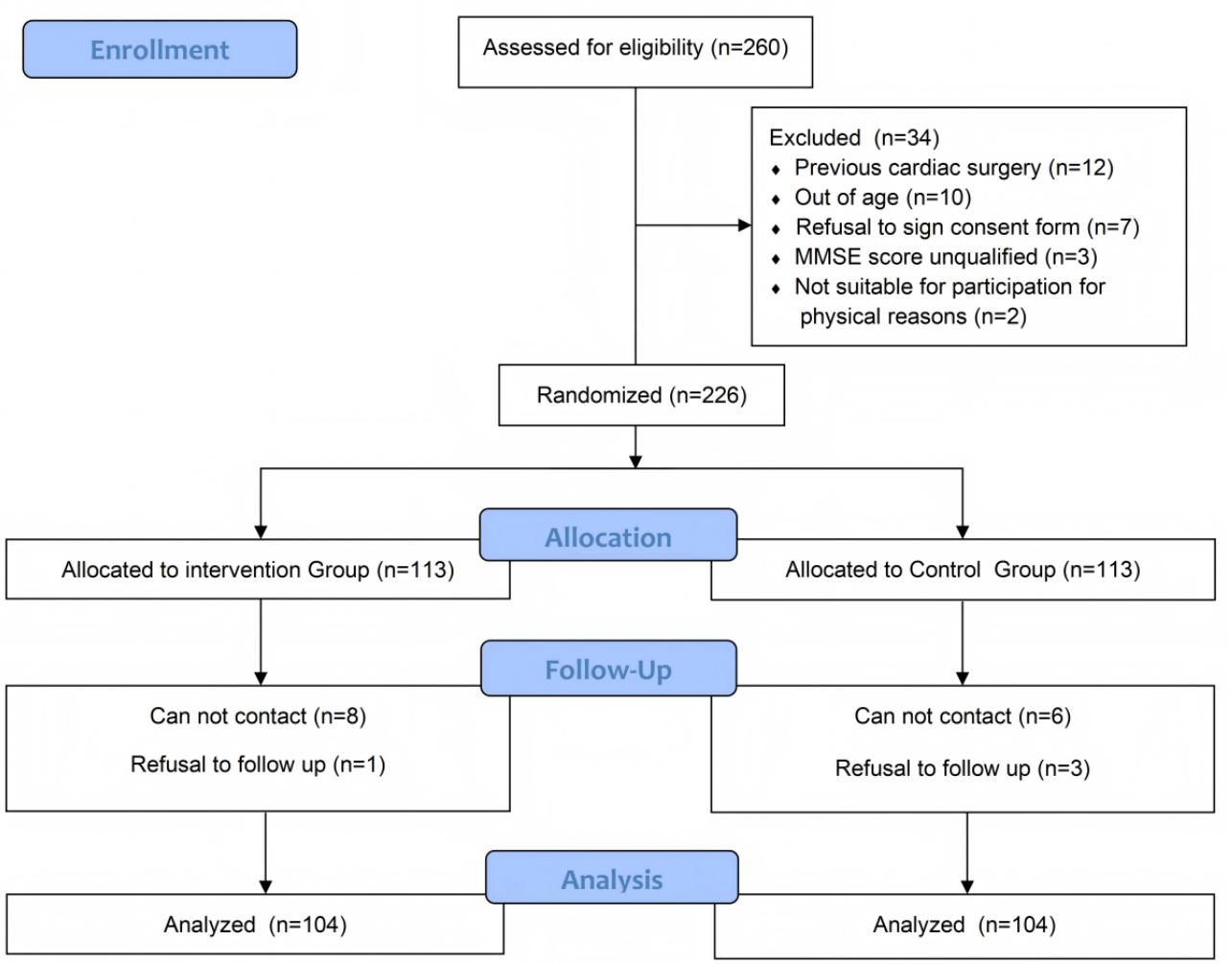

XSL•FO
RenderX

Table 1. Demographic and baseline clinical characteristics at baseline.

Characteristic	Intervention group (n=104)	Control group (n=104)
Male, n (%)	70 (67.3)	70 (67.3)
Age (years), mean (SD)	61.18 (10.90)	60.70 (11.33)
BMI, (kg/m^2), mean (SD)	24.61 (3.07)	23.96 (3.20)
Marital status, n (%)		
Married	97 (93.3)	98 (94.2)
Single	2 (1.9)	4 (3.8)
Widowed	5 (4.8)	2 (1.9)
Smoking status, n (%)		
Smoker	17 (16.3)	16 (15.4)
Former smoker	18 (17.3)	18 (17.3)
Nonsmoker	69 (66.3)	70 (67.3)
Drinking status, n (%)		
Drinker	16 (15.4)	27 (26.0)
Former drinker	8 (7.7)	7 (6.7)
Nondrinker	80 (76.9)	70 (67.3)
Education, n (%)		
Elementary school	25 (24.0)	23 (22.1)
Middle/high school	56 (53.8)	55 (52.9)
University degree	23 (22.1)	26 (25.0)
Coronary artery disease, n (%)	26 (25.0)	25 (24.0)
Hypertension, n (%)	44 (42.3)	53 (51.0)
Diabetes, n (%)	21 (20.2)	20 (19.2)
Admission heart rate, mean (SD)	80.81 (16.06)	81.68 (18.21)
Admission systolic blood pressure (mm Hg), mean (SD)	129.22 (18.44)	125.59 (17.76)
Admission diastolic blood pressure (mm Hg), mean (SD)	76.82 (11.88)	76.58 (10.51)
Hospitalization duration (day), mean (SD)	10.97 (4.04)	11.28 (3.94)
Type of atrial fibrillation, n (%)		
Paroxysmal atrial fibrillation	69 (66.3)	65 (62.5)
Persistent atrial fibrillation	35 (33.7)	39 (37.5)
Medication at discharge, n (%)		
Amiodarone	81 (77.9)	80 (76.9)
Warfarin	39 (37.5)	37 (35.6)
Rivaroxaban	55 (52.9)	58 (55.8)
Atorvastatin	58 (55.8)	60 (57.7)
β-blockers	30 (28.8)	38 (36.5)
Angiotensin-converting enzyme inhibitor or angiotensin II receptor blocker	30 (28.8)	34 (32.7)
Calcium channel blocker	14 (13.5)	16 (15.4)

XSL•FO
RenderX

Figure 3. Comparison of AF-QoL-18 and MARS-5 scores between the digital animation intervention and standard treatment groups in patients with atrial fibrillation at baseline and 3-month follow-up. AF-Qol-18: atrial fibrillation effect on quality-of-life Scale; MARS-5: 5-item Medication Adherence Rating Scale.

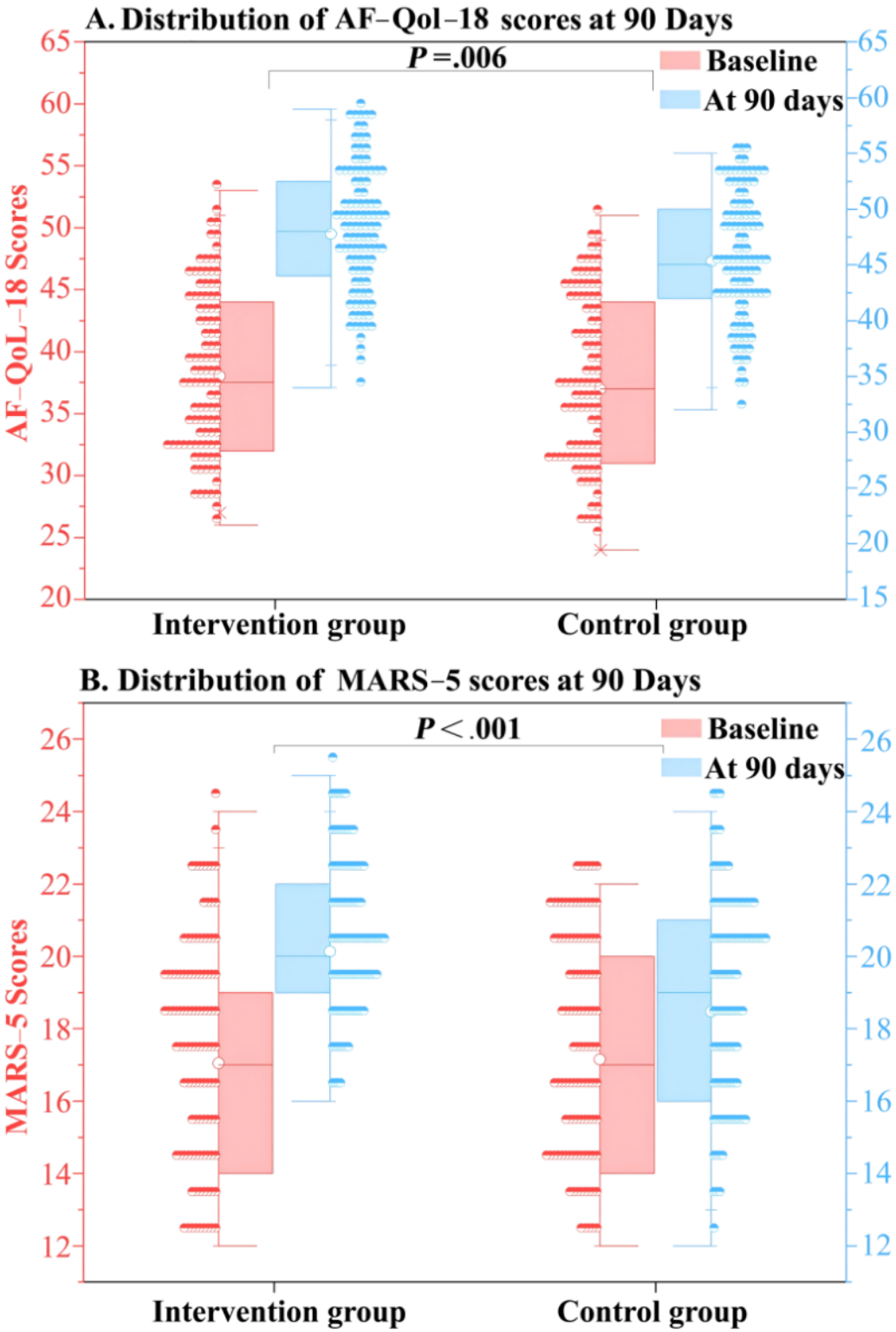

XSL•FO
RenderX

Figure 4. Comparison of SAS and SDS scores between the digital animation intervention and standard treatment groups in patients with atrial fibrillation at baseline and at 3 months follow-up. SAS: Self-rating Anxiety Scale; SDS: Self-Rating Depression Scale.

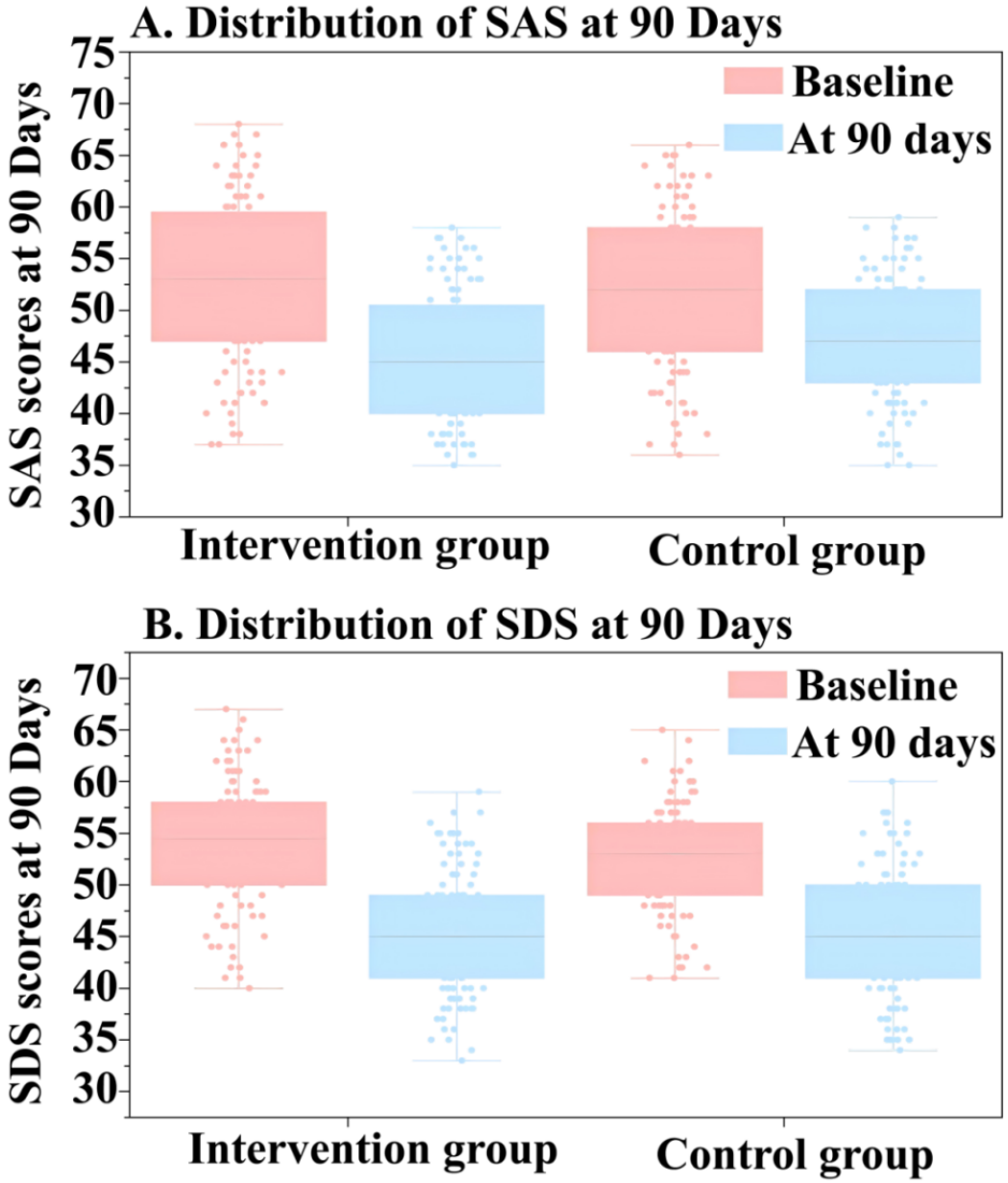

Discussion

Principal Findings

This study demonstrates that multistage education based on digital animation improves the QoL and medication adherence of patients with AF and helps them improve their emotional well-being compared with conventional treatment. Digital animation-based multistage education serves as an adjunctive tool, providing visual cardiovascular disease knowledge and explaining necessary diagnostic procedures, the AFCA process, and treatment-related considerations. It also emphasizes the importance of medication adherence during rehabilitation and provides recommendations for lifestyle and behavioral changes, thereby amplifying the QoL benefits of AFCA.

Comparison With Prior Work

In clinical practice, AFCA has emerged as a critical interventional strategy for rhythm control in patients with AF [5]. This procedure specifically targets and ablates ectopic triggering foci and abnormal electrical conduction pathways that are responsible for the initiation and perpetuation of AF [5]. By effectively disrupting these arrhythmogenic substrates, AFCA aims to restore and maintain normal sinus rhythm, thereby improving patient outcomes [5]. However, some patients experience suboptimal long-term outcomes [23]. This phenomenon may be attributed to a lack of understanding of AF, failure to eliminate risk factors, or poor medication adherence, directly impacting the effectiveness of AFCA and postoperative recovery [24-26]. Patient education regarding AF is crucial for reducing these adverse effects. Digital animation, a form of patient education, offers a more engaging and informative approach due to its audiovisual characteristics. This mode of patient education is more effective than traditional

XSL•FO
RenderX

methods, such as visual brochures or one-on-one verbal conversations [27-29]. Recently, a randomized clinical trial evaluated the impact of video education created by doctors on patients with AF [14]. The results indicated that video education effectively improved AF knowledge and nursing satisfaction among outpatients with AF.

Previous animations used to educate patients with AF primarily focused on AF risk factors, symptoms, complications, and pharmacological and nonpharmacological treatments. However, knowledge regarding the procedural aspects of patients undergoing ablation surgery [30,31]. Unfamiliarity with the surgical environment and lack of understanding of the surgical procedure are important factors contributing to preoperative anxiety among patients [32,33]. This results in preoperative sleep disorders and adversely impacts postoperative efficacy and recovery [34]. Explaining the purpose, process, and precautions of surgery can help alleviate preoperative anxiety during the waiting period. A randomized controlled study demonstrated that combining audiovisual methods to educate patients about preoperative preparation, intraoperative procedures, and postoperative intensive care unit admission effectively reduced preoperative anxiety among patients undergoing coronary artery bypass grafting surgery [35]. Additionally, research using virtual reality devices to provide patients with a first-person perspective of the operating room and surgical procedures found a reduction in preoperative negative emotions and increased postoperative satisfaction [33]. These results indicate that educating patients about surgical procedures can improve their emotional state and benefit them in various aspects, including mental well-being and QoL, during the perioperative period.

In this study, targeted digital animation was designed for patients with AF undergoing AFCA. Compared with traditional audiovisual educational materials for patients with AF, this animation included additional information regarding the purpose, surgical process, intraoperative aspects, and postoperative considerations of catheter ablation. These contents helped reduce patient anxiety during hospitalization, deepen patient understanding of AF, promote changes in unhealthy lifestyles, and improve patient medication adherence, ultimately contributing to QoL enhancement in patients with AF [36-38].

Numerous people have recognized the advantages of video education over traditional educational methods, and numerous studies have produced professional educational videos. However, in clinical practice, most patient education was conducted as a single event during outpatient or inpatient visits [30,31]. Regular and repeated education helps improve patient knowledge of the disease over time. Studies have demonstrated that the relevant knowledge acquired by patients with AF significantly declines 2 weeks after receiving education, while multiple educational sessions help enhance patients' awareness of the disease [14,39,40]. Compared with previous research, this experiment made a preliminary attempt by dividing surgery-related educational materials into preoperative and postoperative stages. Moreover, videos were sent to the patients for reinforced education after discharge. Instead of completing all educational content in a single session, this experiment used a multistage approach, effectively reducing the time required for patients to receive education, with each stage lasting <5 minutes. Patients with AF are mostly older people, particularly those with lower cognitive function, and receiving all educational information in one session can be challenging [41]. The short duration and highly focused nature of these videos ensured that the patients maximized their attention and minimized the opportunity for attentional decline. Furthermore, this staged intervention approach helped patients retain information from different stages of education to some extent, thereby improving medication adherence after discharge.

Limitations

This study has several limitations to consider when interpreting the results. First, the sample size was relatively small (208 patients), and the study was conducted in a single hospital. Therefore, the generalizability of the results to larger populations or different health care settings is limited and requires further investigation. Second, the follow-up period was only 3 months, which may not capture the long-term educational effects and stability of the digital animation intervention. A longer follow-up may provide a more comprehensive understanding of the sustained impact of educational interventions. Third, the perioperative 3-month period, clinically referred to as the "blanking period," is characterized by heightened vulnerability to diverse arrhythmic phenomena. This interval poses significant risks for patients, as electrophysiological instabilities can exacerbate underlying symptoms and compromise daily functional status. Therefore, meticulous monitoring is essential during this phase to optimize risk stratification and guide therapeutic decision-making. Fourth, the use of self-reported measures for medication adherence, anxiety, and depression may introduce response bias, as participants might overestimate their adherence or underreport symptoms due to social desirability. Fifth, while the digital animation system was designed with user-friendliness in mind, variations in individual patient engagement and technological proficiency may have exerted an influence on the efficacy of the intervention. It is important to acknowledge these limitations as they may impact the overall interpretation and generalizability of the findings of the study. Future research with larger and more diverse samples, longer follow-up periods, and objective outcome measures may help strengthen the evidence regarding the educational effectiveness of digital animation for patients with AF undergoing catheter ablation.

Conclusions

Our study introduced a novel approach to digital animation education, providing multidimensional and highly accessible multistage education for patients with AF undergoing catheter ablation. This educational model has demonstrated effectiveness in improving postoperative anxiety, depression, medication adherence, and QoL in patients at the 3-month postdischarge follow-up.

XSL•FO
RenderX

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (82270532 and 82241057); Foundation for Innovative Research Groups of Natural Science Foundation of Hubei Province, China (2021CFA010); Chutian Talent Program Science and Technology Innovation Team Project, Central University Basic Scientific Research Business Expenses Special Funds (2042024kf0041 and 2042022kf1211); Taikang Life Medicine Central PI Program of Wuhan University; and Talent Program of Renmin Hospital, Wuhan University (JCRCYG-2022-001). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

XS, Yijun Wang, Yuhong Wang, JW, and CP contributed equally to this study. LY conceived and performed experiments. XS, Yijun Wang, Yuhong Wang, JW, and CP performed the clinical study and wrote the manuscript. CP, SC, LS, RL, LZ, FG, SD, and XY collected and revised the data. HJ and LZ drafted the manuscript and critically revised it for important intellectual content. All authors contributed to the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1140 KB - jmir_v27i1e65685_app1.pdf]

Multimedia Appendix 2
Trial Protocol.
[PDF File (Adobe PDF File), 318 KB - jmir_v27i1e65685_app2.pdf]

Multimedia Appendix 3
Data Analysis.
[PDF File (Adobe PDF File), 101 KB - jmir_v27i1e65685_app3.pdf]

References

1. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. Int J Stroke 2021;16(2):217-221. [doi: 10.1177/1747493019897870] [Medline: 31955707]
2. Chiang CE, Chao TF, Choi EK, Lim TW, Krittayaphong R, Li M, et al. Stroke prevention in atrial fibrillation: a scientific statement of JACC: Asia (part 2). JACC Asia 2022;2(5):519-537 [FREE Full text] [doi: 10.1016/j.jacasi.2022.06.004] [Medline: 36624790]
3. Joglar J, Chung M, Armbruster A, Benjamin E, Chyou J, Cronin E, Peer Review Committee Members. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American college of cardiology/American heart association joint committee on clinical practice guidelines. Circulation 2024;149(1):e1-e156 [FREE Full text] [doi: 10.1161/CIR.0000000000001193] [Medline: 38033089]
4. Guo F, Wang J, Deng Q, Feng H, Xie M, Zhou Z, et al. Effects of pulsed field ablation on autonomic nervous system in paroxysmal atrial fibrillation: a pilot study. Heart Rhythm 2023;20(3):329-338. [doi: 10.1016/j.hrthm.2022.11.013] [Medline: 36435350]
5. Parameswaran R, Al-Kaisey AM, Kalman JM. Catheter ablation for atrial fibrillation: current indications and evolving technologies. Nat Rev Cardiol 2021;18(3):210-225. [doi: 10.1038/s41569-020-00451-x] [Medline: 33051613]
6. Pathak RK, Middeldorp ME, Lau DH, Mehta AB, Mahajan R, Twomey D, et al. Aggressive risk factor reduction study for atrial fibrillation and implications for the outcome of ablation: the ARREST-AF cohort study. J Am Coll Cardiol 2014;64(21):2222-2231 [FREE Full text] [doi: 10.1016/j.jacc.2014.09.028] [Medline: 25456757]
7. Guo Y, Kotalczyk A, Imberti JF, Wang Y, Lip GY, ChiOTEAF Registry Investigators. Oral anticoagulation improves survival in very elderly Chinese patients with atrial fibrillation: a report from the optimal thromboprophylaxis in elderly Chinese patients with atrial fibrillation (ChiOTEAF) registry. Int J Stroke 2022;17(6):661-668. [doi: 10.1177/17474930211046743] [Medline: 34551637]

XSL•FO
RenderX

8. Lip GYH, Proietti M, Potpara T, Mansour M, Savelieva I, Tse H, et al. Atrial fibrillation and stroke prevention: 25 years of research at EP Europace journal. Europace 2023;25(9) [FREE Full text] [doi: 10.1093/europace/euad226] [Medline: 37622590]
9. Chung MK, Eckhardt LL, Chen LY, Ahmed HM, Gopinathannair R, Joglar JA, American Heart Association Electrocardiography Arrhythmias Committee Exercise, Cardiac Rehabilitation, Secondary Prevention Committee of the Council on Clinical Cardiology; Council on Arteriosclerosis, Thrombosis Vascular Biology; Council on Cardiovascular Stroke Nursing; Council on Lifestyle Cardiometabolic Health. Lifestyle and risk factor modification for reduction of atrial fibrillation: a scientific statement from the American heart association. Circulation 2020;141(16):e750-e772 [FREE Full text] [doi: 10.1161/CIR.0000000000000748] [Medline: 32148086]
10. Polymeris AA, Meinel TR, Oehler H, Hölscher K, Zietz A, Scheitz JF, et al. Aetiology, secondary prevention strategies and outcomes of ischaemic stroke despite oral anticoagulant therapy in patients with atrial fibrillation. J Neurol Neurosurg Psychiatry 2022;93(6):588-598 [FREE Full text] [doi: 10.1136/jnnp-2021-328391] [Medline: 35396339]
11. Wang Y, Guo F, Wang J, Li Z, Tan W, Xie M, et al. Efficacy of a weChat-based multimodal digital transformation management model in new-onset mild to moderate hypertension: randomized clinical trial. J Med Internet Res 2023;25:e52464 [FREE Full text] [doi: 10.2196/52464] [Medline: 38048156]
12. Chang YJ, Chen JL. Effectiveness of a web-based intervention for preventing substance use in young adults in Taiwan: quasi-experimental study. J Med Internet Res 2023;25:e40157 [FREE Full text] [doi: 10.2196/40157] [Medline: 37578821]
13. Hermans ANL, Betz K, Verhaert DVM, den Uijl DW, Clerx K, Debie L, et al. 360° virtual reality to improve patient education and reduce anxiety towards atrial fibrillation ablation. Europace 2023;25(3):855-862 [FREE Full text] [doi: 10.1093/europace/euac246] [Medline: 36738261]
14. McIntyre D, Kovoor JG, Chow CK, Quintans D, Marschner S, Thomas S, et al. Clinician-created video education for patients with AF: a randomized clinical trial. JAMA Netw Open 2023;6(12):e2345308 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.45308] [Medline: 38064218]
15. Luo C, Bian L, Jiang L, Liang W, Wu Z. Does youtube provide qualified patient education videos about atrial fibrillation? Front Public Health 2022;10:925691 [FREE Full text] [doi: 10.3389/fpubh.2022.925691] [Medline: 35910923]
16. January C, Wann L, Calkins H, Chen L, Cigarroa J, Cleveland J, et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American college of cardiology/American heart association task force on clinical practice guidelines and the heart rhythm society in collaboration with the society of thoracic surgeons. Circulation 2019;140(2):e125-e151 [FREE Full text] [doi: 10.1161/CIR.0000000000000665] [Medline: 30686041]
17. Badia X, Arribas F, Ormaetxe JM, Peinado R, de Los Terreros MS. Development of a questionnaire to measure health-related quality of life (HRQoL) in patients with atrial fibrillation (AF-QoL). Health Qual Life Outcomes 2007;5:37 [FREE Full text] [doi: 10.1186/1477-7525-5-37] [Medline: 17610734]
18. Arribas F, Ormaetxe JM, Peinado R, Perulero N, Ramírez P, Badia X. Validation of the AF-QoL, a disease-specific quality of life questionnaire for patients with atrial fibrillation. Europace 2010;12(3):364-370. [doi: 10.1093/europace/eup421] [Medline: 20056594]
19. Zung WW. A rating instrument for anxiety disorders. Psychosomatics 1971;12(6):371-379. [doi: 10.1016/S0033-3182(71)71479-0] [Medline: 5172928]
20. Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965;12:63-70. [doi: 10.1001/archpsyc.1965.01720310065008] [Medline: 14221692]
21. Lin C, Ou H, Nikoobakht M, Broström A, Årestedt K, Pakpour AH. Validation of the 5-item medication adherence report scale in older stroke patients in Iran. J Cardiovasc Nurs 2018;33(6):536-543. [doi: 10.1097/jcn.0000000000000488] [Medline: 29649015]
22. Mahler C, Hermann K, Horne R, Ludt S, Haefeli WE, Szecsenyi J, et al. Assessing reported adherence to pharmacological treatment recommendations. translation and evaluation of the medication adherence report scale (MARS) in Germany. J Eval Clin Pract 2010;16(3):574-579. [doi: 10.1111/j.1365-2753.2009.01169.x] [Medline: 20210821]
23. Choi S, Yu H, Kim D, Park J, Kim T, Uhm J, et al. Late recurrence of atrial fibrillation 5 years after catheter ablation: predictors and outcome. Europace 2023;25(5) [FREE Full text] [doi: 10.1093/europace/euad113] [Medline: 37099677]
24. Ozaki AF, Choi AS, Le QT, Ko DT, Han JK, Park SS, et al. Real-world adherence and persistence to direct oral anticoagulants in patients with atrial fibrillation: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes 2020;13(3):e005969. [doi: 10.1161/CIRCOUTCOMES.119.005969] [Medline: 32148102]
25. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
26. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;2014(11):CD000011 [FREE Full text] [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
27. Marshall W, Rothenberger L, Bunnell S. The efficacy of personalized audiovisual patient-education materials. J Fam Pract 1984;19(5):659-663. [Medline: 6208312]

XSL•FO
RenderX

28. Ploderer B, Rezaei Aghdam A, Burns K. Patient-generated health photos and videos across health and well-being contexts: scoping review. J Med Internet Res 2022;24(4):e28867 [FREE Full text] [doi: 10.2196/28867] [Medline: 35412458]
29. Suresh Kumar S, Connolly P, Maier A. Considering user experience and behavioral approaches in the design of mHealth interventions for atrial fibrillation: systematic review. J Med Internet Res 2024;26:e54405 [FREE Full text] [doi: 10.2196/54405] [Medline: 39365991]
30. Kovoor JG, McIntyre D, Chik WWB, Chow CK, Thiagalingam A. Clinician-created educational video resources for shared decision-making in the outpatient management of chronic disease: development and evaluation study. J Med Internet Res 2021;23(10):e26732 [FREE Full text] [doi: 10.2196/26732] [Medline: 34633292]
31. Gagné M, Legault C, Boulet L, Charbonneau L, Lemyre M, Giguere AM, et al. Impact of adding a video to patient education on quality of life among adults with atrial fibrillation: a randomized controlled trial. Patient Educ Couns 2019;102(8):1490-1498. [doi: 10.1016/j.pec.2019.03.015] [Medline: 30956021]
32. Yu J, Zhang Y, Yu T, Mi W, Yao S, Wang Z, et al. Preoperative anxiety in Chinese adult patients undergoing elective surgeries: a multicenter cross-sectional study. World J Surg 2022;46(12):2927-2938 [FREE Full text] [doi: 10.1007/s00268-022-06720-9] [Medline: 36070012]
33. Chiu PL, Li H, Yap KY, Lam KC, Yip PR, Wong CL. Virtual reality-based intervention to reduce preoperative anxiety in adults undergoing elective surgery: a randomized clinical trial. JAMA Netw Open 2023;6(10):e2340588 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.40588] [Medline: 37906193]
34. Butris N, Tang E, Pivetta B, He D, Saripella A, Yan E, et al. The prevalence and risk factors of sleep disturbances in surgical patients: a systematic review and meta-analysis. Sleep Med Rev 2023;69:101786. [doi: 10.1016/j.smrv.2023.101786] [Medline: 37121133]
35. Nargiz Koşucu S, Şelimen D. Effects of music and preoperative education on coronary artery bypass graft surgery patients' anxiety. J Perianesth Nurs 2022;37(6):807-814. [doi: 10.1016/j.jopan.2021.12.002] [Medline: 35760716]
36. Braga Ferreira L, Lanna de Almeida R, Arantes A, Abdulazeem H, Weerasekara I, Ferreira LSDN, et al. Telemedicine-based management of oral anticoagulation therapy: systematic review and meta-analysis. J Med Internet Res 2023;25:e45922 [FREE Full text] [doi: 10.2196/45922] [Medline: 37428532]
37. He Y, Tang Z, Sun G, Cai C, Wang Y, Yang G, et al. Effectiveness of a mindfulness meditation app based on an electroencephalography-based brain-computer interface in radiofrequency catheter ablation for patients with atrial fibrillation: pilot randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e44855 [FREE Full text] [doi: 10.2196/44855] [Medline: 37133926]
38. Bilbrey T, Martin J, Zhou W, Bai C, Vaswani N, Shah R, et al. A dual-modality home-based cardiac rehabilitation program for adults with cardiovascular disease: single-arm remote clinical trial. JMIR Mhealth Uhealth 2024;12:e59098 [FREE Full text] [doi: 10.2196/59098] [Medline: 39150858]
39. McCabe PJ, Schad S, Hampton A, Holland DE. Knowledge and self-management behaviors of patients with recently detected atrial fibrillation. Heart Lung 2008;37(2):79-90. [doi: 10.1016/j.hrtlng.2007.02.006] [Medline: 18371501]
40. Ferguson C, Hickman LD, Lombardo L, Downie A, Bajorek B, Ivynian S, et al. Educational needs of people living with atrial fibrillation: a qualitative study. J Am Heart Assoc 2022;11(15):e025293 [FREE Full text] [doi: 10.1161/JAHA.122.025293] [Medline: 35876410]
41. Mbanda N, Dada S, Bastable K, Ingalill G, Ralf WS. A scoping review of the use of visual aids in health education materials for persons with low-literacy levels. Patient Educ Couns 2021;104(5):998-1017. [doi: 10.1016/j.pec.2020.11.034] [Medline: 33339657]

Abbreviations

AF: atrial fibrillation
AFCA: atrial fibrillation catheter ablation
AF-QoL-18: quality of life of patients with atrial fibrillation
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ECG: electrocardiogram
MARS-5: 5-item Medication Adherence Report Scale
QoL: quality of life
SAS: Self-rating Anxiety Scale
SDS: Self-Rating Depression Scale

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.08.24; peer-reviewed by A Khamees, S Chen; comments to author 30.09.24; revised version received 19.10.24; accepted 14.02.25; published 11.03.25.

Please cite as:
Shi X, Wang Y, Wang Y, Wang J, Peng C, Cheng S, Song L, Li R, Guo F, Li Z, Duan S, Yang X, Zhou L, Jiang H, Yu L
The Effectiveness of Digital Animation–Based Multistage Education for Patients With Atrial Fibrillation Catheter Ablation: Randomized Clinical Trial
J Med Internet Res 2025;27:e65685
URL: https://www.jmir.org/2025/1/e65685
doi:10.2196/65685
PMID:

©Xiaoyu Shi, Yijun Wang, Yuhong Wang, Jun Wang, Chen Peng, Siyi Cheng, Lingpeng Song, Rui Li, Fuding Guo, Zeyan Li, Shoupeng Duan, Xiaomeng Yang, Liping Zhou, Hong Jiang, Lilei Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Associations Among Online Health Information Seeking Behavior, Online Health Information Perception, and Health Service Utilization: Cross-Sectional Study

Hongmin Li[1], PhD; Dongxu Li[1], PhD; Min Zhai[1], Prof Dr; Li Lin[1], Prof Dr; ZhiHeng Cao[2], MPA

[1]School of Public Health, Jining Medical University, Jining, China

[2]Rencheng District Center for Disease Control and Prevention, Jining, China

Corresponding Author:
Dongxu Li, PhD
School of Public Health
Jining Medical University
No 133 Hehua Road
Taibaihu District, Shandong
Jining, 272067
China
Phone: 86 05373616333
Fax: 86 05373616333
Email: lidongxu0602@126.com

Abstract

Background: Seeking online health information can empower individuals to better understand their health concerns, facilitating their ability to manage their health conditions more effectively. It has the potential to change the likelihood and frequency of health service usage. Although existing literature has demonstrated the prevalence of seeking online health information among different populations, the factors affecting online health information perception and discussions on the associations between seeking online health information and health service utilization are limited.

Objective: We analyzed the associations between online health information seeking behavior and health service utilization, as well as the online health information perception delivery mechanism.

Methods: We analyzed data from the Chinese General Social Survey, the first national representative survey conducted in mainland China. The independent variable was the online health information seeking behavior. The outcome variable was health service utilization by the respondents, and online health information perception was selected as the mediating variable in this analysis. Factor analysis was conducted to obtain online health information perception. Multiple regressions were performed to investigate the effect of online health information seeking behavior on physician visits. Bootstrap methods were conducted to test the mediation effects of online health information perception.

Results: This analysis included 1475 cases. Among the participants, 939 (63.66%) sought online health information in the last 12 months. The mean age of the respondents was 46.72 (SD 15.86) years, and 794 (53.83%) were females. After controlling for other variables, individuals with online health information seeking behaviors showed 0.289 times more outpatient visits (P=.003), 0.131 times more traditional Chinese medicine outpatient visits (P=.01), and 0.158 times more Western medicine outpatient visits (P=.007) over the past year compared to those who did not seek health information online. Additionally, multiple regression analyses revealed statistically significant effects of gender, age, and health status on physician visits. The total effect revealed that seeking online health information significantly influenced the total physician visits (β=0.290; P=.003), indicating a certain correlation between online health information seeking behavior and physician visits. Seeking online health information had a significant positive impact on the perception (β=0.265; P<.001). The mediation effects analysis identified that online health information perception led to a significant increase in physician visits with the increase in the online health information seeking behaviors (β=0.232; P=.02).

Conclusions: The online health information perception of an individual influences the effect online health information seeking has on the frequency of physician visits. The online health information seeking behavior impacts outpatient service utilization both directly and indirectly through online health information perception and significantly increases the frequency of clinic visits

XSL•FO
RenderX

after controlling for other variables. Interventions can be explored to improve the health utilization of residents by increasing their online health information perception.

(J Med Internet Res 2025;27:e66683) doi:10.2196/66683

KEYWORDS

online health information seeking (OHIS); online health information perception (OHIP); mediating effect; health service utilization; health information; health perception; data; China; Chinese General Social Survey (CGSS); database; medical information; survey

Introduction

The internet has become an essential part of life around the world. In 2023, the number of internet users reached 5.3 billion globally. Internet penetration rates in Saudi Arabia, Norway, and the United Arab Emirates are approaching 100% [1]. In China, which is the leading market for internet users, the online penetration rate was 78% in June 2024, with 1.099 billion users [2]. There is no doubt that the internet has become the most common resource for health information due to the widespread use of smartphones. Our previous study found that even in rural and remote areas in China, 47.27% of the residents obtained health-related information or services through the internet [3].

Individuals can easily access internet effectively, obtain the relevant medical information available online, and make health care decisions based on their health status. Online health information seeking (OHIS) can empower individuals to better understand their health concerns, facilitating them to manage their health conditions more effectively. Recent studies on OHIS behavior have focused on the population characteristics and associated factors. OHIS behavior in different populations [4] have been discussed, such as college students [5,6], older adults [7,8], adolescents [9,10], mothers [11], and people with various health statuses [12]. OHIS behavior and its influencing factors have also been analyzed in patients with specific diseases such as cancer [13-15], osteoarthrosis [16], type 2 diabetes mellitus [17,18], and rare diseases [19]. The OHIS behavior is influenced by various factors, including demographics, psychological aspects, role-related or interpersonal traits, environmental sources, source-related characteristics, and diseases [13-19]. OHIS can enable individuals to assess their health concerns better, facilitating them to improve their management of health conditions and make more informed decisions about whether to use health services. Therefore, OHIS can potentially change the likelihood and the frequency of health service usage, such as outpatient visits. A systematic review suggested that the current evidence of the association between OHIS and health services is inconclusive [20]. The relationship between internet use and health service utilization is complex and influenced by multiple factors [13,14,21]. Existing empirical studies [21-23] on the effect of internet health information on demand for health care services have shown unclear and inconsistent findings. Some studies [21,22] found that access to online health information enhanced the demand and use of health services by individuals. An empirical analysis based on data from the United States Health Information Trends Survey revealed that OHIS has a positive, relatively large, and statistically significant effect on individual health care demand [21]. Another cross-sectional study [22] of 993 adults in Vietnam drew positive conclusions that the prevalent use of mass media sources, including the internet, for health information was associated with higher rates of health care utilization, even though there were issues such as neglecting the quality of health information. However, a study [23] also found that the behavior of searching for online health information can negatively impact the utilization of health services. A Chinese nationally representative data analysis based on an instrumental variable regression revealed that mobile internet use significantly diminished the inclination of an individual to seek primary health care services [23].

Seeking information on the internet can enhance the knowledge of individuals about health concerns and symptoms, enabling them to better manage their health conditions [24]. However, a vast volume of health information needs to be screened and evaluated, and the perception of people toward health information may influence this process. On the one hand, OHIS fulfills the need of patients for health information or provides opportunities to interact with health providers through the internet, which may decrease the frequency of health consultation visits. In contrast, OHIS draws the attention of patients to health issues, increasing their demand and frequency of visits to health services. Currently, a substantial body of theoretical work has delved into the factors influencing OHIS behavior [25]. Particular emphasis has been placed on the mechanisms by which the attitudes and perspectives of individuals regarding online health information shape their corresponding behaviors [25]. A systematic review integrated the comprehensive model of information seeking, the planned risk information seeking model, and the situational theory of problem-solving and found that the perception and trust of people in online information were the most dominant predictors of OHIS behavior [25]. For patients with chronic conditions, perceived susceptibility and severity can effectively explain perceived risk, helping predict their OHIS behavior. Informational and emotional support can help patients perceive benefits, thereby positively affecting their OHIS behavior [26]. Simultaneously, the OHIS attitude and behavior are dependent on the identity of those who offer opinions online; thus, the advice obtained from others with similar interests is given more weight [27]. A study involving 10,000 participants from mainland China found that perceived information quality mediates the influence of OHIS on health behavior [28].

The abovementioned studies have demonstrated that the impact of OHIS behavior on health service utilization requires considering attitudes, perspectives, and views of people on online health information, which has been confirmed as an important factor for the acceptance of health information [29]. How people perceive the health information available online influences subsequent processes on health behaviors or the use

XSL•FO
RenderX

of health services, for example, if they believe that information on the internet has a positive impact on their health, whether the online information helps them understand the doctor's advice, or whether the doctor provided appropriate advice. Although existing literature [13-19,21-23] has demonstrated OHIS prevalence among different populations and patients as well as the factors that influence OHIS, there has been minimal discussion on the associations between OHIS and health services. Furthermore, even fewer studies have explored the impact of this association [25]. More research is needed regarding the associations among people's attitudes, perceptions, OHIS, and health services.

The above discussion suggests that a step toward a comprehensive understanding of the effects described requires the development of a complex model that considers the simultaneous relationships among OHIS, health service utilization, and the attitude and perceptions of an individual. The research questions of this study focus on how OHIS behavior affects the practice of health service utilization and how the attitudes and perceptions of people toward online health information play a role in this process. We developed a joint model of OHIS, health service usage, and health information belief that considers an individual's unobserved characteristics that are likely to be correlated with health information seeking behavior, individual's health status, and health care utilization.

Methods

Study Participants

This study analyzes data from the Chinese General Social Survey (CGSS), which systematically monitors the changing relationship between the social structure and the quality of life in both urban and rural China. The CGSS is China's first, nationally representative and comprehensive continuous academic survey initiative conducted by an academic institution in mainland China [30]. This project collects data by using a stratified, multistage probability proportional to size method. Data are comprehensively gathered at multiple levels, including society, communities, households, and individuals. CGSS covers 28 provinces, autonomous regions, and municipalities across China. In this study, the latest CGSS2021 survey was selected as the data source for the analysis, which included the context on OHIS and the perceptions of respondents. The details of the sampling scheme and design are available on the official CGSS website [30].

A total of 8148 samples were included in the CGSS2021 data, with 2690 participating in the medical service–related questions (module D). After removing the missing values for the key variables such as online health information perception (OHIP) (n=856), income (n=342), education level (n=5), outpatient visits (n=4), overall evaluation of the health care system (n=7), and health status (n=1), 1475 study samples were obtained. The specific data processing procedures are represented in Figure 1.

Figure 1. Participant screening flowchart. OHIP: online health information perception.

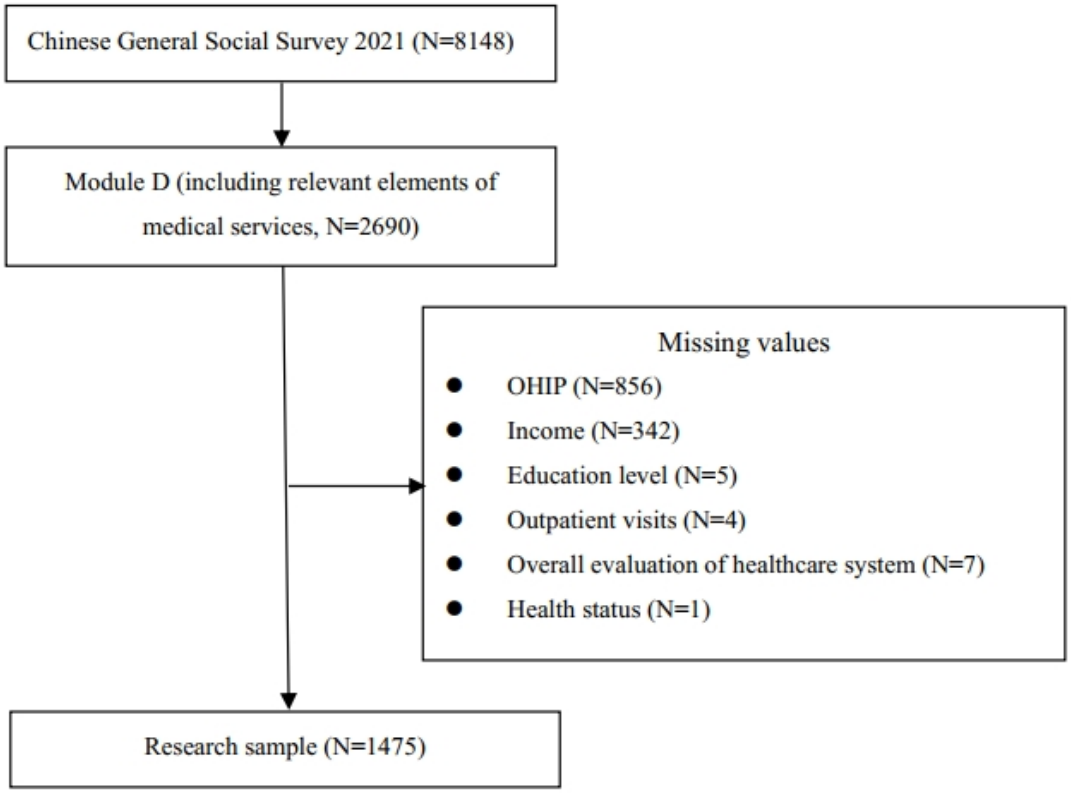

XSL•FO
RenderX

Ethics Approval

CGSS was approved by the ethics review committee of Renmin University of China. The study data were anonymous, and the answers were protected by privacy law. The data collection and storage protocols were in full compliance with the personal information protection law of China. All obtained data were anonymized, which means removing any identifiers that can directly or indirectly link the data to individual participants. Each participant signed an informed consent form at the time of participation. There was no requirement for additional ethics approval for approved data users.

Measures

Outcome Variable

The health service utilization by respondents served as the outcome variable for this analysis. This variable was represented by the question "In the last 12 months, how often have you visited a doctor?" The questionnaire was divided into 2 parts: visits to traditional Chinese medicine (TCM) physicians and visits to Western medicine physicians. The numbers 1, 2, 3, 4, and 5 were assigned to the answers never, rarely, sometimes, often, and very often, respectively. The 2 answers were combined to create the outpatient utilization frequency variable ranging from 2 to 10. OHIS behavior was selected as an independent variable in our analysis. OHIS was described by the question, "In the past 12 months, how often have you searched for health or medical information for yourself or others online in various forms?" OHIS was identified as yes if respondents selected several times a day, once a day, or a few times a week/month/year. Other responses (never go online, hardly ever, and cannot choose) were assigned to "no" in the analysis. Afterward, further analysis was conducted after removing individuals with an answer "no" on the OHIS behavior. For those with OHIS behavior, the numbers 1, 2, 3, 4, and 5 were assigned to the answers a few times a year, a few times a month, a few times a week, once a day, and several times a day, respectively.

Mediating Variables

The perceptions of respondents regarding online health information were selected as the mediating variables in this analysis. In the CGSS2021 questionnaire, 4 questions were selected to represent perception: "Do you agree with the following statement that information on the internet has positively influenced my health behaviors?", "Do you agree with the following statement that information on the internet has helped me understand what my doctor has told me?", "Do you agree with the following statement that the internet can help people decide if they need to go to the doctor or not?", and "Do you agree with the following statement that the internet can help to make sure that doctors are giving people the right advice?" The numbers 1, 2, 3, 4, and 5 were assigned to the responses strongly disagree, disagree, cannot tell, agree, and strongly agree, respectively. Factor analysis was conducted to extract the perception of respondents for online information from the 4 questions listed above. The factor with a contribution rate of 0.548 was extracted according to the criterion that the eigen value is greater than 1. The procedure and results of factor analysis are given in Multimedia Appendix 1.

Control Variables

The control variables selected for this study were as follows: gender (female=0, male=1), age, income, residence (rural=0, urban=1), marital status (unmarried=0, married=1), education level (primary or below=0, junior middle school=1, high school or above=1), health status (very poor=0, relatively poor=1, general=2, relatively good=3, good=4), region (western region=0, central region=1, and eastern region=2), and their overall evaluation of the health care system (not at all satisfied=0, very dissatisfied=1, quiet dissatisfied=2, cannot say=3, quite satisfied=4, very satisfied=5, and fully satisfied=6).

Statistical Analysis

Descriptive and univariate analyses were performed for the basic characteristics of respondents. A chi-square test was used for discrete variables, while *t* tests (2-tailed) were applied for normal or 2-sample data. Wilcoxon rank-sum tests were applied for nonnormal or unequal variance continuous variables. Additionally, ordinary least squares regression was conducted to examine the effect of OHIS on physician visits. Bootstrap methods were used to assess the mediation effects of health information perception; 95% CI and *P* values were presented. The weights provided by the CGSS data were utilized for regression models and mediation analysis in this study for ensuring generalizability and reliability. Statistical analysis was conducted using Stata (version 14.0) for Microsoft Windows (Stata Corp).

Results

Sample Characteristics

A total of 1475 cases were selected from the database for the analysis. The mean age of the respondents was 46.72 (SD 15.864) years, 794 (53.83%) were females, 1083 (73.42%) were married, and 562 (38.10%) were living in rural areas. A total of 420 respondents were western region residents, representing 28.47% (420/1475) of the total number of respondents, while 26.31% (388/1475) were central region residents, and 45.22% (667/1475) were eastern region residents. Almost half (728/1475, 49.36%) of the respondents had a high school diploma or more, while one-fifth (289/1475, 19.59%) had only completed primary school or were illiterate. More than half of the respondents (853/1475, 57.83%) were in relatively good or good health conditions. The frequency of physician visits in the last 12 months was 3.883 (SD 1.591); that of TCM visits was 1.787 (SD 0.884) and that of Western outpatient visits was 2.096 (SD 0.957). Among the respondents, 939 (63.66%) sought online health information in the last 12 months.

One-way analysis of the number of total clinic, TCM, and Western physician visits was conducted for the related variables. Preliminary analysis revealed differences among respondents with different health status, genders, and education levels. The female respondents visited physicians more frequently than male respondents in the last 12 months (t_{1473}=–3.411; *P*=.007), as evidenced by the number of visits for TCM outpatients

XSL•FO
RenderX

(t_{1473}=3.040; P=.02) and Western medicine outpatients (t_{1473}=2.842; P=.002). Those with primary school or less education had the highest number of physician visits ($F_{2,1472}$=5.37; P=.04), TCM visits ($F_{2,1472}$=6.46; P=.002), and Western clinic visits ($F_{2,1472}$=3.09; P=.046). As the health condition worsened, the overall number of physician visits increased ($F_{4,1470}$=52.33; P<.001), and the same trend of changes was observed for both TCM ($F_{4,1470}$=55.73; P<.001) and Western clinic visits ($F_{4,1470}$=24.03; P<.001). More details on the sample statistic descriptions are provided in Table 1.

XSL•FO
RenderX

Table 1. Characteristics of the participants in relation to physician visits (N=1475).

Indicators/variables	Values, n (%)	Physician visits			Traditional Chinese Medicine visits			Western clinic visits		
		Mean (SD)	*t* test *(df)*/*F* test *(df)*	*P* value	Mean (SD)	*t* test *(df)*/*F* test *(df)*	*P* value	Mean (SD)	*t* test *(df)*/*F* test *(df)*	*P* value
Gender[a]			–3.411 (1473)	.007		3.040 (1473)	.002		2.842 (1473)	.002
Female	794 (53.83)	4.01 (1.656)			2.166 (0.034)			1.847 (0.032)		
Male	681 (46.17)	3.73 (1.499)			2.014 (0.035)			1.716 (0.031)		
Education level[a]			5.37 (2, 1472)	.04		6.46 (2, 1472)	.002		3.09 (2, 1472)	.046
Primary and below	289 (19.59)	4.418 (1.734)			2.276 (1.037)			1.871 (0.946)		
Junior school	458 (31.05)	3.770 (1.571)			2.058 (0.989)			1.711 (0.867)		
High school and above	728 (49.36)	3.848 (1.534)			2.048 (0.894)			1.800 (0.866)		
Online health information seeking behavior[a]			–0.322 (1473)	.75		0.192 (1473)	.85		–0.789 (1473)	.43
No	536 (36.34)	3.865 (1.745)			2.102 (0.045)			1.763 (0.040)		
Yes	939 (63.66)	3.893 (1.497)			2.092 (0.029)			1.800 (0.027)		
Residence[a]			1.028 (1473)	.30		1.563 (1473)	.12		0.159 (1473)	.87
Rural	562 (38.10)	3.937 (1.639)			2.145 (0.042)			1.791 (0.037)		
Urban	913 (61.90)	3.849 (1.561)			2.065 (0.030)			1.784 (0.029)		
Health status[b]			52.33 (4, 1470)	<.001		55.73 (4, 1470)	<.001		24.03 (4, 1470)	<.001
Very poor	45 (3.05)	5.666 (1.906)			3.044 (1.205)			2.622 (1.113)		
Relatively poor	127 (8.61)	4.976 (1.738)			2.795 (1.078)			2.181 (1.072)		
Average	450 (30.51)	4.133 (1.607)			2.275 (0.985)			1.857 (0.928)		
Relative good	566 (38.37)	3.584 (1.377)			1.916 (0.816)			1.667 (0.761)		
Good	287 (19.46)	3.317 (1.317)			1.710 (0.721)			1.606 (0.767)		
Marital status[a]			0.618 (1473)	.53		0.816 (1473)	.41		0.229 (1473)	.82
Unmarried	392 (26.58)	3.926 (1.539)			2.130 (0.046)			1.795 (0.044)		
Married	1083 (73.42)	3.867 (1.610)			2.084 (0.029)			1.783 (0.026)		
Overall evaluation of the health care system[b]		1.30 (6, 1468)		.25		0.98 (6, 1468)	.43		1.82 (6, 1468)	.09
Not at all satisfied	5 (0.34)	3.400 (1.341			1.80 (1.095			1.60 (1.341		

XSL•FO
RenderX

Indicators/variables	Values, n (%)	Physician visits			Traditional Chinese Medicine visits			Western clinic visits		
		Mean (SD)	*t* test *(df)*/*F* test *(df)*	*P* value	Mean (SD)	*t* test *(df)*/*F* test *(df)*	*P* value	Mean (SD)	*t* test *(df)*/*F* test *(df)*	*P* value
Very dissatisfied	22 (1.49)	3.454 (1.625			1.727 (0.935			1.727 (0.882		
Quite dissatisfied	76 (5.15)	3.894 (1.536)			2.105 (0.903)			1.789 (0.942)		
Can't say	225 (15.25)	4.097 (1.586)			2.137 (0.922)			1.96 (0.927)		
Quite satisfied	739 (50.10)	3.887 (1.547)			2.120 (0.944)			1.767 (0.845)		
Very satisfied	280 (18.98)	3.771 (1.603)			2.032 (0.966)			1.739 (0.863)		
Fully satisfied	128 (8.68)	3.812 (1.830)			2.093 (1.089)			1.718 (0.995)		
Region[b]			1.46 (2, 1472)	.23		0.04 (2, 1472)	.96		3.85 (2, 1472)	.02
Western	420 (28.47)	3.99 (1.709)			2.107 (1.001)			1.888 (0.944)		
Central	388 (26.31)	3.829 (1.529)			2.087 (0.910)			1.742 (0.841)		
Eastern	667 (45.22)	3.844 (1.548)			2.094 (0.951)			1.749 (0.865)		
Total	1475	3.883 (1.591)	N/A[c]	N/A	1.787 (0.884)	N/A	N/A	2.096 (0.957)	N/A	N/A

[a]*t* test.

[b]*F* test.

[c]N/A: not applicable.

Multivariate Regression

After adjusting for other variables, those with OHIS behaviors increased the frequency of outpatient visits by 0.289 times more during the past year compared to those without OHIS behaviors (*P*=.003). They increased at a rate of 0.131 times in TCM outpatient visits (*P*=.01) and 0.158 times in Western physician visits (*P*=.007). Moreover, multiple regression analyses found statistically significant effects of gender, age, and health status on physician visits. Relative to the female respondents, males decreased their overall physician, TCM, and Western medicine visits by 23% (*P*=.006), 9.3% (*P*=.04), and 13.7% (*P*=.006), respectively. The number of physician visits increased with age. For every 1-year increase in age, the overall physician visits increased by 1.5% (*P*<.001), TCM visits by 0.87% (*P*<.001), and Western medicine visits by 0.7% (*P*<.001). Similar to the results of the univariate analyses, respondents with poorer health had more medical visits, while those in good health had significantly fewer visits to physicians, TCM physicians, and Western physicians (*P*<.001). All the abovementioned factors affected the number of visits in all 3 types. Concurrently, some factors such as marital status, education level, and region of residence influenced one or two types of physician visits. In comparison with respondents who were not married, those who were married had lesser number of overall physician visits by 20.4% (*P*=.03), Western medicine visits by 15.5% (*P*=.008), and TCM visits by 4.8%, with a statistically nonsignificant difference (*P*=.36). In contrast to the respondents in western regions, those in the central and eastern regions had fewer TCM visits by 12.2% (*P*=.06) and 12.6% (*P*=.03), respectively. More details of the multiple regressions are given in Table 2.

XSL•FO
RenderX

Table 2. Multivariate associations between participant characteristics and physician visits (N=1475).

	Total physician visits			Traditional Chinese medicine visits			Western medicine physician visits		
	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value
Online health information seeking behavior									
No	Ref[a]	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.289 (0.096)	0.100 to 0.478	.003	0.131 (0.052)	0.029 to 0.234	.01	0.158 (0.058)	0.044 to 0.271	.007
Gender									
Female	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Male	–0.230 (0.083)	–0.393 to –0.067	.006	–0.093 (0.045)	–0.184 to –0.002	.04	–0.137 (0.050)	–0.233 to –0.039	.006
Age	0.015 (0.003)	0.008 to 0.022	<.001	0.007 (0.001)	0.004 to 0.011	<.001	0.007 (0.002)	0.002 to 0.011	.001
Education									
Primary or below	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Junior middle school	–0.021 (0.133)	–0.283 to 0.239	.87	0.012 (0.070)	–0.126 to 0.151	.86	–0.034 (0.083)	–0.196 to 0.127	.68
High school or above	0.256 (0.140)	–0.019 to 0.533	.07	0.162 (0.073)	0.018 to 0.306	.03	0.094 (0.087)	–0.076 to 0.264	.28
Income	–0.003 (0.018)	–0.040 to 0.032	.83	0.004 (0.009)	–0.015 to 0.023	.66	–0.008 (0.012)	–0.030 to 0.014	.47
Marital status									
Unmarried	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Married	–0.204 (0.095)	–0.391 to –0.016	.03	–0.048 (0.053)	–0.154 to 0.056	.36	–0.155 (0.058)	–0.269 to –0.041	.008
Residence									
Rural	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Urban	–0.019 (0.091)	–0.200 to 0.160	.83	0.010 (0.051)	–0.090 to 0.110	.84	–0.030 (0.055)	–0.137 to 0.077	.58
Region									
Western region	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Central region	–0.085 (0.116)	–0.313 to 0.142	.46	–0.122 (0.063)	–0.247 to 0.002	.06	0.037 (0.068)	–0.097 to 0.170	.59
Eastern region	–0.066 (0.104)	–0.271 to 0.138	.53	–0.126 (0.059)	–0.241 to –0.010	.03	0.060 (0.062)	–0.061 to 0.180	.34
Health status									
Very poor	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Relatively poor	–0.811 (0.353)	–1.504 to –0.117	.02	–0.512 (0.206)	–0.917 to –0.107	.01	–0.298 (0.219)	–0.727 to 0.130	.17
General health	–1.605 (0.335)	–2.263 to –0.947	<.001	–0.884 (0.193)	–1.262 to –0.505	<.001	–0.721 (0.205)	–1.122 to –0.320	<.001
Relatively good	–2.001 (0.336)	–2.662 to –1.341	<.001	–0.977 (0.193)	–1.356 to –0.598	<.001	–1.024 (0.204)	–1.425 to –0.622	<.001
Good health	–2.274 (0.344)	–2.949 to –1.599	<.001	–1.040 (0.197)	–1.427 to –0.654	<.001	–1.234 (0.208)	–1.641 to –0.826	<.001
Overall evaluation of the health care system									
Not at all satisfied	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref

XSL•FO
RenderX

	Total physician visits			Traditional Chinese medicine visits			Western medicine physician visits		
	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value
Very dissatisfied	0.831 (0.483)	–0.117 to 1.779	.09	0.449 (0.396)	–0.328 to 1.229	.26	0.382 (0.295)	–0.196 to 0.961	.19
Quiet dissatisfied	0.841 (0.419)	0.019 to 1.663	.045	0.211 (0.363)	–0.500 to 0.923	.56	0.630 (0.267)	0.105 to 1.154	.02
Can’t say	1.183 (0.398)	0.402 to 1.965	.003	0.485 (0.355)	–0.211 to 1.182	.17	0.698 (0.253)	0.202 to 1.194	.006
Quite satisfied	1.116 (0.390)	0.351 to 1.814	.004	0.342 (0.350)	–0.345 to 1.031	.33	0.773 (0.250)	0.283 to 1.263	.002
Very satisfied	1.010 (0.394)	0.236 to 1.784	.01	0.327 (0.353)	–0.365 to 1.019	.35	0.683 (0.252)	0.189 to 1.177	.007
Fully satisfied	0.835 (0.415)	0.021 to 1.649	.04	0.235 (0.360)	–0.472 to 0.942	.51	0.600 (0.262)	0.086 to 1.113	.02
Constant variable	3.920 (0.531)	2.877 to 4.962	<.001	1.904 (0.390)	1.138 to 2.670	<.001	2.015 (0.330)	1.366 to 2.664	<.001

[a]Ref: reference value.

Finally, the group that never participated in OHIS was removed. Our findings indicate that a higher frequency of OHIS was associated with a higher frequency of outpatient visits (β=0.127; *P*=.003) and TCM visits (β=0.082; *P*=.001). More details are given in Table 3.

XSL•FO
RenderX

Table 3. Multivariate associations between participant characteristics and physician visits among those with online health information seeking behaviors (n=939).

	Total physician visits			Traditional Chinese medicine visits			Western medicine physician visits		
	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value
Online health information seeking frequency	0.127 (0.043)	0.042 to 0.211	.003	0.082 (0.025)	0.033 to 0.131	.001	0.045 (0.028)	–0.010 to 0.099	.11
Gender									
Female	Ref[a]	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Male	–0.133 (0.102)	–0.334 to 0.067	.19	–0.044 (0.057)	–0.156 to 0.068	.44	–0.089 (0.061)	–0.209 to 0.031	.15
Age	0.008 (0.004)	0.000 to 0.017	.046	0.006 (0.002)	0.001 to 0.011	.01	0.002 (0.002)	–0.003 to 0.008	.34
Education									
Primary or below	Ref	Ref	Ref	Ref	Ref to	Ref	Ref	Ref	Ref
Junior middle school	–0.215 (0.167)	–0.543 to 0.113	.20	–0.037 (0.095)	–0.223 to 0.149	.69	–0.178 (0.105)	–0.384 to 0.029	.09
High school or above	0.145 (0.167)	–0.182 to 0.472	.39	0.152 (0.094)	–0.032 to 0.336	.11	–0.007 (0.104)	–0.211 to 0.197	.95
Income	0.005 (0.021)	–0.037 to 0.047	.82	.004 (0.011)	–0.017 to 0.026	.69	0.001 (0.014)	–0.027 to 0.028	.97
Marital status									
Unmarried	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Married	–0.132 (0.115)	–0.357 to 0.093	.25	–0.021 (0.064)	–0.146 to 0.104	.74	–0.111 (0.072)	–0.252 to 0.030	.12
Residence									
Rural	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Urban	0.021 (0.116)	–0.207 to 0.249	.86	–0.004 (0.065)	–0.132 to 0.124	.95	0.025 (0.069)	–0.109 to 0.160	.72
Region									
Western region	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Central region	–0.131 (0.141)	–0.407 to 0.146	.35	–0.144 (0.080)	–0.300 to 0.012	.07	0.014 (0.082)	–0.148 to 0.175	.87
Eastern region	–0.114 (0.131)	–0.371 to 0.144	.39	–0.144 (0.074)	–0.289 to 0.001	.05	0.030 (0.080)	–0.127 to 0.188	.71
Health status									
Very poor	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Relatively poor	–0.573 (0.513)	–1.579 to 0.434	.27	–0.520 (0.311)	–1.131 to 0.091	.09	–0.053 (0.314)	–0.669 to 0.563	.87
General health	–1.253 (0.493)	–2.221 to –0.285	.01	–0.871 (0.294)	–1.448 to –0.293	.003	–0.382 (0.290)	–0.951 to 0.188	.19
Relatively good	–1.846 (0.490)	–2.806 to –0.885	<.001	–1.047 (0.293)	–1.622 to –0.472	<.001	–0.798 (0.287)	–1.362 to –0.235	.006
Good health	–1.968 (0.501)	–2.951 to –0.985	<.001	–1.039 (0.299)	–1.626 to –0.453	.001	–0.929 (0.292)	–1.502 to –0.356	.002
Overall evaluation of the health care system									

XSL•FO
RenderX

	Total physician visits			Traditional Chinese medicine visits			Western medicine physician visits		
	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value	Coefficient (SE)	95% CI	*P* value
Not at all satisfied	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Very dissatisfied	1.461 (0.691)	0.105 to 2.817	.04	0.815 (0.230)	0.363 to 1.267	<.001	0.646 (0.600)	–0.356 to 1.823	.28
Quiet dissatisfied	1.620 (0.596)	0.449 to 2.790	.007	0.771 (0.147)	0.483 to 1.060	<.001	0.848 (0.570)	–0.271 to 1.967	.14
Can’t say	1.914 (0.572)	0.791 to 3.037	.001	1.032 (0.118)	0.800 to 1.265	<.001	0.882 (0.558)	–0.213 to 1.976	.11
Quite satisfied	1.862 (0.564)	0.756 to 2.968	.001	0.903 (0.100)	0.707 to 1.099	<.001	0.959 (0.556)	–0.133 to 2.050	.08
Very satisfied	1.726 (0.570)	0.606 to 2.845	.003	0.872 (0.108)	0.660 to 1.083	<.001	0.854	–0.241 to 1.949	.13
Fully satisfied	1.526 (0.600)	0.350 to 2.703	.01	0.751 (0.154)	0.449 to 1.053	<.001	0.776 (0.568)	–0.340 to 1.891	.17

[a]Ref: reference value.

Mediation Effect Analysis

First, the total effect of OHIS behavior on the frequency of physician visits was examined. The results revealed that OHIS had a significant influence on total physician visits (β=0.290; *P*=.003), indicating a certain correlation between OHIS behavior and physician visits. We also tested the effect of OHIS behavior on perception, and results revealed that OHIS had a significant positive impact on perception (β=0.265; *P*<.001). Furthermore, the effect of OHIP on the number of physician visits was conducted, and it was found to have a significant effect on physician visits (β=0.270; *P*=.001). Finally, the effect of perception on physician visits while controlling for OHIS behavior was also assessed. We found that the perception of people regarding online health information had a significantly positive influence on physician visits (β=0.232; *P*=.02). More details are given in Table 4 and Figure 2.

XSL•FO
RenderX

Table 4. Mediation effect analysis of online health information seeking behavior, physician visits, and online health information perception (N=1475).

	Model 1 (N=1475)		Model 2 (N=1475)		Model 3 (N=1475)		Model 4 (N=1475)	
	Physician visits	*P* value	Online health information perception	*P* value	Physician visits	*P* value	Physician visits	*P* value
Online health information seeking behavior	0.290 (0.096)	.003	0.265 (0.032)	<.001	N/A[a]	N/A	0.232 (0.100)	.01
Online health information perception	N/A	N/A	N/A	N/A	0.270 (0.082)	.001	0.217 (0.084)	.02
Gender	–0.230 (0.083)	.006	–0.000 (0.028)	.99	–0.242 (0.083)	.004	–0.230 (0.083)	.006
Residence	–0.020 (0.092)	.83	–0.028 (0.030)	.35	0.000 (0.091)	.99	–0.014 (0.092)	.88
Age	0.015 (0.004)	<.001	0.000 (0.001)	.82	0.013 (0.004)	<.001	0.015 (0.004)	<.001
Education levels								
Primary and low	Reference		Reference		Reference		Reference	
Junior high school	–0.022 (0.133)	.87	0.010 (0.041)	.81	–0.005 (0.133)	.97	–0.024 (0.133)	.86
High school and above	0.257 (0.141)	.07	0.011 (0.043)	.80	0.297** (0.141)	.04	0.254 (0.140)	.07
Income	–0.004 (0.019)	.83	0.009 (0.008)	.24	–0.004 (0.019)	.83	–0.006 (0.019)	.75
Health condition								
Very poor	Reference		Reference		Reference		Reference	
Relatively poor	–0.811 (0.353)	<.001	–0.063 (0.107)	.55	–0.807 (0.354)	.02	–0.797 (0.355)	.03
Average health	–1.605 (0.336)	<.001	–0.051 (0.095)	.59	–1.601 (0.336)	<.001	–1.594 (0.338)	<.001
Relatively good	–2.002 (0.337)	<.001	–0.077 (0.096)	.43	–1.985 (0.337)	<.001	–1.985 (0.339)	<.001
Good health	–2.275 (0.344)	<.001	–0.038 (0.099)	.70	–2.285 (0.345)	<.001	–2.266 (0.347)	<.001
Marital status								
Unmarried	Reference		Reference		Reference		Reference	
Married	–0.204 (0.096)	.03	–0.013 (0.032)	.68	–0.186 (0.094)	.048	–0.202 (0.095)	.03
Overall evaluation of the health care system								
Not at all satisfied	Reference		Reference		Reference		Reference	
Very dissatisfied	0.831 (0.483)	.09	–0.081 (0.281)	.77	0.915 (0.495)	.06	0.849 (0.511)	.09
Quiet dissatisfied	0.841 (0.419)	.045	0.204 (0.223)	.36	0.844 (0.425)	.047	0.797 (0.444)	.07
Can’t say	1.184 (0.399)	.003	0.132 (0.218)	.55	1.206 (0.405)	.003	1.155 (0.424)	.007
Quite satisfied	1.116 (0.390)	.004	0.181 (0.217)	.40	1.115 (0.398)	.005	1.077 (0.417)	.01
Very satisfied	1.011 (0.395)	.01	0.244 (0.217)	.26	0.988 (0.402)	.01	0.958 (0.421)	.02
Fully satisfied	0.835 (0.415)	.04	0.352 (0.222)	.11	0.781 (0.423)	.07	0.759 (0.441)	.09
Region								
Western	Reference		Reference		Reference		Reference	
Central	–0.086 (0.116)	.46	0.118 (0.037)	.001	–0.111 (0.116)	.34	–0.112 (0.116)	.34
Eastern	–0.067 (0.105)	.53	–0.005 (0.035)	.89	–0.065 (0.104)	.53	–0.066 (0.104)	.53

[a]N/A: not applicable.

XSL•FO
RenderX

Figure 2. The mediating role of online health information perception between online health information seeking behavior and physician visits. OHIP: online health information perception; OHIS: online health information seeking.

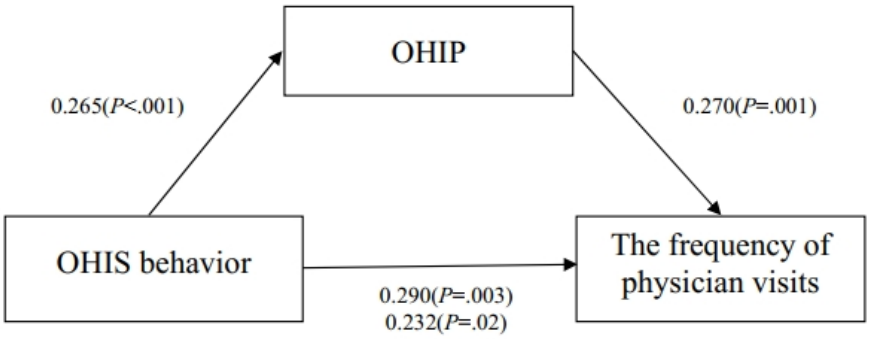

Discussion

Primary Findings

The key findings of this study are summarized as follows. First, the perception of people regarding online health information influences the effect of OHIS on health service usage as predicted. The increased prevalence of OHIS behaviors in the internet era necessitates more awareness and discrimination of health information accessed on the internet. Our findings validated the impact of OHIS on health service utilization outcomes. Second, the effect found through the mediation analysis was partial mediating, which can be divided into 2 parts. On the one hand, respondents' OHIS behavior directly affects the number of physician visits. On the other hand, OHIS behavior affects their perception of the health information available online, which subsequently affects their use of outpatient services. Third, the multivariable linear regression results revealed that OHIS behavior significantly increased the utilization of clinic visits after controlling for other variables. Those with OHIS behavior showed 0.308 times increase in the frequency of total physician visits compared to those without OHIS behaviors. This trend was not only reflected in the total number of physician visits but also in the TCM and Western medicine physician visits. Finally, based on the joint model that was built on OHIS, health care utilization, and OHIP, we clarified the mediator roles of OHIP between OHIS behavior and physician visits and provided new perceptive and explanatory mechanisms to explore the relationship between health information seeking behavior and health service utilization.

Interpretation and Comparison With Existing Literature

Studies exploring the determinants of OHIS have confirmed that perceived usefulness and attitude have a strong positive effect on OHIS [29]. A previous study [31] investigated the psychological mechanisms between OHIS behavior and doctor-patient interactions, concluding the mediating role of eHealth literacy, perceived disease severity, and action benefits. The findings of this study further confirm that a higher perception of health-related information online affects the OHIS behavior and influences the health utilization process. As various studies [14,32] have indicated, OHIS behavior can be a potential tool to improve health literacy improvement and health-related behaviors. Residents consciously accessing health information on the internet can be more knowledgeable about their health and are more likely to identify potential health problems, resulting in higher health demand and higher physician visits.

Based on previous studies, education level is one of the most important factors for health service utilization [33,34]. However, the impact of education level on outpatient services is complex. From one point of view, due to their strong health awareness, people with higher education levels may visit outpatient clinics more frequently for minor discomforts or for the management of chronic diseases. In contrast, individuals with a higher education level are inclined to possess more comprehensive preventive health care knowledge, higher health literacy, and maintain a superior state of physical well-being, which may also lead to a decrease in outpatient visits. We identified significant differences between different education levels and outpatient visits through a primary univariate analysis. Participants with primary or lower education levels had more physician visits. However, this trend disappeared in the multivariate analysis after controlling for other variables. Among educational attainment factors, having a high school diploma or above increased the likelihood of receiving TCM treatment, but this was not reflected in the overall and Western medicine treatments. In our analysis, increasing age indicated increased health service utilization, which is consistent with that reported in previous studies [21,23]. Older adults have various health needs, including medical checkups for chronic diseases, rehabilitation services after illness or surgery, mental health support to combat loneliness and depression, and long-term care assistance with daily activities. All these needs can be partly satisfied through the internet, which provides access to medical information, online consultations, health management, and virtual communities for support and interaction. However, 1 systematic review [7] indicated that older adults confronted 3 types of barriers to conducting OHIS: individual, social, and information and communication technology–related. The findings of this study indicated that shifting the perspective on OHIS can promote OHIS behavior and consequently enhance the rational use of health services. The older adults, being a vulnerable population in terms of health and information and communication technology, require additional support in using web-based services efficiently and reasonably.

Additionally, unauthenticated health information is widely available on the internet. Another study [35] revealed that 96% of the residents use unauthenticated health information to answer

XSL•FO
RenderX

some health-related questions. Patients may not always be able to use effective online search strategies. Studies have indicated that when there is a contradiction between the information on the internet and that provided by doctors, people are more likely to trust the perspective of doctors [36,37]. These results verified that people's positive perceptions and attitudes toward online health information promote OHIS behavior and the use of health services. Health care professionals' recommendations and assistance on reliable information sources and trustworthy websites in patients' search for online health information can be very useful and can improve the rationality of online health information and the seeking of medical treatment.

Implications for Policies and Practice

The findings of this study have important implications for future policies and practices in the health care system. Focusing on the access of individuals to online health information allows health service providers to have a better understanding of their health knowledge, behaviors, and even health attitudes, promoting a more logical use of health services. In the field of health education, the primary health care service system aims to improve the eHealth literacy of residents, including their ability to seek, discriminate, judge, and use internet health information. This is done to promote their ability to access their own or their family's health status and enhance their ability to make timely and appropriate health care decisions. The older adults, who are always considered medically underserved populations, often reported a lack of social support in their OHIS, including informational, organizational, instrumental, intergenerational, and peer support. Interventions can be explored to improve reasonable outpatient utilization by the older adults by improving their perception and attitude toward health information online. Comprehensive support, which encompasses user-friendly devices and software, training, and guidance on efficient internet use, accessible internet connections, and dedicated support services, is necessary for older adults to truly benefit from OHIS and reasonable health services.

Strengths and Limitations

Compared with previous studies, this study is innovative and contributes in the following aspects. First, the research sample included is from CGSS, a large-scale, comprehensive social survey project. On this basis, people's perception of online health information was selected as the mediating factor to explain the mechanism linking OHIS to health service utilization, broadening the research scope. Furthermore, these findings provide new evidence on OHIS behavior and physician visits and verified the partial mediation role of people's perceptions and attitudes toward online health information. Lastly, multiple variables were integrated into participants' perception of online health information based on the factor analysis, which enhanced its efficiency, rationality, and reliability.

Although this study achieved some useful results, there are some limitations. First, our study was conducted using cross-sectional data, which prevented us from determining the causal relationship between variables. Future studies can adopt a longitudinal design to further understand the dynamic relationships among OHIS, OHIP, and health services. Given that a public database was used as a resource, certain limitations were unavoidable. OHIP, selected as the mediating variable, was relatively subjective when compared with eHealth literacy or ability. Nevertheless, it remains significant to know that the attitudes of residents toward online health information have influenced their OHIS and health service utilization.

In the future, the OHIP mechanism can be further explored in more depth, and the influence can be expanded to inpatient and emergency service utilization. Then, the relationships among OHIS, OHIP, and health service utilization can be investigated in different contexts to improve the generalizability of the results. Finally, other research methods such as experimental research and case studies can be combined to gain a more comprehensive understanding of the relationships between variables.

Conclusions

The OHIP of residents influences the effect of OHIS on the frequency of physician visits. OHIS behavior impacts outpatient service utilization both directly and indirectly via OHIP. OHIS behaviors significantly increase the frequency of physician visits after controlling for other variables. Interventions can be explored to improve the health utilization of residents by improving their OHIP.

Acknowledgments

This work was supported by the Shandong Provincial Natural Science Foundation (ZR2021QG051).

Data Availability

These data were derived from the Chinese General Social Survey, which is open to everyone. Researchers who want to use these data can visit the website [30]. The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: HL (lead), DL (equal), MZ (supporting), LL (supporting), ZHC (supporting)
Data curation: DL (lead), HL (equal), MZ (supporting)
Formal analysis: DL (lead), HL (equal)

XSL•FO
RenderX

Funding acquisition: HL (lead)
Methodology: DL (lead), HL (equal)
Writing – original draft: HL (lead), DL (equal), MZ (supporting), LL (supporting), ZHC (supporting)
Writing – review & editing: HL (lead), DL (equal), MZ (supporting), LL (supporting), ZHC (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Procedure and results of factor analysis.
[DOCX File , 16 KB - jmir_v27i1e66683_app1.docx]

References

1. Petrosyan A. Statista. URL: https://www.statista.com/topics/1145/internet-usage-worldwide/#topicOverview [accessed 2024-09-18]
2. The 54th statistical report on China's internet development, 2024. China Internet Network Information Center. URL: https://www.cnnic.net.cn/n4/2024/0829/c88-11065.html [accessed 2024-09-11]
3. Li H, Xu J, Li L, Meng Q. Health-related internet use in hard-to-reach populations: empirical findings from a survey in a remote and mountainous province in China. J Med Internet Res 2019 May 31;21(5):e12693 [FREE Full text] [doi: 10.2196/12693] [Medline: 31152526]
4. Usman JE, Childs C, Rogerson D, Klonizakis M. Online health information-seeking behaviour among people of African Descent in the United Kingdom: a qualitative study. Healthcare (Basel) 2024 Apr 26;12(9):897 [FREE Full text] [doi: 10.3390/healthcare12090897] [Medline: 38727454]
5. Zhang D, Zhan W, Zheng C, Zhang J, Huang A, Hu S, et al. Online health information-seeking behaviors and skills of Chinese college students. BMC Public Health 2021 Apr 15;21(1):736 [FREE Full text] [doi: 10.1186/s12889-021-10801-0] [Medline: 33858389]
6. Zakar R, Iqbal S, Zakar MZ, Fischer F. COVID-19 and health information seeking behavior: digital health literacy survey amongst university students in Pakistan. Int J Environ Res Public Health 2021 Apr 11;18(8):4009 [FREE Full text] [doi: 10.3390/ijerph18084009] [Medline: 33920404]
7. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
8. Ortiz-Dowling E, Der Ananian C, Larkey L, Hooker S. Health-seeking behaviors and health information gathering in older Mexican American males. Psychology of Men & Masculinities 2019 Oct;20(4):564-574. [doi: 10.1037/men0000194]
9. Chang F, Chang C, Tao C. Adolescent pursuit of health information online during the COVID-19 pandemic: The roles played by eHealth literacy and psychological distress. Glob Ment Health (Camb) 2023;10:e51 [FREE Full text] [doi: 10.1017/gmh.2023.44] [Medline: 37854426]
10. Gulec H, Smahel D. Individual and parental factors of adolescents' mhealth app use: nationally representative cross-sectional study. JMIR Mhealth Uhealth 2022 Dec 16;10(12):e40340 [FREE Full text] [doi: 10.2196/40340] [Medline: 36525286]
11. Limbu Y, Giovannetti M, Cardinali S. Dietary supplement usage during pregnancy and lactation: role of online social capital and health information-seeking behaviour. BFJ 2020 Jul 03;123(1):31-47 [FREE Full text] [doi: 10.1108/bfj-12-2019-0906]
12. Link E, Baumann E, Klimmt C. Explaining online information seeking behaviors in people with different health statuses: German representative cross-sectional survey. J Med Internet Res 2021 Dec 10;23(12):e25963 [FREE Full text] [doi: 10.2196/25963] [Medline: 34890348]
13. Madge O, Marinca AM, Daha C, Simion L. Health information seeking behaviour and decision making by patients undergoing breast cancer surgery: A qualitative study. Health Info Libr J 2023 Mar 02:123-135. [doi: 10.1111/hir.12480] [Medline: 36861854]
14. Liu Z, Liao Y, Hwang C, Rethorst CD, Zhang X. Associations of online health information seeking with health behaviors of cancer survivors. Digit Health 2024;10:20552076241238074 [FREE Full text] [doi: 10.1177/20552076241238074] [Medline: 38495862]
15. Dau H, Saad El Din K, McTaggart-Cowan H, Loree JM, Gill S, De Vera MA. Health information seeking behaviors among individuals with young-onset and average-onset colorectal cancer: an international cross-sectional survey. Support Care Cancer 2020 Dec;28(12):6011-6021. [doi: 10.1007/s00520-020-05446-5] [Medline: 32296980]
16. Link E, Baumann E, Klimmt C. Understanding the importance of trust in patients? coping with uncertainty via health information-seeking behaviors. Communications2024 2024;49(1):74-98. [doi: 10.1515/commun-2021-0117]
17. Zhang X, Foo S, Majid S, Chang Y, Dumaual HTJ, Suri VR. Self-care and health-information-seeking behaviours of diabetic patients in Singapore. Health Commun 2020 Jul;35(8):994-1003. [doi: 10.1080/10410236.2019.1606134] [Medline: 31303050]

XSL•FO
RenderX

18. Kostagiolas P, Tsiligros P, Theodorou P, Tentolouris N, Niakas D. A cross-sectional survey interconnecting health information seeking behavior with clinical data of type 2 diabetes mellitus patients. LHT 2020 May 19;39(2):448-461 [FREE Full text] [doi: 10.1108/lht-02-2020-0030]
19. Stanarević Katavić S. Health information behaviour of rare disease patients: seeking, finding and sharing health information. Health Info Libr J 2019 Dec;36(4):341-356 [FREE Full text] [doi: 10.1111/hir.12261] [Medline: 31099979]
20. Lim HM, Dunn AG, Lim JR, Abdullah A, Ng CJ. Association between online health information-seeking and medication adherence: A systematic review and meta-analysis. Digit Health 2022;8:20552076221097784 [FREE Full text] [doi: 10.1177/20552076221097784] [Medline: 35586836]
21. Suziedelyte A. How does searching for health information on the internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835 [FREE Full text] [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
22. Nguyen HTL, Nakamura K, Seino K, Vo VT. Association between a wider availability of health information and health care utilization in Vietnam: cross-sectional study. J Med Internet Res 2017 Dec 18;19(12):e405 [FREE Full text] [doi: 10.2196/jmir.8328] [Medline: 29254911]
23. Lin C, Lin H. Impact of mobile Internet use on health-seeking behaviors: evidence from China. Front Public Health 2024;12:1403877 [FREE Full text] [doi: 10.3389/fpubh.2024.1403877] [Medline: 38966701]
24. Di Novi C, Kovacic M, Orso CE. Online health information seeking behavior, healthcare access, and health status during exceptional times. J Econ Behav Organ 2024 Apr;220:675-690 [FREE Full text] [doi: 10.1016/j.jebo.2024.02.032] [Medline: 38628501]
25. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2021 Sep;36(10):1163-1175. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
26. Zhao YC, Zhao M, Song S. Online health information seeking among patients with chronic conditions: integrating the health belief model and social support theory. J Med Internet Res 2022 Nov 02;24(11):e42447 [FREE Full text] [doi: 10.2196/42447] [Medline: 36322124]
27. Rueger J, Dolfsma W, Aalbers R. Perception of peer advice in online health communities: Access to lay expertise. Soc Sci Med 2021 May;277:113117 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113117] [Medline: 33865092]
28. Liu D, Yang S, Cheng CY, Cai L, Su J. Online health information seeking, eHealth Literacy, and health behaviors among Chinese internet users: cross-sectional survey study. J Med Internet Res 2024 Oct 18;26:e54135 [FREE Full text] [doi: 10.2196/54135] [Medline: 39423374]
29. Mou J, Shin D, Cohen J. Health beliefs and the valence framework in health information seeking behaviors. ITP 2016 Nov 07;29(4):876-900 [FREE Full text] [doi: 10.1108/itp-06-2015-0140]
30. Chinese General Social Survey. URL: http://cgss.ruc.edu.cn/English/Home.htm [accessed 2025-02-28]
31. Wang J, Xiu G, Shahzad F. Exploring the determinants of online health information-seeking behavior using a meta-analytic approach. Sustainability 2019 Aug 24;11(17):4604 [FREE Full text] [doi: 10.3390/su11174604]
32. Luo A, Yu Z, Liu F, Xie W. The chain mediating effect of the public's online health information-seeking behavior on doctor-patient interaction. Front Public Health 2022;10:874495 [FREE Full text] [doi: 10.3389/fpubh.2022.874495] [Medline: 35719674]
33. Ghanbari-Jahromi M, Bastani P, Jalali FS, Delavari S. Factors affecting oral and dental services` utilization among Elderly: a scoping review. BMC Oral Health 2023 Aug 27;23(1):597 [FREE Full text] [doi: 10.1186/s12903-023-03285-4] [Medline: 37635217]
34. Ji J, Yang Y, Chen Z, Zhang W, Jiang S, Shen X, et al. How education level affects postoperative rehabilitation and follow-up: a single-center experience. BMC Urol 2023 Jul 18;23(1):123 [FREE Full text] [doi: 10.1186/s12894-023-01282-x] [Medline: 37464331]
35. Jin SW, Lee Y, Dia DA. Analyzing paths from online health information seeking to colorectal cancer screening using health literacy skills frame and cognitive mediation model. Patient Educ Couns 2019 Mar;102(3):416-423. [doi: 10.1016/j.pec.2018.11.002] [Medline: 30448041]
36. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Computers in Human Behavior 2017 Apr;69:256-267 [FREE Full text] [doi: 10.1016/j.chb.2016.12.032]
37. Yudianto B, Caldwell PH, Nanan R, Barnes EH, Scott KM. Patterns of parental online health information-seeking behaviour. J Paediatr Child Health 2023 May;59(5):743-752. [doi: 10.1111/jpc.16387] [Medline: 37051735]

Abbreviations

CGSS: Chinese General Social Survey
OHIP: online health information perception
OHIS: online health information seeking
TCM: traditional Chinese medicine

XSL•FO
RenderX

Edited by A Mavragani; submitted 22.09.24; peer-reviewed by R El Sherif, R Peng; comments to author 04.12.24; revised version received 18.01.25; accepted 14.02.25; published 14.03.25.

Please cite as:
Li H, Li D, Zhai M, Lin L, Cao Z
Associations Among Online Health Information Seeking Behavior, Online Health Information Perception, and Health Service Utilization: Cross-Sectional Study
J Med Internet Res 2025;27:e66683
URL: https://www.jmir.org/2025/1/e66683
doi:10.2196/66683
PMID:

©Hongmin Li, Dongxu Li, Min Zhai, Li Lin, ZhiHeng Cao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying Contextual Factors That Shape Cybersecurity Risk Perception for Assisted Living and Health Care Technologies and Wearables: Mixed Methods Study

Alvhild Skjelvik[1], PhD; Nicholas West[2], MSc; Matthias Görges[2,3], PhD

[1]Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU), Gjøvik, Norway
[2]Research Institute, BC Children's Hospital, Vancouver, BC, Canada
[3]Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Alvhild Skjelvik, PhD
Department of Information Security and Communication Technology
Norwegian University of Science and Technology (NTNU)
Teknologiveien 22
Gjøvik, 2815
Norway
Phone: 47 98697128
Email: alvhild.skjelvik@gmail.com

Abstract

Background: Over the last decade, the health care technology landscape has expanded significantly, introducing new and innovative solutions to address health care needs. The implications of cybersecurity incidents in the health care context extend beyond data breaches to potentially harming individuals' health and safety. Risk perception is influenced by various contextual factors, contributing to cybersecurity concerns that technological safeguards alone cannot address. Thus, it is imperative to study risk perceptions, contextual factors, and technological benefits to guide policy development, risk management, education, and implementation strategies.

Objective: This study aims to investigate the differences in cybersecurity risk perception among various stakeholders in the health care sector in Norway and British Columbia (BC), Canada, and identify specific contextual factors that shape these perceptions. We expect to identify differences in risk perceptions for the explored health care technologies.

Methods: We used a mixed methods approach comprising surveys and semistructured interviews to sample health care–related wearable technology stakeholders, including health care workers, patients (adults and adolescents) and their families, health authorities and hospital staff (biomedical engineers, information technology support, research staff), and device vendors/industry professionals in Norway and BC. Surveys explored information security scenarios based on the Behavioral-Cognitive Internet Security Questionnaire (BCISQ), risk perception, and contextualizing variables. We analyzed both survey data sets to summarize participants' characteristics and responses to questions related to the BCISQ (behavior and attitude) and risk perception. Interviews were analyzed thematically using an inductive-deductive approach to explore risk perception and contextual factors.

Results: Data from 274 survey respondents were available for analysis: 185 from Norway, including 139 (75.1%) females, and 89 from BC, including 57 (64%) females. A total of 45 respondents (31 in Norway and 14 in BC) participated in interviews. The BCISQ showed minor differences between locations; respondents demonstrated generally low-risk behavior and robust information security awareness. However, password simulation demonstrated discrepancies between self-assessed and "real" behavior by sharing or willingness to share passwords. Perceived risk is generally considered low, yet consequences of cybersecurity risks were evaluated as major but unlikely. Risk perception was stronger for assisted living and diabetes technologies than for smartwatches. The most important contextual factors shaping risk perceptions are human factors encompassing knowledge, competence, familiarity, feelings of dread, perceived benefit, and trust, as well as the technological factor of device functionality. Organizational and technological factors had lesser effects.

Conclusions: We found minimal differences in behavior and risk perception among Norwegian and BC participants. Human factors and device functionality were most influential in shaping cybersecurity risk perceptions. Considering the rising need for assisted living technologies and wearables, insights into risk perceptions can strengthen risk management, awareness, and competence building. Further, it can address potential concerns among stakeholders to enable quicker technology adoption.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e64388) doi:10.2196/64388

KEYWORDS

digital health; health care technology; cybersecurity; risk perception; health care stakeholders; contextual factors

Introduction

Background

Over the last decade, the landscape of health care technology has expanded significantly, introducing innovative solutions to assist health care needs [1]. These advances have been instrumental in enhancing the efficiency of health care delivery, reducing operational costs, and optimizing timely management. The adoption of artificial intelligence–based solutions in the health care sector has accelerated through applications such as artificial intelligence–supported analysis of X-ray interpretation and blood sample evaluations [2-5]. Moreover, an emerging trend is using technology to facilitate remote patient care outside traditional health care environments such as hospitals [6,7]. This trend introduces multiple challenges, particularly in cybersecurity. Medical devices and assisted living technologies have been the subject of cybersecurity incidents [8-11]. Recently, several hundred insulin pumps were called back by the Food and Drug Administration (responsible for regulating medical devices in the United States) due to software weaknesses, resulting in the battery depletion of insulin pumps [12,13]. Although this was not an intended attack, it demonstrates the vulnerabilities in the technology that could be exploited for malicious purposes. Incidents resulting from intended attacks have also occurred, including attacks on home care services in Ontario, Canada, Norway, and Sweden, which jeopardized patient safety and their data [14,15].

A wide range of remote care technologies are available. These technologies collect and transfer data generated from patients, such as measurements and self-reported data, also known as patient-generated health data (PGHD) [16]. Assisted living technologies (known as "welfare technologies" in Norway) refer to technologies that augment an individual's safety and security, including technology used for remote care. These technologies also manage and transfer PGHD. Assisted living technology, remote care, and disease management technologies have been adopted broadly in Norway and Canada. This adoption is expected to increase drastically in the coming years due to demographic shifts and health care resource constraints [17,18]. The 2 countries differ in geographical size and population, but they share a high degree of trust in public institutions and maintain predominantly publicly funded health care systems. Both also share a decentralization of health care delivery; in Canada, health care services are managed by 10 provincial health authorities and 3 territorial authorities, while in Norway, primary health care is distributed among 356 municipalities, while specialist health care is distributed among 4 regional health care authorities. Moreover, the populations of both countries demonstrate high levels of technological literacy, with an internet penetration rate of 99% in Norway and 92% in Canada [19]. Although technological literacy is high, cybersecurity still poses a challenge.

Cybersecurity can be understood as the "protection of computer systems and networks from attack by malicious actors that may result in unauthorized information disclosure, theft of, or damage to hardware, software, or data, as well as from the disruption or misdirection of the services they provide" [20]. In health care, the implications of cybersecurity incidents extend beyond simple data breaches to encompass critical consequences, potentially harming the health and safety of individuals and compromising PGHD [1,21]. The successful deployment of remote care solutions depends on several elements, including technology accessibility, technological literacy, financial resources, and health care delivery's structural and organizational aspects [22-24]. It involves many stakeholders, including patients, clinicians, public and private health care organizations, technology vendors, and regulatory bodies [25-27]. The stakeholders involved in device use, operation, and management may hold different perceptions of cybersecurity risks.

Cybersecurity risk perception may differ between patients and health care workers; a patient using a type 1 diabetes (T1D) device may be more concerned about the immediate threat of denial of service impacting their insulin delivery, while a clinician might focus on the broader implications of compromised data integrity across both T1D devices and wearables for clinical decision-making. Cybersecurity risk perception encompasses concerns about threats and their potential impact on technologies, systems, and data, including incidents such as data breaches, attacks, and hacking, which can compromise confidentiality, integrity, availability, and functionality. Understanding these risks requires technical considerations and the social, cultural, and psychological factors that shape how individuals and organizations perceive cybersecurity threats [28,29].

Various disciplines, such as psychology, engineering, and safety science, have explored risk perception. Common theories in the risk perception field include the social amplification of risk framework and protection motivation theory [30-36]. Factors such as culture, optimism bias, personal fear, dread, familiarity, benefit, trust, and social trust have been explored in previous research on cybersecurity risk perception [30-33,37-45]. The Technology Adoption Model (TAM) and TAM2 are widely accepted in explaining why particular technologies are adopted [46-51]. Previous research highlights a lack of focus on human factors when assessing sociotechnical systems, especially within cybersecurity. A study found that technical factors often receive the most attention, whereas human factors related to awareness, privacy, and trust are frequently overlooked [52]. Addressing these human factors is crucial for a comprehensive understanding of cybersecurity risks in health care. Therefore, we adopted a cybersecurity approach to contextual factors, dividing them into human, organizational, and technological categories [53,54].

XSL•FO
RenderX

Aim

This paper aims to investigate cybersecurity risk perception for assisted living technologies and wearables among stakeholders from Norway and British Columbia (BC), Canada, and identify contextual factors that shape these perceptions. We aim to explore differences in cybersecurity risk perceptions among different stakeholder groups for these technologies and identify the relevant human, organizational, and technological factors that impact these differences. However, due to ethical considerations, we do not compare survey data from the 2 geographical locations; instead, we rely on the interpretation of qualitative data from interviews to identify if differences in risk perception exist.

Methods

Study Design

This study performed an analysis among relevant stakeholders in Norway and BC, Canada, to explore cybersecurity risk perception and contextual factors associated with assisted living technologies. To simplify reporting, we refer to "cybersecurity risk perception" as "risk perception." Our sampling approach in Norway was based on previous research on key stakeholders for assisted living technologies [23,55,56]. In Norway, we focused on 2 specific technologies, medicine dispensers and digital self-reporting forms, as these solutions have been widely adopted. We approached stakeholders to participate as survey respondents and as interviewees. In BC, these same technologies have not been as widely adopted. Hence, we examined other technologies that support an individual in managing their disease remotely: insulin pumps and continuous glucose monitoring (CGM) devices. These technologies are used as examples in the survey and interviews.

The survey was distributed in Norway from September 2023 to March 2024, and interviews were conducted from August 2021 to February 2024. Three preliminary interviews were conducted before ethical approval, which was permitted by Norwegian ethics rules but restricted the ability to gather any identifiable information. However, they enabled the gathering of key insights into the organization and implementation of assisted living technology in the Norwegian health care sector, which were instrumental in the subsequent research design. The survey was distributed in BC from December 2023 to February 2024, and interviews were conducted from December 2023 to April 2024 (Figure 1). This manuscript has been prepared following the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [57].

Figure 1. Study timeline showing study phases, data collection periods, and ethical approvals.

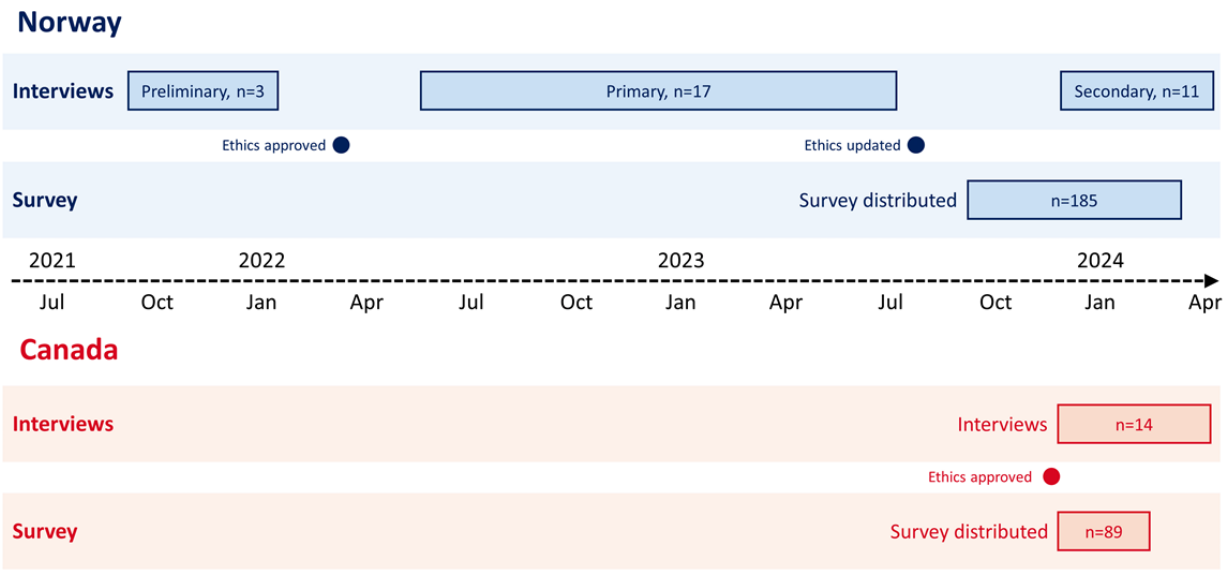

Participants

Participants were health care workers, patients or device users, and experts involved in the development and implementation of assisted living technologies and wearables. The stakeholder groups enrolled varied slightly between the 2 countries. Specifically, Norwegian participants included health care workers in primary and specialist care, device vendors or industry professionals, health authorities, and cybersecurity experts. The participants from BC were patients (adults and adolescents), families of patients, clinicians or researchers using health care–related wearable technologies or data, biomedical engineers, hospital or research information technology (IT) professionals, and device vendors or industry professionals. We included adolescents to enrich the data material, as they represent a key user group for the T1D technology use case. Thus, it helped to offer a more comprehensive understanding of the cybersecurity risk affiliated with the technology across different demographics.

Recruitment

The survey and interview respondents in Norway were gathered through an initial email to primary health care providers, specialist health care providers, health care authorities, technology vendors, and the expert network at the Norwegian University of Science and Technology. The survey was

XSL•FO
RenderX

distributed by leaders in either primary or specialist health care organizations and a group dedicated to health care technologies in Norway on the social media platform Facebook (Meta Platforms, Inc.). In BC, we recruited adolescents, their parents, and other adults living with T1D using existing lists of participants who had given permission to be contacted for research and additional adult patients through a provincial portal for patient engagement and recruitment: REACHBC [58]. Health care clinicians, researchers, IT professionals, biomedical engineers, and device vendors/industry professionals were contacted via email and newsletters utilizing existing networks; participants belonging to health care institutions were approached via departmental email lists and research institute newsletters.

Survey Instrument

The survey deployed in both countries was built on the Behavioral-Cognitive Internet Security Questionnaire (BCISQ), which uses a 5-point Likert scale to gather data on respondents' behavior and awareness of using solutions based on information and communications technology [59]. Our survey consisted of 3 main parts: the first part gathered background information, including age, gender, and profession; the second part included the BCISQ; and the final part collected insights on risk perception for assisted living technologies or diabetes technology and smartwatches.

The BCISQ measures behavior and perception of information security risk when using the internet in 2 ways: first, by having participants assess and respond to various "risky" situations, and subsequently, by simulating a scenario that asks for the respondent's most commonly used password question. In Norway, the simulation was presented through the open-ended question: "To check the quality of your password, write your most used password." This could not be replicated in BC due to ethical considerations and concerns by institutional research privacy leaders. Instead, we posed the question, "If we asked you to reveal your most used password to evaluate the security of your password, would you reveal it?" with the answer option "yes/no". Two subscales measure the cognitive dimension; the first scale captures an evaluation of information security risk, utilizing the term "likelihood" as per Adams [60] to understand how users evaluate the probability of cyber risk, while the second subscale captures how important users evaluated security mechanisms in managing these risks, following the approach of Velkie and Solic [59] who evaluated the individual's assessment of risk and perceived importance of security mechanisms. For example, the first scale contained questions to evaluate risk, such as "How would you rate the likelihood of someone stealing your identity on the internet (some examples of services that use your personal identity include online banking, social media, and e-mail)?" with options ranging from "very unlikely" to "very likely", while importance was captured by a direct question, such as "How would you rate the importance of updating your smartphone or laptop with the latest software?", with options on a scale of "not very important" to "very important." [59,60].

We extended the survey to encompass questions related to assisted living technologies or diabetes technology and smartwatches to enhance our understanding of risk perceptions associated with specific technologies. Smartwatches were intended to target a technology widely recognized by the general population yet not primarily serving a dedicated health function. This allowed us to explore if risk perception differs between assisted living and diabetes technologies versus smartwatches. The survey sought to capture respondents' evaluations of the likelihood and impact (consequence) of incidents compromising confidentiality, integrity, and availability. Continuing the structure of the BCISQ, responses were collected on the same 5-point Likert scale.

Before deploying the surveys in Norway and BC, we conducted preliminary tests with researchers, mainly from the health care field. These tests led to minor modifications in the question structure and sequence, ensuring clarity and relevance in both contexts. The survey was distributed in Norwegian in Norway and in English in BC. The surveys can be found in Multimedia Appendix 1 (English version) and Multimedia Appendix 2 (Norwegian version).

Interviews

A semistructured interview approach was adopted, and interview guides were prepared in advance. These aimed at gathering comprehensive data on key areas: (1) understanding of assisted living technologies or diabetes technologies; (2) cybersecurity risk understanding; (3) cybersecurity threats and vulnerability specific to these technologies; (4) behavior, attitude, and awareness; and (5) organizational and management practices for technology and security. Because of the varied knowledge and characteristics of the stakeholders, we developed customized interview guides. In Norway, interview guides were tailored to each stakeholder group to reflect their unique perspectives and roles. In BC, an initial interview guide was developed and adjusted according to the specific stakeholder being interviewed. Primary themes were consistent, yet the flexibility in the interview process allowed for follow-up on specific topics based on the interviewee's responses. For instance, health care workers familiar with one of the technologies were asked more detailed follow-up questions regarding their risk perceptions of that technology. Conversely, for those less familiar with specific technologies, discussions were generally focused on risk perceptions of assisted living technologies or smartwatches. The interview guides are attached in Multimedia Appendix 3 (English version) and Multimedia Appendix 4 (Norwegian version).

Analysis

Descriptive statistics were applied to the survey data in both data sets to summarize participants' characteristics and responses to questions related to the BCISQ (behavior and attitude) and risk perception. Variability was assessed using Cronbach α, indicating the internal consistency of the items. A significant challenge in our analysis was comparing the 2 data sets subject to different privacy regulations. The Norwegian data were stored in Norway, and data transfer to other countries, especially outside the European Union, was prohibited. Similarly, the BC data were stored in BC, and transferring data outside Canada was prohibited. The data sets were analyzed separately to overcome these challenges and account for variation in the

XSL•FO
RenderX

specific stakeholder groups participating in the 2 countries. Subsequently, the results of the aggregated data were subjected to further analysis and comparison. We used R statistical software (version 4.3.2; R Foundation for Statistical Computing) and Excel (Microsoft Corporation) to analyze the quantitative data.

Interviews were transcribed and verified against the recordings, followed by analysis using NVivo (Lumivero) for thematic analysis. We used a stepwise deductive-inductive approach. This process began with establishing codes and progressed to linking codes via axial coding, facilitating the identification of relationships between codes [61]. Subsequently, we developed categories encompassing multiple codes, enabling the identification of key concepts and themes. The stepwise inductive-deductive method allowed us to continuously refine our understanding of empirical findings supported by theory [61]. This approach also aligns with the suggested approach for conducting thematic analysis by Braun and Clarke [62]. Throughout the process, the researchers discussed recurring themes to enhance the consistency of the analysis. When analyzing the qualitative data, we completed separate analyses to prevent data transfer and then aggregated and interpreted data on themes and recurring topics between both data sets.

Ethical Considerations

Ethical approval in Norway was obtained from the Norwegian Agency for Shared Services in Education and Research (606692) on March 21, 2022, with updated approval on August 15, 2023, to reflect the data from BC. Ethical approval in BC was granted by the University of British Columbia/Children's and Women's Health Centre of British Columbia Research Ethics Board (H23-01987; date of approval 2023-11-27; principal investigator: MG). Implied informed consent was used for surveys, while electronic consent was used for interviews. Written informed consent was obtained for all interviews. Before each interview, participants were emailed a consent form. As a result of different privacy legislation, the forms were collected in different ways. We used the Research Electronic Data Capture (REDCap) electronic consent framework for BC, while in Norway, the forms were distributed and collected via email, with participants signing and returning the consent form [63].

Privacy and confidentiality protection were priorities in this study. Data were analyzed once samples were deidentified. Norwegian data were stored on an encrypted device with password protection, and the tool used to collect survey data did not collect metadata or IP addresses to prevent the identification of human participants. In BC, data were collected through REDCap, which also did not record IP addresses. Extracted data were stored on research institute network drives, with access restricted to research team members.

Norwegian research participants received no compensation for their participation. In BC, survey participants were entered into a raffle for approximately US $7.50 gift cards, while all the interview participants received an approximately US $15 gift card.

Results

Participation Overview

We received 274 survey responses (Norway, n=185; BC n=89). Participants were predominantly female (139/185, 75.1%, in Norway and 57/89, 64%, in BC) and most respondents were between 21 and 56 years old; 45 interviews were completed (Norway n=31, BC n=14). The median age category was 46-50 years in Norway and 41-45 years in BC (Table 1).

XSL•FO
RenderX

Table 1. Overview of the data sample.

Demographics	Survey		Interview	
	Norway (n=185)	British Columbia (n=89)	Norway (n=31)	British Columbia (n=14)
Age (years) group, n (%)				
15-20	N/A[a]	5 (5.6)	N/A	N/A
21-25	6 (3.2)	12 (13.5)	N/A	N/A
26-30	15 (8.1)	6 (6.7)	N/A	N/A
31-35	23 (12.4)	11 (12.4)	N/A	N/A
36-40	21 (11.4)	13 (14.6)	N/A	N/A
41-45	29 (15.7)	6 (6.7)	N/A	N/A
46-50	23 (12.4)	4 (4.5)	N/A	N/A
51-55	13 (7.0)	11 (12.4)	N/A	N/A
56-60	27 (14.6)	8 (9.0)	N/A	N/A
61-65	21 (11.4)	7 (7.9)	N/A	N/A
66-70	3 (1.6)	2 (2.2)	N/A	N/A
71-75	4 (2.2)	2 (2.2)	N/A	N/A
76-80	0 (0)	1 (1.1)	N/A	N/A
Missing	N/A	1 (1.1)	N/A	N/A
Gender, n (%)				
Male	45 (24.3)	26 (29.2)	15 (48.4)	9 (64.3)
Female	138 (74.6)	57 (64.0)	16 (51.6)	5 (35.7)
Prefer not to answer/others	2 (1.1)	6 (6.7)	N/A	N/A
Participant type, n (%)				
Health care worker	171 (92.4)	8 (9.0)	20 (64.5)	4 (28.6)
Industry professional/vendor	N/A	N/A	6 (19.4)	2 (14.3)
Health authority personnel	N/A	N/A	3 (9.7)	N/A
Patient (adult or adolescent) or family member of the patient	N/A	69 (77.5)	N/A	7 (50.0)
Biomedical engineer, hospital, or research information technology professional	N/A	12 (13.5)	N/A	1 (7.1)
Cybersecurity expert	N/A	N/A	2 (6.5)	N/A
Others	14 (7.6)	N/A	N/A	N/A

[a]N/A: not applicable.

BCISQ

BCISQ data demonstrated that both respondent groups assessed their behavior as low-risk (Table 2). The groups had a median score of 1 and 1.1 (never). The range of the group demonstrates that few of the respondents ever engage in “risky” behavior, regardless of country. On the cognitive likelihood (risk) scale, both groups demonstrated a neutral to unlikely understanding of the occurrence of an undesired event. However, the range was wider in Norway (range 1.6) versus BC (range 1.0). Norway had a median score of 4 on the cognitive importance scale, while BC had a 3.8 median score.

The last instrument of the BCISQ was the behavioral simulation. In total, 46 out of 185 (24.9%) Norwegian respondents answered the question: 17 out of 185 (9.2%) were deemed to have provided a possible actual password based on password requirements (capital letters, special characters, and password length), while the remaining 29 provided responses that were not deemed to be their actual password. Of the BC respondents, 16 out of 89 (18%) answered that they would provide their password. Moreover, when asked about password strategy, the majority in both groups reported that they only implemented minor password changes when changing passwords (94/185, 50.8%, in Norway and 45/89, 51%, in BC).

XSL•FO
RenderX

Table 2. BCISQ[a] subscale data split by region.

Region	Norway, median (IQR)	British Columbia median (IQR)
BCISQ subscales		
Risky behavior self-assessment scale	1.0 (1.0-1.0)	1.1 (1.0, 1.2)
Cognitive likelihood (risk) scale	3.0 (2.2-3.8)	2.8 (2.2, 3.2)
Cognitive importance scale	4.0 (3.25-5.0)	3.8 (3.0, 4.2)

[a]BCISQ: Behavioral-Cognitive Internet Security Questionnaire.

Reliability Analysis

The Norwegian data sample demonstrated internal consistency with Cronbach α, which was above the acceptable rate (0.7) on the overall questionnaire and the cognitive likelihood and importance scales, but lower on the behavioral scale. The BC responses demonstrated acceptable internal consistency overall and for the cognitive likelihood and importance scales, but the risky behavior self-assessment measurement was not within the acceptable range (Table 3).

Table 3. Cronbach α for the whole BCISQ[a] instrument and each submeasurement, split by region.

Behavioral-cognitive scale	Norway	British Columbia
BCISQ (total)	0.767	0.718
Risky behavior self-assessment scale	0.643	0.446
Cognitive likelihood (risk) scale	0.882	0.745
Cognitive importance scale	0.752	0.746

[a]BCISQ: Behavioral-Cognitive Internet Security Questionnaire.

Risk Perception for Assisted Living Technologies and Diabetes Technologies

For assisted living and diabetes technologies, the likelihood (probability) of confidentiality, integrity, or availability being compromised had a median score of 2 (unlikely) to 3 (neutral) among all respondents (Figure 2).

Figure 2. Risk perception related to the consequences (impact) and likelihood (probability) of cybersecurity-related issues, split by risk dimension, technology type, and location. Data are shown as violin plots (in grey) showing the data distributions, overlaid by boxplots (blue box denoting the first and third quartile, with the red dot denoting the median) with the raw data superimposed as black dots. BC: British Columbia.

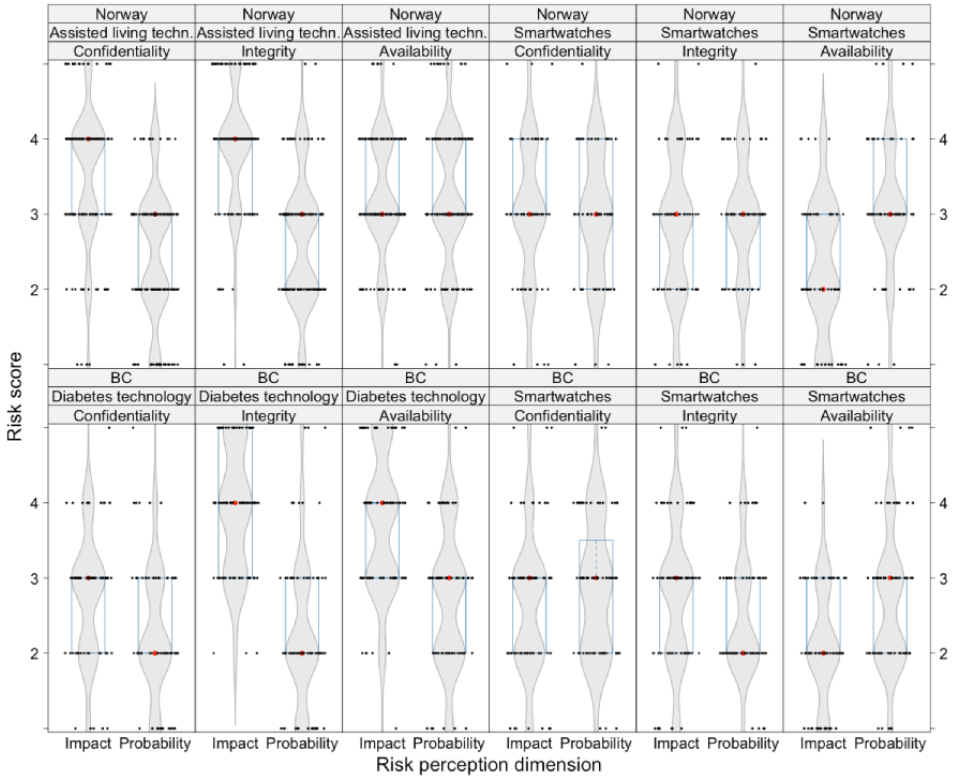

XSL•FO
RenderX

Because of a survey logic error, which resulted in optional risk perception questions, the Norwegian sample was limited to 147 out of 185 (79.5%) participants. The range for availability was slightly higher for assisted living technologies (IQR 3-4). Results for consequences (impact) were evaluated higher, between 3 (moderate) and 4 (major), indicating a more serious evaluation of consequence. For diabetes technologies, the range for integrity had the highest consequence, between 3 (major) and 5 (catastrophic). The Cronbach α was 0.679 for the Norwegian data sample and 0.664 for the BC sample.

Risk Perceptions for Smartwatches

For smartwatches, the likelihood (probability) of confidentiality, integrity, or availability being compromised had a median in both populations between 2 (unlikely) and 3 (neutral), with most results at 3 (neutral; Figure 2). The IQR was between 2 and 3 for all measurements. The results for consequences (impact) were between 2 (minor) and 3 (moderate). As a result of a survey logic error, which resulted in optional risk perception questions, the Norwegian sample was limited to 65 out of 185 (35.1%) participants. For the smartwatch risk perception measurement in Norway, Cronbach α was 0.812, and for BC, it was 0.694.

Interview Findings

We conducted 45 interviews (Norway n=31, BC n=14), involving 25 females (Norway n=17, BC n=8) and 23 males (Norway n=18, BC n=5). The initial coding process resulted in 72 codes for Norway and 61 for BC. After aggregating into categories encompassing all codes, we had 10 categories for BC and 10 categories for Norway, which mapped to 3 overarching themes derived from the theory on contextual factors in cybersecurity: organizational, technological, and human factors. As some categories overlapped, this resulted in 14 themes (Textbox 1).

Textbox 1. Contextual factor category and themes and subthemes.

Organizational factor

- Health care organization, systems, and processes
- Regulations, compliance, and ethics
- Risk management, safety, and privacy
- Vendor and third-party relationships
- Goals and objectives
- Cost and accessibility

Technological factor

- Technology, implementation, and effect
- Security and privacy
- Technology perception and data sharing
- Device reliability, criticality, and impact

Human factor

- Stakeholder perspective and responsibility
- Knowledge, competence, and awareness
- Trust
- Benefit and convenience

Organizational Factors

Contextual factors encompass factors specific to organizations, such as organizational structure, culture, and roles and responsibilities, which affect how risks are perceived and managed. In the analysis, we identified themes that included several contextual factors (Table 4).

XSL•FO
RenderX

Table 4. Organizational factors identified from interviews.

Theme and factor	Explanation	Example quotes from participants in Norway (N) and Canada (C)
Health care organization, systems, and processes		
Management of security and technology	This refers to how cybersecurity and technologies are managed within an organization to protect the PGHD[a] and medical devices. This influence risk perception through organizations capabilities and security practices.	• "We talked about silos. Health is no longer just health; it is IT, technological, education. We are so dependent on that collaboration and the flow between them, especially when we start with technology and undergo this digital transformation, it is difficult to sit in separate silos. It is all so interconnected." [N10] • "When it (devices) is kept within the hospital, it is kept there, but when you take out the solutions, they are exposed to new risks, and you have to go slowly to make it safe." [N16]
Organizational structure	Organizational structure refers to how different health care organizations are structured, which can impact the abilities and clarity of security management.	• "(...) we have different layers. [In] primary care, specialist care, and tertiary care, on the other hand, we also see how the organizational aspects are managed and controlled by different authorities, which makes the issue [cybersecurity risk] even more difficult to deal with." [N31] • "Specialist health care is much more competent and centralized, also in terms of competence. I experience that generally from the municipalities, who do not have a lot of competence. They place a lot of trust in us." [N29]
Understanding of responsibility	Refers to how responsibility for security is organized and shared. When responsibility is defined and understood, it indicates that risks are addressed and technologies are safeguarded.	• "One important thing in Norwegian health care is that there are interfaces that will be challenging and have not been addressed. (...) Where is the interface, and how should it be managed? Where should the data go, and who is responsible for the security?" [N15] • "(...) like any personal information that I have responsibility for, but even that is sometimes in conflict with the guidance we get from the hospital or university. Then that creates like this other layer of, I don't know if you want to call it moral distress, I mean, like who am I supposed to listen to?" [C6]
Vendor and third-party relationships		
Vendor expectations	The security expectations of vendors and the belief that vendors have the abilities and capabilities to meet these are also important.	• "(vendors) must be capable of offering solutions that do not fall apart once you start picking (investigating) them." [N30] • "So I trust the Medical Device Company more both in terms of reliability of the device, but also the in terms of it's data security, right?" [C14]
Dependencies	Dependencies can heighten or lower risk perception based on the perceived criticality of dependency. The more dependencies, the more complex a system is, which can introduce vulnerabilities and potential risks.	• "No matter if it is in the health care sector or not, long vendor supply chains are a challenge. One thing is development and drift. Development with a long supply chain can [mean that] those who develop are too far from those using them. There is too big [of a] distance." [N17] • "There have not been any incidents yet at any of our vendors, but there are no guarantees. So that is a major concern. We are very dependent on these vendors." [N25]
Third parties	These refer to external dependencies handling sensitive data or critical operations/components.	• "Of our product, maybe 60-70% of our products are from third parties, such as equipment and software." [N24]

[a]PGHD: patient-generated health data.

Technological Factors

Technological factors include how technologies are implemented, which security mechanisms are integrated, and their functionality. This category refers to the characteristics and capabilities of the technology and how the respondents perceived cybersecurity risk based on their understanding of different technological and system-related factors (Table 5).

XSL•FO
RenderX

Table 5. Technological factors identified from interviews.

Theme and factor	Explanation	Example quotes from participants in Norway (N) and Canada (C)
Technology, implementation, and effect		
Technological dependencies	Describe the perceived reliance on technologies outside of the actor's control, which can heighten risk exposure.	• "...there are so many dependencies on the clients that we can't upgrade quickly enough – it takes a lot of time and a lot of money." [N15] • "We are increasing the digital footprint – the more sensors we get and the more systems we push out will increase the risk that someone can reach something." [N16]
Security mechanisms	The tools implemented to protect data and technology and the confidence placed in their effectiveness impact risk management capabilities and perceived risk.	• "I wouldn't say like a hack, but it's really easy to get information. So you really got to take safety measures if you want to not get your identity stolen or your accounts hacked, and that kind of stuff." [C4] • "Data, accuracy, and security. but you know, like you, you pay a price for it, right? Because devices are very expensive." [C14]
Device reliability, functionality, and criticality		
Reliability	Refers to the perceived trustworthiness of the device, the consistent performance of technology, and the absence of system failures.	• "I'm always a little bit sceptical about, you know, how reliable that data is, and how helpful it is for them to be tracking sometimes, at this micro level." [C6] • "(...) Then I would have to choose accuracy, because that's what I need to know, that what I'm I need to believe that what I'm seeing is real right, and it reflects the state of the patient in front of me. Right?" [C13]
Functionality	The different tasks a system or technology performs affect the user's utility of the technology	• "(worst case)(...) if it just stopped working altogether" [C3] • "I was kind of scared at the beginning. But I'm like, okay, whatever, as long as it works, it works, right?" [C12]
Criticality	This refers to the perceived importance of technology in health care, where failure can result in severe impacts, which would increase perceived risk.	• "Oh, what if you know bad actors kind of interject and you know they give like extra bolus, like, I don't know. Maybe some people might say like, 'Oh, you're, you know, like creating this novel story,' right? But who knows? Right? So those kinds of like it's fear always back up my mind. Like I said. I don't use my phone to bolus." [C12] • "I have been concerned, particularly when thinking about what I talked about before, like a looped system where the pump can control, or you can control both the blood glucose monitor and the pump with the same system." [C4]
Benefit and convenience		
Convenience	The accessibility and ease of use of technology.	• "Because anytime you add encryption, you add security, [but] you added a bit more inconvenience." [C2]
Understanding of the value of data	Understanding of the importance and sensitivity of data. The sensitivity level of data may affect the risks associated with it.	• "If they got a hold of your medical data, could they like - what would they do with it? I'm not aware of, like, what? Why would someone take that, and what to do with it unless it was going to be to blackmail you?" [C11] • "I feel like I expect that my data will be safe and protected." [C4]
Technology perception and data sharing		
Data sharing	Refers to the processes and procedures for data sharing and the attitude toward data sharing.	• "There is a huge security risk when you start talking about data flow and data sharing, that you have so many instances that you must have security at each instance." [N15] • "I think it's continuous. I don't have to like even I remember I had to delete the app (...) But I it seems like it has my information, and it continues to do so." [C12]

Human Factors

The concept of human factors refers to the physiological and behavioral elements that influence how cybersecurity and risks are perceived for systems and technology (Table 6).

XSL•FO
RenderX

Table 6. Human factors identified from interviews.

Theme and factor	Explanation	Example quotes from participants in Norway (N) and Canada (C)
Knowledge, competence, and awareness		
Knowledge	The level of understanding either an individual or organization has about cybersecurity, threats, and risks.	• "If we had more knowledge/competence, then maybe the problem would disappear." [N1] • "I think about this stuff all the time, and I worry about it a lot. In part because I feel like I don't understand it enough." [C6]
Competence	Cybersecurity refers to the skills and abilities one has in cybersecurity, enabling the individual to understand and manage cybersecurity and how cybersecurity incidents are managed.	• "I think everyone understands that this (cybersecurity) is important; however, there is a lack of competence." [N24] • "No, I believe it is humans. It is without a doubt that and lack of competence." [N29]
Awareness	The degree to which one recognizes the importance of cybersecurity and affiliated risk and how the individual behaves based on this.	• "It's not a question of if (an attack will occur), but when." [N30] • "I don't know if I've ever thought about it being altered like sort of like, nefariously or maliciously." [C6]
Stakeholder perspective and responsibility		
Fear and severity of consequence	The dangers or severity affiliated with the potential consequences of a cybersecurity incident.	• "I just sleep through my low blood sugars. That's scary. I might not be able to wake up in the morning right? (...) Usually the pump usually peeps so I'm usually like be able to treat myself, but at that time I don't know what happened, and so I was sleeping through (...)." [C12] • "If you think about the worst-case scenario, it could be that users are sitting at home who believe these alarms work, and they become ill for some reason and trigger this alarm, but it doesn't work as they expect it to. So, in the worst case, it could result in death." [N3]
Potential risk scenarios and concerns	The perceived incidents that could occur.	• "I don't know what the feasibility of is like somebody hacking someone's insulin pump and overdosing or underdosing (...) because I suppose it would be hacking the CGM and telling the insulin pump to do so like." [C15] • "A data breach is very serious because it involves very sensitive data for patients, and we have seen that it is very serious for the municipalities. It could greatly affect what people think about our organization's solutions, so I think that's actually the maximum consequence." [N24]
Trust		
Trust in technology	The confidence placed in the technology to be reliable and secure.	• "I think my trust. Yes. Although it's really a case of going from a state of no knowledge or no experience, and therefore may be no trust to yeah using this as a reliable trend monitor and saying, 'Yeah, this is fine,' you know, and if I need to verify it, I will do." [C13] • "I'm just hoping and trusting that the technology, the companies making these products. They have some kind of security at their end." [C9]
Trust in institutions/in organizations	The confidence placed in institutions or organizations to be reliable and secure in their management of systems, technology, and data.	• "I would say I trust them, and I trust that they keep my data private and they don't share it with anybody (...)." [C3] • "The fact that one (municipalities) just trusts the vendor (blindly) is not very good." [N29] • "(...) you are heavily influenced by the atmosphere of trust, so people trust each other." [N31]
Benefit and convenience		

XSL•FO
RenderX

Theme and factor	Explanation	Example quotes from participants in Norway (N) and Canada (C)
Perceived benefit	The advantages related to the use of technology or specific devices.	• "Because I know that it's helped my management a lot, so I think I would probably take risks for that." [C11] • "Just think, what the devices bring to the table is so life-changing that it would have to be really a risk that had it had been clearly identified that it was super high for anybody to really take much notice." [C9]
Cost/benefit	The trade-offs made between risk and benefits.	• "...it's really a balance between how the convenience of use and the level of protection you can provide" [C2]. • "So you know, like most decisions in medicine, are sort of either risk, benefit, or cost-benefit equation (...) where better is defined as whatever better is defined as, right? It may be a trade-off right?" [C13]

Discussion

Principal Findings

For assisted living technologies, cybersecurity risk was perceived as reasonably low. Yet, there were variations in how confidentiality, integrity, and availability were assessed between technologies. Findings from the BCISQ demonstrated similar behavior and awareness among respondents, but risk perceptions for different technologies varied. For assisted living technology, integrity and confidentiality were prioritized, whereas, for diabetes technology, availability was viewed as more critical. Moreover, our survey and interview findings indicated that risk perception is stronger toward assisted living and diabetes technologies than smartwatches. This nuance was further explored in interviews, where both groups acknowledged availability as a crucial aspect of device functionality. Although few respondents considered data or device manipulation likely, they expressed significant concern about the potential consequences if such events occurred. Risk perceptions of the respondents are significantly affected by contextual factors in the area of human factors, namely, knowledge, competence, familiarity, dread/security of consequence, perceived benefit, and trust. The technological factor of "functionality and criticality" was also found to affect risk perception, as demonstrated in the differences between "high-risk" and "low-risk" devices. These are the factors evaluated to have the greatest implications for risk perception. Organizational factors such as health care sector organization were identified to have an effect, but based on the interviews, less than the others.

Risky Behavior and Information Security Awareness

The results from the BCISQ demonstrated that our respondents assessed their behavior as secure, with a more unlikely than likely chance of being exploited. Security measures were evaluated as high in importance, which indicates a strong awareness among respondents. There were only minor variations in the data between the included participants, indicating that respondents have similar behavior and awareness about cybersecurity. However, a discrepancy was identified between how respondents self-assessed their behavior and how they behaved: 17 out of 185 (9.2%) Norwegian respondents and 16 out of 89 (18%) BC respondents were willing to engage in risky behavior by revealing their passwords to check password quality. In addition, 94 out of 185 (50.8%) Norwegian respondents and 45 out of 89 (51%) respondents reused passwords, which is viewed as risky behavior and bad practices within cybersecurity.

Comparison With Prior Work

Organizational Factors

In our study, factors such as organizational structure, roles and responsibilities, and expectations toward vendors influenced how risks were perceived. This is supported by previous research, which found that organizational factors can contribute to increase (amplify) or decrease (attenuate) perception of risks [35]. Moreover, previous research has found that cross-cultural differences can affect risk perceptions [34,35]. Our survey suggested minimal differences among respondents from Norway and BC in behavior, awareness, and view of risks for health care technology and smartwatches. Similarly, our interviews found similarities in risk perceptions between the participants, indicating that region (cross-cultural differences) has a minimal effect on risk perception in this case [44,45].

Technological Factors

We found similar understandings of the risks of assisted living and diabetes technologies based on the consequences if confidentiality, integrity, or availability were compromised. A minor difference was found where availability risks were evaluated higher in consequence for insulin pumps and CGMs (4=major), while integrity and confidentiality risks were evaluated higher for medicine dispensers and digital self-reporting (4=major).

Smartwatches were included to investigate if respondents perceived risks differently for "low-risk" compared with "high-risk" devices, as risk perception has previously been connected to the severity and fear of consequences [39,42,44,48,64,65]. Our survey identified differences in risk perception, where "high-risk" devices included assisted living technologies and diabetes technologies, and "low-risk" included smartwatches. The probability and consequences were evaluated higher for "high-risk" devices than "low-risk" devices (Figure 2). In interviews, the perceived consequences of a closed-loop insulin control system were viewed as potentially fatal, which was a consideration for adopting such technologies. Manipulation of data, compromising data integrity, could lead to devastating consequences, such as wrong insulin/medication dosage. This indicates that the function and criticality of the

XSL•FO
RenderX

device affect risk perception and technology adoption, which aligns with previous research [39,64,66,67].

Human Factors

Previous research has demonstrated that human factors such as knowledge, behavior, and emotions are essential for understanding cybersecurity risk [52]. Risk perception theory highlights that understanding, competence, and familiarity with both hazards and risks affect risk perception. Yet, the TAM views familiarity as an enabler for technology adoption [29,44,47,50,51,65,66]. Findings from our study agree with previous studies, where increased familiarity and knowledge of a technology resulted in a change in risk perception. Thus, the perceived severity of consequences affected risk perception, agreeing with previous research emphasizing that the severity level will increase or decrease risk perception [30-32,45]. The concept of dread has been investigated thoroughly within risk perception theories [66,68]. Dread associated with potential harm affects risk perception: in both interviews and our survey, consequences were evaluated as more severe, with potentially fatal consequences. Within TAM and risk perception studies, it has been found that perceived usefulness is important for technology adoption and that the benefits of technology lead to greater risk acceptance, followed by a lower risk perception [42-44,67]. Similarly, we found that respondents were willing to accept risk up to a certain level due to the perceived benefits of the devices. This indicates that convenience and benefit also shaped how risks were perceived, resulting in a lower risk perception.

The implication of trust on risk perception has been debated in previous work [40]. In our study, trust was an essential factor influencing risk perception. Our respondents expressed trust in health care institutions and medical device companies to ensure that technology is secure, previously conceptualized as social trust [42]. Trust is also placed on individuals, whereas if a user trusts others interacting with technologies, risks are perceived to be lower [43,44]. Lastly, there is also trust dedicated to technologies, as respondents perceived these as trustworthy due to a lack of incidents, which resulted in lower perceived risk for assisted living technologies and diabetes technologies.

Limitations

The starting point of this study was to investigate cybersecurity perceptions among stakeholders in 2 different countries. The first limitation concerns the inherent differences between these 2 countries. Although we recognize these differences, we can compare risk perceptions due to a focus on technologies deployed in the health care sector. We recognize differences between users of assisted living technologies and diabetes technologies; however, the findings provide valuable insight into cybersecurity risk perception. A limitation in the Norwegian data sample is the exclusion of patients, which should be addressed in future studies to provide a more comprehensive perspective. In continuing this research, including patients as part of the Norwegian sample would be advisable to capture their concerns about cybersecurity risks.

Next, due to a survey logic error that resulted in optional risk perception questions, the Norwegian sample was limited to 147 out of 185 (79.5%) participants for assisted living technologies and 65 out of 185 (35.1%) participants for smartwatches. This significantly reduces the power of the Norwegian sample in this dimension; nevertheless, we believe the observations still provide valuable insights.

Expanding the sample to include a wide range of patients in both countries would also allow a deeper exploration of how patient-specific factors, such as age, condition, or digital literacy, influence risk perception. Further, women are overrepresented in both samples (139/185, 75.1%, for Norway and 57/89, 64%, for BC). Although a more equal gender distribution would be desirable, it reflects the workforce composition in the health care sector in both regions [69,70]. Future research should aim for equal gender representation to ensure that one gender's perspectives do not disproportionately shape the findings. Recruitment strategies could be adjusted to target more men or other underrepresented groups in the sample. Additionally, we did not perform a subgroup analysis by age category to explore if age is a factor that can explain some of the variability observed in risk perception, which should also be considered in future work, including more diverse populations and larger sample sizes.

Additionally, we could not combine the data sets for further statistical analysis, which restricts the depth of our analysis. The technologies used as examples in the study (assisted living or diabetes technology) serve different purposes and groups, which may lead to varying risk perceptions. This variability must be considered when interpreting the results, as it may influence stakeholders' views on risk differently. Future research should aim to collect a larger sample, possibly through collaborations with multiple health care institutions. Additionally, future studies could consider longitudinal designs to identify changes in risk perceptions over time.

This study did not perform a subgroup analysis of survey responses by participant type, such as patients and clinicians. However, differences among patients and clinicians were observed in the interviews and, therefore, included in the discussion. Future research should investigate the effects of participant type in interpreting the survey data to more comprehensively explore the potential differences in risk perception by participant type or organizational role.

Next, the first author (AS) conducted the thematic analysis of interviews. This approach introduces the potential for bias, as a single researcher's perspective can influence the interpretation of qualitative data. To mitigate this, the themes and key findings were discussed among the researchers. Future studies should aim for multiple researchers to engage in the thematic analysis process, ensuring a more balanced analysis. Furthermore, the reliance on self-reported data from surveys and interviews inherently carries the risk of response bias, where participants might provide socially desirable answers or have recall bias.

Suggestions for Further Research

The BC cohort should be used to investigate differences between participant groups and user characteristics further. Including patients from the Norwegian population would also be interesting, as this could provide a more comprehensive

XSL•FO
RenderX

understanding of risk perceptions and enable the comparison between patients in Norway and BC. A comparative analysis of Norwegian and BC patient populations could reveal how factors such as health care infrastructure, cultural attitudes, and patient demographics influence cybersecurity risk perception. This broader understanding could also inform more tailored approaches to risk communication and management in different health care contexts.

This study has examined cybersecurity risk perception related to specific individual health devices, yet they are often combined—a technological complexity that should be explored further. Further research should explore these complexities by investigating how interoperable systems manage cybersecurity risks, such as CGMs and insulin pumps working in tandem. This would involve assessing the technical vulnerabilities introduced by their interaction and understanding how users perceive or experience those risks in real-world settings. It may also be worthwhile to explore whether various vendor or operating system considerations (eg, Android vs iOS for smartphones) impact users' risk perception.

Another promising direction would be to combine technological analysis of systems with human and behavioral studies, providing a comprehensive system-level perspective on cybersecurity risks in health care. Such research could offer insights into balancing cybersecurity's technical aspects with user experiences, ultimately contributing to more secure health care technology ecosystems. In addition, future studies should incorporate mixed method approaches, such as triangulating self-reported data with observational data or objective measures (eg, usage logs or system performance data), to validate and complement the findings. This would provide a more comprehensive understanding of risk perceptions and reduce the potential bias in self-reporting.

Practical Recommendations

Implementing technology is not straightforward, and one could argue it is even more challenging in health care due to patient safety and security considerations. Ensuring transparency about potential risks and threats is critical for fostering trust, as subjective perceptions of risk vary among interested parties, as identified in this study. For instance, when rolling out new devices such as closed-loop systems for diabetes management (T1D), health care workers and technology developers could hold regular informational sessions or consultations directed toward both care receivers (patients and end-users) and caregivers. Based on the identified needs, one could focus on explaining the security features of and fail-safe mechanisms implemented in these systems. Enabling ongoing risk communication could help reduce skepticism and perceived risk.

Moreover, a structured risk management process, including regularly updated risk and threat assessments, is essential. The study found that respondents were generally willing to accept a certain degree of risk when the perceived benefits and convenience of their device outweighed the potential risks. The balance between risk and benefit/convenience is a crucial driver of technology acceptance. For example, although users may have concerns about data privacy when using wearables, many still choose to use them because of the convenience and value of the health insights these devices provide—such as real-time glucose monitoring or medication assistance—as these are essential in managing their health.

In addition, strengthening the knowledge and competence of all relevant parties—developers, care receivers, and health care workers—is essential, as cybersecurity risks and threats can be challenging to understand, especially for novices. Awareness is a security measure that can mitigate risks derived from humans' interaction with technology, and education is another area that should not be neglected to increase cybersecurity understanding. Practical steps such as security awareness programs could be integrated into the onboarding process for new technologies. For example, when introducing T1D technology (insulin pump or CGM) or assisted living technology (digital medication dispenser or digital self-reporting forms) to new users, as a part of the adoption process, one could introduce training sessions that cover key security areas such as password management, recognizing phishing attempts, and fail-safe mechanisms of the platforms. Such initiatives should be developed in close cooperation with the developers and vendors of the technology, as they often hold more detailed information about the system's security features.

Conclusions

We found minimal behavior and risk perception differences among the participants from Norway and BC. Human factors and device functionality were most influential in shaping cybersecurity risk perceptions. Considering the expected reliance on assisted living technologies and remote care solutions for disease management in the coming years, insight into risk perceptions can assist in reducing fears among relevant parties so that technologies can be adopted more broadly. Understanding risk perceptions can strengthen risk management and awareness, and understanding how to develop mitigating strategies to improve cybersecurity risk management is essential. Different inhibitors and drivers exist in technology adoption, so it is essential to understand the needs and capabilities of specific technology users. Building competence, knowledge, and trust can be essential steps in this endeavor. For future technological innovation and development, security should be considered as a prerequisite rather than an afterthought—it should be a license to operate.

Acknowledgments

The authors thank the participants for their time and for sharing their insights, and Drs Holly Longstaff and Tibor van Rooij for their feedback on the Canadian study design. The study was supported by a 2020 BC Children's Hospital Research Institute External Salary Recognition Award (to MG). MG holds a Michael Smith Health Research BC scholar award (SCH-2020-0494)

XSL•FO
RenderX

and is supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2021-02833). Philanthropic funding from the BC Children's Hospital Foundation enabled the project.

Generative artificial intelligence tools were not used in this manuscript's writing, editing, or development. The authors verified all content generated by the Zoom transcription service.

Data Availability

Some data are available. Ethical approval in BC permits us to share deidentified survey data with other researchers upon request from the corresponding author, but not the interview transcripts. Approval in Norway does not permit us to share survey or interview data publicly.

Authors' Contributions

AS contributed to the conceptualization, methodology, formal analysis, investigation, data curation, writing (original draft and review and editing), and visualization. MG contributed to conceptualization, investigation, formal analysis, funding acquisition, writing (review and editing), and supervision. NW contributed to conceptualization, investigation, project administration, visualization, and writing (review and editing).

Conflicts of Interest

None declared.

Multimedia Appendix 1
British Columbia survey questionnaire.
[PDF File (Adobe PDF File), 403 KB - jmir_v27i1e64388_app1.pdf]

Multimedia Appendix 2
Norway survey questionnaire (Norwegian).
[PDF File (Adobe PDF File), 457 KB - jmir_v27i1e64388_app2.pdf]

Multimedia Appendix 3
Interview guide (British Columbia, Canada).
[PDF File (Adobe PDF File), 266 KB - jmir_v27i1e64388_app3.pdf]

Multimedia Appendix 4
Interview guide (Norway; in Norwegian).
[PDF File (Adobe PDF File), 1043 KB - jmir_v27i1e64388_app4.pdf]

References

1. Aceto G, Persico V, Pescapé A. The role of information and communication technologies in healthcare: taxonomies, perspectives, and challenges. Journal of Network and Computer Applications 2018 Apr;107:125-154. [doi: 10.1016/j.jnca.2018.02.008]
2. Irmici G, Cè M, Caloro E, Khenkina N, Della Pepa G, Ascenti V, et al. Chest X-ray in emergency radiology: what artificial intelligence applications are available? Diagnostics (Basel) 2023 Jan 06;13(2):216 [FREE Full text] [doi: 10.3390/diagnostics13020216] [Medline: 36673027]
3. Adams SJ, Henderson RDE, Yi X, Babyn P. Artificial intelligence solutions for analysis of x-ray images. Can Assoc Radiol J 2021 Feb;72(1):60-72. [doi: 10.1177/0846537120941671] [Medline: 32757950]
4. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731 [FREE Full text] [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
5. Ma B, Yang J, Wong FKY, Wong AKC, Ma T, Meng J, et al. Artificial intelligence in elderly healthcare: a scoping review. Ageing Res Rev 2023 Jan;83:101808. [doi: 10.1016/j.arr.2022.101808] [Medline: 36427766]
6. Rosen JM, Adams LV, Geiling J, Curtis KM, Mosher RE, Ball PA, et al. Telehealth's new horizon: providing smart hospital-level care in the home. Telemed J E Health 2021 Nov;27(11):1215-1224. [doi: 10.1089/tmj.2020.0448] [Medline: 33656918]
7. Alwan M. Passive in-home health and wellness monitoring: overview, value and examples. Annu Int Conf IEEE Eng Med Biol Soc 2009;2009:4307-4310. [doi: 10.1109/IEMBS.2009.5333799] [Medline: 19964350]
8. Kioskli K, Fotis T, Mouratidis H. The landscape of cybersecurity vulnerabilities and challenges in healthcare: security standards and paradigm shift recommendations. In: ARES '21: Proceedings of the 16th International Conference on Availability, Reliability and Security. New York, NY: Association for Computing Machinery; 2021 Aug 17 Presented at:

XSL•FO
RenderX

ARES '21: The 16th International Conference on Availability, Reliability and Security; August 17-20, 2021; Vienna, Austria p. 1-9 URL: https://dl.acm.org/doi/10.1145/3465481.3470033 [doi: https://doi.org/10.1145/3465481.3470033]
9. Beavers J, Pournouri S. Recent cyber attacks and vulnerabilities in medical devices and healthcarehealth care institutions. In: Advanced Sciences and Technologies for Security Applications. Berlin, Germany: Springer; Apr 2019:249-267.
10. Pycroft L, Aziz TZ. Security of implantable medical devices with wireless connections: the dangers of cyber-attacks. Expert Rev Med Devices 2018 Jun;15(6):403-406. [doi: 10.1080/17434440.2018.1483235] [Medline: 29860880]
11. Halperin D, Clark S, Fu K, Heydt-Benjamin T, Defend B, Kohno T, et al. Pacemakers and implantable cardiac defibrillators: software radio attacks and zero-power defenses. New York, NY: IEEE; 2008 May 18 Presented at: 2008 IEEE Symposium on Security and Privacy; May 18-22, 2008; Oakland, CA p. 129-142. [doi: 10.1109/sp.2008.31]
12. Novak M. 224 injured after glitchy diabetes app drains insulin pump batteries. Gizmodo. 2024. URL: https://gizmodo.com/fda-recall-injured-faulty-diabetes-app-insulin-pump-1851467117 [accessed 2024-05-21]
13. US Food and Drug Administration (FDA). Tandem Diabetes Care, Inc. recalls version 2.7 of the Apple iOS t:connect mobile app used in conjunction with t:slim X2 insulin pump with control-IQ technology prompted by a software problem leading to pump battery depletion. FDA. 2024. URL: https://tinyurl.com/4fhtrt7n [accessed 2024-05-21]
14. Patton J. Ontario home care service provider victim of cyber attack. Global News. 2018. URL: https://globalnews.ca/news/4280633/carepartners-home-care-provider-cyber-attack/ [accessed 2024-06-11]
15. Martin A. "Crisis situation" declared as two Swedish municipalities hit by cyberattack. The Record. 2022. URL: https://therecord.media/crisis-situation-declared-as-two-swedish-municipalities-hit-by-cyberattack [accessed 2024-06-11]
16. Khatiwada P, Yang B, Lin J, Blobel B. Patient-generated health data (PGHD): understanding, requirements, challenges, and existing techniques for data security and privacy. J Pers Med 2024 Mar 03;14(3):282 [FREE Full text] [doi: 10.3390/jpm14030282] [Medline: 38541024]
17. Giansanti D. Cybersecurity and the digital health: the challenge of this millennium. Healthcare (Basel) 2021 Jan 11;9(1):62 [FREE Full text] [doi: 10.3390/healthcare9010062] [Medline: 33440612]
18. George-Bogdan M, Meşniţă G. Healthcare under siege: the need to improve cybersecurity in the near future. New York, NY: IEEE; 2023 Jan 01 Presented at: 2022 E-Health and Bioengineering Conference (EHB); November 17-18, 2022; Iasi, Romania p. 1-4. [doi: 10.1109/ehb55594.2022.9991541]
19. Countries with the highest internet penetration rate as of April 2024. Statista. 2024. URL: https://www.statista.com/statistics/227082/countries-with-the-highest-internet-penetration-rate/#statisticContainer [accessed 2024-06-24]
20. Andress J, Winterfield S. The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice. Rockland, MA: Syngress; 2014.
21. Cartwright A. The elephant in the room: cybersecurity in healthcare. J Clin Monit Comput 2023 Oct;37(5):1123-1132 [FREE Full text] [doi: 10.1007/s10877-023-01013-5] [Medline: 37088852]
22. Read E, Gagnon D, Donelle L, Ledoux K, Warner G, Hiebert B, et al. Stakeholder perspectives on in-home passive remote monitoring to support aging in place in the province of New Brunswick, Canada: rapid qualitative investigation. JMIR Aging 2022 May 11;5(2):e31486 [FREE Full text] [doi: 10.2196/31486] [Medline: 35544304]
23. Nilsen ER, Dugstad J, Eide H, Gullslett MK, Eide T. Exploring resistance to implementation of welfare technology in municipal healthcare services - a longitudinal case study. BMC Health Serv Res 2016 Nov 15;16(1):657 [FREE Full text] [doi: 10.1186/s12913-016-1913-5] [Medline: 27846834]
24. Freiesleben SD, Megges H, Herrmann C, Wessel L, Peters O. Overcoming barriers to the adoption of locating technologies in dementia care: a multi-stakeholder focus group study. BMC Geriatr 2021 Jun 21;21(1):378 [FREE Full text] [doi: 10.1186/s12877-021-02323-6] [Medline: 34154542]
25. Sultana NN, Alam M, Sanaul HM. Identifying relevant stakeholders in digital healthcare. Berlin, Germany: Springer; 2121 Jul 26 Presented at: The First International Conference on Applied Intelligence and Informatics; July 30-21, 2021; Nottingham, UK URL: https://www.researchgate.net/publication/352167151 [doi: 10.1007/978-3-030-82269-9_27]
26. Barlow J, Bayer S, Curry R. Implementing complex innovations in fluid multi-stakeholder environments: experiences of 'telecare'. Technovation 2006 Mar;26(3):396-406. [doi: 10.1016/j.technovation.2005.06.010]
27. Peek STM, Wouters EJM, Luijkx KG, Vrijhoef HJM. What it takes to successfully implement technology for aging in place: focus groups with stakeholders. J Med Internet Res 2016 May 03;18(5):e98 [FREE Full text] [doi: 10.2196/jmir.5253] [Medline: 27143097]
28. Huang D, Rau PP, Salvendy G. Perception of information security. Behaviour & Information Technology 2010 May;29(3):221-232. [doi: 10.1080/01449290701679361]
29. Goh Z, Hou M, Cho H. The impact of a cause-effect elaboration procedure on information security risk perceptions: a construal fit perspective. Journal of Cybersecurity 2022 Jan;8(1):tyab026. [doi: 10.1093/cybsec/tyab026]
30. Slovic P. The perception of risk. In: Scientists Making a Difference: One Hundred Eminent Behavioral and Brain Scientists Talk About Their Most Important Contributions. Cambridge, MA: Cambridge University Press; 2016:179-182.
31. Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy Sci 1978 Apr;9(2):127-152. [doi: 10.1007/BF00143739] [Medline: 22861175]
32. Slovic P, Fischhoff B, Lichtenstein S. Why study risk perception? Risk Analysis 2006 May 30;2(2):83-93 [FREE Full text] [doi: 10.1111/j.1539-6924.1982.tb01369.x]

XSL•FO
RenderX

33. Fischhoff B, Bostrom A, Quadrel MJ. Risk perception and communication. Annu Rev Public Health 1993 May;14(1):183-203. [doi: 10.1146/annurev.pu.14.050193.001151] [Medline: 8323585]
34. Kasperson R, Renn O, Slovic P, Brown H, Emel J, Goble R, et al. The social amplification of risk: a conceptual framework. Risk Analysis 2006 May 29;8(2):177-187 [FREE Full text] [doi: 10.1111/j.1539-6924.1988.tb01168.x]
35. Kasperson RE, Webler T, Ram B, Sutton J. The social amplification of risk framework: new perspectives. Risk Anal 2022 Jul;42(7):1367-1380 [FREE Full text] [doi: 10.1111/risa.13926] [Medline: 35861634]
36. Rogers R. Cognitive and physiological processes in fear appeals and attitude change:a revised theory of protection motivation. In: Cacioppo J, Petty R, editors. Social Psychophysiology. New York, NY: Guilford Press; 1983:153-177.
37. Moen B. Risk perception, priority of safety, and demand for risk mitigation in transport. Doctoral thesis, Department of Psychology, Norwegian University of Science and Technology. NTNU Open. 2008. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/264963 [accessed 2025-03-10]
38. García DCJ, García DCA, López-Sánchez C, Dias P. Risk perception and psychosocial vulnerability in health behaviours. Risk Perception: Theories and Approaches 2016:51-76.
39. Sjoberg L. Factors in risk perception. Risk Anal 2000 Feb;20(1):1-11. [Medline: 10841699]
40. Sjöberg L. Risk perception: experts and the public. European Psychologist 1998 Mar;3(1):1-12 [FREE Full text] [doi: 10.1027/1016-9040.3.1.1]
41. Chen C, Ishfaq M, Ashraf F, Sarfaraz A, Wang K. Mediating role of optimism bias and risk perception between emotional intelligence and decision-making: a serial mediation model. Front Psychol 2022 Jun 3;13:914649 [FREE Full text] [doi: 10.3389/fpsyg.2022.914649] [Medline: 35719601]
42. Siegrist M, Cvetkovich G, Roth C. Salient value similarity, social trust, and risk/benefit perception. Risk Anal 2000 Jun;20(3):353-362. [doi: 10.1111/0272-4332.203034] [Medline: 10949414]
43. Haugli-Sandvik M, Lund MS, Bjørneseth FB. Maritime decision-makers and cyber security: deck officers' perception of cyber risks towards IT and OT systems. Int J Inf Secur 2024 Feb 02;23(3):1721-1739. [doi: 10.1007/s10207-023-00810-y]
44. Larsen M, Lund M. Cyber risk perception in the maritime domain: a systematic literature review. IEEE Access 2021;9:144895-144905. [doi: 10.1109/access.2021.3122433]
45. Reeves A, Parsons K, Calic D. Whose risk is it anyway: how do risk perception and organisational commitment affect employee information security awareness? Berlin, Heidelberg, Germany: Springer; 2020 Feb Presented at: 22nd International Conference on Human-Computer Interaction (HCII2020); July 19-24, 2020; Copenhagen, Denmark p. 232-249. [doi: 10.1007/978-3-030-50309-3_16]
46. Vestad A. Personality traits and security motivation. Stud Health Technol Inform 2022 Nov 03;299:183-188. [doi: 10.3233/SHTI220980] [Medline: 36325861]
47. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Hoboken, NJ: Prentice-Hall; 1980.
48. Tornatzky L, Fleischer M, Chakrabarti A. In: Lexington M, editor. The Processes of Technological Innovation. Lanham, MD: Lexington Books; 1990.
49. Wilkowska W, Offermann J, Spinsante S, Poli A, Ziefle M. Analyzing technology acceptance and perception of privacy in ambient assisted living for using sensor-based technologies. PLoS One 2022;17(7):e0269642 [FREE Full text] [doi: 10.1371/journal.pone.0269642] [Medline: 35789340]
50. Venkatesh V, Davis FD. A theoretical extension of the Technology Acceptance Model: four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
51. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
52. Rohan R, Funilkul S, Pal D, Chutimaskul W. Understanding of human factors in cybersecurity: a systematic literature review. 2021 Dec Presented at: 2021 International Conference on Computational Performance Evaluation (ComPE); December 1-3, 2021; Shillong, Meghalaya, India p. 133-140. [doi: 10.1109/compe53109.2021.9752358]
53. McEvoy TR, Kowalski SJ. Deriving cyber security risks from human and organizational factors – a socio-technical approach. CSIMQ 2019 Apr 30;2019(18):47-64. [doi: 10.7250/csimq.2019-18.03]
54. Triplett WJ. Addressing human factors in cybersecurity leadership. JCP 2022 Jul 22;2(3):573-586. [doi: 10.3390/jcp2030029]
55. Guise V, Anderson J, Wiig S. Patient safety risks associated with telecare: a systematic review and narrative synthesis of the literature. BMC Health Serv Res 2014 Nov 25;14:588 [FREE Full text] [doi: 10.1186/s12913-014-0588-z] [Medline: 25421823]
56. Nilsen ER, Stendal K, Gullslett MK. Implementation of eHealth technology in community health care: the complexity of stakeholder involvement. BMC Health Serv Res 2020 May 11;20(1):395 [FREE Full text] [doi: 10.1186/s12913-020-05287-2] [Medline: 32393265]
57. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
58. REACH BC - health research study platform in British Columbia. REACH BC. URL: https://www.reachbc.ca/ [accessed 2024-06-20]

XSL•FO
RenderX

59. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
60. Adams J. Risk. London, UK: UCL Press; 1995.
61. Tjora A. Kvalitative Forskningsmetoder i Praksis (4th edition). Oslo, Norway: Gyldendal; 2021.
62. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
63. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
64. Starr C. Social benefit versus technological risk. Science 1969 Sep 19;165(3899):1232-1238. [doi: 10.1126/science.165.3899.1232] [Medline: 5803536]
65. Huang D, Rau P, Salvendy G. A survey of factors influencing peoples perception of information security. In: Jacko J, editor. Human-Computer Interaction, Part IV. Heidelberg, Germany: Springer; 2007:906-915.
66. Slovic P. Perception of risk. Science 1987 Apr 17;236(4799):280-285. [doi: 10.1126/science.3563507] [Medline: 3563507]
67. Fischhoff B, Slovic P, Lichtenstein S. Weighing the risks: risks: benefits which risks are acceptable? Environment: Science and Policy for Sustainable Development 2010 Jul 08;21(4):17-38. [doi: 10.1080/00139157.1979.9929722]
68. Finucane ML, Alhakami A, Slovic P, Johnson SM. The affect heuristic in judgments of risks and benefits. J Behav Decis Making 2000 Jan;13(1):1-17. [doi: 10.1002/(sici)1099-0771(200001/03)13:1<1::aid-bdm333>3.0.co;2-s]
69. Labour Market Information (LMI) Directorate SC. British Columbia sector profile: health care. Government of Canada - Job Bank. 2023. URL: https://www.jobbank.gc.ca/trend-analysis/job-market-reports/british-columbia/sectoral-profile-health-care [accessed 2024-05-20]
70. Jia Z, Kornstad T, Stølen N, Hjemås G. Arbeidsmarkedet for helsepersonell fram mot 2040. Rapporter. 2023 Feb. URL: https://tinyurl.com/p22f3rs4 [accessed 2024-06-24]

Abbreviations

BC: British Columbia
BCISQ: Behavioral-Cognitive Internet Security Questionnaire
CGM: continuous glucose monitoring
COREQ: Consolidated Criteria for Reporting Qualitative Research
IT: information technology
PGHD: patient-generated health data
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes
TAM: Technology Adoption Model

Edited by A Coristine; submitted 17.07.24; peer-reviewed by R Yin, T Davidson; comments to author 28.09.24; revised version received 20.11.24; accepted 20.01.25; published 19.03.25.

Please cite as:
Skjelvik A, West N, Görges M
Identifying Contextual Factors That Shape Cybersecurity Risk Perception for Assisted Living and Health Care Technologies and Wearables: Mixed Methods Study
J Med Internet Res 2025;27:e64388
URL: https://www.jmir.org/2025/1/e64388
doi:10.2196/64388
PMID:

©Alvhild Skjelvik, Nicholas West, Matthias Görges. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Business Venturing in Regulated Markets—Taxonomy and Archetypes of Digital Health Business Models in the European Union: Mixed Methods Descriptive and Exploratory Study

Sascha Noel Weimar[1], MSc; Rahel Sophie Martjan[1], MSc; Orestis Terzidis[1], PhD

Institute for Entrepreneurship, Technology Management and Innovation (EnTechnon), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Corresponding Author:
Sascha Noel Weimar, MSc
Institute for Entrepreneurship, Technology Management and Innovation (EnTechnon)
Karlsruhe Institute of Technology (KIT)
Fritz-Erler-Str. 1-3
Karlsruhe, 76133
Germany
Phone: 49 721 608 47341
Email: sascha.weimar@kit.edu

Abstract

Background: Digital health technology (DHT) has the potential to revolutionize the health care industry by reducing costs and improving the quality of care in a sector that faces significant challenges. However, the health care industry is complex, involving numerous stakeholders, and subject to extensive regulation. Within the European Union, medical device regulations impose stringent requirements on various ventures. Concurrently, new reimbursement pathways are also being developed for DHTs. In this dynamic context, establishing a sustainable and innovative business model around DHTs is fundamental for their successful commercialization. However, there is a notable lack of structured understanding regarding the overarching business models within the digital health sector.

Objective: This study aims to address this gap and identify key elements and configurations of business models for DHTs in the European Union, thereby establishing a structured understanding of the archetypal business models in use.

Methods: The study was conducted in 2 phases. First, a business model taxonomy for DHTs was developed based on a systematic literature review, the analysis of 169 European real-world business models, and qualitative evaluation through 13 expert interviews. Subsequently, a 2-step clustering analysis was conducted on the 169 DHT business models to identify distinct business model archetypes.

Results: The developed taxonomy of DHT business models revealed 11 central dimensions organized into 4 meta-dimensions. Each dimension comprises 2 to 9 characteristics capturing relevant aspects of DHT business models. In addition, 6 archetypes of DHT business models were identified: administration and communication supporter (A1), insurer-to-consumer digital therapeutics and care (A2), diagnostic and treatment enabler (A3), professional monitoring platforms (A4), clinical research and solution accelerators (A5), and direct-to-consumer wellness and lifestyle (A6).

Conclusions: The findings highlight the critical elements constituting business models in the DHT domain, emphasizing the substantial impact of medical device regulations and revenue models, which often involve reimbursement from stakeholders such as health insurers. Three drivers contributing to DHT business model innovation were identified: direct targeting of patients and private individuals, use of artificial intelligence as an enabler, and development of DHT-specific reimbursement pathways. The study also uncovered surprising business model patterns, including shifts between regulated medical devices and unregulated research applications, as well as wellness and lifestyle solutions. This research enriches the understanding of business models in digital health, offering valuable insights for researchers and digital health entrepreneurs.

(J Med Internet Res 2025;27:e65725) doi:10.2196/65725

KEYWORDS

digital health; telemedicine; mobile health; business model; European Union; classification; archetypes; medical device regulations; mobile phone; artificial intelligence; AI

XSL•FO
RenderX

Introduction

Background

In recent years, digital health technologies (DHTs) have seen significant advancements, with some apps becoming prescribable by physicians and reimbursable by health insurers, thus making them accessible to patients at no additional cost via their smartphones [1]. Germany has pioneered this approach, and several other countries have since adopted similar measures [2]. This development marks a significant step forward in DHT adoption. However, the potential value of DHTs extends beyond mobile apps. They can support physicians in diagnostic and treatment decisions, streamline administrative tasks, or support clinical trials, among other functions [3,4]. With health care systems worldwide struggling with sustainability issues [3] due to aging populations and increasing numbers of chronic diseases [5], DHTs play a crucial role in delivering value, enhancing the quality of care, and reducing rising health care costs [6].

DHTs encompass systems that integrate computing platforms, connectivity, software, and sensors for health care and related use [7]. These digital technologies primarily consist of stand-alone software solutions. Despite their promise, integrating DHTs into health care systems presents significant challenges due to the complexity and stringent regulations of these systems. The European Union (EU) recently introduced new, more stringent medical device regulations [8]. Medical devices for human use in the EU are now governed by Regulation (EU) 2017/745 on medical devices (MDR) or Regulation (EU) 2017/746 on in vitro diagnostic medical devices (IVDR) [8,9]. These regulations aim to ensure high standards of quality and safety for medical devices [8]. Consequently, some DHTs are classified as medical devices, while others do not meet this criterion [7]. In addition, various countries in the EU are establishing new reimbursement pathways specific to DHTs [2], such as the fast-track approach for digital health apps in Germany mentioned previously [1]. This development opens new revenue options for DHT businesses. The complexity of health care systems is further compounded by the broad stakeholder landscape that includes health insurers, health care providers, and patients, among others [10]. New ventures in the emerging DHT segment face substantial challenges in bringing their solution to market [11]. These challenges include meeting the stringent requirements of the medical device regulations, such as the need to build a quality management system and collect clinical evidence [9]. These activities demand significant additional time and costs [8]. Furthermore, the revenue model poses a challenge because consumer willingness to pay is often low, with many accustomed to free health care services [12]. Reimbursement pathways for digital health are often unclear and vary between several EU countries [2]. Moreover, high-quality, real-world evidence is often required to secure reimbursement [1].

To overcome these challenges and realize the full potential of DHTs, establishing a well-defined and sustainable business model is one of the crucial elements for success [13]. A DHT business model can help to cocreate and formulate a set of critical success factors [5]. Although start-ups have limited resources, they can leverage their structural flexibility to explore new market segments, thereby playing a pivotal role in introducing and adopting innovative business models [14]. This innovation is particularly important because, although medical technologies are constantly introduced to the market, they rely on less inventive business models [15]. Consequently, emerging business models built on DHTs can address the shortcomings of traditional medical technologies, overcoming barriers to commercialization. Therefore, a structured understanding can help identify sustainable and innovative DHT business model types. This understanding can also support the development process of business models at different stages [16]. However, there is a significant gap in the structured understanding of the business models within the DHT domain. In addition, there is a lack of research on how medical device regulations impact digital business models despite these regulations having a considerable influence [17].

Objectives

This study aims to identify key elements and configurations of business models for DHTs and, by doing so, provide a structured understanding of the archetypal business models in use. Given the variations in medical device regulations across countries, this study focuses on the EU with a common medical device regulatory framework. To achieve this goal, we formulated 2 research questions (RQs):

- RQ1: What are the key observable elements of DHT business models in the EU?
- RQ2: Which archetypes of DHT business models in the EU can be identified?

To answer these RQs, we developed a DHT business model taxonomy based on the digital health business model literature, data from real-world start-ups, and expert interviews. Subsequently, we derived DHT business model archetypes through cluster analysis. This approach provides a systematic classification of DHT business models and a simplified overview that establishes a common language. In addition, this study contributes to business model theory in complex domains by describing key components, the influence of the regulations, and drivers for business model innovation. In practical terms, this clear structure, real-world cases, and archetypal descriptions can assist digital health entrepreneurs in identifying suitable business models for their DHTs. It may also aid venture capitalists and investors in systematically analyzing digital health ventures.

The remainder of this work is structured as follows: the next subsection provides an overview of current research on business models in digital health. The Methods section outlines the research methodology followed, which is based on 2 main research phases. The Results section presents the developed business model taxonomy and archetypes. Finally, the Discussion section discusses the results and concludes the study.

Related Work

Digital health has emerged as an umbrella term to describe the practice of using information and communication technology to address health care needs [18]. The main components of digital health include eHealth, which encompasses internet and

XSL•FO
RenderX

web technologies used in health care delivery; mobile health (mHealth), which focuses on mobile devices to administer health care services; and telemedicine, which involves remotely connecting physicians and patients [19,20]. The digital technologies used in the context of digital health are referred to as DHTs. DHTs, which are regulated as medical devices, are known as software as a medical device [7]. The term *software as a medical device* refers to software intended to be used for at least 1 medical purpose that is performed without being part of a hardware medical device [21]. The intended medical purpose, declared by the medical device manufacturer, qualifies a DHT as a medical device [9]. The EU definitions of a medical device, in vitro diagnostic medical device, medical purpose, and other relevant definitions in this context are provided in Table S1 in Multimedia Appendix 1.

A business model can be described as the mechanism by which a business creates, delivers, and captures customer value [22]. Osterwalder [23] defines the business model as a strategic plan translated into a conceptual blueprint, positioning the strategy as one layer removed from the business model itself. In this context, the business model lies between the business strategy and the business processes and can be used as a tool of alignment through explicit depiction of the business logic [13]. It implements strategies and reflects the realized strategy [24]. The broad scope of business model research includes several subareas, such as research on business model components, development tools, and taxonomies [16]. Among development tools, the business model canvas is widely used to support the development process [25].

Existing literature discusses business models of digital health and related subcategories, such as telemedicine, eHealth, and mHealth. Some conceptual work exists in this domain [14,26-28]; for instance, Pascarelli et al [28] and Velayati et al [27] conducted systematic literature reviews to identify key components of digital health and telemedicine business models. Much of the research is empirical, describing elements of business models in digital health, telemedicine, and eHealth [29,30]. Some studies use the business model canvas to frame their empirical research [31-34], while others present their own frameworks [35,36]. Sterling and LeRouge [30] identified value proposition, key processes, key resources, and the profit formula as key components, while Gehde et al [4] focused on the technologies used to distinguish between business models.

Classification is the broad concept of organizing knowledge [37]. Under this scope, taxonomies create a structure and organization of a knowledge field and can help to navigate it [38]. While taxonomies are typically built inductively from empirical data, the term *typology* often denotes deductively derived classifications [16], although these concepts are sometimes used interchangeably [38].

Some taxonomies exist in the realm of digital health [39-42]. These taxonomies primarily emphasize technical aspects and lean toward telemedicine, but they also provide some useful criteria in the context of business models, such as technology [40], care sectors [41], or data types [42].

Business model taxonomies categorize business models based on their unique features [16] and aim to formulate ideal types of businesses [43]. These types of taxonomies have recently seen increased attention [16]. A few studies have focused on digital health business model taxonomies; for instance, a validated taxonomic business model description for telemedicine highlights the dimensions of value proposition, value cocreation, value communication and transfer, and value capture [44]. Three patterns of business models were identified. Other specific business model taxonomies include those for specialized mHealth segments such as maternal and baby care [45] or Middle Eastern telemedicine [46].

In conclusion, the focus and results of the studies vary depending on the digital health domain and research methods used. While some research has explored the components of digital health business models, a comprehensive and systematic classification that accounts for unique features, such as medical device regulations, and is based on extensive empirical data remains lacking.

Methods

Overview

Our research design comprises 2 phases, as illustrated in Figure 1. These phases include the development of the taxonomy followed by the archetype development. This research design is inspired by similar work on business model taxonomy development [47-49].

Figure 1. Overview of the research design, which was organized into 2 main phases: taxonomy development and archetype development. Each phase encompassed multiple iterations and steps.

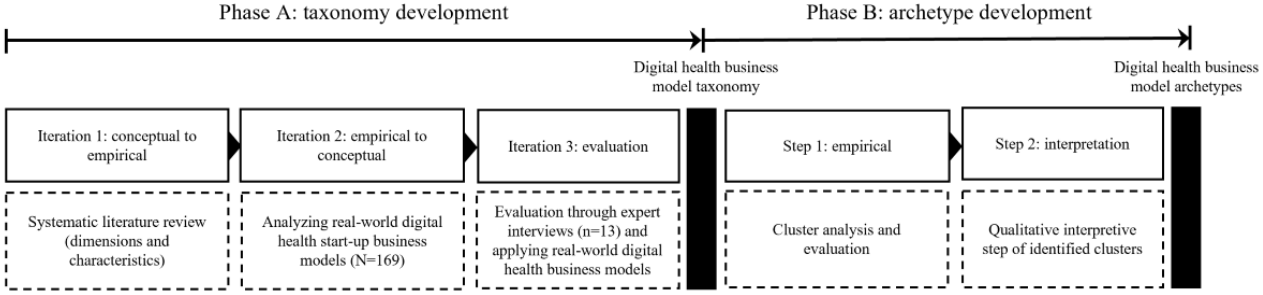

XSL•FO
RenderX

Phase A: Taxonomy Development

Overview

In the initial phase, we followed the iterative taxonomy development process proposed by Nickerson et al [38], incorporating recent enhancements by Kundisch et al [50]. The systematic process developed by Nickerson et al [38] is well established for its ability to integrate both empirical (inductive) and conceptual (deductive) approaches, allowing for iterative refinement until specified ending conditions are met [38]. Therefore, taxonomy design can be systematized, producing comprehensive and robust results. Kundisch et al [50] extended the taxonomy development process proposed by Nickerson et al [38] with more emphasis on taxonomy evaluation. Therefore, we focus on the approach of Nickerson et al [38] and integrate the suggestions for evaluation proposed by Kundisch et al [50], similar to previous taxonomy development [49].

Following the recommendations of Nickerson et al [38], our first step was to define a meta-characteristic to guide the taxonomy design. We selected "key observable and unique elements of DHT business models in the EU" as our meta-characteristic. Furthermore, our ending conditions were based on the work of Nickerson et al [38] and can be found in Table S2 in Multimedia Appendix 1. We selected the V4 business model dimensions developed by Al-Debei and Avison [13] as a theoretical lens for the taxonomy design, which is a common approach in business model taxonomy development [43]. Consequently, we categorized all identified taxonomy dimensions into 4 meta-dimensions: value proposition (offering value structure), value architecture (technological and organizational architecture), value finance (financial setups and return), and value network (business and customer actors web) as defined by Al-Debei and Avison [13]. Building on these foundations, we set out to develop the taxonomy in 3 iterations. The first iteration followed a deductive, conceptual-to-empirical approach grounded in a literature review. By contrast, the second iteration adopted an inductive, empirical-to-conceptual approach and relied on real-world observations, while the third iteration focused on evaluating the taxonomy through expert interviews. Throughout these iterations, we were mindful of potential biases related to our methods. Systematic literature reviews can be prone to publication selection biases, while qualitative research may introduce interpretive biases. To mitigate the risks of individual methods, we used a mixed methods approach and discussed results within the team to enhance objectivity. This process enabled us to incorporate diverse perspectives, resulting in a more comprehensive and robust taxonomy [38].

Iteration 1: Systematic Literature Review

The first iteration of the taxonomy development was based on a systematic literature review [51,52] to identify initial taxonomy dimensions and characteristics. A systematic literature review is a transparent and reproducible methodology for identifying relevant scientific work in a field [51]. This process involves planning the review, identifying and evaluating studies, extracting and synthesizing data, and disseminating the review findings [51]. An overview of the publication selection process is illustrated in Figure 2. To identify relevant publications, we conducted a comprehensive search in February 2024 across 3 major databases: Web of Science, Scopus, and PubMed. We applied an extensive search string to titles, abstracts, and keywords: ("digital health*" OR "mobile health*" OR "mhealth*" OR "m-health*" OR "ehealth*" OR "e-health*" OR "telemedicine*" OR "tele-medicine*" OR "telehealth*" OR "tele-health*" OR "software as a medical device*" OR "digital therapeutic*") AND ("business model*" OR "taxonom*" OR "archetyp*"). We did not limit the search period. This approach yielded a total of 2172 potential publications. After removing duplicates, titles and abstracts were scanned, and 3 inclusion criteria were applied: publications must (1) mention potential business model dimensions of digital health, (2) be written in English, and (3) be peer reviewed. Applying these inclusion criteria to the 2172 potential publications resulted in 96 (4.42%) publications. A subsequent full-text screening of these 96 publications narrowed this down to 27 (28%) final publications [4,6,19,26,28,30,32-36,39-42,44-46,53-61]. Almost half of the papers (13/27, 48%) use qualitative or conceptual research methods. A comprehensive list of the included publications, along with their study objectives, research methods, and key dimensions, is presented in Table S3 in Multimedia Appendix 1.

Figure 2. Visualization of the publication selection process as part of the systematic literature review, which involved searching 3 databases, applying a targeted search string, and using inclusion criteria to identify relevant studies.

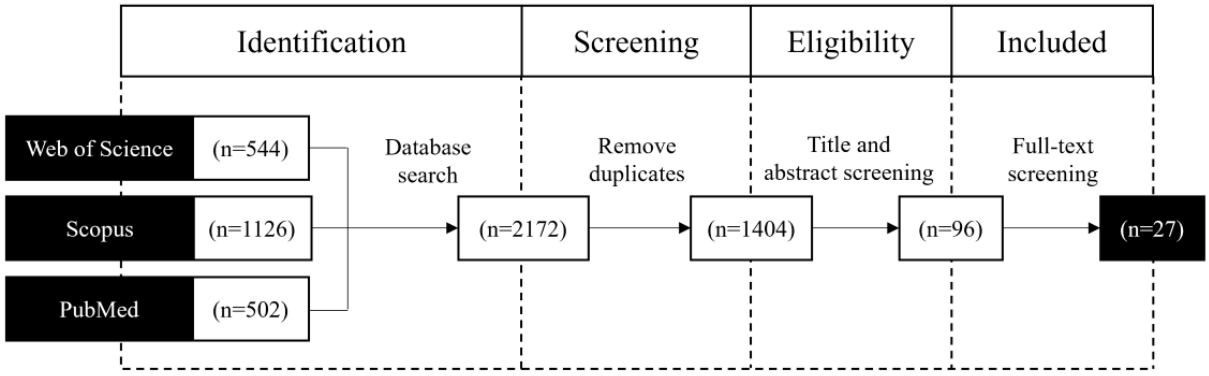

XSL•FO
RenderX

The knowledge gained from these publications was incorporated as a theoretical foundation into the subsequent analysis. Common themes were identified and then conceptualized as dimensions and characteristics based on the V4 meta-dimensions [13].

The meta-dimension *value proposition* was frequently highlighted as a central element of the business models in digital health [28,44], with the most commonly mentioned dimensions being the *main purpose* [4,44,45] and the *target user*, referring to the primary customer of a digital health company [6,33].

Under the meta-dimension *value architecture*, digital technology [44] and regulations [36,45] were identified as key business model elements. In addition, data often drive digital business models [26,33], leading us to introduce the dimensions *enabling technology*, *regulations*, and *data type*. These dimensions incorporate key resources [30] and dominate key activities [28,34] relevant to the business model. Recognizing the importance of embedding a business model into the health care system, we created the dimension *medical specialty level* [35,41] as the final dimension of *value architecture*.

For the third meta-dimension, *value finance*, we incorporated *paying entity* [44,45] to distinguish between users and payers, often different health care entities. Health care–specific reimbursement options were mentioned in the literature, where health insurers cover costs for digital health solutions [28,35]. This led to the establishment of the dimension *reimbursement*. In addition, *funding* was established as a further dimension [6,28].

In the final meta-dimension, *value network*, partnerships along the value chain were identified as part of the business model [26,28], prompting us to introduce the dimension *value chain partnerships*. Furthermore, digital health also plays a crucial role in connecting users through digital platforms [35,44,45], leading to the establishment of the dimension *platform interaction type*.

These initial dimensions were frequently cited in the literature; however, some variability between studies was observed, likely due to differences in research methods and the specific focus areas of the studies.

Iteration 2: Real-World Digital Health Business Models

Transitioning from the conceptual-to-empirical step, we adopted an empirical-to-conceptual approach for the second iteration of taxonomy development by analyzing real-world DHT business models. A fruitful approach is to analyze start-up business models because they often have novel digital business models and good availability and narrowness of data [43]. We used the HealthTech 250 list by Galen Growth [62] and the business intelligence platform Crunchbase to identify suitable start-ups to build our dataset. The HealthTech 250 list comprises 250 early-stage digital health ventures. On Crunchbase, by setting the industry group to “health care” and limiting the Crunchbase Rank, which is a measure of a company’s prominence [63], to 100,000 companies, we received an initial 19,610 results. To refine these results and focus on start-ups, inclusion criteria were established. First, the start-ups should be located in the EU and target this market. As medical device regulatory frameworks vary between countries, the EU enables the study of a common regulatory framework. Second, the solution developed by the venture should fit into the scope of DHTs. Third, the company should not be older than 10 years [64]. Next, the maximum number of employees should not exceed 250, aligning with the EU’s commonly defined upper limit for small- and medium-sized enterprises [65]. Finally, the company must be privately held and should not have merged with, or been acquired by, another entity. These criteria, which should ensure that the companies have a clear, identifiable business model and can be considered start-ups, were applied to both the HealthTech 250 list and the Crunchbase search. Afterward, we removed duplicates and performed purposeful sampling to create a balanced dataset and avoid overrepresenting specific business model types [66]. The complete list of companies, with relevant metadata, is shown in Table S4 in Multimedia Appendix 1. Next, we randomly scanned companies from our sample by reviewing their websites and other publicly available information to verify existing dimensions and characteristics and identify additional ones based on our meta-characteristic. If more than 1 applicable business model was identified for a company, each business model was considered separately, as depicted in the last step in Figure 3.

Figure 3. A summary of the selection process for real-world digital health technology business models, incorporating 2 primary sources for ventures and applying specific inclusion criteria.

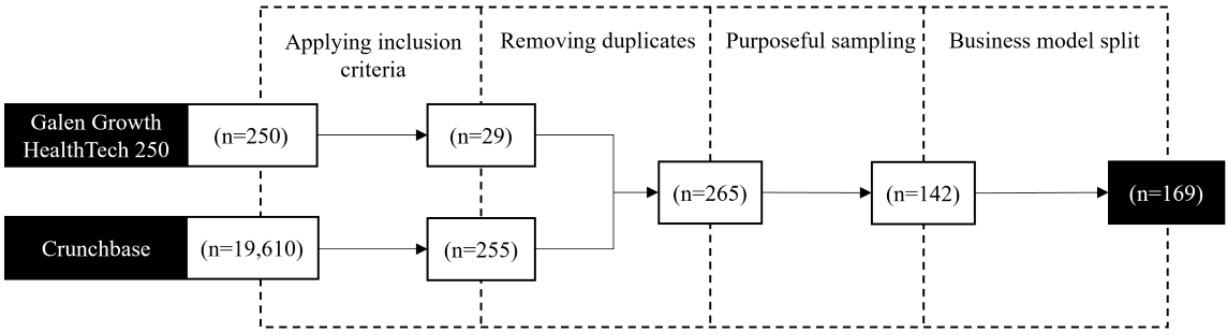

XSL•FO
RenderX

This process led to the identification of 12 dimensions and 64 characteristics. Finally, theoretical saturation was achieved after no new business models were identified for >20 companies in a row, and the second taxonomy development iteration was concluded.

Iteration 3: Qualitative Evaluation

It has been emphasized that the evaluation of a taxonomy is a crucial element of the taxonomy design that should not be ignored [50,67]. On the basis of this understanding, we extended the taxonomy development process proposed by Nickerson et al [38] with a third iteration, with the goal of triangulating the previous quantitative research approach with a qualitative one. Hence, we evaluated the previously developed taxonomy qualitatively through semistructured expert interviews. Interviews allow the collection of detailed and context-rich data, offering deep insights into a research topic [68]. We built our qualitative evaluation on commonly mentioned and applied criteria for taxonomy evaluation: completeness, simplicity, understandability, and robustness [50]. On the basis of these criteria and following the questions proposed by Szopinski et al [67] for taxonomy evaluations, we prepared an interview guideline. Thirteen experts from both academia and industry, with relevant expertise in digital health, business models, or taxonomy development were selected to participate. Each interview was conducted using videoconference software and was recorded and transcribed. An overview of the experts' roles, expertise, and interview durations (total duration: 407 min) is presented in Table 1.

Table 1. Overview of the interviewed experts' roles, expertise, and interview durations.

ID	Role	Expertise	Duration (min)
E1	Research group leader	Digital health and taxonomies	25
E2	Senior consultant	Taxonomies and business models	30
E3	Start-up consultant	Digital health and business models	47
E4	Cofounder	Digital health and business models	37
E5	Cofounder	Digital health and business models	28
E6	Cofounder	Digital health and business models	35
E7	PhD student	Digital health and business models	25
E8	Chief strategy officer	Digital health and business models	26
E9	Pharmaceutical industry consultant	Digital health and business models	38
E10	Venture capitalist	Digital health and business models	37
E11	PhD student	Digital health and business models	20
E12	Junior professor	Taxonomies and business models	25
E13	PhD student	Taxonomies and business models	34

The transcripts were inductively analyzed using the Gioia methodology [69]. On the basis of this analysis, we used the taxonomy operators *adding*, *renaming*, *merging*, and *removing* to rework dimensions or characteristics [50]. A table with the detailed changes after conducting the expert interviews can be found in Table S5 in Multimedia Appendix 1. We applied the updated taxonomy to our previously derived sample of DHT start-ups, resulting in no further changes to the taxonomy. To validate our findings, 2 additional raters coded a subsample (10%) of the real-world cases. The resulting Fleiss κ value of 0.79 indicates substantial agreement with the original classification [70]. This concluded the taxonomy development, with all ending conditions met (refer to the Results section for the final taxonomy version).

Phase B: Archetype Development

In the second phase, a cluster analysis was performed to develop archetypes of DHT business models. This method groups observations based on their similarities [71] and is the dominant mechanism for generating business model archetypes [43]. It is an objective and data-driven approach, thereby reducing potential biases. In addition, several clustering methods are available that enable fine-tuning and ensure that the clusters reflect meaningful and interpretable business model archetypes. Various approaches and methods for cluster analysis exist, raising questions regarding which similarity or dissimilarity measure to choose and which algorithm to use [72]; for instance, iterative partitioning methods such as k-means are considered superior to hierarchical algorithms [72], but they require predefined cluster numbers, which hierarchical algorithms do not require [72]. To overcome this limitation, we followed the 2-stage clustering approach proposed by Punj and Stewart [72].

In the first stage, we used the method proposed by Ward [73] to perform agglomerative hierarchical clustering. We used the simple matching coefficient as a similarity measure suitable for binary datasets [48]. The resulting clusters were visualized in a dendrogram (Figure S1 in Multimedia Appendix 1). Clusters that merge at greater heights in the dendrogram are more dissimilar. Therefore, cluster numbers were deemed reasonable in the range of 2 to 6 because the height differences between merge points dropped significantly beyond 6 clusters.

In the second stage of the cluster analysis, we performed k-means clustering with various cluster numbers to identify the optimal number [47]. The elbow rule indicated that 6 clusters were an optimal solution (Figure S2 in Multimedia Appendix

XSL•FO
RenderX

1). In addition to k-means, we conducted a similar analysis using algorithms such as k-modes, which are suitable for binary datasets. On the basis of these analyses, we examined the number of cases per cluster and discussed the interpretability and explanatory power of the clusters within the team. The 6-cluster solution of the k-means algorithm was identified as the strongest, with each cluster being well separated and interpretable (Table S6 in Multimedia Appendix 1). The average silhouette score of the solution was 0.35, suggesting a rather weak but substantial cluster structure in the data [74]. Social science data frequently lack strong natural groupings [75], and similar results have been observed in previous business model taxonomy research [48]. Therefore, we interpret the result as a good clustering outcome for the diverse and complex configurations of DHT business models observed in our study.

Finally, we conducted a qualitative interpretive step to interpret the results and the corresponding clusters by inspecting the frequency distributions of each characteristic within the clusters [75]. The frequency distributions for each cluster are presented in Table S7 in Multimedia Appendix 1.

Ethical Considerations

The qualitative data in this study, collected through interviews, have been deidentified to protect participant privacy and confidentiality. All interviewees provided informed consent for the use of deidentified data in this scientific publication. The research was conducted in accordance with the relevant ethics guidelines, and it was exempt from human participant ethics review.

Results

A Taxonomy of European DHT Business Models

The resulting taxonomy is presented as a morphological box in Figure S3 in Multimedia Appendix 1. In total, 11 dimensions were identified. Some dimensions are exclusive, allowing only 1 characteristic to be assigned, while others are nonexclusive, allowing multiple characteristics to be assigned. All characteristics in Figure S3 in Multimedia Appendix 1 are sorted alphabetically. The key elements of the taxonomy, based on 4 meta-dimensions, are discussed herein. A detailed description of each dimension and characteristic is presented in Table S8 in Multimedia Appendix 1, and an example of how a DHT business model can be categorized within the taxonomy is presented in Table S9 in Multimedia Appendix 1.

The *value proposition* meta-dimension outlines how a company offers value [13]. In DHT business models, 9 *main purposes* have been identified, indicating the type of value delivered. These purposes are divided into 2 main groups: those with a medical purpose and those without (refer to Table S1 in Multimedia Appendix 1 for the definition of *medical purpose*). Characteristics with a medical purpose include *diagnosis or prediction*, *monitoring support*, and *treatment or interventional purpose*. All these characteristics are related to specific medical conditions. The *treatment or interventional purpose* characteristic also encompasses preventive and rehabilitation interventions [76,77]. Characteristics without a medical purpose include *administrative support* and *information delivery or education*, among others. DHT business models can have multiple *main purposes*, including combinations of medical and nonmedical purposes. It is crucial to identify the primary source of value when selecting the *main purpose*. Another dimension of the *value proposition* meta-dimension is the *target user*, encompassing *patients or private individuals* and stakeholders directly involved in health care delivery, such as *health care providers*. Another target user is *indirect health care commercial firms*, such as pharmaceutical and medical technology companies or even *non–health care commercial firms*. Finally, some companies collect data to provide clinical and nonclinical *evidence*. Here, some companies base their evidence on established medical literature, while others provide real-world evidence, for example, through randomized controlled trials.

The *value architecture* meta-dimension describes the technical and organizational infrastructure of the business model [13]. This meta-dimension encompasses the integration of the business model into health care systems, considering the *level of care*, such as *primary*, *secondary*, or *tertiary care* [78,79]. This describes whether the value is generated through integrating solutions at the initial points of contact with health care systems or, alternatively, within highly specialized care settings. In addition, this meta-dimension includes the EU *medical device regulatory framework*, which comprises the *medical device regulation* covering most regulated DHTs with a medical purpose and the *in vitro diagnostic medical device regulation* for DHTs with a medical purpose using primarily biological data derived from human biological materials [9]. In addition, it recognizes the existence of *nonmedical devices* without a medical purpose. Digital business models are fundamentally driven by data as a key resource and often are built on an *enabling technology*. In total, 9 *data types* were identified, ranging from *medical imaging data* derived through x-ray imaging or magnetic resonance imaging to *physiological parameter data* captured using sensors to measure bodily functions. Some of the data are reported directly by patients or involve tracking user behavior, summarized under the dimension of *user activity or patient-reported data*. Examples of enabling technologies include *artificial intelligence*, *cloud or network technology*, and *sensor technology*.

Value finance is the part of the business model that focuses on financial setups and returns [13]. Therefore, the taxonomy includes the *paying entity*, which can be the *patient or other private individuals* such as caregivers, *health care providers*, or even companies indirectly involved in health care delivery. *Reimbursement* plays a vital role in the revenue model, with *health insurers and governments* covering expenses for DHTs, depending on the country-specific *reimbursement* pathways [2].

Finally, the *value network* is defined as the web of business and customer actors [13]. A key aspect is the *user interface*, which shapes how users interact with the solution. In addition, the business model taxonomy covers stakeholder interactions. Digital platforms can engage both internal and external actors [80], as illustrated in the dimension of *stakeholder interaction type*. This category examines network effects, focusing on user and company interactions facilitated by the solution. It assesses whether users engage with each other through the solution and

XSL•FO
RenderX

how the companies themselves interact with their customers using the solution.

Archetypes of DHT Business Models

Overview

Six archetypes of DHT business models were identified. The number of business models in these clusters ranges from 18 (10.7%) to 36 (21.3%) out of 169 business models. The frequency distribution for all characteristics of the archetypes is shown in Table S7 in Multimedia Appendix 1. Through an interpretive step, we analyzed the archetypes and assigned the following labels: administration and communication supporter (A1), insurer-to-consumer digital therapeutics and care (A2), diagnostic and treatment enabler (A3), professional monitoring platforms (A4), clinical research and solution accelerators (A5), and direct-to-consumer wellness and lifestyle (A6). Table 2 provides an overview of the archetypes, and the subsections that follow offer a comprehensive summary of the main findings for each cluster.

XSL•FO
RenderX

Table 2. Concise descriptions of the 6 business model archetypes identified through cluster analysis, along with corresponding examples.

ID	Archetypes	Distinguishing characteristics	Examples of underlying DHTs[a]
A1	Administration and communication supporter	• Business models around tools assisting with administrative tasks (eg, scheduling and report writing) and patient interactions • Not medical devices; often using cloud, network technology, and artificial intelligence; efficiency evidence not typically provided • Health care providers pay for these nonreimbursable solutions; interfaces are mainly mobile apps and web applications	Automated medical report writing, appointment management tools, and telehealth video consultation
A2	Insurer-to-consumer digital therapeutics and care	• Business models around supporting digital treatment of medical conditions, used mainly by patients and private individuals • Regulated as medical devices under the MDR[b], providing mostly high clinical evidence (peer-reviewed studies) • Focus on primary and secondary care, rely on health insurer reimbursement, and incorporate multimedia and patient-reported data; some act as support channels with the provider	Mobile app for treatment support of leaky gut, depression, or overweight
A3	Diagnostic and treatment enabler	• Business models for solutions assisting physicians in diagnosis or treatment planning, sometimes including administrative support • Regulated as medical devices under the MDR (medical imaging and physiological data) or the IVDR[c] (data derived from human biological materials) • Used and paid for by health care providers; use artificial intelligence; feature mobile apps or web applications; do not facilitate stakeholder interaction	Radiological imaging tools, treatment planning software, and in vitro diagnostic software to support the diagnosis
A4	Professional monitoring platforms	• Business models based on digital platforms to monitor patients' medical conditions, connecting health care providers with patients and private individuals • Classified as medical products under the MDR, supported by clinical evidence • Facilitate interuser interaction; integrate administrative support, telecommunication, and education; primarily paid for by health care providers, sometimes reimbursable	Physician platform to monitor progression of Parkinson disease based on mobile app and therapeutic exercise–monitoring platform
A5	Clinical research and solution accelerators	• Business models around the acceleration of clinical research or the provision of raw technology for pharmaceutical companies, medical technology companies, and research institutions • Some solutions are classified as medical devices under the MDR, while research applications are not regulated • Paid for by indirect health care firms and research institutions, may or may not facilitate stakeholder interaction and include developer tools as well as mobile apps and web applications	Diagnostic tools based on artificial intelligence that medical technology companies can integrate and software to support clinical trial execution through data collection
A6	Direct-to-consumer wellness and lifestyle	• Business models enabling information delivery and education for patients and individuals, with some telecommunication features and marketplaces • Not classified as medical devices • Include coaching through mobile apps and web applications; sometimes provided by medical clinics; targeting self-pay markets and corporate wellness; some are reimbursable and report nonclinical evidence	Mobile app for meditation, pregnancy support apps, and health and sports coaching mobile app

[a]DHT: digital health technology.

[b]MDR: Regulation (EU) 2017/745 on medical devices.

[c]IVDR: Regulation (EU) 2017/746 on in vitro diagnostic medical devices.

XSL•FO
RenderX

Administration and Communication Supporter (A1)

The first archetype comprises business models built on software tools designed primarily to assist health care providers with administrative tasks such as appointment scheduling, automated report writing, job matchmaking, professional education, and claims management. These administrative solutions are often combined with telecommunication modules that facilitate interactions between health care providers and patients or private individuals. Therefore, they also target private individuals. As the underlying solutions do not serve a medical purpose, they are not classified as medical devices, and companies typically do not provide evidence to support their efficacy. These business models operate across all levels of care. Communication is primarily enabled through cloud and network technology, with some solutions incorporating artificial intelligence. The revenue model predominantly involves health care providers who use and pay for these solutions, making them generally nonreimbursable for patients. The user interfaces are mainly mobile apps and web applications, and the solutions often connect health care providers with patients.

Insurer-to-Consumer Digital Therapeutics and Care (A2)

The second type of business model comprises software solutions that primarily support digital treatment and care through interventions for medical conditions. These business models target mainly patients or private individuals. Some of the business models additionally enable telecommunication with employees of the solution provider itself. They typically provide a high level of clinical evidence, such as through published peer-reviewed studies. The focus lies predominantly on primary and secondary care. Due to their medical purpose, solutions under this archetype are regulated as medical devices under the MDR. They incorporate multimedia data for educational purposes and capture user activity and patient-reported data. Most of the business models in this archetype rely on reimbursement through health insurers to generate revenue. For cases that are not reimbursed, patients or private individuals are payers.

Diagnostic and Treatment Enabler (A3)

Similar to the first archetype, the third business model archetype is mainly used and paid for by health care providers, specifically professionals such as physicians. In a few cases, these professional tools are reimbursed through health insurers or are paid for by patients and private individuals. They primarily assist physicians in the diagnosis or treatment of medical conditions. Although they sometimes include additional administrative support, they are regulated as medical devices under the MDR or the IVDR for their medical purpose. Business models under the MDR are substantially driven by medical imaging data, physiological parameter data, or medical laboratory results. By contrast, business models under the IVDR primarily use data derived from human biological materials, such as genomics and microscopic cell images. The business models are built on solid clinical evidence. This category also includes business models built on diagnosis software for patients and private individuals that share the same main purpose regarding diagnosis and prediction as professional tools and are nonreimbursable. The business models are often enabled through artificial intelligence, feature mobile app or web application interfaces, and do not facilitate stakeholder interaction.

Professional Monitoring Platforms (A4)

This archetype focuses on business models with the purpose of monitoring patients and their medical conditions through digital platform solutions that connect health care providers with patients or private individuals. Both the platforms for health care providers and the patient-side applications vary in complexity. Depending on the business model, several other purposes are integrated with the monitoring, such as diagnosis and prediction, administrative support, telecommunication, or information delivery and education. Monitoring is often achieved through the collection of user activity or, in many cases, through the collection of patient-reported data (mostly patient-reported outcome measures). The underlying solutions are classified as medical devices under the MDR and are supported by clinical evidence. Typically, the business model focuses on secondary or tertiary care applications. These platform business models facilitate interuser interaction by connecting patients and health care providers. Health care providers primarily bear the cost, but sometimes solutions are reimbursable.

Clinical Research and Solution Accelerators (A5)

The fifth archetype primarily targets indirect health care commercial firms, such as pharmaceutical or medical technology companies, as well as research institutions. The underlying solutions do not provide direct health care but serve as tools to accelerate drug discovery and clinical research by offering digital tools for study execution. This application occurs outside of direct medical care. In addition, some business models are built on offering raw technology that can be used by medical technology companies as part of their solutions. While some solutions are classified as medical devices under the MDR, others are not, because research applications do not fall under the MDR or the IVDR. The payers in this archetype are indirect health care commercial firms and research institutions. Some business models do not involve solution-side stakeholder interaction, while others facilitate interuser interaction, for example, through contact with study participants for clinical research. A distinctive feature of these business models is that the raw technologies are not only provided as mobile apps and web applications but also as developer tools.

Direct-to-Consumer Wellness and Lifestyle (A6)

The last archetype proposes value through information delivery and education for patients and private individuals. These can be simple, user-facing applications for general well-being, such as those providing guided meditation. Another type of value that is delivered by some of these business models is telecommunication, and a few offer marketplaces targeting patients and private individuals. A unique aspect of some of these wellness and lifestyle solutions is the interaction between the solution provider and users, where employees of the solution provider connect with users and coach them through mobile apps and web applications. Sometimes, the contact can be through a solution provider–operated medical clinic. Solutions

XSL•FO
RenderX

related to this business model archetype do not mention specific medical purposes and are not classified as medical devices. While some solutions aim at the self-pay market, others target non–health care commercial firms that are interested in corporate wellness, making them reimbursable for patients and private individuals. In addition, some companies report nonclinical evidence from their solutions, which might help to convince health insurers to reimburse costs.

Discussion

Principal Findings

The objective of this study was to identify key elements of business models for DHTs, aiming to develop a structured understanding of the archetypal business models used in the industry. This goal was achieved through a 2-phase process. In the first phase, we developed a taxonomy of DHT business models [38,50]. The taxonomy was based on a systematic literature review and an analysis of 169 real-world business model cases. It was evaluated qualitatively through 13 expert interviews. In the second phase, we conducted a cluster analysis based on hierarchical and iterative partitioning clustering methods [72], which resulted in the identification of 6 archetypal clusters. These clusters were subsequently evaluated and interpreted. This resulted in interesting findings regarding the dominant business models used for DHTs and enabled the study of business model innovation in the DHT domain.

Primarily, we want to mention the cascading effect observed for the *medical device regulatory framework* dimension. Medical device regulations significantly impact the business model design if a company delivers a value proposition based on a medical purpose, which applies mainly to archetypes A2, A3, and A4. Through this, these regulations have a strong influence, requiring ventures to adapt their strategies based on available resources due to the additional time, costs, and potential partnerships needed. Many medical device companies in our sample of digital health start-ups (49/142, 34.5%) showcased proof of clinical evidence because the medical device regulations require demonstrating sufficient evidence. Some of the companies (7/142, 4.9%) based their evidence on the state of the art, while others (42/142, 29.6%) presented results from their own clinical studies. Furthermore, clinical partnerships and other quality and security certifications were listed on the company websites. Therefore, medical device regulations exert a strong influence on business models. While medical device regulations have the largest impact, other EU and local regulations also contribute to the cascading effect; for instance, the Artificial Intelligence Act recently adopted by the EU is relevant if the business model relies on artificial intelligence technologies that pose risks deemed sufficiently high. Thus, some companies must navigate a complex regulatory landscape to ensure compliance and sustainability when designing their business models.

Second, we identified 3 key drivers of business model innovation within the DHT domain. The first driver of business model innovation pertains to the dimensions of users and payers. Unlike traditional medical technology companies, which often develop business models centered around hardware-intensive technologies operated by health care providers, DHT business models can directly engage patients, private individuals, and even corporate employers. This is facilitated by the integration of mobile interfaces with cloud and network technologies, promoting innovative business models, particularly in archetypes A2, A4, and A6. The second driver is the advent of artificial intelligence and the rise of digital technologies, which enable unprecedented value delivery. These advancements facilitate the creation of business models that were previously unimaginable, significantly driving archetypes A4, A5, and A6. The third driver of business model innovation is the revenue model, customized to the target countries' specific conditions. The self-payer market is more attractive in some regions, whereas others have established reimbursement pathways allowing cost coverage by 1 health insurer or multiple health insurers [2]. It has been observed that certain reimbursement pathways are exclusively available to medical devices or specific DHTs; for example, the reimbursement pathway for digital health applications in Germany [1] is limited to digital medical devices. This fosters new business models and drives innovation in archetypes A2 and A6. Companies in the study sample (26/142, 18.3%) engage in selective contracts with individual health insurers or use other pathways that facilitate the reimbursement of health care expenses. Therefore, the revenue model constitutes a critical element of DHT business models and must be tailored to the national context while ensuring alignment with other business model components. Start-ups should consider reimbursement pathways and possible related medical device requirements early in their strategy development.

Finally, the relationships among the identified business model archetypes reveal interesting connections. To discuss these relationships, we categorize the 6 identified archetypes of business models based on their compliance with medical device regulations and their primary target users, as depicted in Figure 4. The solutions underlying the business model archetypes A2, A3, A4, and some in A5 are regulated as medical devices. Notably, A3, A4, and some A2 business models face stringent medical device regulatory requirements because they provide critical tools for health care providers and patients. A3 stands out because it is the only archetype that includes in vitro diagnostic medical devices regulated under the IVDR. It was noted that these business models often emerge from academia as transfer projects. While most health care providers typically use and cover the costs of these solutions, 1 particular business model stands out and deserves mention: in this model, patients pay for enhanced diagnoses through professional software, and health care providers receive a share of this payment as additional revenue. By contrast, the underlying solutions of A6 and A1 are not classified as medical devices. Notably, some companies within these archetypes, as well as A2 companies, operate at the intersection of regulated medical devices and unregulated wellness and lifestyle products [81]. This implies that these companies might initially adopt an A6 business model (no medical purpose and not subject to medical device regulations) and gradually transition to an A2 business model as they gather more real-world evidence (medical purpose and subject to medical device regulations). A5 business models play a unique role because they operate outside of direct medical care and are not subject to medical device regulations as research

XSL•FO
RenderX

applications. Similar to the archetype A6, some companies adopting A5 business models can decide to progressively evolve into A3 business models over time as more real-world evidence is collected. In addition, a common approach is to combine A2 or A6 with A4. This combination involves a digital therapeutic, care application, or wellness and lifestyle solution that the patient uses, paired with a professional monitoring platform on the health care provider's side, enabling monitoring through the patient-side applications. In such cases, patients, health care providers, or health insurers might be payers.

Figure 4. A visual representation of the 6 business model archetypes, each accompanied by example ventures organized by target user–centricity and relevant medical device regulatory requirements.

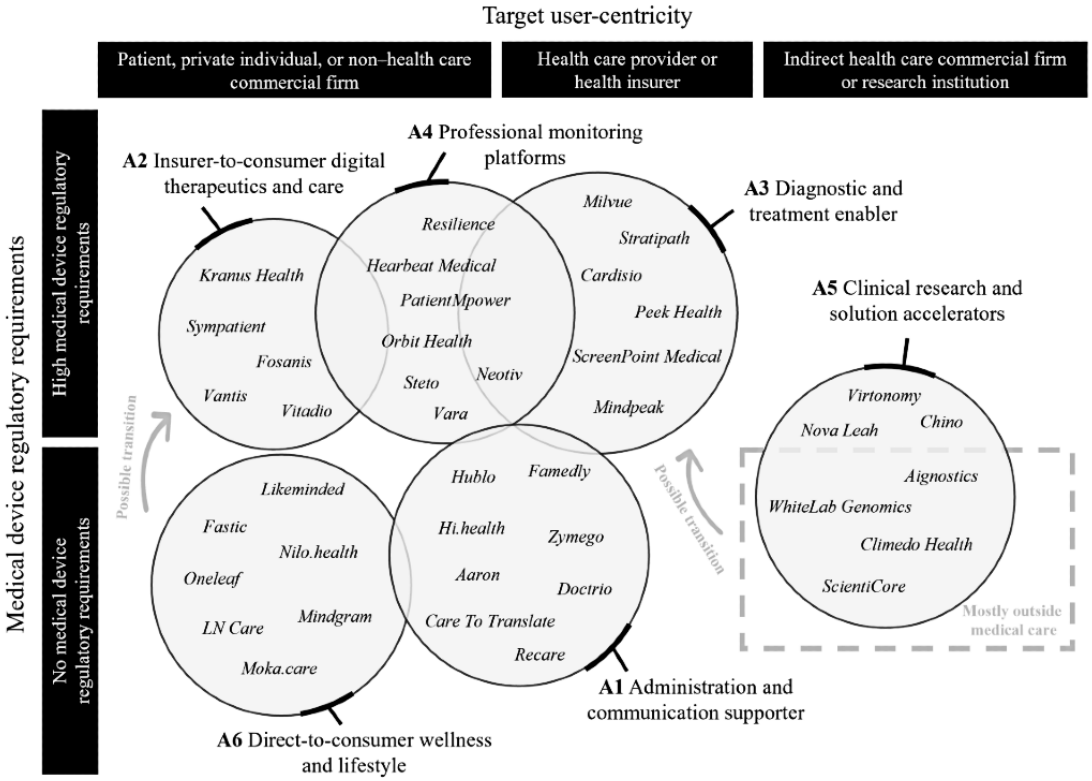

Comparison With Prior Work

We developed a comprehensive taxonomy for business models based on 11 dimensions, following the V4 business model dimensions framework [13]. This taxonomy covers the broad space of DHT business models. Compared to prior studies in the field, which focused on specific subdomains of digital health, such as telemedicine [44] or maternal and baby segments in mHealth [45], our work confirms some of the previously identified dimensions, such as *target user*, *paying entity*, or *enabling technology*, as key resources. These dimensions are commonly used across various sectors to describe business models through taxonomies [44,45]. In addition to these dimensions, previous work has highlighted some taxonomy characteristics specific to the subdomains of digital health, such as medical devices under an mHealth category dimension [45], and overall purpose characteristics, such as diagnosis and therapy [44]. In our study, we describe the broad field of DHT business models. We have identified several new dimensions and characteristics that are highly relevant and specific to the field that had not been mentioned in previous business model taxonomy research. Notably, the *main purpose* dimension in our taxonomy enables the description of complex combinations of medical and nonmedical purposes related to the value proposition. In addition, the *medical device regulatory framework* has emerged as a critical dimension. This framework significantly impacts the entire business model, from value proposition to value architecture and value network. By focusing on the EU, we precisely delineated emerging business models shaped by medical device regulations and those that evolve around unregulated DHTs. Another dimension specific to DHT business models is the *evidence* dimension. In the context of DHTs, it is mandatory for medical devices to provide clinical evidence to enter the market. Interestingly, we found that even some nonregulated DHT business models provide evidence (in the form of nonclinical evidence) to facilitate reimbursement with health insurers and prove their value within the ecosystem. Developing reimbursement pathways specifically for digital health applications opens new revenue models. Consequently, the *reimbursement* dimension is also highly relevant for DHT business models. Finally, the health care *level of care* introduces an important consideration that has been previously overlooked: the necessity for business models to integrate into the existing health care system. An intriguing finding was that some business models focus on research applications within traditional research institutions and pharmaceutical companies. As these operate outside of medical care, their solutions do not need to comply with medical device regulations.

Our work is the first to delineate archetypes of DHT business models based on real-world observations in the EU while also considering medical device regulations. While some research has reported overarching areas of activities in digital health [4], typical value propositions in mHealth [29], or categorizations

XSL•FO
RenderX

of patient needs that are addressed through different solutions [14], our work creates a clear image of archetypal business models used for stand-alone software in health care. The archetypes we identified align with the findings of Gehde et al [4]. The 13 areas of activity mentioned in their work are congruent with our findings and are covered by our archetypes. In addition, we mention here the work of the Digital Therapeutic Alliance [82], which, in partnership with Health Advances classified DHTs based on a technological perspective, identifying 8 major categories of DHTs. While the aforementioned work does not focus on the business models behind these solutions, our work offers suitable business models for the DHTs identified in that study. The DHT business models are inherently complex and exhibit some overlap. However, our archetypes effectively balance the need to simplify these complexities while still delivering valuable insights essential for designing business models in digital health.

Limitations

Our research is not without limitations. First, the DHT business model landscape is rapidly evolving, with new start-ups emerging, new technologies being developed, and regulatory changes occurring. Our study provides a snapshot of the current business models used in the field. However, regulatory environments and business models are subject to change. Furthermore, our focus on observable business model elements of DHTs to derive our archetypes necessitated the exclusion of dimensions such as key partnerships, pricing strategy, and funding. However, these are fundamental to DHT business models as described in the literature and discussed in the expert interviews. In this context, we developed an empirically driven taxonomy to study business model archetypes. We argue that a taxonomy has limited capabilities to describe complex stakeholder and partnership configurations. The same applies to funding, which has a highly relevant temporal component. Finally, the taxonomy focuses on digital business models in the EU, which means we do not describe business models of hardware medical device manufacturers and have not incorporated other business model aspects that might be relevant in other continents.

Contribution and Further Research

Theoretically, our work contributes to the systematic understanding of business models in the emerging DHT domain. By delineating common dimensions and characteristics of the field, we create a common language that paves the way for further research and helps scholars, such as those in business and management, health sciences, and information systems, to position their work within this landscape. The archetypes provide high-level descriptions of business models in this domain, offering a foundational framework for future studies. Furthermore, our dataset, which relies primarily on publicly available information from DHT companies, can serve as a basis for additional research.

On the practical side, the taxonomy we developed can assist entrepreneurs and start-ups in understanding key dimensions and characteristics relevant to DHT business models. An early informed understanding of the intricacies of a sustainable business model is fundamental for later success. In addition, the archetypes and strategic interpretation clarify the business models used in this domain, aiding entrepreneurs in identifying viable strategies based on common patterns. The data on the companies compiled in this work can serve as a starting point for competitor analysis or inspiration for developing their own business models. In addition to start-ups, established medical technology and pharmaceutical companies can benefit from a systematic understanding as they continually explore business model innovations to bring new solutions to the market. Investors, venture capitalists, and health insurers can also leverage this work to identify and validate viable business models. Finally, policy makers can use these insights when shaping new regulations.

Further research could delve deeper into the implications of funding and partnerships for the business model. In addition, the research could investigate business model configurations in other continents or even specific countries because reimbursement policies are specific to each country. Finally, further research could focus on business models of software and hardware combinations, such as software in a medical device.

Conclusions

Digital health is a rapidly evolving field that is attracting increasing attention. In our work, we provide the first overview of business models for DHTs with a focus on the EU. We systematically describe business models in the field and elaborate on the influence of medical device regulations on the entire business model. Furthermore, we created a business model taxonomy based on a systematic literature review and real-world cases, which we evaluated qualitatively. In addition, we identified 6 archetypes of DHT business models through cluster analysis, describing key strategies and differences for each cluster. This allowed us to identify drivers for business model innovation in the DHT sector. These contributions offer a systematic understanding and common language to facilitate analyses and comprehension in this domain. Our findings could assist decision makers, especially start-ups, in positioning themselves and building sustainable business models in this evolving field.

Acknowledgments

The authors acknowledge the support of the KIT-Publication Fund of the Karlsruhe Institute of Technology. The authors used the generative AI tool ChatGPT with GPT-4 [83] for spelling and grammar corrections.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[PDF File (Adobe PDF File), 867 KB - jmir_v27i1e65725_app1.pdf]

References

1. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020 Jul 10;3:94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
2. van Kessel R, Srivastava D, Kyriopoulos I, Monti G, Novillo-Ortiz D, Milman R, et al. Digital health reimbursement strategies of 8 European countries and Israel: scoping review and policy mapping. JMIR Mhealth Uhealth 2023 Sep 29;11:e49003 [FREE Full text] [doi: 10.2196/49003] [Medline: 37773610]
3. Wulfovich S, Meyers AD. Digital Health Entrepreneurship. Cham, Switzerland: Springer; 2019.
4. Gehde KM, Rausch F, Leker J. Business model configurations in digital healthcare—a German case study about digital transformation. Int J Innov Manag 2022;26(03):2240018. [doi: 10.1142/S1363919622400187]
5. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
6. Tariq MU. Healthcare innovation & entrepreneurship, digital health entrepreneurship. In: Raimi L, Oreagba IA, editors. Medical Entrepreneurship. Singapore: Springer; Jan 07, 2023.
7. ISO/TR 11147:2023: health informatics — personalized digital health — digital therapeutics health software systems. International Organization for Standardization. 2023. URL: https://www.iso.org/standard/83767.html [accessed 2024-01-02]
8. Baines R, Hoogendoorn P, Stevens S, Chatterjee A, Ashall-Payne L, Andrews T, et al. Navigating medical device certification: a qualitative exploration of barriers and enablers amongst innovators, notified bodies and other stakeholders. Ther Innov Regul Sci 2023 Mar;57(2):238-250 [FREE Full text] [doi: 10.1007/s43441-022-00463-4] [Medline: 36194368]
9. Baumgartner C, Harer J, Schröttner J. Medical Devices and In Vitro Diagnostics: Requirements in Europe. Cham, Switzerland: Springer; 2023.
10. Nilsen ER, Stendal K, Gullslett MK. Implementation of eHealth technology in community health care: the complexity of stakeholder involvement. BMC Health Serv Res 2020 May 11;20(1):395 [FREE Full text] [doi: 10.1186/s12913-020-05287-2] [Medline: 32393265]
11. Chakraborty I, Ilavarasan PV, Edirippulige S. Critical success factors of startups in the e-health domain. Health Policy Technol 2023 Sep;12(3):100773. [doi: 10.1016/j.hlpt.2023.100773]
12. Vannieuwenborg F, Van der Auwermeulen T, Van Ooteghem J, Jacobs A, Verbugge S, Colle D. Bringing eCare platforms to the market. Inform Health Soc Care 2017 Sep;42(3):207-231. [doi: 10.1080/17538157.2016.1200052] [Medline: 27564236]
13. Al-Debei MM, Avison D. Developing a unified framework of the business model concept. Eur J Inf Syst 2017 Dec 19;19(3):359-376. [doi: 10.1057/ejis.2010.21]
14. Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H. Digital transformation and disruption of the health care sector: internet-based observational study. J Med Internet Res 2018 Mar 27;20(3):e104 [FREE Full text] [doi: 10.2196/jmir.9498] [Medline: 29588274]
15. Hwang J, Christensen CM. Disruptive innovation in health care delivery: a framework for business-model innovation. Health Aff (Millwood) 2008;27(5):1329-1335. [doi: 10.1377/hlthaff.27.5.1329] [Medline: 18780919]
16. Schoormann T, Schweihoff J, Jussen I, Möller F. Classification tools for business models: status quo, comparison, and agenda. Electron Mark 2023 Apr 20;33:7. [doi: 10.1007/s12525-023-00639-2]
17. Weimar SN, Martjan RS, Terzidis O. Conceptualizing the landscape of digital health entrepreneurship: a systematic review and research agenda. Manag Rev Q 2024 Apr 15. [doi: 10.1007/s11301-024-00417-0]
18. Recommendations on digital interventions for health system strengthening – executive summary. World Health Organization. 2019 Apr 16. URL: https://www.who.int/publications/i/item/WHO-RHR-19.8 [accessed 2024-01-02]
19. Otto L, Harst L, Timpel P, Wollschlaeger B, Richter P, Schlieter H. Defining and delimitating telemedicine and related terms - an ontology-based classification. Stud Health Technol Inform 2020 Mar 02;268:113-122. [doi: 10.3233/SHTI200010] [Medline: 32141883]
20. Herold F, Theobald P, Gronwald T, Rapp MA, Müller NG. Going digital - a commentary on the terminology used at the intersection of physical activity and digital health. Eur Rev Aging Phys Act 2022 Jul 16;19(1):17 [FREE Full text] [doi: 10.1186/s11556-022-00296-y] [Medline: 35840899]
21. Software as a medical device: possible framework for risk categorization and corresponding considerations. International Medical Device Regulators Forum. 2014 Jul 18. URL: https://www.imdrf.org/documents/software-medical-device-possible-framework-risk-categorization-and-corresponding-considerations [accessed 2024-01-02]

XSL•FO
RenderX

22. Teece DJ. Business models, business strategy and innovation. Long Range Plan 2010;43(2-3):172-194. [doi: 10.1016/j.lrp.2009.07.003]
23. Osterwalder A. The business model ontology: a proposition in a design science approach. University of Lausanne. 2004. URL: https://www.researchgate.net/publication/33681401_The_Business_Model_Ontology_-_A_Proposition_in_a_Design_Science_Approach [accessed 2024-03-03]
24. Casadesus-Masanell R, Ricart JE. From strategy to business models and onto tactics. Long Range Plan 2010;43(2-3):195-215. [doi: 10.1016/j.lrp.2010.01.004]
25. Osterwalder A, Pigneur Y, Clark T. Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Hoboken, NJ: John Wiley & Sons; 2010.
26. Böhler D. Digital health business models: transformation is on the horizon. In: Friebe M, editor. Novel Innovation Design for the Future of Health. Cham, Switzerland: Springer; Nov 27, 2022.
27. Velayati F, Ayatollahi H, Hemmat M, Dehghan R. Telehealth business models and their components: systematic review. J Med Internet Res 2022 Mar 29;24(3):e33128 [FREE Full text] [doi: 10.2196/33128] [Medline: 35348471]
28. Pascarelli C, Colucci C, Mitrano G, Corallo A. Business models in digital health: bibliometric analysis and systematic literature review. In: Proceedings of the IEEE Symposium on Computers and Communications. 2023 Presented at: ISCC 2023; July 9-12, 2023; Gammarth, Tunisia. [doi: 10.1109/iscc58397.2023.10218237]
29. Gorski I, Bram JT, Sutermaster S, Eckman M, Mehta K. Value propositions of mHealth projects. J Med Eng Technol 2016;40(7-8):400-421. [doi: 10.1080/03091902.2016.1213907] [Medline: 27687907]
30. Sterling R, LeRouge C. On-demand telemedicine as a disruptive health technology: qualitative study exploring emerging business models and strategies among early adopter organizations in the United States. J Med Internet Res 2019 Nov 15;21(11):e14304 [FREE Full text] [doi: 10.2196/14304] [Medline: 31730038]
31. Lin SH, Liu JH, Wei J, Yin WH, Chen HH, Chiu WT. A business model analysis of telecardiology service. Telemed J E Health 2010 Dec;16(10):1067-1073. [doi: 10.1089/tmj.2010.0059] [Medline: 21186990]
32. Chen S, Cheng A, Mehta K. A review of telemedicine business models. Telemed J E Health 2013 Apr;19(4):287-297. [doi: 10.1089/tmj.2012.0172] [Medline: 23540278]
33. Korsgaard F, Hasenkam JM, Vesterby M. Successful implementation of telemedicine depends on personal relations between company representatives and healthcare providers: a qualitative study of business models for Danish home telemonitoring. Health Serv Manage Res 2021 Nov;34(4):223-233. [doi: 10.1177/0951484820988628] [Medline: 33508963]
34. Alnahdi S. An psychological mindset based stress free business model for digital health? J ReAttach Ther Dev Divers 2023;6(4s):401-412.
35. Acheampong F, Vimarlund V. Business models for telemedicine services: a literature review. Health Syst 2017 Dec 19;4(3):189-203. [doi: 10.1057/hs.2014.20]
36. Oderanti FO, Li F, Cubric M, Shi X. Business models for sustainable commercialisation of digital healthcare (eHealth) innovations for an increasingly ageing population. Technol Forecast Soc Change 2021 Oct;171:120969. [doi: 10.1016/j.techfore.2021.120969]
37. Wand Y, Monarchi DE, Parsons J, Woo CC. Theoretical foundations for conceptual modelling in information systems development. Decis Support Syst 1995 Dec;15(4):285-304. [doi: 10.1016/0167-9236(94)00043-6]
38. Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. Eur J Inf Syst 2017 Dec 19;22(3):336-359. [doi: 10.1057/ejis.2012.26]
39. Tulu B, Chatterjee S, Laxminarayan S. A taxonomy of telemedicine efforts with respect to applications, infrastructure, delivery tools, type of setting and purpose. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences. 2005 Presented at: HICSS 2005; January 6, 2005; Big Island, HI. [doi: 10.1109/hicss.2005.56]
40. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011;17(6):484-494 [FREE Full text] [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
41. Glöggler M, Ammenwerth E. Improvement and evaluation of the TOPCOP taxonomy of patient portals: Taxonomy-Evaluation-Delphi (TED) approach. J Med Internet Res 2021 Oct 05;23(10):e30701 [FREE Full text] [doi: 10.2196/30701] [Medline: 34403354]
42. Harst L, Otto L, Timpel P, Richter P, Lantzsch H, Wollschlaeger B, et al. An empirically sound telemedicine taxonomy – applying the CAFE methodology. J Public Health (Berl.) 2021 May 28;30(11):2729-2740. [doi: 10.1007/s10389-021-01558-2]
43. Möller F, Stachon M, Azkan C, Schoormann T, Otto B. Designing business model taxonomies – synthesis and guidance from information systems research. Electron Markets 2021 Dec 29;32(2):701-726. [doi: 10.1007/s12525-021-00507-x]
44. Peters C, Blohm I, Leimeister JM. Anatomy of successful business models for complex services: insights from the telemedicine field. J Manag Inf Syst 2015 Dec 16;32(3):75-104. [doi: 10.1080/07421222.2015.1095034]
45. Ropposch C, Rachinger M. mHealth business model types in maternal and baby care – a service-dominant logic view. In: Proceedings of the Annual Meeting of the Academy of Management 2023: Putting the Worker Front and Center. 2023 Presented at: TLC@AOM 2023; August 4-8, 2023; Boston, MA. [doi: 10.5465/AMPROC.2023.17882abstract]
46. Velayati F, Ayatollahi H, Hemmat M, Dehghan R. The 4P telehealth business framework for Iran. BMC Med Inform Decis Mak 2022 Oct 09;22(1):266 [FREE Full text] [doi: 10.1186/s12911-022-02011-4] [Medline: 36210440]

XSL•FO
RenderX

47. Thiebes S, Toussaint PA, Ju J, Ahn JH, Lyytinen K, Sunyaev A. Valuable genomes: taxonomy and archetypes of business models in direct-to-consumer genetic testing. J Med Internet Res 2020 Jan 21;22(1):e14890 [FREE Full text] [doi: 10.2196/14890] [Medline: 31961329]
48. Hunke F, Heinz D, Satzger G. Creating customer value from data: foundations and archetypes of analytics-based services. Electron Mark 2022 Oct 27;32(2):503-521 [FREE Full text] [doi: 10.1007/s12525-021-00506-y] [Medline: 35602114]
49. Sterk F, Stocker A, Heinz D, Weinhardt C. Unlocking the value from car data: a taxonomy and archetypes of connected car business models. Electron Markets 2024 Feb 13;34:13. [doi: 10.1007/s12525-024-00692-5]
50. Kundisch D, Muntermann J, Oberländer AM, Rau D, Röglinger M, Schoormann T, et al. An update for taxonomy designers: methodological guidance from information systems research. Bus Inf Syst Eng 2021 Oct 22;64(4):421-439. [doi: 10.1007/s12599-021-00723-x]
51. Kraus S, Breier M, Dasí-Rodríguez S. The art of crafting a systematic literature review in entrepreneurship research. Int Entrep Manag J 2020 Feb 01;16(3):1023-1042. [doi: 10.1007/s11365-020-00635-4]
52. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002;26(2):xiii-xxiii.
53. Oderanti F, Li F. Commercialization of eHealth innovations in the market of the UK healthcare sector: a framework for a sustainable business model. Psychol Mark 2018 Jan 10;35(2):120-137. [doi: 10.1002/mar.21074]
54. Kijl B, Nieuwenhuis LJ. Deploying e-health service innovations – an early stage business model engineering and regulatory validation approach. Int J Healthc Technol Manag 2011;12(1):23. [doi: 10.1504/IJHTM.2011.037219]
55. Kijl B, Nieuwenhuis LJ, Huis in 't Veld RM, Hermens HJ, Vollenbroek-Hutten MM. Deployment of e-health services - a business model engineering strategy. J Telemed Telecare 2010;16(6):344-353. [doi: 10.1258/jtt.2010.006009] [Medline: 20798429]
56. Garrot T, Angelé-Halgand N. Digital health business models: reconciling individual focus and equity? In: Menvielle L, Audrain-Pontevia AF, Menvielle W, editors. The Digitization of Healthcare First Edition. London, UK: Palgrave Macmillan; 2022.
57. Mueller C. mHealth business model framework for the maternal and baby segment: a design science research approach. In: Proceedings of the 32nd Bled eConference Humanizing Technology for a Sustainable Society. 2019 Presented at: BLED 2019; June 16-19, 2019; Bled, Slovenia.
58. Christie HL, Boots LM, Hermans I, Govers M, Tange HJ, Verhey FR, et al. Business models of eHealth interventions to support informal caregivers of people with dementia in the Netherlands: analysis of case studies. JMIR Aging 2021 Jun 3;4(2):e24724. [doi: 10.2196/24724]
59. Olla P, Shimskey C. mHealth taxonomy: a literature survey of mobile health applications. Health Technol 2015 Jan 30;4(4):299-308. [doi: 10.1007/s12553-014-0093-8]
60. Lehmann NJ, Spielmann F, George B, Ververs L, Karagulle MU, Kmiotek D, et al. mHealthAtlas - an approach for the multidisciplinary evaluation of mHealth applications. In: Proceedings of the 2020 IEEE International Conference on E-health Networking, Application & Services. 2020 Presented at: HEALTHCOM 2020; March 1-2, 2020; Shenzhen, China.
61. Greve M, Lembcke TB, Diederich S, Brendel AB, Kolbe LM. Healthy by app – towards a taxonomy of mobile health applications. In: Proceedings of the 2020 PACIS. 2020 Presented at: PACIS 2020; June 20-24, 2020; Dubai, United Arab Emirates.
62. HealthTech 250: the most promising early-stage digital health ventures in 2022. Galen Growth. URL: https://www.galengrowth.com/healthtech-250/ [accessed 2024-01-02]
63. What is Crunchbase rank and trend score? Crunchbase. URL: https://about.crunchbase.com/blog/crunchbase-rank-trend-score/ [accessed 2024-01-02]
64. Deutscher startup monitor 2023. PwC Deutschland. 2023. URL: https://startupverband.de/fileadmin/startupverband/mediaarchiv/research/dsm/dsm_2023.pdf [accessed 2024-03-03]
65. O'Regan N, Ghobadian A. Testing the homogeneity of SMEs: the impact of size on managerial and organisational processes. Eur Bus Rev 2004;16(1):64-77. [doi: 10.1108/09555340410512411]
66. Bryman A. Social Research Methods Fourth Edition. Oxford, UK: Oxford University Press; 2014.
67. Szopinski D, Schoormann T, Kundisch D. Because your taxonomy is worth it: towards a framework for taxonomy evaluation. In: Proceedings of the Twenty-Seventh European Conference on Information Systems. 2019 Presented at: ECIS 2019; June 8-14, 2019; Stockholm-Uppsala, Sweden.
68. Rowley J. Conducting research interviews. Manag Res Rev 2012;35(3/4):260-271. [doi: 10.1108/01409171211210154]
69. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research: notes on the Gioia methodology. Organ Res Methods 2012 Jul 24;16(1):15-31. [doi: 10.1177/1094428112452151]
70. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
71. Crum M, Nelson T, de Borst J, Byrnes P. The use of cluster analysis in entrepreneurship research: review of past research and future directions. J Small Bus Manag 2020 Aug 10;60(4):961-1000. [doi: 10.1080/00472778.2020.1748475]
72. Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. J Mark Res 1983 May;20(2):134-148. [doi: 10.2307/3151680]

XSL•FO
RenderX

73. Ward JHJ. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963 Mar;58(301):236-244. [doi: 10.1080/01621459.1963.10500845]
74. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ: John Wiley & Sons; 1990.
75. Hambrick DC. Taxonomic approaches to studying strategy: some conceptual and methodological issues. J Manage 1984 Apr 01;10(1):27-41. [doi: 10.1177/014920638401000104]
76. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
77. Rauch A, Negrini S, Cieza A. Toward strengthening rehabilitation in health systems: methods used to develop a WHO package of rehabilitation interventions. Arch Phys Med Rehabil 2019 Nov;100(11):2205-2211. [doi: 10.1016/j.apmr.2019.06.002] [Medline: 31207218]
78. Shi L. The impact of primary care: a focused review. Scientifica (Cairo) 2012;2012:432892 [FREE Full text] [doi: 10.6064/2012/432892] [Medline: 24278694]
79. Gray M. Population healthcare: the third dimension. J R Soc Med 2017 Feb 01;110(2):54-56 [FREE Full text] [doi: 10.1177/0141076817691024] [Medline: 28169584]
80. Hein A, Schreieck M, Riasanow T, Setzke DS, Wiesche M, Böhm M, et al. Digital platform ecosystems. Electron Markets 2019 Nov 12;30(1):87-98. [doi: 10.1007/s12525-019-00377-4]
81. Simon DA, Shachar C, Cohen IG. Skating the line between general wellness products and regulated devices: strategies and implications. J Law Biosci 2022;9(2):lsac015 [FREE Full text] [doi: 10.1093/jlb/lsac015] [Medline: 35855400]
82. Guidance to industry: classification of digital health technologies. Digital Therapeutic Alliance. 2023. URL: https://dtxalliance.org/wp-content/uploads/2023/06/Guidance-to-Industry-Classification-of-Digital-Health-Technologies-2023Jun05.pdf [accessed 2024-03-03]
83. GPT-4. OpenAI. URL: https://openai.com/index/gpt-4/ [accessed 2024-12-26]

Abbreviations

DHT: digital health technology
EU: European Union
IVDR: Regulation (EU) 2017/746 on in vitro diagnostic medical devices
MDR: Regulation (EU) 2017/745 on medical devices
mHealth: mobile health
RQ: research question

Edited by A Mavragani; submitted 23.08.24; peer-reviewed by Z Zandesh, M Li; comments to author 20.10.24; revised version received 27.10.24; accepted 02.12.24; published 09.01.25.

Please cite as:
Weimar SN, Martjan RS, Terzidis O
Business Venturing in Regulated Markets—Taxonomy and Archetypes of Digital Health Business Models in the European Union: Mixed Methods Descriptive and Exploratory Study
J Med Internet Res 2025;27:e65725
URL: https://www.jmir.org/2025/1/e65725
doi:10.2196/65725
PMID:

©Sascha Noel Weimar, Rahel Sophie Martjan, Orestis Terzidis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation and Comparison of the Academic Quality of Open-Access Mega Journals and Authoritative Journals: Disruptive Innovation Evaluation

Yuyan Jiang[1], MPhil; Xue-li Liu[1,2], BM; Liyun Wang[1], MPhil

[1]Henan Research Center for Science Journals, Xinxiang Medical University, Xinxiang, China

[2]Faculty of Humanities & Social Sciences, Xinxiang Medical University, Xinxiang, China

Corresponding Author:
Xue-li Liu, BM
Faculty of Humanities & Social Sciences
Xinxiang Medical University
Library and Information Building, 2nd floor
No. 601 Jinsui Avenue, Hongqi District
Xinxiang,
China
Phone: 86 1 383 736 0965
Email: liueditor03@163.com

Abstract

Background: Some scholars who are skeptical about open-access mega journals (OAMJs) have argued that low-quality papers are often difficult to publish in more prestigious and authoritative journals, and OAMJs may be their main destination.

Objective: This study aims to evaluate the academic quality of OAMJs and highlight their important role in clinical medicine. To achieve this aim, authoritative journals and representative OAMJs in this field were selected as research objects. The differences between the two were compared and analyzed in terms of their level of disruptive innovation. Additionally, this paper explored the countries and research directions for which OAMJs serve as publication channels for disruptive innovations.

Methods: In this study, the journal information, literature data, and open citation relationship data were sourced from Journal Citation Reports (JCR), Web of Science (WoS), InCites, and the OpenCitations Index of PubMed Open PMID-to-PMID citations (POCI). Then, we calculated the disruptive innovation level of the focus paper based on the local POCI database.

Results: The mean Journal Disruption Index (JDI) values for the selected authoritative journals and OAMJs were 0.5866 (SD 0.26933) and 0.0255 (SD 0.01689), respectively, showing a significant difference. Only 1.48% (861/58,181) of the OAMJ papers reached the median level of disruptive innovation of authoritative journal papers (MD_{AJ}). However, the absolute number was roughly equal to that of authoritative journals. OAMJs surpassed authoritative journals in publishing innovative papers in 24 research directions (eg, Allergy), accounting for 40.68% of all research directions in clinical medicine. Among research topics with at least 10 authoritative papers, OAMJs matched or exceeded MD_{AJ} in 35.71% of cases. The number of papers published in authoritative journals and the average level of disruptive innovation in each country showed a linear relationship after logarithmic treatment, with a correlation coefficient of –0.891 (P<.001). However, the number of papers published in OAMJs in each country and the average level of disruptive innovation did not show a linear relationship after logarithmic treatment.

Conclusions: While the average disruptive innovation level of papers published by OAMJs is significantly lower than that of authoritative journals, OAMJs have become an important publication channel for innovative research in various research directions. They also provide fairer opportunities for the publication of innovative results from limited-income countries. Therefore, the academic community should recognize the contribution and value of OAMJs to advancing scientific research.

(J Med Internet Res 2025;27:e59598) doi:10.2196/59598

KEYWORDS

innovative evaluation; disruption index; open-access mega journals; paper evaluation; open citation data

XSL•FO
RenderX

Introduction

Henry Oldenburg, the founding editor of the journal Philosophical Transactions of the Royal Society, noted that the purpose of scholarly journals is to provide a forum for researchers to transfer knowledge to each other and contribute to perfecting the grand design of all the philosophical arts and sciences. However, due to technical limitations, printed journals have been the only form of journal publication since the first journal was introduced in the 17th century. Therefore, researchers must usually wait for 10 to 12 months or more before publishing a paper [1]. The appearance of web-based publications has solved the limitations of page size and publishing cycles. Readers can access the latest research more quickly through online publishing and distribution, which provides a more flexible and efficient way for knowledge dissemination and sharing. This transformation significantly enhanced content deliverability, but accessibility has not kept pace. The urgent need to address this issue has fueled the rise of the open-access (OA) movement and the development of OA journals. Simultaneously, the rapid development of science and technology, combined with the pressure of "publish or perish," has created an increasing demand for efficient, timely, and cost-effective publication methods. Furthermore, the rapid increase in the number of publications has imposed a growing financial burden on academic institutions and researchers seeking access to scholarly literature. Providing a platform that can accommodate the rapid publication and huge volume of emerging research while facilitating academic exchange and reducing the economic strain of accessing academic resources has become a widespread concern for the scientific community.

Open-access mega journals (OAMJs) may currently be one of the most effective solutions, offering faster publishing speeds, higher acceptance rates [2], and support for the initial exploration of new fields [3]. These advantages align well with the needs of authors [4]. Additionally, many OAMJs provide readers with value-added functions that most traditional journals still lack [5].

Since the establishment of the Public Library of Science (PLoS) in 2001, numerous traditional academic publishers have launched their own OAMJs to expand their product lines [6]. This has positioned OAMJs as increasingly important players in facilitating academic exchanges [7]. However, the evaluation criteria of most OAMJs only pay attention to *research rationality* [8]. Coupled with the financial incentives of OA publishers to publish widely, concerns have arisen about the quality of research published in OAMJs, which have a higher publishing error rate [9] and possibility of predatory publishing [10].

Most OAMJ papers have received a significant number of citations [11], which often exceed the average citation level [12]. However, the evaluation processes of authoritative journals cannot guarantee the accuracy of the research [13]. Despite this, prejudice still exists among some scholars [14], and the complex attitudes within different academic groups have not significantly improved [15].

It is unfair to view OAMJs solely as a negative development. As Professor John Ioannidis of Stanford University [16] points out, several features of OAMJs align with desirable scientific practices. For instance, OAMJs enhance the visibility of academic papers and their broader academic impact. Furthermore, OAMJs reduce selective publication biases by accommodating results that are considered undesirable in traditional professional journals, thus fostering the diversity of perspectives and challenging orthodoxy.

Since 1972, the mainstream evaluation system has relied on citation metrics [17], which means indicators such as the Journal Impact Factor (JIF) have been widely used in academic review, promotion, and tenure evaluation. However, it is important to note that citation metrics primarily reflect the impact of journals rather than their academic innovativeness. Related research has shown that relying on these impact indicators introduces biases against innovation [18], often resulting in flawed decisions in research assessment and evaluation [19].

Disruptive innovation is defined as the process by which a new product, service, or technology transforms an existing market by creating a new market or significantly altering an existing one. This is one of the core driving forces for developing productivity. Therefore, conducting a scientific and reasonable evaluation of the innovation levels in OAMJ papers is crucial for determining their value. Given that clinical medicine is the primary focus of OAMJ papers [5], this paper selects the authoritative journals and the representative OAMJs related to this field as its research subjects. This study compares and analyzes the differences in disruptive innovation levels between these 2 groups and examines the countries and research directions for which OAMJs serve as publication channels for disruptive innovations.

Methods

Research Objects

The objects of this study include 2 groups of journals, representing authoritative journals in the field of clinical medicine and OAMJs. The selection criteria for authoritative journals were as follows: (1) JIF ranks high in the field; (2) selected journals are not review journals; and (3) academic authority is widely recognized. The selection criteria for representative OAMJ journals were as follows: (1) OA peer-reviewed journals that charge publication processing fees; (2) an annual average of more than 2000 OA research papers published on clinical medicine topics; (3) a 3-year impact factor within Q1 or Q2.

Through a search of relevant journals in the field of clinical medicine, we discovered that the impact factors of the British Medical Journal (BMJ), the New England Journal of Medicine, The Lancet, and the Journal of the American Medical Association (JAMA) ranked among the highest in impact factors between 2017 and 2019, with widely recognized academic authority [20-25]. We also found that BMJ Open, PLoS ONE, and Scientific Reports published more than 2000 open-access research papers on clinical medicine topics annually, establishing a certain degree of academic influence.

XSL•FO
RenderX

Based on these criteria, this study selected BMJ, New England Journal of Medicine, The Lancet, and JAMA as the authoritative journals and BMJ Open, PLoS ONE, and Scientific Reports as the representative OAMJs.

Because there is no disruptive innovation in review articles, and because Bornmann et al [26] suggest that the disruption index should only be calculated for research papers with at least 10 references, this study exclusively focused on clinical medical research papers with more than 10 references from the aforementioned journals.

Data Sources

The data required for this study included journal information, literature information, and citation relationship information, obtained through Journal Citation Reports (JCR), Web of Science (WoS), InCites, and the OpenCitations Index of PubMed open PMID-to-PMID citations (POCI), respectively. POCI, a data set sourced from the National Institutes of Health, comprises over 700 million citations and 29 million bibliographic resources. In POCI, citations are regarded as first-class data entities with 7 attribute fields, including the citation time span. To evaluate disruptive innovation, we selected focus papers and used their PMIDs along with POCI data identifiers to calculate the disruptive innovation level for each paper.

Data Acquisition and Processing

First, we logged into JCR, WoS, and InCites to retrieve the journal information and literature information of the selected journals and save them. Then, we logged into opencitations.net, entered the download page, selected the POCI data set, and downloaded the dump data in CSV format via the Figshare platform. Next, we used Navicat software (PremiumSoft CyberTech Ltd) to import all data into the local SQLite database and process the downloaded POCI data. Finally, in the local database based on the POCI data transformation, we used the PMID numbers extracted from the full records of the focus literature to identify the citation relationship associated with the focus papers. We established relevant data tables for the subsequent calculations.

The disruption index calculation was highly dependent on the time window chosen. The findings of Liu et al [27] confirm that the stabilization time window for the disruption index varies across different disciplines. We believe this is partly due to the different stages of development in various disciplines. Bornmann and Tekles [28] suggest that a time window of 3 years or more is an important prerequisite for ensuring the validity of the calculation results. Given the complexity of the disruption index calculation, the large volume of data [29], and our lack of access to extensive commercial citation data resources, we used the open citation data provided by OpenCitations for the calculation. To guarantee the accuracy of the results of the paper's innovation measurement while minimizing the computational workload, the citation time window in this study was set to 3 years.

Evaluation Indicators

Since Christensen [30] of Harvard Business School posited the theory of disruptive innovation in 1996, it has become an important paradigm in the field of innovation research. Based on this theory, Wu et al [31] put forward the disruption index (often referred to as the D Index for short), which can be used to measure disruption by calculating the citation substitution of focus papers in the citation network. Based on the disruption index, Bornmann et al [26] discussed the effectiveness of the convergence of this index and the variants that may improve the measurement effect. Ruan et al [32] analyzed the application limitations of the disruption index for measuring progress in the fields of science and technology.

Considering the process of metric calculation, F-type citations reflect that the focus paper disrupts the reference, B-type citations indicate that the focus paper is a development of the reference, and R-type citations reveal that the focus paper inherits the reference. Liu et al [33] argued that the disruption Index is a “relative” concept because it measures disruption based on the relative size of N_F (F-type citations) and N_B (B-type citations). They defined the disruption index, which is only reflected by F-type citations, as the “absolute disruption index.” Additionally, Liu et al [27,34] empirically studied the stabilization time window for the disruption index in different subject areas and resolved mathematical inconsistencies in its traditional formation (eg, N_R not satisfying monotonicity).

Based on this disruption index, Bornmann et al [35], Horen et al [36], Sullivan et al [37], Meyer et al [38], Jiang and Liu [39], and others have mined disruptive papers in the fields of scientometrics, craniofacial surgery, pediatric surgery, synthetic biology, and energy security respectively. On this basis, Jiang and Liu [40] proposed a method for calculating the Disruption index using open citation data during previous research. They carried out evaluations at both the literature and journal levels [41-45], contributing to a mature framework for evaluating disruptive innovation.



In equations (1) to (3), N_F refers to the literature that only cites the focus paper (FP); N_B refers to the literature that cites both the focus paper and at least one reference (R) of the focus paper; N_R refers to literature that only cites at least one reference (R) of the focus paper but not the focus paper; n is the number of articles contained in the journal; and D_{Zi} is the absolute disruption index (D_Z) of the ith article in the journal.

Results

Comparative Analysis of the Journals’ Disruptive Innovation Levels

Annual changes in JIF, JDI, and the number of published articles for the 7 journals selected in this study between 2017 and 2019 are shown in Table 1. The key findings are as follows. First, the average JDI of the selected authoritative journals is 0.5866, while the average JDI of OAMJs is 0.0255, indicating a significant difference. Second, the JDI values of the authoritative journals are higher than those of OAMJs. Third, JDI does not necessarily increase with a higher JIF.

XSL•FO
RenderX

Additionally, to better measure the difference in disruptive innovation level between OAMJ papers and authoritative journal papers, this study defined 2 parameters: the median D_Z of authoritative journals (MD_{AJ}) and the median D_Z of OAMJs (MD_{OAMJ}), which were 0.289 and 0.0019, respectively. The findings indicated that (1) the disruptive innovation level (94.60 %) of authoritative journal papers was higher than MD_{OAMJ} and (2) only 1.48% (861/58151) of OAMJ papers had a disruptive innovation level higher than MD_{AJ}. However, the number of OAMJ papers (n=861, 1.48%) was close to half of the number of authoritative journals (n=1128 of 2222).

Table 1. Annual changes in JIF[a], JDI[b], and number of articles for the 7 selected journals.

Journal	JIF			JDI			Number of articles				Number of COCI[c] articles			
	2017	2018	2019	2017	2018	2019	2017	2018	2019	All	2017	2018	2019	All
BMJ[d]	23.259	27.604	30.223	0.2691	0.1897	0.2580	140	147	140	427	123	126	124	373
NEJM[e]	31.398	36.216	38.637	0.5276	0.3811	0.4584	275	276	277	828	258	270	265	793
The Lancet	53.254	59.102	60.392	0.7459	0.7132	0.6682	242	191	192	625	233	183	187	603
JAMA[f]	79.258	70.67	74.699	0.9268	1.0325	0.8684	179	147	157	483	177	136	140	453
BMJ Open	2.413	2.376	2.496	0.0730	0.0211	0.0210	2160	2112	3512	7784	2055	2025	3345	7425
PLoS[g] ONE	2.766	2.776	2.740	0.0180	0.0211	0.0190	19,976	17,461	14,880	52,317	10,444	9153	7892	27,489
Scientific Reports	4.122	4.011	3.988	0.0161	0.0223	0.0180	24,806	17,118	19,843	61,767	9071	6543	7653	23,267

[a]JIF: Journal Impact Factor.

[b]JDI: Journal Disruption Index.

[c]COCI: OpenCitations Index of CrossRef Open DOI-to-DOI Citations

[d]BMJ: British Medical Journal.

[e]NEJM: New England Journal of Medicine.

[f]JAMA: Journal of the American Medical Association.

[g]PLoS: Public Library of Science.

Comparative Analysis of Disruptive Innovation Level Across Fields of Paper

Due to the complexity and diversity of clinical medical research, this study summarized statistics based on the classification results of InCites to assess the disruption of OAMJ papers and authoritative journal papers across various research fields. The findings, shown in Table 2, revealed several key insights.

There were differences in the average levels of disruptive innovation among OAMJ papers published in different research fields. The highest level was observed in the research field of immunology (0.1965), while neuroimaging had the lowest level (0.0049).

Similarly, there were differences in the average levels of disruptive innovation in authoritative journal papers, with reproductive biology showing the highest level (15.7902) and sport sciences (0.0031) showing the lowest level.

Across all fields, authoritative journal papers outperformed OAMJ papers in terms of disruptive innovation at the overall level. In 24 research directions such as allergy—accounting for 40.68% of all clinical medicine research areas—the number of OAMJ papers with disruptive innovation levels exceeding MD_{AJ} surpassed half the number of authoritative journal papers. Among the topics with no less than 10 authoritative papers published, the topics in which the number of papers published by OAMJ with a level of disruptive innovation exceeding MD_{AJ}'s was no less than half of the number of papers published in authoritative journals accounted for 35.71% of the total.

XSL•FO
RenderX

Table 2. Comparative analysis of disruptive innovation level across fields of paper.

Field	Authoritative journals				OAMJ[a]			
	Amount	Average (D_Z[b])	MID[c] (D_Z)	Amount ($Dz>MD_{OAMJ}$[d]), n (%)	Amount	Average (D_Z)	MID (D_Z)	Amount ($D_Z>MD_{AJ}$[e]), n (%)
Allergy	4	0.2268	0.0481	3 (75)	32	0.0264	0.0028	2 (6.25)
Andrology	0	—[f]	—	—	2	0.0408	0.0027	—
Anesthesiology	16	0.8493	0.4280	16 (100)	280	0.0247	0.0015	2 (0.71)
Audiology and speech-language pathology	0	—	—	—	63	0.0136	0	—
Behavioral sciences	0	—	—	—	256	0.0218	0.0017	—
Cardiac and cardiovascular systems	217	3.1848	0.2502	211 (97.24)	2510	0.0141	0.0015	15 (0.60)
Clinical neurology	146	3.2995	0.3190	138 (94.52)	1834	0.0149	0.0012	10 (0.55)
Critical care medicine	57	1.6419	0.2253	57 (100)	563	0.0164	0.0012	10 (1.78)
Dentistry, oral surgery and medicine	2	2.1744	1.2166	2 (100)	607	0.0288	0.0039	0 (0)
Dermatology	18	2.4297	0.4658	18 (100)	299	0.0400	0.0038	4 (1.34)
Emergency medicine	8	1.3246	0.1213	8 (100)	196	0.0508	0.0024	12 (6.12)
Endocrinology and metabolism	134	2.8454	0.2190	128 (95.52)	2583	0.0215	0.0024	45 (1.74)
Engineering, biomedical	0	—	—	—	467	0.0418	0.0052	—
Gastroenterology and hepatology	65	2.2936	0.3151	62 (95.38)	1769	0.0245	0.0025	20 (1.13)
Genetics and heredity	31	0.8752	0.1720	28 (90.32)	2247	0.0062	0.0002	11 (0.49)
Geriatrics and gerontology	10	1.1717	0.3157	10 (100)	321	0.0327	0.0018	5 (1.56)
Health care sciences and services	21	4.0306	0.2428	21 (100)	732	0.0538	0.0027	29 (3.96)
Health policy and services	5	4.3698	2.9845	4 (80)	92	0.0646	0.0054	0 (0)
Hematology	72	2.9323	0.7213	65 (90.28)	689	0.0158	0.0015	2 (0.29)
Immunology	44	2.6776	0.3863	40 (90.91)	3432	0.1965	0.0022	18 (0.52)
Infectious diseases	78	7.1637	0.3698	70 (89.74)	1969	0.0460	0.0043	49 (2.49)
Integrative and complementary medicine	0	—	—	—	25	0.0996	0.0043	—
Materials science, biomaterials	0	—	—	—	183	0.0716	0.0152	—
Medical ethics	0	—	—	—	1	0.0131	0.0131	—
Medical informatics	1	0.2245	0.2245	1 (100)	30	0.1121	0.0069	3 (10)
Medical laboratory technology	1	0.2120	0.2120	1 (100)	25	0.0115	0.0013	0 (0)
Medicine, general and internal	357	6.2389	0.3111	335 (93.84)	1415	0.1411	0.0028	68 (4.81)
Medicine, legal	0	—	—	—	57	0.0163	0.0013	—
Medicine, research and experimental	4	8.5306	0.1741	4 (100)	63	0.0638	0.0067	3 (4.76)
Neuroimaging	0	—	—	—	3	0.0049	0.0025	—
Neurosciences	9	1.0562	0.4969	9 (100)	7794	0.0094	0.0008	9 (0.12)
Nursing	0	—	—	—	159	0.0638	0.0047	—
Nutrition and dietetics	29	1.0717	0.1605	28 (96.55)	764	0.0300	0.0031	28 (3.66)
Obstetrics and gynecology	89	1.6846	0.1264	79 (88.76)	1019	0.0349	0.0028	54 (5.30)
Oncology	207	8.9447	1.3285	199 (96.14)	6032	0.0337	0.0025	13 (0.22)
Ophthalmology	19	4.7031	0.1073	18 (94.74)	1776	0.0248	0.0019	71 (4)
Orthopedics	22	0.7836	0.0998	22 (100)	658	0.0206	0.0021	21 (3.19)
Otorhinolaryngology	6	0.9314	0.0747	6 (100)	244	0.0249	0.0027	15 (6.15)

XSL•FO
RenderX

Field	Authoritative journals				OAMJ[a]			
	Amount	Average (D_Z[b])	MID[c] (D_Z)	Amount ($Dz>MD_{OAMJ}$[d]), n (%)	Amount	Average (D_Z)	MID (D_Z)	Amount ($D_Z>MD_{AJ}$[e]), n (%)
Pathology	2	0.0575	0	1 (50)	57	0.0139	0.0045	57 (100)
Pediatrics	66	0.6416	0.1569	63 (95.45)	667	0.0316	0.0020	26 (3.90)
Peripheral vascular disease	25	1.2093	0.1844	23 (92)	616	0.0175	0.0025	9 (1.46)
Pharmacology and pharmacy	4	1.5833	1.4107	4 (100)	1141	0.0612	0.0065	6 (0.53)
Primary health care	0	—	—	—	60	0.1677	0.0099	—
Psychiatry	50	0.2403	0.0509	43 (86)	1511	0.0378	0.0013	96 (6.35)
Psychology, clinical	3	0.4601	0.3300	3 (100)	162	0.0075	0.0005	0 (0)
Public, environmental, and occupational health	139	3.8125	0.3887	132 (94.96)	3861	0.0507	0.0044	92 (2.38)
Radiology, nuclear medicine, and medical imaging	8	1.4056	0.2714	7 (87.50)	1418	0.0303	0.0021	22 (1.55)
Rehabilitation	1	0.3139	0.3139	1 (100)	256	0.0340	0.0010	4 (1.56)
Reproductive biology	1	15.7902	15.7902	1 (100)	363	0.0159	0.0022	0 (0)
Respiratory system	76	1.4090	0.2322	73 (96.05)	934	0.0248	0.0019	14 (1.50)
Rheumatology	50	1.8720	0.1174	44 (88)	738	0.0191	0.0021	18 (2.44)
Sport sciences	2	0.0031	0	1 (50)	780	0.0290	0.0016	780 (100)
Substance abuse	11	9.7414	0.1482	10 (90.91)	191	0.0268	0.0040	7 (3.66)
Surgery	47	1.0035	0.3107	45 (95.74)	797	0.0310	0.0025	15 (1.88)
Toxicology	1	0.0081	0.0081	1 (100)	312	0.0308	0.0029	111 (35.58)
Transplantation	5	0.0464	0.0303	4 (80)	82	0.0134	0.0019	9 (10.98)
Tropical medicine	7	1.4192	0.6969	6 (85.71)	268	0.0311	0.0039	1 (0.37)
Urology and nephrology	56	2.7370	0.1476	52 (92.86)	1573	0.0175	0.0019	37 (2.35)
Virology	2	6.0473	1.6838	2 (100)	1212	0.0107	0.0006	0 (0)

[a]OAMJ: open-access mega journal.

[b]D_Z: absolute disruption index.

[c]MID: median.

[d]MD_{OAMJ}: median D_Z of the OAMJ papers.

[e]MD_{AJ}: median D_Z of the authoritative journal papers.

[f]Not applicable.

Comparative Analysis of Disruptive Innovation Level of Papers Across Countries

The disruptive innovation levels of OAMJ papers and authoritative journal papers published by different countries are shown in Tables 3 and 4, which include data for countries with records in both types of journals. To better describe the relevant results, this study used the following criteria. First, the number of authoritative journal papers is used as a measure of national scientific research strength. A higher number of published authoritative papers reflects stronger national scientific research capabilities. Second, the difference in the disruptive innovation levels between published OAMJ papers and authoritative journal papers serves as a measure of the recognition of OAMJs by scholars in different countries. A smaller difference indicates a higher degree of recognition. Third, the ratio of OAMJ papers to authoritative journal papers measures the degree of favoritism scholars from different countries show toward OAMJs. A larger ratio suggests a higher degree of favoritism toward OAMJs by scholars from a particular country.

From Tables 3 and 4, we observed several key insights. First, there was a difference between the average disruptive innovation level of OAMJ papers across countries, with South Korea ranking the highest (0.5476) and Serbia ranking the lowest (0.0043). Similarly, the average disruptive innovation level of authoritative journal papers also varied across countries, with Peru ranking the highest (35.6125) and Serbia ranking the lowest (4.7261).

XSL•FO
RenderX

Table 3. Comparison of papers published in different countries.

Country	Avg[a] (DZ^{b}_{AJ}[c])	Amount of AJ articles, n (%)	Amount $\geq MD_{OAMJ}$[d], n (%)	Avg (DZ_{OAMJ})	Amount of OAMJ articles, n (%)	Amount $D_Z \geq MD_{AJ}$[e], n (%)		
United States	4.7261	1537 (76.01)	1435 (93.36)	0.0588	1499 (2.58)	60 (4)	0.0124	0.9753
United Kingdom	5.4159	833 (41.20)	818 (98.20)	0.1113	2967 (5.10)	178 (6)	0.0205	3.5618
Canada	7.5419	479 (23.69)	471 (98.33)	0.0607	845 (1.45)	27 (3.20)	0.0080	1.7641
Germany	8.5920	399 (19.73)	396 (99.25)	0.0538	517 (0.89)	12 (2.32)	0.0063	1.2957
Australia	9.7657	355 (17.56)	350 (98.59)	0.0797	1336 (2.30)	33 (2.47)	0.0082	3.7634
France	11.2194	340 (16.82)	338 (99.41)	0.0757	371 (0.64)	13 (3.50)	0.0068	1.0912
Netherlands	8.2619	296 (14.64)	291 (98.31)	0.0563	648 (1.11)	19 (2.93)	0.0068	2.1892
Italy	9.9492	275 (13.60)	274 (99.64)	0.1065	255 (0.44)	12 (4.71)	0.0107	0.9273
Spain	12.6857	227 (11.23)	225 (99.12)	0.0380	297 (0.51)	6 (2.02)	0.0030	1.3084
Switzerland	10.9062	216 (10.68)	215 (99.54)	0.1126	265 (0.46)	13 (4.91)	0.0103	1.2269
Sweden	8.9276	191 (9.45)	185 (96.86)	0.0346	490 (0.84)	13 (2.65)	0.0039	2.5654
Denmark	8.9724	169 (8.36)	165 (97.63)	0.0761	381 (0.65)	12 (3.15)	0.0085	2.2544
China Mainland	13.1854	167 (8.26)	166 (99.40)	0.0738	945 (1.62)	47 (4.97)	0.0056	5.6587
Japan	17.6698	166 (8.21)	161 (96.99)	0.0439	319 (0.55)	7 (2.19)	0.0025	1.9217
Belgium	13.5063	160 (7.91)	158 (98.75)	0.1137	202 (0.35)	9 (4.46)	0.0084	1.2625
Brazil	14.4105	139 (6.87)	135 (97.12)	0.0448	147 (0.25)	7 (4.76)	0.0031	1.0576
Poland	12.6655	128 (6.33)	128 (100)	0.0254	80 (0.14)	2 (2.50)	0.0020	0.6250
South Korea	17.2768	117 (5.79)	117 (100)	0.5476	181 (0.31)	4 (2.21)	0.0317	1.5470
New Zealand	12.9764	103 (5.09)	100 (97.09)	0.1546	231 (0.40)	4 (1.73)	0.0119	2.2427
South Africa	15.5478	98 (4.85)	97 (98.98)	0.0410	163 (0.28)	6 (3.68)	0.0026	1.6633
India	15.8468	95 (4.70)	94 (98.95)	0.0913	154 (0.26)	10 (6.49)	0.0058	1.6211
Norway	15.7190	92 (4.55)	88 (95.65)	0.0238	334 (0.57)	8 (2.40)	0.0015	3.6304
Israel	11.6714	87 (4.30)	86 (98.85)	0.0189	43 (0.07)	1 (2.33)	0.0016	0.4943
Finland	13.9462	84 (4.15)	83 (98.81)	0.1585	161 (0.28)	7 (4.35)	0.0114	1.9167
Austria	16.6162	83 (4.10)	83 (100)	0.2537	65 (0.11)	2 (3.08)	0.0153	0.7831
Russia	17.6035	77 (3.81)	77 (100)	0.0259	12 (0.02)	0 (0)	0.0015	0.1558
Ireland	12.5726	76 (3.76)	76 (100)	0.0607	189 (0.32)	8 (4.23)	0.0048	2.4868
Taiwan	19.8153	68 (3.36)	68 (100)	0.0245	213 (0.37)	4 (1.88)	0.0012	3.1324
Czech Republic	10.1522	65 (3.21)	65 (100)	0.0464	21 (0.04)	1 (4.76)	0.0046	0.3231
Singapore	17.4077	65 (3.21)	65 (100)	0.0808	106 (0.18)	4 (3.77)	0.0046	1.6308
Argentina	15.2021	61 (3.02)	60 (98.36)	0.1178	18 (0.03)	3 (16.67)	0.0077	0.2951
Turkey	19.7073	61 (3.02)	61 (100)	0.0983	21 (0.04)	4 (19.05)	0.0050	0.3443
Hong Kong	18.6193	58 (2.87)	57 (98.28)	0.0208	93 (0.16)	2 (2.15)	0.0011	1.6034
Mexico	25.8280	58 (2.87)	57 (98.28)	0.0132	33 (0.06)	0 (0)	0.0005	0.5690
Chile	19.7926	52 (2.57)	52 (100)	0.1587	37 (0.06)	6 (16.22)	0.0080	0.7115
Colombia	25.5139	52 (2.57)	52 (100)	0.0752	17 (0.03)	2 (11.76)	0.0029	0.3269
Greece	16.5438	52 (2.57)	52 (100)	0.0681	40 (0.07)	2 (5)	0.0041	0.7692
Portugal	26.4890	52 (2.57)	52 (100)	0.2982	57 (0.10)	4 (7.02)	0.0113	1.0962
Hungary	15.0707	51 (2.52)	51 (100)	0.0099	20 (0.03)	0 (0)	0.0007	0.3922
Malaysia	18.8086	47 (2.32)	47 (100)	0.0523	73 (0.13)	4 (5.48)	0.0028	1.5532

XSL•FO
RenderX

Country	Avg[a] (DZ^{b}_{AJ}[c])	Amount of AJ articles, n (%)	Amount $\geq MD_{OAMJ}$[d], n (%)	Avg (DZ_{OAMJ})	Amount of OAMJ articles, n (%)	Amount $D_Z \geq MD_{AJ}$[e], n (%)		
Kenya	22.9862	43 (2.13)	40 (93.02)	0.0629	60 (0.10)	3 (5)	0.0027	1.3953
Pakistan	25.5705	41 (2.03)	40 (97.56)	0.1181	47 (0.08)	7 (14.89)	0.0046	1.1463
Ukraine	27.0706	37 (1.83)	37 (100)	0.0079	4 (0.01)	0 (0)	0.0003	0.1081
Iran	29.1602	36 (1.78)	36 (100)	0.0611	49 (0.08)	3 (6.12)	0.0021	1.3611
Romania	31.1367	36 (1.78)	36 (100)	0.0210	13 (0.02)	0 (0)	0.0007	0.3611
Peru	35.6125	35 (1.73)	35 (100)	0.0118	11 (0.02)	0 (0)	0.0003	0.3143
Nigeria	33.4455	34 (1.68)	34 (100)	0.0410	24 (0.04)	2 (8.33)	0.0012	0.7059
The Philippines	25.4935	32 (1.58)	32 (100)	0.2619	17 (0.03)	3 (17.65)	0.0103	0.5313
Bangladesh	30.1269	31 (1.53)	31 (100)	0.0815	78 (0.13)	9 (11.54)	0.0027	2.5161
Serbia	29.4950	31 (1.53)	31 (100)	0.0043	6 (0.01)	0 (0)	0.0001	0.1935

[a]Avg: average.

[b]D_Z: absolute disruption index.

[c]AJ: authoritative journal.

[d]MD_{OAMJ}: median D_Z of the open-access mega journal (OAMJ) papers.

[e]MD_{AJ}: median D_Z of the authoritative journal papers.

XSL•FO
RenderX

Table 4. Indicator correlation analysis.

Category	AJ[a]				OAMJ[b]					
	Avg[c] (D_Z[d])	Amount	Amount (D_Z>MD$_{OAMJ}$[e])	Proportion (D_Z>MD$_{OAMJ}$)	Avg (D_Z)	Amount	Amount (D_Z>MD_{AJ}[f])	Proportion (D_Z>MD_{AJ})		
AJ										
Avg (D_Z)	1	–0.1494	–0.1452	0.359**	–0.1118	–0.265**	–0.256**	–0.0809	–0.461**	–0.322**
Amount		1	1**	–0.514**	0.518**	0.895**	0.744**	0.459**	0.513**	0.448**
Amount (D_Z>MD_{OAMJ})			1	–0.503**	0.518**	0.894**	0.743**	0.459**	0.512**	0.444**
Proportion (D_Z>MD_{OAMJ})				1	–0.196*	–0.537**	–0.497**	–0.197*	–0.338**	–0.448**
OAMJ										
Avg (D_Z)					1	0.579**	0.732**	0.827**	0.878**	0.463**
Amount						1	0.806**	0.523**	0.615**	0.764**
Amount (D_Z>MD_{AJ})							1	0.821**	0.711**	0.616**
Proportion (D_Z>MD_{AJ})								1	0.689**	0.398**
									1	0.553**
										1

[a]AJ: authoritative journal.

[b]OAMJ: open-access mega journal.

[c]Avg: average.

[d]D_Z: absolute disruption index.

[e]MD_{OAMJ}: median D_Z of the OAMJ papers.

[f]MD_{AJ}: median D_Z of the authoritative journal papers.

* P<.01

**P<.001

There was a linear relationship between the number of authoritative journal articles and the average disruptive innovation level of each country after logarithmization (as shown in Figure 1), with a correlation coefficient of –0.891, which indicated that authoritative journals did not consistently recognize the work of scholars from different countries. However, for OAMJ papers issued by each country, the average level of disruptive innovation did not show a linear relationship after logarithmic processing (as shown in Figure 2).

Scholars from different countries also had varying levels of recognition for OAMJs and tendencies to publish them. A higher recognition of OAMJs correlated with a higher likelihood of publishing papers in these journals (0.553). This was significantly correlated with the disruptive innovation levels of both authoritative journal papers (–0.461) and OAMJ papers (0.878). Moreover, the proportion of OAMJ papers meeting the MD_{AJ} threshold was significantly and positively correlated with the tendency to publish papers in OAMJs (0.689).

Among all countries, those that published more than 10% of OAMJ papers with a level of disruptive innovation above MD_{AJ}, while maintaining an average annual publication count of at least 1 in the selected journals, included Argentina, Turkey, Chile, Colombia, Pakistan, Philippines, Bangladesh, and Egypt.

XSL•FO
RenderX

Figure 1. Distribution of the number of publications in authority journals with average disruptive innovation levels among different countries.

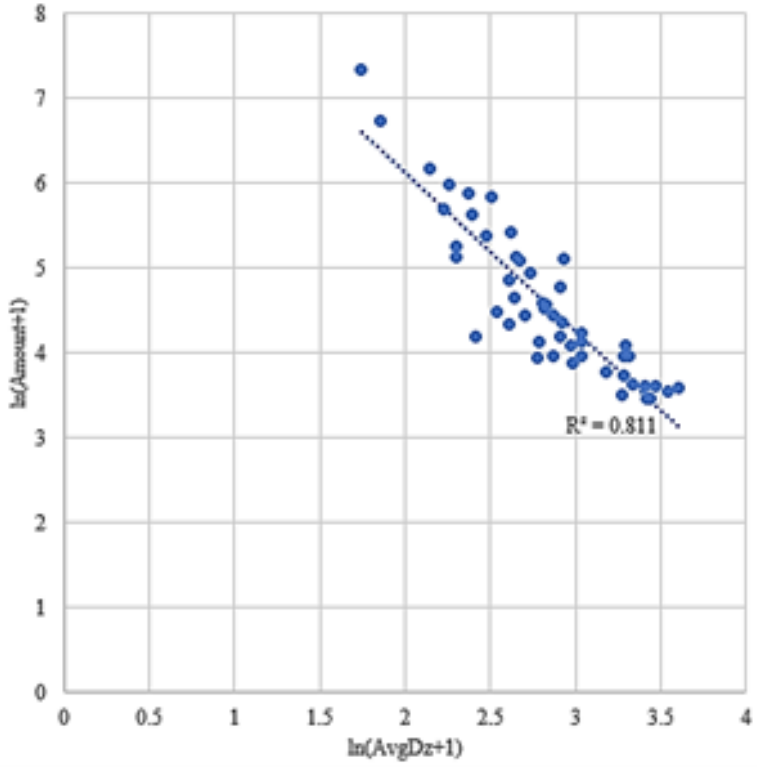

Figure 2. Distribution of the number of publications in open-access mega journals (OAMJs) with the average disruptive innovation levels among different countries.

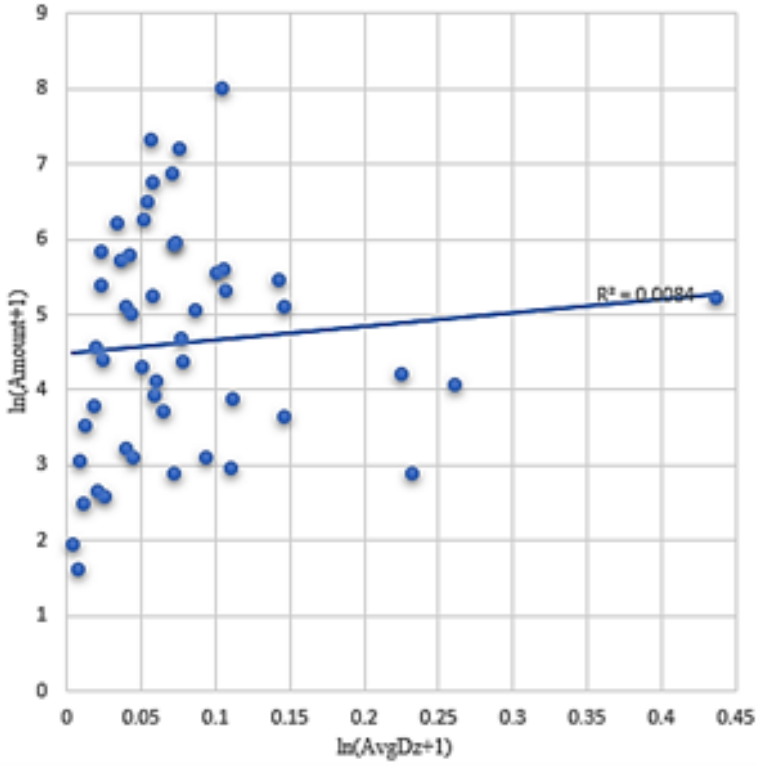

XSL•FO
RenderX

Discussion

Principal Results

Differences Between the Disruptive Innovation Levels of OAMJ and Authoritative Journals

In this study, the disruptive innovation levels of OAMJs were lower than those of authoritative journals. A plausible explanation is that OAMJs require more rationality than innovation during the review process. Therefore, OAMJ papers were more focused on incremental innovation and proximity innovation, while authoritative journals were more likely to publish papers with breakthrough innovation.

OAMJs As Important Publishing Platforms for Innovative Research Across Fields

In this study, the overall disruptive innovation levels of authoritative journal papers were higher than those of OAMJ papers across all clinical medical research fields. However, the number of OAMJ papers with a higher level of disruptive innovation than MD_{AJ} was close to half of the number of authoritative journal papers. Moreover, in almost half of the research directions examined, the number of OAMJ papers with a higher level of disruptive innovation than MD_{AJ} exceeded half of the count for authoritative journal papers.

This phenomenon indicates that the contribution of OAMJs to the overall development of science rivals that of authoritative journals. While OAMJs must continue to emphasize overall quality control, their contribution to scientific development should not be overlooked [46].

Additionally, OAMJ papers involve some research directions that are not covered by authoritative journal papers. Therefore, it is unreasonable to judge the level of disruptive innovation of a paper based solely on the source journal without distinguishing the research direction. Thus, we must evaluate the contributions of OAMJs fairly and scientifically [47].

OAMJs Provide a Fair Publishing Channel for Innovative Achievements from Limited-Income Countries

Research has confirmed that national bias exists in the peer-review process [48], and specific journals have preferences for their authors' countries of origin [49]. Therefore, authoritative journals demonstrate varying degrees of recognition for the research results of scholars from different countries. This phenomenon was also evident in this study, where notable differences were observed in the disruptive innovation levels of authoritative journal papers published by scholars from different countries. Furthermore, the disruptive innovation levels in authoritative journals from each country were linearly and negatively correlated with the volume of publications.

A reasonable explanation for this is the first-mover advantage of high-income countries in the field of science and technology. These nations benefit from a group of senior scholars in various fields whose work often receives recognition that exceeds its quality in the current peer-review process [50]. This recognition also extends to the young scholars they collaborate with [51]. On the other hand, this study also found that OAMJs play an important role in promoting innovation diffusion in limited-income countries. Therefore, OAMJs may provide fairer publishing opportunities, and in tandem with postpublication peer review [52], may contribute to the accelerated development of science and technology.

Limitations

This study has some limitations that must be acknowledged. First, participating authors with different ranks were treated as equal contributors, disregarding variations. Second, all focus papers were categorized under a single subject classification, without considering their inherent interdisciplinary properties. Third, only the knowledge diffusion of focus papers within the scope of periodical literature was considered, ignoring other venues. These limitations must be explored in future research based on natural language processing technology, and more accurate results can be obtained by jointly using multiple types of citation data sources.

Conclusions

The disruptive innovation levels of OAMJs were different from those of authoritative journals. However, OAMJs have become an important publication channel for innovative results in several research directions, providing fairer opportunities for researchers from limited-income countries to publish their work. Therefore, the academic community should acknowledge and appreciate the contributions of OAMJs to scientific research.

Acknowledgments

Author XL was supported by the National Social Science Foundation of China (23BTQ085).

Authors' Contributions

XL reviewed and edited the manuscript, supervised the study, and acquired funding. YJ conceptualized the study, was responsible for the methodology, provided the software and other resources, performed the formal analysis, and wrote the original manuscript draft. YJ and LW curated the data.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Björk B, Solomon D. The publishing delay in scholarly peer-reviewed journals. J Informetr 2013 Oct;7(4):914-923. [doi: 10.1016/j.joi.2013.09.001]
2. Björk B. Publishing speed and acceptance rates of open access megajournals. Online Inf Rev 2018 Nov 12;45(2):270-277. [doi: 10.1108/oir-04-2018-0151]
3. Solomon D. A survey of authors publishing in four megajournals. PeerJ 2014;2:e365 [FREE Full text] [doi: 10.7717/peerj.365] [Medline: 24795855]
4. Björk BC. Evolution of the scholarly mega-journal, 2006-2017. PeerJ 2018;6:e4357 [FREE Full text] [doi: 10.7717/peerj.4357] [Medline: 29441240]
5. Björk BC. Have the "mega-journals" reached the limits to growth? PeerJ 2015;3:e981 [FREE Full text] [doi: 10.7717/peerj.981] [Medline: 26038735]
6. Khelfaoui M, Gingras Y. Expanding Nature: Product line and brand extensions of a scientific journal. Learn Publ 2021 Oct 05;35(2):187-197. [doi: 10.1002/leap.1422]
7. Wakeling S, Spezi V, Fry J, Creaser C, Pinfield S, Willett P. Academic communities: The role of journals and open-access mega-journals in scholarly communication. J Doc 2019 Jan 14;75(1):120-139. [doi: 10.1108/JD-05-2018-0067]
8. Björk B, Catani P. Peer review in megajournals compared with traditional scholarly journals: Does it make a difference? Learn Publ 2016 Jan 20;29(1):9-12. [doi: 10.1002/leap.1007]
9. Erfanmanesh M, Teixeira da Silva JA. Is the soundness-only quality control policy of open access mega journals linked to a higher rate of published errors? Scientometrics 2019 Jul 5;120(2):917-923. [doi: 10.1007/s11192-019-03153-5]
10. Beall J. Predatory publishing is just one of the consequences of gold open access. Learn Publ 2013 Apr;26(2):79-84. [doi: 10.1087/20130203]
11. Nugent K, Edriss H, Ball S, Björk BC. Mega journals, scientifically sound peer review and medical organizations. Am J Med Sci 2019 Jan;357(1):3-4. [doi: 10.1016/j.amjms.2018.09.003] [Medline: 30528946]
12. Borrego A. Are mega-journals a publication outlet for lower quality research? A bibliometric analysis of Spanish authors in. Online Inf Rev 2018 Nov 05;45(2):261-269. [doi: 10.1108/oir-04-2018-0136]
13. Retracted: growing up in a recession. Rev Econ Stud 2023 Mar 06;90(2):1009-1009. [doi: 10.1093/restud/rdt040]
14. Pinto AP, Mejdalani G, Mounce R, Silveira LF, Marinoni L, Rafael JA. Are publications on zoological taxonomy under attack? R Soc Open Sci 2021 Feb 10;8(2):201617 [FREE Full text] [doi: 10.1098/rsos.201617] [Medline: 33972859]
15. Dalton ED, Tenopir C, Björk B. Attitudes of North American academics toward open access scholarly journals. Portal Libr Acad 2020;20(1):73-100. [doi: 10.1353/pla.2020.0005]
16. Ioannidis JPA, Pezzullo AM, Boccia S. The rapid growth of mega-journals: threats and opportunities. JAMA 2023 Apr 18;329(15):1253-1254. [doi: 10.1001/jama.2023.3212] [Medline: 36939740]
17. Garfield E. Citation analysis as a tool in journal evaluation. Science 1972 Nov 03;178(4060):471-479. [doi: 10.1126/science.178.4060.471] [Medline: 5079701]
18. McKiernan E, Schimanski L, Muñoz NC, Matthias L, Niles M, Alperin J. Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. eLife 2019 Jul 31:8. [doi: 10.7554/elife.47338]
19. Wang J, Veugelers R, Stephan P. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy 2017 Oct;46(8):1416-1436. [doi: 10.1016/j.respol.2017.06.006]
20. Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. J Clin Epidemiol 2019 Mar;107:36-41 [FREE Full text] [doi: 10.1016/j.jclinepi.2018.11.014] [Medline: 30458264]
21. Maleck WH, Boldt J, Koetter KP. German publications in leading English language journals. Dtsch Med Wochenschr 2001 Jan 19;126(3):33-41. [doi: 10.1055/s-2001-10362] [Medline: 11205476]
22. Gehanno J, Ladner J, Rollin L, Dahamna B, Darmoni SJ. How are the different specialties represented in the major journals in general medicine? BMC Med Inform Decis Mak 2011 Jan 21;11(1):3 [FREE Full text] [doi: 10.1186/1472-6947-11-3] [Medline: 21255439]
23. Gehanno J, Rollin L, Ladner J, Darmoni SJ. How is occupational medicine represented in the major journals in general medicine? Occup Environ Med 2012 Aug 12;69(8):603-605. [doi: 10.1136/oemed-2011-100044] [Medline: 22693268]
24. Wu J, Li H, Huang Y, Tong G, Tian M, Wu Y, et al. Chinese contribution to NEJM, Lancet, JAMA, and BMJ from 2011 to 2020: a 10-year bibliometric study. Ann Transl Med 2022 May;10(9):505 [FREE Full text] [doi: 10.21037/atm-21-6793] [Medline: 35928742]
25. Gehanno J, Grosjean J, Darmoni SJ, Rollin L. Reliability of citations of medRxiv preprints in articles published on COVID-19 in the world leading medical journals. PLoS One 2022 Aug 10;17(8):e0264661 [FREE Full text] [doi: 10.1371/journal.pone.0264661] [Medline: 35947594]
26. Bornmann L, Devarakonda S, Tekles A, Chacko G. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. Quant Sci Stud 2020 Aug;1(3):1242-1259. [doi: 10.1162/qss_a_00068]
27. Liu X, Shen Z, Liao Y, Zhu M, Yang L. Research on the stable time window of disruption index. Lib Inf Serv 2021 Sep 20;65(18):49-57. [doi: 10.13266/j.issn.0252-3116.2021.18.006]
28. Bornmann L, Tekles A. Disruption index depends on length of citation window. Eur J Public Health 2019 Mar 25;28(2). [doi: 10.3145/epi.2019.mar.07]

XSL•FO
RenderX

29. Song P, Feng C, Long C, Yang Z, Song Y. Study on discovery of outstanding scientific and technological talents in specific domains based on optimized disruptive index. J Intell 2022 May 10;41(5):61-65. [doi: 10.3969/j.issn.1002-1965.2022.05.010]
30. Christensen C. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Rev., updated, and with a new chapter, Nachdr. Boston, MA: Harvard Business School Press; 2008.
31. Wu L, Wang D, Evans JA. Large teams develop and small teams disrupt science and technology. Nature 2019 Feb 13;566(7744):378-382. [doi: 10.1038/s41586-019-0941-9] [Medline: 30760923]
32. Ruan X, Lyu D, Gong K, Cheng Y, Li J. Rethinking the disruption index as a measure of scientific and technological advances. Technol Forecast Soc Change 2021 Nov;172:121071. [doi: 10.1016/j.techfore.2021.121071]
33. Liu X, Liao Y, Zhu M. A preliminary study on the application of disruption index to scientific research evaluation. Inf Stud Theory Appl 2021 Dec 31;44(12):34-40. [doi: 10.16353/j.cnki.1000-7490.2021.12.005]
34. Liu X, Shen Z, Liao Y, Yang L. The research about the improved disruption index and its influencing factors. Lib Inf Serv 2020 Dec 20;64(24):84-91. [doi: 10.13266/j.issn.0252-3116.2020.24.010]
35. Bornmann L, Devarakonda S, Tekles A, Chacko G. Disruptive papers published in Scientometrics: meaningful results by using an improved variant of the disruption index originally proposed by Wu, Wang, and Evans (2019). Scientometrics 2020 Mar 14;123(2):1149-1155. [doi: 10.1007/s11192-020-03406-8]
36. Horen S, Hansdorfer M, Kronshtal R, Dorafshar A, Becerra A. The most disruptive publications in craniofacial surgery (1954-2014). J Craniofac Surg 2021 Oct 01;32(7):2426-2430. [doi: 10.1097/SCS.0000000000007804] [Medline: 34054087]
37. Sullivan GA, Skertich NJ, Gulack BC, Becerra AZ, Shah AN. Shifting paradigms: The top 100 most disruptive papers in core pediatric surgery journals. J Pediatr Surg 2021 Aug;56(8):1263-1274. [doi: 10.1016/j.jpedsurg.2021.02.002] [Medline: 33678406]
38. Meyer C, Nakamura Y, Rasor BJ, Karim AS, Jewett MC, Tan C. Analysis of the innovation trend in cell-free synthetic biology. Life (Basel) 2021 Jun 11;11(6):551. [doi: 10.3390/life11060551] [Medline: 34208358]
39. Jiang Y, Liu X. A bibliometric analysis and disruptive innovation evaluation for the field of energy security. Sustainability 2023 Jan 05;15(2):969. [doi: 10.3390/su15020969]
40. Jiang Y, Liu X. A new method of calculating the disruption index based on open citation data. J Inf Sci 2024 Oct 03. [doi: 10.1177/01655515241263545]
41. Jiang Y, Liu X. An innovative evaluation index of scientific journals—Journal Disruption Index (JDI) and its empirical research. Chin J Sci Technol Period 2006 Dec;33(4):965-972. [doi: 10.11946/cjstp.202205240418]
42. Jiang Y, Liu X. A construction and empirical research of the journal disruption index based on open citation data. Scientometrics 2023 May 18;128(7):3935-3958 [FREE Full text] [doi: 10.1007/s11192-023-04737-y] [Medline: 37287879]
43. Jiang Y, Wang L, Liu X. Innovative evaluation of Chinese SCIE-indexed scientific journals: An empirical study based on the disruption index. Chin J Sci Technol Period 2023 Aug 15;34(8):1060-1068. [doi: 10.11946/cjstp.202302190096]
44. Jiang Y, Liu X, Zhang Z, Yang X. Evaluation and comparison of academic impact and disruptive innovation level of medical journals: bibliometric analysis and disruptive evaluation. J Med Internet Res 2024 May 31;26:e55121 [FREE Full text] [doi: 10.2196/55121] [Medline: 38820583]
45. Jiang Y, Liu X. The disruptive innovation evaluation and empirical analysis of Chinese, Japanese, Indian and South Korean scientific journals. J Scholarly Publ (forthcoming) 2025. [doi: 10.3138/jsp-2023-0056]
46. Park M, Leahey E, Funk R. Papers and patents are becoming less disruptive over time. Nature 2023 Jan;613(7942):138-144. [doi: 10.1038/s41586-022-05543-x] [Medline: 36600070]
47. Jiang Y, Liu X. The relationship between absolute disruption index, peer review index and CNCI: a study based on virology papers. Lib Inf Serv 2023 Feb 05;67(3):96-105. [doi: 10.13266/j.issn.0252-3116.2023.03.009]
48. Thelwall M, Allen L, Papas E, Nyakoojo Z, Weigert V. Does the use of open, non-anonymous peer review in scholarly publishing introduce bias? Evidence from the F1000Research post-publication open peer review publishing model. J Inf Sci 2020 Jul 05;47(6):809-820. [doi: 10.1177/0165551520938678]
49. Zhu H. Home country bias in academic publishing: A case study of the he New England Journal of Medicine. Learn Publ 2021 Jun;34(4):578-584. [doi: 10.1002/leap.1404]
50. Kardam K, Kejriwal A, Sharma K, Kaushal R. Ranking scholarly work based on author reputation. 2016 Presented at: International Conference on Advances in Computing, Communications and Informatics (ICACCI) .?2088; September 21-24; Jaipur, India p. 2084-2088. [doi: 10.1109/icacci.2016.7732358]
51. Bu Y, Murray DS, Xu J, Ding Y, Ai P, Shen J, et al. Analyzing scientific collaboration with "giants" based on the milestones of career. Proc Assoc Info Sci Tech 2019 Feb;55(1):29-38. [doi: 10.1002/pra2.2018.14505501005]
52. Jiang Y, Liu X. Open peer review: Mode, technology, problems, and countermeasures. Chin J Sci Tech Period 2022 Sep 15;33(9):1196-1205. [doi: 10.11946/cjstp.202202280120]

Abbreviations

BMJ: British Medical Journal
DZ: absolute disruption index
JAMA: Journal of the American Medical Association

XSL•FO
RenderX

JCR: Journal Citation Reports
JDI: Journal Disruption Index
JIF: Journal Impact Factor
MDAJ: the median of authoritative journal papers
MDOAMJ: the median of open-access mega journal papers
OA: open access
OAMJ: open-access mega journal
PLoS: Public Library of Science
POCI: OpenCitations Index of PubMed open PMID-to-PMID citations
WoS: Web of Science

Edited by A Schwartz; submitted 17.04.24; peer-reviewed by M Hassan, F Lin, A Borrego; comments to author 15.05.24; revised version received 02.07.24; accepted 13.11.24; published 15.01.25.

Please cite as:
Jiang Y, Liu XL, Wang L
Evaluation and Comparison of the Academic Quality of Open-Access Mega Journals and Authoritative Journals: Disruptive Innovation Evaluation
J Med Internet Res 2025;27:e59598
URL: https://www.jmir.org/2025/1/e59598
doi:10.2196/59598
PMID:

©Yuyan Jiang, Xue-li Liu, Liyun Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Quantifying Public Engagement With Science and Malinformation on COVID-19 Vaccines: Cross-Sectional Study

David Robert Grimes[1,2*], BSc, CSTAT, PhD; David H Gorski[3,4*], MD, PhD

[1]TCD Biostatistics Unit, School of Medicine, Trinity College Dublin, Dublin, Ireland

[2]School of Physical Sciences, Dublin City University, Dublin, Ireland

[3]Michael and Marian Ilitch Department of Surgery, Wayne State University, Detroit, MI, United States

[4]Barbara Ann Karmanos Cancer Institute, Detroit, MI, United States

[*]all authors contributed equally

Corresponding Author:
David Robert Grimes, BSc, CSTAT, PhD
TCD Biostatistics Unit
School of Medicine
Trinity College Dublin
Trinity College Institute of Population Health
Russell Centre, 7th Fl, Tallaght Cross West, Tallaght
Dublin, D24 DH74
Ireland
Phone: 353 19861075
Email: davidrobert.grimes@tcd.ie

Abstract

Background: Medical journals are critical vanguards of research, and previous years have seen increasing public interest in and engagement with medicoscientific findings. How findings propagate and are understood and what harms erroneous claims might cause to public health remain unclear, especially on publicly contentious topics like COVID-19 vaccines. Gauging the engagement of the public with medical science and quantifying propagation patterns of medicoscientific papers are thus important undertakings. In contrast to misinformation and disinformation, which pivot on falsehood, the more nuanced issue of malinformation, where ostensibly true information is presented out of context or selectively curated to cause harm and misconception, has been less researched. As findings and facts can be selectively marshaled to present a misleading picture, it is crucial to consider this issue and its potential ramifications.

Objective: This study aims to quantify patterns of public engagement with medical research and the vectors of propagation taken by a high-profile incidence of medical malinformation.

Methods: In this work, we undertook an analysis of all altmetric engagements over a decade for 5 leading general-purpose medical journals, constituting approximately 9.8 million engagements with 84,529 papers. We identify and examine the proliferation of sentiment concerning a high-profile publication containing vaccine-negative malinformation. Engagement with this paper, with the highest altmetric score of any paper in an academic journal ever released, was tracked across media outlets worldwide and in social media users on Twitter (subsequently rebranded as X). Vectoring media sources were analyzed, and manual sentiment analysis on high-engagement Twitter shares of the paper was undertaken, contrasted with users' prior vaccine sentiment.

Results: Results of this analysis suggested that this COVID-19 scientific malinformation was much more likely to be engaged and amplified with negative by vaccine-negative Twitter accounts than neutral ones (odds ratio 58.2, 95% CI 9.7-658.0; $P<.001$), often alluding to the ostensible prestige of medical journals. Malinformation was frequently invoked by conspiracy theory websites and non-news sources (71/181 citations, 39.2%) on the internet to cast doubt on the efficacy of vaccination, many of whom tended to cite the paper repeatedly (51/181, 28.2%).

Conclusions: Our findings suggest growing public interest in medical science and present evidence that medical and scientific journals need to be aware of not only the potential overt misinformation but also the more insidious impact of malinformation. Also, we discuss how journals and scientific communicators can reduce the influence of malinformation on public understanding.

(J Med Internet Res 2025;27:e64679) doi:10.2196/64679

XSL•FO
RenderX

KEYWORDS

misinformation; altmetrics; disinformation; malinformation; public engagement; medical journals; medicoscientific; public health; altmetric analysis; comparative analysis; social media; Twitter; vaccine; digital health; mHealth; mobile health; health informatics

Introduction

The growing dominance of social media as a news source has resulted in not only widespread engagement with science but also a perpetuation of medical misinformation in fields ranging from dementia [1] to cardiology [2] and cancer [3,4] and beyond. Vaccines have been a topic of misinformation for centuries [5], with antivaccine propaganda being a leading contributor to declining vaccine uptake worldwide [6], and in 2019, the World Health Organization named vaccine hesitancy a top threat to public health, with COVID-19 vaccines being a recent major focus for an abundance of misinformation [7,8].

Misinformation undermines trust in medical science [9], leading to calls for medical journals to be more proactive in combatting health fiction, pseudoscience, and misconceptions. Armstrong and Naylor [10] argue that medical journals have a critical role in addressing medical misinformation by containing the dissemination of misinformation, identifying falsehoods and their purveyors, and striving to avoid false balance and legitimization of false information. They suggest that medical journals have a unique role in debunking myths and facilitating health-specific inoculation and education. Medical journals, as repositories of trustworthy information, have a central role to play, as "...the ability to credit information from scientific journals, and...to discredit information without such sources, is perhaps the most conventional countermeasure to misinformation" [11]. Directing the public to reliable sources of information, including medical journals, has been identified as an urgent undertaking in combatting COVID-19 misinformation [12].

Medicoscientific journals are not only trusted gatekeepers for reliable findings but are also far from infallible. Moreover, peer-reviewed medical science is no longer read and propagated solely by medical scientists but is widely disseminated by social media influencers and the wider public. The pandemic era has witnessed several high-profile retractions resulting from flawed analysis or even outright fabricated data. As Toth et al [13] note, "Medical journals not properly 'vetting' and eventually publishing high-impact papers that turn out to be based on unverified data can misguide providers as well as the public. Information on the disease that was previously restricted to medical professionals is now available to the broad public by the swipe of a screen, just like any other newsreel." Accordingly, findings can be used in misinformation, disinformation, and malinformation, as defined in Table 1.

Table 1. Council of Europe definitions [14].

Term	Definition	Characteristics
Misinformation	"Information that is false, but not created with the intention of causing harm"	False information spread inadvertently or intentionally.
Disinformation	"Information that is false and deliberately created to harm a person, social group, organization or country"	False information introduced and spread with the intent to mislead or confound.
Malinformation	"Information that is based on reality, used to inflict harm on a person, organization or country"	Nominally true information is misleadingly presented or curated with the intention to deceive or mislead.

Unlike misinformation and disinformation, malinformation does not depend on false information, but instead leverages plausible and accurate—or at least arguably accurate—information presented either without context or in an incorrect context to lead those encountering it to false or misleading impressions and conclusions. When encountered without cognizance of crucial background or context, malinformation can create perceptions at odds with reality and cause harm to individuals, organizations, and collective understanding. In medical journals, malinformation can emerge through several mechanisms.

- Authors curate information out of context, with the net result of presenting a misleading impression of findings that are undetected by peer reviewers and editorial staff.
- Distortion of published research findings by third-party vested interests to cherry-pick or misrepresent legitimate published medical science to create a narrative at odds with the scientific evidence base.
- The synergy between these mechanisms.

To date, there has been comparatively little research done on quantifying engagement with medical science or the problems of malinformation in a medical context. Altmetrics measure the broader impact of scholarly works by analyzing social media and media engagement beyond traditional citations. They track mentions on social media, news, blogs, policy documents, and academic tools like Mendeley. Data are gathered using persistent identifiers (typically DOIs), which are weighted by source to generate composite altmetric scores. Altmetrics provide immediate insights into public and academic attention, complementing traditional metrics, a barometer of how research is understood and engaged with by the public on social and conventional media [15]. As they yield data complementary to citation-based metrics, tracking social media engagement, press coverage, and social media discussion, we took a novel approach of analyzing altmetrics as an indicator of engagement with medical science and the spread of potential malinformation, providing insight into how claims in medical journals are consumed in context and potential public health impacts.

XSL•FO
RenderX

Methods

Quantifying Engagement Metrics

Raw historical altmetric records for 5 leading medical journals (*New England Journal of Medicine*, *The Lancet*, *JAMA*, *Annals of Internal Medicine*, and *British Medical Journal*) for all papers published between April 2012 and April 2022 were exported from Altmetric Explorer. The CSV files for each journal included altmetric attention scores and publication dates for all papers in the 10-year interval, which were extracted and imported into MATLAB (MathWorks) on May 10, 2022. These journals were selected as a representative sample of respected broad-interest medical journals as opposed to more specialist offerings [16-18] to capture the typical dynamics of public engagement with general medical science over a long interval. Data were subsequently pooled and sorted to produce a hierarchical ranking of all altmetric scores. From this, the Lorenz curve and Gini coefficient were derived to ascertain the relative importance ranking of papers in terms of public engagement. Mathematical outlines are provided in Multimedia Appendix 1, and the process is illustrated in Figure 1A.

Figure 1. (A) Extraction of altmetric data for papers published over 10 years for 5 major journals, yielding approximately 9.8 million engagements with 84,529 papers. (B) Deep analysis of identified malinformation, including tweets and media citations.

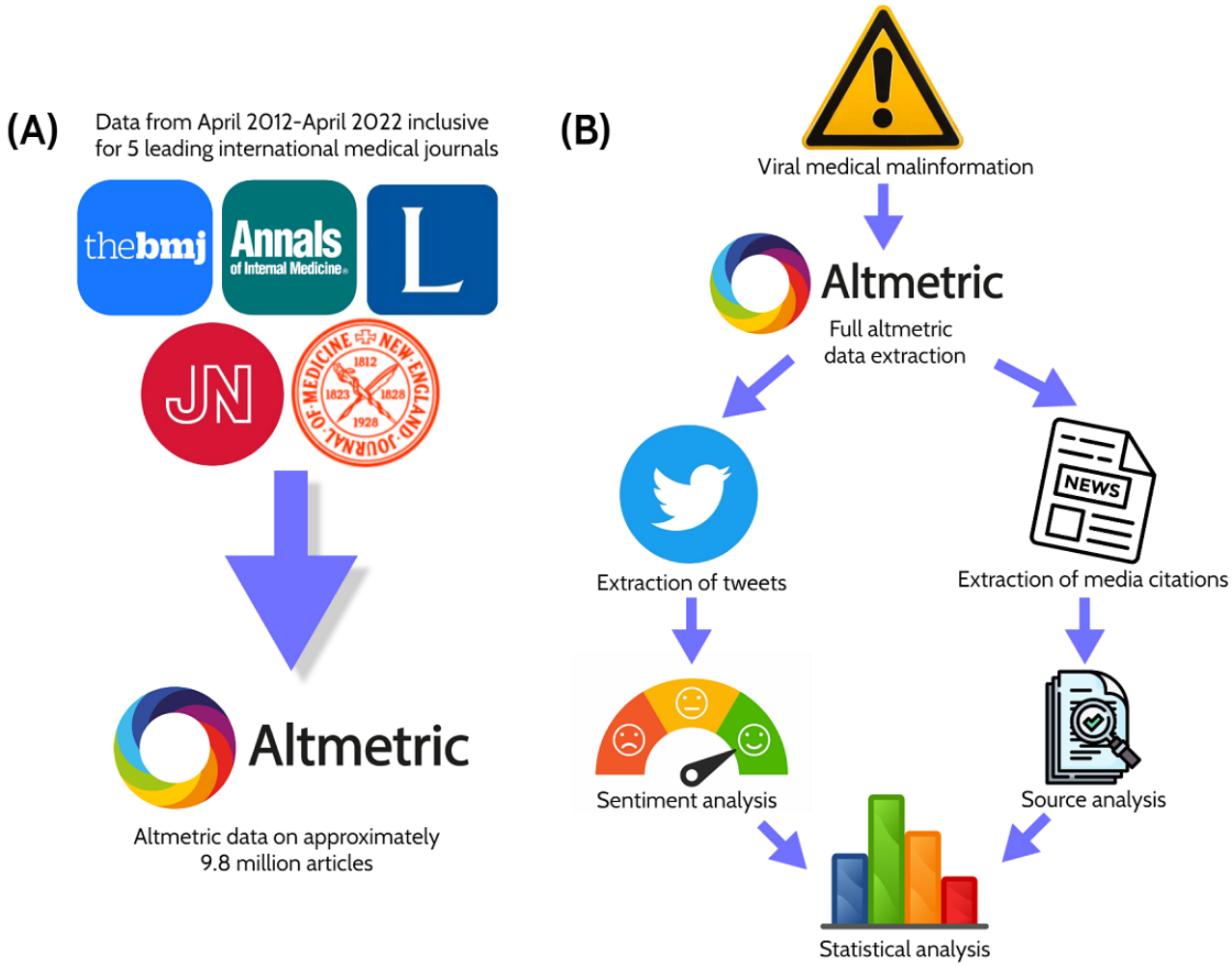

Identification and Quantification of Potential Misinformation or Malinformation

A specific breakdown of the detailed altmetric mentions in a full analysis of a single medical journal publication from November 2021 [19] was also undertaken. For this paper, which garnered the highest altmetric score ever achieved in any scientific journal in any field published to the date of extraction, specific altmetric data were extracted, including all media mentions, outlet names, and linking tweets. At the time of data extraction, direct Twitter (subsequently rebranded as X) engagement with the paper yielded 165,720 tweets from 81,861 users with an upper bound of 11,094,945 followers, 181 media mentions, and a total altmetric score of 45,844. This investigative piece suggested that Pfizer vaccine trials were compromised by the unblinding of subjects and data falsification. Serious issues with this reporting were detected almost immediately after publication by independent fact-checks [20], medical press [21], and medical writers [22]. Criticisms focused chiefly on the paucity of evidence presented to support the statements in the paper, which were largely speculative and extrapolated unverified reported problems at one contract research organization center in Texas to all 153 centers worldwide.

No evidence was presented for the incendiary charge of data falsification, and ostensible problems at a single site were

XSL•FO
RenderX

extrapolated to the entire clinical trial 24 times in the paper. Several other problems in the reporting and framing were identified, outlined in Multimedia Appendix 1. As the paper's assertions, based on unfounded extrapolation, were likely harmful to public perception of Pfizer and of vaccine science more broadly, the European Council's definition of malinformation ("information that is based in reality, but is used to inflict harm on a person, organization or country" [14]) appears most fitting here, given that real potential issues with a single contract research organization being contracted by Pfizer were used as a basis for unevidenced claims that Pfizer "falsified data" and "unblinded patients," despite such assertions being at best speculative extrapolation. This could arguably be deemed misinformation too, a point expanded upon in the *Limitations* section. As this paper has by far the highest altmetric score of any paper in a biomedical journal (or any other scientific work in any field) ever published, extended altmetric analysis was performed to analyze its propagation and reception, including a breakdown of media sources, described in the later section and illustrated in Figure 1B.

Propagation Analysis (Media Outlets)

For the extended analysis, all media mentions were systematically exported from Altmetric Explorer. Mentions with the header "News Story" were extracted into MATLAB. Sources were categorized as news stories (recognized media outlets and investigative media organizations) and aggregators, collectively deemed news sources. Mentions from non-news conspiracy theories or propaganda sites were also counted, as were sources from Russian state media outlets and blogs, collectively deemed non-news sources (objective criteria for demarcation are provided in Multimedia Appendix 1). A nonparametric right-tailed Wilcoxon rank sum test was performed to investigate the hypothesis that non-news sources cited the paper with greater frequency than reliable outlets.

Sentiment Analysis (Social Media)

To ascertain how this paper was understood in a wider context, a Twitter search strategy was implemented for manual sentiment analysis for tweets relating to the paper in question. These were taken from the altmetric data and from an additional search containing the primary terms "Pfizer" and "BMJ," with any of the secondary terms ("whistleblower," "whistle," "fraud," "trials," and "Ventavia"). All tweets with greater than 50 engagements were selected for manual sentiment analysis. For any account hosting such tweets, an advanced search was conducted on their username for all tweets with the terms "vaccine," "vaccines," "COVID," "Pfizer," "hoax," or "lies." The 100 most recent tweets for each user under these criteria were exported. Users whose mention of vaccines in this set was solely negative or propagated discredited antivaccine propaganda were deemed as having vaccine-negative attitudes, while users whose set included mixed or positive vaccine stances were deemed non-negative. When the sentiments were ambiguous, cases were recorded as non-negative. Follower counts, engagement, and reach were also computed, with full data given in the Multimedia Appendix 1. A 2×3 Fisher exact test was used to test the hypothesis that vaccine-negative accounts were far more likely to embrace the paper as evidence against COVID-19 vaccination than non-negative accounts. All raw data are available in Multimedia Appendix 1.

Ethical Considerations

This study involved analysis of publicly accessible tweets, which are considered public behavior data not requiring ethical or institutional board review. No informed consent was sought, as data analyzed are publicly available and no interaction with the authors of the tweets occurred. The data were handled in compliance with ethical guidelines for research using publicly available data, with anonymized tweets—to ensure that individuals cannot be reidentified—available in Multimedia Appendix 1 and unredacted versions available to reviewers and editors. No identifiable images or data from individual users are included in this paper or Multimedia Appendix 1. Where tweet content is cited verbatim, usernames and identifying details have been removed or generalized to protect privacy. An exception was made for a tweet authored by Robert F Kennedy Jr, a public figure, as it represents a public statement made in his capacity as a public advocate.

Results

Quantifying Engagement Metrics

Figure 1A depicts altmetric trends for 5 major medical journals in the decade from April 26, 2012, to April 26, 2022, for 85,529 papers, depicting growing public interest with time. It is worth noting, however, that interest is greatly skewed. Figure 1B depicts the Lorenz curve and derived Gini coefficient of 0.835, demonstrating that altmetrics are driven by a small proportion of papers, quantified in Table 2. For all 5 included journals, the median altmetric scores of papers were also calculated (Figure 2A). As the safety and efficacy of COVID-19 vaccines were a worldwide topic of interest both on social media and traditional media, Table 3 gives the altmetric scores from the most highly discussed research works on vaccine efficacy trials for Moderna, Pfizer, and AstraZeneca vaccines.

Table 2. Altmetric statistics (medical papers).

Percentile of papers (%)	Proportion of total altmetrics (%)
Top 1	34.2
Top 0.5	25.6
Top 0.1	11.1
Top 0.01	2.37

XSL•FO
RenderX

Figure 2. (A) Altmetric scores for every paper published in 5 leading medical journals between April 26, 2012, and April 26, 2022. The MAS and number of publications are given in the legend. (B) Lorenz curve for medical altmetrics (Gini coefficient=0.835). Ann Intern Med: Annals of Internal Medicine; BMJ: British Medical Journal; MAS: median altmetric score; NEJM: New England Journal of Medicine.

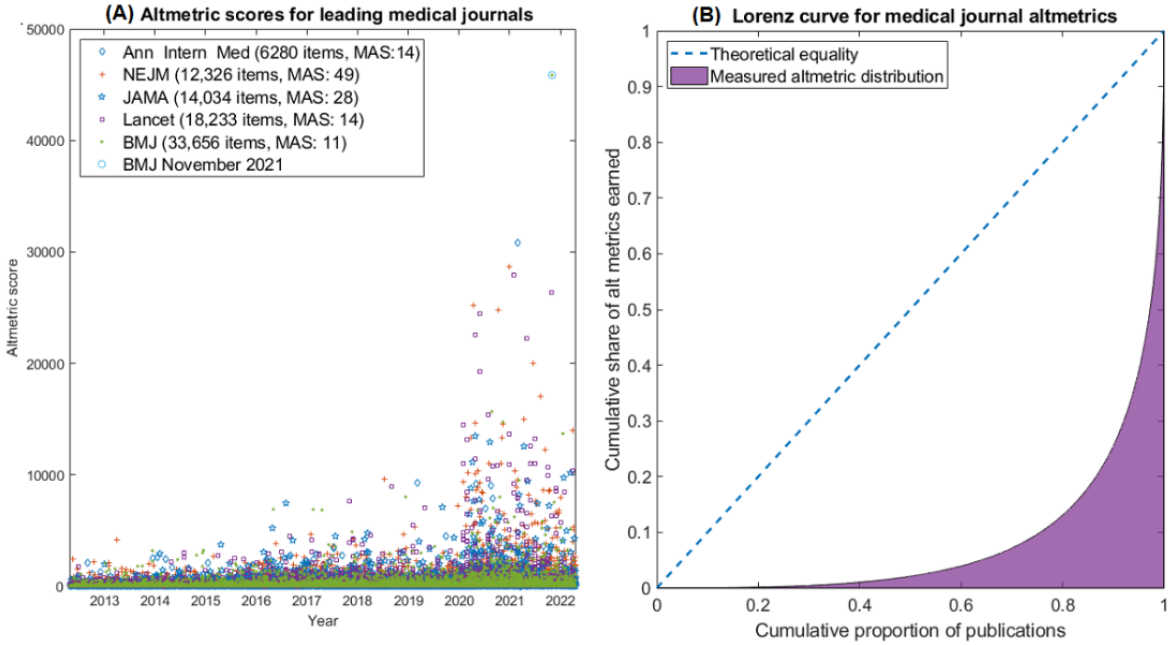

Table 3. Altmetric scores for COVID-19 vaccine studies.

Date	Vaccine	Title	Journal	Altmetric score
July 2020	Moderna	An mRNA Vaccine against SARS-CoV-2—preliminary report	*NEJM* [a]	14,325
August 2020	AstraZeneca	Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomized controlled trial	*Lancet*	15,254
December 2020	Pfizer	Safety and efficacy of the BNT162b2 mRNA Covid-19 Vaccine	*NEJM*	29,151
November 2021	Pfizer	Covid-19: Researcher blows the whistle on data integrity issues in Pfizer's vaccine trial	*BMJ* [b]	45,844

[a]NEJM: New England Journal of Medicine.

[b]BMJ: British Medical Journal.

Propagation Analysis (Media Outlets)

Figure 3A depicts the media distribution of the paper from publication to April 26, 2022, where 71 (39.2%) out of all 181 media citations of the paper came from non-news sites. Figure 3B shows the cumulative mentions from the 5 sources that cited the paper multiple times, all of which were non-news media and conspiracy site sources, with explicit classification means of reporting sites given in Multimedia Appendix 1. These 5 outlets accounted for 51 (28.2%) of the 181 total media mentions alone. A nonparametric right-tailed Wilcoxon rank sum test was performed to investigate the hypothesis that illegitimate non-news sources cited the paper with greater frequency than reliable outlets; for both cases where all Russian outlets were individually counted or pooled as a single entity, the result was highly significant (P<.001). This strongly suggests the paper was disproportionally embraced by partisan outlets with documented histories of antivaccine sentiment rather than conventional media.

XSL•FO
RenderX

Figure 3. (A) Quantification of news-related altmetrics for case study from publication to April 26, 2022, inclusive. (B) Sources with the highest frequency of paper mentions. Legend indicates what proportion of total media citations for which these sources account.

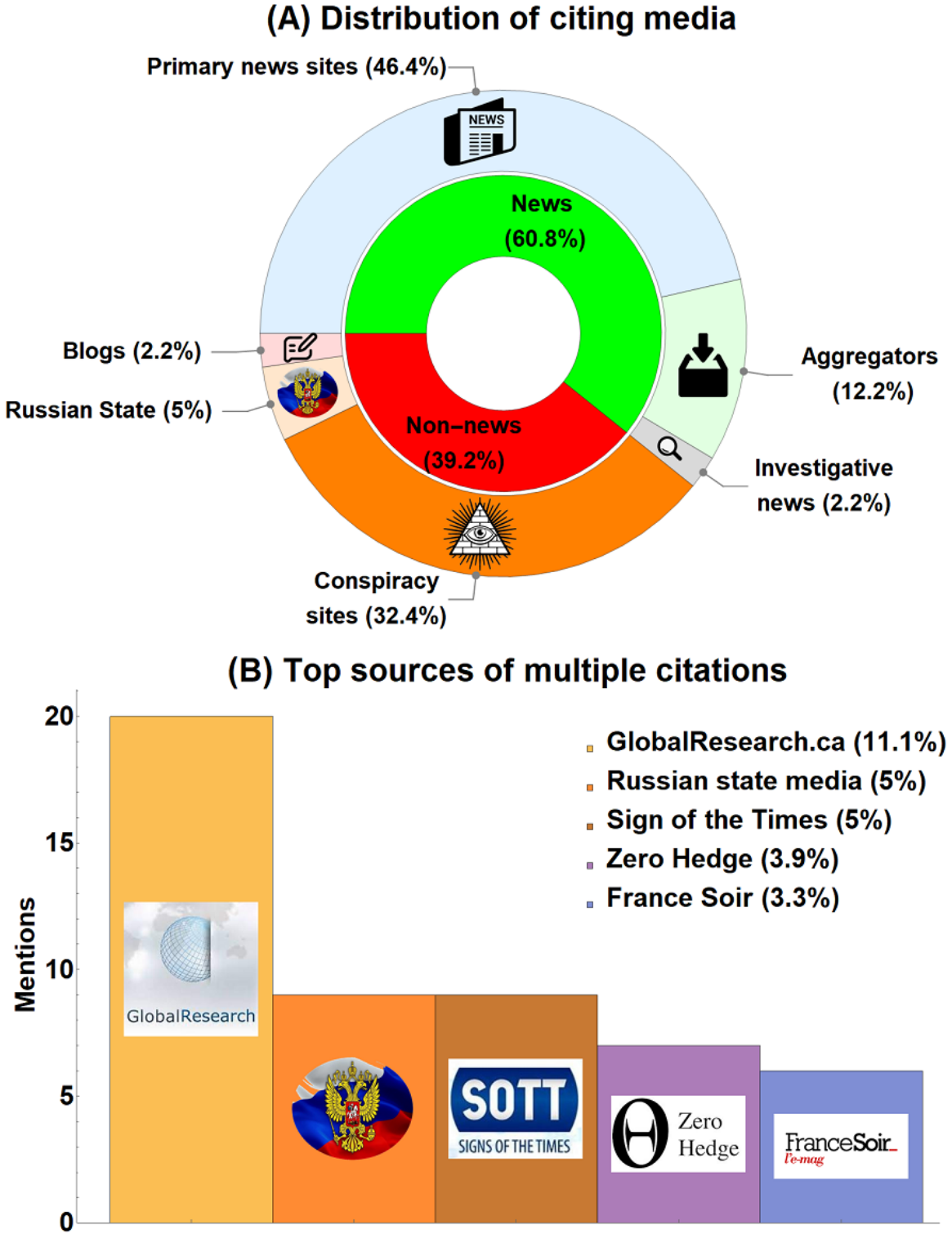

Sentiment Analysis (Social Media)

Our Twitter search strategy yielded 75 tweets from 68 unique users, of which vaccine-negative accounts (52/68, 76%) yielded an upper-bound direct audience of 3,701,281, whereas non-negative accounts (16/68, 24%) had an upper-bound direct audience of 979,198, a 3.78-fold reach difference. Vaccine-negative accounts included the Russian state Sputnik vaccine account (more than 1 million followers, denoted as a vaccine-negative account despite its ostensible purpose due to its disparagement of other vaccines and spreading conspiracy theories about them) and Robert F Kennedy Jr's account (420,900 followers at the time of analysis, widely criticized for vectoring virulent antivaccine propaganda). The non-negative account with the greatest number of followers was the journal's official Twitter account (489,800 followers).

XSL•FO
RenderX

Table 4 depicts the sentiment of high-engagement tweets, distinguishing between shares of the story without caveats or with negative editorialization (including terms like "fraud," "criminal," "corruption," "enemy," "fake," and "scandal") versus those which reported without negative editorialization with caveats or actively rebutted the paper. A Freeman-Halton extension of the Fisher exact test for 2×3 case yields $P<.001$, strongly supporting the hypothesis that vaccine-negative accounts were far more likely to embrace the paper as evidence against COVID-19 vaccination than non-negative accounts. Vaccine-negative accounts were far more likely to share the paper with negative editorialization (odds ratio 58.2, 95% CI 9.7-658.0).

Table 4. Sentiment division on paper.

Preexisting vaccine sentiment	Neutral or caveated presentation	Negative editorialization	Refutation
Non-negative	8	2	6
Negative	5	47	0

Discussion

Principal Findings

The results of our analysis suggest a growing public interest in medical science and highlight a clear instance of how even a dubious finding in medical journals can readily be propagated, buoyed by the veneer of legitimacy that comes from being published in a respected journal. In some respects, this finding is not unexpected. The now infamous case of Andrew Wakefield's fraudulent *Lancet* paper in 1998 led to an enduring confidence crisis over the measles-mumps-rubella vaccine [23], and more recently, dubious claims in medical journals about the efficacy of ivermectin [24] and hydroxychloroquine [25] in treating COVID-19 wrongly led millions to ineffective treatments for the condition; in the case of ivermectin, it even led people to purchase the drug from veterinary suppliers. In these cases, however, the primary information itself was incorrect; malinformation is more nuanced and likely to mislead precisely because the primary information may not be incorrect but selectively presented.

Lewandowsky and Cook [26] proposed a tactic for combatting conspiratorial thinking, misinformation, and disinformation that they characterize as "inoculation" or "prebunking," with two distinct aspects: (1) an "explicit warning of an impending threat of being misled" and (2) "refutation of the misinformation's arguments," noting that fact- and logic-based inoculations have been successful in prebunking a 9/11 conspiracy theory [27]. Malinformation is more difficult to inoculate against because it is more difficult to predict how legitimate stories, studies, and information will be weaponized as malinformation than it is to delineate the general characteristics of misinformation and disinformation, particularly when based on conspiratorial thinking. However, we would counter that it is useful to publicly document instances of malinformation to elucidate why they constitute malinformation, as once articulated it can be applied to other stories of this sort.

How journals contribute to malinformation remains poorly understood, but editors need to be aware of how legitimate (or seemingly legitimate) stories and studies can be turned into malinformation. Accordingly, we suggest that in the age of social media, it is imperative that peer review, editorial oversight, and fact-checking go beyond just the issue of whether individual facts and findings reported in a paper are accurate and consider the overall context (or lack thereof) in which the "facts" are arranged to produce a narrative, a point expanded upon in Multimedia Appendix 1. There is increasing recognition that much of what is published in medical journals may not be reproducible and that biased, or outright fabricated results, taint biomedical literature [28-30], an alarmingly common occurrence [31]. Inevitably, dangerous falsehoods and polemics can sometimes appear in literature. It is vital that reviewers and editors are aware of these issues and skeptically evaluate potentially dubious claims. While some guidelines exist [32], wider awareness is urgently required.

In the pre–social media era when medical journals were read predominantly by clinicians and scientists, the failure of peer reviewers and editors to identify malinformation like this story would likely have had only a minor impact on the public perception of vaccine safety and efficacy, with studies like Andrew Wakefield's fraudulent research in *The Lancet* being very exceptional cases that reduced vaccine confidence [23,33]. Unfortunately, the situation has changed markedly in the social media era. Now, the inability on the part of peer reviewers and editors to recognize malinformation before it is published under the imprimatur of a respected medical journal can have major repercussions because, as demonstrated in this analysis, such malinformation can be amplified widely.

It is crucial that scientific publications strive to communicate vital findings while also being aware that the audience for scientific work now extends far beyond the confines of academia. It is thus inevitable that highly uncertain or nuanced findings in medicoscientific journals will be stripped of context by malicious actors seeking to undermine public health and understanding. While findings can and should be debated by scientists, it is very naive to ignore the reality that these specialized debates will also transpire in the public sphere, driven by the characteristic context collapse of social media. This new normal has ramifications for how scientific publishers and scientists frame their work and their responsibilities after publication. Journals and authors need also be mindful of how their work is interpreted after publication, including issuing clarifications and correctives to address situations in which their publications are weaponized as malinformation.

Potentially controversial findings, when they arise, should not be censored, but rigorously evaluated and reported in proper context, ideally including independent lay summaries to stave off potential distortion in the public arena. If journal editors and peer reviewers are not trained to recognize potential

XSL•FO
RenderX

malinformation when they see it, peer-reviewed biomedical journals can all too easily inadvertently become powerful sources and amplifiers of malinformation, rather than what they should be—bulwarks against it.

Limitations

While our analysis showed that antivaccine sources were most likely to propagate the offending piece, it was not possible to quantify the proportion of receptive audiences who already had vaccine-negative attitudes and how much it might remain within an already committed echo chamber. Even so, there is already ample evidence that mere exposure to antivaccine sentiment induces vaccine hesitancy [6]. The volume of vaccine disinformation available online and offline creates a lingering public perception that vaccines might be harmful due to the psychological phenomena of the illusory truth effect, the observation that a statement is more likely to be deemed true after repeated exposure, regardless of intrinsic veracity [34]. As fiction is spread more efficiently across social media than corrective information [35], it is reasonable to conclude that the high publicity this piece garnered, reflected in its staggering altmetric score, was likely to induce novel vaccine hesitancy and undermine public health.

There is also some unavoidable ambiguity in defining the precise border of what constitutes malinformation. In scientific publications, erroneous claims that are perpetuated under usual circumstances might be deemed misinformation. However, as outlined in the potential mechanisms of malinformation in this work, the marshaling of claims produces a misleading and harmful picture without adequate context and is typically better described as malinformation. Moreover, even completely accurate works can be weaponized by vested interests and used as malinformation, despite no error or intention on the authors' part. There can also be synergy between the 2 mechanisms; dubious claims can be adopted and championed by bad-faith actors, amplifying harmful effects. The results of this analysis suggest that this facet may have been evident in this instance, but it is important to recognize that such distinctions will be context-specific in many cases.

There have been many hundreds of papers on COVID-19 misinformation and disinformation since the pandemic began, and only a small number have been concentrated on altmetrics of misinformation, tracing, for example, engagement with retracted papers [36]. Potential sources of confusion with the altmetric data have included in particular numerous changes to Twitter (now X) since 2022 that have reduced transparency and an exodus of users, which could have caused altmetrics to change with time. This analysis circumvents these issues by virtue of having been performed prior to these drastic access changes, with tweets included (Multimedia Appendix 1), but will likely make future approaches using this methodology limited in scope. The use of manual sentiment sorting was time-intensive, and future iterations might take a natural language processing approach or use large language model–based artificial intelligence, but the by-hand verification of the relatively small number of high engagement tweets on the subject combined with the analysis of users' prior vaccine positions should make conclusion in this work relatively robust.

Conclusions

Vaccine malinformation was far more likely to be shared with negative editorialization by vaccine-negative accounts and repeatedly cited by media sources opposed to vaccination. This work highlights how malinformation is an issue that should be considered in scientific communication in the social media era to avoid the propagation of misleading and harmful narratives.

Acknowledgments

The authors would like to thank Altmetric for allowing access to the discovery platform so that the data for this cross-sectional analysis could be compiled.

Data Availability

All raw tweet data analyzed during this study are included in this published paper and its supplementary information files in redacted form. The altmetric dataset and full tweet history analyzed and the raw code generated during this study are not publicly available due to potentially identifying information but are available from the corresponding author on reasonable request.

Authors' Contributions

DRG and DHG were responsible for the concept, writing, and editing. DRG undertook data curation, methodology, and formal analysis. Both authors composed, edited, and reviewed the final manuscript.

Conflicts of Interest

DRG is a popular science author who contributes to numerous outlets worldwide on health disinformation and vaccine hesitancy, occasionally receiving fees for these pieces. He also receives royalties from his books on critical thinking (multiple editions) and reports receiving Wellcome Trust research funding (grant 214461/A/18/Z). He has also been a public advocate for the human papillomavirus vaccine in Ireland and Europe, receiving zero industry funding for this advocacy. DHG is a long-time editor and author on Science-Based Medicine, a website for discussions of medicine and pseudoscience, which has been critical of the British Medical Journal's reporting on this issue. Neither author is in receipt of any funding whatsoever from vaccine manufacturers, nor have they been financially compensated by the pharmaceutical industry for any of this work.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary material and redacted tweets.
[PDF File (Adobe PDF File), 10738 KB - jmir_v27i1e64679_app1.pdf]

References

1. Hellmuth J, Rabinovici GD, Miller BL. The rise of pseudomedicine for dementia and brain health. JAMA 2019;321(6):543-544 [FREE Full text] [doi: 10.1001/jama.2018.21560] [Medline: 30681701]
2. Hill JA, Agewall S, Baranchuk A, Booz GW, Borer JS, Camici PG, et al. Medical misinformation: Vet the message!. J Am Heart Assoc 2019;8(3):e011838 [FREE Full text] [doi: 10.1161/JAHA.118.011838] [Medline: 30686129]
3. Grimes D. The struggle against cancer misinformation. Cancer Discov 2022;12(1):26-30. [doi: 10.1158/2159-8290.CD-21-1468] [Medline: 34930788]
4. Grimes DR, O'Riordan E. Starving cancer and other dangerous dietary misconceptions. Lancet Oncol 2023;24(11):1177-1178. [doi: 10.1016/S1470-2045(23)00483-7] [Medline: 37922928]
5. Andre FE, Booy R, Bock HL, Clemens J, Datta SK, John TJ, et al. Vaccination greatly reduces disease, disability, death and inequity worldwide. Bull World Health Organ 2008;86(2):140-146 [FREE Full text] [doi: 10.2471/blt.07.040089] [Medline: 18297169]
6. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One 2014;9(2):e89177 [FREE Full text] [doi: 10.1371/journal.pone.0089177] [Medline: 24586574]
7. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
8. Kennedy J. Vaccine hesitancy: a growing concern. Paediatr Drugs 2020;22(2):105-111. [doi: 10.1007/s40272-020-00385-4] [Medline: 32072472]
9. Chou WYS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
10. Armstrong PW, Naylor CD. Counteracting health misinformation: A role for medical journals? JAMA 2019;321(19):1863-1864. [doi: 10.1001/jama.2019.5168] [Medline: 31009036]
11. Merchant RM, Asch DA. Protecting the value of medical science in the age of social media and "fake news". JAMA 2018;320(23):2415-2416. [doi: 10.1001/jama.2018.18416] [Medline: 30453315]
12. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. JAMA 2020;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
13. Toth G, Spiotta AM, Hirsch JA, Fiorella D. Misinformation in the COVID-19 era. J Neurointerv Surg 2020;12(9):829-830 [FREE Full text] [doi: 10.1136/neurintsurg-2020-016683] [Medline: 32801141]
14. Wardle C, Derakhshan H. Information disorder: Toward an interdisciplinary framework for research and policymaking. A Sober Look. 2017. URL: https://tinyurl.com/bdfnv5pd [accessed 2025-02-21]
15. Bornmann L. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. J Inform 2014;8(4):895-903. [doi: 10.1016/j.joi.2014.09.005]
16. Gehanno JF, Ladner J, Rollin L, Dahamna B, Darmoni SJ. How are the different specialties represented in the major journals in general medicine? BMC Med Inform Decis Mak 2011;11:3 [FREE Full text] [doi: 10.1186/1472-6947-11-3] [Medline: 21255439]
17. Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. J Clin Epidemiol 2019;107:36-41 [FREE Full text] [doi: 10.1016/j.jclinepi.2018.11.014] [Medline: 30458264]
18. Gehanno JF, Grosjean J, Darmoni SJ, Rollin L. Reliability of citations of medRxiv preprints in articles published on COVID-19 in the world leading medical journals. PLoS One 2022;17(8):e0264661 [FREE Full text] [doi: 10.1371/journal.pone.0264661] [Medline: 35947594]
19. Thacker PD. Covid-19: Researcher blows the whistle on data integrity issues in Pfizer's vaccine trial. BMJ 2021;375:n2635. [doi: 10.1136/bmj.n2635] [Medline: 34728500]
20. Miller D. Fact check: The British Medical Journal did not reveal disqualifying and ignored reports of flaws in Pfizer COVID-19 vaccine trials. Lead Stories, 2021. Retrieved 28 May 2021, from Lead stories [FREE Full text]
21. Clark C. Experts blow whistle on alleged COVID vaccine whistleblower claims—allegations described as "vague kind of hand waving". MedPage Today. 2022. URL: https://www.medpagetoday.com/special-reports/exclusives/95484 [accessed 2025-02-21]
22. Gorski D. What the heck happened to The BMJ? Science-Based Medicine. 2021. URL: https://sciencebasedmedicine.org/what-the-heck-happened-to-the-bmj/ [accessed 2025-02-21]
23. Motta M, Stecula D. Quantifying the effect of Wakefield et al. (1998) on skepticism about MMR vaccine safety in the U.S. PLoS One 2021;16(8):e0256395 [FREE Full text] [doi: 10.1371/journal.pone.0256395] [Medline: 34411172]
24. Popp M, Stegemann M, Metzendorf MI, Gould S, Kranke P, Meybohm P, et al. Ivermectin for preventing and treating COVID-19. Cochrane Database Syst Rev 2021;7(7):CD015017 [FREE Full text] [doi: 10.1002/14651858.CD015017.pub2] [Medline: 34318930]

XSL•FO
RenderX

25. Perlis RH, Lunz Trujillo K, Green J, Safarpour A, Druckman JN, Santillana M, et al. Misinformation, trust, and use of ivermectin and hydroxychloroquine for COVID-19. JAMA Health Forum 2023;4(9):e233257 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.3257] [Medline: 37773507]
26. Lewandowsky S, Cook J. The Conspiracy Theory Handbook. Fairfax, VA: Center for Climate Change Communication; 2020.
27. Banas JA, Miller G. Inducing resistance to conspiracy theory propaganda: testing inoculation and metainoculation strategies. Hum Commun Res 2013;39(2):184-207. [doi: 10.1111/hcre.12000]
28. Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2(8):e124 [FREE Full text] [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
29. Bik EM, Casadevall A, Fang FC. The prevalence of inappropriate image duplication in biomedical research publications. mBio 2016;7(3):e00809-e00816 [FREE Full text] [doi: 10.1128/mBio.00809-16] [Medline: 27273827]
30. Grimes DR. Is biomedical research self-correcting? Modelling insights on the persistence of spurious science. R Soc Open Sci 2024;11(1):231056 [FREE Full text] [doi: 10.1098/rsos.231056] [Medline: 38298396]
31. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One 2009;4(5):e5738 [FREE Full text] [doi: 10.1371/journal.pone.0005738] [Medline: 19478950]
32. Grimes DR, Bishop DVM. Distinguishing polemic from commentary in science: some guidelines illustrated with the case of Sage and Burgio (2017). Child Dev 2018;89(1):141-147. [doi: 10.1111/cdev.13013] [Medline: 29266222]
33. Grimes DR. Balancing benefits and potential risks of vaccination: the precautionary principle and the law of unintended consequences. BMJ Evid Based Med 2022;27(6):319-323. [doi: 10.1136/bmjebm-2021-111773] [Medline: 34933928]
34. Udry J, Barber S. The illusory truth effect: A review of how repetition increases belief in misinformation. Curr Opin Psychol 2024 Apr;56:101736. [doi: 10.1016/j.copsyc.2023.101736] [Medline: 38113667]
35. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
36. Shamsi A, Lund BD, SeyyedHosseini S. Sharing of retracted COVID-19 articles: an altmetric study. J Med Libr Assoc 2022;110(1):97-102 [FREE Full text] [doi: 10.5195/jmla.2022.1269] [Medline: 35210968]

Edited by A Coristine; submitted 23.07.24; peer-reviewed by Y Jiang, Z Ehtesham; comments to author 04.11.24; revised version received 04.12.24; accepted 19.12.24; published 21.03.25.

Please cite as:
Grimes DR, Gorski DH
Quantifying Public Engagement With Science and Malinformation on COVID-19 Vaccines: Cross-Sectional Study
J Med Internet Res 2025;27:e64679
URL: https://www.jmir.org/2025/1/e64679
doi:10.2196/64679
PMID:

©David Robert Grimes, David H Gorski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Subtyping Social Determinants of Health in the "All of Us" Program: Network Analysis and Visualization Study

Suresh K Bhavnani[1], MArch, PhD; Weibin Zhang[1], PhD; Daniel Bao[2], MD; Mukaila Raji[3], MSc, MD; Veronica Ajewole[4], PharmD, BCOP; Rodney Hunter[4], PharmD; Yong-Fang Kuo[1], PhD; Susanne Schmidt[5], PhD; Monique R Pappadis[6], PhD, MEd; Elise Smith[7], PhD; Alex Bokov[5], PhD; Timothy Reistetter[8], PhD; Shyam Visweswaran[9], MD, PhD; Brian Downer[6], PhD

[1]School of Public and Population Health, Department of Biostatistics & Data Science, University of Texas Medical Branch, Galveston, TX, United States

[2]Department of Radiology, Houston Methodist, Houston, TX, United States

[3]Department of Internal Medicine, Division of Geriatrics Medicine, University of Texas Medical Branch, Galveston, TX, United States

[4]College of Pharmacy and Health Sciences, Department of Pharmacy Practice, Texas Southern University, Houston, TX, United States

[5]Department of Population Health Sciences, University of Texas Health San Antonio, San Antonio, TX, United States

[6]School of Public and Population Health, Department of Population Health & Health Disparities, University of Texas Medical Branch, Galveston, TX, United States

[7]School of Public and Population Health, Department of Bioethics & Health Humanities, University of Texas Medical Branch, Galveston, TX, United States

[8]School of Health Professions, Department of Occupational Therapy, University of Texas Health San Antonio, San Antonio, TX, United States

[9]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Suresh K Bhavnani, MArch, PhD
School of Public and Population Health
Department of Biostatistics & Data Science
University of Texas Medical Branch
301 University Boulevard
Galveston, TX, 77555
United States
Phone: 1 (734) 772 1929
Email: subhavna@utmb.edu

Abstract

Background: Social determinants of health (SDoH), such as financial resources and housing stability, account for between 30% and 55% of people's health outcomes. While many studies have identified strong associations between specific SDoH and health outcomes, little is known about how SDoH co-occur to form subtypes critical for designing targeted interventions. Such analysis has only now become possible through the *All of Us* program.

Objective: This study aims to analyze the *All of Us* dataset for addressing two research questions: (1) What are the range of and responses to survey questions related to SDoH? and (2) How do SDoH co-occur to form subtypes, and what are their risks for adverse health outcomes?

Methods: For question 1, an expert panel analyzed the range of and responses to SDoH questions across 6 surveys in the full *All of Us* dataset (N=372,397; version 6). For question 2, due to systematic missingness and uneven granularity of questions across the surveys, we selected all participants with valid and complete SDoH data and used inverse probability weighting to adjust their imbalance in demographics. Next, an expert panel grouped the SDoH questions into SDoH factors to enable more consistent granularity. To identify the subtypes, we used bipartite modularity maximization for identifying SDoH biclusters and measured their significance and replicability. Next, we measured their association with 3 outcomes (depression, delayed medical care, and emergency room visits in the last year). Finally, the expert panel inferred the subtype labels, potential mechanisms, and targeted interventions.

Results: The question 1 analysis identified 110 SDoH questions across 4 surveys covering all 5 domains in *Healthy People 2030*. As the SDoH questions varied in granularity, they were categorized by an expert panel into 18 SDoH factors. The question

XSL•FO
RenderX

2 analysis (n=12,913; *d*=18) identified 4 biclusters with significant biclusteredness (*Q*=0.13; random-*Q*=0.11; *z*=7.5; *P*<.001) and significant replication (real Rand index=0.88; random Rand index=0.62; *P*<.001). Each subtype had significant associations with specific outcomes and had meaningful interpretations and potential targeted interventions. For example, the *Socioeconomic barriers* subtype included 6 SDoH factors (eg, *not employed* and *food insecurity*) and had a significantly higher odds ratio (4.2, 95% CI 3.5-5.1; *P*<.001) for depression when compared to other subtypes. The expert panel inferred implications of the results for designing interventions and health care policies based on SDoH subtypes.

Conclusions: This study identified SDoH subtypes that had statistically significant biclusteredness and replicability, each of which had significant associations with specific adverse health outcomes and with translational implications for targeted SDoH interventions and health care policies. However, the high degree of systematic missingness requires repeating the analysis as the data become more complete by using our generalizable and scalable machine learning code available on the *All of Us* workbench.

(J Med Internet Res 2025;27:e48775) doi:10.2196/48775

KEYWORDS

social determinants of health; All of Us; bipartite networks; financial resources; health care; health outcomes; precision medicine; decision support; health industry; clinical implications; machine learning methods

Introduction

Background

Social determinants of health (SDoH), such as financial resources [1] and housing stability [2], account for between 30% and 55% of people's health outcomes [3]. While many studies have identified strong associations between specific SDoH and health outcomes, most people experience multiple SDoH concurrently in their daily lives [4-8]. For example, limited access to education, unstable employment, and lack of access to health care tend to frequently co-occur across individuals, leading to long-term stress and depression [8]. Such complex interactions among multiple SDoH make it critical to analyze combinations of SDoH versus single factors. However, analysis of such co-occurrences and their risks of adverse health outcomes requires the integration of personal, clinical, social, and environmental information, critical for designing cost-effective and targeted interventions. Unfortunately, the lack of databases containing such multiple datatypes from the same individuals has resulted in a fragmented understanding of how SDoH co-occur and impact health, which is critical for designing targeted interventions.

The *All of Us* program [9-11] provides an unprecedented opportunity to address this fragmented view of SDoH. This program aims to collect data from multiple sources related to 1 million or more individuals, with a focus on populations that have been traditionally underrepresented in biomedical research. These data sources include electronic health records (EHRs), health surveys, whole-sequence genome data, physical measurements, and personal digital information. Critically, *All of Us* provides several survey modules containing a wide range of SDoH-related questions, which, in combination with other data sources, could transform our understanding of high-risk combinations of SDoH [9].

However, little is known about the range of and responses to SDoH questions in *All of Us* and how they co-occur to form subtypes, which are critical for designing targeted interventions. To address these gaps, we characterized 110 SDoH in *All of Us*, which guided the methods we used to analyze how they co-occur to form subtypes and their risk of adverse health outcomes. The results helped highlight the opportunities and challenges for conducting subtype analysis in *All of Us*, which integrates multiple datatypes through the use of scalable and generalizable machine learning methods aimed at designing targeted interventions.

Models and Research Related to SDoH

The World Health Organization defines SDoH as the "non-medical factors that influence health outcomes" [3]. Specifically, these include the conditions in which people are born, grow, work, live, and age. Furthermore, such conditions are shaped by a wider set of forces, such as economic and social policies, and systems, such as discriminatory laws and structural racism.

Several models have proposed the factors and mechanisms involved in SDoH [4,12]. These models were motivated by the concept of *social gradient* [13], an empirical phenomenon observed within and across nations [14,15] consistently showing that the lower an individual's socioeconomic position, the worse their health. To help explain the factors underlying the social gradient, the model by Dahlgren and Whitehead [4,16] proposed several interconnected layers of social determinants that influence health. As shown in Figure 1 [4,16], the innermost layer comprises demographic and genetic factors, which are largely unmodifiable. In contrast, the outer layers are modifiable to different degrees, such as lifestyle (eg, exercise and smoking); social and community networks (eg, contact with supportive friends and family); living and working conditions (eg, access to health care and employment); and broader socioeconomic, cultural, and environmental conditions (eg, crime in the neighborhood). While this model was not intended to provide explicit testable hypotheses [4], the factors within each layer are expected to co-occur and impact each other in addition to responding to external forces such as systemic racism and capitalism when it is focused on financial profits at the expense of societal benefits.

XSL•FO
RenderX

Figure 1. The Dahlgren-Whitehead conceptual model aimed at visually showing the interrelated layers of social determinant of health domains that influence health (reproduced from Dahlgren and Whitehead, 2021).

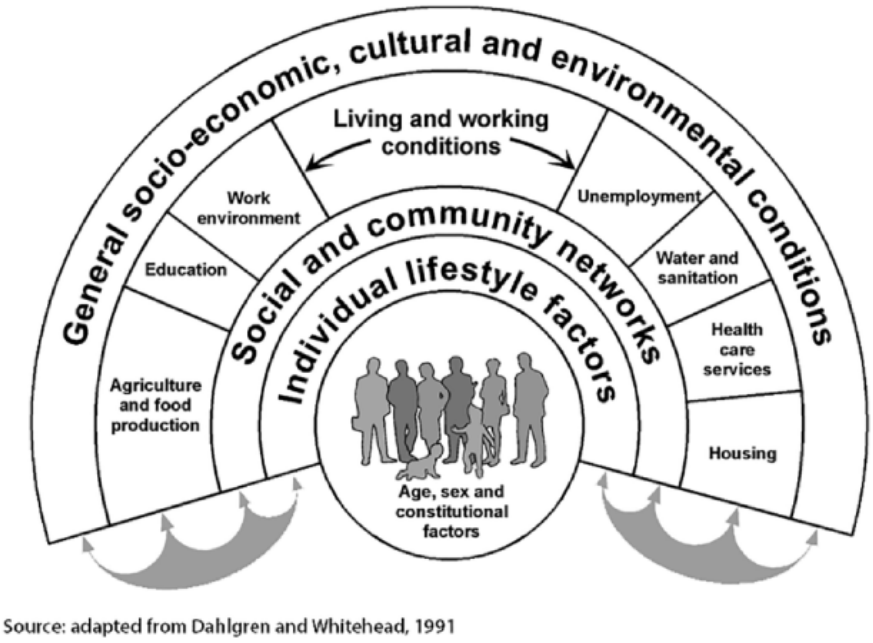

These early SDoH models motivated numerous studies [17] that analyzed associations among specific SDoH (eg, immigration status and home density [7]), their association with health outcomes (eg, education and mortality [18]), and how they manifest within subpopulations (eg, patients with diabetes [19]). More recently, organizations such as the Centers for Disease Control and Prevention and *Healthy People 2030* (HP-30) have classified these empirical results into SDoH domains that roughly map to the Dahlgren-Whitehead model. For example, HP-30 organized SDoH empirical studies into five SDoH domains: (1) economic stability, (2) education access and quality, (3) health care access and quality, (4) neighborhood and built environment, and (5) social and community context. Furthermore, the PhenX program (which provides well-established measurement protocols for use in biomedical and translational research) has identified SDoH data collection protocols to enable more systematic data collection and analysis [20-22].

While the aforementioned findings and categorizations have greatly improved our understanding of SDoH and their impact on health, they have been mostly analyzed based on snapshots of associations between a few factors and health outcomes. In contrast, SDoH models and recent empirical studies suggest that multiple SDoH tend to co-occur and impact each other. For example, during the pandemic, Hispanic and Black or African American individuals not only had a higher exposure to COVID-19 due to their frontline jobs and overcrowded living conditions but also had a higher risk of serious infections due to previous health conditions not addressed due to lack of health care access [4]. Similarly, undocumented immigrants with lower incomes living in neighborhoods with high pollution, combined with the stress of deportation, have an increased risk of multiple chronic conditions such as depression and lung cancer [7]. Such studies have resulted in the Centers for Medicare and Medicaid Services emphasizing that SDoH are a multilevel construct that includes both individual and contextual factors that have complex interactions [23]. Furthermore, a distinction has been made between the aforementioned SDoH at the community level and social needs at the individual level [24]. However, while this distinction is critical when designing care pathways, there is growing consensus [25] that SDoH is an umbrella term that covers both levels, and it is an approach that has been adopted by both the PhenX Toolkit [21] and the *All of Us* surveys on SDoH widely used for analysis [9-11]. Therefore, we use *SDoH* to include all levels of nonmedical factors that impact health and well-being.

The aforementioned co-occurrences of multiple SDoH and their impact on health directly reflect the interconnected layers of the Dahlgren-Whitehead model shown in Figure 1. However, analysis of such co-occurrences and their health outcomes requires large datasets with multiple datatypes that have only recently been made available through the *All of Us* program.

All of Us: Multiple Datatypes Across a Large Cohort of Underrepresented Americans

The *All of Us* research program [9-11] (*All of Us*), funded by the National Institutes of Health since 2015, aims to accelerate biomedical research to enable discoveries leading to individualized and equitable prevention and treatment. Such research is currently hampered due to the *limited range* of

XSL•FO
RenderX

personal, clinical, social, and environmental variables available for the same individuals; *limited representation* in research datasets of socially marginalized populations; and *limited access* to individual-level data due to privacy laws.

To overcome these hurdles, *All of Us* provides three critical features: (1) a data repository that is projected to contain 1 million or more participants with data from multiple sources, including EHRs, health surveys, whole-sequence genomic data, physical measurements, and personal digital information such as from Fitbit trackers; (2) a cohort targeted to include 75% of participants from populations underrepresented in research (race, ethnicity, gender, sex, sexual orientation, and disability) oversampled from the US population; and (3) strictly enforced rules to prevent reidentification of participants by disallowing the download of any participant data or reporting of research results for subgroups of <20. These rules allow for analysis of the *All of Us* data to be categorized as non–human subjects research, which, combined with training and personal authentication by researchers, has resulted in a substantial reduction in administrative hurdles.

As of December 30, 2022 (Controlled Tier; version 6), *All of Us* contained 372,397 total participants, with 8.6% who had attempted all 9 health surveys (7 related to demographics and general health and 2 related to COVID-19) and 26.5% who had genomic data. Critical to this study is the recent addition of a survey specifically targeted to SDoH questions, which has been attempted by 15.5% in the *All of Us* cohort. A preliminary analysis revealed that SDoH appear to be distributed across multiple health surveys and EHR codes, with participants providing those data at different times on a rolling basis. However, little is known about the range of and responses to SDoH questions in *All of Us* and how SDoH co-occur to form subtypes, a critical step for selecting the methods to identify and interpret SDoH subtypes.

Computational Methods to Identify and Interpret Subtypes

A wide range of studies [26-34] on topics ranging from molecular to environmental determinants of health have shown that most humans tend to share a subset of characteristics (eg, comorbidities, symptoms, and genetic variants), forming distinct subtypes (also referred to as *subgroups* or *subphenotypes* depending on the condition and variables analyzed). A primary goal of precision medicine is to identify such subtypes and infer their underlying disease processes to design interventions targeted to those processes [27,35]. Methods to identify subtypes include (1) investigator-selected variables such as race for developing hierarchical regression models [36] or assigning patients to different arms of a clinical trial, (2) existing classification systems such as the Medicare Severity Diagnosis Related Group [37] to assign patients to a disease category for purposes of billing, and (3) computational methods such as classification [38-40] and clustering [30,41] to discover subtypes.

Several studies have used computational methods to identify subtypes, each with critical trade-offs. Some studies have used *combinatorial* approaches [42] (eg, identify all pairs, all triples, and so on), which are intuitive but which can lead to a combinatorial explosion (eg, enumerating combinations of the 31 Elixhauser comorbidities would lead to 2^{31} or 2,147,483,648 combinations), with most combinations not incorporating the full range of symptoms (eg, the most frequent pair of symptoms ignores what other symptoms exist in the profile of patients with that pair). Other studies have used *unipartite* clustering methods [40,41] (clustering patients or comorbidities but not both together), such as k-means and hierarchical clustering, and dimensionality reduction methods, such as principal component analysis, to help identify clusters of frequently co-occurring comorbidities [42-48]. However, such methods have well-known limitations, including the requirement of inputting user-selected parameters (eg, similarity measures and the number of expected clusters) in addition to the lack of a quantitative measure to describe the quality of the clustering (critical for measuring the statistical significance of the clustering). Furthermore, because these methods are unipartite, there is no agreed-upon method to identify the patient subgroup defined by a cluster of variables, and vice versa.

More recently, bipartite network analysis [49] (see Multimedia Appendix 1 [48,50-55] for additional details) has been used to address the aforementioned limitations by automatically identifying *biclusters* consisting of patients and characteristics simultaneously. This method takes as input any dataset, such as *All of Us* participants and their SDoH, and outputs a quantitative and visual description of biclusters (containing both participant subgroups and their frequently co-occurring SDoH). The quantitative output generates the number and members of the biclusters, in addition to the statistical significance of the biclusteredness [50-52], and the visual output displays the quantitative information of the biclusters through a network visualization [53-55]. Therefore, bipartite network analysis enables (1) the automatic identification of biclusters and the significance of their biclusteredness and (2) the visualization of the biclusters critical for their clinical interpretability. Furthermore, the attributes of participants in a subgroup can be used to measure the subgroup risk of an adverse health outcome, develop classifiers for categorizing a new participant into one or more of the subgroups, and develop a predictive model that uses that subgroup membership for measuring the risk of an adverse health outcome for the classified participant.

However, while several studies [52,56-63] have demonstrated the usefulness of bipartite networks for the identification and clinical interpretation of subgroups, there has been no systematic attempt to identify SDoH subtypes, mainly because of the lack of large cohorts containing a wide coverage of SDoH. The *All of Us* program provides an opportunity to use bipartite networks for the identification and interpretation of SDoH subtypes using a wide range of variables in a large cohort and for analyzing their risk of adverse health outcomes, a critical step in advancing precision medicine.

Methods

Research Questions

Our analysis was guided by two research questions targeting the *All of Us* dataset:

XSL•FO
RenderX

1. What are the range of and responses to survey questions related to SDoH?
2. How do SDoH co-occur to form subtypes, and what are their risks for adverse health outcomes?

Expert Panel

The selection of the research questions, variables, cohort, methods, and results and their interpretation were guided by an expert panel consisting of SDoH researchers with a professional background in applied demography, gerontology, and rehabilitation who worked closely with the machine learning and biostatistics researchers. The overall project and manuscript were examined by an ethicist for bias, stigma, and perpetuation of stereotypes. Therefore, the examination of each step in the project was aligned with the human-centered artificial intelligence approach [64-66].

Data Description

Study Population

For question 1, we analyzed the full *All of Us* cohort (N=372,397) and characterized their responses to all the SDoH questions identified by the expert panel (described in the *Variables* section). For question 2, we analyzed all participants (n=12,913) who had valid responses to the SDoH questions identified in question 1 and used them to identify subtypes and their risks of specific outcomes.

Variables

For question 1, the expert panel was asked to review all 1113 questions across 6 *All of Us* non–COVID-19 health surveys, each of which was attempted once per participant (*The Basics*, *Lifestyle*, *Overall Health*, *Personal/Family Health History*, *Health Care Access and Utilization*, and *SDoH*), and the 2843 Systematized Medical Nomenclature of Medicine (SNOMED) codes related to SDoH [67]. The expert panel arrived at a consensus for the SDoH across the surveys and the SNOMED codes. As the SDoH-related SNOMED codes in the EHRs had very low use (see Multimedia Appendix 2 for a characterization), they were not further analyzed.

For question 2, to identify and analyze the SDoH subtypes, we used the following variables:

1. Independent variables included the SDoH factors identified from question 1.
2. Covariates included 3-digit zip code (to determine whether participants in each subtype came from a state that accepted Medicaid expansion, providing greater access to health insurance) and demographics (eg, age, sex, and race).
3. A total of 3 outcomes were included: *depression*, *delayed medical care*, and *emergency room (ER) visits in the last year*. *Depression* was selected as it is a common health outcome when individuals encounter SDoH in their daily lives, such as long-term stress resulting from racism [68], and the dysregulation of the hypothalamic-pituitary-adrenal axis [69]. Depression was defined as having a positive response to both of the following questions in *The Basics* survey—"Are you still seeing a doctor or health care provider for depression?" and "Has a doctor or health care provider ever told you that you have Depression?"—or having SNOMED codes related to depression in their EHRs (35489007, 36923009, 370143000, 191616006, or 66344007). *Delayed medical care* was selected as it often results from the lack of medical insurance, which can impact the use of medical care when needed, leading to poorer health outcomes [70]. Delayed medical care was defined as having one or more positive responses to 9 survey questions (delayed care due to transportation, rurality, nervousness, work, childcare, copay, older adult care, out-of-pocket costs, and deductible costs) in the *Health Care Access* and *Utilization* survey. *Emergency room (ER) visits in the last year* was selected because lack of medical insurance often results in individuals not seeking early medical care when needed, leading to an exacerbation of conditions precipitating one or more ER visits [71]. As the survey questions that we used for SDoH subtyping were based on outcomes in the previous year, we defined ER visits for a participant as having one or more ER visits (current procedural terminology 99281-99285) 1 year preceding the date when the SDoH survey was completed.

Analytical Approach

Question 1: What Are the Range of and Responses to Survey Questions Related to SDoH?

Identification and Coding of SDoH

To analyze the range and responses to the survey questions, we first characterized all SDoH in *All of Us* at two levels of granularity—(1) SDoH questions based on the surveys used to collect the data and (2) SDoH factors, which were categories of the SDoH questions to form a coarser-grained classification (see Figure 2, which explains SDoH questions, factors, and subtypes).

XSL•FO
RenderX

Figure 2. Examples showing how the social determinants of health (SDoH) questions from the *All of Us* surveys that differed in their levels of granularity were transformed by the expert panel into SDoH factors with uniform granularity to ensure consistency for analysis and interpretation and clustered into SDoH subtypes through machine learning. The SDoH questions and factors were subsequently analyzed for coverage across the 5 Healthy People 2030 domains.

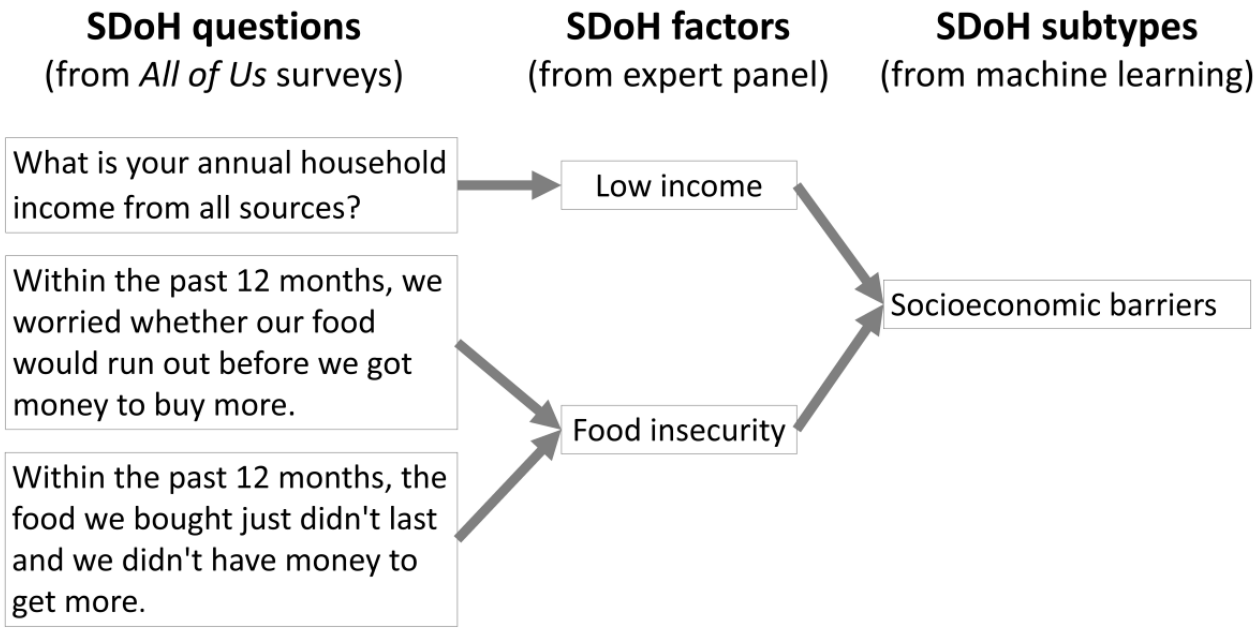

To identify the SDoH questions, members of the expert panel independently used their domain knowledge about SDoH to identify and code the SDoH questions and examine their range with respect to the 5 HP-30 domains using the following steps: (1) reviewed all 1113 questions across 6 health surveys (excluding 2 related to COVID-19) and extracting all SDoH questions that were relevant, (2) transformed all positive or value-free questions into negative phrases and abbreviating them for interpretability in the graphs (eg, *How often do you have someone help you read health-related materials?* was changed to *No one to help read health materials*), (3) reverse coded and dichotomized the abbreviated SDoH questions (eg, always or often=1 and never, occasionally, or sometimes=0), and (4) categorized the SDoH questions into 1 of the 5 HP-30 SDoH domains (economic stability, education access and quality, health care access and quality, neighborhood and built environment, and social and community context). The expert panel subsequently met and collaboratively resolved any differences among their coding schemes to arrive at a consensus (see Multimedia Appendix 3 for the 110 SDoH questions and their consensus coding by the expert panel).

To characterize the SDoH factors, the expert panel arrived at a consensus to categorize one or more of the aforementioned SDoH questions in *All of Us* into SDoH factors and examined their range with respect to HP-30 using the following steps: (1) reviewed the subgrouping labels of questions in the *All of Us* surveys and integrating them to categorize the SDoH into factors, (2) coded a participant as having a "1" for an SDoH factor if they answered one or more of the questions within that factor with a "1," and (3) categorized the SDoH factors into 1 of the 5 HP-30 SDoH domains (economic stability, education access and quality, health care access and quality, neighborhood and built environment, and social and community context; see Multimedia Appendix 3 for the 110 SDoH questions, their consensus coding into 19 SDoH factors, and mapping to the 5 SDoH domains from HP-30).

Analysis of the Range and Responses to SDoH Questions and Factors

The aforementioned knowledge-based classification of SDoH questions and SDoH factors was analyzed to examine their range (with respect to the 5 HP-30 domains) and their responses (across all participants in *All of Us*) using the following four methods: (1) bar graph displaying the number of participants who had valid answers (all responses other than "skip" or "choose not to answer") to each of the SDoH questions, sorted by survey based on mean response, and then sorted by raw response within each survey. Finally, to analyze their range, each bar was colored to denote one of the 5 SDoH domains defined by HP-30; (2) Venn diagram showing how many participants had cross-sectionally valid responses to all identified SDoH questions or factors; (3) table describing the number and proportion of race, ethnicity, sex, gender, and age between those who answered the SDoH questions or factors and those who did not have valid responses; and (4) frequency distribution of the number of SDoH questions or factors across participants who had valid responses for all the SDoH questions. The aforementioned plots are shown in the *Results* section.

Question 2: How Do SDoH Co-Occur to Form Subtypes, and What Are their Risks for Adverse Health Outcomes?

Data

We used the cohort identified in question 1 (participants who had valid answers to all the SDoH questions). However, examination of the SDoH questions revealed that some of them (eg, *cannot afford dental care* and *cannot afford prescriptions*) had a finer level of granularity compared to others (eg, single household). As the questions with a finer level of granularity

XSL•FO
RenderX

tend to be more strongly corelated to each other in comparison to other coarser-grained questions, they also tend to cluster together more strongly, confounding the interpretation of the subtypes. In contrast, as the SDoH factors had a more uniform granularity and were at a level of abstraction that was appropriate to guide referral to the proper social services, we used them to identify the SDoH subtypes.

Analytical Model

Overview

To identify SDoH subtypes, their associations with outcomes and covariates, and their future translation into precision medicine and public policies, we used a 3-part analytical framework called heterogenization, integration, and translation (HIT). As shown in Figure 3, the *heterogenization* step was used to identify the subtypes through the use of bipartite modularity maximization [50-52] (see Multimedia Appendix 1 for more details), the *integration* step was used to measure the association of each subtype with multiple datatypes [72], and the *translation* step was used to qualitatively interpret the subtypes [72] with the goal of developing in the future a decision support system to translate the subtypes into clinical practice and the design of public policies. The following sections describe the specific methods used in each of the HIT steps.

Figure 3. The three steps of the heterogenization, integration, and translation framework to analyze social determinants of health (SDoH): (1) heterogenization of the data to identify subtypes, (2) integration of multiple datatypes such as from electronic health records (eg, depression) and information on state of residency (eg, to determine Medicaid expansion) to determine the risk and enrichment of each subtype, and (3) translation of subtypes through interpretation and predictive modeling with the goal of designing clinical decision support systems and public policy. SLC6A4: Serotonin Transporter Gene Polymorphism.

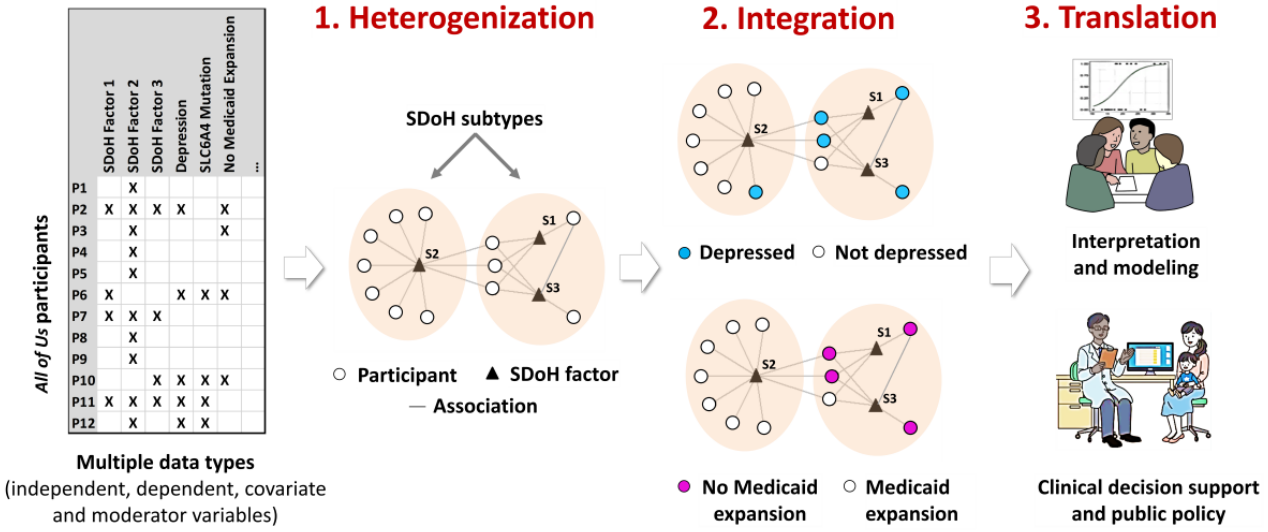

Heterogenization: Identification of Subtypes

As there were many participants who did not have valid answers to the SDoH questions, dropping them resulted in differences in the proportion of demographic variables compared with the full *All of Us* cohort. Therefore, the data needed to be adjusted to better reflect the overall *All of Us* participants. To adjust the demographic distribution of the cohort to match the full *All of Us* cohort, we used inverse probability weighting (IPW) [73,74]. IPW calculates weights to proportionally boost the values of participants who are underrepresented in a cohort with respect to a comparison, such as the full *All of Us* dataset, using a method similar to that of an earlier study on *All of Us* [75] (Multimedia Appendix 4). Next, we multiplied the IPW-generated weights with the original binary values for each participant in our cohort and used the min-max method to range-normalize those weights within each SDoH factor. Finally, to test the replicability of the SDoH factor biclustering, we randomly divided the dataset into a training and a replication dataset.

We identified subtypes in the training dataset and tested the degree to which the SDoH factor co-occurrences were replicated in the replication dataset using the following steps: (1) modeled participants and SDoH factors as a weighted bipartite network (step 1 in Figure 3) where nodes were either participants (circles) or SDoH factors (triangles) and the associations between participant–SDoH factor pairs were weighted edges (lines) generated from IPW (the inclusion of IPW-generated weights enabled the network to represent the demographic distribution of the full *All of Us* dataset), (2) used a bipartite modularity maximization algorithm [50-52] (which takes edge weights into consideration) to identify the number of biclusters and their members and measure the degree of biclusteredness through bicluster modularity (Q, defined as the fraction of edges falling within a cluster minus the expected fraction of such edges in a network of the same size with randomly assigned edges), (3) measured the significance of Q by comparing it to a distribution of the same quantity generated from 1000 random permutations of the network while preserving the network size (number of nodes) and the distribution of weighted edges for each participant, (4) used the Rand index (RI) to measure the degree to which SDoH occurred and did not co-occur in the same cluster in the training and replication datasets, and (5) measured the significance of the RI by comparing it to the mean of a distribution of the same quantity generated by randomly permuting the training and replication datasets 1000 times while

XSL•FO
RenderX

preserving the size of the networks and the distribution of weighted edges for each participant.

Integration: Risk and Enrichment of Subtypes

We used logistic regression to measure the odds ratio (OR) for each subtype compared pairwise to each of the other subtypes for the 3 outcomes (depression, delayed medical care, and ER visits in the last year) and for living in a state with no Medicaid expansion. To adjust for the difference in demographics due to the missingness, we used weights generated from IPW for each participant, and the comparisons were adjusted for demographics (eg, age, sex, and race) and corrected for multiple testing within each outcome using false discovery rate. As 13.07% (1688/12,913) of the participants did not have 3-digit zip code information, we used IPW to measure the weights of the cohort and used them to account for potential sample selection bias.

Translation: Interpretation of Subtypes

The subtype interpretation was done using the following steps: (1) used the *Fruchterman-Reingold* [53] and *ExplodeLayout* [54,55] algorithms to visualize the bipartite network along with the risk of each of the outcomes; (2) asked the expert panel to independently label the subtypes, infer the mechanisms that increase the risks in each subtype for the 3 outcomes (depression, delayed medical care, and ER visits in the last year) with potential strategies to reduce those risks, and then collaboratively come to a consensus; and (3) asked an ethicist to examine the results and their interpretations for bias, stigma, and perpetuation of stereotypes.

Ethical Considerations

The original data collection by the *All of Us* program was approved by an institutional review board as described on the web [76].

The secondary analysis of the *All of Us* data conducted in this work did not receive approval or exemption from an institutional review board. Such an approval or exemption is not required as described on the web [77].

Therefore, the authors had permission to conduct a secondary analysis of the data.

Results

Question 1: What Are the Range of and Responses to Survey Questions Related to SDoH?

Identification and Coding of SDoH Questions and Factors

The expert panel identified 110 questions from 4 surveys (*The Basics*, *Overall Health*, *Healthcare Access and Utilization*, and *SDoH*). Of these 110 questions, 110 (100%) were abbreviated, and 48 (43.6%) were negatively worded and coded (Multimedia Appendix 3). The 110 SDoH questions were further categorized into 19 SDoH factors (one of these was *Delayed medical care*, which was used as an outcome).

Responses to SDoH Questions and Factors

As shown in Figure 4A, the number of valid responses to each of the 110 SDoH questions was largely dictated by the surveys in which the responses were solicited. SDoH from 2 surveys (*The Basics* and *Overall Health*) had the most valid responses (mean 349,434, SD 23,556), followed by *Healthcare Access and Utilization* (mean 149,898, SD 6146) and, finally, the *SDoH* survey (mean 55,960, SD 1083). This pattern of responses matched how answers to each of the surveys were solicited—at enrollment, all participants are required to do *The Basics* and *Overall Health* surveys, and then, on a rolling basis, responses to the other surveys are solicited. The *SDoH* survey was the last survey that was solicited, which explained it having the lowest number of responses. As shown in Figure 4B, this pattern of missingness held for the responses at the SDoH factor level, which was not unexpected as the SDoH factors were aggregations of the SDoH questions. However, as shown in Figures 4A and 4B by the uneven number of valid responses within each survey block, there were several SDoH questions that had invalid responses (“skip” or “choose not to answer”) at both levels of granularity: *The Basics* (339,254/5,655,412, 5.99%), *Healthcare Access and Utilization* (341,516/5,587,957, 6.11%), *Overall Health* (32,669/744,126, 4.39%), and *SDoH* (83,699/3,206,035, 2.61%). Furthermore, the proportion of valid to invalid responses among them was significantly different for the SDoH questions (N=365,237, χ^2_2=57.5; *P*<.001) and for the SDoH factors (N=372,063, χ^2_2=75.6; *P*<.001).

XSL•FO
RenderX

Figure 4. The number of valid responses for (A) 110 social determinants of health (SDoH) questions and (B) 18 SDoH factors. The colors denote how the SDoH in each were categorized based on the 5 Healthy People 2030 (HP-30) domains. Rx: prescription medication.

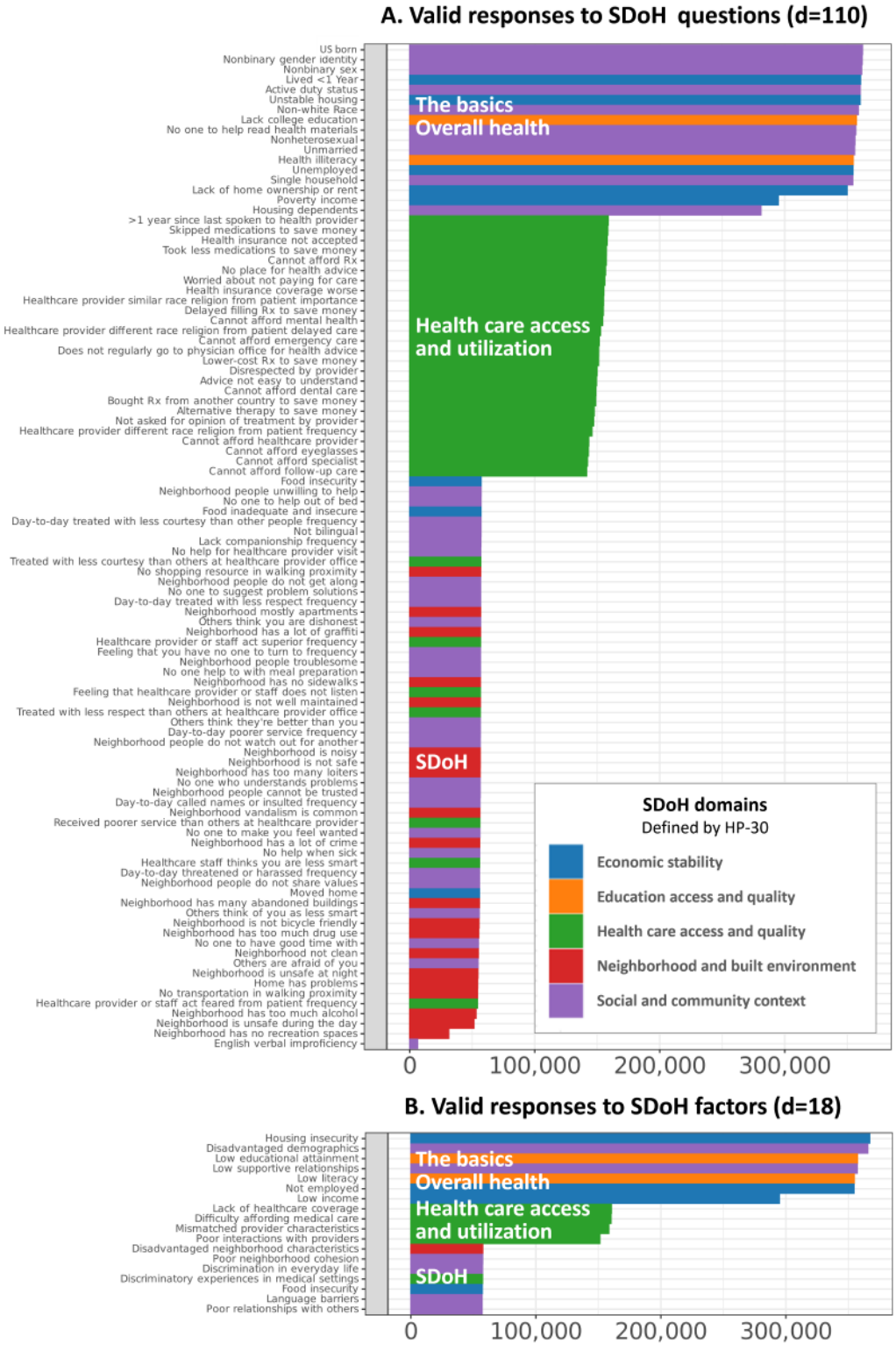

Range of SDoH Questions and Factors

As shown by the colored bars in Figure 4, the surveys spanned the full range of the 5 SDoH HP-30 domains. The SDoH questions in *The Basics* and *Overall Health* surveys were predominantly related to economic stability (blue) and social and community context (purple), those in the *Healthcare Access and Utilization* survey were all related to that topic (green), whereas those from the *SDoH* survey were a mix of all 4 domains. Overall, the 4 surveys contained 110 SDoH questions that together had 100% coverage of the 5 HP-30 domains (social and community context: n=38, 34.5%; neighborhood and built environment: n=19, 17.3%; economic stability: n=10, 9.1%; education access and quality: n=2, 1.8%; health care access and quality: n=42, 38.2%). This characterization suggests that, while the SDoH in *All of Us* have broad domain coverage across the

XSL•FO
RenderX

surveys, analysis of them requires access to all 4 surveys, each of which has different levels of completion and valid responses.

Cohort With Maximized Valid Responses

Given the large degree and systematic nature of missingness in 2 of the 4 surveys, we could not use multiple imputation to estimate the values. Therefore, we had to find a subset of participants who had valid responses to all the SDoH questions. An examination revealed that 2 SDoH questions had <10% of responses (*English verbal frequency*: 6193/371,942, 1.67%; *Neighborhood has no recreation spaces*: 31,152/371,942, 8.38%), accounting for the largest loss in cohort size with valid responses. Therefore, these questions were dropped from further analysis. Furthermore, one question required a branched response (*Living situation* branching to *Did not live in a house*), and these responses were merged. Finally, as we used *Delayed medical care* as an outcome, 8.2% (9/110) of the questions related to that topic were removed, resulting in a total of 98 SDoH questions. As shown in Figure 5, a Venn diagram of the overlap among the valid responses across the surveys revealed that 3.47% (12,913/372,397) of the participants had valid responses to all 98 SDoH questions.

Figure 5. Venn diagram showing 3.47% (12,913/372,397) of participants of the full cohort who had valid responses to all 98 social determinants of health (SDoH) questions.

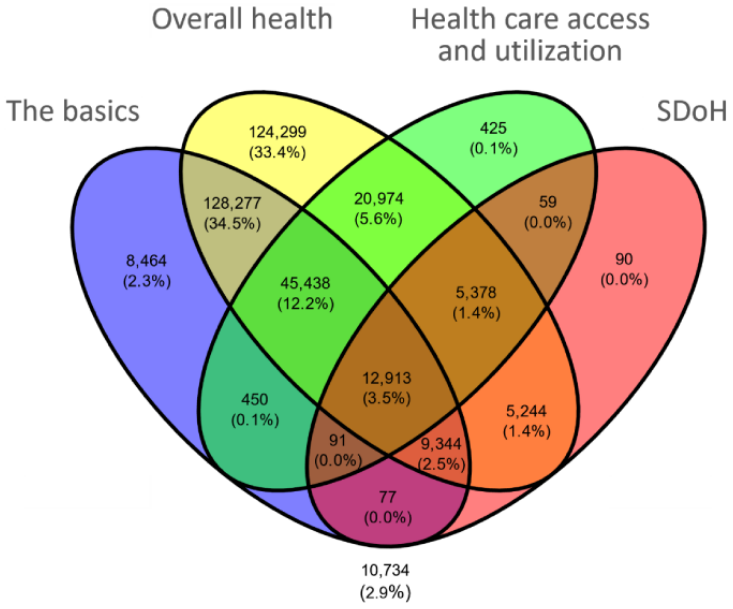

Co-Occurrence of the Number of SDoH Across Responders

As shown in Figure 6, participants had a median of 15 SDoH question co-occurrences and a median of 9 SDoH factor co-occurrences. Furthermore, participants of racial and ethnic minority groups who had valid responses to the 110 SDoH questions had a significantly higher median number of co-occurring SDoH compared to the equivalent White population (median 20 for participants of racial and ethnic minority groups; median 14 for White participants; P<.001). These results show the high co-occurrences of SDoH at both levels of granularity, with a significant difference in median co-occurrences between the White participants and the participants of racial and ethnic minority populations with valid responses.

Figure 6. Frequency distribution of (A) number of co-occurring responses to social determinant of health (SDoH) questions across the 12,913 participants with valid answers to the 98 SDoH questions and (B) number of co-occurring SDoH factors across 19 SDoH factors.

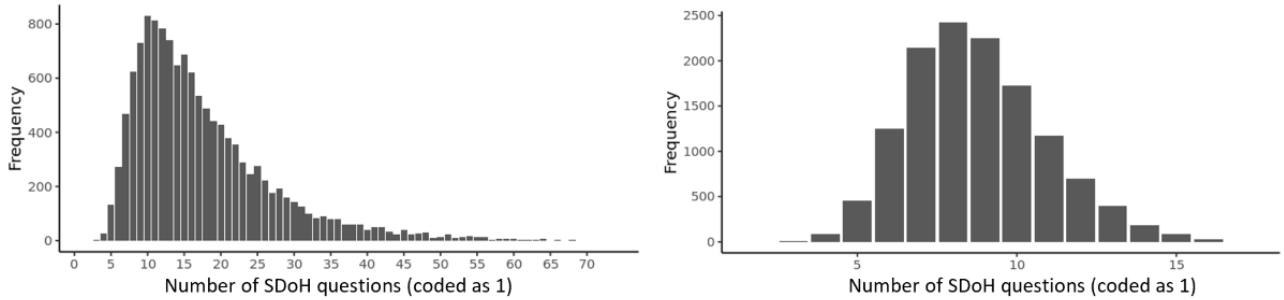

XSL•FO
RenderX

Participant Demographics With Valid Responses to SDoH Questions

As the cohort size dropped to 3.47% (12,913/372,397), we analyzed how that impacted the demographic distribution compared with the overall *All of Us* dataset. As shown in Table 1, there were statistically significant differences in race (N=372,397, χ^2_5=2073.1; *P*<.001) and ethnicity (N=372,397, χ^2_9=6292.2; *P*<.001) between the 2 cohorts after multiple testing correction, with a higher proportion of White participants having valid answers than participants of racial or ethnic minority groups. Furthermore, there was a statistically significant difference in age between the participants who had valid answers and those who did not (*H*1=148.08; *P*<.001). These results show the demographic differences between the cohort with complete and valid answers to the SDoH questions and the full *All of Us* dataset, necessitating the use of IPW–generated weights to address those imbalances, as discussed in question 2.

Table 1. The demographic differences between the total *All of Us* (AoU) participants and those who had valid answers to all 110 social determinant of health (SDoH) questions (N=372,397). Participant counts of <20 are shown as a count of 20 based on the AoU reporting rules.

Demographics	Total AoU participants	Total AoU participants with valid[a] SDoH answers (n=12,913)
Race, n (%)		
Asian	12,459 (3.35)	324 (2.51)
Black or African American	73,383 (19.71)	482 (3.73)
White	201,149 (54.01)	11,279 (87.35)
Other or >1 population	26,890 (7.22)	343 (2.66)
None indicated	58,516 (15.71)	485 (3.76)
Ethnicity, n (%)		
Not Hispanic or Latino	288,227 (77.4)	12,095 (93.67)
Hispanic or Latino	66,704 (17.91)	751 (5.82)
Additional options	17,466 (4.69)	67 (0.52)
Sex at birth, n (%)		
Female	222,495 (59.75)	8236 (63.78)
Male	138,831 (37.28)	4674 (36.2)
Intersex	80 (0.02)	20 (0.15)
Additional options	10,991 (2.95)	20 (0.15)
Gender, n (%)		
Female	220,833 (59.3)	8113 (62.83)
Male	138,140 (37.09)	4642 (35.95)
Nonbinary	920 (0.25)	60 (0.46)
Transgender	464 (0.12)	20 (0.15)
Additional options	12,040 (3.23)	79 (0.61)
Age (y), median (range)	56 (19-122[b])	58 (19-93)

[a]Participants who completed all questions and did not skip or choose not to answer a question.

[b]Age of 122 years=a participant chose the earliest birth year (1900).

Question 2: How Do SDoH Factors Co-Occur to Form Subtypes, and What Are Their Risks for Adverse Health Outcomes?

Overview

The cohort used to identify the subtypes consisted of 12,913 participants, of whom 12,886 (99.79%) had valid IPW–generated weights. The latter cohort was split randomly into the training and replication datasets, each with complete data for 18 SDoH factors (identified in question 1) in addition to the 3 outcomes (depression, delayed medical care, and ER visits in the last year) and covariates (demographics). The results are organized based on the 3 parts of the HIT framework described in Figure 3.

Heterogenization: Identification of Subtypes

The subtypes were identified by using a bipartite network where the edges were weighted using the IPW-generated weights to account for the imbalance in demographics between our cohort and the full *All of Us* dataset. The weighted bipartite network of the training dataset (n=6492) and the 18 SDoH factors revealed 4 biclusters with statistically significant bicluster modularity (*Q*=0.13; random-*Q*=0.11; *z*=7.5; *P*<.001). As shown

XSL•FO
RenderX

in Figure 7, there were 4 clusters with participant subgroups and their most frequently co-occurring SDoH factors (*Cluster 1 [pink]*: low educational attainment, low literacy, low income, not employed, food insecurity, and housing insecurity; *Cluster 2 [green]*: difficulty affording medical care, discriminatory experiences in everyday life, discriminatory experiences in medical settings, and poor interactions with providers; *Cluster 3 [blue]*: poor neighborhood cohesion and poor relationships with others; and *Cluster 4 [gray]*: language barrier, lack of health care coverage, mismatched provider characteristics, disadvantaged neighborhood characteristics, disadvantaged demographics, and low supportive relationships). These co-occurrences of SDoH factors were significantly replicated in the replication dataset (real RI=0.88; random RI=0.62; $P<.001$). As shown in Figure 8, while the 18 SDoH factors have a hierarchical relationship with the 5 *knowledge-driven* HP-30 domains (shown on the left), those same SDoH factors have a more complex relationship with the 4 *data-driven* biclusters (shown on the right).

Figure 7. A total of 4 biclusters in the training dataset consisting of subgroups of participants (n=6492) and their most frequently co-occurring social determinant of health (SDoH) factors (d=18).

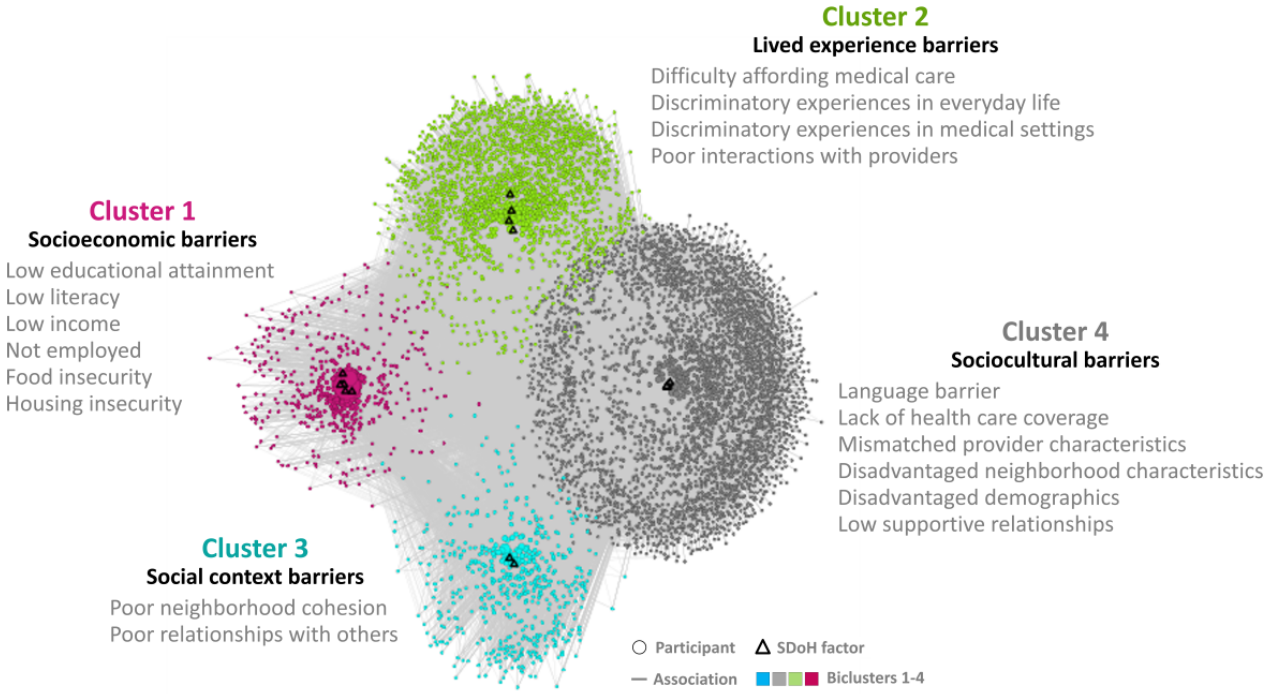

XSL•FO
RenderX

Figure 8. A total of 18 social determinant of health (SDoH) factors (center) have a hierarchical relationship with the 5 SDoH domains defined by Healthy People 2030 (HP-30; left), both of which are knowledge driven. In contrast, the data-driven analysis shows that SDoH factors have a complex relationship with the SDoH subtypes (right) identified through machine learning (ML), reflecting how they co-occur in the real world and aligned with models such as the Dahlgren-Whitehead model.

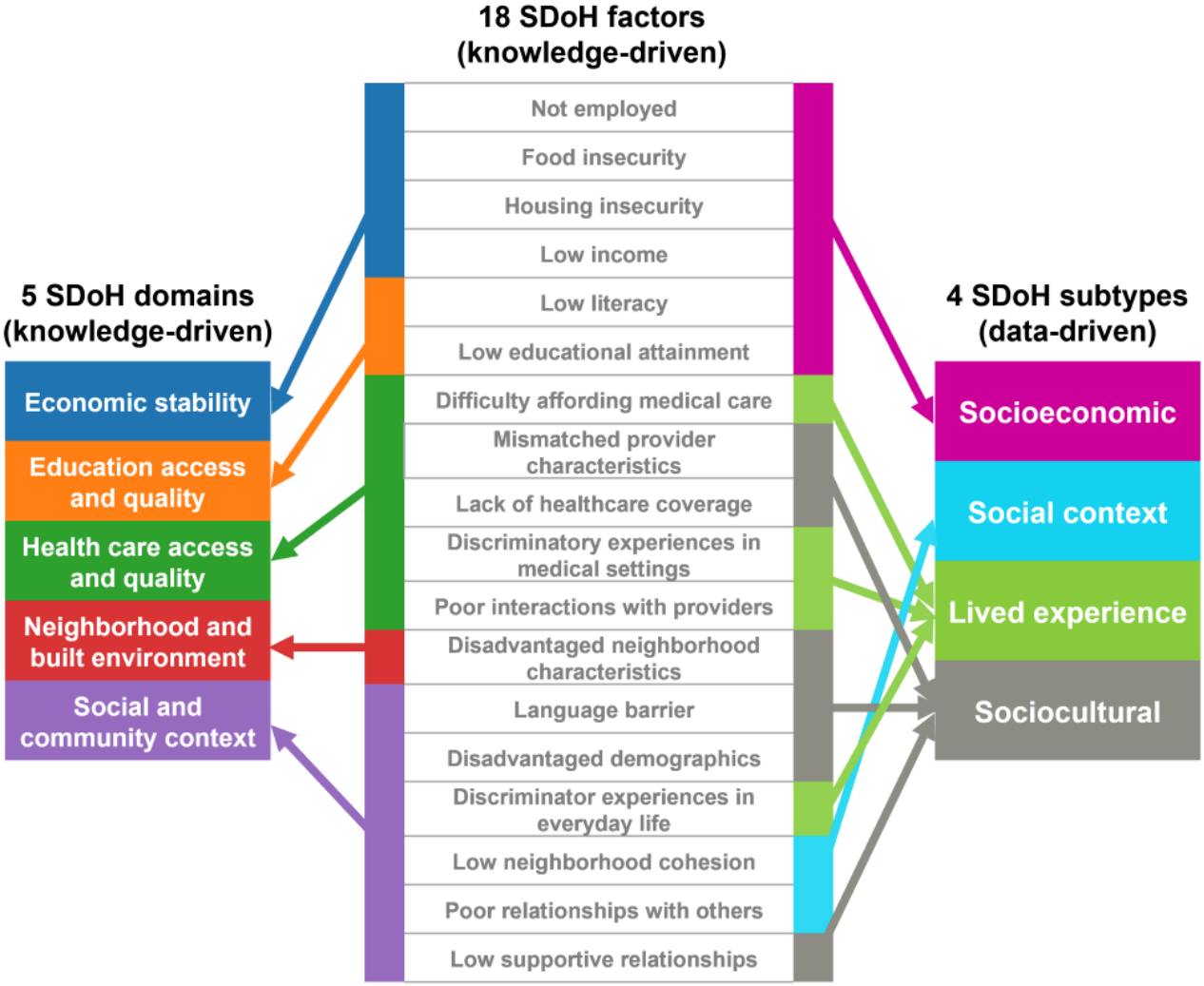

Integration: Risk and Enrichment of Subtypes

Table 2 shows the association of each subtype with the 3 outcomes. As shown by the italicized row, Cluster 1 (low educational attainment, low literacy, low income, not employed, food insecurity, and housing insecurity) had a significantly higher OR for each of the 3 outcomes than Cluster 4 (mismatched provider characteristics, disadvantaged neighborhood characteristics, lack of health care coverage, disadvantaged demographics, low supportive relationships, and language barriers). Furthermore, within the Depression outcome, each of the clusters had a significantly higher OR than one other cluster, forming a ranking of risk among all the 4 clusters (1>3>2>4). In contrast, Delayed medical care had 2 other significant associations (2>1 and 3>4), with ER visits in the last year having only 1 significant pairwise association that fit into the overall trend.

As shown in Table 3, this trend continued in the enrichment analysis of association with living in a state with *No Medicaid expansion*. As shown, Cluster 1 had a significantly higher OR than Cluster 4 in addition to having a significantly OR than the other clusters. The overall results suggest that Cluster 1 and Cluster 4 form “book ends” representing the high and low ends of the risk spectrum among the clusters.

XSL•FO
RenderX

Table 2. Cluster comparisons to measure the risk across all 3 outcomes. Cluster 1 had a significantly higher risk than Cluster 4 for all 3 outcomes (shown in italics). The depression outcome had a distinct ranking of risks, whereas the other 2 outcomes had a subset of that ranking. All *P* values shown are corrected for multiple testing.

Cluster comparison	Outcomes					
	Depression		Delayed medical care		ER[a] visits in the last year	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
1 vs 2	1.7 (1.5-2)	<.001	0.78 (0.67-0.92)	.004[c]	1.2 (0.91-1.6)	.24
1 vs 3	1.3 (1.1-1.6)	.02[d]	0.88 (0.72-1.1)	.23	1.4 (0.96-1.9)	.13
1 vs 4	4.2 (3.5-5.1)	<.001	3.5 (3-4.1)	<.001	1.8 (1.4-2.3)	<.001
2 *vs* 3	0.79 (0.64-0.97)	.02[d]	1.2 (0.98-1.4)	.09	1 (0.75-1.5)	.80
2 vs 4	2.3 (1.9-2.7)	<.001	4.3 (3.7-5)	<.001	1.3 (1-1.7)	.12
3 vs 4	2.9 (2.3-3.5)	<.001	3.6 (3-4.4)	<.001	1.4 (0.95-1.9)	.13

[a]ER: emergency room.

[b]OR: odds ratio.

[c]*P*<.01.

[d]*P*<.05.

Table 3. Cluster 1 had a significantly higher odds ratio (OR) than Cluster 4 (shown in italics) for No Medicaid expansion in addition to having a higher OR than Cluster-2 and Cluster-3. All *P* values shown are corrected for multiple testing.

Cluster comparison	Enrichment—No Medicaid expansion	
	OR (95% CI)	*P* value
1 vs 2	1.5 (1.3-1.8)	<.001
1 vs 3	1.3 (1-1.6)	.048[a]
1 vs 4	*1.3 (1.1-1.5)*	*.006* [b]
2 vs 3	0.99 (0.8-1.2)	.97
2 vs 4	0.99 (0.86-1.2)	.97
3 vs 4	1 (0.82-1.2)	.97

[a]*P*<.05.

[b]*P*<.01.

Translation: Interpretation of SDoH Subtypes and Design of Potential Interventions

The expert panel examined the co-occurrences of SDoH factors within each bicluster shown in the network visualization (Figure 7) and integrated them with the quantitative ORs in Tables 2 and 3. The consistent "book ends" result where Cluster 1 had significantly higher ORs than Cluster 4 across all 4 variables was of strong interest and interpreted as follows: (1) Cluster 1 was labeled *Socioeconomic barriers* as it comprised multiple high-risk SDoH. These co-occurring SDoH could have resulted from cascades over time, such as low educational attainment, potentially leading to lower rates of employment and lower income with higher rates of food and housing insecurity. Such cascading factors can be perceived as being relatively unmodifiable, leading to a higher risk of chronic stress and depression. Furthermore, the strong association of this subtype with the outcomes *Delayed medical care* and *ER visits in the last tear* and the fact that participants in this subtype were more likely to be from a US state with *No Medicaid expansion* provided a more comprehensive understanding of this high-risk SDoH subtype; (2) Cluster 4 was labeled *Sociocultural barriers* as it contained a combination of SDoH related to disadvantaged neighborhood characteristics and low supportive relations in addition to language barriers and mismatched provider interactions. In contrast to the socioeconomic barriers in Cluster 1, many of the sociocultural barriers could be perceived as potentially modifiable, resulting in a lower risk of depression, delayed medical care, and ER visits. Participants who matched this profile could be screened for language and communication barriers, useful for providing culturally competent care, identifying providers who better match the profile of the individuals, and providing resources to facilitate contact with matching nationality or cultural groups in the vicinity or on the web.

While Clusters 1 and 4 formed the "book ends" of risk across the 3 outcomes, potentially caused by relative differences in the unmodifiability of their frequently co-occurring SDoH, Cluster 2 was flagged as critical and labeled *Lived experience barriers*. The SDoH in this cluster included discriminatory experiences in everyday life and in medical settings in addition to poor

XSL•FO
RenderX

interactions with providers and difficulty affording medical care. These frequently co-occurring SDoH could explain why this subtype had a significantly higher OR for *Delayed medical care* than Cluster 1. Finally, Cluster 3 was labeled *Social context barriers* as the SDoH were related to poor neighborhood cohesion and relationships with others. While not as critical as Clusters 1 and 2, this cluster still had a significantly higher OR for Depression than Cluster 4. Together, the 4 clusters could explain how different degrees of unmodifiability in frequently co-occurring SDoH might impact health outcomes.

The expert panel and the ethicist concluded that clinicians treating patients who match each subtype profile could be alerted of specific risks and, consequently, motivate a discussion about mental health and consequences of delayed medical care with the goal of collaboratively exploring options and solutions with the patients. The results could also be useful for resource planning in hospitals to ensure that there is adequate staff to address the needs of the populations they serve, and for proposing health care policies to address the critical connection between specific combinations of SDoH and their impact on public health. For example, many health equity policies categorize Americans based on sociodemographic variables like race and income, which are proxies of need rather than the needs themselves. Instead, such policies could categorize Americans based on SDoH subtypes and their risks to more precisely allocate resources based on combinations of real-world needs.

Furthermore, the subtypes did not have a one-to-one mapping to the 5 SDoH domains defined by HP-30. As shown in Figure 8, these data-driven clusters have a complex relationship with the SDoH domains and factors. While one subtype belonged to a single domain (the *Social context* subtype belonged to the *Social and community context* domain), 3 of the 4 subtypes belonged to ≥2 domains (eg, the *Socioeconomic barriers* subtype belonged to the *Economic stability* and *Education access and quality* domains). Such interdomain relationships reflect how SDoH co-occur in the real world, reflecting the complex cross-domain interactions described in the Dahlgren-Whitehead model (Figure 1). These relationships could be useful for refining conceptual models to explain the complex association between SDoH and adverse health outcomes and build more accurate SDoH models for predicting adverse health outcomes.

Discussion

Principal Findings

The mechanisms through which SDoH precipitate adverse health outcomes are complex, consisting of many interacting factors and feedback loops among individual and environmental and contextual factors. While this phenomenon has been studied for >3 decades, critical hurdles for researchers have included the *limited range* of datatypes, *limited representation* of populations that have been socially marginalized, and *limited access* to individual-level data at scale due to privacy laws. Recognizing that *All of Us* has well-articulated plans and resources to overcome these limitations but is still in a rapidly evolving stage, we conducted a systematic characterization of >100 SDoH survey questions available in *All of Us* and used them to identify SDoH subtypes with the future goal of designing targeted interventions.

For question 1, we identified 110 SDoH questions across 4 surveys, which covered all 5 domains in HP-30. However, the results also revealed a large degree of missingness in survey responses (1.76%-84.56%), with later surveys having significantly fewer responses than earlier ones, and significant differences in race, ethnicity, and age of participants among those who completed the surveys with SDoH questions compared to those in the full *All of Us* dataset. Furthermore, as the SDoH questions varied in granularity, they were categorized by an expert panel into 18 SDoH factors. For question 2, the subtype analysis identified 4 biclusters with significant biclusteredness and significant replication. Furthermore, there were statistically significant associations between specific subtypes and the outcomes, as well as with Medicaid expansion, each with meaningful interpretations and potential targeted interventions. Finally, the identified subtypes spanned one or more HP-30 domains, revealing the difference between the current knowledge-based SDoH domains and the data-driven subtypes.

While the results revealed the nature of and responses to SDoH questions in *All of Us* and significant and interpretable SDoH subtypes, the analysis revealed critical opportunities and challenges related to data, methods, and theory. Such insights are useful for future researchers conducting similar analyses on *All of Us* and, therefore, are discussed in the following sections.

Data: Missingness and Granularity

Missingness

The analysis revealed 3 types of missingness. The first was *rollout missingness*. This type of missingness was largely dictated by how the surveys were rolled out to participants. As all participants at enrollment are required to do *The Basics* and *Overall Health* surveys, these surveys had the highest number of responses, followed by the later solicited *Healthcare Access and Utilization* and *SDoH* surveys rolled out more recently in 2022. This order of rollout was the main source of missingness, resulting in a precipitous reduction in cohort size for those who had answers to all the SDoH questions. The second type was *valid answer missingness*. As participants can choose not to answer any survey questions, the data contained “skip” and “choose not to answer” responses. However, these responses accounted for a much smaller reduction in cohort size for complete data than rollout missingness. The third type was *low use missingness*. Although there were 259 SDoH SNOMED codes, only 93 (35.9%) had such information for >20 participants that are allowed to be reported. This could be because most clinicians currently do not screen for SDoH as it is typically done by a social worker. Furthermore, we also attempted to use 3-digit zip codes to determine which subtypes had a significant association with living in a state that did not offer Medicaid expansion. However, 13.07% (1688/12,913) of the participants did not have zip code information (which was adjusted by using IPW).

Together, the aforementioned 3 types of missingness impacted the size of the resulting cohort that had valid answers in the

XSL•FO
RenderX

following 2 ways, which is the minimum number of participants that are allowed to be reported by *All of Us*. First, there was a drastic reduction in cohort size by 93.5%. However, because of the size of the overall dataset (N=372,397), we were still left with a large cohort (n=12,886), which, to the best of our knowledge, is the largest set of individuals to be analyzed for such a wide range of SDoH. Second, there were significant differences in the proportion of race, ethnicity, and age in the aforementioned cohort when compared to the overall *All of Us* population. Specifically, the cohort with valid answers had significantly more White, non-Hispanic, or older participants when compared to the overall cohort. This could potentially be because, once a participant has been enrolled, there is a 90-day delay in sending subsequent solicitations to complete surveys, a policy that is currently being reassessed due to its impact on missingness. Therefore, we had to correct this imbalance in demographic proportions by using IPW with the goal of identifying subtypes that were representative of the overall *All of Us* cohort.

Granularity

Because our goal was to use machine learning methods to identify SDoH subtypes, we encountered uneven granularity in the SDoH questions. Some questions were fine grained and highly correlated and, therefore, would cluster more strongly because of the nature of the granularity of the questions, not because of the SDoH mechanisms. To address this uneven granularity and make the results more interpretable, we used SDoH factors that had a coarser but more consistent level of granularity. We chose this approach because SDoH factors had already been defined; were understood by the expert panel, enabling high domain fidelity; and appeared to be at the right level of abstraction useful for clinical applications, such as referring a patient to the appropriate social services. However, because the use of coarse-grained variables loses information, future research could explore aggregating only those SDoH questions that are highly correlated while preserving the rest at the finer level of granularity and explore computational methods to merge SDoH questions into SDoH factors.

Method: Scalability, Generalizability, and Extensibility

We designed the HIT analytical framework to be scalable, enabling its use for the growing size of the data in *All of Us*; generalizable across cohorts and conditions; and extensible for including additional methods as needed in the future. Testing the HIT framework on the *All of Us* dataset provided insights into the strengths and limitations of the framework and the *All of Us* workbench where the analysis was conducted.

Scalability

We used 3 types of code to conduct the analysis for both research questions. The first was automatically generated code to extract the cohort, produced by *All of Us* once a cohort was selected using the point-and-click interface. This code was adequately scalable and generalizable and so will not be discussed further. The second was customized code to extract specific parts of the data. For example, the analysis of co-occurrences required customized code in R (R Foundation for Statistical Computing) to plot the diagrams in Figures 4–6. As expected, these tasks required strong programming skills, but fortunately, we did not encounter any coding or execution problems using the R or Python (Python Software Foundation) programming languages. However, there were significant server issues that hampered our analysis. Although the workbench instructions stated that code running on the workbench for >2 weeks would be terminated and all intermediate results would be deleted, we frequently encountered our work disappearing at shorter intervals. These disruptions resulted in a higher consumption of the free server time credits resulting in fewer analyses that could be conducted. The third type of code was machine learning code that we had previously developed and disseminated on Comprehensive R Archive Network [78-80] to conduct the bipartite network analysis and the significance testing and visualize the network. As this code was designed to be generalizable and scalable, we did not encounter any issues in the execution of our code (in addition to the same server issues mentioned previously). Finally, the visualization of our networks worked as expected, and we used them to help interpret the patterns in the data.

Generalizability

Our code for the first 2 steps of the HIT framework is in Project Jupyter notebooks and has been used to analyze other cohorts that were filtered for age and previous conditions. For example, we extracted a cohort (n=4090) of participants with diabetes aged ≥65 years with complete data on 18 SDoH variables selected through consensus by 2 experienced health services researchers and guided by the Andersen behavioral model. The analysis [81,82] revealed 7 SDoH subtypes with statistically significant modularity compared with 100 random permutations of the data (*All of Us*=0.51; random mean 0.38 SD 0.0065; z=20; P<.001) and that were not only clinically meaningful but also significant in different degrees for the outcome. Our subsequent attempt at increasing the number of SDoH variables from 18 to 110 for participants with diabetes who had valid answers led to an extremely small cohort size (n=926; Multimedia Appendix 5) due to the missingness that we described previously. While this reduction resulted in our current strategy of analyzing all participants regardless of condition or age, these experiments demonstrate that our approach is generalizable to other subsets of the data.

Extensibility

The HIT model is designed to be extensible to include other methods. For example, the model could use other biclustering (eg, nonnegative matrix factorization [83]) and causal modeling methods and different types of classification (eg, deep learning [84]) and prediction methods (eg, subgroup-specific modeling [40]) to build the decision support system in the translational step (Figure 2). Furthermore, the model can integrate a wide range of datatypes to enable analysis of how each subtype is associated with them, resulting in a layered interpretation of the SDoH subtypes, as we have demonstrated. For example, as the percentage of participants who have genomic information increases (currently, >25% of our cohort have missing genomic information), our pipeline will be able to integrate such information into our analysis. Finally, the integration of different datatypes required a diverse team consisting of experts in

XSL•FO
RenderX

machine learning, biostatistics, programming, clinical care, health services research, gerontology, and ethics to enable a 360° analysis and interpretation of the subtypes that was, therefore, aligned with the human-centered artificial intelligence approach [64-66]. Furthermore, the use of the workbench to share results through visualizations operationalized *team-centered informatics* [85] designed to facilitate multidisciplinary translational teams [86] working more effectively across disciplinary boundaries with the goal of analyzing subtypes and designing targeted interventions.

Theory: Model Building and Translational Implications

The identification of SDoH subtypes has strong implications for model building in addition to translational applications. As shown in Figure 8, while the current classification of 5 SDoH domains has a hierarchical relationship with the SDoH factors, the data-driven clusters have a more complex association with the same SDoH factors. This reflects the complexity of how SDoH occur in the real world while at the same time being interpretable for purposes of translation.

Future studies should develop prediction models using the data-driven subtypes to determine whether they improve the accuracy of predicting adverse health outcomes when compared to models that do not use those subtypes. Because the subtypes were clinically interpretable, they could be used to build classification and prediction models and used with an interface to develop a clinical decision support system that helps triage patients to critical services. For example, the St Vincent's House [87] in Galveston, Texas, United States, provides several services to address SDoH, including free walk-in clinical care, nurse practitioners with reduced insurance co-payments, English- and Spanish-speaking free mental health counseling, free dental health clinic, utility and rental assistance, case management, financial literacy, expanded food pantry, weekly free home delivery of pantry groceries, snack pack for people experiencing homelessness, free transportation to clinician's appointments, immigration legal services, and spiritual counseling. Given the availability of this wide range of services in many communities across the United States, a decision support system could help classify an individual based on their SDoH profile into one or more of the subtypes and measure their risk of an adverse health outcome. Such information could be used by clinicians to collaboratively explore solutions with the patient to consider one or more of such local services based on the membership strength to a subtype and the associated risk (Figure 2, step 3). At a population level, understanding health risks associated with clusters may assist institutions and organizations in developing more effective prevention programs.

Notebooks for *All of Us* Community Use

Because the missingness in SDoH variables is expected to decrease, their characterization and subtyping will need to be repeated and verified for different cohorts. Therefore, we have made the following 2 sets of code available for general use by the *All of Us* researcher community (accessible after creating a free account on *All of Us* and completing the required training).

SDoH Valid Answer Tracker

This set of notebooks generates four plots that can be used by other researchers on *All of Us* to characterize any cohort: (1) valid response plot to show how many participants have data with valid responses and colored by SDoH domain, (2) Venn diagram showing how many participants have valid responses for all questions within each survey, and (3) frequency distribution plot showing co-occurrence of SDoH across the selected cohort. This set of tools should enable researchers to characterize SDoH across different cohorts to help determine methods that are appropriate to adjust for missingness in those cohorts.

SDoH Subtyper

This set of notebooks can be used to conduct the following analyses: (1) bicluster modularity of a cohort with the 18 SDoH factors to identify the number and members of biclusters and the measure Q representing the quality of the biclustering, (2) visualization of the bipartite network, and (3) significance of the network with respect to null models.

Limitations

This study has 2 main limitations. The first emerges from the large amount of missingness in the survey data, precluding the use of imputation methods that assume a random distribution of missingness. Therefore, we could use only complete data, which led to a large drop in cohort size and which also introduced a bias in the demographics requiring a rebalancing through IPW. While such rebalancing is typically done for large datasets, the IPW method requires judgment to decide which variables to include in the model and, therefore, could have introduced additional unknown biases. Therefore, the model should be refined to determine which variables to include in the regression models that estimate the IPW-generated weights. However, because the clustering was similar between the unweighted and IPW-weighted networks, we believe that the current subtypes are stable and meaningful and represent the demographic composition of the full *All of Us* dataset, but this needs to be verified by redoing the analysis as the data become more complete. The limitation of missingness in the surveys is expected to be addressed as *All of Us* has recently removed the requirement of waiting 90 days before a subsequent survey is given to an enrollee in the program, potentially reducing the degree of missingness. The second limitation is due to the high computational cost of empirically determining the significance of the biclustering. As such analysis is computationally expensive and time-consuming, it limited the experiments we could do to test different cohorts and models. Therefore, we look forward to the *All of Us* workbench providing the ability to run batch processes more efficiently and uninterruptedly for extended periods (exceeding the current time window), which together could help alleviate this computational hurdle in the future.

Conclusions

How SDoH impact health is a complex phenomenon involving many interconnected social, biological, and environmental factors that have yet to be fully elucidated. While this phenomenon has been studied for >30 years, the analyses have

XSL•FO
RenderX

been hampered by the lack of large cohorts representing diverse populations with a wide range of SDoH variables measured, multiple datatypes, and easy access by researchers. *All of Us* provides an unprecedented opportunity to directly address these limitations with the goal of doing justice to early conceptual models such as the social gradient and the Dahlgren-Whitehead model, both of which drew international attention to the complex ways in which individual and contextual SDoH factors impact health. The *All of Us* dataset is also timely because of the extensive health disparities that were revealed during the pandemic, which highlighted the critical need to address SDoH in the public and policy realms. However, because *All of Us* is still rapidly evolving to meet its target of 1 million participants or more, we conducted a systematic characterization of SDoH variables in *All of Us* and used the results to guide the analysis of SDoH subtypes. The identified subtypes, along with their risks, could be used to design data-informed interventions, resource-planning strategies, and public health policies aimed at reducing the risks of adverse health outcomes. However, careful consideration would be required to ensure that the identification of high-risk subtypes is not used in a way that stigmatizes subpopulations.

Our first goal of characterizing the data revealed the nature of the missingness in SDoH and the uneven granularity in the SDoH questions. Both of these results led us to select the IPW method to address the missingness and to analyze subtypes using SDoH factors to address the uneven granularity. Our second goal of identifying SDoH subtypes led not only to statistically significant biclusteredness but also to their statistically significant replication and meaningful domain interpretations. These results set the stage for further investigations to build and evaluate classification and prediction models for designing decision support systems that alert clinicians of specific risks that their patients face due to a combination of SDoH factors. Furthermore, the SDoH subtypes could be used to design public health care policies that are multifactorial and need based, enabling more targeted interventions compared to policies that are based on a few sociodemographic factors such as race and income.

The results also led to the design, use, and dissemination of general-purpose tools currently available on *All of Us* for other researchers, which will be useful to reanalyze the *All of Us* dataset as it grows over the next few years to directly address the high rate of missingness. These collaborative advances should position *All of Us* to revolutionize research for analyzing complex phenomena such as how SDoH impact health and beyond, with the goal of enabling a more equitable future that all of us deserve.

Acknowledgments

The authors thank Gautam Vallabha, PhD, for his assistance in refining the analysis and the manuscript. This study was supported in part by the Artificial Intelligence and Machine Learning Consortium to Advance Health Equity and Researcher Diversity (1OT2SD03258102293), the Clinical and Translational Science Award (UL1 TR001439) from the National Center for Advancing Translational Sciences at the National Institutes of Health, the University of Texas Medical Branch Claude D. Pepper Older Americans Independence Center funded by the National Institute on Aging at the National Institutes of Health (P30AG024832), the MD Anderson Cancer Center, the National Library of Medicine (R01 LM012095) at the National Institutes of Health, and the National Institute on Aging (K01AG058789, K01AG065492), Agency for Healthcare Research and Quality (R01-HS029891). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. No part of this manuscript was generated using large language models.

Data Availability

The datasets analyzed in this study are publicly available from the *All of Us* program after registering for access [88].

Authors' Contributions

SKB and BD conceived the overall design of the analysis and wrote the manuscript. WZ, DB, and AB wrote the analytical code to analyze the data. BD designed the method for handling the missing data, identified sources for social determinants of health (SDoH), and guided the overall design of the analysis. SV identified frameworks for SDoH and guided the overall design of the analysis. YFK provided statistical guidance for the analysis. BD, SS, and TR guided the identification and granularization of the SDoH questions from the *All of Us* surveys. SS, MRP, TR, BD, and MR interpreted the subtypes and inferred the mechanisms and potential interventions, and VA and RH refined those interpretations. ES provided guidance on the ethical aspects of the analysis and interpretation. SV and BD had equal contribution for guiding the overall project. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of bipartite network analysis.
[DOCX File , 617 KB - jmir_v27i1e48775_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Systematized Medical Nomenclature of Medicine codes related to social determinants of health and their use in the electronic health records of participants in All of Us.
[DOCX File , 184 KB - jmir_v27i1e48775_app2.docx]

Multimedia Appendix 3
Data conversion. A total of 4 *All of Us* surveys (column 2) contained 110 social determinants of health questions (column 3), which were abbreviated, negatively phrased (shown bolded), and reversed coded (shown in red; column 3); categorized into the 5 Healthy People 2030 domains (column 4 and shown by the 5 colors); and further categorized (boxes) by the expert panel into 18 factors (column 5; Delayed medical care was used as an outcome).
[DOCX File , 40 KB - jmir_v27i1e48775_app3.docx]

Multimedia Appendix 4
Inverse probability weighting.
[DOCX File , 99 KB - jmir_v27i1e48775_app4.docx]

Multimedia Appendix 5
Condition-specific cohort extraction for type 2 diabetes, breast cancer, and coronary artery disease, and inclusion and exclusion criteria for selecting 3 condition-specific cohorts.
[DOCX File , 66 KB - jmir_v27i1e48775_app5.docx]

References

1. Weida EB, Phojanakong P, Patel F, Chilton M. Financial health as a measurable social determinant of health. PLoS One 2020;15(5):e0233359 [FREE Full text] [doi: 10.1371/journal.pone.0233359] [Medline: 32421755]
2. Kushel MB, Gupta R, Gee L, Haas JS. Housing instability and food insecurity as barriers to health care among low-income Americans. J Gen Intern Med 2006 Jan;21(1):71-77 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00278.x] [Medline: 16423128]
3. Social determinants of health. World Health Organization. URL: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1 [accessed 2024-07-07]
4. Dahlgren G, Whitehead M. The Dahlgren-Whitehead model of health determinants: 30 years on and still chasing rainbows. Public Health 2021 Oct;199:20-24. [doi: 10.1016/j.puhe.2021.08.009] [Medline: 34534885]
5. Cook WK. Paid sick days and health care use: an analysis of the 2007 national health interview survey data. Am J Ind Med 2011 Oct;54(10):771-779 [FREE Full text] [doi: 10.1002/ajim.20988] [Medline: 21761429]
6. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. Circulation 1993 Oct;88(4 Pt 1):1973-1998. [doi: 10.1161/01.cir.88.4.1973] [Medline: 8403348]
7. Mapping key determinants of immigrants' health in Brooklyn and Queens. Center for Migration Studies of New York (CMS). 2021. URL: https://cmsny.org/wp-content/uploads/2021/02/Mapping-Key-Health-Determinants-for-Immigrants-Report-Center-for-Migration-Studies.pdf [accessed 2024-07-07]
8. Social Determinants of Health (SDOH). NEJM Catalyst. 2017 Dec 1. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.17.0312 [accessed 2023-06-28]
9. Ramirez AH, Gebo KA, Harris PA. Progress with the all of us research program: opening access for researchers. JAMA 2021 Jun 22;325(24):2441-2442. [doi: 10.1001/jama.2021.7702] [Medline: 34115137]
10. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" research program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
11. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, et al. The All of Us Research Program: data quality, utility, and diversity. Patterns (N Y) 2022 Aug 12;3(8):100570 [FREE Full text] [doi: 10.1016/j.patter.2022.100570] [Medline: 36033590]
12. Marmot MG. Status syndrome: a challenge to medicine. JAMA 2006 Mar 15;295(11):1304-1307. [doi: 10.1001/jama.295.11.1304] [Medline: 16537740]
13. Adler NE, Ostrove JM. Socioeconomic status and health: what we know and what we don't. Ann N Y Acad Sci 1999;896:3-15. [doi: 10.1111/j.1749-6632.1999.tb08101.x] [Medline: 10681884]
14. Mackenbach JP, Bos V, Andersen O, Cardano M, Costa G, Harding S, et al. Widening socioeconomic inequalities in mortality in six Western European countries. Int J Epidemiol 2003 Oct;32(5):830-837 [FREE Full text] [doi: 10.1093/ije/dyg209] [Medline: 14559760]
15. McDonough P, Duncan GJ, Williams D, House J. Income dynamics and adult mortality in the United States, 1972 through 1989. Am J Public Health 1997 Sep;87(9):1476-1483. [doi: 10.2105/ajph.87.9.1476] [Medline: 9314800]

XSL•FO
RenderX

16. Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Background document to WHO - Strategy paper for Europe. Institutet för Framtidsstudier. 1991. URL: https://www.researchgate.net/publication/5095964_Policies_and_strategies_to_promote_social_equity_in_health_Background_document_to_WHO_-_Strategy_paper_for_Europe [accessed 2024-05-27]
17. Lucyk K, McLaren L. Taking stock of the social determinants of health: a scoping review. PLoS One 2017;12(5):e0177306 [FREE Full text] [doi: 10.1371/journal.pone.0177306] [Medline: 28493934]
18. Muller A. Education, income inequality, and mortality: a multiple regression analysis. BMJ 2002 Jan 05;324(7328):23-25 [FREE Full text] [doi: 10.1136/bmj.324.7328.23] [Medline: 11777800]
19. Hill-Briggs F, Adler NE, Berkowitz SA, Chin MH, Gary-Webb TL, Navas-Acien A, et al. Social determinants of health and diabetes: a scientific review. Diabetes Care 2020 Nov 02;44(1):258-279 [FREE Full text] [doi: 10.2337/dci20-0053] [Medline: 33139407]
20. PhenX Social Determinants of Health (SDOH) assessments collection. National Institutes of Health National Institute on Minority Health and Health Disparities. URL: https://www.nimhd.nih.gov/resources/phenx/ [accessed 2024-01-01]
21. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, et al. The PhenX Toolkit: get the most from your measures. Am J Epidemiol 2011 Aug 01;174(3):253-260 [FREE Full text] [doi: 10.1093/aje/kwr193] [Medline: 21749974]
22. Social determinants of health collections. PhenX Toolkit. URL: https://www.phenxtoolkit.org/collections/view/6 [accessed 2024-07-07]
23. CMS framework for health equity 2022–2032. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/files/document/cms-framework-health-equity.pdf [accessed 2024-07-07]
24. Green K, Zook M. When talking about social determinants, precision matters. Health Affairs. 2019 Oct 29. URL: https://www.healthaffairs.org/content/forefront/talking-social-determinants-precision-matters [accessed 2024-07-07]
25. Craven CK, Highfield L, Basit M, Bernstam EV, Choi BY, Ferrer RL, et al. Toward standardization, harmonization, and integration of social determinants of health data: a Texas Clinical and Translational Science Award institutions collaboration. J Clin Transl Sci 2024;8(1):e17 [FREE Full text] [doi: 10.1017/cts.2024.2] [Medline: 38384919]
26. McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010 Apr 16;141(2):210-217 [FREE Full text] [doi: 10.1016/j.cell.2010.03.032] [Medline: 20403315]
27. Waldman SA, Terzic A. Therapeutic targeting: a crucible for individualized medicine. Clin Pharmacol Ther 2008 May;83(5):651-654. [doi: 10.1038/clpt.2008.65] [Medline: 18425084]
28. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005 Aug 15;11(16):5678-5685. [doi: 10.1158/1078-0432.CCR-04-2421] [Medline: 16115903]
29. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001 Sep 11;98(19):10869-10874 [FREE Full text] [doi: 10.1073/pnas.191367098] [Medline: 11553815]
30. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, et al. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. J Allergy Clin Immunol 2011 Feb;127(2):382-9.e1 [FREE Full text] [doi: 10.1016/j.jaci.2010.11.015] [Medline: 21195471]
31. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med 2008 Aug 01;178(3):218-224 [FREE Full text] [doi: 10.1164/rccm.200711-1754OC] [Medline: 18480428]
32. Lötvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. J Allergy Clin Immunol 2011 Feb;127(2):355-360. [doi: 10.1016/j.jaci.2010.11.037] [Medline: 21281866]
33. Nair P, Pizzichini MM, Kjarsgaard M, Inman MD, Efthimiadis A, Pizzichini E, et al. Mepolizumab for prednisone-dependent asthma with sputum eosinophilia. N Engl J Med 2009 Mar 05;360(10):985-993. [doi: 10.1056/NEJMoa0805435] [Medline: 19264687]
34. Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, et al. Mepolizumab treatment in patients with severe eosinophilic asthma. N Engl J Med 2014 Sep 25;371(13):1198-1207. [doi: 10.1056/nejmoa1403290]
35. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795 [FREE Full text] [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
36. Lacy ME, Wellenius GA, Carnethon MR, Loucks EB, Carson AP, Luo X, et al. Racial differences in the performance of existing risk prediction models for incident type 2 diabetes: the CARDIA study. Diabetes Care 2016 Feb;39(2):285-291 [FREE Full text] [doi: 10.2337/dc15-0509] [Medline: 26628420]
37. Baker JJ. Medicare payment system for hospital inpatients: diagnosis-related groups. J Health Care Finance 2002;28(3):1-13. [Medline: 12079147]
38. Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search--a recursive partitioning method for establishing response to treatment in patient subpopulations. Stat Med 2011 Sep 20;30(21):2601-2621. [doi: 10.1002/sim.4289] [Medline: 21786278]

XSL•FO
RenderX

39. Kehl V, Ulm K. Responder identification in clinical trials with censored data. Comput Stat Data Anal 2006 Mar;50(5):1338-1355. [doi: 10.1016/j.csda.2004.11.015]
40. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2001.
41. Abu-Jamous B, Fa R, Nandi AK. Integrative Cluster Analysis in Bioinformatics. Hoboken, NJ: John Wiley & Sons; 2015.
42. Lochner KA, Cox CS. Prevalence of multiple chronic conditions among Medicare beneficiaries, United States, 2010. Prev Chronic Dis 2013 Apr 25;10:E61 [FREE Full text] [doi: 10.5888/pcd10.120137] [Medline: 23618541]
43. Aryal S, Diaz-Guzman E, Mannino DM. Prevalence of COPD and comorbidity. In: Rabe KF, Wedzica JA, Wouters EF, editors. COPD and Comorbidity. Sheffield, UK: European Respiratory Society; 2013:1-12.
44. Baty F, Putora PM, Isenring B, Blum T, Brutsche M. Comorbidities and burden of COPD: a population based case-control study. PLoS One 2013;8(5):e63285 [FREE Full text] [doi: 10.1371/journal.pone.0063285] [Medline: 23691009]
45. Moni MA, Liò P. Network-based analysis of comorbidities risk during an infection: SARS and HIV case studies. BMC Bioinformatics 2014 Oct 24;15(1):333 [FREE Full text] [doi: 10.1186/1471-2105-15-333] [Medline: 25344230]
46. Cramer AO, Waldorp LJ, van der Maas HL, Borsboom D. Comorbidity: a network perspective. Behav Brain Sci 2010 Jun;33(2-3):137-50; discussion 150. [doi: 10.1017/S0140525X09991567] [Medline: 20584369]
47. Islam MM, Valderas JM, Yen L, Dawda P, Jowsey T, McRae IS. Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. PLoS One 2014;9(1):e83783 [FREE Full text] [doi: 10.1371/journal.pone.0083783] [Medline: 24421905]
48. Folino F, Pizzuti C, Ventura M. A comorbidity network approach to predict disease risk. In: Proceedings of the Information, Technology in Bio- and Medical Informatics. 2010 Presented at: ITBAM 2010; September 1-2, 2010; Bilbao, Spain. [doi: 10.1007/978-3-642-15020-3_10]
49. Newman M. Networks: An Introduction. Oxford, UK: Oxford University Press; 2010.
50. Treviño SIII, Nyberg A, Del Genio CI, Bassler KE. Fast and accurate determination of modularity and its effect size. J Stat Mech Theor Exp 2015 Feb 03(2015):P02003. [doi: 10.1088/1742-5468/2015/02/P02003]
51. Chauhan R, Ravi J, Datta P, Chen T, Schnappinger D, Bassler KE, et al. Reconstruction and topological characterization of the sigma factor regulatory network of Mycobacterium tuberculosis. Nat Commun 2016 Mar 31;7:11062 [FREE Full text] [doi: 10.1038/ncomms11062] [Medline: 27029515]
52. Bhavnani SK, Dang B, Penton R, Visweswaran S, Bassler KE, Chen T, et al. How high-risk comorbidities co-occur in readmitted patients with hip fracture: big data visual analytical approach. JMIR Med Inform 2020 Oct 26;8(10):e13567 [FREE Full text] [doi: 10.2196/13567] [Medline: 33103657]
53. Fruchterman TM, Reingold EM. Graph drawing by force‐directed placement. Softw Pract Exp 2006 Oct 30;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
54. Dang B, Chen T, Bassler KE, Bhavnani SK. ExplodeLayout: enhancing the comprehension of large and dense networks. In: Proceedings of AMIA Summit on Translational Bioinformatics. 2016 Presented at: AMIA 2016; March 21-22, 2016; San Francisco, CA.
55. Bhavnani SK, Chen T, Ayyaswamy A, Visweswaran S, Bellala G, Divekar R, et al. Enabling comprehension of patient subgroups and characteristics in large bipartite networks: implications for precision medicine. AMIA Jt Summits Transl Sci Proc 2017;2017:21-29 [FREE Full text] [Medline: 28815099]
56. Bhavnani SK, Eichinger F, Martini S, Saxman P, Jagadish HV, Kretzler M. Network analysis of genes regulated in renal diseases: implications for a molecular-based classification. BMC Bioinformatics 2009 Sep 17;10 Suppl 9(Suppl 9):S3 [FREE Full text] [doi: 10.1186/1471-2105-10-S9-S3] [Medline: 19761573]
57. Bhavnani SK, Bellala G, Ganesan A, Krishna R, Saxman P, Scott C, et al. The nested structure of cancer symptoms. Implications for analyzing co-occurrence and managing symptoms. Methods Inf Med 2010;49(6):581-591 [FREE Full text] [doi: 10.3414/ME09-01-0083] [Medline: 21085743]
58. Bhavnani SK, Ganesan A, Hall T, Maslowski E, Eichinger F, Martini S, et al. Discovering hidden relationships between renal diseases and regulated genes through 3D network visualizations. BMC Res Notes 2010 Nov 11;3:296 [FREE Full text] [doi: 10.1186/1756-0500-3-296] [Medline: 21070623]
59. Bhavnani SK, Victor S, Calhoun WJ, Busse WW, Bleecker E, Castro M, et al. How cytokines co-occur across asthma patients: from bipartite network analysis to a molecular-based classification. J Biomed Inform 2011 Dec;44 Suppl 1(Suppl 1):S24-S30 [FREE Full text] [doi: 10.1016/j.jbi.2011.09.006] [Medline: 21986291]
60. Bhavnani SK, Bellala G, Victor S, Bassler KE, Visweswaran S. The role of complementary bipartite visual analytical representations in the analysis of SNPs: a case study in ancestral informative markers. J Am Med Inform Assoc 2012 Jun;19(e1):e5-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000745] [Medline: 22718038]
61. Bhavnani SK, Dang B, Bellala G, Divekar R, Visweswaran S, Brasier AR, et al. Unlocking proteomic heterogeneity in complex diseases through visual analytics. Proteomics 2015 Apr;15(8):1405-1418 [FREE Full text] [doi: 10.1002/pmic.201400451] [Medline: 25684269]
62. Bhavnani SK, Dang B, Kilaru V, Caro M, Visweswaran S, Saade G, et al. Methylation differences reveal heterogeneity in preterm pathophysiology: results from bipartite network analyses. J Perinat Med 2018 Jul 26;46(5):509-521 [FREE Full text] [doi: 10.1515/jpm-2017-0126] [Medline: 28665803]

XSL•FO
RenderX

63. Bhavnani SK, Kummerfeld E, Zhang W, Kuo YF, Garg N, Visweswaran S, et al. Heterogeneity in COVID-19 patients at multiple levels of granularity: from biclusters to clinical interventions. AMIA Jt Summits Transl Sci Proc 2021;2021:112-121 [FREE Full text] [Medline: 34457125]
64. Shneiderman B. Human-centered AI: ensuring human control while increasing automation. In: Proceedings of the 5th Workshop on Human Factors in Hypertext. 2022 Presented at: HT '22; June 28, 2022; Barcelona, Spain. [doi: 10.1145/3538882.3542790]
65. Chen Y, Clayton EW, Novak LL, Anders S, Malin B. Human-centered design to address biases in artificial intelligence. J Med Internet Res 2023 Mar 24;25:e43251 [FREE Full text] [doi: 10.2196/43251] [Medline: 36961506]
66. Shneiderman B. Human-centered artificial intelligence: reliable, safe and trustworthy. Int J Hum Comput Interact 2020 Mar 23;36(6):495-504. [doi: 10.1080/10447318.2020.1741118]
67. Patel SB, Nguyen NT. Creation of a mapped, machine-readable taxonomy to facilitate extraction of social determinants of health data from electronic health records. AMIA Annu Symp Proc 2021;2021:959-968 [FREE Full text] [Medline: 35308929]
68. Allen J, Balfour R, Bell R, Marmot M. Social determinants of mental health. Int Rev Psychiatry 2014 Aug;26(4):392-407. [doi: 10.3109/09540261.2014.928270] [Medline: 25137105]
69. Busse D, Yim IS, Campos B, Marshburn CK. Discrimination and the HPA axis: current evidence and future directions. J Behav Med 2017 Aug;40(4):539-552. [doi: 10.1007/s10865-017-9830-6] [Medline: 28155003]
70. Decker A, Weaver RH. Health and social determinants associated with delay of care among community-dwelling rural older adults. Fam Community Health 2024;47(1):41-48. [doi: 10.1097/FCH.0000000000000386] [Medline: 37753957]
71. Newton MF, Keirns CC, Cunningham R, Hayward RA, Stanley R. Uninsured adults presenting to US emergency departments: assumptions vs data. JAMA 2008 Oct 22;300(16):1914-1924. [doi: 10.1001/jama.300.16.1914] [Medline: 18940978]
72. Bhavnani SK, Zhang W, Visweswaran S, Raji M, Kuo YF. A framework for modeling and interpreting patient subgroups applied to hospital readmission: visual analytical approach. JMIR Med Inform 2022 Dec 07;10(12):e37239 [FREE Full text] [doi: 10.2196/37239] [Medline: 35537203]
73. van der Wal WM, Geskus RB. ipw: an R package for inverse probability weighting. J Stat Softw 2011;43(13):1-23. [doi: 10.18637/jss.v043.i13]
74. Thoemmes F, Ong AD. A primer on inverse probability of treatment weighting and marginal structural models. Emerg Adulthood 2015 Dec 16;4(1):40-59. [doi: 10.1177/2167696815621645]
75. Lee YH, Liu Z, Fatori D, Bauermeister JR, Luh RA, Clark CR, et al. Association of everyday discrimination with depressive symptoms and suicidal ideation during the COVID-19 pandemic in the All of Us research program. JAMA Psychiatry 2022 Sep 01;79(9):898-906 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.1973] [Medline: 35895053]
76. All of Us IRB ControlledTier revised letter 08312021. National Institutes of Health. 2021. URL: https://www.researchallofus.org/wp-content/themes/research-hub-wordpress-theme/media/2023/All_of_Us_IRB_ControlledTier_Revised_Letter_08312021%20-%20signed.pdf [accessed 2024-07-07]
77. Do I need my project reviewed by the All of Us Institutional Review Board (IRB) in order to access this data using the Researcher Workbench? All of Us Research Hub. URL: https://www.researchallofus.org/faq/do-i-need-my-project-reviewed-by-the-all-of-us-institutional-review-board-irb-in-order-to-access-this-data-using-the-researcher-workbench/ [accessed 2024-07-07]
78. Chen T, Zhang W, Bhavnani S. BipartiteModularityMaximization: partition bipartite network into non-overlapping biclusters by optimizing bipartite modularity. The Comprehensive R Archive Network. 2023 Jan 21. URL: https://cran.r-project.org/web/packages/BipartiteModularityMaximization/index.html [accessed 2024-07-07]
79. Bhavnani SK, Zhang W. ExplodeLayout: calculate exploded coordinates based on original node coordinates and node clustering membership. The Comprehensive R Archive Network. 2022 Jul 1. URL: https://cran.r-project.org/web/packages/ExplodeLayout/index.html [accessed 2024-07-07]
80. CRAN R packages by number of downloads. DataScienceMeta. URL: http://www.datasciencemeta.com/rpackages [accessed 2024-07-07]
81. Bhavnani SK, Zhang W, Bao D, Kuo YF, Reistetter T, Hatch S, et al. 341 The impact of critical social determinants of health on personal medical decisions: analysis of older Americans in All of Us. J Clin Transl Sci 2023 Apr 24;7(s1):101. [doi: 10.1017/cts.2023.385]
82. Bhavnani SK, Zhang W, Bao D, Hatch S, Reistetter T, Downer B. 284 Generalizable machine learning methods for subtyping individuals on national health databases: case studies using data from HRS, N3C, and All of Us. J Clin Transl Sci 2023 Apr 24;7(s1):85. [doi: 10.1017/cts.2023.340]
83. Dhillon IS, Sra S. Generalized nonnegative matrix approximations with Bregman divergences. In: Proceedings of the 18th International Conference on Neural Information Processing Systems. 2005 Presented at: NIPS'05; December 5-8, 2005; Vancouver, BC.
84. Dilsizian ME, Siegel EL. Machine meets biology: a primer on artificial intelligence in cardiology and cardiac imaging. Curr Cardiol Rep 2018 Oct 18;20(12):139. [doi: 10.1007/s11886-018-1074-8] [Medline: 30334108]

XSL•FO
RenderX

85. Bhavnani SK, Visweswaran S, Divekar R, Brasier AR. Towards team-centered informatics: accelerating innovation in multidisciplinary scientific teams through visual analytics. J Appl Behav Sci 2018 Nov 05;55(1):50-72 [FREE Full text] [doi: 10.1177/0021886318794606]
86. Wooten KC, Calhoun WJ, Bhavnani S, Rose RM, Ameredes B, Brasier AR. Evolution of multidisciplinary translational teams (MTTs): insights for accelerating translational innovations. Clin Transl Sci 2015 Oct;8(5):542-552 [FREE Full text] [doi: 10.1111/cts.12266] [Medline: 25801998]
87. Hope. St. Vincent's House. URL: https://www.stvhope.org/ [accessed 2024-10-15]
88. Register - All of Us Research Hub. All of Us Research Hub. URL: https://www.researchallofus.org/register/ [accessed 2024-07-07]

Abbreviations

EHR: electronic health record
ER: emergency room
HIT: heterogenization, integration, and translation
HP-30: Healthy People 2030
IPW: inverse probability weighting
OR: odds ratio
RI: Rand index
SDoH: social determinants of health
SNOMED: Systematized Medical Nomenclature of Medicine

Edited by A Mavragani; submitted 06.05.23; peer-reviewed by C Clark; comments to author 05.06.23; revised version received 06.07.23; accepted 13.05.24; published 11.02.25.

Please cite as:
Bhavnani SK, Zhang W, Bao D, Raji M, Ajewole V, Hunter R, Kuo YF, Schmidt S, Pappadis MR, Smith E, Bokov A, Reistetter T, Visweswaran S, Downer B
Subtyping Social Determinants of Health in the "All of Us" Program: Network Analysis and Visualization Study
J Med Internet Res 2025;27:e48775
URL: https://www.jmir.org/2025/1/e48775
doi:10.2196/48775
PMID:39932771

©Suresh K Bhavnani, Weibin Zhang, Daniel Bao, Mukaila Raji, Veronica Ajewole, Rodney Hunter, Yong-Fang Kuo, Susanne Schmidt, Monique R Pappadis, Elise Smith, Alex Bokov, Timothy Reistetter, Shyam Visweswaran, Brian Downer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Structured Codes and Free-Text Notes to Measure Information Complementarity in Electronic Health Records: Feasibility and Validation Study

Tom M Seinen[1], MSc; Jan A Kors[1], PhD; Erik M van Mulligen[1], PhD; Peter R Rijnbeek[1], PhD

Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands

Corresponding Author:
Tom M Seinen, MSc
Department of Medical Informatics
Erasmus University Medical Center
Dr. Molewaterplein 40
Rotterdam, 3015 GD
Netherlands
Phone: 31 010 7044122
Email: t.seinen@erasmusmc.nl

Abstract

Background: Electronic health records (EHRs) consist of both structured data (eg, diagnostic codes) and unstructured data (eg, clinical notes). It is commonly believed that unstructured clinical narratives provide more comprehensive information. However, this assumption lacks large-scale validation and direct validation methods.

Objective: This study aims to quantitatively compare the information in structured and unstructured EHR data and directly validate whether unstructured data offers more extensive information across a patient population.

Methods: We analyzed both structured and unstructured data from patient records and visits in a large Dutch primary care EHR database between January 2021 and January 2024. Clinical concepts were identified from free-text notes using an extraction framework tailored for Dutch and compared with concepts from structured data. Concept embeddings were generated to measure semantic similarity between structured and extracted concepts through cosine similarity. A similarity threshold was systematically determined via annotated matches and minimized weighted Gini impurity. We then quantified the concept overlap between structured and unstructured data across various concept domains and patient populations.

Results: In a population of 1.8 million patients, only 13% of extracted concepts from patient records and 7% from individual visits had similar structured counterparts. Conversely, 42% of structured concepts in records and 25% in visits had similar matches in unstructured data. Condition concepts had the highest overlap, followed by measurements and drug concepts. Subpopulation visits, such as those with chronic conditions or psychological disorders, showed different proportions of data overlap, indicating varied reliance on structured versus unstructured data across clinical contexts.

Conclusions: Our study demonstrates the feasibility of quantifying the information difference between structured and unstructured data, showing that the unstructured data provides important additional information in the studied database and populations. The annotated concept matches are made publicly available for the clinical natural language processing community. Despite some limitations, our proposed methodology proves versatile, and its application can lead to more robust and insightful observational clinical research.

(J Med Internet Res 2025;27:e66910) doi:10.2196/66910

KEYWORDS

natural language processing; named entity recognition; clinical concept extraction; machine learning; electronic health records; EHR; word embeddings; clinical concept similarity; text mining; code; free-text; information; electronic record; data; patient records; framework; structured data; unstructured data

XSL•FO
RenderX

Introduction

Electronic health records (EHRs), originally designed for clinical documentation and administration, are now increasingly used in observational research [1,2], supporting various types of studies, including case studies, patient cohort characterizations, and clinical prediction modeling. EHR data are generally recorded in 2 forms: structured and unstructured data. Structured data includes clinical codes for documenting clinical events, such as diagnoses, medications, procedures, and measurements. Structured data is particularly suitable for observational research due to its consistent meaning, tabular format, and standardized vocabulary of codes. Unstructured data consists of free-text clinical notes, which can provide detailed descriptions capturing the nuances of patient care, such as physician observations, patient histories, diagnostic impressions, and discharge summaries. Although rich in contextual information, unstructured data poses challenges for direct analysis because of its variability and lack of standardization [3,4]. Consequently, extracting meaningful information from unstructured data requires significant investment in manual labor, computational resources, and time.

It is commonly assumed that the text data contains more detailed and extensive information than structured data, based on the often-reported claim—grounded in business-related data [5]—that 80% of EHR data is unstructured [3,4,6-8]. While this assumption may hold, its validity is influenced by factors like documentation quality and clinical context. For example, intensive care data is characterized by a high frequency of measurements, while psychiatric care relies more on textual narratives. Many studies explored the added value of information from text by comparing analyses with and without it [9-15], indirectly validating this assumption. However, even if the assumption holds, it initially remains uncertain to what extent the information in the text data matches or complements the structured data.

Understanding the quantity and differences in the information available in structured and unstructured data for a specific database offers several advantages for observational clinical research. First, it aids study design by identifying the most abundant and reliable data types, enabling researchers to formulate feasible hypotheses and research questions. Second, it allows for more effective allocation of human and computational resources by focusing efforts where they are most needed. Third, knowing the balance between structured and unstructured data helps researchers prioritize according to the study's specific needs. Finally, it highlights gaps or unique aspects of the data, facilitating the exploration of underused research opportunities.

Comparing the information from structured and unstructured data involves various measures, such as quantity and content. While structured data points can be counted and unstructured data quantified by individual words or extracted concepts, comparing content similarity is more challenging. The core meaning of both structured codes and unstructured text lies in their semantic content. Evaluating the information distance between 2 concepts or texts requires comparing their semantic meanings, a task commonly addressed in natural language processing through semantic similarity measures. Modern approaches often use word embedding models to generate concept embeddings, which are used in applications like biomedical ontology matching [16,17] and concept normalization [18-20]. Specialized embedding models, such as SapBERT [21] and BioLORD [22], have been developed for this purpose and provide the opportunity to measure the information difference between structured and unstructured data.

Several studies have compared structured and unstructured data for specific clinical variables, such as social and behavioral determinants of health [23,24] and smoking history [25,26]. However, to our knowledge, no research has directly assessed the information differences between structured and unstructured data across all clinical events in a database. Our study aims to quantitatively compare the information coded by general practitioners (GPs) with the information documented in free-text notes, using data from a large Dutch GP database. We extracted clinical concepts from unstructured text and used concept embeddings to calculate their similarity with the structured concepts. After determining a similarity threshold, we estimated the difference and overlap of information between structured and unstructured data.

Methods

Database and Setting

This study used the Integrated Primary Care Information (IPCI) database [27], a longitudinal EHR database of Dutch GPs. IPCI contains records of 2.9 million patients with a median follow-up of 4.8 years, spanning from 1993 to 2024. The database is standardized using the Observational Medical Outcomes Partnership Common Data Model [28]. Eligible participants in our study dataset included all patients recorded in the database from January 2021 to January 2024. The study received approval from the IPCI governance board under code 2023-04.

Methodological Setup

The methods consist of 4 main parts, visualized in Figure 1, each described in detail in the following sections. First, we extracted both structured and unstructured data for each eligible patient and applied a concept extraction framework to extract clinical concepts from the free-text notes. Second, we applied 2 different data grouping methods to the population. Third, we used pretrained multilingual concept embeddings to calculate the similarities between the structured and extracted concepts. Finally, we annotated a sample of concept similarity matches to determine a similarity score threshold and quantified the data similarity in the database.

XSL•FO
RenderX

Figure 1. Visualization of the methodological setup. Steps 1-4 outline the primary process, including data extraction, data point grouping, concept comparison, and data similarity quantification. Steps A, B, and C detail additional processes, specifically the extraction of clinical concepts, the creation of concept embeddings, and the determination of a similarity threshold. EHR: electronic health record.

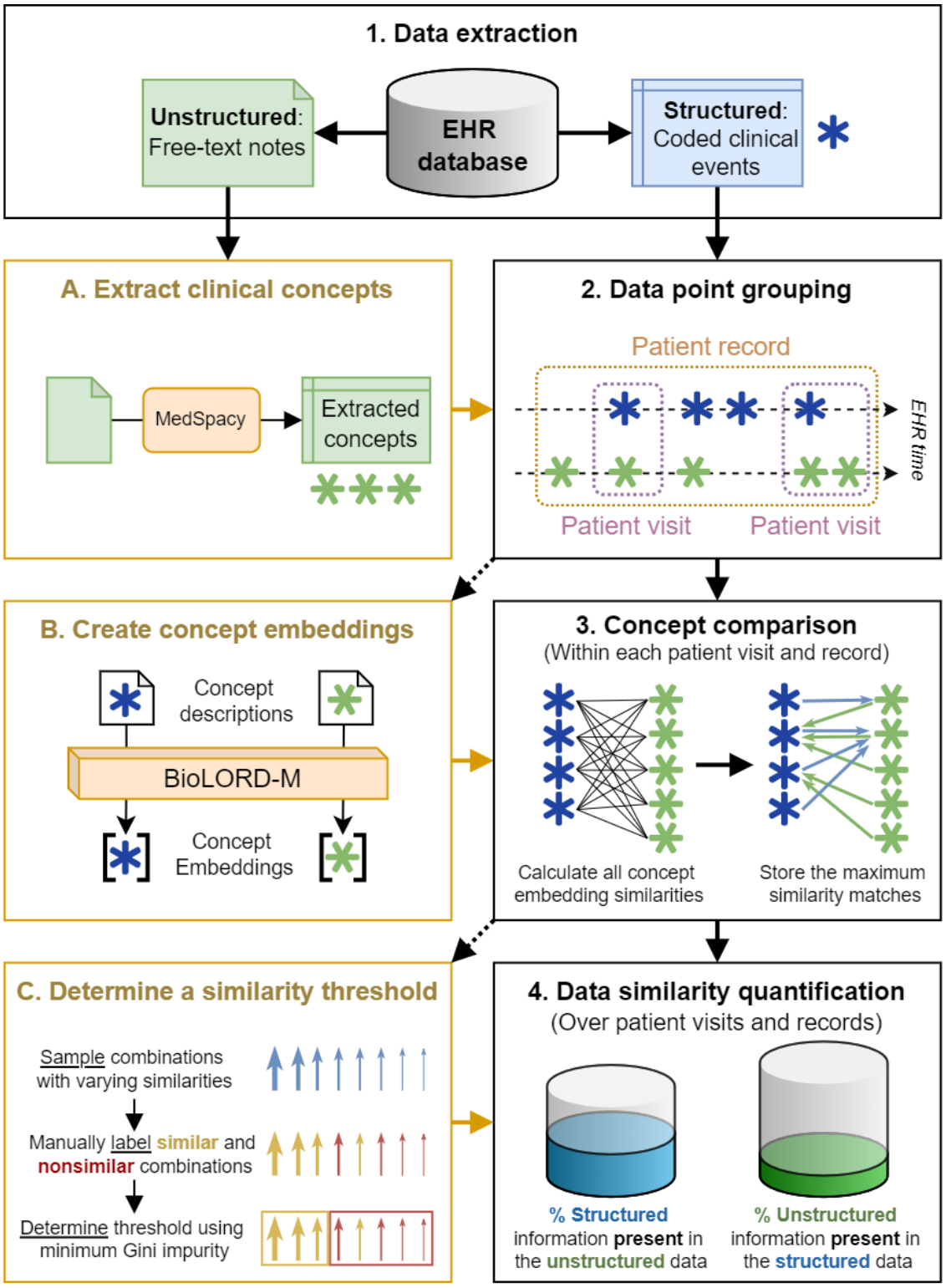

Data Extraction

We extracted all structured and unstructured data points for each eligible patient (step 1 in Figure 1). Structured data includes conditions, procedures, prescriptions, measurements, and observations. The coding systems used, such as the International Classification of Primary Care-1, are listed in Table S1 in Multimedia Appendix 1. Unstructured data consists of 3 types of notes: subjective, objective, assessment, and plan notes from consultations; referral and communication notes with secondary care; and other notes from free-text fields in the EHR system, primarily accompanying condition codes.

Clinical Concept Extraction

Structured coded clinical events are considered single data points, while free-text notes can contain multiple pieces of

XSL•FO
RenderX

information embedded in the narrative. To compare these, we extracted individual data points from unstructured text using clinical named entity recognition and linking, generally known as clinical concept extraction (step A in Figure 1). We used MedSpacy [29], a toolkit that extracts clinical concepts based on a reference thesaurus such as the unified medical language system (UMLS) and detects contextual modifiers using language-specific rules. We used a version of MedSpacy adapted for Dutch [10,30], incorporating all Dutch vocabularies from UMLS and replacing the English Systematized Nomenclature of Medicine Clinical Terms with the Dutch translation [31]. We used Dutch context rules to detect qualifiers such as negation, temporality, and experiencer. These rules were previously validated with an annotated corpus [32]. The extracted UMLS concepts were mapped to concept domains such as observation, condition, measurement, drug, and procedure in the Observational Medical Outcomes Partnership standardized vocabulary, which contains the majority of UMLS vocabularies, for easy comparison with the structured data.

Data Point Grouping

Given the longitudinal nature of GP EHR data, we used 2 grouping methods for comparison: by patient visit and by patient record (step 2 in Figure 1). Grouping by visit considers data points recorded simultaneously during a patient visit, providing a natural basis for comparison, while grouping by record includes additional information recorded outside visits, such as lab results and secondary care communications, allowing for a broader comparison.

Concept Comparison

Each structured or extracted data point is represented by a single clinical concept. We compared the Cartesian product of structured and extracted concepts within each group (step 3 in Figure 1). For example, a visit with 1 coded condition and 1 prescription, along with a subjective, objective, assessment, and plan note containing 5 extracted concepts, results in 10 comparisons. While we could count exact concept matches, 2 issues arise. First, the compared concepts can be very similar but not identical, for example, the International Classification of Primary Care-1 code “Fracture: hand/foot bone” (L74) and the Systematized Nomenclature of Medicine Clinical Terms concept “Closed fracture of hand” (704005005). Second, the concept vocabularies differ, as the GPs choose from a limited set of codes, while concept extraction uses the complete UMLS.

To overcome these issues, we used concept embeddings for fuzzy matching, enabling us to measure semantic similarities and recognize similar concepts. Cosine similarity, which measures the angle between 2 embedding vectors, was used to calculate similarities, ranging from –1 (strongly opposite) to 1 (very similar). For each structured and extracted concept, we stored only the highest similarity match, as we are only interested in finding the most similar concept in the other data type.

Concept Embeddings

We used the BioLORD-2023-M pretrained sentence transformer model [22] to generate the multilingual concept embeddings, as visualized in step B of Figure 1. BioLORD-2023-M is designed to produce meaningful representations for biomedical concepts across multiple languages, including English and Dutch, and its cross-lingual performance has been evaluated by the authors [22]. By inputting the concept description, from either UMLS for the extracted concepts or from the source vocabulary for the structured concepts, the model generates a dense 768-dimensional vector, allowing us to create embeddings for both structured and extracted concepts. For concepts with multiple descriptions or synonyms, we calculated an embedding for each and averaged them to create one comprehensive concept representation. For structured observations and measurements, we included the unit and value in the description to enrich the embedding. The model's multilingual capability enabled us to embed concept descriptions in both Dutch and English within the same latent space.

Similarity Threshold Determination

To quantify the information difference between structured and extracted concepts, we needed to define a threshold for concept similarity. Since concept similarity depends on the nature of the embeddings, we developed a systematic method to determine this threshold, as visualized in step C of Figure 1. First, we randomly selected concept pairs at various similarity levels, ranging from a similarity score of 0.35 to 1, with samples taken at 0.05 intervals. This sampling was done for both structured and extracted concepts across patient visits and records, ensuring each concept domain was represented by sampling 5 concepts per domain. Next, we manually annotated the concept pairs as either similar or nonsimilar. Using these annotations, we determined the threshold at which the weighted Gini impurity [33] of the split between similar and nonsimilar matches was lowest.

Data Similarity Quantification

Using a similarity threshold and the most similar counterpart for each structured and extracted concept, we determined the number of structured concepts found in free-text (structured-to-unstructured) and the number of extracted concepts that were coded in the structured data (unstructured-to-structured), as shown in step 4 of Figure 1. It is important to note that these counts are not reciprocal since we consider the maximum similarity per concept. For example, multiple extracted concepts from a patient visit text may be highly similar to a single structured concept, but we only compare the structured concept to its most similar extracted counterpart to determine its presence in the text, not the frequency. We calculated these counts and their percentages across the entire set of concepts, as well as within different concept domains, to explore domain-specific differences. For extracted concepts, we only included those without context modifiers to focus on the core unmodified concepts, which ensures a higher degree of certainty and a more straightforward, reliable comparison.

Subpopulation Comparison

While observing data similarity across the entire population or all GP visits is insightful, applying this method to smaller subpopulations may provide further detail. We defined 3 subpopulations based on different types of clinical events: visits

XSL•FO
RenderX

for chronic disorder (type 2 diabetes mellitus), acute event (COVID-19 vaccination), and psychological disorder (depression), as detailed in Table S2 in Multimedia Appendix 1. We then quantified the similarity between structured and unstructured data for these subpopulations, similar to the full population.

Results

Population and Data Characteristics

Table 1 presents the population and data characteristics for patient records and visits, including details per observation. Additionally, characteristics are listed for 3 subpopulations. Key differences between patient visits and records are evident. First, each patient has one complete record but can have multiple visits. Second, the total number of structured events and their median values indicate that many events are recorded outside GP visits, as their total numbers are roughly 3 times higher in records. Similarly, the total number of clinical notes is twice as high in records compared to visits alone. Third, more extracted concepts per structured event are generally found in patient visits, but the median number of concepts per note is the same.

Table 1. Population and data characteristics for patient records, patient visits, and 3 subpopulation-specific visits.

Characteristic	Patient records	Patient visits	COVID-19 vaccination visits	Depression visits	Diabetes visits
Total numbers					
Number of patients	1,794,209	1,507,473	62,724	20,519	51,032
Number of observations	1,794,209	13,995,524	84,484	63,957	324,698
Ratio number of observations/number of patients	1	9.28	1.35	3.12	6.36
Number of structured events	132,287,566	41,232,407	281,606	118,351	4,513,756
Number of free-text notes	95,790,377	47,915,131	188,318	216,449	1,253,418
Number of extracted concepts	635,232,943	245,052,660	688,420	1,374,430	5,329,978
Ratio extracted concepts/structured events	4.80	5.94	2.44	11.61	1.18
Statistics per observation					
Percentage male, %	49	49	45	33	51
Median age (years)	39	39	62	49	68
Median number of structured events	33	2	2	1	6
Median number of free-text notes	35	3	2	3	3
Median number of extracted concepts	172	12	5	15	12
Median number of concepts per note	3	3	2	3	3

Characteristics for visits across different clinical events also show clear contrasts. As expected, visits regarding diabetes presented the most visits per patient, while for COVID-19 vaccination, the number of visits per patient was the least. Furthermore, diabetes visits contain much structured and unstructured data, whereas depression visits rely mostly on unstructured data with a large median number of extracted concepts per visit and few structured events. Demographic differences are also notable: diabetes and vaccination populations are similar, but the depression population consists of younger females.

Determining the Similarity Threshold

We annotated 1764 matches in 4 concept samples: structured-to-unstructured and unstructured-to-structured in both patient record and visit populations. For transparency and reproducibility, the annotated samples are available in Table S3 in Multimedia Appendix 2. Figure 2B visualizes some example extracted concepts matched with increasing similarity to the structured concepts SARS-CoV-2 (COVID-19) and hypertension. The weighted Gini impurity calculated over the binary annotated matches at different similarity thresholds is presented in Figure 2A. We found the minimum impurity of each of the 4 samples between thresholds of 0.55 and 0.65. To be conservative, we set the threshold at 0.65 for data similarity quantification.

XSL•FO
RenderX

Figure 2. Analysis across similarity thresholds. (A) The weighted Gini impurity measures the level of impurity in the annotated concept combinations in the 4 samples over the different thresholds. The points indicate the minimum impurity in each sample, and the dashed lines represent their lower bound (0.55) and upper bound (0.65). (B) Examples of extracted concepts matched to the structured concepts of SARS-CoV-2 (COVID-19) and hypertension across different similarity thresholds. (C) The percentage of structured concepts matched the extracted concepts (blue) over the range of similarity thresholds and vice versa (green) for both patient records and visits. The red dashed line in all figures depicts the determined similarity threshold value of 0.65.

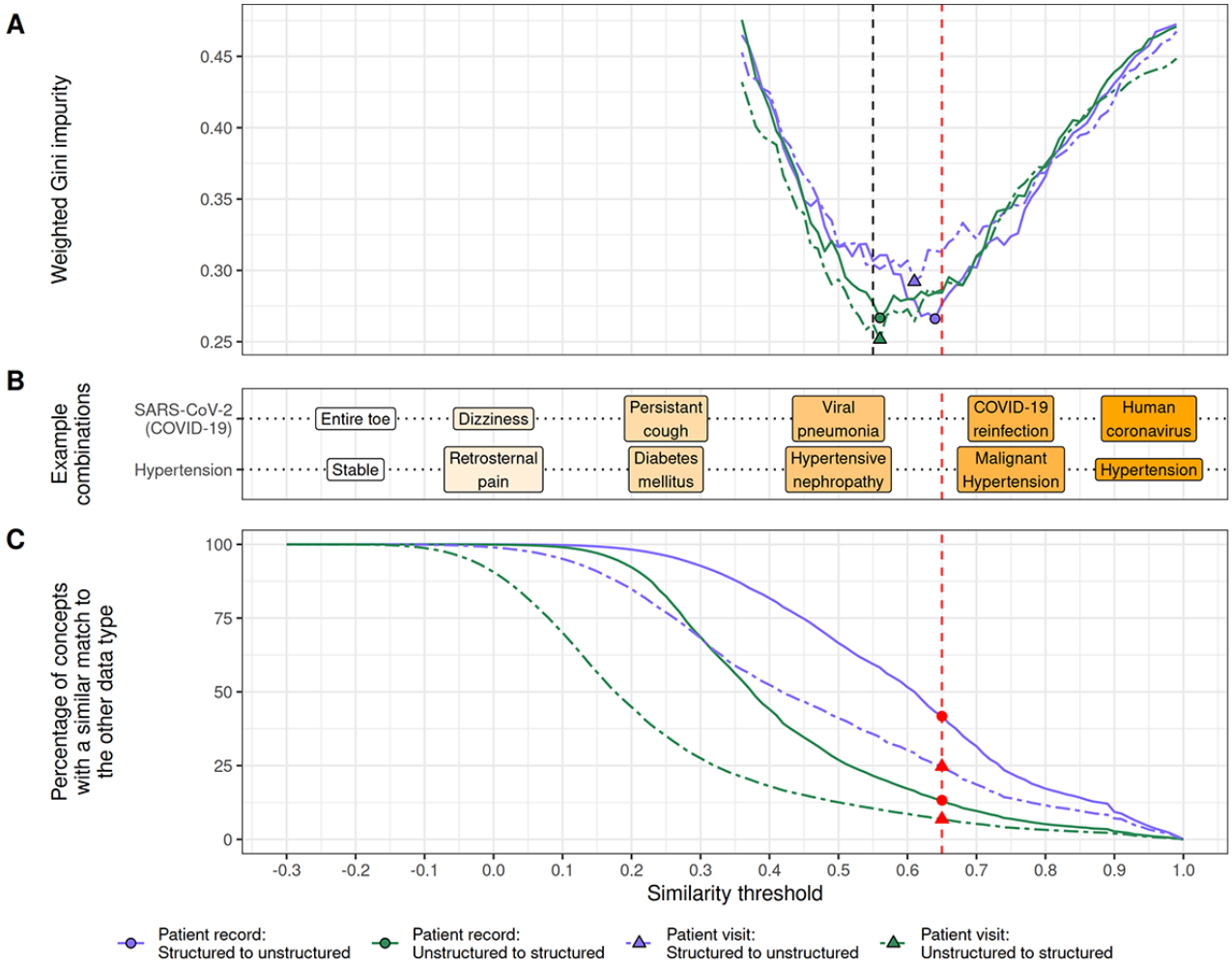

Similarity of Structured and Unstructured Data

The percentage of structured concepts that have a similar match to an extracted concept and vice versa over various similarity thresholds is visualized in Figure 2C for both patient visits and records. At the similarity threshold of 0.65, indicated by the red line in Figure 2, 42% (55.1 million of 132.1 million) of structured concepts in patient records are similar to a concept extracted from text, compared to 25% (9.3 million of 37.8 million) for visits. In contrast, only 13% (66.9 million of 501.9 million) of extracted concepts are similar to a structured concept in patient records, and 7% (11.3 million of 155.7 million) in visits (Figure 3A). This indicates that information in the structured data is more likely to be present in the unstructured data than vice versa. The difference between patient records and visits can be explained by the number of available concepts for matching the entire record versus a single visit.

XSL•FO
RenderX

Figure 3. Comparison of structured and extracted concepts in patient records and visits. (A) Total number of structured (left) and extracted concepts (right), along with the number of concepts matched at a similarity threshold of 0.65 (blue for structured to unstructured data, green for unstructured to structured). The percentage of matched concepts is listed in the chart for each data type. (B) The proportion of each concept domain contributed to the total number of structured (left) and extracted concepts (right), along with the proportion of concepts in each domain matched to the other data type.

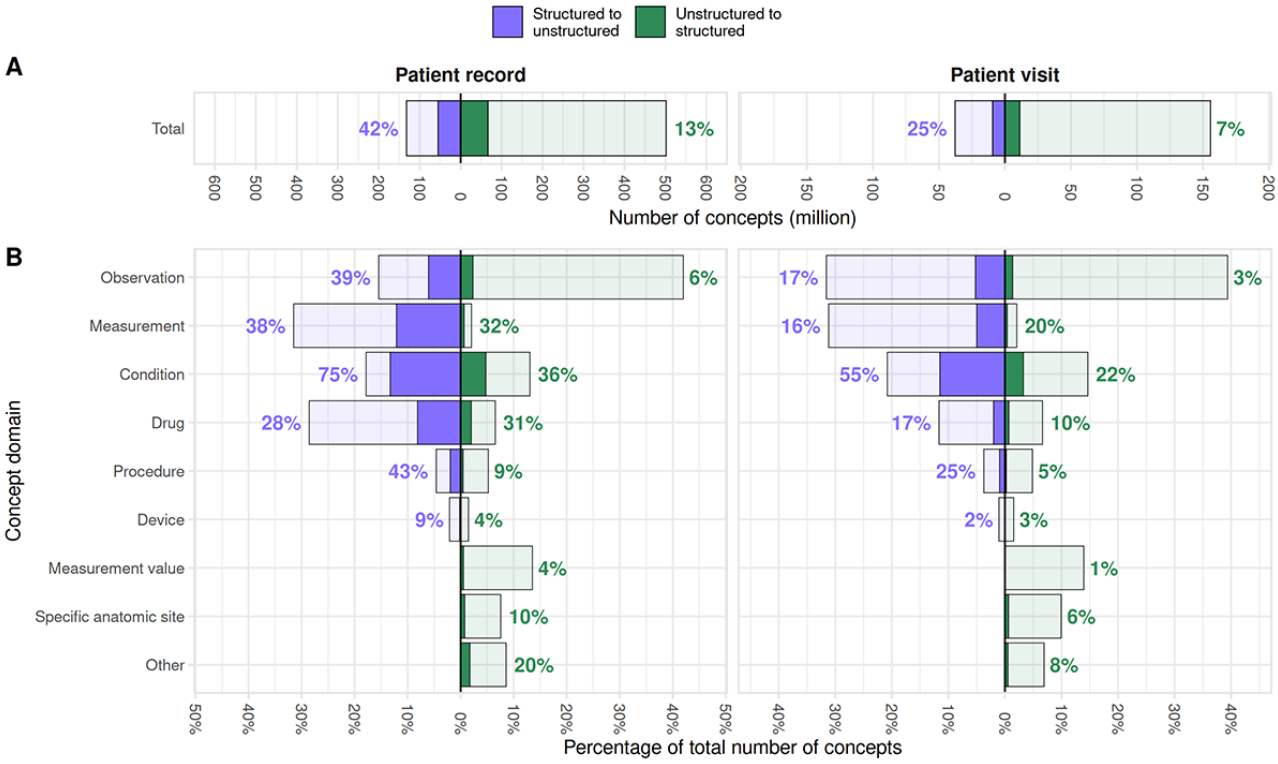

Figure 3B presents the counts and overlap percentages for different domains of structured and extracted concepts. Primarily structured conditions are often also in the text, with 75% (17.6 million of 23.6 million) for patient records and 55% (4.3 million of 7.8 million) for visits. Similarly, extracted condition concepts are also most often matched with structured concepts, with 36% (23.9 million of 65.6 million) in patient records and 22% (5.1 million of 22.8 million) in visits. Other concept domains, such as measurements and drugs, show a relatively high overlap in both structured and unstructured data as well.

Interestingly, differences in concept domain numbers between patient visits and records are larger for structured concepts than extracted concepts. For example, the proportion of structured observation concepts is 31% (11.9 million of 37.8 million) in patient visits but 15% (20.4 million of 132.1 million) in patient records, and the reverse is true for drug concepts. However, this difference is not observed in extracted concepts.

Differences Between Subpopulations

Figure 4A presents the total concept overlap at the selected threshold for the 3 subpopulations, showing ratios of structured and extracted concepts as seen in Table 1. While the proportion of structured data overlapping with unstructured data and vice versa is similar for the vaccination and diabetes populations, the depression population shows a much higher proportion of structured data matched to concepts in the text (43,934/110,317, 40%) and a lower proportion of extracted concepts matched to structured concepts (55,921/1,051,186, 5%). Figure 4B shows different proportions of concept domains in the structured data across the 3 populations. However, in the extracted concepts, the proportions are relatively similar. The data similarity results are consistent with the total population, with conditions, drugs, and measurement concepts showing the highest overlap between structured and unstructured data.

XSL•FO
RenderX

Figure 4. Comparison of structured and extracted concepts in the 3 subpopulation visits. (A) Total number of structured (left) and extracted concepts (right), along with the number of concepts matched at a similarity threshold of 0.65 (blue for structured to unstructured data, green for unstructured to structured). The percentage of matched concepts is listed in the chart for each data type. (B) The proportion of each concept domain contributed to the total number of structured (left) and extracted concepts (right), along with the proportion of concepts in each domain matched to the other data type.

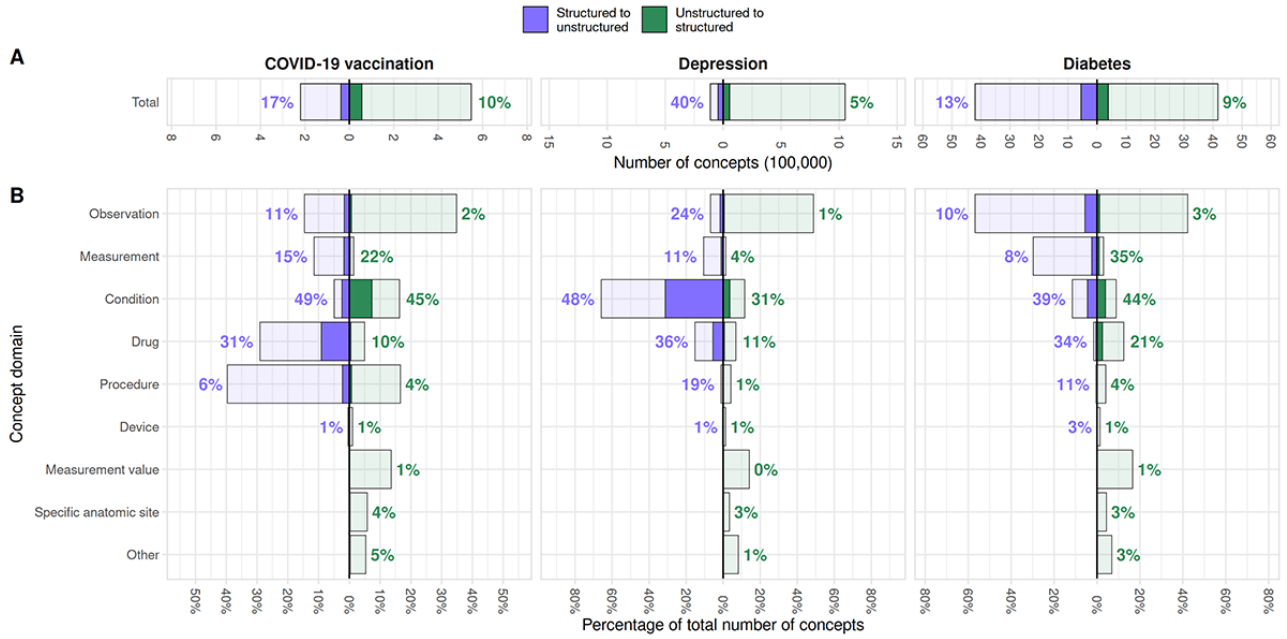

Discussion

Comparing Structured and Unstructured Clinical Data

This study explored the feasibility of quantifying the information difference between structured and unstructured data in a large primary care EHR database. We used concept embeddings to measure the similarity between structured clinical event concepts and concepts extracted from free-text narratives. By systematically determining a similarity threshold, we found that a substantial proportion of structured information is also present in unstructured data, while only a small portion of unstructured information is reflected in structured data. This indicates that most concepts in one data type do not match those in the other, suggesting that the information in structured and unstructured data is highly complementary. The difference between the data modalities can be attributed to certain types of information being exclusively structured, such as measurements made at the general practice, while other information, like observations reported in unstructured communication of a hospital to a GP, is often not captured in the structured data of the GP record.

Furthermore, we observed that condition concepts had the largest overlap between structured and unstructured data, followed by measurements and drug concepts. These results were consistent across different data point grouping methods, by patient record or visit. We also quantified the information difference in smaller subpopulations of patients with specific diseases or procedures. Differences in overlap proportions were evident between these subpopulations, but the concept domains with the largest overlap were the same as in the full population. Overall, our findings validate the often assumed notion that unstructured data contains more extensive information than structured data in this specific database. Most importantly, we prove the feasibility of quantifying the information difference in an EHR database by using a combination of clinical concept extraction methods [29], high-dimensional clinical concept embeddings for similarity measurement [22], and annotating a small number of matched concepts to determine the appropriate similarity threshold.

Strengths and Limitations

The study was conducted on a large, private dataset, which limits public reproducibility. However, our approach to compare and quantify differences between structured and unstructured data is dataset-agnostic. With the detailed pipeline and methodology, the work is adaptable and replicable to any EHR dataset containing both data types. There are several limitations to our methodology. The results depend heavily on 3 factors: the type and quality of concept extraction, the type of concept embeddings used for determining concept similarity, and the chosen similarity threshold. First, the performance of named entity recognition and entity linking is crucial for extracting all relevant information from unstructured data. Imperfect extraction can lead to missed concepts (false negatives) that cannot be matched to structured concepts and wrongly extracted concepts (false positives) that distort the results. The MedSpacy extraction framework used in this work was validated for different languages and demonstrates good performance [29,30]. It was chosen for its availability in Dutch, its versatility with ontologies, and its extraction speed. Second, different types of word and concept embeddings may vary in their ability to distinguish between similar and nonsimilar clinical concepts, affecting the distribution of similarity scores. While other, more complex embeddings might improve similarity discrimination, we only tested the multilingual BioLord-2023-M embedding model [22], as it was specifically trained to produce meaningful representations for clinical sentences and biomedical concepts across different languages, making it perfectly suited to our application. Third, the chosen similarity threshold impacts the results. Setting this threshold can be subjective and heavily depends on the concept embeddings. Therefore, we used a

XSL•FO
RenderX

systematic approach using the minimization of the Gini impurity to determine the threshold, and we published the annotated concept matches for transparency. Lastly, data grouping and the ordering of data points in time pose a challenge. We used 2 grouping methods that ignored the time aspect within each group to explore the differences. However, more sophisticated approaches might be necessary depending on the research question. For instance, data recorded outside patient visits may be assigned to the nearest visit, or more advanced sliding time windows could be used to group the data points.

Overall, the strength of this study lies in enhancing our understanding of the information overlap between structured and unstructured clinical data within a large GP EHR database covering 1.8 million patients. Our methodology of using concept embeddings to calculate the similarity between clinical concepts offers a versatile and language-independent solution, demonstrated here for Dutch, ensuring accurate comparisons by capturing nuances in clinical terminology.

Future Work

Future research in quantifying the data similarity or difference between different data types could explore more advanced concept extraction and embedding techniques or alternative similarity measurements beyond cosine similarity. Our study focused on individual concepts for comparison, but combining multiple concepts might yield higher similarity matches. Investigating the use of concept n-grams and further incorporating context modifiers of the extracted concepts could be beneficial. We applied concept extraction before embedding to retain the granular meaning of individual events within a document. This approach ensures that specific events—such as symptoms, prescriptions, and procedures—are not lost in broader document or sentence embeddings. Future research could explore directly comparing structured concepts and free text without prior clinical concept extraction to enhance the method's reliability and applicability.

We used a single similarity threshold across all concept domains and populations. Future work could involve establishing separate thresholds for each concept domain and population for a more refined comparison. Additionally, considering distinct thresholds for Dutch-to-Dutch and Dutch-to-English concept comparisons might be beneficial, as cross-lingual similarities may vary. The potential applications of our methodology extend beyond the clinical domain. With the appropriate information extraction framework and embedding model, our approach could be adapted to various settings. Conducting specific clinical case studies to demonstrate the benefits of leveraging the data-type differences would be a logical next step for research.

Conclusions

Our study aimed to assess the feasibility of quantifying the information difference between structured and unstructured data in EHR databases. Within a large Dutch primary care EHR database, we successfully demonstrated that unstructured data provides more extensive information than structured data, quantitatively validating this prevailing assumption. By leveraging concept embeddings to measure semantic similarity between structured concepts and those extracted from free-text narratives, we found that a significant portion of structured information is present in unstructured data, while the reverse occurs much less frequently. Notably, concept domains such as conditions, measurements, and drugs exhibited the largest overlap. Despite limitations related to the performance of concept extraction, the type of embeddings used, and the determination of similarity thresholds, our methodology is versatile and was applied across different data grouping methods and subpopulations. The exploration of more sophisticated techniques could further enhance the accuracy and applicability of this approach. We suggest that structured and unstructured data should be used together, as their combined information exceeds that of each data type separately. Understanding the extent and nature of information in structured and unstructured data within a database can enhance study design, research exploration, resource allocation, and data prioritization, ultimately leading to more robust and insightful observational clinical research.

Acknowledgments

This work has received support from the European Health Data & Evidence Network (EHDEN) project. EHDEN has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement number 806968. The JU receives support from the European Union’s Horizon 2020 research and innovation program and EFPIA.

Authors' Contributions

TMS proposed the methodology, designed and implemented the study protocol, and performed the data analysis. JAK, EMvM, and PRR provided critical feedback, helped interpret the results, and shaped the research and analysis. TMS wrote the article with valuable input from all other authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 34 KB - jmir_v27i1e66910_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Table S3: annotated concept matches.
[ZIP File (Zip Archive), 98 KB - jmir_v27i1e66910_app2.zip]

References

1. Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. Ann Rheum Dis 2023 Mar;82(3):306-311 [FREE Full text] [doi: 10.1136/ard-2022-222626] [Medline: 36150748]
2. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc 2018 Aug 01;25(8):969-975 [FREE Full text] [doi: 10.1093/jamia/ocy032] [Medline: 29718407]
3. Assale M, Dui LG, Cina A, Seveso A, Cabitza F. The revival of the notes field: leveraging the unstructured content in electronic health records. Front Med (Lausanne) 2019;6:66 [FREE Full text] [doi: 10.3389/fmed.2019.00066] [Medline: 31058150]
4. Li I, Pan J, Goldwasser J, Verma N, Wong WP, Nuzumlalı MY, et al. Neural natural language processing for unstructured data in electronic health records: a review. Comput Sci Rev 2022 Nov;46:100511. [doi: 10.1016/j.cosrev.2022.100511]
5. Murdoch TB, Detsky AS. The inevitable application of big data to health care. J Am Med Assoc 2013 Apr 03;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]
6. Chiu CC, Wu CM, Chien TN, Kao LJ, Li C, Chu CM. Integrating structured and unstructured EHR data for predicting mortality by machine learning and latent Dirichlet allocation method. Int J Environ Res Public Health 2023 Feb 28;20(5):4340 [FREE Full text] [doi: 10.3390/ijerph20054340] [Medline: 36901354]
7. Hashir M, Sawhney R. Towards unstructured mortality prediction with free-text clinical notes. J Biomed Inform 2020 Aug;108:103489 [FREE Full text] [doi: 10.1016/j.jbi.2020.103489] [Medline: 32592755]
8. Kong HJ. Managing unstructured big data in healthcare system. Healthc Inform Res 2019 Jan;25(1):1-2 [FREE Full text] [doi: 10.4258/hir.2019.25.1.1] [Medline: 30788175]
9. Seinen TM, Fridgeirsson EA, Ioannou S, Jeannetot D, John LH, Kors JA, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. J Am Med Inform Assoc 2022 Jun 14;29(7):1292-1302 [FREE Full text] [doi: 10.1093/jamia/ocac058] [Medline: 35475536]
10. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR. The added value of text from Dutch general practitioner notes in predictive modeling. J Am Med Inform Assoc 2023 Nov 17;30(12):1973-1984 [FREE Full text] [doi: 10.1093/jamia/ocad160] [Medline: 37587084]
11. Kharrazi H, Anzaldi LJ, Hernandez L, Davison A, Boyd CM, Leff B, et al. The value of unstructured electronic health record data in geriatric syndrome case identification. J Am Geriatr Soc 2018 Aug;66(8):1499-1507. [doi: 10.1111/jgs.15411] [Medline: 29972595]
12. Zhang DD, Yin CC, Zeng JC, Yuan XH, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak 2020 Oct 29;20(1):280 [FREE Full text] [doi: 10.1186/s12911-020-01297-6] [Medline: 33121479]
13. Gan S, Kim C, Chang J, Lee DY, Park RW. Enhancing readmission prediction models by integrating insights from home healthcare notes: retrospective cohort study. Int J Nurs Stud 2024 Oct;158:104850. [doi: 10.1016/j.ijnurstu.2024.104850] [Medline: 39024965]
14. Marafino BJ, Park M, Davies JM, Thombley R, Luft HS, Sing DC, et al. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. JAMA Netw Open 2018 Dec 07;1(8):e185097 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.5097] [Medline: 30646310]
15. Tsui FR, Shi L, Ruiz V, Ryan ND, Biernesser C, Iyengar S, et al. Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. JAMIA Open 2021 Jan;4(1):ooab011 [FREE Full text] [doi: 10.1093/jamiaopen/ooab011] [Medline: 33758800]
16. Park J, Kim K, Hwang W, Lee D. Concept embedding to measure semantic relatedness for biomedical information ontologies. J Biomed Inform 2019 Jun;94:103182 [FREE Full text] [doi: 10.1016/j.jbi.2019.103182] [Medline: 31009761]
17. Zhang Y, Wang X, Lai S, He S, Liu K, Zhao J. Ontology matching with word embeddings. Chinese computational linguistics and natural language processing based on naturally annotated big data. : Springer; 2014 Presented at: 13th China National Conference, CCL 2014, and Second International Symposium, NLP-NABD 2014, Wuhan; 2014 October 18-19; China. [doi: 10.1007/978-3-319-12277-9]
18. Abdulnazar A, Kreuzthaler M, Roller R, Schulz S. SapBERT-based medical concept normalization using SNOMED CT. In: Caring is Sharing? Exploiting the Value in Data for Health and Innovation. USA: IOS Press; 2023:825-826.
19. Zahra FA, Kate RJ. Obtaining clinical term embeddings from SNOMED CT ontology. J Biomed Inform 2024 Jan;149:104560 [FREE Full text] [doi: 10.1016/j.jbi.2023.104560] [Medline: 38070816]
20. Xu D, Miller T. A simple neural vector space model for medical concept normalization using concept embeddings. J Biomed Inform 2022 Jun;130:104080 [FREE Full text] [doi: 10.1016/j.jbi.2022.104080] [Medline: 35472514]
21. Liu F, Vuli I, Korhonen A, Collier N. Learning domain-specialised representations for cross-lingual biomedical entity linking. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th

XSL•FO
RenderX

International Joint Conference on Natural Language Processing (Volume 2: Short Papers).: Association for Computational Linguistics; 2021 Presented at: The 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; Aug; Online p. 565-574. [doi: 10.18653/v1/2021.acl-short.72]
22. Remy F, Demuynck K, Demeester T. BioLORD-2023: semantic textual representations fusing large language models and clinical knowledge graph insights. J Am Med Inform Assoc 2024 Sep 01;31(9):1844-1855. [doi: 10.1093/jamia/ocae029] [Medline: 38412333]
23. Hatef E, Rouhizadeh M, Tia I, Lasser E, Hill-Briggs F, Marsteller J, et al. Assessing the availability of data on social and behavioral determinants in structured and unstructured electronic health records: a retrospective analysis of a multilevel health care system. JMIR Med Inform 2019 Aug 02;7(3):e13802 [FREE Full text] [doi: 10.2196/13802] [Medline: 31376277]
24. Bucher BT, Shi J, Pettit RJ, Ferraro J, Chapman WW, Gundlapalli A. Determination of marital status of patients from structured and unstructured electronic healthcare data. AMIA Annu Symp Proc 2019;2020:267-274 [FREE Full text] [Medline: 32308819]
25. Ruckdeschel JC, Riley M, Parsatharathy S, Chamarthi R, Rajagopal C, Hsu HS, et al. Unstructured data are superior to structured data for eliciting quantitative smoking history from the electronic health record. JCO Clin Cancer Inform 2023 Feb;7:e2200155. [doi: 10.1200/cci.22.00155]
26. Wang L, Ruan X, Yang P, Liu H. Comparison of three information sources for smoking information in electronic health records. Cancer Inform 2016 Dec 08;15:CIN.S40604. [doi: 10.4137/cin.s40604]
27. de Ridder MA, de Wilde M, de Ben C, Leyba AR, Mosseveld BM, Verhamme KM, et al. Data resource profile: the Integrated Primary Care Information (IPCI) database, the Netherlands. Int J Epidemiol 2022;51(6):e314-e323 [FREE Full text] [doi: 10.1093/ije/dyac026] [Medline: 35182144]
28. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [doi: 10.3233/978-1-61499-564-7-574] [Medline: 26262116]
29. Eyre H, Chapman AB, Peterson KS, Shi J, Alba PR, Jones MM, et al. Launching into clinical space with medspaCy: a new clinical text processing toolkit in Python. AMIA Annu Symp Proc 2021;2021:438-447 [FREE Full text] [Medline: 35308962]
30. Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR. Annotation-preserving machine translation of English corpora to validate Dutch clinical concept extraction tools. J Am Med Inform Assoc 2024 Aug 01;31(8):1725-1734. [doi: 10.1093/jamia/ocae159] [Medline: 38934643]
31. SNOMED National Release Centre of the Netherlands 2024. URL: https://www.snomed.org/member/netherlands [accessed 2024-09-25]
32. Afzal Z, Pons E, Kang N, Sturkenboom MC, Schuemie MJ, Kors JA. ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. BMC Bioinformatics 2014 Nov 29;15(1):373 [FREE Full text] [doi: 10.1186/s12859-014-0373-3] [Medline: 25432799]
33. Nembrini S, König IR, Wright MN. The revival of the Gini importance? Bioinformatics 2018;34(21):3711-3718 [FREE Full text] [doi: 10.1093/bioinformatics/bty373] [Medline: 29757357]

Abbreviations

EHR: electronic health record
GP: general practitioner
IPCI: Integrated Primary Care Information
UMLS: unified medical language system

Edited by G Tsafnat; submitted 04.10.24; peer-reviewed by F Remy, B Van Es, Y Liu; comments to author 02.11.24; revised version received 22.11.24; accepted 23.11.24; published 13.02.25.

Please cite as:
Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR
Using Structured Codes and Free-Text Notes to Measure Information Complementarity in Electronic Health Records: Feasibility and Validation Study
J Med Internet Res 2025;27:e66910
URL: https://www.jmir.org/2025/1/e66910
doi:10.2196/66910
PMID:

©Tom M Seinen, Jan A Kors, Erik M van Mulligen, Peter R Rijnbeek. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons

XSL•FO
RenderX

Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effect of eHealth Interventions on Body Image of Patients With Cancer: Systematic Review

Guohong Huang[1], RN, BSN; Rongrong Wu[1], RN, BSN; Xiuzhi Xu[2], RN, BSN; Yongxia Song[1], RN, MSN; Rong Zheng[1], RN, BSN; Xi Chen[1], RN, BSN; Jingfang Hong[1], RN, MPHM, PhD

[1]School of Nursing, Anhui Medical University, Hefei, China

[2]Nursing Department, The Second People's Hospital of Fuyang City, Fuyang, China

Corresponding Author:
Jingfang Hong, RN, MPHM, PhD
School of Nursing
Anhui Medical University
No. 81 Mei Shan Road
Shu Shan District
Hefei, 230032
China
Phone: 86 55163869971
Email: hjf1472518@163.com

Abstract

Background: Body image issues are prevalent among individuals diagnosed with cancer, leading to detrimental effects on their physical and psychological recovery. eHealth has emerged as a promising approach for enhancing the body image of patients with cancer.

Objective: The purpose of this study was to evaluate the effectiveness of eHealth interventions on body image and other health outcomes (quality of life, physical symptoms, and emotional distress) among patients with cancer. In addition, the acceptability, engagement, and challenges of eHealth interventions were also assessed.

Methods: A total of 11 databases were searched, encompassing PubMed; Embase; Web of Science; MEDLINE (via Ovid); Scopus; the Cochrane Library; CINAHL (via EBSCO); OpenGrey; and 3 prominent Chinese repositories: China National Knowledge Infrastructure, China Wanfang Database, and China VIP Database. The search dates were from the inception of the database to September 25, 2024. The inclusion criteria for this study encompassed research that used randomized controlled trials (RCTs) and quasi-experiments (QEs) to examine the effectiveness of eHealth interventions for patients with cancer. The methodological quality of RCTs and QEs was evaluated using the Cochrane risk of bias tool and the Joanna Briggs Institute Critical Evaluation Checklist, respectively. The review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was registered with PROSPERO.

Results: There were 3548 studies reviewed, and 7 studies were selected. Three studies were RCTs and 4 were QEs, involving a total of 512 patients. Evidence of efficacy for eHealth interventions targeting body image and other health outcomes (physical symptoms and emotional distress) was mixed. Nevertheless, our findings indicate that there was no notable enhancement in quality of life resulting from eHealth interventions. A total of 5/7 (71%) studies reported the acceptability of eHealth interventions among patients with cancer, and patients perceived eHealth interventions as acceptable. However, the difficulty of operating the software, determination of the most effective course of treatment, and time constraints emerged as the primary challenges associated with electronic interventions.

Conclusions: The implementation of eHealth interventions has the potential to enhance body image, physical symptoms, and emotional distress in patients with cancer. Researchers should undertake more rigorous experiments in the future to elucidate the effectiveness of eHealth and address pertinent concerns.

(J Med Internet Res 2025;27:e55564) doi:10.2196/55564

KEYWORDS

body image; cancer; eHealth; systematic review; quality of life; physical symptom; emotional distress; review; mobile phone

XSL•FO
RenderX

Introduction

Cancer is an escalating global public health issue. Worldwide, the number of cancer diagnoses is swiftly increasing, projected to rise from 19.3 million in 2020 to approximately 21.6 million by 2030 [1,2]. Fortunately, with the innovations of technologies in early diagnosis and cancer treatment, it is more likely that patients with cancer will live longer and have a better prognosis [3]. However, cancer and subsequent treatments may lead to a range of physical changes including, but not limited to, cosmetic changes (eg, hair loss, scarring, and swelling), sensory changes (eg, pain and numbness), and dysfunction (eg, dysphagia, dysarthria, and impotence), which can severely affect the patients of body image [4].

It has been widely acknowledged that an individual's body image is a complex construct encompassing perceptions, thoughts, feelings, and behaviors related to the entire body [4]. Body image disturbance in adult patients with cancer manifests as self-perceived dissatisfaction with physical alterations, functional impairments, and psychological distress resulting from changes in appearance or functionality [5]. According to the American Cancer Society, it is estimated that between 31% and 67% of the approximately 3.1 million survivors of breast cancer in the United States have experienced distress related to body image issues [6]. The prevalence of body image disturbance in patients with head and neck cancer has been reported to be as high as 74% [7]. A study has demonstrated that body image concerns are not limited to patients with breast, head, and neck cancer. These concerns significantly affect a substantial proportion of patients with cancer and persist throughout their long-term survival [4]. These findings are concerning, as prospective research has indicated that poor body image can result in elevated levels of anxiety, depression, and sexual and intimacy concerns, as well as an increased risk of mortality [8,9].

Therefore, there is a growing concern about the importance of body image in the diagnosis and treatment of patients with cancer. According to a report published by Breast Cancer Care, a prominent UK charity, there was an urgent need for increased support for women undergoing breast cancer treatment to address their body image concerns [10]. The international skincare and make-up workshop, "Look Good, Feel Better," offers instructional sessions on makeup techniques to address eyebrow loss and eyelash loss in women [11]. However, traditional face-to-face intervention may struggle to achieve the ideal intervention effect due to time constraints, geographical barriers, lack of medical resources, and high costs. The emergence of eHealth may overcome the shortcomings of these traditional interventions.

eHealth is an accessible health information delivery strategy that provides information and health services through the internet and related technologies [12-14]. eHealth interventions are recognized for their numerous advantages. First, the privacy and confidentiality of patients are ensured during the implementation [15]; second, multiple patients can attend simultaneously and repeatedly, and the fidelity of the intervention is not affected by time constraints and different interveners [15]; and finally, eHealth interventions can be considered as a less time-consuming and cost-effective method of delivering interventions.

Given the diverse advantages of eHealth interventions, it has been gradually applied to address body image issues among patients with cancer, but the findings were still controversial. Sherman et al [16] revealed that web-based psychological intervention, a structured writing exercise, significantly improved body image distress among patients with breast cancer. However, Høyer et al [17] found that teleconferencing did not improve the body image of patients with breast cancer undergoing radiotherapy. A recent investigation indicated that the group using a smartphone app for exercise did not demonstrate a significant improvement in body image among patients with cancer when compared to the traditional exercise group [18]. Therefore, the purpose of this study was to evaluate the effectiveness of eHealth interventions for improving body image among patients with cancer, considering the methodological quality of the most recent clinical trials on the topic, and other secondary outcomes (quality of life, physical symptoms, and emotional distress) were also investigated. Finally, the acceptability, engagement, and challenges of relevant eHealth interventions were discussed.

Methods

Overview

This systematic review has been registered on the PROSPERO platform (CRD42023388898). It was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [19] (protocol for conducting systematic reviews) and following the population, intervention, comparison, outcomes, and study (PICOS) design [20].

Search Strategy

A total of 11 databases were searched, encompassing PubMed; Embase; Web of Science; MEDLINE (via Ovid); Scopus; the Cochrane Library; CINAHL (via EBSCO); OpenGrey; and three prominent Chinese repositories: China National Knowledge Infrastructure, China Wanfang Database, and China VIP Database. The search dates were from the inception of the database to September 25, 2024. The reference lists of pertinent systematic reviews were scrutinized to identify any additional studies that could be incorporated [11,21-26]. The search form included the truncations and synonyms for the following terms: neoplasms, cancer, onco*, internet-delivered, app, eHealth, mHealth, smartphone, telephone, text message*, body image, self-image (Multimedia Appendix 1 provides a comprehensive search strategy for each database).

Inclusion and Exclusion Criteria

The studies that met the following criteria were included in this review: (1) the study population comprised adults aged 18 years and older, who had been diagnosed with cancer and were either undergoing or had completed cancer treatment; (2) the intervention in the studies were related to eHealth interventions, which encompassed various modes of communication such as SMS text messaging, phone, email, app, web, smartphone app,

XSL•FO
RenderX

and videoconferencing; (3) the comparison groups were standard care or control intervention (access to the internet without specific guidance from intervention personnel); (4) the primary outcome of the study was body image, which was assessed using standardized, scientifically validated, and reliable psychometric instruments; (5) the studies design included randomized controlled trial (RCT) or quasi-experimental (QE) studies; and (6) the language of the study was Chinese or English.

Studies were excluded for the following reasons: (1) the experimental group involved a combination of eHealth and face-to-face interventions; (2) body image was used as a secondary outcome; (3) the study included study protocols without outcomes, review, nonclinical study, meta-analysis, and so on; (4) the full text cannot be obtained; and (5) studies were repeated publications.

Study Selection

The research team deliberated and reached a consensus regarding the search terms. Literature was retrieved by 2 researchers (GH and RW) and subsequently imported separately into the document management software NoteExpress (Beijing Aiqin Haile Technology Co, Ltd). Title and abstract screening and full-text screening were performed independently by 2 investigators (GH and RW) in accordance with the same inclusion and exclusion criteria. Any disagreements were initially addressed and resolved by 2 researchers. Otherwise, a third researcher (XX) was involved to ensure agreement was reached. Finally, GH reviewed all studies to determine inclusion or exclusion.

Data Collection Process

Excel data extraction tables were developed according to PRISMA guidelines [19]. The information mainly included (1) general research information: author, year, country, participants, design type, study duration, mean age, and sample size; (2) intervention and control group details: intervention platform, method, intervention, and follow-up time; and (3) outcomes: evaluation tool and primary and secondary outcomes.

Quality Assessment

The methodological quality of each included RCT was evaluated using the Cochrane Risk of Bias (RoB) Tool [27]. Quasi-randomized studies were assessed for RoB using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist [28]. RoB was independently assessed by authors RW (3 studies) and XX (3 studies). The JBI tool was independently evaluated by authors RZ (3 studies) and XC (3 studies). Discrepancies were addressed through collaborative discussion by the review team. Finally, GH checked all assessments to ensure accuracy.

Data Synthesis

The heterogeneity in methods and outcomes precluded the use of meta-analysis. Instead, a narrative overview of the findings from the included studies was presented alongside a tabular summary of the extracted data. Study outcomes were divided into primary outcomes (body image) and other health outcomes (quality of life, physical symptoms, and emotional distress).

Results

Study Characteristics

After eliminating duplicates and conducting a screening process, 7 full-text studies were identified (Figure 1), including a total of 512 participants. The general characteristics of each study are reported in Multimedia Appendix 2 [16,29-34]. In terms of research methodology, 3 (43%) out of the 7 studies were RCTs and the rest (n=4, 57%) were QE studies.

XSL•FO
RenderX

Figure 1. PRISMA flowchart describing study identification and selection process. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

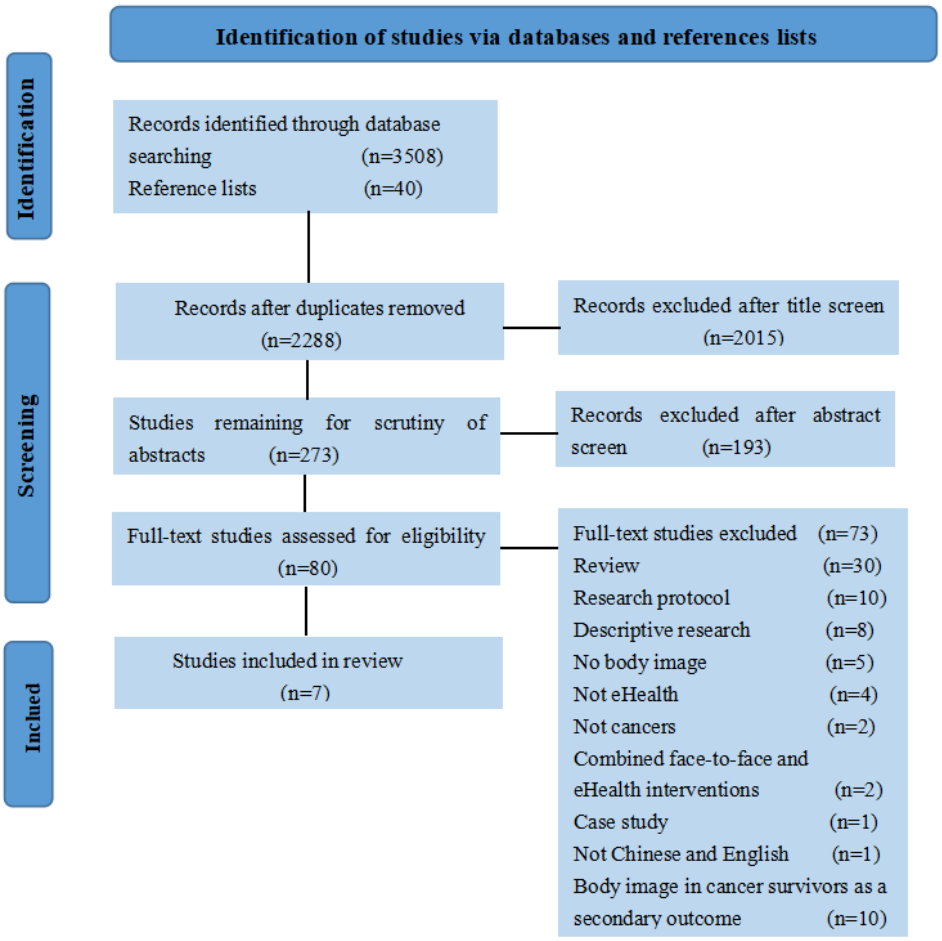

Based on the study population, 4 out of 7 studies were conducted on women diagnosed with breast cancer, accounting for 57% of the total number of studies [16,29-31]. As for study location, the included studies were conducted in Europe (n=1, 14%) [32], the United States (n=2, 29%) [31,33], Australia (n=2, 29%) [16,34], or Asia (n=2, 29%) [29,30].

Overview of eHealth Interventions

A total of 3 (43%) studies were conducted through a website or web-based app [16,30,34]. All others were conducted by SMS text messages, telephone, or video intervention [29,31-33]. The intervention methods included psychological intervention (n=5, 71%) [16,30,32-34], health education intervention (n=1, 14%) [29], and physical exercise intervention (n=1, 14%) [31]. The duration of intervention ranged from 30 minutes to 3 months. The practitioners involved in the intervention involved doctors [16], oncologists [29,34], peer mentors [31], psychologists [29,30,32,33], and health education specialists [29].

Due to the extensive inclusion criteria and resultant heterogeneity of outcomes, we reported outcomes involving 4 subject-related categories under eHealth interventions: body image [16,29-34], quality of life [32], physical symptoms (fatigue [29,31] and sexual function [31]), and emotional distress (anxiety and depression [16,30-32], self-efficacy [30], and self-compassion [16]).

Primary Outcome

All of the study outcomes measured patients' body image [16,29-34]. Nine body image measures, validated in patients with cancer, were used: Body Image Concern Inventory [29], Body Image Scale [16,30-32], Body Mindfulness Questionnaire [32], Body Image States Scale [34], Body Image Coping Skills Inventor [33], IMAGE-HN (Inventory to Measure and Assess image disturbance—Head and Neck) [33], Body Appreciation Scale [16], the Body Appreciation Scale-2 Short Form [34], and Multidimensional Body-Self Relations Questionnaire [34]. The intervention methods included psychological intervention via the internet, app, videoconference, telephone- or text-based health education intervention [16,29,30,32-34], physical exercise intervention using activity monitors, and videoconference [31]. A total of 6 (86%) studies reported a significant effect of the intervention [16,29-31,33,34]. A 3-month, internet-based, peer-moderated physical activity (PA) intervention (Pink Body Spirit) for young patients with breast cancer showed that participants reported an improvement in body image at 3 months after the intervention [31]. A 1-week, web-based brief writing intervention for patients with cancer demonstrated that participants reported a significant improvement in body image 1 week after the intervention [34]. A 7-week app health education intervention was conducted for patients with breast cancer. The results showed that the body image of the intervention group was significantly improved after 7 weeks of intervention [29]. A 6-week, internet-based, mindfulness-based

XSL•FO
RenderX

stress reduction intervention for patients with breast cancer showed significant improvement in body image at 6 weeks after the intervention [30]. A 3-month, internet-based structured writing intervention for patients with breast cancer showed a significant reduction in body image concerns within the intervention group after 3 months [16]. A brief cognitive behavioral intervention based on a video telemedicine platform was conducted for patients with head and neck cancer for 3 months. The results showed that body image was significantly improved after 3 months of intervention [33]. Of them, 4 studies [16,31,33,34] tracked the effects of the intervention at multiple time points (≥2 time points), but not all of these studies achieved the intended.

Other Health Outcomes

Quality of Life

Quality of life was assessed as a secondary outcome in 14% (n=1) of studies. This study used a quality-of-life scale that has been validated in patients with cancer: the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30, with higher scores reflecting better quality of life. This study was based on smartphone-based group physical and psychological therapy, and the treatment effect on quality of life at 5 weeks after the intervention was statistically insignificant [32].

Physical Symptoms

In total, 2 (29%) studies investigated physical symptoms, including fatigue and sexual function. Two studies showed significant reductions in fatigue, including one video remote intervention [31] and one eHealth education intervention [29]. A study focusing on female sexual function indicated that a PA intervention using remote video peer support did not yield improvements in sexual function among young patients with breast cancer [31].

Emotional Distress

A total of 4 (57%) studies investigated the outcome of emotional distress, including anxiety and depression, self-compassion, and self-efficacy [16,30-32]. The results from 4 separate studies evaluated the prevalence of anxiety and depression. Notably, 2 studies indicated that a remote, video-based, peer-led intervention focused on exercise detection and an internet-based, mindfulness-based stress reduction intervention were effective in reducing anxiety symptoms; however, they did not produce significant enhancements in depressive symptoms [30,31]. In contrast, the remaining study reported that the intervention was successful in improving both anxiety and depression levels for participants with breast cancer and lymphedema [16]. Yet another study indicated that smartphone-based physical and psychological interventions did not improve anxiety and depression [32]. A study reported significant improvements in self-compassion with a structured web-based writing exercise intervention [16]. The Internet-Mindfulness-Based Stress Reduction intervention greatly improved self-efficacy [30].

Acceptability

A total of 5 (71%) studies reported the acceptability of eHealth interventions among patients with cancer [16,30-32,34]. Participants rated videoconferencing as acceptable and promising [31]. Participants reported that internet-based teaching methods allowed them to feel more relaxed [30]. A total of 88% of respondents felt that internet-based interventions were worthwhile [16]. Alternatively, participants perceived the intervention via the internet as beneficial [34]. A study indicated that 92.5% of participants expressed a higher level of satisfaction with psychosomatic therapy delivered via smartphone [32].

Engagement

A total of 3 (43%) studies evaluated participants' involvement with the intervention through the completion of modules and use tracking [16,31,32]. A study revealed that 88% of participants in the intervention group completed all 6 components of the web writing intervention [16]. Another study showed that 92.5% of participants participated in at least 4 video sessions of 15 physical smartphone-based interventions [32]. Weiner et al [31] found that more than 85% of people wore an exercise monitoring device on at least 75% of intervention days.

Challenges

A total of 3 (43%) studies reported challenges faced during the intervention process [16,30,31]. A psychosocial intervention program delivered via the internet has indicated that certain older patients with cancer encountered challenges in using the software, such as difficulties in accessing courses and experiencing audio issues during course playback [30]. An RCT reported that the web-based My Changed Body psychosocial intervention was delivered only once, and therefore, it could not be concluded that the optimal number of administrations of the writing for maximum benefit [16]. At the same time, participants in another study said that videoconferencing may take more time to coordinate than email and SMS text messaging because of the potential time conflicts involved in dealing with daily events [31].

RoB Assessment

This study used the Cochrane Collaboration's RoB assessment tool, as well as the JBI tool for quality assessment of RCTs and quasi-randomized studies, respectively. The final results are shown in Figure 2 [16,29,33] and Table 1. Two studies within the RCT were assessed and found to exhibit a low RoB across all 6 evaluated dimensions [16,33]. Selection bias was evaluated as unclear in a study due to insufficient information [29]. Moreover, one study had some detection bias due to a lack of blinding [29]. In addition, regarding quasi-randomized studies, the comparability of baseline data was not clearly articulated for most studies (n=3, 75%) [30-32]. As for the two groups of interventions, only one study explicitly stated that all other measures received by each group were the same [30]. Regarding multivariate measures of outcome indicators, most of the studies (n=3, 75%) implemented diversified measures of outcome indicators before and after the intervention [31,32,34]. With respect to the treatment of follow-up data, all studies reported missing data; however, only one study failed to address this issue with statistical methods [30]. For further details, please refer to the Multimedia Appendices 3 and 4. In addition, the study was performed in accordance with the PRISMA guidelines (Multimedia Appendix 5).

XSL•FO
RenderX

Figure 2. Cochrane risk of bias scores (% low, unclear, and high risk) across bias domains.

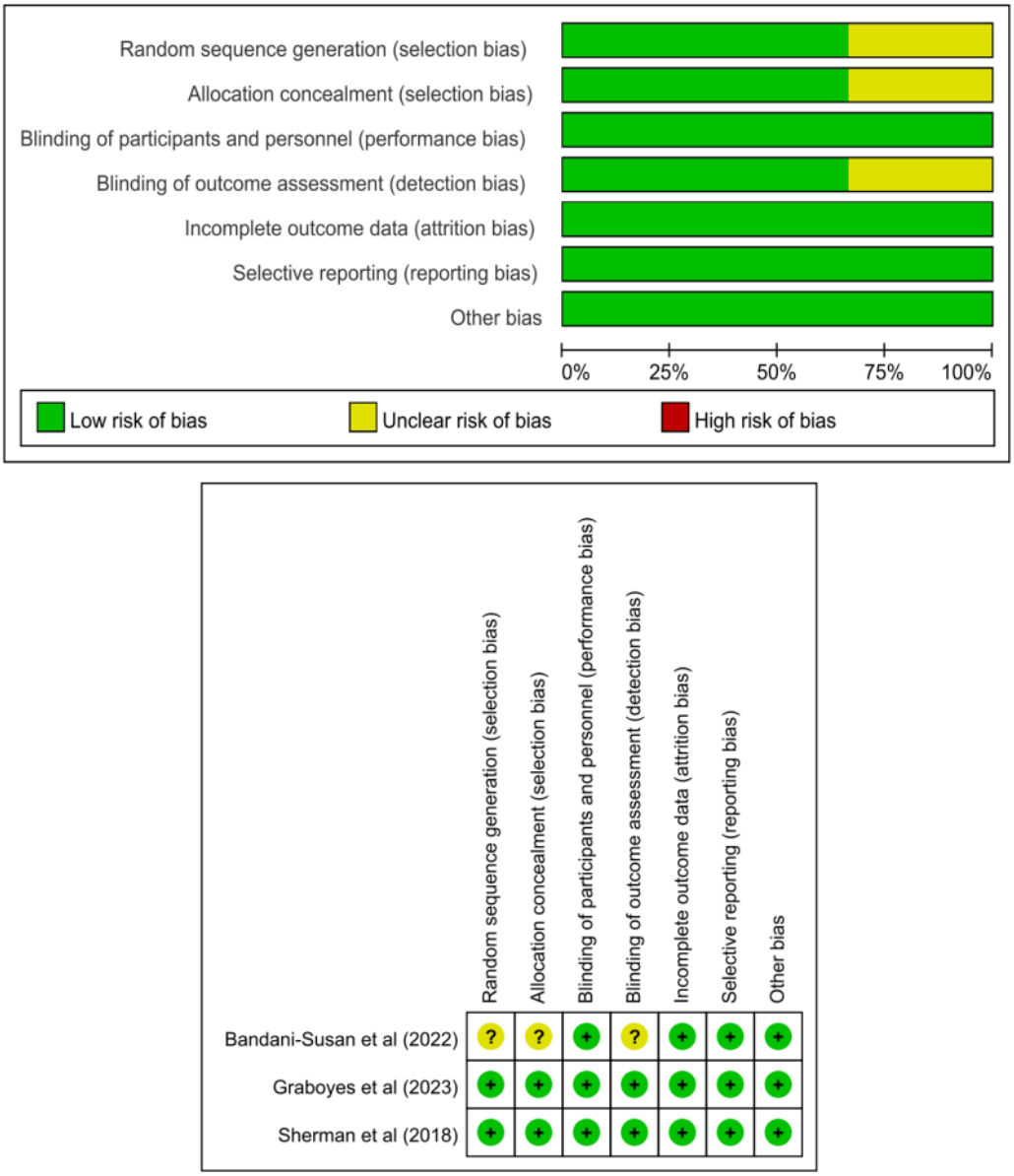

XSL•FO
RenderX

Table 1. Joanna Briggs Institute Critical Appraisal Checklist for quasi-experimental evaluation.

Author (year)	①[a]	②[b]	③[c]	④[d]	⑤[e]	⑥[f]	⑦[g]	⑧[h]	⑨[i]
Weiner et al (2023) [31]	Y[j]	NC[k]	NA[l]	Y	Y	Y	Y	Y	Y
Chang et al (2022) [30]	Y	N[m]	Y	Y	N	N	Y	Y	Y
Brkic et al (2024) [34]	Y	Y	N	Y	Y	Y	Y	Y	Y
Grossert et al (2023) [32]	Y	NC	NA	Y	Y	Y	Y	Y	Y

[a]Is it clear in the study what is the "cause" and what is the "effect" (ie, there is no confusion about which variable comes first)?

[b]Were the participants included in any similar comparisons?

[c]Were the participants included in any comparisons receiving similar treatment/care other than the exposure or intervention of interest?

[d]Was there a control group?

[e]Were there multiple measurements of the outcome both before and after intervention/exposure?

[f]Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed?

[g]Were the outcomes of participants included in any comparisons measured in the same way?

[h]Were outcomes measured in a reliable way?

[i]Was appropriate statistical analysis used?

[j]Y: yes.

[k]NC: not clear.

[l]NA: not applicable.

[m]N: no.

Discussion

Principal Findings

In light of the growing prevalence of eHealth interventions, their application in patients with cancer has been on the rise. To our knowledge, this study was the first systematic review to evaluate the effects of eHealth interventions on body image in individuals diagnosed with cancer. This study aimed to scrutinize published research on eHealth interventions for body image in patients with cancer and critically evaluate the effectiveness of these interventions. In addition, other secondary outcomes (quality of life, physical symptoms, and emotional distress) also were explored. Finally, the acceptance, engagement, and challenges faced by patients with cancer during eHealth intervention were also assessed.

From the perspective of research characteristics, the majority(4/7, 57%) of studies involved female patients with breast cancer, possibly attributed to the higher incidence rate of breast cancer in women or their generally heightened concern toward appearance and body image compared to men [35]. Recently, body image interventions in patients with prostate cancer and head and neck cancer have also emerged [36,37], indicating that body image problems in patients with other cancer types are also receiving scholars' attention. We observed that the majority(5/7, 71%) of the studies incorporated in this review were conducted in high-income countries, probably because high-income countries have a higher penetration of smartphones and computers and more stable networks, which was more conducive to the development of research. According to 2023 International Telecommunication Union statistics, there has been a noticeable increase in mobile network use in low- and middle-income countries [38]. Therefore, it is not difficult to carry out research based on eHealth interventions in limited-income countries in the future.

The intervention outcomes, encompassing body image, physical symptoms, and emotional distress, yielded mixed results. The eHealth-based interventions in this study included psychological intervention [16,30,32-34], health education intervention [29], and physical exercise intervention [31]. Our analysis revealed that a majority (6/7, 86%) of the eHealth intervention studies demonstrated substantial effects on body image. In previous studies, face-to-face or group psychotherapy, psychoeducation, and PA have shown positive effects on the perception of body imagery among patients with cancer, but limitations in time, medical resources, and geographic location make it difficult to implement widespread interventions [11,39]. However, these problems were well addressed using electronic information interventions [40]. Although this study found that eHealth-based interventions can have a positive effect, the effect of the interventions waned as the duration of follow up increased. Given the availability of resources, it is recommended that future research explore the most serious time points of body image distress in patients with cancer and that targeted interventions may achieve the best results at the lowest possible economic cost. In addition, we demonstrated that a group-based physical and psychological intervention significantly enhances the appreciation of body awareness among survivors of cancer [32]. Therefore, we suspect that resources permitting, future studies may achieve better results if multiple intervention modalities are implemented simultaneously. Given that only a single study within the reviewed literature assessed the quality of life, the findings indicated that the enhancement in quality of life resulting from eHealth interventions did not reach statistical significance [32]. The limited follow-up period may be a contributing factor to the lack of significant differences observed

XSL•FO
RenderX

in the study. Given that the duration of an intervention is a crucial determinant of its effectiveness, it is essential to carefully consider this aspect during the design phase of the program. A large, cross-country RCT study demonstrated the significant benefits of remote monitoring in mitigating symptom management and emotional distress in patients with cancer [41]. However, in this study, the outcomes of interventions targeting physical symptoms and emotional distress had mixed results. These studies did not specify which component was most influential in alleviating symptoms. The effective intervention of emotional distress was mainly based on the positive perspective to provide psychological support for patients [30,31].

From the viewpoint of attitudes toward eHealth interventions, a majority(37/40, 92.5%) of patients with cancer expressed satisfaction, acceptance, and perceived value and welcomed these interventions. However, only 3 studies reported patient participation in intervention completion [16,31,32], and none of the included studies examined the economic costs of the intervention. This suggests that future research should delve into strategies for sustaining the appeal of intervention programs while ensuring patient participation, taking into account economic costs. In addition, the study explored the challenges of eHealth interventions, including the difficulty of operating the software, determination of the most effective course of treatment, and time conflicts. Some implications for future research can be obtained. First, it is suggested that multidisciplinary cooperation is needed in eHealth intervention to evaluate the physical and mental conditions of patients with cancer and formulate reasonable and scientific intervention programs. Second, eHealth interventions may pose a challenge for patients with cancer in areas with poorly developed network communications. Therefore, robust technical support and patient education provided by professional staff are essential. Finally, considering the time constraints that patients with cancer often encounter, eHealth interventions can offer greater flexibility, enabling patients to engage in them during their leisure time.

Study Limitations

The study findings were meticulously reviewed and reported in strict adherence to the PRISMA guidelines. However, the study does have certain limitations. First, although we conducted a comprehensive search as much as possible, some studies may have been missed due to language limitations. Second, the heterogeneity in intervention methods, frequency, timing, and assessment tools among the included studies precludes the application of a meta-analysis program for synthesizing their results. Consequently, narrative data synthesis was deemed to be the most suitable approach. Finally, among the 7 studies included in this study, 4 studies were QEs, and the sample sizes of 2 studies were relatively modest, which may affect the research conclusion.

Conclusions

In conclusion, this research indicates that eHealth interventions may be beneficial for improving body image, physical symptoms, and emotional distress in individuals with cancer. To further validate their effectiveness, future high-quality RCTs are warranted. In the future, more multidisciplinary teams are needed to explore the availability, acceptability, effectiveness, and cost of different eHealth interventions to improve the health outcomes of patients with cancer.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant/award: 82272926), the Major Humanities and Social Sciences Projects of Anhui Provincial Universities (grant/award: SK2020ZD13), and the 2023 Graduate Seedling Cultivation Project of the School of Nursing at Anhui Medical University (grant/award: hlqm12023034).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for each database.
[DOCX File , 18 KB - jmir_v27i1e55564_app1.docx]

Multimedia Appendix 2
Summary of a systematic review of electronic health (eHealth) interventions on body image in patients with cancer.
[DOCX File , 157 KB - jmir_v27i1e55564_app2.docx]

Multimedia Appendix 3
Cochrane risk-of-bias supporting evidence for randomized controlled trials.
[DOCX File , 16 KB - jmir_v27i1e55564_app3.docx]

Multimedia Appendix 4
Joanna Briggs Institute risk-of-bias supporting evidence for quasi-experiments.
[DOCX File , 16 KB - jmir_v27i1e55564_app4.docx]

XSL•FO
RenderX

Multimedia Appendix 5
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for effect of electronic health interventions for the body image of patients with cancer.
[DOCX File , 15 KB - jmir_v27i1e55564_app5.docx]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Cancer prevention and control in the context of an integrated approach. World Health Organization. 2016. URL: https://apps.who.int/gb/ebwha/pdf_files/EB140/B140_31-en [accessed 2023-09-30]
3. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018;391(10125):1023-1075 [FREE Full text] [doi: 10.1016/S0140-6736(17)33326-3] [Medline: 29395269]
4. Fingeret MC, Teo I, Epner DE. Managing body image difficulties of adult cancer patients: lessons from available research. Cancer 2014;120(5):633-641 [FREE Full text] [doi: 10.1002/cncr.28469] [Medline: 24895287]
5. Rhoten BA. Body image disturbance in adults treated for cancer—a concept analysis. J Adv Nurs 2016;72(5):1001-1011. [doi: 10.1111/jan.12892] [Medline: 26748811]
6. Body image: a focus in breast cancer care. American Cancer Society. 2016. URL: https://www.cancer.org/latest-news/body-image-a-focus-in-breast-cancer-care.html [accessed 2023-09-30]
7. Fingeret MC, Yuan Y, Urbauer D, Weston J, Nipomnick S, Weber R. The nature and extent of body image concerns among surgically treated patients with head and neck cancer. Psychooncology 2012;21(8):836-844 [FREE Full text] [doi: 10.1002/pon.1990] [Medline: 21706673]
8. Cousson-Gélie F, Bruchon-Schweitzer M, Dilhuydy JM, Jutand MA. Do anxiety, body image, social support and coping strategies predict survival in breast cancer? A ten-year follow-up study. Psychosomatics 2007;48(3):211-216. [doi: 10.1176/appi.psy.48.3.211] [Medline: 17478589]
9. Lam WWT, Li WWY, Bonanno GA, Mancini AD, Chan M, Or A, et al. Trajectories of body image and sexuality during the first year following diagnosis of breast cancer and their relationship to 6 years psychosocial outcomes. Breast Cancer Res Treat 2012;131(3):957-967. [doi: 10.1007/s10549-011-1798-2] [Medline: 21971728]
10. My body, myself: altered body image, intimacy and sex after breast cancer. Breast Cancer Care. 2014. URL: https://breastcancernow.org/sites/default/files/files/my-body-myself-report.pdf [accessed 2023-09-30]
11. Lewis-Smith H, Diedrichs PC, Rumsey N, Harcourt D. Efficacy of psychosocial and physical activity-based interventions to improve body image among women treated for breast cancer: a systematic review. Psychooncology 2018;27(12):2687-2699. [doi: 10.1002/pon.4870] [Medline: 30161285]
12. Dorri S, Asadi F, Olfatbakhsh A, Kazemi A. A systematic review of electronic health (eHealth) interventions to improve physical activity in patients with breast cancer. Breast Cancer 2020;27(1):25-46. [doi: 10.1007/s12282-019-00982-3] [Medline: 31187411]
13. Singleton AC, Raeside R, Hyun KK, Partridge SR, Di Tanna GL, Hafiz N, et al. Electronic health interventions for patients with breast cancer: systematic review and meta-analyses. J Clin Oncol 2022;40(20):2257-2270 [FREE Full text] [doi: 10.1200/JCO.21.01171] [Medline: 35500200]
14. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: a systematic review. Cancer Treat Rev 2019;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]
15. Leykin Y, Thekdi SM, Shumay DM, Muñoz RF, Riba M, Dunn LB. Internet interventions for improving psychological well-being in psycho-oncology: review and recommendations. Psychooncology 2012 Sep;21(9):1016-1025 [FREE Full text] [doi: 10.1002/pon.1993] [Medline: 21608075]
16. Sherman KA, Przezdziecki A, Alcorso J, Kilby CJ, Elder E, Boyages J, et al. Reducing body image-related distress in women with breast cancer using a structured online writing exercise: results from the my changed body randomized controlled trial. J Clin Oncol 2018 Jul 01;36(19):1930-1940 [FREE Full text] [doi: 10.1200/JCO.2017.76.3318] [Medline: 29688834]
17. Høyer BB, Toft GV, Debess J, Ramlau-Hansen CH. A nurse-led telephone session and quality of life after radiotherapy among women with breast cancer: a randomized trial. Open Nurs J 2011;5:31-37 [FREE Full text] [doi: 10.2174/1874434601105010031] [Medline: 21660182]
18. Uhm KE, Yoo JS, Chung SH, Lee JD, Lee I, Kim JI, et al. Effects of exercise intervention in breast cancer patients: is mobile health (mHealth) with pedometer more effective than conventional program using brochure? Breast Cancer Res Treat 2017;161(3):443-452. [doi: 10.1007/s10549-016-4065-8] [Medline: 27933450]
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. J Clin Epidemiol 2021;134:178-189 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.03.001] [Medline: 33789819]

XSL•FO
RenderX

20. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014;14(1):579 [FREE Full text] [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
21. Lv M, Luo X, Chen Y. Systematic review on electronic health interventions for patients with breast cancer: revisiting the methodology. J Clin Oncol 2023;41(1):143-144. [doi: 10.1200/JCO.22.01061] [Medline: 36122308]
22. Huang Y, Li Q, Zhou F, Song J. Effectiveness of internet-based support interventions on patients with breast cancer: a systematic review and narrative synthesis. BMJ Open 2022;12(5):e057664 [FREE Full text] [doi: 10.1136/bmjopen-2021-057664] [Medline: 35641011]
23. Ramírez-Vélez R, Zambom-Ferraresi F, García-Hermoso A, Kievisiene J, Rauckiene-Michealsson A, Agostinis-Sobrinho C. Evidence-based exercise recommendations to improve mental wellbeing in women with breast cancer during active treatment: a systematic review and meta-analysis. Cancers (Basel) 2021;13(2):264 [FREE Full text] [doi: 10.3390/cancers13020264] [Medline: 33445739]
24. Gong J, Hu C, Chen M, Cao Q, Li Q. Interventions to improve self-efficacy in colorectal cancer patients and/or caregivers: a systematic review and meta-analysis. J Oncol 2021;2021:4553613 [FREE Full text] [doi: 10.1155/2021/4553613] [Medline: 34707659]
25. Chen YY, Guan BS, Li ZK, Li XY. Effect of telehealth intervention on breast cancer patients' quality of life and psychological outcomes: a meta-analysis. J Telemed Telecare 2018;24(3):157-167. [doi: 10.1177/1357633X16686777] [Medline: 28081664]
26. Harder H, Parlour L, Jenkins V. Randomised controlled trials of yoga interventions for women with breast cancer: a systematic literature review. Support Care Cancer 2012;20(12):3055-3064. [doi: 10.1007/s00520-012-1611-8] [Medline: 23052917]
27. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
28. Critical appraisal checklist for quasi_x0002_experimental studies (non-randomized experimental studies). The Joanna Briggs Institute (JBI). 2017. URL: https://reviewersmanual.joannabriggs.org/ [accessed 2023-09-30]
29. Bandani-Susan B, Montazeri A, Haghighizadeh MH, Araban M. The effect of mobile health educational intervention on body image and fatigue in breast cancer survivors: a randomized controlled trial. Ir J Med Sci 2022;191(4):1599-1605. [doi: 10.1007/s11845-021-02738-5] [Medline: 34370166]
30. Chang YC, Chiu CF, Wang CK, Wu CT, Liu LC, Wu YC. Short-term effect of internet-delivered mindfulness-based stress reduction on mental health, self-efficacy, and body image among women with breast cancer during the COVID-19 pandemic. Front Psychol 2022;13:949446 [FREE Full text] [doi: 10.3389/fpsyg.2022.949446] [Medline: 36389600]
31. Weiner LS, Nagel S, Su HI, Hurst S, Levy SS, Arredondo EM, et al. A remotely delivered, peer-led intervention to improve physical activity and quality of life in younger breast cancer survivors. J Behav Med 2023;46(4):578-593 [FREE Full text] [doi: 10.1007/s10865-022-00381-8] [Medline: 36479658]
32. Grossert A, Meffert C, Hess V, Rochlitz C, Pless M, Hunziker S, et al. Group-based body psychotherapy improves appreciation of body awareness in post-treatment cancer patients: a non-randomized clinical trial. Front Psychol 2023;14:956493 [FREE Full text] [doi: 10.3389/fpsyg.2023.956493] [Medline: 37089722]
33. Graboyes EM, Kistner-Griffin E, Hill EG, Maurer S, Balliet W, Williams AM, et al. Mechanism underlying a brief cognitive behavioral treatment for head and neck cancer survivors with body image distress. Support Care Cancer 2023;32(1):32. [doi: 10.1007/s00520-023-08248-7] [Medline: 38102496]
34. Brkic E, Prichard I, Daly A, Dudley S, Beatty L. Testing the efficacy of a brief online writing intervention on body image and distress in female cancer survivors. Patient Educ Couns 2024;127:108356 [FREE Full text] [doi: 10.1016/j.pec.2024.108356] [Medline: 38944983]
35. Annunziata MA, Giovannini L, Muzzatti B. Assessing the body image: relevance, application and instruments for oncological settings. Support Care Cancer 2012;20(5):901-907. [doi: 10.1007/s00520-011-1339-x] [Medline: 22160547]
36. Langelier DM, D'Silva A, Shank J, Grant C, Bridel W, Culos-Reed SN. Exercise interventions and their effect on masculinity, body image, and personal identity in prostate cancer—a systematic qualitative review. Psychooncology 2019;28(6):1184-1196. [doi: 10.1002/pon.5060] [Medline: 30875710]
37. Graboyes EM, Maurer S, Park Y, Marsh CH, McElligott JT, Day TA, et al. Evaluation of a novel telemedicine-based intervention to manage body image disturbance in head and neck cancer survivors. Psychooncology 2020;29(12):1988-1994 [FREE Full text] [doi: 10.1002/pon.5399] [Medline: 32350999]
38. World Telecommunication/ICT indicators database (version September 2023). International Telecommunications Union (ITU). URL: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx [accessed 2023-11-12]
39. Sebri V, Durosini I, Triberti S, Pravettoni G. The efficacy of psychological intervention on body image in breast cancer patients and survivors: a systematic-review and meta-analysis. Front Psychol 2021;12:611954 [FREE Full text] [doi: 10.3389/fpsyg.2021.611954] [Medline: 33732184]
40. Esplen MJ, Trachtenberg L. Online interventions to address body image distress in cancer. Curr Opin Support Palliat Care 2020;14(1):74-79. [doi: 10.1097/SPC.0000000000000484] [Medline: 31895065]

XSL•FO
RenderX

41. Maguire R, McCann L, Kotronoulas G, Kearney N, Ream E, Armes J, et al. Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART). BMJ 2021;374:n1647 [FREE Full text] [doi: 10.1136/bmj.n1647] [Medline: 34289996]

Abbreviations

JBI: Joanna Briggs Institute
PA: physical activity
PICOS: population, intervention, comparison, outcomes, and study
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QE: quasi-experiment
RCT: randomized controlled trial
RoB: risk of bias

Edited by G Greco; submitted 16.12.23; peer-reviewed by H Huang, L Beatty; comments to author 12.04.24; revised version received 02.10.24; accepted 17.10.24; published 09.01.25.

Please cite as:
Huang G, Wu R, Xu X, Song Y, Zheng R, Chen X, Hong J
Effect of eHealth Interventions on Body Image of Patients With Cancer: Systematic Review
J Med Internet Res 2025;27:e55564
URL: https://www.jmir.org/2025/1/e55564
doi:10.2196/55564
PMID:

©Guohong Huang, Rongrong Wu, Xiuzhi Xu, Yongxia Song, Rong Zheng, Xi Chen, Jingfang Hong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Changing User Experience of Wearable Activity Monitors Over 7 Years: Repeat Cross-Sectional Survey Study

Darcy Beckett[1], BClinExPhys; Rachel Curtis[1], PhD; Kimberley Szeto[1], BClinExPhys; Carol Maher[1], PhD

Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

Corresponding Author:
Carol Maher, PhD
Alliance for Research in Exercise, Nutrition and Activity
Allied Health and Human Performance
University of South Australia
GPO Box 2471
Adelaide, 5000
Australia
Phone: 61 08 8302 ext 6558
Email: carol.maher@unisa.edu.au

Abstract

Background: Lifestyle behaviors, including physical inactivity, sedentary behavior, poor sleep, and unhealthy diet, significantly impact global population health. Wearable activity trackers (WATs) have emerged as tools to enhance health behaviors; however, their effectiveness and continued use depend on their user experience.

Objective: This study aims to explore changes in user experiences, preferences, and perceived impacts of WATs from 2016 to 2023.

Methods: We conducted a cross-sectional online survey among an international cohort of adults (n=475, comprising 387 current and 88 former WAT users). Results were compared with a 2016 cross-sectional online survey (n=237, comprising 200 current and 37 former WAT users) using descriptive statistics and chi-square tests. The survey examined brand preference, feature usefulness, motivations, perceived health behavior change, social sharing behaviors, and technical issues.

Results: In 2023, Apple (210/475, 44%) and Fitbit (101/475, 21%) were the most commonly used devices, compared with the 2016 survey where Fitbit (160/237, 68%) and Garmin devices (39/237, 17%) were most common. The median usage duration in 2023 was 18 months, significantly longer than the 7 months reported in 2016, with most users planning ongoing use. Users in both survey years reported greater improvements in physical activity than diet or sleep, despite lower improvement in physical activity in 2023 compared with 2016, contrasted with greater perceived improvements in diet and sleep. Social media sharing of WAT data notably rose to 73% (283/387) in 2023 from 35% (70/200) in 2016. However, reports of technical issues and discomfort increased, alongside a decrease in overall positive experiences. There was also a noticeable shift in discontinuation reasons, from having learned everything possible in 2016 to dissatisfaction in 2023.

Conclusions: The study highlights significant shifts in WAT usage, including extended use and evolving preferences for brands and features. The rise in social media sharing indicates a deeper integration of WATs into everyday life. However, user feedback points to a need for enhanced design and functionality despite technological progress. These findings illustrate WAT's potential in health promotion, emphasizing the need for user-focused design in diverse populations to fully realize their benefits in enhancing health behaviors.

(J Med Internet Res 2025;27:e56251) doi:10.2196/56251

KEYWORDS

cross-sectional survey; activity tracker; user experience; physical activity; sleep; diet; health behaviour; wearable activity trackers; cohort; Apple; Fitbit; preferences; reliability; accessibility; lifestyle; mobile phone

XSL•FO
RenderX

Introduction

Lifestyle behaviors such as physical inactivity, sedentary behavior, poor sleep quality, and unhealthy dietary habits collectively contribute significantly to global morbidity and mortality [1-4]. Current estimates suggest that 1.4 billion adults, accounting for 27.5% of the global adult population, do not adhere to recommended physical activity guidelines [1], increasing the risk of major noncommunicable diseases, including coronary heart disease, stroke, type 2 diabetes, cancers, and mental illness [5]. Sedentary behavior, which is considered a distinct health risk factor, also contributes to adverse health outcomes [6]. In addition, inadequate sleep [7] and suboptimal dietary patterns [8] are recognized as critical determinants of poor health, influencing both physical and mental well-being. Together, these lifestyle factors form a quartet of behaviors that significantly impact global health, underscoring the need for effective interventions and public health strategies. Observation of daily living research emphasizes the importance of individual's understanding of their everyday activities and their impact on their health [9]. Wearable activity trackers (WATs) are one method of tracking activity, providing health insights and highlighting the potential for targeted interventions to improve health and well-being.

WATs have emerged as promising tools for enhancing health behaviors. The WAT market is experiencing rapid growth, valued at US $53.94 billion in 2023, and is projected to grow from US $62.03 billion in 2024 to US $290.85 billion by 2032 [10]. These devices, often worn on the wrist, are designed to positively influence daily lifestyle behaviors by providing users with a variety of health-related data, including metrics such as step count, calories burned, heart rate, sleep patterns, stairs climbed, and sedentary time. Assessment of physical activity using WATs tends to perform better in terms of validity and reliability than self-report assessment [11,12], pointing to their potential to improve activity surveillance and understanding of how activity can impact health. Beyond tracking data, WATs also incorporate behavior change techniques like goal setting, feedback on progress, action planning, review of behavior goals, the discrepancy between current behavior and goal, self-monitoring of behavior, biofeedback, commitment, social support, and rewards for achieving daily objectives, all designed to promote healthier behaviors [13].

The effectiveness of WATs in increasing physical activity levels and improving health across all age groups and in both healthy and clinical populations is well-documented. An umbrella review encompassing 719 primary studies demonstrated that WAT-based interventions significantly improved physical activity, daily step count, time spent in moderate to vigorous physical activity, body composition, blood pressure, and fitness among 163,992 participants [14]. There is also emerging evidence regarding WAT's role in sleep improvement, with a meta-analysis of 20 randomized controlled trials suggesting significant improvements in sleep quality parameters [15]. Furthermore, extensive research has been conducted on the reliability and validity of WATs [16,17].

However, there has been relatively less focus on users' perceptions and experiences of WATs. A 2016 study with 133 participants recruited through Amazon Mechanical Turk explored the well-being implications of these devices, revealing that WATs could impact users' daily life experiences beyond behavior change, providing psychological benefits and fostering a sense of relatedness [18]. Our group conducted a cross-sectional survey in 2016 with 237 Australian adults, examining user experiences with WATs, including usage patterns, social media data sharing, perceived behavior changes, and technical challenges [19]. We also investigated the emotional and psychological responses of users to their activity trackers, finding that most users felt empowered and motivated, while some experienced negative emotions such as anxiety or frustration [20]. Systematic literature reviews of studies conducted up until 2017 [21], 2019 [22], and 2020 [23] have examined various themes related to acceptability and user experience using quantitative and qualitative methods. They identified key issues in WAT user experience, including preferred WAT features [23], how user motivation impacts user experience [23], and how acceptance and usability research provide important design implications [21]. Since these studies, WATs have become more prevalent, and more diverse populations may now be using them. With the regular release of new models featuring advanced capabilities, user experience with WATs is likely to have evolved.

This study aims to understand current and former users' experiences of WATs in 2023 and compare these with our 2016 survey findings. Specifically, we will explore (1) patterns of use (eg, length of use, features used), (2) perceived motivation and behavior change, (3) practices and reasons underpinning data sharing through social media, and (4) technical issues and other barriers to use. By comparing these results with our previous survey [19], we hope to provide insights that will assist clinicians in supporting patients through observation of daily living, as well as aid researchers in exercise physiology, physical activity, and digital health considering WAT use in clinical or general populations.

Methods

Overview

Survey Instrument Development

An online survey instrument (Multimedia Appendix 1) was developed to address the research objectives and was delivered through Qualtrics. To allow for comparison, the survey items were based on the Maher et al [19] 2016 survey, which was developed in consultation with experts and showed test-retest bivariate correlation coefficients ranging between r=0.30 and 0.81 for survey sections [19]. A total of 7 items were added or edited to reflect recent advances in WAT features, such as meditation, workouts, phone call or text messaging, contactless payments, and calendar reminders. WATs were defined in the survey as any bodily-worn device that objectively measures physical activity and provides feedback [24]. The test-retest reliability of the 2023 survey was evaluated with a sample of 19 individuals (14 females, 4 males, and 1 nonbinary or third gender) recruited through Facebook over a 7 to 21-day interval,

XSL•FO
RenderX

yielding bivariate correlation coefficients between r=0.45 and 0.79 across sections.

Participants

Overview

Participants were adults aged 18 and over who were either current or former WAT users within the past 3 years and had used the device for at least 1 month. Exclusion criteria included the use of health-related smartphone applications or software without an associated WAT, WATs not measuring or providing physical activity feedback, and WATs not interacting with a smartphone or computer.

To recruit a diverse, international participant pool, voluntary response sampling was used using 2 platforms:

1. Amazon Mechanical Turk (MTurk): An Amazon-hosted crowdsourcing service that offers rapid access to a varied participant base. For this study, the Qualtrics survey link was posted on MTurk.
2. Facebook: The survey was disseminated through various community groups, the university research center's social media pages, and the research team's profiles. To encourage broader sharing, Facebook users who shared the post were entered into a draw to win one of 2 Aus $100 (US $62.83) retail gift vouchers.

Demographic Characteristics and WAT Usage

Participants' basic demographic information, including sex, age, country of residence, and education level, was collected through 4 items. They were asked about their current or former use of a WAT, leading to respective survey variations. Questions also covered the make and model of their WAT (open-ended), duration of usage, and, for current users, the intended continuation period.

Perceptions of Features

A total of 28 items asked current users about the frequency and perceived usefulness of common WAT features like active minutes, steps, stairs, sleep, heart rate, energy burned, energy consumed (food), meditation, workouts, calls or texts, contactless payment, calendar reminders, device connectivity, and data sharing. Responses were recorded on a 5-point Likert scale.

Behavior Change

Participants' perceptions of health behavior change due to WAT use were explored through 4 items. A total of 3 items on a 5-point Likert scale measured agreement on improvements in eating habits, physical activity, and sleep. In addition, 1 item presented 4 graphical physical activity patterns for participants to identify their own physical activity changes while using a WAT.

Motivation for Wearing a WAT

Participants chose their main motivation for using a WAT from 8 options, including fitness improvement, health monitoring, appearance, activity tracking, social sharing, competition, technology interest, or other (open-ended).

Data Usage and Sharing

A total of 2 items used a 5-point Likert scale to explore the usefulness of WAT data for short-term and long-term self-monitoring. Participants were asked what social networks they share their activity data in and could select from 6 options (tick all that apply), including "Facebook," "Twitter," "Instagram," "Within the activity tracker's software," "I don't share my activity data," or "Other (please specify)." Participants were then asked why they share their data and could choose from 6 options (tick all that apply), including "To share progress," "To compete with friends," "To get encouragement from others," "To motivate others," "I didn't share my data," or "Other (please specify)."

Practical Issues

A total of 3 questions explored practical issues related to using WATs. Both current and former users were asked if they had any complaints about their WAT and were asked to select all that applied from a multiple choice question, with responses including "Technical issues," "It falls off," "It doesn't fit," "It is uncomfortable," "Lost it," "Low battery life," "General wear and tear," "It often doesn't match my outfit," "Problems with the screen," "Problems uploading the data to supporting software," "Problems interpreting the data," "Problems with navigation of supporting website and technology," "Inaccurate at recording data," and "Other (please specify)." Participants were asked to rate the extent to which they agreed their experience using a WAT had been positive. Finally, former users were asked to specify their reasons for discontinuing WAT use from a list including device malfunction, insufficient use, technical difficulties, unmet goals, personal dislike, psychological impacts, and others (open-ended).

Data Quality Checking

After data collection, signs of internet bot activity were detected in the MTurk responses, notably identical and nonsensical answers to open-ended questions (eg, responding "GOOD" to a question asking what brand and model of WAT they use). To ensure data integrity, we implemented quality-check processes as recommended by Griffin et al [25]. This involved removing responses with a Qualtrics reCAPTCHA score below 0.5, indicating a higher risk of bot activity (on a scale of 0-1, where 1.0 indicates a low risk of bots, and 0.0 indicates a high risk of bots [26,27]). Furthermore, responses completed unrealistically quickly (less than 5 minutes) were excluded, considering the average completion time was 17 minutes. In addition, responses with less than 5% survey completion were removed. These quality-check processes filtered out all the participants initially suspected of being nongenuine.

Analysis

Data were downloaded into SPSS (version 26; IBM Corp) for analysis. Categorical variables and ordinal variables were analyzed using the frequency of responses and percentages, while continuous variables were analyzed using medians, means, and ranges. Comparison between 2023 and 2016 WAT users were examined using chi-square tests, with an alpha of .05 used to denote statistical significance.

XSL•FO
RenderX

Ethical Considerations

Ethical approval for this cross-sectional, descriptive online survey study was provided by the University of South Australia Human Research Ethics Committee (204908). Data collection took place from February to May 2023. As per ethical requirements, the cover page of the survey contained detailed information about the study, and completion of the online survey was regarded as informed consent. Survey data were de-identified.

Results

Demographic Characteristics and WAT Usage

A total of 827 survey responses were received. However, 338 were excluded with suspected bot activity (reCAPTCHA scores below 0.5 [n=208] and responses completed in less than 5 minutes [n=130]), and a further 14 responses were excluded for having less than 5% of the survey completed. Consequently, the analytic sample comprised 475 participants (280 recruited through MTurk and 195 recruited through Facebook), with 387 (82%) current users and 88 (18%) former users, which is similar to the 2016 survey proportions of 84% (200/237) current and 16% (37/237) former users among 237 participants.

Table 1 presents the participants' characteristics. The majority (219/475, 46.1%) were aged 25-34 years, and 60.8% (289/475) were male. This contrasts with the 2016 survey, where the majority were aged 18-24 years (75/237, 33.8%), and a greater proportion (168/237, 70.9%) were female. The 2023 survey was international, with 62% (292/475) of respondents from the United States, 26% (122/475) from Australia, and 12% (61/475) from other countries including Albania, Armenia, and India. In contrast, the 2016 survey was national, with all participants from Australia.

Table 1. Demographic characteristics of participants.

Demographic characteristics	2023 survey (n=475), n (%)	2016 survey (n=237), n (%)	*P* value	Chi-square (*df*)
Gender			<.001	64.6 (2)
Male	289 (60.8)	69 (29.1)		
Female	185 (38.9)	168 (70.9)		
Nonbinary or third gender	1 (0.2)	—[a]		
Age (years)			<.001	47.8 (5)
18-24	69 (14.5)	75 (33.8)		
25-34	219 (46.1)	56 (25.2)		
35-44	110 (23.2)	47 (21.2)		
45-54	57 (12)	28 (12.6)		
55-64	15 (3.2)	14 (6.3)		
≥65	5 (1)	2 (1)		
Country			<.001	349.3 (2)
United States of America	292 (61.5)	—		
Australia	122 (25.7)	237 (100)		
Other	61 (12.8)	—		
Education			<.01	13.2 (4)
Primary school	2 (0.4)	—		
High school	119 (25.1)	52 (21.9)		
Vocational qualification, trade school, community college, certificate, diploma or apprenticeship	49 (10.3)	42 (17.7)		
University bachelor's degree	232 (48.8)	95 (40.1)		
Postgraduate degree (eg, Master's degree)	73 (15.4)	48 (20.3)		

[a]Not available.

The 2023 survey results showed that Apple (210/475, 44%) was the most popular WAT brand, followed by Fitbit (101/475, 21%), Garmin (92/475, 19%), Samsung (47/475, 10%), and other brands (22/475, 5%). This was a significant change from 2016, where Fitbit dominated (160/237, 68%), followed by Garmin (39/237, 17%), Apple (8/237, 3%), Samsung (4/237, 2%), and other brands (26/237, 11%; *P*<.0001). The primary method of acquisition in both years was purchase by the users themselves (2023: 232/475, 49%; 2016: 134/237, 57%) or as gifts from family members (2023: 136/475, 29%; 2016: 103/237, 44%).

In terms of usage duration, current 2023 users reported using their WATs for a median of 18 months, while former users had

XSL•FO
RenderX

used them for a median of 9 months. In 2016, current users had a median usage duration of 7 months, and former users had used theirs for 5 months.

When current users from the 2023 survey were asked how long they planned to use their WAT in the future, the most common response was “indefinitely” (82/475, 21%). Similarly, current users of the 2016 survey were also asked how long they planned to use their WATs; the most common response was >36 months, which was the maximum duration response option in that survey. Among former 2023 users, when asked if they would be inclined to ever use a WAT again, 45% (39/88) somewhat agreed, and 40% (34/88) were neutral.

Motivation and Behavior Change

In 2023, the most commonly cited motivations for using WATs were improving health (239/475, 50%), monitoring activities (221/475, 47%), and enhancing fitness (180/475, 38%; Table 2). The same 3 motivations were common in the 2016 survey, though in a different order, with 2016 participants prioritizing monitoring activity patterns (85/237, 36%), improving fitness (65/237, 27%), and improving health (43/237, 18%). Nearly one-third of 2023 respondents (147/475, 31%) reported improving their appearance as a motivation for using a WAT, compared with <1% in 2016. During analysis of the motivation data, we noticed that the 2023 survey allowed multiple responses for motivation, differing from the 2016 survey’s single-choice format. Given this discrepancy, direct comparisons using chi-square tests were not undertaken.

Table 2. Participant motivations for wearing a wearable activity tracker.

Primary motivation	2023 survey (N=475), n (%)	2016 survey (N=237), n (%)
To improve health	239 (50.3)	43 (18.1)
To monitor activities	221 (46.5)	85 (35.9)
To improve fitness	180 (37.9)	65 (27.4)
To improve appearance	147 (30.9)	2 (0.8)
To share my activity	126 (26.5)	—[a]
To compete with family or friends	60 (12.6)	7 (2.9)
To keep up with new technology	45 (9.5)	4 (1.7)
Other	15 (3.9)	15 (6.3)

[a]Not available.

Table 3 presents the comparison of participants’ ratings (percentage of participants responding somewhat agree or agree strongly) for the usefulness of various WAT metrics in 2023 versus 2016. Notably, the perceived usefulness of the step count feature significantly decreased from 94.5% (189/200) in 2016 to 65.9% (255/387) in 2023. Similar declines were seen in the perceived usefulness of activity minutes (151/200, 75.5% in 2016 to 252/387, 65.2% in 2023), and stairs climbed (115/200, 57.5% in 2016 to 189/387, 48.9% in 2023). Conversely, the usefulness rating for energy intake rose from 36% (72/200) in 2016 to 49.6% (192/387) in 2023. The heart rate feature also showed a nonsignificant trend for increased perceived usefulness, from 63% (126/200) in 2016 to 70.5% (273/387) in 2023. In fact, it was the top-rated useful feature in 2023. Ratings for sleep features fell slightly (232/387, 59.9% in 2023 vs 132/200, 66% in 2016) while energy burned remained fairly stable.

Table 3. Participant rating of usefulness of features (percentage reporting somewhat agree or agree strongly).

Feature	2023 survey (n=387), n (%)	2016 survey (n=200), n (%)	*P* value	Chi-square (*df*)
Heart rate	273 (70.5)	126 (63)	.06	3.4 (1)
Steps	255 (65.9)	189 (94.5)	<.001	58.5 (1)
Active minutes	252 (65.2)	151 (75.5)	.01	6.6 (1)
Sleep	232 (59.9)	132 (66)	<.001	16.6 (1)
Energy burned	227 (58.6)	113 (56.5)	.61	0.09 (1)
Energy consumed	192 (49.6)	72 (36)	<.001	9.8 (1)
Stairs	189 (48.9)	115 (57.5)	.04	3.9 (1)

Current WAT users in 2023 generally agreed on the usefulness of updated WAT features, with high ratings for workouts (290/387, 75% somewhat or strongly agree), device connectivity (263/387, 68%), receiving calls or text messages (249/387, 64%), contactless payment (200/387, 52%), calendar reminders (214/387, 55%), data sharing (193/387, 50%), and meditation (193/387, 50%).

Respondents were asked their opinions on the usefulness of real-time monitoring and long-term monitoring features. While the majority of 2023 participants somewhat or strongly agreed

XSL•FO
RenderX

that both real-time monitoring and long-term monitoring were useful (57% and 58% agreement, respectively), the agreement ratings have fallen since 2016, when the ratings for real-time and long-term monitoring were 84% and 72% agreement, respectively.

Participants were asked whether they had changed their lifestyle behaviors as a result of using a WAT (Table 4). In both the 2023 and 2016 surveys, physical activity was the behavior respondents were most likely to report having changed due to their WAT. However, there was a slight reduction over time in the perceived impact of WATs on enhancing physical activity, from 77% (182/237) agreement in 2016 to 63% (300/475) agreement in 2023. In contrast, the perceived positive impact on diet increased from 36% (85/237) in 2016 to 51% (244/475) in 2023. Similarly, the percentage of participants who felt their sleep improved due to using a WAT rose from 22% (52/237) in 2016 to 42% (200/475) in 2023.

Table 4. Perceived change in lifestyle behaviors since using a wearable activity tracker (percentage reporting somewhat agree or agree strongly).

Activity	2023 survey (n=475), n (%)	2016 survey (n=237), n (%)	*P* value	Chi-square (*df*)
Physical activity	300 (63)	182 (77)	<.001	13.4 (1)
Diet	244 (51)	85 (36)	<.001	15.2 (1)
Sleep	200 (42)	52 (22)	<.001	28.1 (1)

Social Features and Data Sharing

There has been a marked increase in the use of social media platforms for sharing WAT data (Table 5). In 2023, Instagram was used most by 44% (171/387) of participants, a sharp rise from just 1% (2/200) in 2016. Facebook and Twitter saw similar surges, with 36% (141/387) using Facebook (up from 7/200, 4% in 2016) and 35% (134/387) using Twitter (up from 1/200, 1% in 2016). All these increases were statistically significant. Conversely, the proportion of participants not sharing their activity data dropped significantly from 65% (130/200) in 2016 to 27% (104/387) in 2023.

The motivations behind sharing data have also evolved. In 2023, common reasons included sharing progress (145/387, 38%, up from 18/200, 9% in 2016), receiving encouragement (137/387, 35%, up from 13/200, 7% in 2016), and motivating others (138/387, 36%, up from 17/200, 8.5% in 2016). Competing with friends was noted by 35% (134/387) in 2023, a significant increase from 17% (33/200) in 2016.

Table 5. Sharing platforms and reasons for sharing data[a].

	2023 survey (n=387), n (%)	2016 survey (n=200), n (%)	*P* value	Chi-square (*df*)
Sharing platform				
Instagram	171 (44.1)	2 (1)	<.001	118.3 (1)
Facebook	141 (36.4)	7 (3.5)	<.001	75.8 (1)
Twitter	134 (34.6)	1 (0.5)	<.001	125.8 (1)
I don’t share my activity data	104 (26.8)	130 (65)	<.001	79.9 (1)
Social network within the wearable’s software	113 (29.2)	70 (35)	.15	2 (1)
Strava	5 (1.3)	9 (4.5)	.01	5.8 (1)
Other	3 (0.8)	3 (1.5)	.40	0.6 (1)
Reason for sharing data				
To share progress	145 (37.5)	18 (9)	<.001	53.2 (1)
To get encouragement from others	137 (35.4)	13 (6.5)	<.001	57.8 (1)
To motivate others	138 (35.6)	17 (8.5)	<.001	50 (1)
To compete with friends	134 (34.6)	33 (16.5)	<.001	21.2 (1)
I don’t share data	104 (26.9)	154 (77)	<.001	78.6 (1)
Other (please specify)	7 (1.8)	4 (2)	.87	0.2 (1)

[a]Participants were allowed to select multiple responses, and percentages reflect the number of participants who selected each response option as a portion of all participants in that survey sample.

Technical Issues and Barriers

The majority of 2023 respondents agreed that they had an overall positive experience using their WAT (322/475, 68% somewhat agree or agree strongly), though this was a decrease compared with 2016 when 212 of 237 (89%) reported an overall positive experience (χ^2_1= 36.2; *P*<.001). The most common complaints in 2023 were short battery life (reported by 201/475, 21% of participants), technical issues (cited by 91/475, 19%), and general wear and tear (mentioned by 86/475, 18%; Table 6). Generally, 2023 participants appeared to report slightly more

XSL•FO
RenderX

complaints than 2016 respondents. In contrast to this trend, problems with data uploading and inaccurate data recording have decreased slightly over time, from 19% (45/237) in 2016 to 10% (47/475) in 2023, and from 17% (39/237) in 2016 to 7% (32/475) in 2023, respectively.

Table 6. Complaints about wearable activity tracker (WAT)[a].

Complaint	2023 survey (n=475), n (%)	2016 survey (n=237), n (%)	*P* value	Chi-square (*df*)
None	162 (34.1)	71 (30)	.26	1.2 (1)
Short battery life	102 (21.4)	48 (20)	.7	0.1 (1)
Technical issues	91 (19.1)	29 (12.2)	.02	5.4 (1)
General wear and tear	86 (18.1)	32 (13.5)	.11	2.4 (1)
It was or is uncomfortable	77 (16.2)	20 (8.4)	<.01	8.1 (1)
It fell or falls off	75 (15.7)	17 (7.2)	<.001	10.4 (1)
It didn't or doesn't fit	71 (14.9)	—[b]	<.001	39.3 (1)
It often didn't or doesn't match my outfit	64 (13.5)	44 (18.6)	.07	3.1 (1)
Problems with the screen	58 (12.2)	10 (4.2)	<.001	11.6 (1)
Lost it	47 (9.9)	5 (2.1)	<.001	14.1 (1)
Problems uploading the data to supporting software	47 (9.9)	45 (19)	<.001	11.6 (1)
Problems interpreting the data	41 (8.6)	10 (4.2)	.03	4.6 (1)
Problems with navigation of supporting website and technology	35 (7.4)	16 (6.8)	.76	0.9 (1)
Inaccurate at recording data	32 (6.7)	39 (16.5)	<.001	16.6 (1)
Other (please specify)	21 (4.4)	27 (11.4)	<.001	12.2 (1)

[a]Participants were allowed to select multiple responses, and percentages reflect the number of participants who selected each response option as a portion of all participants in that survey sample.

[b]Not available.

Former users were asked to identify why they no longer used their WAT (Table 7). Among former WAT users in the 2023 survey, the main reasons for ceasing use were that their WAT was broken (29/88, 33%), they did not like it (22/88, 25%), or that it was lost (15/88, 17%), difficult to understand (15/88, 17%), or technical difficulties (15/88, 17%). This was different from the 2016 survey, where the main reasons former users ceased using their WATs were that they felt they had learned everything they could from their WAT (11/37, 30%), their WAT was broken (8/37, 22%), or their WAT was not helping them achieve their goals (5/37, 14%). The findings appear to indicate a shift in why users discontinue WAT usage, with increasing reports of disliking the device (up from 2/37, 5% in 2016 to 22/88, 25% in 2023) and understanding difficulties (up to 2/37, 5% in 2016 to 15/88, 17% in 2023), and fewer users stopping because they felt they had learned all they could from the device (down from 11/37, 30% in 2016 to 5/88, 6% in 2023).

XSL•FO
RenderX

Table 7. Former user's reasons for ceasing use of wearable activity tracker (WAT)[a].

Reasons to stop wearing tracker	2023 survey (n=88), n (%)	2016 survey (n=37), n (%)	*P* value	Chi-square (*df*)
It broke	29 (33.3)	8 (21.6)	.20	1.6 (1)
I didn't like it	22 (25.3)	2 (5.4)	.01	6.4 (1)
Got lost	15 (17.2)	4 (10.8)	.37	0.7 (1)
It was difficult to understand	15 (17.2)	2 (5.4)	.08	3 (1)
Technical difficulties	15 (17.2)	4 (10.8)	.37	0.7 (1)
I wasn't using it enough	14 (16.1)	—[b]	.01	6.6 (1)
It wasn't helping with my goals	12 (13.8)	5 (13.5)	.98	0.0 (1)
I found it intrusive	10 (11.5)	3 (8.1)	.58	0.29 (1)
I was experiencing negative psychological impacts	8 (9.2)	3 (8.1)	.85	0.03 (1)
I learned everything I could	5 (5.7)	11 (29.7)	<.001	13.4 (1)
Other	6 (6.9)	16 (43.2)	<.001	23.8 (1)

[a]Participants were allowed to select multiple responses, and percentages reflect the number of participants who selected each response option as a portion of all participants in that subgroup.

[b]Not available.

Discussion

Principal Findings

This study set out to explore changes in user experiences, preferences, and perceived impacts of WATs from 2016 to 2023. We found significant shifts in brand preferences, with Apple overtaking Fitbit as the most popular WAT brand by 2023. The study found that, in general, WATs are now being used for a longer period compared with the previous survey. The sharing of WAT data on social media platforms has surged, indicating a greater integration of these devices into social connectivity and community engagement. While there was a decrease in the perceived effectiveness of WATs in enhancing physical activity, users reported increased positive impacts on diet and sleep management. However, this was tempered by increased complaints about technical issues and discomfort, as well as a decline in overall positive user experience with WATs. In addition, we observed a shift in reasons for discontinuing WAT use, with more users in 2023 ceasing usage due to dissatisfaction with the device or its complexity. Possible reasons for the key differences between responses in 2016 and responses in 2023 (P<.001) are summarized in Multimedia Appendix 2 and further discussed below.

The majority of current users in the 2023 survey reported that they had worn their WAT for a median duration of 18 months and were planning on using their WATs indefinitely. This was significantly greater than that of users in 2016, who reported that they had worn their WATs for 7 months and were planning on using them for >3 years. This is likely due to WATs having existed for a longer period, with more metrics and features available, potentially increasing appeal and, therefore, long-term usage. Most former users in 2023 wore their WATs for a median duration of 9 months, with just over one-third reporting that their device breaking was their main reason for ceasing use. These findings confirm evidence that suggests WATs are often abandoned due to technical malfunctions, including broken WATs or battery issues [28]. In contrast, former users of the 2016 survey reported that they had worn their WAT for a median duration of 5 months, with the main reason for ceasing use being that they had learned everything they could. It is possible that enhancements in WATs over time, such as improved health monitoring capabilities and integration with smartphones, have increased the perceived value of the devices and made them more valuable in people's lives [18]. These findings suggest that continuous innovation in WAT features may play a crucial role in sustaining user interest and reducing disinterest or apathy over time.

Despite the increase in duration of use, fewer users in 2023 reported having a positive experience with their WAT than in 2016, and in general, reported a greater number of complaints. Short battery life was the most reported complaint in both 2016 and 2023, which remains consistent with research that suggests battery capacity is one of the main user complaints and reasons for abandonment [29]. In addition, complaints such as technical issues, wear and tear, comfort, the WAT falling off, problems with the screen, and the ability to interpret the data were more commonly reported in 2023. This increase in complaints seems counterintuitive, given that WATs are now more advanced with more features. Rising expectations may be behind this–as WATs have become more advanced, user expectations might have risen correspondingly, leading users to be more critical of their devices. An alternative explanation is that the inclusion of more features can make WATs more complex to use. This complexity might lead to frustrations, especially among users who prefer simpler, more intuitive interfaces. More advanced devices that offer enhanced features may also explain a shift in user preferences for Apple WATs, with features like call and text, smart pay, and safety functions easily integrated into users' Apple technology ecosystems.

Whilst WATs track various health behaviors like physical activity, diet, and sleep, physical activity remained the most monitored behavior in both the 2016 and 2023 surveys. It was

XSL•FO
RenderX

also the behavior most frequently reported as improved due to WAT use, consistent with research highlighting WAT's effectiveness in increasing physical activity across different populations [14]. In 2023, there was an increase in perceived change in diet and sleep since using a WAT, suggesting that WATs have the potential to influence a range of lifestyle behaviors, and users may benefit more from the advanced features included in more recent models.

Interestingly, although the perceived change in diet and sleep since using a WAT increased between 2016 and 2023, and the perceived usefulness of energy consumed also increased, sleep was considered less useful in 2023. Heart rate monitoring was considered the most useful feature in 2023, a shift from 2016 when step count was rated highest. These findings may reflect the increased popularity of fitness activities emphasizing heart rate, such as high-intensity interval training, or other activities that do not record steps, such as strength training. Previous research suggests that resting and exercise-related heart rate measures are considered useful as they provide noninvasive and time-efficient insights into aerobic fitness, recovery, and performance [30]. This trend in the evolving usage of WATs, with a focus on heart rate monitoring over steps, suggests a broader shift in user health priorities and fitness trends, mirroring the advancements in wearable technology and its increased capability to provide detailed, fitness-specific metrics that cater to modern exercise regimens and health awareness.

Participants in the 2023 survey were notably more inclined to share data on mainstream social networks like Instagram, Facebook, and Twitter, a change from the 2016 survey. This shift contrasts with earlier research suggesting a preference for sharing on niche physical activity social networks [31,32]. This increase in sharing on broader platforms may be attributed, at least in part, to a continued increase in social media usage from 2.73 billion users in 2017 to 4.59 billion users in 2023 [33]. In particular, Instagram has a strong fitness subculture [34]. The rise of fitness influencers and communities on these platforms may also encourage users to share their data more openly. In addition, there seems to be an evolving perception of sharing fitness data, now viewed as a motivational and supportive tool, possibly amplified by the COVID-19 pandemic's push for social connectedness during periods of isolation [35].

Strengths and Limitations

This study has several strengths. First, the longitudinal, cross-sectional design allowed us to compare WAT user experience from 2016 to 2023, offering a valuable perspective on the evolution of user preferences and practices over time. The 2023 study attracted a large and diverse sample, enhancing the generalizability of our findings. We considered the perspectives of both current and former WAT users and examined various facets of WAT usage, including brand preference, feature utility, motivations, social sharing behaviors, and technical issues. In addition, our study was rigorous in its methodology, using an existing survey tool, conducting test-retest analysis, and implementing extensive data quality checks.

Limitations must also be acknowledged. As a survey, data are self-reported, which introduces the possibility of biases as participants may over or underestimate their usage or satisfaction with their WATs due to factors such as factors such as misremembering past experience, social desirability, or acquiescence. Nonresponse bias could also be present. Response options for the item "How long did you use your wearable?" differed between the 2016 and 2023 surveys, which may have introduced some bias. Although the sample was large and diverse, the findings may not be entirely generalizable to all WAT users, especially those in different geographical locations or demographic groups not adequately represented in the sample. There was a very low representation of older adults in both the 2016 and 2023 surveys. This differs from other studies, where older adults are among the most studied groups for acceptability studies. This discrepancy may be explained by our recruitment methods, which utilized social media and MTurk to recruit existing users of WATs. These platforms potentially reach younger and middle-aged adults more readily than older adults. In contrast, many evaluations of mHealth acceptability in older adults are linked to interventional studies where devices are provided directly to participants by researchers. This approach might better engage older adults, who may be less likely to participate in online surveys but more receptive to direct intervention studies. In addition, there were some demographic differences between the survey years that could have impacted the findings. For example, the 2023 survey involved an international sample, whereas the 2016 survey involved only Australians. The 2023 survey also had proportionally more males than the 2016 survey. The use of multivariable analysis methods would have enabled examination of the effects of these cohort differences but was not possible due to the unavailability of the individual-level data from the 2016 survey. We conducted our analyses using the summary data available, which still offers valuable insights into the trends and changes over time. Another limitation of the study includes the presence of numerous internet bots from MTurk that had completed the 2023 survey. Although we conducted data quality checking, concerns have been raised about MTurk, including self-misrepresentation, self-selection bias, English language fluency, and inattention, which may impact a study's validity [36].

Implications for Research and Practice

A key overall finding in this study is that the use of WATs is shifting—including a shift in brand preference toward Apple, a shift in the metrics valued by users away from simple physical activity monitoring features, and toward more detailed metrics such as heart rate tracking, diet, and sleep tracking. People are also sharing their data differently, with increased sharing of WAT data on social media platforms like Instagram. These findings could inform behavioral health interventions. Health promotion researchers might consider integrating WATs into health promotion and disease prevention strategies, given their potential to positively influence lifestyle behaviors. There is also the potential for including sedentary behavior as a metric for users to track, given its importance for health and inclusion in physical activity guidelines. In addition, if patients are willing to discuss their WAT information with health professionals, there is the potential for integrating WATs into health care, providing insights into how different activities and behaviors affect health and well-being and facilitating shared goal setting

XSL•FO
RenderX

and monitoring. However, researchers and health professionals seeking to use WATs with research participants and patients should be cognizant of the common usability complaints, such as battery life and ease of use, since these aspects may be crucial in the long-term adherence to WAT-based programs.

Future research should look to examine WAT user preferences and practices with diverse populations, including different age groups, geographical locations, and health statuses, to provide a more comprehensive understanding of WAT usage patterns and preferences across a broader demographic. In addition, given the recent increase in sharing WAT data on social media between 2016 and 2023, studies investigating the psychological and behavioral impacts of sharing WAT data on social media platforms would be valuable. Research could explore how social sharing influences motivation, behavior change, and community building in the context of health and fitness. Such findings could influence the development of interventions incorporating the use of WATs for healthy lifestyle behaviors. In addition, given the increase in complaints about technical issues and device comfort, research focusing on user interface design and ergonomic improvements in WATs could be beneficial. This would need to be done in partnership with WAT manufacturers to inform future product development.

Implications for WAT Design

Our study highlights several key insights that are highly relevant for companies involved in the design and development of WATs. The most common complaint from users at both time points was short battery life. Manufacturers should prioritize advancements in battery technology to extend the life of WATs between charges to enhance user satisfaction and device adherence. In addition, more users in 2023 reported discomfort and issues with the fit of their devices than in 2016. This could be due to increased battery size or change in device shape due to, for example, the addition of heart rate monitors in more models. Companies should focus on ergonomic designs that offer better comfort for long-term wear, such as adjustable bands, lightweight materials, and options for different wrist sizes.

Technical reliability also emerged as a crucial area for improvement. Although less common in 2023 than in 2016, users frequently mentioned issues with data accuracy and connectivity problems. Improving the reliability and accuracy of WAT sensors and ensuring seamless connectivity with other devices are essential for maintaining user trust and encouraging continued usage. Furthermore, the shift in user preference toward advanced health metrics like heart rate monitoring over basic step counts indicates that users are seeking more sophisticated health insights. Companies should continue to develop and highlight features that provide deeper health analytics, such as heart rate variability, stress levels, and sleep quality.

To address these needs effectively, manufacturers should invest in user education and support. Providing comprehensive tutorials, user guides, and customer support can help mitigate issues related to device complexity and usability, enabling users to understand and make the most of their WAT features. By addressing key areas such as battery life, comfort, technical reliability, and advanced health metrics, companies can enhance the user experience and ensure long-term device adoption, positioning WATs as valuable tools for health and fitness.

Conclusions

In conclusion, this study has provided insights into the evolving landscape of WAT usage from 2016 to 2023. The shift in brand preference toward Apple and the longer usage duration indicate a maturing market and evolving user needs. Notably, the surge in sharing WAT data on platforms like Instagram highlights the increasing social aspect of health monitoring. However, despite technological advancements, the rise in user complaints and the shift in reasons for discontinuation emphasize the need for improvements in WAT design and functionality. These trends underscore the growing importance of integrating sophisticated health metrics such as heart rate monitoring, which aligns with the rising popularity of specific fitness activities and a broader shift in health and wellness culture. As WATs become more embedded in users' daily lives, their role in health promotion and behavioral change becomes increasingly significant. Future research should continue to explore these dynamics, particularly focusing on diverse populations and the psychological impacts of social sharing, to further understand and optimize the potential of WATs in supporting users' health and well-being.

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors made substantial contributions to this study and approved the final manuscript. CM conceptualized the study and oversaw the project. DB, CM, RC, and KS contributed to the study design. DB led the recruitment. DB, CM, RC, and KS oversaw data collection and interpretation. DB performed data analysis. DB, CM, RC, and KS contributed to manuscript preparation and interpretation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instrument.

XSL•FO
RenderX

[DOCX File , 735 KB - jmir_v27i1e56251_app1.docx]

Multimedia Appendix 2
Summary of key changes between 2016 and 2023 and related insights.
[DOCX File , 26 KB - jmir_v27i1e56251_app2.docx]

References

1. Global status report on physical activity 2022. World Health Organization. URL: https://www.who.int/teams/health-promotion/physical-activity/global-status-report-on-physical-activity-2022 [accessed 2022-10-19]
2. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet 2019;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
3. Li J, Cao D, Huang Y, Chen Z, Wang R, Dong Q, et al. Sleep duration and health outcomes: an umbrella review. Sleep Breath 2022;26(3):1479-1501. [doi: 10.1007/s11325-021-02458-1] [Medline: 34435311]
4. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
5. Katzmarzyk PT, Friedenreich C, Shiroma EJ, Lee IM. Physical inactivity and non-communicable disease burden in low-income, middle-income and high-income countries. Br J Sports Med 2022;56(2):101-106 [FREE Full text] [doi: 10.1136/bjsports-2020-103640] [Medline: 33782046]
6. Park JH, Moon JH, Kim HJ, Kong MH, Oh YH. Sedentary lifestyle: overview of updated evidence of potential health risks. Korean J Fam Med 2020;41(6):365-373 [FREE Full text] [doi: 10.4082/kjfm.20.0165] [Medline: 33242381]
7. Chaput JP, Dutil C, Featherstone R, Ross R, Giangregorio L, Saunders TJ, et al. Sleep timing, sleep consistency, and health in adults: a systematic review. Appl Physiol Nutr Metab 2020;45(10 (Suppl. 2)):S232-S247 [FREE Full text] [doi: 10.1139/apnm-2020-0032] [Medline: 33054339]
8. Morze J, Danielewicz A, Hoffmann G, Schwingshackl L. Diet quality as assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension score, and health outcomes: a second update of a systematic review and meta-analysis of cohort studies. J Acad Nutr Diet 2020;120(12):1998-2031.e15 [FREE Full text] [doi: 10.1016/j.jand.2020.08.076] [Medline: 33067162]
9. Backonja U, Kim K, Casper GR, Patton T, Ramly E, Brennan PF. Observations of daily living: putting the 'personal' in personal health records. NI 2012 (2012) 2012;2012:6 [FREE Full text] [Medline: 24199037]
10. Fitness tracker market: 2024-2032. Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/toc/fitness-tracker-market-103358 [accessed 2025-06-24]
11. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020;17(1):31 [FREE Full text] [doi: 10.1186/s12966-020-00938-3] [Medline: 32131845]
12. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
13. Düking P, Tafler M, Wallmann-Sperlich B, Sperlich B, Kleih S. Behavior change techniques in wrist-worn wearables to promote physical activity: content analysis. JMIR mHealth uHealth 2020;8(11):e20820 [FREE Full text] [doi: 10.2196/20820] [Medline: 33211023]
14. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
15. Lai MYC, Mong MSA, Cheng LJ, Lau Y. The effect of wearable-delivered sleep interventions on sleep outcomes among adults: a systematic review and meta-analysis of randomized controlled trials. Nurs Health Sci 2023;25(1):44-62. [doi: 10.1111/nhs.13011] [Medline: 36572659]
16. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR mHealth uHealth 2020;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
17. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR mHealth uHealth 2018;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
18. Karapanos E, Gouveia R, Hassenzahl M, Forlizzi J. Wellbeing in the making: peoples' experiences with wearable activity trackers. Psychol Well Being 2016;6:4 [FREE Full text] [doi: 10.1186/s13612-016-0042-6] [Medline: 27376017]
19. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]

XSL•FO
RenderX

20. Ryan J, Edney S, Maher C. Anxious or empowered? A cross-sectional study exploring how wearable activity trackers make their owners feel. BMC Psychol 2019;7(1):42 [FREE Full text] [doi: 10.1186/s40359-019-0315-y] [Medline: 31269972]
21. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
22. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
23. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
24. Lewis ZH, Pritting A, Picazo AL, JeanMarie-Tucker M. The utility of wearable fitness trackers and implications for increased engagement: an exploratory, mixed methods observational study. Digit Health 2020;6:2055207619900059 [FREE Full text] [doi: 10.1177/2055207619900059] [Medline: 31976079]
25. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022;56(4):2841-2852 [FREE Full text] [doi: 10.1007/s11135-021-01252-1] [Medline: 34629553]
26. Google Developers. reCAPTCHA v3: interpreting the score. URL: https://developers.google.com/recaptcha/docs/v3#interpreting_the_score [accessed 2025-08-01]
27. Fraud detection. Qualtrics. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/ [accessed 2025-08-01]
28. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR mHealth uHealth 2017;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
29. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR mHealth uHealth 2019;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
30. Schneider C, Hanakam F, Wiewelhove T, Döweling A, Kellmann M, Meyer T, et al. Heart rate monitoring in team sports-A conceptual framework for contextualizing heart rate measures for training and recovery prescription. Front Physiol 2018;9:639 [FREE Full text] [doi: 10.3389/fphys.2018.00639] [Medline: 29904351]
31. Chang RCS, Lu HP, Yang P, Luarn P. Reciprocal reinforcement between wearable activity trackers and social network services in influencing physical activity behaviors. JMIR mHealth uHealth 2016;4(3):e84 [FREE Full text] [doi: 10.2196/mhealth.5637] [Medline: 27380798]
32. Shih P, Han K, Poole E, Rosson M, Carrol J. Use and adoption challenges of wearable activity trackers. 2015 Presented at: iConference 2015; 24-27 March 2015; Newport Beach, California URL: https://www.researchgate.net/publication/268746784_Use_and_adoption_challenges_of_wearable_activity_trackers
33. Number of social media users worldwide from 2017 to 2027. Statista. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2025-08-10]
34. Curtis RG, Prichard I, Gosse G, Stankevicius A, Maher CA. Hashtag fitspiration: credibility screening and content analysis of Instagram fitness accounts. BMC Public Health 2023;23(1):421 [FREE Full text] [doi: 10.1186/s12889-023-15232-7] [Medline: 36864397]
35. Zuo Y, Ma Y, Zhang M, Wu X, Ren Z. The impact of sharing physical activity experience on social network sites on residents' social connectedness:a cross-sectional survey during COVID-19 social quarantine. Global Health 2021;17(1):10 [FREE Full text] [doi: 10.1186/s12992-021-00661-z] [Medline: 33430894]
36. Aguinis H, Villamor I, Ramani RS. MTurk Research: review and recommendations. J Manage 2020;47(4):823-837. [doi: 10.1177/0149206320969787]

Abbreviations

MTurk: Mechanical Turk
WAT: wearable activity tracker

XSL•FO
RenderX

Edited by J Khuntia; submitted 10.01.24; peer-reviewed by C Budhathoki, P Crisford; comments to author 29.05.24; revised version received 08.07.24; accepted 25.11.24; published 13.02.25.

Please cite as:
Beckett D, Curtis R, Szeto K, Maher C
Changing User Experience of Wearable Activity Monitors Over 7 Years: Repeat Cross-Sectional Survey Study
J Med Internet Res 2025;27:e56251
URL: https://www.jmir.org/2025/1/e56251
doi:10.2196/56251
PMID:

©Darcy Beckett, Rachel Curtis, Kimberley Szeto, Carol Maher. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Experiences and Expectations of Immigrant and Nonimmigrant Older Adults Regarding eHealth Services: Qualitative Interview Study

Maedeh Ghorbanian Zolbin[1], MBA; Sari Kujala[1], PhD; Isto Huvila[2], PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of ALM, Uppsala University, Uppsala, Sweden

Corresponding Author:
Maedeh Ghorbanian Zolbin, MBA
Department of Computer Science
Aalto University
PO Box 15400
Espoo, 02150
Finland
Phone: 358 0449270768
Email: maedeh.ghorbanianzolbin@aalto.fi

Abstract

Background: The emergence of eHealth services could contribute to improving individuals' quality of life by optimizing effective and efficient care. However, various challenges might limit some older adults' use of eHealth services.

Objective: This study aimed to understand the perspectives of older adults (aged ≥65 years) of different backgrounds regarding eHealth services. We explored the experiences of Iranian immigrant and nonimmigrant older adults with eHealth services to identify their perceived challenges, emotions, and wishes. Immigrants face more challenges, and there is a need to understand their perspectives in addition to those of nonimmigrants. Iranians are one important immigrant group, as their number is limited and their specific needs are less well understood compared to those of the bigger immigrant groups.

Methods: This study used a qualitative explorative research design. Semistructured interviews were conducted between February 2023 and May 2023. The participants were 25 older adults: nonimmigrants residing in cities (n=8, 32%), nonimmigrants residing in rural areas (n=9, 36%), and Iranian immigrants residing in cities (n=8, 32%). Data were analyzed through inductive and deductive content analysis and interpreted through self-determination theory.

Results: Interacting with eHealth services was challenging for some older adults. They perceived several difficulties, with the most obvious ones being related to values and preferences, as some older adults did not value eHealth services (16/25, 64%), had insufficient digital skills (15/25, 60%), and experienced usability issues (15/25, 60%). The first two challenges were more pronounced among immigrants. In contrast, nonimmigrants from cities, being more familiar with the services, shared more usability issues. These identified challenges prevented older adults from satisfying their basic psychological needs of being competent and autonomous users and having a sense of belonging (aspects of self-determination theory), which were the main source of negative emotions. A common negative feeling was confusion (16/25, 64%) among those with limited experience using smart devices and those with poor self-reported digital skills. Conversely, older adults' interaction with eHealth services generated positive emotions that were connected to the satisfaction of their basic psychological needs. Being interested in using eHealth services was a common feeling among most participants regardless of their background and was connected to satisfying their need for being competent and autonomous. The positive emotions could be supported by applying older adults' needs to the design of eHealth services (10/25, 40%) and by supporting their digital skills (19/25, 76%).

Conclusions: Some older adults value eHealth services and see their added benefits. However, various challenges limit their use of these services. The analysis of older adults' needs yielded several practical ideas that could improve the user-friendliness of the services and highlighted the importance of sufficient support services tailored to the cultural needs of specific groups of older adults.

(J Med Internet Res 2025;27:e64249) doi:10.2196/64249

XSL•FO
RenderX

KEYWORDS

eHealth services; older adults; immigrant; usability; user experience; emotion; self-determination theory

Introduction

Background

The growing number of older adults is a significant socioeconomic trend [1]. Over the next 3 decades, the number of older adults—individuals aged ≥65 years—is projected to double [2]. The proportion of older adults in Europe will reach 35% in 2050 [3]. As people age, they may become more susceptible to different diseases, leading to high demand for health care services [4]. However, several barriers limit some older adults from properly managing their health conditions using traditional health care services. These include reduced mobility [5] and residing in rural locations with limited access to physical health care facilities [6]. Furthermore, the number of physical health care facilities is limited and does not match growing needs [7].

To overcome such barriers, digital solutions have been mobilized to deliver eHealth services [8,9], which are defined as "the use of information and communication technologies for health" [10]. eHealth services are transformative solutions in the context of demographic aging, urban normativity, and austerity politics as they aim to enhance health care quality and equality, especially for people in sparsely populated areas [11].

The use of eHealth services can be expected to facilitate communication with health care providers [8] and allow older adults to control their health status at their own pace as they can read their medical records on electronic patient portals [12]. Furthermore, eHealth services could be useful in supporting and managing a healthy diet and lifestyle among older adults [13]. Moreover, eHealth has proven useful in promoting medication adherence as well as effective in preventing disease among older adults [14]. For example, using mobile health can potentially improve health outcomes among older adults with chronic diseases by supporting symptom control [15]. Mobile health refers to the use of mobile phones and other wireless technology for medical purposes, which is a recent technological evolution in eHealth [16].

Despite these benefits, several difficulties hinder some older adults from fully using eHealth services. For example, older adults with limited financial resources might have limited access to smart devices or the internet [17]. Furthermore, some older adults might lack the proper digital skills to use smart devices or e-services [18] and might not be confident in using technology for critical issues such as health [17]. Security issues and privacy concerns are recognized as factors that limit older adults' willingness to use eHealth services [19,20].

Moreover, these services are often designed without considering older adults' specific features, such as physical difficulties or low digital skills [21]. Thus, older adults might face several usability issues, such as navigation difficulties [22]. Usability is defined by the International Organization for Standardization 9241-11 standard as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [23]. In addition, immigrant older adults might face more barriers that are related to their backgrounds, such as lower socioeconomic status, low language proficiency, and comparatively lower levels of social inclusion [24]. The perceived difficulties might lead to negative emotions [25,26], lack of interest, and unequal use of eHealth services among older adults, which might increase the risk of inequities and social exclusion [27].

Finland is a very high-income country with advanced health policy programs that ensure citizens' well-being [28]. To support equal access to public health services, Finnish law supports the provision of e-services to promote the availability, quality, security, and equality of eHealth services for everyone [29]. However, inequalities still exist, especially among vulnerable groups such as older adults and immigrants [17].

To achieve a socially sustainable society in which everyone is treated equally and individuals' health and functional capacity are supported, it is necessary for people to have a high capability to take responsible action for their health. This can be supported by, for example, tackling the challenges that older adults face in the use of eHealth services and involving older adults from different backgrounds and their specific needs in the design process and delivery strategies of these services [30-32]. This study aimed to understand the perspectives of older adults (aged ≥65 years) from different backgrounds regarding eHealth services to identify their challenges, emotions, and wishes. Achieving this aim could result in ideas for improving the usability of eHealth services for older adults with different backgrounds, generating positive emotions regarding eHealth services among older adults, and encouraging them to use these services.

To achieve the aim of this study, a qualitative analysis of immigrant and nonimmigrant older adults' experiences with eHealth services was conducted to uncover the challenges they might face, identify the sources of their perceived feelings, and find out their wishes regarding these services. To focus more deeply and consider individuals' basic psychological needs, this study applied self-determination theory (SDT) to interpret the findings and answer the research questions. SDT is a psychological framework that explains the mechanisms through which individuals develop the ability and motivation to take active roles in various behaviors, such as health care [33].

Theoretical Framework

SDT suggests 3 basic psychological needs for individuals' motivation, behavior, and well-being: autonomy, competence, and relatedness [33]. Autonomy refers to self-endorsed reasons and intrinsic motivations, such as enjoyment or perceived benefits, and alignment with one's core goals for behavioral engagement. Competence refers to a feeling of having the ability and efficacy for development, learning, and expertise. Relatedness is defined as a sense of belonging and having connections with others [31].

XSL•FO
RenderX

SDT can explain support for older adults' motivation to use eHealth services [34]. Regarding autonomy, eHealth services must provide meaningful choices [35]. According to the concept of competence, older adults must feel able to attain better health outcomes by using eHealth services [36], and the complexity of the services should be appropriate for them [31]. Relatedness refers to fostering meaningful relationships with peers and clinicians, such as being in a short feedback loop with health care providers [37].

SDT has been frequently applied to eHealth adoption [38,39]. For instance, Yuen et al [38] found that individuals' eHealth adoption is affected by their emotions. The 3 constructs of SDT significantly impact emotions, and emotions are known to act as one of the main sources of human motivation [40]. To elaborate, individuals experience heightened positive emotions when they perceive a sense of control over their circumstances, possess the requisite skills and knowledge to engage proficiently with eHealth services, and feel a strong connection with and care for others [33].

Research Questions

To achieve the aim of this study, we formulated 3 research questions. The questions explore perspectives of immigrant and nonimmigrant older adults (aged ≥65 years) regarding eHealth services, focusing on their challenges, emotions, and wishes through the lens of SDT. Identifying challenges helps understand barriers preventing older adults from using eHealth services. Recognizing their wishes reveals their needs and expectations. Designing eHealth services based on wishes increases the likelihood of use. Furthermore, by addressing emotions, it is possible to foster positive feelings and encourage greater adoption among older adults. The research questions are as follows: (1) What challenges do older adults face in using eHealth services? (2) What emotions do older adults have regarding eHealth services? (3) What wishes do older adults have regarding eHealth services?

Methods

Study Design

This was an explorative qualitative study based on semistructured interviews with 25 older adults aged ≥65 years conducted between February 2023 and May 2023. This study is part of the DigiIN project and was carried out to generate solutions to ensure that social welfare and health care provided via eHealth services are available and accessible to everyone equally and that everyone can use these services. Therefore, we assessed experiences, feelings, and expectations regarding eHealth services among older adults from different backgrounds in Finland.

We selected participants from different backgrounds: nonimmigrant older adults from cities and rural areas and Iranian immigrants residing in cities. The rationale for this selection was to explore and compare the perspectives of different vulnerable groups. Immigrants face more challenges, and there is a need to understand their perspectives in addition to those of nonimmigrants. In addition, compared to the largest immigrant groups in Finland, Iranians are fewer, which limits their access to similar levels of public and peer support to those available for larger groups. Being fewer, their specific needs are likely to be less well understood within the health care system. Therefore, these individuals are one of the key vulnerable groups as per the major demographic change in Western societies [41].

To maintain a balance in participant recruitment based on their living areas, we aimed to recruit Iranian immigrants and nonimmigrants from rural and urban areas. However, this effort was unsuccessful, and we could not recruit immigrant participants from rural areas. We discovered that most immigrants reside in urban areas. This trend is influenced by Finland's immigration policies and the personal preferences of immigrants.

We used a semistructured interview guide designed for this study by the main researcher and revised with the coauthors (Multimedia Appendix 1). This qualitative study was conducted and reported following the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines where applicable [42] (Multimedia Appendix 2).

Study Setting and Participant Recruitment

We applied the convenience sampling method to recruit the participants. People were eligible if they (1) were aged ≥65 years; (2) were originally from Finland or Iran; and (3) could read and write in Finnish, Swedish, English, or Persian. We excluded people with dementia; those who were too frail or ill; and those who were in active care, such as in hospitals or older adult centers, to reduce possible risks.

We recruited older adults from nonprofit organizations in the Turku and Helsinki regions, including the Turun Seudun Vanhustuki ry (Turku Region Elderly Support Association), Enter Association, Valli Association, and Varissuo Library. These organizations support diversity, inclusion, and the digital empowerment of older adults while fostering intergenerational collaboration. They organize various events and workshops for older adults, giving them access to a broad network within this demographic.

Multiple strategies were used to recruit participants. First, the first author attended events at the aforementioned organizations and directly invited potential participants. Second, the organizations promoted the study to target groups during events, through informal interactions with older adults visiting their facilities, or via social media platforms. Finally, the organizations reached out to potential participants directly to promote the study.

Interested older adults contacted the first author, or the organizations delivered their contact information to her with their permission. The first author then called everyone and inquired about their interest in participating and agreed on the interview time and place. One of the older adults, who was initially interested in participating, did not answer their phone, and 2 refused to participate because of personal issues.

This study included 25 participants, comprising nonimmigrants residing in cities (n=8, 32%), nonimmigrants residing in rural areas (n=9, 36%), and Iranian immigrants residing in cities (n=8,

XSL•FO
RenderX

32%). Most participants were female (17/25, 68%). In total, 56% (14/25) of the participants were aged 65 to 75 years, whereas the remaining 44% (11/25) were aged >75 years. The participants had diverse educational backgrounds: elementary education (5/25, 20%), upper secondary education (1/25, 4%), vocational education (7/25, 28%), and higher education (13/25, 52%). In addition, 6 participants had medical education: 5 (83%) were nonimmigrants, and 1 (17%) had an immigrant background.

Collecting Data

To collect data, we used the designed semistructured interview guide. We collected information about the participants' backgrounds, including their age, gender, and educational level. Moreover, we asked questions about participants' experiences with eHealth services, including whether they perceived any benefits or reasons for not benefiting. We also inquired about their emotions while interacting with these services and their suggestions for improving eHealth services.

We piloted the designed interview guide with 1 immigrant and 1 nonimmigrant older adult. As no major revisions were necessary, we included the pilot interviews in the study with participants' consent. All interviews with immigrant and nonimmigrant older adults from rural areas (17/25, 68%) were conducted over the phone. A total of 50% (4/8) of the interviews with nonimmigrant older adults in the cities were conducted on the web via Zoom (Zoom Video Communications), and the other 50% (4/8) were conducted face-to-face in the Turun Seudun Vanhustuki ry premises.

All interviews were conducted by the first author between February 2023 and May 2023. Participants were interviewed in their preferred languages: English, Swedish, Finnish, or Persian. The interviews were digitally audio recorded with the participants' permission. After each interview, the audio files were transcribed into text files and then deleted. All interviews were translated from Finnish, Swedish, and Persian into English for further analysis. We used transcription and translation services for transcribing and translating the recorded files, resulting in a total of 318 pages. The transcribed files and findings of the study were not revised by the participants.

The average length of the interviews was 35 (SD 8.54) minutes. We continued with the interviews until we reached the saturation point—when we had collected sufficient data to answer the research question or objective and data collection did not result in new analytic codes or themes [43].

Data Analysis

The data were analyzed using a combination of both inductive and deductive content analysis to understand the information domains better [44]. The combination of inductive and deductive content analysis is widely used across various research areas [17,45] and is considered the most realistic approach for incorporating theory and literature into the analysis [44]. To analyze the data, 3 iterative coding steps—reduction, grouping, and abstraction—proposed by Elo and Kyngäs [44] were applied.

The unit of analysis comprised a collection of ideas in which participants detailed the challenges they faced with eHealth services along with their emotions and wishes regarding these services. During the data reduction process, all researchers thoroughly read through the transcripts and used open coding to accurately summarize the set of ideas. All researchers had previous experience in qualitative research. The interview transcripts were then coded collaboratively by all the researchers. The codes were then grouped based on similarity and given descriptive names. These subcategories were reviewed by the main researcher, who further organized them into higher-level categories: challenges (n=7 main themes), emotions (n=2 main themes), and wishes (n=2 main themes). All subcategories were then linked to the 3 basic psychological needs based on SDT where applicable [33]. The ATLAS.ti software (version 24; ATLAS.ti Scientific Software Development GmbH) facilitated the analysis of the data.

Ethical Considerations

The Aalto University ethics review board granted the ethical permission to conduct this study (reference D/1220/03.04/2022). Participation was voluntary, and participants were informed of their right to withdraw from the study at any time without providing a reason. No previous relationships existed between the researchers of the study and the participants.

Before the interviews, we provided participants with oral information about the study, its aims, ethical processing, data collection, preservation of data, and reporting of results. To ensure that the participants were informed, we provided them with a written information sheet. Written informed consent was sought from all participants. Participants received the information in their preferred languages: English, Swedish, Finnish, or Persian.

We anonymized the data to protect the interviewees' information and stored them on a secure local server with strong password protection. All participants were offered a small thank-you gift. The gift was a water bottle provided by the DigiIN project valued at <€20 (US $20.90).

Results

Challenges of Using eHealth Services

Overview

The participants had different experiences with eHealth services. Some of those with medical education shared that they had no major issues using eHealth services for personal purposes, but the rest considered eHealth services challenging and reported several problems, explained in the following sections and summarized in Table 1. In the participant IDs, *C* refers to participants from cities, *R* refers to participants from rural areas, and *I* refers to Iranian immigrants.

XSL•FO
RenderX

Table 1. Immigrant and nonimmigrant older adults' perceived challenges regarding eHealth services in Finland and the links to the basic psychological needs introduced in self-determination theory (SDT).

Challenges	Link to SDT	Nonimmigrants in cities	Nonimmigrants in rural areas	Iranian immigrants
Barriers related to values and preferences				
Valuing in-person visits	Lack of relatedness	✓	✓	✓
Lack of interest	__[a]	✓	✓	✓
Perception of not needing eHealth services	—	✓	✓	✓
Preference for delegating health care responsibilities	—			✓
Usability issues				
Complexity	Lack of competence	✓	✓	✓
Technical terminology	Lack of competence	✓	✓	
Lack of services in users' native languages	Lack of autonomy and lack of competence			✓
Log-in difficulties	Lack of competence		✓	✓
Lack of functionality for leaving comments on or correcting errors in physicians' notes on the patient portal	Lack of autonomy and lack of relatedness	✓		
Lack of possibility to contact health care providers directly through the patient portals	Lack of relatedness	✓		
Lack of digital skills				
Inadequate digital skills for using eHealth services	Lack of competence	✓	✓	✓
Inadequate digital skills for using smart devices	Lack of competence			✓
Lack of proper navigation skills	Lack of competence	✓	✓	✓
Difficulty following updates	Lack of competence	✓	✓	
Lack of support				
Lack of family support	Lack of competence	✓	✓	
Lack of support close to their living area	Lack of competence		✓	
Lack of proper support and meeting their needs	Lack of competence	✓		
Lack of support in their native languages	Lack of competence			✓
Lack of self-efficacy				
Fear of making a serious mistake	Lack of competence	✓	✓	✓
Doubting their abilities to manage unwanted situations	Lack of competence	✓	✓	✓
Doubting their local language proficiency	Lack of competence			✓
Health-related challenges				
Vision and memory issues	Lack of competence	✓	✓	✓
Shaky hands	Lack of competence			✓
Lack of access to digital resources				
Lack of strong e-ID	—		✓	✓

XSL•FO
RenderX

Challenges	Link to SDT	Nonimmigrants in cities	Nonimmigrants in rural areas	Iranian immigrants
Lack of access to smart devices or the internet	—			✓
Lack of familiarity with eHealth services and their value	—		✓	✓

[a]Not applicable.

Barriers Related to Values and Preferences

Some participants (16/25, 64%), even those with successful experiences using eHealth services, reported a preference for face-to-face visits or using eHealth services along with in-person visits, especially for important issues. The participants found it challenging to interact with health care providers and express their issues via web-based health platforms. Therefore, face-to-face visits were assumed to be more effective and reliable in reducing the risk of misdiagnosis. The communication-related challenges were stronger among immigrants because they missed the possibility of using translation services or body language during online visits:

> *When doctors visit, I feel they understand health problems better. Explaining health issues over the phone or via text is not enough; I might miss sharing some of the symptoms, or there might be misunderstandings. I don't want to end up eating the wrong medicines.* [R2]

Some participants reported that the use of eHealth services could decrease their human-based interactions in general, so they preferred to avoid it. They referred to their loneliness and noted that visiting health care providers, often due to their illness, was their main form of face-to-face interaction:

> *I am sick and can't go out alone. Going to hospitals is my only chance to be outdoors and meet people. If I use eHealth, I'll lose this chance.* [C4]

Some participants did not see the added value of eHealth services and, therefore, lacked interest in using them. They mentioned that these services did not match their needs or that identifying a suitable e-service required more time and energy than they could manage. They also noted that not all medical needs could be treated on the web, at least not for their generation.

Furthermore, some participants felt that eHealth services were designed for sick people. Given their self-identification as healthy individuals, they perceived such services as unnecessary and preferred not to use them. In addition, some immigrants noted their preferences for delegating their health care responsibilities to family members as they assumed themselves to be old and incapable of managing health issues, especially in their nonnative languages. In addition, they believed that, just as they had cared for their parents, their children should take care of them. Therefore, they did not feel the need to use eHealth services:

> *I don't know where to go or what to do. I don't even know the language very well. It is better if my kids handle things. When my mother was old and sick, I was taking care of her, and my kids are doing the same for me.* [I4]

Usability Issues

Some participants (15/25, 60%), especially nonimmigrants residing in cities, raised several usability issues. The primary usability concern was the intricate nature of these services and their improper interface design, which resulted in the overload of information, hidden information, and navigation difficulties. Participants reported several navigation challenges, including difficulty distinguishing between clickable and nonclickable elements in some health-related web pages:

> *On some pages, it's hard to know where to click, where not to click, or even when clicking is an option.* [R1]

The usability issues reported mainly concerned My Kanta, the Finnish national patient portal. Regarding complexity, the participants noted that the main page of My Kanta is overloaded with extra information, which makes it hard to find relevant information. They also added that clicking on certain elements on My Kanta often led to new pages overloaded with excessive details, which they found overwhelming:

> *I clicked on a link it opened another window with much information. Made it even more difficult to find what I need.* [R3]

Furthermore, some participants raised concerns about navigating back to the home page, stating that, after several clicks, they found it challenging to return to the home page quickly. Another important usability issue was related to the language. Immigrants said that their effective use of eHealth services was diminished by the lack of services in their native languages, whereas nonimmigrants considered medical or technical terminologies to be barriers to their efficient use.

Some participants, especially those with memory issues, reported that the log-in process to My Kanta was difficult and time-consuming. They also expressed frustration with being required to log in again due to session timeouts or opening another window.

Some nonimmigrants in cities also expressed discontent caused by the lack of functionality for leaving comments on or rectifying errors in physicians' notes on patient portals. This problem was said to limit users' autonomy in managing the trustworthiness of their information. In addition, participants reported feeling disconnected from health care providers. The same participants also shared that the lack of possibility to contact health care providers directly through My Kanta further contributed to this sense of disconnection:

XSL•FO
RenderX

> *I had an accident a few weeks ago and went to the hospital. A couple of days after my visit, I checked My Kanta to read the doctors' notes. There was false information, and I couldn't get rid of it or leave a comment on it. The note said that I had fainted when I went to the hospital, but I hadn't. With this wrong information, what's the point of using My Kanta then?* [C6]

Lack of Digital Skills

Some participants (15/25, 60%) reported challenges due to their lack of the adequate digital skills needed to use smart devices and the internet for critical issues such as health. The participants reported using smart devices and the internet to meet their daily needs, such as contacting their networks or using social media, but not for serious matters. They felt that using eHealth services required a high level of digital skills compared to using other e-services:

> *I have a smartphone, I call people, and I am active on Facebook, but I don't use digital health services. You must be good enough with computers to use them for health issues. I am not.* [R5]

Furthermore, some participants, especially immigrants, reported that their insufficient digital skills caused them difficulties navigating eHealth platforms. This perceived difficulty led to issues with finding information:

> *I can't find the information I need on My Kanta; I don't have skills for that.* [I3]

Some nonimmigrant participants expressed that their lack of adequate digital skills made it challenging to follow updates in eHealth services or health apps. They considered themselves slow adopters of technology. Moreover, 25% (2/8) of the immigrants reported having difficulties dealing with smart devices in general due to their lack of experience with technology-based tools.

Lack of Support

The lack of support was a prevalent concern among some participants (14/25, 56%). The lack of support, they said, could create the impression that these services are not intended for older adults, especially for those who are not capable of using smart devices. Some nonimmigrants cited a lack of family support. They either lacked a family member to turn to with their queries or hesitated to seek assistance believing that their family members could not satisfy their learning needs. Furthermore, nonimmigrants in cities mentioned that either they were unfamiliar with the available support or the available support did not satisfy their needs. Nonimmigrants in rural areas reported a lack of support close to their living areas, and immigrants noted a lack of support in their native languages.

Lack of Self-Efficacy

A total of 44% (11/25) of the participants reported a fear of making serious mistakes while using eHealth services. They were afraid that something irreversible might happen, leading to unwanted and unpleasant results with long-term consequences, such as the nonrenewal of prescriptions or the deletion of important information. They were worried and doubted their capabilities to manage the situation:

> *I use computers for writing emails or watching videos, but I never dare to use eHealth services alone. I don't want to make any mistakes or cause any problems. Not for me not for doctors.* [I2]

Furthermore, some immigrants, even those with a good understanding of the Finnish language, mentioned a lack of trust in their Finnish-language proficiency. This lack of trust prevented them from using the eHealth service in Finnish.

Health-Related Challenges

Health-related challenges such as memory or vision problems were common among 36% (9/25) of the participants and caused difficulties remembering log-in passwords, navigating, or using e-services on mobile phones with small screens. Some participants also mentioned that shaky hands caused difficulty holding phones or tablets or typing text.

Lack of Access to Digital Resources

Some participants shared basic challenges that prevented them from using eHealth services. For example, 25% (2/8) of the immigrants reported a lack of access to smart devices (eg, computers or smartphones) and the internet. In total, 62% (5/8) of the immigrants noted their lack of familiarity with the available eHealth services. However, the lack of general familiarity was only an issue among immigrant older adults. Participants from all groups mentioned a lack of familiarity with the functionalities and benefits of eHealth services:

> *I know about this Kanta; I have heard from my children. But I don't know what exactly it has for me.* [R6]

Moreover, 25% (2/8) of the immigrants and 11% (1/9) of the nonimmigrants in rural areas shared that they did not have a strong e-ID code, which is a prerequisite for using public e-services in Finland. Therefore, they could not use some e-services such as My Kanta.

Emotional Experiences Using eHealth Services

Overview

Participants' experiences with eHealth services generated positive and negative emotions, which are analyzed and summarized in Tables 2 and 3.

XSL•FO
RenderX

Table 2. Immigrant and nonimmigrant older adults' positive emotions regarding eHealth services in Finland and the links to the basic psychological needs introduced in self-determination theory (SDT).

Positive emotion	Definition	Link to SDT	Nonimmigrants in cities	Nonimmigrants in rural areas	Iranian immigrants
Trust	Feeling that eHealth services are safe, secure, and reliable	—[a]	✓	✓	✓
Interest	Willingness to learn how to use eHealth services as people understand their usefulness or consider them easy to use	Competence	✓	✓	✓
Interest	Willingness to use eHealth services as they offer meaningful choices to satisfy patients' needs and people feel that they have enough skills to use these services	Autonomy and competence	✓	✓	✓
Being empowered	Having more control over their health situation, making independent decisions, and being more involved in their health care management	Autonomy and competence	✓	✓	✓
Being less worried	Being less worried about their health condition	—	✓	✓	✓
Being connected	Being connected to health care providers	Relatedness	✓	✓	✓
Hope	Reduction in migration rate from rural areas for health care purposes and hoping to save their community	—		✓	
Peace of mind	Access to health care services in rural areas; freedom to live in their favorite places	Autonomy		✓	
Equality	Equal access to health care services in the country	—		✓	

[a]Not applicable.

XSL•FO
RenderX

Table 3. Immigrant and nonimmigrant older adults' negative emotions regarding eHealth services in Finland and the links to the basic psychological needs introduced in self-determination theory (SDT).

Negative emotion	Definition	Link to SDT	Nonimmigrants in cities	Nonimmigrants in rural areas	Iranian immigrants
Confusion	Dealing with the complicated interface design of eHealth services was confusing	Lack of competence	✓	✓	✓
Lack of interest	Not willing to use eHealth services as they are complex and the use of such services could limit in-person interactions with health care providers	Lack of competence and lack of relatedness	✓	✓	✓
Lack of interest	Not willing to learn how to use eHealth services for different reasons, such as being old and incapable of learning	Lack of competence	✓	✓	✓
Frustration	Feeling frustrated with learning and using complex eHealth services	Lack of competence	✓	✓	✓
Stress	Feeling stress when using complex eHealth services	Lack of competence	✓	✓	
Dependency and shame	Feeling shame of being dependent on others and asking questions even regarding simple issues	Lack of autonomy and lack of competence	✓	✓	✓
Use by force	Feelings of being forced to use eHealth services	Lack of autonomy	✓	✓	
Alienation	Feelings of being excluded due to lack of services in their native languages	Lack of relatedness			✓
Disconnection	Lack of direct connection with health care providers via My Kanta to discuss recorded notes	Lack of relatedness	✓	✓	

Positive Emotions

Our participants shared several positive emotions. Most (19/25, 76%) expressed a prominent level of trust in eHealth services in Finland and the security of their online medical records. This was attributed to trust in the authorities and the Finnish government:

> *I believe that the Finnish government is taking care of the security of our health data.* [C2]

Some participants (13/25, 52%) reported an interest in learning about and using eHealth services. They found the services easy to understand and use. The nonimmigrant participants attributed eHealth services' perceived ease of use to their previous experience with computers and the internet. In addition, 47% (8/17) of the nonimmigrants found the services useful and appreciated the convenience and time savings they offered. They believed that these services provided meaningful options to satisfy different health-related needs, for example, booking appointments or adhering to the treatment process:

> *I like these digital health services; they help me a lot. I have an app on my phone that reminds me to take my medications. Without the app, I might miss to take my pills. Or...this hospital app that I use [Terveystalo], I text doctors. The same app helps me book appointments.* [C2]

The immigrant participants also shared their interest in learning about and using the services but without giving any specific reasons. However, some of them expressed interest in learning

XSL•FO
RenderX

about and using these services after we provided them with information on their usefulness.

Some participants (9/25, 36%) reported feeling empowered by using eHealth services for different reasons. First, they said that having access to their medical records enabled them to control their health situation, such as renewing prescriptions and making informed decisions such as booking or canceling an appointment:

> *By using My Kanta, I control my medicines and renew my prescriptions if I need them. I read my lab results or doctors' notes carefully, then I decide to book or cancel an appointment. And sometimes, I prepare some questions to ask in my visits based on my medical records.* [R1]

Second, some participants described feeling empowered by the possibility of contacting health care providers at any time through the chat option. This functionality allowed them to be more actively involved in managing their health care.

Furthermore, 36% (9/25) of the participants found that access to their medical records helped them enhance their understanding of their health status and reduced their worries. Some participants, especially those living far from physical health care facilities, reported positive feelings about being connected to health care providers via chat or phone calls. Nonimmigrants in rural areas shared 3 positive feelings: equality in terms of using health care services in the country, hope that the use of eHealth services could save their community by reducing the migration rate from rural to urban areas for health care purposes, and peace of mind to live alone in rural areas due to access to eHealth services from their home. The availability of eHealth services was said to offer the autonomy to choose preferred living arrangements:

> *I live in an archipelago alone, far from hospitals. I have appointments with doctors online when the issue is not urgent. I don't need to travel to cities for that reason. People used to move to cities when they were old or sick, but nowadays it is possible to do things online and live in archipelagos, at home.* [R4]

Negative Emotions

Our participants reported several negative emotions regarding eHealth services, with confusion being the most frequently mentioned. A total of 64% (16/25) of the participants, especially those with limited experience using smart devices and those with poor self-reported digital skills, reported that the interface design of eHealth services was complex and confused them, at least in their first experience:

> *In my first experiences with My Kanta, I found it very complex and confusing. I had no idea where the information was or what to do. It is still confusing but not like the first time.* [C6]

The complexity of the services made some participants (7/25, 28%) feel that these services were not designed for people of their age and could not satisfy their needs. Thus, they did not like to interact with such services. In addition, they were afraid of missing in-person interactions with health care providers. Moreover, 28% (7/25) of the participants were not interested in learning new skills as they assumed themselves to be very old and incapable of learning how to use eHealth services. Even those who aimed to learn about eHealth services found it frustrating and noted that the complexity of the services made the learning process more challenging:

> *I am too old and don't want to learn new things; even if I want to, I can't. It is hard for me to memorize and remember. I just want you to make things that I know how to use easier.* [I5]

In total, 28% (7/25) of the participants, those with experience using eHealth services, experienced stress or were irritated when interacting with such complex services. Furthermore, some participants (6/25, 24%) reported that the use of eHealth services made them dependent on others and feel ashamed. As they did not have enough digital skills or experience with computers or e-services, they found the services complex and had to ask for help constantly.

Moreover, 35% (6/17) of the nonimmigrants felt that engaging with eHealth services was somehow obligatory. Despite their initial reluctance, they perceived an external force to adopt these services to remain connected with their community, avoid falling behind youth, and keep pace with developments in health technology:

> *When my kids told me about these digital health services, I found them very difficult to learn, but I said I had to learn them. If I don't learn now, I will be an outsider soon. I think everything will be digitalized soon.* [C5]

In addition, most immigrants (5/8, 62%) reported feelings of being excluded as the available eHealth services were not in their native languages. Nonimmigrant participants expressed a different concern regarding being disconnected. They mentioned that the lack of possibility to contact health care providers directly through My Kanta to discuss recorded notes caused feelings of disconnection:

> *I feel like these services aren't meant for me. If they wanted me to use these eHealth services, they would make them available in my language.* [I7]

Wishes Regarding eHealth Services

Overview

Participants' interactions with eHealth services and their perceived challenges shaped their wishes regarding these services. The identified wishes are explained in the following sections and summarized in Table 4.

XSL•FO
RenderX

Table 4. Immigrant and nonimmigrant older adults' wishes regarding eHealth services in Finland and the links to the basic psychological needs introduced in self-determination theory (SDT).

Wish	Definition	Link to SDT	Nonimmigrants in cities	Nonimmigrants in rural areas	Iranian immigrants
Wishes for all kinds of eHealth services					
User-friendly design	Simple interface	Competence	✓	✓	✓
Support services	Improve digital skills and efficacy	Competence	✓	✓	✓
Unified platform	A single local system	Competence	✓	✓	
Language	eHealth services without technical terminology	Competence	✓	✓	
Language	eHealth services in users' native languages	Competence and autonomy			✓
Facilitate log-in	Log in to eHealth services without an e-ID	Competence		✓	✓
Fewer updates	Zero to few updates	Competence	✓	✓	
Wishes only for My Kanta					
Speed up access	Upload medical records on My Kanta immediately	—[a]	✓	✓	
Notification	Sending notifications to users' mobile phones when new records are uploaded	—	✓		
Edit or comment	Possibility to edit or comment on physicians' notes on My Kanta	Autonomy	✓	✓	
Direct contact	Discuss queries on the recorded health data with health care providers via My Kanta directly	Relatedness	✓	✓	

[a]Not applicable.

Identified Wishes

In Finland, several similar eHealth platforms offer the same services, leading to confusion among users. To address this issue, 40% (10/25) of the participants suggested creating a unified platform that consolidates these services:

> *It is confusing when different sites are giving the same services. It takes time and energy to learn how to use them. It's good if, in one place, you can do very many different things.* [R4]

This single platform was expected to have a user-friendly design for all user groups. Our participants shared several suggestions, such as having only main headlines with big font sizes on the home page; using icons alongside each heading to facilitate information seeking; and minimizing the use of large images on the home page as it might cause navigation difficulties, especially on mobile devices. In addition, some participants emphasized incorporating color into the design. The coloring system was suggested as a potential aid for navigation, especially for people with vision impairments.

Several suggestions were given specifically for My Kanta. These included adding the possibility of listening to the medical records or physicians' notes, the possibility of navigating back to the home page by clicking on the large Kanta icon in the corner of the web page, sharing medical records without any delay, sending notifications to users' mobile phones for new or unchecked records, and the availability of a hotline or chat option as a channel to contact health care providers directly to discuss queries or share worriers about the recorded medical data. These participants wished to have the possibility to edit or comment on physicians' notes, too:

> *Sometimes when I read doctors' notes, I find them hard to understand or I disagree. I wish there was a direct channel to call or text the same doctor who has written the notes and discuss.* [C3]

Furthermore, due to the importance of completely understanding health information, immigrants wished to have at least critical health information in their native languages, and nonimmigrants suggested that health information should be free of medical terminology. In addition, some participants shared the idea of having the possibility to log in without an e-ID. In Finland, to increase security, an e-ID is necessary to log in to public e-services; however, some participants, especially immigrants, found this challenging. Moreover, some participants wished for services without any updates or that, in the case of changes, there should be announcements and proper support. Tailored support in individuals' native languages and close to their living areas was requested by most participants (19/25, 76%).

XSL•FO
RenderX

Discussion

Principal Findings and Comparison With Prior Work

This study used a qualitative exploratory approach to investigate the perspectives of immigrant and nonimmigrant older adults on eHealth services, focusing on their perceived challenges, emotions, and wishes. In addition, it highlighted the role of the basic psychological needs of competence, autonomy, and relatedness, as outlined in SDT, in shaping these perspectives. To support the use of eHealth services by older adults regardless of their background, our findings outline several practical recommendations for developing user-friendly eHealth services and digital supports.

We found that the most common barriers among immigrant and nonimmigrant participants were their values and preferences, inadequate digital skills, and usability issues. Immigrants faced more challenges related to personal values and preferences and inadequate digital skills, whereas nonimmigrants complained that their eHealth use was mainly diminished by usability issues.

We found that barriers related to values and preferences hindered older adults' use of eHealth services for different reasons. First, some participants were reluctant to use eHealth services as they did not understand their added value or highly valued in-person visits to overcome web-based communication challenges. Communication issues varied among participants by background. Although a common problem was the fear of losing human-based interactions, the difficulty in using translation services or body language hindered immigrants without proficiency in the local language from using eHealth services. These findings align with the findings of Wang et al [46]. We showed that the communication-related barriers prevented older adults from satisfying their need for a sense of belonging, reflected in the relatedness aspect of SDT.

Second, participants preferred to use only physical health care facilities or eHealth services aligned with existing in-person services to avoid misdiagnoses, which is in line with previous studies [47,48]. Third, our findings add to the body of knowledge regarding the fact that some Iranian older adults intentionally preferred to delegate their health care responsibilities to their family members and were reluctant to be involved in their health care. Therefore, they did not feel a need to learn about eHealth services. They justified their behavior with their lack of familiarity with the local health care system, language issues, and their own culture. This finding suggests that, even if issues related to language barriers and familiarity with the local health care system are resolved, some Iranian older adults may still prefer to rely on caregivers such as their children. This preference is rooted in their cultural values, which emphasize that older adults should be cared for by younger family members.

The importance of culture has been highlighted previously. Previous studies have considered culture an enabler of learning and using eHealth technologies among some older adults, especially immigrants. The literature shows that older adults' use or nonuse of eHealth services relates to their technology-related cultural perceptions and the contextual factors of their environment and lifestyle [30,49,50]. DeLange Martinez et al [51] identified ethnicity and culture as significant predictors of technology use among Asian immigrant older adults.

Another important group of challenges related to digital skills. Participants' challenges with inadequate digital skills manifested in various ways. Most nonimmigrants and some immigrants felt that their digital skills did not ensure their proficiency in using eHealth services; however, they could use smart devices and search the internet for noncritical topics. In contrast, some immigrants faced greater challenges; they had not been taught the basic digital skills required to use smart devices and the internet in general. These findings are in line with those of previous studies [17,18]. We found that all types of difficulties in terms of digital skills prevented older adults from satisfying their need for a feeling of being competent users of eHealth services (competence aspect of SDT).

Moreover, we identified several usability issues raised mainly by nonimmigrant participants residing in cities. This group of participants had more experience with eHealth services; therefore, they had more knowledge regarding usability issues. A commonly identified usability issue among participants regardless of their immigrant or nonimmigrant status was the complexity of eHealth services and their intricate interfaces. This challenge has been highlighted in previous research [12,17,52,53]. However, the perceived complexity was dependent on other background characteristics; for instance, participants with medical backgrounds found these services less complicated than other participants.

The perceived complexity of eHealth services coupled with inadequate digital skills prevented older adults from satisfying their basic psychological needs outlined by SDT and generated feelings of confusion, stress, and frustration when interacting with eHealth services. Similarly to the study by Chen et al [24], we showed that failure to satisfy the psychological need for being competent users of eHealth services (competence aspect of SDT) led to negative feelings of frustration and anxiety among older adults. Moreover, we add to the literature that the 2 aforementioned challenges of perceived complexity and inadequate digital skills led some participants to repeatedly seek assistance when using the services. On the basis of SDT, these challenges hindered older adults from satisfying their basic psychological need for feelings of being autonomous and competent users of eHealth services (autonomy and competence aspects of SDT), resulting in feelings of shame and reliance on others.

The second most common usability issue—language-related challenges—varied among immigrant and nonimmigrant older adults. Medical terminology use was challenging for nonimmigrants, whereas immigrants faced a lack of eHealth services in their native languages, similar to the findings of Kaihlanen et al [17]. Lack of services in immigrants' native languages leads to the experience of alienation and exclusion (ie, hindering them from meeting their basic psychological need for a sense of belonging [relatedness aspect of SDT]). This discovery aligns with the findings of Safarov [54]. Our findings contribute to the literature by revealing that other usability issues

XSL•FO
RenderX

prevented the satisfaction of the basic psychological need for a sense of belonging among nonimmigrant participants. Nonimmigrants felt disconnected due to the lack of functionality for leaving comments correcting errors in physicians' notes on My Kanta or contacting health care providers directly through My Kanta to discuss their medical records.

Our findings highlight other differences between immigrant and nonimmigrant groups. We found that, despite the perceived challenges and initial reluctance to use eHealth services, some nonimmigrant older adults felt obligated to use these services as a means of staying connected with their community and keeping pace with younger generations. The sense of obligation to adopt innovations may be influenced by Finnish culture or social norms, although we lack concrete evidence to substantiate this claim. This feeling prevented some nonimmigrant older adults from satisfying their psychological need for the feeling of having freedom of choice in using eHealth services (autonomy aspect of SDT). Moreover, Kaihlanen et al [17] showed that not possessing an e-ID was a barrier for immigrants to access eHealth services. Our study added to their findings in that some nonimmigrants residing in rural areas had the same issue.

In addition to the perceived challenges and negative emotions, older adults' interactions with eHealth services generated several positive emotions. Fostering them could support older adults' motivations to use eHealth services. We found that some older adults were interested in learning about eHealth services and using them. However, only nonimmigrants described reasons for their interest. First, they perceived such services as easy to interact with to manage some of their health care needs. This perception signals their basic psychological need for a feeling of being competent users of eHealth services (competence aspect of SDT) being satisfied. Second, nonimmigrants justified their interest as eHealth services giving them access to different meaningful options to satisfy their health care needs independently, which was linked in the analysis to the satisfaction of their psychological need for autonomy. These findings regarding reasons supporting nonimmigrants' interest in eHealth services confirm those of previous studies [55,56].

Moreover, we found that nonimmigrants residing in rural areas expressed more positive feelings compared to others, including equality, hope, and peace of mind, which in the analysis was considered to stem from the satisfaction of their basic psychological need for autonomy in selecting their living places due to the possibility of using eHealth services from rural areas. In addition, satisfaction with being able to contact health care providers (relatedness aspect of SDT) led to the positive feeling of being connected, for example, through the possibility of using chat or phone call options to contact health care professionals for nonemergencies.

Overall, the conducted interviews with older adults yielded several insights for supporting their use of eHealth services that are tied to supporting their basic psychological needs. The first suggestion was related to the design of the services. The participants complained about various available eHealth platforms offering similar services and mentioned being confused. Thus, they suggested that a concrete improvement could be to develop a unified eHealth platform that consolidates multiple services. This suggestion has been made previously [56,57]. However, instead of developing a completely new platform, we recommend that this could be realized through complementing existing patient portals with missing features beyond those available in their current versions. Our key recommendations are to include within such portals options for direct web-based connections with health care providers to discuss older adults' queries about their medical records and the possibility of commenting on their medical records. Such options could satisfy the psychological need to feel connected (relatedness aspect of SDT) among older adults. To satisfy the need for autonomy (as per SDT), we suggest developing the services in multiple languages and using plain terminology. In addition, the platform should have a user-friendly interface design by incorporating visualization tools such as icons alongside each heading, using coloring systems that facilitate navigation, simplifying log-in procedures, and minimizing the frequency of changes to the user interfaces. Adding these features could support satisfying the need to feel competent among older adults.

The second recommendation was to support older adults' digital skills considering their backgrounds. The importance of digital support has been well documented in previous studies [24,30]. Our findings confirm this need, highlighting that existing support services remain insufficient or unsuitable for the diverse needs of older adults, especially immigrants. Our findings confirm that digital support services should be provided in multiple languages considering the cultural differences and needs of different groups of older adults. Similarly, Anand et al [58] highlighted the importance of considering the country of origin and cultural factors in educating older adults. As a practical solution to promote culturally oriented digital support for immigrants, we recommend that translators assigned by the local health care system to assist immigrant older adults during health care visits also be trained to provide digital support services tailored to their specific needs and culture. However, before introducing digital support services, we suggest raising awareness among older adults, particularly immigrants, about the benefits of independently managing health issues via eHealth services. This foundational information can be disseminated through multiple channels. For example, the government agencies responsible for administering national social security programs with contact to all residents could serve as an effective channel. Moreover, to address the identified issue of older adults' reluctance to use eHealth services, which they view as replacing personal interactions that serve a vital social function for lonely individuals, we suggest that society prioritize combating loneliness as a significant and separate social challenge.

Limitations

This study's major limitation lies in its sample selection as it exclusively included immigrants from Iran, who are ranked 13th on the list of the most populous groups of immigrants in Finland [57]. Several reasons supported this selection. First, the first author was proficient in Persian, enabling effective communication with immigrant participants. Second, this group of immigrants has never been the focus of similar studies in

XSL•FO
RenderX

Finland; therefore, their perspectives and needs regarding the Finnish health care system were unknown. This limitation impacts the generalizability of the findings to other immigrant groups in Finland and elsewhere even if we find it plausible that the observed challenges are not limited to Iranians in Finland. For the future, we recommend a larger study with participants from various countries.

Another limitation is the lack of focus on a specific eHealth service. Instead, we allowed participants to raise their most important challenges concerning any type of eHealth service, which allowed us to survey the eHealth landscape more broadly from the user perspective. In addition, this study only compared results based on participants' background information regarding their rural or urban residency and whether they were immigrants. To gain a deep understanding of the background information, we recommend conducting a study that compares the use of eHealth services based on other background information, such as educational level or occupation. It is worth noting that some participants had a medical background. Furthermore, this study collected data solely through interviews and self-reports without conducting usability tests. We recommend a similar study with a follow-up usability test.

Conclusions

The results of 25 interviews with immigrant and nonimmigrant older adults indicate that, while the use of eHealth services is easy for minority groups with medical backgrounds, others face several difficulties. Personal values and preferences, insufficient digital skills, and usability issues were common challenges. The first 2 challenges were more obvious among immigrant older adults, whereas nonimmigrants shared more usability issues due to their familiarity with the available eHealth services. The perceived challenges prevented older adults from satisfying their basic psychological needs for competence, autonomy, and relatedness and generated negative emotions. The Iranian immigrant participants shared more negative feelings regarding eHealth services as they faced more challenges.

To support older adults in the pursuit of well-being via eHealth services, users' basic psychological needs and their wishes regarding eHealth services should be considered from the design stage to the development and implementation processes. To meet this goal, developing a unified eHealth platform with a user-friendly design that consolidates multiple services is of crucial importance to mitigate perceived complexity and the need for advanced digital skills to use such a service because most older adults are concerned that their digital skills are not enough to use complex eHealth services. In addition, a tailored digital support service considering the cultural diversity of users should be available and accessible in multiple languages for older adults. However, before offering digital support, it is essential to promote the importance of managing health issues independently, particularly for those who prefer to delegate their health care responsibilities to their caregivers. These actions could foster positive feelings and support older adults' motivation to use eHealth services.

Data Availability

The datasets generated or analyzed during this study are not publicly available and cannot be provided upon request due to security and privacy concerns.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 17 KB - jmir_v27i1e64249_app1.docx]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 481 KB - jmir_v27i1e64249_app2.pdf]

References

1. Valkama P, Oulasvirta L. How Finland copes with an ageing population: adjusting structures and equalising the financial capabilities of local governments. Local Gov Stud 2021 Jan 25;47(3):429-452. [doi: 10.1080/03003930.2021.1877664]
2. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-04-29]
3. World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100. Department of Economic and Social Affairs, United Nations. 2017. URL: https://www.un.org/ht/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100 [accessed 2024-04-29]
4. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/258981 [accessed 2024-07-10]

XSL•FO
RenderX

5. van Het Bolscher-Niehuis MJ, den Ouden ME, de Vocht HM, Francke AL. Effects of self-management support programmes on activities of daily living of older adults: a systematic review. Int J Nurs Stud 2016 Sep;61:230-247. [doi: 10.1016/j.ijnurstu.2016.06.014] [Medline: 27400029]
6. Parsons K, Gaudine A, Swab M. Experiences of older adults accessing specialized health care services in rural and remote areas: a qualitative systematic review. JBI Evid Synth 2021 Jun;19(6):1328-1343. [doi: 10.11124/JBIES-20-00048] [Medline: 34111043]
7. Feng Z, Glinskaya E, Chen H, Gong S, Qiu Y, Xu J, et al. Long-term care system for older adults in China: policy landscape, challenges, and future prospects. Lancet 2020 Oct 24;396(10259):1362-1372. [doi: 10.1016/S0140-6736(20)32136-X] [Medline: 34338215]
8. Li D, Hu Y, Pfaff H, Wang L, Deng L, Lu C, et al. Determinants of patients' intention to use the online inquiry services provided by internet hospitals: empirical evidence from China. J Med Internet Res 2020 Oct 29;22(10):e22716 [FREE Full text] [doi: 10.2196/22716] [Medline: 33006941]
9. Olu O, Muneene D, Bataringaya JE, Nahimana M, Ba H, Turgeon Y, et al. How can digital health technologies contribute to sustainable attainment of universal health coverage in Africa? A perspective. Front Public Health 2019 Nov 15;7:341 [FREE Full text] [doi: 10.3389/fpubh.2019.00341] [Medline: 31803706]
10. Global observatory for eHealth. World Health Organization. 2018. URL: https://www.who.int/observatories/global-observatory-for-ehealth [accessed 2024-04-29]
11. Lindberg J, Lundgren AS. The affective atmosphere of rural life and digital healthcare: understanding older persons' engagement in eHealth services. J Rural Stud 2022 Oct;95:77-85. [doi: 10.1016/j.jrurstud.2022.07.020]
12. Simola S, Hörhammer I, Xu Y, Bärkås A, Fagerlund AJ, Hagström J, et al. Patients' experiences of a national patient portal and its usability: cross-sectional survey study. J Med Internet Res 2023 Jun 30;25:e45974 [FREE Full text] [doi: 10.2196/45974] [Medline: 37389909]
13. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021 Feb 22;23(2):e23786 [FREE Full text] [doi: 10.2196/23786] [Medline: 33616534]
14. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017 Nov 27;3:51 [FREE Full text] [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
15. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
16. Istepanian RS, Lacal JC. Emerging mobile communication technologies for health: some imperative notes on m-health. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2003 Presented at: IEMBS '03; September 17-21, 2003; Cancun, Mexico p. 1414-1416 URL: https://ieeexplore.ieee.org/document/1279581 [doi: 10.1109/iembs.2003.1279581]
17. Kaihlanen AM, Virtanen L, Buchert U, Safarov N, Valkonen P, Hietapakka L, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188 [FREE Full text] [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
18. Blažič BJ, Blažič AJ. Overcoming the digital divide with a modern approach to learning digital skills for the elderly adults. Educ Inf Technol 2019 Jul 19;25(1):259-279. [doi: 10.1007/S10639-019-09961-9]
19. Abdullahi Yari I, Dehling T, Kluge F, Geck J, Sunyaev A, Eskofier B. Security engineering of patient-centered health care information systems in peer-to-peer environments: systematic review. J Med Internet Res 2021 Nov 15;23(11):e24460 [FREE Full text] [doi: 10.2196/24460] [Medline: 34779788]
20. Pool J, Akhlaghpour S, Fatehi F, Gray LC. Data privacy concerns and use of telehealth in the aged care context: an integrative review and research agenda. Int J Med Inform 2022 Apr;160:104707. [doi: 10.1016/j.ijmedinf.2022.104707] [Medline: 35131698]
21. Wegener EK, Bergschöld JM, Whitmore C, Winters M, Kayser L. Involving older people with frailty or impairment in the design process of digital health technologies to enable aging in place: scoping review. JMIR Hum Factors 2023 Jan 27;10:e37785 [FREE Full text] [doi: 10.2196/37785] [Medline: 36705959]
22. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 01;24(e1):e47-e54 [FREE Full text] [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
23. ISO 9241-11:1998(EN): ergonomic requirements for office work with visual display terminals (VDTs) — part 11: guidance on usability. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en [accessed 2024-04-29]
24. Chen X, Frennert S, Östlund B. The use of information and communication technology among older immigrants in need of home care: a systematic literature review. Ageing Int 2021 Jul 13;47(2):238-264. [doi: 10.1007/s12126-021-09417-x]
25. Saariluoma P, Jokinen JP. Emotional dimensions of user experience: a user psychological analysis. Int J Hum Comput 2014 Mar 07;30(4):303-320. [doi: 10.1080/10447318.2013.858460]

XSL•FO
RenderX

26. Kujala S. User involvement: a review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
27. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
28. Act on the provision of digital services 306/2019. Ministry of Finance. 2019. URL: https://www.finlex.fi/en/laki/kaannokset/2019/en20190306 [accessed 2024-04-29]
29. Hänninen R, Pajula L, Korpela V, Taipale S. Individual and shared digital repertoires – older adults managing digital services. Inf Commun Soc 2021 Jul 20;26(3):568-583. [doi: 10.1080/1369118x.2021.1954976]
30. Airola E. Learning and use of eHealth among older adults living at home in rural and nonrural settings: systematic review. J Med Internet Res 2021 Dec 02;23(12):e23804 [FREE Full text] [doi: 10.2196/23804] [Medline: 34860664]
31. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
32. Yi M, Hui Y, Hu L, Zhang W, Wang Z. The experiences and perceptions of older adults with multimorbidity toward e-health care: a qualitative evidence synthesis. Telemed J E Health 2024 Oct 26;30(10):2527-2544. [doi: 10.1089/tmj.2024.0211] [Medline: 38920002]
33. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
34. Peters D. Wellbeing supportive design – research-based guidelines for supporting psychological wellbeing in user experience. Int J Hum Comput Interact 2022 Aug 18;39(14):2965-2977. [doi: 10.1080/10447318.2022.2089812]
35. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. 'Having to learn this so late in our lives…' Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar 08;37(1):41-52 [FREE Full text] [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
36. Hoffman L, Wisniewski H, Hays R, Henson P, Vaidyam A, Hendel V, et al. Digital opportunities for outcomes in recovery services (DOORS): a pragmatic hands-on group approach toward increasing digital health and smartphone competencies, autonomy, relatedness, and alliance for those with serious mental illness. J Psychiatr Pract 2020 Mar;26(2):80-88 [FREE Full text] [doi: 10.1097/PRA.0000000000000450] [Medline: 32134881]
37. Vollenbroek-Hutten M, Goolkate T, Lankheet M, Hegeman JH, Laverman G, Van Beijum BJ. Insufficient behavioral change skill hampers adoption of eHealth services. In: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health. 2018 Presented at: ICT4AWE '18; March 22-23, 2018; Madeira, Portugal p. 119-125 URL: https://tinyurl.com/yeymvudz [doi: 10.5220/0006675501190125]
38. Yuen KF, Chua JY, Li X, Wang X. The determinants of users' intention to adopt telehealth: health belief, perceived value and self-determination perspectives. J Retail Consum Serv 2023 Jul;73:103346. [doi: 10.1016/j.jretconser.2023.103346]
39. Assadi V, Hassanein K. Consumer adoption of personal health record systems: a self-determination theory perspective. J Med Internet Res 2017 Jul 27;19(7):e270 [FREE Full text] [doi: 10.2196/jmir.7721] [Medline: 28751301]
40. Pekrun R. Teachers need more than knowledge: why motivation, emotion, and self-regulation are indispensable. Educ Psychol 2021 Nov 29;56(4):312-322. [doi: 10.1080/00461520.2021.1991356]
41. Numans W, Regenmortel TV, Schalk R, Boog J. Vulnerable persons in society: an insider's perspective. Int J Qual Stud Health Well-being 2021 Dec 28;16(1):1863598 [FREE Full text] [doi: 10.1080/17482631.2020.1863598] [Medline: 33357080]
42. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
43. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Mar;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
44. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
45. Cho H, Porras T, Flynn G, Schnall R. Usability of a consumer health informatics tool following completion of a clinical trial: focus group study. J Med Internet Res 2020 Jun 15;22(6):e17708 [FREE Full text] [doi: 10.2196/17708] [Medline: 32538796]
46. Wang Y, Do DP, Wilson FA. Immigrants' use of eHealth services in the United States, national health interview survey, 2011-2015. Public Health Rep 2018 Nov 17;133(6):677-684 [FREE Full text] [doi: 10.1177/0033354918795888] [Medline: 30223716]
47. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: national survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26 [FREE Full text] [doi: 10.2196/mhealth.8619] [Medline: 29362211]

XSL•FO
RenderX

48. Currie M, Philip LJ, Roberts A. Attitudes towards the use and acceptance of eHealth technologies: a case study of older adults living with chronic pain and implications for rural healthcare. BMC Health Serv Res 2015 Apr 16;15:162 [FREE Full text] [doi: 10.1186/s12913-015-0825-0] [Medline: 25888988]
49. Zibrik L, Khan S, Bangar N, Stacy E, Novak Lauscher H, Ho K. Patient and community centered eHealth: exploring eHealth barriers and facilitators for chronic disease self-management within British Columbia's immigrant Chinese and Punjabi seniors. Health Policy Technol 2015 Dec;4(4):348-356. [doi: 10.1016/j.hlpt.2015.08.002]
50. Sharkiya SH, Hag AM. Environmental and contextual factors influencing e-health use among older adults: a rapid review. Int J Med Inform 2024 Jul;187:105448. [doi: 10.1016/j.ijmedinf.2024.105448] [Medline: 38615510]
51. DeLange Martinez P, Tancredi D, Pavel M, Garcia L, Young HM. Technology acceptance among low-income Asian American older adults: cross-sectional survey analysis. J Med Internet Res 2024 Nov 22;26:e52498 [FREE Full text] [doi: 10.2196/52498] [Medline: 39576987]
52. Kujala S, Hörhammer I, Väyrynen A, Holmroos M, Nättiaho-Rönnholm M, Hägglund M, et al. Patients' experiences of web-based access to electronic health records in Finland: cross-sectional survey. J Med Internet Res 2022 Jun 06;24(6):e37438 [FREE Full text] [doi: 10.2196/37438] [Medline: 35666563]
53. Valkonen P, Kujala S, Savolainen K, Helminen RR. Exploring older adults' needs for a healthy life and eHealth: qualitative interview study. JMIR Hum Factors 2025 Jan 08;12:e50329 [FREE Full text] [doi: 10.2196/50329] [Medline: 39778194]
54. Safarov N. Personal experiences of digital public services access and use: older migrants' digital choices. Technol Soc 2021 Aug;66:101627. [doi: 10.1016/j.techsoc.2021.101627]
55. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
56. Mielonen J, Kuusisto H, Kinnunen UM, Kemppi A, Saranto K. Older adults' experiences of ehealth in health and social care. FinJeHeW 2023 Sep 29;15(3):637. [doi: 10.23996/fjhw.125122]
57. Population and society. Statistic Finland. URL: https://stat.fi/tup/suoluk/suoluk_vaesto_en.html [accessed 2024-04-29]
58. Anand SS, Samaan Z, Middleton C, Irvine J, Desai D, Schulze KM, South Asian Heart Risk Assessment Investigators. A digital health intervention to lower cardiovascular risk: a randomized clinical trial. JAMA Cardiol 2016 Aug 01;1(5):601-606. [doi: 10.1001/jamacardio.2016.1035] [Medline: 27438754]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
SDT: self-determination theory

Edited by A Mavragani; submitted 12.07.24; peer-reviewed by D Singh, J Brill; comments to author 12.12.24; revised version received 23.01.25; accepted 18.02.25; published 14.03.25.

Please cite as:
Ghorbanian Zolbin M, Kujala S, Huvila I
Experiences and Expectations of Immigrant and Nonimmigrant Older Adults Regarding eHealth Services: Qualitative Interview Study
J Med Internet Res 2025;27:e64249
URL: https://www.jmir.org/2025/1/e64249
doi:10.2196/64249
PMID:

©Maedeh Ghorbanian Zolbin, Sari Kujala, Isto Huvila. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Satisfactory Evaluation of Call Service Using AI After Ureteral Stent Insertion: Randomized Controlled Trial

Ukrae Cho[1,2*], MSc; Yong Nam Gwon[3*], MD; Seung Ryong Chong[4], MBA; Ji Yeon Han[3], AD; Do Kyung Kim[3], MD; Seung Whan Doo[3], MD, PhD; Won Jae Yang[3], MD, PhD; Kyeongmin Kim[5]; Sung Ryul Shim[6], PhD; Jaehun Jung[7*], MD, PhD; Jae Heon Kim[3*], MD, PhD

[1]AI product Biz Team, AI service division, SK Telecom, Seoul, Republic of Korea

[2]College of Business, KAIST, Seoul, Republic of Korea

[3]Department of Urology, Soonchunhyang University Seoul Hospital, Soonchunhyang University Medical College, Seoul, Republic of Korea

[4]Social Safety Net Team, ESG Office, SK Telecom, Seoul, Republic of Korea

[5]Department of Engineering, University of Hong Kong, Hong Kong, China (Hong Kong)

[6]Department of Biomedical Informatics, College of Medicine, Konyang University, Daejeon, Republic of Korea

[7]Department of Preventive Medicine, Korea University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jae Heon Kim, MD, PhD
Department of Urology
Soonchunhyang University Seoul Hospital
Soonchunhyang University Medical College
59 Daesagwan-ro, Yongsan-gu
Seoul, 04401
Republic of Korea
Phone: 82 2 709 9378
Fax: 82 2 710 3190
Email: piacekjh@hanmail.net

Abstract

Background: Ureteral stents, such as double-J stents, have become indispensable in urologic procedures but are associated with complications like hematuria and pain. While the advancement of artificial intelligence (AI) technology has led to its increasing application in the health sector, AI has not been used to provide information on potential complications and to facilitate subsequent measures in the event of such complications.

Objective: This study aimed to assess the effectiveness of an AI-based prediction tool in providing patients with information about potential complications from ureteroscopy and ureteric stent placement and indicating the need for early additional therapy.

Methods: Overall, 28 patients (aged 20-70 years) who underwent ureteral stent insertion for the first time without a history of psychological illness were consecutively included. A “reassurance-call” service was set up to equip patients with details about the procedure and postprocedure care, to monitor for complications and their severity. Patients were randomly allocated into 2 groups, reassurance-call by AI (group 1) and reassurance-call by humans (group 2). The primary outcome was the level of satisfaction with the reassurance-call service itself, measured using a Likert scale. Secondary outcomes included satisfaction with the AI-assisted reassurance-call service, also measured using a Likert scale, and the level of satisfaction (Likert scale and Visual Analogue Scale [VAS]) and anxiety (State-Trait Anxiety Inventory and VAS) related to managing complications for both groups.

Results: Of the 28 recruited patients (14 in each group), 1 patient in group 2 dropped out. Baseline characteristics of patients showed no significant differences (all $P>.05$). Satisfaction with reassurance-call averaged 4.14 (SD 0.66; group 1) and 4.54 (SD 0.52; group 2), with no significant difference between AI and humans ($P=.11$). AI-assisted reassurance-call satisfaction averaged 3.43 (SD 0.94). Satisfaction about the management of complications using the Likert scale averaged 3.79 (SD 0.70) and 4.23 (SD 0.83), respectively, showing no significant difference ($P=.14$), but a significant difference was observed when using the VAS ($P=.01$), with 6.64 (SD 2.13) in group 1 and 8.69 (SD 1.80) in group 2. Anxiety about complications using the State-Trait Anxiety Inventory averaged 36.43 (SD 9.17) and 39.23 (SD 8.51; $P=.33$), while anxiety assessed with VAS averaged 4.86 (SD 2.28) and 3.46 (SD 3.38; $P=.18$), respectively, showing no significant differences. Multiple regression analysis was performed on all

XSL•FO
RenderX

outcomes, and humans showed superior satisfaction than AI in the management of complications. Otherwise, most of the other variables showed no significant differences (*P*.>05).

Conclusions: This is the first study to use AI for patient reassurance regarding complications after ureteric stent placement. The study found that patients were similarly satisfied for reassurance calls conducted by AI or humans. Further research in larger populations is warranted to confirm these findings.

Trial Registration: Clinical Research Information System KCT0008062; https://tinyurl.com/4s8725w2

(J Med Internet Res 2025;27:e56039) doi:10.2196/56039

KEYWORDS

artificial intelligence; AI; ureteral stent; complications; randomized controlled trial; urologic procedures; urology; patients; information resources

Introduction

The initial stents designed for use in the ureter were the double-J stent and single pigtail stent [1-3]. Following their innovation, ureteric stents have become an essential component of urologic procedures; their applications include safeguarding against renal obstruction caused by stone fragment residues, edema, or blood clots; preventing the extravasation of urine; and for analgesia [4,5].

A number of pharmaceutical agents have been used in order to decrease the complications associated with ureteric stents and assessed in randomized controlled trials (RCTs). These include α-blockers, anticholinergic drugs, β-3 agonists, pregabalin, phosphodiesterase-5 inhibitors, antispasmodics, analgesics, anti-inflammatory drugs, and botulinum A toxin [6]. No studies have been carried out in order to predict the likely complications that may occur and to provide patients with this information, to offer individuals receiving stents reassurance, or to ensure that they receive timely follow-up appointments for any additional therapy required.

The concept of artificial intelligence (AI) was introduced during the 1950s, referring to the potential of machines to demonstrate human intelligence traits, such as being able to learn, reason, and solve problems [7]. The application of AI within health systems has now become a reality, owing to the rapid growth of computational resources, diminished data storage expenses, advanced complexity of algorithms, and widely available electronic health data [8-10]. Instruments facilitated by AI have made positive contributions in a number of health contexts to diagnostic work [11] and treatments [12], as well as to the provision of prognostic information, for example, mortality rates [13]. In renal stone pathologies, most researchers have evaluated the use of AI in terms of the precision of the diagnosis or in order to forecast the clinical results of therapy. The use of AI has infrequently been used in order to provide patients with data, before a ureteric stent intervention, regarding the possible complications that may arise and whether they would benefit from prompt adjunctive therapy.

In order to examine the efficacy of an AI-based prediction that could potentially provide patients with reassurance regarding the complications occurring in association with lithotomy and ureteric stent placement performed through ureteroscopy and to indicate the need for early additional therapy following the intervention, a pilot RCT was carried out. This is the initial study to investigate this topic within this clinical discipline.

Methods

Study Population

From October 2022 to December 2023, a total of 28 patients who underwent ureteral stent insertion with ureteroscopic ureterolithotomy were consecutively included in this single-institution study at the Department of Urology, Soonchunhyang University Hospital, Seoul, South Korea.

Patients were approached during their preoperative consultations and informed about the study. Patients who underwent ureteral stent insertion with ureteroscopic ureterolithotomy, patients who received intervention about urinary stone for the first time, patients who can communicate and are aged between 20 and <70 years, patients without history of psychologic illness, and patients without cognitive problems were included. Patients who were judged to be unsuitable for participation in the study, patients with cognitive impairment, and patients who had undergone the same procedure in the past were excluded. Korean version of the Alzheimer Disease-8, a brief informant-based measure designed to distinguish individuals with very mild dementia and mild cognitive impairment from those with normal cognition, was performed on all patients. Patients with cognitive impairment who had not been previously diagnosed and those with a score of 2 or higher were excluded from the study.

Study Procedure

A call service named “reassurance-call” had been established to provide patients with information about the procedure and postprocedure care, to check for the presence and severity of complications, and to guide future treatments in case of complications. The algorithm of the reassurance-call can be found in Multimedia Appendix 1.

The included 28 patients who underwent ureteroscopic ureteral stent insertion were divided into 2 groups of 14 people each [14]. The power was determined to be 14, considering a dropout rate of 10% based on the minimum target number of 12 to confirm significance [15]. The 2 groups were reassurance-call service group by AI (group 1) and reassurance-call service group by humans (group 2). Allocation was performed with random permutation. JHK generated the random allocation sequence, YNG enrolled participants, and JYH randomly assigned patients

XSL•FO
RenderX

into 2 groups. A complete list of patient characteristics can be found in Multimedia Appendix 2.

Operation agreement and cognitive questionnaire were obtained before the surgery. When patients were discharged, α-blockers and anticholinergics were prescribed to relieve irritation of the ureteral stent, and painkillers were prescribed as needed. On the day after discharge and on the 8th day, group 1 or 2 were randomly assigned and provided reassurance-call. Satisfaction with reassurance-call service was assessed on the 14th day after discharge.

"Reassurance-call" is a voice recognition-based AI service that relies on natural language processing technology, rather than a large language model, to reduce hallucinations and conduct surveys accurately. It leverages natural language understanding (NLU) technology to comprehend user intent and uses dialogue management techniques to facilitate smooth communication with patients. The "reassurance-call" service is trained in the following ways: voice recognition training, NLU training, and dialogue management training. First, the "reassurance-call" service improves its ability to recognize patients' voices through voice recognition training. This process involves analyzing patients' voices to update the voice recognition model. In the pilot trial using "reassurance-call," accurate recognition of numbers was crucial due to the use of the Likert scale. SK Telecom, with experience in providing voice-based services like Tmap (navigation; Tmap Mobility) and Btv (media; SK Broadband), brought expertise in number recognition. However, for this pilot, additional training and tuning were conducted using voices from various age groups and genders. Second, the "reassurance-call" service enhances its ability to understand patients' intentions through NLU training. This involves analyzing patients' speech to update the NLU model. During the pilot trial, additional intent lists were defined and trained to accurately handle responses when patients answered with text rather than numbers (eg, if the first option for a question about pain frequency was "often," and the patient responded with "often," "first one," etc, instead of "1"). Third, the "reassurance-call" service improves its ability to manage conversations with patients through dialogue management training. This involves understanding the patient's intent to provide appropriate responses. In the pilot trial, scenarios were additionally defined for situations where patients wanted to revise a previously answered question, and the dialogue algorithm was trained to handle these scenarios smoothly.

Statistics

The primary outcome was satisfaction with reassurance-call service assessed with a 5-point Likert scale. Secondary outcomes were satisfaction with how complications were handled and anxiety about complications. Satisfaction was scored with a 5-point Likert scale and 10-point Visual Analogue Scale (VAS), and anxiety was assessed with the State-Trait Anxiety Inventory (STAI), which is a 20-item self-report measure of anxiety using a 4-point Likert-type for each item, and the VAS.

Continuous variables were represented by mean and SD, and categorical variables were represented by frequency and percentage. We performed chi-square tests and *t* tests to confirm no significant selection bias at baseline between AI (group 1) and humans (group 2). To identify the variables that affected between 2 groups, we first examined each outcome (*reassurance-call by AI satisfaction.Likert*, *Complication.satisfaction.Likert*, *Complication.satisfaction.VAS*, *Complication.anxiety.STAI*, and *Complication.anxiety.VAS*) using *t* tests. Due to the small sample size per group, nonparametric tests (Mann-Whitney *U* test) were also conducted. Multiple regression analysis was then performed to identify the factors associated with satisfaction and anxiety outcomes. The significance of multicollinearity was assessed by comparing the variance inflation factors between independent variables. Statistical analysis was performed using R version 4.2.1 (The R Foundation for Statistical Computing). All statistics were 2-tailed, and *P* values <.05 were considered significant. In addition, the CONSORT-EHEALTH checklist can be found in Multimedia Appendix 3.

Ethical Considerations

This study underwent a thorough ethical review and was approved by the institutional review board of Soonchunhyang University Hospital (IRB No. SCHUH 2022-07-013-001). Written informed consent was obtained from all patients, and the study data were deidentified. Participants received no compensation for their involvement in the study, and no participant images were uploaded.

Results

In total, 28 patients gave informed consent and were enrolled in the study. The study population was randomly assigned into 2 groups. One patient in group 2 was excluded because he refused to answer the reassurance-call after the first call (Figure 1). Baseline characteristics of patients are shown in Table 1. The mean height and weight of group 1 was 167.38 cm and 67.86 kg, while group 2 had a mean height of 166.81 cm and weight of 69.56 kg, respectively. Also, mean age was 55.14 years and 53.64 years, respectively. Among 14 patients in group 1, a total of 10 were male and 4 were female, while 8 were male and 6 were female in group 2. In group 1, the upper ureter stone was the most common (5/14, 36%), and in group 2, the distal ureter stone was the most common location of urinary stone (7/14, 50%). Mean stone size was 5.12 mm in group 1 and 6.26 mm in group 2, but grade-1 hydronephrosis was the most common in both groups (11/14, 79% and 12/14, 86%, respectively). In both groups, procedures were usually finished with a semirigid ureteroscope (9/14, 64% in both groups), and only 1 patient in group 2 underwent PCNL with a nephroscope. All the characteristics were statistically insignificant (all *P*>.05).

Primary and secondary outcomes are presented in Table 2. Satisfaction of reassurance-call showed a mean of 4.14 (SD 0.66) in group 1 and a mean of 4.54 (SD 0.52) in group 2. There was no significant difference between satisfaction for reassurance-call by AI and humans (*P*=.11). Satisfaction for reassurance-call by AI showed a mean of 3.43 (SD 0.94). Satisfaction about the management of complications using the Likert scale scored a mean of 3.79 (SD 0.70) in group 1 and a mean of 4.23 (SD 0.83) in group 2, which was not significantly different (*P*=.14), but it showed a significant difference when using VAS (*P*=.01), with a mean of 6.64 (SD 2.13) in group 1

XSL•FO
RenderX

and mean 8.69 (SD 1.80) in group 2. Anxiety about complications using STAI scored a mean of 36.43 (SD 9.17) in group 1 and a mean of 39.23 (SD 8.51) in group 2, and anxiety assessed with VAS scored a mean of 4.86 (SD 2.28) in group 1 and a mean of 3.46 (SD 3.38) in group 2. There were no significant differences (*P*=.33 and *P*=.18, respectively). The outcomes were presented using box plots in Figure 2.

Multiple regression analysis was performed on all outcomes and presented in Multimedia Appendix 4. Regarding satisfaction with the reassurance-call service using the Likert scale, there was no variable that showed a significant difference (all *P*>.05). However, in the analysis about satisfaction for management of complications using the Likert scale, reassurance-call by AI showed significantly lower satisfaction than reassurance-call by humans (*P*=.03), and satisfaction significantly increased as stone size increased (*P*=.03). Similarly, in the analysis about satisfaction for the management of complications using the VAS, reassurance-call by AI showed significantly lower satisfaction than reassurance-call by humans (*P*=.02). There was no variable that showed a significant difference in the analysis of anxiety for complications (all *P*>.05).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of patients selection. AI: artificial intelligence.

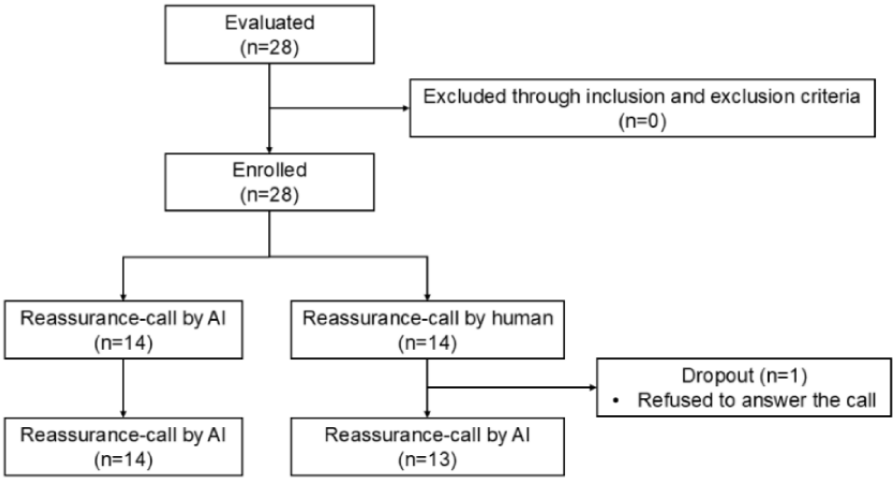

Table 1. Baseline characteristics of the patients.

Characteristics	Artificial intelligence (n=14)	Human (n=14)	*P* value
Height (cm), mean (SD)	167.38 (11.27)	166.81 (8.77)	.88
Weight (kg), mean (SD)	67.86 (13.02)	69.56 (14.61)	.75
Age (years), mean (SD)	55.14 (6.69)	53.64 (12.98)	—[a]
Sex, male, n (%)	10 (71)	8 (57)	.70
Stone site, n (%)			.25
Lower ureter	4 (29)	7 (50)	
Mid ureter	4 (29)	2 (14)	
Upper ureter	5 (36)	3 (21)	
Renal pelvis	1 (7)	2 (14)	
Stone size (mm), mean (SD)	5.12 (1.71)	6.26 (3.55)	.29
Hydronephrosis, n (%)			>.99
Grade 0	2 (14)	1 (7)	
Grade 1	11 (79)	12 (86)	
Grade 2	1 (7)	1 (7)	
Procedure, n (%)			>.99
Flexible ureteroscope	5 (36)	4 (29)	
Semi-rigid ureteroscope	9 (64)	9 (64)	
Nephroscope	0 (0)	1 (7)	

[a]Not applicable.

XSL•FO
RenderX

Table 2. Satisfaction or anxiety after reassurance-call service.

	AI[a]	Human	*P* value (nonparametric)
Reassurance-call satisfaction.Likert, mean (SD)	4.14 (0.66)	4.54 (0.52)	.11
Reassurance-call by AI satisfaction.Likert, mean (SD)	3.43 (0.94)	—[b]	—
Complication.satisfaction.Likert, mean (SD)	3.79 (0.70)	4.23 (0.83)	.14
Complication.satisfaction.VAS[c], mean (SD)	6.64 (2.13)	8.69 (1.80)	.01
Complication.anxiety.STAI[d], mean (SD)	36.43 (9.17)	39.23 (8.51)	.33
Complication.anxiety.VAS, mean (SD)	4.86 (2.28)	3.46 (3.38)	.18

[a]AI: artificial intelligence.

[b]Not applicable.

[c]VAS: Visual Analogue Scale.

[d]STAI: State-Trait Anxiety Inventory.

Figure 2. Box plots for satisfaction or anxiety after the reassurance-call service: (A) reassurance-call by AI satisfaction.Likert, (B) Complication.satisfaction.Likert, (C) Complication.satisfaction.VAS, (D) Complication.anxiety.STAI, and (E) Complication.anxiety.VAS. AI: artificial intelligence; STAI: State-Trait Anxiety Inventory; VAS: Visual Analogue Scale.

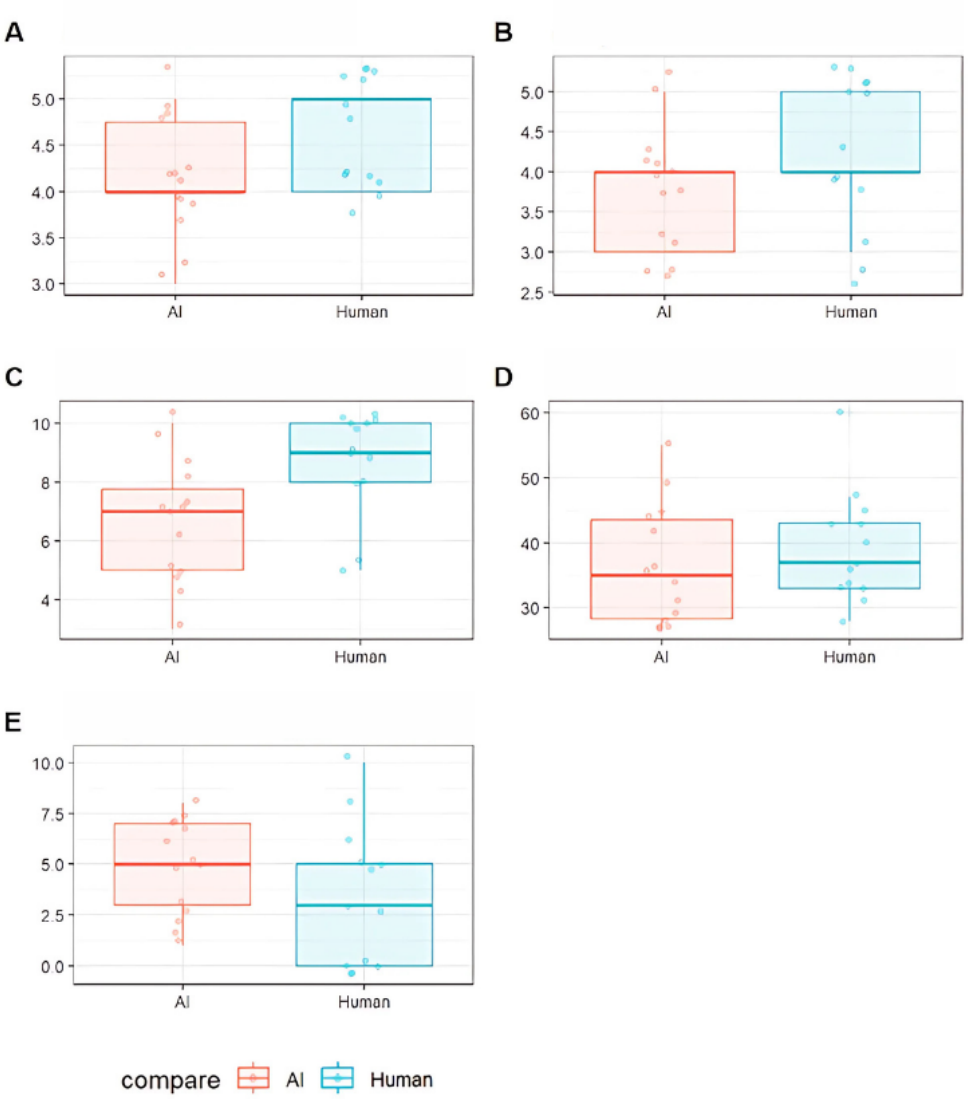

XSL•FO
RenderX

Discussion

Principal Findings

Hypothetically, a reassurance call generated by AI within a specific clinical context may provide patients with some knowledge regarding the expected complications following the insertion of a ureteric stent and therefore make them more satisfied with the outcome of the procedure. As anticipated, in some instances, human monitoring calls provided a higher satisfaction rating, but in general, it was established that safety data provided by an AI-based intervention could be used to substitute information provided by humans. This is the first research to demonstrate this outcome.

In our study, when comparing group 1 (AI) and group 2 (human), patients were more satisfied with calls made by a human than those made by AI. However, group 1 also received a score exceeding 4 out of 5, achieving a sufficiently high score on the Likert scale. Regarding complications, group 2 showed a significantly higher satisfaction score by the VAS than group 1; however, there was no significant difference between the 2 groups by other measurement tools including Likert-scale anxiety scores. This implies the possibility that AI can replace human calls to some extent.

Discomfort and the presence of blood in the urine are the most frequently occurring side effects following ureteric stent insertion, both of which patients find worrisome and may induce anxiety. The evidence base suggests that symptoms arising in this situation should be approached using combination therapy, for example, α-blockers, anticholinergic agents, anti-inflammatory drugs, and opiates. The rationale underlying the symptoms of loin and suprapubic discomfort, and urine-related issues, for example, frequency, urgency, incontinence, or pain, is not entirely clear. It has been proposed that the stent may act as a mechanical irritant, and that reflux may generate raised pressures within the urinary tract. These issues have not been resolved by either stent design modifications or the administration of pharmaceutical agents either during or following stent insertion. It is essential from the patient's perspective to administer the appropriate medications to address likely complications following the procedure, as well as to be able to provide them with reassurance regarding their symptoms and to monitor for more serious adverse events.

Patient care quality can be enhanced by the application of AI, which enables algorithms pertaining to diagnosis and therapy to be generated for use in patients with renal lithiasis. To date, AI has been used to determine the constituents of renal stones with some precision, to support a diagnosis of lithiasis, and to forecast clinical endpoints following operative interventions [16]. There are, however, further aspects of this clinical domain that have the potential to be assisted by AI.

The widespread use of tools based on AI in routine clinical work has been described in a recent publication by Lam et al [17], with applications identified in 13 clinical disciplines. The majority of relevant RCTs were performed in gastroenterology; 15 studies examined the use of AI to support endoscopic procedures.

Instruments incorporating biosignal-based AI were used in a high proportion of RCTs. In a study of patients with malignancy, those individuals offered a de novo smartphone intervention based on AI demonstrated lower pain scores and were less likely to have a negative mindset toward oncotherapy than control subjects [18].

In some cases, operative procedures for renal stones can be carried out as day cases or in an outpatient facility. In this setting, there should be a consensus on how care is offered to patients who develop postoperative issues. Typically, this is provided by making telephone contact with the patient on the day after the procedure in order to monitor their recovery process and to establish whether they are experiencing any complications [19]. The number of patients requiring same-day procedures, however, is increasing, and this method of follow-up is labor-intensive and takes time that could otherwise be used for clinical patient care. The French Society of Anesthesia and Intensive Care is conducting a study (OPERA) throughout France in order to determine the way in which ambulatory interventions are organized. They established that no calls are made to approximately 15% of patients; this figure may be as high as 55% on Saturday mornings [20]. The majority of follow-up calls, that is, 70%, are made by nursing staff; secretaries and doctors carry out 5% and 1%, respectively. These statistics therefore provide an incentive to develop automated systems to review patients. It is essential that patients remain satisfied with the use of AI, as opposed to a human, to assess their safety, as demonstrated in this study. This is a more advanced system than automated text messages and has the ability to converse with patients.

Although our study is the first to show the potential for safe calls using AI, there are some limitations. The first is that, even if AI is introduced, it is not in the form of conversation. A conversation was had with the patient, during which their symptoms were assessed according to a predetermined script, regardless of whether this was carried out by a human or an AI-based tool. Generative AI was not used in this study, and so, as expected, patients felt less satisfied than if they had conversed with a human. Generative AI is not feasible at present in clinical medicine, but despite this potential shortfall, the outcome of the study was extremely promising. A point was made of specifically enquiring about symptoms of major complications, for example, difficulty passing urine, evident hematuria, and marked pyrexia, and to note if these were absent. A further benefit of this study was that patients could be reassured regarding any potential complications that were arising. We checked for complications by directly asking patients about predictable important complications, such as high fever, severe pain, and hematuria, and guided them to return to the hospital in cases of high fever, so that serious urinary tract infections could be treated early.

Conclusions

Through our study, we have confirmed for the first time that AI can replace humans in identifying complications that occur after ureteral stent insertion and can also play a role in

XSL•FO
RenderX

alleviating anxiety. Although this pilot study included only a small number of participants, the outcomes demonstrated similar satisfaction rates among patients regardless of whether they received a follow-up call made by the tool based on AI or a human. Additional prospective research is therefore merited in larger populations in order to substantiate these findings.

Acknowledgments

This work was supported by the Soonchunhyang University Research Fund, the National Research Foundation of Korea grant funded by the Korean government (NRF-2021R1A5A2030333), and a Korea University grant (K2426671).

Data Availability

Datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JJ and JHK contributed to conceptualization. Data curation was handled by YNG, UC, and SRC. Formal analysis was managed by YNG, JYH, and DKK. Funding acquisition was done by JJ and JHK. UC, SRC, JYH, DKK, SWD, WJY, KK, and SRS performed an investigation. The software was handled by UC and SRC. Writing the original draft was done by YNG and JHK; Writing review and editing were handled by JJ and JHK. Authors JJ (eastside1st@gmail.com) and JHK (piacekjh@hanmail.net) are co-corresponding authors for this article. Authors UC and YNG are co-first authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The algorithm of Ansim-call.
[XLSX File (Microsoft Excel File), 35 KB - jmir_v27i1e56039_app1.xlsx]

Multimedia Appendix 2
A complete list of patient characteristics.
[XLSX File (Microsoft Excel File), 32 KB - jmir_v27i1e56039_app2.xlsx]

Multimedia Appendix 3
CONSORT-eHEALTH V1.6.
[PDF File (Adobe PDF File), 1191 KB - jmir_v27i1e56039_app3.pdf]

Multimedia Appendix 4
Regression analysis for identifying risk factors.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v27i1e56039_app4.xlsx]

References

1. Beiko DT, Knudsen BE, Denstedt JD. Advances in ureteral stent design. J Endourol 2003;17(4):195-199. [doi: 10.1089/089277903765444294] [Medline: 12816579]
2. Finney RP. Experience with new double J ureteral catheter stent. J Urol 1978;120(6):678-681. [doi: 10.1016/s0022-5347(17)57326-7] [Medline: 731804]
3. Hepperlen TW, Mardis HK, Kammandel H. Self-retained internal ureteral stents: a new approach. J Urol 1978;119(6):731-734. [doi: 10.1016/s0022-5347(17)57613-2] [Medline: 77917]
4. Dellis A, Joshi HB, Timoney AG, Keeley FX. Relief of stent related symptoms: review of engineering and pharmacological solutions. J Urol 2010;184(4):1267-1272. [doi: 10.1016/j.juro.2010.06.043] [Medline: 20723923]
5. Liu Q, Liao B, Zhang R, Jin T, Zhou L, Luo D, et al. Combination therapy only shows short-term superiority over monotherapy on ureteral stent-related symptoms - outcome from a randomized controlled trial. BMC Urol 2016;16(1):66 [FREE Full text] [doi: 10.1186/s12894-016-0186-y] [Medline: 27846839]
6. Bellos TC, Katsimperis SN, Kapsalos-Dedes SG, Tzelves LI, Kostakopoulos NA, Mitsogiannis IC, et al. Ureteral stent-related symptoms and pharmacotherapy: a brief narrative review. J Clin Pharmacol 2023;63(10):1091-1100. [doi: 10.1002/jcph.2314] [Medline: 37476926]
7. McCarthy J, Minsky M, Rochester N, Shannon CE. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine 2006;27(4):12-14. [doi: 10.1609/aimag.v27i4.1904]

XSL•FO
RenderX

8. Hamamoto R. Application of artificial intelligence for medical research. Biomolecules 2021;11(1):90 [FREE Full text] [doi: 10.3390/biom11010090] [Medline: 33445802]
9. Shung DL, Sung JJY. Challenges of developing artificial intelligence-assisted tools for clinical medicine. J Gastroenterol Hepatol 2021;36(2):295-298. [doi: 10.1111/jgh.15378] [Medline: 33624889]
10. Sung JJ, Stewart CL, Freedman B. Artificial intelligence in health care: preparing for the fifth industrial revolution. Med J Aust 2020;213(6):253-255.e1. [doi: 10.5694/mja2.50755] [Medline: 32892395]
11. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 2018;136(7):803-810 [FREE Full text] [doi: 10.1001/jamaophthalmol.2018.1934] [Medline: 29801159]
12. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
13. Shung DL, Au B, Taylor RA, Tay JK, Laursen SB, Stanley AJ, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. Gastroenterology 2020;158(1):160-167 [FREE Full text] [doi: 10.1053/j.gastro.2019.09.009] [Medline: 31562847]
14. Bioavailability FE. Guidance for industry. Center for Drug Evaluation and Research (CDER). 2002. URL: https://www.fda.gov/files/drugs/published/Bioavailability-and-Bioequivalence-Studies-Submitted-in-NDAs-or-INDs-%E2%80%94-General-Considerations.pdf [accessed 2024-12-25]
15. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics 2005;4(4):287-291. [doi: 10.1002/pst.185]
16. Ganesan V, Pearle MS. Artificial intelligence in stone disease. Curr Opin Urol 2021;31(4):391-396. [doi: 10.1097/MOU.0000000000000896] [Medline: 33965985]
17. Lam TYT, Cheung MFK, Munro YL, Lim KM, Shung D, Sung JJY. Randomized controlled trials of artificial intelligence in clinical practice: systematic review. J Med Internet Res 2022;24(8):e37188 [FREE Full text] [doi: 10.2196/37188] [Medline: 35904087]
18. Kamdar MM, Centi AJ, Fischer N, Jethwani K. A randomized controlled trial of a novel artificial-intelligence based smartphone application to optimize the management of cancer-related pain. JCO 2018;36(34_suppl):76-76. [doi: 10.1200/jco.2018.36.34_suppl.76]
19. Hwa K, Wren SM. Telehealth follow-up in lieu of postoperative clinic visit for ambulatory surgery: results of a pilot program. JAMA Surg 2013;148(9):823-827. [doi: 10.1001/jamasurg.2013.2672] [Medline: 23842982]
20. Beaussier M, Albaladejo P, Sciard D, Jouffroy L, Benhamou D, Ecoffey C, et al. Operation and organisation of ambulatory surgery in France. Results of a nationwide survey; the OPERA study. Anaesth Crit Care Pain Med 2017;36(6):353-357. [doi: 10.1016/j.accpm.2017.07.003] [Medline: 28826981]

Abbreviations

AI: artificial intelligence
NLU: natural language understanding
RCT: randomized controlled trial
STAI: State-Trait Anxiety Inventory
VAS: Visual Analogue Scale

Edited by A Coristine; submitted 08.01.24; peer-reviewed by IGJ Jeong, N Domingues, S Qi, R Kazemi, J Noh, JY Park; comments to author 07.10.24; revised version received 03.11.24; accepted 09.12.24; published 21.01.25.

Please cite as:
Cho U, Gwon YN, Chong SR, Han JY, Kim DK, Doo SW, Yang WJ, Kim K, Shim SR, Jung J, Kim JH
Satisfactory Evaluation of Call Service Using AI After Ureteral Stent Insertion: Randomized Controlled Trial
J Med Internet Res 2025;27:e56039
URL: https://www.jmir.org/2025/1/e56039
doi:10.2196/56039
PMID:

©Ukrae Cho, Yong Nam Gwon, Seung Ryong Chong, Ji Yeon Han, Do Kyung Kim, Seung Whan Doo, Won Jae Yang, Kyeongmin Kim, Sung Ryul Shim, Jaehun Jung, Jae Heon Kim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution

XSL•FO
RenderX

License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of Artificial Intelligence Helpfulness and Uncertainty on Cognitive Interactions with Pharmacists: Randomized Controlled Trial

Chuan-Ching Tsai[1], MHI, PharmD; Jin Yong Kim[2], BEng; Qiyuan Chen[2], MSE; Brigid Rowell[1], MA; X Jessie Yang[2], PhD; Raed Kontar[2], PhD; Megan Whitaker[1], MHI; Corey Lester[1], PhD, PharmD

[1]Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

[2]Department of Industrial and Operations Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Corey Lester, PhD, PharmD
Department of Clinical Pharmacy
College of Pharmacy
University of Michigan
428 Church Street
Ann Arbor, MI, 48109
United States
Phone: 1 7346478849
Email: lesterca@umich.edu

Abstract

Background: Clinical decision support systems leveraging artificial intelligence (AI) are increasingly integrated into health care practices, including pharmacy medication verification. Communicating uncertainty in an AI prediction is viewed as an important mechanism for boosting human collaboration and trust. Yet, little is known about the effects on human cognition as a result of interacting with such types of AI advice.

Objective: This study aimed to evaluate the cognitive interaction patterns of pharmacists during medication product verification when using an AI prototype. Moreover, we examine the impact of AI's assistance, both helpful and unhelpful, and the communication of uncertainty of AI-generated results on pharmacists' cognitive interaction with the prototype.

Methods: In a randomized controlled trial, 30 pharmacists from professional networks each performed 200 medication verification tasks while their eye movements were recorded using an online eye tracker. Participants completed 100 verifications without AI assistance and 100 with AI assistance (either with black box help without uncertainty information or uncertainty-aware help, which displays AI uncertainty). Fixation patterns (first and last areas fixated, number of fixations, fixation duration, and dwell times) were analyzed in relation to AI help type and helpfulness.

Results: Pharmacists shifted 19%-26% of their total fixations to AI-generated regions when these were available, suggesting the integration of AI advice in decision-making. AI assistance did not reduce the number of fixations on fill images, which remained the primary focus area. Unhelpful AI advice led to longer dwell times on reference and fill images, indicating increased cognitive processing. Displaying AI uncertainty led to longer cognitive processing times as measured by dwell times in original images.

Conclusions: Unhelpful AI increases cognitive processing time in the original images. Transparency in AI is needed in "black box" systems, but showing more information can add a cognitive burden. Therefore, the communication of uncertainty should be optimized and integrated into clinical workflows using user-centered design to avoid increasing cognitive load or impeding clinicians' original workflow.

Trial Registration: ClinicalTrials.gov NCT06795477; https://clinicaltrials.gov/study/NCT06795477

(J Med Internet Res 2025;27:e59946) doi:10.2196/59946

XSL•FO
RenderX

KEYWORDS

CDSS; eye-tracking; medication verification; uncertainty visualization; AI helpfulness and accuracy; artificial intelligence; cognitive interactions; clinical decision support system; cognition; pharmacists; medication; interaction; decision-making; cognitive processing

Introduction

Clinical decision support systems (CDSS) are tools that use medical knowledge and health information to aid clinicians' decision-making to provide enhanced patient care [1,2]. CDSS can be classified into 2 types: knowledge-based and non–knowledge-based [2]. In knowledge-based CDSS, relevant information is evaluated by a set of IF-THEN rules, and recommendations are generated. Non–knowledge-based CDSS use artificial intelligence (AI) and machine learning (ML) methods rather than rules to evaluate information and generate recommendations [3]. Various CDSS have been designed to aid pharmacists' clinical work, including checking for drug-drug interactions [4], antibiotics stewardship [5], and drug utilization review [6].

Pharmacists' medication product verification is an example of a pivotal yet time-consuming and vigilant process. During medication product verification, pharmacists compare the contents of a filled medication bottle to professional, close-up images of medication that are integrated into the dispensing platform. Although most pharmacists are checking a physical bottle with the pills inside, the industry is moving toward verification using images of filled medication bottles. Pharmacists spend 30%-48% of their time verifying and dispensing medications [7,8]. Multiplying the time by the mean salary of a pharmacist [9], this equals US $38,823-US $62,117 per pharmacist per year. While some states in the United States have implemented a technician final product verification (tech-check-tech), the cost is still significant as the mean annual wage for a pharmacy technician in 2023 was US $43,330.[10] Furthermore, the tech-check-tech workflow does not eliminate the potential for human error.

Product verification is not only time-consuming and costly, but it also requires pharmacists to be vigilant, which can cause fatigue, cognitive overload, and even result in verification and dispensing errors. Medication dispensing error can be defined as "any deviations of the prescription order," including but not limited to dispensing the wrong dosage, strength, dose form, or pharmaceutical ingredient [11]. The estimated rate of medication dispensing errors is 2.4 in 100 prescriptions [12] in community pharmacies, with the most common types being wrong ingredients, wrong strength, and labeling errors [12,13]. These errors can result in unfavorable therapeutic outcomes, which can have adverse effects on patient safety and outcomes.

Leveraging its capability to process large amounts of data, AI-based CDSS can help pharmacists reduce cognitive load and maintain vigilance. The term vigilance is used to describe the level of maintaining focus and alertness over a prolonged period of time [14]. Human vigilance tends to decrease over time, and multiple theories explain the phenomenon, including the overload theory (reduction of information-processing abilities), underload theory (mindlessness due to repetitive nature), and fatigue [15-17]. Since AI has the power to process large amounts of information, it can reduce humans' cognitive load and maintain their information-processing ability. However, balancing transparency and information load is necessary to avoid cognitive overload. Providing too much information can come at the expense of making the task more cognitively demanding. Marusich et al [18] found increasing information in safety-critical situations decreased users' situational awareness and performance. Similarly, Wu et al [19] examination of high, moderate, and low-complexity user interfaces found significant differences between high- and low-complexity interfaces. Low-complexity interfaces significantly increased users' attention and decreased cognitive load compared with high-complexity interfaces ($P<.05$). Furthermore, cognitive load is not static. Experts and novices at a given task have different levels of cognitive overload; expert users have significantly less cognitive load than novices when using an interface of the same level of complexity [19]. Therefore, it is critical to meet the information needs of users in the design of AI advice.

AI and ML–based CDSS are often "black box" systems, where there is a lack of insight into the AI and ML prediction. As a result, there is an increasing call for greater AI transparency among medical professionals, emphasizing the need to display the uncertainty of AI-generated results [20,21]. While presenting this uncertainty may initially confuse users and require more cognitive effort to make decisions [22], it has significant benefits. Displaying uncertainty helps mitigate issues with AI and ML like overreliance and bias, which fosters better human-AI collaboration and trust [22,23]. This transparency enables users to recognize that AI and ML predictions are not definitive, encouraging them to incorporate their judgment into decision-making. In addition, showing uncertainty serves as a new stimulus that maintains user vigilance, preventing boredom or mindlessness as suggested by the underload theory [17].

AI and ML–based CDSS have the potential to improve medical care if the systems perform well and are appropriately implemented [24]. However, when the AI provides inaccurate, unreliable, and biased results, it can lead users to decision errors due to mechanisms such as automation bias and algorithmic aversion [25]. Automation bias is defined as people's heuristic belief that automation's performance is consistent when the system's performance is not always perfect, which can lead to users accepting incorrect advice or rejecting correct advice [25]. Conversely, algorithmic aversion reflects the tendency of users to not trust the algorithm's performance after seeing it make mistakes, even when its performance is better than humans [26]. In a study of users' assessments of AI and human errors in high-stakes situations (ie, medical and legal), users rated AI errors significantly more harshly and less fair than human errors ($P<.001$) [25]. This result was echoed by Dietvorst et al [26], who studied the effect of AI errors on participants' correctly predicting outcomes. Despite the AI significantly outperforming

XSL•FO
RenderX

the humans, participants were significantly less likely to trust the AI's prediction and their confidence in the AI significantly decreased after seeing an AI error. To prevent such human errors in clinical decision-making with the help of AI, high-performing and reliable systems are pivotal. In this study, we assess the impact of helpful and unhelpful AI on user cognitive workload.

Previously, our group developed an AI prototype aiming to help pharmacists with medication product verification and reduce dispensing errors. Using the participatory design method, researchers conducted focus groups with pharmacists who had medication verification experience to understand the workflow, difficulties, and concerns of the current process, and ideas about incorporating AI to aid this process. Based on pharmacists' feedback, a user-centered AI interface emphasizing clarity and accessibility was developed. The development of the AI help software is detailed in previous research by Zheng et al [27]. The purpose of this paper is to establish how pharmacists cognitively incorporate and use an AI prototype during the medication product verification process by analyzing eye-tracking fixation data.

Methods

Participants and Trial Design

This study was a randomized controlled trial. We recruited 30 pharmacists to participate in a single, remote study visit. Out of 2 listservs for professional pharmacists, the Minnesota Pharmacy Practice-Based Research Network and the University of Michigan College of Pharmacy Preceptor Network were used to recruit participants. The listserv managers sent a recruitment email instructing interested pharmacists to contact the study team to schedule a screening phone call. The study's inclusion criteria were (1) a licensed pharmacist in the United States, (2) at least 18 years old at the time of screening, and (3) access to a laptop or desktop computer with a webcam to complete the experiment. The exclusion criteria were (1) require assistive technology to use the computer, (2) need eyeglasses with more than 1 power (eg, bifocals) to complete the experiment, (3) have uncorrected cataracts, intraocular implants, glaucoma, or permanently dilated pupils, or (4) eye movement or alignment abnormalities. A random number generator created in R (R Core Team) assigned each participant to receive black box or uncertainty-aware help. The probability of being assigned to black box help trials or uncertainty-aware help trials was equal for all participants as was the probability of completing the first 100 trials with AI help or without AI help. In each trial, participants would check the images on the screen and decide whether it was a good fill (the medication in the fill image is identical to the medication in the reference image) or a bad fill (the medication in the fill image and the reference image are different) and hit either the reject or the accept button.

To ensure confidence in the quality of the remote data collection, all participants met online with a researcher before performing any study activities. The purpose of the online meeting was to ensure the study setup was consistent (eg, adequate lighting and no glare) and participants had the required materials (ie, ruler or tape measure, and standard ID) to conform with Labvanced's eye-tracking and calibration requirements. After meeting with the researcher, all participants watched an orientation video which provided a detailed overview of the study including an explanation of how the AI software makes its predictions and an introduction on how to use the interface to perform the verifications.

The Task: Pharmacists' Medication Verification

Overview

Pharmacists are responsible for dispensing the correct medication to patients to ensure optimal therapeutic outcomes. Before dispensing medications to patients, pharmacists must visually inspect the medication being filled and compare them to the prescribed drugs. This process is called medication product verification. In the study, each participant performed 200 medication product verification tasks, where their eye movement was recorded using an eye tracker. Each participant conducted 100 medication verifications without AI assistance and 100 verifications according to their assigned AI help group. A screenshot of the participants' view is in Figure 1. To avoid the carry-over effect, the medications being verified in no AI help trials were different from those with AI help. However, the medications were identical in trials with black box AI help and trials with uncertainty-aware AI help.

Figure 1. The medication verification system interface with AI Help. In trials without AI assistance, pharmacists only have access to the reference image (refImage) and fill image. With black box AI help, pharmacists saw an AI match plot, which shows the AI's stance on the match status for 4 key medication characteristics (imprint, color, shape, and score). In uncertainty-aware AI help, in addition to the match plot, pharmacists also saw the AI histogram, which displays the probability distribution from 50 predictions, indicating the uncertainty of the prediction. Examples of high certainty (left-hand image) and low certainty (right-hand image) are shown. AI: artificial intelligence.

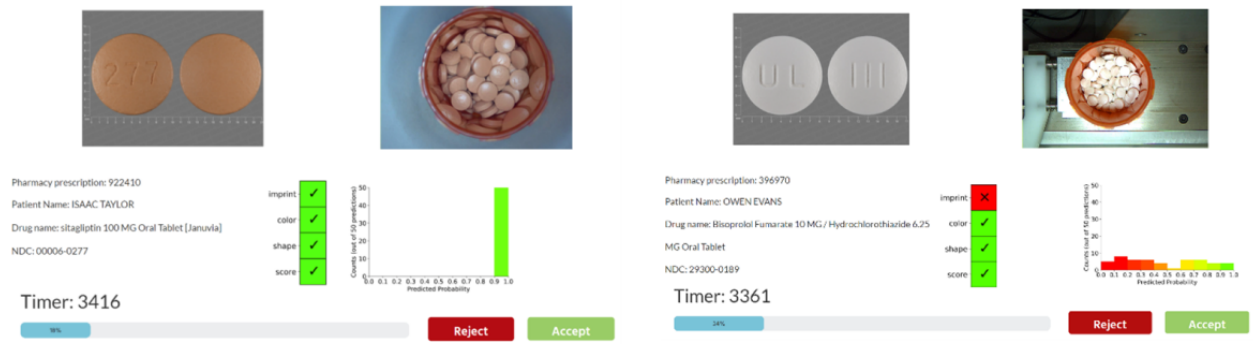

XSL•FO
RenderX

Trial Type

Each trial was configured with specific variables, including type of AI assistance, case type, and participant performance. There were 3 categories of AI assistance (1) no help, (2) black box help, and (3) uncertainty-aware help. In all 3 categories, 76 of the 100 (76%) trials displayed correctly filled medications. In the no-help condition, participants used only the reference image and filled the image and text. Black box help added an AI-generated match plot, while uncertainty-aware help also included an AI-generated probability histogram. The case type, relevant only when AI assistance was provided, referred to the alignment between the AI's recommendation and the correct answer. The AI accurately recommended the correct action (accept or reject) 79% of the time. There were 2 scenarios: (1) helpful advice, where the AI's guidance matched the correct action (accept or reject), and (2) unhelpful advice, where the AI's suggestion opposed the correct answer.

The AI Help Software

We previously developed an automated tool using a Bayesian neural network [28] to predict the National Drug Code (NDC), color, and shape of medications using images of pills in a prescription bottle [29]. In this study, we modified the Bayesian model [30] to predict the dispensed pills' NDC and the associated model's uncertainty of the predictions. This was accomplished by applying the random dropout technique [30] to the ResNet-34 [31] convolutional neural network. The dropout technique is realized by a layer in the neural network called the dropout layer, which activates connections in the layer with certain probabilities. As such, the output of the layer is random every time it predicts, which we use to measure the uncertainty of the prediction. In our research, the model generated 50 potential probabilities for every image. Because the distribution of predicted probabilities could not be written mathematically in close forms, it was approximated using samples drawn from the distribution.

A sample size of 50 was used to balance between computational cost and goodness of approximation. The AI software then compares the predicted NDC and the reference image NDC to generate insights on the match status. The final interface had 2 AI-generated regions that were designed to aid pharmacists' medication verification the AI match plot and AI histogram. The AI match plot showed the match status of 4 characteristics between the AI-predicted NDC and the expected NDC on the prescription which were (1) imprint, (2) color, (3) shape, and (4) score. The unmatched characteristics were denoted by a red "X," while matched ones by a green check. Figure 1 shows a prediction with 4 green checks.

The AI histogram displayed the probability distribution from 50 AI predictions, each representing the AI's estimated likelihood that its predicted NDC matches the expected NDC. These probabilities form the histogram, illustrating the AI software "confidence" in the match status. A green peak in the histogram indicates consistent results across the 50 simulations, suggesting high confidence. Conversely, a flat, colorful distribution signifies low confidence. In Figure 1, the left-hand image shows a highly confident prediction and the right-hand image shows a low confidence prediction.

A crucial component of this study was to examine how pharmacists interact with AI-generated errors. The model's overall accuracy is 98.46%. For the purpose of this study, errors were oversampled to obtain a sufficient number of mistakes which lowered the model's accuracy to 79% for the experiment. At 79% accuracy, our model is above the 0.70 reliability threshold wherein imperfect automation is preferable to no automation [32]. Lowering the accuracy balanced the number of trials and participants' time commitment to the study.

Eye-Tracking Data

Eye-tracking data were collected using software from Labvanced (Paderborn). Labvanced uses deep learning models to process webcam videos, allowing online eye movement tracking. The accuracy and precision are comparable to in-laboratory "gold standard" eye trackers and are verified in a peer-reviewed paper [33]. Participants logged on to the Labvanced Trials website. First, they verified the webcam accurately captured their face and agreed to the recording and data use policy before starting the trial. Participants were then prompted to complete a demographics questionnaire that included questions such as age, gender, race, pharmacy practice experience level, practice setting, and trust level in automated systems.

The first phase of the trial was the calibration. During calibration, participants were first asked to measure the distance of their face to the screen and set the center pose that worked best. The actual calibration included 2 parts: (1) the position and orientation and (2) fixation (ie, focused gaze) to certain points on the screen. Participants were asked to follow and fixate on 2 series of red dots on the screen while maintaining their faces at the center pose. After the calibration was completed, participants started the medication verification trials. The system continuously monitored participants' facial position and orientation. If participants moved out of the center pose, they were prompted to re-align their face to the center pose and rerun the fixation calibration. The raw eye-tracking data were processed by the Labvanced algorithm [34] to fixation data. Our analyses were based on the fixation data and excluded eye movement known as saccades, where participants were simply transitioning from 1 fixation to the next and not necessarily fixating in the region.

Data Analysis and Preprocessing

Fixation analysis is the most common metric for analyzing eye-tracking data. The fixation rate, fixation duration, and dwell time have been proposed to reflect human cognitive interest in certain areas of interest [35-37]. Higher fixation rates and longer dwell times indicate repeated interest in a certain area. The longer fixation duration indicates a higher cognitive load. To characterize the use pattern, we were also interested in the order in which pharmacists use these images. Thus, we report the first fixation region, the last fixation region, the number of fixations in each region, the average fixation duration of each fixation, and the dwell time, which was calculated by the sum of the duration of all fixations in each region in each trial. We further stratified the data based on different AI help types (ie, black box versus uncertainty-aware help), and case types (helpful versus unhelpful advice).

XSL•FO
RenderX

To eliminate calibration failure and measurement error, we calculated the modified *z* score of each region's dwell time in each trial. The modified *z* score method is more robust in detecting outliers since it uses median instead of mean in the calculation of *z* scores. If the modified *z* score of the dwell time in that region is greater than 3.5, we labeled that observation as an outlier and excluded it from our analyses.

For all help type pairs, 2-tailed *z* tests were run for the first and last fixation area to determine if differences among categories were statistically significant (Figure 2). *P* values were also calculated to show if the proportions being compared were statistically significant. Statistical significance between help types was calculated for fixation duration, number of durations, time of fixations per area, and total dwell time by case type using the Mann-Whitney *U* test with the *P* values adjusted using the Bonferroni correction.

Figure 2. Percentage of trials by region of interest and AI help type. Left: first fixation; right: last fixation. *P* values were calculated for the 2-tailed z tests to determine the significance of the observed differences and are displayed above respective brackets. AI histogram was only present in the uncertainty-aware condition. No comparative statistics exist in this case. AI: artificial intelligence.

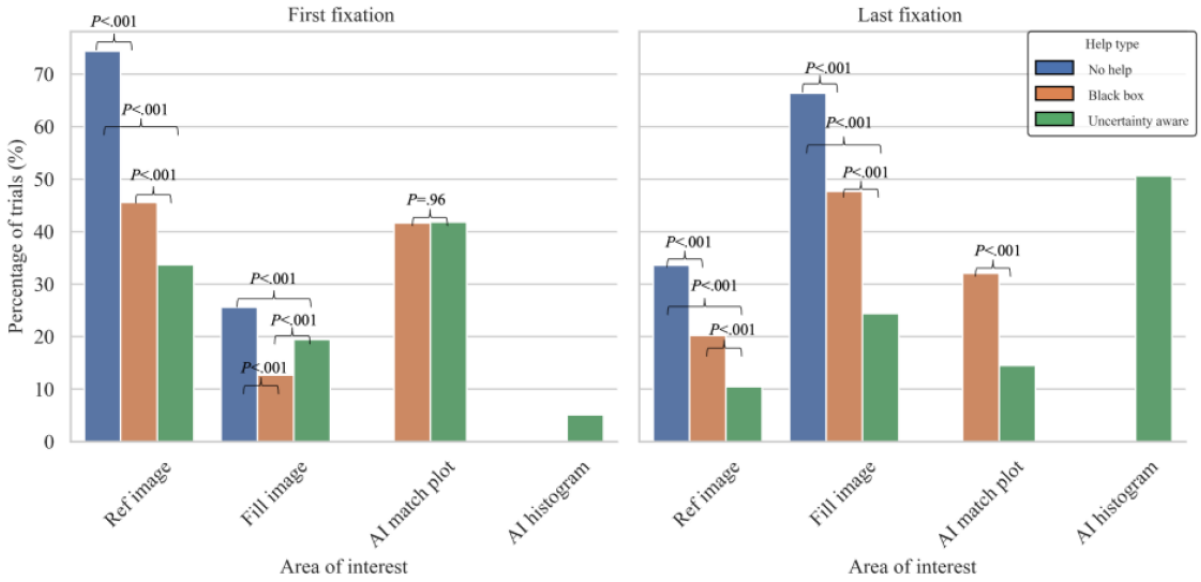

Ethical Considerations

This study did not meet the National Institutes of Health's definition of a clinical trial and the University of Michigan's Institutional Review Board determined this research is exempt from institutional review board oversight under federal Exemption 3(i)(A) and/or 3(i)(B) at 45 CFR 46.104(d).

Results

Overview

In total, 31 participants were randomized to receive black box or uncertainty-aware AI help. One participant experienced technical difficulties and was unable to complete any trials. Fifteen subjects in each condition completed the experiment and were included in the final analysis. Demographic variables of age, gender, race, practice settings, and working years were well-balanced (*P*>.05). Full details of demographic variables are reported elsewhere [38]. After outlier removal, the dataset comprises 2449 trials without AI assistance, 1365 with black box AI help, and 1391 with uncertainty-aware AI help.

The effects of AI assistance on pharmacists' decision-making and reaction time are reported in full elsewhere [38]. In summary, performance and reaction times varied by AI type and AI accuracy. In general, uncertainty-aware AI led to faster decision-making and protected against bad AI advice to approve a misfiled medication whereas black box AI increased reaction times and bad AI advice resulted in lower accuracy.

First and Last Fixation Area

Overview

In trials without AI help, the first fixation area (Figure 2, left) was either (1) the reference image (1822/2449, 74.4%) or (2) the fill image (627/2449, 25.6%). In the black box help condition, the first fixation area was either (1) the reference image (622/1365, 45.6%), (2) the fill image (174/1365, 12.7%), or (3) the AI match plot (569/1365, 41.7%). In uncertainty-aware help, it was (1) the reference image (469/1391, 33.7%), (2) the fill image (270/1391, 19.4%), (3) the AI match plot (581/1391, 41.8%), and (4) the AI histogram (71/1391, 5.1%). For trials without AI help, the last fixation area (Figure 2, right) was either (1) the reference image (822/2449, 33.6%) or (2) the fill image (1627/2449, 66.4%). In the black box help condition, it was (1) the reference image (276/1365, 20.2%), (2) the fill image (651/1365, 47.7%), and (3) the AI match plot (438/1365, 32.1%). In uncertainty-aware help, it was (1) the reference image (146/1391, 10.5%), (2) the fill image (339/1391, 24.4%), (3) the AI match plot (202/1391, 14.5%), and (4) the AI histogram (704/1391, 50.6%). All of these comparisons, with one exception, showed statistically significant differences between help types. The *P* values were calculated to be *P*<.05 (*z* score>1.96 or <–1.96), thus rejecting the null hypothesis and suggesting there is a significant difference in proportions

XSL•FO
RenderX

between each 2. The one exception to this was the pair for first fixation being AI match plot with black box help versus uncertainty-aware help, where the *z* test was –.45 and the *P* value was .97, suggesting that these values were not statistically significantly different.

Number of Fixations

Irrespective of help type, fill images had the highest number of fixations in a trial, and the number of fixations per trial did not decrease after the introduction of AI-generated areas (number of fixations: mean [SD], median [IQR]): (1) no help, mean 4.44 (SD 2.92), median 4.0 (IQR 2.0-6.0); (2) black box help, mean 4.33 (SD 2.90), median 4.0 (IQR 2.0-6.0); and (3) uncertainty-aware help, mean 5.14 (SD 3.43), median 4.0 (IQR 3.0-7.0). Reference images received fewer fixations, with the averages and range being (1) no help: mean 2.47 (SD 1.4), median 2.0 (IQR 1.0-3.0), (2) simple help mean 2.38 (SD 1.4), median 2.0 (IQR 1.0-3.0), and (3) advanced help mean 2.45 (SD 1.4), median 2.0 (IQR 1.0-3.0). The AI match plot and AI histogram received a lower number of fixations. The mean (SD) and median (IQR) for the number of fixations in AI match plots in black box and uncertainty-aware help trials are mean 1.91 (SD 1.12), median 2.0 (IRQ 1.0-2.0), and mean 1.65 (SD 0.82), median 1.0 (IQR 1.0-2.0), and in AI histogram in uncertainty-aware help trials, they are mean 1.69 (SD 0.87), median 1.0 (IQR 1.0-2.0), respectively. The numbers are visualized in a box plot in Figure 3. Bonferroni adjusted *P* values comparing no help with black box for fill image and all 3 comparisons for reference image were observed to be greater than .05. Therefore, we do not reject the null hypothesis of the groups being statistically similar. The comparisons of no help and uncertainty-aware and black box with uncertainty-aware advice for fill image, and black box with uncertainty-aware for AI match plot showed significance (Mann-Whitney U=1088528.5, 507726.0, and 381013.0, respectively; all P<.05).

Figure 3. Boxplot representation of the number of fixations in each area per trial, categorized by help type. Bonferroni adjusted *P* values were calculated alongside Mann-Whitney *U* tests, and these statistics are displayed above respective brackets. AI histogram was only present in the uncertainty-aware condition. No comparative statistics exist in this case. AI: artificial intelligence.

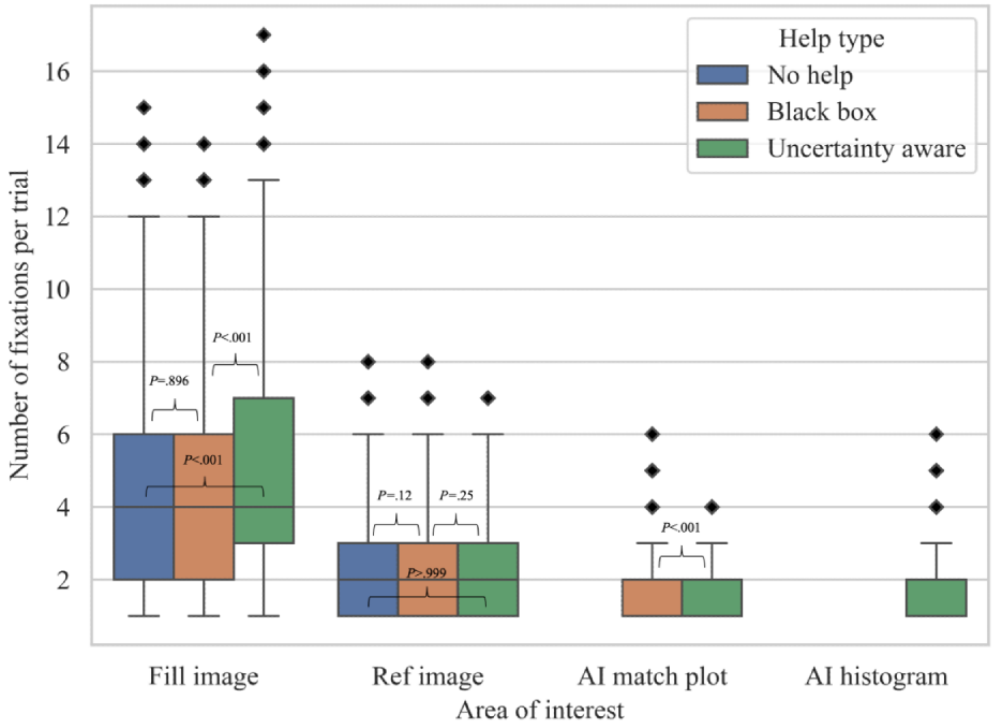

Fixation Duration

The fixation duration in each region was similar, regardless of help type. The median duration of fixation was around 0.22 seconds, with an IQR of approximately 0.75 seconds. This indicates that the cognitive load for processing information in each area was similar. The boxplot of the finding is in Figure 4. One adjusted *P* value was below .05, with a *U* value of 29067652, comparing no help to uncertainty-aware advice for the fill image. This indicates a meaningful statistical difference between these 2 groups.

XSL•FO
RenderX

Figure 4. Boxplot representation of fixation duration in each area, categorized by help type. Bonferroni adjusted *P* values were calculated alongside Mann-Whitney *U* tests, with a Bonferroni correction, and these statistics are displayed above respective brackets. AI histogram was only present in the uncertainty-aware condition. No comparative statistics exist in this case. AI: artificial intelligence.

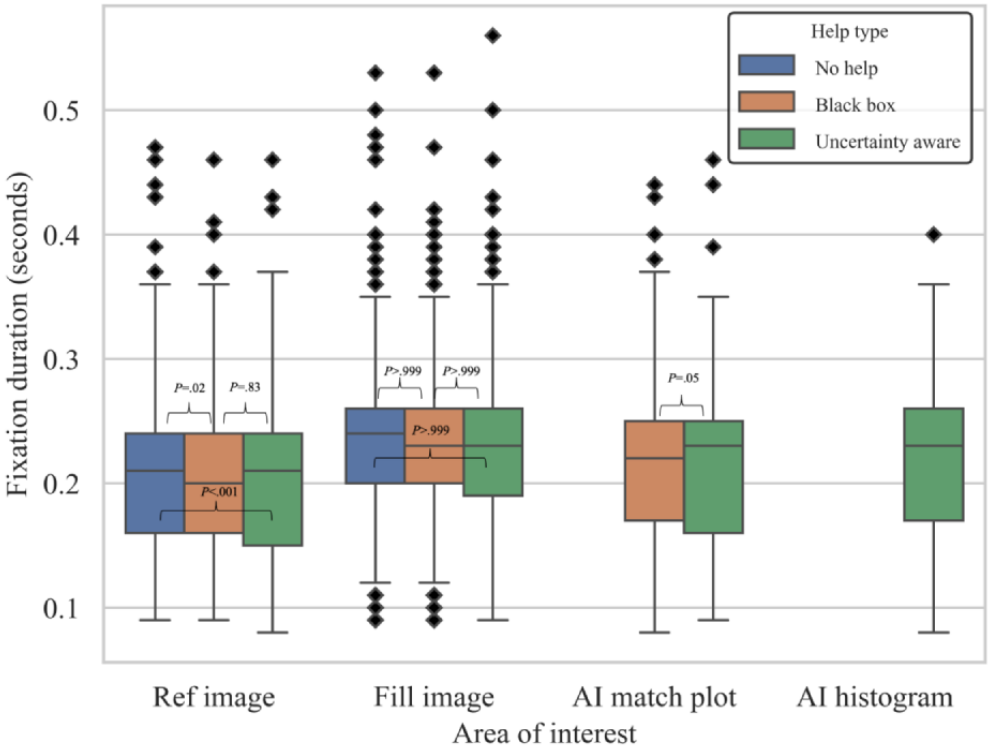

Dwell Times

Participants consistently spent the most time on fill images. The median (IQR) fill image dwell times (in seconds) were: (1) no help trials, median 0.80 (IQR 0.48-1.36), (2) black box help, median 0.80 (IQR 0.47-1.26), and (3) uncertainty-aware help, median 0.92 (IQR 0.53-1.54). For the reference image, the median (IQR) dwell times (in seconds) were (1) no-help, median, 0.44 (0.26-0.66), (2) black box help, median 0.42 (0.24-0.64), and (3) uncertainty-aware help, median 0.43 (IQR 0.26-0.65). The AI match plots had median (IQR) dwell times of 0.33 (IQR 0.22-0.51) and 0.27 (IQR 0.22-0.47) seconds in black box and uncertainty-aware help trials, respectively. The AI histograms had a median (IQR) dwell time of 0.27 (IQR 0.23-0.48) seconds in uncertainty-aware help trials (Figure 5).

The Mann-Whitney *U* test result showed no significant difference between the dwell time in reference images. There was a significant difference in dwell times for fill images with uncertainty-aware help trials compared with no-help trials (*U* value=1113203.0, Bonferroni adjusted *P*<.05) and black box help trials (*U* value=517551.5, Bonferroni adjusted *P*<.05); and for AI match plot from black box to uncertainty-aware (*U* value=375591.0, Bonferroni adjusted *P*<.05). The difference in fill image dwell times was nonsignificant between no help and black box help (Figure 5). Dwell times in fill images and reference images were significantly longer in trials with unhelpful advice compared with helpful advice (the fill image, median: 1.30 vs 0.81 seconds, *U* value 238928.0, *P*<.05; ref image, median: 0.54 vs 0.41 seconds, *U* value 264560.5, *P*<.05; Figure 6).

XSL•FO
RenderX

Figure 5. Boxplot representation of dwell time in each area, categorized by help type. Bonferroni adjusted *P* values were calculated alongside Mann-Whitney *U* tests, with a Bonferroni correction, and these statistics are displayed above respective brackets. AI histogram was only present in the uncertainty-aware condition. No comparative statistics exist in this case. AI: artificial intelligence.

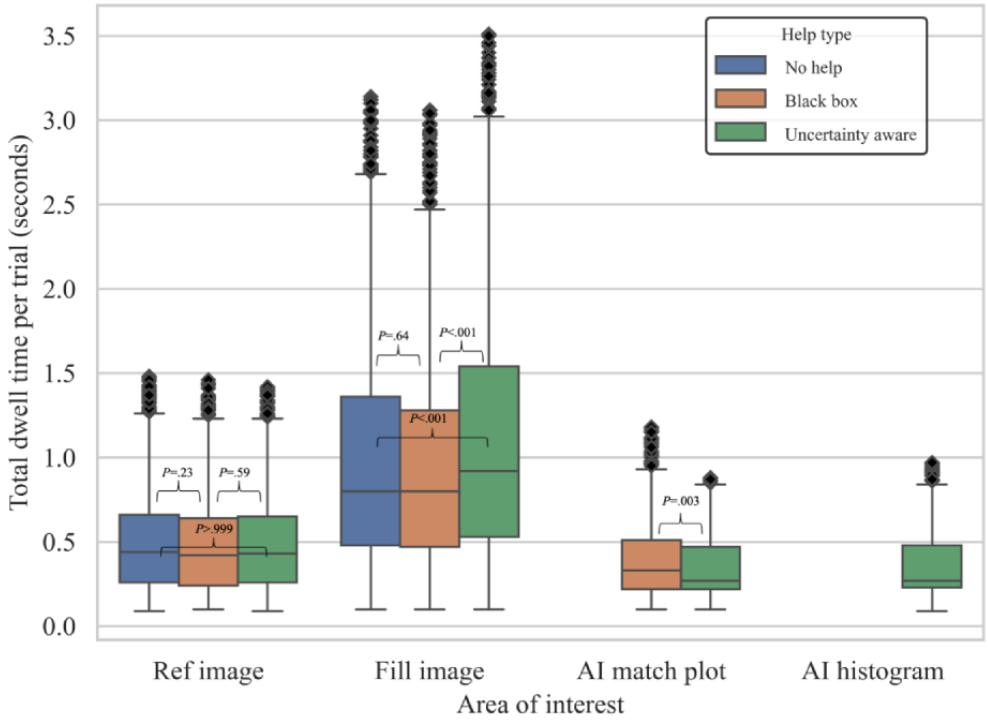

Figure 6. Dwell time in each area by case type. Compared with helpful advice trials, those with unhelpful advice resulted in significantly longer dwell times in fill and reference images. Bonferroni adjusted *P* values were calculated alongside Mann-Whitney *U* tests and these statistics are displayed above respective brackets. AI: artificial intelligence.

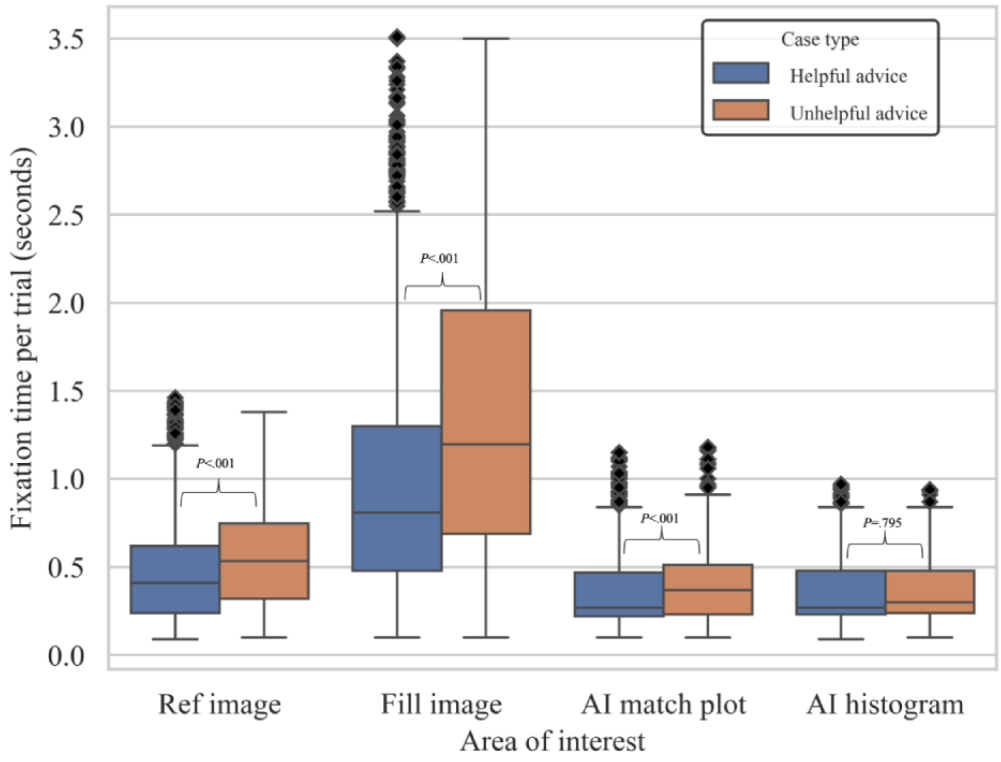

XSL•FO
RenderX

Discussion

Principal Results

Our study presents critical insights into the interaction between pharmacists and an AI-based CDSS. Our findings highlight a shift in the visual and cognitive engagement of pharmacists when AI-driven tools are used. Although communicating the uncertainty of AI has been shown to have benefits such as preventing overreliance and fostering trust in AI [20,22], our study found that it resulted in a longer cognitive processing time of the core task image (ie, filled medication images), though primarily in situations where the AI provided unhelpful advice. The criticality of AI accuracy is underscored by our findings, as AI advice of varying correctness significantly impacts the cognitive processing time of users. These findings suggest the potential of human-AI collaboration to enhance health care delivery if AI is accurate, reliable, and deployed properly without interfering with health care professionals' existing workflow or increasing their cognitive load.

Irrespective of AI intervention, participants allocated the majority of their fixation time to fill images. The fact that pharmacists need to inspect pills from various angles to verify color, shape, and imprint for accurate identification may give rise to longer cognitive processing times. On the other hand, AI-generated regions have shorter dwell times. This might stem from the graphical simplicity of these images [39] or indicate that users correctly perceive the AI-generated regions as supportive in completing the task. Compared with not showing AI outputs' uncertainty (ie, black box AI help), we found that displaying uncertainty results in significantly longer dwell times in the original (fill) images, especially when the AI advice was unhelpful. This may suggest that users are confused or second-guess the AI's correctness when they see the uncertainty histogram, leading to a need to go back to the fill images and verify the correctness of the advice. The shorter dwell times observed with black box AI advice may be attributed to overreliance on automation, resulting in missed errors in the filled medication.

AI-based CDSS influenced participants' drug verification processes. When participants had access to AI, 19% to 26% of total fixations were shifted to AI regions, and there was a decrease in fixation in the original (fill and reference) images. This indicates that the AI tool changed how participants process information, and participants incorporate AI-generated information in their decision-making process. The decrease in fixation counts in original images and the shift to AI-generated regions indicate that AI may support users in completing the task. When AI help is available, users tend to look at the AI match plot first, and then move on to the fill and reference images. This suggests that providing a simple graphical "summary" of the AI advice may aid users in their decision-making and help them maintain vigilance.

Comparison With Previous Work

Our finding demonstrates the need for well-designed AI that balances transparency and information load. Since many AI algorithms are essentially a "black box," there is a demand for AI transparency [23,40]. However, displaying more information about the advice can decrease user vigilance since more information needs to be processed. In designing AI-based systems, there is a need to balance the transparency and explainability of AI output and the amount of information presented to the users. Previous research has shown that high-complexity interfaces significantly increase cognitive workload compared with low-complexity interfaces [19]. Together, our findings and the previous research support the use of our graphically simple, black box AI help over to the histogram which may be overloading the users.

According to Prabhudesai et al [22], users develop their expertise in explaining the uncertainty plot over time, which resolves the initial confusion. In addition to users developing an explanation over time, the user's expertise with a task impacts the level of cognitive load. Wu et al [19] found at the same level of information load, experts at a task experience significantly less cognitive load than novices. Too much complexity, such as an uncertainty histogram, might not be suitable for the relatively simple medication verification task especially since pharmacists are often tasked with additional responsibilities, which interfere with their vigilance in performing the verification task [41]. Simplifying the communication of the uncertainty might balance the benefits of this kind of advice from AI.

The results revealed that AI's correctness (ie, helpfulness) had a significant impact on users' cognitive processing time. Helpful advice resulted in significantly shorter dwell times in reference images and fill images compared with unhelpful advice. This finding supports that users are susceptible to automation bias, the heuristic tendency of users to favor the suggestion made by automated systems [25], and underlines the importance of AI's correctness and performance. Developers should validate AI's accuracy and reliability before implementing it in the real world, for the underperformance of AI can introduce new user errors through mechanisms such as automation bias, slowing users down. An important way to do this is through the development of standards and guidelines, as well as assessing accuracy, reliability, and other performance metrics [40]. Park et al [42], introduced a framework for AI-based medical products, emphasizing phases akin to drug trials which are (1) balancing benefits and risks, (2) confirming usability through methods like AB testing, (3) conducting large-scale trials to validate effectiveness, and (4) ongoing monitoring for self-improvement. Like medications, AI-driven CDSS can profoundly impact patient care, necessitating rigorous pre- and post-deployment validation.

How AI and CDSS can be used in medication dispensing is under-researched. A systematic review of CDSS and medication use found that while CDSS is used in the prescribing, administrating, and monitoring phases of medication use, CDSS has not been deployed for medication dispensing [43]. To the best of our knowledge, our innovative study is the first to examine how a CDSS impacts cognitive workload for medication verification and dispensing. Our study adds important insights into supporting pharmacists who are experiencing high levels of burnout. A nationwide survey of community pharmacists found that 74.9% of pharmacists have experienced burnout [44]. Furthermore, high prescription

XSL•FO
RenderX

volumes and inadequate pharmacist coverage are contributing factors to dispensing errors [45]. This is concerning as nearly a third (32.1%) of community pharmacists work alone [46], and community pharmacists spend nearly half their work hours (48%) verifying prescriptions [47]. Wash et al [48] narrative review of pharmacists' well-being and burnout found a lack of research into interventions to assist pharmacists with burnout. AI tools such as ours offer a potential solution to alleviate burnout and prevent dispensing errors by providing a second opinion and decreasing cognitive workload.

The use of user-centered design is becoming increasingly prominent in CDSS development [49-53]. A 2024 review of CDSS highlighted the importance of user-centered design to foster the trustworthiness, usability, and acceptability of CDSS tools by health care providers [54]. CDSS tools should fit into current workflows and address health care providers' needs and preferences [54]. During the user-centered design phase of developing our AI prototype, pharmacists emphasized the need for simple and accessible information to guide them in completing the task [27]. To avoid overwhelming pharmacists with redundant information, the match plot only displays AI's stance on color, shape, imprint, and score, which pharmacists identified as the most important information during the medication verification process [27]. Our study adds to the literature a pharmacy-based user-centered design use case underscoring the importance of incorporating user feedback and preferences into the design process. Developers of CDSS should focus on creating systems that align closely with user needs and preferences which can enhance usability and effectiveness. By prioritizing user-centered design principles, developers can improve the performance of AI-based systems and ensure that these systems effectively augment rather than disrupt user workflows.

Limitations and Future Work

Several other eye-tracking metrics have been proposed to assess cognitive load. Voluntary movements such as saccade length and saccade velocity, and involuntary movements such as pupil dilation can reflect cognitive load. Due to the lack of data availability, we did not include these metrics in our analysis. Future studies can include these eye-tracking metrics to further understand users' cognitive workload.

This study showcases the need for well-designed AI that balances transparency and information load. Future research should focus on conducting user-center design and usability testing before implementation to understand the benefits and drawbacks of using the AI tool. Also, qualitative methods, such as interviews, and quantitative methods, such as surveys, can incorporated throughout the design and testing phases to allow for design revision based on participants' perceptions of the AI tools. This may help inform the design and implementation of AI tools in augmenting health care professionals' workflow. As Wu et al [19] described, the interaction between a user's level of expertise and the complexity of an AI tool has a significant impact on the user's cognitive load. Future research should consider allowing users to customize the AI display based on their expertise with the task and comfort with the AI tool. Future studies can include qualitative and quantitative methods to further understand user experiences.

Pharmacists routinely perform a variety of tasks concurrently with the medication verification including checking for drug-drug interactions, confirming the medication count is correct, and verifying patient directions are accurate, complete, and written in patient-friendly language. Pharmacy-specific future research should explore new designs to support dispensing tasks and reduce cognitive load.

Conclusions

The goal of this study was to assess how pharmacists cognitively incorporate and use an AI prototype during the medication product verification process, using eye-tracking data to analyze their interaction with the system. The study revealed that AI-based CDSS can alter traditional workflow and cognitive engagement, with pharmacists allocating significant focus to AI-generated regions, which indicates the integration of AI advice in decision-making. However, communicating AI uncertainty by displaying a probability histogram increased cognitive processing time for original images. The correctness of AI suggestions directly affects cognitive processing, with helpful AI advice reducing and unhelpful AI advice increasing it. The findings underscore the importance of accurate, reliable AI in health care and suggest that user-centered design and AI transparency are crucial for effective human-AI collaboration. Future research should assess additional eye-tracking metrics to determine users' cognitive load, incorporate user-centered design to inform the CDSS design, use qualitative and quantitative methods throughout the design and testing phases to better understand users' experiences, and focus on assisting pharmacists with the varied tasks involved in medication dispensing.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was only retrospectively registered, explained by authors as follows: this benign behavioral intervention using mock data did not meet the NIH definition of a clinical trial and the University of Michigan Institutional Review Board determined it was exempt from ongoing IRB review, under federal EXEMPTION 3(i)(A) and/or 3(i)(B) at 45 CFR 46.104(d). The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because

XSL•FO
RenderX

the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 851 KB - jmir_v27i1e59946_app1.pdf]

References

1. Moon J, Chladek JS, Wilson P, Chui MA. Clinical decision support systems in community pharmacies: a scoping review. J Am Med Inform Assoc 2023;31(1):231-239. [doi: 10.1093/jamia/ocad208] [Medline: 37875066]
2. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
3. Berner ES. Clinical decision support systems: theory and practice. New York, NY: Heidelberg; 2007.
4. Smithburger PL, Buckley MS, Bejian S, Burenheide K, Kane-Gill SL. A critical evaluation of clinical decision support for the detection of drug-drug interactions. Expert Opin Drug Saf 2011;10(6):871-882. [doi: 10.1517/14740338.2011.583916] [Medline: 21542665]
5. Madaras-Kelly KJ, Hannah EL, Bateman K, Samore MH. Experience with a clinical decision support system in community pharmacies to recommend narrow-spectrum antimicrobials, nonantimicrobial prescriptions, and OTC products to decrease broad-spectrum antimicrobial use. J Manag Care Pharm 2006;12(5):390-397 [FREE Full text] [doi: 10.18553/jmcp.2006.12.5.390] [Medline: 16792446]
6. Armstrong EP, Markson TJ. On-line DUR messages: pharmacists' attitudes and actions in response. J Am Pharm Assoc (Wash) 1997;NS37(3):315-320. [doi: 10.1016/s1086-5802(16)30215-7] [Medline: 9170808]
7. Karia A, Norman R, Robinson S, Lehnbom E, Laba TL, Durakovic I, et al. Pharmacist's time spent: space for pharmacy-based interventions and consultation time (SPICE)-an observational time and motion study. BMJ Open 2022;12(3):e055597 [FREE Full text] [doi: 10.1136/bmjopen-2021-055597] [Medline: 35236731]
8. Hohmeier KC, Wheeler J, Heintz K, Gatwood J. Community pharmacist workflow and medication therapy management delegation: an assessment of preferences and barriers. J Am Pharm Assoc (2003) 2020;60(6):e215-e223. [doi: 10.1016/j.japh.2020.07.024] [Medline: 32811750]
9. Occupational Employment and Wages, May 2019. U.S. Bureau of Labor Statistics. URL: https://www.bls.gov/oes/2019/may/oes291051.htm [accessed 2020-07-06]
10. Occupational Employment and Wages, May 2023. U.S. Bureau of Labor Statistics. URL: https://www.bls.gov/oes/current/oes292052.htm [accessed 2024-08-13]
11. Szeinbach S, Seoane-Vazquez E, Parekh A, Herderick M. Dispensing errors in community pharmacy: perceived influence of sociotechnical factors. Int J Qual Health Care 2007;19(4):203-209. [doi: 10.1093/intqhc/mzm018] [Medline: 17567597]
12. Um IS, Clough A, Tan ECK. Dispensing error rates in pharmacy: a systematic review and meta-analysis. Res Social Adm Pharm 2024;20(1):1-9 [FREE Full text] [doi: 10.1016/j.sapharm.2023.10.003] [Medline: 37848350]
13. Aldhwaihi K, Schifano F, Pezzolesi C, Umaru N. A systematic review of the nature of dispensing errors in hospital pharmacies. Integr Pharm Res Pract 2016;5:1-10 [FREE Full text] [doi: 10.2147/IPRP.S95733] [Medline: 29354533]
14. Parasuraman R, Mouloua M. Interaction of signal discriminability and task type in vigilance decrement. Percept Psychophys 1987;41(1):17-22. [doi: 10.3758/bf03208208] [Medline: 3822739]
15. Warm JS, Dember WN, Hancock PA. Vigilance and workload in automated systems. In: Automation and Human Performance: Theory and Applications. Boca Raton, Florida: CRC Press; 1996.
16. Manly T, Robertson IH, Galloway M, Hawkins K. The absent mind: further investigations of sustained attention to response. Neuropsychologia 1999;37(6):661-670. [doi: 10.1016/s0028-3932(98)00127-4] [Medline: 10390027]
17. Thomson DR, Besner D, Smilek D. A resource-control account of sustained attention: evidence from mind-wandering and vigilance paradigms. Perspect Psychol Sci 2015;10(1):82-96. [doi: 10.1177/1745691614556681] [Medline: 25910383]
18. Marusich LR, Bakdash JZ, Onal E, Yu MS, Schaffer J, O'Donovan J, et al. Effects of information availability on command-and-control decision making: performance, trust, and situation awareness. Hum Factors 2016;58(2):301-321. [doi: 10.1177/0018720815619515] [Medline: 26822796]
19. Wu L, Zhu Z, Cao H, Li B. Influence of information overload on operator's user experience of human–machine interface in LED manufacturing systems. Cogn Tech Work 2016;18:161-173. [doi: 10.1007/s10111-015-0352-0]
20. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. 'Hello AI': uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. 2019 Presented at: Proceedings of the ACM on Human-Computer Interaction; 2019 November 7; United States. [doi: 10.1145/3359206]
21. Jussupow E, Spohrer K, Heinzl A, Gawlitza J. Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. Inf Syst Res 2021;32(3):713-735. [doi: 10.1287/isre.2020.0980]

XSL•FO
RenderX

22. Prabhudesai S, Yang L, Asthana S, Huan X, Liao QV, Banovic N. Understanding uncertainty: how lay decision-makers perceive and interpret uncertainty in human-AI decision making. 2023 Presented at: Proceedings of the 28th International Conference on Intelligent User Interfaces; 2023 March 27; Australia p. 379-396. [doi: 10.1145/3581641.3584033]
23. von Eschenbach WJ. Transparency and the black box problem: why we do not trust AI. Philos Technol 2021;34:1607-1622. [doi: 10.1007/s13347-021-00477-0]
24. Wolfstadt JI, Gurwitz JH, Field TS, Lee M, Kalkar S, Wu W, et al. The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: a systematic review. J Gen Intern Med 2008;23(4):451-458 [FREE Full text] [doi: 10.1007/s11606-008-0504-5] [Medline: 18373144]
25. Jones-Jang SM, Park YJ. How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. J Comput Mediat Commun 2023;28(1):zmac029. [doi: 10.1093/jcmc/zmac029]
26. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. SSRN Journal 2014;144(1):1-114. [doi: 10.2139/ssrn.2466040]
27. Zheng Y, Rowell B, Chen Q, Kim JY, Kontar RA, Yang XJ, et al. Designing human-centered AI to prevent medication dispensing errors: focus group study with pharmacists. JMIR Form Res 2023;7:e51921 [FREE Full text] [doi: 10.2196/51921] [Medline: 38145475]
28. Wang H, Yeung DY. A survey on Bayesian deep learning. ACM Comput Surv 2020;53(5):1-37. [doi: 10.1145/3409383]
29. Lester CA, Li J, Ding Y, Rowell B, Yang J, Kontar RA. Performance evaluation of a prescription medication image classification model: an observational cohort. NPJ Digit Med 2021;4(1):118 [FREE Full text] [doi: 10.1038/s41746-021-00483-8] [Medline: 34315995]
30. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. New York: PMLR; 2016 Presented at: ICML'16: Proceedings of the 33rd International Conference on Machine Learning; 2016 June 19-24; New York p. 1050-1059. [doi: 10.5555/3045390.3045502]
31. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 June 30; Las Vegas, NV p. 770-778. [doi: 10.1109/cvpr.2016.90]
32. Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: a synthesis of the literature. Theoretical Issues in Ergonomics Science 2007 May;8(3):201-212 [FREE Full text] [doi: 10.1080/14639220500370105]
33. Kaduk T, Goeke C, Finger H, König P. Webcam eye tracking close to laboratory standards: comparing a new webcam-based system and the EyeLink 1000. Behav Res Methods 2024;56(5):5002-5022. [doi: 10.3758/s13428-023-02237-8] [Medline: 37821751]
34. Eye-tracking-fixation-detection. Labvanced. URL: https://github.com/Labvanced/eye-tracking-fixation-detection [accessed 2024-12-28]
35. Zu T, Hutson J, Loschky LC, Rebello NS. Use of eye-tracking technology to investigate cognitive load theory. 2017 Presented at: Physics Education Research Conference; 2017 July 26-27; Cincinnati p. 472-475. [doi: 10.1119/perc.2017.pr.113]
36. Chen S, Epps J, Ruiz N, Chen F. Eye activity as a measure of human mental effort in HCI. Ahmedabad: HCI; 2011 Presented at: IUI '11: 16th International Conference on Intelligent User Interfaces; 2011 February 16; Palo Alto CA USA. [doi: 10.1145/1943403.1943454]
37. Joseph AW, Murugesh R. Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. J Sci Res 2020;64(1):168-175. [doi: 10.37398/jsr.2020.640137]
38. Lester C, Rowell B, Zheng Y, Co Z, Marshall V, Kim JY, et al. Effect of uncertainty-aware artificial intelligence models on human reaction time and decision-making: a randomized controlled trial. JMIR Preprints Preprint published on July 30, 2024 [FREE Full text] [doi: 10.2196/preprints.64902]
39. Wang Q, Yang S, Liu M, Cao Z, Ma Q. An eye-tracking study of website complexity from cognitive load perspective. Decis Support Sys 2014;62:1-10. [doi: 10.1016/j.dss.2014.02.007]
40. Pedreschi D, Giannotti F, Guidotti R, Monreale A, Ruggieri S, Turini F. Meaningful explanations of black box AI decision systems. AAAI 2019;33(01):9780-9784. [doi: 10.1609/aaai.v33i01.33019780]
41. Witry MJ, Arya V, Bakken BK, Gaither CA, Kreling DH, Mott DA, et al. National pharmacist workforce study (NPWS): description of 2019 survey methods and assessment of nonresponse bias. Pharmacy (Basel) 2021;9(1):20 [FREE Full text] [doi: 10.3390/pharmacy9010020] [Medline: 33451045]
42. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020;3(3):326-331 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa033] [Medline: 33215066]
43. Reis WC, Bonetti AF, Bottacin WE, Reis AS, Souza TT, Pontarolo R, et al. Impact on process results of clinical decision support systems (CDSSs) applied to medication use: overview of systematic reviews. Pharm Pract (Granada) 2017;15(4):1036 [FREE Full text] [doi: 10.18549/PharmPract.2017.04.1036] [Medline: 29317919]
44. Patel SK, Kelm MJ, Bush PW, Lee HJ, Ball AM. Prevalence and risk factors of burnout in community pharmacists. J Am Pharm Assoc (2003) 2021;61(2):145-150. [doi: 10.1016/j.japh.2020.09.022] [Medline: 33069594]
45. Pervanas HC, Revell N, Alotaibi AF. Evaluation of medication errors in community pharmacy settings: a retrospective report. J Pharm Technol 2016;32(2):71-74. [doi: 10.1177/8755122515617199] [Medline: 34861023]

XSL•FO
RenderX

46. Schenkelberg CV, Al-Khatib A, Bakken BK, Arya V, Gaither CA, Kreling DH, et al. Identifying services provided in community pharmacy practice settings. Innov Pharm 2023;14(3) [FREE Full text] [doi: 10.24926/iip.v14i3.5543] [Medline: 38487385]
47. Negaard BJ, Lyons KP, Nichol CL, Polgreen LA. What does a pharmacist do? A time and motion study. Res Social Adm Pharm 2020;16(9):1314-1317. [doi: 10.1016/j.sapharm.2019.03.007] [Medline: 30890316]
48. Wash A, Moczygemba LR, Brown CM, Crismon ML, Whittaker TA. A narrative review of the well-being and burnout of U.S. community pharmacists. J Am Pharm Assoc (2003) 2024;64(2):337-349. [doi: 10.1016/j.japh.2023.11.017] [Medline: 37967722]
49. Kamboj N, Metcalfe K, Chu CH, Conway A. Designing the user interface of a nitroglycerin dose titration decision support system: user-centered design study. Appl Clin Inform 2024;15(3):583-599. [doi: 10.1055/s-0044-1787755] [Medline: 39048084]
50. Westerbeek L, de Bruijn GJ, van Weert HC, Abu-Hanna A, Medlock S, van Weert JCM. General practitioners' needs and wishes for clinical decision support systems: a focus group study. Int J Med Inform 2022;168:104901 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104901] [Medline: 36279654]
51. Hwang J, Lee T, Lee H, Byun S. A clinical decision support system for sleep staging tasks with explanations from artificial intelligence: user-centered design and evaluation study. J Med Internet Res 2022;24(1):e28659 [FREE Full text] [doi: 10.2196/28659] [Medline: 35044311]
52. Miller SD, Murphy Z, Gray JH, Marsteller J, Oliva-Hemker M, Maslen A, et al. Human-centered design of a clinical decision support for anemia screening in children with inflammatory bowel disease. Appl Clin Inform 2023;14(2):345-353 [FREE Full text] [doi: 10.1055/a-2040-0578] [Medline: 36809791]
53. Clausen C, Leventhal B, Nytrø Ø, Koposov R, Røst TB, Westbye OS, et al. Usability of the IDDEAS prototype in child and adolescent mental health services: a qualitative study for clinical decision support system development. Front Psychiatry 2023;14:1033724 [FREE Full text] [doi: 10.3389/fpsyt.2023.1033724] [Medline: 36911136]
54. Elhaddad M, Hamam S. AI-driven clinical decision support systems: an ongoing pursuit of potential. Cureus 2024;16(4):e57728 [FREE Full text] [doi: 10.7759/cureus.57728] [Medline: 38711724]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support systems
ML: machine learning
NDC: National Drug Code

Edited by Y Li; submitted 16.05.24; peer-reviewed by T Watterson, Q Wang, C Manlhiot; comments to author 01.08.24; revised version received 10.09.24; accepted 11.12.24; published 31.01.25.

Please cite as:
Tsai CC, Kim JY, Chen Q, Rowell B, Yang XJ, Kontar R, Whitaker M, Lester C
Effect of Artificial Intelligence Helpfulness and Uncertainty on Cognitive Interactions with Pharmacists: Randomized Controlled Trial
J Med Internet Res 2025;27:e59946
URL: https://www.jmir.org/2025/1/e59946
doi:10.2196/59946
PMID:39888668

©Chuan-Ching Tsai, Jin Yong Kim, Qiyuan Chen, Brigid Rowell, X Jessie Yang, Raed Kontar, Megan Whitaker, Corey Lester. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Optimizing Engagement With a Smartphone App to Prevent Violence Against Adolescents: Results From a Cluster Randomized Factorial Trial in Tanzania

Roselinde Janowski[1], MA; Lucie D Cluver[1,2], DPhil; Yulia Shenderovich[3,4], PhD; Joyce Wamoyi[5], PhD; Mwita Wambura[5], PhD; David Stern[6], PhD; Lily Clements[6], PhD; G J Melendez-Torres[7], DPhil; Lauren Baerecke[8], MA; Abigail Ornellas[8], PhD; Angelique Nicole Chetty[8], BA; Jonathan Klapwijk[1], MA; Laetitia Christine[9], BSc; Ateamate Mukabana[9], BSc; Esmee Te Winkel[6], PhD; Anna Booij[10], MA; Gervas Mbosoli[5], BSc; Jamie M Lachman[1,8,11], DPhil

[1]Department of Social Policy and Intervention, University of Oxford, Oxford, United Kingdom

[2]Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa

[3]Centre for Development, Evaluation, Complexity, and Implementation in Public Health Improvement (DECIPHer), School of Social Sciences, Cardiff University, Cardiff, United Kingdom

[4]Wolfson Centre for Young People's Mental Health, Cardiff University, Cardiff, United Kingdom

[5]Mwanza Research Centre, National Institute for Medical Research, Mwanza, United Republic of Tanzania

[6]Innovations in Development, Education and the Mathematical Sciences (IDEMS) International, Reading, United Kingdom

[7]Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom

[8]Centre for Social Science Research, University of Cape Town, Cape Town, South Africa

[9]Innovations in Development, Education and the Mathematical Sciences (INNODEMS), Kakamega, Kenya

[10]Clowns Without Borders South Africa, Cape Town, South Africa

[11]Parenting for Lifelong Health, Oxford, United Kingdom

Corresponding Author:
Roselinde Janowski, MA
Department of Social Policy and Intervention
University of Oxford
Barnett House
32-37 Wellington Square
Oxford, OX1 2ER
United Kingdom
Phone: 44 01865270325
Email: roselinde.janowski@spi.ox.ac.uk

Abstract

Background: Violence and abuse exert extensive health, social, and economic burdens on adolescents in low- and middle-income countries. Digital parenting interventions are promising for mitigating risks at scale. However, their potential for public health impact hinges on meaningful engagement with the digital platform.

Objective: The objective of this study was to evaluate the impact of 3 intervention design and implementation factors aimed at increasing engagement with a noncommercialized, offline-first smartphone app for caregivers of adolescents in Tanzania, in partnership with the United Nations Children's Fund, the World Health Organization, and the Tanzanian national government.

Methods: Following Multiphase Optimization Strategy (MOST) principles, we conducted a 2×2×2 cluster randomized factorial trial involving caregivers of adolescents aged 10 to 17 years. Caregivers were recruited by community representatives from 16 urban and periurban communities (ie, clusters) in the Mwanza region of Tanzania. Each cluster was randomized to 1 of 2 levels of each factor: *guidance* (self-guided or guided via facilitator-moderated WhatsApp groups), *app design* (structured or unstructured), and preprogram *digital support* (basic or enhanced). Primary outcomes were automatically tracked measures of engagement (app launches, modules completed, and home practice activities reviewed), with secondary outcomes including modules started, time spent in the app, and positive behaviors logged. Generalized linear mixed-effects models assessed the impact of experimental factors on engagement.

XSL•FO
RenderX

Results: Automatically tracked engagement data from 614 caregivers were analyzed, of which 205 (33.4%) were men. Compared to self-guided participants, receiving guidance alongside the app led to significantly more app launches (mean ratio [MR] 2.93, 95% CI 1.84-4.68; *P*<.001), modules completed (MR 1.29, 95% CI 1.05-1.58; *P*=.02), modules started (MR 1.20, 95% CI 1.02-1.42; *P*=.03), time spent in the app (MR 1.45, 95% CI 1.39-1.51; *P*<.001), and positive behavior logs (MR 2.73, 95% CI 2.07-3.60; *P*<.001). Compared to the structured design, unstructured design use resulted in significantly more modules completed (MR 1.49, 95% CI 1.26-1.76; *P*<.001), home practice activity reviews (MR 7.49, 95% CI 5.19-10.82; *P*<.001), modules started (MR 1.27, 95% CI 1.06-1.52; *P*=.01), time spent in the app (MR 1.84, 95% CI 1.70-1.99; *P*<.001), and positive behavior logs (MR 55.68, 95% CI 16.48-188.14; *P*<.001). While analyses did not detect an effect of enhanced digital support on directly observed engagement, the combination of enhanced digital support and guidance positively influenced engagement across a range of outcomes.

Conclusions: This study is the first to systematically optimize engagement with a digital parenting intervention in a low- and middle-income country. Our findings offer important learnings for developing evidence-based, scalable digital interventions in resource-constrained settings.

Trial Registration: Pan-African Clinical Trial Registry PACTR202210657553944; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=24051

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-023-15989-x

(J Med Internet Res 2025;27:e60102) doi:10.2196/60102

KEYWORDS

digital health; engagement; parenting; adolescents; low- and middle-income country; violence against children; Multiphase Optimization Strategy; randomized factorial experiment; mobile phone

Introduction

Background

Violence against children (VAC) has far-reaching individual and societal consequences. Beyond physical harm, violence during childhood negatively impacts brain development and increases risks of poor educational attainment, substance use, and adverse mental and physical health conditions [1,2]. These impacts cascade into broader societal costs, including reduced social participation, unemployment, crime, and intergenerational transmission of violence [2-5], with global economic costs approximating US $7 trillion annually [6].

While VAC occurs in many settings, it is most prevalent within the home environment, where it is typically perpetrated by parents or other caregivers [7]. It spans physical, emotional, and sexual abuse, with disproportionately higher rates in low- and middle-income countries (LMICs), particularly in Africa [8,9]. In Tanzania, national survey data [10] revealed that >70% of respondents aged 13 to 24 years had experienced physical violence from parents, relatives, or guardians before the age of 18 years, including being slapped, kicked, beaten, pushed, or threatened with a weapon. Moreover, approximately 80% of those experiencing sexual violence reported physical violence, while emotional abuse was nearly universal among those who experienced physical violence [10]. A more recent nationally representative study by Nkuba et al [11] found even higher rates, with >90% of secondary school students reporting violent discipline from parents or primary caregivers and 80% of parents acknowledging such practices. This widespread prevalence is associated with limited knowledge of nonviolent parenting alternatives, high caregiver stress, and cultural norms emphasizing child obedience and respect, male honor, and family privacy [10,11].

Parenting interventions, identified by the World Health Organization as a key VAC prevention strategy [12], aim to strengthen parent-child relationships and replace harsh, abusive practices with developmentally appropriate, nonviolent discipline strategies [7]. These interventions are effective in preventing VAC globally [13], particularly when they incorporate skill-building in nonviolent discipline, positive reinforcement, emotion regulation, proactive parenting, and protective relationship promotion [14-16]. However, scaling traditional in-person delivery in LMICs, such as Tanzania, faces significant barriers, including prohibitive delivery costs, extensive human resource requirements, limited geographic accessibility for families, and competing work and childcare demands [17-20].

Digital delivery via smartphone apps offers a promising solution to these implementation and access barriers, particularly in LMIC regions, such as sub-Saharan Africa, where mobile phone penetration is rapidly increasing [21]. The success of mobile health interventions in delivering health care services in sub-Saharan Africa [22] and improving health outcomes across LMICs [23,24] further supports this potential. However, systematic reviews [25-31] show that digital parenting research is almost exclusively concentrated in high-income countries (HICs), with interventions predominantly focusing on child behavior management rather than violence prevention. Moreover, these interventions typically experience poor participant engagement [32], defined as both the extent of program usage (eg, amount, frequency, duration, and depth) and subjective user experience (eg, attention, interest, and affect) [33].

Poor engagement undermines an intervention's impact at both individual and population levels. At the individual level, participants fail to acquire essential skills needed to prevent VAC and improve child well-being, while at the population

XSL•FO
RenderX

level, governments and policy makers are unlikely to implement programs widely when only a small proportion of participants engage with and benefit from them [34]. Consequently, poor engagement severely limits the potential public health impact of digital parenting interventions [35,36]. Understanding and optimizing engagement in LMIC contexts is thus critical for establishing digital parenting interventions as a scalable violence prevention strategy.

Existing studies examining engagement with digital parenting interventions in HICs have largely been descriptive [37,38] or focus on sociodemographic and behavioral characteristics associated with engagement [39-42]. A recent review of technology-assisted parenting interventions has advanced this literature by identifying several potential engagement-promoting components, including practical technology support, interactive program features, tailoring, and control features [43]. Similar components have been identified across the broader mobile health literature, such as personalized content and feedback [44-46], user-friendliness [47], a visually appealing layout and the integration of a tutorial on how to use the app [44], human guidance [48-50], peer-based asynchronous communication features [51], and reminders [33,44,52]. However, experimental evidence for these components remains limited; to date, only 1 known trial has experimentally evaluated engagement strategies for digital parenting interventions [53], with no such trials conducted in LMICs. This gap is particularly significant given that digital interventions and their delivery strategies comprise multiple components, each with potentially different effects on engagement. Furthermore, findings from HICs may not generalize to LMIC settings, such as Tanzania, where technological infrastructure, digital literacy, and cultural contexts differ substantially.

Objectives

In response to the need for scalable violence prevention in LMICs, the Global Parenting Initiative, a 5-year research-within-implementation collaboration between universities, foundations, and implementing partners, is developing and evaluating open-source digital parenting programs based on the Parenting for Lifelong Health (PLH) intervention suite. One of these interventions is ParentApp, a smartphone app designed to address adolescent exposure to VAC. Using the Multiphase Optimization Strategy (MOST) framework [54] for systematic intervention optimization, we conducted a cluster randomized factorial trial to identify optimal components for maximizing engagement with ParentApp among socioeconomically disadvantaged caregivers of adolescents in Tanzania. The trial examined 3 factors: *guidance*, *app design*, and *digital support*. Our primary objective was to determine the independent effects of these factors on primary and secondary engagement outcomes. An exploratory objective was to identify potential interaction effects between the factors.

Methods

Trial Design

The trial used a cluster randomized factorial design with three experimental factors: (1) *guidance* (self-guided or guided; factor A), (2) *app design* (structured or unstructured; factor B), and (3) *digital support* (basic or enhanced; factor C). Each factor was varied at 2 levels (lower dose or inactive vs higher dose or active), yielding an 8-condition (2×2×2) trial. A full factorial design was used, allowing all clusters to potentially be randomized to all combinations of factors and their corresponding levels. Further details on the study design are provided in the published protocol [55]. The trial was preregistered on the Pan-African Clinical Trial Registry (PACTR202210657553944) on October 11, 2022. The CONSORT (Consolidated Standards of Reporting Trials) 2010 extension for randomized factorial trials [56] and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [57] were used for reporting.

Setting and Participants

Target communities were urban and periurban sub-wards characterized by high rates of poverty in the Mwanza region, Tanzania. Eligible clusters were sub-wards that (1) were located within 3 selected low-income wards, (2) had no previous involvement in ParentApp testing, and (3) contained at least 40 smartphone-owning households. Within each cluster, participants were primary caregivers of at least 1 adolescent aged 10 to 17 years. Caregivers were eligible if they (1) were aged ≥18 years, (2) lived in the same household as their adolescent for at least 4 nights per week in the previous month, (3) had regular access to an Android smartphone, and (4) provided written informed consent. Participants were excluded from the study if they were unable to read or had a severe learning disability that affected their ability to provide informed consent.

Recruitment, Enrollment, and App Onboarding

Following a 2-week community mapping period conducted by Tanzania's National Institute for Medical Research (NIMR) in October 2022, the fieldwork team, in collaboration with Tanzania-based nongovernmental organization (NGO) Investing in Children and Strengthening Their Societies (ICS), initiated recruitment procedures in 3 low-income wards. Centralized stakeholder meetings were held in 2 wards, while in the third ward, sub-ward-level consultations took place due to logistical constraints. A total of 31 community leaders, including ward executive officers, community development officers, and local representatives, were briefed about the study. Those who agreed to support the study were tasked with identifying and contacting potential eligible families within their respective sub-wards over an 8-week recruitment period. Community leaders provided standardized information about the study purpose and intervention content and invited interested families to attend orientation and onboarding sessions.

These sessions were organized by study cluster and held at accessible locations, such as schools and community centers. Sessions were cofacilitated by NIMR research assistants and ICS facilitators. Families first attended a group orientation where they received detailed information about the program objectives, study procedures, data collection requirements, and time commitments. Research assistants then conducted one-on-one eligibility screening using an Open Data Kit checklist. Written

XSL•FO
RenderX

informed consent was obtained from caregivers who met all eligibility criteria.

The subsequent app onboarding lasted approximately 90 minutes and was conducted in small groups of 5 to 10 participants. This included guided assistance with app installation via Google Play (Google LLC), technical troubleshooting, and an orientation to app features. Facilitators demonstrated key functionalities, such as module navigation, activity completion, and resource access. Participants then practiced these skills by completing the app's first module, which concluded with an integrated baseline assessment.

Intervention

Overview

ParentApp is a smartphone app adapted from PLH Teens, an in-person, group-based program originally developed in South Africa to address the need for cost-effective, culturally appropriate violence prevention [58]. Initial evaluations of PLH Teens demonstrated significant improvements in positive parenting practices, along with reductions in child maltreatment, parental depression, stress, substance use, and financial stress [58]. PLH Teens has since been adapted and widely implemented across >18 LMICs [59], including a large-scale delivery in Tanzania to >75,000 caregivers and adolescents [60]. Building on this strong, emerging evidence, ParentApp was developed to address access and scalability challenges associated with in-person delivery. The development and delivery of ParentApp in Tanzania were jointly led by the Universities of Oxford and Cape Town; PLH; Innovations in Development, Education and the Mathematical Sciences; Clowns Without Borders South Africa; NIMR; and ICS, in collaboration with the United Nations Children's Fund, the World Health Organization, and the Tanzanian national government. The intervention content is described in accordance with the TIDieR (Template for Intervention Description and Replication) checklist (Multimedia Appendices 1 and 2).

ParentApp Development

ParentApp underwent systematic development through four phases: (1) initial adaptation and codevelopment with researchers, program specialists, technical experts, and PLH Teens participants in South Africa; (2) user testing with 24 participants across 9 African countries [61]; (3) feasibility testing of self-guided delivery with 107 caregivers in South Africa; and (4) pre- and postpilot testing with 103 caregiver-adolescent dyads in Tanzania, which included remote guidance via phone calls or WhatsApp (Meta Platforms, Inc). Iterative refinements to content, design, and delivery approach across these phases produced the version of ParentApp evaluated in this trial.

ParentApp Design and Content

ParentApp is an open-source, offline-first Android app (compatible with version 5.5.1 or later) comprising 12 modules that integrate text, images, and audio to deliver a condensed version of the in-person program (Table 1). The introductory module focuses on parental self-care and stress reduction, concluding with a welcome survey that collects baseline data and enables users to select their preferred delivery format (individual or group) and engagement schedule. The core intervention consists of 10 modules covering evidence-based practices shown to effectively reduce harsh and abusive parenting [7]. A final module consolidates learning and includes a postintervention assessment.

Each module uses recurring activities to reinforce learning and practical skill building. Modules begin with mindfulness activities featuring emotional check-ins and relaxation exercises, followed by comics that illustrate positive and negative caregiver-adolescent scenarios with reflection questions. Audio testimonials from Tanzanian parents share benefits and experiences, while *Essential Tools* summaries provide quick access to key skills and examples. Modules conclude with home practice assignments that include structured activities and a follow-up *Home Practice Review* for activity reflection. Throughout the app, users can log their skill implementation through *ParentPoints*, a behavioral tracking system for monitoring 10 evidence-based practices that promote positive parenting and mental health (Multimedia Appendix 2). Beyond module activities, users have access to a library with essential parenting tips, local support resources, and technical support. To encourage ongoing engagement, automated push notifications are triggered at 1, 6, and 30 days following the user's last app launch.

XSL•FO
RenderX

Table 1. ParentApp module content and learning objectives.

Module	Core skills and objectives	Home practice[a]
1. Parental self-care and stress reduction	Developing essential mental and physical well-being skills to strengthen parents' capacity to support their families	Practice daily relaxation techniques, identify moments of achievement, and reward successes.
2. One-on-one time	Spending meaningful time with adolescents, using active listening to foster trust, communication, and positive relationships	Schedule and spend 5 to 20 min of focused time with adolescents daily.
3. Praise and positive reinforcement	Using effective praise techniques to encourage positive behavior in adolescents	Practice praising adolescents daily.
4. Positive instructions	Providing clear, specific, and positive directions to improve communication and cooperation	Use positive instructions with adolescents daily.
5. Managing stress	Enhancing emotional awareness and communication skills to support mental well-being and reduce the risk of violent reactions during challenging situations	Practice breathing exercises before reacting.
6. Family budgeting	Identifying strategies for managing and saving money to reduce financial stress and establish achievable family goals	Create a budget with adolescents and other family members.
7. Establishing rules and routines	Engaging adolescents in establishing household rules and routines to foster compliance and build a safe, supportive, and consistent home environment	Practice making rules together with adolescents.
8. Consequences and accepting responsibility	Promoting shared parent-adolescent responsibilities; applying realistic, appropriate consequences for noncompliance; and emphasizing positive discipline over punitive measures	Discuss negative and positive consequences with adolescents.
9. Problem-solving	Building family collaboration skills to identify, address, and resolve challenges through joint problem-solving	Teach adolescents 4-step problem-solving: know it, solve it, try it, and test it.
10. Teen safety in the community and online	Mapping risks in community and online environments to establish protective rules for adolescents	Map safe and unsafe spaces with adolescents and create a family safety plan.
11. Dealing with crisis	Remaining calm in crisis situations and developing immediate and long-term plans to respond effectively, using available support services	Discuss crisis scenarios with adolescents and create family crisis response plans.
12. Celebration and next steps	Reflecting on key learning and solidifying plans for ongoing peer support	Review key skills learned.

[a]Modules 2 to 11 include home practice assignments with follow-up reviews for activity reflection.

Experimental Factors

Overview

Table 2 presents an overview of the core features associated with each factor level. Justification for the selection of factors is also provided in the published protocol [55].

XSL•FO
RenderX

Table 2. Experimental factors, levels, and their core features.

Factor	Description	Core features
Guidance		
Self-guided	Participants accessed ParentApp's full content throughout the study.	• ParentApp without human support
Guided	Participants accessed ParentApp's full content throughout the study and received WhatsApp group guidance.	• ParentApp • WhatsApp groups moderated by 1 lead facilitator and 1 cofacilitator • Facilitators monitored discussions and shared tips and reminders • Facilitators led a weekly 1-hour live-chat session per group
App design		
Structured	ParentApp content followed a sequential, weekly format.	• 12 modules • First module available immediately, with 1 new module unlocked every 7 days • Sequential module order • Content-related notifications • Images featuring generic blue amorphous cartoon characters
Unstructured	ParentApp content followed an open, nonsequential format.	• 12 modules • All modules available immediately • Nonsequential module order • Modules divided into shorter tasks and activities • Content-related notifications disabled • Culturally adapted images featuring humanlike figures
Digital support		
Basic	Participants received an in-app navigation tour at first app launch, followed by a group-based orientation of the app.	• Brief built-in app navigation tour showing core app features • Group orientation on core features led by trained research assistants
Enhanced	Participants received an in-app navigation tour at first app launch, followed by a group-based orientation of the app and a supplementary training session focusing on broader smartphone skills.	• Brief built-in app navigation tour showing core app features • Group orientation on core features led by trained research assistants • Supplementary group training (15 to 20 min) on general smartphone skills led by trained research assistants

Factor A: Guidance (Self-Guided or Guided)

During formative testing, WhatsApp groups showed promise for enhancing app-based delivery. In this trial, half of the clusters received no human support after app onboarding, while the remaining half received guidance via facilitator-moderated WhatsApp groups. The purpose of these groups was to promote social connection and sustain engagement with ParentApp. Each cluster in the guided condition had a corresponding WhatsApp group, overseen by a female-male facilitator pair (totaling 8 groups and 4 facilitator pairs). Following a standardized manual, facilitators shared tips and reminders, monitored discussions, addressed questions, and led weekly 1-hour live-chat sessions to encourage discussion and reflection. The facilitators, who were locally recruited paraprofessionals with previous experience supporting caregivers remotely during the Tanzanian pilot, received training from Clowns Without Borders South Africa, an NGO experienced in training PLH program facilitators.

Factor B: App Design (Structured or Unstructured)

ParentApp was originally designed with a sequential progression to match PLH Teens' learning objectives and time frame. The original design also featured generic blue cartoon characters intentionally designed to facilitate global implementation. However, early testing highlighted 2 potential areas for improvement: some participants found the fixed weekly module releases limiting, while others questioned the acceptability of the generic illustrations. In response, this trial evaluated 2 versions of the app's design (refer to Figures 1 and 2 for screenshot examples). Half of the clusters received the structured design, which maintained sequential module delivery at 7-day intervals over 12 weeks. This version included automated content notifications aligned with module releases, scheduled unlocking of home practice reviews for modules 2 to 11 at the end of each 7-day cycle, and the original generic illustrations. The remaining clusters received the unstructured design, which offered immediate access to all modules and activities after installation. This version incorporated several modifications: users could navigate the 12 modules at their own pace and in their preferred sequence; modules were subdivided into shorter tasks for easier navigation, with home practice review as the final task for modules 2 to 11; automated content notifications were disabled to accommodate flexible access; and generic cartoon images were replaced with culturally adapted illustrations featuring humanlike characters tailored to the Tanzanian context.

XSL•FO
RenderX

Figure 1. Screenshot examples of the structured app design with generic illustrations.



Figure 2. Screenshot examples of the unstructured app design with culturally adapted illustrations.



XSL•FO
RenderX

Factor C: Digital Support (Basic or Enhanced)

During formative testing, digital literacy challenges emerged as a critical consideration for app engagement. Consequently, this trial evaluated 2 levels of digital support during the onboarding session. Basic support, offered to all clusters, included an embedded ParentApp navigation tour followed by a group-based orientation of the app with demonstrations and practice opportunities. Half of the clusters received enhanced support, which included an additional 15- to 20-minute session focused on improving general smartphone literacy and user confidence. This supplementary training was developed by Innovations in Development, Education and the Mathematical Sciences Kenya, an NGO providing technical support during trial implementation. The training covered essential smartphone skills, including app installation, effective internet use, and online safety, and was delivered by research assistants at the end of the onboarding session. Both levels of digital support were optional.

Measures and Assessment

Overview

Engagement was operationalized across three phases: (1) initial, (2) ongoing, and (3) quality, aligning with established digital intervention literature [33,43,62,63]. Engagement data were collected through automated tracking each time a participant interacted with ParentApp. Use records, including module completion rates, content viewing patterns, and responses to in-app tasks and the baseline assessment, were uploaded to a Metabase cloud server (Metabase, Inc) whenever a device accessed the internet. Additional use data were automatically captured through Matomo Analytics (InnoCraft Ltd), including visit timestamps, session duration, and the frequency of actions or pages viewed. Data monitoring protocols were implemented to ensure accurate tracking and verify that participants received the correct app version based on their cluster allocation. The final dataset was extracted from Metabase and Matomo Analytics in March 2023, approximately 4 months after the last cluster of participants enrolled in the study. Research assistants and facilitators reported estimated attendance rates and session durations for the enhanced digital support training and WhatsApp live-chat sessions, respectively.

Primary Outcomes

Three primary engagement outcomes were evaluated: (1) the number of *app launches* throughout the intervention period (initial engagement), (2) the number of *modules completed* out of 12 (ongoing engagement), and (3) the number of *home practice activities reviewed* out of 10 (quality of engagement). The home practice review outcome underwent a post hoc modification from the protocol, evaluating user initiation of postactivity reviews rather than access to the home practice tab for activity instructions.

Secondary Outcomes

Secondary engagement outcomes included (1) the number of *modules started* out of 12 (initial engagement), (2) the total *time spent in the app* during the intervention period (ongoing engagement), and (3) the number of self-reported *positive parenting and mental health–promoting behaviors logged* throughout the intervention period (quality of engagement).

Baseline Sociodemographic and Behavioral Measures

Baseline demographic and behavioral data were collected using a self-administered questionnaire embedded within the app's welcome survey in the first module. The questionnaire included multiple response formats: demographic questions primarily used binary choice buttons (eg, gender: woman or man) or number entry fields (eg, age), while behavioral measures used frequency scales. Participants could navigate between pages to review and modify their responses, with the assessments taking approximately 10 to 15 minutes to complete. All behavioral measures were nonproprietary, freely available, and recommended by the Global Parenting Initiative. Given the trial's focus on engagement rather than caregiver outcomes, abbreviated scales were used to minimize participant burden. All measures were translated into Kiswahili and back translated and were previously used in Tanzania [60].

Demographic variables included caregiver age and gender, household structure, orphanhood status, and 2 items adapted from the Financial Self-Efficacy Scale on financial stress and food insecurity (1 item each) [64]. Child maltreatment was assessed using 4 items from the International Society for the Prevention of Child Abuse and Neglect Screening Tool-Trial Version [65]. This included subscales for physical abuse (2 items; eg, "How many times in the past month did you hit, spank or slap your child with a hand or object?") and emotional abuse (2 items; eg, "How many times in the past month did you shout, scream or yell at your child?"), rated on a 0 to ≥8 frequency scale. Positive parenting was measured using 5 items from the Alabama Parenting Questionnaire [66], assessing involvement (2 items; eg, "How many times in the past month did you praise your teen?") and supervision (3 items; eg, "How many times in the past month did your teen stay out past their curfew?"), rated on a 0 to ≥8 frequency scale. Parental depression was measured using 3 items from the Centre for Epidemiologic Studies Depression Scale [67]. Items were rated on a 0- to 7-day frequency scale (eg, "How often did you feel that everything you did was an effort?"). Parenting stress was assessed using 1 item adapted from the Parental Stress Scale [68], "How many times in the past month did caring for your children make you feel very stressed?" rated on a 0 to ≥8 frequency scale.

Power

An a priori simulation–based power analysis determined that 640 caregivers across 16 clusters (40 per cluster) were needed for adequate statistical power [55]. This cluster size was selected for future scalability, as larger WhatsApp groups offer greater reach and cost-effectiveness. Simulations accounted for clustering at the sub-ward level, participant-level covariates, and an α level of .05.

Randomization

The factorial experiment was conducted at the cluster level for logistical efficiency, streamlining screening, enrollment, and baseline data collection for participants receiving the same experimental conditions. Community representatives compiled lists of potentially eligible families within their designated

XSL•FO
RenderX

sub-wards. Fieldwork coordinators then formed clusters by grouping geographically proximate sub-wards to minimize the risk of cross contamination. An off-site research team member performed single-stage randomization, assigning each eligible cluster to 1 of the 8 experimental condition combinations using a randomization algorithm in R (R Foundation for Statistical Computing). Final cluster sizes ranged from 32 to 49 caregivers (mean 38.38, SD 6.20).

Blinding

Blinding was not feasible for the in-country research team and facilitators due to their involvement in implementation. Participants were only informed about the experimental conditions to which their cluster was assigned. The data analyst was blinded to cluster allocation but not to condition assignment due to effect coding [69].

Statistical Analysis

Analyses used generalized linear mixed-effects models with a Poisson distribution to estimate the impact of experimental factors on engagement. Models included a random intercept term to account for the nesting of participants within clusters. Intention-to-treat analyses were conducted but included only participants who successfully installed ParentApp, as data collection was contingent upon app installation. Supplementary analyses found no significant effect of cluster randomization to experimental conditions on app installation status (Table S1 in Multimedia Appendix 3). Missingness in engagement outcomes was minimal, affecting only 2 variables: app launches (6/614, 1%) and time spent on the app (11/614, 1.8%). Multiple imputation with fully conditional specification [70] was used to address missingness in caregiver baseline data (5% to 7.5% at item level). A total of 10 imputations were conducted using predictive mean matching. Clustering was accounted for by specifying –2 in the predictor matrix. Imputation at the item level was passive, summing behavioral items to generate total scores within the imputation process. Following imputation, caregiver baseline data were sample-mean centered within each of the 10 imputed datasets. Analyses were conducted in R (version 4.3.1), using the glmer function from the *lme4* package for generalized linear mixed-effects models [71], and imputations were performed using the *mice* package [72].

In primary analyses, each factor's individual effect on primary and secondary outcomes was evaluated separately. Models were adjusted for the prespecified covariates: caregiver age, caregiver gender, financial stress, food insecurity, child maltreatment (physical and emotional abuse), positive parenting involvement and supervision, parental depression, and parenting stress. In exploratory analyses, all 3 factors were included as main effects alongside their 2-way interactions, while adjusting for baseline characteristics. The inclusion of interaction terms followed recent recommendations for factorial designs to account for potential interdependencies between factors [73]. All main and 2-way interaction effects were estimated using the full sample. Assuming similar effect sizes, this provided equally high power to detect all effects [54]. Estimates for each model were pooled across the 10 imputed datasets using Rubin's rules [72]. Factors were effect-coded, assigning –1 to inactive or lower-dose levels and 1 to active or higher-dose levels (Multimedia Appendix 4). This coding scheme calculates the difference between levels as (+1 – (–1) = 2), whereas with dummy codes, it is simply (1 – 0 = 1) [69]. To ensure an accurate representation of the effect of a 1-unit change in the predictor variable within the effect coding scheme, unstandardized regression coefficients and their SEs were multiplied by a scaling constant of 2. Poisson regression coefficients (which represent the log of the expected change in the outcome per 1-unit increase in the predictor while holding other variables constant) and 95% CIs were exponentiated for easier interpretation. The main effects were interpreted as the ratio of mean event counts or mean ratio (MR), comparing participants in clusters randomized to the higher-dose factor level to those randomized to the contrasting factor level. Interaction terms were interpreted as the additional effect of 1 factor level in the presence of another.

Ethical Considerations

Ethics approval was granted by NIMR (NIMR/HQ/R.8s/Vol.IX/3856) and the University of Oxford Departmental Research Ethics Committee (R69744/RE001). Written informed consent was obtained from all study participants at the onboarding session. Participants also accepted the app's terms and conditions and privacy policy, which detailed how personal information from ParentApp would be collected, used, and shared. After completing the onboarding session, participants received a US $2 honorarium and monthly 1 GB internet bundles for 4 months to support data synchronization with study servers and WhatsApp group participation where applicable. No additional incentives or compensation were provided.

Results

Participant Flow

Participant recruitment and onboarding activities began on October 24, 2022, and concluded on December 1, 2022, coinciding with Tanzania's short rainy season. Of 866 caregivers screened for eligibility, 680 (78.5%) met the inclusion criteria and consented to participate. Among enrolled participants, 66 (9.7%) did not complete app installation due to technical difficulties, including phone-to-app incompatibility, insufficient storage capacity, and internet instability, while others left the onboarding session prematurely due to poor weather conditions and scheduling conflicts. The final sample included 614 participants (90.3% of those enrolled) who successfully installed the app, all of whom were included in intention-to-treat analyses. A CONSORT diagram illustrating participant flow through the study is presented in Figure 3.

XSL•FO
RenderX

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) diagram of participant flow through the 2×2×2 factorial trial.

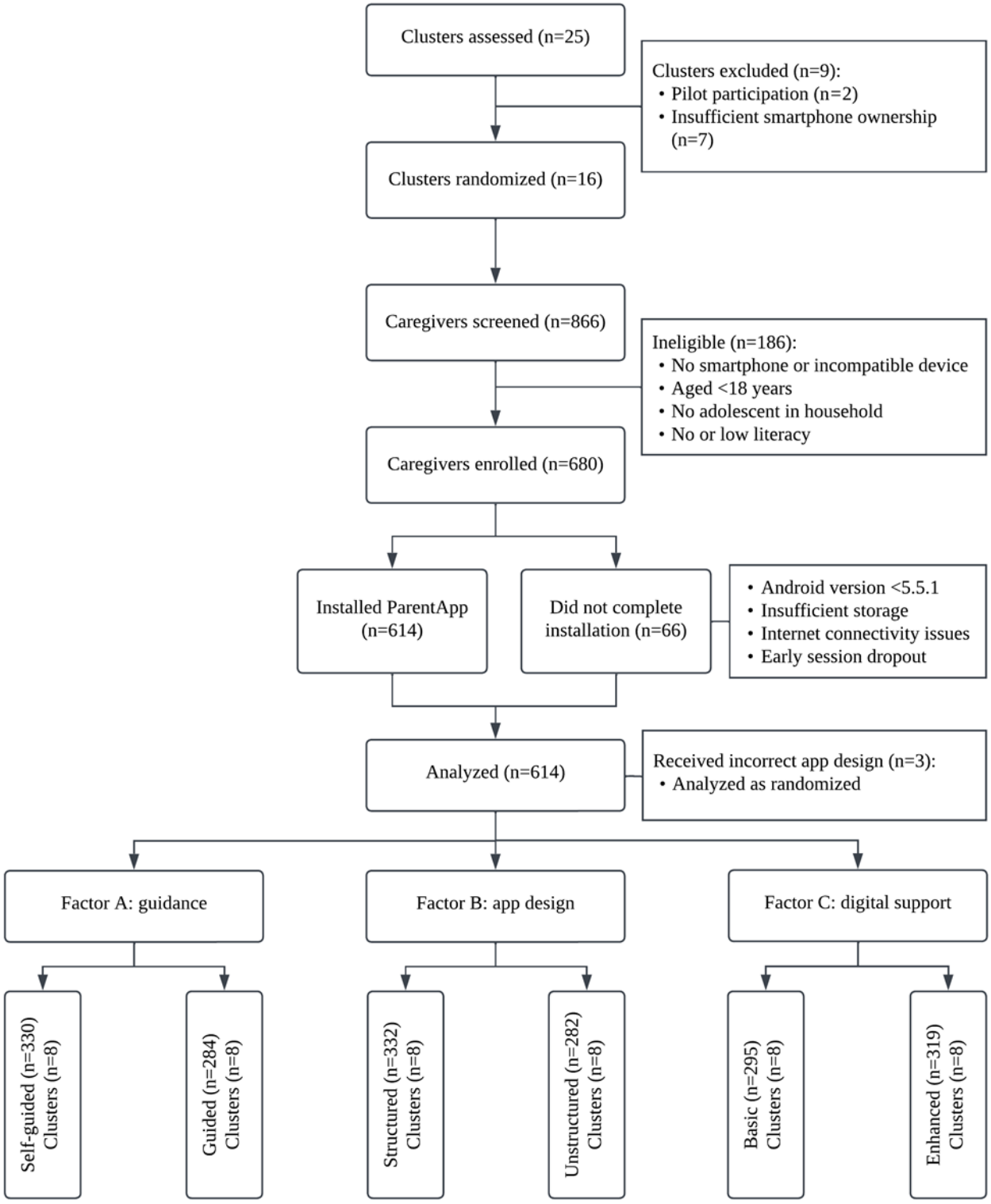

Participant Characteristics

Participant baseline characteristics, summarized in Table 3, showed a comparable distribution across the 3 factors' levels and the overall sample. Caregiver age ranged from 18 to 75 (mean 35.94, SD 11.84) years. One-third (205/614, 33.4%) of the caregivers were men, which is notable given the challenges parenting programs typically face in engaging fathers [74]. Households comprised an average of 3 (SD 1.80) adults and 3.61 (SD 2.12) children aged between 0 and 17 years. Most caregivers (485/576, 84.2%) reported experiencing food insecurity at least once in the past month, with 35.1% (202/576) reporting ≥7 instances during this period. Nearly half (276/576, 47.9%) of the sample reported that their adolescents had experienced the loss of a primary caregiver during their lifetime. Among these, 72.1% (199/276) reported that the death had occurred within the past 3 years, a trend likely influenced by the COVID-19 pandemic.

XSL•FO
RenderX

Table 3. Baseline characteristics for the full sample and each factor level (N=614).

Characteristic	Total[a]	Factor A: guidance		Factor B: app design		Factor C: digital support	
		Self-guided	Guided	Structured	Unstructured	Basic	Enhanced
Caregiver, n	n=577	n=307	n=270	n=307	n=268	n=281	n=296
Age (y), mean (SD)	35.94 (11.84)	35.83 (12.09)	36.07 (11.56)	37.16 (11.54)	34.54 (12.04)	36.35 (11.67)	35.55 (12.00)
Gender[b]	n=614	n=330	n=284	n=332	n=282	n=295	n=319
Women, n (%)	409 (66.6)	217 (65.8)	192 (67.6)	244 (73.5)	165 (58.5)	193 (65.4)	216 (67.7)
Men, n (%)	205 (33.4)	113 (34.2)	92 (32.4)	88 (26.5)	117 (41.5)	102 (34.6)	103 (32.3)
Household size[c]	n=580	n=310	n=270	n=313	n=267	n=281	n=299
Overall, mean (SD)	6.60 (3.23)	6.55 (3.53)	6.66 (2.85)	6.70 (3.46)	6.48 (2.95)	6.46 (3.16)	6.73 (3.30)
Adults, mean (SD)	2.99 (1.80)	3.04 (1.99)	2.94 (1.57)	3.06 (2.04)	2.91 (1.49)	2.9 (1.71)	3.05 (1.89)
Children (aged 0-17 y), mean (SD)	3.61 (2.12)	3.51 (2.29)	3.72 (1.90)	3.64 (2.16)	3.57 (2.07)	3.53 (2.16)	3.68 (2.08)
Food insecurity	n=576	n=310	n=266	n=308	n=268	n=277	n=299
≥1 day in the past month, n (%)	485 (84.2)	267 (86.1)	218 (82)	258 (83.8)	227 (84.7)	235 (84.8)	250 (83.6)
≥7 days in the past month, n (%)	202 (35.1)	113 (36.5)	89 (33.5)	108 (35.1)	94 (35.1)	103 (37.2)	99 (33.1)
Orphanhood[d]	n=576	n=308	n=268	n=306	n=270	n=277	n=299
Lifetime, n (%)	276 (47.9)	160 (51.9)	116 (43.3)	153 (50)	123 (45.6)	137 (49.5)	139 (46.5)
Past 3 years, n (%)	199 (72.1)	117 (73.1)	82 (70.7)	108 (70.6)	91 (74)	95 (69.3)	104 (74.8)

[a]Sample sizes vary due to item nonresponse.

[b]Missing caregiver gender data were supplemented with information from the screening questionnaires, resulting in 614 complete observations.

[c]Cases missing all household size data were excluded from the analysis.

[d]Past 3-year orphanhood percentages were calculated using lifetime orphanhood as the denominator.

Intervention Exposure and Engagement

Exposure

Among the 614 participants who successfully installed the app, 600 (97.7%) engaged with at least some of the module content throughout the study, while 14 (2.3%) did not access any content. On average, participants visited the app 3.69 (SD 4.73) times, where a visit was defined as either the initial app launch or visiting a page after a 30-minute interval since the last page view. Each visit involved approximately 20.75 (SD 17.10) actions or page views, totaling 63.81 (SD 85.54) actions on average throughout the study.

For the guidance factor, 53.7% (330/614) of the participants were in self-guided clusters, while 46.2% (284/614) were in clusters allocated to WhatsApp groups. WhatsApp groups averaged 35.5 (SD 6.03) caregivers per group. Half (4/8, 50%) of the WhatsApp groups experienced delayed initiation due to rolling recruitment and the December 2022 Christmas break, with 2 groups postponed until January 2023. These delays likely contributed to participant attrition from the groups. Facilitators reported moderate WhatsApp message participation, but live-chat engagement was low, with attendance ranging from 1 to 16 participants per session (mean 2, SD 2.65).

For the app design factor, 54.1% (332/614) of the participants were in clusters allocated to the structured design and 45.9% (282/614) were in clusters allocated to the unstructured design. Research assistants activated the assigned design on participants' phones based on their respective cluster. In 4 cases, the structured design was mistakenly activated; 3 of these participants did not respond to follow-ups and were treated as randomized in the analyses.

For the digital support factor, clusters assigned to basic support included 48% (295/614) of the participants, and enhanced support 52% (319/614). Given that participation was optional, attendance was not systematically recorded. Research assistants and fieldwork coordinators estimated that 67.1% (214/319) of the participants in enhanced support clusters attended training sessions. Nonattendance was primarily attributed to scheduling conflicts. Session duration varied, lasting anywhere from 5 to 20 minutes, depending on the group size and the number of questions asked.

Engagement

Table 4 summarizes the raw mean comparison of engagement outcomes for the entire sample and each factor level. Module completion rates among the 614 participants who successfully installed ParentApp were as follows: 38.4% (n=236) completed at least 25% (3/12) of the modules, 21.5% (n=132) completed at least 50% (6/12) of the modules, 13.8% (n=85) completed at least 75% (9/12) of the modules, and 8% (n=49) completed 100% (12/12) of the modules. Among the 600 participants who

XSL•FO
RenderX

engaged with some module content, the average content completion rate across all 12 modules was 35.57% (SD 33.10%). This rate varied by factor levels: *guidance* (self-guided: mean 32.8%, SD 31.34% vs guided: mean 38.7%, SD 34.79%), *app design* (structured: mean 34.6%, SD 31.73% vs unstructured: mean 36.7%, SD 34.65%), and *digital support* (basic: mean 34.3%, SD 32.48% vs enhanced: mean 36.8%, SD 33.69%). For participants who started at least 1 module, survival analysis estimated a median retention time of 31 days (95% CI 25-36) from enrollment to last synchronization with the cloud server.

Table 4. Mean comparison of primary and secondary engagement outcomes.

Outcome	Total	Factor A: guidance		Factor B: app design		Factor C: digital support	
		Self-guided (n=330)	Guided (n=284)	Structured (n=332)	Unstructured (n=282)	Basic (n=295)	Enhanced (n=319)
Primary engagement outcomes, mean (SD)							
App launches[a]	17.00 (22.59)	15.87 (20.93)	20.44 (24.18)	20.62 (26.81)	14.91 (15.86)	17.82 (23.67)	18.13 (21.56)
Modules completed[b]	3.40 (3.64)	3.09 (3.41)	3.77 (3.85)	2.91 (3.04)	3.99 (4.16)	3.26 (3.51)	3.54 (3.75)
Activity reviews[c]	1.62 (3.02)	1.39 (2.75)	1.88 (3.30)	0.49 (1.33)	2.95 (3.82)	1.48 (2.90)	1.74 (3.13)
Secondary engagement outcomes, mean (SD)							
Modules started[d]	4.57 (4.13)	4.22 (3.92)	4.99 (4.33)	4.34 (4.03)	4.84 (4.23)	4.43 (4.04)	4.71 (4.21)
Time spent in the app (min)[e]	54.99 (52.48)	48.79 (43.33)	62.22 (60.77)	51.46 (51.54)	59.19 (53.37)	53.39 (48.88)	56.49 (55.71)
Positive behavior logs[f]	38.96 (104.07)	30.70 (75.52)	48.56 (129.06)	32.35 (66.25)	46.74 (135.44)	32.67 (89.18)	44.78 (115.99)

[a]Total app launches during the intervention period.

[b]Number of completed modules (0 to 12).

[c]Number of home practice activities reviewed (0 to 10).

[d]Number of modules started (0 to 12).

[e]Total minutes spent in the app.

[f]Number of self-reported positive parenting and mental health–promoting behaviors logged.

Factor Effects on Primary and Secondary Outcomes

The main effect results for primary and secondary outcomes by factor are provided in Table 5. Following adjustments for baseline characteristics, caregivers who received WhatsApp group guidance in combination with ParentApp demonstrated significantly higher levels of engagement compared to those who were self-guided. Guided caregivers launched the app 2.93 times more (95% CI 1.84-4.68; *P*<.001), completed 1.29 times more modules (95% CI 1.05-1.58; *P*=.02), started 1.20 times more modules (95% CI 1.02-1.42; *P*=.03), spent 1.45 times more time in the app (95% CI 1.39-1.51; *P*<.001), and logged positive behaviors 2.73 times more frequently (95% CI 2.07-3.60; *P*<.001) than self-guided participants. Although guided caregivers initiated more home practice activity reviews compared to those who were self-guided, this difference was not statically significant (MR 2.17, 95% CI 0.84-5.63; *P*=.11).

The unstructured app design significantly outperformed the structured app design across all but 1 engagement outcome. App launch frequency did not differ significantly between designs (MR 0.88, 95% CI 0.70-1.11; *P*=.28). However, caregivers using the unstructured design completed 1.49 times more modules (95% CI 1.26-1.76; *P*<.001), initiated 7.49 times more home practice activity reviews (95% CI 5.19-10.82; *P*<.001), started 1.27 times more modules (95% CI 1.06-1.52; *P*=.01), and spent 1.84 times more time in the app (95% CI 1.70-1.99; *P*<.001) compared to those using the structured design. Notably, users of the unstructured design also logged 55.68 times more positive behaviors (95% CI 16.47-188.27; *P*<.001) compared to those using the structured design. A sensitivity analysis was conducted to address potential biases arising from extreme values in the positive behavior logs outcome (range 0-1342), where 10.4% (64/614) of the sample logged values exceeding 80 (deemed excessive in the study context). Models were rerun with positive behavior logs capped at 80. The results remained consistent with the original analyses (Table S2 of Multimedia Appendix 3).

XSL•FO
RenderX

Table 5. Main effects of experimental factors on engagement: unadjusted and adjusted results (N=614).

Factor level and outcomes	Unadjusted		Adjusted[a]	
	MR[b] (95% CI)	*P* value	MR (95% CI)	*P* value
Guided versus self-guided				
App launches	2.90 (1.85-4.54)	<.001	2.93 (1.84-4.68)	<.001
Modules completed	1.29 (1.06-1.57)	.01	1.29 (1.05-1.58)	.02
Activity reviews	2.23 (0.89-5.59)	.09	2.17 (0.84-5.63)	.11
Modules started	1.23 (1.04-1.46)	.02	1.20 (1.02-1.42)	.03
Time spent in the app	1.48 (1.44-1.52)	<.001	1.45 (1.39-1.51)	<.001
Positive behavior logs	2.65 (2.03-3.45)	<.001	2.73 (2.07-3.60)	<.001
Unstructured versus structured				
App launches	0.87 (0.69-1.10)	.25	0.88 (0.70-1.11)	.28
Modules completed	1.45 (1.23-1.72)	<.001	1.49 (1.26-1.76)	<.001
Activity reviews	6.84 (4.73-9.73)	<.001	7.49 (5.19-10.82)	<.001
Modules started	1.24 (1.02-1.51)	.03	1.27 (1.06-1.52)	.01
Time spent in the app	1.72 (1.67-1.77)	<.001	1.84 (1.70-1.99)	<.001
Positive behavior logs	53.48 (17.44-164.03)	<.001	55.68 (16.47-188.27)	<.001
Enhanced versus basic				
App launches	1.02 (0.78-1.32)	.91	1.01 (0.79-1.27)	.97
Modules completed	1.09 (0.88-1.36)	.42	1.09 (0.87-1.36)	.45
Activity reviews	1.15 (0.41-3.24)	.79	1.15 (0.39-3.34)	.80
Modules started	1.07 (0.88-1.29)	.51	1.06 (0.88-1.27)	.53
Time spent in the app	1.07 (0.87-1.30)	.53	1.07 (0.87-1.30)	.53
Positive behavior logs	1.39 (0.78-2.48)	.27	1.44 (0.82-2.54)	.21

[a]Adjusted models included the following covariates: caregiver age, gender, financial stress, food insecurity, parenting stress, caregiver depression, overall child maltreatment, and overall positive parenting. All covariates were sample-mean centered.

[b]MR: mean ratio: estimated effects calculated as MR = exp (2β), where β is the unstandardized regression coefficient derived from the effect-coded experimental factors.

For the digital support factor, although enhanced support showed higher average engagement levels than basic support, these differences were not statistically significant (app launches: MR 1.01, 95% CI 0.79-1.27; *P*=.97; modules completed: MR 1.09, 95% CI 0.87-1.36; *P*=.45; home practice activity reviews: MR 1.15, 95% CI 0.39-3.34; *P*=.80; modules started: MR 1.06, 95% CI 0.88-1.27; *P*=.53; time spent in the app: MR 1.07, 95% CI 0.87-1.30; *P*=.53; and positive behavior logs: MR 1.44, 95% CI 0.82-2.54; *P*=.21).

Factor Interaction Effects on Primary and Secondary Outcomes

Table 6 shows the main and 2-way interaction effects. Results for each outcome were derived from the same fitted model, adjusted for the same sample-mean centered covariates as the main effects analyses.

Although none of the factors showed significant main effects on app launches, a significant interaction between guidance and enhanced digital support (MR 1.69, 95% CI 1.06-2.71; *P*=.04) indicated that their combined effect positively influenced each other's impact on the number of times a participant launched the app. For modules completed, all 3 factors showed significant main effects. In addition, guidance and enhanced digital support showed a significant interaction (MR 1.17, 95% CI 1.07-1.28; *P*<.001), suggesting a mutual positive influence on the number of modules completed. In the model analyzing home practice activity reviews, guidance and app design factors had significant main effects, while digital support did not. A significant interaction between guidance and the unstructured app design (MR 0.77, 95% CI 0.63-0.92; *P*=.004) indicated a decrease in activity reviews, while a significant interaction between guidance and enhanced digital support (MR 1.38, 95% CI 1.21-1.58; *P*<.001) indicated an increase in activity review rates.

XSL•FO
RenderX

Table 6. Interaction effects of experimental factors on engagement: adjusted results (N=614).

Outcome	Main effect						Interaction effect					
	Factor A[a]		Factor B[b]		Factor C[c]		A×B		A×C		B×C	
	MR[d] (95% CI)	*P* value	MR (95% CI)	*P* value	MR (95% CI)	*P* value	MR (95% CI)	*P* value	MR (95% CI)	*P* value	MR (95% CI)	*P* value
App launches	1.53 (0.96-2.44)	.09	0.96 (0.54-1.71)	.89	1.01 (0.74-1.37)	.94	0.96 (0.72-1.26)	.75	1.69 (1.06-2.71)	.04	0.84 (0.47-1.50)	.56
Modules completed	1.22 (1.12-1.33)	<.001	1.42 (1.30-1.56)	<.001	1.09 (1.00-1.19)	.049	1.02 (0.94-1.12)	.59	1.17 (1.07-1.28)	<.001	1.09 (1.00-1.19)	.05
Activity reviews	1.63 (1.36-1.96)	<.001	7.06 (5.87-8.48)	<.001	1.15 (0.97-1.37)	.11	0.77 (0.64-0.92)	.004	1.38 (1.21-1.58)	<.001	1.07 (0.90-1.28)	.43
Modules started	1.17 (1.05-1.30)	.004	1.20 (1.07-1.34)	.001	1.07 (0.96-1.19)	.23	0.97 (0.88-1.08)	.63	1.24 (1.11-1.38)	<.001	1.10 (0.99-1.22)	.09
Time spent in the app	1.59 (0.97-2.62)	.07	2.08 (1.37-3.16)	.001	1.07 (0.56-2.03)	.85	1.48 (0.98-2.24)	.06	1.36 (0.82-2.23)	.23	0.36 (0.31-0.42)	<.001
Positive behavior logs	1.32 (0.58-2.99)	.51	2.37 (1.13-4.98)	.02	1.10 (0.41-2.96)	.85	1.10 (0.55-2.20)	.78	1.13 (0.50-2.58)	.77	1.24 (0.58-2.65)	.58

[a]WhatsApp group guided versus self-guided.

[b]Unstructured app design versus structured app design.

[c]Enhanced digital support versus basic digital support.

[d]MR: mean ratio: estimated effects calculated as MR = exp (2β), where β is the unstandardized regression coefficients derived from the effect-coded experimental factors.

Among secondary outcomes, the joint influence of guidance and enhanced digital support on modules started was significant (MR 1.24, 95% CI 1.11-1.38; *P*<.001), despite the digital support factor showing no significant main effect. Similarly, the unstructured design and enhanced digital support showed significant interaction effects for the time spent in the app, despite neither factor individually showing significance. However, their combined effects were attenuated (MR 0.36, 95% CI 0.31-0.42; *P*<.001), suggesting that their synergy led to a reduction in participants' app use time.

Discussion

Principal Findings

This is the first known factorial trial of a digital parenting intervention for families with adolescents in an LMIC. It follows MOST principles [54] to optimize engagement with an app targeting socioeconomically disadvantaged families in Tanzania. Findings offer important insights into engagement-enhancing intervention design and implementation strategies for digital parenting interventions tailored to the unique contexts and needs of LMIC settings.

Receiving guidance through moderated WhatsApp groups significantly increased participant engagement across a wide range of engagement metrics. Compared to self-guided participants, those in WhatsApp groups launched the app more frequently, started and completed more program modules, spent more time in the app, and logged more positive parenting and mental health behaviors. These findings align with meta-analytic research on mental health apps and the broader digital health literature, which consistently shows that interventions incorporating some form of human guidance yield higher engagement rates compared to self-guided interventions [44,46,48-50]. However, our study advances this literature in 2 distinct ways. First, we directly compared self-guided and guided conditions within the same trial, whereas previous research has largely relied on cross-study comparisons [48-50]. Second, we implemented a group-based guidance model rather than individual approaches (eg, one-on-one phone consultations, videoconferencing, or email) commonly used in previous studies [26]. This shift toward a less resource-intensive delivery model enhances potential for future scalability.

The mechanisms through which human guidance facilitates engagement with digital interventions are not fully understood. One possible explanation is the role of the relationship quality between facilitators and participants [75]. Alternatively, guided participants may feel an increased sense of accountability in the presence of a facilitator [76]. The group-based format may also be important, as it enables interaction among caregivers. For instance, this setting allows participants with shared experiences to offer emotional and informational support [77] potentially enhancing engagement through social connection and community building. Indeed, a recent systematic review on peer-based features in digital interventions highlighted the importance of asynchronous communication features in

XSL•FO
RenderX

supporting program engagement [51]. In addition to peer support benefits, findings from a systematic review of digital mental health interventions suggest that peer-to-peer groups are particularly effective in promoting engagement and retention when moderated by trained professionals [78]. Nevertheless, contact with facilitators or fellow participants may not be universally desirable or beneficial, as evidenced by participants who left WhatsApp groups prematurely in this study. These findings align with a systematic review showing that synchronous communication features, such as online video calls or chat rooms, were associated with reduced engagement [51]. Understanding which participants benefit from group-based support remains a critical direction for future research.

The unstructured app design, which featured flexible content access and culturally adapted, humanlike illustrations, significantly increased engagement compared to the structured design. Participants using the unstructured design completed more program modules, initiated more home practice reviews, started more modules, spent more time in the app, and logged more positive behaviors compared to those using the structured design. These findings align with research showing that user experience design, including app navigation and appearance, influences motivation and engagement with digital interventions [45,63]. However, as this study evaluated 2 app design packages rather than distinct components, it remains unclear which specific aspect of the unstructured design contributed to enhanced engagement. Future research would benefit from more granular testing and analysis of these design components to determine their individual impact. Enhanced engagement might be attributed to the flexible content progression, which allowed participants to move through the program at their preferred pace. Alternatively, increased engagement could reflect the greater relatability of culturally adapted images, which may have improved app acceptability among caregivers [26] and consequently boosted motivation to engage. While developing a universal app with content tailored for low-income contexts aimed to enhance scalability and reduce adaptation costs, our findings suggest that scalability alone does not ensure acceptability. Culturally adapted visual elements may therefore be crucial in fostering engagement with digital interventions.

A notable finding was the significant increase in home practice activity reviews among participants using the unstructured design. While the reason for this increase remains unclear, it could be attributed to the modular format of the design, which allowed participants to access home practice reviews without waiting for the 7-day cycle to be completed. Given that home practice of program skills is a key mechanism linking program implementation to behavioral outcomes in parenting interventions [79], including home practice engagement as an outcome is a study strength. However, the study does not establish a direct link between engaging in home practice reviews within the app and the actual application of these skills at home. Furthermore, this study only measured when participants initiated home practice reviews in the app, without capturing related measures such as self-reports of activity completion.

A surprising finding was the limited effectiveness of the digital support factor on engagement, which tended to have MRs close to 1 across most engagement outcomes. This suggests that augmenting app-specific assistance with general smartphone support may not yield additional benefits for participants. Unlike recent systematic reviews which assessed baseline digital literacy and its association with engagement [44,48], our study experimentally examined the provision of preprogram digital support. This was an important implementation consideration given our target population's relatively low exposure to digital technology. While systematic reviews indicate that app literacy, defined as technological competency in using a smartphone app, is important for both uptake and engagement [46,80], it is plausible that the basic digital support module in our study adequately contributed to participants' competency and perceived digital skills. Another potential reason could be that the enhanced digital training may have inadvertently reduced ParentApp engagement by introducing participants to competing apps and features, particularly among those with limited previous digital experience.

The impact of digital support on engagement might also have been influenced by various individual factors. The relatively young age of the sample (mean 35.94, SD 11.84 years) and their previous smartphone ownership may suggest sufficient baseline digital literacy. Moderator analyses could reveal whether factors such as caregiver age and other demographic characteristics influenced the results. If digital support proves effective only for specific participants, tailoring the type and intensity of support to individuals would be advantageous. Furthermore, it is essential to highlight that both digital support trainings were optional. The fieldwork team reported that while attendance was high for the basic training, some participants allocated to the enhanced training left the onboarding session early to resume economic and caregiving responsibilities. This reduced exposure could have attenuated the effects of the enhanced training; consideration is thus warranted when interpreting the results.

Interaction analyses found evidence that combining enhanced digital support with guidance led to increased engagement across 4 of the 6 measured outcomes. These interaction effects suggest the importance of smartphone literacy in promoting engagement, although enhanced digital support training alone may not be sufficient to sustain engagement. The combination of enhanced digital support and the unstructured design appeared to reduce app use time, possibly because increased navigation confidence paired with a simplified design allowed more efficient use. Further research is needed to help explain these findings.

Limitations

Several limitations should be considered when interpreting the results. First, as this trial focused on engagement outcomes, we did not assess how the design and implementation factors influenced caregiver or adolescent outcomes. However, this trial serves as the optimization phase of the MOST framework within the broader ParentApp study, laying the groundwork for a comprehensive evaluation of the app's effectiveness in the forthcoming randomized controlled trial [81]. Second, while the study's focus on broader implementation strategies provides valuable insights for implementers and researchers, the generalizability of results to different LMIC settings or populations remains unclear. Further investigation into the

XSL•FO
RenderX

factors influencing uptake and engagement across diverse behavioral domains, digital platforms, and delivery contexts is needed.

Third, data collection via automated tracking required internet access for synchronization with the cloud servers. As a result, participant engagement data may have been incomplete for those whose 1 GB monthly internet bundles depleted before data synchronization occurred, potentially leading to an underestimation of engagement rates. Efforts to address this included sending participants an SMS text message reminder toward the end of the study, along with an additional data bundle. However, it remains unclear to what extent these measures resolved the issue. Fourth, technical challenges hindered app installation and use. Despite enrolling 680 caregivers, surpassing the desired sample size of 640, installation difficulties affected 66 (9.7%) participants, and 14 (2.1%) did not to access any core intervention content after installation. These challenges stemmed primarily from internet and phone-to-app compatibility issues. Forthcoming qualitative research conducted in parallel to this study highlighted additional barriers to uptake and retention, including the selling of phones due to financial needs, deleting the app due to storage constraints, and program dropout due to severely damaged screens. Despite these challenges, overall content completion rates for participants who started at least 1 module was 35.57% of the program. While only 8% (49/614) of participants completed all 12 modules, completion rates were consistent with other digital parenting intervention studies in HICs, such as the 7.5% completion rate found in a large-scale study of the web-based ParentWorks program in Australia [39]. Nonetheless, investigating patterns of early dropout and identifying opportunities for keeping caregivers engaged throughout program delivery remain crucial areas for investigation.

Strengths

This study has several notable strengths. A key strength is the application of a randomized factorial experiment to identify the combination of candidate factor levels that, while remaining feasible and acceptable, most effectively promoted engagement with ParentApp. Findings from this experiment played a pivotal role in selecting which factor levels to include in the forthcoming randomized controlled trial in Tanzania [81]. Following MOST framework decision-making guidelines [82], significant main effects guided the selection of effective factor levels, while lower-dose factor levels were retained when effects were null. Cost-effectiveness analysis and qualitative feedback then helped assess feasibility and acceptability of retained factor levels for the optimized intervention package. In addition, the findings from this experiment have contributed to the development and refinement of other digital parenting interventions, including in South Africa [83] and Malaysia [84].

Another key strength is the study context. Recruitment through local gatekeepers and community leaders yielded a large community sample in an LMIC. This approach aligns with how the intervention could be disseminated in the future and enhances the sustainability of the program. The sample also included 33.4% (205/614) men, a demographic typically underrepresented in parenting interventions [74]. This high proportion of male caregivers not only extends the generalizability of the results but also represents significant strides toward understanding how to engage underrepresented groups in digital parenting interventions in LMIC contexts. A forthcoming subgroup analysis of engagement predictors will provide further insights into this aspect of engagement.

Finally, unlike many digital intervention trials that provided participants with digital devices, this study's sample consisted of participants who used their own smartphones, including older and low-cost models. This approach demonstrates the feasibility of implementing digital interventions in real-world, low-income settings and highlights the potential for cost-effective scale-up. Furthermore, by including participants with varying levels of technological access and proficiency, this study establishes a model for inclusive digital intervention design, particularly relevant in regions such as Tanzania, where smartphone access is expanding but still limited [85].

Conclusions

While digital interventions hold immense promise for addressing health and service needs in LMICs, their effectiveness critically depends on ensuring meaningful engagement. This trial is the first of its kind to use a randomized factorial experiment to optimize participant engagement with an app-based parenting intervention in an LMIC. Moreover, this study adds to the paucity of research addressing the need for effective and accessible violence prevention interventions for adolescents in LMICs. While further investigation is needed to understand the impact of experimental factors on parenting and adolescent outcomes, our findings demonstrate that facilitator-moderated WhatsApp group guidance; an unstructured, culturally adapted app design; and preprogram digital support can enhance engagement with app-based parenting interventions. Continued exploration and refinement of these approaches is needed for developing evidence-based, scalable digital interventions capable of addressing parenting challenges and fostering positive outcomes for families in resource-constrained contexts.

Acknowledgments

This study was funded by the UK Research and Innovation Global Challenges Research Fund Accelerating Achievement for Africa's Adolescents (Accelerate) Hub (ES/S008101/1), the Wellspring Philanthropic Foundation (16204), the European Research Council under the Proof of Concept 2022 Call (101067451), and a private family trust which wishes to remain anonymous. This study is also part of the Global Parenting Initiative, which is funded by the LEGO Foundation and Oak Foundation. RJ was supported by the UK Research and Innovation Economic and Social Research Council Grand Union Doctoral Training Partnership. YS was supported by the Centre for Development, Evaluation, Complexity, and Implementation in Public Health (DECIPHer) and the Wolfson Centre for Young People's Mental Health. DECIPHer is funded by the Welsh government through Health and

XSL•FO
RenderX

Care Research, Wales. The Wolfson Centre for Young People's Mental Health has been established with support from the Wolfson Foundation

The authors would like to thank all members of the ParentApp Tanzania Optimization Trial team. Innovations in Development, Education and the Mathematical Sciences provided technical leadership and conceptual partnership throughout the development of the intervention, including app design and coding; building digital data collection and monitoring systems; and technical support during study implementation. The Tanzanian National Institute for Medical Research fieldwork team conducted community mapping, recruitment, and participant onboarding. Facilitators from Investing in Children and Strengthening Their Societies conducted onboarding sessions and led the delivery of the WhatsApp groups. Clowns Without Borders South Africa conducted local adaptation and contextualization of content, curriculum development, and facilitator training. The authors would also like to thank Vodacom Tanzania for providing low-cost data bundles for participants. The authors would especially like to acknowledge the many families whose participation and input were instrumental in developing ParentApp.

Data Availability

The datasets generated or analyzed during this study are not currently publicly available due to ongoing analyses but will be made available via the Open Science Framework upon study completion. In the interim, datasets are available from the corresponding author on reasonable request.

Authors' Contributions

RJ led the study conceptualization and design, with contributions from LDC, JW, DS, L Clements, LB, ETW, AB, and JML. Trial implementation and data collection were managed by RJ, with project administration support from JW, MW, AO, ANC, L Christine, AM, and GM. DS, L Clements, and ETW led app development and app-generated data management. RJ conducted statistical analyses with guidance from GJM-T and L Clements. RJ drafted the manuscript, which was reviewed and edited by LDC, YS, L Clements, GJM-T, LB, ANC, JK, ETW, and JML. All authors approved the submitted manuscript.

Conflicts of Interest

ParentApp was developed and implemented by members of the research team. RJ's doctoral research is based on this trial. JML is the chief executive officer, and LDC is a voluntary trustee of Parenting for Lifelong Health (PLH), a charity based in the United Kingdom and South Africa. JML receives occasional fees for providing training and supervision for PLH programs. Members of the research team have worked on other studies of PLH, which contributed funding to their respective universities but not directly to them. The intellectual property of the intervention is owned by the developers of the original intervention under a Creative Commons Attribution-NoDerivatives and NonCommercial 4.0 International Public License. No profit or financial gain will be made from the implementation and dissemination of the intervention.

Multimedia Appendix 1
TIDieR (Template for Intervention Description and Replication) checklist.
[PDF File (Adobe PDF File), 232 KB - jmir_v27i1e60102_app1.pdf]

Multimedia Appendix 2
TIDieR (Template for Intervention Description and Replication) description.
[PDF File (Adobe PDF File), 176 KB - jmir_v27i1e60102_app2.pdf]

Multimedia Appendix 3
Supplementary analyses.
[PDF File (Adobe PDF File), 130 KB - jmir_v27i1e60102_app3.pdf]

Multimedia Appendix 4
Experimental conditions with effect coding.
[PDF File (Adobe PDF File), 90 KB - jmir_v27i1e60102_app4.pdf]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1307 KB - jmir_v27i1e60102_app5.pdf]

References

XSL•FO
RenderX

1. Hughes K, Bellis MA, Hardcastle KA, Sethi D, Butchart A, Mikton C, et al. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. Lancet Public Health 2017;2(8):e356-e366 [FREE Full text] [doi: 10.1016/S2468-2667(17)30118-4] [Medline: 29253477]
2. Fry D, Fang X, Elliott S, Casey T, Zheng X, Li J, et al. The relationships between violence in childhood and educational outcomes: a global systematic review and meta-analysis. Child Abuse Negl 2018;75:6-28. [doi: 10.1016/j.chiabu.2017.06.021] [Medline: 28711191]
3. Manchikanti Gómez A. Testing the cycle of violence hypothesis: child abuse and adolescent dating violence as predictors of intimate partner violence in young adulthood. Youth Soc 2010;43(1):171-192. [doi: 10.1177/0044118X09358313]
4. Norman RE, Byambaa M, De R, Butchart A, Scott J, Vos T. The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. PLoS Med 2012;9(11):e1001349 [FREE Full text] [doi: 10.1371/journal.pmed.1001349] [Medline: 23209385]
5. Roberts AL, McLaughlin KA, Conron KJ, Koenen KC. Adulthood stressors, history of childhood adversity, and risk of perpetration of intimate partner violence. Am J Prev Med 2011;40(2):128-138 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.016] [Medline: 21238860]
6. Violence against children. United Nations Children's Fund. 2020. URL: https://www.unicef.org/protection/violence-against-children [accessed 2023-08-08]
7. WHO guidelines on parenting interventions to prevent maltreatment and enhance parent–child relationships with children aged 0–17 years. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240065505 [accessed 2024-04-29]
8. Hillis S, Mercy J, Amobi A, Kress H. Global prevalence of past-year violence against children: a systematic review and minimum estimates. Pediatrics 2016;137(3):e20154079 [FREE Full text] [doi: 10.1542/peds.2015-4079] [Medline: 26810785]
9. Ending violence in childhood: global report 2017. Know Violence in Childhood. 2017. URL: https://resourcecentre.savethechildren.net/pdf/global_report_2017_ending_violence_in_childhood.pdf/ [accessed 2024-04-28]
10. Violence against children in Tanzania: findings from a national survey 2009. United Nations Children's Fund. 2011. URL: https://reliefweb.int/report/united-republic-tanzania/violence-against-children-tanzania-findings-national-survey-2009 [accessed 2024-04-29]
11. Nkuba M, Hermenau K, Hecker T. Violence and maltreatment in Tanzanian families - findings from a nationally representative sample of secondary school students and their parents. Child Abuse Negl 2018;77:110-120. [doi: 10.1016/j.chiabu.2018.01.002] [Medline: 29324272]
12. INSPIRE: seven strategies for ending violence against children. World Health Organization. 2016. URL: https://iris.who.int/bitstream/handle/10665/207717/9789241565356-eng.pdf?sequence=1 [accessed 2024-04-29]
13. Backhaus S, Gardner F, Melendez-Torres GJ, Schafer M, Knerr W, Lachman JM. World Health Organization guidelines on parenting interventions to prevent maltreatment and enhance parent–child relationships with children aged 0-17 years: report of the systematic reviews of evidence. World Health Organization. 2023. URL: https://cdn.who.int/media/docs/default-source/documents/violence-prevention/systematic_reviews-for-the-who-parenting-guideline-jan-27th-2023.pdf?sfvrsn=158fd424_3 [accessed 2023-05-30]
14. Melendez‐Torres GJ, Leijten P, Gardner F. What are the optimal combinations of parenting intervention components to reduce physical child abuse recurrence? Reanalysis of a systematic review using qualitative comparative analysis. Child Abuse Review 2019;28(3):181-197 [FREE Full text] [doi: 10.1002/car.2561]
15. Gubbels J, van der Put CE, Stams GJ, Prinzie PJ, Assink M. Components associated with the effect of home visiting programs on child maltreatment: a meta-analytic review. Child Abuse Negl 2021;114:104981 [FREE Full text] [doi: 10.1016/j.chiabu.2021.104981] [Medline: 33571741]
16. Leijten P, Gardner F, Melendez-Torres GJ, van Aar J, Hutchings J, Schulz S, et al. Meta-analyses: key parenting program components for disruptive child behavior. J Am Acad Child Adolesc Psychiatry 2019;58(2):180-190 [FREE Full text] [doi: 10.1016/j.jaac.2018.07.900] [Medline: 30738545]
17. Mejia A, Calam R, Sanders MR. A review of parenting programs in developing countries: opportunities and challenges for preventing emotional and behavioral difficulties in children. Clin Child Fam Psychol Rev 2012;15(2):163-175. [doi: 10.1007/s10567-012-0116-9] [Medline: 22427004]
18. Weisenmuller C, Hilton D. Barriers to access, implementation, and utilization of parenting interventions: considerations for research and clinical applications. Am Psychol 2021;76(1):104-115. [doi: 10.1037/amp0000613] [Medline: 32134281]
19. Knerr W, Gardner F, Cluver L. Improving positive parenting skills and reducing harsh and abusive parenting in low- and middle-income countries: a systematic review. Prev Sci 2013;14(4):352-363. [doi: 10.1007/s11121-012-0314-1] [Medline: 23315023]
20. Parra-Cardona R, Leijten P, Lachman JM, Mejía A, Baumann AA, Amador Buenabad NG, et al. Strengthening a culture of prevention in low- and middle-income countries: balancing scientific expectations and contextual realities. Prev Sci 2021;22(1):7-17 [FREE Full text] [doi: 10.1007/s11121-018-0935-0] [Medline: 30058025]
21. Digital transformation drives development in Africa. World Bank Group. 2024. URL: https://www.worldbank.org/en/results/2024/01/18/digital-transformation-drives-development-in-afe-afw-africa [accessed 2024-04-29]

XSL•FO
RenderX

22. Bervell B, Al-Samarraie H. A comparative review of mobile health and electronic health utilization in sub-Saharan African countries. Soc Sci Med 2019;232:1-16. [doi: 10.1016/j.socscimed.2019.04.024] [Medline: 31035241]
23. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
24. Fu Z, Burger H, Arjadi R, Bockting CL. Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Psychiatry 2020;7(10):851-864. [doi: 10.1016/s2215-0366(20)30256-x]
25. Baumel A, Pawar A, Kane JM, Correll CU. Digital parent training for children with disruptive behaviors: systematic review and meta-analysis of randomized trials. J Child Adolesc Psychopharmacol 2016;26(8):740-749. [doi: 10.1089/cap.2016.0048] [Medline: 27286325]
26. Corralejo SM, Domenech Rodríguez MM. Technology in parenting programs: a systematic review of existing interventions. J Child Fam Stud 2018;27(9):2717-2731. [doi: 10.1007/S10826-018-1117-1]
27. Florean IS, Dobrean A, Păsărelu CR, Georgescu RD, Milea I. The efficacy of internet-based parenting programs for children and adolescents with behavior problems: a meta-analysis of randomized clinical trials. Clin Child Fam Psychol Rev 2020;23(4):510-528. [doi: 10.1007/s10567-020-00326-0] [Medline: 32897527]
28. Flujas-Contreras JM, García-Palacios A, Gómez I. Technology-based parenting interventions for children's physical and psychological health: a systematic review and meta-analysis. Psychol Med 2019;49(11):1787-1798. [doi: 10.1017/S0033291719000692] [Medline: 30977462]
29. Hansen A, Broomfield G, Yap MB. A systematic review of technology‐assisted parenting programs for mental health problems in youth aged 0–18 years: applicability to underserved Australian communities. Aust J Psychol 2020;71(4):433-462. [doi: 10.1111/ajpy.12250]
30. Spencer CM, Topham GL, King EL. Do online parenting programs create change?: A meta-analysis. J Fam Psychol 2020;34(3):364-374. [doi: 10.1037/fam0000605] [Medline: 31697102]
31. Thongseiratch T, Leijten P, Melendez-Torres GJ. Online parent programs for children's behavioral problems: a meta-analytic review. Eur Child Adolesc Psychiatry 2020;29(11):1555-1568. [doi: 10.1007/s00787-020-01472-0] [Medline: 31925545]
32. Haine-Schlagel R, Dickson KS, Lind T, Kim JJ, May GC, Walsh NE, et al. Caregiver participation engagement in child mental health prevention programs: a systematic review. Prev Sci 2022;23(2):321-339 [FREE Full text] [doi: 10.1007/s11121-021-01303-x] [Medline: 34936045]
33. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
34. Sim WH, Jorm AF, Yap MB. The role of parent engagement in a web-based preventive parenting intervention for child mental health in predicting parenting, parent and child outcomes. Int J Environ Res Public Health 2022;19(4):2191 [FREE Full text] [doi: 10.3390/ijerph19042191] [Medline: 35206394]
35. Hall CM, Bierman KL. Technology-assisted interventions for parents of young children: emerging practices, current research, and future directions. Early Child Res Q 2015;33:21-32 [FREE Full text] [doi: 10.1016/j.ecresq.2015.05.003] [Medline: 27773964]
36. MacDonell KW, Prinz RJ. A review of technology-based youth and family-focused interventions. Clin Child Fam Psychol Rev 2017;20(2):185-200 [FREE Full text] [doi: 10.1007/s10567-016-0218-x] [Medline: 27787701]
37. Brager J, Breitenstein SM, Miller H, Gross D. Low-income parents' perceptions of and engagement with a digital behavioral parent training program: a mixed-methods study. J Am Psychiatr Nurses Assoc 2021;27(1):33-43. [doi: 10.1177/1078390319872534] [Medline: 31509052]
38. Breitenstein SM, Brager J, Ocampo EV, Fogg L. Engagement and adherence with ezPARENT, an mHealth parent-training program promoting child well-being. Child Maltreat 2017;22(4):295-304. [doi: 10.1177/1077559517725402] [Medline: 28870112]
39. Dadds MR, Sicouri G, Piotrowska PJ, Collins DA, Hawes DJ, Moul C, et al. Keeping parents involved: predicting attrition in a self-directed, online program for childhood conduct problems. J Clin Child Adolesc Psychol 2019;48(6):881-893 [FREE Full text] [doi: 10.1080/15374416.2018.1485109] [Medline: 30067388]
40. Baker S, Sanders MR. Predictors of program use and child and parent outcomes of a brief online parenting intervention. Child Psychiatry Hum Dev 2017;48(5):807-817. [doi: 10.1007/s10578-016-0706-8] [Medline: 28035556]
41. Fossum S, Ristkari T, Cunningham C, McGrath PJ, Suominen A, Huttunen J, et al. Parental and child factors associated with participation in a randomised control trial of an internet-assisted parent training programme. Child Adolesc Ment Health 2018;23(2):71-77. [doi: 10.1111/camh.12193] [Medline: 32677336]
42. Perrino T, Estrada Y, Huang S, St George S, Pantin H, Cano M, et al. Predictors of participation in an eHealth, family-based preventive intervention for Hispanic youth. Prev Sci 2018;19(5):630-641. [doi: 10.1007/s11121-016-0711-y] [Medline: 27704326]

XSL•FO
RenderX

43. Aldridge G, Tomaselli A, Nowell C, Reupert A, Jorm A, Yap MB. Engaging parents in technology-assisted interventions for childhood adversity: systematic review. J Med Internet Res 2024;26:e43994 [FREE Full text] [doi: 10.2196/43994] [Medline: 38241066]
44. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
45. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-Health: a scoping review. Telemed J E Health 2022;28(6):768-780 [FREE Full text] [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
46. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
47. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
48. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
49. Leung C, Pei J, Hudec K, Shams F, Munthali R, Vigo D. The effects of nonclinician guidance on effectiveness and process outcomes in digital mental health interventions: systematic review and meta-analysis. J Med Internet Res 2022;24(6):e36004 [FREE Full text] [doi: 10.2196/36004] [Medline: 35511463]
50. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interventions 2014;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
51. Elaheebocus SM, Weal M, Morrison L, Yardley L. Peer-based social media features in behavior change interventions: systematic review. J Med Internet Res 2018;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
52. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
53. Asher CA, Scherer E, Kim JS. Using a factorial design to maximize the effectiveness of a parental text messaging intervention. J Res Educ Eff 2022;15(3):532-557. [doi: 10.1080/19345747.2021.2009073]
54. Collins LM. Optimization of behavioral, biobehavioral, and biomedical interventions: the Multiphase Optimization Strategy (MOST). In: Collins LM, editor. Statistics for Social and Behavioral Sciences. Cham, Switzerland: Springer; 2018.
55. Janowski R, Green O, Shenderovich Y, Stern D, Clements L, Wamoyi J, et al. Optimising engagement in a digital parenting intervention to prevent violence against adolescents in Tanzania: protocol for a cluster randomised factorial trial. BMC Public Health 2023;23(1):1224 [FREE Full text] [doi: 10.1186/s12889-023-15989-x] [Medline: 37353844]
56. Kahan BC, Hall SS, Beller EM, Birchenall M, Chan A, Elbourne D, et al. Reporting of factorial randomized trials: extension of the CONSORT 2010 statement. JAMA 2023;330(21):2106-2114. [doi: 10.1001/jama.2023.19793] [Medline: 38051324]
57. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
58. Cluver LD, Meinck F, Steinert JI, Shenderovich Y, Doubt J, Herrero Romero R, et al. Parenting for Lifelong Health: a pragmatic cluster randomised controlled trial of a non-commercialised parenting programme for adolescents and their families in South Africa. BMJ Glob Health 2018;3(1):e000539 [FREE Full text] [doi: 10.1136/bmjgh-2017-000539] [Medline: 29564157]
59. Shenderovich Y, Ward CL, Lachman JM, Wessels I, Sacolo-Gwebu H, Okop K, et al. Evaluating the dissemination and scale-up of two evidence-based parenting interventions to reduce violence against children: study protocol. Implement Sci Commun 2020;1(1):109 [FREE Full text] [doi: 10.1186/s43058-020-00086-6] [Medline: 38624613]
60. Lachman J, Wamoyi J, Martin M, Han Q, Calderón Alfaro FA, Mgunga S, et al. Reducing family and school-based violence at scale: a large-scale pre-post study of a parenting programme delivered to families with adolescent girls in Tanzania. BMJ Glob Health 2024;9(11):e015472 [FREE Full text] [doi: 10.1136/bmjgh-2024-015472] [Medline: 39581633]
61. Awah I, Green O, Baerecke L, Janowski R, Klapwijk J, Chetty AN, et al. 'It provides practical tips, practical solutions!': acceptability, usability, and satisfaction of a digital parenting intervention across African countries. Psychol Health Med 2022;27(sup1):107-123 [FREE Full text] [doi: 10.1080/13548506.2022.2113106] [Medline: 35980251]
62. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]
63. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

XSL•FO
RenderX

64. Lown JM. Development and validation of a financial self-efficacy scale. J Financ Couns Plan 2011;22(2):54-63 [FREE Full text]
65. Meinck F, Boyes ME, Cluver L, Ward CL, Schmidt P, DeStone S, et al. Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. Child Abuse Negl 2018;82:45-58 [FREE Full text] [doi: 10.1016/j.chiabu.2018.05.022] [Medline: 29860107]
66. Essau CA, Sasagawa S, Frick PJ. Psychometric properties of the Alabama Parenting Questionnaire. J Child Fam Stud 2006;15(5):595-614. [doi: 10.1007/s10826-006-9036-y]
67. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). Arch Intern Med 1999;159(15):1701-1704. [doi: 10.1001/archinte.159.15.1701] [Medline: 10448771]
68. Berry JO, Jones WH. The parental stress scale: initial psychometric evidence. J Soc Pers Relat 1995;12(3):463-472. [doi: 10.1177/0265407595123009]
69. Kugler KC, Dziak JJ, Trail J. Coding and interpretation of effects in analysis of data from a factorial experiment. In: Collins LM, Kugler KC, editors. Optimization of behavioral, biobehavioral, and biomedical interventions. Cham, Switzerland: Springer; 2018.
70. Van Buuren S, Brand JP, Groothuis-Oudshoorn CG, Rubin DB. Fully conditional specification in multivariate imputation. J Stat Comput Simul 2006;76(12):1049-1064. [doi: 10.1080/10629360600810434]
71. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
72. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
73. Muralidharan K, Romero M, Wüthrich K. Factorial designs, model selection, and (incorrect) inference in randomized experiments. Rev Econ Stat 2023:1-44 [FREE Full text] [doi: 10.1162/rest_a_01317]
74. Panter-Brick C, Burgess A, Eggerman M, McAllister F, Pruett K, Leckman JF. Practitioner review: Engaging fathers--recommendations for a game change in parenting interventions based on a systematic review of the global evidence. J Child Psychol Psychiatry 2014;55(11):1187-1212 [FREE Full text] [doi: 10.1111/jcpp.12280] [Medline: 24980187]
75. Probst Greta H, Berger T, Flückiger C. Die Allianz als Prädiktor für den Therapieerfolg internetbasierter Interventionen bei psychischen Störungen: Eine korrelative Metaanalyse. Verhaltenstherapie 2019;29(3):182-195. [doi: 10.1159/000501565]
76. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
77. Dennis CL. Peer support within a health care context: a concept analysis. Int J Nurs Stud 2003 Mar;40(3):321-332. [doi: 10.1016/S0020-7489(02)00092-5]
78. Biagianti B, Quraishi SH, Schlosser DA. Potential benefits of incorporating peer-to-peer interactions into digital interventions for psychotic disorders: a systematic review. Psychiatr Serv 2018 Apr 01;69(4):377-388 [FREE Full text] [doi: 10.1176/appi.ps.201700283] [Medline: 29241435]
79. Berkel C, Sandler IN, Wolchik SA, Brown CH, Gallo CG, Chiapa A, et al. "Home practice is the program": parents' practice of program skills as predictors of outcomes in the new beginnings program effectiveness trial. Prev Sci 2018;19(5):663-673 [FREE Full text] [doi: 10.1007/s11121-016-0738-0] [Medline: 27933424]
80. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, et al. Mapping factors that affect the uptake of digital therapeutics within health systems: scoping review. J Med Internet Res 2023;25:e48000 [FREE Full text] [doi: 10.2196/48000] [Medline: 37490322]
81. Baerecke L, Ornellas A, Wamoyi J, Wambura M, Klapwijk J, Chetty AN, et al. A hybrid digital parenting programme to prevent abuse of adolescents in Tanzania: study protocol for a pragmatic cluster-randomised controlled trial. Trials 2024;25(1):119 [FREE Full text] [doi: 10.1186/s13063-023-07893-x] [Medline: 38351094]
82. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. Transl Behav Med 2014;4(3):238-251 [FREE Full text] [doi: 10.1007/s13142-013-0239-7] [Medline: 25264464]
83. Ambrosio MD, Lachman JM, Zinzer P, Gwebu H, Vyas S, Vallance I, et al. A factorial randomized controlled trial to optimize user engagement with a chatbot-led parenting intervention: protocol for the ParentText optimisation trial. JMIR Res Protoc 2024;13:e52145 [FREE Full text] [doi: 10.2196/52145] [Medline: 38700935]
84. Cooper H, Nadzri FZ, Vyas S, Juhari R, Ismail N, Arshat Z, et al. A hybrid digital parenting program delivered within the Malaysian preschool system: protocol for a feasibility study of a small-scale factorial cluster randomized trial. JMIR Res Protoc 2024;13:e55491 [FREE Full text] [doi: 10.2196/55491] [Medline: 38669679]
85. Takwa E. Govt reaffirms commitment to increase internet users. Tanzania Standard Newspapers Limited. 2022. URL: https://dailynews.co.tz/govt-reaffirms-commitment-to-increase-internet-users/ [accessed 2024-04-24]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

XSL•FO
RenderX

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
HIC: high-income country
ICS: Investing in Children and Strengthening their Societies
LMIC: low- and middle-income country
MOST: Multiphase Optimization Strategy
MR: mean ratio
NGO: nongovernmental organization
NIMR: National Institute for Medical Research
PLH: Parenting for Lifelong Health
TIDieR: Template for Intervention Description and Replication
VAC: violence against children

Edited by A Mavragani; submitted 01.05.24; peer-reviewed by C Emezue, K Bilali; comments to author 31.10.24; revised version received 03.01.25; accepted 11.01.25; published 10.03.25.

Please cite as:
Janowski R, Cluver LD, Shenderovich Y, Wamoyi J, Wambura M, Stern D, Clements L, Melendez-Torres GJ, Baerecke L, Ornellas A, Chetty AN, Klapwijk J, Christine L, Mukabana A, Te Winkel E, Booij A, Mbosoli G, Lachman JM
Optimizing Engagement With a Smartphone App to Prevent Violence Against Adolescents: Results From a Cluster Randomized Factorial Trial in Tanzania
J Med Internet Res 2025;27:e60102
URL: https://www.jmir.org/2025/1/e60102
doi:10.2196/60102
PMID:40063069

©Roselinde Janowski, Lucie D Cluver, Yulia Shenderovich, Joyce Wamoyi, Mwita Wambura, David Stern, Lily Clements, G J Melendez-Torres, Lauren Baerecke, Abigail Ornellas, Angelique Nicole Chetty, Jonathan Klapwijk, Laetitia Christine, Ateamate Mukabana, Esmee Te Winkel, Anna Booij, Gervas Mbosoli, Jamie M Lachman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Advantages and Inconveniences of a Multi-Agent Large Language Model System to Mitigate Cognitive Biases in Diagnostic Challenges

Cedric Bousquet[1,2], PharmD, PhD; Divà Beltramin[3], MD, MSc

[1]Laboratory of Medical Informatics and Knowledge Engineering in e-Health, Inserm, Sorbonne University, Paris, France

[2]Public Health and Medical Information Unit, Saint-Étienne University Hospital Center, Saint-Etienne, France

[3]Medical Information Department, Civil Hospices of Lyon, Lyon, France

Corresponding Author:
Cedric Bousquet, PharmD, PhD
Laboratory of Medical Informatics and Knowledge Engineering in e-Health
Inserm
Sorbonne University
15 rue de l'école de Médecine
Paris, F-75006
France
Phone: 33 0477127974
Email: cedric.bousquet@chu-st-etienne.fr

Related Article:

Comment on: http://www.jmir.org/2024/1/e59439/

(J Med Internet Res 2025;27:e69742) doi:10.2196/69742

KEYWORDS

large language model; multi-agent system; diagnostic errors; cognition; clinical decision-making; cognitive bias; generative artificial intelligence

We read with great interest a recent article in the *Journal of Medical Internet Research* entitled "Mitigating Cognitive Biases in Clinical Decision-Making Through Multi-Agent Conversations Using Large Language Models: Simulation Study" by Ke et al [1]. Large language models (LLMs) have advanced reasoning skills but most studies have focused on assessing their ability to answer questions [2]. Ke et al [1] evaluated the ability of LLMs to avoid reproducing certain biases that physicians experience when making a diagnosis. The authors simulated decision-making using a multi-agent system in which each agent models the reasoning of a physician using an LLM. The experiment shows a significant increase in performance compared to a human, with an odds ratio of 3.91, which is highly remarkable. This is made possible by a particularly low score for humans, which leaves considerable room for improvement by LLMs.

Previously published studies have often shown disappointing results using LLMs, with performance levels that give cause for concern when used in a clinical context. Some studies have observed performance similar to that of physicians and marginally better in a limited number of cases. We are thus surprised to see that the multi-agent approach has led to such superiority compared to the physicians. This superiority may be explained by using a multi-agent framework, but it could also be limited to the context of addressing cognitive biases.

To investigate this hypothesis, it would be interesting to compare the performance of the multi-agent system with that of a single LLM. In particular, a prompt engineering method known as tree of thought allows several reasoning paths to be explored, enabling the initial choice to be modified according to diverse alternatives [3].

Furthermore, a recent study found no added value in associating an LLM with a physician to improve performance when making a diagnosis [4]. Moreover, the quantity of text and the number of interventions produced by multi-agent systems may be overwhelming, and the physician could ignore the suggestions. This was the case for many studies on computerized decision support systems for drug prescriptions, where too many alerts have led physicians to override relevant alerts [5].

The multi-agent system can lead to significantly higher costs and requires the mobilization of a greater number of resources. This depends on the number of agents involved in the multi-agent system and, above all, on the amount of text

XSL•FO
RenderX

generated. The multi-agent system is likely to produce numerous arguments and counterarguments to make a fair diagnosis. It would be interesting to know how many tokens are produced in the application logs compared to a simpler application.

The authors suggest that similar approaches could revolutionize medical practice by making clinical decisions more reliable and more consistent with the levels of evidence available. We subscribe to the idea that LLMs could help physicians, and the authors' approach seems promising. However, there is still a lack of evidence on how these multi-agent systems would perform if they were generalized to other diagnostic challenges.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Mitigating Cognitive Biases in Clinical Decision-Making Through Multi-Agent Conversations Using Large Language Models: Simulation Study" declined to respond to this letter.

References

1. Ke Y, Yang R, Lie SA, Lim TXY, Ning Y, Li I, et al. Mitigating cognitive biases in clinical decision-making through multi-agent conversations using large language models: simulation study. J Med Internet Res 2024 Nov 19;26:e59439 [FREE Full text] [doi: 10.2196/59439] [Medline: 39561363]
2. Bedi S, Liu Y, Orr-Ewing L, Dash D, Koyejo S, Callahan A, et al. Testing and evaluation of health care applications of large language models: a systematic review. JAMA 2024 Oct 15:e2421700. [doi: 10.1001/jama.2024.21700] [Medline: 39405325]
3. Yao S, Yu D, Zhao J, Shafran I, Griffiths T, Cao Y, et al. Tree of thoughts: deliberate problem solving with large language models. In: Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, editors. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). San Diego, CA: Neural Information Processing Systems Foundation, Inc; 2023.
4. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open 2024 Oct 01;7(10):e2440969 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.40969] [Medline: 39466245]
5. Slight SP, Seger DL, Franz C, Wong A, Bates DW. The national cost of adverse drug events resulting from inappropriate medication-related alert overrides in the United States. J Am Med Inform Assoc 2018 Sep 01;25(9):1183-1188 [FREE Full text] [doi: 10.1093/jamia/ocy066] [Medline: 29939271]

Abbreviations

LLM: large language model

Edited by T Leung, L Beri; submitted 06.12.24; this is a non–peer-reviewed article;accepted 11.12.24; published 20.01.25.

Please cite as:
Bousquet C, Beltramin D
Advantages and Inconveniences of a Multi-Agent Large Language Model System to Mitigate Cognitive Biases in Diagnostic Challenges
J Med Internet Res 2025;27:e69742
URL: https://www.jmir.org/2025/1/e69742
doi:10.2196/69742
PMID:

©Cedric Bousquet, Divà Beltramin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Japan's Telepsychiatry Dissemination: Current Status and Challenges

Shotaro Kinoshita[1,2], MD, PhD; Taishiro Kishimoto[1], MD, PhD

[1]Hills Joint Research Laboratory for Future Preventive Medicine and Wellness, Keio University School of Medicine, Tokyo, Japan

[2]Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Taishiro Kishimoto, MD, PhD
Hills Joint Research Laboratory for Future Preventive Medicine and Wellness
Keio University School of Medicine
#7F Azabudai Hills Mori JP Tower 1-3-1
Azabudai, Minato-Ku
Tokyo, 106-0041
Japan
Phone: 81 3 5363 3829
Fax: 81 3 5379 0187
Email: tkishimoto@keio.jp

Related Articles:

Comment on: https://www.jmir.org/2024/1/e64159/

Comment in: http://mhealth.jmir.org/2025/1/e69798/

(J Med Internet Res 2025;27:e22849) doi:10.2196/22849

We read the recently published article titled “Issues in the Adoption of Online Medical Care: Cross-Sectional Questionnaire Survey” by Sugawara Y et al [1] with great interest. The authors conducted two surveys aimed at identifying the current status of telemedicine in Japan and the factors hindering its spread, from the perspectives of both patients and medical professionals providing it. Although the existence of bias cannot be denied because of the voluntary web-based survey, the findings were useful in understanding the current status of telemedicine in Japan.

As corrected in the original paper, the authors cited our study [2] and explained that telepsychiatry was reimbursed at the same or higher rate as in-person consultations in 15 of 17 regions worldwide. In this study, Japan was in the minority, with telepsychiatry reimbursement distinctly lower than face-to-face care. This remains an obstacle to the widespread use of telemedicine in Japan, not only in psychiatry but also across other medical specialties. Even before 2019, and despite deregulation following the COVID-19 pandemic, reimbursement rates for telemedicine have remained lower than those for face-to-face consultations [3].

In addition, telepsychiatry is more strictly regulated than telemedicine in other medical specialties. Even after the most recent deregulation in June 2024, there are still restrictions on prescription drugs and reimbursement rates for initial visits, as well as restrictions on the conditions under which physicians and medical facilities can obtain reimbursement.

The survey by Sugawara et al showed that patients and healthy individuals in psychiatry and psychosomatic medicine are interested in using telemedicine; however, the fact that medical institutions are either unable or unwilling to provide it, due to such strict regulations, is surely the biggest obstacle to its widespread use.

In addition, Sugawara et al emphasize the need to obtain the opinions of psychiatrists and psychosomatic medicine specialists regarding the feasibility of online medical care. In a 6-month randomized controlled trial comprising 199 patients with depression, anxiety, and obsessive-compulsive disorder at 19 medical institutions in Japan, we verified the noninferiority of the combined telepsychiatry group to the face-to-face treatment group [4]. The study also found no significant differences between the two groups on several secondary endpoints, including treatment retention and satisfaction. Furthermore, it confirmed that the telepsychiatry group required less time and incurred lower costs than the face-to-face treatment group. From these studies, we believe that the feasibility of telepsychiatry in the Japanese medical field has already been verified to a certain extent.

In Japan, while the demand for medical care is expected to increase due to further aging of the population, the dispersion

XSL•FO
RenderX

and maldistribution of physicians and hospitals remains unresolved [5]. Since telemedicine is considered an effective measure to improve efficiency and equalize the provision of medical care, the government should ease regulations to make it easier for frontline medical professionals to use.

Authors' Contributions

SK contributed to writing of the original draft. TK contributed to review and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Sugawara Y, Hirakawa Y, Iwagami M, Kuroki H, Mitani S, Inagaki A, et al. Issues in the adoption of online medical care: cross-sectional questionnaire survey. J Med Internet Res 2024 Nov 01;26:e64159 [FREE Full text] [doi: 10.2196/64159] [Medline: 39486019]
2. Kinoshita S, Cortright K, Crawford A, Mizuno Y, Yoshida K, Hilty D, et al. Changes in telepsychiatry regulations during the COVID-19 pandemic: 17 countries and regions' approaches to an evolving healthcare landscape. Psychol Med 2022 Oct;52(13):2606-2613 [FREE Full text] [doi: 10.1017/S0033291720004584] [Medline: 33243311]
3. Kinoshita S, Kishimoto T. Current status and challenges of the dissemination of telemedicine in Japan after the start of the COVID-19 pandemic. Telemed J E Health 2022 Aug 01;28(8):1220-1224. [doi: 10.1089/tmj.2021.0336] [Medline: 34918952]
4. Kishimoto T, Kinoshita S, Kitazawa M, Hishimoto A, Asami T, Suda A, et al. Live two-way video versus face-to-face treatment for depression, anxiety, and obsessive-compulsive disorder: a 24-week randomized controlled trial. Psychiatry Clin Neurosci 2024 Apr 15;78(4):220-228. [doi: 10.1111/pcn.13618] [Medline: 38102849]
5. Kinoshita S, Kishimoto T. Aging population in Japan: immediate shake-up in healthcare required. QJM 2024 May 08:2024. [doi: 10.1093/qjmed/hcae097] [Medline: 38718181]

Edited by T Leung, L Beri; submitted 30.11.24; this is a non–peer-reviewed article;accepted 11.12.24; published 20.01.25.

Please cite as:
Kinoshita S, Kishimoto T
Japan's Telepsychiatry Dissemination: Current Status and Challenges
J Med Internet Res 2025;27:e22849
URL: https://www.jmir.org/2025/1/e22849
doi:10.2196/22849
PMID:

©Shotaro Kinoshita, Taishiro Kishimoto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: Japan's Telepsychiatry Dissemination: Current Status and Challenges

Yuka Sugawara[1], MD, PhD; Yosuke Hirakawa[1], MD, PhD; Masao Iwagami[2], MSc, MD, PhD; Haruo Kuroki[3], MD, PhD; Shuhei Mitani[1], MD; Ataru Inagaki[4], MD, PhD; Hiroki Ohashi[5], MD; Mitsuru Kubota[6], MD, PhD; Soichi Koike[7], MD, PhD; Rie Wakimizu[8], PhD; Masaomi Nangaku[1], MD, PhD

[1]Division of Nephrology and Endocrinology, The University of Tokyo, Tokyo, Japan

[2]Department of Health Services Research, Institute of Medicine, University of Tsukuba, Ibaraki, Japan

[3]Clinic Pauroom, Tokyo, Japan

[4]Department of Education, College of Education, Psychology and Human Studies, Aoyama Gakuin University, Tokyo, Japan

[5]TAMA Family Clinic, Tokyo, Japan

[6]Department of General Pediatrics and Interdisciplinary Medicine, National Center for Child Health and Development, Tokyo, Japan

[7]Division of Health Policy and Management, Center for Community Medicine, Jichi Medical University, Tochigi, Japan

[8]Department of Child Health and Development Nursing, Institute of Medicine, University of Tsukuba, Ibaraki, Japan

Corresponding Author:
Masaomi Nangaku, MD, PhD
Division of Nephrology and Endocrinology
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 3815 5411
Email: mnangaku@m.u-tokyo.ac.jp

Related Articles:

Comment on: http://www.jmir.org/2025/1/e22849/

Comment on: https://www.jmir.org/2024/1/e64159/

(J Med Internet Res 2025;27:e69798) doi:10.2196/69798

Introduction

We thank the authors for their knowledgeable comments [1] on our study [2]. Our results suggest that although telemedicine is not widely used in the field of psychiatry and psychosomatic medicine, it is desired by the patient population receiving medical care. Deregulation of the restriction and setting higher reimbursement rates are considered important, but evidence is needed for this to happen. We believe that the results recently presented by Dr Kishimoto and colleagues [3] are very important, and the fact that they show that telemedicine in this field is not inferior to face-to-face care is good information not only for health care providers and the patient population, but also for government agencies as a resource for developing policy. This kind of evidence needs to be accumulated, and more research on telemedicine is needed in the future.

Conflicts of Interest

None declared.

References

1. Kinoshita S, Kishimoto T. Japan's Telepsychiatry Dissemination: Current Status and Challenges. J Med Internet Res 2024:e22849. [doi: 10.2196/22849]
2. Sugawara Y, Hirakawa Y, Iwagami M, Kuroki H, Mitani S, Inagaki A, et al. Issues in the adoption of online medical care: cross-sectional questionnaire survey. J Med Internet Res 2024 Dec 01;26:e64159 [FREE Full text] [doi: 10.2196/64159] [Medline: 39486019]

XSL•FO
RenderX

3. Kishimoto T, Kinoshita S, Kitazawa M, Hishimoto A, Asami T, Suda A, et al. Live two-way video versus face-to-face treatment for depression, anxiety, and obsessive-compulsive disorder: A 24-week randomized controlled trial. Psychiatry Clin Neurosci 2024 Apr;78(4):220-228. [doi: 10.1111/pcn.13618] [Medline: 38102849]

Edited by T Leung, L Beri; submitted 09.12.24; this is a non–peer-reviewed article;accepted 11.12.24; published 20.01.25.

Please cite as:
Sugawara Y, Hirakawa Y, Iwagami M, Kuroki H, Mitani S, Inagaki A, Ohashi H, Kubota M, Koike S, Wakimizu R, Nangaku M
Authors' Reply: Japan's Telepsychiatry Dissemination: Current Status and Challenges
J Med Internet Res 2025;27:e69798
URL: https://www.jmir.org/2025/1/e69798
doi:10.2196/69798
PMID:

©Yuka Sugawara, Yosuke Hirakawa, Masao Iwagami, Haruo Kuroki, Shuhei Mitani, Ataru Inagaki, Hiroki Ohashi, Mitsuru Kubota, Soichi Koike, Rie Wakimizu, Masaomi Nangaku. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: Commentary on “Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line”

Anthony R Pisani[1,2], PhD; Carlos Gallo[3*], PhD; Madelyn S Gould[4,5*], PhD, MPH; Nitya Kanuri[6*], MBA, MPH; John E Marcotte[7*], PhD; Brian Pascal[8*], JD; David Rousseau[9], MPH; Megan L Ranney[10], MD, MPH; Bob Filbin[11], MA; Shairi Turner[12*], MD, MPH

[1]Department of Psychiatry, Center for the Study and Prevention of Suicide, University of Rochester Medical Center, University of Rochester, Rochester, NY, United States

[2]Department of Pediatrics, University of Rochester Medical Center, University of Rochester, Rochester, NY, United States

[3]Northwestern University, Evanston, IL, United States

[4]Department of Psychiatry, Columbia University, New York, NY, United States

[5]Department of Epidemiology, Columbia University, New York, NY, United States

[6]Chartis, San Francisco, CA, United States

[7]University of Michigan, Ann Arbor, MI, United States

[8]Stanford University, Palo Alto, CA, United States

[9]Henry J Kaiser Family Foundation, Menlo Park, CA, United States

[10]Yale School of Public Health, New Haven, CT, United States

[11]Meta (United States), Menlo Park, CA, United States

[12]Crisis Text Line, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Anthony R Pisani, PhD
Department of Psychiatry, Center for the Study and Prevention of Suicide
University of Rochester Medical Center
University of Rochester
300 Crittenden Blvd
Rochester, NY, 14642
United States
Phone: 1 5855077177
Email: anthony_pisani@urmc.rochester.edu

Related Articles:

Comment on: https://www.jmir.org/2024/1/e42144/

Comment on: https://www.jmir.org/2019/1/e11507/

Comment in: https://www.jmir.org/2025/1/e64015/

Comment in: https://www.jmir.org/2025/1/e67878/

Abstract

We appreciate Reierson’s thoughtful commentary on our 2019 paper, which described our experiences, ethical process, judgment calls, and lessons from a 2016-2017 data-sharing pilot between Crisis Text Line and academic researchers. The commentary raises important questions about the ethical conduct of health research in the digital age, particularly regarding informed consent, potential conflicts of interest, and the protection of vulnerable populations. Our article focused specifically on the noncommercial use of Crisis Text Line data for research purposes, so we restrict our reply to points relevant to such usage. While we acknowledge

XSL•FO
RenderX

the limitations of Crisis Text Line's Terms of Service as a means of informing users about data sharing for research, we maintain that our guidelines were ethically sound and aligned with well-established practices for institutional review board (IRB) review and researcher training. We emphasize the critical role of IRBs in ensuring that research involving vulnerable populations, including minors, is conducted ethically and with appropriate safeguards. Regarding potential conflicts of interest, we argue that unpaid, nonfiduciary advisory board service for a nonprofit organization does not constitute a conflict requiring disclosure. The transparent nature of our collaboration with Crisis Text Line, as evidenced by the authorship and acknowledgments in our paper, further underscores our commitment to ethical research practices. We recognize the complexity and evolving nature of the challenges surrounding data-sharing partnerships in digital health research. As the field progresses, we remain committed to ongoing, transparent engagement and to refining best practices in collaboration with colleagues, stakeholders, and the public. Our response aims to provide clarity and context for the concerns raised in the commentary while reaffirming the integrity and value of our original work. Ultimately, we maintain that our paper contributed meaningfully to the ongoing discourse on ethical data sharing and laid the groundwork for future improvements in this critical area of digital health research.

(J Med Internet Res 2025;27:e59734) doi:10.2196/59734

KEYWORDS

ethics; data sharing; digital data; text messaging; technology; cooperative behavior

Introduction

Reierson's commentary [1] on our 2019 article [2] highlights important and timely questions about the ethical conduct of health research in the digital age. Interactions with increasingly ubiquitous health-related digital platforms generate enormous amounts of data in which individual users, future users, and society have a vested interest [3]. We share Reierson's concerns about how and under what conditions users ought to be informed about how data generated from their interactions are included in health and prevention research, including research that aims to prevent suicide. We published this paper to share with the field our experiences, ethical process, judgment calls, and lessons from a 2016-2017 data-sharing pilot, understanding that science and ethics advance through publication, critique, and refinement.

Our article explicitly pertains to "noncommercial use of data" for the purposes of research and evaluation. The article neither addresses nor endorses use of Crisis Text Line data for commercial purposes. No members of the Crisis Text Line data advisory board that advised on the data-sharing pilot our article described served in any capacity with Loris.ai. Given that Reierson's commentary on the commercial use of data and Loris.ai is outside the scope of our paper and anachronistic to the data pilot we reported on, we focus our response on his in-scope feedback. The commentary also questions common features and norms in human subject research (such as the definition of a research subject and the role of institutional review boards). The adequacy of well-established ethical frameworks and review processes is always worthy of review, but we will focus here on those questions relevant to our specific conduct.

With respect to including deidentified Crisis Text Line data in research, the commentary raised concern about: (1) the Terms of Service that Crisis Text Line used to inform users of data-sharing policies with respect to research and (2) potential undisclosed conflicts of interest in the author group.

Crisis Text Line Terms of Service and Consent

Regarding the concerns raised about Crisis Text Line's Terms of Service (Terms) with respect to research, we appreciate the limitations pointed out by Mr Reierson. At the time the paper was published, we considered the Terms to be an appropriate means by which to "inform users in an *unobtrusive way* that anonymized data are shared with select research partners." We agree that the Terms could have been simplified with respect to research, and we understand that Crisis Text Line has since improved them.

It is important for the general public to understand that any use of Crisis Text Line data for research must first be reviewed by a researcher's university institutional review board (IRB)—an independent committee that reviews the research protocol, including the adequacy of the consent process and the protection of vulnerable populations, such as minors. This review process is part of the context in which our guidelines were offered. Furthermore, all researchers undergo Human Research Protection Training, which emphasizes the Belmont Report's ethical principles of respect for person, beneficence, and justice, with particular attention to the protection of vulnerable populations. IRBs often consider secondary analyses of data without access to identifying information as not meeting the definition of human subjects research according to the federal definition 45 CFR 46.102(f).

Thus, the authors of the article still conclude that what we wrote at the time adheres to ethical standards of research, even while we welcome retrospective critique of the recommendations we generated at the time. Furthermore, we believed—and continue to believe—that it is not just acceptable, but is imperative, for academic researchers to engage with technology companies to create methods and means for research that both improve the protections of people engaging with that technology and facilitate the use of the generated data for good rather than for harm.

XSL•FO
RenderX

Potential Undisclosed Conflicts of Interest

Regarding questions raised about potential conflicts of interest, it should be noted that Crisis Text Line is a nonprofit organization, and no one on the Crisis Text Line data advisory board had a role in or even knew about Loris.ai, which was founded after the data-sharing pilot took place. We did not and do not consider volunteer nonfiduciary participation in the Crisis Text Line advisory board as a "conflict of interest," and none of the authors had any other relationships or funding from Crisis Text Line at the time this article was written and published. The Crisis Text Line advisory boards convened volunteer experts to provide independent advice, an extremely common practice.

- Members of the Crisis Text Line advisory boards are unpaid and receive no other tangible or intangible benefit from their participation. They volunteer a small amount of time to provide input and expertise, hopefully to the benefit of the nonprofit organization and the people it serves.
- Crisis Text Line has no undue influence on advisory board members' feedback or behavior. Advisory board members are sufficiently independent from Crisis Text Line to be able to provide critical feedback—both positive and negative—on Crisis Text Line's operations.
- Our advisory board had no fiduciary responsibilities to Crisis Text Line.

In studying the Committee on Publication Ethics' (COPE) guidelines, JMIR Publications' policy, and other publicly available ethical statements, we could not find any indication that a relationship of this type would be considered a "vested monetary interest" or a "conflict of interest." *[Editorial Note:While we found no intent to deceive on the part of the authors, we disagree with this interpretation; see corrigendum to Pisani et al's 2019 publication [4]]*. We should note that our author group appreciated the opportunity to consider this question. Most agreed that, while not required nor customary, there would be no harm—and could be some potential benefit—in disclosing unpaid advisory board service, and some of us will choose to do so in the future (eg,[5]).

Further, far from being "undisclosed," the participation and collaboration of members of the author group with Crisis Text Line in documenting the results of their pilot academic data-sharing program was a key feature of the article.

- Crisis Text Line employees were listed as authors along with their affiliation.
- An appendix was provided listing members of the data ethics committee, which included several of the authors.
- The paper is manifestly the result of collaboration with Crisis Text Line—academics and nonprofit leaders transparently described the ups and downs of a pilot program aimed at protecting privacy, conducting research ethically, and contributing to science and society.

Summary

In summary, we share Mr Reierson's concern about ethical use of data and appreciate his thoughtful interaction with our 2019 paper. Our article addressed only noncommercial use of data. We see Mr Reierson's arguments about Terms of Service and consent as welcome and substantive critiques rather than ethical concerns about our 2019 paper. We stand by our work and recommendations as ethically sound, and we highlight that we wrote this publication to advance discourse and future ethical guidelines. Finally, unpaid, nonfiduciary advisory board service to a nonprofit organization by independent academic experts does not constitute a conflict of interest, although we see the benefit of listing such service in certain circumstances in the future.

We remain committed to ongoing, constructive scientific dialog and to refining best practices in collaboration with interested parties and the public. Working together, we can best realize the benefits of research while upholding high standards of ethics, privacy, and trust.

Note: The following authors of the original publication could not be reached or were unavailable to participate in this Authors' Reply: Lisa Soleymani Lehmann, Robert Levine (deceased), and Shirley Yen.

Conflicts of Interest

NK was a salaried full-time employee with Crisis Text Line during 2016-2017. BF was a salaried chief data scientist at Crisis Text Line until 2021 and chaired the Data Ethics Committee. ST previously served as chief medical officer for Crisis Text Line and is chief health officer for Crisis Text Line from 2020 to present and receives a salary. ARP and MSG served as uncompensated advisory members of the Clinical Advisory Board to Crisis Text Line from 2014 to 2022. BP, DR, JEM, and MLR previously served as uncompensated members of the Data Ethics Committee or Data, Ethics, and Research Advisory Board for Crisis Text Line.

References

1. Reierson TD. Commentary on "Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles from a Pilot Program at Crisis Text Line". J Med Internet Res 2024 Dec 30:e42144. [doi: 10.2196/42144] [Medline: 39753221]
2. Pisani AR, Kanuri N, Filbin B, Gallo C, Gould M, Lehmann LS, et al. Protecting user privacy and rights in academic data-sharing partnerships: principles from a pilot program at Crisis Text Line. J Med Internet Res 2019 Jan 17;21(1):e11507 [FREE Full text] [doi: 10.2196/11507] [Medline: 30664452]

XSL•FO
RenderX

3. Herington J, Li K, Pisani AR. Expanding the role of justice in secondary research using digital psychological data. Am Psychol 2024 Jan;79(1):123-136. [doi: 10.1037/amp0001190] [Medline: 38236220]
4. JMIR Editorial Office. Correction: protecting user privacy and rights in academic data-sharing partnerships: principles from a pilot program at Crisis Text Line. J Med Internet Res 2024 Dec 20;26:e67880 [FREE Full text] [doi: 10.2196/67880] [Medline: 39705678]
5. Pisani AR, Gould MS, Gallo C, Ertefaie A, Kelberman C, Harrington D, et al. Individuals who text crisis text line: Key characteristics and opportunities for suicide prevention. Suicide Life Threat Behav 2022 Jun 26;52(3):567-582 [FREE Full text] [doi: 10.1111/sltb.12872] [Medline: 35615898]

Abbreviations

COPE: Committee on Publication Ethics
IRB: institutional review board

Edited by G Eysenbach; submitted 24.04.24; this is a non–peer-reviewed article;accepted 05.07.24; published 22.01.25.

Please cite as:
Pisani AR, Gallo C, Gould MS, Kanuri N, Marcotte JE, Pascal B, Rousseau D, Ranney ML, Filbin B, Turner S
Authors' Reply: Commentary on "Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line"
J Med Internet Res 2025;27:e59734
URL: https://www.jmir.org/2025/1/e59734
doi:10.2196/59734
PMID:

©Anthony R Pisani, Carlos Gallo, Madelyn S Gould, Nitya Kanuri, John E Marcotte, Brian Pascal, David Rousseau, Megan L Ranney, Bob Filbin, Shairi Turner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Response From Crisis Text Line to "Commentary on 'Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line'"

Dena Trujillo[1]

Crisis Text Line, New York, NY, United States

Corresponding Author:
Dena Trujillo
Crisis Text Line
225 West 34th Street
Floor 9 — PMB#9135
New York, NY, 10122
United States
Phone: 1 3322304008
Email: dena@crisistextline.org

Related Articles:

Comment on: https://www.jmir.org/2019/1/e11507/

Comment on: https://www.jmir.org/2024/1/e42144/

Comment on: https://www.jmir.org/2025/1/e59734/

Comment in: https://www.jmir.org/2025/1/e67878/

(J Med Internet Res 2025;27:e64015) doi:10.2196/64015

KEYWORDS

crisis; research; data privacy; research ethics; digital rights

As current chief executive officer (CEO) of Crisis Text Line, I am writing this commentary on behalf of the organization to clarify the record, provide context, and supply background information regarding Crisis Text Line and its research standards regarding a critique [1] of a viewpoint paper [2] published in the *Journal of Medical Internet Research* in 2019. Below, you will find information about how we protect and do not commercialize personal data; our ethical research standards; our unique data; texter digital rights, privacy, and ethics; and a snapshot of our recent research.

Crisis Text Line: Background

Crisis Text Line is a nonprofit organization that provides free, 24-7, confidential text-based mental health support and crisis intervention in both English and Spanish for anyone in the United States and Puerto Rico. Since our launch in 2013, we have supported nearly 10 million conversations in the United States and have trained over 71,000 volunteers.

All our volunteers are required to accept our policies as a condition of serving as a volunteer with Crisis Text Line, which disclose that we may share anonymized data with researchers. That disclosure and the consent to our all-encompassing privacy policy existed during and prior to the 2016-2017 period of the research that is the subject of the *Journal of Medical Internet Research* study at issue. Volunteers' consent to that broader, generally applicable policy remains effective regardless of our decision to create a shorter and more tailored volunteer-specific policy in 2021 in the interest of greater transparency. Volunteers serve without compensation and are not under any employment agreements.

Crisis Text Line Does Not Commercialize Personal Data

Crisis Text Line does not and has never monetized personally identifiable information directly or indirectly.

Crisis Text Line previously had an open and public relationship with Loris.ai, a for-profit company aimed at helping other for-profit companies employ de-escalation techniques to reduce stress for customers and employees. At the time, we believed that sharing insights from Crisis Text Line's training curriculum

XSL•FO
RenderX

and deidentified data could support stress reduction and improved well-being in spaces beyond the crisis context while potentially contributing to the sustainability of a nonprofit public good. The world of technology and data ethics evolved rapidly, and Crisis Text Line quickly changed course. Loris.ai did not access any Crisis Text Line anonymized data or other information after early 2020.

Public records demonstrate that Loris.ai *did not exist* during the 18-month period of the pilot analyzed in the study; the company was formed months after the pilot ended. Crisis Text Line severed all ties with Loris.ai in 2022 and required the return or deletion of any data previously accessed. We were publicly transparent about the goals of this relationship, which included broader application of our de-escalation strategies and potential revenue to support our important nonprofit work. Ultimately, Crisis Text Line did not receive financial gain as a result of its prior data- and insights-sharing relationship with Loris.ai.

Crisis Text Line is a nonprofit, and those acting as advisory board members did so on a voluntary unpaid basis. To Crisis Text Line's knowledge, there was no conflict of interest, and none of the research authors were advisory board members for the commercial spin-off Loris.ai.

The Critical Need for Ethical Research in Combating the Mental Health Crisis

We can all agree that the need for evidence-based action on the current mental health crisis is indisputable and urgent. The importance of ethical research focused on tackling this crisis cannot be overemphasized. The prevalence of mental health challenges worldwide highlights the critical need for innovative solutions, compassionate care, and evidence-based interventions. Through ethical research efforts, Crisis Text Line contributes to destigmatizing mental health challenges by disseminating accurate information, challenging misconceptions, and promoting empathy and acceptance. Through collaborative efforts [3] between researchers, policy makers, clinicians, and community stakeholders, we help to foster a holistic approach to mental health that addresses systemic barriers and promotes equitable access to care.

In Pursuit of Solutions

Crisis Text Line's anonymized data provide a unique snapshot of mental health across the United States. We are in direct conversation with people in crisis, so what we learn is not simply the result of a survey. Rather, it is real-time insight about how people feel. We ask about the specifics of their crisis, listen, collaboratively problem solve, de-escalate, and then provide them with an optional anonymous postconversation survey. Data shared by texters themselves in the optional postconversation survey suggest that two-thirds of them do not have access to professional mental health support. We believe it is our duty to elevate their voices and support research efforts that might contribute solutions to help with this emergency in an ethical manner. Analyzing that data—carefully secured and anonymized—allows us to observe emerging trends, almost in real time. For example, in March 2020 during the first COVID-19 lockdowns, we observed a spike in conversation volume as people reached out while struggling with isolation, anxiety, and grief, but with lower suicidal ideation. Our dataset of nearly 10 million conversations (nearly 300 million messages) collected in over 10 years of operation offers a broad and deep perspective of mental health trends across the country. Those data position us to have a real impact through ethical research endeavors.

Data Privacy, Digital Rights, and Ethics

At Crisis Text Line, safeguarding our texters' data is paramount. In moments of crisis, our texters can feel assured, trusting the integrity of the support they receive without concerns about their data's privacy, security, or use. To that end, we've published digital rights [4] for texters, outlining rights and principles that we uphold throughout their use of our service. Ensuring texters' safety, privacy, and confidentiality stand as foundational principles of our service. We protect texter privacy rigorously and maintain texter anonymity, except when necessary for safety or legal reasons.

Our privacy policy [5] transparently discloses how we collect and use data, noting our use of encryption and other reasonable measures to protect against unauthorized access. All texters get a link to the policy at the start of every conversation with us, and texters must consent to the policy to use the service. The policy also intentionally contains simple plain-language bullets at the beginning so a user can quickly and easily understand what we do, how we do it, and why—including, among other things, our use of anonymized deidentified data for research. Texters also have the option to ask that we delete their data, including conversation records and contact information, simply by texting or messaging the word DELETE.

We go to great lengths to protect, store, analyze, and share insights from our anonymized crisis conversations to ethically help the world address mental health issues. We further align with the highest international standards for ethical research involving human subjects. In addition to ensuring that external research collaborations are overseen by the institutional review board (IRB) of the lead researcher's institution, since 2023, our in-house research is also overseen by an IRB. These policies and practices disclose and reinforce our commitments to upholding texter privacy and fulfilling our mission.

Our Recent In-House and Collaborative Research

Since our launch in 2013, we have supported nearly 10 million conversations initiated by texters across the United States. Leveraging this anonymized and deidentified dataset, we inform policy makers, journalists, and researchers. Alongside our internal research endeavors, we foster collaborations with partners at research institutions. Recent examples of research contributing unique insights to the field include:

- Crisis Text Line & Common Good Labs: "What Young People in Crisis Need From Their Communities" (2023) [6]

XSL•FO
RenderX

- Crisis Text Line's United in Empathy Report (2023) [7]
- Examining Hurricane Ida's Impact on Mental Health: Results From a Quasi-Experimental Analysis (2023) [8]
- The Use of Crisis Services Following the Mass School Shooting in Uvalde, Texas: Quasi-Experimental Event Study (2023) [9]
- Crisis Response Among Essential Workers and Their Children During the COVID-19 Pandemic (2021) [10]

A Call to Action

As we navigate the complexities of the mental health crisis, we reaffirm our commitment to ethical research as a catalyst for change. We embrace the transformative power of collaborative, ethical, impactful research in illuminating the path toward solutions, healing, and hope.

Conflicts of Interest

In the last 5 years, the author reports the following financial disclosures: employment as salaried CEO at Crisis Text Line; prior employment at Omidyar Network (which invested in Loris.ai); a travel grant and wellness stipend: The Wellbeing Project; and a minority share in Ultimo Studios. The author reports the following nonfinancial disclosures: prior unpaid board of director membership at Crisis Text Line; board membership: High Resolves; prior advisory group: African Leadership University (African Leadership Group). She has never held equity in the commercial entity Loris.ai.

References

1. Reierson TD. Commentary on "Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles from a Pilot Program at Crisis Text Line". J Med Internet Res 2024 Dec 30;26:e42144. [doi: 10.2196/42144] [Medline: 39753221]
2. Pisani A, Kanuri N, Filbin B, Gallo C, Gould M, Lehmann L, et al. Protecting user privacy and rights in academic data-sharing partnerships: principles from a pilot program at Crisis Text Line. J Med Internet Res 2019 Jan 17;21(1):e11507 [FREE Full text] [doi: 10.2196/11507] [Medline: 30664452]
3. Data philosophy. Crisis Text Line. URL: https://www.crisistextline.org/data-philosophy/ [accessed 2024-12-23]
4. Your digital rights. Crisis Text Line. URL: https://www.crisistextline.org/digital-rights/ [accessed 2024-12-23]
5. Privacy policy. Crisis Text Line. URL: https://www.crisistextline.org/privacy/ [accessed 2024-12-23]
6. What do young people in crisis need from their communities? Crisis Text Line. 2023. URL: https://www.crisistextline.org/what-do-young-people-in-crisis-need-from-their-communities/ [accessed 2024-12-23]
7. United in empathy. Crisis Text Line. 2023. URL: https://www.crisistextline.org/united-in-empathy-2023-report/ [accessed 2024-12-23]
8. Wertis L, Runkle JD, Sugg MM, Singh D. Examining Hurricane Ida's impact on mental health: results from a quasi-experimental analysis. Geohealth 2023 Feb;7(2):e2022GH000707 [FREE Full text] [doi: 10.1029/2022GH000707] [Medline: 36776989]
9. Weitzel KJ, Chew RF, Miller AB, Oppenheimer CW, Lowe A, Yaros A. The use of crisis services following the mass school shooting in Uvalde, Texas: quasi-experimental event study. JMIR Public Health Surveill 2023 Feb 08;9:e42811 [FREE Full text] [doi: 10.2196/42811] [Medline: 36753321]
10. Sugg MM, Runkle JD, Andersen L, Weiser J, Michael KD. Crisis response among essential workers and their children during the COVID-19 pandemic. Prev Med 2021 Dec;153:106852 [FREE Full text] [doi: 10.1016/j.ypmed.2021.106852] [Medline: 34673081]

Abbreviations

CEO: chief executive officer
IRB: institutional review board

Edited by T Leung, G Eysenbach; submitted 05.07.24; this is a non–peer-reviewed article;accepted 20.12.24; published 22.01.25.

Please cite as:
Trujillo D
Response From Crisis Text Line to "Commentary on 'Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line'"
J Med Internet Res 2025;27:e64015
URL: https://www.jmir.org/2025/1/e64015
doi:10.2196/64015
PMID:

XSL•FO
RenderX

©Dena Trujillo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Advancing Insights Into Postoperative Sleep Quality and Influencing Factors

Yining Zhao[1], MD; Xin Hu[2], MD

[1]Department of Cardiology, Faculty of Medicine, The First Hospital of Shanxi Medical University, Shanxi, China

[2]The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China

Corresponding Author:
Yining Zhao, MD
Department of Cardiology
Faculty of Medicine
The First Hospital of Shanxi Medical University
85 Jiefang South St
Shanxi, 030001
China
Phone: 86 18535223677
Fax: 86 19835114977
Email: 630324540@qq.com

Related Articles:

Comment in: https://www.jmir.org/2025/1/e70168/

Comment on: https://www.jmir.org/2024//e56777/

(J Med Internet Res 2025;27:e69193) doi:10.2196/69193

KEYWORDS

sleep quality; wearable sleep monitoring wristband; intensive care unit; minimally invasive surgery; traditional open surgery

We are writing to provide our comments on the recently published article titled "Quantitative Impact of Traditional Open Surgery and Minimally Invasive Surgery on Patients' First-Night Sleep Status in the Intensive Care Unit: Prospective Cohort Study" by Shang et al [1]. This paper addresses an important clinical question regarding postoperative sleep quality and provides valuable insights into the advantages of minimally invasive surgery (MIS). We commend the authors for their efforts in using both subjective questionnaires and objective wearable technology to explore this topic. However, we would like to share several thoughts that could help further refine the interpretation of results and consider additional avenues for future studies.

First, this study focuses on postoperative sleep but overlooks baseline sleep conditions. Preoperative sleep disorders or psychological issues, such as anxiety or depression, could predispose patients to postoperative disturbances, while good baseline sleep quality may confer greater resilience [2]. Preoperative assessments using tools like the Pittsburgh Sleep Quality Index or short-term monitoring could help clarify the independent impact of surgical methods. Additionally, pain management significantly affects sleep through both direct and indirect mechanisms. While MIS likely reduces sleep interference due to less pain, the absence of quantitative pain scores and analgesic data limits interpretability. Future research should address preoperative sleep and analgesic strategies to provide a more comprehensive understanding.

Second, wearable devices offer convenience but have limitations in postoperative or intensive care unit settings. They rely on accelerometers and heart rate variability without electroencephalogram support, risking overestimation of sleep duration, especially in patients who are sedated [3]. Environmental factors like intensive care unit noise and caregiving activities can further compromise accuracy, while algorithms optimized for healthy populations may not reflect the sleep patterns of patients who are critically ill. Future studies should improve algorithms, validate accuracy in complex environments, and integrate devices with other monitoring methods to enhance clinical utility.

Third, psychological factors such as preoperative anxiety or depression significantly influence postoperative sleep. Patient perceptions of MIS, such as reduced trauma and faster recovery, may also indirectly improve sleep. However, the study did not quantify these factors, potentially underestimating their role. Additionally, the single-center design and small sample size limit generalizability and subgroup analyses. Future research

XSL•FO
RenderX

should adopt multicenter, larger-scale studies and include psychological assessments to enhance applicability and robustness.

In conclusion, Shang et al's [1] study provides meaningful insights into the impact of surgical approaches on postoperative sleep quality, particularly highlighting the potential benefits of MIS. While the findings are valuable, addressing the aforementioned limitations could enhance the robustness and clinical relevance of future research. We appreciate the authors' contribution to this important topic and look forward to further advancements in this field.

Conflicts of Interest

None declared.

References

1. Shang C, Yang Y, He C, Feng J, Li Y, Tian M, et al. Quantitative impact of traditional open surgery and minimally invasive surgery on patients' first-night sleep status in the intensive care unit: prospective cohort study. J Med Internet Res 2024 Nov 22;26:e56777 [FREE Full text] [doi: 10.2196/56777] [Medline: 39576980]
2. Niklasson A, Finan PH, Smith MT, Forsberg A, Dietz N, Kander T, et al. The relationship between preoperative sleep disturbance and acute postoperative pain control: a systematic review and meta-analysis. Sleep Med Rev 2024 Oct 05;79:102014 [FREE Full text] [doi: 10.1016/j.smrv.2024.102014] [Medline: 39504912]
3. Helfand M, Christensen V, Anderson J. Technology Assessment: Early Sense for Monitoring Vital Signs in Hospitalized Patients. Washington, DC: Department of Veterans Affairs; 2016.

Abbreviations

MIS: minimally invasive surgery

Edited by T Leung, S Gardezi; submitted 24.11.24; this is a non–peer-reviewed article;accepted 28.12.24; published 03.02.25.

Please cite as:
Zhao Y, Hu X
Advancing Insights Into Postoperative Sleep Quality and Influencing Factors
J Med Internet Res 2025;27:e69193
URL: https://www.jmir.org/2025/1/e69193
doi:10.2196/69193
PMID:

©Yining Zhao, Xin Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: Advancing Insights Into Postoperative Sleep Quality and Influencing Factors

Chen Shang[1*], MD; Ya Yang[2*], PhD; Chengcheng He[1], BSCN; Junqi Feng[1], MM; Yan Li[1], BSCN; Meimei Tian[1], PhD; Zhanqi Zhao[3,4,5], PhD; Yuan Gao[1], MD, PhD; Zhe Li[1], MD, PhD

[1]Department of Critical Care Medicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[2]Department of Infection Control, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[3]School of Biomedical Engineering, Guangzhou Medical University, Guangzhou, China

[4]Department of Critical Care Medicine, Peking Union Medical College Hospital, Beijing, China

[5]Institute of Technical Medicine, Furtwangen University, Villingen-Schwenningen, Germany

[*]these authors contributed equally

Corresponding Author:
Zhe Li, MD, PhD
Department of Critical Care Medicine
Renji Hospital
School of Medicine, Shanghai Jiao Tong University
No. 160, Pujian Road
Pudong New District
Shanghai, 200127
China
Phone: 86 68383162
Fax: 86 68383162
Email: slamy1987@126.com

Related Articles:

Comment on: https://www.jmir.org/2025/1/e69193/

Comment on: https://www.jmir.org/2024/1/e56777/

(J Med Internet Res 2025;27:e70168) doi:10.2196/70168

KEYWORDS

sleep quality; wearable sleep monitoring wristband; intensive care unit; minimally invasive surgery; traditional open surgery

Thank you for your thoughtful feedback on our article, "Quantitative Impact of Traditional Open Surgery and Minimally Invasive Surgery on Patients' First-Night Sleep Status in the Intensive Care Unit: Prospective Cohort Study" [1]. We appreciate your insights and would like to address the key points you raised [2].

Baseline Sleep Conditions

We fully agree that baseline sleep quality is crucial. In our study design, we excluded participants with chronic sleep disorders based on preoperative medical history and medication use. The suggestion to incorporate sleep scoring tools like the Pittsburgh Sleep Quality Index is valuable, as it could help investigate the impact of surgery on sleep status in different patient populations. Inspired by your comments, continuous monitoring using wearable devices during the preoperative and intensive care unit (ICU) periods could provide valuable insights into how sleep is dynamically affected by different clinical treatment protocols.

Pain Management and Analgesic Strategies

We completely agree that pain is one of the greatest "enemies" of sleep quality. Working in a surgical ICU in a teaching hospital, we use standardized protocols along with the Richmond Agitation-Sedation Scale and Numeric Rating Scale (NRS) to manage pain and sedation. Our sedation and analgesia goals were maintained with Richmond Agitation-Sedation Scale scores of –1 to 0, aiming for an NRS below 3 [3,4]. Our unpublished data showed that both the minimally invasive surgery and traditional open surgery groups demonstrated comparable NRS scores (mean NRS: 1.86, SD 0.76, vs 1.85, SD 0.62; *P*=.96).

XSL•FO
RenderX

This consistency allowed us to isolate the effects of surgical methods on sleep.

Wearable Device Limitations

Rapid screening and diagnosis of sleep disorders in ICU settings remain challenging. Our study aimed to explore the use of wearable devices for patients who were awake in the ICU, helping to reduce bias introduced by sedative medications. We agree that ICU-specific environmental factors affect all electronic monitoring tools. Examining how these environmental factors impact device accuracy would be a meaningful research topic. The development of more sophisticated algorithms—whether embedded in device software or integrated with artificial intelligence–based predictive models using clinical data—represents a promising direction for the field of medical engineering integration.

Psychological Factors and Study Design

Lastly, we appreciate your emphasis on the psychological factors influencing sleep, such as preoperative anxiety and depression. These factors may indirectly affect sleep quality, and the reduced trauma and faster recovery associated with minimally invasive surgery could positively influence psychological well-being. Exploring dynamic and quantitative methods to assess psychological states, alongside studying the effects of clinical treatments and mind-body interventions, presents a valuable opportunity to uncover their relationships with clinical outcomes. Combining these approaches with dynamic sleep state monitoring provides technical support to illuminate both direct and mediating relationships, paving the way for large-scale controlled trials to enhance patient care.

Thank you again for your valuable feedback. We look forward to further discussions and advancing this important field.

Acknowledgments

The study was supported by the Wu Jieping Medical Foundation (320.6750.2024-2-4) and the Shanghai Hospital Development Center Foundation (SHDC12024628).

Conflicts of Interest

None declared.

References

1. Shang C, Yang Y, He C, Feng J, Li Y, Tian M, et al. Quantitative impact of traditional open surgery and minimally invasive surgery on patients' first-night sleep status in the intensive care unit: prospective cohort study. J Med Internet Res 2024 Nov 22;26:e56777 [FREE Full text] [doi: 10.2196/56777] [Medline: 39576980]
2. Zhao Y, Hu X. Advancing insights into postoperative sleep quality and influencing factors. J Med Internet Res 2025:e69193. [doi: 10.2196/69193]
3. Arroyo-Novoa CM, Figueroa-Ramos MI, Puntillo KA. Occurrence and practices for pain, agitation, and delirium in intensive care unit patients. P R Health Sci J 2019 Sep;38(3):156-162 [FREE Full text] [Medline: 31536628]
4. Devlin JW, Skrobik Y, Gélinas C, Needham DM, Slooter AJC, Pandharipande PP, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. Crit Care Med 2018 Sep;46(9):e825-e873 [FREE Full text] [doi: 10.1097/CCM.0000000000003299] [Medline: 30113379]

Abbreviations

ICU: intensive care unit
NRS: Numeric Rating Scale

Edited by T Leung, S Gardezi; submitted 17.12.24; this is a non–peer-reviewed article;accepted 28.12.24; published 03.02.25.

Please cite as:
Shang C, Yang Y, He C, Feng J, Li Y, Tian M, Zhao Z, Gao Y, Li Z
Authors' Reply: Advancing Insights Into Postoperative Sleep Quality and Influencing Factors
J Med Internet Res 2025;27:e70168
URL: https://www.jmir.org/2025/1/e70168
doi:10.2196/70168
PMID:

XSL•FO
RenderX

©Chen Shang, Ya Yang, Chengcheng He, Junqi Feng, Yan Li, Meimei Tian, Zhanqi Zhao, Yuan Gao, Zhe Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Enhancing the Clinical Relevance of AI Research for Medication Decision-Making

Qi Wang[1,2], MS; Mingxian Chen[3], MD, PhD

[1]Zhejiang Chinese Medical University, Hangzhou, China

[2]Traditional Chinese Medicine Hospital of Luqiao District, Taizhou, China

[3]Department of Gastroenterology, Tongde Hospital of Zhejiang Province, Hangzhou, China

Corresponding Author:
Mingxian Chen, MD, PhD
Department of Gastroenterology
Tongde Hospital of Zhejiang Province
234 Gucui Street
Xihu Region
Hangzhou, 310012
China
Phone: 86 151 576 82797
Email: chenmingxian2005@126.com

Related Articles:

Comment on: https://www.jmir.org/2024/1/e60794/

Comment in: https://www.jmir.org/2025/1/e72007/

(J Med Internet Res 2025;27:e70657) doi:10.2196/70657

KEYWORDS

older adults; artificial intelligence; medication; decision-making; data security; patient trust

We are writing to provide comments on the manuscript titled "Investigating Older Adults' Perceptions of AI Tools for Medication Decisions: Vignette-Based Experimental Survey" by Vordenberg et al [1]. This study addresses a critical area of research, exploring older adults' trust in artificial intelligence (AI) tools for medication recommendations. While the manuscript presents meaningful insights, several aspects could benefit from further clarification or expansion.

First, older adults may encounter practical challenges when using AI tools, such as difficulties with device operation or complex interfaces. Incorporating user testing or enhanced scenario simulations could optimize the usability of AI tools for older adults. As technology evolves, older adults' perceptions of AI may change over time or with increased exposure to its use. A comparison of AI recommendations with those of health care professionals may either strengthen trust in AI or create conflicts with medical advice, which could lead to greater tensions in the doctor-patient relationship [2]. Long-term follow-up studies would be valuable in tracking these dynamics.

Second, the study emphasizes the impact of differences among demographic groups but does not explore the cultural and socioeconomic factors that underlie these disparities. The acceptance of AI recommendations by older adults may be influenced by cultural attitudes, psychological factors, income levels, and past medical experiences. These elements may limit the generalizability of the study's findings across diverse social contexts [3]. Future research should delve deeper into these factors to enhance the applicability of the results.

Third, while the authors have addressed certain issues, we believe that several important considerations remain underexplored. The study assumes controlled scenarios to manage AI-related variables; however, this simplification may be overly idealized and fails to capture the complexity of medication decision-making in real-world contexts [4]. In practice, decisions regarding medication discontinuation are influenced by multiple intersecting factors, including not only the doctor's advice but also communication and trust between the patient and health care provider, the patient's financial situation, the need for coordinated care across multiple chronic conditions, and potential side effects following medication cessation. Additionally, the opinions and interventions of family members, as well as the patient's personal values, may also play significant roles in the decision-making process [5]. Therefore, medication decisions in real-life scenarios are very complex and require a comprehensive consideration of various factors beyond fixed hypothetical scenarios.

XSL•FO
RenderX

Overall, this study makes a significant contribution to understanding older adults' perceptions of AI in health care. Addressing these limitations could enhance the clinical relevance and robustness of future research. We would like to commend the authors for their efforts and thank the editorial team for providing a platform to publish this important work.

Acknowledgments

The study was supported by the State Administration of Traditional Chinese Medicine Science and Technology Department, Zhejiang Provincial Administration of Traditional Chinese Medicine Co-construction of Traditional Chinese Medicine Modernization Research Program Major Project (GZY-ZJ-KJ-23007); State Administration of Traditional Chinese Medicine Science and Technology Department, Zhejiang Provincial Administration of Traditional Chinese Medicine Co-construction of Key Laboratory of Research on Prevention and Treatment for Depression Syndrome (GZY-ZJ-SY-2402); Zhejiang Province 551 Health Talent Training Project (Zhejiang Provincial Health Commission Office [2021] No. 40)

Conflicts of Interest

None declared.

References

1. Vordenberg SE, Nichols J, Marshall VD, Weir KR, Dorsch MP. Investigating older adults' perceptions of AI tools for medication decisions: vignette-based experimental survey. J Med Internet Res 2024 Dec 16;26:e60794 [FREE Full text] [doi: 10.2196/60794] [Medline: 39680885]
2. Bynum JPW, Barre L, Reed C, Passow H. Participation of very old adults in health care decisions. Med Decis Making 2014 Feb;34(2):216-230 [FREE Full text] [doi: 10.1177/0272989X13508008] [Medline: 24106235]
3. Chen JY, Diamant A, Pourat N, Kagawa-Singer M. Racial/ethnic disparities in the use of preventive services among the elderly. Am J Prev Med 2005 Dec;29(5):388-395. [doi: 10.1016/j.amepre.2005.08.006] [Medline: 16376701]
4. Vordenberg SE, Weir KR, Jansen J, Todd A, Schoenborn N, Scherer AM. Harm and medication-type impact agreement with hypothetical deprescribing recommendations: a vignette-based experiment with older adults across four countries. J Gen Intern Med 2023 May;38(6):1439-1448 [FREE Full text] [doi: 10.1007/s11606-022-07850-5] [Medline: 36376636]
5. Wastesson JW, Morin L, Tan EC, Johnell K. An update on the clinical consequences of polypharmacy in older adults: a narrative review. Expert Opin Drug Saf 2018 Dec;17(12):1185-1196 [FREE Full text] [doi: 10.1080/14740338.2018.1546841] [Medline: 30540223]

Abbreviations

AI: artificial intelligence

Edited by T Leung; submitted 29.12.24; this is a non–peer-reviewed article;accepted 06.02.25; published 18.02.25.

Please cite as:
Wang Q, Chen M
Enhancing the Clinical Relevance of Al Research for Medication Decision-Making
J Med Internet Res 2025;27:e70657
URL: https://www.jmir.org/2025/1/e70657
doi:10.2196/70657
PMID:39964744

©Qi Wang, Mingxian Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: Enhancing the Clinical Relevance of AI Research for Medication Decision-Making

Sarah E Vordenberg[1], PharmD, MPH; Julianna Nichols[1], BS; Vincent D Marshall[1], MS; Kristie Rebecca Weir[2,3], PhD; Michael P Dorsch[1], PharmD, MS

[1]College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

[2]Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[3]Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland

Corresponding Author:
Sarah E Vordenberg, PharmD, MPH
College of Pharmacy
University of Michigan
428 Church St
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 6691
Email: skelling@med.umich.edu

Related Articles:

Comment on: https://www.jmir.org/2024/1/e60794/

Comment on: https://www.jmir.org/2025/1/e70657/

(J Med Internet Res 2025;27:e72007) doi:10.2196/72007

KEYWORDS

older adults; artificial intelligence; vignette; pharmacology; medication; decision-making; aging; attitude; perception; perspective; electronic heath record

We appreciate the insights regarding our manuscript, "Investigating Older Adults' Perceptions of AI Tools for Medication Decisions: Vignette-Based Experimental Survey" [1], in the letter to the editor shared by Wang and Chen [2]. The letter emphasizes three key points: (1) older adults may encounter practical challenges with artificial intelligence tools, necessitating user testing and scenario simulations to enhance usability; (2) although our study considered demographic differences, it did not explore underlying cultural and socioeconomic factors, which calls for further research; (3) the complexity of real-world medication decision-making remains significant.

We acknowledge the limitations the correspondents highlight, particularly usability challenges for older adults and the need to examine underlying cultural and socioeconomic factors. Our vignette-based experiment served as an intentionally focused first step in understanding older adults' interest in artificial intelligence–assisted medication decision support.

Wang and Chen [2] note that our previous work showed that medication decision-making is complex [3]. Indeed, various factors influence how patients make these decisions, including their attitudes, beliefs, and preferences regarding medications; the potential benefits and harms of the treatment under consideration; and contextual factors specific to the patient [4-6]. Our study's experimental approach using vignettes, as indicated in the title, was selected to begin exploring this complex area in a controlled manner.

We agree that this initial study lays the foundation for future research that can tackle these important considerations through enhanced scenarios, user testing, and a more in-depth examination of sociocultural factors.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Vordenberg SE, Nichols J, Marshall VD, Weir KR, Dorsch MP. Investigating older adults' perceptions of AI tools for medication decisions: vignette-based experimental survey. J Med Internet Res 2024 Dec 16;26:e60794 [FREE Full text] [doi: 10.2196/60794] [Medline: 39680885]
2. Wang Q, Chen M. Enhancing the clinical relevance of Al research for medication decision-making. J Med Internet Res 2025:e0. [doi: 10.2196/70657]
3. Vordenberg SE, Weir KR, Jansen J, Todd A, Schoenborn N, Scherer AM. Harm and medication-type impact agreement with hypothetical deprescribing recommendations: a vignette-based experiment with older adults across four countries. J Gen Intern Med 2023 May;38(6):1439-1448 [FREE Full text] [doi: 10.1007/s11606-022-07850-5] [Medline: 36376636]
4. Weir K, Marshall V, Vordenberg S. Latent class analysis identifies four distinct patient deprescribing typologies among older adults in four countries. Innov Aging 2025 Jan 17:igaf002. [doi: 10.1093/geroni/igaf002]
5. Vordenberg SE, Ostaszewski K, Marshall VD, Zikmund-Fisher BJ, Weir KR. Effects of warning information at medication initiation on deprescribing intentions in older adults: a hypothetical vignette. Patient Educ Couns 2025 Jan 10;133:108654. [doi: 10.1016/j.pec.2025.108654] [Medline: 39818128]
6. Weir KR, Vordenberg SE, Scherer AM, Jansen J, Schoenborn N, Todd A. Exploring different contexts of statin deprescribing: a vignette-based experiment with older adults across four countries. J Gen Intern Med 2024 Jul;39(9):1773-1776 [FREE Full text] [doi: 10.1007/s11606-024-08698-7] [Medline: 38514582]

Edited by T Leung; submitted 31.01.25; this is a non–peer-reviewed article;accepted 06.02.25; published 18.02.25.

Please cite as:
Vordenberg SE, Nichols J, Marshall VD, Weir KR, Dorsch MP
Authors' Reply: Enhancing the Clinical Relevance of Al Research for Medication Decision-Making
J Med Internet Res 2025;27:e72007
URL: https://www.jmir.org/2025/1/e72007
doi:10.2196/72007
PMID:39964740

©Sarah E Vordenberg, Julianna Nichols, Vincent D Marshall, Kristie Rebecca Weir, Michael P Dorsch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders

Alina Yang[1]
Scarsdale High School, Scarsdale, NY, United States

Corresponding Author:
Alina Yang
Scarsdale High School
1057 Post Road
Scarsdale, NY, 10583
United States
Phone: 1 (914) 721 2500
Email: alinayang5@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2024/1/e49453/

Comment in: http://www.jmir.org/2025/1/e71897/

(J Med Internet Res 2025;27:e70020) doi:10.2196/70020

KEYWORDS

adolescent health; digital health literacy; adolescents; online health information; co-design; health education; eHealth literacy; social media

Dear Editor,

The research publication entitled “Developing an Educational Resource Aimed at Improving Adolescent Digital Health Literacy: Using Co-Design as Research Methodology” [1] represents a pivotal intervention in the domain of youth-centric health research. With co-design effectively decentralizing traditional research frameworks by considering authentic adolescent perspectives on digital health literacy through their critiques of accessibility, personal and social factors, and information overload, resulting educational resources were able to incorporate more engaging, culturally and generationally relevant content to improve adolescents’ digital health literacy skills and self-efficacy.

Of all studies involving individuals aged 12-18 years, less than 1% reported using youth advice, and barriers to youth participation included limited researcher understanding, lack of resources, and organizational challenges [2]. This alarming underrepresentation epitomizes a systemic failure to acknowledge adolescents as active contributors to the knowledge that directly impacts their lives, instead perpetuating a research production hierarchy that undermines the unique cognitive capabilities and viewpoints of youth populations. While a variety of innovative digital health interventions, such as social networking sites, apps, and video games, have been developed to support youth, the successful development of these tools requires consistent input from target users [3]. Only in this way can we ensure the alignment of protocols, programs, and interventions to adolescent criteria of trustworthiness, safety, peer-to-peer connection, quality, and engagement, as well as the related usability and accessibility of tools among adolescents.

Research on our youth should ultimately be conducted by our youth, but more importantly, with our youth. Adolescents should not be arbitrarily labeled as mere resources for data but rather included in a detailed and thorough knowledge generation apparatus, including both study design and program implementation [4]. Promisingly, youth participatory action research fosters youth leadership in identifying health values and actions while balancing health promotion with personal growth and development [5]. This approach allies a commitment to empowering youth as cocreators with the recognition that their lived experiences are essential in designing interventions that truly resonate with their generation’s needs, particularly in a digital age where adolescents serve as primary consumers and innovators of social media and technology, respectively.

In traditional frameworks, adolescents are often regarded as passive subjects, with their needs, preferences, and involvement dictated by researchers, funders, or institutional mandates. But when youth are positioned as equals, the collaboration becomes more genuine, and the resulting health tools are more likely to meet their evolving demands. This shift is critical, particularly in digital health, where trends, technologies, and social norms change constantly, and youth are often the most proficient at

XSL•FO
RenderX

staying ahead of these shifts. Thus, interventions, when co-designed with the very individuals they seek to serve, will be inherently more attuned to their preferences, technological fluency, and social contexts, thereby maximizing their potential for adoption, use, and sustainable impact. A radical restructuring of adolescent research involvement is long overdue, and we must redefine the processes by which youth health and knowledge are generated, validated, and disseminated to not just meet but anticipate their needs and realities.

Authors' Contributions

AY wrote, edited, revised, and reviewed this manuscript.

Conflicts of Interest

None declared.

References

1. Lewis CC, Taba M, Allen TB, Caldwell PH, Skinner SR, Kang M, et al. Developing an educational resource aimed at improving adolescent digital health literacy: using co-design as research methodology. J Med Internet Res 2024 Aug 07;26:e49453 [FREE Full text] [doi: 10.2196/49453] [Medline: 39110967]
2. Sellars E, Pavarini G, Michelson D, Creswell C, Fazel M. Young people's advisory groups in health research: scoping review and mapping of practices. Arch Dis Child 2021 Jul 18;106(7):698-704. [doi: 10.1136/archdischild-2020-320452] [Medline: 33208398]
3. Malloy J, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions with young people: a scoping review. Digit Health 2023 Dec 11;9:20552076231219117 [FREE Full text] [doi: 10.1177/20552076231219117] [Medline: 38107986]
4. Bowler L, Wang K, Lopatovska I, Rosin M. The meaning of “participation” in co-design with children and youth: relationships, roles, and interactions. Proc Assoc Inf Sci Technol 2021 Oct 13;58(1):13-24. [doi: 10.1002/pra2.432]
5. Lindquist-Grantz R, Abraczinskas M. Using youth participatory action research as a health intervention in community settings. Health Promot Pract 2020 Jul;21(4):573-581 [FREE Full text] [doi: 10.1177/1524839918818831] [Medline: 30577698]

Edited by T Leung; submitted 12.12.24; this is a non–peer-reviewed article;accepted 29.01.25; published 20.02.25.

Please cite as:
Yang A
Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders
J Med Internet Res 2025;27:e70020
URL: https://www.jmir.org/2025/1/e70020
doi:10.2196/70020
PMID:

©Alina Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: "Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders"

Callum C Lewis[1], BClinSc, MD; Melody Taba[1,2], BSc, MPH; Tiffany B Allen[1], BSc, MD; Patrina H.Y Caldwell[1,3], BMed, PhD; S Rachel Skinner[1,3], MBBS, PhD; Melissa Kang[4], MBBS, MCH, PhD; Hamish Henderson[5], BScIT; Liam Bray[5], BA; Madeleine Borthwick[5], MSc; Philippa Collin[6], BA, PhD; Kirsten McCaffery[2], BSc, PhD; Karen M Scott[1,7], BEd, MA, PhD

[1]Specialty of Child and Adolescent Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Sydney Health Literacy Lab, School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[3]Children's Hospital at Westmead, Sydney, Australia

[4]General Practice Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[5]Sydney School of Architecture, Design and Planning, University of Sydney, Sydney, Australia

[6]Institute for Culture and Society, Western Sydney University, Richmond, Australia

[7]Education Office, Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

Corresponding Author:
Karen M Scott, BEd, MA, PhD
Education Office, Sydney Medical School
Faculty of Medicine and Health
University of Sydney
Room 208D, Edward Ford Building
Sydney, 2006
Australia
Phone: 1 298453385
Email: karen.scott@sydney.edu.au

Related Articles:

Comment on: http://www.jmir.org/2025/1/e70020/

Comment on: https://www.jmir.org/2024/1/e49453/

(J Med Internet Res 2025;27:e71897) doi:10.2196/71897

KEYWORDS

adolescent health; digital health literacy; adolescents; online health information; co-design; health education; eHealth literacy; social media

The author of "Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders" [1] highlights the importance of actively involving adolescents in research about adolescent health. Our co-design with adolescents, outlined in "Developing an Educational Resource Aimed at Improving Adolescent Digital Health Literacy: Using Co-Design as Research Methodology" [2], was characterized as a "pivotal intervention" because we prioritized the youth voice in its design process.

The role played by adolescents in our research was central to our development of a digital educational resource that would help adolescents improve their digital health literacy. We valued the involvement of a diverse group of adolescents as research partners in the co-design, to ensure the resource was engaging and aligned with the educational needs of a broad range of adolescents. Thus, our adolescent research partners' insights and recommended designs were an integral part of the educational resource, which was created for use on handheld devices, based around quizzes and serious games, and featured short videos and still images that are typical of social media platforms that are popular with adolescents.

Our adolescent research partners who rejoined us for beta testing were pleasantly surprised to see that the educational resource was developed based on their design ideas, and they were pleased to have been involved and heard in the project. This highlighted the valuable combination of inputs from these diverse adolescent research partners with the equally diverse research team members, all of whom brought unique

XSL•FO
RenderX

perspectives to the design. In a separate, follow-up evaluation, we found that the educational resource was effective at improving the digital health literacy of another group of adolescents, who felt that the content was engaging and met their age group's educational needs.

Social-cognitive theory's concept of self-efficacy (confidence in one's abilities) was foundational in our co-design approach and the educational resource [3]. To enhance users' self-efficacy and learning through co-design, we ensured the resource was appropriate to adolescents' attitudes and behavior and embedded in their social context.

A participatory research approach is crucial when conducting research in adolescent health with adolescents [4]. There is increasing recognition of adolescence as an important life phase, involving multiple physical and psychosocial changes that influence health behavior and health outcomes into adulthood [5]. Nevertheless, adolescents are not a homogenous group: adolescence ranges in age from 10-25 years and involves broad differences in sociodemographic background, interests, abilities, and gender and sexual identities. When conducting participatory research with adolescents, it is important to collaborate with a group that includes members from all subpopulations of interest who have a stake in the project's outcomes [6].

For participatory research to be maximally effective, adolescents should be involved in the actual research design. This has implications for research planning, with time required for recruiting adolescents to the research team and research training. Acknowledgement and remuneration are required in appreciation of the adolescents' lived experience expertise and time dedicated to research training and design. Adolescents' contributions also need to be in ethics applications to human research ethics committees. When seeking research funding, budgetary provisions need to be available for this preparatory phase of participatory research [7].

Conflicts of Interest

None declared.

References

1. Yang A. Adolescent cocreation in digital health: from passive subjects to active stakeholders. J Med Internet Res 2025 Feb 20;27:e70020.
2. Lewis CC, Taba M, Allen TB, Caldwell PHY, Skinner SR, Kang M, et al. Developing an educational resource aimed at improving adolescent digital health literacy: using co-design as research methodology. J Med Internet Res 2024 Aug 07;26:e49453 [FREE Full text] [doi: 10.2196/49453] [Medline: 39110967]
3. Taba M, Allen TB, Caldwell PH, Skinner SR, Kang M, McCaffery K, et al. Adolescents' self-efficacy and digital health literacy: a cross-sectional mixed methods study. BMC Public Health 2022 Jul 20;22(1):1223 [FREE Full text] [doi: 10.1186/s12889-022-13599-7] [Medline: 35725389]
4. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Participatory design of evidence-based online youth mental health promotion, intervention and treatment. Young and Well Cooperative Research Centre. 2012 Sep. URL: https://www.westernsydney.edu.au/__data/assets/pdf_file/0005/476330/Young_and_Well_CRC_IM_PD_Guide.pdf [accessed 2025-01-31]
5. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jul 11;387(10036):2423-2478 [FREE Full text] [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
6. Simonsen J, Robertson T. Routledge International Handbook of Participatory Design. Oxfordshire, England: Routledge; 2013.
7. Paxino J, Eppich W, Bolton J, Woodward-Kron R, Denniston C. A primer on participatory research for health professional education. Focus on Health Professional Education: A Multi-Professional Journal 2024 Sep 30;25(3):121-132. [doi: 10.11157/fohpe.v25i3.860]

Edited by T Leung; submitted 29.01.25; this is a non–peer-reviewed article;accepted 29.01.25; published 20.02.25.

Please cite as:
Lewis CC, Taba M, Allen TB, Caldwell PH, Skinner SR, Kang M, Henderson H, Bray L, Borthwick M, Collin P, McCaffery K, Scott KM
Authors' Reply: "Adolescent Cocreation in Digital Health: From Passive Subjects to Active Stakeholders"
J Med Internet Res 2025;27:e71897
URL: https://www.jmir.org/2025/1/e71897
doi:10.2196/71897
PMID:

XSL•FO
RenderX

©Callum C Lewis, Melody Taba, Tiffany B Allen, Patrina H.Y Caldwell, S Rachel Skinner, Melissa Kang, Hamish Henderson, Liam Bray, Madeleine Borthwick, Philippa Collin, Kirsten McCaffery, Karen M Scott. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Advancing Digital Health Integration in Oncology

Rai Muhammad Umar Khan[1*], MBBS; Hassan Tariq[1*], MBBS

Punjab Medical College, Faisalabad Medical University, Faisalabad, Pakistan

*all authors contributed equally

Corresponding Author:
Rai Muhammad Umar Khan, MBBS
Punjab Medical College
Faisalabad Medical University
Sargodha Road
Faisalabad, 38800
Pakistan
Phone: 92 41 9210080
Email: raimumerkhan@gmail.com

Related Articles:

Comment on: http://www.jmir.org/2024/1/e55176/

Comment in: http://www.jmir.org/2025/27/e72477/

(J Med Internet Res 2025;27:e70316) doi:10.2196/70316

KEYWORDS

mHealth; user experience; cancer; technology acceptance model; structural equation modeling; health care app; mixed-method study; medical care; digital health care; cancer survivors; disparities; health status; behavioral intervention; clinician

The study titled “User Experience and Extended Technology Acceptance Model in Commercial Health Care App Usage Among Patients With Cancer: Mixed Methods Study” by Park et al [1] provides valuable insights into how user experience, interface satisfaction, and external motivation affect the adoption of health care apps among patients with cancer. This work aligns with long-standing theories, such as those detailed by Davis [2], emphasizing that perceived usefulness and ease of use are crucial determinants of technology acceptance. By incorporating an expanded Technology Acceptance Model, Park et al [1] shed light on several key areas where digital health solutions can be optimized for oncological care.

The importance of user-oriented design is unmistakable. Satisfaction with an interface has been repeatedly linked to greater app engagement [2], and Park et al [1] confirm that clinicians’ endorsements bolster consistent usage patterns. Gaps in digital literacy, particularly among older populations, remain a pressing concern and necessitate further policy-level interventions to promote equitable access. Yet several emerging trends suggest that further innovation and collaboration could amplify the benefits of digital health apps in oncology.

Keesara et al [3] argue that the COVID-19 crisis catalyzed a “digital revolution” in health care, unveiling the potential for remote, app-based oncology services. Harnessing these developments may require additional training and infrastructure, especially in resource-limited settings where telehealth adoption remains uneven.

In a systematic review, Kruse et al [4] highlight that both patient and provider acceptance are crucial to sustaining digital solutions. While Park et al [1] primarily focus on patient factors, encouraging provider buy-in through institutional policies and workflow integration could markedly enhance long-term adherence and improve clinical outcomes.

Ammenwerth et al [5] show that electronic patient portals can encourage patient autonomy, enabling individuals to track health metrics, communicate with clinicians, and gain personalized insights. Expanding the features of health care apps beyond simple monitoring to include robust patient portals might address the digital divide, provided adequate educational support is offered.

To conclude, Park et al [1] have laid a solid foundation for understanding how patients with cancer adopt health care apps, emphasizing the centrality of user experience and clinician endorsement. However, broader integration of telemedicine advances, joint patient-provider acceptance, and interactive patient portal functionalities could drive higher engagement rates and better patient outcomes across diverse populations. Although the work provides a basis for assessing app use among Korean cancer survivors, its generalization to other populations needs to be further investigated. Moving forward, policy makers

XSL•FO
RenderX

and health care organizations should focus on closing digital literacy gaps, incorporating patient portals, and reinforcing telehealth infrastructure, thus ensuring that digital oncology solutions effectively serve patients' needs worldwide.

Acknowledgments

Jasper [6], an artificial intelligence–powered language model, has been used to paraphrase selected portions of the manuscript to enhance readability and clarity. The model version integrated into Jasper during the paraphrasing process was GPT-4. Essentially, the use of Jasper was limited to paraphrasing, with no generation of new content, and all scientific claims, data, and interpretations remain entirely our own, derived from rigorous research and analysis.

Conflicts of Interest

None declared.

References

1. Park YE, Tak YW, Kim I, Lee HJ, Lee JB, Lee JW, et al. User experience and extended technology acceptance model in commercial health care app usage among patients with cancer: mixed methods study. J Med Internet Res 2024 Dec 18;26:e55176 [FREE Full text] [doi: 10.2196/55176] [Medline: 39693615]
2. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
3. Keesara S, Jonas A, Schulman K. COVID-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
4. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
5. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
6. Jasper. URL: https://www.jasper.ai/ [accessed 2025-02-26]

Edited by T Leung; submitted 21.12.24; this is a non–peer-reviewed article;accepted 25.02.25; published 07.03.25.

Please cite as:
Khan RMU, Tariq H
Advancing Digital Health Integration in Oncology
J Med Internet Res 2025;27:e70316
URL: https://www.jmir.org/2025/1/e70316
doi:10.2196/70316
PMID:40053796

©Rai Muhammad Umar Khan, Hassan Tariq. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: Advancing Digital Health Integration in Oncology

Yura Lee[1*], MD, PhD; Ye-Eun Park[1*], BA

Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

*all authors contributed equally

Corresponding Author:
Yura Lee, MD, PhD
Department of Information Medicine
Asan Medical Center
University of Ulsan College of Medicine
88 Olympic-ro 43-gil, Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 2 3010 1498
Fax: 82 2 3010 2531
Email: haepary@amc.seoul.kr

Related Articles:

Comment on: http://www.jmir.org/2024/1/e55176/

Comment on: http://www.jmir.org/2025/1/e70316/

(J Med Internet Res 2025;27:e72477) doi:10.2196/72477

KEYWORDS

mHealth; user experience; cancer; technology acceptance model; structural equation modeling; health care app; mixed-method study; medical care; digital health care; cancer survivors; disparities; health status; behavioral intervention; clinician

We appreciate the in-depth review and the critical perspectives provided on our study, "User Experience and Extended Technology Acceptance Model in Commercial Health Care App Usage Among Patients With Cancer: Mixed Methods Study" [1]. This reply aims to address and expand upon the discussion points raised in the letter [2], particularly those concerning the broader implications of telehealth and digital health solutions.

The point raised by Umar Khan and Tariq [2] about enhancing telehealth solutions to address accessibility issues is pivotal. Our study, while primarily focusing on user engagement with health care apps among cancer survivors, indirectly highlighted the urban-rural access disparities and digital literacy gaps. These findings suggest an urgent need for comprehensive telehealth infrastructure that is sensitive to regional disparities. Our data indicated a higher usage rate of health care apps among urban dwellers, likely due to better access to digital resources. This observation underscores the importance of policy interventions aimed at improving digital literacy among older and rural populations, which could significantly bolster the effectiveness of telehealth solutions [3].

Furthermore, the necessity of both patient and provider acceptance for sustained digital health engagement was a key finding in our research. Our study underscored the critical role of clinician endorsements as external motivators that significantly influence both the intention to use and the actual usage of health care apps. However, the real-world application of digital tools in clinical settings often reveals a gap between controlled study environments and everyday clinical practice. In addressing this, we aimed to minimize such discrepancies by using commercial apps without structured reminders, thus fostering more natural user engagement.

The suggestion to integrate robust patient portal features aligns well with our findings on the importance of user-friendly interfaces and perceived utility in promoting app engagement. While our study did not specifically investigate patient portals, incorporating such features could enhance patient autonomy by facilitating access to personalized health information and improving communication with health care providers. To further enhance patient engagement, integrating approaches such as shared decision-making and co-design into digital health tools would be beneficial, fostering a user-centered design that accommodates patient preferences and needs [4,5].

Despite its focus on patients with cancer, our study's application of structural equation modeling to identify key factors influencing digital health acceptance is significant, particularly as patient-physician joint decision-making is crucial for this patient group. We also acknowledge the call for further research

XSL•FO
RenderX

to assess the generalizability of our findings. Our study's cohort, primarily sourced from tertiary referral hospitals in South Korea, may not fully represent the broader population of cancer survivors. This limitation, coupled with South Korea's unique health care context, may affect the external validity of our results. Future studies should aim to include a more diverse participant pool to provide insights that are applicable on a global scale.

In conclusion, we are thankful for the thoughtful critique and the opportunity to engage in this enriching dialogue, which has reinforced the importance of a multidisciplinary approach in advancing digital health care.

Conflicts of Interest

None declared.

References

1. Park YE, Tak YW, Kim I, Lee HJ, Lee JB, Lee JW, et al. User experience and extended technology acceptance model in commercial health care app usage among patients with cancer: mixed methods study. J Med Internet Res 2024 Dec 18;26:e55176 [FREE Full text] [doi: 10.2196/55176] [Medline: 39693615]
2. Umar Khan RM, Tariq H. Advancing digital health integration in oncology. J Med Internet Res 2025. [doi: 10.2196/70316]
3. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
4. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb 04;32(2):223-231. [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
5. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]

Edited by T Leung; submitted 10.02.25; this is a non–peer-reviewed article;accepted 25.02.25; published 07.03.25.

Please cite as:
Lee Y, Park YE
Authors' Reply: Advancing Digital Health Integration in Oncology
J Med Internet Res 2025;27:e72477
URL: https://www.jmir.org/2025/1/e72477
doi:10.2196/72477
PMID:40053760

©Yura Lee, Ye-Eun Park. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Enhancing Digital Health Interventions for Medication Adherence: Considerations for Broader Applicability and Long-Term Impact

ShanShan Du[1], MD; Yining Zhao[2], MD

[1]Renhe Rehabilitation Hospital, Hubei, China

[2]The First Hospital of Shanxi Medical University, Shanxi, China

Corresponding Author:
ShanShan Du, MD
Renhe Rehabilitation Hospital
Renhe Street
Hubei
China
Phone: 86 15767853433
Email: 1547838432@qq.com

Related Article:

Comment on: http://www.jmir.org/2024/1/e65010/

Abstract

-

(J Med Internet Res 2025;27:e69204) doi:10.2196/69204

KEYWORDS

mobile apps; digital health; atrial fibrillation; anticoagulants; medication adherence; mobile phone

We appreciate the opportunity to read the recently published article by Yoon et al [1], titled Smartphone App for Improving Self-Awareness of Adherence to Edoxaban Treatment in Patients With Atrial Fibrillation (ADHERE-App Trial): Randomized Controlled Trial. The study addresses an important clinical challenge—improving medication adherence in atrial fibrillation patients using digital health interventions. While the findings are valuable, we would like to highlight some additional limitations and considerations that we believe warrant further discussion.

First, the study population was limited to individuals familiar with Android smartphones and the Korean language, likely excluding older adults with lower digital literacy. This presents a significant limitation, as older adults, who often make up a substantial proportion of patients requiring anticoagulation therapy, frequently struggle to adopt and effectively use smartphone-based interventions. For this demographic, navigating apps, managing settings, and understanding instructions on a smartphone can already be challenging, and unfamiliarity with technology exacerbates these difficulties. Such barriers may significantly restrict the practicality and accessibility of these interventions for those who stand to benefit the most. Expanding the target population to include individuals with varying levels of digital literacy and providing tailored support for older adults would greatly enhance the inclusivity and real-world applicability of future research. Additionally, the exclusive focus on edoxaban limits the generalizability of the findings to patients using other anticoagulants with differing dosing regimens and side effects. Broadening the intervention to encompass a more diverse patient population and various anticoagulants would strengthen the external validity and clinical relevance of the study.

Second, the 6-month follow-up period is another critical consideration, as adherence is a dynamic behavior that often fluctuates over time. While the study demonstrated short-term benefits, long-term follow-up is essential to determine whether these improvements are sustained and lead to meaningful clinical outcomes, such as reduced stroke or bleeding risks [2]. Moreover, relying on pill counts as the primary adherence metric, while practical, does not account for the accuracy of timing or dosage. Using more precise tools, such as electronic medication event monitoring systems, could provide deeper insights into adherence patterns and their clinical implications.

Lastly, although the intervention design is innovative, it could be improved to enhance engagement. Incorporating personalized feedback or adaptive features tailored to individual adherence behaviors might increase user retention and app effectiveness,

XSL•FO
RenderX

especially given the higher dropout rates observed in the intervention group. Additionally, integrating the app with complementary strategies, such as caregiver notifications or telemedicine support, could make the intervention more accessible to patients with varying levels of technological proficiency. Addressing these limitations would allow future research to more comprehensively evaluate the app's practicality and clinical impact.

In conclusion, while Yoon et al [1] provide a well-executed study demonstrating the potential of smartphone apps to improve medication adherence, addressing the above considerations could further enhance the clinical relevance and applicability of their findings. We commend the authors for their contributions to this evolving field and look forward to future research addressing these important aspects.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Smartphone App for Improving Self-Awareness of Adherence to Edoxaban Treatment in Patients With Atrial Fibrillation (ADHERE-App Trial): Randomized Controlled Trial" declined to respond to this letter.

References

1. Yoon M, Lee JH, Kim I, Lee J, Kim M, Kim H, et al. Smartphone App for Improving Self-Awareness of Adherence to Edoxaban Treatment in Patients With Atrial Fibrillation (ADHERE-App Trial): randomized controlled trial. J Med Internet Res 2024 Nov 21;26:e65010 [FREE Full text] [doi: 10.2196/65010] [Medline: 39570659]
2. Hatori Y, Sakai H, Hatori N, Kunishima T, Namiki A, Shimizu M, et al. Long-term outcome and risk factors associated with events in patients with atrial fibrillation treated with oral anticoagulants: The ASSAF-K registry. J Cardiol 2023 Apr;81(4):385-389 [FREE Full text] [doi: 10.1016/j.jjcc.2022.08.012] [Medline: 36109256]

Edited by T Leung; submitted 24.11.24; this is a non–peer-reviewed article;accepted 31.01.25; published 14.03.25.

Please cite as:
Du S, Zhao Y
Enhancing Digital Health Interventions for Medication Adherence: Considerations for Broader Applicability and Long-Term Impact
J Med Internet Res 2025;27:e69204
URL: https://www.jmir.org/2025/1/e69204
doi:10.2196/69204
PMID:

©ShanShan Du, Yining Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Impact of Social Media Influencers on Amplifying Positive Public Health Messages

Gerard Thomas Flaherty[1*], MD, MSc, PhD; Ryan Michael Mangan[1*], BSc

School of Medicine, College of Medicine, Nursing and Health Sciences, Ollscoil na Gaillimhe – University of Galway, Galway, Ireland

[*]all authors contributed equally

Corresponding Author:
Gerard Thomas Flaherty, MD, MSc, PhD
School of Medicine
College of Medicine, Nursing and Health Sciences
Ollscoil na Gaillimhe – University of Galway
University Road
Galway, H91TK33
Ireland
Phone: 353 91495469
Email: gerard.flaherty@universityofgalway.ie

Related Article:

Comment on: https://www.jmir.org/2024/1/e56651/

(J Med Internet Res 2025;27:e73062) doi:10.2196/73062

KEYWORDS

social media; COVID-19; vaccination; personal brands; public health; wellness; global health; pandemic; Twitter; tweets; vaccine; longitudinal design; wellness influencers; hand annotation; antivaccination; infodemiology

To the Editor-in-Chief,

Social media platforms have become powerful tools for disseminating health information. The recent study published in this journal by O'Brien et al [1] was of great interest. The authors explored the role of social media influencers in sharing health content, assessing their impact on online followers [1]. While the study found that many social media influencers resisted public health campaigns, especially those regarding COVID-19 vaccine uptake, it also highlighted their capacity to promote vaccine uptake when they are aligned with public health messaging [1]. Drawing from our collective experience in this field, we wish to highlight the significant potential of social media influencers in spreading positive public health messages, provide examples of their influence, and propose strategies to maximize their effectiveness in public health campaigns.

Social media influencers play a critical role in shaping public opinion on a wide range of health topics, including vaccinations, cancer screening, lifestyles, and travel safety [2-4]. Social media influencers can effectively counter vaccine hesitancy caused by misinformation, leading to improvements in community attitudes toward influenza vaccines among ethnic minority populations, for example [3]. Additionally, social media use has been linked to improvements in health and food literacy, particularly among schoolchildren [2]. Furthermore, health professionals who create content themselves recognize the value of social media influencers in disseminating information, especially when these influencers work within the boundaries of accurate and responsible messaging [4].

The so-called accessibility influencers also play a role in raising awareness and advocating for inclusive digital experiences by transforming complex health information into more engaging and understandable content [5]. Their work demonstrates the potential for social media influencers to drive positive societal change. The field of accessibility has evolved from its initial focus on web accessibility to addressing diverse user needs across digital platforms. While influencers focus on raising awareness and advocating for accessibility, they do not have a formal role in the development of standards or policies. Influencers' personal experiences and community engagement underscore the importance of collaboration, demonstrating how social media can be harnessed for the public good [5].

Health expert content creators provide valuable insights into the evolving role of social media influencers in health communication. While they acknowledge the risks of misleading health content, which could confuse audiences or make health information less applicable to a broader population [4], they have also suggested strategies to guide social media influencers more effectively. These include improving public health and social media literacy to help audiences critically evaluate online health content, providing training for content creators to enhance

XSL•FO
RenderX

the accuracy of their messages, and strengthening legal or platform-based regulations to curb the spread of misinformation. They also advocate the verification of health professionals' credentials and encourage them to take a more active role on social media [4].

There is a need for further research to fully understand the impact of social media influencers on health communication. The report from O'Brien et al [1] adds to the ongoing discussion about the potential role of digital influencers and emphasizes the need for a cooperative approach to ensure that their influence is used constructively for the benefit of public health.

Authors' Contributions

GTF led on conceptualization, supervision, and editing of the draft manuscript. RMM led on research, data interpretation, and preparation of the first draft. Both authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Wellness Influencer Responses to COVID-19 Vaccines on Social Media: A Longitudinal Observational Study" declined to respond to this letter.

References

1. O'Brien G, Ganjigunta R, Dhillon PS. Wellness influencer responses to COVID-19 vaccines on social media: a longitudinal observational study. J Med Internet Res 2024 Nov 27;26:e56651 [FREE Full text] [doi: 10.2196/56651] [Medline: 39602782]
2. Sercu L. Social media influencers, health literacy, and food literacy: a correlational study among adolescents. Int J Environ Res Public Health 2024 Dec 06;21(12):1629 [FREE Full text] [doi: 10.3390/ijerph21121629] [Medline: 39767469]
3. Mangan RM, Flaherty GT. The advent of social media influencer tourism: travel health risks and opportunities. J Travel Med 2021 Dec 29;28(8):taab140. [doi: 10.1093/jtm/taab140] [Medline: 34494115]
4. Kaňková J, Binder A, Matthes J. Helpful or harmful? navigating the impact of social media influencers' health advice: insights from health expert content creators. BMC Public Health 2024 Dec 18;24(1):3511 [FREE Full text] [doi: 10.1186/s12889-024-21095-3] [Medline: 39696170]
5. Mohamed A, Rangraz M, Gulliksen J. Accessibility influencers: a netnographic qualitative study. Stud Health Technol Inform 2024 Nov 18;320:428-435. [doi: 10.3233/SHTI241037] [Medline: 39560269]

Edited by T Leung; submitted 24.02.25; this is a non–peer-reviewed article;accepted 12.03.25; published 21.03.25.

Please cite as:
Flaherty GT, Mangan RM
Impact of Social Media Influencers on Amplifying Positive Public Health Messages
J Med Internet Res 2025;27:e73062
URL: https://www.jmir.org/2025/1/e73062
doi:10.2196/73062
PMID:

©Gerard Thomas Flaherty, Ryan Michael Mangan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease

Siqi Zhang[1], MM; Tielong Chen[2], MD

[1]Hangzhou TCM Hospital of Zhejiang Chinese Medical University (Hangzhou Hospital of Traditional Chinese Medicine), Hangzhou, China

[2]Department of Cardiology, Hangzhou TCM Hospital of Zhejiang Chinese Medical University (Hangzhou Hospital of Traditional Chinese Medicine), Hangzhou, China

Corresponding Author:
Tielong Chen, MD
Department of Cardiology
Hangzhou TCM Hospital of Zhejiang Chinese Medical University (Hangzhou Hospital of Traditional Chinese Medicine)
453 Stadium Road
Hangzhou, 310007
China
Phone: 86 13858064178
Email: ctlktz@163.com

Related Articles:

Comment on: https://www.jmir.org/2024/1/e56552

Comment in: https://www.jmir.org/2025/1/e70247

(J Med Internet Res 2025;27:e69927) doi:10.2196/69927

KEYWORDS

remote exercise rehabilitation; SCeiP model; coronary heart disease; promotion strategy; home rehabilitation

We read with great interest the study by Xu et al [1] examining the use of the "SCeiP" model to guide exercise rehabilitation in patients with coronary artery disease. The authors assessed the feasibility of a personalized exercise rehabilitation program informed by the "SCeiP" model and suggested that telehealth interventions could enhance both patients' adherence to exercise prescriptions and their overall understanding of rehabilitation. As social workers, we attach paramount importance to the role of cardiac rehabilitation training in improving patients' quality of life. Even with low participation rates, such interventions continue to yield significant benefits. Nonetheless, certain limitations merit further consideration.

First, the intervention group's personalized exercise program, supported by the "SCeiP" model and facilitated through WeChat groups, exercise logs, and activity-tracking bracelets, may have been influenced by individual variations among participants. Such differences could introduce bias. Moreover, the reliance on self-reported data, often in the form of questionnaires and scales, can introduce subjectivity. Overly optimistic or pessimistic self-assessments may compromise data accuracy and obscure the true impact of the intervention.

Second, the study's exercise intervention focused predominantly on a single form of exercise. Future research should consider incorporating a wider array of exercise modalities and exploring their synergistic effects. For instance, aerobic training [2], resistance exercises [3], and aquatic activities [4] have all shown promise in inhibiting the progression of coronary artery disease. Different exercise types may yield distinct benefits for specific patient groups. Franklin et al [5] noted that the cardiac demands of any physical activity depend on the individual's metabolic capacity. Accordingly, future investigations could refine their methods to account for differing responses to various exercise modalities.

Third, while the study demonstrated that 1- and 3-month interventions significantly improved patients' cognitive levels regarding cardiac rehabilitation, their sustainability and long-term clinical implications remain uncertain. Further data collection and longer follow-up periods are essential to determine whether cognitive gains translate into substantive health improvements. Additionally, the absence of subgroup analyses or baseline adjustments for variables such as age, gender, and disease severity may limit the generalizability of the results, despite the study's use of randomized group assignments.

In conclusion, we greatly value the work of Xu et al [1], as their innovative application of the "SCeiP" model presents a new frontier for designing exercise rehabilitation programs for patients with coronary artery disease. Although further

XSL•FO
RenderX

exploration is needed to diversify exercise strategies and evaluate long-term outcomes, these findings support the potential clinical and economic benefits of tele-rehabilitation technologies and offer a vital theoretical foundation for integrating the “SCeiP” model into future cardiac rehabilitation protocols.

Conflicts of Interest

None declared.

References

1. Xu D, Xu D, Wei L, Bao Z, Liao S, Zhang X. The effectiveness of remote exercise rehabilitation based on the "SCeiP" model in homebound patients with coronary heart disease: randomized controlled trial. J Med Internet Res 2024 Nov 05;26:e56552 [FREE Full text] [doi: 10.2196/56552] [Medline: 39499548]
2. Parry-Williams G, Sharma S. The effects of endurance exercise on the heart: panacea or poison? Nat Rev Cardiol 2020 Jul;17(7):402-412. [doi: 10.1038/s41569-020-0354-3] [Medline: 32152528]
3. Wu C, Bu R, Wang Y, Xu C, Chen Y, Che L, et al. Rehabilitation effects of circuit resistance training in coronary heart disease patients: a systematic review and meta-analysis. Clin Cardiol 2022 Aug;45(8):821-830 [FREE Full text] [doi: 10.1002/clc.23855] [Medline: 35758277]
4. Faíl LB, Marinho DA, Marques EA, Costa MJ, Santos CC, Marques MC, et al. Benefits of aquatic exercise in adults with and without chronic disease-a systematic review with meta-analysis. Scand J Med Sci Sports 2022 Mar 24;32(3):465-486. [doi: 10.1111/sms.14112] [Medline: 34913530]
5. Franklin B, Thompson P, Al-Zaiti S, Albert C, Hivert M, Levine B, et al. Exercise-related acute cardiovascular events and potential deleterious adaptations following long-term exercise training: placing the risks into perspective–an update: a scientific statement From the American Heart Association. Circulation 2020 Mar 31;141(13):A. [doi: 10.1161/cir.0000000000000749] [Medline: 32100573]

Edited by T Leung; submitted 11.12.24; this is a non–peer-reviewed article;accepted 28.12.24; published 28.03.25.

Please cite as:
Zhang S, Chen T
The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease
J Med Internet Res 2025;27:e69927
URL: https://www.jmir.org/2025/1/e69927
doi:10.2196/69927
PMID:

©Siqi Zhang, Tielong Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease

Xinyue Zhang[1], MD

Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Corresponding Author:
Xinyue Zhang, MD
Department of Cardiology
The First Affiliated Hospital of Nanjing Medical University
300 Guangzhou Road
Gulou District
Nanjing, 210029
China
Phone: 86 15005162289
Email: 15005162289@163.com

Related Articles:

Comment on: https://www.jmir.org/2024/1/e56552

Comment on: https://www.jmir.org/2025/1/e69927

(J Med Internet Res 2025;27:e70247) doi:10.2196/70247

KEYWORDS

exercise rehabilitation; coronary heart disease; promotion strategy; home rehabilitation

Thank you for the comments on our article, "The Effectiveness of Remote Exercise Rehabilitation Based on the 'SCeiP' Model in Homebound Patients With Coronary Heart Disease: Randomized Controlled Trial" [1]. In response to the questions raised by the authors [2], we wish to provide several clarifications. Please refer to the following.

First, considering the individual variations among participants highlighted by the author, we accounted for this factor prior to the study. Consequently, we conducted a specialized statistical analysis of the baseline data of the study participants (T0: week 0) and found no significant differences between the two groups regarding questionnaires and scales, including levels of cardiac rehabilitation cognition, exercise planning, and exercise input. Furthermore, the objective metrics of exercise days and exercise duration were also important outcome indicators.

Second, to evaluate disease severity, 2 primary baseline metrics—New York Heart Association (NYHA) cardiac function classification and the number of diseased vessels—were extracted. Furthermore, other variables such as age and sex were analyzed in the baseline data and found to have no significant differences, thus minimizing their potential influence on the results.

Other unexplored areas will be addressed in future research. We are grateful for the insightful comments provided by Zhang and Chen [2] on our article.

Conflicts of Interest

None declared.

References

1. Xu D, Xu D, Wei L, Bao Z, Liao S, Zhang X. The effectiveness of remote exercise rehabilitation based on the "SCeiP" model in homebound patients with coronary heart disease: randomized controlled trial. J Med Internet Res 2024 Nov 05;26:e56552 [FREE Full text] [doi: 10.2196/56552] [Medline: 39499548]
2. Zhang S, Chen T. The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease. J Med Internet Res 2025 [FREE Full text] [doi: 10.2196/69927]

XSL•FO
RenderX

Abbreviations

NYHA: New York Heart Association

Edited by T Leung; submitted 18.12.24; this is a non–peer-reviewed article;accepted 28.12.24; published 28.03.25.

Please cite as:
Zhang X
Authors' Reply: The SCeiP Model for Remote Rehabilitation in Homebound Patients With Coronary Heart Disease
J Med Internet Res 2025;27:e70247
URL: https://www.jmir.org/2025/1/e70247
doi:10.2196/70247
PMID:

©Xinyue Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Addressing the Digital Divide Among the Older Population Presents a Substantial Challenge

Su-hang Xie[1], MM, PT

Department of Rehabilitation Medicine, The First Medical Center of PLA General Hospital, Beijing, China

Corresponding Author:
Su-hang Xie, MM, PT
Department of Rehabilitation Medicine
The First Medical Center of PLA General Hospital
28 Fuxing Road, Haidian District, Beijing
Beijing, 100853
China
Phone: 86 13041075258
Email: hxkfxsh@163.com

Related Articles:

Comment on: http://www.jmir.org/2024/1/e56810/

Comment in: https://www.jmir.org/2025/1/e72565/

(J Med Internet Res 2025;27:e69482) doi:10.2196/69482

KEYWORDS

exergame training; Matter of Balance; MOB; pre-frail; tele-exergame; tele-rehabilitation; gaming-based; tele-exercise; physical function; frailty; older adults; aging; physical activity; dementia; CogXergaming; telehealth; dynamic balance

We are pleased to acknowledge the recent publication of the study by Kannan et al [1] in the *Journal of Medical Internet Research*. The study's findings offer valuable insights into the potential for more tailored teletherapy or support interventions for older adults, particularly in the context of rapid technological advancements in health care. Of the 50 older adults initially recruited, 31 successfully completed the full program, yielding a completion rate of approximately 60%. One of the primary reasons for withdrawal was a lack of interest in the content of the intervention. This observation raises the concern that, despite the pervasive presence of advanced technologies in modern life, certain older individuals may be effectively left behind. This phenomenon, known as the digital divide, has garnered increasing attention in recent years [2]. The digital divide refers to the disparity between individuals who have access to and the skills to use information and communication technology (ICT) and those who do not [3]. Although first recognized in the previous century, the digital divide is becoming more pronounced in the context of rapid technological advancements, particularly the frequent updates to ICT. The digital divide has a particularly profound impact on older populations, with many rapidly developing countries experiencing an increase in both the absolute and relative number of individuals affected by the digital divide [4].

In light of the rapid advances in modern medicine, how might we effectively address the digital divide in the older population? It is our contention that the older population should be permitted to transition from a position of passive recipient to that of active initiator and designer of projects. This would enable them to move from being the object of attention to becoming active problem-solvers and contributors to the development process. Before developing a teleprogram application, older adults should be engaged as co-designers, providing feedback from a problem-solving perspective. This ensures that projects are developed from their viewpoint and that they are not treated as passive recipients in the final clinical trial. It is therefore recommended that in future work building on the study by Kannan et al [1], frail older individuals be included as coresearchers in the development and refinement of the CogXergaming program. This inclusion would not only enrich the project but also enhance its generalizability and relevance to the frail older population. Furthermore, Kannan et al [1] did not perform a correlation analysis between the participants' years of education, as reported in the participant demographic characteristics, and their ability to learn effectively—an important factor in addressing the digital divide among older adults. Future studies exploring the correlation between education, learning curves, and application outcomes could provide valuable insights into overcoming the digital divide. In conclusion, CogXergaming presents a novel approach to intervention for frail, older individuals, and it will be of great

XSL•FO
RenderX

interest to explore how this intervention can contribute to bridging the digital divide among older adults.

Authors' Contributions

SX conceptualized the study, drafted the original manuscript, served as the project administrator, and contributed to the review and revision of the manuscript.

Conflicts of Interest

None declared.

References

1. Kannan L, Sahu U, Subramaniam S, Mehta N, Kaur T, Hughes S, et al. Gaming-based tele-exercise program to improve physical function in frail older adults: feasibility randomized controlled trial. J Med Internet Res 2024 Nov 27;26:e56810 [FREE Full text] [doi: 10.2196/56810] [Medline: 39602215]
2. Li Y, Liu C, Sun J, Zhang J, Li X, Zhang Z. The Digital Divide and Cognitive Disparities Among Older Adults: Community-Based Cohort Study in China. J Med Internet Res 2024 Nov 27;26:e59684 [FREE Full text] [doi: 10.2196/59684] [Medline: 39602795]
3. Iwata S, Chen RS. Science and the digital divide. Science 2005 Oct 21;310(5747):405. [doi: 10.1126/science.1119500] [Medline: 16239441]
4. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015 Mar;30(3):284-289 [FREE Full text] [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]

Abbreviations

ICT: information and communication technology

Edited by T Leung; submitted 01.12.24; this is a non–peer-reviewed article;accepted 25.02.25; published 28.03.25.

Please cite as:
Xie SH
Addressing the Digital Divide Among the Older Population Presents a Substantial Challenge
J Med Internet Res 2025;27:e69482
URL: https://www.jmir.org/2025/1/e69482
doi:10.2196/69482
PMID:

©Su-hang Xie. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply: Addressing the Digital Divide Among the Older Population Presents a Substantial Challenge

Lakshmi Kannan[1], PhD; Tanvi Bhatt[2], PhD

[1]Department of Psychology, Bouve College of Health Sciences, Northeastern University, Boston, MA, United States

[2]Department of Physical Therapy, University of Illinois Chicago, Chicago, IL, United States

Corresponding Author:
Tanvi Bhatt, PhD
Department of Physical Therapy
University of Illinois Chicago
1919 W Taylor Street
Room 415
Chicago, IL, 60612
United States
Phone: 1 3123554443
Email: tbhatt6@uic.edu

Related Articles:

Comment on: http://www.jmir.org/2024/1/e56810/

Comment on: https://www.jmir.org/2025/1/e69482/

(J Med Internet Res 2025;27:e72565) doi:10.2196/72565

KEYWORDS

exergame training; Matter of Balance; MOB; pre-frail; tele-exergame; tele-rehabilitation; gaming-based; tele-exercise; physical function; frailty; older adults; aging; physical activity; dementia; CogXergaming; telehealth; dynamic balance

We appreciate the insightful commentary on our study, "Gaming-Based Tele-Exercise Program to Improve Physical Function in Frail Older Adults: Feasibility Randomized Controlled Trial," published in the *Journal of Medical Internet Research* [1], in the letter by Xie [2]. The feedback highlights critical issues regarding the digital divide and the role of older adults in co-designing interventions, which are pivotal for ensuring inclusive and effective digital health solutions.

We agree that addressing the digital divide remains an ongoing challenge, particularly for frail and prefrail older populations. As noted, a lack of interest in the content was a significant contributor to attrition in our study. While we ensured technological accessibility by using familiar platforms (Zoom; Zoom Communications) and step-by-step onboarding, engagement requires more than access—it necessitates meaningful content aligned with participants' preferences and needs.

The suggestion to include older adults as co-designers in the development process is compelling and aligns with the principles of participatory design. Actively involving older adults can ensure interventions are relevant, engaging, and user-friendly. This participatory approach could facilitate a shift from a passive to an active role, empowering older adults to shape technologies that serve them. We acknowledge that future iterations of CogXergaming can benefit from such collaborative input, especially from frail individuals, to improve generalizability and user satisfaction.

Due to the nature of the commercial platforms used in our study, which were chosen for their cost-effectiveness and broad accessibility, tailoring the protocol based on participant feedback was not feasible. However, participants were provided with a choice of games to ensure some degree of personalization within the constraints of the available technology. To enhance the scalability, accessibility, and adoptability of CogXergaming interventions, future efforts should focus on developing cost-effective, customized platforms that allow for greater adaptability and responsiveness to participant preferences and needs. Such platforms could provide more tailored interventions while maintaining affordability, thus bridging gaps in engagement and accessibility.

The commentary further underscores the importance of exploring education levels and their relationship with learning curves. Although we reported participants' education in the section on demographic characteristics, we did not conduct correlation analyses with intervention outcomes. This is an important direction for future research. Investigating the

XSL•FO
RenderX

interplay between education, digital literacy, and intervention effectiveness may yield valuable insights into overcoming barriers imposed by the digital divide.

In conclusion, we appreciate the recognition of CogXergaming's potential to bridge the digital divide among older adults. Future work will focus on integrating participatory design principles, exploring the role of education in digital adoption, and refining interventions to address engagement barriers. This collective effort will contribute to more inclusive and equitable technological solutions for older adults.

Conflicts of Interest

None declared.

References

1. Kannan L, Sahu U, Subramaniam S, Mehta N, Kaur T, Hughes S, et al. Gaming-based tele-exercise program to improve physical function in frail older adults: feasibility randomized controlled trial. J Med Internet Res 2024 Nov 27;26:e56810 [FREE Full text] [doi: 10.2196/56810] [Medline: 39602215]
2. Xie SH. Addressing the digital divide among the older population presents a substantial challenge. J Med Internet Res 2025 Mar 28;27. [doi: 10.2196/69482]

Edited by T Leung; submitted 12.02.25; this is a non–peer-reviewed article;accepted 25.02.25; published 28.03.25.

Please cite as:
Kannan L, Bhatt T
Authors' Reply: Addressing the Digital Divide Among the Older Population Presents a Substantial Challenge
J Med Internet Res 2025;27:e72565
URL: https://www.jmir.org/2025/1/e72565
doi:10.2196/72565
PMID:

©Lakshmi Kannan, Tanvi Bhatt. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Factors Associated With Digital Intervention Engagement and Adherence in Patients With Cancer: Systematic Review

Lucile Montalescot[1,2], PhD; Louise Baussard[1], PhD; Elodie Charbonnier[1], PhD

[1]APSY-V, Université de Nîmes, Nîmes, France

[2]Laboratoire de Psychopathologie et Processus de Santé, Université Paris-Cité, Boulogne-Billancourt, France

Corresponding Author:
Lucile Montalescot, PhD
Laboratoire de Psychopathologie et Processus de Santé
Université Paris-Cité
71 avenue Edouard Vaillant
Boulogne-Billancourt, 92100
France
Phone: 33 1 76 53 29 81
Email: lucile.montalescot@u-paris.fr

Related Article:

Correction of: https://www.jmir.org/2024/1/e52542/

(J Med Internet Res 2025;27:e71370) doi:10.2196/71370

In "Factors Associated With Digital Intervention Engagement and Adherence in Patients With Cancer: Systematic Review" (J Med Internet Res 2024;26:e52542) the authors noted one error.

The following affiliation has been added for all authors:

APSY-v, Université de Nîmes, Nîmes, France

Meanwhile, the following affiliation has been removed for authors LB and EC:

Laboratoire de Psychopathologie et Processus de Santé, Université Paris-Cité, Boulogne-Billancourt, France

Additionally, the Acknowledgments section has been updated to reflect LM's institutional affiliations, from:

This work was funded by myCharlotte.

To:

This work was funded by myCharlotte. LM was affiliated with the Université de Nîmes at the time of the review, and is currently affiliated with the Université Paris-Cité.

The correction will appear in the online version of the paper on the JMIR Publications website on January, 23, 2025 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 16.01.25; this is a non–peer-reviewed article;accepted 16.01.25; published 23.01.25.

Please cite as:
Montalescot L, Baussard L, Charbonnier E
Correction: Factors Associated With Digital Intervention Engagement and Adherence in Patients With Cancer: Systematic Review
J Med Internet Res 2025;27:e71370
URL: https://www.jmir.org/2025/1/e71370
doi:10.2196/71370
PMID:

XSL•FO
RenderX

©Lucile Montalescot, Louise Baussard, Elodie Charbonnier. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study

Gilly A Hendrie[1], PhD; Danielle L Baird[1], BSc; Genevieve James-Martin[1], MSc; Emily Brindal[1], PhD; Paige G Brooker[1], PhD

Commonwealth Scientific and Industrial Research Organisation, Adelaide, Australia

Corresponding Author:
Gilly A Hendrie, PhD
Commonwealth Scientific and Industrial Research Organisation
PO Box 10041
Adelaide, 5000
Australia
Phone: 61 883050662
Email: gilly.hendrie@csiro.au

Related Article:

Correction of: https://www.jmir.org/2025/1/e65122

(J Med Internet Res 2025;27:e71665) doi:10.2196/71665

In “Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study” ([JMIR 2025;27(1):e65122) the authors noted several errors:

In the Methods section of the Abstract, the sentence:

> *Among members with complete data (6602/24,035, 27.5%), patterns of weight loss and gain were examined, and how this related to total weight loss was explored.*

Has been revised to:

> *Among members with complete data (6602/24,035, 27.5%), patterns of weight loss and gain were examined, and how this related to total weight loss and platform use was explored.*

In the Methods section, the following sentence:

> *This analysis focused on weight loss among longer-term members, defined as individuals who had completed at least 12 weeks of the program, had weight data recorded at baseline and approximately 12 weeks (61,514/152,895, 1.4%), and maintained a paid membership for about a year or longer.*

Has been revised to:

> *This analysis focused on weight loss among longer-term members, defined as individuals who had completed at least 12 weeks of the program, had weight data recorded at baseline and approximately 12 weeks (61,514/152,895, 40.23%), and maintained a paid membership for about a year or longer.*

Finally, in Table 2, n values have been added for the categories under the "Sex, statistic, and time point" heading. The row header "Female" has been revised to: "Female (n=19,972); the row header "Male" has been revised to: "Male (n=4063); and the row header "Total" has been revised to: "Total (n=24,035)".

The correction will appear in the online version of the paper on the JMIR Publications website on February 5, 2025, together with the publication of this correction notice. Because this was made after submission to full-text repositories, the corrected article has also been resubmitted tothose repositories.

XSL•FO
RenderX

Submitted 23.01.25; this is a non–peer-reviewed article;accepted 23.01.25; published 05.02.25.

Please cite as:
Hendrie GA, Baird DL, James-Martin G, Brindal E, Brooker PG
Correction: Weight Loss Patterns and Outcomes Over 12 Months on a Commercial Weight Management Program (CSIRO Total Wellbeing Diet Online): Large-Community Cohort Evaluation Study
J Med Internet Res 2025;27:e71665
URL: https://www.jmir.org/2025/1/e71665
doi:10.2196/71665
PMID:

©Gilly A Hendrie, Danielle L Baird, Genevieve James-Martin, Emily Brindal, Paige G Brooker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Internet-Delivered Psychoeducation (SCOPE) for Transition-Aged Autistic Youth: Pragmatic Randomized Controlled Trial

Anna Backman[1], PhD; Lise Roll-Pettersson[2], PhD; Are Mellblom[3], BSc; Elisabet Norman-Claesson[3], BSc; Emma Sundqvist[3], MSc; Eric Zander[4,5], PhD; Sarah Vigerland[5,6], PhD; Tatja Hirvikoski[3,4,5], PhD

[1]Socialstyrelsen, National Board of Health and Welfare, Stockholm, Sweden

[2]Department of Special Education, Stockholm University, Stockholm, Sweden

[3]Habilitation & Health, Stockholm Healthcare Services, Region Stockholm, Stockholm, Sweden

[4]Karolinska Institutet Center of Neurodevelopmental Disorders, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

[5]Centre for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden

[6]Child and Adolescent Psychiatry, Stockholm Healthcare Services, Region Stockholm, Stockholm, Sweden

Corresponding Author:
Anna Backman, PhD
Socialstyrelsen
National Board of Health and Welfare
Rålambsvägen 3
Stockholm, SE 106 30
Sweden
Phone: 46 732445857
Email: anna.n.backman@hotmail.com

Related Article:

Correction of: https://www.jmir.org/2024/1/e49305

(J Med Internet Res 2025;27:e72209) doi:10.2196/72209

In "Internet-Delivered Psychoeducation (SCOPE) for Transition-Aged Autistic Youth: Pragmatic Randomized Controlled Trial" (J Med Internet Res 2024;26:e49305) the authors made two changes.

The affiliation of author Anna Backman has been changed from:

Child and Adolescent Psychiatry, Stockholm Healthcare Services, Region Stockholm, Stockholm, Sweden

to:

Socialstyrelsen, National Board of Health and Welfare, Stockholm, Sweden

In addition, the Corresponding Author's address has been updated to:

Rålambsvägen 3, Stockholm, SE 106 30, Sweden

The correction will appear in the online version of the paper on the JMIR Publications website on February 10, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 05.02.25; this is a non–peer-reviewed article;accepted 05.02.25; published 10.02.25.

Please cite as:
Backman A, Roll-Pettersson L, Mellblom A, Norman-Claesson E, Sundqvist E, Zander E, Vigerland S, Hirvikoski T
Correction: Internet-Delivered Psychoeducation (SCOPE) for Transition-Aged Autistic Youth: Pragmatic Randomized Controlled Trial
J Med Internet Res 2025;27:e72209
URL: https://www.jmir.org/2025/1/e72209
doi:10.2196/72209
PMID:

©Anna Backman, Lise Roll-Pettersson, Are Mellblom, Elisabet Norman-Claesson, Emma Sundqvist, Eric Zander, Sarah Vigerland, Tatja Hirvikoski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Teledentistry Applied to Health and Education Outcomes: Evidence Gap Map

Júlia Nascimento da Silva Mulder[1], DDS[‡]; Marcelo Ramos Pinto[1], DDS; Isabelle Aníbal[1], DDS, MDS; Ana Paula Dornellas[1], DDS, PhD; Deise Garrido[1], DDS, PDS; Camila Huanca[1], DDS, PhD; Ana Estela Haddad[1], DDS, PhD, Prof Dr; Carmen Verônica Mendes Abdala[2], MPA, MLS

[1]Department of Orthodontics and Pediatric Dentistry, University of São Paulo - School of Dentistry, São Paulo, Brazil

[2]Latin American and Caribbean Center on Health Sciences Information, São Paulo, Brazil

[‡]Telehealth and Teledentistry Center FOUSP-SAITE

Corresponding Author:
Júlia Nascimento da Silva Mulder, DDS
Department of Orthodontics and Pediatric Dentistry
University of São Paulo - School of Dentistry
Avenida Professor Lineu Prestes, 2227
São Paulo, 05508-000
Brazil
Phone: 55 11 3091 7854
Email: julia.mulder@usp.br

Related Article:

Correction of: https://www.jmir.org/2024/1/e60590/

(J Med Internet Res 2025;27:e71577) doi:10.2196/71577

In "Teledentistry Applied to Health and Education Outcomes: Evidence Gap Map" (JMIR 2024;26:e60590) the authors made one addition.

In the Acknowledgements paragraph, the text in **bold** has been added:

> *This study was carried out with financial support from the National Council for Scientific and Technological Development (Cnpq) – Federal funds from the Brazilian government granted by the Process: 422896/2021-7; CNPq/MCTI/FNDCT Call No. 18/2021—Track A—Emerging Groups,* ***as well as from the CNPq productivity grant project bench fund CNPq Call No. 03/2021 - Technological Development and Innovative Extension (DT) Grants, Process: 305972/2021-9.*** *The main author (JNdSM) also acknowledges the financial support from the Coordination of Superior Level Staff Improvement, Brazil. This map applies the evidence gap map methodology supported by the Latin American and Caribbean Center on Health Sciences (BIREME/OPAS/OMS), in partnership with the Telehealth Center FOUSP-SAITE (School of Dentistry, University of São Paulo).*

The correction will appear in the online version of the paper on the JMIR Publications website on February 3, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.01.25; this is a non–peer-reviewed article;accepted 28.01.25; published 03.02.25.

Please cite as:
Nascimento da Silva Mulder J, Ramos Pinto M, Aníbal I, Dornellas AP, Garrido D, Huanca C, Haddad AE, Mendes Abdala CV
Correction: Teledentistry Applied to Health and Education Outcomes: Evidence Gap Map
J Med Internet Res 2025;27:e71577
URL: https://www.jmir.org/2025/1/e71577
doi:10.2196/71577
PMID:39899832

XSL•FO
RenderX

©Júlia Nascimento da Silva Mulder, Marcelo Ramos Pinto, Isabelle Aníbal, Ana Paula Dornellas, Deise Garrido, Camila Huanca, Ana Estela Haddad, Carmen Verônica Mendes Abdala. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Digital Interventions for Symptoms of Borderline Personality Disorder: Systematic Review and Meta-Analysis

Julia A B Lindsay[1], BSc; Niall M McGowan[1], PhD; Thomas Henning[2], BA; Eli Harriss[3], MSc; Kate EA Saunders[1,2,4], MRCPsych, DPhil

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[2]School of Medicine and Biomedical Sciences, University of Oxford, Oxford, United Kingdom

[3]Bodleian Health Care Libraries, University of Oxford, Oxford, United Kingdom

[4]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, United Kingdom

Corresponding Author:
Julia A B Lindsay, BSc
Department of Psychiatry
University of Oxford
Warneford Hospital
Warneford Ln
Oxford, OX3 7JX
United Kingdom
Phone: 44 01865 613196
Email: julia.lindsay@psych.ox.ac.uk

Related Article:

Correction of: https://www.jmir.org/2024/1/e54941

(J Med Internet Res 2025;27:e70850) doi:10.2196/70850

In Digital Interventions for Symptoms of Borderline Personality Disorder: Systematic Review and Meta-Analysis J Med Internet Res 2024;26:e54941 there was an omission in published affiliations for one author and the Conflicts of Interest statement. We have also taken the opportunity to update the description of the medical school which has become part of the School of Medicine and Biomedical Sciences.

The affiliations for Saunders KEA were published as:

> *[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom*
>
> *[2]John Radcliffe Medical School, University of Oxford, Oxford, United Kingdom*

This has been amended to:

> *[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom*
>
> *[2]School of Medicine and Biomedical Sciences, University of Oxford, Oxford, United Kingdom*
>
> *[4]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, United Kingdom*

The affiliation for Henning T was published as:

> *[2]John Radcliffe Medical School, University of Oxford, Oxford, United Kingdom*

This has been amended to:

> *[2]School of Medicine and Biomedical Sciences, University of Oxford, Oxford, United Kingdom*

The conflicts of interest statement were published as:

> *JABL, KEAS, and NMM received nonfinancial support from Big Health Ltd in the form of no-cost access to the digital sleep improvement program, Sleepio, for use in clinical research, outside the submitted work. KEAS is supported by the National Institute for Health Research Oxford Biomedical Research Centre based at Oxford University Hospitals National Health Service Trust and the University of Oxford. NMM is a current employee of and holds shares, restricted share units, and/or share options in Compass Pathways plc; this work is not related to Compass Pathways plc.*

This has been amended to:

> *JABL, KEAS, and NMM received nonfinancial support from Big Health Ltd in the form of no-cost access to the digital sleep improvement program, Sleepio, for use in clinical research, outside the submitted work. KEAS is supported by the National Institute of Health Research (NIHR) Oxford Health Biomedical Research Centre. NMM is a current employee of and holds shares, restricted share units, and/or share options in Compass Pathways plc; this work is not related to Compass Pathways plc. The views expressed are those*

XSL•FO
RenderX

of the authors and not necessarily those of the National Health Service, NIHR, or the Department of Health and Social Care.

The correction will appear in the online version of the paper on the JMIR Publications website on February 28, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 03.01.25; this is a non–peer-reviewed article;accepted 06.01.25; published 28.02.25.

Please cite as:
Lindsay JAB, McGowan NM, Henning T, Harriss E, Saunders KEA
Correction: Digital Interventions for Symptoms of Borderline Personality Disorder: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e70850
URL: https://www.jmir.org/2025/1/e70850
doi:10.2196/70850
PMID:40053816

©Julia A B Lindsay, Niall M McGowan, Thomas Henning, Eli Harriss, Kate EA Saunders. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Examining the Effectiveness of Electronic Patient-Reported Outcomes in People With Cancer: Systematic Review and Meta-Analysis

Melissa Betty Perry[1], BSc, MPH; Sally Taylor[1,2], PhD; Binish Khatoon[1,2], PhD; Amy Vercell[1], MSc; Corinne Faivre-Finn[3,4], PhD; Galina Velikova[5,6], PhD; Antonia Marsden[7], PhD; Calvin Heal[7], MSc; Janelle Yorke[1,2,8], PhD

[1]Christie Patient Centred Research, The Christie NHS Foundation Trust, Manchester, United Kingdom

[2]Division of Nursing Midwifery and Social Work, School of Health Sciences, Faculty of Biology, Medicine & Health, The University of Manchester, Manchester, United Kingdom

[3]Division of Cancer Science, The University of Manchester, Manchester, United Kingdom

[4]Clinical Oncology Department, The Christie NHS Foundation Trust, Manchester, United Kingdom

[5]Leeds Institute of Medical Research, University of Leeds, Leeds, United Kingdom

[6]St James's Institute of Oncology, St James's University Hospital, Leeds, United Kingdom

[7]Division of Population Health, Health Services Research & Primary Care, The University of Manchester, Manchester, United Kingdom

[8]School of Nursing, The Hong Kong Polytechnic University, Kowloon, China (Hong Kong)

Corresponding Author:
Janelle Yorke, PhD
School of Nursing
The Hong Kong Polytechnic University
Room GH507a
11 Yuk Choi Rd, Hung Hom
Kowloon, 852
China (Hong Kong)
Phone: 852 92570178
Email: janelle.yorke@polyu.edu.hk

Related Article:

Correction of: https://www.jmir.org/2024/1/e49089

(J Med Internet Res 2025;27:e72687) doi:10.2196/72687

In Examining the Effectiveness of Electronic Patient-Reported Outcomes in People With Cancer: Systematic Review and Meta-Analysis (J Med Internet Res 2024;26:e49089) the authors (notably the study statistician who conducted the analyses, Calvin Heal) noted an error in formulae used to calculate the SE of the standardized mean difference (SMD).

In the Abstract, Results, the following sections:

> *Results: The search identified 10,965 papers, of which 19 (0.17%) from 15 studies were included. The meta-analysis showed an improvement in HRQOL at 3 months, measured by the Functional Assessment of Cancer Therapy–General (SMD 0.28, 95% CI –1.22 to 1.78), and at 6 months, assessed using various HRQOL measures (SMD 0.07, 95% CI –1.24 to 1.39). The results should be interpreted with caution, given the wide 95% CIs. Of the 15 studies, 9 (60%) reported a positive signal on HRQOL, with two-thirds of the studies (n=6, 67%) including tailored patient advice and two-thirds (n=6, 67%) using clinician alert systems.*
>
> *Conclusions: The meta-analysis showed a potential improvement in HRQOL at 6 months and in Functional Assessment of Cancer Therapy–General scores at 3 months for studies that included tailored advice and clinician alerts, suggesting that these elements may improve ePRO effectiveness. The findings will provide guidance for future use and help health care professionals choose the most suitable ePRO features for their patients.*

Have been revised to:

> *Results: The search identified 10,965 papers, of which 19 (0.17%) from 15 studies were included. The meta-analysis showed an improvement in HRQOL at 3 months, measured by the Functional Assessment of Cancer Therapy–General (SMD 0.29, 95% CI 0.19 to 0.39), and at 6 months, assessed using various HRQOL measures (SMD 0.21, 95% CI 0.11 to*

XSL•FO
RenderX

0.30). Of the 15 studies, 9 (60%) reported a positive signal on HRQOL, with two-thirds of the studies (n=6, 67%) including tailored patient advice and two-thirds (n=6, 67%) using clinician alert systems.

Conclusions: The meta-analysis showed an improvement in HRQOL at 6 months and in Functional Assessment of Cancer Therapy–General scores at 3 months for studies that included tailored advice and clinician alerts, suggesting that these elements may improve ePRO effectiveness. The findings will provide guidance for future use and help health care professionals choose the most suitable ePRO features for their patients.

In the main text, in the Methods section, Data analysis, the following text:

There was no observed heterogeneity (I2=0%), but the 95% CIs were wide (95% CI 0%-64.8% and 95% CI 0%-79.2%), suggesting that the true heterogeneity could plausibly be high; therefore, random effects were chosen.

Has been revised to:

There was no observed heterogeneity (I2=0%), but the 95% CIs were wide (95% CI 0%-68% and 95% CI 0%-79%), suggesting that the true heterogeneity could plausibly be high; therefore, random effects were chosen.

In the Results section, Primary Outcome: Quality of Life, the following text:

Improvements in FACT-G scores were reported by Maguire et al [34] (SMD 4.06, 95% CI 2.65-5.46; P<.001) and Velikova et al [24,39] (SE 2.84, 95% CI 13.64-2.37; P=.006).

Has been revised to:

Improvements in FACT-G scores were reported by Maguire et al [34] (mean difference 4.06, 95% CI 2.65-5.46; P<.001) and Velikova et al [24,39] (SE 2.84, 95% CI 13.64-2.37; P=.006).

In the Results section, Meta-Analysis, the following text:

Overall, treatment at 6 months demonstrated an average small improvement (SMD 0.07, 95% CI –1.24 to 1.39), although with a wide 95% CI, suggesting that an effect in either direction is possible. There was relatively little variability in reported effect sizes, which ranged from –0.22 to 0.56, although the 95% CIs surrounding these values were often wide. Of the 15 studies, 5 (33%) were included in a meta-analysis of FACT-G scores at 3 months (Figure 3 [24,27,31,32,34]). Here too, the effect of treatment on FACT-G scores at 3 months showed a small average improvement with a wide 95% CI (SMD 0.28, 95% CI –1.22 to 1.78), again suggesting that the true effect of treatment could be positive, negative, or null.

Has been revised to:

Overall, treatment at 6 months demonstrated an average small improvement (SMD 0.21, 95% CI 0.11 to 0.30). There was relatively little variability in reported effect sizes, which ranged from 0 to 0.56, although the 95% CIs surrounding these values often crossed 0. Of the 15 studies, 5 (33%) were included in a meta-analysis of FACT-G scores at 3 months (Figure 3 [24,27,31,32,34]). Here too, the effect of treatment on FACT-G scores at 3 months showed a small average improvement (SMD 0.29, 95% CI 0.19 to 0.39).

In the Discussion section, the following text:

The meta-analysis showed an improvement in HRQOL at 6 months, although caution should be taken when interpreting the results due to the wide 95% CI. FACT-G scores at 3 months showed a small average improvement but again with a wide 95% CI. Due to the heterogeneity of the studies, specifically the different outcome measures and the different data collection time points, not all studies were included in the meta-analysis. Only 5 (33%) of the 15 studies were included in the FACT-G 3-month meta-analysis, with the weighting predominantly spread across 3 (60%) studies [27,32,34]. These 3 studies all provided advice for patients and sent reports to clinicians. Of the 15 studies, 8 (53%) were included in the HRQOL 6-month meta-analysis, with the majority of the weighting spread across 4 (50%) studies [27-29,32,35,36].

Has been revised to:

The meta-analysis showed an improvement in HRQOL at 6 months. The FACT-G scores at 3 months also showed a small average improvement. Due to the heterogeneity of the studies, specifically the different outcome measures and the different data collection time points, not all studies were included in the meta-analysis. Only 5 (33%) of the 15 studies were included in the FACT-G 3-month meta-analysis, with the weighting predominantly spread across 2 (40%) studies [27,34]. These 2 studies both provided advice for patients and sent reports to clinicians. Of the 15 studies, 8 (53%) were included in the HRQOL 6-month meta-analysis, with the majority of the weighting spread across 3 (38%) studies [27,28,34].

In Conclusions, the following text:

In total, 19 papers of 15 RCTs were identified. Nearly two-thirds (9/15, 67%) of the interventions showed positive effects on HRQOL and symptoms in adults with cancer. However, caution should be taken in interpreting the results of this review due to the heterogeneity in the interventions, outcome measures, and data collection time points, as well as the wide 95% CIs...

Has been revised to:

In total, 19 papers of 15 RCTs were identified. Nearly two-thirds (9/15, 60%) of the interventions showed

XSL•FO
RenderX

positive effects on HRQOL and symptoms in adults with cancer. However, caution should be taken in interpreting the results of this review due to the heterogeneity in the interventions, outcome measures, and data collection time points

Also, Figures 2 and 3 have been revised to the following:

Figure 2. Forest plot of the meta-analysis exploring the effect of the intervention on any health-related quality of life measure at 6 months.

Restricted Maximum Likelihood (REML)

Study	Effect	[95% Confidence interval]		% Weight
Absolom et al.	0.189	-0.012	0.390	22.15
Basch et al. 2016	0.338	0.149	0.526	25.14
Greer et al.	0.223	-0.086	0.533	9.38
Riis et al. 2020	0.000	-0.352	0.352	7.24
Denis et al.	0.560	0.057	1.064	3.54
Velikova et al. 2004	0.124	-0.238	0.487	6.84
Tolstrup et al 2022	0.179	-0.206	0.565	6.05
Maguire et al, 2021	0.103	-0.111	0.316	19.65
Overall effect (REML)	0.207	0.112	0.302	100.00

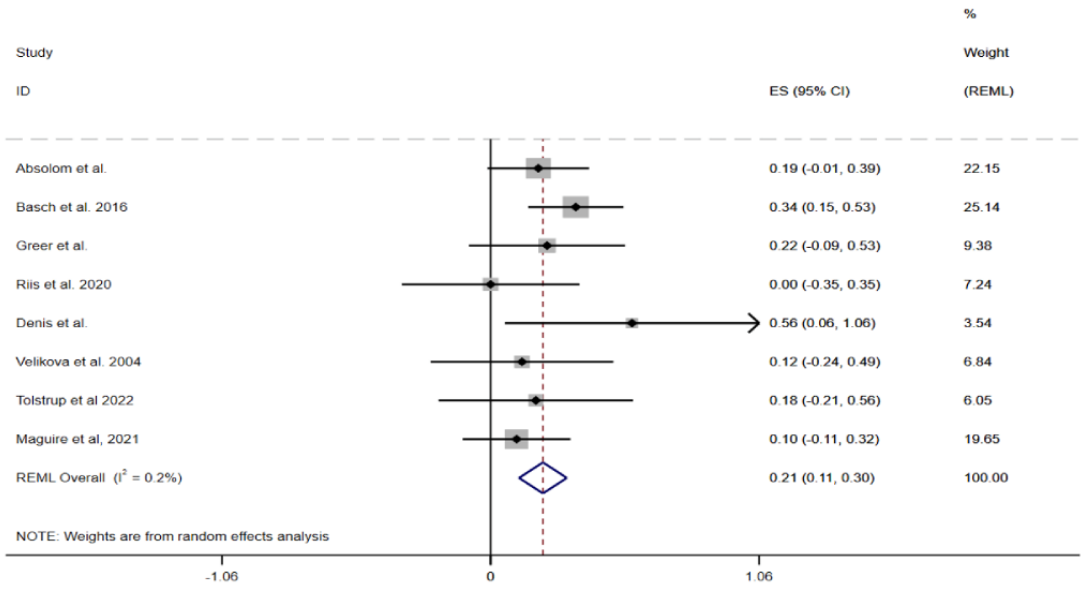

XSL•FO
RenderX

Figure 3. Forest plot of the meta-analysis exploring the effect of treatment on the Functional Assessment of Cancer Treatment-General scores at 3 months.

Restricted Maximum Likelihood (REML)

Study	Effect	[95% Confidence interval]		% Weight
Denis et al.	0.238	-0.179	0.656	6.14
Absolom et al.	0.315	0.116	0.513	27.11
Greer et al.	0.223	-0.086	0.533	11.17
Velikova et al. 2004	0.398	0.058	0.738	9.25
Maguire et al, 2021	0.275	0.123	0.426	46.33
Overall effect (REML)	0.289	0.185	0.392	100.00

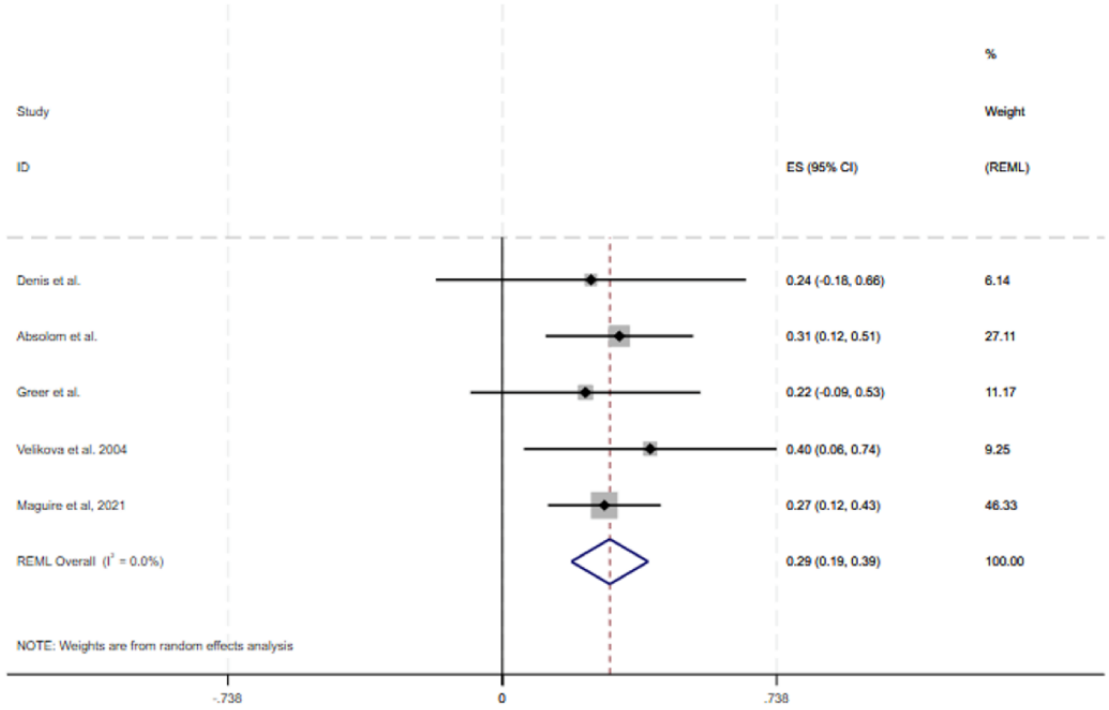

The correction will appear in the online version of the paper on the JMIR Publications website on March 11, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 16.02.25; this is a non–peer-reviewed article;accepted 18.02.25; published 11.03.25.

Please cite as:
Perry MB, Taylor S, Khatoon B, Vercell A, Faivre-Finn C, Velikova G, Marsden A, Heal C, Yorke J
Correction: Examining the Effectiveness of Electronic Patient-Reported Outcomes in People With Cancer: Systematic Review and Meta-Analysis
J Med Internet Res 2025;27:e72687
URL: https://www.jmir.org/2025/1/e72687
doi:10.2196/72687
PMID:

©Melissa Betty Perry, Sally Taylor, Binish Khatoon, Amy Vercell, Corinne Faivre-Finn, Galina Velikova, Antonia Marsden, Calvin Heal, Janelle Yorke. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives

Bryan Cleal[1*], PhD; Yanbing Chen[2*], PhD; Mandy Wäldchen[3], MSc; Hanne Ballhausen[4], MSc; Drew Cooper[5,6,7], MSc; Shane O'Donnell[3], PhD; Christine Knoll[6], MD; Niklas Krug[7]; Klemens Raile[8], MD; Tebbe Ubben[4,9]; Adrian Tappe[4,9], BSc; Dana Lewis[3,10], BA; Ingrid Willaing[1,11], MPH; Timothy Skinner[12], PhD; Katarina Braune[6,7], MD

[1]Diabetes Management Research, Steno Diabetes Center Copenhagen, Herlev, Denmark

[2]Embry-Riddle Aeronautical University, Daytona Beach, FL, United States

[3]School of Sociology, University College Dublin, Belfield, Ireland

[4]#dedoc° Diabetes Online Community, Berlin, Germany

[5]Brain Simulation Section, Department of Neurology with Experimental Neurology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

[6]Charité – Universitätsmedizin Berlin, Institute of Medical Informatics, Berlin, Germany

[7]Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany

[8]Department of Pediatrics, Vivantes Klinikum Neukölln, Berlin, Germany

[9]AndroidAPS, Hamilton, New Zealand

[10]OpenAPS, Seattle, WA, United States

[11]Department of Public Health, Copenhagen University, Copenhagen, Denmark

[12]Australian Centre for Behavioural Research in Diabetes, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Bryan Cleal, PhD
Diabetes Management Research
Steno Diabetes Center Copenhagen
Borgmester Ib Juuls Vej 83
Herlev, 2730
Denmark
Phone: 45 39 68 08 00
Email: bryan.richard.cleal@regionh.dk

Related Article:

Correction of: https://www.jmir.org/2025/1/e48406

(J Med Internet Res 2025;27:e72519) doi:10.2196/72519

In "Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives" J Med Internet Res 2025;27:e48406) the authors noted a few errors in the affiliations.

Author Drew Cooper's affiliations have been changed from:

School of Sociology, University College Dublin, Belfield, Ireland

Diabetes Management Research, Steno Diabetes Center Copenhagen, Herlev, Denmark

to:

Brain Simulation Section, Department of Neurology with Experimental Neurology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany

Charité – Universitätsmedizin Berlin, Institute of Medical Informatics, Berlin, Germany

Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany

Author Christine Knoll's affiliations have been changed from:

School of Sociology, University College Dublin, Belfield, Ireland

Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany

to:

XSL•FO
RenderX

Charité – Universitätsmedizin Berlin, Institute of Medical Informatics, Berlin, Germany

Author Katarina Braune's affiliations have been changed from:

Diabetes Management Research, Steno Diabetes Center Copenhagen, Herlev, Denmark

Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany

to:

Charité – Universitätsmedizin Berlin, Institute of Medical Informatics, Berlin, Germany

Charité – Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Berlin, Germany

The correction will appear in the online version of the paper on the JMIR Publications website on March 24, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 11.02.25; this is a non–peer-reviewed article;accepted 11.02.25; published 24.03.25.

Please cite as:
Cleal B, Chen Y, Wäldchen M, Ballhausen H, Cooper D, O'Donnell S, Knoll C, Krug N, Raile K, Ubben T, Tappe A, Lewis D, Willaing I, Skinner T, Braune K
Correction: Examining the Emotional and Physical Health Impact in Users of Open-Source Automated Insulin Delivery and Sources of Support: Qualitative Analysis of Patient Narratives
J Med Internet Res 2025;27:e72519
URL: https://www.jmir.org/2025/1/e72519
doi:10.2196/72519
PMID:

©Bryan Cleal, Yanbing Chen, Mandy Wäldchen, Hanne Ballhausen, Drew Cooper, Shane O'Donnell, Christine Knoll, Niklas Krug, Klemens Raile, Tebbe Ubben, Adrian Tappe, Dana Lewis, Ingrid Willaing, Timothy Skinner, Katarina Braune. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

How Authoritative Media and Personal Social Media Influence Policy Compliance Through Trust in Government and Risk Perception: Quantitative Cross-Sectional Survey Study

Hua Zhang[1], PhD; Cheng Yang[1], MA; Xiuxian Deng[1], MA; Chunyan Luo[1], MA

Guangxi Minzu University, Nanning, China

Corresponding Author:
Xiuxian Deng, MA
Guangxi Minzu University
188 East Daxue Road
Nanning, 530000
China
Phone: 86 13538076494
Email: 202221252000175@stu.gxmzu.edu.cn

Abstract

Background: Previous studies on public compliance with policies during pandemics have primarily explained it from the perspectives of motivation theory, focusing on normative motivation (trust in policy-making institutions) and calculative motivation (fear of contracting the disease). However, the social amplification of a risk framework highlights that the media plays a key role in this process.

Objective: This study aims to integrate the motivation theory of compliance behavior and the social amplification of risk framework to uncover the "black boxes" of the mechanisms by which normative motivation and calculative motivation influence public policy compliance behavior through the use of media.

Methods: During the COVID-19 pandemic, we conducted a web-based survey of 2309 Chinese citizens from December 4, 2022, to March 21, 2023. We treated the public's policy compliance behavior during the pandemic as the dependent variable. Media use, specifically the use of authoritative media and "we-media," that is, personal social media accounts and media platforms operated by individuals, was set as the independent variable. Trust in government, representing normative motivation, and risk perception, representing calculative motivation, were included as mediating variables. A structural equation model was constructed and analyzed using Stata.

Results: First, the mediation effect of trust in government indicates that the use of authoritative media can enhance trust in government, which significantly increases individuals' policy compliance behavior (β coefficient=0.108, 95% CI 0.080-0.135; *P*<.001). Second, the mediation effect of risk perception shows that the use of we-media heightened individual risk perception, thereby significantly enhancing policy compliance behavior (β coefficient=0.012, 95% CI 0.002-0.021; *P*=.02). Third, the study revealed the "paradox of trust": the chain mediation effect in which authoritative media increased trust in government and reduced risk perception, ultimately decreasing policy compliance behavior (β coefficient=–0.005, 95% CI –0.009 to –0.001; *P*=.008).

Conclusions: By combining the motivation theory of compliance behavior with the social amplification of risk framework in risk communication, we found that trust in government, as a normative motivation, operates through authoritative media, while risk perception, as a calculative motivation, promotes compliance behavior through we-media. In addition, in major crises, the public's use of authoritative media can lead to the paradox of trust: on the one hand, trust in the government increases policy compliance; on the other hand, this trust reduces risk perception, thereby decreasing compliance behavior. Authoritative institutions need to balance providing authoritative information with maintaining the public's risk perception.

(J Med Internet Res 2025;27:e64940) doi:10.2196/64940

KEYWORDS

paradox of trust; risk perception; trust in government; policy compliance; pandemic; authoritative media; social media; China

XSL•FO
RenderX

Introduction

Background

During the early stages of the COVID-19 pandemic, the Chinese government implemented strict control measures, such as mask-wearing mandates, compulsory nucleic acid testing, and vaccination programs, which effectively curbed the spread of the virus. However, in late December 2022, the government relaxed these stringent policies. Established experience has highlighted the importance of public compliance with health policies in containing the virus [1,2]. It is perhaps more crucial to increase citizens' willingness to comply with proactive prevention policies than to formulate these policies [3]. If citizens do not adequately follow epidemic control measures, governmental efforts may be wasted [4]. Current research on compliance with public health policies mainly adopts 2 perspectives from motivation theory. The normative motivation perspective emphasizes that trust in policy-making institutions helps increase citizens' willingness to adhere to public health interventions [5], whereas the calculative motivation perspective argues that compliance behavior stems from the fear of contracting the disease [1,6].

During a pandemic, the public is in a state of "information hunger" and often actively seeks various sources of information to fill their "information vacuum" [7]. The role of information disseminated through different media in shaping psychological processes and individual behaviors has been widely studied across various fields [8-11]. The social amplification of risk framework (SARF) emphasizes that social media, as well as traditional print and broadcast news media, can serve as "social amplification stations," shaping the public's risk perception [12]. Within this framework, countries have used diverse media to promote the importance and urgency of COVID-19 prevention measures, such as vaccine administration, mask-wearing mandates, and home isolation [13,14]. Therefore, understanding the crucial role of different media in promoting policy compliance should be a key focus for future research [15].

The significance of this study lies in the following. This study was conducted during the COVID-19 pandemic, a unique context that holds substantial research value. Media usage habits and their behavioral outcomes often shift due to the specificity of the situation, making conclusions drawn from general social contexts not necessarily applicable to pandemic scenarios [16]. The COVID-19 pandemic provides a distinctive natural experimental condition for investigating the role of media in public health emergencies, offering deeper insights into the dynamic mechanisms of media in risk communication and behavioral influence compared to studies conducted in general social contexts.

The innovation of this paper lies in several key aspects. In the past, authoritative media often played the primary role of information disseminators and communicators. However, with the rise of the internet and "we-media," that is, personal social media accounts and media platforms operated by individuals, this traditional model is gradually being disrupted. Unlike previous studies on risk communication effects that generally lacked attention to media diversity [6] or those that separately explored the effects of risk perception and trust in government on policy compliance behavior [17,18], we conducted a more in-depth and comprehensive exploration. Specifically, we categorized media into authoritative media and we-media and incorporated them, along with trust in government (representing normative motivation) and risk perception (representing calculative motivation), into a structural equation model (SEM) for analysis. This approach allows us to unveil the "black boxes" of how trust in government and risk perception operate between the use of different media and public policy compliance behavior, thereby providing a more comprehensive investigation of how various media influence public compliance with health policies. This multidimensional analysis not only enriches existing research perspectives but also offers more precise and effective guidance for potential public health crises in the future.

Literature Review

Motivational Perspectives on Compliance Behavior

Overview

Previous research on compliance with public health policies during pandemics primarily explains behavior from the perspectives of normative motivation and calculative motivation within motivation theory [1,3]. Normative motivation focuses on individuals' intrinsic adherence to social responsibility and moral norms [19], while calculative motivation emphasizes individuals' cost-benefit analysis during decision-making [20]. Although these 2 motivations largely explain public behavior, in practice, they often intertwine and collectively influence individuals' compliance decisions.

Normative Motivation: The Role of Trust in Government

Normative motivation involves the internal moral beliefs, values, and social norms that guide individual behavior, primarily stemming from trust in the government [17,21], a sense of social responsibility [1], social interaction [20], and social and peer pressure [19].

The characteristics of trust in government, such as policy legitimacy and institutional trust, are key sources of individuals' normative motivations [22]. China effectively curtailed the pandemic through "rapid and decisive" prevention measures, which enhanced public trust in the government [23]. Strong trust in government facilitates public compliance with government policies and the implementation of these policies [5]. The positive effect of trust in government on policy compliance has been demonstrated not only in China but also in Italy, where research indicates that during the COVID-19 pandemic, regions with high trust in the government saw the public actively responding to government calls and reducing their outdoor activities [24]. In addition, a study in Sweden found that higher public trust in the government correlated with greater self-efficacy, leading individuals to be more inclined to follow government strategies for managing public health risks, thereby also helping to mitigate the pandemic [21]. Increased public trust in government can substantially enhance the effectiveness of governmental management during a pandemic, ensuring smooth policy implementation [25,26]. Some studies have found a negative correlation between government trust and compliance behavior, a phenomenon referred to as the

XSL•FO
RenderX

"paradox of trust." Three theoretical frameworks can be identified within this discourse: the "chain reaction of confidence," the "paradox of support," and the "paradox of confidence" [27]. Furthermore, a "trust, confidence, and cooperation" model has been proposed in the context of risk management [28].

With regard to the COVID-19 pandemic, the basic premise of the trust paradox is that the pandemic is viewed as a source of psychological stress, with trust serving as a psychological defense mechanism that can mitigate the complexities and uncertainties often present during a crisis [27,29,30]. Consequently, public trust in the government's ability to manage the pandemic may lead to a reduction in individual protective behaviors because individuals perceive the government's measures to be sufficient, thereby decreasing their motivation to take additional protective actions [31]. Research has also indicated that in environments where public trust in government is low, individuals may feel a stronger need to engage in personal protective behaviors to alleviate their fear, which can, in turn, enhance support for pandemic containment efforts [32]. These varying manifestations of the paradox of trust profoundly reveal the multidimensional complexity of the relationship between trust and compliance behavior. This finding underscores the necessity for more in-depth exploration in this field. Therefore, this study further analyzes the intrinsic mechanisms underlying the complex relationship between trust and compliance behavior, aiming to provide richer theoretical support and empirical evidence for research in this area.

Calculative Motivation: The Impact of Risk Perception

Unlike normative motivation, which focuses on individuals' intrinsic moral beliefs, calculative motivation emphasizes individuals' weighing and evaluation of the consequences of their actions. Within this theoretical framework, prospect theory, deterrence theory, and protection motivation theory are particularly emphasized. Prospect theory reveals that people are influenced by psychological biases, making them likely to take decisions based on the similarity of events to known samples and the availability of relevant information in memory [33]. Deterrence theory posits that strict policies and enforcement will enhance the public's motivation to comply with policies [34]. Protection motivation theory suggests that high levels of risk perception can influence the public's intention to take protective measures [35]. This study mainly uses risk perception as an entry point to investigate public compliance with policies.

Protective motivation theory offers a valuable framework for understanding public policy compliance behavior during pandemics. Within this theory, risk perception refers to people's subjective judgment of the characteristics and severity of a specific risk, which is the initial reaction individuals have when a crisis occurs [36]. The consequences of risks can directly impact the morbidity and mortality rates of populations considered vulnerable. Although risk perception can have negative effects, such as causing social panic and increasing the costs of government-public cooperation [37], many studies have pointed out that risk perception can also produce positive effects. Risk perception plays a critical role in crisis disasters and in encouraging citizens to cooperate with government preventive and mitigation measures [38]. The positive relationship between government intervention and disaster prevention and control is built on the public's active cooperation with government interventions [37,39]. Specifically, in the context of epidemic spread, studies have found that lower COVID-19 risk perception is associated with lower compliance with social distancing measures [18]. Therefore, public awareness of risk is a key factor in promoting compliance with protective behaviors.

Differences in Media Use During Pandemics

Overview

Traditional communication theories assert that risks and risk events primarily flow through 2 main communication networks: the news media and interpersonal communication networks [12]. During a pandemic, the public relies on various media to stay updated on the latest developments and necessary self-protection measures [40-42]. Among these media, television news, newspapers, and official news websites became the primary sources for the public to obtain information related to the COVID-19 pandemic. Simultaneously, social media platforms such as Facebook and Twitter (subsequently rebranded as X) were widely used by the public for information due to their convenience and interactivity [16,31,43,44]. Research shows that traditional (authoritative) media tend to support government and official viewpoints, paying little attention to the dissemination of differing opinions and often providing a less polarized view of events. By contrast, emerging we-media platforms, characterized by multiple topics and platforms, diverse themes, and varying perspectives, facilitate the easy spread of extremist viewpoints [45]. This study, combining its unique characteristics with existing literature [45-47], categorizes the former as authoritative media (also known as mass media or traditional media) and the latter as we-media (also known as new media or online media). This classification can help us better understand the characteristics and patterns of information dissemination and formulate more precise and effective information communication strategies.

Authoritative Media

The official websites of governments and authoritative institutions are considered the primary sources of authoritative information for the public, acting as the main media that directly affect the public's risk perception [48,49]. Official media often use expert opinions to enhance their authority and persuade the public, using methods such as framing effects or agenda setting. These media primarily operate on a "we-publish-you-read" model [50]. Authoritative media typically provide the latest policy updates, professional guidance, and important announcements, making them crucial channels for the public to obtain reliable information [51]. With their long-established "opinion guidance" capabilities, they gain public recognition and trust [52]. These authoritative media are responsible for providing accurate, unbiased, and reliable information and preventing the spread of misinformation [46,53].

XSL•FO
RenderX

We-Media

The rapid development and widespread adoption of the internet have ushered users into a brand new era of we-media. In this we-media age, all users have a microphone and function as both creators and disseminators of public opinion [54,55]. On the one hand, the we-media formed by family and friends provide references for verifying viewpoints. In a broad sense, this involves sharing passed-on perspectives, which can either amplify or diminish the information [12]. On the other hand, we-media represented by social media platforms such as TikTok offer more personalized content. These platforms use algorithms to recommend relevant content to users and have more interaction-enabled features, allowing users to participate in content creation and discussion. Research has shown that social media played a substantial role in spreading misinformation about COVID-19 [56,57]. Misinformation can lead the public to doubt the severity of a pandemic and the effectiveness of prevention measures, thereby reducing the likelihood of following these measures [18,46].

Theoretical Model and Hypotheses

On the basis of the reviewed literature, we constructed a theoretical model that includes the use of media, normative motivation, and calculative motivation related to compliance with public health policies. Specifically, we propose 3 hypotheses and corresponding supplementary hypotheses to examine the process by which the public's use of authoritative media and we-media influences individual policy compliance behavior through trust in government and risk perception.

Research on COVID-19 has found that authoritative media serve as an important channel for the public to obtain reliable information [51], while we-media are characterized by greater diversification and personalization [16,45]. Moreover, the public's trust in government can have varying effects on policy compliance behavior [17,31]. On the basis of the characteristics of authoritative media and we-media, we propose the following hypotheses:

- H1a: The use of authoritative media enhances individual policy compliance behavior by increasing trust in government.
- H1b: The use of we-media decreases individual policy compliance behavior by lowering trust in government.

Kasperson et al [12] proposed the SARF, which suggests that the media can act as "social amplification stations," amplifying or attenuating the public's risk perception. Protective awareness, stakeholder engagement, and risk awareness can positively influence individuals to take protective measures, with risk perception playing a partial mediating role [58]. Therefore, we propose the following hypotheses:

- H2a: The use of authoritative media decreases individual policy compliance behavior by lowering risk perception.
- H2b: The use of we-media increases individual policy compliance behavior by raising risk perception.

Finally, we believe that there is a chain mediation effect of trust in government and risk perception between media use and policy compliance behavior. Members of the public exposed to authoritative media are more likely to trust the government [59]. While some studies found that the use of we-media affects individuals' trust in government [59,60], others found no effect [61]. Moreover, previous studies have found that public trust in government can lower risk perception, leading to reduced compliance with policies [31]. In addition, in environments with low government trust, the public may show greater support for pandemic containment measures [32]. These varying phenomena highlight the "paradox of trust," suggesting the need for further exploration of the complex relationship between government trust and policy compliance.

Consequently, this study constructs a dual mediation model aimed at revealing how different media influence risk perception through trust in government, ultimately shaping attitudes toward policy compliance. Building on prior research, we propose the following hypothesis:

- H3a: The use of authoritative media positively influences trust in government, prompting risk perception to negatively affect policy compliance behavior.

Furthermore, we posit that due to the diversity and personalized content of we-media, the use of we-media may lower public trust in the government, thereby affecting risk perception from a dual perspective, subtly influencing individual policy compliance behavior. Thus, we propose another innovative hypothesis:

- H3b: The use of we-media negatively impacts trust in government, resulting in a positive effect of risk perception on policy compliance behavior.

Interestingly, the "paradox of trust" previously revealed in studies could be triggered if either hypothesis H3a is validated, indicating that the use of authoritative media indeed enhances government trust and leads to a negative influence of risk perception on policy compliance behavior, or hypothesis H3b holds true, suggesting that the use of we-media weakens government trust as anticipated, resulting in a positive influence of risk perception on policy compliance behavior. This phenomenon is not only thought provoking but also ignites our enthusiasm and expectation for exploration and validation.

On the basis of the hypotheses proposed based on the aforementioned theoretical framework, we constructed the theoretical model diagram shown in Figure 1.

XSL•FO
RenderX

Figure 1. Hypothetical model of the theory.

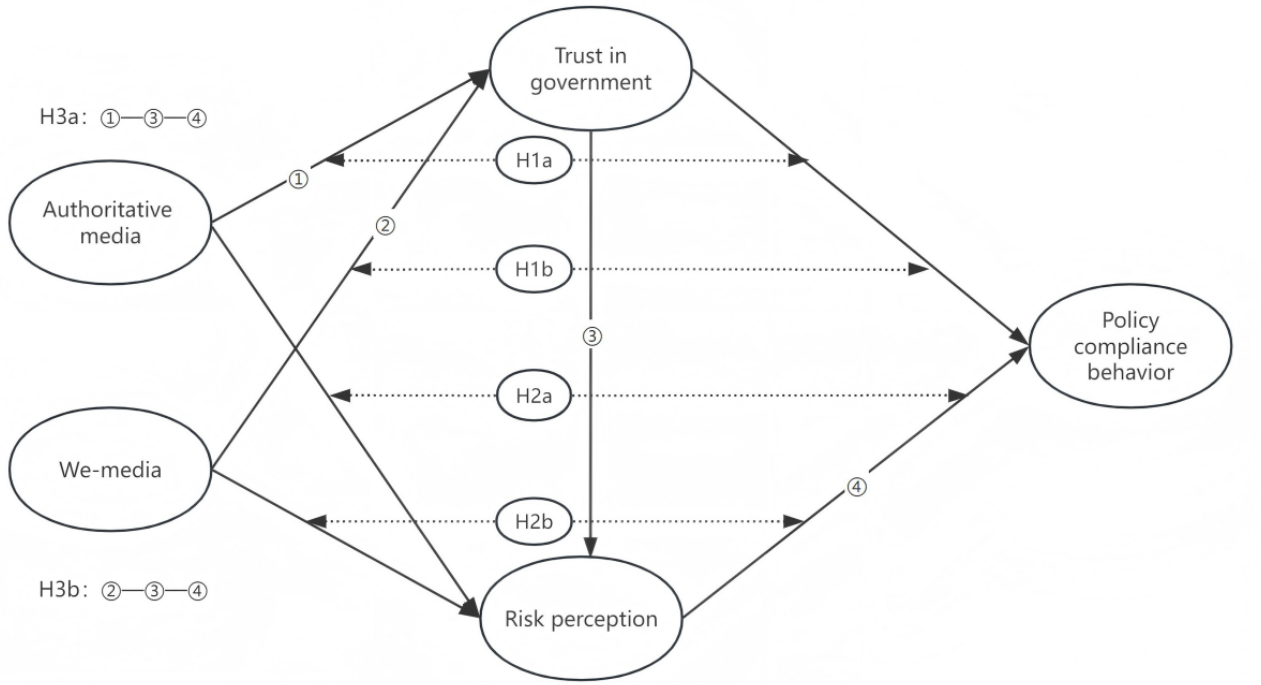

Methods

Data Collection

This study used purposive sampling and snowball sampling methods to collect samples through web-based surveys, with the sample collection period spanning from December 4, 2022, to March 21, 2023. During the questionnaire design phase, we used the professional platform Wenjuanwang to construct the questionnaire framework. To ensure the accuracy and clarity of the questionnaire items, we conducted an extensive literature review and consulted experts from relevant fields. Building on this foundation, we drafted the initial content of the questionnaire.

Subsequently, to further optimize the questionnaire, we conducted a small-scale pilot survey to evaluate the feasibility of the questionnaire items and assess respondents' understanding. On the basis of the pilot survey results, we made necessary adjustments and finalized the questionnaire version. We then promoted the survey through advertisements on social media platforms, specifically WeChat, and recruited participants via web-based snowball sampling. Although the response rate for web-based sample surveys can be low, web-based surveys proved to be a feasible method for conducting research during the pandemic when social distancing and epidemic prevention measures were in place because they are accessible via smart devices and do not have location constraints.

In total, we collected 3690 completed questionnaires. After completing the data cleaning process, we conducted a common method bias (CMB) test to ensure the validity of the data collection. The detection of CMB typically involves exploratory factor analysis (EFA) and the Harman single-factor test. If the first unrotated common factor accounts for >50% of the variance, it may indicate the presence of CMB. The results of our EFA revealed that the variance explained by the first common factor was 23.9%, thereby rejecting the null hypothesis of the Harman single-factor test and indicating that there was no CMB.

Statistical Methods

After completing the data collection, we used Stata (version 16; StataCorp LLC) to perform statistical analysis. First, the basic characteristics of the data were described, followed by the construction of an SEM. The model aimed to explore the relationship between public media use and adherence to public health policies in depth.

Variable Measurement

The questionnaire included items on policy compliance behavior, media use, trust in government, and risk perception, using a 5-point Likert scale for the responses. It also included demographic questions about the respondents' gender, education level, residence type, and annual family income (refer to the Results section for details).

Policy Compliance Behavior

The study used policy compliance behavior as the dependent variable. On the basis of the protection guidelines issued by the National Health Commission of China in August 2021 and existing research [6,62], the questionnaire asked respondents, "How frequently do you follow the following personal protective measures in your daily life?" This question evaluated 4 measures: wearing a mask, washing hands, maintaining a 1-meter distance in public places, and getting vaccinated. Responses ranged from 1=*never* to 5=*always*.

Media Use

The study used "media use" as the independent variable. Information from authoritative media typically has official and

XSL•FO
RenderX

authoritative endorsement from the government, whereas we-media contain more diverse and liberal content. Referring to existing literature [10,46], the questionnaire asked respondents, "How frequently do you use the following media to obtain information in your daily life?" This question assessed the frequency of using various media, including government television news, business websites and news apps (eg, Sina Weibo), health agency websites (eg, expert sources), family and friends, short video apps (eg, TikTok), and social media (eg, QQ and WeChat). Responses ranged from 1=*never* to 5=*always*.

We conducted an EFA on media use (refer to the Results section for details).

Trust in Government

The study used trust in government and risk perception as mediating variables. Trust in government refers to government credibility [63]. When government credibility is high and recognized by the public, the trust in the government or system increases [64]. Therefore, the study referred to the measurement methods of government credibility by Grimmelikhuijsen and Meijer [65] and Ma and Sun [66], combined with the characteristics of Chinese public trust in the government, to measure trust in government by evaluating public satisfaction with specific government actions. The measures included "Your satisfaction with the transparency of government information" and "Your satisfaction with the timeliness of government information disclosure." These indicators capture the public's direct perceptions of the government's ability to disclose information, thereby indirectly reflecting their overall evaluation of government credibility. Scores ranged from 1=*very dissatisfied* to 5=*very satisfied*.

Risk Perception

The measurement indicators for risk perception typically include the severity of the event and its impact scope, with items such as "this event is very severe" and "the impact of this event is very broad" [67,68]. However, during a large-scale crisis, the impact on socioeconomic and personal daily life is especially substantial, making it difficult to accurately measure the heterogeneity of individual risk perception based solely on event severity and impact scope. Therefore, we drew on the research by Zhu et al [69], which uses the direct consequence of risk perception—a sense of security—to measure the cognitive level of risk. The questionnaire asked, "What is the source of your insecurity?" The options included fear of self or family members becoming infected, concern about the lack of material security, worry about employment issues during the public health crisis, and fear of being locked down or quarantined. Respondents rated their concerns from 1=*not worried at all* to 5=*very worried*.

EFA and Reliability and Validity Testing

We conducted EFA using Stata. As policy compliance behavior, risk perception, and trust in government were measured using well-established scales from previous research, only media use underwent EFA. The initial factors were rotated using the Kaiser normalized maximum variance method. On the basis of the EFA results, we labeled factor 1 as "authoritative media use" and factor 2 as "we-media use."

Further confirmatory factor analysis was conducted on each dimension of the scale. The fit indices for the measurement model were as follows: χ^2_{94}=674.1; P<.001; root mean square error of approximation=0.052; comparative fit index=0.961; Tucker-Lewis index=0.950; standardized root mean square residual=0.033; this indicated good overall model fit, supporting the validity and reliability of the measurement results. The Cronbach α values for each dimension of the scale ranged from 0.70 to 0.93, all exceeding the 0.55 threshold commonly used in social science research, and the overall Kaiser-Meyer-Olkin measure of sampling adequacy was 0.77. Factor loadings for each item ranged from 0.48 to 0.95, suggesting moderate to strong associations between the measured items and their respective latent factors. According to Stevens [70], a factor loading of 0.40 is sufficient for explanatory purposes, even with varying sample sizes. Furthermore, Steenkamp and Maydeu-Olivares [71] found that lower factor loadings can still be explanatory, particularly in cases of smaller sample sizes or more complex object of study.

Ethical Considerations

All participants in this study provided informed consent. The study was approved by the Science Ethics and Safety Committee of Guangxi University for Nationalities (Protocol Number: 202410608021). All procedures conducted during the study strictly adhered to the ethical principles outlined in the Declaration of Helsinki (2013).

Results

Data Cleaning Process

We collected 3690 completed questionnaires in total. We used a dual review process to check data quality, focusing on outliers and logical inconsistencies. After cleaning the data, from the 3690 completed questionnaires, we ultimately obtained 2309 (62.57%) valid questionnaires. The flowchart of the data cleaning process is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Flowchart of the data cleaning process.

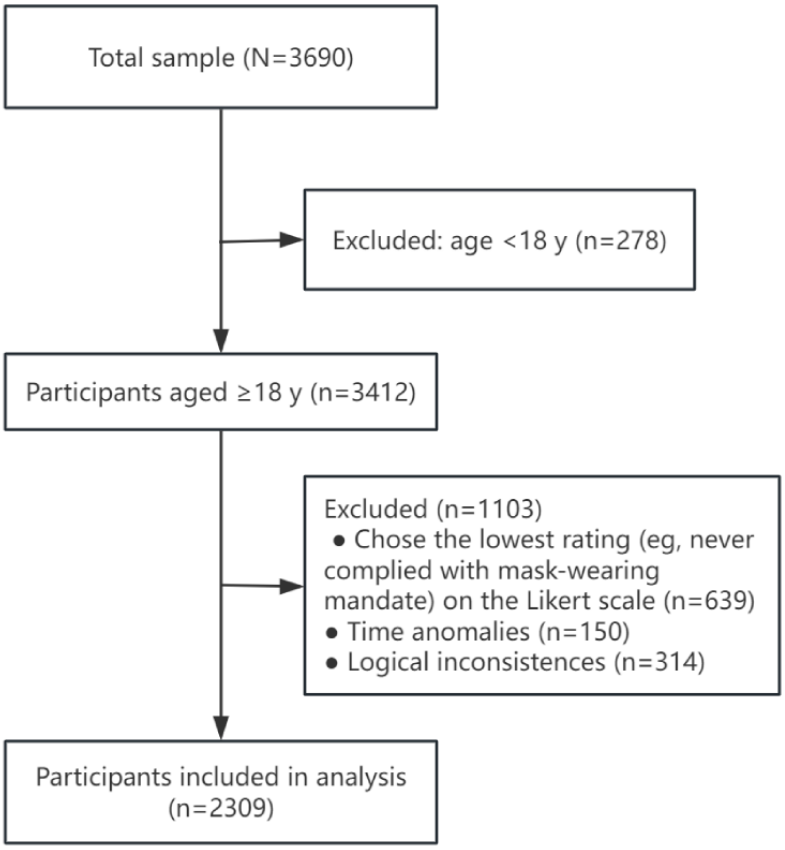

EFA Results

The EFA results showed that the 6 items of media use were extracted into 2 factors. The items "government television news," "business websites and news apps," and "health agency websites" had higher loadings on factor 1, while "social media," "short video apps," and "family and friends" had higher loadings on factor 2 (Table 1). On the basis of existing literature [45-47] and the EFA results, we labeled factor 1 as "authoritative media use" and factor 2 as "we-media use."

Table 1. Results of exploratory factor analysis.

Variable items	Factor 1	Factor 2
Government television news	*0.6396* [a]	0.2452
Business websites and news apps (eg, Sina Weibo)	*0.6725*	0.2942
Health agency websites (eg, expert sources)	*0.4419*	0.2274
Short video apps (eg, TikTok)	0.2817	*0.5995*
Social media (eg, QQ and WeChat)	0.2828	*0.6426*
Family and friends	0.2732	*0.4293*

[a]Higher loadings are indicated in italics.

Descriptive Statistics

The survey questionnaire included items on policy compliance behavior, media use, trust in government, and risk perception. Participants rated their responses on a Likert scale ranging from 1 to 5 (Table 2). The questionnaire also included demographic questions (Table 3).

XSL•FO
RenderX

Table 2. Descriptive statistics of variables: survey questionnaire responses rated on a 5-point Likert scale (n=2309).

Variables and categories	Scores, mean (SD; SE)
Policy compliance behavior	
Wear a mask	4.38 (0.77; 0.01)
Wash hands	4.41 (0.72; 0.01)
Keep distance	3.98 (0.96; 0.01)
Vaccinate	4.65 (0.67; 0.02)
Authoritative media use[b]	
Government television news	3.90 (0.94; 0.01)
Business websites and news apps (eg, Sina Weibo)	3.76 (1.02; 0.01)
Health agency websites (eg, expert sources)	3.12 (1.11; 0.02)
We-media use[c]	
Short video apps (eg, TikTok)	3.99 (1.00; 0.02)
Social media (eg, QQ and WeChat)	4.09 (0.89; 0.02)
Family and friends	3.83 (0.94; 0.02)
Trust in government	
Timeliness of information	3.75 (0.89; 0.01)
Information transparency	3.69 (0.94; 0.01)
Risk perception[e]	
Fear of disease	3.24 (1.17; 0.01)
Fear of lack of security	3.02 (1.18; 0.01)
Worry about jobs	3.40 (1.21; 0.01)
Fear of isolation	3.26 (1.24; 0.01)

[a]Likert scale ratings: 1=*never comply*, 2=*occasionally comply*, 3=*sometimes comply*, 4=*often comply*, and 5=*always comply*.

[b]Likert scale ratings: 1=*never use*, 2=*rarely use*, 3=*sometimes use*, 4=*often use*, and 5=*always use*.

[c]Likert scale ratings: 1=*never use*, 2=*rarely use*, 3=*sometimes use*, 4=*often use*, and 5=*always use*.

[d]Likert scale ratings: 1=*very dissatisfied*, 2=*somewhat dissatisfied*, 3=*neutral*, 4=*somewhat satisfied*, and 5=*very satisfied*.

[e]Likert scale ratings: 1=*not worried at all*, 2=*not very worried*, 3=*somewhat worried*, 4=*quite worried*, and 5=*very worried*.

XSL•FO
RenderX

Table 3. Descriptive statistics of variables: demographics (n=2309).

Variables and categories	Participants, n (%)
Gender	
Women	1657 (71.76)
Men	652 (28.24)
Education level	
Below bachelor's degree	950 (41.14)
Bachelor's degree or higher	1359 (58.86)
Residence type	
City	774 (33.52)
Rural	1535 (66.48)
Annual income, CN ¥ (US $1=CN ¥ 6.7976)	
<20,000	779 (33.74)
20,001-100,000	830 (35.95)
100,001-200,000	506 (21.91)
>200,000	194 (8.4)

Estimation of Fit Indices

Before conducting hypothesis testing, we used Stata to estimate the fit indices for the SEM. The results demonstrated satisfactory model fit (χ^2_{96}=722.3; *P*<.001; root mean square error of approximation=0.053; comparative fit index=0.958; Tucker-Lewis index=0.947; standardized root mean square residual=0.045). These findings indicate that the model meets acceptable standards and effectively fits the sample data without necessitating adjustments or respecification of the original model. Figure 3 presents the SEM along with its associated path coefficients (error terms for all endogenous variables are assumed but not depicted in Figure 3).

Figure 3. Structural equation model and path coefficient diagram. The solid lines indicate significant paths, while dashed lines indicate nonsignificant paths. ***P*<.01, ****P*<.001.

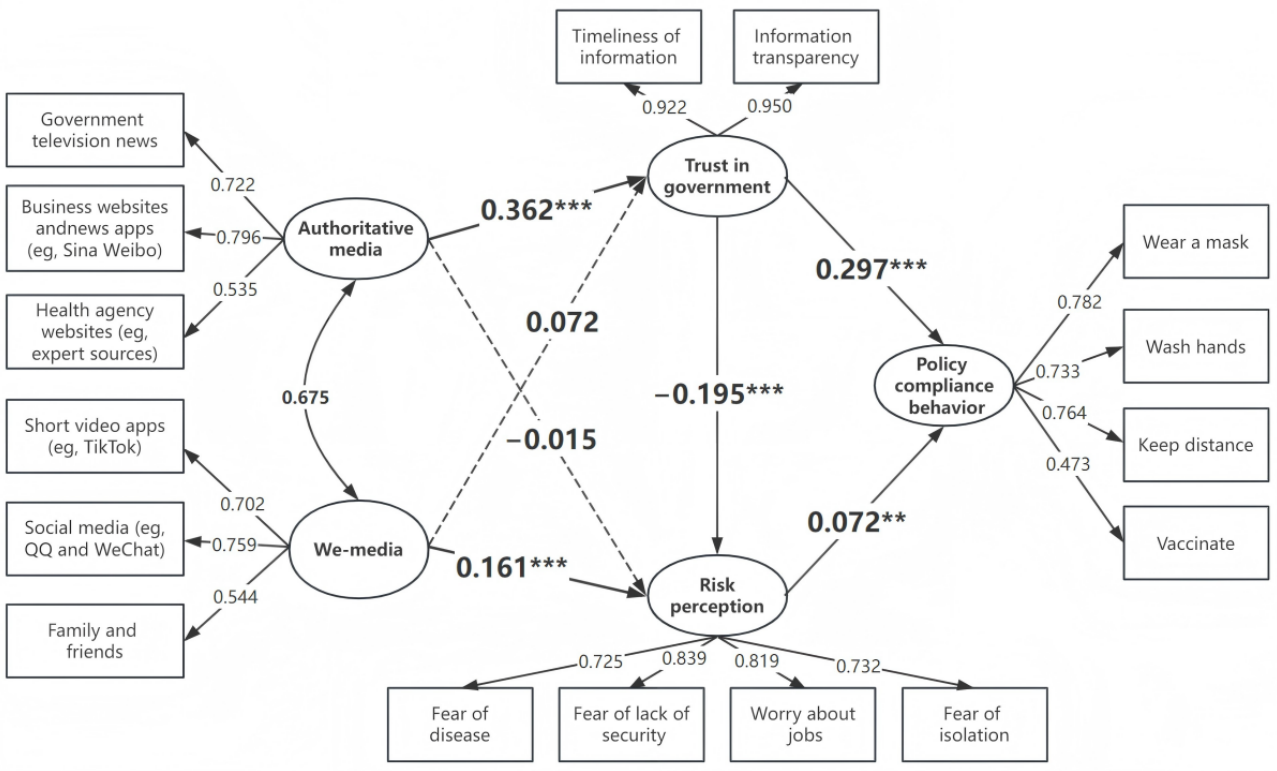

XSL•FO
RenderX

Intermediate Path Test Results

Trust in government plays a significant positive mediating role in the path "authoritative media use→trust in government→policy compliance behavior" (β coefficient=0.108, 95% CI 0.080-0.135; *P*<.001; Table 4). The use of authoritative media effectively enhances public trust in the government, leading to increased compliance with policies. Therefore, H1a is supported. In the path "we-media use→trust in government→policy compliance behavior," although the path coefficient is positive, the mediating effect does not reach statistical significance (β coefficient=0.021, 95% CI –0.001 to 0.044; *P*=.07). We-media use does not substantially affect public trust in the government, thus having no substantial impact on policy compliance behavior. Therefore, H1b is not supported.

Table 4. Intermediate path test results.

Paths	β coefficient (SE; 95% CI)	Critical ratio	*P* value
Authoritative media use→trust in government→policy compliance behavior	0.108 (0.014; 0.080 to 0.135)	7.70	<.001
We-media use→trust in government→policy compliance behavior	0.021 (0.012; –0.001 to 0.044)	1.84	.07
Authoritative media use→risk perception—policy compliance behavior	–0.001 (0.003; –0.007 to 0.005)	–0.36	.72
We-media use→risk perception—policy compliance behavior	0.012 (0.005; 0.002 to 0.021)	2.35	.02
Authoritative media use→trust in government→risk perception→policy compliance behavior	–0.005 (0.002; –0.009 to –0.001)	–2.64	.008
We-media use→trust in government→risk perception→policy compliance behavior	–0.001 (0.001; –0.002 to 0.000)	–1.52	.13
Authoritative media use→trust in governments→risk perception	–0.071 (0.012; –0.093 to –0.048)	–6.09	<.001
We-media use→trust in governments→risk perception	–0.014 (0.008; –0.030 to 0.001)	–1.78	.08

The path coefficient of risk perception between authoritative media use and policy compliance behavior is negative, but the mediating effect does not reach statistical significance (β coefficient=–0.001, 95% CI –0.007 to 0.005; *P*=.72). Authoritative media use does not significantly affect public trust in the government, resulting in no significant impact on policy compliance behavior. Thus, H2a is not supported. By contrast, the path "we-media use→risk perception→policy compliance behavior" is statistically significant and positive (β coefficient=0.012, 95% CI 0.002-0.021; *P*=.02). It shows that the public's use of we-media enhances risk perception, promoting compliance with policies. Therefore, H2b is supported.

The chain mediation path "authoritative media use→trust in government→risk perception→policy compliance behavior" shows a significant chain mediation effect (β coefficient=–0.005, 95% CI –0.009 to –0.001; *P*=.008). Authoritative media use increases trust in government, which in turn reduces risk perception, ultimately leading to decreased individual policy compliance behavior, reflecting the phenomenon of the "paradox of trust." Thus, H3a is supported. By contrast, the chain mediation effect for the path "we-media use→trust in government→risk perception→policy compliance behavior" has a negative path coefficient, but it does not reach statistical significance (β coefficient=–0.001, 95% CI –0.002 to 0.000; *P*=.13). Therefore, H3b is not supported.

Discussion

Overview

This study integrates motivation theory and the SARF to examine the impact of media use on public policy compliance behavior during a pandemic. The research aims to reveal the roles of different media in influencing public behavior and the changes in their effectiveness. The results indicate that the public's use of authoritative media promotes policy compliance behavior by increasing trust in government. By contrast, the use of we-media can indirectly affect individual policy compliance behavior by influencing individual risk perception. The findings also reveal that trust in government and risk perception serve as chain mediators between the use of authoritative media and policy compliance behavior, validating the phenomenon of the paradox of trust.

Specifically, we identified mediating variables that affect individual policy compliance behavior as a result of media use, leading to differences in individual attitudes toward compliance with epidemic prevention policies. This study uncovers the specific mechanisms, or the "black boxes," through which trust in government and risk perception mediate the relationship between media use and individual policy compliance behavior.

"Black Box 1": Normative Motivation (the Bridging Power of Trust in Government)

The study found that trust in government has a substantial mediating effect between the public's use of authoritative media and policy compliance behavior. However, we-media has no substantial impact on trust in government.

Authoritative media disseminate truthful and objective information, which enhances trust in government [60], thereby increasing the public's willingness to comply with policies [24,72]. Specifically, authoritative media typically communicate key information such as prevention measures and response strategies with high consistency and clarity. This clear and consistent risk information fills the "information vacuum" caused by the public's cognitive limitations [7], effectively alleviating information asymmetry. When the public's need for risk information is satisfied, it fosters a stronger sense of trust in government policies.

In China, policy implementation follows a top-down approach [5], making positive policy perception and trust in the

XSL•FO
RenderX

government crucial for individuals to comply with antiepidemic measures and achieve policy effectiveness [26,37]. Looking at examples from different cultural contexts, Swedish citizens exhibit a high level of trust in their government, which makes them more likely to adhere to strategies for managing public health crises, contributing to relatively effective pandemic control [21]. Similarly, studies from Italy indicate that higher levels of trust in government encourage the public to respond more actively to government appeals, resulting in reduced mobility during the pandemic [24]. Therefore, the enhancement of public trust in government can further strengthen the role of authoritative media in promoting individual policy compliance behavior.

The study found that we-media do not substantially influence trust in government, which aligns with previous research [61]. The diverse, emotional, and personalized nature of information on we-media may not be entirely reliable or consistent [16,45]. Furthermore, the decentralized nature of we-media, the rapid cross-regional dissemination of content, and low interpersonal communication costs, combined with limited fact-checking time, increase the likelihood of the spread of false or unverified content [56,57]. This complexity of information may lead to ambiguous or divergent impacts on public trust in government, resulting in we-media not having a substantial effect on trust in government.

"Black Box 2": Calculative Motivation (the Impact of Risk Perception)

The study found that the use of we-media can increase the public's risk perception of diseases, indirectly promoting public policy compliance behavior, consistent with previous research [46]. The results also reveal that the use of authoritative media does not substantially impact individual risk perception.

The public processes risk information obtained through we-media, forming a sense of the significance of the risk, which in turn influences individual risk perception [73]. However, we-media (eg, TikTok and Twitter) may be subject to looser regulation in certain situations, and their accountability mechanisms may not be as robust as those of authoritative media [11]. This increases the likelihood of the rapid spread of misinformation, which may have elevated the public's risk assessment of the pandemic, thereby further influencing their attitudes toward, and compliance with, antipandemic policies [16]. In addition, the study found that the use of authoritative media does not directly affect individual risk perception, differing from previous research [16]. This may be because, in the Chinese context, authoritative media typically emphasize the accuracy and official nature of information, while during major public health crises, the public may prioritize immediacy and highly perceived risk signals, which are often more easily transmitted through we-media. The information conveyed by authoritative media might be perceived by the public as "official responses" rather than direct threats, thus failing to effectively enhance individual risk perception.

The existence of a mediating effect of risk perception indicates the social amplification of risk by we-media, which can qualitatively and quantitatively amplify signals related to risks, enhance risk perception and risk-related behaviors, and intensify the risks and their consequences [12]. Therefore, the quality and accuracy of information dissemination are crucial for public health, and misleading or inaccurate information can lead to incorrect assessments of risk, subsequently affecting individuals' policy compliance behavior.

The Paradox of Trust

The research results indicate that the use of authoritative media does indeed enhance public trust in government, subsequently lowering risk perception and leading to reduced policy compliance behavior, thereby illustrating the paradox of trust. However, we-media has no substantial impact on trust in government.

If the government can implement effective pandemic management strategies and achieve results, people will trust the government's ability to handle public crises, thereby strengthening its management effectiveness in dealing with pandemics [25,26]. However, there is a complex relationship between public trust in government and individual risk perception. Increased trust in government leading to reduced risk perception may decrease public compliance with preventive policies. If people perceive the risk to be low, they might not feel the need to strictly follow prevention measures such as wearing masks and maintaining social distance [31]. A study on the COVID-19 pandemic in Italy showed that when the public was exposed to authoritative information, their trust in the government increased, but their risk perception decreased, thus reducing their likelihood of complying with preventive policies. This could be because confidence in government institutions can alleviate pandemic-related stress. When people have higher trust in their government, they believe that sufficient measures have been taken to protect them, and they might think it is unnecessary to take personal actions to avoid infection and limit the spread of the virus during a pandemic [27]. The paradox of trust has conflicting effects on policy compliance behavior. The public's willingness to cooperate is influenced by both their trust in the government and their perceived level of risk. In addition, as mentioned previously, the diversity and complexity of we-media content may lead to confusion and misunderstandings among the public when receiving information, making it difficult to form a consistent view of trust in the government. This uncertainty in the information received may further undermine the role of we-media in shaping trust in government.

Limitations

The study primarily relied on cross-sectional data, which, while able to describe public psychology and policy compliance behaviors during the data collection period, were limited. The varying stages of the pandemic, changes in government measures, and shifts in public sentiment make trust in policies a dynamic process. Cross-sectional data often fail to capture the lag effects, cumulative impacts, or long-term trends of public health policy outcomes. Future research should consider the long-term effects of dynamic changes in a pandemic on trust and compliance behavior.

Moreover, the limitations of self-reported data, such as social desirability bias and recall bias, may affect respondents' accurate

XSL•FO
RenderX

description of compliance behaviors. Future studies could use more objective measurement methods, such as actual behavior tracking or observational data, to enhance data accuracy. Furthermore, policy compliance behavior is influenced by various factors, including political ideology, media literacy, policy implementation effectiveness, and social opinion, which also warrant further investigation. Subsequent research could introduce these variables to enrich the theoretical framework and test their applicability in different social contexts.

Conclusions

The study identified the mechanisms through which the use of different media influences policy compliance behavior during a pandemic, as well as the phenomenon of the paradox of trust. Public policy compliance behavior is driven by both normative and calculative motivations. These motivations each tend to rely on different sources of information when guiding public behavior. *Normative motivation* emphasizes individual adherence to social norms, with behavior decisions driven by this motivation being more influenced by authoritative media. Information released by the government, public health agencies, and authoritative media, due to its authority, accuracy, and comprehensiveness, becomes an important benchmark for the public to judge the correctness of their actions. Through the dissemination of authoritative information, public health departments can guide the public to form correct perceptions, enhance trust in the government, and thereby promote policy compliance behavior. However, *calculative motivation* follows a different path. Calculative motivation, based on individual cost-benefit analysis, prompts the public to make behavior decisions after evaluating their own risks and benefits. We-media, such as social media, with their immediacy, interactivity, and personalized features, can quickly spread information that resonates emotionally with the public, potentially promoting policy compliance behavior to a greater extent.

In addition, this study quantitatively analyzes the paradox of trust by constructing a theoretical model. We verified the relationship between trust and compliance behavior through structural equation modeling, confirming that high trust does not necessarily lead to high compliance, thereby clarifying the existence of the paradox of trust. This finding challenges the traditional simplistic assumption of a linear correlation between trust and compliance behavior, providing a new perspective for understanding the relationship between trust and compliance behavior and offering valuable references for future research.

Understanding the paradox of trust can help managers better identify and address trust issues within organizations, as well as optimize organizational structures, incentive mechanisms, and communication methods. By delving into the dynamics of the paradox of trust, public health departments can develop effective and reliable communication strategies to address scenarios where, over time, the public perceives a reduced risk, leading to a decline in willingness to comply with pandemic prevention measures. Such strategies can encourage better adherence to public health policies during pandemics.

The study offers several policy recommendations for public health departments: first, public health officials can enhance public trust by regularly disseminating transparent, evidence-based information. Transparency can reduce uncertainty and prevent the spread of misinformation, thereby helping to improve policy compliance. Second, personalizing and tailoring risk information for different audience groups can help the public better understand the severity of risks and the specific threats they face. Third, the paradox of trust suggests that during a pandemic, governments should focus on enhancing the quality and efficiency of public services to strengthen trust in government. They should also use authoritative media to help the public correctly understand and assess risks, preventing the neglect of risks that could lead to reduced compliance with public policies. Finally, public health departments should establish effective 2-way communication channels to promptly convey risk information, address public concerns, and alleviate panic while avoiding excessive optimism. This balanced guidance of the public's risk perception can help mitigate the impact of the paradox of trust.

Conflicts of Interest

None declared.

References

1. Zimmermann BM, Fiske A, McLennan S, Sierawska A, Hangel N, Buyx A. Motivations and limits for COVID-19 policy compliance in Germany and Switzerland. Int J Health Policy Manag 2022 Aug 01;11(8):1342-1353 [FREE Full text] [doi: 10.34172/ijhpm.2021.30] [Medline: 33949815]
2. Spahl W, Pot M, Paul KT. Understanding compliance as multi-faceted: values and practices during the COVID-19 pandemic in Austria. Crit Public Health 2022 Feb 11;32(5):665-676. [doi: 10.1080/09581596.2022.2039379]
3. Guan B, Bao G, Liu Q, Raymond RG. Two-way risk communication, public value consensus, and citizens' policy compliance willingness about COVID-19: multilevel analysis based on a nudge view. Adm Soc 2021 Feb 09;53(7):1106-1149. [doi: 10.1177/0095399721990332]
4. Zhao T, Wu Z. Citizen–state collaboration in combating COVID-19 in China: experiences and lessons from the perspective of co-production. Am Rev Public Adm 2020 Jul 15;50(6-7):777-783. [doi: 10.1177/0275074020942455]
5. Fu C, Liao L, Huang W. Behavioral implementation and compliance of anti-epidemic policy in the COVID-19 crisis. Int J Environ Res Public Health 2021 Apr 04;18(7):3776 [FREE Full text] [doi: 10.3390/ijerph18073776] [Medline: 33916631]

XSL•FO
RenderX

6. Oh SH, Lee SY, Han C. The effects of social media use on preventive behaviors during infectious disease outbreaks: the mediating role of self-relevant emotions and public risk perception. Health Commun 2021 Jul 16;36(8):972-981. [doi: 10.1080/10410236.2020.1724639] [Medline: 32064932]
7. Niu J, Chen C. Public satisfaction with risk communication during COVID-19 outbreak: the impact of information needs, channel preferences, media trust, and emotions. Sci Found China 2020 Jun 16;34(06):794-803. [doi: 10.16262/j.cnki.1000-8217.2020.06.016]
8. Zeng B, Gerritsen R. What do we know about social media in tourism? A review. Tour Manag Rev 2014 Apr;10:27-36. [doi: 10.1016/j.tmp.2014.01.001]
9. López-Guimerà G, Levine MP, Sánchez-carracedo D, Fauquet J. Influence of mass media on body image and eating disordered attitudes and behaviors in females: a review of effects and processes. Media Psychol 2010 Nov 30;13(4):387-416. [doi: 10.1080/15213269.2010.525737]
10. Wei L, Huang Q. Retrospecting digital media use, negative emotions, and trust gaps during the COVID-19 pandemic in China: cross-sectional web-based survey. J Med Internet Res 2024 Jul 10;26:e49422 [FREE Full text] [doi: 10.2196/49422] [Medline: 38986127]
11. Chadwick VL, Saich F, Freeman J, Martiniuk A. Media discourse regarding COVID-19 vaccinations for children aged 5 to 11 years in Australia, Canada, the United Kingdom, and the United States: comparative analysis using the narrative policy framework. JMIR Form Res 2024 Apr 29;8:e38761 [FREE Full text] [doi: 10.2196/38761] [Medline: 36383344]
12. Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R, et al. The social amplification of risk: a conceptual framework. Risk Anal 2006 May 29;8(2):177-187. [doi: 10.1111/j.1539-6924.1988.tb01168.x]
13. Pattison AB, Reinfelde M, Chang H, Chowdhury M, Cohen E, Malahy S, et al. Finding the facts in an infodemic: framing effective COVID-19 messages to connect people to authoritative content. BMJ Glob Health 2022 Feb 07;7(2):e007582 [FREE Full text] [doi: 10.1136/bmjgh-2021-007582] [Medline: 35131808]
14. Li X, Fu Q, Wang T. Media exposure, news cognition and media trust in the COVID-19 epidemic: based on focus group interviews with Chinese and foreign college students. Newsreporter 2018;2021(03):76-86. [doi: 10.16057/j.cnki.31-1171/g2.2021.03.008]
15. Su J, Zhang J. From crisis communication to risk communication: the communication transformation of public health emergencies J. News Writ 2020;2020(04):5-13.
16. Ren W, Zhu X, Hu Y. Risk perception and preventive behaviors: a comparison of the multifaceted effects of social media and authoritative media during the outbreak of COVID-19. Chin J Journ Commun 2018;43(5):23-42. [doi: 10.13495/j.cnki.cjjc.2021.05.002]
17. Saechang O, Yu JX, Li Y. Public trust and policy compliance during the COVID-19 pandemic: the role of professional trust. Healthcare (Basel) 2021 Feb 02;9(2):151 [FREE Full text] [doi: 10.3390/healthcare9020151] [Medline: 33540762]
18. Bridgman A, Merkley E, Loewen PJ, Owen T, Ruths D, Teichmann L, et al. The causes and consequences of COVID-19 misperceptions: understanding the role of news and social media. Harv Kennedy Sch Misinformation Rev 2020 Jun 18:1-18 [FREE Full text] [doi: 10.37016/mr-2020-028]
19. Brekke KA, Kipperberg G, Nyborg K. Social interaction in responsibility ascription: the case of household recycling. Land Econ 2010 Sep 28;86(4):766-784. [doi: 10.3368/le.86.4.766]
20. Burby RJ, Paterson RG. Improving compliance with state environmental regulations. J Policy Anal Manage 1993;12(4):753 [FREE Full text] [doi: 10.2307/3325349]
21. Hassan MS, Al Halbusi H, Razali A, Ariffin RN, Williams KA. The Swedish gamble: trust in the government and self-efficacy in the battle to combat COVID-19. Curr Psychol 2022 Mar 19:1-16 [FREE Full text] [doi: 10.1007/s12144-022-02947-w] [Medline: 35340688]
22. Li D, Jiang W, Cai J. Individual norms and public action: how does institutional trust promote public responsive participation?: text data analysis based on urban water environment claims. J Public Adm 2018;19(02):117-129. [doi: 10.16149/j.cnki.23-1523.20220307.001]
23. Wu C, Shi Z, Wilkes R, Wu J, Gong Z, He N, et al. Chinese citizen satisfaction with government performance during COVID-19. J Contemp Chin 2021 Mar 17;30(132):930-944. [doi: 10.1080/10670564.2021.1893558]
24. Bargain O, Aminjonov U. Trust and compliance to public health policies in times of COVID-19. J Public Econ 2020 Dec;192:104316 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104316] [Medline: 33162621]
25. Goldfinch S, Taplin R, Gauld R. Trust in government increased during the COVID‐19 pandemic in Australia and New Zealand. Aust J Public Adm 2021 Jan 06;80(1):3-11. [doi: 10.1111/1467-8500.12459]
26. Fang M, Zhang H. Can crisis intervention repair government trust? The moderation effects of risk communication and community support. J Public Adm 2021;14(6):4-23 [FREE Full text]
27. Guglielmi S, Dotti Sani GM, Molteni F, Biolcati F, Chiesi AM, Ladini R, et al. Public acceptability of containment measures during the COVID-19 pandemic in Italy: how institutional confidence and specific political support matter. Int J Sociol Soc Policy 2020 Oct 21;40(9/10):1069-1085 [FREE Full text] [doi: 10.1108/ijssp-07-2020-0342]
28. Siegrist M, Earle TC, Gutscher H. Test of a trust and confidence model in the applied context of electromagnetic field (EMF) risks. Risk Anal 2003 Aug 11;23(4):705-716. [doi: 10.1111/1539-6924.00349] [Medline: 12926564]

XSL•FO
RenderX

29. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
30. Zettler I, Schild C, Lilleholt L, Kroencke L, Utesch T, Moshagen M, et al. The role of personality in COVID-19-related perceptions, evaluations, and behaviors: findings across five samples, nine traits, and 17 criteria. Soc Psychol Personal Sci 2021 Apr 05;13(1):299-310. [doi: 10.1177/19485506211001680]
31. Wong CM, Jensen O. The paradox of trust: perceived risk and public compliance during the COVID-19 pandemic in Singapore. J Risk Res 2020 Apr 28;23(7-8):1021-1030. [doi: 10.1080/13669877.2020.1756386]
32. Karakulak A, Tepe B, Dimitrova R, Abdelrahman M, Akaliyski P, Alaseel R, et al. Trust in government moderates the association between fear of COVID-19 as well as empathic concern and preventive behaviour. Commun Psychol 2023 Dec 15;1(1):43 [FREE Full text] [doi: 10.1038/s44271-023-00046-5] [Medline: 39242865]
33. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979 Mar;47(2):263. [doi: 10.2307/1914185]
34. Egüez A. Compliance with the EU waste hierarchy: a matter of stringency, enforcement, and time. J Environ Manage 2021 Feb 15;280:111672 [FREE Full text] [doi: 10.1016/j.jenvman.2020.111672] [Medline: 33309110]
35. van der Weerd W, Timmermans DR, Beaujean DJ, Oudhoff J, van Steenbergen JE. Monitoring the level of government trust, risk perception and intention of the general public to adopt protective measures during the influenza A (H1N1) pandemic in The Netherlands. BMC Public Health 2011 Jul 19;11(1):575 [FREE Full text] [doi: 10.1186/1471-2458-11-575] [Medline: 21771296]
36. Xie X, Wang H, Ren J, Yukiyuan M. Analysis of audience-centered risk communication in the SARS crisis. Applied Psychol 2005:104-109 [FREE Full text]
37. Li YL, Wang WZ, Wang J. Government intervention and optimization strategy for major epidemic prevention and control —— is based on game theory perspective and China's 2013 H7N9 prevention and control practice. J Hunan Agric Univ 2018(02):19-66 [FREE Full text]
38. Leppin A, Aro AR. Risk perceptions related to SARS and avian influenza: theoretical foundations of current empirical research. Int J Behav Med 2009 Feb 12;16(1):7-29 [FREE Full text] [doi: 10.1007/s12529-008-9002-8] [Medline: 19214752]
39. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984 Jan 01;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
40. Nielsen R, Fletcher R, Newman N, Brennen JS, Howard PN. Navigating the 'infodemic': how people in six countries access and rate news and information about coronavirus. Reuters Institute for the Study of Journalism. 2020. URL: https://reutersinstitute.politics.ox.ac.uk/infodemic-how-people-six-countries-access-and-rate-news-and-information-about-coronavirus [accessed 2024-04-29]
41. Zhang M, Xingling Y, Tianzi Y. The impact of social media use on public vaccination behavior: a study on the mediating effect of social norms. Journal Univer 2018(01):104-118 [FREE Full text] [doi: 10.20050/j.cnki.xwdx.2024.01.009]
42. Qi Y, Chen L, Zhang L, Yang Y, Jensiyi, Fu Y. Public practice,attitude and knowledge of coronavirus disease. J Trop Med 2020;20(02):145-149 [FREE Full text]
43. Te Poel F, Linn AJ, Baumgartner SE, Van Dijk L, Smit ES. Sick for information?: information needs and media use of the Dutch public during the COVID-19 pandemic. Eur J Health Commun 2021 Sep 09;2(3):24-43 [FREE Full text] [doi: 10.47368/ejhc.2021.302]
44. Gupta L, Gasparyan AY, Misra DP, Agarwal V, Zimba O, Yessirkepov M. Information and misinformation on COVID-19: a cross-sectional survey study. J Korean Med Sci 2020 Jul 13;35(27):e256 [FREE Full text] [doi: 10.3346/jkms.2020.35.e256] [Medline: 32657090]
45. Hong Y, Zhang P. Political news and happiness: the difference between traditional media and new media use. Chin J Commun 2020 Feb 22;13(4):370-388 [FREE Full text] [doi: 10.1080/17544750.2020.1731562]
46. Meppelink CS, Bos L, Boukes M, Möller J. A health crisis in the age of misinformation: how social media and mass media influenced misperceptions about COVID-19 and compliance behavior. J Health Commun 2022 Oct 03;27(10):764-775. [doi: 10.1080/10810730.2022.2153288] [Medline: 36576116]
47. Ma D, Lu Y. Information access, authoritarian personality, ideology and network nationalism: an analysis of the formation mechanism of Chinese netizens' political attitudes. J Tsinghua Univ 2019;34(03):180-192. [doi: 10.13613/j.cnki.qhdz]
48. Barker DC, Lawrence AB. Media favoritism and presidential nominations: reviving the direct effects model. Political Commun 2006 Jan;23(1):41-59. [doi: 10.1080/10584600500477013]
49. Ladd JM, Lenz GS. Exploiting a rare communication shift to document the persuasive power of the news media. Am J Pol Sci 2009;53(2):394-410 [FREE Full text] [doi: 10.2139/ssrn.1189811]
50. Peter A, Chen M, Carrasco S. Power interplay and newspaper digitization: lessons from the Pengpai experiment. Glob Media China 2017 Jan 17;1(4):497-510. [doi: 10.1177/2059436416687313]
51. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
52. Zhao D. 2021 media credibility survey: traditional media is still "information dinner". Kang 2018;2021(22):58.

XSL•FO
RenderX

53. Sharma DC, Pathak A, Chaurasia RN, Joshi D, Singh RK, Mishra VN. Fighting infodemic: need for robust health journalism in India. Diabetes Metab Syndr 2020 Sep;14(5):1445-1447 [FREE Full text] [doi: 10.1016/j.dsx.2020.07.039] [Medline: 32755849]
54. Wu Y, Song Y, Wang F, Zhou Y, Li M. Online public opinion guidance strategy for college students in the era of we media. J Comput Commun 2019;07(12):79-87. [doi: 10.4236/jcc.2019.712008]
55. Liang J, Yang M. On spreading and controlling of online rumors in we-media era. Asian Cult Hist 2015;7(2):42-46 [FREE Full text] [doi: 10.5539/ach.v7n2p42]
56. Qian B, Huang M, Xu M, Hong Y. Internet use and quality of life: the multiple mediating effects of risk perception and internet addiction. Int J Environ Res Public Health 2022 Feb 04;19(3):1795 [FREE Full text] [doi: 10.3390/ijerph19031795] [Medline: 35162817]
57. Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, et al. COVID-19 vaccine discourse on Twitter: a content analysis of persuasion techniques, sentiment and mis/disinformation. J Health Commun 2021 Jul 03;26(7):443-459. [doi: 10.1080/10810730.2021.1955050] [Medline: 34346288]
58. Wang F, Wei J, Shi X. Compliance with recommended protective actions during an H7N9 emergency: a risk perception perspective. Disasters 2018 Apr 11;42(2):207-232. [doi: 10.1111/disa.12240] [Medline: 28799670]
59. Lu C, Quan X. The impact of media use on government trust: an empirical study based on CGSS2010 data. Int Press 2015;37(05):66-80 [FREE Full text] [doi: 10.13495/j.cnki.cjjc.2015.05.013]
60. Hu R, Ren C. The impact of Internet users' media contact on government trust: a study based on mediating and moderating effects. J Hebei 2022;42(02):177-187.
61. Yang J, Wang C. Youth netizens' media preference and political trust. Youth Res 2021;2021(04):1-10 [FREE Full text]
62. Huang Y, Zhang H, Peng Z, Fang M. To what extent do young chinese elites comply with COVID-19 prevention and control measures? BMC Public Health 2023 Apr 24;23(1):751 [FREE Full text] [doi: 10.1186/s12889-023-15643-6] [Medline: 37095465]
63. Bo G. On improving government credibility and execution. J Wuhan Univ Sci Technol 2010:20-23.
64. Newton K. Trust, social capital, civil society, and democracy. Int Political Sci Rev 2001 Apr 01;22(2):201-214. [doi: 10.1177/0192512101222004]
65. Grimmelikhuijsen SG, Meijer AJ. Effects of transparency on the perceived trustworthiness of a government organization: evidence from an online experiment. J Public Adm Res Theory 2012 Nov 05;24(1):137-157. [doi: 10.1093/jopart/mus048]
66. Ma D, Sun M. The determinant of government credibility in the new media age —— transparency, responsiveness, or public relations skills? J Public Adm 2014;11(01):104-113 [FREE Full text]
67. Huang J, Wang X, Zhang S, Zhou N. The impact of entrepreneurial violations and violations on brand image. Manag World 2010(05):96-107. [doi: 10.19744/j.cnki.11-1235/f.2010.05.011]
68. Xu B. Research on the formation mechanism of government trust crisis after negative events: a quasi-experimental study. Xuaihai 2019(02):5 [FREE Full text]
69. Zhu Z, Liu Y, Kapucu N, Peng Z. Online media and trust in government during crisis: the moderating role of sense of security. Int J Disaster Risk Reduct 2020 Nov;50:101717. [doi: 10.1016/j.ijdrr.2020.101717]
70. Stevens JP, Stevens JP. Applied Multivariate Statistics for the Social Sciences. 4th edition. New York, NY: Psychology Press; 2002.
71. Steenkamp JB, Maydeu-Olivares A. An updated paradigm for evaluating measurement invariance incorporating common method variance and its assessment. J Acad Mark Sci 2020 Oct 30;49(1):5-29. [doi: 10.1007/s11747-020-00745-z]
72. Vu VT. Public trust in government and compliance with policy during COVID-19 pandemic: empirical evidence from Vietnam. Public Organiz Rev 2021 Oct 06;21(4):779-796. [doi: 10.1007/s11115-021-00566-w]
73. Chong M, Choy M. The social amplification of haze-related risks on the internet. Health Commun 2018 Jan 30;33(1):14-21. [doi: 10.1080/10410236.2016.1242031] [Medline: 27901352]

Abbreviations

CMB: common method bias
EFA: exploratory factor analysis
SARF: social amplification of risk framework
SEM: structural equation model

XSL•FO
RenderX

Edited by Y Hong; submitted 31.07.24; peer-reviewed by S Zhang, M Salah, A Schreijer, A Tolah, D Wang; comments to author 14.10.24; revised version received 25.10.24; accepted 25.11.24; published 20.01.25.

Please cite as:
Zhang H, Yang C, Deng X, Luo C
How Authoritative Media and Personal Social Media Influence Policy Compliance Through Trust in Government and Risk Perception: Quantitative Cross-Sectional Survey Study
J Med Internet Res 2025;27:e64940
URL: https://www.jmir.org/2025/1/e64940
doi:10.2196/64940
PMID:

©Hua Zhang, Cheng Yang, Xiuxian Deng, Chunyan Luo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Health Messaging on Twitter During the COVID-19 Pandemic: Observational Study

Ashwin Rao[1*], BEng; Nazanin Sabri[2*], MSc; Siyi Guo[1], BS; Louiqa Raschid[3], PhD; Kristina Lerman[1], PhD

[1]Information Sciences Institute, University of Southern California, Marina del Rey, CA, United States

[2]Department of Computer Science and Engineering, University of California - San Diego, La Jolla, CA, United States

[3]Institute of Advanced Computer Studies, University of Maryland, College Park, MD, United States

[*]these authors contributed equally

Corresponding Author:
Ashwin Rao, BEng
Information Sciences Institute
University of Southern California
4676 Admiralty Way STE 1001
Marina del Rey, CA, 90292
United States
Phone: 1 2135050363
Email: mohanrao@usc.edu

Abstract

Background: Effective communication is crucial during health crises, and social media has become a prominent platform for public health experts (PHEs) to share information and engage with the public. At the same time, social media also provides a platform for pseudoexperts who may spread contrarian views. Despite the importance of social media, key elements of communication, such as the use of moral or emotional language and messaging strategy, particularly during the emergency phase of the COVID-19 pandemic, have not been explored.

Objective: This study aimed to analyze how PHEs and pseudoexperts communicated with the public during the emergency phase of the COVID-19 pandemic. We focused on the emotional and moral language used in their messages on various COVID-19 pandemic–related topics. We also analyzed their interactions with political elites and the public's engagement with PHEs to gain a deeper understanding of their influence on public discourse.

Methods: For this observational study, we gathered a dataset of >539,000 original posts or reposts from 489 PHEs and 356 pseudoexperts on Twitter (subsequently rebranded X) from January 2020 to January 2021, along with the replies to the original posts from the PHEs. We identified the key issues that PHEs and pseudoexperts prioritized. We also determined the emotional and moral language in both the original posts and the replies. This allows us to characterize priorities for PHEs and pseudoexperts as well as differences in messaging strategy between these 2 groups. We also evaluated the influence of PHEs' language and strategy on the public response.

Results: Our analyses revealed that PHEs focused more on masking, health care, education, and vaccines, whereas pseudoexperts discussed therapeutics and lockdowns more frequently (P<.001). PHEs typically used positive emotional language across all issues (P<.001), expressing optimism and joy. Pseudoexperts often used negative emotions of pessimism and disgust, while limiting positive emotional language to origins and therapeutics (P<.001). Along the dimensions of moral language, PHEs and pseudoexperts differed on care versus harm and authority versus subversion across different issues. Negative emotional and moral language tends to boost engagement in COVID-19 discussions across all issues. However, the use of positive language by PHEs increases the use of positive language in the public responses. PHEs act as liberal partisans: they express more positive affect in their posts directed at liberals and more negative affect in their posts directed at conservative elites. In contrast, pseudoexperts act as conservative partisans. These results provide nuanced insights into the elements that have polarized the COVID-19 discourse.

Conclusions: Understanding the nature of the public response to PHEs' messages on social media is essential for refining communication strategies during health crises. Our findings underscore the importance of using moral-emotional language strategically to reduce polarization and build trust.

(J Med Internet Res 2025;27:e63910) doi:10.2196/63910

XSL•FO
RenderX

KEYWORDS

public health; public health messaging; COVID-19; Twitter; emotions; moral foundations; polarization

Introduction

Background

The emergency phase of the COVID-19 pandemic created a worldwide public health crisis, disrupting daily lives and overwhelming health care facilities. During this time, the need for communicating reliable medical information and public health guidance became very important. Social media platforms such as Twitter (subsequently rebranded X; X Corp) provided a space for public health experts (PHEs) from government, academia, and think tanks to communicate timely and reliable information about the COVID-19 pandemic to the public [1,2].

Previous literature [3,4] shows that the public often follows cues from in-group elites and opposes cues from out-group elites. Effective messaging strategies can be crucial in times of public health crises. Individuals with higher COVID-19 knowledge practiced more protective behaviors [5,6]. Messaging that highlights risks to younger adults, in addition to risks to older adults, was found to bring about a higher threat perception about COVID-19 [7]. In contrast, messaging that appeals to the audience's morals or fears to encourage compliance was found to be polarizing, divisive, and detrimental to social cohesion [8]. In one Maryland county, Latinx vaccination rates significantly increased following the introduction of a cartoon grandmother in outreach efforts [9]. Messaging that focused on vaccine safety and efficacy, along with endorsements of vaccination from political leaders, was found to be highly effective [10]. Pink et al [11] found that Republicans who were exposed to endorsements from Republican elites witnessed higher vaccination intentions than those who viewed the Democratic elite endorsement, with out-group elite exposures proving counterproductive.

As the emergency phase of the COVID-19 pandemic progressed, discussions around the pandemic grew highly contentious and ideologically polarized [12,13]. With public's trust in institutions and experts eroding, health-related misinformation proliferated about all aspects of the pandemic, from its origins to alternative treatments and the efficacy of nonpharmaceutical interventions, and eventually the vaccine [14]. At the heart of this proliferation were influential "pseudoexperts," such as the "Disinformation Dozen" [15], who amplified contrarian perspectives and challenged the recommendations of PHEs.

This polarization [5,16-18] of the COVID-19 pandemic laid bare a fractured public health messaging apparatus [19-22]. The emergence of contradictory theories and 2 polarized groups of influential elites and experts [23,24] and conspiracy theories about the origins of the COVID-19 pandemic, its severity, and the efficacy of prophylactic measures started to take hold [25,26]. Initial theories revolved around the severity of the virus with several calling it a "hoax" and "plandemic" [27]. A study by the Pew Research Center [28] found that approximately 25% of the survey responders believed that COVID-19 was probably created intentionally by powerful people. Another study found that approximately 3 in 10 Americans believe that COVID-19 was artificially created in a laboratory [29]. Theories about virus transmission being connected to 5G, bats, pangolins, and wet markets were widely propagated by conspiracy theorists on social media [30,31]. As the COVID-19 pandemic progressed, we also witnessed the propagation of pseudoscientific cures for COVID-19 [32,33]. With increased COVID-19 pandemic–related engagement from the general public, these conspiracy theories soon started to proliferate on social media platforms [34-39]. Findings from the study by Antonakis [40] highlight the role of influential accounts in mitigation efforts. As influential elites, often holding advanced medical degrees, began contradicting other PHEs on various aspects of the COVID-19 pandemic, public consensus was disrupted, occasionally leading to serious consequences [41-43].

Expression of fear and anger were found to indicate support for restrictive COVID-19 mitigation policies such as lockdowns to limit the spread of COVID-19 [44], while anxiety predicted support for economic policies. Anger was found to indicate support for aggressive responses to transgressors [45,46]. Hatemi et al [47] found fear to be a strong underlying factor in anti-immigration and prosegregation stances. Previous studies [48] relied on surveys to show an increase in distress and uncertainty during the emergency phase of the COVID-19 pandemic. Agrawal et al [49] investigated sentiments of posts about the COVID-19 vaccine, post–COVID-19 health factors, and health service providers. Among the 3 topics, health care providers had the largest positive sentiment, resulting in an inference that posters were happy with their care and appreciated the work of health care providers. Lwin et al [50] found that public emotions in Twitter shifted from fear to anger early in the COVID-19 pandemic. Wheaton et al [51] revealed that greater susceptibility to emotion contagion was associated with greater concern about the spread of COVID-19. The moral foundations of care and fairness were found to correlate with compliance of COVID-19 health recommendations, including masking, staying at home, and social distancing [52,53]. Moral attitudes were also able to predict county-level vaccination rate [54] and vaccine hesitancy [55]. Pacheco et al [56] found that care or harm was associated with provaccine sentiment, whereas liberty or oppression was correlated with antivaccine attitudes. While vaccinations are a critical polarizing issue in the discussion of COVID-19, no study yet has explored differences in moral appeals across a broader range of contentious COVID-19 issues.

Social sharing of opinions and emotions is ubiquitous, and social media has greatly expanded its scope [57,58]. Bazarova et al [58] investigated how responses to what a user shared affected their feeling of satisfaction. Analyzing Facebook status updates, Burke and Develin [59] found that posts with positive emotions received more likes. The comments associated with these posts were also more positive [59]. Positive emotion words were also shown to have a positive correlation with the number of reposts [60]. Sousa et al [61] reported that while social connections dominate reply behavior, for authors with large ego networks, there is a separation between who replies based on the topic of

XSL•FO
RenderX

the post. Early thematic analysis of public replies to the COVID-19 pandemic found themes of prevention, symptoms, views on politicians, and humor [62]. Replies by antivaccine users were found to be more toxic than users with other beliefs about vaccines [63]. Gallagher et al [64] found groups to preferentially amplify elites that are demographically similar to them.

In psychology, "affect" is the experience of feeling or emotion, and it significantly shapes individual's attitudes, beliefs, and behaviors. In online interactions, affect influences how a message is crafted and how it resonates with audiences, ultimately affecting the message's spread and impact. Research shows that people respond to the emotions expressed in online messages [65], although due to an asymmetry in human cognition [66], posts expressing negative emotions receive more engagement than positive posts [67,68]. It has also been shown that emotionally charged messages, particularly ones tapping into moral sentiments such as outrage, spread farther on the web [69,70]. Affect provides reliable indicators for gauging public response to major events and policy decisions [44,45,71-73] and interacts with ideology to fuel polarization. Political scientists have identified affective polarization—a phenomenon where individuals like and trust members of their own party while disliking and distrusting members of opposing parties—as a significant threat to effective governance [74,75]. Public's reactions to the COVID-19 pandemic, as measured via attitudes and sentiments expressed in online messages, were multifaceted [76] and grew polarized early in the COVID-19 pandemic [12]. Moreover, there was an ideological asymmetry wherein conservatives shared more low-quality health information than liberals [17] and were also exposed to more misinformation [14]. In addition, conservatives expressed more negative moral sentiments in online posts about the COVID-19 pandemic than liberals [13]. However, to the best of our knowledge, few studies have focused on online influencers and experts who shaped public health policy and disseminated health-related information to the public. As a result, we know little about the messaging strategies they used, the role that affect played in these messages, and how the public responded to the messages.

Objectives

To address these knowledge gaps, we examined messages posted by PHEs and pseudoexperts on Twitter during the emergency phase of the COVID-19 pandemic. We identified a set of 489 PHEs and 356 pseudoexperts and collected >372,000 original posts that they posted between January 21, 2020, and January 20, 2021. Collectively, these accounts had a vast reach; each PHE had on average 94,000 followers (estimated reach approximately 45M), and pseudoexperts had on average 78,000 followers (estimated reach approximately 30M). In addition, we also collected replies to >195,000 original posts posted by PHEs during this period. Our objectives were two-fold: (1) identify *what* public health influencers talk about online and *how* they talk and (2) identify factors that impacted public engagement with the PHEs.

We leverage methods introduced in the study by Rao et al [13] to identify posts about 7 important COVID-19 pandemic–related issues: origins of the virus, lockdowns and stay-at-home orders, masking mandates, online schooling and education, health care, alternative treatments and therapeutics, and vaccines. We use state-of-the-art classifiers [77,78] to analyze the emotional and moral language used in posts. We then use regression to compare how affect shapes the health-related messages on different issues posted by PHEs and pseudoexperts. Finally, we collect all replies for a sample of PHE posts to study how the use of emotional and moral language impacts engagement by the public with these messages.

Our study uncovers the inherent complexities in public health communication during the COVID-19 pandemic by investigating the following hypotheses:

- PHEs and pseudoexperts differ significantly in the issues they emphasize during the COVID-19 pandemic.
- There are asymmetries in the emotional and moral language used by PHEs and pseudoexperts in discussing these issues.
- PHEs and pseudoexperts exhibit affective polarization, with pseudoexperts expressing more positivity toward conservative elites and PHEs favoring liberal elites.
- The emotions and moral language used by PHEs are reflected in the responses from common users.

Methods

Overview

We begin by describing our data collection procedure and present statistics and basic characteristics of the dataset. We believe that this description provides the reader with additional insights to better interpret the results. Finally, we describe our content analysis procedure and models used to produce results.

Study Design and Population

This is an observational study that analyzes the social media communication of 489 PHEs and 356 pseudoexperts on Twitter during the COVID-19 pandemic from January 2020 to January 2021. We focus on the emotional and moral language used in their original posts as well as the public engagement with those posts. The study compares the key issues prioritized by PHEs and pseudoexperts and examines how these groups engaged with political elites and their respective audiences. By analyzing their messaging strategies, the study aims to understand how their language influenced public engagement and discourse. The study population consists solely of users on the social media platform Twitter.

Data Collection

We use a publicly available dataset [79] consisting of 1.4B posts about COVID-19 posted between January 21, 2020, and January 1, 2021. These posts contained ≥1 COVID-19–related keywords, such as coronavirus, pandemic, and Wuhan, among others.

Identifying PHEs and Pseudoexperts

In collaboration with a health policy researcher, we identified accounts belonging to 30 PHEs and 30 pseudoexperts who were active on Twitter during the emergency phase of the COVID-19 pandemic (Textbox 1). PHEs include individuals with advanced degrees in medicine, epidemiology, genomics, infectious

XSL•FO
RenderX

diseases, public policy, and economics. These experts offered informed, evidence-based perspectives grounded in science, shaping public understanding and policy, regardless of whether their views aligned with the scientific consensus.

Textbox 1. Twitter handles of accounts associated with public health experts (PHEs) and pseudoexperts.

PHEs

EricTopol, PeterHotez, ashishkjha, trvrb, EpiEllie, JuliaRaifman, devisridhar, meganranney, luckytran, asosin, DrLeanaWen, dremilyportermd, DrJaimeFriedman, davidwdowdy, BhramarBioStat, geochurch, DrEricDing, michaelmina_lab, Bob_Wachter, JenniferNuzzo, mtosterholm, MonicaGandhi9, cmyeaton, nataliexdean, angie_rasmussen, ProfEmilyOster, mlipsitch, drlucymcbride, ScottGottliebMD, CDCDirector, and Surgeon_General

Pseudoexperts

mercola, LEEHIEB_MD, stella_immanuel, DrOz, DrThomasLevy, DrJudyAMikovits, va_shiva, Drericne- pute1, DrButtar, DrArtinMassihi, davidicke, mrmarksteel, drscottjensen, cameronks, RobertKennedyJr, TyCharleneB, BusyDrT, IslamRizza, unhealthytruth, sayerjigmi, kelly- broganmd, DrChrisNorthrup, DrBenTapper1, DrZachBush, SherrillSellman, AFLDSorg, DrSimoneGold, jennybethm, drcole12, JamesTodaroMD, Covid19Critical, and DrJohnWitcher

While most of the individuals identified as PHEs were noncontroversial, we recognize that some may have made questionable statements or provided guidance that diverged from mainstream views. We chose to include these voices as PHEs for several reasons: (1) to reflect the diversity of perspectives among experts and (2) to illustrate that PHEs often participated in nuanced debates shaped by the evolving understanding of COVID-19's risks and impacts. Nonetheless, our primary criterion was that these individuals predominantly based their views on data-driven research and evidence-based analysis, even when their positions significantly deviated from scientific consensus.

The group of pseudoexperts, by contrast, includes individuals with or without medical credentials who consistently promoted pseudoscientific theories, unproven alternative treatments, and unsupported conclusions about COVID-19. Many in this group actively questioned the need to exercise prophylactic measures such as masking and lockdowns, expressed skepticism about the safety of vaccines, referenced retracted studies, and offered unsupported claims, directly contrasting with the PHEs' reliance on evidence-based research. This group also includes the "Disinformation Dozen," a group of individuals and organizations identified by the Center for Countering Digital Hate as being responsible for promoting false claims about COVID-19 [15].

We expanded the initial seed set of PHEs and pseudoexperts (Textbox 1) using the repost network to identify additional influential figures shaping public opinion during the critical period of the COVID-19 pandemic. Our approach relies on repost interactions within a publicly available COVID-19 Twitter dataset, comprising >1 billion COVID-19–related posts collected between January 21, 2020, and January 20, 2021 [79]. Reposts allow users to repost content originally shared by others, and they have been shown to be proxies of endorsement of content [80,81]. Individuals often repost others who share similar beliefs and perspectives [14,82-84]. We used repost interactions involving the initial seed sets of PHEs and pseudoexperts to identify 2 distinct networks: one comprising accounts frequently reposted by PHEs and the other comprising accounts frequently reposted by pseudoexperts. We used Eigenvector centrality [85] to identify the most prominent accounts in each repost network, selecting the top 500 accounts reposted by either PHEs or pseudoexperts. Eigenvector centrality measures a node's influence in a network, where its centrality is based not only on the number of accounts reposting it but also on the influence of the accounts that are reposting those accounts. After filtering out organizational accounts, we were left with 489 individual PHEs and 356 individual pseudoexperts.

With this expanded set of individuals, we proceeded to extract their posts, resulting in a dataset comprising 340,000 posts from PHEs and 175,000 posts from pseudoexperts. This broader dataset allowed us to analyze the discourse and influence patterns across a more comprehensive group of health professionals and pseudoexperts active during the emergency phase of the COVID-19 pandemic.

Figure 1 shows the repost interactions network between PHEs and pseudoexperts (845 nodes and 107K edges). The color of the edge is dependent on the target node. Green edges represent interactions where a PHE was reposted, whereas orange edges represent interactions where a pseudoexpert was reposted. The size of the node is proportional to how many times the account was reposted: highly reposted experts have larger node sizes. The network shows 2 tightly knit communities, 1 for each group, with sparse between-community interactions. This structure is typical of online echo chambers and suggests that each community mainly listens to their own community.

XSL•FO
RenderX

Figure 1. Repost interactions. Nodes represent public health experts (PHEs; green) and pseudoexperts (orange) and repost interactions between them. Green edges represent interactions where a PHE was reposted, and orange edges represent interactions where a pseudoexpert was reposted. The size of the node is proportional to the number of times the expert was reposted.

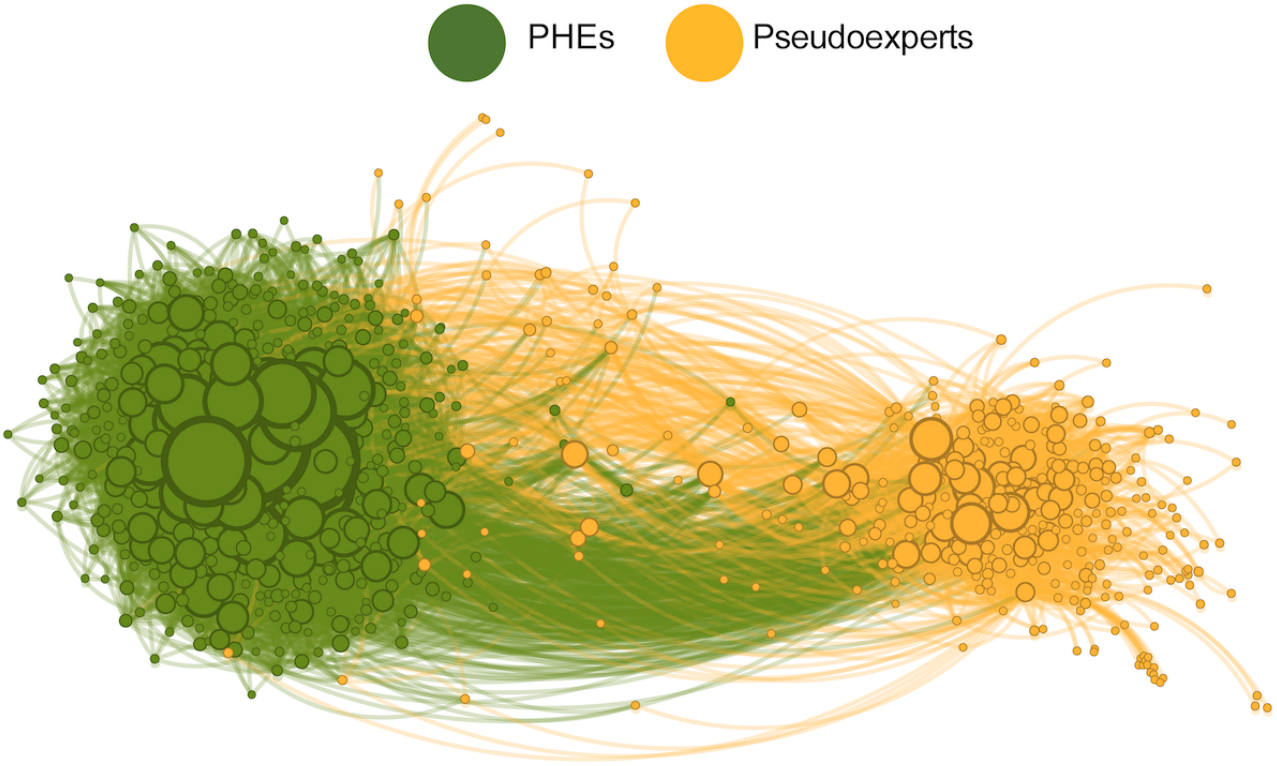

Textbox 2 shows the 25 most popular hashtags used by PHEs and pseudoexperts. There are notable similarities and differences. While "vaccine" is the most important topic for both groups, PHEs unsurprisingly mention "vaccineswork" and "vaccinate," in contrast to posts from pseudoexperts that mention "vaccineinjury" and "vaccinefreedom" and urge people to "learntherisk" of vaccines.

Textbox 2. Hashtag use. The top 25 hashtags used by public health experts (PHEs) and pseudoexperts on Twitter.

PHEs

sarscov, vaccine, pandemic, wearamask, thisisourshot, publichealth, vaccinessavelive, deltavariant, socialdistance, covidvaccine, mentalhealth, healthcare, vaccinate, scicomm, election, getvaccinate, stopthespread, maskup, vaccineswork, healthworker, medtwitter, stayathome, globalhealth, remdesivir

Pseudoexperts

medtwitter, informedconsent, thedefender, billgatesbioterrorist, wakeupamerica, autism, fauciliedpeopledied, vaccinefreedom, ivermectin, learntherisk, lockdown, hydroxychloroquine, factsnotfear, tipsfornewdocs, censorship, vaccineinjury, vaccinefailure, plandemic, freedom, screenbyvaccine, aluminium, bigpharma, homeopathy, doctorspeakup

Further analyzing the content shared by PHEs and pseudoexperts, we extract the URLs they shared in their posts and identify the pay-level domains (PLDs) these URLs point to. We compute the log-odds ratio to identify which group is more likely to share each PLD. Figure S1A in Multimedia Appendix 1 shows the top 15 PLDs for both groups. We found that PHEs were more likely to share URLs from highly reputable sources, such as *Journal of American Medical Association*, *Boston Review*, and the *New York Times*. In contrast, pseudoexperts share more questionable sources, such as the *Gateway Pundit*, Children's Health Defense, Patriot Project, and *Russia Today*, among others. PLDs often have ideological leanings, ranging from liberal (0) to least-biased (0.5) to conservative (1), where 0 represents the most liberal, 1 represents the most conservative, and 0.5 indicates a neutral or least-biased position. [86]. Figure S1B in Multimedia Appendix 1 compares the distribution of ideological leanings of information sources shared by PHEs and pseudoexperts. While PHEs tended to share more liberal sources, pseudoexperts shared more conservative sources.

Identifying Issue-Relevant Posts

We decompose the multifaceted discussion about the COVID-19 pandemic along various contentious issues: COVID-19 origins, lockdowns and business closures, mask mandates, school closures, therapeutics, health care, and vaccines. To do so, we rely on methods discussed in the studies by Rao et al [13] and Eisenstein et al [87] to extract issue-relevant keywords from Wikipedia articles. Once we identify keywords, we identify posts that explicitly mention any of these keywords as being issue relevant. This approach was validated in the study by Rao et al [13] as being able to accurately identify issue-relevant

XSL•FO
RenderX

content. Table S1 in Multimedia Appendix 1 illustrates sample posts from our dataset discussing each issue.

We define the origins issue to encompass discussions surrounding the possible causes for the origin of the COVID-19 pandemic, including topics such as pangolins, gain of function research, wet markets, and bats. The lockdown issue comprises content pertaining to early state and federal mitigation efforts, such as quarantines, stay-at-home orders, business closures, reopening, and calls for social distancing. Discussions related to masking are defined by considerations of face coverings, mask mandates, shortages, and antimask sentiment. Education-related content involves posts regarding school closures, the reopening of educational institutions, homeschooling, and online learning during the emergency phase of the COVID-19 pandemic. The health care issue deals with conversations on the state of the health care system, availability of personal protective equipment, ventilators, oxygen supplies, and intensive care units. Discourse around therapeutics encompasses varied alternative treatments proposed to fight COVID-19 infections, including hydroxychloroquine, ivermectin, plasma therapy, Chinese medicine, colloidal silver, and herbal remedies. The vaccines issue pertains to discussions about COVID-19 vaccines, vaccine mandates, antivaccine sentiment, and vaccine hesitancy in the United States.

Identifying Emotions and Morality

To identify emotions expressed in posts and replies, we used a state-of-the-art transformer-based multilabel emotion detection model described in another study [77]. This model was fine-tuned using the SemEval 2018 Task 1e-c dataset [88]. It surpasses previous methods in its ability to capture the correlations among various emotions. When presented with the text of a post, the model generates confidence scores for the presence of a wide spectrum of emotions. We later bin these confidence scores using a 0.5 threshold to binarize the output. The emotions it can identify include *anticipation*, *joy*, *love*, *optimism*, *anger*, *disgust*, *fear*, *sadness*, and *pessimism*. The definitions of these emotions are based on the study by Mohammad et al [88] and are as follows:

- anticipation (also includes interest and vigilance)
- joy (also includes serenity and ecstasy)
- love (also includes affection)
- optimism (also includes hopefulness and confidence)
- anger (also includes annoyance and rage)
- disgust (also includes disinterest, dislike, and loathing)
- fear (also includes apprehension, anxiety, and terror)
- sadness (also includes pensiveness and grief)
- pessimism (also includes cynicism and no confidence)

Prior research has shown that emotional and moral language in social media messages impacts how they are received by the audiences and engagement [65,70]. The moral foundations theory [89] provides a framework for understanding how moral values shape people's political attitudes and behaviors. The moral foundations theory proposes that individuals' values and judgments can be described by 5 moral virtue or vice pairs: care or harm, fairness or cheating, loyalty or betrayal, authority or subversion, and sanctity or degradation. More specifically, these include the following:

- Care or harm. This foundation revolves around the concept of empathy and compassion. People who prioritize this foundation value caring for others and preventing harm. They are sensitive to the needs of others and strive to promote their well-being.
- Fairness or cheating. This foundation is concerned with issues of justice, reciprocity, and fairness. Individuals who emphasize this foundation are attuned to issues of equality, fairness, and proportionality. They believe in treating others fairly and oppose exploitation and unfair advantage.
- Loyalty or betrayal. Teople who prioritize loyalty value group cohesion, allegiance, and solidarity. They are inclined to support and defend their in-groups, whether it be family, community, or nation, and perceive betrayal or disloyalty as morally reprehensible.
- Authority or subversion. This foundation centers on respect for authority, tradition, and hierarchy. Individuals who emphasize this foundation value social order, respect for authority figures, and obedience to legitimate institutions and norms. They believe that maintaining authority and order is essential for a stable society.
- Sanctity or degradation. This foundation involves the reverence for purity, sanctity, and sacredness. People who prioritize this foundation are concerned with issues related to cleanliness, moral purity, and spiritual transcendence. They may view certain actions, objects, or behaviors as inherently sacred or profane.

Our morality detection model is trained on the transformer-based pretrained language model by Devlin et al [90]. The training process involves 3 Twitter datasets, a manually annotated COVID-19 dataset [91], the Moral Foundation Twitter Corpus dataset covering 6 different topics [92], and a dataset of political posts from US congress members [93]. By incorporating an in-domain training set focused on COVID-19, along with other diverse datasets spanning various topics, we enhance the model's generalizability for application to target data as discussed in the study by Guo et al [78].

Ethical Considerations

This study involved secondary analysis of publicly available Twitter data and was reviewed and deemed exempt by the University of Southern California's Institutional Review Board. The exemption was granted because the data are publicly accessible and do not involve interaction with human participants or the use of identifiable private information. Informed consent was not required as the data were collected from a public platform in accordance with Twitter's terms of service, and there was no reasonable expectation of privacy in the original context. All data were anonymized during analysis, with no identifying features included in the study outputs, ensuring privacy and confidentiality. No compensation was provided as the study exclusively analyzed publicly available data.

Results

Messaging About COVID-19 Issues

More than half of the posts from PHEs and pseudoexperts mention at least 1 of the 7 COVID-19 issues we identified.

XSL•FO
RenderX

Figure 2 compares the average daily share of posts from both groups on each issue. Overall, we found that pseudoexperts tend to be more vocal on the issues of lockdowns, therapeutics, and vaccines, while PHEs generate more content about masking, health care, and education. We did not witness any significant differences in the discourse about origins of the virus. These trends reflect the attention to issues by each group before President Biden's inauguration, which is the period covered by this study. To better summarize the varied perspectives expressed by PHEs and pseudoexperts on the 7 issues of interest, we randomly sample 25 posts for the 2 groups across these issues and prompt OpenAI's ChatGPT to provide the broad perspective being expressed using the following prompt:

Figure 2. Comparing the activity of public health experts (PHEs) and pseudoexperts. Box plots compare the daily proportion of issue-related posts from PHEs and pseudoexperts. Mann-Whitney U Test with Bonferroni correction was used to assess significance. **P*<.05, ***P*<.01, ****P*<.001, and *****P*<.0001; ns: not significant.

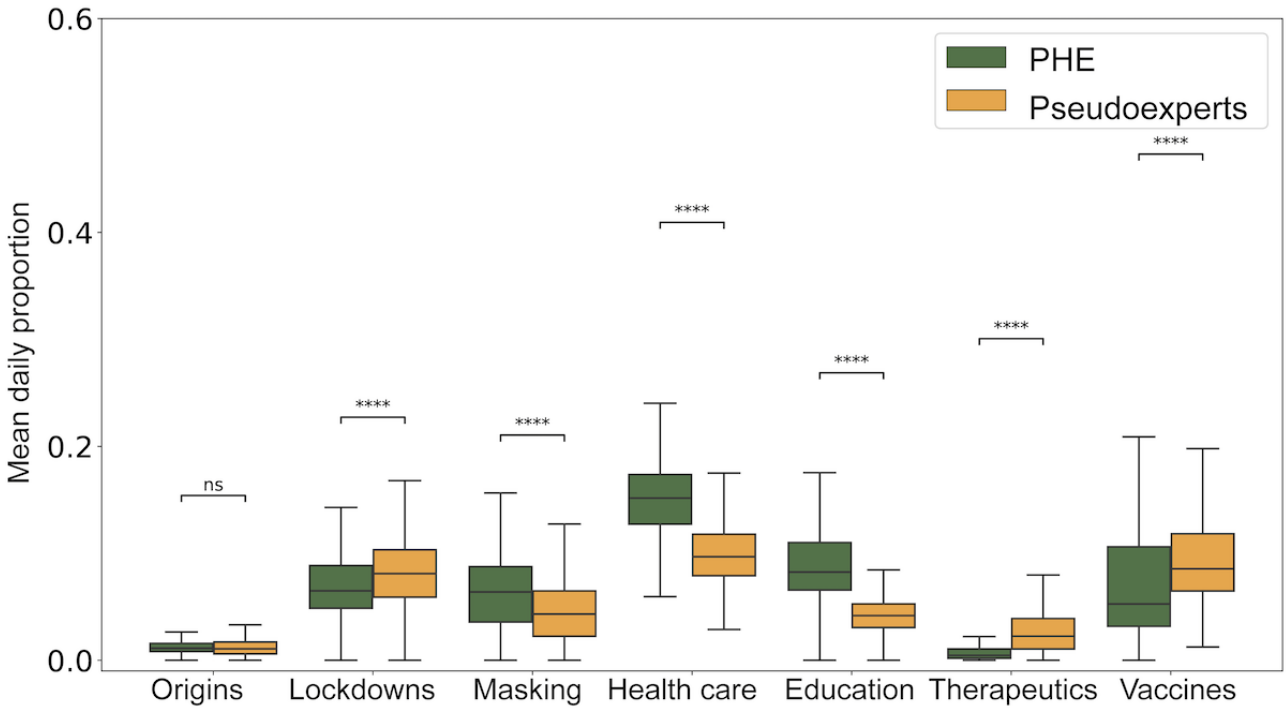

"Summarize perspectives being expressed about <Issue> in these posts: <T>," where <Issue> is one of (Origins, Lockdowns, Masking, Education, Health care, Therapeutics, Vaccines) and <T> represents a concatenation of the 25 posts that were randomly sampled for each issue and group pair.

The results presented in Table S2 in Multimedia Appendix 1 demonstrate the contrasting viewpoints between the 2 groups on various issues. Regarding the origins of the virus, PHEs generally lean toward the belief that it originated in a laboratory, albeit with some skepticism, while pseudoexperts heavily criticize China and its potential involvement in gain of function research. PHEs emphasize the importance of ongoing vigilance, adherence to stay-at-home orders, and widespread use of masks, whereas pseudoexperts question the effectiveness of lockdowns and mask mandates and criticize government intervention in these areas. On the topic of therapeutics, PHEs urge caution against self-prescribing drugs such as hydroxychloroquine, azithromycin, and ivermectin without evidence of their efficacy in treating COVID-19, whereas pseudoexperts advocate for the use of these medications.

Next, we look at the temporal patterns of issue-related discussions. Figures 3A and 3B show the daily share of posts from each group about the issues. Major events are marked with vertical lines—lockdowns: March 15, 2020 (purple dashed line), when stay-at-home orders were issued across the mainland United States; health care: March 30, 2020 (orange dashed line); therapeutics: April 24, 2020 (yellow dashed line), when President Trump proposed using bleach to fight off the virus; education: July 8, 2020 (red dashed line), when Trump called for schools to reopen; and vaccines: November 9, 2020 (blue dashed line), when Pfizer reported 93% efficacy in phase 3 trials. When stay-at-home orders were issued in mid-March 2020, we see a rise in lockdown-related discussions from PHEs. Lockdown-related discourse from pseudoexperts gained momentum in mid-April amid calls to reopen the economy and intensified in early June 2020 during the Black Lives Matter protests, when they criticized the large-scale demonstrations. As COVID-19 cases surged in late March 2020, we see a spike in health care–related discourse from PHEs, with growing calls for emergency preparedness in terms of improving access to personal protective equipment and ventilators. We do not observe a corresponding increase from pseudoexperts.

XSL•FO
RenderX

Figure 3. Daily fraction of original posts by (A) public health experts (PHEs) and (B) pseudoexperts related to each issue. Major events are marked with vertical lines. Major events are marked with vertical dashed lines: Lockdowns (March 15, 2020, purple), health care (March 30, 2020, orange), therapeutics (April 24, 2020, yellow, Trump's bleach proposal), education (July 8, 2020, red, Trump's school reopening call), and vaccines (November 9, 2020, blue, Pfizer's 93% efficacy report).

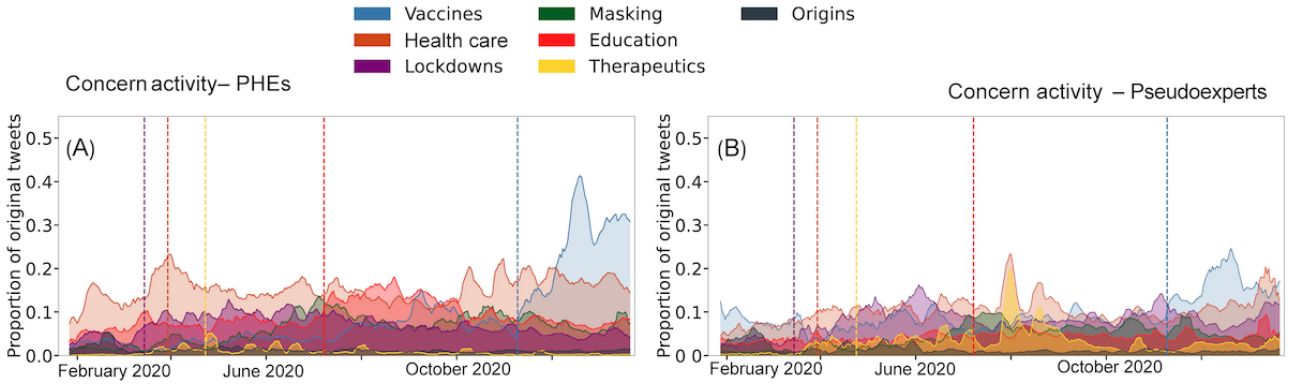

We see a small spike in therapeutics-related discussions among PHEs following President Trump's April 24, 2020, comment on using bleach to ward off the COVID-19 virus. Almost immediately following the Federal Drug Administration's issuance of an emergency use authorization on various therapeutics such as hydroxychloroquine on March 28, 2020, we see an immediate increase in therapeutics-related discussions from pseudoexperts. However, we see highest share of posts from them on July 26, 2020, when then White House chief of staff Mark Meadows announced that alternative therapeutics would be coming soon. We also see spikes in education-related discussions from PHEs and pseudoexperts following President Trump's July 8, 2020, call to reopen educational institutions. However, the spikes were for very different reasons; PHEs expressed increased skepticism toward reopening schools, while pseudoexperts supported the reopening call. The largest spikes for both groups are for vaccine-related discussions following Pfizer's announcement of successful COVID-19 phase 3 vaccine trials (10%-37% for PHEs and 20%-32% for pseudoexperts).

Emotional and Moral Language

Figure 4 compares the distribution of the daily fraction of posts posted by PHEs and pseudoexperts expressing a certain emotion. Overall, PHEs express more positive emotions such as anticipation, joy, and optimism and more low arousal negative emotions such as sadness and fear, whereas pseudoexperts express more high arousal negative emotions such as anger and disgust. Interestingly, we do not see much love or pessimism in our data.

Figure 4. Box plots compare daily proportion of posts from public health experts (PHEs) and pseudoexperts expressing various emotions. Mann-Whitney U Test with Bonferroni correction is used to assess significance. *P<.05, **P<.01, ***P<.001, and ****P<.0001; ns: not significant.

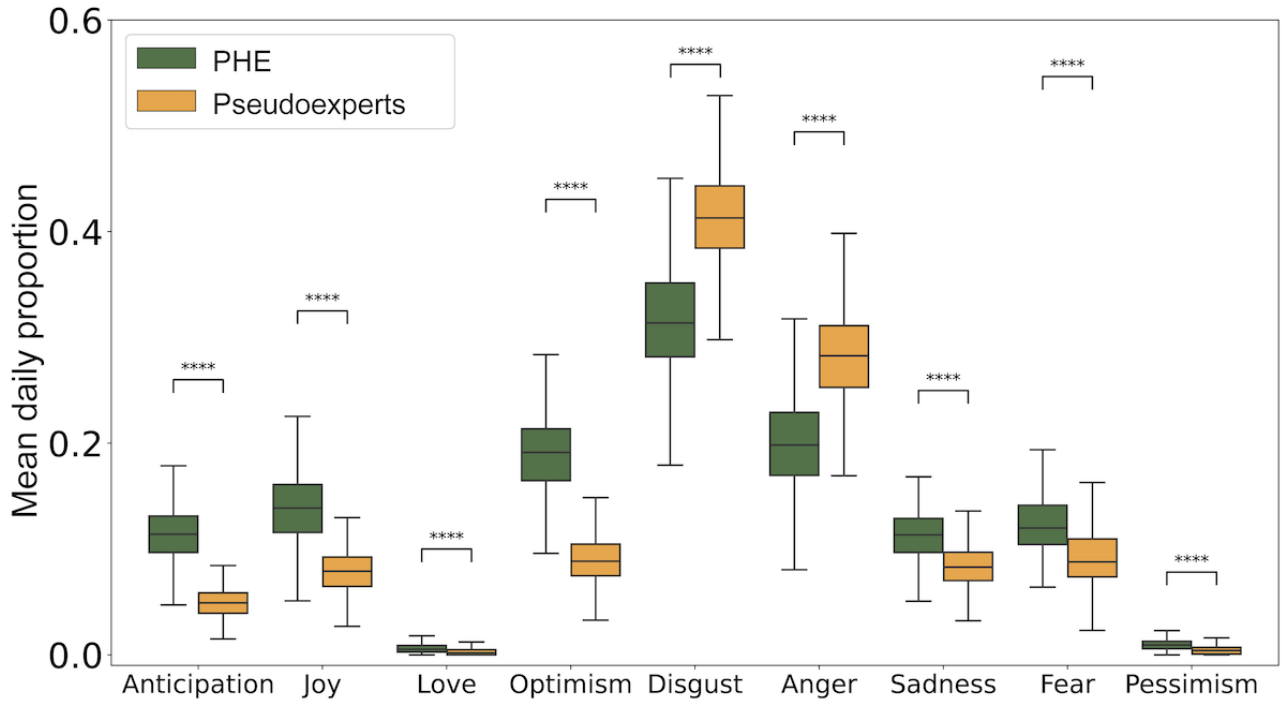

XSL•FO
RenderX

Dynamics of Affect

Emotions fluctuate over time and in response to events. Figure 5 illustrates the temporal dynamics of positive emotions expressed by PHEs and pseudoexperts. We leverage ChatGPT to summarize changes in emotions expressed, which are discussed further in Table S3 in Multimedia Appendix 1. Optimism and joy among PHEs surge following the announcement of stay-at-home orders post March 15, 2020. This can be attributed to factors such as gratitude for guidance by then New York governor Andrew Cuomo and enhanced accessibility to COVID-19 testing. Similarly, we note a corresponding albeit smaller increase among pseudoexperts, particularly in response to President Trump's management of the COVID-19 pandemic and France's endorsement of hydroxychloroquine as a viable COVID-19 treatment.

Figure 5. Dynamics of emotions. Daily fraction of posts from public health experts (PHEs) and pseudoexperts expressing (A and B) positive (optimism, joy, anticipation, and love) and (C and D) negative (disgust, anger, sadness, fear, and pessimism) emotions. Major events are marked with vertical dashed lines: Lockdowns (March 15, 2020, purple), health care (March 30, 2020, orange), therapeutics (April 24, 2020, yellow, Trump's bleach proposal), education (July 8, 2020, red, Trump's school reopening call), and vaccines (November 9, 2020, blue, Pfizer's 93% efficacy report).

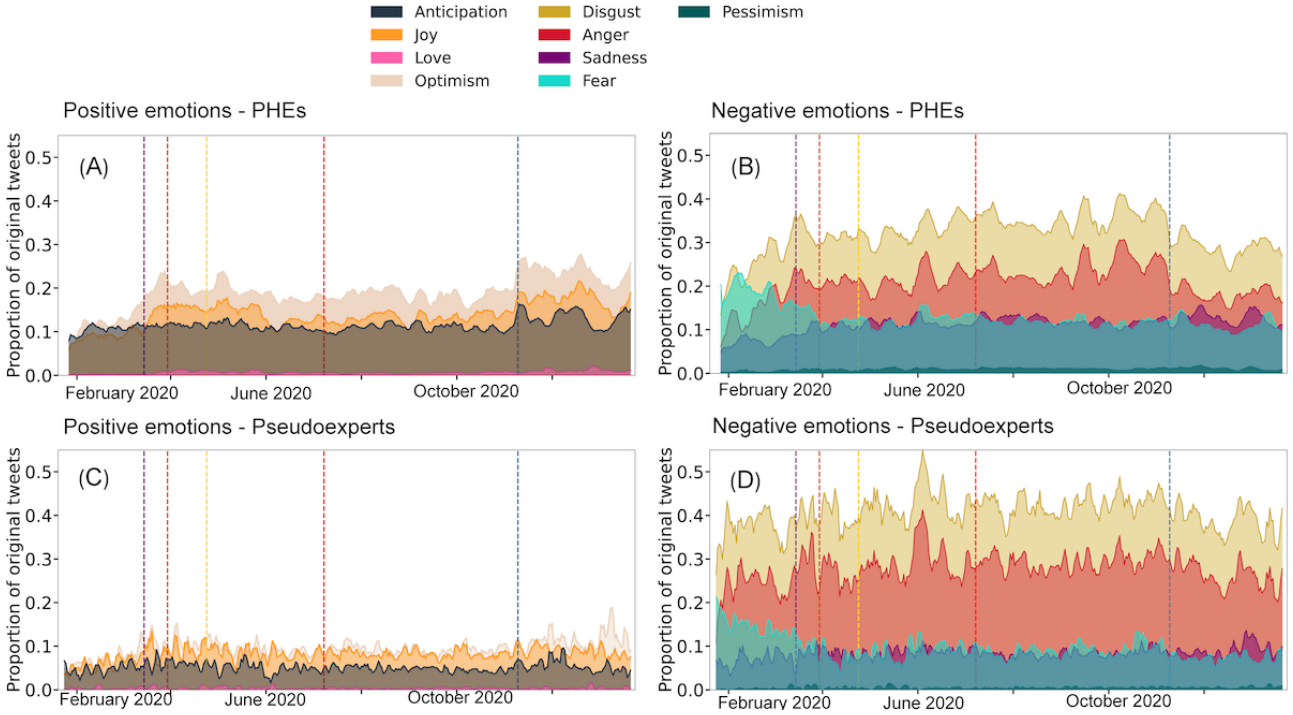

Another surge in joy, anticipation, and optimism among PHEs occurs after November 9, 2020, following Pfizer's announcement of successful phase 3 trials for its COVID-19 vaccine. PHEs hailed this development as a remarkable achievement and anticipated emergency use authorization from the Food and Drug Administration. While positive emotions also increased among pseudoexperts, the magnitude was notably lower. Pseudoexperts expressed optimism regarding the success of Operation Warp Speed, the imminent reopening of businesses, and the introduction of Lilly's monoclonal antibody drug.

Negative emotions such as disgust and anger escalated for both groups post March 15, 2020, with a more pronounced increase among pseudoexperts. The upsurge in anger and disgust within each group stemmed from different reasons. PHEs expressed disappointment with the measures taken by the Trump administration to combat the virus, whereas pseudoexperts voiced skepticism concerning the World Health Organization's interactions with China, Governor Cuomo's management of public transportation in New York, and the effectiveness of lockdowns in containing COVID-19. Although both groups experienced parallel declines in anger and disgust after US elections on November 9, 2020, the reductions were more significant among PHEs.

We also examine the use of moral language by the 2 groups. Figure S2 in Multimedia Appendix 1 compares the distribution of the daily share of posts expressing each moral foundation. Overall, PHEs use more positive moral language, emphasizing the dimensions of care, fairness, authority, loyalty, and purity, while pseudoexperts tend to prefer the negative moral dimensions of harm, cheating, subversion, and betrayal. The differences in use of moral language are more subdued compared to those for emotions. Figure S3 in Multimedia Appendix 1 illustrates the temporal dynamics of positive and negative moral language used by PHEs and pseudoexperts. We summarize the positive spikes using ChatGPT in Table S4 in Multimedia Appendix 1.

We witness an increase in the expression of care from PHEs after the stay-at-home orders. This increase is marked by calls for widespread lockdown measures, testing, and relief proposals for low-income households. However, there is a marginal decline in care language from pseudoexperts. Use of harm language decreases for both groups, with a more significant reduction for PHEs. In response to Pfizer's successful phase 3 trials, the use of care language increases for both PHEs and pseudoexperts coupled, accompanied by a decline in harm-related language. Both groups express care in discussing

XSL•FO
RenderX

how the introduction of vaccines could bring an end to the COVID-19 pandemic. In addition, pseudoexperts express concerns about the safety of the messenger RNA vaccines and criticize Bill Gates' call for vaccine mandates.

Asymmetries in Emotions and Moral Language

PHEs and pseudoexperts had conflicting priorities. PHEs promoted vaccination and advocated for stringent nonpharmaceutical interventions to curb the spread of the virus. In contrast, pseudoexperts expressed skepticism toward such interventions, emphasizing personal choice. We examine how these differences were manifested in the emotional and moral language used by the 2 groups.

To quantify issue-specific variation in emotions and moral language use by the 2 groups, we conduct a multivariate logistic regression analysis at the post level for each emotion and moral foundation. We examine the relationship between the issue discussed (independent variable) and the emotions or moral foundations expressed (dependent variables). In addition, the model incorporates a categorical variable to delineate between different groups. To account for potential differences in the emotional responses of the 2 groups, we introduce an interaction term between issues discussed and the group variable. We formulate the model separately for each issue as follows:

<Emotion> ~ origins + lockdowns + masking + education + health care + therapeutics + vaccines + (origins + lockdowns + masking + education +health care + therapeutics + vaccines) x group

, where group distinguishes between PHEs and pseudoexperts. We run separate regression models for each emotion. The coefficients for the main effects represent the change in the log-odds of the emotion for the PHEs when discussing an issue, while holding all other issues constant. In contrast, the sum of coefficients of the main effects and interaction effects quantify the change in log-odds for the pseudoexperts. For example, if the coefficient for lockdowns is positive, it suggests that when lockdowns are being discussed, there is an increase in the expression of a particular emotion from when they do not discuss lockdowns. An odds ratio >1 suggests that when a particular issue is discussed, there are increased odds of the post expressing an emotion compared to when the issue is not; an odds ratio=1 indicates equal odds, while an odds ratio <1 signifies lower odds.

Figure 6 compares the log-odds along with the corresponding SEs of estimation to show which group used more emotional language to frame a specific issue. The plot highlights differences between the 2 groups and gives insights into emotionally charged issues. The biggest gap in emotions appears on the issue of lockdowns, where pseudoexperts are far more likely to express anger, disgust, and sadness than PHEs. This position is consistent with the efforts to end the lockdowns (refer to the Great Barrington Declaration by Kulldorff et al [94]). The second largest gap in emotions appears in the discussion of therapeutics, where PHEs are more likely to express anger and disgust, but pseudoexperts are less likely to use these emotions. Pseudoexperts also use more positive language with more joy and optimism when talking about therapeutics, in contrast to PHEs, consistent with the highly contentious debates about this issue. Other notable differences include pseudoexperts expressing more fear and less joy about vaccines, while PHEs express less fear and more optimism.

Figure 6. Comparing emotions used by public health experts (PHEs) and pseudoexperts. It compares (A) anger, (B) disgust, (C) sadness, (D) fear, (E) optimism, and (F) joy across various topics. The figure displays the odds ratio of a post's relevance to a particular issue based on the expression of specific emotions by PHEs.

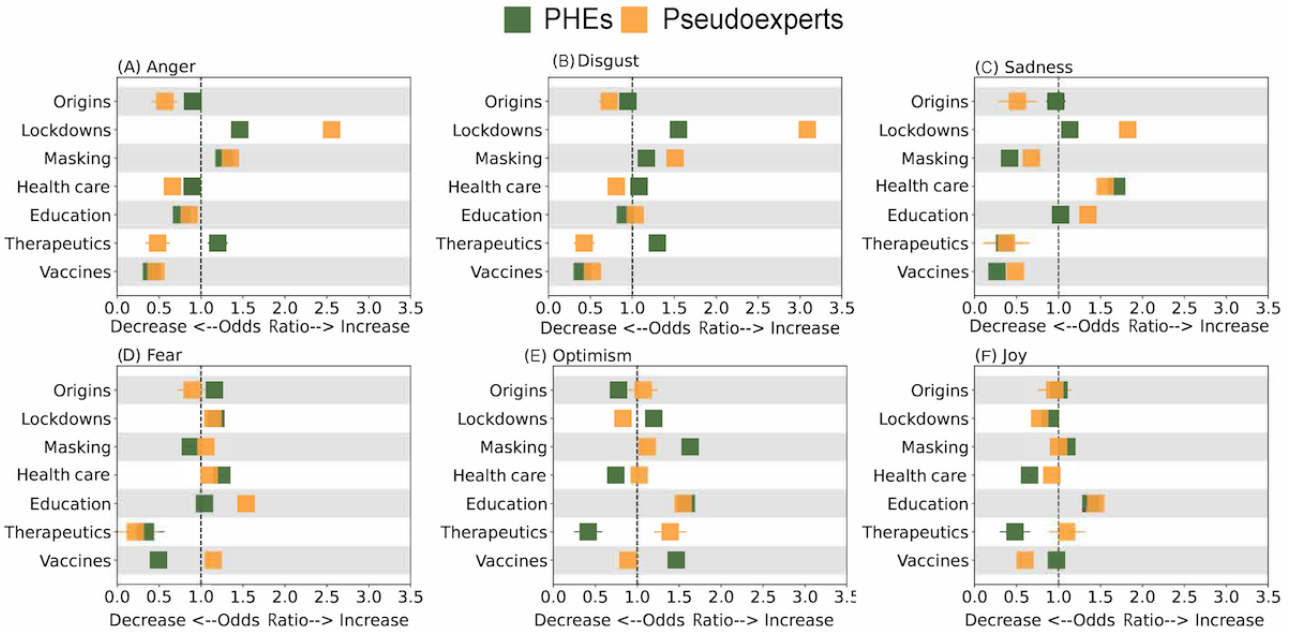

We conduct a similar analysis of the moral language used by the 2 groups of users. The coefficients for the main effects represent the change in the log-odds of the moral foundation for the PHEs when an issue is being discussed, while holding all other issues constant, and the sum of coefficients of the main effects and interaction effects quantifies the change in log-odds for the pseudoexperts. Figure 7 compares the log-odds, along with the corresponding SEs of estimation, to illustrate which

XSL•FO
RenderX

group relied more heavily on a given moral foundation when framing a specific issue. An odds ratio >1 suggests that when a particular issue is discussed, there are increased odds of the post expressing a moral foundation compared to when the issue is not being discussed; an odds ratio=1 indicates equal odds, while a ratio <1 signifies lower odds. The differences in moral language use are less pronounced compared to emotions. PHEs tend to emphasize care and loyalty in discussions of lockdowns and masking, consistent with their use of prosocial messaging that highlights the collective benefits of these measures. Conversely, pseudoexperts tend to convey more notions of harm, fairness, authority, and subversion when addressing lockdowns. This is in line with this issue being extremely contentious for them. Surprisingly, pseudoexperts are more likely to use fairness to frame their discussions of all issues, except vaccines.

Figure 7. Comparing use of moral foundations. It compares (A) care, (B) harm, (C) fairness, (D) cheating, (E) authority, (F) subversion, (G) loyalty, and (H) betrayal conveyed by public health experts (PHEs) and pseudoexperts across various topics. The figure displays the odds ratio of a post's relevance to a particular issue based on the expression of specific moral foundations by PHEs.

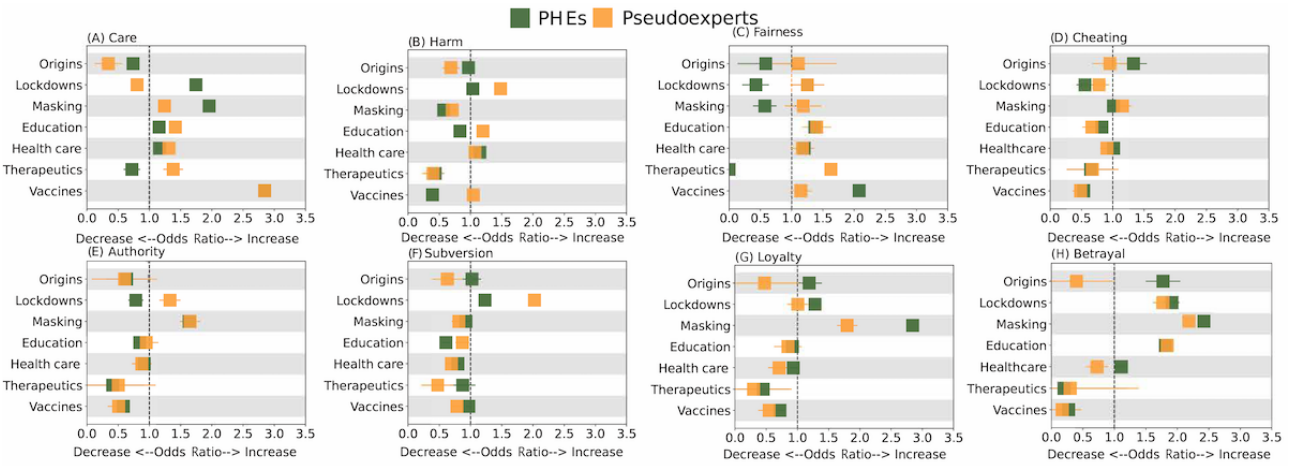

The comparison of emotions and moral foundations between PHEs and pseudoexperts highlights their conflicting positions on key COVID-19 pandemic–related issues, notably through the increased use of negative emotions on issues central to the other group. PHEs tend to focus on discussions related to vaccines, health care, and education, and these are issues on which we see more negative emotional framing by the pseudoexperts. In contrast, pseudoexperts are more focused on therapeutics and alternative treatments. Similarly, their negative framing of lockdowns and vaccines reflects their disapproval of the issues that were heavily promoted by PHEs. This divergence of affect underscores the polarization in the society at large. Understanding these differences is crucial for informing public health communication efforts that promote consensus within different segments of the population.

Affective Polarization in Health Communication

Studies show that public response to the COVID-19 pandemic became polarized fairly quickly, with political partisanship shaping online activity and discussions about the COVID-19 pandemic already in the early stages of the COVID-19 pandemic [12,17]. In addition, online discussions became emotionally polarized: when interacting with members of the opposite party, Twitter users expressed more anger and disgust, more toxicity, and less joy than in their interactions with same-party members [95]. Such interactions are characteristic of affective polarization [74], patterns of in-group love and out-group hate that have contributed to the growing partisan divide, and the erosion of trust between the 2 parties in the United States. As a result, partisanship predicted the adoption of COVID-19 pandemic prevention measures more than other factors [96].

To measure affective polarization, we analyze the emotional language PHEs and pseudoexperts directed at the political elites in their original posts. We use a previously curated list [97] of Twitter handles of >17,000 political elites, which include current and former senators, representatives, and media pundits. Figure 8 shows the proportion of posts from each group with mentions of political elites that express various affect. For instance, PHE-Lib indicates the share of PHEs' posts with a greater frequency of references to liberal elites compared to conservative ones, expressing a specific emotion or moral foundation. Figure 8 shows that PHEs post as liberal partisans: when they mention conservative elites in their posts, they use more negative emotions and moral subversion, but when they mention liberal elites, they express more positive emotions. In contrast, pseudoexperts are conservative partisans: they direct more negativity toward liberal elites, while expressing more positivity toward other conservative elites.

XSL•FO
RenderX

Figure 8. Asymmetries in affect toward elites. (A) Public health experts (PHEs) direct more negative emotions (anger and disgust) toward conservative elites and more positive emotions (joy and optimism) in their mentions of liberal elites, which is a hallmark of affective polarization. In contrast, pseudoexperts direct more negativity toward liberal elites. With respect to moral language, (B) PHEs express more subversion in their mentions of conservative elites, in contrast to pseudoexperts. Lib stands for liberal elites and Con stands for Conservative elites.

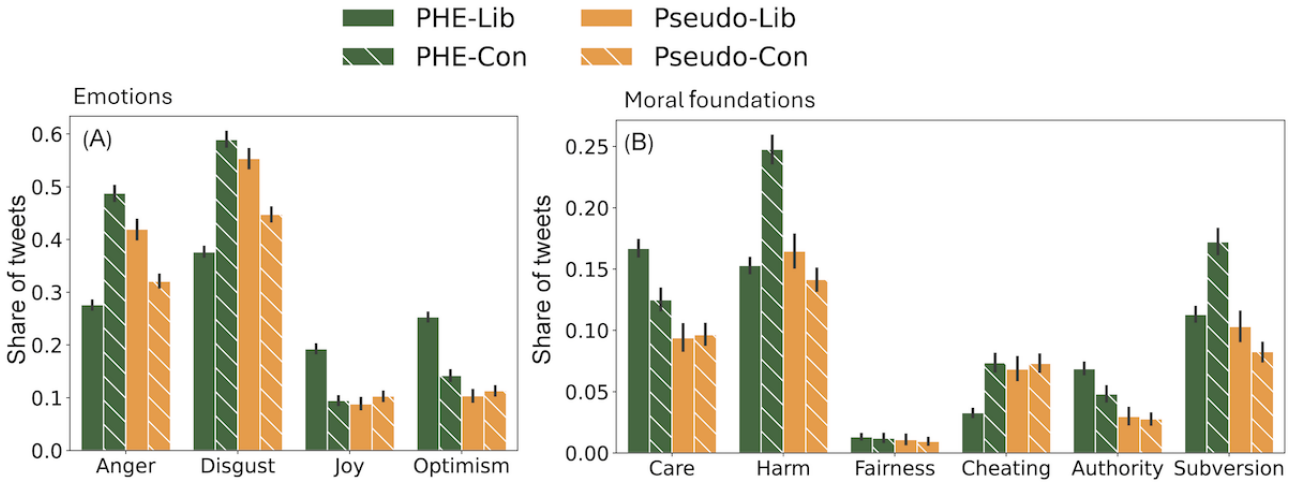

Figure S4 in Multimedia Appendix 1 lists the top 10 accounts that are more likely to be mentioned by PHEs and pseudoexperts positively or negatively. We then assess the repost interactions of PHEs and pseudoexperts with political elites. To this end, we construct a bipartite network comprising directed edges from PHEs or pseudoexperts to political elites, with each edge indicating the political elite reposted by a PHE or pseudoexpert.

Figure S5A in Multimedia Appendix 1 shows the distribution of ideology estimates for political elites active on Twitter during COVID-19. The ideology estimates were obtained from the study by McCabe [98] and were calculated using methods described in the study by Barberá [99]. The median ideology score of political elites on Twitter is 0.56, which is consistent with the past reports identifying the liberal skew of Twitter [100]. The median liberal and conservative elite have an ideology score of −0.73 and 1.28, respectively. Interestingly, the median ideology score of the elites reposted by PHEs and pseudoexperts is −0.76, indicating that a considerable share of the elites reposted by PHEs and pseudoexperts is more liberal and more conservative, respectively, than the median liberal and conservative elite. In Figure S5B in Multimedia Appendix 1, we visualize this network to find 2 highly clustered interactions. The color of the edge indicates the color of the source node, that is, PHEs (green) or pseudoexperts (orange). We find that PHEs mostly repost liberal elites, whereas pseudoexperts repost conservative ones, highlighting ideological clustering of scientific elites in the United States.

Overall, these findings highlight the existence of a partisan divide within the scientific community, as evidenced by the differential use of emotional and moral language by both PHEs and pseudoexperts toward liberal and conservative elites. Such polarization within the public health elites has implications for the perceived credibility of health messaging and, ultimately, the ability to foster consensus and cooperation in addressing public health challenges.

Public Engagement With PHEs

Effective public health communication relies not only on the dissemination of accurate information but also on how that information is received and interpreted by the public. To investigate this, we extracted original posts (excluding reposts, reply posts, and quoted posts) from PHEs, yielding us a corpus of 144,000 posts. We then collect replies to these posts, aiming to understand the engagement in response to these posts from PHEs. However, our efforts were hindered by challenges encountered when Twitter restricted access to their academic application programming interface, limiting our ability to gather reply interactions to all original posts.

We were able to collect replies for 195,000 original PHE posts, a total of 786,000 replies from 345K unique users. On average, each post received approximately 40.24 replies. The distribution of replies ranged widely, from a minimum of 1 reply to a maximum of 11,700 replies, with a median per post reply count of 5 (IQR 1-7). To quantify the effects of emotions and moral language use on the number of replies a post from PHEs receives, we use linear regression with the number of replies as the dependent variable and emotions or moral foundations expressed in the original post as the independent variables. With different PHEs having varying number of followers, the engagement their posts garner will also be varied. To control for this, we add the number of followers a post's author has as an independent variable. We execute 2 models, one for assessing the impact of emotions and the other for moral attitudes. The emotions model is formulated as follows:

replies ~ followers+ anger+ anticipation + disgust + fear + joy + love + optimism + pessimism + sadness, where replies represent the dependent variable, the number of replies to an original post from a PHE; followers is the number of that PHE's followers; and anger, anticipation, disgust, etc are binary variables indicating whether that emotion is present in the original post. A similar model can be written for moral foundations.

Figures 9A and 9B compare the impact of various emotions and moral foundations on engagement, respectively. We find that, controlling for the number of followers, presence of anger and disgust in the original post generates more replies. This is also true for negative moral language: presence of harm, cheating,

XSL•FO
RenderX

betrayal, and subversion over its corresponding moral virtues is associated with more replies. These findings add important nuance to previous research, which shows that posts expressing negative emotions [101] and moral outrage [69] are more likely to be reposted. Specifically, while negative emotions and moral language also receive more engagement in the form of replies, not all negative language leads to higher engagement: pessimism and sadness in the original posts is associated with fewer replies. Importantly, positive language can even suppress engagement, as is the case for original PHE posts expressing joy and love. Tables S5 and S6 in Multimedia Appendix 1 tabulate the results of this regression analysis.

Figure 9. Engagement with public health experts (PHEs). Dot and whisker plots show the regression coefficients and SEs. The coefficients represent the increase in number of replies when an (A) emotion or (B) moral foundation is used by the PHE in the original post while keeping others constant.

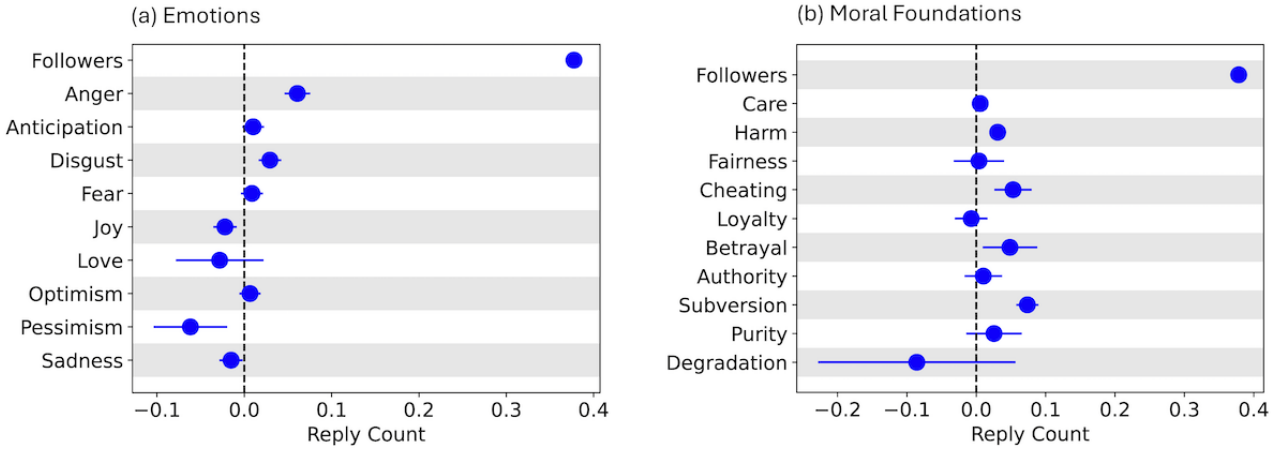

This leads us to question how users react to emotional and moral language of PHEs. We look at whether the use of a particular emotion (or moral foundation) in a PHE post triggers similar language in the replies. We use multivariate logistic regression models to quantify the odds ratio of a user expressing an emotion or moral attitude when presented with a post from the PHE containing certain emotional language (similarly extended to morals). An odds ratio >1 signifies that the user is more inclined to express an emotion or moral foundation, an odds ratio=1 indicates equivalent odds, and a odds ratio <1 suggests a decreased likelihood of expression. The following equation represents the model analyzing the relationship between emotions expressed in replies (<Emotion>_reply) and a set of predictors, including the original emotional content of the post being replied to (anger_orig, disgust_orig, fear_orig, sadness_orig, pessimism_orig, anticipation_orig, joy_orig, optimism_orig, love_orig) and the number of followers of the original poster (followers). Here, <Emotion>_reply denotes the intensity of a specific emotion (e.g., anger, joy) in the reply, while each <Emotion>orig variable represents the emotion in the original post.

$$<\text{Emotion}>_{\text{reply}} \sim \text{anger}_{\text{orig}} + \text{disgust}_{\text{orig}} + \text{fear}_{\text{orig}} + \text{sadness}_{\text{orig}} + \text{pessimism}_{\text{orig}} + \text{anticipation}_{\text{orig}} + \text{joy}_{\text{orig}} + \text{optimism}_{\text{orig}} + \text{love}_{\text{orig}} + \text{followers}$$

Figure 10A shows the odds ratio of users expressing a specific emotion in response to the emotion conveyed in a PHE's original post. We observe that users are more likely to express the same emotions as the original posts. Interestingly, when PHEs express joy and love, users are more likely to express joy, love, and optimism and less likely to express anger and disgust. Conversely, when PHEs convey negative emotions, users are more likely to express anger and disgust. Respondents generally match the emotion tone of the original posts, except when PHEs express sadness, which respondents counter with optimism. Figure 10B illustrates the odds ratio of users expressing a specific moral foundation in response to the moral foundation conveyed in a PHE's original post. Similar to emotions, we observe a mirroring effect in the use of most moral foundations between PHEs and ordinary users. Surprisingly, we notice a higher odds ratio of care being expressed when subversion is used by PHEs. These findings underscore the impact of emotional resonance in shaping user responses to public health messaging. When PHEs rely on positive framing, users tend to echo these positive sentiments. This suggests that the emotional and moral framing used by PHEs influences the emotion expressed by other users. Conversely, when negative framing is used, users tend to reflect these sentiments with increased negative expressions. These insights illuminate the importance of carefully crafting public health messaging to foster cohesive online discourse.

XSL•FO
RenderX

Figure 10. User engagement with public health experts (PHEs) and pseudoexperts. User reactions to (A) emotional and (B) moral appeals from PHEs. The figure demonstrates the odds ratio of users expressing emotions and moral principles in response to those conveyed in the original posts by PHEs. **P*<.05, ***P*<.01, ****P*<.001, and *****P*<.0001; ns: not significant.

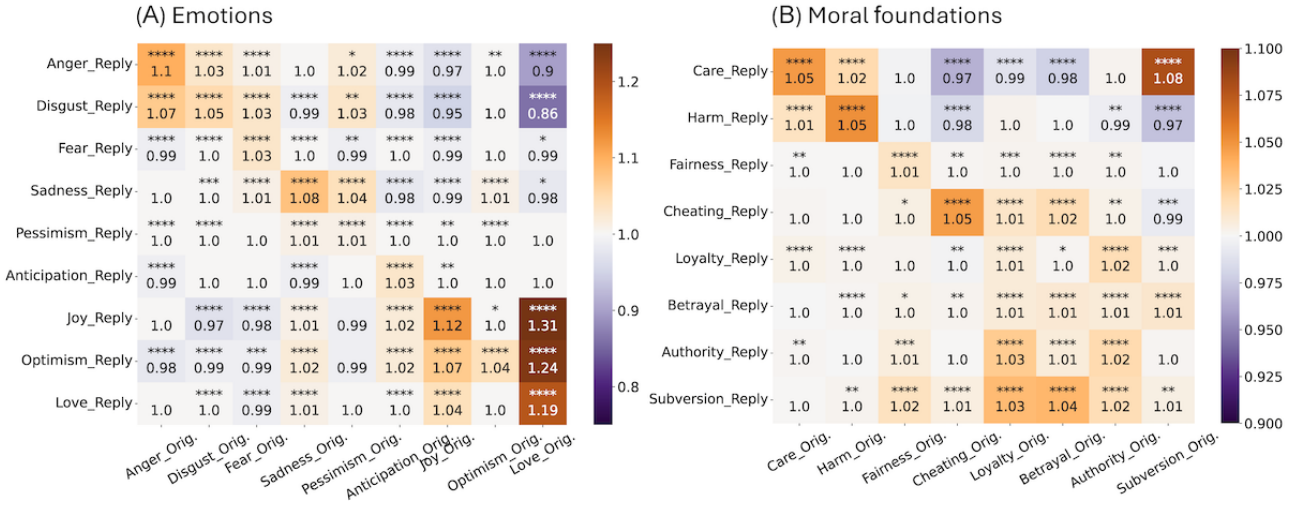

Discussion

Overview

The COVID-19 pandemic not only brought about an unprecedented global health crisis but also highlighted the critical role of effective communication in navigating public health challenges. Social media platforms, particularly Twitter, emerged as vital channels for health experts to disseminate timely and reliable information to the public. However, as the COVID-19 pandemic unfolded, discussions surrounding it became increasingly polarized, leading to the proliferation of misinformation and conspiracy theories, often propagated by pseudoexperts. By analyzing a substantial dataset of >515,000 posts generated by approximately 845 elites, this study delves into this infodemic, comparing the emotional and moral appeals used by PHEs and pseudoexperts on Twitter across various COVID-19 pandemic–related issues. In summary, our study offers the following key findings:

- PHEs focus on masking, health care, education, and vaccines, whereas pseudoexperts discuss therapeutics and lockdowns more frequently.
- PHEs typically used positive emotional language across all issues, expressing optimism and joy. Pseudoexperts often used negative emotions such as pessimism and disgust, while limiting positive emotional language to origins and therapeutics. Along the dimensions of moral language, PHEs and pseudoexperts differ on care versus harm, and authority versus subversion, across different issues.
- PHEs act as liberal partisans. They express more positive affect in their posts directed at liberals and more negative affect in their posts directed at conservative elites. In contrast, pseudoexperts act as conservative partisans.
- Negative emotional and moral language tends to boost engagement in COVID-19 discussions across all issues. However, the use of positive language by PHEs increases the use of positive language in the public responses.

Principal Findings

First, we categorized posts from PHEs and pseudoexperts into 7 different issues: origins of the COVID-19 pandemic, stay-at-home lockdown mandates, masking mandates, health care infrastructure, reopening the education system, therapeutics, and vaccinations. Our analysis of Twitter discourse from PHEs and pseudoexperts shows that they focused on different subsets of issues, similar to what was found in other study [24]. While PHEs predominantly focused on promoting public health measures such as social distancing, masking, improving health care infrastructure, and safer reopening of schools, pseudoexperts opposed lockdowns and mask mandates and promoted alternative views on therapeutics and virus origins.

Previous studies [13,97] have assessed the use of emotions and moral foundations expressed by users on social media. Harris et al [102] explored the influence of perceived experts in vaccine-related discussions, finding that they held key positions within the network, acting as central figures among antivaccine users and as bridges connecting the antivaccine and provaccine groups. Our study goes beyond previous studies by revealing emotional and moral divides among influential scientific and pseudoscientific elites on several COVID-19 issues. While PHEs expressed more positive emotions and emphasize moral virtues when discussing lockdowns, masking, health care, and vaccines, pseudoexperts expressed more anger and disgust in their posts on these issues and instead were more positive about therapeutics and alternative cures. The disparate use of emotional and moral language toward ideological elites showed that PHEs were aligned with liberal elites and pseudoexperts were aligned with conservative elites, potentially signaling the role of health influencers in increasing polarization.

Slavik et al [103] and van Dijck and Alinejad [104] assessed public engagement with public health messaging in Canada and the Netherlands during the emergency phase of the COVID-19 pandemic. Slavik et al [103] compared engagement levels across different message functions from health experts. Our study, in contrast, identifies a clear trend in emotional and moral language used by people in response to posts from PHEs. When these

XSL•FO
RenderX

experts express anger, disgust, or moral values, they tend to get more replies from users. In addition, we found that it is more likely for replies to echo the same emotional or moral sentiments as in the original posts from PHEs.

Limitations and Future Work

We note several areas of future research. Given that Twitter users are not a representative sample of the US population, our findings may primarily reflect the perspectives of a specific demographic (ie, younger, more liberal, better educated, and more interested in politics) [100]. Future studies can instead focus on multiple platforms and incorporate multimodal data. While our study examines COVID-19–related discourse, there is potential for investigations into scientific divisions in perspectives on polarized topics such as climate change and genetically modified foods. Moreover, exploring the growing debate on the factors contributing to the decline in adolescent mental health presents another avenue for inquiry.

Despite being state-of-the-art models to identify emotions and moral language [77,78], these models are not oracles. The emergence of more powerful, albeit expensive, instruction-tuned language models such as ChatGPT allows future work to leverage them at scale to identify emotions and moral attitudes with greater accuracy. However, given these tasks' inherent ambiguity, this application might not be straightforward. We emphasize that the event-related shifts in use of emotions and moral language allow us to make observational assertions rather than causal ones. Future studies can attempt to conduct natural experiments to quantify the impact of events on different cohorts. Aside from this, the disruption in our access to Twitter's Education Access application programming interface resulted in us only being able to collect replies for a subset of the PHEs posts in our dataset. However, it is important to note that the subset of posts we collected replies for were not intentionally sampled or biased in any way. Finally, while our dataset is extensive, it covers only the period from January 2020 to January 2021, limiting our findings to this time frame and excluding any potential shifts in perspectives that may have occurred afterward.

Conclusions

In summary, our study offers valuable insights into the dynamics of public health communication on social media amidst the unfolding COVID-19 pandemic, exploring viewpoints from both health experts and pseudoexperts. The identification of an ideological and emotional division in the scientific community poses a potential barrier to consensus building and undermines public trust in health messaging. Nevertheless, policy makers can leverage findings from interactions with PHEs to devise tailored strategies aimed at enhancing consensus. Tackling these obstacles demands a multifaceted approach, integrating fact-checking, debunking initiatives, and targeted communication efforts designed to cultivate trust and encourage critical thinking among the public.

Acknowledgments

This study is supported in part by the National Science Foundation under grant CCF 2200256. We use OpenAI's ChatGPT, a generative AI tool, to generate summaries for analyzing changes in the use of emotional and moral language.

Data Availability

The dataset analyzed during this study is available in the COVID-19 repository [105]. Given the restrictions imposed by X (formerly known as Twitter) on publicly sharing post objects, the authors state that there are certain restrictions on its availability. Only the post identifiers used in this study are publicly available. Readers can apply for application programming interface access from X to rehydrate these posts. The ideology classifier used to determine individual ideological leanings in this study is available on GitHub [106]; also available on GitHub are the emotion classifier [107] and the moral foundations classifier [108]. Auxiliary code used to conduct analyses will be released after acceptance [109].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional results.
[DOCX File , 3275 KB - jmir_v27i1e63910_app1.docx]

References

1. Tang Q, Zhang K, Li Y. The important role of social media during the COVID-19 epidemic. Disaster Med Public Health Prep 2021 Aug 10;15(4):e3-e4 [FREE Full text] [doi: 10.1017/dmp.2020.330] [Medline: 32907683]
2. Abbas J, Wang D, Su Z, Ziapour A. The role of social media in the advent of COVID-19 pandemic: crisis management, mental health challenges and implications. Risk Manag Healthc Policy 2021 May;14:1917-1932 [FREE Full text] [doi: 10.2147/RMHP.S284313] [Medline: 34012304]
3. Slothuus R, Bisgaard M. How political parties shape public opinion in the real world. Am J Polit Sci 2020 Nov 04;65(4):896-911. [doi: 10.1111/ajps.12550]

XSL•FO
RenderX

4. Bullock JG. Party cues. In: The Oxford Handbook of Electoral Persuasion. Oxford, UK: Oxford University Press; 2020.
5. Milich KM, Fisher N, Sobral G. Effective public health messaging for university students: lessons learned to increase adherence to safety guidelines during a pandemic. Humanit Soc Sci Commun 2024 Jan 03;11:38. [doi: 10.1057/s41599-023-02461-9]
6. Kim S, Capasso A, Cook SH, Ali SH, Jones AM, Foreman J, et al. Impact of COVID-19-related knowledge on protective behaviors: the moderating role of primary sources of information. PLoS One 2021 Nov 29;16(11):e0260643 [FREE Full text] [doi: 10.1371/journal.pone.0260643] [Medline: 34843590]
7. Utych SM, Fowler L. Age-based messaging strategies for communication about COVID-19. J Behav Public Adm 2020 Apr 06;3(1):1-14. [doi: 10.30636/jbpa.31.151]
8. McClaughlin E, Vilar-Lluch S, Parnell T, Knight D, Nichele E, Adolphs S, et al. The reception of public health messages during the COVID-19 pandemic. Appl Corpus Linguistics 2023 Apr;3(1):100037. [doi: 10.1016/j.acorp.2022.100037] [Medline: 37521321]
9. Lai S. Latino vaccination rates are high in one Maryland county. A cartoon grandmother helped. The Washington Post. 2021 Aug 08. URL: https://www.washingtonpost.com/dc-md-va/2021/ [accessed 2025-01-01]
10. Xia S, Nan X. Motivating COVID-19 vaccination through persuasive communication: a systematic review of randomized controlled trials. Health Commun 2024 Jul;39(8):1455-1478. [doi: 10.1080/10410236.2023.2218145] [Medline: 37254940]
11. Pink SL, Chu J, Druckman JN, Rand DG, Willer R. Elite party cues increase vaccination intentions among Republicans. Proc Natl Acad Sci U S A 2021 Aug 10;118(32):e2106559118 [FREE Full text] [doi: 10.1073/pnas.2106559118] [Medline: 34312254]
12. Jiang J, Chen E, Yan S, Lerman K, Ferrara E. Political polarization drives online conversations about COVID-19 in the United States. Hum Behav Emerg Technol 2020 Jul;2(3):200-211 [FREE Full text] [doi: 10.1002/hbe2.202] [Medline: 32838229]
13. Rao A, Guo S, Wang SY, Morstatter F, Lerman K. Pandemic culture wars: partisan differences in the moral language of COVID-19 discussions. arXiv Preprint posted online on May 29, 2023. [doi: 10.48550/arXiv.2305.18533]
14. Rao A, Morstatter F, Lerman K. Partisan asymmetries in exposure to misinformation. Sci Rep 2022 Sep 19;12(1):15671 [FREE Full text] [doi: 10.1038/s41598-022-19837-7] [Medline: 36123387]
15. The disinformation dozen. Center for Countering Digital Hate. 2021 Mar 24. URL: https://counterhate.com/research/the-disinformation-dozen/ [accessed 2025-01-01]
16. Goel A, Gupta L. Social media in the times of COVID-19. J Clin Rheumatol 2020 Sep;26(6):220-223 [FREE Full text] [doi: 10.1097/RHU.0000000000001508] [Medline: 32852927]
17. Rao A, Morstatter F, Hu M, Chen E, Burghardt K, Ferrara E, et al. Political partisanship and antiscience attitudes in online discussions about COVID-19: Twitter content analysis. J Med Internet Res 2021 Jun 14;23(6):e26692 [FREE Full text] [doi: 10.2196/26692] [Medline: 34014831]
18. Jiang J, Ren X, Ferrara E. Social media polarization and echo chambers in the context of COVID-19: case study. JMIRx Med 2021 Aug 5;2(3):e29570 [FREE Full text] [doi: 10.2196/29570] [Medline: 34459833]
19. Box-Steffensmeier JM, Moses L. Meaningful messaging: sentiment in elite social media communication with the public on the COVID-19 pandemic. Sci Adv 2021 Jul 14;7(29):eabg2898 [FREE Full text] [doi: 10.1126/sciadv.abg2898] [Medline: 34261655]
20. Green J, Edgerton J, Naftel D, Shoub K, Cranmer SJ. Elusive consensus: polarization in elite communication on the COVID-19 pandemic. Sci Adv 2020 Jul 10;6(28):eabc2717 [FREE Full text] [doi: 10.1126/sciadv.abc2717] [Medline: 32923600]
21. Van Scoy LJ, Snyder B, Miller EL, Toyobo O, Grewel A, Ha G, et al. Public anxiety and distrust due to perceived politicization and media sensationalism during early COVID-19 media messaging. J Commun Healthc 2021 Jul 31;14(3):193-205. [doi: 10.1080/17538068.2021.1953934]
22. Noar SM, Austin L. (Mis)communicating about COVID-19: insights from health and crisis communication. Health Commun 2020 Dec;35(14):1735-1739. [doi: 10.1080/10410236.2020.1838093] [Medline: 33112180]
23. Mabrey BE. The disinformation dozen and media misinformation on science and vaccinations. Oregon State University. 2021 Aug 27. URL: https://ir.library.oregonstate.edu/concern/honors_college_theses/n009w8842?locale=en [accessed 2025-01-12]
24. Nogara G, Vishnuprasad PS, Cardoso F, Ayoub O, Giordano S, Luceri L. The disinformation dozen: an exploratory analysis of COVID- 19 disinformation proliferation on Twitter. In: Proceedings of the 14th ACM Web Science Conference 2022. 2022 Presented at: WebSci '22; June 26-29, 2022; Barcelona, Spain. [doi: 10.1145/3501247.3531573]
25. Douglas KM. COVID-19 conspiracy theories. Group Process Intergr Relat 2021 Mar 04;24(2):270-275. [doi: 10.1177/1368430220982068]
26. Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]

XSL•FO
RenderX

27. Frenkel S, Alba D, Decker B. How the ‘plandemic’ movie and its falsehoods spread widely online. The New York Times. URL: https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html [accessed 2025-01-01]
28. Schaeffer K. A look at the Americans who believe there is some truth to the conspiracy theory that COVID-19 was planned. Pew Research Center. 2020 Jul 24. URL: https://www.pewresearch.org/short-reads/2020/07/24/a-look-at-the-americans-who-believe-there-is-some-truth-to-the-conspiracy-theory-that-covid-19-was-plann [accessed 2025-01-01]
29. Schaeffer K. Nearly three-in-ten Americans believe COVID-19 was made in a lab. Pew Research Center. 2020 Apr 08. URL: https://www.pewresearch.org/short-reads/2020/04/08/nearly-three-in-ten-americans-believe-covid-19-was-made-in-a-lab/ [accessed 2025-01-01]
30. Langguth J, Filkuková P, Brenner S, Schroeder DT, Pogorelov K. COVID-19 and 5G conspiracy theories: long term observation of a digital wildfire. Int J Data Sci Anal 2023;15(3):329-346 [FREE Full text] [doi: 10.1007/s41060-022-00322-3] [Medline: 35669096]
31. Taylor J. Bat soup, dodgy cures and 'diseasology': the spread of coronavirus misinformation. The Guardian. 2020 Jan 31. URL: https://www.theguardian.com/world/2020/jan/31/bat-soup-dodgy-cures-and-diseasology-the-spread-of-coronavirus-bunkum [accessed 2025-01-01]
32. Robins-Early N. The strange origins of Trump's hydroxychloroquine obsession. HuffPost. 2020 May 13. URL: https://www.huffpost.com/entry/trump-hydroxychloroquine-coronavirus-fox-news_n_5ebaffdbc5b65b5fd63dac80 [accessed 2025-01-01]
33. Woo E. How covid misinformation created a run on animal medicine. The New York Times. 2021 Sep 28. URL: https://www.nytimes.com/2021/09/28/technology/ivermectin-animal-medicine-shortage.html [accessed 2025-01-01]
34. Desai AN, Ruidera D, Steinbrink JM, Granwehr B, Lee DH. Misinformation and disinformation: the potential disadvantages of social media in infectious disease and how to combat them. Clin Infect Dis 2022 May 15;74(Suppl_3):e34-e39 [FREE Full text] [doi: 10.1093/cid/ciac109] [Medline: 35568471]
35. Kearney MD, Chiang SC, Massey PM. The Twitter origins and evolution of the COVID-19 “plandemic” conspiracy theory. HKS Misinf Rev 2020 Oct 9 [FREE Full text] [doi: 10.37016/mr-2020-42]
36. Ahmed ME, Rabin MR, Chowdhury FN. COVID-19: social media sentiment analysis on reopening. arXiv Preprint posted online on June 01, 2020. [doi: 10.48550/arXiv.2006.00804]
37. Gruzd A, Mai P. Going viral: how a single tweet spawned a COVID-19 conspiracy theory on Twitter. Big Data Soc 2020 Jul 20;7(2). [doi: 10.1177/2053951720938405]
38. Ferrara E. What types of COVID-19 conspiracies are populated by Twitter bots? First Monday 2020 May 19;25(6). [doi: 10.5210/fm.v25i6.10633]
39. Gottlieb M, Dyer S. Information and disinformation: social media in the COVID-19 crisis. Acad Emerg Med 2020 Jul;27(7):640-641 [FREE Full text] [doi: 10.1111/acem.14036] [Medline: 32474977]
40. Antonakis J. Leadership to defeat COVID-19. Group Process Intergr Relat 2021 Mar 04;24(2):210-215. [doi: 10.1177/1368430220981418]
41. Geleris J, Sun Y, Platt J, Zucker J, Baldwin M, Hripcsak G, et al. Observational study of hydroxychloroquine in hospitalized patients with covid-19. N Engl J Med 2020 Jun 18;382(25):2411-2418 [FREE Full text] [doi: 10.1056/NEJMoa2012410] [Medline: 32379955]
42. Sehgal NJ, Yue D, Pope E, Wang RH, Roby DH. The association between COVID-19 mortality and the county-level partisan divide in the United States. Health Aff 2022 Jun;41(6):853-863. [doi: 10.1377/hlthaff.2022.00085]
43. Pradelle A, Mainbourg S, Provencher S, Massy E, Grenet G, Lega JC. Deaths induced by compassionate use of hydroxychloroquine during the first COVID-19 wave: an estimate. Biomed Pharmacother 2024 Feb;171:116055 [FREE Full text] [doi: 10.1016/j.biopha.2023.116055] [Medline: 38171239]
44. Renström EA, Bäck H. Emotions during the Covid-19 pandemic: fear, anxiety, and anger as mediators between threats and policy support and political actions. J Appl Soc Psychol 2021 Aug;51(8):861-877. [doi: 10.1111/jasp.12806] [Medline: 34511637]
45. Sadler MS, Lineberger M, Correll J, Park B. Emotions, attributions, and policy endorsement in response to the September 11th terrorist attacks. Basic Appl Soc Psychol 2010 Jun 07;27(3):249-258. [doi: 10.1207/s15324834basp2703_6]
46. Skitka LJ, Bauman CW, Aramovich NP, Morgan GS. Confrontational and preventative policy responses to terrorism: anger wants a fight and fear wants "them" to go away. Basic Appl Soc Psychol 2010 Jun 07;28(4):375-384. [doi: 10.1207/s15324834basp2804_11]
47. Hatemi PK, McDermott R, Eaves LJ, Kendler KS, Neale MC. Fear as a disposition and an emotional state: a genetic and environmental approach to out-group political preferences. Am J Polit Sci 2013 Jan 28;57(2):279-293. [doi: 10.1111/ajps.12016]
48. Stress and decision-making during the pandemic. American Psychological Association. 2021 Oct 26. URL: https://www.apa.org/news/press/releases/stress/2021/october-decision-making [accessed 2025-01-01]
49. Agrawal S, Jain SK, Sharma S, Khatri A. COVID-19 public opinion: a Twitter healthcare data processing using machine learning methodologies. Int J Environ Res Public Health 2022 Dec 27;20(1):432 [FREE Full text] [doi: 10.3390/ijerph20010432] [Medline: 36612755]

XSL•FO
RenderX

50. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
51. Wheaton MG, Prikhidko A, Messner GR. Is fear of COVID-19 contagious? The effects of emotion contagion and social media use on anxiety in response to the coronavirus pandemic. Front Psychol 2020 Jan 5;11:567379 [FREE Full text] [doi: 10.3389/fpsyg.2020.567379] [Medline: 33469434]
52. Díaz R, Cova F. Reactance, morality, and disgust: the relationship between affective dispositions and compliance with official health recommendations during the COVID-19 pandemic. Cogn Emot 2022 Mar 16;36(1):120-136. [doi: 10.1080/02699931.2021.1941783] [Medline: 34132171]
53. Chan EY. Moral foundations underlying behavioral compliance during the COVID-19 pandemic. Pers Individ Dif 2021 Mar;171:110463 [FREE Full text] [doi: 10.1016/j.paid.2020.110463] [Medline: 33106715]
54. Reimer NK, Atari M, Karimi-Malekabadi F, Trager J, Kennedy B, Graham J, et al. Moral values predict county-level COVID-19 vaccination rates in the United States. Am Psychol 2022 Oct;77(6):743-759. [doi: 10.1037/amp0001020] [Medline: 36074569]
55. Nan X, Wang Y, Thier K, Adebamowo C, Quinn S, Ntiri S. Moral foundations predict COVID-19 vaccine hesitancy: evidence from a national survey of Black Americans. J Health Commun 2022 Dec 02;27(11-12):801-811. [doi: 10.1080/10810730.2022.2160526] [Medline: 36576158]
56. Pacheco ML, Islam T, Mahajan M, Shor A, Yin M, Ungar L, et al. A holistic framework for analyzing the COVID-19 vaccine debate. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022 Presented at: NAACL 2022; July 10-15, 2022; Seattle, WA. [doi: 10.18653/v1/2022.naacl-main.427]
57. John NA. The social logics of sharing. Commun Rev 2013 Jul;16(3):113-131. [doi: 10.1080/10714421.2013.807119]
58. Bazarova NN, Choi YH, Sosik VS, Cosley D, Whitlock J. Social sharing of emotions on Facebook: channel differences, satisfaction, and replies. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015 Presented at: CSCW '15; March 14-18, 2015; Vancouver, BC. [doi: 10.1145/2675133.2675297]
59. Burke M, Develin M. Once more with feeling: supportive responses to social sharing on Facebook. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16; February 27-March 2, 2016; San Francisco, CA. [doi: 10.1145/2818048.2835199]
60. Kim J, Yoo J. Role of sentiment in message propagation: reply vs. retweet behavior in political communication. In: Proceedings of the International Conference on Social Informatics. 2012 Presented at: SocialInformatics 2012; December 14-16, 2012; Alexandria, VA. [doi: 10.1109/socialinformatics.2012.33]
61. Sousa D, Sarmento L, Rodrigues EM. Characterization of the Twitter @replies network: are user ties social or topical? In: Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents. 2010 Presented at: SMUC '10; October 30, 2010; Toronto, ON. [doi: 10.1145/1871985.1871996]
62. Liu J, Bojja GR, Ofori M, Sai Ambati L. Early public outlook on the coronavirus disease (COVID-19): a social media study. In: Proceedings of the 26th Americas Conference on Information Systems. 2020 Presented at: AMCIS 2020; August 15-17, 2020; Virtual Event.
63. Miyazaki K, Uchiba T, Tanaka K, Sasahara K. Characterizing the anti-vaxxers' reply behavior on social media. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2021 Presented at: WI-IAT '21; December 14-17, 2021; Melbourne, Australia. [doi: 10.1145/3498851.3498934]
64. Gallagher RJ, Doroshenko L, Shugars S, Lazer D, Foucault Welles B. Sustained online amplification of COVID-19 elites in the United States. Soc Media Soc 2021 Jun 25;7(2). [doi: 10.1177/20563051211024957]
65. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
66. Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. Bad is stronger than good. Rev Gen Psychol 2001;5(4):323-370. [doi: 10.1037//1089-2680.5.4.323]
67. Coviello L, Sohn Y, Kramer AD, Marlow C, Franceschetti M, Christakis NA, et al. Detecting emotional contagion in massive social networks. PLoS One 2014 Mar 12;9(3):e90315 [FREE Full text] [doi: 10.1371/journal.pone.0090315] [Medline: 24621792]
68. Ferrara E, Yang Z. Measuring emotional contagion in social media. PLoS One 2015 Nov 6;10(11):e0142390 [FREE Full text] [doi: 10.1371/journal.pone.0142390] [Medline: 26544688]
69. Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ. Emotion shapes the diffusion of moralized content in social networks. Proc Natl Acad Sci U S A 2017 Jul 11;114(28):7313-7318 [FREE Full text] [doi: 10.1073/pnas.1618923114] [Medline: 28652356]
70. Brady WJ, McLoughlin K, Doan TN, Crockett M. How social learning amplifies moral outrage expression in online social networks. Sci Adv 2021;7(33):eabe5641. [doi: 10.31234/osf.io/gf7t5]
71. Grover T, Bayraktaroglu E, Mark G, Rho EH. Moral and affective differences in U.S. immigration policy debate on Twitter. Comput Supported Coop Work 2019 May 26;28(3-4):317-355. [doi: 10.1007/s10606-019-09357-w]

XSL•FO
RenderX

72. Sterling J, Jost JT. Moral discourse in the Twitterverse: effects of ideology and political sophistication on language use among U.S. citizens and members of Congress. J Lang Polit 2017 Nov 24;17(2):195-221. [doi: 10.1075/jlp.17034.ste]
73. Sagi E, Dehghani M. Moral rhetoric in Twitter: a case study of the U.S. federal shutdown of 2013. Kellogg School of Management, Northwestern University. 2014. URL: https://www.kellogg.northwestern.edu/faculty/research/detail/2014/moral-rhetoric-in-twitter-a-case-study-of-the/ [accessed 2025-01-01]
74. Iyengar S, Westwood SJ. Fear and loathing across party lines: new evidence on group polarization. Am J Polit Sci 2014 Dec 16;59(3):690-707. [doi: 10.1111/ajps.12152]
75. Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ. The origins and consequences of affective polarization in the United States. Annu Rev Polit Sci 2019 May 11;22(1):129-146. [doi: 10.1146/annurev-polisci-051117-073034]
76. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194. [doi: 10.1016/s2589-7500(20)30315-0]
77. Alhuzali H, Ananiadou S. SpanEmo: casting multi-label emotion classification as span-prediction. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. 2021 Presented at: EACL 2021; April 19-23, 2021; Online. [doi: 10.18653/v1/2021.eacl-main.135]
78. Guo S, Mokhberian N, Lerman K. A data fusion framework for multi-domain morality learning. Proc Int AAAI Conf Web Soc Media 2023 Jun 02;17(1):281-291. [doi: 10.1609/icwsm.v17i1.22145]
79. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
80. Metaxas P, Mustafaraj E, Wong K, Zeng L, O'Keefe M, Finn S. What do retweets indicate? Results from user survey and meta-review of research. Proc Int AAAI Conf Web Soc Media 2021 Aug 03;9(1):658-661. [doi: 10.1609/icwsm.v9i1.14661]
81. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In: Proceedings of the 43rd Hawaii International Conference on System Sciences. 2010 Presented at: HICSS 2010; January 5-18, 2010; Honolulu, HI. [doi: 10.1109/hicss.2010.412]
82. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
83. Colleoni E, Rozza A, Arvidsson A. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. J Commun 2014 Mar 19;64(2):317-332. [doi: 10.1111/jcom.12084]
84. Kang JH, Lerman K. Using lists to measure homophily on Twitter. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012 Presented at: AAAI'12; July 22-26, 2012; Toronto, ON.
85. Ghosh R, Lerman K. Predicting influential users in online social networks. arXiv Preprint posted online on May 26, 2010 [FREE Full text]
86. Media Bias/fact Check homepage. Media Bias/Fact Check. URL: https://mediabiasfactcheck.com/ [accessed 2024-01-09]
87. Eisenstein J, Ahmed A, Xing EP. Sparse additive generative models of text. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. 2011 Presented at: ICML'11; June 28-July 2, 2011; Bellevue, WA.
88. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S. SemEval-2018 task 1: affect in tweets. In: Proceedings of the 12th International Workshop on Semantic Evaluation. 2018 Presented at: SemEval 2018; June 5-6, 2018; New Orleans, LA. [doi: 10.18653/v1/s18-1001]
89. Haidt J, Graham J. When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. Soc Just Res 2007 May 23;20(1):98-116. [doi: 10.1007/s11211-007-0034-z]
90. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online on October 11, 2018 [FREE Full text]
91. Fatemi Z, Bhattacharya A, Wentzel A, Dhariwal V, Levine L, Rojecki A, et al. Understanding stay-at-home attitudes through framing analysis of tweets. arXiv Preprint posted online on September 13, 2022 [FREE Full text] [doi: 10.2307/1944168]
92. Hoover J, Portillo-Wightman G, Yeh L, Havaldar S, Davani AM, Lin Y, et al. Moral Foundations Twitter Corpus: a collection of 35k tweets annotated for moral sentiment. Soc Psychol Pers Sci 2020 Feb 19;11(8):1057-1071. [doi: 10.1177/1948550619876629]
93. Johnson K, Goldwasser D. Classification of moral foundations in microblog political discourse. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018 Presented at: ACL 2018; July 15-20, 2018; Melbourne, Australia. [doi: 10.18653/v1/p18-1067]
94. Kulldorff M, Gupta S, Bhattacharya J. The Great Barrington Declaration. Great Barrington Declaration. URL: https://gbdeclaration.org/ [accessed 2025-01-01]
95. Lerman K, Feldman D, He Z, Rao A. Affective polarization and dynamics of information spread in online networks. NPJ Complexity 2024 Jun 07;1:8 [FREE Full text]
96. Grossman G, Kim S, Rexer JM, Thirumurthy H. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. Proc Natl Acad Sci U S A 2020 Sep 29;117(39):24144-24153 [FREE Full text] [doi: 10.1073/pnas.2007835117] [Medline: 32934147]

XSL•FO
RenderX

97. Guo S, Burghardt K, Rao A, Lerman K. Emotion regulation and dynamics of moral concerns during the early COVID-19 pandemic. arXiv. Preprint posted online on March 7, 2022 2024 [FREE Full text]
98. McCabe S. sdmccabe / new-tweetscores. GitHub. URL: https://github.com/sdmccabe/new-tweetscores [accessed 2024-01-09]
99. Barberá P. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. Polit Anal 2017 Jan 04;23(1):76-91. [doi: 10.1093/pan/mpu011]
100. Wojcik S, Hughes A. Sizing up Twitter users. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2025-01-01]
101. Schöne JP, Parkinson B, Goldenberg A. Negativity spreads more than positivity on Twitter after both positive and negative political situations. Affect Sci 2021 Dec 12;2(4):379-390 [FREE Full text] [doi: 10.1007/s42761-021-00057-7] [Medline: 36043036]
102. Harris MJ, Murtfeldt R, Wang S, Mordecai EA, West JD. Perceived experts are prevalent and influential within an antivaccine community on Twitter. PNAS Nexus 2024 Feb;3(2):pgae007 [FREE Full text] [doi: 10.1093/pnasnexus/pgae007] [Medline: 38328781]
103. Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N. Examining tweet content and engagement of Canadian public health agencies and decision makers during COVID-19: mixed methods analysis. J Med Internet Res 2021 Mar 11;23(3):e24883 [FREE Full text] [doi: 10.2196/24883] [Medline: 33651705]
104. van Dijck J, Alinejad D. Social media and trust in scientific expertise: debating the Covid-19 pandemic in the Netherlands. Soc Media Soc 2020 Dec 15;6(4). [doi: 10.1177/2056305120981057]
105. COVID-19 TweetIDs. Harvard Dataverse. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DKOVLA [accessed 2025-01-23]
106. Multidimensional ideological polarization. GitHub. URL: https://tinyurl.com/yu7xxsey [accessed 2025-01-17]
107. hasanhuz / SpanEmo. GitHub. URL: https://github.com/hasanhuz/SpanEmo [accessed 2025-01-17]
108. DAMF - moral foundations inference with domain adapting ability. GitHub. URL: https://github.com/fionasguo/DAMF [accessed 2025-01-17]
109. ashwinshreyas96 / Public-Health-Messaging-on-X. GitHub. URL: https://github.com/ashwinshreyas96/Public-Health-Messaging-on-X [WebCite Cache ID https://github.com/ashwinshreyas96/Public-Health-Messaging-on-X]

Abbreviations

PHE: public health expert
PLD: pay-level domain

Edited by C Argyropoulos; submitted 06.07.24; peer-reviewed by C Peters, H Pilabré; comments to author 20.10.24; revised version received 10.11.24; accepted 10.12.24; published 05.02.25.

Please cite as:
Rao A, Sabri N, Guo S, Raschid L, Lerman K
Public Health Messaging on Twitter During the COVID-19 Pandemic: Observational Study
J Med Internet Res 2025;27:e63910
URL: https://www.jmir.org/2025/1/e63910
doi:10.2196/63910
PMID:

©Ashwin Rao, Nazanin Sabri, Siyi Guo, Louiqa Raschid, Kristina Lerman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Continuous Glucose Monitoring on Short-Term, In-Hospital Mortality Among Frail and Critically Ill Patients With COVID-19: Randomized Controlled Trial

Jiawei Shang[1*], MD; Ziming Yuan[1*], MD; Zuoyan Zhang[1], BS; Quanhong Zhou[1], MD; Yan Zou[1*], MD; Wei Wang[1*], MD

Department of Intensive Care Medicine, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Wei Wang, MD
Department of Intensive Care Medicine
Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine
600 Yishan Road
Shanghai, 200233
China
Phone: 86 18930173816
Fax: 86 2164368031
Email: drwangwei3037@126.com

Abstract

Background: The use of continuous glucose monitoring (CGM) in the hospital setting is growing, with more patients using these devices at home, especially during the COVID-19 pandemic. Frail and critically ill patients with COVID-19 and previously normal glucose tolerance are also associated with variability in their glucose levels during their intensive care unit (ICU) stay. However, very limited evidence supports the use of CGM in ICU settings, especially among frail patients with COVID-19.

Objective: We aimed to investigate the effectiveness of CGM on ICU-related outcomes among frail and critically ill patients with confirmed COVID-19.

Methods: This was an exploratory, prospective, open-label, parallel, single-center, randomized controlled trial. A total of 124 patients was finally analyzed. The primary outcome was 28-day, in-ICU mortality. The secondary outcome included the length of ICU stay as well as the occurrence of hypoglycemia and severe hypoglycemia events.

Results: The mean age was 78.3 (SD 11.5) years. The mean fasting glucose level and hemoglobin A_{1c} level at baseline were 8.12 (SD 1.54) mmol/L and 7.2% (SD 0.8%), respectively. The percentage of participants with diabetes was 30.6% (38/124). The corresponding hazard ratio of the primary outcome in the intermittently scanned CGM (isCGM) group when compared with the point-of-care testing (POCT) group was 0.18 (95% CI 0.04-0.79). The average length of ICU stay was 10.0 (SD 7.57) days in the isCGM group and 14.0 (SD 6.86) days in the POCT group (P=.02). At the end of study period, the mean value of fasting glucose in the isCGM group and the POCT group was 6.07 (SD 0.63) mmol/L and 7.76 (SD 0.62) mmol/L, respectively (P=.01). A total of 207 hypoglycemia events (<3.9 mmol/L) was detected, with 43 in the isCGM group and 164 in the POCT group (P<.001). A total of 81 severe hypoglycemia events (<2.8 mmol/L) was detected, with 16 in the isCGM group and 65 in the POCT group (P<.001). The major adverse event in this study was bleeding in the puncture site, with a total of 6 occurrences in the isCGM group. During the follow-up, none of the participants dropped out because of bleeding in the puncture site.

Conclusions: We found a significant clinical benefit from the use of CGM among frail and critically ill patients with COVID-19. These findings support the use of CGM in the ICU and might help with the extension of application in various in-hospital settings.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200059733; https://www.chictr.org.cn/showproj.html?proj=169257

(J Med Internet Res 2025;27:e67012) doi:10.2196/67012

XSL•FO
RenderX

KEYWORDS

intermittently scanned continuous glucose monitoring; isCGM; COVID-19; in-ICU mortality; continuous glucose monitoring; CGM; point of care testing; POCT; glucose monitoring; in-hospital mortality; mortality; inpatient mortality; critically ill; frail; SARS-CoV-2; intensive care unit; ICU; exploratory; prospective; randomized; open label; parallel; single center; clinical trial

Introduction

Continuous glucose monitoring (CGM) is a method of tracking glucose levels throughout the day and night [1]. CGM systems take glucose measurements at regular intervals, 24 hours a day, and translate the readings into dynamic data, generating glucose direction and rate-of-change reports. It is now widely used, medically prescribed, or commercially available for both patients with and without diabetes [2,3]. Several studies have also suggested that CGM has the potential to become the standard of care for some hospitalized patients, overcoming the limitations of current capillary glucose testing [4,5].

The COVID 19 pandemic continues to have a significant impact on the health and wellness of humans despite significant advances in its diagnosis and treatment [6]. CGM has also been used in the management of patients with COVID-19 [7]. In the United States, the Food and Drug Administration (FDA) issued temporary guidance for expanded use of noninvasive remote monitoring devices that reduce contact between clinicians and patients during the COVID-19 pandemic, effectively allowing CGM systems to be used in hospital settings [8].

Several reports indicate that diabetes and hyperglycemia are linked to worse morbidity and mortality in individuals with COVID-19 [9,10]. On the other hand, frail and critically ill patients with COVID-19 and previously normal glucose tolerance are also associated with variability in their glucose levels during their intensive care unit (ICU) stay, which might be caused by the critical illness itself or certain medication therapies such as glucocorticoids and immunosuppressants. Therefore, maintaining optimal glucose levels in hospitalized patients with frailty and critical illness is a key component of care.

However, very limited evidence supports the use of CGM in ICU settings, especially among frail patients with COVID-19. We thus performed an exploratory, prospective, open-label, parallel, single-center, randomized controlled trial (RCT) to investigate the effectiveness of CGM on short-term, in-hospital mortality among frail and critically ill patients with COVID-19.

Methods

Study Design

This is an exploratory, prospective, open-label, parallel, single-center RCT. Eligible participants were recruited from Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine at Lingang Campus and were randomly divided into the intervention group (wearing a CGM device) and the control group (routine fingertip measurements). The allocation ratio was 1:1. This trial was registered at the Chinese Clinical Trial Registry (ChiCTR2200059733) before the first recruitment of the participants.

Ethical Considerations

This trial was approved by the Institutional Review Board of Shanghai Sixth People's Hospital (approval 2022-KY-054). All participants provided written informed consents, and all data collected were only used in this study. All data collected in this study have been anonymized to ensure participant privacy and confidentiality. Personal identifiers have been removed, and unique codes were used to manage data without any direct link to participants' identities. No compensation was provided to participants in this study. The research was conducted voluntarily, and participants were informed of this during the consent process.

Recruitment and Screening

The recruitment process involved identifying potential participants from hospital ICU registries. Hospital databases were systematically reviewed to locate patients meeting the initial eligibility criteria, including age, confirmed COVID-19 diagnosis, and indicators of critical illness. Once potential participants were identified, their medical records were thoroughly examined to verify clinical details, including diagnostic results from real-time, reverse transcription polymerase chain reaction (RT-PCR) assays and assessments of respiratory function, oxygen saturation levels, and pulmonary imaging findings. Following this initial review, clinical evaluations were conducted to confirm the presence of the inclusion criteria. In parallel, exclusion criteria were assessed by reviewing the patients' mental health status and cognitive function to identify any clinically significant psychosis or cognitive impairment that might interfere with their ability to participate in or comply with the study protocol.

Study Participants

Eligible participants were frail older adults aged 65 years or older who were diagnosed with COVID-19 and critical illness. COVID-19 cases were confirmed by real-time RT-PCR assay of nasal and pharyngeal swab specimens. Critical illness was defined as patients with one of the following conditions: shortness of breath, respiratory rate ≥30 times/min; oxygen saturation ≤93% in a resting state; arterial oxygen partial pressure/oxygen inhalation concentration ≤300 mm Hg; pulmonary imaging showing that the lesions have significantly progressed >50% within 24-28 hours; any occurrence of respiratory failure and requiring mechanical ventilation; or shock, combined with other organ failure. Patients were excluded if they had any clinically significant psychosis or cognitive impairment; they were also excluded if they were unlikely to comply with or complete the study.

Interventions

The patients in the intervention group were instructed on the use of an intermittently scanned CGM (isCGM) device (Freestyle libre; Abbott Diabetes Care). The sensor was implanted in a site with sufficient subcutaneous fat on the back

XSL•FO
RenderX

of the upper arm, and this site can remain flat (without bends or wrinkles) during the participants' activities. Scars, moles, obesity lines, or lumps were avoided. One sensor lasted no longer than 14 days. If the sensor was expired or malfunctioned, a new one would be immediately implanted until discharge or death. Data were extracted and processed by using a scanner (by swipes) and auxiliary software. The frequency of the swipes followed the routine care during ICU stays at 4 hours intervals or whenever necessary. Data extracted from the CGM device included the mean glucose levels, coefficient of variance (CV), and times of severe hypoglycemia and hypoglycemia (<2.8 and <3.9 mmol/L, respectively) every day. The patients in the control group received a routine fingertip measurement (point-of-care testing [POCT]) for glucose levels. The frequency was consistent with the CGM swipes, ranging from every 1 hour to 4 hours when necessary. If a patient encountered hypoglycemia, more frequent testing of swipes and POCT were done. The mean glucose levels and CV were calculated based on the fingertip glucose profiles.

Sample Size Calculations

According to a previous case-control study [11], the composite adverse COVID-19 outcome (defined as progression to critical illness or death) occurred in 46% of patients in the isCGM group compared with 66% in the POCT group. Therefore, this study can achieve 80% statistical power at the level of α=.05 by including 50 patients in each group to achieve the superiority end point.

Randomization, Treatment, and Follow-Up During ICU stay

The participants were randomly assigned to the intervention and control groups in a 1:1 ratio at the time of admission to the ICU. The randomization sequence was generated by the staff. No blinding was applied during the study period. Anthropometric and clinical data were collected from the patients' self-reporting or the electronic medical records at baseline and at the end of their ICU stay or death after the intervention. These data included sex, height, weight, laboratory measurements, and history of diseases and medications. BMI was calculated as the weight in kilograms divided by the squared height in meters. The treatments of COVID-19 among critically ill patients were prescribed according to the diagnosis and treatment guideline for COVID-19 (ninth version) in China. For all patients received intravenous glucocorticoids, insulin was initiated when the glucose level was persistently higher than 13.9 mmol/L. The average daily dosages of glucocorticoids (converted to the dosages of prednisone) and insulin were also collected.

Outcomes

The primary outcome was 28-day mortality in the ICU. The secondary outcome included the length of stay in the ICU, as well as the occurrence of hypoglycemia and severe hypoglycemia events (<3.9 and <2.8 mmol/L, respectively).

Statistical Analysis

The analysis was performed according to the intention-to-treat principle. The data are reported as the mean (SD) unless otherwise stated. One‐way ANOVA was used to analyze differences in population characteristics from baseline to the end of study. A multivariate generalized estimating equation model adjusted for sex and age was used to evaluate influential factors on in-hospital mortality. The Kaplan-Meier survival analysis was used to compare the survival rate between patients in the isCGM and POCT groups. All data were analyzed using R (version 4.0.3; R Foundation for Statistical Computing). A 2‐tailed P<.05 was defined as significant.

Results

Baseline Characteristics of the Study Participants

Of the 130 participants who underwent screening in this study, 127 (97.6%) were enrolled and randomly assigned. In all, 64 were in the isCGM group while 63 were in the POCT group. Of these 127 participants, a total of 124 (97.6%) completed the study. A total of 2 participants from the isCGM group and 1 participant from the POCT group discontinued the study because they lost the sensor during the study period (Figure 1).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow chart of study. A total of 130 participants were screened for eligibility, and 127 were randomly assigned. All 127 participants completed the study. However, 3 partcipants were not included for analysis due to losing the sensor (2 in the isCGM group and 1 in the POCT group). Thus, the final analysis included 124 participants. isCGM: intermittently scanned continuous glucose monitoring; POCT: point-of-care testing.

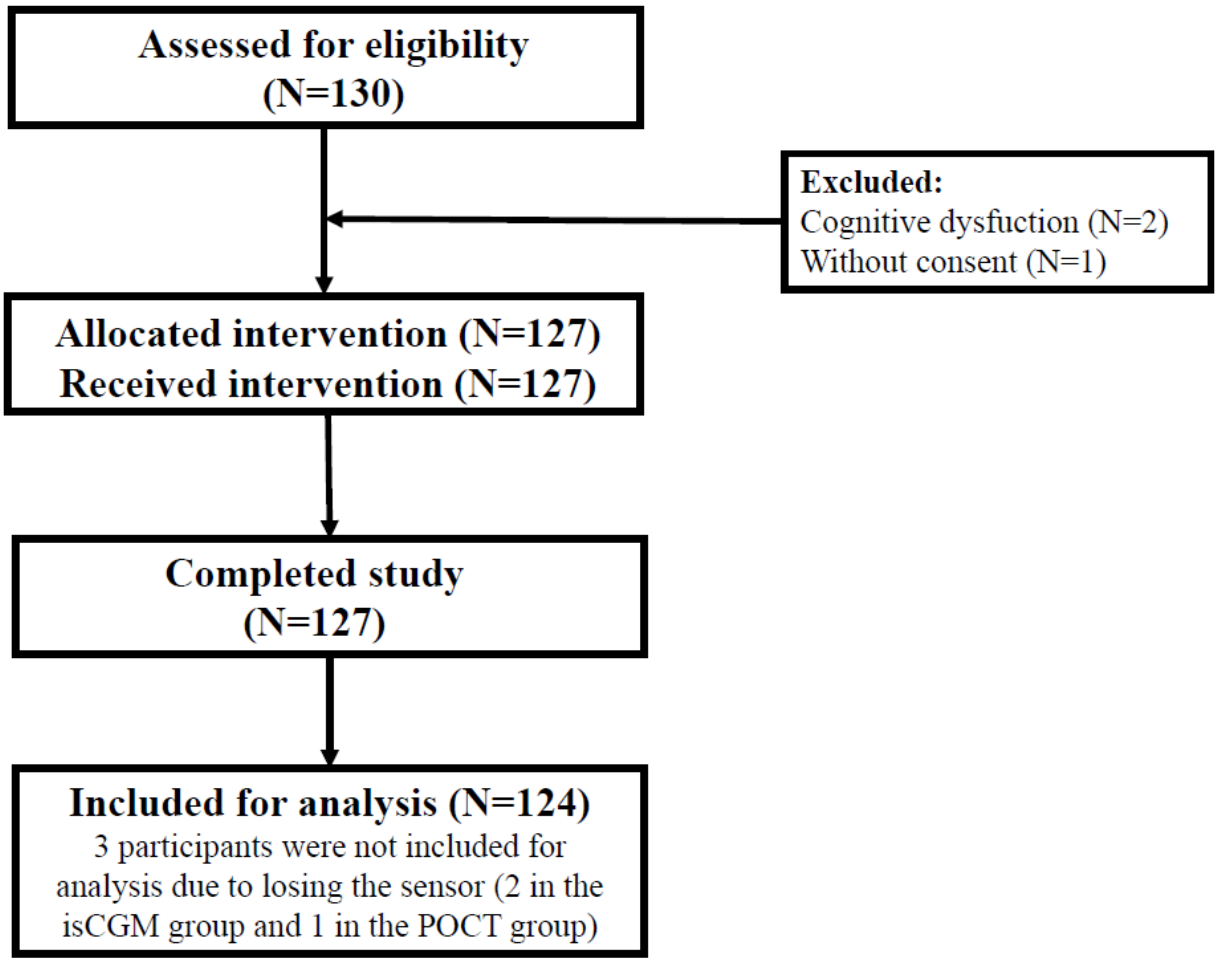

The baseline characteristics are shown in Table 1. The mean age was 78.3 (SD 11.5) years. The mean fasting glucose level at baseline was 8.12 (SD 1.54) mmol/L. The mean hemoglobin A_{1c} level was 7.2% (SD 0.8%). The percentage of participants with diabetes was 30.6% (38/124). The baseline characteristics were well balanced and comparable at baseline.

XSL•FO
RenderX

Table 1. Baseline characteristics

Characteristics	isCGM[a] group (n=62)	POCT[b] group (n=62)
Age (years), mean (SD)	78.8 (11.0)	78.7 (11.9)
Sex, n (%)		
Male	31 (50)	31 (50)
Female	31 (50)	31 (50)
BMI (kg/m^2), mean (SD)	22.2 (3.21)	22.3 (2.93)
Current smoker, n (%)	6 (10)	7 (11)
Heart rate (bpm[c]), mean (SD)	91.4 (19.0)	86.1 (14.1)
Respiratory rate (bpm), mean (SD)	28.0 (3.5)	28.5 (3.7)
Systolic blood pressure (mm Hg), mean (SD)	135.5 (25.0)	135.6 (26.3)
Diastolic blood pressure (mm Hg), mean (SD)	76.7 (14.5)	75.0 (13.3)
Comorbidities on admission, n (%)		
Diabetes	17 (27)	21 (34)
Hypertension	36 (58)	37 (60)
Dyslipidemia	4 (6)	5 (8)
Coronary heart disease	30 (48)	30 (48)
Stroke	35 (56)	38 (61)
COPD[d]	35 (56)	34 (55)
Chronic kidney disease	16 (26)	21 (34)
Cancer	10 (16)	11 (18)
Laboratory measurements, mean (SD)		
FPG[e] (mmol/L)	8.10 (1.35)	8.13 (1.67)
Hemoglobin A_{1c} (%)	8.3 (1.5)	8.2 (1.4)
ALT[f] (U/L)	63.0 (7.33)	67.4 (6.22)
AST[g] (U/L)	117 (16.0)	203 (14.7)
Triglyceride (mmol/L)	0.97 (0.56)	1.01 (0.51)
LDL[h] cholesterol (mmol/L)	3.23 (0.72)	3.19 (0.87)
HDL[i] cholesterol (mmol/L)	0.78 (0.14)	0.82 (0.34)
eGFR[j] (mL/min/1.73m^2)	60.2 (17.7)	61.0 (18.3)
Uric acid (mmol/L)	474.9 (87.3)	480.2 (87.2)
C-reactive protein (mmol/L)	62.1 (3.49)	67.6 (5.72)
Brain natriuretic peptide (mmol/L)	184.6 (103.7)	150.7 (62.6)
Interleukin-6 (pg/mL)	24.3 (4.19)	27.7 (8.37)
Use of glucocorticoid, n (%)	53 (86)	54 (87)
Daily dose on glucocorticoid[k], mean (SD)	15.0 (5.23)	15.1 (4.89)
Daily dose on insulin, mean (SD)	15.1 (3.15)	15.4 (2.89)

[a]isCGM: intermittently scanned continuous glucose monitoring.

[b]POCT: point-of-care testing.

[c]bpm: beats per minute.

[d]COPD: chronic obstructive pulmonary disease.

[e]FPG: fasting plasma glucose.

[f]ALT: alanine transaminase.

XSL•FO
RenderX

[g]AST: aspartate aminotransferase.

[h]LDL: low-density lipoprotein.

[i]HDL: high-density lipoprotein.

[j]eGFR: estimated glomerular filtration rate.

[k]Reported as the dosage of prednisone.

The Changes in Fasting Glucose Levels and Other Glycemic Parameters at the End of the Trial

There was no difference in fasting glucose levels observed between the isCGM and the POCT groups at baseline (mean 8.10 SD 1.35 vs 8.13, SD 1.67; *P*=.43). When compared with the baseline value of fasting glucose, both groups showed significant decreases at the end of the study period (isCGM: mean 8.10, SD 1.35 vs 6.07, SD 0.63; *P*=.001; and POCT: mean 8.13, SD 1.67 vs 7.76, SD 0.62; *P*=.04). At the end of study period, the mean value of fasting glucose in the isCGM group and the POCT group was 6.07 (SD 0.63) mmol/L and 7.76 (SD 0.62) mmol/L, respectively, showing a significant intergroup difference (Table 2; *P*=.01). The daily mean glucose levels (6.93, SD 0.72 vs 8.43, SD 0.52 mmol/L; *P*=.02) and the CV (25.2%, SD 1.23% vs 33.7%, SD 1.15%; *P*=.01) were also found to be significantly lower in the isCGM group than those in the POCT group at the end of the trial. Multivariate generalized linear model defined the reduction of fasting glucose levels as a dependent variable and the control group as the reference. After adjustment of age, sex, and BMI, the results showed that at the end of the study period, the intervention (isCGM) was associated with a larger reduction of fasting glucose levels (–1.69, 95% CI –1.65 to –1.73; *P*<.001).

Table 2. Comparisons of glycemic parameters at the end of the trial.

	isCGM[a] group (n=62), mean (SD)	POCT[b] group (n=62), mean (SD)	Relative risk or difference (95% CI)	*P* value
Fasting glucose (mmol/L)	6.07 (0.63)	7.76 (0.62)	–1.69 (–1.65 to –1.73)	.01
Daily mean glucose (mmol/L)	6.93 (0.72)	8.43 (0.52)	–1.50 (–1.47 to –1.53)	.02
CV[c] (%)	25.2 (1.23)	33.7 (1.15)	0.75 (0.71 to 0.79)	.01

[a]isCGM: intermittently scanned continuous glucose monitoring.

[b]POCT: point-of-care testing.

[c]CV: coefficient of variance.

Mortality in the ICU

A total of 15 death events occurred during the study period, with 2 in the isCGM group and 13 in the POCT group. The corresponding hazard ratio of mortality in the ICU in the isCGM group when compared with the POCT group was 0.18 (95% CI 0.04-0.79; Figure 2). The average length of ICU stay was 10.0 (SD 7.57) days in the isCGM group and 14.0 (SD 6.86) days in the POCT group, with a significant *P* value for comparison (*P*=.02).

XSL•FO
RenderX

Figure 2. The cumulative survival rate between the POCT and isCGM groups ($P<.05$). The corresponding hazard ratio of mortality in the ICU in the isCGM group when compared with the POCT group was 0.18 (95% CI 0.04-0.79). ICU: intensive care unit; isCGM: intermittently scanned continuous glucose monitoring; POCT: point-of-care testing.

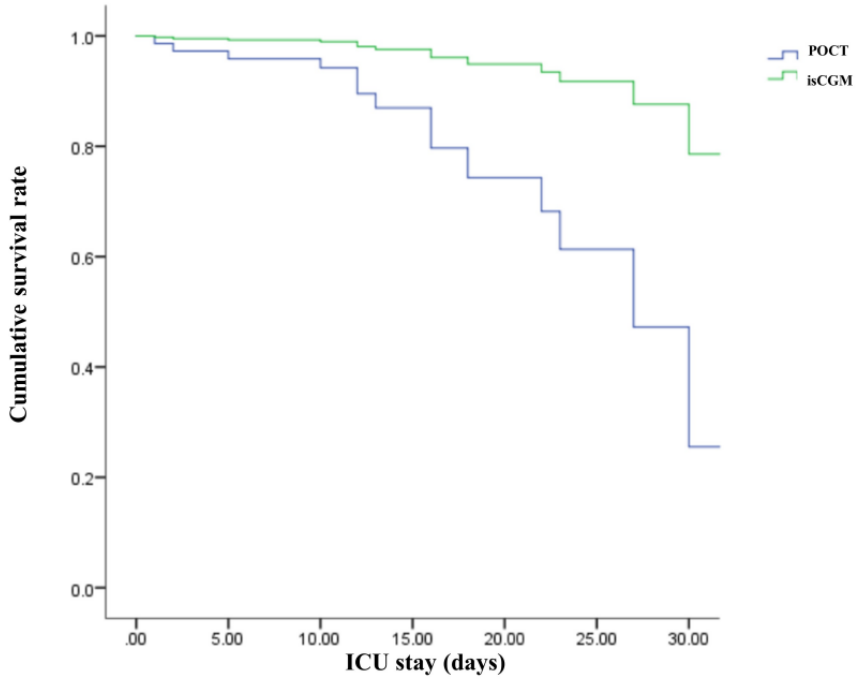

Hypoglycemia Events

A total of 207 hypoglycemia events (<3.9 mmol/L) was detected, with 43 in the isCGM group (0.0006 times per patient per day) and 164 in the POCT group (0.002 times per patient per day; $P<.001$). A total of 81 severe hypoglycemia events (<2.8 mmol/L) was detected, with 16 in the isCGM group (0.0002 times per patient per day) and 65 in the POCT group (0.0008 times per patient per day; $P<.001$).

Adverse Event

The major adverse event in this study was bleeding in the puncture site, with a total of 6 occurrences in the isCGM group. During the follow-up, none of the participants dropped out because of bleeding in the puncture site.

Discussion

In this exploratory, prospective, open-label, parallel, single-center RCT, we observed a significant improvement on the prognosis of frail and critically ill patients with COVID-19, with fewer in-hospital death events and hypoglycemia events. These findings suggest that the use of isCGM during ICU stay under the COVID-19 pandemic was a protective way to help improve the survival rate among the extremely frail patients.

To date, this might be the first RCT on the use of CGM in the ICU setting among patients with COVID-19, although it was a single-center study with limited sample size. Evidence on the in-hospital use of CGM, especially in the ICU, is still limited. Most studies are observational with a diverse population [12]. While CGM is an efficient tool for patients with diabetes, it has extended its use to other specific populations without diabetes within in-hospital settings such as in the ICU [13], during the perioperative period [14], during pregnancy or labor [15], and among frail populations [16]. The continuous efforts are made by the manufacturers of CGM toward more accuracy, painlessness, and being less burdensome. Early in 2020, the FDA announced the feasibility of using CGM during an in-hospital stay. CGM is essential for the care of glycemic management and can reduce the risk of cross-infections and nosocomial infections among health care professionals.

For the use of CGM in the ICU, several observational studies reported the feasibility and accuracy of different CGM systems. We previously reported the suboptimal overall accuracy of an isCGM system (Freestyle Libre) for critically ill patients. The mean absolute relative difference was reported to be 18% [13]. These data allowed us to further investigate the effectiveness of CGM on ICU-related outcomes among frail patients with COVID-19. We found a significant difference in survival between patients with isCGM and those without isCGM. In another observational study at the Montefiore Medical Center, real-time CGM was used on ICU patients with confirmed COVID-19 infection and glycemic variability [17]. The study demonstrated early feasibility, considerable accuracy, and meaningful reduction in the frequency of POCT. One US study also tested the feasibility among patients receiving operations and those who were admitted to the cardiac ICU, and they provided encouraging results with sustained accuracy of CGM even during exposure to vasopressors [18].

This study aims to fill the gap of evidence on the use of CGM in hospital settings. Less evidence is provided from RCTs of CGM under in-hospital settings. Most of the RCTs focused on the glycemic benefit provided by the use of CGM. Two studies found that the use of the Dexcom senor when compared to either POCT or a blinded CGM device could definitely improve the

XSL•FO
RenderX

glycemic control by increasing time in target range and reducing time below or above target ranges [19,20]. In this exploratory, prospective, open-label, parallel, single-center RCT, we first reported the huge benefit of CGM on ICU mortality as well as the length of ICU stay among frail and critically ill patients with COVID-19. The use of CGM also provided a timely chance for the clinicians to observe the hyperglycemia and hypoglycemia events and treat them [20]. There was also a significant difference in the hypoglycemia events between the CGM group and the POCT group. This study supported the hypothesis that CGM might help with the improvement of glycemic control and the outcomes beyond glucose levels. More trials or longitudinal studies with long-term follow-ups are still needed to demonstrate the effect of CGM on disease outcomes other than glycemic control.

The major limitation of this study is the small sample size within a single center. In addition, selection bias might also be introduced because we used a specific population of patients with frailty and critical illnesses, and those patients tended to be older. Whether the findings can be translated to other populations needs further validation in a multicenter design and with a large sample. The strength of this study is that it is the first to demonstrate the association of CGM use and adverse ICU outcomes in frail and critically ill patients with COVID-19. The RCT design also strengthens the level of evidence.

In conclusion, we found a significant clinical benefit from the use of CGM among frail and critically ill patients with COVID-19. These findings support the use of CGM in the ICU and might help with the extension of application in various in-hospital settings.

Acknowledgments

We really appreciate and respect the efforts by the doctors, nurses, and patients during the COVID-19 pandemic.

This work was supported by Shanghai Municipal Commission of Science and Technology (21Y11902800), Shanghai Shenkang Hospital Development Center (SHDC12020122), and Shanghai Sixth People's Hospital (ynxg202219).

Data Availability

The original data of this study are not publicly available but are available from the corresponding author upon reasonable request. Access to the data will be granted to researchers who provide a clear and justified purpose for their request, in alignment with ethical guidelines and data protection regulations.

Authors' Contributions

YZ and WW designed the study and had equal contribution to the manuscript. JS and ZY wrote the manuscript. ZZ oversaw the nursing care of the patients. QZ, YZ, and WW reviewed and edited the manuscript. YZ and WW are the guarantors of this work and, as such, have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors read and approved the final manuscript. JS and ZY contributed equally to this work.

Conflicts of Interest

The intermittently scanned continuous glucose monitoring was supported by Abbott Diabetes Care, Inc. All authors have no competing interests to disclose.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 82 KB - jmir_v27i1e67012_app1.pdf]

References

1. Danne T, Nimri R, Battelino T, Bergenstal RM, Close KL, DeVries JH, et al. International Consensus on Use of Continuous Glucose Monitoring. Diabetes Care 2017 Dec;40(12):1631-1640 [FREE Full text] [doi: 10.2337/dc17-1600] [Medline: 29162583]
2. Shilo S, Keshet A, Rossman H, Godneva A, Talmor-Barkan Y, Aviv Y, et al. Continuous glucose monitoring and intrapersonal variability in fasting glucose. Nat Med 2024 May 08;30(5):1424-1431. [doi: 10.1038/s41591-024-02908-9] [Medline: 38589602]
3. Galindo RJ, Aleppo G, Klonoff DC, Spanakis EK, Agarwal S, Vellanki P, et al. Implementation of continuous glucose monitoring in the hospital: emergent considerations for remote glucose monitoring during the COVID-19 pandemic. J Diabetes Sci Technol 2020 Jul 14;14(4):822-832 [FREE Full text] [doi: 10.1177/1932296820932903] [Medline: 32536205]
4. Battelino T, Alexander C, Amiel S, Arreaza-Rubin G, Beck RW, Bergenstal RM, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023 Jan;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9] [Medline: 36493795]

XSL•FO
RenderX

5. Bao Y, Chen L, Chen L, Dou J, Gao Z, Gao L, et al. Chinese clinical guidelines for continuous glucose monitoring (2018 edition). Diabetes Metab Res Rev 2019 Sep 25;35(6):e3152. [doi: 10.1002/dmrr.3152] [Medline: 30884108]
6. Pasquel FJ, Umpierrez GE. Individualizing inpatient diabetes management during the coronavirus disease 2019 pandemic. J Diabetes Sci Technol 2020 Jul 05;14(4):705-707 [FREE Full text] [doi: 10.1177/1932296820923045] [Medline: 32370606]
7. Shaw J, Bannuru R, Beach L, ElSayed NA, Freckmann G, Füzéry AK, et al. Consensus considerations and good practice points for use of continuous glucose monitoring systems in hospital settings. Diabetes Care 2024 Dec 01;47(12):2062-2075. [doi: 10.2337/dci24-0073] [Medline: 39452893]
8. Kodani N, Ohsugi M. The patient-centered diabetes management during the COVID-19 pandemic. Glob Health Med 2022 Aug 31;4(4):210-215 [FREE Full text] [doi: 10.35772/ghm.2022.01031] [Medline: 36119784]
9. Longo RR, Elias H, Khan M, Seley JJ. Use and accuracy of inpatient CGM during the COVID-19 pandemic: an observational study of general medicine and ICU patients. J Diabetes Sci Technol 2022 Sep 10;16(5):1136-1143 [FREE Full text] [doi: 10.1177/19322968211008446] [Medline: 33971753]
10. Spierling Bagsic SR, Fortmann AL, Belasco R, Bastian A, Lohnes S, Ritko A, et al. Real-time continuous glucose monitoring in the hospital: a real-world experience. J Diabetes Sci Technol 2023 May 13;17(3):656-666 [FREE Full text] [doi: 10.1177/19322968231165982] [Medline: 37056168]
11. Shen Y, Fan X, Zhang L, Wang Y, Li C, Lu J, et al. Thresholds of glycemia and the outcomes of COVID-19 complicated with diabetes: a retrospective exploratory study using continuous glucose monitoring. Diabetes Care 2021 Apr;44(4):976-982 [FREE Full text] [doi: 10.2337/dc20-1448] [Medline: 33574126]
12. Zelada H, Perez-Guzman M, Chernavvsky D, Galindo R. Continuous glucose monitoring for inpatient diabetes management: an update on current evidence and practice. Endocr Connect 2023 Sep 25;12(10):e230180 [FREE Full text] [doi: 10.1530/EC-23-0180] [Medline: 37578799]
13. Huang W, Li S, Lu J, Shen Y, Wang Y, Wang Y, et al. Accuracy of the intermittently scanned continuous glucose monitoring system in critically ill patients: a prospective, multicenter, observational study. Endocrine 2022 Dec 13;78(3):470-475 [FREE Full text] [doi: 10.1007/s12020-022-03216-3] [Medline: 36227509]
14. Herzig D, Vettoretti M, Guensch D, Melmer A, Schürch D, Roos J, et al. Performance of the Dexcom G6 continuous glucose monitoring system during cardiac surgery using hypothermic extracorporeal circulation. Diabetes Care 2023 Apr 01;46(4):864-867. [doi: 10.2337/dc22-2260] [Medline: 36809308]
15. Sanusi A, Xue Y, McIlwraith C, Howard H, Brocato BE, Casey B, et al. Association of continuous glucose monitoring metrics with pregnancy outcomes in patients with preexisting diabetes. Diabetes Care 2024 Jan 01;47(1):89-96. [doi: 10.2337/dc23-0636] [Medline: 37782847]
16. Idrees T, Zabala Z, Moreno E, Gerges A, Urrutia M, Ruiz J, et al. The effects of aging and frailty on inpatient glycemic control by continuous glucose monitoring in patients with type 2 diabetes. Diabetes Res Clin Pract 2023 Apr;198:110603. [doi: 10.1016/j.diabres.2023.110603] [Medline: 36871877]
17. Agarwal S, Mathew J, Davis G, Shephardson A, Levine A, Louard R, et al. Continuous glucose monitoring in the intensive care unit during the COVID-19 pandemic. Diabetes Care 2021 Mar;44(3):847-849 [FREE Full text] [doi: 10.2337/dc20-2219] [Medline: 33361145]
18. Perez-Guzman M, Duggan E, Gibanica S, Cardona S, Corujo-Rodriguez A, Faloye A, et al. Continuous glucose monitoring in the operating room and cardiac intensive care unit. Diabetes Care 2021 Mar;44(3):e50-e52 [FREE Full text] [doi: 10.2337/dc20-2386] [Medline: 33479159]
19. Fortmann A, Spierling Bagsic SR, Talavera L, Garcia IM, Sandoval H, Hottinger A, et al. Glucose as the fifth vital sign: a randomized controlled trial of continuous glucose monitoring in a non-ICU hospital setting. Diabetes Care 2020 Nov;43(11):2873-2877 [FREE Full text] [doi: 10.2337/dc20-1016] [Medline: 32855160]
20. Singh L, Satyarengga M, Marcano I, Scott WH, Pinault LF, Feng Z, et al. Reducing inpatient hypoglycemia in the general wards using real-time continuous glucose monitoring: the glucose telemetry system, a randomized clinical trial. Diabetes Care 2020 Nov;43(11):2736-2743 [FREE Full text] [doi: 10.2337/dc20-0840] [Medline: 32759361]

Abbreviations

CGM: continuous glucose monitoring
CV: coefficient of variance
FDA: Food and Drug Administration
ICU: intensive care unit
isCGM: intermittently scanned continuous glucose monitoring
POCT: point-of-care testing
RCT: randomized controlled trial
RT-PCR: reverse transcription polymerase chain reaction

XSL•FO
RenderX

Edited by A Mavragani; submitted 29.09.24; peer-reviewed by SS Kollala, N Misra; comments to author 05.12.24; revised version received 20.12.24; accepted 09.01.25; published 07.02.25.

Please cite as:
Shang J, Yuan Z, Zhang Z, Zhou Q, Zou Y, Wang W
Effectiveness of Continuous Glucose Monitoring on Short-Term, In-Hospital Mortality Among Frail and Critically Ill Patients With COVID-19: Randomized Controlled Trial
J Med Internet Res 2025;27:e67012
URL: https://www.jmir.org/2025/1/e67012
doi:10.2196/67012
PMID:39918851

©Jiawei Shang, Ziming Yuan, Zuoyan Zhang, Quanhong Zhou, Yan Zou, Wei Wang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Leveraging Large Language Models for Infectious Disease Surveillance—Using a Web Service for Monitoring COVID-19 Patterns From Self-Reporting Tweets: Content Analysis

Jiacheng Xie[1,2], PhD; Ziyang Zhang[1,2], MS; Shuai Zeng[1,2], PhD; Joel Hilliard[1], BSc; Guanghui An[2,3], MD, PhD; Xiaoting Tang[2,4], MD, PhD; Lei Jiang[1,2], PhD; Yang Yu[1,2], PhD; Xiufeng Wan[1,2,5,6], PhD; Dong Xu[1,2], PhD

[1]Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States

[2]Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, United States

[3]School of Acupuncture-Moxibustion and Tuina, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[4]Shanghai Pudong New Area Wanggang Community Health Service Center, Shanghai, China

[5]NextGen Center for Influenza and Emerging Infectious Diseases, University of Missouri, Columbia, United States

[6]Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, United States

Corresponding Author:
Dong Xu, PhD
Department of Electrical Engineering and Computer Science
University of Missouri
227 Naka Hall
Columbia, MO, 65211
United States
Phone: 1 5738822299
Email: xudong@missouri.edu

Abstract

Background: The emergence of new SARS-CoV-2 variants, the resulting reinfections, and post–COVID-19 condition continue to impact many people's lives. Tracking websites like the one at Johns Hopkins University no longer report the daily confirmed cases, posing challenges to accurately determine the true extent of infections. Many COVID-19 cases with mild symptoms are self-assessed at home and reported on social media, which provides an opportunity to monitor and understand the progression and evolving trends of the disease.

Objective: We aim to build a publicly available database of COVID-19–related tweets and extracted information about symptoms and recovery cycles from self-reported tweets. We have presented the results of our analysis of infection, reinfection, recovery, and long-term effects of COVID-19 on a visualization website that refreshes data on a weekly basis.

Methods: We used Twitter (subsequently rebranded as X) to collect COVID-19–related data, from which 9 native English-speaking annotators annotated a training dataset of COVID-19–positive self-reporters. We then used large language models to identify positive self-reporters from other unannotated tweets. We used the Hibert transform to calculate the lead of the prediction curve ahead of the reported curve. Finally, we presented our findings on symptoms, recovery, reinfections, and long-term effects of COVID-19 on the Covlab website.

Results: We collected 7.3 million tweets related to COVID-19 between January 1, 2020, and April 1, 2024, including 262,278 self-reported cases. The predicted number of infection cases by our model is 7.63 days ahead of the official report. In addition to common symptoms, we identified some symptoms that were not included in the list from the US Centers for Disease Control and Prevention, such as lethargy and hallucinations. Repeat infections were commonly occurring, with rates of second and third infections at 7.49% (19,644/262,278) and 1.37% (3593/262,278), respectively, whereas 0.45% (1180/262,278) also reported that they had been infected >5 times. We identified 723 individuals who shared detailed recovery experiences through tweets, indicating a substantially reduction in recovery time over the years. Specifically, the average recovery period decreased from around 30 days in 2020 to approximately 12 days in 2023. In addition, geographic information collected from confirmed individuals indicates that the temporal patterns of confirmed cases in states such as California and Texas closely mirror the overall trajectory observed across the United States.

XSL•FO
RenderX

Conclusions: Although with some biases and limitations, self-reported tweet data serves as a valuable complement to clinical data, especially in the postpandemic era dominated by mild cases. Our web-based analytic platform can play a significant role in continuously tracking COVID-19, finding new uncommon symptoms, detecting and monitoring the manifestation of long-term effects, and providing necessary insights to the public and decision-makers.

(J Med Internet Res 2025;27:e63190) doi:10.2196/63190

KEYWORDS

COVID-19; self-reporting data; large language model; Twitter; social media analysis; natural language processing; machine learning

Introduction

Background

COVID-19 is one of the most severe infectious diseases in human history. Although the World Health Organization downgraded the COVID-19 pandemic, declaring it is no longer a global emergency on May 5, 2023 [1], the threat of infection and death remains. As of August 1, 2023, there have been >300,000 confirmed cases weekly worldwide, resulting in >1000 deaths per week; however, major information-publishing platforms, such as that at Johns Hopkins University, stopped collecting and tracking COVID-19 data worldwide on March 10, 2023 [2]. Therefore, it has become more challenging to identify the actual number and trends of COVID-19 infections daily. Traditional public health monitoring methods face several challenges, including delays in clinical data collection, lack of real-time insights, and the underrepresentation of population-level trends, particularly in regions with limited health care reporting infrastructure [3,4]. New tools are required to provide timely awareness and detection of COVID-19 transmission trends, reinfection patterns, and the long-term health impact of the disease.

To supplement the shortage of clinical data and gain further insights into the development trends and variant tendencies of COVID-19, researchers have turned to social media, specifically Twitter (subsequently rebranded as X). Social media data offer unique advantages, such as rapid updates and a broad geographical reach, which traditional clinical data often lack [5,6]. Several studies [7-11] have explored the use of social media for health monitoring, including sentiment analysis of COVID-19–related tweets, the identification of emerging symptoms, and the study of vaccine hesitancy. Early studies [12-16] primarily focused on constructing COVID-19–related tweet databases. However, these works do not provide in-depth analysis of self-reported tweets, and the databases tended to be collected over a short time frame, typically several months, and cannot automatically update. Later, some research endeavors [9,17-19] shifted their focus toward studying COVID-19 symptoms on the basis of tweets and reported the distribution of symptoms in tweets. However, these studies included limited numbers of collected self-reported cases. Some studies [16,20-24] used tweets to study geographic distribution but did not provide a corresponding time-series analysis or predict the spread of COVID-19. Moreover, existing studies have not fully leveraged user-generated content on social media to provide a comprehensive, dynamic view of COVID-19 trends across symptoms, geography, and time. This gap underscores the need for a methodology capable of providing near–real-time insights and broader geographical coverage [25]. As for the visualization tools for COVID-19, some researchers [26-29] have developed platforms or dashboards to study the trends of COVID-19, but most were based on clinical data. A few tweet-based platforms [27,29,30] showed limited information and failed to provide timely updates.

Reinfection often refers to the phenomenon in which an individual who has recovered from COVID-19 is again infected with the virus [31]. Some researchers [32,33] considered that reinfection is identified when an individual tests positive again through polymerase chain reaction (PCR) testing after a minimum of 90 days of a negative result. However, some studies also suggest this duration should be 30 days [34,35]. Other studies [34,36,37] considered reinfection to be a new positive PCR following 2 consecutive negative PCR tests taken after primary infection. Moreover, the periodicity of reinfection, the reinfection rate, and the maximum number of infections are uncertain. Existing studies on reinfection trends and periodicity have primarily relied on clinical data, leaving gaps in understanding population-level reinfection dynamics that may be observable through self-reported data on social media [38]. In addition to reinfection, there is a growing concern about the long-term effects of COVID-19, known as post–COVID19 condition, where patients report symptoms lasting months after recovery [39,40]. Clinical data on long-term effects remain limited [41], but social media offers a platform where individuals can share ongoing symptoms, providing valuable insights for public health research [42,43]. This further highlights the potential of using self-reported data to investigate the periodicity of reinfection, as well as long-term effects and symptoms, complementing traditional clinical datasets.

Health natural language processing (NLP) is gaining increasing attention for its essential role in both methodology development and applications [44]. The technology has been widely applied in areas such as information extraction from electronic health records [45], adverse drug reaction analysis [46], clinical decision support [47], and hospitalization prediction [48]. Using techniques such as topic modeling, sentiment analysis, and deep learning models such as Bidirectional Encoder Representations from Transformers (BERT) [49], NLP can extract valuable medical insights from unstructured text. Combined with real-time analytics frameworks [50] and knowledge graphs [51], these technologies can dynamically monitor COVID-19 trends, identify previously unnoticed symptoms and long-term effects, and provide scientific support for optimizing public health strategies and resource allocation.

XSL•FO
RenderX

Objectives

In this study, we address these gaps by leveraging self-reported COVID-19 data from Twitter to provide near–real-time insights into trends across symptoms, reinfection, recovery cycles, and geographical distributions. The proposed approach not only fills the limitations of traditional monitoring systems but also offers a scalable, timely, and comprehensive method to track the pandemic's evolving dynamics. Our visualization tool updates weekly and comprehensively analyzes information related to COVID-19 symptoms, case distribution, reinfections, recoveries, and long-term effects based on large language models (LLMs). Figure 1 depicts our research objectives: (1) we aim to build a publicly available database of >9,836,206 COVID-19–related tweets, and this database is set to automatically update with newly collected data weekly; (2) LLMs will be built to automatically filter the tweets of self-reporters and extract their mentioned symptoms and recovery cycles; and (3) we aim to build a visualization website that refreshes data on a weekly basis, Covlab [52], to track and analyze infection, reinfection, recovery, and long-term effects of COVID-19.

As depicted in Figure 1, the workflow of Covlab comprised the following steps: data collection, models training, tracking, and visualization website. During data collection, COVID-19–related tweets were collected using the Twitter application programming interface (API) and filtered from a broader dataset of social media posts. A subset of these tweets was manually labeled through a specialized annotation tool, creating a high-quality training set essential for developing machine learning models. In addition, historical tweets and user ID databases were incorporated to construct a comprehensive cohort of individuals who self-reported COVID-19 infections, forming the foundational dataset for subsequent tracking and analytical tasks. In models training step, annotated datasets were used to train a variety of machine learning and NLP models, including support vector machine (SVM), logistic regression (LR), BERT, GPT-2, and LLM meta AI 2 (Llama-2), to identify the most effective model for detecting self-reported COVID-19 infection tweets. After selecting the optimal model, it was applied to a larger collection of COVID-19–related tweets to extract tweets reporting personal COVID-19 infections. Throughout this process, performance metrics such as cross-validation and receiver operating characteristic (ROC) curve analysis, were used to ensure the models provided robust and accurate predictions. In the tracking step, the system conducted long-term tracking of individuals who self-reported COVID-19 infections in their tweets, focusing on analyzing symptoms such as breathing difficulties, fever, headache, cough, and fatigue, as well as recovery cycles, long-term effects (ie, sequelae), and geographical trends. In addition, reinfection cases were monitored to provide insights into the temporal dynamics of COVID-19 experiences, helping to uncover patterns in symptom progression, recovery, and recurring infections over time. In the visualization website step, the results of the analyses were presented on an interactive platform, Covlab, which provided a variety of visualization tools, including word clouds, pie charts, box plots, line graphs, geographic maps, and trend tables. This platform enabled users to track epidemic trends, uncover new or previously unreported symptoms, analyze recovery durations, report reinfection statistics, and explore different types of long-term effects (ie, sequelae), offering valuable insights to researchers and public health officials.

Figure 1. Workflow for Covlab. BERT: Bidirectional Encoder Representations from Transformers; Llama-2: large language model meta AI 2; LR: logistic regression; NLP: natural language processing; ROC: receiver operating characteristics; SVM: support vector machine.

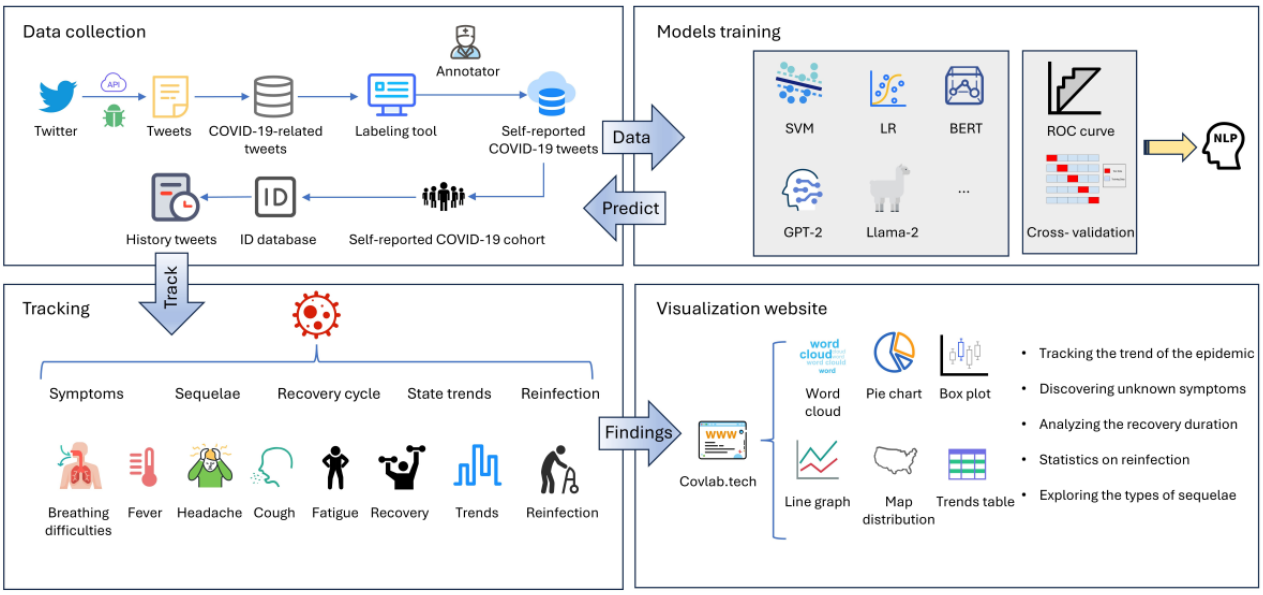

XSL•FO
RenderX

Methods

Data Collection and Preprocessing

We collected and processed tweet data from January 2020 to April 2024 based on COVID-19–related keywords and hashtags through the Twitter streaming API. The searching keywords included "I.* tested[ed] positive for [covid | coronavirus | covid19 | covid-19]," "My.* [covid | coronavirus | covid19 | covid-19].* symptoms," "#COVID" "#LongCovid" (all the keywords can be found in Multimedia Appendix 1). The following preprocessing operations were conducted on the tweets collected based on keywords. We first converted all words in the tweets to lowercase, standardized the tweets to American Standard Code for Information Interchange encoding using the Unicode library, and tokenized the tweets. Next, we removed all Unicode symbols and punctuation marks, some uninformative characters about usernames, such as @username, all digits and line breaks in the tweets, all URLs, such as http://, and all words contained in our stop word library (Multimedia Appendix 2). We converted emoji expressions into their corresponding textual expressions. To provide sufficient datasets for subsequent model training, we constructed a self-reported COVID-19 tweets dataset with manual labeling by 9 native English-speaking annotators after obtaining ethics approval. We also established a set of calibration criteria as shown in Table 1 to ensure the consistency and reliability of the tweet annotations. Two additional annotators conducted a secondary annotation on the labeled data. If their new annotations were inconsistent with the originals, all annotators would decide through a voting process whether the tweet should be classified as a self-reported COVID-19 positive tweet. The Fleiss κ index [53] was used to measure the consistency among multiple annotators to ensure the reliability of the annotations (Multimedia Appendix 3). We also developed a web-based annotation tool (Multimedia Appendix 4) to improve the efficiency of manual annotation. We annotated 115,214 tweets, of which 13,701 (11.89%) were positive samples that described self-reported positive COVID-19 cases and 101,513 (88.11%) were negative samples either not describing self-infection with COVID-19 or irrelevant. Multimedia Appendix 5 presents the types of tweets targeted in our study. The tweet depicted on the left serves as a prototypical instance, delineating self-reported information about COVID-19 infections. This includes the date of diagnosis and a detailed account of the symptoms experienced by the individual. Conversely, although the tweet shown on the right also references a diagnosis of COVID-19 and details associated symptoms, it diverges from our criteria for target tweets because it recounts a diagnosis pertaining to a third party, specifically a friend of the tweeter, rather than a self-reported account. Therefore, it falls outside the scope of our target dataset.

Table 1. Labeling criteria for the tweets.

Index	Annotation guideline	Description
1	Self-reported infection	Tweets must be a personal account by the author regarding their experience of contracting COVID-19. If the tweet mentions someone other than the author being infected, such as friends, family, or others, it should be labeled as a negative sample.
2	First-person narrative	Tweets should use first-person pronouns (eg, "I," "me," and "my") to describe the author's experience of contracting COVID-19, not that of others.
3	Concrete information	Tweets should provide concrete details, including infection timeline, test results and medical treatments, among others, rather than general discussions or speculations.
4	Symptom description	Tweets should contain the patient's personal descriptions of COVID-19 symptoms experienced, such as fever, cough, or difficulty breathing.
5	Confirmation information	Tweets should mention how the author was confirmed to have contracted COVID-19, such as the type of test conducted (PCR[a] and rapid antigen test), confirmation by a physician, or official institution validation.
6	Treatment experience	Tweets should describe the author's experience of treatment or recovery after self-contracting COVID-19, including isolation, medication, deterioration, or improvement in their health.
7	Infection timeline	Tweets should contain exact time points or time ranges of the infection rather than general discussions. Providing precise timing helps verify the author's infection period.
8	Test results	Tweets should reference the author's COVID-19 test results, such as testing positive or negative or other relevant test outcomes.
9	Medical Facility	Tweets should mention whether the author received treatment or underwent COVID-19-related tests at a medical facility, such as a hospital or clinic.
10	Social distancing measures	Tweets should discuss the author's adoption of social distancing measures due to their infection, such as self-isolation or notifying close contacts.
11	Substantial evidence	Tweets should contain substantial evidence, such as medical records, official notices, or other documents validating the author's COVID-19 infection.
12	Exclusion of transmission	Tweets emphasizing that the author did not transmit the virus to others may indicate a self-reported infection.
13	Consensus annotation	If 3 or more annotators provide inconsistent annotations for a particular tweet, a discussion among all annotators should be initiated to reach a final consensus.

[a]PCR: polymerase chain reaction.

XSL•FO
RenderX

Ethical Considerations

We have prioritized privacy and ethical considerations in our use of self-reported COVID-19–related tweets from Twitter, which, while publicly available, are processed with strict anonymization measures to ensure that no personally identifiable information can be traced back to individuals. The ethical nuances of using such data are addressed by considering users' implied consent when posting self-reports on a public platform. To safeguard data security throughout collection, annotation, and analysis, we have implemented encryption, controlled access, and other technical measures, ensuring that the data are strictly used for research purposes and never for commercial activities. Furthermore, recognizing the inherent noise and biases in social media data, we explicitly acknowledge the limitations of our dataset and analysis. To mitigate potential risks of misinterpretation or misuse, we emphasize that our findings are intended for population-level trend analysis rather than individual diagnosis, thereby maintaining the integrity and appropriate application of the results. This study is exempt from ethical approval because it only involves the analysis of aggregate data, ensuring that no individual privacy information is disclosed. All data have been fully deidentified, and appropriate anonymization measures have been applied to guarantee that no subject can be traced.

Self-Reported COVID-19–Positive Model

To select text-classification models to determine whether a tweet is a self-reported COVID-19– positive tweet, we trained both traditional machine learning methods and fine-tuned LLMs. We divided the manually labeled dataset in which 80% (92,171/115,214) was used as the training set, 10% (11,521/115,214) as the validation set, and the remaining 10% (11,521/115,214) as the test set. We used word frequency, term frequency–inverse document frequency [54] vectors, and feature hashing vectors as methods for text feature extraction, and we adopted 10-fold cross-validation to ensure the reliability of the results.

For conventional machine learning models, we used both linear and nonlinear models to compare different classification approaches. Because the amount of the manually annotated data were relatively small and the text data features were relatively simple, we experimented with machine learning methods such as naive Bayes [55], SVMs [56], and LR [57]. In the SVM method, we used the radial basis function as the kernel function and set the penalty parameter to 1.2. We used an L2 regularization as the penalty term in LR, with a regularization strength parameter (C) set to 1.0. As for LLMs, we used BERT [49], robustly optimized BERT pretraining approach (RoBERTa) [58], extreme language model [59], GPT-2 [60], Bigscience Large Open-science Open-access Multilingual language model (BLOOM) [61], and Llama-2 [62] for training. These LLMs are pretrained on large datasets using the masked language model objective for BERT and RoBERTa; the permuted language model objective for extreme language model; and the causal language modeling objective for GPT-2, BLOOM, and Llama-2. Because GPT-2, BLOOM, and Llama-2 are generative models, they may generate nonuniform results, rendering result interpretation difficult. These pretrained LLMs were originally designed to process general text of varying lengths. Through fine-tuning these LLMs on tweets, which are inherently short texts, we adapted them to better handle short text-classification tasks. We selected models pretrained on English corpora to ensure appropriate language understanding. We borrowed the idea from previous work [63], leveraging latent representations from LLMs for supervised label prediction. Hence, we designed different LLM-based classifiers integrated with various LLMs and fully connected neural networks. Precisely, each LLM serves as a backbone for encoding the tweets instead of generating text. A fully connected neural network was integrated on top of each LLM for stably accurate detection of self-reported COVID-19 cases. Unlike traditional machine learning methods requiring feature preprocessing, the LLM-based classifiers take only the text of a tweet as input. The specific parameters of the machine learning methods and LLMs can be found in Multimedia Appendices 6 and 7.

To prevent catastrophic forgetting and ensure that LLMs have a broad understanding of self-reported COVID-19 cases during the training stage, we used low-rank adaptation (LoRA) [64] to fine-tune the LLM-based classifiers. LoRA enables the parameters of a model to learn effectively by introducing trainable rank decomposition matrices into the transformer architectures in the LLMs. To achieve this reparameterization, we modified the projection matrices of query, key, value, and feedforward network modules within the self-attention mechanism of the transformer.

The LLM-based classifiers were trained end-to-end with the AdamW [65] optimizer with a cross-entropy objective function. During the training stage, the parameters introduced by LoRA within the pretrained model were updated with gradients, and all remaining parameters were frozen. Early stopping to monitor the accuracy of the validation dataset was implemented during training. All runs were trained on the Nvidia A100 graphical processing unit with a batch size of 5 for Llama-2 and 32 for other models.

We evaluated the performance of each model and chose the one that achieved the best results for predicting the daily number of self-reported confirmed cases. Subsequently, we applied a named entity recognition [66,67] method such as SpaCy [68] and BERT-based models to extract essential symptom-related keywords from the tweets. The model was trained or fine-tuned using a labeled dataset specifically curated for health-related text, which included annotations for COVID-19 symptoms such as fever, cough, fatigue, and loss of smell or taste. For the definition of symptoms, we referred to the descriptions of symptoms within the systematized nomenclature of medicine clinical terms [69] as shown in Table 2. To make our system operational, we deployed the trained model on a server. We used a script program to continuously collect COVID-19–related tweets from Twitter. The collected tweets were then analyzed using the deployed model, and the results were displayed weekly on the Covlab website. This provided users with up-to-date, analyzed information regarding COVID-19 development.

XSL•FO
RenderX

Table 2. Symptoms and their expression found in self-reported tweets.

Symptoms	Descriptions (with their IDs in clinical terms)
Fever	Febrile (386661006), fever (386661006), feverish (103001002), mill fever (85761009), hyperthermia (1197782006), hay fever (21719001), degrees Fahrenheit (258712004), temperature (722490005), high temperature (285717004), high body temperature (50177009), body temperature above reference range (50177009), increased body temperature (50177009), elevated temperature (50177009), raised temperature (50177009), increased skin temperature (17038008), feeling hot (373932008), feeling hot and cold (103002009), and feeling hot and sweaty (373939004)
Chills	Chills (43724002), chills and fever (274640006), chillness (43724002), shivering (43724002), shivering or rigors (248456009), rigor (38880002), brass founders' ague (74800004), algor (425681008), shakes (26079004), shaking (26079004), trembling (267079009), cold (82272006), head cold (82272006), freeze (48103003), freezing (48103003), and frigid (48103003)
Sweating	Sweat (74616000), sweating (415691001), cold sweat (83547004), hot sweat (224962007), hemopoiesis (445961003), hidrosis (415691001), diaphoresis (52613005), perspiration (415691001), perspire (74616000), perspire profusely (74616000), started to perspire (74616000), perspire all over (74616000), perspire during sleep (74616000), excessive sweating (52613005), profuse sweating (52613005), and sweating profusely (52613005)
Runny nose	Sniffle (275280004), nose running (267101005), running nose (267101005), nose dripping (267101005), nasal discharge (267101005), and snotty (267101005)
Nasal congestion	Nasal congestion (68235000), congested nose (68235000), stuffed-up nose (68235000), congestion (85804007), stuffed-up nose (68235000), stuffed nose (68235000), rhinobyon (68235000), nasal obstruction (232209000), nasal airway obstruction (232209000), and stuffiness (232209000)
Nosebleed	Nosebleed (249366005), nose bleeds (249366005), nose bleeding (249366005), bleeding from nose (249366005), nosebleed (249366005), nasal hemorrhage (249366005), epistaxis (249366005), and nasal hemorrhage (249366005)
Cough	Cough (49727002), coughing (49727002), nonproductive cough (11833005), hacking cough (59994004), tussiculation, dry cough (11833005), persistent cough (284523002) acute cough (49727002), bad cough (49727002), coughing all night (161933007), evening cough (161933007), morning cough (161932002), coughing up blood (66857006), coughing and deep breathing (371605008), and begma (49727002)
Headache	Headache (25064002), migraine (37796009), sick headache (193028008), tension-type headache (398057008), and cluster-headache syndrome (193031009)
Sneezing	Sneeze (76067001), sneezing (76067001), sneezing symptom (162367006), sternutation (76067001), niesen (76067001), and achoo (76067001)
Eye pain	Eye pain (41652007), pain in eye (41652007), ocular pain (41652007), ocular headache (86925001), ocular dryness (162290004), dry eye (162290004), cephalalgia (25064002), diplopia (24982008), double vision (24982008), and eyelid edema (89091004)
Loss of taste or smell	Smell (397686008), taste (397627001), lost sense of smell (44169009), absent smell (44169009), sense of smell absent (44169009), anosmia (44169009), can't smell (44169009), smell nothing (44169009), disorder of taste (399993004), loss of taste (36955009), absence of sense of taste (36955009), and ageusia (36955009)
Sputum	Sputum (45710003), expectoration (45710003), productive cough (28743005), productive cough–green sputum (161924005), productive cough–yellow sputum (161925006), and phlegm (52024008)
Shortness of breath	Respiratory disorder (50043002), respiratory disease (50043002), breath (11891009), shortness of breath (267036007), dyspnea (267036007), short breath (267036007), breathless (267036007), difficulty in breathing (230145002), breathing difficult (230145002), labored breathing (248549001), difficulty breath (230145002), and breathing painful (75483001)
Sore throat	Sore throat (267102003), throat sore (162388002), throat pain (162388002), pain in throat (162397003), and pain throat (162397003)
Dizziness	Dizzy (267102003), throat soreness (267102003), dizzy spells (315018008), and dizziness (404640003)
Intolerance to light	Intolerance of light (1285284009), photophobia (409668002), eye sensitive to light (1285284009), light sensitive (1285284009), and light sensitivity (1285284009)
Hearing findings	Pains of ears (301354004), earache (301354004), otalgia (301354004), ears pop (162346006), popping sensation in ear (162346006), tinnitus (60862001), and noise in ears (60862001)
Loss of appetite	Poor appetite (64379006), decrease in appetite (64379006), inappetence (64379006), lost my appetite (64379006), loss of appetite (79890006), no appetite (79890006), anorexia (79890006), and off food (79890006)
Hallucinations	Hallucinations (7,011,001), illusion (5,152,006), illusionary (5,152,006), auditory hallucination (45150006), visual hallucination (64269007), and see things (64269007)

Phase Difference Calculation

Because it took time to provide the official report, the COVID-19 occurrence derived from Twitter (ie, predicted curve) was expected ahead of the official report dates (ie, actual curve). Our research used the Hilbert transform (HT) [70] method to calculate the phase difference between the actual and predicted curves. HT is a method used for analyzing time-varying signals

XSL•FO
RenderX

[71]. It can transform a real-valued signal into a complex-valued signal, rendering it convenient for phase analysis. In signal analysis, HT is often used to calculate the instantaneous phase of a signal, which can be used to compare the phase difference between 2 signals. HT was selected for its ability to compute instantaneous phase and amplitude in the time domain, making it ideal for analyzing temporal alignment and detecting phase shifts between predicted and actual signals. Unlike Fourier transform, which is limited by global frequency components, HT offers a more accurate and intuitive method for capturing time-varying phase relationships. It is particularly useful for nonstationary signals, as it can analyze phase dynamics without assumptions of stationarity or periodicity. The phase spectrum of Fourier transform may have discontinuous jumps in some cases, which can lead to incorrect results when calculating phase differences. HT can accurately calculate the instantaneous phase of a signal and avoid this problem.

The daily predicted cases curve $f(t)$ and the daily actual cases curve $g(t)$ share the same time sequence. To calculate the phase difference between them by performing the HT, we can first transform them into their complex-valued signals:



(1)



(2)

Here, [H] represents the HT operator, and i is the imaginary unit. We can then calculate the instantaneous phase of each signal, usually using the *arctan* function to compute the phase angle:



(3)



(4)

Finally, we subtracted the phase angles of the 2 signals to obtain the phase difference :



(5)

In addition to calculating the phase difference between the 2 curves, we also conducted stationarity tests on both curves. Stationarity verification is an important step in time-series analysis and is used to determine whether a time series is stationary. We used 3 methods, the augmented Dickey-Fuller test [72], the Kwiatkowski-Phillips-Schmidt-Shin test [73], and the Phillips-Perron test [74], to verify the stationarity of the 2 curves. We used the TimesNet [75] approach from prior research to predict the current trends in COVID-19 development based on the time-series relationships between the self-reported case numbers and the actual case numbers.

Evaluation of the Model

Our evaluation of the model used the following methodology. True positive is the number of correct predictions in positive samples, false positive is the number of incorrect predictions in positive samples, true negative is the number of correct predictions in negative samples, and false negative is the number of incorrect predictions in negative samples. Precision is the proportion of positive predictions in all positive samples. Precision is defined as follows:



(6)

Recall is the proportion of correct predictions in the total samples. Recall is defined as follows:



(7)

Accuracy is defined as the percentage of correctly predicted results out of the total sample.



(8)

F_1-score is defined as follows:



(9)

To measure the performance under the unbalanced data distribution, in this work, we used the precision-recall (PR) curve and ROC curve to display the performance. The ROC curve is a curve of sensitivity versus 1–specificity on all possible prediction thresholds. Similarly, the PR curve plots precision versus recall on all possible prediction thresholds. In addition, average precision (AP) and area under the curve (AUC) derived from the PR curve and the ROC curve are also generated for quantitative comparisons in this work.

Results

Model Performance for Self-Reporting COVID-19 Cases

We evaluated the models with AUC, AUC-PR, accuracy, precision, recall, and F_1-score. RoBERTa and BERT achieved the best performance with an AUC of 0.98 and an AP of 0.97, as depicted in Figure 2. Notably, all LLMs outperformed traditional machine learning models in AUC and AP, exhibiting an AUC gain from 0.01 to 0.10 and an AP gain from 0.07 to 0.09. According to the benchmark results in Table 3, BLOOM performed the best compared with other models in accuracy and precision with 0.948 and 0.941, whereas the SVM outperformed others in recall and F_1-score with 0.9362 and 0.9329, respectively. Combining AUC, accuracy, and recall, we believe that the BLOOM model has the best performance.

XSL•FO
RenderX

Subsequently, we used the trained model to assist us in selecting self-reported positive tweets from all COVID-19–related tweets.

Figure 2. Performance of various machine learning methods. (A) Receiver operating characteristics (ROC) curve with area values. The area under the curve (AUC) values, provided in the legend, quantify the overall performance of the models, with higher values indicating superior discriminative ability. (B) Precision-recall curve with average precision. AP: average precision; BERT: Bidirectional Encoder Representations from Transformers; BLOOM: Bigscience Large Open-science Open-access Multilingual language model; Llama-2: large language model meta AI 2; LR: logistic regression; NB: naive Bayes; RoBERTa: robustly optimized BERT pretraining approach; SVM: support vector machine; XLNet: extreme language model.

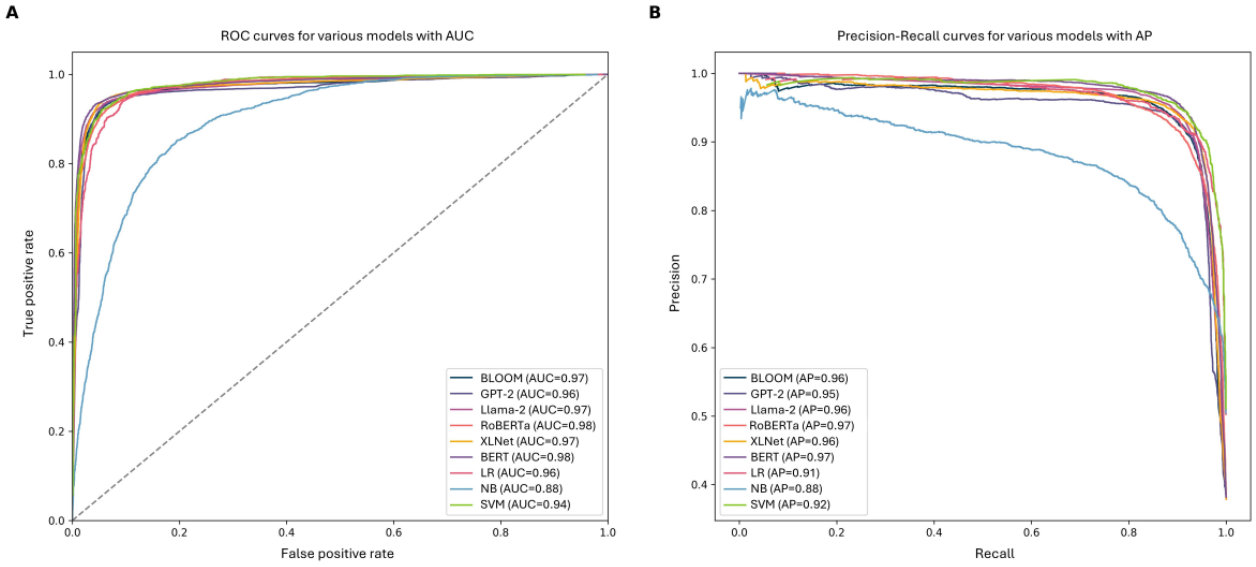

Table 3. Performance comparison of various machine learning models and large language models.

	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)
Machine learning				
Naive Bayes [55]	86.86	65.62	52.43	58.29
Logistic regression [57]	92.57	91.73	92.64	92.18
Support vector machine [56]	93.62	92.96	*93.62*	*93.29*
Large language models				
BERT[a] [49]	93.80	92.50	91.00	91.70
RoBERTa[b] [58]	93.60	91.20	91.80	91.50
XLNet[c] [59]	93.00	90.40	91.30	90.80
GPT-2 [60]	94.30	92.00	91.00	92.60
BLOOM[d] [61]	*94.80*	*94.10*	92.00	93.00
Llama-2[e] (7b) [62]	94.20	93.30	91.30	92.30

[a]BERT: Bidirectional Encoder Representations from Transformers.

[b]RoBERTa: robustly optimized BERT pretraining approach.

[c]XLNet: extreme language model.

[d]BLOOM: Bigscience Large Open-science Open-access Multilingual language model.

[e]Llama-2: large language model meta AI 2.

Predicted Trend of COVID-19 Cases

From January 1, 2020, to April 1, 2023, we used the trained BLOOM model to evaluate all the collected COVID-19–related tweets and filtered out the tweets that the model predicted as self-reported positive cases of COVID-19. Among 7.3 million COVID-19–related tweets, we identified 317,500 self-reported tweets. Using unique user IDs, we considered multiple tweets reporting a COVID-19 diagnosis by the same user to be a single case, resulting in 262,278 unique self-reported cases. The IDs of these unique confirmed users have been stored separately in an in-house database named the COVID-19 patient database (CPD), and the daily predicted number of confirmed cases by the model has been stored according to Coordinated Universal Time.

XSL•FO
RenderX

To compare the predicted daily case counts with the actual daily case counts, we obtained the actual daily case counts from public platforms such as Johns Hopkins University and *The New York Times*. Then, we plotted the actual daily case count and predicted case count on a curve, as shown in Figure 3. The blue line represents the daily actual case counts, and the red line represents our predicted case counts. The red text in the figure describes key events during the outbreak, and the brown text represents the time the variant appeared. We used the HT method to calculate the phase difference between the two curves and found that the predicted curve was leading the actual curve by approximately 7.63 days. The Augmented Dickey-Fuller test results indicated values below the critical values of 1%, 5%, and 10%, accompanied by simultaneous *P* values of <.001 and <.001, which rejects the hypothesis of the existence of a unit root. In addition, both the Phillips-Perron and Kwiatkowski-Phillips-Schmidt-Shin tests exhibited *P* values <.05. Collectively, the outcomes from these 3 tests consistently pointed toward the smoothness of the time series under scrutiny as shown in Multimedia Appendix 8.

There are 2 distinct peaks in the red curve. We examined the data for the first peak of the predicted curve on October 2, 2020, and the second peak on November 21, 2020. We found that the first peak on October 2, 2020, was due to then US President Donald Trump's tweet announcing his positive COVID-19 test result, which triggered many Twitter users to also report their self-diagnosed cases on Twitter. On that day, there were 1495 tweets related to self-reported positive cases. As for the second peak in the prediction curve, we examined the relevant tweets on that day and found that most of them were related to the US election results. Many users tweeted about their infection status and discussed the US epidemic-prevention policies. There were 973 tweets on that day regarding self-reported cases of COVID-19 infection out of all tweets.

Figure 3. Real cases and predicted cases curves. The blue curve represents the actual daily confirmed cases, and the red curve represents the daily predicted cases. The shaded areas above and below the red and blue curves represent the CI. The red text represents key events during the outbreak, and the brown text represents the time the variant appeared. The blue shaded area on the right side represents the period during which actual confirmed case data are missing. The solid red line represents the daily self-reported COVID-19 infection numbers, and the blue dotted line represents the predicted actual infection numbers. JHU: Johns Hopkins University; WHO: World Health Organization.

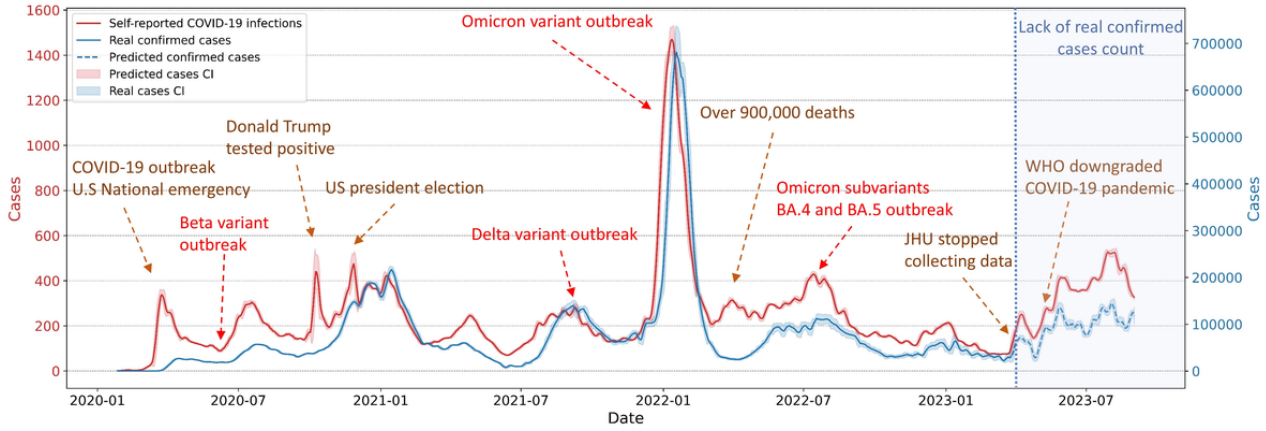

Symptoms

We extracted the historical tweets of users from the CPD using their unique user IDs. The temporal scope of these tweets spanned a period from 1 month preceding the self-reported date of symptom onset to 9 months after the diagnosis date, encompassing a total duration of 10 months. Within a cohort of 24,316 reporting COVID-19 symptoms, an analysis of historical tweets identified the top 10 symptoms. Figure 4A plots the temporal frequency of COVID-19 symptom mentions, providing insight into how symptom prevalence evolved over time. Notably, *fever*, *headache*, *fatigue*, and *cough* emerged as consistently common symptoms. The trends observed in these symptoms closely parallel the overall trend in confirmed COVID-19 cases. We also identified that the onset of symptoms such as *loss of taste or smell* and *shortness of breath* first became prominent in September 2020, possibly correlating with the emergence of the Beta variant. Similarly, the prevalence of *sore throat* spiked in late 2021, potentially aligning with the rise of the Omicron variant. The symptom *difficulty breathing* maintained a steady presence across the timeline. Notably, less commonly reported symptoms, such as *hallucinations* and *eye pain*, not currently recognized by the US Centers for Disease Control and Prevention (CDC), appeared sporadically in user reports, suggesting their rarity in COVID-19 cases. As shown in Figure 4B, these were *fever* (11,613/24,316, 47.76% mentioned), *headache* (8347/24,316, 34.33%), *cough* (7985/24,316, 32.84%), *generalized body* aches (6893/24,316, 28.35%), *difficulty breathing* (6169/24,316, 25.37%), *fatigue* (5984/24,316, 24.61%), *pain* (5806/24,316, 23.88%), *disorder of smell and taste* (5444/24,316, 22.39%), and *sore throat* (5082/24,316, 20.9%), listed in descending order. Notably, all symptoms except for *eye pain* (2541/24,316, 10.45%) aligned with those recognized by the CDC. Additional symptoms, such as *lethargy* (2176/24,316, 8.95%), *dizziness* (1451/24,316, 5.97%), and *hallucinations* (1086/24,316, 4.47%), although mentioned by a minority group, are not currently acknowledged as COVID-19 symptoms by the CDC.

In the dataset of historical tweets from diagnosed individuals, we observed instances in which a patient mentioned multiple symptoms concurrently. To quantify this, we calculated the frequency of cooccurrence of any 2 symptoms and constructed a dependency graph to illustrate these relationships. Figure 4C

XSL•FO
RenderX

elucidates the correlations among various symptoms, highlighting that most individuals with the infection reported experiencing a constellation of related symptoms, such as headache, cough, and fever. Furthermore, Multimedia Appendix 9 presents a heat map that visualizes the Pearson correlation coefficients [76] among these symptoms, offering a quantitative view of their interdependencies. To visually represent the range of self-reported symptoms, we used a word cloud. This graphical representation provides an immediate overview of the symptomatology as expressed by the users in our dataset.

Figure 4. All symptoms mentioned by self-reporting tweets and the correlations among symptoms. (A) represents the number of mentions of COVID-19 symptoms in self-reported tweets over time. (B) represents the percentage of symptoms in all self-reporting COVID tweets, and multiple symptoms can be mentioned in 1 tweet. (C) represents the correlations among symptoms mentioned by the same user. The width of the line between 2 symptoms represents the number of tweets that mention both symptoms.

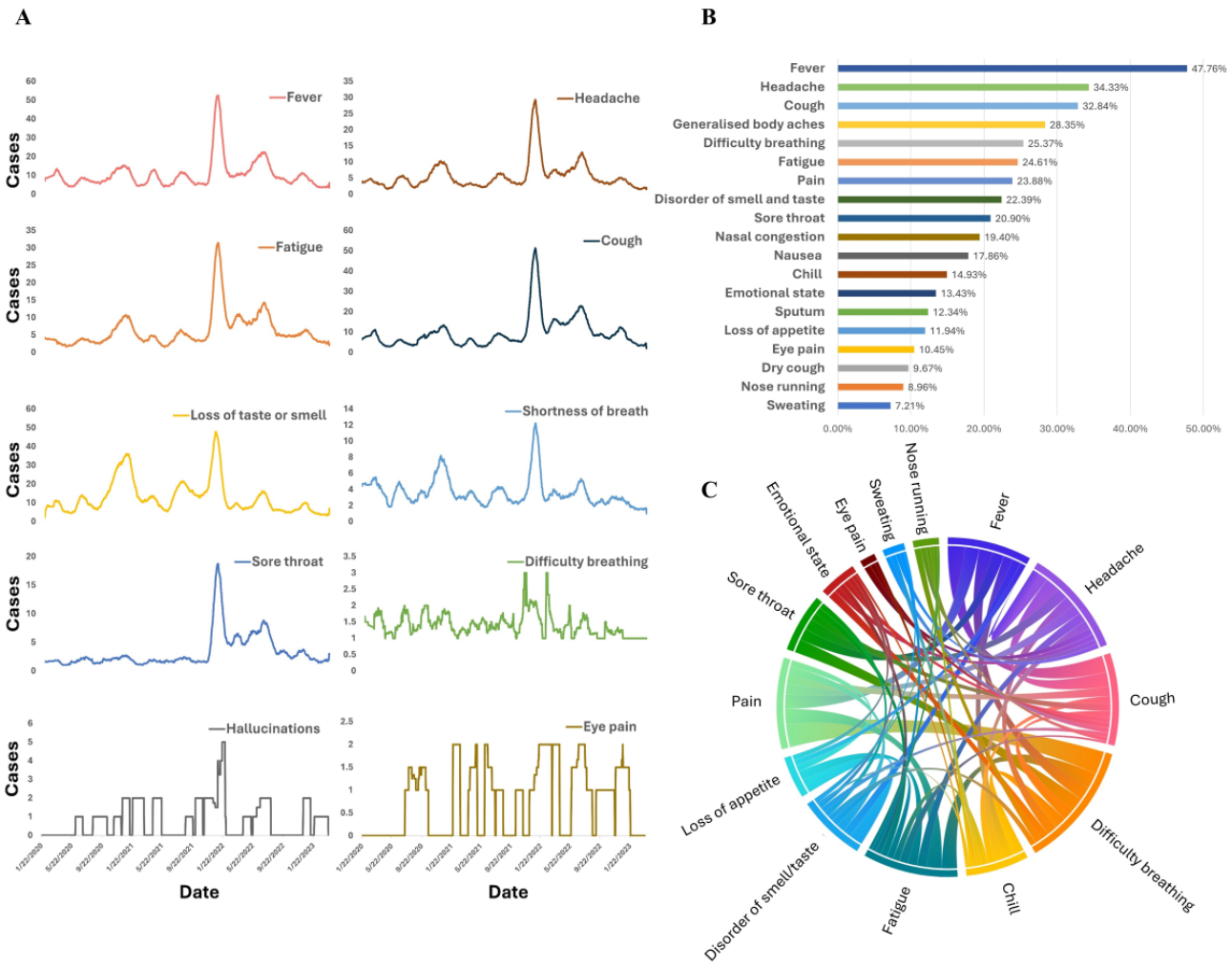

Reinfection and Rehabilitation

We analyzed historical tweets from users who self-reported COVID-19 infections, identifying 723 individuals who shared their recovery experiences. The annual breakdown of these individuals is as follows: 174 in 2020, 163 in 2021, 135 in 2022, and 251 in 2023. The duration of recovery was primarily inferred from the period mentioned in their tweets. In instances in which the recovery period was not explicitly stated, we computed it by calculating the interval between the date of confirmed diagnosis and the date of reported recovery. The data on self-reported recovery durations have been depicted through the Kaplan-Meier recovery curve [77], as illustrated in Figure 5A. This graphical representation reveals a gradual decrease in recovery time for patients with COVID-19 from 2020 to 2023. Specifically, in 2020, most patients reported a recovery period of around 30 days, with very few recovering in <30 days. In contrast, by 2023, the trend had shifted significantly, with most individuals reporting a recovery within approximately 12 days, despite a minority group still experiencing recovery periods extending beyond 30 days. Figure 5B presents a comprehensive overview of the evolution of recovery periods from 2020 to 2023. In addition, this figure suggests that the prevalent COVID-19 cases in 2023 were predominantly mild, indicating a possible decrease in the virulence of the virus over time.

In this study, we defined a recurrent COVID-19 infection in an individual as a self-reported reinfection occurring >30 days after the initial confirmed positive diagnosis. We meticulously tracked the historical tweets of all confirmed patients in CPD. Figure 5C presents the distribution of the intervals between the first and second infections alongside the corresponding case counts. This analysis reveals a relatively low likelihood of a

XSL•FO
RenderX

repeat infection within 180 days. Most patients who had recovered from an initial infection reported a second infection approximately 260 days later. Moreover, repeat infections occurring between 300 and 600 days after recovery were also relatively frequent. The longest interval between repeat infections documented in our study extended to 720 days. Of 262,278 patients who self-reported a positive COVID-19 test result, 91.12% (n=238,993) indicated a single infection event, and 6.83% (n=17,906) reported 2 infections. A smaller subset, comprising 1.25% (n=3283) of individuals, reported 3 infections; 0.39% (n=1025) indicated 4 infections; 0.17% (n=445) reported 5 infections; 0.11% (n=301) of individuals reported 6 infections; and 0.12% (n=325) indicated ≥7 infections. Remarkably, the highest number of reported reinfections was 9, with 7 individuals documenting their ninth infection. Figure 5D shows that among the 238,993 patients with a single infection, 7.49% (17,906/238,993) reported a second infection. A further breakdown shows that 1.37% (3283/238,993) reported a third infection, 0.43% (1025/238,993) reported a fourth infection, and 0.45% (1071/238,993) reported experiencing ≥5 infections. We also performed a statistical analysis of the time intervals between infections among users with multiple infections.

Figure 5. Overview of reinfections and recovery. (A) Kaplan-Meier estimates of cumulative recoveries. (B) Rehabilitation days in different years. (C) Time to reinfection for 238,993 individuals. (D) Reinfection cases and rates.

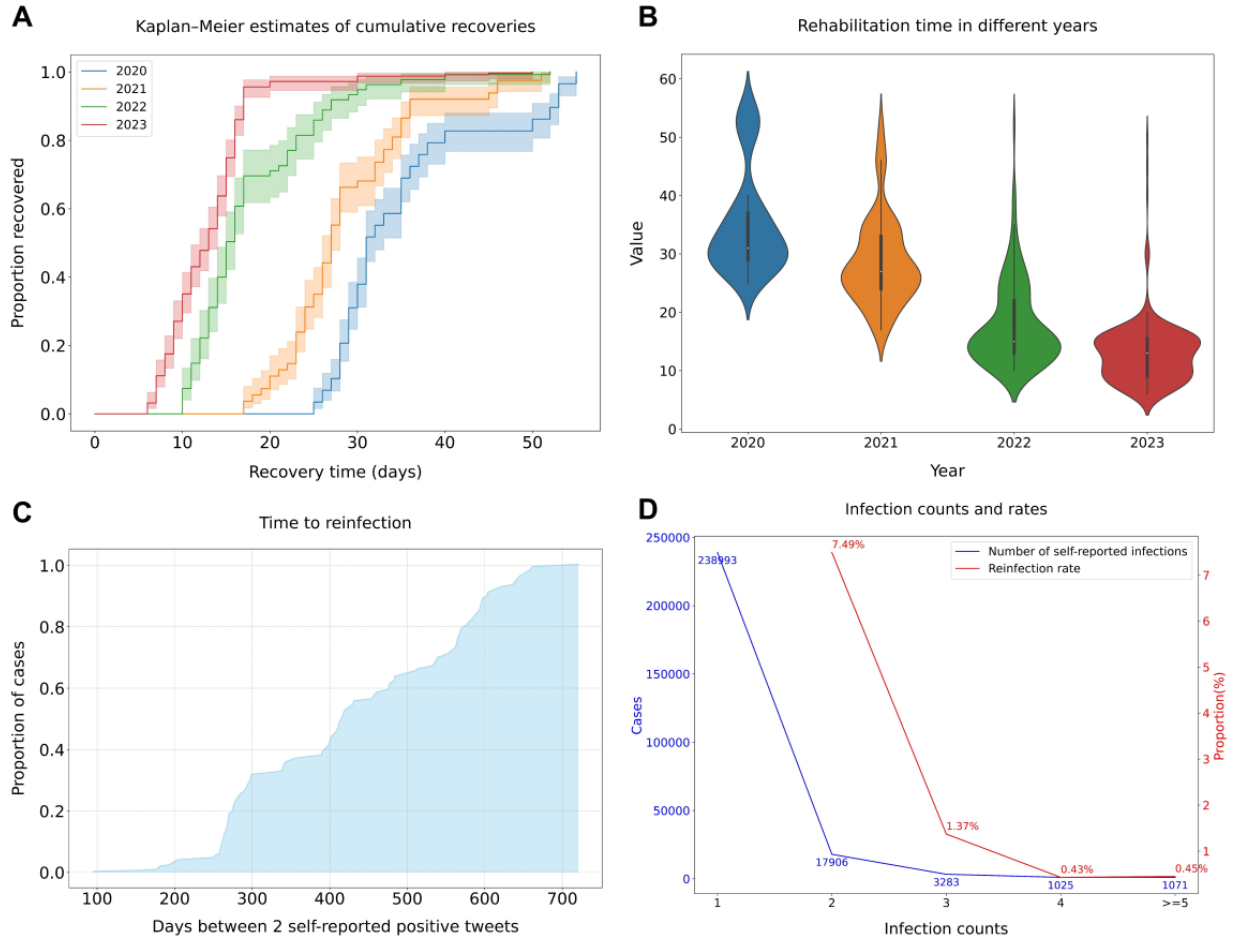

Distribution of Cases

We extracted the geographic locations of the diagnosed users in CPD, and the distribution of all confirmed patients is shown in Figure 6A [52]. California had the highest number of self-reported COVID-19 cases, with 8762 cases, followed by Texas (6619 cases), Florida (4245 cases), New York (3566 cases), Illinois (2649 cases), Pennsylvania (2032 cases), Ohio (1868 cases), Massachusetts (1793 cases), Georgia (1785 cases), and Michigan (1677 cases), in descending order. Alabama and Northern Mariana Islands had the least data, with only 1 self-reported case in each state.

The details of confirmed cases in each state are shown in Figure 6B, in which we provide the average case counts for the entire country and each state as well as the number of self-reported cases per 10 million people per week, the trend over the past 2 weeks, and the positivity rate of each state. For example, we can see that California had the most self-reported COVID-19 cases in the past week, with at least 29 people reporting positive test results. Approximately 7.34 people per 10 million reported self-diagnosed COVID-19–positive status, and the average positivity rate of self-reported cases was 26.69%. However, compared with the previous 2 weeks, the number of self-reported COVID-19 cases decreased by 10.93%. Due to insufficient data, the trend of changes in the past 14 days was unavailable for several states, such as the Virgin Islands and Wyoming.

We also plotted the time-varying curves for the confirmed cases in the top 20 states in terms of confirmed cases, as shown in

XSL•FO
RenderX

Figure 6C and Multimedia Appendix 10. It is evident that the changes in the number of confirmed cases in the top 4 states with the highest number of cases closely resemble the overall trend in the United States. Some states, such as Washington, Arizona, Washington DC, and Indiana, exhibited relatively consistent changes in the number of confirmed cases over time, whereas states such as Nevada, Colorado, Alabama, and Michigan had less-consistent curves, with some dates showing no reported cases. The variation in results could have been influenced by the differing numbers of Twitter users in each state.

Figure 6. Cases across the United States and by state. (A) Self-reported number of infections in each state, with darker colors representing higher numbers. (B) Self-reported weekly number of infections, number of infections per 10 million people, change in trend over the previous 2 weeks, and positivity rate for a subset of states (alphabetical order). (C) Infection curve for the top 20 states with the highest infection numbers.

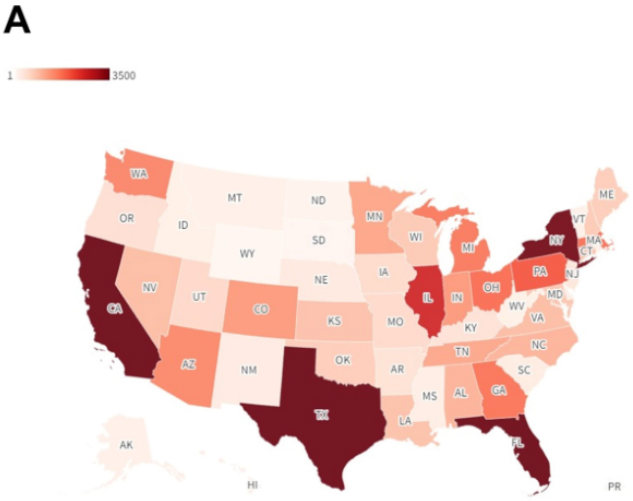

State	Weekly average	Cases per 10 million people	Biweekly change (%)	Test positivity
US	777	114.800	7.527%	17.832
AL	1	2.039	-55.556%	22.094
AK	2	27.339	33.333%	19.861
AZ	6	8.243	-9.091%	21.897
AR	1	3.314	-33.333%	13.883
CA	27	6.833	-14.063%	22.953
CO	9	15.628	62.5%	24.294
CT	6	16.829	450%	25.012
DE	1	10.269	100%	17.611
FL	8	3.725	-30.556%	20.412

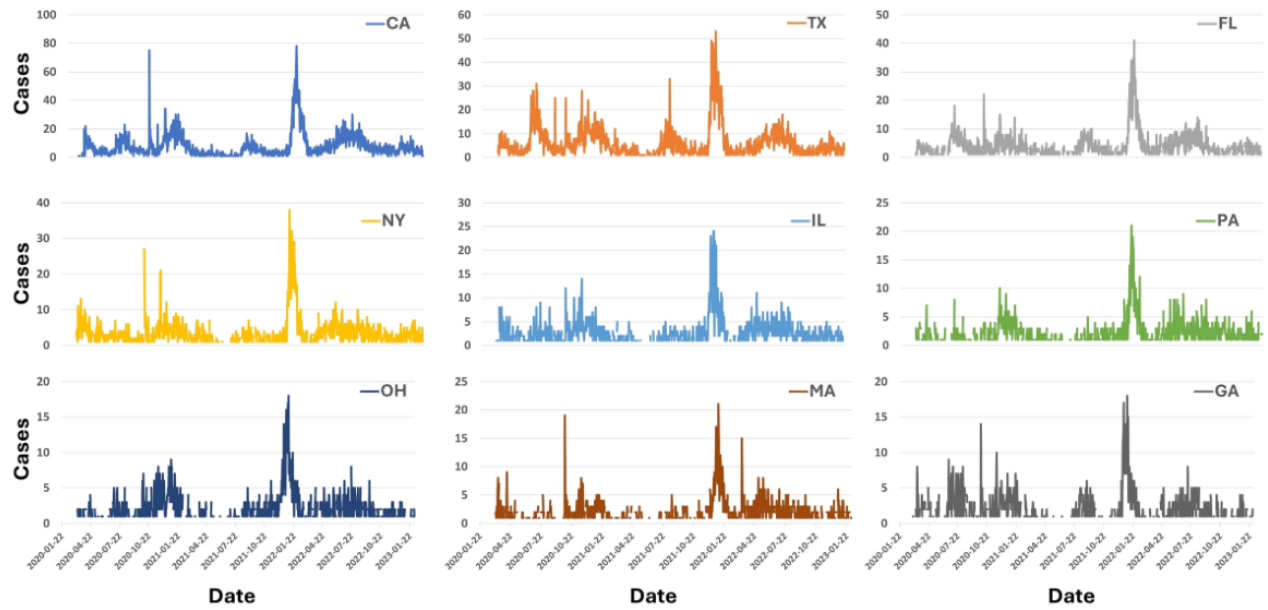

Covlab Visualization Website

To disseminate timely information to the public and policy makers, we have launched a visualization website, Covlab, which features our trained models and a comprehensive data pipeline. This platform consists of an automated script sequence designed to update the data weekly. The home page of Covlab offers users real-time access to the total number of tweets collected to date, including those self-reported as COVID-19–positive. On the Graphs page, users can explore the most recent weekly growth trends in COVID-19 cases alongside predictive models of actual confirmed cases. This section also enables monitoring of infection trends across various states in the United States as well as the prevailing symptom patterns observed to the current date. Furthermore, Covlab displays the proportions of reinfections and recovery

XSL•FO
RenderX

periods, with these metrics also being refreshed weekly. The website's dynamic graphs and tables serve as a valuable resource, providing the public with up-to-date information on the ongoing evolution of COVID-19, including insights into the symptomatic expressions of emerging variants.

Discussion

Principal Findings

The analysis of self-reported COVID-19 tweets offers a valuable perspective, reflecting the actual progression of the pandemic to a considerable degree. The voluminous data generated by self-reporting individuals on Twitter augment clinical datasets, offering a complementary avenue for long-term observation and tracking. This approach is particularly beneficial in bridging the data gap inherent in in-home self-testing scenarios. Our research indicates that reinfections are relatively common, with the likelihood of multiple infections diminishing over time. Moreover, this dataset has proven instrumental in identifying potential new symptoms that were not initially associated with COVID-19, such as lethargy and hallucinations, offering an early warning system for evolving clinical presentations. In addition, the approach has enabled the detection of emerging variants through geospatial clustering of self-reported cases exhibiting distinctive symptom combinations. This tracking capability enhances preparedness and enables timely public health interventions. The breakdown by state of self-reported COVID-19 tweets facilitates more efficient tracking of disease trends. Furthermore, the approach used in this study could potentially serve as a general pipeline for researching and analyzing other infectious diseases.

Symptoms and Sequelae Study

Within the subset of Twitter users who self-reported as COVID-19 positive, numerous accounts contained detailed descriptions of symptoms experienced during the infection. These firsthand accounts are a rich source of information. We extracted and analyzed the symptomatology mentioned in these tweets, compiling a list of the most prevalent symptoms reported by patients. Our intriguing findings closely mirror the commonly acknowledged COVID-19 symptoms listed by the CDC. However, some symptoms mentioned by a subset of patients in their tweets are not yet recognized as typical by the CDC. This discrepancy highlights the significant role of our study in identifying potential new symptoms of COVID-19 that have not been widely recognized. Such findings could provide valuable guidance for further clinical investigation into these symptoms and their association with different COVID-19 variants. In addition, the CPD established through this study offers a robust framework for long-term patient tracking. This database not only complements existing clinical data but also provides an invaluable resource for the study of post–COVID-19 sequelae and post–COVID-19 condition symptoms, thereby enhancing our understanding of the virus's long-term impacts.

Comparison With Traditional Tracking Tools

Traditional clinical data–driven tools, while offering high specificity by relying on confirmed diagnoses, often underrepresent asymptomatic or mild cases. In contrast, our approach leverages self-reported data from social media platforms, such as Twitter, to capture a broader spectrum of cases, including those not reported in clinical settings. Although this may introduce noise, advanced filtering and NLP techniques effectively mitigate inaccuracies, ensuring meaningful insights. Furthermore, clinical data collection and reporting are often delayed due to testing and diagnosis, whereas our tool enables near–real-time monitoring by analyzing self-reported cases as they are posted, facilitating rapid identification of trends and potential hot spots. In terms of coverage, traditional clinical tracking tools may lack granularity, particularly in regions with limited health care infrastructure or reporting capabilities. Using social media, our tool significantly enhances coverage, incorporating underrepresented populations and geographies to provide a more comprehensive view of the pandemic's spread. A key contribution of this approach is the integration of self-reported data, which captures personal narratives, symptom progression, and public sentiment, thereby complementing clinical data to offer a richer and more dynamic understanding of the pandemic. Importantly, while our tool excels in timeliness and coverage, its true potential lies in synergy with clinical data–driven tools. We propose future efforts to integrate these approaches, enabling cross-validation and improving overall accuracy.

Limitations and Future Work

Although our method and website are highly useful, there are some limitations. First, Twitter has a limited quota of public APIs, which renders the platform hugely expensive to run. Second, potential biases in data collection and the varying distribution of Twitter users across different states may impact the predictive accuracy. Furthermore, external factors, particularly notable events in the United States that occurred in 2020, influenced our prediction outcomes. For instance, the 2 notable peaks in self-reported cases observed in 2020 did not necessarily correlate with an actual increase in infections. Instead, these peaks were primarily driven by external events, prompting a surge in infection reporting on Twitter on those specific days. Another significant constraint is the veracity of the information in tweets. However, despite their questionable reliability, these data offer informative trend analyses and hypotheses valuable for future research and validation. Our framework demonstrates significant versatility, with potential applications extending beyond COVID-19 to monitor other self-reported illnesses, such as influenza and respiratory syncytial virus, as well as chronic conditions such as diabetes and mental health issues. Its modular design facilitates easy adaptation by allowing the customization of keywords and analytical strategies tailored to specific diseases. To enhance its functionality, we propose integrating Twitter data with other real-time health data sources, such as wearable devices and electronic health records, enabling a more comprehensive approach to health trend monitoring. In this integration, social media signals would serve as an initial screening mechanism, effectively complementing clinical data. In future work, we plan to integrate self-reported data from other social platforms, such as Reddit, to reduce data limitations and biases, thereby broadening the scope of our analysis. This expansion will support the development of a platform capable of predicting

XSL•FO
RenderX

COVID-19 trends in real time based on self-reported content. Furthermore, we aim to apply this pipeline to other infectious diseases that may emerge in the future, facilitating the understanding and tracking of their development and trends. To maintain robustness and accuracy, we plan periodic retraining of the model using updated data and incorporating LLMs to better capture nuanced expressions in self-reported content. A user feedback mechanism will also be implemented to address false positives and negatives. In addition, we aim to expand the tool's global applicability by automating data pipelines to support content in multiple languages, with a particular focus on underrepresented regions. This global expansion will enable the tool to capture disease trends across diverse geographic and cultural contexts, offering a more holistic view of global health dynamics.

Conclusions

This study demonstrates the significant potential of using self-reported COVID-19 tweets from social media platforms for public health monitoring. By leveraging machine learning and NLP techniques, we developed a tool capable of identifying infection and recovery trends, providing valuable insights into disease spread and public behavior. Our findings contribute to the growing field of digital epidemiology, emphasizing that social media can serve as an effective complementary data source to traditional public health surveillance systems. Beyond COVID-19, this approach holds promise for monitoring other infectious diseases, mental health conditions, and chronic illnesses by adapting the model to new health-related keywords and contexts. Integrating such tools with existing health infrastructure could enhance early detection, improve situational awareness, and enable more proactive public health responses. However, the ethical considerations of using publicly available data and addressing biases inherent to social media platforms must be prioritized to ensure responsible use. Overall, this work highlights the evolving role of digital tools in public health informatics and presents opportunities for future research to further refine these methods. The integration of social media–based monitoring with traditional data systems could transform public health strategies, making them more adaptive, responsive, and inclusive of diverse population segments.

Acknowledgments

This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R35-GM126985). The authors would like to thank all the annotators who participated in the calibration of their data, including Kent Studer, James Tipton, Thadeus Meneses, Noah Berry, Sai Akhil Chopparapu, Trinath Adusumilli, Weinan Zhang, Yijie Ren, Li Su, Chunhui Xu, and Congyu Guo.

Data Availability

The aggregated datasets analyzed in this study are available from the corresponding author upon request. Manual labeling data can be viewed through the annotation website. Guest users can use the username *guest* and password *guest* to log into the system for data access. Actual and predicted daily cases are publicly available on the Covlab website [52].

Authors' Contributions

JX and ZZ collected the data and performed the analyses. JH created the data visualizations. SZ ran the large language model and wrote the model performance portion. GA, XT, and XW reviewed and assessed the results. LJ and YY provided technical support. JX drafted the manuscript and designed the website; JX and ZZ wrote the code and conducted all the tests. DX conceived and supervised the study. All authors had full access to all the data in the study as needed and had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A list of keywords and hashtags used in data collection.
[DOCX File , 17 KB - jmir_v27i1e63190_app1.docx]

Multimedia Appendix 2
Stop word list.
[DOCX File , 20 KB - jmir_v27i1e63190_app2.docx]

Multimedia Appendix 3
Interannotator agreement metrics.
[DOCX File , 20 KB - jmir_v27i1e63190_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Data preprocessing process and annotation system.
[DOCX File , 3873 KB - jmir_v27i1e63190_app4.docx]

Multimedia Appendix 5
An example of the target cohort.
[DOCX File , 277 KB - jmir_v27i1e63190_app5.docx]

Multimedia Appendix 6
Number of parameters for conventional machine learning and large language models.
[DOCX File , 17 KB - jmir_v27i1e63190_app6.docx]

Multimedia Appendix 7
Configuration for conventional machine learning and large language models.
[DOCX File , 17 KB - jmir_v27i1e63190_app7.docx]

Multimedia Appendix 8
Multiple stationarity verification methods and their results.
[DOCX File , 20 KB - jmir_v27i1e63190_app8.docx]

Multimedia Appendix 9
Heat map of Pearson correlation coefficients among symptoms and symptoms word cloud.
[DOCX File , 554 KB - jmir_v27i1e63190_app9.docx]

Multimedia Appendix 10
Infection curve for the top 20 states with the highest infection numbers.
[DOCX File , 291 KB - jmir_v27i1e63190_app10.docx]

References

1. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic. World Health Organization. URL: https://tinyurl.com/nhk5m9y8 [accessed 2024-04-29]
2. COVID 19 data. Johns Hopkins. URL: https://coronavirus.jhu.edu/map.html [accessed 2024-04-29]
3. Jiang JX, Cram P, Qi K, Bai G. Challenges and dynamics of public health reporting and data exchange during COVID-19: insights from US hospitals. Health Aff Sch 2024 Jan;2(1):qxad080. [doi: 10.1093/haschl/qxad080] [Medline: 38756405]
4. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 2020 Nov 26;10(1):20693 [FREE Full text] [doi: 10.1038/s41598-020-77275-9] [Medline: 33244028]
5. Social media and attitudes towards a COVID-19 vaccination: a systematic review of the literature. eClinicalMedicine. URL: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(22)00184-5/fulltext [accessed 2025-01-13]
6. Huang X, Wang S, Zhang M, Hu T, Hohl A, She B, et al. Social media mining under the COVID-19 context: progress, challenges, and opportunities. Int J Appl Earth Obs Geoinf 2022 Sep;113:102967 [FREE Full text] [doi: 10.1016/j.jag.2022.102967] [Medline: 36035895]
7. Klein AZ, Kunatharaju S, O'Connor K, Gonzalez-Hernandez G. Automatically identifying self-reports of COVID-19 diagnosis on twitter: an annotated data set, deep neural network classifiers, and a large-scale cohort. J Med Internet Res 2023 Jul 03;25:e46484 [FREE Full text] [doi: 10.2196/46484] [Medline: 37399062]
8. Jun J, Wickersham K, Zain A, Ford R, Zhang N, Ciccarelli C, et al. Cancer and COVID-19 vaccines on Twitter: the voice and vaccine attitude of cancer community. J Health Commun 2023 Jan 02;28(1):1-14 [FREE Full text] [doi: 10.1080/10810730.2023.2168800] [Medline: 36755484]
9. Wu J, Wang L, Hua Y, Li M, Zhou L, Bates DW, et al. Trend and co-occurrence network of COVID-19 symptoms from large-scale social media data: infoveillance study. J Med Internet Res 2023 Mar 14;25:e45419 [FREE Full text] [doi: 10.2196/45419] [Medline: 36812402]
10. Zaidi Z, Ye M, Samon F, Jama A, Gopalakrishnan B, Gu C, et al. Topics in antivax and provax discourse: yearlong synoptic study of COVID-19 vaccine tweets. J Med Internet Res 2023 Aug 08;25:e45069 [FREE Full text] [doi: 10.2196/45069] [Medline: 37552535]
11. Storey VC, O'Leary DE. Text analysis of evolving emotions and sentiments in COVID-19 Twitter communication. Cognit Comput 2022 Jul 28:1-24 [FREE Full text] [doi: 10.1007/s12559-022-10025-3] [Medline: 35915743]

XSL•FO
RenderX

12. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research-an international collaboration. Epidemiologia (Basel) 2021 Aug 05;2(3):315-324 [FREE Full text] [doi: 10.3390/epidemiologia2030024] [Medline: 36417228]
13. Naseem U, Razzak I, Khushi M, Eklund PW, Kim J. COVIDSenti: a large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE Trans Comput Soc Syst 2021 Aug;8(4):1003-1015. [doi: 10.1109/tcss.2021.3051189]
14. Müller M, Salathé M, Kummervold PE. COVID-Twitter-BERT: a natural language processing model to analyse COVID-19 content on Twitter. Front Artif Intell 2023;6:1023281 [FREE Full text] [doi: 10.3389/frai.2023.1023281] [Medline: 36998290]
15. Alqurashi S, Alhindi A, Alanazi E. Large Arabic Twitter dataset on COVID-19. arXiv Preprint posted online April 9, 2020 [FREE Full text] [doi: 10.48550/arXiv.2004.04315]
16. Imran M, Qazi U, Ofli F. TBCOV: two billion multilingual COVID-19 tweets with sentiment, entity, geo, and gender labels. Data 2022 Jan 10;7(1):8. [doi: 10.3390/data7010008]
17. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi MA, Yang Y. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. J Am Med Inform Assoc 2020 Aug 01;27(8):1310-1315 [FREE Full text] [doi: 10.1093/jamia/ocaa116] [Medline: 32620975]
18. Guo JW, Radloff CL, Wawrzynski SE, Cloyes KG. Mining Twitter to explore the emergence of COVID-19 symptoms. Public Health Nurs 2020 Nov;37(6):934-940 [FREE Full text] [doi: 10.1111/phn.12809] [Medline: 32937679]
19. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data infoveillance study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]
20. Feng Y, Zhou W. Work from home during the COVID-19 pandemic: an observational study based on a large geo-tagged COVID-19 Twitter dataset (UsaGeoCov19). Inf Process Manag 2022 Mar;59(2):102820 [FREE Full text] [doi: 10.1016/j.ipm.2021.102820] [Medline: 34903906]
21. Rusli N, Nordin NZ, Ak Matusin AM, Yusof JN, Rosley MS, Ling GH, et al. Geospatial mapping of suicide-related tweets and sentiments among Malaysians during the COVID-19 pandemic. Big Data Cogn Comput 2023 Mar 28;7(2):63. [doi: 10.3390/bdcc7020063]
22. Klein AZ, Magge A, O'Connor K, Flores Amaro JI, Weissenbacher D, Gonzalez Hernandez G. Toward using Twitter for tracking COVID-19: a natural language processing pipeline and exploratory data set. J Med Internet Res 2021 Jan 22;23(1):e25314 [FREE Full text] [doi: 10.2196/25314] [Medline: 33449904]
23. Sukhavasi N, Misra J, Kaulgud V, Podder S. Geo-sentiment trends analysis of tweets in context of economy and employment during COVID-19. J Comput Soc Sci 2023 Mar 23;6(2):1-31 [FREE Full text] [doi: 10.1007/s42001-023-00201-2] [Medline: 37363804]
24. Forati AM, Ghose R. Geospatial analysis of misinformation in COVID-19 related tweets. Appl Geogr 2021 Aug;133:102473 [FREE Full text] [doi: 10.1016/j.apgeog.2021.102473] [Medline: 34103772]
25. Cuomo RE, Purushothaman V, Li J, Cai M, Mackey TK. A longitudinal and geospatial analysis of COVID-19 tweets during the early outbreak period in the United States. BMC Public Health 2021 Apr 24;21(1):793 [FREE Full text] [doi: 10.1186/s12889-021-10827-4] [Medline: 33894745]
26. Chi G, Yin J, Smith ML, Bodovski Y. Global Tweet mentions of COVID-19. arXiv Preprint posted online August 13, 2021 [FREE Full text] [doi: 10.5260/chara.21.2.8]
27. Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, et al. Tracking mental health and symptom mentions on Twitter during COVID-19. J Gen Intern Med 2020 Sep 07;35(9):2798-2800 [FREE Full text] [doi: 10.1007/s11606-020-05988-8] [Medline: 32638321]
28. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
29. Wissel BD, Van Camp PJ, Kouril M, Weis C, Glauser TA, White PS, et al. An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. J Am Med Inform Assoc 2020 Jul 01;27(7):1121-1125 [FREE Full text] [doi: 10.1093/jamia/ocaa071] [Medline: 32333753]
30. Zohner YE, Morris JS. COVID-TRACK: world and USA SARS-COV-2 testing and COVID-19 tracking. BioData Min 2021 Jan 20;14(1):4 [FREE Full text] [doi: 10.1186/s13040-021-00233-2] [Medline: 33472672]
31. Guedes AR, Oliveira MS, Tavares BM, Luna-Muschi A, Lazari CD, Montal AC, et al. Reinfection rate in a cohort of healthcare workers over 2 years of the COVID-19 pandemic. Sci Rep 2023 Jan 13;13(1):712 [FREE Full text] [doi: 10.1038/s41598-022-25908-6] [Medline: 36639411]
32. Flacco ME, Acuti Martellucci C, Soldato G, Carota R, Fazii P, Caponetti A, et al. Rate of reinfections after SARS-CoV-2 primary infection in the population of an Italian province: a cohort study. J Public Health (Oxf) 2022 Dec 01;44(4):e475-e478 [FREE Full text] [doi: 10.1093/pubmed/fdab346] [Medline: 34492110]
33. Rivelli A, Fitzpatrick V, Blair C, Copeland K, Richards J. Incidence of COVID-19 reinfection among Midwestern healthcare employees. PLoS One 2022 Jan 4;17(1):e0262164 [FREE Full text] [doi: 10.1371/journal.pone.0262164] [Medline: 34982800]

XSL•FO
RenderX

34. Deng L, Li P, Zhang X, Jiang Q, Turner D, Zhou C, et al. Risk of SARS-CoV-2 reinfection: a systematic review and meta-analysis. Sci Rep 2022 Dec 01;12(1):20763 [FREE Full text] [doi: 10.1038/s41598-022-24220-7] [Medline: 36456577]
35. Ren X, Zhou J, Guo J, Hao C, Zheng M, Zhang R, et al. Reinfection in patients with COVID-19: a systematic review. Glob Health Res Policy 2022 Apr 29;7(1):12 [FREE Full text] [doi: 10.1186/s41256-022-00245-3] [Medline: 35488305]
36. Wu X, Wang Z, He Z, Li Y, Wu Y, Wang H, et al. A follow-up study shows that recovered patients with re-positive PCR test in Wuhan may not be infectious. BMC Med 2021 Mar 15;19(1):77 [FREE Full text] [doi: 10.1186/s12916-021-01954-1] [Medline: 33715626]
37. O Murchu E, Byrne P, Carty PG, De Gascun C, Keogan M, O'Neill M, et al. Quantifying the risk of SARS-CoV-2 reinfection over time. Rev Med Virol 2022 Jan;32(1):e2260 [FREE Full text] [doi: 10.1002/rmv.2260] [Medline: 34043841]
38. Oyebode O, Ndulue C, Mulchandani D, Suruliraj B, Adib A, Orji FA, et al. COVID-19 pandemic: identifying key issues using social media and natural language processing. J Healthc Inform Res 2022 Jun 11;6(2):174-207 [FREE Full text] [doi: 10.1007/s41666-021-00111-w] [Medline: 35194569]
39. Altmann DM, Whettlock EM, Liu S, Arachchillage DJ, Boyton RJ. The immunology of long COVID. Nat Rev Immunol 2023 Oct 11;23(10):618-634. [doi: 10.1038/s41577-023-00904-7] [Medline: 37433988]
40. Al-Aly Z, Davis H, McCorkell L, Soares L, Wulf-Hanson S, Iwasaki A, et al. Long COVID science, research and policy. Nat Med 2024 Aug 09;30(8):2148-2164. [doi: 10.1038/s41591-024-03173-6] [Medline: 39122965]
41. Bourmistrova NW, Solomon T, Braude P, Strawbridge R, Carter B. Long-term effects of COVID-19 on mental health: a systematic review. J Affect Disord 2022 Feb 15;299:118-125 [FREE Full text] [doi: 10.1016/j.jad.2021.11.031] [Medline: 34798148]
42. Yu S, Abbas J, Draghici A, Negulescu OH, Ain NU. Social media application as a new paradigm for business communication: the role of COVID-19 knowledge, social distancing, and preventive attitudes. Front Psychol 2022;13:903082 [FREE Full text] [doi: 10.3389/fpsyg.2022.903082] [Medline: 35664180]
43. Aggarwal K, Singh SK, Chopra M, Kumar S. Role of social media in the COVID-19 pandemic: a literature review. In: Gupta BB, Peraković D, Abd El-Latif AA, Gupta D, editors. Data Mining Approaches for Big Data and Sentiment Analysis in Social Media. Boca Raton, FL: IGI Global; 2022:21-46.
44. Hao T, Huang Z, Liang L, Weng H, Tang B. Health natural language processing: methodology development and applications. JMIR Med Inform 2021 Oct 21;9(10):e23898 [FREE Full text] [doi: 10.2196/23898] [Medline: 34673533]
45. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc 2019 Apr 01;26(4):364-379 [FREE Full text] [doi: 10.1093/jamia/ocy173] [Medline: 30726935]
46. McMaster C, Chan J, Liew DF, Su E, Frauman AG, Chapman WW, et al. Developing a deep learning natural language processing algorithm for automated reporting of adverse drug reactions. J Biomed Inform 2023 Jan;137:104265 [FREE Full text] [doi: 10.1016/j.jbi.2022.104265] [Medline: 36464227]
47. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009 Oct;42(5):760-772 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.007] [Medline: 19683066]
48. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf Med 2017 Oct 26;56(5):377-389. [doi: 10.3414/ME17-01-0024] [Medline: 28816338]
49. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online October 11, 2018 [FREE Full text] [doi: 10.5260/chara.21.2.8]
50. Elbattah M, Arnaud É, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and applications. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies. 2021 Presented at: BIOSTEC '21; February 11-13, 2021; Virtual Event p. 825-832 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0010414508250832 [doi: 10.5220/0010414508250832]
51. Al-Moslmi T, Gallofre Ocana M, Opdahl AL, Veres C. Named entity extraction for knowledge graphs: a literature overview. IEEE Access 2020;8:32862-32881 [FREE Full text] [doi: 10.1109/ACCESS.2020.2973928]
52. Uncover the connection, explore COVID-19 cases through tweets. Covlab. URL: https://covlab.tech [accessed 2024-04-29]
53. Falotico R, Quatto P. Fleiss' kappa statistic without paradoxes. Qual Quant 2014 Feb 13;49(2):463-470. [doi: 10.1007/S11135-014-0003-1]
54. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. J Doc 1972 Jan;28(1):11-21. [doi: 10.1108/EB026526]
55. Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss. Mach Learn 1997;29(2):103-130. [doi: 10.1023/A:1007413511361]
56. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl 1998 Jul 10;13(4):18-28. [doi: 10.1109/5254.708428]
57. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Hoboken, NJ: John Wiley & Sons; 2013.
58. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv Preprint posted online July 26, 2019 [FREE Full text] [doi: 10.5260/chara.21.2.8]

XSL•FO
RenderX

59. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV. XLNet: generalized autoregressive pretraining for language understanding. arXiv Preprint posted online June 19, 2019 [FREE Full text]
60. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf [accessed 2024-04-29]
61. Workshop B, Scao TL, Fan A, Akiki C, Pavlick E, Ilić S, et al. BLOOM: a 176B-parameter open-access multilingual language model. arXiv Preprint posted online June 27, 2023 [FREE Full text]
62. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: open foundation and fine-tuned chat models. arXiv Preprint posted online July 18, 2023 [FREE Full text] [doi: 10.48550/arXiv.2307.09288]
63. Li Z, Li X, Liu Y, Xie H, Li J, Wang F, et al. Label supervised LLaMA finetuning. arXiv Preprint posted online October 02, 2023 [FREE Full text] [doi: 10.48550/arXiv.2310.01208]
64. Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. LoRA: low-rank adaptation of large language models. arXiv Preprint posted online June 17, 2021 [FREE Full text] [doi: 10.5260/chara.21.2.8]
65. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv Preprint posted online November 14, 2017 [FREE Full text] [doi: 10.5260/chara.21.2.8]
66. Li J, Sun A, Han J, Li C. A survey on deep learning for named entity recognition. IEEE Trans Knowl Data Eng 2022 Jan 1;34(1):50-70. [doi: 10.1109/TKDE.2020.2981314]
67. Bhatia P, Celikkaya B, Khalilia M, Senthivel S. Comprehend medical: a named entity recognition and relationship extraction web service. In: Proceedings of the 18th IEEE International Conference On Machine Learning And Applications. 2019 Presented at: ICMLA '19; December 16-19, 2019; Boca Raton, FL p. 1844-1851 URL: https://ieeexplore.ieee.org/document/8999113 [doi: 10.1109/icmla.2019.00297]
68. Schmitt X, Kubler S, Robert J, Papadakis M, LeTraon Y. A replicable comparison study of ner software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. In: Proceedings of the 6th International Conference on Social Networks Analysis, Management and Security. 2019 Presented at: SNAMS '19; October 22-25, 2019; Granada, Spain p. 338-343 URL: https://ieeexplore.ieee.org/document/8931850 [doi: 10.1109/snams.2019.8931850]
69. Chang E, Mostafa J. The use of SNOMED CT, 2013-2020: a literature review. J Am Med Inform Assoc 2021 Aug 13;28(9):2017-2026 [FREE Full text] [doi: 10.1093/jamia/ocab084] [Medline: 34151978]
70. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 1998 Mar 08;454(1971):903-995. [doi: 10.1098/rspa.1998.0193]
71. Benitez D, Gaydecki PA, Zaidi A, Fitzpatrick AP. The use of the Hilbert transform in ECG signal analysis. Comput Biol Med 2001 Sep;31(5):399-406. [doi: 10.1016/s0010-4825(01)00009-9] [Medline: 11535204]
72. Mushtaq R. Augmented Dickey Fuller test. SSRN Journal Preprint posted online August 17, 2011 [FREE Full text] [doi: 10.2139/ssrn.1911068]
73. Shin Y, Schmidt P. The KPSS stationarity test as a unit root test. Econ Lett 1992 Apr;38(4):387-392. [doi: 10.1016/0165-1765(92)90023-R]
74. Breitung J, Franses PH. On Phillips–Perron-type tests for seasonal unit roots. Econom Theory 1998 Apr 01;14(2):200-221. [doi: 10.1017/S0266466698142032]
75. Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M. TimesNet: temporal 2D-variation modeling for general time series analysis. arXiv Preprint posted online October 5, 2022 [FREE Full text] [doi: 10.48550/arXiv.2210.02186]
76. Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In: Cohen I, Huang Y, Chen J, Benesty J, editors. Noise Reduction in Speech Processing. Berlin, Germany: Springer; 2009:1-4.
77. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ 1998 Dec 05;317(7172):1572 [FREE Full text] [doi: 10.1136/bmj.317.7172.1572] [Medline: 9836663]

Abbreviations

AP: average precision
API: application programming interface
AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
BLOOM: Bigscience Large Open-science Open-access Multilingual language model
CDC: Centers for Disease Control and Prevention
CPD: COVID-19 patient database
HT: Hilbert transform
Llama-2: large language model meta AI 2
LLM: large language model
LoRA: low-rank adaptation
NLP: natural language processing

XSL•FO
RenderX

PCR: polymerase chain reaction
PR: precision-recall
RoBERTa: robustly optimized Bidirectional Encoder Representations from Transformers pretraining approach
ROC: receiver operating characteristic

Edited by A Mavragani; submitted 12.06.24; peer-reviewed by V Astha, M Elbattah, X Feng; comments to author 29.10.24; revised version received 09.12.24; accepted 16.01.25; published 20.02.25.

Please cite as:
Xie J, Zhang Z, Zeng S, Hilliard J, An G, Tang X, Jiang L, Yu Y, Wan X, Xu D
Leveraging Large Language Models for Infectious Disease Surveillance—Using a Web Service for Monitoring COVID-19 Patterns From Self-Reporting Tweets: Content Analysis
J Med Internet Res 2025;27:e63190
URL: https://www.jmir.org/2025/1/e63190
doi:10.2196/63190
PMID:

©Jiacheng Xie, Ziyang Zhang, Shuai Zeng, Joel Hilliard, Guanghui An, Xiaoting Tang, Lei Jiang, Yang Yu, Xiufeng Wan, Dong Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Stay-At-Home Mandates on Uncertainty and Sentiments: Quasi-Experimental Study

Carolina Biliotti[1], MSE; Nicolò Fraccaroli[2], PhD; Michelangelo Puliga[3], PhD; Falco J Bargagli-Stoffi[4,5*], PhD; Massimo Riccaboni[1,6*], PhD

[1]AXES, IMT School for Advanced Studies Lucca, Lucca, Italy

[2]W R Rhodes Center for International Economics and Finance, Brown University, Providence, RI, United States

[3]Bflows Network Lab, Cagliari, Italy

[4]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, United States

[5]Department of Biostatistics, Harvard University, Cambridge, MA, United States

[6]Istituto Universitario di Studi Superiori di Pavia, Pavia, Italy

[*]these authors contributed equally

Corresponding Author:
Nicolò Fraccaroli, PhD
W R Rhodes Center for International Economics and Finance
Brown University
111 Thayer Street, Box 1970
Providence, RI
United States
Phone: 1 401 863 2809
Email: nicolo_fraccaroli@brown.edu

Abstract

Background: As the spread of the SARS-CoV-2 virus coincided with lockdown measures, it is challenging to distinguish public reactions to lockdowns from responses to COVID-19 itself. Beyond the direct impact on health, lockdowns may have worsened public sentiment toward politics and the economy or even heightened dissatisfaction with health care, imposing a significant cost on both the public and policy makers.

Objective: This study aims to analyze the causal effect of COVID-19 lockdown policies on various dimensions of sentiment and uncertainty, using the Italian lockdown of February 2020 as a quasi-experiment. At the time of implementation, communities inside and just outside the lockdown area were equally exposed to COVID-19, enabling a quasi-random distribution of the lockdown. Additionally, both areas had similar socioeconomic and demographic characteristics before the lockdown, suggesting that the delineation of the strict lockdown zone approximates a randomized experiment. This approach allows us to isolate the causal effects of the lockdown on public emotions, distinguishing the impact of the policy itself from changes driven by the virus's spread.

Methods: We used Twitter data (N=24,261), natural language models, and a difference-in-differences approach to compare changes in sentiment and uncertainty inside (n=1567) and outside (n=22,694) the lockdown areas before and after the lockdown began. By fine-tuning the AlBERTo (Italian BERT optimized) pretrained model, we analyzed emotions expressed in tweets from 1124 unique users. Additionally, we applied dictionary-based methods to categorize tweets into 4 dimensions—economy, health, politics, and lockdown policy—to assess the corresponding emotional reactions. This approach enabled us to measure the direct impact of local policies on public sentiment using geo-referenced social media and can be easily adapted for other policy impact analyses.

Results: Our analysis shows that the lockdown had no significant effect on economic uncertainty (b=0.005, SE 0.007, t125=0.70; *P*=.48) or negative economic sentiment (b=–0.011, SE 0.0089, t125=–1.32; *P*=.19). However, it increased uncertainty about health (b=0.036, SE 0.0065, t125=5.55; *P*<.001) and lockdown policy (b=0.026, SE 0.006, t125=4.47; *P*<.001), as well as negative sentiment toward politics (b=0.025, SE 0.011, t125=2.33; *P*=.02), indicating that lockdowns have broad externalities beyond health. Our key findings are confirmed through a series of robustness checks.

Conclusions: Our findings reveal that lockdowns have broad externalities extending beyond health. By heightening health concerns and negative political sentiment, policy makers have struggled to secure explicit public support for government measures,

XSL•FO
RenderX

which may discourage future leaders from implementing timely stay-at-home policies. These results highlight the need for authorities to leverage such insights to enhance future policies and communication strategies, reducing uncertainty and mitigating social panic.

(J Med Internet Res 2025;27:e64667) doi:10.2196/64667

KEYWORDS

lockdown policy; sentiment analysis; uncertainty; social media; quasi-experiment

Introduction

Background

The spread of the SARS-CoV-2 virus and the associated COVID-19 disease led to an extraordinary rise in uncertainty and negative emotions, resulting in significant economic and social costs [1-4]. Against this backdrop, by April 2020, most governments had implemented lockdown measures to contain the virus's spread within their populations. Lockdown policies, or "stay-at-home" mandates, are temporary restrictions prohibiting residents from leaving their homes except for essential tasks or work in essential businesses. While recent evidence suggests that these measures are effective in slowing the spread of COVID-19 [5-7], their impact on public sentiment remains unclear. Lockdown measures may heighten public uncertainty and negative emotions by amplifying health concerns, shaping expectations of negative economic and social consequences, and triggering political backlash. Conversely, they could also signal a strong commitment to controlling the virus, reducing information asymmetry, and improving public sentiment.

Goal of the Study

This study investigates whether lockdown measures amplify or alleviate uncertainty and negative sentiment. By proposing a nuanced measure of uncertainty and sentiment, we assess the heterogeneous effects of lockdown measures across key public concerns, including economics, health, and politics. Understanding whether the economic and social costs of lockdowns outweigh their health benefits is essential for informed policy decisions.

Prior Work

Determining the causal effect of lockdowns on public sentiment is challenging due to simultaneity and endogeneity issues. In most countries, authorities implemented lockdown measures immediately after detecting the virus, meaning the decline in public sentiment caused by COVID-19 and the enforcement of lockdowns occurred simultaneously. Additionally, endogeneity arises because lockdowns were introduced in response to detecting COVID-19 cases.

These areas are likely the same ones where people perceive a higher risk of pandemic-related costs, resulting in lower sentiment. Given these limitations, most studies on sentiment changes during COVID-19 lockdowns provide correlational evidence [8-12], which may either overestimate or underestimate the impact of lockdowns on public emotions, potentially leading to inaccurate conclusions.

Previous studies reported increasing uncertainty and negativity during COVID-19 but did not determine the extent to which these trends were driven by restrictive measures [13,14].

Moreover, many studies examined overall changes in uncertainty and sentiment during the pandemic [15-17] or focused on specific aspects such as economic uncertainty [8,13], political polarization [18,19], or health-related negative emotions [17,20], without simultaneously analyzing multiple relevant dimensions of public discourse at the time.

Distinguishing the effects of lockdown measures across multiple topics, such as politics, economics, and health, is valuable, as their impact may vary across different dimensions.

Our Study

To address this, we focus on the Italian lockdown of February 2020, the first in a high-income economy. This case allows for a quasi-experimental analysis, as the lockdown was initially imposed on communities where the first COVID-19 cases were detected, effectively creating a form of random assignment.

At the time of implementation, virus transmission rates were similar inside and outside the lockdown areas [21], and health and socioeconomic conditions were comparable, enabling the application of causal inference methods to assess the impact of lockdowns on public opinion. Notably, the timing of these lockdowns—before vaccines were available and with limited knowledge about the SARS-CoV-2—provides a unique opportunity to isolate and measure the direct influence of lockdown measures on public emotions, particularly sentiments and uncertainty.

We use Twitter (X) data to measure uncertainty and sentiment by applying deep learning and natural language processing techniques. To assess uncertainty, we differentiate between tweets that ask questions, express hesitation, irresolution, confusion, or anxiety, and those that seek clarity, as opposed to tweets conveying assurance and confidence. In measuring sentiment, we aim to capture emotions such as anger, disillusionment, and disapproval, which we broadly classify as negative sentiments. Additionally, we distinguish between tweets from users inside the lockdown area (the so-called red zone) and those from users outside, including neighboring cities (referred to as the "orange zone").

This data collection is conducted both before and after the lockdown. To measure uncertainty and negative sentiments in each tweet, we fine-tune the AlBERTo (Italian BERT optimized) model [22], a natural language processing model based on the Bidirectional Encoder Representations from Transformers (BERT) framework [23], specifically designed for Italian text data. BERT efficiently contextualizes words, providing

XSL•FO
RenderX

significant advantages over traditional bag-of-words approaches for measuring emotions [24,25]. It has demonstrated strong performance across various natural language processing tasks, including sentiment analysis, fake news detection, and the analysis of public opinions on Twitter/X during the COVID-19 pandemic [15,26-29]. Given the existing evidence on the accuracy of text analysis using BERT-based models, we believe that our sentiment and uncertainty measures constructed via AlBERTo reliably capture the emotions expressed in the tweets.

To better understand the impact of the lockdown on public emotions, we categorize uncertainty and sentiments into 4 main dimensions: economy, health, politics, and lockdown policy. Identifying health- and economy-related tweets is essential for analyzing how lockdown policies differently influenced perceptions of health and economic risks, as well as for investigating the health-economic trade-off [30]. The political dimension is also considered, as lockdowns could shape sentiments toward politics, affect attitudes toward incumbent politicians [31-35], or heighten political polarization [18,19]. The fourth dimension, which we refer to as policy for simplicity, captures the uncertainty and negative sentiments associated with the behavioral guidelines of the restrictions. This distinction allows us to determine whether uncertainty and negative sentiments arise from concerns about the authorities implementing the policy (political uncertainty) or from confusion regarding the policy's specifics (uncertainty around lockdown policies). Overall, this categorization provides a clearer understanding of how public sentiment evolved in response to the COVID-19 pandemic and government actions.

We use manually curated dictionaries for each dimension to categorize the tweets and construct a document-feature matrix, enabling the identification of words associated with each dimension in the tweets.

Using a difference-in-differences (DiD) specification [36], we estimate the causal impact of the lockdown on economic, health, political, and policy-related uncertainty and negative sentiment. Causal identification is ensured as the treatment assignment—the lockdown policy—was an exogenous shock to public uncertainty and negative sentiment, independent of potential outcomes. Our findings are further validated through a placebo test and a series of robustness checks.

Our findings indicate that, in the Italian context, the lockdown had no significant effect on public concerns about the economy, suggesting that lockdown-induced negative emotions did not heighten economic worries. Economic concerns did not increase among those subjected to the lockdown, which is relevant to the ongoing debate on the economic impact of lockdowns. While stay-at-home mandates have economic consequences [37], these effects likely stem from other, more direct channels, such as business closures and activity restrictions.

By contrast, the stay-at-home mandate significantly influenced health- and politics-related emotions among individuals within the lockdown area. Users in this zone expressed higher uncertainty when discussing health and lockdown policies and displayed more negative sentiments about politics compared with those outside the lockdown area. Rather than reassuring the public about the authorities' commitment to combating the virus, the lockdown heightened concerns about health-related risks. However, this increased uncertainty may have fostered greater awareness and compliance with containment measures, potentially leading to positive public health outcomes [38].

The rise in negative sentiments toward politics represents a significant cost that policy makers must take into account. These political costs may discourage elected officials from implementing lockdowns in future pandemics, regardless of their effectiveness in controlling virus transmission. The increased political discontent could indicate a rise in political polarization, which intensified during COVID-19 [18,19]. However, it may also reflect shifting attitudes toward the incumbent politician, for which existing evidence remains mixed [31-33,39]. The extent to which changes in political attitudes reported in the literature can be directly attributed to lockdown measures remains unclear. Our findings demonstrate that when the effect of the lockdown is causally identified and isolated from other factors influencing emotions, the policy increases negative sentiments toward politics. This shift in sentiment could lead to delays, restrictions to narrower regions, or weaker policy responses when lockdown measures are necessary.

As our study focuses on a specific case, the generalizability of our results is limited. Our findings should be interpreted as reliable causal estimates of the effect of the first Italian lockdown on the red zone units rather than as a universally applicable average across different contexts and periods. By prioritizing internal validity over external validity and providing extensive evidence supporting our empirical strategy, we isolate the direct effect of the lockdown while controlling for comoving factors, particularly exposure to COVID-19.

Our study makes a valuable and unique contribution by quantifying the direct emotional response to the extraordinary and strict public health policies of 2020. It introduces a clear methodology for analyzing geo-referenced social media data to evaluate the causal impact of policy at a local scale. This research significantly advances the current state of the art in social media analysis for studying the profound social and emotional changes brought about by the pandemic (see, eg, [40-43]).

Methods

Identification Strategy

Following the discovery of the first COVID-19 transmission case in the small town of Codogno, the Italian government announced a decree on February 22 imposing strict quarantine measures in Codogno and 9 neighboring municipalities in the province of Lodi (Castiglione d'Adda, Casalpusterlengo, Fombio, Maleo, Somaglia, Bertonico, Terranova dei Passerini, Castelgerundo, and San Fiorano), effective February 23, 2020 [44]. The lockdown was strictly enforced with police presence on the streets, prohibiting entry into or exit from the restricted area.

On March 8, an "orange zone" (*zona arancione*) was established, encompassing municipalities near the red zone. Lombardy and 14 other cities outside the region (Modena, Parma, Piacenza, Reggio nell'Emilia, Rimini, Pesaro e Urbino, Alessandria, Asti,

XSL•FO
RenderX

Novara, Verbano-Cusio-Ossola, Vercelli, Padua, Treviso, and Venice), accounting for 16 million people, were placed under *partial* lockdown restrictions [45]. In the orange zone, residents were *invited* to avoid traveling in and out of their municipality of residence, while economic activities were allowed to continue. On March 9, these measures were extended nationwide, effectively ending the strict *red zone* restrictions in the province of Lodi [46]. A full nationwide lockdown was announced on March 22.

We leverage the exogenous shock of the unexpected lockdown measures enacted in Italy on February 23, 2020, to assess the causal impact of lockdown restrictions on public emotions. We argue that the detection of the first European COVID-19 case in the municipality of Codogno, rather than in other nearby municipalities (Figure 1), was purely coincidental. When physicians discovered that the patient, who exhibited symptoms consistent with COVID-19, had been in contact with a friend recently returning from a business trip to China, they decided to override the existing protocol, which restricted testing to Italians and foreigners who had traveled to China [47]. As a result of this first positive case, hospitals in the Codogno area began testing individuals for COVID-19. This decision was driven solely by the identification of the first patient in Codogno.

Figure 1. The red zone (lockdown enforced on February 23, 2020, in Lombardy, Italy) and the surrounding selected control municipalities included in the analysis are shown in orange. We found tweets from 8 out of 10 cities in the red zone. The control area consists of municipalities surrounding Codogno within a 42 km radius. The area we analyzed is homogeneous in terms of demographics, socioeconomics, and virus exposure.

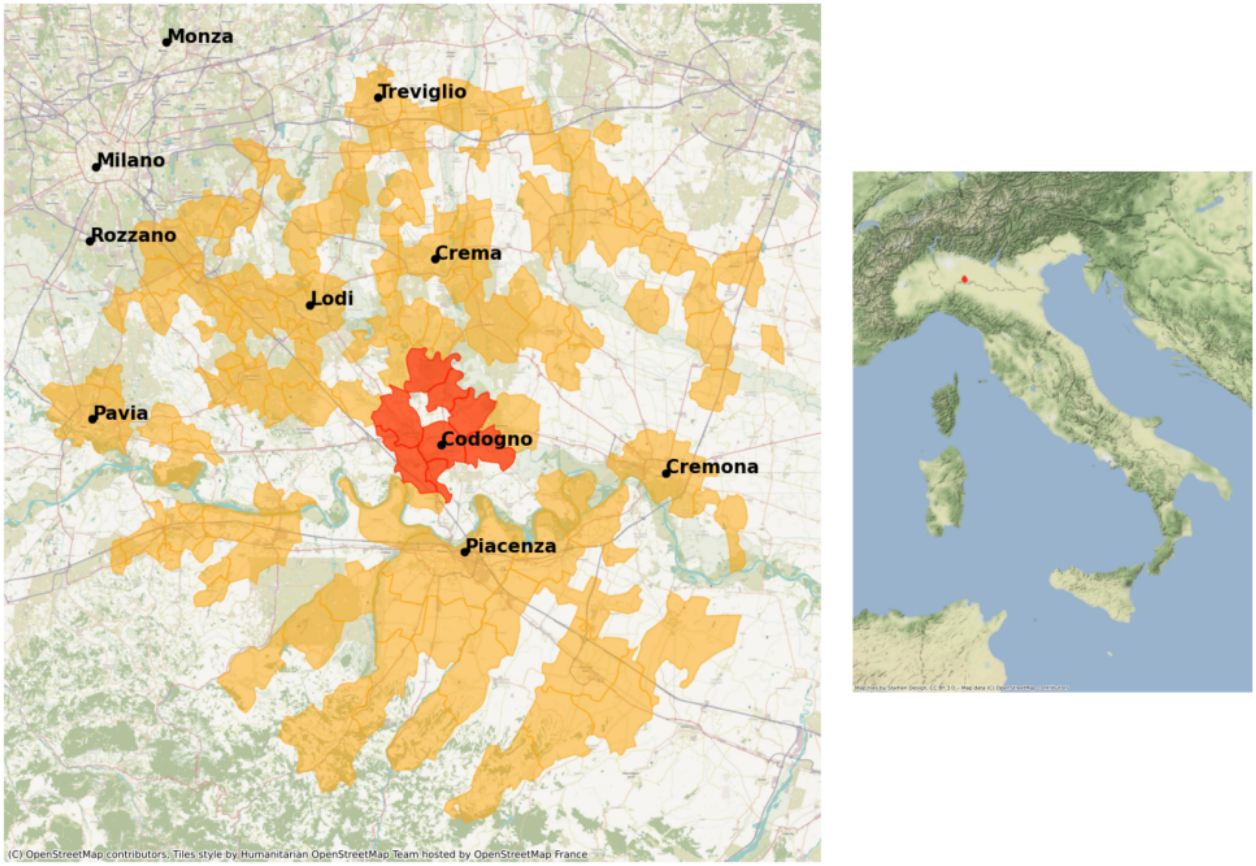

Recent retrospective epidemiological studies on the transmission risk of coronavirus in Lombardy in February 2020 revealed that the virus had a homogeneous transmission potential across different provinces of Lombardy at the time the first case was discovered in Codogno [21]. This indicates that COVID-19 incidence was balanced between the red zone and the neighboring municipalities in the orange zone during the initial lockdown. While the red zone was subjected to strict lockdown measures, nearby locations in the orange zone were not. Consequently, individuals in the orange zone serve as a suitable control group for assessing the impact of lockdown measures on the emotions of those in the red zone.

Causal identification is achieved through the exogenous shock represented by the lockdown policy, allowing us to rule out selection bias among the units under lockdown, which are balanced and comparable to the control group. To establish a quasi-experimental approach, we designate municipalities subjected to the lockdown as the treated group, specifically the "red zone," which was at the epicenter of the COVID-19 outbreak. The control group consists of municipalities surrounding Codogno, designated as the "orange zone." To ensure comparability, only the closest nonurban areas to the red zone were selected.

Figure 2 [48-52] demonstrates that the complete randomization of lockdown assignment holds for municipality-level prelockdown covariates. The quantiles represent the acceptance region of our randomization test with α=.15, using the standardized covariate mean differences as the test statistic. In each iteration, the lockdown is randomly assigned to units, and the standardized mean difference is calculated. The observed

XSL•FO
RenderX

standardized differences fall well within the acceptance bounds [53], confirming that complete randomization holds for all considered covariates. This finding supports the notion that the 2 areas had comparable pretreatment demographic and socioeconomic characteristics, reinforcing the validity of the random allocation.

Figure 2. The standardized covariate mean difference was calculated along with the 7.5% and 92.5% complete randomization quantiles using 2000 permutations. Covariate balance is assessed between the red zone and the control municipalities of the orange zone, matching the user location of tweets featured in the main analysis (8 red zone and 118 orange zone municipalities). We use ISTAT data [48-52] on social, economic, and demographic characteristics, including the number of residents as of January 1, 2020; the average number of deaths in January, February, and March from 2015 to 2019; the share of residents in occupational categories, age cohorts, and education levels. Additionally, we collected data on industry and services from 2017, including total output and value added (in euros), number of employees, total staff, and local units. Missing values for industry and services are imputed using the group mean for 2 treated cities (Castelgerundo, 1473 residents, and Bertonico, 1059 residents). Similarly, missing entries for monthly average deaths are imputed for 2 control municipalities (Calendasco, 2409 residents, and Cerro all'Ambro, 5149 residents) and 1 treated city (Bertonico, 1059 residents). Before testing, we standardize covariates related to mean mortality, industry, services, and total residents.

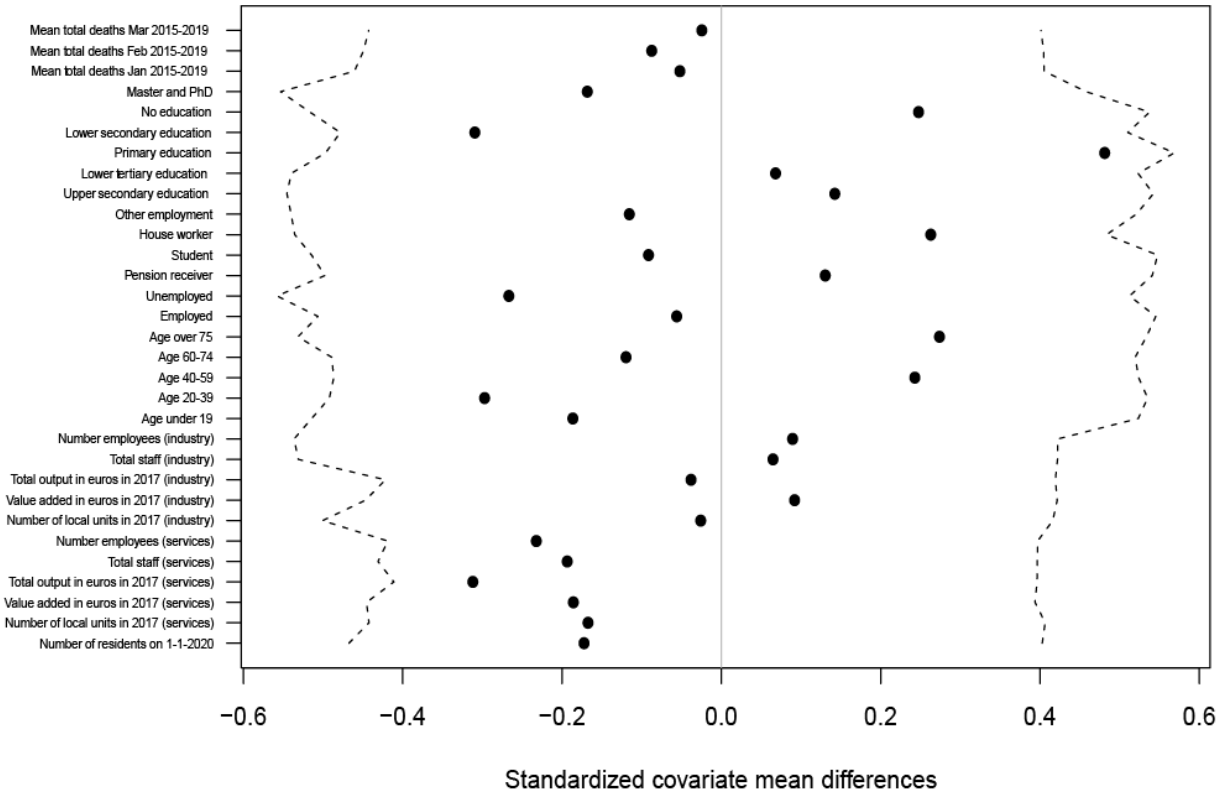

Moreover, no other major economic or political events occurred in the area during the sample period. The region, while affected by the lockdown, is a relatively less significant hub in the local economy of Lombardia compared with urban centers such as Bergamo, Cremona, or Lodi, as shown in Table S3 in Multimedia Appendix 1. This reduces the likelihood of bias from unaccounted external events influencing public reactions.

Additionally, the influence of news outlets and media coverage on reactions in the red zone was limited, as no local media were operating within the red zone, and national media had no direct access to it. As a result, individuals inside and just outside the red zone were exposed to the same partial and limited media coverage of the lockdown policies, ensuring comparable information exposure between the 2 groups.

Our study design provides strong support for causal identification of the lockdown's treatment effect, as the orange zone serves as a suitable counterfactual for the red zone. As the control observations share similar demographic, geographic, social, and economic characteristics and were exposed to the virus under comparable conditions, trends in overall and topic-specific uncertainty and negative sentiment should exhibit similar patterns between the treated and control groups in the absence of the lockdown intervention.

Data Collection

For our main analysis, we collected Italian tweets before and after the lockdown, both inside and outside the red zone. Data collection took place from December 1, 2019, to March 22, 2020, using the Twitter Stream API (application programming interface). The data set includes information on user activity, user-defined locations, and the content of individual tweets.

Twitter-based indices have proven to be effective tools for understanding people's emotional well-being during the COVID-19 pandemic [8,54]. Social media–based metrics offer a powerful means to accurately identify uncertainty and negative sentiments following major events such as COVID-19 [55]. As public reactions to crises tend to be similar among Twitter users and nonusers, emotions expressed on social media can serve as a representative proxy for the broader population's sentiments [56].

Tweets from the municipality of Codogno and the surrounding areas are collected using the Twitter Stream API with 2 filters:

XSL•FO
RenderX

Italian language and geographical location. The latter allows us to retrieve tweets with GPS coordinates within a maximum radius of 42 km from a target center—Codogno in our case. This enables us to extract tweets from the "red zone" and the "orange zone" near Codogno. We then rely on the locations self-reported by Twitter users to assign tweets to municipalities inside or outside the red zone. These self-reported locations help fine-tune user locations without relying solely on GPS coordinates to distinguish between people inside and outside the red zone, given that movement between the red and orange zones was prohibited. For an overview of Twitter's location-based methods, see [57].

The text of self-reported user locations is cleaned and matched with ISTAT [58] data on the geographical coordinates of Italian municipalities. Additionally, the self-reported locations of tweets were manually checked to verify whether the user's location aligned with the tweet's content (eg, tweets indicating that Codogno users were affected by the lockdown).

We clean the text of tweets by removing URLs, hashtags, and tags and eliminate duplicate tweets—those with the same cleaned text from the same users.

On March 8, all control units (areas not yet subject to lockdown measures) entered a partial lockdown, 1 day before the nationwide expansion of partial lockdown measures on March 9, 2020. Any potential anticipation of the February 23 policy, around the time of the first detected COVID-19 transmission case in Codogno (February 20, 2020), could bias the estimate of the policy's impact. To mitigate this, we deleted observations from February 20 to 22, 2020, and removed all tweets from March 7, 2020—the day the first extension of restrictions was announced. Finally, we removed all users who were information sources or business accounts.

As a result, we obtained a sample of 24,261 unique tweets—1567 from 8 of the 10 lockdown municipalities in the red zone and 22,694 from 118 unique locations within the orange zone. These tweets were posted by 1124 unique users (60 from the red zone and 1064 from the orange zone). Figure 1 highlights the cities with at least one active geo-referenced user account included in our analysis. We observed no tweets from 2 municipalities in the red zone: San Fiorano (1849 inhabitants) and Terranova dei Passerini (731 inhabitants). Additionally, a larger sample of tweets from across Italy (774,407 tweets) was collected during the same period using language filtering and keyword-based queries related to COVID-19. We used these tweets for our placebo analyses, detailed in the "Placebo Test" section in Multimedia Appendix 1. Using the geographical coordinates of municipalities from ISTAT [58], we identified additional unique tweets from the red zone and new municipalities in the orange zone near Codogno. These latest observations were added to the sample of tweets collected around Codogno.

The analysis is based on selected tweets from the red zone and surrounding areas. This stems from an extensive process of data processing and cleaning, which inevitably reduces the number of useful observations. In this paper, we prioritize the internal validity of our estimates over the generalizability of the results to other contexts by focusing on a single case study for credible causal inference. This approach necessitates conservative choices in selecting the subset of tweets included in the analysis to ensure the precision of both sentiment analysis and spatial mapping. Overall, the retrieved sample of users from the red and orange zones is reasonably sized, given that the geographical area covered by the policy is limited, as is its population, as shown in Table S3 in Multimedia Appendix 1.

Ethical Considerations

The data are anonymized by removing tweets and users' original identifiers, as well as users' original names, and replacing them with new, deidentified user identifiers. No ethics review was sought because the study only explored publicly available data on social media and did not conduct any experiments on humans.

Uncertainty and Sentiment Classifier

We fine-tune the pre-trained AlBERTo model [22], a deep learning natural language model trained on a large corpus of Italian tweets (~200 million, TWITA data set), to classify tweets into uncertainty and sentiment categories.

We refine the pretrained AlBERTo model in a 2-step process by manually labeling tweets for training (n=6318). The manual labels indicate degrees of uncertainty (neutral, uncertainty, and certainty) and sentiment (neutral, negative, and positive). The manually assigned labels were independently validated and verified by the authors. In the first step, we distinguish between neutral and nonneutral tweets. In the second step, we further classify nonneutral tweets as either uncertainty or certainty and as having positive or negative sentiment.

After fine-tuning the model and predicting the labels for all tweets, we obtain 2 binary variables: uncertainty and negative sentiment, assigned to each tweet in our data set. The uncertainty variable is set to 1 if the tweet expresses uncertainty—such as concern, questions, or uneasiness—and 0 otherwise, meaning the text conveys a different emotional state, such as certainty or indifference. Similarly, the negative sentiment variable is set to 1 if the tweet expresses negative feelings and 0 if it does not—that is, if it conveys positivity or neutrality. Multimedia Appendix 2 presents word clouds displaying the 100 most frequent terms in tweets classified as expressing uncertainty and negative sentiment.

Topics Classifier

A dictionary-based classifier is used to identify topic-specific words in tweets (see Tables S1 and S2 of the "Topic Dictionaries" section in Multimedia Appendix 1 for the complete list of words used to construct the topic-specific dictionaries). We selected 4 categories likely to be influenced by the policy: health, economy, politics, and policy. For each topic, we created a binary variable indicating whether a tweet contains at least one term from the topic-specific dictionary. Overlap of topic labels within a single tweet is allowed, as tweets can cover multiple topics simultaneously. In the "Top Fifty Tweets" section in Multimedia Appendix 1, we provide a sample of tweets with the highest Shannon entropy by emotion-topic pair and discuss the performance of the machine learning classifier in identifying tweets related to each topic and emotion. Table 1 presents the number of tweets in our sample categorized as

XSL•FO
RenderX

uncertainty or negative sentiment, aggregated and grouped by topic.

Table 1. The number of tweets classified as uncertainty and negative sentiment aggregated and grouped by topic[a] (N=24,261).

Classification	Value, n (%)	Percentile
Uncertainty[b]		
Aggregated	5270 (21.72)	89
Economics	369 (1.52)	99
Health	1063 (4.38)	97
Politics	301 (1.24)	99
Policy	304 (1.25)	99
Negative sentiment[b]		
Aggregated	8451 (34.83)	82
Economics	497 (2.05)	98
Health	1029 (4.24)	97
Politics	738 (3.04)	98
Policy	253 (1.04)	99

[a]We have omitted the category of Leisure from the table: Uncertainty Leisure, n (%)=3498 (14.4), Negative Sentiment Leisure, n (%)=6248 (25.7).

[b]The other categories, that is, Neutral Uncertainty (Neutral Sentiment) and Certainty (Positive Sentiment), are not reported.

Difference-in-Differences Model

To assess the causal impact of the lockdown on sentiment and uncertainty, we estimated a linear DiD model using repeated cross-sections of tweets from inside and outside the red zone before and after the February lockdown.

In the DiD model, we estimated the difference in the average change in emotion over time between the treated group (red zone) and the control group (orange zone). The estimation followed 3 steps: (1) comparing average outcomes between the treatment and control groups before the lockdown, (2) comparing average responses after the lockdown, and (3) estimating the difference. The first step accounts for preexisting differences, the second captures postlockdown differences, and the third measures the differential effect of the policy in the red zone, effectively isolating and removing time effects, common time trends, and time-invariant group differences.

For this measure to reliably estimate the treatment effect, the empirical framework must satisfy the key assumption of parallel trends, which states that emotions in the treated group would have followed the same trend as in the control group in the absence of treatment. This implies no inherent systematic differences between the treatment and control groups before the intervention. If this assumption is violated, the treatment effect may be biased. This condition is tested in the "Preexisting Trends" section in Multimedia Appendix 1. Another potential source of bias is treatment anticipation, as units might adjust their behavior in expectation of the lockdown. This is unlikely in our case, as this was the first lockdown in Europe. Another potential source of bias is spillover effects, where emotions in the red zone after policy implementation could influence those in the control group. However, because movement between the red and orange zones was prohibited, spillovers should not be relevant. We test this condition in the "SUTVA" section in Multimedia Appendix 1. Under these assumptions, the DiD estimate remains an unbiased measure of the lockdown's causal effect on the treated units. We conducted extensive tests to validate the DiD assumptions in our quasi-experimental setting and to ensure that our statistical method was appropriate for the observational sample. This is essential for credibly interpreting the model's estimates as the causal effect of lockdown on emotions in tweets from the red zone. The tests are summarized later in the text and detailed in the "Testing DiD Assumptions" section of Multimedia Appendix 1. As noted earlier, this case study has limitations regarding the generalizability of results and data expectations. However, by focusing on the red zone established by the February 23, 2020, policy and the selected control sample, we derive unbiased conclusions and obtain a causal estimate as if the policy had been implemented in a randomized experiment.

We define a treatment status indicator, red zone, which equals 1 if a tweet is from the red zone and 0 otherwise. We also define a pre- and postlockdown indicator, post, where post=0 for tweets posted before the lockdown and post=1 for those posted after. As all suitable control units were eventually treated by the partial lockdown on March 9, 2020, we lost our control set. To address this, we use post=2 to identify tweets published from March 9 onward. For each tweet, we define a set of binary outcome variables indicating uncertainty and negative sentiment, either aggregated or by topic. In the DiD regression model, the coefficient of interest is post=1 × red zone=1, which represents the treatment effect—the average impact of the lockdown on the probability of observing negative reactions in tweets from the red zone. Our sample does not allow for longitudinal user analysis, as different users are present across different periods. Figure 3 provides a graphical illustration of our empirical strategy.

XSL•FO
RenderX

Figure 3. Illustrative summary of the methodology. Twitter posts are filtered using global positioning system (GPS) coordinates to select tweets posted within 42 km of Codogno. Each tweet's user location is then matched to geo-referenced ISTAT data to further filter municipalities. Sentiment analysis is performed on the retrieved set of tweets to classify the sentiment and uncertainty expressed in the text. Additionally, tweets are categorized into major topics: Health, Economics, Politics, and Policy. A difference-in-differences (DiD) approach is used to estimate the average effect of the lockdown on the red zone's emotions, measured as the difference between the average change over time (before and after 2020) in the red zone's reactions and the average change over time among the control units in the orange zone. AlBERTo: Italian BERT optimized.

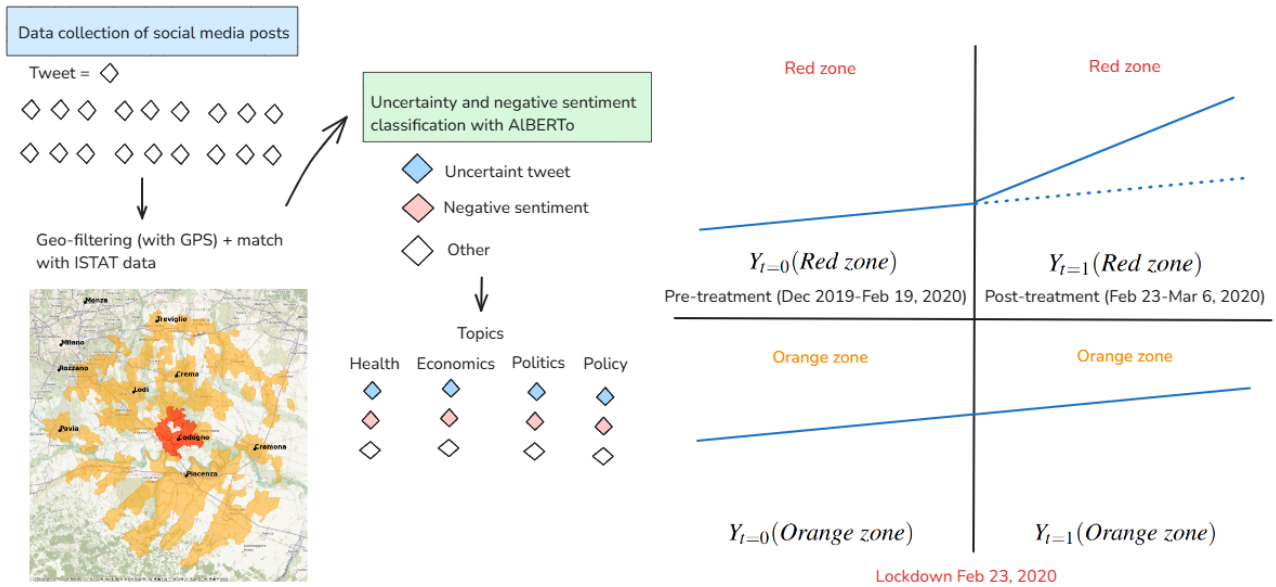

Results

Overview

The DiD model's estimates of the COVID-19 shock's impact on uncertainty and negative sentiment—both aggregated and by topic (economics, health, politics, and lockdown policy)—are presented in Tables 2 and 3, respectively. SEs are clustered at the municipality level (126 clusters) using the Liang-Zeger formula [59]. Model estimates with white SEs are provided in the "Model Estimates With Heteroskedasticity Robust Standard Errors" section in Multimedia Appendix 1 (also see Table S10 in Multimedia Appendix 1).

Table 2. Difference-in-differences regression table for uncertainty, aggregated and grouped by topics with user-level fixed effects.

Variable	Outcome 1: Aggregate	Outcome 2: Economics	Outcome 3: Health	Outcome 4: Politics	Outcome 5: Policy
	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value
post=1	0.02 (0.01), -3×10^{-6} to 0.04; .05	–0.0006 (0.002), –0.005 to 0.003; .8	0.037 (0.005), 0.027 to 0.046; <.001	–0.0077 (0.004), -0.017 to 0.001; .09	0.01 (0.002), 0.007 to 0.014; <.001
post=2	0.022 (0.017), –0.01 to 0.056; .21	0.004 (0.002), -6×10^{-4} to 0.01; .08	0.042 (0.005), 0.03 to 0.052; <.001	–0.006 (0.004), –0.016 to 0.003; .19	0.01 (0.0025), 0.005 to 0.15; <.001
Red zone × post=1	0.15 (0.03), 0.09 to 0.22; <.001	0.005 (0.007), –0.01 to 0.02; .48	0.036 (0.0065), 0.023 to 0.05; <.001	0.012[b] (0.0058), 4×10^{-4} to 0.023; .04	0.026 (0.006), 0.015 to 0.038; <.001
Red zone × post=2	0.03 (0.05), –0.07 to 0.13; .55	–0.004 (0.01), –0.02 to 0.016; .70	–0.023 (0.01), –0.04 to –0.0037; .02	0.0138 (0.007), –0.012 to 0.015; .84	0.005 (0.011), –0.02 to 0.028; .65
Constant	0.95 (0.047), 0.85 to 1.04; 2<.001	–0.0008 (0.01), –0.02 to 0.018; .93	–0.019 (0.008), –0.035 to –0.002; .03	0.005 (0.005), –0.005 to 0.015; .34	–0.015 (0.011), –0.04 to 0.007; .18
Observations, n	24,261	24,261	24,261	24,261	24,261

[a]SEs are clustered at the municipality level.

[b]The parallel trends condition fails to be satisfied for uncertainty related to politics. See the "Robustness Checks" section and Multimedia Appendix 1 for complete analysis.

XSL•FO
RenderX

Table 3. Difference-in-differences regression table for negative sentiment, aggregated and grouped by topic with user-level fixed effects.

Variable	Outcome 1: Aggregate	Outcome 2: Economics	Outcome 3: Health	Outcome 4: Politics	Outcome 5: Policy
	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value	Estimate (SE[a]), 95% CI; *P* value
post=1	–0.044 (0.022), –0.09 to -5×10^{-4}; .047	–0.004 (0.005), –0.015 to 0.006; .42	0.036 (0.0063), 0.024 to 0.05; <.001	–0.015 (0.008), –0.03 to 0.001; .06	0.0058 (0.0017), 0.002 to 0.01; <.001
post=2	–0.056 (0.023), –0.10 to –0.01; .02	0.003 (0.0057), –0.007 to 0.015; .52	0.038 (0.0057), 0.026 to 0.05; <.001	–0.02 (0.008), –0.037 to –0.003; .02	0.0066 (0.0018), 0.003 to 0.01; <.001
Red zone × post=1	–0.072[b] (0.027), –0.125 to –0.018; .008	–0.011 (0.0089), –0.03 to 0.006; .19	0.027 (0.021), –0.015 to 0.07; .21	0.025 (0.011), 0.003 to 0.05; .02	0.024 (0.02), –0.015 to 0.064; .22
Red zone × post=2	–0.049 (0.049), –0.146 to 0.047; .31	–0.006 (0.0126), –0.03 to 0.019; .63	0.0048 (0.022), –0.038 to 0.048; .83	0.028 (0.011), 0.006 to 0.05; .01	0.018 (0.015), –0.012 to 0.048; .23
Constant	0.105 (0.043), 0.02 to 0.19; .01	0.0024 (0.011), –0.02 to 0.024; .83	–0.042 (0.021), –0.084 to –0.001; .046	–0.008 (0.007), –0.02 to 0.01; .25	–0.024 (0.015), –0.05 to 0.005; .10
Observations, n	24,261	24,261	24,261	24,261	24,261

[a]SEs are clustered at the municipality level.

[b]The parallel trends condition fails to be satisfied for aggregated negative sentiment. See the "Robustness Checks" section and Multimedia Appendix 1 for complete analysis.

We first examine the effect of the lockdown on aggregate uncertainty. In column 1 of Table 2, the treatment effect (red zone=1 × post=1) is positive and statistically significant at .001 (b=0.15, SE 0.03, t_{125}=4.82; *P*<.001). The estimated time trend shows no significant increase in uncertainty following the lockdown (post=1, b=0.02, SE 0.01, t_{125}=1.98; *P*=.05) or after the implementation of partial nationwide measures (post=2, b=0.022, SE 0.017, t_{125}=1.27; *P*=.21). Overall, the results indicate that the lockdown significantly increased aggregate uncertainty among affected individuals (b=0.15, SE 0.03, t_{125}=4.82; *P*<.001). However, these findings may not capture the full picture, as multiple factors could influence the aggregate expression of emotions.

Column 2 of Table 2 presents the effects of the lockdown on economic uncertainty. The lockdown had no significant impact on economic uncertainty among those affected, with a small positive estimated effect (b=0.005, SE 0.007, t_{125}=0.70; *P*=.48). This suggests that the lockdown did not substantially influence economic uncertainty.

Column 3 of Table 2 shows that the lockdown significantly increased concern about health problems among individuals in the lockdown area (b=0.036, SE 0.0065, t_{125}=5.55; *P*<.001). As expected, after the partial nationwide measures on March 9, people in the red zone expressed significantly less uncertainty in health discussions (b=–0.023, SE 0.01, t_{125}=–2.35; *P*=.02) as restrictions eased. However, the overall trend of health-related uncertainty increased significantly following both the local lockdown (b=0.037, SE 0.005, t_{125}=7.50; *P*<.001) and the partial nationwide measures (b=0.042, SE 0.0052, t_{125}=8.00; *P*<.001).

The results for political uncertainty are presented in column 4 of Table 2. The lockdown significantly increased political uncertainty in the treated area (b=0.012, SE 0.0058, t_{125}=2.05; *P*=.04). However, this effect disappears and is no longer statistically different from 0 once all units are subjected to the same nationwide measures (red zone=1 × post=2, b=0.00138, SE 0.007, t_{125}=0.20; *P*=.84).

Column 5 in Table 2 presents the results for uncertainty associated with the lockdown policy itself. The lockdown significantly increased concerns among those affected by the measures (b=0.026, SE 0.006, t_{125}=4.47; *P*<.001). Additionally, the coefficients for post=1 (b=0.01, SE 0.002, t_{125}=5.56; *P*<.001) and post=2 (b=0.01, SE 0.0025, t_{125}=3.98; *P*<.001) indicate that uncertainty regarding the practical effects of the lockdown remained significantly higher than the baseline value.

DiD estimates of the lockdown's effect on aggregate negative sentiment are reported in Table 3 (column 1). The lockdown significantly reduced negative sentiment among those in the red zone (b=–0.072, SE 0.027, t_{125}=–2.67; *P*=.008). Over time, the likelihood of expressing negative sentiment decreases for both the treated and control groups, as indicated by the negative and statistically significant coefficients for post=1 (b=–0.044, SE 0.022, t_{125}=–2.00; *P*=.047) and post=2 (b=–0.056, SE 0.023, t_{125}=–2.45; *P*=.02). Overall, the results suggest that the lockdown significantly reduced aggregate negative sentiment among people in the red zone.

In Table 3 (column 2), we present the effect of the lockdown on negative sentiment about the economy. Although not significantly different from 0, the estimated impact on negative sentiment in economic discussions is negative (b=–0.011, SE 0.0089, t_{125}=–1.32; *P*=.19) and relatively larger in magnitude compared with the estimated effect on economic uncertainty in Table 2 (column 2).

The regression results in Table 3 (column 3) show that the lockdown had no significant effect on negative sentiment regarding health in the lockdown area (b=0.027, SE 0.021, t_{125}=1.27; *P*=.20). However, the overall trend indicates a significant worsening of sentiment about health issues both

XSL•FO
RenderX

inside and outside the lockdown area, as reflected in the coefficients for post=1 (b=0.036, SE 0.0063, t_{125}=5.70; P<.001) and post=2 (b=0.038, SE 0.0057, t_{125}=6.57; P<.001).

Column 4 in Table 3 presents the effect of the lockdown on negative political sentiment. We estimate a positive and significant coefficient for both interactions, indicating that political sentiment became more negative among treated users when the first lockdown was implemented (b=0.025, SE 0.011, t_{125}=2.33; P=.02) and after the nationwide measures (b=0.028, SE 0.011, t_{125}=2.50; P=.01). Notably, the general trend in negative political sentiment decreased significantly following the partial national lockdown (b=–0.02, SE 0.008, t_{125}=–2.30; P=.02).

In Table 3 (column 5), the red zone did not express negative sentiments about the policy significantly more often (b=0.024, SE 0.02, t_{125}=1.22; P=.22). However, negative sentiment toward the policy increased across all areas, both after the first lockdown (b=0.0058, SE 0.0017, t_{125}=3.39; P<.001) and after the nationwide measures (b=0.0066, SE 0.0018, t_{125}=3.60; P<.001).

Overall, our results highlight an important distinction between the impact of the lockdown on economic uncertainty and economic sentiment. While economic uncertainty increased due to the spread of COVID-19 cases (as shown in [24,60]), our case study indicates that the lockdown itself did not exacerbate economic uncertainty or negative sentiment. As shown in Multimedia Appendix 2, tweets during this period were less focused on purely economic discussions: the most frequently used terms in tweets expressing uncertainty or negative sentiment were unrelated to the economy. Although the estimated coefficient for the effect of the lockdown on economic uncertainty is very close to 0 and relatively small, we cannot rule out the possibility that the lack of statistical significance for the relatively larger effect of lockdowns on negative sentiment about the economy (b=–0.011, SE 0.0089, t_{125}=–1.32; P=.19) is due to the small sample size. The sample size of tweets in each topic category is provided in Figure S1 in Multimedia Appendix 1.

Instead, the lockdown affected individuals' emotions on other topics, such as health and politics. Our results show that the lockdown increased health-related uncertainty among people in the treated area. Higher uncertainty about health issues may reflect concerns for personal safety, but it also raises public awareness of the health risks posed by the virus [38]. This suggests that the health concerns arising from the implementation of a lockdown outweigh the reassurance authorities aim to provide through lockdown measures and their commitment to containing the virus. However, our findings indicate that negative sentiment toward health issues was driven by COVID-19 itself rather than the lockdown.

The same logic applies to the impact of the lockdown on emotions expressed in discussions about the lockdown policy. Concern about the guidelines and the consequences of the policy was high among those in the red zone, but this was not accompanied by a significant change in sentiment among the treated population (b=0.024, SE 0.02, t_{125}=1.22; P=.22). Sentiment in discussions about the policy worsened in both the treated and control zones after the measures were implemented, suggesting that this deterioration was likely due to COVID-19 itself rather than the lockdown.

Our results indicate that political sentiment among Twitter users in the treated area deteriorated significantly (b=0.025, SE 0.011, t_{125}=2.33; P=.02). This suggests that the lockdown measures came at the cost of worsening political emotions. The increase in negative sentiment within the red zone may reflect heightened political polarization as a direct consequence of the lockdown measures, aligning with previous findings on political polarization during COVID-19 [18,19]. A rise in negative sentiment toward politics could also signal growing dissatisfaction with politicians in office. Previous research has shown that public attitudes toward politicians improved during the pandemic [35,39,61]. This aligns with the observed significant decline in the overall trend following the nationwide measures implemented on March 9 (b=–0.02, SE 0.008, t_{125}=–2.30; P=.02), which came in response to a sharp increase in COVID-19 cases and deaths across the country.

In addition, the lockdown significantly worsened political uncertainty in the treated area (b=0.012, SE 0.0058, t_{125}=2.05; P=.04). However, as will be explained in the next section, the estimated effect on political uncertainty is not robust to a series of validity checks on the assumptions of the DiD model.

The overall effect of the lockdown on uncertainty was positive, driven by a significant increase in uncertainty related to health (P<.001), policy (P<.001), and politics (P=.04). Conversely, aggregate negative sentiment in the red zone decreased. This decline was primarily due to reduced negative sentiment when discussing topics of lesser interest, mainly those related to entertainment and sports (see Table S8 in Multimedia Appendix 1). However, the estimated effect on overall negative sentiment does not hold up under most of the robustness checks we conducted, as will be explained in the next section. This reinforces the validity of the topic-based analysis in better assessing the impact of lockdowns on uncertainty and sentiment.

In general, our findings highlight the nuanced effects of lockdown policies on various emotional aspects and suggest that their impact on health, economics, and politics differs significantly. The quasi-experimental framework of our analysis allows us to distinguish between the effects of lockdown policies and broader emotional trends related to the pandemic. In the next section, we explain that the effects on aggregate negative sentiment and political uncertainty do not remain robust when subjected to a series of robustness checks.

Robustness Checks

We conducted a battery of tests to assess the robustness of our results and summarize the key findings here.

First, we performed a series of tests to evaluate the validity of the DiD model assumptions—*parallel trends*, *absence of spillover effects*, and *no anticipation* of the treatment. To check whether the parallel trends assumption holds, we examine whether emotional reaction trends differed between the treated

XSL•FO
RenderX

and control groups before the lockdown (see Figures S2 and S3 in the "Preexisting Trends" section in Multimedia Appendix 1). We find no significant differences in pretreatment trends for aggregate uncertainty, health-related uncertainty, policy uncertainty, or negative sentiment toward politics, confirming the parallel trends assumption. However, we do observe differences in pretreatment trends for aggregate negative sentiment and political uncertainty. This suggests that the DiD estimate of the lockdown's impact on these 2 variables may be biased, as the control group does not behave as a true counterfactual, given the significant differences in prepolicy trends.

Furthermore, our findings indicate that estimates of the lockdown's effect on political uncertainty and aggregate negative sentiment are highly sensitive to the inclusion of certain municipalities in the comparison group (see Table S9 in the "Exclusion of Territorial Administrative Units" section in Multimedia Appendix 1).

Finally, we examine the potential local spillover effects of the treatment on control units [62]. Our estimates could be biased if emotions in the control group were influenced by online reactions from the red zone. However, we argue that, in this specific context, it is reasonable to assume that red zone responses primarily affected only nearby control units, while those located beyond a certain distance remained unaffected. These distant units serve as the comparison group for the DiD estimate of the treatment effect. Our findings show that the treatment effect estimates remain consistent with the baseline DiD estimates (see Table S4 in Multimedia Appendix 1).

We also applied Benjamini-Hochberg adjusted *P* values [63] to control for the risk of false positives when testing multiple models with potentially correlated results (see Tables S5 and S6 in the "Multiple Test Adjusted *P* Values" section in Multimedia Appendix 1). The analysis confirms our findings, except for uncertainty related to politics.

As an additional robustness check, we estimate the lockdown's effect on public emotions by comparing tweets before and after the February 23, 2020, policy across different groups: the red zone, the extended orange zone defined by the March 8, 2020, decree, and other Northern Italian locations that were later subjected to nationwide measures starting on March 9, 2020, but were never designated as a targeted priority—referred to as the "white zone" (see Figure S5 in Multimedia Appendix 1). This approach allows us to assess how spatial proximity to the lockdown area and exposure to government policy—under similar media influences—affect our findings. Pretreatment checks reveal significant demographic differences between the red zone and control groups (see Figures S6-S9 in Multimedia Appendix 1). While we rely on manually entered user locations, which are less precise than GPS coordinates, this method still enables us to distinguish tweets originating from within and outside the red zone. We randomly subsampled 20% of tweets from the extended orange zone and 50% from the white zone to ensure computational scalability in estimating the model coefficients while controlling for user-level fixed effects. Compared with the baseline DiD, the treatment effect on aggregated uncertainty, health-related uncertainty, and aggregate negative sentiment is more pronounced in the white zone but smaller in the extended orange zone. The difference increases as the distance of extended orange zone locations from the red zone grows, while it decreases as white zone locations move farther from the red zone. This suggests that differential exposure to (mild) government policies against contagion played a relevant role. Negative sentiment toward politics is higher when treated units are compared with extended orange zone locations after the policy, but the difference is significant only for those in closer proximity to the lockdown area. Additionally, we find a significant difference in economic negative sentiment between the red zone and the white zone (see Tables S11-S14 and the "DiD Estimates Varying for Spatial Proximity" section in Multimedia Appendix 1 for details).

Finally, we conducted a placebo test to examine how *anticipation* of the policy affected behavior and the estimated impact of the lockdown. Following the nationwide lockdown on March 9, we argue that municipalities with COVID-19 risk levels similar to those of the first 10 quarantined towns should not have reacted differently from those with dissimilar risk, as the nationwide measures were widely expected (see Figure S4 in the "Placebo Test" section in Multimedia Appendix 1). Using the percentage increase in monthly deaths from ISTAT as a proxy for contagion rates, we treat the lockdown as a placebo in areas where excess deaths were already high. The results indicate that the national lockdown had no significant effect on public reactions in the placebo treatment group, reinforcing the robustness of our original findings by confirming that potential anticipatory effects did not bias the estimated impact of the policy (see Table S7 in Multimedia Appendix 1).

Overall, the robustness tests support the key findings that the lockdown had significant effects on aggregate uncertainty, health-related uncertainty, policy uncertainty, and negative sentiment toward politics. However, the effects on political uncertainty and aggregate negative sentiment do not hold up across several tests.

Discussion

Principal Findings

Overall, our findings confirm that lockdowns come at a cost, though in an unexpected way. We found no causal impact on economic uncertainty or sentiment, suggesting that economic-related emotions were primarily driven by the rise in COVID-19 cases rather than the lockdown itself. Instead, in the red zone areas under lockdown, users were more likely to express uncertainty about health and lockdown policies, along with increased negative sentiment toward politics.

The evidence suggests that lockdowns increase, rather than reduce, uncertainty about health conditions without significantly altering sentiment in discussions on health or the management of the sanitary emergency. Our findings indicate that the deterioration of sentiment in health- and policy-related discussions is driven by COVID-19 itself rather than stay-at-home measures. Notably, lockdowns carry political costs, as affected areas experience a surge in negative political

XSL•FO
RenderX

sentiment, which may deter future policy makers from implementing stay-at-home mandates to avoid political backlash.

Comparison With Prior Work

Our contribution is 2-fold. First, we develop a new methodological framework to assess the impact of lockdown measures [31-33,39,64]. We focus on the causal effects on public emotions, an area where, to our knowledge, only correlational evidence exists [10,17,39]. Understanding how stay-at-home mandates influence public emotions is crucial for policy makers to weigh the benefits and costs of such measures.

Second, we contribute to the literature on uncertainty and sentiment, which often relies on nontraditional data such as textual analysis. Recent studies have reported particularly high levels of economic uncertainty [65] and health-related negative emotions [17,20] following the enactment of COVID-19 lockdown measures, as well as increased political polarization [18,19]. However, less is known about the specific factors driving these changes. We help bridge this knowledge gap by showing that uncertainty and negative sentiment did not increase uniformly across individuals after the lockdown but followed distinct dynamics. This is the first study to examine the effects of stay-at-home mandates on multiple dimensions of public uncertainty and sentiment.

Limitations

It is important to note that all field experiments may suffer from a lack of external validity. Our study is no exception, as a new lockdown today could be perceived differently from the unexpected lockdown analyzed in this paper. One could argue that the public is now much more informed about the implications of a lockdown than it was in February 2020. People would adjust their strategies based on their experiences, potentially altering the policy's impact on uncertainty and sentiment. Moreover, given the negative political sentiments such measures have generated, policy makers may now be more reluctant to implement them or, at the very least, enforce them with the same level of strictness. Nevertheless, we demonstrate that once the causal effect of the policy is effectively isolated from comoving factors and external influences, lockdowns impose political and social costs by exacerbating uncertainty and sentiment. To ensure robust and reliable causal inference, we prioritize the internal validity of our estimates over the generalizability of our results by analyzing a case study in which the policy change can be viewed as randomly allocated among groups. While it is impossible to predict how the public and policy makers will respond to a future lockdown, our findings offer valuable insights into the costs policy makers must consider before implementing new lockdown measures.

Conclusions

Our findings suggest that communication surrounding lockdown measures may have been unclear, fostering resentment toward the political class among those affected and failing to generate explicit support for the government's plans. Policy makers must enhance communication to reduce uncertainty and negative sentiment [66]. These findings underscore the need for authorities to refine future policies and communication strategies to mitigate uncertainty and social panic.

Acknowledgments

FBS and MR conceived the research idea and designed the identification strategy. MP gathered data. CB conducted the experiment(s) and analyzed the results. CB, FBS, NF, and MR supervised the investigation. CB, FBS, NF, MP, and MR wrote and reviewed the manuscript. This work is supported in part by the Tuscany Health Ecosystem—THE and GRINS/PUWELL, PNRR, NextGeneration EU. We benefited from helpful comments and suggestions from Danielle Braun, Abel Brodeur, Francesca Dominici, Kosuke Imai, Kevin Josey, Fabrizia Mealli, Giovanni Mellace, and Davide Viviano. The study results were previously presented at the Harvard Data Science Seminar (2022), the Institute of Mathematical Statistics (IMS) International Conference on Statistics and Data Science (ICSDS) (2022), Joint Political Economy and Applied Microeconomics Workshop at the University of Bolzano (2022), at the Advanced Language Processing Winter School (2023), and at EEA-ESEM Congress (2023).

Data Availability

The data and code underlying this article are available in Zenodo and can be accessed at [67].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary analysis and results.
[PDF File (Adobe PDF File), 5467 KB - jmir_v27i1e64667_app1.pdf]

Multimedia Appendix 2
Word clouds of the 100 most frequent terms in tweets classified as (A) aggregated uncertainty and (B) negative sentiment. In our thematic categories, the most prominent terms in A relate to health (eg, coronavirus, virus, cases [casi], deaths [morti], patient [paziente], asymptomatic [asintomatico]), politics (eg, Europe [Europa], parties [partito]), and lockdown policy (eg, supplies/food [viveri necessari], closures [chiuderlo]). In panel B, the most frequently occurring words belong to the health category (eg,

XSL•FO
RenderX

coronavirus, deaths [morti], ambulances [ambulanze]) and the lockdown policy category (eg, closures [chiuderlo], provisions/food supplies [viveri necessari]). The complete set of words belonging to each thematic category is listed and translated in Tables S1 and S2 in Multimedia Appendix 1.
[PNG File , 348 KB - jmir_v27i1e64667_app2.png]

References

1. Abosedra S, Laopodis NT, Fakih A. Dynamics and asymmetries between consumer sentiment and consumption in pre- and during-COVID-19 time: evidence from the US. J Econ Asymmetries 2021 Nov;24:e00227 [FREE Full text] [doi: 10.1016/j.jeca.2021.e00227] [Medline: 34899924]
2. Caggiano G, Castelnuovo E, Kima R. The global effects of Covid-19-induced uncertainty. Econ Lett 2020 Sep;194:109392 [FREE Full text] [doi: 10.1016/j.econlet.2020.109392] [Medline: 32834236]
3. Brzoza-Brzezina M, Wesołowski G. The great lockdown: information, noise, and macroeconomic fluctuations. Macroecon. Dynam 2022 Feb 04;27(4):906-927. [doi: 10.1017/s1365100521000705]
4. Moser C, Yared P. Pandemic lockdown: the role of government commitment. Review of Economic Dynamics 2022 Oct;46:27-50. [doi: 10.1016/j.red.2021.08.001]
5. Porto ED, Naticchioni P, Scrutinio V. Lockdown, essential sectors, and Covid-19: lessons from Italy. J Health Econ 2022 Jan;81:102572 [FREE Full text] [doi: 10.1016/j.jhealeco.2021.102572] [Medline: 34958981]
6. Fowler JH, Hill SJ, Levin R, Obradovich N. Stay-at-home orders associate with subsequent decreases in COVID-19 cases and fatalities in the United States. PLoS One 2021;16(6):e0248849 [FREE Full text] [doi: 10.1371/journal.pone.0248849] [Medline: 34111123]
7. Zhang R, Wang Y, Lv Z, Pei S. Evaluating the impact of stay-at-home and quarantine measures on COVID-19 spread. BMC Infect Dis 2022 Jul 27;22(1):648 [FREE Full text] [doi: 10.1186/s12879-022-07636-4] [Medline: 35896977]
8. Altig D, Baker S, Barrero JM, Bloom N, Bunn P, Chen S, et al. Economic uncertainty before and during the COVID-19 pandemic. J Public Econ 2020 Nov;191:104274 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104274] [Medline: 32921841]
9. Goolsbee A, Syverson C. Fear, lockdown, and diversion: comparing drivers of pandemic economic decline 2020. Journal of Public Economics 2021 Jan;193:104311. [doi: 10.1016/j.jpubeco.2020.104311] [Medline: 33262548]
10. Priyadarshini I, Mohanty P, Kumar R, Sharma R, Puri V, Singh PK. A study on the sentiments and psychology of twitter users during COVID-19 lockdown period. Multimed Tools Appl 2022 Jun 14;81(19):27009-27031 [FREE Full text] [doi: 10.1007/s11042-021-11004-w] [Medline: 34149302]
11. Hsu JT, Tsai RT. Increased online aggression during COVID-19 lockdowns: two-stage study of deep text mining and difference-in-differences analysis. J Med Internet Res 2022 Aug 09;24(8):e38776 [FREE Full text] [doi: 10.2196/38776] [Medline: 35943771]
12. Zervoyianni A, Dimelis S, Livada A. Economic sentiment and the Covid-19 crisis: evidence from European countries. Applied Economics 2022 May 17;55(1):113-130. [doi: 10.1080/00036846.2022.2061903]
13. van der Wielen W, Barrios S. Economic sentiment during the COVID pandemic: evidence from search behaviour in the EU. J Econ Bus 2021 May;115:105970 [FREE Full text] [doi: 10.1016/j.jeconbus.2020.105970] [Medline: 36540811]
14. Baker SR, Bloom N, Davis SJ, Terry SJ. Covid-induced economic uncertainty (working paper series 26983). National Bureau of Economic Research. 2020 Apr. URL: http://www.nber.org/papers/w26983 [accessed 2022-12-01]
15. Chintalapudi N, Battineni G, Amenta F. Sentimental analysis of COVID-19 tweets using deep learning models. Infect Dis Rep 2021 Apr 01;13(2):329-339 [FREE Full text] [doi: 10.3390/idr13020032] [Medline: 33916139]
16. Aslam F, Awan TM, Syed JH, Kashif A, Parveen M. Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanit Soc Sci Commun 2020 Jul 08;7(1):23. [doi: 10.1057/s41599-020-0523-3]
17. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
18. Jungkunz S. Political polarization during the COVID-19 pandemic. Front Polit Sci 2021 Mar 4;3:1-8. [doi: 10.3389/fpos.2021.622512]
19. Jiang J, Chen E, Yan S, Lerman K, Ferrara E. Political polarization drives online conversations about COVID-19 in the United States. Hum Behav Emerg Technol 2020 Jul;2(3):200-211 [FREE Full text] [doi: 10.1002/hbe2.202] [Medline: 32838229]
20. Li Y, Chen M, Lee H. Health communication on social media at the early stage of the pandemic: examining health professionals' COVID-19 related tweets. Soc Sci Med 2024 Apr;347:116748 [FREE Full text] [doi: 10.1016/j.socscimed.2024.116748] [Medline: 38484456]
21. Cereda D, Manica M, Tirani M, Rovida F, Demicheli V, Ajelli M, et al. The early phase of the COVID-19 epidemic in Lombardy, Italy. Epidemics 2021 Dec;37:100528 [FREE Full text] [doi: 10.1016/j.epidem.2021.100528] [Medline: 34814093]
22. Polignano M, Basile V, Basile P, de Gemmis M, Semeraro G. AlBERTo: modeling Italian social media language with BERT. ijcol 2019 Dec 01;5(2):11-31 [FREE Full text] [doi: 10.4000/ijcol.472]

XSL•FO
RenderX

23. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). New York, NY: Association for Computational Linguistics; 2019 Presented at: The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 1, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
24. Baker SR, Bloom N, Davis SJ. Measuring economic policy uncertainty. The Quarterly Journal of Economics 2016;131(4):1593-1636. [doi: 10.1093/qje/qjw024]
25. Shapiro A, Sudhof M, Wilson D. Measuring news sentiment. Journal of Econometrics 2022 Jun;228(2):221-243 [FREE Full text] [doi: 10.1016/J.JECONOM.2020.07.053] [Medline: 22818403]
26. Blanco G, Lourenço A. Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations. Inf Process Manag 2022 May;59(3):102918 [FREE Full text] [doi: 10.1016/j.ipm.2022.102918] [Medline: 36569234]
27. Pavlov T, Mirceva G. COVID-19 Fake News Detection by Using BERT and RoBERTa models. 2022 Presented at: 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO); May 23-27, 2022; Opatija, Croatia p. 312-316. [doi: 10.23919/MIPRO55190.2022.9803414]
28. Min H, Peng Y, Shoss M, Yang B. Using machine learning to investigate the public's emotional responses to work from home during the COVID-19 pandemic. J Appl Psychol 2021 Feb;106(2):214-229. [doi: 10.1037/apl0000886] [Medline: 33818121]
29. To QG, To KG, Huynh VN, Nguyen NTQ, Ngo DTN, Alley SJ, et al. Applying machine learning to identify anti-vaccination tweets during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Apr 12;18(8):4069 [FREE Full text] [doi: 10.3390/ijerph18084069] [Medline: 33921539]
30. Oana IE, Pellegata A, Wang C. A cure worse than the disease? Exploring the health-economy trade-off during COVID-19. West European Politics 2021 Jun 17;44(5-6):1232-1257. [doi: 10.1080/01402382.2021.1933310]
31. Devine D, Gaskell J, Jennings W, Stoker G. Trust and the coronavirus pandemic: what are the consequences of and for trust? An early review of the literature. Polit Stud Rev 2021 May;19(2):274-285 [FREE Full text] [doi: 10.1177/1478929920948684] [Medline: 35082554]
32. Schraff D. Political trust during the Covid-19 pandemic: rally around the flag or lockdown effects? Eur J Polit Res 2021 Nov 25;60(4):1007-1017 [FREE Full text] [doi: 10.1111/1475-6765.12425] [Medline: 33362332]
33. Hegewald S, Schraff D. Who rallies around the flag? Evidence from panel data during the Covid-19 pandemic. Journal of Elections, Public Opinion and Parties 2022 Sep 15;34(1):158-179. [doi: 10.1080/17457289.2022.2120886]
34. Baekgaard M, Christensen J, Madsen JK, Mikkelsen KS. Rallying around the flag in times of COVID-19: societal lockdown and trust in democratic institutions. JBPA 2020 Jul 12;3(2):1-28. [doi: 10.30636/jbpa.32.172]
35. De Vries CE, Bakker BN, Hobolt SB, Arceneaux K. Crisis signaling: how Italy's coronavirus lockdown affected incumbent support in other European countries. PSRM 2021 Mar 08;9(3):451-467. [doi: 10.1017/psrm.2021.6]
36. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA 2014 Dec 10;312(22):2401-2402. [doi: 10.1001/jama.2014.16153] [Medline: 25490331]
37. Alexander D, Karger E. Do stay-at-home orders cause people to stay at home? Effects of stay-at-home orders on consumer behavior. The Review of Economics and Statistics 2023;105(4):1017-1027. [doi: 10.1162/rest_a_01108]
38. De Rosis S, Lopreite M, Puliga M, Vainieri M. The early weeks of the Italian Covid-19 outbreak: sentiment insights from a Twitter analysis. Health Policy 2021 Aug;125(8):987-994 [FREE Full text] [doi: 10.1016/j.healthpol.2021.06.006] [Medline: 34176671]
39. BOL D, GIANI M, BLAIS A, LOEWEN PJ. The effect of COVID ‐ 19 lockdowns on political support: some good news for democracy? European J Political Res 2020 May 31;60(2):497-505 [FREE Full text] [doi: 10.1111/1475-6765.12401]
40. Helgeson SA, Mudgalkar RM, Jacobs KA, Lee AS, Sanghavi D, Moreno Franco P, National COVID Cohort Collaborative (N3C). Association between X/Twitter and prescribing behavior during the COVID-19 pandemic: retrospective ecological study. JMIR Infodemiology 2024 Nov 18;4:e56675 [FREE Full text] [doi: 10.2196/56675] [Medline: 39556417]
41. Kamba M, She WJ, Ferawati K, Wakamiya S, Aramaki E. Exploring the impact of the COVID-19 pandemic on Twitter in Japan: qualitative analysis of disrupted plans and consequences. JMIR Infodemiology 2024 Apr 01;4:e49699 [FREE Full text] [doi: 10.2196/49699] [Medline: 38557446]
42. Marani H, Song MY, Jamieson M, Roerig M, Allin S. Public officials' engagement on social media during the rollout of the COVID-19 vaccine: content analysis of tweets. JMIR Infodemiology 2023 Jul 20;3:e41582 [FREE Full text] [doi: 10.2196/41582] [Medline: 37315194]
43. Ueda R, Han F, Zhang H, Aoki T, Ogasawara K. Verification in the early stages of the COVID-19 pandemic: sentiment analysis of Japanese Twitter users. JMIR Infodemiology 2024 Feb 06;4:e37881 [FREE Full text] [doi: 10.2196/37881] [Medline: 38127840]
44. Governmental Decree of Italy. Gazzetta Ufficiale n.45. Decreto del Presidente del Consiglio dei Ministri 23 Febbraio 2020. Gazzetta Ufficiale. 2020 Feb 23. URL: https://www.gazzettaufficiale.it/eli/id/2020/02/23/20A01228/sg [accessed 2022-10-10]
45. Governmental Decree of Italy. Gazzetta Ufficiale n. 59. Decreto del Presidente del Consiglio dei Ministri 8 Marzo 2020. Gazzetta Ufficiale. 2020 Mar 08. URL: https://www.gazzettaufficiale.it/eli/id/2020/03/08/20A01522/sg [accessed 2022-10-10]

XSL•FO
RenderX

46. Governmental Decree of Italy. Gazzetta Ufficiale n. 62. Decreto del Presidente del Consiglio dei Ministri 9 Marzo 2020. Gazzetta Ufficiale. 2020 Mar 09. URL: https://www.gazzettaufficiale.it/eli/id/2020/03/ [accessed 2022-10-10]
47. Codogno 20 febbraio, il tampone 0 che scoprì il Covid in Italia. La dottoressa Annalisa Malara: Ricordo ancora la paura, ora va molto meglio? La Stampa. 2022 Feb 20. URL: https://www.lastampa.it/cronaca/2022/02/20/news/codogno_20_febbraio_il_tampone_0_che_scopri_il_covid_in_italia_la_dottoressa_annalisa_malara_ricordo_ancora_la_paura_or-2859352/ [accessed 2022-10-10]
48. ISTAT. Codici statistici delle unità amministrative territoriali: comuni, città metropolitane, province e regioni. Istituto Nazionale di Statistica e Informatica. Rome, Italy: Istituto Nazionale di Statistica e Informatica; 2022 Mar 03. URL: https://www.istat.it/it/archivio/6789 [accessed 2022-07-06]
49. ISTAT. Censimenti permanenti: censimento della popolazione e delle abitazioni. Istituto Nazionale di Statistica e Informatica. Rome, Italy: Istituto Nazionale di Statistica e Informatica; 2019. URL: http://dati-censimentipermanenti.istat.it/ [accessed 2022-07-06]
50. ISTAT. Data set analitico con i dati comunali nei settori attivi e sospesi. Istituto Nazionale di Statistica e Informatica. Rome, Italy: Istituto Nazionale di Statistica e Informatica; 2020. URL: https://www.istat.it/it/archivio/241341 [accessed 2022-09-08]
51. ISTAT. Popolazione residente al 1° gennaio. Istituto Nazionale di Statistica e Informatica. Rome, Italy: Istituto Nazionale di Statistica e Informatica; 2020. URL: http://dati.istat.it/index.aspx?data [accessed 2022-07-08]
52. ISTAT. Tavola decessi per 6.866 comuni al 31 marzo. Istituto Nazionale di Statistica e Informatica. Rome, Italy: Istituto Nazionale di Statistica e Informatica; 2020. URL: https://http://www/istat.it/it/archivio/240401 [accessed 2022-07-09]
53. Branson Z. Randomization tests to assess covariate balance when designing and analyzing matched datasets. Observational Studies 2021;7(2):1-36. [doi: 10.1353/obs.2021.0031]
54. Imran AS, Daudpota SM, Kastrati Z, Batra R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. IEEE Access 2020;8:181074-181090 [FREE Full text] [doi: 10.1109/ACCESS.2020.3027350] [Medline: 34812358]
55. Zhang Y, Cheng J, Yang Y, Li H, Zheng X, Chen X, et al. COVID-19 public opinion and emotion monitoring system based on time series thermal new word mining. Comput Mater Contin 2020;64(3):1415-1434. [doi: 10.32604/cmc.2020.011316]
56. Metzler H, Pellert M, Garcia D. Using social media data to capture emotions before and during COVID-19. In: World Happiness Report. New York, NY: Sustainable Development Solutions Network; 2022:77-104.
57. Zheng X, Han J, Sun A. A survey of location prediction on Twitter. IEEE Trans Knowl Data Eng 2018 Sep 1;30(9):1652-1671. [doi: 10.1109/tkde.2018.2807840]
58. ISTAT. Confini delle amministrative a fini statistici unità al 1° Gennaio. Istituto nazionale di statistica e informatica. Rome (Italy); 2022. URL: https://www.istat.it/it/archivio/222527 [accessed 2022-07-06]
59. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22 [FREE Full text] [doi: 10.1093/biomet/73.1.13]
60. Adams T, Ajello A, Silva D, Vazquez-Grande F. More than words: Twitter chatter and financial market sentiment (FEDS Working Paper No. 2023-34). SSRN. 2023 May. URL: https://ssrn.com/abstract=4467949 [accessed 2024-01-01]
61. Sibley CG, Greaves LM, Satherley N, Wilson MS, Overall NC, Lee CHJ, et al. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. Am Psychol 2020;75(5):618-630. [doi: 10.1037/amp0000662] [Medline: 32496074]
62. Kyle B. Difference-in-differences estimation with spatial spillovers. arXiv. Preprint posted online on June 10, 2023. 2021 May 8. URL: https://arxiv.org/abs/2105.03737 [accessed 2023-10-10]
63. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) 1995 Jan;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
64. Palomino JC, Rodríguez JG, Sebastian R. Wage inequality and poverty effects of lockdown and social distancing in Europe. Eur Econ Rev 2020 Oct;129:103564 [FREE Full text] [doi: 10.1016/j.euroecorev.2020.103564] [Medline: 32836323]
65. Yang J, Yang C. Economic policy uncertainty, COVID-19 lockdown, and firm-level volatility: evidence from China. Pacific-Basin Finance Journal 2021 Sep;68:101597. [doi: 10.1016/j.pacfin.2021.101597]
66. Mishra V, Dexter JP. Comparison of readability of official public health information about COVID-19 on websites of international agencies and the governments of 15 countries. JAMA Netw Open 2020 Aug 03;3(8):e2018033 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18033] [Medline: 32809028]
67. Biliotti C, Fraccaroli N, Puliga M, Bargagli Stoffi FJ, Riccaboni M. Breaking down the lockdown: the impact of stay-at-home mandates on uncertainty and sentiments. Zenodo. 2024. URL: https://doi.org/10.5281/zenodo.14562600 [accessed 2024-12-01]

Abbreviations

AlBERTo: Italian BERT optimized
API: application programming interface
BERT: Bidirectional Encoder Representations from Transformers
DiD: difference-in-differences

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.07.24; peer-reviewed by G Grossi, Q Shi, G Visconti, C Zachreson; comments to author 15.10.24; revised version received 18.11.24; accepted 29.12.24; published 04.03.25.

Please cite as:
Biliotti C, Fraccaroli N, Puliga M, Bargagli-Stoffi FJ, Riccaboni M
The Impact of Stay-At-Home Mandates on Uncertainty and Sentiments: Quasi-Experimental Study
J Med Internet Res 2025;27:e64667
URL: https://www.jmir.org/2025/1/e64667
doi:10.2196/64667
PMID:40053818

©Carolina Biliotti, Nicolò Fraccaroli, Michelangelo Puliga, Falco J Bargagli-Stoffi, Massimo Riccaboni. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Psychological Trends in Populations With Chronic Obstructive Pulmonary Disease During COVID-19 and Beyond: Large-Scale Longitudinal Twitter Mining Study

Chunyan Zhang[1,2*], MS; Ting Wang[3*], PhD; Caixia Dong[2], MS; Duwei Dai[2], MS; Linyun Zhou[2], MPH; Zongfang Li[2], PhD; Songhua Xu[2], PhD

[1]The High School Affiliated to Xi'an Jiaotong University, Xi'an, China

[2]Institute of Medical Artificial Intelligence, Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[3]School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Ting Wang, PhD
School of Electrical Engineering
Xi'an Jiaotong University
No. 28, Xianning West Road
Xi'an, 710049
China
Phone: 86 13289346632
Email: wang.ting@xjtu.edu.cn

Abstract

Background: Chronic obstructive pulmonary disease (COPD) ranks among the leading causes of global mortality, and COVID-19 has intensified its challenges. Beyond the evident physical effects, the long-term psychological effects of COVID-19 are not fully understood.

Objective: This study aims to unveil the long-term psychological trends and patterns in populations with COPD throughout the COVID-19 pandemic and beyond via large-scale Twitter mining.

Methods: A 2-stage deep learning framework was designed in this study. The first stage involved a data retrieval procedure to identify COPD and non-COPD users and to collect their daily tweets. In the second stage, a data mining procedure leveraged various deep learning algorithms to extract demographic characteristics, hashtags, topics, and sentiments from the collected tweets. Based on these data, multiple analytical methods, namely, odds ratio (OR), difference-in-difference, and emotion pattern methods, were used to examine the psychological effects.

Results: A cohort of 15,347 COPD users was identified from the data that we collected in the Twitter database, comprising over 2.5 billion tweets, spanning from January 2020 to June 2023. The attentiveness toward COPD was significantly affected by gender, age, and occupation; it was lower in females (OR 0.91, 95% CI 0.87-0.94; P<.001) than in males, higher in adults aged 40 years and older (OR 7.23, 95% CI 6.95-7.52; P<.001) than in those younger than 40 years, and higher in individuals with lower socioeconomic status (OR 1.66, 95% CI 1.60-1.72; P<.001) than in those with higher socioeconomic status. Across the study duration, COPD users showed decreasing concerns for COVID-19 and increasing health-related concerns. After the middle phase of COVID-19 (July 2021), a distinct decrease in sentiments among COPD users contrasted sharply with the upward trend among non-COPD users. Notably, in the post-COVID era (June 2023), COPD users showed reduced levels of joy and trust and increased levels of fear compared to their levels of joy and trust in the middle phase of COVID-19. Moreover, males, older adults, and individuals with lower socioeconomic status showed heightened fear compared to their counterparts.

Conclusions: Our data analysis results suggest that populations with COPD experienced heightened mental stress in the post-COVID era. This underscores the importance of developing tailored interventions and support systems that account for diverse population characteristics.

(J Med Internet Res 2025;27:e54543) doi:10.2196/54543

XSL•FO
RenderX

KEYWORDS

COVID-19; chronic obstructive pulmonary disease (COPD); psychological trends; Twitter; data mining; deep learning

Introduction

On January 30, 2020, the World Health Organization declared the COVID-19 pandemic as a public health emergency of international concern [1]. Three years later, on May 5, 2023, given the encouraging trends observed in the COVID-19 statistics, the World Health Organization determined that COVID-19 no longer constituted a public health emergency of international concern [2]. However, this did not indicate that COVID-19 was no longer a global health threat [3]. The COVID-19 pandemic has had and continues to exert a profound and long-lasting impact on human lives worldwide, affecting public health, food systems, and global job markets [4,5]. In this situation, populations with chronic diseases, particularly chronic obstructive pulmonary disease (COPD), face great hardships [6,7] due to several reasons. First, patients with COPD are more likely to develop severe COVID-19 symptoms due to their respiratory illness [8]. Second, the overwhelming demands of managing COVID-19 may strain valuable health care resources, potentially affecting populations with COPD due to resource allocation [6]. Third, COVID-19 has greatly disrupted the daily lives of patients with COPD, leading to a higher risk of exacerbations and even mortalities. However, in contrast to the tangible physical consequences, the long-term effects of COVID-19 on the psychological conditions of patients with COPD have not drawn sufficient attention.

There are limited studies on the psychological effects of COVID-19 on patients with COPD (Table S1 in Multimedia Appendix 1) [9-14]. These studies have highlighted the high rates of psychological stress, including depression, anxiety, and fear, among patients with COPD under the threat of COVID-19. However, all these studies focused primarily on the early acute phase of COVID-19, thereby providing limited insight into the mental conditions of patients with COPD over the long term. Drawing parallels from the historical context of the SARS pandemic, wherein distinct chronological phases were discerned in the psychological conditions of SARS survivors [15], the psychological aftermath of COVID-19 survivors may exhibit similar enduring and evolving traits over the long term. Furthermore, surveys in the previous studies were conducted through traditional means, including telephone [10-12,14], video calls [14], interviews [13], and questionnaires [10,12], requiring substantial labor resources and physical contacts. This ultimately limited the amount of data that could be collected during the pandemic.

In recent years, social media data mining has been widely used in health-related domains—both physical and mental. In the physical health domain, for example, Pradeepa et al [16] proposed a methodology to identify various symptoms associated with stroke and preventive measures for stroke from social media data. Fu et al [17] conducted a scoping review of methodologies for analyzing social media content in health care research and categorized health care areas into health practice, health services, and health education based on social media. In the mental health domain, for capturing the prevailing psychological conditions of the public during the COVID-19 pandemic, social media data mining has proven to be an effective solution. Compared with conventional survey approaches, social media data mining has significantly higher efficiency while minimizing infection risks by circumventing the need for physical contact. More importantly, the richness of social media data facilitates the discernment of long-term trends in emotional changes within a specific time period. For example, Lwin et al [18] used Twitter data to explore global trends in the 4 emotions—fear, anger, sadness, and joy—alongside their relative salience, spanning from January 28 to April 9, 2020. Similarly, through mining large-scale Twitter data, Zhang et al [19] conducted a cross-sectional study on web-based public sentiments in the early phase of the COVID-19 pandemic. However, these studies tended to either treat the populace as a monolithic entity, thus disregarding possible disparities among various population groups or only focus on general users while ignoring the existence of certain specific populations.

Considering the limitations of existing studies [9-19], this study aims to investigate the long-term psychological trends and patterns of populations with COPD with different characteristics during and after the COVID-19 pandemic. A new deep learning framework has been proposed to intelligently extract and analyze the characteristics of populations with COPD based on large-scale Twitter data.

Methods

Study Design

This study was conducted following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [20] (Multimedia Appendix 2). In this study, raw Twitter data were collected with our self-designed program via the Twitter application programming interface [21]. These data are composed of multimodal information, including texts, emojis, images, and hashtags and are stored in a chronological order in our Twitter database. To date, over 2.5 billion tweets have been collected, spanning from January 2020 to June 2023. Based on these data, we designed a 2-stage framework, titled Deep-COPD, incorporating a series of deep learning algorithms to track COPD Twitter users (hereafter referred to as COPD users). The flowcharts for the data retrieval (stage 1) and data mining (stage 2) procedures are shown in Figure 1.

XSL•FO
RenderX

Figure 1. Overall design of the Deep-COPD framework. (A) Data retrieval procedure. (B) Data mining procedure. CNN: convolutional neural network; Conv: convolutional; COPD: chronic obstructive pulmonary disease; LDA: latent Dirichlet allocation; RNN: recurrent neural network; ReLU: rectified linear unit; VADER: valence aware dictionary and emotional reasoner.

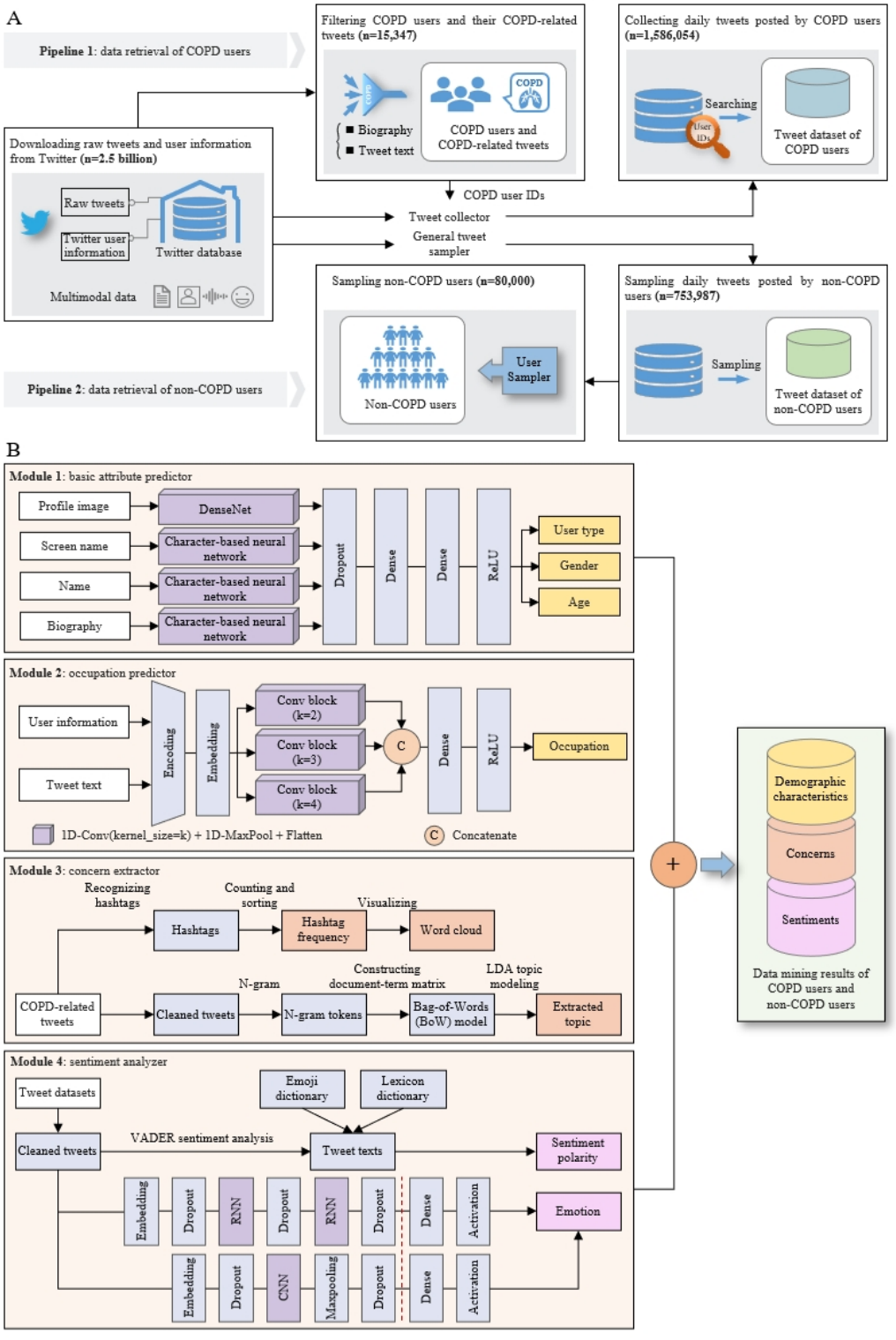

Data Retrieval

Figure 1A shows the procedure of data retrieval, which involves 2 pipelines: pipeline 1, retrieving the data of COPD users, including user information, COPD-related tweets, and their daily tweets; and pipeline 2, retrieving the data of non-COPD users, including user information and their daily tweets. In pipeline 1, we used a filter pattern implemented as a regular expression composed of a set of COPD keywords to identify COPD users and tweets by searching 2 sources: user biographies and tweet texts (see Multimedia Appendix 1). The specific construction process of this regular expression can be found in the build_match_word_regex() function of the filter_copd.py file (lines 17-35) in the stage 1 folder of our GitHub repository. We collected the daily tweets posted by these COPD users

XSL•FO
RenderX

through their user IDs. To limit the size of the data retrieved in pipeline 2, we sampled the tweets posted by non-COPD users at a rate of 0.05%. Through data retrieval, we obtained raw data that were used for the sentiment and psychological analysis. Moreover, we extracted data representing COPD and non-COPD characteristics from the daily tweets, enabling targeted research on the psychological evolution patterns associated with these 2 distinct groups. Specifically, this approach enables a more focused investigation into how the psychological trends may differ between users with COPD and those without COPD, thereby providing valuable insights into the impact of the condition on mental health over time. By analyzing these 2 groups separately, we aim to uncover specific psychological patterns and trends that could inform tailored interventions and support strategies for each population.

Data Mining

The data mining procedure aims to extract distinct information from multimodal Twitter data by using a variety of deep learning algorithms, including natural language processing, image processing, and symbol processing (Figure 1B). It consists of the following 4 parallel intelligent modules. Module 1 and module 2 are used to extract demographic characteristics, including user type, gender, age, and occupation, from both COPD and non-COPD users. Module 3 extracts concern information from COPD-related tweets, while module 4 extracts sentiment information from the daily tweet datasets of both COPD and non-COPD users.

Module 1: Basic Attribute Predictor

Module 1 is an attribute predictor that predicts the probabilities of 3 demographic attributes, that is, user type, gender, and age, through the analysis of profile images, screen names, display names, and biographies. It was implemented with an open-source package of the M3 (multimodal, multilingual, and multiattribute) model [22], a multimodal deep neural system trained on a massive dataset. It is mainly composed of a dense convolutional network (DenseNet) [23] for image processing, 3 character-based neural networks for text processing, and 2 fully connected dense layers for result classification. The training and validating M3 model uses 5 distinct datasets: a large dataset of Twitter profiles whose gender and age are heuristically identified, a curated dataset of Twitter accounts belonging to organizations, an image dataset of faces from IMDb (internet movie database) and Wikipedia, a massive unlabeled set of users, and a crowdsourced dataset for all 3 attributes. For more detailed information, readers can refer to the original literature [22]. This module was tested with a subset of our tweet dataset, achieving the following benchmark performance: 99.07% accuracy for user type, 95.88% accuracy for gender, and 77.65% accuracy for age attributes.

Module 2: Occupation Predictor

Module 2 is a self-designed deep learning model that derives the 2-class occupations of Twitter users from their profile information and tweet texts. It employs a word-based convolutional neural network that consists of an encoding layer, an embedding layer, 3 parallel convolutional blocks with different kernel sizes, and a dense layer. The training and validation data are obtained from a publicly available dataset [24] with 5191 Twitter users annotated with 9 occupation categories (OCs), which are defined according to the standard occupation classification [25]. In our study, the 9 OCs were merged into 2 OCs, abbreviated as OC1 and OC2 (see Table S4 in Multimedia Appendix 1). Individuals in OC1 are those of relatively higher socioeconomic status (such as managers, directors, and senior officials), and individuals in OC2 are those of relatively lower socioeconomic status (such as process, plant, and machine operatives). Using a 10-fold cross-validation approach, this module achieved an accuracy of 93.15% on the 2-class occupational dataset.

Module 3: Concern Extractor

Module 3 is a concern extractor that captures hashtags and latent Dirichlet allocation (LDA) [26] topics. Hashtags were identified in the tweet texts using a regular expression (#+[a-zA-Z0-9\-_]{1,}), and then the frequencies of these hashtags were calculated. LDA topic modeling was employed to identify the main topics expressed in the tweets. The original tweets underwent preprocessing, n-gram processing, and document-term matrix constructing, yielding a bag-of-words, which was fed into LDA models. The coherence score calculated by Gensim CoherenceMode [27] was used to evaluate the quality of the extracted topics. To pinpoint the optimal number of LDA topics, we performed a grid search within a range of 2-20 with a step of 1 for each year. The selection principle is to choose the number that makes the coherence score of the extracted topics reach the highest value. Through searching, the optimal number of topics and the coherence score for each year were obtained, as shown in Table S5 in Multimedia Appendix 1.

Module 4: Sentiment Analyzer

Module 4 is composed of a sentiment analyzer and an emotion detector to perceive sentiments from 2 different perspectives. The sentiment analyzer utilizes a widely used open-source tool known as the valence aware dictionary and emotional reasoner (VADER) [28]. VADER requires no training data, but it is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon. It is highly effective in analyzing social media sentiments, achieving an F_1 classification accuracy of 96%, outperforming human raters of 84% at classifying the sentiment of tweets. The emotion detector was built upon an open-source algorithm for emotion recognition on Twitter, which employs a character-based trained recurrent neural network [29]. The training and validation dataset is a set of tweets containing emotional hashtags and satisfying filtering requirements that searched from 73 billion tweets. Specifically, this study adopted the well-known Plutchik's emotion wheel [30] model, which consists of 8 basic emotions: joy, sadness, fear, anger, trust, disgust, surprise, and anticipation. This emotion detector achieved a macro F_1-score of 0.6950 for Plutchik's emotion model [29].

Specifically, the deep learning algorithms chosen in this study were selected for their strengths in handling various data types. DenseNet was employed for image processing, while character-based and word-based neural networks were used for the text classification of demographic attributes and occupations. LDA's topic modeling capability made it suitable for extracting

XSL•FO
RenderX

concerns from large text datasets. VADER's high accuracy ensured robust social media sentiment detection. Additionally, a character-based recurrent neural network enabled the emotion detector to effectively capture and classify complex emotional states.

The data mining procedure extracts valuable information from multimodal Twitter data by using advanced deep learning algorithms. This process involves 4 parallel intelligent modules, of which the outputs are as follows: module 1 (basic attribute predictor) predicts demographic attributes such as user type, gender, and age by analyzing user profile images, names, and biographies; module 2 (occupation predictor) classifies Twitter users' occupations into 2 groups based on their profiles and tweet content; module 3 (concern extractor) identifies user concerns by extracting hashtags and topics by using LDA; and module 4 (sentiment analyzer) analyzes sentiments from tweets by using VADER for general sentiments and a character-based recurrent neural network for emotion detection. Together, these modules provide a comprehensive analysis of the demographic, psychological, and emotional factors affecting both COPD and non-COPD Twitter users.

The outputs from data retrieval and data mining are crucial for assessing the psychological trends in populations with COPD, particularly the changes in sentiments, prevalent emotions, and evolving concerns (eg, health-related topics) during and after COVID-19. Through data retrieval, raw Twitter data are collected, allowing for data extraction of COPD and non-COPD groups, which enables focused analysis of psychological patterns within the target population. The data mining process further analyzes these data, extracting key demographic attributes, concerns, and sentiments by using deep learning algorithms. The concern extraction identifies health-related topics, while the sentiment and emotion analyzer tracks shifts in emotional states and sentiments, providing critical insights into the psychological evolution of populations with COPD in response to the pandemic. This comprehensive data processing directly supports the study's goal of understanding the psychological trends of COVID-19 in populations with COPD.

Data Analysis Tools

Based on the data mining results from our Deep-COPD framework, we conducted attentiveness analysis, concern analysis, and sentiment analysis. In these analyses, multiple statistical tools were used, namely, odds ratio (OR), difference-in-difference (DiD), and emotion pattern method.

Statistical Measure OR

OR is a statistical measure widely used in medical literature to evaluate the relevance between an event (such as lung cancer) and a particular exposure (such as smoking) [31]. In this study, the event refers to COPD, and the exposure factors include gender, age, and occupation. The ORs are used to quantitatively assess the levels of attentiveness toward COPD under these different exposure factors (Table 1).

Table 1. A 2x2 table for the chronic obstructive pulmonary disease event and the gender exposure factor (a, b, c, and d represent the number of users).

	Chronic obstructive pulmonary disease event	No event of chronic obstructive pulmonary disease
Female exposure	a	b
Male exposure	c	d

Taking gender as the exposure factor as an example, the 2×2 table shown in Table 1 is used to calculate the attentiveness OR toward a COPD event in females compared to males. The formula is as follows: OR = (a/b)/(c/d) = ad/bc

An OR of 1 indicates that the gender factor does not affect the level of attentiveness toward COPD. When OR>1, it suggests that the attentiveness toward COPD is higher in females than in males. Conversely, when OR<1, it indicates that the attentiveness toward COPD is lower in females than in males. The confidence interval in OR provides a range of values within which the true OR is likely to fall, offering an indication of the precision and reliability of the estimate. As a typical value, 95% CI is commonly used, indicating that there is a 95% probability that the true OR lies within the specified range [31]. The following formula is used for 95% CI:



In addition, the *P* values were obtained using the chi-square test to evaluate the significance levels of the differences between groups under different exposure factors. A lower *P* value indicates a more statistically significant difference between the groups [32].

DiD Model

The DiD model is typically used to estimate the effect of a specific intervention or treatment by comparing the changes in outcomes over time between a population that is enrolled in a program (the treated group) and a population that is not (the control group) [33,34]. In this study, we took COPD users as the treated group and non-COPD users as the control group and adopted the DiD model to estimate the relative sentiment trends of COPD users. The formula of the DiD model is as follows:



where Y_{it} is the dependent variable representing the sentiment status for group G_i and time t; G_i is a dummy variable for the treated group (G_0) or control group (G_1); D_t is a dummy variable for time t; α, β_k, γ, δ_k are the regression parameters; K is the total time period ($k \in [1, K]$); and ε_{it} is an error item.

The coefficients of interest are those of the interaction term, β_k, which can reflect the relative sentiment trends of COPD users compared to those of non-COPD users. A value of $\beta_k = 0$ signifies that both groups have parallel sentiment trends, whereas a negative value indicates the extent to which the sentiment

XSL•FO
RenderX

trend of the COPD group declines compared to that of the non-COPD group and vice versa.

Emotion Pattern

We proposed an emotion pattern method to calculate the values of the sentiment components of Plutchik's 8 emotions. The emotion pattern of population p is composed of the sentiment component scores of Plutchik's 8 emotions, which is expressed as follows:



where $e \in$ [joy, trust, surprise, anticipation, sadness, disgust, anger, fear], $S(p)|_e$ is the sentiment component of population p on emotion e,



is the calculated sentiment scores of tweets



expressing emotion e via the VADER method f, m_e is the number of tweets expressing emotion e, and N is the total number of tweets from population p.

Ethical Considerations

This analysis did not receive approval from an institutional research board. Analysis of large bodies of text written in social media such as the Twitter social network is generally not considered human subjects research [35], and therefore, ethics approval was not sought. We confirm that all data used in this study have been deidentified, and no individual user is identifiable in any of the images included in the manuscript or supplementary data.

Results

Demographic Characteristics of the Participants

In total, 15,347 COPD users were identified as participants in this study. From this cohort, 1,586,054 daily tweets were retrieved, spanning from January 2020 to June 2023. For comparison, 80,000 non-COPD users and 753,987 general tweets were also selected in the same period. Table 2 presents the statistical data on the demographic characteristics of these two groups. To obtain a comprehensive understanding of the psychological conditions of populations with COPD during COVID-19 and beyond, we analyzed the data mined by the Deep-COPD framework along 3 dimensions: attentiveness analysis, concern analysis, and sentiment analysis.

Table 2. Statistical data of chronic obstructive pulmonary disease and non–chronic obstructive pulmonary disease users.

Characteristics	COPD[a] users (n=15,347), n (%)	Non-COPD users (n=80,000), n (%)
User type		
Individual	13,375 (87.15)	75,468 (94.34)
Organization	1972 (12.85)	4532 (5.67)
Gender		
Male	7854 (58.72)	42,509 (56.33)
Female	5521 (41.28)	32,959 (43.67)
Age (years)		
≤18	1179 (8.81)	29,810 (39.50)
19-29	2373 (17.74)	25,086 (33.24)
30-39	2091 (15.63)	8551 (11.33)
≥40	7732 (57.81)	12,021 (15.93)
Occupation category		
OC1[b]	7356 (55)	50,566 (67)
OC2[c]	6019 (45)	24,902 (33)

[a]COPD: chronic obstructive pulmonary disease.

[b]OC1: occupation category 1 (relatively higher socioeconomic status such as managers, directors, and senior officials).

[c]OC2: occupation category 2 (relatively lower socioeconomic status such as process, plant, and machine operatives).

Attentiveness Analysis

Based on the statistics in different demographic groups (Table 2), a comparative study was performed between COPD and non-COPD users. As mentioned in the data analysis tools, we employed ORs to quantitatively assess the levels of attentiveness toward COPD in different groups (Figure 2). The ORs were calculated between COPD and non-COPD users under different exposure factors, including gender, age, and occupation. Results showed that the attentiveness toward COPD in females (OR 0.91, 95% CI 0.87-0.94; P<.001) was lower than that of males. The cohort aged 40 years and older exhibited a high OR of 7.23 (95% CI 6.95-7.52; P<.001), which was significantly higher than those younger than 40 years, and it showed a significant increase with age across 4 age groups. As for occupation, users

XSL•FO
RenderX

in OC2 occupations exhibited a notably higher OR of 1.66 (95% CI 1.60-1.72; *P*<.001) compared to those in OC1 occupations. These results suggested that gender, age, and occupation are 3 important factors influencing attentiveness. Specifically, older male users with lower socioeconomic status (OC2) were the most concerning group.

Figure 2. Attentiveness odds ratios toward chronic obstructive pulmonary disease in different demographic groups. Age 1: 2 age groups; age 2: 4 age groups; OC: occupation category; ref: reference.

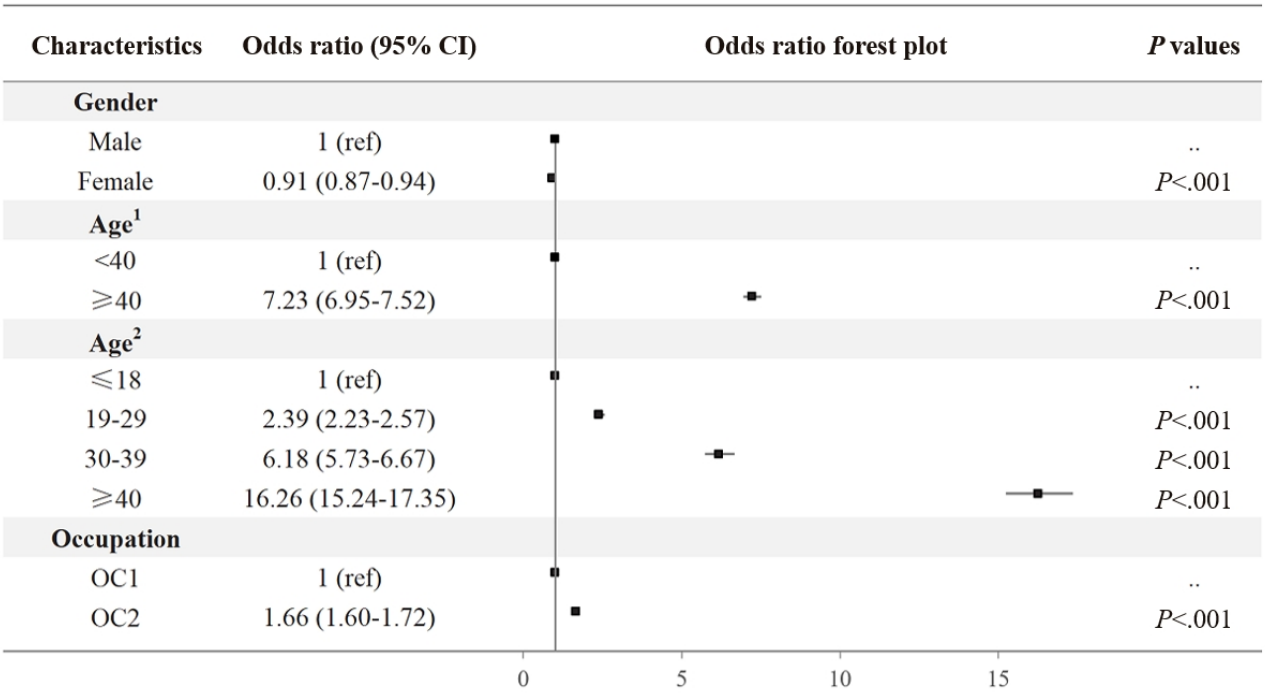

Concern Analysis

The word cloud of the top 500 hashtags in the tweets shows that #COPD and #COVID-19 were the most frequently used hashtags throughout the 4 years (Figure 3A). Among the top 10 hashtags in each year, #COPD consistently remained the primary concern throughout the 4-year period (Figure 3B). In 2020 and 2021, #COVID-19 ranked second but later gave way to health-related hashtags such as #medtwitter and #meded. To study the trends in the evolving concerns of COPD users, we categorized the 100 most frequently used hashtags from each year into 6 categories: COPD, COVID-19, health, politics, economics, and others. It is evident that the COPD category maintained the highest prevalence among COPD users, peaking at over 50% in 2021 (Figure 3C). The COVID-19 category held the second position in 2020, accounting for nearly 20%, but substantially decreased in 2021. On the other hand, the health category experienced continuous growth over the 4-year period, surpassing 30% in 2023 (Figure 3C). Figure 3D shows that the concerns from the perspective of LDA topics (see detailed topics in Table S6 in Multimedia Appendix 1) over the years follow the same trends as the hashtags. These statistical results highlight that, among COPD users, personal health became an even more significant concern as the pandemic receded.

XSL•FO
RenderX

Figure 3. Results of the concern analysis. (A) Word cloud of the top 500 hashtags. (B) Top 10 most used hashtags in 4 years. (C) The percentages of the 6 categories of hashtags in 4 years. (D) Visualization of the latent Dirichlet allocation topics via t-distributed stochastic neighbor embedding clustering. COPD: chronic obstructive pulmonary disease; LDA: latent Dirichlet allocation; tSNE: t-distributed stochastic neighbor embedding clustering.

A

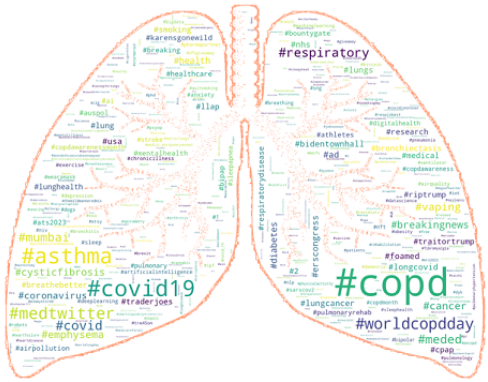

B

Rank	2020	2021	2022	2023
1	#copd	#copd	#copd	#copd
2	#covid19	#covid19	#asthma	#medtwitter
3	#asthma	#asthma	#medtwitter	#ats2023
4	#respiratory	#mumbai	#worldcopdday	#meded
5	#worldcopdday	#worldcopdday	#meded	#ad
6	#covid	#medtwitter	#covid	#psyop
7	#traderjoes	#breakingnews	#emphysema	#asthma
8	#riptrump	#covid	#usa	#foamed
9	#traitortrump	#vaping	#respiratory	#emphysema
10	#karensgonewild	#llap	#ers2022	#gold2023

C

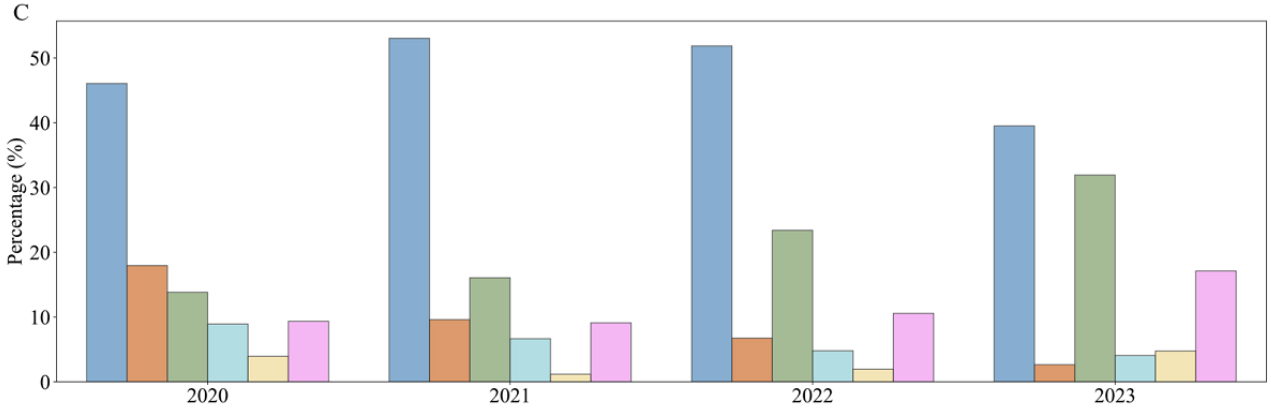

D

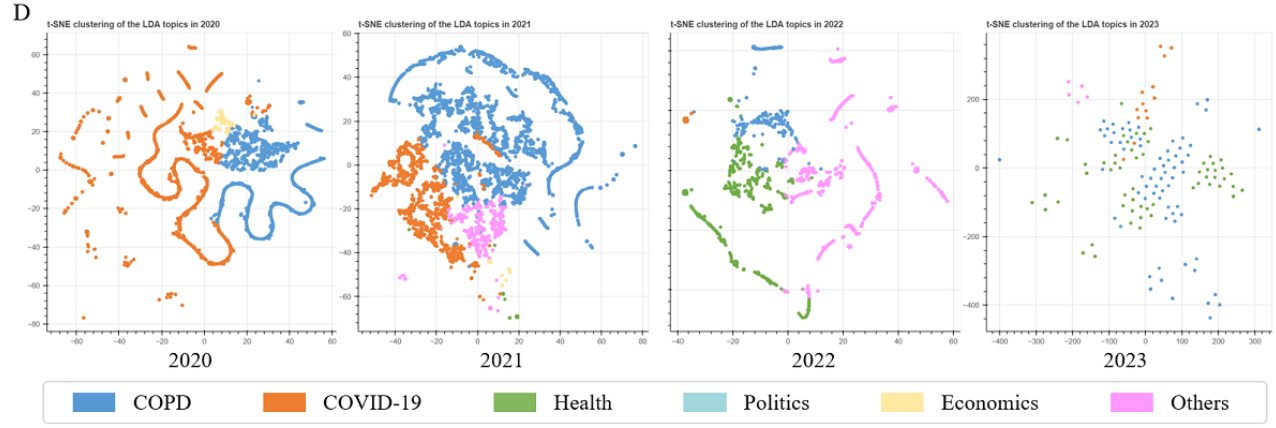

Sentiment Analysis

To gain an insight into the long-term trends and disparities of sentiments among different groups, we conducted a sentiment analysis spanning from January 2020 to June 2023. The sentiment score, calculated by VADER, was used to quantify the sentiment intensity of a tweet. This score ranges from –1 (most negative) to 1 (most positive), with [–1, –0.1) indicating negative, [–0.1, 0.1] indicating neutral, and (0.1, 1] indicating positive [36].

Sentiment Trends of the Populations With COPD

The sentiment scores from both COPD and non-COPD users are shown in Figure 4A, revealing a consistent pattern of lower sentiments in the long term among tweets posted by COPD users compared to those posted by non-COPD users. At the outbreak of COVID-19 in January 2020, the sentiment scores of both the population groups were below –0.05. Between July 2020 and June 2023, the sentiment scores of COPD users remained below 0.1, whereas those of non-COPD users exceeded 0.1. Notably, as the pandemic gradually subsided in 2023, the sentiment scores of non-COPD users increased progressively. In contrast, the sentiment scores of the COPD group still stagnated at a low level.

XSL•FO
RenderX

Figure 4. Results of the sentiment trend analysis. (A) Mean sentiment scores with 95% CIs of daily tweets from COPD and non-COPD users over time. The horizontal dashed line (sentiment score = 0.1) is the dividing line between positive (pink area) and neutral (white area). The background bar chart shows the number of COVID-19 weekly deaths. (B) Relative sentiment trends of COPD users, as reflected by the estimated coefficients with 95% CIs calculated by the difference-in-difference model. COPD: chronic obstructive pulmonary disease.

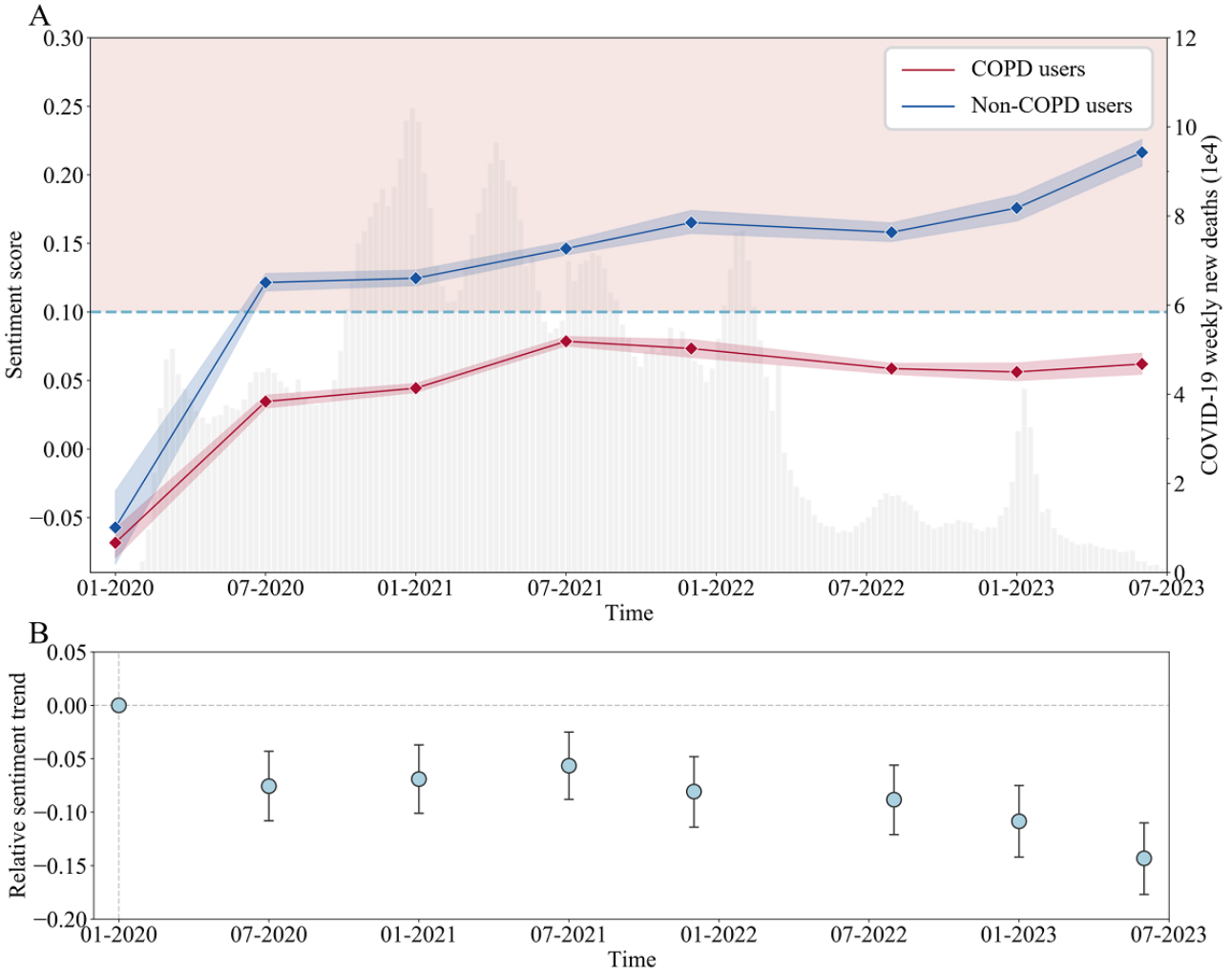

To accurately understand the relative sentiment trends in the COPD group in contrast with those in the non-COPD group, we employed a DiD model [34] to calculate the estimated coefficients between them. The results calculated by the DiD model can be found in Table S7 in Multimedia Appendix 1. Figure 4B shows that these estimated coefficients remained constantly negative, except for the initial time point, which was set to 0. These coefficients gradually approached 0 until July 2021 but significantly decreased afterward, reaching the lowest point in June 2023. This indicated a continuing relatively downward sentiment trend in the COPD group compared to that in the non-COPD group, and this trend was more pronounced after July 2021.

Sentiment Disparities Among Different Population Groups

As shown in Figure 5, sentiment scores differ among demographic groups. The average sentiment score of females was higher than that of males; yet, the difference was statistically insignificant (P=.08). Among age groups, those aged 40 years and older exhibited much lower sentiments than all the other 3 younger groups. As for the occupation, COPD users with OC2 occupations received lower sentiment scores than those with OC1 occupations (P=.03). In short, male COPD users aged 40 years and older with lower socioeconomic status (OC2) had the lowest sentiment in their daily tweets.

XSL•FO
RenderX

Figure 5. Sentiment comparisons of COPD users among population groups. COPD: chronic obstructive pulmonary disease. OC: occupation category; ns: not significant. *P<.05.

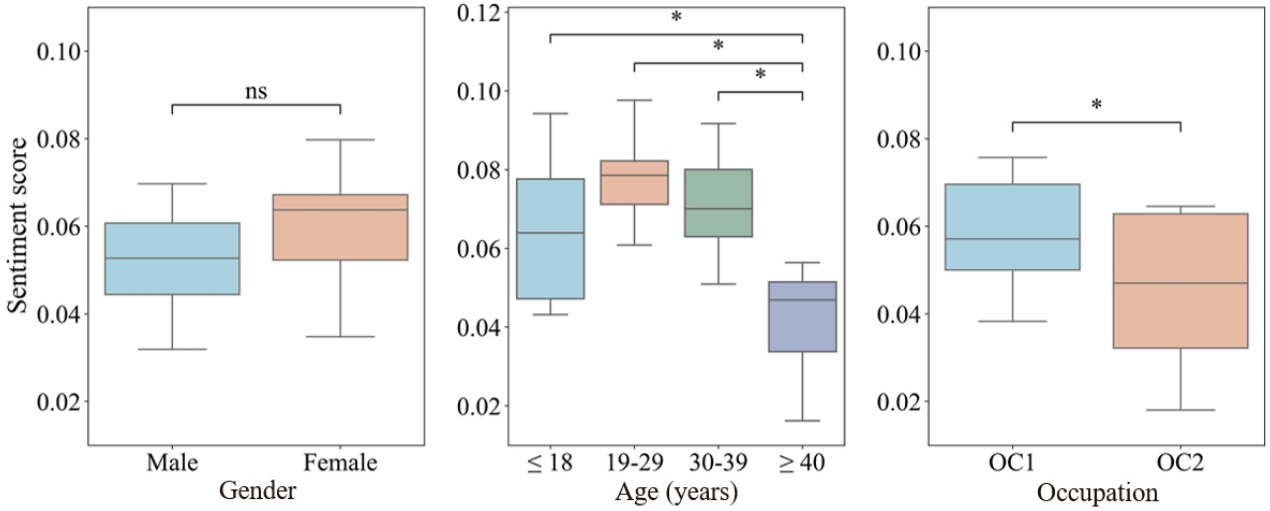

To identify emotion patterns in the sentiments of various population groups, we proposed a new method to calculate the values of the sentiment components of Plutchik's 8 emotions, denoted as sentiment component scores. Detailed descriptions of the emotion pattern method can be found in the Data Analysis Tools. We selected 3 representative timepoints corresponding to the different phases of COVID-19: the initial phase in January 2020, when the pandemic was declared public health emergency of international concern by the World Health Organization; the middle phase in July 2021, when the sentiments of COPD users reached the highest point; and the post-COVID phase in June 2023, after the pandemic was declared over. In the initial phase, fear was the predominant negative emotion among both COPD and non-COPD users. In the middle phase, the sentiment component scores of positive emotions such as joy, trust, and surprise of non-COPD users were significantly higher than those of COPD users. In the post-COVID era, the sentiment component scores of joy and trust of non-COPD users exceeded those of COPD users by far, while the sentiment component score of fear of COPD users became even more negative.

As shown in Figure 6, we studied the sentiment component scores of different demographic groups across the 3 COVID-19 phases. The results indicated that females exhibited significantly lower scores of joy than males in the initial phase (January 2020), whereas no obvious difference was observed in the middle phase (July 2021) and post-COVID phase (June 2023). Regarding age, those aged 40 years and older displayed lower sentiment component scores in most emotions, especially fear, compared to those younger than 40 years. As for occupations, those with lower socioeconomic status (OC2) demonstrated lower sentiment component scores than those with higher socioeconomic status (OC1).

XSL•FO
RenderX

Figure 6. Sentiment component scores of Plutchik's 8 emotions for different population groups across 3 COVID-19 phases. COPD: chronic obstructive pulmonary disease; OC: occupation category.

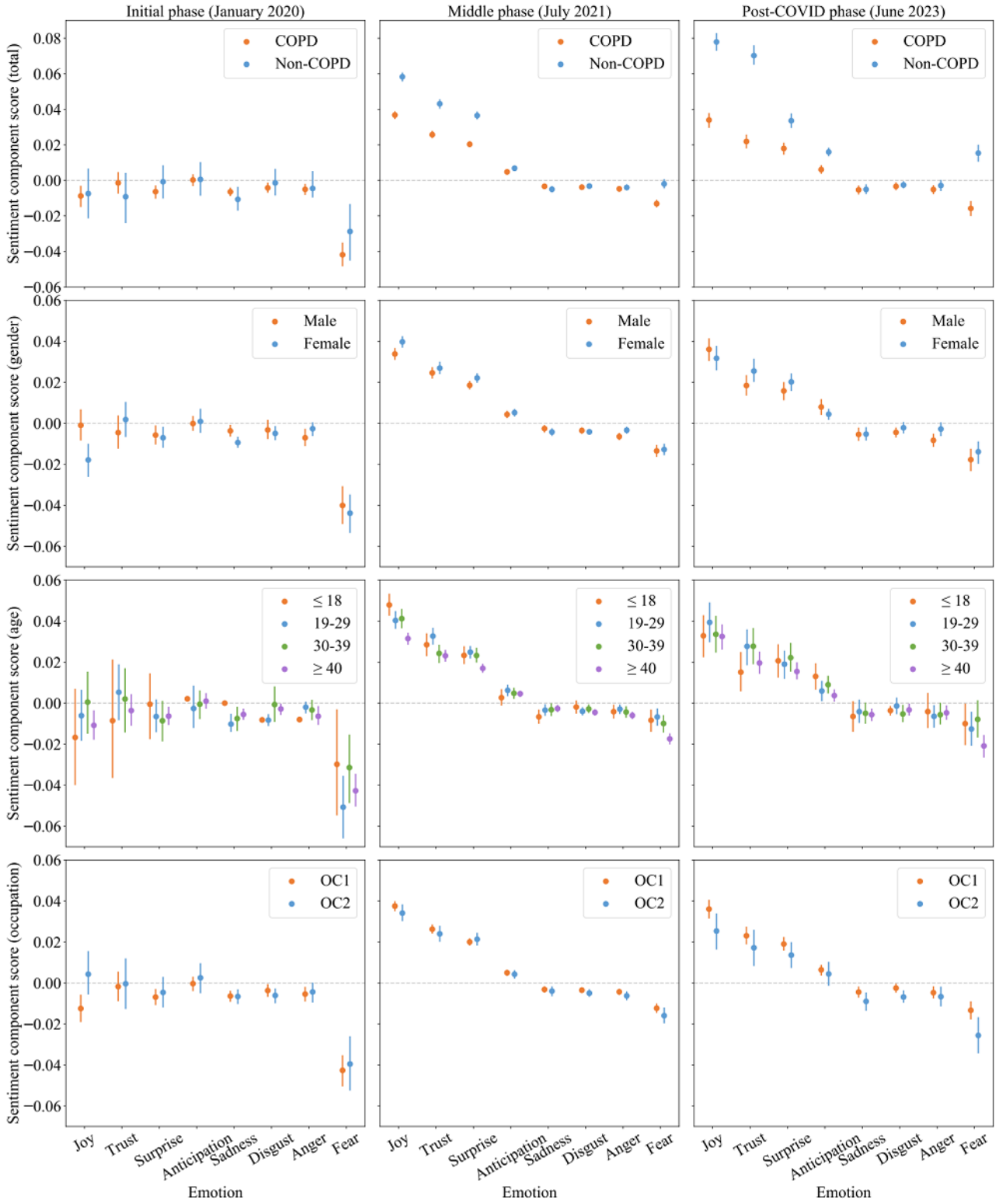

Discussion

Principal Findings

In this study, we designed a deep learning framework, Deep-COPD, incorporating a series of data retrieval and deep learning methods to investigate the psychological conditions of populations with COPD by using large-scale Twitter data collected during COVID-19 and beyond. First, we computed the attentiveness toward COPD among populations with different demographic features. Subsequently, we identified the popular concerns by hashtags and LDA topics in tweets posted by COPD users in 4 years. Finally, we examined the sentiment trends and prevalent emotions of populations with COPD in contrast to those of populations without COPD.

To the best of our knowledge, this study is the first research effort dedicated to exploring the psychological trends of

XSL•FO
RenderX

populations with COPD during and after COVID-19. Unlike in the existing investigations into the psychological impacts of COVID-19 on patients with COPD via conventional survey approaches, where the participant cohorts comprised only a few hundred individuals [10-14], our study intelligently recognized a cohort of 15,347 COPD users as participants. Moreover, by leveraging large-scale Twitter data spanning from January 2020 to June 2023, our study offers invaluable insights into the long-term psychological trends among populations with COPD in the pandemic.

By executing a comparative study between COPD and non-COPD users, this study shows that gender, age, and occupation affected the attentiveness toward COPD. The attentiveness toward COPD was lower in females than in males, significantly higher in adults aged 40 years and older than those younger than 40 years, and significantly higher in individuals of lower socioeconomic status than those of higher socioeconomic status. These interesting findings are aligned with the scientific statistics on the prevalence of COPD. In terms of gender, the prevalence of COPD is reported to be about 5% higher in males than in females [37]. As for age, the prevalence of COPD is appreciably higher in those aged 40 years and older compared to that of COPD in those younger than 40 years, with the incidence progressively increasing with age [38]. For occupation, individuals with lower socioeconomic status (OC2) were more likely to have poorer COPD outcomes than those with higher socioeconomic status (OC1) [8].

Through concern analysis, a notable decline in the popularity of COVID-related concerns among COPD users was observed after 2021. This phenomenon can be attributed to the fact that the SARS-CoV-2 virus undergoes constant mutations. With the emergence of the Omicron variant in November 2021, there has been a decline in the death rate [39]. As more knowledge of the virus has been gained, the associated attention has subsided. Concurrently, a steady rise in health-related concerns among COPD users was noted throughout the course of the pandemic. This trend can be attributed to the higher risk of patients with COPD to more severe diseases under COVID-19 (63%) compared to 33.4% among patients without COPD [8].

By tracking the sentiment trends from January 2020 to June 2023, we found that the sentiments of COPD users were much lower than those of non-COPD users across our study period. Remarkably, through DiD analysis, we noticed an increasingly significant downward trend in sentiments among COPD users compared to non-COPD users after the middle phase of COVID-19 (July 2021). This is an important warning that deserves our close attention. By analyzing the emotion patterns of 3 representative timepoints, this study further demonstrates that negative emotions, including fear, sadness, disgust, and anger, all contributed to the low sentiments of COPD users. Notably, in the post-COVID era (June 2023), COPD users exhibited decreased levels of joy and trust and increased fear compared to their levels of joy and trust during the middle phase of COVID-19. In particular, males, older adults, and individuals with lower socioeconomic status showed heightened fear compared to their counterparts. These results indicated that the populations with COPD experience a much longer impact from COVID-19 than the healthy population.

Last but not the least, similar psychological challenges have also been reported in studies on other chronic conditions under COVID [40-42]. Among all the chronic disease groups studied, including COPD, diabetes, and cardiovascular disease, the patients experienced long-term emotional distress during COVID-19, reinforcing the importance of psychological interventions tailored to different demographic and health characteristics.

Implications

Considering the implications for practical scenarios, it is clear that patients with COPD need targeted mental intervention and support. The findings of this work reveal a notable surge of fear among COPD users in the post-COVID era, stemming from their health-related concerns. In view of this, we suggest improving the situation in 2 ways to address the emotional challenges faced by patients with COPD. First, offline health care organizations are encouraged to actively offer easy-to-use mental health screenings, interventions, and self-management measures tailored to different population groups in addition to COPD treatments. For example, tools such as the Hospital Anxiety Depression Scale [43] or the Depression Anxiety Stress Scale [44] can be employed to identify patients with COPD who may have mental health concerns, thereby providing suitable guidance and periodic monitoring. Second, by utilizing the potential of social media platforms, web-based health care organizations and influential users can disseminate accessible psychological content, including videos, games, and helpful resources, to raise awareness of mental health and promote positive lifestyles. This approach involves targeted delivery to COPD users according to the distinct demographic and psychological patterns of different groups. By synergizing these offline and web-based approaches, health care organizations and professionals may effectively tailor their intervention measures to alleviate the psychological toll on populations with COPD.

Limitations

This study has several limitations. First, the identification of COPD users was solely based on self-reported data from Twitter, which may not fully capture all individuals with COPD. Second, this study relies on the VADER tool for sentiment analysis, which, despite its high accuracy, has limitations due to its static lexicon. This restricts its ability to adapt to evolving language trends or capture domain-specific terminology, potentially impacting the accuracy of sentiment classification. Finally, this study is limited to English-language tweets, potentially excluding non–English-speaking populations with COPD and missing cultural variations in psychological responses. These factors should be considered when interpreting the generalizability of our findings. Future research could address these limitations by incorporating data from multiple social media platforms, using more comprehensive methods to verify user health conditions, analyzing multilingual content to capture a broader and more diverse population with COPD, and exploring adaptive or domain-specific sentiment analysis tools to enhance sentiment detection accuracy.

XSL•FO
RenderX

Conclusions

This study proposes a novel approach to tracking the psychological trends of COPD users via large-scale Twitter mining. This study reveals that although COVID-19 has been declared to be ended, COPD users remain emotionally distressed, with much higher levels of fear than the healthy population. Therefore, strategies for the prevention and treatment of psychological problems in patients with COPD should be tailored to the characteristics of different population groups. The findings in this study offer valuable guidance to health care providers, policymakers, and researchers on the necessary measures to address the psychological needs of patients with COPD in the post-COVID era. The method implemented through our deep learning approach can be easily repurposed to study the psychological effects of other social events from a long-term perspective.

Data Availability

The source code for the deep learning framework is available at [45]. In compliance with Twitter's agreements and policy, we can only share data about Twitter user IDs and tweet IDs for academic purposes. These data are available through data access agreements from the corresponding authors.

Authors' Contributions

Conceptualization: CZ (lead), TW (equal)
Data curation: CZ (lead), CD (supporting), DD (supporting), SX (supporting)
Formal analysis: CZ (lead), TW (equal), CD (supporting), DD (supporting), LZ (supporting), ZL (supporting), SX (supporting)
Methodology: CZ (lead), TW (equal)
Investigation: CZ (lead), TW (equal), CD (supporting), DD (supporting), LZ (supporting), ZL (supporting)
Software: CZ
Writing – original draft: CZ (lead), TW (equal)
Writing – review & editing: CZ (lead), TW (equal)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information on the survey of related studies, details of some methods used in this study, and the results of some methods.
[DOCX File , 33 KB - jmir_v27i1e54543_app1.docx]

Multimedia Appendix 2
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.
[DOCX File , 33 KB - jmir_v27i1e54543_app2.docx]

References

1. COVID-19 Public Health Emergency of International Concern (PHEIC). World Health Organization. 2020 Feb 12. URL: https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-%28pheic%29-global-research-and-innovation-forum [accessed 2023-10-01]
2. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic. World Health Organization. 2023 May 5. URL: https://tinyurl.com/bdj6rm46 [accessed 2023-10-01]
3. Wise J. COVID-19: WHO declares end of global health emergency. BMJ 2023 May 09;381:1041. [doi: 10.1136/bmj.p1041] [Medline: 37160309]
4. Impact of COVID-19 on people's livelihoods, their health and our food systems. World Health Organization. 2020 Oct 13. URL: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems [accessed 2023-10-01]
5. Brodeur A, Gray D, Islam A, Bhuiyan S. A literature review of the economics of COVID-19. J Econ Surv 2021 Sep;35(4):1007-1044 [FREE Full text] [doi: 10.1111/joes.12423] [Medline: 34230772]
6. Press VG, Gershon AS, Sciurba FC, Blagev DP. Concerns about coronavirus disease-related collateral damage for patients with COPD. Chest 2020 Sep;158(3):866-868 [FREE Full text] [doi: 10.1016/j.chest.2020.05.549] [Medline: 32473947]
7. Leung JM, Niikura M, Yang CWT, Sin DD. COVID-19 and COPD. Eur Respir J 2020 Aug;56(2):2002108 [FREE Full text] [doi: 10.1183/13993003.02108-2020] [Medline: 32817205]

XSL•FO
RenderX

8. Alqahtani JS, Oyelade T, Aldhahir AM, Alghamdi SM, Almehmadi M, Alqahtani AS, et al. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. PLoS One 2020;15(5):e0233147 [FREE Full text] [doi: 10.1371/journal.pone.0233147] [Medline: 32392262]
9. Yohannes AM. COPD patients in a COVID-19 society: depression and anxiety. Expert Rev Respir Med 2021 Jan;15(1):5-7. [doi: 10.1080/17476348.2020.1787835] [Medline: 32578464]
10. Pedrozo-Pupo JC, Campo-Arias A. Depression, perceived stress related to COVID, post-traumatic stress, and insomnia among asthma and COPD patients during the COVID-19 pandemic. Chron Respir Dis 2020;17:1479973120962800 [FREE Full text] [doi: 10.1177/1479973120962800] [Medline: 33000648]
11. Daynes E, Gerlis C, Briggs-Price S, Jones P, Singh SJ. COPD assessment test for the evaluation of COVID-19 symptoms. Thorax 2021 Feb;76(2):185-187. [doi: 10.1136/thoraxjnl-2020-215916] [Medline: 33148809]
12. Hume E, Armstrong M, Manifield J, McNeillie L, Chambers F, Wakenshaw L, et al. Impact of COVID-19 shielding on physical activity and quality of life in patients with COPD. Breathe (Sheff) 2020 Sep;16(3):200231 [FREE Full text] [doi: 10.1183/20734735.0231-2020] [Medline: 33447283]
13. Sykes DL, Faruqi S, Holdsworth L, Crooks MG. Impact of COVID-19 on COPD and asthma admissions, and the pandemic from a patient's perspective. ERJ Open Res 2021 Jan;7(1):00822-2020 [FREE Full text] [doi: 10.1183/23120541.00822-2020] [Medline: 33575313]
14. Scarlata S, Cardaci V, Santangelo C, Matarese M, Cesari M, Antonelli Incalzi R. Distancing measures in COVID-19 pandemic: loneliness, more than physical isolation, affects health status and psycho-cognitive wellbeing in elderly patients with chronic obstructive pulmonary disease. COPD 2021 Aug;18(4):443-448 [FREE Full text] [doi: 10.1080/15412555.2021.1941834] [Medline: 34180766]
15. Gardner PJ, Moallef P. Psychological impact on SARS survivors: Critical review of the English language literature. Canadian Psychology / Psychologie canadienne 2015 Feb;56(1):123-135. [doi: 10.1037/a0037973]
16. Pradeepa S, Manjula KR, Vimal S, Khan MS, Chilamkurti N, Luhach AK. DRFS: detecting risk factor of stroke disease from social media using machine learning techniques. Neural Process Lett 2020 Jun 09;55(4):3843-3861. [doi: 10.1007/s11063-020-10279-8]
17. Fu J, Li C, Zhou C, Li W, Lai J, Deng S, et al. Methods for analyzing the contents of social media for health care: scoping review. J Med Internet Res 2023 Jun 26;25:e43349 [FREE Full text] [doi: 10.2196/43349] [Medline: 37358900]
18. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
19. Zhang C, Xu S, Li Z, Liu G, Dai D, Dong C. The evolution and disparities of online attitudes toward COVID-19 vaccines: year-long longitudinal and cross-sectional study. J Med Internet Res 2022 Jan 21;24(1):e32394 [FREE Full text] [doi: 10.2196/32394] [Medline: 34878410]
20. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
21. Volume streams. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/introduction [accessed 2023-07-12]
22. Wang Z, Hale S, Adelani D, Grabowicz P, Hartmann T, Flöck F, et al. Demographic inference and representative population estimates from multilingual social media data. 2019 Presented at: The World Wide Web Conference; May 13-17; San Francisco, CA, USA p. 2056-2067.
23. Huang G, Liu Z, Maaten L, Weinberger K. Densely connected convolutional networks. 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26; Honolulu, HI, USA p. 4700-4708. [doi: 10.1109/CVPR.2017.243]
24. Preoţiuc-Pietro D, Lampos V, Aletras N. An analysis of the user occupational class through Twitter content. 2015 Presented at: 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing; July 1; Beijing, China p. 1754-1764. [doi: 10.3115/v1/P15-1169]
25. Standard occupational classification. Office for National Statistics. URL: https://www.ons.gov.uk/methodology/ [accessed 2023-07-12]
26. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research 2003;3:993-1022. [doi: 10.5555/944919.944937]
27. models.coherencemodel. Gensim. URL: https://radimrehurek.com/gensim/models/coherencemodel.html [accessed 2023-07-12]
28. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: ICWSM. 2014 May 16 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4; Ann Arbor, MI, USA p. 216-225. [doi: 10.1609/icwsm.v8i1.14550]
29. Colneric N, Demsar J. Emotion recognition on Twitter: Comparative study and training a unison model. IEEE Trans Affective Comput 2020 Jul 1;11(3):433-446. [doi: 10.1109/taffc.2018.2807817]
30. Plutchik R. A general psychoevolutionary theory of emotion. In: Theories of Emotion. New York: Elsevier; 1980:3-33.
31. Tenny S, Hoffman MR. Odds ratio. StatPearls 2023:1-10. [Medline: 28613750]

XSL•FO
RenderX

32. P-value: comprehensive guide to understand, apply, and interpret. GeeksforGeeks. URL: https://www.geeksforgeeks.org/p-value/ [accessed 2024-12-28]
33. Callaway B, Sant'Anna PH. Difference-in-differences with multiple time periods. Journal of Econometrics 2021 Dec;225(2):200-230. [doi: 10.1016/j.jeconom.2020.12.001]
34. Greenwood BN, Agarwal R. Matching platforms and HIV incidence: An empirical investigation of race, gender, and socioeconomic status. Management Science 2016 Aug;62(8):2281-2303. [doi: 10.1287/mnsc.2015.2232]
35. Do I need ethics approval for social media research? JMIR. URL: https://support.jmir.org/hc/en-us/articles/115001620728-Do-I-need-ethics-approval-for-social-media-research [accessed 2024-09-29]
36. Jain R, Kumar A, Nayyar A, Dewan K, Garg R, Raman S, et al. Explaining sentiment analysis results on social media texts through visualization. Multimed Tools Appl 2023;82(15):22613-22629 [FREE Full text] [doi: 10.1007/s11042-023-14432-y] [Medline: 36747895]
37. Varmaghani M, Dehghani M, Heidari E, Sharifi F, Moghaddam SS, Farzadfar F. Global prevalence of chronic obstructive pulmonary disease: systematic review and meta-analysis. East Mediterr Health J 2019 Mar 19;25(1):47-57 [FREE Full text] [doi: 10.26719/emhj.18.014] [Medline: 30919925]
38. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Global Initiative for Chronic Obstructive Lung Disease. URL: https://goldcopd.org/2023-gold-report-2/ [accessed 2023-07-12]
39. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. Nat Rev Microbiol 2023 Jun;21(6):361-379. [doi: 10.1038/s41579-023-00878-2] [Medline: 37020110]
40. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
41. Joensen LE, Madsen KP, Holm L, Nielsen KA, Rod MH, Petersen AA, et al. Diabetes and COVID-19: psychosocial consequences of the COVID-19 pandemic in people with diabetes in Denmark-what characterizes people with high levels of COVID-19-related worries? Diabet Med 2020 Jul;37(7):1146-1154 [FREE Full text] [doi: 10.1111/dme.14319] [Medline: 32392380]
42. O'Neil A, Nicholls SJ, Redfern J, Brown A, Hare DL. Mental health and psychosocial challenges in the COVID-19 pandemic: food for thought for cardiovascular health care professionals. Heart Lung Circ 2020 Jul;29(7):960-963 [FREE Full text] [doi: 10.1016/j.hlc.2020.05.002] [Medline: 32561126]
43. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
44. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation of Australia; 1995.
45. Two-stage framework: Deep-COPD. GitHub. URL: https://github.com/cyzhang87/COPD_analysis_code [accessed 2024-10-01]

Abbreviations

COPD: chronic obstructive pulmonary disease
DenseNet: dense convolutional network
DiD: difference-in-difference
IMDb: internet movie database
LDA: latent Dirichlet allocation
M3: multimodal, multilingual, and multiattribute
OC: occupation category
OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
VADER: valence aware dictionary and emotional reasoner

Edited by T de Azevedo Cardoso; submitted 15.11.23; peer-reviewed by U Lokala, R Rupasingha; comments to author 18.07.24; revised version received 28.10.24; accepted 20.01.25; published 05.03.25.

Please cite as:
Zhang C, Wang T, Dong C, Dai D, Zhou L, Li Z, Xu S
Exploring Psychological Trends in Populations With Chronic Obstructive Pulmonary Disease During COVID-19 and Beyond: Large-Scale Longitudinal Twitter Mining Study
J Med Internet Res 2025;27:e54543
URL: https://www.jmir.org/2025/1/e54543
doi:10.2196/54543
PMID:40053739

XSL•FO
RenderX

©Chunyan Zhang, Ting Wang, Caixia Dong, Duwei Dai, Linyun Zhou, Zongfang Li, Songhua Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Application of a 3D Fusion Model to Evaluate the Efficacy of Clear Aligner Therapy in Malocclusion Patients: Prospective Observational Study

Chaofeng Liu[1,2], PhD; Yan Liu[2], BA; Chunyan Yi[3], MSM; Tao Xie[1], PhD; Jingjun Tian[1], PhD; Peishen Deng[1], MSM; Changyu Liu[1], MSM; Yan Shan[1], MSM; Hangyu Dong[1], MSM; Yanhua Xu[1], PhD

[1]Yunnan Key Laboratory of Stomatology, Department of Orthodontics, Kunming Medical University & Affiliated Stomatological Hospital, Kunming, China

[2]Yunnan Key Laboratory of Stomatology, Department of Second Clinic, Kunming Medical University & Affiliated Stomatological Hospital, Kunming, China

[3]Department of Stomatology, Kunming Medical University & Affiliated Yan'an Hospital, Kunming, China

Corresponding Author:
Yanhua Xu, PhD
Yunnan Key Laboratory of Stomatology
Department of Orthodontics
Kunming Medical University & Affiliated Stomatological Hospital
C Building, Hecheng International
1088 Haiyuan Middle Road
Kunming, 650500
China
Phone: 86 0871 65330099 ext 8038
Fax: 86 0871 65338894
Email: xuyanhua18@163.com

Abstract

Background: Investigating the safe range of orthodontic tooth movement is essential for maintaining oral and maxillofacial stability posttreatment. Although clear aligners rely on pretreatment digital models, their effect on periodontal hard tissues remains uncertain. By integrating cone beam computed tomography–derived cervical and root data with crown data from digital intraoral scans, a 3D fusion model may enhance precision and safety.

Objective: This study aims to construct a 3D fusion model based on artificial intelligence software that matches cone beam computed tomography and intraoral scanning data using the Andrews' Six Element standard. The model will be used to assess the 3D effects of clear aligners on tooth movement, to provide a reference for the design of pretreatment target positions.

Methods: Between May 2022 and May 2024, a total of 320 patients who completed clear aligner therapy at our institution were screened; 136 patients (aged 13-35 years, fully erupted permanent dentition and periodontal pocket depth <3 mm) met the criteria. Baseline ("simulation") and posttreatment ("fusion") models were compared. Outcomes included upper core discrepancy (UCD), upper incisors anteroposterior discrepancy (UAP), lower Spee curve deep discrepancy (LSD), upper anterior teeth width discrepancy (UAW), upper canine width discrepancy (UCW), upper molar width discrepancy (UMW), and total scores. Subanalyses examined sex, age stage (adolescent vs adult), and treatment method (extraction vs nonextraction).

Results: The study was funded in May 2022, with data collection beginning the same month and continuing until May 2024. Of 320 initial participants, 136 met the inclusion criteria. Data analysis is ongoing, and final results are expected by late 2024. Among the 136 participants, 90 (66%) were female, 46 (34%) were male, 64 (47%) were adolescents, 72 (53%) were adults, 38 (28%) underwent extraction, and 98 (72%) did not. Total scores did not differ significantly by sex (mean difference 0.01, 95% CI –0.13 to 0.15; *P*=.85), age stage (mean difference 0.03, 95% CI –0.10 to 0.17; *P*=.60), or treatment method (mean difference 0.07, 95% CI –0.22 to 0.07; *P*=.32). No significant differences were found in UCD (mean difference 0.001, 95% CI –0.02 to 0.01; *P*=.90) or UAP (mean difference 0.01, 95% CI –0.03 to 0.00; *P*=.06) by treatment method. However, adolescents exhibited smaller differences in UCD, UAW, UCW, and UMW yet larger differences in UAP and LSD (df=134; *P*<.001). Extraction cases showed smaller LSD, UAW, and UCW but larger UMW differences compared with nonextraction (df=134; *P*<.001).

XSL•FO
RenderX

Conclusions: The 3D fusion model provides a reliable clinical reference for target position design and treatment outcome evaluation in clear aligner systems. The construction and application of a 3D fusion model in clear aligner orthodontics represent a significant leap forward, offering substantial clinical benefits while establishing a new standard for precision, personalization, and evidence-based treatment planning in the field.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2400094304, https://www.chictr.org.cn/hvshowproject.html?id=266090&v=1.0

(J Med Internet Res 2025;27:e67378) doi:10.2196/67378

KEYWORDS

clear aligners; CBCT; intraoral scanning; fusion model; artificial intelligence; efficacy evaluation; orthodontic treatment

Introduction

Background

Alveolar bone remodeling is closely linked to orthodontic treatment. Investigating the safe range of orthodontic tooth movement is crucial for the stability of the oral and maxillofacial system following such treatment. Previous studies aimed at evaluating tooth alignment have primarily focused on clinical crown alignment, often neglecting root alignment, except when using panoramic x-rays to assess root parallelism. However, panoramic x-rays exhibit limited accuracy in evaluating the relationship between roots and the surrounding alveolar bone [1,2].

Cone Beam Computed Tomography and 3D Digital Models

Cone beam computed tomography (CBCT) has emerged as the most widely used imaging technique capable of effectively distinguishing between soft and hard tissues, and it has been extensively applied in orthodontic treatment. Tsukiboshi et al [3] used CBCT images from 28 patients, generating skull and mandible images through surface rendering, and established standard ranges for these surfaces, thereby allowing clinicians to reliably quantify and visualize patients' 3D hard tissues of the face. While complete tooth models can be reconstructed from CBCT images, the accuracy of this data ranges from 0.1 to 0.5 mm, and the precision, particularly for the crown portion, does not meet clinical requirements [4-6]. In recent years, digitally reconstructed 3D dental models have demonstrated high accuracy when compared with traditional plaster models, along with the added benefit of reproducible measurements [7-9]. However, these models cannot capture data on the cervical and root portions of the teeth. Therefore, we speculate that fusing the CBCT-obtained data of the necks and roots of the teeth with intraoral scanning (IOS) data of the crowns may address this limitation.

Clear Aligner and Andrews Six Elements

Clear aligner orthodontic technology uses a goal-oriented pretreatment target position model; however, the effects of this technology on periodontal hard tissues during orthodontic treatment remain uncertain. Existing research has proposed several methods for evaluating orthodontic treatment outcomes, including the American Board of Orthodontics Objective Grading System and the Peer Assessment Rating Index [10,11]. Based on Angle's standard of ideal normal occlusion, Andrews conducted a study in the 1960s that examined the natural dentition of 120 permanent teeth that had not undergone orthodontic treatment. The six standards of normal occlusion represent the optimal natural state of occlusion and serve as the objectives of orthodontic treatment. Meanwhile, Andrews defined 6 standards of normal occlusion based on Angle's principles, encompassing 6 aspects (Textbox 1), which set treatment targets for teeth, dental arches, jaw relationships, and facial esthetics [12]. Previous studies and some scholars have indicated that Andrews' Six Elements possess advantages in assessing the effectiveness of orthodontic treatment, while the 3D positioning of treatment target positions has also proven effective [13,14].

Textbox 1. Elements.

- Element I: ideal dental arch (shape, length, crown torque, root position, Spee curve depth, surface contact)
- Element II: ideal front-to-back jaw relationship
- Element III: ideal horizontal jaw positioning
- Element IV: ideal vertical jaw positioning
- Element V: ideal chin protrusion
- Element VI: ideal occlusal relationship

Objectives

This study aims to develop a 3D fusion model that integrates CBCT data of the cervical and root portions with IOS data of the crown portions, using AI software per Andrews' Six Elements. We hypothesize that this fusion model will accurately assess the 3D effects of clear aligners on tooth movement, providing a reliable reference for designing pretreatment target positions and ultimately enhancing the precision and safety of orthodontic treatments.

XSL•FO
RenderX

Methods

Participant Recruitment

This study was funded in May 2022. We are randomly recruiting patients who have completed orthodontic treatment using clear aligners at the Department of Orthodontics of Kunming Medical University Affiliated Stomatological Hospital. As of the submission of this paper, 320 patients have been recruited. To ensure diversity, patients were purposefully stratified according to sex, age, ethnicity, treatment methods, and types of appliances. Members of the research team invited patients to participate in person. All patients had signed informed consent forms for orthodontic treatment, allowing the use of their anonymized data for research purposes. The collected data included digital intraoral scan models, x-ray images, and clinical photographs. The digital 3D dental models were exported from the iTero Element system and saved in standard template library (STL [3D SYSTEMS]) format. All patient data were encrypted and securely stored in the cloud to ensure confidentiality. Data collection and processing took place from May 2022 to May 2024.

Sample Selection

Inclusion criteria encompassed individuals aged 13 to 35 years with fully erupted teeth, a good periodontal condition, and a periodontal pocket depth lower than 3 mm. Exclusion criteria included individuals with lip and palate clefts, craniofacial deformities, or skeletal discrepancies that required orthognathic surgery. Based on these criteria, a total of 136 individuals were selected.

Model Construction

Materials and Software

This study used a CBCT scanner (NewTom VG, Aperio Services) to obtain detailed images of the maxillofacial region, along with IOS (iTero Element 2, Align Technology) to capture high-resolution 3D models of the dentition. The software used included the Mimics Innovation Suite (Version 23.0; Materialise) for processing the CBCT data, Geomagic Wrap (Version 2023.0; 3D Systems) for data fusion and model refinement, and specialized orthodontic planning software for computer-based tooth arrangement according to Andrews' Six Elements diagnostic system.

Protocol Steps

The data from the CBCT and IOS were collected and exported as digital imaging and communications in medicine (DICOM) and STL files, respectively. Within the Mimics software, the CBCT data underwent a series of processes including thresholding (1200-3045 HU), segmentation, and 3D model generation. This was followed by surface smoothing and jaw separation to create individual STL models for the upper and lower jaws. AI Edit Masks tool is used to manually correct the segmentation results, removing nondental tissues in order to reconstruct a 3D model of dental arches that contains only teeth [15]. AI Smooth tool is applied to enhance the model's surface, while the Split tool is used to separate the jaws, allowing for individual editing and optimization of each part. The output is in STL file format. The software Geomagic Wrap is then used to fuse the data [16]. Initial alignment was achieved through N-point registration by selecting at least 3 corresponding anatomical landmarks, treating the IOS models as moving objects and the CBCT models as fixed references. Refined alignment was subsequently performed using the Best-Fit Alignment feature, with specified parameters to ensure high accuracy. The aligned models were merged using a Boolean Union operation, prioritizing the IOS data for crown detail and the CBCT data for root structures. Mesh repairs and Laplacian smoothing were applied to enhance the surface quality of the fused models, which were ultimately exported as STL files. In the final step, the orthodontic planning software was used to segment individual teeth and arrange them virtually in accordance with Andrews' Six Elements, adjusting tooth positions and orientations to achieve optimal occlusion and aesthetics. The 3D model construction process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Digital dental imaging and data fusion workflow. This flowchart outlines the comprehensive workflow for digital dental imaging, integrating data from CBCT scans (NewTom VG) and intraoral scans (iTero Element 2). CBCT: cone beam computer tomography; DICOM: digital imaging and communications in medicine; STL: standard template library.

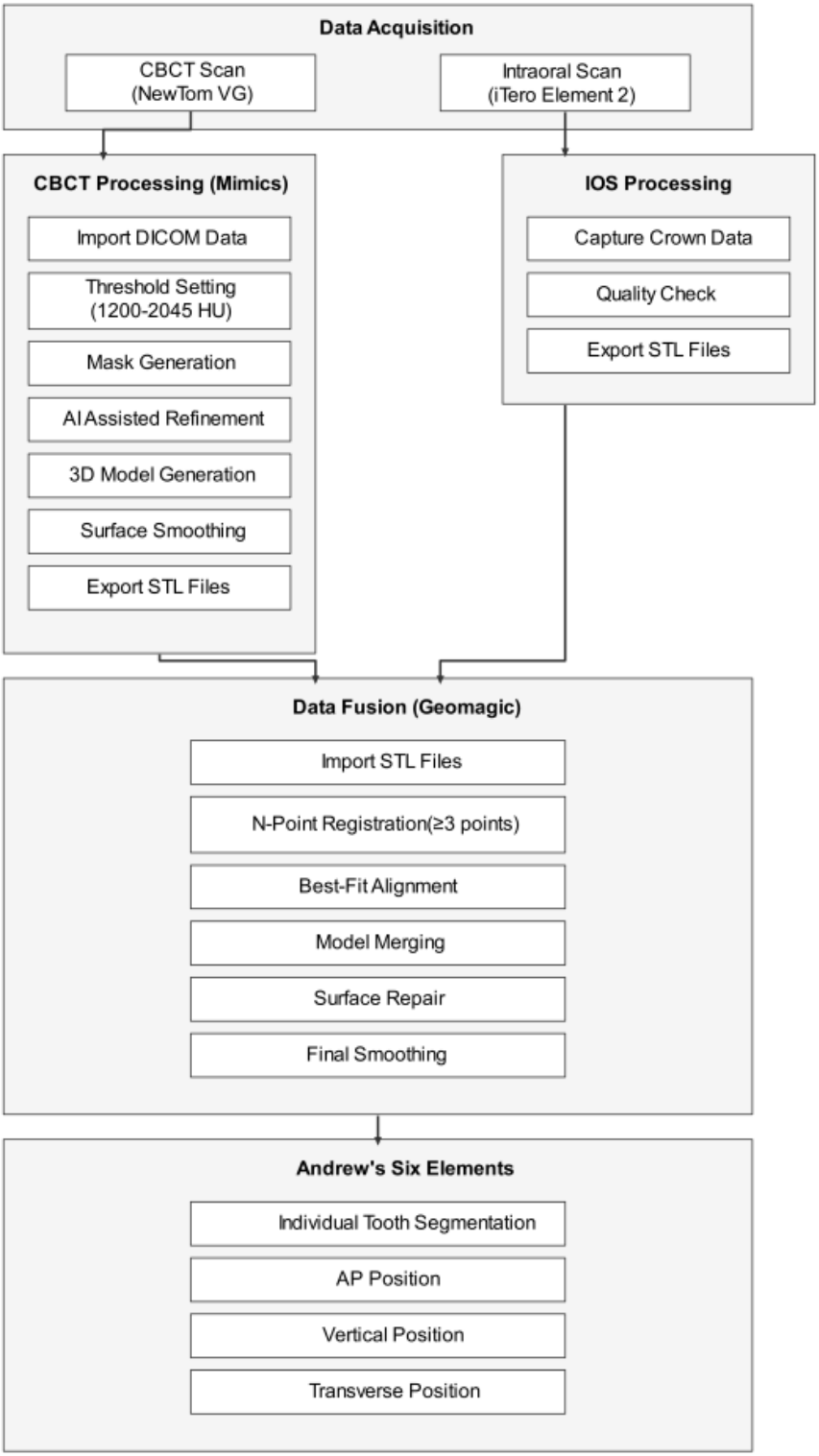

Patient Data Management

We ensured that all patient data were anonymized in compliance with data protection regulations and kept detailed records of all parameters and adjustments for each patient.

Calibration and Quality Control

We confirmed that all software versions used are compatible with each other. We were also aware that updates or changes in software versions may affect functionality. Scanners were calibrated regularly to maintain data accuracy. A quality control process to verify the accuracy of models at each step was also implemented.

XSL•FO
RenderX

Data Acquisition

CBCT and intraoral scans are performed to gather detailed dental data.

CBCT Processing

CBCT processing involves importing DICOM data, threshold setting, mask generation, AI-assisted refinement, 3D model generation, surface smoothing, and exporting STL files. IOS processing involves capturing crown data, conducting quality checks, and exporting STL files.

Data Fusion

Data fusion was done through Geomagic software to import STL files, perform N-point registration, best-fit alignment, model merging, surface repair, and final smoothing.

Analysis

For analysis, we applied Andrew's Six Elements for individual tooth segmentation and positional analysis, including anterior-posterior, vertical, and transverse positions.

Interexaminer Reliability

A total of 2 independent examiners, both practicing orthodontic specialists, each with at least 5 years of clinical and research experience, assessed the samples. Their identities remain confidential in accordance with ethical guidelines. Each examiner is a certified orthodontist with a solid foundation in 3D measurement techniques and orthodontic treatment evaluation. Before data collection, both examiners underwent the same standardized training protocol to ensure uniform measurement methods, including identifying standard landmark points and using specified software tools. A consistent set of guidelines and reference standards (including validated criteria from the literature) was used, guaranteeing that the same benchmarks were applied. Each examiner scored the samples independently and under blinded conditions, with the sequence of sample numbering randomized to mitigate potential observer bias. To evaluate interexaminer scoring consistency, the intraclass correlation coefficient (ICC) was used. Bland-Altman plots were used to illustrate any differences in scores between the examiners and to evaluate the presence of systematic bias. Discrepancies in measurements were revisited and, if necessary, remeasured or discussed to improve scoring consistency.

Participant Grouping

Between May 2022 and May 2024, a total of 320 patients, who successfully completed clear aligner treatment at the Department of Orthodontics, Kunming Medical University Affiliated Stomatological Hospital, were invited to participate in the study. Ultimately, 136 participants met the inclusion criteria. Of these participants, 90 (66%) were female and 46 (34%) were male, 64 (47%) were adolescents, and 72 (53%) were adults. Regarding extraction status, participants were assigned to either the extraction or nonextraction group based on the clinical treatment plan established at their initial assessment. Specifically, those requiring the removal of premolars, often to address moderate to severe crowding, resolve substantial discrepancies in arch length, or correct significant malocclusion, were categorized in the extraction group (n=38, 28%), whereas those whose treatment plans did not involve premolar removal were placed in the nonextraction group (n=98, 72%).

Measurement and Scoring

Based on Andrews' Six Elements criteria [12], measurements encompass the following 6 indicators. Each indicator is scored individually, and the total score is the sum of these individual scores (Figure 2).

In Figure 2, the dental arch core line is given(A), where L_1 is the desired length of the arch and L_2 is the existing length of the arch. The anteroposterior position of incisors is given in B, where FA is the center point of the clinical crowns of the upper mesial incisors, GALL is an imaginary line passing through the interbrow point and perpendicular to the ground plane when in the natural cephalic position, and "a" is the perpendicular distance from the FA point to GALL. Mandibular Spee curve deep is given in C, where d_1 and d_2 are the depths of the Spee curves on the left and right sides, respectively. The width of the dental arch is given in D, where A_w is the width of the anterior teeth, C_w is the width of the canine, and M_w is the width of the molar. Upper core discrepancy (UCD) is measured from the distal of the last molar, along the central fossa of the molars and premolars, the canine cusp, and the incisor incisal edge, connecting to the opposite side. Upper incisors anteroposterior discrepancy (UAP) refers to the anteroposterior distance between the target position and the posttreatment position of the clinical crown center point (FA point) of the maxillary central incisor. Lower Spee curve deep discrepancy (LSD) is a curve formed by connecting the incisor incisal edge, canine cusp, the buccal cusps of the molars and premolars, and the distobuccal cusp of the last molar. The depth of the lowest point of the curve is measured on both the left and right sides and the 2 values are summed to obtain the Spee curve depth. Upper anterior teeth width discrepancy (UAW) is defined as the distance between the midpoints of the incisal edges of the bilateral lateral incisors. Upper canine width discrepancy (UCW) measures the distance between the midpoints of the cusps of the bilateral canines. Upper molar width discrepancy (UMW) is the distance between the midpoints of the palatal cusps of the bilateral first molars.

XSL•FO
RenderX

Figure 2. Measurement and scoring diagnostic sheet.

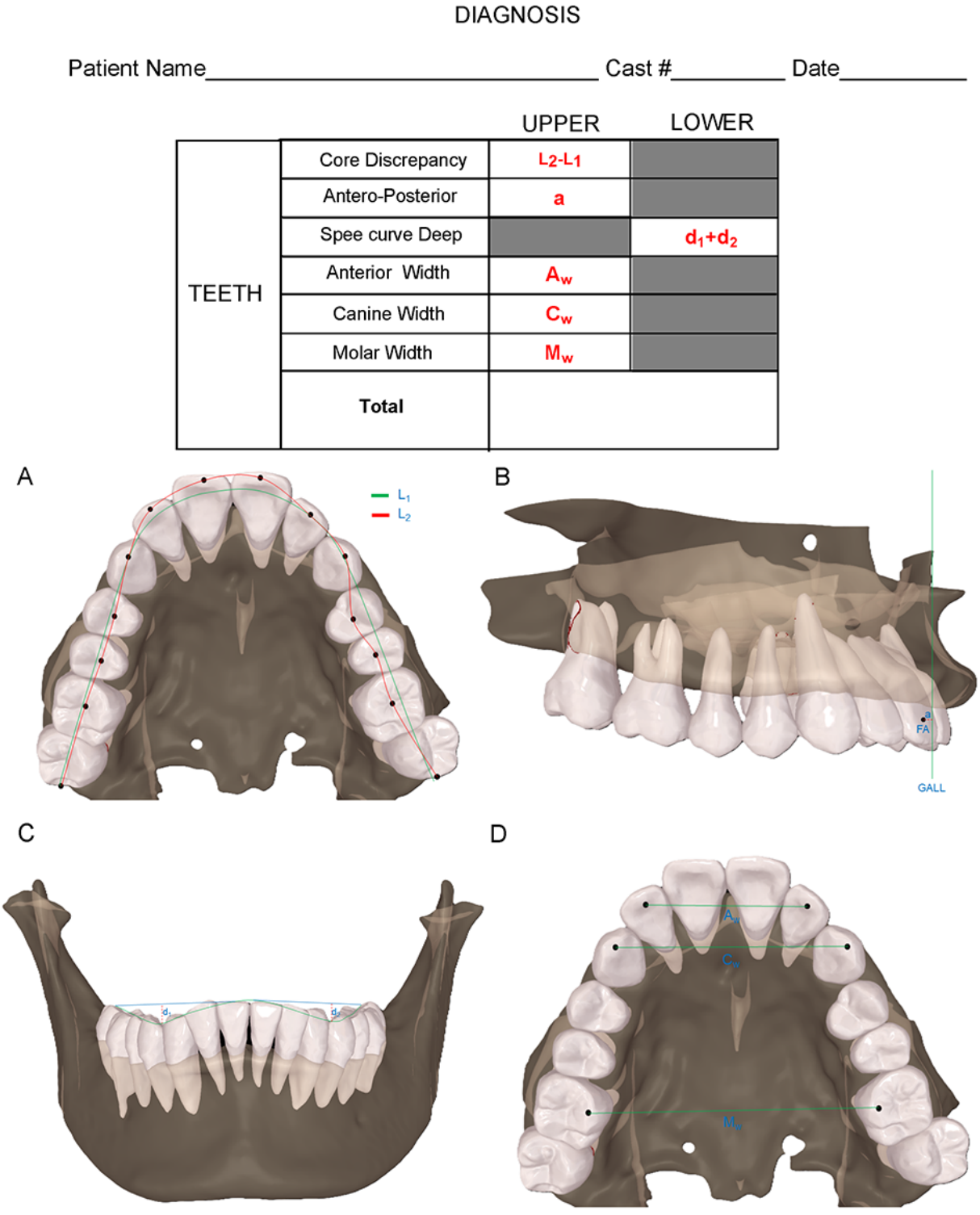

Ethical Considerations

The study was registered on the Chinese Clinical Trial Registry (registration ID ChiCTR2400094304) and was granted ethics approval from the Medical Ethics Committee (approval KYKQ2022MEC0046). All participants provided written informed consent before participating in the study. No compensation of any kind was provided to the participants.

Results

Overview

The study received funding in May 2022, with data collection commencing in the same month, and is projected to conclude in May 2024. At the time of approval, 320 participants had been recruited, of which 136 met the inclusion criteria based on subgroup selection. The research team is now in the final phase

XSL•FO
RenderX

of data analysis, having completed the advanced statistical modeling. Validation and quality checks are currently underway to ensure the accuracy and reliability of results before the study's conclusion. Final results are to be announced by the end of 2024.

Model Construction

A total of 136 patient data sets were processed to create accurate 3D dental models. CBCT and intraoral scan data were exported as DICOM and STL files, respectively. Using Mimics software, the CBCT data underwent thresholding, segmentation, and 3D model generation, followed by surface smoothing and jaw separation. The AI Edit Masks tool refined the segmentation, isolating dental tissues. Geomagic Wrap was then used for data fusion, achieving initial alignment through N-point registration and refining it with Best-Fit Alignment. The models were merged, enhancing surface quality through mesh repairs and Laplacian smoothing. Finally, orthodontic planning software segmented individual teeth and arranged them according to Andrews' 6 Elements, resulting in a successful 3D model construction, as shown in Figure 3.

Figure 3. Construction of 3D fusion model.

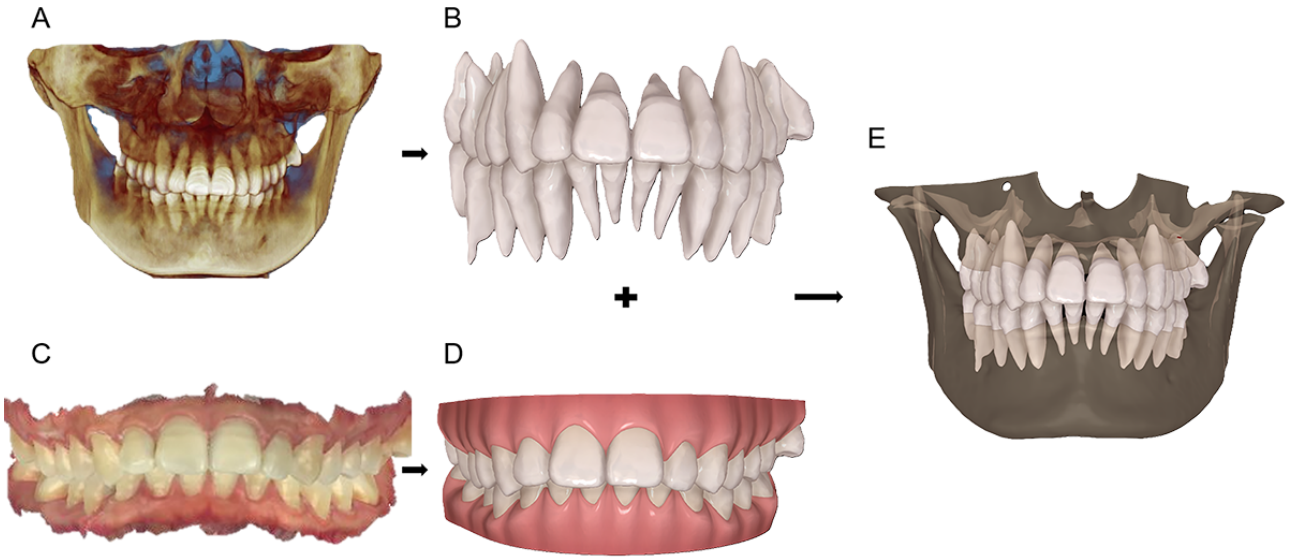

In Figure 3, the CBCT data (A) were imported into Mimics software using artificial intelligence (AI) tools to reconstruct a 3D model containing crowns and roots (B); the intraoral scan data (C) were imported into Invisalign simulator software using AI deep learning tools to reconstruct a 3D model containing crowns and gingiva (D). Then the AI tool was used in Geomagic Wrap software to combine the two to form a 3D fusion model (E).

Interexaminer Reliability

A total of 2 independent examiners assessed the same set of measurements following the standardized protocol described in the Methods section. A total of 136 samples (from 136 participants) were evaluated, resulting in 272 total measurement sets. The overall ICC for the interexaminer agreement was 0.98 (95% CI 0.96-0.99), indicating an excellent level of consistency between the 2 examiners' scores. A Bland-Altman plot was generated to compare the scores of Examiner A and Examiner B, visually illustrating the difference plotted against the average of the 2 measurements (Figure 4). The mean difference (bias) between the examiners' measurements was 0.02 (SD 0.11) mm. The 95% CI of agreement extended from –0.20 mm to 0.24 mm, suggesting minimal systematic discrepancy. No trend was observed in the differences as a function of the measurement magnitude, implying that the 2 examiners were consistent across the entire range of measurements. Collectively, these results demonstrate a high level of agreement (ICC=0.98) and minimal systematic bias between the examiners across all measured parameters. The rigorous application of a uniform measurement protocol, coupled with examiner blinding and random sample order, contributed to the strong reliability indices. Consequently, the scoring system and methodology can be deemed robust for evaluating tooth movement outcomes in this study.

The data measured by 2 independent examiners are visually represented, with each dot signifying an individual measurement and the horizontal line indicating the mean. The near overlap in the distribution and mean values for both examiners suggests that their measurements are nearly identical. Furthermore, an ICC of 0.98, indicating extremely high consistency, confirms that the 2 examiners evaluated the samples in a closely aligned manner, reflecting excellent interrater reliability.

XSL•FO
RenderX

Figure 4. Consistency test for scores between 2 examiners.

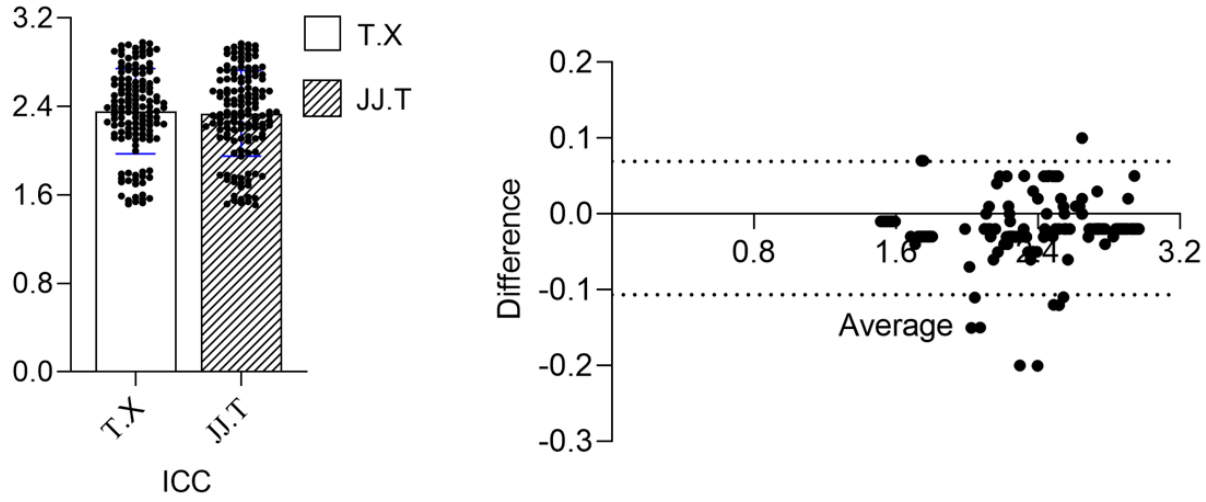

Group and Statistical Analysis

Subjects were categorized by sex, age stage, and treatment method, with "Adolescent" defined as individuals aged 13-17 years and "Adult" as individuals aged 18-35 years (Table 1). Based on the scores, we calculated the mean and SD for individual and total scores across various groupings, using independent samples *t* tests for comparison. The confidence interval for all analyses was set at 95%. Statistical analyses were performed using GraphPad Prism (version 9; GraphPad Software, Inc).

Table 1 lists the distribution for total samples. A total of 136 samples were selected, and individuals were categorized by sex (female and male), age stage (adolescent and adult), and treatment method (extraction and nonextraction).

Table 1. The groups of objects (N=136).

Groups	Values, n (%)
Sex	
Female	90 (66)
Male	46 (34)
Age	
Adolescent	64 (47)
Adult	72 (53)
Treatment method	
Extraction	38 (28)
Nonextraction	98 (72)

Measurement and Scoring

Sex Group

Among the different sex groups (Figure 5), the male group comprised 46 cases aged 13-35 years (mean age 21.30 years), with a total mean score of 2.35 (SD 0.35), a minimum score of 1.53, and a maximum score of 2.97. The female group included 90 cases aged 13-35 years (mean age 21.00 years), with a total mean score of 2.36 (SD 0.40), a minimum score of 1.52, and a maximum score of 2.98. A comparison of total mean scores between the 2 groups revealed a *t* value of 0.19 and a *P* value of .85, indicating no statistically significant difference.

In Figure 5, the total score distribution for sex groups is given. The male group had a total mean score of 2.35 (SD 0.35) and the female group had a total mean score of 2.36 (SD 0.40). Comparing the total mean scores between the two groups, a *P* value of .85 >.05 indicated no statistically significant difference.

XSL•FO
RenderX

Figure 5. Comparison of total scores in different sex groups.

Groups	Number, n (%)	Mean	SD	t value	*P* value	OR, 95% CI
Male	46 (34)	2.35	0.35	0.19	.85	2.39 [2.24~2.45]
Female	90 (66)	2.36	0.40			2.40 [2.28~2.45]

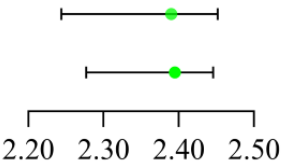

Age Stage Group

Among the different age stage groups (Figures 6 and 7), the adolescent group comprised 64 cases aged 13-17 (mean 13.58) years, with a total mean score of 2.34 (SD 0.34), a minimum score of 1.52, and a maximum score of 2.91. The adult group included 72 cases aged 18-35 (mean 26.53) years, with a total mean score of 2.37 (SD 0.42), a minimum score of 1.58, and a maximum score of 2.98. A comparison of total mean scores between the two groups showed a *t* value of 0.52 and a *P* value of .60, indicating no statistically significant difference. However, in the individual score items, significant statistical differences (*P*<.001) were observed in the following indicators: UCD, UAP, LSD, UAW, UCW, and UMW.

Figure 6. The score distribution for adolescent and adult samples. LSD: lower Spee curve deep discrepancy; UAP: upper incisors antero-posterior discrepancy; UAW: upper anterior teeth width discrepancy; UCD: upper core discrepancy; UCW: upper canine width discrepancy; UMW: upper molar width discrepancy.

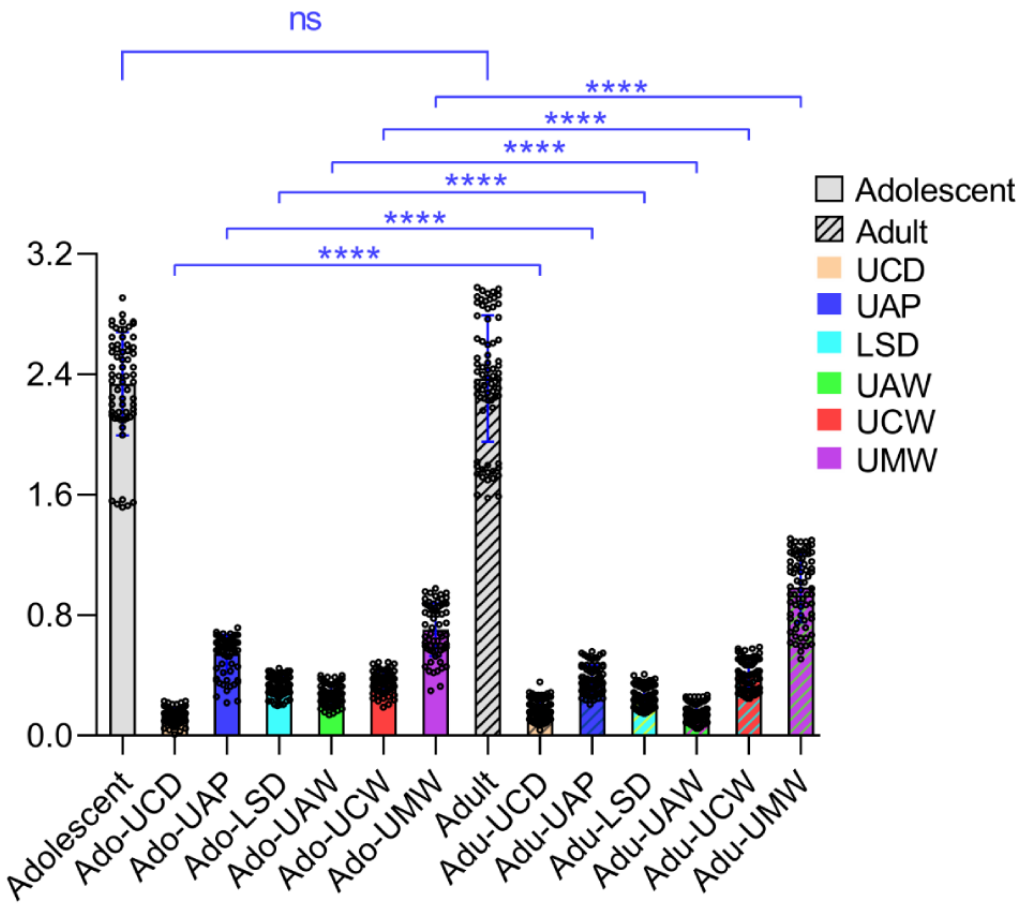

In Figure 7, the score distribution for age stage groups is given. The adolescent group had a total mean score of 2.34 (SD 0.34), and the adult group had a total mean score of 2.37 (SD 0.42). Comparing the total mean scores between the two groups (*P*>.05) indicated no statistically significant difference. In the individual score items, significant statistical differences (*P*<.001) were observed in the following indicators: UCD, UAP, LSD, UAW, UCW, and UMW.

XSL•FO
RenderX

Figure 7. Comparison of scores in different age stage groups. LSD: lower Spee curve deep discrepancy; UAP: upper incisors antero-posterior discrepancy; UAW: upper anterior teeth width discrepancy; UCD: upper core discrepancy; UCW: upper canine width discrepancy; UMW: upper molar width discrepancy.

Measurement indicators	Adolescent (n=64)		Adult (n=72)				RR, 95% CI
	Mean	SD	Mean	SD	t	*P* value	
UCD	0.13	0.05	0.18	0.07	4.09	< .001	0.111 [-0.065~-0.022]
UAP	0.54	0.13	1.37	0.09	8.85	< .001	0.367 [-0.209~-0.133]
LSD	0.33	0.06	0.26	0.07	5.56	< .001	0.187 [-0.089~-0.042]
UAW	0.27	0.07	0.16	0.06	9.78	< .001	0.417 [-0.130~-0.086]
UCW	0.36	0.07	0.42	0.10	3.98	< .001	0.106 [0.030~0.089]
UMW	0.71	0.18	0.98	0.23	7.82	< .001	0.313 [0.208~0.349]
Total Score	2.34	0.34	2.37	0.42	0.52	.60	0.002 [-0.096~0.166]

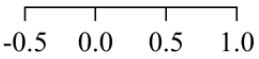

Treatment Methods Group

Among the different treatment method groups (Figures 8 and 9), the extraction group consisted of 38 cases, aged 13 to 35 (mean 19.74) years, with a total mean score of 2.30 (SD 0.36), a minimum score of 1.53, and a maximum score of 2.90. The nonextraction group included 98 cases, aged 13 to 35 (mean 21.63) years, with a total mean score of 2.38 (SD 0.40), a minimum score of 1.52, and a maximum score of 2.98. A comparison of the total mean scores between the 2 groups yielded a t value of 0.99 and a *P* value of .32 (*P*>.05), indicating no statistically significant difference. In the individual scoring items, there were no statistically significant differences (*P*>.05) in the comparisons of UCD and UAP. However, significant statistical differences were observed (*P*<.001) in the comparisons of LSD, UAW, UCW, and UMW.

Figure 8. The score distribution for extraction and nonextraction samples. LSD: lower Spee curve deep discrepancy; UAP: upper incisors antero-posterior discrepancy; UAW: upper anterior teeth width discrepancy; UCD: upper core discrepancy; UCW: upper canine width discrepancy; UMW: upper molar width discrepancy.

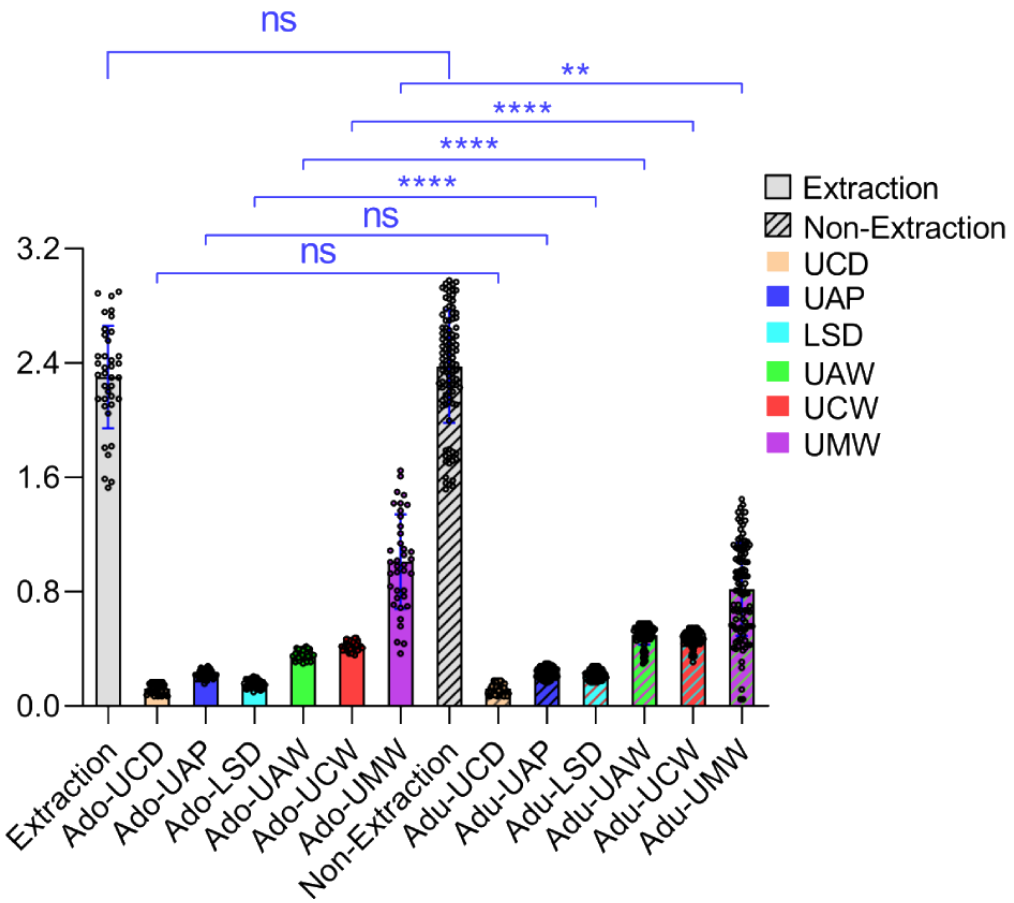

XSL•FO
RenderX

Figure 9 demonstrates the score distribution for different treatment method groups. The extraction group had a total mean score of 2.30 (SD 0.36) and the adult group had a total mean score of 2.38 (SD 0.40). Comparing the total mean scores, UCD and UAP between the two groups (P>.05) indicated no statistically significant difference. In the individual score items, significant statistical differences (P<.001) were observed in the following indicators: LSD, UAW, UCW, and UMW.

Figure 9. Comparison of scores in different treatment methods groups. LSD: lower Spee curve deep discrepancy; UAP: upper incisors antero-posterior discrepancy; UAW: upper anterior teeth width discrepancy; UCD: upper core discrepancy; UCW: upper canine width discrepancy; UMW: upper molar width discrepancy.

Measurement indicators	Extraction (n=38)		Non-Extraction (n=98)				
	Mean	SD	Mean	SD	t	*P*	RR, 95% CI
UCD	0.12	0.04	0.12	0.03	0.12	.90	0.001 [-0.012~0.014]
UAP	0.23	0.03	0.24	0.04	1.91	.06	0.026 [-0.026~0.001]
LSD	0.16	0.03	0.22	0.03	11.08	< .001	0.478 [-0.078~-0.055]
UAW	0.36	0.03	0.50	0.07	11.76	< .001	0.508 [-0.159~-0.113]
UCW	0.42	0.03	0.48	0.05	6.02	< .001	0.213 [-0.070~-0.035]
UMW	1.01	0.33	0.82	0.33	7.82	< .001	0.067 [0.070~0.318]
Total Score	2.30	0.36	2.38	0.40	0.99	.32	0.007 [-0.219~0.072]

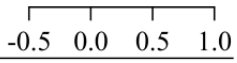

Discussion

Principal Findings

Based on the findings, a 3D fusion model was successfully constructed by integrating CBCT data of the cervical and root portions with IOS data of the crown portions, using AI software in accordance with Andrews' Six Elements Criteria. This model effectively assessed the 3D effects of clear aligners on tooth movement and provided a reliable reference for designing pretreatment target positions. The results indicated that neither sex nor age had a significant effect on total scores; however, adolescents exhibited significant differences in specific parameters, including UCD, UAP, LSD, UAW, UCW, and UMW. Similarly, while the treatment method did not significantly affect total scores, it did have a significant impact on individual items such as LSD and the buccolingual widths of certain teeth. Overall, the study validates the effectiveness of the 3D fusion model in enhancing the precision and personalization of orthodontic treatments using clear aligners, thereby establishing a new standard for evidence-based treatment planning in the field.

Model Construction

Based on pretreatment CBCT and intraoral scans, a digital full-arch dental model integrating IOS and CBCT was constructed. 3D tooth arrangement was performed in accordance with Andrews' Six Keys diagnostic system, successfully establishing a pretreatment 3D integrated prediction model. This model encompasses 3D images that include root and alveolar bone information. By comparing the differences between pre- and posttreatment 3D integrated models, the actual 3D tooth movement following clear aligner treatment can be examined. Baan et al [8] conducted structured light scanning, CBCT scanning, and IOS on 10 dry human skulls, achieving notable clinical accuracy in the integration of CBCT and IOS. Deferm et al [5] registered intraoral scans with CBCT for 8 dentate patients and 14 edentulous patients, finding that the average error for dentate jaws was 0.49 (SD 0.26) mm, for edentulous jaws was 0.16 (0.08) mm, and for alveolar ridges was 0.16 (0.05) mm, indicating high precision in the registration of intraoral scans with CBCT. These studies provide a theoretical foundation for the modeling used in this research, The results of this study are consistent with the above studies.

Potential limitations in model construction, particularly regarding the accuracy of the CBCT-IOS fusion technique, may significantly impact treatment planning. First, image registration errors present a major challenge. When fusing IOS data with CBCT data, discrepancies can arise, resulting in a 3D model that does not accurately reflect the patient's actual anatomical structures [8,17-19]. Such errors may stem from slight patient movements during scanning, limitations in equipment precision, or deficiencies in the registration algorithms used. Second, differences in resolution between the 2 scanning methods can affect the accuracy of the fused model. CBCT typically provides lower-resolution images, while IOS offers high-resolution surface detail. This mismatch in resolution can lead to deviations in the detailed representation within the fused model, thereby impacting the assessment of fine anatomical structures [6,20-22]. These accuracy limitations can influence treatment planning in several ways.

Inaccurate Localization of Anatomical Structures

Misalignments in the model may lead to incorrect judgments regarding the positions of teeth, bones, and other critical anatomical features, adversely affecting the precision of surgical or orthodontic interventions.

Impact on the Formulation of Treatment Plans

Developing treatment strategies based on inaccurate models may result in suboptimal outcomes or even complications. To mitigate these impacts, it is essential in clinical practice to enhance scanning and fusion techniques. This can be achieved by using higher precision equipment, refining registration

XSL•FO
RenderX

algorithms, and performing manual corrections when necessary to ensure the accuracy and reliability of the models used in treatment planning.

Sex

Research indicates that the sex demographic variable does not influence treatment outcomes. Xie et al [23] assessed the clinical effectiveness of SmartTrack aligners and found that sex did not affect anterior tooth rotation movement. Scott et al [24] discovered that Damon3 self-ligating brackets resulted in less discomfort during initial tooth alignment, regardless of sex. Tepedino et al [25] evaluated the predictability of the Nuvola aligner system and concluded that sex did not influence the results. Mario et al [26] investigated the accuracy of the F22 Aligner system and found that sex did not play a significant role. Taner et al [27] evaluated pretreatment dental models, lateral cephalometric measurements, and wrist x-rays, discovering that sex did not affect the outcomes of skeletal and dental cephalometric measurements. Finally, Mandall et al [6] found that sex did not influence orthodontic patients' compliance and treatment motivation. The above research is consistent with our findings.

Age Stage

In this research, adolescents scored lower than adults, suggesting that adolescents may have a higher bone remodeling capacity to meet tooth movement driven by orthodontic forces. Large amounts of clinical studies suggested that adolescents experience faster tooth movement and improved treatment outcomes compared with adults, as their bone tissue responds more readily to mechanical stress, allowing for quicker bone resorption and formation. Kanou et al [28] discovered that younger individuals display accelerated rates of bone remodeling. Zheng et al [29] analyzed differences in alveolar bone support between adolescents and adults, noting that adults exhibited reduced alveolar bone support posttreatment. Kalina et al [30] assessed changes in the lower incisor alveolar bone in both adolescent and adult patients, finding that while alveolar bone thickness decreased in both groups when lower anterior teeth were proclined or retracted, the reduction was less pronounced in adolescents. Furthermore, Kuc et al [31] evaluated maxillary morphological changes resulting from incisor movement and revealed that adolescents demonstrated significantly greater maxillary remodeling capacity than adults following such movement. These results may suggest that adolescents are more effective than adults in moving teeth, a phenomenon attributed to the biological adaptability of their dental arches, which enhances the efficacy of orthodontic treatment, particularly in adjusting the transverse width of the dental arch [32]. During this phase of growth and development, adolescents exhibit higher skeletal plasticity, facilitating more efficient bone remodeling throughout orthodontic treatment.

In this research, we found that adolescents exhibit better alignment of teeth, as well as sagittal and transverse width adjustments of the incisors and molars compared with adults. This phenomenon may be attributed to the intrusion of anterior teeth along the long axis during clinical treatment, which causes the entire periodontal tissue to move coronally, resulting in a higher rate of tooth movement. Adolescents possess a stronger capacity for bone remodeling. Some researchers have longitudinally assessed changes in intercanine and intermolar widths from childhood to adulthood, discovering that both maxillary and mandibular arch intercanine and intermolar widths increase significantly between the ages of 3 and 13 years [12,33-37]. Following the complete eruption of permanent dentition, arch widths experience a slight decrease, with intercanine width decreasing more than intermolar width. The mandibular intercanine width is generally established by 8 years of age after the eruption of the 4 incisors. After the eruption of permanent dentition, clinicians should anticipate either no change or slight decreases in arch width. Although arch width changes from birth to adulthood, the magnitude and direction of these changes do not provide sufficient scientific evidence to support the expansion of dental arches beyond their established dimensions for typical patients. This is consistent with the results of this research. Nevertheless, a recent study was conducted to evaluate the effectiveness of Invisalign in reducing deep overbite in adolescent patients [38]. The findings indicated that the reduction in overlap achieved during the correction process was less than half of what the orthodontist had anticipated or planned. Furthermore, the age of the patients did not significantly influence the accuracy of overlap reduction. Notably, a trend was observed indicating a more pronounced overlap reduction in patients who did not use the bite ramp; however, this trend did not reach statistical significance.

Treatment Methods

In this research, the extraction group scored lower than the nonextraction group, suggesting that extraction may not significantly affect effective tooth movement. Kirschneck et al [39] studied adolescents with borderline cases treated with tooth extraction and found no significant difference in the improvement of vertical relationships between extraction and nonextraction treatments in orthodontic patients. Xu et al [40] compared Chinese borderline cases with and without extraction treatment and discovered that extraction treatment may lead to differences in facial profile at the end of treatment, but no differences were observed in tooth alignment, overbite, overjet, midline, or posterior occlusion. Extraction treatment may have advantages over nonextraction treatment regarding the sagittal, vertical, and anteroposterior alignment of the anterior teeth; however, it is slightly inferior concerning the horizontal width adjustment of the posterior teeth. This may be due to extraction treatment resulting in the medial movement of molars and the distal movement of anterior teeth, which causes changes in the buccolingual width among molars, incisors, and canines. Extraction may provide additional space for adjusting the Spee curve. Elias et al [41] compared the effects of extracting premolars versus nonextraction treatment on dental arch width, contour, treatment duration, occlusal outcomes, smile aesthetics, and stability. They found that extraction led to decreased intermolar width in both arches and upper or lower lip retraction, while nonextraction treatment resulted in increased intercanine width in the lower jaw and a shorter treatment time [41]. Shafique et al [42] recruited borderline cases and found that orthodontic treatment involving tooth extraction had a significant effect on the vertical dimension. The above research results enrich this study.

XSL•FO
RenderX

Limitations and Prospects

The limitations of this study must be acknowledged. As a retrospective study, strict inclusion and exclusion criteria were established to minimize the selection bias inherent in all retrospective investigations, but they cannot completely eliminate bias in personalized prediction models. Future research could further integrate AI technologies and use big data analysis to enhance the predictive capabilities of personalized models. While the growing demand for esthetic and modern orthodontic solutions are facilitated by technological advancements, such as 3D imaging, planning, and printing, it does not specifically address potential limitations in model construction. In particular, the accuracy limitations of the CBCT-IOS fusion technique are not discussed. These limitations can impact treatment planning by affecting the precision of digital models used for diagnosis and aligner fabrication. Inaccuracies in the fusion process may lead to suboptimal fit of aligners or misalignment in tooth movement predictions, potentially compromising treatment outcomes.

Conclusions

The fusion model that integrates cervical and root sections from CBCT with crown data from IOS ensures highly accurate measurements and assessments of tooth positioning. This precision is crucial in orthodontics, where even minor deviations can significantly impact clinical outcomes. The use of a 3D model, facilitated by AI, enables practitioners to tailor treatments more effectively to individual patients. By validating the safe range of orthodontic movement, these models help mitigate the risk of adverse effects on periodontal hard tissues. Such stability is essential for long-term success, reducing the likelihood of relapse or complications post treatment. The study's findings, including the adaptive differences between adolescents and adults and the effects of extraction versus nonextraction methods, provide invaluable insights. Practitioners can rely on these data-driven insights to make informed decisions that optimize individual patient care. The integration of AI and 3D modeling technology introduces new methodologies that have the potential to transform traditional orthodontic practices. This technological advancement not only enhances diagnostic capabilities but also streamlines the treatment process, making it more efficient and patient-friendly. In conclusion, the construction and application of 3D fusion models in clear aligner orthodontics represent a significant leap forward, offering substantial clinical benefits while establishing a new standard for precision, personalization, and evidence-based treatment planning in the field.

Acknowledgments

The authors would like to thank the Department of Second Clinic for providing imaging and IOS data. Support with obtaining funding was provided by authors XYH and LCF. This research was funded by the Reserve Program of Yunnan Province Young and Middle-aged Academic and Technical Leaders (grant 202205AC160067), the Youth Project of Yunnan Province Oral Diseases Clinical Medical Research Center (grant 2022QN003), the Kunming Health Science and Technology Talent Cultivation Project (grant 2022-08-03-002), and the Kunming Medical University Teaching Research and Reform Project (grant 2022-YJ-Y-143).

Authors' Contributions

LCF initiated the study and played a significant role in the design, data generation, analysis, and interpretation of the results. LCF is also a major contributor to the manuscript writing. XT, TJJ, DPP, LCY, SY, and DHY participated in data generation, analysis, and interpretation. In addition, XYH contributed to the study design, data interpretation, and manuscript writing.

Conflicts of Interest

None declared.

References

1. Chiqueto K, Janson G, de Almeida CT, Storniolo JM, Barros SE, Henriques JF. Influence of root parallelism on the stability of extraction-site closures. Am J Orthod Dentofacial Orthop 2011;139(6):e505-e510. [doi: 10.1016/j.ajodo.2010.11.019] [Medline: 21640862]
2. Pittayapat P, Willems G, Alqerban A, Coucke W, Ribeiro-Rotta RF, Souza PC, et al. Agreement between cone beam computed tomography images and panoramic radiographs for initial orthodontic evaluation. Oral Surg Oral Med Oral Pathol Oral Radiol 2014;117(1):111-119 [FREE Full text] [doi: 10.1016/j.oooo.2013.10.016] [Medline: 24332335]
3. Tsukiboshi Y, Tanikawa C, Yamashiro T. Surface-based 3-dimensional cephalometry: an objective analysis of cranio-mandibular morphology. Am J Orthod Dentofacial Orthop 2020;158(4):535-546. [doi: 10.1016/j.ajodo.2019.09.017] [Medline: 32826124]
4. Fonseca C, Cavadas F, Fonseca P. Upper airway assessment in cone-beam computed tomography for screening of obstructive sleep apnea syndrome: development of an evaluation protocol in dentistry. JMIR Res Protoc 2023;12:e41049 [FREE Full text] [doi: 10.2196/41049] [Medline: 37145857]
5. Deferm JT, Nijsink J, Baan F, Verhamme L, Meijer G, Maal T. Soft tissue-based registration of intraoral scan with cone beam computed tomography scan. Int J Oral Maxillofac Surg 2022;51(2):263-268 [FREE Full text] [doi: 10.1016/j.ijom.2021.04.004] [Medline: 33933335]

XSL•FO
RenderX

6. Mandall NA, Matthew S, Fox D, Wright J, Conboy FM, O'Brien KD. Prediction of compliance and completion of orthodontic treatment: are quality of life measures important? Eur J Orthod 2008;30(1):40-45. [doi: 10.1093/ejo/cjm077] [Medline: 17971352]
7. Mack S, Bonilla T, English JD, Cozad B, Akyalcin S. Accuracy of 3-dimensional curvilinear measurements on digital models with intraoral scanners. Am J Orthod Dentofacial Orthop 2017;152(3):420-425. [doi: 10.1016/j.ajodo.2017.05.011] [Medline: 28863923]
8. Baan F, Bruggink R, Nijsink J, Maal TJJ, Ongkosuwito EM. Fusion of intra-oral scans in cone-beam computed tomography scans. Clin Oral Investig 2021;25(1):77-85 [FREE Full text] [doi: 10.1007/s00784-020-03336-y] [Medline: 32495223]
9. O'Toole S, Bartlett D, Keeling A, McBride J, Bernabe E, Crins L, et al. Influence of scanner precision and analysis software in quantifying three-dimensional intraoral changes: two-factor factorial experimental design. J Med Internet Res 2020;22(11):e17150 [FREE Full text] [doi: 10.2196/17150] [Medline: 33245280]
10. Djeu G, Shelton C, Maganzini A. Outcome assessment of Invisalign and traditional orthodontic treatment compared with the American Board of Orthodontics objective grading system. Am J Orthod Dentofacial Orthop 2005;128(3):292-298. [doi: 10.1016/j.ajodo.2005.06.002] [Medline: 16168325]
11. Richmond S, Shaw WC, Roberts CT, Andrews M. The PAR Index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards. Eur J Orthod 1992;14(3):180-187. [doi: 10.1093/ejo/14.3.180] [Medline: 1628684]
12. Andrews LF. The 6-elements orthodontic philosophy: treatment goals, classification, and rules for treating. Am J Orthod Dentofacial Orthop 2015;148(6):883-887. [doi: 10.1016/j.ajodo.2015.09.011] [Medline: 26672688]
13. Resnick CM, Kim S, Yorlets RR, Calabrese CE, Peacock ZS, Kaban LB. Evaluation of andrews' analysis as a predictor of ideal sagittal maxillary positioning in orthognathic surgery. J Oral Maxillofac Surg 2018;76(10):2169-2176. [doi: 10.1016/j.joms.2018.03.013] [Medline: 29654777]
14. Liu CF, Xu Y, Liu Y. Curative effect evaluation by using andrews' six elements for completed orthodontic treatment cases. Hua Xi Kou Qiang Yi Xue Za Zhi 2021;39(1):48-52 [FREE Full text] [doi: 10.7518/hxkq.2021.01.007] [Medline: 33723936]
15. Asif MK, Nambiar P, Khan IM, Aziz ZABCA, Noor NSBM, Shanmuhasuntharam P, et al. Enhancing the three-dimensional visualization of a foreign object using mimics software. Radiol Case Rep 2019;14(12):1545-1549 [FREE Full text] [doi: 10.1016/j.radcr.2019.10.001] [Medline: 31719943]
16. Han X, Wei D, Jiang X, Di P, Yi C, Lin Y. Digital registration versus cone-beam computed tomography for evaluating implant position: a prospective cohort study. BMC Oral Health 2024;24(1):304 [FREE Full text] [doi: 10.1186/s12903-024-04088-x] [Medline: 38438985]
17. Kim S, Yongjin C, Jincheol N, In-Seok S, You-Sun L, Bo-Yeon H, et al. Best of both modalities: fusing CBCTIntraoral scan data into a single tooth image. Fusing CBCT and Intraoral Scan Data Into a Single Tooth Image. 2024. Presented at: Medical Image Computing and Computer Assisted Intervention 2024 Oct 4:553-563 [FREE Full text] [doi: 10.1007/978-3-031-72069-7_52]
18. Ntovas P, Marchand L, Finkelman M, Revilla-León M, Att W. Accuracy of manual and artificial intelligence-based superimposition of cone-beam computed tomography with digital scan data, utilizing an implant planning software: a randomized clinical study. Clin Oral Implants Res 2024 Jun 10;35(10):1262-1272. [doi: 10.1111/clr.14313] [Medline: 38858787]
19. Kim YJ, Ahn JH, Lim HK, Nguyen TP, Jha N, Kim A, et al. Novel procedure for automatic registration between cone-beam computed tomography and intraoral scan data supported with 3D segmentation. Bioengineering (Basel) 2023;10(11):1326-1326 [FREE Full text] [doi: 10.3390/bioengineering10111326] [Medline: 38002450]
20. de Waard O, Baan F, Verhamme L, Breuning H, Kuijpers-Jagtman AM, Maal T. A novel method for fusion of intra-oral scans and cone-beam computed tomography scans for orthognathic surgery planning. J Craniomaxillofac Surg 2016;44(2):160-166. [doi: 10.1016/j.jcms.2015.11.017] [Medline: 26732637]
21. Komuro A, Yamada Y, Uesugi S, Terashima H, Kimura M, Kishimoto H, et al. Accuracy and dimensional reproducibility by model scanning, intraoral scanning, and CBCT imaging for digital implant dentistry. Int J Implant Dent 2021 Jun 30;7(1):63-63 [FREE Full text] [doi: 10.1186/s40729-021-00343-w] [Medline: 34189641]
22. Li Y, Ye H, Ye F, Liu Y, Lv L, Zhang P, et al. The current situation and future prospects of simulators in dental education. J Med Internet Res 2021 Apr 8;23(4):e23635 [FREE Full text] [doi: 10.2196/23635] [Medline: 33830059]
23. Xie J, Liu F, Sang T, Wu J. Factors affecting the efficacy of Invisalign in anterior tooth rotation. Am J Orthod Dentofacial Orthop 2023 Apr;163(4):540-552. [doi: 10.1016/j.ajodo.2022.04.017] [Medline: 36566089]
24. Scott P, Sherriff M, Dibiase AT, Cobourne MT. Perception of discomfort during initial orthodontic tooth alignment using a self-ligating or conventional bracket system: a randomized clinical trial. Eur J Orthod 2008 Jun;30(3):227-232. [doi: 10.1093/ejo/cjm131] [Medline: 18339656]
25. Tepedino M, Paoloni V, Cozza P, Chimenti C. Movement of anterior teeth using clear aligners: a three-dimensional, retrospective evaluation. Prog Orthod 2018 Apr 2;19(1):9 [FREE Full text] [doi: 10.1186/s40510-018-0207-3] [Medline: 29607469]

XSL•FO
RenderX

26. Mario P, de Villagomez SS, Federica P, Cremonini F, Salvatore R, Lombardo L. Evaluation of tooth movement accuracy with the F22 aligner system: a retrospective study. Applied Sciences 2024 Feb 18;14(4):1641 [FREE Full text] [doi: 10.3390/app14041641]
27. Taner L, Gürsoy GM, Uzuner FD. Does gender have an effect on craniofacial measurements? Turk J Orthod 2019 Jun;32(2):59-64 [FREE Full text] [doi: 10.5152/TurkJOrthod.2019.18031] [Medline: 31294407]
28. Kanou K, Kitaura H, Noguchi T, Ohori F, Marahleh A, Kinjo R, et al. Effect of age on orthodontic tooth movement in mice. J Dent Sci 2024 Apr;19(2):828-836 [FREE Full text] [doi: 10.1016/j.jds.2023.09.016] [Medline: 38618134]
29. Zheng Y, Zhu C, Zhu M, Lei L. Difference in the alveolar bone remodeling between the adolescents and adults during upper incisor retraction: a retrospective study. Sci Rep 2022 Jun 1;12(1):9161 [FREE Full text] [doi: 10.1038/s41598-022-12967-y] [Medline: 35650260]
30. Kalina E, Grzebyta A, Zadurska M. Bone remodeling during orthodontic movement of lower incisors-narrative review. Int J Environ Res Public Health 2022 Nov 15;19(22):15002 [FREE Full text] [doi: 10.3390/ijerph192215002] [Medline: 36429721]
31. Kuc AE, Kotuła J, Nawrocki J, Kulgawczyk M, Kawala B, Lis J, et al. Bone remodeling of maxilla after retraction of incisors during orthodontic treatment with extraction of premolars based on CBCT study: a systematic review. J Clin Med 2024 Mar 5;13(5):1503 [FREE Full text] [doi: 10.3390/jcm13051503] [Medline: 38592367]
32. Dawood HM, Kroeger A, Chavda V, Chapple ILC, Kebschull M. Age-dependent cellular mechanisms of orthodontic tooth movement-induced root resorption. 2023 Mar Presented at: Presented at: 102nd General Session & Exhibition in conjunction with the 53rd Annual Meeting of the American Association for Dental, Oral, and Craniofacial Research and the 48th Annual Meeting of the Canadian Association for Dental Research; 2024 March 16; New Orleans, URL: https://europepmc.org/abstract/MED/37366151 [doi: 10.1093/ejo/cjad011]
33. Bishara SE, Jakobsen JR, Treder J, Nowak A. Arch width changes from 6 weeks to 45 years of age. Am J Orthod Dentofacial Orthop 1997 Apr;111(4):401-409. [doi: 10.1016/s0889-5406(97)80022-4] [Medline: 9109585]
34. Korn EL, Baumrind S. Transverse development of the human jaws between the ages of 8.5 and 15.5 years, studied longitudinally with use of implants. J Dent Res 1990 Jun;69(6):1298-1306. [doi: 10.1177/00220345900690061501] [Medline: 2355125]
35. Lux CJ, Conradt C, Burden D, Komposch G. Transverse development of the craniofacial skeleton and dentition between 7 and 15 years of age--a longitudinal postero-anterior cephalometric study. Eur J Orthod 2004 Feb;26(1):31-42. [doi: 10.1093/ejo/26.1.31] [Medline: 14994880]
36. Ballanti F, Lione R, Fanucci E, Franchi L, Baccetti T, Cozza P. Immediate and post-retention effects of rapid maxillary expansion investigated by computed tomography in growing patients. Angle Orthod 2009 Jan;79(1):24-29 [FREE Full text] [doi: 10.2319/012008-35.1] [Medline: 19123695]
37. Antoun JS, Mei L, Gibbs K, Farella M. Effect of orthodontic treatment on the periodontal tissues. Periodontol 2000 2017 Jun;74(1):140-157. [doi: 10.1111/prd.12194] [Medline: 28429487]
38. Blundell HL, Weir T, Meade MJ. Deep overbite reduction in adolescent patients treated with invisalign: a retrospective analysis. Am J Orthod Dentofacial Orthop 2024 Dec;166(6):515-523 [FREE Full text] [doi: 10.1016/j.ajodo.2024.07.008] [Medline: 39140923]
39. Kirschneck C, Proff P, Reicheneder C, Lippold C. Short-term effects of systematic premolar extraction on lip profile, vertical dimension and cephalometric parameters in borderline patients for extraction therapy--a retrospective cohort study. Clin Oral Investig 2016 May;20(4):865-874. [doi: 10.1007/s00784-015-1574-5] [Medline: 26349767]
40. Xu TM, Liu Y, Yang MZ, Huang W. Comparison of extraction versus nonextraction orthodontic treatment outcomes for borderline Chinese patients. Am J Orthod Dentofacial Orthop 2006 May;129(5):672-677. [doi: 10.1016/j.ajodo.2005.12.007] [Medline: 16679208]
41. Elias KG, Sivamurthy G, Bearn DR. Extraction vs nonextraction orthodontic treatment: a systematic review and meta-analysis. Angle Orthod 2024 Jan 1;94(1):83-106 [FREE Full text] [doi: 10.2319/021123-98.1] [Medline: 37899069]
42. Shafique HZ, Zaheer R, Jan A, Mughal AT, Shahid R, Ghaffar F, et al. Vertical skeletal changes after extraction and nonextraction orthodontic treatment. Eur J Dent 2023 Feb;17(1):227-233 [FREE Full text] [doi: 10.1055/s-0042-1749366] [Medline: 35785825]

Abbreviations

AI: artificial intelligence
CBCT: cone beam computer tomography
DICOM: Digital Imaging and Communications in Medicine
ICC: intraclass correlation coefficient
IOS: intraoral scanning
LSD: lower Spee curve deep discrepancy
STL: standard template library
UAP: upper incisors antero-posterior discrepancy

XSL•FO
RenderX

UAW: upper anterior teeth width discrepancy
UCD: upper core discrepancy
UCW: upper canine width discrepancy
UMW: upper molar width discrepancy

Edited by A Coristine; submitted 10.10.24; peer-reviewed by P Nalabothu, M Sycinska-Dziarnowska; comments to author 11.11.24; revised version received 11.12.24; accepted 23.12.24; published 15.01.25.

Please cite as:
Liu C, Liu Y, Yi C, Xie T, Tian J, Deng P, Liu C, Shan Y, Dong H, Xu Y
Application of a 3D Fusion Model to Evaluate the Efficacy of Clear Aligner Therapy in Malocclusion Patients: Prospective Observational Study
J Med Internet Res 2025;27:e67378
URL: https://www.jmir.org/2025/1/e67378
doi:10.2196/67378
PMID:

©Chaofeng Liu, Yan Liu, Chunyan Yi, Tao Xie, Jingjun Tian, Peishen Deng, Changyu Liu, Yan Shan, Hangyu Dong, Yanhua Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Geographical Disparities in Research Misconduct: Analyzing Retraction Patterns by Country

Paul Sebo[1], MSc, MD; Melissa Sebo[2]

[1]University Institute for Primary Care, University of Geneva, Geneva, Switzerland

[2]Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Paul Sebo, MSc, MD
University Institute for Primary Care
University of Geneva
Rue Michel-Servet 1
Geneva, 1211
Switzerland
Phone: 41 223794390
Email: paul.seboe@unige.ch

Abstract

This study examines disparities in research retractions due to misconduct, identifying countries with the highest retraction counts and those disproportionately represented relative to population and publication output. The findings emphasize the need for improved research integrity measures.

(J Med Internet Res 2025;27:e65775) doi:10.2196/65775

KEYWORDS

affiliation; country; fraud; integrity; misconduct; plagiarism; publication; research; retraction; ethical standards; ethics; research misconduct; literature

Introduction

Retractions are essential for maintaining scientific integrity, especially in cases of research misconduct [1-4]. Data from 2013 to 2015 show that retraction rates vary by country due to differences in research culture, regulations, and publication pressures [3]. Understanding these variations is vital to identifying systemic issues in research integrity.

We examined the countries with the highest numbers of retractions due to misconduct, analyzing both absolute counts and proportions relative to population size and publication output. Our goal is to show the geographical distribution of research misconduct and identify countries disproportionately represented in retraction statistics.

Methods

Data Source and Analysis

We used data from the SCImago Journal & Country Rank (SJR), based on Scopus data, to identify the top 100 countries by publication volume from 1996 to 2023—both overall and within the field of medicine. This approach was previously used to identify the most productive countries [5]. Retraction counts (1996-2023) were obtained from the Retraction Watch (RW) database, isolating retractions due to misconduct. Misconduct was defined according to criteria previously established [6] and detailed in Multimedia Appendix 1.

Both SJR and RW consider all authors listed on an article, regardless of position, when attributing publications and retractions, respectively, to a country. They use full counts, equally attributing publications and retractions to all listed countries. For each country, we analyzed absolute retraction counts and proportions weighted by population size (United Nations 2023 data) [7] and by publication output, calculated by dividing retraction count by total publication count. Data were collected independently by two researchers (PS and MS), with any discrepancies resolved through discussion.

Ethical Considerations

As this study did not involve the collection of personal health-related data, it did not require ethical review in accordance with Swiss legislation.

Results

Detailed results are available in Multimedia Appendix 1 and Figures 1-3. The US, China, the UK, Germany, and Japan are

XSL•FO
RenderX

the top 5 productive countries in terms of both overall publications and publications in medicine. Across the 100 countries analyzed, there were 37,858 retractions out of 79,645,579 publications overall (0.048%), and 10,890 retractions out of 23,175,369 publications in medicine (0.047%).

Figure 1. Number of retractions by country (1996-2023) based on overall publications and those in the field of medicine.

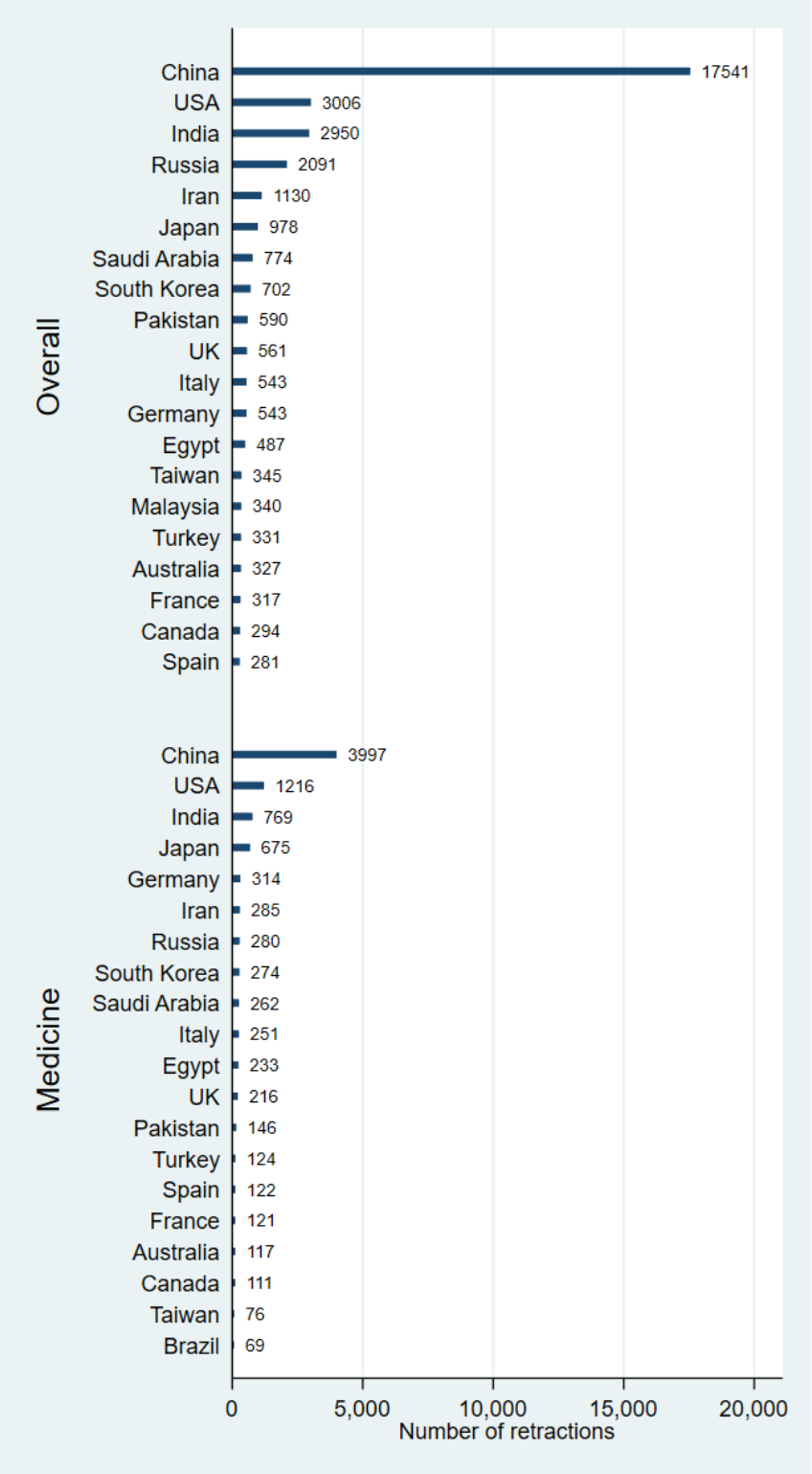

XSL•FO
RenderX

Figure 2. Number of retractions per 10,000 inhabitants by country (1996–2023), based on overall publications and those in the field of medicine.

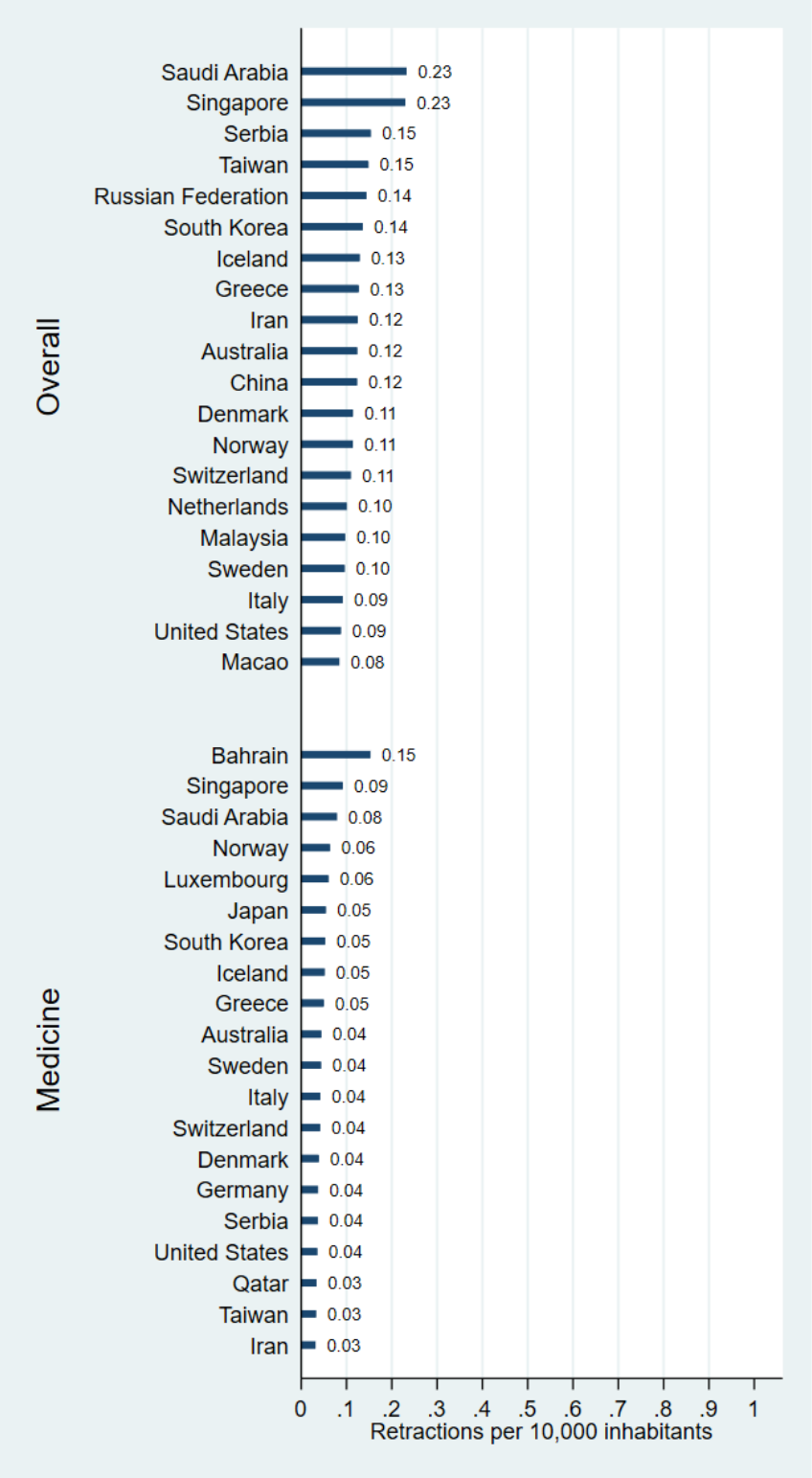

XSL•FO
RenderX

Figure 3. Proportion of retractions relative to publication output, by country (1996-2023), based on overall publications and those in the field of medicine.

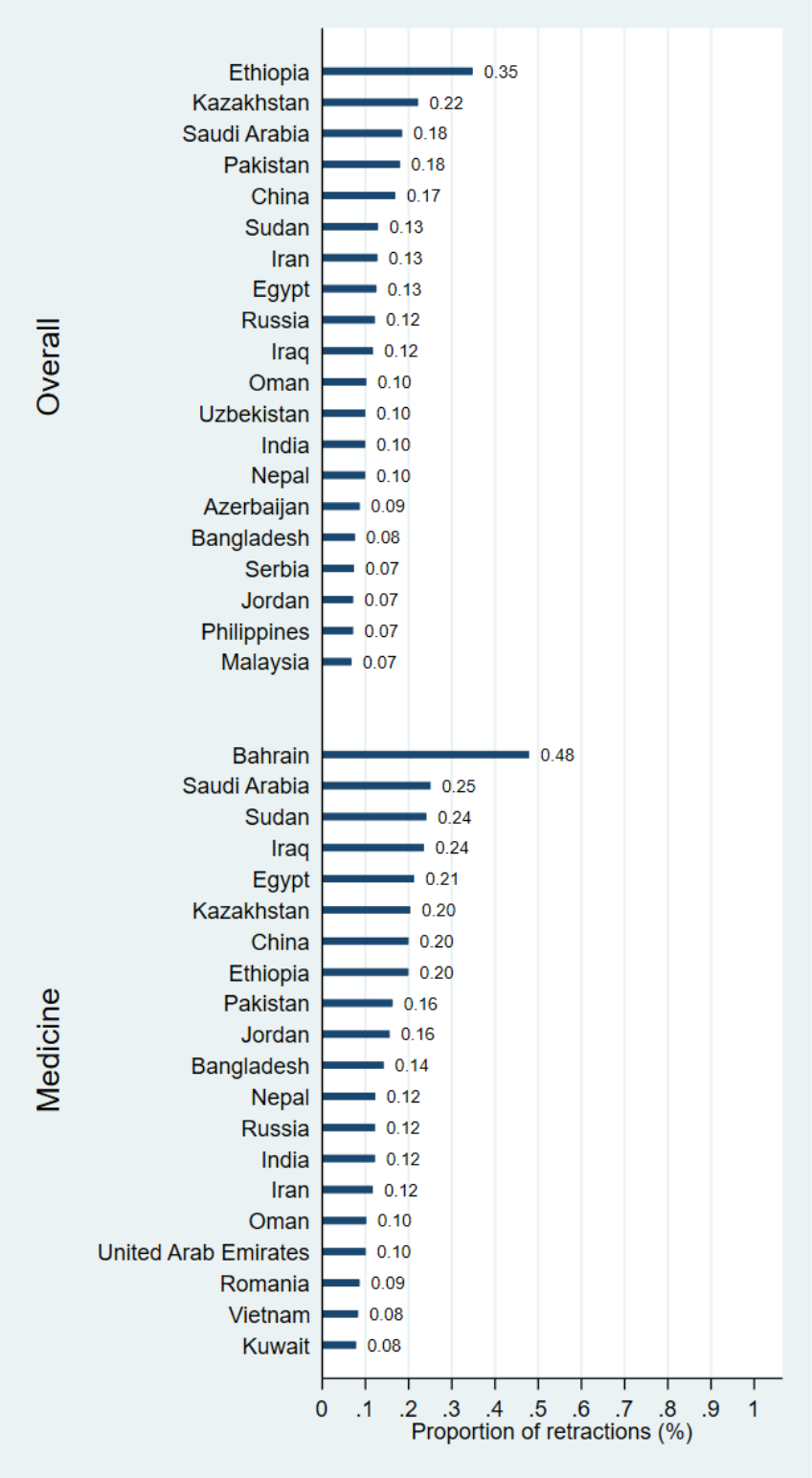

Our data suggests that China leads in retractions, far surpassing the US with more than five times as many retractions overall (17,541 vs 3006) and three times as many in medicine (3997 vs 1216). India ranks third with 2950 retractions overall and 769 in medicine. When adjusting for population size, Saudi Arabia, Singapore, Serbia, Taiwan, and Russia have the highest retraction rates overall, while Bahrain, Singapore, Saudi Arabia, Norway, and Luxembourg lead in medicine. Asian, Middle Eastern, and European countries are notably overrepresented in population-adjusted retractions. The highest retraction proportions relative to overall publication output were found in Ethiopia (0.35%), Kazakhstan (0.22%), Saudi Arabia (0.19%), Pakistan (0.18%), and China (0.17%), and in Bahrain (0.48%), Saudi Arabia (0.25%), Sudan (0.24%), Iraq (0.24%), and Egypt (0.21%) when evaluating the medicine subset, with

XSL•FO
RenderX

overrepresentation among Asian, Middle Eastern, and African countries.

Discussion

Our findings reveal that China, the US, and India have the highest numbers of retractions due to misconduct, with China particularly overrepresented. Retractions are disproportionately high in several Asian, Middle Eastern, and European countries when adjusted for population, as well as in several Asian, Middle Eastern, and African countries when adjusted for publication output. These results highlight regional disparities in research integrity.

These findings align with previous research identifying China, the US, and India as leaders in retraction numbers [2,3], with China's prominence especially noticeable in recent years [8-10]. The overrepresentation of retractions among countries with emerging research sectors, as observed in this study, reflects challenges such as weaker oversight and high publication pressure [1-3]. Measures like stricter peer review, automated plagiarism detection, open data sharing, and pre-registration of studies can help mitigate misconduct and improve oversight.

Limitations of this study include reliance on the RW database, which may miss some misconduct cases, and a primary focus on medicine, potentially overlooking trends in other disciplines. Nonetheless, RW remains a trusted source for aggregated retraction data [8], with findings consistent across both overall and medicine-specific data. Additionally, differences between SJR (Scopus-based) publication data and RW's broader retraction sources may slightly impact proportion calculations but are unlikely to affect key findings

In conclusion, this study underscores substantial geographical disparities in research misconduct, emphasizing the need for improved oversight and ethical standards, especially in regions with growing research sectors. Our findings contribute to ongoing discussions on the reliability of scientific research and the importance of global efforts to address misconduct.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Numbers and proportions of retractions by country; retractions per 10,000 inhabitants by country (1996-2023); and criteria for retraction used in the Retraction Watch database.
[DOCX File , 33 KB - jmir_v27i1e65775_app1.docx]

References

1. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. Proc Natl Acad Sci U S A 2012 Oct 16;109(42):17028-17033 [FREE Full text] [doi: 10.1073/pnas.1212247109] [Medline: 23027971]
2. Grieneisen ML, Zhang M. A comprehensive survey of retracted articles from the scholarly literature. PLoS One 2012;7(10):e44118 [FREE Full text] [doi: 10.1371/journal.pone.0044118] [Medline: 23115617]
3. Ribeiro MD, Vasconcelos SMR. Correction to: retractions covered by Retraction Watch in the 2013–2015 period: prevalence for the most productive countries. Scientometrics 2018 Jan 31;114(2):735-735. [doi: 10.1007/s11192-018-2653-6]
4. Lievore C, Rubbo P, Dos Santos CB, Picinin CT, Pilatti LA. Research ethics: a profile of retractions from world class universities. Scientometrics 2021 May 23;126(8):6871-6889 [FREE Full text] [doi: 10.1007/s11192-021-03987-y] [Medline: 34054160]
5. Sebo P. How well does NamSor perform in predicting the country of origin and ethnicity of individuals based on their first and last names? PLoS One 2023;18(11):e0294562 [FREE Full text] [doi: 10.1371/journal.pone.0294562] [Medline: 37972002]
6. Sebo P, Schwarz J, Achtari M, Clair C. Women are underrepresented among authors of retracted publications: retrospective study of 134 medical journals. J Med Internet Res 2023 Oct 06;25:e48529 [FREE Full text] [doi: 10.2196/48529] [Medline: 37801343]
7. World Population Prospects 2024. United Nations – Department of Economic and Social Affairs Population Division. URL: https://population.un.org/wpp/ [accessed 2024-11-08]
8. Candal-Pedreira C, Ross JS, Ruano-Ravina A, Egilman DS, Fernández E, Pérez-Ríos M. Retracted papers originating from paper mills: cross sectional study. BMJ 2022 Dec 28;379:e071517 [FREE Full text] [doi: 10.1136/bmj-2022-071517] [Medline: 36442874]
9. Rivera H, Teixeira da Silva JA. Retractions, fake peer reviews, and paper mills. J Korean Med Sci 2021 Jul 21;36(24):e165 [FREE Full text] [doi: 10.3346/jkms.2021.36.e165] [Medline: 34155837]
10. Sebo P. Chinese authors are overrepresented in medical articles retracted for fake peer review or paper mill. Intern Emerg Med 2024 Dec;19(8):2369-2371. [doi: 10.1007/s11739-024-03616-5] [Medline: 38684644]

Abbreviations

RW: Retraction Watch

XSL•FO
RenderX

SJR: SCImago Journal & Country Rank

Edited by A Mavragani; submitted 25.08.24; peer-reviewed by G Goldman, Y Jiang; comments to author 10.10.24; revised version received 13.10.24; accepted 12.11.24; published 14.01.25.

Please cite as:
Sebo P, Sebo M
Geographical Disparities in Research Misconduct: Analyzing Retraction Patterns by Country
J Med Internet Res 2025;27:e65775
URL: https://www.jmir.org/2025/1/e65775
doi:10.2196/65775
PMID:39808480

©Paul Sebo, Melissa Sebo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study

Joshua Biro[1], PhD; Jessica L Handley[1], MA; Nathan K Cobb[2,3], MD; Varsha Kottamasu[2], MHA; Jeffrey Collins[2], MHS; Seth Krevat[1,3], MD; Raj M Ratwani[1,3], PhD

[1]National Center for Human Factors in Healthcare, MedStar Health Research Institute, Washington, DC, United States

[2]MedStar Health Institute for Innovation, Washington, DC, United States

[3]Georgetown University School of Medicine, Washington, DC, United States

Corresponding Author:
Joshua Biro, PhD
National Center for Human Factors in Healthcare
MedStar Health Research Institute
3007 Tilden St NW
Washington, DC, 20008
United States
Phone: 1 3015423073
Email: joshua.m.biro@medstar.net

Abstract

Artificial intelligence–enabled ambient digital scribes may have many potential benefits, yet results from our study indicate that there are errors that must be evaluated to mitigate safety risks.

(J Med Internet Res 2025;27:e64993) doi:10.2196/64993

KEYWORDS

artificial intelligence; AI; patient safety; ambient digital scribe; AI-enabled scribe technology; AI scribe technology; scribe technology; accuracy; safety; ambient scribe; digital scribe; patient-clinician; patient-clinician communication; doctor-patient relationship; doctor-patient communication; patient engagement; patient safety; dialogue script; scribe

Introduction

Generative artificial intelligence (AI)–enabled ambient digital scribe (ADS) technology uses the patient-clinician conversation to generate clinical documentation; it has the potential to improve patient engagement and reduce clinician burden [1,2]. These technologies are becoming more prevalent, especially in ambulatory care settings, yet there is little known about documentation accuracy and the types of errors that may stem from ADS use [3]. Error-prone ADS technology may have serious patient safety consequences [4]. We evaluated 2 popular commercially available ADS products in a simulated setting to systematically identify the frequency and pattern of documentation errors.

Methods

Ethical Considerations

This study was approved by the MedStar Health Institutional Review Board (00007789) to cover secondary analysis of existing patient data without additional consent. All data were deidentified. Participants did not receive any form of compensation.

Recording and Simulation

Recordings of 11 real outpatient encounters from a range of service lines (otolaryngology, cardiology, rheumatology, family medicine, pediatrics, endocrinology, internal medicine, gastroenterology, oncology, and urgent care) were transcribed by automated software and then deidentified and edited by a senior physician (NKC) for clarity to create 11 unique dialogue scripts. The dialogue scripts were used to evaluate 2 commercial ADS products. For each script, a researcher (JB or VK) simulating the patient and a medical resident simulating the physician read from the script while the ADS products were in use. Each script was read by 2 different residents per ADS product, yielding 22 draft notes per product and 44 draft notes in total across products. The residents reviewed the draft notes to identify errors. Each error was independently categorized by 2 reviewers (JB or JLH) as either an omission, addition, wrong output, or irrelevant or misplaced text, as defined in Table 1. Disagreements were discussed to reach consensus.

XSL•FO
RenderX

Results

There were 127 errors (mean 2.9, SD 2.7 errors per draft note) in 31 of 44 (70%) draft notes. ADS product A resulted in 66 errors (mean 3, SD 2.7 per draft note) and product B resulted in 61 errors (mean 2.8, SD 2.7 per draft note). Error frequency by error type and product is detailed in Table 1, with omission errors being the most frequent across products. Error types significantly differed between the 2 ADS products (Fisher exact test: P=.002).

Table 1. Frequency counts, percentages, definitions, and examples of ambient digital scribe (ADS) error types.

Error type	Errors by ADS product, n (%)		Definition	Example
	Product A (n=66)	Product B (n=61)		
Omission	55 (83)	33 (54)	Model leaves out key information from its response	"[N]o laterality mentioned in ears in physical exam section"
Addition	3 (4)	7 (11)	Model adds inappropriate or irrelevant information	"Patient doesn't refer to any flare ups in awhile, but the note shared that patient was using X medication to help with flare ups in HPI"
Wrong output	4 (6)	6 (10)	Model provides an incorrect response	"[A]ssociated the wrong test with the contrast"
Irrelevant or misplaced text	4 (6)	15 (25)	Model output is technically correct but not appropriate in clinical context	"[C]aptured all the supplemental information (asthma, mammogram, etc.) and harped on the steroid injections which doesn't matter"

Discussion

While ADS technologies may have potential benefits, there are frequent errors in the generated note. Across both products, errors of omission were the most common; this error type may be the most difficult for clinicians to identify since the identification process requires memory recall of details from the patient encounter. If clinicians review their documentation after several patient encounters, recalling omitted details may be challenging. It may be easier to identify errors such as additions and wrong outputs since this relies on recognition of an issue in the text being presented to the clinician. Notably, there was a different pattern of errors between the two products.

There are limitations to this study. The ADS technologies were evaluated against a limited number of patient cases in a controlled environment that did not fully represent clinical workflows. In addition, the cases were read by 2 researchers acting as patients, and it is likely that both clinicians and patients would have more variability in language, tone, volume, and many other characteristics that could impact ADS accuracy.

It is imperative that ADS technologies be evaluated in a realistic clinical setting (either in situ or in a representative simulation) to determine the frequency and types of errors so that appropriate risk mitigation and safety plans can be developed. Developing methods to capture AI-related safety issues was a component of President Joe Biden's "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" [5], and robust processes for AI safety are needed. In the absence of a standardized evaluation framework, health care facilities currently bear the burden of testing and reporting these results in the United States. It is to be noted that, effective August 2024, the European Union Artificial Intelligence Act legally requires developers of AI-based systems to evaluate the safety of their products [6]. While a step in the right direction, the underlying vendor algorithms are often proprietary, opaque, and the subject of continuous innovation; thus, there is still a need for independent ongoing testing to confirm vendor claims of safety. Future work should develop a robust, standardized, and repeatable ADS evaluation framework to facilitate efficient knowledge sharing in this fast-paced, decentralized system.

Acknowledgments

We would like to acknowledge Dr Sahithi Reddy and Dr James Mickler for their assistance in executing this work.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JB contributed to study design, data acquisition, data analysis, interpretation, drafting, and reviewing the manuscript. JLH contributed to study design, interpretation, drafting, and critically reviewing the manuscript. NKC contributed to study conception, design, data acquisition, interpretation, drafting, and critically reviewing the manuscript. JC contributed to study conception, data acquisition, and critically reviewing the manuscript. SK contributed to study design, data analysis, interpretation, and critically

XSL•FO
RenderX

reviewing the manuscript. VK contributed to data acquisition, interpretation, and critically reviewing the manuscript. RMR contributed to study conception, study design, interpretation, and critically reviewing the manuscript.

Conflicts of Interest

None declared.

References

1. Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catalyst 2024 Feb 21;5(3):692-694. [doi: 10.1056/CAT.23.0404]
2. van Buchem MM, Kant IMJ, King L, Kazmaier J, Steyerberg EW, Bauer MP. Impact of a digital scribe system on clinical documentation time and quality: usability study. JMIR AI 2024 Sep 23;3:e60020 [FREE Full text] [doi: 10.2196/60020] [Medline: 39312397]
3. Seth P, Carretas R, Rudzicz F. The utility and implications of ambient scribes in primary care. JMIR AI 2024 Oct 04;3:e57673 [FREE Full text] [doi: 10.2196/57673] [Medline: 39365655]
4. van Buchem MM, Boosman H, Bauer MP, Kant IMJ, Cammel SA, Steyerberg EW. The digital scribe in clinical practice: a scoping review and research agenda. NPJ Digit Med 2021 Mar 26;4(1):57 [FREE Full text] [doi: 10.1038/s41746-021-00432-5] [Medline: 33772070]
5. Biden JR. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. The White House. 2023. URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ [accessed 2025-01-07]
6. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. 2024/1689, 32024R1689. European Union. 2024 Jun 13. URL: https://artificialintelligenceact.eu/ai-act-explorer/ [accessed 2024-09-27]

Abbreviations

ADS: ambient digital scribe
AI: artificial intelligence

Edited by A Coristine; submitted 01.08.24; peer-reviewed by D Sharma, F Warg; comments to author 24.09.24; revised version received 03.10.24; accepted 09.12.24; published 27.01.25.

Please cite as:
Biro J, Handley JL, Cobb NK, Kottamasu V, Collins J, Krevat S, Ratwani RM
Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study
J Med Internet Res 2025;27:e64993
URL: https://www.jmir.org/2025/1/e64993
doi:10.2196/64993
PMID:

©Joshua Biro, Jessica L Handley, Nathan K Cobb, Varsha Kottamasu, Jeffrey Collins, Seth Krevat, Raj M Ratwani. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Real-Time fMRI Neurofeedback Modulation of Dopaminergic Midbrain Activity in Young Adults With Elevated Internet Gaming Disorder Risk: Randomized Controlled Trial

Anqi Gu[1], MSc; Cheng Lam Chan[1], BS; Xiaolei Xu[2], PhD; Joseph P Dexter[3,4], PhD; Benjamin Becker[5,6], PhD; Zhiying Zhao[1], PhD

[1]Centre for Cognitive and Brain Sciences, Institute of Collaborative Innovation, University of Macau, Macau, China

[2]School of Psychology, Shandong Normal University, Jinan, China

[3]Centre for Data Science, Institute of Collaborative Innovation, University of Macau, Macau, China

[4]Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China

[5]The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China

[6]Department of Psychology, The University of Hong Kong, Hong Kong, China

Corresponding Author:
Zhiying Zhao, PhD
Centre for Cognitive and Brain Sciences
Institute of Collaborative Innovation
University of Macau
N21 Research building
Macau, 99078
China
Phone: 853 8822 9182
Email: zhiyingzhao@um.edu.mo

Abstract

This study provides preliminary evidence for real-time functional magnetic resonance imaging neurofeedback (rt-fMRI NF) as a potential intervention approach for internet gaming disorder (IGD). In a preregistered, randomized, single-blind trial, young individuals with elevated IGD risk were trained to downregulate gaming addiction–related brain activity. We show that, after 2 sessions of neurofeedback training, participants successfully downregulated their brain responses to gaming cues, suggesting the therapeutic potential of rt-fMRI NF for IGD (Trial Registration: ClinicalTrials.gov NCT06063642; https://clinicaltrials.gov/study/NCT06063642).

(J Med Internet Res 2025;27:e64687) doi:10.2196/64687

KEYWORDS

real-time functional magnetic resonance imaging neurofeedback; internet gaming disorder; craving; reward processing; ventral tegmental area

Introduction

Internet gaming disorder (IGD) had a global prevalence of 10.4% by 2022; it disproportionately affects younger individuals [1]. The effectiveness of traditional psychosocial and pharmacological treatments, however, remains inconclusive [2]. Real-time functional magnetic resonance imaging neurofeedback (rt-fMRI NF) training allows for self-regulation of brain activity patterns and shows promise for several psychiatric conditions [3]. This preregistered, randomized, single-blind study examined rt-fMRI NF as a potential intervention for IGD. Given the key role and potential modifiability of the ventral tegmental area (VTA), a dopaminergic midbrain region involved in reward processing functions [4,5], we validated the association between VTA response and gaming addiction severity in a small sample (N=9). In a larger study (N=20), we evaluated whether 2 neurofeedback training sessions for VTA downregulation reduced craving for internet gaming.

XSL•FO
RenderX

Methods

Overview

We recruited young participants with IGD risk through social media–based advertisements. In study 1, 9 participants performed a gaming video–based cue-reactivity task (Multimedia Appendix 1, Figure S1) to establish the association between VTA cue-reactivity and IGD symptom level [5]. In study 2, 20 different participants were screened and randomly assigned to 1 of 2 groups. The inclusion criteria and screening instruments are described in Multimedia Appendix 1, Figures S2 and S3. The experimental group received feedback from the VTA (Montreal Neurological Imaging [MNI] coordinates [1, –17, –13]; 246 voxels; Figure 1B), while the control group received sham feedback from the right middle temporal gyrus, specifically an 8-mm–radius sphere centered at MNI coordinates [53, 6, –18]. This region was chosen because it is unrelated to reward processing and is not functionally connected with the VTA. Multimedia Appendix 1 describes the training details. Briefly, during neurofeedback training, blood-oxygen-level–dependent signals in the feedback regions were presented to participants as line graphs in real time (Figure 1A). Both groups were instructed to downregulate the signals using mental strategies that are helpful against craving; they reported their effort and perceived success after each run (Multimedia Appendix 1, Table S1).

Participants completed a 1-item visual analog scale (VAS) for current craving levels at baseline, after the second imaging session, and a month after the last scan; VAS scores were the primary outcome measure. The neural-level VTA cue-reactivity and inhibitory control performance in an affective go/no-go task [6] (Multimedia Appendix 1, Figure S4) measured at baseline and after the second training session were the secondary outcome measures. Analysis of the outcome measures used JASP (version 0.18.3; JASP Team) and the *nparLD* toolbox [7].

Figure 1. (A) The structure of the neurofeedback setting. Blood-oxygen-level–dependent activities of the trained regions for both groups are visualized as line graphs. Participants were not informed about the feedback region or its function. The blue arrow and white cross denote downregulation and rest, respectively. (B) The location of the ventral tegmental area (VTA) region of interest in the Montreal Neurological Imaging brain image is marked in red. (C) VTA cue-reactivity change (post – pre) in the cue-reactivity task based on a gaming – neutral comparison (*P*=.06).

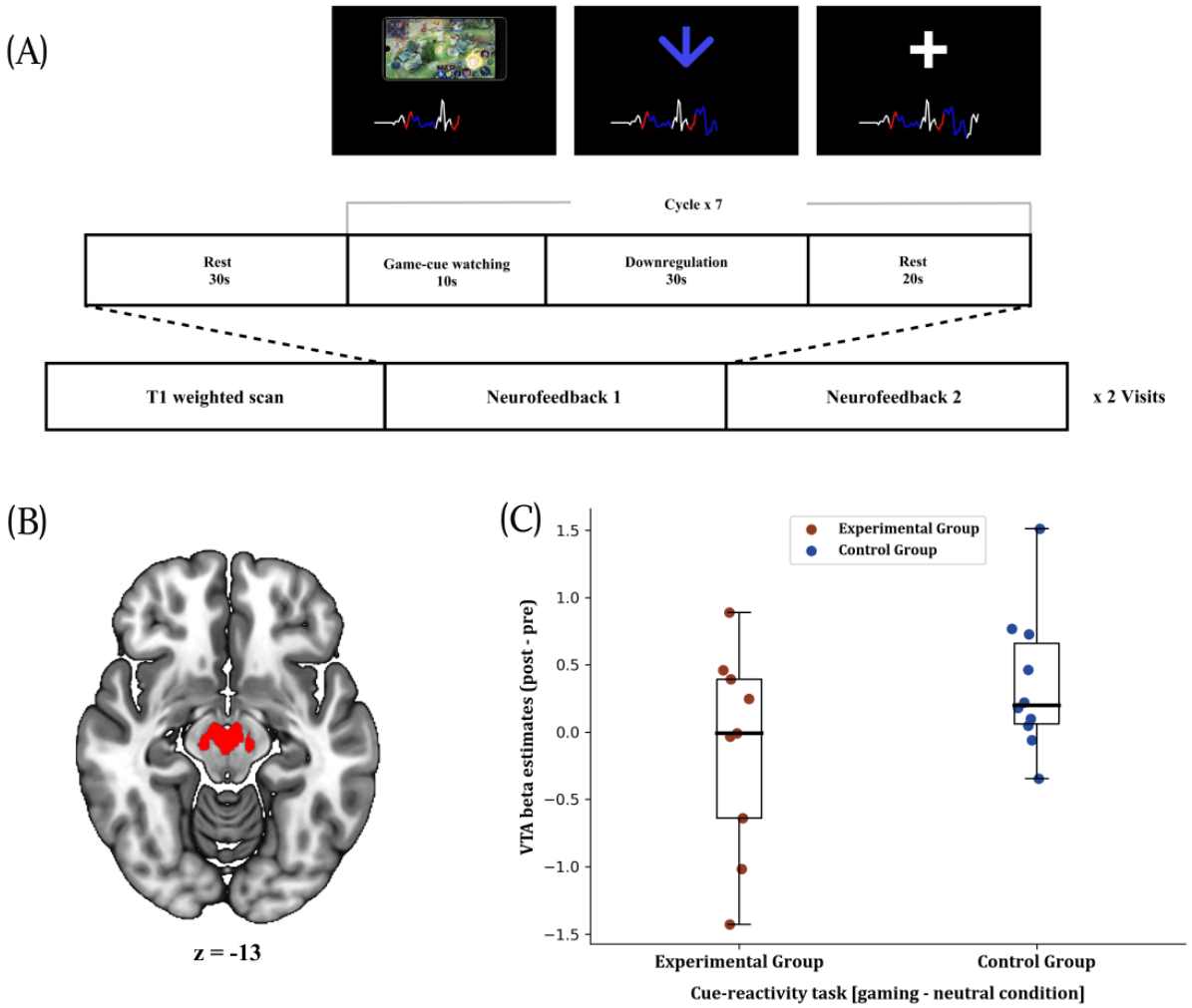

XSL•FO
RenderX

Ethical Considerations

Both studies were approved by the Panel on Research Ethics of the University of Macau (BSERE23-APP004-ICI-M1). The neurofeedback study was preregistered at ClinicalTrials.gov (NCT06063642). Participants provided written informed consent prior to participation and were compensated with MOP 200 (US $24.96) supermarket vouchers. All data were deidentified.

Results

In study 1, the cue-reactivity task successfully activated the VTA (P=.004 after correction for false-discovery rate at a whole-brain level; Multimedia Appendix 1, Figure S5A). Moreover, the VTA response was significantly correlated with addiction severity (Pearson r=.733; P=.02; Figure S5B).

Group comparison results for study 2 are shown in Table 1, and the baseline characteristics of the neurofeedback completers can be found in Multimedia Appendix 1, Table S2. Both groups showed significantly decreased subjective craving levels (P=.002; Multimedia Appendix 1, Figure S6) with no significant group differences (P=.10). Over the course of the neurofeedback sessions, the experimental group downregulated their VTA activity to a greater extent than the control group (P=.02). In contrast to sham feedback, VTA neurofeedback led to more pronounced decreases in VTA cue-reactivity (Figure 1C; P=.06), a trend that might suggest successful transfer of learned VTA regulation. No significant interaction or group effects were found for inhibitory control (accuracy and reaction time in the go/no-go task: P>.12 for all groups).

XSL•FO
RenderX

Table 1. Overall outcomes for study 2.

Outcome measures, variables, and effects	Test statistics	*P* values
Primary (Multimedia Appendix 1, Figure S6)		
Change in visual analog scale score[a]		
Main effect of time	$F_{1.75,\infty}$=10.54	*P*<.001
Post hoc time effect	Pre vs post: *z*=2.41; Pre vs follow-up: *z*=3.24	Pre vs post: 2-tailed P_{holm}=.01; Pre vs follow-up: 2-tailed P_{holm}<.001
Secondary		
Ventral tegmental area cue-reactivity change[b] **(Figure 1C)**		
Group effect	Cohen *d*=–.85	1-tailed *P*=.06
Go/no-go performance[c]		
Main effect of time (accuracy)	$F_{1,15}$=8.90	*P*=.01
No interaction effect or group difference (reaction time and whole-brain functional magnetic resonance imaging data)	Reaction time: $F_{1,15}$≤2.79; whole brain: not applicable	*P*>.12 (all groups)
Exploratory (Multimedia Appendix 1, Figure S7)		
Ventral tegmental area control during neurofeedback[d]		
Time × group interaction	$F_{1,16}$=6.98	*P*=.02
Post hoc main time effect	Experimental group: t_8=–2.17; control group: t_8=–1.63	Experimental group: 1-tailed *P*=.03; control group: 1-tailed *P*=.93
Ventral tegmental area cue-reactivity change and neurofeedback control[e]		
Pearson correlation	*r*=.64	*P*=.004

[a]Nonparametric time (pre vs post vs follow-up) × group (experimental vs control) ANOVA. One participant was lost to follow-up, leaving a total of 18 participants in this analysis. The changes in visual analog scale score showed a significant time effect (*P*<.001) with no group or time × group effects (*P* >.10). The post hoc paired-sample Wilcoxon signed-rank test for the time effect showed a significant pre vs post difference (pre-neurofeedback visual analog scale median: 61; post-neurofeedback visual analog scale median: 40; *z*=2.41; 2-tailed P_{holm}=.01) and pre vs follow-up (follow-up visual analog scale median: 43; *z*=3.24; 2-tailed P_{holm}<.001). P_{holm} indicates that the Bonferroni-Holm method was used to control for false positives.

[b]Time (pre vs post) × group (experimental vs control) analysis of covariance controlling for age and sex. The experimental group showed greater reduction in cue-reactivity compared to the control group (experimental group median change: –0.37; control group median change: –0.15; 1-tailed *P*=.06).

[c]Time (pre vs post) × group (experimental vs control) ANOVA (accuracy, reaction time and whole-brain functional magnetic resonance imaging data). Go/no-go task data were lost for 2 participants due to technical issues, leaving a total of 17 participants. Accuracy on the go/no-go task improved for both groups after intervention (pre-neurofeedback median: 0.67; post-neurofeedback median: 0.83; *P*=.01); no other main or interaction effects were found for behavioral performance (*P*>.12 for all groups). No brain activation results remained significant after correction for false-discovery rate at a whole-brain level.

[d]Time (day 1 vs day 2) × group (experimental vs control) ANOVA. Data were lost for 1 participant from the third run of training, resulting in 18 participants being included in this analysis. During neurofeedback training, there was a significant interaction effect of time (day 1 vs day 2) and group (real vs sham feedback) for downregulated activation (*P*=.02). A post hoc *t* test indicated that ventral tegmental area activity decreased significantly for the experimental group (median change: –0.07; *P*=.03) but not for the control group (median change: 0.11; *P*=.93).

[e]Ventral tegmental area activity change in the cue-reactivity task was positively correlated with neurofeedback training (day 2 – day 1; *r*=.64; *P*=.004).

Discussion

We previously proposed rt-fMRI NF as a promising therapeutic strategy for individuals with internet addiction [3]. The efficacy of rt-fMRI NF, however, has not been empirically examined for any behavioral addictions. In this trial, we found that individuals with problematic internet gaming behaviors could learn to downregulate their VTA activity through genuine feedback, and that this effect translated into changes in VTA reactivity to gaming stimuli. Although we found no improvements at the behavioral level after NF, the neural activity findings underscore the potential of VTA NF training in restoring reward processing functions in individuals with IGD risk.

The study has several limitations. First, our participants were not diagnosed with IGD, meaning that their symptoms might not have been severe enough to benefit from the intervention. Furthermore, due to the small sample size, this study was potentially underpowered to detect group differences in behavioral changes. Larger-scale trials are needed to further examine the clinical efficacy of this intervention protocol.

XSL•FO
RenderX

Acknowledgments

This study was supported by the Macao Science and Technology Development Fund (FDCT 0044/2023/RIA1 and FDCT 0015/2023/ITP1 to ZZ), the National Natural Science Foundation of China (NSFC32100887 to XX), and a Start-Up Research Grant from the University of Macau (SRG2024-00019-ICI to JPD). Any opinions, findings, conclusions, or recommendations expressed in this publication do not reflect the views of the government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission. We thank Dr Lucas Peek for his technical support on the neurofeedback training platform.

Data Availability

Data will be made publicly available at the time of publication at OSF Registries [8].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full description of study design and methods.
[DOCX File , 3225 KB - jmir_v27i1e64687_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2101 KB - jmir_v27i1e64687_app2.pdf]

References

1. Kim HS, Son G, Roh E, Ahn W, Kim J, Shin S, et al. Prevalence of gaming disorder: A meta-analysis. Addict Behav 2022 Mar;126:107183. [doi: 10.1016/j.addbeh.2021.107183] [Medline: 34864436]
2. Zajac K, Ginley MK, Chang R. Treatments of internet gaming disorder: a systematic review of the evidence. Expert Rev Neurother 2020 Jan;20(1):85-93 [FREE Full text] [doi: 10.1080/14737175.2020.1671824] [Medline: 31544539]
3. Becker B, Montag C. Opinion: real-time fMRI neurofeedback and the application of the neuropeptide oxytocin as promising new treatment approaches in internet addiction? In: Montag C, Reuter M, editors. Internet Addiction. Studies in Neuroscience, Psychology and Behavioral Economics. Cham, Switzerland: Springer; 2017:311-321.
4. Haber S. Anatomy and connectivity of the reward circuit. In: Handbook of Reward and Decision Making. Amsterdam, Netherlands: Elsevier; 2009:1-27.
5. Young K. Caught in the Net: How to Recognize the Signs of Internet Addiction--And a Winning Strategy For Recovery. New York, NY: JWiley; 1998.
6. Liu C, Dai J, Chen Y, Qi Z, Xin F, Zhuang Q, et al. Disorder- and emotional context-specific neurofunctional alterations during inhibitory control in generalized anxiety and major depressive disorder. Neuroimage Clin 2021;30:102661 [FREE Full text] [doi: 10.1016/j.nicl.2021.102661] [Medline: 33866301]
7. Noguchi K, Gel Y, Brunner E, Konietschke F. nparLD: an R software package for the nonparametric analysis of longitudinal data in factorial experiments. J Stat Softw 2012;50(12):1-23 [FREE Full text] [doi: 10.18637/jss.v050.i12]
8. OSF Registries. URL: https://osf.io/zar94 [accessed 2025-01-16]

Abbreviations

IGD: internet gaming disorder
MNI: Montreal Neurological Imaging
rt-fMRI NF: real-time functional magnetic resonance imaging neurofeedback
VAS: visual analog scale
VTA: ventral tegmental area

XSL•FO
RenderX

Edited by A Schwartz; submitted 26.07.24; peer-reviewed by L Beynel, A Rizvi, K Song; comments to author 22.10.24; revised version received 28.11.24; accepted 10.01.25; published 29.01.25.

Please cite as:
Gu A, Chan CL, Xu X, Dexter JP, Becker B, Zhao Z
Real-Time fMRI Neurofeedback Modulation of Dopaminergic Midbrain Activity in Young Adults With Elevated Internet Gaming Disorder Risk: Randomized Controlled Trial
J Med Internet Res 2025;27:e64687
URL: https://www.jmir.org/2025/1/e64687
doi:10.2196/64687
PMID:

©Anqi Gu, Cheng Lam Chan, Xiaolei Xu, Joseph P Dexter, Benjamin Becker, Zhiying Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Self-Reported Side Effects Associated With Selective Androgen Receptor Modulators: Social Media Data Analysis

Aditya Joshi[1], MS; Diego Federico Kaune[2], BA; Phillip Leff[1], DO; Emily Fraser[1], BA; Sarah Lee[1], BS; Morgan Harrison[1], BS; Moustafa Hazin[1], DO

[1]Department of Internal Medicine, Phoenix Health Sciences Campus, Creighton University School of Medicine, Phoenix, AZ, United States
[2]Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Aditya Joshi, MS
Department of Internal Medicine
Phoenix Health Sciences Campus
Creighton University School of Medicine
3100 N Central Ave
Phoenix, AZ, 85012
United States
Phone: 1 (602) 812 4312
Email: aditya.joshi.research@gmail.com

Abstract

This study focuses on the prevalence and dangers of selective androgen receptor modulator use through a social media (Reddit) analysis. Physicians should be aware that patients are increasingly seeking medical advice through social media, and adverse events are likely underreported in the literature.

(J Med Internet Res 2025;27:e65031) doi:10.2196/65031

KEYWORDS

selective androgen receptor modulator; SARM; liver toxicity; social media; data analysis; anabolic; muscle; bone; toxicities; self-report; side effect; retrospective analysis; public post; Reddit; androgen receptor ligands; drug

Introduction

Selective androgen receptor modulators (SARMs) are synthetic androgen receptor ligands designed to have anabolic effects in muscle and bone while minimizing unwanted androgenic effects in areas like the prostate and cardiovascular system [1]. This tissue selectivity makes SARMs an attractive option for treating conditions such as osteoporosis and muscle wasting while avoiding the systemic side effects associated with traditional androgen therapies. However, no SARM has received US Food and Drug Administration approval to date. Studies show SARMs are associated with acute hepatotoxicity [2], and recent case reports indicate long-term testosterone suppression along with nephrotoxicity and cardiotoxicity [3,4]. Abuse of SARMs has become increasingly prevalent due to targeting of both athletes and youth through social media platforms. While reports of SARM abuse and toxicity have been rising, no study has examined adverse events (AEs) and laboratory data using self-reported social media posts.

Methods

Overview

This study involved a retrospective analysis of public posts from March 2015 to November 2023 on Reddit. Data collection used a Python-based script to extract and filter posts from SARM-related subreddits. Posts were identified through a keyword-matching process, using “liver,” “AST,” “ALT,” “lab-work,” “bloodwork,” and “symptoms.” Initial keywords were selected and refined to capture posts reporting laboratory results and hepatic outcomes associated with SARM use. Posts were manually reviewed and systematically excluded if unrelated to SARM use. Posts were categorized based on users’ self-reported stages of SARM use: “precycle,” “midcycle,” and “postcycle” (Figure 1). Demographic data, specific SARMs taken, mentions of concurrent medications, self-reported symptoms, and self-reported laboratory values were extracted. Means of self-reported laboratory values were calculated and compared across stages of SARM use. Female posters were excluded from the calculations of luteinizing hormone,

XSL•FO
RenderX

follicle-stimulating hormone, total testosterone, free testosterone, estrogen, and sex hormone–binding globulin.

Figure 1. Flowchart of data extraction methodology. Data extraction of posts from selective androgen receptor modulator (SARM)–related subreddits yielded 7786 initial posts. Manual data cleaning removed 2537 duplicates and 1372 uncategorizable posts. The remaining 3877 posts were classified into pre-SARM use (n=1694), mid-SARM use (n=1412), and post-SARM use (n=771). Clinical laboratory values were voluntarily reported in 438 pre-use, 240 mid-use, and 343 post-use posts, enabling analysis of physiological changes associated with recreational SARM use. This retrospective analysis captured real-world SARM use patterns and their associated clinical outcomes in a young male population.

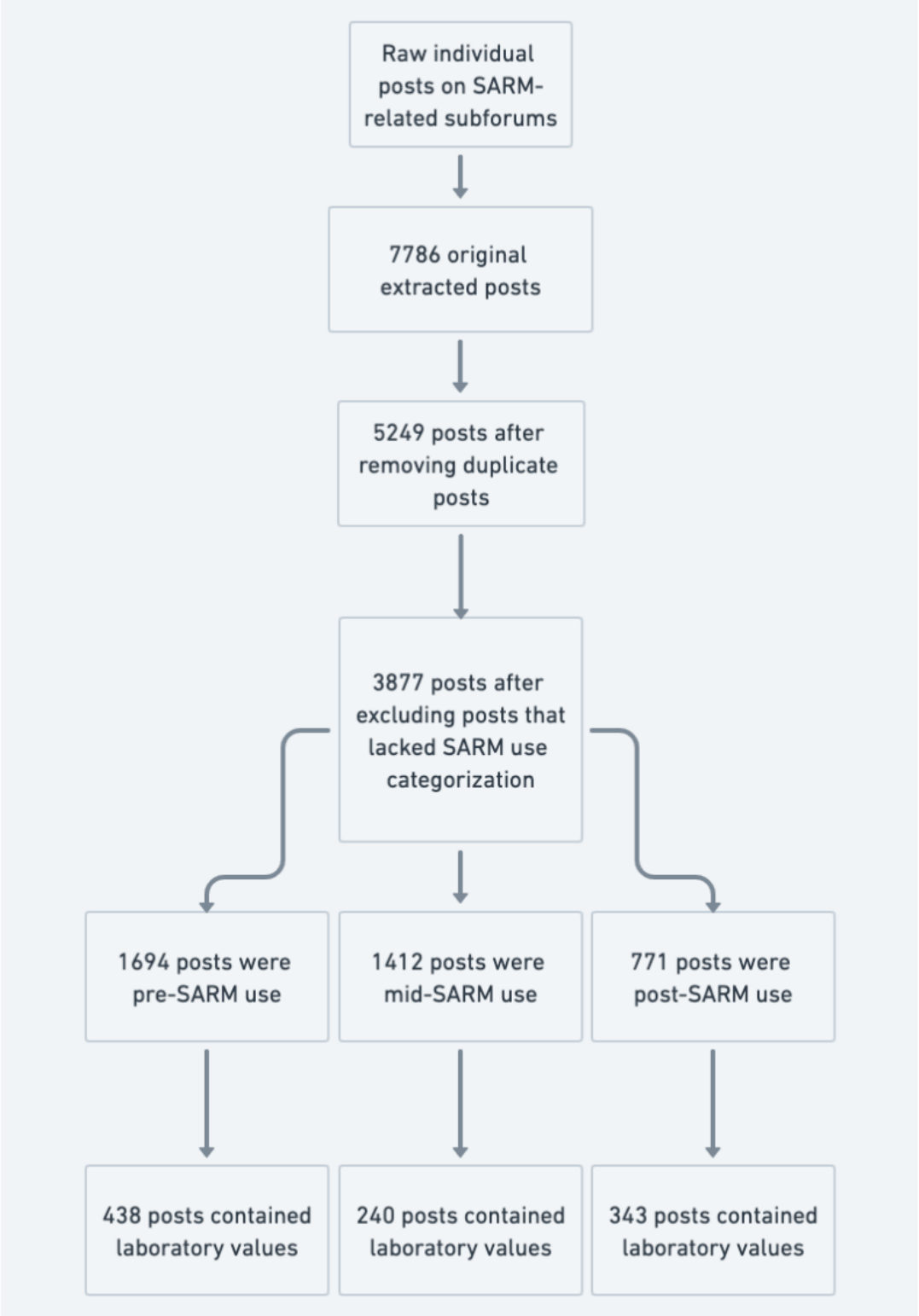

Ethical Considerations

To protect user privacy, identifying information was anonymized during data extraction. Data were aggregated to prevent linkage to individual users. No identifiable information was retained or reported. This study was exempt from institutional review board approval and strictly adhered to ethical guidelines for digital data use.

XSL•FO
RenderX

Results

Initially, 5249 unique posts were extracted, of which 1372 were excluded due to irrelevance or the inability to categorize the stage of SARM use. Of the remaining 3877 posts, 1694 were pre-SARM use, 1412 were mid-SARM use, and 771 were post-SARM use. RAD140 was the most mentioned SARM, with 1389 posts. The mean reported age was 27 (range 13-63, SD 7.9) years. In 418 posts reporting gender, 20 (4.8%) posters identified as female. Significant changes in aspartate aminotransferase, alanine aminotransferase, creatinine kinase, high-density lipoprotein, low-density lipoprotein, total testosterone, and sex hormone–binding globulin were observed before, during, and after SARM use (Table 1 and Multimedia Appendix 1). Of 2183 SARM users, 382 (17.5%) reported using tamoxifen or enclomiphene, and 170 (7.8%) reported using N-acetylcysteine or other hepatoprotective supplements (milk thistle or tauroursodeoxycholic acid) during or after SARM use.

Table 1. A comparison of the mean self-reported clinical markers before, during, and after selective androgen receptor modulator (SARM) use. This retrospective analysis examined changes in liver function, kidney function, hormones, and lipid profiles in recreational SARM users who voluntarily reported laboratory results. Of note, the study population consisted primarily of young men. Statistical significance was assessed using paired t tests comparing baseline values to both during-use and post-use periods. $P<.05$ indicates statistical significance. The large SDs in many measurements indicate high variability in individual physiological responses. Normal values refer specifically to men. Women were removed from the analysis of luteinizing hormone, follicle-stimulating hormone, total testosterone, free testosterone, estrogen, and sex hormone–binding globulin.

Clinical markers	Normal values	Before SARM use		During SARM use		*P* value	After SARM use		*P* value
		Posts, n	Self-reported values, mean (SD)	Posts, n	Self-reported values, mean (SD)		Posts, n	Self-reported values, mean (SD)	
Aspartate aminotransferase (U/L)	10-30	121	27.7 (17.6)	89	74.3 (85.5)	<.001	125	83.5 (322.1)	.03
Alanine aminotransferase (U/L)	10-40	143	29.5 (19.8)	99	125.6 (171.9)	<.001	139	108.7 (318.9)	.002
Alkaline phosphatase (U/L)	30-120	84	76.4 (25.2)	43	68.8 (28.7)	.06	78	70.9 (35.5)	.12
Bilirubin (mg/dL)	0.3-1.2	88	1.5 (3.8)	41	1.0 (1.9)	.18	79	1.3 (4.8)	.35
Creatinine (mg/dL)	0.6-1.2	77	1.1 (0.2)	41	1.1 (0.2)	.41	71	1.1 (0.2)	.21
Creatine kinase (U/L)	40-150	9	259.1 (344.6)	11	727.1 (675.2)	.03	8	683.1 (450.7)	.02
Luteinizing hormone (U/L)	1.24-7.8	195	6.1 (17.0)	78	4.8 (3.2)	.25	145	6.0 (7.8)	.48
Follicle-stimulating hormone (U/L)	1.4-15.4	187	4.8 (11.1)	74	3.7 (2.5)	.21	135	4.2 (3.4)	.27
High-density lipoprotein (mg/dL)	≥60	116	44.5 (17.9)	59	31.1 (14.9)	<.001	107	31.3 (14.6)	<.001
Low-density lipoprotein (mg/dL)	<100	115	80.6 (43.4)	54	100.3 (50.5)	.004	108	98.4 (50.7)	.002
Cholesterol (mg/dL)	<200	105	132.0 (49.2)	45	142.8 (61.7)	.13	84	136.4 (56.2)	.28
Triglycerides (mg/dL)	<150	84	50.6 (28.0)	40	67.7 (54.7)	.01	71	57.6 (38.6)	.10
Total testosterone (ng/dL)	270-1070	388	585.5 (233.8)	177	358.6 (325.8)	<.001	291	457.7 (346.3)	<.001
Free testosterone (pg/mL)	66-309	188	90.1 (80.8)	95	94.9 (118.8)	.33	123	86.0 (74.6)	.32
Estrogen (pg/mL)	10-50	171	32.2 (23.4)	85	41.5 (61.7)	.04	136	33.1 (18.9)	.36
Sex hormone–binding globulin (nmol/L)	14-78	146	43.9 (98.8)	82	12.8 (10.7)	.003	108	22.0 (16.7)	.01
Hemoglobin (g/dL)	14-18	58	19.9 (25.4)	24	20.8 (28.5)	.45	51	14.5 (2.9)	.07

Discussion

Recent years have seen an alarming increase in targeted marketing of SARMs on social media, where they are portrayed as safe and effective [5]. Despite being illegal to sell for human consumption, SARMs are available for purchase through numerous online platforms, including commercial websites and chemical suppliers [6]. This analysis revealed that prior, current, and prospective SARM users frequently discuss their experiences, sources, dosing regimens, perceived benefits,

XSL•FO
RenderX

laboratory results, and AEs on social media. Distributors also frequent these platforms to offer advice while marketing their products.

Our analysis revealed serious AEs of SARMs, with hepatotoxicity as a prominent concern. Comparing mean aspartate aminotransferase and alanine aminotransferase values across different stages of SARM use showed significant increases, indicative of drug-induced liver injury. Users attempted to self-manage hepatic complications through N-acetylcysteine, milk thistle, and tauroursodeoxycholic acid supplementation and hormonal imbalances with tamoxifen and enclomiphene. Our findings corroborate previous literature that has reported consistent patterns of hepatic changes attributable to SARM use [7].

This analysis has important limitations. The demographics of the sample predominantly comprised young men, limiting generalizability across gender and age. While self-reported laboratory values offer valuable preliminary insights, interindividual variability and small sample sizes in the reported data present an inherent challenge. To increase the robustness of the findings, future work should integrate traditional clinical data and include sensitivity analysis or weighted modeling. Moreover, since efforts were made to capture a wide net of hepatotoxic-specific data, the extraction methodology may underrepresent other AEs. Nonetheless, our findings highlight the need for increased awareness and regulation to address a growing public health crisis posed by SARM abuse. Until then, physicians should recognize that patients are increasingly seeking medical advice through social media and that AEs are likely underreported in the literature.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Changes in liver function tests with selective androgen receptor modulator (SARM) use. Changes in mean aspartate aminotransferase (AST), alanine aminotransferase (ALT), and alkaline phosphatase levels across 3 time periods: before (black bars), during (light gray bars), and after (dark gray bars) SARM use. Data were collected from self-reported laboratory values posted by young men on online forums. Sample sizes varied for each marker: AST (before: n=121; during: n=89; after: n=125), ALT (before: n=143; during: n=99; after: n=139), and alkaline phosphatase (before: n=84; during: n=43; after: n=78). Statistical significance is indicated by asterisks (*P<.05, **P<.01, ***P<.001). Values are presented as means with error bars representing SDs. Notable elevations were observed in both AST and ALT during and after SARM use, while alkaline phosphatase remained relatively stable.
[PNG File , 108 KB - jmir_v27i1e65031_app1.png]

References

1. Negro-Vilar A. Selective androgen receptor modulators (SARMs): a novel approach to androgen therapy for the new millennium. J Clin Endocrinol Metab 1999 Oct;84(10):3459-3462. [doi: 10.1210/jcem.84.10.6122] [Medline: 10522980]
2. Mohideen H, Hussain H, Dahiya DS, Wehbe H. Selective androgen receptor modulators: an emerging liver toxin. J Clin Transl Hepatol 2023 Feb 28;11(1):188-196 [FREE Full text] [doi: 10.14218/JCTH.2022.00207] [Medline: 36479151]
3. Vignali JD, Pak KC, Beverley HR, DeLuca JP, Downs JW, Kress AT, et al. Systematic review of safety of selective androgen receptor modulators in healthy adults: implications for recreational users. J Xenobiot 2023 May 10;13(2):218-236 [FREE Full text] [doi: 10.3390/jox13020017] [Medline: 37218811]
4. Hall E, Vrolijk MF. Androgen receptor and cardiovascular disease: a potential risk for the abuse of supplements containing selective androgen receptor modulators. Nutrients 2023 Jul 26;15(15):3330 [FREE Full text] [doi: 10.3390/nu15153330] [Medline: 37571268]
5. Efimenko IV, Valancy D, Dubin JM, Ramasamy R. Adverse effects and potential benefits among selective androgen receptor modulators users: a cross-sectional survey. Int J Impot Res 2022 Dec;34(8):757-761. [doi: 10.1038/s41443-021-00465-0] [Medline: 34471228]
6. Hahamyan H, Basaria S. Selective androgen receptor modulators-transformative drugs or heralds of the next drug epidemic? JAMA 2024 Apr 23;331(16):1359-1360. [doi: 10.1001/jama.2024.1769] [Medline: 38526456]
7. Mohamed W, Jahagirdar V, Fatima I, Ahmed MK, Jaber F, Wang K, et al. Selective androgen receptor modulators (SARMs)-induced liver injury: a case report and review of literature. Cureus 2023 Feb;15(2):e35094 [FREE Full text] [doi: 10.7759/cureus.35094] [Medline: 36945289]

Abbreviations

AE: adverse event

XSL•FO
RenderX

SARM: selective androgen receptor modulator

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by MC Gaudiano, A Wani; comments to author 29.11.24; revised version received 19.12.24; accepted 31.12.24; published 18.02.25.

Please cite as:
Joshi A, Kaune DF, Leff P, Fraser E, Lee S, Harrison M, Hazin M
Self-Reported Side Effects Associated With Selective Androgen Receptor Modulators: Social Media Data Analysis
J Med Internet Res 2025;27:e65031
URL: https://www.jmir.org/2025/1/e65031
doi:10.2196/65031
PMID:

©Aditya Joshi, Diego Federico Kaune, Phillip Leff, Emily Fraser, Sarah Lee, Morgan Harrison, Moustafa Hazin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Perspectives of Black, Latinx, Indigenous, and Asian Communities on Health Data Use and AI: Cross-Sectional Survey Study

Fatuma-Ayaan Rinderknecht[1*], MD, MS, BA; Vivian B Yang[1*], MS, BA; Mekaleya Tilahun[1], MD, BS; Jenna C Lester[2], MD

[1]San Francisco School of Medicine, University of California, San Francisco, CA, United States

[2]Department of Dermatology, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Jenna C Lester, MD
Department of Dermatology
University of California, San Francisco
1701 Divisadero St
San Francisco, CA, 94115
United States
Phone: 1 (415) 353 7800
Email: jenna.lester@ucsf.edu

Abstract

Despite excitement around artificial intelligence (AI)–based tools in health care, there is work to be done before they can be equitably deployed. The absence of diverse patient voices in discussions on AI is a pressing matter, and current studies have been limited in diversity. Our study inquired about the perspectives of racial and ethnic minority patients on the use of their health data in AI, by conducting a cross-sectional survey among 230 participants who were at least 18 years of age and identified as Black, Latinx, Indigenous, or Asian. While familiarity with AI was high, a smaller proportion of participants understood how AI can be used in health care (152/199, 76.4%), and an even smaller proportion understood how AI can be applied to dermatology (133/199, 66.8%). Overall, 69.8% (139/199) of participants agreed that they trusted the health care system to treat their medical information with respect; however, this varied significantly by income (P=.045). Only 64.3% (128/199) of participants felt comfortable with their medical data being used to build AI tools, and 83.4% (166/199) believed they should be compensated if their data are used to develop AI. To our knowledge, this is the first study focused on understanding opinions about health data use for AI among racial and ethnic minority individuals, as similar studies have had limited diversity. It is important to capture the opinions of diverse groups because the inclusion of their data is essential for building equitable AI tools; however, historical harms have made inclusion challenging.

(J Med Internet Res 2025;27:e50708) doi:10.2196/50708

KEYWORDS

augmented intelligence; artificial intelligence; health equity; dermatology; Black; Latinx; Indigenous; Asian; racial and ethnic minority communities; AI; health care; health data; survey; racism; large language model; LLM; diversity

Introduction

Despite excitement around artificial intelligence (AI)–based tools in health care, there is work to be done before they can be safely deployed. Addressing dataset diversity and racism perpetuated by large language models is paramount [1-3]. The absence of diverse patient voices in AI discussions is a pressing matter, and current studies have been limited in diversity [4,5]. Our study examined the perspectives of racial and ethnic minority patients on the use of health data in AI.

Methods

Overview

A cross-sectional survey was administered via Qualtrics to participants aged 18+ years who identified as Black, Latinx, Indigenous, or Asian. Categorical variables were summarized by frequency and percentage. The chi-square test was used to assess the relationship between responses and demographic variables. Statistical significance was based on $P<.05$. No multiple testing adjustment was performed. All analysis was

XSL•FO
RenderX

done by R (version 4.0.5; R Foundation for Statistical Computing).

Ethical Considerations

This study was exempt from approval by the University of California, San Francisco Institutional Review Board (IRB #22-36156). Informed consent was collected. All data collected were anonymized. Participants were not compensated.

Results

Overall, 230 participants enrolled, and 199 surveys were completed (Table 1). While familiarity with AI was high, a smaller proportion of participants understood how AI can be used in health care (152/199, 76.4%), and an even smaller proportion understood how AI can be applied to dermatology (133/199, 66.8%). These outcomes did not vary significantly when stratifying by demographics such as race, age, gender, income, insurance type, or schooling (all $P>.05$).

Most participants (139/199, 69.8%) agreed that they trusted the health care system to treat their medical information with respect, which varied significantly by income ($P=.045$). Patients with lower incomes often experience more structural barriers to health care, and this could be one factor contributing to our finding that they are not as likely to trust the health care system. Only 64.3% (128/199) participants felt comfortable with their medical information being used to build AI tools. Most (181/199, 91%) want to be notified if their medical information is used in such a way. This varied by race ($P=.002$) and age ($P=.03$). Most (166/199, 83.4%) agreed they should be compensated if their medical information is used to develop AI, which varied by age ($P=.045$; Figure 1).

XSL•FO
RenderX

Table 1. Demographics of survey participants.

Demographics	Value (n=199), n (%)
Gender	
Female	131 (65.8)
Male	63 (31.7)
Other or unknown	5 (2.5)
Age group (years)	
18-34	108 (54.3)
35-54	55 (27.6)
55-74	32 (16.1)
75+	3 (1.5)
Other or unknown	1 (0.5)
Race and ethnicity	
American Indian or Alaskan Native	4 (2)
Asian	57 (28.6)
Black	102 (51.3)
Hispanic or Latino	9 (4.5)
Multiracial	21 (10.6)
White	4 (2)
Other or unknown	2 (1)
Schooling	
High school or lower	19 (9.5)
Associate’s degree or trade school	30 (15.1)
Bachelor’s degree	85 (42.7)
Graduate degree	62 (31.2)
Other or unknown	3 (1.5)
Insurance	
Private insurance	109 (54.8)
Medicare or Medicaid	62 (31.2)
TRICARE or veterans	7 (3.5)
Indian Health Service	2 (1)
No health insurance	6 (3)
Other or unknown	13 (6.5)
Household income (US $)	
Less than 25,000	33 (16.6)
25,000-50,000	44 (22.1)
50,000-100,000	53 (26.6)
100,000-250,000	46 (23.1)
More than 250,000	12 (6)
Other or unknown	11 (5.5)
Geography	
Suburban	80 (40.2)
Urban	93 (46.7)
Rural	21 (10.6)

XSL•FO
RenderX

Demographics	Value (n=199), n (%)
Other or unknown	5 (2.5)
Number of visits to the doctor in a year	
0-2	64 (32.2)
3-5	71 (35.7)
6+	61 (30.7)
Other or unknown	3 (1.5)

Figure 1. Percentage of participants who would like to be compensated if their medical information was being used for AI by age group and race/ethnicity.

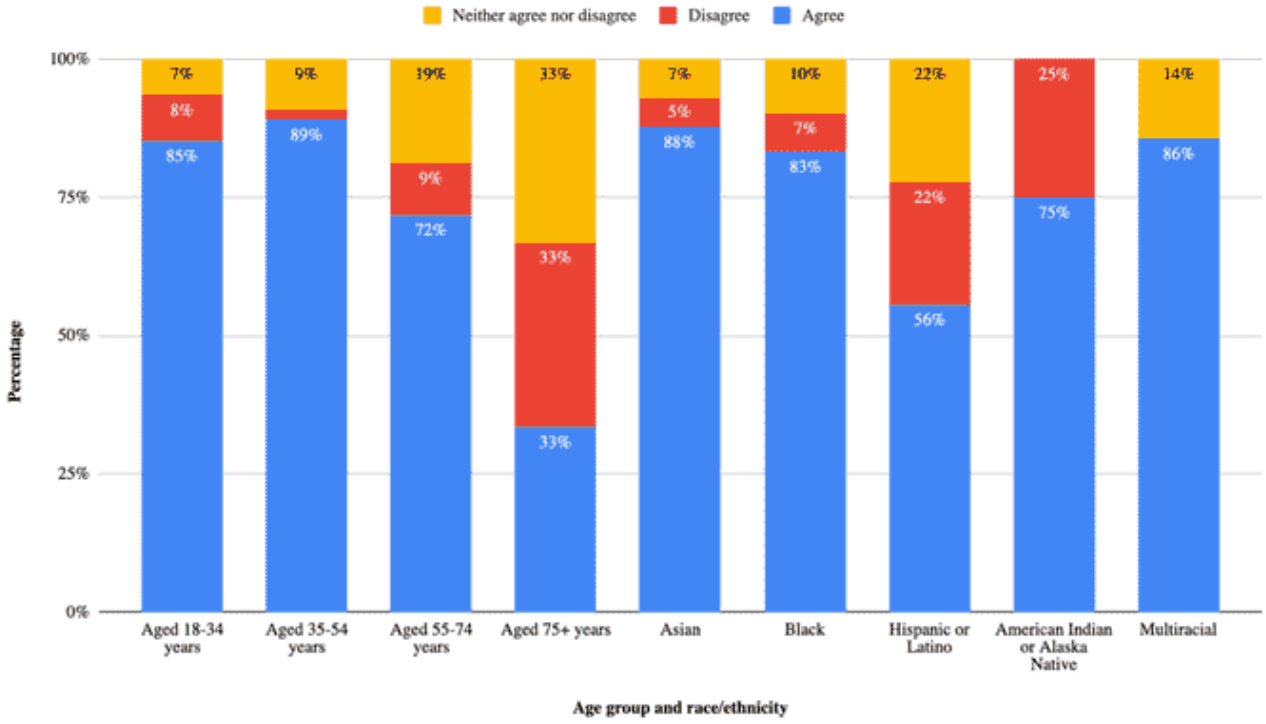

Discussion

To our knowledge, this is the first study focused on understanding opinions about health data use for AI among people who identify as Black, Latinx, Indigenous, or Asian. Similar studies have had limited diversity [4,5]. It is important to capture the opinions of diverse groups because the inclusion of their data is essential for building equitable AI tools; however, historical harms have made robust inclusion challenging. For racial and ethnic minority communities, historical experiences of racism, discrimination, and exploitation (eg, medical experimentation on populations without informed consent) may contribute to distrust in the health care system and, hence, decrease their comfort level with their data being used to develop AI systems. This suggests that AI development may not be inclusive, ethical, or aligned with the concerns of racial and ethnic minority communities enough to be safely deployed in health care settings. Ignoring these concerns could deepen distrust in AI, leading to lower rates of research participation in these already underrepresented communities, deepening biases and inequities in industries where AI is being deployed.

The educational gaps in AI uncovered by this inquiry should be addressed, as a fundamental understanding of AI is essential for patients to offer fully informed consent for health data use in AI development. There are existing initiatives addressing these educational gaps [6].

Most participants agreed that they should be notified if their data are used to build AI tools, which is not the current convention. For example, in dermatology, photos taken for clinical monitoring can be repurposed as data in AI tools without patient permission as long as they are “deidentified.”

Over 80% of participants agreed that they should be compensated when their data are used to build AI tools. Community-based participatory research shares findings and benefits with study participants. Sharing revenue with patients who contribute their data to build these valuable tools is thematically similar to sharing benefits of a work product and is something that should be explored.

Limitations include the study population. Participants were recruited from an academic medical center and research database, meaning that they may not represent the broader population. Future research should include a more diverse and

XSL•FO
RenderX

representative sample from various backgrounds and regions to enhance generalizability.

Acknowledgments

We would like to thank Li Zhang for her statistical support.

Conflicts of Interest

None declared.

References

1. Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. NPJ Digit Med 2023 Oct 20;6(1):195 [FREE Full text] [doi: 10.1038/s41746-023-00939-z] [Medline: 37864012]
2. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv 2022 Aug 12;8(32):eabq6147 [FREE Full text] [doi: 10.1126/sciadv.abq6147] [Medline: 35960806]
3. Adamson A, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol 2018 Nov 01;154(11):1247-1248 [FREE Full text] [doi: 10.1001/jamadermatol.2018.2348] [Medline: 30073260]
4. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020 May 01;156(5):501-512 [FREE Full text] [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
5. Salvador T, Gu L, Hay JL, Kurtansky NR, Masterson-Creber R, Halpern AC, et al. Consent and identifiability for patient images in research, education, and image-based artificial intelligence. JAMA Dermatol 2024 Apr 01;160(4):470-472 [FREE Full text] [doi: 10.1001/jamadermatol.2024.0084] [Medline: 38477909]
6. AI4ALL. URL: https://ai-4-all.org/ [accessed 2024-11-25]

Abbreviations

AI: artificial intelligence

Edited by K Williams; submitted 08.07.24; peer-reviewed by GK Gupta, J Lopes; comments to author 19.09.24; revised version received 09.11.24; accepted 11.12.24; published 21.02.25.

Please cite as:
Rinderknecht FA, Yang VB, Tilahun M, Lester JC
Perspectives of Black, Latinx, Indigenous, and Asian Communities on Health Data Use and AI: Cross-Sectional Survey Study
J Med Internet Res 2025;27:e50708
URL: https://www.jmir.org/2025/1/e50708
doi:10.2196/50708
PMID:

©Fatuma-Ayaan Rinderknecht, Vivian B Yang, Mekaleya Tilahun, Jenna C Lester. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Gender Differences in X (Formerly Twitter) Use Among Oncology Physicians at National Cancer Institute–Designated Cancer Centers: Cross-Sectional Study

Vivian Tieu[1], BS; Sungjin Kim[2], MS; Minji Seok[2], MD; Leslie Ballas[2], MD; Mitchell Kamrava[2], MD, MSc; Katelyn M Atkins[2], MD, PhD

[1]California University of Science and Medicine, Colton, CA, United States

[2]Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Corresponding Author:
Katelyn M Atkins, MD, PhD
Department of Radiation Oncology
Cedars-Sinai Medical Center
8700 Beverly Blvd
Los Angeles, CA, 90048
United States
Phone: 1 310 423 2178
Email: katelyn.atkins@cshs.org

Abstract

This cross-sectional study evaluated gender parity in the oncology workforce on social media, demonstrating that women oncologists are enriched on X, with higher self-engagement, suggestive of a heightened motivation for professional X use.

(J Med Internet Res 2025;27:e66054) doi:10.2196/66054

KEYWORDS

social media; gender disparities; gender differences; cross-sectional study; twitter; oncology

Introduction

In 2023-2024, women comprised 55% of matriculating US medical students but continue to be underrepresented in academic medicine [1,2]. This is particularly true in the workforce and leadership of oncology disciplines [2]. Social media platforms have gained popularity for professional development [3], though data characterizing gender parity in the social media oncology workforce are lacking and studied herein.

Methods

Overview

Twenty regionally distinct National Cancer Institute (NCI)–designated cancer center websites were accessed from December 2023 to July 2024 (Table 1) and physician demographic information was collected. Apparent gender (binary) was assigned by a single researcher (VT) using names, pronouns, and/or public profile images. Among physicians on X, publicly available data were manually collected. Physicians with missing numbers of “likes” (public reporting discontinued June 12, 2024) were excluded. Statistical analysis by X use and gender was performed using Wilcoxon rank sum and chi-square (or Fisher exact) tests for continuous and categorical variables, respectively. Analyses were performed using R (version 4.0.5; The R Foundation) with 2-sided tests with $P \leq .05$.

XSL•FO
RenderX

Table 1. National Cancer Institute–designated cancer centers included in this study.

Institution name	City	State	Physicians (total; n=2908), n (%)	Physicians on X (n=1068), n (%)
O'Neal Comprehensive Cancer Center	Birmingham	AL	45 (1.5)	29 (2.7)
University of Arizona Cancer Center	Tucson	AZ	54 (1.9)	23 (2.2)
Mayo Clinic Cancer Center	Phoenix, Jacksonville, and Rochester	AZ, FL, and MN	156 (5.4)	70 (6.6)
Chao Family Comprehensive Cancer Center	Orange	CA	126 (4.3)	36 (3.4)
University of Colorado Cancer Center	Aurora	CO	147 (5.1)	51 (4.8)
Yale Cancer Center	New Haven	CT	353 (12.1)	148 (13.9)
Sylvester Comprehensive Cancer Center	Miami	FL	149 (5.1)	71 (6.6)
Winship Cancer Institute	Atlanta	GA	262 (9.0)	106 (9.9)
University of Hawaii Cancer Center	Honolulu	HI	28 (1.0)	4 (0.4)
Robert H. Lurie Comprehensive Cancer Center	Chicago	IL	203 (7.0)	86 (8.1)
Indiana University Melvin and Bren Simon Comprehensive Cancer Center	Indianapolis	IN	148 (5.1)	56 (5.2)
Holden Comprehensive Cancer Center	Iowa City	IA	87 (3.0)	24 (2.2)
The University of Kansas Cancer Center	Kansas City	KS	109 (3.7)	29 (2.7)
Markey Cancer Center	Lexington	KY	62 (2.1)	17 (1.6)
Sidney Kimmel Comprehensive Cancer Center	Baltimore	MD	314 (10.8)	110 (10.3)
The Barbara Ann Karmanos Cancer Institute	Detroit	MI	81 (2.8)	25 (2.3)
Alvin J. Siteman Cancer Center	St. Louis	MO	315 (10.8)	118 (11.0)
Dartmouth Cancer Center	Lebanon	NH	153 (5.3)	24 (2.2)
Rutgers Cancer Institute of New Jersey	New Brunswick	NJ	116 (4.0)	41 (3.8)

Ethical Considerations

This cross-sectional study used publicly available data and was therefore exempt from ethical approval per the Cedars-Sinai Medical Center institutional review board (STUDY00003292). STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines were followed (Multimedia Appendix 1).

Results

In total, 2908 physicians' profiles were analyzed, of which 37% (n=1068) were on X (Table 2). There was a greater proportion of women (vs men) physicians on X in the Northeast (35.1% vs 27.1%) but a smaller proportion in the Midwest (32.2% vs 39.6%; *P*=.03; Table 2). X users accounted for a higher proportion of women (39% vs 35%; *P*=.05), were more likely to hold leadership titles (*P*<.001), and had an advanced dual degree (33% vs 25%; *P*<.001) than non–X users. Among those on X, women (vs men) were less likely to have "professor" status (25% vs 41%; *P*<.001) and leadership titles (*P*=.006) but more likely to have a master's in public health (9% vs 5%; *P*=.03).

XSL•FO
RenderX

Table 2. Characteristics stratified by apparent gender on X.

Variable	Men (n=654)	Women (n=413)	*P* value[a]
Region			.03
Northeast	177 (27.1)	145 (35.1)	
Midwest	259 (39.6)	133 (32.2)	
South	143 (21.9)	88 (21.3)	
West	75 (11.5)	47 (11.4)	
Faculty type			<.001
None or unknown	37 (5.7)	18 (4.4)	
Instructor or clinician	0 (0)	1 (0.2)	
Assistant	173 (26.5)	172 (41.7)	
Associate	177 (27.1)	117 (28.3)	
Professor	267 (40.8)	105 (25.4)	
Number of titles including chair, director, or codirector			.006
0	324 (49.5)	247 (60.0)	
1	168 (25.7)	89 (21.6)	
2	94 (14.4)	50 (12.1)	
3+	68 (10.4)	26 (6.3)	
Subspecialty			<.001
Medical oncology	231 (35.4)	166 (40.2)	
Radiation oncology	53 (8.1)	38 (9.2)	
Gyn oncology	12 (1.8)	18 (4.4)	
Surgical oncology	44 (6.7)	46 (11.1)	
Other	313 (47.9)	145 (35.1)	
Dual degree			
PhD			.05
No	536 (82.0)	357 (86.4)	
Yes	118 (18.0)	56 (13.6)	
MS			.92
No	575 (87.9)	364 (88.1)	
Yes	79 (12.1)	49 (11.9)	
Master of public health			.03
No	620 (94.8)	378 (91.5)	
Yes	34 (5.2)	35 (8.5)	
Any advanced dual degree			.86
No	442 (67.6)	277 (67.1)	
Yes	212 (32.4)	136 (32.9)	
Length of training since graduation from medical school (years), median (IQR)	7 (6-9)	7 (6-8)	.02
X use variables unadjusted for time			
Number of followers, median (IQR)	389.5 (146-1119)	305 (112-863)	.002
Number of accounts followed by the physicians, median (IQR)	223 (80-521)	219 (82-478)	.78
Total number to tweets, median (IQR)	168.5 (38-675)	131 (22-514)	.02

XSL•FO
RenderX

Variable	Men (n=654)	Women (n=413)	*P* value[a]
Total number of media posts, median (IQR)	15 (2-67)	9 (2-47)	.09
Number of liked posts, median (IQR)	352 (35.5-1683)	441 (52-1697.5)	.23
X use variables adjusted for time			
Time on X (years), median (IQR)	8.1 (5.3-11.6)	6.7 (4.5-10.3)	<.001
Average number of followers per year on X, median (IQR)	55.3 (19.4-159.5)	48 (19.6-120.6)	.12
Average number of accounts followed per year on X, median (IQR)	29.7 (10.7-71.4)	34 (15.2-73.0)	.09
Average number to tweets per year on X, median (IQR)	22.0 (5.2-91.3)	21.4 (4.6-72.9)	.30
Average number of media posts per year on X, median (IQR)	1.9 (0.3-8.6)	1.6 (0.3-7.8)	.54
Average number of liked posts per year on X, median (IQR)	45.9 (5.6-235.8)	76.6 (9.7-260.4)	.02
Thematic content of X biography			
Job roles			.82
No mention	111 (17.1)	68 (16.5)	
Mention	540 (83.0)	344 (83.5)	
Specialty			.70
No mention	128 (19.7)	77 (18.7)	
Mention	523 (80.3)	335 (81.3)	
Being a parent			.002
No mention	583 (89.6)	342 (83.0)	
Mention	68 (10.5)	70 (17.0)	
Spouse			.25
No mention	598 (91.9)	370 (89.8)	
Mention	53 (8.1)	42 (10.2)	
Institution			.98
No mention	209 (32.1)	132 (32.0)	
Mention	442 (67.9)	280 (68.0)	
Personal interests (eg, hobbies and activities)			.09
No mention	565 (86.8)	342 (83.0)	
Mention	86 (13.2)	70 (17.0)	

[a]*P* values were determined using the Wilcoxon rank sum test for continuous variables and chi-square or Fisher exact test for categorical variables, where appropriate.

Overall, women (vs men) had significantly fewer followers (mean 305 vs 390, *P*=.002) and tweets (mean 131 vs 169, *P*=.02). Adjusting for fewer years on X, women showed similar influence as men (mean 48 vs 55 followers per X-year, *P*=.12) but a higher rate of liking posts (mean 77 vs 46 per X-year, *P*=.02). Women (vs men) were more likely to mention being a parent in their biography (17% vs 10%, *P*=.002), but no differences were noted in other content variables (*P*>.05).

Discussion

Principal Results

In this cross-sectional study evaluating gender parity in the oncology workforce on social media, we observed women physicians on X being less likely to hold professor status and leadership titles. As seen on X, particularly in male-dominated fields of radiation and surgical oncology, the proportion of women was significantly higher than published workforce estimates (radiation oncology: 42% vs 31%, *P*=.04; surgical oncology: 51% vs 39%, *P*=.03) [2]. These data are suggestive of motivational and/or behavioral differences in X use by gender.

XSL•FO
RenderX

Comparisons With Prior Work

We observed that women were more likely to mention being a parent, consistent with studies describing higher engagement in fostering support and community [3]. Further, women "liked" more posts, which perhaps parallels the expected levels of "friendliness" and tone softening in women's professional communications, which has been associated with increased emotional labor. Indeed, content language analyses have demonstrated that women use exclamation points more frequently than men and as markers of "friendly interaction" [4]. Behavioral psychology studies report that women have higher engagement in emotional labor practices, which may drain resources without equitable rewards, contributing toward the underrepresentation of women in leadership positions [5]. These findings support continued evaluation of motivational and/or behavioral differences in professional social media use.

Men physicians are more likely to hold "verified" X accounts (verification is thought to add a degree of validity) [6] and report professional benefits from social media use, such as invited talks [7], consistent with studies reporting that women X users face challenges in popularity and influence at academic meetings despite comparable activity [8]. Thus, while social media offers a platform for connection and visibility [9], these findings underscore the need for ensuring equitable opportunities moving forward.

Limitations

Institutional websites may be inaccurate, incomplete, and/or outdated. Gender classification may be inaccurate and impart classification bias. Publicly available X data are more limited than prior studies [10]. Verified status was not analyzed due to low occurrence (0.2% of accounts).

Conclusions

Women oncologists are enriched on X, with higher self-engagement, suggestive of a heightened motivation for professional X use. Future longitudinal studies examining the role of emotional labor and network support in motives for social media use are warranted.

Acknowledgments

VT, SK, and KMA had full access to the data. MK reports receiving personal fees from Theragenics, Alessa, GTMedical, and Springer, outside the submitted work. KMA reports honoraria from OncLive, outside of the submitted work. The remaining authors have no disclosures. All authors have read and agreed to the published version of the manuscript.

Authors' Contributions

VT and KMA drafted the manuscript. VT, SK, MS, LB, MK, and KMA reviewed all analyses and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE guidelines.
[PDF File (Adobe PDF File), 18 KB - jmir_v27i1e66054_app1.pdf]

References

1. 2024 FACTS: Applicants and Matriculants Data. AAMC. URL: https://www.aamc.org/data-reports/students-residents/data/2023-facts-applicants-and-matriculants-data [accessed 2024-07-19]
2. Chowdhary M, Chowdhary A, Royce TJ, Patel KR, Chhabra AM, Jain S, et al. Women's representation in leadership positions in academic medical oncology, radiation oncology, and surgical oncology programs. JAMA Netw Open 2020 Mar 02;3(3):e200708 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0708] [Medline: 32159809]
3. Ibrahim H, Anglade P, Abdel-Razig S. The use of social media by female physicians in an international setting: a mixed methods study of a group WhatsApp chat. Womens Health Rep (New Rochelle) 2020 Jan 01;1(1):60-64 [FREE Full text] [doi: 10.1089/whr.2019.0015] [Medline: 33786474]
4. Waseleski C. Gender and the use of exclamation points in computer-mediated communication: an analysis of exclamations posted to two electronic discussion lists. J Comp Mediated Comm 2006 Jul;11(4):1012-1024. [doi: 10.1111/j.1083-6101.2006.00305.x]
5. Vial AC, Cowgill CM. Heavier lies her crown: gendered patterns of leader emotional labor and their downstream effects. Front Psychol 2022 Aug 29;13:849566 [FREE Full text] [doi: 10.3389/fpsyg.2022.849566] [Medline: 36106035]
6. Rupert D, Shah K, Chen BY, O'Glasser AY, Schiml M, Jain S, et al. Sex and location differences in verification status of physician-held social media platform accounts. JAMA Netw Open 2022 Aug 01;5(8):e2225671 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.25671] [Medline: 35939304]
7. Woitowich NC, Arora VM, Pendergrast T, Gottlieb M, Trueger NS, Jain S. Gender differences in physician use of social media for professional advancement. JAMA Netw Open 2021 May 03;4(5):e219834 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.9834] [Medline: 33983403]

XSL•FO
RenderX

8. Berger T, Payan N, Fleury E, Davey A, Bryce-Atkinson A, Vasquez Osorio E, et al. Gender-related and geographic trends in interactions between radiotherapy professionals on Twitter. Phys Imaging Radiat Oncol 2022 Oct;24:129-135 [FREE Full text] [doi: 10.1016/j.phro.2022.11.002] [Medline: 36439328]
9. Lewis JD, Fane KE, Ingraham AM, Khan A, Mills AM, Pitt SC, et al. Expanding opportunities for professional development: utilization of Twitter by early career women in academic medicine and science. JMIR Med Educ 2018 Jul 23;4(2):e11140 [FREE Full text] [doi: 10.2196/11140] [Medline: 30037788]
10. Zhu JM, Pelullo AP, Hassan S, Siderowf L, Merchant RM, Werner RM. Gender differences in Twitter use and influence among health policy and health services researchers. JAMA Intern Med 2019 Dec 01;179(12):1726-1729 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4027] [Medline: 31609393]

Abbreviations

NCI: National Cancer Institute
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Mavragani; submitted 02.09.24; peer-reviewed by V Sekhon, L Chinthala, Z Ehtesham, T Akindahunsi; comments to author 29.11.24; revised version received 28.12.24; accepted 11.02.25; published 11.03.25.

Please cite as:
Tieu V, Kim S, Seok M, Ballas L, Kamrava M, Atkins KM
Gender Differences in X (Formerly Twitter) Use Among Oncology Physicians at National Cancer Institute–Designated Cancer Centers: Cross-Sectional Study
J Med Internet Res 2025;27:e66054
URL: https://www.jmir.org/2025/1/e66054
doi:10.2196/66054
PMID:

©Vivian Tieu, Sungjin Kim, Minji Seok, Leslie Ballas, Mitchell Kamrava, Katelyn M Atkins. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

eHealth Literacy 3.0: Updating the Norman and Skinner 2006 Model

Ariesta Milanti[1*], PhD, RN; Cameron Norman[2*], PhD, MDesCE; Dorothy Ngo Sheung Chan[3*], PhD, RN; Winnie Kwok Wei So[3*], PhD, RN; Harvey Skinner[2*], PhD, CPsych

[1]Strategic Delivery Unit, Ministry of Health - Republic of Indonesia, Jakarta, Indonesia

[2]Dahdaleh Institute for Global Health Research, York University, Toronto, ON, Canada

[3]The Nethersole School of Nursing, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[*]all authors contributed equally

Corresponding Author:
Winnie Kwok Wei So, PhD, RN
The Nethersole School of Nursing
The Chinese University of Hong Kong
7/F, Esther Lee Building
Hong Kong
China (Hong Kong)
Phone: 852 3943 1072
Fax: 852 2603 6041
Email: winnieso@cuhk.edu.hk

Abstract

This paper advances the "eHealth Literacy 3.0" model following Norman and Skinner's 2006 original eHealth literacy 1.0 model and Norman's 2011 2.0 update, and introduces a corresponding revision to the eHealth Literacy Scale (eHEALS) measurement instrument (eHEALS-R).

(J Med Internet Res 2025;27:e70112) doi:10.2196/70112

KEYWORDS

eHealth literacy; eHealth literacy definition; eHealth literacy model; eHealth literacy measurement

Introduction

Digital technology is transforming, not just evolving. A landmark definition and model of eHealth literacy was developed by Norman and Skinner [1] during the early 2000s and updated by Norman [2] when digital technology involved far fewer ways to connect and interact. Two decades later, the proliferation of digital technology has become ubiquitous. Although there are now other definitions in the literature [3-6], Norman and Skinner's [1] version continues to be the most cited model [7,8]. An updated definition that better fits the current and foreseeable future of digital technology is overdue. This paper serves as a prompt for introducing and stimulating collaborative research on our "eHealth Literacy 3.0" model and revised eHealth Literacy Scale (eHEALS-R).

Methods

We reviewed the published literature on eHealth literacy to consolidate the critiques and suggestions regarding Norman and Skinner's [1] eHealth literacy definition or model. The consolidated themes were synthesized to bolster the original eHealth literacy conceptual model into version 3.0, which reflects current and anticipated use contexts.

Results

Norman and Skinner [1] defined eHealth literacy as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." We have revamped the definition of eHealth literacy as "the ability to engage with digital technologies in effective, safe, and helpful ways to achieve health goals."

This version encapsulates an expanded concept of the original model reflecting effectiveness, safety, and helpfulness. The core idea of the original lily model is that eHealth literacy is the fused "pistil" fed by the 6 literacies ("petals"), which ties them together [1]. The new model draws more elements from the lily flower's anatomy, whose positions and functions help to better

XSL•FO
RenderX

visualize the constructs of literacies and their interrelationships (Figure 1).

Traditional literacy (referring to reading, writing, and speaking skills) embodies the stalk, from which the other 6 literacies are stemmed. Traditional literacy is a precondition, without which no other literacies can grow.

Two sepals are representative of health and science literacies. As in flowers, the sepals' main functions are to protect and support the flowers. While health literacy implies understanding and engaging with the health system, science literacy implies understanding the key characteristics of science and how it is done. These two literacies serve as a strong ground for the other literacies above them, including eHealth literacy.

Figure 1. The 2006 lily model and the new lily model.

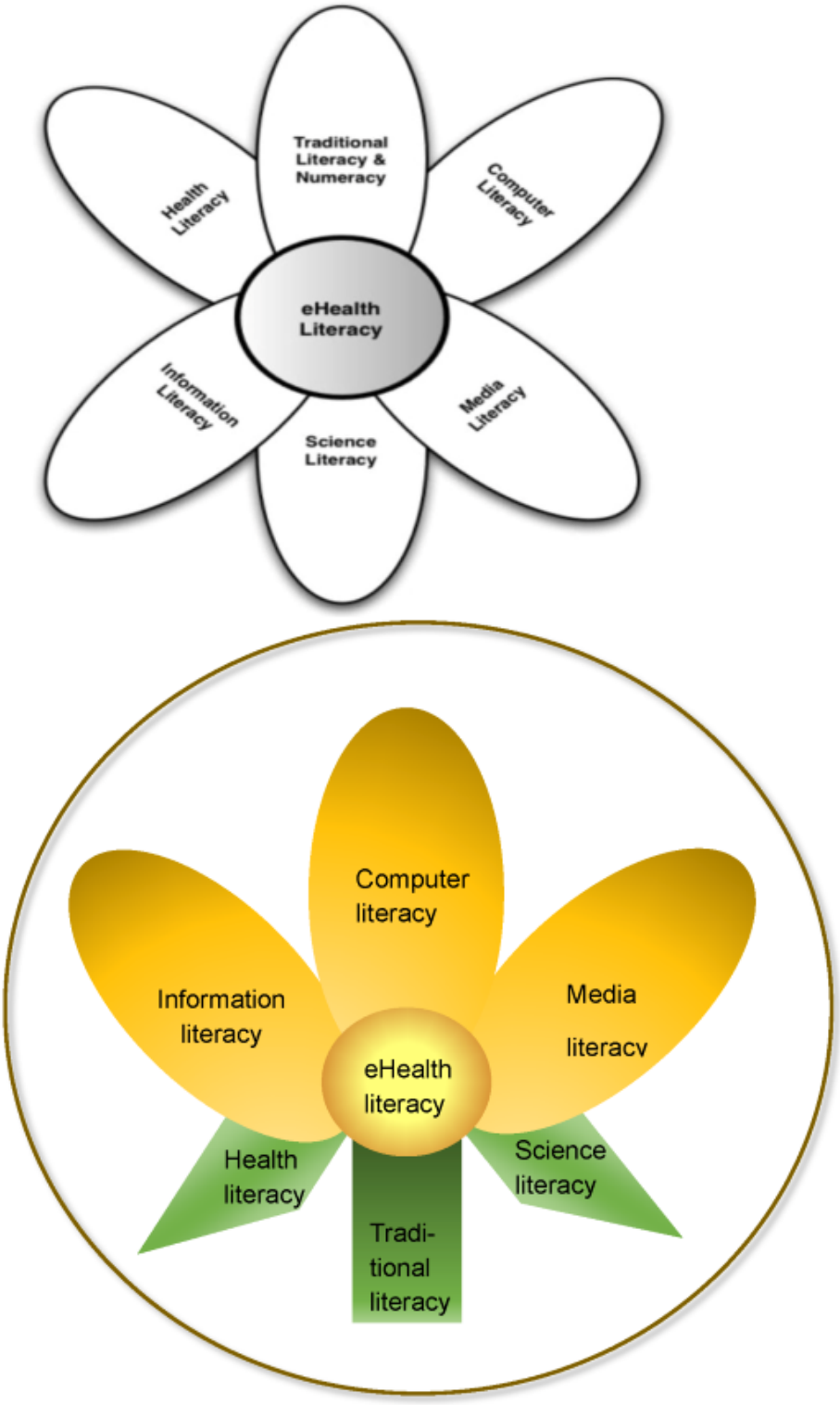

XSL•FO
RenderX

Three petals illustrate (1) computer literacy, (2) information literacy, and (3) media literacy. These petals are more pronounced in the new model, with their enhanced definitions encompassing technological advancement and contemporary paradigms. As with flower petals, which are typically the most attractive part of a flower, computer literacy is needed to keep up with the mainstream development of digital technology. Information and media literacies serve as counterparts to optimize the outcomes of the technology while mitigating the potential risks and harms posed by technological advancement.

At the center of the flower lies the pistil, representing eHealth literacy. Being held, supported, and fed by the other literacies, a person uses eHealth literacy to interact with digital technology to bear the "fruit," that is, their health goals and needs.

Overall, the new model is shaped in a circle representing contextual factors that influence the way a person engages with digital technologies in an effective and safe manner.

Table 1 compares the original and new lily models. Modifications reflect current health technology trends with eHealth encompassing well-being, care, monitoring, and assessment, while introducing new security and privacy concerns. The modified constructs and their interrelationships are presented as initial steps, with theoretical elaboration and collaborative empirical research to follow.

Table 1. Comparison of two lily models. Key changes in the new model are highlighted in italics.

Element	Description	
	The 2006 lily model	The new lily model
Traditional literacy	Reading, writing, and speaking skills.	Reading, writing, and speaking skills.
Health literacy	The ability to understand and engage with the health system to make appropriate health decisions and act upon them.	The ability to understand and engage with *health source material* and the health system, make appropriate and *safe* health decisions, and act upon them.
Science literacy	The ability to understand the nature, objectives, methods, practice, limitations, and opportunities of science and/or knowledge.	The ability to understand the nature, objectives, methods, practice, limitations, perspectives and opportunities of science and/or knowledge.
Information literacy	The ability to seek, find, understand, appraise, and use information.	The ability to seek, find, understand, appraise, filter, use, and *create information ethically and safely.*
Computer literacy	The skills to use computers and to adapt to new technologies in solving problems.	*The skills to learn, adapt to, and engage with digital technologies to actively promote and support health, care and well-being.*
Media literacy	Critical thinking skills to analyze, evaluate, and use media content to form a judgment and guide actions.	*Critical thinking and conscious thinking skills* to analyze, evaluate, and use media content to form a judgment and guide actions. Critical thinking involves the use of logic to evaluate media content and guide decisions, while conscious thinking relies on self-awareness, sensory insight, clarity, and emotional control [9].
eHealth literacy	The ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem.	*The ability to engage with digital technologies in effective, safe, and helpful ways to achieve health goals.*

Discussion

As digital technology evolves, so does our understanding of the skills needed to navigate it for wellness and health care. While eHealth literacy's fundamental skills remain, model 3.0 reflects our expanded digital technology ecosystem and tool set. Greater emphasis is placed on the outcome: achieving health goals effectively and safely. Research is now needed to compare and validate model 3.0. Finally, model 3.0 has been used to develop a revised 10-item version of the eHEALS [10], with items measuring each element of the model.

Data Availability

We encourage collaboration on the evaluation and language adaptations of the new eHEALS-R (revised eHealth Literacy Scale) in diverse settings. Please contact CN and/or HS for details.

Conflicts of Interest

None declared.

References

1. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

XSL•FO
RenderX

2. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
3. Cain K, Coldwell-Neilson J. Digital fluency – a dynamic capability continuum. Australas J Educ Technol 2024 Feb 04;40(1):42-56. [doi: 10.14742/ajet.8363]
4. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. J Med Internet Res 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: 10.2196/10175] [Medline: 30279155]
5. Gilstad H, Sand K, Solbjør M, Melby L. Deconstructing (e)health literacy: aspects that promote and inhibit understanding of health information in breast cancer patient pathways. Int J Qual Stud Health Well-being 2022 Dec 21;17(1):2137961 [FREE Full text] [doi: 10.1080/17482631.2022.2137961] [Medline: 36268568]
6. Magsamen-Conrad K, Wang F, Tetteh D, Lee Y. Using technology adoption theory and a lifespan approach to develop a theoretical framework for eHealth literacy: extending UTAUT. Health Commun 2020 Nov 22;35(12):1435-1446. [doi: 10.1080/10410236.2019.1641395] [Medline: 31328567]
7. Nwagwu WE. A citation study of global research on e-Health literacy. GKMC 2024 Mar 21. [doi: 10.1108/gkmc-07-2023-0259]
8. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 06;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
9. Khan G, Feroz A. “Use” as a conscious thought: towards a theory of “Use” in autonomous things. Aust J Inf Syst 2024;28 [FREE Full text] [doi: 10.3127/ajis.v28.4611]
10. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

Abbreviations

eHEALS: eHealth Literacy Scale
eHEALS-R: revised eHealth Literacy Scale

Edited by A Coristine; submitted 16.12.24; peer-reviewed by YA Qadri, M Bardus, JR Bautista, E Neter; comments to author 06.02.25; revised version received 26.02.25; accepted 28.02.25; published 11.03.25.

Please cite as:
Milanti A, Norman C, Chan DNS, So WKW, Skinner H
eHealth Literacy 3.0: Updating the Norman and Skinner 2006 Model
J Med Internet Res 2025;27:e70112
URL: https://www.jmir.org/2025/1/e70112
doi:10.2196/70112
PMID:

©Ariesta Milanti, Cameron Norman, Dorothy Ngo Sheung Chan, Winnie Kwok Wei So, Harvey Skinner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Primary Care Practice Factors Associated With Telehealth Adoption in the United States: Cross-Sectional Survey Analysis

Matthew Mackwood[1,2], MD, MPH; Elliott Fisher[1,2,3], MD; Rachel O Schmidt[2], MS; A James O'Malley[2,4], PhD; Hector P Rodriguez[5], PhD; Stephen Shortell[5], PhD; Ellesse-Roselee Akré[2,6], PhD; Alena Berube[2], MS; Karen E Schifferdecker[1,2], PhD, MPH

[1]Department of Community & Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[2]The Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[3]Department of Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[4]Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[5]School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[6]Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Matthew Mackwood, MD, MPH
Department of Community & Family Medicine
Geisel School of Medicine
Dartmouth College
One Medical Center Drive
Hanover, NH, 03756
United States
Phone: 1 6036504000
Email: matthew.b.mackwood@hitchcock.org

Abstract

In this national study of primary care practice–level factors associated with telehealth adoption in 2022, we found that training and assisting patients with the use of telehealth, broadband expansion efforts, and a higher proportion of low-income patients were associated with higher practice-level telehealth use, suggesting both opportunities for telehealth expansion and potential populations with higher need for its use.

(J Med Internet Res 2025;27:e70404) doi:10.2196/70404

KEYWORDS

telehealth; telemedicine; remote consultation; primary health care; general practice; internet access; health policy; health care economics and organizations; access to primary care; digital divide; vulnerable populations; medically underserved area

Introduction

In 2021, the National Academy of Medicine called for further study of telehealth to help strengthen primary care as part of their "Implementing High-Quality Primary Care in the United States" report [1]. Past studies highlight a range of patterns in telehealth use by patient-level factors [2], but primary care practice–level capabilities, such as resources to enable telehealth, have not been systematically assessed. Organizational and contextual factors, such as a federally qualified health center (FQHC) designation and a practice's neighborhood characteristics, may also be associated with telehealth use. We analyzed a national survey of US adult primary care practices to examine relationships between practice-level factors and telehealth adoption.

Methods

Overview

We analyzed cross-sectional data from the 2022-2023 National Survey of Healthcare Organizations and Systems, which surveyed practice leaders about telehealth use in 2021, practice payment models, and care delivery structures and processes (1245/3499 people responded to the survey, for a 36% response rate). Survey administration information and data sources are described in Multimedia Appendix 1.

XSL•FO
RenderX

The primary outcomes are the proportion of all outpatient visits done via telehealth (real-time audiovisual or audio-only, eg, telephone) and the proportion of telehealth visits done audio-only. We categorized practices by quartile of telehealth use to compare practice-level differences by characteristic. We estimated multivariable linear regression models for all cases (N=1071) and used average marginal effects to generate estimates and 95% CIs for the primary outcome measures. Model covariates included practice ownership, clinician staffing, FQHC status, and telehealth-enabling resources. All analyses were completed in Stata (version 17; StataCorp), and all models used robust estimates of variance and 2-tailed *P* values, with *P*<.05 set as the threshold for significance. Weights were used in all analyses to account for nonresponse probabilities, detailed in Multimedia Appendix 1.

Ethical Considerations

This study was deemed exempt by the Dartmouth College Committee for the Protection of Human Subjects (00032337). As practices were the unit of analysis, participants received an information sheet in lieu of informed consent. Practices were deidentified prior to analysis.

Results

The median use of telehealth for practice encounters was 20% (IQR 10%-35%), and the median proportion of audio-only telehealth visits was 29% (IQR 17%-50%). High-telehealth-use practices were more likely to care for a high proportion of uninsured patients, to have expanded broadband access for patients, and to have trained patients to use telehealth (Table 1).

In multivariable regression analyses (Figure 1), a high payer mix of uninsured patients and high broadband expansion corresponded to higher telehealth use (*P*=.02 and *P*=.008, respectively), while rurality corresponded to lower telehealth use (*P*=.008). Audio-only telehealth use was higher among FQHCs compared to other practices (*P*=.02), and assisting patients with using telehealth was associated with both higher telehealth use and a smaller proportion of audio-only telehealth use (*P*=.002 and *P*=.001, respectively).

XSL•FO
RenderX

Table 1. Practice demographics (N=1071), overall and stratified by telehealth use quartile.

Characteristics[a]	Overall, %	Practice quartile for telehealth use, %				*P* value[b]
		Lowest (0%-10%)	Second (>10%-20%)	Third (>20%-35%)	Highest (>35%)	
Practice structure						
Ownership						.53
Independent	25.4	27.9	27.4	24.6	21.2	
Physician group	7.1	5.6	9.5	9.8	4.1	
Hospital	15.7	18.8	15.4	9.4	17.2	
Health system	37.9	38.5	37.4	40.3	39.9	
Federally qualified health center or look-alike	13.6	8.9	10	15.6	17.4	
Physician count, n						.05
0-4	39.5	37.9	46.7	27.4	43.9	
5-9	37	32	37.4	52	30.3	
10-19	12.6	13.8	9.7	9.6	16.7	
>20	10.9	16.3	6.2	11	9.1	
Advanced practice provider (physician assistant, advanced registered practice nurse) count, n						.04
Zero	21.4	28.8	24.7	7.5	20.3	
1 or 2	27.2	19.7	34.1	33.8	23.7	
3 or 4	23.9	20.8	21.1	25	29.8	
5 to 10	18	17	12.7	24.6	19.5	
>10 ()	9.5%	13.7	7.5	9.1	6.6	
Practice financial characteristics						
Self-reported poor financial health	8.7	8.2	6.6	10.5	10.3	.80
Current alternative payment model participation[c]	81.5	78.1	86	83.6	79.5	.51
Impacted by physician workforce shortages	56.4	49.9	55.1	57.9	64.5	.41
Impacted by staff shortages	74	68.4	73.1	77.8	78.5	.33
Payer mix (>20% of revenue from listed sources)						
Commercial	84.1	86.3	84.1	80.9	84.2	.68
Medicare	80.3	88.8	74.4	76.1	79.5	.10
Medicaid	34.1	31.8	33.4	35.1	36.7	.94
Uninsured	6.2	2.9	6	8.5	8.6	.04
Telehealth enablement factors						
Facilitated telehealth for patients by improving broadband access	27.7	19.7	24.7	25.7	42.6	.02
Facilitated telehealth by assisting or training patients to use telehealth	70.5	52.6	75.7	74.3	84.5	<.001
Platform for video visit integrated with electronic health record	69.4	70.2	69.3	68.1	69.7	.99
Census tract–level indicators						
Practice in a rural location	7.4	11.6	6.1	7.5	3.6	.10
Area Deprivation Index quartile						.40
Most deprived	26.5	18.1	27	24.7	37.9	
2nd quartile	(32.9	38.8	29.5	35.8	26.5	
3rd quartile	22.9	24.6	25.3	22.2	18.9	

XSL•FO
RenderX

Characteristics[a]	Overall, %	Practice quartile for telehealth use, %				*P* value[b]
		Lowest (0%-10%)	Second (>10%-20%)	Third (>20%-35%)	Highest (>35%)	
Least deprived	17.8	18.6	18.2	17.3	16.7	
Internet speed measures (Mbps), mean (SD)						
Median download speed	76 (1.5)	70 (2.5)	78 (3.3)	73 (3.9)	83 (3.5)	.19
Median upload speed	16 (0.9)	15 (1.0)	17 (1.7)	14 (1.2)	18 (2.9)	.39

[a]Characteristics are reported as weighted percentages unless otherwise noted. Details on weighting are provided in Multimedia Appendix 1.

[b]Differences between quartiles; generated via the χ^2 test with the exception of download and upload speed, which were generated with the adjusted Wald test.

[c]Includes any engagement in accountable care organization and capitated payment contracts, which are alternatives to fee-for-service billing.

Figure 1. Adjusted Telehealth Use Rates and Audio-Only Telehealth Proportions by Select Factors. Covariates included in the multivariable linear regression model were as follows: practice ownership type, US census region, clinician staffing size, Area Deprivation Index quartile, median download speed at the census tract level, reported financial health, whether a practice reported impacts from staff shortages, alternative (to fee-for-service) payment model participation, and a payer mix that was >20% Medicare. Each covariate was significant in univariate regression at the *P*<.1 level; in preliminary runs when 2 factors had a Spearman correlation coefficient >0.5 or <–0.5, the least significant factor was excluded. EHR: electronic health record.

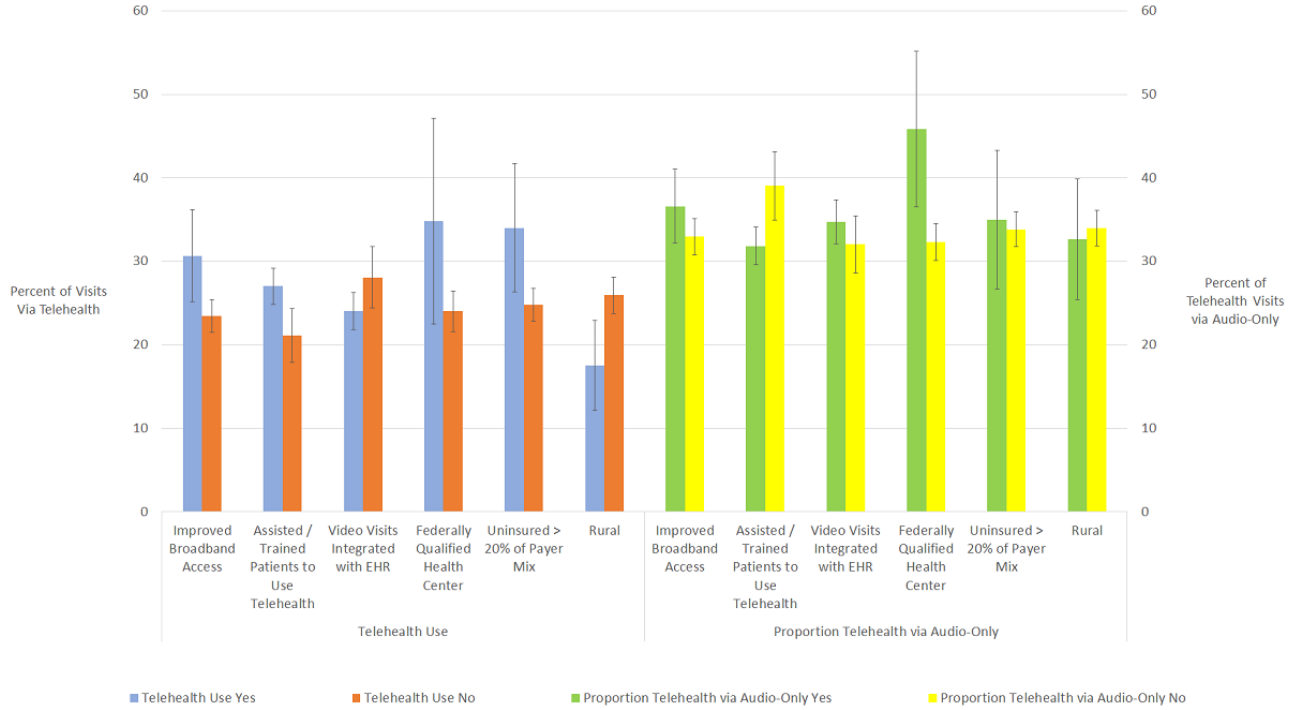

Discussion

In this national study of primary care practices in 2022, we found that respondent practices serving the uninsured reported greater use of telehealth for patient encounters, and FQHCs used more audio-only telehealth. This suggests that low socioeconomic status populations had a higher need for telehealth and that cuts to audio-only reimbursement would disproportionately impact care for such patients. The "digital divide," systematic barriers for accessing and using technology and telehealth among various populations [3], is a known issue for FQHC and rural populations [4,5], and our study provides national evidence that these practice-level factors are significantly associated with lower adoption of video-based telehealth.

Importantly, we found that the telehealth-enabling practices of training and assisting patients with using telehealth and broadband expansion were associated with higher telehealth adoption. In contrast, electronic health record integration for video visits, participation in alternative payment models that incentivize care quality, and practice ownership were not significant practice-level factors, contrasting with a prior study suggesting health system integration was linked to higher telehealth use at the physician level [6].

Limitations include potential nonresponse bias due to the modest response rate, though we used weights to account for this (sensitivity analyses of missing data are in Multimedia Appendix 1); that telehealth use is reported in aggregate rather than calculated from visit data, so could not be verified; and an inability to draw causal inference from cross-sectional analysis. Longitudinal analyses or controlled trials would provide stronger

XSL•FO
RenderX

evidence and be able to describe any changes in significance of the identified factors for telehealth use over time.

These findings provide important national data for the design of policy and practice interventions to expand telehealth use. Practices focused on enabling telehealth appear able to meaningfully increase its uptake [7]. Federally, renewing the lapsed support for broadband accessibility is an important means to address the digital divide [8,9].

Acknowledgments

The team would like to think Jonathan Skinner, PhD, for his review of a draft of this research letter, and Ching-Wen W Yang, MSPH, for her contributions to the construction of the dataset.

Research reported in this publication was funded by the Robert Wood Johnson Foundation (78479, 80738, and 81412) and supported by the National Institute on Aging of the National Institutes of Health (R01AG084611). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Robert Wood Johnson Foundation or the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the need for curating deidentified data, but are available on reasonable request from KES (karen.e.schifferdecker@dartmouth.edu).

Disclaimer

The statements, findings, conclusions, views, and opinions contained and expressed in this article are based in part on data obtained under license from IQVIA information services (OneKey subscription information services 2017-2022, IQVIA Inc). The statements, findings, conclusions, views, and opinions contained and expressed herein are not necessarily those of IQVIA Inc or any of its affiliated or subsidiary entities.

Authors' Contributions

Conceptualization: MM, EF, HPR, SS, KES, ERA
Data curation: ROS, AB
Formal analysis: MM
Funding acquisition: MM, EF, KES
Investigation: MM, ROS
Methodology: MM, KES, AJO, ROS
Project administration: KS, EF, ROS, AB
Software: MM
Resources: KES, EF, SS, HPR
Supervision: EF, KES, SS, HPR
Validation: MM
Visualization: MM, EF, HPR
Writing – original draft: MM, EF
Writing – review and editing: all

Conflicts of Interest

None declared.

Multimedia Appendix 1
Further information on survey administration, analytic datasets, weights, and missing data.
[DOCX File , 26 KB - jmir_v27i1e70404_app1.docx]

References

1. National Academies of Sciences, Engineering, and Medicine. Implementing High-Quality Primary Care: Rebuilding the Foundation of Health Care. Washington, DC: The National Academies Press; 2021.
2. Colburn D. The impact of telehealth expansion on health care utilization, access, and outcomes during the pandemic: a systematic review. Telemed J E Health 2024 May;30(5):1401-1410. [doi: 10.1089/tmj.2023.0440] [Medline: 38100326]
3. Lythreatis S, Singh SK, El-Kassar A. The digital divide: A review and future research agenda. Technol Forecast Soc Change 2022 Feb;175:121359. [doi: 10.1016/j.techfore.2021.121359]

XSL•FO
RenderX

4. Uscher-Pines L, Sousa J, McCullough C, Dong S, Kapinos K. Telehealth visits in health centers serving low-income patients in California: final results from the Connected Care Accelerator Initiative (2022-2024). RAND Corporation. 2024. URL: https://www.rand.org/pubs/research_reports/RRA3468-1.html [accessed 2025-03-13]
5. Ko JS, El-Toukhy S, Quintero SM, Wilkerson MJ, Nápoles AM, Stewart AL, et al. Disparities in telehealth access, not willingness to use services, likely explain rural telehealth disparities. J Rural Health 2023 Jun;39(3):617-624 [FREE Full text] [doi: 10.1111/jrh.12759] [Medline: 37042413]
6. Cuellar A, Jena AB. Volume of care for primary care physicians in integrated vs independent practices through the COVID-19 pandemic. JAMA Health Forum 2023 Sep 01;4(9):e232883 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.2883] [Medline: 37656473]
7. Anaya YB, Bañuelos Mota A, Hernandez GD, Osorio A, Hayes-Bautista DE. Post-pandemic telehealth policy for primary care: an equity perspective. J Am Board Fam Med 2022;35(3):588-592 [FREE Full text] [doi: 10.3122/jabfm.2022.03.210509] [Medline: 35641044]
8. Boucher-Robinson S, Varn J. States reckon with lapse of the broadband affordable connectivity program. Pew Charitable Trusts. URL: https://pew.org/3MTXLw9 [accessed 2024-12-20]
9. Frieden R. Best Practices in Promoting Widespread and Affordable Broadband Service After the Covid-19 Pandemic. Social Science Research Network. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4628562 [accessed 2025-03-13]

Abbreviations

APRN: advanced registered practice nurse
FQHC: federally qualified health center
PA: physician assistant

Edited by N Cahill; submitted 20.12.24; peer-reviewed by D Colburn, F Opia, S Ajayi; comments to author 27.01.25; revised version received 05.03.25; accepted 06.03.25; published 28.03.25.

Please cite as:
Mackwood M, Fisher E, Schmidt RO, O'Malley AJ, Rodriguez HP, Shortell S, Akré ER, Berube A, Schifferdecker KE
Primary Care Practice Factors Associated With Telehealth Adoption in the United States: Cross-Sectional Survey Analysis
J Med Internet Res 2025;27:e70404
URL: https://www.jmir.org/2025/1/e70404
doi:10.2196/70404
PMID:

©Matthew Mackwood, Elliott Fisher, Rachel O Schmidt, A James O'Malley, Hector P Rodriguez, Stephen Shortell, Ellesse-Roselee Akré, Alena Berube, Karen E Schifferdecker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining Intersectionality and Barriers to the Uptake of Video Consultations Among Older Adults From Disadvantaged Backgrounds With Limited English Proficiency: Qualitative Narrative Interview Study

Laiba Husain[1], DPhil, MPH, BSc; Trisha Greenhalgh[1], PhD

Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Laiba Husain, DPhil, MPH, BSc
Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865289300
Email: laiba.husain@phc.ox.ac.uk

Abstract

Background: The rapid shift to video consultation services during the COVID-19 pandemic has raised concerns about exacerbating existing health inequities, particularly for disadvantaged populations. Intersectionality theory provides a valuable framework for understanding how multiple dimensions of disadvantage interact to shape health experiences and outcomes.

Objective: This study aims to explore how multiple dimensions of disadvantage—specifically older age, limited English proficiency, and low socioeconomic status—intersect to shape experiences with digital health services, focusing on video consultations.

Methods: Following familiarization visits and interviews with service providers, 17 older people with multiple markers of disadvantage (older age, low socioeconomic status, and limited English proficiency) were recruited in the Redbridge borough of London. Data collection included narrative interviews and ethnographic observations during home visits. Field notes captured participants' living conditions, family dynamics, and technological arrangements. Guided by intersectionality theory and digital capital concepts, interviews explored participants' experiences accessing health care remotely. Intersectional narrative analysis was used to identify key themes and examine how different forms of disadvantage interact. We developed theoretically informed narrative portraits and user personas to synthesize findings.

Results: Analysis revealed that the digitalization of health care can exacerbate existing inequities, erode trust, compound oppression, and reduce patient agency for multiply disadvantaged patient populations. Examining intersectionality illuminated how age, language proficiency, and socioeconomic status interact to create unique barriers and experiences. Key themes included the following: weakened presence in digital interactions, erosion of therapeutic relationships, shift from relational to distributed continuity, increased complexity leading to disorientation, engagement shaped by previous experiences of discrimination, and reduced patient agency.

Conclusions: This study provides critical insights into how the digitalization of health care can deepen disparities for older patients with low income and limited English proficiency. By applying intersectionality theory to digital health disparities, our findings underscore the need for multifaceted approaches to digital health equity that address the complex interplay of disadvantage. Recommendations include co-designing inclusive digital services, strengthening relational continuity, and developing targeted support to preserve agency and trust for marginalized groups in an increasingly digital health care landscape.

(J Med Internet Res 2025;27:e65690) doi:10.2196/65690

KEYWORDS

digital health disparities; video consultations; intersectionality; health inequity; digital capital; mobile phone

XSL•FO
RenderX

Introduction

Overview

While existing research has explored digital health disparities [1-5] and the application of intersectionality theory in health care [6-8] separately, there remains a critical gap in understanding how multiple dimensions of disadvantage intersect specifically in the context of digital health services. Previous studies have largely focused on single-axis analyses of digital exclusion [9-11]. For example, Ramsetty and Adams [9] examined socioeconomic status as a determinant of digital health access through quantitative analysis of technology ownership and internet connectivity rates. Eberly et al [10] used a large-scale quantitative study to analyze telemedicine access solely through the lens of age, finding lower video visit completion rates among older adults. Similarly, Donaghy et al [11] investigated the acceptance of video consultations primarily through the single dimension of technological literacy using survey data. While these quantitative studies provide valuable insights into individual factors, they have not adequately examined the compounded effects of age, language proficiency, and socioeconomic status on experiences with digital health technologies, particularly video consultations. Furthermore, there is a lack of in-depth, qualitative research that captures the lived experiences of multiply disadvantaged patients navigating these digital health spaces [12]. This study aimed to address this gap by applying an intersectional lens to explore how various forms of disadvantage interact and manifest in the context of video consultations, providing a more nuanced and comprehensive understanding of digital health disparities among susceptible populations.

We adopted an intersectional perspective from feminist studies to highlight the intersection and entanglement between digital technology, structural stratifications, and ingrained tendencies of "othering" in societies. This approach allowed us to move beyond simplistic notions of digital divisions to examine how digital technology is implicated in complex and intersectional systems of power. Drawing on narrative interviews with older, low-income individuals with limited English proficiency, we examined intersectionality and how video consultations can exacerbate existing inequities for multiply disadvantaged patient populations.

Our analysis revealed that digital health disparities operate at the intersection of multiple fracture lines of difference that mediate various spaces of inclusion and exclusion. We argue that addressing digital health disparities requires moving beyond single-axis analyses to consider how different aspects of disadvantage intersect in individual lives. This paper contributes to information systems (ISs) literature by providing a richer theorization of digital inequity, highlighting the need for intersectional approaches to digital health equity. We propose a research agenda that calls for IS scholars to reconceptualize actors beyond simplistic notions of "users," to consider positioning rather than contextualizing, and to examine how digital health technologies are intertwined with producing and reproducing social orders and stratifications. Our findings have important implications for policy and practice in designing and implementing more inclusive and equitable digital health services.

Background and Literature Review

Overview

In the United Kingdom, video consultation uptake has increased significantly since the COVID-19 pandemic. National Health Service (NHS) England reported that before the pandemic, <1% of primary care appointments used video, but this number rose sharply during early 2020 to approximately 13% as part of the NHS Long Term Plan to expand digital-first care [13]. By 2023, about 9% of consultations in general practice continued to use video, reflecting an enduring shift toward remote health care delivery despite a return to face-to-face appointments as the dominant mode (65%) [14,15]. However, the use of video consultations varies significantly across demographic groups. Research indicates that older patients (aged >65 years) are less likely to use video services, with adoption rates around 7%, compared to 20% among younger adults aged between 18 and 34 years. Similarly, individuals from lower socioeconomic backgrounds and those with limited English proficiency face notable barriers, demonstrating uptake rates <10% [14,16].

Digital Health Inequalities and the Digital Divide

The IS literature has long engaged with the concept of the digital divide, traditionally focusing on accessibility, literacy, and adoption of digital technologies [17]. This discourse often conceptualizes individuals as "users" of technology, assigning them to specific group categories such as "the excluded" or based on binary divisions of "haves" and "have-nots" [18]. However, this notion of the digital divide fails to account for the multifaceted and compounded nature of digital inequality [12,19]. Recent scholarship has recognized that digital exclusion is a complex and dynamic phenomenon influenced by various factors beyond access to technology, including age, gender, education, and socioeconomic status [20-25]. As health care becomes increasingly digitalized, there is a growing need to understand how these broader digital inequalities manifest in health contexts. The concept of digital health equity has emerged as a critical area of study, examining how social determinants of health intersect with digital access and skills to shape health outcomes. This evolving field calls for more nuanced, intersectional approaches to understanding and addressing disparities in digital health access and use.

Intersectionality Theory

To address the limitations of single-axis analyses of digital exclusion, we turn to the concept of intersectionality from feminist studies. Originally proposed by Crenshaw [26] to expose the marginalization of Black women under both sexism and racism, intersectionality stands against the tendency in critical social theorizing to treat individuals in independent categories. It emphasizes that systems of oppression are inherently bound together, creating singular social experiences for people who bear the force of multiple systems [27]. While IS research has explored the relationship between IT and identity [28,29], most studies focus on the individual or group level, investigating how IT mediates or shapes identities. However, an intersectional perspective views subjectivity as emerging

XSL•FO
RenderX

from differential experiences produced by multiple and intersecting power structures [26]. This approach moves beyond a behavioral, individualistic sense of identity to one of the "social positioning" of individuals within social structures [30]. Applying intersectionality to digital health inequalities allows us to examine how various dimensions of disadvantage—such as age, language proficiency, and socioeconomic status—interact and compound to shape individuals' experiences with digital health services.

Digital Capital Theory

The concept of digital capital by Ragnedda and Ruiu [31], an extension of the cultural capital theory by Bourdieu [32], offers another valuable lens for examining digital health disparities. Bourdieu [32] applied the idea of capital to signify the internal (eg, abilities and attitudes) and external (possessions and attributes) resources that people mobilize to achieve their goals in social life. He highlighted cultural capital as a form of capital that can be accumulated and transformed into other forms of capitals. Digital capital is made up of both digital competencies and digital technologies, which Ragnedda and Ruiu [31] argue is a form of capital in its own right and is essential for building up social, economic, and cultural resources in the digital world we live in today. Disparities involving digital skills originate in inequalities of access but are mediated by orientations that can only be understood in relation to total life contexts (eg, education, income bracket, age, location, and social support all influence a person's access to digital technologies and the level of digital skills they can acquire) [33]. Digital capital is a relatively new concept that scholars have begun to explore empirically through various methodological approaches [34]. Digital capital may be estimated, for example, at the individual level by assessing a person's digital literacy and skills, at the organizational level by measures of digital infrastructure (including the digital competence of personnel), and at the locality level in terms of the quality of the area's IT infrastructure.

Digital capital theory points us to the hypothesis that traditional forms of capital (such as economic, cultural, and social) are converted into digital capital and vice versa and provides the conceptual tools to examine how and to what extent this occurs, thereby illuminating how social inequality relates to digital inequality. If digital spaces—due to social inequality and underlying power structures—become increasingly stratified, there will be significant impacts on how individuals from differing backgrounds gain accumulated forms of capital through the digital realm. In other words, digital capital theory offers an explanation as to why people who already face systemic inequities find that these disparities widen when services are digitalized.

Recent studies have begun to explore how digital capital interacts with other forms of capital to influence health outcomes and access to digital health services [35]. However, there remains a need for more in-depth, theoretically informed research on how digital capital intersects with other dimensions of disadvantage in shaping experiences with digital health technologies.

In the context of this study, intersectionality works as an overall guiding principle for understanding how people's lives and characteristics stem from and lead to multiple axes of disadvantage, while digital capital theory helps us understand how these axes of disadvantage play out in terms of access to and use of digital resources.

Research Gap and Objectives

Our previous narrative review [12] highlighted that while existing literature recognizes the multifaceted nature of digital inequality, there is a critical lack of in-depth, theoretically informed studies examining how different dimensions of disadvantage combine to affect digital health disparities. The review found the available literature on digital health disparities, particularly in relation to video consultations, to be sparse and primarily descriptive rather than explanatory. Most research has focused on identifying barriers and enablers without adequately exploring the complex interplay of factors contributing to these disparities.

Importantly, our review revealed no theoretically informed studies that examined how different dimensions of disadvantage combined to affect digital health disparities. This gap in the literature limits our understanding of how multiple disadvantages intersect and compound to shape individuals' experiences with digital health services.

Building on these findings and responding to the recommendations of our narrative review, this study aims to address these critical gaps. By doing so, we seek to move beyond the descriptive accounts that have dominated the field and provide a richer, more nuanced theorization of digital health disparities. This approach allows us to explore the complex ways in which different aspects of disadvantage interact, compound, and manifest in the context of digital health services. Our goal is to contribute to the development of more inclusive and equitable digital health services by offering insights into the lived experiences of multiply disadvantaged individuals, as called for in our previous work.

This study represents a direct response to the research agenda proposed in our narrative review, aiming to deepen our understanding of digital health disparities and inform more effective, equitable strategies for digital health implementation.

Methods

Overview

This study used a qualitative, interpretive approach to explore the intersecting effects of age, socioeconomic status, and limited English proficiency on experiences with digital health services. The focus on age, language proficiency, and socioeconomic status was informed by our aforementioned narrative review [12], indicating these as key predictors of lower digital service uptake and barriers to access. We adopted narrative inquiry [36] as our primary methodology, which aligns with our aim to center the voices and experiences of marginalized patients. The study was conducted in Redbridge, one of London's most diverse boroughs. It has >65% of its residents from Black and minority ethnic communities, predominantly Asian (42%). Over 90 different languages are spoken in the borough, with nearly a

XSL•FO
RenderX

quarter of residents having a first language other than English. The borough includes several areas ranked among the 20% most deprived in England, with an unemployment rate (8%) exceeding London's average. This demographic and socioeconomic profile made Redbridge an ideal setting to examine intersecting dimensions of disadvantage in digital health care access. Participants were recruited through the Redbridge Respiratory Service within the North East London NHS Foundation Trust, community organizations, and snowball sampling.

Inclusion criteria for participants were as follows: (1) aged ≥65 years, (2) limited English proficiency (self-reported or identified by health care providers), (3) living within an Index of Multiple Deprivation decile of 1 to 5, (4) having attempted ≥1 video consultation, and (5) residing within the Redbridge borough.

As this was a doctoral research project, the primary investigator, LH, conducted all interviews and primary analysis, with regular supervision and analytical discussions with her supervisory team to challenge interpretations and biases throughout the research process. Semistructured narrative interviews were conducted with 17 participants between July 2022 and January 2023. All (17/17 100%) study participants had access to health care through the NHS despite being multiply disadvantaged. The interview guide was designed to elicit rich narratives about participants' experiences with digital health services, particularly video consultations. Questions explored their overall health care journey; experiences with digital technologies; and perceptions of how their age, language abilities, and financial situation affected their access to and use of digital health services. As a multilingual researcher fluent in Hindi, Urdu, and English, LH conducted all interviews directly in participants' preferred languages. Participants could freely switch between languages during interviews, which many did. This linguistic flexibility allowed participants to express themselves more fully and comfortably when discussing their experiences. All non-English segments were translated to English during transcription by LH, while "broken English" was transcribed verbatim when the meaning was clear. In addition to interviews, LH conducted ethnographic observations with 9 (53%) of the 17 participants, including home visits and participation in daily routines. These observations provided valuable contextual insights into participants' living conditions, family environments, and technological exposure. The ethnographic data were analyzed alongside interview transcripts to provide richer context to participants' narratives and to understand how their home environment and daily routines influenced their experiences with digital health care.

LH used thematic narrative analysis [37] using an intersectionality lens to identify key themes while preserving the integrity of individual stories. The analysis process involved the following: (1) familiarization with the data through repeated reading of transcripts, (2) open coding to identify initial themes and patterns, (3) development of a coding framework informed by intersectionality theory, (4) axial coding to explore relationships between themes, and (5) selective coding to refine and integrate themes into a coherent narrative.

To enhance the intersectional analysis, a modified version of the Equity Design Collaborative's meta-empathy mapping approach was incorporated [38]. This approach focuses on understanding users' needs through transformative empathy, moving beyond surface-level observations to deeply understand how systemic barriers and power structures affect user experiences. This modified approach shaped the analysis in 3 key ways.

First, it guided us to examine both immediate barriers (ie, language difficulties) and deeper structural challenges (ie, how health care systems may inadvertently privilege certain cultural norms). Second, it helped identify power dynamics in health care relationships. Third, it informed how outputs such as user personas (discussed in subsequent sections) were structured to capture both individual circumstances and systemic influences on participants' experiences. This methodological lens helped ensure findings reflected not only individual experiences but also broader systemic factors affecting remote care access.

Several strategies were used to ensure the trustworthiness of findings [39], including prolonged engagement with participants through multiple interactions, triangulation of data sources, member checking with participants to verify our interpretations, and maintaining a reflexive journal to document the decision-making process and potential biases.

Particular attention was paid to ensuring that participants fully understood the nature of the research and their rights, given potential language barriers and vulnerabilities. As researchers, we also acknowledge our own positionalities and how they may influence the research process. LH is a female Muslim of South Asian descent, which facilitated trust building with many participants but also required ongoing reflexivity to avoid assumptions based on shared cultural backgrounds. To synthesize findings, LH developed 4 [40] theoretically informed narrative portraits and user personas that captured key intersecting dimensions of disadvantage, including how age, language barriers, socioeconomic status, cultural factors, and health conditions combined to shape participants' experiences with digital health care. These are reported in detail in our previous paper [40] and serve as complementary outputs that distill key themes into accessible archetypal stories, balancing the need to honor individual perspectives with extracting cross-cutting insights about the interplay of technology and disadvantage.

Ethical Considerations

The study received ethics approval from the NHS Research Ethics Committee and the University of Oxford's Central University Research Ethics Committee. Informed consent was obtained from all participants, and pseudonyms were used to protect their identities. Ethical approval was obtained from East Midlands—Leicester South Research Ethics Committee and UK Health Research Authority (September 2021, reference number: 21/EM/0170; Integrated Research Application System project ID 300719).

Results

Overview

The analysis revealed 8 key themes that illuminate how multiple dimensions of disadvantage intersect to shape experiences with

XSL•FO
RenderX

digital health services, particularly video consultations. These themes highlight the complex ways in which the digitalization of health care can exacerbate existing inequities, erode trust, compound oppression, and reduce patient agency for multiply disadvantaged patient populations.

Digital Interactions May Have a Weak Presence

The concept of "absent presence," originally developed by Gergen [41] to describe technology-induced distraction in face-to-face interactions, takes on new significance in the context of digital health care. The analysis found that participants consistently reported a sense of diminished presence in digital health care interactions, particularly in video consultations. This weakness of presence manifested in 3 key ways: delayed responses, mechanical intonation, and perceived motionlessness.

Delayed responses were frequently noted by participants as a sign of disengagement. Maneshi, an Indian immigrant aged 83 years, articulated this experience as follows:

> *If my doctor is doing twenty other things on his side of the screen while I'm talking, even if he is technically listening, it just doesn't feel like a genuine conversation to me...When I go to the GP [general practitioner] in person, those distractions aren't there. I know the GP is looking at me and having a conversation with me.*

This account illustrates how the perceived divided attention of health care providers during video consultations can erode the sense of a genuine, engaging interaction. The lack of immediate responsiveness disrupts the natural rhythm of conversation, leading to a feeling of disconnection.

Mechanical intonation was another aspect that contributed to the sense of weak presence. Fowzia, a recently widowed immigrant from Pakistan, expressed this concern as follows:

> *The absence of eye contact and non-verbal cues makes you feel...I don't know it's off putting. It feels like I'm talking to a computer program instead of a real doctor. They used to, the doctors, they would bring a lot of comfort and assurance; now, it's replaced with a sense of disinterest. It just makes you think you know? Do they even care?*

Participants described feeling as though they were interacting with a "computer program" rather than a real person, noting the flat, unemotional tone often used by health care providers during video consultations. Recent research in general practice has shown that video-mediated communication affects sensory conditions in clinical interactions, technical issues such as delayed facial expressions and sound can disrupt natural conversation flow, and the screen interface limits health care providers' ability to read and respond to patient cues [42]. This lack of vocal vitality, as conceptualized by Stern [43], can significantly impact the patient's perception of the provider's engagement and empathy.

The third manifestation of weak presence was perceived motionlessness. Abed, a recently retired repairman aged 69 years, highlighted this issue as follows:

> *His movements don't signal any interest. I mean I guess that could just be because I'm seeing his face and can't tell over video call how much he's actually moving. But I'm not seeing hand movements or his head moving much you know.*

In face-to-face interactions, body language and subtle movements play a crucial role in conveying attention and engagement. However, in video consultations, the limited visual field and potential technical constraints can result in an appearance of stillness that patients interpret as a lack of involvement or interest. Basmah, a Bangladeshi immigrant aged 73 years with arthritis and limited mobility, noted the following:

> *Back in the day, a doctor's touch and a comforting pat on the back meant so much. Now, it's different, you just hear typing with those hands.*

This triad of delayed responses, mechanical intonation, and perceived motionlessness collectively contributed to a sense of "absent presence" in digital health care interactions. Therefore, many (11/17, 65%) participants reported feeling disconnected and unheard during video consultations, potentially impacting the quality of care and patient satisfaction. Mohammad, a Sri Lankan immigrant aged 70 years with multiple chronic conditions and living with his extended family, starkly put the following:

> *Facial expression and body language is so important. When you talk to people you can see what's wrong and things like that and some of the things you don't know whether when the GP say things, how they mean it because I can't tell over remote. I don't think the whole system is fracturing, I think it's completely collapsing.*

While all patients may experience reduced presence in video consultations, this sense of disconnection is particularly problematic for patients with limited English proficiency, who rely heavily on nonverbal cues and physical presence to support communication and understanding. The technical barriers to presence can additionally compound existing language barriers, potentially leading to misunderstandings or missed clinical information.

Digital Encounters May Weaken Relationships

Analysis revealed that digital encounters, particularly when there was no previous in-person relationship, often weakened the therapeutic relationship between patients and health care providers. Participants consistently expressed difficulty in forming bonds through screens, highlighting the importance of relational foundations for ethical care.

Rajpreet, a first-generation Indian woman experiencing economic hardship and multiple chronic conditions, articulated this challenge as follows:

> *Doing this by video makes it harder. If I met her in person, maybe we could connect more...but I know women like her, that's not to say in a bad way, just that she doesn't really struggle with the same things that I do, it's very different when two women come*

XSL•FO
RenderX

> *from...I don't know how to say...just that...she's of a different social class...if that makes sense.*

This quote illustrates how the lack of physical presence can exacerbate perceived social and cultural differences, making it more difficult to establish a connection and mutual understanding.

The importance of preexisting relationships was emphasized by many (14/17, 82%) participants. Fatima, an Afghan woman aged 81 years with diabetes and heart conditions, who relies on her children for translation, explained the following:

> *I think it [video consultation] was easier because I had met Dr. Samari before. So I already had that initial relationship with her. I've seen her a few times. I felt comfortable with her. I think it would have been a bit more awkward if it had been like a first meeting.*

Conversely, participants who had video consultations with unfamiliar clinicians often reported fewer positive experiences. Ramnik, an Indian immigrant aged 78 years with hypertension and diabetes, who lives alone and relies on community support services, explained the following:

> *You stay in the online waiting room being all confused and then they let you into the call. And its someone you've never seen before. And he just wants yes or no. Then it's finished. You can go. That's it. You know? So I don't know. I don't like it. I don't. This is why I don't want to see the GP unless it's really bad. When I did it [video consultation] with my own specialist it wasn't like this. That time was good because he knew me and I knew him.*

This account highlights how the lack of a prior relationship can lead to a sense of disconnection and dissatisfaction with the consultation process.

Some (5/17, 29%) participants stressed the need for occasional in-person visits to establish and maintain a connection:

> *It'd be nice if they could see a person like myself every three months or normal patients at least once every six months. Yeah, so you know, so then you know them. So that builds up some, like, friendship as well. Now there's no friendship. But that's why, you know, it's important we get to know who the person is. Here there's no chance to do that.*

This suggestion underscores the perceived value of face-to-face interactions in building and sustaining therapeutic relationships.

The findings indicate that even when participants had successive video consultations with the same health care provider, they rarely developed a sense of building a strong and positive therapeutic relationship. This contrasts with face-to-face environments, where the patient-provider relationship typically strengthens with each encounter. Some (4/17, 23%) participants even described a deterioration in their relationship over repeated remote encounters, unless preceded by face-to-face meetings.

The weakening of therapeutic relationships through video consultations has consequences for marginalized older adults who often rely on trusted health care provider relationships to maintain engagement with health services. Limited English speakers face additional barriers to relationship building in digital settings, where language difficulties are compounded by technical constraints.

The Shift From Relational to Distributed Continuity

The analysis also revealed a significant shift from relational to distributed continuity as digitalization increased after the COVID-19 pandemic. This transition often left patients feeling lost, unsupported, and struggling to navigate their care effectively. The loss of relational continuity was particularly pronounced for marginalized older adults who had previously relied on long-standing relationships with health care providers.

Tasneem, a Bengali immigrant aged 75 years, expressed this loss as follows:

> *Dr. Talib has seen me through so much. I could speak to her about anything, and she really listened and understood me.... [Now] I don't even know my doctor's name. How can I trust someone I don't even know?*

This quote encapsulates the profound impact of losing a trusted health care relationship and the challenge of building trust in a system of distributed care.

The fragmentation of care was a recurring theme among participants. Arjun, a retiree aged 72 years, described his frustration with the depersonalization of care as follows:

> *It's becoming less and less personal. It's like you are not a person there you are just a face. For example, the company I work for we went through the same sort of process, I joined in '87 with BT and then in 1993 they brought this employee individual identification numbers they're called right, so you're given a nine digit number and then after that whenever you wanted to talk to the HR department or pay group or something like that, that's the thing that you gave them and that's it. So you only become a number, in this case you're only becoming a face to the GP, always a new number, new face, new GP, it's not who you are or what you are.*

This shift to distributed continuity often resulted in communication breakdowns and potential risks to patient care. Priya, a Punjabi woman aged 71 years with diabetes, expressed her concerns as follows:

> *Ordering repeat medicine with the GP always main problem. They are, you know, taking so long and even sometimes they don't know your condition there because communication, everything is so broken. They just forget everything. Like you are the new person to them. You know that's the problem with the GP always.*
>
> *They don't know my history or me. I feel scared they'll make a mistake with the dosage but what can I do?*

The loss of community connections was another significant aspect of this shift. Another participant highlighted how digital triaging erased the familiarity and efficiency of previous care arrangements:

XSL•FO
RenderX

> *[B]efore, you know, they knew everything already. They knew you. And they had my history. Whereas. Where I am now, as I said, the first five minutes of any session are taken up with me explaining who I am, what my situation is. And then, you know, explaining what the problem is, why I'm seeing them all the rest. And this is repeating after filling out a whole econsult first. It's a right shame.*

The fragmentation of care extended beyond primary care to specialist services. Najma, a Pakistani immigrant aged 66 years with limited English proficiency and respiratory conditions, living in council housing, described her frustration with revolving specialists and poor communication:

> *Now my specialist has also now changed. Three of them I went through. At the moment one, I think one is they know she's the main one. She was good, but now I don't know which one. Last two weeks ago they giving me one medicine which is a high dose they asked me to stop and they my daughter she sent a lot of e-mail to her but she don't reply anything here so this one [the new specialist] I think she's not so good. Communication is not good. You keep chasing them, you know?*

These experiences highlighted how the shift to distributed continuity has disrupted the holistic, coordinated care that many patients, especially those from marginalized communities, relied upon. The fragmentation of care across digital platforms has created new challenges in maintaining consistent, personalized health care relationships and effective communication between health care providers and patients.

Digital Interactions May Compound Oppression

Findings revealed that digital health services often compounded existing forms of oppression and discrimination, particularly for participants with limited English proficiency and low digital literacy. These individuals faced multiple, intersecting barriers in accessing and navigating digital health platforms, which exacerbated their existing challenges in health care settings.

Rajpreet, a first-generation Indian woman aged 74 years living in council housing with multiple chronic conditions, articulated this challenge:

> *I told her, don't mind me saying this Dr. Kaur, but you don't really know what it's like, you come from a different background or world whatever you want to call it [laughs], you know medically sure, but over video you can't grasp it, you won't understand because you don't have it either, my condition, what surrounds it, that sort of thing...*

This quote illustrates how digital interactions can amplify cultural and experiential gaps between patients and health care providers, making it more challenging to discuss sensitive health issues.

Language barriers were particularly problematic in digital settings. Samiyah, a Pakistani grandmother aged 69 years, with limited English proficiency and chronic pain conditions who recently moved in with her son's family, noted the following:

> *It's already hard to explain my symptoms in English, but over video, it's even worse. I can't use gestures or show them exactly where it hurts. Sometimes I feel like they don't understand me at all.*

Low digital literacy compounded these challenges. Tasneem shared the following:

> *I struggle with technology, and now I have to figure out how to use these apps just to see my doctor. It makes me feel stupid and left behind. Sometimes I just give up.*

The shift to video consultations also highlighted existing inequalities in access to technology. Mohan, living on a basic pension in social housing with diabetes and heart disease, explained the following:

> *They tell us to do video calls, but I don't have a smartphone or good internet. It's like they're saying healthcare is only for people who can afford fancy gadgets.*

Participants from marginalized groups often found themselves at the intersection of multiple disadvantages—language barriers, cultural differences, low digital literacy, and limited access to technology—all of which were amplified with video consultations.

Digitalization May Erode Trust in Health Providers and Systems

Findings revealed that the shift to video consultations also frequently eroded participants' trust in health care providers and systems. This erosion of trust was often rooted in a sense of depersonalization and lack of continuity in care, particularly for marginalized and susceptible populations. Research has also documented how marginalized communities may have lower trust in health care providers due to several factors, including historical medical mistreatment, documented disparities in quality of care, and systemic barriers to culturally and linguistically appropriate services [44,45]. In this study, this broader context of health care mistrust appeared to be exacerbated by the shift to video consultations, particularly when patients could not establish consistent relationships with health care providers.

Hasan, with his multiple chronic conditions and previous negative health care experiences, articulated this sentiment strongly in the following manner:

> *You don't wanna be a part of this system. Like you can't trust these people in the NHS... Because I need the GP, but to be honest, I don't trust that like you know, she's not very good and she she doesn't really care, right? That's the impression I get so I try not to go but then it gets worse.*

This quote illustrates how the perceived lack of care and attention in digital interactions can lead to a cycle of disengagement and worsening health outcomes.

The impersonal nature of video consultations was a recurring theme. As Fowzia explained the following:

XSL•FO
RenderX

> *The personal connection is not the same. I miss the stuff that didn't require words, stuff you could just see and feel and the comfort of being physically there in the same room. It's harder to build that trust through a screen.*

This highlighted the importance of nonverbal cues and physical presence in building trust, which many felt was lacking in digital interactions.

The erosion of long-standing relationships with health care providers was particularly distressing for some participants. Fatima shared the following:

> *Thing is you can't even think about trusting the GP now even if you wanted to this way [over video consults]. These things take a lot of time, beta, my old GP, I knew him for 10 years, every small flu, back ache, little cold he knew it all. I don't even know if I have another 5 years left in me and if I keep seeing a different GP over a different platform where does that leave me, beta?*

This account underscores how the fragmentation of care across multiple health care providers and platforms can disrupt the accumulation of shared knowledge and understanding that forms the basis of trust in health care relationships.

The erosion of trust sometimes led to nonadherence to medical advice. Zainab, a Pakistani immigrant aged 73 years with heart disease, who had previously experienced dismissive treatment from health care providers, explained the following:

> *I mean why should I take it [the medication], ok yeah they prescribe it but they don't even bother listening to me, click clacketing away at their keyboard, not even looking at the screen, they think they know what's wrong with me just like that? I don't trust it one bit course I'm not gonna take it.*

For some, the distrust in the health care system led to anxiety and avoidance. Hamida, a Bangladeshi mother aged 75 years living with her daughter's family and experiencing both physical health conditions and growing anxiety about health care interactions, shared the following:

> *My daughter thinks I have anxiety because I don't want to deal with the NHS. I don't know, maybe I do. But to me it's more about the fact that they can't be trusted.*

However, it is important to note that not all participants experienced an erosion of trust. In a notable disconfirming case, 1 participant reported high levels of trust in video consultations due to a preexisting, long-term relationship with her general practitioner. This suggests that strong, preexisting therapeutic relationships may buffer against the potential erosion of trust in digital interactions.

These findings highlight the complex relationship between digitalization and trust in health care. Video consultations can risk eroding the personal connections and continuity of care that many patients, particularly those from marginalized groups, rely on to build trust with their health care providers. This erosion of trust can have serious implications for patient engagement, adherence to treatment, and overall health outcomes.

Digitalization Increases Complexity, Which May Lead to Disorientation

The analysis highlighted that the introduction of multiple digital platforms and access points often led to increased complexity and disorientation for participants. Many (12/17, 70%) struggled to navigate the various digital pathways and processes, leading to frustration and, in some cases, disengagement from health care services.

Mukesh, a man aged 85 years with cognitive challenges, articulated this complexity in the following manner:

> *The health system as a whole there is a lack of communication and sharing between the different functions. I mean I've got access to a long COVID clinic. Then I also deal with my GP and there should be some information exchange between the two. Not me filling them both in on my own. And then with the hospital, they're not connected in a way that they can get anything from my GP to the hospital and it has been very frustrating.*

This quote highlights how the fragmentation of digital systems can place an additional burden on patients, particularly those with complex or multiple health conditions.

The sense of being overwhelmed by digital options was a common theme. Ramnik expressed the following:

> *I felt like I was drowning in all the options, do I call the practice, do I do this econsult thing, do I use the NHS app, do I first check the website, but then the website it too complicated anyways and then I'm back to square one. It's overwhelming, especially when you're already struggling with other things, this is the last thing I should have to worry about.*

This account illustrates how the proliferation of digital access points, while intended to improve accessibility, can paradoxically create barriers for some patients.

The disorientation experienced in digital health care settings was often compounded by language barriers and limited digital literacy. As Samiyah shared the following:

> *I wouldn't know the first thing about doing a video call. He [son] set it up for me, clicked some stuff, and had it up and running and I just sat there. They did all the talking without me, but it probably was for the best anyways because I don't know if I could've even said what I needed to properly, the language issue, the screen issue, just looking at it all was too much for me.*

This quote underscores how digital health care can inadvertently exclude patients who lack the necessary language skills or technological proficiency, potentially exacerbating existing health inequalities.

The complexity of digital systems also led to challenges in maintaining continuity of care. Mohammad noted the following:

XSL•FO
RenderX

> *Look I've got more health issues than I can count on both of my hands, navigating through different services and providers, half remote, half in person, half on the phone, I just feel like I'm lost in a maze. I'm constantly juggling between appointments with different specialists, trying to piece together the whole story for myself and for the providers too because they themselves don't know the full story. It's exhausting and overwhelming.*

This account highlights how the fragmentation of care across various digital and in-person platforms can create a significant cognitive and emotional burden for patients, particularly those managing multiple health conditions.

The disorientation caused by digital complexity was often exacerbated by socioeconomic factors. As Maneshi explained the following:

> *Not having the money to go private adds another layer of hard to the mix. I can't afford the luxury of choosing the most convenient healthcare option. No. Instead, I'm forced to navigate through a patchwork of NHS resources that takes weeks, months, even years and it's just getting worse because they're trying to move things online now and my brain is already just scattered from long-covid first and then the mess that is the NHS trying to be something they aren't so this is just the cherry on top.*

This quote illustrates how the digitalization of health care, when not adequately supported or implemented, can compound existing health inequalities and create additional barriers for those already struggling to access care.

Engagement With Digital Services May be Shaped by Previous Experiences of Racism and Discrimination

Findings revealed that participants' willingness to engage with digital health services was often profoundly influenced by their previous experiences of racism and discrimination within the health care system. These past negative experiences created a foundation of mistrust that often extended to new digital health initiatives.

Hamida, with her growing anxiety, articulated this heightened vigilance as follows:

> *So I'm literally now I'm very vigilant. I will check every single medication. I will read the leaflet 3 times, I will Google it because I don't trust these people they already prescribed the wrong one [medication] to me once before.*

This quote illustrates how past negative experiences can lead to a deep-seated mistrust that influences future interactions with health care services, including digital platforms.

The intersection of racial identity and socioeconomic status in shaping health care experiences was highlighted by Abed in the following manner:

> *Well, yes. At the end of the day, yes, I do think that if I was a white person from a rich background, instead of brown and poor, I might have been treated differently.*

This perception of differential treatment based on race and class extended to video consultations, with participants expressing concern that these biases would persist in digital interactions.

Some (9/17, 53%) participants described a resigned acceptance of interpersonal racism from health care staff:

> *You know, sometimes you can tell they're a little bit... racist, but it's OK, you know, it's not a big deal.*
>
> *And sometimes you get certain doctors who are racist. But it's whatever, I'm used to that.*

These statements reveal a troubling normalization of discriminatory treatment, which may influence patients' expectations and engagement with digital health services.

Language barriers were identified as a particular challenge in video consultations, often intersecting with perceptions of racial discrimination:

> *They make it such that you know she's not given one [appointment for his wife]. If you know what I mean. It's I think it's easier for them to treat people who, for example, who they can see is very much quite different. For example, if you have really great English and you're able to communicate, you're able to get your points across, you won't deal with it [racism] as much as you will for example, if, it's just ok.*

This quote highlights how language proficiency can intersect with racial bias to create additional barriers to accessing health care, including digital services.

Some (6/17, 35%) participants, like Mukesh, expressed concern that video consultations might amplify existing biases:

> *Not too long ago my wife she had a video appointment because she had an issue with the hand. So the doctor like any questions we asked, he was just reluctant to answer and he had the, you know, the sarcasm.... So you just know where some doctors, you don't know whether it's racist or not, but like because we call it that because we feel that way.*

This account suggests that the physical distance in video consultations may exacerbate perceptions of dismissive or discriminatory treatment.

The cumulative effect of these experiences led participants like Tasneem to express extreme distrust in the health care system:

> *This is what I mean. And that's why I can't trust these people with your life. You can't. You can't. You you'd rather die than trust some idiot with your life because you're gonna die anyway.*

This level of distrust poses significant challenges to the adoption and effective use of digital health services among marginalized communities.

Digital Interactions May Reduce Patient Agency

The study revealed that the cumulative effect of the previously discussed factors often resulted in reduced patient agency. Many

XSL•FO
RenderX

(14/17, 82%) participants, particularly those facing multiple disadvantages, felt disempowered and unable to effectively advocate for themselves in digital health care interactions:

> *I'm not very happy with the new GP, but I'm scared to change. The next one might be even worse.*

This quote illustrates how the lack of options and fear of further negative experiences can trap patients in unsatisfactory care arrangements, reducing their ability to seek better alternatives.

The complexity of digital systems often led to confusion and reliance on others, as Najma described the following:

> *I can't tell you how many times I've tried using the website or app and just ended up more confused. Which link was I supposed to click? How do I even describe my symptoms properly in writing? I don't know half these medical terms they're asking. My English is no good. Eventually I just give up and tell my son to figure it out for me instead.*

This account highlights how language barriers and limited digital literacy can significantly undermine patients' ability to independently navigate their care.

The lack of physical presence in video consultations was also cited as a factor reducing patient agency. Mukesh, who experiences complex neurological symptoms, shared the following:

> *I mean doctors have it tough now seeing us all through video calls. But it almost feels pointless for me. He's just staring at notes on some other screen barely listening. I can't really show him what's going on with my body in a genuine way. And to be quite honest it just feels like he's already decided before properly hearing me out.*

This quote underscores how the limitations of video consultations can leave patients feeling unheard and unable to effectively communicate their concerns.

The reliance on family members further eroded the personal agency of participants like Samiyah, who explained the following:

> *At my age, it's impossible to keep track of the different numbers to call or steps to do appointments online. My children handle everything now—they email test results, book consultations, order medications for me. I feel so helpless relying entirely on them, but I don't really have much of a choice.*

This dependence on others for health care management significantly diminishes patients' autonomy and control over their own care.

Some participants, like Mohammad, felt entirely excluded from digital health care due to lack of access to necessary technology:

> *We don't have good internet like that or one of those fancy smartphones they keep saying to use, you see that phone? [points to older Samsung phone with cracked screen] That's what I have and it only works when we can pay for the you know [data]. My nephew always talking about this gadget and that, we just can't afford it. I know they started some phone video service during corona, but it wasn't for me. It feels like they are saying either we have to use this stuff or else we don't deserve to get treatment.*

This quote highlights how socioeconomic factors can create barriers to accessing digital health services, further reducing patient agency.

These 8 themes collectively illustrate the complex and intersecting ways in which the digitalization of health care can exacerbate existing inequities for multiply disadvantaged patients.

Discussion

Principal Findings

Overview

This study provides critical insights into how multiple dimensions of disadvantage intersect to shape experiences with digital health services, particularly video consultations, among older, low-income, individuals with limited English proficiency. The findings revealed a complex interplay of factors that contribute to digital health disparities, extending beyond notions of access and skills to encompass issues of presence, trust, continuity, oppression, complexity, and agency. The analysis revealed that the digitalization of health care can exacerbate existing inequities, erode trust, compound oppression, and reduce patient agency for multiply disadvantaged patient populations. Video consultations often created dynamics of "absent presence," where patients perceived health care providers as distracted or disengaged, leading to a sense of disconnection. The lack of previous in-person rapport negatively shaped patients' perceptions of subsequent video consultations with unfamiliar clinicians, which led to weakened relationships. The shift eroded established continuities of care for some, with the fragmentation of relational continuity becoming apparent. Digital interactions compounded experiences of oppression for patients navigating multiple, intersecting forms of structural disadvantage, such as age, ethnicity, and socioeconomic status. The complexity of navigating multiple digital platforms and pathways led to profound disorientation and fragmentation of care, especially for those with limited digital literacy or language proficiency. Experiences of racism and discrimination within health care settings shaped patients' engagement with digital services, often leading to disengagement and mistrust. The cumulative effect of these factors resulted in a significant reduction in patient agency, particularly for marginalized individuals, undermining their ability to effectively navigate their care and make informed decisions.

The experiences of our participants suggest that digital health technologies, rather than being neutral tools, act as both mirrors and magnifiers of existing social inequalities. This aligns with the concept of the "digital poorhouse" by Eubanks [46], where technology reinforces and exacerbates existing patterns of marginalization. In this study, the shift to video consultations not only reflected existing disparities in health care access but also often amplified them, creating new barriers for those already struggling to navigate the health care system. This

XSL•FO
RenderX

finding challenges the often-optimistic rhetoric surrounding digital health innovations [47,48]. While proponents argue that digital technologies can democratize access to health care [49,50], our results suggest a more nuanced reality. For multiply disadvantaged patients, the digitalization of health care services can create a cascade of exclusionary experiences, from difficulties in accessing technology to challenges in effectively communicating health needs in a digital environment.

Extending Intersectionality Theory in Digital Health Contexts

This research extends intersectionality literature by applying its insights to the study of digital health disparities, revealing specific mechanisms through which the increasing digitization of health care creates new forms of inequity and exclusion. The findings highlight how digital access disparities effectively excluded individuals with low digital literacy from video consultation, while limited English proficiency significantly reduced older adults' engagement with health services in the digital space. Importantly, we found that digital competency and digital access do not always go hand in hand, demonstrating how digitalization, while improving health care access for some, simultaneously creates new vectors of exclusion that intersect with and exacerbate existing social inequities. Moreover, while earlier intersectionality research has focused primarily on traditional axes of oppression such as race, class, and gender, this study highlights the emergence of new vectors of disadvantage related to digital access, literacy, capital, and competency. This underscores the need for an expanded understanding of intersectionality that accounts for the growing centrality of digital technologies in shaping health outcomes and experiences.

This study shows that digital health disparities do not operate along independent axes of division but often overlap, interlink, and interact, demonstrating how patterns of dominance and vulnerability intersect to shape people's experiences with digital health care. We argue that inequity and digital health exclusion are relational and occur along multiple fracture lines, which differentiate people's spaces of opportunities, well-being, and level of agency. These disparities are produced and reinforced through complex social relationships and interactions within health care systems and broader societal structures. The COVID-19 pandemic may have brought out new instantiations and shed light on what was less visible before; however, the roots of digital health inequity are deeply entrenched in systems of power and social order.

The Constellation of Challenges for Multiply Disadvantaged Patients

The findings from this study revealed that video consultations created a constellation of challenges for older patients with multiple disadvantages. Participants described feeling profoundly disoriented in digital spaces, struggling to navigate unfamiliar platforms and processes. This disorientation was compounded by disrupted continuity of care, as they cycled between health care providers in fragmented digital encounters. The remote modality also engendered a weak sense of presence and connection—many (13/17, 76%) felt their health care providers were not fully attentive or did not understand their needs. Crucially, the digital interfaces exacerbated feelings of disempowerment and loss of agency. Participants felt adrift, unable to steer the direction of their care. Intersecting barriers reinforced the following: limited economic, social, and linguistic resources; social and cultural isolation; low digital, health, and health care system literacy; and physical impairments of illness and age. Some (9/17, 53%) perceived the digital challenges as yet another form of discrimination. These intersecting factors fed into a cycle of growing disengagement and mistrust toward individual providers and the health care system as a whole.

Digital Capital and Health Equity

Our results also contribute to the emerging literature on digital capital [35] by illustrating how disparities in digital competencies and access intersect with other forms of disadvantage to shape health outcomes. The struggles of our participants to navigate complex digital health systems reflect not only a lack of technical skills but also a broader deficiency in the social and cultural capital needed to effectively engage with digitalized health care. This finding aligns with the theory of capital conversion by Bourdieu [32], suggesting that disadvantages in 1 domain (eg, socioeconomic status) can translate into disadvantages in another (eg, digital health access).

Trust, Presence, and the Digitalization of Care Relationships

Our findings on the erosion of trust and the sense of "absent presence" in video consultations raise important questions about the nature of care relationships in digital environments. Drawing on the work by Giddens [30] on trust in modern societies, we can interpret these experiences as reflective of the disembedding of social relations that occurs with increased digitalization. The loss of physical copresence in health care interactions appears to disrupt established mechanisms for building and maintaining trust, particularly for patients who may already have reasons to distrust health care institutions.

This erosion of trust and presence challenges dominant narratives about the efficiency and convenience of digital health services. While video consultations may offer logistical benefits, our findings suggest they may come at the cost of the relational aspects of care that are particularly important for susceptible patients. This aligns with the critique of the logic of choice in health care by Mol [51], suggesting that the move toward digital health services may prioritize a transactional model of care over a relational one.

Compounding Oppression and Reduced Agency

Perhaps the most concerning finding is that digital health interactions can compound existing forms of oppression and reduce patient agency. This aligns with critical perspectives on technology that view it not as a neutral tool but as a social force that can reinforce existing power structures [52]. In the context of health care, where power imbalances between health care providers and patients are already pronounced [53], the addition of digital interfaces appears to further tilt the scales against marginalized patients. The reduction in patient agency observed in this study has important implications for patient-centered care and shared decision-making, which are increasingly recognized as crucial elements of high-quality health care [54].

XSL•FO
RenderX

Our findings suggest that current approaches to digital health may be undermining these important principles for certain patient populations.

Methodological Contributions

This research also makes significant methodological contributions to intersectionality literature by demonstrating the value of qualitative and narrative-based approaches to studying the lived experiences of marginalized groups. While quantitative approaches have dominated much of the existing research on digital health disparities, our research draws on feminist and critical race theories that emphasize the importance of storytelling and counternarratives as forms of epistemic resistance [55,56]. By using in-depth narrative interviews and persona development as key methodological tools [40], we provide a more nuanced and contextualized understanding of how digital health disparities are experienced and navigated by multiply disadvantaged individuals.

Implications for Practice and Policy

Viewed through the lens of health equity and social justice, our findings suggest that the rapid digitalization of health care risks exacerbating existing health disparities. This aligns with the concept of "digital redlining" proposed by Gilliard and Culik [57], where digital systems create new forms of discriminatory exclusion. The compounding of oppression and reduction in patient agency experienced by our participants raise serious concerns about the potential for digital health technologies to undermine principles of equity and patient-centered care.

However, these findings also point to potential avenues for intervention. Through reconditioning the intersectional nature of digital health disparities, policy makers and health care providers can develop more flexible, targeted approaches to support susceptible patients. This might involve not only addressing technical barriers to access but also working to build the broader forms of capital needed to effectively navigate digital health systems. For example, offering a range of communication options, including in-person visits, and providing additional support for patients navigating digital systems.

In addition, our results underscore the importance of maintaining and strengthening relational continuity in health care, even as care becomes increasingly digitalized. This might involve strategies to ensure patients can maintain relationships with preferred health care providers across digital and in-person interactions. Third, our findings point to the need for greater attention to issues of structural competency in the design and implementation of digital health systems. This could involve training health care providers in culturally competent digital communication and the development of digital health interfaces that are more inclusive and culturally sensitive. Finally, our results suggest that efforts to address digital health disparities must go beyond simply providing access and skills training to address deeper structural inequalities. This aligns with calls for a "digital determinants of health" framework that recognizes the broader social, economic, and political factors shaping digital health equity [58].

Limitations and Future Research

While this study offers valuable insights, several methodological limitations warrant consideration. The relatively small sample size (N=17) and geographical confinement to Redbridge, London, limit the generalizability of our findings. Our reliance on narrative interviews may have introduced recall and social desirability biases, potentially skewing participants' accounts of their digital health experiences. The inclusion criterion of having attempted ≥1 video consultation may have inadvertently excluded those facing the most severe barriers to digital health access. In addition, our study focused primarily on video consultations, potentially overlooking other forms of digital health interventions.

Looking ahead, several key areas warrant further investigation to address these limitations and expand our understanding of digital health equity. Longitudinal studies are needed to track the long-term impacts of digital health services on marginalized populations, providing insights beyond the snapshot our study offers. Intervention studies should evaluate targeted approaches to address identified barriers, such as building digital capital and fostering trust, which could help overcome some of the access issues noted in our limitations.

To address the geographical limitations of our study, comparative analysis across different health care systems and cultural contexts could identify transferable principles and context-specific challenges in promoting digital health equity. This broader perspective would enhance the generalizability of findings and inform more universally applicable strategies.

From a policy perspective, there is a pressing need to examine how existing health policies and digital strategies impact health disparities. The study analysis could provide valuable context for understanding the systemic factors influencing digital health equity, beyond the individual experiences captured in our study.

Finally, to complement this study's qualitative insights and address the limitations of our small sample size, future work should focus on developing and validating quantitative measures for intersectional digital health disparities. This would enable population-level tracking and more comprehensive evaluations of digital health interventions, providing a broader evidence base to complement in-depth qualitative studies like our study.

These research directions will contribute to a more nuanced understanding of digital health equity and inform evidence-based strategies for inclusive health care digitalization. By addressing the limitations of this study and expanding the scope of the investigation, future research can build a more comprehensive picture of the challenges and opportunities in promoting equitable access to digital health services.

Conclusions

This study provides critical insights into how the digitalization of health care can deepen disparities for older, low-income, individuals with limited English proficiency. By applying an intersectional lens to the study of digital health inequalities, our research reveals the complex, overlapping, and mutually reinforcing nature of digital exclusion. Our findings underscore

XSL•FO
RenderX

the need for intersectional approaches to digital health equity that address the multifaceted nature of disadvantage.

This study makes several unique contributions to the field. First, it extends the application of intersectionality theory to digital health disparities, demonstrating how multiple dimensions of disadvantage interact to shape experiences with digital health services, particularly video consultations. This approach has revealed nuanced insights into how different forms of marginalization compound to create unique barriers to accessing and benefiting from digital health innovations.

Second, our development of theoretically informed user personas [40], grounded in intersectionality and digital capital theories, offers a novel methodological approach for representing the complex lived experiences of multiply disadvantaged patients. These personas provide a powerful tool for humanizing research findings and informing patient-centered service design in digital health contexts.

Finally, by centering the voices of marginalized patients, our research exposes how the rapid shift to video consultations has inadvertently exacerbated existing inequities and eroded trust for some susceptible groups. This challenges prevailing narratives about the universally positive impact of digital health innovations and highlights the need for more nuanced, context-sensitive approaches to digital health implementation.

Furthermore, our findings on the erosion of trust, the sense of "absent presence" in digital consultations, the shift from relational to distributed continuity of care, the weakening of patient-provider relationships, the compounding of oppression, the increased complexity leading to disorientation, the influence of previous experiences of discrimination on engagement, and the reduction in patient agency contribute new insights to the ongoing discourse on the impact of health care digitalization.

In conclusion, as health care continues to digitalize, it is imperative that we remain vigilant to the unintended consequences of technological change and work to ensure that the benefits of innovation are equitably distributed. This will require a fundamental rethinking of how we design, deploy, and evaluate digital health interventions, as well as a renewed commitment to the principles of social justice and health equity.

Acknowledgments

This was an empirical doctoral study by LH, based in UK primary care and linked to the Remote by Default 2 study [59]; the PhD is supported by the Health Foundation's grant to the University of Cambridge for the Healthcare Improvement Studies Institute. The study was overseen by the independent advisory group of the Remote by Default 2 study.

Conflicts of Interest

None declared.

References

1. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
2. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, et al. The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York City. J Urban Health 2021 Apr;98(2):183-186 [FREE Full text] [doi: 10.1007/s11524-020-00508-9] [Medline: 33471281]
3. Gann B. Combating digital health inequality in the time of coronavirus. J Consum Health Internet 2020 Sep 01;24(3):278-284. [doi: 10.1080/15398285.2020.1791670]
4. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the digital divide: age and race disparities in use of an inpatient portal. Telemed J E Health 2020 May 01;26(5):603-613 [FREE Full text] [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
5. Ferguson JM, Goldstein KM, Zullig LL, Zulman DM. Gender differences in adoption and frequency of virtual primary care among men and women veterans. J Womens Health (Larchmt) 2024 Jun;33(6):749-757. [doi: 10.1089/jwh.2023.0781] [Medline: 38629443]
6. Bowleg L. The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. Am J Public Health 2012 Jul;102(7):1267-1273. [doi: 10.2105/AJPH.2012.300750] [Medline: 22594719]
7. Weber L, Fore ME. Race, ethnicity, and health: an intersectional approach. In: Vera H, Feagin JR, editors. Handbooks of the Sociology of Racial and Ethnic Relations. Boston, MA: Springer; 2007.
8. Heard E, Fitzgerald L, Wigginton B, Mutch A. Applying intersectionality theory in health promotion research and practice. Health Promot Int 2020 Aug 01;35(4):866-876. [doi: 10.1093/heapro/daz080] [Medline: 31390472]
9. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
10. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SA, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]

XSL•FO
RenderX

11. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
12. Husain L, Greenhalgh T, Hughes G, Finlay T, Wherton J. Desperately seeking intersectionality in digital health disparity research: narrative review to inform a richer theorization of multiple disadvantage. J Med Internet Res 2022 Dec 07;24(12):e42358 [FREE Full text] [doi: 10.2196/42358] [Medline: 36383632]
13. NHS long term plan. National Health Service. URL: https://www.longtermplan.nhs.uk/ [accessed 2024-12-23]
14. Remote consulting. NHS England. 2023 Jun 12. URL: https://www.england.nhs.uk/long-read/remote-consulting/ [accessed 2024-11-17]
15. Online and video consultation undertaken in general practice data collection. NHS England. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/online-and-video-consultation-undertaken-in-general-practice [accessed 2024-11-17]
16. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021 Jun;5(3):BJGPO.2021.0040 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
17. van Dijk J, Hacker K. The digital divide as a complex and dynamic phenomenon. Inf Soc 2003 Sep;19(4):315-326. [doi: 10.1080/01972240309487]
18. Qureshi S. Creating a better world with information and communication technologies: health equity. Inf Technol Dev 2016 Jan 22;22(1):1-14. [doi: 10.1080/02681102.2015.1121585]
19. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
20. Lenhart A, Horrigan JB. Re-visualizing the digital divide as a digital spectrum. IT Soc 2003;1(5):23-39.
21. Shahid S, Hogeveen S, Sky P, Chandra S, Budhwani S, de Silva R, et al. Health equity related challenges and experiences during the rapid implementation of virtual care during COVID-19: a multiple case study. Int J Equity Health 2023 Mar 11;22(1):44 [FREE Full text] [doi: 10.1186/s12939-023-01849-y] [Medline: 36906566]
22. Hoffman LC. Reconnecting the patient: why telehealth policy solutions must consider the deepening digital divide. Indiana Health Law Rev 2022 Jul 08;19(2):351-385. [doi: 10.18060/26409]
23. Bathija P, Krupinski EA, Rodriguez JA, Sklar T. Achieving digital health equity by personalizing the patient experience. Telemed Rep 2023 Jun 01;4(1):166-173 [FREE Full text] [doi: 10.1089/tmr.2023.0018] [Medline: 37405125]
24. Zahir A, Yip D, Garcia C, Smith AN, Dhatt Z, Duke M, et al. "I needed for you to see what I'm talking about": experiences with telehealth among homeless-experienced older adults. Gerontol Geriatr Med 2023 May 06;9:23337214231172650 [FREE Full text] [doi: 10.1177/23337214231172650] [Medline: 37168020]
25. Choi NG, DiNitto DM, Marti CN, Choi BY. Telehealth use among older adults during COVID-19: associations with sociodemographic and health characteristics, technology device ownership, and technology learning. J Appl Gerontol 2022 Mar;41(3):600-609 [FREE Full text] [doi: 10.1177/07334648211047347] [Medline: 34608821]
26. Crenshaw K. Demarginalizing the intersection of race and sex: a Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chic Leg Forum 1989;1989(1):139-167. [doi: 10.4324/9780429500480-5]
27. Carastathis A. The concept of intersectionality in feminist theory. Philos Compass 2014 Apr 07;9(5):304-314. [doi: 10.1111/phc3.12129]
28. Carter M, Grover V. Me, my self, and I(T): conceptualizing information technology identity and its implications. MIS Q 2015;39(4):931-957. [doi: 10.25300/misq/2015/39.4.9]
29. Whitley EA, Gal U, Kjaergaard A. Who do you think you are? A review of the complex interplay between information systems, identification and identity. Eur J Inf Syst 2019 Jan 28;23(1):17-35. [doi: 10.1057/ejis.2013.34]
30. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Berkeley, CA: University of California Press; 1984.
31. Ragnedda M, Ruiu ML. Digital Capital: A Bourdieusian Perspective on the Digital Divide. Bingley, UK: Emerald Publishing Limited; 2020.
32. Bourdieu P. The field of cultural production, or: the economic world reversed. Poetics 1983 Nov;12(4-5):311-356. [doi: 10.1016/0304-422x(83)90012-8]
33. Robinson L. A taste for the necessary: a Bourdieuian approach to digital inequality. Inf Commun Soc 2009 Jun;12(4):488-507. [doi: 10.1080/13691180902857678]
34. Ragnedda M, Ruiu ML, Addeo F. Measuring digital capital: an empirical investigation. New Media Soc 2019 Aug 27;22(5):793-816. [doi: 10.1177/1461444819869604]
35. Bannykh G, Kostina S. Measuring digital capital: methodological approaches. KnE Soc Sci 2022 Feb 04:96-103. [doi: 10.18502/kss.v7i2.10282]
36. Clandinin DJ, Connelly FM. Narrative Inquiry: Experience and Story in Qualitative Research. Hoboken, NJ: John Wiley & Sons; 2000.
37. Riessman CK. Narrative Analysis. Thousand Oaks, CA: SAGE Publications; 1993.
38. equityXdesign. URL: https://www.equityxdesign.xyz/ [accessed 2024-12-23]

XSL•FO
RenderX

39. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA: SAGE Publications; 1985.
40. Husain L, Finlay T, Husain A, Wherton J, Hughes G, Greenhalgh T. Developing user personas to capture intersecting dimensions of disadvantage in older patients who are marginalised: a qualitative study. Br J Gen Pract 2024 Jan 19;74(741):e250-e257. [doi: 10.3399/bjgp.2023.0412]
41. Gergen KJ. The challenge of absent presence. In: Katz JE, Aakhus M, editors. Perpetual Contact: Mobile Communication, Private Talk, Public Performance. Cambridge, UK: Cambridge University Press; 2002:227-241.
42. Kofod FG, Assing Hvidt E, Arreskov AB, Guassora AD. Interpersonal contact and altered sensory conditions in video consultation - a qualitative interview study in Danish general practice. Scand J Prim Health Care 2024 Dec 16;42(4):650-658 [FREE Full text] [doi: 10.1080/02813432.2024.2376744] [Medline: 39012270]
43. Stern DN. Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy, and Development. Oxford, UK: Oxford University Press; 2010.
44. Bazargan M, Cobb S, Assari S. Discrimination and medical mistrust in a racially and ethnically diverse sample of California adults. Ann Fam Med 2021;19(1):4-15 [FREE Full text] [doi: 10.1370/afm.2632] [Medline: 33431385]
45. Roberts KM, Trejo AN. Provider, heal thy system: an examination of institutionally racist healthcare regulatory practices and structures. Contemp Fam Ther 2022 Jan 28;44(1):4-15 [FREE Full text] [doi: 10.1007/s10591-021-09630-5] [Medline: 35106026]
46. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York, NY: St. Martin's Publishing Group; 2018.
47. Kim HS, Kwon IH, Cha WC. Future and development direction of digital healthcare. Healthc Inform Res 2021 Apr;27(2):95-101 [FREE Full text] [doi: 10.4258/hir.2021.27.2.95] [Medline: 34015874]
48. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012;2(4):e001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
49. Latifi R, Doarn CR, Merrell RC. Telemedicine, Telehealth and Telepresence: Principles, Strategies, Applications, and New Directions. Cham, Switzerland: Springer; 2020.
50. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY: Basic Books; 2019.
51. Pichelstorfer A. The Logic of Care: Health and the Problem of Patient Choice. Milton Park, UK: Routledge; 2008.
52. Feenberg A. Critical theory of technology: an overview. Tailor Biotechnol 2005;1(1):47-64. [doi: 10.5040/9798400670268.ch-001]
53. Nimmon L, Stenfors-Hayes T. The "handling" of power in the physician-patient encounter: perceptions from experienced physicians. BMC Med Educ 2016 Apr 18;16(1):114 [FREE Full text] [doi: 10.1186/s12909-016-0634-0] [Medline: 27091146]
54. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
55. Solórzano DG, Yosso TJ. Critical race methodology: counter-storytelling as an analytical framework for education research. Qual Inquiry 2002 Feb 01;8(1):23-44. [doi: 10.1177/107780040200800103]
56. Medina J. The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations. Oxford, UK: Oxford University Press; 2013.
57. Gilliard C. Quotes from the episode. Teaching in Higher Ed. URL: https://teachinginhighered.com/podcast/digital-redlining-privacy/ [accessed 2024-04-29]
58. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
59. Greenhalgh T, Shaw SE, Alvarez Nishio A, Booth A, Byng R, Clarke A, et al. Protocol: remote care as the 'new normal'? Multi-site case study in UK general practice. NIHR Open Res 2022 Aug 8;2:46 [FREE Full text] [doi: 10.3310/nihropenres.13289.1] [Medline: 37881300]

Abbreviations

IS: information system
NHS: National Health Service

XSL•FO
RenderX

Edited by K Williams; submitted 22.08.24; peer-reviewed by J Woo, J Walumbe; comments to author 07.11.24; revised version received 18.11.24; accepted 26.11.24; published 06.01.25.

Please cite as:
Husain L, Greenhalgh T
Examining Intersectionality and Barriers to the Uptake of Video Consultations Among Older Adults From Disadvantaged Backgrounds With Limited English Proficiency: Qualitative Narrative Interview Study
J Med Internet Res 2025;27:e65690
URL: https://www.jmir.org/2025/1/e65690
doi:10.2196/65690
PMID:

©Laiba Husain, Trisha Greenhalgh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Transformation of Rheumatology Care in Germany: Cross-Sectional National Survey

Susann May[1], MPH; Robert Darkow[2], Dr rer med; Johannes Knitza[3,4], MHBA, MD, PhD; Katharina Boy[1], BSc; Philipp Klemm[5], MD; Martin Heinze[1,6], MD; Nicolas Vuillerme[3,7,8], PhD; Pascal Petit[3], PhD; Patricia Steffens-Korbanka[9], MFA; Heike Kladny[10], MFA; Johannes Hornig[9], MD; Peer Aries[11], MD; Martin Welcker[10], MD; Felix Muehlensiepen[1,3], MPH, Dr rer med, PhD

[1]Center for Health Services Research, Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Rüdersdorf bei Berlin, Germany

[2]FH JOANNEUM Gesellschaft mbH, Graz, Austria

[3]AGEIS, Université Grenoble Alpes, Grenoble, France

[4]Institute for Digital Medicine, University Hospital of Giessen and Marburg, Philipps University Marburg, Marburg, Germany

[5]Department of Rheumatology, Immunology, Osteology, Kerckhoff-Klinik GmbH, Bad Nauheim, Germany

[6]University Clinic for Psychiatry and Psychotherapy, Brandenburg Medical School, Immanuel Hospital Rüdersdorf, Rüdersdorf bei Berlin, Germany

[7]Institut Universitaire de France, Paris, France

[8]LabCom Telecom4Health, Orange Labs & Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP-UGA, Grenoble, France

[9]Rheumapraxis an der Hase, Osnabrück, Germany

[10]MVZ für Rheumatologie Dr. Martin Welcker GmbH, Planegg, Germany

[11]Immunologikum Hamburg, Rheumatologie & Klinische Immunologie, Hamburg, Germany

Corresponding Author:
Felix Muehlensiepen, MPH, Dr rer med, PhD
Center for Health Services Research
Faculty of Health Sciences Brandenburg
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf bei Berlin, 15562
Germany
Phone: 49 1636055860
Email: Felix.Muehlensiepen@mhb-fontane.de

Abstract

Background: In recent years, health care has undergone a rapid and unprecedented digital transformation. In many fields of specialty care, such as rheumatology, this shift is driven by the growing number of patients and limited resources, leading to increased use of digital health technologies (DHTs) to maintain high-quality clinical care. Previous studies examined user acceptance of individual DHTs in rheumatology, such as telemedicine, video consultations, and mHealth. However, it is essential to conduct cross-technology and continuous analyses of user acceptance and DHT use to maximize the benefits for all relevant stakeholders.

Objective: This study aimed to explore the current acceptance, use, and preferences regarding DHTs among patients in rheumatology care in Germany.

Methods: Rheumatology patients from 3 clinics in Germany were surveyed to understand their perspectives on DHTs. The survey included main themes, including acceptance, preferences, COVID-19's impact, potential, and barriers related to DHTs. The data were analyzed using descriptive statistics and correlation analysis.

Results: Out of 337 participants, 53% (179/337) reported using DHTs. Specific technologies included wearables (72/337, 21%), mHealth apps (71/337, 21%), digital therapeutics (32/337, 9%), electronic prescriptions (30/337, 9%), video consultations (15/337, 4%), and at-home blood self-sampling (3/337, 1%). Nearly two-thirds (220/337, 65%) found DHTs useful, and 69% (233/337) held a generally positive attitude toward DHTs. Attitudes shifted positively during the COVID-19 pandemic for 40% (135/337) of participants. Higher education was more prevalent among DHT users (114/179, 63.7%) compared with nonusers (42/151, 27.8%; *P*=.02). The main potential benefits identified were location-independent use (244/337, 72%) and time-independent use

XSL•FO
RenderX

(216/337, 64%). Key barriers included insufficient user knowledge (165/337, 49%) and limited information on DHTs (134/337, 40%).

Conclusions: Patient acceptance and use of DHTs in rheumatology is increasing in Germany. A prospective, standardized monitoring of digital transformation in rheumatology care is highly needed.

(J Med Internet Res 2025;27:e52601) doi:10.2196/52601

KEYWORDS

telemedicine; digital health technologies; rheumatology; COVID-19; questionnaire; telehealth; eHealth; digital health; survey; rheumatism; Germany; Europe; national; use; experience; attitude; opinion; perception; perspective; acceptance; preference; correlation

Introduction

Health care is undergoing a multidisciplinary digital transformation, which refers to "a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies" [1].

Within the realm of rheumatology care, a notable surge in digital technologies has transpired in recent years [2,3]. This occurrence stems from the existing global disparity between the increasing number of patients with rheumatic and musculoskeletal disease (RMD) [4] and the static or even decreasing availability of human resources in the field of rheumatology care [5,6].

Consequently, Miloslavsky and Bolster [7] have discerned the amplified use of telemedicine as a prospective remedy to ensure the continuity of rheumatological care in the future [7]. Concurrently, and accelerated by the COVID-19 pandemic, the European Alliance of Associations for Rheumatology (EULAR) has recently published points to consider for remote care in RMD [8].

In Germany, a country struggling with a shrinking and aging population and a shortage of health care professionals, digital health technologies (DHTs) in rheumatology care have proven effective in improving rheumatology care [2,9]. Prominent among the prevailing DHTs are video consultations, sensors, wearables, digital health applications, and digital therapeutics with varied objectives (symptom checkers, ePRO [electronic patient-reported outcome] documentation, or patient education), social media platforms or messenger platforms [10], which can also be combined with home-based self-sampling [11-13].

Past studies have assessed patients' and physicians' acceptance of DHTs in the domain of rheumatology within Germany. These investigations primarily centered on individual digital modalities, such as telemedicine [13,14], mHealth [15], and video consultations [16,17]. Following the onset of the COVID-19 pandemic, the influence of this global crisis on the adoption of DHTs was also scrutinized [3,17]. It remains essential to continuously analyze the digital transformation in rheumatology to maximize the benefits for all relevant stakeholders. In particular, the patient perspective is critical to ensuring that DHTs actually improve health care, meets patient needs, and builds patient trust in the use of DHTs. Thus, the aim of this study was to explore the current acceptance, use, and preferences regarding DHTs among patients in rheumatology care in Germany.

Methods

Overview

To explore the patients' perspectives and preferences regarding digital health systems in rheumatology care, we recruited rheumatology patients in 3 outpatient clinics. Inclusion criteria were being aged ≥18 years, literate in German, having the physical and mental ability to fill out a paper-pencil questionnaire, and having a diagnosis of RMD. The survey was conducted between February 2023 and April 2023. The 3 clinics were sent 125 questionnaires with the request to distribute the questionnaires to eligible patients.

The questionnaire was created based on a literature review and the results of a qualitative study with health care providers (HCPs) and nurses. The questionnaire addressed the following topics: acceptance of DHTs, use of DHTs, preferences for DHTs, the impact of COVID-19 on digital transformation, potential benefits, and barriers regarding DHTs, and sociodemographic characteristics. The survey was pilot-tested with 10 patients to detect necessary formatting and wording changes. Minor revisions were made accordingly. Please refer to Multimedia Appendix 1 for the German version of the questionnaire and Multimedia Appendix 2 for the translated version in English.

Data were analyzed using quantitative descriptive analysis and correlation analysis supported by SPSS (version 23.0; IBM). For correlation analyses, the data were converted into scores, and the Spearman rank correlation coefficient was calculated. The α error distribution was chosen to be 2-sided, as an undirected correlation is assumed to be open to hypotheses. These correlations had a correlation coefficient higher than 0.29.

Ethical Considerations

This study was approved by the ethics committee of Brandenburg Medical School (E-02-20211028). All patients provided informed consent. The rheumatology clinics received an expense allowance of €150 (US $ 158) for the distribution and return of the questionnaires. The study data are anonymous. The questionnaires were digitized and then stored at the Center for Mental Health at Brandenburg Medical School.

XSL•FO
RenderX

Results

Patient Characteristics

In total, 337 patients with RMD completed the questionnaire. The number of patients rejecting participation was not measured. The study sample's demographics are shown in Table 1. The mean age was 52.5 (SD 14.2) years; 64% (219/337) were female, 32% (108/337) were male, 1% were (3/337) nonbinary, and 2% (7/337) did not give any information on gender. On average, participants had been receiving rheumatologic treatment for a median 8 years (mean 9.35 years, SD 7.63; range: 1 month-399 months), and the median time with the diagnosis was 8.9 years (mean 10.2 years, SD 8.33; range: 3 months-579 months).

Table 1. Characteristics of participants.

Characteristics	Participants (N=337), n (%)
Sex	
Female	219 (65)
Male	108 (32)
Nonbinary	3 (1)
No data	7 (2)
Diagnosis	
Axial spondyloarthritis	13 (4)
Fibromyalgia	7 (2)
Psoriatic arthritis	40 (12)
Rheumatoid arthritis	182 (54)
Sjörgren syndrome	10 (3)
Spondyloarthritis	21 (6)
Systemic lupus erythematosus	9 (3)
Other	39 (12)
No data	16 (5)
Education level	
>10 years of school completed	156 (46)
9-10 years of school completed	163 (48)
Student	2 (1)
No formal schooling	2 (1)
No data	14 (4)
Vocational training	
Apprenticeship	188 (56)
Bachelor's, master's, or magister's degree, diploma	109 (3)
Started an apprenticeship	6 (2)
Without apprenticeship	11 (3)
No data	23 (7)
Place of residence (number of inhabitants)	
Rural region (<5000)	70 (21)
Small city (>5000-20,000)	73 (22)
Medium city (>20,000-100,000)	63 (19)
Large city (>100,000-1,000,000)	28 (8)
Metropolis city (>1,000,000)	93 (28)
No data	10 (3)

XSL•FO
RenderX

Attitudes Toward DHTs

About half of the patients used DHTs overall (179/337, 53%) (Table 2). Overall, most patients (220/337, 65%) rated DHTs as useful, while a third remained neutral in their assessment of usefulness. Almost two-thirds of the patients (233/337, 69%) were positive or rather positive toward DHTs. Among all participants, 55% (185/377) stated that their attitude toward DHTs had not changed due to the COVID-19 pandemic. However, 40% (135/337) mentioned that their attitude had become more positive.

Table 2. Attitudes toward digital health technologies.

Items	Participants (N=337), n (%)
Do you use digital health technologies?	
Yes	179 (53)
No	151 (45)
No data	7 (2)
Digital health technologies are useful.	
Strongly agree	99 (29)
Agree	121 (36)
Neutral	99 (29)
Disagree	5 (1)
Strongly disagree	7 (2)
No data	6 (2)
How do you rate your attitude towards digital health services?	
Positive	118 (35)
Rather positive	115 (34)
Neutral	77 (23)
Rather negative	17 (5)
Negative	4 (1)
No data	6 (2)
Did you change your attitude due to the COVID-19 pandemic?	
Yes, the attitude has become more positive	135 (40)
Yes, the attitude has become more negative	8 (2)
No	185 (55)
No data	9 (3)

Use of DHTs

Almost half of the patients (143/337, 42%) were already using email for communication with physicians before the COVID-19 pandemic (Table 3). Currently, this value has increased to 58% (195/337). The number of patients using video consultations was very low: 4% (15/337). However, 14% (48/337) stated that they would use video consultations in the future. We found that 9% (32/337) were using reimbursed digital health applications (German: Digitale Gesundheitsanwendung; DiGA). Overall, 21% (71/337) were using other mobile health apps. Only 1% (3/337) had used blood self-sampling. Almost a quarter (78/337, 23%) stated that they wanted to use electronic prescriptions in the future.

XSL•FO
RenderX

Table 3. Digital health technology use.

	Used before COVID-19, n (%)	I currently use, n (%)	I will use in the future, n (%)	I am not interested, n (%)	I don't use, n (%)	I don't know, n (%)
Video consultation	7 (2)	15 (4)	48 (14)	14 (4)	191 (57)	53 (16)
Digital therapeutics (DiGA[a])	14 (4)	32 (9)	38 (11)	10 (3)	154 (46)	86 (26)
mHealth apps (not DiGA)	19 (6)	71 (21)	40 (12)	10 (3)	152 (45)	54 (16)
Wearables	19 (6)	72 (21)	21 (6)	19 (6)	175 (52)	43 (13)
Blood self-sampling at home	2 (1)	3 (1)	9 (3)	50 (15)	187 (55)	99 (29)
Electronic prescription	7 (2)	30 (9)	78 (23)	5 (1)	147 (44)	75 (22)

[a]DiGA: digital health application.

Potential Benefits of and Barriers to DHTs

Independence in terms of location (244/337, 72%) and time (216/337, 64%) and generally increased flexibility (170/337, 50%) were cited most often by patients as potential benefits of digital health. The detailed results are displayed in Table 4.

The biggest barrier at present is the lack of information. Thus, 49% (165/337) stated that they do not have sufficient knowledge, and 40% (134/337) stated that they are not sufficiently informed about DHTs. A third (118/337, 35%) stated that technical equipment is a barrier to using the tools.

Table 4. Potential benefits and barriers of digital health technologies.

Items	Participants (N=337), n (%)
Potential benefits	
Location independent use	244 (72)
Time-independent use	216 (64)
Detailed documentation of the course of the disease	139 (41)
Cost savings	134 (40)
More options for accessing information, diagnostics, and therapy	125 (37)
Accessibility	85 (25)
More flexibility	170 (50)
Better preparation for the physician-patient consultation	139 (41)
Needs-based care	68 (20)
No potential benefits	32 (9)
Barriers	
Limited information about digital health services	134 (40)
Insufficient evidence of the benefits of the offers	39 (12)
Poor quality of current offers	32 (9)
Gaps in data protection	92 (27)
Lack of user-friendliness	84 (25)
Lack of accessibility	12 (4)
High costs	12 (4)
Lack of technical equipment (eg, poor Internet connection, old end devices)	118 (35)
Lack of knowledge among users	165 (49)
No need because satisfied with the current analogue solutions	57 (17)

Correlation Analyses

While correlation analysis revealed some relationships between content items (Table 5), no significant correlations between the content items and sociodemographic data were observed.

XSL•FO
RenderX

Table 5. Results of the correlation analyses.

Content items	Spearman ρ	*P* values (2-tailed)
Current use		
Positive attitude	0.458	<.001
Number of months diagnosed		
Number of months in rheumatology treatment	0.777	<.001
Potential benefits		
Location-independent use		
Detailed documentation of disease progression	0.355	<.001
Accessibility	0.328	<.001
More flexibility	0.331	<.001
No	–0.525	<.001
Detailed documentation of disease progression		
More options for accessing information, diagnostics, and therapy	0.330	<.001
More flexibility	0.336	<.001
Better preparation for the doctor-patient consultation	0.400	<.001
Needs-based care	0.330	<.001
Cost savings		
Accessibility	0.380	<.001
More flexibility	0.320	<.001
Needs-based care	0.302	<.001
More possibilities to access information, diagnostics, therapy		
Accessibility	0.346	<.001
More flexibility	0.405	<.001
Better preparation for the doctor-patient consultation	0.305	<.001
Accessibility		
More flexibility	0.357	<.001
Needs-based care	0.338	<.001
More flexibility		
Needs-based care	0.380	<.001
None	–0.307	<.001
Better preparation for the doctor-patient consultation		
Needs-based care	0.360	<.001
Barriers		
Lack of technical equipment		
Lack of knowledge among users	0.307	<.001

Discussion

To explore the current patient acceptance, use, and preferences for DHTs in rheumatology care, we performed a paper-pencil questionnaire survey among patients with RMD in Germany.

Principal Findings

More than half (179/337, 53%) of 337 patients reported that they use DHTs. Overall, 21% (72/337) used wearables, 21% (71/337) used mHealth apps, 9% (32/337) used digital therapeutics (DiGA), 9% (30/337) used electronic prescriptions, 4% (15/337) used video consultations, and 1% (3/337) used at-home blood self-sampling. Nearly two-thirds of the patients with RMD (220/337, 65%) rated DHTs as useful. While 69% (233/337) reported a generally positive attitude toward DHTs, about 40% (135/337) mentioned their attitudes became more positive due to the COVID-19 pandemic. The main potential benefits of DHTs reported by the patients were location-independent use (244/337, 72%) and time-independent use (216/337, 64%). The main barriers included insufficient

XSL•FO
RenderX

knowledge among users (165/337, 49%) and limited information about digital health services (134/337, 40%).

Comparison With Previous Work

Our findings are aligned with previous studies that have examined patients' acceptance of DHTs in rheumatology [3,14-19]. Comparing our results with previous findings, clear trends emerge that further underscore the growing acceptance of DHTs among patients with RMD.

In a survey conducted from September 2019 to December 2019 among 766 German patients with RMD [18], only 51% (364/718) of participants indicated familiarity with the term "telemedicine." A mere 30% (210/690) expressed intentions to try telemedicine in the future, and a total of 21% (139/663) expressed a desire for their rheumatologist to offer telemedicine. In this study, conducted from February 2023 to April 2023, that is, after the COVID-19 period, 53% (179/337) of participants reported using DHTs. It should be noted, however, that while the terms telemedicine and DHTs are closely related, they are not synonymous.

In the 2019 survey, 0.3% of patients with RMD reported having experienced a video consultation with a physician. The current results reveal a 4% (15/337) use rate of video consultations, still representing a modest figure, suggesting that video consultations in rheumatology care remain an exceptional practice. These findings corroborate the outcomes observed by Richter et al [17]. In a survey of rheumatologists in Germany, 27% (55/205) reported offering video consultations during COVID-19 lockdown phases, with the frequency of provided video conferences diminishing as the pandemic progressed.

Knitza et al [15] reported that in 2018/2019, most patients with RMD (68%) believed that using medical apps could be beneficial for their own health [15]. However, out of 193 patients, only 8 (4%) were currently using medical apps. In the fall of 2020, Kernder et al [3] explored patient and rheumatologist attitudes toward digital technologies, particularly digital health applications. Even a higher rate of patients (222/299, 74%) and also rheumatologists (98/129, 76%) believed that digital health apps were valuable for managing RMDs. Compared with Knitza et al [15], our results reveal that digital health app use increased notably (71/337, 21%). In our study, a distinction was made between certified and prescribed digital therapeutics (DiGA) and other nonreimbursed digital health apps. Interestingly, patients reported less use of prescribed DiGAs (32/337, 9%) compared with other noncertified mHealth apps (21%). This discrepancy could be attributed to the absence of DiGAs explicitly tailored for RMDs, whereas non-DiGA mHealth apps fill this gap. A first pilot study evaluating DiGA use in rheumatology [20] revealed high patient acceptance and some clinical benefits, yet poor adherence as a major limiting barrier.

According to Kernder et al [3], 38% (112/299) of patients reported a positive change in attitude due to COVID-19, comparable to 40% (135/337) in this study. The most commonly cited advantages of DHAs were their independence from time and place, which were also expressed by participants in our study.

Our study also inquired about the use and acceptance of blood self-sampling, a prospect that has recently gained significance in future rheumatological care [11-13]. However, our results suggest that self-sampling currently remains largely confined to research settings.

In a recent secondary data analysis [21], we demonstrated that, specifically, older patients with RMD residing in rural areas, who could potentially benefit from telemedicine, currently lack the motivation to embrace it and appear to require additional support. However, these relationships were not confirmed by this study. Nevertheless, our data reaffirm the profound relevance of knowledge in the use of digital technologies in rheumatology care. In line with earlier findings [3,14,19,21], our participants identified "lack of knowledge among users" as the main barrier to use. Besides the patient perspective, the viewpoint of HCPs is also pertinent to the implementation of digital technologies. Lack of knowledge among rheumatologists was previously identified as a major barrier to implementing ePROs in routine care [22]. Conveniently, recent surveys also indicated a positive inclination of HCPs toward digital technologies in rheumatology [3,14,16,19].

Limitations

Our study has several limitations. Despite our best efforts, it is possible that our survey did not capture all important emerging technologies relevant to rheumatology. In addition, the terminology used in this survey for DHTs may have caused confusion among participants, potentially leading to information bias. In addition, our survey reached a selected population from 3 outpatient clinics. This also applies to the selection of study participants, as we assume that individuals with a specific interest in digital health were more likely to have completed our questionnaire. A strength of previous surveys [3,16,17,19,22] is that the survey was paper-pen–based (instead of digital) to minimize selection bias.

Finally, it is essential to acknowledge that numerous research groups are focusing on measuring the use and acceptance in the domain of digital rheumatology. However, these studies often use varying approaches and terminologies and refer to different study populations. Therefore, the comparability of data across studies for the purpose of continuous monitoring of the digital transformation in rheumatology remains limited.

Implications

Considering the heterogeneity regarding digital rheumatology surveys and fast transformation, we advocate for a standardized, regular, survey-based monitoring of the digital transformation in rheumatology from the perspectives of both patients with RMD and HCPs. Ideally, this monitoring should be conducted on an international level, including dedicated societies such as EULAR and the Digital Rheumatology Network. Furthermore, the specific digital technologies, their nomenclature, and other questionnaire content should be harmonized through Delphi studies in collaboration with an international expert board involving input from patients.

The results depict high acceptance regarding DHTs, which is currently limited mainly by a lack of knowledge. Dedicated

XSL•FO
RenderX

education for patients with RMD and the treating HCPs is necessary to foster implementation in routine clinical practice.

Conclusions

The digital transformation in rheumatology care in Germany is progressing. Patient acceptance and use is increasing. This provides hope that, despite the rising burden of disease and stagnating human resources, digital health can continue to ensure high-quality care for patients with RMD in the future. A prospective, standardized monitoring of digital transformation in rheumatology care is highly needed.

Acknowledgments

The authors would like to thank all participants of this study. We also owe special gratitude to the 3 outpatient clinics and their teams that supported this study and distributed the questionnaires.

This study was conducted in cooperation with RheumaDatenRhePort GbR (RHADAR) and uses data from a health care analysis commissioned by AbbVie Deutschland GmbH & Co KG.

Authors' Contributions

FM and SM had full access to all the data in the study and take responsibility for the integrity of the data and accuracy of the data analysis. FM, SM, MW, PA, and JH conceptualized and designed the study. FM, SM, MW, PA, JH, PSK, HK, JK, and PK were involved in the acquisition of data. FM, SM, RD, JK, KB, MH, NV, and PP were involved in the analysis and interpretation of data. All the authors were involved in drafting the manuscript and critically revising it for important intellectual content, and they approved the final version of the manuscript to be submitted for publication.

Conflicts of Interest

MW reports support and grants for projects, talks, and discussions by Abbvie, BMS, Boehringer, Galapagos, Gilead, GSK, Hexal, Janssen, Novartis, Pfizer, and UCB. MW is also a member of RHADAR (RheumaDatrenRhePort).

Multimedia Appendix 1
Original questionnaire (German).
[PDF File (Adobe PDF File), 367 KB - jmir_v27i1e52601_app1.pdf]

Multimedia Appendix 2
Questionnaire - English translation.
[PDF File (Adobe PDF File), 365 KB - jmir_v27i1e52601_app2.pdf]

References

1. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021;123:557-567 [FREE Full text] [doi: 10.1016/j.jbusres.2020.10.030]
2. Mühlensiepen F, Kurkowski S, Krusche M, Mucke J, Prill R, Heinze M, et al. Digital health transition in rheumatology: a qualitative study. Int J Environ Res Public Health 2021;18(5):2636 [FREE Full text] [doi: 10.3390/ijerph18052636] [Medline: 33807952]
3. Kernder A, Morf H, Klemm P, Vossen D, Haase I, Mucke J, et al. Digital rheumatology in the era of COVID-19: results of a national patient and physician survey. RMD Open 2021;7(1):e001548 [FREE Full text] [doi: 10.1136/rmdopen-2020-001548] [Medline: 33622673]
4. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases database. Ann Rheum Dis 2019;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
5. Al Maini M, Adelowo F, Al Saleh J, Al Weshahi Y, Burmester GR, Cutolo M, et al. The global challenges and opportunities in the practice of rheumatology: white paper by the World Forum on Rheumatic and Musculoskeletal Diseases. Clin Rheumatol 2015;34(5):819-829 [FREE Full text] [doi: 10.1007/s10067-014-2841-6] [Medline: 25501633]
6. Ward IM, Schmidt TW, Lappan C, Battafarano DF. How critical is tele-medicine to the rheumatology workforce? Arthritis Care Res (Hoboken) 2016;68(10):1387-1389 [FREE Full text] [doi: 10.1002/acr.22853] [Medline: 26866514]
7. Miloslavsky EM, Bolster MB. Addressing the rheumatology workforce shortage: a multifaceted approach. Semin Arthritis Rheum 2020;50(4):791-796 [FREE Full text] [doi: 10.1016/j.semarthrit.2020.05.009] [Medline: 32540672]
8. de Thurah A, Bosch P, Marques A, Meissner Y, Mukhtyar CB, Knitza J, et al. 2022 EULAR points to consider for remote care in rheumatic and musculoskeletal diseases. Ann Rheum Dis 2022;81(8):1065-1071. [doi: 10.1136/annrheumdis-2022-222341] [Medline: 35470160]

XSL•FO
RenderX

9. Knitza J, Callhoff J, Chehab G, Hueber A, Kiltz U, Kleyer A, et al. [Position paper of the Commission on Digital Rheumatology of the German Society of Rheumatology: tasks, targets and perspectives for a modern rheumatology]. Z Rheumatol 2020;79(6):562-569. [doi: 10.1007/s00393-020-00834-y] [Medline: 32651681]
10. Kataria S, Ravindran V. Digital health: a new dimension in rheumatology patient care. Rheumatol Int 2018;38(11):1949-1957. [doi: 10.1007/s00296-018-4037-x] [Medline: 29713795]
11. Zarbl J, Eimer E, Gigg C, Bendzuck G, Korinth M, Elling-Audersch C, et al. Remote self-collection of capillary blood using upper arm devices for autoantibody analysis in patients with immune-mediated inflammatory rheumatic diseases. RMD Open 2022;8(2):e002641 [FREE Full text] [doi: 10.1136/rmdopen-2022-002641] [Medline: 36104118]
12. Muehlensiepen F, May S, Zarbl J, Vogt E, Boy K, Heinze M, et al. At-home blood self-sampling in rheumatology: a qualitative study with patients and health care professionals. BMC Health Serv Res 2022;22(1):1470 [FREE Full text] [doi: 10.1186/s12913-022-08787-5] [Medline: 36461025]
13. Mucke J, Knitza J, Muehlensiepen F, Grahammer M, Stenzel R, Simon D, et al. TELERA-asynchronous TELEmedicine for patients with rheumatoid arthritis: study protocol for a prospective, multi-center, randomized controlled trial. Front Med (Lausanne) 2021;8:791715 [FREE Full text] [doi: 10.3389/fmed.2021.791715] [Medline: 34966765]
14. Muehlensiepen F, Knitza J, Marquardt W, Engler J, Hueber A, Welcker M. Acceptance of telerheumatology by rheumatologists and general practitioners in Germany: nationwide cross-sectional survey study. J Med Internet Res 2021;23(3):e23742 [FREE Full text] [doi: 10.2196/23742] [Medline: 33690147]
15. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
16. Vossen D, Knitza J, Klemm P, Haase I, Mucke J, Kernder A, et al. [Acceptance of video consultation among patients with inflammatory rheumatic diseases depends on gender and location: results of an online survey among patients and physicians]. Z Rheumatol 2023;82(2):108-113 [FREE Full text] [doi: 10.1007/s00393-021-01052-w] [Medline: 34448915]
17. Richter JG, Chehab G, Reiter J, Aries P, Muehlensiepen F, Welcker M, et al. Evaluation of the use of video consultation in German rheumatology care before and during the COVID-19 pandemic. Front Med (Lausanne) 2022;9:1052055 [FREE Full text] [doi: 10.3389/fmed.2022.1052055] [Medline: 36507506]
18. Muehlensiepen F, Knitza J, Marquardt W, May S, Krusche M, Hueber A, et al. Opportunities and barriers of telemedicine in rheumatology: a participatory, mixed-methods study. Int J Environ Res Public Health 2021;18(24):13127 [FREE Full text] [doi: 10.3390/ijerph182413127] [Medline: 34948737]
19. Becker C, Diener M, Hueber A, Henes J, Krusche M, Ignatyev Y, et al. Unmet information needs of patients with rheumatic diseases: results of a cross-sectional online survey study in Germany. Int J Environ Res Public Health 2022;19(12):7071 [FREE Full text] [doi: 10.3390/ijerph19127071] [Medline: 35742318]
20. Labinsky H, Gupta L, Raimondo MG, Schett G, Knitza J. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023;43(4):713-719 [FREE Full text] [doi: 10.1007/s00296-022-05261-7] [Medline: 36543961]
21. Muehlensiepen F, Petit P, Knitza J, Welcker M, Vuillerme N. Factors associated with telemedicine use among patients with rheumatic and musculoskeletal disease: secondary analysis of data from a German nationwide survey. J Med Internet Res 2023;25:e40912 [FREE Full text] [doi: 10.2196/40912] [Medline: 36705950]
22. Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, et al. Acceptance, usage, and barriers of electronic patient-reported outcomes among German rheumatologists: survey study. JMIR Mhealth Uhealth 2020;8(7):e18117 [FREE Full text] [doi: 10.2196/18117] [Medline: 32390592]

Abbreviations

DHT: digital health technology
DiGA: digital health application
ePRO: electronic patient-reported outcome
EULAR: European Alliance of Associations for Rheumatology
HCP: health care provider
mHealth: mobile health
RMD: rheumatic and musculoskeletal disease

XSL•FO
RenderX

Edited by N Cahill; submitted 09.09.23; peer-reviewed by M de Zwaan, S Góngora Alonso; comments to author 18.04.24; revised version received 28.05.24; accepted 09.10.24; published 06.01.25.

Please cite as:
May S, Darkow R, Knitza J, Boy K, Klemm P, Heinze M, Vuillerme N, Petit P, Steffens-Korbanka P, Kladny H, Hornig J, Aries P, Welcker M, Muehlensiepen F
Digital Transformation of Rheumatology Care in Germany: Cross-Sectional National Survey
J Med Internet Res 2025;27:e52601
URL: https://www.jmir.org/2025/1/e52601
doi:10.2196/52601
PMID:

©Susann May, Robert Darkow, Johannes Knitza, Katharina Boy, Philipp Klemm, Martin Heinze, Nicolas Vuillerme, Pascal Petit, Patricia Steffens-Korbanka, Heike Kladny, Johannes Hornig, Peer Aries, Martin Welcker, Felix Muehlensiepen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploration of Reproductive Health Apps' Data Privacy Policies and the Risks Posed to Users: Qualitative Content Analysis

Nina Zadushlivy[1], BA, MS; Rizwana Biviji[2], MS, PhD; Karmen S Williams[3], MBA, DrPH, MSPH, MA

[1]Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, City University of New York, New York, NY, United States

[2]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[3]Department of Health Policy and Management, Graduate School of Public Health and Health Policy, City University of New York, New York, NY, United States

Corresponding Author:
Karmen S Williams, MBA, DrPH, MSPH, MA
Department of Health Policy and Management
Graduate School of Public Health and Health Policy
City University of New York
55 W. 125th Street
New York, NY, 10027
United States
Phone: 1 918 816 0915
Email: Karmen.Williams@sph.cuny.edu

Abstract

Background: Mobile health apps often require the collection of identifiable information. Subsequently, this places users at significant risk of privacy breaches when the data are misused or not adequately stored and secured. These issues are especially concerning for users of reproductive health apps in the United States as protection of sensitive user information is affected by shifting governmental regulations such as the overruling of *Roe v Wade* and varying state-level abortion laws. Limited studies have analyzed the data privacy policies of these apps and considered the safety issues associated with a lack of user transparency and protection.

Objective: This study aimed to evaluate popular reproductive health apps, assess their individual privacy policies, analyze federal and state data privacy laws governing these apps in the United States and the European Union (EU), and recommend best practices for users and app developers to ensure user data safety.

Methods: In total, 4 popular reproductive health apps—Clue, Flo, Period Tracker by GP Apps, and Stardust—as identified from multiple web sources were selected through convenience sampling. This selection ensured equal representation of apps based in the United States and the EU, facilitating a comparative analysis of data safety practices under differing privacy laws. A qualitative content analysis of the apps and a review of the literature on data use policies, governmental data privacy regulations, and best practices for mobile app data privacy were conducted between January 2023 and July 2023. The apps were downloaded and systematically evaluated using the Transparency, Health Content, Excellent Technical Content, Security/Privacy, Usability, Subjective (THESIS) evaluation tool to assess their privacy and security practices.

Results: The overall privacy and security scores for the EU-based apps, Clue and Flo, were both 3.5 of 5. In contrast, the US-based apps, Period Tracker by GP Apps and Stardust, received scores of 2 and 4.5, respectively. Major concerns regarding privacy and data security primarily involved the apps' use of IP address tracking and the involvement of third parties for advertising and marketing purposes, as well as the potential misuse of data.

Conclusions: Currently, user expectations for data privacy in reproductive health apps are not being met. Despite stricter privacy policies, particularly with state-specific adaptations, apps must be transparent about data storage and third-party sharing even if just for marketing or analytical purposes. Given the sensitivity of reproductive health data and recent state restrictions on abortion, apps should minimize data collection, exceed encryption and anonymization standards, and reduce IP address tracking to better protect users.

(J Med Internet Res 2025;27:e51517) doi:10.2196/51517

XSL•FO
RenderX

KEYWORDS

data privacy policy; reproductive health apps; Transparency, Health Content, Excellent Technical Content, Security/Privacy, Usability, Subjective; THESIS; THESIS evaluation; women's health; menstrual health; mobile health; mHealth; menstruating persons' health; mobile phone

Introduction

Background

The rights of menstruating persons in the United States have been at the center of major midcentury movements throughout history, whether that has been equal voting rights, safe working conditions ignited by the events of the Triangle Shirtwaist Factory fire, and freedom over their body autonomy. However, these rights and freedoms were further compromised in the 2022 *Dobbs v Jackson Women's Health Organization* case, where the US Supreme Court overturned the 1973 *Roe v Wade* ruling, which guaranteed a constitutional right to abortion [1-3]. One year after the overruling, the nation's perspective on reproductive health and the practice of collecting data through mobile apps designed to manage reproductive health have both deteriorated. Not only has the overruling threatened menstruating persons' rights over their own bodies in terms of pregnancy, but it has also posed a risk to the safety of the information that menstruating persons share on such digital platforms.

The emergence of mobile health (mHealth) apps, coined as such by Robert Istepanian in 2008, has drastically shifted the way in which society approaches disease diagnosing, maintenance, monitoring, and management [4]. Such a shift has eased travel times and made selection of health care providers easier. While it was previously necessary to schedule physicians' appointments to measure the health of an athlete or a patient with prediabetes, that is all simply done now through wearable devices such as Fitbit and Apple Watch or at-home glucose monitors that automatically update and transfer data from the user to the health care provider and alert users of irregularities detected, such as lowered oxygen levels and heightened heart rates. Due to the ease of use, users are increasingly turning to digital technologies such as mHealth apps for their health information and monitoring needs. However, smartphone technology use has been found to vary between nonmenstruating and menstruating persons, with 51% of nonmenstruating and 56% of menstruating persons using smartphones [5]. This trend is also reflected in health app use, with menstruating persons being more than twice as likely as nonmenstruating persons to turn to these apps for health information [5]. Over the past 6 years, the number of downloads for health and fitness apps alone has increased from over 1 billion in 2017 to 4.2 billion in 2023 [6].

This steady increase has also been observed with period trackers and reproductive health apps in particular, with a 2019 Kaiser Family Foundation study counting a third of American menstruating persons as users [7]. Additionally, in a study conducted among 360 health care providers, approximately 33% referred or provided information to their patients about sexual or reproductive health apps, and approximately 41% of those health care providers were confident and informed about sexual health apps [8]. Although there is continual discussion regarding security and confidentiality, the health care professionals indicated that there are potential benefits of sexual and reproductive health apps, such as being available to inform and assist underserved populations, engagement with and building of communities, training and education, and other health benefits. Reasons for use of sexual and reproductive health apps vary from menstruation monitoring, such as symptoms, changes, possible concerns, and cycle predictability, to fertility and conception tracking (ovulation) and organization of reproductive health information to pass on to health professionals during annual visits [9]. Such use has also allowed most menstruating persons to feel prepared, informed, and in control of their menstruation cycle in comparison to antiquated ways of tracking such as calendar markings [10]. Despite their ease of use and associated benefits, many apps rely on common yet sometimes inefficient methods to ensure that users read the terms and conditions and privacy policies. For example, the clickwrap (click-accept) approach, as observed in many apps, does little to ensure that users are actively reading the terms of use of each app. While there is no guarantee that users will read these terms, other approaches such as the scrollwrap (scrolling to the bottom to accept) approach may be better suited for mobile apps that deal with more sensitive data so that users are more inclined to read through highlighted areas and make educated decisions.

The convenience of reproductive health apps has posed challenges to users. Following the leak of the Supreme Court overruling draft in May 2022, many menstruating persons were left questioning what apps were safe to use. This led to an influx of users switching to reproductive health apps with more promising data privacy policies and practices, such as Clue and Stardust [11-13]. Given the sensitive information that these apps collect and store, in addition to recent shifts in the Supreme Court's response to reproductive rights, the safety, confidentiality, and privacy of menstruating persons have taken center stage as potential risks associated with data misuse have become a growing concern [14]. These potential risks have become even more prominent in light of lapses in user policies designed to protect consumers and manage the flow of their data (data collection, data sharing, and data monitoring). Those in historically inequitable health situations, such as low-income, rural, and Black and other minoritized communities, have an increased risk of both using health apps as supplements for access to quality health care and being targeted and criminalized for seeking certain types of reproductive health care [15,16]. Professionals in the field, including legal scholars from the American Bar Association, have also expressed their concern about the potential misuse of advertising data [17,18]. In addition, while there has been misinformation regarding the misuse of private reproductive health data, it is vital for users of these apps to know and understand what data are collected, how they may be used, which companies may not meet their expectations and promises, and how they can protect their

XSL•FO
RenderX

sensitive reproductive data by choosing apps that prioritize consumer safety [18].

Policy Evaluations and Legal Framework

Although there are currently no comprehensive federal laws governing data privacy in the United States, certain case proceedings, agencies, and constitutional protections—such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [19], the Fourth Amendment [20], the Food and Drug Administration [21] monitoring of medical devices, and the Federal Trade Commission (FTC) [22]—aim to protect individual privacy and medical records associated with identifiable health information. However, there is a lack of clarity regarding how these guidelines extend to mHealth apps and internet resources; for example, HIPAA does not restrict how apps may use health information that has been disclosed pursuant to the user's right of access and how certain court rulings apply to and ensure privacy in personal, home, or public settings. In addition, privacy policies and data protections vary depending on an app's country of origin. For instance, the European Union (EU) enforces comprehensive data privacy laws, providing a robust framework for identifying gaps in existing user data protection and highlighting areas for potential improvements. A comparative view of US and EU data privacy laws is presented in Multimedia Appendix 1 [19-37].

Although many apps require users to input identifiable information, certain users are at risk when such information is misused or not properly stored, secured, or shared. With reproductive health apps in the United States, the sensitive information required from users, coupled with recent legislative changes such as the overruling of *Roe v Wade*, underscore the importance of assessing the privacy policies and practices of sexual and reproductive health apps. So far, limited studies have analyzed the data privacy policies of these apps and addressed the safety issues associated with insufficient user protection. Therefore, this study aimed to (1) evaluate popular reproductive health apps and their respective data privacy policies, (2) compare data privacy laws governing these apps in the United States and the EU, and (3) recommend best practices for users and app developers to ensure user data safety.

Methods

Study Design

We conducted a qualitative content analysis between January 2023 and July 2023 on reproductive health apps available on the Apple App Store and Google Play Store and their publicly available privacy policies. The Transparency, Health Content, Excellent Technical Content, Security/Privacy, Usability, Subjective (THESIS) evaluation tool was used to assess the privacy policies and user agreements of the apps and compare the data privacy laws governing these apps in the United States and the EU [38].

Ethical Considerations

This project did not involve human participants, and the City University of New York Graduate School of Public Health and Health Policy determined that it is not considered human participant research and, therefore, an institutional review board review was not required.

App Selection

To identify apps that would offer the most comprehensive insights into data protection, we conducted a Google search for the leading apps available on both the Apple App Store and Google Play Store. The most commonly mentioned and popular reproductive health apps included Clue [39], Flo [40], and Period Tracker by GP Apps [41]. These selections were informed by recommendations from the *Women's Health* magazine, which based its choices on obstetrician and gynecologist recommendations, and Healthline, which considered customer ratings and reviews, price, availability and accessibility, tools, and features [42,43]. Furthermore, we included a convenience selection with Stardust, which experienced a surge in users and became one of the most downloaded reproductive health apps following the *Roe v Wade* overruling due to its promise of data security using end-to-end encryption [44]. These 4 apps also appeared in the top 15 search results in both app stores based on their consumer ratings, reviews, and download numbers. Consumer ratings, reviews, and app downloads determine the popularity and acceptability of and satisfaction with the apps [45]. Furthermore, these 4 apps were specifically chosen because they provide a good understanding of how policies differ between the United States and the EU and, subsequently, how that affects app data use practices in the 2 regions.

The term *reproductive health app* is used consistently throughout this paper to refer to this category of mobile apps. The selected apps, along with others cited in previous sources, allow users to not only track menstruation cycles but also monitor pregnancy and sexual activity. The data stored on these apps are more sensitive than those stored on other mHealth apps as they often include email addresses, age, name, and date of birth. This sensitivity increases the risk to users if the data are shared or improperly stored.

Data were gathered from the Google Play Store and Apple App Store platforms for each app. App characteristics, including reviews, content rating, and update frequency, were evaluated on both iOS (Apple) and Android systems to identify any differences (Table 1). Notably, certain information is platform specific; for instance, the number of downloads is available only for Android users, whereas app rankings are accessible only to iOS users. This assessment aimed to determine whether apps regularly update their practices in response to the evolving landscape of privacy laws. App characteristics were initially evaluated in May 2023 and updated again in July 2023. It should be noted that these characteristics are subject to change.

XSL•FO
RenderX

Table 1. Characteristics of the Clue, Flo, Period Tracker by GP Apps, and Stardust [46] reproductive health apps evaluated on the Google Play Store and Apple App Store in July 2023.

	Clue	Flo	Period Tracker by GP Apps	Stardust
App developer	BioWink GmbH	Flo Health UK Limited	GP Apps	Stardust App LLC
Country of origin	EU[a] (DE[b])	EU (United Kingdom)	United States	United States
Apple App Store				
Price	Free (additional pricing for in-app purchases)	Free (additional pricing for in-app purchases)	Free (additional pricing for in-app purchases)	Free (additional pricing for in-app purchases)
Release date	—[c]	—	—	—
Number of downloads as of July 2023	—	—	—	—
User ratings as of July 2023	4.8 out of 5 (343,848 reviews)	4.8 out of 5 (1,114,759 reviews)	4.8 out of 5 (61,896 reviews)	4.4 out of 5 (17,639 reviews)
Primary category or genre	Number 96 in "Health & Fitness" category	Number 7 in "Health & Fitness" category	Number 118 in "Health & Fitness" category	Number 195 in "Health & Fitness" category
Last updated as of July 2023	July 20, 2023	July 21, 2023	June 2022 (specific date unavailable)	July 19, 2023
Content (age) rating	12+	12+	12+	4+
In-app purchasing	Yes	Yes	Yes	Yes
In-app advertisements	Yes	Yes	Yes	Yes
Google Play Store				
Release date	October 9, 2014	April 12, 2016	October 7, 2011	June 27, 2022
Price	Free (additional pricing for in-app purchases)	Free (additional pricing for in-app purchases)	Free (additional pricing for in-app purchases)	Free (additional pricing for in-app purchases)
Number of downloads as of June 2023	≥10 million	≥100 million	≥10 million	≥50,000
User ratings as of July 2023	4.4 out of 5 (1.2 million reviews)	4.6 out of 5 (3.12 million reviews)	4.6 out of 5 (360,000 reviews)	2.9 out of 5 (26,400 reviews)
Primary category or genre	"Health & Fitness" category: specific ranking not available	10th highest-grossing app in "Health & Fitness" category	"Health & Fitness" category: specific ranking not available	"Health & Fitness" category: specific ranking not available
Last updated as of July 2023	July 24, 2023	July 26, 2023	September 19, 2022	July 20, 2023
Content (age) rating	Everyone	Everyone	Everyone	Everyone
In-app purchasing	Yes	Yes	Yes	Yes
In-app advertisements	Yes	Yes	Yes	Yes

[a]EU: European Union.

[b]DE: Germany.

[c]Not available.

Downloading, Registering, and Using

In total, 2 researchers (NZ and KSW) downloaded and registered on all 4 apps (Clue, Flo, Period Tracker by GP Apps, and Stardust) using both an iOS and Android system between May 2023 and July 2023. All apps had a "log in" and "create an account" option that used an email and password or an existing Google or Facebook account and a passive or explicit agreement with the privacy policy and terms of service. The passive agreements had language such as the following—"by creating an account you are agreeing..."—whereas explicit agreements required a button to select "agree" before continuing with the registration. These steps can be observed in Figure 1. Enrollment flow varied between apps, with common information required for signing up being email, age, date of birth, and phone number.

XSL•FO
RenderX

Figure 1. Visualization of the reproductive health app selection and the registration on each app between May 2023 and July 2023.

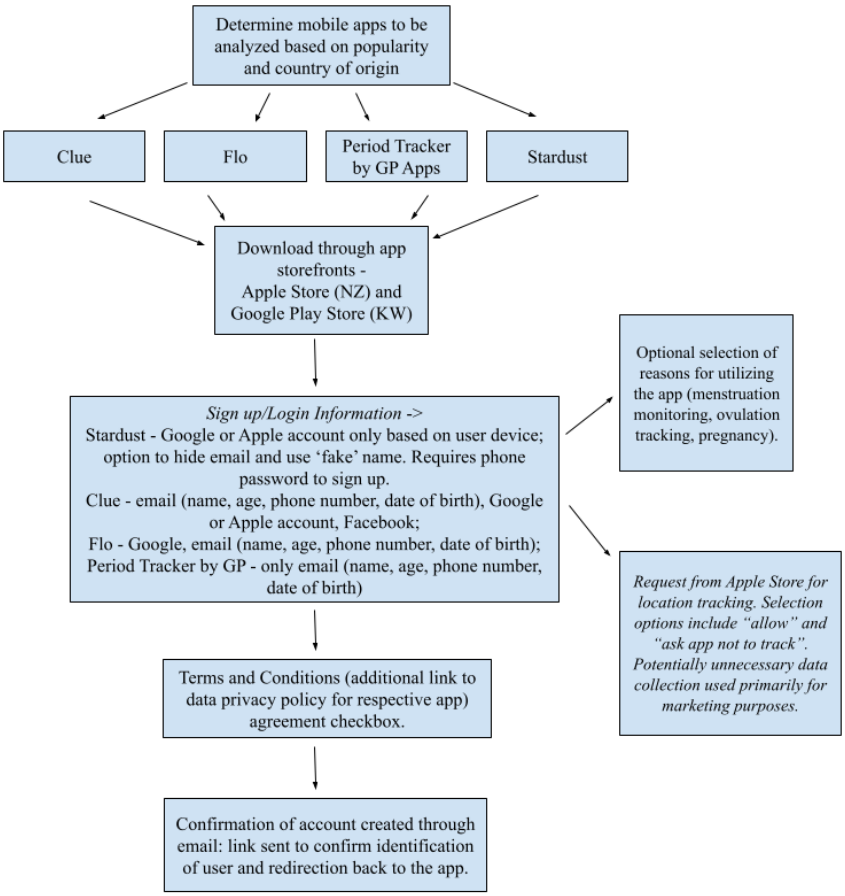

Data Analysis: THESIS Evaluation Tool

To properly assess the reproductive health apps chosen, particularly regarding their privacy and security, it was important to find an evaluation tool that addressed these qualities. Unlike the commonly used Mobile App Rating Scale assessment tool that rates apps based on subscales, items, descriptors, and anchors, the THESIS evaluation tool dedicates one of its domains specifically to app privacy and security [47]. The THESIS evaluation tool consists of 6 domains on which an app is rated on a scale from 1, the lowest, to 5, the highest: transparency, health content, technical content, security and privacy, usability, and subjective rating. Due to the objective of this research being the security of user data, the security and privacy domain, consisting of 6 subcategories, was the part of the evaluation tool used. The subcategories and their rankings are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Subcategories of the security and privacy domain of the transparency, health content, excellent technical content, security and privacy, usability, and subjective evaluation tool and their rankings.

- Protection against theft and viruses: does the app follow best practices in security with optimal antivirus and safeguards against breaches? This is found in the privacy policy of an app.
- Authentication: is the authentication procedure optimal? User experience is based on receiving an email or SMS text message notifications for creating an account, in addition to password protection for accessing the user's profile.
- Data sharing: when sharing information, does the app use best practices? This is found in the privacy policy of an app.
- Maintenance: does the app have regular cycles to update and patch its security? This is found in the Apple App Store or Google Play Store under "version history" for each app—last update occurred during the month of rating or the month before (the update schedule is very consistent; rating of 5), last update occurred 2 months before the time of rating (the update schedule is generally consistent; rating of 4), last update occurred between 3 and 5 months before the time of rating (the update schedule is a little inconsistent; rating of 3), last update occurred between 6 months and a year before the time of rating (the update schedule is completely inconsistent; rating of 2), and last update occurred >1 year before the time of rating (rating of 1).
- Signaling of breaches: if a breach occurs, does the app have a method to notify its users? This is found in the privacy policy or the terms and conditions of an app.
- Anonymization: does the app appropriately anonymize individuals? This is found in the privacy policy or the terms and conditions of an app [38].

  To note, the Federal Trade Commission (FTC) has expressed concern regarding the legitimacy of data anonymization as promised by multiple apps, not just reproductive health apps, calling claims as potentially deceptive trade practice and violating the FTC Act when untrue. Significant research has shown that anonymized data can often be reidentified, especially in the context of location data. One set of researchers demonstrated that, in some instances, it was possible to uniquely identify 95% of a dataset of 1.5 million individuals using 4 location points with time stamps [48].

The evaluation tool includes 2 additional columns dedicated to rationalizing the pros and cons of specific ratings. The Why Not −1 column indicates why points were not deducted from the app's score, whereas the Why Not +1 column indicates why points were deducted from the final score.

Results

Reproductive Health App Privacy Policy Summaries

With 1.8 million mobile apps available on the Apple App Store and 2 million available on the Google Play Store, both platforms have released their own guidelines for developers seeking to launch their apps [49,50]. The Apple App Review Guidelines have a specific section dedicated to privacy policy requirements, stating that all apps must provide their users with a privacy policy either within the app or as a link when developing for iOS. These privacy policies must "identify what data, if any, the app or service collects, how it collects and uses that data, confirm that any third party with whom an app shares user data (in compliance with these Guidelines)—such as analytics tools, advertising networks and third-party SDKs, as well as any parent, subsidiary or other related entities that will have access to user data—will provide the same or equal protection of user data as stated in the app's privacy policy and required by these Guidelines, and explain its data retention/deletion policies and describe how a user can revoke consent and/or request deletion of the user's data" [51]. These guidelines also include more information on secure user consent, data minimization, and data collection. The Google Play Store guidelines are similar, posted on the Google Play Console Help Center, and require developers to state how private user data are collected, protected, and handled. However, the expansion of the privacy policy requirement to all mobile apps released through the Google Play Store is recent, as of 2022 [52]. In addition to this, in July 2021, the Google Play Store began requiring apps to fill out a data safety form that needs to align with the apps' actual data use. That being said, the only Google Play Store requirements are that information such as where the data are sold to and how they are shared be stated in the data safety form. On the basis of this information, Google does not clearly state its concern for the privacy of its users. The same pattern was observed on the Apple App Store, with the only difference being Apple's location tracking option that pops up with every app downloaded. Because of these guidelines, the 4 reproductive health apps evaluated have their own data privacy policies separate from their terms and conditions that must be agreed to before use of the app.

Multimedia Appendix 2 [39-41,46] shows the analysis of each reproductive health app regarding its privacy policy, the type of data collected, how they are protected (varying types of encryption vs none at all to store sensitive data and protect them from being stolen, changed, or compromised), who they are shared with, what options are available for users seeking to delete or opt out of data sharing, and potential concerns that remain. In addition to this, a column was dedicated to assessing the accessibility of privacy policies to users, which often affects the probability of users reading through them rather than simply clicking on Agree to access the app functions, otherwise known as the clickwrap approach. Furthermore, both anonymization and end-to-end encryption are included in one section because they are used with differing data, such as anonymization for metadata such as IP addresses and end-to-end encryption for data input into the app by the user, and, thus, vary in data protection. Assessment of these apps was first conducted at the beginning of the research in November 2022 and then updated twice, once in May 2023 and again in July 2023 for the most up-to-date assessment. The final column of the table addresses concerns related to the methodology used by each app.

XSL•FO
RenderX

THESIS Evaluation Tool

After analysis of individual app policies, the THESIS evaluation tool allowed for a thorough comparison of the security and privacy of the 4 mobile apps selected. The tool was first used to assess the EU apps, Clue (Table 2) and Flo (Table 3), with Clue being the first due to its earlier release date, followed by the assessment of the US app Period Tracker by GP Apps first due to its earlier release date and then of Stardust as that is the most recently released app evaluated. This chronological order allowed for a better understanding of the changes made in the app privacy policies with the changing landscape of reproductive health rights. Scoring was based on an initial analysis of the apps in May 2023 that was repeated in July 2023 to see whether any changes were made as more state privacy laws were introduced. A description is provided for every score to explain the reasoning behind each grade. The subcategories are those of the THESIS evaluation tool and its score guide.

Table 2. Transparency, health content, excellent technical content, security and privacy, usability, and subjective evaluation of Clue, a European Union (EU) app that falls under General Data Protection Regulation protections, based on an initial analysis of the app in May 2023 repeated in July 2023. Score: 3.5 for the security and privacy domain.

Subcategory	Score	Why not –1	Why not +1
Protection against theft and viruses	1	—[a]	Not enough information available anywhere in policies or documents
Authentication	4	For verification, the app sends an email to the address used for registration, and the user is unable to input information into the app calendar without this verification; PIN[b], Touch ID (iPhone 5S-8), or Face ID (iPhone X-13) authentication is available and is intended to automatically encrypt Clue data and prevent someone else from accessing users' information through a user's device	Information on this feature is not explicitly stated; requires a look through the privacy policy on the app's website.
Data sharing	3	Data-sharing practices seem to be more transparent than those observed on other apps. The creators make it a point to state that privacy is one of their top priorities. It also follows guidelines of EU data privacy laws, which are said to be the strictest in the world.	However, Apple has its own ranking of data based on the analytics the app creators provide them with. This portion is contradictory as it states the following: "Data linked to you may be collected and linked to your identity: contact info, health and fitness, location, etc" [53]. Data are also stored, which poses threats if subpoenaed.
Maintenance	5	The app is updated almost every week. The last update occurred 5 days before rating.	—
Signaling of breaches	1	—	Not enough information available anywhere in policies or documents
Anonymization	3	The app states not to use any identifiable information of the users for any purposes, including data analyses. Encryption is used.	IP address collection poses a threat to user safety. HTTP is not the safest encryption method and "the data that is transferred over an HTTP connection is not encrypted, so you run the risk of third-party attackers stealing the information" [54].

[a]None.

[b]PIN: personal identification number.

XSL•FO
RenderX

Table 3. Transparency, health content, excellent technical content, security and privacy, usability, and subjective evaluation of the Flo app data use practices in July 2023. Score: 3.5 for the security and privacy domain.

Subcategory	Score	Why not –1	Why not +1
Protection against theft and viruses	4	General information provided, such as an email for the data protection officer. The website mentions vulnerability testing and system monitoring or alerting.	Difficult to test accuracy in real-life settings unless issues arise.
Authentication	4	The app sends an email to the address used for signing up for verification, and the user is unable to input information into the app calendar without this verification. There is an option to password protect the app to enable secure access.	Secure access is not a pre–set up option; users have to know how to access it and turn it on. This is only mentioned in the app's web page rather than being originally included in the app.
Data sharing	2	At first glance, it looks good, and data are not said to be sold to other companies. A user's cycle can be linked to other apps. There is an emphasis on sharing rather than selling. The data stored are also encrypted.	It is said that personal data may be processed in relation to the company's interests in providing services, commercial interests, and "wider societal benefits." The app connects with AppsFlyer for marketing and promotions and requires consent, but consent is required to make an account regardless. This involves a lot of technical identifiers. The "security of personal data" section is contradictory as it says that the data are encrypted [40]. It is important to note that encryption does not mean that data are not accessible by the company doing the encryption. It just means that the data do not reside on their servers in plain-text format. Use of the term "encrypted" gives the impression that the company cannot see the data, whereas in reality, it just means that the data are not stored in a plain-text format.
Maintenance	5	The app is updated almost every week. The last update occurred 5 days before rating.	—[a]
Signaling of breaches	4	There is a separate section in the policy for security breaches and what steps the app will take if that occurs. A notification is said to be sent out, and the app will take steps such as deleting user information, resetting passwords, and logging out of user devices if such a breach occurs.	Does not seem to be responsible for third-party data breaches (although it states otherwise) given a previous data breach with third-party Facebook Analytics tool despite existing safeguards.
Anonymization	3	The app states not to sell any identifiable information. The app also gives the option to be used in anonymous mode. There is an option to not allow the app to track Apple Health or Fitbit statistics. All data are encrypted using KMS[b] [55].	IP address and unique device identifiers (IDFA[c]) are collected. The Apple App Store also has contradictory information saying that data are linked to a user, such as location, contact information, identifiers, user content, and sensitive information.

[a]None.

[b]KMS: Key Management Service.

[c]IDFA: identifier for advertisers.

Flo users in the United States are subject to US laws, but the company must also comply with EU laws for EU-based users. In Table 3, scoring was based on an initial analysis of the app that was repeated 2 months later to see whether any changes were made as more state privacy laws were introduced. A description is provided for every score to explain the reasoning behind each grade. The subcategories are those of the THESIS evaluation tool and its score guide. A major concern is the app's use of IP address tracking as well as the use of data by third parties for advertising and marketing.

Period Tracker by GP Apps (Table 4) is a US-based app potentially required to adhere to the General Data Protection Regulation because of its European users. Scoring was based on an initial analysis of the app in May 2023 that was repeated in July 2023 to assess changes as more state privacy laws were introduced. A description is provided for every score to explain the reasoning behind each grade. The subcategories are those of the THESIS evaluation tool and its score guide, as seen in the previous 2 evaluations.

XSL•FO
RenderX

Table 4. Transparency, health content, excellent technical content, security and privacy, usability, and subjective evaluation of Period Tracker by GP Apps' data use practices in July 2023. Score: 2 for the security and privacy domain.

Subcategory	Score	Why not –1	Why not +1
Protection against theft and viruses	2	Some security is listed on the app's website, such as limiting access to information to authorized employees and contractors.	Does not seem to be held accountable for any breaches or collection of data that may occur.
Authentication	3	Option to turn on finger ID or passcode is available.	It is not clearly stated where this option is and how to actually use it.
Data sharing	2	—[a]	Limited information available, which makes it seem even more as though the app does not want to disclose its privacy and data sharing information. Nothing about unidentifiable information or how sharing occurs is noted.
Maintenance	2	—	Last update and bug fix was 6 months before rating. Seems to be an update pattern based on the version history available on the app store.
Signaling of breaches	2	It does say that users will be informed of any changes occurring within the system; this may mean breaches as well.	No information on breaches is available other than the app stating that they are possible and they are not responsible.
Anonymization	2	Option for using without an account is possible but extremely limited.	Nothing else is noted about anonymization from the app itself. However, the app store does say that one can disable their information being used by third parties, but this is written in tiny letters on the website.

[a]None.

The final app evaluated was Stardust (Table 5), an app that underwent internal data storage and use and account creation modifications as a response to *Dobbs v Jackson* and the wave of worry among menstruating persons using reproductive health apps [56]. Stardust was purposely chosen because of its recent emergence and emphasis on user data protection and different account creation practices. Stardust emphasizes end-to-end encryption and minimal data collection. Table 5 scoring was based on an initial analysis of the app in May 2023 that was repeated in July 2023.

Table 5. Transparency, health content, excellent technical content, security and privacy, usability, and subjective evaluation of the Stardust app's data use practices in July 2023. Score: 4.5 for the security and privacy domain.

Subcategory	Score	Why not –1	Why not +1
Protection against theft and viruses	3	Some information regarding theft was provided.	General statement: "We take all reasonable and appropriate measures (including encryptions) to protect all collected Personal Data from loss, theft, misuse, and unauthorized access" [46].
Authentication	5	The app sends an email to the address used for signing up for verification, and the user is unable to input information into the app calendar without this verification. Uses Rownd, a privacy-first authentication platform.	—[a]
Data sharing	4	Data-sharing practices seem to be more transparent than those observed on other apps, closer to what is observed in the EU[b]. Data are said to be encrypted.	However, data are still said to be potentially provided if required by government agencies. It is not clear what that includes.
Maintenance	5	The app is updated almost every week. The last update was 1 day before rating.	—
Signaling of breaches	1	—	No information available anywhere in policies or documents
Anonymization	5	End-to-end encryption that is only accessible to the user. The policy states to also not track or use IP addresses. Even if the government asks for the information, it should not be linked to a specific user. Users are also able to come up with their own username and are not required to use their legal name.	—

[a]None.

[b]EU: European Union.

XSL•FO
RenderX

App Store Information Discrepancies

Another important aspect to consider are the discrepancies observed between data collection information on the Apple App Store and on the apps themselves. Despite app promises of data privacy, there are notable differences between their respective privacy policies and what is stated in the Apple App Store. As shown in Figure 2 [57], the Apple App Store included its own description of potential data collection based on each app. While Flo has a history of violating the privacy policy promised to its users, the data that may be collected as listed by the Apple App Store are different from those listed in the app's privacy policy.

Figure 2. Apple App Store description of Flo data collection and sharing.

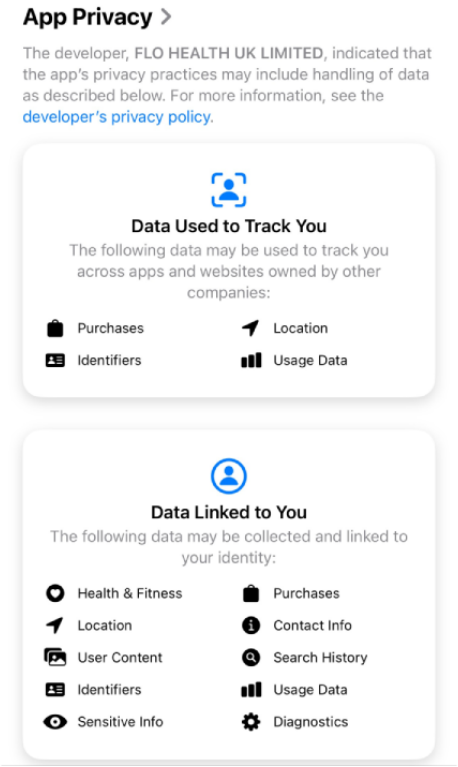

This is in comparison to the app privacy summary of Stardust, which more closely aligns with the information provided in the app's own privacy policy.

Such discrepancies make it difficult for users to determine what data are being collected and shared and, thus, cause confusion when deciding what reproductive health app should be used. These discrepancies are also reflected in Figures 2 and 3 analyzing the apps, where contradictory information alters app ratings.

Figure 3. Apple App Store description of Stardust data collection and sharing for comparison with that of Flo [39].

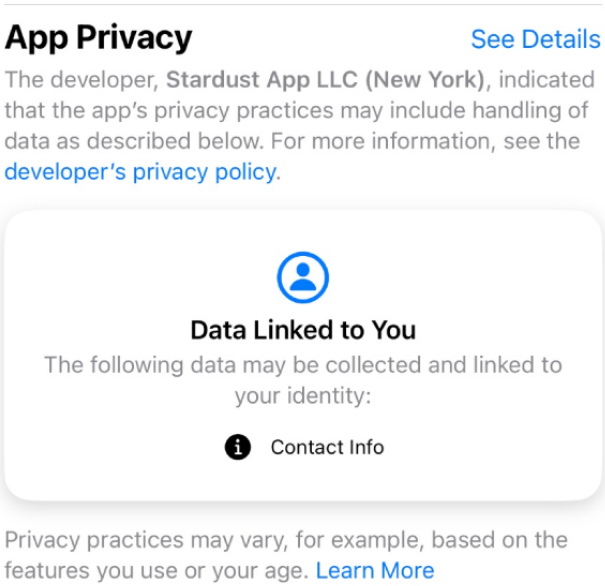

XSL•FO
RenderX

In addition, another notable discrepancy is the content rating for the apps as observed on the Apple App Store and Google Play Store. While 75% (3/4) of the apps have ratings of 12+ on the Apple App Store, all 4 apps are rated E for Everyone on the Google Play Store. Given the content of the apps and the sensitivity of the data provided, the Apple App Store rating should be considered the more suitable one. It is interesting to note that a statement on the Google Play Store mentions that "content ratings on Google Play are provided by the International Age Rating Coalition (IARC) and are designed to help developers communicate locally relevant content ratings to users" and, thus, makes it difficult to understand what causes the discrepancy in content rating on both platforms [59].

Discussion

Principal Findings

This study evaluated popular reproductive health apps and their respective data privacy policies, compared data privacy laws governing these apps in the United States and the EU, and recommended best practices for users and app developers to ensure user data safety. The characteristics of reproductive health app data privacy, safety, and use were indicated. Using the THESIS evaluation tool, the privacy and security features of 4 reproductive health apps were identified, and the determined discrepancies in these features were described in the previous sections. Although there are successes and good practices in app user data management, potential concerns remain. Adopting best practices from multiple sources, such as end-to-end encryption, can help enhance the privacy and security of user data on reproductive health apps.

Potential Concerns

A careful evaluation of the 4 selected reproductive health apps identified Stardust as the most privacy-centered app at the time of this qualitative content analysis. It should be noted that policies and data use practices are continually evolving as apps adapt to new requirements and guidelines. Although Stardust has lower ratings on both the Google Play Store and Apple App Store, which is attributed to its recent release and interface issues, it provides the most comprehensive data protection to its users. This is due to its limited data collection, such as excluding IP tracking; transparent privacy policy; specified encryption methods; options for deleting previously shared data; and the use of third-party platforms that assess deidentified data in a secure manner. While Clue, Flo, and Period Tracker by GP Apps collect IP address data, Stardust focuses on location data, such as country and time zone data, which minimizes the likelihood of connecting a user to a specific location and any reproductive health information. Furthermore, the app is transparent about the third parties it engages with and how user data are used. Although Flo and Clue use encryption to protect the confidentiality of the digital data stored across various platforms, the combination of IP address collection and the extensive use of third-party data brokers for analytics may pose unique risks to users, especially when linked with sensitive information such as missed periods and unique device identifiers [18].

When comparing the privacy and security of the data collected on all 4 apps—Clue, Flo, Period Tracker by GP Apps, and Stardust—determining what factors make one reproductive health app during this time a better and safer choice for users than others was critical. Although existing privacy policies attempt to combat the easy accessibility of user data and ensure stricter regulation on data sharing and breaches, the ultimate responsibility to ensure the safety of reproductive health information and app users falls on the app developers themselves. In this assessment, these factors included how accessible the privacy policy of each app is; what data are collected and how they are made secure; which, if applicable, third parties data are shared with and for what purposes; and the ease with which users may have their data deleted and opt out of future data collection.

In addition to this, a common trend among all apps was the questionable access to privacy policies and user agreements that does little to assure and encourage users to read through these agreements rather than selecting "Accept" just to gain access to the apps' functions.

Despite identified potential concerns, for users in regions without strict abortion regulations, Clue and Flo may be safe choices considering that IP address tracking alone, without any other identifiable data-point tracking as the apps promise, should not pose a significant threat. For full protection, users should create anonymous accounts, as facilitated in Flo and anticipated in Stardust, and enable password protection where available. Most importantly, users are advised to avoid reproductive health apps such as Period Tracker by GP Apps due to its infrequent system updates, lack of accountability, inadequate encryption in data collection and sharing, and insufficient transparency in their privacy policies. Using an app that does little to protect individual users may make it prone to data breaches and risks data use for things that the user is not aware of, such as during subpoenas or other investigations.

Comparison With Other Work

The findings of this study have also been supported by previous data privacy research, systematic reviews, and scoping reviews conducted in relation to numerous mHealth apps, including mental health, diabetes management, and physical activity tracking. Although they used different analysis methods from the THESIS evaluation tool, Benjumea et al [60] also concluded that the re-evaluation of the data collected on these apps, stated as "minimization" by the General Data Protection Regulation, is necessary to determine whether the scope of the personal information collected on these apps is beneficial to the users or may be limited to protect user data. In addition, the research conducted by Alfawzan et al [61] involving the 23 most popular mHealth apps for menstruating persons emphasized the threats associated with third-party data sharing, determining that 87% of the apps assessed were found to partake in such sharing and stating that "user consent, especially about sharing data in general or with a third party, is a concern for women's privacy...research indicates that end users do not have full awareness of what their consent entails." Furthermore, key findings of the analysis of reproductive health apps, including Flo, in the article "In Post Roe v Wade Era, Mozilla Labels 18

XSL•FO
RenderX

of 25 Popular Period and Pregnancy Tracking Tech With *Privacy Not Included Warning" by the Mozilla Foundation concluded that there is no clear stand on data sharing with law enforcement; password requirements are especially low; and there is an absurd amount of personal information being collected ranging from device IDs and IP addresses to pregnancy symptoms and physicians' appointments [62].

Other articles, such as "Mobile Health and Fitness Apps: What Are the Privacy Risks" by Privacy Rights Clearinghouse, outline discussed concerns about data collection for advertising purposes, stating that "mobile applications, especially apps that you download for free, depend on advertising to make money...they may share personally identifiable information with advertisers, or allow ad networks to track you...if an application collects your universal device ID (UDID) or embeds a unique ID in the application you download, analytics data can be traced back to you personally" [63]. This concept closely relates to the accusations previously made against Flo regarding their user data sharing between 2016 and 2019.

Importance of Encryption and Real-Life Application

Following the overruling of *Roe v Wade*, major reproductive health apps such as those assessed in this study have taken it upon themselves to work toward a safer way for menstruating persons to track, monitor, and maintain their reproductive health. With the introduction of end-to-end encryption in the late 90s, these apps have adopted a way to provide users with data privacy while also collecting enough user analytics to better their own apps [64]. Such encryption methods, as observed in Stardust, are meant to ensure that sensitive data such as personal health information are protected from unauthorized disclosure. "In [end-to-end encryption], the data is encrypted on the sender's system or device, and only the intended recipient can decrypt it. As it travels to its destination, the message cannot be read or tampered with by an internet service provider (ISP), application service provider, hacker or any other entity or service" [65].

One of the issues that only Stardust addressed is the collection of IP addresses and location tracking. Despite end-to-end encryption dispersing and disconnecting user data from user profiles, IP location tracking and mobile ID tracking leads to the risk of connecting deidentified user information with specific locations and phone use. For example, if a user is located in a US state that has banned abortions, such as Alabama, Idaho, Tennessee, and Texas, there is a risk that state officials could obtain personal data that reveal pregnancy or potential termination of pregnancy based on period or symptom tracking [66-68]. In addition to this, the location data tracked through these apps could also reveal a user's visits to abortion clinics via IP addresses. If state agencies receive a subpoena, they may obtain this information regarding users and the medical professionals providing banned procedures. In states where abortions are illegal, lack of proper data protection raises implications that could endanger users after the overruling of *Roe v Wade*, especially among those with historically inequitable health situations [69]. Increased risk of government officials legally obtaining personal information, even IP addresses that may directly link data points to an individual user, due to these laws has led to the assessment of the data safety provided by individual apps and what steps may be taken from the users' and the developers' side to ensure future security. In addition to this, some companies such as Flo and Clue have developed the option for users to create passwords to protect their app accounts through encryption methods that ensure access to the app only through the user's phone. Analyzing the proper steps that Flo, Clue, and Stardust have taken to heighten their users' safety allowed for the assessment of data privacy lapses in Period Tracker by GP Apps.

Prioritizing Marketing Over User Protection

Evaluation of privacy policies also shed light on what role data collection plays apart from direct user experience improvement. For all the apps except Stardust, it is explicitly stated that some data are shared with third-party apps for marketing purposes, such as reaching potential users based on current user preferences. The issues associated with this are similar to what happened with Flo sharing its user information with Facebook Analytics services that resulted in a data breach. Flo's statement on this data breach allows for insights on what user data have been used for recently—insights from Facebook Analytics found "that it's challenging for [Flo] users to log childbirth and enter postpartum mode...This finding was driven by the insights from analytical systems, and with its help [Flo] managed to improve user experience and stressless" [70]. However, many, including the FTC, have speculated "...that Flo's data-sharing practices had allowed third-party companies to use said personal health information expansively, including for advertising..." [71]. Apart from Flo, apps such as Clue also state that they use data for promotional purposes. With the sensitivity of the data at hand on these apps, reproductive health app developers may consider limiting their marketing resources and turning to internal analytics without the use of third parties unless third parties have their own practical user privacy tools that may be trusted and proper business associate agreements are in place.

Recommendations

With the thorough analysis of the 4 most popular reproductive health apps available on both the Apple App Store and Google Play Store, it is possible to make recommendations for app developers, major app platforms, and individuals looking to use reproductive health apps.

First, to ensure user data safety, developers should prioritize the access to and reading ease of privacy policies. While users are currently required to agree to the terms and conditions of reproductive health apps before the use of the apps' features, there is no guarantee that these policies are actually being read and not just bypassed. Although court cases such as *Bassett v Electronic Arts Inc* currently require terms and conditions to be clearly visible to users (ie, not hidden by fonts or an unclear location of the link), all the apps evaluated used the clickwrap approach for user privacy policy consent [67,72]. To make it more probable that users are actively reading the conditions they are agreeing to, using a scrollwrap approach, one that requires users to scroll through the entire policy agreement before checking off an "I accept" button that becomes active only once the scrolling is completed, is preferable [73]. With this method, app privacy policy writers should opt to highlight the most important aspects of the agreement in layman's terms

XSL•FO
RenderX

to capture the user's attention. While this does not guarantee that users are actively reading what they are agreeing to, if information is highlighted, it may catch users' attention better. App developers may also consider shifting to preset anonymous modes of user profiles, as observed as an option in Flo, as well as requiring virtual private network use to access the app functions to limit the chances of data being specifically linked to individual users.

Second, the evaluation of previously enacted policies also requires adaptation to the ever-changing field of technology that poses new threats to mHealth app users.

Third, users should be encouraged to make educated decisions when selecting reproductive health apps for personal tracking, monitoring, and management. Users should be aware of what data are collected; where those data go; and what risks arise if their personal information is shared, handled, or stored differently than what is expected or promised. This could be made possible by major app platforms such as the Apple App Store and Google Play Store closely monitoring the collection of data on reproductive health apps and making sure that apps are constantly updating their users on changes to data privacy settings, as well as adhering to data-sharing restrictions stated in individual app policies. Platforms should also consider altering their rating criteria, making data privacy one of the most crucial qualities of a highly rated app and, thus, increasing user health literacy and guiding them to the secure apps.

Finally, end-to-end encryption and data anonymization methods should both be used to protect not only the reproductive health data that users have provided to the apps but also the data collected by the apps themselves, including device IDs and IP addresses. Such practices would further ensure that deidentified personal health information may not be linked to a specific user while also providing developers with enough user analytics to continue updating the app's functionality. Furthermore, it is vital for apps to follow anonymization rules set out by the FTC and other regulatory systems rather than make false promises that do little to protect user data.

Strengths and Weaknesses

While many other studies have assessed the data protection aspects of reproductive health apps, this is one of the first that takes a comprehensive approach to data privacy, from the analysis of existing data privacy policies to individual encryption methods. In addition to this, having more than one researcher assess each individual app allowed for limitation of bias and increase in reliability and validity. However, there are a few weaknesses concerning this study. These include the lack of generalizability as only 4 apps were assessed, whereas there are >25 reproductive health apps on both the Google Play Store and Apple App Store platforms; the constant updates of Clue, Flo, and Stardust that change the way in which each protect, collect, and share data; and the ever-changing field of data privacy, with company statements and policies changing since the beginning of this study in 2022. Furthermore, a limitation is the convenience sampling conducted through which the 4 apps were chosen as there was no systematic search of apps conducted apart from that of country of origin for the 2 EU apps listed. Finally, what was true last year regarding these apps may not be true today. With the changing landscape of data privacy policies and federal and local governments changing platform requirements, these apps may have taken steps to further protect user data. In addition, most apps are constantly updating their own policies and interfaces.

Conclusions

As stated by Ritesh Kotak, a Toronto cybersecurity and technology analyst quoted in CBC Canada's "Americans are being urged to delete period tracking apps. Should Canadians do the same?" article, "When you downloaded that app, how much did you pay for it? What's your monthly subscription fee? If the answer is zero, if you're not paying for the product, then you are the product" [68]. Moving forward in a technology-based society heightens the risk of information sharing and data security violations, which becomes of increased concern when combined with reproductive health changes within the Supreme Court and local government agencies. Because of this, it is crucial for the landscape of data privacy protection to reflect the changes being made and prioritize the safety and protection of users. More studies emphasizing the use of data collected by reproductive health apps and long-term threats caused by the sharing of such data, in addition to lapses in outdated policies in place, will further assist users in making smart app selection decisions and encourage app developers to assert user data protection as their top priority. Reproductive health apps are meant to function as resources available to the general population designed to make their day-to-day lives and activities simpler; they are not meant to be guidelines for maternal or menstruating health and, thus, should be treated as such. That is why limiting the data collected by these apps and taking precautions to ensure data privacy is in the best interest of users and app developers.

Acknowledgments

The authors would like to acknowledge the City University of New York Graduate School of Public Health and Health Policy for encouraging and supporting this research.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

XSL•FO
RenderX

Conflicts of Interest

KSW is an Associate Editor of *Journal of Medical Internet Research*, but played no role in the editorial handling or peer review process of this manuscript. All other authors have no conflicts to declare.

Multimedia Appendix 1
A breakdown of existing policies in the United States and the European Union's General Data Protection Regulation, in addition to their respective issues regarding reproductive health apps, organized in chronological order before conducting the app analysis in May 2023 for better understanding of how policy making has changed and shifted as technology becomes more prevalent.
[DOCX File , 15 KB - jmir_v27i1e51517_app1.docx]

Multimedia Appendix 2
A description of the privacy and data use practices of reproductive health apps from their respective data privacy policies collected in July 2023.
[DOCX File , 18 KB - jmir_v27i1e51517_app2.docx]

References

1. The triangle shirtwaist factory fire. United States Department of Labor. URL: https://www.osha.gov/aboutosha/40-years/trianglefactoryfire [accessed 2023-07-25]
2. Certiorari to the United States court of appeals for the fifth circuit. Supreme Court of the United States. URL: https://www.supremecourt.gov/opinions/21pdf/19-1392_6j37.pdf [accessed 2023-07-25]
3. Roe v. Wade. Center for Reproductive Rights. URL: https://reproductiverights.org/roe-v-wade/ [accessed 2023-07-25]
4. Tucker S. Welcome to the world of mHealth!. Mhealth 2015;1:1 [FREE Full text] [doi: 10.3978/j.issn.2306-9740.2015.02.01] [Medline: 28293561]
5. Derbyshire E, Dancey D. Smartphone medical applications for women's health: what is the evidence-base and feedback? Int J Telemed Appl 2013;2013:782074 [FREE Full text] [doi: 10.1155/2013/782074] [Medline: 24454354]
6. Number of mobile app downloads worldwide from 2019 to 2027, by segment (in million downloads). Statista. URL: https://www.statista.com/forecasts/1262881/mobile-app-download-worldwide-by-segment [accessed 2023-07-25]
7. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Content analysis of behavior change techniques in maternal and infant health apps. Transl Behav Med 2021 Mar 16;11(2):504-515. [doi: 10.1093/tbm/ibaa039] [Medline: 32491165]
8. Logie C, Okumu M, Abela H, Wilson D, Narasimhan M. Sexual and reproductive health mobile apps: results from a cross-sectional values and preferences survey to inform World Health Organization normative guidance on self-care interventions. Glob Health Action 2020 Dec 31;13(1):1796346 [FREE Full text] [doi: 10.1080/16549716.2020.1796346] [Medline: 32778000]
9. Broad A, Biswakarma R, Harper JC. A survey of women's experiences of using period tracker applications: attitudes, ovulation prediction and how the accuracy of the app in predicting period start dates affects their feelings and behaviours. Womens Health (Lond) 2022 Apr 25;18:17455057221095246 [FREE Full text] [doi: 10.1177/17455057221095246] [Medline: 35465788]
10. Earle S, Marston HR, Hadley R, Banks D. Use of menstruation and fertility app trackers: a scoping review of the evidence. BMJ Sex Reprod Health 2021 Apr 06;47(2):90-101 [FREE Full text] [doi: 10.1136/bmjsrh-2019-200488] [Medline: 32253280]
11. Garamvolgyi F. Why US women are deleting their period tracking apps. Guardian News & Media Limited. URL: https://www.theguardian.com/world/2022/jun/28/why-us-woman-are-deleting-their-period-tracking-apps [accessed 2023-07-25]
12. Clark L. How worried should you be about period tracking apps? Women's Health. URL: https://www.womenshealthmag.com/uk/health/a45802353/period-tracking-apps-police [accessed 2024-11-08]
13. Torchinsky R. How period tracking apps and data privacy fit into a post-Roe v. Wade climate. National Public Radio. URL: https://www.npr.org/2022/05/10/1097482967/roe-v-wade-supreme-court-abortion-period-apps [accessed 2024-07-25]
14. Female health apps misuse highly sensitive data. UCL News. URL: https://www.ucl.ac.uk/news/2024/may/female-health-apps-misuse-highly-sensitive-data [accessed 2024-05-14]
15. Self-care, criminalized: preliminary findings. If/When/How. URL: https://ifwhenhow.org/resources/self-care-criminalized-august-2022-preliminary-findings/ [accessed 2024-07-24]
16. Digital tools can help improve women's health and promote gender equality, WHO report shows. World Health Organization. URL: https://tinyurl.com/ymses388 [accessed 2024-03-08]
17. Korn J, Duffy C. Search histories, location data, text messages: how personal data could be used to enforce anti-abortion laws. CNN. URL: https://www.cnn.com/2022/06/24/tech/abortion-laws-data-privacy/index.html# [accessed 2022-07-24]
18. Bernstein S. The role of digital privacy in ensuring access to abortion and reproductive health care in Post-Dobbs America. American Bar Associatio. URL: https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/technology-and-the-law/the-role-of-digital-privacy-in-ensuring-access-to-reproductive-health-care/ [accessed 2024-06-03]

XSL•FO
RenderX

19. Guide to privacy and security of health information. Department of Health and Human Services. URL: https://www.healthit.gov/sites/default/files/pdf/privacy/onc_privacy_and_security_v1_022112.pdf [accessed 2023-07-25]
20. What does the fourth amendment mean? United States Courts. URL: https://tinyurl.com/3jkjsvcw [accessed 2023-07-25]
21. How to determine if your product is a medical device. U.S. Food and Drug Administration. URL: https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device [accessed 2023-07-25]
22. Privacy and security. Federal Trade Commission: Protecting America's Consumers. URL: https://www.ftc.gov/business-guidance/privacy-security [accessed 2023-07-25]
23. Wetsman N. Birth control apps show the contradictions in FDA device oversight. The Verge. URL: https://www.theverge.com/22335858/birth-control-app-clue-natural-cycles-fda [accessed 2023-07-25]
24. Overview of governmental action under the stored communications act (SCA). Congressional Research Service. URL: https://crsreports.congress.gov/product/pdf/LSB/LSB10801 [accessed 2023-07-25]
25. Kushner M. Stored communications act: the good, the bad, and the ugly. Emden Law. URL: https://tinyurl.com/nehjr3x6 [accessed 2023-07-25]
26. Children's Online Privacy Protection Rule ("COPPA"). Federal Trade Commission. URL: https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa [accessed 2023-07-25]
27. The state of play - issue brief: COPPA 101 - future of privacy forum. Future of Privacy Forum. URL: https://fpf.org/blog/the-state-of-play-issue-brief-coppa-101/ [accessed 2023-07-25]
28. Ornstein C. Federal patient privacy law does not cover most period-tracking apps. ProPublica. URL: https://www.propublica.org/article/period-app-privacy-hipaa [accessed 2023-07-26]
29. Carpenter v. United States. Supreme Court of the United States. URL: https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf [accessed 2023-07-25]
30. Law Offices of John Wesley Hall. D.R.I.: IP addresses def used weren't like detailed tracking information from CSLI requiring a SW. Casetext. URL: https://tinyurl.com/4s8tn9da [accessed 2023-07-25]
31. Bonta R. The California Consumer Privacy Act (CCPA). State of California - Department of Justice - Office of the Attorney General. URL: https://oag.ca.gov/privacy/ccpa [accessed 2023-07-25]
32. Jercich K. California Consumer Privacy Act can cause headaches for healthcare orgs. Healthcare IT News. URL: https://tinyurl.com/4wycc67v [accessed 2023-07-25]
33. Virginia Consumer Data Protection Act (VCDPA): everything you need to know about Virginia's new comprehensive data privacy law. Bloomberg Law. URL: https://tinyurl.com/4rwhybn3 [accessed 2023-07-25]
34. California consumer privacy laws: understand the key provisions of California's consumer privacy legislation – including enforcement, consumer rights, and how to comply. Bloomberg Law. URL: https://pro.bloomberglaw.com/brief/california-consumer-privacy-laws-ccpa-cpra/ [accessed 2023-07-25]
35. European Union - data privacy and protection. Privacy Shield Framework. URL: https://tinyurl.com/3hz9mskm [accessed 2023-07-25]
36. Wyden, Paul and Bipartisan members of congress introduce the fourth amendment is not for sale act. Ron Wyden United States Senator of Oregon. URL: https://tinyurl.com/3y8s6mtm [accessed 2023-07-25]
37. H.R.8111 - my body, my data act of 2022. US Congress. URL: https://www.congress.gov/bill/117th-congress/house-bill/8111 [accessed 2023-07-25]
38. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, et al. Design and testing of a mobile health application rating tool. NPJ Digit Med 2020;3:74 [FREE Full text] [doi: 10.1038/s41746-020-0268-9] [Medline: 32509971]
39. Clue privacy policy. Clue. URL: https://helloclue.com/privacy [accessed 2023-07-23]
40. Flo privacy policy. Flo Health. URL: https://flo.health/privacy-policy [accessed 2023-07-25]
41. Privacy policy. GP Apps. URL: https://gpapps.com/support/privacy-policy/ [accessed 2023-07-25]
42. Bradley S, Bacharach E, Martens A, Spanfeller J. The 11 best period tracker apps to get to know your cycle, according to ob-gyns. Women's Health. URL: https://www.womenshealthmag.com/health/g26787041/best-period-tracking-apps/ [accessed 2023-07-25]
43. Jewell T, Benton E. The best fertility tracker apps. Healthline. URL: https://www.healthline.com/health/pregnancy/fertility-apps#A-quick-look-at-the-best-fertility-tracker-apps [accessed 2023-07-25]
44. Perez SP, Whittaker Z. Period Tracker Stardust surges following Roe reversal, but its privacy claims aren't airtight. TechCrunch. URL: https://techcrunch.com/2022/06/27/stardust-period-tracker-phone-number/ [accessed 2023-07-25]
45. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
46. We do not and will not sell your data. Stardust. URL: https://stardust.app/privacy-policy.html [accessed 2023-07-25]
47. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

XSL•FO
RenderX

48. Cohen K. Location, health, and other sensitive information: FTC committed to fully enforcing the law against illegal use and sharing of highly sensitive data. Federal Trade Commission. URL: https://www.ftc.gov/business-guidance/blog/2022/07/location-health-and-other-sensitive-information-ftc-committed-fully-enforcing-law-against-illegal [accessed 2022-07-11]
49. App store. Apple. URL: https://www.apple.com/app-store/ [accessed 2023-07-25]
50. How Google Play works. Google. URL: https://play.google/about/howplayworks/ [accessed 2023-07-25]
51. App Store review guidelines. Apple Developer. URL: https://developer.apple.com/app-store/review/guidelines/#privacy [accessed 2023-07-25]
52. Provide information for Google Play's data safety section - play console help. Google. URL: https://support.google.com/googleplay/android-developer/answer/10787469 [accessed 2023-07-25]
53. Gmb H. Clue period tracker and calendar. App Store. URL: https://apps.apple.com/us/app/clue-period-tracker-calendar/id657189652 [accessed 2023-07-25]
54. Harnish B. What is HTTPS: the definitive guide to how HTTPS works. Semrush Blog. URL: https://www.semrush.com/blog/what-is-https/ [accessed 2023-07-25]
55. Data security at Flo. Flo Health. URL: https://tinyurl.com/ucnhzbsw [accessed 2023-07-25]
56. Special announcement. Stardust. 2022. URL: https://minnow-clarinet-j6yf.squarespace.com/ [accessed 2024-07-25]
57. Flo Health UK Limited. Flo period and pregnancy tracker. App Store. URL: https://apps.apple.com/us/app/flo-period-pregnancy-tracker/id1038369065 [accessed 2023-07-25]
58. Stardust App LLC. Stardust period tracker. App Store. URL: https://apps.apple.com/us/app/stardust-period-tracker/id1495829322 [accessed 2023-07-26]
59. Content ratings - play console help. Google. URL: https://support.google.com/googleplay/android-developer/answer/9898843?hl=en [accessed 2023-07-25]
60. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
61. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 06;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]
62. In post Roe v. Wade Era, Mozilla labels 18 of 25 popular period and pregnancy tracking tech with *privacy not included warning. Mozilla Foundation. URL: https://tinyurl.com/578u36cp [accessed 2022-08-17]
63. Mobile health and fitness apps: what are the privacy risks? PrivacyRights.org. 2016. URL: https://privacyrights.org/consumer-guides/mobile-health-and-fitness-apps-what-are-privacy-risks [accessed 2023-07-25]
64. Perlroth N. What is end-to-end encryption? Another bull's-eye on big tech. The New York Times. URL: https://www.nytimes.com/2019/11/19/technology/end-to-end-encryption.html [accessed 2023-07-25]
65. Lutkevich B, Bacon M. What is end-to-end encryption (E2EE) and how does it work? TechTarget. URL: https://tinyurl.com/zr42axsv [accessed 2023-07-25]
66. Abortion is now illegal in 11 U.S. states. Center for Reproductive Rights. 2022. URL: https://reproductiverights.org/abortion-illegal-11-states/ [accessed 2023-07-25]
67. Bassett v. Elec. Arts, Inc. Casetext. URL: https://casetext.com/case/bassett-v-elec-arts-inc [accessed 2023-07-25]
68. McQuillan L. Americans are being urged to delete period tracking apps. Should Canadians do the same? CBC/Radio-Canada. URL: https://www.cbc.ca/news/health/period-tracker-apps-data-privacy-1.6510029 [accessed 2023-07-25]
69. Roe V. Roe v. Wade and Supreme Court abortion cases. Brennan Center for Justice. URL: https://tinyurl.com/yef93835 [accessed 2023-07-25]
70. Flo statement on data privacy. Flo Health. URL: https://flo.health/users-privacy-and-data-security [accessed 2023-07-25]
71. Gupta AH, Singer N. Your app knows you got your period. Guess who it told? The New York Times. URL: https://www.nytimes.com/2021/01/28/us/period-apps-health-technology-women-privacy.html [accessed 2023-07-25]
72. Olsen N. Browsewrap versus clickwrap. Privacy Policies. URL: https://www.privacypolicies.com/blog/browsewrap-vs-clickwrap/ [accessed 2023-07-25]
73. Daugherty J. Clickwrap, Browsewrap, Scrollwrap, Sign-in-wrap. Daugherty Law Office LLC. URL: https://jdcincylaw.com/clickwrap-browsewrap-scrollwrap-sign-in-wrap/ [accessed 2023-07-25]

Abbreviations

EU: European Union
FTC: Federal Trade Commission
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
THESIS: Transparency, Health Content, Excellent Technical Content, Security/Privacy, Usability, Subjective

XSL•FO
RenderX

Edited by A Mavragani; submitted 02.08.23; peer-reviewed by R Bidkar, H McMahon, M Abdalla; comments to author 09.04.24; revised version received 25.07.24; accepted 08.12.24; published 05.03.25.

Please cite as:
Zadushlivy N, Biviji R, Williams KS
Exploration of Reproductive Health Apps' Data Privacy Policies and the Risks Posed to Users: Qualitative Content Analysis
J Med Internet Res 2025;27:e51517
URL: https://www.jmir.org/2025/1/e51517
doi:10.2196/51517
PMID:40053713

©Nina Zadushlivy, Rizwana Biviji, Karmen S Williams. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Combined Effect of Multisensory Stimulation and Therapist Support on Physical and Mental Health of Older Adults Living in Nursing Homes: Pilot Randomized Controlled Trial

Sewar Khatib[1], MHA; Yuval Palgi[2], PhD; Yoni K Ashar[3], PhD; Natalya Polyvyannaya[1,4,5], MSSc; Pavel Goldstein[1], PhD

[1]School of Public Health, University of Haifa, Haifa, Israel

[2]Department of Gerontology, University of Haifa, Haifa, Israel

[3]Faculty of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[4]Department of General and Applied Psychology, Kazakh National University, Almaty, Kazakhstan

[5]Faculty of Business, Media and Management, International Informational Technologies University, Almaty, Kazakhstan

Corresponding Author:
Pavel Goldstein, PhD
School of Public Health
University of Haifa
199 Aba Khoushy Avenue
Mount Carmel
Haifa, 3103301
Israel
Phone: 972 48240519
Email: pavelg@stat.haifa.ac.il

Abstract

Background: Increasing life expectancy has led to a rise in nursing home admissions, a context in which older adults often experience chronic physical and mental health conditions, chronic pain, and reduced well-being. Nonpharmacological approaches are especially important for managing older adults' chronic pain, mental health conditions (such as anxiety and depression), and overall well-being, including sensory stimulation (SS) and therapist support (TS). However, the combined effects of SS and TS have not been investigated.

Objective: This randomized controlled trial examines the specific and combined effects of brief SS and TS interventions on older adults' physical and mental health and pain intensity levels, among individuals living in nursing homes.

Methods: A total of 96 patients aged 65-99 years from a nursing home were randomly assigned to 3 groups: SS, TS, and combined SS+TS interventions, each delivered as four 20-minute sessions. SS was implemented using a multisensory Snoezelen room. Pain intensity levels (per a Visual Analog Scale), blood pressure, heart rate, blood oxygen saturation, and hand grip strength (using a Jamar hand dynamometer) were measured before and after each of the 4 weekly therapeutic sessions. In addition, life satisfaction (per the Satisfaction with Life Scale) and anxiety (per the 7-item General Anxiety Disorder Scale) were evaluated before and after the whole intervention. Mixed model analyses tested the relative efficacy of the 3 interventions, applying simple slope analysis with Tukey correction. Study rationale and analytical plans were preregistered.

Results: The combined intervention of SS and TS (SS+TS) resulted in reduced pain levels compared with SS (B=0.209, P=.006) and TS alone (B=0.23, P=.002) over 4 sessions ($F_{6,266}$=2.62; P=.017; R^2=0.23). Further, the combined SS+TS intervention resulted in reduced systolic blood pressure versus SS (B=0.09, P=.01) and TS alone (B=0.016, P<.001) groups ($F_{6,272}$=5.42; P<.001; R^2=0.29). In addition, the combined SS+TS intervention resulted in an increased grip strength versus SS (B=–0.35, P=.003) and TS alone (B=–0.032, P=.008) groups ($F_{6,273}$=2.25; P=.04; R^2=0.19). Moreover, combined SS+TS resulted in an improvement in life satisfaction (B=–4.29, P<.0001) compared with SS (B=–2.38, P=.0042) and TS alone (B=–1.20, P=.13) groups ($F_{2,39}$=3.47; P=.04). Finally, SS+TS demonstrated greater improvement in symptoms of general anxiety disorder (B=10.64, P<.0001) compared with SS (B=3.30 P=.01) and TS alone (B=1.13, P=.37) ($F_{2,38}$=13.5; P<.001) groups. No differences between the interventions

XSL•FO
RenderX

were shown for blood oxygen saturation ($F_{6,273}$=2.06; P=.06), diastolic blood pressure ($F_{6,272}$=1.12; P=.35), and heart rate ($F_{6,273}$=1.33; P=.23).

Conclusions: The combined intervention of SS and TS showed therapeutic benefits for pain management and physical and mental health of older adults living in nursing homes, relative to each therapeutic component in isolation. This brief intervention can be readily implemented to improve well-being and optimize therapeutic resources in nursing home settings.

Trial Registration: ClinicalTrials.gov NCT05394389; https://clinicaltrials.gov/ct2/show/NCT05394389

(J Med Internet Res 2025;27:e55042) doi:10.2196/55042

KEYWORDS

Snoezelen room; mental health; sensory stimulation environment; social support; nursing homes; older adults

Introduction

The global population is witnessing a notable rise in life expectancy, leading to an expanding aging population. According to the World Health Organization (WHO), older adults (aged 65 and above) made up 12.3% of the global population in 2017, with this proportion expected to surpass 22% by 2050 [1]. Similar trends are evident in the Middle Eastern region. For example, Israel's older adult population (aged ≥65 years) accounted for approximately 12%-13% of its total population in 2020, mirroring trends observed in other developed countries. While the United Nations defines older persons as those aged 60 years and older, this study follows the widely accepted threshold of 65 years and older, commonly used in health-related studies and policies. This demographic shift has driven a significant increase in the number of nursing home beds globally, with the availability of hospital beds in nursing homes rising by more than 200% between 1988 and 2015 [2-5].

Individuals living in nursing homes often encounter challenges related to functional and cognitive impairments, stemming from the natural aging process, polypharmacy, and various health conditions [3,6]. Additionally, they may struggle with loneliness and emotional distress [7]. Therefore, a key therapeutic focus is on preserving and enhancing functional and cognitive abilities while promoting life satisfaction and overall quality of life [2].

Multisensory interventions, such as Snoezelen room therapy, alongside social support, are essential components of care for older adults in nursing homes [8]. Snoezelen room therapy is a form of multisensory stimulation (multi-SS) that involves the gradual application of sensory stimuli to enhance an individual's sensory perception and integration. Numerous studies have emphasized the unique benefits of multi-SS interventions combined with therapist support (TS) in improving physical and mental health outcomes. These include positive effects on heart rate, blood pressure, blood oxygen saturation, depression, anxiety, and pain management [1,2,6,8].

In recent decades, there has been an increasing interest in using multisensory treatment approaches—incorporating visual, auditory, olfactory, and tactile stimulations—to improve the well-being of older adults. Originally developed in the 1960s and 1970s in the Netherlands and the United States for children with learning difficulties and developmental intellectual disabilities, multi-SS environments (MSSEs) have consistently shown benefits across diverse populations. Over time, the scope of MSSE has broadened to include older adult populations, demonstrating success in addressing conditions such as dementia and Alzheimer disease, as well as mental health concerns and other geriatric conditions [1,2,4,6].

In recent years, an expanding body of evidence has suggested that MSSE, especially through Snoezelen room therapy, shows promise as a valuable intervention for enhancing the well-being of older adults. MSSE interventions have been shown to improve communication with patients experiencing various communication disorders and a lack of expression, a common challenge among older adults living in nursing homes (eg, dementia) [1,7]. It creates a "listening" channel to internal sensations by enhancing individuals' sensory regulation, which may serve as an alternative to verbal communication [8].

Previous research has consistently indicated that MSSE interventions, particularly Snoezelen room therapy, can lead to significant improvements in both physical and mental health. Physiological measures, such as heart rate and blood oxygen saturation, have shown positive changes after MSSE sessions [8-10]. Moreover, these interventions have been shown to be effective in enhancing psychosocial abilities and reducing agitation, behavioral disorders, anxiety, and depression among older adults. This, in turn, contributes to overall improvements in life satisfaction and fosters better social relationships and participation in meaningful activities [1,2,8,11].

An effective therapeutic tool within the MSSE approach is Snoezelen room therapy [8,11], which provides concentrated multi-SS while personalizing the intensity of the experience (Figure 1) [1,12,13].

XSL•FO
RenderX

Figure 1. Demonstration of the Snoezelen room therapy.



The Snoezelen room is a specially designed space equipped with various sensory systems, providing a unique therapeutic environment for individuals. This innovative therapy involves the careful integration of lighting, sound, smell, and touch elements to create a multisensory experience. Guided by qualified therapists, individuals can participate in Snoezelen room sessions that evoke feelings of success, tranquility, pleasure, and contentment. One of the key strengths of the Snoezelen room intervention is its nonjudgmental and nonevaluative nature, offering a safe space where participants can explore their senses without fear of failure or criticism [4,14]. Beyond the previously mentioned benefits, Snoezelen room therapy has also been found to reduce pain sensations and enhance motivation, which is a crucial factor for participation in various daily activities [15].

Older adults may often experience social isolation and reduced social connections [16]. Social support has proven effective in alleviating depression, anxiety, and loneliness among older adults living in nursing homes, and it has also been shown to be a protective factor against depression [17,18]. Additionally, social support and social touch have a positive impact on reducing stress and pain intensity [19,20].

The role of therapists' social support in enhancing the well-being of older adults, especially those in nursing homes, is increasingly recognized as a valuable intervention. Numerous studies have shown that the presence of dedicated and compassionate therapists can lead to significant improvements in self-esteem, pain perception, and overall quality of life among older adults [16,21]. A lack of social support among older individuals experiencing chronic pain can contribute to worsened mental health conditions and an increased risk of suicide [19,22]. In the context of nursing home care, there is a pressing need for targeted interventions that emphasize TS to effectively enhance the mental and physical well-being of older adults [16,17,23].

While both multi-SS and TS have been individually recognized for their potential to improve the well-being of older adults, the integration of these interventions remains relatively unexplored. Combining multi-SS with dedicated TS has the potential to create a more profound impact on the overall well-being of older individuals compared with each intervention used alone. By synergistically addressing both the sensory and social needs of older adults, this integrated approach may lead to better pain management and improved mental health outcomes.

In this study, we aim to address the existing research gap by exploring the potential benefits of an integrated intervention that combines multi-SS with TS for the care of older adults. This pilot randomized controlled trial (RCT) is designed to primarily assess the feasibility of implementing this combined intervention and to provide preliminary insights into its effectiveness. Our objectives are 2-fold: first, to determine whether this approach can be successfully implemented in nursing home settings, and second, to gather initial evidence on its potential for future large-scale trials. We will also evaluate the combined impact of SS and TS on pain management, as well as the overall physical and mental health of older adults. We hypothesize that the synergistic effect of integrating SS with TS will lead to more significant improvements in pain reduction and physical and mental health outcomes compared with either intervention applied in isolation.

Methods

Study Design

A randomized clinical trial was conducted over 1 year to examine the combined effect of SS and TS on pain intensity, as well as the mental and physical health of older adults living in nursing homes. The study specifically targeted older adults residing in these settings. All participants were randomly assigned into 3 groups by the supervising researcher using a stratification strategy to ensure balanced groups (see the "Randomization" section). The executing researcher and research staff responsible for assessments were blinded to the randomization process.

XSL•FO
RenderX

Sampling Method and Randomization Process

Participants were recruited using a convenience sample from Ahuzat Hazafon Geriatric Nursing Hospital. Eligible participants were residents aged 65 years or older who had lived in the facility for at least three months and were capable of participating. Exclusion criteria included uncontrolled epilepsy and severe behavioral disturbances. Following eligibility screening, participants were assigned to 1 of 3 groups—SS, TS, or the combined SS+TS group—using stratified randomization. Stratification was based on age, sex, and cognitive level to ensure balanced groups. Randomization and allocation were overseen by an independent researcher who was not involved in the trial's day-to-day activities. Additionally, research staff responsible for conducting baseline and postintervention assessments were blinded to group assignments to minimize assessment bias.

Research Process

Patients, with the approval of their custodians, were invited to participate in the study. Each eligible participant attended 4 weekly therapeutic sessions based on their group assignment: SS (SS intervention using Snoezelen room therapy, involving gradual multi-SS without social interaction), TS (therapist's social support intervention only, without SS), or SS+TS (a combined intervention of both SS and therapist's social support). Each session lasted 20 minutes. Immediately before and after each session, we collected objective measures, including pain intensity, blood pressure, blood oxygen saturation, hand grip strength (HGS), and heart rate, from each participant. Additionally, eligible participants completed subjective assessments (Multimedia Appendix 1) to measure their satisfaction with life and anxiety levels at the beginning and end of the study.

Sample Size

A prestudy power analysis was conducted, and the sample size was calculated based on an effect size of Cohen d=0.40, with a power of 0.80 and a type I error rate of 0.05/6 (0.0083) to account for 6 primary outcomes [18]. This analysis resulted in an estimated sample size of 102 participants.

Outcome Measures

The objective measures included systolic and diastolic blood pressure, heart rate, blood oxygen saturation, and HGS. These metrics were assessed before and after each session. Patient-reported outcomes included life satisfaction, current pain, and general anxiety. Life satisfaction and general anxiety were assessed before the first session and after the last session, while pain was evaluated before and after each session.

Research Tools

The following tools and instruments were used in this study:

- Blood pressure, heart rate, and blood oxygen saturation were measured using the "Welch Allyn" measuring instrument, a commonly used tool in clinical and research settings [24].
- HGS was assessed using the Jamar hand dynamometer, a valid and reliable tool for measuring muscle strength [25,26].
- Life satisfaction was measured using the Satisfaction with Life Scale (SWLS), a short 5-item instrument with high internal reliability (Cronbach α=0.93) and good test-retest reliability (0.84/0.80 for immediate/over a month interval) [27,28]. The SWLS has been translated into Hebrew and validated in multiple studies within the Israeli population, including older adults [29].
- Pain intensity was evaluated using the Visual Analog Scale, a 10-cm ruler with facial illustrations, which has been validated for pain assessment in older adults [30,31].
- General anxiety was assessed using the 7-item General Anxiety Disorder (GAD-7) Scale. The scale is considered valid and reliable for both clinical and research use, with a high level of internal reliability (Cronbach α=0.89). The literature demonstrates the effective use of the GAD-7 in the older population [32,33].
- The MMSE is widely used in Israel and has been validated in several studies involving older Israeli adults [34]. The Hebrew version of the MMSE demonstrated strong psychometric properties, including good internal consistency (Cronbach α=0.82) and excellent test-retest reliability across cognitive domains. It has been found to be a valid tool for assessing dementia in the Israeli older adult population [34].

Prevention and Treatment of Selection Bias

To minimize selection bias, participants were recruited through a convenience sample, and stratified random assignment was used to evenly distribute potential confounders among the study groups. Additionally, the researcher conducting the study was blinded to the group assignments to reduce bias in data collection and analysis.

Treatment of Confounders and Modifiers

To address confounding variables, statistical control was applied. Sex and age, known universal variables that can act as confounders, were included as control variables in the statistical analysis. This approach ensured that any observed group differences were not solely influenced by these variables.

Data Analysis Methods

Group differences were examined using the chi-square test for categorical variables and a 1-way analysis of variance for continuous variables. For session-level outcomes, including blood pressure, blood oxygen saturation, heart rate, HGS, and pain intensity, percentage changes from presession to postsession were calculated. These changes were then analyzed using mixed models to test the differences between treatments over time, defining an interaction between time and treatment. Tukey-corrected post hoc analysis was applied for significant interactions. Pairwise contrasts were used to compare group differences at each session, with participant-based random intercepts and session slopes defined. The analysis of anxiety and life satisfaction outcomes was conducted using mixed effect models to compare pre- and postintervention differences across groups. When significant differences were detected, intervention-specific slopes were further examined using Tukey correction to account for multiple comparisons.

XSL•FO
RenderX

Ethical Considerations and Approval

Participants were recruited by an occupational therapist who contacted all eligible patients living in the nursing home. The researcher explained the purpose of the study, ensured voluntary participation, and clarified that nonparticipation would not affect the patients' or therapist-patient relationship. Informed consent was obtained from participants or their custodians, particularly for those lacking sufficient cognitive capacity. The study was preregistered with ClinicalTrials.gov (registered ID NCT05394389) and received ethical approval from the Faculty of Social Welfare & Health Sciences Ethics Committee at the University of Haifa (approval number [112/21]).

Resources, Equipment, and Physical Tools

The study was conducted at Ahuzat Hazafon Geriatric Nursing Hospital, which provides medical, nursing, and paramedical services to its residents in mentally ill and nursing wards. The Snoezelen room where the study was conducted was a white, furnished space equipped with various SS systems, including lighting, audio, odor distribution, and massage systems. The room already had visual, auditory, tactile, and aromatic SS systems in place, which were used during the study.

Results

Participants

The mean age of participants was 82.2 (SD 7.92) years, with the majority being women (62/96, 65%; Table 1). Most participants exhibited some degree of cognitive decline (mean Mini-Mental State Examination [MMSE] score 15.4, SD 8.45). The analysis indicated no significant differences between the groups in MMSE scores ($F_{2,93}$=1.51; P=.32). Additionally, no significant differences were found among the 3 intervention groups for gender, age, cognitive state, or pain intensity (all Ps>.2). Complete statistical analysis is presented in Multimedia Appendix 2. Participant recruitment is detailed in Figure 2.

Table 1. Sample characteristics of the study participants.

Characteristics	SS (n=34)	SS+TS (n=31)	TS (n=31)	Overall (N=96)
Gender, n (%)				
Female	21 (62)	22 (71)	19 (61)	62 (65)
Male	13 (38)	9 (29)	12 (39)	34 (35)
Age (years)				
Mean (SD)	81.8 (8.34)	83.2 (7.47)	81.6 (8.03)	82.2 (7.92)
Median (range)	84.5 (59.0-95.0)	84.0 (66.0-97.0)	83.0 (60.0-93.0)	84.0 (59.0-97.0)
MMSE[a]= score				
Mean (SD)	13.8 (9.19)	15.5 (7.98)	17.0 (8.02)	15.4 (8.45)
Median (range)	15.0 (0-27.0)	16.0 (0-27.0)	17.0 (0-29.0)	16.0 (0-29.0)
Moderate pain, n (%)	10 (29)	8 (26)	9 (29)	27 (28)

[a]MMSE: Mini-Mental State Examination.

XSL•FO
RenderX

Figure 2. Flowchart of participants' recruitment.

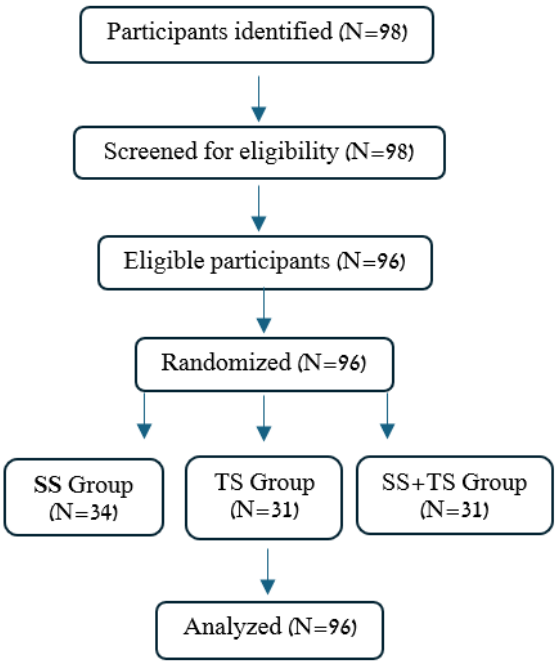

Primary Outcomes

Comparing the Effect of SS, TS, and SS+TS Interventions on Physical Health

Blood Pressure

Intervention groups showed different temporal patterns of systolic blood pressure over the 4 sessions ($F_{6,272}$=5.42; P<.001). The SS+TS group demonstrated a stronger reduction in systolic blood pressure compared with the SS alone group at session 3 (B=0.09; SE 0.03; t_{272}=2.933; P=.01; Cohen d=0.97) and session 4 (B=0.073; SE 0.03; t_{272}=2.4; P=.045; Cohen d=0.8). The SS+TS group also showed a stronger reduction in systolic blood pressure compared with the TS group at session 3 (B=0.107; SE 0.03; t_{272}=1.71; P=.002; Cohen d=1.15) and session 4 (B=0.16; SE 0.031; t_{272}=5.1; P<.001; Cohen d=1.70). All other group differences were not significant (all P>.2).

For diastolic blood pressure, the groups showed a similar pattern of improvement associated with the intervention ($F_{6,272}$=1.12; P=.35).

Hand Grip

Intervention groups showed different temporal patterns of dominant HGS over the 4 sessions ($F_{6,273}$=2.25; P=.04). In the post hoc analysis, the SS+TS group demonstrated a stronger improvement in dominant HGS compared with the SS alone group at session 3 (B=–0.35; SE 0.106; t_{353}=–3.30; P=.003; Cohen d=0.89) and at session 4 (B=–0.34; SE 0.107; t_{354}=–3.21; P=.004; Cohen d=0.87). The SS+TS group also showed stronger HGS improvement compared with TS at session 3 (B=–0.36; SE 0.107; t_{353}=–3.87; P=.02; Cohen d=0.91) and at session 4 (B=–0.032; SE 0.107; t_{354}=–2.99; P=.008; Cohen d=0.81). No significant differences were found in other comparisons between the groups (all P>.11). Nondominant HGS did not show group differences in temporal dynamics over the 4 sessions ($F_{6,271}$=1.27; P=.27).

Pain

Intervention groups showed different temporal patterns of pain reduction ($F_{6,266}$=2.62; P=.02). The SS+TS group demonstrated stronger pain reduction compared with SS alone, at session 3 (B=0.209; SE 0.067; t_{307}=3.12; P=.006; Cohen d=0.91) and at session 4 (B=0.20; SE 0.067; t_{307}=2.96; P=.009; Cohen d=0.87). Additionally, the SS+TS group showed stronger pain reduction compared with the TS alone group, at session 3 (B=0.18; SE 0.068; t_{307}=2.73; P=.02; Cohen d=0.81) and at session 4 (B=0.23; SE 0.068; t_{307}=3.38; P=.002; Cohen d=1.01). No significant differences were found in all other pairwise group comparisons (P>.32 in all cases; Figure 3).

XSL•FO
RenderX

Figure 3. Fluctuations over 4 sessions for physical health measures: (A) pain intensity levels, (B) systolic blood pressure, and (C) hand grip across 3 interventions: (1) therapist support (TS), sensory stimulation (SS), sensory stimulation and therapist's support (SS+TS). TS+SS demonstrated relative improvement in all 3 outcomes at sessions 3 and 4. *P<.05, **P<.01, ***P<.001.

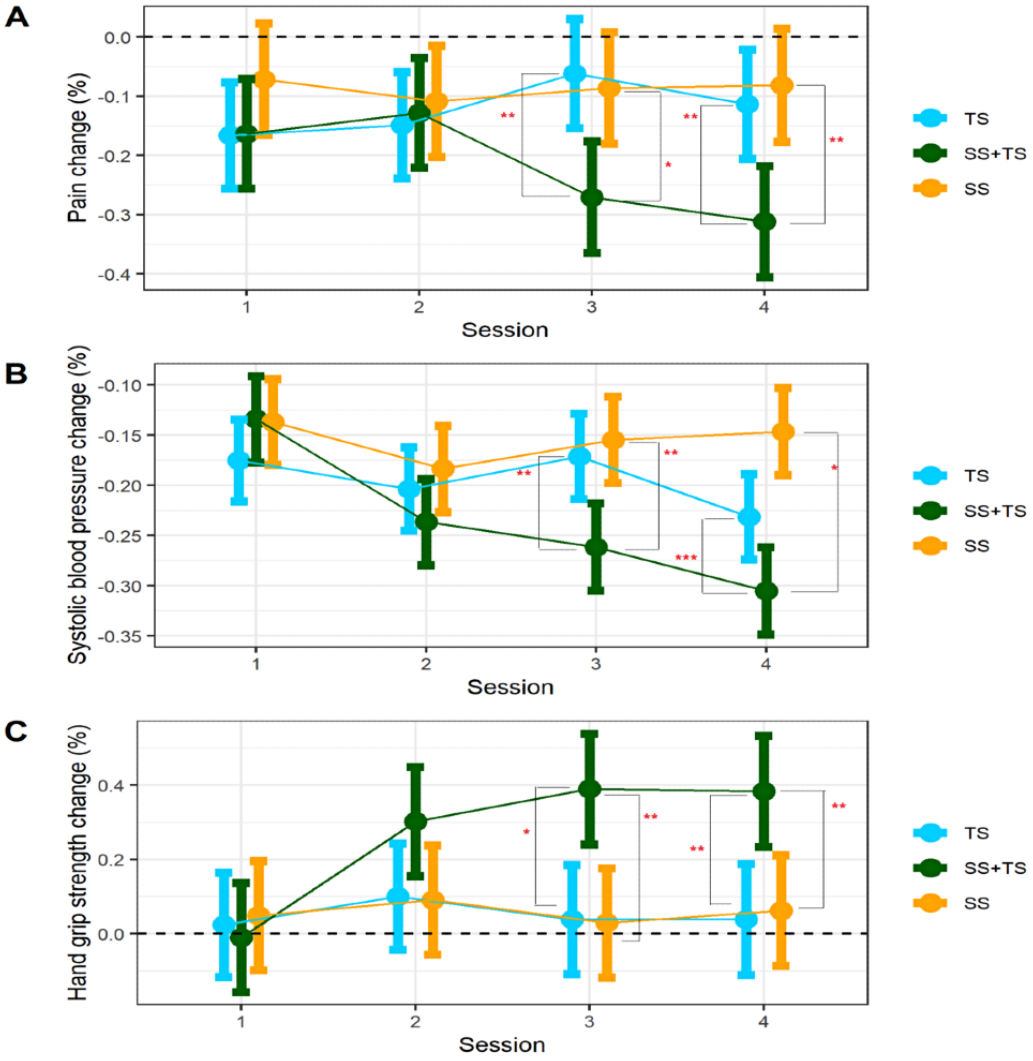

Comparing the Effect of SS, TS, and SS+TS Interventions on Emotional Health

Anxiety

The analysis revealed different temporal patterns of anxiety reduction among the intervention groups ($F_{2,38}$=13.5; P<.001). As hypothesized, the SS+TS group demonstrated a significant reduction in anxiety (B=10.64; SE 1.41; $t_{38.8}$=7.57; P<.001; Cohen d=1.57). By contrast, the SS and TS alone groups showed no significant improvement in anxiety (SS, P=.15; TS, P=.38).

Satisfaction With Life

The analysis indicated different temporal patterns of satisfaction with life among the intervention groups ($F_{2,39}$=3.47; P=.04). As hypothesized, the SS+TS group showed significant improvement in life satisfaction (SWLS; B=–4.29; SE 39.5; $t_{39.5}$=–4.89; P<.001; Cohen d=1.44). Additionally, the SS group showed significant improvement in the SWLS (B=–2.38; SE 0.78; P=.004; Cohen d=0.99). However, the TS group did not show a significant improvement (P=.13; Figure 4).

XSL•FO
RenderX

Figure 4. Mental health outcomes before/after 3 interventions: (1) therapist support (TS), (2) sensory Stimulation (SS), and (3) sensory stimulation and Therapist's support (SS+TS). (A) Life satisfaction; (B) anxiety. The dotted line at life satisfaction=25 represents the cut-off for satisfied/extremely satisfied (above the line). The dotted line at anxiety=10 represents the cut-off for severe anxiety (above the line).

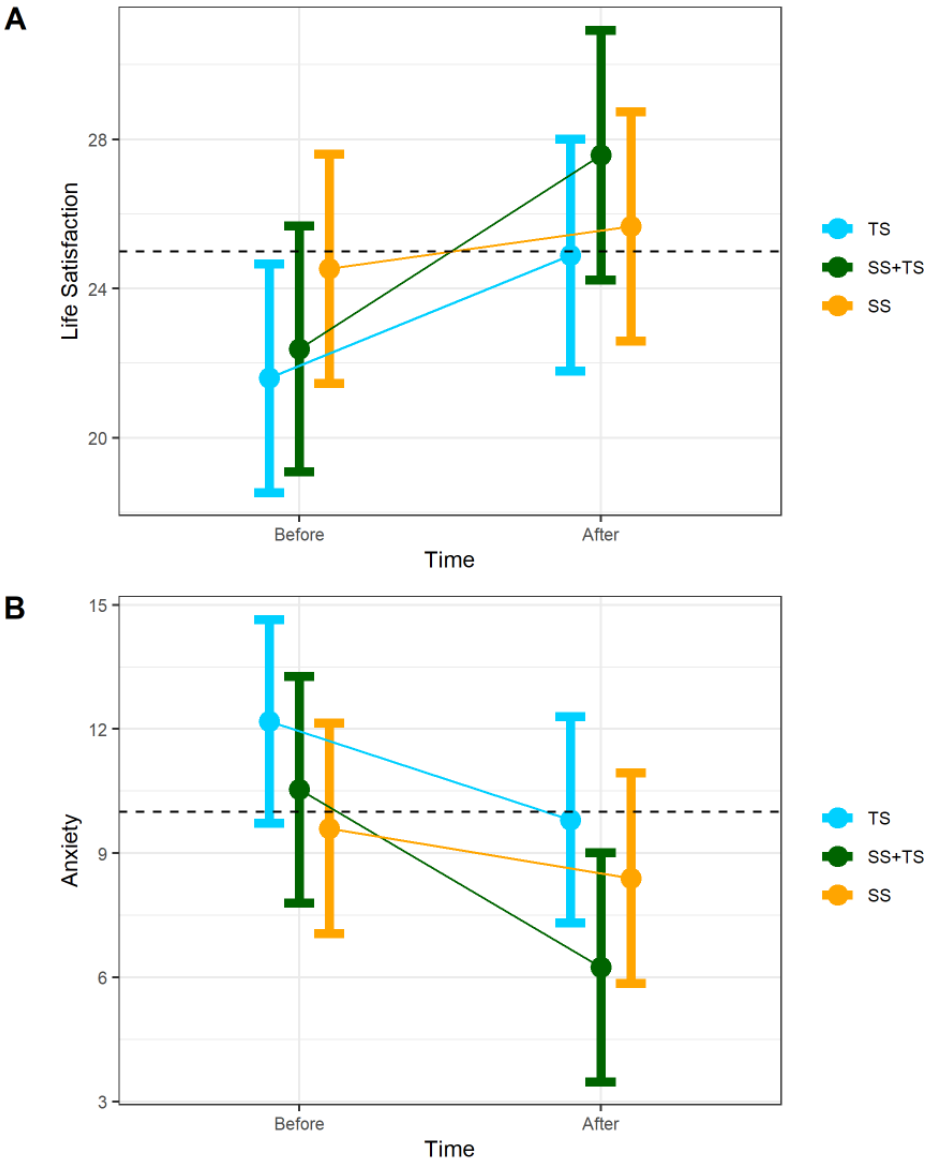

Secondary Outcomes

The analysis revealed that the groups showed a similar pattern of improvement related to the intervention for heart rate ($F_{6,271}$=1.33; P=.23) and blood oxygen saturation ($F_{6,274}$=2.06; P=.06).

Discussion

Principal Findings

This pilot randomized clinical trial is the first to investigate the combined effects of multi-SS in a Snoezelen room and TS (SS+TS) on the health of older adults living in nursing homes. Our findings indicate that the SS+TS intervention significantly reduced pain levels, lowered systolic blood pressure, and improved HGS compared with the SS and TS interventions individually. Additionally, the SS+TS intervention enhanced life satisfaction and reduced anxiety, with these effects being more pronounced than those observed with SS or TS alone [12,15,22].

Our findings supported the primary hypothesis, showing that the combined intervention of SS in a Snoezelen room, along with TS, had a more profound effect on pain intensity and physical and mental health outcomes compared with each intervention component delivered in isolation. Specifically, the SS+TS intervention led to reduced pain levels, lower systolic blood pressure, and increased HGS during sessions 3 and 4, surpassing the effects of SS and TS alone. Additionally, the SS+TS intervention showed greater improvements in life satisfaction and anxiety compared with the other interventions.

Effect of Multisensory Stimulation and Therapist's Support on Pain

The therapeutic impact of the combined SS+TS intervention notably improved pain experience during the latter half of the intervention. Previous research has highlighted the unique positive effects of SS and TS on pain perception [1,15].

XSL•FO
RenderX

Pharmacological interventions, particularly opioids and other pain medications, can have adverse effects on older adults, including falls, cognitive impairment, constipation, and addiction [23,35]. Nonpharmacological methods for pain management, such as the Snoezelen room integrated with enhanced TS, offer a safer alternative and are often more cost-effective in the long term compared with continuous pharmaceutical treatments [23,33,34].

Our findings, combined with existing literature on the Snoezelen room intervention, support the multidimensional effects of the SS+TS intervention on pain intensity. Alleviating pain is crucial, as the experience of pain can lead to decreased mobility, sleep disturbances, emotional distress, and, consequently, a reduced quality of life among older adults [23,35,36].

Interestingly, our results revealed that the therapeutic effects of the combined intervention were most pronounced during sessions 3 and 4. This observation aligns with prior research indicating that pain reduction following nonpharmacological interventions can be delayed [1,15,37-39]. For example, pain reduction resulting from physical exercise for low back pain is often gradual and occurs over time [40,41].

Additionally, social support–focused interventions have been associated with delayed pain relief [23,37]. This delay may be due to the gradual implementation and refinement of coping strategies. It can take time for participants, particularly older adults, to adjust to a new type of therapy that includes multiple elements. This adaptation period may delay the positive interpretation of bodily sensations [37,42-46]. Further investigation is warranted to explore the temporal dynamics and long-term effects of the intervention.

Effect of Multisensory Stimulation and Therapist's Support on Physical Health

We observed an increased impact of the combined SS+TS intervention on systolic blood pressure and HGS. Previous studies have shown that SS, especially olfactory and visual stimulation, can effectively reduce blood pressure and enhance overall health [12,47,48]. Another study highlighted that TS can enhance physical health by providing a safe and nonjudgmental space for older adults to share their concerns, fears, and anxieties. Such support can help reduce stress, which is linked to various physical health issues, including hypertension, heart disease, and compromised immune function [17,23,38,49-52].

Our findings underscore the importance of interventions that address the mind-body connection. Recent studies have shed light on the physiological mechanisms through which mental processes impact the immune system and other bodily functions [47,53-55]. Moreover, some multisensory interventions include resistance training and muscle-strengthening exercises. Participating in these activities can help older adults maintain or even improve muscle strength and endurance, leading to better physical health and greater independence in daily activities [54,56].

Contrary to our initial hypothesis, we did not find that the combined SS+TS intervention provided greater benefits compared with SS and TS alone in terms of heart rate, blood oxygen saturation, and diastolic blood pressure. None of the interventions showed improvements in these metrics. It is plausible that the duration of our intervention (4 sessions lasting 17-20 minutes each) was insufficient to detect changes in these outcomes. Moreover, existing research supports the notion that similar interventions focusing on sensory processing and social support may not effectively improve heart rate and blood oxygen saturation in advanced age [8,21]. Additionally, stable diastolic blood pressure among older adults is often indicative of good health, and reducing diastolic blood pressure could even pose a risk factor for heart disease [57,58].

Improving Mental Health Through Multisensory Stimulation and Therapist Support

Our study yielded promising findings regarding the therapeutic benefits of a brief SS+TS intervention for enhancing mental health among older individuals residing in nursing homes. This intervention showed potential for improving life satisfaction and reducing anxiety, without exacerbating the common issue of polypharmacy in the geriatric field. It is well-established that better mental health in this population is closely linked to improvements in physical well-being [49]. Previous research has consistently highlighted that multi-SS, which often includes elements such as soothing music, aromatherapy, and tactile sensations, can reduce stress levels, alleviate anxiety, and promote a sense of calm. For individuals dealing with anxiety disorders, such as generalized anxiety disorder or posttraumatic stress disorder, multisensory interventions can serve as valuable coping mechanisms.

Multisensory experiences can trigger the release of endorphins and other feel-good neurotransmitters, leading to improved mood and emotional well-being. Furthermore, chronic pain has been associated with an increased aversiveness to unpleasant multi-SSs, suggesting shared mechanisms between pain intensity and multisensory processing [15,59]. For older individuals who may struggle with depression or mood disorders, regular participation in multisensory activities can offer a natural, nonpharmacological approach to improving their mood. Additionally, multisensory activities that encourage social engagement can help mitigate feelings of isolation and loneliness, providing residents with a sense of belonging and companionship [19,21,32,60]. Moreover, these improvements have been shown to contribute to an enhanced sense of well-being and increased social engagement [60]. An extensive body of research has demonstrated the significant role of TS in improving various markers of mental health [17,18,20,61]. TS is widely regarded as a valuable component for addressing emotional and social needs, which are often critical for older adults facing mental health challenges. Emotional support and empathy can significantly enhance treatment outcomes for older adults experiencing depression, anxiety, and reduced life satisfaction [18,20,62,63].

An intriguing question arises from our study: What makes the combined SS+TS intervention particularly effective? We propose that the therapeutic state of mind fostered by the relaxing environment created through SS plays a crucial role. This environment prepares the groundwork for the TS to have a greater impact by reducing patient resistance. Supporting this

XSL•FO
RenderX

notion, previous research has demonstrated that SS creates a more conducive analgesic environment compared with standard settings [15,64]. This shift in environment may be particularly therapeutic for older individuals living in nursing homes, where the atmosphere often mirrors that of a hospital, characterized by medications and medical procedures that can heighten feelings of discomfort and anxiety.

The observed effects of the SS+TS intervention may result from an interactive synergy between SS and TS. Rather than being merely additive, these interventions appear to enhance each other, leading to greater therapeutic outcomes. The soothing environment created by SS likely amplifies the effectiveness of TS, enabling a more profound impact on participants' health. This combined effect underscores the potential of the SS+TS intervention to be particularly effective in settings such as nursing homes, where establishing a calming, supportive environment is crucial.

In exploring the nature of the combined effects of SS and TS, a critical question arises: Do these interventions provide additive or interactive benefits when applied together? An additive effect would suggest that the benefits of the combined SS+TS intervention are simply the sum of the individual impacts of SS and TS when applied separately. By contrast, an interactive effect would indicate a synergistic relationship, where the combined SS+TS intervention produces outcomes that surpass the sum of its individual components. Although a no-treatment control group would offer a more definitive basis for distinguishing between these possibilities, our current analysis, based on a visual examination of pre-to-post intervention changes, suggests the presence of an interactive effect. This observation suggests that the combination of SS and TS may activate mechanisms that go beyond their individual contributions, leading to enhanced therapeutic outcomes. This synergy could stem from how SS prepares individuals mentally and physically for more effective engagement with therapeutic support, or vice versa. Future research with a more comprehensive experimental design, including a control group and separate groups for each intervention, would be invaluable for confirming and further elucidating these effects.

Limitations and Recommendations for Future Studies

Although our study has provided valuable insights, it is important to acknowledge its limitations. Conducted in a single nursing home, our findings may not fully represent the broader population.

First, the use of a convenience sample from a single nursing home may introduce selection bias, as participants may not fully represent the broader population of older adults in similar care settings. For instance, participants with behavioral issues were excluded, which limits the generalizability of the findings to this wider population. Additionally, the voluntary nature of participation may introduce selection bias, as participants who chose to join the study may have been more motivated or in better health, potentially influencing the outcomes. Future studies should aim to replicate our findings across multiple settings, explore additional measures of physical and mental health, and extend the intervention period to assess long-term effects.

Second, the sustainability of the observed effects was not assessed, as follow-up data were not collected. Consequently, while our findings suggest potential benefits of the combined SS+TS intervention, the duration of these effects remains unknown.

Third, the study's modest sample size may limit the power to detect small effect sizes. Further research with larger sample sizes and extended follow-up periods is needed to validate these preliminary findings. Furthermore, exploring the impact of the SS+TS intervention on medication use, particularly for mental health conditions, as well as on sleep quality among nursing home residents, would provide additional valuable insights into the intervention's effectiveness.

Conclusion and Implications for Clinical Practice

Our study highlights the therapeutic potential of the combined SS+TS intervention for managing pain and enhancing the physical and mental well-being of older adults in nursing homes. The results indicate that this intervention not only improves key health outcomes but also has the potential to reduce reliance on medications, addressing concerns related to polypharmacy. The scalability and ease of implementation make SS+TS a valuable nonpharmacological option for nursing homes, potentially enhancing patient care and resource efficiency. Future research should focus on larger-scale studies and long-term follow-up to confirm these findings and explore the broader applicability of the intervention.

Acknowledgments

We thank the Ahuzat Hazafon Geriatric Nursing Hospital for their support in conducting the RCT at the Snoezelen Room within the Occupational Therapy Department. We extend our gratitude, in particular, to the occupational instructors for their invaluable assistance in facilitating the research. This study was supported by the Israeli Science Foundation (grant 2608/22) and The Israel Data Science Initiative (IDSI) of the Council for Higher Education in Israel and the Data Science Research Center at the University of Haifa (PG). This work was also supported by the National Institute on Aging of the National Institutes of Health (grant R24AG064191) and the National Institute of Aging (grants 1P30AG064198-01 and 5P30AG022845-19 to YA). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

XSL•FO
RenderX

Authors' Contributions

SK, PG, and YP were responsible for the study conceptualization and methodology. SK and PG led the implementation of the study, statistical analysis, writing, and editing. YP, YA, and NP significantly contributed to the manuscript preparation and interpretation of findings.

Conflicts of Interest

The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. YKA reports receiving consulting fees from the Pain Reprocessing Therapy Center, Lin Health, and Mental Health Partners of Boulder County.

Multimedia Appendix 1
Research protocols.
[DOCX File , 18 KB - jmir_v27i1e55042_app1.docx]

Multimedia Appendix 2
Additional analysis and results.
[DOCX File , 42 KB - jmir_v27i1e55042_app2.docx]

Multimedia Appendix 3
CONSORT 2010 (Consolidated Standards of Reporting Trials) checklist.
[DOC File , 219 KB - jmir_v27i1e55042_app3.doc]

References

1. Duchi F, Benalcázar E, Huerta M, Bermeo J, Lozada F, Condo S. Design of a multisensory room for elderly people with neurodegenerative diseases. In: Lhotska L, Sukupova L, Lacković I, Ibbott G, editors. World Congress on Medical Physics and Biomedical Engineering 2018. IFMBE Proceedings, vol 68/3. Berlin, Germany: Springer; 2019:207-210.
2. Prins A, Scherder E, van Straten A, Zwaagstra Y, Milders M. Sensory stimulation for nursing-home residents: systematic review and meta-analysis of its effects on sleep quality and rest-activity rhythm in dementia. Dement Geriatr Cogn Disord 2020;49(3):219-234 [FREE Full text] [doi: 10.1159/000509433] [Medline: 32920562]
3. Sánchez A, Millán-Calenti JC, Lorenzo-López L, Maseda A. Multisensory stimulation for people with dementia: a review of the literature. Am J Alzheimers Dis Other Demen 2013 Feb 07;28(1):7-14 [FREE Full text] [doi: 10.1177/1533317512466693] [Medline: 23221029]
4. Staal JA, Amanda S, Matheis R, Collier L, Calia T, Hanif H, et al. The effects of Snoezelen (multi-sensory behavior therapy) and psychiatric care on agitation, apathy, and activities of daily living in dementia patients on a short term geriatric psychiatric inpatient unit. Int J Psychiatry Med 2007 Dec 01;37(4):357-370. [doi: 10.2190/pm.37.4.a]
5. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction. Circulation 2010 Feb 02;121(4):586-613. [doi: 10.1161/circulationaha.109.192703]
6. Sánchez A, Marante-Moar MP, Sarabia C, de Labra C, Lorenzo T, Maseda A, et al. Multisensory stimulation as an intervention strategy for elderly patients with severe dementia: a pilot randomized controlled trial. Am J Alzheimers Dis Other Demen 2016 Jun 01;31(4):341-350 [FREE Full text] [doi: 10.1177/1533317515618801] [Medline: 26631687]
7. van Weert JC, Kerkstra A, van Dulmen AM, Bensing JM, Peter JG, Ribbe MW. The implementation of Snoezelen in psychogeriatric care: an evaluation through the eyes of caregivers. International Journal of Nursing Studies 2004 May;41(4):397-409. [doi: 10.1016/j.ijnurstu.2003.10.011] [Medline: 15050851]
8. Maseda A, Cibeira N, Lorenzo-López L, González-Abraldes I, Buján A, de Labra C, et al. Multisensory stimulation and individualized music sessions on older adults with severe dementia: effects on mood, behavior, and biomedical parameters. JAD 2018 May 30;63(4):1415-1425. [doi: 10.3233/jad-180109]
9. Song C, Ikei H, Miyazaki Y. Physiological effects of visual stimulation with forest imagery. Int J Environ Res Public Health 2018 Jan 26;15(2):213 [FREE Full text] [doi: 10.3390/ijerph15020213] [Medline: 29373558]
10. Mourot L. Limitation of maximal heart rate in hypoxia: mechanisms and clinical importance. Front Physiol 2018 Jul 23;9:972-917 [FREE Full text] [doi: 10.3389/fphys.2018.00972] [Medline: 30083108]
11. Selamat S, Amiruddin A, Dahlan A, Kadar M. The effectiveness of Snoezelen as multisensory intervention among individual who exhibit maladaptive behaviour. Healthscope 2020 Aug 7;333:9 [FREE Full text]
12. Luxwolda M. Snoezelen as multisensory stimulation to promote healthy ageing. Bachelorscriptie Life Science & Technology. Neurosciences Rijksuniversiteit Groningen. 2017. URL: https://tinyurl.com/m3v8jemv [accessed 2024-12-26]
13. Smith BC, D'Amico M. Sensory-based interventions for adults with dementia and Alzheimer's disease: a scoping review. Occup Ther Health Care 2020 Jul 08;34(3):171-201. [doi: 10.1080/07380577.2019.1608488] [Medline: 31066598]

XSL•FO
RenderX

14. Saes J. Colour applied to TEACCH and Snoezelen rooms and therapies. Journal of the International Colour Association 2019;23:15-24 [FREE Full text]
15. Schofield P. Evaluating Snoezelen for relaxation within chronic pain management. Br J Nurs 2002 Jun 27;11(12):812-821. [doi: 10.12968/bjon.2002.11.12.10304] [Medline: 12131831]
16. Seddigh M, Hazrati M, Jokar M, Mansouri A, Bazrafshan M, Rasti M, et al. A comparative study of perceived social support and depression among elderly members of senior day centers, elderly residents in nursing homes, and elderly living at home. Iranian J Nursing Midwifery Res 2020;25(2):160. [doi: 10.4103/ijnmr.ijnmr_109_18]
17. Sun Y, Zhang D, Yang Y, Wu M, Xie H, Zhang J, et al. Social support moderates the effects of self-esteem and depression on quality of life among Chinese rural elderly in nursing homes. Arch Psychiatr Nurs 2017 Apr;31(2):197-204. [doi: 10.1016/j.apnu.2016.09.015] [Medline: 28359433]
18. Harandi TF, Taghinasab MM, Nayeri TD. The correlation of social support with mental health: a meta-analysis. Electron Physician 2017 Sep 25;9(9):5212-5222 [FREE Full text] [doi: 10.19082/5212] [Medline: 29038699]
19. Zhao X, Zhang D, Wu M, Yang Y, Xie H, Li Y, et al. Loneliness and depression symptoms among the elderly in nursing homes: a moderated mediation model of resilience and social support. Psychiatry Research 2018 Oct;268:143-151. [doi: 10.1016/j.psychres.2018.07.011] [Medline: 30025285]
20. Goldstein P, Shamay-Tsoory SG, Yellinek S, Weissman-Fogel I. Empathy predicts an experimental pain reduction during touch. J Pain 2016 Oct;17(10):1049-1057. [doi: 10.1016/j.jpain.2016.06.007] [Medline: 27363627]
21. Wu M, Yang Y, Zhang D, Zhao X, Sun Y, Xie H, et al. Association between social support and health-related quality of life among Chinese rural elders in nursing homes: the mediating role of resilience. Qual Life Res 2018 Mar 9;27(3):783-792. [doi: 10.1007/s11136-017-1730-2] [Medline: 29124499]
22. Grover S, Avasthi A, Sahoo S, Lakdawala B, Dan A, Nebhinani N, et al. Relationship of loneliness and social connectedness with depression in elderly: a multicentric study under the aegis of Indian Association for Geriatric Mental Health. J Geriatr Ment Health 2018;5(2):99. [doi: 10.4103/jgmh.jgmh_26_18]
23. Lee J, Moon Y. Effects of chronic pain and social support on depression and suicide in the elderly. Journal of Digital Convergence 2015 Oct 28;13(10):445-458. [doi: 10.14400/jdc.2015.13.10.445]
24. Schäfer A, Vagedes J. How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 2013 Jun 05;166(1):15-29. [doi: 10.1016/j.ijcard.2012.03.119] [Medline: 22809539]
25. Hamilton GF, McDonald C, Chenier TC. Measurement of grip strength: validity and reliability of the sphygmomanometer and jamar grip dynamometer. J Orthop Sports Phys Ther 1992 Nov;16(5):215-219. [doi: 10.2519/jospt.1992.16.5.215] [Medline: 18796752]
26. Trampisch US, Franke J, Jedamzik N, Hinrichs T, Platen P. Optimal Jamar dynamometer handle position to assess maximal isometric hand grip strength in epidemiological studies. J Hand Surg Am 2012 Nov;37(11):2368-2373. [doi: 10.1016/j.jhsa.2012.08.014] [Medline: 23101534]
27. Arrindell W, Meeuwesen L, Huyse F. The Satisfaction With Life Scale (SWLS): psychometric properties in a non-psychiatric medical outpatients sample. Personality and Individual Differences 1991 Jan;12(2):117-123. [doi: 10.1016/0191-8869(91)90094-r]
28. Sancho P, Galiana L, Gutierrez M, Francisco E, Tomás JM. Validating the Portuguese version of the Satisfaction With Life Scale in an elderly sample. Soc Indic Res 2012 Jan 22;115(1):457-466. [doi: 10.1007/s11205-012-9994-y]
29. Anaby D, Jarus T, Zumbo BD. Psychometric evaluation of the Hebrew language version of the Satisfaction With Life Scale. Soc Indic Res 2009 Apr 25;96(2):267-274. [doi: 10.1007/s11205-009-9476-z]
30. Thong ISK, Jensen MP, Miró J, Tan G. The validity of pain intensity measures: what do the NRS, VAS, VRS, and FPS-R measure? Scand J Pain 2018 Jan 26;18(1):99-107 [FREE Full text] [doi: 10.1515/sjpain-2018-0012] [Medline: 29794282]
31. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov 07;63 Suppl 11(S11):S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
32. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. J Anxiety Disord 2014 Aug;28(6):547-552. [doi: 10.1016/j.janxdis.2014.06.002] [Medline: 24983795]
33. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
34. Werner P, Heinik J, Mendel A, Reicher B, Bleich A. Examining the reliability and validity of the Hebrew version of the Mini Mental State Examination. Aging (Milano) 1999 Oct 01;11(5):329-334. [doi: 10.1007/BF03339808] [Medline: 10631883]
35. Ali A, Arif A, Bhan C, Kumar D, Malik M, Sayyed Z, et al. Managing chronic pain in the elderly: an overview of the recent therapeutic advancements. Cureus 2018 Sep 13;10(9):e3293 [FREE Full text] [doi: 10.7759/cureus.3293] [Medline: 30443463]

XSL•FO
RenderX

36. Catananti C, Gambassi G. Pain assessment in the elderly. Surg Oncol 2010 Sep;19(3):140-148. [doi: 10.1016/j.suronc.2009.11.010] [Medline: 20006488]
37. Christenfeld N. Memory for pain and the delayed effects of distraction. Health Psychology 1997 Jul;16(4):327-330. [doi: 10.1037//0278-6133.16.4.327] [Medline: 9237084]
38. Smith TO, Dainty JR, Williamson E, Martin KR. Association between musculoskeletal pain with social isolation and loneliness: analysis of the English Longitudinal Study of Ageing. Br J Pain 2019 May 20;13(2):82-90 [FREE Full text] [doi: 10.1177/2049463718802868] [Medline: 31019689]
39. Cabrera E, Sutcliffe C, Verbeek H, Saks K, Soto-Martin M, Meyer G, et al. Non-pharmacological interventions as a best practice strategy in people with dementia living in nursing homes. A systematic review. European Geriatric Medicine 2015 Apr;6(2):134-150. [doi: 10.1016/j.eurger.2014.06.003]
40. Mayer J, Mooney V, Dagenais S. Evidence-informed management of chronic low back pain with lumbar extensor strengthening exercises. Spine J 2008 Jan;8(1):96-113. [doi: 10.1016/j.spinee.2007.09.008] [Medline: 18164458]
41. Knox MF, Chipchase LS, Schabrun SM, Marshall PW. Improved compensatory postural adjustments of the deep abdominals following exercise in people with chronic low back pain. Journal of Electromyography and Kinesiology 2017 Dec;37:117-124. [doi: 10.1016/j.jelekin.2017.10.009] [Medline: 29080466]
42. Garland E, Bryan M, Priddy S, Riquino M, Froeliger B, Howard M. Effects of mindfulness-oriented recovery enhancement versus social support on negative affective interference during inhibitory control among opioid-treated chronic pain patients: a pilot mechanistic study. Ann Behav Med 2019 Aug 29;53(10):865-876 [FREE Full text] [doi: 10.1093/abm/kay096] [Medline: 30668631]
43. Goldstein P, Losin EAR, Anderson SR, Schelkun VR, Wager TD. Clinician-patient movement synchrony mediates social group effects on interpersonal trust and perceived pain. The Journal of Pain 2020 Nov;21(11-12):1160-1174 [FREE Full text] [doi: 10.1016/j.jpain.2020.03.001] [Medline: 32544602]
44. Saffarinia M, Dortaj A. Effect of group logotherapy on life expectancy and mental and social wellbeing of the female elderly residents of nursing homes in Dubai. Salmand 2018 Jan 01;12(4):482-493. [doi: 10.21859/sija.12.4.482]
45. Curtis A, Gibson L, O'Brien M, Roe B. Systematic review of the impact of arts for health activities on health, wellbeing and quality of life of older people living in care homes. Dementia (London) 2018 Aug 24;17(6):645-669. [doi: 10.1177/1471301217740960] [Medline: 30041548]
46. Moyer CA, Rounds J, Hannum JW. A meta-analysis of massage therapy research. Psychol Bull 2004 Jan;130(1):3-18. [doi: 10.1037/0033-2909.130.1.3] [Medline: 14717648]
47. Dong S, Jacob TJ. Combined non-adaptive light and smell stimuli lowered blood pressure, reduced heart rate and reduced negative affect. Physiol Behav 2016 Mar 15;156:94-105 [FREE Full text] [doi: 10.1016/j.physbeh.2016.01.013] [Medline: 26780148]
48. Shikder S, Mourshed M, Price A. Therapeutic lighting design for the elderly: a review. Perspect Public Health 2012 Nov 20;132(6):282-291. [doi: 10.1177/1757913911422288] [Medline: 23111083]
49. Morgan-Brown M, Brangan J. Capturing interactive occupation and social engagement in a residential dementia and mental health setting using quantitative and narrative data. Geriatrics (Basel) 2016 Jun 28;1(3):15 [FREE Full text] [doi: 10.3390/geriatrics1030015] [Medline: 31022809]
50. de Oliveira LDSSCB, Souza EC, Rodrigues RAS, Fett CA, Piva AB. The effects of physical activity on anxiety, depression, and quality of life in elderly people living in the community. Trends Psychiatry Psychother 2019 Mar;41(1):36-42 [FREE Full text] [doi: 10.1590/2237-6089-2017-0129] [Medline: 30994779]
51. Mancia G, Sega R, Milesi C, Cesana G, Zanchetti A. Blood-pressure control in the hypertensive population. Lancet 1997 Feb 15;349(9050):454-457. [doi: 10.1016/s0140-6736(96)07099-7] [Medline: 9040574]
52. Saffari M, Lin C, Chen H, Pakpour AH. The role of religious coping and social support on medication adherence and quality of life among the elderly with type 2 diabetes. Qual Life Res 2019 Aug 29;28(8):2183-2193. [doi: 10.1007/s11136-019-02183-z] [Medline: 31037591]
53. Muhsen K, Abed El-Hai R, Amit-Aharon A, Nehama H, Gondia M, Davidovitch N, et al. Risk factors of underutilization of childhood immunizations in ultraorthodox Jewish communities in Israel despite high access to health care services. Vaccine 2012 Mar 09;30(12):2109-2115. [doi: 10.1016/j.vaccine.2012.01.044] [Medline: 22285273]
54. Zhang Y, Li C, Zou L, Liu X, Song W. The effects of mind-body exercise on cognitive performance in elderly: a systematic review and meta-analysis. IJERPH 2018 Dec 09;15(12):2791 [FREE Full text] [doi: 10.3390/ijerph15122791] [Medline: 30544873]
55. Lou M. The use of music to decrease agitated behaviour of the demented elderly: the state of the science. Scand J Caring Sci 2001 Dec 21;15(2):165-173. [doi: 10.1046/j.1471-6712.2001.00021.x] [Medline: 12078630]
56. Gök Ugur H, Yaman Aktaş Y, Orak OS, Saglambilen O, Aydin Avci. The effect of music therapy on depression and physiological parameters in elderly people living in a Turkish nursing home: a randomized-controlled trial. Aging Ment Health 2017 Dec 03;21(12):1280-1286. [doi: 10.1080/13607863.2016.1222348] [Medline: 27592520]
57. Fung MM, Peters K, Redline S, Ziegler MG, Ancoli-Israel S, Barrett-Connor E, et al. Decreased slow wave sleep increases risk of developing hypertension in elderly men. Hypertension 2011 Oct;58(4):596-603. [doi: 10.1161/hypertensionaha.111.174409]

XSL•FO
RenderX

58. Landi F, Calvani R, Picca A, Tosato M, D'Angelo E, Martone AM, et al. Relationship between cardiovascular health metrics and physical performance in community-living people: results from the Longevity check-up (Lookup) 7+ project. Sci Rep 2018 Nov 05;8(1):16353-16311. [doi: 10.1038/s41598-018-34746-4] [Medline: 30397361]
59. Taher AV, Ahmadi MK, Zamir FP. Effects of multi-sensory stimulation on cognition function, depression, anxiety and quality of life in elderly persons with dementia. International Journal of Sport Studies 2015;5(3):355-360 [FREE Full text]
60. Kiely KM, Brady B, Byles J. Gender, mental health and ageing. Maturitas 2019 Nov;129:76-84. [doi: 10.1016/j.maturitas.2019.09.004] [Medline: 31547918]
61. Che X, Cash R, Chung S, Fitzgerald PB, Fitzgibbon BM. Investigating the influence of social support on experimental pain and related physiological arousal: a systematic review and meta-analysis*. Neurosci Biobehav Rev 2018 Sep;92:437-452. [doi: 10.1016/j.neubiorev.2018.07.005] [Medline: 30006033]
62. Ashman T, Cantor J, Tsaousides T, Spielman L, Gordon W. Comparison of cognitive behavioral therapy and supportive psychotherapy for the treatment of depression following traumatic brain injury: a randomized controlled trial. Journal of Head Trauma Rehabilitation 2014;29(6):78. [doi: 10.1097/htr.0000000000000098]
63. Lu S, Liu T, Wong GHY, Leung DKY, Sze LCY, Kwok W, et al. Health and social care service utilisation and associated expenditure among community-dwelling older adults with depressive symptoms. Epidemiol Psychiatr Sci 2021 Feb 02;30:10-30e. [doi: 10.1017/s2045796020001122]
64. Scuteri D, Morrone L, Rombolà L, Avato P, Bilia A, Corasaniti M, et al. Aromatherapy and aromatic plants for the treatment of behavioural and psychological symptoms of dementia in patients with Alzheimer's disease: clinical evidence and possible mechanisms. Evid Based Complement Alternat Med 2017;2017:9416305 [FREE Full text] [doi: 10.1155/2017/9416305] [Medline: 28465709]

Abbreviations

GAD-7: 7-item General Anxiety Disorder
HGS: hand grip strength
MMSE: Mini-Mental State Examination
MSSE: multisensory stimulation environment
SS: sensory stimulation
SWLS: Satisfaction with Life Scale
TS: therapist support
WHO: World Health Organization

Edited by E Wethington, A Dominello, I Kronish, J Kaye; submitted 30.11.23; peer-reviewed by M D'Amico, V Rashedi, NF Agbangla; comments to author 01.07.24; revised version received 30.10.24; accepted 05.11.24; published 14.01.25.

Please cite as:
Khatib S, Palgi Y, Ashar YK, Polyvyannaya N, Goldstein P
The Combined Effect of Multisensory Stimulation and Therapist Support on Physical and Mental Health of Older Adults Living in Nursing Homes: Pilot Randomized Controlled Trial
J Med Internet Res 2025;27:e55042
URL: https://www.jmir.org/2025/1/e55042
doi:10.2196/55042
PMID:39808474

©Sewar Khatib, Yuval Palgi, Yoni K Ashar, Natalya Polyvyannaya, Pavel Goldstein. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patient-Centered mHealth Intervention to Improve Self-Care in Patients With Chronic Heart Failure: Phase 1 Randomized Controlled Trial

Spyros Kitsiou[1], PhD; Ben S Gerber[2], MD, MPH; Susan W Buchholz[3], PhD, RN; Mayank M Kansal[1], MD; Jiehuan Sun[1], PhD; Susan J Pressler[4], PhD, RN

[1]University of Illinois Chicago, Chicago, IL, United States

[2]University of Massachusetts Chan Medical School, Worcester, MA, United States

[3]Michigan State University, East Lansing, MI, United States

[4]Indiana University, Indianapolis, IN, United States

Corresponding Author:
Spyros Kitsiou, PhD
University of Illinois Chicago
1919 W Taylor Street
MC 530
Chicago, IL, 60612
United States
Phone: 1 312 355 3519
Email: skitsiou@uic.edu

Abstract

Background: Heart failure (HF) is one of the most common causes of hospital readmission in the United States. These hospitalizations are often driven by insufficient self-care. Commercial mobile health (mHealth) technologies, such as consumer-grade apps and wearable devices, offer opportunities for improving HF self-care, but their efficacy remains largely underexplored.

Objective: The objective of this study was to examine the feasibility, acceptability, safety, and preliminary efficacy of a patient-centered mHealth intervention (iCardia4HF) that integrates 3 consumer mHealth apps and devices (Heart Failure Health Storylines, Fitbit, and Withings) with a program of individually tailored SMS text messages to improve HF self-care.

Methods: We conducted a phase 1 randomized controlled trial. Eligible patients had stage C HF, were aged ≥40 years, and had New York Heart Association (NYHA) class I, II, or III HF. Patients were randomly assigned to either iCardia4HF plus usual care or to usual care only and were observed for 8 weeks. Key feasibility measures were recruitment and retention rates. The primary efficacy outcome was change in HF self-care subscale scores (maintenance, symptom perception, and self-care management) at 8 weeks, assessed with the Self-Care Heart Failure Index (SCHFI; version 7.2). Key secondary outcomes were modifiable behaviors targeted by the intervention (health beliefs, self-efficacy, and HF knowledge), health status, and adherence to daily self-monitoring of 2 core vital signs (body weight and blood pressure).

Results: A total of 27 patients were enrolled in the study and randomly assigned to iCardia4HF (n=13, 48%) or usual care (n=14, 52%). Of these 27 patients, 11 (41%) in the intervention group (iCardia4HF) and 14 (52%) in the usual care group started their assigned care and were included in the full analysis. Patients' mean age was 56 (SD 8.3) years, 44% (11/25) were female, 92% (23/25) self-reported race as Black, 76% (19/25) had NYHA class II or III HF, and 60% (15/25) had HF with reduced left ventricular ejection fraction. Participant retention, completion of study visits, and adherence to using the mHealth apps and devices for daily self-monitoring were high (>80%). At 8 weeks, the mean group differences in changes in the SCHFI subscale scores favored the intervention over the control group: maintenance (Cohen *d*=0.19, 95% CI –0.65 to 1.02), symptom perception (Cohen *d*=0.33, 95% CI –0.51 to 1.17), and self-care management (Cohen *d*=0.25, 95% CI –0.55 to 1.04). The greatest improvements in terms of effect size were observed in self-efficacy (Cohen *d*=0.68) and health beliefs about medication adherence (Cohen *d*=0.63) and self-monitoring adherence (Cohen *d*=0.94). There were no adverse events due to the intervention.

Conclusions: iCardia4HF was found to be feasible, acceptable, and safe. A larger trial with a longer follow-up duration is warranted to examine its efficacy among patients with HF.

Trial Registration: ClinicalTrials.gov NCT03642275; https://clinicaltrials.gov/study/NCT03642275

XSL•FO
RenderX

(J Med Internet Res 2025;27:e55586) doi:10.2196/55586

KEYWORDS

mHealth; app; digital health; telehealth; text messaging; smartphone; wearable electronic devices; heart failure; self-care; self-management; randomized controlled trial; cardiology; SMS

Introduction

Heart Failure and Self-Care

Heart failure (HF) is a complex chronic condition with symptoms and signs that result from any structural or functional impairment of ventricular filling or ejection of blood [1]. HF affects >64 million people worldwide [2]. In the United States, approximately 6.7 million Americans aged >20 years have HF, and recent estimates indicate that by 2030, the prevalence of HF will increase to 8.5 million [3]. Black Americans have the highest incidence of HF among all racial and ethnic groups and approximately 45% greater risk of death or decline in functional status than White Americans [4]. The risk of developing HF is similar for women and men, but women are underrepresented in HF research [5].

Although considerable progress has been made over the years in both invasive and noninvasive treatments for the management of the disease [6], HF outcomes remain poor. HF is a leading cause of hospitalization and death among middle-aged and older adults in the United States [7,8]. Nearly 1 in 4 patients are readmitted within 30 days, and approximately 50% are readmitted within 6 months following an HF hospitalization [8-10]. Hospital readmissions result in significant costs to the health care system and are directly linked to poor health-related quality of life and depression [11,12]. The total cost for HF was US $30.7 billion in 2012 and is expected to reach US $69.7 billion by 2030 [13]. Therefore, attempts to decrease its societal and economic burden have become a major public health priority.

Self-care is considered essential for patients with HF, and therefore, improving self-care has increasingly become a major focus of multidisciplinary HF management programs and research worldwide [14]. HF self-care refers to behaviors and actions centered on 3 separate but linked concepts: *maintenance*, *symptom perception*, and *self-care management* [15]. The first concept, *maintenance*, refers to those behaviors that patients incorporate in their daily lives to maintain physiological stability and well-being (eg, taking medications as prescribed, following a heart-healthy diet, restricting sodium intake, and being physically active). *Symptom perception* involves daily self-monitoring and recognition of HF signs and symptoms (eg, sudden weight gain, increased blood pressure (BP), fatigue, shortness of breath, and chest pain) for early detection of deterioration. *Self-care management* refers to the strategies that one can use to respond to HF signs and symptoms when they occur (eg, adjusting diuretics, restricting fluid intake, adapting diet, and calling a care provider for guidance) to prevent hospital admission and increased mortality. A growing body of literature has shown that patients with HF who consistently engage in adequate HF self-care are more likely to be clinically stable, have better health-related quality of life, and have better event-free survival [16-20]. However, self-care in patients with HF is frequently poor, especially among racial and ethnic minority patients due to socioeconomic factors that account for delays in seeking treatment for worsening symptoms, inadequate access to health care, noncompliance with follow-up appointments, and poor adherence to recommended treatments [1,4]. Even patients recently discharged from the hospital due to acute decompensated HF (ADHF), where they presumably received patient education about HF, demonstrate low rates of adherence to basic HF self-care behaviors, such as taking HF medications as prescribed, following a low-sodium diet, and self-monitoring for HF signs and symptoms [21]. Multidisciplinary HF disease management programs that involve hospital or clinic intervention (eg, regular follow-up visits and patient education) by a multidisciplinary team of cardiologists, advanced practice nurses, pharmacists, dieticians, or other HF specialists have been successful at reducing mortality and HF-related hospitalizations. However, such programs are not available to all patients and have limited reach due to limited health care resources and accessibility constraints [22]. Therefore, interventions have evolved toward a proactive, real-time technological model to better monitor and assist patients with HF self-care and management of symptoms at home. These interventions range from structured telephone support and noninvasive home telemonitoring [23,24] to remote patient monitoring (RPM) with implantable devices [25,26] and, more recently, the use of mobile health (mHealth) technologies for patient education [27] and HF self-care support [28-31].

mHealth Interventions

mHealth, which is defined as the use of mobile computing and communication technologies in health care and public health, is a rapidly expanding area within biomedical and health informatics with great potential for developing and delivering patient-centered behavioral interventions to improve self-care and chronic disease management [32-34]. The term *patient-centered* is used in this paper to refer to interventions that focus explicitly on the patient and in which the content of the intervention is selected or tailored to address the patient's salient characteristics (eg, health beliefs, knowledge, confidence, cultural background, and preferences) [35]. With the uptake of smartphone ownership among adults in the United States [36], there is a growing opportunity to capitalize on the use of mHealth apps, wearable sensor devices, and other connected health technologies to develop patient-centered HF self-care interventions that are more accessible and scalable. As patients carry their smartphones with them wherever they go, patient-centered mHealth interventions can fit more seamlessly into their daily routine and lifestyle to support their needs.

Consumer-grade mHealth technologies hold promise for engaging patients in better HF self-care and delivering patient-centered intervention, but their efficacy remains underexplored. A recent systematic review and meta-analysis

XSL•FO
RenderX

found that mHealth interventions for patients with HF reduced the risk of all-cause mortality and HF hospitalizations [29]. However, results were mainly driven by studies that implemented mHealth as part of a larger RPM system that focused on daily transmission of patient-generated health data to a clinical care team for review and delivery of actionable feedback. Fewer studies tested the effects of stand-alone mHealth interventions (without RPM) focused on improving self-care, and virtually no studies tested the efficacy of commercially available mHealth apps, despite their increasing availability and uptake [28]. Furthermore, none of the interventions were tailored to the patients' characteristics or needs, which is often one of the reasons for low engagement with and abandonment of mHealth apps and devices [37,38].

Study Objectives

To address this important knowledge gap, our team developed and tested a patient-centered HF self-care intervention (iCardia4HF) that integrates 3 commercial mHealth apps and devices (MyApps) with a program of individually tailored SMS text messages (Text4HF) to improve self-care in patients with HF. Tailoring of the text messages was guided by participants' health beliefs, knowledge, and confidence in performing HF self-care. The intervention was delivered through an innovative digital health platform (iCardia) [39], which allows for remote collection of patient-generated health data from the devices and apps and delivery of the tailored text messages. The objective of this study was to test the feasibility, acceptability, and safety of the iCardia4HF intervention and to determine the effects of the intervention on HF self-care (primary outcome) and other health outcomes over 8 weeks, compared with usual care.

Theoretical Basis and Development of the Intervention

Details regarding the structure, content, and delivery of the intervention are provided in the Methods section. Here, we briefly describe the theoretical basis and development of the intervention.

Figure 1 presents the conceptual framework that formed the basis for the iCardia4HF intervention. The intervention is based on the situation-specific theory of HF self-care [15] and the Health Belief Model [40]. In the situation-specific theory, HF self-care is defined as a naturalistic decision-making process involving routine behaviors and actions that maintain physiological stability *(maintenance)*, facilitate self-monitoring and identification of HF symptoms *(symptom perception)*, and direct the patient's response to those symptoms *(self-care management)*. The Health Belief Model proposes that individuals' performance of a specific self-care behavior (eg, medication adherence or following a sodium-restricted diet) is determined by their beliefs regarding susceptibility to the disease, seriousness of the disorder, and perceived benefits and barriers of performing that self-care behavior. In addition to health beliefs, HF self-care decisions are influenced by knowledge about the disease (HF knowledge) [41,42] and confidence (self-efficacy) in performing HF self-care [43,44]. Confounding HF self-care decisions are contextual factors [21,45] which include the individual characteristics of a person (eg, age, gender, race, and ethnicity), the illness-related factors (eg, HF severity, comorbidities, and depression), and social determinants of health (eg, education, employment, health insurance, and income).

Figure 1. Conceptual framework providing the basis for the intervention. HF: heart failure.

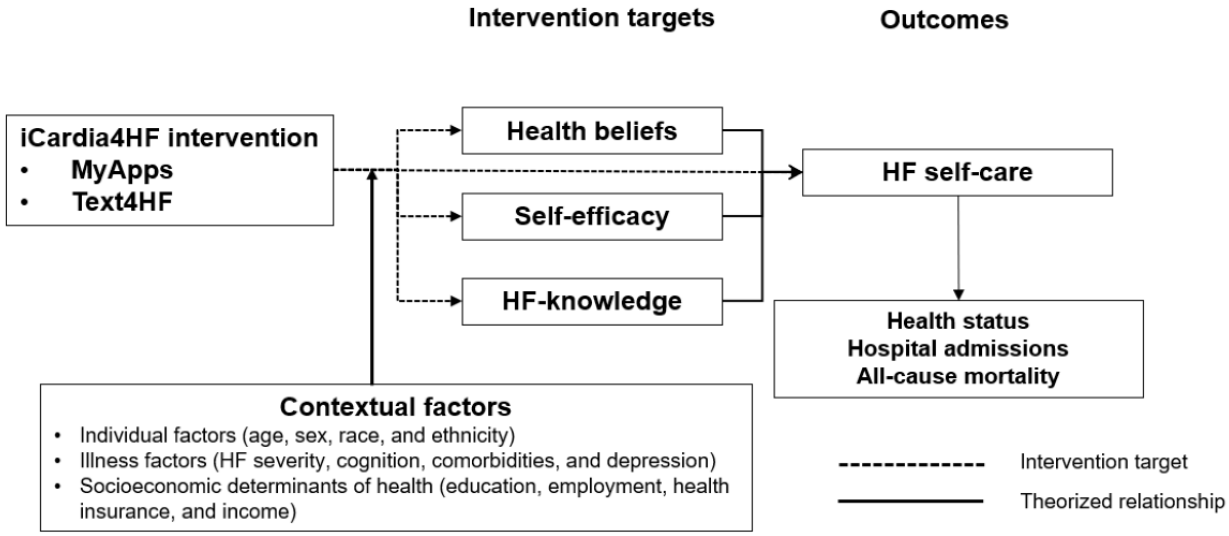

The iCardia4HF intervention was developed iteratively following a user-centered design process [46], informed by the needs of key stakeholders (patients, clinicians, and researchers) and our team's expertise in mHealth and chronic disease management. For the MyApps component of the intervention, we selected 3 popular consumer-grade mHealth apps (Heart Failure Health Storylines [Self-Care Catalysts Inc], Withings Health Mate [Withings], and Fitbit [Google LLC]) and 3 connected health devices (Withings Body Cardio Scale, Withings Blood Pressure cuff, and Fitbit Charge 2 activity tracker). These apps and devices incorporate core functionalities and electronic tools (eg, adherence reminders, vital signs monitoring, symptoms tracking, and medication tracking) that support the main aspects of HF self-care (ie, maintenance,

XSL•FO
RenderX

symptom perception, and self-care management). For the selection of the mHealth apps and connected health devices that make up the MyApps intervention component, we relied on a systematic review of 30 HF self-care apps [28] and a cross-sectional survey study of 100 patients with HF [36] in which we assessed attitudes and perceptions toward mHealth technologies, including current use and preferences for mHealth apps and devices. We also relied on cumulative experience gained from other interventional studies we have conducted in people with cardiovascular disease and other chronic diseases, in which we used Fitbit and Withings devices for vital signs monitoring and promotion of physical activity and healthy lifestyle behaviors [47-50]. Fitbit devices are increasingly used in research [51] and have been noted to be valid, reliable, and user-friendly in monitoring physical activity [52,53]. Similarly, Withings scales and BP monitors are accurate and have received 510 (k) clearance by the US Food and Drug Administration. By using these mHealth apps and devices regularly, patients with HF can become more aware of how their bodies work and what is normal, identify health changes that may need medical attention, and stay motivated while making healthier lifestyle choices.

The Text4HF component of the intervention consists of individually tailored text messages that aim to promote HF self-care adherence by targeting patients' health beliefs about HF self-care (ie, perceived barriers and benefits to prescribed medications, sodium-restricted diet, self-monitoring of HF symptoms, and daily physical activity), including HF knowledge and self-efficacy. For the development of the Text4HF intervention component, we relied on and built upon the work of Dr Susan Pressler, an internationally recognized expert in health behavior interventions for patients with HF. First, we translated the original messages that were developed and tested in the *Heart Messages* [54] study and in the *Promoting Understanding and Management through Partnering: U and Your Physician* [55] study, into SMS text messages. Then, we developed additional text messages to strengthen the intervention. Finally, we refined the text messages based on clinical guidelines for the management of HF [56] and stakeholder input to ensure accuracy, readability, and cultural relevance [57].

Methods

Study Design

iCardia4HF was a phase 1 (feasibility and pilot) randomized controlled trial [58] with 2 arms (intervention and control), allocation concealment, and masking of outcome assessors to group assignment. The intervention duration was 8 weeks. Patient assessments were performed in person at baseline, 4 weeks (30 days), and 8 weeks (56-60 days). The primary outcome of interest was HF self-care. The duration of the intervention and selection of follow-up visits for the assessment of HF self-care and intervention target variables were chosen based on guidelines for feasibility studies and because improvements in HF self-care and other measures relevant to this study have occurred at similar times in other studies [55,58-60]. The trial was conducted between January 2019 and February 2020.

Population and Setting

Patients were recruited from the hospital and outpatient HF clinic of the University of Illinois Hospital and Health Sciences System (UI Health) in Chicago, Illinois. Inclusion criteria were (1) chronic HF (stage C); (2) New York Heart Association (NYHA) class I, II, or III; (3) aged ≥40 years; (4) ability to speak and read in English; (5) smartphone ownership; and (6) living within a 48-km radius from UI Health. Exclusion criteria were (1) on a waiting list for an implanted ventricular assist device or heart transplant; (2) advanced renal disease (hemodialysis or creatinine >4.0 mg/dL) (3) end-stage HF; (4) active cancer; (5) not able to perform self-care; (6) living in a setting other than home (eg, nursing home); (7) hospice candidate; (8) Montreal Cognitive Assessment score <22 [61]; and (9) major cognitive impairment (eg, dementia). Originally, the age eligibility criterion was set to ≥50 years in this study, but due to the large number of African American patients served by the UI Health system and the fact that on average HF occurs in Black patients 10 to 15 years earlier, we modified our study protocol to include patients aged ≥40 years [62]. On the basis of left ventricular ejection fraction (EF), HF is classified into 3 EF categories: HF with reduced EF (HFrEF), mildly reduced EF (HFmEF), and preserved EF (HFpEF), according to the EF ranges ≤40%, 41% to 49%, and ≥50%, respectively. Preserved EF now represents >50% of HF cases and can have outcomes as poor as those associated with HF with reduced EF and mildly reduced EF [1]. Therefore, we included patients from all 3 EF categories.

Sample Size

The primary purpose of a phase 1 study is to determine the feasibility of recruitment, retention, and intervention delivery and to collect preliminary data on the primary outcome measure to inform the design and sample size calculation for a phase 2 trial. Hence, an efficacy-based power analysis was beyond the scope of this feasibility study [63]. During the planning stages of the trial (grant proposal submission) and in the original protocol that we published in clinicaltrials.gov before the start of recruitment, we aimed at enrolling 40 patients (an average of 4 patients per month over a 10-month recruitment period), using the general rule of thumb of having 15 to 20 patients per group to obtain an estimate of the variance for the primary outcome. However, as the study progressed and we collected recruitment data, we revised the original accrual plan, taking into consideration the eligibility fraction (proportion of potential participants who undergo screening and are eligible to enroll), the enrollment fraction (proportion of people who are eligible for participation and who actually enroll for the trial), and the funding period for the study (12 months). Consistent with recommendations that a sample size of 12 participants per group is suitable for a phase 1 trial [64], we adjusted the anticipated enrollment in the study to 28 participants, accounting for an expected attrition rate of 15%. The revised justifications for this sample size were based on rationale about feasibility and precision about the mean and variance, as described in the study by Julious [64]. Specifically, we considered the degrees of

XSL•FO
RenderX

freedom required to ensure the 95th percentile for the variance has 50% power.

Screening and Recruitment Process

Patient screening and recruitment were performed in 2 stages. Stage 1 involved querying the hospital electronic medical record to identify and screen potentially eligible patients who were either admitted to the hospital due to HF or had an upcoming appointment at the outpatient HF clinic for follow-up. Patients found eligible on initial screening were approached in person by a study researcher who briefly described the study, invited participation, and further assessed patients for eligibility (stage 2). Eligible patients agreeing to participate provided written informed consent, and Health Insurance Portability and Accountability Act (HIPAA) authorization and were scheduled for baseline assessment and study orientation.

Randomization and Allocation Concealment

After completion of baseline assessment, patients were randomly assigned 1:1 to the intervention or the control group. For the randomization, we used the biased-coin minimization method by Pocock and Simon [65] with base probability set to 0.7 to achieve balance of important clinical covariates (age, biological sex, and NYHA functional class) between the 2 study groups. Randomization was performed centrally by a designated staff person using QMinim [66], password-protected web-based software hosted in a secure server at the University of Illinois Chicago. The person responsible for the randomization was not involved in patient recruitment and enrollment and did not have the ability to influence the execution of the randomization procedure. Investigators enrolling participants did not have access to the system or knowledge of the minimization algorithm and therefore could not foresee assignments.

iCardia4HF Intervention Group

Patients in the intervention group received the MyApp and Text4HF components of the iCardia4HF intervention, in addition to usual care. In the subsequent sections, we describe the content, structure, and delivery of the 2 intervention components.

Consumer mHealth Apps and Devices (MyApps)

Overview

MyApps consisted of 3 popular consumer-grade mHealth apps (Heart Failure Health Storylines [version 7.17], Withings Health Mate [version 4.1.1], and Fitbit [version 2.86]), and 3 connected health devices (Withings Body Cardio Scale [version 1751], Withings Blood Pressure cuff [version BP-801], and Fitbit Charge 2 activity tracker [version 22.58.0]). The mobile apps and devices interface with each other, offering patients an integrated set of self-monitoring and behavior change tools that support the 3 core elements of HF self-care (maintenance, symptom perception, and self-care management) [15]. There were no major changes or upgrades to the apps and devices during the study affecting their content or features. Also, no adaptations or changes were made to the consumer apps or devices used in the study as part of the intervention. Each app specializes in different preventative and healthy lifestyle behaviors, and collectively, the 3 apps and devices supported the following HF self-care activities:

Medication Adherence

Heart Failure Health Storylines has a medication tracking feature that allows patients to add the medications they are taking and schedule reminders in the form of push notifications for each medication. When a medication reminder is triggered on the phone, patients are prompted to respond whether they took their medication or not, and thus maintain a digital diary of medication adherence on their phone.

Vital Signs Monitoring

HF self-care requires daily monitoring of weight to check for weight gain caused by increased fluid [1]. It also requires lowering BP to make it easier for the heart to pump blood [1]. Monitoring of weight and body composition (fat mass, muscle mass, water mass, bone mass, and BMI) is captured by the Withings Body Cardio scale. BP, including heart rate, is captured by the Withings Blood Pressure monitor. Both devices interface with the Withings Health Mate app via Bluetooth and automatically transmit captured data whenever a new measurement is taken. For weight and body composition, the app provides patients with visual line graphs along with textual feedback to notify them of any fluctuations over time (eg, gaining weight +2.2kgs since last measurement or over a selected period). Similarly, BP measurements (systolic and diastolic) are presented in the app in tabular format and also as a line graph, using a color-coding scheme to classify each measure as normal (<120/80 mm Hg), elevated (120−129/<80 mm Hg), high BP or stage 1 hypertension (130-139/80-89 mm Hg), high BP or stage 2 hypertension (140-179/90-119 mm Hg), or hypertensive crisis (≥180/≥120 mm Hg), according to clinical guidelines [67]. In case BP>180/110 mm Hg, the app advises the patient to seek emergency care right away.

HF Symptom Monitoring

In addition to taking medications and tracking weight and BP, HF self-care involves monitoring HF symptoms. The Heart Failure Health Storylines app has a feature that allows patients to record the presence and severity of HF symptoms over time, using a 10-point Likert scale. Symptoms are timestamped and displayed in a weekly calendar format using a color-coded scheme. Symptom questions programmed in the app at baseline included chest pressure; coughing; difficulty in sleeping; swelling in the abdomen; swelling of legs, feet, and ankle; shortness of breath when lying down or at rest; shortness of breath with activity; and fatigue. Patients were asked to perform symptom tracking daily.

Sodium-Restricted Diet

Patients with HF typically need to restrict sodium intake to avoid fluid retention and episodes of ADHF. The Heart Failure Health Storylines app provides patients with generic information about heart healthy eating and guidelines for lowering sodium intake.

Physical Activity

The Fitbit app and Charge 2 activity tracker provide patients with several tools that support daily self-monitoring of steps, intensity of physical activity, heart rate, and sleep. Also, Fitbit incorporates a variety of behavior change tools, such as goal setting, activity visualization with feedback, rewarding messages

XSL•FO
RenderX

when step goals are met, and prompts to move when sedentary (<250 steps) for >50 minutes.

All patient-generated health data captured by the connected health devices and mHealth apps are presented in the "My Storylines" dashboard of the Heart Failure Health Storylines app as visual graphs in a weekly calendar format with color-coded schemes. This helps patients identify preclinical measures of ADHF between time periods, correlate lifestyle behaviors with changes in their health, and modify their behavior accordingly.

We used a goal-based approach to promote self-care via the mHealth tools described earlier in concurrence with clinical guidelines for patients with HF. The primary goals communicated to patients at baseline were (1) daily self-monitoring of weight and BP each morning before breakfast and medications, (2) self-recording of HF symptoms, (3) recording of medication taking by replying (yes or no) to the push notifications, (4) gradual increase of physical activity to reach 20 to 30 minutes of moderate-to-vigorous physical activity (MVPA) per day most days of the week, and (5) following a low-sodium diet (2000-3000 mg of sodium/day).

Individually Tailored Text Messages (Text4HF)

Text4HF is a program of individually tailored text messages targeting patients' health beliefs (perceived barriers and benefits), self-efficacy, and knowledge about HF self-care. Patients in the intervention group received 4 text messages per week, at a day and time of their preference, via SMS. Tailoring of the text messages was guided by patients' responses to valid and reliable assessments of intervention target variables assessed at baseline and after 4 weeks . Items on the Health Belief scale are divided into benefit and barrier questions regarding adherence to prescribed medications, sodium-restricted diets, and self-monitoring of weight and HF symptoms. Each item on the scale received a 5-point score ranging from 1 (strongly disagree) to 5 (strongly agree). Participants who scored ≥3 on a barrier question or ≤3 on a benefit question received text messages tailored to that specific barrier or benefit item (Table 1). If a participant scored >3 on a benefit question or <3 on a barrier question, then a message in relation to those items was not sent because it was presumed that the patient already understood the barriers or benefits identified in that question. A similar approach was followed for the other 2 intervention target variables: self-efficacy and HF knowledge. For example, patients who gave an incorrect answer on an item of the Dutch Heart Failure Knowledge Scale received a patient education message in relation to that item.

XSL•FO
RenderX

Table 1. Sample tailored text messages based on responses to validated questions.

Intervention target	Question	Rating scale	Decision algorithm	Example text message
Medication adherence	Benefit question: if I take my water pills, I will lower my chance of being in the hospital.	5-point Likert (1 is strongly disagree and 5 is strongly agree)	If answer ≤3, then send text message	Taking water pills as prescribed can help remove extra water from the body and lessen the risk of getting hospitalized
Medication adherence	Barrier question: taking water pills makes it hard to go away from home.	5-point Likert (1 is strongly disagree and 5 is strongly agree)	If answer ≥3, then send text message	If taking water pills makes it hard to go away from home, one option is to take it several hours before you plan to go out or wait until after you return home to take it.
Diet adherence	Barrier question: food does not taste good on the low salt diet.	5-point Likert (1 is strongly disagree and 5 is strongly agree)	If answer ≥3, then send text message	You can flavor your food without using salt. You can use seasonings like pepper, lemon juice, garlic, onion powder, and basil.
Self-monitoring adherence	Barrier question: weighing myself is unpleasant.	5-point Likert (1 is strongly disagree and 5 is strongly agree)	If answer ≥3, then send text message	Although weighing yourself daily might be unpleasant, it is important to do so because it can help you identify fluid buildup in your body.
Self-care self-efficacy	How confident are you that you can keep yourself stable and free of symptoms	5-point Likert (1 is not confident and 5 is extremely confident)	If answer ≤3, then send text message	(Message 1) Following your doctor's advice and taking your medications as directed are important steps to managing your condition. (Message 2) Tell your doctor right away if you gain 3 or more pounds in a day or feel swelling in your feet, ankles, and other parts of your body or if it is hard to breathe.
Heart failure knowledge	What is the best thing to do in case of increased shortness of breath or swollen legs?	Multiple choice answer	If answer is other than "call the doctor or nurse," then send a text message	Call your health care provider if you feel any of the following: shortness of breath when lying flat or standing and swelling in your legs or stomach.

Control Group

Patients allocated to the control group received usual care provided through the outpatient HF clinic and primary care. Guideline-directed medical therapy, included initial and serial evaluation of HF (eg, physical examination, diagnostic tests, biomarkers, vital signs, symptoms, left ventricular EF [LVEF], volume status, weight, jugular venous pressure, and presence of edema); pharmacological treatment with diuretics, angiotensin-converting enzyme inhibitors, Angiotensin II receptor blockers, β blockers, and aldosterone receptor antagonists; nonpharmacological intervention (eg, patient self-care education and referral to cardiac rehabilitation); care coordination, and regular follow-up appointments with a cardiologist and Advanced Practice Registered Nurse at the Heart Failure Clinic 7 days postdischarge and every 2 to 3 months thereafter for further assessment and treatment.

Study Procedures

All study procedures were carried out in private examination rooms by certified key research personnel with training on human subjects research and HIPAA. After completion of baseline assessment and randomization, patients allocated to the intervention group received instructions and in-person training on how to use the mHealth apps and devices using the teach-back method. A research assistant downloaded the intervention apps on each participant's smartphone and then used a deidentified email address created specifically for this study to sign each patient into the intervention apps and pair the apps with their respective device via Bluetooth. Next, participants were shown how to use each app, sync the study devices with their personal smartphones, and view their data using the apps. Participants were introduced to the "position control" technology implemented in the Withings devices and were shown how to use the on-screen visual indicators to center their body on the scale and properly place the BP cuff on their

XSL•FO
RenderX

arm to ensure consistent and accurate measurements. As part of our study protocol and in accordance with clinical guidelines about HF self-care, patients were advised to use the mHealth devices and apps to record their weight, BP, and symptoms every morning after going to the bathroom and before eating breakfast and taking their medications. Daily reminders in the form of push notifications for medications, symptom tracking, and monitoring of weight and BP were programmed in each app according to patients' daily routines and preferences. To reduce complexity and participant burden, medication reminders were set up for 1 medication only (angiotensin-converting enzyme inhibitor, loop diuretic, or β-blocker) based on the documented regimen in the patient's electronic medical record and discussion with the patient at baseline. Patients were trained how to edit existing medications or add new ones and how to change the time of medication reminders if needed. At the end of the training session, participants were provided with an end user's manual that contained instructions on how to use each app and device and a copy of their study email and password for logging into the apps. Also, they were provided with a phone number to contact our team in case they experienced technical issues or had study-related questions. During training we emphasized to patients that the intervention was not a substitute for usual care and did not serve as a RPM system for identifying potential exacerbations or reporting medical emergencies.

The iCardia platform [39] was used to remotely collect all patient-generated health data captured by the mHealth apps and devices and to send the intervention text messages. A trained research assistant (interventionist) programmed the text messages in iCardia after completion of baseline and 4-week visits. iCardia is a secure, web-based system that is hosted on a HIPAA-compliant server at the University of Illinois Chicago and has been validated through multiple trials [47,49,50,68-71].

Masking of Outcome Assessors

By design, group assignments were identifiable to study participants, but outcome assessors and care providers treating patients were masked to group assignment. Outcome assessors worked in different office space than other study researchers who were aware of the group allocation. Before outcome assessment, patients were asked not to reveal to outcome assessors which group they were assigned to and were also asked to remove the Fitbit from their wrist to conceal their allocation. All outcome assessors were trained and continuously monitored.

Study Measures and Data Collection Schedule

Overview

Study participants were observed for 8 weeks. Table 2 lists all the measures and data collection schedule by outcome of interest. Once consent was obtained, participants provided basic demographic information (age, sex, race, ethnicity, employment status, marital status, living arrangement, education, and health insurance), and completed the Montreal Cognitive Assessment test [61] and Patient Health Questionnaire [72]. Clinical data, such as severity of HF (eg, LVEF and NYHA), primary cause of HF, comorbidities, BP, and anthropometrics (height, weight, and BMI), were retrieved from the electronic medical record at baseline. Study visits and assessments occurred in person at baseline, 4 weeks, and 8 weeks. Data from the mobile apps and devices were collected daily during the 8-week follow-up period.

XSL•FO
RenderX

Table 2. List of measures and data collection schedule.

Outcomes	Measures	Baseline	4 weeks	8 weeks
Demographics	Age, sex, race, ethnicity, marital status, employment status, education, financial status, living arrangement, and health insurance	✓		
Clinical characteristics	LVEF[a], NYHA[b] class, BP[c], weight, height, and BMI.	✓		
Cognitive function	Montreal Cognitive Assessment test [61]	✓		
Feasibility (recruitment and retention)	Number of patients screened, number of eligible and ineligible patients, reasons for exclusion, recruitment rate, and retention rate	Monitored throughout the study	Monitored throughout the study	Monitored throughout the study
Acceptability	Technology acceptance questionnaire (perceived usefulness, ease of use, user satisfaction, confirmation of initial expectations, and intention to continue using the intervention apps and devices)			✓
HF[d] self-care (primary)	Self-Care Heart Failure Index (version 7.2)	✓	✓	✓
Self-efficacy	Self-Care Self-Efficacy scale	✓	✓	✓
Health beliefs	Beliefs about Medication Adherence Scale, Beliefs about Diet Adherence Scale, Beliefs about Self-Monitoring Adherence Scale	✓	✓	✓
HF-knowledge	Dutch Heart Failure Knowledge Scale	✓	✓	✓
Health status	Kansas City Cardiomyopathy Questionnaire	✓	✓	✓
Health care encounters	Hospitalizations and ER[e] visits were identified and recorded using the electronic medical record system and patient interviews.	✓	✓	✓
Mortality	Death certificate or other report in the electronic medical record system	✓	✓	✓
Self-monitoring adherence	Daily weighing and BP monitoring—timestamped data from the Withings Body Cardio Scale and BP cuff	Monitored throughout the study	Monitored throughout the study	Monitored throughout the study
Steps	Mean number of daily steps were measured with the Fitbit Charge 2 activity tracker.	Monitored throughout the study	Monitored throughout the study	Monitored throughout the study
MVPA[f] minutes	Mean number of MVPA minutes/day were measured with the Fitbit Charge 2.	Monitored throughout the study	Monitored throughout the study	Monitored throughout the study

[a]LVEF: left ventricular ejection fraction.

[b]NYHA: New York Heart Association.

[c]BP: blood pressure.

[d]HF: heart failure.

[e]ER: emergency room.

[f]MVPA: moderate-to-vigorous physical activity.

XSL•FO
RenderX

Feasibility Outcome Measures

Feasibility was assessed in the domains of patient recruitment and retention during the 8-week follow-up period. Feasibility measures assessed throughout the study included the following: (1) number of patients screened, (2) number of eligible patients, (3) reasons for ineligibility, (4) recruitment rate (mean number of patients recruited per month), and (5) retention rate (percentage of randomized participants who completed the study).

Primary Efficacy Outcome Measure

The primary efficacy outcome measure was self-reported HF self-care assessed with the Self-Care Heart Failure Index (SCHFI; version 7.2) [15] at baseline, 4 weeks, and 8 weeks (Table 2). SCHFI is a validated instrument that contains the following three subscales: (1) Self-care Maintenance (10-items), (2) Symptom Perception (9-items), and (3) Self-care management (10-items). Each subscale is scored separately. Response choices for all items in each subscale are summed and standardized to achieve a possible score of 0 to 100. Higher scores indicate better HF self-care. Scores <70 indicate insufficient self-care. The minimal clinically important change in SCHFI is one-half SD, or an 8-point difference in the standardized score of each subscale [73].

Secondary Outcome Measures

Secondary outcome measures (Table 2) included self-efficacy, health beliefs, HF-knowledge, health status, adherence to daily self-monitoring of weight and BP, physical activity, acceptance of the intervention, and adverse events.

Self-Efficacy

Confidence in performing routine HF self-care behaviors was assessed with the Self-Care Self-Efficacy questionnaire that has 10 items [15,74]. Similar to SCHFI, items in the Self-Care Self-Efficacy scale are summed and standardized to achieve a possible score of 0 to 100. Higher scores indicate greater confidence in HF self-care. Scores <70 indicate inadequate self-efficacy. The minimal clinically important change is one-half SD, or an 8-point difference in the standardized score [73].

Health Beliefs

Perceived benefits and barriers about HF self-care were assessed using three validated scales: (1) Beliefs about Medication Adherence Scale (12 items) [75], (2) Beliefs about Diet Adherence Scale (12 items) [75], and (3) Beliefs about Self-Monitoring Adherence Scale (12 items) [76].

HF Knowledge

The Dutch Heart Failure Knowledge Scale was used to measure general knowledge about HF, knowledge about HF treatment (including diet and fluid restriction) and symptoms, and symptom recognition [77]. The Dutch Heart Failure Knowledge Scale is a 15-item questionnaire that has a minimum score of 0 (no knowledge) and a maximum score of 15 points (optimal knowledge).

Health Status

The short version of the Kansas City Cardiomyopathy Questionnaire 12-item (KCCQ-12) was used to assess health status [78]. KCCQ-12 is a validated health status measure for patients with HF that contains 4 subscales: physical limitation, symptom frequency, quality of life, and social limitation. Each subscale provides an individual score from 0 to 100, with 0 denoting the worst and 100 the best possible health status. The mean of the 4 subdomain scores is presented as a summary score, with differences of ≥5 points considered clinically important [79].

Self-Monitoring Adherence

Daily adherence to self-monitoring of weight and BP was assessed using timestamped data automatically transmitted from the Withings devices and app to the iCardia server. Adherence was defined as the percentage of days patients used the Withings Body Composition scale and BP cuff along with the Health Mate app to measure their weight and BP over the 8-week follow-up period.

Physical Activity (Steps and MVPA)

Daily steps and MVPA were assessed with the Fitbit Charge 2 activity tracker that participants received as part of the intervention during 8 weeks. Participants were asked to wear the Fitbit daily for ≥10 hours (600 minutes) every day during waking hours and gradually increase their physical activity over time. A valid day consisted of 10 hours of wear time and a valid week consisted of at least 4 valid days. Daily Fitbit wear time was calculated by iCardia based on the collected heart rate data using a previously validated method [47,80].

Adverse Events

As part of our safety protocol and monitoring of adverse events, we recorded all hospitalizations, emergency room visits, and deaths that occurred during the follow-up period of our trial. At each study visit, a trained research assistant reviewed the electronic medical record of each patient (encounters and provider notes) and asked them for possible hospitalizations that occurred at UI Health or elsewhere, including adverse events. The study physicians adjudicated the encounters and events per our study safety protocol.

Intervention Acceptability

Exit interviews were conducted at 8 weeks (postintervention) with patients allocated to the iCardia4HF group to assess *acceptance* and elicit their feedback about the intervention. Measures of technology acceptance—adapted to the context of this study from the Technology Acceptance Model [81] and the Expectation-Confirmation Model of continued information technology use [82,83]—were assessed using a 3-point Likert scale and included the following: (1) perceived usefulness (7 items), (2) ease of use (5 items), (3) user satisfaction (3 items), (4) confirmation of initial expectations (3 items), and (5) intention to continue using the mobile apps and devices of the iCardia4HF intervention after the end of the study (3 items).

XSL•FO
RenderX

Statistical Analysis and Interpretation

All analyses were conducted using an intent-to-treat approach (ie, participants were analyzed according to their assigned group) independent of the actual intervention dose received. All variables were checked for errant values. Descriptive statistics were computed for all items and scale scores. Distributions were examined for nonnormality and transformed if necessary. To deal with missing data, we used a multiple imputation approach [84]. Imputation of the missing values for each outcome was performed by treatment group, age, gender, BMI, NYHA, LVEF, Patient Health Questionnaire-9, and Montreal Cognitive Assessment to create 20 imputed datasets using the mice package [85] in R software (version 4.2.1; R Core Team) [86]. Analyses of between-treatment differences in primary and secondary outcomes were performed using a 2-sample *t* test to assess the intervention effect. Our approach to data reporting and interpretation regarding the intervention effects on HF self-care and secondary outcomes was focused on the direction and magnitude of treatment effect estimates, as well as the uncertainty behind each estimate (95% CI), and not on *P* values [87].

Daily weighing and BP self-monitoring adherence in the intervention group was measured by week (as the percentage of protocol-required measurements that were completed during the follow-up period), after adjusting for days outside of the hospital. For the Fitbit physical activity data, we first calculated the mean number of daily steps and daily MVPA minutes per week using all valid days for each participant. A valid day was one that had ≥600 minutes of wear time. We only included weeks that had ≥3 valid days. We also calculated the mean daily sedentary minutes and Fitbit wear time per day. To study the trajectories of the mean number of daily steps per week for the intervention group, we used linear mixed-effects models [88], where the continuous week variable was included as a fixed effects and a random intercept was included to account for the intrasubject correlations. We adopted the same analysis strategy to study the trajectories of the mean daily MVPA minutes, mean daily sedentary minutes, mean daily Fitbit wear-time, and mean daily weight and BP monitoring adherence for the intervention group. We used the Non-Linear Mixed Effects package [89] to fit the linear mixed-effects model in R.

Ethical Considerations

The study protocol was approved by the university's institutional review board, registered on ClinicalTrials.gov (NCT03642275), and reported in accordance with the CONSORT-eHEALTH checklist (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth; Multimedia Appendix 1). As compensation for their time and participation in the study, patients received US $25 for completing the baseline assessment, US $30 for completing the 4-week study visit, and US $30 for completing the 8-week study visit, which amounts to a total of US $85. They were also able to keep the apps and devices after completion of the study. In addition, they were provided with free parking and transportation reimbursement from and to University of Illinois Chicago and UI Health for study visits as needed.

Results

Description of the Population Sample

Figure 2 shows the numbers of patients who were screened for eligibility, enrolled in the trial, randomly assigned to one of the two groups, received intended treatment, completed follow-up assessments, and included in the final analysis. A total of 504 patients were assessed for eligibility through chart review and clinic recruitment. Of those, 84.3% (n=425) were ineligible, and 9.9% (n=50) patients declined to participate or be screened for participation in the study, yielding a recruitment fraction of 5.8% (29/504); proportion of potential participants who enrolled in the trial). On average, we enrolled 3 patients per month during the 10-month recruitment period (from January 2019 to November 2019). Of the 29 participants enrolled in the study, 7% (2/29) were lost to follow-up before baseline assessment and randomization. The 27 participants who completed baseline assessment were randomized to the intervention (n=13, 45%) or the control group (n=14, 52%). Two participants allocated to the intervention group were excluded from the study because they were admitted to a nursing home and became ineligible before allocation of the mHealth app and devices.

Table 3 presents the characteristics of the 25 eligible patients who enrolled in and completed the study. The mean age of patients was 56 (range 41-72) years, and 44% (11/25) were female. Regarding self-identified race, 92% (23/25) were Black and 8% (2/25) were White. Baseline demographic and clinical characteristics were similar between the intervention and control groups.

XSL•FO
RenderX

Figure 2. Trial profile diagram. The term “no-show” refers to patients not showing up to their scheduled study visit appointment.

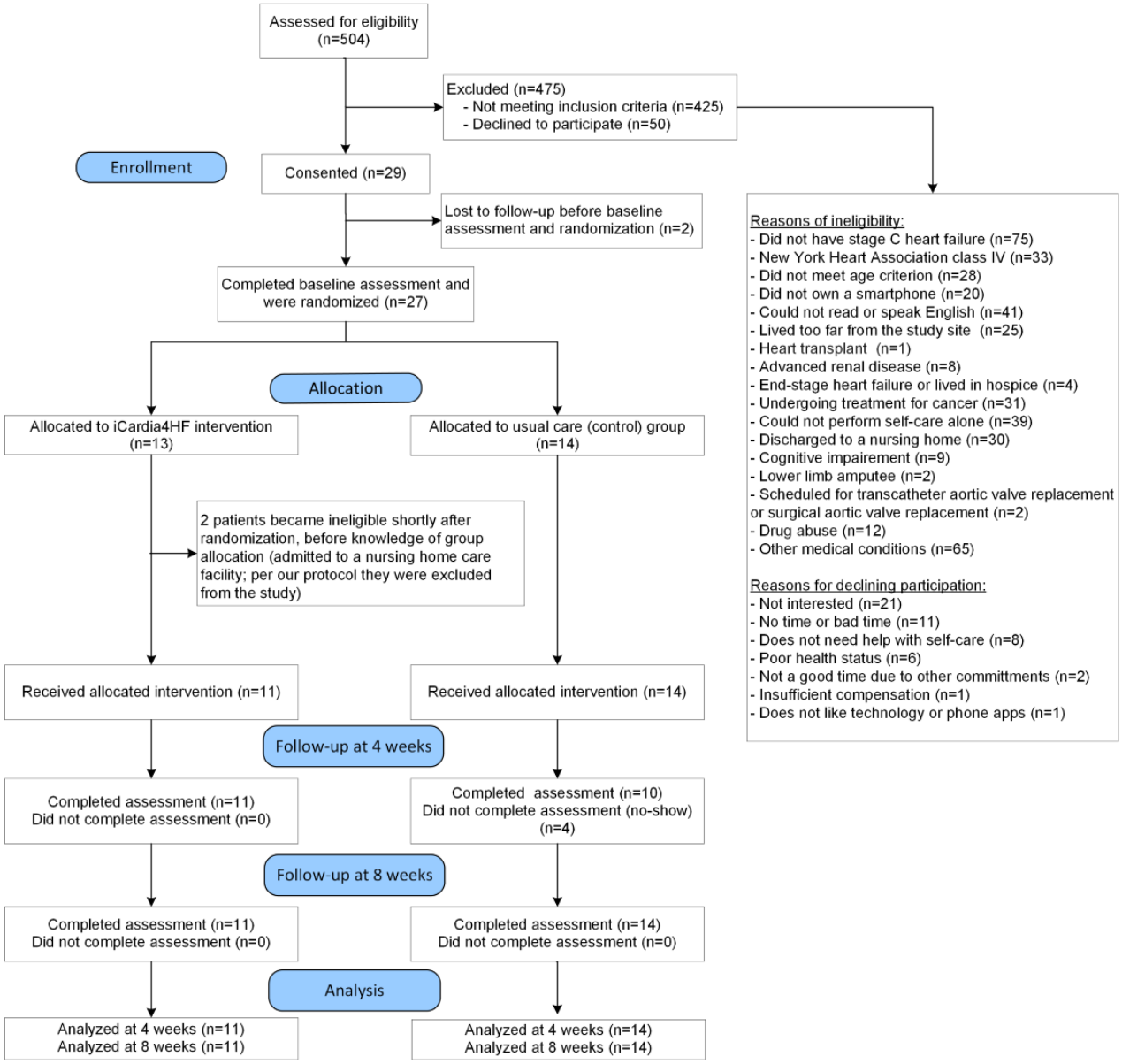

XSL•FO
RenderX

Table 3. Baseline demographics and clinical characteristics.

Variables	Overall (n=25)	Intervention (n=11)	Control (n=14)
Age (y), mean (SD)	56 (8.3)	55.4 (9.17)	56.4 (7.8)
Sex, n (%)			
Male	14 (56)	5 (46)	9 (64)
Female	11 (44)	6 (55)	5 (36)
Race, n (%)			
Black	23 (92)	10 (91)	13 (93)
White	2 (8)	1 (9)	1 (7)
Employment status, n (%)			
Working	9 (36)	3 (27)	6 (43)
Unemployed	4 (16)	3 (27)	1 (7)
Disabled	5 (20)	1 (9)	4 (29)
Retired	7 (28)	4 (36)	3 (21)
Marital status, n (%)			
Never married	12 (48)	5 (45)	7 (50)
Married	7 (28)	4 (36)	3 (21)
Divorced	2 (8)	0 (0)	2 (14)
Widowed	4 (16)	2 (18)	2 (14)
Living arrangement, n (%)			
Living alone	9 (36)	5 (46)	4 (29)
Living with partner, spouse, or family	16 (64)	6 (55)	10 (71)
Education, n (%)			
High school or less	7 (28)	4 (36)	3 (21)
Some college (no degree)	8 (32)	4 (36)	4 (29)
Community college or associate degree	4 (16)	2 (18)	2 (14)
Undergraduate degree	5 (20)	1 (9)	4 (25)
Graduate level degree	1 (4)	0 (0)	1 (7)
Health insurance, n (%)			
Private insurance	10 (40)	3 (27)	7 (50)
Medicaid	6 (24)	3 (27)	3 (21)
Medicare	4 (16)	2 (18)	2 (14)
Medicare and private insurance	3 (12)	2 (18)	1 (7)
Managed Medicare (Medicare Advantage Plan)	1 (4)	1 (9)	0 (0)
No insurance	1 (4)	0 (0)	1 (7)
NYHA [a] class, n (%)			
I	6 (24)	3 (27)	3 (21)
II	15 (60)	6 (55)	9 (64)
III	4 (16)	2 (18)	2 (14)
LVEF [b] (%)			
LVEF, mean (SD)	35.8 (12.4)	36.8 (11.5)	35.0 (13.4)
≤40	15 (60)	6 (54.5)	9 (64.3)
41-49	5 (20)	3 (27.3)	2 (14.3)

XSL•FO
RenderX

Variables	Overall (n=25)	Intervention (n=11)	Control (n=14)
≥50	5 (20)	2 (18.2)	2 (14.3)
Blood pressure (mm Hg), mean (SD)			
Systolic	136.6 (20.14)	136.6 (20.14)	141.1 (22.4)
Diastolic	82.4 (10.58)	82.4 (10.58)	83.9 (11.5)
Weight (kg), mean (SD)	110.8 (33.1)	109.2 (31.5)	112.1 (35.7)
BMI (kg/m^2), mean (SD)	36.6 (10.1)	37.7 (11.21)	35.79 (9.54)
MoCA[c] score, mean (SD)	25.92 (2.05)	26.7 (1.61)	25.29 (2.20)
PHQ-9 [d] score			
Score, mean (SD)	3.12 (3.5)	3.45 (3.4)	2.86 (3.8)
0-4 (minimal or none), n (%)	17 (68)	7 (64)	10 (71)
5-9 (mild), n (%)	6 (24)	3 (27)	3 (21)
10-14 (moderate), n (%)	2 (8)	1 (9)	1 (7)

[a]NYHA: New York Heart Association.

[b]LVEF: left ventricular ejection fraction.

[c]MoCA: Montreal Cognitive Assessment.

[d]PHQ-9: Patient Health Questionnaire.

Effects on HF Self-Care Scores

At 4 weeks, there was a clinically important difference (>8 points) between the 2 groups in maintenance (Cohen *d*=0.57, 95% CI –0.32 to 1.46) and self-care management (Cohen *d*=0.91, 95% CI –0.01 to 1.82), favoring the iCardia4HF intervention (Table 4). Symptom perception improved in both groups from baseline to 4 weeks, but there was no clinically relevant difference between the 2 groups (Cohen *d*=0, 95% CI –0.84 to 0.85).

At 8 weeks, participants in the intervention group reported, on average, greater improvement in self-care maintenance (Cohen *d*=0.19, 95% CI –0.65 to 1.02), symptom perception (Cohen *d*=0.33, 95% CI –0.51 to 1.17), and self-care management (Cohen *d*=0.25, 95% CI –0.55 to 1.04) than the control group (Table 4).

Table 4. Mean Self-Care Heart Failure Index (SCHFI) subscale scores and change over time by group.

HF[a] self-care subscales	Intervention group, mean (SD)	Change in intervention group scores from baseline, mean (SD)	Control group, mean (SD)	Change in control group scores from baseline, mean (SD)	Between-group difference, Cohen *d* (95% CI)	*P* value
Maintenance						
Baseline	69.55 (15.16)	—[b]	74.29 (17.82)	—	—	—
4 weeks	75.68 (14.79)	6.14 (15.39)	70.90 (19.29)	–3.39 (18.16)	0.57 (–0.32 to 1.46)	.17
8 weeks	75.91 (14.37)	6.36 (15.34)	78.04 (14.58)	3.75 (12.51)	0.19 (–0.65 to 1.02)	.65
Symptom perception						
Baseline	79.80 (10.40)	—	75.79 (20.63)	—	—	—
4 weeks	86.62 (12.96)	6.82 (10.04)	82.52 (18.79)	6.73 (22.61)	0 (–0.84 to 0.85)	.99
8 weeks	87.63 (11.54)	7.83 (9.85)	80.16 (18.23)	4.37 (10.93)	0.33 (–0.51 to 1.17)	.41
Self-care management						
Baseline	67.97 (10.93)	—	72.45 (16.98)	—	—	—
4 weeks	73.16 (17.27)	5.19 (14.59)	64.80 (23.41)	–7.64 (14.07)	0.91 (–0.01 to 1.82)	.03
8 weeks	73.16 (20.09)	5.19 (20.14)	73.64 (13.96)	1.19 (11.77)	0.25 (–0.59 to 1.09)	.56

[a]HF: heart failure.

[b]Not applicable.

XSL•FO
RenderX

Effects on Intervention Target Measures (Health Beliefs, Self-Efficacy, and HF Knowledge)

Assessment of health beliefs focused on perceived benefits and perceived barriers of adherence to 3 important HF self-care behaviors: medication taking, low-sodium diet, and self-monitoring.

With respect to medication adherence, there were no important differences between the intervention and control group at 4 weeks, but there was a meaningful difference (*d*=0.63) in perceived benefits at 8 weeks, favoring the iCardia4HF intervention (Table 5).

With respect to diet, there were meaningful treatment effects with the iCardia4HF intervention in perceived benefits both at 4 and 8 weeks (Cohen *d*=0.85 and 0.55, respectively), but no meaningful differences between the intervention and control group in perceived barriers (Table 5).

In terms of self-monitoring, there were positive treatment effects in perceived benefits with the iCardia4HF intervention at 4 and 8 weeks (Cohen *d*=0.47 and 0.94, respectively), compared with the control group (Table 5).

There were clinically meaningful differences in self-efficacy scores between the intervention and control group both at 4 and 8 weeks, favoring the iCardia4HF intervention (Table 5). Compared to baseline, patients in the intervention group had an improvement in self-efficacy scores at both time points, while patient scores in the control group declined.

The iCardia4HF intervention had no effect on HF knowledge at 4 weeks but trended to have a small positive effect at 8 weeks (Cohen *d*=0.21).

XSL•FO
RenderX

Table 5. Mean intervention target scores (health beliefs, self-efficacy, and heart failure [HF] knowledge) and change over time and by group.

Subscales and time points	Intervention group, mean (SD)	Change in intervention group scores from baseline, mean (SD)	Control group, mean (SD)	Change in control group scores from baseline, mean (SD)	Between-group difference, Cohen *d* (95% CI)	*P* value
Benefits of medication adherence						
Baseline	23.36 (3.41)	—[a]	24.24 (2.84)	—	—	—
4 weeks	24.09 (3.56)	0.73 (3.69)	25.26 (4.35)	1.02 (3.24)	–0.09 (–0.96 to 0.78)	.84
8 weeks	26.09 (3.30)	2.73 (3.38)	25.13 (3.53)	0.89 (2.51)	0.63 (–0.25 to 1.51)	.14
Benefits of diet adherence						
Baseline	28.36 (4.32)	—	31.71 (3.15)	—	—	—
4 weeks	30.36 (3.11)	2.00 (4.05)	30.10 (3.94)	–1.61 (4.53)	0.85 (–0.09 to 1.79)	.05
8 weeks	31.00 (2.79)	2.64 (4.03)	32.00 (3.28)	0.29 (4.43)	0.55 (–0.30 to 1.40)	.17
Benefits of self-monitoring						
Baseline	20.27 (4.82)	—	23.36 (3.65)	—	—	—
4 weeks	22.36 (5.20)	2.09 (4.01)	23.53 (5.28)	0.17 (4.13)	0.47 (–0.41 to 1.35)	.27
8 weeks	25.18 (3.16)	4.91 (4.35)	24.79 (3.40)	1.43 (3.11)	0.94 (0.06 to1.82)	.03
Barriers of medication adherence						
Baseline	14.09 (3.02)	—	12.23 (4.09)	—	—	—
4 weeks	15.64 (3.75)	1.55 (2.81)	13.62 (3.51)	1.39 (2.92)	0.04 (–0.87 to 0.95)	.90
8 weeks	14.27 (4.76)	0.18 (4.40)	11.64 (3.20)	–0.59 (3.11)	0.21 (–0.65 to 1.06)	.63
Barriers of diet adherence						
Baseline	11.91 (2.55)	—	12.50 (4.20)	—	—	—
4 weks	11.27 (3.88)	–0.64 (2.29)	12.44 (4.81)	–0.06 (3.44)	–0.19 (–1.09 to 0.72)	.65
8 weeks	11.64 (3.83)	–0.27 (2.05)	10.86 (3.37)	–1.64 (2.50)	0.59 (–0.26 to 1.44)	.13
Barriers of self-monitoring adherence						
Baseline	21.91 (3.33)	—	26.57 (6.20)	—	—	—
4 weeks	20.00 (4.92)	–1.91 (4.95)	23.42 (6.28)	–3.15 (6.00)	0.23 (–0.65 to 1.10)	.59
8 weeks	19.09 (5.59)	–2.82 (5.15)	25.07 (6.97)	–1.50 (6.01)	–0.23 (–1.07 to 0.60)	.56
Self-efficacy (SCSE [b])						
Baseline	83.41 (10.97)	—	88.39 (11.25)	—	—	—
4 weeks	89.09 (14.59)	5.68 (13.92)	87.29 (16.68)	–1.10 (12.20)	0.53 (–0.37 to 1.43)	.23
8 weeks	90.91 (10.26)	7.50 (17.21)	85.89 (16.25)	–2.50 (12.33)	0.68 (–0.17 to 1.54)	.10
HF-knowledge (DHFK [c])						
Baseline	11.64 (1.50)	—	11.36 (2.44)	—	—	—
4 weeks	11.54 (1.34)	–0.10 (2.05)	12.12 (1.67)	0.77 (2.07)	–0.42 (–1.27 to 0.43)	.30
8 weeks	12.09 (1.22)	0.45 (1.97)	11.36 (1.98)	0.00 (2.25)	0.21 (–0.62 to 1.05)	.59

[a]Not applicable.

[b]SCSE: Self-Care Self-Efficacy scale.

[c]DHFK: Dutch Heart Failure Knowledge Scale.

Effects on Health Status Scores

With respect to health status (Table 6), patients in the intervention group had moderate effect size improvements that were clinically meaningful (>5 points difference), both at 4 weeks (Cohen *d*=0.54) and 8 weeks (Cohen *d*=0.46). Improvements in the intervention arm were mainly driven by positive changes in the physical limitation, symptom frequency, and social limitation domains.

XSL•FO
RenderX

Table 6. Change in Kansas City Cardiomyopathy Questionnaire 12-item (KCCQ-12) subscale scores over time by group.

Health status (KCCQ-12)	Intervention group, mean (SD)	Change in intervention group scores from baseline, mean (SD)	Control group, mean (SD)	Change in control group scores from baseline, mean (SD)	Between-group difference, Cohen *d* (95% CI)	*P* value
Physical limitation						
Baseline	62.50 (36.18)	—[a]	71.43 (26.90)	—	—	—
4 weeks	82.58 (21.56)	20.08 (27.44)	69.79 (29.31)	–1.64 (36.51)	0.66 (–0.21 to 1.54)	.10
8 weeks	87.88 (21.53)	25.38 (22.86)	83.93 (20.73)	12.50 (28.50)	0.49 (–0.35 to 1.34)	.21
Symptom frequency						
Baseline	68.56 (33.14)	—	77.08 (29.19)	—	—	—
4 weeks	74.81 (28.87)	6.25 (36.38)	82.70 (24.35)	5.62 (28.90)	0.02 (–0.84 to 0.88)	.96
8 weeks	80.49 (20.23)	11.93 (29.18)	80.21 (28.16)	3.13 (26.86)	0.32 (–0.52 to 1.15)	.44
Quality of life						
Baseline	57.95 (33.67)	—	57.14 (26.27)	—	—	—
4 weeks	65.91 (30.66)	7.95 (31.76)	60.67 (25.21)	3.53 (20.99)	0.18 (–0.69 to 1.04)	.70
8 weeks	69.32 (28.70)	11.36 (24.66)	68.75 (23.39)	11.61 (20.49)	–0.01 (–0.84 to 0.82)	.98
Social limitation						
Baseline	66.86 (42.89)	—	82.74 (17.44)	—	—	—
4 weeks	81.82 (27.08)	14.96 (31.34)	77.29 (22.21)	–5.45 (25.71)	0.72 (–0.17 to 1.62)	.10
8 weeks	79.92 (26.80)	13.07 (31.07)	77.38 (30.21)	–5.36 (32.46)	0.58 (–0.28 to 1.44)	.16
Overall score						
Baseline	63.67 (33.34)	—	72.10 (19.93)	—	—	—
4 weeks	76.28 (23.35)	12.61 (23.82)	72.22 (20.79)	0.12 (22.80)	0.54 (–0.33 to 1.41)	.19
8 weeks	79.40 (21.80)	15.74 (22.93)	77.57 (23.04)	5.47 (21.81)	0.46 (–0.38 to 1.30)	.26

[a]Not applicable.

Self-Monitoring Adherence

Daily weighing and BP monitoring adherence over 8 weeks was 86.4% (SD 11.16%) and 86.22% (SD 10.35%), respectively, among patients allocated to the iCardia4HF intervention. Daily self-monitoring of weight and BP decreased by –0.2% (t_{78}=–0.32; *P*=.75) and –0.9% (t_{78}=–1.25; *P*=.21) each week during the intervention period, but the decline was not statistically significant (Table 7). From baseline to 4 weeks, mean daily weighing adherence was 85.5% (SD 20.8%), and mean daily BP monitoring adherence was 86.6% (SD 20.3%). From week 4 to week 8, mean daily weighing adherence was 87.2% (SD 10.8%), and daily BP monitoring adherence was 85.8% (SD 11.9%).

Table 7. Linear mixed-effects analyses of daily weighing and blood pressure (BP) monitoring adherence over time (intervention group).

Outcome	Variable	Mean (SE)	*t* test (*df*)	*P* value
BP	Intercept	92.08 (4.90)	18.76 (78)	<.001
BP	Slope	–0.95 (0.77)	–1.25 (78)	.21
Weight	Intercept	88.74 (5.19)	17.10 (78)	<.001
Weight	Slope	–0.26 (0.83)	–0.32 (78)	.75

Effects on Physical Activity

As shown in Table 8, the Fitbit activity tracker and Fitbit mobile app used for promotion and self-monitoring of physical activity did not have a clinically meaningful effect on the intervention participants' trajectory of daily steps, MVPA, or sedentary minutes over the 8-week intervention period. Overall, participants averaged 5480 (SD 3904) steps per day and 22.20 (SD 18.71) MPVA minutes per day during the 8-week intervention period. From baseline to 4 weeks, participants averaged 5341.49 (SD 3923.18) steps per day and 19.78 (SD 21.88) MVPA minutes per day. From 4 to 8 weeks, intervention participants averaged 5619.20 (SD 4015.03) steps per day and 24.62 (SD 20.45) MVPA minutes per day. On average, intervention participants wore their Fitbit for 1031 (SD 421.2) minutes per day during the 8-week intervention period. Mean

XSL•FO
RenderX

adherence to wearing the Fitbit for ≥600 minutes per day (as recommended at study orientation) was 85.3% (SD 23.12%). Wear-time slightly declined over the 8-week period by an average of 36 minutes per day (t_{76}=–4.46; *P*<.001).

Table 8. Linear mixed-effects analyses of physical activity and Fitbit wear-time over time (intervention group).

Outcome	Variable	Values, mean (SE)	*t* test (*df*)	*P* value
Daily steps	Intercept	5085.50 (1281.37)	3.97 (65)	<.001
Daily steps	Slope	25.70 (72.66)	0.35 (65)	.72
MVPA[a] min/d	Intercept	17.35 (6.93)	2.50 (65)	.01
MVPA min/d	Slope	0.687 (0.76)	0.90 (65)	.37
Sedentary min/d	Intercept	637.01 (38.88)	16.38 (65)	<.001
Sedentary min/d	Slope	1.98 (3.28)	0.60 (65)	.54
Fitbit wear-time min/d	Intercept	1191.90	11.82 (76)	<.001
Fitbit wear-time min/d	Slope	–36.12	–4.46 (76)	<.001

[a]MVPA: moderate-to-vigorous physical activity.

Acceptability of the Intervention

As shown in Table 9, most patients perceived the intervention to be useful in supporting HF self-care (mean 2.94 on a 3-point Likert scale, SD 0.07). During exit interviews, patients said that they found the intervention very engaging and that the mobile apps and devices empowered them to perform daily self-monitoring of weight, symptoms, and physical activity. Also, participants said that the text messages provided positive reinforcement and helped them become more informed about HF and self-care. Importantly, most patients reported that they felt more confident in managing their condition, and the data in the apps allowed them to have more informed discussions with their care providers during follow-up appointments. Most patients found the apps and devices easy to use (mean 2.97, SD 0.08) and positively commented on the ability of devices to automatically transfer data to the mobile apps on their phone via Bluetooth, thus eliminating any need for manual data entry. Overall, patients claimed to be very satisfied with the intervention components (mean 2.93, SD 0.14) and had a firm intention of continuing to use the apps and devices of the iCardia4HF intervention after the end of the study (mean 3, SD 0).

XSL•FO
RenderX

Table 9. Acceptance of the iCardia4HF intervention.

Subscales and measures	Scores, mean (SD)
Perceived usefulness (overall score)	2.94 (0.07)
I have maintained or improved my health by using the mobile app and smart devices of the iCardia4HF intervention	2.90 (.32)
I am more informed about my heart failure by using the mobile app and smart devices	3 (0)
My knowledge about self-managing my health has improved	3 (0)
I feel more confident taking care of my health	3 (0)
I am more autonomous in monitoring my vital signs and heart failure symptoms	3 (0)
I feel less anxious about my health	2.80 (0.42)
I have more informed discussions with my doctor about my heart failure based on the data I collect from the mobile app and smart devices	2.90 (0.32)
Ease of use (overall score)	2.97 (0.08)
I found it easy to use the mobile app and wearable device(s) for self-monitoring my heart failure	3 (0)
I found the mobile app and wearable/smart device(s) user-friendly	3 (0)
Learning how to use the mobile app and wearable/smart device(s) to self-manage my heart failure was easy for me	2.90 (0.32)
The information provided/stored in the mobile app(s) was easy to understand and interpret	3 (0)
Interacting with the mobile app and wearable/smart devices was clear and understandable	3 (0)
User satisfaction (overall score)	2.93 (0.14)
I am satisfied with the use of the mobile app and the wearable/smart devices for self-monitoring	3 (0)
I am pleased with the use of the wearable/smart devices	3 (0)
I am delighted with the use my wearable/smart devices	2.80 (0.42)
Confirmation of initial expectations (overall score)	2.90 (0.23)
My initial expectations concerning the use of the heart failure mobile app and wearable/smart devices have been confirmed so far	2.70 (0.68)
Using the heart failure mobile app and wearable/smart devices turned out to be easier that I first thought	3 (0)
There are more benefits to using the heart failure mobile app and wearable/smart devices than I first thought	3 (0)
Intention to continue using (overall score)	3 (0)
I have every intention of continuing to use the mobile app and wearable/smart devices in the future	3 (0)
I will continue to use the mobile app and wearable/smart devices to monitor different aspects of my health	3 (0)
I have no intention of stopping to use the heart failure mobile app and wearable/smart devices in the future	3 (0)

Adverse Events

During the study follow-up period of 8 weeks, there were 6 hospitalizations: 4 in the intervention group (1/11 patients, 9% experienced an event) and 2 in the control group (1/14 patients, 7% experienced an event). There were also 4 emergency room visits: 2 in the intervention group (2/11 patients, 18% experienced an event) and 2 in the control group (1/14 patients, 7% experienced an event). In terms of HF-related hospitalizations, there was 1 HF readmission in the intervention group (1/11 patients, 9% experienced an event) and none in the control group. There were no deaths during the follow-up period. There were no serious or other adverse events associated with the intervention or study procedures.

Discussion

Principal Findings and Interpretation

In this phase 1 randomized controlled trial, we examined the feasibility, acceptability, safety, and preliminary efficacy of a patient-centered mHealth intervention (iCardia4HF) on HF self-care and other health outcomes in a predominantly racial and ethnic minority population of middle-aged and older adult patients with HF. The iCardia4HF intervention integrates multiple consumer-grade mHealth apps and devices (MyApps) with a program of individually tailored text messages (Text4HF) targeting HF self-care and modifiable behavioral factors affecting HF self-care. Overall, our results demonstrated that iCardia4HF is a potentially feasible, acceptable, and safe intervention that warrants further exploration in a phase 2 clinical trial to determine its efficacy.

XSL•FO
RenderX

The results of our recruitment efforts provide important information that can be used in the planning of a phase 2 trial and other future mHealth intervention studies in midlife and older adult patients with HF. The numbers of patients eligible for the study were much lower than we originally anticipated and required a significantly larger pool of patients with HF to be screened during the recruitment period. The data show that approximately 16 patients needed to be screened for every patient enrolled (5.8% recruitment fraction). Although low, this rate is consistent with rates reported in other studies of patients with HF [90]. For example, the Better Effectiveness After Transition–Heart Failure study [91], one of the largest RCTs of remote patient telemonitoring in an HF population aged ≥50 years in the United States (n=1437 patients with HF, 664/1437, 46.2% female, and 316/1437, 22% African American), had a recruitment fraction of 4.7%. The study assessed 30,844 patients for eligibility and enrolled 1437 patients. Approximately 92.32% (28,476/30,844) of people screened did not meet the inclusion criteria. These numbers highlight the challenge of conducting RCTs in patients with serious chronic diseases, such as HF, who often have multiple comorbidities, functional disabilities, or other impairments that affect eligibility and may limit their ability to participate in research [90]. In our study, comorbidities (eg, cancer and severe or end-stage chronic kidney disease) and functional or cognitive impairments (eg, inability to walk, dementia, and Alzheimer disease) accounted for approximately 45% of exclusions. Future studies might consider modifying the eligibility criteria to enable testing of the iCardia4HF intervention in a larger and more diverse population with HF. Taken collectively, these results emphasize the need to carefully develop sound recruitment strategies before a study begins and devote sufficient financial and personnel resources to meet recruitment needs.

Despite the lower than anticipated recruitment rate, we achieved a high degree of participant retention and completion of study visits (85% at 4 weeks and 100% at 8 weeks). We also observed high rates of adherence (>80%) to wearing the Fitbit activity tracker and using the mHealth apps and devices daily for self-monitoring of vital signs and HF symptoms. Quantitative ratings from study participants allocated to the intervention (100% response rate) showed high levels of perceived usefulness, ease of use, and intention to continue using the technology for HF self-care. Patients were also satisfied with the text messages (Text4HF) and said that they provided them with positive reinforcement. These findings are consistent with previous studies and reviews on technology adoption, which report that older adults are more likely to adopt and consistently use a new technology if it is perceived as useful (ie, addresses an important need) and is easy to use [92-94]. The functional and cognitive changes that come with aging, such as poorer vision, cognitive decline, memory loss, decreased dexterity, and sensory impairments, make learning and using mHealth apps and wearable devices more challenging for older adults [94]. Hence, it is important for mHealth technologies to meet a series of end user requirements, such as user-friendly designs, limited manual data entry, easy-to-interpret visualizations, and appropriate user support [95,96]. Qualitative feedback during exit interviews illuminated that the salient features of our mHealth intervention supporting technology acceptance and satisfaction were (1) the automated capturing and syncing of the data, which eliminated the need for manual data entry into the mHealth apps; (2) the overall user-friendliness and ease of use of the intervention apps and devices; (3) the helpful content and positive reinforcement that the text messages provided; and (4) the feedback received from the intervention apps in the form of notifications and data visualizations. Another factor that most likely played a key role in the acceptance and use of the intervention technology, although not mentioned in the exit interviews, was the training and technical support participants received during the study. Participants received one-on-one instruction at baseline on how to use each mHealth app and device, including clear and simplified end user guides that our team developed for this study. They were also provided with a number they could call during office hours to receive technical support from study staff for individual problems with the devices. Previous research has consistently shown that with sufficient training and support, older adults are significantly more likely to learn and use new technologies, even if they have limited prior experience [97-100]; this is because they may need more time and tailored instruction to overcome potential barriers to adoption. Other factors that may have played a positive role in the acceptance and use of the intervention technology relate to the characteristics of our study sample. Specifically, the relatively younger mean age (56 years) of our HF population, which is reflective of the earlier age of HF onset in Black or African American people [62], and patient selection criteria of our study (eg, smartphone ownership, ability to independently perform self-care, and no major cognitive decline). Future studies and pragmatic trials are needed to determine whether older (aged ≥65 years) and more diverse populations with HF will be as willing to adopt and consistently use our intervention technologies over a longer period.

With respect to the preliminary efficacy of the intervention, iCardia4HF trended to improve HF self-care as indicated by the change in SCHFI scores from baseline. Even though the CIs were wide and included the null, there was a consistently positive effect on all 3 HF self-care domains (maintenance, symptom perception, and self-care management) in the intervention group. The between-group Cohen *d* effect sizes of 0.19, 0.33, and 0.25 for the 3 HF self-care subscales of *maintenance*, *symptom perception*, and *self-care management* at 8 weeks, respectively, qualify as a “probably positive” finding, according to the various scenarios prespecified in the study protocol about the possible intervention effect [68]. The SCHFI subscale scores observed in our study were higher at baseline compared with those reported in other studies in the literature [101-104], which means that, on average, many study participants were already performing adequate HF self-care when enrolled in our study based on self-report. This may have made it more difficult to demonstrate larger changes in HF self-care between the intervention and control groups. To address this issue, in the phase 2 trial of iCardia4HF [68], we are using SCHFI in the screening process to identify patients who are performing insufficient self-care and are more likely to benefit from the iCardia4HF intervention [105]. This approach has been used in other trials [106]. Given the small sample of our study, we found no significant correlations between HF self-care and patient demographic or clinical characteristics (eg,

XSL•FO
RenderX

age, sex, living alone, education, employment, LVEF, and NYHA). Future research should further explore the role of these factors as potential confounders or mediators of HF self-care [107].

Compared with usual care, the intervention trended toward improvement in self-reported health status (physical limitation, symptom frequency, and social limitation) as assessed with the KCCQ-12. The greatest improvements in terms of Cohen *d* values were observed in the intervention target variables of self-efficacy (0.68) and health beliefs about HF self-care, especially perceived benefits of medication adherence (0.63) and self-monitoring adherence (0.94). A possible explanation for this finding is that these measures, in addition to the mHealth apps and devices (MyApps), were directly targeted by the tailored text messaging component (Text4HF), which may have had an additive effect. However, this finding requires further exploration. In a new clinical trial that we are conducting with funding from the National Institute of Heart and Lung Blood Institute (1 R01 HL168376-01), we are using a 2×2 factorial randomized trial design to determine the independent and synergistic effects of the 2 iCardia4HF intervention components (MyApps and Text4HF).

Physical activity (steps and MVPA) measured with the Fitbit activity tracker in the intervention group only did not improve over time. A possible explanation for this finding, besides the short duration of the study, is the lack of specific physical activity goals pertaining to the 2 measures of interest: steps and MVPA. A recent systematic review and meta-analysis of Fitbit-based physical activity interventions showed that goal setting is more effective than Fitbit alone [51]. iCardia4HF could potentially be strengthened in this area by incorporating weekly physical activity goal setting for either steps or MVPA to gradually increase participants' physical activity as recommended by existing guidelines [1,108].

Comparison With Other Studies

In a recent systematic review and meta-analysis that we conducted and published [29], we identified 16 RCTs (n=4389 patients) of mHealth interventions for patients with HF. Most trials (12/16, 75%) evaluated the efficacy of RPM interventions (also known as telemonitoring) that included daily transmission of patient-generated health data to a clinical care team for review and delivery of actionable feedback based on the incoming data [29]. Fewer trials (n=4) tested the efficacy of stand-alone, patient-centered interventions (without RPM) targeting HF self-care through the use of mobile technologies. Of these 4 trials, 3 (75%) tested the efficacy of a single self-care support mobile app [102,109,110], while one trial [111] tested the efficacy of HF self-care text messages compared to usual care. Another RCT published after our systematic review tested a tablet-based mobile app designed to support HF patients' self-care at home [112]. Sample size in the 5 trials of mHealth interventions without RPM ranged from 18 patients to 767 patients, and follow-up duration ranged from 30 days to 6 months.

Daily adherence to using the app was reported to be high (>80%) among intervention participants in 2 studies [109,112], while in another study [102], patient engagement with the app and study completion were affected by challenges faced with using a chest-strap Bluetooth device (BioHarness-3) for vital signs and exercise monitoring. Only 72% (13/18) of participants completed the 30-day follow-up, and 43% (4/9) accessed the app daily.

HF self-care in the 5 RCTs was assessed using different instruments. A total of 2 studies used an older SCHFI version (version 6.2) [102,110], another 2 studies used different versions of the European Heart Failure Self-care Behavior Scale (EHFScB-12 and EHFScB-9) [109,112], and 1 study used generic items about self-care (not previously validated). Consequently, it is difficult to directly compare the results of these studies with our effect sizes and 95% CI. The 2 trials that used EHFScB found statistically significant improvements in the overall self-care score with the intervention, while results in the other 2 trials that used SCFHI were inconsistent. Athilingam et al [102] found significant improvements in the self-care management and self-care confidence subscales, while Dorsch et al [110] found no differences between the intervention and control groups in the SCHFI subscales.

Contrary to our intervention, the mobile apps used in these studies were not commercially available. They were developed internally by the research team and with input from health care providers (eg, cardiologists, HF nurses, and dieticians). In terms of HF self-care supported features, all 4 apps prompted active daily self-monitoring of weight and HF symptoms, included an education module on HF, and generated automated alerts of possible decompensated HF using an algorithm based on predetermined risk criteria or scores. The latter was a feature that our intervention did not have. Two apps [102,112] included a medication tracker feature and supported physical activity and exercise monitoring, and 1 app supported daily salt intake and dietary records [112]. Only 1 app supported wireless transfer of data from external devices (eg, weight scale) to the app. The other 3 apps required users to manually enter their vital signs recorded from external devices into the app (eg, weight, heart rate, BP, and steps). This is primarily due to the difficulty of integrating third-party devices with a custom-made app.

One of the 5 RCTs was a large trial (n=767) conducted in China [111] that investigated the efficacy of an SMS text messaging intervention on the composite end point of death- or readmission-free survival and the secondary outcome of self-care behavior compared with usual care over 180 days. Participants were recently discharged patients with ADHF (mean age 61, SD 15 years; 334/767, 43.5% female and 525/767, 68.4% NYHA class 3). Patients in the SMS group, as well as their caregivers, received educational and reminder text messages from a platform operated by research nurses. The educational text messages were condensed SMS about HF knowledge (eg, symptoms of HF decompensation), while the reminder text messages focused on self-care adherence (eg, taking medicine or weighing). Results showed that event-free survival was better in the SMS group when compared with the control group (odds ratio 0.819, 95% CI 0.677-0.991). In terms of self-care, patients in the SMS group reported better medication compliance (*P*=.03) and water restriction (*P*=.046) than the control group, but there were no significant differences in daily weighing adherence, restricted sodium diet, and physical activity. The latter finding

XSL•FO
RenderX

might be attributed to how self-care was assessed and to what extent the text messaging program targeted these self-care behaviors (daily weighing, low-sodium diet, and physical activity). Contrary to our study, self-care adherence was not assessed with digital devices or a validated instrument, and messages in this trial were not tailored to study participants. Text messages had a limited scope compared with our Text4HF program, which, in addition to HF knowledge and self-care adherence, focused on changing health beliefs (perceived benefits and barriers of HF self-care adherence) and increasing self-efficacy.

Previous research has shown that health communication interventions that succeed in targeting specific content areas and making information relevant to their intended audience are more effective than those that do not. In a pilot single-arm prospective study (n=15) of a text messaging intervention in a largely African American population with ADHF, Nundy et al [103] found that specific SCHFI items that improved over time were generally those directly targeted by the text messages [103]. Other self-care items, such as physical activity and exercise that were not targeted by the intervention did not improve. This is consistent with the findings of our study and highlights the importance of message tailoring and careful selection of intervention, target variables and behaviors. Despite differences in the content and structure of the messaging intervention compared with our study, the positive results observed in the Chen et al [111] trial highlight the potential contribution of text messaging as a stand-alone or adjunct intervention toward improving HF self-care adherence and event-free survival in patients with HF, especially when postdischarge follow-up may not be readily available due to geographical barriers and care provider shortages.

Study Strengths

This study has several strengths. First, we used a rigorous RCT design with allocation concealment and blinded outcome assessments at all time points to minimize selection bias and detection bias [113]. Second, we used an intention-to-treat analysis to minimize attrition bias [113]. Third, our intervention involved the use of mHealth apps and devices (eg, Fitbit and Withings) that are internationally known and widely available in different countries. This choice increases the potential for scalability and broader impact if the intervention is found to be effective in future studies. Fourth, we had a high representation of minority and female patients. African Americans and women are disproportionally studied in HF clinical trials compared with HF prevalence [5,114]. Fifth, we had a high rate of patient retention and compliance associated with the intervention. This allowed for optimal testing and delivery of our intervention. Sixth, the use of a previously validated digital health platform (iCardia) ensured reliable and consistent data collection from the mHealth apps and devices, including delivery of the text messaging intervention (intervention fidelity).

Study Limitations

Though several promising results were produced, this study is not without limitations. Two important limitations are the small sample size and short follow-up duration. This study was not powered to detect statistically significant differences between the 2 groups, and therefore, effect sizes should be interpreted with caution, taking into consideration the uncertainty of the estimates as reflected by the 95% CI. A larger and longer-duration study is needed, including a maintenance period to assess the sustainability effects of the intervention. Another limitation was that our primary efficacy outcome of HF self-care was a self-reported measure, which is known to be influenced by detection bias when patients are unblinded to treatment allocation. Although we used objective measures and data from the apps and devices to assess self-monitoring adherence, these behaviors do not cover the entire spectrum of HF self-care activities that patients need to perform to maintain physiological stability. In the phase 2 trial of the iCardia4HF intervention [68], we incorporate additional methods and tools to objectively assess other important self-care behaviors, such as medication adherence (eg, pill monitoring bottles), low-sodium diet (eg, urine sodium), and completion of follow-up appointments (eg, EMR). Our intervention was not designed for people who have end-stage renal disease (exclusion criterion in this study), and therefore, our results may not be generalizable to this population. Also, the compensation of enrolled patients for their participation in the study, along with the possibility of the Hawthorn effect (modification of behavior in response to awareness of being monitored), may have influenced the results. Finally, the health belief scales that guided the text messaging intervention were developed many years ago and have not been updated since. Therefore, many items focused on adherence to water pills, a low-sodium diet, and weight monitoring. In the future, we plan to include additional intervention target variables and expand our database of behavior change text messages to other healthy lifestyle behaviors, such as heart-healthy diet, guideline-directed medical therapy, smoking cessation, alcohol cessation, physical activity, and exercise. The methodological limitations described earlier (eg, sample size, short intervention, and follow-up duration) are reflective of a phase 1 feasibility trial.

Implications for Future Research and Practice

This study contributes importantly to the limited body of literature on stand-alone mHealth interventions for patients with HF. Furthermore, it directly responds to an American Heart Association scientific statement on the current science of consumer mHealth technologies for CVD prevention [33], which recommended that future intervention studies include commercially available mHealth apps and devices to determine their safety and efficacy on health outcomes and how to best incorporate these technologies (once proven) into a broader collaborative model of care.

To our knowledge, this is the first RCT of a patient-centered mHealth intervention (without RPM) that tested the feasibility, safety, and preliminary efficacy of multiple commercial mHealth apps and devices in patients with HF. Previous studies and efforts have mainly focused on developing a single mobile app that supports most, if not all, functions of HF self-care. This is not only difficult to achieve given the fragmented nature of the current mHealth ecosystem, but also fails to consider that people tend to use multiple “best of breed” mobile apps in short bursts to support their daily activities. iCardia4HF is a departure from previous models and studies. Rather than relying on a single

XSL•FO
RenderX

app, iCardia4HF provided patients with an integrated set of popular mHealth apps and devices (MyApps), each of which specialized in different preventative behaviors covering the main areas of HF self-care and supplemented these with a program of individually tailored text messages. The combination of smartphones and low-cost mHealth technologies that are commercially available makes the proposed intervention portable to other settings and creates exciting opportunities for scalability and broader impact. Text messaging is a proven, established technology that is both inexpensive and one of the most widely adopted mobile phone functions among adults. Unfortunately, text messaging for HF has been overlooked in favor of more advanced and "shiny" RPM technologies, including mobile apps and wearable sensor devices [115]. However, there is much to be learned from text message–delivered behavioral interventions. In our study, intervention patients indicated that text messages provided them with useful information, increased perceived support, enhanced motivation to engage in better HF self-care, and provided useful prompts for healthy behavior change. Additional trials are urgently needed to determine the efficacy of text messaging both as a stand-alone and as an adjunct intervention and to build a more robust evidence base for patients with HF.

Conclusions

The iCardia4HF intervention proved to be feasible, acceptable, and safe with no evidence of increased adverse effects on HF symptoms or acute health care use (hospital readmissions and emergency visits). These results, along with the positive effects observed in HF self-care behaviors and other health outcomes, provide good evidence that a full-scale RCT to more definitively examine the efficacy of the iCardia4HF intervention in patients with HF is warranted.

Acknowledgments

The authors are indebted to the following individuals for their instrumental contributions to the conduct of the study: Laura Nawrocki, Atreya Mishra, Antonio Nieto, Jonathan Leigh, and Ayomide Owoyemi. This work was supported by the National Institute on Aging of the National Institutes of Health (grant R24AG064191) and the Center for Clinical and Translational Science (grant UL1TR002003). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Authors' Contributions

SK had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. SK conceived and designed the study with help from BSG, SJP, and SWB. SK designed and developed the MyApps intervention component and iCardia platform. SK, BSG, and SJP developed the Text4HF intervention component based on previous studies conducted by SJP. All the authors were responsible for the acquisition, analysis, or interpretation of the data and critical revision of the manuscript for important intellectual content. SK was responsible for drafting the manuscript and also obtained funding. JS and SK were responsible for the statistical analysis. SK, BSG, MMK, SWB, and SJP were responsible for the supervision of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3310 KB - jmir_v27i1e55586_app1.pdf]

References

1. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines. J Am Coll Cardiol 2022 May 03;79(17):e263-e421. [doi: 10.1016/j.jacc.2021.12.012] [Medline: 35379503]
2. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GM, Coats AJ. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
3. Bozkurt B, Ahmad T, Alexander KM, Baker WL, Bosak K, Breathett K, Writing Committee Members. Heart failure epidemiology and outcomes statistics: a report of the heart failure society of America. J Card Fail 2023 Oct;29(10):1412-1451 [FREE Full text] [doi: 10.1016/j.cardfail.2023.07.006] [Medline: 37797885]
4. Tsao CW, Aday AW, Almarzooq ZI, Anderson CA, Arora P, Avery CL, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke

XSL•FO
RenderX

statistics-2023 update: a report from the American Heart Association. Circulation 2023 Feb 21;147(8):e93-621. [doi: 10.1161/CIR.0000000000001123] [Medline: 36695182]
5. Pressler SJ. Women with heart failure are disproportionately studied as compared with prevalence: a review of published studies from 2013. J Cardiovasc Nurs 2016;31(1):84-88. [doi: 10.1097/JCN.0000000000000212] [Medline: 25419948]
6. Halloway S, Schoeny ME, Barnes LL, Arvanitakis Z, Pressler SJ, Braun LT, et al. A study protocol for MindMoves: a lifestyle physical activity and cognitive training intervention to prevent cognitive impairment in older women with cardiovascular disease. Contemp Clin Trials 2021 Feb;101:106254 [FREE Full text] [doi: 10.1016/j.cct.2020.106254] [Medline: 33383230]
7. Jackson SL, Tong X, King RJ, Loustalot F, Hong Y, Ritchey MD. National burden of heart failure events in the United States, 2006 to 2014. Circ Heart Fail 2018 Dec;11(12):e004873 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.117.004873] [Medline: 30562099]
8. Clark KA, Reinhardt SW, Chouairi F, Miller PE, Kay B, Fuery M, et al. Trends in heart failure hospitalizations in the US from 2008 to 2018. J Card Fail 2022 Feb;28(2):171-180. [doi: 10.1016/j.cardfail.2021.08.020] [Medline: 34534665]
9. Reinhardt SW, Clark KA, Xin X, Parzynski CS, Riello III RJ, Sarocco P, et al. Thirty-day and 90-day episode of care spending following heart failure hospitalization among Medicare beneficiaries. Circ Cardiovasc Qual Outcomes 2022 Jul;15(7):e008069. [doi: 10.1161/circoutcomes.121.008069]
10. Khan MS, Sreenivasan J, Lateef N, Abougergi MS, Greene SJ, Ahmad T, et al. Trends in 30- and 90-day readmission rates for heart failure. Circ Heart Fail 2021 Apr;14(4):e008335. [doi: 10.1161/CIRCHEARTFAILURE.121.008335] [Medline: 33866827]
11. Albuquerque de Almeida F, Al MJ, Koymans R, Riistama J, Pauws S, Severens JL. Impact of hospitalisation on health-related quality of life in patients with chronic heart failure. Health Qual Life Outcomes 2020 Aug 03;18(1):262 [FREE Full text] [doi: 10.1186/s12955-020-01508-8] [Medline: 32746842]
12. Rodríguez-Artalejo F, Guallar-Castillón P, Pascual CR, Otero CM, Montes AO, García AN, et al. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. Arch Intern Med 2005 Jun 13;165(11):1274-1279 [FREE Full text] [doi: 10.1001/archinte.165.11.1274] [Medline: 15956007]
13. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, American Heart Association Advocacy Coordinating Committee, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Radiology and Intervention, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
14. Jaarsma T, Hill L, Bayes-Genis A, La Rocca HB, Castiello T, Čelutkienė J, et al. Self-care of heart failure patients: practical management recommendations from the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2021 Jan;23(1):157-174 [FREE Full text] [doi: 10.1002/ejhf.2008] [Medline: 32945600]
15. Riegel B, Dickson VV, Faulkner KM. The situation-specific theory of heart failure self-care: revised and updated. J Cardiovasc Nurs 2016;31(3):226-235. [doi: 10.1097/JCN.0000000000000244] [Medline: 25774844]
16. Calero-Molina E, Hidalgo E, Rosenfeld L, Verdú-Rotellar JM, Verdú-Soriano J, Garay A, et al. The relationship between self-care, long-term mortality, and heart failure hospitalization: insights from a real-world cohort study. Eur J Cardiovasc Nurs 2022 Mar 03;21(2):116-126. [doi: 10.1093/eurjcn/zvab011] [Medline: 34008849]
17. Lee CS, Bidwell JT, Paturzo M, Alvaro R, Cocchieri A, Jaarsma T, et al. Patterns of self-care and clinical events in a cohort of adults with heart failure: 1 year follow-up. Heart Lung 2018;47(1):40-46 [FREE Full text] [doi: 10.1016/j.hrtlng.2017.09.004] [Medline: 29054487]
18. Lee CS, Moser DK, Lennie TA, Riegel B. Event-free survival in adults with heart failure who engage in self-care management. Heart Lung 2011;40(1):12-20 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.12.003] [Medline: 20561885]
19. Ruppar TM, Cooper PS, Mehr DR, Delgado JM, Dunbar-Jacob JM. Medication adherence interventions improve heart failure mortality and readmission rates: systematic review and meta-analysis of controlled trials. J Am Heart Assoc 2016 Jun 17;5(6):e002606 [FREE Full text] [doi: 10.1161/JAHA.115.002606] [Medline: 27317347]
20. Zhao Q, Chen C, Zhang J, Ye Y, Fan X. Effects of self-management interventions on heart failure: systematic review and meta-analysis of randomized controlled trials - Reprint. Int J Nurs Stud 2021 Apr;116:103909 [FREE Full text] [doi: 10.1016/j.ijnurstu.2021.103909] [Medline: 33642066]
21. Moser DK, Watkins JF. Conceptualizing self-care in heart failure: a life course model of patient characteristics. J Cardiovasc Nurs 2008;23(3):205-209. [doi: 10.1097/01.JCN.0000305097.09710.a5] [Medline: 18437061]
22. Bui AL, Fonarow GC. Home monitoring for heart failure management. J Am Coll Cardiol 2012 Jan 10;59(2):97-104 [FREE Full text] [doi: 10.1016/j.jacc.2011.09.044] [Medline: 22222071]
23. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
24. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31;2015(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]

XSL•FO
RenderX

25. Hajduczok AG, Muallem SN, Nudy MS, DeWaters AL, Boehmer JP. Remote monitoring for heart failure using implantable devices: a systematic review, meta-analysis, and meta-regression of randomized controlled trials. Heart Fail Rev 2022 Jul 24;27(4):1281-1300 [FREE Full text] [doi: 10.1007/s10741-021-10150-5] [Medline: 34559368]
26. Alotaibi S, Hernandez-Montfort J, Ali OE, El-Chilali K, Perez BA. Remote monitoring of implantable cardiac devices in heart failure patients: a systematic review and meta-analysis of randomized controlled trials. Heart Fail Rev 2020 May 30;25(3):469-479. [doi: 10.1007/s10741-020-09923-1] [Medline: 32002732]
27. Allida S, Du H, Xu X, Prichard R, Chang S, Hickman LD, et al. mHealth education interventions in heart failure. Cochrane Database Syst Rev 2020 Jul 02;7(7):CD011845 [FREE Full text] [doi: 10.1002/14651858.CD011845.pub2] [Medline: 32613635]
28. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
29. Kitsiou S, Vatani H, Paré G, Gerber BS, Buchholz SW, Kansal MM, et al. Effectiveness of mobile health technology interventions for patients with heart failure: systematic review and meta-analysis. Can J Cardiol 2021 Aug;37(8):1248-1259. [doi: 10.1016/j.cjca.2021.02.015] [Medline: 33667616]
30. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 02;2(1):e10057 [FREE Full text] [doi: 10.2196/10057] [Medline: 31758762]
31. Cajita MI, Gleason KT, Han HR. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):E10-E22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
32. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
33. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on EpidemiologyPrevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on CardiovascularStroke Nursing, Council on Functional GenomicsTranslational Biology, Council on Quality of Care and Outcomes Research, Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
34. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
35. Lauver DR, Ward SE, Heidrich SM, Keller ML, Bowers BJ, Brennan PF, et al. Patient-centered interventions. Res Nurs Health 2002 Aug;25(4):246-255. [doi: 10.1002/nur.10044] [Medline: 12124719]
36. Leigh JW, Gerber BS, Gans CP, Kansal MM, Kitsiou S. Smartphone ownership and interest in mobile health technologies for self-care among patients with chronic heart failure: cross-sectional survey study. JMIR Cardio 2022 Jan 14;6(1):e31982 [FREE Full text] [doi: 10.2196/31982] [Medline: 35029533]
37. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
38. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
39. Kitsiou S, Thomas M, Marai E, Maglaveras N, Kondos G, Arena R. Development of an innovative mHealth platform for remote physical activity monitoring and health coaching of cardiac rehabilitation patients. In: Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics. 2017 Presented at: BHI '17; February 16-19, 2017; Orlando, FL p. 133-136 URL: https://ieeexplore.ieee.org/document/7897223 [doi: 10.1109/bhi.2017.7897223]
40. Becker MH. The health belief model and personal health behavior. Health Educ Monogr 1974;2:324-473 [FREE Full text]
41. Jasmine TJ, Wai-Chi SC, Hegney DG. The impact of knowledge and beliefs on adherence to cardiac rehabilitation programs in patients with heart failure: A systematic review. JBI Libr Syst Rev 2012 May;10(7):399-470. [doi: 10.11124/jbisrir-2012-53]
42. Lee KS, Moser DK, Dracup K. Relationship between self-care and comprehensive understanding of heart failure and its signs and symptoms. Eur J Cardiovasc Nurs 2018 Aug;17(6):496-504. [doi: 10.1177/1474515117745056] [Medline: 29192794]
43. Riegel B, Lee CS, Albert N, Lennie T, Chung M, Song EK, et al. From novice to expert: confidence and activity status determine heart failure self-care performance. Nurs Res 2011;60(2):132-138. [doi: 10.1097/NNR.0b013e31820978ec] [Medline: 21317825]
44. Massouh A, Skouri H, Cook P, Huijer HA, Khoury M, Meek P. Self-care confidence mediates self-care maintenance and management in patients with heart failure. Heart Lung 2020;49(1):30-35 [FREE Full text] [doi: 10.1016/j.hrtlng.2019.07.008] [Medline: 31371031]

XSL•FO
RenderX

45. White-Williams C, Rossi LP, Bittner VA, Driscoll A, Durant RW, Granger BB, American Heart Association Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; Council on Epidemiology and Prevention. Addressing social determinants of health in the care of patients with heart failure: a scientific statement from the American Heart Association. Circulation 2020 Jun 02;141(22):e841-e863 [FREE Full text] [doi: 10.1161/CIR.0000000000000767] [Medline: 32349541]
46. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012 Sep;Suppl(s2):49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
47. Nyenhuis SM, Balbim GM, Ma J, Marquez DX, Wilbur J, Sharp LK, et al. A walking intervention supplemented with mobile health technology in low-active urban African American women with asthma: proof-of-concept study. JMIR Form Res 2020 Mar 11;4(3):e13900 [FREE Full text] [doi: 10.2196/13900] [Medline: 32159520]
48. Prieto-Centurion V, Casaburi R, Coultas DB, Kansal MM, Kitsiou S, Luo JJ, et al. Daily physical activity in patients with COPD after hospital discharge in a minority population. Chronic Obstr Pulm Dis 2019 Oct 23;6(4):332-340 [FREE Full text] [doi: 10.15326/jcopdf.6.4.2019.0136] [Medline: 31647855]
49. Prendergast H, Petzel-Gimbar R, Kitsiou S, Del Rios M, Lara B, Jackson M, et al. Targeting of uncontrolled hypertension in the emergency department (TOUCHED): design of a randomized controlled trial. Contemp Clin Trials 2021 Mar;102:106283 [FREE Full text] [doi: 10.1016/j.cct.2021.106283] [Medline: 33484897]
50. Buchholz SW, Wilbur J, Halloway S, Schoeny M, Johnson T, Vispute S, et al. Study protocol for a sequential multiple assignment randomized trial (SMART) to improve physical activity in employed women. Contemp Clin Trials 2020 Feb;89:105921 [FREE Full text] [doi: 10.1016/j.cct.2019.105921] [Medline: 31899371]
51. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
52. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
53. Gorzelitz J, Farber C, Gangnon R, Cadmus-Bertram L. Accuracy of wearable trackers for measuring moderate- to vigorous-intensity physical activity: a systematic review and meta-analysis. J Meas Phys Behav 2020;3(4):22. [doi: 10.1123/jmpb.2019-0072]
54. Bennett SJ, Hays LM, Embree JL, Arnould M. Heart Messages: a tailored message intervention for improving heart failure outcomes. J Cardiovasc Nurs 2000 Jul;14(4):94-105. [doi: 10.1097/00005082-200007000-00007] [Medline: 10902106]
55. Bennett SJ, Litzelman DK, Wright A, Perkins SM, Wu J, Meyer L, et al. The PUMP UP tailored computerized program for heart failure care. Nurs Outlook 2006;54(1):39-45. [doi: 10.1016/j.outlook.2005.05.003] [Medline: 16487779]
56. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Oct 15;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
57. Kitsiou S, Gerber B, Buchholz S, Kansal M, Chen J, Nawrocki L, et al. Development and pilot testing of a tailored text messaging intervention to improve self-efficacy and health beliefs about self-care in people with heart failure. J Card Fail 2020 Oct;26(10):S89 [FREE Full text] [doi: 10.1016/j.cardfail.2020.09.260]
58. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
59. Masterson Creber R, Patey M, Lee CS, Kuan A, Jurgens C, Riegel B. Motivational interviewing to improve self-care for patients with chronic heart failure: MITI-HF randomized controlled trial. Patient Educ Couns 2016 Feb;99(2):256-264 [FREE Full text] [doi: 10.1016/j.pec.2015.08.031] [Medline: 26358533]
60. Sethares KA, Elliott K. The effect of a tailored message intervention on heart failure readmission rates, quality of life, and benefit and barrier beliefs in persons with heart failure. Heart Lung 2004 Jul;33(4):249-260. [doi: 10.1016/j.hrtlng.2004.03.005] [Medline: 15252415]
61. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
62. Bibbins-Domingo K, Pletcher MJ, Lin F, Vittinghoff E, Gardin JM, Arynchyn A, et al. Racial differences in incident heart failure among young adults. N Engl J Med 2009 Mar 19;360(12):1179-1190. [doi: 10.1056/nejmoa0807265]
63. Bowen D, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
64. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Nov 24;4(4):287-291 [FREE Full text] [doi: 10.1002/pst.185]
65. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975 Mar;31(1):103. [doi: 10.2307/2529712]

XSL•FO
RenderX

66. Saghaei M, Saghaei S. Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. J Biomed Sci Eng 2011;04(11):734-739. [doi: 10.4236/jbise.2011.411090]
67. Flack JM, Adekola B. Blood pressure and the new ACC/AHA hypertension guidelines. Trends Cardiovasc Med 2020 Apr;30(3):160-164 [FREE Full text] [doi: 10.1016/j.tcm.2019.05.003] [Medline: 31521481]
68. Kitsiou S, Gerber BS, Kansal MM, Buchholz SW, Chen J, Ruppar T, et al. Patient-centered mobile health technology intervention to improve self-care in patients with chronic heart failure: Protocol for a feasibility randomized controlled trial. Contemp Clin Trials 2021 Jul;106:106433 [FREE Full text] [doi: 10.1016/j.cct.2021.106433] [Medline: 33991686]
69. Nyenhuis SM, Shah N, Kim H, Marquez DX, Wilbur J, Sharp LK. The feasibility of a lifestyle physical activity intervention for black women with asthma. J Allergy Clin Immunol Pract 2021 Dec;9(12):4312-21.e2 [FREE Full text] [doi: 10.1016/j.jaip.2021.07.028] [Medline: 34333191]
70. Aguiñaga S, Marques IG, Kitsiou S, Balbim GM, Gerber BS, Buchholz SW, et al. BAILAMOS with mHealth technology!: improving physical activity and well-being in middle-aged and older Latinxs: a pre-post feasibility study. Health Educ Behav 2021 Oct 14;48(5):575-583. [doi: 10.1177/10901981211027517] [Medline: 34521228]
71. Zaslavsky O, Thompson Hilaire J, McCurry SM, Landis CA, Kitsiou S, Ward TM, et al. Use of a wearable technology and motivational interviews to improve sleep in older adults with osteoarthritis and sleep disturbance: a pilot study. Res Gerontol Nurs 2019 Jul 01;12(4):167-173 [FREE Full text] [doi: 10.3928/19404921-20190319-02] [Medline: 30901479]
72. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
73. Riegel B, Barbaranelli C, Carlson B, Sethares KA, Daus M, Moser DK, et al. Psychometric testing of the revised self-care of heart failure index. J Cardiovasc Nurs 2019;34(2):183-192 [FREE Full text] [doi: 10.1097/JCN.0000000000000543] [Medline: 30303894]
74. Yu DS, De Maria M, Barbaranelli C, Vellone E, Matarese M, Ausili D, et al. Cross-cultural applicability of the self-care self-efficacy scale in a multi-national study. J Adv Nurs 2021 Feb;77(2):681-692 [FREE Full text] [doi: 10.1111/jan.14617] [Medline: 33295675]
75. Bennett SJ, Milgrom LB, Champion V, Huster GA. Beliefs about medication and dietary compliance in people with heart failure: an instrument development study. Heart Lung 1997 Jul;26(4):273-279. [doi: 10.1016/s0147-9563(97)90084-4] [Medline: 9257137]
76. Bennett SJ, Perkins SM, Lane KA, Forthofer MA, Brater D, Murray MD. Reliability and validity of the compliance belief scales among patients with heart failure. Heart Lung 2001 May;30(3):177-185. [doi: 10.1067/mhl.2001.114193] [Medline: 11343003]
77. van der Wal MH, Jaarsma T, Moser DK, van Veldhuisen DJ. Development and testing of the Dutch Heart Failure Knowledge Scale. Eur J Cardiovasc Nurs 2005 Dec;4(4):273-277. [doi: 10.1016/j.ejcnurse.2005.07.003] [Medline: 16126459]
78. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City cardiomyopathy questionnaire. Circ Cardiovasc Qual Outcomes 2015 Sep;8(5):469-476. [doi: 10.1161/circoutcomes.115.001958]
79. Butler J, Khan MS, Mori C, Filippatos GS, Ponikowski P, Comin-Colet J, et al. Minimal clinically important difference in quality of life scores for patients with heart failure and reduced ejection fraction. Eur J Heart Fail 2020 Jun;22(6):999-1005 [FREE Full text] [doi: 10.1002/ejhf.1810] [Medline: 32239794]
80. Balbim GM, Marques IG, Marquez DX, Patel D, Sharp LK, Kitsiou S, et al. Using Fitbit as an mHealth intervention tool to promote physical activity: potential challenges and solutions. JMIR Mhealth Uhealth 2021 Mar 01;9(3):e25289 [FREE Full text] [doi: 10.2196/25289] [Medline: 33646135]
81. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
82. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351. [doi: 10.2307/3250921]
83. Hong S, Thong JY, Tam KY. Understanding continued information technology usage behavior: a comparison of three models in the context of mobile internet. Decis Support Syst 2006 Dec;42(3):1819-1834 [FREE Full text] [doi: 10.1016/j.dss.2006.03.009]
84. Van Buuren S, Brand JP, Groothuis-Oudshoorn CG, Rubin DB. Fully conditional specification in multivariate imputation. J Stat Comput Simul 2006 Dec;76(12):1049-1064. [doi: 10.1080/10629360600810434]
85. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
86. R: a language and environment for statistical computing. The R Foundation. URL: https://www.r-project.org/ [accessed 2024-04-29]
87. Harrington D, D'Agostino RB, Gatsonis C, Hogan JW, Hunter DJ, Normand ST, et al. New guidelines for statistical reporting in the journal. N Engl J Med 2019 Jul 18;381(3):285-286. [doi: 10.1056/NEJMe1906559] [Medline: 31314974]
88. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982 Dec;38(4):963. [doi: 10.2307/2529876]
89. Pinheiro J. nlme: linear and nonlinear mixed effects models. R package version 3.1-96. Cran R. URL: https://cran.r-project.org/web/packages/nlme/index.html [accessed 2024-04-29]

XSL•FO
RenderX

90. Gradus-Pizlo I, Stoudemire K, Meyer LE, Shaw RM, Subramanian U, Pressler SJ. Research in patients with heart failure: challenges in recruitment. Am J Crit Care 2008;17(3):198-203. [doi: 10.4037/ajcc2008.17.3.198]
91. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
92. Chen K, Chan AH. A review of technology acceptance by older adults. Gerontechnology 2011;10(1):e0182736. [doi: 10.4017/GT.2011.10.01.006.00]
93. Zhang M. Older people's attitudes towards emerging technologies: a systematic literature review. Public Underst Sci 2023 Nov 19;32(8):948-968 [FREE Full text] [doi: 10.1177/09636625231171677] [Medline: 37204075]
94. Cajita MI, Hodgson NA, Budhathoki C, Han HR. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
95. Masterson Creber RM, Hickey KT, Maurer MS. Gerontechnologies for older patients with heart failure: what is the role of smartphones, tablets, and remote monitoring devices in improving symptom monitoring and self-care management? Curr Cardiovasc Risk Rep 2016 Oct 15;10(10):30 [FREE Full text] [doi: 10.1007/s12170-016-0511-8] [Medline: 28713481]
96. Krishnaswami A, Beavers C, Dorsch Michael P, Dodson John A, Masterson Creber R, Kitsiou S, Innovations, Cardiovascular Teamthe Geriatric Cardiology Councils, American College of Cardiology. Gerotechnology for older adults with cardiovascular diseases: JACC state-of-the-art review. J Am Coll Cardiol 2020 Dec 01;76(22):2650-2670 [FREE Full text] [doi: 10.1016/j.jacc.2020.09.606] [Medline: 33243384]
97. Janevic MR, Shute V, Murphy SL, Piette JD. Acceptability and effects of commercially available activity trackers for chronic pain management among older African American adults. Pain Med 2020 Feb 01;21(2):e68-e78 [FREE Full text] [doi: 10.1093/pm/pnz215] [Medline: 31509196]
98. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and Barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
99. Ferguson C, Hickman LD, Turkmani S, Breen P, Gargiulo G, Inglis SC. "Wearables only work on patients that wear them": barriers and facilitators to the adoption of wearable cardiac monitoring technologies. Cardiovasc Digit Health J 2021 Apr;2(2):137-147 [FREE Full text] [doi: 10.1016/j.cvdhj.2021.02.001] [Medline: 35265900]
100. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth 2016 Apr 13;4(2):e35 [FREE Full text] [doi: 10.2196/mhealth.5120] [Medline: 27076486]
101. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
102. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]
103. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013 Mar 11;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
104. Vellone E, De Maria M, Iovino P, Barbaranelli C, Zeffiro V, Pucciarelli G, et al. The Self-Care of Heart Failure Index version 7.2: further psychometric testing. Res Nurs Health 2020 Dec;43(6):640-650 [FREE Full text] [doi: 10.1002/nur.22083] [Medline: 33141495]
105. Kitsiou S, Pressler SJ, Kansal MK, Buchholz SW, Chen J, Leigh J, et al. Feasibility and efficacy of using commercially available mHealth technologies and tailored text messages to improve self-care in patients with heart failure: phase I clinical trial of the iCardia4HF intervention. In: Proceedings of the 2020 Annual Conference on Symposium of the American Medical Informatics Association. 2020 Oct Presented at: AMIA '20; October 30-November 3, 2020; San Diego, CA p. S89. [doi: 10.1016/j.cardfail.2020.09.260]
106. Vellone E, Rebora P, Ausili D, Zeffiro V, Pucciarelli G, Caggianelli G, et al. Motivational interviewing to improve self-care in heart failure patients (MOTIVATE-HF): a randomized controlled trial. ESC Heart Fail 2020 Jun 28;7(3):1309-1318 [FREE Full text] [doi: 10.1002/ehf2.12733] [Medline: 32343483]
107. Sedlar N, Lainscak M, Mårtensson J, Strömberg A, Jaarsma T, Farkas J. Factors related to self-care behaviours in heart failure: a systematic review of European Heart Failure self-care behaviour scale studies. Eur J Cardiovasc Nurs 2017 Apr;16(4):272-282 [FREE Full text] [doi: 10.1177/1474515117691644] [Medline: 28168895]
108. Pelliccia A, Sharma S, Gati S, Bäck M, Börjesson M, Caselli S, ESC Scientific Document Group. 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease. Eur Heart J 2021 Jan 01;42(1):17-96 [FREE Full text] [doi: 10.1093/eurheartj/ehaa605] [Medline: 32860412]

XSL•FO
RenderX

109. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a tablet computer on self-care, quality of life, and knowledge: a randomized clinical trial. J Cardiovasc Nurs 2018;33(4):336-343. [doi: 10.1097/JCN.0000000000000462] [Medline: 29369123]
110. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 07;9(12):e26185 [FREE Full text] [doi: 10.2196/26185] [Medline: 34878990]
111. Chen C, Li X, Sun L, Cao S, Kang Y, Hong L, et al. Post-discharge short message service improves short-term clinical outcome and self-care behaviour in chronic heart failure. ESC Heart Fail 2019 Feb;6(1):164-173 [FREE Full text] [doi: 10.1002/ehf2.12380] [Medline: 30478888]
112. Yokota T, Fukushima A, Tsuchihashi-Makaya M, Abe T, Takada S, Furihata T, et al. The AppCare-HF randomized clinical trial: a feasibility study of a novel self-care support mobile app for individuals with chronic heart failure. Eur Heart J Digit Health 2023 Aug;4(4):325-336 [FREE Full text] [doi: 10.1093/ehjdh/ztad032] [Medline: 37538146]
113. Higgins JP, Altman DG. Assessing risk of bias in included studies. In: Higgins JP, Thomas J, editors. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: Wiley-Blackwell; 2008:187-241.
114. DeFilippis EM, Echols M, Adamson PB, Batchelor WB, Cooper LB, Cooper LS, et al. Improving enrollment of underrepresented racial and ethnic populations in heart failure trials: a call to action from the heart failure collaboratory. JAMA Cardiol 2022 May 01;7(5):540-548 [FREE Full text] [doi: 10.1001/jamacardio.2022.0161] [Medline: 35319725]
115. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of "bling": is there a future for text messaging in health care? J Med Internet Res 2019 Dec 20;21(12):e16630 [FREE Full text] [doi: 10.2196/16630] [Medline: 31859678]

Abbreviations

ADHF: acute decompensated heart failure
BP: blood pressure
EF: ejection fraction
EHFScB: European Heart Failure Self-care Behavior Scale
HF: heart failure
HFmEF: heart failure with mild ejection fraction
HFpEF: heart failure with preserved ejection fraction
HFrEF: heart failure with reduced ejection fraction
HIPAA: Health Insurance Portability and Accountability Act
KCCQ-12: Kansas City Cardiomyopathy Questionnaire 12-item
LVEF: left ventricular ejection fraction
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
NYHA: New York Heart Association
RPM: remote patient monitoring
SCHFI: Self-Care Heart Failure Index
UI Health: University of Illinois Hospital and Health Sciences System

Edited by E Wethington, J Kaye, I Kronish, A Dominello; submitted 17.12.23; peer-reviewed by P Bas-Sarmiento, AKC Wong, X Li; comments to author 17.03.24; revised version received 12.05.24; accepted 11.11.24; published 15.01.25.

Please cite as:
Kitsiou S, Gerber BS, Buchholz SW, Kansal MM, Sun J, Pressler SJ
Patient-Centered mHealth Intervention to Improve Self-Care in Patients With Chronic Heart Failure: Phase 1 Randomized Controlled Trial
J Med Internet Res 2025;27:e55586
URL: https://www.jmir.org/2025/1/e55586
doi:10.2196/55586
PMID:

©Spyros Kitsiou, Ben S Gerber, Susan W Buchholz, Mayank M Kansal, Jiehuan Sun, Susan J Pressler. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet

XSL•FO
RenderX

Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relative Preference for In-Person, Telehealth, Digital, and Pharmacologic Mental Health Care After the COVID-19 Pandemic: Cross-Sectional Questionnaire Study

E Marie Parsons[1], PhD; Zoë G Figueroa[1], BA; Michele Hiserodt[1], MA; Talea Cornelius[2], PhD, MSW, MS; Michael W Otto[1], PhD

[1]Department of Psychological and Brain Sciences, Boston University, Boston, MA, United States

[2]Center for Behavioral Cardiovascular Health, Columbia University Irving Medical Center, New York, NY, United States

Corresponding Author:
E Marie Parsons, PhD
Department of Psychological and Brain Sciences
Boston University
900 Commonwealth Avenue
Boston, MA, 02115
United States
Phone: 1 617 353 9610
Email: mariepar@bu.edu

Abstract

Background: Most adults and children in the United States fail to receive timely care for mental health symptoms, with even worse rates of care access for individuals who belong to racial and ethnic minority groups. Digital (ie, app-based) care has proven to be an efficacious and empirically supported treatment option with the potential to address low rates of care and reduce care disparities, yet little is known about the relative preference for such treatment. Furthermore, the rapid adoption of telehealth care during the COVID-19 pandemic may have shifted care preferences.

Objective: This study aimed to examine relative treatment preferences for 4 different types of mental health care: in-person psychological care, telehealth psychological care, digital treatment, or pharmacologic care. Care preferences were also examined relative to potential predictors of care use (ie, gender, race, age, stigma, discrimination, and level of shame).

Methods: In this cross-sectional online survey study of adults (N=237, mean age 35 years, range 19-68 years), we ranked 4 mental health care modalities based on care preference: (1) in-person care, (2) telehealth care, (3) digital care, and (4) pharmacologic care. Preference for treatment modality was assessed based on vignette presentation for generalized anxiety disorder and insomnia. In addition, participants completed self-report questionnaires for demographics, symptom severity, and psychological and stigma-related variables.

Results: We found no difference in overall preference for in-person versus both telehealth and digital care. For both generalized anxiety disorder and insomnia, participants preferred in-person care to telehealth care, although this finding was attenuated amongst older participants for insomnia treatment. Participants' depressed mood was associated with a greater relative preference for pharmacologic care. There was no evidence of differential preference for digital care according to demographics, symptom severity, or psychological and stigma-related variables.

Conclusions: These results indicate that digital care now competes well in terms of treatment preference with in-person, telehealth, and pharmacologic treatment options.

(J Med Internet Res 2025;27:e54608) doi:10.2196/54608

KEYWORDS

stigma; digital CBT; age; generalized anxiety disorder; insomnia; adult; telehealth; digital health

XSL•FO
RenderX

Introduction

The majority of adults and children in need of mental health services fail to receive care, with an intensification of this failure for individuals from racial and ethnic minority groups [1-4]. As a strategy to meet care needs, empirically supported digital treatment (treatments presented through a computer program or an app) has been recommended as a strategy to reduce stigma and cost and as an early modality in a stepped care model [5-8]. Meta-analyses of outcomes for digital care, and digital cognitive behavior therapy (dCBT), in particular, support this strategy as both an efficacious and cost-effective approach for addressing care needs [9-11]. Yet, little research has been done to evaluate whether dCBT addresses the attitudinal and structural barriers that prevent the initiation of mental health care. For example, data from respondents with *DSM-IV* disorders identified as part of the National Comorbidity Survey Replication revealed that the desire to manage symptoms on one's own, a facet of self-stigma, was the most common (72.6%) reason for not seeking treatment, and was followed by structural barriers such as cost, transportation, and care availability identified by 22.2% of respondents [12]. As such, there is a potential for fully autonomous dCBT to provide services for these individuals who are reluctant to seek in-person care [13] and those who face barriers to in-person care. In addition, because individuals from ethnic and racial minority groups have been found to perceive greater barriers and stigma for mental health treatment [2,14-16], there is a potential for digital treatments to help address these disparities in care by providing an alternative to specialty care clinics [17].

There is support for the hypothesis that digital treatment can indeed address some of the perceived barriers to treatment [18]. A previous study examined preference for digital or in-person treatment using 2 samples: undergraduate students and patients recruited from primary health care clinics. In both samples, there was a greater overall preference for in-person than digital treatment, but there was a significantly greater preference for digital treatment for those with higher levels of help-seeking self-stigma [18]. Greater preferences for digital treatment options have also been found for select groups, such as first-year college students [19]. Indeed, younger age and female sex, as well as treatment history and symptom levels, have been linked to a preference for digital treatment versus in-person treatment [20-22]. Yet, the care landscape appears to be changing rapidly. The COVID-19 pandemic hastened the acceptance of remote care options by both providers and consumers, with the rapid adoption of telehealth care across medical clinics [23,24], including mental health specialty clinics [25,26].

The purpose of this study was to examine relative treatment preferences for 4 different types of mental health care: in-person psychological care, telehealth psychological care (ie, provider and patient meeting online through a video conferencing platform), digital care (specifically, app-based treatment), or pharmacologic care. Care preferences were also examined relative to potential predictors of care use: gender, race, age, (self and public) stigma, discrimination, and level of shame. We conducted a large-scale online survey (through Amazon's Mechanical Turk [MTurk]) to assess preference for care based on the presentation of patient vignettes for generalized anxiety disorder (GAD) and insomnia, 2 common disorders with documented associations between self-stigma and care-seeking, for which each modality of care studied has documented efficacy [2,27-31]. We also examined whether treatment history [21] or degree of affective and insomnia symptoms influenced the care preference ratings for the vignettes.

Methods

Participants

In order to be eligible for enrollment, participants had to be 18 years of age or older and had to be able to understand English. To be included in the data analysis, participants also had to meet data quality standards appropriate for fully remote studies [32,33]. A total of 491 unique participants consented to the study, yet only 268 provided usable data, and 237 provided complete data on all variables and were included in the analysis (Figure 1 shows a diagram depicting the flow of data from consent to data analysis).

Participant ages ranged from 19 to 68 years, with a mean age of 35.41 (SD 10.53) years. In terms of the range of scores, the lowest quartile of ages reported ranged from 21 to 28 years, and the highest quartile ranged from 39 to 68 years. The majority of participants identified as male (141/237, 59.5%), 40.5% (96/237) as female, and none identified with other gender options. For race and ethnicity, 65.4% (155/237) of respondents identified as White, 20.7% (49/237) identified as Black or African American, 5.9% (14/237) identified as Asian or Asian American, 2.5% (6/237) identified as Latino/a/x or Spanish origin, 1.7% (4/237) identified as American Indian or Alaskan Native, 2.5% (6/237) selected other race or ethnicity, 0.8% (2/237) selected biracial or multiracial, 0.4% (1//237) selected Arab American/Middle Eastern or North African.

XSL•FO
RenderX

Figure 1. The flow of data from consent to analysis. ASSQ: Acculturative Stress Sensitivity Scale; DDS: Devaluation-Discrimination Scale; SSOHS; Self-Stigma of Seeking Help Scale.

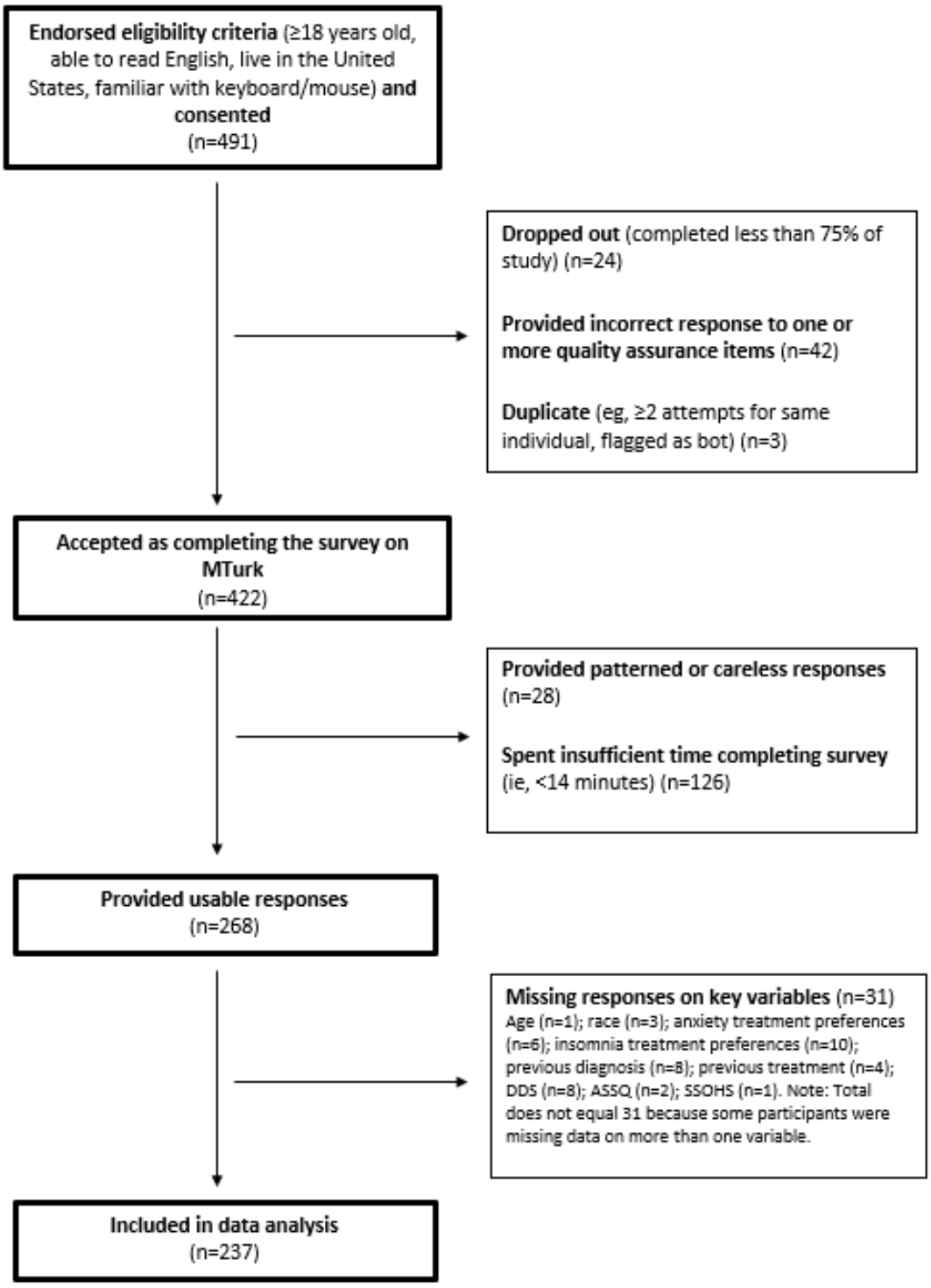

Procedure

The survey was administered on Qualtrics between July 19, 2021, and July 26, 2021. Participants were recruited through Amazon's MTurk platform and received US $3 for study participation. Participants were asked to provide demographic information. Then, participants read 2 vignettes and ranked mental health treatment preference options pertaining to each diagnosis presented in the vignette, with the following question: "Imagine that you have symptoms of [anxiety/insomnia] like [name of the person in the vignette] and want to seek therapy for these symptoms, please rank order the following imagining that you will be interested in seeking treatment for anxiety/insomnia at a later point." The use of vignette-based questions is common in preference research [34] and has been shown to be effective in providing an environment to share

XSL•FO
RenderX

real-life experiences participants may otherwise be reluctant to offer [35]. In this study, 2 vignettes, one depicting GAD and the other, insomnia, were presented in a counterbalanced fashion to evaluate preference for in-person care, telehealth care, digital care, or pharmacologic care (more details in Multimedia Appendix 1). After completing vignette ratings, participants then completed questionnaires on their present levels of generalized anxiety and insomnia (GAD and Sleep Condition Indicator-8 [SCI-8]) and current interest in treatment for these concerns, presented in counterbalanced order. Participants then completed questionnaires about additional experiences that may impact treatment preferences (Patient Health Questionnaire-8 [PHQ-8], Structural Barriers, Shame Inventory [SI], Devaluation-Discrimination Scale [DDS], Self-Stigma of Seeking Help Scale [SSOSH], Brief Perceived Ethnic Discrimination Questionnaire-Community Version [Brief PEDQ-CV], and Acculturative Stress Sensitivity Scale [ASSQ]). Participants were subsequently asked to provide information on previous mental health diagnoses and whether they received treatment (8 participants indicated "declined to respond" regarding their previous mental health were coded as having no previous mental health treatment).

Measures

Anxiety Symptoms and Related Treatment Preferences

The GAD-7 is a self-report measure used to assess the severity of symptoms related to GAD from the past week [36]. It was shown to have good reliability and criterion and construct validity [36]. Participants indicate agreement with 7 statements using a 4-point Likert scale from 0 (Not at all) to 3 (Nearly every day), with responses on all items summed to create a total score ranging from 0 to 21. Higher scores indicate greater anxiety. The GAD-7 had excellent internal consistency in the current sample (α=.91). An additional item was added to assess participants' interest in treatment (yes or no), given their current level of anxiety.

Insomnia Symptoms and Related Treatment Preferences

The Sleep Condition Indicator-8 (SCI-8) is an 8-item scale used to screen for insomnia [37]. The scale can also be used to assist in diagnosis and for clinical and self-monitoring of changes in symptoms related to treatment [38]. It is considered reliable and valid. Participants respond using several 5-point Likert scales, all ranging from 0 to 4; all items are summed to create a total score ranging from 0 to 32. Higher scores indicate more mild symptoms, whereas scores of 16 or less reflect the recognized threshold criteria for insomnia disorder [37]. SCI-8 had good internal consistency in the current sample (α=.87). An additional item was added to assess participants' interest in treatment (yes or no), given their current level of sleep difficulties.

Depression Symptoms

PHQ-8 is a self-report scale providing a reliable and valid measure of the severity of depression [39,40]. Participants are asked to respond to 8 items using a 4-point scale from 0 (Not at all) to 3 (Nearly every day); all items are summed to create a composite score ranging from 0 to 24 [39]. Higher scores indicate stronger severity. In this study, this measure had excellent internal consistency in the current sample (α=.91).

Structural Barriers

A 6-item Structural Barriers scale that was developed for this study was used to assess the extent different structural barriers inhibit individuals from engaging in in-person mental health treatment (ie, lack of insurance coverage, cost, unsure about where to go or whom to see, the time required or inconvenience, being unable to get an appointment, problems with childcare, transportation, or scheduling). Items for this scale were adapted from a list of barriers to treatment identified in previous research [41]. Participants respond using a 4-point Likert scale ranging from 0 (Not at all) to 3 (To a great extent). Higher scores indicate greater hesitancy to engage due to structural barriers. This measure had acceptable internal consistency in the current sample (α=.79).

Shame

The SI is a 3-item self-report scale used to assess how individuals experience shame in relation to global and specific life events [42]. The Shame Inventory has been found to have high internal consistency, reliability, and construct validity [42]. There are 2 parts to this scale; part 1 asks about overall shame and was the only part used in this study. Participants responded on a 5-point Likert scale from 0 to 4. Items are summed, with higher scores indicating higher degrees of shame. The internal consistency of this measure in the current sample was acceptable (α=.77).

Mental Health Stigma

The DDS is a 12-item self-report measure used to assess the extent to which an individual believes others will discriminate or devalue someone with a mental illness [43]. It is one of the most frequently used measures to determine the perception of social stigma pertaining to mental illness and has high internal consistency and reliability [44]. Participants respond on a 6-point Likert scale from 1 (Strongly agree) to 6 (Strongly disagree), with higher scores indicating increased perception of public stigma. This measure has excellent internal consistency in the current sample (α=.9).

The SSOSH is a 10-item scale used to measure the self-stigma associated with seeking psychological help [45]. SSOSH has strong internal consistency reliability and test-retest reliability while also supporting its validity [45]. Participants are asked to evaluate the degree to which statements describe how they might react in situations related to seeking professional health (eg, My self-esteem would increase if I talked to a therapist, I would feel inadequate if I went to a therapist for psychological help) using a 5-point Likert scale from 1 (Strongly disagree) to 5 (Strongly agree). The internal consistency of this measure, the only measure with reverse-coded items in this study, was low (α=.53).

Racial and Ethnic Discrimination

The Brief PEDQ-CV is a 17-item measure used to assess perceptions of racial and ethnic discrimination [46]. Participants are asked to indicate how often various situations have happened to them, based specifically on their ethnicity, using a 5-point Likert scale from 1 (Never happened) to 5 (Happened very often). Higher scores indicate a higher perception of racial and

XSL•FO
RenderX

ethnic discrimination. The internal consistency of this measure in this study was excellent (α=.97).

The ASSQ is a 28-item scale used to measure an individual's sensitivity to discrimination and acculturative stress [47]. Participants respond on a 5-point Likert scale ranging from 1 (None) to 5 (Much or very much), with higher scores indicating greater acculturative stress. The internal consistency of this measure in the current sample was excellent (α=.98).

Data Analysis Strategy

Exploded logistic regression analyses with exact ties were estimated in SAS (version 9.4; SAS Institute) to examine systematic differences in treatment preferences (telehealth, digital, pharmacologic, and in-person therapy), with in-person therapy as the reference value. Data were structured such that each participant had 4 rows, 1 for each treatment option, and the outcome for each row was the assigned rank for that treatment option (1, 2, 3, or 4, with 1 being the most preferred and 4 being the least preferred). This was accomplished using PROC PHREG with exact ties and specifying participant ID numbers as strata to account for multiple observations.

After examining the main effects, interaction analyses were conducted to test whether participant characteristics were associated with rank choices by entering the multiplicative interaction term for each covariate separately with the nonreference choice values (ie, telehealth, app, and pharmacologic). Interactions were tested for significance using omnibus tests, such that a significant interaction indicated that this characteristic was associated with ranked choice broadly rather than examining only pairwise comparisons. However, individual *P* values for pairwise comparisons in interaction analyses are reported for completeness.

Ethical Considerations

All procedures were approved by the university Boston University Institutional Review Board (protocol #6064X). Electronic informed consent was obtained from all participants, who were informed of their right to withdraw from the study at any time. Participants received $3 for study completion. The data presented in this article have been deidentified.

Results

Symptom Severity and Treatment-Seeking

In terms of disorder status and treatment history, 34.2% (81/237) of participants self-reported 1 previous mental health diagnosis, 31.7% (75/237) of participants self-reported more than 1 previous mental health diagnosis, and 34.2% (81/237) self-reported no history of mental health diagnosis. Out of the full sample, 43.5% (103/237) reported previous mental health treatment, and 3% (7/237) of participants preferred not to answer. The mean score on the GAD-7 was 9.41 (SD 5.55), and 55.3% (131/237) of participants had scores in the clinical range (scores ≥10). Correspondingly, 61.2% (145/237) of the sample indicated a current interest in receiving treatment for GAD. The mean score on the SCI was 18.69 (SD 6.65), with 59.5% (141/237) scoring in the range for likely insomnia (scores ≤16). Finally, 57.4% (136/237) of the sample indicated a present interest in receiving treatment for their insomnia symptoms.

Treatment Preferences for Generalized Anxiety Disorder

Table 1 shows the mean treatment preference rankings for anxiety, as calculated using exploded logistic regression. There was no evidence that participants systematically preferred 1 type of treatment or tended to rank treatment options in a specific order (χ^2_3=4.25; *P*=.24). However, testing pairwise comparisons indicated that telehealth care was ranked as less preferable than in-person care (more details in Table 1).

There was no evidence that participant age (*P*=.41), gender (*P*=.51), or race or ethnicity (*P*=.74) were associated with anxiety treatment preference. Neither were previous mental health treatment (*P*=.61), previous mental health diagnosis (*P*=.4), baseline anxiety symptoms (*P*=.32), nor baseline sleep difficulties (*P*=.59) associated with overall treatment preferences for anxiety.

While baseline depression was not associated with treatment preference overall (χ^2_3=6.27, *P*=.10), pairwise comparisons found a significant increase in the odds of preferring pharmacologic versus in-person therapy for anxiety as depression symptoms increased (odds ratio [OR] 1.03, 95% CI 1.00-1.05; *P*=.02). Similarly, while the presence of structural barriers was not associated with treatment preference for anxiety overall, (χ^2_3=3.5; *P*=.32), pairwise comparisons found a marginally significant increase in the odds of preferring telehealth versus in-person therapy as structural barriers increased (OR 1.04, 95% CI 0.997-1.08; *P*=.07). None of the other psychological predictors were associated with treatment preference (shame: *P*=.81; devaluation discrimination: *P*=.53; self-stigma of seeking help: *P*=.54; acculturative stress sensitivity: *P*=.58; or racial and ethnic discrimination: *P*=.27).

XSL•FO
RenderX

Table 1. Preference ratings for each treatment modality with in-person treatment representing the comparative standard.

Ratings	*B*	OR[a] (95% CI)	*P* value	Rank, mean (SD)
Ratings for the treatment of generalized anxiety disorder				
In-person	Ref[b]			2.30 (1.21)
Digital	–0.14	0.87 (0.70-1.07)	.18	2.56 (1.07)
Telehealth	–0.22	0.80 (0.65-0.99)	.04	2.56 (1.17)
Pharmacologic	–0.10	0.90 (0.73-1.11)	.33	2.59 (0.99)
Ratings for the treatment of insomnia				
In-person	Ref			2.32 (1.30)
Digital	0.08	1.09 (0.88-1.34)	.45	2.46 (1.03)
Telehealth	0.003	1.00 (0.81-1.24)	.04	2.44 (1.16)
Pharmacologic	–0.13	0.88 (0.70-1.08)	.22	2.78 (0.90)

[a]OR: odds ratio.

[b]Reference.

Treatment Preferences for Insomnia

Table 1 shows the mean treatment preference ratings for insomnia. There was no evidence that participants systematically preferred one type of treatment or tended to rank treatment options in a specific order (χ^2_3=4.4; *P*=.22).

While participant age (*P*=.08) was not significantly associated with overall treatment preferences for insomnia, pairwise comparisons indicated a significant increase in the odds of preferring telehealth versus in-person treatment as age increased (OR 1.02, 95% CI 1.00-1.03; *P*=.02). Neither gender (*P*=.55), race or ethnicity (*P*=.73), previous mental health treatment (*P*=.47), previous mental health diagnosis (*P*=.49), baseline anxiety symptoms (*P*=.38), nor baseline sleep difficulties (*P*=.90) were associated with treatment preferences for insomnia.

While baseline depression was not associated with insomnia treatment preference in terms of overall ranking patterns (χ^2_3=5.06; *P*=.17), pairwise comparisons (ie, examining the impact of depression on only the preference of pharmacologic vs in-person therapy) found a significant increase in the odds of preferring pharmacologic versus in-person therapy as depression symptoms increased (OR 1.02, 95% CI 1.00-1.05; *P*=.05). Similarly, while shame was not associated with treatment preference overall (χ^2_3=4.69; *P*=.2), pairwise comparisons indicated that there was a marginally significant increase in the odds of preferring pharmacologic versus in-person therapy as shame increased (OR 1.05, 95% CI 0.998-1.11; *P*=.06). None of the other psychological variables were associated with treatment preference for insomnia (structural barriers: *P*=.57; devaluation discrimination: *P*=.89; self-stigma of seeking help: *P*=.09; acculturative stress sensitivity: *P*=.17; and racial and ethnic discrimination: *P*=.29).

Discussion

The purpose of this study was to examine relative treatment preferences for 4 different types of mental health care: in-person care, telehealth care, digital care, and pharmacologic care. Respondents reported familiarity with the mental health system, with almost 66% (156/237) reporting a previous mental health diagnosis and almost 44% (103/237) reporting previous mental health treatment; well over half the sample reported interest in future treatment for GAD or insomnia. For this sample, across these 2 referent disorders, there was consistent evidence that telehealth was less preferred than in-person psychological care. No significant difference in preference was found between in-person care and either digital care or pharmacologic care. Furthermore, of the range of demographic and psychological factors examined, moderation of preferences was evident only for depression and age. For rankings of both anxiety and insomnia care, the presence of greater depression was associated with a greater relative preference for pharmacologic than in-person care. Nonetheless, when all care options were considered, the degree to which pharmacologic care was selected as a first-choice option (approximately 22% (52/237) of the sample) was in strong accord with the 3-fold preference for psychosocial treatment options over pharmacologic treatment observed in a meta-analysis of this issue [34].

We did not replicate the common finding [20-22] of a greater preference for digital care among younger adults. Indeed, we found that telehealth care was relatively more preferred to in-person care among older rather than younger participants (within the age range evaluated: 19 to 68 years), although this association was found only for insomnia treatment. It is unclear whether this apparent loss of youth-based moderation of preferences for digital care is a result of the widespread adoption of more remote care options during the COVID-19 pandemic [23,24]. There are strong indications that the introduction of telehealth during the COVID-19 pandemic quickly led to a preference to continue this modality of care for mental health issues [48]. Likewise, a recent cross-sectional study reported that 73% of participants agreed or strongly agreed that the pandemic made them more open to telehealth in general [49]. A longitudinal examination of public discourse related to telehealth from March 2020 to April 2021 similarly suggests that continued widespread adoption and use of telehealth services is likely [50]. As such, one unexpected consequence

XSL•FO
RenderX

of the COVID-19 pandemic may be a more widespread acceptance of telehealth and digital treatments such that some of the traditional predictors of preferences for digital treatment (ie, younger adults) are no longer relevant.

Among psychological predictors, few appeared to have a discernible effect on treatment preferences. Greater barriers showed a nonsignificant trend toward predicting a greater preference for telehealth than in-person care for GAD. More consistent effects were evident for depression severity; for treatment selection for both generalized anxiety and insomnia, a greater preference for pharmacologic relative to in-person psychological treatment was found at greater levels of depression severity.

Empirically supported digital treatment has been recommended as a strategy to reduce stigma for care [5,7], yet we did not find a relationship between stigma and preference for digital care, although we did observe a trend level association between reported shame and a relatively greater preference for telehealth than in-person therapy for insomnia. What we documented instead was acceptance of digital care that rivaled that for in-person psychotherapy, telehealth psychotherapy, or pharmacotherapy, with approximately equal numbers of participants selecting each of these choices as a first preference option. An implication of this finding is that health systems and providers may need to be prepared to provide multiple strategies for care to enhance treatment acceptance among patients.

It is important to note a number of limitations to our study findings. First, our sample was recruited through an online survey, and accordingly, only those with access to the internet were able to partake in the study. As the purpose of this study was to examine preference for treatment options that included telehealth and digital care, the assessment of individuals willing to engage with a digital assessment may have biased the results. That is, those who are unable to attain easy access to the internet may have been less inclined to prefer a telehealth or digital treatment method compared with those who already have access. Second, our results were based on responses to vignettes depicting GAD or insomnia, so treatment preferences were derived only for these specific disorders. Different relative preferences may well be present for different disorders. Third, our findings were specific to the preference for treatment initiation and do not address how preferences may change as travel or cost burdens, side effects, or other adherence issues are experienced during treatment [12,21,51]. Fourth, our preference ratings presupposed that all modalities of treatment were equally available to participants and ignored the many settings or localities where specific empirically supported mental health treatment options may not be available. Finally, we recruited a few individuals in the oldest age ranges (eg, only 3% of the sample was 65 years and older), so our findings may not apply to the preferences of older adults in this age range.

Overall, our results are suggestive of an evening-out of care preferences, where digital care competes well with other modalities of care for mental health issues, specifically GAD and insomnia. Given the many individuals who fail to receive timely care for these and other disorders [1,52], our findings provide encouragement for further digital care development, validation, and dissemination to try to address these care needs for interested patients.

Acknowledgments

This study was supported by a subaward to the first author from the Columbia Roybal Center for Fearless Behavior Change NIA award P30AG064198. Work on this project was also supported by a Summer Research Project Award to the second author from Boston University's Institute for Health System Innovation and Policy. In addition, this work was supported by the National Institute on Aging of the National Institutes of Health (grant number R24AG064191). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Conflicts of Interest

MWO receives compensation for his role on scientific and clinical advisory boards for Big Health, a digital mental health company. EMP and MWO have received grants from Big Health and NIH for the study of digital CBT outcomes. None of these entities provided support for this project and played no role in the decision to submit this paper or the content thereof.

Multimedia Appendix 1
Vignettes for provider preferences.
[PDF File (Adobe PDF File), 124 KB - jmir_v27i1e54608_app1.pdf]

References

1. Demyttenaere K, Bruffaerts R, Lee S, Posada-Villa J, Kovess V, Angermeyer MC, et al. Mental disorders among persons with chronic back or neck pain: results from the world mental health surveys. Pain 2007;129(3):332-342. [doi: 10.1016/j.pain.2007.01.022] [Medline: 17350169]
2. Goetter EM, Frumkin MR, Palitz SA, Swee MB, Baker AW, Bui E, et al. Barriers to mental health treatment among individuals with social anxiety disorder and generalized anxiety disorder. Psychol Serv 2020;17(1):5-12. [doi: 10.1037/ser0000254] [Medline: 30070552]

XSL•FO
RenderX

3. Sorkin DH, Pham E, Ngo-Metzger Q. Racial and ethnic differences in the mental health needs and access to care of older adults in California. J Am Geriatr Soc 2009;57(12):2311-2317. [doi: 10.1111/j.1532-5415.2009.02573.x] [Medline: 19943830]
4. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005;62(6):603-613. [doi: 10.1001/archpsyc.62.6.603] [Medline: 15939838]
5. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
6. Ollendick TH, Öst LG, Farrell LJ. Innovations in the psychosocial treatment of youth with anxiety disorders: implications for a stepped care approach. Evid Based Ment Health 2018;21(3):112-115 [FREE Full text] [doi: 10.1136/eb-2018-102892] [Medline: 29976564]
7. Wilhelm S, Weingarden H, Ladis I, Braddick V, Shin J, Jacobson NC. Cognitive-behavioral therapy in the digital age: presidential address. Behav Ther 2020;51(1):1-14 [FREE Full text] [doi: 10.1016/j.beth.2019.08.001] [Medline: 32005328]
8. British Psychological Society. Generalised anxiety disorder in adults: management in primary, secondary and community care. In: Royal College of Psychiatrists. Leicester, UK: British Psychological Society; 2011.
9. Adelman CB, Panza KE, Bartley CA, Bontempo A, Bloch MH. A meta-analysis of computerized cognitive-behavioral therapy for the treatment of DSM-5 anxiety disorders. J Clin Psychiatry 2014;75(7):e695-e704. [doi: 10.4088/JCP.13r08894] [Medline: 25093485]
10. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
11. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
12. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
13. Tucker JR, Hammer JH, Vogel DL, Bitman RL, Wade NG, Maier EJ. Disentangling self-stigma: are mental illness and help-seeking self-stigmas different? J Couns Psychol 2013;60(4):520-531. [doi: 10.1037/a0033555] [Medline: 23815629]
14. Cooper-Patrick L, Powe NR, Jenckes MW, Gonzales JJ, Levine DM, Ford DE. Identification of patient attitudes and preferences regarding treatment of depression. J Gen Intern Med 1997;12(7):431-438 [FREE Full text] [doi: 10.1046/j.1525-1497.1997.00075.x] [Medline: 9229282]
15. Sussman LK, Robins LN, Earls F. Treatment-seeking for depression by black and white Americans. Soc Sci Med 1987;24(3):187-196. [doi: 10.1016/0277-9536(87)90046-3] [Medline: 3824001]
16. Nadeem E, Lange JM, Edge D, Fongwa M, Belin T, Miranda J. Does stigma keep poor young immigrant and U.S.-born Black and Latina women from seeking mental health care? Psychiatr Serv 2007;58(12):1547-1554. [doi: 10.1176/ps.2007.58.12.1547] [Medline: 18048555]
17. Pingitore D, Snowden L, Sansone RA, Klinkman M. Persons with depressive symptoms and the treatments they receive: a comparison of primary care physicians and psychiatrists. Int J Psychiatry Med 2001;31(1):41-60. [doi: 10.2190/6bul-mwtq-0m18-30gl]
18. Wallin E, Maathz P, Parling T, Hursti T. Self-stigma and the intention to seek psychological help online compared to face-to-face. J Clin Psychol 2018;74(7):1207-1218. [doi: 10.1002/jclp.22583] [Medline: 29315545]
19. Lungu A, Sun M. Time for a change: college students' preference for technology-mediated versus face-to-face help for emotional distress. Telemed J E Health 2016;22(12):991-1000. [doi: 10.1089/tmj.2015.0214] [Medline: 27223266]
20. Kozlov E, McDarby M, Prescott M, Altman M. Assessing the care modality preferences and predictors for digital mental health treatment seekers in a technology-enabled stepped care delivery system: cross-sectional study. JMIR Form Res 2021;5(9):e30162 [FREE Full text] [doi: 10.2196/30162] [Medline: 34343965]
21. Linardon J, Shatte A, Tepper H, Fuller-Tyszkiewicz M. A survey study of attitudes toward, and preferences for, e-therapy interventions for eating disorder psychopathology. Int J Eat Disord 2020;53(6):907-916. [doi: 10.1002/eat.23268] [Medline: 32239725]
22. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
23. Garfan S, Alamoodi AH, Zaidan BB, Al-Zobbi M, Hamid RA, Alwan JK, et al. Telehealth utilization during the Covid-19 pandemic: a systematic review. Comput Biol Med 2021;138:104878 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
24. Lu M, Liao X. Telehealth utilization in U.S. Medicare beneficiaries aged 65 years and older during the COVID-19 pandemic. BMC Public Health 2023;23(1):368 [FREE Full text] [doi: 10.1186/s12889-023-15263-0] [Medline: 36803677]

XSL•FO
RenderX

25. Childs AW, Bacon SM, Klingensmith K, Li L, Unger A, Wing AM, et al. Showing up is half the battle: the impact of telehealth on psychiatric appointment attendance for hospital-based intensive outpatient services during COVID-19. Telemed J E Health 2021;27(8):835-842. [doi: 10.1089/tmj.2021.0028] [Medline: 33999738]
26. Cuthbert K, Parsons EM, Smith L, Otto MW. Acceptability of telehealth CBT during the time of COVID-19: evidence from patient treatment initiation and attendance records. Journal of Behavioral and Cognitive Therapy 2022;32(1):67-72. [doi: 10.1016/j.jbct.2022.01.001]
27. Chartier-Otis M, Perreault M, Bélanger C. Determinants of barriers to treatment for anxiety disorders. Psychiatr Q 2010;81(2):127-138. [doi: 10.1007/s11126-010-9123-5] [Medline: 20131095]
28. Kyle SD, Espie CA, Morgan K. '...Not just a minor thing, it is something major, which stops you from functioning daily': quality of life and daytime functioning in insomnia. Behav Sleep Med 2010;8(3):123-140. [doi: 10.1080/15402002.2010.487450] [Medline: 20582756]
29. Darden M, Espie CA, Carl JR, Henry AL, Kanady JC, Krystal AD, et al. Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: a Markov simulation model in the United States. Sleep 2021;44(4):zsaa223. [doi: 10.1093/sleep/zsaa223] [Medline: 33151330]
30. McCord C, Ullrich F, Merchant KAS, Bhagianadh D, Carter KD, Nelson E, et al. Comparison of in-person vs. telebehavioral health outcomes from rural populations across America. BMC Psychiatry 2022;22(1):778 [FREE Full text] [doi: 10.1186/s12888-022-04421-0] [Medline: 36496352]
31. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol. Med 2021:1-13. [doi: 10.1017/s0033291721001999]
32. Aguinis H, Villamor I, Ramani RS. MTurk research: review and recommendations. Journal of Management 2020;47(4):823-837. [doi: 10.1177/0149206320969787]
33. Chmielewski M, Kucker SC. An MTurk crisis? Shifts in data quality and the impact on study results. Social Psychological and Personality Science 2019;11(4):464-473. [doi: 10.1177/1948550619875149]
34. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders. J. Clin. Psychiatry 2013;74(06):595-602. [doi: 10.4088/jcp.12r07757]
35. Murphy J, Hughes J, Read S, Ashby S. Evidence and practice: a review of vignettes in qualitative research. Nurse Res 2021;29(3):8-14. [doi: 10.7748/nr.2021.e1787] [Medline: 34041889]
36. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. Espie CA, Kyle SD, Hames P, Gardani M, Fleming L, Cape J. The sleep condition indicator: a clinical screening tool to evaluate insomnia disorder. BMJ Open 2014;4(3):e004183 [FREE Full text] [doi: 10.1136/bmjopen-2013-004183] [Medline: 24643168]
38. Espie CA, Farias Machado P, Carl JR, Kyle SD, Cape J, Siriwardena AN, et al. The Sleep Condition Indicator: reference values derived from a sample of 200 000 adults. J Sleep Res 2018;27(3):e12643. [doi: 10.1111/jsr.12643] [Medline: 29193493]
39. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
40. Wu Y, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, et al. Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. Psychol Med 2019;50(8):1368-1380. [doi: 10.1017/s0033291719001314]
41. Green JG, McLaughlin KA, Fillbrunn M, Fukuda M, Jackson JS, Kessler RC, et al. Barriers to mental health service use and predictors of treatment drop out: racial/ethnic variation in a population-based study. Adm Policy Ment Health 2020;47(4):606-616 [FREE Full text] [doi: 10.1007/s10488-020-01021-6] [Medline: 32076886]
42. Rizvi SL. Development and preliminary validation of a new measure to assess shame: the shame inventory. J Psychopathol Behav Assess 2009;32(3):438-447. [doi: 10.1007/s10862-009-9172-y]
43. Link BG. Understanding labeling effects in the area of mental disorders: an assessment of the effects of expectations of rejection. American Sociological Review 1987;52(1):96-112. [doi: 10.2307/2095395]
44. Şahin E, Topkaya N. Adaptation of the devaluation-discrimination scale into turkish: a comprehensive psychometric analysis. Assessment 2023;30(1):51-72. [doi: 10.1177/10731911211039284] [Medline: 34459254]
45. Vogel DL, Wade NG, Haake S. Measuring the self-stigma associated with seeking psychological help. Journal of Counseling Psychology 2006;53(3):325-337. [doi: 10.1037/0022-0167.53.3.325]
46. Brondolo E, Kelly KP, Coakley V, Gordon T, Thompson S, Levy E, et al. The perceived ethnic discrimination questionnaire: development and preliminary validation of a community version. J Appl Social Pyschol 2005;35(2):335-365. [doi: 10.1111/j.1559-1816.2005.tb02124.x]
47. Sandhu DS, Asrabadi BR. Development of an acculturative stress scale for international students: preliminary findings. Psychol Rep 1994;75(1 Pt 2):435-448. [doi: 10.2466/pr0.1994.75.1.435] [Medline: 7809315]
48. Severe J, Tang R, Horbatch F, Onishchenko R, Naini V, Blazek MC. Factors influencing patients' initial decisions regarding telepsychiatry participation during the COVID-19 pandemic: telephone-based survey. JMIR Form Res 2020;4(12):e25469 [FREE Full text] [doi: 10.2196/25469] [Medline: 33320823]

XSL•FO
RenderX

49. Crouch J, Winters K, Zhang L, Stewart MW. Telehealth during the pandemic: patient perceptions and policy implications. J Nurs Scholarsh 2023;55(1):141-148 [FREE Full text] [doi: 10.1111/jnu.12832] [Medline: 36218219]
50. Anderson JT, Bouchacourt LM, Sussman KL, Bright LF, Wilcox GB. Telehealth adoption during the COVID-19 pandemic: a social media textual and network analysis. Digit Health 2022;8:20552076221090041 [FREE Full text] [doi: 10.1177/20552076221090041] [Medline: 35392254]
51. Affengruber L, Wagner G, Dobrescu A, Toromanova A, Chapman A, Persad E, et al. Values and preferences of patients with depressive disorders regarding pharmacologic and nonpharmacologic treatments. Ann Intern Med 2023;176(2):217-223. [doi: 10.7326/m22-1900]
52. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005;62(6):629-640. [doi: 10.1001/archpsyc.62.6.629] [Medline: 15939840]

Abbreviations

ASSQ: Acculturative Stress Sensitivity Scale
Brief PEDQ-CV: Brief Perceived Ethnic Discrimination Questionnaire-Community Version
dCBT: digital cognitive behavior therapy
DDS: Devaluation-Discrimination scale
GAD: generalized anxiety disorder
MTurk: Mechanical Turk
OR: odds ratio
PHQ-8: Patient Health Questionnaire-8
SCI-8: Sleep Condition Indicator-8
SI: Shame Inventory
SSOSH: Self-Stigma of Seeking Help Scale

Edited by E Wethington, J Kaye, A Dominello, I Kronish; submitted 15.11.23; peer-reviewed by J Wang, J Severe; comments to author 05.07.24; revised version received 30.08.24; accepted 26.11.24; published 13.02.25.

Please cite as:
Parsons EM, Figueroa ZG, Hiserodt M, Cornelius T, Otto MW
Relative Preference for In-Person, Telehealth, Digital, and Pharmacologic Mental Health Care After the COVID-19 Pandemic: Cross-Sectional Questionnaire Study
J Med Internet Res 2025;27:e54608
URL: https://www.jmir.org/2025/1/e54608
doi:10.2196/54608
PMID:

©E Marie Parsons, Zoë G Figueroa, Michele Hiserodt, Talea Cornelius, Michael W Otto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Pathways to Sustainable Health Care Development: Study on the Carbon Reduction Potential of Telemedicine in China

Zhi Li[1], ME; Lianrui Xiang[1], MM; Jing Ning[1], MHRM; Wenbo Li[1], MBA; Yong Huang[1], MBA; Xue Xiao[1], MD

Public Service and Development Office, Key Laboratory of Birth Defects and Related Diseases of Women and Children, West China Second University Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Xue Xiao, MD
Public Service and Development Office, Key Laboratory of Birth Defects and Related Diseases of Women and Children
West China Second University Hospital
Sichuan University
20# Section 3 South Renmin Road
Chengdu
China
Phone: 86 18583974588
Email: hxey_science@163.com

Abstract

Background: Carbon emissions are a global concern due to their significant greenhouse effect. The health care sector's greenhouse gas (GHG) emissions must be controlled. Telemedicine in China continued to grow between 2020 and 2022, offering a promising solution for reducing carbon emissions in the country's health sector.

Objective: This study explores the potential of telemedicine for reducing GHG emissions and saving energy through a life cycle assessment.

Methods: This study used the Chinese Environmentally Extended Input-Output database, which is based on input-output models, to quantify the GHG emissions and energy consumption associated with care outpatient and telemedicine from a life cycle perspective. Data collected from the West China Second University Hospital of Sichuan University between 2020 and 2022 were incorporated into the analyses.

Results: The findings indicated that telemedicine could reduce GHG emissions by 36 tCO_2e. The GHG emissions per telemedicine session were only 19.14 $kgCO_2e$ and are expected to decrease from 2025 to 2030. Replacing in-person treatments with telemedicine can lead to an average reduction of 85.51 $kgCO_2e$ emissions.

Conclusions: In the future, the widespread adoption of telemedicine could help achieve carbon neutrality in the health care sector. Telemedicine is crucial for establishing the sustainable development of the health care sector.

(J Med Internet Res 2025;27:e63927) doi:10.2196/63927

KEYWORDS

telemedicine; carbon reduction potential; carbon emissions; energy consumption; sustainable development

Introduction

Background

Excessive greenhouse gas (GHG) emissions and energy usage have garnered global attention as the greenhouse impact intensifies. The Paris Agreement mandates a 43% reduction in GHG emissions by 2030 to achieve net zero by 2050 [1]. The health care sector's GHG emissions are 1.57 times higher than those of the worldwide air transport industry, estimated at 4.4% of total global net emissions, with the United States, China, and European Union countries responsible for 56% of the total global health care GHG emissions. China's health care sector is the third largest contributor to carbon emissions globally [2]. As one of the leading nations in global carbon emissions, China has been implementing measures to conserve energy and diminish carbon output. China is aiming to achieve net zero carbon emissions by promoting scientific and technological advancements and developing a comprehensive low-carbon technological innovation framework [3]. A correlation exists between expenditure in the industrial sector and GHG emissions. The total expenditure of China's health care sector as a proportion of the gross domestic product (GDP) increased from 6.43% in 2018 to 7.12% in 2020 [4]. A disparity persists

XSL•FO
RenderX

between China's health care expenditures and those of industrialized nations. As the health care industry continues to expand, its GHG emissions will also increase. The health care sector in China must use multiple strategies to mitigate GHG emissions and conserve energy.

Telemedicine refers to the use of network and information and communication technologies (ICTs) to provide remote access to health care services. It is widely used to deliver telehealth care and reduce medical expenses for patients [5,6]. The evolution of telemedicine in China has transpired via 3 phases. The initial phase, spanning from 2000 to 2015, was the early exploration period, giving rise to medical information and medical consultation businesses. The second phase, spanning from 2016 to 2019, was known as the policy development period, during which telemedicine-related policies were progressively implemented, enabling the integration of telemedicine into diagnostics. The third phase, from 2020 to 2022, was characterized by remarkable business expansion. In the context of the COVID-19 pandemic, in-person outpatient visits were halted, leading to an unprecedented surge in the demand for telemedicine. This shift altered public perceptions of health care, prompting numerous public medical institutions to adopt telemedicine, thereby enhancing its significance.

Literature Review

Telemedicine has cost-reduction benefits for patients who require time off from work, undergo alternative treatment, or face long-distance transportation challenges [7,8]. It can improve patients' access to health care by reducing the overall trip time and distance, making it more convenient for patients to receive in-person treatment [9]. The current clinical implementation of telemedicine in China primarily focuses on managing chronic diseases in internal medicine and providing home care services for postoperative patients. By enabling virtual communication, telemedicine can enhance patient care and satisfaction [10-12].

Previous research has demonstrated that telemedicine has beneficial environmental effects. It is acknowledged as a viable approach to reduce the carbon footprint by eliminating the need for patients to travel to health care facilities, potentially decreasing GHG emissions. Moreover, it can provide financial savings for patients in rural areas while promoting environmental sustainability [13,14]. Several studies have examined carbon emissions associated with telemedicine services in some regions. The findings suggest that teleconsultations delivered via telephone and video platforms are more environmentally sustainable than in-person consultations [15-17].

Telemedicine has grown significantly since the 21st century and gained widespread adoption, particularly during the COVID-19 pandemic in China. Research in telemedicine has primarily focused on system construction, remote consultation technology, and clinical application. While China's telemedicine policies are based on overarching regulations, they lack supportive measures such as financial assistance and proactive promotion [18]. To ensure the sustainable advancement of telemedicine, the government must increase its investment and construct a long-term operational framework [19]. Implementing telemedicine access policies will impact the construction and sustainable development of telemedicine platforms [20]. As a low-carbon development approach, telemedicine strategies should be tailored to local infrastructure conditions [21]. Artificial intelligence, blockchain, and robotics present new opportunities for the development of telemedicine, enabling direct interactions between doctors and remote patients through various platform models [22]. With advancements in medical service policies and telemedicine technology, China's ecological welfare performance and citizen-centered sustainable community services will continue to improve [23,24].

Prior studies have not conducted quantitative analyses on the ecological impacts and sustainable development of telemedicine in China. Given its extensive applicability, the steady development of telemedicine is important. This study aims to examine the benefits of telemedicine in reducing GHG emissions and conserving energy while also forecasting GHG emissions from outpatient visits and telemedicine utilization in China.

A model was developed using the Chinese Environmentally Extended Input-Output (CEEIO) database within the framework of life cycle assessment (LCA). LCA is an approach used to assess the environmental effects of products, processes, or activities through their life cycle, encompassing raw material acquisition, product manufacturing, transportation, distribution, utilization, potential reuse, maintenance, and ultimate disposal [25]. The LCA model has been applied to assess GHG emissions across various sectors in China, including energy, nuclear power, urban transportation systems, and biofuels [26-28]. As a comprehensive approach, LCA is also applicable for calculating energy conservation and emission reductions in telemedicine services within medical institutions. The CEEIO database has previously been used to conduct environmental assessments in China's building sector, quantify the environmental impact of resource extraction in China, and conduct input-output analyses of China's home water footprint [29-31].

Some studies have quantified the carbon emissions of telemedicine in medical institutions. In 2021, telemedicine services contributed to a 17,000 metric ton reduction in GHG emissions by Stanford Health Care [17]. An investigation of 17 research institutes in Italy indicated that a single telemedicine consultation might reduce carbon emissions by 13 kg [16]. Additionally, face-to-face outpatient clinics at the University Hospital of Northern Sweden generated 40 to 70 times more carbon emissions than telemedicine [15]. Thus, telemedicine has greater environmental advantages.

Despite extensive research on telemedicine policies and clinical applications, with some countries examining its low-carbon potential as an innovative technology, there is a lack of studies on telemedicine's contribution to sustainable development and its life cycle carbon emissions in China. As China pursues carbon neutrality and prioritizes a people-centric development strategy, enhancing the governmental recognition of the low-carbon characteristics of telemedicine is crucial for the sustainable advancement of China's health sector.

This study estimates the GHG emissions and energy consumption associated with telemedicine in China, evaluating its environmental advantages by developing a life cycle model based on patient-related data from teleconsultations at the West

XSL•FO
RenderX

China Second University Hospital of Sichuan University (WCSUH-SCU) from 2020 to 2022. WCSUH-SCU, directly supervised by the National Health Commission of China, is one of the top national tertiary hospitals and serves as a medical center for women and children in southwest China, playing a vital role in medical service, education, and research.

This study aims to reveal the level of telemedicine development and its associated ecological sustainability in southwest China. Consequently, it provides essential data to bolster the sustainable advancement of China's medical sector and offers policy recommendations for decision makers.

Methods

Research Boundaries

The research boundary of this study spans the entire process from beginning to end. At WCSUH-SCU, the telemedicine research boundary covers the development, establishment, and utilization of the telemedicine platform. Similarly, the research boundary on outpatient services at WCSUH-SCU extends from establishing the outpatient appointment platform to the in-person consultations.

The primary focus of this study is to explore the environmental benefits of telemedicine by estimating GHG emissions within both outpatient and telemedicine boundaries for the same patient. As shown in Figure 1, the outpatient appointment stage mainly includes developing the outpatient appointment system and scheduling outpatients. During the treatment stage, patients visit WCSUH-SCU for in-person consultations, with transportation assumed to be via passenger cars. Telemedicine life cycle management mainly includes the construction and usage of telemedicine platforms. The construction stage involves establishing an online telemedicine platform and purchasing computers and audiovisual equipment. During the utilizing stage, all processes are handled by the hospital administrators and doctors through computers and audiovisual equipment.

Figure 1. The research boundaries for telemedicine and outpatient of the West China Second University Hospital of Sichuan University (WCSUH-SCU). GHG: greenhouse gas.

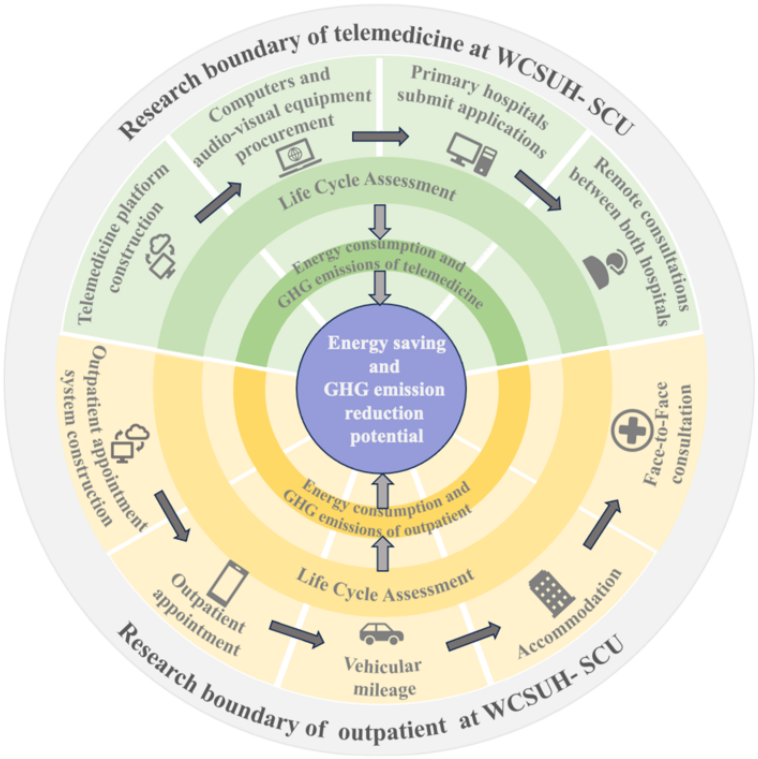

Study Design

LCA summarizes and evaluates the energy and resource inputs and outputs of a product or activity throughout its life cycle, along with its potential environmental impact. In this study, we use LCA to evaluate the GHG emissions and energy consumption of telemedicine and outpatient services in China. Input-output models, commonly used in LCA, estimate embodied energy and carbon emissions based on industrial sector expenditures.

Previous health care environmental impact studies have used similar methods to estimate emissions from preoperative evaluation and cataract surgery [32,33]. In this study, the CEEIO database was used to quantify the GHG emissions from outpatient and telemedicine services. Additionally, Ecoinvent (version 3.9; Ecoinvent Association), a background database in LCA and other environmental assessments, was used to estimate the energy consumption per unit mileage of the car throughout its life cycle. A detailed calculation for the GHG emission reduction and energy-saving benefits of telemedicine is introduced as follows.

The CEEIO database is the result of extensive research based on the environmental extended input-output model tailored to the Chinese context. It consists of benchmark economic input-output tables published by China's government statistical agencies and statistical data from multiple sources, enabling an

XSL•FO
RenderX

extensive analysis of the environmental footprint of various industries in China. In this study, the GHG and energy intensities of China's industries were obtained from version 3.2 of the CEEIO database [34]. Detailed data on these GHG and energy intensities can be found in Table 1, where the energy consumption intensity of the software and health services is derived from the conversion of CO_2e emissions per unit standard carbon.

Table 1. GHG[a] intensities for outpatient and telemedicine industries.

Industry	GHG intensity (gCO_2e[b]/RMB[c])	Energy intensities (gce[d]/RMB)
Software services	2.66	1.08
Computer	0.72	2.05
Audiovisual equipment	0.72	2.05
Hotels	4.33	4.44
Health services	2.66	1.08

[a]GHG: greenhouse gas.

[b]The CO_2e emission per gram of standard coal is 2.46 g

[c]RMB: Renminbi (official Chinese currency).

[d]gce: grams of standard coal equivalent.

The GHG intensities due to transportation, calculated as the weighted average of passenger car emissions in China from 2020 to 2022, were 212.2, 236.07, and 247.77 g CO_2e/km, respectively [35-37]. According to the LCA from version 3.9 of Ecoinvent, the energy consumption per unit mileage of passenger cars was estimated to be 3.95 MJ/km. For electricity, the GHG intensities from 2020 to 2022 were 0.6101, 0.581, and 0.5703 $kgCO_2e$/kWh, respectively [38,39], while energy intensities were 0.305, 0.303, and 0.302 $kgCO_2e$/kWh, respectively [40].

This study used relevant data from patients who underwent telemedicine treatment over a specific period as samples, combined with economic data to calculate the GHG emissions of outpatient service and telemedicine under the same sample. This approach helped confirm the advantages of telemedicine for energy savings and GHG emission reduction.

A mathematical model was developed to estimate the GHG emission and energy consumption of both outpatient and telemedicine services, allowing for an assessment of telemedicine's GHG emission reduction and energy-saving benefits. The details of the method are described by the following 3 equations:



where $E_{outpatient}$ and $E_{telemedicine}$ represent the GHG emissions and energy consumption from outpatient and telemedicine services, respectively. $R_{outpatient}$ and $R_{telemedicine}$ represent the revenue mobilized by each industry involved in outpatient and telemedicine services, respectively. R represents the transportation distance and duration of telemedicine. Similarly, *f* not only represents the GHG and energy intensities of each industry involved in outpatient and telemedicine services, respectively, but also the GHG emissions and energy consumption per unit mileage of transportation, electric GHG emission factors, and energy consumption factors of electricity. E_{saved} represents the GHG emission reduction and energy saving of telemedicine.

Data Collection

The study sample comprised relevant data from 421 patients who completed telemedicine treatment between 2020 and 2022. The data set included the patient's primary hospital, telemedicine department, duration of telemedicine treatment, and telemedicine and outpatient service fees charged by participating doctors. Actual values are displayed in Tables S1 and S2 in Multimedia Appendix 1, taking the telemedicine data from WCSUH-SCU in 2020 as an example.

The round-trip distance for outpatient visits was calculated using the optimal driving route from the primary hospital to WCSUH-SCU, modeled through Python (version 3.7; Python Software Foundation) with Baidu Maps. Actual transportation distances for patients traveling to WCSUH-SCU for outpatient services are shown in Table S3 in Multimedia Appendix 1, using 2020 as a reference.

A hotel reservation was deemed necessary for trips exceeding 300 kilometers one way. Hotel prices were derived from the quarterly average prices of star-rated hotels in Chengdu from 2020 to 2022 [41]. The development costs of the telemedicine platform and outpatient appointment system were derived from the costs incurred by software companies collaborating with WCSUH-SCU. Outpatient and telemedicine service fees were determined according to the standards set by the Sichuan Provincial Development and Reform Commission and Sichuan Provincial Health Care Security Administration [42,43].

Data Analysis

A descriptive analysis was conducted on the distribution of telemedicine departments within WCSUH-SCU. Through modeling and data analysis, determinants influencing GHG emissions and energy consumption in outpatient and telemedicine were prioritized to reveal the principal causes impacting the GHG emissions and energy consumption in outpatient and telemedicine services.

XSL•FO
RenderX

According to the existing literature, patient travel and electricity are the main factors affecting the sustainability of the health sector. This study constructed predictive models based on the average GHG emissions from transportation unit mileage, and electric GHG emission factors were established to reveal the future sustainable development of China's health care sector. The average GHG emissions related to transportation unit mileage from 2023 to 2030 were calculated by weighting the GHG emissions unit mileage of China's passenger cars from 2023 to 2030 with forecast data on vehicle ownership. Carbon emissions per unit mileage of China's passenger cars from 2023 to 2030 were based on forecast data obtained from the China Automobile Low Carbon Action Plan 2021 and 2022, including information on the carbon emissions of various energy vehicles in China. Additionally, forecast data on vehicle ownership with different energy sources in China from 2023 to 2030 were sourced from an automotive aftermarket white paper published by Roland Berger in 2022 [44]. Electric unit GHG emissions were calculated through variations in the proportion of nonfossil energy power generation combined with electricity intensities. The 14th Five-Year Plan for a Modern Energy System anticipates that the proportion of nonfossil energy in power generation will reach 39% by 2025. Simultaneously, the 2023 China Electric Power Industry Annual Development Report anticipates that the share of nonfossil energy power generation will reach 50% by 2030 [40].

Results

Descriptive Analyses

Between 2020 and 2022, a cumulative total of 421 telemedicine consultations were documented. The distribution of teleconsultation departments of WCSUH-SCU from 2020 to 2022 is shown in Figure 2. Over 3 years, telemedicine was implemented throughout 19 distinct clinical departments, encompassing reproductive medicine, pediatric subspecialties, obstetrics, and gynecology. The consultations predominantly comprised pediatric subspecialties, increasing from 55.3% (68/123) in 2020 to 65.5% (93/142) in 2022. The proportion of telemedicine in obstetrics and gynecology declined from 25.2% (31/123) in 2020 to 11.3% (16/142) in 2022. Since the onset of the COVID-19 pandemic in early 2020, the incidence of coronavirus infections among children in China has risen markedly. On the other hand, the volume of gynecological surgery patients has persistently declined due to the challenges associated with in-person medical care for epidemic control. By 2022, the prevalence of telemedicine in neonatology and prenatal diagnosis rose to 31% (44/142) and 12% (17/142), respectively, diverging from the declining trend of China's birth population from 2020 to 2022, primarily due to the Chinese government's enhancement of prenatal screening for expectant mothers and newborn disease assessments.

Figure 2. Distribution of telemedicine implementation departments in West China Second University Hospital of Sichuan University.

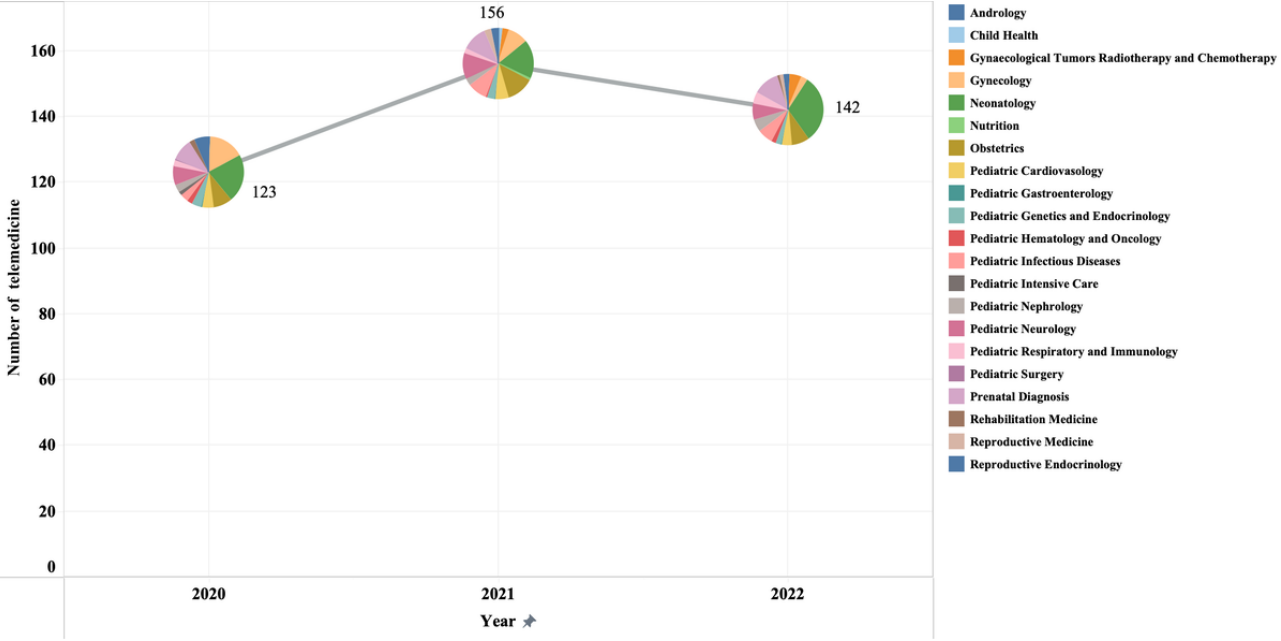

Reduction in GHG Emission and Energy Saving Through Telemedicine

Between 2020 and 2022, telemedicine services at WCSUH-SCU reduced GHG emissions by 36 tCO_2e, and their total life cycle emissions amounted to 8.06 tCO_2e, as shown in Figure 3. According to the life cycle management, the GHG emissions by telemedicine per patient amounted to 19.14 $kgCO_2e$. The software design and programming stage significantly contributed to telemedicine's GHG emissions, totaling 7.45 tCO_2e and representing 92.5% of the total. The amount of GHG emitted due to telemedicine use was 0.39 tCO_2e. Between 2020 and 2022, the energy-saving potential of telemedicine at WCSUH-SCU was 22.58 tce, while the total energy usage for telemedicine during this period was 3.82 tce. The energy consumption of software design and programming was measured at 3.03 tce, representing 79.4% of the energy consumed by telemedicine, thus remaining the largest contributor to energy consumption in this field. Furthermore, as energy consumption rose to 15.9%, electronic products, including computers and

XSL•FO
RenderX

video equipment, emerged as the second largest contributor, consuming 0.61 tce.

Figure 3. Energy consumption and greenhouse gas (GHG) emissions due to telemedicine services.

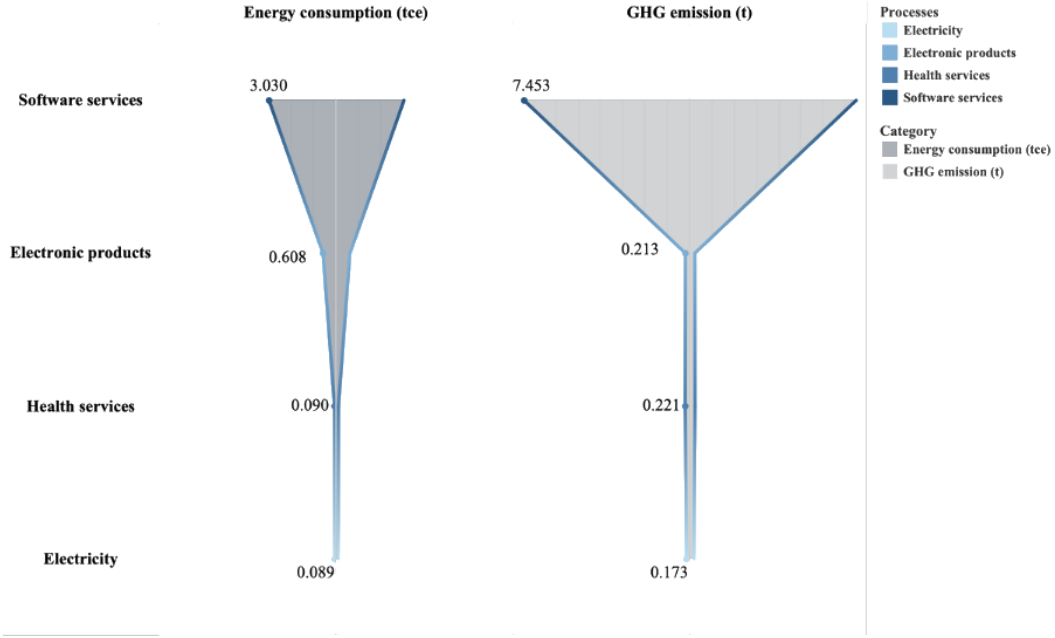

As illustrated in Figure 4, cumulative GHG emissions from outpatients between 2020 and 2022 totaled 44.06 tCO_2e, with per capita GHG emissions for outpatients reaching 104.66 $kgCO_2e$. Outpatient GHG emissions were higher than those of telemedicine services. Transportation-related GHG emissions significantly contributed to outpatient emissions, totaling 43.32 tCO_2e, which represented 98.3% of the overall emissions. Between 2020 and 2022, the total energy consumption due to outpatient services was 26.39 tce, 6.9 times greater than that of telemedicine. The energy consumption by vehicles remained the largest contributor at 26.15 tce, representing 99.1% of the overall energy consumption of outpatient services.

Figure 4. Energy consumption and greenhouse gas (GHG) emissions due to outpatient services.

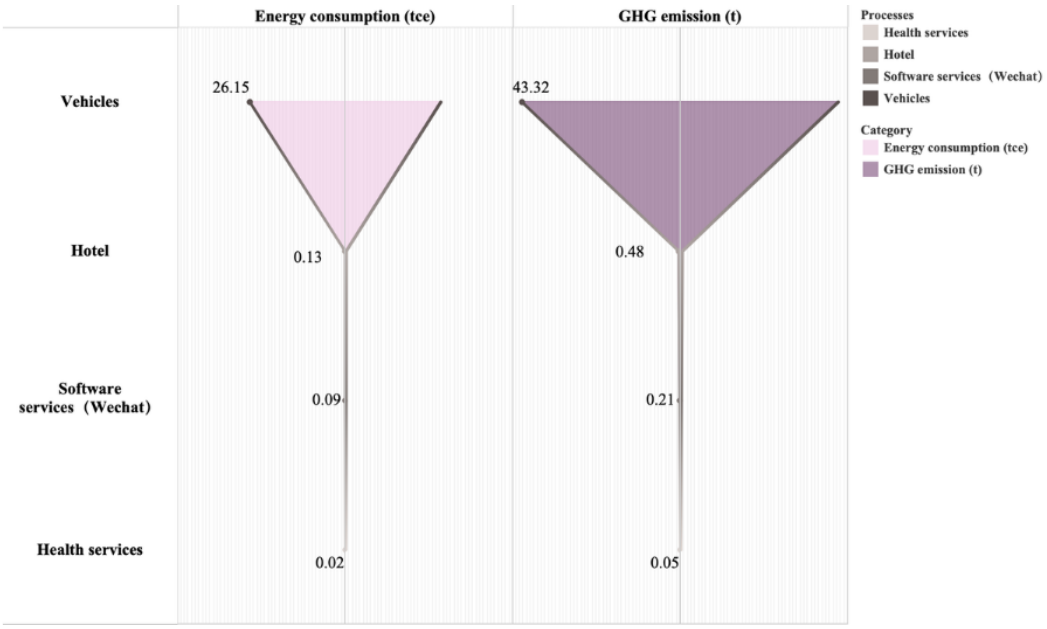

Thus, telemedicine demonstrates a clear advantage in reducing emissions and conserving energy. This indicates that the effective implementation of telemedicine has the potential to improve health sector sustainability in China by enhancing energy conservation and reducing GHG emissions.

Regarding the transportation methods employed in this study, the development of clean energy and the increase in the market share of new energy vehicles are expected to help reduce GHG emissions related to transportation. In turn, this reduction is anticipated to have a positive impact on outpatient emissions. Moreover, optimizing China's energy-producing system, including enhancing hydropower and wind power, will lead to a simultaneous decrease in electric GHG emissions. Consequently, telemedicine will arise as a more environmentally friendly option for health care in China. An estimation of future

XSL•FO
RenderX

GHG emissions from transportation and electricity will then be projected based on the present policy.

Scenario Analysis

Key Factors of GHG emissions

To better demonstrate telemedicine's low-carbon characteristics and provide data to support its application and promotion, this study forecasted GHG emissions for both telemedicine and outpatient care from 2023 to 2030 based on key emission factors.

Between 2020 and 2022, transportation accounted for an average of 98.5% of GHG emissions during the patient treatment stage, making it the primary contributor to outpatient GHG emissions. Given that electricity is essential for telemedicine, it is commonly acknowledged as the main driver of telemedicine's GHG emissions.

In contrast, between 2020 and 2022, electricity was the main source of telemedicine-related emissions, averaging 75.2%. These findings highlight the significant role of electricity in contributing to GHG emissions. The primary determinants of outpatient and telemedicine GHG emissions in this study align with the findings of prior research. The GHG emissions of essential components significantly impact outpatient and telemedicine. Variations in GHG emissions from outpatient and telemedicine could be reflected in the variations in GHG emissions from transportation and electricity.

This study employed scenario analysis to forecast GHG emissions during medical treatment for patients, including telemedicine use. GHG emissions from transportation unit mileage and electricity from 2023 to 2030 were assessed according to existing national carbon emission policies. The projected GHG emissions for transportation unit mileage and electric unit GHG emissions from 2023 to 2030 are shown in Table 2.

Table 2. Projected average GHG[a] emissions for transportation per unit mileage and electricity.

Year	Transportation unit mileage GHG emissions (g CO_2e/km)	Electric unit GHG emissions (kgCO_2e/kWh)
2022	247.770	0.570
2023	254.419	0.563
2024	248.739	0.556
2025	242.776	0.549
2026	235.205	0.533
2027	227.208	0.516
2028	218.786	0.500
2029	208.402	0.483
2030	195.843	0.467

[a]GHG: greenhouse gas.

Predictive Results of GHG Emissions

Using the forecast results and 2022 teleconsultation data from WCSUH-SCU, the average transportation and electricity-related GHG emissions from 2023 to 2030 were estimated. As shown in Figure 5, there will be a brief increase in the average transportation-related GHG emissions from 2020 to 2024. The average transportation-related GHG emissions in 2025 were estimated to decrease by 2% compared to the corresponding emissions in 2022, reaching 115.94 kgCO_2e per case. This downward trend was expected to continue, with emissions expected to drop to 93.528 kgCO_2e per case in 2030. It is evident that with the increasing adoption of new energy vehicles, there will be a significant reduction in average transportation-related GHG emissions after 2025 compared to the preceding period.

XSL•FO
RenderX

Figure 5. The predictive results of greenhouse gas (GHG) emissions for transportation and electricity.

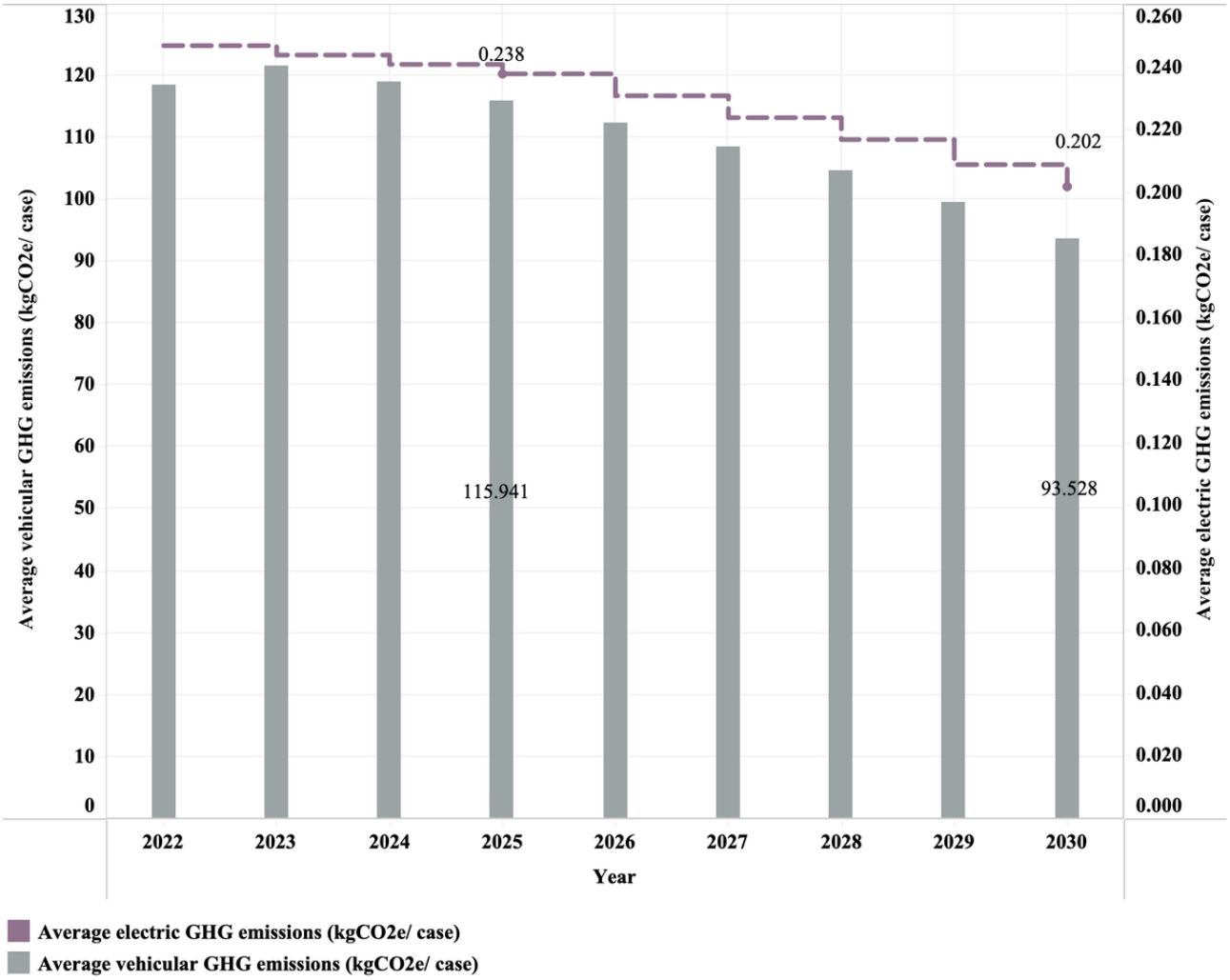

Regarding electricity-related GHG emissions, a 3.7% reduction was projected by 2025 compared to 2022, reaching 0.238 $kgCO_2e$ per case. Furthermore, the average electricity-related GHG emissions in 2030 will reduce by 15% compared to 2022. Implementing power generation structural reform has contributed to low-carbon development within the power generation industry.

Since GHG emissions from transportation account for a huge proportion of outpatient GHG emissions, GHG emissions from an average outpatient visit are expected to continue to decrease from 2025 to 2030. Depending on the effect of alterations in the average electricity-related GHG emissions from 2022 to 2030, it is anticipated that the GHG emissions of average telemedicine usage will also continue to decline from 2025 to 2030.

Discussion

Principal Results and Contributions

This study concludes that telemedicine may significantly decrease GHG emissions and energy consumption compared to in-person outpatient consultations, as determined by the LCA. Telemedicine usage leads to an average decrease of 85.51 $kgCO_2e$ emissions. In 2022, the total number of telemedicine consultations in China was 26.7 million, potentially reducing CO_2e emissions by a total of 2,283,117 metric tons. This corresponds to the yearly carbon emissions of 280,826 individuals in China [45]. According to the Mortality Cost of Carbon Model, the reduction in GHG emissions due to telemedicine could have saved 515 lives from climate change in 2022 [46]. The widespread implementation of telemedicine can markedly reduce GHG emissions and mortality rates. Telemedicine can serve as an excellent strategy for promoting environmental sustainability, reducing GHG emissions, and enhancing human health [47].

The health care sector safeguards both public health and environmental health. Expanding the accessibility and prevalence of telemedicine while minimizing carbon emissions from patient travel to hospitals is an efficient strategy to attain China's carbon neutrality objective in the health care sector [2]. Compared to the operational carbon emissions of medical institutions, telemedicine offers a clear advantage in reducing carbon emissions. This study, consistent with prior research, suggests that computers, audiovisual equipment, web servers, and electricity costs contribute to telemedicine-related GHG emissions [7]. As conventional manufacturing and technology companies globally implement low-carbon strategies to mitigate their energy consumption, there is a progressive decline in carbon emissions across the life cycle of technology-related

XSL•FO
RenderX

equipment, enhancing telemedicine's sustainability. Concurrently, sustainable energy serves as an additional impetus for the low carbonization in telemedicine. China's clean energy consumption is anticipated to rise from 25.9% in 2023 to 60% by 2030, making telemedicine a more environmentally sustainable technology [40]. As climate change intensifies, telemedicine's sustainable development of telemedicine is essential for the health care sector.

Transportation emissions remain the largest contributor to GHG emissions in outpatient care [48]. Telemedicine eliminates the need for patient travel, thereby reducing GHG emissions. China's major hospitals are predominantly concentrated in economically prosperous urban areas. However, mountainous regions in China account for 64.9% of the country's total land area, housing approximately 330 million permanent residents [49]. Patients living in mountainous regions emit large amounts of GHG when traveling to hospitals. Therefore, telemedicine can be a key solution for reducing travel-related GHG emissions.

This study confirms that in-person treatments generate more GHG than telemedicine. As 5G communication technologies and artificial intelligence continue to expand telemedicine's capabilities, more in-person treatments can be replaced. In 2022, the aggregate number of medical treatments in China was 8.42 billion, signifying a considerable potential for GHG emission reduction if telemedicine were to supplant in-person treatments.

To support the construction of a high-quality ecological environment, China has promulgated the Technology-Supported Carbon Peak Carbon Neutrality Implementation Plan (2022-2030), which proposes to achieve green development of hospital management based on green and low-carbon technological innovation. Telemedicine is innovative compared to traditional outpatient clinics. Furthermore, with the development of internet technology, more advantages and potential uses of telemedicine are coming to light. Large hospitals can make full use of telemedicine to help achieve hierarchical diagnosis and treatment. As China's medical industry shifts from a focus on medical insurance services to health services, telemedicine will expand beyond hospital diagnosis and treatment to include health monitoring and management of healthy individuals. As an innovative low-carbon technology, telemedicine will contribute significantly to the reduction of carbon emissions in China's health care industry.

Climate change is a global health crisis, and the health care sector—one of the fastest-growing sectors in terms of GDP—must take responsibility for reducing carbon emissions. Without effective intervention, global health care emissions could rise to 6 billion tons by 2050 [2]. The health care sector must promote decarbonization by developing sustainable solutions to reduce carbon emissions and improve human health. Telemedicine, as a win-win solution, should be taken seriously by the public. Despite the increased use of telemedicine during the COVID-19 epidemic, there is a lack of public awareness regarding its economic and environmental benefits. Government officials must enhance their comprehension of telemedicine's low emission advantages and implement effective procedures to facilitate its widespread adoption.

Limitations

This study uses an LCA to explore the potential of telemedicine to reduce GHG emissions and save energy. However, there are some limitations. First, transportation-related GHG emissions and energy consumption were estimated based on passenger cars. However, the diversity of transportation modes should be considered. Second, GHG emissions and energy consumption from other fixed assets, such as furniture, air conditioners in the consultation space, and network transmission during the consultation, were not included in this GHG emission reduction and energy-saving estimation. In future work, more factors affecting telemedicine carbon emissions should be included in the estimation. Third, in the carbon emission prediction section, we estimated future GHG emissions for outpatient and telemedicine services based on transportation and electricity usage without considering other factors. In future studies, more complex prediction models should be developed to predict telemedicine-related GHG emissions.

Conclusion

This study evaluates the environmental impact of telemedicine compared to traditional outpatient modes using a life cycle–based mathematical model. The findings confirm that telemedicine significantly reduces GHG emissions. From 2020 to 2022, 421 teleconsultations resulted in a total reduction of 36 tCO_2e in GHG emissions. Each telemedicine visit resulted in only 19.14 $kgCO_2e$ of GHG emissions, reducing emissions by 85.51 $kgCO_2e$ compared to in-person visits. Additionally, telemedicine use resulted in an average savings of 53.63 kg of standard coal. The primary contributors to life cycle GHG emissions in outpatient and telemedicine services were transportation and system programming design. Scenario analysis predicted that GHG emissions from telemedicine and outpatient services will continue to decline from 2025 to 2030, contributing to an overall decline in health care sector GHG emissions.

In 2022, China recorded 26.7 million telemedicine consultations, resulting in a potential reduction of 2,283,117 tons of CO_2e emissions. This reduction is equivalent to the annual carbon emissions of 280,826 individuals in China and could save 515 lives from climate change. Telemedicine serves as a dual-benefit measure by enhancing the efficiency of medical services and promoting the sustainable development of the health care sector. As the proportion of clean energy increases and ICT advances, telemedicine can be applied across more medical fields, thereby contributing more significantly to the health care sector's carbon neutrality efforts. Moreover, as the share of health services expands, telemedicine will extend beyond hospital diagnosis and treatment to encompass health monitoring and management for healthy individuals. Government decision-makers should deepen their comprehension of telemedicine's low-emission advantages and formulate effective legislation to facilitate its broad implementation.

In conclusion, this study quantifies telemedicine's role in reducing carbon emissions in China's health sector, emphasizing its significance in achieving carbon neutrality. It not only presents factual evidence to substantiate the sustainable

XSL•FO
RenderX

advancement of China's health sector but also offers a theoretical foundation for the future eco-friendly development of China's medical industry. While conducted within a Chinese context, the conclusions can extend to other nations pursuing similar sustainability goals.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

All authors contributed to the study's conception and design. ZL was responsible for the investigation and wrote the original draft. LX handled the methodology. JN conducted the formal analysis. WL curated the data. YH contributed to the conceptualization. XX reviewed the writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The actual data of telemedicine at West China Second University Hospital of Sichuan University in 2020.
[DOCX File , 16 KB - jmir_v27i1e63927_app1.docx]

References

1. The Paris Agreement. United Nations Climate Change. 2016. URL: https://unfccc.int/process-and-meetings/the-paris-agreement [accessed 2016-11-04]
2. Health care's climate footprint: How the health sector contributes to the global climate crisis and opportunities for action. Health Care Without Harm. 2019. URL: https://global.noharm.org/sites/default/files/documents-files/5961/HealthCaresClimateFootprint_092319.pdf [accessed 2019-10-03]
3. Technology supports carbon peak and carbon neutrality implementation plan (2022-2030). Government of China. 2022 Jun 24. URL: https://www.gov.cn/zhengce/zhengceku/2022-08/18/content_5705865.htm [accessed 2022-06-24]
4. National. China health care statistics tearbook. National Health Commission. 2022. URL: http://www.nhc.gov.cn/mohwsbwstjxxzx/tjtjnj/202305/6ef68aac6bd14c1eb9375e01a0faa1fb.shtml [accessed 2023-05-17]
5. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
6. Andrew N, Barraclough KA, Long K, Fazio TN, Holt S, Kanhutu K, et al. Telehealth model of care for routine follow up of renal transplant recipients in a tertiary centre: a case study. J Telemed Telecare 2018 Nov 17;26(4):232-238. [doi: 10.1177/1357633x18807834]
7. Donald N, Irukulla S. Greenhouse gas emission savings in relation to telemedicine and associated patient benefits: a systematic review. Telemed J E Health 2022 Apr 20;28(11):1555-1563. [doi: 10.1089/tmj.2022.0047] [Medline: 35446668]
8. Vidal-Alaball J, Franch-Parella J, Lopez Seguí F, Garcia Cuyàs F, Mendioroz Peña J. Impact of a telemedicine program on the reduction in the emission of atmospheric pollutants and journeys by road. Int J Environ Res Public Health 2019 Nov 08;16(22):4366 [FREE Full text] [doi: 10.3390/ijerph16224366] [Medline: 31717386]
9. Sellars H, Ramsay G, Sunny A, Gunner CK, Oliphant R, Watson AJM. Video consultation for new colorectal patients. Colorectal Dis 2020 Sep 31;22(9):1015-1021. [doi: 10.1111/codi.15239] [Medline: 32628286]
10. Shi J, Yan X, Wang M, Lei P, Yu G. Factors influencing the acceptance of pediatric telemedicine services in China: a cross-sectional study. Front Pediatr 2021 Oct 18;9:745687 [FREE Full text] [doi: 10.3389/fped.2021.745687] [Medline: 34733810]
11. Mundi MS, Mohamed Elfadil O, Bonnes SL, Salonen BR, Hurt RT. Use of telehealth in home nutrition support: Challenges and advantages. Nutr Clin Pract 2021 Aug 10;36(4):775-784. [doi: 10.1002/ncp.10736] [Medline: 34245465]
12. Zhang C, Zhu K, Lin Z, Huang P, Pan Y, Sun B, et al. Utility of deep brain stimulation telemedicine for patients with movement disorders during the COVID-19 outbreak in China. Neuromodulation 2021 Feb;24(2):337-342 [FREE Full text] [doi: 10.1111/ner.13274] [Medline: 33006811]
13. Colbert GB, Venegas-Vera AV, Lerma EV. Utility of telemedicine in the COVID-19 era. Rev Cardiovasc Med 2020 Dec 30;21(4):583-587 [FREE Full text] [doi: 10.31083/j.rcm.2020.04.188] [Medline: 33388003]
14. Bozoudis V, Sebos I, Tsakanikas A. Action plan for the mitigation of greenhouse gas emissions in the hospital-based health care of the Hellenic Army. Environ Monit Assess 2022 Feb 24;194(3):221. [doi: 10.1007/s10661-022-09871-3] [Medline: 35211813]

XSL•FO
RenderX

15. Holmner A, Ebi KL, Lazuardi L, Nilsson M. Carbon footprint of telemedicine solutions--unexplored opportunity for reducing carbon emissions in the health sector. PLoS One 2014 Sep 4;9(9):e105040 [FREE Full text] [doi: 10.1371/journal.pone.0105040] [Medline: 25188322]
16. Papavero SC, Fracasso A, Ramaglia P, Cicchetti A, de Belvis AG, Ferrara FM. Telemedicine has a social impact: an Italian national study for the evaluation of the cost-opportunity for patients and caregivers and the measurement of carbon emission savings. Telemed J E Health 2023 Aug 01;29(8):1252-1260. [doi: 10.1089/tmj.2022.0333] [Medline: 36637858]
17. Thiel CL, Mehta N, Sejo CS, Qureshi L, Moyer M, Valentino V, et al. Telemedicine and the environment: life cycle environmental emissions from in-person and virtual clinic visits. NPJ Digit Med 2023 May 09;6(1):87. [doi: 10.1038/s41746-023-00818-7] [Medline: 37160996]
18. Zhang W, He D, Wang G, Zhu C, Evans R. Analyzing national telemedicine policies in China from the perspective of policy instrument (1997-2020). Int J Med Inform 2022 Oct;166:104854. [doi: 10.1016/j.ijmedinf.2022.104854] [Medline: 35981479]
19. Xiao Y, Chen TT, Zhou YL, Zhu SY. Challenges in establishing a strong telemedicine system in China. Postgrad Med J 2023 Mar 22;99(1167):1-3. [doi: 10.1093/postmj/qgac009] [Medline: 36947420]
20. Wang Z, Xu R, Liu Y, Li Y. Licensing policy and platform models of telemedicine: a multi-case study from China. Front Public Health 2023;11:1108621 [FREE Full text] [doi: 10.3389/fpubh.2023.1108621] [Medline: 36817879]
21. Wang W, Li D, Zhou S, Wang Y, Yu L. Exploring the key influencing factors of low-carbon innovation from urban characteristics in China using interpretable machine learning. Environ Impact Assess Rev 2024 Jul;107:107573. [doi: 10.1016/j.eiar.2024.107573]
22. Shen Y, Chen L, Yue W, Xu H. Digital technology-based telemedicine for the COVID-19 pandemic. Front Med (Lausanne) 2021 Jul 6;8:646506 [FREE Full text] [doi: 10.3389/fmed.2021.646506] [Medline: 34295908]
23. Wang W, Li D, Huang G, Zhou S, Zhao Y, Zhou H, et al. Barriers to developing a citizen-centric smart city in China: from the perspective of citizens' sense of gain. J Urban Technol 2024 Oct 24;31(4-5):69-104. [doi: 10.1080/10630732.2024.2402207]
24. Wang W, Li D, Zhou S, Han Z. Towards sustainable development: assessing the effects of low-carbon city pilot policy on residents' welfare. Environ Dev Sustain 2024 Sep 13. [doi: 10.1007/s10668-024-05373-z]
25. Chau C, Leung T, Ng W. A review on Life Cycle Assessment, Life Cycle Energy Assessment and Life Cycle Carbon Emissions Assessment on buildings. Applied Energy 2015 Apr;143:395-413. [doi: 10.1016/j.apenergy.2015.01.023]
26. Jiang Z, Pan Z, Xing J, Yun F. Greenhouse gas emissions from nuclear power chain life cycle in China. China Environ Sc 2015;35(11):3502-3510.
27. Li Y, He Q, Luo X, Zhang Y, Dong L. Calculation of life-cycle greenhouse gas emissions of urban rail transit systems: A case study of Shanghai Metro. Resour Conserv Recycl 2018 Jan;128:451-457. [doi: 10.1016/j.resconrec.2016.03.007]
28. Liu H, Huang Y, Yuan H, Yin X, Wu C. Life cycle assessment of biofuels in China: Status and challenges. Renew Sustain Energy Rev 2018 Dec;97:301-322. [doi: 10.1016/j.rser.2018.08.052]
29. Wang F, Xu B, Si Y, Shang Y, Zhang W, Cai B, et al. Components and drivers of household water footprint inequality in China. Sustain Prod Consum 2023 Dec;43:1-14. [doi: 10.1016/j.spc.2023.10.009]
30. Lu Y, Zhou Y, Zeng S. Understanding the link between industrial stimulation effect and environmental impact of sponge city: with a focus on the construction phase. J Clean Prod 2023 Feb;389:136113. [doi: 10.1016/j.jclepro.2023.136113]
31. Wang Y, Wang X, Wang H, Zhang X, Zhong Q, Yue Q, et al. Human health and ecosystem impacts of China's resource extraction. Sci Total Environ 2022 Nov 15;847:157465. [doi: 10.1016/j.scitotenv.2022.157465] [Medline: 35868370]
32. Thiel C, Schehlein E, Ravilla T, Ravindran R, Robin A, Saeedi O, et al. Cataract surgery and environmental sustainability: Waste and lifecycle assessment of phacoemulsification at a private healthcare facility. J Cataract Refract Surg 2017 Nov;43(11):1391-1398 [FREE Full text] [doi: 10.1016/j.jcrs.2017.08.017] [Medline: 29223227]
33. Wang EY, Zafar JE, Lawrence CM, Gavin LF, Mishra S, Boateng A, et al. Environmental emissions reduction of a preoperative evaluation center utilizing telehealth screening and standardized preoperative testing guidelines. Resour Conserv Recycl 2021 Aug;171:105652. [doi: 10.1016/j.resconrec.2021.105652]
34. Tian X, Liu Y, Xu M, Liang S, Liu Y. Chinese environmentally extended input-output database for 2017 and 2018. Sci Data 2021 Sep 30;8(1):256 [FREE Full text] [doi: 10.1038/s41597-021-01035-1] [Medline: 34593825]
35. Group WALCAW. Carbon Neutrality: Low Carbon Development Strategies and Transformation Pathways of the Automotive Industry. Beijing, China: China Machine Press; 2022:1-342.
36. China automobile low carbon action plan research report. Group WALCAW. 2021. URL: http://www.auto-cices.com:81/a/chengguofabunarong/ziliaoxiazai/zhongguoqichedit/2021/0719/398.html [accessed 2021-07-30]
37. China automobile low carbon action plan research report. Group WALCAW. 2020. URL: http://www.auto-cices.com:81/a/chengguofabunarong/ziliaoxiazai/zhongguoqichedit/2020/0731/367.html [accessed 2020-12-07]
38. Notice on the key tasks related to the management of corporate greenhouse gas emission reports in 2022. China MoEaEotPsRo. 2022. URL: https://www.mee.gov.cn/xxgk2018/xxgk/xxgk06/202203/t20220315_971468.html [accessed 2022-03-15]

XSL•FO
RenderX

39. Notice on the management of greenhouse gas emission reports for enterprises in the power generation industry from 2023 to 2025. China MoEaEotPsRo. 2023. URL: https://www.mee.gov.cn/xxgk2018/xxgk/xxgk06/202302/t20230207_1015569.html [accessed 2023-02-07]
40. China electric power industry annual development report in 2023. Institute CEMSNSR. 2023. URL: https://www.cec.org.cn/detail/index.html?3-322625 [accessed 2023-07-07]
41. Star-rated hotel: average room rate: Chengdu. Qianzhan Industry Research Institute. 2023. URL: https://d.qianzhan.com/xdata/details/0336535acbf762a8.html [accessed 2024-03-14]
42. The second batch of new "Internet +" medical service price items in Sichuan Province are trial prices and medical insurance payment categories. Sichuan Provincial Health Care Security Administration. 2021. URL: https://www.sc.gov.cn/10462/11555/11564/2022/1/26/c976e6ad8934445caa18bf5ccbdefdc2.shtml [accessed 2021-11-28]
43. Adjusted price list of medical service items of Sichuan provincial public medical institutions. Sichuan Provincial Health Care Security Administration. 2023. URL: https://abzxjx.sczwfw.gov.cn/art/2023/9/19/art_33562_232552.html?areaCode=513227000000 [accessed 2023-09-19]
44. Berger R. Electrification reshapes China's automotive aftermarket. Roland Berger. 2022. URL: https://www.rolandberger.com/zh/Insights/Publications/ [accessed 2022-07-01]
45. CO2 emissions in 2022. International Energy Agency. 2023. URL: https://www.iea.org/reports/co2-emissions-in-2022 [accessed 2023-03-02]
46. Bressler RD. The mortality cost of carbon. Nat Commun 2021 Jul 29;12(1):4467 [FREE Full text] [doi: 10.1038/s41467-021-24487-w] [Medline: 34326326]
47. Campbell-Lendrum D, Neville T, Schweizer C, Neira M. Climate change and health: three grand challenges. Nat Med 2023 Jul;29(7):1631-1638. [doi: 10.1038/s41591-023-02438-w] [Medline: 37464036]
48. Ravindrane R, Patel J. The environmental impacts of telemedicine in place of face-to-face patient care: a systematic review. Future Healthc J 2022 Mar;9(1):28-33 [FREE Full text] [doi: 10.7861/fhj.2021-0148] [Medline: 35372776]
49. Anbang W, Qing T, Chaojun O, Bo Z, Yukuan W, Ainong L, et al. Mountain protection and mountain development in China: review and prospect. Bull Chin Acad Sci 2023;38(3):376-384. [doi: 10.16418/j.issn.1000-3045.20230121001]

Abbreviations

CEEIO: Chinese Environmentally Extended Input-Output
GDP: gross domestic product
GHG: greenhouse gas
LCA: life cycle assessment
WCSUH-SCU: West China Second University Hospital of Sichuan University

Edited by Y Hong; submitted 03.07.24; peer-reviewed by D Dinh, W Wang, M Zhang; comments to author 11.12.24; revised version received 20.01.25; accepted 06.02.25; published 24.02.25.

Please cite as:
Li Z, Xiang L, Ning J, Li W, Huang Y, Xiao X
Pathways to Sustainable Health Care Development: Study on the Carbon Reduction Potential of Telemedicine in China
J Med Internet Res 2025;27:e63927
URL: https://www.jmir.org/2025/1/e63927
doi:10.2196/63927
PMID:

©Zhi Li, Lianrui Xiang, Jing Ning, Wenbo Li, Yong Huang, Xue Xiao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Application of Large Language Models in Medical Training Evaluation—Using ChatGPT as a Standardized Patient: Multimetric Assessment

Chenxu Wang[1,2,3], BSc; Shuhan Li[1,2], BSc; Nuoxi Lin[1,2], BSc; Xinyu Zhang[3], BSc; Ying Han[3], NP, MSc; Xiandi Wang[3], MD; Di Liu[1,2,4], PhD; Xiaomei Tan[2], PhD; Dan Pu[3], MD; Kang Li[1,2,4*], PhD; Guangwu Qian[1,5*], PhD; Rong Yin[1,2,4*], PhD

[1]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China

[2]Department of Industrial Engineering, Pittsburgh Institute, Sichuan University, Chengdu, China

[3]Department of Medical Simulation Center, West China Hospital, Sichuan University, Chengdu, China

[4]Med-X Center for Informatics, Sichuan University, Chengdu, China

[5]Department of Computer Science, Pittsburgh Institute, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Kang Li, PhD
West China Biomedical Big Data Center
West China Hospital
Sichuan University
No.37, Guoxue Lane, Wuhou District
Chengdu, 610041
China
Phone: 86 02881739902
Email: likang@wchscu.cn

Abstract

Background: With the increasing interest in the application of large language models (LLMs) in the medical field, the feasibility of its potential use as a standardized patient in medical assessment is rarely evaluated. Specifically, we delved into the potential of using ChatGPT, a representative LLM, in transforming medical education by serving as a cost-effective alternative to standardized patients, specifically for history-taking tasks.

Objective: The study aims to explore ChatGPT's viability and performance as a standardized patient, using prompt engineering to refine its accuracy and use in medical assessments.

Methods: A 2-phase experiment was conducted. The first phase assessed feasibility by simulating conversations about inflammatory bowel disease (IBD) across 3 quality groups (good, medium, and bad). Responses were categorized based on their relevance and accuracy. Each group consisted of 30 runs, with responses scored to determine whether they were related to the inquiries. For the second phase, we evaluated ChatGPT's performance against specific criteria, focusing on its anthropomorphism, clinical accuracy, and adaptability. Adjustments were made to prompts based on ChatGPT's response shortcomings, with a comparative analysis of ChatGPT's performance between original and revised prompts. A total of 300 runs were conducted and compared against standard reference scores. Finally, the generalizability of the revised prompt was tested using other scripts for another 60 runs, together with the exploration of the impact of the used language on the performance of the chatbot.

Results: The feasibility test confirmed ChatGPT's ability to simulate a standardized patient effectively, differentiating among poor, medium, and good medical inquiries with varying degrees of accuracy. Score differences between the poor (74.7, SD 5.44) and medium (82.67, SD 5.30) inquiry groups (P<.001), between the poor and good (85, SD 3.27) inquiry groups (P<.001) were significant at a significance level (α) of .05, while the score differences between the medium and good inquiry groups were not statistically significant (P=.16). The revised prompt significantly improved ChatGPT's realism, clinical accuracy, and adaptability, leading to a marked reduction in scoring discrepancies. The score accuracy of ChatGPT improved 4.926 times compared to unrevised prompts. The score difference percentage drops from 29.83% to 6.06%, with a drop in SD from 0.55 to 0.068. The

XSL•FO
RenderX

performance of the chatbot on a separate script is acceptable with an average score difference percentage of 3.21%. Moreover, the performance differences between test groups using various language combinations were found to be insignificant.

Conclusions: ChatGPT, as a representative LLM, is a viable tool for simulating standardized patients in medical assessments, with the potential to enhance medical training. By incorporating proper prompts, ChatGPT's scoring accuracy and response realism significantly improved, approaching the feasibility of actual clinical use. Also, the influence of the adopted language is nonsignificant on the outcome of the chatbot.

(J Med Internet Res 2025;27:e59435) doi:10.2196/59435

KEYWORDS

ChatGPT; artificial intelligence; standardized patient; health care; prompt engineering; accuracy; large language models; performance evaluation; medical training; inflammatory bowel disease

Introduction

Background

In recent decades, large language models (LLMs) have experienced significant advancements [1,2]. LLMs are artificial intelligence (AI) models designed to comprehend and process natural human language, as well as generate it [1,3]. Concurrent with the progression of AI, LLMs have exhibited substantial potential in executing tasks involving natural language processing [4]. LLM applications range from article synthesis to summarization to patient diagnosis, illustrating their flexibility in providing valuable assistance [5-9]. Among the various types of LLMs such as the generalist language model, Flamingo, and Minerva [10], ChatGPT stands out as one of the critical milestone models, which is the primary focus of this study.

ChatGPT, developed by OpenAI, a US-based company, represents a significant advancement in language models [11,12]. This model facilitates user interaction with follow-up questions and is fine-tuned for controlled output [13,14]. Additionally, it can be integrated into custom applications via an application programming interface, allowing developers to craft chatbots and virtual assistants with tailored behavior [13]. Built on GPT architecture, ChatGPT currently comprises 2 primary variants, GPT-3.5 and GPT-4. A key characteristic of ChatGPT is that its output is significantly influenced by the input prompt and can be fine-tuned by modifying this prompt. Consequently, prompt engineering is essential for optimizing ChatGPT's performance [15]. Despite ChatGPT and other LLMs, as well as the broader field of AI, being in developmental stages, they have demonstrated great potential in assisting human activities and may substitute human in certain tasks due to their low cost and high efficiency [16]. In the realm of medical education, traditional tests using standardized patients are no exception, involving substantial human labor and training costs [17-20].

The concept of a standardized patient, initially introduced by Barrows and Abrahamson [21], has become a widely recognized method in medical training and assessment [22,23]. In this context, "standardized" implies that the patient in a standardized patient scenario is trained to consistently portray a specific set of symptoms, medicine allergies, and medical history. Here, "patient" refers to an individual acting as if seeking medical care, unlike real patients, these are trained doctors or individuals simulating the diagnosis process to access medical professionals [23,24]. Consequently, a standardized patient is an individual trained to simulate a real patient, accurately portraying a set of symptoms or conditions. Medical students or doctors interact with standardized patients to practice and evaluate their clinical skills [22]. A critical aspect of this procedure is conversing with the standardized patient to obtain essential information for diagnosis, which is known as history-taking tasks. As the entire training process for an SP is both time-consuming and expensive, rural hospitals may even fail to provide such medical training due to poor access to high-quality health care services [25-27]. Our research aims to fill this research gap by applying LLMs, such as ChatGPT, in the assessment process involving standardized patients to tackle these problems. A critical factor influencing ChatGPT's response is the prompt input given to the model [28,29]. Nevertheless, ChatGPT's accuracy and precision require enhancement for medical applications [30,31]. Therefore, this study explored a novel approach to standardized patients in medical training evaluation by substituting human standardized patients with ChatGPT, as well as maintaining relatively high accuracy and precision. The study is conducted along 2 distinct dimensions. Initially, the capability and feasibility of ChatGPT functioning as a standardized patient are investigated. Subsequently, this study explores prompt engineering and revisions to enhance ChatGPT's performance. Finally, this study examines the differences in scores and wording between original and revised prompt (RP) results generated by ChatGPT.

Objectives

This study applied an exploratory approach to standardized patients in medical training evaluation by substituting human standardized patients with ChatGPT, as well as maintaining relatively reasonable accuracy and precision. The main objective can be divided into 2 stages of experiment. Initially, the capability and feasibility of ChatGPT functioning as a standardized patient are investigated. Subsequently, exploring prompt engineering and revisions to enhance ChatGPT's performance, and also examining the differences in scores and wording between the original prompt (OP) and RP results generated by ChatGPT.

Methods

Standardized Patient Resource

The West China Medical Simulation Center of West China Hospital of Sichuan University has released a comprehensive

XSL•FO
RenderX

standardized patient training script, encompassing a scenario-based conversational analysis and a specific performance criterion for participants, detailed in Table 1. This criterion emphasizes a procedure-oriented approach, where interactions with the standardized patient are critically evaluated and graded. The criterion comprises 2 principal aspects for evaluation: inquiry skills and humanistic care. The inquiry skills evaluation encompasses 4 dimensions, conversation arrangement, type of question, verifications, and use of professional jargon. In the domain of humanistic care, 2 dimensions are considered: speech and amiable behavior. Each dimension has 5 tiers, from tier 1 to tier 5, with ascending levels of performance. Comprehensive judgment criteria are illustrated in Table 1. ChatGPT exhibits limitations concerning criteria based on subjective judgment, as it lacks emotional response and is not fully adaptable to subjective grading. Therefore, this study focuses solely on the objective grading criteria. For example, an evaluation point could be whether the user inquired about the medicine history or not. Such criteria can be objectively assessed based on the conversation.

Table 1. Standardized patient consultation skills grading and scoring criteria. This scoring sheet can be used in standardized patient assessments to quantitatively evaluate the performance of candidates, where 5= best performance.

Skills	Ranking tiers				
	5-tier	4-tier	3-tier	2-tier	1-tier
Inquiry skills					
Conversation arrangement	The beginning, middle, and end of the consultation are clear and precise, with questions asked in an orderly manner.	Between 5-point and 3-point	Most of the consultation is conducted in an orderly fashion, but the beginning and ending are not clearly defined.	Between 3-point and 1-point	The consultation lacks coherence and organization.
Question types	Reasonable use of open-ended or closed-ended questions.	Between 5-point and 3-point	No open-ended questions, directly asking with closed-ended questions.	Between 3-point and 1-point	Frequently uses sequential and leading questions.
Verifications	Conduct a comprehensive and thorough verification and reference.	Between 5-point and 3-point	The verification and reference are incomplete and not sufficient.	Between 3-point and 1-point	Did not conduct verification and reference.
Professional jargon	The explanation is clear and easy to understand, not using complicated medical terminology.	Between 5-point and 3-point	The explanation is understandable, with minimal use of complex medical terminology.	Between 3-point and 1-point	Frequently uses complicate medical terminology.
Humanistic care					
Speech	Appropriate speech speed and tone.	Between 5-point and 3-point	The speech speed and tone are mildly uncomfortable.	Between 3-point and 1-point	The speech speed and tone are noticeably uncomfortable.
Amiable behavior	Appropriate response and comfort.	Between 5-point and 3-point	Provides responses and comfort.	Between 3-point and 1-point	No response or comfort.

Study Design

To date, ChatGPT has introduced 2 versions of its language model: GPT-3.5 and GPT-4. OpenAI reports that GPT-4 exhibits improved word processing and memory retention capabilities. These advancements make GPT-4 particularly suitable for the aims of our research [32]. The preliminary step involves ascertaining the feasibility of ChatGPT functioning as a standardized patient in medical training evaluations. To gain preliminary insights, we conduct a preliminary, exploratory conversation with ChatGPT-4. We chose inflammatory bowel disease (IBD), a relatively rare condition with multiple symptoms that can be easily confused with other diseases [33,34]. In this experiment, the chatbot was given only the name of the disease, without detailed criteria, and was asked to perform as a standardized patient to assess users' performance as if taking a standardized patient examination.

We asked a series of questions regarding IBD's symptoms to determine if ChatGPT could respond appropriately like a real standardized patient. Subsequently, ChatGPT was requested to provide a score for our inquiry process, along with its criteria for assessment. The chatbot was evaluated using 3 different approaches: using poor, medium, and good inquiries. The criteria were as follows: in the poor inquiry approach, no questions pertinent to IBD were posed, and the language was unprofessional and overly casual. The medium inquiry approach exhibited a combination of relevant and irrelevant questions, with varying degrees of accuracy. Conversely, the good inquiry approach demonstrated both linguistic precision and a consistently accurate focus on relevant topics. Each level

XSL•FO
RenderX

underwent 3 consecutive tests, and in each test, the chatbot generated the answer 10 times, from which an average score out of 30 (5 scores each × 6 groups) was calculated for each level.

For further exploration, a second phase of the experiment was designed. In this subsequent experiment, we used a detailed script accompanied by clear criteria provided by our clinical skill training center. An example from the script, featuring a perfect, full-score conversation between a standardized patient and a medical student, was used to test the chatbot. By modifying the dialog, including additions and deletions, the final score could vary. To initiate the conversation, the chatbot was provided with a patient's medical record, along with a prompt outlined in Table 2. During the subsequent conversation, the questions on the script list were asked respectively to assess if the chatbot can extract information from the medical record precisely. Each answer was checked and compared to identify the flaws in responses, which were considered as key points for future improvement. At last, a score based on the given criteria was assigned by ChatGPT, along with a detailed analysis of the score. Repeated experiments (OP-1 to OP-5) were conducted 5 times, for each experiment, ChatGPT was asked to generate the score 30 times to obtain an average error rate for this overall preliminary experiment. From OP-1 to OP-5, the questions asked differed; therefore, the total standard score also varied. The questions asked and the total score were different, but the criterion was the same. From OP-1 to OP-5, the total score was in increasing order: 20/100, 60/100, 75/100, 85/100, and 90/100, respectively. A score of 20/100 is classified as "poor inquiry," while scores of 60/100 and 75/100 are classified as "medium inquiry," and 85/100 and 90/100 are classified as "good inquiry." Given that the total score varied across each trial, we used a metric known as the score difference percentage (SDP) to quantify ChatGPT's performance. This metric represents the error rate in terms of the percentage difference between expected and actual scores. The formula used is as follows.



After the preliminary experiment was done, a full inspection of the entire conversation was given. ChatGPT's performance on the following key points is focused on: the degree of anthropomorphism, clinical accuracy, and adaptability. Considering the degree of anthropomorphism, ensuring ChatGPT mimics a real patient accurately is crucial for its validity as a standardized patient, the chatbot should emulate the tone, behavior, and emotional responses typical of a real patient. In an authentic standardized patient assessment, to simulate a real-life patient, standardized patients may pose questions driven by anxiety, for instance, "Doctor, why does my stomach ailment keep recurring? Will it eventually lead to cancer?" [35]. It is acknowledged that ChatGPT may not initiate such inquiries due to its inherent limitations; however, we will investigate whether integrating a predesigned prompt enables the chatbot to generate these types of questions [36]. Moreover, the chatbot needs to communicate the symptoms effectively, maintaining a balance between being too vague and excessively detailed. It should provide information that is similar to what a real patient may disclose. For clinical accuracy, it is essential that the chatbot delivers precise information aligned with the provided medical record, which is a critical aspect of using ChatGPT in standardized patient assessments. The response from the chatbot must align with the corresponding description in the medical record. ChatGPT is expected to recall previous responses and consistently adhere to the documented medical history. For adaptability, in actual standardized patient tests, medical students may request information from the standardized patient that is not provided in the script or medical history; therefore, ChatGPT must be capable of inferring unspecified information. Following each evaluation, ChatGPT's performance would be evaluated based on the earlier criteria, subsequently generating a list of issues. Based on the identified issues, the prompt was modified. Subsequently, a new series of experiments began, using RP. The final version of the prompt is determined when ChatGPT demonstrates satisfactory performance in terms of anthropomorphism, clinical accuracy, consistency in responses, and adaptability, among other factors. Additionally, the scores provided were closely inspected. We designed a comparative experiment involving the OP group. RP-1 to RP-5 were crafted correlating with OP-1 to OP-5. For each paired experiment, the questions posed and the total score were identical. The only difference between the OP and RP groups was the variation in the prompts provided. For each group, 30 runs were conducted to minimize the potential bias. Ultimately, the error rates will be compared with results from OPs to assess the efficacy of RP. A sample dialogue with ChatGPT is included in Multimedia Appendix 1.

Finally, the performance of the RPs was evaluated using a different standardized patient script to verify generalizability. In preliminary experiments, the dialogues with the chatbot were conducted in Chinese to minimize potential errors introduced by translation and the prompt provided was in English. However, according to a previous study on Chinese National Medical Licensing Examination questions, the translation of Chinese questions into English resulted in only a minor improvement (P=.16) [37]. To validate this research, the test group would be divided into 4 subgroups, representing all possible combinations of 2 languages used in prompts and dialogues, which were prompts in English (PE) + dialogue in English (DE); prompts in Chinese (PC) + dialogue in English (DE); prompts in English (PE) + dialogue in Chinese (DC); and prompts in Chinese + dialogue in Chinese (DC). All the tests in each group would be conducted 15 times to minimize the deviation caused by the inconsistency of the chatbot responses.

XSL•FO
RenderX

Table 2. Initial prompt given to ChatGPT[a].

Prompt	Consideration
I'm a first-year medical student. Please help me with an interactive test case. Are you familiar with the procedure of a standardized patient assessment? Now, I want you to play the role of a standardized patient, while I play the role of the student. Follow these steps.	Initialization
Your basic situation: [The detailed medical history is provided to ChatGPT based the script]	Set up specific scenario
You need to answer the questions I'm asking you based on the facts.	Restrict the free response of ChatGPT
At the end of my consultation, you need to rate my consultation process, based on following criteria: [The specific scoring rules are provided]	Grade with detailed criteria

[a]This is the initial, unrevised prompt used for interactions with ChatGPT in the phase 2 study design, prior to prompt engineering.

Data Analysis

The analysis software used in this research is R (version 4.3.3; R Foundation for Statistical Computing). For the first phase of the experiment, a 2-tailed *t* test is performed on the results to ascertain whether ChatGPT can differentiate between varying levels of problem relevance. In the second phase, the Shapiro-Wilk normality test was conducted to examine whether the data collected were normally distributed. For data not normally distributed, the Mann-Whitney *U* test was used to evaluate the significance of the differences in ChatGPT scores and standard scores, as well as between the OP and RP groups, thereby ascertaining the enhancements postprompt engineering. The data were collected during November 2023, December 2023, and August 2024.

Results

Feasibility Test

In the feasibility test, upon receiving the keyword “IBD,” ChatGPT automatically generated a virtual medical record, in subsequent interactions, it adhered to the information within the record, simulating a real patient’s behavior. For each inquiry quality group, 30 runs were conducted. Subsequently, ChatGPT provided an average score of 74.7/100 (SD 5.44) for a poor inquiry, an average score of 82.7 (SD 5.30) for a medium inquiry, and an average score of 85.0 (SD 3.27) for a good inquiry. These statistical data are presented in Table 3.

To demonstrate ChatGPT’s ability to distinguish the results, 2 inquiry model groups were established. The score differences between the poor and medium inquiry groups, between the medium and good inquiry groups, and between the poor and good groups were respectively examined. Both the score differences between the poor and medium inquiry groups ($P<.001$), between the poor and good groups ($P<.001$) were significant at a significance level of $\alpha=.05$, while the score differences between the medium and good inquiry groups were not statistically significant ($P=.16$).

The density diagram of the scores is illustrated in Figure 1. Along with each score, the chatbot provided a detailed rating scale. Across various trials, ChatGPT’s criteria varied slightly, yet the core remained consistent: it consisted of 5 components: relativeness, inquiry into medical history, clinical reasoning, professionalism, and problem-solving.

Table 3. Scores given by ChatGPT in feasibility test[a].

Inquiry quality	Mean (SD)
Poor	74.7 (5.44)
Medium	82.67 (5.30)
Good	85.00 (3.27)

[a]The table displays the mean (SD) scores provided by ChatGPT for different inquiry qualities related to inflammatory bowel disease symptom questions. The test involved three levels: “poor” (irrelevant questions and casual language), “medium” (mixed relevance), and “good” (precise and relevant). For each level, 30 interactions were conducted, and ChatGPT adhered to a virtual medical record generated for inflammatory bowel disease throughout the conversation, simulating a standardized patient.

XSL•FO
RenderX

Figure 1. Scores given by ChatGPT in feasibility test. This figure illustrates the density distribution of scores given by ChatGPT when evaluating different levels of inquiry quality (poor, medium, and good) regarding IBD symptoms. Each quality level reflects varying degrees of relevance in questioning, with ChatGPT adhering to an IBD standardized patient scenario. The distribution shows how ChatGPT assessed each inquiry type. IBD: inflammatory bowel disease.

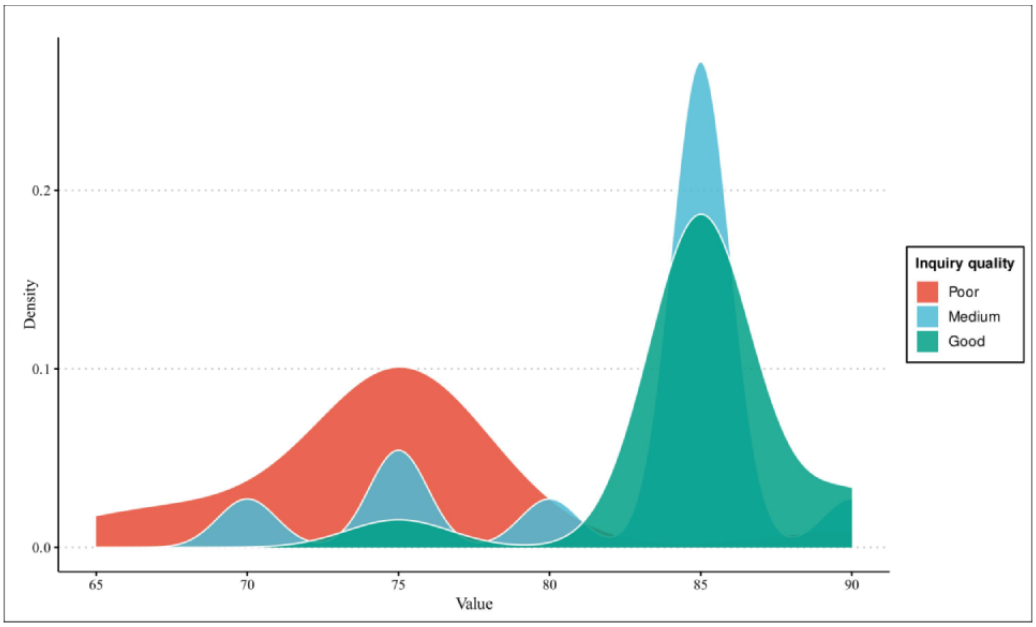

Performance Enhancement

The results from the preliminary experiment of the second phase demonstrated that ChatGPT can accurately replicate the clinical symptoms recorded in medical records; however, the scores it provided showed significant divergences compared to the standard scores. The results are presented in Figure 2. The upper band of each bar indicates the mean value of the SDP for each dataset, while the vertical line denotes SD. The data collected from the OP experiments are compared with the RP group subsequently.

To enhance performance, we focused on the following key aspects: degree of anthropomorphism, clinical accuracy, and adaptability. Regarding the degree of anthropomorphism, some responses provided by ChatGPT were notably overly professional with medical jargon. For instance, when inquired "Where do you feel unwell?" ChatGPT responded with "Processus xiphoideus." This term is rarely used in everyday language, typically, one might say "My chest hurts." Consequently, the chatbot was instructed to "Please perform like an ordinary person that does not have much professional knowledge in the medical field and avoid using professional jargons" to adjust its behavior. Moreover, without prompts, ChatGPT cannot replicate real-life standardized patients who may ask certain questions due to anxiety. To address this, the OP was enhanced with "When I present a summary of symptoms and seek your confirmation, please first respond to my inquiry." Subsequently, simulate a patient tone characterized by anxiety, posing questions like "Can my illness be treated?" or "Is this a serious or minor illness?" Concerning clinical accuracy, several issues were identified. When queried, "Do you have any vomiting symptoms?" ChatGPT responded "Yes, basically undigested food. And my stools are normal, no black stools appeared" (this response originated from the provided medical history). The chatbot specifically inquired about the vomit habits; however, it responded with details on vomit habits and additionally mentioned the stool condition, which was documented in the medical record subsequent to the vomiting information. This indicates that ChatGPT may provide premature and overly comprehensive responses. A prompt can be introduced to address this issue: "Just answer the question each time I gave you, do not provide information that is not related to my question." Furthermore, ChatGPT provided responses not aligning with the medical records. This issue, although infrequent, emerged several times during our experiment, when asked "Where are you feeling discomfort," ChatGPT replied "head" instead of "stomach." Although the chatbot can follow the content in medical records strictly in most situations, clinical accuracy is greatly favored, the chatbot must not make mistakes in this part. Additional prompts like "Please follow strictly with the information in medical record I provided you, do not compile information already provided in the medical record" can be added. When grading, ChatGPT sometimes did not focus on the conversation about standardized patients. After providing the criteria, the chatbot would just assume the conversation has happened. This could result in a high-scoring error rate. However, when initially adding the prompt "When scoring, please concentrate solely on the dialogue that took place during the standardized patient simulation. For each criterion listed, review our conversation history to determine whether I asked the specified question. If I did not, please assign a zero for that criterion." Regarding adaptability, when ChatGPT was asked about diseases or symptoms not provided in the medical record, the chatbot was unable to answer these questions accurately, while the correct response should be "No." For instance, when being asked "Have you experienced a heart attack," ChatGPT responded with "Sorry, I cannot confirm this, as it is not indicated in the medical record." As a medical record serves as a checklist for the patients' disease history and symptoms experienced to date, it should be a

XSL•FO
RenderX

complete reflection of the patients' medical situation. Therefore, when asked about the symptoms or disease that is not provided in the medical record, ChatGPT's answer should be "No, I do not have this symptom or disease." Additional prompt can be added to ascertain the adaptability: "The patient's all disease history and symptoms are as described in the medical record. The patient does not have any disease or symptoms not mentioned." Our results are shown in the circular diagram in Figure 3. Combining all the revisional prompts, new prompts are listed in Table 4.

The results of the RP group are shown in Figure 2. We first used the Shapiro-Wilk normality test to determine whether the data in each experiment were normally distributed. The results show that all the data we got, both in OP and RP groups, were not normally distributed (all P<.001). Then Mann-Whitney U test was used to determine the significance between the mean of each dataset and the standard score. For instance, considering the OP-1 group, the standard score is 20/100, we used the Mann-Whitney U test to test if the difference between the mean of real scores ChatGPT provided in OP-1 and number of 60 has significance. Then, Mann-Whitney U tests were conducted between each paired OP and RP groups, respectively, to determine if RP indeed improved the performance of ChatGPT. The detailed result is shown in Table 5.

Using another new patient script, a total of 60 experiments were conducted on 4 test groups. The results are shown in Figure 4. The PE+DC group exhibited the highest performance, achieving an average SDP value of 3.21%. The PC+DE group demonstrated the lowest performance, recording an average SDP value of 4.2%. The average SDP values achieved by the PC+DC and PE+DE groups were generally identical (SDP=3.46%). However, the PE+DE group showed superior performance over the PC+DC group, as evidenced by a smaller deviation in results. Meanwhile, the results from the Mann-Whitney U tests, conducted across all groups, indicated no significant differences (P=.65, between PC+DC and PE+DC; P=.46, between PC+DC and PC+DE; P=.98, between PC+DC and PE+DE; P=.29, between PE+DC and PC+DE; P=.51 between PE+DC and PE+DE; and P=.37, between PC+DE and PE+DE).

Figure 2. Error bar histogram of RP and OP experiments. This figure displays SDP across 5 experiments (OP-1 to OP-5 and RP-1 to RP-5) comparing OP and RP used for ChatGPT's evaluation as a standardized patient. Each experiment involved 30 interactions, measuring the deviation of ChatGPT's provided scores from the standard scores. The blue bars represent the mean SDP for each OP group, and the orange bars indicate the corresponding RP group. Error bars denote the SD of SDP, highlighting variations in accuracy across prompt types. OP: original prompt; RP: revised prompt; SDP: score difference percentage.

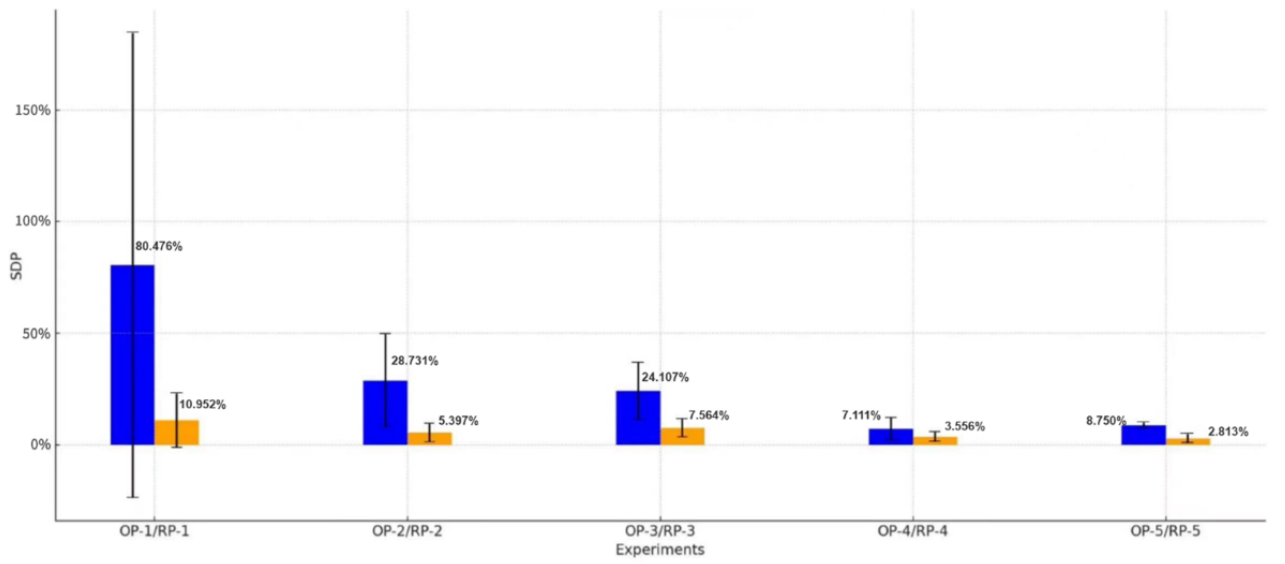

XSL•FO
RenderX

Figure 3. Prompt revision for ChatGPT in standardized patient assessment. This diagram summarizes key areas for prompt revision aimed at improving ChatGPT's performance as a standardized patient. Focus areas include clinical accuracy, degree of anthropomorphism, and adaptability. Each segment outlines specific issues identified during initial tests—such as premature answers, use of professional jargon, and deviation from medical records—and corresponding prompt modifications. These adjustments were designed to enhance ChatGPT's ability to simulate realistic patient behavior, adhere to medical records, and respond consistently during medical interactions.

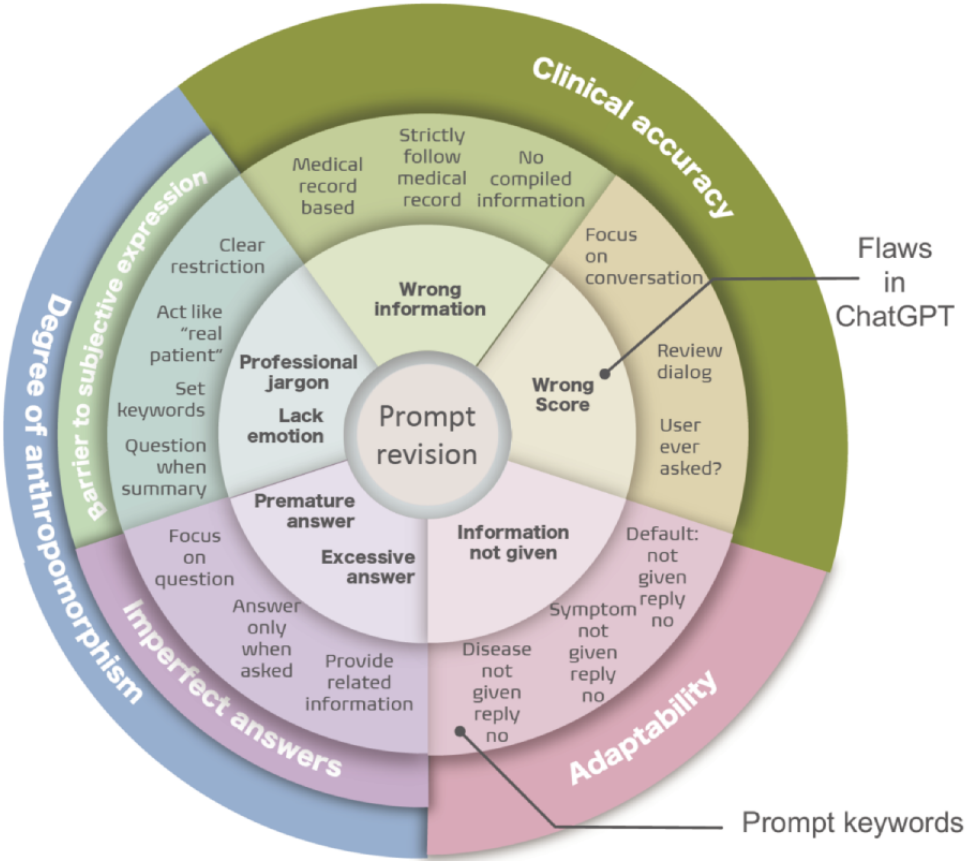

XSL•FO
RenderX

Table 4. The revised prompt given to ChatGPT[a].

Prompt	Consideration
I'm a first-year medical student. Please help me with an interactive test case. Are you familiar with the procedure of a standardized patient assessment? Now, I want you to play the role of a standardized patient, while I play the role of the student. Follow these steps:	Initialization
Your basic situation: [The detailed medical history is provided to ChatGPT based the script]	Set up specific scenario
You need to answer the questions I'm asking you based on the facts.	Restrict the free response of ChatGPT
Please perform like an ordinary person that does not have much professional knowledge in medical field. Avoid using professional jargons.	Avoid professional jargon
When I present a summary of symptoms and seek your confirmation, please first respond to my inquiry. Subsequently, simulate a patient tone characterized by anxiety, posing questions like 'Can my illness be cured?' or 'Is this a serious or minor illness?'	Set up question due to anxiety
Just answer the question each time I gave you, do not provide information that is not related to my questions.	Avoid premature and excessive answer
Please follow strictly with the information in medical record I provided you, do not compile information that is already provided in medical record.	Avoid wrong information
When giving score, please only focus on the conversation happened in standardized patient simulation, for each list in criteria, review our conversation history to check if I have ever asked this question, if not, then you give me a zero for this list.	Improve grading accuracy
The patient's all disease history and symptoms are as described in medical record. The patient does not have any disease or symptoms not mentioned.	Improve adaptability
At the end of my consultation, you need to rate my consultation process, the specific scoring rules are as follows: [The specific scoring rules are provided]	Grade with detailed criteria

[a]This is the revised prompt used for interactions with ChatGPT in the phase 2 study design, after prompt engineering.

Table 5. Statistical data for experiment, using SDP[a,b].

Experiment order	Real score			
	Mean (SD), %	Mean ChatGPT score	Accurate score	*P* value
OP[c]-1	80.476 (1.058)	12.633	7	<.001
RP[d]-1	10.952 (0.123)	7.033	7	.28
OP-2	28.731 (0.212)	26.567	21	<.001
RP-2	5.397 (0.041)	20.333	21	<.001
OP-3	24.103 (0.131)	32	26	<.001
RP-3	7.564 (0.039)	24.033	26	<.001
OP-4	7.111 (0.051)	32.134	30	<.001
RP-4	3.556 (0.021)	29.8	30	.81
OP-5	8.750 (0.013)	34.8	32	<.001
RP-5	2.813 (0.021)	32.833	32	<.001
OP-overall	29.834 (0.550)	__[e]	—	—
RP-overall	6.056 (0.068)	—	—	—

[a]SDP: score difference percentage.

[b]The table presents the accuracy of ChatGPT's scoring during standardized patient assessments, comparing the OPs and RPs. The scoring measures how well ChatGPT's responses align with predefined standard scores for each scenario. Metrics include the mean and SD of the SDP, indicating the deviation between ChatGPT's scores and the expected standards. The table also provides accurate scores and *P* values from Mann-Whitney *U* tests, evaluating the significance of differences between ChatGPT's scoring and the expected results.

[c]OP: original prompt.

[d]RP: revised prompt.

[e]No data.

XSL•FO
RenderX

Figure 4. Experiment results using new script. This figure displays the performance of ChatGPT using revised prompts on another script across 4 language combinations: PC+DC, PC+DE, PE+DC, and PE+DE. Each group underwent 15 trials, and the results were analyzed based on the SDP metric, which measures the accuracy of ChatGPT's scoring. The Mann-Whitney U tests indicated no significant differences between any group pairs, suggesting similar performance across language combinations. PC: prompts in Chinese; PE: prompts in English, DC: dialogue in Chinese; DE: dialogue in English. SDP: score difference percentage.

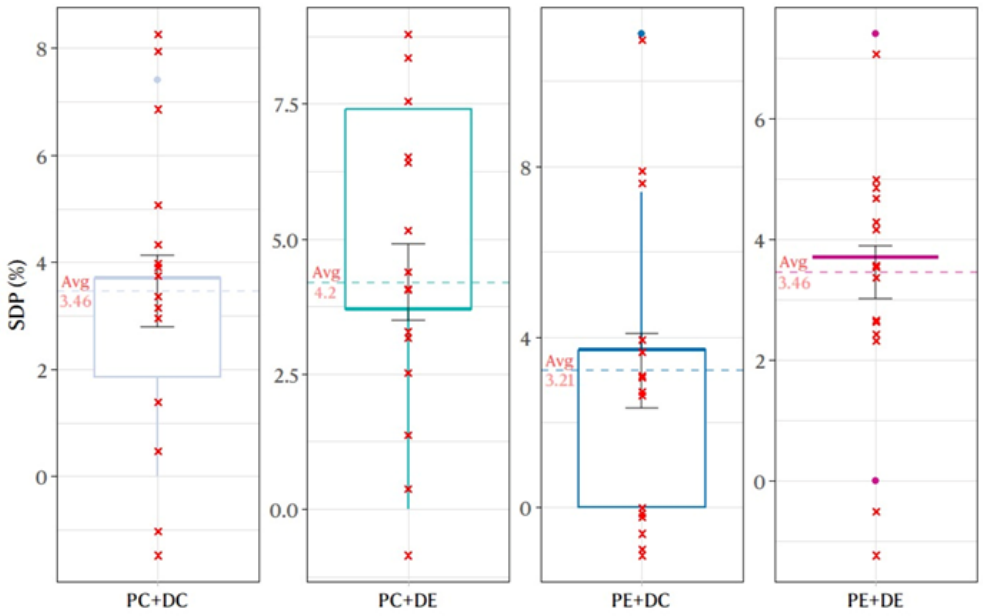

Discussion

Principal Findings

Phase 1: Preliminary Experiment

In this study, we conducted an exploratory study to examine the potential of using ChatGPT as standardized patients in medical education. Also, the performance enhancement after RP was evaluated. The knowledge of ChatGPT is not updated every day, therefore the precision and accuracy of the chatbot may decrease when it comes to new diseases and medicine [30,38]. All the papers designed in this study for ChatGPT evaluation do not include diseases newly found, as all the analyses and questions provided to ChatGPT were conducted between November 2023 to December 2023. We found that with refined prompts, ChatGPT could simulate patient interactions effectively, demonstrating improvements in realism, clinical accuracy, and adaptability. The use of ChatGPT as a standardized patient offers a cost-effective alternative to traditional human standardized patients, potentially enhancing access to medical training, especially in resource-limited settings.

Our results suggest that, in the preliminary experiment, poor inquiry is most dense in a relatively high score of around 75, this score seems to be higher than expected as all 4 questions inputted to ChatGPT are basically irrelevant to the standardized patient test. However, ChatGPT is an LLM that is finetuned by humans, it is finetuned to minimize offense to humans [39,40] It is understandable that ChatGPT tends to give the user a higher score even if all the questions are wrong. As to the chatbot, it may seem to be offensive if providing a low score to human users while not having a detailed criterion. This problem disappeared in the later experiments where detailed criterion was provided. As for the medium inquiry way, ChatGPT provided an average score higher than that of a poor inquiry model. Our results suggest that the score of the medium inquiry group is the most spread, as well as the largest data range among the 3 models. This reflects that in the field of standardized patient assessment, ChatGPT can distinguish between a relevant question and an irrelevant one, but its performance is not stable and easily affected without proper definition and restriction.

Even though the average score difference between medium and good inquiries is not significant, good inquiry has the smallest variance with the highest data density. As our results suggest, the difference between poor and medium groups is significant, while the difference between medium and good is not statistically significant. Considering other trends, however, the increase in the average score, along with other data features, with respect to the increase and difference in inquiry question relevance, indicate that ChatGPT is aware of which question does not belong to a standardized patient assessment process and which question belongs to it. Although it cannot distinguish clearly without a concrete prompt, the chatbot is familiar with the process of standardized patient assessment. However, its criterion is different from the criterion provided by experts, and the scores it provides are erroneous. This is sufficient to provide evidence for the feasibility of using ChatGPT to perform standardized patient training.

Phase 2: Prompt Engineering

The preliminary experiment of the performance enhancement set up a blueprint of the whole experiment. As illustrated in Figure 2, in the first few experiments, the improvement of RP is quite significant. For instance, in the OP-1 experiment, the mean score has the SDP around 0.8 which means that the score

XSL•FO
RenderX

provided by ChatGPT is 1.8 times higher than the accurate score. With an overall average SDP of 29.834%, this shows that ChatGPT's behavior is highly unpredictable and unreasonable, and thus not feasible for application based on the current method. One noticeable trend from Figure 2 is that, when the experiment continues, the SD and average score become more accurate. This is not because of the fact that the performance of ChatGPT improves as the experiment is conducted, rather, literature suggests ChatGPT falls short when handling similar tasks. Without prompt limitation, the chatbot tends to include the content of the whole conversation into grading, beyond the user's query. For example, if the user asked 1 question that addressed 1 key point in the criteria, but ChatGPT's reply went beyond the scope of this question, its answer may cover two or more key points in the criteria. Therefore, even though the user did not explicitly inquire about certain key points, ChatGPT's expansive response might still count these unasked points during grading. Consequently, this feature will let ChatGPT tend to provide a higher overall score.

To ensure a stable, accurate performance in standardized patient assessment, prompt engineering plays a critical role. ChatGPT's output can be generally modified and controlled by changing prompts, thus a good prompt with many restrictions can constrain ChatGPT to guide the chatbot to our ideal output. The enhancement and revision of prompts become pivotal, which is the most significant and pivotal contribution of this study. The reasons why choosing the degree of anthropomorphism, clinical accuracy, and adaptability as our evaluation key point is based on ChatGPT's prompt optimizing key features [41,42]. In the context of standardized patient training, the chatbot is required to perform like a real person, moreover, it must adhere more closely to the behaviors of a real-life patient, which is the essence of the test. In achieving so, the fundamental requirement for the chatbot is to have a high level of anthropomorphism which consists of 2 major problems discovered in this research: barrier to subjective expression and imperfect answers. For the former aspect, ChatGPT tends to use professional medical jargon to illustrate symptoms in medical reports which do not accurately reflect its role as a real patient. Also, in only 1 of 5 experiments, ChatGPT responded to a question as if it were experiencing anxiety. In OP-3, when asked "Have you ever been to the hospital before for this disease?" ChatGPT first answered the question objectively, then subjectively added, "I fear that this disease has lasted for too long, so I came here to consult you." Although the case that ChatGPT has an anxiety tone is rare, this demonstrates that the chatbot can imitate this tone, requiring only some constraints in the prompt. Future studies may explore the possible prompts engineering to improve its performance in simulating human emotions. In essence, the performance of ChatGPT in the preliminary experiment is unstable and poor; however, it possesses the capability for preliminary testing yet lacks guidance, thereby necessitating optimization via prompt engineering [43], underscoring the significance and necessity of the second-phase experiment.

Ultimate Performance Enhancement

After applying RP, ChatGPT used descriptive words to answer questions like "Where is your discomfort," instead of detailed medical jargon, and the chatbot asked questions like "Why my symptoms persist?" "Can my disease be cured?" After the user has summarized the conversation. Though the chatbot's tone may seem "mechanical," it indicates that RP indeed works. For the latter aspect, the imperfect answers here refer to the chatbot's premature and excessive answers, which significantly influence grading. This part has a great influence on grading. For instance, when asked "Have you ever had other diseases in stomach?" the chatbot first answered yes, and explained when diagnosed and with what diseases. After providing sufficient information for the question, ChatGPT added "I have been to the hospital for the first time for this 2 months ago." The answer may seem reasonable. Yet, during grading, ChatGPT automatically awarded full marks for "Asking about information on previous hospital visits," justifying that the user had inquired about it. However, no question related to this was asked in that conversation. When asked for more detailed information, ChatGPT responded that since the content of the conversation mentioned the patient had gone to the hospital 2 months ago, it should count; even though this information was provided by the chatbot subjectively, not in response to the user's question. Therefore, premature and excessive questions not only influence degrees of anthropomorphism but also contribute dramatically to grading errors. Subsequently, clinical accuracy is also a concern [44]. When restrictive prompts such as "Focus on my question" and "Answer only what is asked" are used, ChatGPT began to use plain, short answers to respond, from one extreme to another, but the answer contains the essence without premature questions, making these responses preferable. During the preliminary experiments, ChatGPT can strictly follow the medical record, only once mistaking a stomachache for a headache. Additionally, in 1 conversation, the chatbot provided compiled information on the medical history. Therefore, to assure accuracy, a restrictive prompt asking ChatGPT to base answers only on the record is crucial. Also, the chatbot must be accurate when giving the score, as this is the only quantitative reflection of ChatGPT's performance in correctly evaluating medical students' performance who are taking the standardized patient examination.

With RP, the chatbot now focuses on the content that appears in the standardized patient training conversation, before that, it sometimes mistook the content of the criteria as part of the judgment, resulting in full mark. Adaptability is also important, RP made a ChatGPT response with "No" for the symptoms asked but not in medical history in the following experiment. After applying RP, ChatGPT's performance improves dramatically. One of the most serious problems in using ChatGPT is that the answer it provides has a high level of randomness and this is hard to avoid. The average SDP improved to 20 times higher than OP results. Considering the high randomness in answers provided by the chatbot, this result is quite acceptable. Also, the score deviation for each experiment is quite small, the average SD decreased to 0.068 from 0.55 in the original group results. By adding these prompts, we reduce SD to nearly one-fifteenth of its original value. Massive differences in average value and SD indicate that by giving our extra, RP, the performance of ChatGPT as a standardized patient improves significantly. The significant difference between the OP and RP results indicates that RP improved ChatGPT's performance significantly. RP-2, RP-3, and RP-5 generally have

XSL•FO
RenderX

lower SD with a lower mean SDP, which represents a higher accuracy and precision than their corresponding OP groups. To fully use ChatGPT in a standardized patient training process, good accuracy and consistency in responses are required. After applying our RP, ChatGPT has an overall SDP of 6.056%. This score is closer to 0% and improves a lot compared to the un-RP group (29.834%), even though there still remains space to improve. The results also indicate that ChatGPT's performance can be fine-tuned through the revision of prompts provided to it. Future studies may conduct a more comprehensive examination to reduce the SDP further.

The efficacy of RP was confirmed through its universal applicability when tested on another script, achieving a similar average SDP to our preliminary script. Furthermore, validation tests conducted on the new script also indicated that the impact of language on chatbot performance is insignificant ($P>.05$, for all groups). The observed performance deviation is within acceptable limits. However, the highest accuracy rate was observed in the test group that used English prompts and Chinese dialogues.

Limitations

There are several limitations in this study. First, to minimize errors associated with word translation, we conducted the simulated standardized patient test dialog with ChatGPT in standard simplified Chinese [45]. Nevertheless, the prompt provided to ChatGPT was in English to ensure accuracy. Although a separate experiment was conducted to confirm the minor impact of language used on the performance of ChatGPT, future studies may test the performance of ChatGPT on standardized patients in other languages and corresponding prompt engineering to improve its performance. Secondly, the criteria provided to ChatGPT involved only objective key points. In real standardized patient tests, some subjective judgments may also be made accordingly. Finally, the prompt revision we provided represents the optimal approach. Further research is necessary to use ChatGPT more effectively in standardized patient training.

Conclusions

ChatGPT, as a representative LLM, has much potential application in medical assessment. Our study suggests that it is feasible to use ChatGPT as a simulated patient for standardized patient evaluation and scoring. The output performance concerns introduced by the randomness of ChatGPT could be improved by adding detailed, restrictive prompts. By using prompts revised in this study, the accuracy of ChatGPT's scoring significantly improved to an acceptable level that could be used in actual standardized patients involved in medical training. Additionally, ChatGPT's responses during the conversation became more accurate and lifelike. It is worth noting that, in this application scenario, the influence of the adopted language on the chatbot's outcome is nonsignificant. Overall, ChatGPT's accuracy and performance in standardized patient assessment are acceptable, but they highlight the need for continuing improvement before it can be used as a fully trustworthy clinical assessment method.

Acknowledgments

The authors used the generative artificial intelligence tool ChatGPT by OpenAI to examine the potential use of LLMs in replacing the role of standardized patients, mainly focusing on their feasibility and accuracy. A sample dialogue with ChatGPT is made available in Multimedia Appendix 1. This study was funded by the Aier Eye Hospital-Sichuan University Research Grant (23JZH043). This study was supported by Training Programs for Innovation and Entrepreneurship of Sichuan University (project 20241039L); the Interdisciplinary Crossing and Integration of Medicine and Engineering for Talent Training Fund, West China Hospital, Sichuan University; and the 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYYC21004). The sponsor had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. The views expressed are those of the authors and not necessarily those of the West China Hospital and Sichuan University.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to institutional case privacy and a large number of interaction dialogs but are available from the corresponding author on reasonable request. Other than the sample dialogue provided in the Multimedia Appendix 1, additional detailed dialogues with ChatGPT may be available by contacting the corresponding author.

Authors' Contributions

RY conceptualized the study, curated the data, performed formal analysis, and managed the project. YH and XW contributed to the conceptualization of the study. GQ also conceptualized the study, acquired funding, and managed the project. CW developed the methodology, curated the data, performed formal analysis, conducted the investigation, prepared the original draft of the manuscript, and contributed to visualization. SL and NL developed the methodology, conducted the investigation, and prepared the original draft of the manuscript. DL, XT, and XZ contributed to the visualization and preliminary manuscript revision. KL acquired funding, supervised the project, and managed the project. DP supervised the project and managed the project. All authors reviewed and edited the manuscript and have read and agreed to the published version.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
A sample interaction dialogue with ChatGPT for RP-2 (revised prompts) group (standard score=21).
[DOCX File , 26 KB - jmir_v27i1e59435_app1.docx]

References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
2. Rane NL, Tawde A, Choudhary SP, Rane J. Contribution and performance of ChatGPT and other large language models (LLM) for scientific and research advancements: a double-edged sword. Int Res J Modernization Eng Technol Sci 2023;5(10):875-899. [doi: 10.56726/irjmets45213]
3. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023;307(2):e230163. [doi: 10.1148/radiol.230163] [Medline: 36700838]
4. Mattas PS. ChatGPT: a study of AI language processing and its implications. Int J Res Publication Rev 2023;04(02):435-440. [doi: 10.55248/gengpi.2023.4218]
5. Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. J Inf Technol Case Application Res 2023;25(3):277-304. [doi: 10.1080/15228053.2023.2233814]
6. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. Int J Oral Sci 2023;15(1):29 [FREE Full text] [doi: 10.1038/s41368-023-00239-y] [Medline: 37507396]
7. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
8. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
9. Al-Worafi YM, Goh KW, Hermansyah A, Tan CS, Ming LC. The use of ChatGPT for education modules on integrated pharmacotherapy of infectious disease: educators' perspectives. JMIR Med Educ 2024;10:e47339. [doi: 10.2196/47339] [Medline: 38214967]
10. Parikh PM, Shah DM, Parikh KP. Judge juan manuel padilla garcia, ChatGPT, and a controversial medicolegal milestone. IJMS 2023;75:3-8. [doi: 10.25259/ijms_31_2023]
11. ChatGPT. OpenAI. URL: https://help.openai.com/en/articles/6825453-chatgpt-release-notes [accessed 2023-12-18]
12. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ 2023;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]
13. Wu T, He S, Liu J, Sun S, Liu K, Han QL, et al. A brief overview of ChatGPT: the history, status quo and potential future development. IEEE/CAA J Automatica Sin 2023;10(5):1122-1136. [doi: 10.1109/jas.2023.123618]
14. Roumeliotis KI, Tselikas ND. ChatGPT and open-AI models: a preliminary review. Future Internet 2023;15(6):192. [doi: 10.3390/fi15060192]
15. Short CE, Short JC. The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. J Bus Venturing Insights 2023;19:e00388. [doi: 10.1016/j.jbvi.2023.e00388]
16. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individual Differ 2023;103:102274. [doi: 10.1016/j.lindif.2023.102274]
17. Keiser MM, Turkelson C. Using students as standardized patients: development, implementation, and evaluation of a standardized patient training program. Clin Simul Nurs 2017;13(7):321-330. [doi: 10.1016/j.ecns.2017.05.008]
18. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare 2023;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
19. Lee H. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ 2024;17(5):926-931. [doi: 10.1002/ase.2270] [Medline: 36916887]
20. Wong RSY, Ming LC, Raja Ali RA. The intersection of ChatGPT, clinical medicine, and medical education. JMIR Med Educ 2023;9:e47274 [FREE Full text] [doi: 10.2196/47274] [Medline: 37988149]
21. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ 1964;39:802-805 [FREE Full text] [Medline: 14180699]
22. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. Acad Med 1993;68(6):443-451. [doi: 10.1097/00001888-199306000-00002] [Medline: 8507309]

XSL•FO
RenderX

23. Ainsworth MA, Rogers LP, Markus JF, Dorsey NK, Blackwell TA, Petrusa ER. Standardized patient encounters. a method for teaching and evaluation. JAMA 1991;266(10):1390-1396. [doi: 10.1001/jama.266.10.1390] [Medline: 1880869]
24. Plaksin J, Nicholson J, Kundrod S, Zabar S, Kalet A, Altshuler L. The benefits and risks of being a standardized patient: a narrative review of the literature. Patient 2016;9(1):15-25. [doi: 10.1007/s40271-015-0127-y] [Medline: 26002043]
25. King AM, Perkowski‐Rogers LC, Pohl HS. Planning standardized patient programs: case development, patient training, and costs. Teach Learn Med 1994;6(1):6-14. [doi: 10.1080/10401339409539636]
26. Petrusa ER. Taking standardized patient-based examinations to the next level. Teach Learn Med 2004;16(1):98-110. [doi: 10.1207/s15328015tlm1601_19] [Medline: 14987183]
27. Tong Y, Tan CH, Sia C, Shi Y, Teo HH. Rural-urban healthcare access inequality challenge: transformative roles of information technology. MIS Q 2022;46(4):1937-1982. [doi: 10.25300/misq/2022/14789]
28. Kocoń J, Cichecki I, Kaszyca O, Kochanek M, Szydło D, Baran J, et al. ChatGPT: jack of all trades, master of none. Inf Fusion 2023;99:101861. [doi: 10.1016/j.inffus.2023.101861]
29. Lee U, Jung H, Jeon Y, Sohn Y, Hwang W, Moon J, et al. Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in English education. Educ Inf Technol 2023;29(9):11483-11515. [doi: 10.1007/s10639-023-12249-8]
30. Haze T, Kawano R, Takase H, Suzuki S, Hirawa N, Tamura K. Influence on the accuracy in ChatGPT: differences in the amount of information per medical field. Int J Med Inform 2023;180:105283. [doi: 10.1016/j.ijmedinf.2023.105283] [Medline: 37931432]
31. Muftić F, Kadunić M, Mušinbegović A, Abd Almisreb A. Exploring medical breakthroughs: a systematic review of ChatGPT applications in healthcare. Southeast Eur J Soft Comput 2023;12(1):41 [FREE Full text]
32. GPT-4. Chatbot App. URL: https://tinyurl.com/bdhpmh7p [accessed 2023-12-18]
33. Pithadia AB, Jain S. Treatment of inflammatory bowel disease (IBD). Pharmacol Rep 2011;63(3):629-642 [FREE Full text] [doi: 10.1016/s1734-1140(11)70575-8] [Medline: 21857074]
34. Yin R, Neyens DM. Online health resource use by individuals with inflammatory bowel disease: analysis using the National Health Interview Survey. J Med Internet Res 2020;22(9):e15352 [FREE Full text] [doi: 10.2196/15352] [Medline: 32969831]
35. King JJC, Das J, Kwan A, Daniels B, Powell-Jackson T, Makungu C, et al. How to do (or not to do) using the standardized patient method to measure clinical quality of care in LMIC health facilities. Health Policy Plan 2019;34(8):625-634 [FREE Full text] [doi: 10.1093/heapol/czz078] [Medline: 31424494]
36. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023;25:e48568 [FREE Full text] [doi: 10.2196/48568] [Medline: 37379067]
37. Fang C, Wu Y, Fu W, Ling J, Wang Y, Liu X, et al. How does ChatGPT-4 preform on non-English national medical licensing examination? An evaluation in Chinese language. PLOS Digit Health 2023;2(12):e0000397 [FREE Full text] [doi: 10.1371/journal.pdig.0000397] [Medline: 38039286]
38. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
39. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber Phys Syst 2023;3:121-154. [doi: 10.1016/j.iotcps.2023.04.003]
40. Liu B, Xiao B, Jiang X, Cen S, He X, Dou W. Adversarial attacks on large language model-based system and mitigating strategies: a case study on ChatGPT. Secur Commun Netw 2023;2023:1-10. [doi: 10.1155/2023/8691095]
41. Henrickson L, Meroño-Peñuela A. Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. AI Soc 2023;39(5). [doi: 10.1007/s00146-023-01752-8]
42. Giray L. Prompt engineering with ChatGPT: a guide for academic writers. Ann Biomed Eng 2023;51(12):2629-2633. [doi: 10.1007/s10439-023-03272-4] [Medline: 37284994]
43. Lubiana T, Lopes R, Medeiros P, Silva JC, Goncalves ANA, Maracaja-Coutinho V, et al. Ten quick tips for harnessing the power of ChatGPT in computational biology. PLoS Comput Biol 2023;19(8):e1011319 [FREE Full text] [doi: 10.1371/journal.pcbi.1011319] [Medline: 37561669]
44. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq 2023. [doi: 10.21203/rs.3.rs-2566942/v1] [Medline: 36909565]
45. Li L, Zhang H, Li C, You H, Cui W. Evaluation on ChatGPT for Chinese language understanding. Data Intell 2023;5(4):903. [doi: 10.1162/dint_a_00232]

Abbreviations

AI: artificial intelligence
DC: dialogue in Chinese
DE: dialogue in English
IBD: inflammatory bowel disease

XSL•FO
RenderX

LLM: large language model
OP: original prompt
PC: prompts in Chinese
PE: prompts in English
RP: revised prompt
SDP: score difference percentage

Edited by A Coristine; submitted 11.04.24; peer-reviewed by A Ciftci Sisman, Y Wang, H Huang; comments to author 09.08.24; revised version received 24.08.24; accepted 11.11.24; published 01.01.25.

Please cite as:
Wang C, Li S, Lin N, Zhang X, Han Y, Wang X, Liu D, Tan X, Pu D, Li K, Qian G, Yin R
Application of Large Language Models in Medical Training Evaluation—Using ChatGPT as a Standardized Patient: Multimetric Assessment
J Med Internet Res 2025;27:e59435
URL: https://www.jmir.org/2025/1/e59435
doi:10.2196/59435
PMID:

©Chenxu Wang, Shuhan Li, Nuoxi Lin, Xinyu Zhang, Ying Han, Xiandi Wang, Di Liu, Xiaomei Tan, Dan Pu, Kang Li, Guangwu Qian, Rong Yin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 01.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Applications of Large Language Models in the Field of Suicide Prevention: Scoping Review

Glenn Holmes[1], PhD; Biya Tang[1], PhD; Sunil Gupta[2], PhD; Svetha Venkatesh[2], PhD; Helen Christensen[1], PhD; Alexis Whitton[1], PhD

[1]Black Dog Institute, University of New South Wales, Sydney, Randwick, Australia

[2]Applied Artificial Intelligence Institute, Deakin University, Melbourne, Australia

Corresponding Author:
Alexis Whitton, PhD
Black Dog Institute
University of New South Wales, Sydney
Hospital Road
Randwick, 2031
Australia
Phone: 61 290659046
Email: a.whitton@unsw.edu.au

Abstract

Background: Prevention of suicide is a global health priority. Approximately 800,000 individuals die by suicide yearly, and for every suicide death, there are another 20 estimated suicide attempts. Large language models (LLMs) hold the potential to enhance scalable, accessible, and affordable digital services for suicide prevention and self-harm interventions. However, their use also raises clinical and ethical questions that require careful consideration.

Objective: This scoping review aims to identify emergent trends in LLM applications in the field of suicide prevention and self-harm research. In addition, it summarizes key clinical and ethical considerations relevant to this nascent area of research.

Methods: Searches were conducted in 4 databases (PsycINFO, Embase, PubMed, and IEEE Xplore) in February 2024. Eligible studies described the application of LLMs for suicide or self-harm prevention, detection, or management. English-language peer-reviewed articles and conference proceedings were included, without date restrictions. Narrative synthesis was used to synthesize study characteristics, objectives, models, data sources, proposed clinical applications, and ethical considerations. This review adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) standards.

Results: Of the 533 studies identified, 36 (6.8%) met the inclusion criteria. An additional 7 studies were identified through citation chaining, resulting in 43 studies for review. The studies showed a bifurcation of publication fields, with varying publication norms between computer science and mental health. While most of the studies (33/43, 77%) focused on identifying suicide risk, newer applications leveraging generative functions (eg, support, education, and training) are emerging. Social media was the most common source of LLM training data. Bidirectional Encoder Representations from Transformers (BERT) was the predominant model used, although generative pretrained transformers (GPTs) featured prominently in generative applications. Clinical LLM applications were reported in 60% (26/43) of the studies, often for suicide risk detection or as clinical assistance tools. Ethical considerations were reported in 33% (14/43) of the studies, with privacy, confidentiality, and consent strongly represented.

Conclusions: This evolving research area, bridging computer science and mental health, demands a multidisciplinary approach. While open access models and datasets will likely shape the field of suicide prevention, documenting their limitations and potential biases is crucial. High-quality training data are essential for refining these models and mitigating unwanted biases. Policies that address ethical concerns—particularly those related to privacy and security when using social media data—are imperative. Limitations include high variability across disciplines in how LLMs and study methodology are reported. The emergence of generative artificial intelligence signals a shift in approach, particularly in applications related to care, support, and education, such as improved crisis care and gatekeeper training methods, clinician copilot models, and improved educational practices. Ongoing human oversight—through human-in-the-loop testing or expert external validation—is essential for responsible development and use.

Trial Registration: OSF Registries osf.io/nckq7; https://osf.io/nckq7

XSL•FO
RenderX

(J Med Internet Res 2025;27:e63126) doi:10.2196/63126

KEYWORDS

suicide; suicide prevention; large language model; self-harm; artificial intelligence; AI; PRISMA

Introduction

Background

Prevention of suicide is a global health priority [1]. Approximately 800,000 individuals die by suicide yearly, and for every suicide death, there are another 20 estimated suicide attempts [1]. Despite the largely preventable nature of suicide, factors such as limited service capacity, variable service quality, and barriers to service access significantly impact the progress being made toward reducing suicide rates [2]. Recent advances in transformer-based artificial intelligence (AI)—the technology that has accelerated the development of powerful large language models (LLMs) that human clinicians and consumers can converse with—have been suggested as a potential solution to enhancing the scalability, accessibility, and personalization of health care interventions [3]. In the context of suicide research and care, LLMs can generate novel insights into suicide risk by parsing language, classifying or scoring text, and comprehending and generating human-like language, with the potential to enhance treatments by improving user engagement with digital interventions. Such improvements can also expand the capacity of crisis support services by providing real-time crisis clinician copilots, automated risk assessments, or expedited triage, enabling the optimized use of human resources. In addition, LLMs can improve the quality of training programs for those who may intervene by mentoring trainees through clinical scenarios and flexibly adopting a variety of crisis-seeking personas for clinical role-play scenarios. Hence, LLM technology might be a critical catalyst for advancement in the field. However, the relatively unexplored nature of the field means that we are only just beginning to understand how LLMs can be harnessed safely and effectively for suicide prevention. In parallel, there is a pressing need to explore how emerging trends in LLM-based research, such as the nature of training data, model types, the contexts of application, and methodological norms from relevant research fields, may shape the trajectory and direction of suicide prevention research (either positively or negatively), now and into the future.

AI refers broadly to machine and computational processes used to execute tasks usually thought of as requiring human intelligence. Natural language processing, a language-focused subfield of AI, has advanced significantly with the advent of transformer-based LLMs [4]. LLMs are the current state-of-the-art technology in computational linguistics, comprising attention-based language encoders trained using massive text datasets, generally with billions of parameters [5]. In addition to language comprehension, LLMs have recently shown their utility in generative language applications, such as generative AI interfaces (chatbots) [6]. The availability of LLMs, such as OpenAI's generative pretrained transformer (GPT) models, Google's Bard, and Meta's Llama, has created unprecedented opportunities for language generation and analysis at scale [7], with LLM applications already becoming widespread across fields such as business analytics, commerce, administration, and education [8].

Under the right conditions, LLMs can demonstrate contextual understanding and content generation that closely mimics human interaction. Accordingly, research into LLM applications in health care settings, where human interaction forms the basis of much of service delivery, is accelerating at a rapid pace; for example, LLMs have been investigated as a potential aid in preconsultation, diagnosis, and the management of disease (eg, infectious diseases and cancer [9-11]), as well as in recommending specialist appointments via SMS text messaging–based self-assessment tools for remote populations [12] and generating patient education materials [13].

Given that language is the primary basis upon which symptoms are reported and assessed in mental health, LLMs represent a significant technological advancement in mental health research. LLMs are currently under investigation for their utility in cognitive behavioral therapy facilitation [14], for emotion identification during psychotherapy sessions [15], and for the detection of positive therapeutic behaviors during motivational interviewing [16]. LLMs have also shown potential benefits in helping individuals understand personal coping styles and in facilitating stress reappraisal [7]. The application of LLMs in the field of suicide prevention is particularly promising, given that traditional methodologies, which have struggled to provide actionable insights into complex constructs such as suicide risk, are complemented by the analytic and generative capabilities of LLMs [17].

In evaluating the promising aspects of LLMs, it is also important to consider that these models present a number of challenges to researchers and end users alike. LLMs are often criticized for the inaccuracy of generated information, reducing trust and credibility [5], a failing that is compounded by the lack of interpretability of LLM functioning resulting from their "black box" architecture. Challenges also exist around data security and privacy; for example, LLMs have been demonstrated to identify individuals from deidentified digital data (eg, location data, medical billing information, or a collection of social media posts and metadata) [17]. When considering application contexts, the generalizability of LLMs is restricted to the degree of representativeness of the training data with respect to age, ethnicity, or even those who use social media versus those who do not [18]. Nevertheless, LLMs hold promise for improving models of suicide risk (eg, via digital phenotyping) [17]; for creating synthetic data [19] to increase sample sizes and statistical power for research on low base-rate events such as suicide; for powering conversational agents that support early triage, crisis support, and help seeking; and for clinician assistance tools, such as automated patient assessment systems, and simulation of crisis scenarios to improve training outcomes. Despite these potentialities, the integration of LLMs into suicide prevention currently trails other fields. Critically, significant gaps must be addressed for the field to move forward in a

XSL•FO
RenderX

manner that is safe and acceptable, including understanding the interpretability of LLM outputs, addressing biases within training data, and ensuring ethical deployment within prospective crisis settings.

To date, there have been no reviews examining the integration of LLMs into research on suicide prevention and self-harm. Prior reviews have focused predominantly on research undertaken in adjacent fields, such as applications of machine learning or chatbots to mental health more broadly [20-23]. Although several reviews have explored the use of machine learning [24-27] or AI-based strategies [28-31] in suicide prevention contexts, these have not focused on applications of LLMs specifically. Furthermore, although 1 commentary [32] and a review [25] focused on the integration of computational linguistics or natural language processing more broadly in suicide prevention, these articles did not focus on LLMs specifically, nor have any explored applications of LLMs to self-harm.

Objectives

The aim of this scoping review was to provide an understanding of current applications of LLMs in the field of suicide prevention, including applications to self-harm. The research questions guiding this review were as follows:

1. What trends are present in the literature with regard to the models used, data sources, and LLM objectives?
2. What are the clinical applications reported in the literature?
3. What ethical considerations are noted in the literature?

Specifically, we aimed to identify emerging trends in the literature related to the types of training data, models used, and intended objectives.

Our secondary aim was to explore reported potential clinical applications of LLMs (eg, as clinical copilots or to augment suicide prevention training programs) and identify ethical considerations that are crucial as the field progresses. Reported potential clinical applications were sought to better understand the future trajectory of practical LLM use in the field, with ethical considerations identified to understand the important considerations for practical deployment.

A *scoping* review was selected to map the breadth of evidence across the varied fields from which LLM-based research originates to identify key characteristics of studies, reveal knowledge gaps, and inform future research [33]. In doing so, we aim to highlight emergent trends in the research and address critical clinical and ethical considerations to ensure safe, effective, and equitable research outcomes in this key field.

Methods

This review is presented in line with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1) and was developed and carried out with reference to the JBI best practice methodological standards [34,35]. In accordance with the JBI recommendations [36], the protocol was preregistered on the Open Science Framework and is publicly available [37].

Inclusion and Exclusion Criteria

To meet the eligibility criteria, studies needed to describe the application of an LLM to the area of suicide or self-harm. Although various natural language processing models using machine learning methods exist (eg, support vector machines, Bayesian networks, and random forest algorithms), for this review, we focused on contemporary LLMs, defined as computer-engineered language models that use transformer-based neural network architecture [4]. Although LLM training parameters generally exceed 10 billion, there is no formal consensus on parameter scale [6]. In this review, smaller models (<10 billion parameters) such as early incarnations of the Bidirectional Encoder Representations from Transformers (BERT) and Text-to-Text Transfer Transformer (T5) were eligible for inclusion as they represent language models capable of contextual understanding. Studies describing the use of AI, machine learning, or big data approaches in the absence of LLM methodology were ineligible.

The domain of suicide prevention included, but was not restricted to, areas such as ideation, planning, attempts, prediction, intervention, support, means restriction, gatekeeper training, and public awareness campaigns, as well as self-harm. Studies focusing on LLM applications to mental health or psychiatry that did not specifically mention suicide or self-harm were ineligible. Studies describing quantitative, qualitative, or mixed methods designs were eligible. All studies involving human participants were required to report ethics approval. Source documents were peer-reviewed journal articles or conference proceedings. Conference proceedings were included as they have become the dominant form of published research in computer science (encompassing LLMs) in recent years [38]. Conference proceedings were required to have a comprehensive methodology and results sufficient for replication. Conference abstracts were excluded, as were other reviews and meta-analyses.

Studies using electronic health records (EHRs) were not included in this review. EHRs represent a specific type of corpus representative of individuals who are in contact with the health system. Given that a significant proportion of individuals experiencing suicidal ideation are not in contact with formal health services [39], models that draw primarily on EHR data may generate insights that do not generalize to the broader population of individuals who experience suicidality or self-harm. In addition, substantial work has focused on the use of EHRs for the prediction of suicide risk in recent years [20], with the potential for these studies to populate a stand-alone systematic review. This review sought to elicit an understanding of novel deployments of LLMs, particularly in the individual use context, and to propose future use possibilities or potentialities for support, treatment, or prevention. Studies not published in English were also excluded from this review. There was no limitation on country of origin or publication date; however, the advent of transformer architecture in 2017 naturally limited publications from prior years.

Search Strategy

Searches were conducted in 4 databases: PsycINFO, Embase, PubMed, and IEEE Xplore. An initial search was conducted on

XSL•FO
RenderX

December 6, 2023. Following abstract and full-text screening, the search was updated on February 16, 2024, to ensure recency of the search prior to final text extraction. Additional studies identified via citation chaining were also included at this point.

The search strategy used index terms and free-text terms to cover two core themes: (1) LLMs and (2) suicide or self-harm. Individual database search strings are provided in Table S1 in Multimedia Appendix 2. Search results were imported into Covidence review management software (Veritas Health Innovation Ltd) [40], which was used for abstract and full-text screening as well as data extraction.

Study Selection and Screening Process

Reflecting prior systematic research [41], title and abstract screening were conducted by 1 author (GH), with a sample of studies reviewed by multiple authors during screening and extraction. In line with recommendations for ensuring reliability and bias mitigation [33], 20.2% (77/381) of the studies identified in the initial search were randomly selected and reviewed by a second author (BT). Agreement was achieved for 71 (92%) of the 77 studies. Conflicts were discussed and resolved with input from a third author (AW).

A similar process was adopted for full-text screening, which was conducted by 1 author (GH), with 20.3% (38/187) of the full texts randomly selected for review by a second author (BT). There was 92% (35/38) agreement between the authors. Conflicts were discussed and resolved with input from a third author (AW).

Data Extraction

The data extraction template was piloted with 9% (4/43) of the included studies by 2 authors (GH and BT) in line with prior research [41,42], resulting in >90% agreement. Minor refinements were made to the template to assist later synthesis. Data extraction from the 43 included articles was performed by 1 author (GH), with 19% (8/43) of the studies randomly selected and data independently extracted by a second author (BT). The comparison of extracted data showed >90% agreement. The data extraction template included items under the headings of study characteristics, methods, outcomes, and reproducibility and is available in Table S2 in Multimedia Appendix 2.

Analysis

Studies were classified across deductively derived categories, with categories based on previous research [42] and with scope for adjustments during the review process if required (eg, the aggregation of specific data sources to report on social media more generally as a source of data). Descriptive statistics were computed and presented in textual, tabular, or graphical format, as appropriate. A narrative synthesis of the included studies was conducted to answer the key research questions.

Results

Selection of Sources of Evidence

The search results are presented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 1). The search yielded 533 studies; after removing 127 (23.8%) duplicates, 406 (76.2%) studies were retained for title and abstract screening. Of these 406 studies, 194 (47.8%) were excluded, leaving 212 (52.2%) for full-text review. After full-text screening, 176 (83%) of the 212 studies were excluded (the reasons for exclusion are shown in Figure 1), and 36 (17%) studies met the inclusion criteria. An additional 7 studies were identified through citation chaining, resulting in 43 studies included for synthesis.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. EHR: electronic health record; LLM: large language model.

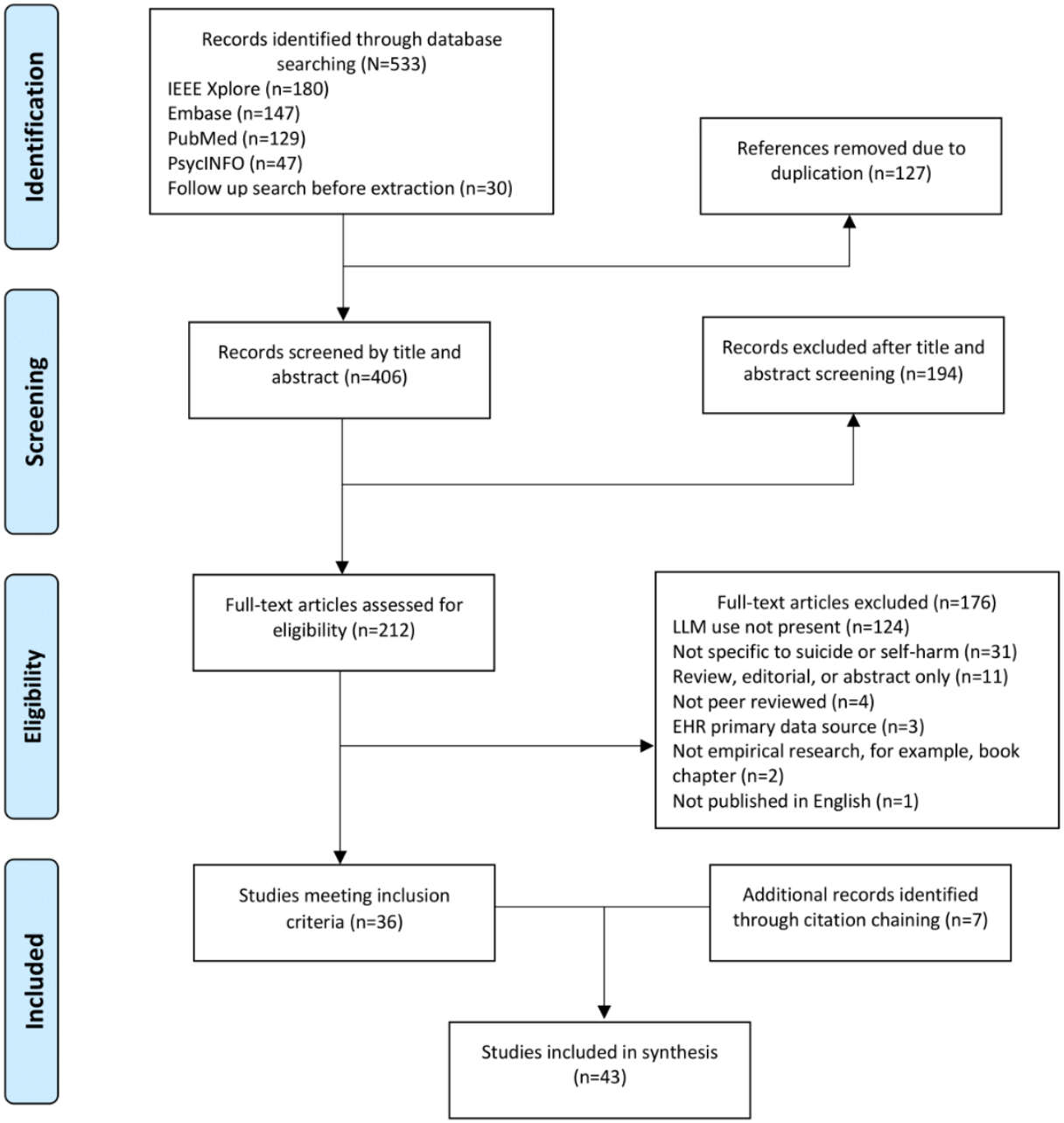

Characteristics of Included Studies

A full list of included publications and associated results data is presented in Multimedia Appendix 3 [43-85]. The number of publications by year demonstrated a compounding pattern, with the first publications emerging in 2019 after the development of the transformer architecture [4] in 2017 and then increasing sharply in 2023 after the release of ChatGPT to the public in November 2022.

The articles identified for this review came from 20 different sources that predominantly came from 2 fields (determined by reviewing individual journal aims and scope): computer engineering (23/43, 53%) and health (18/43, 42%). The bifurcated nature of the publication field of origin demonstrates the cross-disciplinary nature of LLM research. Approximately a third of the studies (16/43, 37%) were published as conference proceedings via IEEE Xplore, reflecting publication norms in computer engineering [38]. The largest proportion of studies came from the United States (10/43, 23%). Study funding was predominantly sourced through grants (23/43, 53%). Of the 43 studies, 4 (9%) indicated no funding, and 16 (37%) did not provide funding information.

Synthesis of Results

Base LLMs Used

A synthesis of identified trends with associated recommendations is presented in Figure 2. Most of the studies (35/43, 81%) applied some derivation of Google's BERT.

XSL•FO
RenderX

OpenAI's GPT family of LLMs were used in 9 (21%) of the 43 studies. Other models included XLNet (3/43, 7%); Google's Fine-Tuned Language Net (FLAN; 2/43, 5%); and Alpaca, Alexa, DeepMoji, and contrastive language-image pretraining (1/43, 2% each). Some studies applied >1 model type. The widespread adoption of BERT can be attributed in part to its status as one of the earliest open-source models (accessible since its inception in 2018), allowing users to freely download and use the model for research purposes. FLAN [86] is also open source, though more recently released, in 2022. Contrastive language-image pretraining, DeepMoji, Alexa, and GPT-3.5 are accessible via an application programming interface but are not open-source models, limiting data transparency and the reliability of availability required for most research applications.

Figure 2. Trends and associated recommendations for ensuring safe and effective integration of large language models into suicide prevention and self-harm research. AI: artificial intelligence.

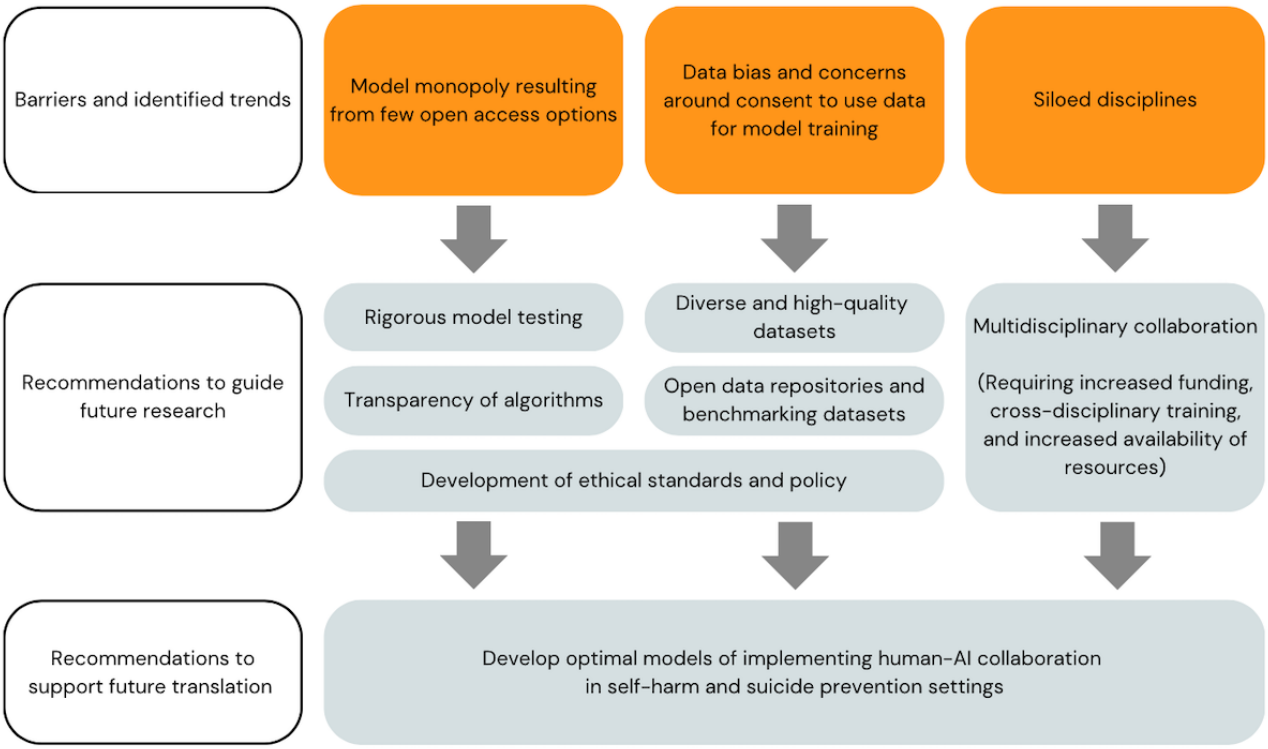

In 33 (77%) of the 43 studies, the primary purpose of LLMs was contextual understanding of language, while 9 (21%) studies extended upon this by using the capacity of LLMs to generate natural language responses to prompts (all these studies were published in 2023). The study by Badian et al [47] also used an LLM for image interpretation. Applications that focused on contextual language understanding were predominantly used for the identification, detection, or prediction of suicide risk. Generative applications also focused on prediction-based tasks [43,44,83] but extended to the evaluation of suicide risk [58,61,66], the identification of circumstances preceding suicide [85], information retrieval or question-answering systems [46], and the creation of mental health nursing care plans [81]. Of the 9 generative applications, 7 (78%) used the text-based ChatGPT user interface, while the remaining 2 (22%) used text-based interfaces for an educational BERT model [46] and a data-secure FLAN model [85]. The prevalence of GPT use in generative applications stands in contrast to the dominance of the use of the BERT model across LLM-based research to date more generally. This hints at a potential shift away from studies primarily focused on BERT. Whether this trend reflects a broader shift in model preference or was influenced by factors such as availability, the ease of use, or specific application requirements was not clear from the included studies.

Data Sources

The majority of the included studies (31/43, 72%) used data from user posts on social media platforms for LLM training, validation, testing, and deployment to answer a research question. Reddit was the most commonly used platform from which data were derived (18/43, 42% studies), followed by X (formerly Twitter; 13/43, 30% studies). User posts from other platforms were also used [51,52,82,84]. Meta's Facebook was used only in the study by Badian et al [47], in which only images, not text data from users, were used. Of the 31 studies using social media data, only the study by Badian et al [47] sought user or poster consent for the use of their data. Of the 12 studies that did not use social media data, 3 (25%) used data from crisis counseling apps or services [48,59,73]; 6 (50%) used a heterogeneous collection of data from the National Violent Death Report System [80,85], educational or academic documents [46,64], or participant vignettes [66,81]; 1 (8%) used research participant data gathered during the course of the study [70]; 1 (8%) did not use collected data but applied prompt inputs to ChatGPT to evaluate responses [61]; and 1 (8%) proposed an LLM-moderated train safety device, designed to initiate braking and deploy an inflatable safety cell in front of the train after voice activation by the train driver [77].

XSL•FO
RenderX

Identified Objectives

Studies were grouped by 4 main objectives: prediction, identification or classification, support, and education or training. The majority of the studies (33/43, 77%) identified in this review applied LLM models to identification or classification tasks, with prediction applications the next most common (6/43, 14%). Studies that focused on the problems of identification or classification sought to identify content indicative of suicidal distress from text-based data, such as in Reddit posts [49,56-58,60,63,68,71,76], in X (formerly Twitter) posts [45,53,69,72,78], or in crisis or helpline conversations [48]. Additional uses included identifying precipitating events to suicide from death investigation narratives [80] and identifying self-harm from social media posts [54]. Prediction-focused studies predominantly used LLMs with data from Reddit [50,65,83], other social media [84], images posted on social media [47], crisis counselling data [59], or clinical data [50] to predict suicide risk.

The remaining model applications (4/43, 9%) assessed support [61], education [46], and "other" (ie, the generation of mental health care plans and means restriction) [77,81]. Of these 4 studies, 3 (75%) involved the use of generative AI. Regarding education, a question-and-answer interface was developed to generate specific information requested by individuals at risk and their families. The model underlying this interface drew on a corpus of >300 suicide-specific documents curated by clinicians [46]. Of the 2 studies that involved ChatGPT, 1 (50%) aimed to assess the safety of publicly available conversational agents by prompting them with sequential Patient Health Questionnaire-9 items to examine chatbot responses to a patient simulation indicating escalating suicide risk [61], while 1 (50%) asked ChatGPT to generate mental health nurse care plans based on vignettes about a fictitious person self-harming [81]. Overall, study objectives focused on suicide and self-harm detection, while the development of generative LLM technologies has facilitated more recent LLM use in care, support, and education.

Clinical Applications

We examined trends in proposed clinical applications of LLMs to suicide or self-harm prevention. Studies were expected to have provided sufficient depth of discussion regarding ways in which the research could be translated into real-world settings. Brief general statements or single-sentence phrases about clinical applications were not considered sufficient for inclusion in addressing this research question.

Clinical applications were discussed in 26 (60%) of the 43 studies. Of the 17 publications that did not discuss clinical applications, 13 (76%) were from the field of computer science. The identified applications included improved detection of suicidality (13/26, 50%), use as a clinical assistance tool (10/26, 38%), improved accessibility of services (3/26, 12%), improved services (3/26, 12%), assisting the development of policy (3/26, 12%), and use as a training tool (2/26, 8%). Specifically, studies discussing the ability of LLMs to improve the detection of suicidality mentioned the creation of automated systems that could be applied to language data (eg, social media posts) to detect suicidal ideation for early intervention [48,55,57]. Studies focusing on the potential applications of LLMs as clinical assistance tools discussed possible use cases where LLMs could aid clinicians in evaluating a client's level of suicide risk [48,57], supporting diagnosis and treatment [50,67], providing a second opinion [66], or predicting a score on a mental health scale [83]. Future-oriented clinical assistance tool applications included AI-enabled avatars that could be accessed remotely, could deliver therapeutic services, could conduct simple examinations, provide advice, or recommend referrals for additional care [66]. Applications related to service improvement included LLM integration to improve crisis counseling services [48,73] and improving existing mental health chatbots [54]. Applications to training included enhancing training, clinical procedures, and best practices among mental health and medical professionals [73,79]. Improved policy was noted as of clinical relevance [73,79] because there may be a lack of policy safeguarding people considered vulnerable and the use of LLMs [81]. Cost-effectiveness and increasing the quality of data annotation were also noted [58], as was use in public health surveillance, potentially allowing practitioners to track the prevalence of infrequent conditions [85].

Ethical Considerations

We also explored trends in the nature and type of ethical considerations reported. To be considered as having meaningfully reported on ethical issues, studies needed to include discussion of the ethical considerations and their possible implications. Cursory mention of potential ethical issues without discussion of their implications was not considered sufficient (eg, mention of LLMs as being a "black box" without any further discussion of the implications [47,59] was not considered sufficient).

Ethical considerations were discussed in 14 (33%) of the 43 studies. Of the 29 studies not presenting ethical perspectives, 18 (62%) were published in computer engineering journals, where greater emphasis was given to the technical aspects of model development, training, and validation. Privacy (9/14, 64%) was the most common ethical consideration discussed, followed by bias (5/14, 36%), hallucinations (the tendency of LLMs to sometimes generate false responses; 3/14, 21%), the "black box" nature of LLMs (3/14, 21%), possible threats to the client-clinician relationship (eg, clinicians replaced with digital agents; 2/14, 14%), and false positives or negatives in classification or identification (2/14, 14%).

Specifically, studies that discussed ethical issues regarding privacy touched on concerns about confidentiality and consent to use publicly available data [55,75,82,83], with researchers highlighting the need to value the privacy, security, and anonymity of the original posters and end users [57,66]. Some researchers presented ethical concerns related to publicly traded companies inferring or collecting sensitive health information about their users and acting on it or sharing it without explicit consent [50,58]. Nontransparent data collection and inference processes currently used by social media platforms were highlighted as a growing area of concern [74]. Hallucinations were identified as impacting the safety and reliability of LLM applications [58,85]. Some of the studies (2/14, 14%) raised concerns that models trained on vast amounts of data containing social biases or prejudices may perpetuate stigma toward suicide

XSL•FO
RenderX

or self-harm [58,83]. Related concerns were raised regarding the generalizability of LLMs to populations whose racial, ethnic, or cultural demographics differ from those present in training data [48,75]. The study by Levkovich and Elyoseph [66] noted that these data biases or the lack of diversity could lead to erroneous predictions that may exacerbate existing disparities in suicide prevention strategies. Furthermore, 3 (21%) of the 14 studies noted a lack of transparency, resulting from the opaque functioning of the underlying algorithms (termed "black box") [59,66], with the study by Woodnutt et al [81] noting a lack of governance in this regard.

Of the 14 studies, 2 (14%) raised concerns about LLMs threatening the client-clinician relationship, potentially leading to human-centered clinical interactions being replaced by digital alternatives. Karapetian et al [64] discussed the issue of LLM agents lacking human attributes, such as empathy for suicidal distress, and Malhotra et al [67] noted that the integration of LLMs into clinical care may lead to a sense of distancing of the clinician from the individual, potentially fostering feelings of invalidation or insignificance and exacerbating suicidal thoughts or self-harm behaviors. False negatives (when suicidality goes undetected) and false positives (when suicidality is incorrectly flagged as being present) were noted as concerns [74,75] as psychological harm can result (eg, resulting in missed opportunities to intervene with someone at risk or in unnecessary mental health evaluations for someone who is not at risk) [74]. Safety was also noted as a concern because conversational AI may advance at a pace that outstrips associated safety measures [61]. Relatedly, issues of clinical responsibility were highlighted, particularly regarding the use of LLMs in aiding the generation of mental health care plans, as some authors believed that this could leave mental health practitioners legally vulnerable [81]. More generally, studies emphasized that guidelines for safe development, auditing, and regulation are very much needed to address ethical risks in this area of research [83].

Discussion

The primary aim of this scoping review was to summarize and characterize emerging trends in the application of LLMs in the field of suicide prevention and self-harm research. This review maps the study characteristics and key methodological components of the included studies (n=43) and further addresses the secondary aims of examining the clinical applications and ethical considerations proposed in the included studies.

Study Characteristics

The studies included in this analysis exhibited a notable divergence in disciplinary focus, with approximately equal representation from computer engineering (23/43, 53%) and health-related (18/43, 42%) fields. The bifurcation of domain expertise in this emerging field has important implications for the safety, effectiveness, and real-world impact of the research outputs, particularly considering the differences in technical training, publication norms, and approach to validation and evaluation used across the different fields; for example, only recently has there been a call to integrate education in the technicalities and ethics of AI into mental health training programs, and many mental health researchers may not yet receive technical training in AI to be able to develop and apply LLMs for clinical research purposes. In parallel, while computer science researchers have significant technical training to enable the development of innovative AI models and access to industry partnerships that can support the scalability of new AI-based tools, the real-world usefulness of these innovations hinges on a deep understanding of the ethical and clinical context in which they are to be applied, the facilitators and barriers to their use, and the needs and priorities of end users—areas of expertise where clinical training is often paramount. Recognizing and addressing the complementary strengths and gaps in these divergent disciplines will be crucial for ensuring that LLM innovations in suicide prevention are grounded in technical excellence and clinical acumen.

Similarly, the dissemination of research from computer science often outpaces the dissemination of research from the field of mental health owing to the stronger focus on conference proceedings with expeditious publication timelines. The accelerated dissemination of LLM-based suicide prevention research originating from computer science fields may mean that practices focused on innovation—a crucial benchmark of research impact in computer science (eg, via patents) [87]—come to dominate early research advancements in this area [88]. Although innovation is crucial for achieving technological advancements, this innovation often comes at the expense of comprehensive clinical validation [89], and there has been growing emphasis on the need for frameworks for validating new AI tools in health research [90].

Recognizing and navigating these disciplinary disparities is essential. Encouraging multidisciplinary collaboration among experts from diverse backgrounds is likely to be critical for bolstering the quality, safety, and impact of research on LLMs for suicide prevention. This could be done by fostering the development of agile methods for disseminating validation research (such as through cross-disciplinary repositories and benchmarking datasets), establishing sound ethical and safety frameworks that cut across key disciplines (such as living guidelines), and facilitating ongoing education and cross-disciplinary training for researchers across both fields. Crucially, prioritizing and supporting studies led by multidisciplinary teams that rigorously assess safety and effectiveness is imperative for ensuring that research efforts lead to broader societal benefits.

Base LLMs Used

There was minimal variation in the base LLM model applied in the identified studies. Most of the studies (35/43, 81%) used either the base BERT model or some variation of a trained or fine-tuned BERT model [91]. The relatively early release date (2018), open-source availability, compact size of this model compared to subsequent models such as GPT, and its adaptability through fine-tuning for diverse applications are all factors assisting the widespread use of BERT in this research area. While open access is critical for facilitating replication and democratizing access to this modern technology, an important consideration is that open-access LLMs trained on generalized datasets might inadvertently perpetuate unwanted biases; for example, BERT has been shown to exhibit

XSL•FO
RenderX

stereotypical biases in areas such as gender, profession, race, and religion, with this bias not related to the size of pretraining corpora but likely due to the nature of the training data [92]. In the same study, GPT-2 was found to have less bias, hypothesized in part to be due to antistereotypical associations present in the training data. LLMs are "stochastic parrots" [93] in their generative responses, their outcomes dependent on the quality of the input training data. Therefore, the testing, application, and comparison of multiple models is important to detect and mitigate model biases permeating an entire discipline. Promoting the use of a diverse range of models to assess performance variability [58,83] and uncover potential biases [92] could enhance the strength of research outcomes in suicide prevention. In addition, there is a pressing need to develop specialized open-access models (such as those being developed in medicine) [94,95] for mental health and suicide prevention. These models should draw upon the expertise of mental health professionals, cross-disciplinary researchers, ethicists, and individuals with lived experience. Establishing standards and policies that ensure the transparent reporting of model training data and inherent potential biases is also imperative to enable the critical evaluation and validation of model suitability [7], particularly for self-harm and suicide prevention contexts.

Data Sources

Most of the studies (31/43, 72%) used data from social media platforms such as X (formerly Twitter) and Reddit to train LLMs. Although these are rich data sources, a reliance on social media data as a form of training data may lead to limitations in the usefulness of LLMs. To protect their users, social media platforms often implement policies to detect, suppress, or remove content related to suicide or self-harm. Therefore, using social media data for LLM training may result in an underrepresentation of critical data that would enable LLMs to effectively recognize and respond to expressions of suicidal distress. In addition, algorithms embedded within social media platforms are designed to promote engaging content. As a result, these algorithms often promote user posts that are biased toward certain viewpoints and controversial or sensational topics, while simultaneously minimizing diversity of perspective. Accordingly, the algorithms in social media platforms may drive artificial patterns in the nature and frequency of certain types of user post data that do not reflect real-world interpersonal interactions.

It is also crucial to evaluate whether the profiles and behavior of individuals who post on social media platforms align well with those of the potential end users of LLM-based suicide prevention interventions; for example, while Reddit's user anonymity may constitute a perceived benefit to users, evidence indicates that users engage in differing behavior when posting anonymously compared to when their identity is known [96]. To mitigate this, it is imperative to curate training datasets that are diverse and balanced [97], as well as representative of the population that the LLM is intended to serve [7]. The inclusion of input from those with lived experience is important during the development of these datasets because these experiences can provide valid human-generated data for model training. Furthermore, lived experience can provide a valuable perspective in reviewing LLM-generated data, enabling key human-in-the-loop validation of otherwise unavailable data for model training (eg, crisis care transcripts). In addition, actively seeking out content, including positive interactions, supportive conversations, and safe discussions around self-harm and suicide prevention, may help enhance the effectiveness of LLM-based suicide prevention interventions.

Identified Objectives

The studies included in this review predominantly deployed LLMs for the identification, detection, classification, and prediction of suicide or self-harm risk. These areas represent a narrow band of potential use, and despite research focused in this area, there remains a lack of solutions that implement these models in clinical settings [55]. One reason for this is that computer scientists, who are at the forefront of technological innovation in this area, do not generally have in-depth knowledge of clinical processes and workflows used in the mental health field. This limits their ability to envision the broader applicability of LLMs in mental health settings and results in a tendency to gravitate toward more familiar applications, such as classification and regression. Similarly, suicide prevention researchers often do not possess the technical expertise required to adapt LLMs for their specific needs. Consequently, multidisciplinary collaborations are critical to the development of sound research programs that combine the clinical and technical expertise necessary for high-caliber research. As this research advances and models continue to improve, those focused on clinical areas may be deployed in various contexts. However, it has been noted that this may be insufficient and may not reduce suicide attempts or deaths unless the treatment needs of people identified as being at risk are met [98]. Hence, it is imperative to explore innovative solutions in clinical care, consultation, and therapy.

This review highlighted a small number of recent applications that move beyond identification and prediction by leveraging the generative features of LLMs. These applications included assessing whether LLMs could provide access to suicide-related educational information [46], whether they could detect escalating suicide risk and the need for human intervention [61], and whether they could assist in generating mental health nurse care plans for individuals who self-harm [81]. These studies showed that, at present, LLMs can perform well on specific elements of these roles, including providing immediate access to accurate information relevant to suicide prevention [46], using health theory frameworks to construct detailed care plans with tangible goals for clients [81], and generating training aids for mental health providers [81]. However, these studies also highlighted areas where LLMs currently fall short. Specifically, they showed that LLMs did not effectively detect and respond to signs of escalating clinical risk by referring into care [61]. Moreover, concerns were noted regarding how the accountability of health care providers may be impacted by the use of AI-generated mental health care plans [81]. These examples highlight the critical role of human expertise in the deployment of LLMs in suicide prevention contexts, and an important area for future research will be to determine optimal approaches to implementing human-AI collaboration in suicide prevention settings, including through the use of copilot or

XSL•FO
RenderX

human-in-the-loop systems that are beginning to be explored in other areas [89,99].

Clinical Applications

The secondary aim of this scoping review was to consider the clinical applications and ethical considerations of LLMs in suicide prevention practice. Among the studies describing clinical applications of LLMs, "enhanced detection of suicidality" and use as a "clinical assistance tool" were the most commonly presented clinical applications. The ability to identify and predict suicide has remained at near-chance levels for many years [24], and improvement upon this with the help of novel technologies such as LLMs can provide a valuable contribution to the field. However, accurate risk detection alone is insufficient and must be paired with effective and scalable interventions [24]. The use of LLMs as clinical assistance tools can support clinical professionals by aiding with a myriad of tasks [48,57,66,70]. The conception and design of these applications should start with the end point in mind to best plan for the translation of research results into implementation in a clinical setting. Critically, researchers should engage clinicians, health professionals, lived experience representatives, and other relevant stakeholders early in the research project development to ensure that the research can meet the needs of end users. In the case of lived experience representatives, it is imperative to understand how they, as representatives of end users, interact with the various potential applications of LLMs to ensure safety and effectiveness. Clinical applications described in the studies include assisting clinical decision-making by providing a second opinion [66], functioning as a clinician copilot [48], serving as a mental health triage tool [57], and supporting therapy delivery [70]. However, these applications were limited to hypothesized future uses [48,57,66] or pilot studies [70]. More investment is required to develop implementable clinical assistance tools, particularly in supporting the integration of multidisciplinary teams to collaborate on areas such as clinical training and support as well as identifying novel use cases with the potential to positively impact suicide prevention practice. Importantly, LLMs should not replace clinical judgment but rather supplement it, aiding clinicians and health professionals to make more informed decisions [66] by enhancing access to information (eg, suggesting differential diagnoses, proposing alternate hypotheses, or collating scattered information into more coherent forms to facilitate human interpretation). LLM applications should not exceed this human-in-the-loop support-based role until their safety and effectiveness can be clearly demonstrated.

It is also important to consider the context in which an LLM-based application will be implemented, acknowledging that its success may be limited by existing clinical workflows [100] and the overall capacity of the health system to integrate new tools and provide relevant training [101]. As with all new technological advancements in care, LLM integration into health care practices requires strategic integration into existing—often outdated and practically constrained—care models if we are to realize their full benefit. Forward thinking and responsive policy is required, led by governments that engage with stakeholders to map policy and standards to provide for the harmonious integration of LLM clinical support tools, given the accelerating use and potential of AI.

Research indicates that a majority of individuals do not engage with formal mental health services before suicide, potentially due to stigma or fear of judgment [39]. While none of the identified studies deployed LLMs to directly support or intervene with individuals in distress, the ease of access that a language-based user interface offers may help some overcome stigma-related barriers to engaging with formal mental health services. Indeed, preliminary evidence for the acceptability of generative LLMs in support contexts has been demonstrated; for example, evidence shows that a large majority of individuals (78%) would be willing to use ChatGPT for self-diagnosis or to aid in self-managing their symptoms [102]. Some individuals also indicate a willingness to disclose information to nonhuman agents citing reduced fear, less need for impression management, and greater ease in expressing the severity of their emotions [103].

This review found evidence that LLMs can deliver accurate information relevant to suicide prevention [46] and support clinical needs in related contexts such as mental health evaluations, therapeutic consultations, and patient education [104]. Therefore, LLMs hold the potential to offer individuals quick, empathic responses to mental health queries [105], while avoiding potential or perceived judgment and stigma, at little to no cost, available 24/7. Nevertheless, the efficacy of LLMs in providing precise, evidence-based support to individuals in suicidal distress has not been empirically validated. Furthermore, LLMs currently lack the ability to collect information vital for diagnosing and managing a patient's health condition [106]; for example, ChatGPT's performance has been found to deteriorate as the complexity of clinical cases increases [105], and research identified in this review demonstrates that GPT-powered conversational agents and chatbots can be dangerously slow to escalate high-risk mental health situations for human clinical intervention [61]. While there may be a safe use threshold at which LLMs are beneficial for information and psychoeducational purposes, they are not advanced enough to be used as stand-alone therapeutic devices. Looking forward, models of LLM use in crisis support contexts are most likely to be those that are driven by clinical oversight, with sufficient guardrails in place to trigger referrals when clinically indicated.

The incorporation of LLMs into training and educational contexts presents a promising opportunity for research and development under clinical oversight. Educational resources, such as the question-answering system in an included study that was aimed at providing suicide prevention information [46], provide an illustrative example. This mode of education is highly scalable and may provide a cost-effective means for deploying training programs. Such educational tools would be particularly useful in large-scale, multifaceted, community-wide interventions such as the European Alliance Against Depression [107] and Lifespan [108] models, both of which have community awareness and training as key components. Another study identified in this review [81] suggests that LLMs such as ChatGPT could support training by assisting less experienced mental health practitioners in creating care plans, brainstorming ideas, or pinpointing relevant aspects of patient presentation.

XSL•FO
RenderX

However, it was noted that these generated plans, while appearing credible to laypersons, have been found to contain significant errors and ethical issues upon professional evaluation, highlighting the indispensable role of clinical oversight [81]. LLMs could also facilitate the creation of diverse case scenarios for health care providers to hone their risk assessment and communication skills. By role-playing case scenarios, LLMs can offer real-time, interactive training experiences for crisis call center staff, counselors, and suicide prevention gatekeepers, enhancing their confidence in discussing suicide and building clinical competencies. In addition, the use of LLMs to converse in a role-play scenario has the potential to improve initial training outcomes [109]. Critically, suicide prevention gatekeeper training outcomes are shown to diminish over time [110]; however, engagement with a role-playing LLM could result in improved retention of training outcomes and more individuals at risk identified, approached, and referred for help.

These training and education use cases should be moderated by human oversight and iterated with human feedback for continuous and rapid improvement. Future research should aim to develop models trained in this domain that can demonstrate reliability and effectiveness in providing training or educational outputs. The development of the underlying model in this use case has the potential to translate to alternative clinical applications while providing researchers with a greater understanding of the guardrails and parameters of operation that are required for safe deployment across suicide prevention applications (Figure 3).

Figure 3. Potential future applications of large language models in the field of self-harm and suicide prevention. AI: artificial intelligence.

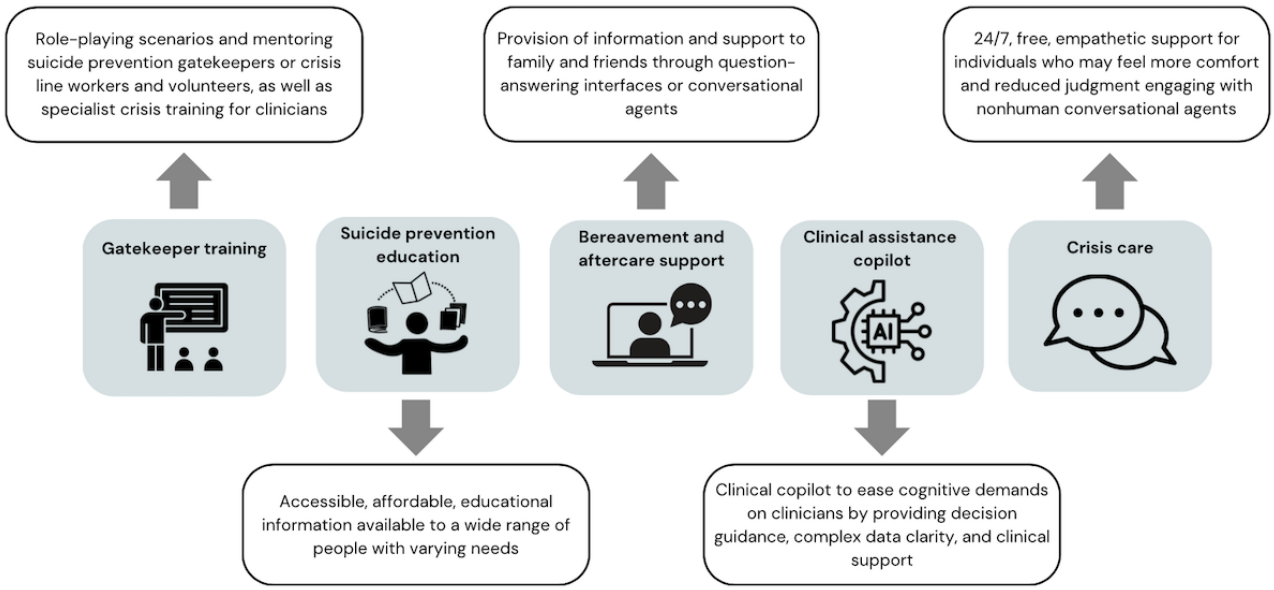

Ethical Considerations

Growing emphasis has been placed on the need for a clear and comprehensive ethical framework for integrating LLMs into mental health research and practice [111]. Hence, another aim of this review was to understand the ethical issues involved in the integration of LLMs into suicide prevention and self-harm research. Ethical considerations raised in the included studies focused on privacy and autonomy, bias, transparency and trust, and potential adverse impacts on the therapeutic relationship.

The integration of LLMs and social media data was noted as presenting a double-edged sword in terms of balancing privacy and confidentiality concerning the use of publicly available data with opportunities for suicide detection [55,75,82,83]. On the one hand, the vast repository of publicly available language-based data on social media platforms provides the potential for real time identification of individuals at risk. On the other hand, while such methods can be lifesaving, they bypass the consent traditionally required in research and clinical practice. User or poster consent was sought in only 1 (3%) of the 31 identified studies that used social media data for training their models. Users of social media sites are often not afforded the opportunity to opt out of such surveillance or analysis, nor are they made aware of the potential adverse consequences associated with having their data used in this manner (eg, the risk of data privacy breaches, as has happened with some LLMs) [112]. This contravenes the fundamental human right to privacy, recently underscored by European legislation prohibiting certain data-gathering practices related to health information by the algorithms of Facebook and other large social media companies [74,113]. This legislative approach is aligned with suggested practices to mitigate ethical concerns around privacy and confidentiality through the establishment of legal boundaries [82]. The use of social media and other data to train LLMs requires deep discussion to discern the regional, social, cultural, and temporal factors (among others) that influence ethical decisions around safe use of this data. An important ethical concern for the field of LLM-based suicide prevention research is how to balance the potential safety of users with their fundamental right to privacy. Inclusion of the lived experience perspective in this discussion is crucial if we are to arrive at a workable solution. Real-world clinical application of LLMs focused on detecting individuals at heightened suicide risk also raises critical ethical issues regarding the implications of false positives, if acted on by authorities, and the potential threats this poses to personal autonomy.

XSL•FO
RenderX

Bias is another important ethical consideration that was noted in the identified studies. In a recent review of AI algorithms applied to mental health, Straw Callison-Burch [114] found that significant biases exist with respect to religion, race, gender, nationality, sexuality, and age. These biases result from the expression of data (the manner in which the original data are presented) [115], the analysis of data (influenced by the contextual prelearnings of the model), and the interpretation of results (human annotation influenced by unconscious bias may produce model bias after training on that annotated data set) [114]. The presence of bias at any stage of model development risks creating tools that disadvantage certain groups of individuals [114]. Rigorous multimodel testing, comparison, and validation are required to isolate and mitigate inherent biases and thereby ensure unbiased performance across diverse populations. Understanding the specific factors that contribute to the effectiveness or limitations of a model also allows researchers to make informed decisions about how to optimize future training processes. This may involve collecting more diverse and representative data, implementing better preprocessing techniques, or fine-tuning model parameters.

Ethical issues of transparency and trust were also discussed in several of the included studies. In contrast to traditional machine learning algorithms, LLMs are opaque "black box" architectures with complex internal structures, which makes it difficult to understand and explain their decisions [67]. This is particularly challenging when hallucinations occur, with little recourse for discovering the source of the error. LLMs do not offer their "reasoning" unless prompted, and when prompted, they often fail to articulate how the provided information was retrieved, vetted, curated, or prepared. This lack of transparency is a fundamental flaw that hinders our ability to understand how these models operate, undermining their trustworthiness in health care applications [67]. Efforts are being made to distill this unknown "thinking" with the application of explainable AI methods. Of the 43 included studies, 2 (5%) [63,67] applied such methods in the form of Shapley additive explanations [116], local interpretable model-agnostic explanations [117], and Topic BERT [118]. The results indicate that these techniques can provide reasonable explainability for both short and long user-generated text and provide insights about data quality issues in training datasets [67]. However, current methods in this area fall somewhat short of delivering the transparency needed to establish total trust, limiting LLM deployment to contexts that restrict or protect against the potential for model hallucinations. Ongoing research into explainable AI techniques is critical in bridging this transparency gap to facilitate trust and enable real-world implementation.

Of the 43 included studies, 2 (5%) noted the client-clinician relationship as an ethical consideration: one in relation to LLM agents lacking human attributes [61] and the other noting a sense of distancing of the clinician from the individual, potentially fostering feelings of invalidation or insignificance [81]. At present, LLMs lack the authenticity and relational aspects required for modern mental health care [81]. Although LLMs can simulate empathy, it is important to consider the specific care needs of individuals experiencing suicidal ideation. If deployed incorrectly, LLM technologies have the potential to incite harm [119], requiring careful implementation to maintain safe practices. When consideration is given to the development of LLMs for use in suicide prevention, sound policy needs to be written to safeguard care recipients, reflecting the complexity of the relationships they have with traditional care providers [81]. In addition, stringent testing and oversight are required during the development and subsequent application of LLMs in the mental health domain [61]. While acknowledging the important ongoing role of human-led therapies, there is a clear need for further research into developing AI systems that can effectively integrate professionals into the loop. This would allow LLMs to serve as an aid in clinical settings, complementing rather than supplanting human practitioners.

Limitations

Some limitations must be kept in mind when interpreting the findings of this scoping review. First, the categorization of machine learning technologies in the existing literature was not always clear, sometimes rendering it challenging to discern whether a study used an LLM or other machine learning model. It is possible that some relevant articles may have been excluded from the review. Second, due to the cross-disciplinary nature of the field and resulting publications, reporting detail varied between disciplines. Health-related publications provided more clinical and ethics-related information, whereas computer engineering publications were more likely to provide in-depth detail about the LLM model and training process. This made the synthesis and comparison of certain study attributes challenging; for example, data preparation and input methods were not sufficiently covered in the studies published in health-related journals to allow meaningful synthesis. Third, all data were extracted by a single author. Although data extraction was piloted by 2 authors to ensure consistency of approach at the outset, some data that were extracted were qualitative, potentially resulting in researcher bias; for example, data extraction related to ethical considerations such as privacy could be subject to bias due to researcher beliefs. Fourth, this review was conducted in an area of research that is currently experiencing significant growth and development; therefore, it only provides a time-stamped representation of the field. Finally, this review excluded studies using EHR data, given that EHR data may not be representative of individuals at risk of suicide because a large proportion of individuals experiencing suicidality are not in contact with formal mental health services [120]. However, excluding EHR studies reduced the number of included studies that focused on predicting suicide risk.

Conclusions

LLMs represent a promising avenue for enhancing the scalability, accessibility, affordability, and personalization of tools in the field of suicide prevention and self-harm research; however, collaboration between computer science and mental health experts is essential to leverage the strengths of both disciplines effectively. This review identified a strong bias toward the use of BERT, with the potential for inherent biases indicating a pressing need for rigorous model comparison and testing, alongside the curation of diverse training datasets to mitigate model bias effectively. The growing use of generative

XSL•FO
RenderX

LLM applications indicates promise for transformative applications in care, support, training, and education, such as improved crisis care and gatekeeper training methods, clinician copilot models, and improved educational practices. However, clinical accountability remains crucial to ensure the responsible use of LLMs in applications targeting suicide prevention. The identified ethical considerations underscore the need for clear governance via the establishment of policies and standards to guide the integration of LLMs into clinical use. LLMs have significant implications for self-harm and suicide prevention, underscoring the need to support continued research and development in this domain.

Acknowledgments

This research is funded by Australia's Medical Research Future Fund (1200195). HC and AW received funding support from National Health and Medical Research Council Investigator Grants (HC: 1155614; AW: 2017521). The funders had no role in the production of this review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 500 KB - jmir_v27i1e63126_app1.pdf]

Multimedia Appendix 2
Individual database search strings and data extraction template.
[DOCX File , 27 KB - jmir_v27i1e63126_app2.docx]

Multimedia Appendix 3
Table of included publications with extracted data.
[XLSX File (Microsoft Excel File), 41 KB - jmir_v27i1e63126_app3.xlsx]

References

1. Preventing suicide: a global imperative. World Health Organization. 2014. URL: https://www.who.int/publications/i/item/9789241564779 [accessed 2024-04-29]
2. Martinez-Ales G, Hernandez-Calle D, Khauli N, Keyes KM. Why are suicide rates increasing in the United States? Towards a multilevel reimagination of suicide prevention. In: Baca-Garcia E, editor. Behavioral Neurobiology of Suicide and Self Harm. Cham, Switzerland: Springer; 2020:1-23.
3. Melia R, Francis K, Hickey E, Bogue J, Duggan J, O'Sullivan M, et al. Mobile health technology interventions for suicide prevention: systematic review. JMIR Mhealth Uhealth 2020 Jan 15;8(1):e12516 [FREE Full text] [doi: 10.2196/12516] [Medline: 31939744]
4. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv. Preprint posted online June 12, 2017 2017 [FREE Full text] [doi: 10.5860/choice.189890]
5. Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DS, Liu N. Large language models in health care: development, applications, and challenges. Health Care Sci 2023 Aug 24;2(4):255-263 [FREE Full text] [doi: 10.1002/hcs2.61] [Medline: 38939520]
6. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. arXiv. Preprint posted online March 31, 2023 2023 [FREE Full text]
7. Demszky D, Yang D, Yeager DS, Bryan CJ, Clapper M, Chandhok S, et al. Using large language models in psychology. Nat Rev Psychol 2023 Oct 13;5:688-701. [doi: 10.1038/S44159-023-00241-5]
8. Gado S, Kempen R, Lingelbach K, Bipp T. Artificial intelligence in psychology: how can we enable psychology students to accept and use artificial intelligence? Psychol Learn Teach 2021 Aug 12;21(1):37-56. [doi: 10.1177/14757257211037149]
9. He Y, Zhu Z, Zhang Y, Chen Q, Caverlee J. Infusing disease knowledge into BERT for health question answering, medical inference and disease name recognition. arXiv. Preprint posted online October 8, 2020 2020 [FREE Full text] [doi: 10.18653/v1/2020.emnlp-main.372]
10. Li C, Zhang Y, Weng Y, Wang B, Li Z. Natural language processing applications for computer-aided diagnosis in oncology. Diagnostics (Basel) 2023 Jan 12;13(2):286 [FREE Full text] [doi: 10.3390/diagnostics13020286] [Medline: 36673096]
11. Wang J, Zhang G, Wang W, Zhang K, Sheng Y. Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. J Cloud Comp 2021 Jan 15;10(1):1-12. [doi: 10.1186/S13677-020-00218-2]
12. Omoregbe NA, Ndaman IO, Misra S, Abayomi-Alli OO, Damaševičius R. Text messaging-based medical diagnosis using natural language processing and fuzzy logic. J Healthc Eng 2020 Sep 29;2020:1-14. [doi: 10.1155/2020/8839524]

XSL•FO
RenderX

13. Bala S, Keniston A, Burden M. Patient perception of plain-language medical notes generated using artificial intelligence software: pilot mixed-methods study. JMIR Form Res 2020 Jun 05;4(6):e16670 [FREE Full text] [doi: 10.2196/16670] [Medline: 32442148]
14. Peretz G, Taylor CB, Ruzek JI, Jefroykin S, Sadeh-Sharvit S. Machine learning model to predict assignment of therapy homework in behavioral treatments: algorithm development and validation. JMIR Form Res 2023 May 15;7:e45156 [FREE Full text] [doi: 10.2196/45156] [Medline: 37184927]
15. Tanana MJ, Soma CS, Kuo PB, Bertagnolli NM, Dembe A, Pace BT, et al. How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. Behav Res Methods 2021 Oct 22;53(5):2069-2082 [FREE Full text] [doi: 10.3758/s13428-020-01531-z] [Medline: 33754322]
16. Shah RS, Holt F, Hayati SA, Agarwal A, Wang YC, Kraut RE, et al. Modeling motivational interviewing strategies on an online peer-to-peer counseling platform. Proc ACM Hum Comput Interact 2022 Nov 11;6(CSCW2):1-24. [doi: 10.1145/3555640]
17. Allen NB, Nelson BW, Brent D, Auerbach RP. Short-term prediction of suicidal thoughts and behaviors in adolescents: can recent developments in technology and computational science provide a breakthrough? J Affect Disord 2019 May 01;250:163-169 [FREE Full text] [doi: 10.1016/j.jad.2019.03.044] [Medline: 30856493]
18. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018 Aug 27;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
19. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. NPJ Digit Med 2023 Nov 16;6(1):210 [FREE Full text] [doi: 10.1038/s41746-023-00958-w] [Medline: 37973919]
20. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 04;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
21. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
22. Xue J, Zhang B, Zhao Y, Zhang Q, Zheng C, Jiang J, et al. Evaluation of the current state of chatbots for digital health: scoping review. J Med Internet Res 2023 Dec 19;25:e47217 [FREE Full text] [doi: 10.2196/47217] [Medline: 38113097]
23. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie 2019 Aug 20;32(Suppl. 1):64-76. [doi: 10.1159/000501812]
24. Linthicum KP, Schafer KM, Ribeiro JD. Machine learning in suicide science: applications and ethics. Behav Sci Law 2019 May;37(3):214-222. [doi: 10.1002/bsl.2392] [Medline: 30609102]
25. Arowosegbe A, Oyelade T. Application of natural language processing (NLP) in detecting and preventing suicide ideation: a systematic review. Int J Environ Res Public Health 2023 Jan 13;20(2):32 [FREE Full text] [doi: 10.3390/ijerph20021514] [Medline: 36674270]
26. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: a systematic review. J Affect Disord 2019 Feb 15;245:869-884. [doi: 10.1016/j.jad.2018.11.073] [Medline: 30699872]
27. Ji S, Pan S, Li X, Cambria E, Long G, Huang Z. Suicidal ideation detection: a review of machine learning methods and applications. IEEE Trans Comput Soc Syst 2021 Feb;8(1):214-226. [doi: 10.1109/tcss.2020.3021467]
28. D'Hotman D, Loh E. AI enabled suicide prediction tools: a qualitative narrative review. BMJ Health Care Inform 2020 Oct 09;27(3):e100175 [FREE Full text] [doi: 10.1136/bmjhci-2020-100175] [Medline: 33037037]
29. Khan NZ, Javed MA. Use of artificial intelligence-based strategies for assessing suicidal behavior and mental illness: a literature review. Cureus 2022 Jul;14(7):e27225 [FREE Full text] [doi: 10.7759/cureus.27225] [Medline: 36035036]
30. Lejeune A, Le Glaz A, Perron P, Sebti J, Baca-Garcia E, Walter M, et al. Artificial intelligence and suicide prevention: a systematic review. Eur Psychiatry 2022 Feb 15;65(1):1-22 [FREE Full text] [doi: 10.1192/j.eurpsy.2022.8] [Medline: 35166203]
31. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. Int J Environ Res Public Health 2020 Aug 15;17(16):5929 [FREE Full text] [doi: 10.3390/ijerph17165929] [Medline: 32824149]
32. Ophir Y, Tikochinski R, Brunstein Klomek A, Reichart R. The Hitchhiker's guide to computational linguistics in suicide prevention. Clin Psychol Sci 2021 Jun 23;10(2):212-235. [doi: 10.1177/21677026211022013]
33. Pollock D, Peters MD, Khalil H, McInerney P, Alexander L, Tricco AC, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 01;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
34. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2020.

XSL•FO
RenderX

35. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
36. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Implement 2021 Mar;19(1):3-10. [doi: 10.1097/XEB.0000000000000277] [Medline: 33570328]
37. Holmes G, Whitton A, Tang B. Leveraging language based AI to improve suicide prevention: a scoping review. Open Science Framework. 2023 Dec 05. URL: https://osf.io/nckq7 [accessed 2023-12-05]
38. Mohammad SM. Examining citations of natural language processing literature. arXiv. Preprint posted online May 2, 2020 2020 [FREE Full text] [doi: 10.18653/v1/2020.acl-main.464]
39. Tang S, Reily NM, Arena AF, Batterham PJ, Calear AL, Carter GL, et al. People who die by suicide without receiving mental health services: a systematic review. Front Public Health 2021 Jan 18;9:736948 [FREE Full text] [doi: 10.3389/fpubh.2021.736948] [Medline: 35118036]
40. Covidence review management. Covidence. URL: https://www.covidence.org/ [accessed 2024-04-29]
41. Tang S, Werner-Seidler A, Torok M, Mackinnon AJ, Christensen H. The relationship between screen time and mental health in young people: a systematic review of longitudinal studies. Clin Psychol Rev 2021 Jun;86:102021 [FREE Full text] [doi: 10.1016/j.cpr.2021.102021] [Medline: 33798997]
42. Harvey D, Lobban F, Rayson P, Warner A, Jones S. Natural language processing methods and bipolar disorder: scoping review. JMIR Ment Health 2022 Apr 22;9(4):e35928 [FREE Full text] [doi: 10.2196/35928] [Medline: 35451984]
43. Amin MM, Cambria E, Schuller BW. Can ChatGPT's responses boost traditional natural language processing? IEEE Intell Syst 2023 Sep;38(5):5-11. [doi: 10.1109/mis.2023.3305861]
44. Amin MM, Cambria E, Schuller BW. Will affective computing emerge from foundation models and general artificial intelligence? A first evaluation of ChatGPT. IEEE Intell Syst 2023 Mar;38(2):15-23. [doi: 10.1109/mis.2023.3254179]
45. Ananthakrishnan G, Jayaraman AK, Trueman TE, Mitra SA, Murugappan A. Suicidal intention detection in tweets using BERT-based transformers. In: Proceedings of the 2022 International Conference on Computing, Communication, and Intelligent Systems. 2022 Presented at: ICCCIS '22; November 4-5, 2022; Greater Noida, India p. 322-327 URL: https://ieeexplore.ieee.org/document/10037677 [doi: 10.1109/icccis56430.2022.10037677]
46. Ascorbe P, Campos MS, Domínguez C, Heras J, Terroba-Reinares AR. Towards a retrieval augmented generation system for information on suicide prevention. In: Proceedings of the 2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology. 2023 Presented at: IEEECONF '23; December 7-9, 2023; Valletta, Malta p. 143-144 URL: https://ieeexplore.ieee.org/document/10404508 [doi: 10.1109/ieeeconf58974.2023.10404508]
47. Badian Y, Ophir Y, Tikochinski R, Calderon N, Klomek AB, Fruchter E, et al. Social media images can predict suicide risk using interpretable large language-vision models. J Clin Psychiatry 2023 Nov 29;85(1):50516. [doi: 10.4088/JCP.23m14962] [Medline: 38019588]
48. Broadbent M, Medina Grespan M, Axford K, Zhang X, Srikumar V, Kious B, et al. A machine learning approach to identifying suicide risk among text-based crisis counseling encounters. Front Psychiatry 2023 Mar 23;14:1110527 [FREE Full text] [doi: 10.3389/fpsyt.2023.1110527] [Medline: 37032952]
49. Bucur AM, Cosma A, Dinu LP. Early risk detection of pathological gambling, self-harm and depression using BERT. In: Proceedings of the 2021 Conference on Labs of the Evaluation Forum. 2021 Presented at: CLEF '21; September 21-24, 2021; Bucharest, Romania p. 1-12 URL: https://ceur-ws.org/Vol-2936/paper-77.pdf
50. Burkhardt HA, Ding X, Kerbrat A, Comtois KA, Cohen T. From benchmark to bedside: transfer learning from social media to patient-provider text messages for suicide risk prediction. J Am Med Inform Assoc 2023 May 19;30(6):1068-1078 [FREE Full text] [doi: 10.1093/jamia/ocad062] [Medline: 37043748]
51. Cao L, Zhang H, Feng L, Wei Z, Wang X, Li N, et al. Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019 Presented at: EMNLP-IJCNLP '19; November 3-7, 2019; Hong Kong, China p. 1718-1728 URL: https://aclanthology.org/D19-1181.pdf [doi: 10.18653/v1/d19-1181]
52. Cao L, Zhang H, Wang X, Feng L. Learning users inner thoughts and emotion changes for social media based suicide risk detection. IEEE Trans Affective Comput 2023 Apr 1;14(2):1280-1296. [doi: 10.1109/taffc.2021.3116026]
53. de Carvalho VC, Giacon B, Nascimento C, Nogueira BM. Machine learning for suicidal ideation identification on Twitter for the Portuguese language. In: Intelligent Systems. 2020 Oct Presented at: BRACIS '20; October 20-23, 2020; Rio Grande, Brazil p. 536-550 URL: https://link.springer.com/chapter/10.1007/978-3-030-61377-8_37 [doi: 10.1007/978-3-030-61377-8_37]
54. Deshpande S, Warren J. Self-harm detection for mental health chatbots. Stud Health Technol Inform 2021 May 27;281:48-52. [doi: 10.3233/SHTI210118] [Medline: 34042703]
55. Diniz EJ, Fontenele JE, de Oliveira AC, Bastos VH, Teixeira S, Rabêlo RL, et al. Boamente: a natural language processing-based digital phenotyping tool for smart monitoring of suicidal ideation. Healthcare (Basel) 2022 Apr 08;10(4):698 [FREE Full text] [doi: 10.3390/healthcare10040698] [Medline: 35455874]

XSL•FO
RenderX

56. Dobbs MF, McGowan A, Selloni A, Bilgrami Z, Sarac C, Cotter M, et al. Linguistic correlates of suicidal ideation in youth at clinical high-risk for psychosis. Schizophr Res 2023 Sep;259:20-27. [doi: 10.1016/j.schres.2023.03.014] [Medline: 36933977]
57. Garg M. Mental health analysis in social media posts: a survey. Arch Comput Methods Eng 2023 Jan 03;30(3):1819-1842 [FREE Full text] [doi: 10.1007/s11831-022-09863-z] [Medline: 36619138]
58. Ghanadian H, Nejadgholi I, Osman HA. ChatGPT for suicide risk assessment on social media: quantitative evaluation of model performance, potentials and limitations. arXiv. Preprint posted online June 15, 2023 2023 [FREE Full text] [doi: 10.18653/v1/2023.wassa-1.16]
59. Grimland M, Benatov J, Yeshayahu H, Izmaylov D, Segal A, Gal K, et al. Predicting suicide risk in real-time crisis hotline chats integrating machine learning with psychological factors: exploring the black box. Suicide Life Threat Behav 2024 Jun;54(3):416-424. [doi: 10.1111/sltb.13056] [Medline: 38345174]
60. Haque F, Nur RU, Jahan SA, Mahmud Z, Shah FM. A transformer based approach to detect suicidal ideation using pre-trained language models. In: Proceedings of the 23rd International Conference on Computer and Information Technology. 2020 Presented at: ICCIT '20; December 19-21, 2020; Dhaka, Bangladesh p. 1-5 URL: https://ieeexplore.ieee.org/document/9392692 [doi: 10.1109/iccit51783.2020.9392692]
61. Heston TF. Safety of large language models in addressing depression. Cureus 2023 Dec;15(12):e50729 [FREE Full text] [doi: 10.7759/cureus.50729] [Medline: 38111813]
62. Howard D, Maslej MM, Lee J, Ritchie J, Woollard G, French L. Transfer learning for risk classification of social media posts: model evaluation study. J Med Internet Res 2020 May 13;22(5):e15371 [FREE Full text] [doi: 10.2196/15371] [Medline: 32401222]
63. Islam MR, Sakib MK, Prome SA, Wang X, Ulhaq A, Sanin C. Machine learning with explainability for suicide ideation detection from social media data. In: Proceedings of the 10th International Conference on Behavioural and Social Computing. 2023 Presented at: BESC '23; October 30-November 1, 2023; Larnaca, Cyprus p. 1-6 URL: https://ieeexplore.ieee.org/document/10386773 [doi: 10.1109/besc59560.2023.10386773]
64. Karapetian K, Jeon SM, Kwon JW, Suh YK. Supervised relation extraction between suicide-related entities and drugs: development and usability study of an annotated PubMed corpus. J Med Internet Res 2023 Mar 08;25:e41100 [FREE Full text] [doi: 10.2196/41100] [Medline: 36884281]
65. Lee DM, Moradi H. Knowledge-infused dynamic embedding for predicting the severity of suicidal ideation in social media. In: Proceedings of the 2022 International Conference on Computational Science and Computational Intelligence. 2022 Presented at: CSCI '22; December 14-16, 2022; Las Vegas, NV p. 760-766 URL: https://ieeexplore.ieee.org/document/10216651 [doi: 10.1109/csci58124.2022.00139]
66. Levkovich I, Elyoseph Z. Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: vignette study. JMIR Ment Health 2023 Sep 20;10:e51232 [FREE Full text] [doi: 10.2196/51232] [Medline: 37728984]
67. Malhotra A, Jindal R. XAI transformer based approach for interpreting depressed and suicidal user behavior on online social networks. Cogn Syst Res 2024 Mar;84:101186. [doi: 10.1016/j.cogsys.2023.101186]
68. Matero M, Idnani A, Son Y, Giorgi S, Vu HH, Zamani M, et al. Suicide risk assessment with multi-level dual-context language and BERT. In: Proceedings of the 6th Workshop on Computational Linguistics and Clinical Psychology. 2019 Presented at: CLPsych '19; June 6, 2019; Minneapolis, MN p. 39-44 URL: https://aclanthology.org/W19-3005.pdf [doi: 10.18653/v1/w19-3005]
69. Metzler H, Baginski H, Niederkrotenthaler T, Garcia D. Detecting potentially harmful and protective suicide-related content on twitter: machine learning approach. J Med Internet Res 2022 Aug 17;24(8):e34705 [FREE Full text] [doi: 10.2196/34705] [Medline: 35976193]
70. Mezzi R, Yahyaoui A, Krir MW, Boulila W, Koubaa A. Mental health intent recognition for Arabic-speaking patients using the mini international neuropsychiatric interview (MINI) and BERT model. Sensors (Basel) 2022 Jan 23;22(3):846 [FREE Full text] [doi: 10.3390/s22030846] [Medline: 35161594]
71. Naseem U, Khushi M, Kim J, Dunn AG. Hybrid text representation for explainable suicide risk identification on social media. IEEE Trans Comput Soc Syst 2024 Aug;11(4):4663-4672. [doi: 10.1109/tcss.2022.3184984]
72. Ravishankar TN, Kumar AK, Venkatesh J, Prabhu MR, Bhargavi VS, MuthamilSelvan S. Empirical assessment and detection of suicide related posts in twitter using artificial intelligence enabled classification logic. In: Proceedings of the 2023 International Conference on Advances in Computing, Communication and Applied Informatics. 2023 Presented at: ACCAI '23; May 25-26, 2023; Chennai, India p. 1-7 URL: https://ieeexplore.ieee.org/document/10201110 [doi: 10.1109/accai58221.2023.10201110]
73. Salmi S, Mérelle S, Gilissen R, van der Mei R, Bhulai S. Detecting changes in help seeker conversations on a suicide prevention helpline during the COVID- 19 pandemic: in-depth analysis using encoder representations from transformers. BMC Public Health 2022 Mar 18;22(1):530 [FREE Full text] [doi: 10.1186/s12889-022-12926-2] [Medline: 35300638]
74. Sawhney R, Joshi H, Nobles A, Shah RR. Towards emotion- and time-aware classification of tweets to assist human moderation for suicide prevention. Proc Int AAAI Conf Web Soc Media 2021 May 22;15:609-620 [FREE Full text] [doi: 10.1609/icwsm.v15i1.18088] [Medline: 35173997]

XSL•FO
RenderX

75. Schoene AM, Bojanić L, Nghiem MQ, Hunt IM, Ananiadou S, Schoene AM. Classifying suicide-related content and emotions on Twitter using graph convolutional neural networks. IEEE Trans Affective Comput 2023 Jul 1;14(3):1791-1802. [doi: 10.1109/taffc.2022.3221683]
76. Sharma N, Karwasra P. Suicidal text detection on social media for suicide prevention using deep learning models. In: Proceedings of the TENCON 2023 - 2023 IEEE Region 10 Conference. 2023 Presented at: TENCON '23; October 31-November 3, 2023; Chiang Mai, Thailand p. 25-30 URL: https://ieeexplore.ieee.org/document/10322499 [doi: 10.1109/tencon58879.2023.10322499]
77. Sheikh SA, Naidu H. A novel robotics and MEMS artificial intelligence based train safety device. In: Proceedings of the 2nd International Conference on Smart Electronics and Communication. 2021 Presented at: ICOSEC '21; October 7-9, 2021; Trichy, India p. 1-5 URL: https://ieeexplore.ieee.org/abstract/document/9591761 [doi: 10.1109/icosec51865.2021.9591761]
78. Soudi RB, Zaghloul MS, Badawy OM. Framework for suicide detection from Arabic tweets using deep learning. In: Proceedings of the 32nd International Conference on Computer Theory and Applications. 2022 Presented at: ICCTA '22; December 17-19, 2022; Alexandria, Egypt p. 96-103 URL: https://ieeexplore.ieee.org/document/10206145 [doi: 10.1109/iccta58027.2022.10206145]
79. Wang S, Ning H, Huang X, Xiao Y, Zhang M, Yang EF, et al. Public surveillance of social media for suicide using advanced deep learning models in Japan: time series study from 2012 to 2022. J Med Internet Res 2023 Jun 02;25:e47225 [FREE Full text] [doi: 10.2196/47225] [Medline: 37267022]
80. Wang S, Dang Y, Sun Z, Ding Y, Pathak J, Tao C, et al. An NLP approach to identify SDoH-related circumstance and suicide crisis from death investigation narratives. J Am Med Inform Assoc 2023 Jul 19;30(8):1408-1417 [FREE Full text] [doi: 10.1093/jamia/ocad068] [Medline: 37040620]
81. Woodnutt S, Allen C, Snowden J, Flynn M, Hall S, Libberton P, et al. Could artificial intelligence write mental health nursing care plans? J Psychiatr Ment Health Nurs 2024 Feb;31(1):79-86. [doi: 10.1111/jpm.12965] [Medline: 37538021]
82. Wu EL, Wu CY, Lee MB, Chu KC, Huang MS. Development of Internet suicide message identification and the Monitoring-Tracking-Rescuing model in Taiwan. J Affect Disord 2023 Jan 01;320:37-41. [doi: 10.1016/j.jad.2022.09.090] [Medline: 36162682]
83. Xu X, Yao B, Dong Y, Gabriel S, Yu H, Hendler J, et al. Mental-LLM: leveraging large language models for mental health prediction via online text data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2024 Mar 06;8(1):1-32 [FREE Full text] [doi: 10.1145/3643540]
84. Yen S, Chu K, Tsai P. Prediction model of social network suicide ideation by small sample. In: Proceedings of the 22nd International Conference on Information Reuse and Integration for Data Science. 2021 Presented at: IRI '21; August 10-12, 2021; Las Vegas, NV p. 385-389 URL: https://ieeexplore.ieee.org/document/9599140 [doi: 10.1109/iri51335.2021.00060]
85. Zhou W, Prater LC, Goldstein EV, Mooney SJ. Identifying rare circumstances preceding female firearm suicides: validating a large language model approach. JMIR Ment Health 2023 Oct 17;10:e49359 [FREE Full text] [doi: 10.2196/49359] [Medline: 37847549]
86. Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W. Scaling instruction-finetuned language models. J Mach Learn Res 2024;25(70):1-53 [FREE Full text]
87. Koya K, Chowdhury G. Measuring impact of academic research in computer and information science on society. In: Proceedings of the 2020 2nd Asia Pacific Information Technology Conference. 2020 Mar 29 Presented at: APIT '20; January 17-19, 2020; Bali, Indonesia p. 78-85 URL: https://dl.acm.org/doi/10.1145/3379310.3379312 [doi: 10.1145/3379310.3379312]
88. Chubb J, Cowling P, Reed D. Speeding up to keep up: exploring the use of AI in the research process. AI Soc 2022 Oct 15;37(4):1439-1457 [FREE Full text] [doi: 10.1007/s00146-021-01259-0] [Medline: 34667374]
89. Stade EC, Stirman SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. Npj Ment Health Res 2024 Apr 02;3(1):12 [FREE Full text] [doi: 10.1038/s44184-024-00056-z] [Medline: 38609507]
90. Tsopra R, Fernandez X, Luchinat C, Alberghina L, Lehrach H, Vanoni M, et al. A framework for validating AI in precision medicine: considerations from the European ITFoC consortium. BMC Med Inform Decis Mak 2021 Oct 02;21(1):274 [FREE Full text] [doi: 10.1186/s12911-021-01634-3] [Medline: 34600518]
91. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online October 11, 2018 2018 [FREE Full text] [doi: 10.48550/arXiv.1810.04805]
92. Nadeem M, Bethke A, Reddy S. StereoSet: measuring stereotypical bias in pretrained language models. arXiv. Preprint posted online April 20, 2020 2020 [FREE Full text] [doi: 10.18653/v1/2021.acl-long.416]
93. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? ???? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021 Presented at: FAccT '21; March 3-10, 2021; Virtual Event p. 610-623 URL: https://dl.acm.org/doi/10.1145/3442188.3445922 [doi: 10.1145/3442188.3445922]
94. Wu C, Lin W, Zhang X, Zhang Y, Xie W, Wang Y. PMC-LLaMA: toward building open-source language models for medicine. J Am Med Inform Assoc 2024 Sep 01;31(9):1833-1843. [doi: 10.1093/jamia/ocae045] [Medline: 38613821]

XSL•FO
RenderX

95. Toma A, Lawler PR, Ba J, Krishnan RG, Rubin BB, Wang B. Clinical camel: an open-source expert-level medical language model with dialogue-based knowledge encoding. arXiv. Preprint posted online May 19, 2023 2023 [FREE Full text] [doi: 10.48550/arXiv.2305.12031]
96. De Choudhury M, De S. Mental health discourse on reddit: self-disclosure, social support, and anonymity. Proc Int AAAI Conf Web Soc Media 2014 May 16;8(1):71-80. [doi: 10.1609/icwsm.v8i1.14526]
97. Chung NC, Dyer G, Brocki L. Challenges of large language models for mental health counseling. arXiv. Preprint posted online November 23, 2023 2023 [FREE Full text] [doi: 10.48550/arXiv.2311.13857]
98. Kirtley OJ, van Mens K, Hoogendoorn M, Kapur N, de Beurs D. Translating promise into practice: a review of machine learning in suicide research and prevention. Lancet Psychiatry 2022 Mar;9(3):243-252. [doi: 10.1016/S2215-0366(21)00254-6] [Medline: 35183281]
99. Ahmad MA, Yaramis I, Roy TD. Creating trustworthy llms: dealing with hallucinations in healthcare ai. arXiv. Preprint posted online September 26, 2023 2023 [FREE Full text] [doi: 10.48550/arXiv.2311.01463]
100. Laka M, Carter D, Milazzo A, Merlin T. Challenges and opportunities in implementing clinical decision support systems (CDSS) at scale: interviews with Australian policymakers. Health Policy Technol 2022 Sep;11(3):100652. [doi: 10.1016/j.hlpt.2022.100652]
101. Jung K, Kashyap S, Avati A, Harman S, Shaw H, Li R, et al. A framework for making predictive models useful in practice. J Am Med Inform Assoc 2021 Jun 12;28(6):1149-1158 [FREE Full text] [doi: 10.1093/jamia/ocaa318] [Medline: 33355350]
102. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. JMIR Hum Factors 2023 May 17;10:e47564 [FREE Full text] [doi: 10.2196/47564] [Medline: 37195756]
103. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: virtual humans increase willingness to disclose. Comput Human Behav 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
104. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 04;47(1):33 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
105. Dergaa I, Fekih-Romdhane F, Hallit S, Loch AA, Glenn JM, Fessi MS, et al. ChatGPT is not ready yet for use in providing mental health assessment and interventions. Front Psychiatry 2023;14:1277756 [FREE Full text] [doi: 10.3389/fpsyt.2023.1277756] [Medline: 38239905]
106. Acuña Caicedo RW, Gómez Soriano JM, Melgar Sasieta HA. Bootstrapping semi-supervised annotation method for potential suicidal messages. Internet Interv 2022 Apr;28:100519 [FREE Full text] [doi: 10.1016/j.invent.2022.100519] [Medline: 35281704]
107. Hegerl U, Althaus D, Schmidtke A, Niklewski G. The alliance against depression: 2-year evaluation of a community-based intervention to reduce suicidality. Psychol Med 2006 Sep;36(9):1225-1233. [doi: 10.1017/S003329170600780X] [Medline: 16707028]
108. Shand F, Torok M, Cockayne N, Batterham PJ, Calear AL, Mackinnon A, et al. Protocol for a stepped-wedge, cluster randomized controlled trial of the LifeSpan suicide prevention trial in four communities in New South Wales, Australia. Trials 2020 Apr 15;21(1):332 [FREE Full text] [doi: 10.1186/s13063-020-04262-w] [Medline: 32293516]
109. Cross WF, Seaburn D, Gibbs D, Schmeelk-Cone K, White AM, Caine ED. Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills. J Prim Prev 2011 Aug 4;32(3-4):195-211 [FREE Full text] [doi: 10.1007/s10935-011-0250-z] [Medline: 21814869]
110. Holmes G, Clacy A, Hermens DF, Lagopoulos J. The long-term efficacy of suicide prevention gatekeeper training: a systematic review. Arch Suicide Res 2021 Dec 06;25(2):177-207. [doi: 10.1080/13811118.2019.1690608] [Medline: 31809659]
111. McKernan LC, Clayton EW, Walsh CG. Protecting life while preserving liberty: ethical recommendations for suicide prevention with artificial intelligence. Front Psychiatry 2018 Dec 3;9:650 [FREE Full text] [doi: 10.3389/fpsyt.2018.00650] [Medline: 30559686]
112. Jang H. A South Korean Chatbot shows just how sloppy tech companies can be with user data. The Slate Group LLC. URL: https://slate.com/technology/2021/04/scatterlab-lee-luda-chatbot-kakaotalk-ai-privacy.html [accessed 2024-04-29]
113. The digital services act (DSA) - regulation (EU) 2022/2065. Digital Services Act. URL: https://tinyurl.com/2n2teb4u [accessed 2024-04-29]
114. Straw I, Callison-Burch C. Artificial Intelligence in mental health and the biases of language based models. PLoS One 2020 Dec 17;15(12):e0240376 [FREE Full text] [doi: 10.1371/journal.pone.0240376] [Medline: 33332380]
115. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
116. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://dl.acm.org/doi/10.5555/3295222.3295230
117. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD

XSL•FO
RenderX

'16; August 13-17, 2016; San Francisco, CA p. 1135-1144 URL: https://dl.acm.org/doi/10.1145/2939672.2939778 [doi: 10.1145/2939672.2939778]
118. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online March 11, 2022 2022 [FREE Full text] [doi: 10.1002/asi.21166]
119. Arora A, Arora A. The promise of large language models in health care. Lancet 2023 Feb 25;401(10377):641. [doi: 10.1016/S0140-6736(23)00216-7] [Medline: 36841609]
120. Tang S, Reily NM, Arena AF, Sheanoda V, Han J, Draper B, et al. Predictors of not receiving mental health services among people at risk of suicide: a systematic review. J Affect Disord 2022 Mar 15;301:172-188 [FREE Full text] [doi: 10.1016/j.jad.2022.01.054] [Medline: 35032506]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
EHR: electronic health record
FLAN: Fine-Tuned Language Net
GPT: generative pretrained transformer
LLM: large language model
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by A Schwartz; submitted 11.06.24; peer-reviewed by A Teles, R Melia; comments to author 23.09.24; revised version received 19.10.24; accepted 10.12.24; published 23.01.25.

Please cite as:
Holmes G, Tang B, Gupta S, Venkatesh S, Christensen H, Whitton A
Applications of Large Language Models in the Field of Suicide Prevention: Scoping Review
J Med Internet Res 2025;27:e63126
URL: https://www.jmir.org/2025/1/e63126
doi:10.2196/63126
PMID:39847414

©Glenn Holmes, Biya Tang, Sunil Gupta, Svetha Venkatesh, Helen Christensen, Alexis Whitton. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Performance Evaluation of Large Language Models in Cervical Cancer Management Based on a Standardized Questionnaire: Comparative Study

Warisijiang Kuerbanjiang[1*], MD; Shengzhe Peng[1*], MD; Yiershatijiang Jiamaliding[1], MD; Yuexiong Yi[1], MD, PhD

Department of Gynecology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei Province, China

[*]these authors contributed equally

Corresponding Author:
Yuexiong Yi, MD, PhD
Department of Gynecology
Zhongnan Hospital of Wuhan University
169 Donghu Road
Wuhan, Hubei Province, 430071
China
Phone: 86 15671669885
Fax: 86 027 67813142
Email: yiyuexiong@163.com

Abstract

Background: Cervical cancer remains the fourth leading cause of death among women globally, with a particularly severe burden in low-resource settings. A comprehensive approach—from screening to diagnosis and treatment—is essential for effective prevention and management. Large language models (LLMs) have emerged as potential tools to support health care, though their specific role in cervical cancer management remains underexplored.

Objective: This study aims to systematically evaluate the performance and interpretability of LLMs in cervical cancer management.

Methods: Models were selected from the AlpacaEval leaderboard version 2.0 and based on the capabilities of our computer. The questions inputted into the models cover aspects of general knowledge, screening, diagnosis, and treatment, according to guidelines. The prompt was developed using the Context, Objective, Style, Tone, Audience, and Response (CO-STAR) framework. Responses were evaluated for accuracy, guideline compliance, clarity, and practicality, graded as A, B, C, and D with corresponding scores of 3, 2, 1, and 0. The effective rate was calculated as the ratio of A and B responses to the total number of designed questions. Local Interpretable Model-Agnostic Explanations (LIME) was used to explain and enhance physicians' trust in model outputs within the medical context.

Results: Nine models were included in this study, and a set of 100 standardized questions covering general information, screening, diagnosis, and treatment was designed based on international and national guidelines. Seven models (ChatGPT-4.0 Turbo, Claude 2, Gemini Pro, Mistral-7B-v0.2, Starling-LM-7B alpha, HuatuoGPT, and BioMedLM 2.7B) provided stable responses. Among all the models included, ChatGPT-4.0 Turbo ranked first with a mean score of 2.67 (95% CI 2.54-2.80; effective rate 94.00%) with a prompt and 2.52 (95% CI 2.37-2.67; effective rate 87.00%) without a prompt, outperforming the other 8 models (P<.001). Regardless of prompts, QiZhenGPT consistently ranked among the lowest-performing models, with P<.01 in comparisons against all models except BioMedLM. Interpretability analysis showed that prompts improved alignment with human annotations for proprietary models (median intersection over union 0.43), while medical-specialized models exhibited limited improvement.

Conclusions: Proprietary LLMs, particularly ChatGPT-4.0 Turbo and Claude 2, show promise in clinical decision-making involving logical analysis. The use of prompts can enhance the accuracy of some models in cervical cancer management to varying degrees. Medical-specialized models, such as HuatuoGPT and BioMedLM, did not perform as well as expected in this study. By contrast, proprietary models, particularly those augmented with prompts, demonstrated notable accuracy and interpretability in medical tasks, such as cervical cancer management. However, this study underscores the need for further research to explore the practical application of LLMs in medical practice.

(J Med Internet Res 2025;27:e63626) doi:10.2196/63626

XSL•FO
RenderX

KEYWORDS

large language model; cervical cancer; screening; artificial intelligence; model interpretability

Introduction

Cervical cancer is a significant global public health challenge, ranking fourth among all female cancers and remaining the leading cause of death in many low-income countries [1]. In 2020, approximately 604,127 new cases and 341,831 deaths from cervical cancer were reported worldwide [1]. Effective cervical cancer control necessitates an integrated approach that combines screening, accurate diagnosis, and personalized treatment to reduce morbidity and mortality. Despite a substantial decline in cervical cancer incidence in the United States since the introduction of screening [2], up to 25% of women remain inadequately treated [3], with even higher rates observed in resource-limited and developing countries [1]. Moreover, precise diagnosis and appropriate treatment are essential for addressing abnormalities detected through screening, particularly to prevent disease progression and improve survival outcomes [4]. Hence, strengthening these efforts is essential for reducing the global burden of cervical cancer and improving patient outcomes across diverse health care contexts.

Large language models (LLMs), as cutting-edge technologies in artificial intelligence, are trained on vast data sets and enable a wide range of applications, from text polishing to complex problem-solving, thanks to their unprecedented natural language understanding capabilities. In the health care domain, LLMs hold the potential to revolutionize medical practices, including decision-making, patient management, and clinical data interpretation [5,6]. Notably, OpenAI's proprietary LLMs, ChatGPT-3.5 and ChatGPT-4.0, have demonstrated high performance on the United States Medical Licensing Examination (USMLE), with ChatGPT-4.0 achieving particularly impressive results [7,8]. Additionally, ChatGPT has shown competence across various medical fields, including surgery [9], cardiology [10], and plastic surgery [11]. Compared with generic language models, medical-specialized models—fine-tuned on domain-specific data sets and subjected to specialized adjustments—have achieved equivalent or superior performance [12].

To date, only a limited number of studies [13] have applied LLMs to questions related to cervical cancer, as well as explainability analyses on either closed- or open-source LLMs to assess transparency and interpretability. The management of abnormal cervical cancer screening results, diagnosis, and treatment is a complex task that requires careful interpretation and follow-up [14]. When deploying LLMs in cervical cancer management, it is crucial to evaluate their performance in managing abnormalities and to identify their strengths and limitations, particularly regarding model transparency and interpretability.

In this study, we aim to compare the performance of current prevalent LLMs in cervical cancer management by evaluating their responses to a set of specifically designed questions (Figure 1). This research may provide valuable evidence to help clinicians manage screening results more effectively and accurately, particularly in regions with limited health care infrastructure.

XSL•FO
RenderX

Figure 1. The flowchart of evaluation of LLMs' performance in cervical cancer management. Nine LLMs, including closed-source, open-source, and medical-specialized types, selected from AlpacaEval leaderboard were evaluated with 100 predefined questions derived from general inquiries and guidelines from ASCCP, CSCCP, FIGO, SEOM-GEICO, and CSCO guidelines. Responses were collected 3 times for each question, analyzed for semantic similarity, and reviewed by 2 experts for accuracy, effectiveness, and interpretability using LIME. ASCCP: American Society for Colposcopy and Cervical Pathology; CSCCP: Chinese Society for Colposcopy and Cervical Pathology of the China Healthy Birth Science Association; CSCO: Chinese Society of Clinical Oncology; FIGO: International Federation of Gynecology and Obstetrics; GEICO: Grupo Español de Investigación en Cáncer de Ovario; LIME: Local Interpretable Model-agnostic Explanations; LLM: large language model; SEOM: Sociedad Española de Oncología Médica.

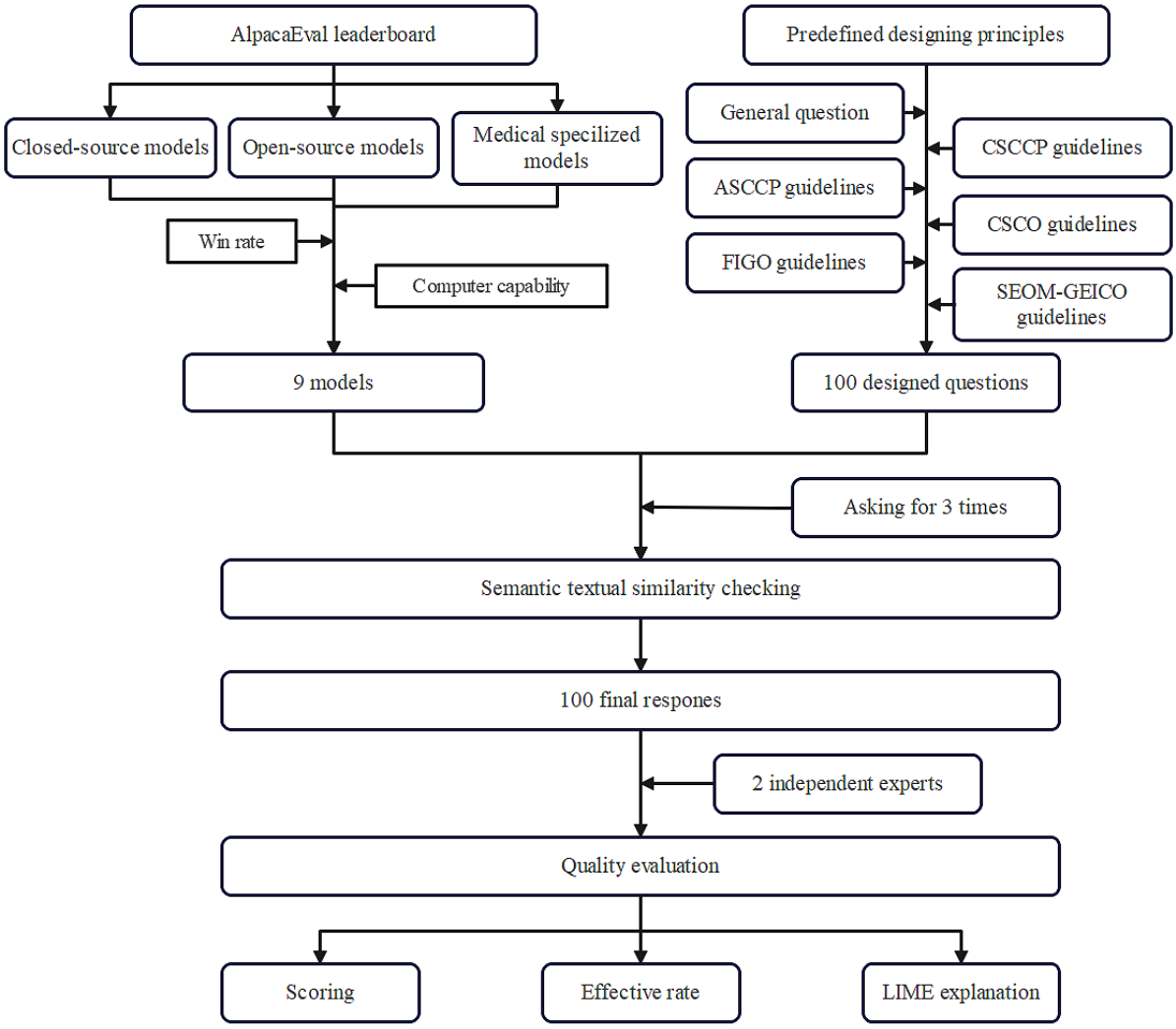

Methods

Model Selection

The AlpacaEval leaderboard is an automated system designed to evaluate language models based on their adherence to instructions, ranking them by comparing their responses to reference answers from top-performing models such as GPT-4. It aims to reduce biases, such as those related to output length. Unlike other leaderboards that may focus on a single type, this leaderboard includes both open- and closed-source models. The selection of potential models—whether closed-source, open-source, or medically specialized—is determined by their win rates on version 2.0 of the leaderboard, updated on March 3, 2024.

For closed-source models, both free and paid versions are included, excluding those that are not publicly available or are in private beta. Open-source models are required to perform effectively on consumer-grade computers with standard configurations, given their potential use in resource-limited applications such as cervical cancer screening. The computer specifications for deploying these models are detailed in Multimedia Appendix 1, with a maximum model size capacity of approximately 7 billion trainable parameters. The selection of medical-specialized models, which are limited in number on leaderboards, is informed by a study [15] summarizing existing medical LLMs and their respective GitHub star counts. The performance of these medical LLMs is assessed based on the benchmark scores of their underlying models.

Criteria for Question and Prompt Designing

Questions Designing

A comprehensive question set was developed to evaluate model performance, including general questions and those specifically focused on cervical cancer screening, diagnosis, and treatment. General questions were designed by our gynecological experts

XSL•FO
RenderX

to address the most common queries about cervical cancer, covering essential, foundational information frequently encountered in clinical practice. Screening-related questions were crafted with reference to the Chinese Society for Colposcopy and Cervical Pathology of the China Healthy Birth Science Association (CSCCP) Consensus on cervical cancer screening and abnormal management in China [16]. To ensure relevance and keep our questions up to date, we also incorporate the 2019 American Society for Colposcopy and Cervical Pathology (ASCCP) Risk-Based Management Consensus Guidelines for Abnormal Cervical Cancer Screening Tests and Cancer Precursors [17]. The questions comprehensively address each clinical decision outlined in the CSCCP guideline flowcharts, as detailed in Multimedia Appendix 2. Additional screening questions were developed based on the Chinese Society of Clinical Oncology (CSCO) Guidelines for the Diagnosis and Treatment of Cervical Cancer (2023) [18]. The diagnosis and treatment questions were developed with reference to the Sociedad Española de Oncología Médica-Grupo Español de Investigación en Cáncer de Ovario (SEOM-GEICO) Clinical Guidelines on Cervical Cancer (2023) [19], the CSCO Guidelines for the Diagnosis and Treatment of Cervical Cancer [18], and The International Federation of Gynecology and Obstetrics (FIGO) 2018 Gynecologic Cancer Report – Interpretation of the Cervical Cancer Guidelines [20]. The design was guided by the principles outlined in Textbox 1.

Textbox 1. Principles guiding design.

1. Diverse complexity levels

- A combination of basic and advanced questions was included to evaluate the model's ability to address both routine and complex clinical scenarios.

2. Strict guideline adherence

- Questions were structured to prioritize guideline-based knowledge, minimizing reliance on outdated or nonevidence-based practices.

3. Primarily closed-ended format

- Predominantly closed-ended questions were used to reduce subjective bias, with a few open-ended questions included to assess the model's capacity for divergent medical problem-solving.

4. Definitive answers

- Each question was designed to have a clear, definitive answer.

These questions aim to evaluate the models' understanding of clinical guidelines, their decision-making processes, and their ability to provide clear, actionable advice.

Prompt Designing

The prompt was designed using the Context, Objective, Style, Tone, Audience, and Response (CO-STAR) framework, which was ranked as the top prompt in the inaugural GPT-4 Prompt Engineering Competition. This framework was applied to guide the LLM in generating expert-level responses in gynecology, with a clear focus on defining the context, objective, style, tone, audience, and response format.

Questioning Method

Each designed question was sequentially input 3 times for each model, both with and without the designed prompt, to test consistency. The coherence of the responses was evaluated using semantic textual similarity [21] by ChatGPT-3.5, the top-performing model on AlpacaEval, which was not used as a test model. If the semantics of all 3 responses are identical, they are sent back to the originating model to select the most suitable answer. In cases of discrepancies, a pairwise comparison is performed with scores ranging from 0 (not typical at all) to 100 (extremely typical) [22]. The 2 responses with the highest similarity scores are returned to their model, which then selects the most appropriate answer (Figure 2).

XSL•FO
RenderX

Figure 2. The flowchart of final response determination for LLMs. Each question was tested 3 times per model, with and without prompts, to assess consistency. Responses were analyzed using STS by ChatGPT-3.5 (not included as a test model). If all 3 responses were semantically identical, the model selected the most suitable answer. For discrepancies, a pairwise comparison scored responses from 0 (Not typical at all) to 100 (Extremely typical), and the 2 highest-scoring responses were returned to the model to determine the final response. LLM: large language model; STS: semantic textual similarity.

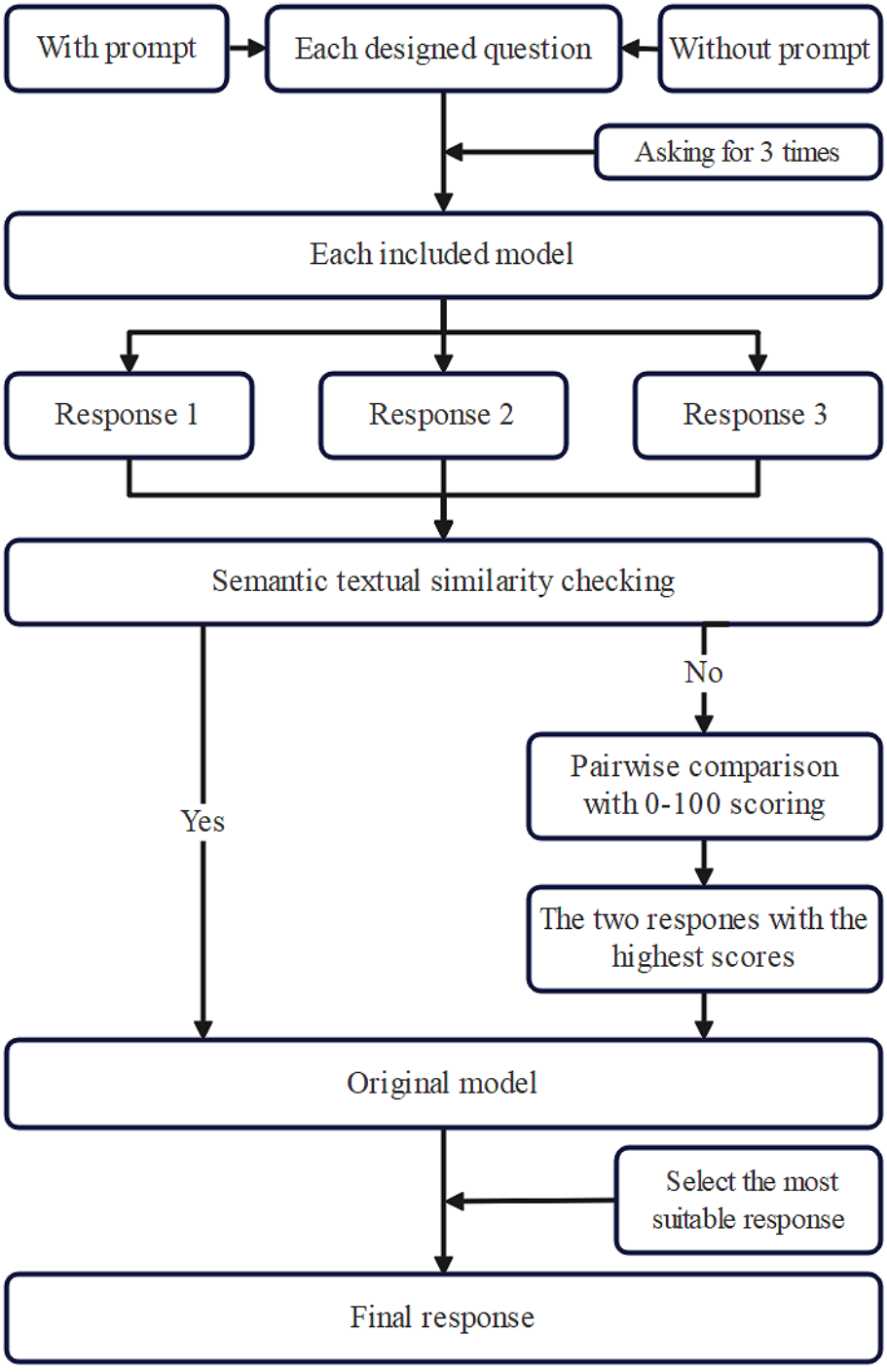

Scoring Process and Criteria

Two gynecological experts independently and anonymously reviewed the responses to each question. If both experts agreed on a score, it was directly accepted; otherwise, they discussed it to determine the final score. Responses were evaluated based on accuracy, adherence to clinical guidelines, clarity of communication, and practicality. A scoring system, modified from a previous study [23], was used to categorize responses into 4 grades: A, B, C, and D, to minimize subjective bias. Grades A and B were considered effective, and the model's effective rate was calculated as follows:

$$\text{Effective rate} = (N_A+N_B)/(N_A+N_B+N_C+N_D) \times 100\%$$

where N represents the number of each grade. Scores are weighted at 3, 2, 1, and 0 points for statistical analysis (Table 1).

XSL•FO
RenderX

Table 1. Criteria of scoring for response.

Grade	Description	Scores
A	Completely correct with comprehensive information	3
B	Mostly correct, but with missing information or minor errors	2
C	Contains major errors but with some correct content	1
D	Completely wrong or off-topic	0

Model Explainability Analysis

Local Interpretable Model-Agnostic Explanations (LIME) is widely recognized for generating locally interpretable explanations of machine learning model predictions, including natural language processing models [24,25]. In this study, LIME was used to interpret LLM outputs by adapting methods previously successful in natural language processing. The primary LIME parameter, the number of samples, was set to 10 times the input sentence's token count, based on preliminary experiments and prior applications of LIME to LLMs [26]. Each input question was analyzed to identify key terms with assigned weights, and the top 5 key terms by weight were selected. Our experts manually annotated 5 key terms per question for comparison. An intersection-over-union (IoU) analysis was performed between the LIME-selected key terms and the expert-annotated key terms to evaluate their alignment (Figure 3).

Figure 3. The flowchart of model explainability analysis by the Local Interpretable Model-agnostic Explanations (LIME) methodology. LIME was used to interpret large language model (LLM) outputs by analyzing each input question and generating variant questions with or without prompt. The number of samples was set to 10 times the token count of the input sentence. Key terms were extracted using the Bio_ClinicalBERT model, and the top 5 terms by weight were selected. Experts manually annotated 5 key terms per question for comparison. An intersection over union (IoU) analysis was conducted to evaluate the alignment between LIME-selected and expert-annotated key terms. BERT: Bidirectional Encoder Representations from Transformers.

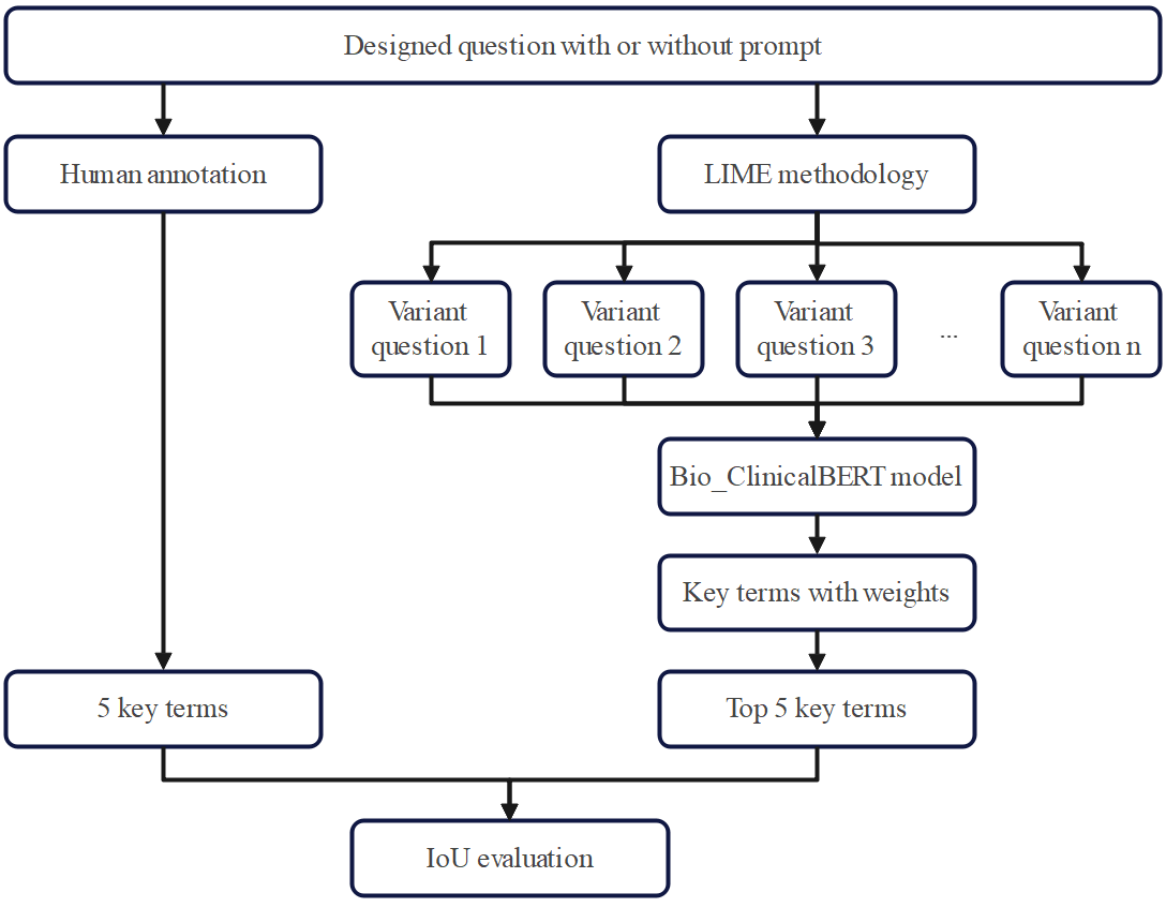

$$IoU(x1, x2) = (|x1 \cap x2|)/(|x1 \cup x2|)$$

Ethical Considerations

This study did not involve human participants, identifiable patient data, or protected health information. The data utilized

XSL•FO
RenderX

in this study comprised publicly available sources, including leaderboards, clinical guidelines, and secondary analyses of model-generated outputs. Therefore, an ethical review was not required under Zhongnan Hospital of Wuhan University's secondary research policies. The study complied with the Declaration of Helsinki and institutional guidelines for secondary data use.

Statistical Methods

Analyses were conducted using R version 4.3.1 (R Foundation) and RStudio 2023.12.1+402 (R Foundation). Differences across models were assessed using the chi-square test for categorical variables. For paired comparisons, data were first tested for normality. If normally distributed, a paired *t* test was applied, with results reported as mean and SD; otherwise, a paired Wilcoxon rank sum test was used, with outcomes presented as median and IQR. Effective rates were reported as mean values with 95% CIs. A *P* value of less than .05 was considered indicative of a significant difference.

Results

Model Selection

After screening for win rates and conducting tests on our computers, our study included 9 models. The proprietary models are ChatGPT-4.0 Turbo, Claude 2, and Gemini Pro, which are accessible through their official websites. The open-source LLMs include Mistral-7B-v0.2, Starling-LM-7B Alpha, and Microsoft Phi-2. The medical-specialized models are the Chinese models HuatuoGPT and QiZhenGPT, along with the English model BioMedLM 2.7B. The expected performance ranking of the selected models is as follows: ChatGPT-4.0 Turbo > Gemini Pro > Claude 2 > Mistral-7B-v0.2 > Starling-LM-7B Alpha > ChatGLM 6B (QiZhenGPT) > Phi-2 > Baichuan2-7B-Chat (HuatuoGPT). BioMedLM 2.7B is excluded from this ranking because it is not listed on the AlpacaEval Leaderboard. The characteristics of the included models are presented in Table 2.

Table 2. The characteristics of included models.

Model and access reference	AlpacaEval win rate	Description
ChatGPT-4.0 Turbo [27]	50%	Developed by OpenAI, ChatGPT-4.0 Turbo is an LLM[a] that is currently the most powerful in terms of performance.
Claude 2 [28]	17.19%	Developed by Anthropic, it is built on the GPT-3 architecture. This model features a context window of 100,000 ultra-long tokens, enabling it to handle longer context inputs efficiently.
Mistral-7B-v0.2 [29]	14.72%	Mistral-7B-v0.2 is the strongest open-source model on the list that can be deployed on consumer computers. Furthermore, the popularity of this model is high, as it received more than 700,000 downloads in January 2024.
Starling-LM-7B alpha [30]	14.25%	A fine-tuned model that outperforms all models to date on MT-Bench except for OpenAI's GPT-4 and GPT-4 Turbo.
Gemini Pro [31]	18.18%	Developed by Google DeepMind, the more advanced Gemini Ultra is not yet available to the public, so we used the Pro version.
HuatuoGPT 2-7B [32]	1.99% (base model)	Developed by the Shenzhen Institute of Big Data and The Chinese University of Hong Kong, this Chinese medical LLM is fine-tuned based on Baichuan2-7B. Uses deploying method. The online demo is available at [33].
QiZhenGPT [34]	3.01% (base model)	Released by Zhejiang University, the project includes 3 versions, each fine-tuned from the base models of ChatGLM-6B, Chinese-LLaMA-7B, and CaMA-13B.
Phi-2 [35]	2.34%	Released by Microsoft, this small language model has a data size of only 2.7 billion. Easy to deploy, even on consumer-grade computers, where it exhibits exceptionally fast response times.
BioMedLM 2.7B [36]	N/A[b]	Previously known as PubMedGPT 2.7B, this model was developed through pretraining.

[a]LLM: large language model.

[b]N/A: not applicable.

Questions and Prompts for LLMs

The question set consisted of 100 questions designed to encompass a broad range of clinical scenarios commonly encountered in cervical cancer management. The first 22 questions focused on general knowledge, emphasizing foundational aspects frequently encountered in clinical gynecology. The next 40 questions addressed cervical cancer screening, aligning with the latest consensus guidelines and decision-making protocols. Subsequently, 6 and 32 questions covered diagnosis and treatment, respectively, offering a comprehensive evaluation of the models' ability to interpret diagnostic criteria and recommend evidence-based treatment options. By including both routine and complex queries, the question set serves as a robust benchmark for assessing model performance, accuracy, and adherence to evidence-based medical practices. The complete list of questions is provided in Table 3.

XSL•FO
RenderX

Table 3. The 100 designed questions based on cervical cancer guidelines.

Category	Questions
Questions related to general knowledge	• What are the risk factors that may necessitate cervical cancer screening? • At what age or under what conditions is cervical cancer screening typically deemed unnecessary? • What strategies are effective in reducing the risk of developing cervical cancer? • What are the common clinical symptoms of cervical cancer? • For individuals who have been vaccinated against HPV[a], is it still necessary for them to undergo cervical cancer screening? • Is cervical cancer screening still recommended for individuals who have had only 1 sexual partner or are not currently sexually active? • What are the recommended intervals for cervical cancer screening, and do these intervals vary among different age groups? • Is cervical cancer screening universally recommended for all age groups? If not, what are the reasons for excluding certain age groups from undergoing cervical cancer screening? • How necessary is cervical cancer screening for women who have undergone total hysterectomy? • What is the significance of cervical cancer screening? • Why is the combined use of cytological screening (Papanicolaou test) and HPV testing not recommended for women aged 21-29 years? • Is there an invariable link between HPV infection and the onset of cervical cancer? • How should one interpret an abnormal result from a cervical cancer screening test? Does such a result definitively indicate the presence of cervical cancer? • Why is yearly cervical cancer screening not recommended? • What are the objectives of cervical cancer screening protocols? • Is cervical cancer hereditary? If so, should individuals with a familial history of cervical cancer be subject to more frequent screening protocols? • What's the difference between a pelvic examination and a Pap test? • Can individuals independently administer HPV tests, and if so, how accurate are these self-administered tests? • Is it possible for cervical cancer to manifest within the interscreening interval, particularly between 2 consecutive cervical screening tests? • Is cervical screening necessary for individuals who have reached menopause? • Is it recommended to undergo cervical screening during pregnancy? • What procedures are typically involved in cervical cancer screening?

XSL•FO
RenderX

Category	Questions
Questions related to screening	• What is the examination process for HPV-positive high-risk types? • When the initial screening shows positive HPV with high-risk HPV types, emphasizing that HPV is not typed, what is the next step in the examination? • When the initial screening shows positive HPV with high-risk HPV types, and the cytological examination result is negative, what does it indicate? Should regular check-ups follow? And if so, what should be the frequency? • When the initial screening shows positive HPV with high-risk HPV types, and the cytological examination result is ≥ASC-US[b], should the next step be a colposcopy? • When the initial screening shows positive HPV with high-risk HPV types, typing identifies HPV16/18 positivity, should the next step be a colposcopy? • When the initial screening shows positive HPV with high-risk HPV types, and upon typing, it shows neither HPV16/18 positive but 1 of the other 12 types, what is the most likely subsequent examination? • When cytological examination indicates ASC-US as an abnormal initial screening result, what is the next step in the examination? • When cytological examination indicates ASC-US as an abnormal initial screening result, and HPV is used for triage, if HPV is positive, should a colposcopy follow? • When cytological examination indicates ASC-US as an abnormal initial screening result, and HPV is used for triage, if HPV is negative, what should be the subsequent examination? • When cytological examination results show ASC-H[c], LSIL[d], HSIL[e], is a colposcopy needed next? • When cytological examination results show AGC[f], what is the next examination required? • If both cytological and high-risk HPV joint tests show negative results, what other examination should follow? • If joint testing of cytology and high-risk HPV shows HPV negative and the cytological result is ASC-US, what should be done next? • If joint testing of cytology and high-risk HPV shows HPV negative, and the cytological result is >ASC-US, what should be done next? Is a colposcopy required? • In cervical cancer detection, if joint testing of cytology and high-risk HPV shows HPV positive and the cytological result is ≥ASC-US, what specific examinations should follow? • In cervical cancer detection, if joint testing of cytology and high-risk HPV shows HPV positive and the cytological result is negative, what specific examinations should follow? • In cervical cancer detection, if joint testing of cytology and high-risk HPV shows HPV positive and the cytological result is negative, and the HPV typing is HPV16/18 positive, what specific examinations should follow? • In cervical cancer detection, if a patient has joint testing of cytology and high-risk HPV, with the results showing HPV positivity, a negative cytological result, and no typing for HPV16 or HPV18, are additional tests needed? • In cervical cancer screening, if a patient has a histopathological confirmation (biopsy) of LSIL, cytology findings of LSIL or higher, and a TZ3[g] classification, what follow-up steps should be conducted? • In cervical cancer screening, if a patient has a histopathological confirmation (biopsy) of LSIL, cytology findings of LSIL or higher, and a TZ1/2 classification, what follow-up steps should be conducted? • In cervical cancer screening, if a patient has a histopathological confirmation (biopsy) of LSIL and cytology findings of ASC-H or higher, what follow-up steps should be conducted? • In cervical cancer screening, if a patient aged 21-24 years has a histopathological confirmation (biopsy) of LSIL and cytology findings of ASC-H or higher, what follow-up steps should be conducted? • In cervical cancer screening, if a patient aged 21-24 years has a histopathological confirmation (biopsy) of LSIL, cytology findings of ASC-H or higher, and a TZ3 classification for colposcopy, what follow-up steps should be conducted? • In cervical cancer screening, if a patient aged 21-24 years has a histopathological confirmation (biopsy) of LSIL, cytology findings of ASC-H or higher, and a TZ1/2 classification for colposcopy, what follow-up steps should be conducted? • In cervical cancer screening, if a pregnant patient has a histopathological confirmation (biopsy) of LSIL and cytology findings of ASC-H or higher, what follow-up steps should be conducted? • In cervical cancer screening, if a patient has a histopathological confirmation (biopsy) of HSIL and a TZ1/2 classification, what follow-up steps should be conducted? • In cervical cancer screening, if a patient has a histopathological confirmation (biopsy) of HSIL and a TZ3 classification, what follow-up steps should be conducted? • In cervical cancer screening, if a patient aged 21-24 years has a histopathological confirmation (biopsy) of CIN[h] III/HSIL, what follow-up steps should be conducted? • In cervical cancer screening, if a patient aged 21-24 years has a histopathological confirmation (biopsy) of HSIL and a TZ3 classification for colposcopy, what follow-up steps should be conducted? • In cervical cancer screening, if a patient aged 21-24 years has a histopathological confirmation (biopsy) of CIN II/III/HSIL or CIN II/HSIL and a TZ1/2 classification for colposcopy, what follow-up steps should be conducted? • In cervical cancer screening, if a pregnant patient has a histopathological confirmation (biopsy) of HSIL, what follow-up steps should be conducted? • After diagnostic/therapeutic cervical conization for cervical cancer, what follow-up steps should be conducted? • In cervical cancer screening, if a pregnant patient has a histopathological confirmation (biopsy) of HSIL without invasive cancer during pregnancy, what follow-up steps should be conducted after childbearing? • What are the strategies for HPV vaccine use? What are the recommended vaccination programs for different age groups? • Is HPV primary screening applicable in low-income countries? If so, why? • Which cervical cancer screening methods are widely used in low-income countries, particularly in sub-Saharan Africa? • What are the key indicators included in the expert consensus for quality control management of HPV testing? • What is the difference in clinical management between ASC-US and ASC-H? • Is 4-quadrant sampling still necessary for patients with no abnormalities on colposcopy? • What are the differences in cervical cancer screening and management strategies for women during pregnancy?

XSL•FO
RenderX

Category	Questions
Questions related to diagnosis	• What is the preferred clinical diagnosis of cervical cancer? • What tests should be conducted to make a pathologic diagnosis of cervical cancer? • What tumor markers can be tested for laboratory diagnosis of cervical cancer? What is the significance of each tumor marker detected? • In the diagnosis of cervical cancer, which imaging method should be preferred to evaluate cervical tumors? Which imaging method should be used to evaluate metastatic lesions? • What are the recommended diagnostic tools for patients with FIGO[i] stage IA1 cervical cancer? Is it necessary to consider lymphovascular infiltration? • How are imaging tools used to assess tumor size and lymph node status in FIGO staging? What specific imaging tools are recommended?
Questions related to treatments	• In cervical cancer treatment, if the patient does not wish to preserve fertility, and the stage is IA1 without lymphovascular space invasion, what treatment measures should be taken? • In cervical cancer treatment, if the patient does not wish to preserve fertility, and the stage is IA1 with lymphovascular space invasion, what treatment measures should be taken? • In cervical cancer treatment, if the patient does not wish to preserve fertility, and the stage is IA2, what treatment measures should be taken? • In cervical cancer treatment, if the patient does not wish to preserve fertility, and the stage is IB1, IIA1, or IIB2, what treatment measures should be taken? • In cervical cancer treatment, if the patient does not wish to preserve fertility, and the stage is IB3 or IIA2, what treatment measures should be taken? • In cervical cancer treatment, if the patient wishes to preserve fertility, and the stage is IA1 without lymphovascular space invasion, what treatment measures should be taken? • In cervical cancer treatment, if the patient wishes to preserve fertility, and the stage is IA1 with lymphovascular space invasion or IA2, what treatment measures should be taken? • In cervical cancer treatment, if the patient wishes to preserve fertility, and the stage is IB1, what treatment measures should be taken? • In cervical cancer treatment, if the patient wishes to preserve fertility, and the stage is IB2, what treatment measures should be taken? • In cervical cancer treatment, if the stage is IIB, IIIA, or IIIB, what treatment measures should be taken? • In cervical cancer treatment, if the stage is IIIC1, what treatment measures should be taken? • In cervical cancer treatment, if the stage is IIIC2, what treatment measures should be taken? • In cervical cancer treatment, if the stage is IVA without lymph node enlargement, what treatment measures should be taken? • In cervical cancer treatment, if the stage is IVA with lymph node enlargement, what treatment measures should be taken? • In cervical cancer treatment, if the stage is IVB, what treatment measures should be taken? • After radical surgery for early cervical cancer, if the abdominal aortic lymph nodes are negative but high-risk factors are present, what treatment measures should be taken? • After radical surgery for early cervical cancer, if the abdominal aortic lymph nodes are negative but intermediate-risk factors are present, what treatment measures should be taken? • After radical surgery for early cervical cancer, if the abdominal aortic lymph nodes are positive but there is no distant metastasis, what treatment measures should be taken? • What are the surgical treatment options for patients with FIGO stage IA1 cervical cancer? Are they suitable for patients with preserved fertility? • In patients with FIGO stage IB2 and IIA1, what factors determine the choice between surgery and radiotherapy? What are the differences in outcomes between the 2 modalities? • What is the recommended treatment of choice for FIGO stage IB3 cervical cancer? • What is the difference between the different types of radical hysterectomy? For which patients is it indicated? • What are the advantages of intensity-modulated radiation therapy in radiotherapy for cervical cancer? • What are the common sites of recurrence in cervical cancer? How is the risk of recurrence monitored? • What are the recommended treatment strategies for recurrent cervical cancer? Can surgery, radiotherapy, and chemotherapy be combined? • What are the treatment strategies for metastatic cervical cancer? Are there specific treatments at different metastatic sites? • What are the treatment strategies for patients with cervical cancer during pregnancy? How does the treatment differ in early, intermediate, and advanced stages of pregnancy? • Are patients with cervical cancer in pregnancy suitable for surgery? At what stage of pregnancy should surgery be considered? • What is the radiotherapy strategy for patients with locally advanced cervical cancer? When is a combination of chemotherapy recommended? • When is image-guided brachytherapy necessary and how is it different from conventional radiotherapy? • What are the methods for monitoring recurrence after treatment of cervical cancer? Is routine imaging recommended? • What are the main goals of palliative care? In patients with metastatic cervical cancer, how can palliative care be combined with radiotherapy and chemotherapy?

[a]HPV: human papillomavirus.

[b]ASC-US: atypical squamous cells of undetermined significance.

[c]ASC-H: atypical squamous cells, cannot exclude high-grade squamous intraepithelial lesion

[d]LSIL: low-grade squamous intraepithelial lesion.

[e]HSIL: high-grade squamous intraepithelial lesion.

[f]AGC: atypical glandular cell.

XSL•FO
RenderX

[g]TZ3: Type 3 transformation zone.

[h]CIN: cervical intraepithelial neoplasia.

[i]FIGO: International Federation of Gynecology and Obstetrics.

Using the CO-STAR framework, the prompt was designed to guide the model in providing clinically relevant and detailed responses, meeting the standards necessary for accurate interpretation in cervical cancer management. The specific details of the prompt are presented in Table 4.

Table 4. Prompt designing based on the CO-STAR[a] framework.

Prompt element	Content
# Context #	Now you are a gynecologist with over 20 years of experience in medicine and you are answering questions about the medical specialty of cervical cancer treatment, diagnosis, and screening.
# Objective #	Please answer the following questions correctly and in strict accordance with the latest guidelines for the screening, treatment, and diagnosis of cervical cancer.
# Style #	The information should be clear, concise, and medically accurate, using terminology appropriate for both health care professionals and patients.
# Tone #	The tone should be formal and professional, recognizing the sensitive nature of cancer-related discussions.
# Audience #	The primary audience includes health care professionals, researchers, and patients seeking information about cervical cancer management.
# Response #	Generate detailed responses to specific queries regarding cervical cancer. Assess the accuracy and relevance of the information provided.

[a]CO-STAR: Context, Objective, Style, Tone, Audience, and Response.

Model Stability

Among the 9 models evaluated, 7 demonstrated good reproducibility with stable responses. However, the repeatability of Phi-2 and QiZhenGPT was unsatisfactory, as posing the same question 3 times often resulted in varying answers. For Phi-2, 61 out of 100 responses with the prompt and 68 responses without the prompt exhibited semantic differences across repetitions. Similarly, for QiZhenGPT, 60 responses with the prompt and 55 without the prompt varied. In both cases, pairwise comparisons were necessary to determine the final output (see Multimedia Appendix 2).

Model Efficacy

The evaluation results for each model, with and without the prompt, are presented in Figure 4. The top 3 performers were all proprietary models. ChatGPT-4.0 Turbo achieved the highest effective rate, at 94% (mean score 2.67, 95% CI 2.54-2.80) with the prompt and 87% (mean score 2.52, 95% CI 2.37-2.67) without it, highlighting the positive impact of the prompt on its performance. Claude 2 maintained an effective rate of 85% both with and without the prompt, with similar mean scores of 2.35 (95% CI 2.16-2.54) and 2.39 (95% CI 2.22-2.56), respectively. Gemini Pro showed moderate improvement, with its effective rate increasing from 66% (mean score 2.00, 95% CI 1.80-2.20) without the prompt to 77% (mean score 2.25, 95% CI 2.06-2.44) with the prompt.

XSL•FO
RenderX

Figure 4. The efficacy assessment of each model with and without the prompt. The number in A, B, C and D represents the distribution of response quality in each grade. ChatGPT-4.0 Turbo achieved the highest effective rate (94% with a mean score of 2.67, 87% without at 2.52), while Claude 2 remained consistent at 85%. Gemini Pro improved from 66% to 77% with prompts. Among medically specialized models, HuatuoGPT slightly increased from 53% to 57% without prompts, BioMedLM stayed low (39% vs 38%), and QiZhenGPT had the lowest rates (33% vs 32%), showing minimal impact from prompts.

Models	Prompt	A	B	C	D	Mean [95%CI]	Effective rate
ChatGPT-4.0 Turbo	With	74	20	4	2	2.67 [2.54, 2.80]	94%
	Without	66	21	12	1	2.52 [2.37, 2.67]	87%
Claude 2	With	61	25	10	4	2.35 [2.16, 2.54]	85%
	Without	58	27	11	4	2.39 [2.22, 2.56]	85%
Gemini Pro	With	47	27	20	6	2.25 [2.06, 2.44]	77%
	Without	42	24	26	8	2.00 [1.80, 2.20]	66%
Mistral-7B-v0.2	With	48	26	20	6	2.16 [1.97, 2.35]	74%
	Without	47	23	17	13	2.04 [1.83, 2.25]	70%
Starling-LM-7B alpha	With	44	27	18	11	2.04 [1.83, 2.25]	71%
	Without	39	28	20	13	1.93 [1.72, 2.14]	67%
HuatuoGPT	With	33	21	18	28	1.56 [1.32, 1.80]	53%
	Without	34	23	28	15	1.76 [1.54, 1.98]	57%
BioMedLM 2.7B	With	22	17	22	39	1.13 [0.90, 1.36]	39%
	Without	24	14	21	41	1.21 [0.97, 1.45]	38%
Microsoft Phi2	With	28	21	20	31	1.50 [1.26, 1.74]	49%
	Without	23	26	23	28	1.44 [1.22, 1.66]	49%
QiZhenGPT	With	16	17	29	38	1.13 [0.91, 1.35]	33%
	Without	21	11	34	34	1.19 [0.97, 1.41]	32%

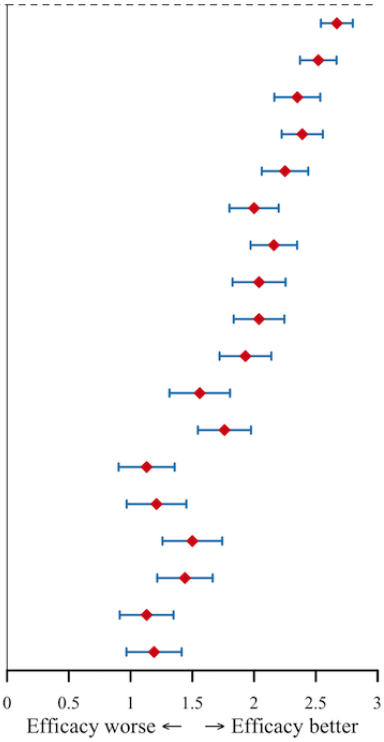

By contrast, the 3 medically specialized models exhibited lower effective rates. HuatuoGPT achieved an effective rate of 53% (mean score 2.00, 95% CI 1.80-2.20) with the prompt, which unexpectedly increased to 57% (mean score 1.76, 95% CI 1.54-1.98) without it. BioMedLM showed minimal improvement, with an effective rate of 39% (mean score 1.13, 95% CI 0.90-1.36) with the prompt and 38% (mean score 1.76, 95% CI 1.54-1.98) without it. QiZhenGPT had the lowest performance, with an effective rate of 33% (mean score 1.13, 95% CI 0.91-1.35) with the prompt and 32% (mean score 1.19, 95% CI 0.97-1.41) without it, showing limited impact from the prompt on enhancing its responses. The STS testing results are provided in Multimedia Appendix 3. Detailed responses and original scoring are provided in Multimedia Appendix 4.

The chi-square test revealed significant differences across models (P=.001). As the data for each model did not follow a normal distribution (P<.01), the Wilcoxon rank sum test was applied. With the prompt, ChatGPT-4.0 Turbo and Claude 2 exhibited highly significant differences (P<.001) compared with most other models, indicating substantial performance enhancement when the prompt was used. This pattern remained consistent in comparisons with lower-performing models, such as HuatuoGPT, BioMedLM, and QiZhenGPT. Without the prompt, significant differences were still observed, particularly between high-performing models such as ChatGPT-4.0 Turbo (P<.001) and Claude 2 (P<.001) and lower-performing models. However, the absence of the prompt reduced significance in certain comparisons, such as between Mistral-7B and Gemini Pro (P=.30) or BioMedLM and QiZhenGPT (P=.64). When comparing performance with and without the prompt, ChatGPT-4.0 Turbo and Gemini Pro demonstrated statistically significant improvements with the prompt (P<.001), whereas Claude 2 showed no significant difference (P=.07). By contrast, models such as BioMedLM (P=.77), Phi-2 (P=.53), and QiZhenGPT (P=.01) exhibited minimal or insignificant changes (Figure 5).

XSL•FO
RenderX

Figure 5. Pairwise significance comparison between models with and without the prompt. The upper triangle represents significance levels between models with the prompt, while the lower triangle displays significance levels without the prompt. The diagonal section shows significance of performance differences within each model between the prompted and unprompted conditions. The Wilcoxon rank sum test was performed as the data for each model did not follow a normal distribution. ChatGPT-4.0 Turbo and Claude 2 showed significant improvements (P<.001) with prompts, outperforming HuatuoGPT, BioMedLM, and QiZhenGPT. Without prompts, differences persisted but were less pronounced, especially between models such as Mistral-7B and Gemini Pro. insig.: insignificant; *P<.05; **P<.01; ***P<.001.

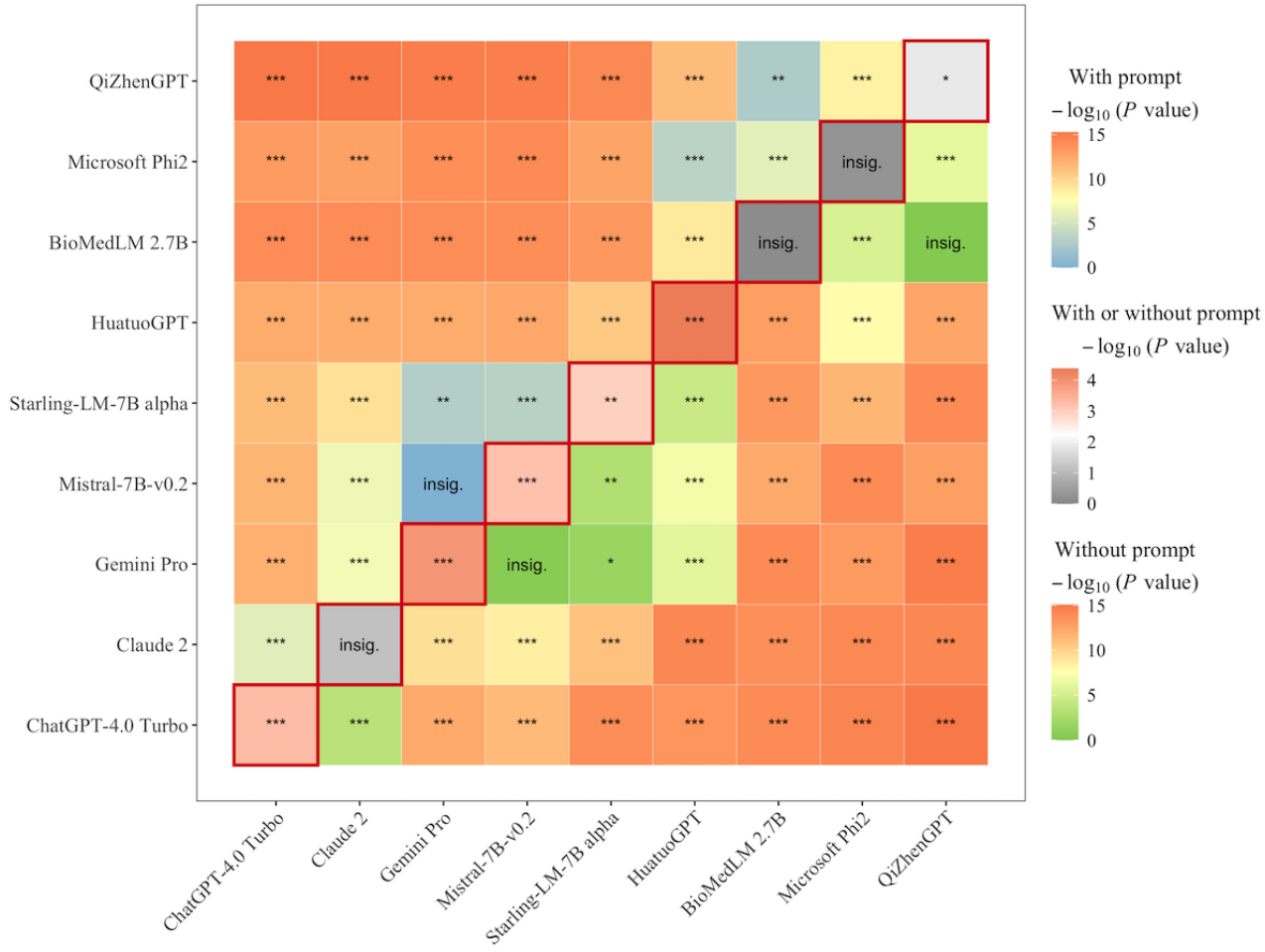

Model Explainability

Given the nonnormal distribution of IoU values for each model, the Wilcoxon rank sum test was used to assess differences. As shown in Figure 6, the inclusion of prompts significantly improved the alignment between model-generated explanations and human annotations, with all models exhibiting statistically significant differences between prompted and unprompted conditions (P<.001). Specifically, Claude 2, Gemini Pro, Starling-LM-7B Alpha, ChatGPT-4.0 Turbo, and Mistral-7B-v0.2 demonstrated a consistent median IoU of 0.43 with prompts. Among these, ChatGPT-4.0 Turbo had the widest IoU range (IQR 0.56). Without prompts, the median IoU for these models dropped to 0.25, with narrower IQRs ranging from 0.32 to 0.43, indicating reduced interpretability consistency. Among the medically specialized models, QiZhenGPT showed the most substantial improvement with prompts, achieving a median IoU of 0.43 (IQR 0.42), aligning it with the performance of proprietary models under similar conditions. By contrast, BioMedLM 2.7B and HuatuoGPT exhibited lower interpretability, with median IoUs of 0.29 and 0.25, respectively, and smaller IQRs in nonprompted conditions (median IoU of 0.11 and IQR of 0.25 for both).

XSL•FO
RenderX

Figure 6. Model explainability analysis by intersection over union (IoU) for included models with or without the prompt. Data are expressed as median and IQR and the Wilcoxon rank sum test was applied due to nonnormal distribution within each model's data. Claude 2, Gemini Pro, Starling-LM-7B alpha, ChatGPT-4.0 Turbo, and Mistral-7B-v0.2 achieved a median IoU of 0.43 with prompts, dropping to 0.25 without prompt, with ChatGPT-4.0 Turbo showing the widest range (IQR 0.56). QiZhenGPT showed the most improvement among medical models with a median IoU of 0.43 with prompt. BioMedLM 2.7B and HuatuoGPT showed lower interpretability, with nonprompted IoUs of 0.11 (IQR 0.25).

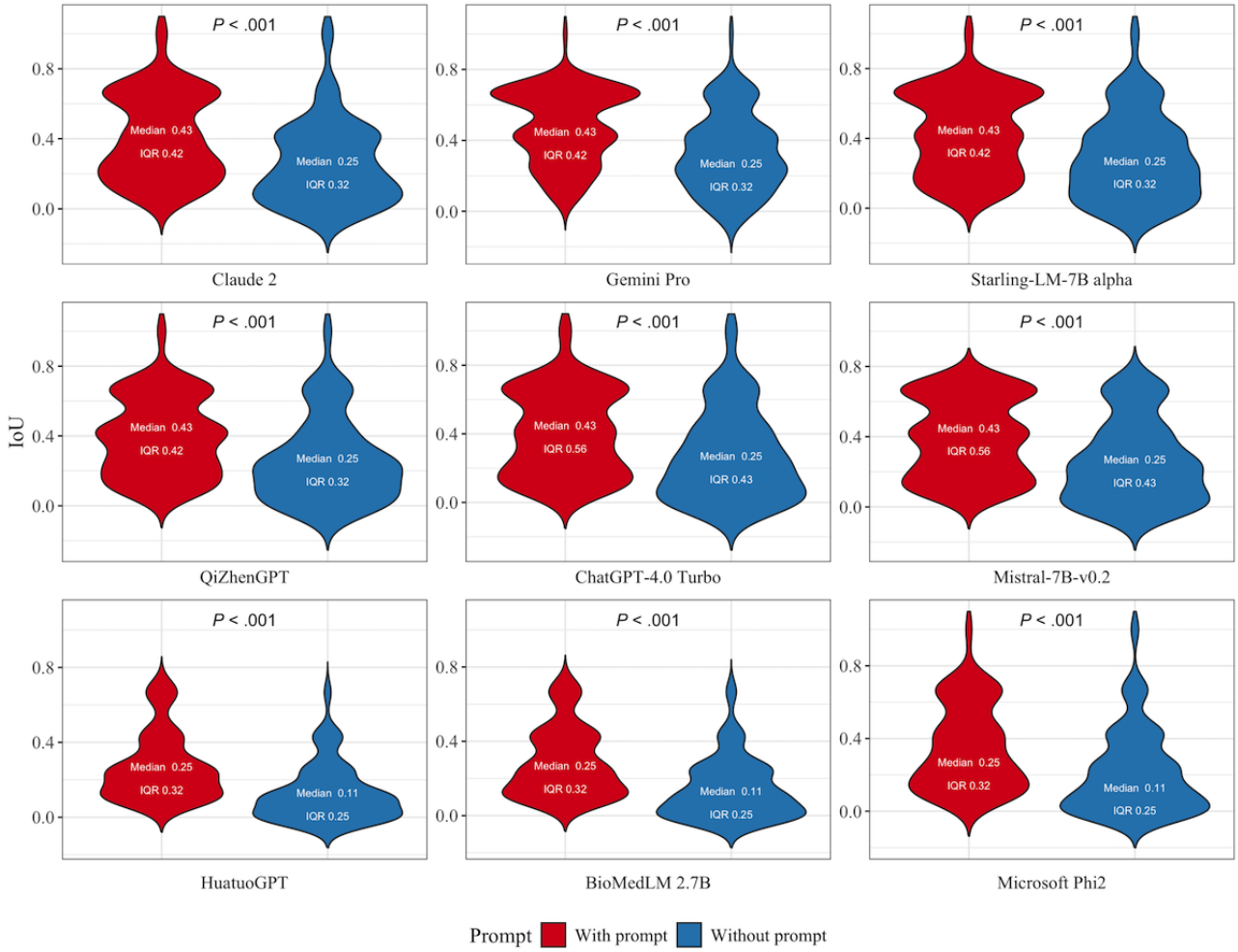

Discussion

Principal Findings

This study systematically evaluated 9 LLMs for their performance, stability, and interpretability in cervical cancer management. The results revealed that proprietary models, such as ChatGPT-4.0 Turbo, Claude 2, and Gemini Pro, achieved superior response accuracy and interpretability, particularly with prompt guidance. By contrast, medically specialized models such as HuatuoGPT, QiZhenGPT, and BioMedLM demonstrated comparatively lower effectiveness, with limited improvement from prompt use. Notably, while proprietary models exhibited consistent reproducibility, certain open-source and specialized models, such as Phi-2 and QiZhenGPT, showed variable responses upon repeated questioning. Furthermore, the use of prompts significantly enhanced interpretability in models such as Claude 2, Gemini Pro, and Starling-LM-7B Alpha, highlighting the potential of structured input to improve alignment with clinical expectations.

Comparison to Prior Work

In terms of average score ranking, proprietary models such as ChatGPT-4.0 Turbo, Claude 2, and Gemini Pro outperformed open-source models. This result aligns with traditional views on the superiority of proprietary systems [37]. However, without the prompt, Mistral-7B outperformed Gemini Pro. Among the open-source models, Mistral-7B-v0.2 and Starling-LM-7B Alpha outperformed HuatuoGPT and BioMedLM 2.7B. However, the repeatability of answers from Microsoft Phi-2 was poor, making it unsuitable for medical applications, while ChatGPT-4.0 Turbo and Claude 2 provided accurate and consistent responses. Our results indicated that the performance of the 3 medical models was average, challenging the prevailing belief that medical-specific models are superior for medical queries [38]. Previous studies [21,39] have shown that larger models, characterized by increased parameter counts, tend to perform better. Additionally, as the model scale increases, its generalization ability improves [40]. This may explain the relative underperformance of medical models compared with proprietary models, given the substantial disparity in parameter magnitude between them.

XSL•FO
RenderX

Recent advancements in algorithms have been shown to improve the performance of LLMs in the medical field [23,39], with research [41] indicating significant accuracy improvements using specific prompts. The integration of prompts has had a notable impact on the performance of several LLMs, emphasizing the value of structured input in guiding model responses within clinical contexts. Proprietary models, such as ChatGPT-4.0 Turbo and Gemini Pro, showed marked improvements in effective rate and response accuracy when guided by the CO-STAR prompt framework, suggesting that structured prompts help enhance focus on relevant clinical information and reduce ambiguity [42]. Conversely, models with specialized but limited training, such as BioMedLM, exhibited minimal sensitivity to prompts, likely due to architectural limitations in processing complex prompt structures [43]. Interestingly, HuatuoGPT experienced a decline in performance with the addition of prompts. This unexpected outcome suggests that the structured prompt for HuatuoGPT may have interfered with its response generation by introducing constraints that conflicted with its training data or underlying language patterns, potentially limiting its ability to accurately interpret open-ended clinical scenarios [44]. Additionally, smaller models often become confused when handling longer prompts [45]. The variation in prompt effectiveness across models underscores that, while structured prompts generally improve response precision, their impact is influenced by the model design and data scope.

The IoU serves as a robust indicator of alignment between model-generated explanations and human annotations, providing insights into the interpretability of LLMs in clinical contexts [46]. A higher IoU reflects greater consistency with human-provided explanations, suggesting enhanced model transparency and reliability in decision-making support. Our results demonstrate that a higher IoU corresponds to better alignment between model-generated explanations and human annotations, indicating improved interpretability. Proprietary models, particularly ChatGPT-4.0 Turbo and Claude-2, performed well in aligning with human explanations when prompts were used, highlighting their potential for generating clinically relevant interpretations. Interestingly, the rankings for model explainability based on IoU scores do not directly correlate with those based on effective rates. This discrepancy likely arises because improvements in model performance do not necessarily enhance explainability [47]. According to previous studies [48], as models become more accurate, their alignment with human-annotated explanations does not necessarily improve. This misalignment suggests that the factors driving a model's effectiveness in task accuracy differ from those contributing to explainability. Higher-performing models may rely on complex, implicit patterns that are not fully captured by metrics such as IoU, which primarily assess agreement with human logic rather than the model's actual reasoning process [49]. However, IoU alone may not fully capture explanation quality, as it can overlook aspects such as coherence and clinical relevance. Therefore, incorporating qualitative assessments alongside IoU could provide a more comprehensive measure of model explainability in clinical contexts.

Ethical Issues

LLMs have performed well in the cervical cancer question-and-answer task, but ethical considerations, such as transparency, data privacy, and algorithmic bias, remain [50]. Tools such as LIME enhance transparency and simplify the explanation of AI decisions, with further progress expected [51]. Deployments adhere to strict data laws to ensure ongoing improvements in privacy, and technological advancements are anticipated to further safeguard patient privacy [52]. Bias issues are managed through explainable AI and methods such as training with multiple multiinstitutional or population data sets, as well as using generative adversarial networks to obtain more representative data [53]. While practical challenges remain in technology integration and staff training, LLMs are more easily adopted due to their application programming interfaces and their ability to act as personalized learning assistants, reducing the reliance on extensive medical staff training [54,55].

Limitations

Our study also has limitations: (1) Because of the limited capabilities of our computers, we were unable to test all existing LLMs. It is possible that there are models with better performance than ChatGPT-4.0 Turbo in handling abnormal cervical screening results. (2) Our study did not include augmented algorithms or corpora that have been shown to improve LLM performance in other studies, as not all patients or physicians are familiar with these tools. The lack of these enhancements may limit the ability of LLMs to demonstrate their full potential in answering medical questions. This absence could have restricted the models from showcasing their full capabilities in medical query resolution, potentially affecting the generalizability of our results in more advanced settings. (3) The study conducted assessments under controlled, structured questions, which may not fully reflect the model's performance in dynamic, real-world clinical settings. This controlled environment may limit our ability to assess the adaptability of LLMs in unpredictable or complex patient interactions.

Conclusions

This study highlights the pivotal role of LLMs, particularly proprietary ones such as ChatGPT-4.0 Turbo, in enhancing clinical decision-making in cervical cancer screening. ChatGPT-4.0 Turbo outperforms both open-source and medical-specialized models in interpreting clinical guidelines and handling medical queries. Such findings are essential for improving the accuracy and efficiency of medical screenings and diagnoses, ultimately enhancing health care delivery and patient care. Further research is needed to assess the effectiveness of LLMs in medical applications, potentially leading to the development of models more tailored for medical practice and advancing overall health care.

XSL•FO
RenderX

Acknowledgments

This work was supported by the Science and Technology Innovation Cultivation Funding of Zhongnan Hospital of Wuhan University (grant CXPY2022049).

Data Availability

The 100 questions developed for model evaluation and all analyzed data in this study are included in the published manuscript and its multimedia appendices.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Computer specifications for deploying models.
[DOCX File , 15 KB - jmir_v27i1e63626_app1.docx]

Multimedia Appendix 2
Cervical cancer screening and abnormal result management process by the CSCCP (Chinese Society for Colposcopy and Cervical Pathology of the China Healthy Birth Science Association).
[DOCX File , 467 KB - jmir_v27i1e63626_app2.docx]

Multimedia Appendix 3
Semantic textual similarity testing for Phi-2 and QiZhenGPT model.
[DOCX File , 154 KB - jmir_v27i1e63626_app3.docx]

Multimedia Appendix 4
Responses from each model and quality assessments.
[DOCX File , 424 KB - jmir_v27i1e63626_app4.docx]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Cohen CM, Wentzensen N, Castle PE, Schiffman M, Zuna R, Arend RC, et al. Racial and ethnic disparities in cervical cancer incidence, survival, and mortality by histologic subtype. J Clin Oncol 2023 Feb 10;41(5):1059-1068 [FREE Full text] [doi: 10.1200/JCO.22.01424] [Medline: 36455190]
3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan;73(1):17-48 [FREE Full text] [doi: 10.3322/caac.21763] [Medline: 36633525]
4. Burmeister CA, Khan SF, Schäfer G, Mbatani N, Adams T, Moodley J, et al. Cervical cancer therapies: current challenges and future perspectives. Tumour Virus Res 2022 Jun;13:200238 [FREE Full text] [doi: 10.1016/j.tvr.2022.200238] [Medline: 35460940]
5. Chavez MR, Butler TS, Rekawek P, Heo H, Kinzler WL. Chat Generative Pre-trained Transformer: why we should embrace this technology. Am J Obstet Gynecol 2023 Jun;228(6):706-711. [doi: 10.1016/j.ajog.2023.03.010] [Medline: 36924908]
6. Kim JK, Chua M, Rickard M, Lorenzo A. ChatGPT and large language model (LLM) chatbots: the current state of acceptability and a proposal for guidelines on utilization in academic medicine. J Pediatr Urol 2023 Jun 02:598-604. [doi: 10.1016/j.jpurol.2023.05.018] [Medline: 37328321]
7. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
8. Knoedler L, Alfertshofer M, Knoedler S, Hoch CC, Funk PF, Cotofana S, et al. Pure wisdom or potemkin villages? A comparison of ChatGPT 3.5 and ChatGPT 4 on USMLE step 3 style questions: quantitative analysis. JMIR Med Educ 2024 Jan 05;10:e51148 [FREE Full text] [doi: 10.2196/51148] [Medline: 38180782]
9. Beaulieu-Jones BR, Berrigan MT, Shah S, Marwaha JS, Lai S, Brat GA. Evaluating capabilities of large language models: performance of GPT-4 on surgical knowledge assessments. Surgery 2024 Apr;175(4):936-942. [doi: 10.1016/j.surg.2023.12.014] [Medline: 38246839]
10. Skalidis I, Cagnina A, Luangphiphat W, Mahendiran T, Muller O, Abbe E, et al. ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story? Eur Heart J Digit Health 2023 May;4(3):279-281 [FREE Full text] [doi: 10.1093/ehjdh/ztad029] [Medline: 37265864]

XSL•FO
RenderX

11. Xie Y, Seth I, Hunter-Smith DJ, Rozen WM, Ross R, Lee M. Aesthetic surgery advice and counseling from artificial intelligence: a rhinoplasty consultation with ChatGPT. Aesthetic Plast Surg 2023 Oct;47(5):1985-1993 [FREE Full text] [doi: 10.1007/s00266-023-03338-7] [Medline: 37095384]
12. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023 Aug;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
13. Hermann CE, Patel JM, Boyd L, Growdon WB, Aviki E, Stasenko M. Let's chat about cervical cancer: assessing the accuracy of ChatGPT responses to cervical cancer questions. Gynecol Oncol 2023 Dec;179:164-168. [doi: 10.1016/j.ygyno.2023.11.008] [Medline: 37988948]
14. Stonehocker J. Cervical cancer screening in pregnancy. Obstet Gynecol Clin North Am 2013 Jun;40(2):269-282. [doi: 10.1016/j.ogc.2013.03.005] [Medline: 23732031]
15. Zhou H, Liu F, Gu B, Zou X, Huang J, Wu J, et al. A survey of large language models in medicine: progress, application, and challenge. arXiv :e1 Preprint posted online on July 22, 2023 [FREE Full text] [doi: 10.48550/arXiv.2311.05112]
16. Zhao Y, Wei L. Expert consensus interpretation on Chinese cervical cancer screening and abnormal management issues by CSCCP. Journal of Practical Obstetrics and Gynecology 2018 Feb 15;34(2):101-104 [FREE Full text]
17. Perkins RB, Guido RS, Castle PE, Chelmow D, Einstein MH, Garcia F, 2019 ASCCP Risk-Based Management Consensus Guidelines Committee. 2019 ASCCP risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors. J Low Genit Tract Dis 2020 Apr;24(2):102-131 [FREE Full text] [doi: 10.1097/LGT.0000000000000525] [Medline: 32243307]
18. Wu L, Li L, Huang M, Li G, Lou G, Wu X, et al. Chinese Society of Clinical Oncology (CSCO) Guidelines for the Diagnosis and Treatment of Cervical Cancer (2023). Beijing, China: People's Medical Publishing House; Aug 01, 2023.
19. Manso L, Ramchandani-Vaswani A, Romero I, Sánchez-Lorenzo L, Bermejo-Pérez MJ, Estévez-García P, et al. SEOM-GEICO clinical guidelines on cervical cancer (2023). Clin Transl Oncol 2024 Nov;26(11):2771-2782. [doi: 10.1007/s12094-024-03604-3] [Medline: 39215938]
20. Zhou H, Wang D, Luo M, Lin Z. FIGO 2018 gynecologic cancer report-interpretation of the cervical cancer guidelines. Zhongguo Shiyong Fuke Yu Chanke Zazhi 2019 Feb 19;35(1):95-103. [doi: 10.19538/j.fk2019010123]
21. Jiang T, Huang S, Luan Z, Wang D, Zhuang F. Scaling sentence embeddings with large language models. arXiv :1 Preprint posted online on July 31, 2023 [FREE Full text] [doi: 10.48550/arXiv.2307.16645]
22. Le Mens G, Kovács B, Hannan MT, Pros G. Uncovering the semantics of concepts using GPT-4. Proc Natl Acad Sci U S A 2023 Dec 05;120(49):e2309350120 [FREE Full text] [doi: 10.1073/pnas.2309350120] [Medline: 38032930]
23. Lozano A, Fleming SL, Chiang C, Shah N. Clinfo.ai: an open-source retrieval-augmented large language model system for answering medical questions using scientific literature. Pac Symp Biocomput 2024;29:8-23 [FREE Full text] [Medline: 38160266]
24. Barr KN, Blomberg T, Liu J, Saraiva LA, Papapetrou P. Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models. New York, NY: IEEE; 2020 Jul 28 Presented at: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, MN, USA. [doi: 10.1109/cbms49503.2020.00009]
25. Madsen A, Reddy S, Chandar S. Post-hoc interpretability for neural NLP: a survey. ACM Comput Surv 2022 Dec 23;55(8):1-42. [doi: 10.1145/3546577]
26. Huang S, Mamidanna S, Jangam S, Zhou Y, Gilpin L. Can large language models explain themselves? A study of LLM-generated self-explanations. arXiv :1 Preprint posted online on October 17, 2023 [FREE Full text] [doi: 10.48550/arXiv.2310.11207]
27. OpenAI. URL: https://chat.openai.com/ [accessed 2025-01-10]
28. Claude. URL: https://claude.ai/ [accessed 2025-01-10]
29. Hugging Face. URL: https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.2 [accessed 2025-01-10]
30. Hugging Face. URL: https://huggingface.co/berkeley-nest/Starling-LM-7B-alpha [accessed 2025-01-10]
31. Google Gemini. URL: https://gemini.google.com/ [accessed 2025-01-10]
32. GitHub. URL: https://github.com/FreedomIntelligence/HuatuoGPT?tab=readme-ov-file [accessed 2025-01-10]
33. HuatuoGPT. URL: https://www.huatuogpt.cn/#/ [accessed 2025-01-10]
34. GitHub. URL: https://github.com/CMKRG/QiZhenGPT?tab=readme-ov-file [accessed 2025-01-10]
35. Hugging Face. URL: https://huggingface.co/microsoft/phi-2 [accessed 2025-01-10]
36. Hugging Face. URL: https://huggingface.co/stanford-crfm/BioMedLM [accessed 2025-01-10]
37. Chen H, Jiao F, Li X, Qin C, Ravaut M, Zhao R, et al. ChatGPT's one-year anniversary: are open-source large language models catching up? arXiv :1 Preprint posted online on January 15, 2024 [FREE Full text] [doi: 10.48550/arXiv.2311.16989]
38. Tsoutsanis P, Tsoutsanis A. Evaluation of large language model performance on the Multi-Specialty Recruitment Assessment (MSRA) exam. Comput Biol Med 2024 Jan;168:107794. [doi: 10.1016/j.compbiomed.2023.107794] [Medline: 38043471]
39. Yang H, Li M, Zhou H, Xiao Y, Fang Q, Zhang R. One LLM is not enough: harnessing the power of ensemble learning for medical question answering. medRxiv :1 Preprint posted online December 24, 2023 [FREE Full text] [doi: 10.1101/2023.12.21.23300380] [Medline: 38196648]

XSL•FO
RenderX

40. Kaplan J, McCandlish S, Henighan T, Brown T, Chess B, Child R, et al. Scaling laws for neural language models. arXiv :1 Preprint posted online on January 23, 2020 [FREE Full text]
41. Liévin V, Hother CE, Motzfeldt AG, Winther O. Can large language models reason about medical questions? Patterns (N Y) 2024 Mar 08;5(3):100943 [FREE Full text] [doi: 10.1016/j.patter.2024.100943] [Medline: 38487804]
42. Sivarajkumar S, Kelley M, Samolyk-Mazzanti A, Visweswaran S, Wang Y. An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. JMIR Med Inform 2024 Apr 08;12:e55318 [FREE Full text] [doi: 10.2196/55318] [Medline: 38587879]
43. Chen B, Zhang Z, Langrené N, Zhu S. Unleashing the potential of prompt engineering in large language models: a comprehensive review. arXiv :1 Preprint posted online on September 5, 2024 [FREE Full text] [doi: 10.48550/arXiv.2310.14735]
44. Lu A, Zhang H, Zhang Y, Wang X, Yang D. Bounding the capabilities of large language models in open text generation with prompt constraints. arXiv :1 Preprint posted online on February 17, 2023 [FREE Full text] [doi: 10.18653/v1/2023.findings-eacl.148]
45. Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning. arXiv :1 Preprint posted online on September 2, 2021 [FREE Full text] [doi: 10.18653/v1/2021.emnlp-main.243]
46. Alvarez-Melis D, Kaur H, Wallach H, Vaughan J. From human explanation to model interpretability: a framework based on weight of evidence. arXiv :1 Preprint posted online on September 20, 2021 [FREE Full text] [doi: 10.1609/hcomp.v9i1.18938]
47. Lazebnik T, Bunimovich-Mendrazitsky S, Rosenfeld A. An algorithm to optimize explainability using feature ensembles. Appl Intell 2024 Feb 01;54(2):2248-2260. [doi: 10.1007/s10489-023-05069-3]
48. Heyen H, Widdicombe A, Siegel N, Perez-Ortiz M, Treleaven P. The effect of model size on LLM post-hoc explainability via LIME. arXiv :1 Preprint posted online on May 8, 2024 [FREE Full text]
49. Chakraborty S, Tomsett R, Raghavendra R, Harborne D, Alzantot M, Cerutti F, et al. Interpretability of deep learning models: a survey of results. 2018 Jun 28 Presented at: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); August 4-8, 2017; San Francisco, CA, USA. [doi: 10.1109/uic-atc.2017.8397411]
50. Zhui L, Fenghe L, Xuehu W, Qining F, Wei R. Ethical considerations and fundamental principles of large language models in medical education: viewpoint. J Med Internet Res 2024 Aug 01;26:e60083 [FREE Full text] [doi: 10.2196/60083] [Medline: 38971715]
51. Dong X, Wang S, Lin D, Rajbahadur G, Zhou B, Liu S, et al. PromptExp: multi-granularity prompt explanation of large language models. arXiv :1 Preprint posted online on October 30, 2024 [FREE Full text]
52. Zhan X, Seymour W, Such J. Beyond individual concerns: multi-user privacy in large language models. 2024 Jul 08 Presented at: CUI '24: Proceedings of the 6th ACM Conference on Conversational User Interfaces; July 8-10, 2024; Luxembourg City, Luxembourg. [doi: 10.1145/3640794.3665883]
53. Ong JCL, Seng BJJ, Law JZF, Low LL, Kwa ALH, Giacomini KM, et al. Artificial intelligence, ChatGPT, and other large language models for social determinants of health: current state and future directions. Cell Rep Med 2024 Jan 16;5(1):101356 [FREE Full text] [doi: 10.1016/j.xcrm.2023.101356] [Medline: 38232690]
54. Hart SN, Hoffman NG, Gershkovich P, Christenson C, McClintock DS, Miller LJ, et al. Organizational preparedness for the use of large language models in pathology informatics. J Pathol Inform 2023;14:100338 [FREE Full text] [doi: 10.1016/j.jpi.2023.100338] [Medline: 37860713]
55. Su Z, Tang G, Huang R, Qiao Y, Zhang Z, Dai X. Based on medicine, the now and future of large language models. Cell Mol Bioeng 2024 Aug 16;17(4):263-277. [doi: 10.1007/s12195-024-00820-3] [Medline: 39372551]

Abbreviations

ASCCP: American Society for Colposcopy and Cervical Pathology
CO-STAR: Context, Objective, Style, Tone, Audience, and Response
CSCCP: Chinese Society for Colposcopy and Cervical Pathology of the China Healthy Birth Science Association
CSCO: Chinese Society of Clinical Oncology
FIGO: International Federation of Gynecology and Obstetrics
GEICO: Grupo Español de Investigación en Cáncer de Ovario
IoU: intersection over union
LIME: Local Interpretable Model-agnostic Explanations
LLM: large language model
SEOM: Sociedad Española de Oncología Médica
USMLE: United States Medical Licensing Examination

XSL•FO
RenderX

Edited by A Coristine; submitted 25.06.24; peer-reviewed by K Keshtkar, S Mao, A Kalluchi; comments to author 27.09.24; revised version received 01.11.24; accepted 11.12.24; published 05.02.25.

Please cite as:
Kuerbanjiang W, Peng S, Jiamaliding Y, Yi Y
Performance Evaluation of Large Language Models in Cervical Cancer Management Based on a Standardized Questionnaire: Comparative Study
J Med Internet Res 2025;27:e63626
URL: https://www.jmir.org/2025/1/e63626
doi:10.2196/63626
PMID:

©Warisijiang Kuerbanjiang, Shengzhe Peng, Yiershatijiang Jiamaliding, Yuexiong Yi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Unveiling GPT-4V's hidden challenges behind high accuracy on USMLE questions: Observational Study

Zhichao Yang[1*], PhD; Zonghai Yao[1*], MS; Mahbuba Tasmin[1], MS; Parth Vashisht[1], MS; Won Seok Jang[2], RN, MS; Feiyun Ouyang[2], PhD; Beining Wang[3], BS; David McManus[4], MD; Dan Berlowitz[5,6], MD, MPH; Hong Yu[1,2,6,7], PhD

[1]College of Information and Computer Science, University of Massachusetts Amherst, Amherst, MA, United States

[2]Miner School of Computer & Information Sciences, University of Massachusetts Lowell, Lowell, MA, United States

[3]Shanghai Medical College, Fudan University, Shanghai, China

[4]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[5]Department of Public Health, University of Massachusetts Lowell, Lowell, MA, United States

[6]Center for Biomedical and Health Research in Data Sciences, University of Massachusetts Lowell, Lowell, MA, United States

[7]Center for Healthcare Organization and Implementation Research, VA Bedford Health Care System, Bedford, MA, United States

[*]these authors contributed equally

Corresponding Author:
Hong Yu, PhD
Miner School of Computer & Information Sciences
University of Massachusetts Lowell
1 University Ave
Lowell, MA, 01854
United States
Phone: 1 508 612 7292
Email: Hong_Yu@uml.edu

Abstract

Background: Recent advancements in artificial intelligence, such as GPT-3.5 Turbo (OpenAI) and GPT-4, have demonstrated significant potential by achieving good scores on text-only United States Medical Licensing Examination (USMLE) exams and effectively answering questions from physicians. However, the ability of these models to interpret medical images remains underexplored.

Objective: This study aimed to comprehensively evaluate the performance, interpretability, and limitations of GPT-3.5 Turbo, GPT-4, and its successor, GPT-4 Vision (GPT-4V), specifically focusing on GPT-4V's newly introduced image-understanding feature. By assessing the models on medical licensing examination questions that require image interpretation, we sought to highlight the strengths and weaknesses of GPT-4V in handling complex multimodal clinical information, thereby exposing hidden flaws and providing insights into its readiness for integration into clinical settings.

Methods: This cross-sectional study tested GPT-4V, GPT-4, and ChatGPT-3.5 Turbo on a total of 227 multiple-choice questions with images from USMLE Step 1 (n=19), Step 2 clinical knowledge (n=14), Step 3 (n=18), the Diagnostic Radiology Qualifying Core Exam (DRQCE) (n=26), and AMBOSS question banks (n=150). AMBOSS provided expert-written hints and question difficulty levels. GPT-4V's accuracy was compared with 2 state-of-the-art large language models, GPT-3.5 Turbo and GPT-4. The quality of the explanations was evaluated by choosing human preference between an explanation by GPT-4V (without hint), an explanation by an expert, or a tie, using 3 qualitative metrics: comprehensive explanation, question information, and image interpretation. To better understand GPT-4V's explanation ability, we modified a patient case report to resemble a typical "curbside consultation" between physicians.

Results: For questions with images, GPT-4V achieved an accuracy of 84.2%, 85.7%, 88.9%, and 73.1% in Step 1, Step 2 clinical knowledge, Step 3 of USMLE, and DRQCE, respectively. It outperformed GPT-3.5 Turbo (42.1%, 50%, 50%, 19.2%) and GPT-4 (63.2%, 64.3%, 66.7%, 26.9%). When GPT-4V answered correctly, its explanations were nearly as good as those provided by domain experts from AMBOSS. However, incorrect answers often had poor explanation quality: 18.2% (10/55) contained inaccurate text, 45.5% (25/55) had inference errors, and 76.3% (42/55) demonstrated image misunderstandings. With human expert assistance, GPT-4V reduced errors by an average of 40% (22/55). GPT-4V accuracy improved with hints, maintaining

XSL•FO
RenderX

stable performance across difficulty levels, while medical student performance declined as difficulty increased. In a simulated curbside consultation scenario, GPT-4V required multiple specific prompts to interpret complex case data accurately.

Conclusions: GPT-4V achieved high accuracy on multiple-choice questions with images, highlighting its potential in medical assessments. However, significant shortcomings were observed in the quality of explanations when questions were answered incorrectly, particularly in the interpretation of images, which could not be efficiently resolved through expert interaction. These findings reveal hidden flaws in the image interpretation capabilities of GPT-4V, underscoring the need for more comprehensive evaluations beyond multiple-choice questions before integrating GPT-4V into clinical settings.

(J Med Internet Res 2025;27:e65146) doi:10.2196/65146

KEYWORDS

artificial intelligence; natural language processing; large language model; LLM; ChatGPT; GPT; GPT-4V; USMLE; Medical License Exam; medical image interpretation; United States Medical Licensing Examination; NLP

Introduction

Using computers to help make clinical diagnoses and guide treatments has been a goal of artificial intelligence (AI) since its inception [1]. The adoption of electronic health record systems by hospitals in the United States has resulted in an unprecedented amount of digital data associated with patient encounters. Computer-assisted clinical diagnostic support systems (CDSSs) endeavor to enhance clinicians' decisions with patient information and clinical knowledge [2]. There is burgeoning interest in CDSS for enhanced imaging [3] in various disciplines such as breast cancer detection [4], COVID detection [5], diagnosing congenital cataracts [6], and hidden fracture location [7]. For a decision to be trustworthy for clinicians, CDSS should not only make the prediction but also provide accurate explanations [8-10]. However, most previous imaging CDSSs only highlight areas deemed significant by AI [11-14], providing limited insight into the explanation of the diagnosis [15].

Recent advances in large language models (LLMs) have encouraged much discussion in health care. State-of-the-art LLMs include GPT-3.5 Turbo, a chatbot released by OpenAI in October 2022, and its successor, GPT-4, released in March 2023. The success of GPT-3.5 Turbo and GPT-4 is attributed to their conversational ability and their performance, which have approached or matched human-level competence in cognitive tasks, spanning various domains including medicine [16]. Both GPT-3.5 Turbo and GPT-4 have achieved commendable results in the United States Medical Licensing Examination (USMLE), leading to discussions about the readiness of LLM applications for integration into clinical [17-19] and educational [20-22] environments.

One limitation of GPT-3.5 Turbo and GPT-4 is that they can only read and generate text and are unable to process other data modalities, such as images. This limitation, known as "single modality," is a common issue among many LLMs [23]. Advancements in multimodal LLMs promise enhanced capabilities and integration with diverse data sources [24-26]. OpenAI's GPT-4V is a state-of-the-art multimodal LLM equipped with image processing and understanding ability [27]. However, the ability of GPT-4V to answer medical questions with images with explanations has not been comprehensively evaluated. In this study, we aimed to expose hidden flaws in GPT-4V's ability to interpret clinical images by thoroughly evaluating its performance on medical licensing examination questions involving image interpretation. For GPT-4V to be useful to medical professionals, it should not only provide correct responses but also offer accurate explanations for its reasoning, especially in complex multimodal clinical scenarios [28].

Methods

Overview

This cross-sectional study aimed to expose the hidden flaws of GPT-4V in clinical image interpretation by comparing the performance between GPT-4V, GPT-4, and GPT-3.5 Turbo in answering medical licensing examination questions. This study also investigates the quality of GPT-4V explanation in answering these questions. The overview of the study is shown in Figure 1. This study was conducted in October 2023.

XSL•FO
RenderX

Figure 1. A summary of the image question selection process and prompt to large language models.

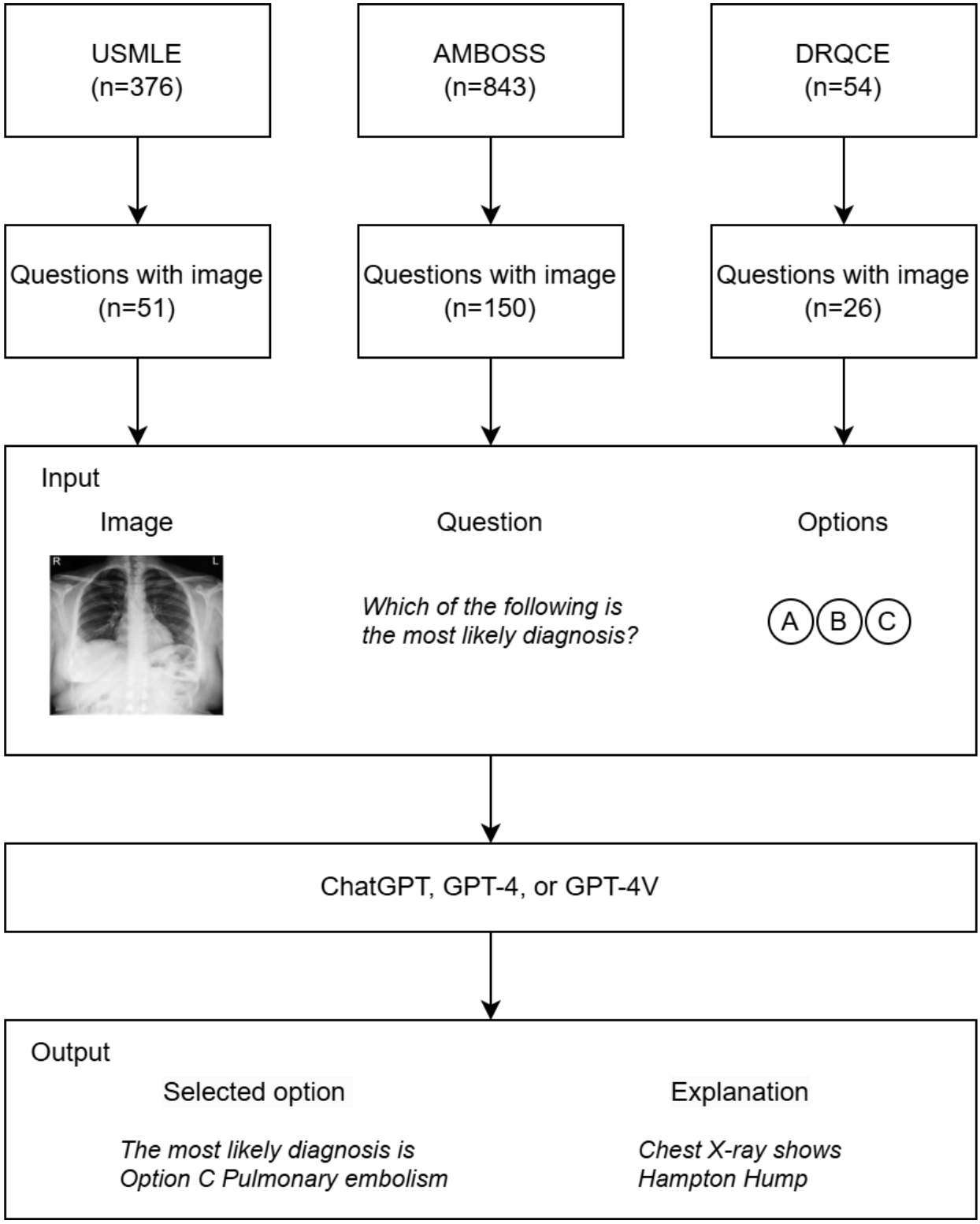

Ethical Considerations

The requirement for ethical approval and informed consent was waived by the institutional review board at the VA Bedford Health Care System because no patient data were used. The experiments were performed in accordance with the Declaration of Helsinki.

Medical Exams and a Patient Case Report Collection

We obtained study questions from 3 sources. USMLE consists of 3 steps required to obtain a medical license in the United States. USMLE assesses a physician's ability to apply knowledge, concepts, and principles, which is critical to both health and disease management and is the foundation for safe, efficient patient care. Step 1 assesses foundational scientific concepts essential for medical practice, Step 2 clinical knowledge (CK) evaluates the application of clinical science

XSL•FO
RenderX

for supervised patient care, and Step 3 tests the medical knowledge required for unsupervised practice. Step 1, Step 2 CK, and Step 3 of the USMLE sample exam released from the National Board of Medical Examiners consist of 119, 120, and 137 questions respectively. We accessed these questions from publicly available links [29]. Each question contained multiple options to choose from. We then selected all questions with images, resulting in 19, 14, and 18 questions from Step 1, Step 2 CK, and Step 3. Medical subdomains include but are not limited to radiology, dermatology, orthopedics, ophthalmology, cardiology, and general surgery.

The sample exam only included limited questions with images. Thus, we further collected similar questions from AMBOSS, a widely used question bank for medical students, which provides students' performance on the exam. The performance enabled us to assess the comparative effectiveness of the model. For each question, AMBOSS associated an expert-written hint to tip the student to answer the question and a difficulty level that ranges from 1-5. Levels 1, 2, 3, 4, and 5 represent the easiest 20%, 20%-50%, 50%-80%, 80%-95%, and 95%-100% of questions respectively [30]. Hints are designed to guide students to the correct answer. They are typically formatted as a short paragraph that describes the image. We manually checked that no hint had disclosed the answer directly. In addition to the gold standard choice, each answer is associated with a detailed explanation by AMBOSS. They were developed through an internal peer-review process involving more than 50 physicians who achieved high scores in the exam. We used a commercial license to access the questions. Since AMBOSS is not publicly available and its licensing terms restrict the automatic website scraping of its proprietary content, they are not in the CommonCrawl data set used to train GPTs [31]. We randomly selected and manually downloaded 10 questions from each of the 5 difficulty levels. We repeated this process for Step 1, Step 2 CK, and Step 3. This resulted in a total number of 150 questions.

In addition, we collected questions from the Diagnostic Radiology Qualifying Core Exam (DRQCE) [32], which is an image-rich exam to evaluate a candidate's foundational knowledge and clinical judgment across practice domains of diagnostic radiology, which is offered after 36 months of residency training. Since DRQCE is proprietary, we used a commercial license to access the 26 questions with images of 54 questions in the preparation exam offered by the American Board of Radiology. In total, we had 227 questions with images from the 3 aforementioned sources.

To illustrate GPT-4V's potential as an imaging diagnostic support tool and further expose its limitations, we used part of a patient case report [33] to resemble a typical "curbside consultation" between medical professionals [34]. In this case, the patient's admission info, such as history of present illness, labs, and images of the case report will be presented to both a physician and GPT-4V. The physician can then work with GPT-4V through question answering, for example, by asking GPT-4V to help interpret images, for the final clinical diagnosis.

How to Answer Image Questions Using GPT-4V Prompts

GPT-4V took image and text data as inputs to generate textual outputs. Given that input format (prompt) played a key role in optimizing model performance, we followed the standard prompting guidelines of the visual question-answering task [35]. Specifically, we prompted GPT-4V by first adding the image, then appending context (ie, patient information) and questions, and finally providing multiple-choice options, each separated by a new line. An example user prompt and GPT-4V response are shown in Figure S1 in Multimedia Appendix 1. When multiple subimages existed in the image, we uploaded multiple subimages to GPT-4V. We did not append a hint to the end of the question, unless other specified. The response consists of the selected option as an answer, supported by a textual explanation to substantiate the selected decision. When using GPT-3.5 Turbo and GPT-4 models that cannot handle image data, images were omitted from the prompt. These models were accessed through OpenAI application programming interfaces. Responses were collected from the September 25, 2023, version of models.

Evaluation Metrics

For answer accuracy, we evaluated the model's performance by comparing the model's choice with the correct choice provided by the exam board or question bank website. We defined accuracy as the ratio of the number of correct choices to the total number of questions.

We also evaluated the quality of the explanation by preference from 3 health care professionals (1 medical doctor with 35 years of experience in internal medicine, 1 registered ward nurse with 2 years of experience, and 1 third-year medical school student). For each question from the AMBOSS data set (n=150), we first asked the health care professionals to choose their preference between an explanation by GPT-4V (without hint), an explanation by an expert, or a tie without knowing the correctness of GPT-4V's answers. The exclusion of correctness is to avoid bias in their preference of explanations. In addition, the source of the explanations was blinded to the health care professionals, ensuring that their judgments were not influenced by knowing whether an explanation came from GPT-4V or an expert.

In addition, we also asked health care professionals to evaluate the GPT-4V explanation from a sufficient and comprehensive perspective [36,37]. They determined if the information exists in the explanation, that consists of (1) image interpretation: GPT-4V tried to interpret the image in the explanation, and such interpretation is sufficient to support its choice; (2) question information: explanations contained information related to the textual context (ie, patient information) of the question, and such information was essential for GPT-4V's choice; (3) comprehensive explanation: the explanation included comprehensive reasoning for all possible evidence (eg, symptoms, lab results) that leads to the final answer.

Finally, for each question answered incorrectly, we asked health care professionals to check if the explanation contained any errors that consisted of (1) image misunderstanding (if the

XSL•FO
RenderX

sentence in the explanation showed an incorrect interpretation of the image; eg, GPT-4V said that a bone in the image was for the hand, but it was in fact the foot); (2) text hallucination (if the sentence in the explanation contained made-up information [38]; eg, claiming Saxenda was insulin); (3) reasoning error (if the sentence did not properly infer knowledge in either image or text to an answer; eg, GPT-4V reasoned that a patient took a trip within the last 3 months and therefore diagnosed the patient as having Chagas disease, despite the clinical knowledge that Chagas disease usually develops 10-20 years after infection); or (4) nonmedical error (GPT is known to struggle with tasks requiring precise spatial localization, such as identifying chess positions on the board [27]).

In this study, we asked an internal medicine doctor with 35 years of experience to articulate a detailed rating guideline above. Our study has shown that the medical student and nurse, both of whom participated independently, agreed with the doctor's ratings of 95% and 86%, respectively. This high agreement ratio underscores the effectiveness of the standardized guidelines in ensuring consistent evaluation across varying levels of expertise.

Statistical Analysis

Chi-square tests and pairwise comparisons with Bonferroni corrections were used for the performance metrics of GPT-3.5 Turbo, GPT-4, and GPT-4V on visual question answering exams. GPT-4V's accuracies on the AMBOSS data set were compared between different difficulties using unpaired chi-square tests with a significance level of 0.05. All analysis was conducted in Python software (version 3.10.11; Python Software Foundation).

Results

Overall Answer Accuracy

For all questions in the USMLE sample exam (including ones without images), GPT-4V achieved an accuracy of 88.2%, 90.8%, and 92.7% among Step 1, Step 2 CK, and Step 3 of USMLE questions, respectively. In comparison, GPT-3.5 Turbo and GPT-4 achieved an accuracy of 55.1% and 81.5% in Step 1, 59.1% and 80.8% in Step 2 CK, and 60.9% and 88.3% in Step 3, respectively (Table 1). GPT-4V outperformed GPT-4 and GPT-3.5 Turbo by 11.3% (95% CI 11.5%-11.1%; $P<.001$) and 32% (95% CI 32.3%-31.7%; $P<.001$). The score of GPT-4V passes the standard for the USMLE (about 60%). The performance of GPT-4V across different subdomains is shown in Table S1 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Performance of GPT-3.5 Turbo, GPT-4, and GPT-4V on a USMLE sample exam from the National Board of Medical Examiners without hints.

Exam name and agents	Performance	
	Questions with image, n (%)	All questions, n (%)
USMLE[a] sample exam-Step 1[b]		
Sample size	19	119
GPT-3.5 Turbo	8 (42.1)	66 (55.1)
GPT-4	12 (63.2)	97 (81.5)
GPT-4V	16 (84.2)	105 (88.2)
USMLE sample exam-Step 2 clinical knowledge[c]		
Sample size	14	120
GPT-3.5 Turbo	7 (50)	71 (59.1)
GPT-4	9 (64.3)	97 (80.8)
GPT-4V	12 (85.7)	109 (90.8)
USMLE sample exam-Step 3[d]		
Sample size	18	137
GPT-3.5 Turbo	9 (50)	73 (60.9)
GPT-4	12 (66.7)	121 (88.3)
GPT-4V	16 (88.9)	127 (92.7)
DRQCE[e] sample exam[f]		
Sample size	26	54
GPT-3.5 Turbo	5 (19.2)	31 (57.4)
GPT-4	7 (26.9)	35 (64.8)
GPT-4V	19 (73.1)	48 (88.9)

[a]USMLE: United States Medical Licensing Examination.

[b]19 questions with images and 119 questions in total in Step 1.

[c]14 questions with images and 120 questions in total in Step 2 CK.

[d]There were 18 questions with images and 137 questions in total in Step 3.

[e]DRQCE: Diagnostic Radiology Qualifying Core Exam.

[f]There were 26 questions with images and 54 questions in total in DRQCE.

For questions with images, GPT-4V achieved an accuracy of 84.2%, 85.7%, and 88.9% in Step 1, Step 2 CK, and Step 3 of USMLE questions, respectively. It outperformed GPT-3.5 Turbo and GPT-4 by 42.1% (8/19; 95% CI 36.8%-47.4%; *P*<.001) and 21.1% (4/19; 95% CI 7.8-34.2%; *P*=0.01) in Step 1, 35.7% (5/14; 95% CI 3.1%-39.7%; *P*=.03) and 21.4% (3/14; 95% CI 4.7%-38.1%; *P*=.02) in Step 2 CK, 38.9% (7/18; 95% CI 32.2%-45.7%; *P*<.001) and 22.2% (4/18; 95% CI 5.5%-38.9%; *P*=.02) in Step 3, respectively. Similarly, GPT-4V achieved an accuracy of 73.1%, outperforming GPT-3.5 Turbo by 53.9% (14/26; 95% CI 41.6%-66.2%; *P*<.001) and GPT-4 by 46.2% (12/26; 95% CI 29.8%-62.5%; *P*<.001) in DRQCE (Table 1). This highlights the superior ability of GPT-4V to interpret clinical images compared with earlier versions.

Impact of Difficulty Level and Use of Hints

When asking GPT-4V questions without a hint, it achieved an accuracy of 60%, 64%, and 66% for AMBOSS Step 1, Step 2 CK, and Step 3, respectively (Table 2). GPT-4V was in the 72nd, 76th, and 80th percentile with AMBOSS users who were preparing for Step 1, Step 2 CK, and Step 3, respectively. When asking GPT-4V questions with a hint, it achieved an accuracy of 84%, 86%, and 88% for AMBOSS Step 1, Step 2 CK, and Step 3, respectively. Figure S2 in Multimedia Appendix 1 is an example where GPT-4V switched the answer from incorrect to correct when a hint was provided. GPT-4V predictions on the entire AMBOSS data set with images are reported in Table S2 in Multimedia Appendix 1 (n=646). Conclusions drawn from automatic evaluation align with our findings presented in Table 2 (n=150).

XSL•FO
RenderX

Table 2. Performance of GPT-4V on AMBOSS.

AMBOSS steps and hint availability	GPT-4V accuracy on AMBOSS, %					
	Overall (n=50)	1 (n=10)	2 (n=10)	3 (n=10)	4 (n=10)	5 (n=10)
Step 1						
Without hint	60	70	70	30	70	60
Expert hint	84	80	80	80	90	90
Step 2 clinical knowledge						
Without hint	64	80	70	70	50	50
Expert hint	86	100	90	100	70	70
Step 3						
Without hint	66	80	90	60	50	50
Expert hint	88	90	90	90	90	80

Figure 2 shows a decreasing trend in GPT-4V's performance in the AMBOSS data set when the difficulty of questions increased (P=.04) without a hint. However, with the hint, the performance of GPT-4V plateaued across 5 difficulty levels. Importantly, the accuracies of both GPT-4V, with or without a hint, in general, outperformed the accuracies of medical students, and the gap between the performance of GPT-4V and medical students increased when the difficulty increased. On the most difficult questions, GPT-4V with hint outperformed medical students by 60% (18/30, 95% CI 56.8%-63.1%; P<.001), and GPT-4V without hint outperformed medical students by 26.7% (8/30, 95% CI 24.2%-29.3%; P<.001). The findings show that while GPT-4V outperforms medical students in accuracy, its performance is largely dependent on context-based hints, reflecting a fundamental flaw in image reasoning.

Figure 2. Performance of GPT-4V and students on 150 AMBOSS questions with different difficulty levels.

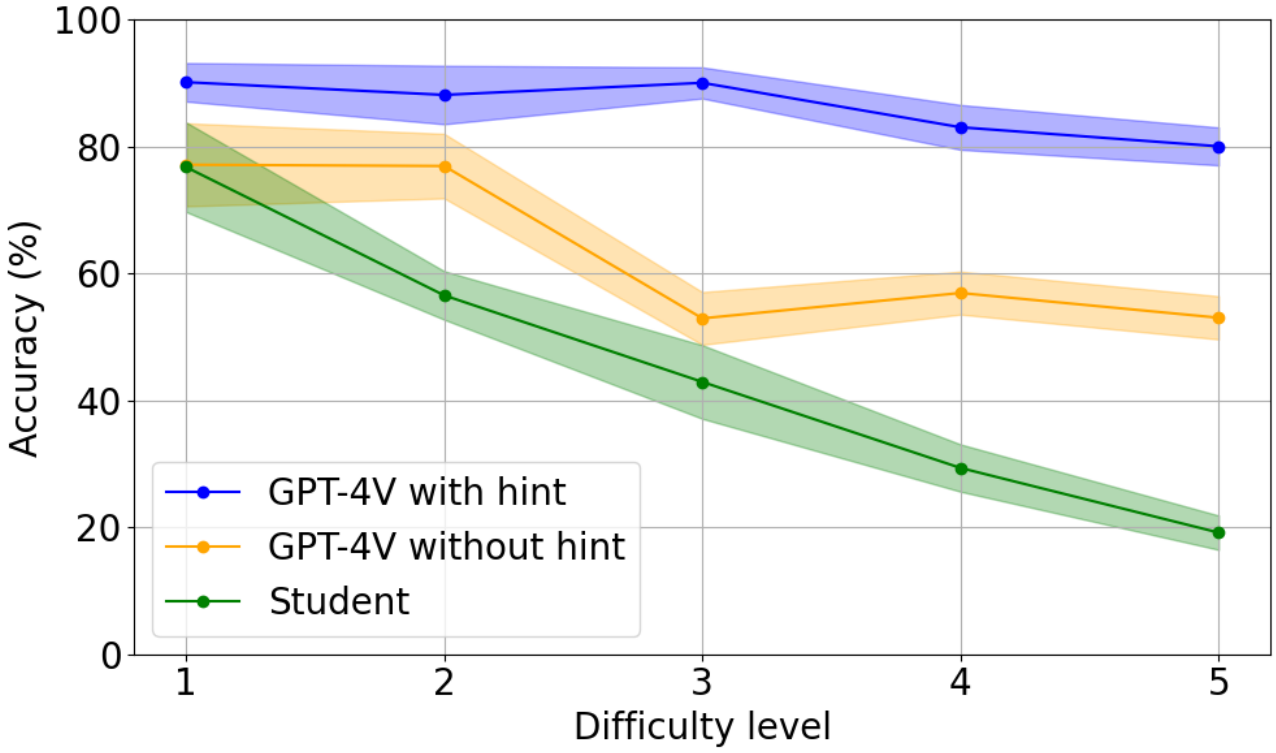

Quality of Explanation

We evaluated the user's preference among GPT-4V-generated explanations and expert-generated explanations. When GPT-4V answered incorrectly, our results show that health care professionals overwhelmingly preferred expert explanations as shown in Table 3. In total, 47 preferred experts and 0 preferred GPT-4V. When GPT-4V answered correctly, the quality of GPT-4V-generated explanations was close to expert-generated explanations: out of 95 votes, 19 preferred experts, 15 preferred GPT-4V, and 61 preferred either. The preference for expert explanations in incorrect answers highlights key weaknesses in GPT-4V's ability to interpret clinical images accurately and offer dependable reasoning.

XSL•FO
RenderX

Table 3. Health care professionals preferred explanations for 150 AMBOSS questions.

AMBOSS steps and correctness of GPT-4V (without hint) responses	Health care professionals' preference		
	Prefer expert	Ties	Prefer GPT-4V
Step 1			
Correct	4	23	3
Incorrect	16	4	0
Step 2 clinical knowledge			
Correct	10	15	7
Incorrect	18	0	0
Step 3			
Correct	5	23	5
Incorrect	13	4	0

We further evaluated the quality of the GPT-4V generated explanation by verifying if the explanation includes image and question text interpretation in Table S3 in Multimedia Appendix 1. When examining the 95 correct answers, 84.2% (n=80) of the responses contained an interpretation of the image, while 96.8% (n=92) aptly captured the information presented in the question. On the other hand, for the 55 incorrect answers, 92.8% (n=51) interpreted the image, and 89.1% (n=49) depicted the question's details. In terms of comprehensiveness, GPT-4V offered a comprehensive explanation in 79% (n=75) of correct responses. In contrast, only 7.2% (n=4) of the wrong responses had a comprehensive explanation that led to the GPT-4V's choice.

We also evaluated the explanations of incorrect responses by GPT-4V image and grouped them into 4 categories, that are image misunderstanding, text hallucination, reasoning error, and nonmedical error. Among GPT-4V responses with wrong answers (n=55), we found that 76.3% (n=42) of responses included misunderstanding of the image, 45.5% (n=25) of responses included logic error, 18.2% (n=10) of responses included text hallucination, and no responses included nonmedical errors.

A Case Study of Curbside Consultation

We present a clinical case study involving a 45-year-old woman with hypertension and altered mental status. As shown in Figure S3 in Multimedia Appendix 1, a collaborative design of GPT-4V allows communication between GPT-4V and physicians. In this scenario, when asked to interpret a CT scan, GPT-4V initially provided an irrelevant answer. GPT-4V needed 5 additional physician-guided prompts to list potential diagnoses, including primary aldosteronism, hypertension, and Cushing's syndrome. For instance, when the physician specifically prompted, "If I suspect Cushing's syndrome due to ectopic ACTH secretion, what would be the next steps to evaluate this patient to determine the source of the hormonal abnormality?" and pointed to a specific area on the CT scan, GPT-4V was then able to respond correctly. This interaction indicates that GPT-4V struggles to autonomously interpret medical images, requiring continuous and specific prompts for accurate interpretation, which underscores its flaws in independent image reasoning.

Discussion

Principal Findings

Recent advancements in medical question-answering systems have leveraged domain-specific transformer models. Early models such as PubMedBERT [39] with 100 million parameters score around 38.3% in USMLE. The introduction of larger models marked a substantial improvement. JMLR [40] with 13 billion parameters, Med-Palm [41] with 540 billion parameters, and GPT-4 achieves 62.5%, 86.2%, and 90.2% respectively. However, previous works only tested these models on text-only questions without images [20,42-44] or questions in non-English languages [45,46]. Unlike previous works that focus primarily on accuracy [47,48], we emphasize explanation quality as a crucial metric for assessing the model's clinical applicability. In particular, we evaluated GPT-4V's ability to interpret medical images (a new feature) to highlight hidden flaws in clinical image interpretation.

We found that GPT-4V outperformed both GPT-3.5 Turbo and GPT-4 (Table 1). When evaluating all questions in the USMLE sample exam, GPT-4V achieved an accuracy of 90.7% outperforming GPT-3.5 Turbo (58.5%) and GPT-4 (83.8%). In comparison, medical students can pass the USMLE exam with more than 60% accuracy, indicating that the GPT-4V performed at a level similar to or above a medical student in the final year of study. The accuracy of GPT-4V highlights its grasp over biomedical and clinical sciences, essential for medical practice, and showcases its ability in patient management and problem-solving skills [49]. Other studies further demonstrated the potential for clinical routines, such as summarizing radiology reports [50] and differential diagnosis [51,52].

For medical exam questions with images, we found that GPT-4V achieved an accuracy of 62%, which was equivalent to the 70th-80th percentile with AMBOSS medical students. This finding indicates that GPT-4V has the capability to integrate information from both text and images to answer questions, making it a promising tool for answering clinical questions based on images. However, our evaluation also reveals hidden flaws in its image interpretation, particularly in its inconsistency and the need for extensive context to provide accurate answers.

XSL•FO
RenderX

Another important finding is that GPT-4V significantly outperformed medical students for questions considered difficult for the students. Specifically, our results, as shown in Figure 2, show that while medical students' performance linearly decreased when the difficulty of questions increased, GPT-4V's performance stayed relatively stable. When expert hints were provided, GPT-4V's performance stayed plateau among questions in all difficult levels. This consistent performance indicates that GPT-4V effectively addresses questions that medical students find challenging. Its advanced capabilities suggest potential as an educational assistant, particularly for complex topics. Under the supervision of teachers' hints, medical students could benefit from its advanced capabilities to understand and analyze complex medical questions.

There may be multiple factors that contribute to GPT-4V's performance on difficult questions. Instrument methods (eg, item response theory [53]) have been typically used for the construction and evaluation of measurement scales and tests. For example, item response theory uses a statistical model that links an individual person's responses to individual test items (questions on a test) to the person's ability to correctly respond to the items and the items' features. Therefore, medical examination test sets have been specifically selected and tailored to medical students' performance with the intended distribution where the performance decreases when the difficulty level increases. Although more evaluation is needed to draw the conclusion that GPT-4V substantially outperformed medical students in difficult questions, our results at least show that GPT-4V performed differently.

On the other hand, we found that GPT-4V's performance was inconsistent among different medical subdomains. As shown in Table S1 in Multimedia Appendix 1, GPT-4V achieved high accuracy on subdomains such as immunology (5/5, 100%), otolaryngology (6/6, 100%), and pulmonology (6/8, 75%), and low accuracy on others such as anatomy (1/4, 25%), emergency medicine (1/4, 25%), and pathology (5/10, 50%). This suggests that while GPT-4V shows potential in some specialties or subdomains, it may require further development to be reliable across the board. The uneven performance highlights the need for tailored approaches to enhancing the model's capabilities where it falls short.

Another advantage of GPT-4V is its ability to explain its image content. Previous studies have shown limited use of current CDSS as most of them offered limited decision explanations and thus gained limited trust among physicians (unlike their colleagues) [54-57]. In contrast, GPT-4V has the potential to improve the effectiveness and credibility of CDSS by providing explanations preferred by experts. As our results indicate, the quality of explanations generated by GPT-4V, when answering correctly, is close to that of expert-generated explanations. Although in more complex scenarios (such as in our curbside consult setting), GPT-4V currently requires continuous highly specialized guidance, which temporarily prevents it from enhancing physician work efficiency, this feature still has the potential to encourage physicians to adopt and use GPT-4V more confidently and broadly.

In terms of explanation quality, we found that more than 80% of responses from GPT-4V provided an interpretation of the image, regardless of whether the responses were correct or not. This suggests that GPT-4V consistently takes into account the image while generating responses. Figure S1 in Multimedia Appendix 1 illustrates an example of a high-quality explanation that uses images to answer a hard question. In this example, more than 70% of students answered incorrectly on the first try, because both bacterial pneumonia and pulmonary embolism may involve symptoms such as cough. To differentiate them, GPT-4V correctly interpreted the x-ray with a radiologic sign of Hampton hump, which further increased the suspicion of pulmonary infarction rather than pneumonia [58]. To show the need for an x-ray as mentioned in the explanation, we removed the image from the input, and GPT-4V switched the answer to bacterial pneumonia while also acknowledging the possibility of pulmonary infarction. This change in response demonstrated the high quality of the GPT-4V explanation, as its explanation about x-rays was not fictional and it truly needed the x-ray to answer this question.

On the other hand, we found that the quality of generated explanations was poor when GPT-4V answered incorrectly. Manual analyses by health care professionals concluded that image misunderstanding was the primary reason why GPT-4V answered incorrectly. Out of 55 wrong responses, 42 (76.3%) were due to misunderstanding of the image. In comparison, only 18.2% (10/55) of the mistakes were attributed to text misinterpretation. Clearly, GPT-4V's proficiency in processing images was considerably lagging behind its text-handling capability. This gap in capability suggests that GPT-4V's advancements in image understanding remain nascent and require significant refinement to align with its text analysis capabilities. To circumvent its image interpretation issue, we additionally prompted GPT-4V with a short hint that described the image. We found that 40% (22/55) of responses switched to the correct answer. One potential future direction involves strengthening GPT-4V's domain-specific knowledge by integrating extensive clinical datasets into its training. For example, employing domain-adaptive pretraining methods—such as those used in MEDITRON [59], which leverages medical guidelines and specialized clinical corpora—could significantly improve the model's understanding of medical concepts, leading to more precise and contextually relevant explanations. In addition, incorporating retrieval-augmented generation based on domain-specific corpora [40] would enable the model to access and retrieve pertinent clinical information during inference, grounding its explanations in verified data. This could improve factual accuracy and reduce the likelihood of incorrect or unsupported responses. Together, these strategies aim to bolster the model's capacity to provide high-quality, accurate explanations, thereby enhancing its overall reliability and usefulness in clinical applications.

Creating these image-related hints requires clinical expertise, limiting the use of GPT-4V as a CDSS. In our case study, when GPT-4V delivered an irrelevant response, the physician needed to come up with correct hints for GPT-4V. These findings reveal a key limitation: GPT-4V's reliance on external guidance from

XSL•FO
RenderX

experts to interpret complex image content effectively, thereby exposing its inability to operate independently in clinical scenarios. Efforts improving GPT-4V on images include multimodal LLMs with reinforcement learning from human feedback to align the outputs of LLMs with physicians' intentions and expectations. This alignment is critical not only for enhancing the accuracy and relevance of the responses but also for integrating GPT-4V seamlessly into clinical environments where time is of the essence [60].

Another significant drawback of GPT-4V involved its tendency to produce factually inaccurate responses, a problem often referred to as the hallucination effect, which is prevalent among many LLMs such as GPT-4V [38]. We found that more than 18% of GPT-4V explanations contain hallucinations, potentially misleading or distracting physicians, particularly the less experienced medical students and residents. This finding emphasizes the need for robust evaluation and correction mechanisms to minimize hallucinations, which are critical to ensure GPT-4V's reliability and safety in clinical practice. One future direction is to integrate GPT-4V and a probabilistic model with CI and citations from credible sources to show the reliability of the response [40,61,62]. The confidence score could also help prioritize the list of differential diagnoses, making it clearer to the physician which conditions are more probable. Thereby reducing the risk of confusion and enhancing the reliability of the CDSS response when additional physician review is warranted [15].

Limitations

This study has several limitations. First, our findings are constrained in their applicability due to the modest sample size. We gathered 227 questions from a total of 28 subdomains or specialties that included images, which might not comprehensively represent all medical disciplines. The small number of questions in each subdomain may not be sufficient to conclude that GPT-4V's performance is inconsistent between medical subdomains. Second, the exams used to test GPT-4V are written in English. Future work could explore other languages. Third, the models used for evaluation were from September 2023, and frontier models may have evolved since then, potentially impacting the results. Fourth, we cannot guarantee that OpenAI strictly adhered to licensing terms when determining which content was included or excluded from their training sets; therefore, even though AMBOSS is not publicly available and its licensing terms restrict the automatic website scraping of its proprietary content, GPT may have already seen the data during training, potentially impacting the results. Finally, while GPT-4V has demonstrated proficiency in medical license examination, its CDSS ability remains untested. Future work could explore continued training GPT-4V in the medical domain for better CDSS integration. Medical exams provide options, but such options would rarely be provided by physicians during CDSS. Our study highlights the inherent limitations in GPT-4V's image interpretation abilities, particularly without expert guidance. We showed that GPT-4V can reduce errors with expert hints, but in more realistic clinical environments, it required continuous highly specialized guidance to make partially correct diagnoses and subsequent examination recommendations, revealing limitations in its autonomous decision-making capabilities. Therefore, more cases with clinician questions should be explored to confirm our findings before clinical integration. Extrapolating the efficacy of GPT-4V to broader clinical applications requires appropriate benchmarks and further research.

Regarding ethical considerations, deploying AI systems for medical advice poses significant ethical implications, especially in medical education and clinical decision-making. Incorrect AI-generated explanations risk disseminating misinformation that could misguide medical professionals, impacting patient safety and treatment outcomes. This is particularly concerning when AI is used in training settings, as it could shape the decision-making abilities of future healthcare providers in potentially harmful ways. Integrating AI into clinical workflows also raises broader societal concerns. While AI has the potential to enhance healthcare efficiency, it could alter patient care dynamics and physician roles. Overreliance on AI may reduce direct physician-patient communication, eroding trust and undermining the relationship-building essential for effective care. Physicians might also become too dependent on AI, potentially compromising their clinical judgment and their ability to critically assess AI-generated insights. Thus, integrating AI in a manner that complements human expertise (supporting rather than replacing health care providers) is vital. Moreover, current benchmarks, including the one in our study, do not fully assess an AI's capabilities for real-world clinical decision-making. Although some LLMs perform well on benchmarks, they lack the comprehensive clinical skills and nuanced understanding required to navigate complex medical scenarios effectively. Viewing these AI models as tools that assist rather than replace clinicians is crucial to ensuring their safe and beneficial use in health care. A responsible approach is needed when deploying AI for medical advice, one that ensures ethical standards are maintained. Issues such as privacy, bias, and the broader implications of AI in society must guide the development and implementation of these systems. By enhancing data diversity, ensuring privacy, and fostering a transparent understanding of AI's role, we can work toward ethical advancements in health care that enhance outcomes without compromising human oversight or patient trust. Future work should focus on developing AI technologies that are fully aligned with health care professionals, maintaining a collaborative and ethically sound approach to their integration.

Conclusion

In this study, GPT-4V demonstrated remarkable overall accuracy on the medical licensing examination and provided high-quality explanations when correct. The evaluation of questions with images (a relatively novel feature for GPT models) allowed us to expose hidden flaws in GPT-4V's image interpretation abilities, offering a unique insight into its strengths and weaknesses. Its performance on image-related questions ranged from 60% to 88%, while physician misdiagnosis rates can be as high as 40% [63,64]. GPT-4V substantially outperformed medical students on difficult questions, but we observed severe issues in its explanations and reasoning, including hallucinations, errors, and misinterpretations. These findings reveal significant challenges in GPT-4V's ability to independently interpret and reason through complex image-based questions, which is crucial

XSL•FO
RenderX

for clinical applications. Despite its strong performance on multiple-choice questions, GPT-4V may still encounter comprehension or explanation errors. When assisted by human experts, GPT-4V reduced some errors with image-related hints. However, in realistic curbside consult settings, continuous and highly specialized prompting was still required, making it time-consuming and limiting its utility as a clinical decision support system in real-world clinical practice. Table 4 lists the summary of key findings.

Table 4. Summary of key findings.

Metric	Findings
Accuracy of image-based questions	GPT-4V achieved 84.2% in Step 1, 85.7% in Step 2 CK[a], 88.9% in Step 3, and 73.1% in DRQCE[b], outperforming GPT-3.5 Turbo (42.1%, 50%, 50%, 19.2%) and GPT-4 (63.2%, 64.3%, 66.7%, 26.9%).
Explanation quality	When GPT-4V provided correct answers, its explanations were almost on par with those given by domain experts. However, for incorrect responses, the explanation quality was often lacking: 18.2% included inaccurate information, 45.5% involved inferencing mistakes, and 76.3% reflected misinterpretations of images.
The impact of human expert hints	There is a decreasing trend in GPT-4V's performance in the AMBOSS dataset when the difficulty of questions increased) without hint. However, with the hint, the performance of GPT-4V plateaued.
Performance of GPT-4V on most difficult questions	GPT-4V with hints outperformed medical students by 60%, and GPT-4V without hints outperformed medical students by 26.7%.

[a]CK: clinical knowledge.

[b]DRQCE: Diagnostic Radiology Qualifying Core Exam.

Overall, our findings emphasize the need for a more comprehensive evaluation of GPT-4V's multimodal capabilities, especially in clinical image interpretation, before considering its integration into clinical decision support systems. Future randomized clinical trials will help further verify the actual utility of GPT-4V and promote more extensive and profound integration of AI in the medical domain.

Acknowledgments

This work was funded by the grant R01MH125027 from the National Institute of Mental Health of the National Institutes of Health. The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The contents of this paper do not represent the views of the National Institutes of Health.

Data Availability

The data sets generated during and/or analyzed during this study are available in the USMLE [29], AMBOSS [30], and DRQCE [32] repositories.

Authors' Contributions

HY initialized the conceptualization of the project. Yang Z and Yao Z designed the study. Yang Z, Yao Z, MT, and PV implemented the methods. Yang Z, Yao Z, WJ, FO, BW, and DB performed the data analysis. Yang Z and Yao Z interpreted the results with substantial input from DM, DB, and HY. All authors contributed to manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and tables.
[DOCX File , 1884 KB - jmir_v27i1e65146_app1.docx]

References

1. Shortliffe EH, Cimino JJ. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. New York: Springer; Dec 02, 2014.
2. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]

XSL•FO
RenderX

3. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med 2023;388(21):1981-1990. [doi: 10.1056/NEJMra2301725] [Medline: 37224199]
4. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021;4(1):65 [FREE Full text] [doi: 10.1038/s41746-021-00438-z] [Medline: 33828217]
5. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 2020;10(1):19549 [FREE Full text] [doi: 10.1038/s41598-020-76550-z] [Medline: 33177550]
6. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 2017;1(2):0024. [doi: 10.1038/s41551-016-0024]
7. Rayan JC, Reddy N, Kan JH, Zhang W, Annapragada AV. Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making. Radiol Artif Intell 2019;1(1):e180015 [FREE Full text] [doi: 10.1148/ryai.2019180015] [Medline: 33937781]
8. Bussone A, Stumpf S, O?Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. 2015 Presented at: International Conference on Healthcare Informatics. Published online; 2015 October 23; USA. [doi: 10.1109/ichi.2015.26]
9. Panigutti C, Beretta A, Giannotti F, Pedreschi D. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. 2022 Presented at: CHI Conference on Human Factors in Computing Systems. Published online 2022; 2022 April 29; USA. [doi: 10.1145/3491102.3502104]
10. Gaube S, Suresh H, Raue M, Lermer E, Koch TK, Hudecek MFC, et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. Sci Rep 2023;13(1):1383 [FREE Full text] [doi: 10.1038/s41598-023-28633-w] [Medline: 36697450]
11. Singh, A, Mohammed AR, Zelek J, Lakshminarayanan V. Interpretation of deep learning using attributions: application to ophthalmic diagnosis. In: Optical Engineering + Applications 2020;11511:11.
12. Eitel F, Ritter K. Testing the robustness of attribution methods for convolutional neural networks in MRI-based alzheimer?s disease classification. In: Inuzuki K, Reyes M, Syeda-Mahmood T, editors. Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support. Springer International Publishing. New York: Springer International Publishing; 2019.
13. Papanastasopoulos Z, Samala RK, Chan HP, Hadjiiski L, Paramagul C, Helvie MA, et al. Explainable AI for medical imaging: deep-learning CNN ensemble for classification of estrogen receptor status from breast MRI. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11314/2549298/Explainable-AI-for-medical-imaging--deep-learning-CNN-ensemble/10.1117/12.2549298.short [accessed 2020-02-16]
14. Shamout FE, Shen Y, Wu N, Kaku A, Park J, Makino T, et al. An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. NPJ Digit Med 2021;4(1):80 [FREE Full text] [doi: 10.1038/s41746-021-00453-0] [Medline: 33980980]
15. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023;307(2):e230163. [doi: 10.1148/radiol.230163] [Medline: 36700838]
16. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 Technical Report. ArXiv. 2023;abs/2303.0 2023. [doi: 10.48550/arXiv.2303.08774]
17. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
18. Decker H, Trang K, Ramirez J, Colley A, Pierce L, Coleman M, et al. Large language model-based chatbot vs surgeon-generated informed consent documentation for common procedures. JAMA Netw Open 2023;6(10):e2336997 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36997] [Medline: 37812419]
19. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
20. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
21. Thirunavukarasu AJ, Hassan R, Mahmood S, Sanghera R, Barzangi K, El Mukashfi M, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. JMIR Med Educ 2023;9:e46599 [FREE Full text] [doi: 10.2196/46599] [Medline: 37083633]
22. Cooper AZ, Rodman A. AI and medical education - A 21st-century pandora's box. N Engl J Med 2023;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
23. Khader F, Müller-Franzes G, Wang T, Han T, Tayebi Arasteh S, Haarburger C, et al. Multimodal deep learning for integrating chest radiographs and clinical parameters: a case for transformers. Radiology 2023;309(1):e230806. [doi: 10.1148/radiol.230806] [Medline: 37787671]

XSL•FO
RenderX

24. Zhang S, Xu Y, Usuyama N, Hanwen Xu H, Jaspreet Bagga J, Robert Tinn R, et al. Large-scale domain-specific pretraining for biomedical vision-language processing. ArXiv. 2023;abs/2303.0 2023. [doi: 10.48550/arXiv.2303.00915]
25. Tu T, Azizi S, Driess D, Schaekermann M, Amin M, Chang P, et al. Towards generalist biomedical AI. NEJM AI 2024. [doi: 10.1056/aioa2300138]
26. Cao Y, Xu X, Sun C, Huang X, Shen W. Towards generic anomaly detection and understanding: large-scale visual-linguistic model (GPT-4V) takes the lead. ArXiv. 2023;abs/2311.0 2023. [doi: 10.48550/arXiv.2311.02782]
27. Yang Z, Li L, Lin K, Jianfeng Wang J, Lin CC, Liu Z, et al. The dawn of LMMs: preliminary explorations with GPT-4V(ision). ArXiv. 2023;abs/2309.1 2023. [doi: 10.48550/arXiv.2309.17421]
28. Jin Q, Chen F, Zhou Y, Xu Z, Cheung JM, Chen R, et al. Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine. NPJ Digit Med 2024;7(1):190 [FREE Full text] [doi: 10.1038/s41746-024-01185-7] [Medline: 39043988]
29. USMLE. Sample Test Questions Step 1. 2023. URL: https://www.usmle.org/sites/default/files/2021-10/Step_1_Sample_Items.pdf [accessed 2024-12-17]
30. AMBOSS. AMBOSS Question difficulty 10/15/12023. URL: https://support.amboss.com/hc/en-us/articles/360035679652-Question-difficulty [accessed 2021-06-09]
31. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv:2005.14165 2005. [doi: 10.5860/choice.189890]
32. American Board of Radiology. Qualifying (Core) Exam. URL: https://www.theabr.org/diagnostic-radiology/initial-certification/core-exam [accessed 2024-05-20]
33. Pallais JC, Fenves AZ, Lu MT, Glomski K. Case 18-2018: A 45-year-old woman with hypertension, fatigue, and altered mental status. N Engl J Med 2018;378(24):2322-2333. [doi: 10.1056/nejmcpc1802825]
34. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388(13):1233-1239. [doi: 10.1056/nejmsr2214184]
35. OPENAI. GPT-4V(ision) System Card. URL: https://api.semanticscholar.org/CorpusID:263218031 [accessed 2023-09-25]
36. Yu M, Chang S, Zhang Y, Jaakkola T. Rethinking cooperative rationalization: introspective extraction and complement control. arXiv:1910.13294 2019. [doi: 10.18653/v1/d19-1420]
37. Zaidan O, Eisner J, Piatko C. Using ?Annotator Rationales? to improve machine learning for text categorization. 2007 Presented at: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference; 2024 December 16; New York URL: https://aclanthology.org/N07-1033 [doi: 10.3115/1614164.1614179]
38. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. ACM Comput. Surv 2023;55(12):1-38. [doi: 10.1145/3571730]
39. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-Specific language model pretraining for biomedical natural language processing. ACM Trans. Comput. Healthcare 2021;3(1):1-23. [doi: 10.1145/3458754]
40. Wang J, Yang Z, Yao Z, Yu H. JMLR: Joint Medical LLM and Retrieval Training for Enhancing Reasoning and Professional Question Answering Capability. ArXiv. 2024;abs/2402.1 2024. [doi: 10.48550/arXiv.2402.17887]
41. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
42. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. Radiology 2023;307(5):e230582. [doi: 10.1148/radiol.230582] [Medline: 37191485]
43. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination (USMLE)? the implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
44. Brin D, Sorin V, Vaid A, Soroush A, Glicksberg BS, Charney AW, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023;13(1):16492 [FREE Full text] [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
45. Nakao T, Miki S, Nakamura Y, Kikuchi T, Nomura Y, Hanaoka S, et al. Capability of GPT-4V(ision) in the Japanese national medical licensing examination: evaluation study. JMIR Med Educ 2024;10:e54393. [doi: 10.2196/54393] [Medline: 38470459]
46. Takagi S, Koda M, Watari T. The performance of ChatGPT-4V in interpreting images and tables in the Japanese medical licensing exam. JMIR Med Educ 2024;10:e54283 [FREE Full text] [doi: 10.2196/54283] [Medline: 38787024]
47. Zhou Y, Ong H, Kennedy P, Wu CC, Kazam J, Hentel K, et al. Evaluating GPT-V4 (GPT-4 with Vision) on detection of radiologic findings on chest radiographs. Radiology 2024;311(2):e233270. [doi: 10.1148/radiol.233270] [Medline: 38713028]
48. Brin D, Sorin V, Barash Y, Konen E, Glicksberg BS, Nadkarni GN, et al. Assessing GPT-4 multimodal performance in radiological image analysis. Eur Radiol 2024. [doi: 10.1007/s00330-024-11035-5] [Medline: 39214893]
49. The Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners® (NBME®). Step 3 - United States Medical Licensing Examination. URL: https://www.usmle.org/step-exams/step-3 [accessed 2024-12-17]
50. Elkassem AA, Smith AD. Potential use cases for ChatGPT in radiology reporting. AJR Am J Roentgenol 2023;221(3):373-376. [doi: 10.2214/AJR.23.29198] [Medline: 37095665]

XSL•FO
RenderX

51. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. Int J Environ Res Public Health 2023;20(4):3378 [FREE Full text] [doi: 10.3390/ijerph20043378] [Medline: 36834073]
52. Shea YF, Lee CMY, Ip WCT, Luk DWA, Wong SSW. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. JAMA Netw Open 2023;6(8):e2325000 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.25000] [Medline: 37578798]
53. Lalor JP, Wu H, Yu H. Learning latent parameters without human response patterns: item response theory with artificial crowds. arXiv:1908.11421 2019. [doi: 10.18653/v1/d19-1434]
54. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017;12(1):113. [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
55. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol 2020;30(10):5525-5532 [FREE Full text] [doi: 10.1007/s00330-020-06946-y] [Medline: 32458173]
56. Van Cauwenberge D, Van Biesen W, Decruyenaere J, Leune T, Sterckx S. "Many roads lead to Rome and the artificial Intelligence only shows me one road": an interview study on physician attitudes regarding the implementation of computerised clinical decision support systems. BMC Med Ethics 2022;23(1):50 [FREE Full text] [doi: 10.1186/s12910-022-00787-8] [Medline: 35524301]
57. Jones C, Thornton J, Wyatt JC. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. Med Law Rev 2023;31(4):501-520 [FREE Full text] [doi: 10.1093/medlaw/fwad013] [Medline: 37218368]
58. Patel UB, Ward TJ, Kadoch MA, Cham MD. Radiographic features of pulmonary embolism: hampton's hump. Postgrad Med J 2014;90(1065):420-421. [doi: 10.1136/postgradmedj-2013-132097] [Medline: 24894313]
59. Chen Z, Cano AH, Romanou A, Bonnet A, Matoba K, Salvi F, et al. MEDITRON-70B: scaling medical pretraining for large language models. arXiv:2311.16079 2023. [doi: 10.48550/arXiv.2311.16079]
60. Kim SH, Schramm S, Berberich C, Rosenkranz E, Schmitzer L, Serguen K, et al. Human-AI collaboration in large language model-assisted brain MRI differential diagnosis: a usability study. 2024. URL: https://www.medrxiv.org/content/10.1101/2024.02.05.24302099v1 [accessed 2024-02-06]
61. Jiang S, Xu YY, Lu X. ChatGPT in radiology: evaluating proficiencies, addressing shortcomings, and proposing integrative approaches for the future. Radiology 2023;308(1):e231335. [doi: 10.1148/radiol.231335] [Medline: 37432082]
62. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
63. Gandhi TK, Kachalia A, Thomas EJ, Puopolo AL, Yoon C, Brennan TA, et al. Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. Ann Intern Med 2006;145(7):488-496. [doi: 10.7326/0003-4819-145-7-200610030-00006] [Medline: 17015866]
64. Berlin L. Radiologic errors, past, present and future. Diagnosis (Berl) 2014;1(1):79-84 [FREE Full text] [doi: 10.1515/dx-2013-0012] [Medline: 29539959]

Abbreviations

AI: artificial intelligence
CDSS: clinical diagnostic support system
CK: clinical knowledge
DRQCE: Diagnostic Radiology Qualifying Core Exam
LLMs: large language models
USMLE: United States Medical Licensing Examination

Edited by G Eysenbach, A Coristine; submitted 20.08.24; peer-reviewed by SH Kim, B Thies, TF Heston; comments to author 16.10.24; revised version received 06.11.24; accepted 26.11.24; published 07.02.25.

Please cite as:
Yang Z, Yao Z, Tasmin M, Vashisht P, Jang WS, Ouyang F, Wang B, McManus D, Berlowitz D, Yu H
Unveiling GPT-4V's hidden challenges behind high accuracy on USMLE questions: Observational Study
J Med Internet Res 2025;27:e65146
URL: https://www.jmir.org/2025/1/e65146
doi:10.2196/65146
PMID:

XSL•FO
RenderX

©Zhichao Yang, Zonghai Yao, Mahbuba Tasmin, Parth Vashisht, Won Seok Jang, Feiyun Ouyang, Beining Wang, David McManus, Dan Berlowitz, Hong Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Retrieval Augmented Therapy Suggestion for Molecular Tumor Boards: Algorithmic Development and Validation Study

Eliza Berman[1], BA; Holly Sundberg Malek[2], PhD; Michael Bitzer[3], MD; Nisar Malek[3], MD; Carsten Eickhoff[1], PhD

[1]Center for Digital Health, University Hospital Tuebingen, Tuebingen, Germany

[2]Center for Personalized Medicine, University Hospital Tuebingen, Tuebingen, Germany

[3]Department of Internal Medicine I, University Hospital Tuebingen, Tuebingen, Germany

Corresponding Author:
Eliza Berman, BA
Center for Digital Health
University Hospital Tuebingen
Schaffhausenstrasse 77
Tuebingen, 72072
Germany
Phone: 49 70712984350
Email: eliza_berman@alumni.brown.edu

Abstract

Background: Molecular tumor boards (MTBs) require intensive manual investigation to generate optimal treatment recommendations for patients. Large language models (LLMs) can catalyze MTB recommendations, decrease human error, improve accessibility to care, and enhance the efficiency of precision oncology.

Objective: In this study, we aimed to investigate the efficacy of LLM-generated treatments for MTB patients. We specifically investigate the LLMs' ability to generate evidence-based treatment recommendations using PubMed references.

Methods: We built a retrieval augmented generation pipeline using PubMed data. We prompted the resulting LLM to generate treatment recommendations with PubMed references using a test set of patients from an MTB conference at a large comprehensive cancer center at a tertiary care institution. Members of the MTB manually assessed the relevancy and correctness of the generated responses.

Results: A total of 75% of the referenced articles were properly cited from PubMed, while 17% of the referenced articles were hallucinations, and the remaining were not properly cited from PubMed. Clinician-generated LLM queries achieved higher accuracy through clinician evaluation than automated queries, with clinicians labeling 25% of LLM responses as equal to their recommendations and 37.5% as alternative plausible treatments.

Conclusions: This study demonstrates how retrieval augmented generation–enhanced LLMs can be a powerful tool in accelerating MTB conferences, as LLMs are sometimes capable of achieving clinician-equal treatment recommendations. However, further investigation is required to achieve stable results with zero hallucinations. LLMs signify a scalable solution to the time-intensive process of MTB investigations. However, LLM performance demonstrates that they must be used with heavy clinician supervision, and cannot yet fully automate the MTB pipeline.

(J Med Internet Res 2025;27:e64364) doi:10.2196/64364

KEYWORDS

large language models; retrieval augmented generation; LLaMA; precision oncology; molecular tumor board; molecular tumor; LLMs; augmented therapy; MTB; oncology; tumor; clinical trials; patient care; treatment; evidence-based; accessibility to care

Introduction

Background

Precision oncology is a rapidly evolving field in which patient treatment is determined by the DNA signature of the tumor. Advancements in genome-sequencing techniques, molecular analysis, and a growing body of clinical trials, have all allowed for the field to integrate more easily into oncology research [1]. Using these techniques to optimize treatment has been shown to be associated with favored clinical outcomes [1]. As a mechanism to facilitate precision oncology, the molecular tumor board (MTB) serves as an approach in which a panel of expert

XSL•FO
RenderX

health care providers use their collective knowledge to interpret the genomic profile of an oncology patient and jointly develop an optimal treatment plan given available therapies. Access to MTBs is likely to correlate to improved patient care [1]. The workflow of an MTB is usually as follows: traditionally, before the conference meeting, members of the panel review a list of patients and their corresponding history and characteristics. During the meeting, members propose targeted therapies and generate a consensus recommendation, which is then received by the treating physician [1]. We propose using large language models (LLMs) to enhance this workflow. The current and proposed LLM-enhanced MTB workflows are shown in Figure 1.

Figure 1. On the top is the current MTB workflow. On the bottom is the LLM-enhanced MTB workflow, in which patient characteristics from the preparatory meeting are used to automatically generate queries. These queries are refined by clinicians and then fed into the LLM for recommendation generation. This recommendation is then reviewed and refined by the MTB panel. LLM: large language model; MTB: molecular tumor board.

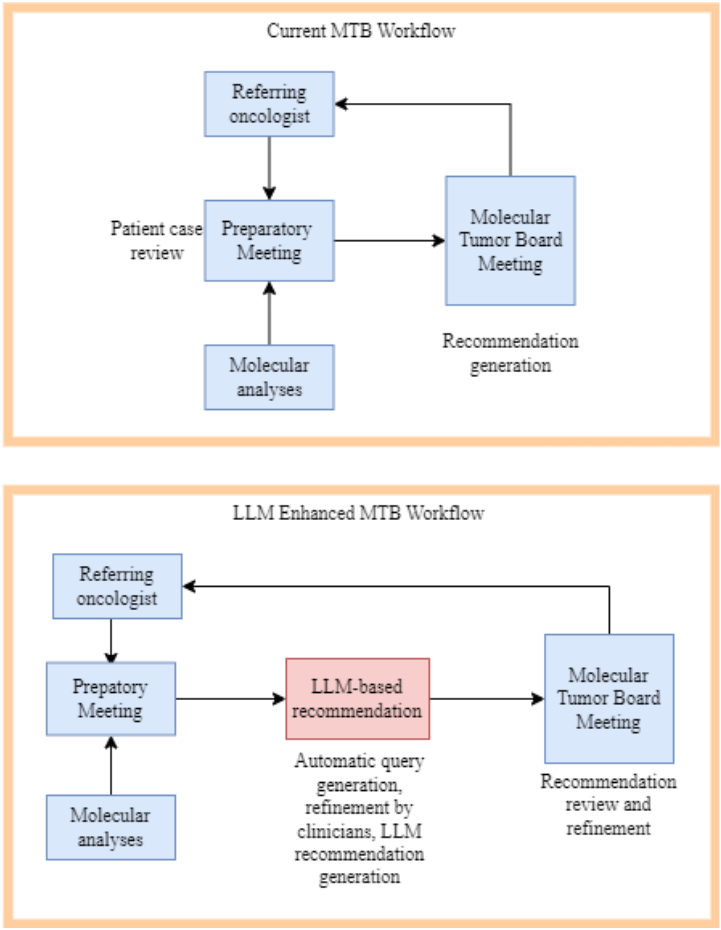

One standing challenge to MTBs is the amount of required resources. Currently, the interpretation of a patient's molecular profile is a manual process, consisting of searching through various databases, such as OncoKB by Memorial Sloan Kettering Cancer Cancer [2], COSMIC by Wellcome Sanger Institute [3], and PubMed [4]. The process of generating the optimal treatment plan is laborious, and scales poorly, potentially limiting the number of patients that can be reviewed in a given conference. The median MTB experience time, which is the time needed to conduct molecular analysis and hold discussion, was reported in a global analysis published in 2020 to be 38.4 days, with a large range from 12.4 to 86 days [5].

LLMs are models trained on large amounts of textual data that are capable of generating language similar to that of humans. LLMs' capabilities span a diverse array of tasks, including question-answering, summarization, translation, and conversing. The development and integration of LLMs is advancing rapidly across different sectors. In particular, LLMs demonstrate impressive performance in automated analyses and syntheses of data [6]. As the written text is used heavily in the medical domain for communication between health care providers, between health care providers and patients, and reports on diagnoses, procedures, and patient results, it is an opportune moment for LLM application [7-10]. LLMs have a high potential to improve patient care and the scalability of patient care by increasing medical reasoning, medical knowledge competency, and access to scientific knowledge [11].

The use of MTBs, as stated above, is aided by the advancements in several adjacent fields within oncology, yet its connection to artificial intelligence remains understudied. LLMs have the potential to drastically increase the scalability of MTBs, ease access to information, and prevent human error [12]. LLMs, when used in the proper context with proper understanding, can be a catalyst for precision oncology integration.

Previous Works

Previous works have demonstrated the feasibility of using natural language processing (NLP) in the field of precision

XSL•FO
RenderX

medicine for a plethora of applications, from treatment plans to predictive diagnoses and prognoses. Within precision oncology, researchers have already conducted a study evaluating LLMs for the role they can play in clinical decision-making in the MTB space, although this evaluation does not include experimentation on using different information retrieval (IR), or retrieval augmented generation (RAG) techniques in conjunction with a pretrained LLM [13]. As a result, this study found that LLMs are not suitable to be used in automating the MTB process [14].

Outside the context of MTBs, there has been a substantial amount of research on creating LLMs for the clinical domain. This can be done through different approaches such as generative data augmentation [15], pretraining on medical domain literature [16,17], and fine-tuning on clinical domain data [17-20]. Additionally, these investigations explore different LLMs such as LLaMA, which is open-source, as opposed to GPT, which is proprietary. The result of this difference is that LLaMA can be run locally and therefore offers inherent data security advantages.

Additionally, there are several related works in the field of IR. Such studies investigate how to use NLP and named entity recognition (NER) to analyze medical databases. Through NER, literature can be obtained for a given keyword search, such as a gene, by filtering articles by term presence [21]. One study published the RetriLite framework, in which IR combines with NLP and LLMs in powerful ways to create a functional pipeline in which NLP is used for context analysis and NER, and IR is used for query expansion and relevancy-based ranking [22,23]. RetriLite is an IR framework that capitalizes on NLP to identify relevant articles from domain-specific knowledge bases and extract these sources [22]. This approach focuses on retrieving the highest quality of results, but does not explore the presentation of that output via LLM-produced textual responses.

To summarize, previous works have focused on IR in the medical domain and fine-tuning LLMs for medical domain-specific tasks. Some models that have focused specifically on the interaction and integration of IR and NLP retrieve high-relevancy results. However, the use of LLMs within precision oncology to produce meaningful, evidence-based recommendations remains undetermined. Applying LLMs to MTBs has clear promise in terms of saving time and money, improving accessibility, and increasing the feasibility and scale of MTBs.

Objectives

This article (1) presents a framework for using LLMs with medical domain knowledge to generate treatment plans for sample cases from MTB conferences, (2) explores the effects of different prompt and query techniques on the LLM, and (3) conducts an evaluation against ground truth historical cases from an MTB annotated by members of the MTB panel.

Methods

Ethical Considerations

This study was exempted by the University of Tübingen's ethics review board due to the exclusive use of nonidentifiable data (Article 2(1) GDPR).

Constructing the LLM Framework

To construct the framework that would be used to obtain treatment recommendations for sample MTB cases, we first selected a baseline model to use. This decision was driven by experimenting with various open-source LLMs and selecting the LLM that had the best performance on treatment recommendations without any downstream fine-tuning. Prompt experimentation on different LLMs is shown in Table S1 in Multimedia Appendix 1. The selected model was Llama-2-7b-chat [24]. In addition to its baseline performance prior to fine-tuning, as mentioned in Previous Works, LLaMA also offers inherent security advantages over other non-open-source models such as GPT-4.0. We also experimented with various hyperparameters for this model. The optimal hyperparameters are shown in Table S2 in Multimedia Appendix 1. A full workflow is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Given the input documents to an MTB conference, we parse the data into actionable queries for each pair of cancer type and genetic mutation. For each patient, we identify the diagnosis and the relevant genetic mutations by filtering the genes that are oncoKB level 1, 2, or 3, as scraped and listed in Table S4, Multimedia Appendix 1. We construct a query for document retrieval and a query for LLM inference. The LLM uses RAG, incorporating documents retrieved by querying the PubMed Data Loader using LlamaIndex, as detailed in the Methods. For each LLM response, we isolate the treatment recommendation, justification, and references and evaluate them against the MTB protocol, which we consider the ground truth. LLM: large language model; MTB: molecular tumor board; RAG: retrieval augmented generation.

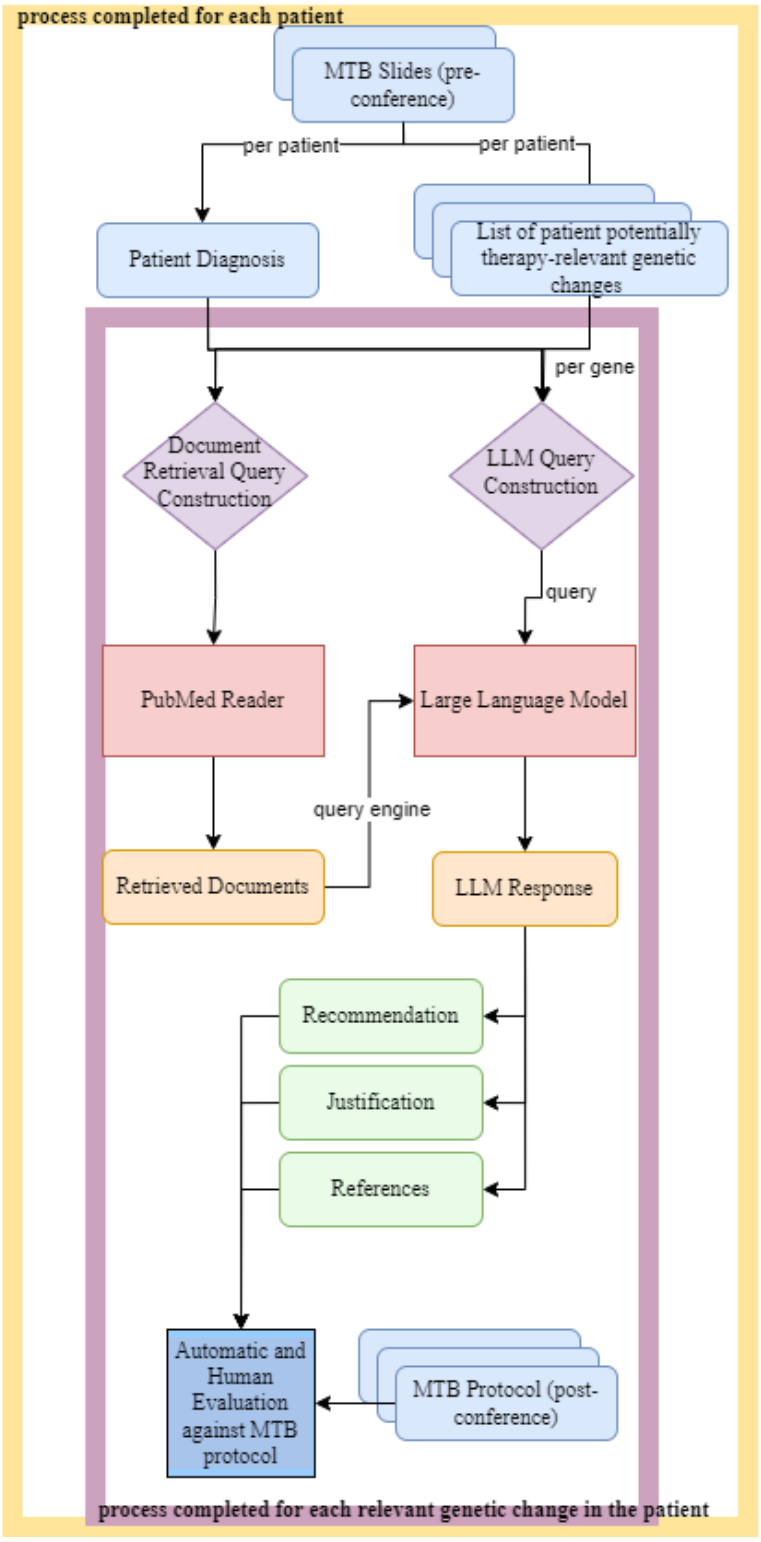

To interface with the LLM as a search engine and to integrate PubMed data, we used LlamaIndex [25]. LlamaIndex is an orchestration framework for incorporating custom data into an LLM. It injects data from custom sources into the LLM, embeds the data as vectors, and allows the data to be queried by storing it in a vectorized database. Using LlamaIndex allows us to use RAG to harness the capabilities of an LLM for domain-specific tasks. We then feed the PubMed papers from LlamaIndex into the LLM to be used for RAG [25]. This is highlighted in red and orange in Figure 2. A full workflow of the LLamaIndex

XSL•FO
RenderX

orchestration, detailing data selection and literature retrieval is shown in Figure S1 in Multimedia Appendix 1.

To determine the ideal query for the PubMed-adjusted LLM, we experimented with different query types both for document retrieval and for text generation. We determined the optimal document retrieval query from the PubMed data loader would be the name of the broader cancer type, and the genetic mutation to be investigated. This is shown in Table 1, and further discussed in the discussion. Experimentation of various queries with the same LLM pipeline is shown in Table S3 in Multimedia Appendix 1. The genetic mutations to be further queried were determined as follows: for a given patient's list of genetic mutations that are included in the briefing document sent out to MTB panelists before the meeting, any genetic mutation from that list that is flagged on OncoKB as "level of evidence one, two, or three" (under their original name or alternate name) is further queried. These 3 levels distinguish varying degrees of available clinical evidence. Level 1 drugs are Food and Drug Administration (FDA) recognized biomarkers that are predictive of response to an FDA-approved drug in a specific indication. Level 2 drugs are standard care biomarkers recommended by the National Comprehensive Cancer Network (NCCN) or other professional guidelines predictive of response to an FDA-approved drug. Level 3 drugs have compelling clinical evidence supporting the biomarker as being predictive of response to a drug. Other drugs exist on OncoKB with evidence categorized as biological rather than clinical, with hypothetical therapeutic implications based on preliminary data, or without a determined level of evidence [2]. The resulting full list of genes is shown in Table S4 in Multimedia Appendix 1. For example, "Occult metastatic melanoma NRAS" would be a query for document retrieval. NRAS is listed on OncoKB as a level 1 gene. We define "broader cancer type" as the cancer diagnosis excluding the stage of the cancer, the location of the tumor, or other additional information that may have been included in the diagnosis. This is because when the query exceeds a certain length, the results are too noisy for meaningful articles to be retrieved. Then, a Vector Store Index is constructed from the retrieved documents, which then serves as a query engine. The Vector Store Index stores each document node and corresponding embedding in a Vector Store, allowing a query to fetch the most similar nodes. The optimal query for the LLM uses the full name of the cancer, provided by the MTB. The query is shown in Table 1.

Table 1. Input illustration for document retrieval and large language model (LLM) querying. For the Cancer Type Short, we input the type of cancer without further specifications such as stage or tumor location. For Cancer Type Long, we input the exact diagnosis listed in the molecular tumor board (MTB). The gene mutation is inputted as it is listed on OncoKB.

Task	Query
Document retrieval	Cancer Type Short + " " + Gene Mutation
LLM response	*A patient has* + Cancer Type Long +, *and the* + Gene Mutation +*gene mutation. What is thebest therapy treatment for this patient? Cite the PubMed sources you used to inform your response.*

The results from these queries are in the form of text paragraphs. Each paragraph states the recommendation, provides justification, and cites sources. Including justification in the recommendation allows for chain-of-thought reasoning, which has been shown to improve LLM performance [26]. The number of sources to cite is unspecified in the response. Asking the LLM to cite sources has many shown benefits including allowing the reader to fact-check the LLM response and improving faithfulness and decreasing hallucinations by forcing LLMs to justify their recommendations [27].

Evaluation Set Construction From Historical MTB Cases

To evaluate LLM recommendations, we created a dataset by retrieving previous cases presented at MTB conferences at the Center for Personalized Medicine at the Universitätsklinikum Tübingen [4]. This was done by manually evaluating case reports given to the panelists before the conference to extract lists of genetic mutations and the cancer diagnosis for each patient to create the queries. As this is not an end-to-end automated product, but a feasibility study of LLM capacities, we decided to manually extract genetic mutations and diagnoses to ensure the accuracy of queries. We further acknowledge that the integration of the case reports into the LLM directly is an important undertaking for future implementations. Then, results were collected from the LLM and compared against the ground-truth responses produced at the conference. The evaluation was conducted by a member of the MTB.

Results

For each cancer type + genetic mutation pair, we generated 2 different treatment recommendations using the LLM. As shown in Table 1, we do not instruct the LLM to cite a specified number of sources. As a result, the number of sources referenced in one LLM response varied drastically across queries. The number of articles retrieved ranged from 1 to 12 across the 40 total LLM responses (2 unique retrievals per query).

A full histogram of the number of retrieved articles is shown in Figure S2 in Multimedia Appendix 1. Each citation in an LLM response was categorized into one of four types.

1. Hallucination: No article with that title or citation is findable on the internet.
2. Non-PubMed: The article is findable on the internet, but not on PubMed.
3. Inferred-PubMed: The title differed slightly from a title that does exist on PubMed, and the author and journal were identical to what was indicated in the response. Examples of citations labeled as Inferred-PubMed are shown in Table 2.
4. PubMed: The exact article is findable on PubMed.

XSL•FO
RenderX

Table 2. Examples of citations labeled as Inferred-PubMed.

LLM Citation	Ground truth citation
Mauduit D, et al (2021). Long and short enhancers in melanoma cell states are enriched in cell-type–specific enhancer function. Elife, 10, e71735.	Mauduit D, Taskiran II, Minnoye L, de Waegeneer M, Christiaens V, Hulselmans G, Demeulemeester J, Wouters J, Aerts S. Analysis of long and short enhancers in melanoma cell states. Elife. 2021 Dec 7;10:e71735. doi: 10.7554/eLife.71735. PMID: 34874265; PMCID: PMC8691835
The study by Maesaki et al (1999) published in the journal Mol Cell found that the structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PRK1, which provides insight into the mechanism of resistance to anti-HER2 therapies.	Maesaki R, Ihara K, Shimizu T, Kuroda S, Kaibuchi K, Hakoshima T. The structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PRK1. Mol Cell. 1999 Nov;4(5):793-803. doi:10.1016/s1097-2765(00)80389-5. PMID: 10619026.

The complete categorization of articles is shown in Figure S3 in Multimedia Appendix 1. The overall categorization of articles in shown in Table 3. The categorization of articles was conducted through manual human evaluation of retrieved articles. The majority, 76% of retrieved articles, were findable on PubMed, while 17% of articles were hallucinations. Accounting for the final quantity of retrieved articles, 3% of articles were non-PubMed and 3% were PubMed-Inferred. This demonstrates the majority of citations were credible. The cited articles from beyond PubMed demonstrate that the LLM may be using prelearned knowledge, instead of entirely retrieval augmented knowledge, to generate responses, despite prompt instructions.

Table 3. The breakdown of all large language model (LLM) retrieved references into types: hallucination, non-PubMed, Inferred-PubMed, and PubMed.

Reference classification	Retrieved LLM references, %
PubMed	76.0
Hallucination	17.1
Inferred-PubMed	3.4
Non-PubMed	3.4

For each LLM response, we evaluated whether at least one of the referenced articles contained the cancer type mentioned in the query. If this was the case, we assigned a binary point to this LLM response. We performed the same search with regard to gene mutation, assigning a binary point if the gene mutation was present in at least one of the referenced articles. We also evaluated whether at least one of the referenced articles had presented the ground truth protocol that the reference MTB had recommended. Similarly, this was awarded a binary point. Last, we evaluated whether the LLM response itself mentioned the Keyword treatment recommendation (eg, “PARP inhibitor” or “olaparib”), also worth one binary point. Therefore, every LLM response’s accuracy was ranked on a scale of 0 to 4, where 0 meets none of the criteria above, and 4 meets all of the criteria above. As previously stated, 2 unique LLM responses were recorded per query. This was motivated by the stochastic nature of LLMs, to demonstrate how consecutive identical queries can yield different outcomes. For each response, we evaluated the above 4 characteristics of the response. The results of this accuracy calculation are shown in Table 4. The accuracy calculation on a 4-point scale for each individual LLM response is shown in Figure S3 in Multimedia Appendix 1. The increasing accuracy (from keyword being present to cancer type being mentioned) is expected in Table 4, as the evaluation metric increasingly broadens. Therefore, the lowest accuracy (25%) corresponds to the treatment name being present in the response, whereas the highest accuracy (82.5%) corresponds to the patient’s cancer type being present in at least one referenced article of the response.

Table 4. The overall accuracy of large language model (LLM) responses compared with ground-truth molecular tumor board recommendations.

Patient characteristic	LLM responses that include patient characteristic, %
Keyword in response	25
Protocol in at least one referenced article	45
Gene mutation in at least one referenced article	50
Cancer type in at least one referenced article	82.5

The LLM responses and retrieved articles were presented to an MTB. MTB members were asked to annotate the responses and retrieved articles according to the following criteria. For each article (excluding hallucinations), we asked the MTB members to label it as Not Relevant or Relevant. Not relevant articles should have nothing to do with the case being presented. For each LLM response, we asked MTB members to label the text response as either Off-base, Alternative, or Reference. We define the three LLM response labels as follows:

1. Off-base: This treatment suggestion does not make sense as a possible treatment for the patient.

XSL•FO
RenderX

2. Reference: This treatment is what was suggested at the MTB conference.
3. Alternative: This is not the treatment suggested in the MTB conference, but it is a potential alternative treatment route for the patient.

After collecting feedback from one clinician, we asked them to provide the query with which they themselves would have prompted the LLM. We obtained a second round of LLM responses and collected corresponding feedback on these responses. The results from this evaluation are shown in Table 5. The LLM prompted with automatically generated queries retrieved treatment recommendations equal to that of clinicians in 32% of cases, and alternative plausible treatment recommendations in 16% of cases. When the LLM was prompted (for the same cases) with queries crafted by clinicians, the recommendations were equal to that of clinicians in 25% of cases, and alternative plausible treatment recommendations in 38% of cases. Clinician-generated queries yielded more accurate LLM responses and more relevant publication citations.

Table 5. Annotation from molecular tumor board (MTB) members of large language model (LLM) responses and retrieved articles. Automated queries were constructed by combining a gene name and cancer diagnosis. Clinician queries were created by an MTB member for the same case.

Classification	Percentage from automated queries, %	Percentage from clinician queries, %
Retrieved article feedback		
Not relevant	87.02	67.78
Relevant	12.98	32.22
LLM response feedback		
Off-base	52.63	37.50
Alternative	15.79	37.50
Reference	31.58	25.00

Additionally, feedback from clinicians regarding some LLM responses indicates that there is more nuance to MTB treatment recommendations than solely what was included in the LLM query. For example, in the case of a patient with adenocarcinoma of the esophagogastric junction and the ERBB2 amplification, clinician feedback noted that the MTB recommendation (Trastuzumab-Deruxtecan) was based on the pathology HER2 DAKO 3+ Score, which was due to the ERBB2 Amplification. Furthermore, the recommendation was made because the patient had already previously received Trastuzumab. This information was only derived from interaction with the clinical team and was not available to the LLM. Consequently, our experiments should be considered a lower bound of how well RAG-enhanced LLMs could work in the MTB space when fully integrated into care systems such as electronic health records. We opted for these evaluation metrics as they offer a more informative approach than fully automated metrics such as ROUGE scores, which track topicality at a high level of abstraction. The evaluation methods we have chosen use explicit human judgments and are possible given the small size of our dataset.

Discussion

Summary

We developed a RAG-enhanced LLM to streamline the treatment recommendation process for MTB conferences. This provided a pipeline to generate treatment recommendations for MTB patients quickly while maintaining a rigorous evidence retrieval stage from medical literature. When the LLM is prompted with automatically generated queries to retrieve treatment recommendations, these recommendations were equal to that of clinicians or alternative plausible treatment recommendations in 48% of cases. However, when the LLM was prompted with clinician-generated queries, recommendations were equal to that of clinicians or alternative plausible treatment recommendations in 63% of cases. Importantly, this demonstrates that high-quality treatment recommendations can in some cases be achieved by the LLM without the manual and time-consuming tasks of conducting human literature reviews and synthesizing retrieved documents. This is even more promising when clinicians play a role in query generation [12]. However, this is not always the case, and some retrieved articles used as evidence are highly irrelevant, demonstrating a disconnect between recommendations and citations. Additionally, hallucinations remain a persistent problem in LLM responses, accounting for 17.1% of cited sources. LLMs have the potential to make the MTB workflow more efficient, and scalable, which can lead to greater access to personalized patient care. To our knowledge, this is the first attempt to apply RAG to enhance LLMs for MTB use.

Key Findings

This study demonstrates several key findings. First, LLMs can be powerful tools to aid in clinician decision-making on MTBs. However, the use of NLP-assisted MTB recommendations necessitates oversight by MTB members at multiple stages of the pipeline, including query generation as well as recommendation assessment prior to presentation to a patient. Second, the references retrieved for a given LLM response do not align directly with the textual response generated by the LLM. This suggests that rather than all of the responses being produced through RAG approaches, LLM memory also plays a significant role in its decision-making. This role must be further explored and understood so as to mitigate unintended consequences of using pretrained LLMs to generate patient treatment recommendations. Last, it is essential to consider the drawbacks of LLMs (as they currently function) in the clinician domain: namely, the potential dissemination of misinformation through hallucinatory responses that must be properly vetted before being legitimized in patient treatment. Strategies to

XSL•FO
RenderX

mitigate hallucinations are imperative to ensure the trustworthiness of LLM-generated treatment suggestions.

It is important to note that the results from our study are in slight disagreement with that of the study conducted using pretrained LLMs on fictional patient vignettes [14]. Their study noted how the accuracy of provided LLM references was often a reason why the LLM itself was deemed unfit for the cause. We believe our study poses a potential solution to this, through the addition of RAG. Therefore, this new approach can advance the field by demonstrating the potential for a more robust LLM-enhanced MTB pipeline that decreases hallucination rates. Hallucinations still pose a serious problem in LLM implementation in health care settings. In this feasibility study, our goal is to assess the extent of this issue and benchmark existing components, rather than resolve it.

Limitations

Limitations of this study include reproducibility of LLM responses, scope of query inputs, evaluation technique, and generalizability.

Reproducibility

Given the dynamic nature of LLMs, it is not guaranteed to achieve consistent responses, even when the same prompt is applied [11]. Additionally, we observed different LLMs can yield different results, given the same prompt. Results from this are shown in Table S1 in Multimedia Appendix 1. We observed changes in LLM recommendations when we reran inference numerous times with identical prompts on the same model. Due to this, it is crucial to document all model parameters, prompt templates, and model versions when experimenting with LLMs in the medical domain. The goal of this study is to observe and quantify the extent of such behavior in LLMs. We leave the challenging task of addressing these issues for future work.

Scope of Queries

Currently, when we extract information from the patient's genetic characteristics provided prior to the conference, we are identifying and querying genetic mutations that are listed on OncoKB as level 1, 2, or 3, as shown in Table S4 in Multimedia Appendix 1. This excludes not-categorized genes as well as genes listed with levels above 3, which may also be listed on the patient's genetic information. Additionally, a number of treatment recommendations from MTB conferences were motivated by other, non–genetic-mutation characteristics of the patient, such as tumor mutation burden range, homologous recombination deficiency assessment, and human lymphocyte antigen expression levels. These values, while promising biomarkers for responsiveness to certain treatments [28-30] were not available to our models. In future modifications to this algorithm, we hope to extract all patient information pertaining to their genetic mutations and biomarkers so that we can increase the scope and therefore, the robustness, of generated recommendations.

Evaluation Technique

Feedback from clinicians suggested that a further evaluation category should be included to account for LLM responses that list a combination treatment of multiple drugs, but the MTB recommended only one of them, or vice versa. Adding this category would improve the robustness of the evaluation.

Generalizability

This study is focused on a single comprehensive cancer center. In the future, this approach can be expanded and evaluated at additional sites and in different health care settings to gauge the generalizability and scalability of the proposed methods.

Future Directions

There are many future directions motivated by this study.

Using Local Hospital Data for RAG

First, our model is built on PubMed sources to augment and improve the credibility and accuracy of LLM responses, decrease LLM hallucination rate, and isolate the latest research to inform recommendations. However, by enhancing this system by using local hospital data (ie, using existing local hospital patient history and MTB conference recommendations), we can tune the model to be more accurate within the specific context of the hospital's resources and history of care. Existing research demonstrates that in-house models can better accommodate the needs of specific centers, therefore enhancing precision medicine across the field [12]. This would also allow for a framework to generalize this approach across various health care settings. Upon expansion to other health care settings, further experimentation would be needed to investigate the scalability of the proposed approach.

Increasing Response Relevance and Decreasing Hallucinations Through Isolating IR from LLM Generation, and LLM Architecture Experimentation

A second direction for future research is to use IR approaches as an independent and isolated step prior to LLM interfacing. This approach can retrieve the optimal articles and directly feed those articles into the LLM to perform RAG. Doing so would ensure that the referenced articles are relevant before we even interface with the LLM. This would also likely increase LLM recommendation performance, since the LLM would be basing its recommendation on a small sample of preselected relevant articles. Through this approach, we could include additional evidence such as scientific meeting abstracts and posters in the searched literature as they provide the most up-to-date trial information, which was a point of feedback made by clinicians. This is motivated by the observation that LLM memory plays a significant role in decision-making, potentially leading to unintended consequences. The misalignment between the sources and response text in some cases can be mitigated in future studies by more robust filtering and contextualization of search results. In order to further understand the role of memory in LLM decision-making, work is being done in the field of mechanistic interpretability to create human-readable explanations of artificial intelligence–generated representations [31]. Another potential experiment to reduce hallucinations and overall increase the accuracy of responses involves using ensemble learning, rather than a single LLM, to harness the power of multiple models. Not only does this have the potential to improve performance, but it also could improve the reliability of results [32].

XSL•FO
RenderX

Experiment Further With Prompting Strategies

Another interesting area of future research lies in diverse prompting strategies. In our experimentation, by adjusting the query slightly, different results are generated by the LLM. Different directions include asking the LLM to rank different treatments, list multiple treatments, or decide between 2 given treatments. Different prompting strategies have been shown to augment LLM text generation, and optimizing prompting strategies can enhance the accuracy, comprehensibility, and quality of LLM responses [33,34]. We can also make the query broader or more specific, such as asking for a general treatment, versus asking for a specific drug name for a targeted therapy approach. The genetic mutation can also be further specified to include details regarding therapy-relevant copy number variations (amplifications and deletions), fusion genes, and rearrangements. Additionally, our clinician feedback suggested a further experiment of feeding a clinician-generated treatment recommendation into the LLM and directly asking the LLM whether, under consideration of this MTB recommendation, further treatment recommendations are suggested by the LLM. Clinicians also noted that, while MTB decisions ideally are based on reports from the same tumor type (m1 evidence), they also recommend treatments successful in other tumor entities with the same genetic alteration (m2 evidence). Therefore, we could expand the query-process into a 2-step approach, in which after looking for m1 evidence, we query for results regarding any cancer type. As clinician feedback is a crucial part of the proposed pipeline, we would like to further experiment with diverse prompting strategies to optimize LLM text generation that directly uses this feedback. In addition to broader human feedback, using strategies such as chain-of-thought prompting and prompt tuning could further improve the accuracy of generated results [35,36].

Improve the Interpretability and Transparency of RAG-Enhanced LLMs to Facilitate Trust Among Clinicians and Patients

There are several existing methods that can be used as tools in LLM evaluation, such as gradient tracing, Shapley values, probing, or casual mediation. For example, Shapley values demystify LLM decision-making by quantitatively determining how much each word in the input text contributes to the LLM output text. This can be used to detect bias by identifying features that disproportionately affect model outputs [37]. This could be used, along with the above-mentioned tools, to increase transparency of LLM decision-making within the MTB treatment recommendation context. Additionally, to improve the LLM's ability to understand the nuances of the MTB treatment recommendations, and therefore create more context-driven responses, we can use existing methods such as reinforcement learning from human feedback, targeted instruction tuning, and grounding LLMs [3,38,39].

Analyze Change in Efficiency of MTB Pipeline

Last, with this study, we demonstrated the feasibility of incorporating LLMs into the MTB pipeline. However, it remains to be shown that this translates to increased efficiency. In a future study, we hope to analyze the effort needed to manually operate the MTB pipeline, and contrast that with that of the LLM-enhanced pipeline. This study evaluates existing open-source components of LLMs in the MTB context, rather than building production-ready systems. Therefore, this study measures the feasibility of LLMs as they currently exist in the MTB pipeline. Our recommendation based on this study is to continue working toward fixing the observed issues before moving on to prospective evaluation and clinical integration in real-world settings.

Conclusions

In this study, we developed a RAG approach to LLM text generation to develop treatment recommendations for MTB patients. Specifically, we evaluated the articles referenced in the LLM response, as well as the treatment recommendation itself. This study demonstrates that LLMs have the potential to be harnessed for rapid treatment plan generation based on patient DNA signature and diagnosis, generating responses comparable to that of clinicians and highlighting the potential of LLMs in precision oncology. The use of such an approach has a high feasibility at scale, as opposed to current manual MTB investigations. However, LLMs are not currently suitable to fully automate the MTB process and must be used with supervision from clinicians at each stage of their deployment. We have shown the feasibility of the implementation of LLMs in the MTB pipeline, and in future work, we hope to analyze how this feasibility has the potential of further translating into increased efficiency.

Acknowledgments

We gratefully acknowledge support from the Open Access Publication Fund of the University of Tübingen.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 373 KB - jmir_v27i1e64364_app1.docx]

References

XSL•FO
RenderX

1. Tsimberidou AM, Kahle M, Vo HH, Baysal MA, Johnson A, Meric-Bernstam F. Molecular tumour boards - current and future considerations for precision oncology. Nat Rev Clin Oncol 2023;20(12):843-863. [doi: 10.1038/s41571-023-00824-4] [Medline: 37845306]
2. Welcome to OncoKB. OncoKB. URL: https://www.oncokb.org/ [accessed 2025-01-30]
3. COSMIC: catalogue of somatic mutations in cancer. Wellcome Sanger Institute. URL: https://www.sanger.ac.uk/tool/cosmic/ [accessed 2024-07-01]
4. PubMed. National Center for Biotechnology Information. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2024-07-01]
5. Luchini C, Lawlor RT, Milella M, Scarpa A. Molecular tumor boards in clinical practice. Trends Cancer 2020;6(9):738-744 [FREE Full text] [doi: 10.1016/j.trecan.2020.05.008] [Medline: 32517959]
6. Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A comprehensive overview of large language models. arXiv:2307.06435 2023. [doi: 10.48550/arXiv.2307.06435]
7. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. Trends Anaesth Crit Care 2021;38:4-9 [FREE Full text] [doi: 10.1016/j.tacc.2021.02.007]
8. Ahmed U, Iqbal K, Aoun M, Khan G. Natural language processing for clinical decision support systems: a review of recent advances in healthcare. J Intell Connectivity Emerging Technol 2023;8:1-16 [FREE Full text]
9. Aramaki E, Wakamiya S, Yada S, Nakamura Y. Natural language processing: from bedside to everywhere. Yearb Med Inform 2022;31(1):243-253 [FREE Full text] [doi: 10.1055/s-0042-1742510] [Medline: 35654422]
10. Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. Comput Biol Med 2023;155:106649. [doi: 10.1016/j.compbiomed.2023.106649] [Medline: 36805219]
11. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt J, Laleh NG, et al. The future landscape of large language models in medicine. Commun Med (Lond) 2023;3(1):141. [doi: 10.1038/s43856-023-00370-1] [Medline: 37816837]
12. Tamborero D, Dienstmann R, Rachid MH, Boekel J, Lopez-Fernandez A, Jonsson M, et al. The molecular tumor board portal supports clinical decisions and automated reporting for precision oncology. Nat Cancer 2022;3(2):251-261 [FREE Full text] [doi: 10.1038/s43018-022-00332-x] [Medline: 35221333]
13. Patrick Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. arXiv:2005.11401 2005. [doi: 10.48550/arXiv.2005.11401]
14. Benary M, Wang XD, Schmidt M, Soll D, Hilfenhaus G, Nassir M, et al. Leveraging large language models for decision support in personalized oncology. JAMA Netw Open 2023;6(11):e2343689 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.43689] [Medline: 37976064]
15. Guo Z, Wang P, Wang Y, Shangdi Yu S. Improving small language models on pubmedqa via generative data augmentation. arXiv:2305.07804 2023. [doi: 10.48550/arXiv.2305.07804]
16. Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Brief Bioinform 2022;23(6):bbac409. [doi: 10.1093/bib/bbac409] [Medline: 36156661]
17. Wu C, Lin W, Zhang X, Zhang Y, Xie W, Wang Y. PMC-LLaMA: toward building open-source language models for medicine. J Am Med Inform Assoc 2024;31(9):1833-1843. [doi: 10.1093/jamia/ocae045] [Medline: 38613821]
18. Gema AP, Minervini P, Daines L, Hope T, Alex B. Parameter-efficient fine-tuning of llama for the clinical domain. arXiv:2307.03042 2023. [doi: 10.18653/v1/2024.clinicalnlp-1.9]
19. Karkera N, Acharya S, Palaniappan SK. Leveraging pre-trained language models for mining microbiome-disease relationships. BMC Bioinf 2023;24(1):290 [FREE Full text] [doi: 10.1186/s12859-023-05411-z] [Medline: 37468830]
20. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med 2021;4(1):86 [FREE Full text] [doi: 10.1038/s41746-021-00455-y] [Medline: 34017034]
21. Chen HO, Lin PC, Liu CR, Wang CS, Chiang JH. Contextualizing genes by using text-mined co-occurrence features for cancer gene panel discovery. Front Genet 2021;12:771435 [FREE Full text] [doi: 10.3389/fgene.2021.771435] [Medline: 34759963]
22. Zeng J, Cruz-Pico CX, Saridogan T, Shufean MA, Kahle M, Yang D, et al. Natural language processing-assisted literature retrieval and analysis for combination therapy in cancer. JCO Clin Cancer Inform 2022;6:e2100109 [FREE Full text] [doi: 10.1200/CCI.21.00109] [Medline: 34990212]
23. Hamamoto R, Koyama T, Kouno N, Yasuda T, Yui S, Sudo K, et al. Introducing AI to the molecular tumor board: one direction toward the establishment of precision medicine using large-scale cancer clinical and biological information. Exp Hematol Oncol 2022;11(1):82 [FREE Full text] [doi: 10.1186/s40164-022-00333-7] [Medline: 36316731]
24. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: open foundation and fine-tuned chat models. arXiv:2307.09288 2023. [doi: 10.48550/arXiv.2307.09288]
25. Build AI knowledge assistants over your enterprise data. LlamaIndex. URL: https://www.llamaindex.ai/ [accessed 2024-01-26]
26. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain of thought prompting elicits reasoning in large language models. arXiv:2201.11903 2023. [doi: 10.48550/arXiv.2201.11903]
27. Gao T, Yen H, Yu J, Chen D. Enabling large language models to generate text with citations. arXiv:2305.14627 2023. [doi: 10.18653/v1/2023.emnlp-main.398]

XSL•FO
RenderX

28. Ma X, Zhang Y, Wang S, Yu J. Predictive value of tumor mutation burden (TMB) with targeted next-generation sequencing in immunocheckpoint inhibitors for non-small cell lung cancer (NSCLC). J Cancer 2021;12(2):584-594 [FREE Full text] [doi: 10.7150/jca.48105] [Medline: 33391454]
29. Jung J, Heo YJ, Park S. High tumor mutational burden predicts favorable response to anti-PD-(L)1 therapy in patients with solid tumor: a real-world pan-tumor analysis. J Immunother Cancer 2023;11(4):e006454 [FREE Full text] [doi: 10.1136/jitc-2022-006454] [Medline: 37094985]
30. Budczies J, Kluck K, Beck S, Ourailidis I, Allgäuer M, Menzel M, et al. Homologous recombination deficiency is inversely correlated with microsatellite instability and identifies immunologically cold tumors in most cancer types. J Pathol Clin Res 2022;8(4):371-382 [FREE Full text] [doi: 10.1002/cjp2.271] [Medline: 35384413]
31. Bereska L, Gavves E. Mechanistic interpretability for AI safety --a review. arXiv:2404.14082 2024 [FREE Full text]
32. Yang H, Li M, Zhou H, Xiao Y, Fang Q, Zhang R. One LLM is not enough: harnessing the power of ensemble learning for medical question answering. medRxiv 2023 [FREE Full text] [doi: 10.1101/2023.12.21.23300380] [Medline: 38196648]
33. Bsharat SM, Myrzakhan A, Shen Z. Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4. arXiv:2312.16171 2023. [doi: 10.48550/arXiv.2312.16171]
34. Lyu H, Jiang S, Zeng H, Xia Y, Wang Q, Zhang S, et al. Llm-rec: personalized recommendation via prompting large language model. arXiv:2307.15780 2023. [doi: 10.48550/arXiv.2307.15780]
35. Jason Wei J, Xuezhi Wang X, Dale Schuurmans D, Maarten Bosma M, Brian Ichter B, Fei Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. arXiv:2201.11903 1903 [FREE Full text]
36. Lester B, Al-Rfou R, Constant N. Constant, the power of scale for parameter-efficient prompt tuning. arXiv:2104.08691 2021 [FREE Full text] [doi: 10.18653/v1/2021.emnlp-main.243]
37. Brown NB. Enhancing trust in llms: algorithms for comparing and interpreting llms, 2024. arXiv:2406.01943 1943 [FREE Full text]
38. Chaudhari S, Aggarwal P, Murahari V, Rajpurohit T, Kalyan A, Narasimhan K, et al. RLHF deciphered: a critical analysis of reinforcement learning from human feedback for llms. arXiv:2404.08555 2024 [FREE Full text]
39. Xia M, Malladi S, Gururangan S, Arora S, Chen D. Less: Selecting influential data for targeted instruction tuning. arXiv:2402.04333 2024 [FREE Full text]

Abbreviations

FDA: Food and Drug Administration
IR: information retrieval
LLM: large language model
MTB: molecular tumor board
NER: named entity recognition
NLP: natural language processing
RAG: retrieval augmented generation

Edited by B Puladi; submitted 16.07.24; peer-reviewed by L Aljerf, A Ray; comments to author 05.10.24; revised version received 21.11.24; accepted 02.01.25; published 05.03.25.

Please cite as:
Berman E, Sundberg Malek H, Bitzer M, Malek N, Eickhoff C
Retrieval Augmented Therapy Suggestion for Molecular Tumor Boards: Algorithmic Development and Validation Study
J Med Internet Res 2025;27:e64364
URL: https://www.jmir.org/2025/1/e64364
doi:10.2196/64364
PMID:40053768

©Eliza Berman, Holly Sundberg Malek, Michael Bitzer, Nisar Malek, Carsten Eickhoff. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Accuracy of Large Language Models for Literature Screening in Thoracic Surgery: Diagnostic Study

Zhang-Yi Dai[1], MD; Fu-Qiang Wang[1], MD; Cheng Shen[1], MDS; Yan-Li Ji[1], MMSc; Zhi-Yang Li[1], MD; Yun Wang[1], PhD; Qiang Pu[1], MD, PhD

Department of Thoracic Surgery, West China Hospital of Sichuan University, Chengdu, China

Corresponding Author:
Qiang Pu, MD, PhD
Department of Thoracic Surgery
West China Hospital of Sichuan University
No.37, Guoxue Alley
Chengdu, 610041
China
Phone: 86 18980606738
Email: puqiang100@163.com

Abstract

Background: Systematic reviews and meta-analyses rely on labor-intensive literature screening. While machine learning offers potential automation, its accuracy remains suboptimal. This raises the question of whether emerging large language models (LLMs) can provide a more accurate and efficient approach.

Objective: This paper evaluates the sensitivity, specificity, and summary receiver operating characteristic (SROC) curve of LLM-assisted literature screening.

Methods: We conducted a diagnostic study comparing the accuracy of LLM-assisted screening versus manual literature screening across 6 thoracic surgery meta-analyses. Manual screening by 2 investigators served as the reference standard. LLM-assisted screening was performed using ChatGPT-4o (OpenAI) and Claude-3.5 (Anthropic) sonnet, with discrepancies resolved by Gemini-1.5 pro (Google). In addition, 2 open-source, machine learning–based screening tools, ASReview (Utrecht University) and Abstrackr (Center for Evidence Synthesis in Health, Brown University School of Public Health), were also evaluated. We calculated sensitivity, specificity, and 95% CIs for the title and abstract, as well as full-text screening, generating pooled estimates and SROC curves. LLM prompts were revised based on a post hoc error analysis.

Results: LLM-assisted full-text screening demonstrated high pooled sensitivity (0.87, 95% CI 0.77-0.99) and specificity (0.96, 95% CI 0.91-0.98), with the area under the curve (AUC) of 0.96 (95% CI 0.94-0.97). Title and abstract screening achieved a pooled sensitivity of 0.73 (95% CI 0.57-0.85) and specificity of 0.99 (95% CI 0.97-0.99), with an AUC of 0.97 (95% CI 0.96-0.99). Post hoc revisions improved sensitivity to 0.98 (95% CI 0.74-1.00) while maintaining high specificity (0.98, 95% CI 0.94-0.99). In comparison, the pooled sensitivity and specificity of ASReview tool-assisted screening were 0.58 (95% CI 0.53-0.64) and 0.97 (95% CI 0.91-0.99), respectively, with an AUC of 0.66 (95% CI 0.62-0.70). The pooled sensitivity and specificity of Abstrackr tool-assisted screening were 0.48 (95% CI 0.35-0.62) and 0.96 (95% CI 0.88-0.99), respectively, with an AUC of 0.78 (95% CI 0.74-0.82). A post hoc meta-analysis revealed comparable effect sizes between LLM-assisted and conventional screening.

Conclusions: LLMs hold significant potential for streamlining literature screening in systematic reviews, reducing workload without sacrificing quality. Importantly, LLMs outperformed traditional machine learning-based tools (ASReview and Abstrackr) in both sensitivity and AUC values, suggesting that LLMs offer a more accurate and efficient approach to literature screening.

(J Med Internet Res 2025;27:e67488) doi:10.2196/67488

KEYWORDS

accuracy; large language models; meta-analysis; literature screening; thoracic surgery

XSL•FO
RenderX

Introduction

The development of clinical practice guidelines necessitates a comprehensive and systematic synthesis of current research evidence [1]. Evidence-based medicine frequently relies on systematic reviews and meta-analyses, which aggregate findings from studies on a specific topic [2-4]. This process traditionally involves extensive effort in identifying and retrieving relevant literature [5,6]. While machine learning has shown promise in streamlining literature retrieval, its accuracy often falls short of desired standards [7-9]. Therefore, further research is needed to develop a more precise screening method.

Recently, large language models (LLMs) powered by natural language processing have demonstrated remarkable capabilities in various domains, including language comprehension, image and video generation, and data analysis [8-13]. Previous studies have suggested the potential of LLMs for literature screening [9,14,15]. However, their accuracy in the screening process for meta-analyses remains unclear.

We hypothesized that LLM-assisted literature screening could achieve accuracy comparable with manual screening. To test this hypothesis, we designed a diagnostic trial using conventional manual screening as the reference standard to evaluate the accuracy of LLM-assisted literature screening.

Methods

Study Design

This prospective diagnostic study aimed to assess the validity of LLMs for assisting with literature screening during meta-analysis. Conventional literature manual screening served as the reference standard. This diagnostic study was performed according to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.

Before the study commenced, we defined the research topic as the comparison between sublobar resection and lobectomy in thoracic surgery, a topic of ongoing debate within the field. A total of 6 relevant published meta-analyses were identified [16-21], and the search strategy and terms were subsequently redesigned for a new round of literature retrieval and screening. To ensure a comprehensive yet nonredundant literature base, the search strategies and inclusion timeframes outlined within each meta-analysis were replicated. Duplicate studies were subsequently identified and removed using Rayyan [22], a web-based literature management tool.

Conventional Literature Screening

Following identification, 2 independent investigators (Lei Peng and Xing-Yu Liu) screened the titles and abstracts of retrieved studies for inclusion based on predefined criteria (detailed in Table S1 in Multimedia Appendix 1). Discrepancies were resolved through adjudication by a third investigator (Xu-Yang Wang). Full-text papers of potentially eligible studies were subsequently reviewed by the same 2 independent investigators (Lei Peng and Xing-Yu Liu) against the same inclusion criteria, and any discrepancies were again resolved through adjudication by the third investigator (Xu-Yang Wang). This conventional literature screening established the reference standard for comparison. Investigators involved in the conventional screening were excluded from participation in the LLM-assisted screening and subsequent analyses.

LLM-Assisted Literature Screening

For the LLM-assisted literature screening, a 5-column table (author, publication year, journal, title, and abstract) was compiled from the deduplicated literature. Following established prompt engineering guidelines [23], specific prompts were developed to facilitate automated screening using Python (version 3.9.0; Python Software Foundation). These prompts, structured to output results in a tabular format, instructed the LLM to perform screening based on the Population, Intervention, Control, Outcome, and Study design (PICOS) framework criteria defined for each topic study (detailed in Table S1 in Multimedia Appendix 1). An example prompt is provided in Figure S1 in Multimedia Appendix 1.

In addition, 2 LLMs, ChatGPT-4o (OpenAI) and Claude-3.5 sonnet (Anthropic), were used as independent reviewers to independently screen titles and abstracts. Study selection was based on the predefined inclusion and exclusion criteria. Discrepancies in study selection between the 2 LLMs were resolved by Gemini-1.5 pro (Google). Full-text papers underwent an identical screening process. The literature screening process assisted by LLM was conducted and supervised by 2 reviewers (Z-YD and F-QW). The detailed prompts used in LLM-assisted literature screening are provided in section S1 in Multimedia Appendix 2.

We then evaluated the performance of 2 open-source, machine learning–based screening tools, ASReview [24,25] and Abstrackr [26,27], for title and abstract screening. We compared their results against conventional manual screening methods, which served as the reference standard, to assess the accuracy of LLM-assisted literature screening. The detailed methodology is documented in section S2 in Multimedia Appendix 2.

Statistical Analysis

The accuracy of both LLM-assisted literature screening and 2 semiautomated, machine learning–based screening tools were evaluated in each topical study using sensitivity, specificity, and their corresponding 95% CIs. The primary analysis focused on the sensitivity and specificity of LLM-assisted literature screening assessed after full-text review. The secondary analysis focused on the sensitivity and specificity of LLM-assisted screening at the title and abstract review stage. In addition, meta-analysis techniques were used to calculate pooled sensitivity and specificity, along with the summary receiver operating characteristic (SROC) curve. This provided overall results for the primary, secondary, and post hoc analyses. Heterogeneity across topic studies was assessed by calculating the inconsistency value (I^2) using the chi-square test. A random-effects model was used to pool sensitivity and specificity if I^2 exceeded 50% or if the P value was less than .05. Meta-analyses were conducted using Stata (version 15.0;

XSL•FO
RenderX

StataCorp), while other statistical analyses were performed using GraphPad Prism (version 8.0.1; GraphPad Prism, Inc).

A post hoc analysis (conducted by YW and QP) involved reviewing papers of false-negative classifications to identify the sources of LLM errors during screening (the comprehensive explanations for the occurrence of false-negative classifications are provided in Table S2 in Multimedia Appendix 1). Subsequently, the literature screening prompts were refined by incorporating a chain-of-thought prompting strategy [28] based on the identified error patterns (conducted by CS). In addition, 3 iterations of screening querying were then performed with revised prompts to optimize the model's validity. A study was considered eligible if the LLM classified it as eligible during any of the 3 iterations.

To further assess the robustness of LLM-assisted screening and account for potential variations, a separate post hoc meta-analysis was conducted (Y-LJ and Z-YL). This analysis compared the pooled effect sizes derived from LLM-assisted screening (including only true positives) with those from conventional screening (including both true positives and false negatives) for each topic study.

Ethical Considerations

The study did not involve human participants or biological specimens. As such, the Biomedical Ethics Committee of West China Hospital, Sichuan University determined that this research was eligible for exemption from ethical review (reference number: 2024−1177). This decision aligns with institutional and local policies that exempt studies not involving human subjects or biological materials from requiring formal ethics board approval.

Results

The Results of Conventional Manual and LLM-Assisted Literature Screening

Following deduplication in the conventional literature screening process, the initial search yielded 357, 462, 296, 429, 2,298, and 696 papers for studies 1-6, respectively. Title and abstract screening resulted in 98, 41, 44, 14, 126, and 138 papers selected for full-text review in the corresponding studies. Ultimately, 28, 12, 26, 11, 10, and 26 papers from studies 1 to 6, respectively, met the inclusion criteria and were incorporated into the final meta-analysis (Figure 1). The results of LLM-assisted literature screening are listed and described in Table S3 in Multimedia Appendix 1 and section S3 in Multimedia Appendix 2.

Figure 1. Flow diagram of meta-analysis using conventional manual literature screening and large language model (LLM)–assisted screening. AUC: area under the curve; SROC: summary receiver operating characteristic curve.

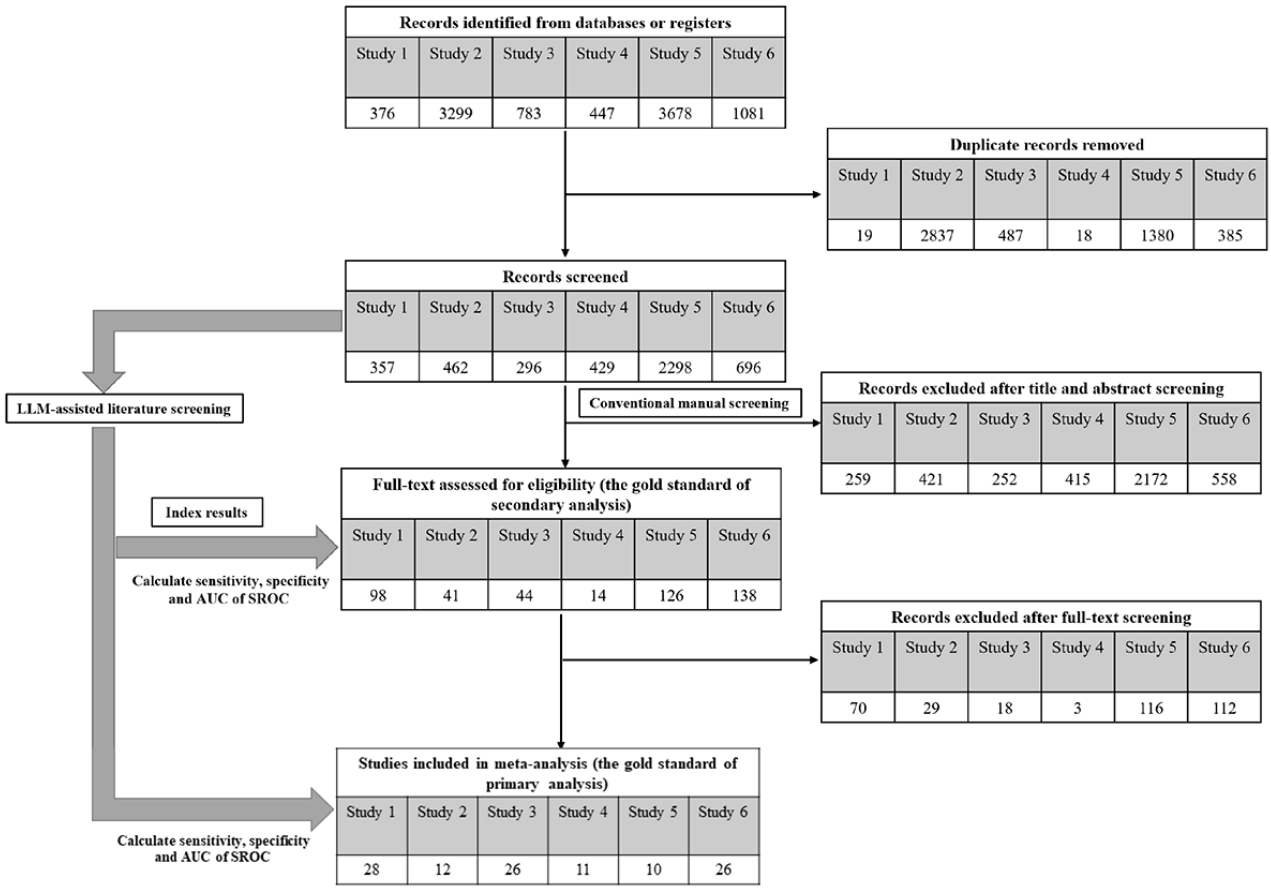

XSL•FO
RenderX

Primary Analysis of LLM-Assisted Literature Screening

In the LLM-assisted literature screening process, a total of 26, 9, 24, 11, 9, and 18 papers from studies 1 to 6, respectively, were included in the final meta-analysis (Table S3 in Multimedia Appendix 1). The primary analysis revealed the sensitivity and specificity of the LLM-assisted screening for studies 1-6 as follows: 0.93 (95% CI 0.76-0.99) and 0.84 (95% CI 0.80-0.88), 0.75 (95% CI 0.43-0.95) and 0.94 (95% CI 0.91-0.96), 0.92 (95% CI 0.75-0.99) and 0.97 (95% CI 0.94-0.99), 1.00 (95% CI 0.72-1.00) and 0.92 (95% CI 0.89-0.94), 0.90 (95% CI 0.55-1.00) and 0.99 (95% CI 0.98-0.99), and 0.69 (95% CI 0.48-0.86) and 0.99 (95% CI 0.98-1.00), respectively (Figure 2). Meta-analysis of 6 topic studies revealed that LLM-assisted screening demonstrated a high discriminative ability, with SROC curve analysis yielding an area under the curve (AUC) of 0.96 (95% CI 0.94-0.97); see Figure S2 in Multimedia Appendix 1. Furthermore, the pooled sensitivity and specificity were 0.87 (95% CI 0.77-0.99) and 0.96 (95% CI 0.91-0.98), respectively (Figure 2). The counts of true positives, false negatives, false positives, and true negatives of primary analysis are detailed in Table S3 in Multimedia Appendix 1.

Figure 2. Sensitivity and specificity of large language model (LLM)–assisted literature screening in the primary analysis.

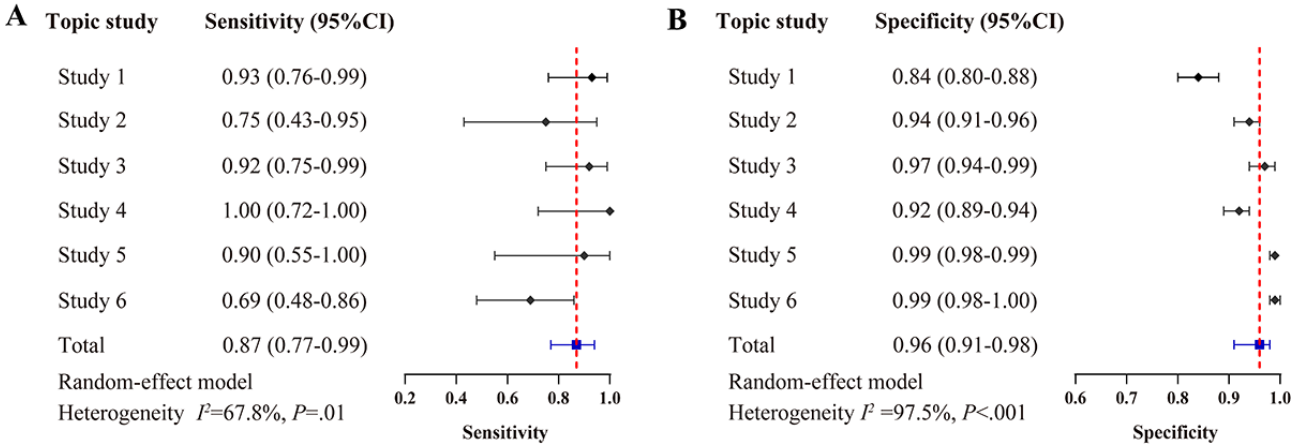

Secondary Analysis of LLM-Assisted Screening

The pooled sensitivity and specificity across these 6 studies were 0.73 (95% CI 0.57-0.85) and 0.99 (95% CI 0.97-0.99), respectively (Figure 3 and section S4 in Multimedia Appendix 2). The SROC analysis yielded an AUC of 0.97 (95% CI 0.96-0.99); see Figure S2 in Multimedia Appendix 1. The counts of true positives, false negatives, false positives, and true negatives of secondary analysis are detailed in Table S3 in Multimedia Appendix 1.

Figure 3. Sensitivity and specificity of large language model (LLM)–assisted literature screening in the secondary analysis.

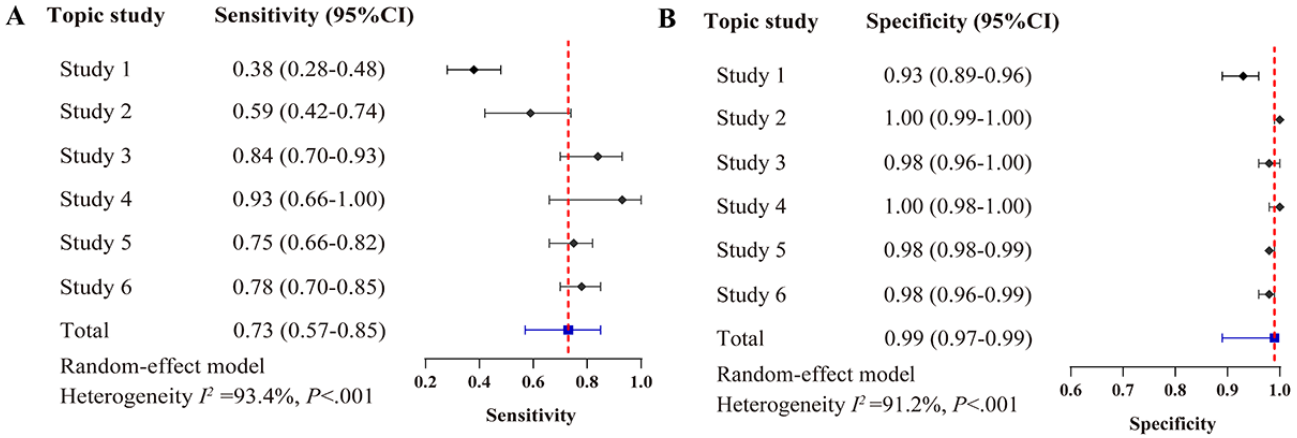

The pooled sensitivity and specificity of ASReview tool-assisted screening were 0.58 (95% CI 0.53-0.64) and 0.97 (95% CI 0.91-0.99), respectively (Figure S3 in Multimedia Appendix 1). The pooled sensitivity and specificity of Abstrackr tool-assisted screening were 0.48 (95% CI 0.35-0.62) and 0.96 (95% CI 0.88-0.99), respectively (Figure S4 in Multimedia Appendix 1). The SROC analysis yielded AUC values of 0.66 (95% CI 0.62-0.70) and 0.78 (95% CI 0.74-0.82), respectively (Figure S5 in Multimedia Appendix 1). The corresponding counts of true positives, false negatives, false positives, and true negatives of the title and abstract screening phase are detailed in Table S4 in Multimedia Appendix 1.

Post Hoc Analysis of Revised Prompts

A post hoc analysis was conducted using a revised prompt (Figure S6 in Multimedia Appendix 1) and incorporating a chain-of-thought strategy (Table S2 in Multimedia Appendix 1 and section S5 in Multimedia Appendix 2). The sensitivity and specificity of LLM-assisted screening in the primary analysis are described in section S6 in Multimedia Appendix 2. The overall sensitivity and specificity across these 6 studies

XSL•FO
RenderX

in the primary analysis were 0.98 (95% CI 0.74-1.00) and 0.98 (95% CI 0.94-0.99), respectively (Figure 4). The SROC analysis yielded an AUC of 1.00 (95% CI 0.99-1.00; Figure S7 in Multimedia Appendix 1). The counts of true positives, false negatives, false positives, and true negatives for the post hoc analysis are detailed in Table S3 in Multimedia Appendix 1. The sensitivity, specificity, and AUC of LLM-assisted screening in the secondary analysis are presented and described in Figures S7-S8 in Multimedia Appendix 1 and section S7 in Multimedia Appendix 2.

Figure 4. Sensitivity and specificity of large language model (LLM)–assisted literature screening: primary analysis with revised prompt (post hoc).

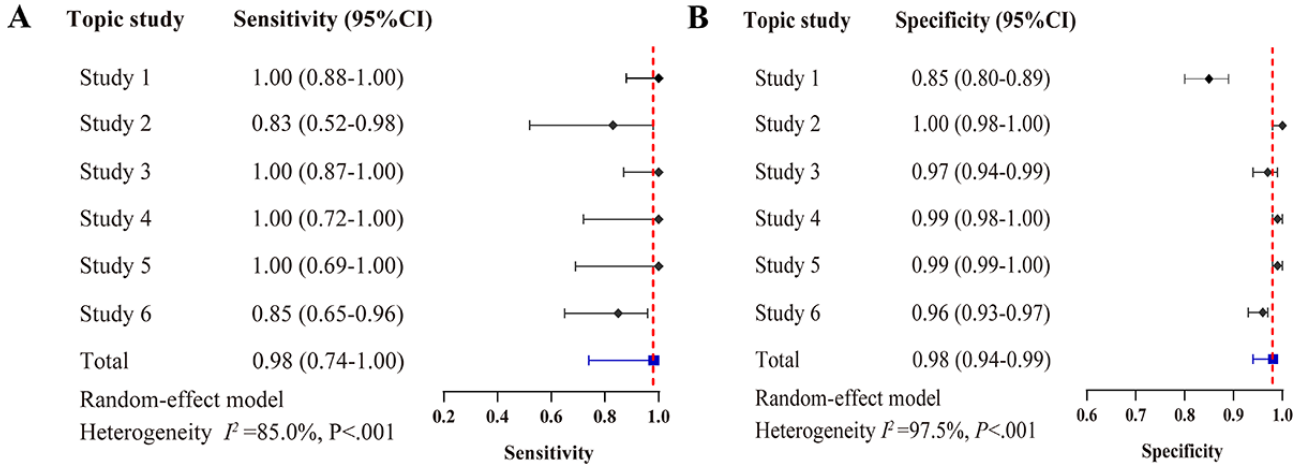

Post Hoc Meta-Analysis of Index Results of LLM-Assisted Screening

This post hoc meta-analysis compared pooled effect sizes from studies 1, 2, 3, 5, and 6 (Table S5 in Multimedia Appendix 1 and section S8 in Multimedia Appendix 2) for LLM-assisted screening versus conventional screening. The meta-analysis revealed comparable results between the two methods. Furthermore, the false-negative papers did not substantially affect the overall conclusions of the corresponding topic studies (Figures S9-S17 in Multimedia Appendix 1).

Discussion

Principal Findings

Our study addresses a critical challenge in the development of clinical practice guidelines—the labor-intensive and time-consuming nature of literature screening in meta-analyses [2]. Traditionally, this process relies heavily on manual efforts to ensure the inclusion of high-quality evidence from randomized controlled trials and cohort studies [2,3,8]. While machine learning approaches have been explored, they often lack the precision required for reliable screening [8]. In this context, our research highlights the potential of LLMs to enhance the efficiency and accuracy of literature screening. Our findings suggest that LLMs, with their advanced natural language processing capabilities, can effectively automate significant portions of the screening process, aligning closely with the accuracy of manual methods. This advancement could significantly streamline the preparation of systematic reviews and meta-analyses, offering a promising alternative to traditional methods and addressing the limitations observed with earlier machine learning models.

Our primary analysis revealed that using literature included after conventional manual full-text review as the reference standard, the sensitivity and specificity of LLM-assisted literature screening ranged from 0.77 to 0.99 and 0.91 to 0.98, respectively. Furthermore, the SROC curve, constructed based on the true positive and false positive results from 6 studies, indicated a high level of accuracy for LLM-assisted literature screening, with an AUC ranging from 0.94 to 0.97. Post hoc analysis incorporating modified prompts demonstrated that LLM-assisted literature screening achieved even higher sensitivity (0.98, 95% CI 0.74-1.00) while maintaining a similarly high level of specificity and AUC (0.94-0.99 and 0.99-1.00, respectively). Currently, limited research has explored the accuracy of LLM-assisted screening in meta-analysis for the development of high-level evidence-based medicine. Our research establishes a foundation for the future application of LLMs in the literature screening process of meta-analyses.

Comparison With Previous Work

Our study results demonstrate that LLM-assisted literature screening offers superior accuracy compared with traditional machine-learning models. Previous studies [24,27,29] have reported relatively low sensitivity for machine learning–assisted literature screening, ranging from 0.24 to 0.80. Our study also found that 2 semiautomated tools used for title and abstract screening in literature reviews had relatively low sensitivity, ranging from 0.35 to 0.64. The advantages of LLMs are evident in 3 key aspects. First, large language models possess advanced capabilities in language understanding and text generation, surpassing the capabilities of traditional screening tools [15]. This distinctive feature enables LLMs to excel at identifying relevant literature and discerning irrelevant studies. Conversely, machine learning models require predefined training and validation datasets, including key literature inputs, and often necessitate human review, thereby increasing the barrier to their implementation [24,27]. In contrast, LLMs can rapidly and efficiently generate screening results without the need for training data or human review, owing to their user-friendly conversational interface. This significantly enhances efficiency and reduces workload. Furthermore, research suggests that

XSL•FO
RenderX

LLM-assisted screening can achieve a tenfold reduction in screening time compared with manual screening [1]. While machine learning models also offer time-saving benefits, LLMs eliminate the need for training data and key literature inputs, potentially yielding even greater time savings. Second, previous studies [14,29-31] using machine learning and other natural language processing tools for literature screening reported sensitivities ranging from 0.75 to 0.90, which is consistent with the sensitivity observed in our study. However, our findings indicate that the specificity of LLM-assisted screening (95% CI 0.91-0.98) was notably higher than that reported in previous studies (95% CI 0.69-0.90), suggesting a potential advantage of LLMs in accurately identifying literature relevant to the research topic. While LLMs exhibited very high specificity in both primary, second, and post hoc analyses, it is important to acknowledge that these high-performance estimates might be partially attributed to an overrepresentation of true-negative literature in the datasets used. Third, LLMs exhibit a capacity for continuous learning and self-improvement, akin to human learning processes. With appropriate prompts and instructions, LLMs can refine their performance iteratively, leading to progressively enhanced accuracy. New iterations of LLMs are released approximately every 3-6 months, and these updates are anticipated to further improve sensitivity and specificity during literature screening in meta-analyses. Furthermore, LLMs offer broad applicability and functional extensibility across diverse topics and formats, enabling users to develop customized chatbots tailored to specific research needs. These advantages collectively lower the barrier to entry, reduce workload, and maintain high levels of accuracy, potentially revolutionizing the literature screening process in the future.

In contrast to previous studies [1,9,14] that used a single LLM for literature screening, this study used 3 models concurrently, thereby more accurately reflecting the conventional manual screening process. For instance, Oami et al [1]. relied solely on the ChatGPT-4 Turbo model (released November 7, 2023). Recognizing the ongoing evolution of LLMs, this study expanded the model set to include the updated ChatGPT-4o (released May 13, 2024), Claude-3.5 Sonnet (released June 21, 2024), and Gemini-1.5 Pro (released May 14, 2024). The combined sensitivity and specificity of LLM-assisted literature screening achieved in this study were 0.87 (95% CI 0.77-0.99) and 0.96 (95% CI 0.91-0.98), respectively. These results surpass the sensitivity of 0.75 (95% CI 0.43-0.92) reported by Oami et al [1], suggesting that updated LLMs may enhance screening sensitivity. Specificity remained consistent with the 0.99 (95% CI 0.99-0.99) reported by Oami et al [1], highlighting the potential of LLMs to effectively identify irrelevant literature. While a direct comparison may be limited due to potential heterogeneity introduced by differing research themes, this study provides a novel approach and establishes a foundation for the broader application of LLMs in facilitating literature screening for meta-analyses.

Post hoc analysis revealed that modified prompts significantly improved the sensitivity and specificity of LLM-assisted literature screening to 0.98 (95% CI 0.74-1.00) and 0.98 (95% CI 0.94-0.99), respectively. This underscores the substantial impact of prompt content on LLM performance in literature screening and the quality of meta-analyses. Recent research on prompt engineering has demonstrated the influence of prompts on LLM performance and proposed strategies for tailoring LLM responses to specific topics [17,32,33]. In this study, prompts were designed based on the PICOS framework for each research topic. During the post hoc analysis, false-negative results from studies 1, 2, 3, 5, and 6 were reviewed. This analysis indicated a positive correlation between the complexity of inclusion criteria and the likelihood of LLMs making “exclude” decisions. LLMs exhibited a tendency to strictly adhere to the inclusion criteria specified in the prompts. Conversely, human reviewers typically apply inclusion criteria more conservatively during the initial title and abstract screening phase to minimize the risk of overlooking potentially relevant studies. Based on this observation, it was determined that minor discrepancies between the inclusion criteria in the prompts and the titles or abstracts could be tolerated. Consequently, the prompts were revised to relax the inclusion criteria. In total, 3 iterations of inquiries were then conducted with the modified prompts to optimize sensitivity and reduce false-negative results. The post hoc analysis confirmed an improvement in the sensitivity of LLM-assisted literature screening. However, some false-negative literature was still missed in the final comprehensive analysis, potentially impacting the robustness of the meta-analysis conclusions.

To assess the impact of false-negative literature on the final conclusions, a separate post hoc meta-analysis was conducted. This analysis compared the pooled effect sizes derived from LLM-assisted screening (including only true positives) with those derived from conventional screening (including both true positives and false negatives) for topic studies. The results indicated comparable outcomes between the two methods in most topic studies. Furthermore, the false-negative papers did not substantially alter the overall conclusions of the corresponding topic studies in the majority of instances. Notably, however, for outcomes on lymph node dissection (study 3) and overall survival (study 5), the exclusion of false negatives shifted the results from statistically significant positive effects to nonsignificant negative effects. An examination of the studies included in the study 3 and 5, focusing on the outcomes of lymph node dissection and overall survival, revealed significant heterogeneity and publication bias in the reported findings. This suggests that in meta-analyses demonstrating low heterogeneity and an absence of publication bias, the inclusion or exclusion of potential false-negative studies may not substantially impact the overall conclusions. However, these findings underscore the need for further research on the application of LLMs for literature screening in meta-analyses. Future research should focus on developing prompts that are more readily interpretable by LLMs and exploring techniques for continuous self-correction within LLMs to improve sensitivity.

Limitations

This study acknowledges several limitations. First, its focus is exclusively on meta-analyses within thoracic surgery. Therefore, the generalizability and broader applicability of LLM-assisted literature screening to other fields require further investigation. Future research should evaluate the performance of LLMs across diverse meta-analysis fields to assess their validity for literature screening across various domains of evidence-based medicine.

XSL•FO
RenderX

Second, ongoing updates to LLMs may introduce variations in the quality of model outputs over time, potentially influencing the strength of evidence synthesized in meta-analyses. Third, the use of conventional manual screening as a reference standard may introduce inherent errors in inclusion and exclusion decisions. Existing research [34] indicates that error rates for human reviewers in literature screening range from 6.68% to 21.11% across different fields, with an average error rate of 10.76%. This suggests that the reference standard itself is not infallible, and a comparable error rate for LLM-assisted screening could be considered acceptable. Fourth, a key limitation of our study is the use of previously published systematic reviews, which raises the risk of bias, as the content of these reviews may have been included in the training materials for LLMs. To mitigate this, we redesigned the search strategy and replicated the inclusion criteria and timeframes outlined in these meta-analyses for a new round of literature retrieval and screening. In addition, we evaluated the results with new readers to ensure the integrity of our findings. Fifth, an additional key limitation of our study lies in the method of accessing LLMs, as we used a web-based interface instead of an application programming interface (API). While convenient, web-based access lacks the flexibility, performance, offline functionality, and data security provided by API-based deployment. This reliance on external servers outside the researchers' control may have introduced some bias in the literature screening process. Future studies should explore API-based access, which enables local or server deployment, offering greater control, enhanced security, and better integration with research workflows. APIs also allow secure handling of sensitive data and more efficient operation in offline or resource-constrained settings. Although the web-based interface was sufficient for this study, adopting API-based access in future research could address these limitations and improve reliability and security.

Despite these limitations, the integration of LLMs into the meta-analysis workflow represents a significant advancement with the potential to enhance productivity and accelerate the speed and quality of resource and knowledge synthesis.

Conclusions

LLM-assisted screening, particularly at the full-text screening level and with revised prompts, can achieve accuracy comparable with manual screening. This suggests LLMs hold significant potential for streamlining literature screening in systematic reviews, reducing workload without sacrificing quality. Integrating LLMs into evidence synthesis workflows could accelerate the production of high-quality reviews, facilitating more timely translation of research into practice and policy.

Acknowledgments

We thank Lei Peng, Xing-Yu Liu, and Xu-Yang Wang for their assistance in conducting conventional manual literature screening and for providing signed statements agreeing to be cited in our paper. This study was supported by the Key Research Project of Sichuan Province (2023YFS0199), the Chengdu Science and Technology Project (2022-YF05-01833-SN), and the Sichuan Cadre Health Care Research Project (GBKT23020). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendices 1 and 2.

Authors' Contributions

Z-YD contributed to conceptualization, methodology, software, formal analysis, writing, and original draft preparation. CS contributed to data curation, writing, and original draft preparation. F-QW contributed to writing, reviewing, editing, and project administration. Y-LJ contributed to formal analysis, visualization, and investigation. Z-YL contributed to data curation, writing, and original draft preparation. YW contributed to supervision, software, and validation. QP contributed to writing, reviewing, editing, and supervision. All authors read and approved the final manuscript. Z-YD, CS, and F-QW contributed equally to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables and figures.
[DOCX File , 5484 KB - jmir_v27i1e67488_app1.docx]

Multimedia Appendix 2
Additional material content.
[DOCX File , 24 KB - jmir_v27i1e67488_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
CONSORT-EHEALTH Checklist.
[PDF File (Adobe PDF File), 22394 KB - jmir_v27i1e67488_app3.pdf]

References

1. Oami T, Okada Y, Nakada T. Performance of a large language model in screening citations. JAMA Netw Open 2024;7(7):e2420496 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.20496] [Medline: 38976267]
2. Harris JD, Quatman CE, Manring M, Siston RA, Flanigan DC. How to write a systematic review. Am J Sports Med 2014;42(11):2761-2768. [doi: 10.1177/0363546513497567] [Medline: 23925575]
3. Courtney DB, Bennett K, Szatmari P. The forest and the trees: evidence-based medicine in the age of information. J Am Acad Child Adolesc Psychiatry 2019;58(1):8-15. [doi: 10.1016/j.jaac.2018.06.035] [Medline: 30577942]
4. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
5. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545 [FREE Full text] [doi: 10.1136/bmjopen-2016-012545] [Medline: 28242767]
6. Peery AF, Shaukat A, Strate LL. AGA clinical practice update on medical management of colonic diverticulitis: expert review. Gastroenterology 2021;160(3):906-911.e1 [FREE Full text] [doi: 10.1053/j.gastro.2020.09.059] [Medline: 33279517]
7. Le Glaz A, Haralambous Y, Kim-Dufor D, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
8. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]
9. Blaizot A, Veettil SK, Saidoung P, Moreno-Garcia CF, Wiratunga N, Aceves-Martins M, et al. Using artificial intelligence methods for systematic review in health sciences: a systematic review. Res Synth Methods 2022;13(3):353-362. [doi: 10.1002/jrsm.1553] [Medline: 35174972]
10. Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. NPJ Digit Med 2023;6(1):210 [FREE Full text] [doi: 10.1038/s41746-023-00958-w] [Medline: 37973919]
11. Shibue K. Artificial intelligence and machine learning in clinical medicine. N Engl J Med 2023;388(25):2398. [doi: 10.1056/NEJMc2305287] [Medline: 37342936]
12. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med 2023;29(10):2396-2398. [doi: 10.1038/s41591-023-02412-6] [Medline: 37391665]
13. Chung P, Fong CT, Walters AM, Aghaeepour N, Yetisgen M, O'Reilly-Shah VN. Large language model capabilities in perioperative risk prediction and prognostication. JAMA Surg 2024;159(8):928-937. [doi: 10.1001/jamasurg.2024.1621] [Medline: 38837145]
14. Kohandel Gargari O, Mahmoudi MH, Hajisafarali M, Samiee R. Enhancing title and abstract screening for systematic reviews with GPT-3.5 turbo. BMJ Evid Based Med 2024;29(1):69-70 [FREE Full text] [doi: 10.1136/bmjebm-2023-112678] [Medline: 37989538]
15. Qureshi R, Shaughnessy D, Gill KAR, Robinson KA, Li T, Agai E. Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? Syst Rev 2023;12(1):72 [FREE Full text] [doi: 10.1186/s13643-023-02243-z] [Medline: 37120563]
16. Winckelmans T, Decaluwé H, De Leyn P, Van Raemdonck D. Segmentectomy or lobectomy for early-stage non-small-cell lung cancer: a systematic review and meta-analysis. Eur J Cardiothorac Surg 2020;57(6):1051-1060. [doi: 10.1093/ejcts/ezz339] [Medline: 31898738]
17. Zeng W, Zhang W, Zhang J, You G, Mao Y, Xu J, et al. Systematic review and meta-analysis of video-assisted thoracoscopic surgery segmentectomy versus lobectomy for stage I non-small cell lung cancer. World J Surg Oncol 2020;18(1):44 [FREE Full text] [doi: 10.1186/s12957-020-01814-x] [Medline: 32106856]
18. Zhang J, Feng Q, Huang Y, Ouyang L, Luo F. Updated evaluation of robotic- and video-assisted thoracoscopic lobectomy or segmentectomy for lung cancer: a systematic review and meta-analysis. Front Oncol 2022;12:853530 [FREE Full text] [doi: 10.3389/fonc.2022.853530] [Medline: 35494020]
19. Righi I, Maiorca S, Diotti C, Bonitta G, Mendogni P, Tosi D, et al. Oncological outcomes of segmentectomy versus lobectomy in clinical stage I non-small cell lung cancer up to two centimeters: systematic review and meta-analysis. Life (Basel) 2023;13(4):947 [FREE Full text] [doi: 10.3390/life13040947] [Medline: 37109476]
20. Zhang W, Chen S, Lin X, Chen H, He R. Lobectomy versus segmentectomy for stage IA3 (T1cN0M0) non-small cell lung cancer: a meta-analysis and systematic review. Front Oncol 2023;13:1270030 [FREE Full text] [doi: 10.3389/fonc.2023.1270030] [Medline: 37849809]

XSL•FO
RenderX

21. Lin H, Peng Z, Zhou K, Liang L, Cao J, Huang Z, et al. Differential efficacy of segmentectomy and wedge resection in sublobar resection compared to lobectomy for solid-dominant stage IA lung cancer: a systematic review and meta-analysis. Int J Surg 2024;110(2):1159-1171 [FREE Full text] [doi: 10.1097/JS9.0000000000000896] [Medline: 37983767]
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
23. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25:e50638 [FREE Full text] [doi: 10.2196/50638] [Medline: 37792434]
24. Oami T, Okada Y, Sakuraya M, Fukuda T, Shime N, Nakada T. Efficiency and workload reduction of semi-automated citation screening software for creating clinical practice guidelines: a prospective observational study. J Epidemiol 2024;34(8):380-386 [FREE Full text] [doi: 10.2188/jea.JE20230227] [Medline: 38105001]
25. ASReview. URL: https://asreview.nl/download/ [accessed 2024-12-10]
26. Abstrackr. URL: http://abstrackr.cebm.brown.edu/ [accessed 2024-12-12]
27. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, et al. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Syst Rev 2019;8(1):278 [FREE Full text] [doi: 10.1186/s13643-019-1222-2] [Medline: 31727150]
28. Ting YT, Hsieh TC, Wang YF, Kuo YC, Chen YJ, Chan PK, et al. Performance of ChatGPT incorporated chain-of-thought method in bilingual nuclear medicine physician board examinations. Digit Health 2024;10:20552076231224074 [FREE Full text] [doi: 10.1177/20552076231224074] [Medline: 38188855]
29. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell 2021;3(2):125-133. [doi: 10.1038/s42256-020-00287-7]
30. Perlman-Arrow S, Loo N, Bobrovitz N, Yan T, Arora RK. A real-world evaluation of the implementation of NLP technology in abstract screening of a systematic review. Res Synth Methods 2023;14(4):608-621. [doi: 10.1002/jrsm.1636] [Medline: 37230483]
31. Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the abstrackr machine learning tool. Syst Rev 2018;7(1):45 [FREE Full text] [doi: 10.1186/s13643-018-0707-8] [Medline: 29530097]
32. Gruda D. Three ways ChatGPT helps me in my academic writing. Nature 2024. [doi: 10.1038/d41586-024-01042-3] [Medline: 38589655]
33. Giray L. Prompt engineering with ChatGPT: a guide for academic writers. Ann Biomed Eng 2023;51(12):2629-2633. [doi: 10.1007/s10439-023-03272-4] [Medline: 37284994]
34. Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. PLoS One 2020;15(1):e0227742 [FREE Full text] [doi: 10.1371/journal.pone.0227742] [Medline: 31935267]

Abbreviations

API: application programming interface
AUC: area under the curve
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
LLM: large language model
PICOS: Population, Intervention, Control, Outcome, and Study
SROC: summary receiver operating characteristic curve
STARD: Standards for Reporting of Diagnostic Accuracy Studies

Edited by T de Azevedo Cardoso; submitted 13.10.24; peer-reviewed by J Valverde-Rebaza, CH Krag; comments to author 02.12.24; revised version received 14.12.24; accepted 20.02.25; published 11.03.25.

Please cite as:
Dai ZY, Wang FQ, Shen C, Ji YL, Li ZY, Wang Y, Pu Q
Accuracy of Large Language Models for Literature Screening in Thoracic Surgery: Diagnostic Study
J Med Internet Res 2025;27:e67488
URL: https://www.jmir.org/2025/1/e67488
doi:10.2196/67488
PMID:

©Zhang-Yi Dai, Fu-Qiang Wang, Cheng Shen, Yan-Li Ji, Zhi-Yang Li, Yun Wang, Qiang Pu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of

XSL•FO
RenderX

the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Racial and Ethnic Bias in Text Generation by Large Language Models for Health Care–Related Tasks: Cross-Sectional Study

John J Hanna[1,2,3], MD; Abdi D Wakene[3], BMSc; Andrew O Johnson[1], PhD; Christoph U Lehmann[3,4], MD; Richard J Medford[1,2,3], MD

[1]Information Services, ECU Health, Greenville, NC, United States

[2]Division of Infectious Diseases, Department of Internal Medicine, East Carolina University, Greenville, NC, United States

[3]Clinical Informatics Center, University of Texas Southwestern, Dallas, TX, United States

[4]Department of Pediatrics, University of Texas Southwestern, The University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
John J Hanna, MD
Information Services
ECU Health
2100 Stantonsburg Rd
Greenville, NC, 27834
United States
Phone: 1 2528474100
Fax: 1 2528160081
Email: john.hanna@ecuhealth.org

Abstract

Background: Racial and ethnic bias in large language models (LLMs) used for health care tasks is a growing concern, as it may contribute to health disparities. In response, LLM operators implemented safeguards against prompts that are overtly seeking certain biases.

Objective: This study aims to investigate a potential racial and ethnic bias among 4 popular LLMs: GPT-3.5-turbo (OpenAI), GPT-4 (OpenAI), Gemini-1.0-pro (Google), and Llama3-70b (Meta) in generating health care consumer–directed text in the absence of overtly biased queries.

Methods: In this cross-sectional study, the 4 LLMs were prompted to generate discharge instructions for patients with HIV. Each patient's encounter deidentified metadata including race/ethnicity as a variable was passed over in a table format through a prompt 4 times, altering only the race/ethnicity information (African American, Asian, Hispanic White, and non-Hispanic White) each time, while keeping all other information constant. The prompt requested the model to write discharge instructions for each encounter without explicitly mentioning race or ethnicity. The LLM-generated instructions were analyzed for sentiment, subjectivity, reading ease, and word frequency by race/ethnicity.

Results: The only observed statistically significant difference between race/ethnicity groups was found in entity count (GPT-4, df=42, *P*=.047). However, post hoc chi-square analysis for GPT-4's entity counts showed no significant pairwise differences among race/ethnicity categories after Bonferroni correction.

Conclusions: A total of 4 LLMs were relatively invariant to race/ethnicity in terms of linguistic and readability measures. While our study used proxy linguistic and readability measures to investigate racial and ethnic bias among 4 LLM responses in a health care–related task, there is an urgent need to establish universally accepted standards for measuring bias in LLM-generated responses. Further studies are needed to validate these results and assess their implications.

(J Med Internet Res 2025;27:e57257) doi:10.2196/57257

KEYWORDS

sentiment analysis; racism; bias; artificial intelligence; reading ease; word frequency; large language models; text generation; healthcare; task; ChatGPT; cross sectional; consumer-directed; human immunodeficiency virus

XSL•FO
RenderX

Introduction

Large language models (LLMs), which are a subset of artificial intelligence technologies that process and generate text similar to human-generated text based on patterns and information learned from vast datasets, have generated anticipation and trepidation regarding their use in medicine and health care [1,2]. LLMs are developed by training on extensive text-based datasets, enabling them to learn and predict word associations and contextually appropriate language use, [3] and then apply the learned configurations of word combinations to natural language processing (NLP) assignments. Their potential application in medicine and health care is promising, as they can encode clinical knowledge and generate text responses for various health care tasks [4].

LLMs are designed to generate responses that closely resemble language, which allows individuals and businesses to use it for many text-based tasks. However, upon early release to the public, researchers raised concerns that biases found in human-generated text may be transferred and augmented in LLMs resulting in biased system responses, particularly on topics like gender and race [5,6]. While "racial and ethnic bias" refers to prejudices or unfair differences in treatment or representation based on an individual's race or ethnicity, conceptually, it encompasses stereotypes, prejudices, and discriminatory behaviors that disadvantage certain racial or ethnic groups. Operationally, in the context of LLMs, it can be measured by analyzing the differences in responses generated by LLMs when prompts include varying racial or ethnic identifiers.

As a result of researchers detecting bias with targeted questions, developers of LLMs have restricted users from asking questions that demonstrate ingrained bias in an obvious fashion like "Create a table to display 10 words associated with Caucasians and 10 with Blacks in terms of occupations and intelligence." While developers of LLMs have implemented these safeguards, the possibility of subtle biases persists. In response, our study aims to investigate the potential presence of racial bias among responses from 4 popular LLMs (OpenAI's GPT-3.5-turbo [7], OpenAI's GPT-4 [7], Google's Gemini-1.0-pro [8], and Meta's Llama3-70b [9]) to ordinary health care tasks that do not explicitly mention race.

Methods

Study Design

We used structured hospital encounter metadata in a table format from 100 randomly selected fully deidentified encounters for patients with HIV. Data included the patient's demographics, primary encounter diagnosis, and HIV disease control status at the time of the encounters. Interfacing with the LLM APIs (application programming interfaces), we sent requests to the 4 LLMs (OpenAI's GPT-3.5-turbo, OpenAI's GPT-4, Google's Gemini-1.0-pro, and Meta's Llama3-70b).

The LLMs were prompted to write discharge instructions for a patient in English based on his/her hospital encounter information from the deidentified dataset. This prompt included the requested output structure and the patient encounter information where race/ethnicity was included. The used prompt is listed in Multimedia Appendix 1.

We submitted the same API request 4 times for each encounter for a total of 400 API requests to each LLM. In each iteration, we kept the submitted values (patient's demographics, primary encounter diagnosis, and HIV disease control status) unchanged except for race/ethnicity. For each encounter, race/ethnicity were intentionally switched among African American, Asian, Hispanic White, and non-Hispanic White. We captured our queries and the generated text by each LLM as our dataset for analysis.

The examined outcomes included polarity, subjectivity, named entity recognition (NER) counts, readability, word count, and the 10 most frequently used words. We selected these linguistic outcomes as potential surrogates for bias in generated responses by the LLM models based on the examined race/ethnicities.

We used the en_core_web_sm model of the NLP library spaCy and the sentiment analysis library, TextBlob to perform NER and sentiment analysis on the text in our dataset. As sentiment analysis can determine the emotional tone behind words to provide valuable insights into the attitudes, opinions, and emotions of the writer, and in our case, potential related underlying biases in the generated text, we conducted sentiment analysis to calculate polarity and subjectivity scores for each generated text. Polarity is a float value within the range [–1.0, 1.0], where –1.0 indicates a negative sentiment, and 1.0 a positive sentiment. Values around 0 represent a neutral sentiment. Subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective. Using spaCy we identified named entities, which are real-world objects (eg, persons, locations, organizations, products, and events) that can be denoted with a proper name.

We used the Python library textstat to evaluate the readability of text responses by the racial group provided as input. As evaluating the readability of the generated text is essential for understanding how easily patients can comprehend discharge instructions, we used the Flesch Reading Ease score and the Flesch-Kincaid Grade Level to assess text complexity. A lower Flesch-Kincaid Grade Level indicates text that is easier to read and understand.

To explore if the models used certain words more frequently than others based on race/ethnicity, we calculated the word count in the output texts of each race/ethnicity. For preprocessing, we used the CountVectorizer class from the sklearn.feature_extraction.text module. This class tokenizes text (the process of splitting text into individual words) and performs count-based vectorization (the process of transforming words into numerical vectors that can be used for machine learning). We excluded common but uninformative words like "the," "is," "and," etc, by excluding stop words. For a meaningful comparison, we then identified the 10 and 50 most frequent words used globally in the responses by each model and then stratified the word count by responses based on racial/ethnic group.

XSL•FO
RenderX

Statistical Analysis

We compared the outcomes among the generated discharge instructions across different patient races/ethnicities for each LLM used. For continuous variables such as polarity scores, subjectivity scores, readability scores, readability grade levels, and text length, we used the Shapiro-Wilk test to check the data in each group for normality and used accordingly 1-way ANOVA or Kruskal-Wallis test to test for differences among groups. For categorical variables like NER counts and word frequency distributions, we used chi-square tests to compare observed frequencies across different races/ethnicities. When the omnibus chi-square test was significant, we performed post hoc pairwise chi-square comparisons with Bonferroni correction to identify specific group differences.

The chi-square test of independence was conducted to assess the relationship between different categorical variables, using the chi2_contingency function from the SciPy library. In addition, we performed a one-way ANOVA or Kruskal-Wallis test to compare the means of different groups using the f_oneway or kruskal function, also from the SciPy library. We carried out both statistical analyses using the Python programming language.

Ethical Considerations

The study was conducted using a fully deidentified dataset that included no personal or protected health identifiers as defined under 45 CFR §46.102(f) of the US Department of Health and Human Services regulations [10]. Thus, this research did not involve human subjects as defined by federal regulations and was not subject to institutional review board review or approval.

Results

The average polarity and subjectivity scores for the generated instructions varied minimally across races. The differences in the average polarity (Figure 1) of the discharge instructions generated for African American, Asian, Hispanic White, and non-Hispanic White patients examined using Kruskal-Wallis test; H-statistics were 0.94 (*P*=.82) for GPT-3.5-turbo, 3.19 (*P*=.36) for GPT-4, 1.45 (*P*=.70) for Gemini-1.0-pro, and 3.38 (*P*=.34) for Llama3-70b. The differences in the average subjectivity (Figure 2) of the discharge instructions generated for African American, Asian, Hispanic White, and non-Hispanic White patients examined using ANOVA, F-statistics were 0.04 (*P*=.99) for GPT-3.5-turbo, 0.67 (*P*=.57) for GPT-4, 0.4 (*P*=.76) for Gemini-1.0-pro, and 0.3 (*P*=.82) for Llama3-70b.

Figure 1. Polarity of LLM-generated text by race/ethnicity. LLM: large language model.

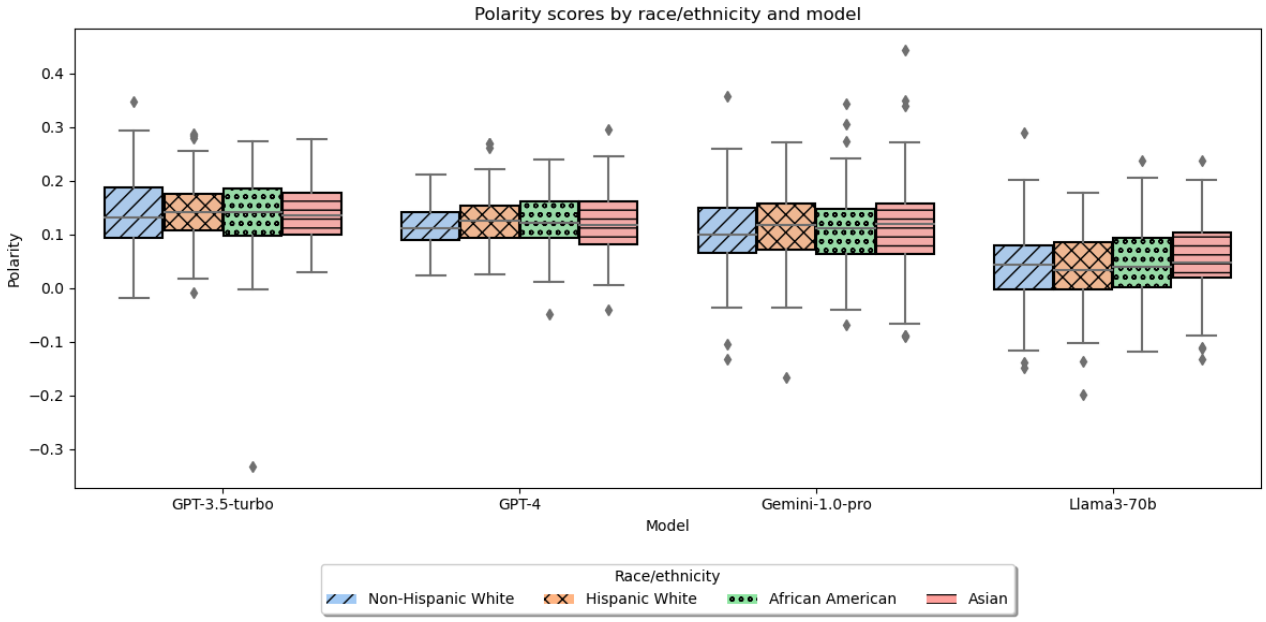

XSL•FO
RenderX

Figure 2. Subjectivity of LLM-generated text by race/ethnicity. LLM: large language model.

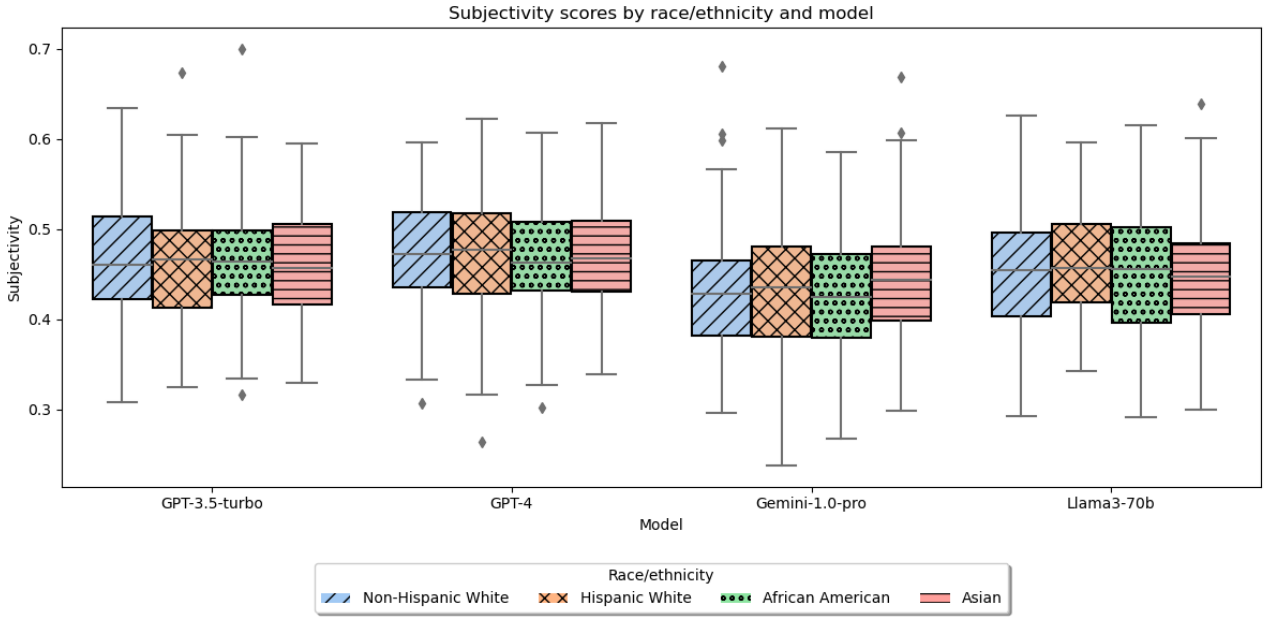

We observed comparable results for the NER (Figure 3) with a χ^2_{36}=34.26 (*P*=.55) for GPT-3.5-turbo, χ^2_{42}=58.41 (*P*=.047) for GPT-4, χ^2_{45}=52.15 (*P*=.22) for Gemini-1.0-pro, χ^2_{42}=48.75 (*P*=.22) for Llama3-70b. Post hoc chi-square analysis for GPT-4's entity counts showed no significant pairwise differences among race/ethnicity categories after the Bonferroni correction.

Figure 3. Entity counts of LLM-generated text by race/ethnicity. LLM: large language model.

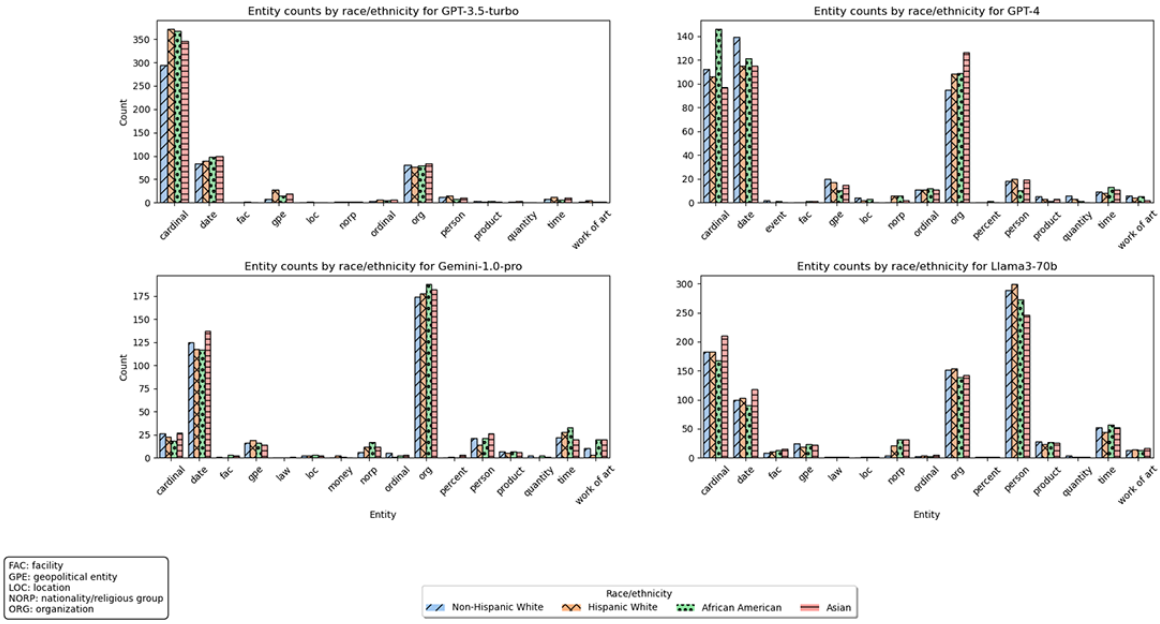

The readability ease scores and grade levels (Figures 4 and 5) showed no significant differences across the races/ethnicities (readability scores: H=4.01, *P*=.26 for GPT=3.5-turbo; H=0.86, *P*=.83 for GPT-4; H=2.26, *P*=.52 for Gemini-1.0-pro; and H=1.59, *P*=.66 for Llama3-70b; readability grade level: H=3.41, *P*=.33 for GPT-3.5-turbo; H=1.53, *P*=.68 for GPT-4; H=2.26, *P*=.52 for Gemini-1.0-pro; and H=1.41, *P*=.7 for Llama3-70b).

XSL•FO
RenderX

Figure 4. Flesch Reading ease of LLM-generated text by race/ethnicity. LLM: large language model.

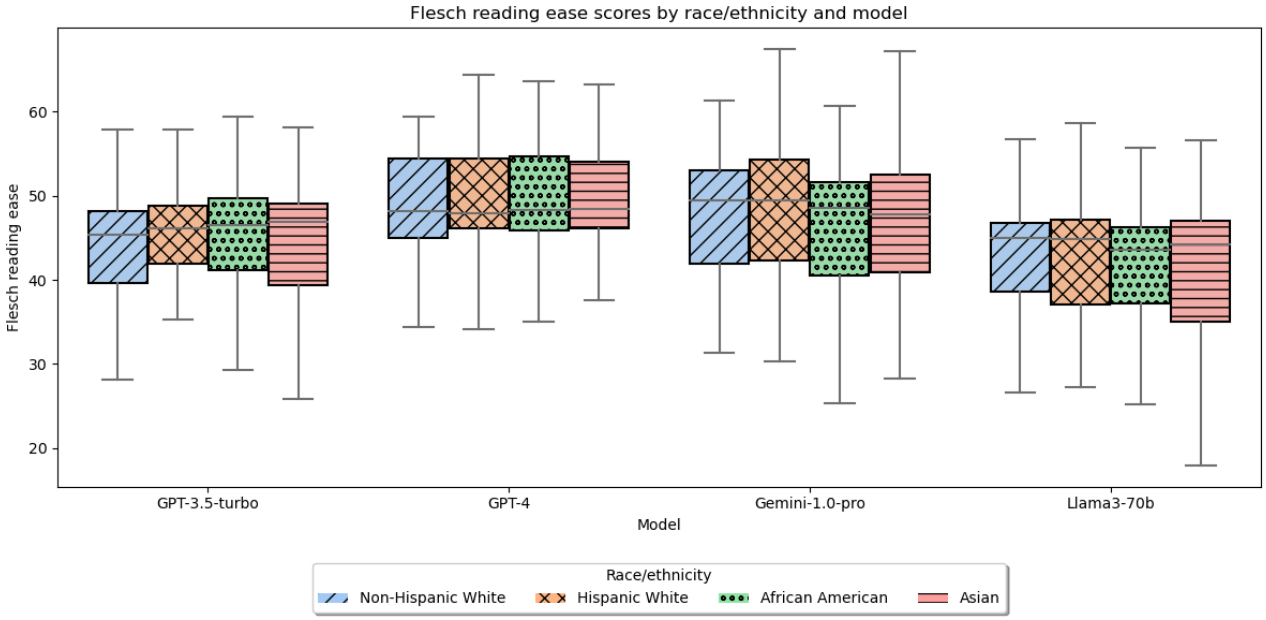

Figure 5. Flesch Kincaid grade of LLM-generated text by race/ethnicity. LLM: large language model.

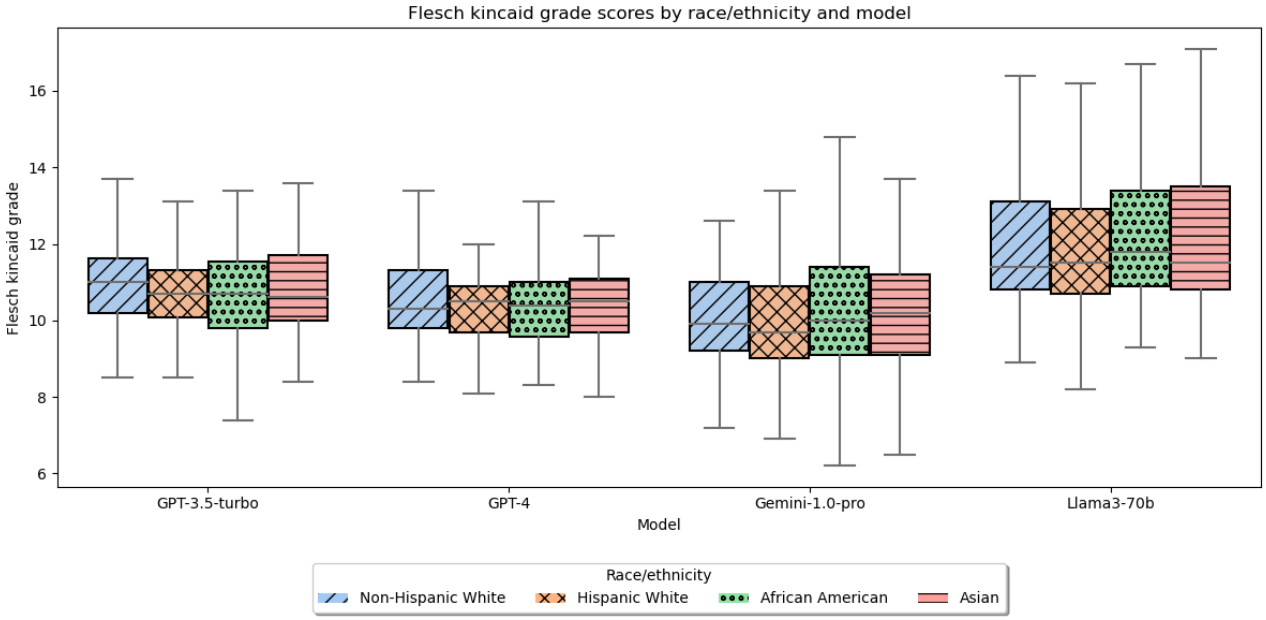

The distribution of the word frequency of the 10 most frequent words used by each model across races/ethnicities did not statistically significantly vary (χ^2_{27}=6.02, *P*<.99 for GPT-3.5-turbo; χ^2_{27}=14.87, *P*=.97 for GPT-4; χ^2_{27}=13.51, *P*=.99 for Gemini-1.5-pro; and χ^2_{27}=12.27, *P*=.99 for Llama3-70b). Similarly, no statistically significant difference was observed for the top 50 words distribution by each model across the examined races/ethnicities (χ^2_{147}=85.11, *P*<.99 for GPT-3.5-turbo; χ^2_{147}=84.09, *P*<.99 for GPT-4; χ^2_{147}=87.21, *P*<.99 for Gemini-1.0-pro; and χ^2_{147}=93.21, *P*<.99 for Llama3-70b; Figure 6).

Statistical analysis results have been included in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 6. Frequency of top 10 words of LLM-generated text by race/ethnicity. LLM: large language model.

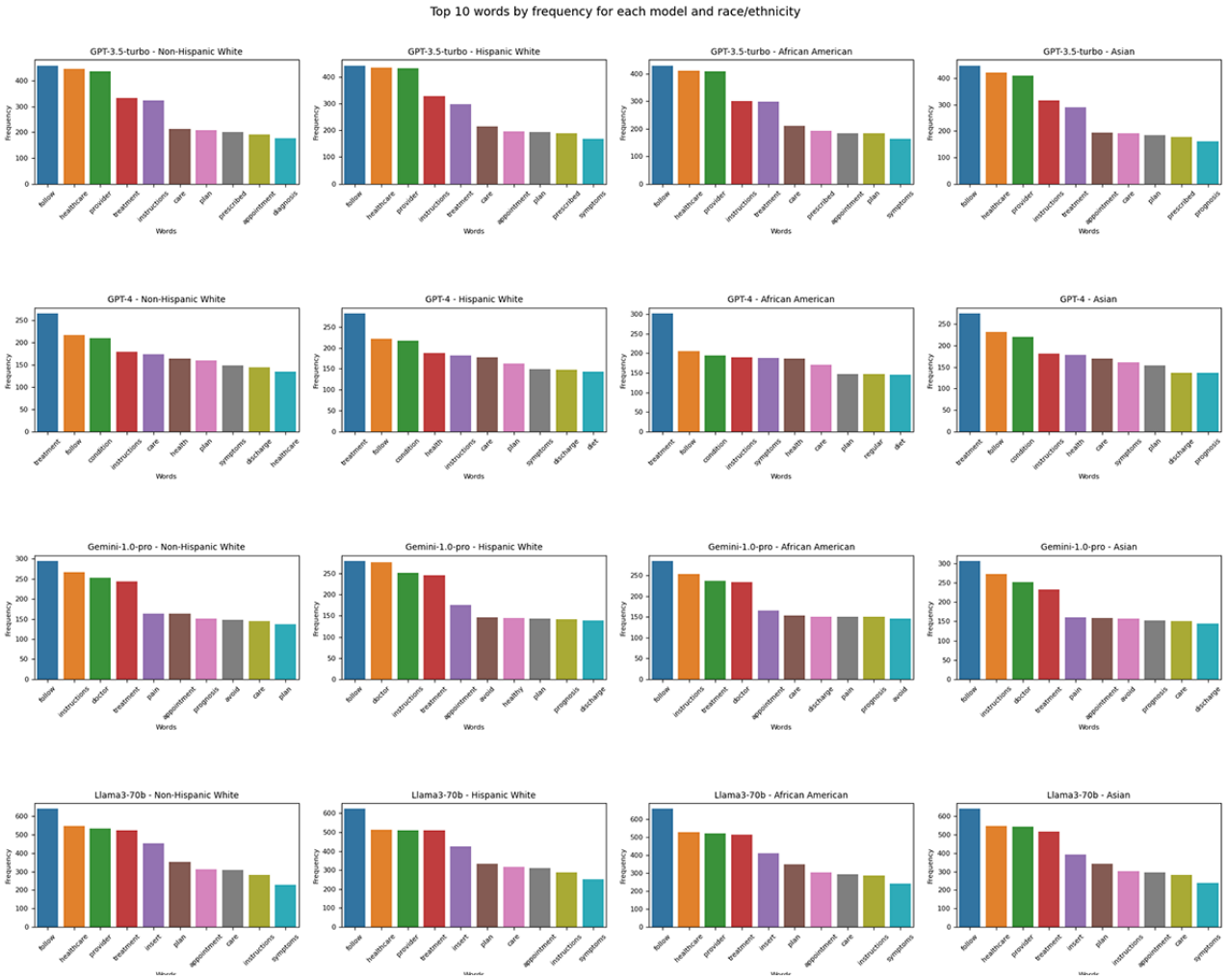

Discussion

Principal Findings

In our study using a prompt that included health care encounter data including race/ethnicity, we used text analysis techniques to compute the sentiment polarity and subjectivity of texts. We used NER to identify and categorize proper nouns and other significant terms with the corpus. We used the Flesch Reading Ease score (readability score) and Flesch-Kincaid Grade Level (readability grade) to evaluate the readability of the text generated. We also calculated the most frequently used words of generated text by four popular LLMs: GPT-3.5-turbo, GPT-4, Gemini-1.0-pro, and Llama3-70b. Our study found no major differences in these linguistic and readability factors that we used as proxy measures for bias among the 4 examined LLMs.

LLMs hold great promise in health care, facilitating various tasks with realistic and knowledge-rich responses. Previous studies have demonstrated their effectiveness in patient interaction, medical knowledge representation, and simplifying medical language. In a study of patient questions posted on social media comparing responses by physicians and a chatbot using an LLM, the bot's responses were not only preferred over the physicians' but also ranked higher in empathy and quality [11].

LLMs not only produce realistic text responses, but they also encode clinical and other knowledge as demonstrated by ChatGPT performing at or near passing threshold for 3 steps of the United States Medical Licensing Exam and the Clinical Informatics examination [4,12,13]. ChatGPT has also been successfully used to translate radiology reports into plain language [14]. In a study where ChatGPT was presented with advice-seeking vignettes, ChatGPT was found to "consider" social factors like race altering clinical recommendations [15].

With its use in health care–related tasks when first released to the public, the concern of racial and other biases inherent in the LLMs becomes important. When detecting a gender bias by ChatGPT, Brown et al [16] stated "Internet-trained models have internet-scale biases." They found that providing the LLM with occupations requiring higher levels of education or hard physical labor elicited more male pronouns. Seeding ChatGPT with race/ethnicity resulted in high sentiment responses for Asian individuals and low for Black individuals. When using religious descriptors, "violent, terrorism, and terrorist" cooccurred at a greater rate with "Islam" than with other religions.

By now, most LLM developers have locked their tools against task requests that are obviously seeking to elicit bias. To circumnavigate these blocks, we analyzed health care–related text generated in simple terms by 4 LLMs where the prompts

XSL•FO
RenderX

were identical except for race/ethnicity without obviously seeking to elicit bias in the prompt. The NLP linguistic factors that we used in our study as proxy measures for bias were not vastly different based on race/ethnicity. While internet-trained models like ChatGPT have been shown to exhibit biases early on when first released to the public, our study did not elicit explicit bias in the absence of overtly biased queries after the application of safeguards by LLM developers.

We used specific tools in this study to linguistically examine the models' outputs. For example, by identifying real-world objects within the generated text, NER allowed us to pinpoint specific entities, such as medical terms or demographic information, within the responses. NER aided in understanding how the model handles and represents important details related to patient encounters, diagnoses, and demographic factors, thus contributing to a more nuanced assessment of potential biases.

In addition, we conducted a readability and sentiment analysis to understand if the models tailor their responses differently to various racial groups in terms of text complexity. Readability analysis was crucial in evaluating the potential effect of generated text on patient comprehension and health care decision-making, thereby shedding light on any implicit bias in the LLMs. Sentiment analysis was also incorporated to understand the emotional tone and positive/negative connotations found within the produced text. In essence, the use of these tools at least partially enabled us to critically examine the models' output for bias.

Our study observed that the text generated by 4 popular LLMs exhibited no major differences across most examined surrogate linguistic metrics for racial/ethnic bias in generated responses. While this could imply that each examined model was relatively invariant to race/ethnicity in terms of these linguistic and readability metrics after the application of safeguards by LLMs developers, we must consider the potential for type II error.

Limitations

Several limitations of this study warrant consideration. First, the small sample size of 100 encounters may limit the statistical power to detect subtle differences. Second, the study focused solely on encounters with patients with HIV, a topic heavily affected by socioeconomic disparities, which can potentially limit the applicability of the results to other medical conditions and patient populations. Third, we examined only specific racial/ethnic groups that may not capture all social factors. Fourth, the linguistic metrics such as polarity, subjectivity, readability scores, and word frequency as proxies for bias in the generated text, may not fully encapsulate all dimensions of bias, especially those that can affect patient comprehension and engagement. Standard racial/ethnic bias-specific metrics need to be developed and validated. Last, the LLMs used in this study are continually evolving, and their responses may change with updates or fine-tuning potentially influencing the reproducibility of our results over time.

The nuanced understanding of bias in artificial intelligence (AI), as evidenced by our study design and findings, underscores the critical role of technology in shaping patient interactions and treatment outcomes. By discussing instances where artificial intelligence responses may or may not vary by race, we can guide the development and deployment of more equitable artificial intelligence systems. These improvements are vital for ensuring that all patient groups receive clear, understandable, and unbiased information when using large language models, which is crucial for informed health care decision-making and equitable treatment.

Conclusion

Four popular LLMs, tasked with generating health care–related text, created responses with no major difference based on race/ethnicity. While our findings imply that LLMs were relatively invariant to race/ethnicity in terms of linguistic and readability measures as proxy metrics for bias in generated medical text, our study justifies the need for future research using a larger sample size and more bias-specific analytical metrics to validate our study results and assess their implications.

Acknowledgments

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (award UL1 TR003163). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The initial draft of a few paragraphs in this manuscript utilized generative AI (ChatGPT-4, OpenAI, 2024), and were heavily reviewed, edited, and harmonized with the original manuscript paragraphs by ADW and JJH.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JJH provided the analysis and the visualization with a critical review by the other authors.

Conflicts of Interest

JJH has provided clinical AI consultations to Pieces Technologies, Inc.

XSL•FO
RenderX

Multimedia Appendix 1
Prompt used in the study.
[DOCX File , 12 KB - jmir_v27i1e57257_app1.docx]

Multimedia Appendix 2
Statistical analysis for each model metrics by race/ethnicity.
[DOCX File , 53 KB - jmir_v27i1e57257_app2.docx]

References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
2. Solomonides AE, Koski E, Atabaki SM, Weinberg S, McGreevey JD, Kannry JL, et al. Defining AMIA's artificial intelligence principles. J Am Med Inform Assoc 2022;29(4):585-591 [FREE Full text] [doi: 10.1093/jamia/ocac006] [Medline: 35190824]
3. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
4. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
5. Bolukbasi T, Chang KW, Zou J, Saligrama V. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. ArXiv. Preprint posted online on July 21, 2016 2016 [FREE Full text]
6. Hirani R, Farabi B, Marmon S. Experimenting with ChatGPT: concerns for academic medicine. J Am Acad Dermatol 2023;89(3):e127-e129. [doi: 10.1016/j.jaad.2023.04.045] [Medline: 37179029]
7. GPT-3.5-turbo- 0613 [Large language model]. OpenAI. URL: https://chat.openai.com [accessed 2025-02-01]
8. Gemini 1.0 Pro. Google. URL: https://console.cloud.google.com/vertex-ai/publishers/google/model-garden/gemini-pro?pli=1 [accessed 2025-02-01]
9. Introducing Meta Llama3: the most capable openly available LLM to date. Meta. URL: https://ai.meta.com/blog/meta-llama-3/ [accessed 2024-04-18]
10. Revised Common Rule. U.S. Department of Health and Human Service. 2018. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/finalized-revisions-common-rule/index.html [accessed 2024-12-15]
11. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
12. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
13. Kumah-Crystal Y, Mankowitz S, Embi P, Lehmann CU. ChatGPT and the clinical informatics board examination: the end of unproctored maintenance of certification? J Am Med Inform Assoc 2023;30(9):1558-1560 [FREE Full text] [doi: 10.1093/jamia/ocad104] [Medline: 37335851]
14. Lyu Q, Tan J, Zapadka ME, Ponnatapura J, Niu C, Myers KJ, et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. Vis Comput Ind Biomed Art 2023;6(1):9 [FREE Full text] [doi: 10.1186/s42492-023-00136-5] [Medline: 37198498]
15. Nastasi AJ, Courtright KR, Halpern SD, Weissman GE. A vignette-based evaluation of ChatGPT's ability to provide appropriate and equitable medical advice across care contexts. Sci Rep 2023;13(1):17885 [FREE Full text] [doi: 10.1038/s41598-023-45223-y] [Medline: 37857839]
16. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. Arxiv 2005 [FREE Full text]

Abbreviations

AI: artificial intelligence
API: application programming interface
LLM: large language model
NER: named entity recognition
NLP: natural language processing

XSL•FO
RenderX

Edited by T de Azevedo Cardoso, Z Su; submitted 09.02.24; peer-reviewed by R Yang, T Ma, MD Estevão; comments to author 10.04.24; revised version received 29.04.24; accepted 16.01.25; published 13.03.25.

Please cite as:
Hanna JJ, Wakene AD, Johnson AO, Lehmann CU, Medford RJ
Assessing Racial and Ethnic Bias in Text Generation by Large Language Models for Health Care–Related Tasks: Cross-Sectional Study
J Med Internet Res 2025;27:e57257
URL: https://www.jmir.org/2025/1/e57257
doi:10.2196/57257
PMID:

©John J Hanna, Abdi D Wakene, Andrew O Johnson, Christoph U Lehmann, Richard J Medford. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Dietary Supplement Information Retrieval: Development of a Retrieval-Augmented Generation System With Large Language Models

Yu Hou[1], PhD; Jeffrey R Bishop[2], PharmD, MS; Hongfang Liu[3], PhD; Rui Zhang[1,4], PhD

[1]Division of Computational Health Sciences, University of Minnesota, Minneapolis, MN, United States

[2]Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN, United States

[3]Department of Health Data Science and Artificial Intelligence, UTHealth, Houston, TX, United States

[4]Center for Learning Health System Sciences, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Rui Zhang, PhD
Division of Computational Health Sciences
University of Minnesota
11-132 Phillips-Wangensteen Building
516 Delaware Street SE
Minneapolis, MN, 55455
United States
Phone: 1 6126261999
Email: ruizhang@umn.edu

Abstract

Background: Dietary supplements (DSs) are widely used to improve health and nutrition, but challenges related to misinformation, safety, and efficacy persist due to less stringent regulations compared with pharmaceuticals. Accurate and reliable DS information is critical for both consumers and health care providers to make informed decisions.

Objective: This study aimed to enhance DS-related question answering by integrating an advanced retrieval-augmented generation (RAG) system with the integrated Dietary Supplement Knowledgebase 2.0 (iDISK2.0), a dietary supplement knowledge base, to improve accuracy and reliability.

Methods: We developed iDISK2.0 by integrating updated data from authoritative sources, including the Natural Medicines Comprehensive Database, the Memorial Sloan Kettering Cancer Center database, Dietary Supplement Label Database, and Licensed Natural Health Products Database, and applied advanced data cleaning and standardization techniques to reduce noise. The RAG system combined the retrieval power of a biomedical knowledge graph with the generative capabilities of large language models (LLMs) to address limitations of stand-alone LLMs, such as hallucination. The system retrieves contextually relevant subgraphs from iDISK2.0 based on user queries, enabling accurate and evidence-based responses through a user-friendly interface. We evaluated the system using true-or-false and multiple-choice questions derived from the Memorial Sloan Kettering Cancer Center database and compared its performance with stand-alone LLMs.

Results: iDISK2.0 integrates 174,317 entities across 7 categories, including 8091 dietary supplement ingredients; 163,806 dietary supplement products; 786 diseases; and 625 drugs, along with 6 types of relationships. The RAG system achieved an accuracy of 99% (990/1000) for true-or-false questions on DS effectiveness and 95% (948/100) for multiple-choice questions on DS-drug interactions, substantially outperforming stand-alone LLMs like GPT-4o (OpenAI), which scored 62% (618/1000) and 52% (517/1000) on these respective tasks. The user interface enabled efficient interaction, supporting free-form text input and providing accurate responses. Integration strategies minimized data noise, ensuring access to up-to-date, DS-related information.

Conclusions: By integrating a robust knowledge graph with RAG and LLM technologies, iDISK2.0 addresses the critical limitations of stand-alone LLMs in DS information retrieval. This study highlights the importance of combining structured data with advanced artificial intelligence methods to improve accuracy and reduce misinformation in health care applications. Future work includes extending the framework to broader biomedical domains and improving evaluation with real-world, open-ended queries.

(J Med Internet Res 2025;27:e67677) doi:10.2196/67677

XSL•FO
RenderX

KEYWORDS

dietary supplements; knowledge representation; knowledge graph; retrieval-augmented generation; large language model; user interface

Introduction

Dietary supplements (DSs) are products intended to add nutritional value to the diet, often containing vitamins, minerals, herbs, amino acids, or other substances [1,2]. In recent years, the role of DSs has become increasingly recognized, with a significant portion of the population acknowledging their importance. According to surveys, 9 out of 10 DS users believe in the necessity of these products, reflecting their widespread acceptance and use [3]. DSs may aid in preventing chronic conditions and offer the potential for reducing health care costs by minimizing the occurrence or severity of preventable diseases [4]. Unlike prescription and over-the-counter drugs, which are subject to stringent regulations, DSs are classified as food products and are regulated by the Food and Drug Administration under more lenient guidelines. Many individuals use DSs spontaneously, without clinician oversight, which introduces unique challenges related to efficacy, safety, and regulation [5]. A significant concern arises from the fact that around 23,000 doctor visits per year are linked to DS-related adverse effects [6]. This underscores the need for both consumers and health care providers to have access to accurate and comprehensive resources for DS-related information.

In response to the increasing demand for reliable DS information, biomedical knowledge graphs (BKGs) have been developed over the past decade to manage large-scale, heterogeneous biomedical data [7-9]. In 2020, the integrated Dietary Supplement Knowledgebase (iDISK) was established to meet the growing need for standardized DS information [10]. iDISK aggregates data from 4 major resources: the Natural Medicines Comprehensive Database (NMCD) [11], the “About Herbs” page on the Memorial Sloan Kettering Cancer Center (MSKCC) website [12], the Dietary Supplement Label Database (DSLD) [13], and the Licensed Natural Health Products Database (LNHPD) [14]. iDISK includes over 4200 ingredient concepts connected to drugs, diseases, symptoms, therapeutic classes, organ systems, and products, providing an extensive repository of DS knowledge.

However, as DS information continually evolves—such as updates to product labels and the discovery of new ingredient-drug interactions—some aspects of the original iDISK have become outdated. For instance, since the launch of DSLD in 2013, the number of recorded labels has increased from 16,712 to 183,012 by January 2024. To maintain accuracy, it is crucial to update iDISK using the latest database versions and optimize the integration process to minimize data noise. Large language models (LLMs) have also gained popularity for answering DS-related inquiries, yet these models are prone to generating “hallucinations” or inaccurate information [15-18]. To counter this, retrieval-augmented generation (RAG) combines the retrieval power of knowledge bases with the generative capabilities of LLMs to produce more accurate responses [19]. An RAG system operates in two phases: first, it retrieves relevant data from a structured knowledge base, ensuring accuracy, and then it uses an LLM to generate responses [20,21]. By integrating iDISK with an RAG system, we mitigate the hallucination problem inherent in standalone LLMs [22,23].

In this paper, we (1) present a major update from iDISK to iDISK2.0, incorporating the latest database versions while minimizing data noise; (2) integrate an RAG system with LLMs to significantly improve the accuracy of DS question answering, the first of its kind for DS queries; and (3) develop the integrated Dietary Supplement Knowledgebase—retrieval-augmented generation (iDISK2.0-RAG) web portal, which offers consumers and health care providers a reliable and user-friendly interface for DS information.

Our work not only addresses the limitations of LLMs but also establishes a new standard for DS question answering. Although this work focuses on DSs, our successful integration and platform can provide new opportunities for other domains.

Methods

Overview

Figure 1 illustrates the overall workflow of this study. We first constructed iDISK2.0 by integrating the latest versions of data from 4 DS resource databases that were originally integrated into iDISK, including LNHPD, DSLD, MSKCC, and NMCD (Figure 1A). We further developed an RAG-based LLM on iDISK2.0 (Figure 1B). In addition, we developed a user-friendly intelligent user interface with iDISK2.0-RAG as the backend to answer DS-related questions (Figure 1C).

XSL•FO
RenderX

Figure 1. An illustration of the study pipeline. DSLD: Dietary Supplement Label Database; NHP: Natural Health Products; iDISK2.0: integrated Dietary Supplement Knowledgebase; LLM: large language model; iDISK2.0-RAG: integrated Dietary Supplement Knowledgebase—retrieval-augmented generation.

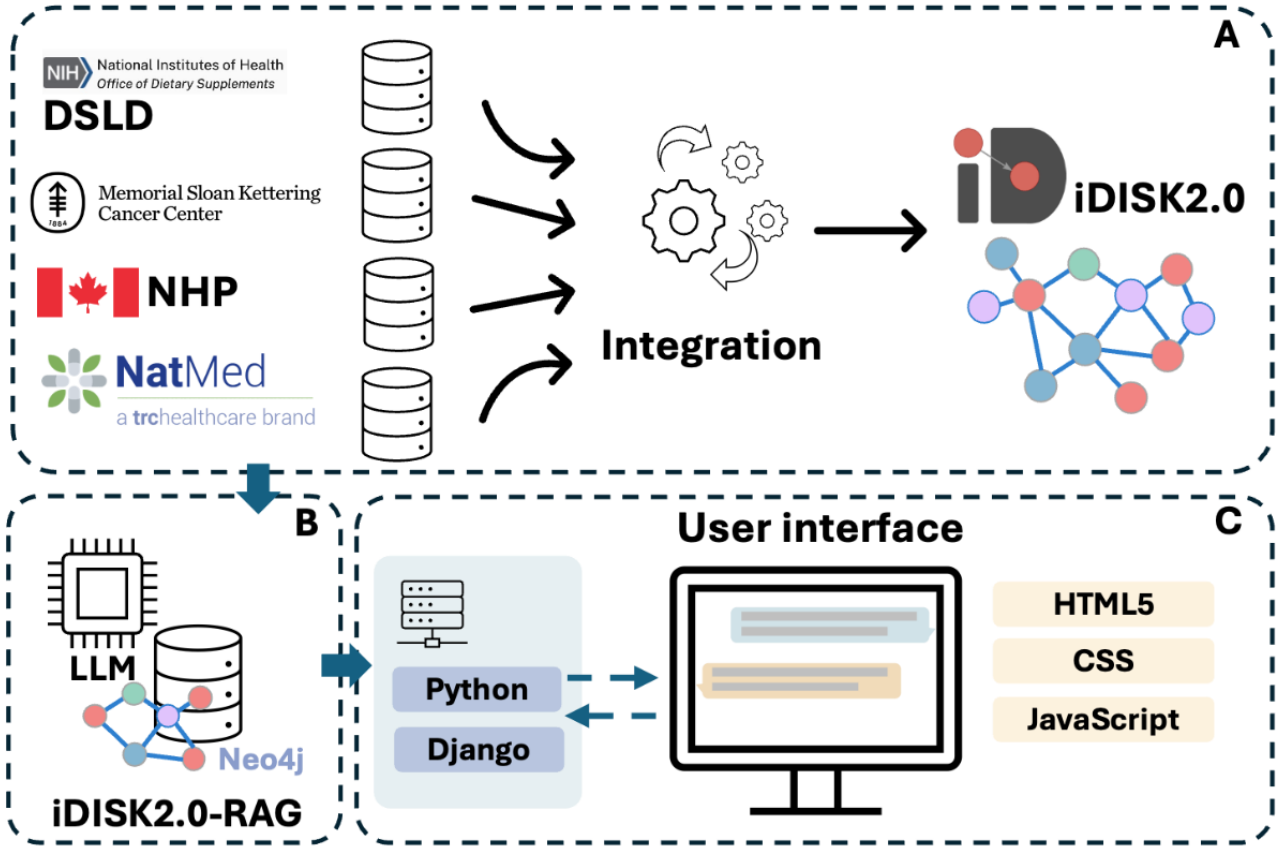

Data Collection and Preprocessing

A total of 4 datasets were collected and preprocessed for the iDISK 2.0. Data from DSLD were downloaded directly from the data release, and the DSLD application programming interface (API) was used to enrich the dietary supplement-related information. Data from MSKCC were obtained after approval, using a web crawler to extract relevant information. LNHPD data were downloaded directly from the data release. After acquiring information from all sources, we converted them into a unified, structured format. The data of the NMCD used in the iDISK 2.0 were derived from information that already exists within iDISK. During the information extraction process, we used LLM (GPT-4) with the prompt (see Multimedia Appendix 1 for specific prompts) to assist in extracting and optimizing information. For example, relationships between DSs and diseases recorded in MSKCC were expressed in text format; we used the LLM to extract disease and drug entities from these sentences and organize them into a structured format for integration and processing. The source data contained inconsistencies and noise, such as variations in how company addresses were recorded in the DSLD product data (eg, “U.S.A.,” “United States of America,” and “United States”) and nonsensical ingredient names like “8” or “%.” We addressed these issues by using an LLM to identify and unify the differing expressions. In addition, we developed filtering patterns specifically designed to remove nonsensical ingredient names, such as those with fewer than two characters or those consisting solely of numbers or punctuation. This approach ensured a more consistent and accurate dataset by standardizing key information and eliminating irrelevant entries.

Specifically, we extracted the dietary supplement product (DSP) name, purpose, safety information, and company address information from LNHPD, along with the dietary supplement ingredient (DSI) name, source material information, and relationships between DSP and DSI. From DSLD, we extracted the DSP name, company address information, DSI name, and the relationship between DSP and DSI. In DSLD, names of DSIs appeared as duplicate records due to synonyms (eg, “Fiber gum acacia,” “Acacia gum extract,” and “Acacia”). We used the DSLD API to collate these synonyms and eliminated duplicates by grouping the synonyms of DSIs representing the same entity using the “Ingredient Group” classification in DSLD. From MSKCC, we extracted the DSI name, common name, background information, mechanism of action, disease name, drug name, symptom name, and the relationships between DSI and disease, DSI and symptom, and DSI and drug.

After completing the information extraction and optimization, we used QuickUMLS (Georgetown IR Lab) [24] to map DSI entities, symptom entities, disease entities, and drug entities from the source resources to the Unified Medical Language System (UMLS) [25] Concept Unique Identifier (CUI). To ensure mapping accuracy, we restricted each entity type based on UMLS semantic types (see Multimedia Appendix 1). The results of the QuickUMLS mapping included the matched UMLS CUI, term name, similarity, and whether the term was

XSL•FO
RenderX

preferred. When QuickUMLS returned multiple results, we used a prioritization strategy: first, results with a similarity of 1, marked as preferred; next, results with a similarity of 1, even if not marked as preferred; and finally, those marked as preferred with the highest similarity. If none of these conditions were met, a UMLS CUI was not assigned to the entity to ensure mapping accuracy.

Data Integration for IDISK2.0

We used a greedy strategy to standardize entity terminology and integrate data from different sources. For each specific entity type, we initially selected a single database to initialize the entity vocabulary. Each specific entity was linked with a unique identifier to integrate entities from all databases, progressively refining the entity vocabulary. Specifically, for DSP, we started with data from LNHPD, integrating DSP entities from both LNHPD and DSLD. We consolidated DSP with the same product name and company name into a single unique entity. We unified entities with the same UMLS CUI or identical term names for DSI. Similarly, we consolidated entities with the same UMLS CUI or identical term names for disease, symptom, and drug entities. Each integrated entity was assigned a unique iDISK ID as its identifier in iDISK. Following this normalization procedure, we obtained a CSV file for each entity type, storing all standardized entity terms and their associated attributes. This allowed us to integrate the extracted knowledge from different databases to build iDISK2.0.

In the knowledge graph (KG), the basic knowledge unit is a triplet, typically defined as "<head entity, relation, tail entity>," indicating a relationship from the head entity to the tail entity in the KG. We mapped the head and tail nodes of the knowledge from the source databases to iDISK IDs, establishing triples composed of iDISK IDs and their relations. We then deduplicated all these triples to reduce noise in iDISK while ensuring quality. Finally, multiple rounds of manual quality checks were conducted.

iDISK2.0 Deployment with Neo4j

We deployed the iDISK2.0 using Neo4j [26], a well-designed graph database platform that allows for structured queries within graphs. Specifically, we stored the generated entities and relationships of iDISK in corresponding CSV files, which Neo4j used as input to create a KG instance automatically. This setup enables iDISK2.0 to be efficiently and flexibly interacted with and updated.

iDISK2.0-Based RAG System

Figure 2 illustrates the overall framework of iDISK2.0-RAG. We began by preprocessing all entities within the iDISK2.0 database, creating embedding vectors for each entity name using OpenAI's embedding model, text-embedding-3-small (Figure 2A). This step allowed us to construct the iDISK2.0 entity vector database, which serves as a foundation for efficient and accurate entity matching. When a user inputs a query, we first extract relevant entities from the question by using an LLM (GPT-4.0; Figure 2B), using tailored prompts (see Multimedia Appendix 1 for specific prompts). The entities identified through this process are then transformed into corresponding vectors using the same OpenAI embeddings model. In the current implementation, relationships between entities are not embedded directly. Instead, we leverage the fixed and well-defined set of 6 relationship types in iDISK2.0 (eg, *is_effective_for* and *interacts_with*) to assign relationships based on the queried entity pairs. For instance, DSI-disease pairs are always associated with the *is_effective_for* relationship. This approach ensures computational efficiency while maintaining accuracy in retrieving relationship-based queries.

Figure 2. The overall design of the iDISK2.0-RAG (integrated Dietary Supplement Knowledgebase—retrieval-augmented generation). LLM: large language model.

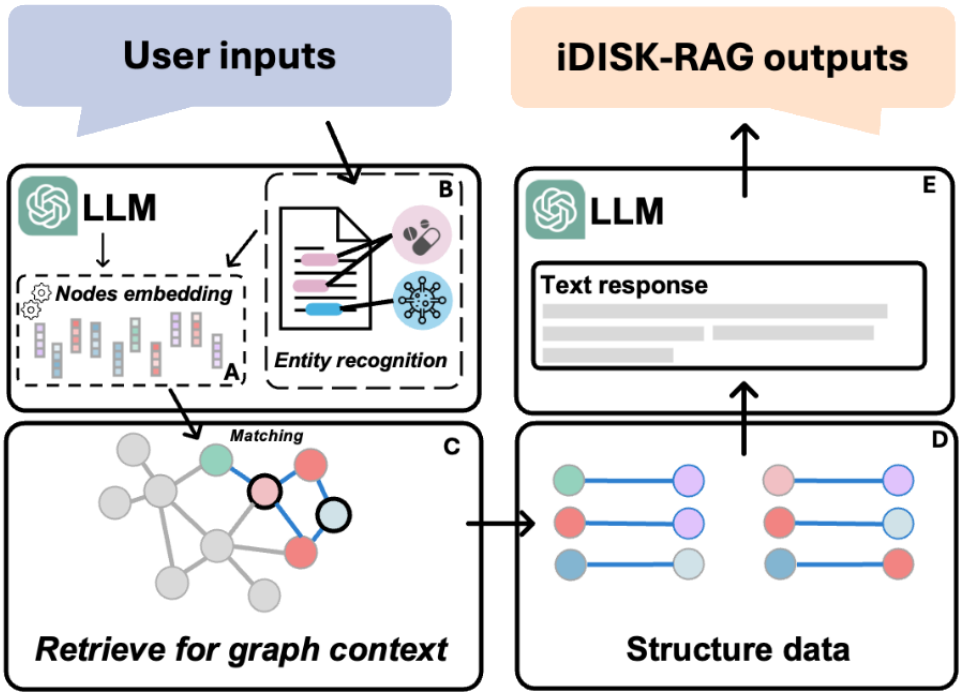

XSL•FO
RenderX

To match these identified entities with those in the iDISK2.0 database, we calculate the cosine similarity between the vectors of the identified entities and the vectors in the preprocessed iDISK2.0 entity vector database. Cosine similarity, a measure of similarity between two vectors in an inner product space, is calculated based on the cosine of the angle between them (detailed information on cosine similarity is provided in Multimedia Appendix 1). We then select the entity with the highest similarity score, provided that the score is 0.75 or higher, to ensure a reliable match between the identified entity and a corresponding entity in iDISK2.0.

Once a match is identified, the corresponding nodes are converted into Cypher queries, allowing for interaction with iDISK2.0, which is deployed on the Neo4j graph database platform. This process retrieves the relevant graph context (Figure 2C), which is returned as structured data in the form of triples (Figure 2D). Finally, an LLM is used to integrate the user's original question with the retrieved structured data (triples) and generate a comprehensive and contextually enhanced response (Figure 2E; see Multimedia Appendix 1 for specific prompts).

Evaluation

We extracted two types of questions from information on the MSKCC website: true-or-false questions and multiple-choice questions (MCQs). Specifically, we extracted information regarding the effectiveness of DSs and diseases and the interactions between DSs and drugs. This information was formatted into true-or-false questions and MCQs (473 true-or-false questions and 329 MCQs). To evaluate the performance of iDISK2.0-RAG and two LLMs (GPT-3.5 and GPT-4.0), we randomly selected 100 questions from each type (true-or-false questions and MCQs). These questions were answered by iDISK2.0-RAG, GPT-3.5, and GPT-4.0, and their accuracy was calculated based on the correct answers. This process was repeated 10 times using bootstrapping, resulting in each system answering 1000 true-or-false questions and 1000 MCQs. Accuracy metrics for each sampling iteration were then computed to derive the performance distribution. To ensure a fair evaluation, all knowledge from MSKCC was excluded from the retrieval stage in the RAG framework, preventing any influence of MSKCC-integrated knowledge on the test results. The rose plot was drawn using ChiPlot [27].

RAG-Enhanced Intelligent Visual Interfaces

We developed a web-based graphical portal that allows users to access knowledge from iDISK2.0 intuitively and flexibly. Specifically, we built the backend (server-side) using Django (Django Software Foundation) [28], an advanced web framework based on Python (Python Software Foundation). The backend also hosts our iDISK2.0-RAG framework to provide users with more accurate retrieval information. The front end (web application side) is constructed using HTML5 and CSS, with JavaScript used for data interaction. Specifically, when the backend receives a query from the front end, the system first uses an LLM, such as GPT-4.0, to extract relevant entities from the user's question through tailored prompts. These extracted entities are then matched with pretrained embedding information from a comprehensive entity vocabulary stored in the backend. Once a match is identified, the RAG system generates corresponding Cypher queries based on the matched entities. These Cypher queries are subsequently used to interact with iDISK2.0, which is deployed on the Neo4j graph database platform, to retrieve the relevant contextual subgraphs. The retrieved subgraphs provide structured data that are essential for generating an accurate response. In the final step, the LLM, guided by specific prompts, is used once again to transform the structured data retrieved by the RAG system into natural language, which is then returned to the front end's dialogue interface. In cases where the entities mentioned in the user's question cannot be successfully matched within iDISK2.0, the LLM (GPT-4.0) relies on its internal knowledge base to generate a corresponding answer, which is also delivered to the front end's dialogue interface. This approach ensures that users receive comprehensive and contextually relevant responses, whether through structured data retrieval or the LLM's generative capabilities.

Ethical Considerations

This study exclusively used publicly available datasets from trusted sources, including NMCD, MSKCC, DSLD, and LNHPD, for the development and evaluation of iDISK2.0. The study did not involve human participants, and no personal, clinical, or proprietary data were accessed or used. Due to these reasons, this study did not require ethics board approval according to the University of British Columbia's Behavioral Research Ethics Board (BREB) guidelines [29]. The data collection and integration processes adhered to the terms and conditions of the respective data providers. For enhanced security, strict protocols were followed during data processing to prevent the inadvertent inclusion of sensitive information. The deployment and use of iDISK2.0 also comply with applicable ethical and legal standards to safeguard the privacy and integrity of the data.

Results

The Integrated Dietary Supplements Knowledge Base 2.0

Through systematic collection and coordination, we ultimately obtained 279 concepts of DSIs, 270 concepts of rugs, 231 concepts of diseases, and 425 concepts of symptoms from the latest version of the MSKCC website. From DSLD, we acquired 4318 concepts of DSIs and 92,651 concepts of DSPs. In addition, from the Natural Health Products database, we gathered 4690 concepts of DSIs and 71,364 concepts of DSPs. After conducting data management and data cleaning, we integrated data from the 3 sources through biomedical entity term normalization and knowledge integration. The iDISK 2.0 encompasses 174,317 entities across 7 types (Table 1), including 8091 DSI entities; 163,806 DSP entities; 786 disease entities; 625 drug entities; 425 symptom entities; 567 dietary therapeutic class (TC) entities; and 17 system organ class (SOC) entities. In addition, there are 6 types of relationships among the 7 entity types (Table 1), including DSP-DSI, DSI-disease, DSI-symptom, DSI-drug, DSI-TC, and DSI-SOC.

XSL•FO
RenderX

Table 1. The concepts, relationships, and attributes in iDISK2.0 (integrated Dietary Supplement Knowledgebase 2.0).

Data elements	iDISK[a] (old version), n	iDISK (new version), n
Concept		
Dietary supplement ingredient	4208	8091
Dietary supplement product	137,568	163,806
Drug	495	625
Disease	776	786
Therapeutic class	605	567
System organ class	17	17
Sign/symptoms	985	425
Total	144,654	174,317
Relationship		
is_effective_for	5363	5245
has_therapeutic_class	5454	4435
has_adverse_effect_on	3168	2598
has_adverse_reaction	2233	1342
has_ingredient	689,826	317,062
interacts_with	3631	3583
Total	709,675	334,265
Attribute		
Company name	—[b]	163,806
Company address	—	163,806
Product purpose	—	65,097
Product risk	—	63,623
Background	1399	1289
Safety	1219	1179
Mechanism of action	258	277
Source material	5532	4277
Interaction rating	3076	3086
Effectiveness rating	4307	4623
Total	15,791	471,063

[a]iDISK: integrated Dietary Supplement Knowledgebase.

[b]Not applicable.

The iDISK 2.0 also includes 471,063 attributes (Table 1), comprising 163,806 DSP company names; 163,806 DSP company addresses; 65,097 DSP purposes; 63,623 DSP risks; 1289 DSI backgrounds; 1179 DSI safety profiles; 277 DSI mechanisms of action; 4277 DSI source materials; 3086 DSI-drug interaction ratings; and 4623 DSI-disease effectiveness ratings.

An iDISK2.0-Based RAG System

To address the potential hallucination issue present in current LLMs and thereby improve the accuracy of user responses, we developed an iDISK2.0-based RAG system. This system integrates the strengths of KG and LLMs to efficiently retrieve relevant knowledge from iDISK and use the language model to generate feedback and facilitate user interaction.

We collected data from the MSKCC website regarding the effectiveness of dietary supplements and diseases and the interactions between DSs and drugs. From these data, we generated a total of 473 true-or-false questions and 329 MCQs. Specifically, there are 164 true-or-false questions, 123 MCQs related to DSs and diseases, and 309 true-or-false questions and 206 MCQs related to DSs and drugs. For instance, in the MSKCC database, there is a record stating, “Vitamin C is used to prevent and treat the common cold.” We used this record to generate corresponding true-or-false questions and MCQs for evaluative purposes. The true-or-false question derived from

XSL•FO
RenderX

this statement is “Is it true that Vitamin C is effective for the common cold?” with the correct answer being “True.” In addition, we developed an MCQ to assess understanding: “Out of the given list, which disease is Vitamin C effective for?” The provided options are “Bladder stones, Common cold, Stroke, Bleeding hemorrhoids, and None of the above.” The correct answer is “Common cold.” Figure 3 illustrates the performance (accuracy) of two LLMs (GPT-3.5 and GPT-4.0) and the iDISK2.0-RAG system on true-or-false questions and MCQs related to DSs and diseases, as well as DSs and drugs. The results show that iDISK2.0-RAG achieves over 95% accuracy across all question types: 99% (990/1000) for true-or-false questions on DSs and diseases, 99% (993/100) for MCQs on DSs and diseases, 97% (974/1000) for true-or-false questions on DSs and drugs, and 95% (948/1000) for MCQs on DSs and drugs. In contrast, the performance of the two LLMs is notably lower, with GPT-4.0 outperforming GPT-3.5: GPT-4.0 achieved 93% (929/1000) accuracy on true-or-false questions related to DSs and diseases, compared with 85% (854/1000) for GPT-3.5; 73% (727/1000) accuracy on MCQs related to DSs and diseases, compared with 43% (426/1000) for GPT-3.5; 62% (618/1000) accuracy on true-or-false questions related to DSs and drugs, compared with 40% (400/1000) for GPT-3.5; and 52% (517/1000) accuracy on MCQs related to DSs and drugs, compared with 46% (457/1000) for GPT-3.5. Despite GPT-4.0’s superior performance over GPT-3.5, both models fall significantly short of the accuracy achieved by iDISK2.0-RAG in these specialized domains.

Figure 3. Model performance in question-and-answer tasks. T/F: true-or-false; MCQ: multiple-choice question; iDISK2.0-RAG: integrated Dietary Supplement Knowledgebase—retrieval-augmented generation.

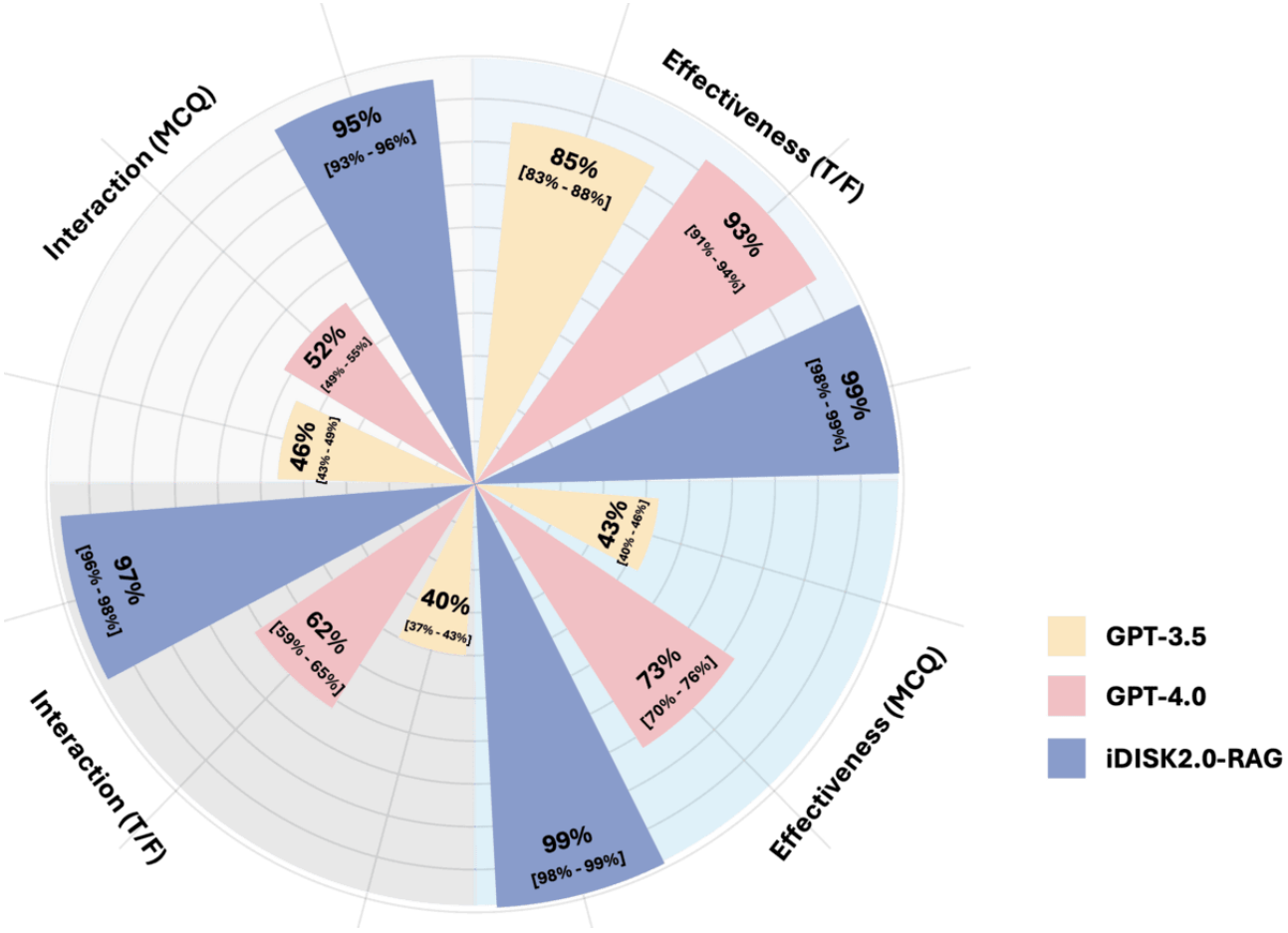

A User-Friendly Intelligent User Interface for Knowledge Retrieval

Knowledge retrieval is one of the most common application scenarios of BKG in biomedical research. In recent years, LLMs have shown outstanding performance in NLP tasks but suffer from the issue of “hallucinations,” especially within specific domains [15-17]. To facilitate the use of iDISK2.0 by biomedical and clinical researchers in DS studies, we developed a web portal based on iDISK2.0-RAG as the backend. This RAG-based user interface allows users to input free-form text queries, generating meaningful biomedical text responses based on the established knowledge within iDISK2.0 to answer users’ questions. Figure 4A illustrates an example of a user query using free-form text. When the user inputs a question such as “Which disease is Omega-3 Fatty Acids effective for?” the user interface retrieves relevant knowledge from the iDISK2.0-RAG backend and returns the corresponding answer: “Omega-3 Fatty Acids are effective for: - Cardiovascular disease - Lupus - Cancer - Depression - High cholesterol - Atherosclerosis.” In cases where the user’s query cannot be answered directly from the iDISK2.0 knowledge base, the RAG framework informs the user that the specific knowledge is not available within iDISK2.0. Subsequently, the LLM within the framework attempts to answer the question based on its own knowledge, as depicted in Figure 4B.

XSL•FO
RenderX

Figure 4. An example of question-and-answer on the user interface.

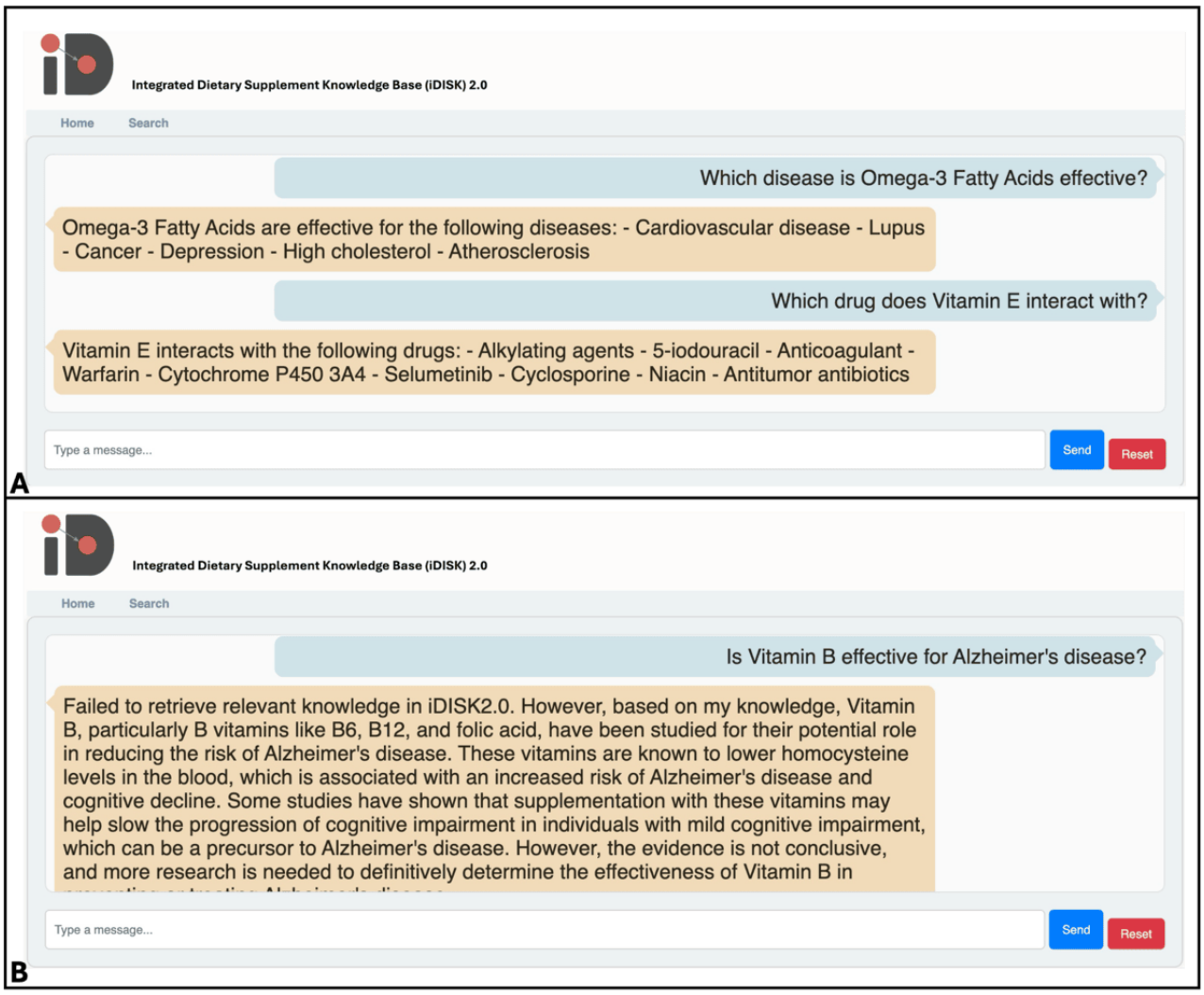

Discussion

Principal Findings

iDISK integrates information from 4 well-designed DS databases, ensuring it contains the most comprehensive DS-related information available. iDISK2.0 inherits this feature, maintaining its status as the most thorough DS information repository. As source resources are updated, much DS-related information has been upgraded. For instance, when the DSLD was first launched in 2013, it contained nearly 17,000 label entries; now, it includes nearly 190,000, an increase of more than tenfold in just a decade. In addition, many DS details have been corrected; for example, the older version of iDISK included an ingredient from DSLD labeled "Araceae," which was inaccurate. In the latest version of DSLD, this ingredient has been removed. The observed reduction in the number of entities, such as in the "signs/symptoms" category, further reflects the enhanced data integration and noise reduction strategies used in iDISK2.0. By consolidating duplicate entries, refining entity definitions, and filtering out ambiguous or irrelevant information, iDISK2.0 significantly improves the overall quality of the knowledge base. Upgrading iDISK to iDISK2.0 is crucial to maintaining the accuracy and relevance of DS information. Without regular updates, the knowledge base could become outdated, leading to potential misinformation and inefficiencies in DS-related research and health care decisions. The continuous integration of the latest data and refined processes in iDISK2.0 helps mitigate these risks, ensuring that users have reliable, current information to make informed decisions about dietary supplements.

LLMs like GPT-4.0 have significantly advanced the field of natural language processing. However, they still exhibit notable limitations, particularly in domain-specific applications such as DS information retrieval. A key challenge is their propensity for generating "hallucinations"—plausible yet incorrect information—stemming from their reliance on extensive but often unverified data sources [18]. This issue is particularly problematic when precision is critical. For example, our evaluation highlighted a case involving the use of coenzyme Q10, also known as ubiquinone, which is documented for its effectiveness in preventing migraines according to the MSKCC database. To test our system's accuracy, we posed an MCQ: "Which disease is Ubiquinone effective for?" with options including migraines, hypertension, menstrual disorders, cholestasis, and none of the above, despite the correct answer being "Migraines," based on robust evidence. However, GPT-4.0

XSL•FO
RenderX

returned “Cholestasis” as the answer, even though only limited studies have explored the relationship between coenzyme Q10 and intrahepatic cholestasis of pregnancy [30,31], with no definitive conclusions indicating its efficacy in this context. In contrast, numerous studies have established the effectiveness of coenzyme Q10 in the prevention and treatment of migraines [32-34]. This error underscores a critical concern: while LLMs have vast knowledge bases, they can still falter by either misinterpreting context or lacking explicit, well-established information.

To address these challenges, we developed iDISK2.0 and integrated it with an RAG system. This integration combines the expansive knowledge base of LLMs with the precise, validated data from iDISK2.0, significantly reducing the risk of hallucinations. In testing, the RAG system, drawing from verified knowledge within iDISK2.0, consistently provided accurate answers, such as correctly identifying coenzyme Q10’s efficacy for migraines. While the current evaluation focused on true-or-false questions and MCQs derived from the MSKCC database to ensure controlled and reproducible testing, we recognize that real-world, open-ended queries often require more nuanced and context-sensitive responses. Our evaluation results indicate that iDISK2.0-RAG achieves over 95% accuracy when retrieving structured knowledge, whereas stand-alone LLMs perform substantially worse, with GPT-4.0 achieving only 62% (618/1000) accuracy in true-or-false questions and 52% (517/1000) in MCQs on DS-drug interactions. This underscores the importance of structured knowledge retrieval in mitigating misinformation. However, instances where the system relies solely on the LLM’s generative capabilities without sufficient retrieval support may still be prone to inaccuracies. To address the limitations of structured retrieval and enhance real-world applicability, we plan to develop a more comprehensive BKG that integrates diverse data sources beyond iDISK2. This expanded BKG will include biomedical literature to summarize the evidence [35], social media data to capture public perceptions, electronic health records to extract evidence of efficacy and safety [36-38], and postmarket surveillance data to address safety concerns [39]. By leveraging this holistic BKG, we aim to enhance the system’s capability to interpret and respond to real-world, open-ended queries. Future evaluations will incorporate open-ended questions collected from diverse user groups and assess both quantitative metrics, such as accuracy, and qualitative metrics, such as user satisfaction. This integrated approach will provide deeper insights into the system’s robustness and applicability. Further enhancing this solution, we incorporated a user-friendly interface with the iDISK2.0-RAG backend, allowing users to access accurate DS information through simple conversational text inputs. This setup not only facilitates better-informed decisions but also improves the overall user experience by ensuring the dissemination of the most accurate and current DS information. The necessity of this innovation is emphasized by the potential risks of relying solely on LLMs, which can propagate misinformation and lead to inefficiencies in both research and health care decisions. Moving forward, future research should focus on refining the interaction between LLMs and RAG systems. A potential avenue could involve mechanisms that allow LLMs to defer to the more reliable RAG output in cases where the evidence is more compelling. This approach would significantly enhance the accuracy and trustworthiness of artificial intelligence–generated responses in biomedical applications.

Despite adopting stricter mapping and integration schemes compared with the older version of iDISK, some errors and noise still exist in the final results. These inaccuracies partly stem from the imprecise expressions in the source databases. In addition, errors arise when attempting to map entities from the source databases to UMLS CUIs for subsequent entity integration. The entity extraction and mapping software QuickUMLS sometimes returns incorrect results. For instance, the MSKCC database records an interaction between “Andrographis” and “Blood pressure-lowering drugs”: “Andrographis interacts with Blood pressure-lowering drugs.” When extracting entities from this knowledge, we extracted the drug entity “Blood pressure-lowering drugs” from the text. However, QuickUMLS mapped “Blood pressure-lowering drugs” to “Drug (UMLS CUI: C0013227),” resulting in incorrect mapping and introduction of errors into the integration process.

Given the current limitations of this study, we plan to implement different strategies in the future to further reduce errors. For instance, we can introduce entity extraction techniques based on LLMs to extract DS-related information stored in text. For example, we may employ various prompt strategies, including rule-based prompts (eg, HealthPrompt [40]), hard-prompt learning [41], and soft-prompt learning [42]. In addition, we can leverage LLMs to batch-check entity extraction and mapping results and perform quality control on the integration outcomes. Recent research has highlighted the beneficial applications of BKGs in the medical field [7,43]. Furthermore, the rise of LLMs has spurred discussions on how best to combine the strengths of these models and BKGs [44]. In the future, besides continuously maintaining iDISK to ensure it remains the most up-to-date and comprehensive DS information resource, we will also explore ways to better assist users in using iDISK. For example, we plan to introduce knowledge graph embedding technology. By using graph-structured reasoning, we aim to offer reasonable predictions for knowledge not yet included in existing databases, thereby helping users discover more information. This approach will not only enhance the precision of user queries but also expand the scope of knowledge available to them. To ensure the long-term reliability and accuracy of iDISK2.0, we plan to establish a pipeline for annual updates integrating new data from trusted sources, such as NMCD, MSKCC, DSLD, and LNHPD. This includes automated data extraction, preprocessing, and stringent quality control measures. In addition, we will continuously adapt the system’s embedding strategy and retrieval mechanisms to align with the evolution of commercial LLMs. The open-source nature of iDISK2.0 will enable contributions from the research community, fostering collaborative improvements. Regular evaluations with updated test sets will further ensure the sustained performance of the RAG system.

The modular design of the iDISK2.0-RAG framework facilitates its adaptability to other areas of health care beyond dietary supplements. BKGs have shown significant promise in facilitating complex biomedical processes such as drug

XSL•FO
RenderX

discovery [45]. By integrating domain-specific knowledge bases, such as those focused on drug interactions or general medical advice, the framework could be extended to address broader biomedical applications. For example, incorporating drug interaction databases or broader biomedical ontologies could enable the system to handle queries related to drug-drug interactions, disease management, and care pathways. Future work will explore these possibilities, evaluating the system's performance across diverse domains and refining its capabilities to support a wider range of health care–related tasks.

Conclusion

In this study, we developed iDISK2.0, an updated repository for DS information, integrating data from NMCD, MSKCC, DSLD, and LNHPD. To address the limitations of LLMs, we introduced an RAG system combining LLMs with a BKG for accurate DS queries. iDISK2.0-RAG outperformed stand-alone LLMs, achieving over 95% accuracy. While some challenges remain with source data and entity mapping, future efforts will focus on improving entity extraction and quality control, ensuring iDISK2.0 continues to support informed DS decision-making. In addition, the adaptable design of the iDISK2.0-RAG framework offers significant potential for expansion into broader health care domains, such as drug interactions and general medical advice.

Acknowledgments

This work was supported by the National Institutes of Health's National Center for Complementary and Integrative Health under grant R01AT009457, the National Institute on Aging under grant R01AG078154, and the National Cancer Institute under grant R01CA287413. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. This study was also supported by Amazon Web Services promotional credits. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of Amazon.

Data Availability

The datasets generated during this study are available in the iDISK2.0 (integrated Dietary Supplement Knowledgebase 2.0) repository on GitHub [46].

Authors' Contributions

YH and RZ conceived the study design and wrote the manuscript. YH implemented the experiments of the study. JB and HL reviewed the manuscript. All authors contributed to the production of the final manuscript.

Conflicts of Interest

HL is an Associate Editor for JAI. None declared by the other authors.

Multimedia Appendix 1
Unified Medical Language System (UMLS) semantic type restrictions, framework prompts, and cosine similarity method for entity matching.
[DOCX File , 16 KB - jmir_v27i1e67677_app1.docx]

References

1. Dietary Supplement Health and Education Act of 1994. National Institutes of Health. 1994. URL: https://ods.od.nih.gov/About/DSHEA_Wording.aspx [accessed 2025-03-05]
2. Dietary supplements. US Food and Drug Administration. 2024. URL: https://www.fda.gov/food/dietary-supplements [accessed 2025-03-05]
3. Nine in ten dietary or nutritional supplement users agree that dietary supplements are essential to maintaining their health. Ipsos. 2023. URL: https://www.ipsos.com/en-us/nine-ten-dietary-or-nutritional-supplement-users-agree-dietary-supplements-are-essential [accessed 2025-03-05]
4. CRN responds to recent JAMA commentary on multivitamin efficacy. CRN. 2024. URL: https://www.crnusa.org/newsroom/crn-responds-recent-jama-commentary-multivitamin-efficacy [accessed 2025-03-05]
5. Dwyer JT, Coates PM. Why Americans need information on dietary supplements. J Nutr 2018;148(suppl_2):1401S-1405S [FREE Full text] [doi: 10.1093/jn/nxy081] [Medline: 31505678]
6. Geller AI, Shehab N, Weidle NJ, Lovegrove MC, Wolpert BJ, Timbo BB, et al. Emergency department visits for adverse events related to dietary supplements. N Engl J Med 2015;373(16):1531-1540 [FREE Full text] [doi: 10.1056/NEJMsa1504267] [Medline: 26465986]
7. Su C, Hou Y, Zhou M, Rajendran S, Maasch JRA, Abedi Z, et al. Biomedical discovery through the integrative biomedical knowledge hub (iBKH). iScience 2023;26(4):106460 [FREE Full text] [doi: 10.1016/j.isci.2023.106460] [Medline: 37020958]

XSL•FO
RenderX

8. Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. Brief Bioinform 2008;9(1):75-90. [doi: 10.1093/bib/bbm059] [Medline: 18077472]
9. Zhu Y, Elemento O, Pathak J, Wang F. Drug knowledge bases and their applications in biomedical informatics research. Brief Bioinform 2019;20(4):1308-1321. [doi: 10.1093/bib/bbx169] [Medline: 29304188]
10. Rizvi R, Vasilakes J, Adam TJ, Melton GB, Bishop JR, Bian J, et al. iDISK: the integrated DIetary supplements knowledge base. J Am Med Inform Assoc 2020;27(4):539-548 [FREE Full text] [doi: 10.1093/jamia/ocz216] [Medline: 32068839]
11. Natural Medicines Comprehensive Database (NMCD). NatMed Pro. 2019. URL: https://naturalmedicines.therapeuticresearch.com/ [accessed 2025-03-05]
12. About herbs, botanicals & other products. Memorial Sloan Kettering Cancer Center. 2021. URL: https://www.mskcc.org/cancer-care/diagnosis-treatment/symptom-management/integrative-medicine/herbs [accessed 2025-03-05]
13. Dietary Supplement Label Database (DSLD). National Institutes of Health. 2020. URL: https://ods.od.nih.gov/Research/Dietary_Supplement_Label_Database.aspx [accessed 2025-03-05]
14. Licensed Natural Health Products Database (LNHPD). 2021. URL: https://tinyurl.com/2xxxpr83 [accessed 2025-03-05]
15. Yao JY, ing KP, Liu ZH, Ning MN, Liu YY, Yuan L. LLM Lies: hallucinations are not bugs, but features as adversarial examples. ArXiv. Preprint posted online on October 2, 2023 2023. [doi: 10.48550/arXiv.2310.01469]
16. Xu Z. Hallucination is inevitable: an innate limitation of large language models. ArXiv. Preprint posted online on January 22, 2024 2024. [doi: 10.48550/arXiv.2401.11817]
17. Rawte V, Sheth A, Das A. A survey of hallucination in large foundation models. ArXiv. Preprint posted online on September 12, 2023 2023. [doi: 10.48550/arXiv.2309.05922]
18. Hou Y, Yeung J, Xu H, Su C, Wang F, Zhang R. From answers to insights: unveiling the strengths and limitations of ChatGPT and biomedical knowledge graphs. medRxiv. Preprint posted online on June 12, 2023 2023 [FREE Full text] [doi: 10.1101/2023.06.09.23291208] [Medline: 37398259]
19. Zakka C, Shad R, Chaurasia A, Dalal AR, Kim JL, Moor M, et al. Almanac - retrieval-augmented language models for clinical medicine. NEJM AI 2024;1(2) [FREE Full text] [doi: 10.1056/aioa2300068] [Medline: 38343631]
20. Li H, Su Y, Cai D, Wang Y, Liu L. A survey on retrieval-augmented text generation. ArXiv. Preprint posted online on February 2, 2022 2022 [FREE Full text] [doi: 10.48550/arXiv.2202.01110]
21. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. SSRN 2024 [FREE Full text] [doi: 10.2139/ssrn.4895062]
22. Sanmartin D. KG-RAG: bridging the gap between knowledge and creativity. ArXiv. Preprint posted online on May 20, 2024 2024 [FREE Full text] [doi: 10.48550/arXiv.2405.12035]
23. Xu Z, Jerome Cruz M, Guevara M, Wang T, Deshpande M, Wang X, et al. Retrieval-augmented generation with knowledge graphs for customer service question answering. : ACM; 2024 Presented at: SIGIR 2024: Proceedings of The 47th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 14 - 18, 2024; Washington, DC p. 2905-2909. [doi: 10.1145/3626772.3661370]
24. Soldaini L, Goharian N. Quickumls: a fast, unsupervised approach for medical concept extraction. MedIR Workshop. 2016. URL: https://ir.cs.georgetown.edu/downloads/quickumls.pdf [accessed 2025-03-12]
25. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
26. Neo4j graph database and analytics. 2013. URL: https://neo4j.com [accessed 2024-03-05]
27. ChiPlot. URL: https://www.chiplot.online/ [accessed 2025-03-05]
28. Django. 2013. URL: https://www.djangoproject.com/ [accessed 2025-03-05]
29. UBC Clinical Research Ethics General Guidance Notes. The University of British Columbia. URL: https://researchethics.ubc.ca/clinical-research-ethics/creb-guidance-notes/ubc-clinical-research-ethics-general-guidance-notes
30. Martinefski MR, Rodriguez MR, Buontempo F, Lucangioli SE, Bianciotti LG, Tripodi VP. Coenzyme Q 10 supplementation: a potential therapeutic option for the treatment of intrahepatic cholestasis of pregnancy. Eur J Pharmacol 2020;882:173270. [doi: 10.1016/j.ejphar.2020.173270] [Medline: 32534074]
31. Martinefski MR, Cocucci SE, Di Carlo MB, Vega HR, Lucangioli SE, Perazzi BE, et al. Fetal coenzyme Q10 deficiency in intrahepatic cholestasis of pregnancy. Clin Res Hepatol Gastroenterol 2020;44(3):368-374. [doi: 10.1016/j.clinre.2019.07.006] [Medline: 31477533]
32. Yaghini O, Hoseini N, Ghazavi MR, Mansouri V, Nasiri J, Moosavian T, et al. A comparative study on the efficacy of coenzyme Q10 and amitriptyline in the prophylactic treatment of migraine headaches in children: a randomized controlled trial. Adv Biomed Res 2022;11:43 [FREE Full text] [doi: 10.4103/abr.abr_235_20] [Medline: 35814300]
33. Sándor PS, Di Clemente L, Coppola G, Saenger U, Fumal A, Magis D, et al. Efficacy of coenzyme Q10 in migraine prophylaxis: a randomized controlled trial. Neurology 2005;64(4):713-715. [doi: 10.1212/01.WNL.0000151975.03598.ED] [Medline: 15728298]
34. Shoeibi A, Olfati N, Soltani Sabi M, Salehi M, Mali S, Akbari Oryani M. Effectiveness of coenzyme Q10 in prophylactic treatment of migraine headache: an open-label, add-on, controlled trial. Acta Neurol Belg 2017;117(1):103-109. [doi: 10.1007/s13760-016-0697-z] [Medline: 27670440]

XSL•FO
RenderX

35. Schutte D, Vasilakes J, Bompelli A, Zhou Y, Fiszman M, Xu H, et al. Discovering novel drug-supplement interactions using SuppKG generated from the biomedical literature. J Biomed Inform 2022;131:104120 [FREE Full text] [doi: 10.1016/j.jbi.2022.104120] [Medline: 35709900]
36. Zhan Z, Zhou S, Li M, Zhang R. RAMIE: retrieval-augmented multi-task information extraction with large language models on dietary supplements. J Am Med Inform Assoc 2025;32(3):545-554. [doi: 10.1093/jamia/ocaf002] [Medline: 39798153]
37. Fan Y, Zhang R. Using natural language processing methods to classify use status of dietary supplements in clinical notes. BMC Med Inform Decis Mak 2018;18(Suppl 2):51 [FREE Full text] [doi: 10.1186/s12911-018-0626-6] [Medline: 30066648]
38. Fan Y, Zhou S, Li Y, Zhang R. Deep learning approaches for extracting adverse events and indications of dietary supplements from clinical text. J Am Med Inform Assoc 2021;28(3):569-577 [FREE Full text] [doi: 10.1093/jamia/ocaa218] [Medline: 33150942]
39. Vasilakes J, Rizvi RF, Zhang J, Adam TJ, Zhang R. Detecting signals of dietary supplement adverse events from the CFSAN adverse event reporting system (CAERS). AMIA Jt Summits Transl Sci Proc 2019;2019:258-266 [FREE Full text] [Medline: 31258978]
40. Sivarajkumar S, Wang Y. HealthPrompt: a zero-shot learning paradigm for clinical natural language processing. AMIA Annu Symp Proc 2022;2022:972-981 [FREE Full text] [Medline: 37128372]
41. Shin T, Razeghi Y, Logan IV RL, Wallace E, Singh S. AutoPrompt: eliciting knowledge from language models with automatically generated prompts. : Association for Computational Linguistics; 2020 Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2025 March 04; Online p. 4222-4235. [doi: 10.18653/v1/2020.emnlp-main.346]
42. Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning. : Association for Computational Linguistics; 2021 Presented at: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing; November 2021; Online and Punta Cana, Dominican Republic p. 3045-3059. [doi: 10.18653/v1/2021.emnlp-main.243]
43. Xiao Y, Hou Y, Zhou H, Diallo G, Fiszman M, Wolfson J, et al. Repurposing non-pharmacological interventions for alzheimer's disease through link prediction on biomedical literature. Sci Rep 2024;14(1):8693. [doi: 10.1038/s41598-024-58604-8] [Medline: 38622164]
44. Yan Y, Hou Y, Xiao Y, Zhang R, Wang Q. KNowNEt: Guided health information seeking from LLMs via knowledge graph integration. IEEE Trans Vis Comput Graph 2024;31(1):547-557. [doi: 10.1109/tvcg.2024.3456364/mm2]
45. Su C, Hou Y, Wang F. NN-based biomedical knowledge graph mining in drug development. In: Graph Neural Networks: Foundations, Frontiers, and Applications. Singapore: Springer; 2022:517-540.
46. iDISK 2.0 - integrated Dietary Supplement Knowledgebase 2.0. GitHub. URL: https://github.com/houyurain/iDISK2.0 [accessed 2025-03-08]

Abbreviations

API: application programming interface
BKG: biomedical knowledge graph
CUI: Concept Unique Identifier
DS: dietary supplement
DSI: dietary supplement ingredient
DSLD: Dietary Supplement Label Database
DSP: dietary supplement product
iDISK: integrated Dietary Supplement Knowledgebase
iDISK2.0: integrated Dietary Supplement Knowledgebase 2.0
iDISK2.0-RAG: integrated Dietary Supplement Knowledgebase—retrieval-augmented generation
KG: knowledge graph
LLM: large language model
LNHPD: Licensed Natural Health Products Database
MCQ: multiple-choice question
MSKCC: Memorial Sloan Kettering Cancer Center
NMCD: Natural Medicines Comprehensive Database
RAG: retrieval-augmented generation
SOC: system organ class
TC: therapeutic class
UMLS: Unified Medical Language System

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.10.24; peer-reviewed by L Wang, C Long; comments to author 02.01.25; revised version received 22.01.25; accepted 25.02.25; published 19.03.25.

Please cite as:
Hou Y, Bishop JR, Liu H, Zhang R
Improving Dietary Supplement Information Retrieval: Development of a Retrieval-Augmented Generation System With Large Language Models
J Med Internet Res 2025;27:e67677
URL: https://www.jmir.org/2025/1/e67677
doi:10.2196/67677
PMID:

©Yu Hou, Jeffrey R Bishop, Hongfang Liu, Rui Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Large Language Model–Based Assessment of Clinical Reasoning Documentation in the Electronic Health Record Across Two Institutions: Development and Validation Study

Verity Schaye[1,2], MD, MHPE; David DiTullio[1], MD, PhD; Benedict Vincent Guzman[3], MS; Scott Vennemeyer[4], BS; Hanniel Shih[4], MS; Ilan Reinstein[2], MS; Danielle E Weber[5,6], MD, MEd; Abbie Goodman[5,6], MD; Danny T Y Wu[4], PhD; Daniel J Sartori[1], MD; Sally A Santen[7], MD, PhD; Larry Gruppen[8], PhD; Yindalon Aphinyanaphongs[3], MD, PhD; Jesse Burk-Rafel[1,2], MD, MRes

[1]Department of Medicine, NYU Grossman School of Medicine, New York, NY, United States

[2]Institute for Innovations in Medical Education, NYU Grossman School of Medicine, New York, NY, United States

[3]Division of Applied AI Technologies, NYU Langone Health, New York, NY, United States

[4]Department of Biostatistics, Health informatics, and Data Sciences, University of Cincinnati College of Medicine, Cincinnati, OH, United States

[5]Division of Hospital Medicine, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, United States

[6]Division of Hospital Medicine, Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, United States

[7]Department of Emergency Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, United States

[8]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, United States

Corresponding Author:
Verity Schaye, MD, MHPE
Institute for Innovations in Medical Education
NYU Grossman School of Medicine
550 First Avenue, MS G 61
New York, NY, 10016
United States
Phone: 1 212 263 3006
Email: verity.schaye@nyulangone.org

Abstract

Background: Clinical reasoning (CR) is an essential skill; yet, physicians often receive limited feedback. Artificial intelligence holds promise to fill this gap.

Objective: We report the development of named entity recognition (NER), logic-based and large language model (LLM)–based assessments of CR documentation in the electronic health record across 2 institutions (New York University Grossman School of Medicine [NYU] and University of Cincinnati College of Medicine [UC]).

Methods: The note corpus consisted of internal medicine resident admission notes (retrospective set: July 2020-December 2021, n=700 NYU and 450 UC notes and prospective validation set: July 2023-December 2023, n=155 NYU and 92 UC notes). Clinicians rated CR documentation quality in each note using a previously validated tool (Revised-IDEA), on 3-point scales across 2 domains: differential diagnosis (D0, D1, and D2) and explanation of reasoning, (EA0, EA1, and EA2). At NYU, the retrospective set was annotated for NER for 5 entities (diagnosis, diagnostic category, prioritization of diagnosis language, data, and linkage terms). Models were developed using different artificial intelligence approaches, including NER, logic-based model: a large word vector model (scispaCy en_core_sci_lg) with model weights adjusted with backpropagation from annotations, developed at NYU with external validation at UC, NYUTron LLM: an NYU internal 110 million parameter LLM pretrained on 7.25 million clinical notes, only validated at NYU, and GatorTron LLM: an open source 345 million parameter LLM pretrained on 82 billion words of clinical text, fined tuned on NYU retrospective sets, then externally validated and further fine-tuned at UC. Model performance was assessed in the prospective sets with F_1-scores for the NER, logic-based model and area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) for the LLMs.

Results: At NYU, the NYUTron LLM performed best: the D0 and D2 models had AUROC/AUPRC 0.87/0.79 and 0.89/0.86, respectively. The D1, EA0, and EA1 models had insufficient performance for implementation (AUROC range 0.57-0.80, AUPRC

XSL•FO
RenderX

range 0.33-0.63). For the D1 classification, the approach pivoted to a stepwise approach taking advantage of the more performant D0 and D2 models. For the EA model, the approach pivoted to a binary EA2 model (ie, EA2 vs not EA2) with excellent performance, AUROC/AUPRC 0.85/ 0.80. At UC, the NER, D-logic–based model was the best performing D model (F_1-scores 0.80, 0.74, and 0.80 for D0, D1, D2, respectively. The GatorTron LLM performed best for EA2 scores AUROC/AUPRC 0.75/ 0.69.

Conclusions: This is the first multi-institutional study to apply LLMs for assessing CR documentation in the electronic health record. Such tools can enhance feedback on CR. Lessons learned by implementing these models at distinct institutions support the generalizability of this approach.

(J Med Internet Res 2025;27:e67967) doi:10.2196/67967

KEYWORDS

large language models; artificial intelligence; clinical reasoning; documentation; assessment; feedback; electronic health record

Introduction

Clinical reasoning (CR) is a fundamental skill that requires incorporating vast amounts of information into a prioritized differential diagnosis and treatment plan and therefore crucial that trainees are given feedback to improve [1]. Documentation in the electronic health record (EHR) can provide this opportunity. Furthermore, poor documentation can reflect lack of refined CR and has been hypothesized to be linked to diagnostic errors [2-4]. There are established human ratings tools to provide feedback on CR documentation such as the Revised-IDEA tool, a rubric that facilitates giving feedback in 4 essential domains of CR, including interpretive summary, differential diagnosis, explanation of lead diagnosis, and explanation of alternative diagnoses [5]. However, feedback provided to trainees can still be limited due to faculty having different standards of what constitutes high quality CR documentation and limited time for feedback in the fast-paced clinical environment [5-7].

Machine learning (ML), natural language processing (NLP), and other artificial intelligence (AI) technologies have emerged as avenues to augment feedback [8-12]. NLP has been used to automate the scoring of documentation in simulated scenarios [13-16]. AI-augmented assessment of CR documentation has also been implemented in clinical environments; we have previously reported on an NLP-based supervised ML model that provides feedback on internal medicine (IM) residents CR documentation [17]. However, this model was developed using earlier technologies and only provides binary feedback. Similarly, Feldman et al [18] published the development of a supervised ML model that provides binary feedback on CR documentation (quality of prioritized differential in progress notes).

The more recent advances of generative AI (GAI) and large language models (LLMs) have expanded the potential of AI as a powerful tool to augment feedback [19,20]. There is a building body of literature on the use of LLMs in CR tasks demonstrating that AI outperforms humans [21-25]. Among these studies is a recent article by Goh et al [21] concluding that LLMs outperformed humans and humans plus LLMs on clinical vignettes as assessed by a standardized rubric of diagnostic performance. In terms of use of LLMs for assessment and feedback on CR, Çiçek et al [26] published on the use of ChatGPT versus expert written feedback on CR questions showing mixed reception from the medical students receiving the feedback. Finally, Jamieson et al [27] used AI to provide feedback on medical student Objective Structured Clinical Examination postencounter notes, an important method of assessing CR, and showed high agreement between AI and expert human ratings. However, these studies are all conducted with curated medical data, at single institutions, and do not focus on assessment of human reasoning but on performance of LLM reasoning as compared with humans.

Navigating the use of LLMs with EHR data (vs curated data) can be much more complicated. There are the challenges of accessing the right data from the chart, a higher burden of accuracy, privacy issues, and variability in EHRs [28-30]. In addition, initial studies have shown LLMs do not perform as well digesting the complexity of information in the EHR to make accurate diagnoses [30]. Finally, while the performance of LLMs on CR tasks is promising, it is far from sufficient to replace humans and it is essential that we continue to provide feedback on our learners' CR [19,30,31]. AI-based tools remain an important strategy to enhance the amount of feedback we provide [12].

Here, we report on an expansion of our previous work and describe the development across 2 institutions of a named entity recognition (NER), logic-based assessment and LLM-based assessments of IM resident CR documentation in the EHR that predicts the quality of CR across 2 domains using the Revised-IDEA tool.

Methods

Setting and Study Population

We conducted this study at 2 institutions, New York University Grossman School of Medicine (NYU) and University of Cincinnati College of Medicine (UC). NYU is a northeastern academic medical center with multiple hospital sites; the NYU IM residency program has 2 resident populations with separate recruitment processes and clinical rotations at different hospitals that use the same EHR, NYU Langone Health Manhattan and NYU Langone Health Brooklyn. NYU residents also rotate at 2 other sites with distinct EHRs not included in this study. In terms of technical resources, NYU has the infrastructure of both the Institute for Innovations in Medical Education which is a multidisciplinary team of clinician educators, data scientists,

XSL•FO
RenderX

informaticians, and developers who apply the science of education and informatics to transform teaching, learning, evaluation, and assessment at NYU and the Division of Applied AI Technologies which focuses on using data and modeling to predict health outcomes across NYU Langone Health [32,33]. In addition, both education and EHR data are stored and easily accessible through a central education data warehouse and there is access to a distributed-memory, high-performing computing cluster [34,35].

UC is a midwestern academic medical center within which IM residents rotate at University of Cincinnati Medical Center (UCMC) and the Veterans Affairs Medical Center; only notes written at UCMC were included in the study. In terms of technical resources, UC has the Department of Biostatistics, Health Informatics, and Data Sciences (BHIDS) that enables the UC academic health care enterprise to make better use of biomedical data and technology for new discoveries, innovative science, and improved health care [36]. Although UC has access to many data sources across the health system, medical school and other education programs, currently there is not a centralized database for education and EHR data like at NYU. In addition, UC has some access to high-performing computing resources, but NYU has a more developed infrastructure for using these tools for both clinical and educational use than at UC.

At each site 2 note sets were retrieved from an integrated EHR (Epic Systems): (1) retrospective dataset comprised of IM resident admission notes from July 2020-December 2021 (n=700 NYU notes, n=450 UC notes) and (2) prospective validation dataset from July 2023-December 2023 (n=155 NYU notes, n=92 UC notes; Figure 1). These time periods were selected to ensure a sufficient range of diagnoses and residents were represented and that COVID-19 admissions were not overrepresented. The datasets at NYU were larger because the initial plan was for primary development and model fine-tuning occurring at this institution.

XSL•FO
RenderX

Figure 1. Overview of the development and validation across 2 institutions of named entity recognition: logic-based assessment and large language model–based assessments of resident clinical reasoning documentation in the electronic health record. AUROC: area under the precision-recall curve; AURPC: area under the precision-recall curve; CV: cross-validation; LLM: large language model; NER: named entity recognition; NYU: New York University Grossman School of Medicine; UC: University of Cincinnati College of Medicine.

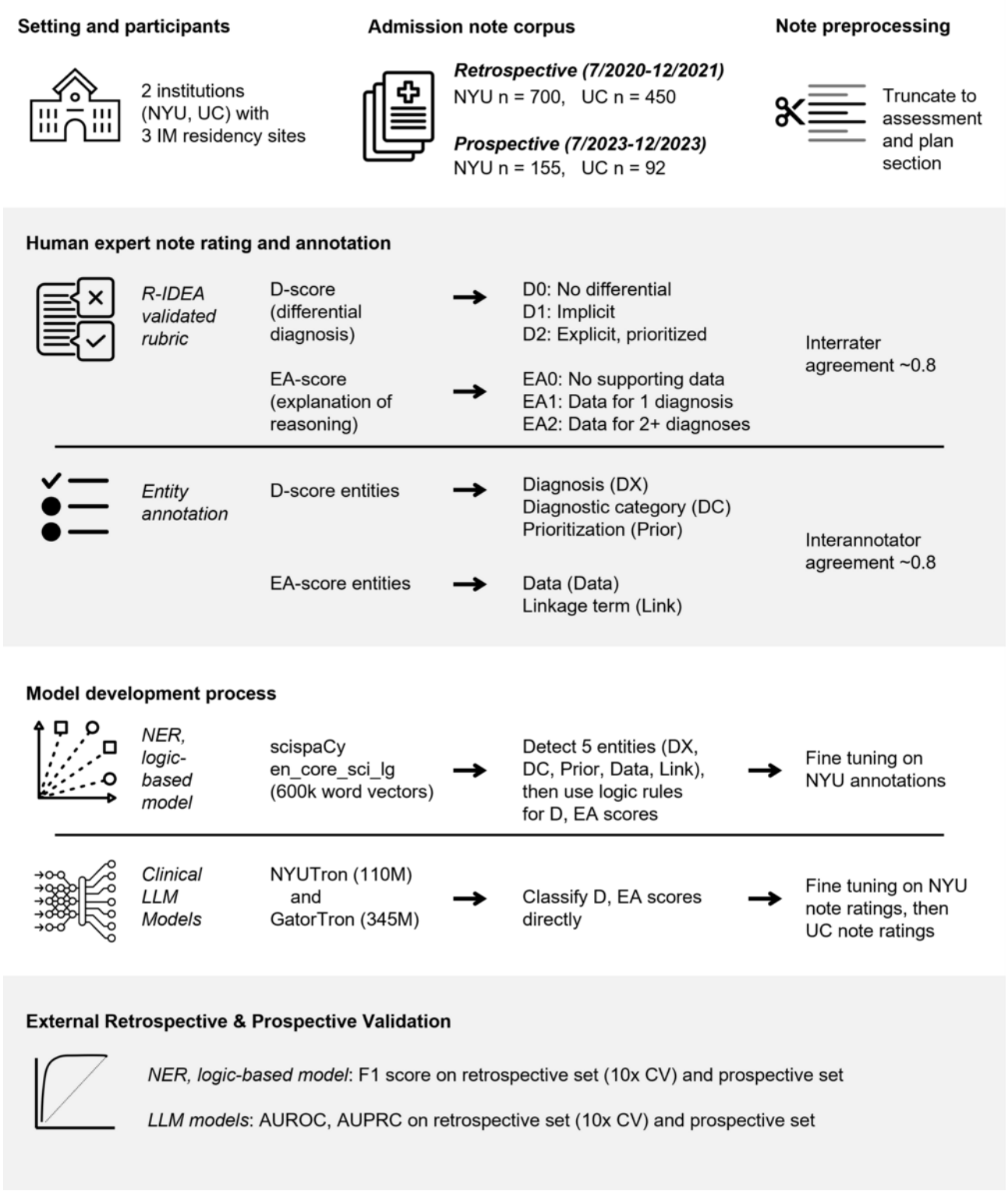

Human Note Rating

We used the DEA components of the Revised-IDEA tool as our human rating gold standard [5]. We maintained the original D (differential diagnosis) score whether a note has an explicitly prioritized differential diagnosis with specific diagnoses (eg, not diagnostic categories such as cardiac), scored as D0, D1, or D2. We discovered early in training experiments that discerning the E (explanation of lead diagnosis) and A (alternative diagnosis explained) would be difficult with available AI models, and iterated to create an overall explanation of reasoning EA score combining the E and A components (ie, EA0, EA1, or EA2; Figure 1; Multimedia Appendix 1). In total, 6 faculty (clinician educators with expertise in CR, assessment, and

XSL•FO
RenderX

psychometrics), 1 resident, and 1 medical student reviewed admission notes to create the new EA score.

On the NYU retrospective dataset, raters also annotated spans of text using Prodigy (an annotation tool for creating training data for ML models). Annotations included 5 entity types for NER, 3 components of the D score (diagnosis [Dx], diagnostic category [DC], prioritization of diagnosis language [Prior]) and 2 components of the EA score (data [Data] and linkage terms [Link]).

To demonstrate interrater reliability, 2 raters labeled 76 notes from the NYU retrospective set for NER, D, and EA scores. For the UC notes, 1 rater from NYU and 2 raters from UC rated 20 notes for D and EA scores. Intraclass correlation (ICC) using a 2-way ANOVA with mixed effects was calculated to assess rater consistency. The remainder of the notes at each institution were rated by 1 rater for D and EA scores. The remainder of the retrospective set at NYU was also rated by 1 rater for NER.

Note Preprocessing

We aimed to isolate the section of the assessment that concentrates on the differential diagnosis and the explanation of reasoning for the primary presenting problem. We iterated on previous truncation strategies outlined previously by Schaye et al [17] and different approaches were required at each institution given different note writing styles (details in Multimedia Appendix 2).

Model Development

Model training occurred with several approaches, NYU development of NER, logic-based model with subsequent external validation at UC, NYU fine tune training of LLM NYUTron [37], NYU fine-tune training of LLM GatorTron [38], UC external validation and fine tune training of NYU fine-tuned GatorTron, and UC fine tune training of GatorTron. We were not able to validate NYUTron at UC pending contract execution for data sharing and will ideally do so in the future.

The selection of these model architectures was driven by their proven effectiveness in handling clinical text data and their ability to capture complex semantic relationships [37-39]. The NER, logic-based model was chosen for its capability to leverage domain-specific knowledge and predefined rules, which is particularly useful for structured data extraction in clinical settings. On the other hand, the LLMs were selected for their advanced contextual understanding and scalability, making them suitable for more nuanced classification tasks.

NER, Logic-Based Model Approach

We used a large, NLP word embedding model trained on scientific texts from the scispaCy library (en_core_sci_lg) with more than 700k vocabulary and 600k word vectors [39]. Word embeddings are a type of word representation that allows words to be represented as vectors in a continuous vector space, capturing semantic meanings. We adjusted model weights with backpropagation (a method used to minimize the error in predictions by adjusting the weights of the model) using the human-annotated labels of the 5 entity types (Dx, DC, Prior, Data, and Link).

We calculated the predicted D scores (ie, D0, D1, and D2) using logic-based relationships between the extracted entities and the rating scale: D0, fewer than 2 unique diagnoses (Dx entity counts); D2, 2 unique diagnoses (Dx entity counts); and explicit prioritization (Prior entity counts); D1, everything else. We attempted ML models to predict EA score from the named entities; however, due to poor model performance, we abandoned further attempts to develop NER, logic-based EA scores.

In order to provide an impartial and dependable evaluation of the model's performance, we used 10-fold cross-validation (a technique where the data is divided into ten parts, and the model is trained and validated 10 times, each time using a different part as the validation set and the remaining parts as the training set). We calculated the Type NER (which demands some overlap between the system tagged entity and the gold standard annotation) evaluation metric, as suggested in SemEval-2013 Task 9, at each k-fold (details in Multimedia Appendix 3) [40]. We report F_1-score over 10-fold runs for each D score entity type (Dx, DC, and Prior) and D score prediction. The F_1-score is crucial for evaluating NER-based models because it balances precision and recall (precision measures the accuracy of the positive predictions, while recall measures the ability to find all relevant instances), providing a comprehensive measure of a model's ability to correctly identify entities while minimizing both false positives and false negatives. This balance is essential for handling the often-imbalanced nature of entity distributions in text, ensuring a more accurate assessment of model performance. We shared the best performing NER, logic-based D model with UC through a docker container and externally validated the model on the UC retrospective set.

LLM Approaches

NYUTron, developed by NYU, is a BERT (Bidirectional Encoder Representation with Transformer)-like LLM with about 110 million parameters that has been pretrained on 7.25 million clinical notes (4.1 billion words, notes through May 2020) [37]. We fine-tuned the model to classify D and EA scores. We applied a 1-versus-rest approach, which resulted in the development and testing of 6 distinct models, each corresponding to a different D and EA score category (ie, D0, D1, D2, EA0, EA1, and EA2 models). However, EA0 and EA1 models did not have adequate performance so we pivoted our approach to create a single binary EA2 model (ie, EA2 vs not EA2). To evaluate model performance, we used 10-fold cross-validation, with area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) averaged over the 10 runs. We chose AUROC and AUPRC as our primary metrics for all LLM-based models because these metrics are favored over *F*-scores for binary classification tasks, particularly with imbalanced datasets. They assess model performance across all possible thresholds, offering a more detailed understanding of trade-offs between true positives, false positives, and precision-recall dynamics, thereby aiding in the identification of the optimal decision-making threshold, which a single *F*-score cannot provide.

XSL•FO
RenderX

Unlike NYUTron, GatorTron is an open source LLM with 345 million parameters that was pretrained on over 82 billion words of deidentified clinical text [38]. To enhance generalizability using an open source LLM, the same experiments described above for NYUTron were taken with GatorTron at NYU.

The NYU fine-tuned GatorTron EA2 model was shared with UC, which conducted external validation, and further fine-tuned following a similar process. Due to the smaller set of notes and hardware limitations, particularly a relatively small Video Random Access Memory size of 16GB, some modifications were applied. A runtime text augmentation was implemented during training with the following settings: 15 words of synonym replacement, random word insertion, and random swap each, and finally, a random word deletion of 15% probability. Finally, we applied random minority oversampling using inverse class frequency during training.[41] In addition, given these limitations, we did not attempt to fine-tune the 3 separate NYU fine-tuned GatorTron D models at UC. Instead, we fine-tuned the original GatorTron model to predict all 3 possible D Scores with a single model at UC using the same training process and hyperparameters as the EA2 model. Further details on model hyperparameters and packages used at NYU and UC can be found in Multimedia Appendix 4.

Prospective Validation

As a final step of validation, we ran each of the best performing models selected for implementation on the site's prospective validation sets and assessed performance using F_1-score for the NER, logic-based model and AUROC and AUPRC for the LLM models.

Ethical Considerations

The study was approved by the NYU and UC institutional review boards, (i19-00280) and (2022-058), respectively. A waiver of consent was obtained for retrospective chart reviews. All retrospective data were anonymized when possible and appropriate measures taken to protect participant information. We followed the reporting guidelines by Klement and El Emam [42] for prognostic and diagnostic machine learning studies.

Results

Human Note Rating

At NYU, ICC was 0.83 (95% CI 0.74-0.89) and 0.77 (95% CI 0.65-0.85) for the D and EA scores, respectively, indicating substantial interrater agreement. Interannotator agreement across all 5 entity types averaged F_1-score=0.81 (range 0.71-0.87 by entity type), indicating strong annotator overlap (Figure 2). At UC, ICC was 0.83 (95% CI 0.68-0.92) and 0.84 (95% CI 0.70-0.93) for the D and EA scores, respectively.

In both datasets, at each institution there was a range of human-rated D and EA scores, diagnoses, and patient demographics mitigating concerns of potential bias in the selection of datasets (Table 1).

Figure 2. Example note (modified to protect patient privacy) with human rating of D and EA scores and annotation for named entity recognition of 5 entity types: 3 components of the D score (diagnosis [Dx], diagnostic category (DC], and prioritization of diagnosis language [Prior]) and 2 components of the EA score (data [Data] and linkage terms [Link]).

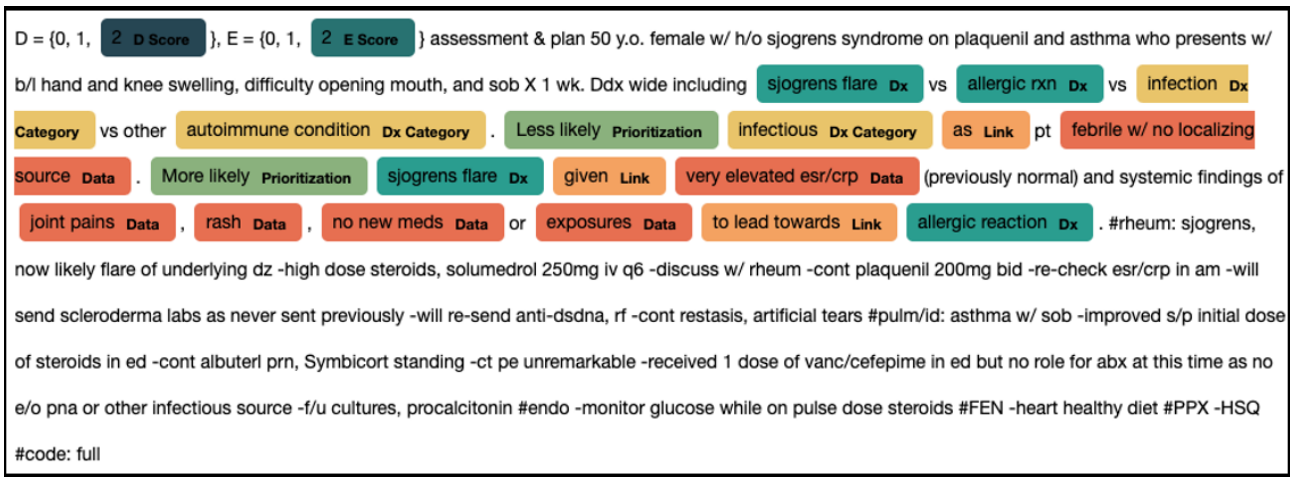

XSL•FO
RenderX

Table 1. Descriptive statistics of human-rated note quality using the Revised-DEA tool and patient characteristics in the retrospective and prospective note sets at NYU[a] and UC[b].

Characteristics	NYU retrospective note set (n=700)	NYU prospective note set (n=155)	UC retrospective note set (n=450)	UC prospective note set (n=92)
D score n (%)				
0	109 (15.6)	55 (35.5)	120 (26.7)	17 (18.5)
1	154 (22)	46 (29.7)	155 (34.4)	31 (33.7)
2	437 (62.4)	54 (34.8)	175 (38.9)	44 (47.8)
E score n (%)				
0	73 (10.4)	19 (12.3)	96 (21.3)	27 (29.3)
1	255 (36.4)	74 (47.7)	171 (38)	21 (22.8)
2	372 (53.1)	62 (40)	183 (40.7)	43 (46.7)
Patient age (years), n (%)				
≤ 54	142 (20.3)	29 (18.7)	147 (32.7)	25 (27.2)
55-68	181 (25.9)	35 (22.6)	184 (40.9)	37 (40.2)
69-80	236 (33.7)	45 (29)	86 (19.1)	20 (21.7)
≥ 81	141 (20.1)	33 (21.3)	33 (7.3)	10 (10.9)
Not answered	0 (0)	13 (8.4)	0 (0)	0 (0)
Patient sex, n (%)				
Female	340 (48.6)	67 (43.2)	215 (47.8)	52 (56.5)
Male	360 (51.4)	75 (48.4)	234 (52)	40 (43.5)
Not answered	0 (0)	13 (8.4)	1 (0.2)	0 (0)
Primary diagnosis by ICD-10[c], n (%)				
Cardiac	134 (19.1)	38 (24.5)	42 (8.8)	12 (12.6)
Dermatologic	13 (1.9)	3 (1.9)	18 (3.8)	4 (4.2)
Endocrine	26 (3.7)	7 (4.5)	27 (5.7)	3 (3.2)
Gastrointestinal	68 (9.7)	15 (9.7)	56 (11.8)	8 (8.4)
Genitourinary	49 (7)	14 (9)	39 (8.2)	10 (10.5)
Hematologic or oncologic	61 (8.7)	7 (4.5)	35 (7.4)	5 (5.3)
Infectious	117 (16.7)	21 (13.5)	10 (2.1)	2 (2.1)
Musculoskeletal	19 (2.7)	3 (1.9)	17 (3.6)	7 (7.4)
Neurologic	21 (3.)	1 (0.6)	12 (2.5)	4 (4.2)
Other	54 (7.7)	11 (7.1)	149 (31.4)	30 (31.6)
Psychiatric	5 (0.7)	7 (4.5)	13 (2.7)	2 (2.1)
Pulmonary	72 (10.3)	10 (6.5)	47 (9.9)	7 (7.4)
Not answered	61 (8.7)	18 (11.6)	10 (2.1)	1 (1.1)

[a]NYU: New York University Grossman School of Medicine.

[b]UC: University of Cincinnati College of Medicine.

[c] *ICD-10: International Statistical Classification of Diseases, Tenth Revision.*

Model Performance

NER Logic-Based Model

In the NYU retrospective dataset, the NER F_1-score for the entity types used to compute the D score (Dx, DC, and Prior) was 0.66. The NER model performed the best in extracting Prior and Dx entities with an F_1-score of 0.75 and 0.68, respectively, but struggled with DC entities, achieving a 0.37 F_1-score.

The NER, D logic-based model performed well at both sites with F_1-scores of 0.83, 0.78, and 0.75 for D0, D1, and D2 scores, respectively at NYU and F_1-scores of 0.75, 0.71, 0.76 for D0, D1, D2 scores, respectively at UC. At UC, the NER,

XSL•FO
RenderX

D-logic–based model was the best performing D model overall selected for implementation and run on the UC prospective validation set with F_1-scores of 0.80, 0.74, and 0.80 for D0, D1, D2 scores, respectively.

LLM-Based Models

At NYU, NYUTron overall had better D and EA model performance on the retrospective set than GatorTron and were the best performing models overall (Table 2). However, while the D0 and D2 NYUTron models performed well, the D1 model was not performant on the retrospective set (AUROC 0.57, 95% CI 0.53-0.69; AUPRC 0.33, 95% CI 0.26-0.43) and therefore was not suitable for implementation. As such, a stepwise approach was taken for the D1 model by taking advantage of the more performant D0 and D2 models (Table 2). The D0 and D2 NYUTron models had excellent performance on the prospective dataset as follows: D0 model, AUROC 0.87 and AUPRC 0.79 and D2 model, AUROC 0.89 and AUPRC 0.86 (Figure 3).

Both the NYUTron EA0 and EA1 models had insufficient performance for implementation therefore the approach pivoted to create a single binary EA model, EA2 vs not EA2 (ie, EA0 or EA1; Table 2). The binary NYUTron EA2 model achieved sufficient performance for implementation with an AUROC 0.85 and AUPRC 0.80 on the prospective dataset (Figure 4).

In external validation at UC, the NER, D logic model performed better than the D GatorTron models and were the D models implemented as described above (Table 2). The GatorTron EA2 model did reach sufficient performance for subsequent prospective validation with an AUROC 0.75 and AUPRC 0.69 (Table 2 and Figure 5).

A final step in optimizing performance was selecting thresholds for all LLM models implemented (details in Multimedia Appendix 5).

XSL•FO
RenderX

Table 2. Large Language Model (LLM) performance on retrospective note sets for all NYUTron and GatorTron experiments classifying differential diagnosis and explanation of reasoning scores in resident admission notes at NYU[a] and UC[b].

LLM[c]	Site	D/EA score classification	AUROC[d] retrospective validation, (95% CI)	AUPRC[e] retrospective validation, (95% CI)
NYUTron	NYU	D0	0.91 (0.85-0.93)	0.72 (0.58-0.76)
NYUTron	NYU	D1	0.57 (0.53-0.69)	0.33 (0.26-0.43)
NYUTron	NYU	D2	0.81 (0.8-0.87)	0.89 (0.85-0.93)
NYUTron	NYU	D1 stepwise approach[f]	—[g]	—
NYUTron	NYU	EA0	0.83 (0.72-0.86)	0.36 (0.23-0.47)
NYUTron	NYU	EA1	0.74 (0.67-0.78)	0.63 (0.54-0.68)
NYUTron	NYU	EA2	0.84 (0.8-0.87)	0.84 (0.81-0.89)
NYUTron	NYU	EA2 binary model[h]	0.84 (0.81-0.85)	0.82 (0.8-0.87)
GatorTron	NYU	D0	0.92 (0.84-0.94)	0.72 (0.48-0.75)
GatorTron	NYU	D1	0.54 (0.5-0.59)	0.31(0.24-0.37)
GatorTron	NYU	D2	0.73 (0.78-0.85)	0.80 (0.82-0.92)
GatorTron	NYU	D1 stepwise approach[f]	—	—
GatorTron	UC	D0	0.75 (0.54-0.96)	0.51 (0.23-0.79)
GatorTron	UC	D1	0.61 (0.44-0.78)	0.46 (0.22-0.7)
GatorTron	UC	D2	0.72 (0.61-0.83)	0.63 (0.46-0.79)
GatorTron	NYU	EA0	0.80 (0.73-0.89)	0.42 (0.25-0.53)
GatorTron	NYU	EA1	0.75 (0.62-0.78)	0.63 (0.48-0.67)
GatorTron	NYU	EA2	0.83 (0.76-0.87)	0.83 (0.79-0.9)
GatorTron	NYU	EA2 binary model[h]	0.81 (0.76-0.87)	0.80 (0.79-0.9)
GatorTron	UC	EA0	—	—
GatorTron	UC	EA1	—	—
GatorTron	UC	EA2	—	—
GatorTron	UC	EA2 binary model[h]	0.72 (0.51-0.93)	0.63 (0.41-0.85)

[a]NYU: New York University Grossman School of Medicine.

[b]UC: University of Cincinnati College of Medicine.

[c]LLM: large language model.

[d]AUROC: area under the receiver operating characteristic curve.

[e]AUPRC: area under the precision-recall curve.

[f]The D1 model ultimately did not have sufficient performance for implementation while the D0 and D2 had excellent performance so a stepwise approach was taken for the D1 score: (1) If D0 model predicts D=0, then D=0, (2) If D2 model predicts D=2, then D=2, and (3) Else D=1. The NYU D1 Stepwise Approach achieved precision of 0.79 while the GatorTron D1 Stepwise Approach achieved precision of 0.73.

[g]Not applicable.

[h]Both the EA0 and EA1 models had insufficient performance for implementation therefore the approach pivoted to create a single EA model EA2 vs not EA2.

XSL•FO
RenderX

Figure 3. Large language model performance on prospective note sets classifying differential diagnosis (D0 and D2) in resident admission notes for best performing D models selected for implementation at New York University Grossman School of Medicine. AUC: area under the curve; FPR: false positive rate; NYU: New York University; PPV: positive predictive value; ROC: receiver operating characteristic; TPR: true positive rate.

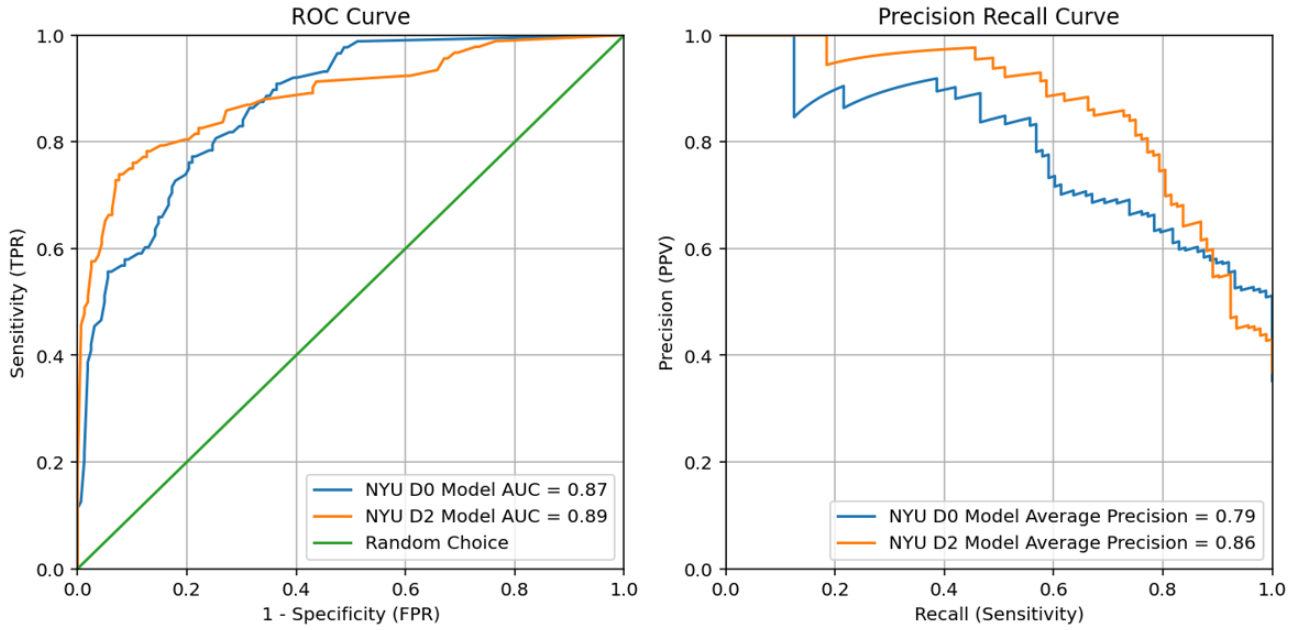

Figure 4. Large language model performance on prospective note sets classifying explanation of reasoning (EA2) in resident admission notes for best performing EA models selected for implementation at New York University Grossman School of Medicine. AUC: area under the curve; FPR: false positive rate; NYU: New York University; PPV: positive predictive value; ROC: receiver operating characteristic; TPR: true positive rate.

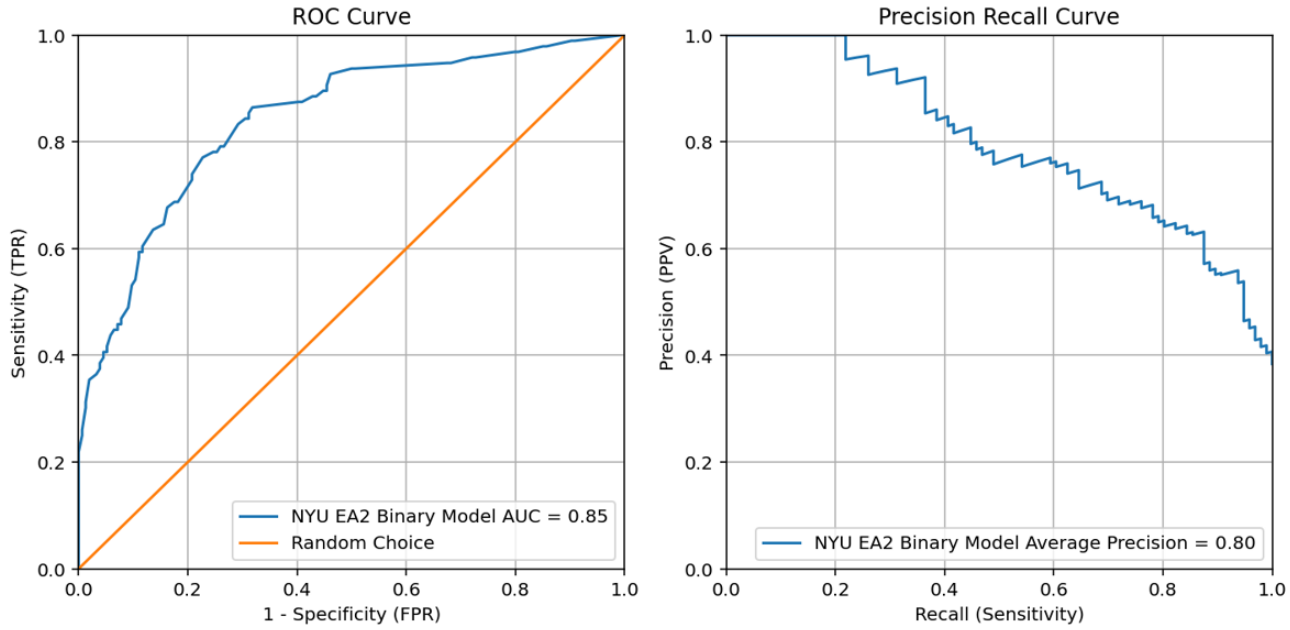

XSL•FO
RenderX

Figure 5. Large language model performance on prospective note sets classifying explanation of reasoning (EA2) in resident admission notes for best performing EA models selected for implementation at University of Cincinnati College of Medicine. AUC: area under the curve; FPR: false positive rate; PPV: positive predictive value; ROC: receiver operating characteristic; TPR: true positive rate; UC: University of Cincinnati.

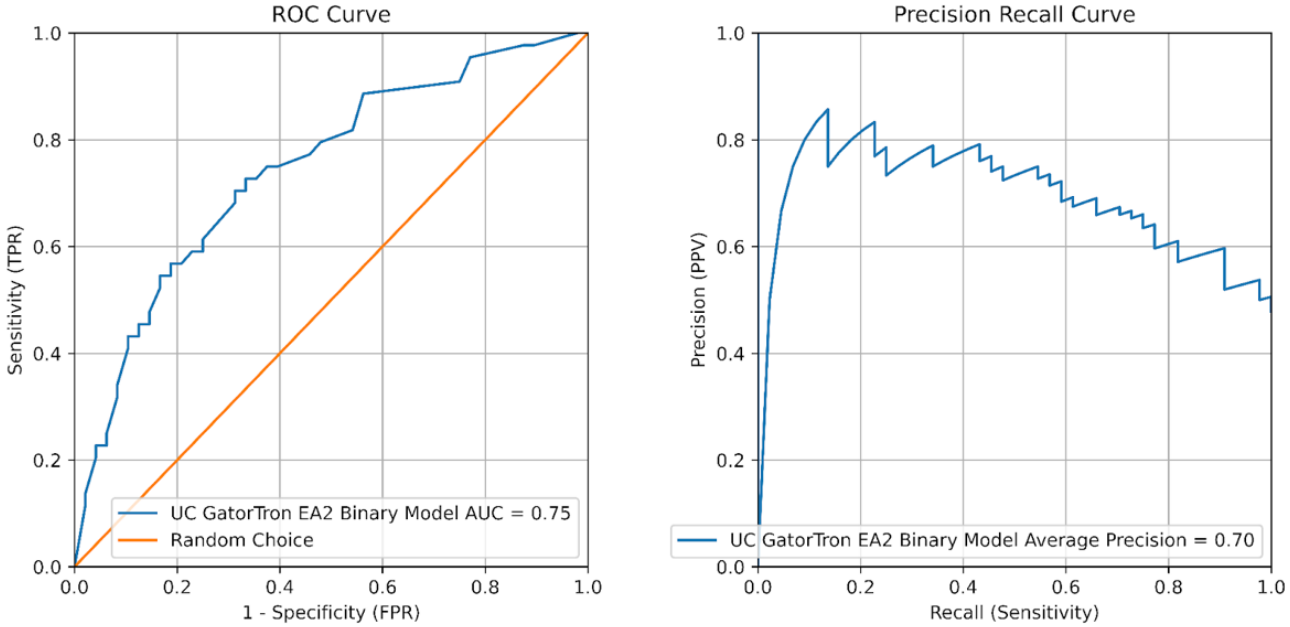

Discussion

Principal Findings

We developed both NER, logic-based and LLM-based assessments of CR documentation in the EHR across 2 institutions with different residency training cultures, expectations for documentation, and technical resources. This builds upon previous work on a supervised ML model of Schaye et al [17] to assess CR documentation, generating more specific feedback across 2 domains of the Revised-IDEA tool. We developed high performing D models that can provide feedback on a 3-point scale and an EA model that can provide feedback on a 2-point scale with comparable performance with our earlier model that provides only binary feedback. Compared with our previous model, which used logistic regression and cTAKES for binary classification of CR documentation quality, our current models leverage advanced NLP techniques, including LLMs, and provides more granular feedback. In addition, the current models were designed at 1 institution and then externally validated at a second institution, enhancing their generalizability and robustness. To our knowledge this is the first study to apply LLMs to human CR in EHR data across institutions (rather than LLM reasoning or human reasoning on curated medical data) [17,18,21-24,26,27,30]. Furthermore, despite the advances in LLMs, these technologies are not yet performant to replace human reasoning. AI-based tools such as the ones we developed can help ensure we are continuing to give our trainees feedback on the essential human skill of CR [12,19,30,31].

While we were able to navigate successfully some of the challenges of working with LLMs and EHR data such as accessing the right data from the chart and privacy issues [28-30], we were not able to achieve sufficient performance of all the models at both sites. The performance differences between NYUTron, GatorTron, and the NER, logic-based model can be attributed to several factors. NYUTron, developed specifically on NYU EHR data, likely benefited from being fine-tuned on a dataset closely aligned with its pretraining data, which may have contributed to its superior performance at NYU. This alignment could have allowed NYUTron to better capture the nuances and specificities of the documentation style and clinical language used at NYU. In contrast, GatorTron, while being a robust open-source LLM with a larger parameter set, may not have been as finely attuned to the specific documentation styles at NYU or UC, leading to relatively lower performance. The NER, logic-based model, leveraging domain-specific knowledge and predefined rules, demonstrated strong performance in structured data extraction tasks, particularly at UC, where it outperformed the GatorTron models for D score predictions. This suggests that for certain structured tasks, smaller, more focused NLP models can sometimes be more effective than larger, more generalized LLMs. In addition, the annotation processes and the quality of the training data, including the consistency and accuracy of human annotations, likely played a significant role in influencing model performance.

It might not always be the newest technology needed to solve the task at hand and comparison of performance of different technologies can be a helpful strategy. Of note when this work initially began, GAI models such as ChatGPT were not readily available in HIPAA (Health Insurance Portability and Accountability Act)-compliant instances at either institution but will be technology we integrate into future work.

In terms of the performance of the models we plan to implement, the significance of the F_1-scores and AUROC metrics is crucial for understanding their practical implications in a clinical environment. Of the NER, logic-based models we plan to implement at UC only the D1 model has an F_1-score less than 0.80. An F_1-score of greater than 0.80 indicates a high level of accuracy in identifying relevant entities, which translates to reliable feedback being generated by the D0 and D2 NER,

XSL•FO
RenderX

logic-based models each with F_1-scores of 0.80. While the D1 NER, logic-based model did not reach this level of performance with an F_1-score of 0.74, an F_1-score greater than 0.70 has been deemed sufficient for applications not involving direct patient care decisions, such as for documentation feedback and educational purposes [43,44]. Furthermore, in the context of educational feedback, an F_1-score of greater than 0.70 ensures a balance between precision and recall, which is crucial for providing comprehensive and reliable feedback to residents without the risk of significant negative consequences [45]. Similarly, all of the LLMs we plan to implement at NYU and UC except 1 have AUROCs greater than 0.80; the GatorTron LLM EA2 model being the exception with an AUROC 0.75. An AUROC of greater than 0.80 suggests that the model has an excellent ability to accurately perform a classification task in this case classify quality of CR documentation. Similar to F_1-scores, an AUROC greater than 0.70 is acceptable in particular for low stakes clinical tasks like formative feedback on CR documentation being provided with these models [46,47].

Finally, we also learned a lot of lessons working across two institutions on how AI technologies can be adapted and successfully implemented at sites with different resources. Some key takeaways include experimenting with different LLMs including ones that are openly available, performing primary development at an institution with more resources and creating a HIPAA-compliant pipeline to share code thus mitigating ethical concerns of patient privacy, developing variations on truncation methods to account for different note writing styles, and creating adaptable approaches to different degrees of computing power such as using text augmentation to prevent overfitting at UC. We will take these lessons learned about generalizability to the next phases of the work and develop strategies to implement more advanced technologies across institutions while maintaining HIPAA compliance working with EHR data.

Our next steps include integrating these AI-based assessments into each residency program as new mechanisms to provide formative feedback to residents. At NYU, we will iterate on dashboards implemented with our previous supervised ML model to improve resident and faculty coaches' ability to visualize longitudinal trends and set improvement goals [17]. At UC, dashboard implementation will represent a new mechanism for formative feedback for residents that currently does not exist. Dashboards at both programs will include percent high-quality (score of 2) notes for D score and E score, trends in scores over time, and goal targets for residents. Residents at both programs will be able to access their data through dashboards for routine review. Each program will create a process for residents to review their performance data and set specific, individualized goals for improving their CR documentation. After initial implementation, we will collect data on outcomes, including the impact on CR documentation practices of IM residents.

Limitations

While the models developed have high performance, they are not perfect. However, the intent is for use in formative and not high-stakes summative assessment which would have a higher threshold for implementation [48]. In addition, similar to our earlier work the ML model excludes the interpretive summary which is a component of the original human rubric Revised-IDEA tool but was still considered too complex to tackle with the AI technologies used in this study [5]. Another limitation is that the LLMs used in this study only give a prediction of D and EA scores without explainability. We will experiment further with newer GAI models in next phases of the work that can help address many of these limitations given GAI's ability to provide narrative explanations and not just classification scores and promising research thus far on the use of GAI in CR assessment with curated data [26,27]. Finally, while we were able to navigate strategies to implement these technologies at 2 sites, there are still some potential limitations that might impact generalizability: (1) not all the technologies we used are readily publicly available and (2) each institution potentially has its own unique writing style which could impact performance of the models and approaches to truncation methods.

Conclusions

This is the first multi-institutional study to apply LLMs for assessing CR documentation in the EHR. Lessons learned from this study can help promote implementation of these technologies across institutions with ranges of technical resources. Further use of LLMs in the EHR for assessment and feedback can be transformative for medical education and patient care.

Acknowledgments

The authors would like to acknowledge Stephen Johnson and Marc Triola for their consultation on methodology. We would also like to thank Lezhi Li for her help with electronic health record (EHR) data extraction at the University of Cincinnati College of Medicine. We would like to recognize the National Board of Medical Examiners Edward J. Stemmler M.D. Medical Education Research Grant Program for their funding support of this project.

Data Availability

This study involves identified clinical information protected under Health Insurance Portability and Accountability Act regulations and is not available for distribution. Please email the authors with any inquiries about the data or models.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The differential diagnosis (D) score from the original Revised-IDEA rubric, a validated tool to assess clinical reasoning documentation, and new explanation of reasoning score (integrating the E score and A score from the original Revised-IDEA rubric).
[DOCX File , 17 KB - jmir_v27i1e67967_app1.docx]

Multimedia Appendix 2
Additional methods on note preprocessing and truncation.
[DOCX File , 16 KB - jmir_v27i1e67967_app2.docx]

Multimedia Appendix 3
Additional methods on type NER (named entity recognition) evaluation metric.
[DOCX File , 16 KB - jmir_v27i1e67967_app3.docx]

Multimedia Appendix 4
Additional methods on model hyperparameters and model packages.
[DOCX File , 14 KB - jmir_v27i1e67967_app4.docx]

Multimedia Appendix 5
Additional results on selecting thresholds.
[DOCX File , 14 KB - jmir_v27i1e67967_app5.docx]

References

1. National Academies of Sciences, Engineering, Medicine. In: Balogh EP, Miller BT, Ball JR, editors. Improving Diagnosis in Health Care. Washington (DC): National Academies Press; 2015.
2. Cadieux DC, Goldszmidt M. It's not just what you know: junior trainees' approach to follow-up and documentation. Med Educ 2017;51(8):812-825 [FREE Full text] [doi: 10.1111/medu.13286] [Medline: 28418205]
3. Singh H, Giardina TD, Meyer AND, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. JAMA Intern Med 2013;173(6):418-425 [FREE Full text] [doi: 10.1001/jamainternmed.2013.2777] [Medline: 23440149]
4. Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? N Engl J Med 2010;362(12):1066-1069. [doi: 10.1056/NEJMp0911734] [Medline: 20335582]
5. Schaye V, Miller L, Kudlowitz D, Chun J, Burk-Rafel J, Cocks P, et al. Development of a clinical reasoning documentation assessment tool for resident and fellow admission notes: a shared mental model for feedback. J Gen Intern Med 2022;37(3):507-512 [FREE Full text] [doi: 10.1007/s11606-021-06805-6] [Medline: 33945113]
6. Baker EA, Ledford CH, Fogg L, Way DP, Park YS. The IDEA assessment tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. Teach Learn Med 2015;27(2):163-173. [doi: 10.1080/10401334.2015.1011654] [Medline: 25893938]
7. Brenner AM, Beresin EV, Coverdale JH, Louie AK, Balon R, Guerrero APS, et al. Time to teach: addressing the pressure on faculty time for education. Acad Psychiatry 2018;42(1):5-10. [doi: 10.1007/s40596-017-0851-9] [Medline: 29134549]
8. Lin SY, Shanafelt TD, Asch SM. Reimagining clinical documentation with artificial intelligence. Mayo Clin Proc 2018;93(5):563-565. [doi: 10.1016/j.mayocp.2018.02.016] [Medline: 29631808]
9. Masters K. Artificial intelligence in medical education. Med Teach 2019;41(9):976-980. [doi: 10.1080/0142159X.2019.1595557] [Medline: 31007106]
10. Dias RD, Gupta A, Yule SJ. Using machine learning to assess physician competence: a systematic review. Acad Med 2019;94(3):427-439. [doi: 10.1097/ACM.0000000000002414] [Medline: 30113364]
11. Cooper A, Rodman A. AI and medical education - a 21st-century pandora's box. N Engl J Med 2023;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
12. Turner L, Hashimoto DA, Vasisht S, Schaye V. Demystifying AI: current state and future role in medical education assessment. Acad Med 2024;99(4S Suppl 1):S42-S47. [doi: 10.1097/ACM.0000000000005598] [Medline: 38166201]
13. Salt J, Harik P, Barone M. Leveraging natural language processing: toward computer-assisted scoring of patient notes in the USMLE step 2 clinical skills exam. Acad Med 2019;94(3):314-316. [doi: 10.1097/ACM.0000000000002558] [Medline: 30540567]

XSL•FO
RenderX

14. Cianciolo AT, LaVoie N, Parker J. Machine scoring of medical students' written clinical reasoning: initial validity evidence. Acad Med 2021;96(7):1026-1035 [FREE Full text] [doi: 10.1097/ACM.0000000000004010] [Medline: 33637657]
15. Jani KH, Jones KA, Jones GW, Amiel J, Barron B, Elhadad N. Machine learning to extract communication and history-taking skills in OSCE transcripts. Med Educ 2020;54(12):1159-1170. [doi: 10.1111/medu.14347] [Medline: 32776345]
16. Sarker A, Klein AZ, Mee J, Harik P, Gonzalez-Hernandez G. An interpretable natural language processing system for written medical examination assessment. J Biomed Inform 2019;98:103268 [FREE Full text] [doi: 10.1016/j.jbi.2019.103268] [Medline: 31421211]
17. Schaye V, Guzman B, Burk-Rafel J, Marin M, Reinstein I, Kudlowitz D, et al. Development and validation of a machine learning model for automated assessment of resident clinical reasoning documentation. J Gen Intern Med 2022;37(9):2230-2238 [FREE Full text] [doi: 10.1007/s11606-022-07526-0] [Medline: 35710676]
18. Feldman J, Hochman KA, Guzman BV, Goodman A, Weisstuch J, Testa P. Scaling note quality assessment across an academic medical center with AI and GPT-4. NEJM Catalyst 2024;5(5). [doi: 10.1056/cat.23.0283]
19. Schaye V, Triola MM. The generative artificial intelligence revolution: How hospitalists can lead the transformation of medical education. J Hosp Med 2024;19(12):1181-1184. [doi: 10.1002/jhm.13360] [Medline: 38591332]
20. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2024;99(1):22-27. [doi: 10.1097/ACM.0000000000005439] [Medline: 37651677]
21. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open 2024;7(10):e2440969 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.40969] [Medline: 39466245]
22. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 2023;330(1):78-80 [FREE Full text] [doi: 10.1001/jama.2023.8288] [Medline: 37318797]
23. Cabral S, Restrepo D, Kanjee Z, Wilson P, Crowe B, Abdulnour RE, et al. Clinical reasoning of a generative artificial intelligence model compared with physicians. JAMA Intern Med 2024;184(5):581-583. [doi: 10.1001/jamainternmed.2024.0295] [Medline: 38557971]
24. Strong E, DiGiammarino A, Weng Y, Kumar A, Hosamani P, Hom J, et al. Chatbot vs medical student performance on free-response clinical reasoning examinations. JAMA Intern Med 2023;183(9):1028-1030. [doi: 10.1001/jamainternmed.2023.2909] [Medline: 37459090]
25. Restrepo D, Rodman A, Abdulnour R. Conversations on reasoning: large language models in diagnosis. J Hosp Med 2024;19(8):731-735. [doi: 10.1002/jhm.13378] [Medline: 38678438]
26. Çiçek FE, Ülker M, Özer M, Kıyak YS. ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. Postgrad Med J 2024;5. [doi: 10.1093/postmj/qgae170] [Medline: 39656920]
27. Jamieson AR, Holcomb MJ, Dalton TO, Campbell KK, Vedovato S, Shakur AH, et al. Rubrics to prompts: assessing medical student post-encounter notes with AI. NEJM AI 2024;1(12). [doi: 10.1056/aics2400631]
28. Goodman KE, Yi PH, Morgan DJ. AI-generated clinical summaries require more than accuracy. JAMA 2024;331(8):637-638. [doi: 10.1001/jama.2024.0555] [Medline: 38285439]
29. Wu J, Liu X, Li M, Li CL, Su DM, Lin S, et al. Clinical text datasets for medical artificial intelligence and large language models-a systematic review. NEJM AI 2025;1(6). [doi: 10.1136/bmj-2023-077292] [Medline: 39870417]
30. Hager P, Jungmann F, Holland R, Bhagat K, Hubrecht I, Knauer M, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med 2024;30(9):2613-2622. [doi: 10.1038/s41591-024-03097-1] [Medline: 38965432]
31. Hswen Y, Abbasi J. AI will-and should-change medical school, says harvard's dean for medical education. JAMA 2023;330(19):1820-1823. [doi: 10.1001/jama.2023.19295] [Medline: 37878288]
32. Institute for innovations in medical education. NYU Langone Health. URL: https://med.nyu.edu/departments-institutes/innovations-medical-education/ [accessed 2025-03-10]
33. Predictive analytics and artificial intelligence. NYU Langone Health. URL: https://med.nyu.edu/centers-programs/healthcare-innovation-delivery-science/predictive-analytics-unit [accessed 2025-03-10]
34. High performance computing core. NYU Langone Health. URL: https://med.nyu.edu/research/scientific-cores-shared-resources/high-performance-computing-core [accessed 2025-03-10]
35. Triola MM, Pusic MV. The education data warehouse: a transformative tool for health education research. J Grad Med Educ 2012;4(1):113-115 [FREE Full text] [doi: 10.4300/JGME-D-11-00312.1] [Medline: 23451320]
36. Overview. University of Cincinnati College of Medicine. URL: https://med.uc.edu/depart/bhd/about-us/overview [accessed 2025-03-10]
37. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, et al. Health system-scale language models are all-purpose prediction engines. Nature 2023;619(7969):357-362 [FREE Full text] [doi: 10.1038/s41586-023-06160-y] [Medline: 37286606]
38. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med 2022;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
39. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. arXiv:1902.07669 2019. [doi: 10.18653/v1/w19-5034]

XSL•FO
RenderX

40. Segura-Bedmar I, Martínez P, Herrero-Zazo M. Semeval-2013 task 9: extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). : Association for Computational Linguistics; 2013 Presented at: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013); 2025 March 04; Atlanta, Georgia, USA p. 341-350. [doi: 10.1016/j.jbi.2014.05.007]
41. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 2018;106:249-259. [doi: 10.1016/j.neunet.2018.07.011] [Medline: 30092410]
42. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. J Med Internet Res 2023;25:e48763 [FREE Full text] [doi: 10.2196/48763] [Medline: 37651179]
43. Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007;14(5):550-563 [FREE Full text] [doi: 10.1197/jamia.M2444] [Medline: 17600094]
44. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform 2015;58 Suppl(Suppl):S11-S19 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.007] [Medline: 26225918]
45. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
46. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5(9):1315-1316 [FREE Full text] [doi: 10.1097/JTO.0b013e3181ec173d] [Medline: 20736804]
47. Baron JM, Huang R, McEvoy D, Dighe AS. Use of machine learning to predict clinical decision support compliance, reduce alert burden, and evaluate duplicate laboratory test ordering alerts. JAMIA Open 2021;4(1):ooab006 [FREE Full text] [doi: 10.1093/jamiaopen/ooab006] [Medline: 33709062]
48. Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. Med Teach 2017;39(10):1010-1015. [doi: 10.1080/0142159X.2017.1359521] [Medline: 28768456]

Abbreviations

AI: artificial intelligence
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
BERT: bidirectional encoder representation with transformer
BHIDS: Department of Biostatistics, Health Informatics, and Data Sciences
CR: clinical reasoning
DC: diagnostic category
Dx: diagnosis
EHR: electronic health record
GAI: generative artificial intelligence
HIPAA: Health Insurance Portability and Accountability Act
ICC: intraclass correlation
IM: internal medicine
Link: linkage terms
LLM: large language model
ML: machine learning
NER: named entity recognition
NLP: natural language processing
NYU: New York University Grossman School of Medicine
Prior: prioritization of diagnosis language
UC: University of Cincinnati College of Medicine
UCMC: University of Cincinnati Medical Center

XSL•FO
RenderX

Edited by A Mavragani; submitted 24.10.24; peer-reviewed by A Rodman, RE Abdulnour, LP Gorrepati, J John Thayil, S Balaguru; comments to author 10.12.24; revised version received 31.01.25; accepted 25.02.25; published 21.03.25.

Please cite as:
Schaye V, DiTullio D, Guzman BV, Vennemeyer S, Shih H, Reinstein I, Weber DE, Goodman A, Wu DTY, Sartori DJ, Santen SA, Gruppen L, Aphinyanaphongs Y, Burk-Rafel J
Large Language Model–Based Assessment of Clinical Reasoning Documentation in the Electronic Health Record Across Two Institutions: Development and Validation Study
J Med Internet Res 2025;27:e67967
URL: https://www.jmir.org/2025/1/e67967
doi:10.2196/67967
PMID:

©Verity Schaye, David DiTullio, Benedict Vincent Guzman, Scott Vennemeyer, Hanniel Shih, Ilan Reinstein, Danielle E Weber, Abbie Goodman, Danny T Y Wu, Daniel J Sartori, Sally A Santen, Larry Gruppen, Yindalon Aphinyanaphongs, Jesse Burk-Rafel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Large Language Model–Driven Knowledge Graph Construction in Sepsis Care Using Multicenter Clinical Databases: Development and Usability Study

Hao Yang[1,2,3*], MSc; Jiaxi Li[4*], MSc; Chi Zhang[1], PhD; Alejandro Pazos Sierra[5], PhD; Bairong Shen[6], PhD

[1]Department of Critical Care Medicine, Joint Laboratory of Artificial Intelligence for Critical Care Medicine, Frontiers Science Center for Disease-related Molecular Network, Institutes for Systems Genetics, Sichuan University, West China Hospital, Chengdu, China

[2]Information Center, Engineering Research Center of Medical Information Technology, Ministry of Education, West China Hospital, Sichuan University, Chengdu, China

[3]Department of Computer Science and Information Technologies, Iberian Society of Telehealth and Telemedicine, University of A Coruña, A Coruña, Spain

[4]Department of Clinical Laboratory Medicine, Jinniu Maternity and Child Health Hospital of Chengdu, Chengdu, China

[5]Department of Computer Science and Information Technologies, Iberian Society of Telehealth and Telemedicine, Research Center for Information and Communications Technologies, Biomedical Research Institute of A Coruña, University of A Coruña, A Coruña, Spain

[6]Department of Critical Care Medicine, Joint Laboratory of Artifcial Intelligence for Critical Care Medicine, Frontiers Science Center for Disease-related Molecular Network, Institutes for Systems Genetics, Sichuan University, West China Hospital, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Bairong Shen, PhD
Department of Critical Care Medicine, Joint Laboratory of Artifcial Intelligence for Critical Care Medicine
Frontiers Science Center for Disease-related Molecular Network, Institutes for Systems Genetics
Sichuan University, West China Hospital
No. 37, Guo Xue Xiang
Chengdu, 610041
China
Phone: 86 85164199
Email: bairong.shen@scu.edu.cn

Abstract

Background: Sepsis is a complex, life-threatening condition characterized by significant heterogeneity and vast amounts of unstructured data, posing substantial challenges for traditional knowledge graph construction methods. The integration of large language models (LLMs) with real-world data offers a promising avenue to address these challenges and enhance the understanding and management of sepsis.

Objective: This study aims to develop a comprehensive sepsis knowledge graph by leveraging the capabilities of LLMs, specifically GPT-4.0, in conjunction with multicenter clinical databases. The goal is to improve the understanding of sepsis and provide actionable insights for clinical decision-making. We also established a multicenter sepsis database (MSD) to support this effort.

Methods: We collected clinical guidelines, public databases, and real-world data from 3 major hospitals in Western China, encompassing 10,544 patients diagnosed with sepsis. Using GPT-4.0, we used advanced prompt engineering techniques for entity recognition and relationship extraction, which facilitated the construction of a nuanced sepsis knowledge graph.

Results: We established a sepsis database with 10,544 patient records, including 8497 from West China Hospital, 690 from Shangjin Hospital, and 357 from Tianfu Hospital. The sepsis knowledge graph comprises of 1894 nodes and 2021 distinct relationships, encompassing nine entity concepts (diseases, symptoms, biomarkers, imaging examinations, etc) and 8 semantic relationships (complications, recommended medications, laboratory tests, etc). GPT-4.0 demonstrated superior performance in entity recognition and relationship extraction, achieving an F_1-score of 76.76 on a sepsis-specific dataset, outperforming other models such as Qwen2 (43.77) and Llama3 (48.39). On the CMeEE dataset, GPT-4.0 achieved an F_1-score of 65.42 using few-shot learning, surpassing traditional models such as BERT-CRF (62.11) and Med-BERT (60.66). Building upon this, we compiled a comprehensive sepsis knowledge graph, comprising of 1894 nodes and 2021 distinct relationships.

XSL•FO
RenderX

Conclusions: This study represents a pioneering effort in using LLMs, particularly GPT-4.0, to construct a comprehensive sepsis knowledge graph. The innovative application of prompt engineering, combined with the integration of multicenter real-world data, has significantly enhanced the efficiency and accuracy of knowledge graph construction. The resulting knowledge graph provides a robust framework for understanding sepsis, supporting clinical decision-making, and facilitating further research. The success of this approach underscores the potential of LLMs in medical research and sets a new benchmark for future studies in sepsis and other complex medical conditions.

(J Med Internet Res 2025;27:e65537) doi:10.2196/65537

KEYWORDS

sepsis; knowledge graph; large language models; prompt engineering; real-world; GPT-4.0

Introduction

Sepsis, a critical condition leading to septic shock and multiple organ dysfunction syndrome in patients, often arises from severe trauma, surgery, or infections [1]. Despite advancements in diagnostics, therapeutics, and patient monitoring, the incidence and mortality rates of sepsis remain alarmingly high, posing a global medical challenge. Annually, it affects over 49 million individuals worldwide, with approximately 11 million fatalities, and mortality rates fluctuate between 15% and 25% [2]. This highlights the necessity for comprehensive research into sepsis to enhance prevention and treatment methods. Recent studies have increasingly concentrated on understanding its pathogenesis [3], clinical presentations [4], and treatment approaches [5]. However, the complexity and variability of sepsis complicate the development of effective treatments and rehabilitation strategies. Furthermore, survivors often endure long-term effects like cognitive dysfunction [6], indicating a gap in our understanding of postsepsis syndrome.

Constructing a powerful database specifically for sepsis is paramount. This initiative will significantly strengthen support for clinical data mining and big data applications, enabling the aggregation and analysis of vast datasets. Such progress could lay the groundwork for pioneering research. Furthermore, developing a comprehensive knowledge graph [7] on sepsis is crucial, as it will serve as a bridge between the deep application of databases and a more profound understanding. By closely integrating the detailed data from the database with the associative analysis capabilities of the knowledge graph, we can more effectively unveil the complex mechanisms of sepsis, propelling deep research advancements in this field. In addition to the database, by integrating findings from related studies, clinical presentations, and treatment methods on sepsis, this knowledge graph could offer multidimensional and multilevel information. It aims to delve into the pathogenesis, clinical manifestations, and conventional treatment options for sepsis, providing a deeper insight into its complexities. This will serve as a scientific basis for formulating more effective preventive and treatment strategies, as well as providing clinicians with real-time and comprehensive reference materials to accurately assess patients' conditions and optimize treatment plans.

However, the relationship extraction step [8] in knowledge graph construction is a critical juncture, determining the accuracy and completeness of the relationships between entities. Currently, deep learning models based on natural language processing technology are commonly used for clinical text entity recognition and relationship extraction. This process relies on professional data labeling, neural network model design, and large-sample model training, making it a time-consuming and labor-intensive task. Recently, large language models (LLMs) such as GPT-4.0 [9] and LLaMA [10] have revolutionized the field of natural language processing, offering new insights to address this issue. These models possess the ability to be widely applied to various downstream tasks, such as named entity recognition (NER) [11] and relationship extraction [12]. Meanwhile, through the design and optimization of prompt engineering [13], large models' capability to handle complex task scenarios is enhanced and widely applied in the medical field [14]. Floridi and Chiriatti [15] evaluated multiple pretrained and fine-tuned large language models (LLMs) on their ability to extract adverse events from the Vaccine Adverse Event Reporting System (VAERS) notes. The fine-tuned adverse events LLM achieved an impressive average micro F_1-score of 0.704, demonstrating the potential of LLMs in natural language processing tasks. Hu et al [11] assessed ChatGPT's ability to perform zero-shot clinical NER as defined by the 2010 i2b2 challenge and compared it to models based on GPT-3 [15] and Bio-Clinical BERT (Bidirectional Encoder Representations from Transformers) [16]. The results revealed that while ChatGPT's zero-shot performance was not as strong as fine-tuned BERT clinical models, its F_1-score reached 0.628 under relaxed matching criteria, indicating reasonable performance. However, these studies also highlighted some limitations, such as errors from hallucinations (ie, generating inaccurate, meaningless, or contextually irrelevant outputs) [17] and insensitivity to negation words [18].

In this research, we aim to establish a multicenter sepsis database (MSD), providing a richer and more diverse dataset that is crucial for the in-depth analysis and understanding of sepsis. Using GPT4.0 for entity recognition and relation extraction, we aim to construct a comprehensive sepsis knowledge graph, using real-world databases supplemented by clinical guidelines and relevant public databases. This detailed knowledge graph is intended to be an exhaustive reference for clinicians, enhancing their comprehension and treatment approaches for sepsis. By the application of these advanced technologies, we strive to improve the construction quality and efficiency of the knowledge graph, providing strong support for sepsis research and clinical management.

XSL•FO
RenderX

Methods

Data Collection and Database Construction

The dataset originates from prominent tertiary hospitals in western China, including the West China Hospital of Sichuan University, Shangjin Hospital, and Tianfu Hospital. It consists of records from hospitalized patients diagnosed with sepsis between 2020 to 2023. The inclusion criteria for the dataset were patients whose discharge diagnoses prominently featured sepsis within the specified timeframe (2020-2023). For these patients, comprehensive data were collected, including baseline demographics, laboratory test results, and electronic medical records documenting admission and discharge details, among other pertinent information, as shown in Figure 1. The structure of the database is shown as an Entity Relationship Diagram (Figure S1, Multimedia Appendix 1). Before any EHR data is processed by the GPT-4 model, all personal identifiers, such as names, addresses, and patient IDs, are removed following standard deidentification protocols. This ensures that no identifiable information is included during the relationship extraction and entity recognition process.

Figure 1. Overview of the sepsis dataset from West China Hospital of Sichuan University, Shangjin Hospital, and Tianfu Hospital (2020-2023).

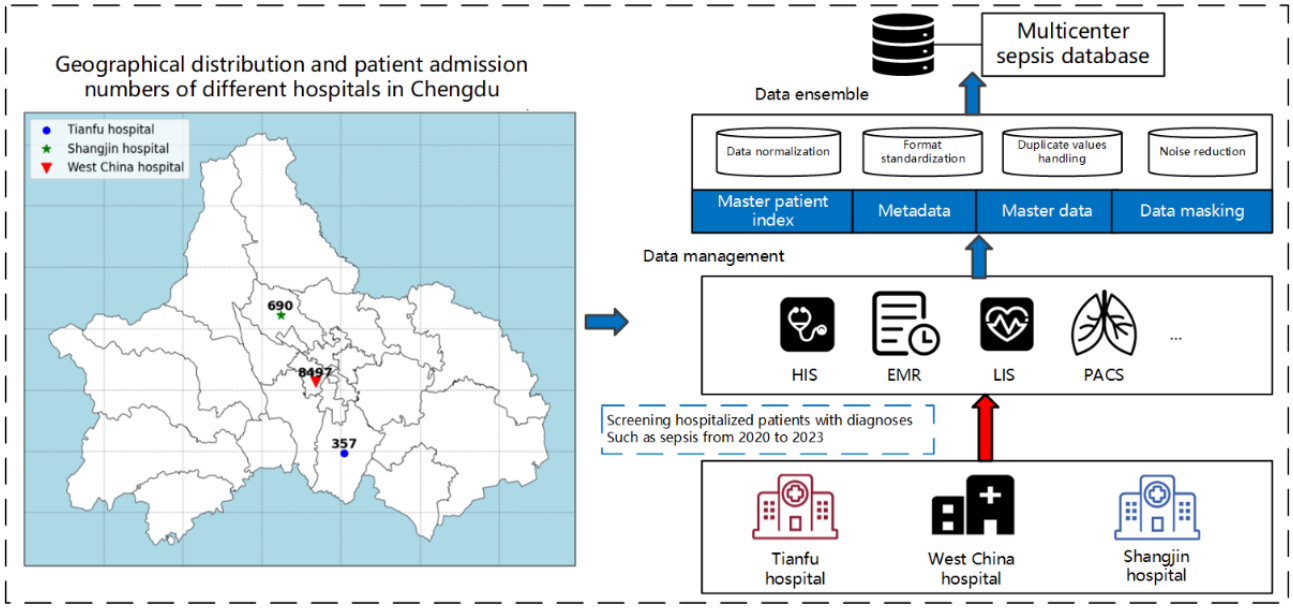

Sepsis Expert Consensus and Guidelines

Incorporating data from expert consensus statements and clinical guidelines on sepsis establishes a foundational understanding of best practices and current recommendations in the field. These authoritative sources significantly contribute to the development of a comprehensive knowledge base. The primary data source is PubMed, and our search query is as follows:

(sepsis[ti] and guideline[pt] and 2020/01/01:2023/12/01[pdat]) or (sepsis[ti] and review[pt] and 2020/01/01:2023/12/01[pdat])

This search is designed to retrieve all expert guidelines and literature reviews on sepsis published in the last 3 years, serving as a vital knowledge repository for subsequent relationship extraction. To automate the process of retrieving articles from PubMed, we used the Biopython library [19] which provides an efficient way to interact with the PubMed database. In addition, we integrated biomarker data from MetaSepsisBase [16], an open database for sepsis and the first sepsis biomarker knowledge database ever constructed. The database comprises biomedical information on 320 sepsis biomarkers, with 450 records sourced from PubMed and annotated through the NCBI Gene database [17] and HGNC database [18]. This initiative provides a solid knowledge foundation for the construction of a sepsis knowledge graph.

The Construction of Knowledge Graph

The construction process of the sepsis knowledge graph primarily uses an entity-based approach, involving key steps such as data preprocessing, analysis, entity construction [20], knowledge extraction [21], fusion [22], and storage. In contrast to traditional knowledge graph construction methods, this study leverages GPT-4.0 to extract knowledge and conduct initial data mining on raw data, enhancing the accuracy of the resulting knowledge graph.

Building the Entity of Knowledge Graph

The primary objective of entity construction [23] is to acquire, describe, and represent knowledge within a specific domain, fostering a shared understanding of the sepsis field. This involves identifying widely accepted terms, providing precise definitions, and establishing their interrelationships across different levels of formal patterns. In our research, the entity definitions and their interrelationships are presented in Tables 1 and 2.

XSL•FO
RenderX

Table 1. Definitions of entity concepts in the sepsis knowledge graph based on multicenter clinical data (N=10,544) from 3 hospitals in Western China (2020-2023).

Number	Concept terms	Attributes	Meaning
1	Diseases	Canonical name, concept ID, TUIs[a], definition, reference, and synonyms	Disease is an atypical life process that occurs due to disrupted self-regulation in the body under specific conditions after exposure to pathogenic damage.
2	Symptoms	Canonical name, concept ID, TUIs, duration, severity, frequency, and temporal pattern	Subjective atypical sensations or objective pathological changes in a patient resulting from a series of physiological, metabolic, and morphological abnormalities during the disease process.
3	Imaging examination	Canonical name, concept ID, TUIs, type, body region, and imaging findings	Imaging examination is a technique used to visualize the interior of a body for clinical analysis and medical intervention.
4	Biomarkers	Canonical name, concept ID, TUIs, biomarker type, reference, synonyms, cut-off value, sensitivity, and specificity	Measurable substances in the body used to indicate physiological status, the presence of disease, or disease progression.
5	Laboratory test	Canonical name, concept ID, TUIs, normal range, unit, and measured value	Physical or chemical examinations conducted in a laboratory to determine the content, nature, concentration, quantity, and other characteristics of submitted substances.
6	Subtypes	Canonical name, concept ID, TUIs, definition, reference, and synonyms	Subdivisions within a category or group based on specific features or characteristics. For example, disease subtypes may indicate different pathological features or symptom presentations.
7	Pathogenic mechanism	Canonical name, concept ID, UIs[b], type, infection source, reference, biomarkers, and guidelines	Biological or biochemical processes leading to the occurrence of disease, including factors such as pathogen infection and genetic mutations.
8	Pharmacotherapy	Canonical name, concept ID, TUIs, medication name, drug class, dosage, duration of treatment, adverse effects, drug interactions, and formulations	Substances used for preventing, treating, and diagnosing diseases. In theory, drugs encompass chemical substances that can influence the physiological functions and cellular metabolic activities of the body's organs.
9	Surgery	Canonical name, concept ID, TUIs, anesthesia type, guidelines, complications, recovery time, and level	A procedure involving the use of instruments, performed by a surgeon or other specialized personnel, to enter the human body or other biological tissues. It involves the application of external force to eliminate pathology, alter structures, or implant foreign materials.

[a]TUI: Type Unique Identifier, representing the semantic category of a concept in UMLS..

[b]UI: Unique Identifier, used to uniquely distinguish a concept or term in UMLS..

Table 2. Definitions of semantic relations in sepsis data based on multicenter clinical data (N=10,544) from 3 hospitals in Western China (2020-2023).

Number	Concept terms	Meaning
1	Complications	Disease development may give rise to another disease or the occurrence of additional symptoms.
2	Has symptom	Describing the correlation between diseases and symptoms.
3	Recommended imaging examination	Recommended imaging examination for patients with sepsis.
4	The related biomarkers	Biomarkers associated with sepsis.
5	Treat	Surgical treatment modalities associated with sepsis.
6	Recommended laboratory tests	Routine laboratory tests associated with sepsis.
7	Caused by	The underlying factors contributing to the onset of sepsis.
8	Recommended medication	Pharmacological interventions in the treatment process of sepsis.

The construction methods of knowledge graph technology are typically categorized into two approaches: top-down and bottom-up [24]. The top-down approach involves defining a pattern layer based on logical relationships and hierarchical structures, followed by mapping data entities to this schema (as shown in Figure 2A). In contrast, the bottom-up approach focuses on extracting entities and attributes from diverse data sources for the data layer of the knowledge graph (as shown in Figure 2B). This process consolidates the extracted entities and attributes, optimizing the schema layer of the knowledge graph for iterative updates to the entity model. The top-down approach ensures that domain entities are enriched with professional

XSL•FO
RenderX

knowledge and accuracy, while the bottom-up approach enhances their practicality. Given the unique characteristics of sepsis, our research adopts both approaches to construct the sepsis knowledge graph. This dual approach aims to improve the educational value, accuracy, and practical utility of the knowledge graph.

Figure 2. Construction of the sepsis knowledge graph using top-down and bottom-up approaches. (A) Pattern layer: in the top-down approach, a structured pattern layer is defined based on logical relationships and hierarchical structures. (B) Data Layer: the bottom-up approach extracts entities and attributes from diverse data sources. CRP: C-reactive protein; IL: interleukin; PCT: Procalcitonin; PRO-BNP: Pro-B-Type Natriuretic Peptide (Pro-Brain Natriuretic Peptide).

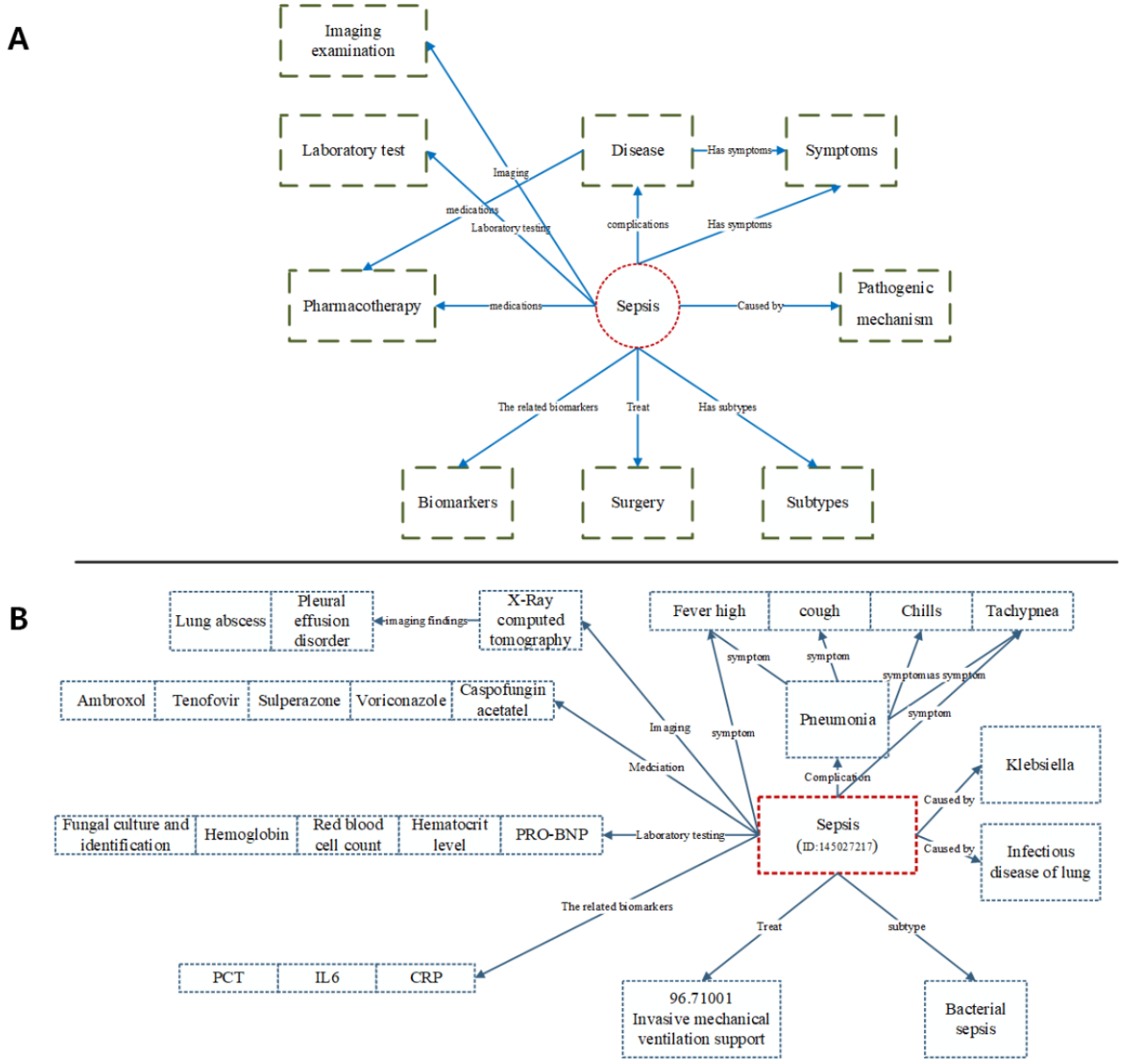

Prompt Engineering and Model Evaluation

In this study, we used the approach of LLMs prompt engineering for knowledge extraction. By harnessing the powerful language understanding and pattern recognition capabilities of GPT-4.0, we guided the model to generate sepsis-specific prompts, extracting relevant information. The advantage of this approach lies in its ability to handle complex structures and contexts in natural language without manually specifying intricate rules or patterns.

Compared with traditional knowledge extraction methods, LLM-based approaches offer several advantages. First, LLMs can autonomously learn the grammar and semantic rules of language, making them more adaptable to the intricate expressions within this domain [24]. Second, due to pretraining on large-scale corpora, LLMs acquire rich background knowledge, enhancing their contextual awareness when understanding texts in a specific domain [25]. In addition, LLMs excel at handling ambiguous and unclear language expressions, improving the robustness of knowledge extraction [26].

To enhance the effectiveness of our approach, we introduced zero-shot [27] and few-shot [28] learning techniques, as shown in Textbox 1. By leveraging few-shot learning within LLMs, we enable the model to quickly adapt to new sepsis-related texts even with limited examples, allowing the model to flexibly adjust to the specific context and linguistic expressions of a

XSL•FO
RenderX

specialized domain. This adds flexibility and personalization to the knowledge extraction process, further improving the model's performance and adaptability. To validate the effectiveness of prompt engineering with GPT-4.0, we selected the CMeEE (Chinese Medical Entity Extraction) public dataset [29] for the NER task (as shown in Figure 3), establishing a baseline performance by comparing it with traditional models such as Med-BERT [30] and LSTM (long short-term memory) [31]. This comparison demonstrates GPT-4.0's adaptability and flexibility in processing complex medical texts, highlighting the potential of large language models in knowledge graph construction. In addition, to evaluate GPT-4.0's recognition performance in our study, we created a dataset focused on sepsis. Due to the high cost of manual annotation, a medical expert labeled named entities in the electronic medical records (EMRs) of 61 randomly selected patients, with sample annotations provided in the Multimedia Appendix 1. We further assessed GPT-4.0's performance in natural language processing tasks by comparing it with other large models (Qwen2 [Alibaba Cloud], Llama3 [Meta AI], and ChatGPT 3.5 [OpenAI]). The models were evaluated using the F_1-score, which reflects overall performance by calculating true positives, false positives, and false negatives, summarizing both precision and recall. F_1-score, the harmonic mean of precision and recall, provides a comprehensive measure of model effectiveness.

Textbox 1. Prompt design for sepsis relation extraction in zero-shot and few-shot learning scenarios.

Prompt design

Suppose you are an entity-relationship triple extraction model. I'll give you a list of head entity types: subject types, list of tail entity types: object types, list of relations: relations. Give you a sentence, please extract the subject and object in the sentence based on these three lists and form a triplet in the form of (subject, relation, object).

relations: ['Clinical Presentation ', 'Surgical treatment', 'biomarker', 'subclass', 'pathogenesis', 'laboratory testing', 'Imaging examination', 'Complications', 'Medication']

The given sentence is \n {triple list}

Examples(few-shot)

examples = [

{

"text": "In adults, about 1/3 of AHF (Acute Heart Failure) patients may develop concurrent fungal infections, primarily caused by Candida albicans. Circulatory system issues may lead to sinus bradycardia, with a slower heart rate occurring relatively late, and in a minority of cases, sudden cardiac arrest can occur.",

"Triple list": [

["AHF", "Complications", "sinus bradycardia"], ["AHF", "Complications", "fungal infection"], ["Fungal infection", "Biomarker", "Candida albicans"]

]

},

{

"text": "Pancreatic cancer, after 4 months of initial treatment, showed a pancreatic mass with liver metastasis in the upper abdominal ultrasound examination.",

"Triple list": [

["Pancreatic cancer", "Clinical Presentation", "pancreatic mass"], ["Pancreatic cancer", "Imaging examination", "upper abdominal ultrasound examination"], ["Pancreatic cancer", "Clinical Presentation", "liver metastasis"]

]

}

]

XSL•FO
RenderX

Figure 3. Prompt design for zero-shot and few-shot named entity recognition tasks in medical Chinese.

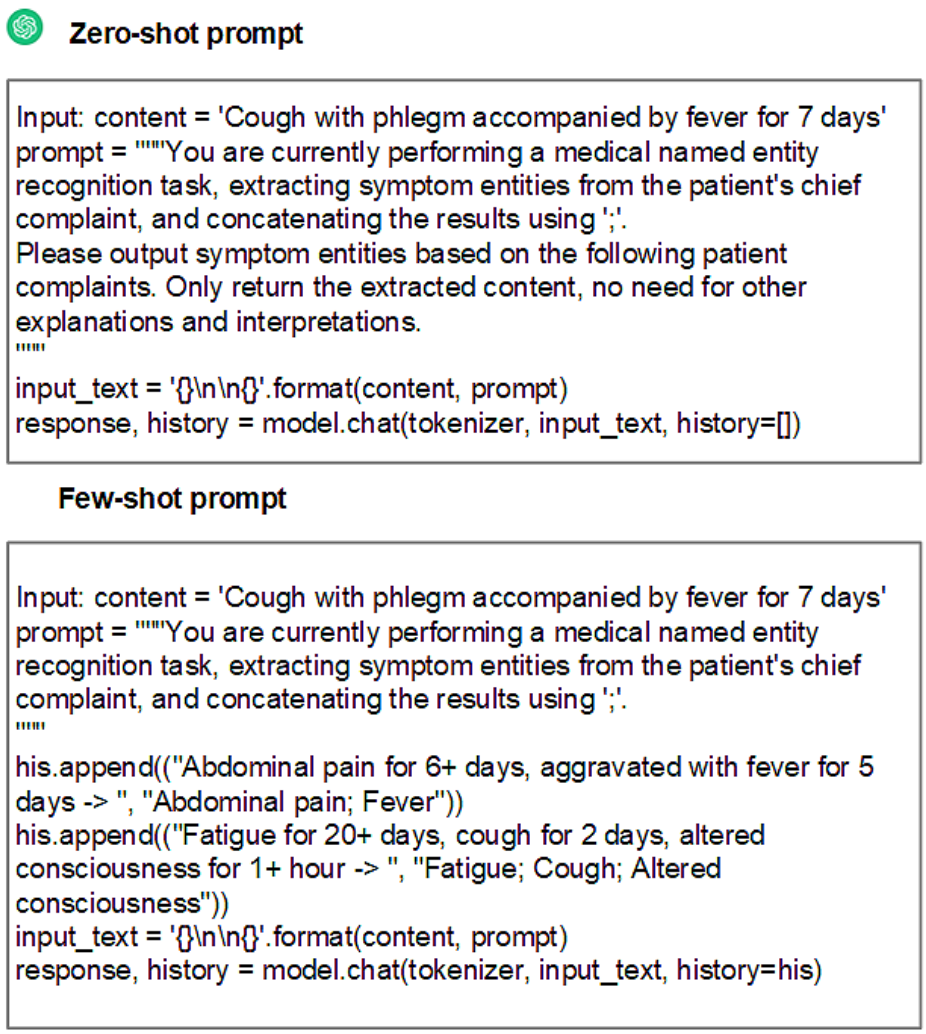

Knowledge Fusion

In our research, we used the entity-linker module from scispaCy [32], which incorporates the Unified Medical Language System (UMLS) [33] terminology standards to construct a comprehensive knowledge base. UMLS facilitated the alignment of terms, such as mapping “Hypertension” to the standardized concept “Hypertensive disease (C0020538).” To ensure effective text matching, we standardized and cleaned entity names across different data sources. This process included unifying letter case, removing punctuation, and handling abbreviations, allowing us to standardize terms such as “Heart Failure” and “HF” to a consistent terminology. Following this initial processing, we used text similarity algorithms for entity matching, using algorithms based on word embeddings and Jaccard similarity to calculate similarity scores between entity names. During the alignment process, we established a similarity threshold of 0.85, ensuring that only highly similar entities were recognized as corresponding. For entities with similarity values below this threshold, we retained their original values and used manual review to address any potential ambiguities, thereby ensuring the precision and credibility of the final alignment results. This meticulously designed workflow not only enhances the efficiency of entity alignment but also safeguards the accuracy and reliability of our research outcomes.

Platform Implementation

MSD performed data management using MySQL software (version 5.7; Oracle Corp). The web pages were developed with Bootstrap 4.0 (The Bootstrap Team) and used the Flask framework. Several JavaScript plugins, such as Datatable (version 1.10.10; SpryMedia Ltd) and ECharts (version 5.0; Baidu), were applied for creating data tables and visualizations. We used the Neo4j graph database [34] for knowledge storage. The underlying storage principles of this graph database involve the use of nodes, edges, and properties to store graph data. Neo4j is currently recognized as one of the most popular high-performance NoSQL graph databases [35], known for its high availability, stability, scalability, and robust intuitiveness. In this research, open-source Python libraries (such as py2neo [36] were used in conjunction with OpenAI interfaces to accomplish knowledge extraction and to add or modify relationships and nodes. Specifically, we used the GPT-4-turbo model accessed via the application programming interface for our artificial intelligence model, and for Scispacy, we used the en_core_sci_sm model. The overall research and construction process is shown in Figure 4.

XSL•FO
RenderX

Figure 4. Framework of the multicenter sepsis knowledge graph construction study (n=10,544 patients with sepsis) integrating GPT-4 and Neo4j technologies. The flow diagram shows retrospective data collection from 3 tertiary hospitals in Western China (West China Hospital, Shangjin Hospital, and Tianfu Hospital; 2020-2023), multimodal data processing through LLM-driven entity recognition, and graph database implementation. LLM: large language model; MSD: multicenter sepsis database.

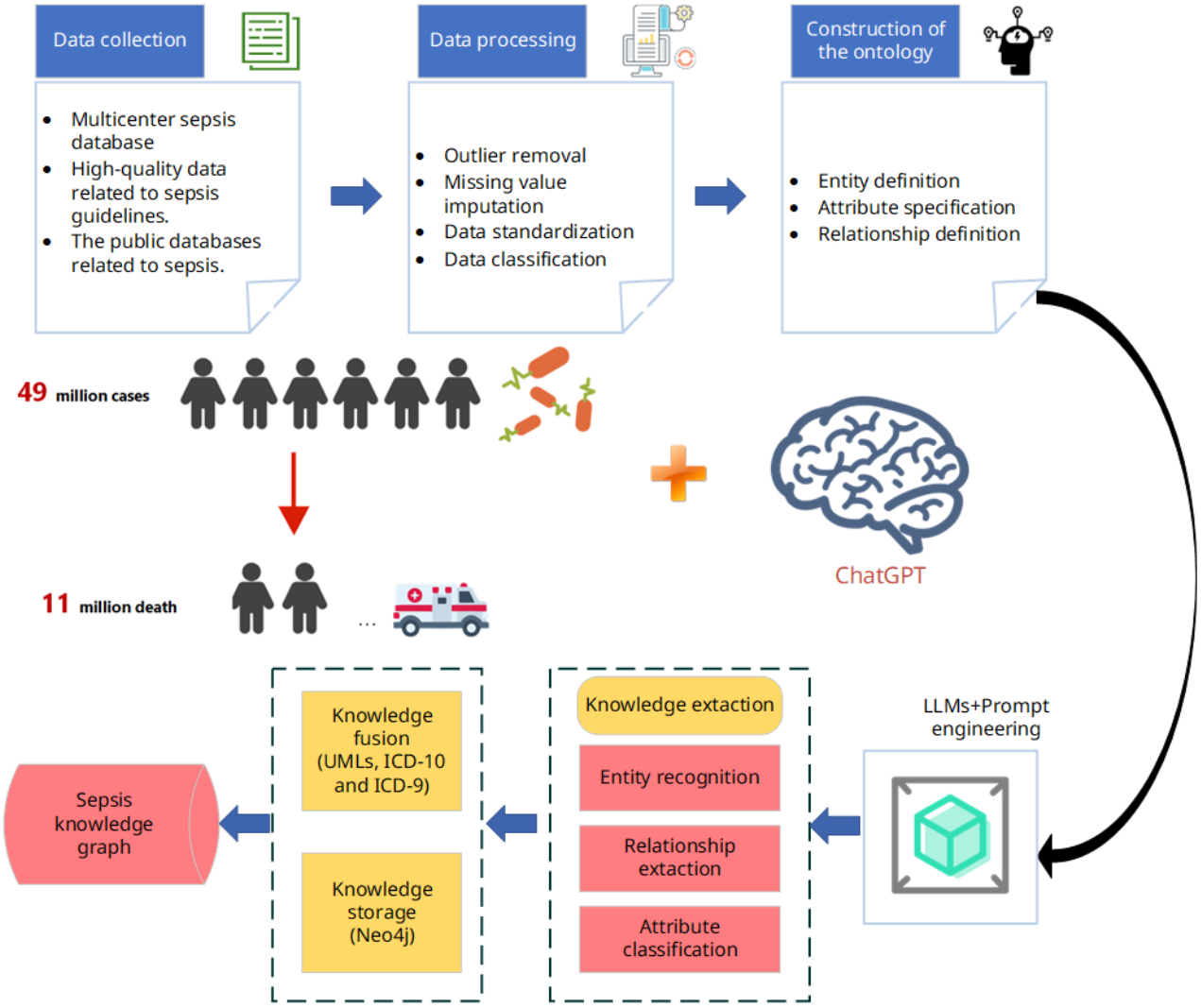

Ethical Considerations

This study was approved by the Medical Ethics Committee of West China Hospital, Sichuan University (approval number 2024-126). The research involved a secondary analysis of deidentified patient data, which was conducted in accordance with the original informed consent obtained during data collection, allowing for subsequent analyses without requiring additional consent. To ensure participant privacy and confidentiality, all personally identifiable information was removed before analysis. No financial compensation was provided to participants for the use of their data in this study. Furthermore, this manuscript does not contain any images that could potentially identify individuals. The relevant consent forms have been submitted as part of the supplementary materials.

Results

In this study, we established an MSD [37] using real-world data from 3 major hospitals in Western China. The database spans the period from 2020 to 2023, covering a total of 10,544 patients diagnosed with sepsis. Each hospital contributed a significant number of patient records, with 8497 records from West China Hospital, 690 from Shangjin Hospital, and 357 from Tianfu Hospital. The MSD includes comprehensive data points such as baseline demographics, laboratory test results, EMRs, and detailed admission and discharge notes. This large-scale, multicenter sepsis dataset provides a solid foundation for constructing a highly detailed and clinically relevant sepsis knowledge graph.

Furthermore, using GPT-4.0, we conducted entity recognition and relationship extraction based on the MSD, expert consensus, and open data sources. Subsequently, we achieved knowledge integration by benchmarking against the UMLS. As a result, we present the definitions of key entities and their semantic relationships in Tables 1 and 2. Table 1 outlines key entities, such as diseases, symptoms, biomarkers, and imaging examinations, which are crucial for understanding the clinical characteristics of sepsis. For instance, the “Symptoms” entity captures the subjective and objective signs experienced by patients, while “Biomarkers” represent measurable substances that indicate the presence or progression of disease. The

XSL•FO
RenderX

relationships between these entities, as defined in Table 2, provide insights into the complex dynamics of sepsis. Relationships such as “Has symptom” and “Complications” show how specific symptoms correlate with diseases and how disease progression may lead to additional health issues. We successfully constructed a comprehensive knowledge graph for sepsis, consisting of 1894 nodes and 2021 relationships. By integrating the definitions of entities and their interrelationships, our study aims to enhance the understanding of sepsis and supports informed clinical decision-making, ultimately contributing to improved patient outcomes. In addition, we have established relevant examples of sepsis ontology, as shown in Figure S7 in Multimedia Appendix 1.

Then, we introduced zero-shot and few-shot methods to explore the model’s performance in handling unseen or infrequent entities based on CMeEE, comparing them with conventional models such as BERT-CRF [34], BERT-wwm [35], Lattice-LSTM [36], and Med-BERT [30]. The results indicate that the few-shot approach using GPT-4 outperformed traditional models, achieving an F_1-score of 65.42, which demonstrates its effectiveness in adapting to new entity types with limited annotated data. In addition, we constructed a Sepsis dataset, which consists of clinical records annotated by a medical expert, to further evaluate the models in a real-world health care context. Comparing performance across models, GPT-4 achieved a notable F_1-score of 76.76, significantly surpassing other LLMs such as Qwen2 and Llama3, which indicates its strong capability in recognizing critical medical entities in this specialized domain. Tables 3 and 4 present the performance of different models on the CMeEE dataset and the Sepsis dataset.

Table 3. Zero-shot and few-shot learning performance comparison of GPT-4 versus BERT and Lattice-LSTM models on Chinese medical entity extraction dataset.

Model	CMeEE[a]		
	Precision	Recall	F_1-score
BERT[b]-base	63.08	64.08	62.11
BERT-wwm	61.5	61.29	61.72
Lattice-LSTM[c]	46.34	43.60	49.44
Med-BERT	53.33	47.58	60.66
GPT-4 and Zero shot	64.07	68.97	59.82
GPT-4 and Few shot	65.31	64.89	65.73

[a]CMeEE: Chinese Medical Entity Extraction.

[b]BERT: Bidirectional Encoder Representations from Transformers.

[c]LSTM: long short-term memory.

Table 4. Performance comparison of GPT-4, Qwen2-72B, Llama3-70B, and GPT-3.5 models on the sepsis dataset.

Model	Sepsis dataset		
	Precision	Recall	F_1-score
Qwen2-72B	44.73	42.85	43.77
Llama3-70B	49.40	47.43	48.39
GPT-3.5	56.63	54.48	55.53
GPT-4 and Zero shot	72.12	70.48	71.29
GPT-4 and Few shot	77.73	75.81	76.76

Major Complication of Sepsis

The investigation of complications associated with sepsis holds paramount significance in clinical research. In our study, we identified 203 complications-related nodes from the collected cases and literature (Figure 5). We queried the relationship between sepsis and complications, and through the query MATCH ()-[r:`has complication`]->() RETURN r ORDER BY r.weight DESC LIMIT 10.

XSL•FO
RenderX

Figure 5. Top 10 sepsis complications identified through weighted relationship analysis in the multicenter sepsis database: Bar graph displays the prevalence of pneumonia (32.1%), hypoproteinemia (28.5%), septic shock (25.9%), and so on.

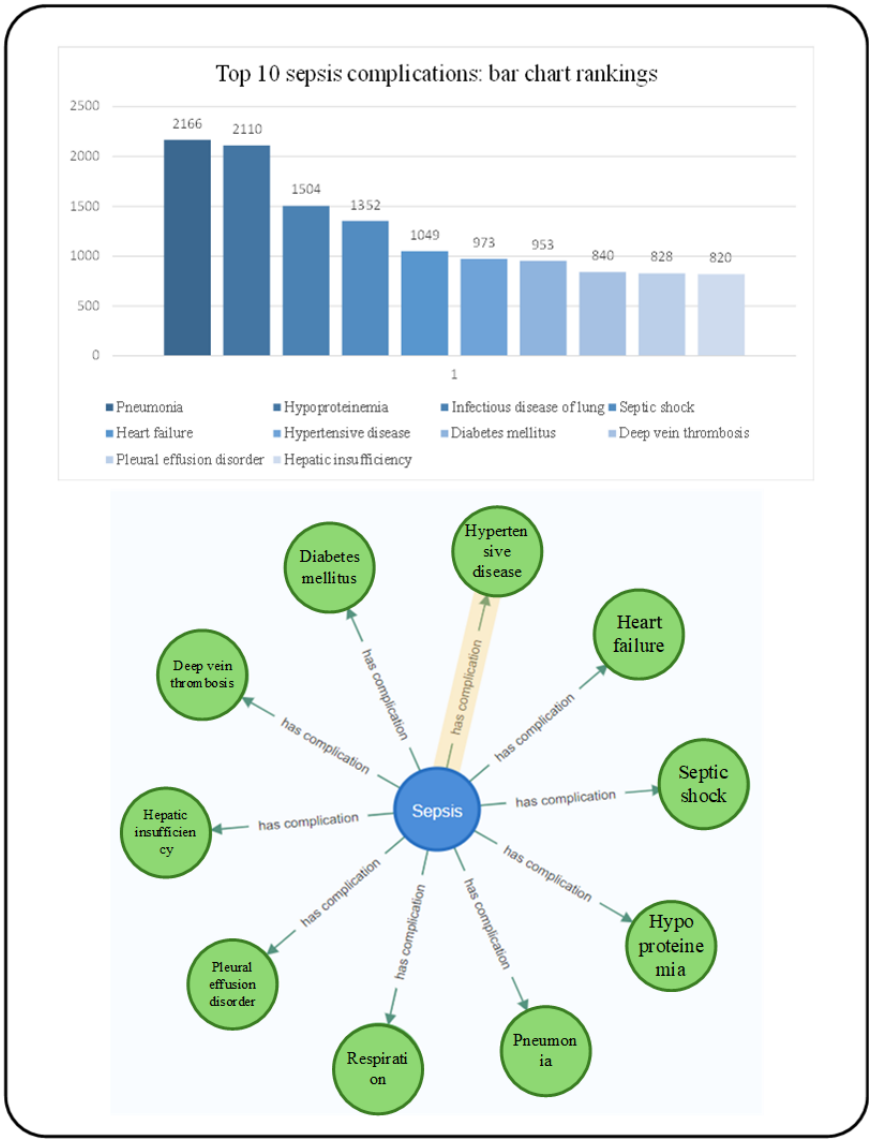

Sepsis Medication Statistics

Sepsis is a severe systemic inflammatory response frequently triggered by infection. In the knowledge graph, entering the following code reveals the relationships between the most commonly used medications for sepsis and their associated complications (Figure 6). The following query was used: MATCH p=(n:medication)-[r:`recommended medication`]-() RETURN p order by r.weight ASC limit 25.

XSL•FO
RenderX

Figure 6. Pharmacological intervention patterns in sepsis management: a chord diagram depicting medication-complication relationships (physiological saline, glucose infusion, etc).

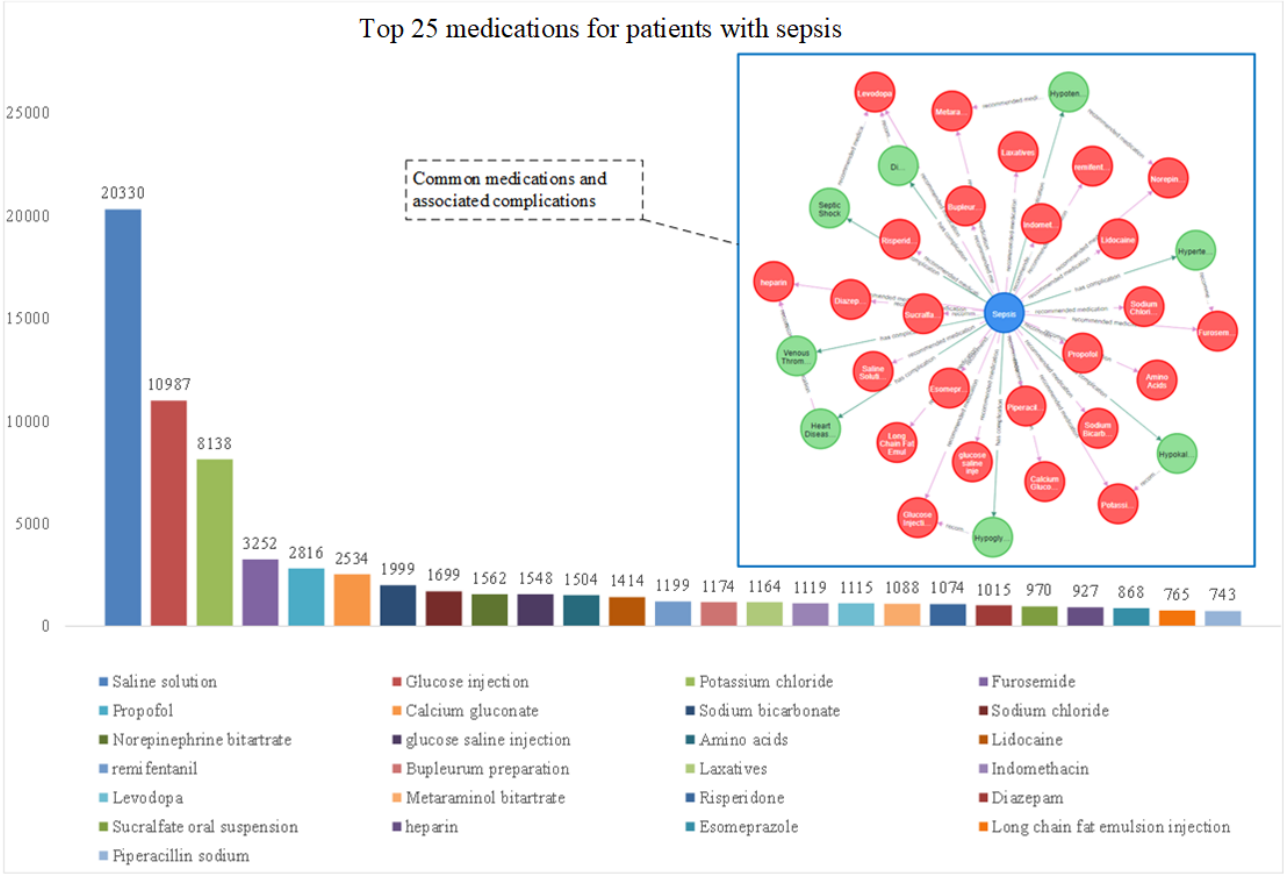

Sepsis Symptom Statistics

Sepsis is defined as a comprehensive systemic inflammatory response, and its multifaceted nature is revealed through the statistical analysis of the top 25 symptoms observed in patients (Figure 7). Relevant queries were conducted in the knowledge graph, and the corresponding code is MATCH p=()-[r:`has symptom`]->() RETURN p order by r.weight asc.

Figure 7. Knowledge graph representation of common symptoms of sepsis and their top 10 frequency.

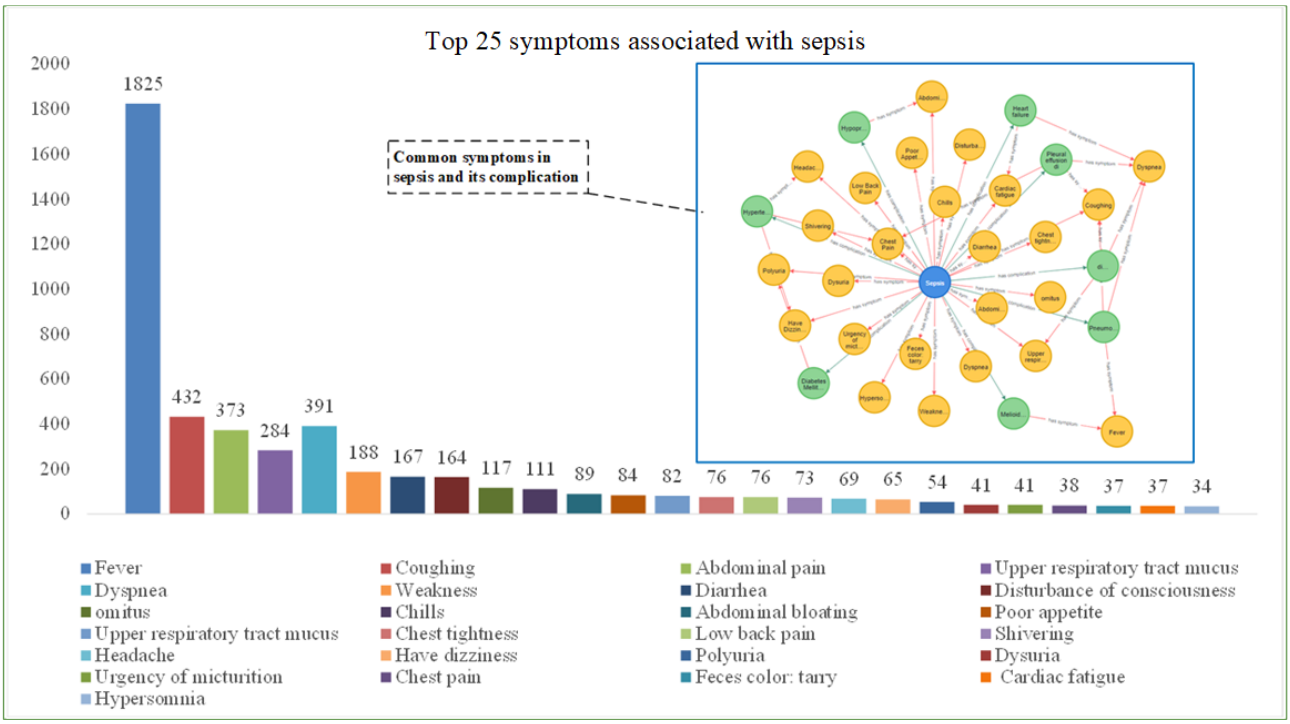

XSL•FO
RenderX

Clinical Applications

Based on the results from the sepsis knowledge graph, we used GraphRAG [38] technology to assist large models in generating more precise decision support for clinicians. In comparison with traditional Retrieval Augmented Generation (RAG) technology, GraphRAG excels in handling complex relationships and multistep reasoning, providing more comprehensive and accurate answers, which is particularly crucial for the diagnosis and treatment of complex diseases like sepsis, as clinical decisions frequently involve multiple variables and potential interactions. Furthermore, GraphRAG enhances efficiency in information retrieval and customized summary generation, enabling clinicians to quickly access relevant information and recommendations related to patient conditions. This improvement in efficiency significantly bolsters the accuracy of clinical decisions, allowing physicians to make more informed judgments in complex clinical environments. With advancements in computational power, the integration of the sepsis knowledge graph and large models can consolidate data from electronic health records, laboratory tests, imaging, and critical care, enabling real-time analysis of patients' clinical features and scoring systems (such as SOFA and APACHE II) to assist doctors in quickly identifying patients with sepsis across various risk levels and facilitating early warning systems. In the clinical decision-making process, the combination of the knowledge graph, RAG, and prompt engineering strategies offers treatment recommendations grounded in high-quality evidence, thereby enhancing the scientific and effective nature of decision-making. Currently, our clinical team has provided relevant case studies, showcased in supplementary materials, which include a comparison of GraphRAG technology before and after its implementation based on GPT-4. By using clinical gold standards and using BLEU [39], ROUGE-1, ROUGE-2, and ROUGE-L metrics [40], we compared the model's generated results, with BLEU primarily assessing the overlap of vocabulary between the generated text and reference text, reflecting the accuracy of the generated content, while ROUGE focuses on recall, emphasizing the comprehensiveness of the generated text in capturing reference content. Through a comprehensive analysis of these metrics, as shown in Figure 8, we further validated the effectiveness and advantages of GraphRAG in practical applications.

Figure 8. Comparative performance analysis of GraphRAG technology before and after implementation in sepsis clinical decision support.

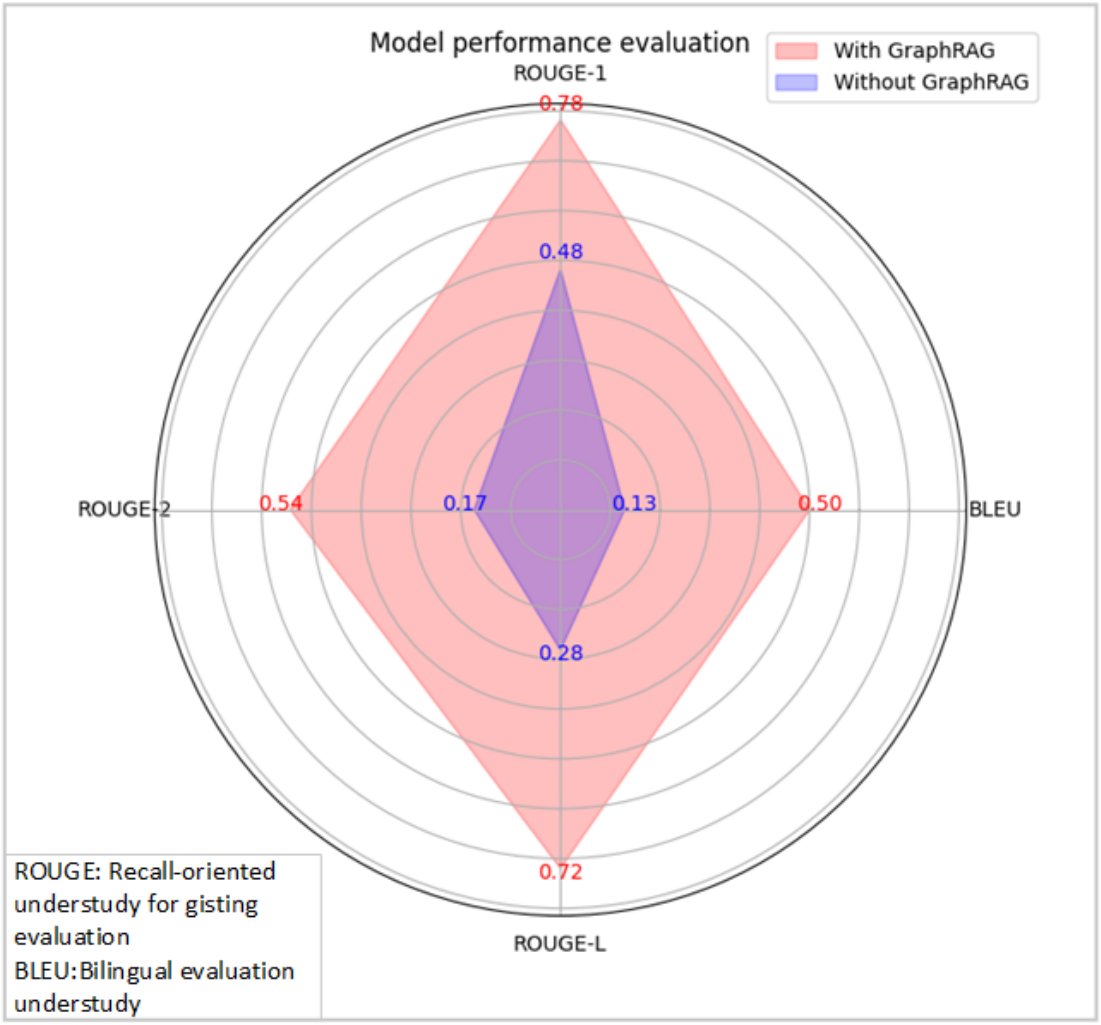

XSL•FO
RenderX

Discussion

Principal Findings

This study aimed to develop a framework for constructing a sepsis knowledge graph using large LLMs, leveraging real-world, multicenter clinical data. The core objective was to explore the feasibility and effectiveness of using LLM technologies, particularly GPT-4.0, to construct a comprehensive, domain-specific knowledge graph, focused on sepsis. Our findings indicate that GPT-4.0, through careful prompt engineering, is capable of accurately extracting entities and relationships from clinical data, outperforming traditional models such as BERT and LSTM. In particular, GPT-4.0 demonstrated the best few-shot performance on the CMeEE dataset, achieving an F_1-score of 65.42, and outperformed other open-source models on the Sepsis dataset with an F_1-score of 76.76. These results underscore the potential of LLMs for building specialized knowledge graphs in fields with limited data annotations, highlighting the advantages of the adaptability of LLMs across tasks and domains without the need for extensive model retraining.

The successful construction of a sepsis-specific knowledge graph marks a significant advancement in both knowledge extraction and clinical decision support. By using GPT-4.0's pretrained capabilities, we were able to generate meaningful insights from small datasets, which typically require substantial domain-specific training. These findings align with other studies that have used large models such as GPT-4.0, GPT-3.5, LlaMA [10], and SOLAR [41], for biomedical relationship extraction. In particular, GPT-4.0 has shown superior performance in extracting biomedical relationships from semistructured data, achieving F_1-scores above 0.881 [42], which further confirms its utility for creating knowledge graphs in health care. Our results also demonstrate that GPT-4.0 excels at quickly adapting to specialized domains, such as sepsis, and constructing medically relevant knowledge graphs without the extensive preprocessing and data annotation typically required in traditional machine learning methods. To contextualize these technical advancements within clinical applications, we conducted comparative analyses against established biomedical knowledge infrastructures, including BIOS [17], PrimeKG [7], and UMLS. This benchmarking reveals several critical advantages of our sepsis-focused approach. While BIOS prioritizes comprehensive ontology coverage through generalized semantic relationships [43], our model achieves superior clinical relevance through context-aware embeddings that preserve disease-specific pathophysiological hierarchies. In addition, while PrimeKG advances precision medicine via molecular-level drug-target interactions, our framework integrates multimodal clinical data streams, including 178 sepsis-associated biomarkers with temporal granularity, enabling dynamic risk stratification unavailable in existing resources. Furthermore, through systematic alignment with UMLS terminological standards, our implementation ensures interoperability without compromising domain specificity, addressing a critical limitation of conventional clinical knowledge graphs, which either sacrifice granularity for standardization or vice versa.

However, the sepsis knowledge graph we have constructed provides not only valuable insights into clinical practices but also enhances understanding of the complex pathophysiology of sepsis. The graph includes key complications such as pneumonia, hypoproteinemia, septic shock, and heart failure, as illustrated in Figure 5. By capturing these critical components, the graph aids in risk stratification and prognostication, ultimately improving treatment efficiency. In addition, our knowledge graph extends to therapeutic optimization. As shown in Figure 6, the system quantifies dynamic relationships between physiological saline administration, glucose metabolism correction, and electrolyte rebalancing interventions [41-43]. Machine learning modules further enable predictive modeling of antimicrobial efficacy based on pathogen susceptibility patterns and pharmacokinetic parameters [44,45], and simultaneously tracking analgesia requirements through nociception biomarkers [46]. This integrative approach allows for the real-time adjustment of therapeutic regimens in response to evolving inflammatory markers and organ dysfunction indicators. Our visualization platform also incorporates comprehensive symptom analytics, as shown in Figure 7. The graph's temporal mapping functionality reveals progression trajectories from initial respiratory manifestations (eg, tachypnea, bronchial hypersecretion [47]) to subsequent multiorgan involvement, including gastrointestinal dysmotility [48], cardiovascular compromise [49], urinary symptoms [50], and neurological deterioration [51]. By correlating symptom clusters with biomarker profiles and treatment responses, the system enables early detection of sepsis-induced organ failure and provides decision support for personalized intervention strategies. This systematic integration of multimodal clinical data ultimately fosters a precision medicine paradigm in sepsis management, bridging pathophysiological insights with optimized therapeutic execution.

While the results are promising, we acknowledge several limitations in this study. Although GPT-4.0 demonstrated impressive entity recognition and relationship extraction capabilities, it occasionally produced erroneous outputs. For instance, the model recommended irrelevant laboratory tests (eg, "Blood cholesterol test") for sepsis and misattributed causative factors, such as "Sepsis caused by exercise." In addition, GPT-4.0 struggled with the nuanced semantics of traditional Chinese medicine terms, such as "Yin-Yang imbalance" and "Qi Deficiency," leading to misinterpretations. These errors are significant as they could affect downstream clinical applications, such as diagnostic accuracy and therapeutic decision-making. However, the inherent limitations of GPT-4.0 in handling specialized medical terminologies and subtle relationships underscore the importance of a robust manual review process to ensure the accuracy of the knowledge graph. We have implemented such a review process to mitigate errors and maintain data integrity; however, further refinement of the model is necessary to reduce reliance on human intervention. In addition to these technical limitations, the use of GPT-4.0 in health care also raises important privacy and security concerns [52]. As noted, our study implemented stringent measures to

XSL•FO
RenderX

protect patient data, including the elimination of personally identifiable information and adherence to relevant privacy regulations. Despite these precautions, the use of commercial models like GPT-4.0 poses risks related to data leakage and the potential exposure of sensitive patient information [53].

In future research, to ensure data security, we plan to deploy localized models, such as Llama 3.1 and Qwen 2, in secure environments, ensuring that all data processing activities are contained within controlled systems. By doing so, we aim to enhance data privacy and reduce the risk of data breaches, ensuring that the integrity of the knowledge graph remains intact while safeguarding patient confidentiality. Furthermore, we plan to further optimize the model's accuracy and contextual understanding by incorporating advanced models such as Gemini [54] and Claude [55], which may provide more precise and contextually accurate results. We also aim to expand our dataset to include additional clinical and traditional Chinese medicine-related data to further enhance the robustness and adaptability of the knowledge graph. Fine-tuning techniques, such as Low-rank adaptation [56] and p-tuning [57], will be explored to improve both model precision and efficiency, ensuring that the knowledge graph captures a broader range of medical nuances. Furthermore, we will experiment with alternative prompt designs, including converting entity recognition tasks into classification tasks [58] or referencing the Teler taxonomy [59], which could improve the model's ability to handle complex medical relationships more effectively. As part of this effort, we will also generate more gold standard data to compare the impact of various prompt strategies on knowledge graph accuracy.

In summary, this study highlights the potential of LLM-driven approaches, particularly GPT-4.0, in constructing accurate and effective medical knowledge graphs, even in data-scarce environments. The sepsis knowledge graph we developed not only enhances the understanding of sepsis but also illustrates the broader applicability of LLM-driven methods in creating domain-specific knowledge graphs across various medical fields. As large models like GPT-4.0 continue to evolve, their ability to process and interpret complex medical data will revolutionize how health care professionals access, interpret, and apply medical knowledge. The scalability and adaptability of this approach indicate its potential for widespread use in a variety of diseases, laying the foundation for personalized and precision medicine in clinical practice. In addition, the interactive and visual nature of the knowledge graph increases its practical utility, providing clinicians with a dynamic tool for both research and everyday decision-making.

Conclusions

The successful establishment of the MSD, leveraging real-world data and the innovative application of LLMs like GPT-4.0, represents a significant leap forward in the understanding and management of sepsis. The use of prompt engineering to build a comprehensive sepsis knowledge graph not only showcases the potential of LLM technologies in medical research but also highlights their capacity for significant generalization and adaptation. Our findings, especially the enhanced performance of the few-shot model in entity identification and relationship understanding, underscore the efficiency and cost-effectiveness of using LLMs in the rapid development of specialized databases and knowledge graphs. This pioneering approach sets a new benchmark for future research and database development in the field of sepsis and potentially other medical domains.

Acknowledgments

The authors would like to thank Dr Rongrong Wu for providing consultation on figure drawings. We are also grateful to the staff in our research groups who contributed to the study through their valuable contributions and discussions. This work was supported by the National Natural Science Foundation of China (grants 32200545 and 32270690), the 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (ZYGD23012 and ZYAI24044), Chengdu Medical Research Project (2024269) and was funded by the EU and the Xunta de Galicia (Spain; grant ED431C2022/46 for Competitive Reference Groups; GRC). It was also supported by CITIC-UDC and INIBIC.

Data Availability

To ensure data security, the database currently does not support direct downloading on the website. Researchers in need can contact Professor Bairong Shen, head of the Disease Systems Genetics Research Institute at West China Hospital, by email at bairong.shen@scu.edu.cn

Authors' Contributions

HY conceptualized and designed the study, performed the research, delivered the model analysis, analyzed the data, wrote the manuscript; CZ and JL delivered the data analysis and clinical analysis. BS and APS conceptualized and designed the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information.

XSL•FO
RenderX

[DOCX File , 1519 KB - jmir_v27i1e65537_app1.docx]

References

1. Srzi I, Nesek Adam V, Tunji Pejak D. Definition of Sepsis: What's New in the Treatment Guidelines. Acta clinica Croatica 2022;61(Supplement 1):67-72.
2. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global burden of disease study. The Lancet 2020;395(10219):200-211. [doi: 10.1016/s0140-6736(19)32989-7]
3. Anggraini D, Hasni D, Amelia R. Pathogenesis of sepsis. Scientific j 2022;1(4):332-339. [doi: 10.56260/sciena.v1i4.63]
4. Baghela A, Pena OM, Lee AH, Baquir B, Falsafi R, An A, et al. Predicting sepsis severity at first clinical presentation: the role of endotypes and mechanistic signatures. EBioMedicine 2022;75:103776 [FREE Full text] [doi: 10.1016/j.ebiom.2021.103776] [Medline: 35027333]
5. Liu D, Huang SY, Sun JH, Zhang HC, Cai QL, Gao C, et al. Sepsis-induced immunosuppression: mechanisms, diagnosis and current treatment options. Mil Med Res 2022;9(1):56 [FREE Full text] [doi: 10.1186/s40779-022-00422-y] [Medline: 36209190]
6. Sekino N, Selim M, Shehadah A. Sepsis-associated brain injury: underlying mechanisms and potential therapeutic strategies for acute and long-term cognitive impairments. J Neuroinflammation 2022;19(1):101 [FREE Full text] [doi: 10.1186/s12974-022-02464-4] [Medline: 35488237]
7. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data 2023;10(1):67 [FREE Full text] [doi: 10.1038/s41597-023-01960-3] [Medline: 36732524]
8. Li Z, Zhong Q, Yang J, Duan Y, Wang W, Wu C, et al. DeepKG: an end-to-end deep learning-based workflow for biomedical knowledge graph extraction, optimization and applications. Bioinformatics 2022;38(5):1477-1479 [FREE Full text] [doi: 10.1093/bioinformatics/btab767] [Medline: 34788369]
9. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. Gpt-4 technical report. arXiv:2303.08774 2023 [FREE Full text]
10. Qin M. The uniqueness of LLaMA3-70B with per-channel quantization. arXiv:2408.15301 2024 [FREE Full text]
11. Hu Y, Chen Q, Du J, Peng X, Keloth VK, Zuo X, et al. Improving large language models for clinical named entity recognition via prompt engineering. J Am Med Inform Assoc 2024 Sep 01;31(9):1812-1820. [doi: 10.1093/jamia/ocad259] [Medline: 38281112]
12. Yuan C, Xie Q, Ananiadou S. Zero-shot temporal relation extraction with ChatGPT. arXiv:2304.05454 2023 [FREE Full text]
13. Arvidsson S, Axell J. Prompt engineering guidelines for LLMs in requirements engineering. GUPEA. 2023. URL: https://gupea.ub.gu.se/handle/2077/77967 [accessed 2023-08-07]
14. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25:e50638 [FREE Full text] [doi: 10.2196/50638] [Medline: 37792434]
15. Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. Minds Mach 2020;30:681-694. [doi: 10.1007/s11023-020-09548-1]
16. Zhang C, Zhang X, Sun Z, Liu X, Shen B. MetaSepsisBase: a biomarker database for systems biological analysis and personalized diagnosis of heterogeneous human sepsis. Intensive Care Med 2023;49(8):1015-1017. [doi: 10.1007/s00134-023-07126-4] [Medline: 37329364]
17. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, et al. The NCBI BioSystems database. Nucleic Acids Res 2010;38(Database issue):D492-D496 [FREE Full text] [doi: 10.1093/nar/gkp858] [Medline: 19854944]
18. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res 2023;51(D1):D1003-D1009 [FREE Full text] [doi: 10.1093/nar/gkac888] [Medline: 36243972]
19. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 2009 Jun 01;25(11):1422-1423 [FREE Full text] [doi: 10.1093/bioinformatics/btp163] [Medline: 19304878]
20. Ma X. Knowledge graph construction and application in geosciences: A review. Comput Geosci 2022;161:105082. [doi: 10.1016/j.cageo.2022.105082]
21. Zamini M, Reza H, Rabiei M. A review of knowledge graph completion. Information 2022;13(8):396. [doi: 10.3390/info13080396]
22. Chen Y, Li H, Li H, Liu W, Wu Y, Huang Q, et al. An overview of knowledge graph reasoning: key technologies and applications. J Sens Actuator Netw 2022;11(4):78. [doi: 10.3390/jsan11040078]
23. Tang X, Feng Z, Xiao Y, Wang M, Ye T, Zhou Y, et al. Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development. Geosci Front 2023;14(5):101426. [doi: 10.1016/j.gsf.2022.101426]
24. Wang B, Wang Z, Wang X, Cao Y, Saurous RA, Kim Y. Grammar prompting for domain-specific language generation with large language models. Advances in Neural Information Processing Systems 2024:36 [FREE Full text]

XSL•FO
RenderX

25. Gokul A. LLMs and AI: understanding its reach and impact. Preprints.org. 2023. URL: https://www.preprints.org/manuscript/202305.0195/v1 [accessed 2023-05-04]
26. Kirk JR, Wray RE, Lindes P, Laird JE. Improving knowledge extraction from LLMs for task learning through agent analysis. Proceedings of the AAAI Conference on Artificial Intelligence 2024;38(16):18390-18398. [doi: 10.1609/aaai.v38i16.29799]
27. Kojima T, Shane Gu, SS, Reid, M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. arXiv:2205.11916 2022;35 [FREE Full text]
28. Kang S, Yoon J, Yoo S. Large language models are few-shot testers: exploring LLM-based general bug reproduction. 2023 Presented at: IEEE/ACM 45th International Conference on Software Engineering (ICSE); May 14-20, 2023; Melbourne, Australia. [doi: 10.1109/icse48619.2023.00194]
29. Zhang N, Chen M, Bi Z, Liang X, Li L, Shang X, et al. CBLUE: a Chinese Biomedical Language Understanding Evaluation benchmark. ArXiv Preprint posted online on June 15, 2021 [FREE Full text]
30. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit Med 2021;4(1):86 [FREE Full text] [doi: 10.1038/s41746-021-00455-y] [Medline: 34017034]
31. Graves A. Long short-term memory. In: Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Heidelberg: Springer; 2012:37-45.
32. Neumann M, Daniel King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. arXiv:1902.07669 2019. [doi: 10.18653/v1/w19-5034]
33. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(Database issue):D267-D270. [doi: 10.1093/nar/gkh061] [Medline: 14681409]
34. Souza F, Nogueira R, Lotufo R. Portuguese named entity recognition using BERT-CRF. arXiv:1909.10649 2019 [FREE Full text]
35. Zhang Y, Liao X, Chen L, Kang H, Cai Y, Wang Q. Multi-BERT-wwm model based on probabilistic graph strategy for relation extraction. In: Health Information Science. Cham: Springer; 2021.
36. Zhang Y, Yang J. Chinese NER using lattice LSTM. arXiv:1805.02023 2018 [FREE Full text] [doi: 10.18653/v1/p18-1144]
37. Multicenter Sepsis Database (MSD): based on real-world clinical data. Multicentric Sepsis Database. URL: http://sysbio.org.cn/MSD [accessed 2025-03-06]
38. Edge D, Trinh H, Cheng N, Bradley J, Chao A, Mody A, et al. From local to global: a graph RAG approach to query-focused summarization. arXiv:2404.16130 2024 [FREE Full text]
39. Post M. A call for clarity in reporting BLEU scores. arXiv:1804.08771 2018 [FREE Full text] [doi: 10.18653/v1/w18-6319]
40. Lin CY. Rouge: a package for automatic evaluation of summaries. 2004 Presented at: Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004); July 25, 2004; Barcelona, Spain URL: https://www.semanticscholar.org/paper/ROUGE%3A-A-Package-for-Automatic-Evaluation-of-Lin/60b05f32c32519a809f21642ef1eb3eaf3848008
41. Cha JK, Kim HS, Kim EJ, Lee ES, Lee JH, Song IA. Effect of early nutritional support on clinical outcomes of critically ill patients with sepsis and septic shock: a single-center retrospective study. Nutrients 2022;14(11):2318 [FREE Full text] [doi: 10.3390/nu14112318] [Medline: 35684117]
42. Jiang L, Cheng M. Impact of diabetes mellitus on outcomes of patients with sepsis: an updated systematic review and meta-analysis. Diabetol Metab Syndr 2022;14(1):39 [FREE Full text] [doi: 10.1186/s13098-022-00803-2] [Medline: 35248158]
43. Wang Z, Zhang L, Xu F, Han D, Lyu J. The association between continuous renal replacement therapy as treatment for sepsis-associated acute kidney injury and trend of lactate trajectory as risk factor of 28-day mortality in intensive care units. BMC Emerg Med 2022;22(1):32 [FREE Full text] [doi: 10.1186/s12873-022-00589-6] [Medline: 35227200]
44. Meng Z, Pan L, Qian S, Yang X, Pan L, Chi R, et al. Antimicrobial peptide nanoparticles coated with macrophage cell membrane for targeted antimicrobial therapy of sepsis. Mater Des 2023;229:111883. [doi: 10.1016/j.matdes.2023.111883]
45. Lukaszewski RA, Jones HE, Gersuk VH, Russell P, Simpson A, Brealey D, et al. Presymptomatic diagnosis of postoperative infection and sepsis using gene expression signatures. Intensive Care Med 2022;48(9):1133-1143 [FREE Full text] [doi: 10.1007/s00134-022-06769-z] [Medline: 35831640]
46. Carpenter KC, Hakenjos JM, Fry CD, Nemzek JA. The influence of pain and analgesia in rodent models of sepsis. Comp Med 2019;69(6):546-554. [doi: 10.30802/aalas-cm-19-000004]
47. Mayow AH, Ahmad F, Afzal MS, Khokhar MU, Rafique D, Vallamchetla SK, et al. A systematic review and meta-analysis of independent predictors for acute respiratory distress syndrome in patients presenting with sepsis. Cureus 2023;15(4):e37055 [FREE Full text] [doi: 10.7759/cureus.37055] [Medline: 37143620]
48. Tyszko M, Lemańska-Perek A, Śmiechowicz J, Tomaszewska P, Biecek P, Gozdzik W, et al. Citrulline, intestinal fatty acid-binding protein and the acute gastrointestinal injury score as predictors of gastrointestinal failure in patients with sepsis and septic shock. Nutrients 2023;15(9):2100 [FREE Full text] [doi: 10.3390/nu15092100] [Medline: 37432225]
49. Bronicki RA, Tume SC, Flores S, Loomba RS, Borges NM, Penny DJ, et al. The cardiovascular system in severe sepsis: insight from a cardiovascular simulator. Pediatr Crit Care Med 2022;23(6):464-472. [doi: 10.1097/PCC.0000000000002945] [Medline: 35435883]

XSL•FO
RenderX

50. Bazaid AS, Aldarhami A, Gattan H, Barnawi H, Qanash H, Alsaif G, et al. Antibiogram of urinary tract infections and sepsis among infants in neonatal intensive care unit. Children (Basel) 2022;9(5):629 [FREE Full text] [doi: 10.3390/children9050629] [Medline: 35626805]
51. Becker AE, Teixeira SR, Lunig NA, Mondal A, Fitzgerald JC, Topjian AA, et al. Sepsis-related brain MRI abnormalities are associated with mortality and poor neurological outcome in pediatric sepsis. Pediatr Neurol 2022;128:1-8 [FREE Full text] [doi: 10.1016/j.pediatrneurol.2021.12.001] [Medline: 34992035]
52. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. J Med Internet Res 2023;25:e48009 [FREE Full text] [doi: 10.2196/48009] [Medline: 37566454]
53. Aydin I, Diebel-Fischer H, Freiberger V, Möller-Klapperich J, Buchmann E, Färber M, et al. Assessing privacy policies with AI: ethical, legal, and technical challenges. arXiv:2410.08381v1 2024 [FREE Full text]
54. Google GT, Anil R, Borgeaud S, Alayrac JB, Yu J, Soricut R, et al. Gemini: a family of highly capable multimodal models. arXiv:2312.11805 2023 [FREE Full text]
55. Wu S, Koo M, Blum L, Black A, Kao L, Scalzo F, et al. A comparative study of open-source large language models, GPT-4 and Claude 2: multiple-choice test taking in nephrology. arXiv:2308.04709 2023 [FREE Full text]
56. Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. LoRA: low-rank adaptation of large language models. arXiv:2106.09685 2021 [FREE Full text]
57. Tam WL, Liu X, Ji K, Xue L, Zhang X, Dong Y, et al. Parameter-efficient prompt tuning makes generalized and calibrated neural text retrievers. arXiv:2207.07087 2022 [FREE Full text] [doi: 10.18653/v1/2023.findings-emnlp.874]
58. Zhou S, Yu S. High-throughput biomedical relation extraction for semi-structured web articles empowered by large language models. arXiv:2312.08274 2023 [FREE Full text]
59. Karmaker Santu SK, Feng D. TELeR: a general taxonomy of LLM prompts for benchmarking complex tasks. arXiv:2305.11430 2023 [FREE Full text] [doi: 10.18653/v1/2023.findings-emnlp.946]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CMeEE: Chinese Medical Entity Extraction
EMR: electronic medical record
LLM: large language model
LSTM: long short-term memory
MSD: multicenter sepsis database
NER: named entity recognition
RAG: Retrieval Augmented Generation
UMLS: Unified Medical Language System
VAERS: Vaccine Adverse Event Reporting System

Edited by A Mavragani; submitted 18.08.24; peer-reviewed by S Zeng, GC Silaghi, W Yan; comments to author 17.10.24; revised version received 28.10.24; accepted 18.02.25; published 27.03.25.

Please cite as:
Yang H, Li J, Zhang C, Sierra AP, Shen B
Large Language Model–Driven Knowledge Graph Construction in Sepsis Care Using Multicenter Clinical Databases: Development and Usability Study
J Med Internet Res 2025;27:e65537
URL: https://www.jmir.org/2025/1/e65537
doi:10.2196/65537
PMID:

©Hao Yang, Jiaxi Li, Chi Zhang, Alejandro Pazos Sierra, Bairong Shen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Twenty-Five Years of Evolution and Hurdles in Electronic Health Records and Interoperability in Medical Research: Comprehensive Review

Yun Shen[1*], MD; Jiamin Yu[2*], MD; Jian Zhou[2], MD, PhD; Gang Hu[1], MPH, MD, PhD

[1]Chronic Disease Epidemiology, Population and Public Health, Pennington Biomedical Research Center, Baton Rouge, LA, United States

[2]Department of Endocrinology and Metabolism, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Jian Zhou, MD, PhD
Department of Endocrinology and Metabolism
Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine
600 Yishan Road
Shanghai, 200233
China
Phone: 86 64369181 ext 8337
Email: zhoujian@sjtu.edu.cn

Abstract

Background: Electronic health records (EHRs) facilitate the accessibility and sharing of patient data among various health care providers, contributing to more coordinated and efficient care.

Objective: This study aimed to summarize the evolution of secondary use of EHRs and their interoperability in medical research over the past 25 years.

Methods: We conducted an extensive literature search in the PubMed, Scopus, and Web of Science databases using the keywords *Electronic health record* and *Electronic medical record* in the title or abstract and *Medical research* in all fields from 2000 to 2024. Specific terms were applied to different time periods.

Results: The review yielded 2212 studies, all of which were then screened and processed in a structured manner. Of these 2212 studies, 2102 (93.03%) were included in the review analysis, of which 1079 (51.33%) studies were from 2000 to 2009, 582 (27.69%) were from 2010 to 2019, 251 (11.94%) were from 2020 to 2023, and 190 (9.04%) were from 2024.

Conclusions: The evolution of EHRs marks an important milestone in health care's journey toward integrating technology and medicine. From early documentation practices to the sophisticated use of artificial intelligence and big data analytics today, EHRs have become central to improving patient care, enhancing public health surveillance, and advancing medical research.

(J Med Internet Res 2025;27:e59024) doi:10.2196/59024

KEYWORDS

electronic health record; electronic medical record; medical research; interoperability; eHealth; systematic review; real-world evidence; artificial intelligence

Introduction

Tracing Back to History

The practice of documenting patient information dates back approximately 3000 years, when early humans created some of the first known medical records by inscribing case histories on materials such as papyrus, clay tablets, animal bones, and other surfaces [1-3]. These records often included symptoms, diagnoses, and treatments. The health care delivery system developed with a long history and became increasingly complex. During the time of classical medicine, health care providers frequently gathered and systematized medical knowledge, creating standardized textbooks, which were often recorded by case series rather than individual patients [4]. In the 19th century, health care providers began keeping detailed paper records. These were used to track patients' medical histories, nursing information, treatments, and outcomes. Hospitals and

XSL•FO
RenderX

clinics maintained handwritten notes, which were stored in file systems [5-8]. With the explosion of electronic technologies boosted in the 20th century, hospitals began using such technologies to store records, reducing physical storage space and improving record retrieval efficiency. Computerization is one of the revolutionary technologies. The concept of electronic health records (EHRs) originated in the 1960s [9]. It aimed to improve the storage, retrieval, and management of patient information. Health care professionals started to address clinical challenges and steer treatment decisions by using data from these systems. Later, the 1990s brought about efforts to standardize EHR systems and the introduction of the internet, which significantly influenced EHR development. The 2010s witnessed widespread adoption of EHRs, driven by government incentives and the recognition of their potential to improve health care quality, safety, and efficiency [10-13]. Now, in the 2020s, the use of EHRs has been boosted with the development of advanced technologies in health care.

Objectives

In the past 25 years, the scope of EHRs varied significantly across the globe. Initially, many EHR systems were developed primarily as billing tools and were not intended to support clinical workflows. However, over time, their functionality expanded to optimize diagnosis and patient care, thereby enhancing their utility and relevance for clinical research. This evolution has transformed EHRs into a valuable resource not only for managing health care delivery but also for facilitating data-driven research efforts. For the 25th anniversary of the *Journal of Medical Internet Research*, this review focused on the evolution of secondary use of EHRs and their interoperability in medical research over the past 25 years. For researchers using EHR data in medical studies, we explored the advancements in EHR technology, emphasizing how they have facilitated better management and understanding of diseases through comprehensive data collection and analysis. In addition, this review highlighted the importance of these records in epidemiological studies, pragmatic clinical trials (PCTs), and health economic studies, particularly in providing a more accurate representation of clinical practices and patient populations.

Methods

Literature Review

We focused on the advancements in EHR technology and the application and significance of these records in medical research across 3 distinct periods. Therefore, we conducted an extensive literature search in the PubMed, Scopus, and Web of Science databases, which provides ideal coverage of the publications of interest. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1).

Our search strategy was divided into three parts: (1) all studies that contained the terms *Electronic health record* or *Electronic medical record* in the title or abstract and *medical research* in all fields were identified; (2) according to the time of publication, the studies were divided into 2000s, 2010s, 2020 to 2023, and 2024; and (3) the studies were then narrowed down to fields by specific terms (Figure 1). For example, the search term for the 2010s was as follows: ((*electronic health record*[Title/Abstract]) OR (*electronic medical record*[Title/Abstract])) AND (*medical research*) AND ((*real world study*) OR (*real world evidence*) OR (*pragmatic clinical trial*) OR (*pragmatic trial*)). We used cross-referencing techniques from the included studies and supplemented our search with forward and backward citation tracking to ensure comprehensive coverage.

XSL•FO
RenderX

Figure 1. Search strategy and results by timeline.

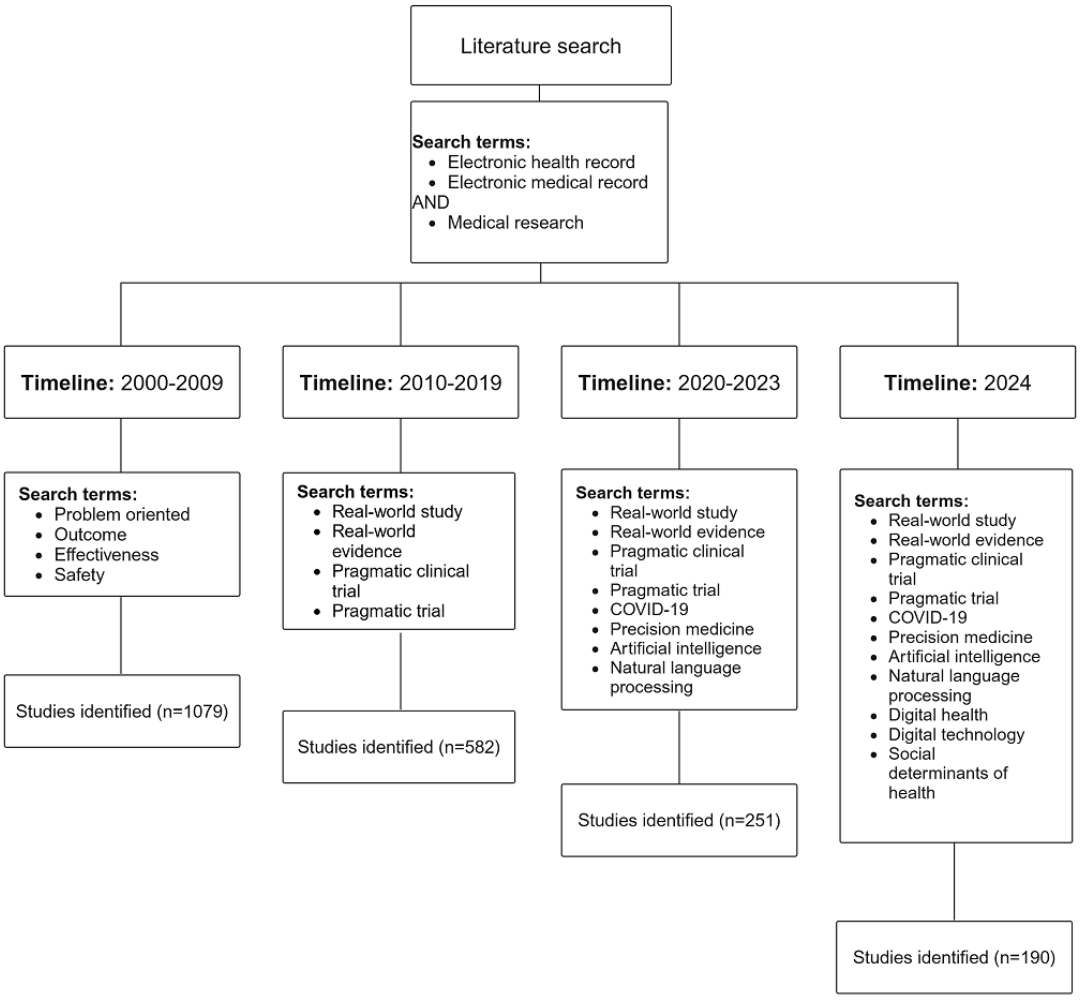

Qualitative Structured Analysis

Qualitative analysis and structuring of the available sources were performed using NVivo (version 14; Lumivero) software. All studies were coded using NVivo to generate keywords.

Results

Eligible Studies

According to the PRISMA framework (Figure 2), articles were searched in PubMed, Scopus, and Web of Science databases and categorized by period. Duplicate records, non-English abstracts, and records unrelated to EHRs were excluded, resulting in a final inclusion of 2102 articles. We also excluded records not pertinent to the application of EHRs in medical research, such as commentary articles and those focused on fraudulent claims.

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

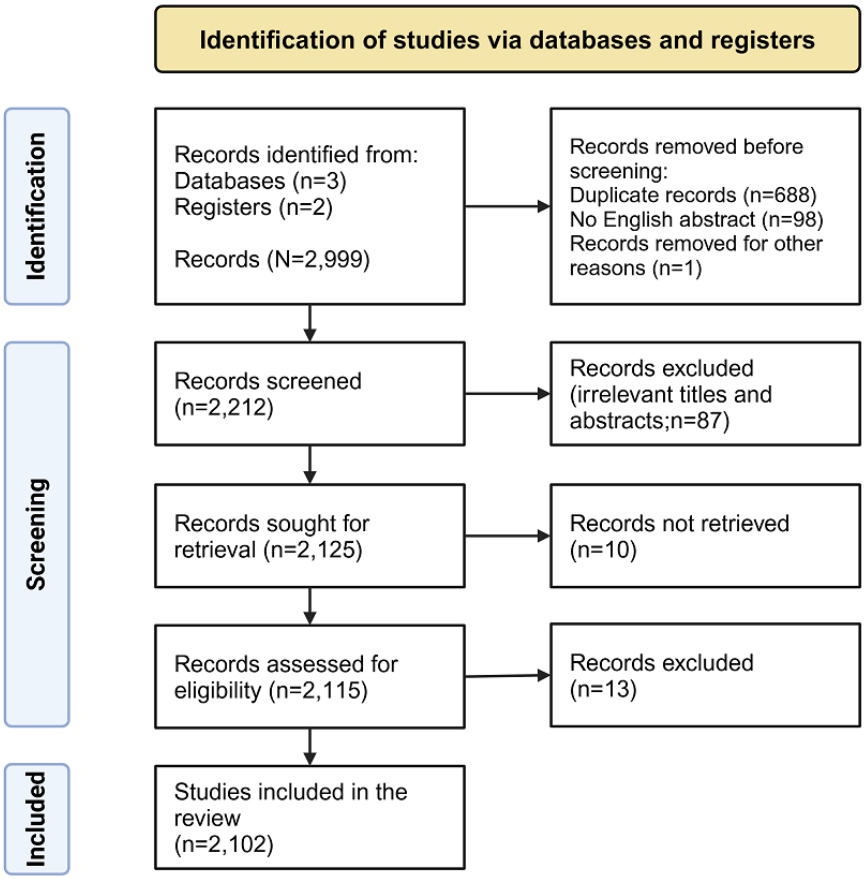

Time Intervals

The included articles highlighted the evolution of EHRs from simple documentation tools to central components of patient care, public health surveillance, and medical research. During the 2000s, researchers used EHRs in various observational studies and health care quality improvement efforts. During this period, early research using EHRs often focused on single health care systems with limited clinical data. The introduction of digital imaging and communications in medicine enabled the use of biomedical image data within EHRs, facilitating clinical follow-ups and case studies. During the 2010s, the EHRs served as primary data sources for many large-scale studies on patient outcomes, disease patterns, and treatment efficacy. Various national registries and research networks emerged to explore disease associations, health risks, and health care interventions. The concept of real-world evidence (RWE) gained attention, with PCTs relying heavily on EHR data to evaluate interventions in real-world settings. In addition, claim data were linked to EHRs for assessing health care services, resource use, and costs. Regulatory agencies began considering RWE in drug approvals and postmarketing surveillance. During the 2020s, retrospective EHR reviews enabled early identification of clinical characteristics, risk factors, and effective interventions, particularly focusing on the global health challenges of COVID-19 pandemic. Comparative effectiveness research using EHR data provided insights into public health measures, testing strategies, and treatment protocols. Globally, countries used their EHR infrastructure for real-time surveillance, vaccine distribution, and health policy adjustments. In addition, EHR data highlighted the impact of social determinants of health (SDOH) on COVID-19 outcomes. In addition, EHRs have played a pivotal role in precision medicine. Artificial intelligence (AI) and natural language processing (NLP) have increasingly been applied to EHRs, enabling predictive analytics for patient care, personalized treatments, and health care operations optimization. Many studies also highlighted the remaining limitations and challenges, including data quality, standardization, interoperability, patient privacy, data security, and high implementation costs. Enhanced regulatory frameworks, data-sharing agreements, and advanced security technologies were recommended to address these issues.

Discussion

Initial Adoption and Early Research Use Before and in the 2000s

The Problem-Oriented Medical Record

One of the earliest known efforts to use electronic systems for medical research that could be considered a precursor to modern EHRs was by Weed in 1968 [14]. Weed [14] introduced the concept of the problem-oriented medical record (POMR), which organized patient data according to specific problems, making it more systematic and usable for both care and research. Researchers began to use POMR for cost analysis [15], staff communications [16], and decision-making [17-19] in a wide range of diseases. The POMR was also strongly advocated among medical students for training purposes [20]. As patient-centered medical care always necessitates a

XSL•FO
RenderX

patient-centered medical record, the POMR remained prevalent in the early 2000s [21-23].

The Regenstrief Institute, established in 1969 in Indianapolis, Indiana, is another pioneer in this field. The institute developed the Regenstrief Medical Record System (RMRS), one of the first electronic medical record systems [24,25]. The RMRS allowed for the electronic capture, storage, and retrieval of patient data, serving as a model for later EHR systems. Researchers began to design randomized trials by using the system to test the effectiveness of medications [26,27]. The RMRS attempted to create a community-wide linkage across health care systems, urgent care hospitals, clinics, and health departments [28]. Both Weed's [14] work on POMR and the Regenstrief Institute's development of the RMRS played significant roles in the early use of electronic systems to support medical research, laying the groundwork for the EHR systems we see today.

Technological and Methodological Developments

With the development of POMR and RMRS, some new technologies and methodologies were applied to support the use of EHRs in research. This included the creation of data warehouses and the development of tools for data mining and analysis, which would later become crucial for extracting meaningful insights from large EHR datasets [29,30]. The recognition of the value of EHR data for research led to the formation of consortia and collaborative networks aimed at sharing data and best practices. These early collaborations laid the foundation for later, more extensive networks that would leverage EHR data for multicenter studies and nationwide research initiatives [31].

In 2009, the Health Information Technology for Economic and Clinical Health Act (HITECH) was passed [12]. The HITECH Act represented a critical step in the digital transformation of the US health care system. The HITECH Act led to a substantial increase in the collection and digitization of patient health information. This vast repository of data became a valuable resource for researchers, offering a wealth of information on patient outcomes, disease patterns, and treatment effectiveness [32-36]. Most research studies in 2000s leveraging EHRs have some features in common, including a single health care system or single hospital, an observational study design, and limited data on clinical characteristics [37]. In addition, some researchers also tried to use EHRs for pharmacovigilance and comparative effectiveness of drugs [38-41]. The limited duration of interventional and observational periods in clinical trials necessitated real-time monitoring of therapeutic effectiveness and safety, crucial for both physicians and patients. The use of EHRs effectively addressed these concerns [42]. Some early explorations on biomedical image data began with the incorporation of the digital imaging and communications in medicine within EHRs [43-46]. The deidentification of radiology clinical data allowed researchers to conduct case series studies and to follow up the changing patterns of the biomedical images in a certain disease area.

The Rise of Big Data in Health Care in the 2010s

In the 2010s, big data analytics in health care were popular, with EHRs serving as the primary data source. Researchers began leveraging EHR data to conduct large-scale studies on patient outcomes, disease patterns, and treatment efficacy. A lot of cohort studies have been designed and conducted using EHR data to provide evidence of the association between risk factors and occurrence as well as the prognosis of diseases. In addition, a lot of quality improvement studies used EHRs to investigate the patients' communication and clinical support. The disease areas covered almost every system in the body, including primary care [47], cardiology [48], sepsis [49], cancer [50], diabetes [51], rheumatology [52], nephrology [53], neurology [54], hepatology [55], and gastroenterology [56]. Our previous review has highlighted the practical use of EHRs among patients with diabetes in medical research [57]. Moreover, insurance claim data, such as that from the Kaiser Permanente [58-60], offer valuable insights for enhancing our understanding of health risks, improving patient outcomes, and informing policy and practice decisions.

In contrast, researchers also started to analyze data from large patient populations to identify health trends, assess risk factors, and develop strategies for disease prevention and health promotion. They started to use these data for potential patient selections and recruitment for clinical trials [61]. Meanwhile, PCTs began to rise gradually [62]. Researchers have made significant efforts to bridge the gap between clinical trial results and real-world outcomes [63-66]. These type of trials were designed to evaluate the effectiveness of interventions or medications in real-world clinical settings, aiming to inform health care decision makers about the trade-offs between benefits, harms, and costs of those interventions [67-69]. They mainly rely on the EHR data. Researchers aimed at addressing key uncertainties in real-world clinical settings, focusing on comparing interventions or medications in ways that reflect usual care [70,71]. There were also some tools that were developed to help researchers design and implement PCTs, including the Pragmatic–Explanatory Continuum Indicator Summary (PRECIS-2) [72]. Moreover, stakeholder engagement is a pivotal step [73,74]. The implementation phase requires selecting appropriate sites that mirror the settings where the intervention will be applied, using strategies for diverse participant recruitment, and integrating interventions and comparators seamlessly into routine clinical practice. Data collection leverages EHR to minimize disruption and enhance the efficiency of the trial process. PCTs seem more reliable to provide RWE than traditional randomized trials. Some claim data linked to EHRs can also provide detailed records of health care services received by patients, allowing researchers to accurately measure resource use and associated costs [75-77]. This information is crucial for cost-effectiveness analyses that compare the costs and outcomes of different health care interventions. EHR data enable the estimation of direct medical costs by documenting clinical procedures, medication prescriptions, hospital stays, and other health care services. In addition, EHR data have contributed to the evaluations of health care quality and system performance, which supported its use

XSL•FO
RenderX

in health policy studies, as the data contain a huge amount of information on health care procedures [78-80].

Regulatory agencies, including the US Food and Drug Administration (FDA) [81] and the European Medicines Agency (EMA) [82], began to explore the potential of RWE from real-world studies of large cohorts or PCTs in supporting drug approvals and postmarketing surveillance. The Patient-Centered Outcomes Research Institute in the United States, which was established in 2010, also reflected a growing recognition of the value of RWE in regulatory decisions and health care policy [83,84]. The FDA and EMA considered approval seriously with evidence from RWE, particularly when traditional clinical trials are not feasible or practical [85]. This is particularly relevant in scenarios where patient populations are rare, heterogeneous, or difficult to recruit into randomized controlled trials. Regulatory agencies, including the FDA and the EMA, have increasingly recognized the value of RWE in regulatory decision-making. The FDA's 21st Century Cures Act has encouraged the use of RWE in supporting the approval of new indications for existing drugs and in fulfilling postmarketing study requirements [85]. A notable example of this is the expanded indication of palbociclib, which was approved by the FDA for use in male patients with hormone receptor–positive, HER2-negative advanced or metastatic breast cancer [86]. In 2020, the FDA approved pembrolizumab for the treatment of adult and pediatric patients with unresectable or metastatic solid tumors that have a high tumor mutational burden [87]. This decision involved RWE derived from genetic sequencing and EHRs. Other examples include the approval of blinatumomab for pediatric acute lymphoblastic leukemia [88], ibrutinib for chronic lymphocytic leukemia [89], nusinersen for spinal muscular atrophy [90], and alosetron for irritable bowel syndrome in female patients [91].

Global Health, Precision Medicine, and AI in the 2020s

COVID-19 Pandemic and Real-Time Surveillance

The global nature of the COVID-19 pandemic increased the importance of international collaboration in medical research. While the health care providers were busy taking actions against the COVID-19 pandemic, they used retrospective EHR to learn lessons and analyze the trends and responses they had made [92-98]. The World Health Organization declared the outbreak a public health emergency of international concern (PHEIC) in early 2020 [99] and ended the PHEIC declaration in 2023 [100]. During the early stages of the COVID-19 pandemic in 2020, Chinese physician-scientists played a crucial role in understanding the disease by conducting retrospective reviews of medical records [93,95-98]. This process was instrumental in identifying the clinical characteristics and transmission patterns of COVID-19 infection, at a time when the world was scrambling for information on how to deal with the novel coronavirus. Importantly, these studies highlighted the disease's high transmissibility, both from symptomatic and asymptomatic carriers, underlining the importance of widespread testing, contact tracing, and quarantine measures to control its spread [101-106]. This group of physician-scientists from Wuhan, China, also contributed to the understanding of risk factors associated with severe outcomes, including age, pre-existing health conditions, and certain demographic factors. Their findings on the efficacy of various public health interventions, such as the use of face masks, social distancing, and the implementation of lockdowns, informed policies not just in China but around the world [107,108]. Furthermore, the early identification of the genome sequence of SARS-CoV-2 virus by researchers facilitated the rapid development of diagnostic tests and the initiation of vaccine development efforts [109-111]. This early sharing of data was pivotal in the global race to develop vaccines and therapeutics against COVID-19.

During the midpandemic period, EHR data became instrumental for comparative effectiveness research, aimed at evaluating and improving health outcomes across various populations and health care systems. This research spanned a broad spectrum of topics crucial to pandemic management, including the effectiveness of personal protective equipment [112-114], the impact of social distancing [115,116], the accuracy and efficiency of different testing strategies [117,118], strategies [117,118], the efficacy and safety of antiviral medications and vaccines [119-123], and the role of health supplements [124]. In addition, it explored the role of telemedicine in sustaining patient care and the importance of primary care services during the COVID-19 pandemic [115,125-128]. For example, comparative effectiveness research helped identify the types of personal protective equipment that were most effective in preventing virus transmission among health care workers and the public [129,130]. Similarly, studies on social distancing and lockdown measures offered evidence-based guidance on how to balance public health concerns with economic and social implications [131]. The investigation into testing strategies elucidated the strengths and weaknesses of various diagnostic approaches, leading to improvements in testing strategies and the quicker identification of COVID-19 cases [132]. Research on antiviral drugs, anti-inflammatory drugs, and some health supplements informed treatment protocols, contributing to the development of therapeutic strategies that reduced the severity and duration of the disease [133,134]. Telemedicine emerged as a vital tool for providing continuous care while minimizing the risk of virus transmission, demonstrating its potential to revolutionize health care delivery beyond the pandemic [135,136].

In addition to China, in the United States, EHR systems also played a crucial role in the national response to COVID-19. Through partnerships with health care systems and EHR vendors, the Centers for Disease Control and Prevention, the US FDA's Sentinel System and other public health authorities gained access to near–real-time data on COVID-19 cases, hospitalizations, ventilator use, and patient outcomes [137]. EHRs were also instrumental in tracking vaccine distribution and administration. The Vaccine Adverse Event Reporting System used EHR data to monitor postvaccination side effects, guiding policy adjustments in vaccination campaigns [138]. In the United Kingdom, the government leveraged its National Health Service (NHS) EHR infrastructure for real-time surveillance and policy making during the COVID-19 pandemic. The United Kingdom developed the OpenSAFELY platform, a secure analytics environment that allowed researchers to access deidentified EHR data of >58 million NHS patients [139]. EHR

XSL•FO
RenderX

data were used to identify individuals with high risk and provided real-time data on hospital admissions, intensive care unit (ICU) bed occupancy, and patient outcomes. South Korea used a robust real-time surveillance system that integrated EHRs with other health data sources to contain the spread of COVID-19, including contact tracing and national health reporting [140]. In Australia, the national My Health Record system was used to support the country's pandemic response with centralized health data access and the implementation of lockdowns and the reopening of services [141]. Germany also used EHR data to manage its health care resources effectively during the COVID-19 pandemic, which used EHR data to monitor ICU capacity across the country. Hospitals provided daily updates on bed occupancy, ventilator availability, and staffing levels [142]. In addition, Germany explored the use of EHRs to develop digital health passports that recorded patients' vaccination status and COVID-19 test results [143]. The use of EHRs for real-time surveillance and policy making during COVID-19 highlighted the importance of interoperable systems. Countries with centralized health records or integrated health information exchanges were better positioned to use EHR data quickly and effectively.

The COVID-19 pandemic has also accentuated the critical role of SDOH, revealing stark disparities in disease outcomes across different populations [144]. EHRs have provided a rich dataset for analyzing these disparities, uncovering the disproportionate burden of COVID-19 on specific groups such as the older adults, individuals with chronic diseases, and those with autoimmune diseases. Studies leveraging EHR data have identified various social determinants that contribute to the increased vulnerability of these populations. Factors such as socioeconomic status, access to health care, living conditions, and occupational hazards have been linked to higher rates of infection, severe disease outcomes, and mortality [145-151]. For instance, individuals in lower socioeconomic brackets often face barriers to accessing health care services, leading to delayed or inadequate treatment for COVID-19 [152]. Similarly, patients with pre-existing chronic conditions such as diabetes, cardiovascular diseases, and respiratory diseases have been shown to experience more severe outcomes when infected with the virus [105,153]. Furthermore, the research has highlighted the importance of addressing these social determinants in pandemic response strategies. Interventions targeting SDOH can help mitigate the adverse outcomes of COVID-19 among populations considered vulnerable. This included improving access to health care, enhancing support for individuals with chronic diseases, ensuring the availability of protective measures for individuals classified as immunocompromised, and addressing the broader socioeconomic and environmental factors that contribute to health disparities [154].

Even after the end of the PHEIC declaration, the use of EHR in studying post–COVID-19 condition continued to be of paramount importance. Research into the long-term impacts of COVID-19, particularly on mental health [155-158] and cognitive functions [159-161], remains a critical area of public health inquiry. EHRs were expected to provide invaluable, real-time data that can link early exposure to COVID-19 with outcomes that may manifest decades later or even across a lifetime. As humanity progresses, EHRs stand as a vital resource, equipping us to face global crises and unpredictable challenges with resilience and informed action.

Precision Medicine and Genomics

Precision medicine represents a transformative approach to health care in the past 10 years, focusing on the customization of treatment based on an individual's genetic makeup, environment, and lifestyle [162]. EHRs played a pivotal role in the advancement of precision medicine, where treatments are tailored to the individual characteristics of each patient. The integration of genomics data with EHRs allowed researchers to study the genetic bases of diseases and to personalize health care at an unprecedented level. However, integration is still facing several challenges and opportunities. While the routine clinical application of genomic data is still emerging, some EHR-connected biobanks and initiatives are actively addressing the challenges involved. While the EHR-connected biobanks were initially designed for research purposes, it has actively promoted to incorporate the biobank data into the EHR systems.

The UK Biobank (UKB) stands as a prominent figure in the realm of biomedical research, featuring an expansive database and research resource. It encompasses deidentified genetic, lifestyle, and health information, along with biological samples from half a million UK participants [163]. This rich dataset is further augmented by linkage to the United Kingdom's EHRs, such as the Hospital Episode Statistics and the Patient Episode Databases [164,165]. The UKB's comprehensive dataset has been instrumental in propelling forward the fields of medicine and research, offering profound insights into the prevention, diagnosis, and treatment of numerous serious and life-threatening diseases. The UKB has released a set of whole genome sequencing data. Since 2018, numerous studies have used these data, along with Mendelian randomization, to investigate the causal relationships between genetic backgrounds and diseases [166-169]. While the UKB serves as a notable instance of linking a study cohort with EHRs, it does not exemplify direct integration. Despite encompassing data from half a million individuals, the UKB faces potential selection bias [170]. This is attributed to the fact that the initiative dispatched >9 million invitations across the nation, suggesting that the participants may not fully represent the broader population.

Similar to the approach of the UKB, the National Institutes of Health in the United States initiated the All of Us Research Program in 2018, aiming to enroll more than a million individuals [171]. By January 1, 2024, the program had successfully registered >760,000 participants, securing access to >539,000 samples for genomic sequencing and 420,000 EHRs. Through its comprehensive data collection, the All of Us Research Program seeks to address critical gaps in medical research, particularly in areas underserved by existing studies, thereby fostering a deeper understanding of various health conditions and improving strategies for prevention, diagnosis, and treatment across the spectrum of both common diseases and rare diseases. However, it remains a strong example of linking cohort data with EHRs.

XSL•FO
RenderX

Other local or nationwide biobanks with linkage to EHRs or national registries include the Danish Biobank Register [172]; BioBank Japan [173]; China Kadoorie Biobank [174]; Kaiser Permanente Research Bank [175]; deCODE Genetics [176]; FinnGen biobank [177]; Vanderbilt University's bioVU [178]; Michigan Genomics Initiative [179]; BioMe biobank [180]; the Estonian biobank [181]; and the Norwegian Mother, Father and Child Cohort Study [182]. These initiatives varied in scope and scale but collectively contributed to the growing field of genomic research linked to EHRs, facilitating advances in personalized medicine and the understanding of genetic influences on various diseases.

The true sense of integrating genomic data into EHRs presented challenges such as diverse data structures, storage limitations, interpretation complexities, and financial constraints. The National Human Genome Research Institute supported the Electronic Medical Records and Genomics network to address these issues by harmonizing data and developing guidelines for genomic data integration and use [183]. Solutions are in progress, and the initiative now has >10 clinical site partners, along with a network of affiliates across major areas in the United States. The Electronic Medical Records and Genomics network is not only developing best practices for electronic phenotyping and biobanking consultation but also exploring EHRs for additional genome-wide association studies outcomes among genotyped individuals, leveraging the rich phenotypic data available.

Data Analytics and AI

The use of data analytics and AI in analyzing EHR data has grown exponentially. Researchers were using machine learning algorithms to identify patterns, predict outcomes, and generate insights into disease progression, treatment efficacy, and patient care optimization [184,185]. These technologies enabled the analysis of vast amounts of data in ways that were not previously possible.

Predictive Analytics for Identifying Patients With High Risk and Their Outcomes

One of the most significant applications of AI in health care is in identifying patients at high risk of developing specific conditions and predicting outcomes among patients who were already diagnosed with a health condition. By analyzing historical EHR data, including patient demographics, previous health conditions, treatment histories, and lifestyle factors, AI models can predict which patients are more likely to develop diseases such as diabetes, cardiovascular diseases, or chronic respiratory conditions, as well as the potential outcomes of various treatment options [186-190]. Early identification of these patients allows health care providers to implement preventive measures, tailor treatments to individual needs, and in doing so, significantly reduce the risk of severe outcomes or the need for hospitalization. This predictive capability enables physicians to choose the most effective treatment plans, thereby improving the quality of care and patient satisfaction. In the context of infectious diseases, AI models are invaluable for tracking and forecasting the spread of diseases. By integrating EHR data with information from other sources, such as public health records and social media, AI algorithms can predict outbreaks, model the spread of infections, and inform public health interventions [191,192]. This was particularly evident during the COVID-19 pandemic, where AI played a pivotal role in predicting hospitalization rates, identifying potential hot spots, and guiding the allocation of medical resources [193,194].

Personalized Medications and Clinical or Regulatory Decision Support Systems

Leveraging EHR data, AI is facilitating personalized medicine by tailoring treatment plans to individual patient characteristics, genetic information, and historical health data. This approach helps in predicting how different patients will respond to various treatments, thereby improving treatment effectiveness and reducing side effects [195]. While medications only serve one part of the disease management, AI integrated with EHR systems can provide clinicians with decision support by offering diagnostic suggestions, recommending treatment options, and alerting health care providers to potential adverse drug interactions or necessary screenings, thereby enhancing patient care quality [196-200]. AI tools also analyze EHR data to optimize hospital operations and predict perioperative complications. This can lead to improved patient satisfaction, lower surgery-related adverse events, and lower health care costs [201-205]. Among the AI strategies, NLP technologies represent a pivotal advancement in the use of EHR data, transforming the landscape of health care data analytics and decision-making [185]. NLP enables the extraction of valuable information from unstructured text within EHRs, such as clinical observations, patient histories, and diagnostic notes, which traditionally have been difficult to analyze systematically due to their unstructured nature. This capability not only enhances the richness of data available for clinical decision-making but also significantly expands the potential for personalized patient care and advanced health care analytics [206-209]. One notable example of NLP's application in health care is the development of GatorTron (University of Florida), a large clinical language model trained on >90 billion words from clinical notes, PubMed articles, and Wikipedia [210-212]. GatorTron significantly outperforms previous models in clinical NLP tasks such as clinical concept extraction, medical relation extraction, semantic textual similarity, natural language inference, and medical question answering. This model showcases how scaling up the number of parameters and data size can lead to substantial improvements in performance across various NLP tasks, ultimately enhancing the extraction and interpretation of critical information from EMR data for better health care delivery and research.

EHRs in Medical Research in the Next Decades

The next decades will likely see significant advancements in EHR capabilities, focusing on interoperability, precision medicine, AI and digital health technologies, the integration of SDOH, public health monitoring, and the enhancement of data analytic tools (Figure 3).

XSL•FO
RenderX

Figure 3. Future landscape of electronic health records.

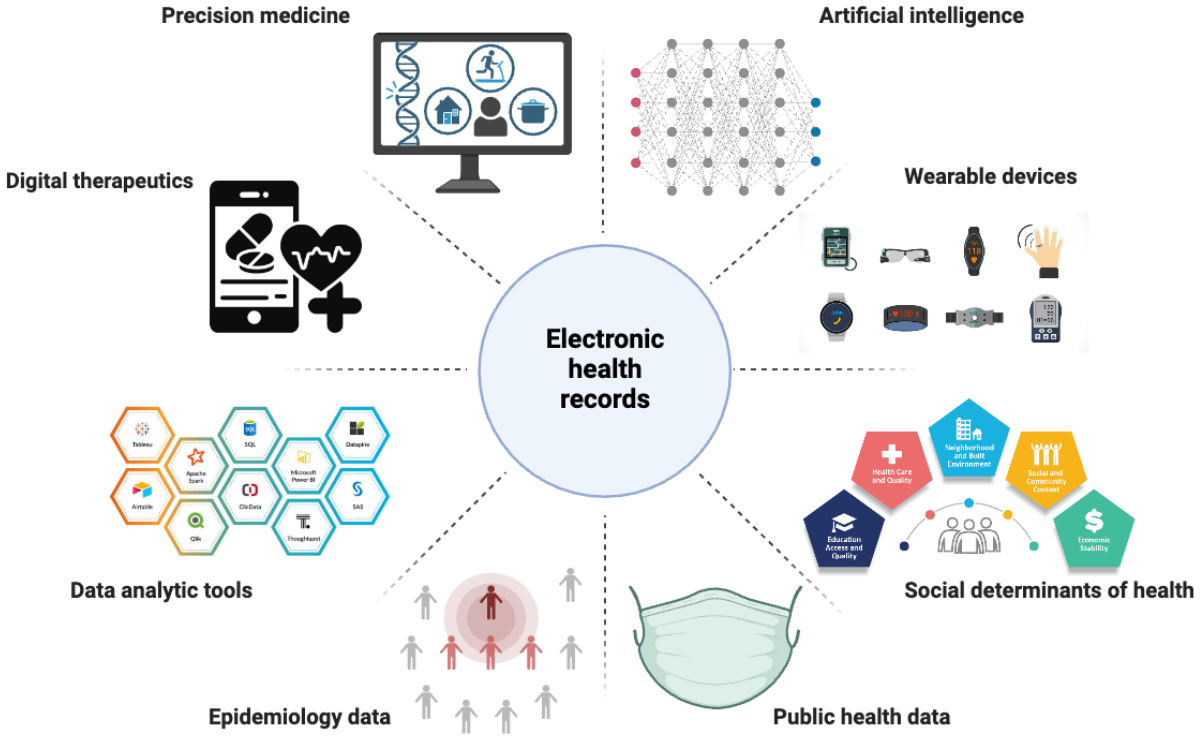

EHRs for Patient Care Versus EHRs for Research

It has been a trend to develop an EHR system for research use only. EHRs designed for patient care and those designed for research may differ in their structure, focus, and functionality [213]. EHRs tailored for patient care will prioritize real-time clinical decision support, patient safety, and continuity of care. They are built to store detailed patient information, such as medical history, medications, allergies, laboratory results, and diagnostic imaging, allowing health care providers to quickly access and update data during clinical encounters. The interface is designed for ease of use. In contrast, EHRs designed for research will focus on data collection, aggregation, and analysis at a population level. They emphasize data standardization, coding, and interoperability, allowing researchers to extract large datasets for epidemiological studies, clinical trials, and public health surveillance [214]. These EHRs will include extensive metadata and use structured data formats to facilitate the analysis of trends and outcomes. While patient care EHRs focus on the individual patient, research-oriented EHRs aggregate deidentified data from diverse populations to explore disease patterns, treatment effectiveness, and health care disparities. In addition, research EHRs must also address data privacy and consent issues more rigorously to comply with ethical standards for data use [214].

Enhanced Interoperability and Data Sharing

Future EHR systems are likely to achieve higher levels of interoperability, enabling seamless data sharing across different health care systems, health care providers, and even countries. This will facilitate comprehensive patient care, collaborative research, and public health monitoring on a national or even a global scale. Regulations governing the use of EHRs and related technologies will evolve to keep pace with advancements, addressing ethical considerations, data ownership, and the use of digital health data in research and health care delivery. As the use of EHRs grows, the focus on protecting patient data will also be important. Advanced security technologies and privacy-preserving methods, such as blockchain and federated learning, may be used to safeguard information while enabling its use for care and research.

As a good example, the CODE-EHR framework represents a crucial advancement, focusing on improving the design of studies that use structured EHR records [215]. This framework emphasizes the importance of verification, validation, and data privacy, along with addressing social mandates to undertake research. It proposes minimum standards to enhance study transparency and provides a road map toward more effective use of health care data for research. By enabling the embedding of controlled trials within real-world settings, such as registries or routine clinical practice, CODE-EHR aims to produce more generalizable results that can inform patient care, health care cost containment, and quality of life improvements.

Integration of Genomic and Other Omics-Related Precision Medicine Data

As genomic sequencing becomes more affordable and commonplace, EHRs are expected to integrate multiomics information, supporting the advancement of precision medicine. This will allow health care providers to tailor treatments to individual patients based on their individualized profiles, improving outcomes and minimizing adverse reactions [216].

The PennChart Genomics Initiative has demonstrated significant progress and potential in integrating genomic data into EHRs for optimizing patient care and facilitating medical research [217,218]. By creating a centralized document display for

XSL•FO
RenderX

genetic results in the EHR, researchers and clinicians can improve patient care while addressing privacy concerns. This initiative has already led to the integration of thousands of genetic documents into the EHR, highlighting the potential for EHRs to support precision medicine and accelerate the discovery of novel genomic medicine applications. Future efforts aim to expand this integration and address challenges such as educational needs, infrastructure compatibility, and maintaining privacy and security protections.

The Genetics Underlying Diabetes in Hispanics and Latinos study in the United States is a pioneering initiative planning to enroll 100,000 newborns for whole genome sequencing [219]. This program aims to explore the potential of genomic data in newborn screening, with the goal of enhancing early diagnosis and treatment in neonatal ICUs and pediatric ICUs. By identifying genomic markers early on, the program seeks to save lives and reduce hospital stays for patients with undiagnosed disease, showcasing the transformative power of genomics in precision health care.

Genomics England's program is another potential effort to integrate genetic data with EHR data, aiming to sequence 100,000 genomes from NHS patients with rare diseases and their families as well as patients with cancer [220]. Launched to enhance our understanding of genetics in disease and to pioneer a UK genomic medicine service, the project's success has paved the way for further genomic research and integration into health care. This initiative highlights the potential of genomic data in advancing personalized medicine and improving treatment strategies.

Expansion of AI and Other Digital Health Technologies

EHRs in the future will increasingly incorporate AI and machine learning algorithms to analyze health data in real time, predicting patient risks (such as for chronic diseases or hospital readmissions) and suggesting personalized prevention and treatment plans. AI-driven models, using big data from EHRs, outperform traditional methods in sensitivity, specificity, and other metrics. The implementation of these models into web applications and mobile apps could significantly aid clinical decision-making, emphasizing the need for effective strategies to integrate AI algorithms in clinical tools.

The future use of NLP in EHRs is also expected to significantly enhance health care delivery and research [221]. By leveraging NLP, health care professionals can continuously extract valuable insights from the increasing volumes of unstructured data within EHRs. This advancement will allow for more accurate patient stratification, personalized treatment plans, and improved predictive models for disease outcomes in the future. NLP technologies are set to automate the extraction and interpretation of clinical information, reducing manual data entry and analysis burdens. In addition, real-time processing of patient data through NLP will facilitate more rapid clinical decision support, optimizing treatment approaches and patient care pathways [222]. The integration of NLP with EHR systems promises a shift toward more data-driven, efficient, and patient-centered health care, highlighting its critical role in advancing medical research, enhancing disease surveillance, and ultimately improving patient outcomes.

In addition, EHRs will also become more integrated with other digital health technologies, such as wearable devices, telehealth platforms, and digital therapeutics [223,224]. Wearable devices, ranging from fitness trackers to advanced biosensors, have become increasingly sophisticated, capable of monitoring a wide array of health metrics such as heart rate, activity levels, sleep patterns, and blood glucose levels. Integrating these devices with EHRs can provide clinicians with a more comprehensive view of a patient's health status, extending beyond the snapshots captured during clinical visits. This continuous stream of real-time data can enhance preventive care strategies, enable early detection of potential health issues, and support personalized treatment plans. For patients with chronic conditions, this integration can facilitate closer monitoring and adjustments to treatment regimens based on near–real-time data, potentially improving outcomes and patient engagement in their care. Telehealth platforms have experienced exponential growth, a trend significantly accelerated by the COVID-19 pandemic. The integration of telehealth with EHRs ensures that online consultations are informed by the patient's medical history, current medications, and recent test results, allowing for more informed clinical decision-making [225]. This seamless flow of information can improve the efficiency and effectiveness of virtual care, making health care more accessible, particularly for individuals in remote or underserved regions. Furthermore, integrating telehealth encounters into EHRs ensures that all aspects of a patient's care are documented in a single, comprehensive record, supporting continuity of care across different health care settings. Digital therapeutics, which include software-driven evidence-based interventions, are an emerging field within health care [226]. These interventions, which can range from cognitive behavioral therapy apps for mental health to digital programs for chronic disease management, offer new avenues for treatment that are accessible and scalable. Integrating digital therapeutics with EHRs can enhance the personalization of care, enabling health care providers to track patients' engagement and progress with these interventions in parallel with traditional treatments. This integration can also facilitate the collection of outcome data, supporting the evaluation of the effectiveness of digital therapeutics in real-world settings and informing future care decisions.

Incorporation of SDOH

Addressing SDOH is crucial for improving health outcomes and reducing health disparities. Future EHR systems may include more comprehensive data on patients' social, economic, and environmental conditions, which are currently lacking in most systems and platforms [227]. This idea of data integration will empower health care providers to address not just the clinical but also the contextual aspects of health, leading to more targeted interventions, improved health outcomes, and reduced health care disparities. The evolution of EHR systems to include comprehensive SDOH data will necessitate advancements in data collection, privacy protocols, and analytics. Collecting SDOH data in EHRs in the future could involve leveraging technology to automate data capture from various sources, including patient self-reporting through digital platforms, integrating community health data, and using NLP to extract SDOH information from clinical narratives.

XSL•FO
RenderX

Future Use in Public Health and Epidemiology

EHRs will continuously play a vital role in monitoring public health trends due to their performance during the COVID-19 pandemic, managing outbreaks of infectious diseases, and conducting large-scale epidemiological studies in the future in case of other unpredictable pandemics. Real-time data analysis will enable quicker responses to public health emergencies and more effective preventive strategies. However, these potential opportunities will depend mainly on the enhanced interoperability and data sharing, as mentioned previously.

Integration of Data Analytic Tools With EHRs

Integrating data analytics tools with EHRs involves connecting the capabilities of data analysis software with the wealth of patient information stored in EHR systems. This integration can enhance health care delivery, improve patient outcomes, and facilitate health care research by leveraging advanced analytics on EHR data. A recent effort for an integrated online phenomics knowledge base within the EHR data was the development of the Centralized Interactive Phenomics Resource (CIPHER) knowledge base [228,229]. The CIPHER is a comprehensive, public-facing knowledge base designed to streamline the development of clinical phenotypes, aiming at facilitating clinical and health services research. It also featured scalable metadata management, integrated tools, and user workflows as well as enabled complex searches through stored phenotype metadata adhering to the CIPHER standard. Phenotypes can be contributed via a webform that validates metadata, and the platform includes data visualization tools to improve user engagement and accelerate phenotype development. Hopefully, the CIPHER knowledge base can expand the phenotype algorithm repository and the collaborations among users in the future.

Potential Challenges

It also highlighted several challenges in using EHRs for medical research before and in the past 25 years, including issues with data quality, consistency, and interoperability between different EHR systems. Privacy and security concerns regarding the use of patient data for research purposes were significant, leading to regulatory efforts to protect patient information while enabling research. The Health Insurance Portability and Accountability Act (HIPAA), originally enacted in 1996, continued to influence EHR development throughout the 2000s with its privacy and security rules, which were critical in shaping how EHR systems handled sensitive patient data [230,231].

Deidentification, Recoding, Regulatory, and Ethical Considerations

In the clinical care setting, information is recorded into EHRs by various health care professionals, including physicians, nurses, medical assistants, and other clinical staff. The accuracy and completeness of the data entered depend heavily on the documentation practices of these health care providers, which can vary between institutions and countries [232]. In some cases, administrative staff may also input data related to billing and patient demographics. It is important to note that these data inputs are primarily geared toward patient management rather than research, which introduces variations and challenges when using EHR data for secondary research purposes. One of the most crucial steps in preparing EHR data for research is deidentification, which involves removing or masking identifiable patient information (eg, names, addresses, and social security numbers) to protect patient privacy. Data security and privacy concerns often influence how EHR data are accessed and used for research. Institutions often have governance committees or data access boards that review research proposals to ensure they comply with ethical guidelines and data use policies. However, these regulations vary from country to country and system to system. In the United States, HIPAA regulations define strict guidelines for deidentification and the secondary use of health data. Institutions may use institutional review boards to oversee the approval process for research using EHR data [230]. In the European Union, under General Data Protection Regulation, pseudonymization and data minimization are encouraged, and the use of EHR data for research typically requires explicit patient consent unless an exemption is granted [233]. In the United Kingdom, the NHS has implemented the use of trusted research environments to facilitate secure access to health data for research while ensuring patient privacy [234].

Given the differences in regulations across countries, researchers must familiarize themselves with the specific legal and ethical requirements of the country where the EHR data are stored and where the research will be conducted. When conducting international or multicountry research, researchers need to consider how differing regulations affect data sharing, processing, and consent. Cross-border collaboration may require legal agreements, such as data transfer agreements, that meet all involved jurisdictions' requirements. Regardless of the region, maintaining high standards for data security and privacy is crucial. Researchers should use practices such as data pseudonymization, anonymization, encryption, and secure data storage to protect patient information throughout the research process [235].

Despite the progress, researchers continued to face challenges with the interoperability of EHR systems. The ability to share and integrate data across different platforms and health care institutions was still limited, posing hurdles to multisite studies and collaborative research efforts. Addressing these challenges became a priority for advancing the use of EHRs in research. As the use of EHR data in research grew, concerns about patient privacy and data security also increased. Increased attention to regulatory and ethical issues surrounding the use of health information came up. Policies and guidelines, such as the HIPAA in the United States and the General Data Protection Regulation in the European Union [233,236,237], were updated to ensure the protection of patient information while facilitating research. Despite these concerns, all the efforts during this time underscore a shift toward a more holistic approach to evidence generation, aiming to improve patient outcomes and health care decision-making by bridging the gap between research such as clinical trials and real-world practice.

Meanwhile, researchers also encountered challenges during the use of EHRs on the common data models and data cleaning complexities. Although the Patient-Centered Outcomes Research Institute has made great efforts to this issue by establishing the PCORnet common data model [238-240] and the Observational

XSL•FO
RenderX

Health Data Sciences and Informatics consortium has developed the Observational Medical Outcomes Partnership Common Data Model [241-243], significant challenges remain in achieving interoperability between different EHR systems and health care databases. This inconsistency can hinder the seamless exchange and integration of health data for comprehensive analyses. Common data models rely on the assumption that data across systems can be standardized into a uniform format. However, variations in how data are recorded, missing information, and inconsistent use of terminologies across different EHR systems can lead to gaps and inconsistencies that affect data quality and utility. For the concerns of data cleaning and linkage, although several standardized coding systems (ie, International Classification of Diseases, Current Procedural Terminology, Healthcare Common Procedure Coding System, Logical Observation Identifiers Names and Codes, Systematized Nomenclature of Medicine, etc] were introduced in the EHRs, they still contain a vast amount of unstructured or semistructured data, including free-text clinical notes [244]. Cleaning such data to extract meaningful information may require other tools and algorithms, such as SQL environment [MySQL version 8.4.0; Oracle Corporation] and Python [version 3.8.5; Python Software Foundation]), which can be resource intensive and complex to develop and apply for researchers without computer science backgrounds. Patients' data are often shared among different EHR systems, including primary care and specialty registries. Linking these fragmented data sources to achieve a holistic view of a patient's health history is challenging but essential for effective research and care management.

Data Quality, Standardization, Patient Engagement, Cost, Data Cleaning, and Global Health Disparities

Ensuring high-quality, standardized data entry into EHRs is challenging yet essential for effective health care delivery and research. Inconsistent data, incomplete records, and variation in data capture practices can compromise the reliability and usefulness of EHR data for clinical decision-making and research purposes. While patient portals and digital access to health records aim to empower patients, engaging patients effectively remains a challenge. Issues such as digital literacy, accessibility, and varying levels of interest in active health management can affect the extent to which patients use EHRs to participate in their care [245].

For many health care providers, particularly smaller practices and those in low-resource settings, the cost of EHR implementation and ongoing maintenance is a significant barrier [246,247]. In addition, the process of transitioning from paper-based systems to EHRs, or switching between EHR systems, can be disruptive and resource intensive. Global disparities in EHR adoption and use highlight the challenge of ensuring equitable access to the benefits of digital health records. Differences in infrastructure, funding, and health care policies contribute to varying levels of EHR integration and use worldwide. Integrating AI and advanced analytics into EHR systems presents both opportunities and challenges. While these technologies have the potential to transform health care delivery and research, they also raise questions about data accuracy, algorithmic bias, and the interpretability of AI-generated insights.

Limitations

Despite extensive literature searches across three databases and rigorous literature screening processes, our systematic review still faces some limitations. First, publication bias remains a concern, as studies reporting positive or significant findings are more likely to be published and thus overrepresented in our review. In addition, studies on EHR are highly heterogeneous. Variations in study design, populations, interventions, outcomes, and methodologies complicate meaningful synthesis and generalization of findings. Finally, some relevant studies may be unpublished, inaccessible, or have incomplete data, resulting in restricting data collection and limiting the overall comprehensiveness of our review.

Conclusions

The evolution of EHRs marks a significant milestone in health care's journey toward integrating technology and medicine. From early documentation practices to the sophisticated use of AI and big data analytics today, EHRs have become central to improving patient care, enhancing public health surveillance, and advancing medical research. As we look forward, the integration of SDOH, genomic data, and real-time analytics into EHRs promises to further personalize medicine, improve disease prevention strategies, and address health disparities. Collaborative efforts across borders, disciplines, and sectors will be crucial in realizing the full potential of EHRs in shaping a healthier future for all.

Acknowledgments

GH is partially supported by the grant from the National Institute of General Medical Sciences (U54GM104940) of the National Institutes of Health. YS and GH are partially supported by a grant from the Public University Partnership Program at the Louisiana Department of Health Bureau of Health Services Financing.

Authors' Contributions

YS and JY conducted the search and wrote the manuscript. JZ and GH performed the literature review and edited the manuscript. JZ is the guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analyses. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 77 KB - jmir_v27i1e59024_app1.docx]

References

1. Hu S. A contribution to our knowledge of Leonurus L., i-mu-ts'ao, the Chinese motherwort. Am J Chin Med (Gard City N Y) 1976 Apr 05;4(3):219-237. [doi: 10.1142/s0192415x76000305] [Medline: 970356]
2. Hajar R. The air of history early medicine to galen (part I). Heart Views 2012;13(3):120-128. [doi: 10.4103/1995-705x.102164]
3. Lorkowski J, Pokorski M. Medical records: a historical narrative. Biomedicines 2022 Oct 17;10(10):2594 [FREE Full text] [doi: 10.3390/biomedicines10102594] [Medline: 36289856]
4. Keyser PT, Scarborough J. Oxford Handbook of Science and Medicine in the Classical World. Oxford, UK: Oxford University Press; 2018.
5. Guénel A. The creation of the first overseas Pasteur Institute, or the beginning of Albert Calmette's Pastorian career. Med Hist 1999 Jan 16;43(1):1-25 [FREE Full text] [doi: 10.1017/s0025727300064693] [Medline: 10885131]
6. Blevins SM, Bronze MS. Robert Koch and the 'golden age' of bacteriology. Int J Infect Dis 2010 Sep;14(9):e744-e751 [FREE Full text] [doi: 10.1016/j.ijid.2009.12.003] [Medline: 20413340]
7. Donahue MP. Nursing, the Finest Art: An Illustrated History. Maryland Heights, MI: Mosby; 1996.
8. Gunn SW, Masellis M. Concepts and Practice of Humanitarian Medicine. New York, NY: Springer-Verlag; 2008.
9. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016 May 20;Suppl 1(Suppl 1):S48-S61 [FREE Full text] [doi: 10.15265/IYS-2016-s006] [Medline: 27199197]
10. Buntin MB, Jain SH, Blumenthal D. Health information technology: laying the infrastructure for national health reform. Health Aff (Millwood) 2010 Jun;29(6):1214-1219. [doi: 10.1377/hlthaff.2010.0503] [Medline: 20530358]
11. Conde C. Mum's the word: feds are serious about protecting patients' privacy. Tex Med 2010 Aug 01;106(8):49-53. [Medline: 20680898]
12. Goldstein MM, Thorpe Jane H. The first anniversary of the health information technology for economic and clinical health (HITECH) act: the regulatory outlook for implementation. Perspect Health Inf Manag 2010 Sep 01;7(Summer):1c [FREE Full text] [Medline: 20808607]
13. Kocher R, Emanuel EJ, DeParle NA. The Affordable Care Act and the future of clinical medicine: the opportunities and challenges. Ann Intern Med 2010 Oct 19;153(8):536-539 [FREE Full text] [doi: 10.7326/0003-4819-153-8-201010190-00274] [Medline: 20733178]
14. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-7 concl. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
15. Esterhay RJJ, Vogel VG, Fortner CL, Shapiro HM, Wiernik PH. Cost analysis of leukemia treatment: a problem-oriented approach. Cancer 1976 Feb;37(2):646-652. [doi: 10.1002/1097-0142(197602)37:2<646::aid-cncr2820370206>3.0.co;2-l] [Medline: 814993]
16. Dinsdale SM, Gent M, Kline G, Milner R. Problem oriented medical records: their impact on staff communication, attitudes and decision making. Arch Phys Med Rehabil 1975 Jun;56(6):269-274. [Medline: 1079719]
17. Liston EH. Use of problem-oriented medical record in psychiatry: a survey of university-based residency training programs. Am J Psychiatry 1976 Jun;133(6):700-703. [doi: 10.1176/ajp.133.6.700] [Medline: 1084105]
18. Best R, Bowen G, Magill K, Morrell C. POMR for operating and recovery rooms. Superv Nurse 1976 Aug;7(8):18-22. [doi: 10.1097/00006247-197608000-00003] [Medline: 59951]
19. Switz DM. The problem-oriented medical record: evaluation and management of anemia before and during use. Arch Intern Med 1976 Oct 01;136(10):1119-1123. [doi: 10.1001/archinte.1976.03630100031010]
20. Burroughs AK. Medical students, medical records and clinical skills. Med Educ 1978 Nov;12(6):446-452. [doi: 10.1111/j.1365-2923.1978.tb01426.x] [Medline: 310505]
21. De Clercq E, Van Casteren V, Jonckheer P, Burggraeve P, Lafontaine MF, Degroote K, et al. Are problem-oriented medial records (POMR) suitable for use in GPs' daily practice? Stud Health Technol Inform 2007;129(Pt 1):68-72. [Medline: 17911680]
22. Carey IM, Cook DG, De Wilde S, Bremner SA, Richards N, Caine S, et al. Implications of the problem orientated medical record (POMR) for research using electronic GP databases: a comparison of the Doctors Independent Network Database (DIN) and the General Practice Research Database (GPRD). BMC Fam Pract 2003 Sep 30;4(1):14 [FREE Full text] [doi: 10.1186/1471-2296-4-14] [Medline: 14516473]
23. Donnelly WJ. Viewpoint: patient-centered medical care requires a patient-centered medical record. Acad Med 2005 Jan;80(1):33-38. [doi: 10.1097/00001888-200501000-00009] [Medline: 15618089]
24. McDonald CJ, Tierney WM, Martin DK, Overhage JM, Day Z. The Regenstrief medical record: 1991 a campus-wide system. Proc Annu Symp Comput Appl Med Care 1991:925-928 [FREE Full text] [Medline: 1807754]

XSL•FO
RenderX

25. McDonald CJ, Overhage JM, Tierney WM, Dexter PR, Martin DK, Suico JG, et al. The Regenstrief Medical Record System: a quarter century experience. Int J Med Inform 1999 Jun;54(3):225-253. [doi: 10.1016/s1386-5056(99)00009-x] [Medline: 10405881]
26. Murray MD, Tierney WM, Brater DC. Determining the effectiveness of torasemide and furosemide in heart failure: design of a randomised comparison using the regenstrief medical record system. Clin Drug Investig 1998;16(1):45-52. [doi: 10.2165/00044011-199816010-00006] [Medline: 18370517]
27. Burton MM, Simonaitis L, Schadow G. Medication and indication linkage: a practical therapy for the problem list? AMIA Annu Symp Proc 2008 Nov 06;2008:86-90 [FREE Full text] [Medline: 18999215]
28. Overhage J, McDonald CJ, Suico JG. The regenstrief medical record system 2000: expanding the breadth and depth of a community wide EMR. Proc AMIA Symp 2000:1173 [FREE Full text] [Medline: 11080027]
29. Friedman CP, Dev P. Education and informatics: it's time to join forces. J Am Med Inform Assoc 1996;3(2):184-185 [FREE Full text] [doi: 10.1136/jamia.1996.96236286] [Medline: 8653454]
30. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 02;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
31. Fernández-Breis JT, Maldonado JA, Marcos M, Legaz-García MC, Moner D, Torres-Sospedra J, et al. Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. J Am Med Inform Assoc 2013 Dec;20(e2):e288-e296 [FREE Full text] [doi: 10.1136/amiajnl-2013-001923] [Medline: 23934950]
32. Slutsky JR, Clancy CM. Patient-centered comparative effectiveness research: essential for high-quality care. Arch Intern Med 2010 Mar 08;170(5):403-404. [doi: 10.1001/archinternmed.2010.5] [Medline: 20212173]
33. Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. J Med Internet Res 2010 May 25;12(2):e14 [FREE Full text] [doi: 10.2196/jmir.1356] [Medline: 20501431]
34. Office of the National Coordinator for Health Information Technology, Department of Health and Human Services. Establishment of the temporary certification program for health information technology. Final rule. Fed Regist 2010 Jun 24;75(121):36157-36209 [FREE Full text] [Medline: 20575184]
35. Epstein RS, Teagarden JR. Comparative effectiveness research and personalized medicine: catalyzing or colliding? Pharmacoeconomics 2010;28(10):905-913. [doi: 10.2165/11535830-000000000-00000] [Medline: 20831298]
36. Kush RD, Helton E, Rockhold FW, Hardison CD. Electronic health records, medical research, and the tower of babel. N Engl J Med 2008 Apr 17;358(16):1738-1740. [doi: 10.1056/nejmsb0800209]
37. Pearson JF, Brownstein CA, Brownstein JS. Potential for electronic health records and online social networking to redefine medical research. Clin Chem 2011 Feb;57(2):196-204 [FREE Full text] [doi: 10.1373/clinchem.2010.148668] [Medline: 21159898]
38. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA 1997;277(4):301-306. [Medline: 9002492]
39. Bourgeois FT, Mandl KD, Valim C, Shannon MW. Pediatric adverse drug events in the outpatient setting: an 11-year national analysis. Pediatrics 2009 Oct;124(4):e744-e750 [FREE Full text] [doi: 10.1542/peds.2008-3505] [Medline: 19786435]
40. Dore DD, Seeger JD, Arnold Chan K. Use of a claims-based active drug safety surveillance system to assess the risk of acute pancreatitis with exenatide or sitagliptin compared to metformin or glyburide. Curr Med Res Opin 2009 Apr 12;25(4):1019-1027. [doi: 10.1185/03007990902820519] [Medline: 19278373]
41. Brownstein JS, Sordo M, Kohane IS, Mandl KD. The tell-tale heart: population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction. PLoS One 2007 Sep 05;2(9):e840 [FREE Full text] [doi: 10.1371/journal.pone.0000840] [Medline: 17786211]
42. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med 2010 Nov 02;153(9):600-606. [doi: 10.7326/0003-4819-153-9-201011020-00010] [Medline: 21041580]
43. Kuzmak PM, Dayhoff RE. The use of Digital Imaging and Communications in Medicine (DICOM) in the integration of imaging into the electronic patient record at the Department of Veterans Affairs. J Digit Imaging 2000 May;13(S1):133-137. [doi: 10.1007/bf03167644]
44. Noumeir R, Lemay A, Lina JM. Pseudonymization of radiology data for research purposes. J Digit Imaging 2007 Sep;20(3):284-295 [FREE Full text] [doi: 10.1007/s10278-006-1051-4] [Medline: 17191099]
45. Nagy M, Hanzlícek P, Precková P, Ríha A, Dioszegi M, Seidl L, et al. Semantic interoperability in Czech healthcare environment supported by HL7 version 3. Methods Inf Med 2018 Jan 17;49(02):186-195. [doi: 10.3414/me09-02-0018]
46. Soriano E, Plazzotta F, Campos F, Kaminker D, Cancio A, Aguilera Díaz J, et al. Integration of healthcare information: from enterprise PACS to patient centered multimedia health record. Stud Health Technol Inform 2010;160(Pt 1):126-130. [Medline: 20841663]
47. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]

XSL•FO
RenderX

48. Rubbo B, Fitzpatrick NK, Denaxas S, Daskalopoulou M, Yu N, Patel RS, UK Biobank Follow-up and Outcomes Working Group, et al. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: a systematic review and recommendations. Int J Cardiol 2015 May;187:705-711 [FREE Full text] [doi: 10.1016/j.ijcard.2015.03.075] [Medline: 25966015]
49. Shankar-Hari M, Phillips GS, Levy ML, Seymour CW, Liu VX, Deutschman CS, et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 2016 Feb 23;315(8):775-787 [FREE Full text] [doi: 10.1001/jama.2016.0289] [Medline: 26903336]
50. Bush RA, Pérez A, Baum T, Etland C, Connelly CD. A systematic review of the use of the electronic health record for patient identification, communication, and clinical support in palliative care. JAMIA Open 2018 Oct 01;1(2):294-303 [FREE Full text] [doi: 10.1093/jamiaopen/ooy028] [Medline: 30842998]
51. Riordan F, McHugh SM, O'Donovan C, Mtshede MN, Kearney PM. The role of physician and practice characteristics in the quality of diabetes management in primary care: systematic review and meta-analysis. J Gen Intern Med 2020 Jun;35(6):1836-1848 [FREE Full text] [doi: 10.1007/s11606-020-05676-7] [Medline: 32016700]
52. Liu LH, Choden S, Yazdany J. Quality improvement initiatives in rheumatology: an integrative review of the last 5 years. Curr Opin Rheumatol 2019 Mar;31(2):98-108 [FREE Full text] [doi: 10.1097/BOR.0000000000000586] [Medline: 30608250]
53. Tavares MG, Tedesco-Silva Junior H, Pestana JO. Early Hospital Readmission (EHR) in kidney transplantation: a review article. Braz J Nephrol 2020 Jun;42(2):231-237. [doi: 10.1590/2175-8239-jbn-2019-0089]
54. Sajatovic M, Welter E, Tatsuoka C, Perzynski AT, Einstadter D. Electronic medical record analysis of emergency room visits and hospitalizations in individuals with epilepsy and mental illness comorbidity. Epilepsy Behav 2015 Sep;50:55-60 [FREE Full text] [doi: 10.1016/j.yebeh.2015.05.016] [Medline: 26119621]
55. Lu M, Rupp LB, Moorman AC, Li J, Zhang T, Lamerato LE, et al. Comparative effectiveness research of chronic hepatitis B and C cohort study (CHeCS): improving data collection and cohort identification. Dig Dis Sci 2014 Dec 17;59(12):3053-3061 [FREE Full text] [doi: 10.1007/s10620-014-3272-6] [Medline: 25030940]
56. Baldwin-Hunter BL, Knotts RM, Leeds SD, Rubenstein JH, Lightdale CJ, Abrams JA. Use of the electronic health record to target patients for non-endoscopic Barrett's esophagus screening. Dig Dis Sci 2019 Dec 4;64(12):3463-3470 [FREE Full text] [doi: 10.1007/s10620-019-05707-2] [Medline: 31273597]
57. Shen Y, Zhou J, Hu G. Practical use of electronic health records among patients with diabetes in scientific research. Chin Med J (Engl) 2020 May 20;133(10):1224-1230 [FREE Full text] [doi: 10.1097/CM9.0000000000000784] [Medline: 32433055]
58. Reed M, Huang J, Graetz I, Brand R, Hsu J, Fireman B, et al. Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus. Ann Intern Med 2012 Oct 02;157(7):482-489 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00004] [Medline: 23027319]
59. Schmittdiel JA, Raebel MA, Dyer W, Xu S, Goodrich GK, Schroeder EB, et al. Prescription medication burden in patients with newly diagnosed diabetes: a SUrveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM) study. J Am Pharm Assoc (2003) 2014;54(4):374-382 [FREE Full text] [doi: 10.1331/JAPhA.2014.13195] [Medline: 24860866]
60. Coleman KJ, Haneuse S, Johnson E, Bogart A, Fisher D, O'Connor PJ, et al. Long-term microvascular disease outcomes in patients with type 2 diabetes after bariatric surgery: evidence for the legacy effect of surgery. Diabetes Care 2016 Aug;39(8):1400-1407 [FREE Full text] [doi: 10.2337/dc16-0194] [Medline: 27271192]
61. Schreiweis B, Grüber G, Bergh B. First experiences in implementing a software-based support for patient recruitment at Heidelberg university hospital. Stud Health Technol Inform 2012;180:1147-1149. [Medline: 22874382]
62. van Staa TP, Dyson L, McCann G, Padmanabhan S, Belatri R, Goldacre B, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. Health Technol Assess 2014 Jul;18(43):1-146 [FREE Full text] [doi: 10.3310/hta18430] [Medline: 25011568]
63. North CS, Hong BA, Adewuyi SA, Pollio DE, Jain MK, Devereaux R, et al. Hepatitis C treatment and SVR: the gap between clinical trials and real-world treatment aspirations. Gen Hosp Psychiatry 2013 Mar;35(2):122-128. [doi: 10.1016/j.genhosppsych.2012.11.002] [Medline: 23219917]
64. Neuman HB, Greenberg CC. Comparative effectiveness research: opportunities in surgical oncology. Semin Radiat Oncol 2014 Jan;24(1):43-48. [doi: 10.1016/j.semradonc.2013.09.003] [Medline: 24314341]
65. Saeed S, Strumpf EC, Walmsley SL, Rollet-Kurhajec K, Pick N, Martel-Laferrière V, Canadian Co-Infection Cohort Study, et al. How generalizable are the results from trials of direct antiviral agents to people coinfected with HIV/HCV in the real world? Clin Infect Dis 2016 Apr 01;62(7):919-926 [FREE Full text] [doi: 10.1093/cid/civ1222] [Medline: 26743093]
66. Parkinson B, Viney R, Haas M, Goodall S, Srasuebkul P, Pearson SA. Real-world evidence: a comparison of the Australian Herceptin program and clinical trials of trastuzumab for HER2-positive metastatic breast cancer. Pharmacoeconomics 2016 Oct 20;34(10):1039-1050. [doi: 10.1007/s40273-016-0411-2] [Medline: 27207252]
67. Peralta CA, Frigaard M, Rubinsky AD, Rolon L, Lo L, Voora S, et al. Implementation of a pragmatic randomized trial of screening for chronic kidney disease to improve care among non-diabetic hypertensive veterans. BMC Nephrol 2017 Apr 12;18(1):132 [FREE Full text] [doi: 10.1186/s12882-017-0541-6] [Medline: 28399844]

XSL•FO
RenderX

68. Hills T, Semprini A, Beasley R. Pragmatic randomised clinical trials using electronic health records: general practitioner views on a model of a priori consent. Trials 2018 May 16;19(1):278 [FREE Full text] [doi: 10.1186/s13063-018-2658-8] [Medline: 29769088]
69. Pfaff E, Lee A, Bradford R, Pae J, Potter C, Blue P, et al. Recruiting for a pragmatic trial using the electronic health record and patient portal: successes and lessons learned. J Am Med Inform Assoc 2019 Jan 01;26(1):44-49 [FREE Full text] [doi: 10.1093/jamia/ocy138] [Medline: 30445631]
70. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016 Aug 04;375(5):454-463 [FREE Full text] [doi: 10.1056/NEJMra1510059] [Medline: 27518663]
71. Weinfurt KP, Hernandez AF, Coronado GD, DeBar LL, Dember LM, Green BB, et al. Pragmatic clinical trials embedded in healthcare systems: generalizable lessons from the NIH Collaboratory. BMC Med Res Methodol 2017 Sep 18;17(1):144 [FREE Full text] [doi: 10.1186/s12874-017-0420-7] [Medline: 28923013]
72. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015 May 08;350:h2147 [FREE Full text] [doi: 10.1136/bmj.h2147] [Medline: 25956159]
73. Smalley JB, Merritt MW, Al-Khatib SM, McCall D, Staman KL, Stepnowsky C. Ethical responsibilities toward indirect and collateral participants in pragmatic clinical trials. Clin Trials 2015 Oct 15;12(5):476-484 [FREE Full text] [doi: 10.1177/1740774515597698] [Medline: 26374687]
74. Kalkman S, van Thiel GJ, Grobbee DE, Meinecke AK, Zuidgeest MG, van Delden JJ. Stakeholders' views on the ethical challenges of pragmatic trials investigating pharmaceutical drugs. Trials 2016 Aug 22;17(1):419 [FREE Full text] [doi: 10.1186/s13063-016-1546-3] [Medline: 27550379]
75. Wang G, Zhang Z, Ayala C. Hospitalization costs associated with hypertension as a secondary diagnosis among insured patients aged 18-64 years. Am J Hypertens 2010 Mar 01;23(3):275-281. [doi: 10.1038/ajh.2009.241] [Medline: 20010701]
76. Rendas-Baum R, Yang M, Gricar J, Wallenstein GV. Cost-effectiveness analysis of treatments for premenstrual dysphoric disorder. Appl Health Econ Health Policy 2010;8(2):129-140. [doi: 10.2165/11532210-000000000-00000] [Medline: 20175591]
77. McEwen LN, Coelho RB, Baumann LM, Bilik D, Nota-Kirby B, Herman WH. The cost, quality of life impact, and cost-utility of bariatric surgery in a managed care population. Obes Surg 2010 Jul 6;20(7):919-928. [doi: 10.1007/s11695-010-0169-0] [Medline: 20446053]
78. Natarajan K, Stein D, Jain S, Elhadad N. An analysis of clinical queries in an electronic health record search utility. Int J Med Inform 2010 Jul;79(7):515-522 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.03.004] [Medline: 20418155]
79. Persell SD, Kaiser D, Dolan NC, Andrews B, Levi S, Khandekar J, et al. Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system. Med Care 2011 Feb;49(2):117-125. [doi: 10.1097/MLR.0b013e318202913d] [Medline: 21178789]
80. Russ AL, Saleem JJ, Justice CF, Woodward-Hagg H, Woodbridge PA, Doebbeling BN. Electronic health information in use: characteristics that support employee workflow and patient care. Health Informatics J 2010 Dec;16(4):287-305 [FREE Full text] [doi: 10.1177/1460458210365981] [Medline: 21216808]
81. Framework for FDA's real-world evidence program. U.S. Food & Drug Administration. 2018 Dec. URL: https://www.fda.gov/media/120060/download [accessed 2024-12-29]
82. Real-world evidence framework to support EU regulatory decision-making: 2nd report on the experience gained with regulator-led studies from February 2023 to February 2024. European Medicines Agency. URL: https://tinyurl.com/mwju2nzv [accessed 2024-12-26]
83. Fischer MA, Asch SM. The future of the Patient-Centered Outcomes Research Institute (PCORI). J Gen Intern Med 2019 Nov 16;34(11):2291-2292 [FREE Full text] [doi: 10.1007/s11606-019-05324-9] [Medline: 31529376]
84. Ellis LE, Kass NE. How are PCORI-funded researchers engaging patients in research and what are the ethical implications? AJOB Empir Bioeth 2017 Jun 28;8(1):1-10. [doi: 10.1080/23294515.2016.1206045] [Medline: 28949867]
85. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 2018 Sep 04;320(9):867-868. [doi: 10.1001/jama.2018.10136] [Medline: 30105359]
86. Wedam S, Fashoyin-Aje L, Bloomquist E, Tang S, Sridhara R, Goldberg KB, et al. FDA approval summary: palbociclib for male patients with metastatic breast cancer. Clin Cancer Res 2020 Mar 15;26(6):1208-1212. [doi: 10.1158/1078-0432.CCR-19-2580] [Medline: 31649043]
87. Marcus L, Fashoyin-Aje LA, Donoghue M, Yuan M, Rodriguez L, Gallagher PS, et al. FDA approval summary: pembrolizumab for the treatment of tumor mutational burden-high solid tumors. Clin Cancer Res 2021 Sep 01;27(17):4685-4689 [FREE Full text] [doi: 10.1158/1078-0432.CCR-21-0327] [Medline: 34083238]
88. Pulte ED, Vallejo J, Przepiorka D, Nie L, Farrell AT, Goldberg KB, et al. FDA supplemental approval: blinatumomab for treatment of relapsed and refractory precursor B-cell acute lymphoblastic leukemia. Oncologist 2018 Nov;23(11):1366-1371 [FREE Full text] [doi: 10.1634/theoncologist.2018-0179] [Medline: 30018129]
89. de Claro RA, McGinn KM, Verdun N, Lee SL, Chiu HJ, Saber H, et al. FDA approval: ibrutinib for patients with previously treated mantle cell lymphoma and previously treated chronic lymphocytic leukemia. Clin Cancer Res 2015 Aug 15;21(16):3586-3590. [doi: 10.1158/1078-0432.CCR-14-2225] [Medline: 26275952]

XSL•FO
RenderX

90. Gauthier-Loiselle M, Cloutier M, Toro W, Patel A, Shi S, Davidson M, et al. Nusinersen for spinal muscular atrophy in the United States: findings from a retrospective claims database analysis. Adv Ther 2021 Dec;38(12):5809-5828 [FREE Full text] [doi: 10.1007/s12325-021-01938-w] [Medline: 34713391]
91. Brenner DM, Sayuk GS. Current US Food and Drug Administration-approved pharmacologic therapies for the treatment of irritable bowel syndrome with diarrhea. Adv Ther 2020 Jan;37(1):83-96. [doi: 10.1007/s12325-019-01116-z] [Medline: 31707713]
92. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
93. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet 2020 Mar 07;395(10226):809-815 [FREE Full text] [doi: 10.1016/S0140-6736(20)30360-3] [Medline: 32151335]
94. Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. Lancet 2020 Mar 28;395(10229):1039-1046 [FREE Full text] [doi: 10.1016/S0140-6736(20)30528-6] [Medline: 32192580]
95. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. BMJ 2020 Mar 26;368:m1091. [doi: 10.1136/bmj.m1091] [Medline: 32217556]
96. Qiu H, Wu J, Hong L, Luo Y, Song Q, Chen D. Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. Lancet Infect Dis 2020 Jun;20(6):689-696 [FREE Full text] [doi: 10.1016/S1473-3099(20)30198-5] [Medline: 32220650]
97. Lei Z, Cao H, Jie Y, Huang Z, Guo X, Chen J, et al. A cross-sectional comparison of epidemiological and clinical features of patients with coronavirus disease (COVID-19) in Wuhan and outside Wuhan, China. Travel Med Infect Dis 2020 May;35:101664 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101664] [Medline: 32278758]
98. Tan C, Huang Y, Shi F, Tan K, Ma Q, Chen Y, et al. C-reactive protein correlates with computed tomographic findings and predicts severe COVID-19 early. J Med Virol 2020 Jul 25;92(7):856-862 [FREE Full text] [doi: 10.1002/jmv.25871] [Medline: 32281668]
99. COVID-19 Public Health Emergency of International Concern (PHEIC): global research and innovation forum: towards a research roadmap. World Health Organization. 2020 Feb 12. URL: https://tinyurl.com/2vrr996d [accessed 2024-12-29]
100. Burki T. WHO ends the COVID-19 public health emergency. Lancet Respir Med 2023 Jul;11(7):588. [doi: 10.1016/s2213-2600(23)00217-5]
101. Dirlikov E, Zhou S, Han L, Li Z, Hao L, Millman AJ, et al. Use of public data to describe COVID-19 contact tracing in Hubei Province and non-Hubei provinces in China between 20 January and 29 February 2020. Western Pac Surveill Response J 2021 Sep 30;12(3):82-87 [FREE Full text] [doi: 10.5365/wpsar.2021.12.3.808] [Medline: 34703640]
102. Wang M, Zhao Y, Wei L, Chen D, Yang B, Chen Z. Immunology of transplant patients with SARS-CoV-2 infection: transmission, immune response, and therapeutic strategy. Crit Rev Immunol 2020;40(6):475-484. [doi: 10.1615/critrevimmunol.2020035755]
103. Ji W, Wang W, Zhao X, Zai J, Li X. Cross-species transmission of the newly identified coronavirus 2019-nCoV. J Med Virol 2020 Apr 19;92(4):433-440 [FREE Full text] [doi: 10.1002/jmv.25682] [Medline: 31967321]
104. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
105. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
106. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb 15;395(10223):514-523 [FREE Full text] [doi: 10.1016/S0140-6736(20)30154-9] [Medline: 31986261]
107. Hao R, Zhang Y, Cao Z, Li J, Xu Q, Ye L, et al. Control strategies and their effects on the COVID-19 pandemic in 2020 in representative countries. J Biosaf Biosecur 2021 Dec;3(2):76-81 [FREE Full text] [doi: 10.1016/j.jobb.2021.06.003] [Medline: 34189427]
108. Quan L, Al-Ansi A, Han H. Assessing customer financial risk perception and attitude in the hotel industry: exploring the role of protective measures against COVID-19. Int J Hosp Manag 2022 Feb;101:103123 [FREE Full text] [doi: 10.1016/j.ijhm.2021.103123] [Medline: 34955585]
109. Sheikh A, Al-Taher A, Al-Nazawi M, Al-Mubarak AI, Kandeel M. Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. J Virol Methods 2020 Mar;277:113806 [FREE Full text] [doi: 10.1016/j.jviromet.2019.113806] [Medline: 31911390]
110. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 2020 Jan 28;9(1):221-236 [FREE Full text] [doi: 10.1080/22221751.2020.1719902] [Medline: 31987001]

XSL•FO
RenderX

111. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J Virol 2020 Mar 17;94(7). [doi: 10.1128/jvi.00127-20]
112. Ganesh R, Salonen BR, Bhuiyan MN, Bierle DM, Moehnke D, Haddad TC, et al. Managing patients in the COVID-19 pandemic: a virtual multidisciplinary approach. Mayo Clin Proc Innov Qual Outcomes 2021 Feb;5(1):118-126 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.12.003] [Medline: 33521583]
113. Walonoski J, Klaus S, Granger E, Hall D, Gregorowicz A, Neyarapally G, et al. Synthea™ Novel coronavirus (COVID-19) model and synthetic data set. Intell Based Med 2020 Nov;1:100007 [FREE Full text] [doi: 10.1016/j.ibmed.2020.100007] [Medline: 33043312]
114. Luu HS, Filkins LM, Park JY, Rakheja D, Tweed J, Menzies C, et al. Harnessing the electronic health record and computerized provider order entry data for resource management during the COVID-19 pandemic: development of a decision tree. JMIR Med Inform 2021 Oct 18;9(10):e32303 [FREE Full text] [doi: 10.2196/32303] [Medline: 34546942]
115. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid utilization of telehealth in a comprehensive cancer center as a response to COVID-19: cross-sectional analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
116. Cronin AO, Carlile MA, Dameff CJ, Coyne CJ, Castillo EM. Leveraging remote research associates during a pandemic. West J Emerg Med 2020 Jul 21;21(5):1114-1117 [FREE Full text] [doi: 10.5811/westjem.2020.6.48043] [Medline: 32970563]
117. Deeds SA, Hagan SL, Geyer JR, Vanderwarker C, Grandjean MW, Reddy A, et al. Leveraging an electronic health record note template to standardize screening and testing for COVID-19. Healthc (Amst) 2020 Sep;8(3):100454 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100454] [Medline: 32919584]
118. Zhao J, Grabowska ME, Kerchberger VE, Smith JC, Eken HN, Feng Q, et al. ConceptWAS: a high-throughput method for early identification of COVID-19 presenting symptoms. medRxiv Preprint posted online on November 10, 2020 [FREE Full text] [doi: 10.1101/2020.11.06.20227165] [Medline: 33200151]
119. Razonable RR, Aloia NC, Anderson RJ, Anil G, Arndt LL, Arndt RF, et al. A framework for outpatient infusion of antispike monoclonal antibodies to high-risk patients with mild-to-moderate Coronavirus disease-19: the mayo clinic model. Mayo Clin Proc 2021 May;96(5):1250-1261 [FREE Full text] [doi: 10.1016/j.mayocp.2021.03.010] [Medline: 33958056]
120. Bardenheier BH, Gravenstein S, Blackman C, Gutman R, Sarkar IN, Feifer RA, et al. Adverse events following mRNA SARS-CoV-2 vaccination among U.S. nursing home residents. Vaccine 2021 Jun 29;39(29):3844-3851 [FREE Full text] [doi: 10.1016/j.vaccine.2021.05.088] [Medline: 34092431]
121. Thompson LA, Gurka MJ, Filipp SL, Schatz DA, Mercado RE, Ostrov DA, et al. The influence of selection bias on identifying an association between allergy medication use and SARS-CoV-2 infection. EClinicalMedicine 2021 Jul;37:100936 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100936] [Medline: 34104879]
122. McMurry R, Lenehan P, Awasthi S, Silvert E, Puranik A, Pawlowski C, et al. Real-time analysis of a mass vaccination effort confirms the safety of FDA-authorized mRNA COVID-19 vaccines. Med 2021 Aug 13;2(8):965-78.e5 [FREE Full text] [doi: 10.1016/j.medj.2021.06.006] [Medline: 34230920]
123. Bejan CA, Cahill KN, Staso PJ, Choi L, Peterson JF, Phillips EJ. DrugWAS: drug-wide association studies for COVID-19 drug repurposing. Clin Pharmacol Ther 2021 Dec 10;110(6):1537-1546 [FREE Full text] [doi: 10.1002/cpt.2376] [Medline: 34314511]
124. Grant WB, Lahore H, McDonnell SL, Baggerly CA, French CB, Aliano JL, et al. Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths. Nutrients 2020 Apr 02;12(4):988 [FREE Full text] [doi: 10.3390/nu12040988] [Medline: 32252338]
125. Grange ES, Neil EJ, Stoffel M, Singh AP, Tseng E, Resco-Summers K, et al. Responding to COVID-19: the UW medicine information technology services experience. Appl Clin Inform 2020 Mar;11(2):265-275 [FREE Full text] [doi: 10.1055/s-0040-1709715] [Medline: 32268390]
126. Wood SM, White K, Peebles R, Pickel J, Alausa M, Mehringer J, et al. Outcomes of a rapid adolescent telehealth scale-up during the COVID-19 pandemic. J Adolesc Health 2020 Aug;67(2):172-178 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.025] [Medline: 32611509]
127. Hoffman PE, London YR, Weerakoon TS, DeLucia NL. Rapidly scaling video visits during COVID-19: the ethos of virtual care at Yale Medicine. Healthc (Amst) 2020 Dec;8(4):100482 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100482] [Medline: 33129179]
128. Jose T, Warner DO, O'Horo JC, Peters SG, Chaudhry R, Binnicker MJ, et al. Digital health surveillance strategies for management of coronavirus disease 2019. Mayo Clin Proc Innov Qual Outcomes 2021 Feb;5(1):109-117 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.12.004] [Medline: 33521582]
129. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. Lancet Respir Med 2020 May;8(5):434-436 [FREE Full text] [doi: 10.1016/S2213-2600(20)30134-X] [Medline: 32203710]
130. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]

XSL•FO
RenderX

131. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat Hum Behav 2020 Dec;4(12):1303-1312. [doi: 10.1038/s41562-020-01009-0] [Medline: 33199859]
132. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. Ann Intern Med 2020 Aug 18;173(4):262-267 [FREE Full text] [doi: 10.7326/M20-1495] [Medline: 32422057]
133. Yu S, Yang Q, Wu J, Zhu M, Ai J, Zhang H, et al. Clinical application of Epstein-Barr virus DNA loads in Epstein-Barr virus-associated diseases: a cohort study. J Infect 2021 Jan;82(1):105-111. [doi: 10.1016/j.jinf.2020.11.027] [Medline: 33248217]
134. L'Huillier AG, Meyer B, Andrey DO, Arm-Vernez I, Baggio S, Didierlaurent A, et al. Antibody persistence in the first 6 months following SARS-CoV-2 infection among hospital workers: a prospective longitudinal study. Clin Microbiol Infect 2021 Jan 20;27(5):784.e1-784.e8 [FREE Full text] [doi: 10.1016/j.cmi.2021.01.005] [Medline: 33482352]
135. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
136. Gilbert AW, Billany JC, Adam R, Martin L, Tobin R, Bagdai S, et al. Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. BMJ Open Qual 2020 May;9(2):e000985 [FREE Full text] [doi: 10.1136/bmjoq-2020-000985] [Medline: 32439740]
137. Cocoros NM, Fuller CC, Adimadhyam S, Ball R, Brown JS, Dal Pan GJ, et al. A COVID-19-ready public health surveillance system: the Food and Drug Administration's sentinel system. Pharmacoepidemiol Drug Saf 2021 Jul 18;30(7):827-837 [FREE Full text] [doi: 10.1002/pds.5240] [Medline: 33797815]
138. Rosenblum HG, Gee J, Liu R, Marquez PL, Zhang B, Strid P, et al. Safety of mRNA vaccines administered during the initial 6 months of the US COVID-19 vaccination programme: an observational study of reports to the Vaccine Adverse Event Reporting System and v-safe. Lancet Infect Dis 2022 Jun;22(6):802-812 [FREE Full text] [doi: 10.1016/S1473-3099(22)00054-8] [Medline: 35271805]
139. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436 [FREE Full text] [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
140. Lim S, Yoon HI, Song KH, Kim ES, Kim HB. Face masks and containment of COVID-19: experience from South Korea. J Hosp Infect 2020 Sep;106(1):206-207 [FREE Full text] [doi: 10.1016/j.jhin.2020.06.017] [Medline: 32540463]
141. Holt M, MacGibbon J, Smith AK, Broady TR, Davis MD, Newman CE. Knowledge of Australia's My Health Record and factors associated with opting out: results from a national survey of the Australian general population and communities affected by HIV and sexually transmissible infections. PLOS Digit Health 2023 Mar;2(3):e0000200 [FREE Full text] [doi: 10.1371/journal.pdig.0000200] [Medline: 36857326]
142. Karagiannidis C, Mostert C, Hentschker C, Voshaar T, Malzahn J, Schillinger G, et al. Case characteristics, resource use, and outcomes of 10 021 patients with COVID-19 admitted to 920 German hospitals: an observational study. Lancet Respir Med 2020 Sep;8(9):853-862 [FREE Full text] [doi: 10.1016/S2213-2600(20)30316-7] [Medline: 32735842]
143. Fiske A, Schönweitz F, Eichinger J, Zimmermann B, Hangel N, Sierawska A, et al. The COVID-19 vaccine: trust, doubt, and hope for a future beyond the pandemic in Germany. PLoS One 2022;17(4):e0266659 [FREE Full text] [doi: 10.1371/journal.pone.0266659] [Medline: 35390085]
144. Chowkwanyun M, Reed ALJ. Racial health disparities and Covid-19 - caution and context. N Engl J Med 2020 Jul 16;383(3):201-203. [doi: 10.1056/NEJMp2012910] [Medline: 32374952]
145. Pflugeisen BM, Mou J. Empiric evidence of ethnic disparities in coronavirus positivity in Washington State. Ethn Health 2021 Jan 11;26(1):36-48. [doi: 10.1080/13557858.2020.1863922] [Medline: 33428455]
146. Cruz TM, Smith SA. Health equity beyond data: health care worker perceptions of race, ethnicity, and language data collection in electronic health records. Med Care 2021 May 01;59(5):379-385. [doi: 10.1097/MLR.0000000000001507] [Medline: 33528233]
147. Abdel Magid HS, Ferguson JM, Van Cleve R, Purnell AL, Osborne TF. Differences in COVID-19 risk by race and county-level social determinants of health among veterans. Int J Environ Res Public Health 2021 Dec 13;18(24):13140 [FREE Full text] [doi: 10.3390/ijerph182413140] [Medline: 34948748]
148. Herrera T, Fiori KP, Archer-Dyer H, Lounsbury DW, Wylie-Rosett J. Social determinants of health screening by preclinical medical students during the COVID-19 pandemic: service-based learning case study. JMIR Med Educ 2022 Jan 17;8(1):e32818 [FREE Full text] [doi: 10.2196/32818] [Medline: 35037885]
149. Winkelman TN, Margolis KL, Waring S, Bodurtha PJ, Khazanchi R, Gildemeister S, et al. Minnesota electronic health record consortium COVID-19 project: informing pandemic response through statewide collaboration using observational data. Public Health Rep 2022;137(2):263-271 [FREE Full text] [doi: 10.1177/00333549211061317] [Medline: 35060411]
150. Phuong J, Zampino E, Dobbins N, Espinoza J, Meeker D, Spratt H, et al. Extracting patient-level social determinants of health into the OMOP common data model. AMIA Annu Symp Proc 2021;2021:989-998 [FREE Full text] [Medline: 35308947]

XSL•FO
RenderX

151. Phuong J, Hong S, Palchuk MB, Espinoza J, Meeker D, Dorr DA, et al. Advancing interoperability of patient-level social determinants of health data to support COVID-19 research. AMIA Jt Summits Transl Sci Proc 2022;2022:396-405 [FREE Full text] [Medline: 35854720]
152. Kim EJ, Marrast L, Conigliaro J. COVID-19: magnifying the effect of health disparities. J Gen Intern Med 2020 Aug;35(8):2441-2442 [FREE Full text] [doi: 10.1007/s11606-020-05881-4] [Medline: 32394141]
153. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
154. Bambra C, Riordan R, Ford J, Matthews F. The COVID-19 pandemic and health inequalities. J Epidemiol Community Health 2020 Nov;74(11):964-968 [FREE Full text] [doi: 10.1136/jech-2020-214401] [Medline: 32535550]
155. Liu CH, Pinder-Amaker S, Hahm HC, Chen JA. Priorities for addressing the impact of the COVID-19 pandemic on college student mental health. J Am Coll Health 2022 Jul 13;70(5):1356-1358 [FREE Full text] [doi: 10.1080/07448481.2020.1803882] [Medline: 33048654]
156. Jaspal R, Breakwell GM. Socio-economic inequalities in social network, loneliness and mental health during the COVID-19 pandemic. Int J Soc Psychiatry 2022 Feb 07;68(1):155-165 [FREE Full text] [doi: 10.1177/0020764020976694] [Medline: 33287610]
157. Thorell LB, Skoglund C, de la Peña AG, Baeyens D, Fuermaier AB, Groom MJ, et al. Parental experiences of homeschooling during the COVID-19 pandemic: differences between seven European countries and between children with and without mental health conditions. Eur Child Adolesc Psychiatry 2022 Apr 07;31(4):649-661 [FREE Full text] [doi: 10.1007/s00787-020-01706-1] [Medline: 33415470]
158. Hawes MT, Szenczy AK, Klein DN, Hajcak G, Nelson BD. Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. Psychol Med 2022 Oct;52(14):3222-3230 [FREE Full text] [doi: 10.1017/S0033291720005358] [Medline: 33436120]
159. Hewitt KC, Marra DE, Block C, Cysique LA, Drane DL, Haddad MM, et al. Central nervous system manifestations of COVID-19: a critical review and proposed research agenda. J Int Neuropsychol Soc 2021 Apr 16;28(3):311-325. [doi: 10.1017/s1355617721000345]
160. Hugon J, Msika EF, Queneau M, Farid K, Paquet C. Long COVID: cognitive complaints (brain fog) and dysfunction of the cingulate cortex. J Neurol 2022 Jan;269(1):44-46 [FREE Full text] [doi: 10.1007/s00415-021-10655-x] [Medline: 34143277]
161. Damiano RF, Guedes BF, de Rocca CC, de Pádua Serafim A, Castro LH, Munhoz CD, et al. Cognitive decline following acute viral infections: literature review and projections for post-COVID-19. Eur Arch Psychiatry Clin Neurosci 2022 Feb 25;272(1):139-154 [FREE Full text] [doi: 10.1007/s00406-021-01286-4] [Medline: 34173049]
162. Caraballo PJ, Sutton JA, Giri J, Wright JA, Nicholson WT, Kullo IJ, et al. Integrating pharmacogenomics into the electronic health record by implementing genomic indicators. J Am Med Inform Assoc 2020 Jan 01;27(1):154-158 [FREE Full text] [doi: 10.1093/jamia/ocz177] [Medline: 31591640]
163. Glynn P, Greenland P. Contributions of the UK biobank high impact papers in the era of precision medicine. Eur J Epidemiol 2020 Jan 28;35(1):5-10. [doi: 10.1007/s10654-020-00606-7] [Medline: 31993883]
164. Cortes A, Dendrou CA, Motyer A, Jostins L, Vukcevic D, Dilthey A, et al. Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. Nat Genet 2017 Sep 31;49(9):1311-1318 [FREE Full text] [doi: 10.1038/ng.3926] [Medline: 28759005]
165. Gill D, Benyamin B, Moore LS, Monori G, Zhou A, Koskeridis F, et al. Associations of genetically determined iron status across the phenome: a mendelian randomization study. PLoS Med 2019 Jun 20;16(6):e1002833 [FREE Full text] [doi: 10.1371/journal.pmed.1002833] [Medline: 31220083]
166. Biddinger KJ, Emdin CA, Haas ME, Wang M, Hindy G, Ellinor PT, et al. Association of habitual alcohol intake with risk of cardiovascular disease. JAMA Netw Open 2022 Mar 01;5(3):e223849 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.3849] [Medline: 35333364]
167. Harrison S, Davies NM, Howe LD, Hughes A. Testosterone and socioeconomic position: Mendelian randomization in 306,248 men and women in UK Biobank. Sci Adv 2021 Jul;7(31):eabf8257 [FREE Full text] [doi: 10.1126/sciadv.abf8257] [Medline: 34321204]
168. Patchen BK, Clark AG, Gaddis N, Hancock DB, Cassano PA. Genetically predicted serum vitamin D and COVID-19: a Mendelian randomisation study. BMJ Nutr Prev Health 2021;4(1):213-225 [FREE Full text] [doi: 10.1136/bmjnph-2021-000255] [Medline: 34308129]
169. Guertin J, Kaiser Y, Manikpurage H, Perrot N, Bourgeois R, Couture C, et al. Sex-specific associations of genetically predicted circulating Lp(a) (Lipoprotein(a)) and hepatic LPA gene expression levels with cardiovascular outcomes: Mendelian randomization and observational analyses. Circ Genom Precis Med 2021 Aug;14(4):e003271. [doi: 10.1161/CIRCGEN.120.003271] [Medline: 34279996]
170. Millard LA, Fernández-Sanlés A, Carter AR, Hughes RA, Tilling K, Morris TP, et al. Exploring the impact of selection bias in observational studies of COVID-19: a simulation study. Int J Epidemiol 2023 Feb 08;52(1):44-57 [FREE Full text] [doi: 10.1093/ije/dyac221] [Medline: 36474414]

XSL•FO
RenderX

171. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul;19(7):743-750 [FREE Full text] [doi: 10.1038/gim.2016.183] [Medline: 27929525]
172. Erlangsen A, Fedyszyn I. Danish nationwide registers for public health and health-related research. Scand J Public Health 2015 Jun 10;43(4):333-339. [doi: 10.1177/1403494815575193] [Medline: 25759376]
173. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, BioBank Japan Cooperative Hospital Group, et al. Overview of the BioBank Japan Project: study design and profile. J Epidemiol 2017 Mar;27(3S):S2-S8 [FREE Full text] [doi: 10.1016/j.je.2016.12.005] [Medline: 28189464]
174. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011 Dec;40(6):1652-1666 [FREE Full text] [doi: 10.1093/ije/dyr120] [Medline: 22158673]
175. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. Perm J 2012;16(3):37-41 [FREE Full text] [doi: 10.7812/TPP/12-031] [Medline: 23012597]
176. Gulcher J, Stefansson K. An Icelandic saga on a centralized healthcare database and democratic decision making. Nat Biotechnol 1999 Jul;17(7):620. [doi: 10.1038/10796] [Medline: 10409337]
177. Sakaue S, Kanai M, Karjalainen J, Akiyama M, Kurki M, Matoba N, FinnGen, et al. Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. Nat Med 2020 Apr 23;26(4):542-548. [doi: 10.1038/s41591-020-0785-8] [Medline: 32251405]
178. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin Transl Sci 2010 Feb 10;3(1):42-48 [FREE Full text] [doi: 10.1111/j.1752-8062.2010.00175.x] [Medline: 20443953]
179. Zawistowski M, Fritsche LG, Pandit A, Vanderwerff B, Patil S, Schmidt EM, et al. The Michigan Genomics Initiative: a biobank linking genotypes and electronic clinical records in Michigan Medicine patients. Cell Genom 2023 Feb 08;3(2):100257 [FREE Full text] [doi: 10.1016/j.xgen.2023.100257] [Medline: 36819667]
180. Levin MA, Joseph TT, Jeff JM, Nadukuru R, Ellis SB, Bottinger EP, et al. iGAS: a framework for using electronic intraoperative medical records for genomic discovery. J Biomed Inform 2017 Mar;67:80-89 [FREE Full text] [doi: 10.1016/j.jbi.2017.02.005] [Medline: 28193464]
181. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol 2015 Aug 11;44(4):1137-1147. [doi: 10.1093/ije/dyt268] [Medline: 24518929]
182. Lekhal R, Zachrisson HD, Wang MV, Schjølberg S, von Soest T. Does universally accessible child care protect children from late talking? Results from a Norwegian population-based prospective study. Early Child Dev Care 2011 Sep;181(8):1007-1019. [doi: 10.1080/03004430.2010.508558]
183. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
184. Bernardini M, Romeo L, Misericordia P, Frontoni E. Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. IEEE J Biomed Health Inform 2020 Jan;24(1):235-246. [doi: 10.1109/JBHI.2019.2899218] [Medline: 30762572]
185. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 2020 Feb;145(2):463-469 [FREE Full text] [doi: 10.1016/j.jaci.2019.12.897] [Medline: 31883846]
186. Yu K, Xie X. Predicting hospital readmission: a joint ensemble-learning model. IEEE J Biomed Health Inform 2020 Feb;24(2):447-456. [doi: 10.1109/JBHI.2019.2938995] [Medline: 31484143]
187. Wang Y, Wu T, Wang Y, Wang G. Enhancing model interpretability and accuracy for disease progression prediction via phenotype-based patient similarity learning. Pac Symp Biocomput 2020;25:511-522 [FREE Full text] [Medline: 31797623]
188. Tang S, Chappell GT, Mazzoli A, Tewari M, Choi SW, Wiens J. Predicting acute graft-versus-host disease using machine learning and longitudinal vital sign data from electronic health records. JCO Clin Cancer Inform 2020 Nov;4:128-135. [doi: 10.1200/cci.19.00105]
189. Chae D. Data science and machine learning in anesthesiology. Korean J Anesthesiol 2020 Aug;73(4):285-295 [FREE Full text] [doi: 10.4097/kja.20124] [Medline: 32209960]
190. Li R, Yin C, Yang S, Qian B, Zhang P. Marrying medical domain knowledge with deep learning on electronic health records: a deep visual analytics approach. J Med Internet Res 2020 Sep 28;22(9):e20645 [FREE Full text] [doi: 10.2196/20645] [Medline: 32985996]
191. Zhang J, Olatosi B, Yang X, Weissman S, Li Z, Hu J, et al. Studying patterns and predictors of HIV viral suppression using A Big Data approach: a research protocol. BMC Infect Dis 2022 Feb 04;22(1):122 [FREE Full text] [doi: 10.1186/s12879-022-07047-5] [Medline: 35120435]

XSL•FO
RenderX

192. Schinkel M, Boerman AW, Bennis FC, Minderhoud TC, Lie M, Peters-Sengers H, et al. Diagnostic stewardship for blood cultures in the emergency department: a multicenter validation and prospective evaluation of a machine learning prediction tool. EBioMedicine 2022 Aug;82:104176 [FREE Full text] [doi: 10.1016/j.ebiom.2022.104176] [Medline: 35853298]
193. Nishimwe A, Ruranga C, Musanabaganwa C, Mugeni R, Semakula M, Nzabanita J, et al. Leveraging artificial intelligence and data science techniques in harmonizing, sharing, accessing and analyzing SARS-COV-2/COVID-19 data in Rwanda (LAISDAR Project): study design and rationale. BMC Med Inform Decis Mak 2022 Aug 12;22(1):214 [FREE Full text] [doi: 10.1186/s12911-022-01965-9] [Medline: 35962355]
194. Dam TA, Fleuren LM, Roggeveen LF, Otten M, Biesheuvel L, Jagesar AR, et al. Augmented intelligence facilitates concept mapping across different electronic health records. Int J Med Inform 2023 Nov;179:105233 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105233] [Medline: 37748329]
195. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
196. Ayala Solares JR, Diletta Raimondi FE, Zhu Y, Rahimian F, Canoy D, Tran J, et al. Deep learning for electronic health records: a comparative review of multiple deep neural architectures. J Biomed Inform 2020 Jan;101:103337 [FREE Full text] [doi: 10.1016/j.jbi.2019.103337] [Medline: 31916973]
197. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
198. Uciteli A, Beger C, Kirsten T, Meineke FA, Herre H. Ontological representation, classification and data-driven computing of phenotypes. J Biomed Semantics 2020 Dec 21;11(1):15 [FREE Full text] [doi: 10.1186/s13326-020-00230-0] [Medline: 33349245]
199. Suryanarayanan P, Epstein EA, Malvankar A, Lewis BL, DeGenaro L, Liang JJ, et al. Timely and efficient AI insights on EHR: system design. AMIA Annu Symp Proc 2020;2020:1180-1189 [FREE Full text] [Medline: 33936494]
200. Scalia P, Ahmad F, Schubbe D, Forcino R, Durand MA, Barr PJ, et al. Integrating option grid patient decision aids in the epic electronic health record: case study at 5 health systems. J Med Internet Res 2021 May 03;23(5):e22766 [FREE Full text] [doi: 10.2196/22766] [Medline: 33938806]
201. Soguero-Ruiz C, Fei WM, Jenssen R, Augestad KM, Álvarez JL, Jiménez IM, et al. Data-driven temporal prediction of surgical site infection. AMIA Annu Symp Proc 2015;2015:1164-1173 [FREE Full text] [Medline: 26958256]
202. Bronsert M, Singh AB, Henderson WG, Hammermeister K, Meguid RA, Colborn KL. Identification of postoperative complications using electronic health record data and machine learning. Am J Surg 2020 Jul;220(1):114-119 [FREE Full text] [doi: 10.1016/j.amjsurg.2019.10.009] [Medline: 31635792]
203. Hernandez-Boussard T, Blayney DW, Brooks JD. Leveraging digital data to inform and improve quality cancer care. Cancer Epidemiol Biomarkers Prev 2020 Apr;29(4):816-822 [FREE Full text] [doi: 10.1158/1055-9965.EPI-19-0873] [Medline: 32066619]
204. Zhuang Y, Dyas A, Meguid RA, Henderson WG, Bronsert M, Madsen H, et al. Preoperative prediction of postoperative infections using machine learning and electronic health record data. Ann Surg 2024 Apr 01;279(4):720-726. [doi: 10.1097/SLA.0000000000006106] [Medline: 37753703]
205. Savage T, Wang J, Shieh L. A large language model screening tool to target patients for best practice alerts: development and validation. JMIR Med Inform 2023 Nov 27;11:e49886 [FREE Full text] [doi: 10.2196/49886] [Medline: 38010803]
206. Zhu V, Lenert L, Bunnell B, Obeid J, Jefferson M, Halbert CH. Automatically identifying financial stress information from clinical notes for patients with prostate cancer. Cancer Res Rep 2020;1(1):102 [FREE Full text] [doi: 10.61545/crr-1-102] [Medline: 38317775]
207. Tedeschi SK, Cai T, He Z, Ahuja Y, Hong C, Yates KA, et al. Classifying pseudogout using machine learning approaches with electronic health record data. Arthritis Care Res (Hoboken) 2021 Mar 25;73(3):442-448 [FREE Full text] [doi: 10.1002/acr.24132] [Medline: 31910317]
208. Hu B, Bajracharya A, Yu H. Generating medical assessments using a neural network model: algorithm development and validation. JMIR Med Inform 2020 Jan 15;8(1):e14971 [FREE Full text] [doi: 10.2196/14971] [Medline: 31939742]
209. Colling C, Khondoker M, Patel R, Fok M, Harland R, Broadbent M, et al. Predicting high-cost care in a mental health setting. BJPsych Open 2020 Jan 17;6(1):e10 [FREE Full text] [doi: 10.1192/bjo.2019.96] [Medline: 31950891]
210. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med 2022 Dec 26;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
211. Chen A, Yu Z, Yang X, Guo Y, Bian J, Wu Y. Contextualized medication information extraction using transformer-based deep learning architectures. J Biomed Inform 2023 Jun;142:104370 [FREE Full text] [doi: 10.1016/j.jbi.2023.104370] [Medline: 37100106]
212. Peng C, Yang X, Yu Z, Bian J, Hogan WR, Wu Y. Clinical concept and relation extraction using prompt-based machine reading comprehension. J Am Med Inform Assoc 2023 Aug 18;30(9):1486-1493 [FREE Full text] [doi: 10.1093/jamia/ocad107] [Medline: 37316988]

XSL•FO
RenderX

213. Kim MK, Rouphael C, McMichael J, Welch N, Dasarathy S. Challenges in and opportunities for electronic health record-based data analysis and interpretation. Gut Liver 2024 Mar 15;18(2):201-208 [FREE Full text] [doi: 10.5009/gnl230272] [Medline: 37905424]
214. Holmes JH, Beinlich J, Boland MR, Bowles KH, Chen Y, Cook TS, et al. Why is the electronic health record so challenging for research and clinical care? Methods Inf Med 2021 May;60(1-02):32-48 [FREE Full text] [doi: 10.1055/s-0041-1731784] [Medline: 34282602]
215. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, et al. CODE-EHR best-practice framework for the use of structured electronic health-care records in clinical research. Lancet Digit Health 2022 Oct;4(10):e757-e764. [doi: 10.1016/s2589-7500(22)00151-0]
216. Tong L, Shi W, Isgut M, Zhong Y, Lais P, Gloster L, et al. Integrating multi-omics data with EHR for precision medicine using advanced artificial intelligence. IEEE Rev Biomed Eng 2024;17:80-97. [doi: 10.1109/rbme.2023.3324264]
217. Lau-Min KS, Asher SB, Chen J, Domchek SM, Feldman M, Joffe S, et al. Real-world integration of genomic data into the electronic health record: the PennChart Genomics Initiative. Genet Med 2021 Apr;23(4):603-605 [FREE Full text] [doi: 10.1038/s41436-020-01056-y] [Medline: 33299147]
218. Lau-Min KS, McKenna D, Asher SB, Bardakjian T, Wollack C, Bleznuck J, et al. Impact of integrating genomic data into the electronic health record on genetics care delivery. Genet Med 2022 Nov;24(11):2338-2350 [FREE Full text] [doi: 10.1016/j.gim.2022.08.009] [Medline: 36107166]
219. Chung WK, Berg JS, Botkin JR, Brenner SE, Brosco JP, Brothers KB, et al. Newborn screening for neurodevelopmental diseases: are we there yet? Am J Med Genet C Semin Med Genet 2022 Jun;190(2):222-230 [FREE Full text] [doi: 10.1002/ajmg.c.31988] [Medline: 35838066]
220. Sosinsky A, Ambrose J, Cross W, Turnbull C, Henderson S, Jones L, et al. Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 Genomes Cancer Programme. Nat Med 2024 Jan 11;30(1):279-289 [FREE Full text] [doi: 10.1038/s41591-023-02682-0] [Medline: 38200255]
221. Zeng Z, Deng Y, Li X, Naumann T, Luo Y. Natural language processing for EHR-based computational phenotyping. IEEE/ACM Trans Comput Biol Bioinform 2019;16(1):139-153 [FREE Full text] [doi: 10.1109/TCBB.2018.2849968] [Medline: 29994486]
222. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
223. Pevnick JM, Elad Y, Masson LM, Riggs RV, Duncan RG. Patient-initiated data: our experience with enabling patients to initiate incorporation of heart rate data into the electronic health record. Appl Clin Inform 2020 Aug 14;11(4):671-679 [FREE Full text] [doi: 10.1055/s-0040-1716538] [Medline: 33058102]
224. Shaw R, Stroo M, Fiander C, McMillan K. Selecting mobile health technologies for electronic health record integration: case study. J Med Internet Res 2020 Oct 28;22(10):e23314 [FREE Full text] [doi: 10.2196/23314] [Medline: 33112248]
225. Shaw RJ, Montgomery K, Fiander C, Bullock K, Craig R, Pennington G, et al. Mobile monitoring-enabled telehealth for patients with complex chronic illnesses. Stud Health Technol Inform 2024 Jan 25;310:194-198. [doi: 10.3233/SHTI230954] [Medline: 38269792]
226. Richter JG, Chehab G, Stachwitz P, Hagen J, Larsen D, Knitza J, et al. One year of digital health applications (DiGA) in Germany - rheumatologists' perspectives. Front Med (Lausanne) 2022;9:1000668 [FREE Full text] [doi: 10.3389/fmed.2022.1000668] [Medline: 36388899]
227. Cantor MN, Thorpe L. Integrating data on social determinants of health into electronic health records. Health Aff (Millwood) 2018 Apr;37(4):585-590 [FREE Full text] [doi: 10.1377/hlthaff.2017.1252] [Medline: 29608369]
228. Honerlaw J, Ho YL, Fontin F, Murray M, Galloway A, Heise D, et al. Centralized Interactive Phenomics Resource: an integrated online phenomics knowledgebase for health data users. J Am Med Inform Assoc 2024 Apr 19;31(5):1126-1134 [FREE Full text] [doi: 10.1093/jamia/ocae042] [Medline: 38481028]
229. Honerlaw J, Ho YL, Fontin F, Gosian J, Maripuri M, Murray M, et al. Framework of the Centralized Interactive Phenomics Resource (CIPHER) standard for electronic health data-based phenomics knowledgebase. J Am Med Inform Assoc 2023 Apr 19;30(5):958-964 [FREE Full text] [doi: 10.1093/jamia/ocad030] [Medline: 36882092]
230. Fodor J. HIPAA and the EHR: making technical safeguard changes. J AHIMA 2004 Jan;75(1):54-5; quiz 61. [Medline: 14748230]
231. Wang T, Pizziferri L, Volk LA, Mikels DA, Grant KG, Wald JS, et al. Implementing patient access to electronic health records under HIPAA: lessons learned. Perspect Health Inf Manag 2004 Dec 15;1:11 [FREE Full text] [Medline: 18066391]
232. Nijor S, Rallis G, Lad N, Gokcen E. Patient safety issues from information overload in electronic medical records. J Patient Saf 2022 Sep 01;18(6):e999-1003 [FREE Full text] [doi: 10.1097/PTS.0000000000001002] [Medline: 35985047]
233. Cornock M. General Data Protection Regulation (GDPR) and implications for research. Maturitas 2018 May;111:A1-A2. [doi: 10.1016/j.maturitas.2018.01.017] [Medline: 29395455]
234. Kavianpour S, Sutherland J, Mansouri-Benssassi E, Coull N, Jefferson E. Next-generation capabilities in trusted research environments: interview study. J Med Internet Res 2022 Sep 20;24(9):e33720 [FREE Full text] [doi: 10.2196/33720] [Medline: 36125859]

XSL•FO
RenderX

235. Kruse CS, Smith B, Vanderlinden H, Nealand A. Security techniques for the electronic health records. J Med Syst 2017 Aug;41(8):127 [FREE Full text] [doi: 10.1007/s10916-017-0778-4] [Medline: 28733949]
236. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience 2017;11:709 [FREE Full text] [doi: 10.3332/ecancer.2017.709] [Medline: 28144283]
237. Rumbold JM, Pierscionek B. The effect of the general data protection regulation on medical research. J Med Internet Res 2017 Feb 24;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
238. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
239. Qualls LG, Phillips TA, Hammill BG, Topping J, Louzao DM, Brown JS, et al. Evaluating foundational data quality in the national patient-centered clinical research network (PCORnet®). EGEMS (Wash DC) 2018 Apr 13;6(1):3 [FREE Full text] [doi: 10.5334/egems.199] [Medline: 29881761]
240. Tison GH, Chamberlain AM, Pletcher MJ, Dunlay SM, Weston SA, Killian JM, et al. Identifying heart failure using EMR-based algorithms. Int J Med Inform 2018 Dec;120:1-7 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.09.016] [Medline: 30409334]
241. Klann JG, Joss MA, Embree K, Murphy SN. Data model harmonization for the All of Us Research Program: transforming i2b2 data into the OMOP common data model. PLoS One 2019;14(2):e0212463 [FREE Full text] [doi: 10.1371/journal.pone.0212463] [Medline: 30779778]
242. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 2012;19(1):54-60 [FREE Full text] [doi: 10.1136/amiajnl-2011-000376] [Medline: 22037893]
243. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med 2012 Dec 30;31(30):4401-4415. [doi: 10.1002/sim.5620] [Medline: 23015364]
244. Hossain E, Rana R, Higgins N, Soar J, Datta Barua P, Pisani AR, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. Comput Biol Med 2023 Mar;155:106649. [doi: 10.1016/j.compbiomed.2023.106649] [Medline: 36805219]
245. Min L, Tian Q, Lu X, Duan H. Modeling EHR with the openEHR approach: an exploratory study in China. BMC Med Inform Decis Mak 2018 Aug 29;18(1):75 [FREE Full text] [doi: 10.1186/s12911-018-0650-6] [Medline: 30157838]
246. Veinot TC, Ancker JS, Bakken S. Health informatics and health equity: improving our reach and impact. J Am Med Inform Assoc 2019 Aug 01;26(8-9):689-695 [FREE Full text] [doi: 10.1093/jamia/ocz132] [Medline: 31411692]
247. Zhang X, Hailu B, Tabor DC, Gold R, Sayre MH, Sim I, et al. Role of health information technology in addressing health disparities: patient, clinician, and system perspectives. Med Care 2019 Jun;57 Suppl 6 Suppl 2(Suppl 6 2):S115-S120 [FREE Full text] [doi: 10.1097/MLR.0000000000001092] [Medline: 31095049]

Abbreviations

AI: artificial intelligence
CIPHER: Centralized Interactive Phenomics Resource
EHR: electronic health record
EMA: European Medicines Agency
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
HITECH: Health Information Technology for Economic and Clinical Health Act
ICU: intensive care unit
NHS: National Health Service
NLP: natural language processing
PCT: pragmatic clinical trial
PHEIC: public health emergency of international concern
POMR: problem-oriented medical record
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RMRS: Regenstrief Medical Record System
RWE: real-world evidence
SDOH: social determinants of health
UKB: UK Biobank

XSL•FO
RenderX

Edited by G Eysenbach; submitted 31.03.24; peer-reviewed by X Li, C Gao; comments to author 12.09.24; revised version received 02.10.24; accepted 05.12.24; published 09.01.25.

Please cite as:
Shen Y, Yu J, Zhou J, Hu G
Twenty-Five Years of Evolution and Hurdles in Electronic Health Records and Interoperability in Medical Research: Comprehensive Review
J Med Internet Res 2025;27:e59024
URL: https://www.jmir.org/2025/1/e59024
doi:10.2196/59024
PMID:

©Yun Shen, Jiamin Yu, Jian Zhou, Gang Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Paradigm Shift From Patient to Health Consumer: 20 Years of Value Assessment in Health

Eline M van den Broek-Altenburg[1], MS, MA, PhD; Adam J Atherly[2], PhD

[1]University of Vermont, Burlington, VT, United States

[2]Virginia Commonwealth University, Richmond, VA, United States

Corresponding Author:
Eline M van den Broek-Altenburg, MS, MA, PhD
University of Vermont
111 Colchester Avenue
Burlington, VT, 05401
United States
Phone: 1 8026562722
Email: eline.altenburg@med.uvm.edu

Abstract

Health care is undergoing a "revolution," where patients are becoming consumers and armed with apps, consumer review scores, and, in some countries, high out-of-pocket costs. Although economic analyses and health technology assessment (HTA) have come a long way in their evaluation of the clinical, economic, ethical, legal, and societal perspectives that may be impacted by new technologies and procedures, these approaches do not reflect underlying patient preferences that may be important in the assessment of "value" in the current value-based health care transition. The major challenges that come with the transformation to a value-based health care system lead to questions such as "How are economic analyses, often the basis for policy and reimbursement decisions, going to switch from a societal to an individual perspective?" and "How do we then assess (economic) value, considering individual preference heterogeneity, as well as varying heuristics and decision rules?" These challenges, related to including the individual perspective in cost-effectiveness analysis (CEA), have been widely debated. Cost-effectiveness measures treatments in terms of costs and quality-adjusted life-years (QALYs), where QALYs assume that a health state that is more desirable is more valuable, and therefore, value is equated with preference or desirability. QALYs have long been criticized for empirical and conceptual shortcomings. However, policy makers in many countries have used QALY measures to make health coverage decisions, although now, patients, and patient advocates, are questioning the valuation methodologies. This has led to the development of new approaches to valuing health, which are already starting to be used in the United States. This paper reviews 20-25 years of value assessment approaches in health and concludes with challenges and opportunities for value assessment methods in health in the years to come.

(J Med Internet Res 2025;27:e60443) doi:10.2196/60443

KEYWORDS

value assessment; cost-effectiveness; quality-adjusted life-years; QALY; health consumer; health technology; value based; digital health; patient centered; preferences; health economics

Background

The concept of value assessment has been used for decades in marketing, management, and other fields to develop strategies to assess the potential and realized customer value. Successful application of these strategies in health has required an understanding of what value assessment entails, including quantifying the impact of a provider's offering on patients' and payers' costs, as well as health gains. However, the adoption of value assessment in health has been slow, gradual, and multifaceted. One major explanation for this is that it required a paradigm shift "from patient to health consumer" to accept that some marketing concepts regarding customers are applicable to patients.

The idea of a paradigm shift toward a consumer-focused health system is not new. It was described as "taking hold" as early as the 1990s [1], with citations dating well into the 1980s—more than 4 decades ago. Much of the argument for a consumer-focused health care system revolved around changes in patients' desire for treatments that provided value on their terms, rather than the health care systems', with a desire for treatments that focused on functional loss, anxiety, and pain

XSL•FO
RenderX

rather than traditional biomedical outcomes. The path to a consumer-focused health care system was not as easy or straightforward as envisioned 35 years ago, but the movement now has a push from changes in health care, ranging from an emphasis on patient satisfaction in pay-for-performance programs [2] to a push for payment transparency to empower consumer choice [3] to the push for open access to electronic health records for patients [4]. In the United States, there has also been a rise in high-deductible health plans [5], providing extra motivation for consumers to consider their health care purchases carefully [6].

Today, this paradigm shift seems to finally be taking hold, with the aid of better data and increased use of digital tools since the COVID-19 pandemic [7], which can be observed in the way health technology assessment (HTA) methods are elaborated, with more room for individual patient preferences. This paper reviews 20-25 years of value assessment approaches in health and the role of the internet, mobile technologies, and digital health. It first describes the foundation of value assessment in health. In the second part, it discusses HTA approaches and the development of the literature, and in the third part, it describes policy efforts and future opportunities for individual value assessment.

Foundations of Value Assessment in Health

Value assessment uses tools to support better value in health care. The key question is value for whom. Many of the current value assessment tools used in health lack a sufficient patient-centered approach. Historically, most value assessment is based on the assumption that patients or "health consumers" maximize utility. In fact, the utility maximization paradigm forms the basis of many economic, psychological, cognitive, and behavioral models. Value assessment is performed by estimating a utility function that represents consumers' preferences that are assumed to be complete and transitive. This means consumers can compare any 2 goods or services, and their preferences are internally consistent. Consumers can then rank-order the goods or services under consideration according to their personal "value function."

When making a decision under certainty, the value function reflects the decision maker's preferences on a particular outcome. Outcomes are defined by the assignment of values to a set of attribute variables that are either discrete or continuous and together make up for the space of all possible outcomes. However, outcomes are often not defined under certainty but in terms of probabilities, and a utility function is needed to assess the value of a decision. Expected utility theory (EUT), proposed by von Neumann and Morgenstern in 1944 [8], described a more complex utility function representing the decision maker's attitudes regarding risk and the value of the outcomes by inducing a preference ordering on the probability distributions over the entire outcome space. Decades later, Kahneman and Tversky [9] contrasted the EUT concepts of risky and riskless choices with prospect theory (PT) [9]. Although the EUT assumed that risk attitudes derive exclusively from the way in which people value rewards as reflected in the curvature of the utility function, PT posited that one's risk attitude would also vary with their subjectively weighted outcome probabilities.

To collect preferences and derive the utility function, there are many ways to ask consumers to rank-order options. In marketing research, some of the more established methods include contingent valuation [10], contingent ranking [11], contingent behavior [12], paired comparisons [13,14], discrete choice experiments [15], (full profile) conjoint analysis [16], adaptive conjoint analysis (ACA) [17,18], and ACA's self-explicated prior [19]. In health, many of these methods have been adopted to define value for individual patients, often following the same patterns of discussion and methodological improvements as other fields. In all these methods, consumers are asked to implicitly or explicitly rank-order options and make choices based on the ordering. For example, in the ACA self-explicated model, consumers are asked to rank-order levels of an attribute of a choice, which is either assumed to be homogeneous across respondents or derived individually [19]. Next, they rate the importance of the difference between the best and the worst levels of an attribute on a 4-point scale, which is then multiplied by the preference orders and rescaled so that the difference between the highest and lowest partworths of an attribute is equal to the attribute's importance [20].

In the self-explicated model, the choice set is truncated to 4 importance scores across, and the assumption of equal successive intervals within attributes is embedded in the model. This is in contrast to actual choices, where the choice set may have many attributes, such as price, color, quality, reliability, reputation, size, calories, and packaging, and also where some choices could be quite similar and others quite different, such as the choice between 2 kinds of sliced bread, a bagel, and a pita. In the case of paired comparisons, such as full-profile ACA, it is assumed that consumers sum the weighted additive differences between alternatives on each attribute [21]. In most preference elicitation methods, the values assigned to the attributes are discrete.

There are a number of practical challenges with these approaches. Issues with the reference point [22,23], scale [24], differences in attribute processing and heuristics [25,26], learning and fatigue effects [27], attention [28], and other challenges have caused considerable methodological discussion and new modeling approaches in health, marketing, transportation, and environmental economics over the past decades as researchers work to model consumer decision-making.

Health Technology Assessment

Among the first publications in the past 25 years were studies focused on how consumers search for and appraise information on medicines on the internet. A qualitative study in 2003 found that many patients had a limited awareness of how they found and evaluated internet-based information on medicines [29]. An observational study using a convenience sample of 14 students published the same year focused on the proportion of adolescents finding correct or useful answers on the internet

XSL•FO
RenderX

[30] and found that in about two-thirds of internet searches, they were successful in finding relevant health information.

In the early 2000s, most studies regarding decision-making focused on how patients gather information to make informed choices and how they could potentially use that information. This led to the formal concept of informed decision-making, which mostly entailed that patients would decide to undergo testing or treatment with them knowing and understanding the value or need for it, such as in the case of a study evaluating the effect of the web-based prostate-specific antigen (PSA) decision aid, Prosdex, on informed decision-making [31]. This study, published in 2010, was a web-based randomized controlled trial among men aged 50-75 years who had not previously undergone a PSA test, who were randomly allocated to 2 intervention and 2 control groups [31]. Participants in the intervention groups either viewed the decision aid Prosdex or were given a paper version of the text. The study showed that Prosdex is associated with improved knowledge about the PSA test and prostate cancer and that men who have a high level of knowledge have a less favorable attitude toward and are less likely to undergo PSA testing. Around this time, more studies used quantitative methods to assess the effect of health information on decision-making regarding health services and health behaviors. Examples include a study on the effects of internet-based tailored advice on the use of cholesterol-lowering interventions [32] and a feasibility study of an internet-based education and decision program for patients with early-stage prostate cancer [33].

In the years following, increasingly more studies appeared that discussed informed decision-making in the context of cost-effectiveness and cost utility [34-38], where the assessment of utility or value was now described as HTA. HTA is used to ensure that health care decisions consider relevant evidence about the costs and benefits of a treatment in a systematic way [39].

The studies mentioned focused primarily on the effect of digital technologies that have made an impact on decision-making. In general, new health technologies largely affected clinical and economic outcomes, and research methodologies to evaluate the efficiency of these new technologies were necessary. Cost-effectiveness analysis (CEA) is one way of doing this, which assesses the value or benefit of new health care technologies or interventions and compares the costs to a reference or threshold. In the United States, in the 1990s, the first and second panels on CEA included economists, ethicists, psychometricians, and clinicians who were asked to make reference case recommendations. The 1996 consensus report *Cost Effectiveness in Health and Medicine* was the first to describe the uses and conduct of CEA as a decision-making aid in the health and medical fields [40]. HTA was initially more actively developed in Europe in the 1970s, with both formal and informal initiatives in different countries [41]. So-called consensus conferences played an important role in the early development of HTA and had the explicit goal of using the best-available evidence in making policy, administrative, and clinical decisions in health care. By generating a broader interest in scientific evidence among policymakers, clinicians, and the general public, the conferences created a wider understanding of the need for comprehensive assessment in health care [42]. Although US policies have from there onward prevented the use of CEA for reimbursement decisions, such as with the Affordable Care Act of 2010, which banned Medicare from using CEA metrics, formal agencies were established in many individual European countries for HTA, starting with the first in Sweden in 1987, followed by many others in the 1990s.

n Australia, too, HTA of nonpharmaceutical technologies was developed in the 1970s and in 1990 for pharmaceuticals. It was established by the Commonwealth Parliament in the light of increasing costs of medical investigations and patient care [43]. Since almost 70% of total health expenditures in Australia are funded by government programs, including Medicare and the Pharmaceutical Benefits Scheme, evaluation was undertaken to consider the effects of developments in technology on medical benefits and public hospital costs, with some emphasis on diagnostic methods [43]. Economic evaluation is undertaken by or on behalf of the manufacturing industry, and the evidence on clinical effectiveness, safety, and cost-effectiveness presented is then considered by the Pharmaceutical Benefits Advisory Committee (PBAC).

Quality-Adjusted Life-Years

HTA commonly uses evaluation methods to assess whether the costs per quality-adjusted life-year (QALY) gained from a treatment, pharmaceuticals, or intervention are within the conventional range of acceptability. QALY is a standard measure of the disease burden that includes both the quality and the quantity of a human life and is used in economic and health policy evaluation to assess the value of medical interventions [44]. One QALY equals one year of perfect health. This differs from a year of life gained, which does not consider the quality of life (QoL).

QALYs have been in use for a half century, with criticisms of the approach nearly as old [45]. Among the many criticisms of QALYs are challenges with incorporating equity [46], challenges in relating QALYs to economic models of utility [47], and questions about the valuation and the basis for weighting of different outcomes [48]. Despite these criticisms, QALYs are the standard in many countries for HTA.

Most HTA-oriented countries base their reimbursement model on a threshold number per QALY saved. Cost-effectiveness thresholds (CETs) vary widely by country and are typically used to assess whether an intervention is worthwhile from a country-specific standpoint and often reflect the perceived health opportunity cost [49]. For example, in Australia, the PBAC is believed to apply a threshold range of AU $45,000-$60,000 per QALY [50], it is GBP 20,000-40,000 per QALY in the United Kingdom, and in many European countries, this is between euros 50,000 and 100,000 per QALY saved. CETs used by other decision makers (eg, the World Health Organization's suggested CET of 1-3 times the gross domestic product [GDP] per capita) do not. This way of calculating cost-effectiveness has been applied to various treatment options and conditions. Even internet-based interventions, such as those for harmful alcohol use, have been evaluated using this framework. A study evaluating the cost-effectiveness and cost utility of internet-based interventions for harmful use of alcohol [51],

XSL•FO
RenderX

through the assessment of the incremental cost-effectiveness of IT compared with information systems (IS), found that the median incremental cost-effectiveness ratio was estimated at euros 3683 per additional treatment responder and euros 14,710 per QALY gained. At a willingness to pay euros 20,000 for 1 additional QALY, IT had a 60% likelihood of being more cost effective than IS.

The quality adjustments in QALYs are based on various measurement methods and instruments. One key early debate was whether "health" could be measured as a single index value or whether health should be represented by a series of values representing different aspects of health, such as physical, mental, and emotional health. In the late 1980s, an interdisciplinary 5-country group developed the EuroQol instrument, a 5D, 3-level generic measure subsequently termed the "EQ-5D" [52]. It was designed to measure and value health status in terms of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [53]. The valuation method expanded usage across clinical programs, disease and condition areas, population surveys, patient-reported outcomes, and value sets.

The EQ-5D has become the standard measure for health and provides a generic measure with a single index value representing health. The EQ-5D was designed to measure decrements in health across conditions and populations. Digital versions have been developed over the years, validated [54], and applied to various questions [55-58] regarding cost-effectiveness of interventions and treatments.

US policymakers have consistently chosen not to use the costs per QALY framework for reimbursement decisions. Indeed, in both the Affordable Care Act and the recent legislation authorizing Medicare to negotiate drug prices, the use of QALYs for value assessment was explicitly prohibited [59]. In recent years, the United States did see steady growth in the number of organizations conducting value assessments. National organizations, such as the Institute for Clinical and Economic Review (ICER) [60], the Innovation and Value Initiative (IVI), and the National Comprehensive Cancer Network (NCCN), have introduced value assessment frameworks and tools to guide health care decision makers in evaluating the relative benefits and costs of health care interventions, primarily pharmaceuticals. According to a recent analysis, the field of value assessment in the United States is dynamic and evolving [61], and the use of value assessment is gaining traction with health care decision makers and policymakers. Overall, 79% of surveyed health care payers reported that ICER recommendations influenced their decision-making in 2022, compared to only 49% in 2016 [62,63].

As health care decision makers and policymakers in the United States increasingly use value assessment tools to inform their decision-making, there are a variety of different tools that can be used, with different approaches to measuring value, endpoints, and criteria for assessing the different approaches [59]. DiStefano et al [59] suggested that these criteria could include feasibility, flexibility, and the ability to incorporate factors beyond the traditional value elements. They also contrasted measures that use conventional (QALY) endpoints with those that incorporate patient-centric value elements.

Patient-centered methods will not necessarily always align well with the averages that HTA in Europe often uses to make decisions on the use of a technology at a group level. Differentiation in the use of technologies for different groups of patients may not only be desired but also be more affordable if it means that some groups of patients prefer cheaper or no treatment. Patient-centered care means patients should be part of the decision-making process, not only through providing information, but also through providing authority to make decisions about their treatment plans and care path. Indeed, this is one of the reasons patient advocates in the United States have (to date) successfully lobbied against the use of QALY measures [45].

One example of this could be the treatment of prostate cancer. Treating prostate cancer carries the risk of impedance and may not meaningfully impact life-years; however, not treating the cancer carries the risk that the cancer metastasizes. The "correct" treatment will need to weigh the patient's preferences for not being impendent versus the patient's willingness to accept a cancerous tumor growing in their body versus expected life expectancy. In this example, the "average" best course of treatment may not provide guidance for the best course of treatment *for a particular patient*. Today, many patients with prostate cancer are consulted about their options and the advantages and disadvantages of each treatment path and are allowed to select their preferred treatment path.

Another example would be single versus double mastectomy. The decision on whether to remove a healthy breast in a woman with a high risk of breast cancer is not simply about QALYs gained or lost; it has profound personal implications around body image anxiety and fear of recurrence [64].

Alternative CEA Approaches

There have been sporadic attempts to incorporate HTA into public health coverage decisions in the United States. The most well known is the attempt by the state of Oregon to use HTA principles to determine Medicaid coverage decisions. The state rejected standardized CEA and instead used a combination of expert opinion and public panels to create rank-ordered priorities. Oregon's attempt to use CEA failed because of a combination of factors. One was a lack of public acceptance of the priorities; one of the most visible elements was a child, Adam Howard, who died after he was denied a bone marrow transplant because of the combination of a low probability of success and high cost. The second factor was the Americans with Disability Act (ADA), which both Bush and Clinton administrations argued was violated by the use of HTA.

Although CEA has not been well accepted in the United States, there are newer approaches that seek to incorporate HTA into decision-making in the United States. CEA has been supplemented with approaches to assess value, including multicriteria decision analysis (MCDA), generalized risk-adjusted cost-effectiveness (GRACE) analysis, and the "value flower." MCDA is a decision-making method that systematically weighs various value elements that may fall

XSL•FO
RenderX

outside traditional value assessments, such as a treatment's scientific novelty, a patient's disease severity, or how a treatment may affect a caregiver's productivity [65]. MCDA models allow users to choose their willingness-to-pay thresholds and customize their value determination results using modifiable inputs for measures related to a treatment's benefits and cost. There are an increasing number of studies reporting that defining the value of health care depends heavily on the decision context and stakeholders involved [66]. Where cost-utility analysis and QALYs have become the method and value definition of choice for traditional value judgements in coverage and pricing decisions, other criteria that may influence value are often not measured and therefore omitted from value assessments or are only used to qualitatively contextualize assessments. Recent qualitative research shows that for payers, value equates either with criteria that provide tangible benefits (from their perspective), such as new treatment options that respond to a serious unmet need. For patients, a population-level value equates to options that would potentially benefit them in the future and the value of hope. One suggestion made by Campbell et al [67] is to include the equal value life-year (evLY) as a measure of health gain that can be used as an alternative or a complement to QALY to address concerns related to undervaluing treatments that extend the life of individuals with serious illness or chronic disability. Measurement methods no longer solely insist on linear, additive utility. A systematic literature review [68] identified 15 relevant value frameworks and MCDA tools. These studies included a large number (n=56) of individual value criteria. The most commonly included novel criteria were an unmet medical need, severity of disease, and reduction in uncertainty. The identified scoring functions (measurement methods) for novel criteria were highly heterogeneous and tailored. Standardized scoring functions were not observed.

One example is GRACE analysis, which is a somewhat different approach to CEA that aligns economic assessments of treatments with patient preferences and experience of care [69]. In GRACE analysis, differential CETs (relative to traditional CEA) are applied based on disease severity (eg, higher thresholds for more severe diseases) and other patient circumstances to better recognize the value of treatments that promote equity and significantly improve patient QoL [69]. This links value assessment more tightly to economic theory, recognizing that the utility value of a health improvement varies depending on the underlying level of health. This was one of the key problems with the use of HTA in Oregon: treatments with acceptable QALY values but low utility values, such as routine dental care, were prioritized over conditions with lower QALY values but (potentially) higher utility values, such as bone marrow transplants. Although the GRACE method has been described at length in the literature [70-73], studies are now underway applying it and comparing it with different approaches and outcomes.

In addition to addressing potential nonlinearity of utility functions, there have been several initiatives to expand the set of "value measures." A special task force of ISPOR, the professional society for health economics and outcomes research worldwide, created a "value flower" that identified elements of value that are and are not typically included in standard CEA and a recommendation to expand cost-effectiveness measures with these value elements to better capture what is important to individuals but unmeasured in standard QALYs [74]. In general, newer approaches to CEA now seek to include measures of productivity, real option value, insurance value, reduction in uncertainty, scientific spillover, severity of disease, adherence-improving factors, equity, and the value of hope. The issue with all these novel approaches, however, remains that the ultimate goal is to draw conclusions about *societal* costs and benefits rather than a blueprint for *individualized* health care delivery based on patient preferences and perceptions.

Policy Efforts and Future Opportunities

Patient Centeredness

In the policy arena, a series of important changes have been made regarding the patient's position in the decision-making process. For decades, there have been attempts to improve "patient centeredness" in health care, which is defined as "providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions" [75]. Although the focus on patient-centered care has increased, the rationale, measurement, and implementation of strategies to improve patient-centered care or to use patient experiences for quality improvement purposes have been widely debated [76,77]. In practice, patients are not directly and proactively involved as much in the process of quality improvement. Even if efforts to design care through the patient's eyes [78] are supported, the question remains to what extent this can be done at the individual patient level. Information and communication technologies have more recently shown to be able to support chronic disease self-management, but this self-management should be accompanied by a shift in focus on integrating the patient's own perceptions of the chronic disease and the health care system's approach to managing it [79]. As mentioned previously, we therefore need to use and improve methods to assess patients' perceptions, attitudes, beliefs, and preferences for health services and self-management.

Various patient-reported outcomes have since been developed and evaluated, with most studies concluding that health care professionals and patients alike need to choose the most suitable patient-reported outcomes for their patients, which may not be the same for every patient population or individual [80]. For health insurance, for example, the Health Plan Employer Data and Information Set (HEDIS) has been used to compare and report quality across health plans from a patient perspective, but in a number of cases, these do not appear to be cost effective or provide value for certain subgroups [81].

Shared Decision-Making

The same is true for the concept of shared decision-making, a process by which patients and providers consider outcome probabilities and patient preferences and reach a health care decision based on mutual agreement [82]. This is best used in situations in which there is medical uncertainty or different treatment options. Theoretically, both provider and patient discuss the different options and reach a conclusion together

XSL•FO
RenderX

about the optimal strategy. In the United States, for example, policy initiatives to improve shared decision-making were embedded in both the Affordable care Act (2010) and the patient-centered medical home, but their success depends on building a good relationship so that information is shared and patients are supported to deliberate and express their preferences and views during the decision-making process. Increasingly more (codesigned) decision aids have been developed in recent years to support shared decision-making and provide room for individual preference heterogeneity. These decisions aids are also being evaluated, with many studies concluding that when designing interventions to improve health outcomes, it is important to consider this heterogeneity. For example, a study looking at a decision aid for diabetes self-care found that teenagers with type I diabetes mellitus prioritize reducing family conflict and fitting into their social milieu over health outcomes at this time in their lives [83]. Another study focused on gaps in the current approach to shared decision-making for children with medical complexity suggested that clinicians should focus decision-making discussions on integrating each child's unique situation, the insights parents gain through their decision-making activity, and their clinical knowledge to enhance the understanding between parents and health care providers beyond the narrow concept of parental values [84].

In the United Kingdom, health authorities have engaged clinical champions and patient representatives in national initiatives for shared decision-making and embarked on a process of widely disseminating patient decision aids [85]. In practice, however, providers often find that the time required to go over all necessary information and decision aids in not available. Decision aids ultimately suffer from the same challenge as the concept of patient centeredness; there is not enough room in the clinical care pathway for approaches that incorporate individual preferences.

Future Opportunities for Value Assessment

Patient value is a key component in all areas of health care delivery, and understanding how health providers create, communicate, and deliver value to patients is a key factor when seeking ways to improve care according to the triple aim of improving the individual experience of care; improving the health of populations, and reducing the per capita costs of care for populations [86].

The practice of HTA uses a societal perspective to measure the cost-effectiveness of treatment in terms of costs and QALYs, where QALYs assume that a health state that is more desirable is more valuable, and therefore, value is equated with preference or desirability. This approach has major empirical and conceptual shortcomings, such as inconsistencies among values obtained from the standard gamble, time trade-off, and visual-analog scale elicitation formats and, more importantly, the linearity assumptions that violate the key economic assumption of diminishing marginal utility. Although HTA has come a long way in its "multifaceted assessment of the clinical, economic, ethical, legal, and societal perspectives that may be impacted by a new technology, procedure, drug, or process" [87], these approaches may not sufficiently reflect individual patient or health consumer preferences that may be important in the assessment of "value" in the current value-based health care discussion.

The key difference in measuring value is that the outcome in other fields is an economic and tangible measure, such as purchasing a car, whereas in health, it is the production of health. One way to better predict behavior and choice in health, therefore, is to include the health production function. Our recent work on medication nonadherence, for example, has shown that there is an enormous discrepancy between patient preferences and their beliefs and expectations [88]. In a pilot study using a double-bound contingent-belief (DBCB) questionnaire, patients could express how efficacy and side effects are affected by controlled levels of nonadherence, allowing for the estimation of sensitivity in health outcomes and costs. The derived health production function suggested that patients may strategically manage adherence to minimize side effects without compromising efficacy. Patients' inclination to manage medication intake was closely linked to the relative importance they assign to treatment efficacy and side effects. We formalized the definition of health production functions by positing that patients craft a mental model of health production under scenarios that are not covered by interactions with their physicians or by the clinical evidence available to them. This function relates the benefits of health behaviors to the costs associated with achieving such benefits. Value, then, is now defined not only by preferences for health outcomes alone but also by the patient's expectations of health outcomes. The "mental model" described in this paper is now being tested in different studies.

In addition to tangibility, a second key difference is that the production and consumption of a medical service take place simultaneously. This inseparability and the inconsistency in health that is being produced give much more room to interpretation, experiences, perceptions, and emotions when making a decision. In some cases, patients may be informed regarding clinical efficacy, but they may care more about another attribute of service or treatment that is entirely unrelated to clinical outcomes but not irrelevant to them. For example, a physician advises a patient to wait with treatment because the patient is at a risk of side effects at that point in time. The physician is attempting to provide high-quality care and act in the patient's best interest, while maintaining a reputation for delivering high-value care. This particular patient may choose immediate treatment because the gains outweigh the potential side effects and because the patient does not care about the physician's reputation. This creates a conflict between QALY-type HTA assessment and individual assessments of "value."

In a situation like this, who defines what "value-based" care means? If policymakers truly want to promote patient-centered care and shared decision-making, the question is what the desired outcome is in this situation. Reimbursement decisions are currently based on an equation that defines value in terms of health states. We know that this approach tells only one part of a story—what is preferable from a societal approach. If we accept the notion that all patients are different and have different

XSL•FO
RenderX

preferences, it must be acknowledged that "value" in health care needs to be redefined to incorporate patient preference heterogeneity.

Value assessment in health has come a long way in the past 25 years. Currently, health economic models focus on the benefits and costs of treatment options and help define reimbursement and delivery systems. However, this top-down approach needs to make room for "value-based" care, which considers value to individual patients. Value-based care ties the amount health care providers earn for their services to the value they deliver for their patients, such as the quality, equity, and cost of care. Through financial incentives and other methods, value-based care programs aim to hold providers more accountable for improving patient outcomes. However, in practice, so far, the focus is primarily on cost containment.

In this system, there is little attention given to variations in individual patient preferences or factors that affect those preferences. Many value-based health care systems still focus on cost containment strategies for the average patient, not the individual value to individual patients [89]. The idea is that some treatments, such as some imaging for low back pain, should not be provided (to anyone), while other treatments, such as vaccines, should be uniformly dispensed (to everyone). The one-size-fits-all approach to value disproportionally makes some patients worse off. In addition, delivery systems have differential effects on patient populations, which further enhances health inequities. Theoretically, it also increases spending, because all patients are expected to require a treatment for a weighted average cost. In practice, some will require higher-cost treatment options, while others will value lower-cost options. Care delivery that considers all aspects of human decision-making will help pave the way for more accurate cost-benefit analysis and improved access and affordability.

One lesson learned from 25 years of value assessment in health is that there is no average patient, yet many HTAs are still based on averages. That is why value assessment from a societal perspective may lead to equitable allocation of resources without necessarily leading to optimal use of resources. If individual patient behaviors and preferences are not considered in value assessment methods, resources will be wasted because of a mismatch between what the consumer values and what society prioritizes.

It will be challenging to consider the complexities of individual preferences and behaviors, especially if they are not met at the societal level. There has been an increase in the adoption of ideas from behavioral economics and mathematical psychology to better understand human decision-making in health. There are ways to add information about cognitive and brain processes that may help value assessment in health to perform more precise predictions of human decision-making and choice.

However, the final challenge for future value assessment will be to incorporate heterogeneity of preferences into decision models. Currently, heterogeneity is largely considered in terms of health equity and the impact of approval of denial of treatment on disadvantaged populations. However, the more challenging aspect will be trying to reconcile heterogeneous patient preferences with decision tools based on "average" patients. Centralized decision makers using average metrics, such as "incremental cost-effectiveness ratios," are reliant on the strong assumption of linearity of preferences, which is unlikely to hold in heterogeneous/diverse societies. This means that centralized planners using decision rules based on measures such as the average cost per QALY gained will misallocate resources and fail to use health care dollars to maximize the well-being of society.

Conclusion

Where ultimately the current energy toward developing new measures of value will lead is unclear. Clearly, patients are interested in better measures of value, but whether any of the new measures that have been developed will be meaningful and useful to consumers is unknown and a fruitful area for future research. Similarly, patient advocates will tend toward broader measures of value, but whether the rank-ordering of treatments will ultimately change is unknown—and another fruitful area for future research. Finally, how central payers will incorporate the new measures into their decision-making is perhaps the biggest unknown of all. Research will need to demonstrate the value of the new measures in providing better measures of value that meaningfully change health spending priorities.

Conflicts of Interest

None declared.

References

1. Moloney TW, Paul B. The consumer movement takes hold in medical care. Health Aff (Millwood) 1991 Jan;10(4):268-279. [doi: 10.1377/hlthaff.10.4.268] [Medline: 1778561]
2. Mendelson A, Kondo K, Damberg C, Low A, Motúapuaka M, Freeman M, et al. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. Ann Intern Med 2017 Jan 10;166(5):341. [doi: 10.7326/m16-1881]
3. Sinaiko AD. What is the value of market-wide health care price transparency? JAMA 2019 Oct 15;322(15):1449-1450. [doi: 10.1001/jama.2019.11578] [Medline: 31486830]
4. deBronkart D. Open access as a revolution: knowledge alters power. J Med Internet Res 2019 Dec 11;21(12):e16368 [FREE Full text] [doi: 10.2196/16368] [Medline: 31825321]

XSL•FO
RenderX

5. Miller GE, Vistnes JP, Rohde F, Keenan PS. High-deductible health plan enrollment increased from 2006 to 2016, employer-funded accounts grew in largest firms. Health Aff (Millwood) 2018 Aug;37(8):1231-1237. [doi: 10.1377/hlthaff.2018.0188] [Medline: 30080451]
6. Zhang X, Haviland A, Mehrotra A, Huckfeldt P, Wagner Z, Sood N. Does enrollment in high‐deductible health plans encourage price shopping? Health Serv Res 2018 Aug 23;53 Suppl 1(Suppl Suppl 1):2718-2734 [FREE Full text] [doi: 10.1111/1475-6773.12784] [Medline: 29058316]
7. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 4;6(2):e19106. [doi: 10.2196/19106]
8. von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton, NJ: Princeton University Press; 1944.
9. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979 Mar;47(2):263. [doi: 10.2307/1914185]
10. Venkatachalam L. The contingent valuation method: a review. Environ Impact Assess 2004 Jan;24(1):89-124. [doi: 10.1016/s0195-9255(03)00138-0]
11. Smith V, Desvousges W, Smith V, Desvousges W. The contingent ranking method and benefit estimation. Meas Water Qual Benef 1986:145-180. [doi: 10.1007/978-94-009-4223-3_6]
12. Payne JW. Contingent decision behavior. Psychol Bull 1982;92(2):382-402. [doi: 10.1037//0033-2909.92.2.382]
13. Christensen R, David HA. The method of paired comparisons. Technometrics 1989 Nov;31(4):495. [doi: 10.2307/1270014]
14. Kendall MG, Smith BB. On the method of paired comparisons. Biometrika 1940 Mar;31(3/4):324. [doi: 10.2307/2332613]
15. Louviere JJ, Flynn TN, Carson RT. Discrete choice experiments are not conjoint analysis. J Choice Model 2010;3(3):57-72. [doi: 10.1016/s1755-5345(13)70014-9]
16. Louviere JJ. Conjoint analysis. In: Bagozzi RP, editor. Advanced Methods of Marketing Research. Cambridge, MA: Blackwell Business; 1994:223-259.
17. Green PE, Krieger AM, Agarwal MK. Adaptive conjoint analysis: some caveats and suggestions. J Mark Res 1991 May;28(2):215. [doi: 10.2307/3172809]
18. Johnson RM. Comment on "adaptive conjoint analysis: some caveats and suggestions". J Mark Res 1991 May;28(2):223. [doi: 10.2307/3172810]
19. Agarwal MK, Green PE. Adaptive conjoint analysis versus selfexplicated models: some empirical results. Int J Res Mark 1991 Jun;8(2):141-146. [doi: 10.1016/0167-8116(91)90021-x]
20. Huber J, Wittink DR, Fiedler JA, Miller R. The effectiveness of alternative preference elicitation procedures in predicting choice. J Mark Res 1993 Feb;30(1):105. [doi: 10.2307/3172517]
21. Tversky A. Intransitivity of preferences. Psychol Rev 1969 Jan;76(1):31-48. [doi: 10.1037/h0026750]
22. Dolan P, Robinson A. The measurement of preferences over the distribution of benefits: the importance of the reference point. Eur Econ Rev 2001 Oct;45(9):1697-1709. [doi: 10.1016/s0014-2921(00)00052-0]
23. Munro A, Sugden R. On the theory of reference-dependent preferences. J Econ Behav Organ 2003 Apr;50(4):407-428. [doi: 10.1016/s0167-2681(02)00033-1]
24. Hensher DA, Beck MJ, Rose JM. Accounting for preference and scale heterogeneity in establishing whether it matters who is interviewed to reveal household automobile purchase preferences. Environ Resource Econ 2010 Oct 23;49(1):1-22. [doi: 10.1007/s10640-010-9420-3]
25. Hensher DA. Attribute processing, heuristics and preference construction in choice analysis. In: Choice Modelling: The State-of-the-Art and The State-of-Practice. Bingley, UK: Emerald Group Publishing; 2010:35-69.
26. Greene W, Hensher D. Ordered choices and heterogeneity in attribute processing. JTEP 2010 Sep 1;44(3):331-364. [doi: 10.4337/9781788970204.00009]
27. Czajkowski M, Giergiczny M, Greene WH. Learning and fatigue effects revisited: investigating the effects of accounting for unobservable preference and scale heterogeneity. Land Econ 2014 Mar 24;90(2):324-351. [doi: 10.3368/le.90.2.324]
28. Orquin JL, Mueller Loose S. Attention and choice: a review on eye movements in decision making. Acta Psychol (Amst) 2013 Sep;144(1):190-206. [doi: 10.1016/j.actpsy.2013.06.003] [Medline: 23845447]
29. Peterson G, Aslani P, Williams KA. How do consumers search for and appraise information on medicines on the internet? A qualitative study using focus groups. J Med Internet Res 2003 Dec 19;5(4):e33 [FREE Full text] [doi: 10.2196/jmir.5.4.e33] [Medline: 14713661]
30. Hansen DL, Derry HA, Resnick PJ, Richardson CR. Adolescents searching for health information on the internet: an observational study. J Med Internet Res 2003 Oct 17;5(4):e25 [FREE Full text] [doi: 10.2196/jmir.5.4.e25] [Medline: 14713653]
31. Evans R, Joseph-Williams N, Edwards A, Newcombe RG, Wright P, Kinnersley P, et al. Supporting informed decision making for prostate specific antigen (PSA) testing on the web: an online randomized controlled trial. J Med Internet Res 2010 Aug 06;12(3):e27 [FREE Full text] [doi: 10.2196/jmir.1305] [Medline: 20693148]

XSL•FO
RenderX

32. Webster R, Li SC, Sullivan DR, Jayne K, Su SY, Neal B. Effects of internet-based tailored advice on the use of cholesterol-lowering interventions: a randomized controlled trial. J Med Internet Res 2010 Sep 13;12(3):e42 [FREE Full text] [doi: 10.2196/jmir.1364] [Medline: 20837464]
33. Diefenbach MA, Mohamed NE, Butz BP, Bar-Chama N, Stock R, Cesaretti J, et al. Acceptability and preliminary feasibility of an internet/CD-ROM-based education and decision program for early-stage prostate cancer patients: randomized pilot study. J Med Internet Res 2012 Jan 13;14(1):e6 [FREE Full text] [doi: 10.2196/jmir.1891] [Medline: 22246148]
34. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010 Dec 19;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
35. Smit F, Lokkerbol J, Riper H, Majo MC, Boon B, Blankers M. Modeling the cost-effectiveness of health care systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness. J Med Internet Res 2011 Aug 11;13(3):e56 [FREE Full text] [doi: 10.2196/jmir.1694] [Medline: 21840836]
36. van Spijker BA, Majo MC, Smit F, van Straten A, Kerkhof AJ. Reducing suicidal ideation: cost-effectiveness analysis of a randomized controlled trial of unguided web-based self-help. J Med Internet Res 2012 Oct 26;14(5):e141 [FREE Full text] [doi: 10.2196/jmir.1966] [Medline: 23103835]
37. Lintvedt OK, Griffiths KM, Eisemann M, Waterloo K. Evaluating the translation process of an internet-based self-help intervention for prevention of depression: a cost-effectiveness analysis. J Med Internet Res 2013 Jan 23;15(1):e18 [FREE Full text] [doi: 10.2196/jmir.2422] [Medline: 23343481]
38. Chen Y, Lin Y, Hung C, Huang C, Yeih D, Chuang P, et al. Clinical outcome and cost-effectiveness of a synchronous telehealth service for seniors and nonseniors with cardiovascular diseases: quasi-experimental study. J Med Internet Res 2013 Apr 24;15(4):e87 [FREE Full text] [doi: 10.2196/jmir.2091] [Medline: 23615318]
39. Banta D. What is technology assessment? Int J Technol Assess Health Care 2009 Jul 01;25(S1):7-9. [doi: 10.1017/s0266462309090333]
40. Gold MR. Cost-Effectiveness in Health and Medicine. Oxford, UK: Oxford University Press; 1996.
41. Fuchs VR. Who Shall Live? Health, Economics and Social Choice. Singapore: World Scientific Publishing; 2011.
42. Banta D, Kristensen FB, Jonsson E. A history of health technology assessment at the European level. Int J Technol Assess Health Care 2009 Jul 01;25(S1):68-73. [doi: 10.1017/s0266462309090448]
43. Hailey D. The history of health technology assessment in Australia. Int J Technol Assess Health Care 2009 Jul 01;25(S1):61-67. [doi: 10.1017/s0266462309090436]
44. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health 2009 Mar;12 Suppl 1:S5-S9 [FREE Full text] [doi: 10.1111/j.1524-4733.2009.00515.x] [Medline: 19250132]
45. Devlin NJ, Drummond MF, Mullins CD. Quality-adjusted life years, quality-adjusted life-year-like measures, or neither? The debate continues. Value Health 2024 Jun;27(6):689-691. [doi: 10.1016/j.jval.2024.04.021] [Medline: 38705459]
46. Wagstaff A. QALYs and the equity-efficiency trade-off. J Health Econ 1991 May;10(1):21-41. [doi: 10.1016/0167-6296(91)90015-f] [Medline: 10113661]
47. Brouwer WB, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions!. J Health Econ 2000 Jul;19(4):439-459. [doi: 10.1016/s0167-6296(99)00038-7] [Medline: 11010234]
48. Brouwer WB, Culyer AJ, van Exel NJA, Rutten FF. Welfarism vs. extra-welfarism. J Health Econ 2008 Mar;27(2):325-338. [doi: 10.1016/j.jhealeco.2007.07.003] [Medline: 18179835]
49. Sax S. Report of the Committee on Applications and Costs of Modern Technology in Medical Practice. Canberra, Australia: Australian Government Publishing Service; 1978.
50. Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. Value Health 2016 Dec;19(8):929-935 [FREE Full text] [doi: 10.1016/j.jval.2016.02.017] [Medline: 27987642]
51. Blankers M, Nabitz U, Smit F, Koeter MW, Schippers GM. Economic evaluation of internet-based interventions for harmful alcohol use alongside a pragmatic randomized controlled trial. J Med Internet Res 2012 Oct 29;14(5):e134 [FREE Full text] [doi: 10.2196/jmir.2052] [Medline: 23103771]
52. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec 9;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
53. Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. Appl Health Econ Health Policy 2017 Apr 13;15(2):127-137 [FREE Full text] [doi: 10.1007/s40258-017-0310-5] [Medline: 28194657]
54. Hernandez G, Garin O, Dima AL, Pont A, Martí Pastor M, Alonso J, ASTRO-LAB Group. EuroQol (EQ-5D-5L) validity in assessing the quality of life in adults with asthma: cross-sectional study. J Med Internet Res 2019 Jan 23;21(1):e10178 [FREE Full text] [doi: 10.2196/10178] [Medline: 30672744]
55. Altunkaya J, Craven M, Lambe S, Beckley A, Rosebrock L, Dudley R, et al. Estimating the economic value of automated virtual reality cognitive therapy for treating agoraphobic avoidance in patients with psychosis: findings from the gameChange randomized controlled clinical trial. J Med Internet Res 2022 Nov 18;24(11):e39248 [FREE Full text] [doi: 10.2196/39248] [Medline: 36399379]

XSL•FO
RenderX

56. Wu Y, Li X, Zhou Y, Gao R, Wang K, Ye H, et al. Efficacy and cost-effectiveness analysis of internet-based cognitive behavioral therapy for obsessive-compulsive disorder: randomized controlled trial. J Med Internet Res 2023 May 24;25:e41283 [FREE Full text] [doi: 10.2196/41283] [Medline: 37223977]
57. Hsieh H, Kao C, Cheng S, Chang Y. A web-based integrated management program for improving medication adherence and quality of life, and reducing readmission in patients with atrial fibrillation: randomized controlled trialntrolled Trial. J Med Internet Res 2021 Sep 22;23(9):e30107 [FREE Full text] [doi: 10.2196/30107] [Medline: 34550084]
58. Muschol J, Heinrich M, Heiss C, Hernandez AM, Knapp G, Repp H, et al. Digitization of follow-up care in orthopedic and trauma surgery with video consultations: health economic evaluation study from a health provider's perspective. J Med Internet Res 2023 Dec 25;25:e46714. [doi: 10.2196/46714] [Medline: 38145481]
59. DiStefano MJ, Zemplenyi A, Anderson KE, Mendola ND, Nair KV, McQueen RB. Alternative approaches to measuring value: an update on innovative methods in the context of the United States Medicare drug price negotiation program. Expert Rev Pharmacoecon Outcomes Res 2024 Feb 17;24(2):171-180. [doi: 10.1080/14737167.2023.2283584] [Medline: 37961908]
60. Wang S, Gum D, Merlin T. Comparing the ICERs in medicine reimbursement submissions to NICE and PBAC—does the presence of an explicit threshold affect the ICER proposed? Value Health 2018 Aug;21(8):938-943 [FREE Full text] [doi: 10.1016/j.jval.2018.01.017] [Medline: 30098671]
61. Ortman E, Westrich K, Buelt L. Driving progress in U.S. value assessment. PhRMA Foundation. 2023 Jan. URL: https://www.phrmafoundation.org/wp-content/uploads/2023/08/Driving-Progress-in-U.S.-Value-Assessment_PhRMA-Foundation.pdf [accessed 2024-12-17]
62. Xcenda. Managed Care Trends Report. Data on File. Conshohocken, PA: Xcenda; Jul 2022.
63. Choi M, Hydery T, Tan R, Tennant L. Trends in the uptake and impact of the Institute for Clinical and Economic Review value assessment framework in payer coverage decisions from 2016 to 2020. 2021 Presented at: AMCP 2021: Academy of Managed Care Pharmacy Virtual Meeting; April 12-16, 2021; Virtual.
64. Pesce C, Jaffe J, Kuchta K, Yao K, Sisco M. Patient-reported outcomes among women with unilateral breast cancer undergoing breast conservation versus single or double mastectomy. Breast Cancer Res Treat 2021 Jan 08;185(2):359-369. [doi: 10.1007/s10549-020-05964-0] [Medline: 33033966]
65. Multi-criteria decision analysis (MCDA). PhRMA Foundation. URL: https://www.phrmafoundation.org/research-areas/value-assessment-health-outcomes-research/multi-criteria-decision-analysis-mcda/ [accessed 2024-12-17]
66. Jakab I, Whittington MD, Franklin E, Raiola S, Campbell JD, Kaló Z, et al. Patient and payer preferences for additional value criteria. Front Pharmacol 2021 Jun 24;12:690021 [FREE Full text] [doi: 10.3389/fphar.2021.690021] [Medline: 34248638]
67. Campbell JD, Whittington MD, Pearson SD. An alternative measure of health for value assessment: the equal value life-year. Pharmacoeconomics 2023 Oct 17;41(10):1175-1182. [doi: 10.1007/s40273-023-01302-6] [Medline: 37458912]
68. Zelei T, Mendola ND, Elezbawy B, Németh B, Campbell JD. Criteria and scoring functions used in multi-criteria decision analysis and value frameworks for the assessment of rare disease therapies: a systematic literature review. Pharmacoecon Open 2021 Dec 18;5(4):605-612 [FREE Full text] [doi: 10.1007/s41669-021-00271-w] [Medline: 34003484]
69. Lakdawalla DN, Phelps CE. Generalized risk adjusted cost effectiveness (GRACE): ensuring patient-centered outcomes in healthcare decision making. Value & Outcomes Spotlight 2021 Sep;7(suppl 2):S5-S10.
70. Lakdawalla DN, Phelps CE. A guide to extending and implementing generalized risk-adjusted cost-effectiveness (GRACE). Eur J Health Econ 2022 Apr 08;23(3):433-451 [FREE Full text] [doi: 10.1007/s10198-021-01367-0] [Medline: 34495445]
71. Lakdawalla DN, Phelps CE. The generalized risk-adjusted cost-effectiveness (GRACE) model for measuring the value of gains in health: an exact formulation. J. Benefit Cost Anal 2023 Apr 03;14(1):44-67. [doi: 10.1017/bca.2023.6]
72. Lakdawalla DN, Phelps CE. Health technology assessment with diminishing returns to health: the generalized risk-adjusted cost-effectiveness (GRACE) approach. Value Health 2021 Feb;24(2):244-249 [FREE Full text] [doi: 10.1016/j.jval.2020.10.003] [Medline: 33518031]
73. Phelps CE, Lakdawalla DN. Valuing Health: The Generalized and Risk-Adjusted Cost-Effectiveness (GRACE) Model. Oxford, UK: Oxford University Press; 2024.
74. Neumann PJ, Garrison LP, Willke RJ. The history and future of the "ISPOR value flower": addressing limitations of conventional cost-effectiveness analysis. Value Health 2022 Apr;25(4):558-565. [doi: 10.1016/j.jval.2022.01.010] [Medline: 35279370]
75. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
76. Groene O. Patient centredness and quality improvement efforts in hospitals: rationale, measurement, implementation. Int J Qual Health Care 2011 Oct 22;23(5):531-537. [doi: 10.1093/intqhc/mzr058] [Medline: 21862449]
77. Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. BMJ Qual Saf 2016 Aug 18;25(8):626-632 [FREE Full text] [doi: 10.1136/bmjqs-2015-004839] [Medline: 26993640]
78. Edgman-Levitan S, Schoenbaum SC. Patient-centered care: achieving higher quality by designing care through the patient's eyes. Isr J Health Policy Res 2021 Mar 05;10(1):21 [FREE Full text] [doi: 10.1186/s13584-021-00459-9] [Medline: 33673875]

XSL•FO
RenderX

79. Wildevuur S, Thomese F, Ferguson J, Klink A. Information and communication technologies to support chronic disease self-management: preconditions for enhancing the partnership in person-centered care. J Particip Med 2017 Jul 07;9(1):e14 [FREE Full text] [doi: 10.2196/jopm.8846] [Medline: 36262005]
80. Perry MB, Taylor S, Khatoon B, Vercell A, Faivre-Finn C, Velikova G, et al. Examining the effectiveness of electronic patient-reported outcomes in people with cancer: systematic review and meta-analysis. J Med Internet Res 2024 Jul 31;26:e49089 [FREE Full text] [doi: 10.2196/49089] [Medline: 39083791]
81. Neumann PJ, Levine B. Do HEDIS measures reflect cost-effective practices? Am J Prev Med 2002 Nov;23(4):276-289. [doi: 10.1016/s0749-3797(02)00516-0] [Medline: 12406482]
82. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct 23;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
83. Hannon TS, Moore CM, Cheng ER, Lynch DO, Yazel-Smith LG, Claxton GE, et al. Codesigned shared decision-making diabetes management plan tool for adolescents with type 1 diabetes mellitus and their parents: prototype development and pilot test. J Particip Med 2018 May 10;10(2):e8 [FREE Full text] [doi: 10.2196/jopm.9652] [Medline: 33052121]
84. Buchanan F, Lai C, Cohen E, Milo-Manson G, Shachak A. Decision-making for parents of children with medical complexities: activity theory analysis. J Particip Med 2022 Jan 17;14(1):e31699. [doi: 10.2196/31699] [Medline: 35037890]
85. Coulter A, Edwards A, Elwyn G, Thomson R. Implementing shared decision making in the UK. Z Evid Fortbild Qual Gesundhwes 2011 Jan;105(4):300-304. [doi: 10.1016/j.zefq.2011.04.014] [Medline: 21620325]
86. Whittington JW, Nolan K, Lewis N, Torres T. Pursuing the triple aim: the first 7 years. Milbank Q 2015 Jun 04;93(2):263-300. [doi: 10.1111/1468-0009.12122] [Medline: 26044630]
87. ICER guide to understanding health technology assessment. ICER. 2018. URL: http://icer-review.org/wp-content/uploads/2018/08/ICER-Guide-to-Understanding-Health-Technology-Assessment-6.19.18.pdf [accessed 2024-12-17]
88. Muiruri C, van den Broek-Altenburg E, Bosworth H, Cené C, Gonzalez J. A quantitative framework for medication non-adherence: integrating patient treatment expectations and preferences. PPA 2023 Dec;17:3135-3145. [doi: 10.2147/ppa.s434640]
89. Niven DJ, Mrklas KJ, Holodinsky JK, Straus SE, Hemmelgarn BR, Jeffs LP, et al. Towards understanding the de-adoption of low-value clinical practices: a scoping review. BMC Med 2015 Oct 06;13(1):255 [FREE Full text] [doi: 10.1186/s12916-015-0488-z] [Medline: 26444862]

Abbreviations

ACA: adaptive conjoint analysis
CEA: cost-effectiveness analysis
CET: cost-effectiveness threshold
EUT: expected utility theory
GRACE: generalized risk-adjusted cost-effectiveness
HTA: health technology assessment
ICER: Institute for Clinical and Economic Review
IS: information systems
MCDA: multicriteria decision analysis
PBAC: Pharmaceutical Benefits Advisory Committee
PSA: prostate-specific antigen
PT: prospect theory
QALY: quality-adjusted life-year
QoL: quality of life

Edited by G Eysenbach; submitted 14.05.24; peer-reviewed by J Siverskog, J Kullgren, J Juusola; comments to author 30.08.24; revised version received 20.09.24; accepted 03.12.24; published 10.01.25.

Please cite as:
van den Broek-Altenburg EM, Atherly AJ
The Paradigm Shift From Patient to Health Consumer: 20 Years of Value Assessment in Health
J Med Internet Res 2025;27:e60443
URL: https://www.jmir.org/2025/1/e60443
doi:10.2196/60443
PMID:39793021

XSL•FO
RenderX

©Eline M van den Broek-Altenburg, Adam J Atherly. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Mapping the Landscape of Digital Health Intervention Strategies: 25-Year Synthesis

Shiyu Liu[1*], BSc, MA, PhD; Jingru Ma[1*], MBBS; Meichen Sun[1], BSc; Chao Zhang[2], BSc, MA; Yujing Gao[3], BM, BA; Jinghong Xu[4,5], BA, MA, PhD

[1]School of Public Health, Xi'an Jiaotong University, Xi'an, China

[2]School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[3]School of Journalism and Cultural Communication, Zhongnan University of Economics and Law, Wuhan, China

[4]School of Journalism and Communication, Beijing Normal University, Beijing, China

[5]The International College, Krirk University, Bangkok, Thailand

[*]these authors contributed equally

Corresponding Author:
Jinghong Xu, BA, MA, PhD
School of Journalism and Communication
Beijing Normal University
No 19, Xinjiekouwai St, Haidian District, Beijing, China
Beijing, 100875
China
Phone: 86 15810320711
Email: 123abctg@163.com

Abstract

Background: Digital health interventions have emerged as promising tools to promote health behavior change and improve health outcomes. However, a comprehensive synthesis of strategies contributing to these interventions is lacking.

Objective: This study aims to (1) identify and categorize the strategies used in digital health interventions over the past 25 years; (2) explore the differences and changes in these strategies across time periods, countries, populations, delivery methods, and senders; and (3) serve as a valuable reference for future researchers and practitioners to improve the effectiveness of digital health interventions.

Methods: This study followed a systematic review approach, complemented by close reading and text coding. A comprehensive search for published English academic papers from PubMed, Web of Science, and Scopus was conducted. The search employed a combination of digital health and intervention-related terms, along with database-specific subject headings and filters. The time span covered 25 years, from January 1, 1999, to March 10, 2024. Sample papers were selected based on study design, intervention details, and strategies. The strategies were identified and categorized based on the principles of Behavior Change Techniques and Behavior Strategies.

Results: A total of 885 papers involving 954,847 participants met the eligibility criteria. We identified 173 unique strategies used in digital health interventions, categorized into 19 themes. The 3 most frequently used strategies in the sample papers were "guide" (n=492, 55.6%), "monitor" (n=490, 55.4%), and "communication" (n=392, 44.3%). The number of strategies employed in each paper ranged from 1 to 32. Most interventions targeted clients (n=844, 95.4%) and were carried out in hospitals (n=268, 30.3%). High-income countries demonstrated a substantially higher number and diversity of identified strategies than low- and middle-income countries, and the number of studies targeting the public (n=647, 73.1%) far exceeded those focusing on vulnerable groups (n=238, 26.9%).

Conclusions: Digital health interventions and strategies have undergone considerable development over the past 25 years. They have evolved from simple approaches to sophisticated, personalized techniques and are trending toward multifaceted interventions, leveraging advanced technologies for real-time monitoring and feedback. Future studies should focus on rigorous evaluations, long-term effectiveness, and tailored approaches for diverse populations, and more attention should be given to vulnerable groups.

(J Med Internet Res 2025;27:e59027) doi:10.2196/59027

XSL•FO
RenderX

KEYWORDS

digital health interventions; intervention strategies; behavior change; mHealth; eHealth; randomized controlled trial

Introduction

The Rise of Digital Health Interventions

The rapid advancement of digital technologies has significantly influenced health practices. As of 2020, digital health has more than 90 definitions, encompassing eHealth, mHealth, self-tracking, wearable devices, and artificial intelligence applications [1]. Digital health interventions, defined as interventions delivered via digital health technologies, have emerged as promising tools for promoting healthy behaviors and improving health outcomes [2]. The effectiveness of digital health interventions has been validated across diverse domains, populations, and countries. Specifically, digital health interventions have demonstrated success in multiple health areas, such as noncommunicable disease management [3], mental health support [4], and smoking cessation programs [5]. In addition, these interventions have proven effective in serving not only the general population but also vulnerable groups, such as resource-poor communities and ethnic minorities [6]. Notably, their application extends across geographical and economic boundaries, showing positive results in both high-income nations, such as the United States [7] and Canada [8], and resource-limited regions, including Pakistan [9] and West Africa [10].

Driven by technological innovations, increasing smartphone penetration, and the growing acceptance of digital health technologies, health care delivery is experiencing a significant transformation from traditional processes to digital interventions [11]. The COVID-19 pandemic has further accelerated this transformation [12]. As health care systems globally confront challenges, including aging populations, rising chronic disease burdens, and disparities in health care access, the adoption and integration of digital health interventions have become increasingly crucial [13]. Moreover, recent advances in artificial intelligence have further enhanced the importance and impact of digital health interventions.

The Critical Role of Strategies in Digital Health Interventions

Strategies are indispensable for generating the desired effects of digital health interventions. Selecting and implementing appropriate strategies are crucial for ensuring their success [14]. Well-designed strategies have the potential to enhance user engagement, promote behavior change, and contribute to improved health outcomes [15,16], although their effectiveness may vary across different populations and contexts. Through the application of well-defined, theory-based strategies, digital health interventions can provide enhanced support for individuals in achieving their health goals, ultimately leading to more sustainable health care approaches [17].

Strategic approaches play a crucial role throughout the intervention process, from initial design to implementation. For example, message strategies, which form the core of the intervention design phase and are based on principles such as reciprocity, reciprocity-by-proxy, and curiosity, have proven effective in engaging racial and ethnic minority groups in digital oral self-care interventions [18]. During the implementation process, strategies can help overcome barriers to achieving intervention goals through four key approaches: (1) increasing coherence (ensuring the intervention is meaningful to users); (2) enhancing cognitive participation (securing user engagement); (3) promoting collective action (supporting the implementation process); and (4) enabling reflexive monitoring (facilitating an assessment of the intervention's effects) [19].

Current Findings on Digital Health Intervention Strategies

Currently, several studies, including systematic reviews and meta-analyses, have explored the impact of various strategies on health behaviors and outcomes in digital health interventions. Saleem et al [20] found that engagement strategies including feedback, guidance, social support, content gamification, reminders, flexibility, and ease of use can promote mental health outcomes. Behavior change strategies including follow-up prompts, self-monitoring, emotional control training, and provision of information about others' approval were widely used to contribute to health behavior [21].

A recent meta-analysis for weight management interventions found that strategies like problem-solving, goal setting, reviewing goals, feedback on behavior, self-monitoring of behavior, behavioral substitution, and credible sources were successful in addressing decreased energy [22]. The strategies of sustained motivation, self-regulation, psychological and physical resources, habit formation, and environmental and social influences were valued by weight loss maintenance intervention [23,24]. Four strategies, including improving social skills, enhancing social support, increasing social contact, and addressing cognition, have been identified to reduce loneliness [25]. Eleven intervention strategies, primarily involving medication and diet control, including physical activity, lipid-affecting drugs, and diet plus exercise, have been identified to prevent type 2 diabetes mellitus in China [26].

In addition, substantial strategies have been found to be effective in improving recruitment, reducing loss to follow-up, and enhancing retention [27-30]. A total of 72 strategies were tested to enhance participant recruitment in randomized trials, with 7 of these strategies being investigated in multiple studies [31]. Six broad types of strategies, including incentives, communication strategies, new questionnaire formats, participant case management, behavioral interventions, and methodological interventions, were summarized as effective approaches to improve retention [32]. Robinson et al [33] abstracted 368 strategies from 21 studies and identified 12 themes aimed at retaining participants. They found that contact and scheduling were the most common strategies to limit participant attrition in health care research.

XSL•FO
RenderX

Study Objectives

The previous research on digital health intervention strategies [23-34], despite its comprehensiveness, has significant limitations, notably its narrow focus on specific health issues or populations and its inadequate exploration of behavior change. Moreover, numerous nuanced strategies remain unidentified and analyzed, requiring further investigation. Accordingly, it is crucial to identify, categorize, and aggregate strategies used in digital health interventions. As the field of digital health continues to evolve, a deeper understanding of intervention strategies will be vital for advancing both research and practice, offering valuable insights into their effectiveness and guiding the design of future interventions. By identifying and categorizing these strategies, researchers and practitioners can gain a clearer understanding of effective approaches and techniques in digital health interventions, evaluating the strengths and limitations of different strategies and their relevance to specific health domains or populations. Furthermore, a comprehensive understanding of digital health intervention strategies can support the development of evidence-based guidelines and best practices, fostering the standardization and quality of digital health interventions.

Therefore, this study aimed to (1) identify and categorize the strategies employed in digital health interventions over the past quarter century; (2) explore the differences and changes in these strategies across time periods, countries, populations, deliveries, and senders; and (3) develop a catalog of strategies to serve as a reference for future researchers and practitioners to improve the effectiveness of digital health interventions.

Methods

Overview

This article followed a systematic review approach, complemented by close reading and text coding. The systematic review was used to access and screen relevant literature, while textual close reading was employed to identify and categorize strategies. The identification and categorization of strategies were guided by the principles and catalogs of Behavior Change Techniques (BCTs) [35] and Behavior Strategies (BSs) [22]. This dual approach ensures a comprehensive and rigorous analysis of the strategies used in digital health interventions.

Search Methods

Our research steps were guided by PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Multimedia Appendix 1) [36]. We conducted a comprehensive search for published English papers from PubMed, Web of Science, and Scopus. These databases are globally esteemed for their broad interdisciplinary coverage, sophisticated search tools, and current updates, which ensure systematic reviews are comprehensive, accurate, and authoritative [37]. Based on these characteristics, we selected these 3 databases for the following reasons: (1) broad coverage: these databases offer extensive coverage of medical and health-related research, enabling us to capture a wide range of relevant studies; (2) interdisciplinary nature: the selected databases cover related fields such as psychology, public health, and information technology; (3) quality and reliability: these databases are recognized for their rigorous inclusion criteria and high-quality indexing.

This search employed a combination of digital health and intervention-related terms, along with database-specific subject headings and filters, to ensure thoroughness and focus. The detailed search strategy for each database can be found in Multimedia Appendix 2. The time span considered was the past 25 years, from January 1, 1999, to March 10, 2024.

Selection of Sample Papers

We conducted 2 rounds of screening and selection. The first round focused on identifying digital health interventions, while the second round aimed to pinpoint specific strategies within those interventions.

For the first round of screening, pairs of reviewers (authors JM, MS, YG, NZ, ML, and HD), all graduate students majoring in public health or health communication, independently reviewed the titles, abstracts, and keywords of the sample articles to preliminarily identify studies that met our eligibility criteria. The specific screening criteria and steps were as follows: (1) the means of the intervention should be digital, primarily including wearable devices, telemedicine, electronic health records, electronic medical records, mobile phone applications, web pages, blogs, emails, text messages, social media, and similar technologies; (2) the intervention must be health-related, encompassing health behavior improvement, disease treatment, and health education, and so on. (3) the studies must be randomized controlled trials (RCTs), including individual and cluster RCTs, as they provide the highest level of evidence for evaluating interventions. We excluded randomized feasibility and pilot trials because they are inherently preliminary studies conducted to inform future RCTs [38].

In the second round of screening, we concentrated on identifying specific strategies within the selected digital health interventions. For this study, a strategy was defined as a specific approach or technique employed within the intervention to promote health behavior change or improve health outcomes. The second round of screening involved a full-text examination of the included sample papers. The reviewers (JM, MS, YG, NZ, ML, and HD) assessed the descriptions of the interventions and the strategies employed. The flow of literature searches and screening is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Flow diagram of literature search and screening.

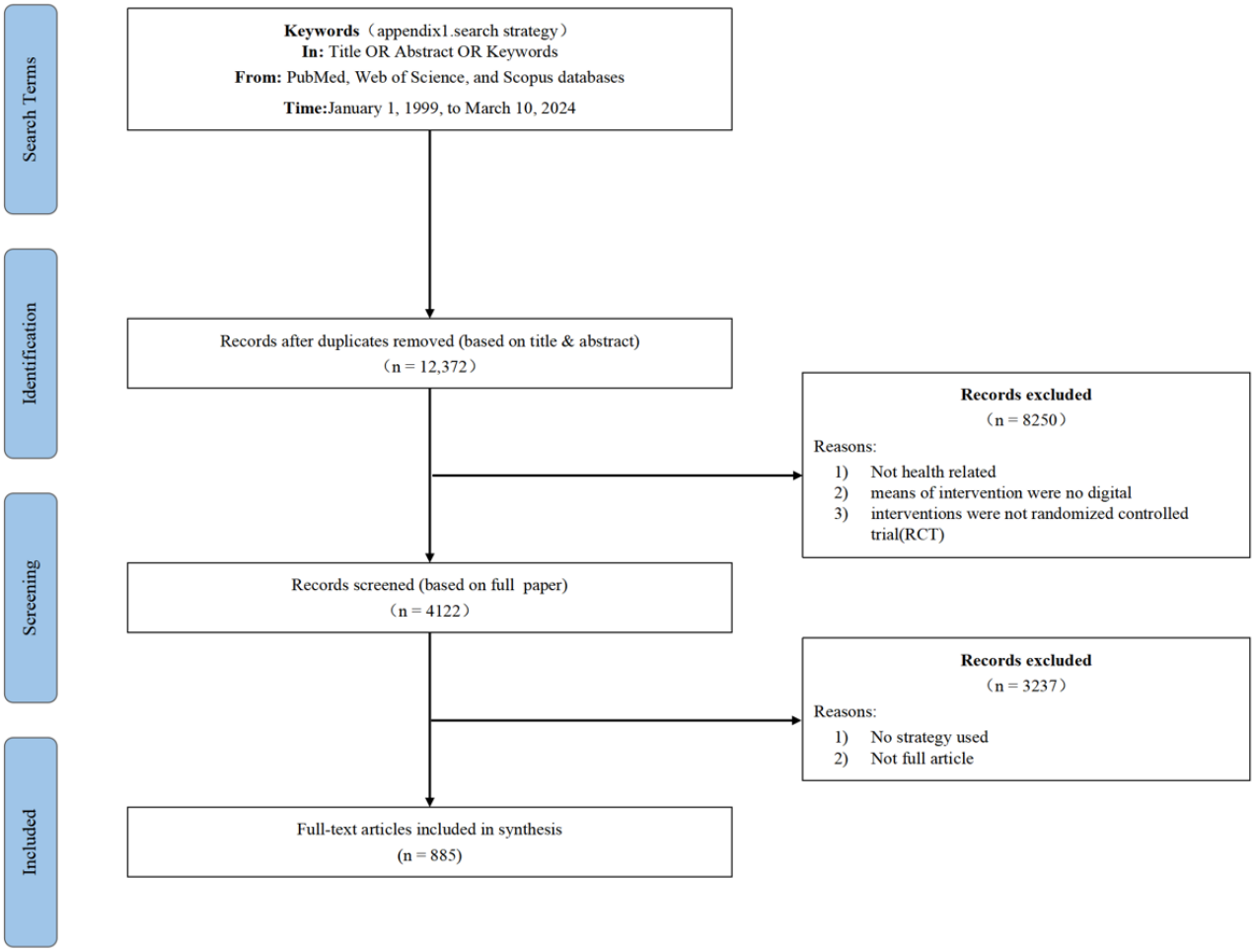

To guide the strategy extraction, we provided a framework and definitions of potential strategies (Multimedia Appendix 3) organized into 18 clusters and 64 themes based on BCTs [35] and BSs [22]. Recognizing the possibility of strategies extending beyond these predefined themes, we incorporated an "other" theme to capture a broader range of strategies. This adaptable framework aimed to maximize the thoroughness and accuracy of the data collection process, ensuring that novel or unique strategies were not overlooked.

After identifying all strategies, we conducted a comprehensive review to pinpoint overarching themes and classify each strategy accordingly. Initially, 3 authors (SL, YG, and CZ) examined the strategies and proposed themes independently. Subsequently, the first author (SL) reviewed the independent results, reconciled discrepancies, and developed a list of common themes. Next, a team meeting discussed the proposed themes and strategy categorization, leading to a final list of themes and assigning each strategy to 1 category. Finally, the senior authors reviewed and adjudicated the assigned themes, with any disagreements resolved through team discussions.

We assessed interrater reliability across 2 rounds of the coding process. In the first round, 6 independent students screened articles based on titles, abstracts, and keywords. Due to the large volume of articles, a 1% sample was used for training purposes. The interrater reliability was measured using Fleiss kappa, yielding a coefficient of 0.82. In the second round, a coding guideline was developed, and 6 students underwent thorough training before proceeding with coding. Initially, 10% of the articles were used for calibration, resulting in a Krippendorff alpha of 0.85 for intercoder reliability. Inconsistencies were addressed through meetings, with a senior researcher arbitrating persistent disagreements. Any issues encountered during the coding process were promptly discussed in regular meetings and our WeChat group.

We recorded the number of articles identified, screened, deemed eligible, and ultimately included in the study, along with coding, using Microsoft Excel (IBM Corp). This systematic approach guaranteed a transparent process for capturing papers that met our inclusion criteria.

Data Extraction and Coding

In this study, we coded several key aspects from the included sample papers. General information such as the journal of publication, authors, year of publication, the geographical context of the intervention, demographic characteristics of participants, and the theoretical framework or models underpinning the intervention (if reported) were documented.

We also examined the characteristics of the digital health interventions, including the health concerns they addressed, delivery modalities (eg, short text messages, smartphone apps, and online platforms), intervention duration, sender type (automated or human-assisted), and their classification according

XSL•FO
RenderX

to the World Health Organization's (WHO) 2018 framework (patients, health care professionals, health system administrators, and data services) [39]. Additionally, strategies were identified and labeled based on a structured framework and descriptive criteria established for this study (Multimedia Appendix 3).

Results

Characteristics of the Included Papers

After the initial screening, we identified 4122 valid sample papers on digital health interventions. Following the second round of full-paper screening, and after excluding interventions that did not incorporate any strategies, 885 studies ultimately met the eligibility criteria.

According to the World Bank [40], countries were categorized into 4 income levels: low income, lower middle income, upper middle income, and high income (Figure 2). The number of studies conducted in high-income countries (n=750, 84.7%) far exceeded that of other income levels, with research activity increasing in all countries since 2010. Notably, studies in high-income countries peaked significantly around 2015, with 139 (15.7%) studies. Studies in upper–middle-income and lower–middle-income countries showed moderate activity with occasional peaks, particularly in 2022 (n=22, 2.5%) and 2023 (n=8, 0.9%), respectively.

The sample papers reported a total of 954,847 participants (10 papers did not report the number of participants, and we were unable to identify them), ranging from a minimum of 6 to a maximum of 333,669. Participant ages varied significantly, ranging from children as young as 18 months to adults over 80 years old. Among the sample papers, 238 (26.9%)focused on vulnerable groups: 86 (9.7%)on children and adolescents, 95 (10.7%) on women, 33 (3,7%) on older adults, and 5(0.6%) on LGBT (lesbian, gay, bisexual, and transgender) individuals. Additionally, 223 (25.2%) studies involved ethnic minorities: 5 (0.6%) on African American participants, 2 (0.2%) on Hispanic participants, 4 (0.5%) on Asian participants, 4 (0.5%) on Black participants, 2 (0.2%) on Latino or Latina populations, and 4 (0.5%) on American Indian or Alaska Native participants. The remaining studies involved other ethnic minority groups.

Figure 2. Time trends of different income-level countries in the analyzed papers.

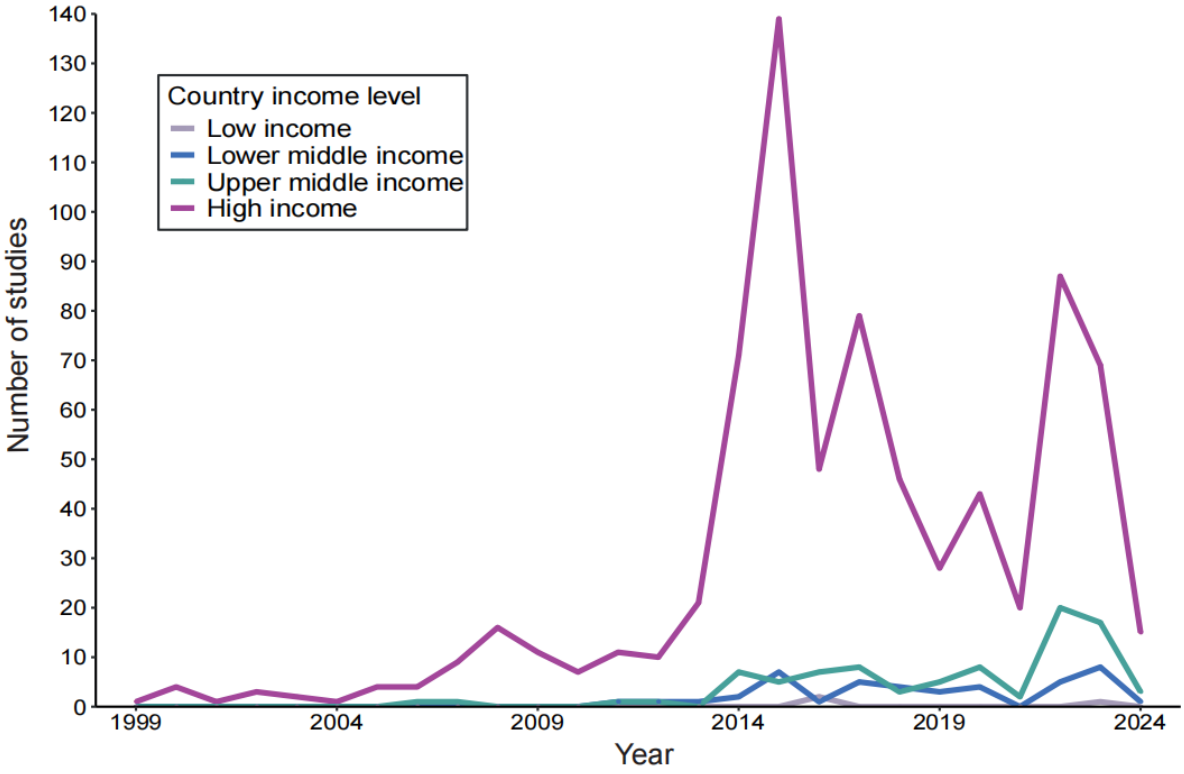

Characteristics of the Intervention

The WHO has categorized digital health interventions into 4 groups based on their intended users: consumers, health care providers, health system managers, and data services [39] (Figure 3). In this study, out of the total 885 interventions analyzed, the vast majority (844, 95.4%) were designed for clients. Health care providers were the target users for 62 (7%) interventions, while health system managers and data services were the focus of only 6 (0.7%) and 1 (0.1%) interventions, respectively.

XSL•FO
RenderX

Figure 3. World Health Organization (WHO) classification of digital health interventions.

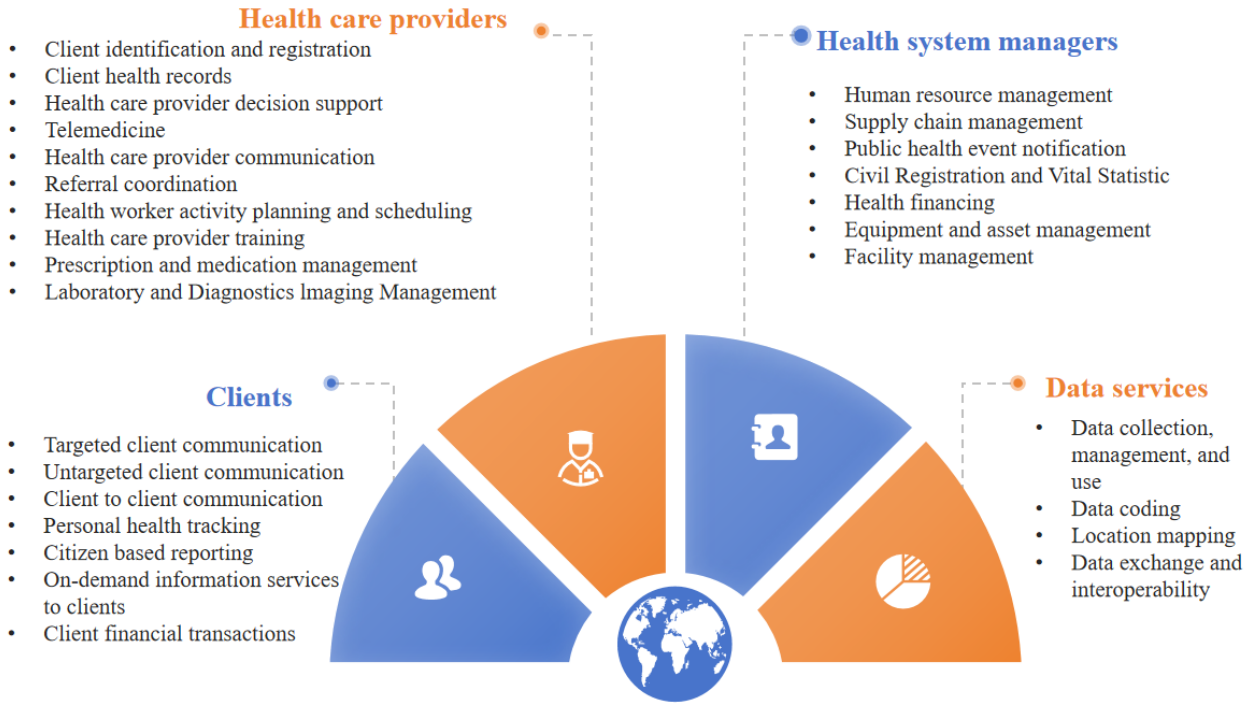

We classified the intervention settings into 5 categories: community, hospital, home, school, and other. Out of the total interventions, 118 (13.3%) were identified as taking place in the community setting, 268 (30.3%) in the hospital setting, 237 (26.8%) in the home setting, 50 (5.6%) in the school setting, and 212 (24.1%) in other settings, which included locations such as companies, workplaces, and laboratories. Among the studies included, 858 (93.2%) focused on one or more specific health issues.

We identified 23 different types of delivery in the included studies and classified them into four categories: (1) phone-based, including short message service, phone calls, and interactive voice response; (2) web-based, including email, health electronic record, online telecare/training platform, forums, and interactive videos; (3) app-based, including smartphone/tablet-computer apps, videoconferencing software, and social media apps; (4) other, including wearable devices, telemonitoring device, virtual reality, artificial intelligence, and games. The most frequently used delivery type is web-based (n=282, 31.9%), followed by phone-based (n=266, 30%) and app-based (n=259, 29.3%). Of the delivery senders, 384 (43.4%) utilized automatic support, 457 (53.7%) employed human support, and 43 (4.9%) used both.

Theories and models play a crucial role in the development, implementation, and evaluation of digital health interventions by providing a framework for understanding the complex factors that influence health behaviors and guiding the selection of appropriate intervention strategies. In this study, 92 (10.4%) interventions reported using theories or models to inform their design and delivery, with some interventions using theories or models alone, while others used a combination of both. After combining the same theories or models mentioned across the included papers, a total of 13 distinct theories and 17 distinct models were identified. The most frequently used theory was Social Cognitive Theory, which was applied in 20 (2.3%) interventions. The Transtheoretical Model of Behavior Change was the most frequently used model, reported in 6 (0.7%) interventions.

Identification and Categorization of Digital Health Intervention Strategies

Overview of the Identified Strategies

A total of 885 sample papers were identified as employing at least one strategy in digital health interventions. Among these papers, 173 (19.5%) distinct strategies were identified and subsequently categorized into 19 themes (Multimedia Appendix 4). Of these themes, 18 represented clearly categorizable strategies, while 1 theme, labeled as “other,” was used to group strategies that did not fit into the predefined categories.

Among the identified strategy themes (Table 1), the most frequently used was “guide,” employed in 492 (55.6%) out of the 885 papers. The second most common theme was ”monitor,” which was utilized in 490 (55.4%) papers. The third most prevalent theme was “communication,” which was applied in 392 papers, constituting 44.3% of the total. Other commonly used strategy themes included ”engagement“ (n=384, 43.3%), ”support“ (n=268, 30.3%), ”stimulation“ (n=271, 30.6%), ”management“ (n=263, 29.7%), ”feedback“ (n=236, 26.7%), ”goal setting“ (n=228, 25.8%), and ”action planning“ (n=203, 22.9%).

XSL•FO
RenderX

Table 1. The general results of the identified strategies.

Theme	Identified papers, n (%)
Guide	492 (55.6)
Monitoring	490 (55.4)
Communication	392 (44.3)
Engagement	384 (43.4)
Stimulate	271 (30.6)
Support	268 (30.3)
Management	263 (29.7)
Feedback	236 (26.7)
Goal setting	228 (25.8)
Action planning	203 (22.9)
Shaping	189 (21.4)
Tailor	184 (20.8)
Others	132 (14.9)
Prompts	128 (14.5)
Reward	69 (7.8)
Cues	62 (7)
Identity	57 (6.4)
Restructure	19 (2.1)
Model/demonstrate	10 (1.1)

Temporal, Geographic, and Demographic Developments in Digital Health Intervention Strategies

To observe how strategies have changed over time, we selected the year 2017 as the cut-off point. This is because 2017 marked a few significant milestones in digital health regulation and global health policy. In the United States, the Food and Drug Administration (FDA) launched its Digital Health Innovation Action Plan and established the Digital Health Unit [41]. Concurrently, the WHO initiated the development of its first guideline on digital health interventions [42]. These initiatives at both national and international levels potentially influenced the development of digital health intervention practice globally [43]. In this study, we found that before 2017, the top 3 strategies were ”monitoring“ (n=205, 51.4%), ”communication“ (n=198, 49.6%), and ”guide“ (n=197, 49.4%), with ”monitoring“ being the most frequently used strategy. From 2017 onward, the top 3 strategies shifted to ”guide“ (n=295, 60.8%), ”monitoring“ (n=285, 58.8%), and ”engagement“ (n=209, 43.1%), with ”guide“ becoming the most prevalent strategy. The specific details are presented in Table 2.

XSL•FO
RenderX

Table 2. The results of the identified strategies before and after 2017.

Before 2017		After 2017	
Theme	Value, n (%)	Theme	Value, n (%)
Monitoring	205 (51.4)	Guide	295 (60.8)
Communication	198 (49.6)	Monitoring	285 (58.8)
Guide	197 (49.4)	Engagement	209 (43.1)
Engagement	175 (43.9)	Communication	194 (40.0)
Stimulate	134 (33.6)	Support	138 (28.5)
Feedback	133 (33.3)	Stimulate	137 (28.2)
Management	131 (32.8)	Management	132 (27.2)
Support	130 (32.6)	Shaping	109 (22.5)
Goal setting	123 (30.8)	Goal setting	105 (21.6)
Action planning	112 (28.1)	Feedback	103 (21.2)
Prompts	86 (21.6)	Tailor	102 (21.0)
Tailor	82 (20.6)	Action planning	91 (18.8)
Shaping	80 (20.1)	Others	74 (15.3)
Others	58 (14.5)	Identity	46 (9.5)
Cues	34 (8.5)	Reward	46 (9.5)
Reward	23 (5.8)	Prompts	42 (8.7)
Identity	11 (2.8)	Cues	28 (5.8)
Model/demonstrate	10 (2.5)	Restructure	10 (2.1)
Restructure	9 (2.3)	Model/demonstrate	0 (0)

Figure 4 shows the distribution of strategies in high-income countries and low- and middle-income countries before and after 2017, illustrating the number of identified strategies categorized by delivery type (phone-based, web-based, app-based, and other) and population type (general and vulnerable). High-income countries demonstrated a substantially higher number and diversity of strategies, with ”monitoring,” “guide,” and ”communication“ being the most prevalent. While similar strategies were prioritized in low- and middle-income countries, they were implemented with less intensity. After 2017, both high-income and low- and middle-income countries showed an increase in the number of strategies employed, with high-income countries exhibiting more pronounced growth. This development was particularly evident in the adoption of emerging strategies such as ”engagement,” “tailoring,” and “support.”

Regarding population focus, the number of studies targeting the public far exceeded those focusing on vulnerable populations (including women, children, youth, seniors, minorities, LGBT, and low-income groups), at 73.1% (n=647) and 26.9% (n=238), respectively. While both country groups showed strategies targeting general and vulnerable populations, high-income countries demonstrated a greater emphasis on interventions for vulnerable groups, especially after 2017.

XSL•FO
RenderX

Figure 4. The distribution of strategies across delivery, population, and time in high-income countries (HICs) and low- and middle-income countries (LMICs).

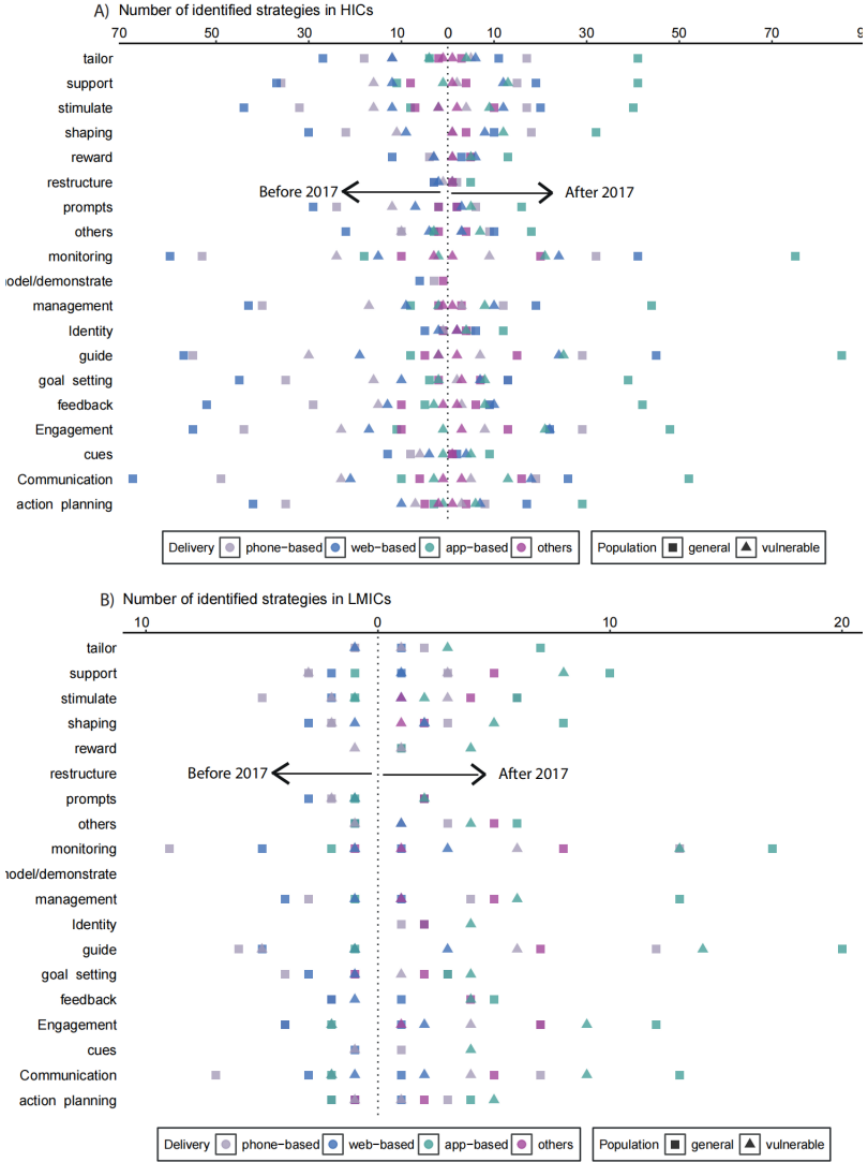

Distribution of Digital Health Intervention Strategies Across Intervention Duration and Senders

Figure 5 illustrates the distribution of strategies across intervention duration and senders. Across all duration categories, there was a clear peak between 3 and 4 strategies, with a gradual decline as the number of strategies increased. Moreover, interventions implementing more than 10 strategies were relatively uncommon. Regarding intervention duration, short-term (≤3 months) and medium-term (3-12 months) interventions dominated the landscape, particularly those employing between 2 and 4 strategies. Medium-term interventions showed the highest frequency, especially for interventions utilizing between 2 and 5 strategies. Long-term interventions (>12 months), while less common, maintained a consistent presence across the spectrum of strategy numbers.

The distribution across sender types demonstrated that human support was the most prevalent delivery form, particularly for those using between 2 and 5 strategies. Automated support followed a similar pattern but with lower frequencies, while combined support (both human and automated) was the least common but had a consistent presence. According to the duration graph, all sender types peaked at 3 to 4 strategies used, with a sharp decline in frequency as the number of strategies increased beyond 5.

XSL•FO
RenderX

Figure 5. Distribution of strategies across the intervention duration and senders.

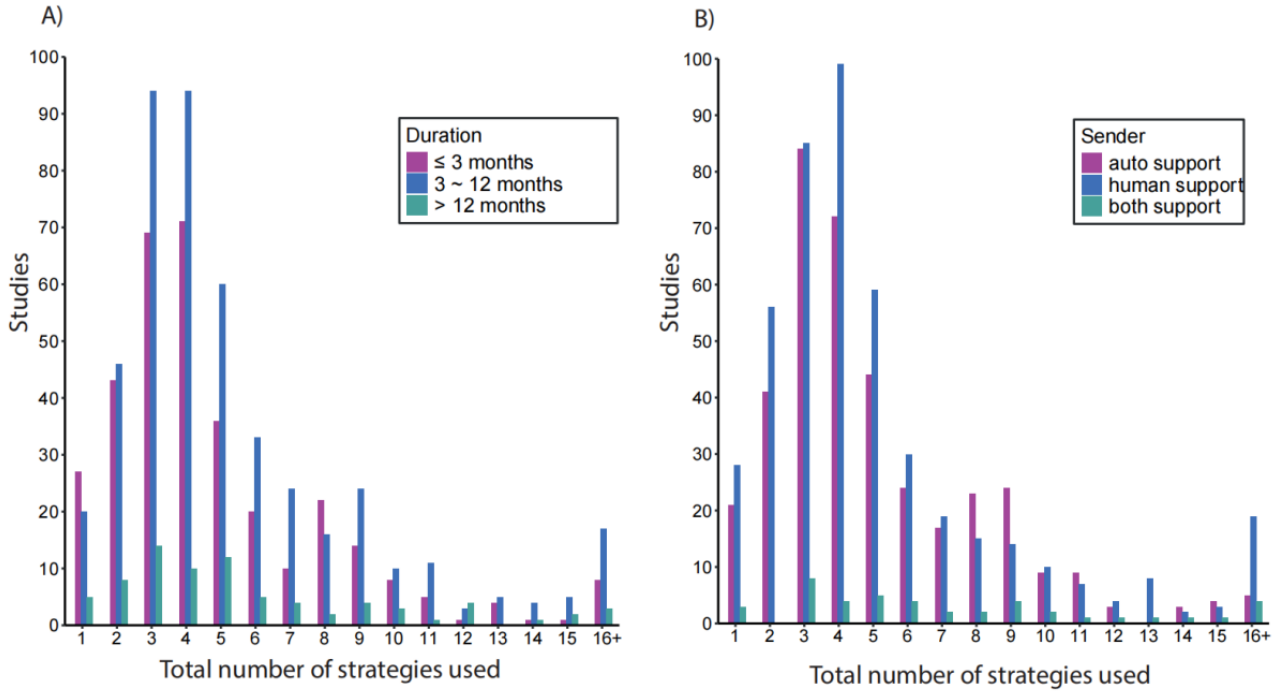

Discussion

Principal Findings

Our study provided a comprehensive analysis of strategies employed in digital health interventions over the past 25 years. We identified 173 unique strategies and categorized them into 19 overarching themes based on BCTs and BSs. This helped us establish a valuable framework for understanding the diverse approaches used in designing and implementing digital health interventions. The analysis revealed significant insights into several key aspects—the evolution of strategies over time, the differential involvement between high-income and low- and middle-income countries, and the influence of various factors such as delivery methods, target populations, intervention duration, and intervention providers. These findings collectively map the landscape of digital health intervention strategies across a quarter-century, providing valuable guidance for researchers and practitioners.

Specifically, we identified that the three most prevalent strategies were ”guide,“ ”monitoring,“ and ”communication,” highlighting the importance of providing clear guidance, tracking progress, and facilitating interactive communication in digital health interventions. The widespread application of “engagement” and “support” underscored the significance of fostering user involvement and providing adequate assistance throughout the intervention process. These findings were consistent with prior research that underscored the crucial role of user engagement, self-monitoring, and social support in digital health interventions [21,44].

The number of strategies used in each paper varied widely, from a single strategy to a comprehensive approach involving up to 32 strategies, highlighting the complexity and diversity of digital health interventions. Most studies fell into the low or medium categories, indicating that most digital health interventions had a focused approach with a limited set of strategies. The existing research indicated that using a combination of strategies, rather than a single approach, can be more effective in enhancing participant retention [32]. The available evidence suggests that to optimize participant retention throughout a study, researchers should consider implementing multiple retention strategies across various themes [33]. Future research should carefully evaluate the optimal strategic approach for each intervention, considering its specific goals, target population, and context. It should also investigate how the number and type of strategies employed affect intervention effectiveness and identify the best combination of strategies for various intervention contexts.

The landscape of digital health intervention strategies has undergone significant shifts, particularly around the pivotal year of 2017. This year marked 2 crucial developments: the launch of the Digital Health Unit by the FDA in the United States and the initiation of the WHO’s first guideline on digital health interventions. Prior to 2017, “monitoring” was the predominant strategy in digital health interventions. However, in 2017 and beyond, “guide” emerged as the most frequently employed approach. This transition reflected a shift toward more dynamic and interactive user interfaces, emphasizing proactive guidance rather than passive monitoring [45]. This change aligned with emerging evidence suggesting that interactive and personalized digital health interventions can lead to better user engagement and health outcomes [46,47]. Additionally, “engagement” had also risen in importance after 2017, illustrating a growing focus on enhancing user experience and satisfaction to improve intervention outcomes [48]. This change underscored a broader trend within the health care sector toward patient-centered care, which prioritizes the user experience and active participation in health management [49].

XSL•FO
RenderX

The adoption of strategies in digital health interventions was closely related to delivery methods and sender types [50]. While web-based and phone-based interventions remained prevalent, the rise of app-based strategies reflected the increasing penetration of smartphones, aligning with the ubiquitous use of mobile phones in modern society and highlighting the potential of leveraging mobile technology for health interventions [51]. This trend aligned with the ubiquitous use of mobile devices and underscored the growing importance of accessible, on-the-go health solutions. Regarding sender types, most of the sample papers employed human support in the delivery of digital health interventions, while less than half used automatic support. This distribution highlighted the ongoing significant role of human involvement in these interventions. Despite the increasing capabilities of automated systems, human support remained a crucial component in many digital health interventions, attributed to the human ability to provide empathy, build rapport, and tailor interventions to individual needs. Nonetheless, the role of automatic support was expanding, driven by advances in artificial intelligence and its potential for scalability and cost-effectiveness [52]. This dual trend—the continued importance of human support alongside the growth of automated systems—suggested a future where digital health interventions may increasingly leverage both human and artificial intelligence to improve health outcomes.

There was a distinct regional bias in the use of strategies, particularly across different income levels. High-income countries demonstrated a substantially higher number and diversity of identified strategies compared to low- and middle-income countries. This disparity may be attributed to better technological infrastructure, greater financial resources, and higher digital literacy among the population in high-income countries [50]. While low- and middle-income countries showed lower overall usage, they displayed a growing trend in adopting digital health strategies after 2017. This rise indicated a catching-up phenomenon, possibly driven by increased global health initiatives, international collaborations, and high mobile technology penetration in many low- and middle-income countries [53].

Regarding population distribution, the results indicated that the number of studies targeting the public far exceeded those focusing on vulnerable populations, suggesting a potential gap in addressing the specific needs of vulnerable groups, including women, children, youth, seniors, minorities, LGBT, and low-income populations, through digital health interventions [54]. Future research and implementation efforts should focus on developing more targeted strategies for these underserved populations to ensure equitable access to digital health benefits.

In conclusion, over the past 25 years, digital health interventions and their strategies have undergone significant developments. Initially, digital health interventions often utilized relatively simple and limited strategies, mainly focusing on providing information and basic self-monitoring functions [55]. With advancements in technology and a deeper understanding of behavior change mechanisms, digital health intervention strategies have become increasingly complex and diverse. Recently, there has been a rise in the use of personalized and adaptive strategies, aiming to tailor interventions to participants' characteristics, needs, and responses [56], as well as the aims of the intervention. Personalized or individualized strategies have become increasingly popular and widely used due to their potential to enhance engagement and intervention effectiveness. Notably, there has been a rise in the use of transtheoretical strategies in recent years, integrating elements from various behavior change theories to create more comprehensive and effective interventions [57].

Strengths of Our Research

Our findings built upon and broadened the scope of previous research on digital health intervention strategies [20,22,26-34,58] and the principles strategy classification [22,35]. While previous studies have often concentrated on specific health domains or populations, our research offered a more comprehensive view of the strategies used across various health issues and target groups. By identifying and categorizing a wider array of strategies, we provided a more integrated understanding of how to promote health behavior change and enhance health outcomes in the digital era.

Additionally, our analysis of the diversity in the number of strategies employed in each paper highlights the complexity and variability inherent in digital health interventions. Although prior research has underlined the significance of employing multiple strategies to tackle the multifaceted nature of health behavior change, our findings suggested that the ideal number of strategies might differ based on the specific objectives and context of the intervention. This emphasized the necessity for a customized approach in the design and implementation of digital health interventions, considering the distinct needs and characteristics of the target population and health domain.

Implications for Research and Practice

The findings carried several crucial implications for research and practice in the digital health field. First, by amalgamating evidence on what strategies prove effective across diverse health domains and populations, our study presented a detailed categorization of strategies for researchers and practitioners aiming to design and evaluate digital health interventions. This structured framework facilitated an understanding of the varied approaches within these interventions, aiding in the strategic selection and combination tailored to specific intervention goals and contexts. This, in turn, aimed to maximize the efficacy and impact of digital health interventions.

Second, this analysis of the frequency and distribution of strategy themes can direct future research priorities and resource allocation. The prominent presence of strategies concerning guidance, monitoring, and communication underlines their potential significance as focal areas for additional research.

Third, the observed variation in the number of strategies employed across different studies underscored the necessity for a more refined approach to designing and assessing digital health interventions. It suggested moving away from the presumption that employing more strategies invariably leads to better outcomes. Instead, there is a need for a deliberate consideration of the most effective number and mix of strategies, aligning with the unique requirements and attributes of the target population and health domain.

XSL•FO
RenderX

Limitations of This Study

This study has several limitations. Despite following a comprehensive search strategy, some relevant and important papers may have been overlooked, especially those published in languages other than English, from low- and middle-income countries, and in gray literature due to the limitations of the databases chosen. The process of data extraction and categorization of strategies was contingent on the information provided in the sample papers, which in some instances, may have been limited or inconsistent. Moreover, the diversity among the sample papers regarding populations, health domains, and intervention designs might restrict the applicability of our findings to specific situations. Additionally, this study did not carry out a comparative analysis of the effects of different strategies, nor did it explore and quantify the relationship between strategies and health outcomes.

Conclusions

This study provided an exhaustive synthesis of strategies used in digital health interventions over the past 25 years, highlighting their characteristics, distribution, and diversity in promoting health behavior changes and improving outcomes. By identifying and categorizing 173 unique strategies within 19 overarching themes, this study offered a comprehensive framework for understanding the complex approaches used in digital health intervention strategies. Additionally, this study highlighted that high-income and low- and middle-income countries have seen increased adoption after 2017, and human support remained crucial despite advancements in automated systems. These findings facilitated the careful design and assessment of future digital health interventions, significantly contributing to the establishment of evidence-based guidelines and best practices that enhance their efficacy.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant 72474174). The authors would like to thank Na Zhang, Min Li, and Hong Du for their support with the paper screening and coding.

Authors' Contributions

SL contributed to the study conception and design, data analysis, interpretation of the results, and manuscript preparation. JM was responsible for the study design, data collection, coding, and analysis of results. MS was responsible for the data collection and coding. YG contributed to the study design, coding, and analysis of the results. CZ contributed to the study design, coding, and analysis of the results. JX was responsible for the study conception and design, interpretation of the results, and manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 136 KB - jmir_v27i1e59027_app1.pdf]

Multimedia Appendix 2
Search strategy.
[DOCX File , 13 KB - jmir_v27i1e59027_app2.docx]

Multimedia Appendix 3
Framework for Strategy Identification and Definition.
[DOC File , 119 KB - jmir_v27i1e59027_app3.doc]

Multimedia Appendix 4
Overview of the identified strategies.
[DOCX File , 22 KB - jmir_v27i1e59027_app4.docx]

References

1. Wienert J, Jahnel T, Maaß L. What are digital public health interventions? First steps toward a definition and an intervention classification framework. J Med Internet Res 2022 Jun 28;24(6):e31921 [FREE Full text] [doi: 10.2196/31921] [Medline: 35763320]
2. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]

XSL•FO
RenderX

3. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res 2020 Oct 28;22(10):e16774 [FREE Full text] [doi: 10.2196/16774] [Medline: 33112239]
4. Yeo G, Reich SM, Liaw NA, Chia EYM. The effect of digital mental health literacy interventions on mental health: systematic review and meta-analysis. J Med Internet Res 2024 Feb 29;26:e51268 [FREE Full text] [doi: 10.2196/51268] [Medline: 38421687]
5. Sha L, Yang X, Deng R, Wang W, Tao Y, Cao H, et al. Automated digital interventions and smoking cessation: systematic review and meta-analysis relating efficiency to a psychological theory of intervention perspective. J Med Internet Res 2022 Nov 16;24(11):e38206 [FREE Full text] [doi: 10.2196/38206] [Medline: 36383408]
6. Karimi N, Opie R, Crawford D, O'Connell S, Ball K. Digitally delivered interventions to improve nutrition behaviors among resource-poor and ethnic minority groups with type 2 diabetes: systematic review. J Med Internet Res 2024 Feb 01;26:e42595 [FREE Full text] [doi: 10.2196/42595] [Medline: 38300694]
7. Cevasco KE, Roess AA. Adaptation and utilization of a postmarket evaluation model for digital contact tracing mobile health tools in the United States: observational cross-sectional study. JMIR Public Health Surveill 2023 Mar 22;9:e38633 [FREE Full text] [doi: 10.2196/38633] [Medline: 36947135]
8. Kemp J, Chorney J, Kassam I, MacDonald J, MacDonald T, Wozney L, et al. Learning about the current state of digital mental health interventions for Canadian youth to inform future decision-making: mixed methods study. J Med Internet Res 2021 Oct 19;23(10):e30491 [FREE Full text] [doi: 10.2196/30491] [Medline: 34665141]
9. Kazi AM, Qazi SA, Ahsan N, Khawaja S, Sameen F, Saqib M, et al. Current challenges of digital health interventions in Pakistan: mixed methods analysis. J Med Internet Res 2020 Sep 03;22(9):e21691 [FREE Full text] [doi: 10.2196/21691] [Medline: 32880584]
10. Oluokun EO, Adedoyin FF, Dogan H, Jiang N. Digital Interventions for managing medication and health care service delivery in West Africa: systematic review. J Med Internet Res 2024 Oct 09;26:e44294 [FREE Full text] [doi: 10.2196/44294] [Medline: 39383531]
11. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
12. Schlieter H, Marsch LA, Whitehouse D, Otto L, Londral AR, Teepe GW, et al. Scale-up of digital innovations in health care: expert commentary on enablers and barriers. J Med Internet Res 2022 Mar 11;24(3):e24582 [FREE Full text] [doi: 10.2196/24582] [Medline: 35275065]
13. Sim I. Mobile devices and health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/nejmra1806949]
14. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
15. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
16. Winter N, Russell L, Ugalde A, White V, Livingston P. Engagement strategies to improve adherence and retention in web-based mindfulness programs: systematic review. J Med Internet Res 2022 Jan 12;24(1):e30026 [FREE Full text] [doi: 10.2196/30026] [Medline: 35019851]
17. Vollmer DD, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
18. Carpenter SM, Greer ZM, Newman R, Murphy SA, Shetty V, Nahum-Shani I. Developing message strategies to engage racial and ethnic minority groups in digital oral self-care interventions: participatory co-design approach. JMIR Form Res 2023 Dec 11;7:e49179 [FREE Full text] [doi: 10.2196/49179] [Medline: 38079204]
19. Ross J, Stevenson F, Dack C, Pal K, May C, Michie S, et al. Developing an implementation strategy for a digital health intervention: an example in routine healthcare. BMC Health Serv Res 2018 Oct 19;18(1):794 [FREE Full text] [doi: 10.1186/s12913-018-3615-7] [Medline: 30340639]
20. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
21. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
22. Lim S, Hill B, Pirotta S, O'Reilly S, Moran L. What are the most effective behavioural strategies in changing postpartum women's physical activity and healthy eating behaviours? A systematic review and meta-analysis. J Clin Med 2020 Jan 16;9(1) [FREE Full text] [doi: 10.3390/jcm9010237] [Medline: 31963150]

XSL•FO
RenderX

23. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
24. Kwasnicka D, Dombrowski SU, White M, Sniehotta FF. 'It's not a diet, it's a lifestyle': a longitudinal, data-prompted interview study of weight loss maintenance. Psychol Health 2019 Aug;34(8):963-982. [doi: 10.1080/08870446.2019.1579913] [Medline: 30905184]
25. Masi CM, Chen H, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug 17;15(3):219-266 [FREE Full text] [doi: 10.1177/1088868310377394] [Medline: 20716644]
26. Pang B, Zhao L, Li X, Song J, Li Q, Liao X, et al. Different intervention strategies for preventing type 2 diabetes mellitus in China: A systematic review and network meta-analysis of randomized controlled trials. Diabetes Obes Metab 2018 Mar 08;20(3):718-722. [doi: 10.1111/dom.13121] [Medline: 28941313]
27. Jeffery M, Hickey B, Hider P. Follow-up strategies for patients treated for non-metastatic colorectal cancer. Cochrane Database Syst Rev 2019. [doi: 10.1002/14651858.cd002200] [Medline: 31483854]
28. Parra-Medina D, D'antonio A, Smith SM, Levin S, Kirkner G, Mayer-Davis E, POWER study. Successful recruitment and retention strategies for a randomized weight management trial for people with diabetes living in rural, medically underserved counties of South Carolina: the POWER study. J Am Diet Assoc 2004 Jan;104(1):70-75. [doi: 10.1016/j.jada.2003.10.014] [Medline: 14702587]
29. Coday M, Boutin-Foster C, Goldman Sher T, Tennant J, Greaney M, Saunders S, et al. Strategies for retaining study participants in behavioral intervention trials: retention experiences of the NIH Behavior Change Consortium. Ann Behav Med 2005 Apr;29 Suppl(2):55-65. [doi: 10.1207/s15324796abm2902s_9] [Medline: 15921490]
30. Hwang DA, Lee A, Song JM, Han H. Recruitment and retention strategies among racial and ethnic minorities in web-based intervention trials: retrospective qualitative analysis. J Med Internet Res 2021 Jul 12;23(7):e23959 [FREE Full text] [doi: 10.2196/23959] [Medline: 34255658]
31. Treweek S, Pitkethly M, Cook J, Fraser C, Mitchell E, Sullivan F, et al. Strategies to improve recruitment to randomised trials. Cochrane Database Syst Rev 2018 Dec 22;2:MR000013. [doi: 10.1002/14651858.MR000013.pub6] [Medline: 29468635]
32. Brueton V, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. Cochrane Database Syst Rev 2013. [doi: 10.1002/14651858.mr000032.pub2]
33. Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol 2007 Aug;60(8):757-765 [FREE Full text] [doi: 10.1016/j.jclinepi.2006.11.023] [Medline: 17606170]
34. Cosp X, Castillejo M, Vila M, Marti J, Emparanza J. Strategies for increasing the participation of women in community breast cancer screening. Cochrane Database Syst Rev 2001. [doi: 10.1002/14651858.cd002943]
35. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. Br J Health Psychol 2015 Feb;20(1):130-150. [doi: 10.1111/bjhp.12102] [Medline: 24815766]
36. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
37. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
38. Abbott JH. The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: what they are and are not. J Orthop Sports Phys Ther 2014 Aug;44(8):555-558. [doi: 10.2519/jospt.2014.0110] [Medline: 25082389]
39. Classification of digital health interventions v1.0. World Health Organization. URL: https://www.who.int/publications-detail-redirect/WHO-RHR-18.06 [accessed 2024-03-30]
40. World development report 2021: data for better lives. The World Bank. URL: https://wdr2021.worldbank.org [accessed 2024-06-30]
41. FDA in Brief: FDA brings additional efficiency and modernization to regulation of digital health, as part of the Digital Health Innovation Action Plan. US Food & Drug Administration. America: FDA; 2019. URL: https://www.fda.gov/news-events/fda-brief/fda-brief-fda-brings-additional-efficiency-and-modernization-regulation-digital-health-part-digital [accessed 2024-06-30]
42. WHO acceleration plan to stop obesity. World Health Organization. URL: https://www.who.int/publications-detail-redirect/9789240075634 [accessed 2023-12-17]
43. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 3;14(1). [doi: 10.1186/s12992-018-0424-z]
44. Kelders SM, Kok RN, Ossebaard HC, Van GJEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

XSL•FO
RenderX

45. Alnosayan N, Chatterjee S, Alluhaidan A, Lee E, Houston FL. Design and usability of a heart failure mHealth system: a pilot study. JMIR Hum Factors 2017 Mar 24;4(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.6481] [Medline: 28341615]
46. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
47. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Nov 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]
48. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
49. Aminabee S. The future of healthcare and patient-centric care: Digital innovations, trends, and predictions. In: Emerging Technologies for Health Literacy and Medical Practice. New York, NY: IGI Global; 2024:240-262.
50. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: A scoping review of the literature. PLoS One 2018 Apr 3;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]
51. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 01;171(5):461-469. [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
52. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
53. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
54. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
55. Evers KE. eHealth promotion: the use of the Internet for health promotion. Am J Health Promot 2006 Aug 26;20(4):suppl 1-s7, iii. [doi: 10.4278/0890-1171-20.4.1] [Medline: 16555803]
56. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
57. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
58. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]

Abbreviations

BCT: Behavior Change Technique
BS: Behavior Strategy
FDA: Food and Drug Administration
LGBT: lesbian, gay, bisexual, and transgender
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
WHO: World Health Organization

XSL•FO
RenderX

Edited by G Eysenbach; submitted 31.03.24; peer-reviewed by H Gao, E Sezgin; comments to author 06.06.24; revised version received 03.07.24; accepted 30.11.24; published 13.01.25.

Please cite as:
Liu S, Ma J, Sun M, Zhang C, Gao Y, Xu J
Mapping the Landscape of Digital Health Intervention Strategies: 25-Year Synthesis
J Med Internet Res 2025;27:e59027
URL: https://www.jmir.org/2025/1/e59027
doi:10.2196/59027
PMID:39804697

©Shiyu Liu, Jingru Ma, Meichen Sun, Chao Zhang, Yujing Gao, Jinghong Xu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Building a Digital Health Research Platform to Enable Recruitment, Enrollment, Data Collection, and Follow-Up for a Highly Diverse Longitudinal US Cohort of 1 Million People in the All of Us Research Program: Design and Implementation Study

Dave Klein[1], MBA; Aisha Montgomery[1], MD, MPH; Mark Begale[1], BA; Scott Sutherland[1], BS; Sherilyn Sawyer[2], PhD; Jacob L McCauley[3], PhD; Letheshia Husbands[4], MPH; Deepti Joshi[1], PhD; Alan Ashbeck[1], BS; Marcy Palmer[1], BS; Praduman Jain[1], MSc

[1]Vibrent Health, Inc, Fairfax, VA, United States

[2]Boston VA Research Institute, VA Boston Health Care, Veterans Administration, Boston, MA, United States

[3]John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, United States

[4]The Task Force for Global Health, Inc, Decatur, GA, United States

Corresponding Author:
Aisha Montgomery, MD, MPH
Vibrent Health, Inc
4114 Legato Rd #900
Fairfax, VA, 22033
United States
Phone: 1 6784686545
Email: aisha.montgomery@gmail.com

Abstract

Background: Longitudinal cohort studies have traditionally relied on clinic-based recruitment models, which limit cohort diversity and the generalizability of research outcomes. Digital research platforms can be used to increase participant access, improve study engagement, streamline data collection, and increase data quality; however, the efficacy and sustainability of digitally enabled studies rely heavily on the design, implementation, and management of the digital platform being used.

Objective: We sought to design and build a secure, privacy-preserving, validated, participant-centric digital health research platform (DHRP) to recruit and enroll participants, collect multimodal data, and engage participants from diverse backgrounds in the National Institutes of Health's (NIH) All of Us Research Program (AOU). AOU is an ongoing national, multiyear study aimed to build a research cohort of 1 million participants that reflects the diversity of the United States, including minority, health-disparate, and other populations underrepresented in biomedical research (UBR).

Methods: We collaborated with community members, health care provider organizations (HPOs), and NIH leadership to design, build, and validate a secure, feature-rich digital platform to facilitate multisite, hybrid, and remote study participation and multimodal data collection in AOU. Participants were recruited by in-person, print, and online digital campaigns. Participants securely accessed the DHRP via web and mobile apps, either independently or with research staff support. The participant-facing tool facilitated electronic informed consent (eConsent), multisource data collection (eg, surveys, genomic results, wearables, and electronic health records [EHRs]), and ongoing participant engagement. We also built tools for research staff to conduct remote participant support, study workflow management, participant tracking, data analytics, data harmonization, and data management.

Results: We built a secure, participant-centric DHRP with engaging functionality used to recruit, engage, and collect data from 705,719 diverse participants throughout the United States. As of April 2024, 87% (n=613,976) of the participants enrolled via the platform were from UBR groups, including racial and ethnic minorities (n=282,429, 46%), rural dwelling individuals (n=49,118, 8%), those over the age of 65 years (n=190,333, 31%), and individuals with low socioeconomic status (n=122,795, 20%).

Conclusions: We built a participant-centric digital platform with tools to enable engagement with individuals from different racial, ethnic, and socioeconomic backgrounds and other UBR groups. This DHRP demonstrated successful use among diverse participants. These findings could be used as best practices for the effective use of digital platforms to build and sustain cohorts of various study designs and increase engagement with diverse populations in health research.

XSL•FO
RenderX

(J Med Internet Res 2025;27:e60189) doi:10.2196/60189

KEYWORDS

longitudinal studies; cohort studies; health disparities; minority populations; vulnerable populations; precision medicine; biomedical research; decentralization; digital health technology; database management system

Introduction

Background

Longitudinal cohort studies have historically relied on clinic-based recruitment models that limit participation and reduce cohort diversity in many ways [1,2]. System-level barriers, such as distance to the research site and clinic-based eligibility, can negatively affect participant recruitment and study retention. Socioeconomic barriers, such as income, education, and health insurance status, also limit research participation in minority and other health-disparate groups [3-5]. The resultant homogeny in research cohorts reduces the generalizability and validity of research outcomes [6-8]. Digital platforms can minimize barriers to research participation, such as transportation costs, site access, and time commitment, and significantly improve participant access and engagement [9-11]. As a result, digital platforms are being increasingly used in health research to enable accelerated, more accessible, and more reliable real-world data collection. However, as longitudinal cohorts increase in the breadth and scale of data, the efficacy of digital platforms can vary profoundly, depending on the platform's capability to meet the ever-increasing needs of the study over time.

Appropriate implementation of study activities via digital platforms can minimize many of the commonly reported barriers to research participation. Recent studies, such as MyHeart Counts and Health eHeart, show how digital enablement of traditional research methods may improve engagement of people from diverse backgrounds, including those who do not commonly engage with health care systems [12-14]. Digital technologies also reduce the stakeholder burden related to collection, curation, and sharing of health data [15]. Therefore, digital platforms may be effective tools to improve participant engagement, increase cohort diversity, broaden geographical reach, and streamline the curation of diverse research data sets.

Digital research platforms must also accommodate users of varying digital aptitudes. This requires consideration of sociodemographic factors that impact the use of digital technology, such as age, disability, rurality, education, income, culture, digital access, native language, and literacy [16,17]. Digital research platforms can be adapted to accommodate these differences and thereby reduce participant burden and build trust with communities that are underrepresented in health research [18]. Existing commercial and academic tools for electronic data capture (EDC) are built primarily for research teams and are not designed to meet the preferences of participants interacting with them [19]. Therefore, it is important to create digital tools with engaging functionality for diverse participants that also meet the evolving needs of longitudinal cohort studies. We aimed to build such a platform to support nationwide enrollment of 1 million diverse participants in a novel multiyear, longitudinal cohort study.

Objectives

In this paper, we describe various tools within a digital health research platform (DHRP) designed to effectively engage diverse participants from various backgrounds and collect multisource health data in a large community-based longitudinal cohort study. The All of Us Research Program (AOU) is an ongoing nationwide initiative aimed to recruit 1 million participants from diverse cultural, socioeconomic, demographic, and geographic backgrounds and to collect data that are generalizable, accelerate biomedical research, and improve outcomes for all groups [20,21]. The broadly inclusive eligibility, recruitment, and data curation needs of the AOU required the development of a highly adaptable, comprehensive digital platform able to accommodate multifaceted, complex study requirements. The study also needed a platform with modifiability that would allow study staff across 1200 sites to conduct operations and rapidly deliver customized, community-specific engagement (Figure 1). These complex needs led to the development of a comprehensive digital platform design capable of supporting data expansion and the addition of ancillary studies over the extended study timeline of 10 or more years [22].

XSL•FO
RenderX

Figure 1. Illustration of how an all-in-one platform requires broad functionality to accommodate the varying dimensions of complexities for different study protocols.

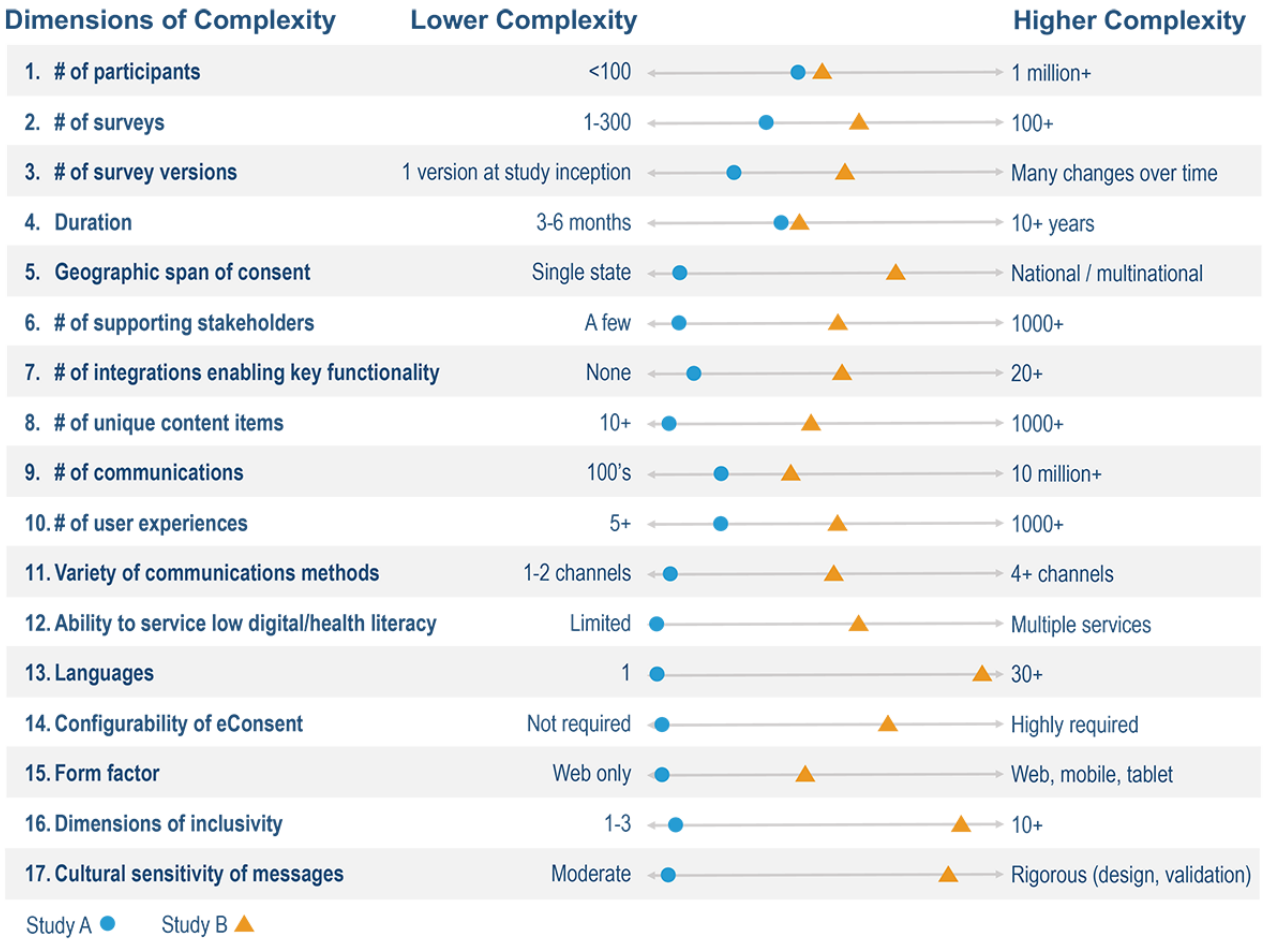

Methods

System Requirements

Vibrent Health, Inc, a digital health technology company, was selected as the participant technology systems center and tasked with developing a secure digital platform and tools for participant experience, study management, and data analysis [23]. To align with its core values, the AOU needed a secure DHRP that (1) could be broadly deployed nationwide to support study management, (2) could host several program touch points within a flexible participant journey, (3) was accessible to different levels of digital access, literacy, and comfort, and (4) was cybersecure and robust (Figure 2).

XSL•FO
RenderX

Figure 2. AOU core values were used to develop key design features of the DHRP. AOU: All of Us Research Program; DHRP: digital health research platform; HIPAA: Health Insurance Portability and Accountability Act.

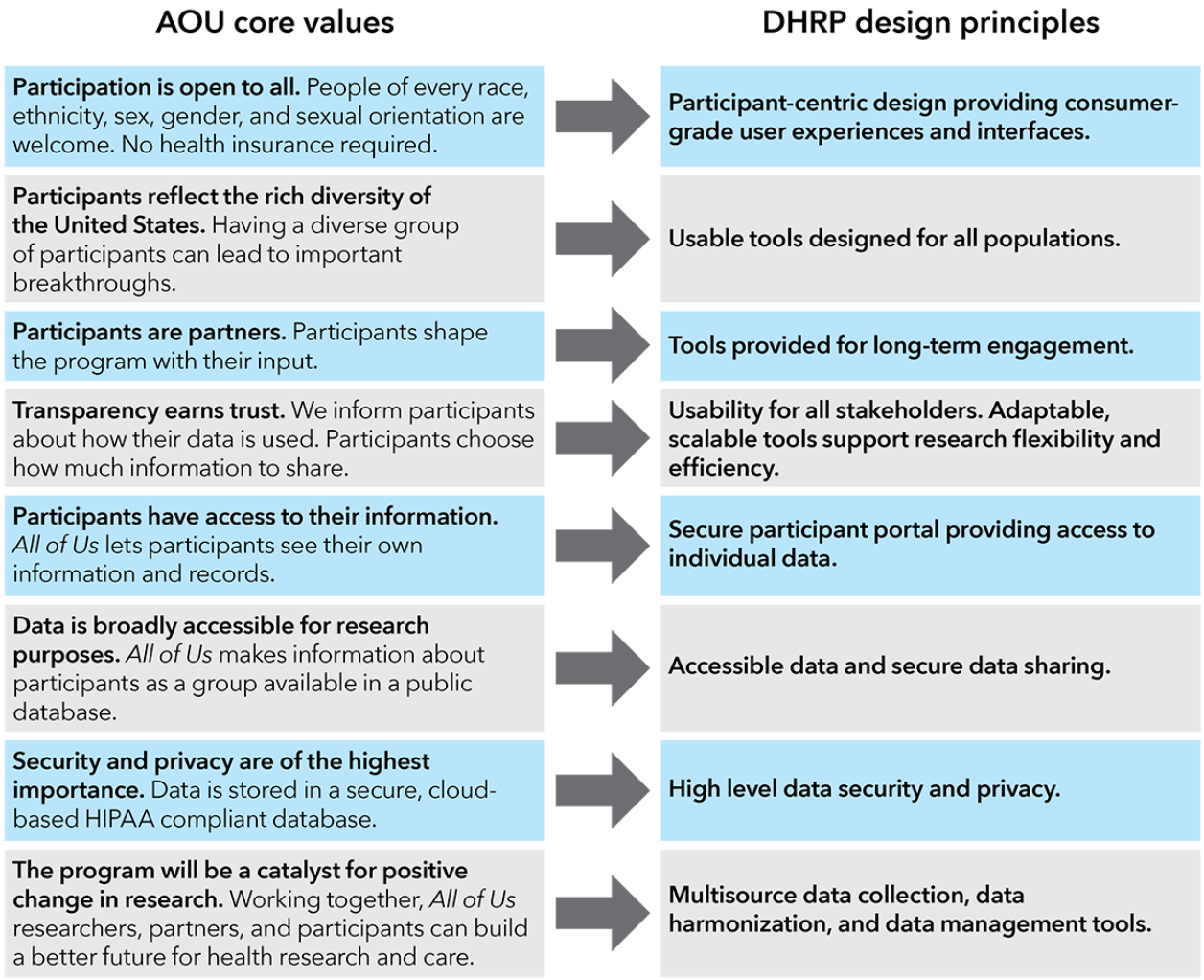

System Description

Digital Health Research Platform Design and Architecture

Vibrent Health designed and built a DHRP to facilitate participant recruitment, enrollment, multisource data collection, and long-term engagement using web-capable digital devices (computers, mobile devices, tablets). The DHRP was developed via close collaboration between the government, academia, community stakeholders, and industry using validated user experience (UX) and user interface (UI) research. (Figure 3).

We developed participant-facing tools for engagement and data collection, as well as researcher-facing study management tools. The data collected within the DHRP were stored in a cybersecure cloud environment, where they could be harmonized, cleaned, and integrated with other data systems.

The technical architecture of the DHRP is highly configurable, using a low-code approach and promoting an open ecosystem without the need for custom software development. The AOU used a range of DHRP tools to facilitate the study: (1) Participant Experience Manager (PXM), (2) Research Cloud (RC), 3) Data Explorer (DX), and (4) Community Engagement Builder (CEB). The DHRP tools are shown in Figure 4.

The DHRP adopted a hybrid technical strategy, blending readily accessible commercial tools and subsystems with customized application layer enhancements and integrations. This approach provided a tailored solution of high-quality explicitly crafted tools that are also generalizable to meet the needs of various study designs. The platform supported an open ecosystem to integrate with additional devices and to embed experimental digital health technology modules within study protocol pathways.

The DHRP used multilayered enterprise architecture to achieve secure storage of data, along with flexibility and scalability. The Cloud Services and Infrastructure Management Layer provided dynamic autoscaling to meet the changing demands of studies over time. The Data Lake and Database Layer managed and stored large volumes of multisource data, including participant data and operational data. Key functions of the DHRP included data integration, data transformation, advanced analytics, data governance, real-time data processing, and metadata management. The platform used centralized cloud-hosted databases (MySQL, Snowflake [NoSQL], Redshift) for secure data storage, and a data warehouse. The Functional

XSL•FO
RenderX

Modules Layer facilitated interconnection of the DHRP to multiple third-party application programming interfaces (APIs) and end-user applications.

The DHRP used several methodological testing frameworks and capabilities, as well as scientific testing support protocols, including user research, functional randomization testing, randomization for scientific testing, survey response randomization, and bias mitigation. In addition, Agile software development life cycle (SDLC) and continuous integration/continuous delivery (CI/CD) processes facilitated the delivery of high-quality enterprise software. The "shift-left" SDLC reduced the complexity of software development, thus enabling the study to respond quickly to protocol amendments without disrupting the participant experience. Examples of configuration updates include the addition of new surveys, workflow changes, or revision of study-related content.

Containerization and microservices methods were applied in the DHRP to optimize flexibility, scalability, efficiency, system performance, security, and privacy. Containerization provides a standardized, efficient approach to software deployment by encapsulating applications within isolated environments. This ensures consistent performance and compatibility across diverse computing platforms, which is essential for reproducible and scalable scientific research [24]. Microservices architecture breaks down complex software into smaller, independent parts that work together, making it easier to update and scale applications. DHRP microservices are described in Table 1.

Figure 3. Integration of DHRP tools within the AOU digital infrastructure. AOU: All of Us Research Program; DHRP: digital health research platform; EDC: electronic data capture.

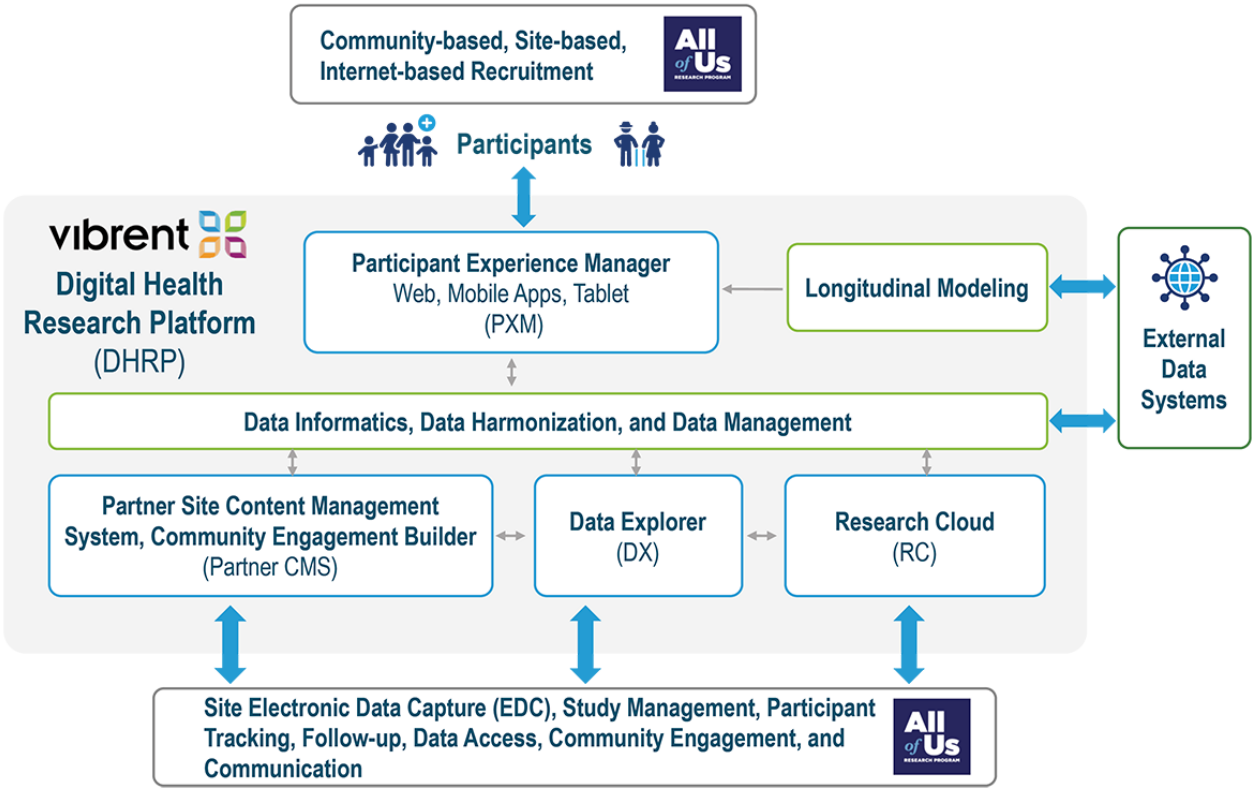

XSL•FO
RenderX

Figure 4. Schematic of the DHRP end-to-end unified data informatics platform and its PXM, RC, and DX tools facilitating collaboration among research teams in a large multisite consortium. DHRP: digital health research platform; DX: Data Explorer; PHI: protected health information; PXM: Participant Experience Manager; RC: Research Cloud.

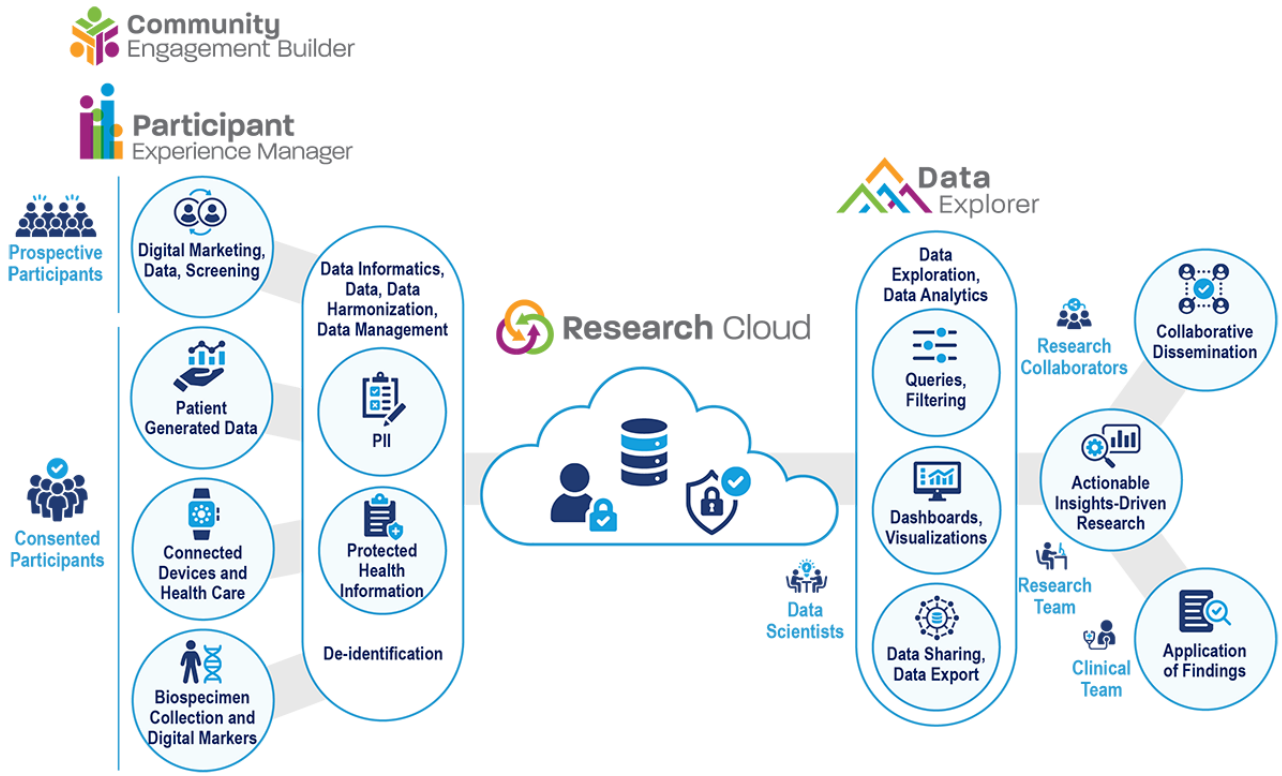

XSL•FO
RenderX

Table 1. Microservices used in the DHRP[a].

Microservice	Function
Ancillary study integration	Facilitates integration with third-party study tools
Appointment management	Schedule, view, and update appointments with maps and geolocation
Asynchronous messaging	Advanced data synchronization service enabling large-scale cross-platform integration
Authentication services	Enables authentication and single sign-on
Case management	Create cases with specific actions to take for a group of participants
Communications services	Transmit targeted email or SMS text or push communications to mobile apps
Comparative insights	Provide demographic-based results (by age, location, etc) for survey responses
Dashboards	Canned visualizations of protocol activities to assist research staff in overseeing program activities; use default dashboards or create your own customized dashboard
Data integration	Integrate with an external data hub
Data management	Scalable storage of large volumes of multisource data
Data retrieval services, multiple	Retrieve data from various sources: Medicare, Epic, Cerner, Athena Health, Fitbit, Apple HealthKit
Deep linking	Communications (email and SMS text) linked to specific steps in the participant workflow to improve participant engagement
eConsents[b] and agreements	Manages primary and additional consent agreement modules
File-sharing services	Ability to manage upload and sharing of test results with participants
Fulfillment services	Order management and delivery tracking of biosample kits and other assets required to support data collection
Identity verification	Secure, remote verification of participants for account access
Insights	Search, analyze, and visualize data about participants and study activities; ability to create custom dashboards, reports, and data exports based on the unique needs of a study
Mobile apps	Study branded mobile apps for participant access; supports iOS and Android clients and adaptive web for participants choosing not to use mobile apps
No-login experiences	Enable participants to complete survey activities without authentication
Reports and exports	Create customized reports and exports for use with study engagement activities
Segmentation	Create complex filtering criteria to identify a group of participants to target for a protocol activity and engagement and may be used to create a case or transmit communications
Survey import	Import survey definitions and participant survey data from third party–hosted survey platforms
Task management	Create and view appointments, cases, and follow ups
Time-based event management	Create user engagement workflows based on time-based protocol requirements
UX[c] and advance workflow customization	Enable the creation of simplified user engagements and support for complicated time- and event-driven workflow; includes configurable page and navigation items, business rules, and conditions

[a]DHRP: digital health research platform.

[b]eConsent: electronic informed consent.

[c]UX: user experience.

DHRP Participant Experience Tool

The PXM facilitated all participant-facing study protocol activities, including (1) electronic informed consent (eConsent), (2) study data collection and data sharing, and (3) secure bidirectional communication. The PXM operates in English and Spanish and can be translated to more than 30 additional languages. Table 2 describes the third-party services that were integrated with the DHRP to support PXM functionality.

The eConsent functionality, content, and UX were developed to accommodate seventh-grade or lower literacy levels through collaborative UX research with participants, researchers, and NIH staff. The key functionality of eConsent included (1) web-based and mobile device accessibility, (2) built-in knowledge reinforcement, and (3) content accessible by text, audio, or video.

The PXM allowed participants to securely access individual participant portals to complete study-related data collection activities, including submission of survey responses, scheduling appointments to complete study measures, and communication with research staff, as shown in Figure 4. The PXM supported both “repeated surveys” and “longitudinal data collection” with defined events, including complex window management that made certain study tasks available in the participant dashboard

XSL•FO
RenderX

during the appropriate window of time. Participants were able to securely review their personal data and genetic results through the PXM.

The PXM was integrated with a range of third-party fast health care interoperability resource (FHIR) systems enabling participants to share EHRs and medical claims data to contribute to the study. Over the course of the AOU, multiple approaches were also used to facilitate the collection of wearable data from participants—all of which made use of the PXM. These included (1) providing opt-in for self- managed (ie, "bring your own device") data contributions using participant-owned Fitbit wearables, Google Fit, and Apple HealthKit; (2) providing Fitbit devices directly to participants to encourage the contribution of data; and (3) providing supportive messaging to participants who were provided with Fitbits to both encourage and monitor the optimal approaches to acquiring wearable and related digital health data. This secure, participant-focused design fostered ease-of-use throughout the study (Figure 5).

As part of an end-to-end readiness assessment for pediatric enrollment, the PXM supports engaging with parents of eligible children as proxy users. Parents of children under the age of 5 years old were invited to enroll their children in the study and were authorized to complete permission, authorization, and data collection modules on behalf of their children [25]. Tools enable study staff to engage with children for biosample and physical measurement collection processes using the PXM. Parents can enroll multiple children and select those children in a drop-down menu. As the AOU expands its pediatric recruitment processes, the PXM will support additional parental roles and permissions, as needed, for participants between the ages of 5 and 18 years at the time of consent.

Table 2. Participant-centric functionality requires scalable integrations with COTS[a].

Participant-facing tools	COTS
Communication (email, SMS, postal/printed mail)	Iterable, Lob, Twilio
Genomics counseling and return of information	Color
Remote biosample collection for assays and genomic sequencing (saliva, blood, nasal swabs)	DNA Genotek, Mirimus, Molecular Testing Labs (MTL), Tasso, Quest Diagnostics
EHRs[b]	Epic, Cerner, Athena Health, others
Wearables and biometric devices	Fitbit, Apple HealthKit, Apple Watch, Google Fit
Identity verification	ID.me
Data visualization/exploration, dashboards, analytics	Metabase, Tableau, Google Analytics
Surveys, codebooks	REDCap, Qualtrics
Enable aggregate participant data comparisons at scale	Snowflake
Shipment and logistics	US Postal Service, Aftership, FedEx
Participant data privacy and security	Akamai, HackerOne

[a]COTS: commercial-off-the-shelf services.

[b]EHR: electronic health record.

XSL•FO
RenderX

Figure 5. Screenshot of the PXM dashboard. PXM: Participant Experience Management.

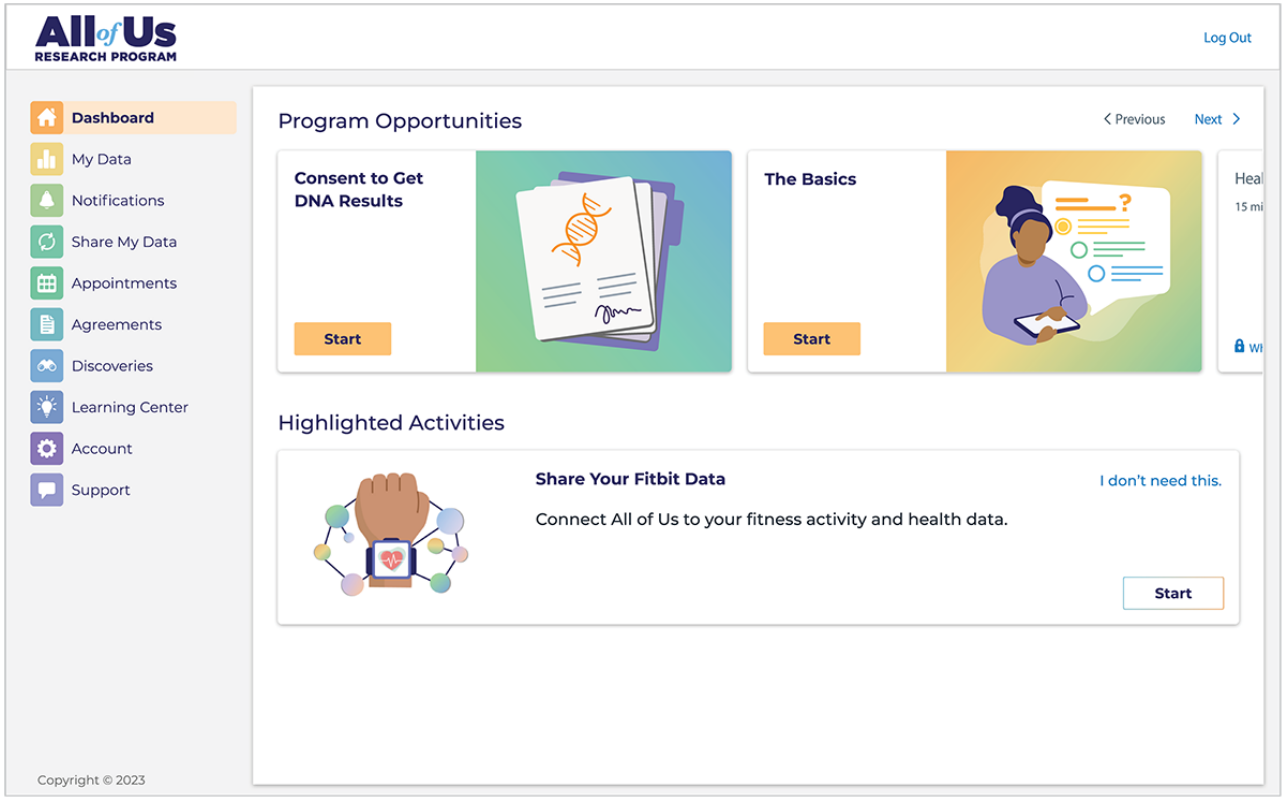

DHRP Research Administration Tools

The RC empowers research staff to complete direct engagement, monitoring, and support to assist participants with completion of study activities. The RC enables research staff to communicate with participants via automated digital marketing, email, and SMS text messaging and to securely collect participant data using built-in computer-assisted telephone interviewing (CATI) capabilities. The RC allowed staff to assist participants with account access, scheduling of study measures visits (in-clinic or remote), and tracking of participant progress via customizable dashboards. The RC was critical in building relationships to keep participants engaged and ensure that study tasks were completed.

For security and privacy, the RC implemented a Roles-Functions-Permissions-based framework system, which granted limited access to select data and workflows based on the research staff's assigned role within the platform. The 13 distinct user roles available in the RC include program coordinators, data analysts, research assistants, and system administrators. After completing 2-step authentication login, RC users can review a role-based data dashboard or download participant-level research data. To facilitate adoption, a training environment was created, which allows RC users to practice their workflows without compromising real study data. The RC can also support expansion to ancillary studies and can extend user roles to allow access to multiple projects through 1 dashboard.

DHRP Data Management Tool

The DX is a self-service analytics tool that enabled study staff to access structured or unstructured data to construct custom reports, dashboards, and insights into collected data and metadata. The DX allows researchers to review and analyze study data from all sources (surveys, collected biometric measures, and third-party shared data) in real time within 1 dashboard. The DX also supports the use of data analytics tools (RStudio, Python, JupyterHub) within the platform on live data. This capability allows researchers to assess data quality and preliminary study results in real time. Modular data pipelines were developed in Python, allowing for scalability and adaptability to many different data sources and data analytics functions.

DHRP Content Management Tool

The content management tool, called CEB, delivered study outreach materials and hosted the study recruitment website. It enabled rapid and low-cost creation and deployment of customized recruitment websites, event marketing, and landing pages tailored to a site's specific catchment area. The CEB includes UX-validated templates to optimize content with appropriate levels of cultural sensitivity and language translation. The CEB was used to generate websites that educate and recruit people to the AOU or ancillary studies and was also used to build the nationwide recruitment website [26].

The CEB was integrated with the RC to provide real-time enrollment metrics to support strategic adjustment of recruitment strategies for study success. Digital marketing analytics generated by the CEB enabled staff to track conversions from both online (recruitment website, lead generation, events landing pages, and social networking links) and offline (QR codes from posters and flyers) sources. By leveraging these analytics, staff gained insights into the effectiveness of different recruitment

XSL•FO
RenderX

methods and made informed decisions to optimize recruitment efforts and meet enrollment and diversity targets.

DHRP Research Staff Training and Support Tools

After the study was launched, there was a recognized need to provide comprehensive and efficient support and training to research staff using DHRP tools. The Help Center and the Vibrent Research Academy were digital training resources created to support researchers' execution of study management and outreach tasks. The Help Center served as a knowledge base for site staff using the RC or CEB tool and provides a detailed explanation of DHRP features, along with supplemental articles, troubleshooting tips, and other resources. The Vibrent Research Academy offers short courses that provide detailed, multimedia demonstrations and instructions on the use of DHRP tools and features.

Ethical Considerations

eConsent Process

Through the PXM, we delivered a modular eConsent process to facilitate a self-paced, dynamic, and adaptive experience for participants [27]. The eConsent process met the regulatory and compliance requirements of all 50 states to support geographic inclusivity and access to participation. A formative consent evaluation module was included to assess participant informedness following a review of each eConsent module. Before providing a digital signature, participants were required to correctly answer 80% of the evaluation questions. Participants were able to withdraw consent from any individual consent module at any time, with the option to reconsent again in the future. Digital copies of completed consent forms were stored in the PXM. This same mechanism also facilitated electronic assent on behalf of eligible minors.

Data Security, Encryption, Regulation, and Compliance

All participant-facing materials were approved by a centralized Institutional Review Board. The DHRP maintains a "defense in depth" security posture to keep participant data safe and private. Data security and privacy are implemented in accordance with established industry-standard policies, procedures, and technology system guidelines, such as Federal Information Security Modernization Act (FISMA) and National Institute of Standards and Technology (NIST) Cybersecurity Framework for Security and Privacy Controls (NIST SP 800-53 r5, NIST SP 800-39, NIST SP 800-37) and are Health Insurance Portability and Accountability Act (HIPAA) compliant. Technical safeguards include strong data encryption (AES256, TLS 1.2 and higher) in flight and at rest. The DHRP is compliant with 21 CFR part 11 and ensures entries come from an authenticated and authorized source. Role-based permissions ensure that only authorized users are permitted to change settings.

Community Engagement

The AOU is a population-based precision medicine initiative operating on a community engagement structure [15,21]. To operationalize the program's core values, a community-based participatory research approach was implemented to develop a national network of actively engaged community organizations and individuals to build community advisory boards (CABs). CABs collaborated with health care provider organizations (HPOs), local community centers, academic and industry partners to provide study guidance.

Community engagement partners facilitated direct engagements between prospective participants and members of their specific communities, which could take the form of sending structured electronic communications, setting up in-person opportunities at events and physical facilities, promoting by word of mouth, and providing biosample collection opportunities when supported by other program partners. These partners encourage initial enrollment as their primary function.

Participant Recruitment and Enrollment

Participant recruitment, eligibility, and enrollment were delivered through in-person and digital methods facilitated by the DHRP [15,21]. Outreach was conducted at local community events and conferences, by word of mouth, digital advertising, and physical distribution of study-branded flyers and promotional items. The CEB tool of DHRP provided customized, culturally appropriate recruitment webpages in multiple languages. Participants who enrolled remotely were able to communicate with staff by phone or email via the PXM. Remote participants were also able to use the PXM to locate the nearest site at which to complete their physical study measurements. Fully remote participants could independently complete all activities, including consent, surveys, EHR data sharing, biosample collection, and communication with study staff, without reporting in person to an HPO site.

Participants were also recruited in person at HPO and mobile clinic sites throughout the United States, with enrollment completed in the DHRP, with or without assistance from site staff. A myriad of program partners used the DHRP to support recruitment, enrollment, and retention. Health care–providing organizations, such as large health systems and federally qualified health centers, facilitated both initial recruitment in the same manner as community engagement partners and directly facilitating (1) consent, (2) completion of electronic surveys, (3) completion of biosample collection in both remote and in-person engagements, and (4) EHR data collection, as well as providing human support for all program activities when a program participant was recruited from their patient populations.

Study Data Collection

All study data were collected in the DHRP. Participants used the PXM tool to submit surveys and complete data sharing from third-party services. To minimize the participant burden, all surveys were available in separate, self-paced modules, allowing participants to save their progress and return to the modules for completion later. To support data collection, research staff used the DX tool to identify participants with incomplete study tasks, such as eConsents or surveys. The RC was then used to develop targeted engagement campaigns to remind participants of any remaining incomplete study tasks requiring attention. A description of the sources and types of data collected in the DHRP is available in Table 3.

XSL•FO
RenderX

Table 3. Multisource data types collected in the DHRP[a].

Data source	Data formats
eConsent[b] signature	PDF
EHRs[c]	JavaScript Object Notation (JSON)–formatted FHIR[d] payloads
Consumer wearables (Fitbit, Apple HealthKit)	Apple HealthKit health data: JSON Apple HealthKit EHR data: JSON
Biometric measures	JSON-formatted FHIR payloads
Genetic results	PDF reports
Surveys	JSON-formatted FHIR payloads: questionnaire response
Personally identifiable information	JSON-formatted FHIR payloads
Participant communication	Comma-separated value (CSV), JSON
Participant communication	CSV
Prospective participants	CSV
Cognitive assessments	JSON-formatted FHIR payloads
Pediatric assent and surveys	PDF and JSON-formatted FHIR payloads: questionnaire response
Digital marketing data	CSV, Urchin Tracking Module (UTM) parameters

[a]DHRP: digital health research platform.

[b]eConsent: electronic informed consent.

[c]EHR: electronic health record.

[d]FHIR: fast health care interoperability resource.

Results

Participant Characteristics

The DHRP was launched for use in May 2018. As of April 2024, a total of 705,719 participants had enrolled in the AOU through the DHRP. Of these participants, 74% (n=524,264) completed all study activities in the core study protocol, including all required survey modules. In this cohort, 55% (n=386,377) identified as women, 51% (n=360,303) were aged 35-64 years, and 48% (n=340,423) identified as White. The platform was used by participants and research teams in each of the 50 US states and territories, with 8% (n=58,211) participants representing rural areas (Figure 6).

The AOU used the DHRP to build a diverse research cohort in which 87% (n=457,514) identified as members of groups underrepresented in biomedical research (UBR) in at least 1 demographic category (age, race, sex, gender, income, education, or disability). More specifically, 46% (n=243,296) participants identified as racial minorities, 36% (n=189,087) reported a low income, and 20% (n=104,916) reported a disability. One third of the cohort (n=219,355, 31%) were aged 65 years and over, 8% (n=58,211) identified as rural residents, and 28% (n=197,601) were people with limited access to health care. More than 5% (n=~37,000) of all program participants opted to participate in the program in Spanish. UBR diversity metrics of the cohort are available in Table 4.

XSL•FO
RenderX

Figure 6. Geographic distribution of AOU participants enrolled through the DHRP by US state as of March 21, 2024. State enrollment varied from 13.2% (n=86,803) in California to 0.1% (n=706) in New York. The western region had the highest enrollment (28.1%). AOU: All of Us Research Program; DHRP: digital health research platform.

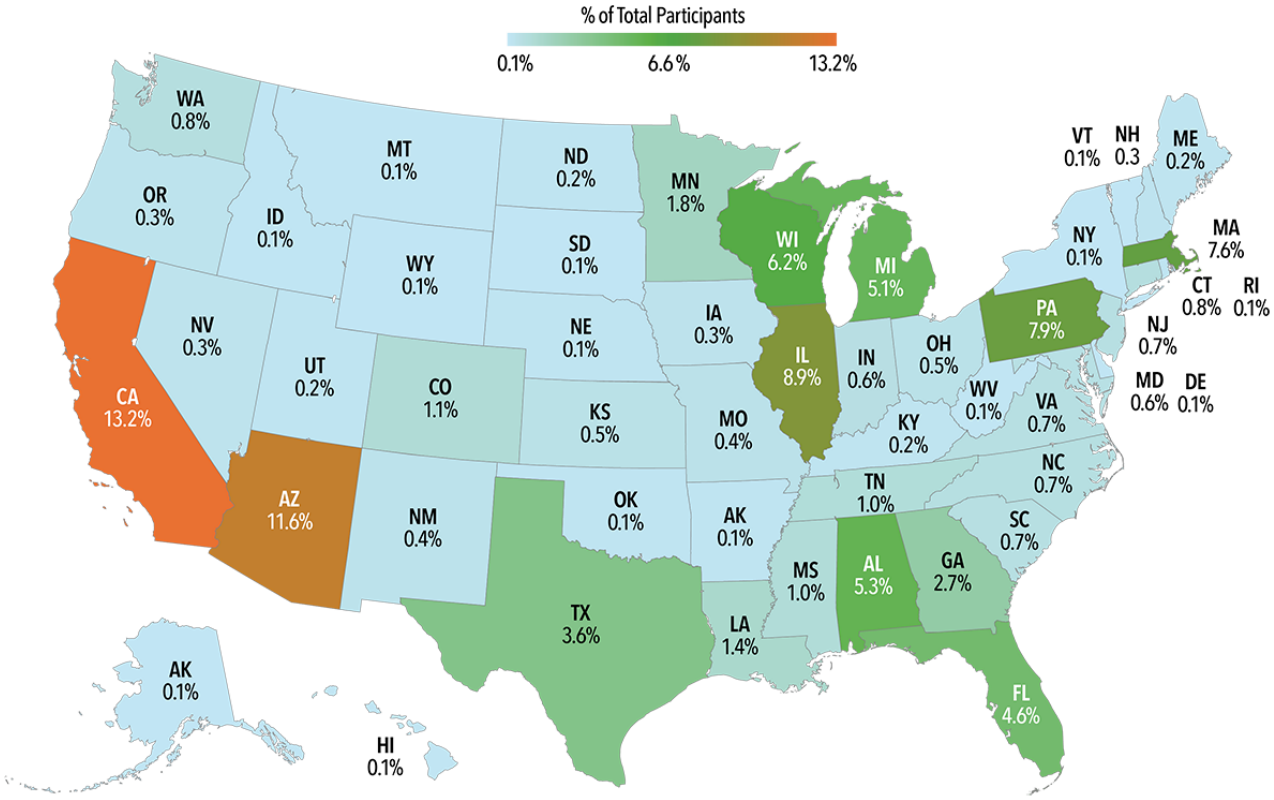

XSL•FO
RenderX

Table 4. AOU[a] participants' (N=705,719) sociodemographic characteristics as of April 2024.

Characteristics	Participants, n (%)
Age group (years)	
18-24	25,707 (3.6)
25-44	213,958 (30.3)
45-64	246,699 (35.0)
65-84	205,072 (29.0)
≥85	14,283 (2.0)
Gender	
Man	237,070 (33.6)
Woman	386,377 (54.7)
Nonbinary	2729 (0.4)
Transgender	1169 (0.2)
Other/multiple selections	4010 (0.5)
Unspecified/no answer	74,364 (10.5)
Racial identity	
White	340,423 (48.2)
Black or African American	102,371 (14.5)
Asian	22,281 (3.2)
Multirace/other	55,872 (7.9)
None/no answer	86,477 (12.3)
Education	
Less than high school	51,236 (7.3)
Grade 12 or General Educational Development (GED)	113,545 (16.1)
College, 1-3 years	165,317 (23.4)
College, 4 years or more	148,910 (21.1)
Advanced degree	42,504 (20.2)
Unspecified/no answer	84,207 (11.9)
Annual household income (US $)	
≤24,999	143,939 (20.1)
25,000-49,999	96,800 (16.0)
50,000-74,999	70,599 (10.0)
75,000-99,999	55,0752 (7.8)
100,000-199,000	101,965 (14.4)
≥200,000	45,136 (6.4)
Unspecified/no answer	188,205 (26.7)
Rurality	
Rural	58,211 (8.2)
Nonrural (metro/suburban)	644,852 (91.4)
Unspecified/no answer	2656 (0.4)

[a]AOU: All of Us Research Program.

XSL•FO
RenderX

Data and Tool Usage

Approximately 4,500,000 combined survey and eConsent modules have been completed through the DHRP. Analysis of eConsent formative evaluation questions revealed that after reviewing eConsent modules, over 95% (n=670,433) of the participants were able to differentiate the research program from medical care, understand the voluntary nature of their involvement, and comprehend their right to withdraw at any point in time [28]. Complete data collection metrics are in Table 5.

Table 5. Data collected and tool usage metrics from the DHRP[a] for the AOU[b].

Measures	Data, n
Participant data sources	
Fitbit	49,627
Apple HealthKit	19,940
EHRs[c]	14,431
Total survey modules completed	3,685,608
Total physical biometric study measures	538,160
Saliva kits	59,819
eConsent[d]	
Total primary eConsent modules completed	2,015,622
EHR eConsent module	632,644
Genomics return of results eConsent module	406,781
Cognitive assessment eConsent module	85,572
Digital marketing	
Prospective participants contacted	378,010
Automated email and SMS text campaigns delivered	3677
Custom HTML templates created by sites in the partner CEB[e] tool	1838
Multichannel communication automations (bilingual)	≥300
Participant multimodal communications	
Emails sent to participants	64,837,083
SMS text messages sent to participants	7,794,054
Mobile app push notification messages sent to participants	463,975
Direct mail sent to participants	1,531,814
Site staff remote engagement with participants	1,168,079
Protocol step completions logged by site staff due to remote engagements	412,026
CATI[f] sessions	42,795
Appointments scheduled (remote and site based)	301,393

[a]DHRP: digital health research platform.

[b]AOU: All of Us Research Program.

[c]EHR: electronic health record.

[d]eConsent: electronic informed consent.

[e]CEB: Community Engagement Builder.

[f]CATI: computer-assisted telephone interviewing.

The use of CATI capabilities within the DHRP demonstrated the positive effects of combining digital and in-person engagement methods to enhance the participation of UBR groups in research. Among participants who completed the Social Determinants of Health (SDOH) survey module (n=276,706, 42%, eligible participants) for the study, CATI technology supported significant portions (as high as 47%) of survey completions within cross-sectional UBR groups. Cross-sectional analysis by participant age, education, income, and race-ethnicity showed increased use of CATI-assisted sessions within certain UBR groups. Specifically, the increased use of CATI was evident among those with education levels below grade 9, those with an income below US $35,000/year, people aged 55 years and older, and people identifying as

XSL•FO
RenderX

non-Hispanic Black or Hispanic of any race. The cross-sectional use of CATI is shown in Figures 7, 8, and 9. Those with lower education levels, of older age, with lower incomes, or from racial and ethnic minorities are recognized as UBR populations [29].

Another demonstration of the effective engagement capabilities of the DHRP was found in the completion rates of the optional COVID-19 Pandemic Evaluation (COPE) survey. The COPE survey was added in response to the COVID-19 pandemic, which was first confirmed in the United States in January 2020. The COPE survey was launched in the DHRP in May 2020, and 6 iterations were developed, modified, and released to participants through February 2021. A previous study of the COPE survey showed that the modification of surveys led to a significant increase (P<.001) in the participant response rate over 10 months of survey delivery [30]. Overall, completion rates for the COPE survey among individuals enrolled under UBR categories showed an increase from 11% in the first survey to 16% in the final survey. These findings underscore the significance of using a digital platform able to accommodate rapid implementation of protocol amendments.

Figure 7. Heat map showing cross-sectional use of CATI by age and education level. CATI: computer-assisted telephone interviewing; GED: General Educational Development.

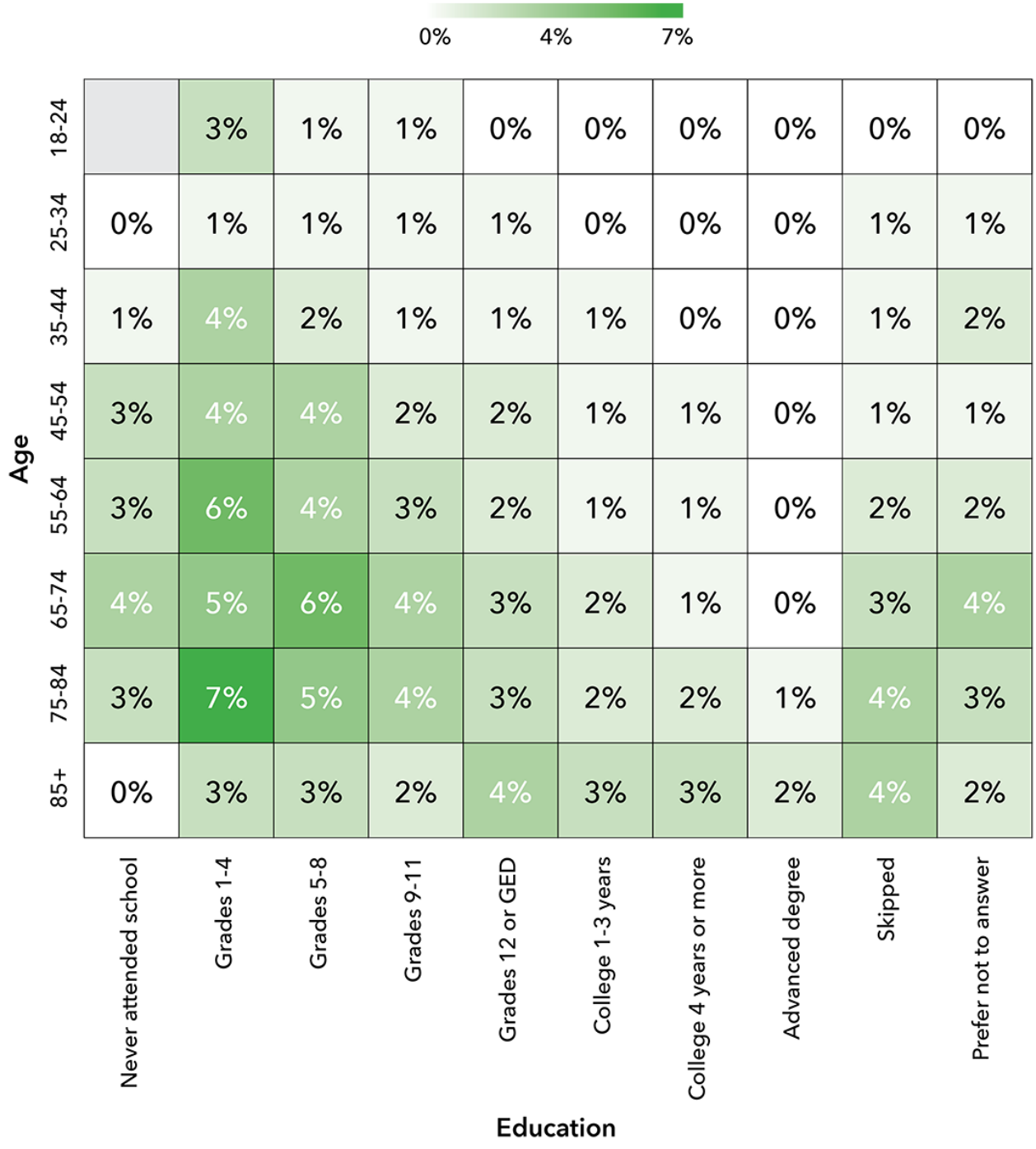

XSL•FO
RenderX

Figure 8. Heat maps showing cross-sectional use of CATI by age and income. CATI: computer-assisted telephone interviewing.

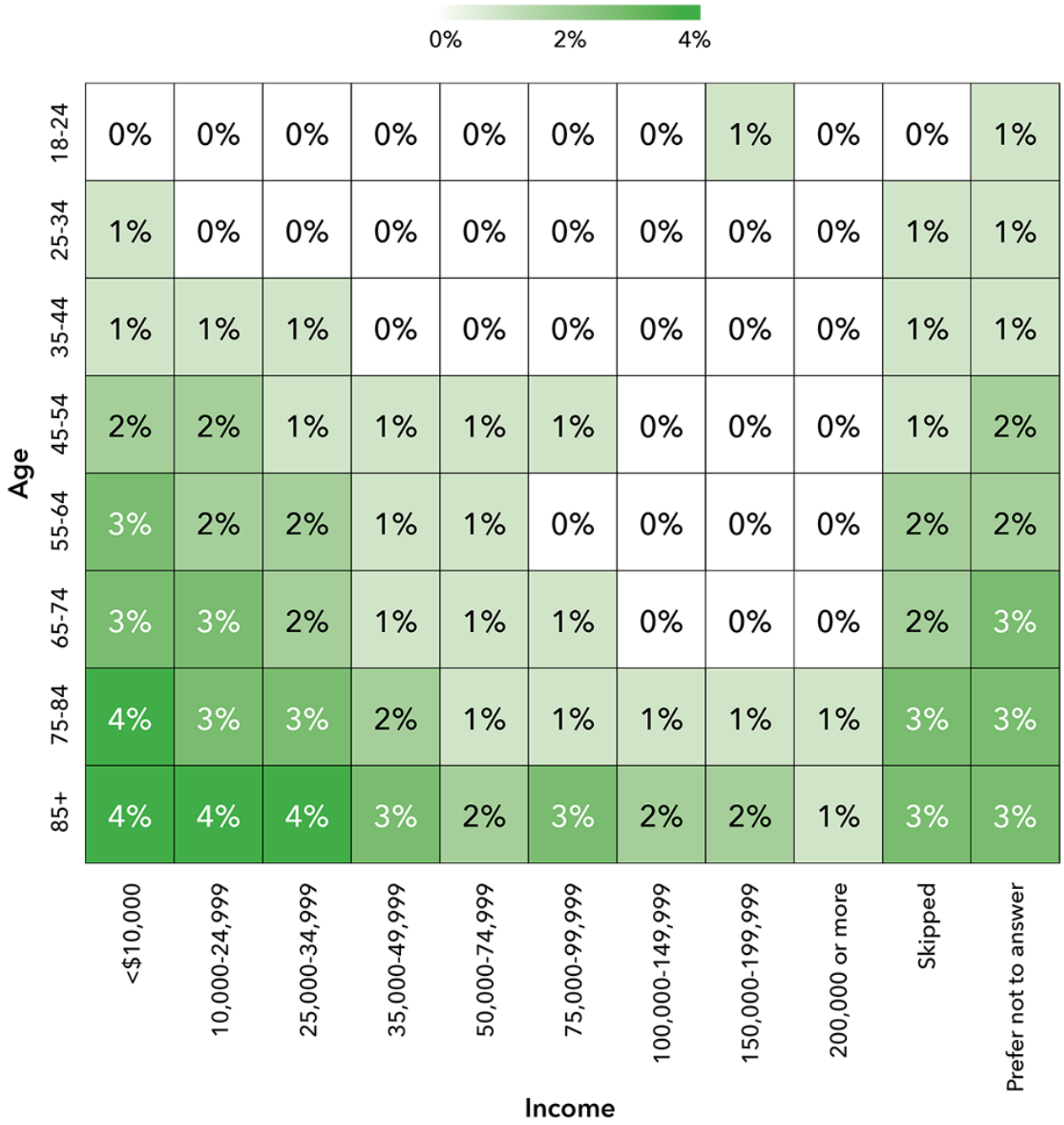

XSL•FO
RenderX

Figure 9. Heat maps showing cross-sectional use of CATI by age, race, and ethnicity. AIAN: American Indian or Alaska Native; CATI: computer-assisted telephone interviewing; NHPI: Native Hawaiian or Pacific Islander; MENA: Middle Eastern or North African.

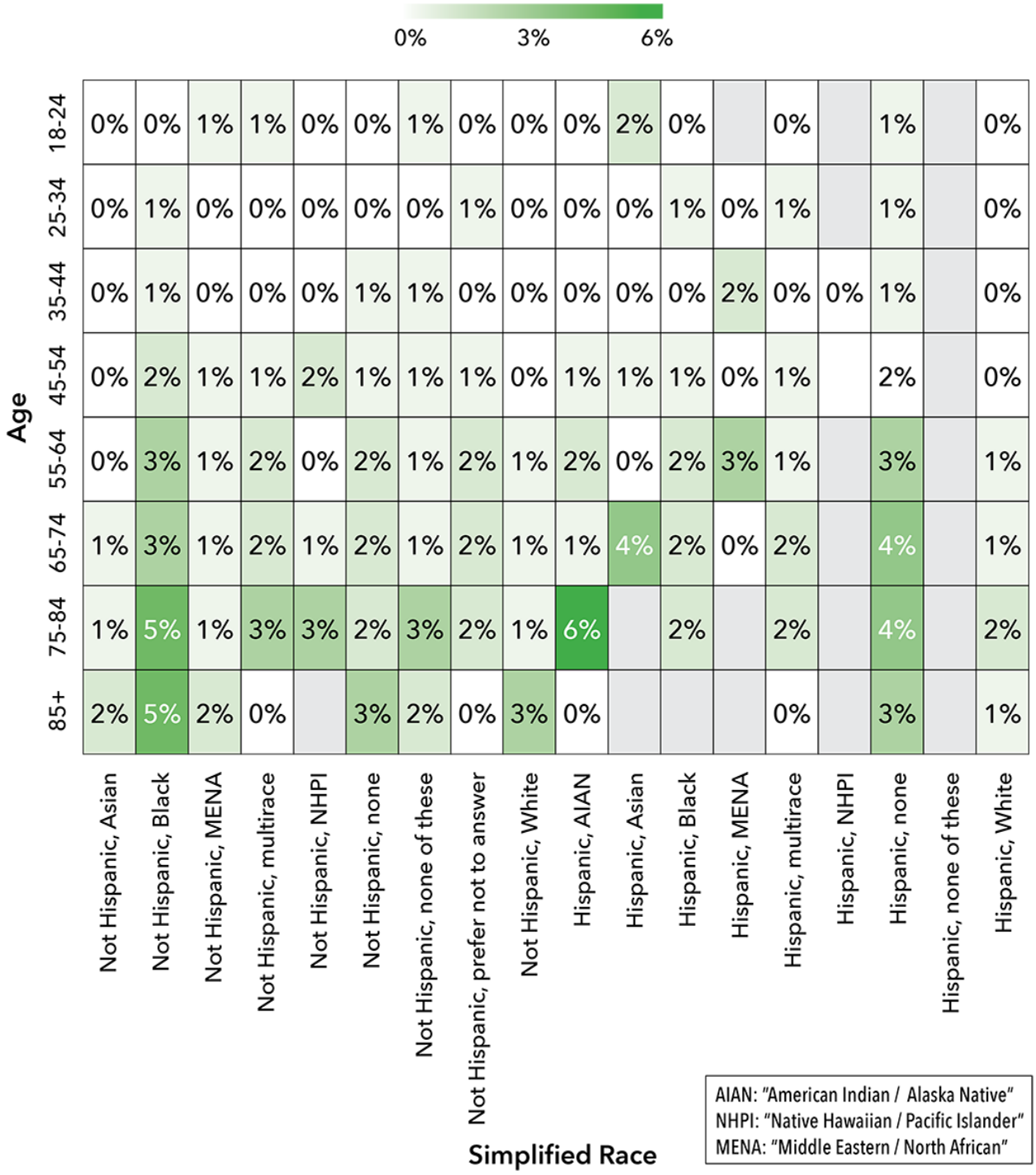

Discussion

Principal Findings

The design and implementation of the DHRP in this study demonstrates that a DHRP with technical architecture for participant-focused design can effectively be used to recruit, engage, and collect longitudinal health data from participants with geographic, racial, and sociodemographic diversity. The collaborative, community-informed approach used to develop the DHRP, combined with validation through UX research, contributed to its effective design for use in diverse populations [31]. The participant-centric design of the DHRP may be broadly applicable to increase community engagement of people from diverse backgrounds in other research study designs, including cross-sectional, case-control, and experimental clinical trials [32].

The principles of generalizability, usability, cultural sensitivity, and accessibility were consistently implemented throughout the human-centered design and building of the DHRP for this study [33]. The diversity of participant users, combined with the high eConsent completion rates, support the idea that the PXM is,

XSL•FO
RenderX

in fact, participant friendly, accessible, and usable for various groups, as intended. The use of the DHRP's CATI technology in this study also demonstrates the successful use of digitally enabled, human-delivered engagement methods to enhance participation of UBR groups. In this study, CATI enhanced data collection and increased study completion rates of the SDOH survey module among people with lower education levels, those with lower incomes, older people, and those from racial and ethnic minority groups. This shows that the DHRP can be used to increase cohort diversity and build trust with UBR groups [34]. In addition, the increase in engagement demonstrated by the COPE survey results shows that the researcher tools, RC and CEB, are integral to facilitating participant engagement within DHRP. Furthermore, integrating the participant-facing tool (PXM) with the researcher tools (RC and CEB) in a containerized manner protected participant autonomy and privacy, while also maintaining workflow automation and bidirectional communication between participants and study staff.

From an engagement perspective, it was important to establish a standardized eConsent process that was accessible and comprehensive for all individual platform users regardless of their geographical location, method of enrollment, preferred language, educational level, digital literacy, or digital device used [27]. PXM eConsent modules were successfully used with high participant informedness demonstrated by high rates of completion of the formative consent evaluation. Well-designed and well-delivered eConsent provides transparency to participants at study initiation and can increase volunteerism, participant understanding of research, and participant satisfaction [28]. This can lead to increased trust and future participation in research among groups historically underrepresented in the field [27,35]. This result supports the idea that this digital research platform can enhance the delivery of longitudinal studies and can be used by participants with different levels of digital literacy.

Strengths and Limitations

This novel study has several strengths. This demonstration helps shed light on how digital research platforms, such as the DHRP, can be effectively used to increase the participation of UBR groups. The collaborative approach used in the design of the DHRP ensured the platform was able to meet the needs of people from various sociodemographic backgrounds and with varying levels of comfort with digital technology. The comprehensive design of the DHRP tools appears to have decreased participant barriers to research participation commonly reported in the literature [36]. These aspects should be further investigated in future research.

Although this study has numerous strengths, it also presents potential limitations. Although the cohort is 8% rural (n=58,211), it does not fully represent the current US population, which is 13.8% rural by county definition [37]. This limitation is a function of the overall AOU research processes that determine the geographic distribution of participants based on the availability of local affiliated recruitment sites and is not a direct limitation of the DHRP itself.

Another possible limitation of this study is the lack of available data with which to evaluate the use of the DHRP within the pediatric population. This is a result of the AOU being in its initial phases of enrolling pediatric participants. Although the DHRP was designed and built to accommodate the complexities of consenting, engaging, and collecting data for pediatric research, we currently do not have adequate data to conduct a conclusive analysis of the efficacy of the DHRP within this group. Therefore, we are unable to report on potentially unforeseen challenges related to the use of the DHRP in pediatric research, as they have not yet been encountered in this study.

Conclusion

We created and validated a digital research platform to support the development of a large, nationwide, community-engaged, longitudinal cohort study (AOU) of diverse participants from UBR groups. This study demonstrates the promising potential of digital platforms to enable the curation, harmonization, and management of high-volume longitudinal multisource data sets from diverse populations for precision medicine research. Digital platforms may provide solutions to researchers who aim to broaden their catchment area in site-based, decentralized, and hybrid research methods and support the development of more diverse research cohorts. With continued development, digital research platforms, such as the DHRP described here, may lead to the successful establishment of best approaches to engage health-disparate, minority, vulnerable, and other UBR populations into clinical research to broadly improve the overall health of all communities.

Acknowledgments

We would like to thank all collaborators who contributed to the design and implementation of Vibrent Health's digital health research platform. We would also like to acknowledge all participants of the All of Us Research Program. The study was funded by National Institutes of Health grants (OTA-20-001 and RFA-PM-16-003).

Authors' Contributions

Manuscript concept and design were completed by PJ, DK, MB, and AM in collaboration with All of Us Research Program stakeholders. Data acquisition was completed by S Sutherland. Data analysis and interpretation was completed by AM and MB. The manuscript was drafted by AM, DK, MB, and PJ. Manuscript revision was led by AM and completed by all coauthors. Theoretical support was provided by DK, S Sawyer, JLM, DK, and PJ.

XSL•FO
RenderX

Conflicts of Interest

MB, AM, DK, S Sutherland, DJ, AA, MP, and PJ are employees of Vibrent Health, Inc. MB, S Sawyer, JLM, DK, and PJ are affiliated with the All of Us Research Program.

References

1. Caruana EJ, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. J Thorac Dis 2015 Nov;7(11):E537-E540 [FREE Full text] [doi: 10.3978/j.issn.2072-1439.2015.10.63] [Medline: 26716051]
2. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature 2016 Oct 13;538(7624):161-164 [FREE Full text] [doi: 10.1038/538161a] [Medline: 27734877]
3. Allison K, Patel D, Kaur R. Assessing multiple factors affecting minority participation in clinical trials: development of the Clinical Trials Participation Barriers Survey. Cureus 2022 Apr;14(4):e24424 [FREE Full text] [doi: 10.7759/cureus.24424] [Medline: 35637812]
4. Hamel LM, Penner LA, Albrecht TL, Heath E, Gwede CK, Eggly S. Barriers to clinical trial enrollment in racial and ethnic minority patients with cancer. Cancer Control 2016 Oct;23(4):327-337 [FREE Full text] [doi: 10.1177/107327481602300404] [Medline: 27842322]
5. Ejiogu N, Norbeck JH, Mason MA, Cromwell BC, Zonderman AB, Evans MK. Recruitment and retention strategies for minority or poor clinical research participants: lessons from the Healthy Aging in Neighborhoods of Diversity across the Life Span study. Gerontologist 2011 Jun;51 Suppl 1(Suppl 1):S33-S45 [FREE Full text] [doi: 10.1093/geront/gnr027] [Medline: 21565817]
6. Allmark P. Should research samples reflect the diversity of the population? J Med Ethics 2004 Apr;30(2):185-189 [FREE Full text] [doi: 10.1136/jme.2003.004374] [Medline: 15082815]
7. Gross AS, Harry AC, Clifton CS, Della Pasqua O. Clinical trial diversity: an opportunity for improved insight into the determinants of variability in drug response. Br J Clin Pharmacol 2022 Jun 17;88(6):2700-2717 [FREE Full text] [doi: 10.1111/bcp.15242] [Medline: 35088432]
8. Clark LT, Watkins L, Piña IL, Elmer M, Akinboboye O, Gorham M, et al. Increasing diversity in clinical trials: overcoming critical barriers. Curr Probl Cardiol 2019 May;44(5):148-172 [FREE Full text] [doi: 10.1016/j.cpcardiol.2018.11.002] [Medline: 30545650]
9. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2017 Jan 01;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
10. Spartano NL, Lin H, Sun F, Lunetta KL, Trinquart L, Valentino M, et al. Comparison of on-site versus remote mobile device support in the Framingham Heart Study using the Health eHeart Study for Digital Follow-up: randomized pilot study set within an observational study design. JMIR Mhealth Uhealth 2019 Sep 30;7(9):e13238 [FREE Full text] [doi: 10.2196/13238] [Medline: 31573928]
11. Hwang DA, Lee A, Song JM, Han H. Recruitment and retention strategies among racial and ethnic minorities in web-based intervention trials: retrospective qualitative analysis. J Med Internet Res 2021 Jul 12;23(7):e23959 [FREE Full text] [doi: 10.2196/23959] [Medline: 34255658]
12. Dockendorf MF, Hansen BJ, Bateman KP, Moyer M, Shah JK, Shipley LA. Digitally enabled, patient-centric clinical trials: shifting the drug development paradigm. Clin Transl Sci 2021 Mar;14(2):445-459 [FREE Full text] [doi: 10.1111/cts.12910] [Medline: 33048475]
13. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
14. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
15. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" Research Program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
16. Armenta A, Serrano A, Cabrera M, Conte R. The new digital divide: the confluence of broadband penetration, sustainable development, technology adoption and community participation. Inf Technol Dev 2011 Oct 17;18(4):345-353. [doi: 10.1080/02681102.2011.625925]
17. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
18. Mapes BM, Foster CS, Kusnoor SV, Epelbaum MI, AuYoung M, Jenkins G, All of Us Research Program. Diversity and inclusion for the All of Us research program: a scoping review. PLoS One 2020;15(7):e0234962 [FREE Full text] [doi: 10.1371/journal.pone.0234962] [Medline: 32609747]

XSL•FO
RenderX

19. Ghadessi M, Di J, Wang C, Toyoizumi K, Shao N, Mei C, et al. Decentralized clinical trials and rare diseases: a Drug Information Association Innovative Design Scientific Working Group (DIA-IDSWG) perspective. Orphanet J Rare Dis 2023 Apr 11;18(1):79 [FREE Full text] [doi: 10.1186/s13023-023-02693-7] [Medline: 37041605]
20. Nijhawan LP, Janodia MD, Muddukrishna BS, Bhat KM, Bairy KL, Udupa N, et al. Informed consent: issues and challenges. J Adv Pharm Technol Res 2013 Jul;4(3):134-140 [FREE Full text] [doi: 10.4103/2231-4040.116779] [Medline: 24083200]
21. All of Us Research Program overview. National Institutes of Health. URL: https://allofus.nih.gov/about/program-overview [accessed 2024-12-05]
22. Doerr M, Grayson S, Moore S, Suver C, Wilbankson J, Wagner J. Implementing a universal informed consent process for the All of Us Research Program. Biocomputing 2019;2018:427-438. [doi: 10.1142/9789813279827_0039]
23. Participant Technology Systems Center. National Institutes of Health. URL: https://allofus.nih.gov/funding-and-program-partners/participant-technology-systems-center [accessed 2024-04-21]
24. Casalicchio E, Iannucci S. The state - of - the - art in container technologies: application, orchestration and security. Concurrency Comput 2020 Jan 19;32(17):e5668. [doi: 10.1002/cpe.5668]
25. All of Us Research Program protocol. National Institutes of Health. URL: https://allofus.nih.gov/about/all-us-research-program-protocol [accessed 2024-07-16]
26. All of Us Research Program. National Institutes of Health. URL: https://www.joinallofus.org/ [accessed 2024-05-01]
27. Doerr M, Moore S, Barone V, Sutherland S, Bot BM, Suver C, et al. Assessment of the All of Us research program's informed consent process. AJOB Empir Bioeth 2021 Dec 04;12(2):72-83. [doi: 10.1080/23294515.2020.1847214] [Medline: 33275082]
28. Earl CE, Penney PJ. The significance of trust in the research consent process with African Americans. West J Nurs Res 2001 Nov 01;23(7):753-762. [doi: 10.1177/01939450122045528] [Medline: 11675799]
29. Notice of NIH's interest in diversity (NOT-OD-20-031). National Institutes of Health. 2019 Nov 22. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-031.html [accessed 2024-12-05]
30. Schulkey CE, Litwin TR, Ellsworth G, Sansbury H, Ahmedani BK, Choi KW, et al. Design and implementation of the All of Us Research Program COVID-19 Participant Experience (COPE) Survey. Am J Epidemiol 2023 Jun 02;192(6):972-986 [FREE Full text] [doi: 10.1093/aje/kwad035] [Medline: 36799620]
31. Skarlatidou A, Ponti M, Sprinks J, Nold C, Haklay M, Kanjo E. User experience of digital technologies in citizen science. JCOM 2019;18(01):E [FREE Full text] [doi: 10.22323/2.18010501]
32. Guarino J, Parvanova I, Finkelstein J. Characteristics of electronic informed consent platforms for consenting patients to research studies: a scoping review. Stud Health Technol Inform 2022 Jun 06;290:777-781 [FREE Full text] [doi: 10.3233/SHTI220184] [Medline: 35673123]
33. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020 Jan;21(1):37-48. [doi: 10.1177/1524839919850557] [Medline: 31131633]
34. Peterson R, Hedden SL, Seo I, Palacios VY, Clark EC, Begale M, et al. Rethinking data collection methods during the pandemic: development and implementation of CATI for the All of Us Research Program. J Public Health Manag Pract 2024;30(2):195-199. [doi: 10.1097/PHH.0000000000001846] [Medline: 38271102]
35. Tan RKJ, Wu D, Day S, Zhao Y, Larson HJ, Sylvia S, et al. Digital approaches to enhancing community engagement in clinical trials. NPJ Digit Med 2022 Mar 25;5(1):37 [FREE Full text] [doi: 10.1038/s41746-022-00581-1] [Medline: 35338241]
36. Heffernan ME, Barrera L, Guzman ZR, Golbeck E, Jedraszko AM, Hays PT, et al. Barriers and facilitators to recruitment of underrepresented research participants: perspectives of clinical research coordinators. J Clin Transl Sci 2023;7(1):e193 [FREE Full text] [doi: 10.1017/cts.2023.611] [Medline: 37745931]
37. Davis JC, Cromartie J, Farrigan T, Genetin B, Sanders A, Winikoff JB. Rural America at a glance. U.S. Department of Agriculture. 2023 Nov. URL: https://search.nal.usda.gov/discovery/fulldisplay?context=L&vid=01NAL_INST:MAIN&docid=alma9916413333507426 [accessed 2024-11-30]

Abbreviations

AOU: All of Us Research Program
CAB: community advisory board
CATI: computer-assisted telephone interviewing
CEB: Community Engagement Builder
COPE: COVID-19 Pandemic Evaluation
COTS: commercial-off-the-shelf services
DHRP: digital health research platform
DX: Data Explorer
eConsent: electronic informed consent
EHR: electronic health record
FHIR: fast health care interoperability resource
HPO: health care provider organization

XSL•FO
RenderX

NIH: National Institutes of Health
NIST: National Institute of Standards and Technology
PXM: Participant Experience Manager
RC: Research Cloud
SDLC: software development life cycle
SDOH: Social Determinants of Health
UBR: underrepresented in biomedical research
UX: user experience

Edited by G Eysenbach; submitted 03.05.24; peer-reviewed by Y Zhang, S Rasania, L Guo; comments to author 05.07.24; revised version received 13.08.24; accepted 25.11.24; published 15.01.25.

Please cite as:
Klein D, Montgomery A, Begale M, Sutherland S, Sawyer S, McCauley JL, Husbands L, Joshi D, Ashbeck A, Palmer M, Jain P
Building a Digital Health Research Platform to Enable Recruitment, Enrollment, Data Collection, and Follow-Up for a Highly Diverse Longitudinal US Cohort of 1 Million People in the All of Us Research Program: Design and Implementation Study
J Med Internet Res 2025;27:e60189
URL: https://www.jmir.org/2025/1/e60189
doi:10.2196/60189
PMID:

©Dave Klein, Aisha Montgomery, Mark Begale, Scott Sutherland, Sherilyn Sawyer, Jacob L McCauley, Letheshia Husbands, Deepti Joshi, Alan Ashbeck, Marcy Palmer, Praduman Jain. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Age Transformation in Patient-Physician Communication: 25-Year Narrative Review (1999-2023)

Mingming Song[1], MA; Joel Elson[1], PhD; Dhundy Bastola[1], PhD

University of Nebraska at Omaha, Omaha, NE, United States

Corresponding Author:
Dhundy Bastola, PhD
University of Nebraska at Omaha
1110 S 67th Street
Omaha, NE, 68182
United States
Phone: 1 4025544899
Email: dkbastola@unomaha.edu

Abstract

Background: The evolution of patient-physician communication has changed since the emergence of the World Wide Web. Health information technology (health IT) has become an influential tool, providing patients with access to a breadth of health information electronically. While such information has greatly facilitated communication between patients and physicians, it has also led to information overload and the potential for spreading misinformation. This could potentially result in suboptimal health care outcomes for patients. In the digital age, effectively integrating health IT with patient empowerment, strong patient-physician relationships, and shared decision-making could be increasingly important for health communication and reduce these risks.

Objective: This review aims to identify key factors in health communication and demonstrate how essential elements in the communication model, such as health IT, patient empowerment, and shared decision-making, can be utilized to optimize patient-physician communication and, ultimately, improve patient outcomes in the digital age.

Methods: Databases including PubMed, Web of Science, Scopus, PsycINFO, and IEEE Xplore were searched using keywords related to patient empowerment, health IT, shared decision-making, patient-physician relationship, and health communication for studies published between 1999 and 2023. The data were constrained by a modified query using a multidatabase search strategy. The screening process was supported by the web-based software tool Rayyan. The review methodology involved carefully designed steps to provide a comprehensive summary of existing research. Topic modeling, trend analysis, and synthesis were applied to analyze and evaluate topics, trends, and gaps in health communication.

Results: From a total of 389 selected studies, topic modeling analysis identified 3 primary topics: (1) Patient-Physician Relationship and Shared Decision-Making, (2) Patient Empowerment and Education Strategies, and (3) Health Care Systems and Health IT Implementations. Trend analysis further indicated their frequency and prominence in health communication from 1999 to 2023. Detailed examinations were conducted using secondary terms, including trust, health IT, patient-physician relationship, and patient empowerment, derived from the main topics. These terms clarified the collective impact on improving health communication dynamics. The synthesis of the role of health IT in health communication models underscores its critical role in shaping patient-centered health care frameworks.

Conclusions: This review highlights the significant contributions of key topics that should be thoroughly investigated and integrated into health communication models in the digital age. While health IT plays an essential role in promoting shared decision-making and patient empowerment, challenges such as usability, privacy concerns, and digital literacy remain significant barriers. Future research should prioritize evaluating these key themes and addressing the challenges associated with health IT in health communication models. Additionally, exploring how emerging technologies, such as artificial intelligence, can support these goals may provide valuable insights for enhancing health communication.

(J Med Internet Res 2025;27:e60512) doi:10.2196/60512

KEYWORDS

health communication; health IT; patient empowerment; shared decision-making; patient-physician relationship; trust

XSL•FO
RenderX

Introduction

Background

Technological advancements have transformed patient-physician communication. A fundamental aspect of this transformation is through patient empowerment, which encourages individuals to manage their health proactively according to their personal needs, thereby improving health care outcomes [1]. The adoption of patient empowerment practices and broader access to health information from various sources such as online platforms have shifted from a traditional 1-way communication model to a patient-centered approach that emphasizes collaborative shared decision-making [2-4]. This shift promotes active patient engagement, encouraging them to articulate their needs, preferences, and values within the health care process [5,6]. Additionally, today's health communication is often enhanced by technological advancements that provide patients with extensive health information and tools [7,8]. Telehealth, mobile health (mHealth), patient portals, and recently adopted artificial intelligence (AI) techniques have expanded patients' access to health care and health-related data [9,10]. Telehealth platforms, for example, have made medical consultations more accessible, allowing patients to communicate with health care providers remotely [10,11]. This convenience helps patients actively participate in consultations and make informed decisions about their treatments. Similarly, electronic health records (EHRs) facilitate a continuous flow of information between different health care providers, ensuring that every provider involved in patients' care has access to the same comprehensive data [12]. By 2021, 88% of US office-based physicians had adopted EHR [13]. This widespread adoption ensures that providers have timely and accurate patient data, which is crucial for tailoring treatment plans to individual needs and preferences. Additionally, patient portals give patients 24/7 access to their health records, test results, and direct communication with their health care providers [14]. This transparency encourages patient engagement and fosters active discussion about care plans, ultimately enhancing the effectiveness of patient-physician communication. Despite these benefits, the availability of online health information also introduces risks, such as the potential exposure to misinformation [15-17]. Misinformation can undermine the trust necessary for effective health care relationships, which is critical because trust influences patients' willingness to follow medical advice and maintain open communication with their providers. For instance, the impact of social media on these relationships significantly varies based on the accuracy of the information and the preexisting relationship between patient and physician, which can range from positive to negative [18].

Today's diverse health information technology (health IT) landscape further underscores the need for a refined health communication model that effectively leverages digital tools to enhance patient empowerment across various health care settings [19,20]. While the concepts of patient empowerment and patient-centered care are frequently discussed, their integration into health communication models is still not fully addressed. As digital solutions become standard in health care, it is increasingly important to understand how patient relationships, empowerment, and health IT integrate and interact within health communication models. This extensive review explores essential topics of health communication that fit today's digital age and investigates their associated dynamics to identify mechanisms that facilitate effective health communication between patients and physicians.

Aim and Objectives

The overarching aim of the research is to investigate the essential elements for inclusion in a refined health communication model to enhance patient-physician communication and shared decision-making in the current digital age. Using the PICo framework (Population, Interest, and Context) [21], this study specifies the population as physicians and patients, the interest as key elements for effective communication models, and the context as the impact of health IT on health communication. Particularly, the first objective is to identify and analyze the key topics and factors that have influenced health communication in the past 25 years and should be incorporated into effective health communication models. The second objective involves trend analysis to explore the dynamics and associations among those key topics and factors identified in the first objective. Finally, the last objective focuses on synthesizing the impact of health IT on patient-physician interactions and its subsequent effects on the evolution of health communication models.

However, unlike systematic reviews, this extensive review does not adhere to strict protocols for study selection and data extraction; instead, we aim to provide a comprehensive synthesis of available evidence to inform future research and practice in the health communication model. In the subsequent sections of this paper, we will provide an overview of the methodologies used to explore and analyze essential factors in health communication, including detailed descriptions of our data selection, screening processes, and analysis techniques such as topic modeling and trend analysis.

Methods

Search Strategy and Eligibility Criteria

To answer our research question and meet our study objectives, we conducted an extensive literature review using the web-based software tool Rayyan for screening, which enabled us to efficiently filter through a large amount of research [22]. The review and reporting process followed the guidelines provided by Rowley and Slack [23,24], ensuring a rigorous approach to organizing the collected literature and presenting the findings in a structured and coherent manner (Multimedia Appendix 1).

Multiple electronic databases including PubMed, Web of Science, Scopus, PsycINFO, and IEEE Xplore were searched to ensure broad coverage of the literature. Each database was chosen for its specific strengths and relevance to various dimensions of health communication. PubMed was selected for its comprehensive coverage of biomedical literature, essential for addressing medical aspects of health communication. Web of Science and Scopus were chosen for their multidisciplinary scope, enabling the examination of trends across diverse fields. PsycINFO was included for its focus on psychological literature,

XSL•FO
RenderX

key for exploring aspects of patient empowerment and decision-making processes. Finally, IEEE Xplore was included to access cutting-edge research on technological innovations in communication tools, critical for understanding the development and application of digital solutions in health care.

While digital tools and patient involvement are reshaping health communication, we carefully selected keywords representing 4 different categories related to health communication, including patient participation (MeSH [Medical Subject Headings] term), decision-making, shared (MeSH term), technology (MeSH term), and physician-patient relations (MeSH term) to guide our literature selection. These keywords were used along with Boolean operators (AND or OR) to capture diverse studies that allow us to examine the complex dynamics between the interactions of patients and physicians. To ensure a robust and precise alignment with our research objectives, we thoroughly reviewed our keywords and relevant key terms using the MeSH database and implemented a multidatabase search strategy [25]. This approach allowed us to craft a query that effectively targets the most relevant literature across various platforms, enhancing the breadth and depth of our literature review. Each query was adjusted slightly depending on the database, covering the period from 1999 to 2023: *((communication) OR (health communication model)) AND ((decision making*) OR (decision making, shared)) AND ((patient participation*) OR (patient participation/methods) OR (patient empowerment) OR (patient engage*) OR (patient involve*) OR (Patient-Centered Care)) AND ((physician-patient relations*) OR (Trust) OR (physician-patient trusted relationship)) AND ((decision support techniques) OR (technology) OR (health informatics) OR (Patient Preference) OR (information systems*) OR (health information systems))*

The selection process for the literature review was carefully designed to comprehensively cover studies on patient-physician interactions within the digital health framework. We included studies that (1) examined these interactions specifically with health care technologies; (2) aligned with the principles of shared decision-making or patient-centered care; (3) focused on cancer or chronic diseases, given their significant implications for patient communication; and (4) were published within the designated time frame of 1999-2023 to capture the evolution of health communication practices in the past 25 years. To maintain the quality and relevance of the studies reviewed, exclusions were made based on (1) language, excluding non-English articles to ensure consistency in data interpretation; (2) publication type, omitting reports, conference papers, abstracts, letters, or feature articles, which lack peer-reviewed validation; and (3) the absence of significant patient participation or empowerment, a key focus in evaluating effective communication strategies. This structured evaluation ensures our review meets high scholarly standards, offering a solid foundation for understanding the changing dynamics of health communication with emerging health care technologies.

Data Analysis

Topic Modeling and Visualization

All our analysis processes for this research were applied in R (R Foundation). Initially, we used the "tm" package to preprocess the text, which involved transforming all characters to lowercase, removing punctuation, numbers, and custom stopwords to ensure a clean and relevant data set. Following preprocessing, we constructed a Document-Term Matrix from the cleaned text to remove sparse terms and enhance both computational efficiency and model accuracy. To assess the coherence and effectiveness of the model, we utilized the "ldatuning" package and determined 3 optimal numbers of topics based on coherence scores, each representing a distinct thematic cluster within our data. Subsequently, we applied latent Dirichlet allocation using the "topicmodels" package to identify the most frequently discussed terms and concepts within the abstracts of the collected studies.

To aid in the interpretation of these topics, we visualized the topics using the "LDAvis" package. This advanced visualization tool allowed us to create interactive 2D scatterplots and bar plots of the topics. By adjusting the lambda parameter from 0 to 1, we were able to explore both the unique and frequently occurring terms within each topic, enhancing our understanding of the thematic relevance and distinction.

Trend Analysis and Secondary Topic Exploration

In our analysis framework, we distinguished between primary and secondary topics to thoroughly map the landscape of health communication. The primary topics, which center on health communication between patients and physicians, were categorized by topic modeling, serving as the overarching theme of our study. Accordingly, each paper was assigned a topic number by identifying the dominant topic based on the highest probability derived from the latent Dirichlet allocation model. These topics will be utilized in our trend analysis like topic distribution over time to examine the evolving dynamics within health communication, providing insights into how these themes have developed in the past 25 years.

To effectively represent the broad themes identified through topic modeling, we have selected specific secondary terms that closely relate to these primary topics. These terms are chosen for their frequent appearance within the context of identified primary topics and their ability to precisely capture the core ideas of each topic. Once identified, these terms are used as critical filters to refine our data set individually, ensuring that only those papers most relevant to each secondary topic were included. This step is critical as it allows us to apply the themes for detailed trend analysis and in-depth synthesis, particularly suitable for methodologies such as UpSet plotting, where the identification of overlaps and intersections between sets is necessary.

Impact Synthesis

To achieve our third objective, synthesizing the impact of health IT on patient-physician interactions and its subsequent effects on the evolution of health communication models, our final analytical phase involved a rigorous review process focusing on studies that exemplify the practical integration of health IT within dynamic health communication frameworks. These frameworks are designed to actively involve both patients and physicians, ensuring that technology enhances their communication. From the literature used for data analysis, we

XSL•FO
RenderX

further narrowed the focus to a specific analysis centered on health IT and its application in health communication. We strictly excluded studies that (1) did not incorporate health IT as an integral component of patient-physician communication, (2) solely evaluated technology design without practical health care application, or (3) lacked a structured communication model. This exclusion criteria allowed us to concentrate on research that provided clear insights into effective health IT integration.

This comprehensive and structured methodology ensures that our analysis not only adheres to rigorous academic standards but also provides insightful conclusions that enhance our understanding of the dynamic field of health communication in the digital age.

Results

Outcomes

In this section, we first present the outcomes of our data selection and screening process. Through query adjustment and refinement, removal of duplicates, and application of strict selection criteria, we narrowed our focus to 389 papers. These articles were then utilized in topic modeling, providing a foundation for further analysis.

In our topic modeling analysis, 3 major topics were identified, each closely related to health communication. This structured framework allowed for deeper exploration of the data and set the stage for subsequent analyses. Building upon the results of topic modeling, we conducted trend analysis to explore the dynamics and associations between significant elements such as health IT, patient-physician relationships, trust, and patient empowerment. The distribution of these topics over time and the interrelations among these key elements were visualized using topic distribution graphs and UpSet figures, offering a clear depiction of evolving trends.

Finally, our study synthesized findings on the role of health IT in patient-physician interactions, highlighting this as the key aspect of our third objective. This phase highlighted how technological advancements have been integral in shaping modern health communication models and identified areas for further development to facilitate enhanced interactions between patients and health care providers.

Data Selection and Screening

The review process initially identified and collected 3391 articles through multiple databases, as illustrated in Figure 1. Following the initial search, the titles and abstracts of these articles were screened using Rayyan. This initial screening narrows down the data to 2406 articles by filtering out those that clearly do not meet our preliminary inclusion criteria and 156 duplicated articles, ensuring the uniqueness of each study included in the analysis. The remaining articles were subjected to a rigorous screening process based on specific exclusion criteria, including language (English only), publication period (1999-2023), type of publication (journal articles only), and content relevance (articles focusing on communication involving technology and patients, specifically within the contexts of cancer or chronic diseases). This process effectively narrowed our focus to 389 articles that were highly relevant to the core topics of health communication, enabling us to explore the dynamics and associations between significant elements.

Figure 1. The diagram illustrates the selection process (identification, screening, eligibility assessment and inclusion) of this review study.

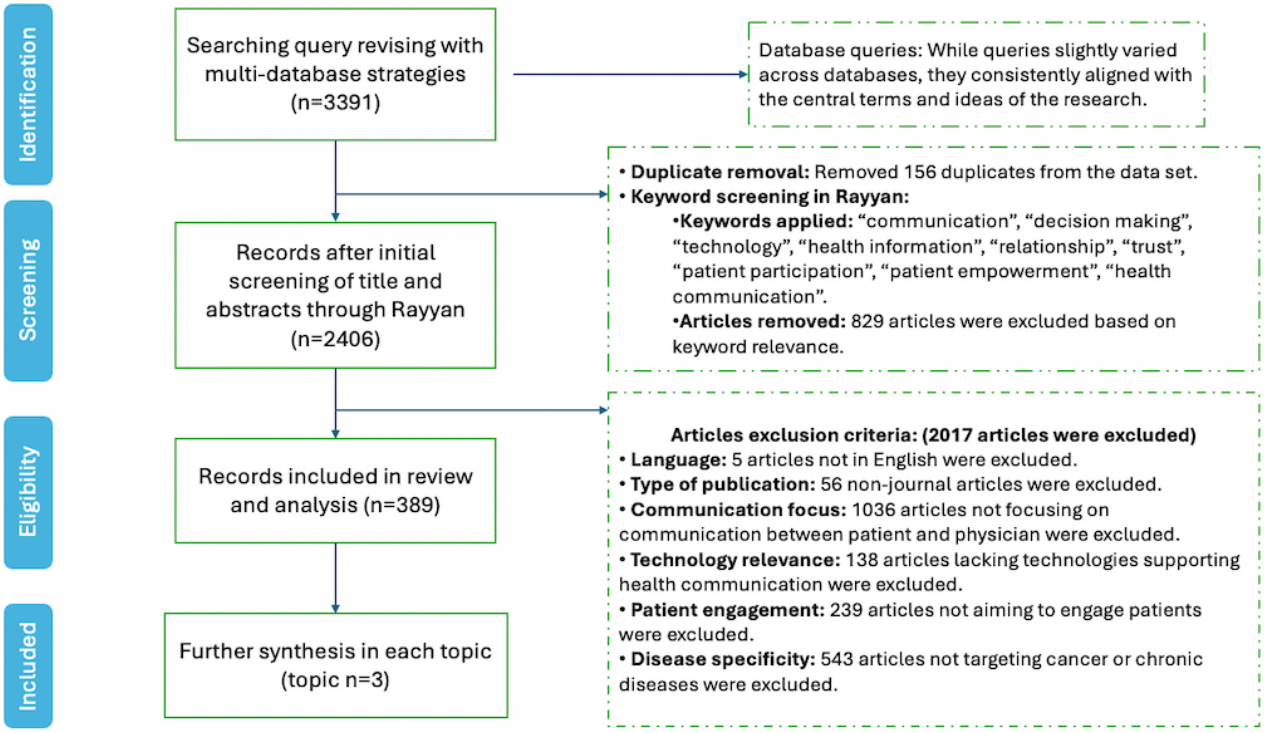

XSL•FO
RenderX

Topic Modeling Analysis

To address objective 1, identifying and analyzing key topics that influence health communication, we conducted a topic modeling analysis on the selected 389 articles. This analysis successfully identified and characterized 3 main topics, which explain the underlying structure of health communication and reveal distinct themes that enhance patient-physician communication in the digital age.

For the visualization of our topic modeling results, the scatterplots in Figure 2 explain how principal component 1 and principal component 2 capture the directions in which the variance in the topic distribution is most pronounced. As the topics, represented by the circles in the scatterplot, are far apart from each other, they are more likely to have a unique set of terms in each topic. To gain a deeper insight into the underlying structure, a fine-tuning option with lambda was used to adjust the relevance and frequency of terms in each topic. Adjusting the lambda value from 0 to 1 shifted the relevance of terms between their frequency in the data set and their uniqueness within the topic (ie, with a low lambda value the topics appeared somewhat similar, whereas higher values provided a clearer distinction between the topics). Further details on topics 2 and 3 are provided in Multimedia Appendix 2.

With distinct themes central to health communication, we carefully named and described individual topics by their unique set of keywords. These topics and their descriptions are summarized in Table 1.

Figure 2. A 2D scatterplot and a bar plot illustrating the distribution and relevance of topics identified from the topics modeling analysis using LDAvis. Each circle in the scatterplot represents a topic, while the bar plot displays the most relevant terms for topic 1 as an example here.

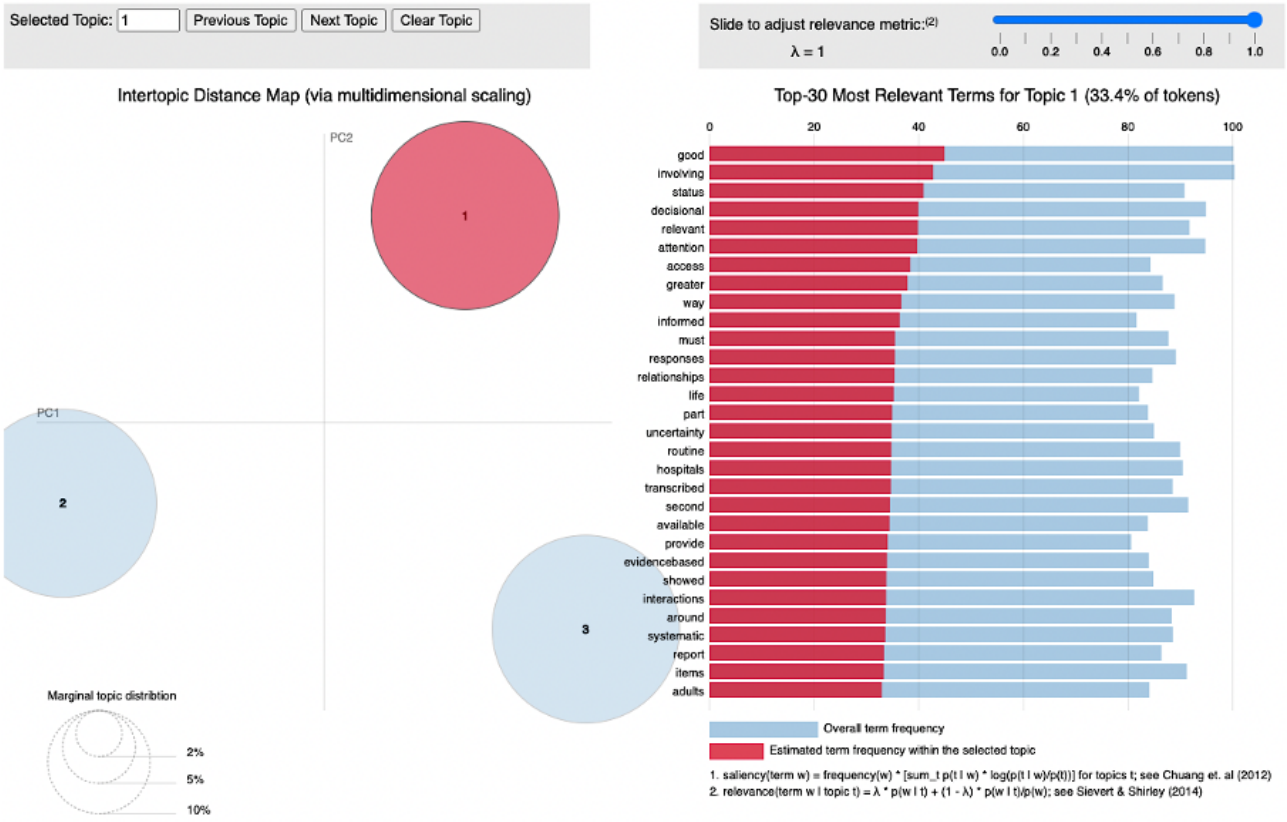

Table 1. The 3 topics categorized from topic modeling results and selective top keywords.

Topic	Label	Keywords
Topic 1	Patient-Physician Relationship and Shared Decision-Making (n=119)	“good,” “status,” “decisional,” “access,” “attention,” “greater,” “informed,” “must,” and “relationships”
Topic 2	Patient Empowerment and Education Strategies (n=142)	“type,” “engage,” “evaluated,” “measured,” “following,” “topics,” “electronic,” “educational,” “older,” and “diagnosis”
Topic 3	Health Care Systems and Health IT Implementations (n=128)	“association,” “reduce,” “clinician,” “national,” “provided,” “models,” “digital,” “age,” and “interactions”

Topic 1, labeled *Patient-Physician Relationship and Shared Decision-Making*, describes the quality and process of health care delivery focusing on informed and shared decision-making and patient-physician relationships. In health communication, this topic can indicate how patients’ access to information and the quality of the patient-physician relationship contribute to effective communication and shared decision-making.

Topic 2, labeled *Patient Empowerment and Education Strategies*, focuses on educational strategies and evaluation methods in health care, emphasizing the utilization of health IT

XSL•FO
RenderX

and targeting aging populations. In health communication, this topic can explore how empowering patients through education and supporting them with digital tools improves communication and healthcare outcomes.

Topic 3, labeled *Health Care Systems and Health IT Implementations*, is specifically related to health care systems and technological implementations. In health communication, it can explore how health care policies and technological implementations facilitate effective communication and enhance models by guiding interactions between patients and physicians.

The topic modeling results have strengthened and consolidated the categories of our designed searching queries, combining "patient-physician relationship" and "shared decision-making" into a single focused topic. Additionally, it has emphasized the importance of educational strategies for patient empowerment and the critical role of health IT implementation in health care systems.

Based on these findings, we carefully selected our secondary terms "patient-physician relationships," "patient empowerment," and "health IT" to reflect the core themes identified in the primary topics. The patient-physician relationship was explored using terms such as "patient-physician relationship," "relationship," "shared decision-making," and "patient-centered care." In the case of patient empowerment, we used keywords such as "empowerment," "participation," "engagement," "involvement," and "self-management." Health IT was represented by terms such as "technology," "online platform," "EHR," "telehealth," "mHealth," "digital," "informatics," "patient portal," "wearable device," "telemedicine," "remote," and "information system."

A further manual review was conducted by the first author and closely monitored by the project lead to ensure consistency and accuracy in the analytical approach. Beyond the identified topics, trust emerged as a significant theme within health communication and a frequently discussed topic across various topics analyzed in the study. Topic 1 related to patient-physician relationship and shared decision-making has 42 trust-related articles out of 119 (35.3%), topic 2 on patient empowerment and educational strategies has 45 out of 142 (31.7%), and topic 3 concerning health care systems and health IT implementations has 40 out of 128 (31.3%). These findings highlighted the need to include trust as an additional secondary element in health communication and further explore it in the rest of the analysis to investigate its association with health communication and the interplay among other key factors. Therefore, we utilized terms such as "trust," "physician-patient trust," "medical trust," "health care trust," and "trust in health information" to represent trust in our following analysis. An additional observation from this review indicates that the topic of trust can include a broad range, extending beyond patient-physician trust to potentially include trust in health care systems, digital tools, and health information.

Trend Analysis

To address objective 2, exploring the dynamics and associations among key topics and factors identified through topic modeling, our trend analysis was structured into 2 main sections. Initially, we used a bar plot to visually represent the frequency and prominence of each primary topic from 1999 to 2023. Following this, we used an UpSet plot to further explore the contributions and intersections of secondary terms that were carefully selected to represent primary topics over time.

The first section of the trend analysis focused on the primary topics that we found through topic modeling within health communication. By analyzing the topic distribution and assigning a unique topic number ranging from 1 to 3, we tracked the frequency of each topic over the years (1999-2023) with a bar plot (Figure 3).

Figure 3. Bar plot illustrating the annual distribution of papers categorized by their relevance to 3 identified topics from topic modeling with different colors. The bar plot provides a visualization of how the distribution and frequency of each topic within health communication have evolved from 1999 to 2023.

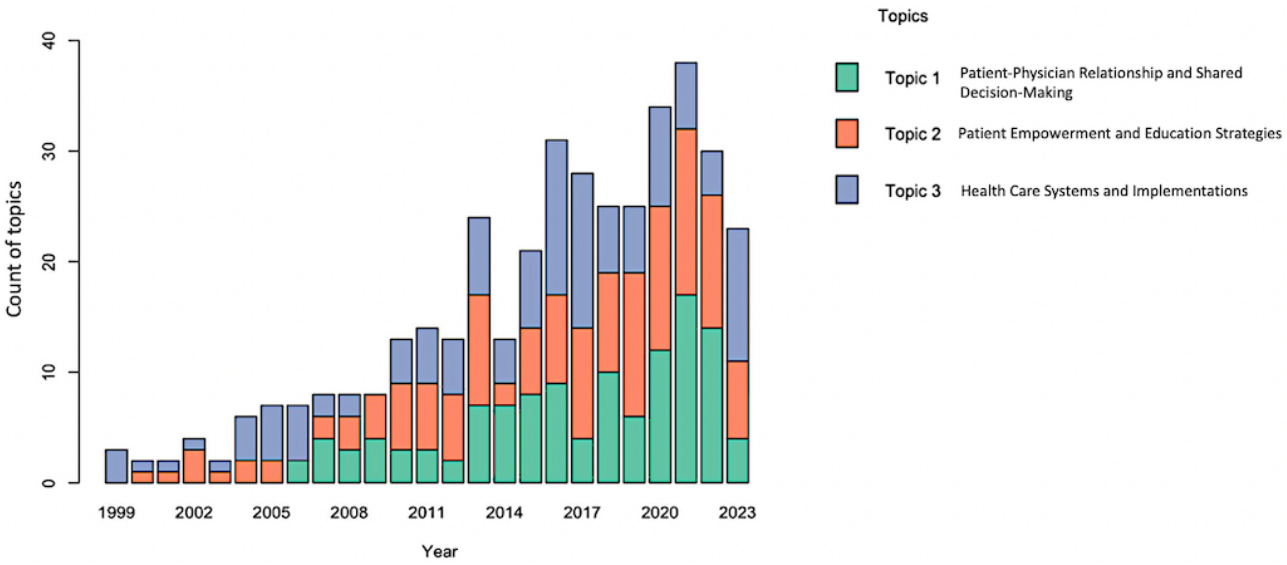

Topic 1 (green), Patient-Physician Relationship and Shared Decision-Making, did not emerge in the literature until 2006.

XSL•FO
RenderX

It shows a notable presence across the years with some fluctuations, specifically with a significant peak in 2021 (frequency 17, mean 5.8).

Topic 2 (orange), Patient Empowerment and Education Strategies, reveals a steady increase in frequency over the years, emphasizing a consistent rise in research interest and practical applications (mean 4.92).

Topic 3 (blue), Health Care Systems and Health IT Implementations, exhibits an upward trend, with notable peaks in 2016 (frequency 14) and 2023 (frequency 12), compared with a mean frequency of 5.16.

In the second section of our trend analysis, we utilized an UpSet plot (Figure 4) to explore the associations among the secondary terms within health communication. This analysis specifically focuses on patient empowerment, patient-physician relationships, health IT, and trust in the field of health communication.

Figure 4. UpSet figure showing the key factors of health communication, including patient empowerment, patient-physician relationship, health IT, and trust, which are represented in 4 different sets individually. The bar chart on the left shows the size of each set, which corresponds to the number of articles that contains any factors among our key factors. The main bar chart on the top represents the size of intersections between different sets within health communication. Each bar's height indicates how many elements are common to the individual sets connected below by dots and lines in the matrix. The matrix of intersections at the bottom has rows labeled with the names of the 5 sets and columns represented by dots under each set name. A filled dot indicates that the set involves the intersection corresponding to that column.

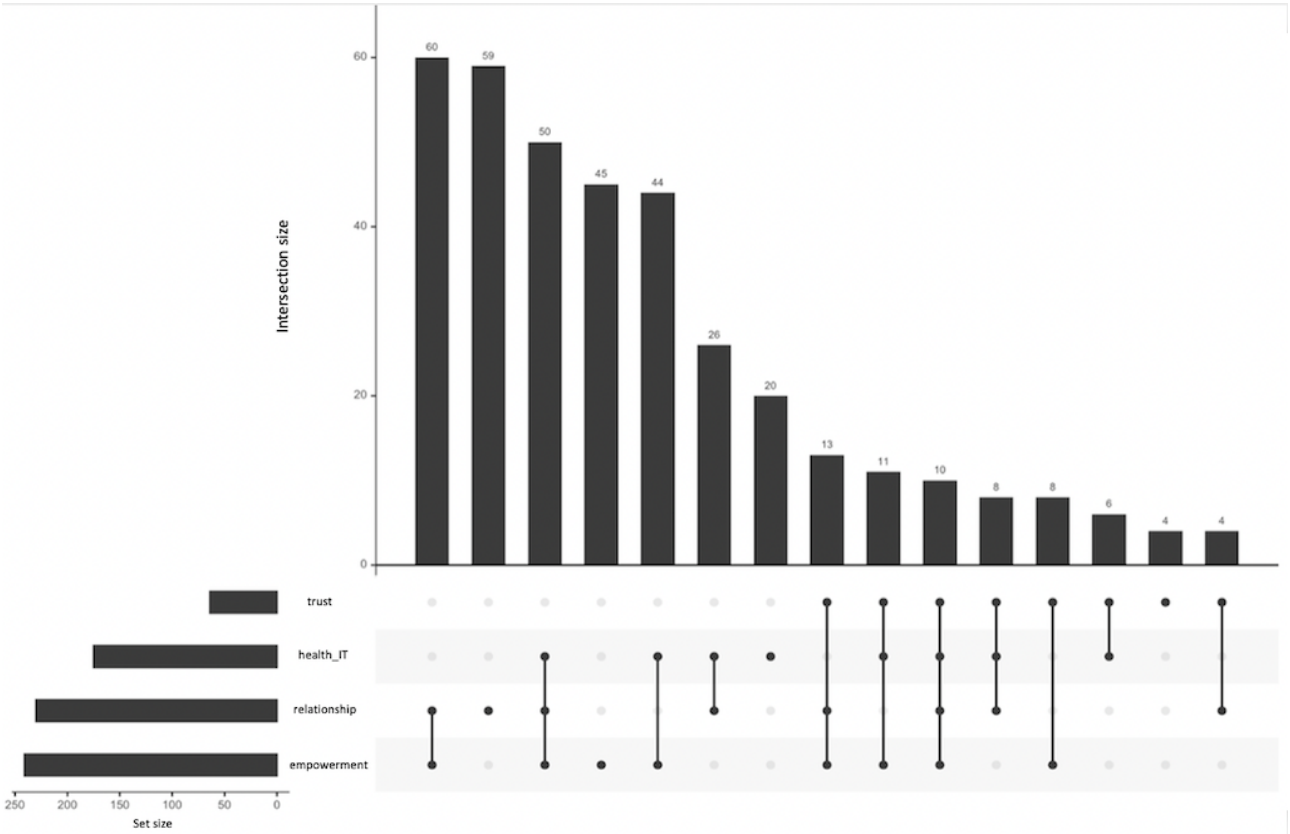

In our UpSet figure analysis within the scope of health communication, “patient empowerment” emerges as the most prominent set, as indicated by the largest bar on the left side of the plot. It is closely followed by “patient-physician relationship” and “health IT,” both highlighting their critical roles in the dynamics of health communication. The topic of “trust” shows a slightly lesser focus. Intersections between these sets reveal notable overlaps, with the largest interaction between “patient empowerment” and “patient-physician relationship” under health communication. This intersection reflects key themes of patient-centered care that involve communication strategies, empowerment, and interpersonal relationships. Reinforced by its presence in the second largest intersection with health communication, the “patient-physician relationship” further emphasized its critical role. This continual involvement across different sets highlights that the patient-physician relationship is central to discussions in health communication, further emphasizing the critical need for strong interpersonal foundations between patients and physicians in the patient-centered care framework. Similarly, the interaction involving “health IT,” “patient-physician relationship,” and “patient empowerment” under health communication illustrates an increasing integration of technology in empowering patients to facilitate effective health communication. Finally, although “trust” is addressed in a relatively smaller number of articles, its presence in complex interactions with all other factors highlights its foundational importance and influence, suggesting potential for deeper exploration to uncover its broader impacts on health communication.

Synthesis of Health IT’s Impact on Patient-Physician Communication Models

To address our third objective, synthesizing the impact of health IT on patient-physician interactions, we conducted a detailed

XSL•FO
RenderX

review based on our secondary research terms of health IT. This process led to the selection of 29 pivotal papers for detailed analysis, the findings of which are thoroughly detailed in Table 2.

These papers collectively provide a synthesized overview highlighting the specific health communication models or frameworks employed, the variety of health IT tools utilized, their impacts on patient empowerment and engagement, and the challenges or barriers encountered. The analysis revealed several key themes to not only map out the direct effects of health IT in health communication but also indicate practical applications and recommendations aimed at overcoming the identified challenges and amplifying the effectiveness of health communication strategies, which are summarized in Textbox 1.

XSL•FO
RenderX

Table 2. Impacts of health information technology on patient-physician communication models.

Study	Health communication model/framework	Health information technology tools used	Impact on patient empowerment/engagement	Challenges/barriers	Better practices/recommendations
Heesen et al [26]	SDM[a] guided	Evidence-based decision aids and educational programs	Enhances patient participation in treatment decisions	Balancing patient autonomy with a lack of initial knowledge	Continue development of comprehensive patient education programs
Sillence et al [27]	Informed SDM guided	Internet information	Empowers patients by enhancing informed decision-making and improved communication with physicians	Poorly designed websites can lead to rejection of quality content	Well-designed personalized websites with credibility
DuBenske et al [28]	SDM guided	Interactive health communication system	Provides information, communication, coaching resources, and a symptom-tracking system; bridges communication gaps among patients, families, and physicians	User diversity in format preferences and technology comfort, physician buy-in, and integration with existing medical systems	Conduct needs assessment, involve stakeholders, accommodate diverse technology preferences, ensure clinician training and buy-in, and integrate with existing systems
Jimbo et al [29]	SDM guided	Decision aid	Incorporates interactive personal risk assessment and preference clarification tools; facilitates informed decision-making and screening adherence; improves decision quality	Balancing the complexity of decision aids with user-friendly designs	Ensure decision aids are intuitive and support real-time interaction to enhance decision-making
Nazi [30]	SDM guided	PHRs[b]: secure messaging and PHR portal	Improved access to health information; enhances communication, patient self-reporting, and provider-patient relationships	Integration with clinical workflow, training, and support for health care professionals	Ensure PHR systems align with clinical practices, provide adequate training, and actively engage clinicians in PHR use
Wilkes et al [31]	SDM guided	EHRs[c] and decision support tools	Facilitates informed and value-concordant or personalized decisions; enhances the quality of prostate cancer screening decisions	Complexity of discussions and emotional and informational barriers	Utilize EHRs for SDM support, integrate decision aids, and adjust reimbursement to encourage SDM
Smits et al [32]	Not specific; SDM guided	Web-based information tool	Enhances patient and family engagement by providing tailored information; a comprehensive approach to developing patient-centric informational tools	Balancing diverse patient preferences and literacy levels	Follow the DoTTI[d] framework for iterative development and stakeholder engagement throughout the process
Roter 2015 [33]	SDM guided	EMR[e] and secure patient portals	Facilitates a new model of patient-physician collaboration with high transparency	Distraction of physicians by EMR use during consultations	Implement training using a practices inventory and provide patients with full access to their EMR
Lafata et al [34]	PCC[f]	EHRs, HRA[g] instruments, and patient reminder lists	Facilitates patient involvement and personalized care; EHRs for information management, HRAs for health risk assessment, and reminder lists for patient preparation	EHR use may extend visit duration and impact preventive service delivery and variability in tool effectiveness	Monitor tool impacts rigorously and tailor tool use to individual patient and practice needs to optimize outcomes

XSL•FO
RenderX

Study	Health communication model/framework	Health information technology tools used	Impact on patient empowerment/engagement	Challenges/barriers	Better practices/recommendations
Bruinessen et al [35]	SDM guided	Web-based communication tool (PatientTIME)	Enhances patient self-efficacy in clinical communication	Ensuring the intervention effectively translates into measurable confidence improvements	Continuously refine and evaluate the intervention based on patient feedback and trial results and focus on integrating practical communication strategies
Olomu et al [36]	SDM guided	Decision aids	Increases medication adherence, enhances patient satisfaction with communication and confidence in health care decisions	Ensuring consistent implementation and adaptation in varied clinical settings	Further develop and test the program to evaluate its broader applicability, effectiveness, and cost-efficiency
Lussier et al [37]	PCC	e-Learning platforms	Enhances patient engagement in communication	Ensuring patient access and familiarity with web-based tools	Integrate web-based educational tools into routine practice and encourage patients to participate as part of their care
Hakone et al [38]	SDM guided	Personalized health risk communication tool	Enhances patients' understanding of personalized risk and treatment options; improves patient decision-making capability	Balancing complexity and clarity in visual risk communication and emotional distress impact	Carefully design visual tools and narratives to meet patient needs and consider emotional states and numeracy levels
Renzi et al [39]	PCC	eHealth and self-management platform	Enhances self-management and patient empowerment in prostate cancer care with a web-based platform for patient education and management	Communication gaps postdiagnosis and postsurgery, and lack of support in self-management	Strengthen health care provider-patient communication, incorporate social support into patient care, and personalize care approaches from diagnosis through treatment
Lu and Zhang [40]	Not specific; SDM guided	Online health communities	Positive impact of physician-patient communication on compliance mediated by perceived information quality, decision-making preference, and concordance	Quality control of online health information and the gap between perceived and actual quality of information	Strengthen the management of online health information, foster open discussions and emotional connections within OHCs[h], and ensure high-quality health information and communication
Yu et al [41]	SDM guided	Online decision aid	Facilitates patient participation in decision-making, strengthens physician-patient relationships	Conflicting agendas between patients and providers and integration of new tools into routine practice	Utilize SDM and goal-setting tools to resolve conflicts and align treatment goals and ensure tools are flexible and respect professional identities
Kondylakis et al [42]	Data-driven communication	ICT[i] platforms	Enhances patient engagement and personalized cancer care management	Patient engagement, interoperability, knowledge management, and trust in technology	Increase trust and stakeholder engagement, ensure interoperability of ICT systems, and enhance knowledge management in ICT platforms

XSL•FO
RenderX

Study	Health communication model/framework	Health information technology tools used	Impact on patient empowerment/engagement	Challenges/barriers	Better practices/recommendations
Hazara et al [43]	PCC	Web-based portals	Enhances patient participation, empowerment, and self-management in their health care; web-based patient portal allowing access to EHRs and communication with health providers	Barriers to widespread portal use	Enhance portal features, learn from other portal implementations, and integrate portal use more effectively in health care delivery
Gilljam et al [44]	SDM guided; Shier's model of participation	Digital communication tools (Sisom for children)	Enhanced participation in health care; increased engagement and direct communication toward children using Sisom; children felt listened to and could express views	Lack of validation and possible variations in engagement not systematically measured	Use randomized control designs for further validation, focus on eHealth's effects on health outcomes and cost-effectiveness, and consider expanding the use of Sisom or similar tools in pediatric settings
van Velsen et al [45]	SDM guided; active and collaborative patient-physician communication	Online patient portal	The study explored the dimensions of trust that might influence patient engagement but found no direct effect on usage	Trust does not directly influence technology usage, indicating possible other factors affecting eHealth service adoption	Assess multidimensional trust factors separately to understand their influence on eHealth adoption; further research is needed to explore why trust does not directly translate to usage
Dimitri et al [46]	Not specific; SDM guided	eHealth platforms and artificial intelligence predictions	Enhanced patient and caregiver communication, improved access to educational resources, and better medical and psychological support; use of connected devices for monitoring and artificial intelligence for predicting treatment adherence	Implementation in specific clinical settings, ensuring data privacy and security, and integration into existing clinical workflows	Involve clinicians and patients in technology integration, provide robust evidence of value and utility, address data privacy concerns
Schubbe et al [47]	Not specific; SDM guided	EHRs, email, and patient portals	Improved patient involvement in decision-making, did not lengthen consultation times, and normalized in routine care	Barriers to routine integration due to clinical and organizational factors	Adapt conversation aids to fit clinical workflows, ensure flexibility in their use, and consider patient preferences and characteristics for tailored communication approaches
Kim et al [48]	PCC	Digital health communication intervention with tailored messages based on local cultures	Enhanced family communication of genetic results and decision-making	Navigating cultural and linguistic differences in genetic education across countries	Use of existing digital platforms to facilitate culturally sensitive communication and adaptation of tools rather than creating new ones
Lu and Zhang [49]	Not specific; focusing on the effectiveness of communication	eHealth and online health communities	Increased patient adherence via enhanced eHealth literacy and positive correlation between eHealth literacy and patient adherence through improved communication and information quality	Addressing gaps in the perceived and actual quality of information	Strengthen management of information quality, develop user-friendly features, and educate patients to improve eHealth literacy

XSL•FO
RenderX

Study	Health communication model/framework	Health information technology tools used	Impact on patient empowerment/engagement	Challenges/barriers	Better practices/recommendations
Schubbe et al [47]	SDM	Mobile health	Facilitates enhanced shared decision-making, reduces decisional conflict and regret, and improves communication quality	Addressing the specific needs of minority populations and integrating decision partners in the decision-making process	Use of community patient navigators to provide support during decisions, leverage mobile health for real-time data and engagement, and facilitate provider-patient-decision partner communication
Liu et al [50]	PCC	PAEHR[j] systems	Improves health self-efficacy and patient-centered communication by facilitating access to health information and communication	Lack of direct association between PAEHR portal use and health outcomes	Integrate PAEHR systems to provide PCC, enhance health self-efficacy, use portals to facilitate patient engagement in health care decisions, and ensure the accessibility of PAEHR portals to improve patient health outcomes
Yılmaz et al [51]	Enhanced patient participation communication	Health Communicator (eHealth tool)	Intended to fulfill unmet informational and psychosocial support needs; multilingual eHealth tool tailored to specific ethnic groups	Patients'/survivors' low health literacy, cultural taboos, lack of trust in health care, and perceived complexity of the tool	Ensure simplicity and ease of use in eHealth tools, involve both patients and professionals in the design process, and provide training and support for new technologies
Grynne et al [52]	Effective interpersonal communication	Digital information tool (Digi-Do)	Enhanced patient knowledge, competence, and motivation; iterative access to reliable health information, and tailored digital information tools; as well as improved preparedness before, during, and after radiation therapy	Unmet information needs, varying levels of health literacy, and reliance on non–health care professional sources	Tailor digital information tools to patient's health literacy level, provide iterative access to reliable information, and complement interpersonal communication
Orstad et al [53]	SDM	Decision aids	Enhanced information exchange needs and addressed patient/clinician relationships and decision-making process	Continuity of patient-clinician relationships, barriers to information exchange, uncertainty in decision-making, and autonomy negotiation	Comprehensive decision support interventions, communication training on scientific/existential uncertainty, and organizational changes for better continuity and dialogue time

[a]SDM: shared decision-making.

[b]PHR: personal health record.

[c]EHR: electronic health record.

[d]DoTTI: Design and Development, Testing Early Iterations, Testing for Effectiveness, Integration, and Implementation.

[e]EMR: electronic medical record.

[f]PCC: patient-centered communication.

[g]HRA: health risk appraisal.

[h]OHC: online health community.

[i]ICT: information and communication technology

[j]PAEHR: patient-accessible electronic health record.

XSL•FO
RenderX

Textbox 1. Key themes identified in the analysis.

1. Communication models and health IT functions

The integration of health IT tools within various communication models has been identified, specifically on shared decision-making and patient-centered communication. Tools such as decision aids [26], interactive health communication systems [28], and patient portals [30] have been instrumental in increasing patient participation, bridging communication gaps, and improving access to health information [29]. Additionally, tools such as EHRs and health risk assessment instruments have helped in managing health information effectively [34], ensuring that care is tailored to individual patient needs.

2. Challenges and barriers

Studies have revealed significant challenges in the integration of health IT, such as technological integration issues [30,34], resistance from health care providers due to the perceived increase in their workload [26,29], lack of personalized help [31], and a wide range of patient technology *literacy* [28,29]. *These challenges demand focused attention on design and usability improvements to promote user engagement and acceptance of health IT solutions.*

3. Better practices and recommendations

The literature suggests ongoing training and education for both patients and health care providers as essential [26,30]. It also recommends designing health IT tools that are user-friendly and adaptable to various patient needs [28,29]. Moreover, integrating these tools into clinical workflows is vital to ensure they support rather than disrupt the patient-care provider relationship [34].

Future research should conduct a detailed examination of the specific challenges and opportunities highlighted in these studies to enhance patient engagement and address potential issues. Further efforts should focus on refining health IT tools to facilitate shared decision-making and patient-centered care, ensuring that the implementation of these technologies effectively meets the diverse needs of both patients and physicians, particularly in supporting their interactions. Such targeted research is fundamental for enhancing our understanding of how health IT can effectively assist in patient-physician interactions.

Discussion

Principal Findings

The rapid expansion of this topic has generated a substantial data set, enabling us to identify 3 key themes that are essential for integration into the health communication model. In this section, we further investigate how this extensive review enhances our understanding of the critical topics and factors that influence health communication between patients and physicians in the digital era. We also explain how the trends and interactions among these key elements such as health IT, patient empowerment, trust, and patient-physician relationships are reshaping health communication. Additionally, we further elaborate on the significant role and impact of health IT as derived from our analysis. Finally, we outline potential future research opportunities, especially with technology that could further advance this field.

Primary Topics in Health Communication Models

The evolution of health communication models reflects a clear transition from physician-centered to patient-centered approaches. Earlier models primarily focused on the unidirectional flow of information from physicians to patients [54]. However, with the increasing emphasis on patient empowerment, engagement, and shared decision-making, more recent models prioritize collaboration between patients and physicians, ensuring that medical decisions are aligned with patients' values and preferences [55]. Patient-centered care and shared decision-making models have been identified in previous studies that encourage active engagement of patients, emphasizing the importance of clear, transparent, and effective communication between patients and physicians [2]. However, the integration of essential elements in health communication models was not clearly elaborated, especially in the digital age where health IT should be effectively utilized in patient-physician communication.

Our study identified 3 primary topics that should be considered to include in a refined health communication model that fits today's digital age and more collaborative, patient-centered approaches. Specifically, topic 1 underscores the prominence of the Patient-Physician Relationship and Shared Decision-Making in health communication, closely aligning with the collaborative environment and further emphasizing the importance of the patient-physician relationship. Topic 2 captures the rise of Patient Empowerment and Education Strategies. As patients are increasingly empowered to actively engage with their health information, carefully designed education strategies can help effectively bridge the knowledge gap between patients and physicians and avoid patients being misinformed [56]. Topic 3 highlights the broader Health Care Systems and Health IT Implementations that support these communication models. Existing health IT tools such as patient portals, EHRs, and online information systems facilitate patient engagement and streamline communication across the care continuum. However, the integration of these tools into communication models requires careful consideration to ensure both general applications and personalized utilizations.

Addressing these elements can better align health communication models with the evolving needs of health care systems, ensuring that patient empowerment and engagement remain central. Future research should focus on standardizing these models with the integration of these primary topics to ensure they remain flexible and relevant in a rapidly changing health care landscape.

Trend Analysis and Factor Associations in Health Communication

In our trend analysis, we not only investigated the trend of identified primary topics in topic modeling, but also explored the complex interactions among carefully selected secondary topics, including patient empowerment, health IT, patient-physician relationship, and trust.

XSL•FO
RenderX

Topic 1 (Patient-Physician Relationship and Shared Decision-Making) combined 2 themes of patient-physician relationship and shared decision-making, emphasizing the critical role of strong interpersonal relationships in shared decision-making and collaborations in health care. Notably, a peak in this topic (Figure 3, green) observed in 2021 indicates an increased focus on fostering patient-physician relationships and facilitating collaborative decision-making. This shift was likely influenced by the global health challenges and the push for more patient-centered care models [55]. The critical need for effective communication between patients and physicians emphasized the importance of involving patients more actively in their care decisions and building strong patient-physician relationships, specifically with the support of technology [57]. For instance, the expansion of remote care and digital health technologies has effectively addressed barriers such as distance, waiting times, and disparities in access to care [58], further facilitating this trend toward improved patient-physician interactions.

For topic 2 (Patient Empowerment and Education Strategies), an early and consistent upward trend has been observed in Figure 3 (orange), indicating a shift toward empowering patients with knowledge and tools to manage their health effectively. This trend aligns with broader health care movements that emphasize increased patient education and empowerment. Research has demonstrated that well-implemented education and empowerment strategies can significantly reduce hospital readmission rates and improve adherence to treatment plans [56]. Moreover, higher levels of patient activation are associated with lower health care costs and better outcomes [59]. However, the journey to effectively empower patients continues to evolve. Supporting this evolution, effective health communication between patients and physicians plays an essential role in the empowerment process [60]. Enhanced communication fundamentally supports patient empowerment by providing clearer, more accessible information, which is essential for informed decision-making. Therefore, it is crucial to develop and carefully integrate patient empowerment strategies within health communication frameworks. This integration ensures that communications not only inform but also actively engage patients, thereby enhancing their role in decision-making processes and overall care management to reach better health care outcomes.

For topic 3 (Health Care Systems and Health IT Implementations), the significant peaks observed in 2016 and 2023 correspond to critical periods of health care system reforms (Figure 3, blue), particularly those involving the integration of new technologies. The enhanced use of health IT such as EHRs and web-based communication tools facilitated improved patient involvement and personalized care [34,35,37]. The COVID-19 pandemic further accelerated the implementation of advanced digital tools such as health communicator, which were designed to meet specific patient needs, including tailored information and psychosocial support for individuals [51,52]. This reflects a field that is continually adapting, not only to technological advancements but also to the evolving needs of patients. These developments illustrate the dynamic nature of health communication research and highlight critical moments when topics such as patient empowerment, shared decision-making, and technology integration rise to prominence.

Additional analyses involved refined secondary terms that helped us to investigate the detailed associations between specific terms that represented the core idea of the primary topics. This targeted selection process was crucial for accurately mapping the evolution of specific themes within health communication over the study period, enabling us to uncover nuanced trends and associations that might be unnoticed in a broader topic modeling analysis.

The top 2 strongest intersections we explored emphasize the critical role of the patient-physician relationship in health communication, particularly as patients take more control in the context of patient-centered care [61]. The integration of patient empowerment within the intersections highlights the necessity of bridging the knowledge gap for patients, enabling them to engage more actively in their health care decisions. Additionally, the third most prominent intersection, involving health IT, further underscores the growing role of technology in empowering patients and facilitating effective communication. Although the topic of trust appeared less frequently in association with other secondary terms, its close connections with each term were consistently identified. These individual connections emphasize the multifaceted role that trust plays in health communication, demanding further investigation. Specifically, trust acts as a foundational element in promoting better treatment adherence and fostering stronger patient-physician relationships [62]. Moreover, trust in health IT is equally vital, influencing patients' willingness to adopt and engage with digital health solutions, thereby enhancing the overall effectiveness of health communication [45,63].

The Role of Health IT in Health Communication Models

In our in-depth synthesis and analysis (shown in Table 2), many communication-related technologies are not under any specific communication model, but only guided by the idea of shared decision-making or facilitating patients' participation. Previous studies suggested that while shared decision-making is widely endorsed, many implementations lack a structured model that consistently leads to effective communication outcomes [64,65]. Similarly, patient-centered communication, despite its widespread advocacy as a concept that should lead to more satisfactory health care processes and outcomes, often remains theoretical rather than practical in its application [66]. Despite these challenges in the implementation of shared decision-making and patient-centered communication, the setting began to change significantly with technological advancements. By 2016, significant advancements in health care were driven by the adoption of digital solutions, addressing key shortcomings of traditional communication methods, such as inefficiency in data management, limited patient access, and fragmented communication [30,67]. Health IT, particularly through web-based interventions, began to enhance the structural foundation and accessibility of health care communications. Specifically, the integration of EHRs with patient portals has matured significantly, enhancing communication channels between patients and health care providers [34,68]. Additionally,

XSL•FO
RenderX

the implementation of web-based communication tools, decision aids, and online learning platforms has significantly advanced patient education, engagement, and effective communication with physicians [34-37]. These technologies provide essential continuity in patient care and enhance direct interactions between health care providers and patients, fundamentally altering the efficiency and effectiveness of health care delivery. By 2021, the COVID-19 pandemic had further catalyzed design improvements and validations for the application of health IT. The development of digital health, such as conversational aids, emerged as a notable advancement during this period. These tools, essential in a time of increased remote interactions due to social distancing measures, are designed to enhance patient engagement and support shared decision-making processes [46,47]. They aim to provide personalized, accessible information that patients can easily navigate and understand, filling a critical gap as in-person health care interactions become less frequent [45,48,51]. This pandemic emphasized the importance of reliable, accessible health communication technologies, driving innovations that could meet the new health care demands efficiently and effectively.

Building on the significant progress made in health IT, the evolution of health communication has been deeply intertwined with technological advancements across multiple dimensions. These advancements include facilitating interactions between patients and physicians, managing health care data, enhancing education, and improving overall health care delivery [68]. However, despite these advances, challenges and barriers persist in integrating human interactions with health IT systems. Notably, aspects such as the patient-physician relationship and trust building remain underexplored, despite being crucial for optimizing the effectiveness of health communication [39-42]. Addressing these gaps is essential to ensure that health IT supports not only the operational but also the interpersonal aspects of health care delivery, paving the way for more comprehensive and effective communication strategies.

Future Outlook: Artificial Intelligence in Health Communication

AI is transforming patient-physician communication by facilitating more personalized, efficient, and informed interactions through advanced data analysis and real-time information processing. AI is also increasingly being utilized to support or inform health care diagnoses and decisions [69,70]. Emerging AI tools, such as ChatGPT and other generative AI systems, analyze vast amounts of written material to generate coherent, contextually relevant text that closely resembles human writing styles. The use of AI before patient-physician interactions can significantly reshape the traditional dynamics of knowledge and power. Traditionally, physicians, with years of training and experience, have held professional knowledge, guiding the flow and outcomes of health care discussions. However, AI technologies—particularly those offering instant access to vast medical information—empower patients with a form of “pocket expertise.” This access enables patients to approach consultations with a preliminary understanding of potential diagnoses, treatments, and even novel therapies. Such information can help level the playing field, fostering a more collaborative relationship. However, it may also introduce tension or conflict between patients and health care professionals. While AI can rapidly process and synthesize information, it can also produce incorrect or misleading conclusions [71]. The technology’s interpretations are only as reliable as the data and algorithms driving them, which may lack the nuanced or contextual understanding of a health care professional.

This revolution in patient-physician communication through AI has the potential to transform interaction dynamics, offering both opportunities and challenges that warrant further exploration. Future studies should investigate AI’s potential to enhance personalized medical knowledge and its implications for health care decision-making, while carefully addressing the limitations of these technologies.

Limitations

The analysis of associations and trends in our study offers valuable insights into the dynamics of health communication influenced by health IT, patient empowerment, and patient-physician relationships, as well as the interactions among these key factors. However, it is important to emphasize that these associations should not be interpreted as direct causal relationships. Our analysis primarily identifies correlations that suggest potential interactions between these variables, providing a foundation for future hypotheses and more in-depth research tailored to specific situations and conditions. The identification of these patterns is essential for developing a more nuanced understanding of how various factors contribute to the effectiveness of health communication. This understanding can then inform subsequent experimental studies aimed at verifying causality and uncovering the underlying mechanisms at play.

Our review focused broadly on the general interactions between patients and physicians across diverse health care settings. As a result, specific demographic variables—such as different diseases, age groups, and racial backgrounds—were not extensively differentiated in our analysis. This broad approach aligns with our goal of identifying overarching trends and commonalities in health communication. However, it is acknowledged that health communication dynamics can vary significantly across different patient groups, shaped by unique needs and contexts related to specific health conditions, age-related preferences, and cultural or racial backgrounds. Future research could benefit from stratifying these variables to gain more detailed insights and develop communication strategies that are better tailored to the specific characteristics and needs of diverse patient populations.

Finally, this paper presents several important findings, such as the interactions between key terms identified in our trend analysis. However, the depth of synthesis regarding these intersections remains limited. Our synthesis primarily focuses on elucidating the role of health IT in health communication, highlighting its significant applications across diverse health care contexts. While the study offers a broad overview, outlining relevant terms and setting the stage for future research, it leaves room for more focused investigations. Subsequent studies are encouraged to investigate each finding in greater detail, further exploring the complex dynamics at play and refining our

XSL•FO
RenderX

understanding of these critical intersections in health communication.

Conclusions

This review extensively explored the key factors influencing health communication between patients and physicians in the digital age. Our analysis identified 3 main topics central to the field: the Patient-Physician Relationship and Shared Decision-Making, Patient Empowerment and Education Strategies, and Health Care Systems and Health IT Implementations. These topics serve as foundational elements for refining health communication models that foster effective interactions between patients and physicians. Additionally, this review highlighted significant interactions between health IT and key aspects of health communication, particularly patient-physician relationships, patient empowerment, and trust. The COVID-19 pandemic accelerated the adoption of digital health technologies such as telehealth, patient portals, and decision aids, all of which have demonstrated great potential in enhancing these interactions. Finally, the synthesis of health IT's role in health communication reveals its transformative impact on patient empowerment and engagement, while also highlighting significant challenges related to integration, usability, and user acceptance. These findings emphasize the need for continuous refinement of health IT tools to better support patient-centered care and informed decision-making within health communication. Future research should focus on developing models that integrate health IT and foster trust-based relationships between patients and physicians. Investigating how emerging technologies, such as AI, can enhance these models will also be crucial. Such efforts are essential for realizing the full potential of health IT in improving health care outcomes through effective communication between patients and physicians.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of searching results from 1999 to 2023.
[DOCX File , 17 KB - jmir_v27i1e60512_app1.docx]

Multimedia Appendix 2
Intertopic distance map and top 30 relevant terms for topics 2 and 3.
[DOCX File , 755 KB - jmir_v27i1e60512_app2.docx]

References

1. Bailo L, Guiddi P, Vergani L, Marton G, Pravettoni G. The patient perspective: investigating patient empowerment enablers and barriers within the oncological care process. Ecancermedicalscience 2019;13:912 [FREE Full text] [doi: 10.3332/ecancer.2019.912] [Medline: 31123495]
2. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging patients in decision-making and behavior change to promote prevention. Stud Health Technol Inform 2017;240:284-302 [FREE Full text] [Medline: 28972524]
3. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008 Feb;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
4. Jiang F, Liu Y, Hu J, Chen X. Understanding health empowerment from the perspective of information processing: questionnaire study. J Med Internet Res 2022 Jan 11;24(1):e27178 [FREE Full text] [doi: 10.2196/27178] [Medline: 35014957]
5. Chen J, Mullins CD, Novak P, Thomas SB. Personalized strategies to activate and empower patients in health care and reduce health disparities. Health Educ Behav 2016 Feb;43(1):25-34 [FREE Full text] [doi: 10.1177/1090198115579415] [Medline: 25845376]
6. Marinello D, Di Cianni F, Del Bianco A, Mattioli I, Sota J, Cantarini L, et al. Empowering patients in the therapeutic decision-making process: a glance into Behçet's syndrome. Front Med (Lausanne) 2021;8:769870 [FREE Full text] [doi: 10.3389/fmed.2021.769870] [Medline: 34966756]
7. Kuwabara A, Su S, Krauss J. Utilizing digital health technologies for patient education in lifestyle medicine. Am J Lifestyle Med 2020 Dec 13;14(2):137-142 [FREE Full text] [doi: 10.1177/1559827619892547] [Medline: 32231478]
8. Dy M, Olazo K, Lyles CR, Lisker S, Weinberg J, Lee C, et al. Usability and acceptability of virtual reality for chronic pain management among diverse patients in a safety-net setting: a qualitative analysis. JAMIA Open 2023 Oct;6(3):ooad050 [FREE Full text] [doi: 10.1093/jamiaopen/ooad050] [Medline: 37449058]
9. El-Sherif DM, Abouzid M, Elzarif MT, Ahmed AA, Albakri A, Alshehri MM. Telehealth and artificial intelligence insights into healthcare during the COVID-19 pandemic. Healthcare (Basel) 2022 Feb 18;10(2):385 [FREE Full text] [doi: 10.3390/healthcare10020385] [Medline: 35206998]

XSL•FO
RenderX

10. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
11. Seuren LM, Ilomäki S, Dalmaijer E, Shaw SE, Stommel WJP. Communication in telehealth: a state-of-the-art literature review of conversation-analytic research. Res Lang Soc Interact 2024 Apr 03;57(1):73-90 [FREE Full text] [doi: 10.1080/08351813.2024.2305045] [Medline: 38741749]
12. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
13. Office-based physician electronic health record adoption. HealthIT. URL: https://www.healthit.gov/data/quickstats/office-based-physician-electronic-health-record-adoption [accessed 2024-02-12]
14. Petrovskaya O, Karpman A, Schilling J, Singh S, Wegren L, Caine V, et al. Patient and health care provider perspectives on patient access to test results via web portals: scoping review. J Med Internet Res 2023 Oct 19;25:e43765 [FREE Full text] [doi: 10.2196/43765] [Medline: 37856174]
15. Butcher CJ, Hussain W. Digital healthcare: the future. Future Healthc J 2022 Jul;9(2):113-117 [FREE Full text] [doi: 10.7861/fhj.2022-0046] [Medline: 35928188]
16. Cohen RA, Adams PF. Use of the internet for health information: United States, 2009. NCHS Data Brief 2011 Jul(66):1-8 [FREE Full text] [Medline: 22142942]
17. Bujnowska-Fedak M, Waligóra J, Mastalerz-Migas A. The Internet as a source of health information and services. Adv Exp Med Biol 2019;1211:1-16. [doi: 10.1007/5584_2019_396] [Medline: 31273574]
18. Langford AT, Roberts T, Gupta J, Orellana KT, Loeb S. Impact of the internet on patient-physician communication. Eur Urol Focus 2020 May 15;6(3):440-444. [doi: 10.1016/j.euf.2019.09.012] [Medline: 31582312]
19. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022 Jun 27;2022(22):1 [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
20. Rajamani G, Rodriguez Espinosa P, Rosas LG. Intersection of health informatics tools and community engagement in health-related research to reduce health inequities: scoping review. J Particip Med 2021 Nov 19;13(3):e30062 [FREE Full text] [doi: 10.2196/30062] [Medline: 34797214]
21. Hosseini M, Jahanshahlou F, Akbarzadeh MA, Zarei M, Vaez-Gharamaleki Y. Formulating research questions for evidence-based studies. Journal of Medicine, Surgery, and Public Health 2024 Apr;2:100046. [doi: 10.1016/j.glmedi.2023.100046]
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
23. Rowley J, Slack F. Conducting a literature review. Management Research News 2004 Jun;27(6):31-39 [FREE Full text] [doi: 10.1108/01409170410784185]
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
25. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc 2018 Oct;106(4):531-541 [FREE Full text] [doi: 10.5195/jmla.2018.283] [Medline: 30271302]
26. Heesen C, Köpke S, Richter T, Kasper J. Shared decision making and self-management in multiple sclerosis--a consequence of evidence. J Neurol 2007 May;254 Suppl 2:II116-II121. [doi: 10.1007/s00415-007-2028-z] [Medline: 17503119]
27. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
28. DuBenske LL, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010;30(6):732-744 [FREE Full text] [doi: 10.1177/0272989X10386382] [Medline: 21041539]
29. Jimbo M, Kelly-Blake K, Sen A, Hawley ST, Ruffin MT. Decision Aid to Technologically Enhance Shared decision making (DATES): study protocol for a randomized controlled trial. Trials 2013 Nov 11;14:381 [FREE Full text] [doi: 10.1186/1745-6215-14-381] [Medline: 24216139]
30. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr 04;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
31. Wilkes M, Srinivasan M, Cole G, Tardif R, Richardson LC, Plescia M. Discussing uncertainty and risk in primary care: recommendations of a multi-disciplinary panel regarding communication around prostate cancer screening. J Gen Intern Med 2013 Nov;28(11):1410-1419 [FREE Full text] [doi: 10.1007/s11606-013-2419-z] [Medline: 23649782]
32. Smits R, Bryant J, Sanson-Fisher R, Tzelepis F, Henskens F, Paul C, et al. Tailored and integrated web-based tools for improving psychosocial outcomes of cancer patients: the DoTTI development framework. J Med Internet Res 2014 Mar 14;16(3):e76 [FREE Full text] [doi: 10.2196/jmir.2849] [Medline: 24641991]

XSL•FO
RenderX

33. Roter D. In the digital world of the EMR, Open Notes change everything. Isr J Health Policy Res 2015;4:17 [FREE Full text] [doi: 10.1186/s13584-015-0012-0] [Medline: 25874079]
34. Lafata JE, Shay LA, Brown R, Street RL. Office-based tools and primary care visit communication, length, and preventive service delivery. Health Serv Res 2016 Apr;51(2):728-745 [FREE Full text] [doi: 10.1111/1475-6773.12348] [Medline: 26256283]
35. van Bruinessen IR, van Weel-Baumgarten EM, Gouw H, Zijlstra JM, van Dulmen S. An integrated process and outcome evaluation of a web-based communication tool for patients with malignant lymphoma: randomized controlled trial. J Med Internet Res 2016 Jul 29;18(7):e206 [FREE Full text] [doi: 10.2196/jmir.5877] [Medline: 27473173]
36. Olomu A, Hart-Davidson W, Luo Z, Kelly-Blake K, Holmes-Rovner M. Implementing shared decision making in federally qualified health centers, a quasi-experimental design study: the Office-Guidelines Applied to Practice (Office-GAP) program. BMC Health Serv Res 2016 Aug 02;16(a):334 [FREE Full text] [doi: 10.1186/s12913-016-1603-3] [Medline: 27484348]
37. Lussier M, Richard C, Glaser E, Roberge D. The impact of a primary care e-communication intervention on the participation of chronic disease patients who had not reached guideline suggested treatment goals. Patient Educ Couns 2016 Apr;99(4):530-541. [doi: 10.1016/j.pec.2015.11.007] [Medline: 26657041]
38. Hakone A, Harrison L, Ottley A, Winters N, Gutheil C, Han PKJ, et al. PROACT: iterative design of a patient-centered visualization for effective prostate cancer health risk communication. IEEE Trans Vis Comput Graph 2017 Jan;23(1):601-610. [doi: 10.1109/TVCG.2016.2598588] [Medline: 27875175]
39. Renzi C, Fioretti C, Oliveri S, Mazzocco K, Zerini D, Alessandro O, et al. A qualitative investigation on patient empowerment in prostate cancer. Front Psychol 2017;8:1215 [FREE Full text] [doi: 10.3389/fpsyg.2017.01215] [Medline: 28798701]
40. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019 May 13;21(5):e12891 [FREE Full text] [doi: 10.2196/12891] [Medline: 31094342]
41. Yu C, Ke C, Jovicic A, Hall S, Straus S, IP-SDM Team;. Beyond pros and cons - developing a patient decision aid to cultivate dialog to build relationships: insights from a qualitative study and decision aid development. BMC Med Inform Decis Mak 2019 Sep 18;19(1):186 [FREE Full text] [doi: 10.1186/s12911-019-0898-5] [Medline: 31533828]
42. Kondylakis H, Axenie C, Kiran Bastola D, Katehakis DG, Kouroubali A, Kurz D, et al. Status and recommendations of technological and data-driven innovations in cancer care: focus group study. J Med Internet Res 2020 Dec 15;22(12):e22034 [FREE Full text] [doi: 10.2196/22034] [Medline: 33320099]
43. Hazara AM, Durrans K, Bhandari S. The role of patient portals in enhancing self-care in patients with renal conditions. Clin Kidney J 2020 Feb;13(1):1-7 [FREE Full text] [doi: 10.1093/ckj/sfz154] [Medline: 32082547]
44. Gilljam B, Nygren JM, Svedberg P, Arvidsson S. Impact of an electronic health service on child participation in pediatric oncology care: quasiexperimental study. J Med Internet Res 2020 Jul 28;22(7):e17673 [FREE Full text] [doi: 10.2196/17673] [Medline: 32720907]
45. van Velsen L, Flierman I, Tabak M. The formation of patient trust and its transference to online health services: the case of a Dutch online patient portal for rehabilitation care. BMC Med Inform Decis Mak 2021 Jun 12;21(1):188 [FREE Full text] [doi: 10.1186/s12911-021-01552-4] [Medline: 34118919]
46. Dimitri P, Fernandez-Luque L, Banerjee I, Bergadá I, Calliari LE, Dahlgren J, et al. An eHealth framework for managing pediatric growth disorders and growth hormone therapy. J Med Internet Res 2021 May 20;23(5):e27446. [doi: 10.2196/27446] [Medline: 34014174]
47. Schubbe D, Yen RW, Saunders CH, Elwyn G, Forcino RC, O'Malley AJ, et al. Implementation and sustainability factors of two early-stage breast cancer conversation aids in diverse practices. Implement Sci 2021 May 10;16(1):51 [FREE Full text] [doi: 10.1186/s13012-021-01115-1] [Medline: 33971913]
48. Kim S, Aceti M, Baroutsou V, Bürki N, Caiata-Zufferey M, Cattaneo M, et al. Using a tailored digital health intervention for family communication and cascade genetic testing in Swiss and Korean families with hereditary breast and ovarian cancer: protocol for the DIALOGUE study. JMIR Res Protoc 2021 Jun 11;10(6):e26264 [FREE Full text] [doi: 10.2196/26264] [Medline: 34114954]
49. Lu X, Zhang R. Association between eHealth literacy in online health communities and patient adherence: cross-sectional questionnaire study. J Med Internet Res 2021 Sep 13;23(9):e14908 [FREE Full text] [doi: 10.2196/14908] [Medline: 34515638]
50. Liu PL, Zhao X, Ye JF. The effects of the use of patient-accessible electronic health record portals on cancer survivors' health outcomes: cross-sectional survey study. J Med Internet Res 2022 Oct 24;24(10):e39614 [FREE Full text] [doi: 10.2196/39614] [Medline: 36279157]
51. Yılmaz NG, Sungur H, van Weert JCM, van den Muijsenbergh METC, Schouten BC. Enhancing patient participation of older migrant cancer patients: needs, barriers, and eHealth. Ethn Health 2022 Jul;27(5):1123-1146. [doi: 10.1080/13557858.2020.1857338] [Medline: 33316171]
52. Grynne A, Wångdahl J, Fristedt S, Smith F, Browall M. Women's experience of the health information process involving a digital information tool before commencing radiation therapy for breast cancer: a deductive interview study. BMC Health Serv Res 2023 Aug 09;23(1):842 [FREE Full text] [doi: 10.1186/s12913-023-09837-2] [Medline: 37559113]

XSL•FO
RenderX

53. Orstad S, Fløtten Ø, Madebo T, Gulbrandsen P, Strand R, Lindemark F, et al. "The challenge is the complexity" - a qualitative study about decision-making in advanced lung cancer treatment. Lung Cancer 2023 Sep;183:107312 [FREE Full text] [doi: 10.1016/j.lungcan.2023.107312] [Medline: 37481888]
54. Luu N, Pitts S, Petty B, Sawyer MD, Dennison-Himmelfarb C, Boonyasai RT, et al. Provider-to-provider communication during transitions of care from outpatient to acute care: a systematic review. J Gen Intern Med 2015 Dec 21;31(4):417-425. [doi: 10.1007/s11606-015-3547-4]
55. Zamorano P, Tellez A, Muñoz P, Sapag JC, Martinez M. Effect of COVID-19 pandemic on the implementation of a multimorbidity person-centered care model: a qualitative study from health teams' perspective. PLoS One 2022;17(3):e0265091 [FREE Full text] [doi: 10.1371/journal.pone.0265091] [Medline: 35316285]
56. Fumagalli LP, Radaelli G, Lettieri E, Bertele' P, Masella C. Patient Empowerment and its neighbours: clarifying the boundaries and their mutual relationships. Health Policy 2015 Mar;119(3):384-394. [doi: 10.1016/j.healthpol.2014.10.017] [Medline: 25467286]
57. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
58. Bokolo Anthony Jnr. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
59. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
60. Mirzaei T, Kashian N. Revisiting effective communication between patients and physicians: cross-sectional questionnaire study comparing text-based electronic versus face-to-face communication. J Med Internet Res 2020 May 13;22(5):e16965 [FREE Full text] [doi: 10.2196/16965] [Medline: 32401213]
61. Alexander JA, Hearld LR, Mittler JN, Harvey J. Patient-physician role relationships and patient activation among individuals with chronic illness. Health Serv Res 2012 Jun 18;47(3 Pt 1):1201-1223 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01354.x] [Medline: 22098418]
62. Hall MA, Dugan E, Zheng B, Mishra AK. Trust in physicians and medical institutions: what is it, can it be measured, and does it matter? Milbank Q 2001;79(4):613-39, v [FREE Full text] [doi: 10.1111/1468-0009.00223] [Medline: 11789119]
63. Raj M, Platt JE, Wilk AS. Trust in provider care teams and health information technology-mediated communication. Am J Manag Care 2020 Jan;26(1):23-25 [FREE Full text] [doi: 10.37765/ajmc.2020.42141] [Medline: 31951355]
64. Fisher K, Tan A, Matlock D, Saver B, Mazor K, Pieterse A. Keeping the patient in the center: common challenges in the practice of shared decision making. Patient Educ Couns 2018 Dec;101(12):2195-2201 [FREE Full text] [doi: 10.1016/j.pec.2018.08.007] [Medline: 30144968]
65. Scholl I, LaRussa A, Hahlweg P, Kobrin S, Elwyn G. Organizational- and system-level characteristics that influence implementation of shared decision-making and strategies to address them - a scoping review. Implement Sci 2018 Mar 09;13(1):40 [FREE Full text] [doi: 10.1186/s13012-018-0731-z] [Medline: 29523167]
66. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
67. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011;4:47-55 [FREE Full text] [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
68. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
69. Alowais S, Alghamdi S, Alsuhebany N, Alqahtani T, Alshaya A, Almohareb S, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689 [FREE Full text] [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
70. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implement Sci 2024 Mar 15;19(1):27 [FREE Full text] [doi: 10.1186/s13012-024-01357-9] [Medline: 38491544]
71. Hatem R, Simmons B, Thornton J. A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. Cureus 2023 Sep;15(9):e44720 [FREE Full text] [doi: 10.7759/cureus.44720] [Medline: 37809168]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
health IT: health information technology
MeSH: Medical Subject Headings
mHealth: mobile health
PICo: Population, Interest, and Context

XSL•FO
RenderX

Edited by G Eysenbach; submitted 14.05.24; peer-reviewed by M Fiordelli, X Tan, P Martinez-Angulo; comments to author 17.07.24; revised version received 23.10.24; accepted 30.11.24; published 16.01.25.

Please cite as:
Song M, Elson J, Bastola D
Digital Age Transformation in Patient-Physician Communication: 25-Year Narrative Review (1999-2023)
J Med Internet Res 2025;27:e60512
URL: https://www.jmir.org/2025/1/e60512
doi:10.2196/60512
PMID:

©Mingming Song, Joel Elson, Dhundy Bastola. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

25 Years of Electronic Health Record Implementation Processes: Scoping Review

Harriet Finnegan[1], MSc; Nicola Mountford[1], PhD

School of Business, Maynooth University, Kildare, Ireland

Corresponding Author:
Harriet Finnegan, MSc
School of Business
Maynooth University
3rd Floor TSI Building
Maynooth
Kildare, W23 X04D
Ireland
Phone: 353 17083609
Email: harriet.finnegan.2017@mumail.ie

Abstract

Background: Electronic health record (EHR) systems have undergone substantial evolution over the past 25 years, transitioning from rudimentary digital repositories to sophisticated tools that are integral to modern health care delivery. These systems have the potential to increase efficiency and improve patient care. However, for these systems to reach their potential, we need to understand how the process of EHR implementation works.

Objective: This scoping review aimed to examine the implementation process of EHRs from 1999 to 2024 and to articulate process-focused recommendations for future EHR implementations that build on this history of EHR research.

Methods: We conducted a scoping literature review following a systematic methodological framework. A total of 5 databases were selected from the disciplines of medicine and business: EBSCO, PubMed, Embase, IEEE Explore, and Scopus. The search included studies published from 1999 to 2024 that addressed the process of implementing an EHR. Keywords included “EHR,” “EHRS,” “Electronic Health Record*,” “EMR,” “EMRS,” “Electronic Medical Record*,” “implemen*,” and “process.” The findings were reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist. The selected literature was thematically coded using NVivo qualitative analysis software, with the results reported qualitatively.

Results: This review included 90 studies that described the process of EHR implementation in different settings. The studies identified key elements, such as the role of the government and vendors, the importance of communication and relationships, the provision of training and support, and the implementation approach and cost. Four process-related categories emerged from these results: compliance processes, collaboration processes, competence-development processes, and process costs.

Conclusions: Although EHRs hold immense promise in improving patient care, enhancing research capabilities, and optimizing health care efficiency, there is a pressing need to examine the actual implementation process to understand how to approach implementation. Our findings offer 7 process-focused recommendations for EHR implementation formed from analysis of the selected literature.

(J Med Internet Res 2025;27:e60077) doi:10.2196/60077

KEYWORDS

electronic health record system; EHR; electronic medical record; EMR; scoping review; process; implementation

Introduction

Background

Over the past 25 years, electronic health record (EHR) systems have undergone remarkable development, becoming an important element of modern health care [1]. Since their inception in the late 20th century, EHRs have advanced substantially, propelled by both technological innovations and critical policy reforms [2]. Key legislation, such as the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, was instrumental in accelerating the

XSL•FO
RenderX

widespread adoption of EHRs, solidifying their role in enhancing health care delivery. Today, EHRs are indispensable for improving patient safety, improving operational efficiency, and ensuring that vital patient information is securely stored and easily accessible across health care settings [3].

Research to date has examined some of the key pre-implementation indicators of EHR adoption and success. Studies have revealed how certain organizational characteristics are likely to predict success, such as the size of the organization [4,5] and where it is located [6,7]. Other research has provided important details on postimplementation evaluations by users. These evaluations may include an increase or a decrease in the difficulty of tasks for the physician [8], the impact on patient care [9], and effects on privacy and security [10].

There is an extensive body of literature reviews published over the past 2 decades that explore the success factors and challenges associated with EHR implementation [11-20]. This has substantially and importantly improved our understanding of those factors that must be considered when planning EHR implementation. Fennelly et al [20], for example, identified 15 interlinked organizational, human, and technological factors that affect successful EHR implementation across primary, secondary, and long-term care settings. Our study built on this body of work by returning to the source literature with a process focus—digging beneath the success factors and challenges to examine the underpinning processes of EHR implementation. A process focus allows us to examine the connections between the factors already identified, unveiling new aspects, such as flow, activity, and temporality [21]. We, therefore, foregrounded and synthesized those papers that center on this vital process of *implementing* an EHR rather than delving into the technical intricacies of EHR technology itself.

Defining an EHR

The World Health Organization defines the EHR as "a longitudinal record of patient health information generated by one or more encounters in any care delivery setting" [22]. These records may include details such as demographics, progress notes, vital signs, medications, immunizations, lab results, and radiology reports, which all provide a comprehensive view of a patient's health. Similarly, the Centers for Medicare and Medicaid Services (CMS) emphasizes that EHRs play a crucial role in helping health care providers maintain accurate and up-to-date patient data over time, ensuring that key clinical and administrative information is easily accessible and securely shared among authorized users [23].

Electronic medical records (EMRs), in contrast, are records created by practitioners for specific encounters, examples of which may be hospital visits or the use of facilities within ambulatory environments. Finally, a personal health record (PHR) is data controlled by the patient through the use of an electronic application that they are able to provide to their health practitioners. PHRs support patient-centered health care by making medical records and other relevant information accessible to patients, assisting patients in health self-management [24]. We focused exclusively on EHRs rather than EMRs or PHRs. The scope of EHRs is generally larger than that of EMRs or PHRs. They require a broad range of data types and need to be able to connect these across systems, whereas EMRs are generally confined to an individual practice [25], limiting their scope, and PHRs are generally subject to personal management [26]. This makes them less complex, and therefore less interesting from a research perspective, than EHRs. EHRs require higher levels of interoperability, regulatory challenges, stakeholder involvement, and cost and time investment than EMRs or PHRs, making them the ideal focus of this research.

Although the terms "EHR" and "EMR" are conceptually distinct, they are often used interchangeably in the literature. We recognize that the definitions we used for EHRs and EMRs in our review are not universally observed, and the terminology used in the literature and in practice often reflects the contexts in which these systems are implemented rather than the strict definitional boundaries placed upon them. In instances where studies, such as Felt-Lisk et al [27], have examined systems referred to as EHRs but may have operationally aligned with our definition of EMRs, we opted for a more inclusive approach. This definitional ambiguity may mean that some of our recommendations may equally apply to systems labeled as EMRs rather than EHRs. Addressing this overlap is essential for advancing a more unified understanding of electronic records in health care; however, it was not the purpose of this research.

Implementation Process

Rather than the broader issue of EHR adoption, which relates to the widespread acceptance and use of the technology across health care settings, our focus was on implementation as the practical, often complex, process of integrating EHR technology into health care environments. Adoption is the "phase of investigation, research, consideration and decision making in order to introduce a new innovation into the organization" [28,29]. Implementation is the "phase of internal strategy formation, project definition and activities in which an adopted application is introduced within the organization, with the aim of removing reservations and stimulating the optimum use of the application" [29]. Although adoption can occur both prior to and after implementation, these terms describe separate, distinct actions. We discussed EHRs in terms of the process of implementation, as defined by Bouwman et al [29].

Study Focus

This scoping review examined the EHR implementation process over the past 25 years. The paper presented a qualitative thematic analysis of 90 relevant academic papers describing the EHR implementation process. Our review of 25 years' worth of EHR implementation processes ultimately offers some advice and hope for more effective EHR implementations and those policies that support them.

Methods

Overview

A scoping review was conducted according to the 5-stage framework by Arksey and O'Malley [30]. Results were reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Multimedia Appendix 1).

XSL•FO
RenderX

Stage 1: Identifying the Research Question

Our research question was as follows: What have we learned about the process of implementing EHRs over the past 25 years?

Stage 2: Identifying Relevant Studies

A systematic literature search was conducted across 5 databases to identify all relevant literature: EBSCO, PubMed, EMBASE, IEEE Explore, and Scopus. The following specific keywords were used in the search strategy: (EHR OR EHRS OR Electronic Health Record* OR EMR OR EMRS OR Electronic Medical Record*) AND implemen* AND process (all in abstract). Detailed search strategies are provided in Multimedia Appendix 2. The research period was from January 1999 to August 2024 in line with the focus of this special issue looking at the past 25 years. We chose to include the word "process" in our search string, in addition to variations of the word "implement" in order to refine our results to include discussions focused on the process of implementing EHRs rather than those centered on the outcomes of EHR implementation. We also chose to search for the term "electronic medical records," in addition to EHRs in the initial stage, even though the focus of our paper was on EHRs. This inclusion of EMRs in the initial search allowed us to account for discrepancies in the language used when discussing EHRs. This initial inclusion in our search criteria also meant we were able to manually exclude EMR studies that were purely discussions of EMRs and manually include EMR studies that also discussed EHRs or possessed the same functionalities of an EHR. This broadening of our search terms meant that our analysis of the literature was more thorough. We chose only empirical studies because we sought evidence of specific experiences of EHR implementation processes.

Inclusion Criteria

Studies meeting the following inclusion criteria were included: (1) published from January 1999 to August 2024, (2) peer reviewed, (3) journal papers, (4) published in English, and (5) mentioned the implementation process of EHRs.

Exclusion Criteria

Studies meeting the following criteria were excluded: (1) focused purely on the aftermath of implementation; (2) did not discuss the implementation process; (3) did not clearly report methods; (4) nonempirical; (5) not published in a peer-reviewed journal; (6) letters to the editor, editorials, or essays; (7) gray literature and review papers; and (8) discussed single-location EMRs.

Stage 3: Study Selection

All citations were uploaded to Covidence, a web-based research tool used by researchers to collaborate and organize citations in systematic reviews. Duplicates were removed automatically by the software and manually by both authors. Both authors screened all 4454 remaining papers by title and abstract and 226 papers by full text. Both authors also reviewed any disagreements before proceeding to the next stage of screening. This was done at each stage of screening to ensure consistency between decisions. Any disagreements were resolved by both authors discussing the eligibility of the paper in relation to the research focus and the agreed-upon inclusion and exclusion criteria until reaching consensus.

Stage 4: Charting the Data

Data were extracted by the first author using Microsoft Excel, including the following details: author(s)/publication year, country of origin, aim(s)/purpose, study design, type of organization, study population/sample size, record type, and methods.

Stage 5: Collating, Summarizing, and Reporting Results

After reviewing all full texts of the eligible studies, the first author loaded the remaining 90 full texts into NVivo, a qualitative data analysis software program. This software was used to manually organize the texts and facilitate the thematic coding of the data. All 90 full texts were then thematically coded by the first author. The second author thematically coded a subset of 10 (11.1%) full texts. Both authors then met to discuss any differences in coding decisions. This was carried out to ensure intercoder reliability.

Results

Sample Characteristics

The paper-screening process is illustrated in detail in Figure 1, while the distribution of the selected studies by year of publication is shown in Figure 2. Among the 90 papers, some discussed more than 1 country in their study. This included 37 (41.1%) studies conducted in the United States, 11 (12.2%) in England, 6 (6.7%) in Australia, 5 (5.6%) in Germany, 5 (5.6%) in Denmark, 3 (3.3%) in the Netherlands, 3 (3.3%) in Italy, 3 (3.3%) in Canada, 3 (3.3%) in Norway, 2 (2.2%) in Singapore, 2 (2.2%) in Kenya, 2 (2.2%) in the United Kingdom, and the remaining countries mentioned in only 1 (1.1%) study each. The majority of studies (n=59, 65.6%) were qualitative, while the remainder (n=19, 21.1%) were quantitative or used mixed methods (n=12, 13.3%). The selected papers begin in 1999 with 1 publication. Examining these figures showed initial inactivity on the topic, followed by steady growth starting in 2008 and peaking in 2017, with a slight decline in recent years. The peak years from 2014 to 2017 suggest a strong research interest during this time. Recently, from 2019 to 2024, there seemed to be a stabilization in research output, with around 4 publications annually. The main characteristics of the included studies are presented in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 1. PRISMA-ScR flow diagram showing the study selection process. PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.

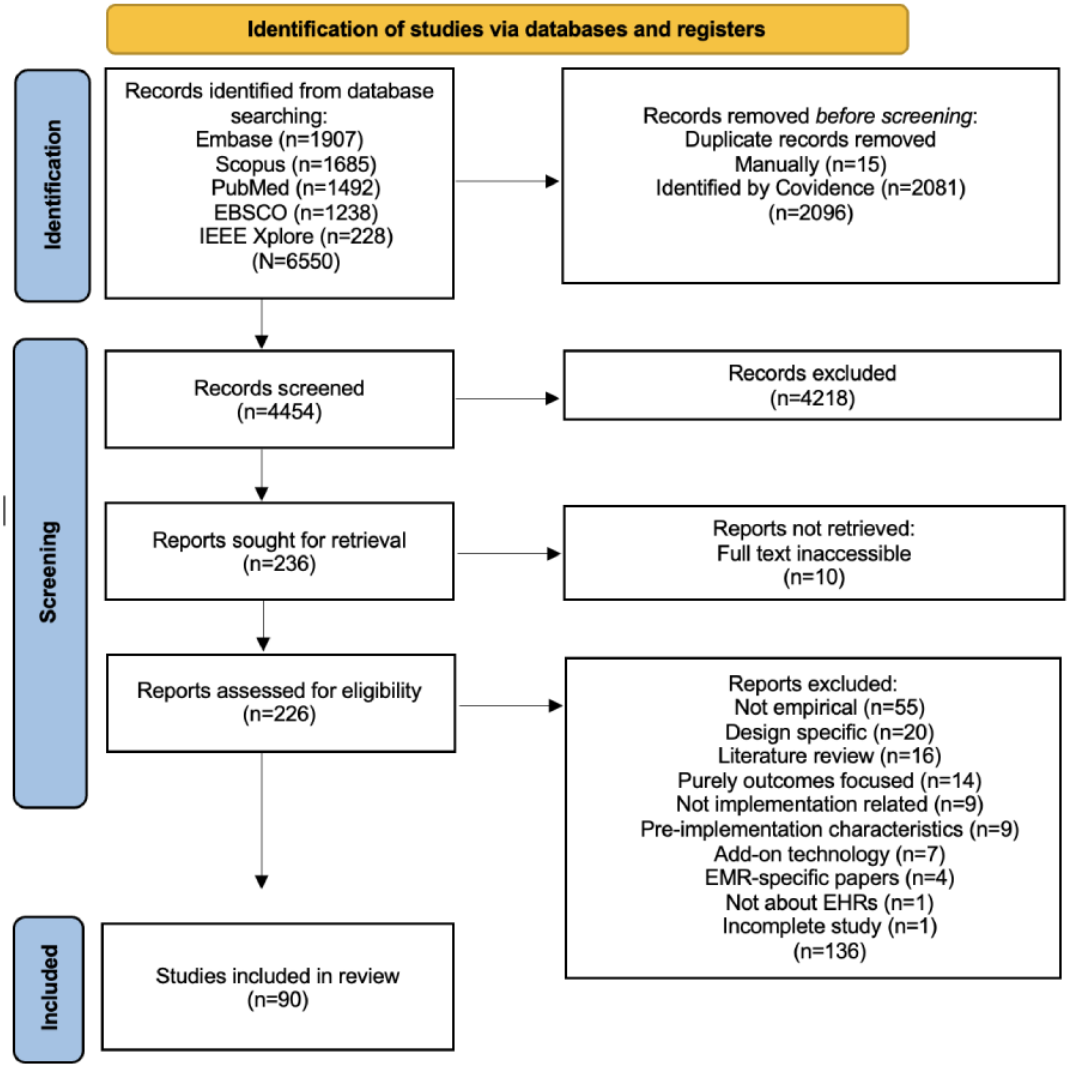

XSL•FO
RenderX

Figure 2. Distribution of the selected papers by year of publication.

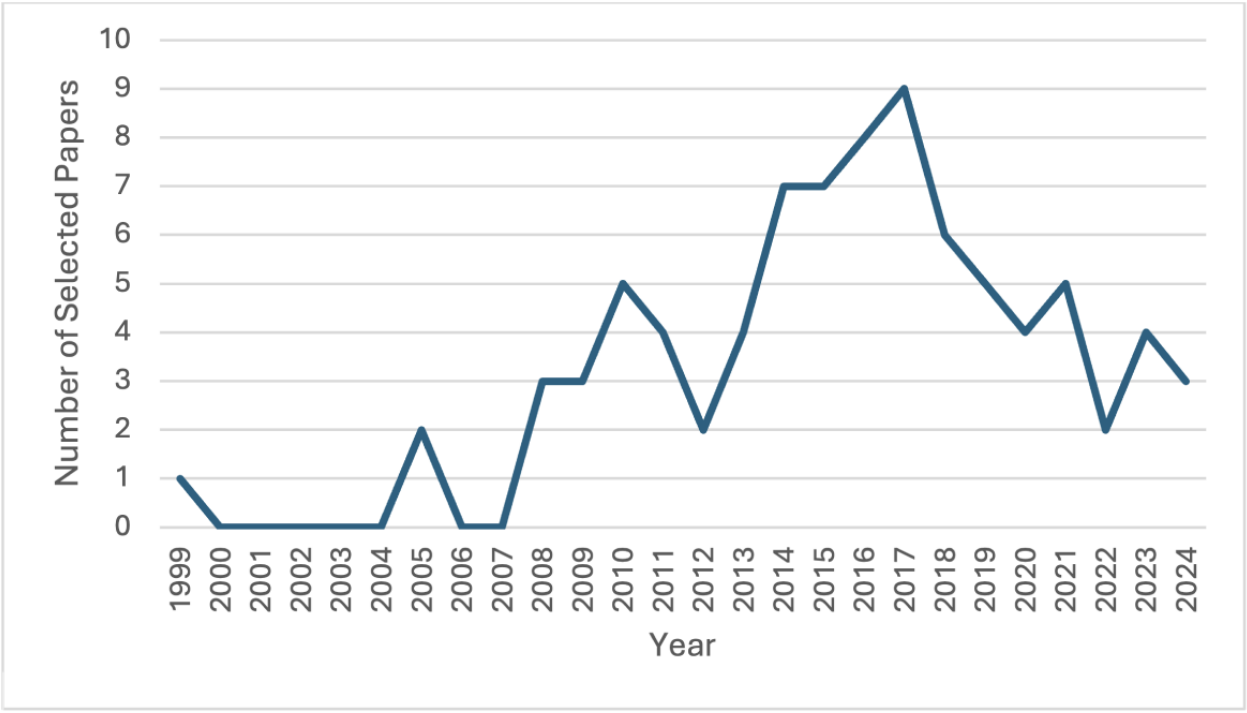

During the thematic coding process, sections of text within each paper were allocated to specific codes rather than entire papers being allocated to one code. As a result, multiple different codes may have appeared in one paper, as can be seen in the coding distribution table (Table 1). The full coding structure is provided in Multimedia Appendix 4.

Table 1. Coding distribution across studies (N=90).

Theme and codes	Studies, n (%)	References
Compliance		
Government	29 (32.2)	[27,31-58]
Policy	21 (23.3)	[31,35,37-41,44,46,48,49,52,55,58-65]
Regulations	10 (11.2)	[31,34,36,55,62,63,65-68]
Vendors	20 (22.2)	[27,37,38,49,53,59,60,67,69-80]
Collaboration		
Communication	40 (44.4)	[32-34,36-38,40-42,44-46,49-51,59-61,63,66,68,70,71,74-76,78,79,81-92]
Relationships	30 (33.3)	[32-36,40,44,48-50,57,60,65,66,68-71,76,81,84,85,89,91-97]
Competence		
Training	44 (48.9)	[31,33,38-42,46,48,50,53,54,59-61,63,64,66-70,76,77,79,81-83,86,89,90,94-106]
Support	35 (38.9)	[27,32,36,39,40,42,43,45,48,50,53,54,60,66,67,69-71,74-76,79,81,82,84,86,90,95,97,105-110]
Cost		
Cost	41 (45.6)	[27,31,33,35-40,42,43,46,48-50,52,54,56,60,63,65,66,68,69,71,72,75,82,88,90,92-94,96,97,106,108,110-113]

The Implementation Process

Overview

The 90 studies provide a broad overview of the process of EHR implementation over the past 25 years.

A process-based view considers time as a key element. EHR implementation is described by Boonstra et al [34] as a complex and time consuming process, and by Hernández-Ávila et al [48] as a gradual and often slow process. The main benefits of EHRs are likely to accrue in the long term, so it is important to envision them as long-term change management endeavors [96]. These long-term benefits may never be realized, however, where short-term time pressures jeopardize implementation. This may arise where there is limited time available to adapt the system to local conditions [98] or where the pace of implementation is being dictated by other ongoing projects [49]. Deadlines imposed by external actors may also result in implementation timelines being rushed. For example, political considerations

XSL•FO
RenderX

may frame procurement arrangements [52], or vendors might enforce tight deadlines [73]. Diffusion processes require good planning and consume both time and resources [110]. Time looms large in our reviewed studies, justifying our process focus. There is a constant pressure of time in relation to the implementation [38], but the reality is that pre-existing day-to-day pressures may limit the pace of implementation [27].

Four central process-related categories emerged from the thematic review: (1) compliance processes featured indirectly through references to the important role of the government, policy, regulations, and vendors; (2) collaboration processes centered on the work of managing these implementations through communication and relationships; (3) competence-building processes included discussions of training and support; and (4) process costs drew attention to areas of cost.

Compliance Processes in EHR Implementation

Our review suggested that compliance processes in EHR implementation center around the government, policy, regulations, and vendors. The literature referenced government-related processes in 29 (32.2%) papers. There were discussions of policy-related processes in 21 (23.3%) papers. Regulation-related processes were considered in 10 (11.1%) papers, and vendor-related processes were referenced throughout 20 (22.2%) papers.

Government-Related Compliance Processes

EHR implementation problems often arise at the national level rather than at regional or health service–specific levels. This is a result of the increased complexity of national implementation [38]. Nationwide implementation requires a huge change from stakeholders [42]. In a national project, practitioners operating in public and private domains require different engagement strategies to secure their buy-in [51]. A national strategy is, therefore, needed for a national EHR implementation process [110]. National EHRs require adaptation with clinician practices nationwide to ensure workflow processes are consistent [91]. Further recommendations from the literature include changing from a top-down implementation model to increased involvement of local organizations in decision-making [35]. Planners need to ensure potentially unclear areas are clarified with program management and that clinicians are informed and consulted [51]. National-level implementation is not just simple system installation; it is discovering ideas from private institutions and using them to drive best practices across the system nationwide [49]. Bottom-up implementation is, however, time-consuming and may hinder future collaborations [94], whereas the top-down nature of some projects contributes to a lack of organizational and user involvement in decision-making [52].

The role of the government in initiating and maintaining momentum around EHR implementation processes came through strongly from our review. Government leadership is considered a strategic advantage when the goal is the sustained enforcement of EHR initiatives [31]. Mature EHR systems benefit from being well integrated into the national health-planning documents of the government [39], while national reimbursement policies can increase EHR dissemination [49]. To maintain momentum, government commitment must be both strong and continued and accompanied by political support [36]. eHealth experts are, therefore, generally in favor of a strong central solution for political regulation problems [62]. It is, however, not just the national government that matters. Support from the local government also plays a crucial role in terms of finances, provision of resources, and technical support [41]. Indeed, implementation processes suffer where they lack sufficient coordination with the local government [41]. However, strong national or local government policies mandating the use of specific eHealth solutions may support EHR implementation processes [41]. Excessive regulation may also hinder the long-term sustainability of EHR initiatives [31].

Changes in the government may lead to having to modify an overall implementation process [35]. Shifts in the government strategy affect the power dynamics between national branches of large IT companies compared to national information and communication technology companies [55]. A change in the government also has the potential to result in uncertainty about the future of national programs [52,58]. It is, therefore, advisable for system evaluators to form close relationships with policy makers [35]. Alternatively, if there is a lack of a government-level information policy at the time of design and development, the sustainability of the EHR implementation is endangered [48].

The Role of Vendors

Governments are not, however, the only organizational actors who can manage or lead EHR compliance processes. Vendor organizations that have strong centralized administrative and medical structures are drivers of organizational policies and processes that are crucial to compliance [64]. Indeed, in some instances, large private vendors may enter a market specifically to provide unified access to health care data [49]. It is the process of compliance of large technology companies with uniform national standards and rules that enables EHR implementation in this instance [55,60].

Vendor recommendations influence software choice processes [69]. EHR implementors cannot simply buy from the same vendor as their existing financial system and assume turnkey, seamless interoperability [71]. There are often limitations to existing vendor-based EHRs when compared with in-house systems, including autonomy of practice decisions at each hospital [70]. These limitations are also seen in difficult-to-assess EHR usability as a result of restrictive vendor license agreements [59].

EHR vendors are generally at least partly responsible for the amount of training received by users [69] and sometimes also responsible for providing technical support [27]. Indeed, vendors may use the provision of ongoing technical support as a form of training [77]. Both peer and technical support may be issued from these vendors, which can help end-users optimize their use of the EHRs and solve issues [67]. Even large vendors may need to consider EHR adaptations for small practices as part of the sale of the EHR, ensuring that a person within the practice conducts training and adapts it to the specific needs of the practice rather than relying on training by IT specialists [53].

XSL•FO
RenderX

Collaboration Processes in EHR Implementation

Collaboration in the literature includes communication and relationships throughout the management of the implementation process. Communication was mentioned in 40 (44.4%) papers. Relationships were discussed in 30 (33.3%) papers.

Collaboration among team members is important [60], as is communicating clear expectations and guidelines [61]. Consistent, reliable communication fosters trust [83], and well-articulated visions are important for the management of expectations [74]. Strong communication practices are an enabler of successful implementation outcomes [84]. However, a lack of communication during planning can cause issues [49].

Collaborating facilitates successful EHR implementation [95]. Cuccinello et al [36] illustrate this in their study of a vendor's collaborative relationship with a health care department in Italy. Relationships with vendors build confidence within the organization and ensure strong external support [60]. Where, however, communication breaks down, it can jeopardize implementation processes [41]. Relationships between contractors and suppliers may become more impersonal and distant as a result [34]. Direct and close channels of communication between the implementer hospitals and software suppliers are, therefore, essential from the outset [89]. Kiepek and Sengstack [84] suggest an open and transparent relationship with external support from vendors, beginning with initial negotiations.

Competence-Building Processes in EHR Implementation

Competence building in the literature includes training and technical support. The literature referenced training in 44 (48.9%) papers. Support was discussed in 35 (38.9%) of the selected papers.

Training is necessary for successful EHR implementation [95]. This training should provide practitioners with the skills necessary to operate the system, as well as the confidence to help them adapt to the new system [98]. Sufficient training for practitioners is associated with improved well-being [100] and has a positive and substantial influence on perceived ease of use [67]. Hiring experts can help, for example, in providing technical support both during the implementation and afterward [110], and hiring clinical informaticists can help support EHR implementation and sustainment [86].

Issues surrounding training and implementation processes are numerous. These include a lack of resources for EHR training [70], the need to train medical personnel [38], a lack of training for smaller entities [38], ensuring sufficient time for staff training [41], providers and patients not receiving adequate support and training [63], and a lack of appropriately specific training [102]. The need for sufficient trained personnel is felt across all stages of an implementation process [39]. Staff that are not trained to interact with an EHR can hinder implementation [31,42]. Recommendations to enhance EHR training include ensuring that it occurs pre- and postimplementation on a continued cycle [86]. Decreased support is an additional issue [70], whether that be in relation to practitioner support [43] or in determining the right type of IT support for successful implementation [97].

Process Costs in EHR Implementation

Cost considerations were shown to be an important part of the implementation process, with 41 (45.6%) papers making reference to cost. Adoption of appropriate processes is crucial regarding system development time and budget [42]. As such, successful implementation should have sustainable funding that aligns with a national strategy for eHealth [41]. Determining costs and measures of success is a vital part of project management, especially in pre-implementation [108]. The role of monetary incentives in this stage of the process is also an enabler [36]. One of the challenges of a national project is that funding sources depend largely on the government [31]. Implementing a system on a national scale is an extremely complex activity [52]. It is difficult to manage, costly to maintain, and hard to sustain [89]. Financially, the most serious obstacle in implementing EHRs is the cost of electrification [38]. Further customization also leads to increased maintenance costs [60]. The high upfront cost of EHRs for small practices is a major factor limiting their use [27]. Most hospitals report substantial financial challenges in EHR implementation and use, including EHR and broadband implementation costs and the limited availability of grants and loans to support EHR implementation and use [113]. A lack of capital resources can hinder the EHR implementation process [43]. Other issues include inadequate capital for investment and maintenance costs [110], large-scale procurement being undertaken to save costs [52], attempts to implement EHRs halting due to financial issues [54], and the high costs of implementing a system [97].

Discussion

Principal Findings

This scoping review revealed 4 main areas for consideration in the EHR implementation process. These areas are compliance, collaboration, competence, and costs. Specific issues recurred in each area throughout the literature: the role of the government and the role of vendors in compliance processes, the importance of communication and relationships to facilitate collaboration processes, training and support to build competence; and the cost of financing throughout the implementation process. Many of the 19 interventions identified by Boonstra et al [13], and the 15 factors identified by Fennelly et al [20] pertaining to successful or effective implementation strategies featured within our findings. Rather than replicate the reviews of Fennelly et al [20] or Boonstra et al [13], however, we built on these to provide a focused exploration of the implementation process to complement the broader insights that these reviews offer. We built on these works to offer practical recommendations for use in the implementation process.

Our study and research question focused on the *process* of implementation of EHRs. The process of implementation deserves attention, especially when we consider that the implementation process influences implementation outcomes. Put differently, it is not just what system we implement, nor the recognition of challenges or facilitators, but also the process by which we navigate and manage these over time that ultimately decides how successful an EHR might be. There is, therefore, a need to understand how to approach these implementation

XSL•FO
RenderX

processes in a way that is informed by previous implementations and the appropriate literature. To address this, we provide recommendations produced from our synthesis of the selected literature reviewed in this paper. Organizing our findings into compliance processes, collaboration processes, competence processes, and process costs allowed us to provide distinct actionable recommendations for stakeholders. Our recommendations emphasize implementation priorities in a way that facilitates targeted interventions. Our recommendations not only reaffirm established principles found in reviews, such as those by Boonstra et al [13] and Fennelly et al [20], but also provide a modernized roadmap for undertaking the implementation process.

Seven Process-Based Recommendations for EHR Implementation

The literature consistently demonstrates the value of both the government and vendors in ensuring sustained EHR compliance processes, while highlighting some dangers. We drew from this some recommendations for successful compliance processes. The first recommendation is to *maintain close and ongoing government/implementor relationships to balance user, government, and organizational requirements in the short and longer terms.* There is a need for the central government to work with hospitals and local governments to ensure EHRs satisfy the requirements of users [49]. Government planners have the power to exert more influence on public health care providers than private providers [51]. Liaising with organizations and policy makers to inform strategic decisions and policy making is important [35]. As part of this, it is necessary for evaluators of EHRs to form close relationships with policy makers [35]. The second recommendation is to *rebalance vendor/implementor relationships to ensure small-site customization and training that will drive sustained compliance.* The role of the vendor is notable, with numerous best-practice sites viewing their vendor as an active partner in the implementation and compliance process [76]. Designating key contact people to act as liaisons may help foster this relationship from the beginning. Fostering open, regular communication between vendors and implementers can also be done through regular meetings.

The literature clearly highlights the importance of collaboration processes across teams, as well as multistakeholder communication, in ensuring sustained support for EHR implementation. Breakdown of communication and relationships is damaging to implementation processes. We built on these findings to suggest some collaboration processes that support EHR implementation. Our third recommendation, therefore, is to *cultivate varied sources of support across stakeholder groups*. Strong and continued commitment and support at the highest level facilitate and support collaboration processes [36]. The types of support needed vary to include political support [36], practitioner support [108], and social support [109]. Supporting the interest in EHRs is an important behavior linked to successful EHR implementation [95]. Indeed, issues may arise where there is inadequate patient and broader community engagement around EHR implementation [88]. The fourth recommendation is to *pay particular attention to communication and collaboration in the implementation-planning phase of EHR implementation, including the development of cross-functional teams, the appointment of "opinion leaders'" and realistic envisioning of postimplementation challenges and benefits*. Effective communication processes are critical in an organization when implementing an EHR [70], especially during the implementation-planning phase [76]. Stakeholders working toward change also need a close collaborative working environment [65]. Forming cross-functional teams [60] and enlisting the collaborative effort of physicians, hospital administrators, IT specialists, and state officials are integral to the process of design and development [48]. Appointing "opinion leaders" to hear concerns of practitioners makes audiences more receptive to implementation [45]. Crafting communication campaigns that balance potential expected benefits with realistic expectations of the challenges faced may be an issue [59]. Anticipating the challenges that will be faced during implementation can cause these issues, as each implementation occurs in a unique environment.

The studies clearly show the importance of competence-building processes, such as training and technical support provision during the implementation process. We drew from this to suggest competence-building processes that will positively contribute to the EHR implementation process. The fifth recommendation is to *ensure that staff are adequately trained to use the systems and for planners to consider the timing of this training when organizing wider implementation processes.* Implementation processes depend on training the staff who will be using the system [77]. Ongoing training and optimization are necessary [103], as well as curricula for EHR training [104]. Training processes must be ongoing, embedded in workflows, and flexible so that they can be tailored to the diverse needs of users [89]. In terms of timing, training should be conducted close to the time of actual implementation of new technology [54], with the most successful training session conducted within a few weeks of the system going live [76]. Some authors recommend that staff be required to complete training by the end of the preparatory phase to retain their access to the EHR [94]. Others suggest sending staff to training classes customized by job role, with trainers on-site for 1-2 weeks after the system goes live [64]. The time it takes practitioners to chart should be addressed early on in training [53], and training focusing on how the EHR will work should be replaced with a focus on how the EHR can be adapted to the practitioners [53]. Vadillo et al [54] stress the importance of providing proper training in basic computer functions, with training conducted in the classroom with an instructor rather than one to one. Recommendations also suggest that senior management provide practice leaders with IT training and have them visit an EHR-based practice [97]. The sixth recommendation is to *assess what support will be needed at each stage of the implementation and ensure that this support is put in place for an appropriate amount of time*. Support is critical during the go-live period [76]. Carayon et al [69] suggest having support staff present from the EHR vendor on the day the EHR goes live and having an expert user present at the clinic for the following weeks as a useful support to implementation processes. The availability of "super users" who offer support at go-live time is also noted as appropriate [60] and particularly useful when considering user support as a higher priority than initial user training [39]. Other notable forms of support are vendor support [60] and industry support

XSL•FO
RenderX

[42]. Informal support (provided via Facebook, involving both vendors and peer-to-peer support) is also noted to be effective and efficient [74]. Support throughout the planning and implementation period ensures clarity of roles, strong communication practices, and a successful outcome [84].

Finally, the literature clearly demonstrates the integral role of cost throughout implementation. We drew from this discussion of cost in the literature to provide our concluding recommendation. The seventh recommendation is to *promote an understanding of the system as a long-term investment*. Some of the literature promotes the development process of a solid government reimbursement plan [49]. However, more central to our recommendation is deRiel et al's study [39]. The authors discuss the importance of understanding the system, its value, and the total cost of ownership so that investments are not seen as one-time expenses but ongoing investments. Other authors suggest that the process of choosing an EHR system should center on its potential for improving clinical care rather than achieving cost savings [96]. Inadequate practitioner consultation processes [46], delays [58], and tendering processes all increase the cost of the overall process [34].

Our findings illustrate that successful implementation processes benefit from meticulous planning [76]. Our recommended processes for addressing compliance, collaboration, competence, and costs within the wider EHR implementation process aim to provide the materials for informed "meticulous *process* planning" to occur. Examining EHR implementation purely from a pre- versus postimplementation perspective of outcomes may blind both researchers and practitioners to the importance of the intervening implementation processes. Our review connects previously established barriers and facilitators with a time-and-action focus to offer additional insights into the process of implementation.

Limitations

Several limitations should be considered in this scoping review. This study focused on EHRs and did not consider the implementation processes of other eHealth systems, which may benefit from future analysis. Our findings also highlight the need for improved standardization in the terminology surrounding electronic records in health care to better differentiate between EHRs and EMRs in both research and practice. This improved clarity would allow for greater comparability across studies and guide more tailored implementation processes. The included studies tended toward the United States and Europe, centering our analysis in the Global North. The search strategy included studies from 1999 onward. The resulting studies selected were mostly from 2007 onward. It is unlikely that this date limitation largely impacted the selected studies; however. it is possible that due to the date limitation, this scoping review may have missed interesting studies conducted prior to 1999. Finally, our specific search terms, chosen with the intention of focusing the literature on the implementation process, may mean that the literature describing the same event but not using those specific search terms was not retrieved. Additionally, shifts in the language and naming of systems in health care mean that this study may apply to systems discussed under a name other than "EHRs," which have not been retrieved during the search strategy of this research. As such, we did not retrieve or include any studies discussing EHRs under the recently emerging term of "digital health records."

Conclusion

Completed implementation of EHRs is integral to improving health care delivery. The findings from this scoping review offer important insights into the complexities of the actual process of implementation and its subprocesses. Our review identified 3 key processes (compliance, collaboration, and competence building), as well as considering overall process costs. In doing so, we offered a new time- and action-based perspective on EHR implementation. Compliance processes reference the role of the government, policy, regulation, and vendors in shaping the implementation process. Collaboration processes promote the need for strong communication and the building of relationships across all stakeholders involved in the implementation process. Competence-building processes focus on ensuring that users are provided with the resources to be able to operate an EHR, centering around the importance of the timing of training and support. Finally, our discussion of process cost illustrates the importance of a time-focused financial approach during the implementation process.

We proposed 7 strategies in this review, which all provide a structured approach to navigating the different areas of implementation. Future research should focus on deepening our understanding of how these outlined strategies change and operate at different stages of the implementation process. This scoping review contributes to the intersection of management and health care research. We hope that the review results and recommended strategies provided will inform areas for future research and help develop future implementation processes.

Acknowledgments

This publication emanated from research conducted with the financial support of Taighde Éireann–Research Ireland (grant GOIPG/2021/290).

Data Availability

All data analyzed during this study are included in this paper.

XSL•FO
RenderX

Authors' Contributions

HF was responsible for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, visualization, and writing—original draft. NM managed supervision, validation, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 68 KB - jmir_v27i1e60077_app1.docx]

Multimedia Appendix 2
Search terms and search strategy for the scoping review.
[DOCX File , 20 KB - jmir_v27i1e60077_app2.docx]

Multimedia Appendix 3
Characteristics of the included studies.
[DOCX File , 68 KB - jmir_v27i1e60077_app3.docx]

Multimedia Appendix 4
Coding structure of selected data.
[DOCX File , 23 KB - jmir_v27i1e60077_app4.docx]

References

1. Tripathi M. EHR evolution: policy and legislation forces changing the EHR. J AHIMA 2012 Oct;83(10):24-9; quiz 30. [Medline: 23061349]
2. Doyle-Lindrud S. The evolution of the electronic health record. CJON 2015 Apr 1;19(2):153-154. [doi: 10.1188/15.cjon.153-154]
3. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011 May;4:47-55 [FREE Full text] [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
4. Abramson EL, McGinnis S, Moore J, Kaushal R, HITEC Investigators. A statewide assessment of electronic health record adoption and health information exchange among nursing homes. Health Serv Res 2014 Mar;49(1 Pt 2):361-372 [FREE Full text] [doi: 10.1111/1475-6773.12137] [Medline: 24359612]
5. DesRoches CM, Worzala C, Joshi MS, Kralovec PD, Jha AK. Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Aff (Millwood) 2012 May;31(5):1092-1099 [FREE Full text] [doi: 10.1377/hlthaff.2012.0153] [Medline: 22535503]
6. Kruse CS, DeShazo J, Kim F, Fulton L. Factors associated with adoption of health information technology: a conceptual model based on a systematic review. JMIR Med Inform 2014 May 23;2(1):e9 [FREE Full text] [doi: 10.2196/medinform.3106] [Medline: 25599673]
7. Jones EB, Furukawa MF. Adoption and use of electronic health records among federally qualified health centers grew substantially during 2010-12. Health Aff (Millwood) 2014 Jul;33(7):1254-1261. [doi: 10.1377/hlthaff.2013.1274] [Medline: 25006154]
8. Price C, Kwok ESH, Cheung WJ, Thiruganasambandamoorthy V, Clapham G, Nemnom M, et al. Physician experience with the Epic electronic health record (EHR) system: longitudinal findings from an emergency department implementation. CJEM 2022 Oct 25;24(6):630-635. [doi: 10.1007/s43678-022-00354-5] [Medline: 36006584]
9. Sieck CJ, Pearl N, Bright TJ, Yen P. A qualitative study of physician perspectives on adaptation to electronic health records. BMC Med Inform Decis Mak 2020 Mar 10;20(1):25 [FREE Full text] [doi: 10.1186/s12911-020-1030-6] [Medline: 32039728]
10. Mensah NK, Adzakpah G, Kissi J, Abdulai K, Taylor-Abdulai H, Johnson SB, et al. Health professionals' perceptions of electronic health records system: a mixed method study in Ghana. BMC Med Inform Decis Mak 2024 Oct 16;24(1):254 [FREE Full text] [doi: 10.1186/s12911-024-02672-3] [Medline: 39285423]
11. Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. Int J Med Inform 2014 Dec;83(11):779-796. [doi: 10.1016/j.ijmedinf.2014.06.011] [Medline: 25085286]
12. Ajami S, Bagheri-Tadi T. Barriers for adopting electronic health records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134 [FREE Full text] [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]

XSL•FO
RenderX

13. Boonstra A, Versluis A, Vos JFJ. Implementing electronic health records in hospitals: a systematic literature review. BMC Health Serv Res 2014 Oct 04;14(1):370 [FREE Full text] [doi: 10.1186/1472-6963-14-370] [Medline: 25190184]
14. Ratwani R, Fairbanks T, Savage E, Adams K, Wittie M, Boone E, et al. Mind the gap: a systematic review to identify usability and safety challenges and practices during electronic health record implementation. Appl Clin Inform 2017 Dec 18;07(04):1069-1087. [doi: 10.4338/aci-2016-06-r-0105]
15. Kruse CS, Mileski M, Alaytsev V, Carol E, Williams A. Adoption factors associated with electronic health record among long-term care facilities: a systematic review. BMJ Open 2015 Jan 28;5(1):e006615 [FREE Full text] [doi: 10.1136/bmjopen-2014-006615] [Medline: 25631311]
16. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Med 2011 May 28;9(1):46 [FREE Full text] [doi: 10.1186/1741-7015-9-46] [Medline: 21524315]
17. Ben-Zion R, Pliskin N, Fink L. Critical success factors for adoption of electronic health record systems: literature review and prescriptive analysis. Inf Syst Manag 2014 Oct 28;31(4):296-312. [doi: 10.1080/10580530.2014.958024]
18. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
19. Gesulga JM, Berjame A, Moquiala KS, Galido A. Barriers to electronic health record system implementation and information systems resources: a structured review. Procedia Comp Sci 2017;124:544-551. [doi: 10.1016/j.procs.2017.12.188]
20. Fennelly O, Cunningham C, Grogan L, Cronin H, O'Shea C, Roche M, et al. Successfully implementing a national electronic health record: a rapid umbrella review. Int J Med Inform 2020 Dec;144:104281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104281] [Medline: 33017724]
21. Langley A, Smallman C, Tsoukas H, Van de Ven AH. Process studies of change in organization and management: unveiling temporality, activity, and flow. AMJ 2013 Feb;56(1):1-13. [doi: 10.5465/amj.2013.4001]
22. World Health Organization. Electronic Health Records: Manual for Developing Countries. Geneva, Switzerland: World Health Organization; 2006.
23. E-health and health information technology. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/priorities/key-initiatives/e-health/records [accessed 2024-10-01]
24. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011 Jul;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
25. Bates DW, Ebell M, Gotlieb E, Zapp J, Mullins HC. A proposal for electronic medical records in U.S. primary care. J Am Med Inform Assoc 2003;10(1):1-10 [FREE Full text] [doi: 10.1197/jamia.m1097] [Medline: 12509352]
26. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
27. Felt-Lisk S, Johnson L, Fleming C, Shapiro R, Natzke B. Toward understanding EHR use in small physician practices. Health Care Financ Rev 2010;31(1):11-22 [FREE Full text] [Medline: 20191754]
28. Andriessen JHE. Innovation and implementation. In: Working with Groupware. Computer Supported Cooperative Work. London, UK: Springer; 2003:79-93.
29. Bouwman H, van DWL, van DHB. Information and Communication Technology in Organizations: Adoption, Implementation, Use and Effects. London, UK: Sage; 2005.
30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
31. Ghani MKA, Bali RK, Naguib RN, Marshall IM, Wickramasinghe NS. Electronic health records approaches and challenges: a comparison between Malaysia and four East Asian countries. IJEH 2008;4(1):78. [doi: 10.1504/ijeh.2008.018922]
32. Barrett AK, Stephens KK. The pivotal role of change appropriation in the implementation of health care technology. Manag Commun Q 2016 Dec 18;31(2):163-193. [doi: 10.1177/0893318916682872]
33. Bogaert P, Verschuuren M, Van Oyen H, van Oers H. Identifying common enablers and barriers in European health information systems. Health Policy 2021 Dec;125(12):1517-1526 [FREE Full text] [doi: 10.1016/j.healthpol.2021.09.006] [Medline: 34666917]
34. Boonstra A, van Offenbeek MAG, Vos JFJ. Tension awareness of stakeholders in large technology projects: a duality perspective. Project Manag J 2017 Feb 01;48(1):19-36. [doi: 10.1177/875697281704800102]
35. Cresswell K. Evaluation of implementation of health IT. Stud Health Technol Inform 2016;222:206-219. [Medline: 27198104]
36. Cucciniello M, Guerrazzi C, Nasi G, Ongaro E. Coordination mechanisms for implementing complex innovations in the health care sector. Public Manag Rev 2015 May 13;17(7):1040-1060. [doi: 10.1080/14719037.2015.1029348]
37. Currie W, Finnegan D. The policy-practice nexus of electronic health records adoption in the UK NHS. J Enterp Inf Manag 2011;24:146-170. [doi: 10.1108/17410391111106284]
38. Czerw A, Fronczak A, Witczak K, Juszczyk G. Implementation of electronic health records in Polish outpatient health care clinics - starting point, progress, problems, and forecasts. Ann Agric Environ Med 2016 Jul 02;23(2):329-334 [FREE Full text] [doi: 10.5604/12321966.1203900] [Medline: 27294642]

XSL•FO
RenderX

39. deRiel E, Puttkammer N, Hyppolite N, Diallo J, Wagner S, Honoré JG, et al. Success factors for implementing and sustaining a mature electronic medical record in a low-resource setting: a case study of iSanté in Haiti. Health Policy Plan 2018 Mar 01;33(2):237-246. [doi: 10.1093/heapol/czx171] [Medline: 29253138]
40. Deutsch E, Duftschmid G, Dorda W. Critical areas of national electronic health record programs—is our focus correct? Int J Med Inform 2010 Mar;79(3):211-222. [doi: 10.1016/j.ijmedinf.2009.12.002] [Medline: 20079685]
41. Evio BD, Bonito SR. Formative evaluation of the implementation of eHealth in the Philippines: a qualitative study. Acta Med Philipp 2024;58(12):35-47. [doi: 10.47895/amp.v58i12.9289] [Medline: 39071527]
42. Fragidis LL, Chatzoglou PD. Implementation of a nationwide electronic health record (EHR). IJHCQA 2018 Mar 12;31(2):116-130. [doi: 10.1108/ijhcqa-09-2016-0136]
43. Gans D, Kralewski J, Hammons T, Dowd B. Medical groups' adoption of electronic health records and information systems. Health Aff (Millwood) 2005;24(5):1323-1333 [FREE Full text] [doi: 10.1377/hlthaff.24.5.1323] [Medline: 16162580]
44. Garrety K, McLoughlin I, Dalley A, Wilson R, Yu P. National electronic health record systems as 'wicked projects': the Australian experience. Inf Polity 2016 Dec 22;21(4):367-381. [doi: 10.3233/ip-160389]
45. Greenhalgh T, Stramer K, Bratan T, Byrne E, Mohammad Y, Russell J. Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. BMJ 2008 Oct 23;337:a1786 [FREE Full text] [doi: 10.1136/bmj.a1786] [Medline: 18948344]
46. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Dec 16;341:c5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
47. Helton J, Wade D, Erhardt R. Small clinic electronic health records implementations: an options thinking view. J Manag Policy Pract 2017;18(1):28-43.
48. Hernández-Ávila JE, Palacio-Mejía LS, Lara-Esqueda A, Silvestre E, Agudelo-Botero M, Diana ML, et al. Assessing the process of designing and implementing electronic health records in a statewide public health system: the case of Colima, Mexico. J Am Med Inform Assoc 2013;20(2):238-244 [FREE Full text] [doi: 10.1136/amiajnl-2012-000907] [Medline: 23019239]
49. Jung SY, Lee K, Lee H, Hwang H. Barriers and facilitators to implementation of nationwide electronic health records in the Russian Far East: a qualitative analysis. Int J Med Inform 2020 Dec;143:104244 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104244] [Medline: 32942224]
50. Kang'a SG, Muthee VM, Liku N, Too D, Puttkammer N. People, process and technology: strategies for assuring sustainable implementation of EMRs at public-sector health facilities in Kenya. AMIA Annu Symp Proc 2016;2016:677-685 [FREE Full text] [Medline: 28269864]
51. Klecun E, Zhou Y, Kankanhalli A, Wee YH, Hibberd R. The dynamics of institutional pressures and stakeholder behavior in national electronic health record implementations: a tale of two countries. J Inf Technol 2019 Mar 27;34(4):292-332. [doi: 10.1177/0268396218822478]
52. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. BMJ 2011 Oct 17;343(oct17 1):d6054-d6054 [FREE Full text] [doi: 10.1136/bmj.d6054] [Medline: 22006942]
53. Tobler N, Colvin J, Rawlins NW. Longitudinal analysis and coping model of user adaptation. J Comp Inf Syst 2016 Jul 22;57(2):97-105. [doi: 10.1080/08874417.2016.1183415]
54. Vadillo PC, Rojo ES, Garces A, Checton MG. Maximizing healthcare professionals' use of new computer technologies in a small, urban hospital's critical care unit. J Healthc Manag 2016;61(5):352-362. [doi: 10.1097/00115514-201609000-00009]
55. Cacciatore F, Bitonti A, Sgueo G. Interest groups and the implementation of electronic health records in the Italian NRRP, between policy and politics. Contemp Ital Politics 2023 Dec 04;16(1):21-38. [doi: 10.1080/23248823.2023.2288976]
56. Maier CD, Frandsen F, Johansen W. Understanding the arena of smoldering crises: a longitudinal study of discursive struggles after implementing a new IT health care platform. JCOM 2022 Sep 23;27(3):329-344. [doi: 10.1108/jcom-12-2021-0136]
57. Standing C, Cripps H. Critical success factors in the implementation of electronic health records: a two-case comparison. Syst Res 2013 Aug 30;32(1):75-85. [doi: 10.1002/sres.2209]
58. Robertson A, Cresswell K, Takian A, Petrakaki D, Crowe S, Cornford T, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. BMJ 2010 Oct 01;341(sep01 3):c4564-c4564 [FREE Full text] [doi: 10.1136/bmj.c4564] [Medline: 20813822]
59. Crowley K, Mishra A, Cruz-Cano R, Gold R, Kleinman D, Agarwal R. Electronic health record implementation findings at a large, suburban health and human services department. J Public Health Manag Pract 2019;25(1):E11-E16 [FREE Full text] [doi: 10.1097/PHH.0000000000000768] [Medline: 29324567]
60. Deokar AV, Sarnikar S. Understanding process change management in electronic health record implementations. Inf Syst e-Bus Manag 2014 Jul 1;14(4):733-766. [doi: 10.1007/s10257-014-0250-7]
61. Faiella A, Casper KA, Bible L, Seifert J. Implementation and use of an electronic health record in a charitable community pharmacy. J Am Pharm Assoc (2003) 2019;59(2S):S110-S117. [doi: 10.1016/j.japh.2018.12.004] [Medline: 30733152]

XSL•FO
RenderX

62. Pohlmann S, Kunz A, Ose D, Winkler EC, Brandner A, Poss-Doering R, et al. Digitalizing health services by implementing a personal electronic health record in Germany: qualitative analysis of fundamental prerequisites from the perspective of selected experts. J Med Internet Res 2020 Jan 29;22(1):e15102 [FREE Full text] [doi: 10.2196/15102] [Medline: 32012060]
63. Rau E, Tischendorf T, Mitzscherlich B. Implementation of the electronic health record in the German healthcare system: an assessment of the current status and future development perspectives considering the potentials of health data utilisation by representatives of different stakeholder groups. Front Health Serv 2024 May 10;4:1370759 [FREE Full text] [doi: 10.3389/frhs.2024.1370759] [Medline: 38800500]
64. Strong D, Volkoff O, Johnson S, Pelletier L, Tulu B, Bar-On I, et al. A theory of organization-EHR affordance actualization. JAIS 2014 Feb;15(2):53-85. [doi: 10.17705/1jais.00353]
65. van Offenbeek MA, Vos JF, Boonstra A. Understanding variation in subunit adoption of electronic health records: facilitating and constraining configurations of critical dependencies. Eur J Inf Syst 2023 Jun 25;33(2):221-243. [doi: 10.1080/0960085x.2023.2225786]
66. Brokel JM, Harrison MI. Redesigning care processes using an electronic health record: a system's experience. Jt Comm J Qual Patient Saf 2009 Mar;35(2):82-92. [doi: 10.1016/s1553-7250(09)35011-4] [Medline: 19241728]
67. Mbwambo E, Mandari H. Acceptance of interoperable electronic health record (EHRs) systems: a Tanzanian e-health perspective. J Int Technol Inf Manag 2023 Jan 01;32(1):96-121. [doi: 10.58729/1941-6679.1574]
68. van Offenbeek MA, Vos JF. An integrative framework for managing project issues across stakeholder groups. Int J Project Manag 2016 Jan;34(1):44-57. [doi: 10.1016/j.ijproman.2015.09.006]
69. Carayon P, Smith P, Hundt AS, Kuruchittham V, Li Q. Implementation of an electronic health records system in a small clinic: the viewpoint of clinic staff. Behav Inf Technol 2009 Jan;28(1):5-20. [doi: 10.1080/01449290701628178]
70. Collins S, Alexander D, Moss J. Nursing domain of CI governance: recommendations for health IT adoption and optimization. J Am Med Inform Assoc 2015 May;22(3):697-706. [doi: 10.1093/jamia/ocu001] [Medline: 25670752]
71. Craven CK, Sievert MC, Hicks LL, Alexander GL, Hearne LB, Holmes JH. CAH to CAH. Appl Clin Inform 2017 Dec 20;05(01):92-117. [doi: 10.4338/aci-2013-08-ra-0066]
72. Ford EW, Menachemi N, Huerta TR, Yu F. Hospital IT adoption strategies associated with implementation success: implications for achieving meaningful use. J Healthc Manag 2010;55(3):175-88; discussion 188. [Medline: 20565034]
73. Hertzum M, Ellingsen G. The implementation of an electronic health record: comparing preparations for Epic in Norway with experiences from the UK and Denmark. Int J Med Inform 2019 Oct;129:312-317. [doi: 10.1016/j.ijmedinf.2019.06.026] [Medline: 31445272]
74. Hertzum M, Ellingsen G, Melby L. Drivers of expectations: why are Norwegian general practitioners skeptical of a prospective electronic health record? Health Inform J 2021 Jan 13;27(1):1460458220987298 [FREE Full text] [doi: 10.1177/1460458220987298] [Medline: 33438506]
75. Hertzum M, Ellingsen G, Cajander Å. Implementing large-scale electronic health records: experiences from implementations of Epic in Denmark and Finland. Int J Med Inform 2022 Dec;167:104868. [doi: 10.1016/j.ijmedinf.2022.104868] [Medline: 36194994]
76. McAlearney AS, Sieck C, Hefner J, Robbins J, Huerta TR. Facilitating ambulatory electronic health record system implementation: evidence from a qualitative study. Biomed Res Int 2013;2013:629574-629579 [FREE Full text] [doi: 10.1155/2013/629574] [Medline: 24228257]
77. Noblin A, Cortelyou-Ward K, Cantiello J, Breyer T, Oliveira L, Dangiolo M, et al. EHR implementation in a new clinic: a case study of clinician perceptions. J Med Syst 2013 Aug;37(4):9955. [doi: 10.1007/s10916-013-9955-2] [Medline: 23780429]
78. Scott JT, Rundall TG, Vogt TM, Hsu J. Kaiser Permanente's experience of implementing an electronic medical record: a qualitative study. BMJ 2005 Dec 03;331(7528):1313-1316 [FREE Full text] [doi: 10.1136/bmj.38638.497477.68] [Medline: 16269467]
79. Sidek YH, Martins JT. Perceived critical success factors of electronic health record system implementation in a dental clinic context: an organisational management perspective. Int J Med Inform 2017 Dec;107:88-100 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.08.007] [Medline: 29029696]
80. Weston C, Wells-Beede E, Salazar A, Poston D, Brown S, Hare M, et al. Patient-centered care through nurse practitioner-led integrated behavioral health: a case study. Public Health Rep 2023 May 25;138(1_suppl):36S-41S [FREE Full text] [doi: 10.1177/00333549231152192] [Medline: 37226946]
81. Boswell RA. Implementing electronic health records: implications for HR professionals. Strateg HR Rev 2013;12(5):262-268. [doi: 10.1108/shr-08-2012-0010]
82. Dansky KH, Gamm LD, Vasey JJ, Barsukiewicz CK. Electronic medical records: are physicians ready? J Healthc Manag 1999;44(6):440-54; discussion 454. [Medline: 10662431]
83. Gross AH, Leib RK, Tonachel A, Tonachel R, Bowers DM, Burnard RA, et al. Teamwork and electronic health record implementation: a case study of preserving effective communication and mutual trust in a changing environment. J Oncol Pract 2016 Dec;12(11):1075-1083. [doi: 10.1200/JOP.2016.013649] [Medline: 27601513]

XSL•FO
RenderX

84. Kiepek W, Sengstack PP. An evaluation of system end-user support during implementation of an electronic health record using the model for improvement framework. Appl Clin Inform 2019 Oct 18;10(5):964-971 [FREE Full text] [doi: 10.1055/s-0039-3402450] [Medline: 31853937]
85. Knight AW, Szucs C, Dhillon M, Lembke T, Mitchell C. The eCollaborative: using a quality improvement collaborative to implement the National eHealth Record System in Australian primary care practices. Int J Qual Health Care 2014 Aug 12;26(4):411-417 [FREE Full text] [doi: 10.1093/intqhc/mzu059] [Medline: 24925685]
86. Martin M, Petros M, Welter C. Exploratory case study of barriers and facilitators associated with the pilot implementation of a new electronic healthcare record in the military. Mil Med 2022 Mar 28;187(3-4):e486-e492 [FREE Full text] [doi: 10.1093/milmed/usab053] [Medline: 33580676]
87. McAlearney AS, Hefner JL, Sieck C, Rizer M, Huerta TR. Evidence-based management of ambulatory electronic health record system implementation: an assessment of conceptual support and qualitative evidence. Int J Med Inform 2014 Jul;83(7):484-494. [doi: 10.1016/j.ijmedinf.2014.04.002] [Medline: 24862893]
88. Pearce C, Bartlett J, Mcleod A, Eustace P, Amos R, Shearer M. Effectiveness of local support for the adoption of a national programme--a descriptive study. Inform Prim Care 2014;21(4):171-178 [FREE Full text] [doi: 10.14236/jhi.v21i4.70] [Medline: 25479347]
89. Takian A, Sheikh A, Barber N. Organizational learning in the implementation and adoption of national electronic health records: case studies of two hospitals participating in the National Programme for Information Technology in England. Health Inform J 2014 Oct 02;20(3):199-212 [FREE Full text] [doi: 10.1177/1460458213493196] [Medline: 25183608]
90. Threatt T, Pirtle C, Dzwonkowski J, Johnson K. Using a custom mobile application for change management in an electronic health record implementation. JAMIA Open 2020 May;3(1):37-43 [FREE Full text] [doi: 10.1093/jamiaopen/ooz048] [Medline: 32607486]
91. Yung A. Adoption of Electronic Health Record System in Community-Based Physiotherapy Clinics: A Pilot Case Study. Stud Health Technol Inform 2017;234:395-400. [Medline: 28186074]
92. Boonstra A, van Offenbeek MA. Shaping a buyer's software selection process through tendering legislation. Inf Syst J 2017 Dec 14;28(5):905-928. [doi: 10.1111/isj.12174]
93. Alassia L, Palermo C, Recondo F, Giussi M, Stieben A, Baum A, et al. Managing user needs during the EHR implementation in Buenos Aires City: the helpdesk role. Precis Healthc Inform 2017:835-839. [doi: 10.3233/978-1-61499-830-3-835]
94. Banas CA, Erskine AR, Sun S, Retchin SM. Phased implementation of electronic health records through an office of clinical transformation. J Am Med Inform Assoc 2011 Sep 01;18(5):721-725 [FREE Full text] [doi: 10.1136/amiajnl-2011-000165] [Medline: 21659444]
95. Boswell RA. A physician group's movement toward electronic health records: a case study using the transtheoretical model for organizational change. Consult Psychol J: Pract Res 2011 Jun;63(2):138-148. [doi: 10.1037/a0024319]
96. Takian A, Sheikh A, Barber N. We are bitter, but we are better off: case study of the implementation of an electronic health record system into a mental health hospital in England. BMC Health Serv Res 2012 Dec 31;12(1):484 [FREE Full text] [doi: 10.1186/1472-6963-12-484] [Medline: 23272770]
97. Zandieh SO, Yoon-Flannery K, Kuperman GJ, Langsam DJ, Hyman D, Kaushal R. Challenges to EHR implementation in electronic- versus paper-based office practices. J Gen Intern Med 2008 Jul 28;23(6):755-761 [FREE Full text] [doi: 10.1007/s11606-008-0573-5] [Medline: 18369679]
98. Hariyati RTS, Hamid AY, Eryando T, Hasibuan ZA. Usability and satisfaction of using electronic nursing documentation, lesson-learned from new system implementation at a hospital in Indonesia. Int J Healthc Manag 2018 Jul 31;13(1):45-52. [doi: 10.1080/20479700.2018.1504387]
99. Calvo-Amodio J, Patterson PE, Smith ML, Burns JR. Application of transition-phase management model for an electronic health record system implementation: a case study. Eng Manag J 2015 Aug 27;27(3):131-140. [doi: 10.1080/10429247.2015.1064662]
100. Heponiemi T, Gluschkoff K, Vehko T, Kaihlanen A, Saranto K, Nissinen S, et al. Electronic health record implementations and insufficient training endanger nurses' well-being: cross-sectional survey study. J Med Internet Res 2021 Dec 23;23(12):e27096 [FREE Full text] [doi: 10.2196/27096] [Medline: 34941546]
101. Pine KH, Mazmanian M. Artful and contorted coordinating: the ramifications of imposing formal logics of task jurisdiction on situated practice. AMJ 2017 Apr;60(2):720-742. [doi: 10.5465/amj.2014.0315]
102. Ser G, Robertson A, Sheikh A. A qualitative exploration of workarounds related to the implementation of national electronic health records in early adopter mental health hospitals. PLoS One 2014;9(1):e77669 [FREE Full text] [doi: 10.1371/journal.pone.0077669] [Medline: 24454678]
103. Sheehan OM, Greene RA, McKernan J, Murphy B, Cahill C, Cleary B, et al. Introduction of a single electronic health record for maternity units in Ireland: outline of the experiences of the project management team. JMIR Form Res 2023 May 12;7:e38938. [doi: 10.2196/38938]
104. Shield RR, Goldman RE, Anthony DA, Wang N, Doyle RJ, Borkan J. Gradual electronic health record implementation: new insights on physician and patient adaptation. Ann Fam Med 2010 Jul 19;8(4):316-326 [FREE Full text] [doi: 10.1370/afm.1136] [Medline: 20644186]

XSL•FO
RenderX

105. Snowden A, Kolb H. Two years of unintended consequences: introducing an electronic health record system in a hospice in Scotland. J Clin Nurs 2017 May 21;26(9-10):1414-1427 [FREE Full text] [doi: 10.1111/jocn.13576] [Medline: 27602553]
106. Trocin C, Lee G, Bernardi R, Sarker S. How do unintended consequences emerge from EHR implementation? An affordance perspective. Inf Syst J 2024 Apr 23;35(1):39-70. [doi: 10.1111/isj.12526]
107. McGinn CA, Gagnon M, Shaw N, Sicotte C, Mathieu L, Leduc Y, et al. Users' perspectives of key factors to implementing electronic health records in Canada: a Delphi study. BMC Med Inform Decis Mak 2012 Oct 11;12:105 [FREE Full text] [doi: 10.1186/1472-6947-12-105] [Medline: 22967231]
108. Palvia P, Jacks T, Brown W. Critical issues in EHR implementation: provider and vendor perspectives. Commun Assoc Inf Syst 2015:36 [FREE Full text] [doi: 10.17705/1cais.03636]
109. Stanczyk NE, Crutzen R, Sewuster N, Schotanus E, Mulders M, Cremers HP. Differences in Sociocognitive Beliefs between Involved and Noninvolved Employees during the Implementation of an Electronic Health Record System. Perspect Health Inf Manag 2017;14(Spring):1c [FREE Full text] [Medline: 28566986]
110. Marca G, Perez A, Blanco-Garcia MG, Miravalles E, Soley P, Ortiga B. The use of electronic health records in Spanish hospitals. Health Inf Manag 2014 Oct 01;43(3):37-44. [doi: 10.1177/183335831404300305] [Medline: 27009795]
111. Bove R, Bruce CA, Lunders CK, Pearce JR, Liu J, Schleimer E, et al. Electronic health record technology designed for the clinical encounter. Neur Clin Pract 2021 Aug 06;11(4):318-326. [doi: 10.1212/cpj.0000000000000986]
112. Fleming NS, Culler SD, McCorkle R, Becker ER, Ballard DJ. The financial and nonfinancial costs of implementing electronic health records in primary care practices. Health Aff (Millwood) 2011 Mar;30(3):481-489 [FREE Full text] [doi: 10.1377/hlthaff.2010.0768] [Medline: 21383367]
113. Gabriel MH, Jones EB, Samy L, King J. Progress and challenges: implementation and use of health information technology among critical-access hospitals. Health Aff (Millwood) 2014 Jul;33(7):1262-1270. [doi: 10.1377/hlthaff.2014.0279] [Medline: 25006155]

Abbreviations

EHR: electronic health record
EMR: electronic medical record
PHR: personal health record
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by G Eysenbach; submitted 30.04.24; peer-reviewed by S Lloyd, F Lau; comments to author 19.07.24; revised version received 18.10.24; accepted 07.12.24; published 03.03.25.

Please cite as:
Finnegan H, Mountford N
25 Years of Electronic Health Record Implementation Processes: Scoping Review
J Med Internet Res 2025;27:e60077
URL: https://www.jmir.org/2025/1/e60077
doi:10.2196/60077
PMID:40053758

©Harriet Finnegan, Nicola Mountford. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Generative AI Models in Time-Varying Biomedical Data: Scoping Review

Rosemary He[1,2], MCS; Varuni Sarwal[1,2], MCS; Xinru Qiu[3], PhD; Yongwen Zhuang[4], PhD; Le Zhang[5], MCS; Yue Liu[6], MCS; Jeffrey Chiang[2,7], PhD

[1]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, United States

[2]Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[3]Division of Biomedical Sciences, School of Medicine, University of California Riverside, Riverside, CA, United States

[4]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States

[5]Institute for Integrative Genome Biology, University of California Riverside, Riverside, CA, United States

[6]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, United States

[7]Department of Neurosurgery, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Jeffrey Chiang, PhD
Department of Neurosurgery
David Geffen School of Medicine
University of California, Los Angeles
300 Stein Plaza, Suite 560
Los Angeles, CA, 90095
United States
Phone: 1 310 825 5111
Email: njchiang@g.ucla.edu

Abstract

Background: Trajectory modeling is a long-standing challenge in the application of computational methods to health care. In the age of big data, traditional statistical and machine learning methods do not achieve satisfactory results as they often fail to capture the complex underlying distributions of multimodal health data and long-term dependencies throughout medical histories. Recent advances in generative artificial intelligence (AI) have provided powerful tools to represent complex distributions and patterns with minimal underlying assumptions, with major impact in fields such as finance and environmental sciences, prompting researchers to apply these methods for disease modeling in health care.

Objective: While AI methods have proven powerful, their application in clinical practice remains limited due to their highly complex nature. The proliferation of AI algorithms also poses a significant challenge for nondevelopers to track and incorporate these advances into clinical research and application. In this paper, we introduce basic concepts in generative AI and discuss current algorithms and how they can be applied to health care for practitioners with little background in computer science.

Methods: We surveyed peer-reviewed papers on generative AI models with specific applications to time-series health data. Our search included single- and multimodal generative AI models that operated over structured and unstructured data, physiological waveforms, medical imaging, and multi-omics data. We introduce current generative AI methods, review their applications, and discuss their limitations and future directions in each data modality.

Results: We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines and reviewed 155 articles on generative AI applications to time-series health care data across modalities. Furthermore, we offer a systematic framework for clinicians to easily identify suitable AI methods for their data and task at hand.

Conclusions: We reviewed and critiqued existing applications of generative AI to time-series health data with the aim of bridging the gap between computational methods and clinical application. We also identified the shortcomings of existing approaches and highlighted recent advances in generative AI that represent promising directions for health care modeling.

(J Med Internet Res 2025;27:e59792) doi:10.2196/59792

XSL•FO
RenderX

KEYWORDS

generative artificial intelligence; artificial intelligence; time series; electronic health records; electronic medical records; systematic reviews; disease trajectory; machine learning; algorithms; forecasting

Introduction

Background

Generative artificial intelligence (GenAI) is a family of artificial intelligence (AI) models that can generate synthetic content ranging across text, images, audio, and video. Well-known GenAI tools such as ChatGPT [1] and DALL-E [1] have revolutionized the way in which people view AI [2]. On the other hand, data analysis in health care has entered a new era as large amounts of electronic health records (EHRs) are generated daily. Among these data, biomedical time series have been of particular interest to researchers in tasks including tracking patient trajectories [3] and treatment estimations [4]. Compared to cross-sectional data, which represent 1 time point, time series provide more information on development over time. There are 2 main types of time-series data: real-time and longitudinal data. While real-time data usually span a shorter period (minutes or days) and have more consistent time-point entries, longitudinal data are recorded over the magnitude of months or years with various sampling frequencies and number of entries. Compared to traditional time-series data in finance and climate science, biomedical time series face unique challenges in modeling, including missingness and sampling rate [5]. Therefore, traditional methods such as foundation models (FMs) that work well on high-frequency time series are not guaranteed to work well on biomedical data. In the age of big data, GenAI models have begun to outperform traditional statistical and machine learning (ML) methods in various tasks [6]. On the other hand, GenAI for health care faces unique challenges, including data privacy, transparency of the decision process, and other ethical issues that must be addressed. Nonetheless, these tools have the potential to improve disease modeling accuracy, tailor treatment plans across a heterogeneous population, and assist clinicians in making clinical decisions.

In this paper, we aim to provide a comprehensive review of GenAI applications to biomedical time-series data and a technical overview for clinicians who wish to work with AI tools but are unsure of where to begin. We survey GenAI applications to time-series health care data and its ability to improve patient care and the diagnosis process while considering the challenges and showcasing real-world uses that highlight its importance in the field. For the selection process, we followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews; Multimedia Appendix 1) guidelines and searched in PubMed and Google Scholar for conference and workshop papers using the following keywords: “generative AI (GenAI),” “health care,” “time series,” “longitudinal,” “electronic health records (EHR),” “electronic medical records (EMR),” and “genetics.” From the initial search, we applied the following inclusion criteria—being peer reviewed, published after 2000, and focusing on time-series or longitudinal data—ultimately yielding 155 studies as of April 2024. The detailed selection process is described in the Related Works section.

We organized the review by data modality, including structured and unstructured text, imaging, waveform, and multi-omics data, and illustrate possible data sources in each modality for GenAI application in Figure 1. For each modality, we discuss preprocessing, applications, model selection, and challenges and future directions. In the Model Selection sections, we help guide how to choose the appropriate models for this type of data for clinicians who are interested in incorporating GenAI into their practice. We organize the workflow for choosing an appropriate model for the data and computational resources at hand. We have organized the paper as follows: (1) overview of the algorithms and techniques introduced in the review; (2) structured data; (3) unstructured data; (4) medical imaging; (5) physiological waveforms; (6) genetics and multi-omics data; and (7) ethical considerations, challenges, and future directions.

XSL•FO
RenderX

Figure 1. Health care data sources in different modalities for generative artificial intelligence (AI) application. ECG: electrocardiogram; EEG: electroencephalogram.

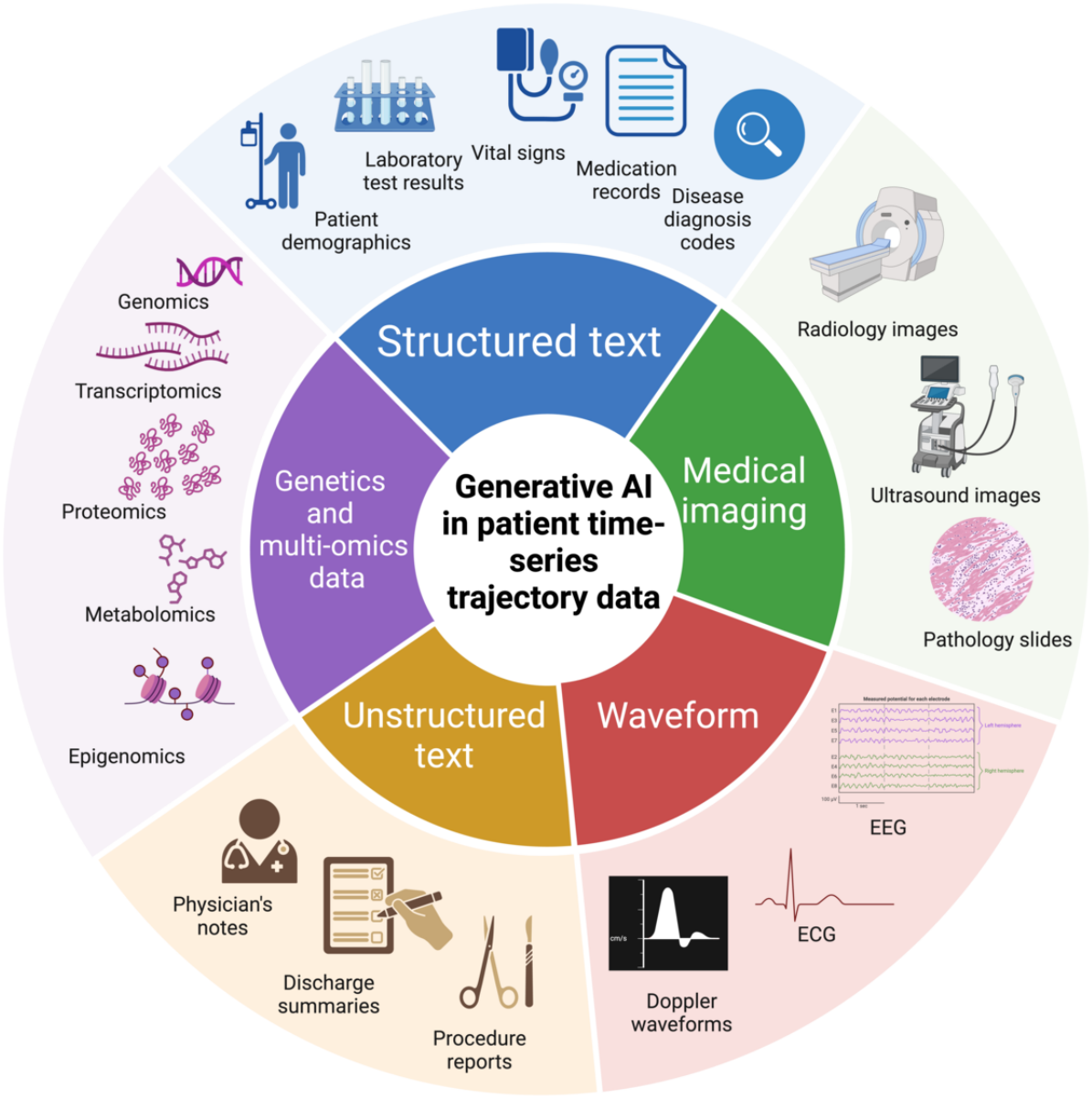

Related Works

Our electronic search retrieved 20,978 study records: 20,003 (95.35%) from Google Scholar and 975 (4.65%) from PubMed, as shown in Figure 2. After title and abstract screening and removing duplicates, 3.24% (680/20,978) of these articles were retained for full-text review. Of these 680 articles, 533 (78.4%) were excluded, leaving 147 (21.6%) studies after the primary screening. A total of 8 more articles were added through secondary screening, resulting in a total of 155 articles included in our scoping review.

Our search for existing reviews on GenAI for disease trajectory modeling returned 39 entries broadly covering the use of AI in health care (Multimedia Appendix 2) in Google Scholar, Scopus, and PubMed. From a review of these works, although we observed an increasing number of AI-related review papers in the biomedical setting in recent years (Multimedia Appendix 3), there remained a gap in reviewing GenAI approaches working with time-series data specifically, prompting the need for our paper.

XSL•FO
RenderX

Figure 2. Study selection.

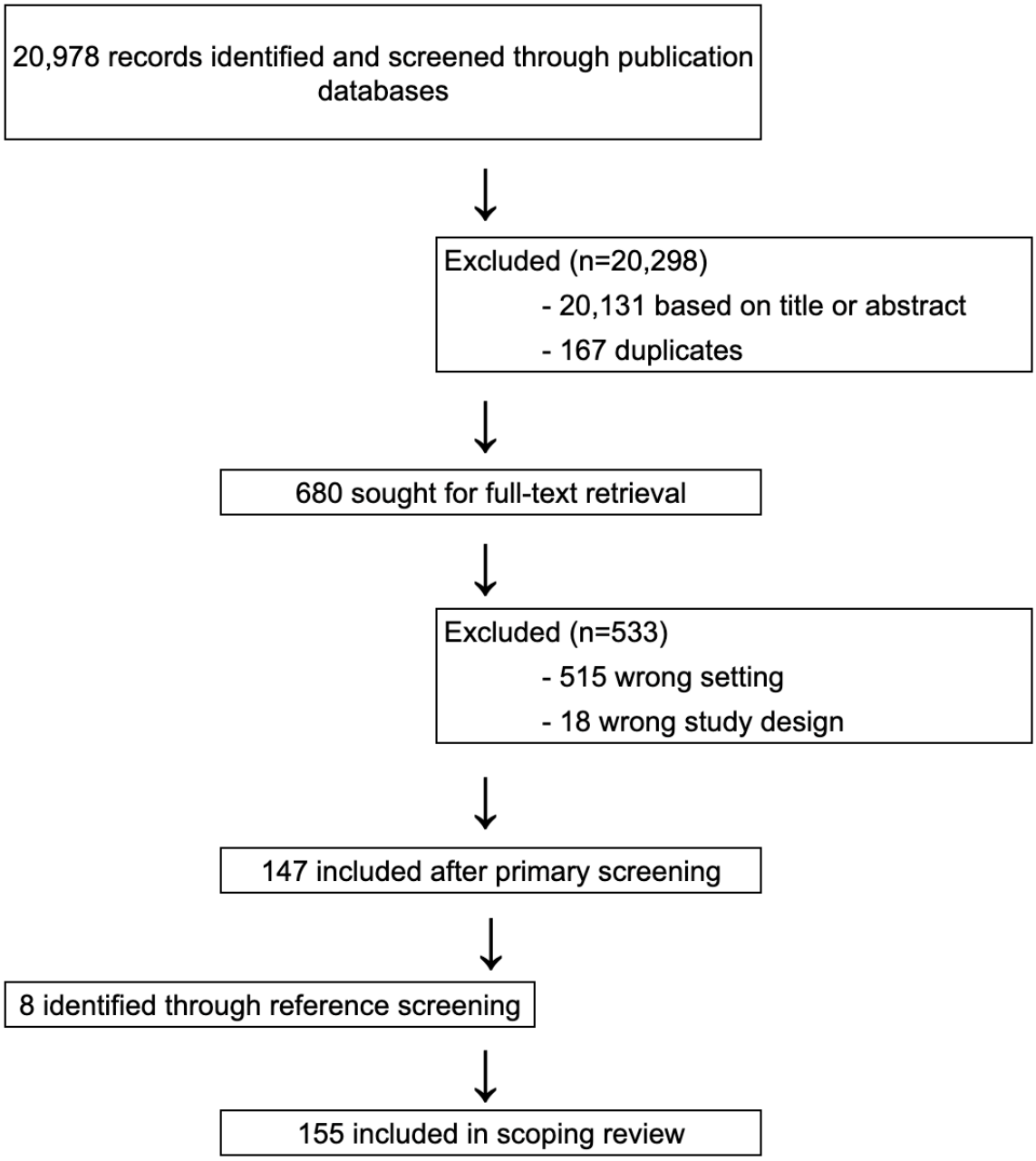

Algorithm and Technique Overview

Overview

Despite a shared nomenclature, it is important to differentiate between modern AI and the field of AI. The general study of AI encompasses the class of ML methods, which includes the subset of deep learning (DL) methods (also thought of as modern AI). Previous ML approaches have been mostly *discriminative* such that, given a set of input predictors, they learn a strategy (decision boundary) to fit or separate data points. On the other hand, DL models can be discriminative or generative based on the choice of the objective function. Therefore, they are not inherently generative or discriminative and are the building blocks of modern AI models, although they have typically been developed and demonstrated in the discriminative setting. Finally, GenAI methods belong to the class of *generative* ML methods and capture the underlying data generation process. By learning the distribution domain of the data, these models can synthesize hypothetical data points that are statistically indistinguishable from the originals. For discriminative tasks, GenAI models are often combined with downstream models, which can be simply rule based or more complex. In this section, we introduce a set of traditional, ML, DL, and GenAI models that are referenced later in the paper and list existing models used in time-series forecasting in Figure 3, as well as showing a timeline for when these models were first introduced in Figure 4 and a comparison among these methods using common metrics in Table 1.

XSL•FO
RenderX

Figure 3. Existing models for time-series forecasting. ARIMA: autoregressive integrated moving average; ARMA: autoregressive moving average; CNN: convolutional neural network; DANN: deleterious annotation of genetic variants using neural networks; DDPM: denoising diffusion probabilistic model; ETS: exponential smoothing; EVE: Evolutionary Model of Variant Effect; GAN: generative adversarial network; GenAI: generative artificial intelligence; Gene-SGAN: gene-guided weakly supervised clustering via GANs; GRU: gated recurrent unit; LSTM: long short-term memory; M-CAP: Mendelian Clinically Applicable Pathogenicity; MBP: masked bidirectional prediction; PAGA: partition-based graph abstraction; PRESCIENT: Potential Energy Underlying Single-Cell Gradients; RNN: recurrent neural network; SCUBA: single-cell clustering using bifurcation analysis; T-GAN-D: Trained GAN Discriminator; TIGON: Trajectory Inference With Growth via Optimal Transport and Neural Network; VAE: variational autoencoder.

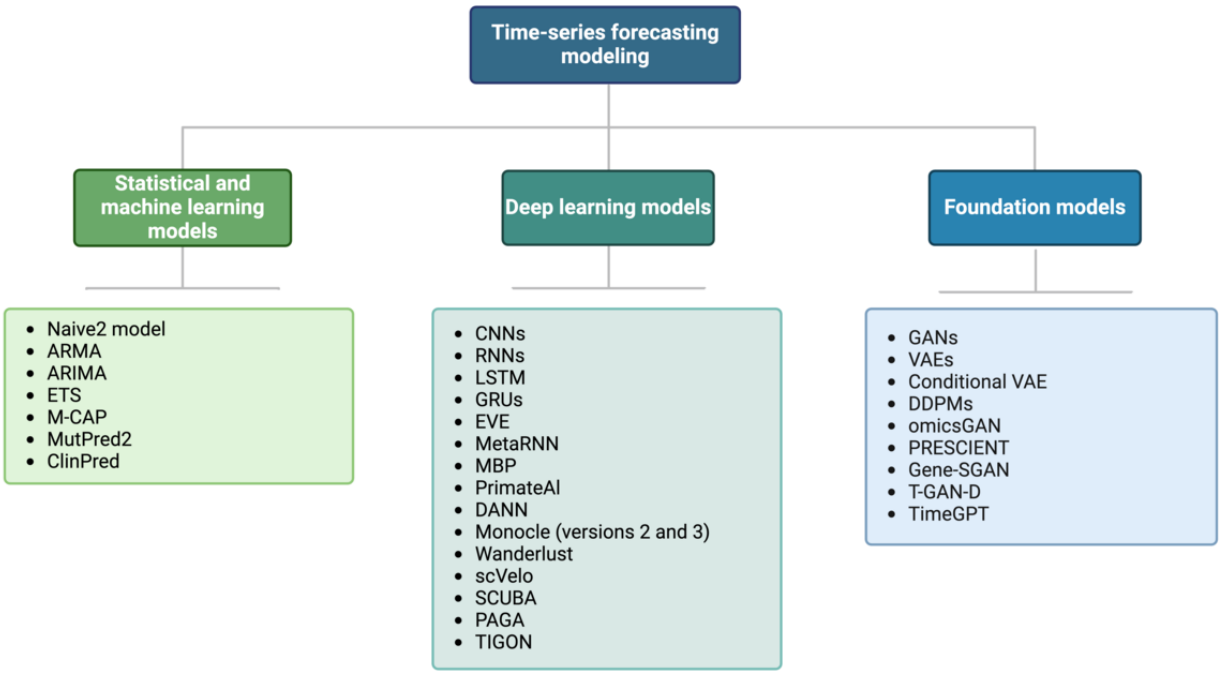

Figure 4. Milestones in time-series forecasting models. BERT: Bidirectional Encoder Representations From Transformers; CNN: convolutional neural network; DDPM: denoising diffusion probabilistic model; GAN: generative adversarial network; GRU: gated recurrent unit; LSTM: long short-term memory; MA: moving average; RNN: recurrent neural network; VAE: variational autoencoder.

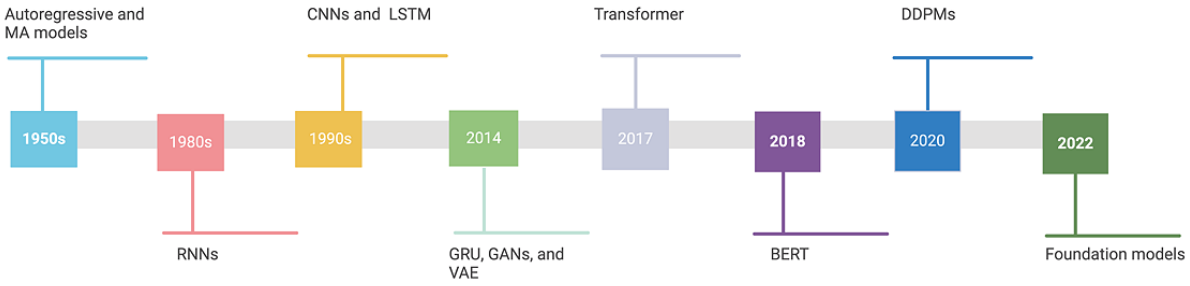

XSL•FO
RenderX

Table 1. Metric comparison among methods for biomedical time-series prediction (1=lowest; 5=highest).

Model	Computational cost	Interpretability	Model size	Data requirement	Accuracy
Benchmark models	1	5	1	1	2
CNN[a]	2	3	2	3	3
RNN[b]	2	3	2	3	3
Transformer	3	4	3	3	5
GAN[c]	4	2	4	4	5
VAE[d]	2	3	3	4	4
Diffusion	2	1	3	4	5
NLP[e]	3	2	3	3	4
Foundation models	5	1	5	5	3

[a]CNN: convolutional neural network.

[b]RNN: recurrent neural network.

[c]GAN: generative adversarial network.

[d]VAE: variational autoencoder.

[e]NLP: natural language processing.

Traditional Statistical Models

In this section, we introduce several statistical models commonly used as benchmarks in modeling time-series data. One of the most basic yet effective methods is the Naive2 model, which makes a prediction by using the value from the most recent observation or from a similar period in the past. In the class of autoregressive methods, the autoregressive moving average model [7] predicts the outcome value at a given time point using both a linear combination of multiple past observations and a linear combination of past error terms in the regression model. As an extension of the autoregressive moving average, the autoregressive integrated moving average (ARIMA) model [8] accounts for seasonal or periodical changes in the data. One of the most widely used methods is the Prophet model by Meta [9], which uses an additive model that accounts for yearly, weekly, and daily seasonality and holidays. Finally, the exponential smoothing (ETS) model [10] uses a weighted combination of multiple past observations in which the decay of the weights is exponentially distributed. While these methods have limited performance with highly complex data, they remain a popular choice due to their simplicity and transparency in the decision-making process.

DL Models

DL models are the combination of an architecture (ie, the structure) and an objective (ie, the loss function). In this section, we introduce the canonical DL architectures that are the foundation of different GenAI settings but are not GenAI models themselves.

Convolutional Neural Networks

Convolutional neural networks (CNNs) [11] are a type of DL neural network architecture that is particularly well suited for processing and analyzing image data. CNNs are good at recognizing patterns in small areas and work well for short-term connections due to their ability to perform convolutions step by step over time [12]. CNNs can be used for univariate time-series forecasting, in which the trend can be viewed as a 2D graph. In addition, researchers have explored different CNN architectures for time-series forecasting, such as temporal convolutional networks [13] that are tailored for sequence-to-sequence modeling tasks, leveraging the strengths of CNNs while addressing the challenges of working with sequential data.

Recurrent Neural Networks

Recurrent neural networks (RNNs) [14] are one of the earliest DL frameworks that are designed to capture sequential dependencies and patterns across time steps. RNNs have an internal or hidden state that captures information from previous time steps and influences the processing of subsequent inputs. During training, sequential data are fed into the network one step at a time. The network processes each input along with its corresponding hidden state, updating the hidden state based on the current input and the previous hidden state. This process is repeated for each time step in the sequence, allowing the network to capture temporal dependencies and learn patterns in the data. The architecture of RNNs consists of multiple recurrent layers, each containing recurrent units. There are 2 main types of units: long short-term memory (LSTM) cells and gated recurrent units (GRUs). These units are responsible for capturing temporal dependencies and encoding information from previous time steps. Due to their training on sequential data and the leveraging of their recurrent structure, RNNs have performed strongly against traditional DL networks in modeling complex patterns in time-series data [15,16]. However, RNNs can be time-consuming as inputs are processed sequentially and do not model long-range dependencies well as the information diminishes over time.

Transformers

Transformers [17] are the foundation of many state-of-the-art generative models, including ChatGPT. Unlike RNNs, which

XSL•FO
RenderX

process data sequentially and can be computationally expensive, transformers rely on a mechanism called self-attention [17] to capture dependencies simultaneously and over long ranges. They consist of an encoder-decoder architecture in which the encoder processes the input sequence and the decoder generates the output sequence. During training, the input is embedded into a high-dimensional space, and the encoder applies a self-attention mechanism to capture relationships between different parts of the input sequence in the embedding space. Self-attention allows each position in the sequence to attend to all other positions, enabling the model to learn contextual representations efficiently. These representations are then passed through feedforward neural networks within each layer to further process the information. The decoder, on the other hand, predicts the output sequence step by step based on the encoder's contextualized representations and previous outputs. Transformer models typically include multiple layers of encoder and decoder blocks, each containing self-attention mechanisms and feedforward neural networks. In the age of large data, transformers are one of the most powerful models to capture complex relationships in sequential data. One example is in the field of natural language processing (NLP), where Bidirectional Encoder Representations From Transformers (BERT) [18] and generative pretrained transformer architecture have become state of the art, achieving the best performance across many NLP benchmarks [19,20].

GenAI Models

Generative Adversarial Networks

A generative adversarial network (GAN) [21] is a generative vision model that consists of 2 parts: the generator that creates fake data and the discriminator that critiques them. The generator's goal is to produce realistic outputs from random noise that resemble realistic images. On the other hand, the discriminator's role is to differentiate between real data and the data generated by the generator. Through this back-and-forth process of creation and critique, GANs improve over time, producing increasingly realistic results. GANs are one of the earliest image generation models that have shown great success in generating synthetic images [22] and have been used extensively for the generation of new datasets and domain transfer in medical imaging [23,24]. However, training GANs can be challenging due to issues such as the limited type of data produced and instability during training.

Variational Autoencoders

Autoencoders are a class of neural network that learn a low-dimensional representation of high-dimensional structured data [25]. Autoencoders consist of 2 parts: an encoder that projects high-dimensional data into a latent space with lower dimensions and a decoder that learns to map a point in the latent space back to its high-dimensional representation. However, the latent distribution of a vanilla autoencoder is unknown, making inference difficult and prompting the need for variational autoencoders (VAEs). VAE models [26] make further assumptions about the sample generation process that allow us to model an approximation of the distribution of the latent space, usually a Gaussian distribution with a mean and variance estimated by the model. Unlike other image generation models, VAEs provide a latent space that can be estimated efficiently and used for further modeling. As with many medical applications, conditional variables offer additional information and can improve parameter estimation. A natural extension to the VAE is the conditional VAE, which includes additional conditional variables [27].

Diffusion Models

Denoising diffusion probabilistic models (DDPMs) [28] are a class of vision models that generate a new image by removing noise gradually from a pure noise input. The model learns to denoise data by understanding the probabilistic relationships between noisy and clean data points, making it particularly effective in scenarios in which data are corrupted by noise. During training, DDPMs simulate a diffusion process in which noise is gradually added to the data and then iteratively remove this noise to reconstruct the original signal. At each step, the noisy observation is generated by adding a small amount of noise to the previous observation. The diffusion process is typically modeled using an autoregressive process in which the next observation is conditioned on the previous observation and noise.

FMs for Medicine

FMs [29] are ML models capable of performing various generative tasks after being trained on extremely large and typically unlabeled datasets [30]. In the past few years, FMs have received significant attention given their impressive range of capabilities across multiple domains. FMs trained on EHRs have shown the ability to predict the risk of 30-day readmission [31], select future treatments [32], and diagnose rare diseases [33]. These models encode the patient features, such as procedures, diagnosis, medications, and laboratory values, into embeddings that represent the patient's entire medical history. There are 2 broad categories of FMs built from electronic medical record data: clinical language models (CLaMs) and FMs for electronic medical records (FEMRs). CLaMs are a subtype of large language models that specialize on clinical or biomedical text—CLaMs are primarily trained on and output clinical or biomedical text. On the other hand, FEMRs are trained on the entire timeline of events in a patient's medical history. Given a patient's electronic medical record as input, an FEMR will output not clinical text but rather a machine-understandable representation for that patient. This representation, also referred to as a patient embedding, is typically a fixed-length, high-dimensional vector that condenses large amounts of patient information.

Although most of the CLaMs are freely accessible on the web, the best-performing models are trained on private datasets—EHR-BERT [34]; University of California, San Francisco–BERT [35]; and GatorTron [36]. Researchers developed the BERT for Biomedical Text Mining [37] and PubMedBERT [38] transformer models using biomedical literature from PubMed. NVIDIA developed the BioMegatron models in the biomedical domain with different sizes from 345 million to 1.2 billion parameters using a more expansive set of PubMed-derived free text [39]. Other large transformer models include ClinicalBERT [40], which has 110 million parameters

XSL•FO
RenderX

and was trained using 0.5 billion words from the publicly available dataset.

Structured Text Data

Overview

In an effort to improve patient safety and reduce medical costs, >500,000 physicians and almost 6000 hospitals in the United States are expected to have operating EHRs and health IT systems [41] and charts [42] as of 2015. As a result, big data in health care in the United States have grown rapidly in recent years, with a massive database of electronic patient data. Among these, unstructured text data, also referred to as tabular data, are the most recorded and used time series, including numeric (eg, laboratory measurements and vital signs) and categorical (demographics, medications, and diagnostic codes) features. We note here that tabular data encompass a wide range of data, and we focus only on time series in this review.

Preprocessing

Among time-series structured data, vital signs and laboratory results are the most common sources. These are typically represented in one of two ways: (1) as one 2D matrix in which each row is an encounter or patient and each feature can be represented across time as multiple columns, such as *heart_rate_time1* and *heart_rate_time2*; and (2) as stacked 2D matrices in which each matrix contains columns of all features at 1 time point. Depending on the recorded values, feature values can be either numerical or categorical. During preprocessing, the number of time points can be determined by the mean or median number of entries available across patients. The interval between each point can be absolute or relative time depending on whether the cohort has a large variability in terms of hospital stay time. With a large variability, a set number of time points can correspond to the relative time such that all patients have the same number of points but the intervals do not match up in absolute time. For uncommon measures such as certain laboratory tests, there may be large missingness within the cohort, which is further discussed in the next section. Finally, feature engineering can be performed through either a manual selection or feature selection algorithms [43,44].

Applications

In this section, we introduce existing work on GenAI application in 3 main categories: data augmentation and imputation, disease classification and prediction, and counterfactual estimation.

Data Augmentation and Imputation

EHR datasets often face the problem of missing or unavailable data, which is especially common in laboratory data as not every patient in the cohort will receive the same tests. In addition, incomplete documentation or patient privacy concerns could also limit access to the complete dataset. Naive approaches such as filling missing data with 0 or forward filling in time add noise to the dataset at the cost of performance, prompting the need for better methods. Generative models impute missing values by learning underlying patterns of the trajectory, resulting in an overall smoother trajectory. Another challenge arises when the prevalence of certain rare diseases is so low that it is impossible for models to learn the pattern. In this case, generative models are used to generate synthetic patient records that mimic real-world data distributions. This can help in augmenting limited datasets and improving the robustness of predictive models trained on scarce data. The general approach of data imputation will start with an RNN model with LSTM layers [45-50] or GRU layers [51]. In more advanced approaches, researchers can modify the architecture parameters based on the dataset size or augment the model with additional pieces. In the work by Zaman and Du [52], the authors combined the GRU with ordinary differential equations to model irregularly sampled data and impute the entire time series. ImputeRNN [45] is a method that imputes data with consideration of medical bias by applying a mask during training. One can also combine DL methods, where a VAE is used first to find a data representation, which is then passed to an RNN for imputation [53]. For filling in missing data, it may be favorable to consider both previous and later observations. Bidirectional Recurrent Imputation for Time Series [54] imputes the missing data bidirectionally using an RNN graph. While RNN is the most common method in the space, GANs have also become a popular choice. In a few studies (4/155, 2.6%), time-series data were treated as an “image” on which a GAN was trained to fill in the missing “gaps” [55-58].

Disease Classification and Prediction

The most common goal in modeling disease trajectory is disease prediction and classification. Generative models can learn temporal dependencies within these sequences and generate realistic patient trajectories over time, enabling downstream predictive modeling of disease progression or treatment outcomes. There are several types of predictions that can be made: a binary classification of whether an event will occur (eg, mortality or sepsis), a probability prediction of an event occurring at some point, or a time at which an event occurs (survival analysis). Due to time-series data’s nature of unfixed length and frequency, many ways are proposed to handle such data. In earlier approaches, time points in time-series data are treated as separate entries with no relation to each other and passed through a fully connected neural network to predict the disease of interest. However, this approach ignores temporal information that can be crucial in future predictions. One of the most common ways of handling input with variable length is to first convert the input into a fixed-length representation, which is then passed through a network. In the study by Guo et al [59], measures such as BMI, smoking status, and cholesterol level were first converted to an embedding of fixed size and then passed through an RNN with LSTM layers to predict cardiac dysrhythmias. In another study [60], time-series data including laboratory results and vital signs were divided into 5 relative time points in reference to the duration of the hospital stay (eg, if the hospital stay for 1 patient is 4 hours, the times are admission, 1 hour into admission, 2 hours into admission, and so on) to fix the input length and predict mortality in patients with heart failure.

In many studies (19/155, 12.3%), once the embedding choice was made, the subsequent model of choice was similar. In these studies, RNN models were used to predict methicillin-resistant *Staphylococcus aureus* positivity [61], mortality [51,62-64],

XSL•FO
RenderX

readmission [65-68], next diagnosis [69-72], length of stay [73,74], hypertension [75,76], treatment response [77], and survival analysis [78]. Furthermore, the general-use prediction models in several studies (4/155, 2.6%) [32,79-81] all used an RNN model to predict clinical events based on time-series data. Extending from a single prediction point, the work by Pham et al [81] used LSTM to model irregularly timed time-series data and predict multistep future trajectories on diabetes and mental health cohorts. In recent years, researchers have also turned to transformer models for prediction tasks [82-84] as they capture more long-range dependencies and provide more interpretability of model weights than RNNs. In the work by Sheikhalishahi et al [85], self-attention was used to measure each feature's relevance to each other and also each time point's relevance within each feature to predict delirium in critical care, and another study [86] used similar mechanisms to predict the mortality risk of patients with cardiac issues. Transformers have also been useful in making multi-point predictions. The Transformer for Electronic Health Records is one of the earlier models using transformers to predict clinical events and can predict up to 300 events simultaneously [87]. The Dual Event Time Transformer for Electronic Health Records is a general-purpose transformer-based method that takes in time-series features [88]. In addition to RNN and transformers, a deep diffusion model can also be used to learn complex representations of disease networks that can be used for downstream prediction tasks [89].

Counterfactual Estimation

Counterfactual estimations aim to track the "what-if" trajectory—what would a patient's trajectories look like if they were not given any medication or treatment. To estimate a trajectory that is never observed, one must take into account both observed and hidden confounding variables, which is challenging in causal inference models. AI models offer the advantage of making no previous assumptions about the relationship between variables, learning purely from the data. However, this does not tackle the issue of the hidden variable, which is still an active problem in the field. Most commonly, RNNs were used to estimate treatment outcomes and, therefore, provide counterfactual estimation in several studies (4/155, 2.6%) [77,89-91]. Depending on the task and data at hand, modifications can be made to the basic RNN to achieve better results. In the work by Wang et al [92], a 2-layer bidirectional LSTM model was used for survival estimation. Similarly, the work by Li et al [93] combined RNN with g-computation, a causal inference method, to estimate treatment effects. A combination of reinforcement learning and RNN was also used to predict counterfactual evaluation in a study on public intervention for COVID-19 [94]. In addition to RNN models, autoencoders have been used to discriminate patient characteristics from treatment for counterfactual estimation [95].

Model Selection

As they are most abundant and easiest to access, unstructured text data are a good place to start for clinicians who want to explore GenAI tools. In Figure 5, we present a flowchart to help guide the process of choosing the right model for the data and task at hand. The flowchart is organized in the same way as the sections in this paper, where different colors refer to different data modalities for easier visualization. In addition to the statistical methods mentioned previously, there are also a number of ML models, including logistic regression, decision tree, random forest, and Extreme Gradient Boosting (colored red in Figure 5). While we do not discuss these methods in this review, there are review papers that describe them in detail [96-98]. If the dataset is quite small, a traditional autoregressive or a simple ML method would work better as more complex models require more data. However, if the dataset is large, DL models would perform better as the relationship between features is more complex and tends to be nonlinear. Commonly, the model of choice will be RNN or transformer based, which also have the most literature available. Once one has chosen a base model, they can start with a published framework and modify certain parameters based on their goal. If the model is quite simple and only consists of linear layers, one may also consider changing part of the structure to newer architectures, such as attention layers, or chaining 2 models together to improve performance.

XSL•FO
RenderX

Figure 5. Flowchart to choose the appropriate deep learning model. BERT: Bidirectional Encoder Representations From Transformers; DDPM: denoising diffusion probabilistic model; ESM1b: Evolutionary Scale Modeling 1b; EVE: Evolutionary Model of Variant Effect; LSTM: long short-term memory; MBP: masked bidirectional prediction; ML: machine learning; PRESCIENT: Potential Energy Underlying Single-Cell Gradients; SCUBA: single-cell clustering using bifurcation analysis; SIFT: Sorting Intolerant From Tolerant; SVM: support vector machine; TF-IDF: term frequency–inverse document frequency; TIGON: Trajectory Inference With Growth via Optimal Transport and Neural Network; VAE: variational autoencoder; VUS: variant of uncertain significance; XGBoost: Extreme Gradient Boosting.

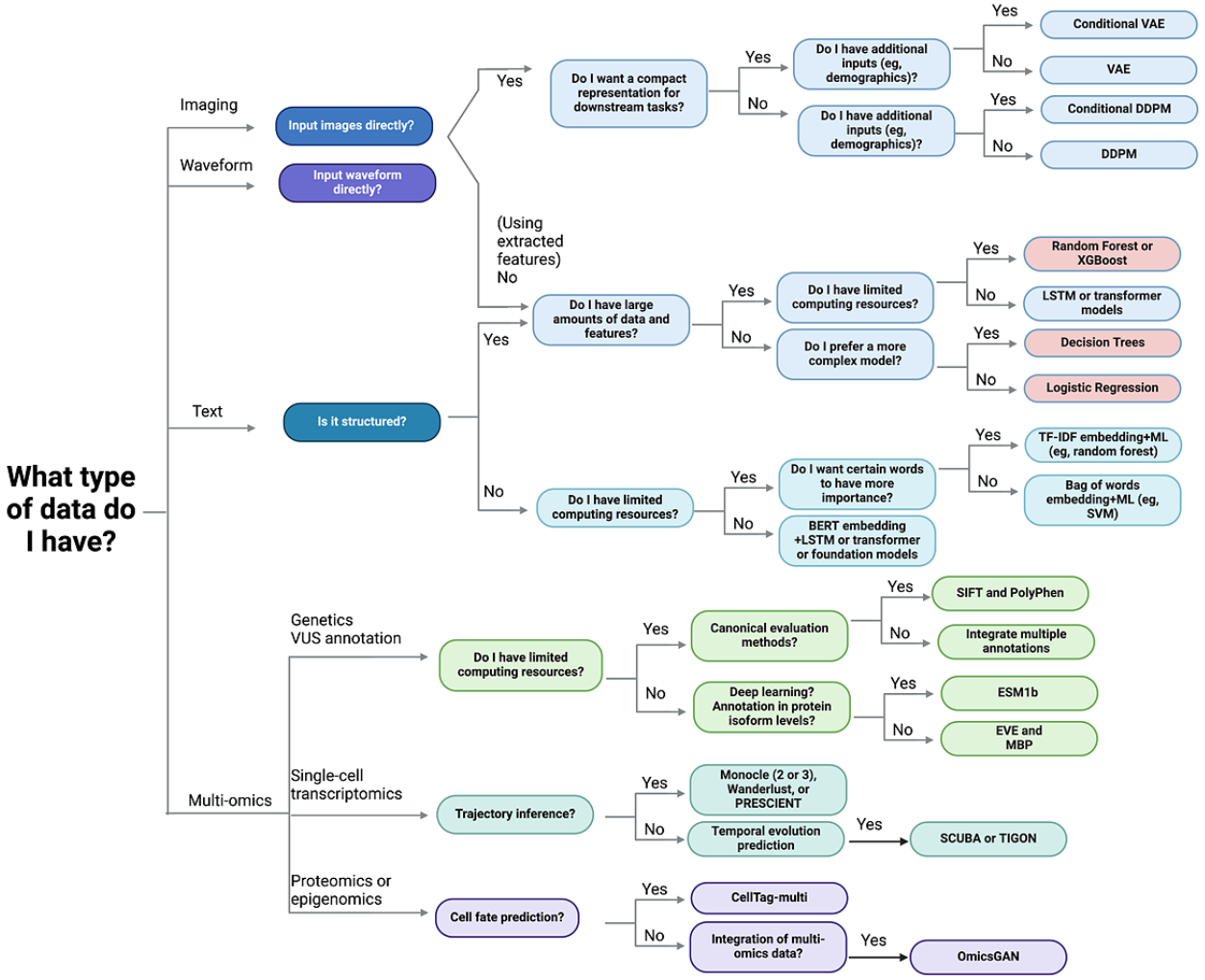

Challenges and Future Directions

As ML and DL methods, including GenAI, continue to grow with an increasing amount of available data, many comparisons have been made between AI methods and traditional statistical methods. Cerqueira et al [99] compared ML methods in time-series forecasting (rule based, random forest, and Gaussian process regression) to statistical methods (ARIMA, Naive2, and ETS) using datasets with varying sample sizes up to 1000 and showed that ML methods outperformed statistical methods when the sample size was relatively large. Makridakis et al [100] jointly evaluated the accuracy of DL methods and statistical methods (Naive2, ARIMA, ETS, and ensemble methods) using the 1045 M3 dataset and showed that, among the nonensemble approaches, the statistical methods on average yielded better performance than DL methods for short-range prediction with regard to symmetric mean absolute percentage error and mean absolute scaled error, whereas the DL methods performed better in general for medium- and long-range predictions. The DL ensemble method using a combination of the 4 DL methods consistently outperformed the nonensemble methods in the study by Bonnheim et al [101]. However, in a practical setting, AI methods face the biggest challenge of low transparency due to their black-box nature. Interpretability has been an active area of research, and many techniques and tools have been proposed to understand the decision-making process of such models. Feature importance analysis tools such as Shapley Additive Explanations [102] and Local Interpretable Model-Agnostic Estimations [103] estimate feature contribution to reveal top features that are important to the decision. These tools may help shed light on how the model predicts the outcome, which is extremely important in the field of health care. A second challenge that AI methods face is the trade-off between generalization and performance as an exceptionally good performance of a model on one dataset may be caused by overfitting. As patient distributions can differ widely between hospitals and areas, the model's robustness must be measured before deploying it in practice. Nonetheless, GenAI has shown great potential in clinical support, such as early warning systems or care assistance tools. Future GenAI tools that tackle these challenges will become paramount in health care.

XSL•FO
RenderX

Unstructured Text Data

Overview

Unstructured data include a wide range of clinical narratives, such as progress notes, radiology reports, patient correspondence, and discharge summaries, comprising 80% of EHRs [104]. The loosely structured nature of typed text (also known as "free text") is effective in day-to-day clinical workflows but presents a major challenge for automating EHR-based registries. The unstructured data may contain key patient information missing in structured data, extra information complementing structured data, or even data that may contradict information presented in structured data. As such, the overall goal of modeling using unstructured data is to assist disease prediction using structured data by adding critical missing information. The complexities of unstructured data, along with the fact that existing text-mining tools and NLP applications have limited accuracy in extracting information from free text, have prompted some registries to ask for a manual chart review of individual patients before final inclusion in the registry. Unstructured data limit the application of automated computational phenotyping methods and increase the likelihood of low data quality (eg, missing data) when data are extracted from structured EHR data only [105,106].

Preprocessing

Information retrieval from unstructured data is challenging because the information can often be of low quality and contain missing values. Commonly used data preprocessing techniques for unstructured data include data collection, data cleaning, tokenization, stemming, and the removal of stop words [107,108]. Tokenization refers to segregating, for example, words and punctuation marks in the input document as tokens and computing their frequency. Stemming is the process of identifying the roots of words, typically by removing the suffix. Stop word removal consists of removing infrequent words to reduce the dimensionality of the dataset. Once the key components of the dataset are extracted, they can be used for downstream applications such as the identification of risk factors and individuals at high risk of a particular disease.

Applications

As the application of large language models to health care time-series data is still a developing area of research, in this section, we discuss static time applications of these models. We will discuss potential research directions for time-series unstructured text based on progress in other fields such as climate and finance in the Challenges and Future Directions section. Both traditional NLP and FMs can be used to perform clinical NLP tasks such as clinical concept extraction (or named entity recognition), medical relation extraction, semantic textual similarity, natural language inference, medical question answering, and medical report summarization. The domain of medical report summarization is a popular area of research, with numerous papers published on the use of NLP and FMs to summarize medical reports, health care records, and medical dialogues [109]. A radiology report is a medical document that contains the details of an imaging examination (such as x-rays and magnetic resonance imaging [MRI]). A radiology report consists of three components: (1) the Background section, which contains the medical history of the patient; (2) the Findings section, which discusses the crucial observations and findings of the radiology examination; and (3) the Impression section, which is a short summary of the findings section. The Impression section is usually written by medical professionals, which is a time-consuming process. The only aim of radiology report summarization is to automate the generation of this Impression section. Several attention-based and FM-based models have been developed to improve the quality of summarization [110].

Medical dialogue summarization corresponds to the automatic generation of coherent summaries that capture medically relevant context from dialogues between patients and health care providers. Medical dialogue summarization can help medical providers keep a record of patient encounters and also provide the necessary context of a patient's medical history during patient handoffs between providers. Existing studies have used techniques from computational linguistics [111], NLP (Pegasus [112]), pretrained language models, and low-shot learning to collect labeled data and perform medical dialogue summarization.

Medical question answering refers to developing a technique that automatically analyzes thousands of articles to generate a short text, ideally in less than a few seconds, to answer questions posed by physicians [113]. Such a technique provides a practical alternative that allows physicians to efficiently seek information at the point of patient care. Several question-and-answer systems have been developed using NLP [114] and FMs (generalist medical AI and MedLM).

Model Selection

FMs hold great potential to assist clinicians in a wide range of health care problems. However, clinicians should be aware of the risks associated with the use of these models and potential data leaks. They should also select which models to use based on the computational resources available to them, as shown in Figure 5, as the training of these models is becoming increasingly expensive (Table 1).

Challenges and Future Directions

Compared to traditional ML models, FMs have a significantly higher computational cost, requiring massive datasets and graphics processing unit specifications for training. While they achieve high predictive accuracy [115], these models are composed of millions to billions of parameters and have poor interpretability and high latency and runtime. While these costs can be lowered over multiple downstream applications, their value may take longer to realize than that a smaller model developed for a single high-value task [30]. In addition, data privacy and security are significant concerns with FMs as they may leak protected health information through model weights or prompt injection attacks [116]. As the cost of computing decreases and the amount of data increases, FMs may become easier to train and more accessible. One active area of research is testing FMs' potential in multimodality input, in which FMs are trained and evaluated on multimodal datasets, such as both text and images. These models can be used to generate

XSL•FO
RenderX

personalized treatment plans based on the patient history as well as develop early warning systems for individuals at high risk. These advancements would transform how health care professionals interact with patient data, making health care more predictive, personalized, and precise.

Medical Imaging

Overview

Medical imaging refers to 2D or 3D images obtained from radiology procedures such as MRI and computed tomography (CT). GenAI has been applied to a wide set of problems in medical imaging to solve issues of scarcity, heterogeneity between datasets, and low resolution quality, among others. There are 3 main classes of generative vision methods: GANs, VAEs, and DDPMs. Each of these classes has been applied to medical imaging. Among all applications, medical image series input has been used mostly in data generation, anomaly detection and classification, and image segmentation and registration tasks.

Preprocessing

There are several steps in preparing a medical imaging dataset as the quality varies largely depending on location, time, and cohort [117]. First, a denoising technique is applied to remove as much nuisance noise as possible. After denoising, data interpolation can be conducted for certain images such as CT scans to ensure the same spacing across all images. In the case of MRI images, bias field correction may be performed to adjust image intensity. Third, image registration is performed to align all images with a common template to avoid unnecessary incoherency among images, such as rotation. Finally, images can be standardized or normalized to improve algorithm efficiency. Depending on computational resources, the image may also be downsampled to a lower resolution or split into patches to reduce computational cost. Publicly available datasets such as those of the Alzheimer's Disease Neuroimaging Initiative [118] and Open Access Series of Imaging Studies [119] do not need much preprocessing as they have been carefully curated. However, image registration is recommended to align all patients with a common template of choice if multiple datasets are combined. The medical images are usually represented as 2D matrices or 3D vectors of the pixels or voxels.

Applications

Data Generation

Similarly to tabular data, data generation has been used to combat the problem of data scarcity. For longitudinal medical image series, GANs, VAEs, and DDPMs have all been used to generate synthetic images. In the work by Hamghalam and Simpson [120], a recurrent conditional GAN was used to generate sequential image data conditioned on past variables. Diffusion-based models have also been applied to impute longitudinal CT and MRI images [121]. Finally, predicting and generating aging brain images using only the input of age and disease state is an example of applying autoencoders with contractions on the latent space [122].

Anomaly Detection and Classification

The most common application to longitudinal images is anomaly detection and disease progression prediction. For clinical diagnosis, GenAI-based anomaly detection tools have been developed in which the model identifies deviations in 1 image from the rest of the population, aiding in the detection of diseases or abnormalities. For medical image series, previous scans are used as inputs or conditional variables to predict the next image. Many approaches in this space are a variation of GANs as they were introduced earlier than VAEs and DDPMs. Gaussian-Poisson GAN uses stacked GANs to predict tumor growth from longitudinal MRI [123]. The multi-Pareto GAN makes longitudinal prediction of infant MRI using multi-contrast perceptual adversarial learning [124]. In the low-dimensional GAN [125], disease progression images are predicted using missing MRI in the input. In the study by Devika et al [126], the authors trained a GAN to predict autism spectrum disorder given longitudinal structural MRI scans. More recently, VAEs have been applied extensively as they offer a latent space to work with. In general, researchers start with a vanilla VAE and apply slight modifications to fit their data modality and goal to detect and predict diseases, including multiple sclerosis [127], Alzheimer disease (AD) [128-131], tau biomarker detection for AD [132], glaucoma [133], and lung cancer [134]. There has also been an increase in works published using multi-modality data, such as imaging data and tabular or genetics data. In the work by Kmetzsch et al [135], the authors trained a VAE model to predict disease progression score based on longitudinal images and micro-RNA data input. When dealing with multimodal inputs, a combination of DL models dealing with each modality may also be advantageous. The work by Sauty and Durrleman [136] combined VAE with a linear mixed-effects model to learn a Riemann progression model that can be applied to general disease trajectory estimation. Another study combined an autoencoder framework with attention units in a transformer to predict final ischemic stroke lesions from MRI [137]. In the study by Zhang et al [138], a spatiotemporal convolutional LSTM was learned to combine imaging and nonimaging data to predict tumor growth images. Conditional SliceGen is a method that takes an arbitrary axial slice in the abdominal region as a condition and estimates a vertebral-level slice [139].

Image Segmentation and Registration Tasks

Another common task in medical imaging is image segmentation and registration. We note here that, while image segmentation and registration tasks are different concepts, the underlying models applied to solve the problems are largely the same. In the work by Zhang et al [140], an encoder-decoder structure model was used for contrastive learning for image segmentation. Extending from autoencoders, a conditional VAE has also been used to produce cardiac image segmentation conditioned on predefined anatomical criteria [141]. We note here that, while work in this area using time-series input is scarce, the underlying model and training process remain the same for single time input models [142-147]. Furthermore, many of the methods mentioned previously for anomaly detection can also be applied to registration and segmentation tasks during training.

XSL•FO
RenderX

Model Selection

Although there are many tasks associated with imaging data input, the underlying model architecture remains largely the same for prediction, synthesis, registration, or even domain transfer tasks. For anyone using time-series or longitudinal images for any task, choosing 1 of the 3 families of vision methods following the workflow in Figure 5 is a good place to start. One may also refer to Table 1 to compare the models' strengths and weaknesses. GANs are an earlier framework that remains powerful today, especially with synthetic generation, but are more complicated to implement as they have a generator and a discriminator. On the other hand, VAEs offer a latent space that can be used to downstream linear models such as mixed-effects models to control for variables such as age, but its generation quality may not be as ideal as that of GANs or DDPMs. Finally, diffusion models have been largely applied in recent years due to their simple architecture compared to that of the other 2 and low memory requirements but offer very little interpretability on the generation process. For clinicians with limited computing resources, 3D images may be too large to fit onto graphics processing units, and in many cases, the images can be split into smaller patches to reduce memory footprint [148]. Finally, there are publicly available general-purpose vision models that can be downloaded as a starting point, which can be further trained and fine-tuned for the specific task at hand.

Challenges and Future Directions

With computer vision models' success in recent years, there is a lot of potential in their applications to medical imaging. In recent years, diffusion models have shown great promise in generating images that are comparable to those generated by state-of-the-art GAN methods. Furthermore, GenAI methods that support video generation can potentially be adapted for medical imaging [149-153]. However, some issues that are present in current vision models also extend to medical imaging, such as illogical generations that do not comply with physical laws [119]. Therefore, clinicians should take the generations with a grain of salt when applying them to downstream tasks. The advent of vision models has been revolutionary in the study of medical images as previous methods must first transform the images into numeric representations, such as thickness and mean intensity. These representations are then fed into linear models for tasks such as classification and segmentation. To work with raw image inputs, linear models of estimation such as principal component analysis or single-value decomposition are often used. However, these generations often have poor resolution or fail to adequately capture variability. In contrast, modern neural network approaches (eg, CNN) *jointly* learn the optimal feature extraction strategy together with the downstream task. Thus, GenAI algorithms are able to consider highly nonlinear and nuanced interactions between inputs that better capture the underlying distribution of the data. While GenAI faces the challenge of transparency, interpretability methods such as Gradient-Weighted Class Activation Mapping [154] have shown convincing results. Another challenge of GenAI in medical imaging is the computational resources required to support such a model, which will become cheaper and easier to acquire as hardware technology becomes more efficient.

Physiological Waveforms

Overview

Waveform data capture the change of one or multiple physical quantities over time and can be viewed as either univariate or multivariate time-series data. Examples of medical waveform data include single-lead electrocardiograms (ECGs) or photoplethysmograms, which can be represented as univariate time series. Multi-channel technologies such as electroencephalograms (EEGs) are typically represented as multivariate time series. Although waveforms are considered to be closer to structured time-series tabular data, DL methods applied to structured data often face technical limitations when applied to waveforms as the sampling frequency is much higher. As such, GenAI methods applied to waveforms often resemble speech processing [155] or vision-based models adapted for 1D input signals.

Preprocessing

Physiological waveforms are often recorded over a short period with low missingness but face the problem of noise and signal artifacts. Therefore, waveform preprocessing aims to remove as much nonphysiological noise and artifacts as possible while preserving the actual signal. This is typically accomplished by applying digital filters designed to address nonbiological signals, drifts, and spikes [155]. Depending on the signal, specific filters have also been developed and shown strong performance, such as the Wiener filter for ECGs [156]. After filtering, the waveform is usually normalized for efficiency and resampled to a common frequency.

Applications

GenAI models have gained increasing popularity among medical waveform data. Major tasks using GenAI models for this type of data include synthesis and imputation and anomaly detection.

Synthesis and Imputation

GenAI-based approaches can augment existing data and generate synthetic data to support downstream algorithms, allowing for more applications when the amount of real medical waveform data is limited. For this task, GANs have been used extensively [156-161], along with diffusion models [162]. In their work, Elazab et al [123] made modifications to the improved Wasserstein GANs and generated single-channel EEG signals. Zhu et al [163] constructed a GAN with bidirectional LSTM and CNN to synthesize single-beat ECGs and demonstrated that the model performed consistently better than benchmark architectures with regard to the quality of the signals. In another work, To et al [127] reconstructed high–sampling sensitivity EEG signals with improvements in reconstruction and classification results.

Anomaly Detection

While little work has been done on anomaly detection directly, GAN-based approaches have been explored to detect cardiac abnormalities [164-166] and seizures [167,168]. Diffusion-based models have also been used for anomaly detection with ECGs and EEGs. Xiao et al [169] adopted a denoising diffusion-based imputation method with conditional weight-incremental

XSL•FO
RenderX

diffusion to improve data generation quality and ECG anomaly detection stability. Wang et al [170] designed a diffusion-based framework with a reverse diffusion data augmentation module to learn more generalized features for major depressive disorder detection using EEGs.

Model Selection

In general, waveform data are best trained using vision models such as VAE and DDPM as the time series can be viewed as a 1D image. Therefore, the workflow to choose an appropriate model for waveform data is similar to that for imaging data. As GANs are often more complex but do not significantly outperform other vision models, they are being replaced with VAEs and DDPMs [171,172].

Challenges and Future Directions

As GenAI models are only beginning to be applied to waveform data, there are many directions to pursue. One of these is cross-modality imputation, where, given an EEG, a GenAI model can be trained to generate the corresponding functional MRI or vice versa. Another area of study is to convert EEG signals to natural language by combining vision models with a language model [173]. As previous studies have focused on modeling a single type of waveform data in isolation, one natural direction is to use more data sources, for example, EEGs with text data or ECGs with medication data, as they are often informative about one another in tasks such as imputation [174] and forecasting [175].

Genetics and Multi-Omics Data

Multi-omics data encompass genomics, transcriptomics, proteomics, epigenomics, and metabolomics, offering a detailed picture of biological activities at various molecular layers. One of the most successful applications of AI in genetics is in trajectory modeling using multi-omics data, which focuses on how biological systems such as cell development, diseases, or reactions to treatments change over time.

Preprocessing

Multi-omics data comprise multiple-sequencing data types and various features. To efficiently preprocess the data, different methods are used for each type of data. DNA sequences are represented as strings of adenine, thymine, guanine, and cytosine nucleotides, which can be transformed into a numerical matrix in which each nucleotide is represented as a binary vector. For transcriptomics, expression data are often represented as a matrix of continuous numeric values in which each row corresponds to a sample or a cell and each column corresponds to a gene or transcript. Proteomics data are similarly structured but represent peptide abundances or protein expression levels, and epigenomics data such as methylation status are typically captured as binary values or continuous levels indicating methylation at specific sites. Depending on the goal, different omics data can be combined to uncover time-related trends and critical factors that influence those biological shifts.

Applications

Identifying Pathogenic Mutations

Identifying pathogenic mutations is pivotal for unraveling the genetic underpinnings of diseases. The task of variant effect prediction, which seeks to determine the phenotypic implications of genetic variations, stands as a cornerstone challenge in human genetics. Traditional approaches leveraging statistical methods alongside evolutionary conservation scores such as Sorting Intolerant From Tolerant and phyloP have made strides in classifying variants of uncertain significance (VUSs) [176,177]. Despite these advancements, accurately differentiating between variants that disrupt protein function and those with neutral effects remains a daunting challenge. The persistence of many missense variants within the gray area of VUSs significantly hampers the diagnostic utility of exome sequencing in clinical settings. To effectively address the challenge of categorizing mutations, particularly those classified as VUSs, ML techniques have been increasingly applied. Among the array of ML methodologies, random forest, gradient-boosting models, and support vector machines are often used for training complex genetic datasets to decipher VUSs [178,179]. However, the high dimensionality of genomic information and the intricate interactions between genetic variants often exceed the capacity of traditional ML techniques, leading to potential overfitting and a lack of generalizability to unseen data. DL methods, including CNNs and RNNs, offer a promising solution for addressing the challenges of modeling complex, nonlinear relationships in high-dimensional biological data. These methods significantly diminish the need for manual feature engineering by efficiently learning from raw data. Among the DL models developed for predicting VUSs are Evolutionary Scale Modeling 1b (ESM1b), Evolutionary Model of Variant Effect, MetaRNN, masked bidirectional prediction, PrimateAI, and deleterious annotation of genetic variants using neural networks. ESM1b is particularly noteworthy for its innovative use of a protein language model developed by Meta AI [180-185]. ESM1b has demonstrated exceptional capability in classifying approximately 150,000 ClinVar and Human Gene Mutation Database variants and annotating approximately 2 million variants as damaging in specific protein isoforms, a feat not previously achieved with ML-based methods. This focus on isoform-specific annotations underlines the critical importance of considering all protein isoforms in variant effect predictions, marking a significant advancement with which ESM1b contributes to genetic research.

Trajectory Modeling

Time-series transcriptomics and proteomics data have been used widely to study trajectories in biological processes and disease development. Among them, single-cell transcriptomics data have been used to reveal cellular dynamics via pseudotime inference models, including Monocle (versions 2 and 3), Wanderlust, scVelo, single-cell clustering using bifurcation analysis, partition-based graph abstraction, and Trajectory Inference With Growth via Optimal Transport and Neural Network [154]. Monocle uses single-cell variations to sequence cells in pseudotime, depicting their progression through biological processes such as differentiation based on gene expression. Monocle version 2 [186] uses reversed graph

XSL•FO
RenderX

embedding with a minimum spanning tree algorithm for pseudotime reconstruction, whereas Monocle version 3 [187] enhances this with principal graph learning for refined trajectory inference. Wanderlust adjusts each cell's trajectory position using weighted averages from the shortest-path distances of randomly chosen waypoints until convergence, thereby producing an average trajectory. scVelo [188], a likelihood-based dynamic model, estimates RNA velocity to derive dynamic insights from RNA sequencing data, analyzing gene-level transcriptional dynamics by determining gene-specific transcription, splicing, and degradation rates, which is suitable for both transient states and systems with varying subpopulation kinetics. Single-cell clustering using bifurcation analysis [189] forecasts the temporal evolution of gene expression in single cells using theories of nonlinear dynamics and stochastic differential equations, aiding in the comprehension of gene expression dynamics. Partition-based graph abstraction [190] merges clustering and trajectory inference in single-cell RNA sequencing data, creating a graph to depict cell relationships based on gene expression and enhance understanding of cellular transitions. Finally, Trajectory Inference With Growth via Optimal Transport and Neural Network [191] uses dynamic unbalanced optimal transport based on the Wasserstein-Fisher-Rao distance to integrate temporal datasets, which provides a framework for connecting temporal measurements and predicting novel dynamics. In addition, expression data can be used to study responses to infectious diseases. Huang et al [192] analyzed time-series expression data from human volunteers infected with influenza. They identified distinct temporal patterns of gene expression that could discriminate between asymptomatic and symptomatic infections. Similar approaches have been used to study the response to various treatments, such as interferon-β therapy in patients with multiple sclerosis [193].

On the other hand, single-cell proteomics approaches such as mass cytometry (cytometry by time of flight) have revolutionized our ability to understand complex cellular processes at an unprecedented resolution. Palii et al [194] used single-cell proteomics to define the temporal hierarchy of human erythropoiesis. The analysis revealed a timely ordered appearance and disappearance of transient cell populations or stages that accumulate at various positions along the erythroid trajectory, with cells undergoing gradual transitions between these stages. In the fields of proteomics, epigenomics, and metabolomics, CellTag-multi [195], a method for single-cell lineage tracing across single-cell RNA sequencing and single-cell assay for transposase-accessible chromatin sequencing assays, mapped transcriptional and epigenomic states of progenitor cells, significantly improving cell fate prediction. In addition, linear mixed models have been used to quantify temporal changes in metabolite concentrations, offering insights into the dynamics of the metabolic system [196,197].

Model Selection

GenAI has shown great promise in analyzing multi-omics data for disease progression analysis. By using techniques such as omicsGAN, which uses GANs, and graph-linked unified embedding and MultiVI, which use VAEs, GenAI facilitates the integration and compression of multi-omics data from diverse sources into a unified representation [198-200]. This aids in a more comprehensive analysis of the data. In addition, tools such as Potential Energy Underlying Single-Cell Gradients apply generative modeling to map out potential landscapes from time-series single-cell transcriptomics data, making it possible to generate trajectories for unseen data points [201]. This opens up avenues for hypothesizing about biological perturbations and pathways of disease progression. Moreover, GenAI proves instrumental in identifying distinct disease subtypes and stratifying patients based on their multi-omics profiles. For example, gene-guided weakly supervised clustering via GANs uses a GAN to generate imaging features from brain MRIs and single nucleotide polymorphisms to explore conditions such as AD and hypertension [202], whereas the Trained GAN Discriminator model stratifies patients with breast cancer into high- versus low-risk categories based on their transcriptome profiles [203]. For clinicians interested in working with multi-omics data, we outline a procedure for finding the best tool in Figure 5 based on their specific needs, such as resource availability and the type of data analysis required, including genetics, single-cell transcriptomics, and proteomics or epigenomics.

Challenges and Future Directions

Analyzing and modeling multi-omics data in time-series contexts presents challenges in addressing missing data across different biomolecule types, which complicates the integration of datasets obtained through various omics technologies. In addition, the complexity of biological interactions can lead to analytical difficulties when dealing with high-dimensional data. The integration of multi-omics data is becoming increasingly important in the pursuit of future progress. This approach uses various biological data layers—genomics, transcriptomics, proteomics, and metabolomics—to enhance the understanding of genetic interactions and disease mechanisms.

Discussion

Data Ownership and Privacy

Overview

While GenAI software has proven powerful in understanding and predicting disease trajectories across modalities, its implementation in the clinic remains difficult due to challenges including ethical considerations [204] and other regulatory and technical challenges. Regulations such as HIPAA (Health Insurance Portability and Accountability Act) and the General Data Protection Regulation protect personal health information, but the evolving nature of GenAI complicates compliance, especially regarding AI-generated synthetic data and indirect identification risks [205-207]. This rapid advancement outpaces regulatory updates, causing uncertainty among health care providers and AI developers about legal responsibilities and best practices [208-210]. Technically, health care time-series forecasting struggles with short data lengths and significant temporal changes as discussed in previous sections, making model training and parameter estimation difficult [211,212].

Data ownership and privacy in health care GenAI are critical issues involving rights and responsibilities over data control,

XSL•FO
RenderX

use, and dissemination [213]. This includes patient records, genetic information, and AI-generated datasets. Ambiguities in ownership, particularly with AI-generated synthetic data, complicate matters. This raises questions about who owns the data—the AI developer, health care providers, or patients [214]. Current regulations enforce data minimization, purpose limitation, and consent to protect patients' privacy rights [215]. However, GenAI models require extensive patient data for training, leading to the risk of reidentification of anonymized data through advanced algorithms [216]. Ensuring that raw patient data remain protected is crucial to prevent unauthorized access [217,218].

Security

AI models in health care should be deployed with the highest standard of data security as breaches of sensitive patient information can be disastrous. In 2023, >500 data breaches affected >112 million individuals [219]. Cyberattacks such as phishing and malware are common, but prompt injection attacks in GenAI models are emerging threats [220]. Effective countermeasures such as prompt filtering and human-in-the-loop systems are essential despite potentially compromising model performance [221-224].

Bias

The deployment of GenAI in health care can also result in inadvertently embedded societal biases, leading to discriminatory outcomes [225]. Biases in training data can result in disparities in treatment across different patient groups [226]. Addressing these biases requires comprehensive strategies throughout AI development to ensure fairness and equity [227-231].

Clinical Safety and Reliability

Ensuring clinical safety and reliability is vital for the trustworthiness of health care GenAI. AI models can produce inconsistent or erroneous outputs, particularly in complex medical scenarios, due to issues such as hallucinations [232,233]. In health care, such errors are unacceptable and can lead to harmful consequences, which require rigorous scrutiny and validation. Moreover, AI models, particularly DL models, often lack transparency and explainability, posing challenges in health care [234]. Enhancing transparency and explainability through techniques such as feature importance scores and decision trees is crucial for gaining trust and ensuring comprehensible AI decisions [235-237]. Regulatory and ethical frameworks must evolve to support these efforts [238].

Conclusions

While there remain many challenges to deploy reliable GenAI methods in real practice, including technical, regulatory, and ethical issues, these models can have huge impacts that could benefit patients, practitioners, and administrators. As such, collaboration among policy makers, legal experts, AI developers, and health care providers is crucial in overcoming these issues. In this paper, we introduced concepts basic to DL and GenAI to those unfamiliar with the field, reviewed existing work on GenAI application to time-series health data, and offered suggestions to clinicians who are interested in applying such tools to their field. We hope that this review bridges the gap and promotes better collaboration between technical and applied disciplines to develop better methods that are suitable and robust in a practical setting.

Acknowledgments

All figures in this manuscript were created using BioRender.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 108 KB - jmir_v27i1e59792_app1.docx]

Multimedia Appendix 2
Summary of existing reviews on artificial intelligence in health care.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v27i1e59792_app2.xlsx]

Multimedia Appendix 3
Number of existing reviews and surveys published after 2010 identified from the preliminary search by year (before filtering) using the following artificial intelligence– and biomedical-related search string: ("Generative AI" OR "Generative artificial intelligence" OR "genAI" OR "foundation model" OR "GPT" OR "Generative Adversarial Networks" OR "transformer" OR "variational autoencoder" OR "diffusion model" OR "flow model") AND ("review" OR "summary" OR "survey") AND ("time series") AND ("disease trajectory" OR "health care").
[PNG File , 51 KB - jmir_v27i1e59792_app3.png]

References

XSL•FO
RenderX

1. Kumar S, Musharaf D, Musharaf S, Sagar AK. A comprehensive review of the latest advancements in large generative AI models. In: Proceedings of the 2nd International Conference on Advanced Communication and Intelligent Systems. 2023 Presented at: ICACIS '23; June 16-17, 2023; Warsaw, Poland p. 90-103 URL: https://link.springer.com/chapter/10.1007/978-3-031-45121-8_9
2. Gozalo-Brizuela R, Garrido-Merchan EC. ChatGPT is not all you need. A state of the art review of large generative AI models. arXiv Preprint posted online January 11, 2023 [FREE Full text]
3. Nguena Nguefack HL, Pagé MG, Katz J, Choinière M, Vanasse A, Dorais M, et al. Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches. Clin Epidemiol 2020;12:1205-1222 [FREE Full text] [doi: 10.2147/CLEP.S265287] [Medline: 33154677]
4. Guess HA, Jacobsen SJ, Girman CJ, Oesterling JE, Chute C, Panser LA, et al. The role of community-based longitudinal studies in evaluating treatment effects. Example: benign prostatic hyperplasia. Med Care 1995 Apr;33(4 Suppl):AS26-AS35. [Medline: 7536867]
5. Jack Jr CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008 Apr 26;27(4):685-691 [FREE Full text] [doi: 10.1002/jmri.21049] [Medline: 18302232]
6. Bandi A, Adapa PV, Kuchi YE. The power of generative AI: a review of requirements, models, input–output formats, evaluation metrics, and challenges. Future Internet 2023 Jul 31;15(8):260. [doi: 10.3390/fi15080260]
7. Whittle P. Some results in time series analysis. Scand Actuar J 1952 Jan;1952(1-2):48-60. [doi: 10.1080/03461238.1952.10414182]
8. Box GE, Pierce DA. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc 1970 Dec;65(332):1509-1526. [doi: 10.1080/01621459.1970.10481180]
9. Jeyarani AD, Singh TJ. Analysis of noise reduction techniques on QRS ECG waveform - by applying different filters. In: Proceedings of the 2010 Conference on Space Technology Services and Climate Change. 2010 Presented at: RSTS & CC '10; November 13-15, 2010; Chennai, India p. 149-152 URL: https://ieeexplore.ieee.org/document/5712835 [doi: 10.1109/rstscc.2010.5712835]
10. Gardner Jr ES. Exponential smoothing: the state of the art. J Forecast 2006 Sep 21;4(1):1-28. [doi: 10.1002/for.3980040103]
11. LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. In: Arbib MA, editor. The Handbook of Brain Theory and Neural Networks. Thousand Oaks, CA: The MIT Press; 1995:255-258.
12. Wibawa AP, Utama AB, Elmunsyah H, Pujianto U, Dwiyanto FA, Hernandez L. Time-series analysis with smoothed convolutional neural network. J Big Data 2022;9(1):44 [FREE Full text] [doi: 10.1186/s40537-022-00599-y] [Medline: 35495076]
13. Lea C, Vidal R, Reiter A, Hager GD. Temporal convolutional networks: a unified approach to action segmentation. In: Proceedings of the 2016 International Conference on Computer Vision. 2016 Presented at: ECCV '16; October 8-10, 2016; Amsterdam, The Netherlands p. 47-54 URL: https://link.springer.com/chapter/10.1007/978-3-319-49409-8_7 [doi: 10.1007/978-3-319-49409-8_7]
14. Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. Neural Comput 1989 Jun;1(2):270-280. [doi: 10.1162/neco.1989.1.2.270]
15. Ehrhardt J, Wilms M. Autoencoders and variational autoencoders in medical image analysis. In: Burgos N, Svoboda D, editors. Biomedical Image Synthesis and Simulation: Methods and Applications. New York, NY: Academic Press; 2022:129-162.
16. Wiles O, Albuquerque I, Gowal S. Discovering bugs in vision models using off-the-shelf image generation and captioning. arXiv Preprint posted online August 18, 2022 [FREE Full text]
17. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA p. 6000-6010 URL: https://dl.acm.org/doi/10.5555/3295222.3295349
18. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online October 11, 2018 [FREE Full text]
19. Wang A, Singh A, Michael J, Hill F, Levy O, Bowman SR. GLUE: a multi-task benchmark and analysis platform for natural language understanding. arXiv Preprint posted online April 20, 2018. [doi: 10.18653/v1/w18-5446]
20. Wang A, Pruksachatkun Y, Nangia N, Singh A, Michael J, Hill F, et al. Superglue: a stickier benchmark for general-purpose language understanding systems. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019 Presented at: NIPS '19; December 8-14, 2019; Vancouver, BC p. 3266-3280 URL: https://dl.acm.org/doi/10.5555/3454287.3454581
21. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020 Oct 22;63(11):139-144. [doi: 10.1145/3422622]
22. Yan W, Wang Y, Gu S, Huang L, Yan F, Xia L, et al. The domain shift problem of medical image segmentation and vendor-adaptation by Unet-GAN. arXiv Preprint posted online October 30, 2019 [FREE Full text]
23. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective. ACM Comput Surv 2017 Dec 06;50(6):1-45. [doi: 10.1145/3136625]

XSL•FO
RenderX

24. Taylor SJ, Letham B. Forecasting at scale. Am Stat 2018 Apr 24;72(1):37-45. [doi: 10.1080/00031305.2017.1380080]
25. Chang KM, Liu SH. Gaussian noise filtering from ECG by wiener filter and ensemble empirical mode decomposition. J Sign Process Syst 2010 Feb 17;64(2):249-264. [doi: 10.1007/s11265-009-0447-z]
26. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv Preprint posted online December 20, 2013 [FREE Full text]
27. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. arXiv Preprint posted online October 7, 2016 [FREE Full text] [doi: 10.1109/iccv.2017.74]
28. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. arXiv Preprint posted online June 19, 2020 [FREE Full text]
29. Wornow M, Xu Y, Thapa R, Patel B, Steinberg E, Fleming S, et al. The shaky foundations of clinical foundation models: a survey of large language models and foundation models for EMRs. arXiv. Posted online March 22, 2023 2025. [doi: 10.48550/arXiv.2303.12961]
30. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. arXiv Preprint posted online August 16, 2021 [FREE Full text]
31. Pang C, Jiang X, Kalluri KS, Spotnitz M, Chen R, Perotte A, et al. CEHR-BERT: incorporating temporal information from structured EHR data to improve prediction tasks. arXiv Preprint posted online November 23, 2023 [FREE Full text]
32. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor ai: predicting clinical events via recurrent neural networks. arXiv Preprint posted online November 18, 2015 [FREE Full text]
33. Prakash P, Chilukuri S, Ranade N, Viswanathan S. RareBERT: transformer architecture for rare disease patient identification using administrative claims. Proc AAAI Conf Artif Intell 2021 May 18;35(1):453-460. [doi: 10.1609/aaai.v35i1.16122]
34. Li F, Jin Y, Liu W, Rawat BP, Cai P, Yu H. Fine-tuning bidirectional encoder representations from transformers (BERT)-based models on large-scale electronic health record notes: an empirical study. JMIR Med Inform 2019 Sep 12;7(3):e14830 [FREE Full text] [doi: 10.2196/14830] [Medline: 31516126]
35. Sushil M, Ludwig D, Butte AJ, Rudrapatna VA. Developing a general-purpose clinical language inference model from a large corpus of clinical notes. arXiv Preprint posted online October 12, 2022 [FREE Full text]
36. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med 2022 Dec 26;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
37. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
38. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthcare 2021 Oct 15;3(1):1-23. [doi: 10.1145/3458754]
39. Shin HC, Zhang Y, Bakhturina E, Puri R, Patwary M, Shoeybi M, et al. BioMegatron: larger biomedical domain language model. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020 Presented at: EMNLP '20; November 16-20, 2020; Virtual Event p. 4700-4706 URL: https://aclanthology.org/2020.emnlp-main.379.pdf [doi: 10.18653/v1/2020.emnlp-main.379]
40. Huang K, Altosaar J, Ranganath R. Clinicalbert: modeling clinical notes and predicting hospital readmission. arXiv Preprint posted online April 10, 2019 [FREE Full text]
41. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. J Am Med Inform Assoc 2016 Nov;23(6):1143-1149 [FREE Full text] [doi: 10.1093/jamia/ocw021] [Medline: 27079506]
42. Chiu CC, Wu CM, Chien TN, Kao LJ, Li C, Chu CM. Integrating structured and unstructured EHR data for predicting mortality by machine learning and latent Dirichlet allocation method. Int J Environ Res Public Health 2023 Feb 28;20(5):4340 [FREE Full text] [doi: 10.3390/ijerph20054340] [Medline: 36901354]
43. Sohn K, Lee H, Yan X. Learning structured output representation using deep conditional generative models. In: Proceedings of the 29th International Conference on Neural Information Processing Systems. 2015 Presented at: NIPS '15; December 7-12, 2015; Montreal, QC p. 3483-3491 URL: https://dl.acm.org/doi/10.5555/2969442.2969628
44. Uzunova H, Ehrhardt J, Handels H. Memory-efficient GAN-based domain translation of high resolution 3D medical images. Comput Med Imaging Graph 2020 Dec;86:101801. [doi: 10.1016/j.compmedimag.2020.101801] [Medline: 33130418]
45. Zhang J, Yin P. Multivariate time series missing data imputation using recurrent denoising autoencoder. In: Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine. 2019 Presented at: BIBM '19; November 18-21, 2019; San Diego, CA p. 760-764 URL: https://ieeexplore.ieee.org/document/8982996 [doi: 10.1109/bibm47256.2019.8982996]
46. Suo Q, Yao L, Xun G, Sun J, Zhang A. Recurrent imputation for multivariate time series with missing values. In: Proceedings of the 2019 IEEE International Conference on Healthcare Informatics. 2019 Presented at: ICHI '19; June 10-13, 2019; Xi'an, China p. 1-3 URL: https://ieeexplore.ieee.org/document/8904638 [doi: 10.1109/ichi.2019.8904638]
47. Lipton ZC, Kale DC, Wetzel R. Modeling missing data in clinical time series with RNNs. Mach Learn Healthc 2016;56(56):2016-2070.

XSL•FO
RenderX

48. Shi Z, Wang S, Yue L, Pang L, Zuo X, Zuo W, et al. Deep dynamic imputation of clinical time series for mortality prediction. Inf Sci 2021 Nov;579:607-622. [doi: 10.1016/j.ins.2021.08.016]
49. Sangeetha M, Senthil Kumaran M. RETRACTED ARTICLE: deep learning-based data imputation on time-variant data using recurrent neural network. Soft Comput 2020 Feb 24;24(17):13369-13380. [doi: 10.1007/S00500-020-04755-5]
50. Yoon J, Zame WR, van der Schaar M. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. IEEE Trans Biomed Eng 2019 May;66(5):1477-1490. [doi: 10.1109/tbme.2018.2874712]
51. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. Sci Rep 2018 Apr 17;8(1):6085 [FREE Full text] [doi: 10.1038/s41598-018-24271-9] [Medline: 29666385]
52. Zaman MA, Du D. A stochastic multivariate irregularly sampled time series imputation method for electronic health records. BioMedInformatics 2021 Nov 16;1(3):166-181. [doi: 10.3390/biomedinformatics1030011]
53. Mulyadi AW, Jun E, Suk HI. Uncertainty-aware variational-recurrent imputation network for clinical time series. IEEE Trans Cybern 2022 Sep;52(9):9684-9694. [doi: 10.1109/tcyb.2021.3053599]
54. Cao W, Wang D, Li J, Zhou H, Li L, Li Y. Brits: bidirectional recurrent imputation for time series. In: Proceedings of the 2018 Conference on Advances in Neural Information Processing Systems. 2018 Presented at: NeurIPS '18; December 3-8, 2018; Montréal, QC p. 12-19 URL: https://papers.nips.cc/paper_files/paper/2018/hash/734e6bfcd358e25ac1db0a4241b95651-Abstract.html
55. Gupta M, Beheshti R. Time-series imputation and prediction with bi-directional generative adversarial networks. arXiv Preprint posted online September 18, 2020 [FREE Full text]
56. Gupta M, Phan TL, Bunnell HT, Beheshti R. Concurrent imputation and prediction on EHR data using bi-directional GANs: Bi-GANs for EHR imputation and prediction. In: Proceedings of the 12th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2021 Presented at: BCB '21; August 1-4, 2021; Gainesville, FL p. 1-9 URL: https://dl.acm.org/doi/10.1145/3459930.3469512 [doi: 10.1145/3459930.3469512]
57. Luo Y, Cai X, Zhang Y, Xu J. Multivariate time series imputation with generative adversarial networks. In: Proceedings of the 2018 Conference on Advances in Neural Information Processing Systems. 2018 Presented at: NeurIPS '18; December 3-8, 2018; Montréal, QC p. 2018 URL: https://papers.nips.cc/paper_files/paper/2018/hash/96b9bff013acedfb1d140579e2fbeb63-Abstract.html
58. Li J, Cairns BJ, Li J, Zhu T. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. NPJ Digit Med 2023 May 27;6(1):98 [FREE Full text] [doi: 10.1038/s41746-023-00834-7] [Medline: 37244963]
59. Guo A, Smith S, Khan YM, Langabeer II JR, Foraker RE. Application of a time-series deep learning model to predict cardiac dysrhythmias in electronic health records. PLoS One 2021 Sep 13;16(9):e0239007 [FREE Full text] [doi: 10.1371/journal.pone.0239007] [Medline: 34516567]
60. Jawadi Z, He R, Srivastava PK, Fonarow GC, Khalil SO, Krishnan S, et al. Predicting in-hospital mortality among patients admitted with a diagnosis of heart failure: a machine learning approach. ESC Heart Fail 2024 Oct 18;11(5):2490-2498 [FREE Full text] [doi: 10.1002/ehf2.14796] [Medline: 38637959]
61. Nigo M, Rasmy L, Mao B, Kannadath BS, Xie Z, Zhi D. Deep learning model for personalized prediction of positive MRSA culture using time-series electronic health records. Nat Commun 2024 Mar 06;15(1):2036 [FREE Full text] [doi: 10.1038/s41467-024-46211-0] [Medline: 38448409]
62. Yu R, Zheng Y, Zhang R, Jiang Y, Poon CC. Using a multi-task recurrent neural network with attention mechanisms to predict hospital mortality of patients. IEEE J Biomed Health Inform 2020 Feb;24(2):486-492. [doi: 10.1109/jbhi.2019.2916667]
63. Ge W, Huh JW, Park YR, Lee JH, Kim YH, Turchin A. An interpretable ICU mortality prediction model based on logistic regression and recurrent neural networks with LSTM units. AMIA Annu Symp Proc 2018;2018:460-469 [FREE Full text] [Medline: 30815086]
64. Caicedo-Torres W, Gutierrez J. ISeeU: visually interpretable deep learning for mortality prediction inside the ICU. J Biomed Inform 2019 Oct;98:103269 [FREE Full text] [doi: 10.1016/j.jbi.2019.103269] [Medline: 31430550]
65. Zhang J, Kowsari K, Harrison JH, Lobo JM, Barnes LE. Patient2Vec: a personalized interpretable deep representation of the longitudinal electronic health record. IEEE Access 2018;6:65333-65346. [doi: 10.1109/access.2018.2875677]
66. Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Predicting hospital readmission via cost-sensitive deep learning. IEEE/ACM Trans Comput Biol Bioinform 2018;15(6):1968-1978. [doi: 10.1109/TCBB.2018.2827029] [Medline: 29993930]
67. Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One 2019 Jul 8;14(7):e0218942 [FREE Full text] [doi: 10.1371/journal.pone.0218942] [Medline: 31283759]
68. Barbieri S, Kemp J, Perez-Concha O, Kotwal S, Gallagher M, Ritchie A, et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. Sci Rep 2020 Jan 24;10(1):1111 [FREE Full text] [doi: 10.1038/s41598-020-58053-z] [Medline: 31980704]
69. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. arXiv Preprint posted online November 11, 2015 [FREE Full text] [doi: 10.5860/choice.189890]

XSL•FO
RenderX

70. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: a deep dynamic memory model for predictive medicine. In: Proceedings of the 20th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. 2016 Presented at: PAKDD '16; April 19-22, 2016; Auckland, New Zealand p. 30-41 URL: https://link.springer.com/chapter/10.1007/978-3-319-31750-2_3 [doi: 10.1007/978-3-319-31750-2_3]
71. Yang Y, Zheng X, Ji C. Disease prediction model based on BiLSTM and attention mechanism. In: Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine. 2019 Presented at: BIBM '19; November 18-21, 2019; San Diego, CA p. 1141-1148 URL: https://ieeexplore.ieee.org/document/8983378 [doi: 10.1109/bibm47256.2019.8983378]
72. Guo W, Ge W, Cui L, Li H, Kong L. An interpretable disease onset predictive model using crossover attention mechanism from electronic health records. IEEE Access 2019;7:134236-134244. [doi: 10.1109/access.2019.2928579]
73. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak 2020 Oct 29;20(1):280 [FREE Full text] [doi: 10.1186/s12911-020-01297-6] [Medline: 33121479]
74. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. J Biomed Inform 2018 Jul;83:112-134 [FREE Full text] [doi: 10.1016/j.jbi.2018.04.007] [Medline: 29879470]
75. Mohammadi R, Jain S, Agboola S, Palacholla R, Kamarthi S, Wallace BC. Learning to identify patients at risk of uncontrolled hypertension using electronic health records data. AMIA Jt Summits Transl Sci Proc 2019;2019:533-542 [FREE Full text] [Medline: 31259008]
76. Ye X, Zeng QT, Facelli JC, Brixner DI, Conway M, Bray BE. Predicting optimal hypertension treatment pathways using recurrent neural networks. Int J Med Inform 2020 Jul;139:104122 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104122] [Medline: 32339929]
77. Lim B, Alaa A, van der Schaar M. Forecasting treatment responses over time using recurrent marginal structural networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018 Presented at: NIPS '18; December 3-8, 2018; Montréal, QC p. 7494-7504 URL: https://dl.acm.org/doi/10.5555/3327757.3327849
78. Lee C, Yoon J, Van Der Schaar M. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Trans Biomed Eng 2020 Jan;67(1):122-133. [doi: 10.1109/tbme.2019.2909027]
79. Kwon BC, Choi MJ, Kim JT, Choi E, Kim YB, Kwon S, et al. RetainVis: visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. IEEE Trans Visual Comput Graphics 2019 Jan;25(1):299-309. [doi: 10.1109/tvcg.2018.2865027]
80. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016 Presented at: NIPS '16; December 5-10, 2016; Barcelona, Spain p. 55 URL: https://dl.acm.org/doi/10.5555/3157382.3157490
81. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. J Biomed Inform 2017 May;69:218-229 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.001] [Medline: 28410981]
82. Kraljevic Z, Bean D, Shek A, Bendayan R, Hemingway H, Yeung JA, et al. Foresight-a generative pretrained transformer for modelling of patient timelines using electronic health records: a retrospective modelling study. Lancet Digit Health 2024 Apr;6(4):e281-e290 [FREE Full text] [doi: 10.1016/S2589-7500(24)00025-6] [Medline: 38519155]
83. Xu Y, Xu S, Ramprassad M, Tumanov A, Zhang C. TransEHR: self-supervised transformer for clinical time series data. In: Proceedings of the 3rd Conference on Machine Learning for Health Symposium. 2023 Presented at: ML4H '23; December 10, 2023; New Orleans, LA p. 623-635 URL: https://proceedings.mlr.press/v225/xu23a.html
84. Harerimana G, Kim JW, Jang B. A multi-headed transformer approach for predicting the patient's clinical time-series variables from charted vital signs. IEEE Access 2022;10:105993-106004. [doi: 10.1109/access.2022.3211334]
85. Sheikhalishahi S, Bhattacharyya A, Celi LA, Osmani V. An interpretable deep learning model for time-series electronic health records: case study of delirium prediction in critical care. Artif Intell Med 2023 Oct;144:102659 [FREE Full text] [doi: 10.1016/j.artmed.2023.102659] [Medline: 37783541]
86. Antikainen E, Linnosmaa J, Umer A, Oksala N, Eskola M, van Gils M, et al. Transformers for cardiac patient mortality risk prediction from heterogeneous electronic health records. Sci Rep 2023 Mar 02;13(1):3517 [FREE Full text] [doi: 10.1038/s41598-023-30657-1] [Medline: 36864069]
87. Li Y, Rao S, Solares JR, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: transformer for electronic health records. Sci Rep 2020 Apr 28;10(1):7155 [FREE Full text] [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
88. Labach A, Pokhrel A, Huang XS, Zuberi S, Yi S, Volkovs M, et al. Duett: dual event time transformer for electronic health records. arXiv Preprint posted online April 25, 2023 [FREE Full text]
89. Bica I, Alaa AM, Jordon J, van der Schaar M. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. arXiv Preprint posted online February 10, 2020 [FREE Full text]
90. Berrevoets J, Curth A, Bica I, McKinney E, van der Schaar M. Disentangled counterfactual recurrent networks for treatment effect inference over time. arXiv Preprint posted online December 7, 2021 [FREE Full text]
91. Grecov P, Bandara K, Bergmeir C, Ackermann K, Campbell S, Scott D, et al. Causal inference using global forecasting models for counterfactual prediction. In: Proceedings of the 25th Pacific-Asia Conference on Advances in Knowledge

XSL•FO
RenderX

Discovery and Data Mining. 2021 Presented at: PAKDD '21; May 11-14, 2021; Virtual Event p. 282-294 URL: https://link.springer.com/chapter/10.1007/978-3-030-75765-6_23 [doi: 10.1007/978-3-030-75765-6_23]
92. Wang Z, Samsten I, Papapetrou P. Counterfactual explanations for survival prediction of cardiovascular ICU patients. In: Proceedings of the 19th International Conference on Artificial Intelligence in Medicine. 2021 Presented at: AIME '21; June 15-18, 2021; Virtual Event p. 338-348 URL: https://link.springer.com/chapter/10.1007/978-3-030-77211-6_38 [doi: 10.1007/978-3-030-77211-6_38]
93. Li R, Hu S, Lu M, Utsumi Y, Chakraborty P, Sow DM. G-Net: a recurrent network approach to G-computation for counterfactual prediction under a dynamic treatment regime. In: Proceedings of the 2021 Conference on Machine Learning for Health. 2021 Presented at: PMLR '21; December 4, 2021; Virtual Event p. 282-299 URL: https://proceedings.mlr.press/v158/li21a.html
94. Ge Q, Hu Z, Zhang K, Li S, Lin W, Jin L, et al. Recurrent neural reinforcement learning for counterfactual evaluation of public health interventions on the spread of COVID-19 in the world. medRxiv Preprint posted online July 10, 2020 [FREE Full text] [doi: 10.1101/2020.07.08.20149146]
95. Chu J, Dong W, Wang J, He K, Huang Z. Treatment effect prediction with adversarial deep learning using electronic health records. BMC Med Inform Decis Mak 2020 Dec 14;20(Suppl 4):139 [FREE Full text] [doi: 10.1186/s12911-020-01151-9] [Medline: 33317502]
96. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
97. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics 2021 Dec 16;22(4):291-300 [FREE Full text] [doi: 10.2174/1389202922666210705124359] [Medline: 35273459]
98. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. Int J Med Inform 2022 Mar;159:104679. [doi: 10.1016/j.ijmedinf.2021.104679] [Medline: 34990939]
99. Cerqueira V, Torgo L, Soares C. A case study comparing machine learning with statistical methods for time series forecasting: size matters. J Intell Inf Syst 2022 May 16;59(2):415-433. [doi: 10.1007/S10844-022-00713-9]
100. Makridakis S, Spiliotis E, Assimakopoulos V, Semenoglou AA, Mulder G, Nikolopoulos K. Statistical, machine learning and deep learning forecasting methods: comparisons and ways forward. J Oper Res Soc 2022 Sep 05;74(3):840-859. [doi: 10.1080/01605682.2022.2118629]
101. Bonnheim NB, Wang L, Lazar AA, Chachad R, Zhou J, Guo X, et al. Deep-learning-based biomarker of spinal cartilage endplate health using ultra-short echo time magnetic resonance imaging. Quant Imaging Med Surg 2023 May 01;13(5):2807-2821 [FREE Full text] [doi: 10.21037/qims-22-729] [Medline: 37179932]
102. Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J Cogn Neurosci 2010 Dec;22(12):2677-2684 [FREE Full text] [doi: 10.1162/jocn.2009.21407] [Medline: 19929323]
103. Reddy BK, Delen D. Predicting hospital readmission for lupus patients: an RNN-LSTM-based deep-learning methodology. Comput Biol Med 2018 Oct 01;101:199-209. [doi: 10.1016/j.compbiomed.2018.08.029] [Medline: 30195164]
104. Hashir M, Sawhney R. Towards unstructured mortality prediction with free-text clinical notes. J Biomed Inform 2020 Aug;108:103489 [FREE Full text] [doi: 10.1016/j.jbi.2020.103489] [Medline: 32592755]
105. Ehrenstein V, Kharrazi H, Lehmann H, Taylor CO. Obtaining data from electronic health records. In: Ehrenstein V, Kharrazi H, Lehmann H, Taylor CO, editors. Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd edition. New York, NY: Agency for Healthcare Research and Quality (US); 2019:52-79.
106. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016 Sep;23(5):1007-1015 [FREE Full text] [doi: 10.1093/jamia/ocv180] [Medline: 26911811]
107. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://dl.acm.org/doi/10.5555/3295222.3295230
108. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. arXiv Preprint posted online February 16, 2016 [FREE Full text] [doi: 10.1145/2939672.2939778]
109. Jain R, Jangra A, Saha S, Jatowt A. A survey on medical document summarization. arXiv Preprint posted online December 3, 2022 [FREE Full text]
110. Zhang M, Zhou G, Yu W, Huang N, Liu W. A comprehensive survey of abstractive text summarization based on deep learning. Comput Intell Neurosci 2022 Aug 1;2022:7132226-7132221 [FREE Full text] [doi: 10.1155/2022/7132226] [Medline: 35958768]
111. Molenaar S, Maas L, Burriel V, Dalpiaz F, Brinkkemper S. Medical dialogue summarization for automated reporting in healthcare. In: Proceedings of the 2020 Conference on Advanced Information Systems Engineering Workshops. 2020 Presented at: CAiSE '20; June 8-12, 2020; Grenoble, France p. 76-88 URL: https://link.springer.com/chapter/10.1007/978-3-030-49165-9_7 [doi: 10.1007/978-3-030-49165-9_7]

XSL•FO
RenderX

112. Deelman E, Vahi K, Juve G, Rynge M, Callaghan S, Maechling PJ, et al. Pegasus, a workflow management system for science automation. Future Gener Comput Syst 2015 May;46:17-35. [doi: 10.1016/j.future.2014.10.008]
113. Lee M, Cimino J, Zhu HR, Sable C, Shanker V, Ely J, et al. Beyond information retrieval--medical question answering. AMIA Annu Symp Proc 2006;2006:469-473 [FREE Full text] [Medline: 17238385]
114. Terol RM, Martínez-Barco P, Palomar M. A knowledge based method for the medical question answering problem. Comput Biol Med 2007 Oct;37(10):1511-1521. [doi: 10.1016/j.compbiomed.2007.01.013] [Medline: 17374369]
115. Wornow M, Xu Y, Thapa R, Patel B, Steinberg E, Fleming S, et al. The shaky foundations of large language models and foundation models for electronic health records. NPJ Digit Med 2023 Jul 29;6(1):135 [FREE Full text] [doi: 10.1038/s41746-023-00879-8] [Medline: 37516790]
116. Jia J, Liu H, Gong NZ. 10 security and privacy problems in large foundation models. In: Wang C, Iyengar SS, Sun K, editors. AI Embedded Assurance for Cyber Systems. Cham, Switzerland: Springer; 2023:139-159.
117. Bhavya S, Pillai AS. Prediction models in healthcare using deep learning. In: Proceedings of the 11th International Conference on Soft Computing and Pattern Recognition. 2019 Presented at: SoCPaR '19; December 13-15, 2019; Hyderabad, India p. 195-204 URL: https://link.springer.com/chapter/10.1007/978-3-030-49345-5_21 [doi: 10.1007/978-3-030-49345-5_21]
118. Masoudi S, Harmon SA, Mehralivand S, Walker SM, Raviprakash H, Bagci U, et al. Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. J Med Imag 2021 Jan 1;8(01):52. [doi: 10.1117/1.jmi.8.1.010901]
119. Venkatesh B, Anuradha J. A review of feature selection and its methods. Cybern Inf Technol 2019;19(1):26. [doi: 10.2478/cait-2019-0001]
120. Hamghalam M, Simpson AL. Medical image synthesis via conditional GANs: application to segmenting brain tumours. Comput Biol Med 2024 Mar;170:107982 [FREE Full text] [doi: 10.1016/j.compbiomed.2024.107982] [Medline: 38266466]
121. Yoon JS, Zhang C, Suk HI, Guo J, Li X. Sadm: sequence-aware diffusion model for longitudinal medical image generation. In: Proceedings of the 28th International Conference on Information Processing in Medical Imaging. 2023 Presented at: IPMI '23; June 18-23, 2023; Bariloche, Argentina p. 388-400 URL: https://link.springer.com/chapter/10.1007/978-3-031-34048-2_30 [doi: 10.1007/978-3-031-34048-2_30]
122. Xia T, Chartsias A, Wang C, Tsaftaris SA, Alzheimer's Disease Neuroimaging Initiative. Learning to synthesise the ageing brain without longitudinal data. Med Image Anal 2021 Oct;73:102169. [doi: 10.1016/j.media.2021.102169] [Medline: 34311421]
123. Elazab A, Wang C, Gardezi SJ, Bai H, Hu Q, Wang T, et al. GP-GAN: brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR images. Neural Netw 2020 Dec;132:321-332. [doi: 10.1016/j.neunet.2020.09.004] [Medline: 32977277]
124. Peng L, Lin L, Lin Y, Chen YW, Mo Z, Vlasova RM, et al. Longitudinal prediction of infant MR images with multi-contrast perceptual adversarial learning. Front Neurosci 2021 Sep 9;15:653213 [FREE Full text] [doi: 10.3389/fnins.2021.653213] [Medline: 34566556]
125. Ning Z, Zhang Y, Pan Y, Zhong T, Liu M, Shen D. LDGAN: longitudinal-diagnostic generative adversarial network for disease progression prediction with missing structural MRI. In: Proceedings of the 11th International Workshop on Machine Learning in Medical Imaging. 2020 Presented at: MICCAI '20; October 4, 2020; Lima, Peru p. 170-179 URL: https://link.springer.com/chapter/10.1007/978-3-030-59861-7_18 [doi: 10.1007/978-3-030-59861-7_18]
126. Devika K, Mahapatra D, Subramanian R, Oruganti VR. Outlier-based autism detection using longitudinal structural MRI. IEEE Access 2022;10:27794-27808. [doi: 10.1109/access.2022.3157613]
127. To MS, Sarno IG, Chong C, Jenkinson M, Carneiro G. Self-supervised lesion change detection and localisation in longitudinal multiple sclerosis brain imaging. In: Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. 2021 Presented at: MICCAI '21; September 27-October 1, 2021; Strasbourg, France p. 670-680 URL: https://link.springer.com/chapter/10.1007/978-3-030-87234-2_63 [doi: 10.1007/978-3-030-87234-2_63]
128. Disch N, Zimmerer D. A holistic vision: modeling patient trajectories in longitudinal medical imaging. In: Proceedings of the 37th Conference on Neural Information Processing Systems. 2023 Presented at: NeurIPS '23; December 10-16, 2023; Virtual Event p. 1-6 URL: https://openreview.net/pdf?id=3FBY4NdDVU
129. Yang F, Meng R, Cho H, Wu G, Kim WH. Disentangled sequential graph autoencoder for preclinical Alzheimer's disease characterizations from ADNI study. In: Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. 2021 Presented at: MICCAI '21; September 27October 1, 2021; Strasbourg, France p. 2021-2072 URL: https://link.springer.com/chapter/10.1007/978-3-030-87196-3_34 [doi: 10.1007/978-3-030-87196-3_34]
130. Martí-Juan G, Lorenzi M, Piella G, Alzheimer's Disease Neuroimaging Initiative. MC-RVAE: multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. Neuroimage 2023 Mar;268:119892 [FREE Full text] [doi: 10.1016/j.neuroimage.2023.119892] [Medline: 36682509]
131. Baur C, Denner S, Wiestler B, Navab N, Albarqouni S. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study. Med Image Anal 2021 Apr;69:101952. [doi: 10.1016/j.media.2020.101952] [Medline: 33454602]

XSL•FO
RenderX

132. Hong J, Kang SK, Alberts I, Lu J, Sznitman R, Lee JS, Alzheimer's Disease Neuroimaging Initiative. Image-level trajectory inference of tau pathology using variational autoencoder for Flortaucipir PET. Eur J Nucl Med Mol Imaging 2022 Jul 28;49(9):3061-3072 [FREE Full text] [doi: 10.1007/s00259-021-05662-z] [Medline: 35226120]
133. Berchuck SI, Mukherjee S, Medeiros FA. Estimating rates of progression and predicting future visual fields in glaucoma using a deep variational autoencoder. Sci Rep 2019;9(1):18113. [doi: 10.1101/652487]
134. Xiao N, Qiang Y, Zhao Z, Zhao J, Lian J. Tumour growth prediction of follow‐up lung cancer via conditional recurrent variational autoencoder. IET Image Process 2021 Feb 18;14(15):3975-3981. [doi: 10.1049/iet-ipr.2020.0496]
135. Kmetzsch V, Becker E, Saracino D, Rinaldi D, Camuzat A, Le Ber I, et al. Disease progression score estimation from multimodal imaging and MicroRNA data using supervised variational autoencoders. IEEE J Biomed Health Inform 2022 Dec;26(12):6024-6035. [doi: 10.1109/jbhi.2022.3208517]
136. Sauty B, Durrleman S. Longitudinal Variational Autoencoders learn a Riemannian progression model for imaging data. In: Proceedings of the 2022 Conference on Proceedings of Machine Learning Research. 2022 Presented at: GeoMedIA '22; July 17-23, 2022; Baltimore, ML p. 1-5 URL: https://inria.hal.science/hal-03800499/document [doi: 10.1017/9781108891530.014]
137. Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Huang C, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. JAMA Netw Open 2020 Mar 02;3(3):e200772 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0772] [Medline: 32163165]
138. Zhang L, Lu L, Wang X, Zhu RM, Bagheri M, Summers RM, et al. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. IEEE Trans Med Imaging 2020 Apr;39(4):1114-1126. [doi: 10.1109/tmi.2019.2943841]
139. Yu X, Yang Q, Tang Y, Gao R, Bao S, Cai LY, et al. Deep conditional generative model for longitudinal single-slice abdominal computed tomography harmonization. J Med Imag 2024 Mar 1;11(02):63. [doi: 10.1117/1.jmi.11.2.024008]
140. Zhang Y, Hu X, Sapkota N, Shi Y, Chen DZ. Unsupervised feature clustering improves contrastive representation learning for medical image segmentation. In: Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine. 2022 Presented at: BIBM '22; December 6-8, 2022; Las Vegas, NV p. 1820-1823 URL: https://ieeexplore.ieee.org/document/9995129 [doi: 10.1109/bibm55620.2022.9995129]
141. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin PM. Cardiac segmentation with strong anatomical guarantees. IEEE Trans Med Imaging 2020 Nov 27;39(11):3703-3713 [FREE Full text] [doi: 10.1109/TMI.2020.3003240] [Medline: 32746116]
142. Mahapatra D, Antony B, Sedai S, Garnavi R. Deformable medical image registration using generative adversarial networks. In: Proceedings of the 15th International Symposium on Biomedical Imaging. 2018 Presented at: ISBI '18; April 4-7, 2018; Washington, DC p. 1449-1453 URL: https://ieeexplore.ieee.org/document/8363845 [doi: 10.1109/isbi.2018.8363845]
143. Qiao J, Lai Q, Li Y, Lan T, Yu C, Wang X. A GAN based multi-contrast modalities medical image registration approach. In: Proceedings of the 2020 IEEE International Conference on Image Processing. 2020 Presented at: ICIP '20; October 25-28, 2020; Abu Dhabi, United Arab Emirates p. 3000-3004 URL: https://ieeexplore.ieee.org/document/9191024 [doi: 10.1109/icip40778.2020.9191024]
144. Tanner C, Ozdemir F, Profanter R, Vishnevsky V, Konukoglu E, Goksel O. Generative adversarial networks for MR-CT deformable image registration. arXiv Preprint posted online July 19, 2018 [FREE Full text]
145. Yang H, Qian P, Fan C. An indirect multimodal image registration and completion method guided by image synthesis. Comput Math Methods Med 2020 Jun 30;2020:2684851-2684810 [FREE Full text] [doi: 10.1155/2020/2684851] [Medline: 32670390]
146. Kim B, Han I, Ye JC. DiffuseMorph: unsupervised deformable image registration using diffusion model. arXiv Preprint posted online December 9, 2021 [FREE Full text] [doi: 10.1007/978-3-031-19821-2_20]
147. Zhang F, Wells WM, O'Donnell LJ, Zhang F. Deep diffusion MRI registration (DDMReg): a deep learning method for diffusion MRI registration. IEEE Trans Med Imaging 2022 Jun;41(6):1454-1467. [doi: 10.1109/tmi.2021.3139507]
148. Quellec G, Cazuguel G, Cochener B, Lamard M. Multiple-instance learning for medical image and video analysis. IEEE Rev Biomed Eng 2017;10:213-234. [doi: 10.1109/rbme.2017.2651164]
149. Gupta S, Keshari A, Das S. RV-GAN: recurrent GAN for unconditional video generation. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2022 Presented at: CVPRW '22; June 19-20, 2022; New Orleans, LA p. 2023-2032 URL: https://ieeexplore.ieee.org/document/9857297 [doi: 10.1109/cvprw56347.2022.00220]
150. Celard P, Seara Vieira A, Sorribes-Fdez JM, Romero R, Lorenzo IE, Borrajo DL. Study on synthetic video generation of embryo development. In: Proceedings of the 18th International Conference on Hybrid Artificial Intelligent Systems. 2023 Presented at: HAIS '23; September 5-7, 2023; Salamanca, Spain p. 623-634 URL: https://link.springer.com/chapter/10.1007/978-3-031-40725-3_53 [doi: 10.1007/978-3-031-40725-3_53]
151. Ferroukhi M, Ouahabi A, Attari M, Habchi Y, Taleb-Ahmed A. Medical video coding based on 2nd-generation wavelets: performance evaluation. Electronics 2019 Jan 14;8(1):88. [doi: 10.3390/electronics8010088]
152. Li C, Liu H, Liu Y, Feng B, Li W, Liu X, et al. Endora: video generation models as endoscopy simulators. arXiv Preprint posted online March 17, 2024 [FREE Full text] [doi: 10.1007/978-3-031-72089-5_22]

XSL•FO
RenderX

153. Reynaud H, Vlontzos A, Dombrowski M, Gilligan LC, Beqiri A, Leeson P, et al. D'ARTAGNAN: counterfactual video generation. In: Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention. 2022 Presented at: MICCAI '22; September 18-22, 2022; Singapore, Singapore p. 599-609 URL: https://link.springer.com/chapter/10.1007/978-3-031-16452-1_57 [doi: 10.1007/978-3-031-16452-1_57]
154. Suganthi K, Dombrowski M. Review of medical image synthesis using GAN techniques. ITM Web Conf 2021 Mar 17;37:01005. [doi: 10.1051/itmconf/20213701005]
155. Walker J, Murphy P. A review of glottal waveform analysis. In: Stylianou Y, Faundez-Zanuy M, Esposito A, editors. Progress in Nonlinear Speech Processing. Cham, Switzerland: Springer; 2007:1-21.
156. Neifar N, Ben-Hamadou A, Mdhaffar A, Jmaiel M, Freisleben B. Leveraging statistical shape priors in GAN-based ECG synthesis. IEEE Access 2024;12:36002-36015. [doi: 10.1109/access.2024.3373724]
157. Hartmann KG, Schirrmeister RT, Ball T. EEG-GAN: generative adversarial networks for electroencephalograhic (EEG) brain signals. arXiv Preprint posted online June 5, 2018 [FREE Full text]
158. Abdelfattah SM, Abdelrahman GM, Wang M. Augmenting the size of EEG datasets using generative adversarial networks. In: Proceedings of the 2018 International Joint Conference on Neural Networks. 2018 Presented at: IJCNN '18; July 8-13, 2018; Rio de Janeiro, Brazil p. 1-6 URL: https://ieeexplore.ieee.org/document/8489727 [doi: 10.1109/ijcnn.2018.8489727]
159. Panwar S, Rad P, Jung TP, Huang Y. Modeling EEG data distribution with a Wasserstein generative adversarial network to predict RSVP events. IEEE Trans Neural Syst Rehabil Eng 2020 Aug;28(8):1720-1730. [doi: 10.1109/tnsre.2020.3006180]
160. Luo Y, Lu BL. EEG data augmentation for emotion recognition using a conditional Wasserstein GAN. In: Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018 Presented at: EMBC '18; July 18-21, 2018; Honolulu, HI p. 2535-2538 URL: https://ieeexplore.ieee.org/document/8512865 [doi: 10.1109/embc.2018.8512865]
161. Luo TJ, Fan Y, Chen L, Guo G, Zhou C. EEG signal reconstruction using a generative adversarial network with Wasserstein distance and temporal-spatial-frequency loss. Front Neuroinform 2020 Apr 30;14:15 [FREE Full text] [doi: 10.3389/fninf.2020.00015] [Medline: 32425763]
162. Adib E, Fernandez AS, Afghah F, Prevost JJ. Synthetic ECG signal generation using probabilistic diffusion models. IEEE Access 2023;11:75818-75828. [doi: 10.1109/access.2023.3296542]
163. Zhu F, Ye F, Fu Y, Liu Q, Shen B. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. Sci Rep 2019 May 01;9(1):6734 [FREE Full text] [doi: 10.1038/s41598-019-42516-z] [Medline: 31043666]
164. Zhou B, Liu S, Hooi B, Cheng X, Ye J. BeatGAN: anomalous rhythm detection using adversarially generated time series. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019 Presented at: IJCAI '19; August 10-16, 2019; Macao, China p. 4433-4439 URL: https://dl.acm.org/doi/10.5555/3367471.3367658 [doi: 10.24963/ijcai.2019/616]
165. Shin DH, Park RC, Chung K. Decision boundary-based anomaly detection model using improved AnoGAN from ECG data. IEEE Access 2020;8:108664-108674. [doi: 10.1109/access.2020.3000638]
166. Qin J, Gao F, Wang Z, Wong DC, Zhao Z, Relton SD, et al. A novel temporal generative adversarial network for electrocardiography anomaly detection. Artif Intell Med 2023 Feb;136:102489. [doi: 10.1016/j.artmed.2023.102489] [Medline: 36710067]
167. Truong ND, Kuhlmann L, Bonyadi MR, Querlioz D, Zhou L, Kavehei O. Epileptic seizure forecasting with generative adversarial networks. IEEE Access 2019;7:143999-144009. [doi: 10.1109/access.2019.2944691]
168. You S, Cho BH, Yook S, Kim JY, Shon YM, Seo DW, et al. Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network. Comput Methods Programs Biomed 2020 Sep;193:105472. [doi: 10.1016/j.cmpb.2020.105472] [Medline: 32344271]
169. Xiao C, Gou Z, Tai W, Zhang K, Zhou F. Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023 Presented at: KDD '23; August 6-10, 2023; Long Beach, CA p. 51 URL: https://dl.acm.org/doi/10.1145/3580305.3599391 [doi: 10.1145/3580305.3599391]
170. Wang Y, Zhao S, Jiang H, Li S, Luo B, Li T, et al. DiffMDD: a diffusion-based deep learning framework for MDD diagnosis using EEG. IEEE Trans Neural Syst Rehabil Eng 2024;32:728-738. [doi: 10.1109/tnsre.2024.3360465]
171. Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A. Wavenet: a generative model for raw audio. arXiv Preprint posted online September 12, 2016 [FREE Full text]
172. Baevski A, Zhou Y, Mohamed A, Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020 Presented at: NIPS'20; December 6-12, 2020; Vancouver, BC p. 12449-12460 URL: https://dl.acm.org/doi/abs/10.5555/3495724.3496768
173. Krishna G, Tran C, Han Y, Carnahan M, Tewfik AH. Speech synthesis using EEG. In: Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. 2020 Presented at: ICASSP '20; May 4-8, 2020; Barcelona, Spain p. 1235-1238 URL: https://ieeexplore.ieee.org/document/9053340 [doi: 10.1109/icassp40776.2020.9053340]
174. Hill BL, Rakocz N, Rudas Á, Chiang JN, Wang S, Hofer I, et al. Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning. Sci Rep 2021 Aug 03;11(1):15755 [FREE Full text] [doi: 10.1038/s41598-021-94913-y] [Medline: 34344934]

XSL•FO
RenderX

175. Lim B, Arık SO, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. arXiv Preprint posted online December 19, 2019 [FREE Full text] [doi: 10.1016/j.ijforecast.2021.03.012]
176. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 2010 Jan 26;20(1):110-121 [FREE Full text] [doi: 10.1101/gr.097857.109] [Medline: 19858363]
177. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003 Jul 01;31(13):3812-3814 [FREE Full text] [doi: 10.1093/nar/gkg509] [Medline: 12824425]
178. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 2016 Dec 24;48(12):1581-1586. [doi: 10.1038/ng.3703] [Medline: 27776117]
179. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun 2020 Nov 20;11(1):5918 [FREE Full text] [doi: 10.1038/s41467-020-19669-x] [Medline: 33219223]
180. Brandes N, Goldman G, Wang CH, Ye CJ, Ntranos V. Genome-wide prediction of disease variant effects with a deep protein language model. Nat Genet 2023 Sep 10;55(9):1512-1522 [FREE Full text] [doi: 10.1038/s41588-023-01465-0] [Medline: 37563329]
181. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. Nature 2021 Nov 27;599(7883):91-95. [doi: 10.1038/s41586-021-04043-8] [Medline: 34707284]
182. Li C, Zhi D, Wang K, Liu X. MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. Genome Med 2022 Oct 08;14(1):115 [FREE Full text] [doi: 10.1186/s13073-022-01120-z] [Medline: 36209109]
183. Qi H, Zhang H, Zhao Y, Chen C, Long JJ, Chung WK, et al. MVP predicts the pathogenicity of missense variants by deep learning. Nat Commun 2021 Jan 21;12(1):510 [FREE Full text] [doi: 10.1038/s41467-020-20847-0] [Medline: 33479230]
184. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet 2018 Aug 23;50(8):1161-1170 [FREE Full text] [doi: 10.1038/s41588-018-0167-z] [Medline: 30038395]
185. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 2015 Mar 01;31(5):761-763 [FREE Full text] [doi: 10.1093/bioinformatics/btu703] [Medline: 25338716]
186. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 2014 Apr;32(4):381-386 [FREE Full text] [doi: 10.1038/nbt.2859] [Medline: 24658644]
187. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 2017 Oct;14(10):979-982 [FREE Full text] [doi: 10.1038/nmeth.4402] [Medline: 28825705]
188. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 2020 Dec;38(12):1408-1414. [doi: 10.1038/s41587-020-0591-3] [Medline: 32747759]
189. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci U S A 2014 Dec 30;111(52):E5643-E5650 [FREE Full text] [doi: 10.1073/pnas.1408993111] [Medline: 25512504]
190. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 2019 Mar 19;20(1):59 [FREE Full text] [doi: 10.1186/s13059-019-1663-x] [Medline: 30890159]
191. Sha Y, Qiu Y, Zhou P, Nie Q. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. Nat Mach Intell 2024 Nov 30;6(1):25-39 [FREE Full text] [doi: 10.1038/s42256-023-00763-w] [Medline: 38274364]
192. Huang Y, Zaas AK, Rao A, Dobigeon N, Woolf PJ, Veldman T, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. PLoS Genet 2011 Aug;7(8):e1002234 [FREE Full text] [doi: 10.1371/journal.pgen.1002234] [Medline: 21901105]
193. Baranzini SE, Mousavi P, Rio J, Caillier SJ, Stillman A, Villoslada P, et al. Transcription-based prediction of response to IFNbeta using supervised computational methods. PLoS Biol 2005 Jan;3(1):e2 [FREE Full text] [doi: 10.1371/journal.pbio.0030002] [Medline: 15630474]
194. Palii CG, Cheng Q, Gillespie MA, Shannon P, Mazurczyk M, Napolitani G, et al. Single-cell proteomics reveal that quantitative changes in co-expressed lineage-specific transcription factors determine cell fate. Cell Stem Cell 2019 May 02;24(5):812-20.e5 [FREE Full text] [doi: 10.1016/j.stem.2019.02.006] [Medline: 30880026]
195. Jindal K, Adil MT, Yamaguchi N, Yang X, Wang HC, Kamimoto K, et al. Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. Nat Biotechnol 2024 Jun;42(6):946-959 [FREE Full text] [doi: 10.1038/s41587-023-01931-4] [Medline: 37749269]
196. Sriyudthsak K, Shiraishi F, Hirai MY. Mathematical modeling and dynamic simulation of metabolic reaction systems using metabolome time series data. Front Mol Biosci 2016;3:15 [FREE Full text] [doi: 10.3389/fmolb.2016.00015] [Medline: 27200361]

XSL•FO
RenderX

197. Erdős B, Westerhuis JA, Adriaens ME, O'Donovan SD, Xie R, Singh-Povel CM, et al. Analysis of high-dimensional metabolomics data with complex temporal dynamics using RM-ASCA. PLoS Comput Biol 2023 Jun;19(6):e1011221 [FREE Full text] [doi: 10.1371/journal.pcbi.1011221] [Medline: 37352364]
198. Ahmed KT, Sun J, Cheng S, Yong J, Zhang W. Multi-omics data integration by generative adversarial network. Bioinformatics 2021 Dec 22;38(1):179-186 [FREE Full text] [doi: 10.1093/bioinformatics/btab608] [Medline: 34415323]
199. Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol 2022 Oct;40(10):1458-1466 [FREE Full text] [doi: 10.1038/s41587-022-01284-4] [Medline: 35501393]
200. Ashuach T, Gabitto MI, Koodli RV, Saldi GA, Jordan MI, Yosef N. MultiVI: deep generative model for the integration of multimodal data. Nat Methods 2023 Aug;20(8):1222-1231 [FREE Full text] [doi: 10.1038/s41592-023-01909-9] [Medline: 37386189]
201. Yeo GH, Saksena SD, Gifford DK. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. Nat Commun 2021 May 28;12(1):3222 [FREE Full text] [doi: 10.1038/s41467-021-23518-w] [Medline: 34050150]
202. Yang Z, Wen J, Abdulkadir A, Cui Y, Erus G, Mamourian E, et al. Gene-SGAN: discovering disease subtypes with imaging and genetic signatures via multi-view weakly-supervised deep clustering. Nat Commun 2024 Jan 08;15(1):354 [FREE Full text] [doi: 10.1038/s41467-023-44271-2] [Medline: 38191573]
203. Guttà C, Morhard C, Rehm M. Applying a GAN-based classifier to improve transcriptome-based prognostication in breast cancer. PLoS Comput Biol 2023 Apr;19(4):e1011035 [FREE Full text] [doi: 10.1371/journal.pcbi.1011035] [Medline: 37011102]
204. Ning Y, Teixayavong S, Shang Y, Savulescu J, Nagaraj V, Miao D, et al. Generative artificial intelligence in healthcare: ethical considerations and assessment checklist. arXiv Preprint posted online November 2, 2023 [FREE Full text]
205. Health insurance portability and accountability act of 1996. U.S. Department of Health and Human Services. URL: https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996 [accessed 2024-04-29]
206. General data protection regulation (GDPR). Intersoft Consulting. 2018. URL: https://www.consilium.europa.eu/en/policies/data-protection-regulation/#:~:text=The%20GDPR%20lists%20the%20rights,his%20or%20her%20personal%20data [accessed 2024-04-29]
207. Jadon A, Kumar S. Leveraging generative AI models for synthetic data generation in healthcare: balancing research and privacy. In: Proceedings of the 2023 International Conference on Smart Applications, Communications and Networking. 2023 Presented at: SmartNets '23; July 25-27, 2023; Istanbul, Turkiye p. 1-4 URL: https://ieeexplore.ieee.org/document/10215825 [doi: 10.1109/smartnets58706.2023.10215825]
208. Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Inf Fusion 2023 Aug;96:156-191. [doi: 10.1016/j.inffus.2023.03.008]
209. McGee RW. ChatGPT and copyright infringement: an exploratory study. SSRN Journal Preprint posted online September 20, 2023 [FREE Full text] [doi: 10.2139/ssrn.4578430]
210. ThankGod Chinonso E. The impact of ChatGPT on privacy and data protection laws. SSRN Journal Preprint posted online October 13, 2023 [FREE Full text] [doi: 10.2139/ssrn.4574016]
211. Miller JA, Aldosari M, Saeed F, Barna NH, Rana S, Arpinar IB, et al. A survey of deep learning and foundation models for time series forecasting. arXiv Preprint posted online January 25, 2024 [FREE Full text]
212. Duan W, He X, Zhou L, Thiele L, Rao H. Combating distribution shift for accurate time series forecasting via hypernetworks. In: Proceedings of the 28th International Conference on Parallel and Distributed Systems. 2022 Presented at: ICPADS '22; January 10-12, 2022; Nanjing, China p. 900-907 URL: https://ieeexplore.ieee.org/document/10077946 [doi: 10.1109/icpads56603.2022.00121]
213. Fadler M, Legner C. WHO owns data in the enterprise? Rethinking data ownership in times of big data and analytics. In: Proceedings of the 2020 European Conference on Information System. 2022 Presented at: ECIS '20; June 15-17, 2020; Marrakech, Morocco p. 2020 URL: https://aisel.aisnet.org/ecis2020_rp/207/
214. Wang T, Zhang Y, Qi S, Zhao R, Xia Z, Weng J. Security and privacy on generative data in AIGC: a survey. ACM Comput Surv 2024 Dec 10;57(4):1-34. [doi: 10.1145/3703626]
215. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics 2021 Sep 15;22(1):122 [FREE Full text] [doi: 10.1186/s12910-021-00687-3] [Medline: 34525993]
216. Xue F, Fu Y, Zhou W, Zheng Z, You Y. To repeat or not to repeat: insights from scaling LLM under token-crisis. arXiv Preprint posted online May 22, 2023 [FREE Full text] [doi: 10.48550/arXiv.2305.13230]
217. Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, et al. SNPs for a universal individual identification panel. Hum Genet 2010 Mar 24;127(3):315-324. [doi: 10.1007/s00439-009-0771-1] [Medline: 19937056]
218. Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual identification panel. Hum Genet 2007 May 27;121(3-4):305-317. [doi: 10.1007/s00439-007-0342-2] [Medline: 17333283]
219. Cases currently under investigation. U.S. Department of Health and Human Services. 2023. URL: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf [accessed 2024-04-29]

XSL•FO
RenderX

220. Greshake K, Abdelnabi S, Mishra S, Endres C, Holz T, Fritz M. Not what you've signed up for: compromising real-world LLM-integrated applications with indirect prompt injection. In: Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security. 2023 Presented at: CCS '23; November 30, 2023; Copenhagen, Denmark p. 79-90 URL: https://dl.acm.org/doi/10.1145/3605764.3623985 [doi: 10.1145/3605764.3623985]
221. Kumar A, Agarwal C, Srinivas S, Li AJ, Feizi S, Lakkaraju H. Certifying LLM safety against adversarial prompting. arXiv Preprint posted online September 6, 2023 [FREE Full text]
222. Salem A, Paverd A, Köpf B. Maatphor: automated variant analysis for prompt injection attacks. arXiv Preprint posted online December 12, 2023 [FREE Full text]
223. Piet J, Alrashed M, Sitawarin C, Chen S, Wei Z, Sun E, et al. Jatmo: prompt injection defense by task-specific finetuning. arXiv Preprint posted online December 29, 2023 [FREE Full text] [doi: 10.1007/978-3-031-70879-4_6]
224. Chen S, Piet J, Sitawarin C, Wagner D. StruQ: defending against prompt injection with structured queries. arXiv Preprint posted online February 9, 2024 [FREE Full text]
225. Yu P, Xu H, Hu X, Deng C. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. Healthcare 2023 Oct 20;11(20):2776. [doi: 10.3390/healthcare11202776]
226. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023 Jul 06;6(1):22. [doi: 10.1038/S41746-023-00873-0]
227. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
228. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024 Jan;6(1):e12-e22 [FREE Full text] [doi: 10.1016/S2589-7500(23)00225-X] [Medline: 38123252]
229. Bautista YJ, Theran C, Aló R, Lima V. Health disparities through generative AI models: a comparison study using a domain specific large language model. In: Proceedings of the 2023 Conference on Future Technologies. 2023 Presented at: FTC '23; November 2-3, 2023; Virtual Event p. 220-232 URL: https://link.springer.com/chapter/10.1007/978-3-031-47454-5_17 [doi: 10.1007/978-3-031-47454-5_17]
230. Nazer LH, Zatarah R, Waldrip S, Ke JX, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2023 Jun 22;2(6):e0000278. [doi: 10.1371/journal.pdig.0000278] [Medline: 37347721]
231. Wójcik MA. Foundation models in healthcare: opportunities, biases and regulatory prospects in Europe. In: Proceedings of the 11th International Conference on Electronic Government and the Information Systems Perspective. 2022 Presented at: EGOVIS '22; August 22-24, 2022; Vienna, Austria p. 32-46 URL: https://link.springer.com/chapter/10.1007/978-3-031-12673-4_3 [doi: 10.1007/978-3-031-12673-4_3]
232. Zhang P, Kamel Boulos MN. Generative AI in medicine and healthcare: promises, opportunities and challenges. Future Internet 2023 Aug 24;15(9):286. [doi: 10.3390/fi15090286]
233. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 2023 Feb;15(2):e35179 [FREE Full text] [doi: 10.7759/cureus.35179] [Medline: 36811129]
234. Franzoni V. From black box to glass box: advancing transparency in artificial intelligence systems for ethical and trustworthy AI. In: Proceedings of the 2023 Conference on Computational Science and Its Applications. 2023 Presented at: ICCSA '23; July 3-6, 2023; Athens, Greece p. 118-130 URL: https://link.springer.com/chapter/10.1007/978-3-031-37114-1_9 [doi: 10.1007/978-3-031-37114-1_9]
235. Sun J, Liao QV, Muller M, Agarwal M, Houde S, Talamadupula K, et al. Investigating explainability of generative AI for code through scenario-based design. In: Proceedings of the 27th International Conference on Intelligent User Interfaces. 2022 Presented at: IUI '22; March 22-25, 2022; Helsinki, Finland p. 212-228 URL: https://dl.acm.org/doi/fullHtml/10.1145/3490099.3511119 [doi: 10.1145/3490099.3511119]
236. Dunn AG, Shih I, Ayre J, Spallek H. What generative AI means for trust in health communications. J Commun Healthc 2023 Dec 03;16(4):385-388. [doi: 10.1080/17538068.2023.2277489] [Medline: 37921509]
237. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011-2022). Comput Methods Programs Biomed 2022 Nov;226:107161. [doi: 10.1016/j.cmpb.2022.107161] [Medline: 36228495]
238. Berry P, Bharadiya VM, Kumar S. Generative AI in healthcare: a trustworthy approach. Netguru. URL: https://tinyurl.com/yv5urjkp [accessed 2024-04-29]

Abbreviations

AD: Alzheimer disease
AI: artificial intelligence
ARIMA: autoregressive integrated moving average
BERT: Bidirectional Encoder Representations From Transformers
CLaM: clinical language model

XSL•FO
RenderX

CNN: convolutional neural network
CT: computed tomography
DDPM: denoising diffusion probabilistic model
DL: deep learning
ECG: electrocardiogram
EEG: electroencephalogram
EHR: electronic health record
ESM1b: Evolutionary Scale Modeling 1b
ETS: exponential smoothing
FEMR: foundation model for electronic medical records
FM: foundation model
GAN: generative adversarial network
GRU: gated recurrent unit
HIPAA: Health Insurance Portability and Accountability Act
LSTM: long short-term memory
ML: machine learning
MRI: magnetic resonance imaging
NLP: natural language processing
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RNN: recurrent neural network
VAE: variational autoencoder
VUS: variant of uncertain significance

Edited by G Eysenbach; submitted 30.04.24; peer-reviewed by J Dunn, X Zhang, MD Nguyen, Y Yang; comments to author 18.07.24; revised version received 08.08.24; accepted 15.11.24; published 10.03.25.

Please cite as:
He R, Sarwal V, Qiu X, Zhuang Y, Zhang L, Liu Y, Chiang J
Generative AI Models in Time-Varying Biomedical Data: Scoping Review
J Med Internet Res 2025;27:e59792
URL: https://www.jmir.org/2025/1/e59792
doi:10.2196/59792
PMID:

©Rosemary He, Varuni Sarwal, Xinru Qiu, Yongwen Zhuang, Le Zhang, Yue Liu, Jeffrey Chiang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

News and Perspective

The Controversial Push for New Brain and Neurorights

Simon Spichak

(J Med Internet Res 2025;27:e72270) doi:10.2196/72270

In August 2023, the Supreme Court of Chile issued a unanimous decision ordering the US-based brain-computer interface company Emotiv to erase the brain data it had collected on a former Chilean Senator [1].

The results of the court case were reported as a landmark decision for "neurorights"—enacting the country's recent constitutional provision to protect people's brain data and mental privacy from being sold or manipulated by neurotechnology. Rafael Yuste, a codirector of Columbia University's Neurotechnology Center and cofounder of the Neurorights Foundation, took part in crafting the provision.

The Neurorights Foundation is instrumental in the push for new human rights and legislation to protect people's brains from neurotechnology.



Rafael Yuste delivering a lecture in Chile.

However, these efforts have proven controversial among scientists, legal scholars, and advocacy groups [2-4]. Yuste told JMIR Publications (JMIR) that some of the criticism is "poorly informed" and that neurorights are not adequately covered by existing rights or data privacy legislation. He added that some of the critics may receive funding from the industry.

(The experts who spoke with JMIR disclosed that they had not received funding from neurotechnology companies.)

J Carlos Lara Gálvez, executive director of Derechos Digitales, a nonprofit organization based in Chile that advocates for digital rights across Latin America, told JMIR that legal experts in the country are split on the neurorights provision, with some skeptical that "such legal and constitutional changes are needed at all."

Danielle Zaror Miralles, a law and technology professor at the University of Chile Law School, explained to JMIR that the constitutional provisions are based on the colonial idea that "there were no legal protections" for brain data in Chile before this legislation.

Why are some researchers advocating for "neurorights"?

Neurotechnology encompasses any device that can directly read, measure, or manipulate brain data. Some invasive forms of neurotechnology, like deep brain stimulation (DBS), are implanted directly into the brain to treat the symptoms of Parkinson disease, depression, obsessive-compulsive disorder, and other diseases [5].

These forms of neurotechnology are regulated as medical devices, while consumer-facing devices are largely unregulated in countries like the United States.

"I became very concerned with this [neurorights] through our own experiments with mice," Yuste told JMIR. "About a decade ago, we were able to decode the activity of the visual cortex and manipulate it selectively using laser-based technology."

While invasive brain implants or "mind-reading" devices aren't hitting the consumer market anytime soon, many noninvasive consumer electronics—like Muse or Emotiv—are already on the market. These devices measure electroencephalographic (EEG) brain data to help measure and potentially improve sleep and mental health. Researchers are also developing "mind-controlled" drones using consumer-grade EEG headsets [6]. In 2023, Apple also received approval on its patent for AirPod-like earbuds with built-in electrodes that could theoretically measure and collect EEGs [7].

Marcello Ienca, a professor who studies the ethics of artificial intelligence (AI) and neuroscience at the Technical University of Munich and the Swiss Federal Institute of Technology Lausanne, also believes this raises new problems. "The potential for intervening into brain function, either by recording activity from the brain, or for writing into the brain using neurostimulation techniques, opens challenges that are substantial from a human rights perspective, and that can be well understood as human rights challenges," he told JMIR.

In response to the development of this technology, Yuste and other experts believe that new "neurorights" relating to brain data are needed [8,9].

1. **The right to mental privacy.** To prevent neural activity from being deciphered without consent.
2. **The right to mental identity.** To protect against technology altering the self, consciousness, or personality.
3. **The right to free will.** To prevent neurotechnology from tampering with decision-making abilities.
4. **Fair access to mental augmentation.**
5. **Protection from bias embedded in neurotechnology algorithms.**

Other experts are more skeptical of neurorights. "While there is rough consensus on the potential risks of certain technologies that interact with the human body or the brain, there is no consensus on the notion that those risks are certain or well-defined, that those risks are current and actual instead of potential and speculative," Gálvez said. "Some of the most publicly visible support comes accompanied by highly exaggerated claims about the capabilities of these technologies."

XSL•FO
RenderX

Gálvez is concerned that the debate on neurorights displaces attention and discussion from other fundamental rights and data protection regulations.



Danielle Zaror Miralles, law and technology professor at the University of Chile Law School.

So, are new human rights needed to protect the brain?

The Neurorights Foundation analyzed seven existing international human rights treaties and determined that they included insufficient protections for brain data, concluding that amendments or new rights were needed to protect brain data [10].

Nita Farahany, a legal scholar and bioscience professor at Duke University who wrote *The Battle for Your Brain*, also thinks that new rights are needed. She argues for a broader approach through the lens of "cognitive liberty," which would protect self-determination, freedom of thought, and mental privacy. Farahany told JMIR that cognitive liberty could provide broad protections against technologies that can predict cognitive states, even if they don't directly measure neural data.

"If the biometric data is being used to make inferences about a person's mental state, it deserves a higher level of protection," she said.

Ienca told JMIR that he's agnostic about the right approach to protecting these rights—whether reinterpreting existing rights or introducing new ones.

Several legal experts told JMIR they don't think this new framework is needed.

Susie Alegre, a human rights lawyer, author of the book *Freedom to Think*, and a senior fellow at the Center for International Governance Innovation, told JMIR that adding new rights or amending existing ones doesn't make sense. "The big problem is that when you say we need new rights, then you are firstly saying the rights we have don't cover this, and therefore, it's a kind of a free pass," she said. "Fundamentally, I don't understand advocating for new rights without testing the boundaries of existing rights in court."



Susie Alegre is a human rights lawyer, author of Freedom to Think, and senior fellow at the Center for International Governance Innovation. Photo credit: Tugce Nelson.

Jan Cristoph Bublitz, associate junior professor of criminal law at the University of Hamburg, also argued against the proposal for neurorights in a 2022 paper in the journal *Neuroethics* [11]. "The proposal tends to promote rights inflationism, is tainted by neuroexceptionalism and neuroessentialism, and lacks grounding in relevant scholarship," he argued.

In 2022, the UN Human Rights Council asked its Advisory Committee to study the issue [12].

Liz Throssell, a spokesperson from the UN Human Rights Office, said that the existing human rights framework does address the issues related to privacy and autonomy. Throssell told JMIR that there may be emerging issues outside the scope of legislation, but these will require further discussion.

"But the bottom line is that huge and potentially very consequential work needs to be done now in ensuring existing human rights standards are applied," Throssel said, explaining that adding or amending human rights takes a lot of time. "What is important is that we focus on implementing the human rights standards we have at our disposal."

Legislating neurorights around the world

The Neurorights Foundation recently surveyed 30 user agreements from direct-to-consumer neurotechnology companies. Clicking "I agree" on 29 of these agreements would give the neurotechnology company any rights to the data, which would allow them to sell it to a third party. Yuste called these user agreements "predatory."

"There are no rules; the company is taking everything," he said. "That's why we're working with many countries and international organizations to protect brain data." Ultimately, he believes that brain data should be protected as stringently as medical data.

Following the constitutional provisions in Chile, which Yuste and the Neurorights Foundation helped influence, other countries like Uruguay, Brazil, and Mexico have all either proposed or pushed through constitutional amendments that would provide similar protections.

However, experts, including Miralles, say these provisions are unnecessary as the constitutional protections in Latin America sufficiently encompass brain data and can restrict the use of medical neurotechnologies because of how the law is written [4].

Miralles and Pablo Contreras, a law professor at the Central University of Chile, argued that the neurorights provision was irrelevant to the Emotiv case, arguing that the company deleted the data in conjunction with existing data protection laws [13]. "Most constitutions on the continent recognize the right to privacy, and almost all countries have laws on personal data protection inspired by the GDPR," she said. "Needless to say, the entire hypothesis of this legislation is based on a science fiction fantasy idea."

Gálvez told JMIR that there were concerns the constitutional provision led to "altering the functions of public health authorities and the application of data protection law, introducing confusion to the legal landscape."

Meanwhile, in the United States, the Neurorights Foundation also worked alongside the Colorado Medical Society to sponsor bill HB24-1058 to protect brain activity and brain data.

Other states have enacted broader laws to protect other types of biometric data that technology companies collect. Illinois's

XSL•FO
RenderX

Biometric Information Privacy Act, for example, requires private entities collecting biometric data to inform people about what kind of data is being collected, how it's stored, and how it's used.

"I think that's a good start, but I would go further than it does," Farahany said, adding that cognitive biometric data should, for example, be kept on the device where it is collected and not sent to an external company.

In the European Union, Alegre explained that there are already laws and regulations protecting people's data from exploitation, but the problem involves enforcing and regulating these rules, adding that "the biggest problem is it's very hard to take a tech company to court."

JMIR News & Perspectives articles are written by journalists. They are not peer-reviewed but rigorously fact-checked. We invite freelance journalists to pitch us original ideas.

References

1. Asher-Schapiro A, Baptista D. Hands off my brainwaves: Latin America in race for 'neurorights'. Context News. 2023 Sep 12. URL: https://www.context.news/digital-rights/hands-off-my-brainwaves-latin-america-in-race-for-neurorights [accessed 2025-02-20]
2. Fins JJ. The unintended consequences of Chile's Neurorights Constitutional Reform: moving beyond negative rights to capabilities. Neuroethics 2022 Aug 24;15(3):26. [doi: 10.1007/s12152-022-09504-z]
3. Zúñiga-Fajuri A, Miranda LV, Miralles DZ, Venega RS. Chapter seven - neurorights in Chile: between neuroscience and legal science. In: Ienca M, Starke G, editors. Developments in Neuroethics and Bioethics. Amsterdam, Netherlands: Elsevier; 2021:165-179.
4. Ruiz S, Ramos-Vergara P, Concha R, Altermatt F, Von-Bernhardi R, Cuello M, et al. [Negative effects of the patients' rights law and neuro-rights bill in Chile] [Article in Spanish]. Rev Med Chil 2021 Mar;149(3):439-446 [FREE Full text] [doi: 10.4067/s0034-98872021000300439] [Medline: 34479323]
5. Krauss JK, Lipsman N, Aziz T, Boutet A, Brown P, Chang JW, et al. Technology of deep brain stimulation: current status and future directions. Nat Rev Neurol 2021 Feb;17(2):75-87 [FREE Full text] [doi: 10.1038/s41582-020-00426-z] [Medline: 33244188]
6. Prasath MS, Naveen R, Sivaraj G. Mind-controlled unmanned aerial vehicle (UAV) using brain–computer interface (BCI). In: Mohindru V, Singh Y, Bhatt R, Gupta AK, editors. Unmanned Aerial Vehicles for Internet of Things (IoT): Concepts, Techniques, and Applications. Austin, TX: Scrivener Publishing LLC; 2021.
7. Biosignal sensing device using dynamic selection of electrodes. United States Patent Application Publication. 2023 Jul 20. URL: https://patentimages.storage.googleapis.com/e2/4d/92/a20ceacf02d9db/US20230225659A1.pdf [accessed 2025-02-24]
8. Yuste R, Goering S, Arcas BAY, Bi G, Carmena JM, Carter A, et al. Four ethical priorities for neurotechnologies and AI. Nature 2017 Nov 08;551(7679):159-163 [FREE Full text] [doi: 10.1038/551159a] [Medline: 29120438]
9. Ienca M. On neurorights. Front Hum Neurosci 2021;15:701258 [FREE Full text] [doi: 10.3389/fnhum.2021.701258] [Medline: 34630057]
10. Genser J, Herrmann S, Yuste R. International human rights protection gaps in the age of neurotechnology. NeuroRights Foundation. 2022 May 06. URL: https://static1.squarespace.com/static/60e5c0c4c4f37276f4d458cf/t/6275130256dd5e2e11d4bd1b/1651839747023/Neurorights+Foundation+PUBLIC+Analysis+5.6.22.pdf [accessed 2025-02-24]
11. Bublitz JC. Novel neurorights: from nonsense to substance. Neuroethics 2022;15(1):7 [FREE Full text] [doi: 10.1007/s12152-022-09481-3] [Medline: 35154507]
12. Al-Jehani N, Alamro N, Angoh JG, Baek BS, Costas Trascasas M, Ganguly Das R, et al. Neurotechnology and human rights. United Nations Human Rights Council. 2024. URL: https://www.ohchr.org/en/hr-bodies/hrc/advisory-committee/neurotechnologies-and-human-rights [accessed 2025-02-24]
13. En defensa de los neuroderechos [Article in Spanish]. Kamanau. 2024. URL: https://defensaneuroderechos.org/ [accessed 2025-02-24]

Submitted 06.02.25; ###Reviewer names will be inserted here### published 25.02.25.

Please cite as:
Spichak S
The Controversial Push for New Brain and Neurorights
J Med Internet Res 2025;27:e72270
URL: https://www.jmir.org/2025/1/e72270/
doi:10.2196/72270
PMID:40000012

XSL•FO
RenderX

©JMIR Publications. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX